آمار و کاربرد آن در مدیریت

جلد اول

دکتر عادل آذر
دکتر منصور مؤمنی

آمار و کاربرد آن در مدیریت

جلد اول

دکتر عادل آذر
دکتر منصور مؤمنی

تهران
۱۳۸۴

سازمان مطالعه و تدوین کتب علوم انسانی دانشگاهها (سمت)

آذر، عادل، ۱۳۴۵ -
آمار و کاربرد آن در مدیریت / عادل آذر، منصور مؤمنی. -- تهران: سازمان مطالعه و تدوین کتب علوم انسانی دانشگاهها (سمت)، ۱۳۸۰.
۲ ج.: مصور، جدول، عکس، نمودار. -- («سمت»؛ ۱۸۹، ۲۷۴: مدیریت؛ ۱۲، ۱۸).
بها: ۱۴۵۰۰ ریال (ج. ۱). ISBN 964-459-189-5
بها: ۱۷۵۰۰ ریال (ج. ۲). ISBN 964-459-274-3
فهرستنویسی بر اساس اطلاعات فیپا.
پشت جلد به انگلیسی: Adel Azar. Statistics and Its Application in Management.
واژه‌نامه.
کتابنامه.
ج. ۲ (چاپ اول: تابستان ۱۳۷۷، چاپ هفتم: بهار ۱۳۸۳).
ج. ۱ (چاپ اول: ۱۳۷۵، چاپ یازدهم: ۱۳۸۴).
۱. آمار. ۲. مدیریت ـ روشهای آماری. ۳. مدیریت ـ آمار. الف. مؤمنی، منصور، گردآورنده. ب. سازمان مطالعه و تدوین کتب علوم انسانی دانشگاهها (سمت). ج. عنوان.
۲آ۴ف/HA۲۹/۵ ۰۰۱/۴۲۲
۱۳۷۷
کتابخانه ملی ایران م۷۷-۶۳۵۹

سازمان مطالعه و تدوین کتب علوم انسانی دانشگاهها (سمت)

آمار و کاربرد آن در مدیریت (جلد اول)
دکتر عادل آذر و دکتر منصور مؤمنی
چاپ اول: ۱۳۷۵
چاپ یازدهم: بهار ۱۳۸۴
تعداد: ۱۷۰۰۰
حروفچینی و لیتوگرافی: سمت
چاپ و صحافی: سازمان چاپ و انتشارات، وزارت فرهنگ و ارشاد اسلامی
قیمت: ۱۴۵۰۰ ریال. در این نوبت چاپ قیمت مذکور ثابت است و فروشندگان و عوامل توزیع مجاز به تغییر آن نیستند.
آدرس ساختمان مرکزی: تهران، بزرگراه جلال آل احمد، غرب پل یادگار امام (ره)، روبروی پمپ گاز، کدپستی ۱۴۶۳۶، تلفن ۲ـ۴۲۴۶۲۵۰.
www.samt.ac.ir info@samt.ac.ir

فهرست مطالب

عنوان	صفحه

عنوان	صفحه

مقدمه

آمار علمی جدید است که شیوه‌ها و روشهای آن در قرن اخیر تکوین و بسط یافته و با آنکه مدت زیادی از پیدایش آن نمی‌گذرد، بسرعت توسعه‌یافته؛ به طوری که قلمرو مفاهیم مورد استفاده در آن از علوم ریاضی فراتر رفته و به کارگیری آن نیز در سایر علوم رواج یافته است.

امروزه، بندرت می‌توان بدون استفاده از روشهای آماری اقدام به تفسیر، تبیین و تحلیل نتایج به دست آمده از تحقیقات و پژوهشهای علمی کرد. به این ترتیب، در قلمرو علم مدیریت نیز کمتر می‌توان بدون استفاده از روشهای آماری تحقیقی درخور توجه انجام داد؛ بنابراین پژوهشگران علم مدیریت نیز باید برای اجرای طرحهای تحقیقاتی خود در حد لزوم به این ابزار مجهز شده، برای برآوردن نیازهای ویژه خود از آن بهره‌مند شوند.

اگر اطلاعات را از مهمترین لوازم تصمیم‌گیری به شمار آوریم؛ علم آمار که با پردازش داده‌ها و تبدیل آنها به اطلاعات مورد نیاز، زمینهٔ اتخاذ تصمیم را فراهم می‌کند، یکی از مهمترین علوم محسوب خواهد شد. آنچه مسلم است هنر مدیران و کارشناسان در نحوهٔ استفاده از روشهای آماری و تحلیل اطلاعات به دست آمده تجلی می‌کند. این امر به حدی اهمیت دارد که امروزه اکثر مدیران و مسئولان سازمانها، از مشاوران آماری استفاده می‌کنند.

این کتاب، حاصل تلاشی برای معرفی علم آمار و نحوهٔ استفاده از آن در علم مدیریت است و هدف اصلی از تدوین آن، برآورده ساختن نیاز دانشجویان رشته‌های مدیریت و حسابداری به یادگیری مفاهیم و روشهای آماری در سطوح کارشناسی و کارشناسی ارشد است. در انتهای هر فصل علاوه بر مسائل، تعدادی سؤالات چند گزینه‌ای نیز آمده است که برای شرکت‌کنندگان در آزمونهای ورودی مقاطع مختلف تحصیلی مفید است؛ زیرا این مجموعه به طور عمده برگزیده‌ای از سؤالات آماری مورد استفاده در آزمونهای سالهای قبل (رشته‌های مدیریت و حسابداری) را دربرمی‌گیرد.

ضرورت استفادهٔ کاربران صنعت، و سازمانهای اداری و خدماتی از روشهای آماری، مؤلفان را بر آن داشت که مطالب کتاب را به گونه‌ای تدوین کنند که این عزیزان نیز بتوانند از آن بهره‌مند گردند. بدین منظور، جلد اول کتاب به معرفی آمار توصیفی و احتمالات اختصاص یافته و سعی شده است که مفاهیم آمار توصیفی جامعتر از سایر کتابهای فارسی موجود، ارائه شود. احتمالات نیز در حد نیاز مطرح شده‌اند و جلد دوم به «تحلیل آماری» اختصاص یافته است؛ به‌طوری که مجموع فصول دو جلد، دربرگیرندهٔ کلیهٔ سرفصلهای مصوب برای دروس «آمار و کاربرد آن در مدیریت (۱)» و «آمار و کاربرد آن در مدیریت (۲)» خواهد بود. در ضمن با توجه به تغییرات مفید و اصولی به عمل آمده در سرفصلهای مصوب شورای‌عالی برنامه‌ریزی وزارت فرهنگ و آموزش‌عالی در سال ۷۶ـ۱۳۷۵ توصیه می‌شود که همهٔ فصول جلد اوّل و فصل هجدهم از جلد دوّم تحت عنوان «فنون تصمیم‌گیری آماری» برای درس «آمار و کاربرد آن در مدیریت (۱)» تدریس شود. فصول ۹ ، ۱۰ ، ۱۱ ، ۱۳ ، ۱۵ و ۱۷ نیز برای تدریس در درس «آمارو کاربرد آن در مدیریت (۲)» توصیه می‌شود. ضمن آنکه با توجه به مطالب جلد دوّم، می‌توان آن را به مثابهٔ یک منبع مفید در درس «تحلیل آماری» در دوره «کارشناسی ارشد» مدیریت و حسابداری مورد استفاده قرار داد.

در چاپ دوّم کتاب، جلد اوّل مورد بازنگری کامل قرارگرفت و سعی شد بسیاری از نارساییها و اشتباهات کتاب در چاپ اوّل برطرف گردد. با وجود این و علی‌رغم تلاش مؤلفان برای تدوین صحیح قواعد و مفاهیم علم آمار، راه زیادی تا ارائه یک مجموعه کامل و بدون نقص وجود دارد. امید است که با عنایت حضرت حقّ و مساعدت مدرسین و دانشجویان بزرگوار، در چاپهای بعدی مجموعه کاملتر و بهتری ارائه گردد.

از آقای دکتر هاشم محلوجی عضو هیأت علمی دانشگاه صنعتی شریف که با دقت، متن این مجموعه را بررسی کرده و نکات علمی ارزنده‌ای را برای اصلاح آن یادآور شده‌اند، کمال تشکر و قدردانی می‌شود. همچنین از آقای علی اصغر پورعزّت که با دقت‌نظر و پیشنهادهای خود موجبات بهبود چاپ دوّم جلد اول را فراهم نموده‌اند، و همچنین از خانم سیمین عارفی که ویراستاری جلد اول و خانم عادله مشایخی که ویراستاری جلد دوم را تقبل نموده‌اند، تشکر و قدردانی می‌شود.

عادل آذر ـ منصور مؤمنی

فصل اول

کلیات

۱ـ۱ آمار و ضرورت آن در مدیریت

کلمهٔ آمار دو معنای کاملاً متفاوت دارد. در یک معنا عبارت است از روشی علمی که برای جمع‌آوری، تلخیص، تجزیه و تحلیل، تفسیر و به طور کلی مطالعه و بررسی مشاهدات به کار برده می‌شود. در معنای دیگر با عبارت «مشاهدات عددی» مترادف است؛ مثلاً وقتی به بعضی از گزارشهای بانک مرکزی یا مرکز آمار ایران اشاره می‌کنیم و آن را آمار می‌نامیم معنای آن این است که در آنها اعداد و ارقام موجود است. این معنای قدیمتر و معمولتر کلمهٔ آمار است. در ابتدا آمار به منظور فراهم نمودن اطلاعاتی که دولتمردان را در کارهای حکومتی یاری می‌کرد، جمع‌آوری می‌گردید. این نوع اطلاعات عددی از زمان ارسطو و در نوشته‌های وی با عنوان «مسائل ایالتی» به کار می‌رفته‌اند. در واقع می‌توان گفت که کلمه «statistics» و «state» از یک ریشه مشتق شده‌اند.[1] از دیرباز، در ممالک متمدن به منظور تحقیق در امور مالی، نظامی، نیروی انسانی و جمعیت‌شناسی، آمار و اطلاعات وسیعی گردآوری می‌کرده‌اند. اکثر افراد هنوز این تصور غلط را درباره آمار دارند که آن را منحصر به ستونهای عددی سرگیجه‌آور و گاهی تعدادی شکلهای مبهوت‌کننده می‌دانند؛ در حالی که نظریه و روشهای جدید آماری، از حدّ ساختن جدولهای اعداد و نمودارها بسیار فراتر رفته‌اند. نمایشهای عددی به صورت جنبه‌ای فرعی از آمار درآمده‌اند و امروزه تعداد بسیار کمی از آمارشناسان حرفه‌ای از طریق فراهم نمودن جدولها و نمودارها کسب درآمد می‌کنند.

آمار به عنوان یک علم، امروزه شامل مفاهیم و روشهایی است که در تمام

---

۱. باتاچاریا، گوری کی. و ریچارد جانسون؛ مفاهیم و روشهای آماری؛ ترجمه مرتضی ابن‌شهرآشوب و فتاح میکائیلی، ص ٦.

پژوهشهایی که مستلزم جمع‌آوری داده‌ها به وسیلهٔ فرایند آزمایش و مشاهده، و استنباط و نتیجه‌گیری به وسیلهٔ تجزیه و تحلیل آن داده‌ها هستند، اهمیت بسیار دارد.

به کارگیری روشهای آماری در علوم اجتماعی و رفتاری کمی قبل از جنگ دوم جهانی شروع شد و تعداد آمارگیریها در زمینه‌های مختلف افزایش یافت و ضرورت تفسیر اطلاعات مربوط به روانشناسی و تعلیم و تربیت آشنایی با علم آمار را اجتناب‌ناپذیر کرد. امروزه موفقیت در بسیاری از زمینه‌های علمی مانند علوم انسانی، پزشکی و فنی و مهندسی بدون داشتن اطلاعات لازم از علم آمار مشکل و گاهی غیرممکن شده است. در چند دههٔ اخیر، برای ساده کردن استفاده از فنون آماری، متخصصان آمار و کامپیوتر به فراهم نمودن بسته‌های کامپیوتری[1] همت گماشته‌اند. نرم‌افزارهای آماری[2]، بیش از آنکه فراگیری دانش آمار را آسان کنند، موجب گسترش و فراگیر شدن استفاده از آن گردیده‌اند؛ بدین ترتیب بسیاری از فنون آماری که کاربرد چندانی نداشتند یا مخصوص تعداد معدودی از رشته‌ها بودند، به بیشتر حوزه‌های علمی کشانده شدند. با وجود چنین ابزارهایی، همچنان ضرورت فراگیری علم آمار برای تحلیل داده‌ها بشدت احساس می‌شود.

از آنجا که آمار ابزار مناسبی برای تبدیل داده[3] به اطلاعات[4] است، استفاده از آن در دانش مدیریت به نحو وسیعی رایج گردیده است. امروزه کمتر سازمانی می‌توان یافت که از متخصصان آمار برای کمک به مدیریت سازمان استفاده نکند. به کارگیری برنامه‌های آماری و کامپیوتری به مدیران یاری‌رسانده تا درکمترین زمان ممکن اطلاعات لازم بسیاری را برای تصمیم‌گیری به دست آورند. بسیاری از داده‌های ذخیره‌شده که تاکنون برای مدیریت زاید محسوب می‌شد، امروزه به کمک بعضی از فنون آماری مثل سریهای زمانی ملاک تصمیم‌گیری و جهت‌گیری برای آینده سازمانها محسوب می‌شود.

استفاده از تکنیک شبیه‌سازی[5] در مدیریت و بخصوص مدیریت سیستمها شاخهٔ جدیدی از علم آمار است که در سالهای اخیر در میان مدیران طرفداران بسیاری پیدا کرده است. شبیه‌سازی تکنیکی است که اساس آن بر تولید اعداد تصادفی[6] از یک رشتهٔ

---

1. package

۲. از مهمترین نرم‌افزارهای آماری می‌توان از SAS، SPSS و Stat, Graph نام برد.

3. data
4. information
5. simulation
6. random numbers

توابع آماری نهفته است؛ بدین ترتیب در مدیریت تمایل به استفاده از داده‌های قطعی فراتر رفته و به داده‌های کاملاً تصادفی نیز کشانده شده است.

به کارگیری فنون پایه‌ای آمار در بررسی صحت و سقم فرضیات یکی دیگر از جنبه‌های کاربرد این علم در گسترهٔ مدیریت است. محقق مدیریتی امروزه بسیاری از فرضیات تحقیق خود را با فنون آماری محک می‌زند و صحت روابط متغیرهای آن را با رویکرد آماری بررسی می‌کند. اعتبار و پایایی همهٔ تحقیقهای پرسشنامه‌ای و مصاحبه‌ای با استفاده از فنون آماری سنجیده می‌شود. ضرورت تصمیم‌گیری منطقی و اساسی در دنیای متلاطم امروز، اهمیت علم آمار و فنون آن را بیش از پیش آشکار می‌کند و استفاده از آن را اجتناب ناپذیر می‌نماید.

۱ـ۲ جامعه و نمونه[1]

عموم افراد وقتی که از جامعه سخن می‌گویند به مجموعه‌ای از موجودات زنده و معمولاً مردم می‌اندیشند؛ در حالی که جامعه ممکن است از اشیا و عناصر بیجان تشکیل شده باشد؛ بدین ترتیب می‌توان جامعه را چنین تعریف کرد: «جامعه بزرگترین مجموعه از موجودات است که در یک زمان معین مطلوب ما قرار می‌گیرد»؛ برای مثال هرگاه اندازه گیری نمرهٔ همهٔ دانشجویان مدیریت که در دانشکده مدیریت دانشگاه تهران ثبت نام کرده‌اند، مطلوب ما باشد، جامعه‌ای خواهیم داشت و اگر فقط نمرهٔ دانشجویان رشتهٔ مدیریت بازرگانی همان دانشکده را در نظربگیریم، جامعهٔ دیگری خواهیم داشت.

اگر مقدار متغیری را برای هریک از عناصر جامعه اندازه گیری کنیم، جامعه مقادیر آن متغیر را به وجود آورده‌ایم؛ از این رو می‌توان، جامعه مقادیر متغیر را چنین تعریف کرد: «جامعهٔ مقادیر، بزرگترین مجموعه از مقادیر قابل اکتساب متغیرهای تصادفی است که در یک زمان معین مطلوب ما قرار می‌گیرد».

محدوده و فضای مطلوب ما جامعه‌های آماری را معین و مشخص می‌کند؛ بنابراین تعریف جامعهٔ آماری عبارت است از: «تعدادی از عناصر مطلوب مورد نظر که حداقل دارای یک صفت مشخصه باشند». صفت مشخصه، صفتی است که بین همهٔ عناصر جامعهٔ آماری مشترک و متمایزکننده جامعهٔ آماری از سایر جوامع باشد. جامعهٔ آماری به

1. population and sample

دو نوع تقسیم می‌شود: محدود و نامحدود. اگر جامعهٔ مقادیر از تعداد محدود و ثابتی تشکیل شود و پایان‌پذیر باشد آن را محدود و در غیراین صورت، وقتی که جامعه از یک ردیف بی‌انتهای مقادیر تشکیل شده باشد، آن را نامحدود گویند.

هر بخشی از جامعهٔ آماری را نمونه گویند؛ مثلاً فرض کنید جامعهٔ ما نمره‌های دانشجویان مدیریت دانشگاههای کشور باشد، هرگاه برای تجزیه و تحلیل، فقط مجموعه‌ای از این نمره‌ها را جمع‌آوری کنیم، یک نمونه از جامعهٔ آماری انتخاب کرده‌ایم؛ بنابراین تعریف نمونه عبارت است از: «تعداد محدودی از آحاد جامعهٔ آماری که بیان‌کنندهٔ ویژگیهای اصلی جامعه باشد». با این تعریف طبیعی است از جامعه‌های آماری می‌توان نمونه‌های متعددی انتخاب کرد.

اندازه‌گیری جامعه برای به دست آوردن برخی از شاخصهاست. این شاخصها، چنانچه با اندازه‌گیری تمامی عناصر جامعهٔ آماری (سرشماری) به دست آمده باشند آنها را پارامتر و اگر با استفاده از بخشی از جامعه به دست آمده باشند آنها را آماره[1] گویند؛ برای مثال اگر متوسط درآمد کارکنان دولت با استفاده از اندازه‌گیری درآمد تمامی کارکنان دولت محاسبه شود آن را پارامتر و اگر با استفاده از اندازه‌گیری درآمد نمونه‌ای از کارکنان به دست آید آن را آماره گویند. تفاوت در نوع و نیز کاربرد شاخصهای آماره و پارامتر تقسیم‌بندی ویژه‌ای را در علم آمار موجب شده و آن را به آمار توصیفی[2] و آمار استنباطی[3] تقسیم کرده است.

۱ـ۳ سیر تحول علم آمار

در قرن شانزدهم، پیشرفتهای شایان توجهی در توسعهٔ علم آمار به دست آمد. در انگلستان، جان گرانت یک رشته مطالعات شبه ریاضی در مورد آمار حیاتی، بیمه و اقتصاد انجام داد. مطالعات او را اندیشمندان دیگر مانند ویلیام پتی و ادموند هالی ادامه دادند. اولین مقالهٔ جامع را دربارهٔ تئوری کلی آمار ژاکوب برنولی نوشت؛ کسی که اساس قانون اعداد بزرگ را پی‌ریزی کرده بود. در سال ۱۷۳۳ توزیع نرمال برای اولین بار به وسیلهٔ دومواور[4] تعریف شد و سپس گوس[5] و لاپلاس آن را توسعه دادند. در آلمان دانشمند دیگری به نام ناپ آمار مرگ و میر را به طور جامع بررسی کرد و همچنین

1. statistic
2. descriptive statistics
3. inferential statistics
4. De Moivre
5. Gauss

لگزیس[1] روشی را توسعه داد که همانند تحلیل واریانس یک عامله[2] امروزی است.

در ربع آخر قرن نوزدهم، فرانسیس گالتون در انگلستان و جانشین بزرگ او کارل پیرسون با طرح مسائل ژنتیک، نظرات وابسته به رگرسیون و همبستگی را گسترش دادند، سپس کارل پیرسون و اسپیرمن این نظرات را کاملتر کردند و در مطالعات و تحقیقات علوم اجتماعی آن را به کار بردند. در همین راستا پیرسون آثار اشتباه در نمونه گیری را نیز به طور جامع مطالعه کرد.

مهمترین ویژگی علم آمار در اوایل قرن بیستم نسبت به قرون گذشته و نیز مسائل امروز و آینده آن به استنباط[3] یا تخمین[4] آماری مربوط می‌شود. تا اوایل قرن بیستم، نتیجهٔ محاسبات آماری، مقادیری قطعی تلقی می‌شدند و فرق میان نمونه و جامعهٔ آماری و شاخصهای مربوط به هرکدام به طور صریح تشخیص داده نمی‌شد. با رواج تدریجی استفاده از محاسبات آماری در تحقیقات علمی این نکتهٔ مورد توجه قرار گرفت که اطلاعات آماری غالباً مربوط به موارد محدود و معین است. واضح بود که اطلاعات نمونه قابل مقایسه با اطلاعات حاصل از کل جامعه نبود. این نکته و مسائل مربوط به آن موجب گسترش نظریهٔ نمونه گیری شد. در این راستا باب دیگری در آمار باز شد که به تخمین پارامترها به کمک آماره‌ها می‌پرداخت. در این باب فیشر و همزمان با او ایگون پیرسون و نیومن مفاهیم اساسی توزیع نمونه‌برداری[5] و حدود اطمینان[6] را در آمار استنباطی وارد کردند. مفهوم مهم دیگری که از رابطهٔ نمونه و جامعه برمی‌خیزد، بیان فرضهای آماری و روش آزمودن آنهاست. ایگون پیرسون و نیومن پس از انجام دادن یک رشتهٔ تحقیقات و نگارش تعدادی مقاله از ۱۹۲۸ به بعد اصول آزمون آماری فرضیات آزمایش را به طور کامل بیان کردند. پیشرفت مهم دیگر در این دوره تدوین اصول «طراحی آزمایشها» به وسیلهٔ فیشر است. طراحی آزمایشها روشی است برای تشکیل گروههای آزمایشی به صورتی که از مقایسهٔ نتایج مربوط به آنها بتوان آثار عوامل مختلف مؤثر در آزمایش را به تفکیک و به طور جداگانه اندازه گیری کرد.

فنون آماری‌ای که پیدایش آنها به اختصار شرح داده شد، همه در مواردی قابل استفاده هستند که داده‌های آماری خصوصیاتی مانند قابلیت اندازه گیری کمّی، پیوستگی

1. Lexis
2. one - way analysis of variance
3. inference
4. estimation
5. sampling distribution
6. confidence limits

تغییرات و فضای نمونه‌ای بی‌نهایت یا دست کم بسیار بزرگ داشته باشند. از آنجا که این خصوصیات در رشته‌های غیرکمّی مثل علوم انسانی کمتر وجود داشت با رواج آمار در این رشته‌ها بنای استفاده از فنون آماری بر فرضیات ساده‌تر و انعطاف‌پذیر گذاشته شد. در اوایل قرن حاضر ویلیام گاست[1] با نام مستعار استیودنت مقالات بسیاری درباره روشهای تجزیهٔ آماری، مخصوصاً تحلیل و تفسیر مشاهداتی که از طریق نمونه گیری به دست آمده بودند، منتشر نمود. او اولین کسی بود که به اهمیت روشهای در حال توسعه برای استخراج اطلاعات قابل اطمینان از نمونه‌های کوچک پی‌برد. روشهایی که گاست ارائه داد، بعدها در انگلستان به وسیله فیشر و همکاران او متداول گردید. آنها نظریه و کاربرد مربوط به این روشها را تا حدی، بخصوص در کشاورزی، توسعه دادند.

تجدید بنای مهمتر در آمار با یک رشتهٔ تحقیقات، بعد از جنگ دوم جهانی آغاز شد و نتیجهٔ آن گشوده‌شدن باب تازه‌ای در علم آمار بود که به مبحث روشهای ناپارامتریک[2] مشهور شده است. روشهای ناپارامتریک به آزمونهایی اطلاق می‌گردد که مشروط به مفروضات آمار کلاسیک نیستند؛ بنابراین جوامع آماری‌ای که از توزیع نرمال[3] برخوردار نیستند و داده‌های غیر کمّی با نمونه‌های کوچک را می‌توان با این فنون بررسی کرد. اغلب این روشها بنا به نیازی که در تحقیقات علوم رفتاری و اجتماعی وجود داشته است، به وسیلهٔ محققان این رشته‌ها، در بیست سال اخیر، تدوین شده‌اند. از صاحبنظران این روشها می توان از کندال[4]، ساندروم، گیبونز، موزس، زیگل[5] و کنوور[6] نام برد.

سیر تحول آمار از نظر موضوعی به سه مرحله تقسیم می‌شود:

۱. آمار توصیفی. این نوع آمار صرفاً به توصیف جامعه می‌پردازد و هدف آن محاسبهٔ پارامترهای جامعه است. چنانچه محاسبهٔ مقادیر و شاخصهای جامعهٔ آماری با استفاده از سرشماری تمامی عناصر آن انجام گیرد آن را آمار توصیفی گویند.

۲. آمار استنباطی. این نوع آمار از اوایل قرن بیستم مطرح شد. در این نوع آمار محقق با استفاده از مقادیر نمونه آماره‌ها را محاسبه می‌کند. سپس به کمک تخمین و آزمون فرض آماری، آماره‌ها به پارامترهای جامعه تعمیم داده می‌شود. به طور کلی در

1. William Gosset
2. non-parametric methods
3. normal distribution
4. Kendall
5. Siegel
6. Konoover

بحث آماری، هرجا سخن از استنباط و استنتاج باشد، آن را آمار استنباطی می‌خوانند.

۳. آمار ناپارامتریک. این نوع آمار در مقابل آمار پارامتریک[1] بیان می‌شود. فرض اساسی در آمار پارامتریک برخوردار بودن مشاهدات از توزیع نرمال است؛ در حالی که در فنون ناپارامتریک این فرض ضرورتی ندارد. در بخش اعظم تحقیقات علوم انسانی و رفتاری که بیشتر متغیرهای آن با مقیاسهای کیفی سنجیده می‌شود ـ این متغیرها فاقد توزیع آماری هستند و آنها را آزاد از توزیع[2] می‌خوانند ـ از این فنون استفاده می‌شود.

۱ـ۴ مراحل پژوهش علمی در آمار

اهمیت بنیادین روش‌شناسی آماری وقتی به بهترین وجه درک می‌شود که در پرتو فرایند عمومی کسب آگاهی، که گاهی روش علمی نامیده می‌شود، به آن بنگریم.

هرچند تحقیق علمی دارای چهارچوب معینی نیست، می‌توان آن را چنین تعریف کرد؛ فرایندی که به کمک آن می‌توان روابط پنهان در پس یک پدیده را که مغشوش به نظر می‌رسند، کشف نمود. در روش علمی ابتدا مدلها یا نظریه‌هایی که به نظر می‌رسد ماهیت پدیده را تبیین می‌کنند، قبول می‌شود، سپس نتیجه‌های منطقی از مدل پذیرفته شده استخراج می‌شود و آنها را با توجه به نتایج یافته‌های واقعی می‌سنجند، مدل تعدیل می‌شود و جستجو به منظور یافتن تبیین بهتر برای پدیده ادامه می‌یابد. ویژگیهای فرایند کسب آگاهی به اندازه انواع زمینه‌های تحقیق متفاوتند. در اینجا تنها بعضی از مراحل اساسی که هستهٔ اصلی اکثر تحقیقات علمی را تشکیل می‌دهند، فهرست‌وار ذکر می‌گردد:

الف) مشخص کردن هدف. هرگاه دانش موجود دربارهٔ موضوع مورد نظر، کافی نباشد، به کمک روشهای تحقیق تلاش برای افزایش آگاهی از موضوع انجام می‌گیرد. امر تحقیق بیشتر ممکن است معطوف به هدفهای مشخصی باشد؛ از قبیل اثبات یک نظریهٔ جدید یا بررسی دقیق نظریهٔ موجود از این لحاظ که تا چه میزانی نتایج منطقی حاصل از آن به وسیلهٔ یافته‌های واقعی تأیید می‌شود. در بعضی موارد، ممکن است هدف تحقیق فقط این باشد که پایه‌ای برای اطلاعات به دست آید که تا حدی منعکس کنندهٔ وضع جاری امور باشد؛ مثلاً اطلاع از وضع درآمد کارکنان یک شرکت که می‌تواند پایه‌ای برای مطالعهٔ درآمدهای جانبی آنها باشد. در موارد دیگری هدف

1. parametric statistics
2. free of distribution

تحقیق ممکن است بسیار جامعتر باشد و علاوه بر ایجاد ادراکی دقیقتر از عوامل عمل‌کننده محیطی، تعیین امکانات و کاربرد آنها در کنترل یا اصلاح امور جنبی یک پدیده را نیز شامل بشود. شناخت عوامل بهداشتی هرزبرگ در مدیریت مثالی از این‌گونه اهداف است. شناخت عوامل نه تنها کسب ادراک دقیقتر از عوامل موجود در محیط را سبب می‌شود، بلکه با استفاده از این شناخت و حفظ آن در یک سطح استاندارد می‌توان مانع عدم رضایت شغلی شد.

ب) جمع‌آوری *داده‌ها*. در هر تحقیقی تهیهٔ داده‌های واقعی، با توجه به مقصودی که از پژوهش داریم، اهمیت اساسی دارد. فرایند گردآوری اطلاعات ممکن است فعالیتهای بسیار متنوعی از قبیل آزمایشهای پیچیده در شرایط کنترل شده، بررسیهای اجتماعی ـ اقتصادی، نظرخواهی، یا حتی بررسی تاریخی را دربرگیرد. امروزه که وسایل نگهداری داده‌ها بسیار پیشرفت کرده و مکانیزه شده است، زیاد شدن کمیت مشاهدات، واقعیتی قابل توجه در زندگی است. داده‌ها، نوعاً به گونه‌ای جمع‌آوری می‌شوند که اندازه‌های عددیِ بعضی از ویژگیها، یا شرح بعضی از صفات کیفی افراد یا عناصر تحت مطالعه، یا هر دو هستند.

ج) تجزیه و تحلیل *داده‌ها*. داده‌هایی که به وسیلهٔ روشهای مناسب آزمایش یا مشاهده یا ... گردآوری می‌شوند، منبع اساسی برای کسب اطلاعات جدید دربارهٔ پدیدهٔ مورد مطالعه هستند. بعد از جمع‌آوری داده‌ها لازم است مجموعهٔ داده‌ها را بررسی کنیم و اطلاعات مربوط به موضوعاتی را که در مرحلهٔ مشخص کردن هدفها مطرح شده‌اند، استخراج کنیم. تجزیه و تحلیل دقیق داده‌ها برای بررسی معلومات جدید و تعیین نقاط قوت و ضعف آنها ضروری است.

د) *بیان یافته‌ها*. مفاد اطلاعاتی که از داده‌ها حاصل می‌شوند، با توجه به هدفهایی که در مرحلهٔ اول تحقیق مشخص شده‌اند، مورد بررسی قرار می‌گیرند. تحلیل داده‌ها برای پاسخگویی به سؤالاتی از این قبیل طرح‌ریزی می‌شود: «از شواهدی که به وسیلهٔ این داده‌ها فراهم می‌آید، چه نکات کلی‌ای دربارهٔ پدیدهٔ تحت مطالعه می‌توان استخراج کرد؟»، «آیا فرضیه یا نظریهٔ موجود با داده‌ها در تناقض است؟» و «آیا داده‌ها نظریهٔ جدیدی را برای تبیین پدیدهٔ تحت مطالعه القا می‌کنند؟». نتایج تجزیه و تحلیل داده‌ها برای جوابگویی به این سؤالات و نیز سنجش میزان عدم قطعیتی که در جوابها وجود دارد، به کار گرفته می‌شوند.

آگاهی حاصل از مراحل پژوهش علمی غالباً به صورت ارائه اصلاحاتی در نظریهٔ موجود بیان می‌شود که خود، ممکن است مستلزم بررسی بیشتری از طریق گردآوری و تجزیه و تحلیل واقعیات باشد؛ بنابراین ماهیت اساسی کسب اطلاع، نوعاً تکرار این دور به شکلهای مختلف است. بندرت ممکن است که در یک یا چند بار تکرارِ این دور حقیقتی پنهان بماند. تغییر شرایط در بسیاری از زمینه‌ها ایجاب می‌کند که فرایند تکرار، به صورت مستمر، ادامه داشته باشد.

۱ـ۵ عناصر اساسی در تحقیقات مدیریتی

در طرح تحقیقات رفتاری و مدیریتی دو عنصر اصلی وجود دارد؛ فرضیه‌های تحقیق و متغیرهایی که برای آزمودن آنها به کار گرفته می‌شوند. فرضیه‌ها از نظریات و تحقیقات گذشته منتج می‌گردند و چراغ راه مطالعات جدید هستند. متغیرها، فرضیه‌ها را به صورتی نشان می‌دهند که محققان رفتاری و مدیریتی بتوانند آنها را مشاهده و اندازه گیری کنند؛ به همین دلیل پیش از توضیح درباره فرضیه‌ها، به توضیح درباره متغیرها می‌پردازیم.

«متغیر خصیصه» متغیری است که مقدار آن از یک فرد به فرد دیگر و یا از عضوی به عضو دیگر جامعهٔ آماری ممکن است تغییر کند؛ برای مثال اندازه سازمان که از کوچک به متوسط و بزرگ تغییر می‌کند یک متغیر است.

«متغیر مستقل و وابسته» دو نوع از متغیرهایی هستند که در تحقیقات رفتاری معمولاً برای آزمودن فرضیات مورد استفاده قرار می‌گیرند. متغیر مستقل که گاه متغیر درونداد یا محرک[1] نیز نامیده می‌شود، علت احتمالی یا فرضی متغیر وابسته است. متغیر وابسته معلول احتمالی یا فرضی‌ای است که گاه متغیر پاسخ یا برونداد[2] نامیده می‌شود.

متغیر مهم دیگری که معمولاً در پژوهشهای مدیریتی به کار می‌رود، متغیر تعدیل‌کننده[3] است. متغیر تعدیل‌کننده در حقیقت متغیری ثانوی است که پژوهشگر می‌خواهد تأثیر آن را در متغیر مستقل اولیه و متغیر وابسته ملاحظه کند. متغیر تعدیل کننده بدین منظور انتخاب می‌شود که روشن شود آیا این متغیر، رابطه بین متغیر مستقل و وابسته را تحت تأثیر قرار می‌دهد یا نه، برای مثال اگر فرضیه ما این باشد که «آزمون شخصیت در گروه زنان دقیقتر از گروه مردان، معدل نمره‌های مسئولیت پذیری را

---

1. input or stimulus variable
2. response or output variable
3. moderator variable

پیش‌بینی می‌کند» مفهوم ضمنی آن این است که برای هر جنس یک معادلهٔ پیش‌بینی جداگانه لازم است. متغیر تعدیل‌کننده یا واسطه‌ای در اینجا جنسیت است که هرچند خود هیچ‌گونه تأثیری در معادلهٔ پیش‌بینی ندارد، نشان می‌دهد که معادلهٔ پیش‌بینی در هر گروه چگونه است.

یکی دیگر از متغیرهایی که در فرضیات مدیریتی جایگاه ویژه‌ای دارد «متغیر کنترل»[1] است. به متغیرهایی که در تحقیق لازم است تأثیر آنها خنثی شود یا از بین برود، متغیرهای کنترل گویند. تفاوت این متغیر با متغیرهای تعدیل‌کننده این است که تأثیرات متغیرهای کنترل از میان می‌رود ولی تأثیرات متغیرهای تعدیل‌کننده مورد مطالعه قرار می‌گیرد.

علاوه بر نوع، مقیاس اندازه گیری متغیرها نیز از اهمیت ویژه‌ای برخوردار است. اندازه گیری عبارت است از نسبت دادن اعداد به خصوصیات اشیا، وقایع یا افراد بر طبق قواعدی منطقی و قابل قبول. برای بیان تفاوتهای موجود بین انواع مختلف متغیرها به قواعد متفاوتی نیاز هست. برای نسبت دادن اعداد به متغیرها چندین قاعدهٔ معروف وجود دارد. مجموعهٔ قواعد مختلف، مقیاسهای مختلف اندازه گیری و مجموعهٔ قواعد مشخص، مقیاسهای اندازه گیری مشخصی را ارائه می‌دهند. چهار نوع مقیاس اندازه گیری مهم برای متغیرها وجود دارد:

الف) مقیاس اسمی[2]. اندازه گیری در ضعیف‌ترین شکل خود وقتی است که اعداد یا سمبولها صرفاً برای طبقه‌بندی اشیا، اشخاص یا خصوصیات به کار رود. وقتی که اعداد یا سمبولها به منظور مشخص نمودن گروههایی که اشیای مختلف متعلق به آن گروهها هستند، به کار رود، در این صورت یک «مقیاس اسمی» یا طبقه‌ای ایجاد شده است؛ برای مثال به منظور مشخص کردن سازمانها می‌توانیم از مقیاس اسمی استفاده کرده، آنها را به الف، ب و ج نامگذاری کنیم. مثال دیگر تقسیم‌بندی سبکهای مدیریت به S_1، S_2، S_3 و S_4 است. در واقع در اینجا ما مدیران را برحسب سبک رهبری‌شان نامگذاری کرده‌ایم.

ب) مقیاس ترتیبی (رتبه‌ای)[3]. مواردی پیش می‌آید که صرف نظر از تفاوت محتویات یک طبقه با محتویات طبقه دیگر، یک نوع ارتباط بین آنها برقرار است. روابط موجود بین طبقات با توجه به نوع مقیاس، بیشتر به صورت «ترجیح دارد به؛

---

1. control variable 2. nominal scale 3. rank scale

و بالاتر، بیشتر، مشکلتر، مغشوشتر یا کاملتر است از» بیان می‌شود. چنین روابطی را معمولاً با علائمی نظیر «>» نشان می‌دهند. چنانچه میان برخی از طبقاتِ یک مقیاس اسمی یعنی گروه متشکل از طبقات هم ارز[1] رابطهٔ «>» برقرار باشد، در آن صورت یک «مقیاس نسبتاً ترتیبی»[2] خواهیم داشت و اگر رابطهٔ «>» برای تمام طبقات به صورت جفت جفت صادق باشد، به طوری که طبقات کاملاً رتبه‌بندی شده باشند، در آن صورت یک «مقیاس ترتیبی» خواهیم داشت. استفاده از اصطلاحاتی چون «بالا، وسط، پایین» و «قوی، متوسط، ضعیف» در تحقیقات بیانگر مقیاس ترتیبی است؛ برای مثال طبقه‌بندی مردم یک کشور به پردرآمد، متوسط و کم درآمد مقیاسی ترتیبی است.

ج) *مقیاس فاصله‌ای*[3]. وقتی که مقیاسی همهٔ خصوصیات یک مقیاس ترتیبی را دارا باشد و علاوه بر آن فاصلهٔ بین هر دو عدد بر روی مقیاس میزان مشخصی داشته باشد، در آن صورت به مقیاسی دست یافته‌ایم به مراتب قوی‌تر که آن را «مقیاس فاصله‌ای» می‌نامند. در این مقیاس تصور ما از طبقات مختلف داده‌ها آنچنان دقیق است که می‌دانیم فواصل بین هر دو عدد بر روی مقیاس دقیقاً چقدر است. هر مقیاس فاصله‌ای با واحد ثابت و مشترکی مشخص می‌شود که یک عدد واقعی را به زوجهای مشاهدات ما در یک مجموعهٔ ترتیبی نسبت می‌دهد. در این نوع اندازه گیری نسبت هر دو فاصله، مستقل از واحد اندازه گیری و مستقل از نقطهٔ صفر است. در مقیاس فاصله‌ای نقطهٔ صفر و واحد اندازه گیری اختیاری و قراردادی است؛ برای مثال با آنکه واحد اندازه گیری و نقطهٔ صفر در دو مقیاس جداگانهٔ سانتیگراد و فارنهایت با یکدیگر متفاوتند، هر دو برای اندازه گیری حرارت به کار می‌روند.

د) *مقیاس نسبتی (نسبی)*[4]. به مقیاسی که علاوه بر همهٔ خصوصیات مقیاس فاصله‌ای دارای نقطهٔ صفر واقعی نیز هست «مقیاس نسبتی» گفته می‌شود. در مقیاس نسبتی، نسبت هر دو نقطه روی مقیاس از واحد اندازه گیری مستقل است؛ برای مثال مقیاسهایی چون پوند و گرم نقطهٔ صفر واقعی دارند و نسبت بین هر دو وزن دلخواه، از واحد اندازه گیری مستقل است. وزن دو شیء مختلف را می‌توان هم با گرم و هم با پوند اندازه گیری کرد و در این صورت نسبت دو وزن برحسب پوند برابر است با نسبت همان دو وزن بر حسب گرم.

---

1. equivalent
2. partially ordered scale
3. interval scale
4. ratio scale

جدول ۱ـ۱ نشان‌دهندهٔ خلاصهٔ مفاهیم ذکر شده درباره مقیاسهای چهارگانه و مقایسهٔ آنهاست. این جدول نشان می‌دهد که عملیات جبری و ریاضی معمول را فقط برای مقیاسهای فاصله‌ای و نسبتی می‌توان به کار گرفت؛ چراکه فقط این مقیاسها نسبت به صفر (قراردادی یا مطلق) سنجیده می‌شوند.

جدول ۱ـ۱ مقایسهٔ مقیاسهای چهارگانه

مراتب / نوع مقیاس	ترتیب	فواصل	مبدأ صفر قراردادی	مبدأ صفر مطلق
اسمی	ندارد	ندارد	ندارد	ندارد
رتبه‌ای	دارد	ندارد	ندارد	ندارد
فاصله‌ای	دارد	دارد	دارد	ندارد
نسبتی	دارد	دارد	دارد	دارد

پس از توضیح درباره متغیرها در اینجا بـه بـحث دربـاره عـنصر اول تـحقیق ـ فرضیه‌ها ـ می‌پردازیم.

«فرضیه»[1] عبارتی آزمایشی است که رابطهٔ قابل انتظار بین دو یا چند متغیر را به صورت دقیق و روشن بیان می‌کند. فرضیه را آزمایشی می‌نامند؛ زیرا باید صحت آن از طریق آزمایش بررسی شود. فـرضیه مـمکن است کـاملاً دقـیق و اسـاسی بـاشد و یـا مجموعه‌ای از حقایق پشتیبانی و تصدیق شود یا پیش‌بینی غیر دقیقی باشد که فقط با چند فرضیهٔ آزمایشی دیگر و یا چند تحقیق پشتیبانی گردد. همچنین فرضیه می‌تواند بهترین حدس ممکن، هنگام فقدان اطلاعات کافی باشد. اگر فـرضیه نشـان‌دهندهٔ انتظـارات پژوهشگر در زمینهٔ روابط بین متغیرهای پـژوهشی بـاشد، آن را «فـرضیهٔ پـژوهشی» می‌نامند. فرضیه از طریق روش قیاسی با استفاده از یک نظریه یا از طریق روش استقرایی با استفاده از یک رشتهٔ مطالعات تحقیقی و یا از طریق روشهای قیاسی و استقرایی با هم، تدوین می‌گردد. فرضیه تدوین شده باید دارای ویژگیهایی باشد؛ این ویژگیها به طـور خلاصه از دید دونالدرای و همکارانش عبارتند از:

۱. فرضیه باید واضح و بدون ابهام باشد و بیشتر در قالب جملات خبری بیان گردد.

۲. فرضیه باید قابل تبیین (علت‌یابی) باشد.

1. hypothesis

۳. فرضیه باید بیان‌کنندهٔ رابطهٔ مورد انتظار بین متغیرها باشد.
٤. فرضیه باید قابل آزمون باشد.

پس از اینکه یک فرضیهٔ پژوهشی تدوین گردید، مرحلهٔ آزمون آن با گردآوری داده‌ها آغاز می‌گردد. فرضیه‌های پژوهشی ـ مدیریتی فقط زمانی قابل آزمون هستند که در قالب فرضیه‌های آماری بیان گردند. نحوه برخورد آماری با فرضیه‌های پژوهشی می‌تواند زوایای مختلفی داشته باشد که نمونه‌هایی از آنها با بیان فرضیه‌های ذیل تشریح می‌گردد:

فرضیه ۱. متوسط نمرهٔ مسئولیت‌پذیری مدیران در ایران حداقل ۵۰ است.

فرضیه ۲. بیش از ٦٠ درصد کارکنان سازمان «الف» از کار خود راضی هستند.

فرضیه ۳. بین میزان ارضای نیازهای اساسی کارکنان سازمان «الف» و عملکرد آنها ارتباط وجود دارد.

فرضیه ۴. روش آموزش متمرکز مدیران از نظر اثربخشی هیچ تفاوتی با روش غیرمتمرکز ندارد.

فرضیه ۵. دوره‌های ضمن خدمت موجب افزایش مهارتهای انسانی مدیران می‌گردد.

فرضیه ۶. عوامل نگهدارنده هرزبرگ به طور یکسان بر تمایل به ترک خدمت (ترک خدمت بالقوه) مؤثرند.

فرضیه ۷. نسبت رضایت از روشهای ارزشیابی کارکنان در تمامی وزارتخانه‌های کشور مساوی است.

فرضیه ۸. بلوغ پیروان در سازمانهای کشور از توزیع نرمال برخوردار است.

۱ـ۵ـ۱ فرضیه‌های توصیفی در مقابل فرضیه‌های استنباطی[1]

تمایز عمدهٔ فرضیه‌ها ـ که در اینجا می‌توان از آن بحث کرد ـ توصیفی یا استنباطی بودن آنهاست. چنانچه مدیران در ایران مورد آزمون مسئولیت‌پذیری قرار گیرند و متوسط نمرهٔ مسئولیت‌پذیری آنان محاسبه گردد، فرضیه ۱ فرضیه‌ای توصیفی خواهد بود. همچنین اگر در آزمون فرضیه ۲ با مراجعه به تمام کارکنان سازمان «الف» نسبت افراد راضی مشخص گردد، این فرضیه توصیفی خواهد بود.

1. descriptive versus inferential hypothesis

حال این سؤال پیش می‌آید که آیا اندازه گیری مسئولیت‌پذیری تمامی مدیران در ایران یا شمارش کارکنان راضی در کل سازمان «الف» امکانپذیر است. اگر پاسخ چنین سؤالی منفی باشد (در خصوص هر فرضیه‌ای با هر نوعی) موضوع در قالب فرضیه‌های استنباطی قابل بررسی است. در اینجا روش معمول این است که به نمونه‌ای از جامعه برای بررسی فرضیه مراجعه شود.

واقعیت این است که بیان فرضیه‌های پژوهشی ـ توصیفی در قالب فرضیه‌های آماری ضرورت ندارد و به تعبیری بیان آنها با فرضیه‌های آماری بی‌معناست؛ چرا که با محاسبهٔ شاخص آزمون فرضیه در جامعه، به راحتی می‌توان فهمید که فرضیه تأیید شده، یا رد گردیده است. فرضیهٔ پژوهشی در صورتی فرضیهٔ آماری محسوب می‌شود که فرضیه‌ای استنباطی باشد؛ چون در این نوع بررسی، استنباط با خطای ناشی از نمونه گیری همراه است و تا حدودی صحت و سقم آن تحت‌تأثیر شانس و تصادف است؛ از این رو تلاش می‌گردد ابتدا فرضیهٔ آماری تعریف و سپس با احتساب درصدی از خطا آن فرضیه تأیید یا رد گردد.

۲ـ۵ـ۱ فرضیه‌های تک متغیره در مقابل فرضیه‌های چند متغیره[1]

چون در هر دو فرضیه ۱ و ۲ فقط با یک متغیر سروکار داریم، چنین فرضیه‌هایی را تک متغیره می‌خوانیم، اما در بیشتر فرضیه‌ها با دو یا چند متغیر روبرو هستیم. فرضیه‌های ۳ و ٤ و ٥ دارای دو متغیر هستند، چنین فرضیه‌هایی را دو متغیره می‌گوییم. در فرضیهٔ ۳ دو متغیر به طور آشکار بیان شده است؛ نیاز اساسی و عملکرد کارکنان دو متغیر فرضیه هستند. در فرضیهٔ ٤ متغیرها عبارتند از آموزش متمرکز و غیرمتمرکز. در فرضیهٔ ٥ نیز دو متغیر وجود دارد؛ زیرا برای بررسی اینکه «آیا دوره‌های ضمن خدمت موجب بهبود مهارتهای انسانی مدیران می‌شود یا نه؟» باید نمره‌های مدیران را قبل از دوره و پس از دوره بدانیم که هریک از آنها را یک متغیر می‌خوانیم.

فرضیهٔ ٦ فرضیه‌ای چند متغیره است. این فرضیه هرچند در ظاهر دارای دو متغیر (عوامل نگهدارنده و تمایل به ترک خدمت) است، در بیان آماری به عنوان یک فرضیه چند متغیره معرفی می‌شود. اگر بپذیریم که عوامل نگهدارنده براساس نظریهٔ هرزبرگ

1. univariate versus multivariate hypothesis

عبارتند از: روابط متقابل افراد، شرایط کار، نظارت فنی، امنیت شغلی، حقوق، و زندگی شخصی؛ بنابراین فرضیهٔ ٦ فرضیه‌ای چند متغیره خواهد بود.

۳ـ۵ـ۱ فرضیه‌های همبستگی در مقابل فرضیه‌های تجربی[1]

برای بررسی فرضیهٔ ۳ ممکن است ما تمامی (نمونه‌ای از) کارکنان سازمان «الف» را مورد پرسش قرار دهیم؛ به گونه‌ای که از یک طرف میزان نیاز اساسی ارضا شده آنها و از طرف دیگر میزان عملکرد هر شخص را مورد سنجش قرار دهیم. در اینجا بر مقدار این متغیرها هیچ کنترلی نداریم. در این موقعیت تحقیق، ما فقط دو متغیر را بدون آنکه بر آنها کنترل داشته باشیم مشاهده می‌کنیم. این‌گونه فرضیه‌ها را فرضیه‌های دو متغیره از نوع همبستگی می‌خوانیم.

در مقابل برای یافتن پاسخی برای فرضیهٔ ٤، بر هر دو متغیر مورد نظر کنترل داریم. با روش اختصاص تصادفی[2] معین می‌کنیم که افراد تحت آموزش با چه روشی آموزش می‌بینند. البته مقصود این نیست که ما آزادی کامل داریم بلکه روش اختصاص تصادفی، فنی است که هدف آن مقابله و مقایسهٔ گروههایی است که با روشهای متمرکز و غیرمتمرکز آموزش دیده‌اند. اما می‌دانیم که همهٔ افراد تحت آموزش یا با روش متمرکز یا با روش غیرمتمرکز آموزش داده می‌شوند. به چنین فرضیه‌ای که آزمایشگر، متغیر (گروه) را دستکاری[3] می‌کند فرضیه (مسأله) تجربی گفته می‌شود. در اینجا متغیرِ مستقل، «گروه متمرکز یا غیرمتمرکز» است که دارای مقیاس اسمی است و متغیرِ وابستهٔ «میزان اثربخشی» است که به عنوان متغیر ضمنی بیان گردیده است.

در پژوهش رفتاری تشخیص فرضیه‌های همبستگی از تجربی حائز اهمیت فراوان است. در مطالعهٔ همبستگی، از هر فرد دست کم درباره دو متغیر ـ بدون اینکه هیچیک از آنها دستکاری یا کنترل شود ـ اطلاعاتی به دست می‌آید؛ مثلاً درمقابل هراندازه نیاز اساسی ارضا شده، یک اندازه مربوط به عملکرد وجود دارد؛ البته لازم نیست بین دوصفت (متغیر) حتماً رابطهٔ خاصی وجود داشته باشد، یا یکی علت دیگری بوده، یا تغییرات یکی تابع تغییرات دیگری باشد، کافی است دو متغیر متعلق به یک فرد

1. correlation versus experimental hypothesis
2. random assignment
3. manipulation

باشند. نکتهٔ مهم این است که مقصود از گردآوری اطلاعات در این گونه مطالعات هرگز کشف روابط ضروری علت و معلولی نیست.

ضریب همبستگی، شاخص رابطه بین دو یا چند متغیر است. یکی از خواص ضریب همبستگی، دو سویه (دولبه) بودن آن است. به تعبیر دیگر اگر انگیزش متغیر مستقل X و کیفیت محصول متغیر وابسته Y باشد، میزان ضریب همبستگی درست برابر با مقدار آن در صورت جا به جا شدن این دو متغیر (کیفیت محصول X و انگیزش Y) است. این خاصیت معمولاً مانع از تفسیر علّیت می‌شود.

بدون داشتن اطلاعات اضافی، ضریب همبستگی نمی‌تواند چیزی در مورد رابطه علّی بین متغیرها در اختیار ما بگذارد. اما مطلب نباید به این صورت تفسیر شود که همبستگی، هیچ چیز در مورد علّیت به ما نمی‌دهد. در تفسیر علّیت فرضیه‌های مقابلِ متعددی وجود دارد که باید آنها را در نظر گرفت و سپس در خصوص علّیت بحث کرد. در مطالعات همبستگی، برای بررسی رابطه‌های علّی از نظریه‌ها و داده‌ها با هم استفاده می‌شود. با توجه به نظریه‌ها می‌توان بعضی از فرضیه‌های رقیب را حذف کرد و برای رابطهٔ علّی پیشنهاد شده، می‌توان در داخل مدل، دلیل منطقی ارائه نمود. سپس با استفاده از داده‌های حاصل از مطالعه می‌توان مشخص نمود که مدل پیشنهاد شده رابطهٔ علّی بین متغیرها را دقیقتر نشان می‌دهد، یا مدلهای رقیب.

۴-۵-۱ فرضیه‌های پژوهشی با گروههای جور شده در مقابل گروههای مستقل[1]

حال فرضیه ٤ را با موقعیت فرضیه ٥ مقایسه کنید. در فرضیه ٥، مجموعه‌ای از اندازه‌ها را در متغیر بهبود مهارتهای انسانی در زمان قبل از دوره و مجموعه‌ای از اندازه‌های همان متغیر را پس از دوره در اختیار داریم؛ بنابراین می‌توان گفت که این دو مجموعه از اندازه‌ها درست مانند مجموعه‌های مربوط به فرضیه ٤ است، با این تفاوت که دو مجموعهٔ مربوط به فرضیه ٥، مستقل از یکدیگر نیستند بلکه وابسته به یکدیگرند؛ یعنی در اینجا دو گروه آزمون شونده را که هر یک جداگانه و به طور تصادفی برگزیده شده باشد، در اختیار نداریم؛ بلکه فقط یک گروه داریم که در آن هر آزمون‌شونده‌ای را از لحاظ یک متغیر واحد دو بار اندازه گیری کرده‌ایم. این دو مجموعه را اندازه‌های وابسته به

1. matched versus independent groups

هم نیز می‌نامند؛ به این دلیل که هر نمره بعد از دوره متناظر و جور شده با نمره قبل از دوره است. از دیگر موارد گروههای جور شده پرسشنامه‌هایی با دو ستون برای هر سؤال است؛ برای مثال اگر یک ستون نشان‌دهنده نمرهٔ وضع موجود و ستون دیگر نشان‌دهندهٔ نمرهٔ وضع مطلوب صفت مورد تحقیق باشد، پاسخهای حاصل از دو ستون گروههای جور شده هستند؛ گروههای فرضیه در اینجا نمره‌های وضع موجود و وضع مطلوب هستند؛ در حالی که پاسخ‌دهنده در گروهها مشترک است.

۵ـ۵ـ۱ فرضیه‌های پارامتریک در مقابل ناپارامتریک[1]

فنون آماری پارامتریک شدیداً تحت تأثیر مقیاس متغیرها و توزیع آماری جامعه (نمونه) تحقیق هستند. اگر متغیرهای فرضیه از نوع اسمی و رتبه‌ای باشند، آزمون آنها صرفاً به کمک فنون ناپارامتریک انجام می‌گیرد؛ از این رو این دسته از فرضیه‌ها را ناپارامتریک خوانند. چنانچه مقیاس سنجش متغیرها، فاصله‌ای یا نسبی باشد، نوع فرضیه بستگی به توزیع آماری خواهد داشت. اگر محقق براساس شواهد در دسترس، فرض نرمال بودن جامعه (نمونه) را بپذیرد، فرضیهٔ پژوهشی او فرضیه‌ای پارامتریک و در غیر این صورت ناپارامتریک خواهد بود.

فرضیه‌های ۷ و ۸ از نوع ناپارامتریک هستند و باید برای آزمودن «تساوی نسبت رضایت از روشهای ارزشیابی»، از آزمون همگنی کای ـ مربع و برای آزمودن «برخوردار بودن بلوغ پیروان از توزیع نرمال» از آزمون کای ـ مربع یا کالمگروف ـ اسمیرنف استفاده کرد که از معروفترین آزمونهای ناپارامتریک هستند.

نوع فرضیه ۱ به پیش فرض نرمال بودن توزیع جامعه (نمونه) بستگی دارد. اگر بپذیریم که توزیع «نمره مسئولیت پذیری مدیران در ایران» نرمال است فرضیه ۱ فرضیهٔ پارامتریک و در غیر این صورت ناپارامتریک خواهد بود.

۱ـ۶ آیا آمار دروغ می‌گوید؟

بسیاری از نویسندگان، گویندگان و گزارش‌دهندگان تمایل دارند که آنچه می‌نویسند، می‌گویند یا گزارش می‌دهند بیش از آنچه واقعیت دارد مورد توجه مردم یا مخاطبشان قرار گیرد. این تمایل بخصوص در متخصصان آمار و کسانی که به نوعی با فنون

---

1. parametric versus nonparametric hypothesis

آماری سروکار دارند به نحوی بارز دیده می‌شود. از آنجا که بعضی از فنون آماری از چنان انعطافی برخوردارند که زمینهٔ هرگونه تعبیر و تفسیر را به همراه توجیه ظاهراً علمی فراهم می‌آورند، برخی از متخصصان بی‌تعهد با تمسک جستن به آنها گزارشهایی تهیه می‌کنند که در واقع دروغی بسیار بزرگ است که برای روشن شدن مطلب یک نمونه ذکر خواهد شد.

ساده‌ترین نوع نمودار آماری، نمودار چند ضلعی [1] است. این نمودار برای درک کاملتر و بهتر روندها و تغییرات مورد استفاده قرار می‌گیرد؛ مثلاً می‌توانیم با نمودار نشان دهیم که چگونه درآمد ملی ۱۰ درصد در سال افزایش یافته است. کار را با ترسیم نمودار روی کاغذ شطرنجی آغاز می‌کنیم. اسامی ماههای سال را روی محور افقی می‌نویسیم و محور عمودی را بر حسب میلیارد ریال مدرج می‌کنیم. نقاط را مشخص کرده، با خطی آنها را به هم وصل می‌کنیم (نمودار ۱ـ۱).

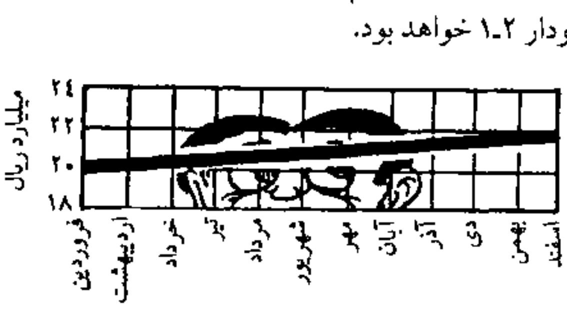
مسأله بسیار واضح است. نمودار آنچه را طی سال اتفاق افتاده، ماه به ماه نشان می‌دهد، حتی کسی که در حال دویدن است می‌تواند آن را ببیند و معنای آن را

نمودار ۱ـ۱ اولین نما از افزایش ۱۰ درصد درآمد ملی در سال

1. polygon chart

بفهمد؛ زیرا نمودار به شکلی مناسب رسم شده وخط صفر یعنی محور افقی را می‌توان برای مقایسه مورد استفاده قرار داد. ۱۰ درصد افزایش همان ۱۰ درصد به نظر می‌آید نه بیشتر و نه کمتر. روندی صعودی وجود دارد که تا حدی قابل ملاحظه است، اما زیاد چشمگیر نیست.

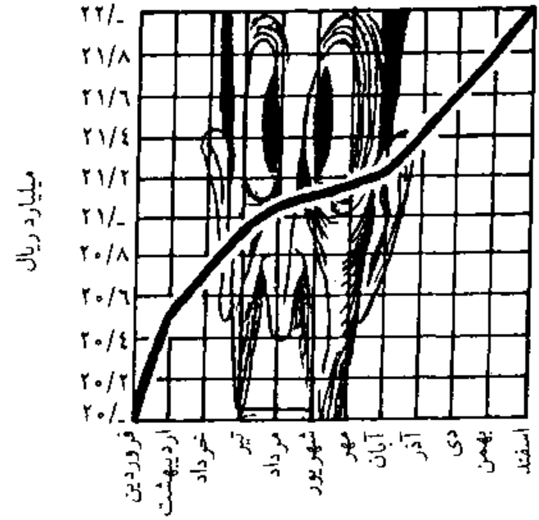
اگر قرار است فقط اطلاعات را انتقال دهید، همه چیز مهیاست، اما فرض کنید که می‌خواهید در یک بحث برنده شوید یا مطلب را به گونه‌ای وانمود کنید که برای خواننده تکان دهنده باشد یا او را وادار به انجام دادن کاری مثلاً خریدن چیزی کند. برای این مقاصد نمودار ۱ـ۱ فاقد گیرایی لازم است. اگر قسمت پایین این نمودار را قیچی کنید حاصل کار نمودار ۲ـ۱ خواهد بود.



نمودار ۲ـ۱ نمای قیچی‌شدهٔ نمودار ۱ـ۱

نمودار جدید مقاصد شما را برآورده می‌سازد. ارقام همان ارقام و منحنی نیز همان منحنی پیشین است. هیچ دروغی اعلام نشده، اما تصویر چیزی را القا می‌کند که دروغین است. آنچه بیننده عجول می‌بیند این است که روند درآمد ملی طی دوازده ماه به میانهٔ نمودار رسیده است. علت آن است که شما بیشترِ نمودار را قیچی کرده و کنار گذاشته‌اید. واضح است به علت خطای دید افزایشی اندک در درآمد ملی، افزایشی بزرگ و قابل ملاحظه به نظر می‌رسد.

شگردهای دیگری در آمار هست که چندین برابر شگرد قبلی کارآیی دارد. این شگردها کاری می‌کند که افزایش کند ۱۰ درصدی درآمد ملی به شکلی درآید که ۱۰۰ درصد به نظر برسد. برای این کار می‌توانید بسادگی نسبت بین دو محور را تغییر دهید. این امر در عین حال که مخالف قاعده و عرف و قانون نیست، سبب می‌شود نمودار ۲ـ۱ به شکل مطلوبتری درآید. تنها کاری که باید انجام دهید این است که مقیاس محور عمودی را ده برابر کوچک کنید. حاصل کار نمودار ۳ـ۱ خواهد بود.



نمودار ۱.۳ نمایی بسیار اغواکننده از رشد ۱۰ درصد درآمد ملی در سال

نمودار ۳ـ۱ اغواکننده است؛ زیرا هر که به آن نگاه کند تصور می‌کند خون ثروت در رگهای کشور به جریان افتاده است. اگر توضیح نمودار از عبارت «درآمد ملی به میزان ۱۰ درصد افزایش یافت» به «درآمد ملی به میزان قابل ملاحظهٔ ۱۰ درصد صعود کرد» تغییر کند، مؤثرتر خواهد بود؛ زیرا هیچ قید یا صفتی برای مخدوش کردن واقعیت در آن وجود ندارد، در عین حال چیزی هم در آن نیست که براساس آن کسی بتواند از شما ایراد بگیرد. مثالهایی از این نوع فراوان است که فقط به همین یک مورد بسنده می‌شود.

از مطالب گفته شده چنین برمی‌آید که اگر فردی بخواهد دروغی را با سفسطه بیامیزد باید به آمار رجوع کند. در حقیقت با آنکه امکان دارد مطالب نادرست بسیاری از طریق آمار گفته شود، در بیشتر حالات ناآگاهی سبب سوء استفاده از روشهای آماری را فراهم می‌کند. اگر فرض کنیم محققان اساساً افرادی درستکار و امانتدارند و از آمار سوء استفاده نمی‌کنند، باید دانست که سوء استفاده غیرعمدی از آمار هم به نتایج بی‌اعتبار منجر می‌شود، ولی اگر فرض ما درست نباشد، در این صورت برای این گونه افراد بسیار آسانتر است که داده‌های خام را دستکاری کنند، به طوری که نتیجهٔ نهایی با فرضیهٔ اصلی توافق حاصل کند. از اینجاست که به گفته معروف ژنرال چارلز اچ. گراسونور[1] در

---

1. General Charles H. Grosvenor

خصوص اعداد و ارقام می‌رسیم که می‌گوید: «ارقام دروغ نمی‌گویند، اما دروغگویان رقم‌سازی می‌کنند».[1]

به هرحال، در هر نوع استفاده از علم آمار باید محتاط بود. باید به این نکته توجه کرد که هیچ روش آماری، به تنهایی، از بروز اشتباهات، نادرستیها، استدلالهای غلط یا نتایج غیرصحیح جلوگیری نمی‌کند. باید مشاهدات آماری دقیق باشد، روشهای آماری به طور صحیح به کار برده شود و نتایج را باید کسی تجزیه و تحلیل و تفسیر نماید که نه تنها با روشهای آماری بلکه با رشته‌ای که آمار در مورد آن به کار رفته است، نیز آشنا باشد. روشهایی که در این کتاب مورد بحث قرار گرفته‌اند، به منزلهٔ وسایل و ابزارهایی هستند که اگر در دست افراد مطلع باشند در موقعیتهای مناسب به‌کار گرفته شوند، ممکن است به نتایج سودمندی منجر شوند ولی به تنهایی هیچ قدرت اعجازی نخواهند داشت.

۱ـ۷ خلاصه

پیشرفتهای چشمگیر علم آمار موجب شده‌است که به آن تنها به منزله ابزاری برای تهیه جدول و نمودار نگریسته نشود. امروزه آمار یکی از مهمترین وسایل مورد نیاز بیشتر رشته‌های علمی است. نفوذ علم آمار در مدیریت تا جایی است که تصمیم‌گیریهای سازمانی را بشدت تحت تأثیر قرار داده است. علاوه بر این صحت و اعتبار بسیاری از تصمیمات و تحقیقات میدانی با آمار محک زده می‌شود.

پیدایش شاخه‌های کاربردی آمار، چون نظریه صف[2]، شبیه‌سازی و طراحی آزمایشها موجب گشته است که توده‌ای از داده‌ها که تا دو دهه قبل زاید جلوه می‌کرد به عنوان با ارزشترین اطلاعات برای مدیران تشخیص داده شود. ساده و انعطاف‌پذیر شدن مفروضات آمار سبب شده است که تحلیل داده‌ها از مقیاس‌های فاصله‌ای و نسبی به مقیاسهای اسمی و رتبه‌ای کشانده شود. تا بیست سال پیش آمار صرفاً به آمار پارامتریک محدود بود ولی امروزه با پدیدار شدن آمار ناپارامتریک بسیاری از تحقیقات رفتاری و مدیریتی که دارای مقیاس اسمی و رتبه‌ای هستند قابل تحلیل و تفسیر شده‌اند.

واقعیت این است که آمار صرفاً ابزاری برای بررسی و تعیین اعتبار تحقیق است؛

---

۱. آذر، عادل؛ «تبیین آماری فرضیات در پژوهشهای رفتاری ـ مدیریتی»، *فصلنامه علمی پژوهشی دانش مدیریت*؛ ش ۲۶، ص ۲۳.

2. queuing theory

یعنی آمار وسیلهٔ تحقیق است نه هدف آن. آمار برای صحت و درستی تحقیق هیچ کمکی نمی‌کند مگر اینکه مفروضات اساسی تحقیق معتبر باشد، داده‌ها بدرستی جمع‌آوری، ثبت و تنظیم شده و سرانجام تحلیل و تعبیر و تفسیر آنها به طور منطقی صورت گرفته باشد. به هرحال، در کاربرد روشهای آماری و نتیجه‌گیری از شواهد آن محدودیتهایی وجود دارد که باید آنها را شناخت. هرگز نباید از روشهای آماری در تحقیقات استفاده شود مگر اینکه تحلیل داده‌ها را با معناتر سازد.

۱.۸ سؤالات و مسائل

سؤالات دوگزینه‌ای

۱. کلمه statistics و state از یک ریشه مشتق شده‌اند. ص ☐ غ ☐

۲. موضوع آمار توصیفی، تحلیل نمونه است. ص ☐ غ ☐

۳. در فنون آماری، از ابتدای پیدایش، بین جامعه و نمونه تفاوت وجود داشته است. ص ☐ غ ☐

۴. آمار پارامتریک برپایهٔ فرض نرمال بودن مشاهدات استوار است. ص ☐ غ ☐

۵. موضوع آمار ناپارامتریک عمدتاً تحلیل مقیاسهای فاصله‌ای و نسبی است. ص ☐ غ ☐

۶. جمع‌آوری داده‌ها اولین گام در هر تحقیق علمی است. ص ☐ غ ☐

۷. نمونه، به تعداد محدودی از افراد جامعهٔ آماری اطلاق می‌شود. ص ☐ غ ☐

۸. آماره شاخصی است که با سرشماری جامعهٔ آماری حاصل می‌شود. ص ☐ غ ☐

۹. دقت پارامتر از آماره بیشتر است. ص ☐ غ ☐

۱۰. روش اختصاص تصادفی مربوط به فرضیه‌های همبستگی است. ص ☐ غ ☐

سؤالات چهارگزینه‌ای

۱۱. چه نوع آماری آزاد از توزیع است؟

الف) توصیفی ج) پارامتریک

ب) استنباطی د) ناپارامتریک

۱۲. کدام یک از این مقیاسها دارای صفر قراردادی است؟

الف) نسبی ج) اسمی

ب) فاصله‌ای د) رتبه‌ای

۱۳. کدام مقیاس ذیل از ویژگیهای بهتری برای اندازه‌گیری برخوردار است؟

الف) نسبی ج) اسمی

ب) فاصله‌ای د) رتبه‌ای

۱٤. اگر برای بررسی فرضیه‌ای از تمام مشاهدات جامعه استفاده شود، آن را چه می‌نامند؟

الف) توصیفی ج) پارامتریک

ب) استنباطی د) همبستگی

۱۵. وزن محصولات تولید شده در یک شرکت، دارای چه نوع مقیاسی است؟

الف) رتبه‌ای ج) فاصله‌ای

ب) اسمی د) نسبی

۱٦. برای بررسی اثربخشی یک دوره آموزش مدیریت از یک گروه گواه و یک گروه آزمایش استفاده شده است. فرضیه‌های این نوع تحقیق را چه می‌نامند؟

الف) گروههای جور شده ج) گروهی همبسته

ب) گروههای مستقل د) توصیفی

۱۷. کدام یک از این متغیرها از این لحاظ که تأثیرات آنها از میان می‌رود با متغیرهای تعدیل کننده تفاوت دارد؟

الف) مستقل ج) پاسخ

ب) وابسته د) کنترل

۱۸. اولین مرحله در تحقیق علمی کدام است؟

الف) جمع‌آوری داده ج) هدفگذاری

ب) فرضیه‌سازی د) تحلیل یافته

۱۹. پارامترها به کمک کدام یک از موارد زیر قابل محاسبه هستند؟

الف) جامعه ج) جامعه و نمونه

ب) نمونه د) به شرایط تحقیق بستگی دارد

۲۰. کدام یک از این گزینه‌ها تعریف صفت مشخصه است؟

الف) از فردی به فرد دیگر تغییر می‌کند.

ب) صفت مشترک بین کلیه افراد جامعه است.

ج) متمایز کننده عناصر جامعه از همدیگر است.

د) عنصر مشترک جوامع آماری مختلف است.

مسائل

۲۱. برای فرضیه‌های تک متغیره و چند متغیره دو مثال بنویسید.
۲۲. برای هریک از فرضیه‌های همبستگی و تجربی دو مثال بنویسید.
۲۳. برای هریک از فرضیه‌های گروههای جور شده و مستقل دو مثال بنویسید.
۲۴. برای هریک از مقیاسها دو مثال بنویسید.
۲۵. مراحل تحقیق علمی را تشریح کنید.

پاسخنامهٔ سؤالات

۱) ص	۲) غ	۳) غ	۴) ص
۵) غ	۶) غ	۷) ص	۸) غ
۹) ص	۱۰) غ	۱۱) د	۱۲) ب
۱۳) الف	۱۴) الف	۱۵) د	۱۶) ب
۱۷) د	۱۸) ج	۱۹) الف	۲۰) ب

فصل دوم

مطالعهٔ توصیفی داده‌های طبقه‌بندی نشده

بسیاری از مطالعات توصیفی با اندازه‌های کوچکی از مقادیر و اعداد ($N \leq ۲۰$) بیان می‌شوند. این گونه جوامع آماری (N) از نوع جوامعی هستند که بدون دستکاری (خلاصه کردن) اندازه‌ها، می‌توان آنها را توصیف کرد. علی‌رغم کوچک بودن اندازه جامعهٔ آماری در این موارد، معمولاً ذهن آدمی نمی‌تواند محتوای کلی اطلاعات ثبت شده در مجموعهٔ داده‌ها را بسرعت درک کند. برای درک بهتر این گونه اندازه‌ها لازم است شاخصهایی برای آنها محاسبه شود که توصیف‌کنندهٔ آنها باشد.

اعدادی را که به منظور بیان کمّی توزیع اندازه‌ها از آنها استفاده می‌شود، شاخصهای عددی می‌نامند. از این‌گونه شاخصها در آمار به مقیاس وسیع استفاده می‌شود. واضح است اعدادی مشخص‌کنندهٔ توزیع مشاهدات خواهند بود که کم و بیش بتوانند خواص آنها را منعکس سازند. قبل از اینکه به بیان این اعداد و شاخصها پرداخته شود، نمادهایی را که از آنها فراوان استفاده خواهیم کرد، معرفی می‌کنیم.

۲ـ۱ نمادگذاری مجموعهٔ مشاهدات و عمل جمع

برای بیان ساده‌تر روابط بهتر است مجموعهٔ مشاهدات را به وسیلهٔ نمادهایی نشان دهند تا بحث محدود به مجموعه‌ای مشخص از اعداد نشود. مجموعهٔ داده‌ها متشکل از تعدادی اندازه است که به طور نمادی به صورت $x_۱$، $x_۲$ ... و x_N نشان داده می‌شود. آخرین اندیس (N) نشان‌دهندهٔ تعداد داده‌هاست و $x_۱$، $x_۲$ و ... به ترتیب نشان‌دهندهٔ اولین مشاهده، دومین مشاهده و ... هستند؛ مثلاً مجموعهٔ نمادهایی را که از ۳ مشاهده ۲/۱ ، ۳/۲ و ٤/۱ تشکیل شده است با نمادهای $x_۱$، $x_۲$ و $x_۳$ نشان می‌دهند که در آن صورت $x_۱ = ۲/۱$ ، $x_۲ = ۳/۲$ و $x_۳ = ٤/۱$ خواهد بود. در مطالعهٔ آمار، همه جا با عمل

جمع کردنِ داده‌ها یا اعداد دیگری که با استفاده از داده‌ها به دست می‌آیند، سروکار داریم. برای اجتناب از نوشتن مکرر علامت جمع (+) شکل بزرگ حرف یونانی زیگما (Σ) را به عنوان نماد ریاضی برای عمل جمع به کار می‌برند. جمله‌ای که بعد از نماد Σ نوشته می‌شود، نشان‌دهندهٔ مقادیری است که باید جمع شوند و نمادهای پایین و بالای Σ دامنهٔ زیرنویس i را معین می‌کنند.

مثال ۱ـ۲

$$\sum_{i=۱}^{۳} x_i = x_۱ + x_۲ + x_۳$$

$$\sum_{i=۱}^{۳} (x_i - ۲) = (x_۱ - ۲) + (x_۲ - ۲) + (x_۳ - ۲)$$

لازم است چند خاصیت اصلی نماد Σ را برای استفاده‌های بعدی ذکر کنیم. اگر a و b اعداد ثابتی باشند، داریم:

$$\sum_{i=۱}^{N} bx_i = b \sum_{i=۱}^{N} x_i \qquad (۱)$$

اثبات: ابتدا معنای عبارت سمت چپ تساوی را می‌نویسیم:

$$= bx_۱ + bx_۲ + \cdots + bx_N$$

$$= b\,(x_۱ + x_۲ + \cdots + x_N)$$

$$= b \sum_{i=۱}^{N} x_i$$

$$\sum_{i=۱}^{N} a = N.a \qquad (۲)$$

اثبات:

$$= \underbrace{a + a + \cdots + a}_{N} = N.a$$

$$\sum_{i=۱}^{N} (x_i + y_i) = \sum_{i=۱}^{N} x_i + \sum_{i=۱}^{N} y_i \qquad (۳)$$

اثبات:

$$= (x_1 + y_1) + (x_2 + y_2) + \cdots + (x_N + y_N)$$

$$= (x_1 + x_2 + \cdots + x_N) + (y_1 + y_2 + \cdots + y_N)$$

$$= \sum_{i=1}^{N} x_i + \sum_{i=1}^{N} y_i$$

تمرین

۱. با توجه به این مقادیر ۱ , ۳− , ٤ , ٥ = x و ۲ , ٥ , ۱ , ۱ , ۱− = Z حاصل هریک از عبارات ذیل را تعیین کنید.

الف) $(\sum x_i)(\sum Z_i)$ ب) $\sum x_i^2$

ج) $(\sum x_i)^2$ د) $(\sum_{i=1}^{3} x_i)(\sum_{j=1}^{2} Z_j)$

۲. ثابت کنید این رابطه برقرار است.

$$\sum (x_i + y_i + Z_i) = \sum x_i + \sum y_i + \sum Z_i$$

۳. صحت این رابطه را نشان دهید.

$$\sum_{i=1}^{10} (3x_i + 2) = 3\sum_{i=1}^{10} x_i + 20$$

٤. ثابت کنید این رابطه برقرار است.

$$\sum_{i=1}^{N} (x_i - b)^2 = \sum_{i=1}^{N} x_i^2 - 2b\sum_{i=1}^{N} x_i + Nb^2$$

۲ـ۲ پارامترهای مرکزی[1]

شاید مهمترین موضوع در مطالعهٔ هر جامعهٔ آماری، تعیین مقدار مرکزی باشد؛ یعنی تعیین مقدار نماینده که مشاهدات در اطراف آن توزیع شده‌اند. هر معیار عددی

1. central parameters

که معرف مرکز مجموعهٔ داده‌ها باشد، پارامتر مرکزی نامیده می‌شود. برای اندازه گیری مرکز جامعه‌های آماری، ملاکهای متعددی وجود دارد که توضیح درباره آنها به ترتیب خواهد آمد.

۱-۲-۲ میانگین[1]

اصلی‌ترین و مورد استفاده‌ترین شاخص مرکزی، میانگین است. اگر داده‌ها بر روی یک محور به صورت منظم ردیف شوند، مقدار میانگین دقیقاً در نقطهٔ تعادل یا مرکز ثقل توزیع قرار می‌گیرد. این محور همانند الاکلنگ است که میانگین، نقطهٔ تعادل آن است. میانگین با توجه به نوع صفتِ اندازه گیری شده دارای انواعی است که به شرح ذیل بیان می‌گردد.

۱-۱-۲-۲ میانگین حسابی[2]

معدل مجموعه‌ای از مشاهدات را میانگین حسابی می‌نامند. این میانگین به وسیلهٔ تقسیم کردن مجموع مشاهدات بر تعداد آنها محاسبه می‌شود. اگر $x_۱$، $x_۲$، ... x_N مشاهده وجود داشته باشد. میانگین حسابی آنها چنین خواهد بود:

$$\mu_x = \frac{\sum_{i=۱}^{N} x_i}{N} \qquad (۲\text{-}۱)$$

در اینجا μ_x نماد یونانی انتخاب شده برای تعریف میانگین جامعه، x_i مشاهدهٔ i ام و N تعداد کل مشاهدات است.

مثال ۲-۲ نمرهٔ مسئولیت‌پذیری پنج مدیر به شرح $x_i =$ ۱۰ , ۱۵ , ۱۴ , ۸ , ۱۳ به دست آمده است. میانگین حسابی این مشاهدات عبارت است از:

$$\mu_x = \frac{۱۰ + ۱۵ + ۱۴ + ۸ + ۱۳}{۵} = \frac{۶۰}{۵} = ۱۲$$

مشخص می‌شود که مرکز ثقل نمرات مسئولیت‌پذیری مدیران ۱۲ است. در این مثال هر یک از مشاهدات تنها یک بار تکرار شده است؛ بنابراین آن را میانگین حسابی

1. mean
2. mathematical mean

ساده[1] می‌نامند. گاهی اوقات، توزیع مشاهدات ممکن است به این صورت بیان شود:

x_i	x_1	x_2	...	x_k	
W_i	W_1	W_2	...	W_k	$\sum W_i = N$

این توزیع نشان می‌دهد که هریک از مشاهدات (x_i) دارای وزنی در حدود W_i است. در این صورت میانگین حسابی را میانگین حسابی موزون[2] می‌نامند که بیان ریاضی آن چنین می‌شود:

$$\mu_w = \frac{\sum_{i=1}^{k} W_i x_i}{\sum_{i=1}^{k} W_i} = \frac{\sum W_i x_i}{N} \qquad (۲\text{-}۲)$$

مثال ۳-۲ نمرات مسئولیت‌پذیری بیست مدیر به این شرح اندازه‌گیری شده است:

x_i (نمره)	۵	۶	۱۰	۱۲	۱۵
W_i (تکرار)	۳	۲	۵	۶	۴

متوسط نمرهٔ مدیران عبارت است از:

$$\mu_w = \frac{(۳ \times ۵) + (۲ \times ۶) + (۵ \times ۱۰) + (۶ \times ۱۲) + (۴ \times ۱۵)}{۳ + ۲ + ۵ + ۶ + ۴} = ۱۰/۴۵$$

میانگین حسابی دارای خواصی است که موجب استفاده بیشتر از آن می‌شود. به دلیل اهمیت این خواص هریک جداگانه بیان شده، اثبات می‌گردد:

خاصیت ۱. جمع جبری اختلاف مجموعه‌ای از اعداد از میانگینشان برابر صفر است؛ یعنی:

$$\sum_{i=1}^{N} (x_i - \mu_x) = ۰$$

1. simple mathematical mean
2. weighted mathematical mean

اثبات: سمت چپ را با استفاده از خواص $\sum$ مجدداً توسعه می‌دهیم:

$$= \sum x_i - \sum \mu_x$$

$$= \sum x_i - N.\mu_x$$

$$= N.\mu_x - N.\mu_x = 0$$

خاصیت ۲. هرگاه هریک از مشاهدات با عدد ثابت a جمع شود، میانگین اعداد حاصل شده برابر میانگین مجموعهٔ اعداد قبلی به اضافه a خواهد بود؛ یعنی اگر:

$$y_i = x_i + a \Rightarrow \mu_y = \mu_x + a$$

اثبات: از تعریف μ_y برای به دست آوردن قضیهٔ جدید استفاده می‌کنیم:

$$\mu_y = \frac{\sum y_i}{N} \Rightarrow \mu_y = \frac{\sum (x_i + a)}{N}$$

$$= \frac{\sum x_i}{N} + \frac{\sum a}{N} \Rightarrow \mu_y = \frac{\sum x_i}{N} + \frac{N.a}{N}$$

$$= \mu_x + a$$

خاصیت ۳. هرگاه هریک از مشاهدات جامعهٔ آماری در عدد ثابت b ضرب شوند، میانگین اعداد حاصل شده، برابر میانگین مجموعهٔ اعداد قبلی ضرب در عدد b خواهد بود؛ یعنی اگر:

$$y_i = bx_i \Rightarrow \mu_y = b\mu_x$$

اثبات: از تعریف μ_y برای به دست آوردن قضیهٔ جدید استفاده می‌کنیم:

$$\mu_y = \frac{\sum y_i}{N} \Rightarrow \mu_y = \frac{\sum (bx_i)}{N}$$

$$\Rightarrow \mu_y = \frac{b\sum x_i}{N}$$

$$\Rightarrow \mu_y = b\mu_x$$

خاصیت ۴. اگر x و z دو مجموعه از مشاهدات باشند و مجموعه y از جمع دو به دو اعداد x و z حاصل شده باشد، میانگین مجموعه y برابر است با جمع

دو میانگین مشاهدات x و z؛ یعنی اگر:

$$y_i = x_i + z_i \Rightarrow \mu_y = \mu_x + \mu_z$$

اثبات:

$$\mu_y = \frac{\sum y_i}{N} \Rightarrow \mu_y = \frac{\sum (x_i + z_i)}{N}$$

$$= \frac{\sum x_i}{N} + \frac{\sum z_i}{N}$$

$$= \mu_x + \mu_y$$

برای تشریح بیشتر این خواص به ذکر یک مثال می‌پردازیم:

مثال ۴ـ۲ با داشتن $x_i = ۱, ۳, ۵, ۷$، $z_i = -۱, ۴, ۳, ۶$، $a = -۲$ و $b = ۳$ می‌خواهیم صحت خواص میانگین حسابی رابررسی کنیم. با توجه به خاصیت ۱ داشتیم که $\sum(x_i - \mu_x) = ۰$؛ بنابراین ابتدا میانگین مقادیر x را محاسبه خواهیم کرد:

$$\mu_x = \frac{۱ + ۳ + ۵ + ۷}{۴} = ۴$$

حال صحت رابطه را به این صورت می‌بینیم:

$$\sum (x_i - \mu_x) = (۱ - ۴) + (۳ - ۴) + (۵ - ۴) + (۷ - ۴) = ۰$$

با توجه به خاصیت ۲ داشتیم که اگر $y_i = x_i + a$ باشد پس $\mu_y = \mu_x + a$ خواهد بود؛ بنابراین ابتدا آرایهٔ متغیر y را براساس تعریف آن به دست می‌آوریم:

x_i	۱	۳	۵	۷	
$y_i = x_i - ۲$	-۱	۱	۳	۵	$\sum y_i = ۸ \Rightarrow \mu_y = \frac{۸}{۴} = ۲$

حال صحت رابطه را خواهیم دید:

$$\mu_y = \mu_x - ۲ \Rightarrow \boxed{۲ = ۴ - ۲}$$

با توجه به خاصیت ۳ داشتیم که اگر $y_i = b\, x_i$ باشد پس $\mu_y = b\, \mu_x$ خواهد بود.

مجدداً براساس تعریف y مقادیر آن را به دست می‌آوریم:

x_i	۱	۳	۵	۷	
$y_i = ۳x_i$	۳	۹	۱۵	۲۱	$\sum y_i = ٤۸ \Rightarrow \mu_y = \frac{٤۸}{۱٤} = ۱۲$

حال صحت رابطه را می‌بینیم:

$$\mu_y = b\,\mu_x \Rightarrow \boxed{۱۲ = ۳\,(٤)}$$

با توجه به خاصیت ٤ داشتیم که اگر $y_i = x_i + z_i$ باشد، پس $\mu_y = \mu_x + \mu_z$ خواهد بود؛ بنابراین جدول ذیل مقادیر y را به دست خواهد داد:

x_i	۱	۳	۵	۷	$\sum x_i = ۱٦ \Rightarrow \mu_x = ٤$
z_i	۱−	٤	۳	٦	$\sum z_i = ۱۲ \Rightarrow \mu_z = ۳$
$y_i = x_i + z_i$	۰	۷	۸	۱۳	$\sum y_i = ۲۸ \Rightarrow \mu_y = ۷$

حال صحت رابطه را می‌بینیم:

$$\mu_y = \mu_x + \mu_z \Rightarrow \boxed{۷ = ٤ + ۳}$$

۲-۱-۲-۲ میانگین پیراسته[1]

غالباً دیده می‌شود که در برخی از تحقیقات، تعداد اندکی از مشاهدات، با بقیه داده‌ها، همخوانی و تجانس ندارند. در نظر داشته باشید که حتی معدودی مشاهده که به طور غیرعادی بزرگ یا کوچک باشند، می‌توانند در میانگین جامعه اثر بگذارند. برای این دسته از داده‌ها می‌توان از میانگین پیراسته استفاده نمود. طرز به دست آوردن میانگین پیراسته به این شرح است:

الف) داده‌ها به صورت صعودی مرتب شود؛

1. trimmed mean

ب) تمام مشاهدات کوچکتر از LN درصد پایین و بزرگتر از LN درصد بالا حذف شود؛

ج) میانگین مشاهدات باقیمانده محاسبه شود.

مشاهده می‌شود که در توزیعهای متقارن، این شاخص تقریباً بخوبی میانگین حسابی است، ولی در صورت وجود مقادیر ناهمگون، میانگین پیراسته از میانگین حسابی بهتر است. وقتی منحنی توزیع یک دنبالهٔ کشیده داشته باشد، این شاخص به جای میانگین حسابی، به عنوان شاخص مرکزی به کار می‌رود.

مثال ۲ـ۵ درآمد نه کارگر در هر ساعت به این شرح به دست آمده است:

۱۳۶/۷ , ۱۰۵/۸ , ۱۳۲/۱ , ۱۲۵ , ۳۵۲/۴ , ۱۱۶/۶ , ۹۳/۹ , ۱۰۶/۵ , ۱۲۸/۳

میانگین پیراسته درآمد افراد در صورتی که ۲۵ ٪ = LN ٪ باشد، عبارت است از:

۹۳/۹ , ۱۰۵/۸ , ۱۰۶/۵ , ۱۱۶/۶ , ۱۲۵ , ۱۲۸/۳ , ۱۳۲/۱ , ۱۳۶/۷ , ۳۵۲/۴

↑ ۲۵ درصد پایین ↑ ۲۵ درصد بالا

میانگین پیراسته:

$$\mu'_x = \frac{\sum x'_i}{N'} = \frac{۱۰۶/۵ + ۱۱۶/۶ + ۱۲۵ + ۱۲۸/۳ + ۱۳۲/۱}{۵} = ۱۲۱/۷$$

نوع دیگری از میانگین پیراسته وجود دارد که در آن به جای مقادیر کوچکتر از LN درصد پایین و بزرگتر از LN درصد بالا، مقدار عددی LN درصد پایین و بالا را قرار می‌دهند و سپس میانگین کل مشاهدات را محاسبه می‌کنند. این نوع میانگین را «میانگین وینزوری» گویند.

مثال ۲ـ۶ برای تعیین میانگین وینزوری در مثال ۲ـ۵ به جای دو مشاهدهٔ پایین ۲۵ درصد پایین، مقدار ۱۰۶/۵ و به جای دو مشاهدهٔ بالای ۲۵ درصد بالا، مقدار ۱۳۲/۱ قرار می‌گیرد. پس از این اصلاحات، میانگین نه مشاهده به این شرح محاسبه می‌شود:

$$\frac{۱۰۶/۵ + ۱۰۶/۵ + ۱۰۶/۵ + ۱۱۶/۶ + ۱۲۵ + ۱۲۸/۳ + ۱۳۲/۱ + ۱۳۲/۱ + ۱۳۲/۱}{۹}$$

میانگین وینزوری $= \frac{۱۰۸۵/۷}{۹} = ۱۲۰/۶$

۳ـ۱ـ۲ـ۲ میانگین هندسی[1]

برای محاسبهٔ میانگین اندازه‌های نسبی همانند نسبتها، درصدها، نرخهای رشد و شاخصها از میانگین هندسی استفاده می‌شود. میانگین هندسی یک رشتهٔ عدد مانند x_1، x_2، ... ، x_N برابر است با ریشهٔ Nام حاصلضرب آن اعداد که فرمول محاسبه آن چنین خواهد بود:

$$\mu_G = (x_1 \times x_2 \times \ldots \times x_N)^{\frac{1}{N}} \qquad (2\text{-}3)$$

در این فرمول μ_G نماد میانگین هندسی و N تعداد مشاهدات است.

مثال ۲ـ۷ نسبت سود شرکت زمزم در سال ۱۳۶۷ به ۱۳۶۶ مساوی ۳؛ سال ۱۳۶۸ به ۱۳۶۷ مساوی ۲ و سال ۱۳۶۹ به ۱۳۶۸ برابر ۴/۵ یعنی $x_1 = 3$، $x_2 = 2$ و $x_3 = 4/5$ است. میانگین هندسی برای نسبت سود شرکت زمزم عبارت است از:

$$\mu_G = (x_1 \times x_2 \times x_3)^{\frac{1}{3}} = (3 \times 2 \times 4/5)^{\frac{1}{3}} = 3$$

فرمول ۲ـ۳ چگونگی محاسبهٔ میانگین هندسی ساده را نشان می‌دهد، در حالی که اگر داده‌ها در این نوع میانگین از وزن برخوردار باشند، ناچار باید از میانگین هندسی موزون استفاده کرد. فرمول ۲ـ۴ روش محاسبهٔ میانگین هندسی موزون را نشان می‌دهد.

$$\mu_G = (\mathop{\pi}_{i=1}^{K} x_i^{W_i})^{\frac{1}{N}} = (x_1^{W_1} \times x_2^{W_2} \times \ldots \times x_k^{W_K})^{\frac{1}{N}} \qquad (2\text{-}4)$$

در این فرمول $\sum_{i=1}^{K} W_i$ مساوی N، و $\mathop{\pi}_{i=1}^{K}$ مُعرف ضرب متغیر x از ۱ تا k است.

مثال ۲ـ۸ درصد کارآیی ۲۰ نفر ماشین‌نویس در این جدول آمده است:

x_i (درصد کارآیی)	۸۸	۸۰	۶۵	۵۵	
W_i (تکرار)	۶	۵	۷	۲	$\sum W_i = 20$

1. geometric mean

میانگین درصد کارآیی به این صورت محاسبه می‌شود:

$$\mu_G = (88^6 \times 80^5 \times 65^7 \times 55^2)^{\frac{1}{20}}$$

برای ساده‌شدن محاسبات لگاریتم دو طرف رابطه گرفته می‌شود:

$$\mathrm{Log}\,\mu_G = \frac{1}{20}\,\mathrm{Log}\,(88^6 \times 80^5 \times 65^7 \times 55^2)$$

براساس خاصیت ضرب لگاریتم می‌توان نوشت:

$$\mathrm{Log}\,\mu_G = \frac{1}{20}\,[\,6\,\mathrm{Log}\,88 + 5\,\mathrm{Log}\,80 + 7\,\mathrm{Log}\,65 + 2\,\mathrm{Log}\,55\,]$$

$$\mathrm{Log}\,\mu_G = (0/3 \times 1/9445) + (0/25 \times 1/9031) + (0/35 \times 1/8129) + (0/10 \times 1/7404)$$

$$\mathrm{Log}\,\mu_G = 1/86768 \Rightarrow \mu_G = 10^{1/86768} = 73/74$$

۲ـ۲ـ۱ـ۴ میانگین هارمونیک[1]

چنانچه مشاهدات جمع‌آوری شده از مقیاس ترکیبی مثل کیلومتر در ساعت یا دور در ثانیه یا نفر ساعت برخوردار باشند، برای محاسبهٔ میانگین آنها از میانگین هارمونیک استفاده می‌شود. این میانگین برای چند اندازه یا مقدار برابر است با عکس میانگین حسابی معکوس آن اندازه‌ها. اگر مشاهدات x_1، x_2، ... ، x_N مفروض باشد، عکس این مقادیر به ترتیب $\frac{1}{x_1}$، $\frac{1}{x_2}$، ... ، $\frac{1}{x_N}$ خواهد بود. میانگین این مقادیر عبارت است از:

$$\text{میانگین معکوس مقادیر} = \frac{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_N}}{N} \qquad (۲ـ۵)$$

میانگین هارمونیک که آن را با μ_H نشان می‌دهند برابر است با:

$$\mu_H = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_N}} = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}} \qquad (۲ـ۶)$$

1. harmonic mean

مثال ۹ـ۲ رانندهای مسافت تهران ـ قم را با سرعت ۸۰ کیلومتر در ساعت میرود و همین مسیر را با سرعت ۱۰۰ کیلومتر در ساعت برمیگردد. متوسط سرعتی که این راننده داشته عبارت است از:

$$\mu_H = \frac{۲}{\frac{۱}{۸۰} + \frac{۱}{۱۰۰}} = ۸۸/۸۹$$

یعنی این راننده به طور متوسط در رفت و برگشت، ساعتی ۸۸/۸۹ کیلومتر راه طی کرده است. برای محاسبهٔ میانگین هارمونیک موزون مشاهدات از این رابطه استفاده میکنیم:

$$\mu_H = \frac{\sum W_i}{\frac{W_۱}{x_۱} + \frac{W_۲}{x_۲} + \cdots + \frac{W_K}{x_K}} = \frac{N}{\sum_{i=۱}^{K} \frac{W_i}{x_i}} \qquad (۲\text{ـ}۷)$$

مثال ۱۰ـ۲ سرعت ۳۰ ماشین بر حسب دور در ثانیه به این شرح است:

x_i (سرعت)	۵	۶	۱۰	۱۵	
W_i (تعداد)	۶	۱۰	۸	۶	$\sum W_i = ۳۰$

میانگین سرعت ماشینها با استفاده از رابطهٔ ۷ـ۲ عبارت است از:

$$\mu_H = \frac{۶ + ۱۰ + ۸ + ۶}{\frac{۶}{۵} + \frac{۱۰}{۶} + \frac{۸}{۱۰} + \frac{۶}{۱۵}} = \frac{۳۰}{۴/۰۶۷} = ۷/۳۷۶$$

ماشینها به طور متوسط دارای سرعتی برابر ۷/۳۷۶ دور در ثانیه بودهاند.

۲ـ۲ـ۲ مد (نما)[1]

مد کلمهای فرانسوی برای رایجترین لباس یا سبک است. در آمار مد به معنای توزیع آماری مقداری است که بیشترین تکرار را در میان مشاهدات جامعه داشته باشد.

1. mode

از این شاخص برای نمایش داده‌هایی استفاده می‌شود که نسبت به سایر مقادیر تکرار بیشتری دارند. مد یکی دیگر از شاخصهای مهم مرکزی است که آن را با Mo نشان می‌دهند. با استفاده از مثالهای ۲ـ۱۱ تا ۲ـ۱۳ مفهوم و کاربرد مد آشکارتر خواهد شد.

مثال ۲ـ۱۱ مشاهدات ۲ , ۲ , ۵ , ۱۵ , ۱۰ , ۸ , ۵ , ۲ را در نظر بگیرید. از آنجا که عدد ۲ بیشترین تکرار را داراست؛ پس Mo = ۲ است.

مثال ۲ـ۱۲ اگر مشاهدات به صورت ۱۵, ۷, ۲, ۱۵, ۱۰, ۸, ۵, ۲ باشد، به این دلیل که ۲ و ۱۵ نسبت به سایر مشاهدات بیشتر تکرار شده، مدهای جامعه ۲ و ۱۵ خواهند بود؛ یعنی جامعه دو مد خواهد داشت.

مثال ۲ـ۱۳ چنانچه کلیهٔ مشاهدات به یک اندازه تکرار شده باشند، جامعه فاقد مد یا مد جامعه تهی خواهد بود؛ به عنوان نمونه این مشاهدات ۸ , ۵ , ۲ , ۸ , ۵ , ۲ , ۸ , ۵ , ۲ فاقد مد هستند که در این صورت $Mo = \phi$ خواهد بود.

۲ـ۲ـ۳ چارکها[1]

اگر جامعهٔ آماری به چهار قسمت مساوی تقسیم شود، چارکهای اول تا سوم مشخص می‌شوند.

چارک اول که با Q_1 نشان داده می‌شود، مقداری است که ۲۵ درصد مشاهدات جامعه پایین‌تر از آن و ۷۵ درصد بالاتر از آن قرار می‌گیرند.

چارک دوم که با Q_2 نشان داده می‌شود، مقداری است که ۵۰ درصد داده‌های جامعه پایین‌تر از آن و ۵۰ درصد بالاتر از آن قرار می‌گیرند. چارک دوم را میانه[2] می‌گویند و با Md نشان می‌دهند. میانه در مقایسه با مد اطلاعات بیشتر و مفیدتری را در زمینه اندازه تمایل به مرکز هر توزیع ارائه می‌دهد. به طور کلی از میانه به عنوان اندازه تمایل به مرکز توزیعهایی که شکل آنها غیرمتقارن است، استفاده می‌شود. این توزیعها معمولاً شامل مقادیر انتهایی، یعنی مقادیری در حد بسیار بالا یا بسیار پایین هستند. از نظر عملی نیز میانه برای اندازه‌های تمایل به مرکز توزیعهای غیرمتقارن مناسبتر است؛ زیرا میانه کمتر تحت تأثیر مقادیر واقع شده در انتهای

---

1. quartiles
2. median

توزیع قرار می‌گیرد. خاصیت مهم میانه آن است که حاصل جمع قدر مطلق تفاضلهای مقادیر از میانه، از حاصل جمع قدر مطلق تفاضلهای مقادیر از هر عدد دیگری مثل C، کوچکتر است؛ یعنی:

$$\sum |x_i - \mathrm{Md}| < \sum |x_i - C| \qquad (۲\_۸)$$

در رابطه ۸ ـ۲، C هر عدد دلخواهی مخالف میانه است $(C \neq \mathrm{Md})$؛ به عبارت دیگر میانه حاصل جمع قدر مطلق تفاضلهای مقادیر از خود را به حداقل می‌رساند؛ یعنی:

$$\sum |x_i - \mathrm{Md}| = \mathrm{Minimum} \qquad (۲\_۹)$$

از این خاصیت میانه برای طرح‌ریزی خطوط راه‌آهن برقی شهرها، اتوبوس و مترو، و در تعیین ایستگاهها، محل پمپهای بنزین و ... می‌توان استفاده کرد.

چارک سوم که با $Q_۳$ نشان داده می‌شود، مقداری است که ۷۵ درصد مشاهدات پایین‌تر از آن و ۲۵ درصد بالاتر از آن واقع می‌شوند.

محاسبهٔ چارکها مراحلی دارد که عبارتند از:

الف) داده‌ها را به طور صعودی مرتب کنید،

ب) داده‌های مرتب شده را از ۱ تا N کدگذاری کنید،

ج) محل چارک aام $(a = ۱, ۲, ۳)$ را با استفاده از رابطهٔ ۱۰ـ۲ محاسبه کنید.

$$C_{Qa} = \frac{aN}{٤} + \frac{۱}{۲} \qquad (۲\_۱۰)$$

در این رابطه $\frac{۱}{۲}$، زوج یا فرد بودن تعداد مشاهدات را معین می‌سازد.

د) با استفاده از محل چارک، مقدار چارک را تعیین نمایید.

از آنجا که استفاده از این مراحل به زوج یا فرد بودن مشاهدات بستگی دارد، دو مثال متفاوت ذکر خواهد شد.

مثال ۱٤ـ۲ فرض کنید مشاهدات به دست آمده عبارتند از: ۸۵, ۱٤۰, ۱۲۰, ۸۰, ۹۰, ۱۰۰ و ما می‌خواهیم مقادیر چارک $Q_۱$ و $Q_۲$ و $Q_۳$ را به دست آوریم؛ بنابراین:

الف) داده‌ها را مرتب می‌کنیم:

۸۰, ۸۵, ۹۰, ۱۰۰, ۱۲۰, ۱٤۰

ب) داده‌های مرتب‌شده را کدگذاری می‌کنیم:

کد مشاهدات: ۶ ۵ ۴ ۳ ۲ ۱

مشاهدات: ۱۴۰ ,۱۲۰ ,۱۰۰ ,۹۰ ,۸۵ ,۸۰

ج) محل چارک aام را محاسبه می‌کنیم:

$$C_{Q_1} = \frac{1 \times 6}{4} + \frac{1}{2} = 2 \qquad a = 1$$

$$C_{Q_2} = \frac{2 \times 6}{4} + \frac{1}{2} = 3/5 \qquad a = 2$$

$$C_{Q_3} = \frac{3 \times 6}{4} + \frac{1}{2} = 5 \qquad a = 3$$

د) مقدار چارک را تعیین می‌کنیم:

$$Q_1 = 85 \qquad Q_2 = \frac{90 + 100}{2} = 95 \qquad Q_3 = 120$$

مثال ۱۵ـ۲ با اضافه کردن مقدار ۱۶۰ به مشاهدات مثال ۱۴ـ۲، مجدداً چارکها را محاسبه می‌کنیم.

۷ ۶ ۵ ۴ ۳ ۲ ۱

۱۶۰ ,۱۴۰ ,۱۲۰ ,۱۰۰ ,۹۰ ,۸۵ ,۸۰

چارک اول: $C_{Q_1} = \frac{1 \times 7}{4} + \frac{1}{2} = 2/25$

عدد ۲/۲۵ بدین معنی است که مقدار چارک اول در فاصلهٔ ۲۵ درصدی مشاهدات ۲ و ۳ قرار دارد که به این صورت محاسبه می‌شود:

$$Q_1 = 85 + \frac{1}{4}(90 - 85) = 86/25$$

چارک دوم (میانه): $C_{Q_2} = \frac{2 \times 7}{4} + \frac{1}{2} = 4$

پس مقدار میانه مساوی با ۱۰۰ خواهد شد.

چارک سوم:
$$C_{Q_3} = \frac{3 \times 7}{4} + \frac{1}{2} = 5/75$$

عدد ۵/۷۵ بدین معنی است که مقدار چارک سوم در فاصله ۷۵ درصدی مشاهدات ۵ و ۶ قرار دارد که به این صورت محاسبه می‌شود:

$$Q_3 = 120 + \frac{3}{4}(140 - 120) = 135$$

تمرین

۱. یک هواپیما فاصله ٤ هزار کیلومتری را با سرعت ۹۰۰ کیلومتر در ساعت و فاصله ۳ هزار کیلومتری را با سرعت ٦۰۰ کیلومتر در ساعت و فاصله ۵ هزار کیلومتری را با سرعت ۸۰۰ کیلومتر در ساعت طی می‌کند، سرعت متوسط این هواپیما را محاسبه نمایید.

۲. سه نفر کارشناس، اولویت اتومبیل پیکان را از نظر سرعت نسبت به رنو تعیین کرده‌اند. مقیاس اولویت‌گذاری از $\frac{1}{9}$ تا ۹ است. نفر اول اولویت پیکان را ۳، نفر دوم ۲ و نفر سوم $\frac{1}{2}$ تعیین کرده‌است. متوسط اولویت اتومبیل پیکان به رنو از نظر سرعت چقدر است؟

۳. مد هریک از این دو گروه را تعیین کنید:

گروه الف: ۳ , ۵ , ۷ , ۸ , ۹ , ۱۰ , ۵ , ۸ , ۵

گروه ب: ۱۰۰ , ۱۵۰ , ۱۱۰ , ۱۰۰ , ۱۵۰ , ۱۱۰

٤. چارک اول، دوم و سوم این مشاهدات را تعیین کنید:

۸۰ , ۹۰ , ۵۰ , ٦۰ , ۱۰۰ , ۷۰ , ۷۵ , ۸۰ , ۷۰ , ۸۰ , ٦۰

۵. تفاوت میانگین پیراسته و وینزوری چیست؟

٦. چرا در حالتی که مشاهدات به صورت نسبت تعریف شوند، از میانگین هندسی استفاده می‌شود؟

۲-۳ پارامترهای پراکندگی[1]

برای اینکه داده‌ها به طور واقعی‌تر توصیف شوند و مجموعه‌های مشاهدات با هم قابل

1. dispersion parameters

مقایسه باشند، باید قدم دیگری برداشته شود و برای سنجش میزان تفاوتهای آنها نیز چند معیار عددی تعریف گردد.

می‌دانیم که مهمترین شاخص مرکزی میانگین است. این پارامتر در همهٔ موارد جوابگوی نیاز تصمیم‌گیرنده نیست؛ چراکه ممکن است جوامع آماری مورد مقایسه دارای میانگین مساوی باشند. در این صورت می‌توان توزیعهای آماری را از نظر شدت نوسان داده‌ها حول مرکز خود، مورد بررسی و مقایسه قرار داد. بعلاوه ضرورتهای ریاضی و کاربردی ایجاب می‌کند که جامعه را از زاویه‌ای دیگر مورد مداقه و توصیف قرار دهیم.

مثال ۲ـ۱۶ تولید کننده‌ای می‌خواهد از بین دو تهیه کنندهٔ مواد اولیه با عنوان آلفا و بتا یکی را انتخاب کند. داده‌های مربوط به تعداد روزهایی که این تهیه کنندگان در تحویل سفارش پنج محموله تأخیر داشته‌اند، به این شرح به دست آمده است:

تهیه کننده آلفا: ۵ , ۵ , ۴ , ۳ , ۳

تهیه کننده بتا: ۲ , ۸ , ۷ , ۲ , ۱

در این مثال اگر تولید کننده بخواهد میانگین را ملاک انتخاب خود قرار دهد، هیچ تفاوتی میان تهیه کنندگان نمی‌بیند؛ چراکه متوسط تأخیر در تحویل محموله برای هر دو مساوی با چهار است. از آنجاکه توزیع آماری این دو جامعه یکسان نیست، تصمیم‌گیرنده به محاسبهٔ پارامترهای پراکندگی خواهد پرداخت که نشان‌دهندهٔ ریسک در تحویل مواد است.

برای شناخت بهتر پراکندگی به شکل ۲ـ۱ که به نمودار نقطه‌ای[1] معروف است، نگاه کنید. هریک از مشاهدات بر روی نمودار با یک نقطه نشان داده شده است. مشاهدات عبارتند از: ۷ , ۵ , ۴ , ۲ , ۱ . میانگین این مشاهدات مساوی با ۳/۸ است. نمودار نشان می‌دهد که یک جامعه را هم با شاخصهای مرکزی و هم با شاخصهای پراکندگی می‌توان توصیف کرد.

1. dot diagram

شکل ۱ـ۲ نمودار نقطه‌ای

برای آنکه اختلاف جوامع آماری با توزیعهای متفاوت آشکار گردد، باید از شاخصهایی استفاده کرد که بتوانند همان ویژگیهایی را که باعث اختلاف دو توزیع شده‌اند، نشان دهند. در مثال ١٦ـ٢ که نشان‌دهنده اختلاف دو توزیع از نظر پراکندگی است، باید شاخصی فراهم کرد که بتواند پراکندگی توزیعها را بخوبی مقایسه کند. شاخصهای پراکندگی انواعی دارند که به ترتیب معرفی می‌شوند.

۱ـ۳ـ۲ دامنهٔ تغییرات[1]

یکی از ساده‌ترین شاخصهای پراکندگی، دامنهٔ تغییرات است. این شاخص با تفاضل کوچکترین مشاهده از بزرگترین مشاهده محاسبه می‌شود. تعریف ریاضی دامنهٔ تغییرات که با R نشان داده می‌شود، چنین است:

$$R = \text{Max}\, x_i - \text{Min}\, x_i \qquad (۲\text{ـ}۱۱)$$

مثال ۱۷ـ۲ دامنهٔ تغییرات برای تهیه کنندهٔ آلفا و بتا در مثال ١٦ـ٢ عبارت است از:

تهیه کننده آلفا: $R = ۵ - ۳ = ۲$

تهیه کننده بتا: $R = ۸ - ۱ = ۷$

دامنهٔ تغییرات برای تهیه کننده بتا بسیار بزرگتر از تهیه کننده آلفا است؛ یعنی داده‌های آلفا پراکندگی کمتری دارد؛ بنابراین تولیدکننده برای سفارش مواد اولیه بهتر است به تهیه کنندهٔ آلفا مراجعه نماید.

---
1. range

دامنهٔ تغییرات علی‌رغم سادگی در محاسبه، از ثبات چندانی برخوردار نیست؛ زیرا اولاً در تعیین این شاخص، از مجموع مشاهدات فقط به کوچکترین و بزرگترین مشاهده توجه می‌شود و ثانیاً با این شاخص چگونگی توزیع سایر مشاهدات نشان داده نمی‌شود؛ به عبارت دیگر اگر داده‌ها به مقدار بزرگتر یا مقدار کوچکتر نزدیک باشند، دامنهٔ تغییرات این ویژگی مشاهدات را نشان نمی‌دهد. دامنهٔ تغییرات کاربرد وسیعی در طبقه‌بندی داده‌های جامعه و کنترل کیفیت آماری دارد. این پارامتر مخصوصاً برای اندازه‌گیری پراکندگی نمونه‌های چهار یا پنج تایی بسیار رایج و به آسانی قابل محاسبه است.

مثال ۲ـ۱۸ حقوق پرداختی ماهیانه کارکنان یک شرکت کوچک به این صورت است:

$$x_i = ۱۵ , ۱۶ , ۱۷ , ۱۷ , ۱۷ , ۲۰ , ۲۰ , ۲۲ , ۶۰$$

دامنهٔ تغییرات در این مثال مساوی با ٤٥ است که بسیار بزرگتر از پراکندگی واقعی داده‌هاست. علت این امر، با آنکه بیشتر داده‌ها به هم نزدیک هستند، این است که مقدار بزرگترین داده با سایر داده‌ها تفاوت چشمگیری دارد و سبب می‌شود پراکندگی بزرگتر از معمول جلوه کند.

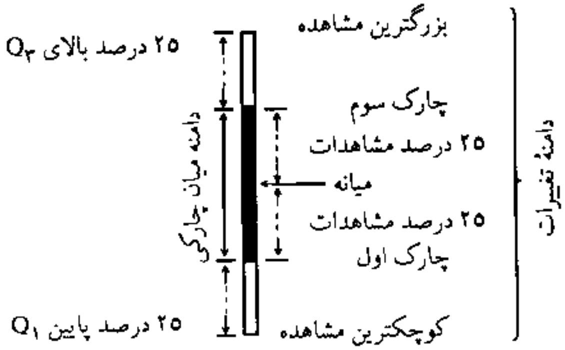
۲ـ۳ـ۲ دامنه میان چارکی[1]

دامنه میان چارکی که با IQR نشان داده می‌شود، دامنهٔ تغییرات ۵۰ درصد از مشاهدات است. در تعریف دامنهٔ میان چارکی، میانه را که ۵۰ درصد داده‌هاست، ملاک قرار داده و از پایین تا ۲۵ درصد و از بالا تا ۷۵ درصد گسترش می‌دهیم؛ به عبارت دیگر IQR از چارک اول تا چارک سوم را شامل می‌شود. برای محاسبه دامنهٔ میان چارکی ابتدا چارک اول و سوم را محاسبه می‌کنیم، سپس از تفاضل چارک اول و سوم دامنهٔ میان چارکی را به دست می‌آوریم؛ یعنی:

$$\mathrm{IQR} = Q_۳ - Q_۱ \qquad (۲\_۱۲)$$

1. interquartile range

شکل ۲ـ۲ بخوبی تفاوت دامنهٔ تغییرات را با دامنهٔ میان چارکی نشان می‌دهد.

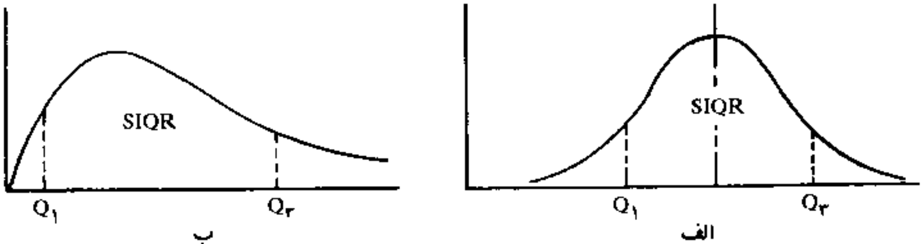


شکل ۲ـ۲ نمایش هندسی دامنهٔ تغییرات و دامنهٔ میان چارکی

در برخی ازمتون به‌جای دامنه میان چارکی از «نیمهٔ میان‌چارکی»[1] یا عنوان «انحراف چارکی»[2] استفاده‌می‌شود.انحراف چارکی مساوی است بانصف دامنه میان چارکی؛ یعنی:

$$SIQR = \frac{Q_3 - Q_1}{2} \qquad (2\text{-}13)$$

اصولاً انحراف چارکی از دامنه تغییرات و دامنه میان چارکی با ثبات‌تر است. شکل ۳ـ۲ نشان‌دهنده انحراف چارکی در توزیعهای متقارن و نامتقارن است. در نمودار این شکل مقادیر انتهایی در مقایسه با مد بر مقدار انحراف چارکی کمتر تأثیر می‌گذارد. در توزیعهایی که دارای تعداد اندکی مقدار در ابتدا و انتها هستند، از انحراف چارکی به عنوان شاخص پراکندگی استفاده می‌شود. در توزیعهای نامتقارن اغلب از میانه به عنوان شاخص مرکزی و از انحراف چارکی به عنوان شاخص پراکندگی استفاده می‌شود.



شکل ۳ـ۲ انحراف چارکی در توزیعهای متقارن و نامتقارن

---

1. semi-interquartile 2. quartile dispersion

مثال ۲ـ۱۹ از آنجا که در مثال ۲ـ۱۵ مقدار $Q_1 = ۸٦/۲۵$ و $Q_3 = ۱۳۵$ بود، پس دامنه میان چارکی و همین‌طور انحراف چارکی عبارتند از:

$$IQR = ۱۳۵ - ۸٦/۲۵ = ٤۸/۷۵$$

$$SIQR = \frac{Q_3 - Q_1}{۲} = \frac{۱۳۵ - ۸٦/۲۵}{۲} = ۲٤/۳۷۵$$

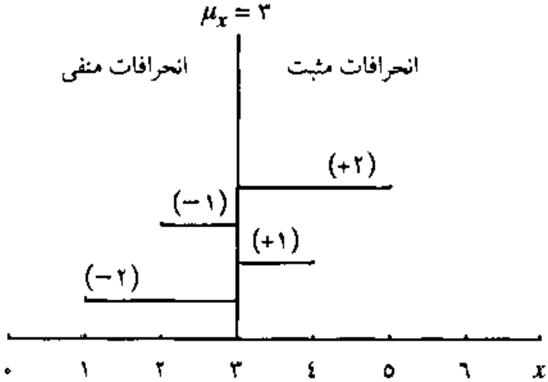
پراکندگی مشاهدات جامعه حدوداً ۲٤/۳۷۵ واحد است.

۲ـ۳ـ۳ انحراف متوسط از میانگین[1]

تمامی شاخصهایی که تاکنون برای سنجش پراکندگی داده‌های جامعه بیان شده‌اند، نواقص مخصوص به خود دارند، به طوری که هیچ کدام قادر به بیان تمامی تغییرات نیستند؛ پس باید به دنبال شاخصی بود که تغییرات کل داده‌ها را اندازه‌گیری کند. طبیعی است تغییر، زمانی مفهوم پیدا می‌کند که هریک از داده‌ها نسبت به یک مبدأ (مرکز) مقایسه شوند. به جرأت می‌توان گفت که بهترین مرکز برای داده‌ها میانگین است. استدلال اولیه در انتخاب این مرکز آن است که میانگین دارای خواص جبری است؛ یعنی اینکه می‌توان هر عمل ریاضی و جبری را روی آن اعمال کرد.

در شکل معمول تعریف پراکندگی داده‌ها از میانگین، باید مجموع پراکندگی داده‌ها را به صورت $\sum_{i=1}^{N}(x_i - \mu_x)$ تعریف کرد. بخاطر داریم که مجموع انحراف داده‌ها از میانگینشان همیشه مساوی صفر است (خاصیت اول میانگین). واضح است که آنچه موجب صفر شدن انحرافات می‌شود، خنثی‌شدن انحرافات منفی با انحرافات مثبت است. شکل ۲ـ٤ نشان‌دهنده توزیع مشاهدات حول میانگین است. در شکل مشخص است که مجموع تفاوت مشاهدات کوچکتر از میانگین، کاملاً مساوی با مجموع تفاوت مشاهدات بالاتر از میانگین است.

1. average deviation of mean



شکل ۲ـ۴ نمایش هندسی تفاوت مشاهدات از میانگین

از آنجا که اهمیت انحرافات مثبت و منفی به لحاظ کاربردی یکسان است؛ پس می‌توان از قدر مطلق انحرافات به این صورت استفاده کرد:

$$|e_i| = |x_i - \mu_x|$$

براین اساس می‌توان به شاخص جدیدی از پراکندگی با عنوان انحراف متوسط از میانگین[1] رسید؛ یعنی:

$$A.D_{\mu} = \frac{\sum |x_i - \mu_x|}{N} \qquad (۲\text{ـ}۱۴)$$

مثال ۲۰ـ۲ مجدداً مشاهدات مربوط به تهیه‌کننده‌های آلفا و بتا را در نظر بگیرید:

تهیه‌کننده آلفا: ۵ , ۵ , ۴ , ۳ , ۳

تهیه‌کننده بتا: ۲ , ۸ , ۷ , ۲ , ۱

علی‌رغم آنکه میانگین آنها یکسان ($\mu = ۴$) است انحراف متوسط از میانگین آنها

---

۱. در برخی از تحقیقات، از انحراف متوسط از مد ($A.D_{Mo} = \frac{\Sigma |x_i - Mo|}{N}$) و انحراف متوسط از میانه ($A.D_{Md} = \frac{\Sigma |x_i - Md|}{N}$) استفاده می‌شود. درباره این شاخصها به لحاظ کاربردهای محدودی که دارند در این نوشته بحث نمی‌شود.

متفاوت و به این صورت خواهد بود:

$$A.D_{\mu_\alpha} = \frac{|۳-۴| + |۳-۴| + |۴-۴| + |۵-۴| + |۵-۴|}{۵} = \frac{۴}{۵} = ۰/۸۰$$

$$A.D_{\mu_\beta} = \frac{|۱-۴| + |۲-۴| + |۷-۴| + |۸-۴| + |۲-۴|}{۵} = \frac{۱۴}{۵} = ۲/۸۰$$

با توجه به مقادیر به دست آمده می‌توان گفت که توزیع شرکت آلفا نسبت به توزیع شرکت بتا از پراکندگی کمتری برخوردار است؛ بنابراین تصمیم‌گیرنده برای سفارش مواد اولیه از شرکت آلفا استفاده خواهد کرد؛ چراکه ریسک در تحویل مواد اولیه در شرکت آلفا کمتر از شرکت بتاست.

انحراف متوسط از میانگین به عنوان شاخص پراکندگی دارای یک نقص اساسی است و آن اینکه هرگاه در برابر تعداد کمی انحرافات بزرگ، تعداد زیادی انحرافات کوچک وجود داشته‌باشد، این شاخص تأثیر انحرافات بزرگ را نشان نمی‌دهد؛ مثلاً اگر از ۱۰۰ مشاهده قدر مطلق انحراف ۹۹ مشاهده، مساوی ۰/۱، و فقط قدرمطلق انحراف یک مشاهده، مساوی ۲ (۲۰ برابر هریک از انحرافات دیگر) باشد، در این صورت انحراف متوسط از میانگین چنین خواهد بود:

$$A.D_{\mu} = \frac{(۹۹ \times ۰/۱) + (۱ \times ۲)}{۱۰۰} = ۰/۱۱۹$$

ملاحظه می‌شود که مقدار انحراف متوسط از میانگین چندان تفاوتی با ۰/۱۰ ندارد؛ در صورتی که از ۲ بیش از ۸ برابر کمتر است. این انحراف بزرگ تأثیر بسیار کمی بر روی شاخص گذاشته است. علاوه بر این، انحراف متوسط از میانگین، برخی از خواص ساده و مطلوب میانگین حسابی مانند خاصیت جمع شاخصها را نیز فاقد است. در نتیجه از نظر کاربردی شاخص مناسبی نیست و کاربرد آن بسیار محدود است و کمتر در آمار و ریاضی مورد استفاده قرار می‌گیرد.

۲-۳-٤ واریانس و انحراف معیار[1]

یکی دیگر از شاخصهای اندازه گیری پراکندگی داده‌ها نسبت به میانگین، واریانس است. هنگام محاسبهٔ انحراف متوسط از میانگین، برای جلوگیری از خنثی‌شدن انحرافات منفی از قدر مطلق استفاده می‌شود. چنانکه گفته شد این روش دارای اشکالات ویژه‌ای است. به منظور رفع آنها می‌توان انحرافات را مجذور کرد. چنانچه از مجذور انحرافات استفاده شود، شاخصی جدید برای سنجش پراکندگی پدید می‌آید که آن را واریانس می‌نامند. اگر $x_۱$، $x_۲$، ... ، x_N مشاهدات جامعهٔ آماری باشند، در این صورت واریانس جامعه چنین خواهد بود:

$$\sigma_x^۲ = \frac{\sum (x_i - \mu_x)^۲}{N} \qquad (۲\text{-}۱۵)$$

مثال ۲-۲۱ سود شرکت ایزایران طی ٦ ماه گذشته به این صورت بوده است:

مقدار سود به ده میلیون ریال: ۲۰ , ۱۳ , ۲۷ , ۲۵ , ۲۰ , ۱۵

برای به دست آوردن واریانس، ابتدا میانگین مشاهدات را حساب می‌کنیم:

$$\mu_x = \frac{\sum x_i}{N} = \frac{۱۲۰}{٦} = ۲۰$$

با استفاده از فرمول، واریانس عبارت است از:

$$\sigma_x^۲ = \frac{(۱۵-۲۰)^۲+(۲۰-۲۰)^۲+(۲۵-۲۰)^۲+(۲۷-۲۰)^۲+(۱۳-۲۰)^۲+(۲۰-۲۰)^۲}{٦} = ۲٤/٦٦۷$$

چنانچه رابطهٔ ۲-۱۵ را بسط دهیم به این رابطهٔ مهم از واریانس خواهیم رسید:

$$\sigma_x^۲ = \frac{۱}{N}\left[\sum (x_i - \mu_x)^۲\right]$$

$$= \frac{۱}{N}\left[\sum (x_i^۲ + \mu_x^۲ - ۲\mu_x . x_i)\right]$$

$$= \frac{۱}{N}\left[\sum x_i^۲ + N . \mu_x^۲ - ۲\mu_x . \sum x_i\right]$$

1. standard deviation

از آنجا که $\sum x_i = N . \mu_x$ است؛ پس خواهیم داشت:

$$= \frac{1}{N} [\sum x_i^2 + N . \mu_x^2 - 2N . \mu_x^2]$$

$$= \frac{1}{N} [\sum x_i^2 - N . \mu_x^2]$$

حال می‌توان تعریف دیگری از واریانس به صورت رابطهٔ ۱۶ـ۲ ارائه داد:

$$\sigma_x^2 = \frac{\sum x_i^2}{N} - \mu_x^2 \qquad (۲\text{ـ}۱۶)$$

زمانی که مشاهدات به صورت مقادیر صحیح تعریف شده باشند، از رابطهٔ ۱۶ـ۲ استفاده خواهد شد.

مثال ۲۲ـ۲ برای مشاهدات ۲۱ـ۲ مجدداً واریانس را با استفاده از رابطهٔ ۱۶ـ۲ محاسبه می‌کنیم:

$$\sigma_x^2 = \frac{(15)^2 + (20)^2 + (25)^2 + (27)^2 + (13)^2 + (20)^2}{6} - (20)^2 = 24/667$$

می‌بینیم که واریانس در هر دو روش مساوی است.

واحد اندازه گیری واریانس، مجذور واحد اصلی متغیراست؛ بنابراین اگر X برحسب ریال اندازه گیری شده باشد، واحد واریانس جامعه ریال به توان دو خواهد بود. از آنجا که واریانس به صورت مجذور واحد اولیه محاسبه می‌شود تفسیر آن زیاد ساده نیست. بعلاوه پراکندگی، مفهوم مشکلتر و نامأنوستری نسبت به مرکزیت است. با این حال این مسأله را با جذر گرفتن از واریانس حل کرده، σ_x را «انحراف معیار جامعه» می‌نامیم. بدین ترتیب معیاری از پراکندگی به دست می‌آید که با همان واحد مشاهدات اصلی بیان می‌شود.

مثال ۲۳ـ۲ انحراف معیار مشاهدات مربوط به سود شرکت ایزایران در مثال ۲۱ـ۲ به این صورت به دست می‌آید:

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{24/667} = 4/967$$

مقدار به دست آمده پراکندگی، میزان تغییرات سود شرکت را نسبت به میانگین به طورمتوسط ۴/۹۶۷ ده‌میلیون ریال نشان می‌دهد. انحراف معیاررا می‌توان دراین مثال ریسک سود نامید.

از آنجا که شاخص واریانس از خواص ریاضی متعددی برخوردار است، استفاده از آن نسبت به انحراف متوسط از میانگین آسانتر می‌نماید؛ بعضی از خواص واریانس عبارتند[1] از:

خاصیت اول. اگر به هریک از مشاهدات، مقدار ثابت a اضافه شود، واریانس مشاهدات جدید تغییر نمی‌کند؛ یعنی اگر:

$$y_i = x_i + a \Rightarrow \sigma_y^2 = \sigma_x^2$$

اثبات: از تعریف σ_y^2 استفاده می‌کنیم:

$$\begin{aligned}\sigma_y^2 &= \frac{\sum (y_i - \mu_y)^2}{N} \\ &= \frac{\sum (x_i + a - \mu_x - a)^2}{N} \\ &= \frac{\sum (x_i - \mu_x)^2}{N} \\ &= \sigma_x^2\end{aligned}$$

خاصیت دوم. اگر هریک از مشاهدات جامعهٔ آماری در عدد ثابت b ضرب شوند، واریانس bx معادل b^2 برابر واریانس قبلی خواهد بود؛ یعنی اگر:

$$y_i = bx_i \Rightarrow \sigma_y^2 = b^2\sigma_x^2$$

اثبات:

$$\begin{aligned}\sigma_y^2 &= \frac{\sum (y_i - \mu_y)^2}{N} \\ &= \frac{\sum (bx_i - b\mu_x)^2}{N} \\ &= \frac{\sum [b^2(x_i - \mu_x)^2]}{N} = \frac{b^2 \sum (x_i - \mu_x)^2}{N} \\ &\Rightarrow \sigma_y^2 = b^2\sigma_x^2\end{aligned}$$

---

۱. فقط خواص ضروری در این فصل طرح می‌شود. واریانس دارای خواص دیگری است که بر حسب ضرورت در فصول بعدی مطرح خواهد شد.

از این خواص برای ساده کردن محاسبهٔ واریانس استفاده می‌شود. گاهی اتفاق می‌افتد که مشاهدات جمع‌آوری شده اشتباهاً مورد استفاده قرار گرفته‌اند که در آن صورت می‌توان به کمک خواص مذکور بدون مراجعه مجدد به مشاهدات اولیه، مقدار واریانس واقعی را محاسبه نمود.

مثال ۲ـ۲۴ با داشتن $x_i = -۱, ۴, ۳, ۶$ ، $a = -۲$ ، $b = ۳$ صحت خواص واریانس را به ترتیب بررسی می‌کنیم. با توجه به خاصیت اول داشتیم که اگر $y_i = x_i + a$ باشد؛ در نتیجه $\sigma_y^۲ = \sigma_x^۲$ خواهد بود.

x_i	-۱	۴	۳	۶	$\sum x_i = ۱۲$ و $\mu_x = ۳$
$(x_i - \mu_x)^۲$	۱۶	۱	۰	۹	$\sum (x_i - \mu_x)^۲ = ۲۶$ و $\sigma_x^۲ = \frac{۲۶}{۴} = ۶/۵$
$y_i = x_i - ۲$	-۳	۲	۱	۴	$\sum y_i = ۴$ و $\mu_y = ۱$
$(y_i - \mu_y)^۲$	۱۶	۱	۰	۹	$\sum (y_i - \mu_y)^۲ = ۲۶$ و $\sigma_y^۲ = ۶/۵$

نتیجه آنکه در این حالت واریانس x و واریانس y با یکدیگر مساوی هستند.

با توجه به خاصیت دوم داشتیم که اگر $y_i = bx_i$ ؛ در نتیجه $\sigma_y^۲ = b^۲ \sigma_x^۲$ خواهد بود.

x_i	-۱	۴	۳	۶	$\sum x_i = ۱۲$
$y_i = ۳x_i$	-۳	۱۲	۹	۱۸	$\sum y_i = ۳۶$ و $\mu_y = ۹$
$(y_i - \mu_y)^۲$	۱۴۴	۹	۰	۸۱	$\sum (y_i - \mu_y)^۲ = ۲۳۴$ و $\sigma_y^۲ = ۵۸/۵$

نتیجه آنکه:

$$\sigma_y^۲ = b^۲ \sigma_x^۲$$

$$۵۸/۵ = (۳)^۲(۶/۵)$$

واریانس و انحراف معیار به عنوان شاخص پراکندگی فاقد نقص انحراف متوسط از میانگین هستند. این شاخصها برخلاف انحراف متوسط از میانگین، انحرافات بسیار

بزرگ را نیز نشان می‌دهند؛ زیرا این گونه انحرافات به توان دوم می‌رسند و مجذور آنها نسبت به مجذور انحرافات کوچک، در مقایسه با انحرافات درجهٔ اول آنها، بیشتر خواهد شد و به همین دلیل نقش انحرافات بزرگ در واریانس افزایش می‌یابد.

۲ـ۳ـ۵ نیمه واریانس[1]

نیمه واریانس یکی دیگر از شاخصهای پراکندگی است که معمولاً برای استخراج انحرافات نامناسب و نامطلوب به کار می‌رود. در داده‌های مربوط به سود، مقادیر کوچکتر از میانگین و در داده‌های مربوط به زیان مقادیر بزرگتر از میانگین نامطلوب هستند. بنابراین نیمه واریانس عبارت است از متوسط مجذور مقادیر نامطلوب. اگر x_1، x_2، ...، x_N معرف مقدار مشاهدات باشد، نیمه واریانس به این صورت تعریف می‌شود:

$$S.V = \frac{\sum_{i=1}^{K}(x_i - \mu_x)^2}{K} \qquad (۲\text{ـ}۱۷)$$

فرض بر آن است که تعداد مشاهدات جامعه N تاست و فقط K تا (K<N) از آنها نامطلوب است.

مثال ۲ـ۲۵ سود ده ماه شرکت سرم‌سازی به این صورت است:

$$x_i = ۰ , ۱ , ۵ , ۷ , ٤ , ۳ , ۲ , ۸ , ٦ , ٤$$

برای محاسبه نیمه واریانس این مشاهدات ابتدا میانگین آنها را محاسبه می‌کنیم:

$$\mu_x = \frac{\sum x_i}{N} = \frac{٤۰}{۱۰} = ٤$$

با توجه به اینکه داده‌ها مربوط به سود است فقط انحرافات کوچکتر از میانگین نامطلوب هستند؛ بنابراین مطابق فرمول نیمه واریانس خواهیم داشت:

$$S.V = \frac{(۰-٤)^2 + (۱-٤)^2 + (۳-٤)^2 + (۲-٤)^2}{٤} = ۷/۵$$

1. semivariance (S. V.)

در این مثال مقدار واریانس مساوی با ۶ خواهد بود:

$$\sigma_x^2 = \frac{\sum_{i=1}^{10}(x_i - \mu_x)^2}{10} = \frac{60}{10} = 6$$

۲-۳-۶ ضریب پراکندگی[1]

در بسیاری از تحقیقات لازم است که برای توصیف داده‌ها، پراکندگی به صورت کسری از میانگین بیان شود. یکی از معیارهای پراکندگی نسبی که ضریب پراکندگی (C.V) نامیده می‌شود به این صورت تعریف می‌گردد:

$$C.V = \frac{\sigma_x}{\mu_x} \qquad (2\text{-}18)$$

ضریب پراکندگی کاربردهایی دارد که واریانس و انحراف معیار فاقد آنها هستند. کاربرد اول در جایی است که در دو یا چند جامعهٔ آماری مورد مقایسه، مشاهدات ناهمگون و نامتجانس وجود داشته باشد؛ برای مثال چنانچه پراکندگی یک جامعه بر حسب متر و پراکندگی دیگری بر حسب کیلوگرم بیان شود، این پراکندگیها به هیچ وجه قابل مقایسه نیستند و برای اینکه قابل مقایسه شوند باید اندازه آنها به صورت کمیّت مجرد بیان گردد. گاهی نیز مقیاس صفت مورد اندازه گیری در دو جامعه یکسان است، ولی بزرگی مشاهدات آنها به طور قابل ملاحظه‌ای تفاوت دارد؛ برای مثال اگر بخواهیم پراکندگی سود و زیان را در صنایع دستی با صنایع سنگین مقایسه کنیم، چاره‌ای جز استفاده از ضریب پراکندگی نخواهیم داشت. کاربرد دوم در جایی است که واریانس دو جامعه یکسان، ولی میانگینهای آنها متفاوت باشد. بنابراین واضح است که انحراف معیار بدون در نظرگرفتن میانگین نمی تواند بیان‌کنندهٔ پراکندگی جامعه باشد و پراکندگی در دو جامعه در مقایسه با یکدیگر در صورتی مفهوم پیدا می‌کند که نسبت به میانگینشان سنجیده شوند.

مثال ۲-۲۶ محققی می‌خواهد بررسی کند متوسط سود شرکتهای کدامیک از دو صنعت x و y متجانس تر است. بدین جهت متوسط سود شرکتهای دو صنعت را طی چهار سال گذشته دریافت کرده است (ارقام به ده میلیون ریال):

$$19, 12, 17, 12 \qquad \mu_x = 15 \qquad \sigma_x^2 = 9/5$$

$$60, 55, 80, 85 \qquad \mu_y = 70 \qquad \sigma_y^2 = 162/5$$

---

1. coefficient of variation

میانگینها نابرابرند از این‌رو ضریب پراکندگی را حساب می‌کنیم:

$$C.V = \frac{\sigma_x}{\mu_x} = \frac{۳/۰۸}{۱۵} = ۰/۲۰۵$$

$$C.V = \frac{\sigma_y}{\mu_y} = \frac{۱۲/۷۵}{۷۰} = ۰/۱۸۲$$

اگر انتظار اولیه آن بود که متوسط سود شرکتهای صنعت x متجانس‌تر است ولی با محاسبه ضریب پراکندگی بدین نتیجه می‌رسیم که متوسط سود شرکتهای صنعت y از پراکندگی کمتری برخوردار است.

تمرین

۱. تفاوت واریانس با نیمه واریانس در چیست؟

۲. موارد کاربرد انحراف چارکی را بیان کنید.

۳. این مشاهدات نشان‌دهنده وزن محصولات تولید شده دریک کارخانه طی یک روزاست:

$$x_i = ۱۰۰ , ۱۵۰ , ۱۲۰ , ۱۳۰ , ۱۴۰ , ۱۳۵ , ۱۴۰ , ۱۲۵ , ۱۳۵ , ۱۴۰$$

الف) انحراف چارکی را محاسبه نمایید.

ب) انحراف معیار را محاسبه و با انحراف چارکی مقایسه کنید.

۴. میزان مهارت مدیران یک سازمان در برنامه‌ریزی به این صورت است:

$$x_i = ۸۰ , ۸۵ , ۶۰ , ۶۵ , ۵۰ , ۴۰ , ۴۵$$

الف) نیمه واریانس را محاسبه نمایید.

ب) انحراف متوسط از میانگین را محاسبه کنید.

۲ـ۴ خلاصه

در این فصل عمدتاً پارامترهای مرکزی مانند میانگین، میانه و مد برای توصیف داده‌های طبقه‌بندی نشده معرفی شدند. این پارامترها علی‌رغم اهمیت و نقش آنها در توصیف جامعهٔ آماری نواقص خاص خود را دارند؛ از جمله اینکه در مقایسه جوامع آماری اغلب ناتوانند.

برای بررسی بیشتر جامعه، مبحثی دیگر از آمار توصیفی با عنوان شاخصهای پراکندگی مطرح شد. این پارامترها علی‌رغم پارامترهای مرکزی که به منظور اندازه گیری مرکز ثقل مشاهدات به کار می‌روند، میزان پراکندگی و نوسان داده‌ها را نشان می‌دهند.

ساده‌ترین شاخصهای پراکندگی دامنهٔ تغییرات (R) ، دامنهٔ میان چارکی (IQR) و انحراف چارکی (SIQR) هستند که به دلیل نقصهای ویژه کمتر در توصیف جامعه مورد استفاده قرار می‌گیرند. مهمترین مشکل این شاخصها آن است که بیان‌کنندهٔ تغییرات تمامی داده‌ها نیستند. در این نوع پارامترها تنها نماینده‌هایی از مشاهدات مانند Q_1 و Q_3 برای بیان تغییرات برگزیده می‌شوند. بعلاوه این شاخصها فاقد خواص ساده جبری هستند؛ به عبارت دیگر روی آنها همانند واریانس و میانگین نمی‌توان عملیات جبری انجام داد.

انحراف متوسط از میانگین ($A.D_{\mu}$) یکی دیگر از پارامترهای پراکندگی است که مشکلات ریاضی خاصی دارد، مانند اینکه نمی‌توان از آن مشتق گرفت و فاقد خواص جبری است؛ به همین دلیل استفاده از آن محدود است.

مهمترین پارامتر پراکندگی، انحراف معیار (σ_x) است که از جذر واریانس (σ_x^2) به‌دست می‌آید. این شاخص، نشان‌دهندهٔ متوسط نوسان مشاهدات از میانگین آنهاست. انحراف معیار به دلیل دارا بودن خواص ریاضی، استفاده‌های فراوانی چه در عمل و چه در آمار ریاضی دارد که در فصل سوم با آنها آشنا خواهید شد.

در مواردی که تصمیم‌گیری به کمک انحراف معیار امکان‌پذیر نباشد، از شاخصی با عنوان ضریب پراکندگی برای مقایسهٔ جوامع آماری ناهمگون یا دارای واریانس مساوی استفاده می‌شود.

۲.۵ سؤالات و مسائل

سؤالات دوگزینه‌ای

۱. مهمترین پارامتر مرکزی میانه است. ص ☐ غ ☐

۲. همهٔ توزیعهای آماری دارای مد هستند. ص ☐ غ ☐

۳. میانه همان چارک دوم است. ص ☐ غ ☐

٤. پارامترهای پراکندگی؛ شاخصهایی برای اندازه‌گیری تغییرات مشاهدات جامعه هستند. ص ☐ غ ☐

۵. انحراف چارکی شاخص مناسبی برای اندازه‌گیری پراکندگی توزیعهای نامتقارن است. ص ☐ غ ☐

٦. نیمهٔ واریانس، شاخصی برای اندازه‌گیری تغییرات مطلوب نسبت به میانگین است. ص ☐ غ ☐

۷. واحد واریانس، مربع واحد اندازه‌گیری مشاهدات است. ص ☐ غ ☐

۸. همیشه نیم دامنه از دامنهٔ تغییرات بزرگتر است. ص ☐ غ ☐

۹. همیشه $\mu_G \le \mu_x \le \mu_H$ است. ص ☐ غ ☐

۱۰. در مفهوم مالی، انحراف معیار همان ریسک است. ص ☐ غ ☐

سؤالات چهارگزینه‌ای

۱۱. اگر میانگین x_1، x_2، ...، x_N برابر μ_x و میانگین y_1، y_2، ...، y_k مساوی μ_y باشد و داشته‌باشیم $\mu_y = a\mu_x$ در آن صورت مقدار $\frac{\sum x_i}{\sum y_i}$ کدام است؟

الف) $\frac{N}{Ka}$ ج) $N.\mu_x$

ب) $N.a$ د) $N.\mu_y$

۱۲. اگر میانگین x_1، x_2، ...، x_N مساوی μ_x باشد، مقدار $\sum(x_i - \mu_x)$ کدام است؟

الف) $N.\mu_x$ ج) N

ب) صفر د) یک

۱۳. معدل یک دانشجو در هفت واحد ۱۳ است. اگر نمره ۱٦ را از نمرات او حذف کنیم، معدل این دانشجو کدام است؟

الف) ۱۲/۵ ج) تغییر نمی‌کند.

ب) ۱۰ د) به ضریب ۱٦ بستگی دارد.

۱٤. اگر μ_x میانگین x_1، x_2، ...، x_N باشد، میانگین $(-\frac{x_1}{۲}+۳)$، $(-\frac{x_۲}{۲}+۳)$، ...، $(-\frac{x_N}{۲}+۳)$ کدام است؟

الف) $-\mu_x + ۳$ ج) $-\frac{۱}{۲}\mu_x + ۳$

ب) $\mu_x + ۳$ د) $\frac{۱}{۲}\mu_x$

۱۵. اگر از هریک از مشاهدات سه واحد کم کنیم، در انحراف معیار مشاهدات چه وضعی پیش می‌آید؟

الف) ثابت می‌ماند. ج) ۹ واحد کاسته می‌شود.

ب) سه واحد کاسته می‌شود. د) هیچ کدام.

۱۶. اگر واریانس مقادیر x_1، x_2، ...، x_N برابر ۱۶ باشد، انحراف معیار $\frac{x_1}{4}$، $\frac{x_2}{4}$، ...، $\frac{x_N}{4}$ کدام است؟

الف) ٤ ج) ١٦

ب) ٢ د) ١

۱۷. اگر ۶۰ = $\sum x_i$، ۴۰۰ = $\sum x_i^2$ و ۱۰ = N باشد، ضریب پراکندگی کدام است؟

الف) ۰/۴۰ ج) ۰/۷۰

ب) ۰/۳۳ د) ۰/۶۲

۱۸. مشاهدات ۹ ، ۹ ، ۹ ، ۷ ، ۷ ، ۷ ، ۵ ، ۵ ، ۵ ، ۲ ، ۲ ، ۲ را در نظر بگیرید. مد کدام است؟

الف) ۹ ، ۷ ، ۵ ، ۲ ج) صفر است.

ب) مد ندارد. د) ۸/۶۷

۱۹. نسبت قیمت برنج به ذرت طی سه سال ۶ ، ۴ و ۵ بوده است، میانگین این نسبتها کدام است؟

الف) ۵ ج) ۴/۸۶

ب) ۴/۹۳ د) ۱۰/۹۵

۲۰. اگر ۱۰ = μ_x و ۲۲ = μ_y و $Z = y - x$ باشد، μ_z کدام است؟

الف) ۳۲ ج) ۱۲

ب) ۱۶ د) ۶

۲۱. چارک اول ، دوم و سوم یک جامعهٔ آماری به ترتیب ۶۰ ، ۷۰ و ۱۰۰ شده است، مقدار انحراف چارکی چقدر است؟

الف) ۴۰ ج) ۲۰

ب) ۳۰ د) ۱۵

۲۲. کدام یک از این پارامترها بیشتر تحت تأثیر انحرافات بزرگ است؟

الف) انحراف متوسط از میانگین ج) انحراف چارکی

ب) واریانس د) نیم‌دامنه

۲۳. مشاهدات ۷ , ۵ , ۳ , ۱ را در نظر بگیرید. اگر انحراف بالای میانگین نامطلوب باشد، نیمه واریانس کدام است؟

الف) ٤ ج) ٥

ب) ۳ د) صفر

۲٤. با بررسی پرسشنامه‌های طرح هزینه و درآمد خانوار شهری در چهار سال متوالی، معلوم شد قیمت نفت سفید یک خانوار به ترتیب ۱/۶، ۱/۸، ۲/۱ و ۲/۵ ریال در لیتر است. اگر خانواری برای هر سال ۲۰ هزار ریال هزینه در نظر بگیرد، متوسط مصرف سوخت سالانه این خانوار برحسب ریال در لیتر چقدر است؟

الف) ۱/۷۵ ج) ۲/۲۵

ب) ۲/۵ د) ۱/۹٤

۲۵. با فرض دراختیارداشتن $\sum_{j=1}^{N}|x_j - a|$ به شرط آنکه a میانه باشد، همواره حاصل این عبارت است:

الف) «حداکثر». ج) مقداری بین «حداقل» و «حداکثر».

ب) «حداقل». د) نامشخص.

مسائل

۲۶. حاصل این عبارات را با استفاده از خواص زیگما بنویسید.

الف) $\sum_{i=1}^{5}(x_i + y_i + z_i)$ ج) $\sum_{i=1}^{3} 4$

ب) $\sum_{i=1}^{4}(x_i - 2)^4$ د) $\sum_{i=1}^{5} x_i^{y_i}$

۲۷. مقادیر $x = 3, -2, 1, -1$ و $y = -3, 2, 4, -1$ را در نظر بگیرید. حاصل این عبارات را محاسبه کنید.

الف) $\sum_{i=1}^{3}(x_i - 3)$ د) $\sum_{i=1}^{4}(y_i^2 - 4y_i + 4)$

ب) $\sum_{i=1}^{3}(x_i y_i)$ ه) $\sum_{i=1}^{3} y_i^3$

ج) $\sum_{i=1}^{4}\sum_{j=1}^{2} x_i y_j$ و) $(\sum_{i=1}^{3} y_i)^3$

۲۸. اگر $\sum x_i = 10$، $\sum y_i = 10$، $\sum x_i y_i = -4$، $\sum x_i^2 = 40$ و $\sum y_i^2 = 60$ باشد، حاصل این عبارات را تعیین کنید.

الف) $(\sum x_i^2)(\sum x_i)^2$ ج) $\sum (x_i - y_i)^2$

ب) $(\sum x_i y_i)(\sum x_i)^3$ د) $\sum (y_i - 2)^2$

۲۹. سود شرکتی در سال ١٣٦٦ نسبت به ١٣٦٥ سه برابر، در سال ١٣٦٧ نسبت به ١٣٦٦ دو برابر و در سال ١٣٦٨ نسبت به ١٣٦٧ چهار برابر شده است. حسابدار شرکت معتقد است که سود شرکت در پایان سال ١٣٦٨ نسبت به ١٣٦٥، ۲۷ برابر شده، آیا این ادعا درست است؟ چرا؟

۳۰. مقدار کار چهار کارگر در کارخانه آجرسازی در روزهای مختلف اندازه‌گیری شده است. کارگر اولی یک مترمربع (١٦ عدد) خشت را به‌طورمتوسط در ۲/۳۰ دقیقه، دومی در دو دقیقه، سومی در سه دقیقه و چهارمی در چهار دقیقه آماده می‌کند. اگر این چهار کارگر با هم کار کنند، به‌طور متوسط یک متر مربع خشت را در چه مدتی آماده می‌کنند؟

۳۱. شاخص قیمت پانزده نوع کالای مختلف در سال ١٣٦١، در این جدول آمده است:

شاخص قیمت	١٢٠	١٣٠	١٣٥	١٣٧	١٤٠
تعداد	٢	٥	٣	٤	١

متوسط شاخص قیمت را محاسبه کنید؟

۳۲. اگر $x_i = 2, 5, 3, 6$، $z_i = -2, 3, 1, 5$، $a = 4$ و $b = \frac{1}{4}$ باشد؛ صحت این روابط را بررسی نمایید.

الف) $\mu_{x+z} = \mu_x + \mu_z$ ج) $\mu_{bx} = b\mu_x$

ب) $\mu_{x+a} = \mu_x + a$ د) $\sum (z_i - \mu_z) = 0$

۳۳. اگر $y_i = x_i - b$ باشد، صحت این روابط را بررسی کنید. (اطلاعات مسأله ۳۲ را در نظر بگیرید).

الف) $\mu_y = \mu_x - b$ ب) $\sigma_y^2 = \sigma_x^2$

۳۴. مشاهدات ۲۵ ، ٥ ، ٥ ، ٥ ، ٥ ، ٥ ، ٥ ، ٥ ، ٥ ، ٥ ، ٥ بیان کنندهٔ وزن یک کالای خاص برحسب کیلوگرم است؛ ضمن محاسبهٔ انحراف متوسط از میانگین و واریانس نقاط قوت و ضعف آنها را ذکر کنید.

۳۵. جمعیت کشور طی چهار سال به ترتیب ۴۵، ۵۰، ۵۳ و ۶۰ میلیون نفر بوده است؛ متوسط میزان رشد جمعیت را محاسبه کنید.

۳۶. هزینهٔ ماهانه یک خانواده تهرانی به این صورت به دست آمده است:

هزینه برحسب صد هزار ریال: ۱۴/۵، ۱۵، ۳۰، ۲۵، ۱۸، ۱۷، ۱۶، ۹، ۲۰، ۱۵، ۸، ۱۰

الف) میانگین و واریانس را محاسبه کنید.

ب) میانگین پیراسته و وینزوری را در صورتی که ۲۵٪ = LN ٪ باشد، محاسبه کنید.

۳۷. فرض کنید خانواده مورد نظر در مسأله ۳۶ فراموش نموده است که هزینهٔ مهد کودک فرزند خود را که ماهانه ۶۰ هزار ریال بوده است، گزارش کند. اثر این هزینه بر میانگین و واریانس مشاهدات اولیه چیست؟ توضیح دهید.

۳۸. تولیدکننده‌ای ناچار است که سه نوع مادهٔ اولیه را با عنوان الف، ب و ج با همدیگر تلفیق کند تا یک محصول ساخته گردد. قیمت هر کیلو مادهٔ اولیه الف ۴ هزار، ب ۲ هزار و ج هزار ریال است. هر واحد از این محصول مجموعاً به یک کیلو مادهٔ اولیه نیاز دارد.

الف) قیمت محصول را در صورتی که مادهٔ اولیه مورد نیاز آن ۳۰ درصد از نوع الف، ۵۰ درصد از نوع ب و ۲۰ درصد از نوع ج فراهم شده باشد، محاسبه کنید.

ب) قیمت محصول را در صورتی که مواد اولیه الف و ب و ج به نسبت یکسان ترکیب شده باشند، محاسبه کنید.

۳۹. مقاومت شکست برای دو نوع بطری بر حسب پوند بر اینچ مربع (PSI) به این صورت اندازه گیری شده است:

نوع اول: ۲۷۰، ۲۱۰، ۲۶۵، ۲۳۰، ۲۴۰، ۲۶۰

نوع دوم: ۲۴۰، ۲۶۵، ۲۴۰، ۲۶۰، ۲۶۵، ۲۴۰، ۳۰۵، ۲۲۸، ۱۹۰

الف) پارامترهای مرکزی را محاسبه نمایید.

ب) انحراف چارکی هر دو جامعه را با یکدیگر مقایسه کنید؛ کدام یک از کیفیت بهتری برخوردار است؟

ج) انحراف متوسط از میانگین هر دو نوع بطری را با یکدیگر مقایسه کنید؛ کدام نوع بهتر است؟

د) انحراف معیار هر دو نوع بطری را با یکدیگر مقایسه کنید؛ کدام نوع بهتر است؟

ه) ضریب پراکندگی مقاومت شکست را برای دو نوع بطری محاسبه نموده، نتیجه را با نتایج بندهای ب، ج و د مقایسه کنید. آیا در تصمیم‌گیری اختلافی حاصل شده است؟ چرا؟

پاسخنامه سؤالات

۱) غ	۲) غ	۳) ص	۴) ص
۵) ص	۶) غ	۷) ص	۸) غ
۹) غ	۱۰) ص	۱۱) الف	۱۲) ب
۱۳) د	۱۴) ج	۱۵) الف	۱۶) د
۱۷) ب	۱۸) ب	۱۹) ب	۲۰) ج
۲۱) ج	۲۲) ب	۲۳) ج	۲۴) د
۲۵) ب			

فصل سوم

طبقه‌بندی و توصیف هندسی مشاهدات جامعهٔ آماری

هدف علم آمار، استنتاج از تعداد زیادی مشاهدات است که از جامعه به دست می‌آیند؛ بنابراین اولین سؤالی که به ذهن هر پژوهشگری می‌رسد این است که آیا می‌توان این مجموعه عظیم از مقادیر را توصیف کرد و چگونه می‌توان مشاهدات زیاد جامعه را در یک چهارچوب مشخص سازماندهی نمود. متون آماری فراوانی در این باب تدوین شده است که روشهای مختلفی را برای تنظیم و سازماندهی داده‌ها و توصیف آنها ارائه می‌دهد. این روشها به روشهای هندسی[1] و روشهای مقداری[2] تقسیم می‌شوند.

در فصل دوم به بیان مهمترین روشهای مقداری پرداخته شد. البته در آن فصل صرفاً به بحث پیرامون روشهای مقداری در زمینه شاخصهای مرکزی و پراکندگی در داده‌های طبقه‌بندی نشده ($N \leq ۲۰$) اکتفا شد. در فصل چهارم نیز به تفصیل روشهای مقداری برای داده‌های طبقه‌بندی شده بیان خواهد شد. در این فصل به بیان روشهای طبقه‌بندی و خلاصه کردن مشاهدات جامعه با تکیه بر جنبه‌های مختلف سؤالات تحقیق پرداخته خواهد شد و سپس عمده‌ترین روشهای هندسی که برای توصیف داده‌ها به کار می‌روند، مطرح و تشریح می‌گردند. روشهای هندسی هم قابل استفاده برای توصیف کل داده‌های جامعه هستند و هم در توصیف نمونه‌ای از جامعه مورد استفاده قرار می‌گیرند.

۳ـ۱ طبقه‌بندی و سازماندهی مشاهدات

منظم کردن داده‌ها، نخستین قدم در مواجه شدن با حجم عظیمی از مشاهدات جمع‌آوری شده است. قدم بعدی خلاصه کردن داده‌هاست؛ به صورتی که اطلاعات مهم برای

1. graphical methods   2. numerical methods

بررسی جدا و اطلاعات کم اهمیت کنار گذاشته شود.

از آنجا که بیشتر تحقیقات با حجم عظیمی از داده‌های جمع‌آوری‌شده، همراهند، به منظور درک بهتر و بیشتر این داده‌های خام لازم است آنها را سازماندهی و خلاصه کنیم. خلاصه کردنِ خصوصیات عمدهٔ داده‌ها، انبوه مشاهدات را به شکل قابل تفسیری درمی‌آورد. سازماندهی مشاهدات را در آمار، توزیع فراوانی[1] گویند. توزیع فراوانی به منزلهٔ وسیله‌ای مناسب برای خلاصه کردن و مشخص نمودن ویژگیهای اصلی داده‌های خام تحقیق است. توزیع فراوانیِ تعدادی از داده‌ها عبارت است از جدول مرتب شدهٔ مقادیر آن داده‌ها که تکرار وقوع هر داده در آن مشخص شده است؛ به عبارت ساده‌تر، توزیع فراوانی جدولی خلاصه شده از داده‌های جمع‌آوری شدهٔ جامعهٔ آماری است.

اگرچه خلاصه کردن داده‌ها محاسن ویژه‌ای دارد بدیهی است در اثر آن بخشی از اطلاعات و داده‌های جمع‌آوری شده از دست می‌رود؛ از این رو باید تمهیداتی فراهم نمود که داده‌های سازماندهی شده حتی‌الامکان ویژگیهای اصلی خود را حفظ نموده، نشان‌دهنده خواص جامعه باشند. با به کارگیری فرایندی خاص می‌توان به این هدف رسید. حال به ذکر مراحل طبقه‌بندی داده‌ها به کمک مثال ۱ـ۳ می‌پردازیم.

مثال ۱ـ۳ عموماً سازمانها در تلاشند مجموعهٔ سرمایه‌گذاری[2] خود را به گونه‌ای حفظ کنند که حداکثر درآمد و حداقل ریسک را برای آنها داشته باشد. یکی از شاخصهایی که نشان‌دهنده درآمد و ریسک بالقوه یک سهم است، نسبت قیمت به درآمد[3] ($\frac{P}{E}$) است. اصولاً سهام با $\frac{P}{E}$ کمتر نسبت به سهام با $\frac{P}{E}$ بیشتر ترجیح دارند. داده‌های ذیل نشان‌دهنده نرخ $\frac{P}{E}$ سهام ۲۵ شرکت در بازار بورس تهران هستند.

٩/٩	٢٤/١	١٥/٦	١٩/٥	٢٠/٥
٢٨/٦	١٧/٠	٥/٠	١٢/٧	١٥/٤
١٨/٤	١١/٨	٢٣/٣	٧/٨	١٦/٩
١٦/٨	٩/٢	١٩/٢	١٤/٣	١٣/٤
١٥/٩	١٢/٦	٢٠/٨	٢٢/١	٨/٨

1. frequency distribution
2. portfolio
3. price-earnings ratio

با بررسی اجمالی این اعداد معلوم می‌شود که بزرگترین و کوچکترین مشاهده ۲۸/۶ و ۵ است؛ اما سؤال اینجاست که ۲۳ عدد باقیمانده چگونه بین این دو مقدار توزیع شده‌اند. برای پاسخ به این سؤال باید فاصلهٔ بین دو مقدار حداکثر و حداقل را به فاصله‌های کوچکتر تقسیم کنیم. به تفاضل بین مقادیر حداکثر و حداقل جامعه، دامنه تغییرات گفته می‌شود که یکی از شاخصهای پراکندگی است. در اینجا فاصله‌های کوچکتر طبقه[1] نامیده می‌شود. به عنوان یک قاعدهٔ سرانگشتی، تعداد طبقات بهتر است بین پنج تا بیست انتخاب شوند؛ طبیعی است مشاهدات بیشتر، طبقات بیشتری را ایجاب می‌نماید. طبقات باید به گونه‌ای تعریف شوند که نسبت به همدیگر مانعةالجمع[2] باشند؛ یعنی هر داده فقط و فقط در یک طبقه قرار گیرد.

برای طبقه‌بندی داده‌های اخیر از شش طبقه استفاده می‌کنیم و سپس فاصلهٔ طبقات را مشخص می‌نماییم. فاصلهٔ طبقات[3] که در اینجا با I نمایش داده می‌شود عبارت است از:

$$I = \frac{R}{K} = \frac{\text{دامنهٔ تغییرات}}{\text{تعداد طبقات}} \qquad (۳\text{-}۱)$$

بنابراین برای مثال ما، فاصلهٔ طبقات به این صورت تعیین می‌گردد:

$$I = \frac{۲۸/۶ - ۵}{۶} = ۳/۹۳۳ \Rightarrow I = ۴$$

در نظر گرفتن فاصلهٔ طبقات به صورت اعشاری ـ دست کم تا یک رقم به اعشار ـ دقیقتر است، ولی برای راحتی کار سعی می‌شود به نزدیکترین مقدار صحیح گرد شود. فاصله طبقات براساس طول طبقه مشخص می‌گردد.

برای تشکیل جدول طبقه‌بندی، حد پایین اولین طبقه را با کوچکترین مشاهده (۵) آغاز می‌کنیم. حدود طبقات را با فاصلهٔ ۴ تشکیل می‌دهیم. حاصل این طبقه‌بندی در جدول ۱ـ۳ آمده است. در جدول ۱ـ۳، ستون اول نشان‌دهنده شماره طبقات است. ستون دوم نشان‌دهنده مرزهای هر طبقه است که به حدود طبقات[4] معروف است و با C_L نشان داده می‌شود. ستون خط و نشان، نشان‌دهنده تعداد مشاهداتی است که در فاصلهٔ «بستهٔ حد

1. class
2. mutually exclusive
3. classes interval
4. classes limits

پایین و بازحد بالای طبقه» قراردارند ـ برای راحتی شمارش، در هر دسته پس از چهارخط، پنجمی را با خط مورب قطع می‌کنیم (卌). ستون چهارم نشان‌دهندهٔ مجموع نشانه‌های خطی هر طبقه است که به آن ستون فراوانی مطلق[1] می‌گوییم و با F_i نشان داده می‌شود.

جدول ۱ـ۳ جدول اعداد طبقه‌بندی شده برای نرخهای $\frac{P}{E}$

طبقه i	حدود طبقات (C - L)	خط و نشان	فراوانی مطلق (F_i)
۱	۵ـ۹	|||	۳
۲	۹ـ۱۳	卌	۵
۳	۱۳ـ۱۷	卌 ||	۷
٤	۱۷ـ۲۱	卌		٦
۵	۲۱ـ۲۵	|||	۳
٦	۲۵ـ۲۹			۱

روشی که برای طبقه‌بندی داده‌ها در جدول ۱ـ۳ به کار رفته است «روش پیوسته» نام دارد. در این روش عرض طبقه[2] (تفاوت حد بالا و پایین طبقه) و طول طبقه[3] با فاصله طبقات مساوی خواهد بود.

تعیین تعداد طبقات به تعداد مشاهدات جامعه و بخصوص تجربهٔ آمارگر بستگی دارد. از آنجا که بسیاری از دانش‌پژوهان در این زمینه مبتدی هستند، ناچاریم در ادامه بحث از این روش کمتر استفاده کنیم و برای طبقه‌بندی داده‌ها الگوریتم مشخصی هرچند غیردقیق ارائه دهیم. با توجه به مثال ۱ـ۳ طبقه‌بندی داده‌ها دارای چهار مرحله است:

مرحله اول. داده‌های جمع‌آوری‌شده را مرتب و دامنهٔ تغییرات را محاسبه می‌کنیم.

$$R = \mathrm{Max}\,x_i - \mathrm{Min}\,x_i \qquad (۳ـ۲)$$

مرحلهٔ دوم. تعداد طبقات (K) را مشخص می‌سازیم. علی‌رغم اینکه برای تعیین تعداد طبقات هیچ‌گونه روش دقیق ریاضی وجود ندارد، برای محاسبهٔ آن می‌توان از این دو فرمول تجربی استفاده کرد. فرمول اول که به وسیلهٔ استورجس ارائه شده عبارت است از:

$$K = ۱ + ۳/۳۲ \,\mathrm{Log}\, N \qquad (۳ـ۳)$$

در این فرمول N تعداد داده‌های جمع‌آوری شده از جامعه است. فرمول دوم که باز یک

1. absolute frequency 2. class width 3. class lengths

روش تقریبی است ، عبارت است از:

$$K = \sqrt{N} \qquad (٣\_٤)$$

واقعیت این است که هیچ یک از این دو روش بر دیگری برتری ندارد و استفاده از آنها به سلیقهٔ آمارگر بستگی دارد.

مرحلهٔ سوم. فاصلهٔ طبقات را محاسبه می‌کنیم.

$$I = \frac{R}{K} \qquad (٣\_٥)$$

مرحلهٔ چهارم. سازماندهی داده‌ها را آغاز می‌کنیم. پس از مشخص شدن تعداد و فاصلهٔ طبقات نوع جدول طبقه‌بندی (سازماندهی) انتخاب می‌شود. نوع سازماندهی به نوع داده‌های جمع‌آوری شده بستگی دارد. اگر داده‌های جمع‌آوری شده، همراه با اعشار باشند، بهترین روش طبقه‌بندی، طبقه‌بندی پیوسته است. در این حالت طول و عرض طبقه با فاصلهٔ طبقات مساوی خواهد بود؛ یعنی جدول طبقه‌بندی داده‌ها از حد پایین اولین طبقه تا حد بالای آخرین طبقه پیوسته است (جدول ١ـ٣). طبیعی است در این روش، فواصل ساخته شده برخی از مقادیری را که جزء مشاهدات جامعهٔ آماری نیستند، نیز در برمی‌گیرد؛ از این رو بهتر است داده‌های اعشاری با تقریبی چون ٠/١٠ ، ٠/٥٠ یا ٠/٠١ تهیه شوند. در چنین مواردی شمارش اعداد با در نظرگرفتن حد پایین و حد بالای طبقه انجام می‌گیرد. مثال ٢ـ٣ نشان‌دهندهٔ چنین کیفیتی است.

مثال ٢ـ٣ این مشاهدات نشان‌دهندهٔ میزان مایع ریخته شده در ٧٠ بطری است (مشاهدات بر حسب میلی‌لیتر اندازه گیری شده‌اند):

٣٦/٩	٣٦/٧	٣٧/٠	٣٦/٤	٣٦/٨	٣٦/٩	٣٦/٩
٣٦/٨	٣٦/٥	٣٦/٩	٣٦/٥	٣٦/٧	٣٧/٠	٣٦/٨
٣٦/٠	٣٧/٠	٣٦/٨	٣٦/٧	٣٦/٩	٣٦/٩	٣٦/٨
٣٦/٥	٣٧/٨	٣٦/٤	٣٦/٩	٣٦/٩	٣٦/٦	٣٧/٠
٣٧/٠	٣٦/٢	٣٧/٠	٣٦/٩	٣٦/٨	٣٦/٧	٣٧/٠
٣٦/٩	٣٦/٧	٣٧/٠	٣٦/٩	٣٧/٠	٣٦/٥	٣٦/٨
٣٦/٨	٣٦/٨	٣٧/٠	٣٦/٩	٣٧/٧	٣٦/٩	٣٧/٠
٣٦/٤	٣٦/٦	٣٦/٦	٣٦/٦	٣٧/٠	٣٦/٩	٣٧/٢
٣٦/٩	٣٦/٩	٣٧/٠	٣٧/٠	٣٧/٠	٣٦/٦	٣٦/٩
٣٧/٠	٣٧/٠	٣٦/٩	٣٦/٠	٣٦/٩	٣٧/٠	٣٧/٠

مشاهدات در ده طبقه با فاصلهٔ طبقاتی ۰/۲۰ سازماندهی شده‌اند. چنانکه جدول ۲ـ۳ نشان می‌دهد، عرض طبقه مساوی ۰/۱۰ و طول طبقه مساوی ۰/۲۰ میلی‌لیتر است. این نوع طبقه‌بندی را، طبقه‌بندی گسسته با تقریب ۰/۱۰ گویند؛ زیرا فاصلهٔ حد بالای هر طبقه با حد پایین طبقهٔ مابعد خود به اندازهٔ ۰/۱۰ است.

مجموع فراوانیهای مطلق همیشه با تعداد مشاهدات (N) مساوی خواهد بود.

چنانچه مشاهدات جامعهٔ آماری مقادیر صحیح باشند، بهتر است جدول طبقه‌بندی آنها با «تقریب واحد» سازماندهی شود؛ یعنی اختلاف طول طبقه با عرض آن یک واحد باشد. به مثال ۳ـ۳ توجه کنید.

جدول ۲ـ۳ جدول طبقه‌بندی با تقریب یک دهم.

طبقه i	حدود طبقات (C - L)	خط و نشان	فراوانی مطلق (F_i)
۱	۳۶/۰۰_۳۶/۱۰	||	۲
۲	۳۶/۲۰_۳۶/۳۰	|	۱
۳	۳۶/۴۰_۳۶/۵۰	𝍸 ||	۷
۴	۳۶/۶۰_۳۶/۷۰	𝍸 𝍸 |	۱۱
۵	۳۶/۸۰_۳۶/۹۰	𝍸 𝍸 𝍸 𝍸 𝍸 ||	۲۷
۶	۳۷/۰۰_۳۷/۱۰	𝍸 𝍸 𝍸 ||||	۱۹
۷	۳۷/۲۰_۳۷/۳۰	|	۱
۸	۳۷/۴۰_۳۷/۵۰	—	۰
۹	۳۷/۶۰_۳۷/۷۰	|	۱
۱۰	۳۷/۸۰_۳۷/۹۰	|	۱

مثال ۳ـ۳ داده‌های مرتب شده ذیل مربوط به سود سالانهٔ ۵۰ شرکت بر حسب ۱۰ میلیون ریال است:

۱۰	۲۱	۳۲	۳۸	۵۰	۵۳	۶۴	۷۳	۸۵	۹۸
۱۰	۲۴	۳۳	۴۹	۵۱	۵۴	۶۴	۷۳	۸۷	۱۰۰
۱۴	۲۵	۳۴	۴۹	۵۱	۵۷	۶۵	۷۵	۸۹	۱۰۶
۱۵	۲۷	۳۵	۵۰	۵۱	۶۰	۶۸	۷۷	۹۲	۱۱۰
۱۷	۳۰	۳۷	۵۰	۵۱	۶۳	۷۰	۸۰	۹۵	۱۲۰

با این مثال مراحل چهارگانه طبقه‌بندی داده‌ها را مرور می‌کنیم.

مرحله اول. دامنهٔ تغییر را محاسبه می‌کنیم:

$$R = ۱۲۰ - ۱۰ = ۱۱۰$$

مرحله دوم. تعداد طبقات را با استفاده از رابطه ۴ـ۳ محاسبه می‌کنیم:

$$K = \sqrt{N} \Rightarrow K = \sqrt{۵۰} = ۷/۰۷۱ \Rightarrow K = ۸$$

مرحله سوم. فاصلهٔ طبقات را محاسبه می‌کنیم:

$$I = \frac{۱۱۰}{۸} = ۱۳/۷۵ \Rightarrow I = ۱۴$$

مرحله چهارم. براساس $K = ۸$ و $I = ۱۴$ جدول طبقه‌بندی ۳ـ۳ را تشکیل می‌دهیم و داده‌ها را سازماندهی می‌کنیم. در این مثال عرض طبقات یک واحد کمتر از طول طبقات است. از آنجا که طبقات نسبت به همدیگر ناسازگارند، برای شمارش، حدود طبقات بسته در نظر گرفته می‌شود.

جدول ۳ـ۳ جدول طبقه‌بندی مشاهدات با تقریب واحد

طبقه i	سود سالانه (C - L)	خط و نشان	فراوانی مطلق (F_i)
۱	۱۰ـ۲۳	~~||||~~ |	۶
۲	۲۴ـ۳۷	~~||||~~ ||||	۹
۳	۳۸ـ۵۱	~~||||~~ ~~||||~~	۱۰
۴	۵۲ـ۶۵	~~||||~~ |||	۸
۵	۶۶ـ۷۹	~~||||~~ |	۶
۶	۸۰ـ۹۳	~~||||~~	۵
۷	۹۴ـ۱۰۷	||||	۴
۸	۱۰۸ـ۱۲۱	||	۲
		$N = ۵۰$	$\sum F_i = ۵۰$

۳ـ۲ طبقه‌بندی مشاهدات ناپیوسته

روشهایی که در قسمت قبل بیان شد برای داده‌هایی مناسب است که از توزیعهای

پیوسته حاصل می‌شوند و طبقه‌بندی آنها در فاصله‌های پیوسته معنادار خواهد بود. بسیاری از مشاهدات ماهیتاً گسسته هستند و تعریف آنها در فاصلهٔ طبقات بی‌معناست؛ برای مثال می‌توان به تعداد غایبین یک شرکت، تعداد زدگیهای یک توپ پارچه و تعداد تصادفات روزانه یک شهر اشاره کرد. بدیهی است طبقه‌بندی این نوع مشاهدات با داده‌های پیوسته تفاوت خواهد داشت. از آنجا که این نوع مشاهدات به صورت عدد صحیح بیان می‌شوند، نمی‌توان آنها را در فاصله طبقات تعریف نمود؛ از این رو تعریف فاصله و تعداد طبقات براساس مفاهیم قسمت ۱ـ۳ بی‌معناست.

برای به دست آوردن توزیع فراوانی برای این دسته از مشاهدات، کافی است که جدول طبقه‌بندی به صورت دو ستون تهیه شود. ستون اول نشان‌دهنده X و ستون دوم تعداد تکرارهای X خواهد بود.

مثال ۴ـ۳ به منظور بررسی تعداد غایبین شرکت «الف» در دو نوبت صبح و عصر، دفاتر حضور و غیاب مورد بررسی قرار گرفت. حاصل بررسی توزیع غایبین نوبت صبح و عصر به این صورت است:

جدول ۴ـ۳ توزیع فراوانی غایبین نوبت صبح و عصر

تعداد روز (F_i)	تعداد غایبین عصر (y_i)	تعداد روز (F_i)	تعداد غایبین صبح (X_i)
۳	۰	۱۰	۰
۵	۱	۱۵	۱
۱۰	۲	۸	۲
۱۲	۳	۱۲	۳
۲۵	۴	۱۰	۴
۵	۵	۵	۵
$\sum F_i = ۶۰$		$\sum F_i = ۶۰$	

جدول ۴ـ۳ نشان می‌دهد که تعداد غایبین شرکت در ۶۰ روز شمارش شده‌اند. این جدول نشان می‌دهد که در نوبت صبح طی ده روز هیچ کس غایب نبوده است؛ در حالی که در نوبت عصر فقط طی سه روز کسی غایب نبوده و همه حضور داشته‌اند.

۳ـ۳ توزیع فراوانی نسبی[1]

چنانچه جدول طبقه‌بندی داده‌ها براساس نسبت داده‌های هر طبقه به کل مشاهدات جامعه تهیه شود، آن را توزیع فراوانی نسبی خوانند. فراوانی نسبی هر طبقه به صورت رابطهٔ ٦ـ۳ تعریف می‌شود:

$$\text{فراوانی نسبی طبقه } i = \frac{\text{فراوانی مطلق طبقه } i}{\text{تعداد کل مشاهدات}} \quad \text{یا} \quad f_i = \frac{F_i}{N} \qquad (۳\text{ـ}٦)$$

چگونگی محاسبه فراوانی نسبی را در مثال ۵ـ۳ می‌توان مشاهده کرد.

مثال ۵ـ۳ جدول طبقه‌بندی نسبت $\frac{P}{E}$ در مثال ۱ـ۳ را در نظر بگیرید. توزیع فراوانی آن به این صورت است:

C - L	F_i	$f_i = \frac{F_i}{N}$
۵ـ۹	۳	۰/۱۲
۹ـ۱۳	۵	۰/۲۰
۱۳ـ۱۷	۷	۰/۲۸
۱۷ـ۲۱	٦	۰/۲٤
۲۱ـ۲۵	۳	۰/۱۲
۲۵ـ۲۹	۱	۰/۰٤
	$\sum F_i = ۲۵$	$\sum f_i = ۱$

به کمک فراوانی نسبی می‌توان درصد تراکم داده‌ها را در هر طبقه مشخص نمود؛ برای مثال در جدول مذکور می‌توان دید که ۱۲ درصد سهام از نسبت $\frac{P}{E}$ ۵ تا ۹ برخوردار هستند. از فراوانی نسبی می‌توان برای یافتن محل تمرکز داده‌ها استفاده کرد.

ممکن است توزیع فراوانی با استفاده از نمایندهٔ (متوسط) طبقات[2] نشان داده شود. متوسط طبقات به این صورت تعریف می‌شود:

$$\text{متوسط طبقه } (X_i) = \frac{(\text{حد پایین طبقه } i) + (\text{حد بالای طبقه } i)}{۲} \qquad (۳\text{ـ}۷)$$

1. relative frequency distribution
2. middle of classes

در این حالت، به جای ستون حدود طبقات، از متوسط طبقات استفاده می‌شود.

مثال ۳ـ۶ جدول ذیل کامل شدهٔ جدول مثال ۳ـ۵ است (ستون متوسط طبقات به جدول اضافه شده است):

C-L	X_i (متوسط طبقه)	F_i	f_i
۵ـ۹	۷	۳	۰/۱۲
۹ـ۱۳	۱۱	۵	۰/۲۰
۱۳ـ۱۷	۱۵	۷	۰/۲۸
۱۷ـ۲۱	۱۹	٦	۰/۲٤
۲۱ـ۲۵	۲۳	۳	۰/۱۲
۲۵ـ۲۹	۲۷	۱	۰/۰٤
		N=۲۵	۱

در این جدول مقادیر $X_۱$، $X_۲$، ...، $X_٦$ عبارتند از:

$$X_۱ = \frac{۵ + ۹}{۲} = ۷ \text{ ،} X_۲ = \frac{۹ + ۱۳}{۲} = ۱۱ \text{ ، ... و } X_٦ = \frac{۲۵ + ۲۹}{۲} = ۲۷$$

۳ـ٤ توزیع فراوانی تجمعی[۱]

فراوانی تجمعی که با FC_i نشان داده می‌شود، نشان دهنده تجمع داده‌ها از حد پایین اولین طبقه تا حد بالای iامین طبقه است؛ بنابراین می‌توان فراوانی تجمعی را چنین تعریف کرد:

$$FC_i = \sum_{i=۱}^{i} F_i \qquad (۳ـ۸)$$

یعنی اینکه فراوانی تجمعی عبارت است از مجموع فراوانی مطلق طبقات تا طبقه iام.

مثال ۳ـ۷ ستون سوم این جدول نشان دهندهٔ فراوانی تجمعی و ستون چهارم نشان‌دهنده فراوانی تجمعی نسبی در هر طبقه است:

---

1. cumulative frequency distribution

C-L	F_i	فراوانی تجمعی (FC_i)	فراوانی تجمعی نسبی (fc_i)
۵_۹	۳	۳	۰/۱۲
۹_۱۳	۵	۸	۰/۳۲
۱۳_۱۷	۷	۱۵	۰/٦۰
۱۷_۲۱	٦	۲۱	۰/۸٤
۲۱_۲۵	۳	۲٤	۰/۹٦
۲۵_۲۹	۱	۲۵	۱

فراوانی نسبی تجمعی[1] از حاصل تقسیم فراوانی تجمعی بر تعداد مشاهدات به دست می‌آید؛ یعنی:

$$fc_i = \frac{FC_i}{N} \qquad (۳_۹)$$

فراوانی نسبی تجمعی درصد مشاهدات واقع شده بین حد پایین اولین طبقه و حد بالای iامین طبقه را نشان می‌دهد. بدیهی است فراوانی تجمعی و فراوانی نسبی تجمعی آخرین طبقه به ترتیب مساوی با N و یک خواهد بود.

تمرین

۱. خسارات ناشی از آتش سوزی (به میلیون ریال) در ۳۰ هفته به این شرح است:

۱۱/۰۰	۱۰/۵۰	۱۰/۲۰	۹/۰۰	۸/۲۰	۷/۱۰
۱۱/۵۰	۱۰/٦۰	۱۰/۳۰	۹/۰۰	۸/٤۰	۷/۲۰
۱۱/٦۰	۱۰/۷۰	۱۰/٤۰	۹/۷۰	۸/٦۰	۷/۸۰
۱۲/۰۰	۱۰/۸۰	۱۰/٤۰	۹/۸۰	۸/۷۰	۸/۰۰
۱۲/۱۰	۱۱/۰۰	۱۰/۵۰	۱۰/۰۰	۸/۹۰	۸/۱۰

جدول توزیع فراوانی نسبی وتجمعی را تهیه نمایید (با تقریب یک دهم طبقه‌بندی کنید).

1. relative cumulative frequency

۲. تعداد اتومبیلهای بیمه شده توسط یک شرکت بیمه در چهل روز بدین شرح است:

۸	۷	۸	۳	٦	٥	٤	۲
٦	٥	۸	۷	۹	۷	۱	٤
۷	٦	۱۰	٤	۳	٦	۳	۹
۹	۱۱	۸	۱۰	٦	۹	۲	٥
٤	٥	۸	٦	٥	۷	٦	۷

جدول توزیع فراوانی نسبی و تجمعی را تهیه نمایید. (داده‌ها از نوع ناپیوسته هستند).
۳. چرا مشاهدات از نوع ناپیوسته را با حدود طبقات دسته‌بندی نمی‌کنیم؟
٤. آیا روشی به عنوان بهترین روش طبقه‌بندی وجود دارد؟ چرا؟

۳ـ۵ نمایش هندسی مشاهدات

برای نمایش توزیعهای فراوانی، اغلب از نمودار استفاده می‌شود. نمودارها کمک می‌کنند که تصویر توزیع به سهولت دیده شود. روشهای نمایش داده‌ها علاوه بر استفاده‌های نظری برای استفاده‌های عملی و کاربردی مورد توجه هستند. به‌کارگیری این روشها در گزارش‌نویسی، یکی از راههای مفید جهت تفهیم موضوع گزارش به خواننده در کمترین زمان ممکن و با ساده‌ترین بیان است. روش نمایش داده‌ها با توجه به نوع مقیاس داده‌ها انتخاب می‌شود. چنانچه مقیاس داده‌ها از نوع فاصله‌ای و نسبی باشد، از نمودارهای کمّی[1] استفاده می‌شود. داده‌هایی که از مقیاس اسمی یا رتبه‌ای برخوردارند نیز به کمک نمودارهای وصفی[2] قابل نمایش هستند.

۳ـ۵ـ۱ نمودارهای کمّی

نمودارهای کمّی برای توزیعهای آماری پیوسته و گسسته قابل استفاده هستند. لازمه استفاده از این نمودارها مقداری بودن مشاهدات است. مهمترین نمودارهای کمّی به ترتیب شرح داده خواهد شد.

1. numerical charts
2. descriptive charts

۱ـ۱ـ۵ـ۳ نمودار بافت نگار[1]

پس از خلاصه کردن مجموعهٔ بزرگ داده‌ها به صورت توزیع فراوانی می‌توان آن را به کمک نمودار بافت نگار نمایش داد. بافت نگار نمایش مناسبی از الگوی توزیع است. برای رسم این نمودار از فراوانی مطلق یا فراوانی نسبی و حدود واقعی طبقات[2] استفاده می‌شود؛ زیرا حدودی که هنگام طبقه‌بندی برای هر طبقه مشخص می‌شود، ممکن است به علت گرد کردن با حدود واقعی آن تفاوت داشته باشد؛ برای مثال به طبقه‌بندی مشاهدات مثال ۲ـ۳ که مجدداً در جدول ۵ـ۳ نشان داده‌شده است، توجه کنید. در این جدول حد پایین طبقه اول ۳۶ میلی‌لیتر در نظرگرفته شده است؛ در حالی که ممکن است حد واقعی در این طبقه ۳۵/۹۵ یا ۳۶/۰۵ بوده باشد. از آنجاکه طبقه تا ۳۶/۰۵ میلی‌لیتر را در برمی‌گیرد؛ بنابراین ۳۵/۹۵ خارج از طبقه می‌ماند. بعلاوه ممکن است برخی از حجمها تا دو رقم اعشار تعریف شوند؛ مثلاً مقادیر ۳۶/۷۵ و ۳۷/۵۵ میلی‌لیتر در حدود طبقات قابل شمارش نیستند. برای برطرف کردن این مشکلات می‌توان حدود واقعی طبقات را تعریف کرد که از این پس این حدود را «حدود کرانه‌ها» می‌نامیم.

جدول ۵ـ۳ حجم مایع ۷۰ شیشهٔ بر حسب میلی‌لیتر

حدود کرانه‌ها	F_i	C-L
۳۵/۹۵ ـ ۳۶/۱۵	۲	۳۶/۰ ـ ۳۶/۱
۳۶/۱۵ ـ ۳۶/۳۵	۱	۳۶/۲ ـ ۳۶/۳
۳۶/۳۵ ـ ۳۶/۵۵	۷	۳۶/۴ ـ ۳۶/۵
۳۶/۵۵ ـ ۳۶/۷۵	۱۱	۳۶/۶ ـ ۳۶/۷
۳۶/۷۵ ـ ۳۶/۹۵	۲۷	۳۶/۸ ـ ۳۶/۹
۳۶/۹۵ ـ ۳۷/۱۵	۱۹	۳۷/۰ ـ ۳۷/۱
۳۷/۱۵ ـ ۳۷/۳۵	۱	۳۷/۲ ـ ۳۷/۳
۳۷/۳۵ ـ ۳۷/۵۵	۰	۳۷/۴ ـ ۳۷/۵
۳۷/۵۵ ـ ۳۷/۷۵	۱	۳۷/۶ ـ ۳۷/۷
۳۷/۷۵ ـ ۳۷/۹۵	۱	۳۷/۸ ـ ۳۷/۹

برای به دست آوردن حدود واقعی توزیعهای فراوانی به کمک حدود طبقات می‌توان قاعدهٔ سرانگشتی «۵ در مکان بعد از آخرین رقم» را به کار برد. به این ترتیب که

1. histogram chart
2. true classes interval

۵ واحد را از محل بعد از آخرین رقم حد پایین طبقه کم می‌کنیم و سپس آن را به مکان بعد از آخرین رقم حد بالای همان طبقه می‌افزاییم؛ بنابراین از روی حدود عنوان شده می‌توانیم به آسانی به حدود کرانه طبقه برسیم؛ برای مثال اگر حدود عنوان شده ۳۶/۰ـ۳۶/۱ باشد حدود واقعی ۳۵/۹۵ـ۳۶/۱۵ خواهد بود.

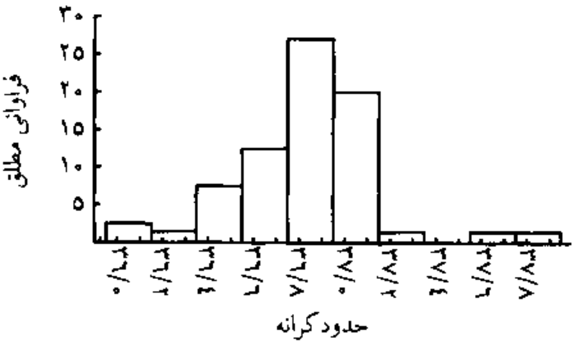
بررسی دقیقتر حدود کرانه‌ها نشان می‌دهد که طول و عرض طبقات با یکدیگر مساوی شده است؛ به عبارت دیگر برای به دست آوردن حدود واقعی طبقات به طبقه‌بندی‌ای مشابه طبقه‌بندیِ روش پیوسته رسیده‌ایم؛ بنابراین چنانچه روش طبقه‌بندی از نوع پیوسته باشد، حدود واقعی طبقات همان حدود طبقات خواهند بود.

در نمودار بافت نگار، محور افقی دستگاه مختصات با حدود واقعی طبقات و محور عمودی با فراوانی نسبی یا مطلق مدرج می‌گردد. روی کرانه‌های هر طبقه، مستطیلی عمودی رسم می‌کنیم که مساحت آن مساوی با فراوانی نسبی آن طبقه باشد؛ به عبارت ساده‌تر:

$$\text{ارتفاع مستطیل} = \frac{\text{فراوانی نسبی طبقه}}{\text{طول طبقه}}$$

بدین ترتیب مساحت هریک از مستطیلهای بافت‌نگار نشان‌دهندهٔ نسبت مشاهدات موجود در هر طبقه است که مستطیل بر آن قرار دارد. برای نمایش فراوانیهای نسبی، استفاده از مساحت مستطیلها به جای ارتفاع آنها فایدهٔ آشکاری دارد. ظاهراً در موقع مقایسه کردن دو قسمت یک بافت نگار، یا دو بافت نگار مختلف، چشم انسان به طور غریزی مساحتها را با هم مقایسه می‌کند. وقتی دو بافت نگار روی طبقات با طولهای متفاوت بنا شده باشند، این خاصیت که مساحت کل برابر یک است، آنها را قابل مقایسه با یکدیگر می‌سازد.

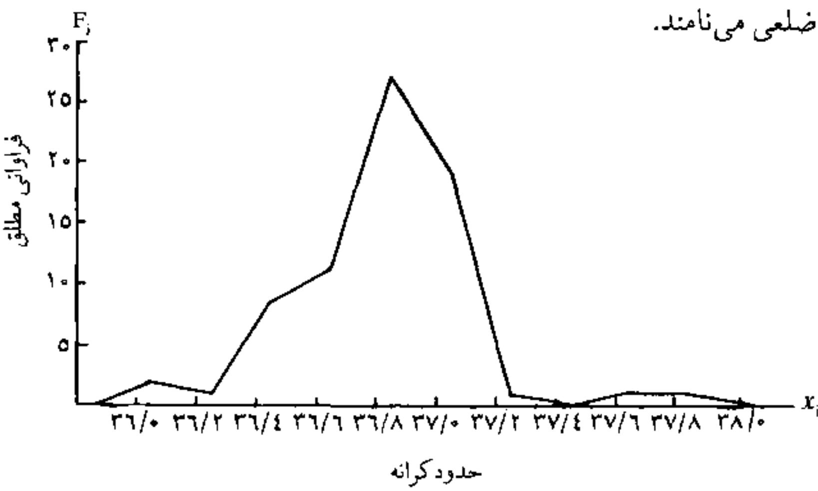
شکل ۱ـ۳ مربوط به بافت نگار داده‌های جدول ۵ـ۳ است. بافت نگار نشان می‌دهد که بیشترین تراکم در فاصلهٔ ۳۶/۷۵ تا ۳۶/۹۵ است. مقادیر روی مستطیلها نشان‌دهنده فراوانی مطلق هر طبقه است. بعلاوه علت اینکه در فاصلهٔ ۳۷/۳۵ـ۳۷/۵۵ مستطیلی دیده نمی‌شود این است که فراوانی مطلق در آن طبقه صفر است.



شکل ۳ـ۱ نمودار بافت نگار فراوانی مطلق حجم مایع بطریها در جدول ۳ـ۵

۳ـ۵ـ۱ـ۲ نمودار چند ضلعی

نمودار چند ضلعی نموداری است که نقطهٔ میانی هر طبقه روی محور افقی، و فراوانی نسبی یا مطلق هریک از نقاط میانی روی محور عمودی آن نشان داده می‌شود. متناظر با هر نمایندهٔ طبقه و فراوانی آن یک نقطه در صفحه مشخص می‌شود که طول آن نماینده طبقه و عرض آن برابر با فراوانی آن طبقه است؛ در نتیجه به تعداد طبقات جدول، در صفحه دستگاه نقطه پدید می‌آید. به نقاط مزبور دو نقطهٔ فرضی دیگر اضافه می‌کنیم؛ اولی مرکز طبقهٔ ماقبل اولین طبقه و دیگری نماینده طبقهٔ مابعد آخرین طبقه است. از اتصال متوالی این نقاط به یکدیگر نموداری حاصل می‌شود که آن را چند ضلعی می‌نامند.



شکل ۳ـ۲ نمودار چند ضلعی حجم مایع بطریها در جدول ۳ـ۵

از نمودارهای چند ضلعی، همانند بافت نگارها، برای رسم توزیعهای فراوانی استفاده می‌شود. تفاوت این دو نمودار آن است که برای نشان دادن رابطهٔ بین سطحهای متغیر و فراوانی آن متغیر، به جای رسم ستونهای متصل به هم، از خطوط مستقیم استفاده می‌شود. انتخاب یکی از این دو نمودار اختیاری است. در تحقیقاتی که شامل دو یا چند جامعهٔ آماری هستند و هدف مقایسهٔ آنهاست، ترجیح داده می‌شود که توزیعهای فراوانی جوامع روی یک نمودار رسم گردد. در چنین حالتی برای سهولت در خواندن و درک اطلاعات، از نمودارهای چند ضلعی فراوانی استفاده می‌شود.

۳ـ۵ـ۱ـ۳ نمودار فراوانی تجمعی[1]

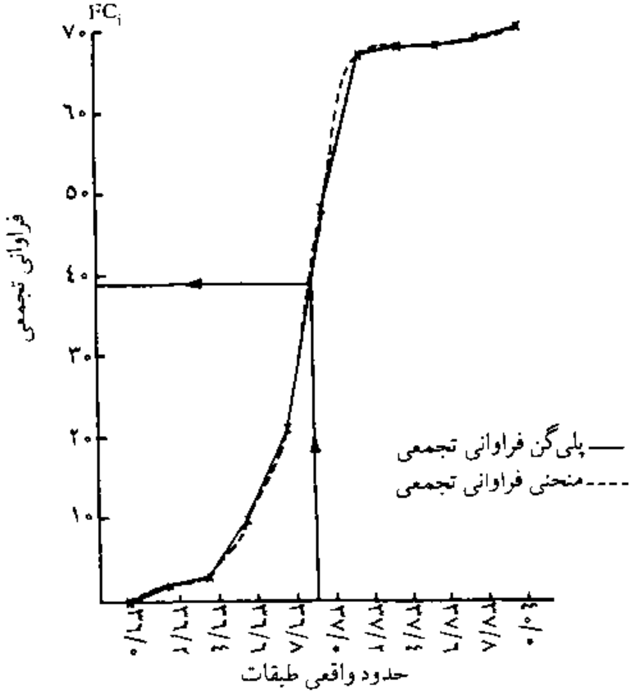
توزیع فراوانی تجمعی، توزیعی است که تعداد مشاهدات قبل از یک نقطهٔ معین را در مقیاس مشاهدات نشان می‌دهد. توزیع فراوانی تجمعی پایه لازم برای محاسبه برخی از پارامترهای آماری همانند صدکها، دهکها و چارکهاست. نمودار فراوانی تجمعی برای مقایسهٔ هندسی دو یا چند توزیع فراوانی به طور همزمان، به کار می‌رود.

نمودار فراوانی تجمعی براساس توزیع فراوانی تجمعی رسم می‌شود. این نمودار به دو روش قابل ترسیم است. در هر دو روش محور عمودی دستگاه مختصات عبارت است از فراوانیهای تجمعی طبقات، ولی محور افقی در هر روش تعریف خاصی دارد. در روش اول محور افقی براساس متوسط طبقات مدرج می‌شود. منحنی به دست آمده در این روش را پلی‌گن فراوانی تجمعی[2] گویند. در روش دوم محور افقی عبارت است از کرانه‌های بالای هر طبقه. منحنی این روش را منحنی فراوانی تجمعی[3] خوانند؛ بدین ترتیب نقاط متناظر فراوانی تجمعی و محور افقی در صفحه دستگاه مشخص شده، سپس با اتصال آنها به یکدیگر نمودار تجمعی حاصل می‌گردد (شکل ۳ـ۳).

بحثهای قابل توجهی پیرامون قابلیت نسبی هریک ازروشهای رسم نمودارفراوانی تجمعی وجود دارد. در پلی‌گن فراوانی تجمعی ارزشهای همهٔ مقادیرداخل هر طبقهٔ

---

1. cumulative frequency chart (ogive)
2. cumulative frequency polygon
3. cumulative frequency curve



شکل ۳.۳ نمودار فراوانی تجمعی حجم مایع بطریها در جدول ۵ـ۳

یکسان[1] فرض شده است؛ بنابراین نقطهٔ میانی هر طبقه با فراوانی تجمعی آن طبقه در نظر گرفته می‌شود، سپس آنها را به هم متصل می‌کنند؛ در حالی که در منحنی فراوانی تجمعی فرض بر عدم یکنواخت[2] بودن ارزش مقادیر هر طبقه است. برهمین اساس سعی می‌شود نمودار به گونه‌ای رسم گردد که این خاصیت بخوبی نشان داده شود. ناچار حد واقعی بالای هر طبقه با فراوانی تجمعی آن طبقه در نظر گرفته می‌شود. علاوه بر نکات فوق از نمودار تجمعی برای پاسخ به سؤالاتی همچون «چه تعداد از بطریها حاوی مایعی کمتر از ۳۶/۹ میلی‌لیتر هستند؟» نیز استفاده می‌شود. برای یافتن پاسخ کافی است که خطی عمودی از ۳۶/۹ به نمودار وصل گردد، سپس با خواندن فراوانی تجمعی آن معلوم می‌گردد که جواب ۳۹ است. در صورتی که محور عمودی نمودار با فراوانی تجمعی نسبی مدرج شده باشد عمل نقطه‌یابی مفیدتر خواهد بود.

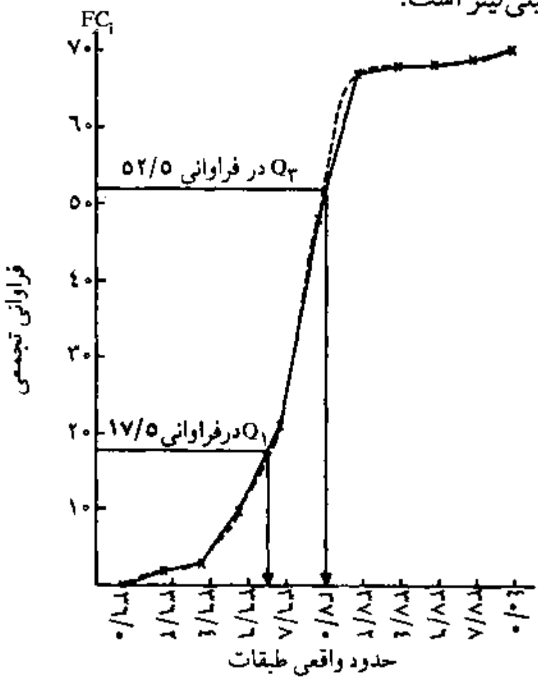
با نمودار فراوانی تجمعی براحتی می‌توان هریک از چندکها را محاسبه کرد. کافی است محل چندک را بر روی محور عمودی منحنی مشخص کرده، از آنجا خطی افقی

---

1. even
2. non-uniform

بر نمودار تجمعی وصل کنیم، سپس از آن نقطه خط را به صورت عمودی بر محور افقی رسم نماییم؛ بدین ترتیب مقدار دقیق چندک مشخص می‌گردد.

شکل ٤ـ۳ چارک اول و سوم را به کمک نمودار تجمعی مشخص می‌سازد. چارک اول جایی است که تجمع داده‌ها به ۲۵ درصد N و چارک سوم جایی است که تجمع داده‌ها به ۷۵ درصد N برسد. چنانکه مشخص است، چارک اول و سوم به ترتیب مساوی ۳٦/٦۸ و ۳۷ میلی‌لیتر است.



شکل ٤ـ۳ نحوه پیدا کردن چارک اول و سوم

این نمودار جهت مقایسه توزیعهای فراوانی تجمعی دو یا چند جامعه که از نظر تعداد با هم مساوی هستند، مفید خواهد بود؛ مثلاً برای مقایسه میزان رشد تورم در دو یا چند کشور می‌توان از این نمودار استفاده کرد. هرچه شباهت نمودارها بیشتر باشد، شباهت رشد تورم در آن کشورها بیشتر است.

٤ـ۱ـ۵ـ۳ تحلیل اکتشافی داده‌ها[1]

بافت نگار، نمایش هندسی بسیار مفیدی است. تصمیم‌گیرنده می‌تواند به کمک آن

1. exploratory data analysis

درک صحیحی از مشاهدات داشته باشد و نیز در نمایش توزیع، مرکزیت و پراکندگی داده‌ها به کار می‌رود، با این حال بافت نگار امکان شناسایی نقاط انفرادی داده‌ها را فراهم نمی‌کند؛ چرا که در این نمودار مشاهدات قرار گرفته در یک طبقه از هم قابل تمیز نیستند. نمودارهای جدیدی وجود دارد که نسبت به بافت نگار اطلاعات بیشتری در اختیار تصمیم‌گیرنده می‌گذارد. این نمودارها اغلب در مراحل اولیه تحلیل داده‌ها مفید هستند؛ به همین دلیل از آنها با عبارت روشهای تحلیل اکتشافی داده‌ها نام برده می‌شود. در این بخش دو نمونه از این روشها را معرفی می‌کنیم.

۱-۴-۱-۵-۳ *نمودار شاخه و برگ*[1]

فرض کنید مشاهدات با X_1، X_2، ...، X_N مشخص شده‌اند و هر مشاهده (X_i) دست کم دو رقمی است. برای تهیه نمودار شاخه و برگ، ارقام مشاهدات را به دو بخش تقسیم می‌کنیم: شاخه شامل یک یا چند رقم اولیه و برگ شامل ارقام باقیمانده؛ مثلاً اگر داده‌ها بین ۰-۱۰۰ باشد، در این صورت می‌توان مقداری مثل ۲۵ را به شاخهٔ ۲ و برگ ۵ تقسیم کرد. به طور کلی باید شاخه‌های نسبتاً کمی در مقایسه با تعداد مشاهدات انتخاب شود. معمولاً بهتر است این تعداد بین ۵ تا ۲۰ شاخه باشد. موقعی که مجموعهٔ شاخه‌ها انتخاب شد، این اعداد در ستون چپ صفحه نوشته می‌شوند و در کنار هر شاخه، تمام برگهای متناظر با مقادیر داده‌ها به ترتیب مشاهده ثبت می‌شود.

مثال ۸-۳ برای رسم نمودار شاخه و برگ، این داده‌ها را که مربوط به زمان انتظار مردان برای تلفن زدن است در نظر بگیرید (زمان بر حسب دقیقه است):

۱۳	۱۳	۰	۰	۱۵	۵	۰	۰
۵	۱۹	۱۰	۱۹	۰	۲۵	۱۸	۰
۱۴	۷	۱۷	۱۷	۱۷	۱۸	۹	۱۵
۸	۱۸	۲۰	۰	۲۱	۰	۰	۳۵
۲۵	۱۴	۰	۴	۰	۱۰	۰	۰
			۲۷	۲۳	۴	۰	۰

1. stem and leaf display

ابتدا مقادیر ۰ ، ۱ ، ۲ و ۳ را به عنوان شاخه انتخاب کرده، سپس داده‌های مربوط به هر شاخه را در جلو آن می‌آوریم. نمودار به این صورت حاصل می‌شود:

شاخه	برگ	فراوانی
۰	۰,۰,۵,۰,۰,۰,۰,۵,۹,۷,۰,۰,۰,۸,۰,۰,۰,۴,۰,۰,۰,۴	۲۲
۱	۵,۳,۳,۸,۹,۰,۹,۵,۸,۷,۷,۷,۴,۸,۰,۴	۱۶
۲	۵,۱,۰,۵,۳,۷	۶
۳	۵	۱

شکل ۳ـ۵ نمودار شاخه و برگ زمان انتظار مردان برای تلفن زدن

برای درک دقیقتر نمودار بهتر است آن را به صورت مرتب شده تهیه کنیم. نمودار مرتب شده شکل ۵ـ۳ در شکل ۶ـ۳ نشان داده شده است.

شاخه	برگ	فراوانی
۰	۰,۰,۰,۰,۰,۰,۰,۰,۰,۰,۰,۰,۰,۰,۰,۴,۴,۵,۵,۷,۸,۹	۲۲
۱	۰,۰,۳,۳,۴,۴,۵,۵,۷,۷,۷,۸,۸,۸,۹,۹	۱۶
۲	۰,۱,۳,۵,۵,۷	۶
۳	۵	۱

شکل ۳ـ۶ نمودار شاخه و برگ مرتب شده زمان انتظار مردان برای تلفن زدن

بررسی اجمالی نمودار شکل ۶ـ۳ مشخص می‌سازد که اکثر زمان انتظارها بین ۰ تا ۱۹ دقیقه بوده است. بعلاوه تمایل داده‌ها به سمت مقادیر کوچکتر است؛ یعنی اینکه پارامترهای مرکزی بخصوص مد، حول این مقادیر متمرکز شده‌اند. بنابراین نمودار شاخه و برگ همانند بافت نگار به ما امکان می‌دهد بعضی از خصوصیات مشاهدات را، که به شکل اولیه قابل شناسایی نیستند، بسرعت تعیین کنیم.

برخلاف بافت نگار در این نمودار اعداد اصلی از بین نمی‌روند. بعلاوه محاسبه چندکها به راحتی امکانپذیر است.

مثال ۹ـ۳ در این مثال به کمک شکل ۶ـ۳ نحوه محاسبه چارک اول، میانه و چارک سوم نشان داده می‌شود. از آنجا که داده‌ها در نمودار شاخه و برگ به صورت صعودی

مرتب شده‌اند، پس به راحتی می‌توان با پیدا کردن محل چارکها، هریک را تعیین نمود.

مکان چارک اول: $C_{Q_1} = \frac{N}{4} + \frac{1}{2} = \frac{45}{4} + \frac{1}{2} = 11/75$

از آنجا که عدد ۱۱/۷۵ در شاخه اول بین یازدهمین و دوازدهمین عدد قرار می‌گیرد، پس مقدار چارک اول صفر خواهد بود.

مکان چارک دوم (میانه): $C_{Md} = \frac{2N}{4} + \frac{1}{2} = \frac{2(45)}{4} + \frac{1}{2} = 23$

پس بیست‌وسومین عدد که درشاخه دوم قرارگرفته است، میانه مشاهدات خواهدبود؛ یعنی $Md = 10$

مکان چارک سوم: $C_{Q_3} = \frac{3N}{4} + \frac{1}{2} = \frac{3(45)}{4} + \frac{1}{2} = 34/25$

بنابراین $Q_3 = 18$

برخی مواقع ممکن است برگ یا شاخهٔ بیشتری مورد نیاز باشد، تأمین این نیاز، تعدیل شاخه‌های اصلی است؛ برای مثال تقسیم شاخهٔ ۱ به دو شاخهٔ جدید *۱ و **۱ که شاخه *۱ دارای برگهای ۰ , ۰ , ۳ , ۳ , ٤ , ٤ , ۵ , ۵ و شاخهٔ **۱ دارای برگهای ۷ , ۷ , ۷ , ۸, ۸ , ۸ , ۹ , ۹ است. بدین ترتیب تعداد شاخه‌های اولیه دو برابر می‌شود. تعیین تعداد شاخه‌ها به تمایل تصمیم‌گیرنده بستگی دارد که ممکن است بیشتر از دو برابر نیز بشود.

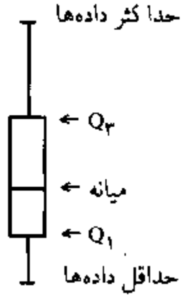
۲-۴-۱-۵-۳ نمودار جعبه‌ای[1]

نمودار جعبه‌ای یکی از مفیدترین نمودارهای اکتشافی برای مقایسهٔ دو یا چند جامعهٔ آماری است. نمودار جعبه‌ای نشان‌دهندهٔ چارکها و حداقل و حداکثر مشاهدات است؛ بدین ترتیب که جعبه شامل اختلاف چارک اول و سوم است. در این نمودار، ابتدای جعبه چارک اول و انتهای آن چارک سوم است (از پایین به بالا). خطی که جعبه را به دو نیم تقسیم می‌کند میانهٔ مشاهدات است. از هر طرفِ جعبه نیز به اندازهٔ نقاط مرزی مشاهدات (حداقل و حداکثر)، خطی ادامه می‌یابد که گاهی ریشه نامیده می‌شود.[2] در مورد جوامع آماری بزرگ این خطها ممکن است تا صدک دهم و صدک نودم یا صدک

1. box plot

۲. از این نمودار گاهی اوقات با عنوان نمودارهای جعبه و ریشه (box and whisker diagrams) نام می‌برند.

پنجم و نود و پنجم کشیده شود. مراحل تهیه، نمودار جعبه‌ای به این شرح است:
الف) حداکثر داده‌ها را پیدا کنید، ب) حداقل داده‌ها را پیدا کنید، ج) میانه را پیدا کنید، د) چارک اول را پیدا کنید، هـ) چارک سوم را پیدا کنید، و) نمودار را رسم کنید.

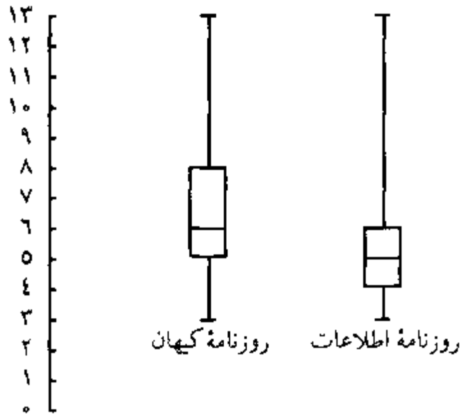


شکل ۳.۷ نمای سادهٔ نمودار جعبه‌ای

به منظور تشریح بیشتر کاربرد نمودار جعبه‌ای به ذکر یک مثال می‌پردازیم.
مثال ۱۰ـ۳ این جدول نشان‌دهندهٔ تعداد لغات انتخاب شده برای عنوانهای درشت دو روزنامهٔ کیهان و اطلاعات طی روزهای مختلف است. نمودار جعبه‌ای هر دو را تهیه کرده، مقایسه می‌کنیم.

تعداد لغات عنوانهای درشت	فراوانی لغات در روزنامه کیهان (روز)	فراوانی لغات در روزنامه اطلاعات (روز)
۳	۱۰	۱۶۶
۴	۴۶	۲۲۶
۵	۸۶	۲۲۵
۶	۹۴	۱۷۸
۷	۷۸	۷۵
۸	۵۰	۲۱
۹	۲۰	۲۰
۱۰	۱۲	۲
۱۱	۱۰	۱
۱۲	۱۲	۰
۱۳	۱۰	۳
	$\sum F_i = ۴۲۸$	$\sum F_i = ۹۱۷$

واضح است که چارک اول، دوم و سوم برای روزنامهٔ کیهان به ترتیب ۵ ، ٦ و ۸ و برای روزنامهٔ اطلاعات ٤ ، ۵ و ٦ است. مقادیر حداقل و حداکثر نیز برای هر دو روزنامه مساوی ۳ و ۱۳ است. نمودار جعبه‌ای این روزنامه‌ها در شکل ۸ـ۳ نشان داده شده است.



شکل ۸ـ۳ نمودار جعبه‌ای تعداد لغات در عنوانهای درشت روزنامهٔ کیهان و اطلاعات

مشخص است که این دو روزنامه فقط در میانه، ارزش نزدیک به هم دارند. لغات استفاده شده در روزنامهٔ کیهان بیشتر از روزنامهٔ اطلاعات است؛ به طوری که در روزنامهٔ اطلاعات بیش از ۲۵ درصد عنوانها از ۳ الی ٤ لغت تشکیل شده‌اند.

۲ـ۵ـ۳ نمودارهای وصفی

گفته شد که این دسته از نمودارها برای نمایش هندسی داده‌های کیفی به کار می‌روند. مشاهداتی از نوع خوب، بد، متوسط و گروه A ، B ، C را می‌توان از نوع مشاهدات کیفی دانست. طبیعی است برای این نوع مشاهدات، تعریف فاصله چنانچه در بافت نگار گفته شد، مفهوم ندارد. به عبارت دیگر، برای ساختن توزیع فراوانی به جای آنکه طبقه را به صورت فاصله در نظر بگیریم، ناچاریم هر مقدار را یک طبقه به شمار آوریم. مهمترین نمودارها برای نمایش داده‌های کیفی به شرح ذیل است.

۱ـ۲ـ۵ـ۳ نمودار ستونی[1]

این نمودار را در یک دستگاه مختصات که محور افقی آن نشان‌دهنده کیفیت مشاهدات

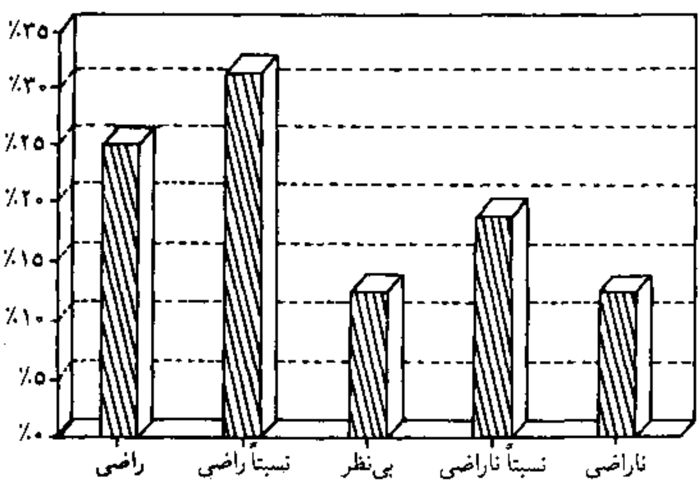
1. bar chart

و محور عمودی آن نشان‌دهنده فراوانی مطلق یا نسبی هر گروه است، رسم می‌کنند. مقادیر متمایز را به صورت نقاطی روی محور افقی مشخص کرده، سپس از نقاط حاصل، خطهایی ضخیم بر محور عمود می‌کنیم. ارتفاع هریک از خطوط براساس فراوانی مشخص می‌گردد. در نمودار میله‌ای، خطوط جایگزین مستطیلها می‌شوند تا بر این موضوع تأکید شود که فراوانیها واقعاً روی فاصله‌ها پخش نشده‌اند.

مثال ۳-۱۱ به منظور اندازه‌گیری رضایت شغلی در سازمانی تحقیقی صورت گرفته و برای این کار از پرسشنامه استفاده شده است. سؤالات پرسشنامه به صورت بسته و با طیف لیکرت (راضی، نسبتاً راضی، بی‌نظر، نسبتاً ناراضی، ناراضی) تهیه شده‌اند. خلاصهٔ اطلاعات جمع‌آوری شده در این جدول آمده است:

نوع پاسخ (X_i)	تعداد نفرات (F_i)	فراوانی نسبی (f_i)
راضی	۲۰۰	۰/۲۵
نسبتاً راضی	۲۵۰	۰/۳۱۲۵
بی‌نظر	۱۰۰	۰/۱۲۵
نسبتاً ناراضی	۱۵۰	۰/۱۸۷۵
ناراضی	۱۰۰	۰/۱۲۵
	N = ۸۰۰	۱

نمودار ستونی مثال ۳-۱۱ در شکل ۳-۹ نشان داده شده است.



شکل ۳-۹ نمودار ستونی رضایت شغلی سازمان

۳ـ۵ـ۲ـ۲ نمودار دایره‌ای[1]

نمودار دایره‌ای یکی دیگر از نمودارهایی است که برای نمایش داده‌های کیفی مورد استفاده فراوان قرار می‌گیرد. این نمودار ابزار مناسبی برای تجسّم مشاهدات است و معمولاً برحسب درصد تهیه می‌گردد. برای رسم نمودار دایره‌ای باید این مراحل را در نظر بگیرید:

الف) فراوانی مطلق را به فراوانی نسبی تبدیل کنید.

ب) با استفاده از رابطهٔ $S_i = ۳٦۰^{\circ} \times f_i$ مساحت هر قطاع از دایره را پیدا کنید.

ج) براساس S_i مساحت دایره را تقسیم کنید.

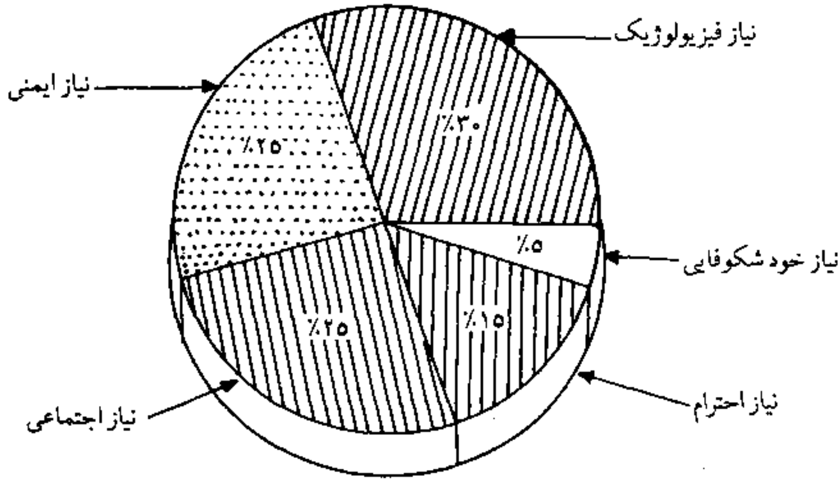
د) نوع مشاهدات و درصد آنها را نسبت به کل مشاهدات بر روی دایره بنویسید.

مثال ۳ـ۱۲ میزان بلوغ کارکنان یک شرکت براساس سطح نیاز آنها در سلسله مراتب مازلو[2] اندازه‌گیری شده است. این جدول نشان‌دهندهٔ سطح نیاز کارکنان شرکت است که برای تهیه نمودار دایره‌ای ستونهای ۳ و ٤ نیز به آن اضافه شده است.

سطح نیاز	تعداد کارکنان (F_i)	فراوانی نسبی (f_i)	$S_i = ۳٦۰^{\circ} \times f_i$
فیزیولوژیک	۳۰۰	٪۳۰	۱۰۸°
ایمنی	۲۵۰	٪۲۵	۹۰°
اجتماعی	۲۵۰	٪۲۵	۹۰°
احترام	۱۵۰	٪۱۵	٥٤°
خود شکوفایی	۵۰	٪۰۵	۱۸°
	$\sum F_i = ۱۰۰۰$	٪۱۰۰	$\sum S_i = ۳٦۰^{\circ}$

چنانکه مشخص است نیاز فیزیولوژیک بیشترین سطح را به خود اختصاص داده و نیاز خود شکوفایی در کمترین سطح قرار دارد؛ به عبارت ساده‌تر در شرکت مورد بحث ۳۰ درصد کارکنان در سطح نیاز فیزیولوژیک، ۳۵ درصد در سطح نیاز ایمنی، ۲۵ درصد در سطح نیاز اجتماعی و ... قرار دارند. این امر نشان می‌دهد که نظریه مازلو درباره سلسله مراتب نیازها در این شرکت صدق می‌کند.

1. pie chart
2. Abraham Maslow



شکل ۳.۱۰ نمودار دایره‌ای سطح نیازهای کارکنان شرکت

۳ـ۲ـ۵ـ۳ نمودار پاره‌تو[1]

نمودار پاره‌تو نوعی نمودار ستونی برای داده‌های وصفی است. در این نمودار، فراوانی هر موضوع روی محور عمودی و نوع آن روی محور افقی آورده می‌شود. نمودار پاره‌تو همیشه به ترتیب نزولی فراوانیها ترسیم می‌شود؛ یعنی پروقوع‌ترین موضوع در سمت چپ قرار گرفته و به همین ترتیب فراوانیهای کمتر در کنار آن جای می‌گیرند.

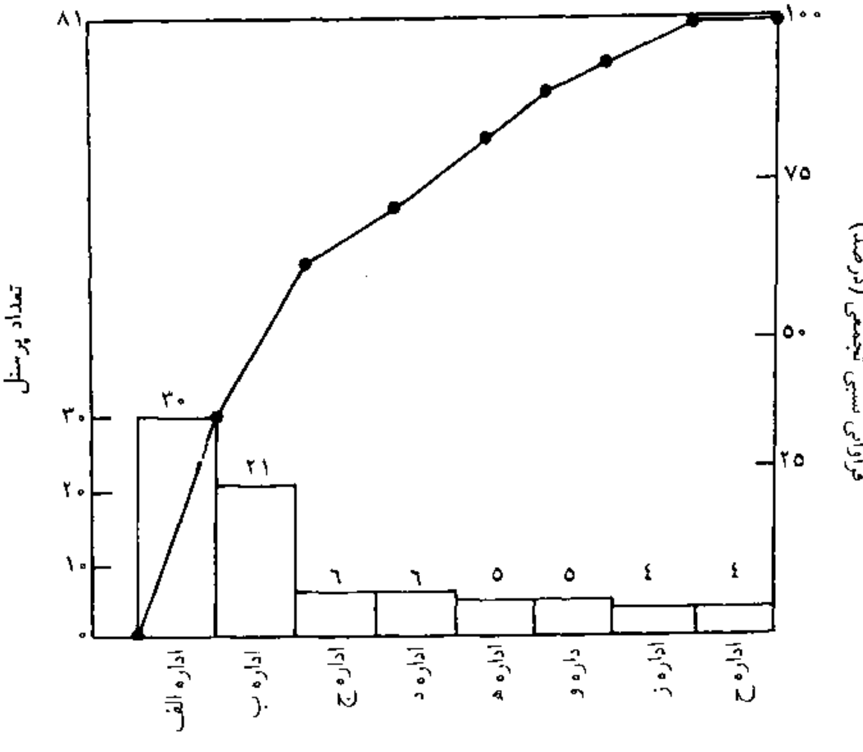
نام نمودار پاره‌تو از نام یک اقتصاددان ایتالیایی گرفته شده است. طبق نظریهٔ او در بعضی از جوامع بیشتر ثروت در دست عده کمی از مردم است. در داده‌های وصفی معمولاً اصل پاره‌تو به وقوع می‌پیوندد؛ به همین دلیل این نام برای نمودار انتخاب شده است.

این نمودار سه محور دارد که محور سوم آن نشان‌دهنده فراوانیهای نسبی تجمعی است و به صورت یک محور عمودی دیگر در انتهای محور افقی آورده می‌شود. منحنی روی نمودار پاره‌تو، درصدهای تجمعی K اُمین طبقه را به هم وصل می‌کند. برای رسم نمودار، ستونها به شکل مستطیل در نظر گرفته می‌شوند.

نمودار پاره‌تو در تحلیل موجودی انبار کالاها، نواقص سیستمها، توزیع درآمد و توزیع کارمندان سازمانها کاربرد فراوانی دارد. در شکل ۱۱ـ۳ نمودار پاره‌تو

1. Pareto chart

نشان‌دهندهٔ فراوانی کارمندان ادارات مختلف یک سازمان است. این سازمان دارای ۸۱ کارمند و هشت اداره است که چگونگی توزیع آنها در ادارات بر روی مستطیلها نشان داده شده است. اهمیت اداراتی که بیشترین سهم را در توزیع کارمندان دارند، بر روی نمودار مشهود است.



شکل ۳.۱۱ نمودار پاره‌تو برای توزیع کارمندان در سازمان

نمودار پاره‌تو یک قسمت مهم از برنامهٔ کنترل کیفیت محسوب می‌شود؛ چراکه مدیران و مهندسان به‌کمک آن می‌توانند توجه خود را به بحرانی‌ترین نواقص محصول یا فرایند معطوف کنند. زمانی که این نواقص بحرانی شناسایی شد، باید اقدامات اصلاحی برای کاهش یا حذف آنها صورت گیرد.

تمرین

۱. این داده‌ها زمان (برحسب ثانیه) دویدن زیگزاگ در مسافت ۲۵ متر را برای ۴۰ دانشجوی

رشتهٔ تربیت بدنی نشان می‌دهد:

۹/۱	۹/۶	۹/۶	۹/۰	۱۰/۵	۹/۷	۹/۶	۱۰/۳
۹/۵	۸/٤	۹/۹	۹/٥	۹/٥	۸/۷	۹/٤	۹/٦
۹/۳	۹/٥	۹/۸	۸/۷	۹/٦	۹/۸	۹/٦	۹/٥
۱۰/۲	۹/۸	۹/۲	۹/۰	۹/۳	۹/٥	۹/٤	۹/٤
۹/۲	۹/٦	۹/۰	۹/٥	۹/۸	۱۰/٤	۱۰/۰	۹/٥

الف) با استفاده از فاصلهٔ طبقات (سی صدم ثانیه)، جدول فراوانی داده‌ها را بسازید.

ب) نمودارهای مناسب را تهیه نمایید.

۲. یکی از صاحبنظران مدیریت با بررسی رفتار مدیران سطوح عالی، اهمیت نسبی انواع گوناگون ارتباطات را به دست آورده است. نتایج این بررسی در این جدول نشان داده شده است:

نوع ارتباط	ملاقاتهای برنامه‌ریزی شده	کارهای جاری	ملاقات بدون برنامه‌ریزی قبلی	پاسخ به تلفن	بازدید
اهمیت نسبی	٪۵۹	٪۲۲	٪۱۰	٪٦	٪۳

الف) نمودار پاره‌تو را تهیه نمایید.

ب) نمودار میله‌ای را تهیه نمایید.

۳. در مدیریت، انسانها را به لحاظ ارتباطات به چهار دسته تصویری، احساسی، صوتی و ارقامی تقسیم می‌کنند. کارکنان یک سازمان از این لحاظ مورد بررسی قرارگرفته‌اند که حاصل تحقیق در این جدول آمده است:

گروه ارتباطی	تصویری	احساسی	صوتی	ارقامی
تعداد کارکنان	۱۰۰	۱۵۰	۳۰۰	۵۰

الف) نمودار پاره‌تو را تهیه نمایید.

ب) نمودار دایره‌ای را تهیه نمایید.

٤. نمودارهای وصفی را با نمودارهای کمّی مقایسه کرده، تفاوت آنها را بنویسید.

۵. چه تفاوتی میان منحنی فراوانی تجمعی و پلی‌گن فراوانی تجمعی وجود دارد؟

۶ـ۳ خلاصه

طبقه‌بندی مشاهدات حجیم یکی از موضوعات بسیار مهم آمار است. طبقه‌بندی داده به نوع مشاهدات بستگی دارد. چنانچه داده‌ها ازنوع پیوسته باشند، بهتر است آنها را با تقریب ۰/۱۰ ، ۰/۰۵ یا ۰/۰۱ طبقه‌بندی کرد. در این نوع طبقه‌بندی ستون دیگری با عنوان حدود کرانه‌ها وجود دارد. حدود کرانه‌ها همان حدود واقعی طبقات است. حدود کرانه‌ها با «قاعده سرانگشتی ۵» ساخته می‌شود.

روش دیگری برای طبقه‌بندی مشاهدات پیوسته وجود دارد که در آن حدود کرانه‌ها و حدود طبقات مساوی هستند؛ به عبارت دیگر؛ در این نوع سازماندهی عرض طبقات با طول طبقات مساوی است؛ در حالی که در سایر روشها طول طبقات با فاصلهٔ طبقات مساوی است ولی عرض طبقات تقریبی از فاصلهٔ طبقات می‌باشد.

برای طبقه‌بندی داده‌های کمّی گسسته و کیفی نیز روشهای ویژه‌ای وجود دارد. در این نوع مشاهدات، طبقه براساس نوع مشاهده تعریف می‌شود؛ یعنی تعداد طبقات در جدول مساوی با تعداد انواع مشاهدات خواهد بود و نوع مشاهدات جایگزین فواصل (در داده‌های پیوسته) خواهد شد.

با پیدا کردن توزیع فراوانی مشاهدات از روشهای هندسی مناسب برای تجسم داده‌ها استفاده می‌شود. روشهای نمایش داده‌ها به مقیاس آنها بستگی دارد. نمودارهای بافت‌نگار، چند ضلعی، شاخه و برگ و جعبه‌ای برای به تصویر کشیدن داده‌هایی که از مقیاس فاصله‌ای و نسبی برخوردارند، مورد استفاده قرار می‌گیرند؛ در حالی که برای نمایش مشاهدات وصفی که از مقیاس اسمی و رتبه‌ای برخوردارند از نمودارهای ستونی، دایره‌ای و پاره‌تو استفاده می‌شود.

۷ـ۳ سؤالات و مسائل

سؤالات دو گزینه‌ای

۱. طبقه‌بندی مشاهدات جامعهٔ آماری در هر حالت بهتر است. ص ☐ غ ☐

۲. روشهای توصیف داده‌ها صرفاً به روشهای هندسی منحصر می‌شود. ص ☐ غ ☐

۳. نمودار بافت نگار یک نمودار کمّی است. ص ☐ غ ☐

۴. نمودار پاره‌تو یک نمودار وصفی است. ص ☐ غ ☐

۵. سازماندهی محاسبات موجب افزایش دقت محاسبات می‌شود. ص ☐ غ ☐

۶. رابطهٔ استورجس دقیقترین رابطهٔ ریاضی برای تعیین تعداد طبقات است.   ص ☐   غ ☐
۷. برای رسم پلی‌گن فراوانی تجمعی از متوسط طبقات استفاده می‌شود.   ص ☐   غ ☐
۸. نمودار شاخه و برگ یکی از نمودارهای تحلیل اکتشافی داده‌هاست.   ص ☐   غ ☐
۹. تعداد طبقات مناسب برای مشاهدات جامعه، حداقل ۵ و حداکثر ۲۰ است.   ص ☐   غ ☐
۱۰. فاصلهٔ طبقات و تعداد طبقات نسبت به هم رابطهٔ معکوس دارند.   ص ☐   غ ☐

سؤالات چهارگزینه‌ای

۱۱. کدام یک از این نمودارها برای تحلیل اکتشافی مشاهدات استفاده می‌شود؟
الف) بافت نگار   ج) پاره‌تو
ب) دایره‌ای   د) جعبه‌ای

۱۲. کدام یک از این نمودارها برای تحلیل مشاهدات کمّی استفاده می‌شود؟
الف) شاخه و برگ   ج) پاره‌تو
ب) دایره‌ای   د) ستونی

۱۳. کدام یک از این نمودارها برای نمایش مشاهداتی با مقیاس رتبه‌ای مناسب است؟
الف) دایره‌ای   ج) بافت نگار
ب) چند ضلعی   د) جعبه‌ای

۱۴. کدام یک از این نمودارها برای نمایش مشاهداتی با مقیاس نسبی مناسب است؟
الف) بافت نگار   ج) جعبه‌ای
ب) چند ضلعی   د) هر سه مورد

۱۵. در رسم نمودارهای بافت نگار، محور X را براساس کدام اندازه مدرج می‌کنند؟
الف) فراوانیهای نسبی   ج) متوسط طبقات
ب) کرانه‌های طبقات   د) فراوانیهای تجمعی

۱۶. در رسم نمودار تجمعی، محور X را براساس کدام اندازه مدرج می‌کنند؟
الف) متوسط طبقات   ج) حدّ پایین طبقات
ب) کرانه‌های طبقات   د) مورد الف و ب

۱۷. در کدام یک از این نمودارها ارزش مشاهدات هر طبقه یکسان تلقی می‌شود؟
الف) منحنی فراوانی تجمعی   ج) پلی‌گن فراوانی تجمعی
ب) بافت نگار   د) هر سه مورد

۱۸. اگر حداکثر و حداقل مشاهدات به ترتیب ۴۰۰ و ۲۰۰ و فاصلهٔ طبقات ۲۵ باشد، تعداد طبقات جدول طبقه‌بندی داده‌ها کدام است؟

الف) ۱۶ ج) ۵

ب) ۸ د) ۲۵

۱۹. اگر ۸۰ـ۸۹ و ۹۰ـ۹۹ دوطبقه ازیک جدول طبقه‌بندی باشند، فاصلهٔ طبقات کدام است؟

الف) ۹ ج) ۹/۵

ب) ۱۰ د) اطلاعات برای اظهار نظر کافی نیست.

۲۰. با توجه به سؤال ۱۹، اندازه طول طبقه کدام است؟

الف) مساوی با عرض طبقه ج) یک واحد بیشتر از عرض طبقه

ب) یک واحد کمتر از عرض طبقه د) مساوی با فاصلهٔ طبقات

مسائل

۲۱. قیمت هر سهم برای ۶۰ شرکت در بازار بورس و اوراق بهادار تهران در فروردین ماه به این شرح به دست آمده است و داده‌ها برحسب ۱۰ ریال است:

۱۰۰۰	۱۲۴۵	۱۹۰۰	۱۸۷۵	۲۰۱۲	۲۲۰۰	۱۹۷۵	۲۱۰۰	۱۲۰۰	۸۰۰
۱۲۰۰	۱۳۷۵	۱۶۵۰	۱۴۵۰	۲۰۱۰	۲۰۸۰	۸۶۷	۹۵۰	۱۱۰۰	۷۵۰
۲۱۰۰	۱۴۸۵	۱۳۷۰	۱۳۲۰	۱۹۷۵	۸۰۰	۹۵۰	۹۶۰	۱۰۹۰	۹۰۰
۲۰۵۰	۱۵۰۰	۱۵۸۰	۱۲۱۰	۱۶۰۰	۹۶۷	۹۵۵	۲۰۲۰	۱۱۱۰	۹۵۰
۱۸۰۰	۱۶۵۰	۱۷۸۰	۱۲۲۵	۱۶۵۰	۹۱۰	۹۸۵	۲۰۱۰	۹۵۰	۹۲۰
۱۹۲۰	۱۵۹۰	۱۸۵۰	۱۷۸۵	۱۴۰۰	۱۰۱۲	۲۱۰۰	۲۰۱۱	۸۷۰	۱۰۰۰

الف) داده‌ها را طبقه‌بندی کنید.

ب) نمودارهای بافت نگار، چند ضلعی، و فراوانی تجمعی را رسم کنید.

۲۲. زمان انتظار ۴۰ راننده که برای زدن بنزین به پمپ بنزین کارگر شمالی مراجعه کرده‌اند بر حسب دقیقه به این شرح به دست آمده است:

۰/۴۰	۲/۱۰	۰/۱۰	۳/۰۰	۰/۷۰	۰/۵۰	۰/۶۰	۱/۳۰
۱/۲۰	۲/۲۰	۰/۳۰	۲/۸۰	۰/۸۰	۱/۵۰	۱/۰۰	۱/۲۰
۰/۰۰	۲/۷۰	۰/۶۰	۲/۴۰	۱/۶۰	۱/۴۰	۳/۰۰	۲/۰۰
۳/۰۰	۲/۶۰	۰/۷۰	۲/۵۰	۱/۷۰	۰/۸۰	۰/۰۰	۲/۵۰
۲/۴۰	۲/۵۰	۰/۹۰	۲/۱۰	۰/۹۰	۰/۲۰	۱/۱۰	۰/۴۰

الف) داده‌ها را با استفاده از فاصلهٔ طبقاتی ۰/٥۰ دسته‌بندی کنید.
ب) نمودار بافت‌نگار را با استفاده از توزیع فراوانی نسبی تهیه کنید.
ج) نمودار تجمعی را با استفاده از فراوانی نسبی تهیه کنید.

۲۳. دانشکده علوم انسانی دانشگاه تربیت مدرس برای اندازه‌گیری عملکرد کارکنان دبیرخانه خود، از زمان‌سنجی استفاده کرده است. مدت زمان ثبت یک نامه در اندیکاتور اندازه‌گیری شده و نتایج آن برحسب دقیقه به این شرح است:

۱/۰۰	۰/۸۰	٥/۰۰	٤/۸۰	۳/۷۰	٤/۸۰	۳/۲۰	۲/۰۰	۳/۲۰	٤/٦۰
۱/٥۰	۲/۲۰	۱/۷۰	٤/۷۰	۳/۹۰	۰/۹۰	۰/۸۰	۲/۲۰	۳/۱۰	٤/۷۰
۲/۰۰	۲/۳۰	۱/۹۰	۳/۱۰	۲/۸۰	۰/۷۰	۰/۹۰	۲/۱۰	۳/۷۰	٤/۷۰
۲/۲۰	۳/٤۰	۳/٥۰	۱/۲۰	٤/٦۰	۱/۷۰	۱/۷۰	۲/۷۰	۳/۹۰	٤/۹۰
۳/۰۰	٤/۲۰	۳/۷۰	٥/۰۰	٤/۷۰	۱/۹۰	۱/۸۰	۲/۸۰	٤/۲۰	٤/۳۰

الف) داده‌ها را با تقریب ۰/۱ طبقه‌بندی کنید.
ب) آیا می‌توان پذیرفت که زمان ثبت نامه‌ها از یک توزیع متقارن برخوردار است؟
ج) نمودار فراوانی تجمعی را با استفاده از دو روش پلی‌گن فراوانی تجمعی و منحنی فراوانی تجمعی تهیه نمایید. نقاط قوت و ضعف دو روش را ذکر کنید.

۲٤. تولیدکننده‌ای عمر مفید یک فیلم عکاسی حساس را بررسی کرده است. داده‌های حاصل برحسب روز به این شرح است:

۱۲٥	۱۲۷	۱٤۰	۱۳٥	۱۲٦	۱۲۰	۱۲۱	۱٤۲	۱٥۱	۱٦۰
۱٤۰	۱۲٥	۱۲٤	۱۲۲	۱۲۷	۱۳۰	۱۳۱	۱٤۱	۱۳۷	۱۲۱
۱۲۱	۱۲۷	۱۲۸	۱۳٤	۱٤۰	۱۲۱	۱۲٦	۱۲٤	۱۲٥	۱۲۷
۱٤۱	۱٤۷	۱٥۰	۱۳۲	۱٤۳	۱۲۱	۱۲٤	۱۳۱	۱۳۱	۱۳۷

با استفاده از رسم یک بافت نگار خصوصیات این داده‌ها را ذکر کنید.

۲٥. تعداد کارکنان یک سازمان برحسب معاونت به این شرح است:

معاونت	اداری	مالی	طرح و برنامه	فرهنگی
تعداد کارکنان	۲٥۰	۲۰۰	۲۰۰	۳٥۰

الف) نمودارهای مناسب را برای نمایش این مشاهدات تهیه کنید.

ب) در صورت امکان پارامترهای مرکزی و پراکندگی را محاسبه کنید.

۲۶. سه نوع طرح طبقه‌بندی برای دسته‌بندی حقوق ساعتی (به ۱۰ ریال) شرکت ایران خودرو به این صورت پیشنهاد شده است:

طرح ۱	طرح ۲	طرح ۳
[۰-۵۰۰)	[۰-۵۵۰]	[۰-۵۰۰)
۵۰۰-۱۰۰۰	۵۵۱-۱۰۰۰	۶۰۰-۱۰۰۰
۱۰۰۰-۱۵۰۰	۱۰۰۱-۱۵۰۰	۱۱۰۰-۱۵۰۰
۱۵۰۰-۲۰۰۰	۱۵۰۱-۲۰۰۰	۱۶۰۰-۲۰۰۰
۲۰۰۰-۲۵۰۰]	≥ ۲۰۰۰	> ۲۰۰۰

این طرحها را مورد نقد و بررسی قرار داده، نظر خود را بنویسید.

۲۷. این مشاهدات نشان‌دهندهٔ میزان کیلومتر طی شده به ازای هر گالن بنزین برای یک ماشین خاص است:

۳۸	۱۹	۳۰	۴۳	۲۱
۳۵	۲۶	۲۶	۲۵	۲۴
۳۷	۲۱	۴۳	۲۳	۲۳
۳۷	۲۵	۲۵	۳۶	۲۴
۲۴	۳۹	۳۷	۳۸	۳۹
۲۴	۳۱	۲۵	۴۹	۴۵
	۲۳	۲۶	۲۴	۳۱
	۲۵	۱۹	۲۲	۳۷
	۲۴	۳۷	۴۳	۲۸

الف) نمودار شاخه و برگ را رسم کرده، آن را تحلیل نمایید.

ب) نمودار جعبه‌ای را رسم کرده، آن را تحلیل کنید.

ج) نمودار بافت نگار را تهیه و آن را با نمودار شاخه و برگ مقایسه کنید.

۲۸. محققی، پرسشنامه‌ای تهیه کرده و به کمک آن سبکهای رهبری را در یک وزارتخانه اندازه گیری کرده است. این جدول نشان‌دهنده تعداد مدیرانی است که از یک سبک خاص پیروی می‌کنند.

سبک	دستوری S_1	تشویقی S_2	مشارکتی S_3	تفویضی S_4
فراوانی (نفر)	۱۰۰۰	۱۵۰۰	۱۵۰۰	۱۰۰۰

نمودارهای مناسب را برای نمایش داده‌های این جدول به‌کار بگیرید.

۲۹. این مشاهدات نشان‌دهنده زمان انتظار زنان برای تلفن‌زدن است. مقادیر بر حسب دقیقه به دست آمده است:

۰	۰	۸	۱۳	۱۵	۰	۰	۰
۱۸	۱۶	۰	۲۰	۰	۲۷	٤	۱۲
۵	۱۰	۱۶	۰	۲۵	۰	۱۵	۰
۹	۸	۲۳	٤	۱۵	۲۲	۱۸	۹
۱۸	۲۱	۵	۱۵	۰	۶	۰	۱۱
۱۸	۱٤	۱٤	۰	۲۰	۰	۱۰	۲۲
۸	۱۹	۰	۰	۸	۱۲	۹	۱۶
۱۳	۰	۲۳	۱٤	۱٤			

الف) نمودار شاخه و برگ این داده‌ها را تهیه کنید و آن را تحلیل نمایید.

ب) نمودار بند الف را با نمودار شاخه و برگ زمان انتظار مردان در شکل ۶ـ۳ مقایسه کرده، تحلیل نمایید.

ج) نمودار جعبه‌ای زمان انتظار مردان و زنان را تهیه کرده، مقایسه کنید.

۳۰. تعداد تصادفات منجر به فوت در دو شهر تهران و اصفهان در ۱۰۰ روز به این شرح به دست آمده‌است:

تعداد تصادف	۰	۱	۲	۳	٤	۵	۶
فراوانی شهر تهران	۲۵	۱۵	۱۲	۱۳	۱۵	۱۵	۵
فراوانی شهر اصفهان	۲۰	۲۵	۱۵	۱٤	۱۶	۸	۲

نمودار جعبه‌ای دو شهر را با همدیگر مقایسه کرده، تحلیل نمایید.

۳۱. درصد ظرفیت اشغال‌شده انبار برحسب گروههای کالا در این جدول آمده‌است:

گروه کالا	A	B	C	D	E	F
درصد	۰/۳۵	۰/۲۵	۰/۲۰	۰/۱۰	۰/۰۸	۰/۰۲

الف) نمودار پاره‌تو را تهیه کنید.

ب) نمودار میله‌ای و دایره‌ای را تهیه کرده، با نمودار پاره‌تو مقایسه کنید.

پاسخنامهٔ سؤالات

۱) غ	۲) غ	۳) ص	٤) ص	٥) غ
٦) غ	۷) ص	۸) ص	۹) ص	۱۰) ص
۱۱) د	۱۲) الف	۱۳) الف	۱٤) د	۱٥) ب
۱٦) د	۱۷) ج	۱۸) ب	۱۹) ب	۲۰) د

فصل چهارم

توصیف مقداری مشاهدات طبقه‌بندی شده

چنین به نظر می‌رسد که اگر مشاهدات و عملکرد جامعهٔ آماری را برحسب مقادیر معنادار توصیف کنیم، مطلوب و مفید است. تفاوتی ندارد که این مشاهدات برای مقاصد پژوهشی و تحقیقاتی یا برای استفاده در موقعیت عملی‌تری مثل آزمایش برنامه‌های مدیریتیِ گردآوری شده، باشد. با وجود اینکه غالباً از جدولهای فراوانی و نمودارها برای نشان دادن توزیع داده‌ها استفاده می‌شود، هرگونه تحلیل یا به کاربردن روشهای آماری مستلزم محاسبهٔ بعضی از پارامترهای توصیفی خواهد بود.

در فصل سوم، تکیه بر توصیف جامعهٔ آماری به کمک روشهای هندسی بود. پس از سازماندهی داده‌ها، جدولی به دست می‌آمد که برای توصیف جامعه با استفاده از روشهای مقداری نیز کارآیی داشت. در این فصل توزیع فراوانی که همان جدول اعداد طبقه‌بندی شده است برای پاسخ به این سؤالات مورد استفاده قرار می‌گیرد:

۱. مرکز توزیع فراوانی کجاست؟
۲. پراکندگی توزیع فراوانی چه مقدار است؟
۳. مشاهدات جامعه به کدام سو تمایل دارند؟
٤. پراکندگی جامعه در مقایسه با توزیعهای مشابه چه وضعیتی دارد؟

سؤال اول به کمک محاسبهٔ پارامترهای مرکزی قابل پاسخگویی است. به سؤال دوم نیز با محاسبهٔ پارامترهای پراکندگی جواب داده می‌شود. مفاهیم بیان شده درباره پارامترهای مرکزی و پراکندگی در فصل دوم، همچنان به قوت خود باقی است، ولی آنچه در این فصل، جدید است روابط (فرمولهای) محاسباتی پارامترهای مرکزی و پراکندگی است. در فصل دوم روابط برای داده‌های طبقه‌بندی نشده، بیان شد، استفاده از داده‌های طبقه‌بندی شده مستلزم تعدادی روابط جدید برای پارامترهای مرکزی و

پراکندگی است. برای پاسخگویی به سؤال سوم از پارامترهایی با عنوان «پارامترهای تعیین انحراف از قرینگی» استفاده می‌شود و به سؤال چهارم با محاسبهٔ شاخص کشیدگی جواب داده می‌شود.

٤ـ۱ پارامترهای مرکزی در داده‌های طبقه‌بندی شده

از آنجاکه درباره کاربرد پارامترهای مرکزی در فصل دوم به تفصیل سخن رفت، در اینجا فقط به روش محاسبهٔ آنها در توزیع فراوانی اشاره می‌شود؛ بدین منظور ابتدا به طرح یک مسأله می‌پردازیم و سپس از این مسأله برای بیان واضحتر روابط استفاده می‌کنیم.

مثال ٤ـ۱ شرکت ایران دارو دارای ۱۰۰ کارمند است که درآمد ماهیانه آنها در این جدول به روش پیوسته طبقه‌بندی شده است:

C-L	F_i
۱۰ـ۲۰	۱۵
۲۰ـ۳۰	۳۰
۳۰ـ٤۰	۲۵
٤۰ـ۵۰	۲۰
۵۰ـ٦۰	۱۰
	N = ۱۰۰

٤ـ۱ـ۱ میانگین

محاسبهٔ میانگین برای داده‌های طبقه‌بندی شده به دو روش صورت می‌گیرد. در روش اول که به روش مستقیم موسوم است میانگین به این صورت تعریف می‌شود:

$$\mu_x = \frac{\sum F_i x_i}{N} \qquad (٤ـ۱)$$

در این فرمول، F_i فراوانی مطلق، x_i متوسط طبقات و N کل مشاهدات است.

در روش دوم که به روش غیر مستقیم (کدگذاری) مرسوم است. میانگین به این صورت تعریف شود:

$$\mu_x = A + \left(\frac{\sum F_i d_i}{N}\right) I \qquad (٤ـ۲)$$

در این فرمول، A عددی دلخواه است که به عنوان تقریبی از میانگین جامعه انتخاب می‌شود. معمولاً تلاش می‌شود که A را از ستون متوسط طبقات، بخصوص طبقات میانی، انتخاب کنند. این عمل موجب تسهیل در محاسبهٔ میانگین و اجتناب از اعمال ریاضی بیش از اندازه است. d_i کد هر طبقه تلقی می‌شود و عبارت است از:

$$d_i = \frac{x_i - A}{I} \qquad (٤\text{-}٣)$$

اثبات: اگر A میانگین اختیاری، I فاصلهٔ طبقات، $d_i = \frac{x_i - A}{I}$ و $x_i = A + I\,d_i$ باشد؛ پس:

$$\mu_x = \frac{\sum F_i x_i}{\sum F_i}$$

$$= \frac{F_1 x_1 + F_2 x_2 + \dots + F_K x_K}{\sum F_i}$$

$$= \frac{F_1 (A + Id_1) + F_2 (A + Id_2) + \dots + F_K (A + Id_K)}{\sum F_i}$$

$$= \frac{A(F_1 + F_2 + \dots + F_K) + I(F_1 d_1 + F_2 d_2 + \dots + F_K d_K)}{\sum F_i}$$

$$= \frac{A \sum F_i}{\sum F_i} + \left(\frac{\sum F_i d_i}{\sum F_i}\right) I$$

$$= A + \left(\frac{\sum F_i d_i}{\sum F_i}\right) I$$

$$= A + \left(\frac{\sum F_i d_i}{N}\right) I$$

چنانکه مشاهده می‌شود آخرین عبارت همان رابطهٔ ۲ـ٤ است. زمانی که مشاهدات حالت اعشار داشته یا به گونه‌ای تعریف شده باشند که محاسبهٔ میانگین با رابطهٔ ۱ـ٤ وقت‌گیر و مشکل‌آفرین باشد، از رابطهٔ ۲ـ٤ استفاده می‌شود.

مثال ۲ـ٤ با استفاده از مثال ۱ـ٤ میانگین را به دو روش مستقیم و غیرمستقیم

محاسبه می‌کنیم. سومین ستون نشان‌دهنده نماینده طبقات است. ستون چهارم جدول از حاصلضرب ستون دوم و سوم به دست آمده است. ستون پنجم نشان‌دهنده کد هر طبقه است که با استفاده از رابطهٔ ۴ـ۳ به دست آمده است. ششمین ستون نیز از حاصلضرب ستون دوم و پنجم به دست آمده است.

C-L	F_i	x_i	$F_i x_i$	d_i	$F_i d_i$
۱۰ـ۲۰	۱۵	۱۵	۲۲۵	۲-	۳۰-
۲۰ـ۳۰	۳۰	۲۵	۷۵۰	۱-	۳۰-
۳۰ـ۴۰	۲۵	۳۵	۸۷۵	۰	۰
۴۰ـ۵۰	۲۰	۴۵	۹۰۰	۱	۲۰
۵۰ـ۶۰	۱۰	۵۵	۵۵۰	۲	۲۰
	$N = ۱۰۰$		$\sum F_i x_i = ۳۳۰۰$	$\sum d_i = ۰$	$\sum F_i d_i = -۲۰$

چنانکه مشاهده می‌شود ۳۵ = A و ۱۰ = I است؛ بنابراین با استفاده از جدول خواهیم داشت:

میانگین به روش مستقیم: $\mu_x \simeq \frac{\sum F_i x_i}{N} = \frac{۳۳۰۰}{۱۰۰} = ۳۳$

میانگین به روش غیرمستقیم: $\mu_x \simeq A + (\frac{\sum F_i d_i}{N}) I = ۳۵ + (\frac{-۲۰}{۱۰۰}) \times ۱۰ = ۳۳$

چنانکه ملاحظه می‌شود میانگین در هر دو روش مساوی است. علت استفاده از علامت ≃ در رابطه‌ها دقیق نبودن پارامترهاست که به واسطهٔ طبقه‌بندی مشاهدات حاصل شده است.

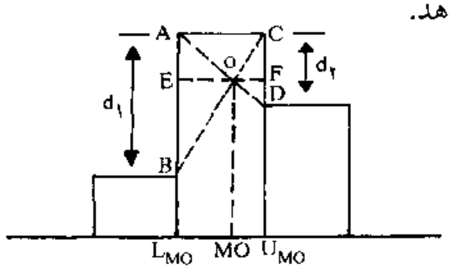
۴ـ۱ـ۲ مد

مد مقداری است که بیشترین تکرار را در میان مشاهدات جامعه داشته باشد. این تعریف برای مد داده‌های گسسته و طبقه‌بندی نشده بخوبی گویا و رساست، ولی در داده‌های طبقه‌بندی شده ناچاریم مد را به کمک رابطهٔ ریاضی ۴ـ۴ محاسبه کنیم؛ چراکه بیشترین تکرار فقط نشان‌دهنده تقریبی طبقهٔ مددار است.

$$Mo \simeq L_{Mo} + (\frac{d_1}{d_1 + d_2}) I \qquad (۴ـ۴)$$

در این فرمول، L_{Mo} حد پایین واقعی طبقهٔ مددار، d_1 تفاضل فراوانی مطلق طبقهٔ ماقبل از فراوانی مطلق طبقهٔ مددار $(d_1 = F_i - F_{i-1})$، d_2 تفاضل فراوانی مطلق طبقهٔ مابعد از فراوانی طبقهٔ مددار $(d_2 = F_i - F_{i+1})$، i طبقهٔ مددار و I فاصلهٔ طبقاتی است.

توجیه هندسی فرمول ٤ـ٤ به شرح ذیل است. تصور کنید که شکل ۱ـ٤ سه مستطیل از بافت نگار توزیع فراوانی درآمد ماهیانهٔ کارمندان شرکت ایران داروست. مستطیل وسط همان طبقهٔ مددار است. مد برابر است با مقداری که خط OMo بر روی محور افقی نشان می‌دهد.



شکل ۱ـ٤ بخشی از بافت نگار درآمد کارمندان در مثال ۱ـ٤

اگر L_{Mo} حد پایین واقعی و U_{Mo} حد بالای واقعی طبقهٔ مددار و d_1 و d_2 به ترتیب اختلاف فراوانی طبقهٔ مددار از طبقهٔ ماقبل و مابعد خود باشد و I فاصلهٔ طبقه را نشان دهد، می‌خواهیم ثابت کنیم که:

$$Mo = L_{Mo} + \left(\frac{d_1}{d_1 + d_2}\right) I$$

دو مثلث AOB و COD متشابه هستند، پس داریم:

$$\frac{AB}{OE} = \frac{CD}{OF} \Rightarrow \frac{d_1}{Mo - L_{Mo}} = \frac{d_2}{U_{Mo} - Mo}$$

می‌دانیم که $U_{Mo} = L_{Mo} + I$، پس اگر این مقدار را به جای U_{Mo} جایگزین کرده، طرفین ـ وسطین کنیم، خواهیم داشت:

$$d_2 (Mo - L_{Mo}) = d_1 (L_{Mo} + I - Mo)$$

$$d_2 Mo - d_2 L_{Mo} = d_1 L_{Mo} + d_1 I - d_1 Mo$$

$$d_2 Mo + d_1 Mo = d_1 L_{Mo} + d_2 L_{Mo} + d_1 I$$

$$Mo (d_1 + d_2) = L_{Mo} (d_1 + d_2) + d_1 I$$

$$Mo = L_{Mo} + \left(\frac{d_1}{d_1 + d_2}\right) I$$

چنانکه ملاحظه می‌شود آخرین عبارت همان رابطهٔ ٤ـ٤ است.

مثال ٤ـ٣ این جدول را که همان جدول مثال ١ـ٤ است مجدداً در نظر بگیرید، می‌خواهیم مد را محاسبه کنیم:

C.L	F_i	
۱۰_۲۰	۱۵	
۲۰_۳۰	۳۰	طبقهٔ مددار
۳۰_٤۰	۲۵	
٤۰_۵۰	۲۰	
۵۰_٦۰	۱۰	
	N = ۱۰۰	

چون طبقهٔ دوم دارای بیشترین تکرار است، پس طبقهٔ مددار، طبقهٔ دوم خواهد بود؛ یعنی i = ۲، پس L_{Mo} = ۲۰، d_1 = ۳۰ − ۱۵ = ۱۵، d_2 = ۳۰ − ۲۵ = ۵ و I = ۱۰؛ بنابراین خواهیم داشت:

$$Mo \simeq ۲۰ + (\frac{۱۵}{۱۵ + ۵}) \times ۱۰ = ۲۷/۵$$

معنای حاصل عبارت این است که بیشترین کارمندان شرکت از حقوق ماهیانه ۲۷۵ هزار ریال برخوردارند. توجه داشته باشید که این مقدار ممکن است با مد واقعی جامعه متفاوت باشد؛ از این رو با علامت تقریب (≃) نشان داده شده است.

٤ـ١ـ٣ چندکها

چندکها مقادیری از مشاهدات هستند که دامنهٔ تغییرات را به فاصله‌های چندکی مورد نیاز تقسیم می‌کنند؛ به طوری که فراوانیها در هریک از این فواصل، درصد معینی از فراوانی کل را تشکیل می‌دهند؛ بنابراین اگر دامنهٔ تغییرات را به چهار قسمت مساوی تقسیم کنیم به چارکها، اگر به ده قسمت مساوی تقسیم کنیم به دهکها[1] و اگر به صد قسمت مساوی تقسیم نماییم به صدکها[2] خواهیم رسید.

چندکها دارای کاربرد فراوانی در علوم بخصوص در اقتصاد کلان، کنترل کیفیت آماری و مدیریت هستند. چندکها در داده‌های طبقه‌بندی نشده و اندک به راحتی قابل

1. decile 2. percentile

مقایسه هستند، ولی در داده‌های طبقه‌بندی شده باید به کمک رابطهٔ ریاضی محاسبه شوند.

برای محاسبهٔ چندکها مراحلی وجود دارد که برای هرگروه به تفکیک در پی خواهد آمد.

۱ـ۳ـ۱ـ٤ چارکها

مراحل محاسبهٔ چارکها به این شرح است:

۱. فراوانی تجمعی (FC_i) را تهیه نمایید.

۲. محل چارک a ام را با استفاده از رابطهٔ ٤ـ٥ پیدا کنید.

$$C_{Qa} = \frac{aN}{٤} \qquad (٤\text{ـ}٥)$$

در این رابطه، a شماره چارک و C_{Qa} محل چارک است.

۳. به کمک مرحله ۱ و ۲ طبقهٔ چارک‌دار را پیدا کنید؛ سپس به کمک رابطهٔ ٤ـ٦ مقدار چارک a ام را محاسبه نمایید.

$$Q_a = L_{Qa} + \left(\frac{\frac{aN}{٤} - FC_{i-1}}{F_i}\right) I \qquad (٤\text{ـ}٦)$$

در این رابطه، L_{Qa} حد پایین واقعی طبقهٔ چارک‌دار و FC_{i-1} فراوانی تجمعی طبقهٔ ماقبل طبقهٔ چارک‌دار و F_i فراوانی مطلق طبقهٔ چارک‌دار است.

مثال ٤ـ٤ با استفاده از اطلاعات این جدول چارک اول تا سوم را محاسبه می‌کنیم.

C-L	F_i	FC_i	
۱۰ـ۲۰	۱۵	۱۵	
۲۰ـ۳۰	۳۰	٤٥	طبقهٔ Q_1
۳۰ـ٤۰	۲٥	۷۰	طبقهٔ Q_2
٤۰ـ٥۰	۲۰	۹۰	طبقهٔ Q_3
٥۰ـ٦۰	۱۰	۱۰۰	
N = ۱۰۰			

چارک اول: $C_{Q1} = \frac{۱ \times ۱۰۰}{٤} = ۲٥$

عدد ۲۵ در طبقهٔ دوم قرار دارد؛ یعنی:

$$Q_1 \simeq ۲۰ + (\frac{۲۵ - ۱۵}{۳۰}) \times ۱۰ = ۲۳/۳۳$$

چارک دوم (میانه):
$$C_{Q_۲} = \frac{۲ \times ۱۰۰}{٤} = ۵۰$$

$$Q_۲ \simeq ۳۰ + (\frac{۵۰ - ٤۵}{۲۵}) \times ۱۰ = ۳۲$$

چارک سوم:
$$C_{Q_۳} = \frac{۳ \times ۱۰۰}{٤} = ۷۵$$

$$Q_۳ \simeq ٤۰ + (\frac{۷۵ - ۷۰}{۲۰}) \times ۱۰ = ٤۲/۵$$

رابطهٔ ٦ـ٤ یک رابطهٔ تقریبی است که براساس درون‌یابی خطی[1] به دست آمده است و تقریباً با چارک aام مساوی است. فرایند درون‌یابی خطی برای پیدا کردن روابط کلیه چندکها قابل استفاده است؛ از این رو قدری درباره آن توضیح می‌دهیم:

فرض کنید می‌خواهیم چارک اول را که مکان آن مساوی ۲۵ درصد N است، محاسبه نماییم. در مثال ما که ۱۰۰ = N است، محل چارک اول مساوی ۲۵=۱۰۰×۰/۲۵ خواهد بود؛ بنابراین با شمردن ۲۵ مشاهده از کوچکترین مشاهده به بالا، چارک اول را می‌توانیم تعیین نماییم. با توجه به ستون فراوانی تجمعی معلوم می‌گردد که چارک اول در طبقهٔ دوم بین ۲۰ـ۳۰ واقع می‌شود؛ بنابراین در بین ۳۰ مشاهده موجود در این طبقه سعی می‌کنیم مکان مشاهدهٔ بیست و پنجم (۱۰ = ۱۵ − ۲۵) را تعیین کنیم. با این فرض که ۳۰ مشاهده موجود در این طبقه به طور یکنواخت پخش شده‌اند، مکان مشاهدهٔ دهم در محلی است که مساحت آن در نقطهٔ ابتدایی طبقه به اندازهٔ $\frac{۱۰}{۳۰}$ عرض طبقه است. عرض طبقه برابر با ۱۰ = ۲۰ − ۳۰ و فاصلهٔ چارک اول از نقطهٔ ابتدایی طبقه، برابر با ۳/۳۳ = ۱۰ × $(\frac{۱۰}{۳۰})$ است؛ بنابراین خواهیم داشت:

$$Q_1 = ۲۰ + (\frac{۱۰}{۳۰}) \times ۱۰ = ۲۳/۳۳$$

---

1. linear interpolation

۲ـ۳ـ۱ـ٤ دهکها

مراحل محاسبهٔ دهکها که با D_a نمایش داده می‌شود به این شرح است:

۱. فراوانی تجمعی را تهیه نمایید.

۲. محل دهک aام را با استفاده از رابطهٔ ٤ـ۷ تهیه کنید.

$$C_{Da} = \frac{aN}{۱۰} \qquad (٤\text{ـ}۷)$$

۳. با استفاده از فراوانی تجمعی، محل دهک، و رابطه ٤ـ۸ مقدار دهک را محاسبه نمایید.

$$D_a = L_{Da} + \left(\frac{\frac{aN}{۱۰} - FC_{i-1}}{F_i}\right) I \qquad (٤\text{ـ}۸)$$

در این رابطه، L_{Da} حد پایین واقعی طبقهٔ دهک‌دار و F_i فراوانی مطلق طبقهٔ دهک‌دار و FC_{i-1} فراوانی تجمعی طبقهٔ ماقبل طبقهٔ دهک‌دار و I فاصلهٔ طبقات است.

مثال ٤ـ۵ با استفاده از این جدول دهک پنجم (میانه) و هفتم را برای مثال ٤ـ۱ محاسبه می‌کنیم.

C.L	F_i	FC_i	
۱۰ـ۲۰	۱۵	۱۵	
۲۰ـ۳۰	۳۰	٤۵	
۳۰ـ٤۰	۲۵	۷۰	محل دهک پنجم
٤۰ـ۵۰	۲۰	۹۰	محل دهک هفتم
۵۰ـ٦۰	۱۰	۱۰۰	
N = ۱۰۰			

از آنجا که محل دهک پنجم مساوی ۵۰ است، پس مقدار دهک به این صورت محاسبه می‌شود:

$$D_۵ = ۳۰ + \left(\frac{۵۰ - ٤۵}{۲۵}\right) \times ۱۰ = ۳۲$$

باید توجه داشت که اگر محل چندک دقیقاً در ستون فراوانی تجمعی دیده شود، آن طبقه

یا طبقهٔ پایین‌تر از آن را می‌توان به عنوان طبقهٔ چندک‌دار در نظر گرفت. در مثال ۵ـ٤ با وجود مشاهده ۷۰ در ستون فراوانی تجمعی، می‌توان طبقهٔ پایین‌تر را به عنوان طبقه‌ای که دهک هفتم در آن قرار گرفته است، در نظر گرفت. بدین ترتیب مقدار دهک هفتم به این صورت محاسبه می‌شود:

$$C_{D_۷} = \frac{۷ \times ۱۰۰}{۱۰} = ۷۰$$

$$D_۷ = ٤۰ + (\frac{۷۰ - ۷۰}{۲۰}) \times ۱۰ = ٤۰$$

۳-۳-۱-٤ صدکها

صدکها را با P_a نشان می‌دهیم. مراحل محاسبهٔ صدک a ام به این صورت است:

۱. ستون فراوانی تجمعی را تهیه نمایید.

۲. محل صدک a ام را با استفاده از رابطهٔ ۹ـ٤ پیدا کنید.

$$C_{P_a} = \frac{aN}{۱۰۰} \qquad (٤\text{-}۹)$$

۳. مقدار صدک را با استفاده از رابطهٔ ۱۰ـ٤ پیدا کنید.

$$P_a = L_{P_a} + (\frac{\frac{aN}{۱۰۰} - FC_{i-۱}}{F_i}) I \qquad (٤\text{-}۱۰)$$

در این رابطه، L_{P_a} حد پایین واقعی طبقهٔ صدک‌دار و $FC_{i-۱}$ فراوانی تجمعی طبقهٔ ماقبل طبقهٔ صدک‌دار و F_i فراوانی مطلق طبقهٔ صدک‌دار و I فاصلهٔ طبقات است.

مثال ٦ـ٤ صدک دهم و نود و پنجم جدول مربوط به مثال ۱ـ٤، را محاسبه می‌کنیم:

C-L	F_i	FC_i	
۱۰_۲۰	۱۵	۱۵	طبقهٔ صدک دهم
۲۰_۳۰	۳۰	٤۵	
۳۰_٤۰	۲۵	۷۰	
٤۰_۵۰	۲۰	۹۰	
۵۰_٦۰	۱۰	۱۰۰	طبقهٔ صدک نود و پنجم
N = ۱۰۰			

صدک دهم: $C_{P_{10}} = \frac{10 \times 100}{100} = 10$

$$P_{10} = 10 + (\frac{10 - 0}{15}) \times 10 = 16/67$$

صدک نود و پنجم: $C_{P_{95}} = \frac{95 \times 100}{100} = 95$

$$P_{95} = 50 + (\frac{95 - 90}{10}) \times 10 = 55$$

قابل ذکر است که تمام محاسبات مربوط به پارامترهای مرکزی با فرض پیوسته بودن حدود طبقات بیان گردیده است و مثال ۱ـ۴ نیز دارای این ویژگی است. چنانچه حدود طبقات با تقریب فراهم شده باشد، ناچار برای محاسبهٔ چندکها و مد باید حدود واقعی طبقات (کرانه‌ها) تهیه گردد. ملاک در رابطه‌های ۴ـ۴ ، ۶ـ۴ ، ۸ـ۴ و ۱۰ـ۴ کرانه پایین طبقه مورد نظر خواهد بود.

توجه: هنگام محاسبهٔ پارامترهای مرکزی؛ مثل مد و چندکها (چارکها، دهکها و صدکها) باید طول و عرض طبقه مساوی باشد؛ به عبارت دیگر چنانچه روش طبقه‌بندی با تقریب صورت گرفته باشد باید از حد پایین کرانه‌ها برای محاسبهٔ این دسته از پارامترها استفاده شود.

در اینجا مثالی می‌آوریم که در آن برای خلاصه کردن داده‌ها از تقریب یک واحد استفاده شده است.

مثال ۷ـ۴ این جدول نشان‌دهنده طبقه‌بندی سود سالانه ۵۰ شرکت برحسب ۱۰ میلیون ریال است. عرض طبقات ۱۳ و طول طبقات ۱۴ است. می‌خواهیم میانگین به روش مستقیم، مد، میانه، دهک اول و صدک نودم را محاسبه کنیم. باید از حدود کرانه‌ها استفاده شود (جدول تمامی ستونهای لازم را داراست).

C-L	حدود گرانه	F_i	FC_i	x_i	$F_i x_i$
۱۰_۲۳	۹/۵_۲۳/۵	۶	۶	۱۶/۵	۹۹
۲۴_۳۷	۲۳/۵_۳۷/۵	۹	۱۵	۳۰/۵	۲۷۴/۵
۳۸_۵۱	۳۷/۵_۵۱/۵	۱۰	۲۵	۴۴/۵	۴۴۵
۵۲_۶۵	۵۱/۵_۶۵/۵	۸	۳۳	۵۸/۵	۴۶۸
۶۶_۷۹	۶۵/۵_۷۹/۵	۶	۳۹	۷۲/۵	۴۳۵
۸۰_۹۳	۷۹/۵_۹۳/۵	۵	۴۴	۸۶/۵	۴۳۲/۵
۹۴_۱۰۷	۹۳/۵_۱۰۷/۵	۴	۴۸	۱۰۰/۵	۴۰۲
۱۰۸_۱۲۱	۱۰۷/۵_۱۲۱/۵	۲	۵۰	۱۱۴/۵	۲۲۹
		N = ۵۰			۲۷۸۵

الف) میانگین به روش مستقیم:

$$\mu_x = \frac{\Sigma F_i x_i}{N} = \frac{۲۷۸۵}{۵۰} = ۵۵/۷$$

ب) مد:

$$Mo = L_{Mo} + (\frac{d_۱}{d_۱ + d_۲}) I$$

$$L_{Mo} = ۳۷/۵ \qquad d_۱ = ۱۰ - ۹ = ۱ \qquad d_۲ = ۱۰ - ۸ = ۲$$

$$Mo = ۳۷/۵ + (\frac{۱}{۱ + ۲}) \times ۱۴ = ۴۲/۱۷$$

ج) میانه:

$$Md = L_{Md} + (\frac{\frac{N}{۲} - FC_{i-۱}}{F_i}) I$$

$$C_{Md} = \frac{N}{۲} = \frac{۵۰}{۲} = ۲۵$$

$$Md = ۵۱/۵ + (\frac{۲۵ - ۲۵}{۸}) \times ۱۴ = ۵۱/۵$$

د) دهک اول:

$$C_{D_۱} = \frac{۱ \times ۵۰}{۱۰} = ۵$$

$$D_۱ = ۹/۵ + (\frac{۵ - ۰}{۶}) \times ۱۴ = ۲۱/۱۷$$

ه) صدک نودم:

$$C_{P_{۹۰}} = \frac{۹۰ \times N}{۱۰۰} = \frac{۹۰ \times ۵۰}{۱۰۰} = ۴۵$$

$$P_{۹۰} = ۹۳/۵ + (\frac{۴۵ - ۴۴}{۴}) \times ۱۴ = ۹۷$$

چنانکه ملاحظه کردید برای محاسبه مد، میانه، دهک اول و صدک نودم از کرانه پایین طبقه استفاده شد.

۴ـ۲ پارامترهای پراکندگی در داده‌های طبقه‌بندی شده

چنانکه گفته شد یکی از خواص میانگین $\sum_{i=1}^{N}(x_i - \mu_x) = 0$ است. براساس همین قضیه پارامترهای پراکندگی مثل انحراف متوسط از میانگین و واریانس پیش از این مطرح شد. از آنجا که این خاصیت برای داده‌های طبقه‌بندی شده نیز ـ البته با اندک تغییر ـ مصداق دارد؛ یعنی: $\sum F_i (x_i - \mu_x) \simeq 0$ ، روابط بیان شده در فصل دوم برای پارامترهای پراکندگی، در این فصل نیز با اندک تغییری به کار گرفته می‌شود.

انحراف متوسط از میانگین را می‌توان به این صورت تعریف کرد:

$$A.D_{\mu} = \frac{\sum F_i \mid x_i - \mu_x \mid}{N} \qquad (۴\text{ـ}۱۱)$$

و برای محاسبهٔ واریانس دو روش مطرح است:

روش اول به روش مستقیم معروف است. در این روش واریانس را می‌توان به کمک روابط ۴ـ۱۲ و ۴ـ۱۳ محاسبه کرد:

$$\sigma_x^2 = \frac{\sum F_i (x_i - \mu_x)^2}{N} \qquad (۴\text{ـ}۱۲)$$

$$\sigma_x^2 = \frac{\sum F_i x_i^2}{N} - \mu_x^2 \qquad (۴\text{ـ}۱۳)$$

رابطهٔ ۴ـ۱۳ را می‌توان از رابطهٔ ۴ـ۱۲ استخراج کرد که در فصل دوم بخش ۴ـ۳ـ۲ توضیح داده شد.

روش دوم به روش غیرمستقیم معروف است. چنانچه میانگین مشاهدات طبقه‌بندی شده اعشاری یا چند رقمی باشد، بیشتر از این روش استفاده می‌شود. در این روش واریانس براساس رابطهٔ ۴ـ۱۴ تعریف می‌شود؛ یعنی:

$$\sigma_x^2 = I^2 \left[\frac{\sum F_i d_i^2}{N} - \left(\frac{\sum F_i d_i}{N} \right)^2 \right] \qquad (۴\text{ـ}۱۴)$$

اثبات: اگر A میانگین اختیاری، I فاصلهٔ طبقات، $d_i = \frac{x_i - A}{I}$ و $x_i = A + Id_i$ باشد، خواهیم داشت:

$$\sigma_x^2 = \frac{\sum F_i (x_i - \mu_x)^2}{N}$$

$$= \frac{\sum F_i [A + Id_i - (A + (\frac{\sum F_i d_i}{N}) I)]^2}{N}$$

$$= \frac{\sum F_i [A + Id_i - A - (\frac{\sum F_i d_i}{N}) I]^2}{N}$$

$$= \frac{\sum F_i [I(d_i - (\frac{\sum F_i d_i}{N}))]^2}{N}$$

$$= \frac{I^2 \sum F_i (d_i - \bar{d})^2}{N}$$

$$= \frac{I^2 \sum F_i (d_i^2 - 2d_i \bar{d} + \bar{d}^2)}{N}$$

$$= \frac{I^2 (\sum F_i d_i^2 - 2\bar{d} \sum F_i d_i + N\bar{d}^2)}{N}$$

$$= \frac{I^2 [\sum F_i d_i^2 - N\bar{d}^2]}{N}$$

$$= I^2 [\frac{\sum F_i d_i^2}{N} - (\frac{\sum F_i d_i}{N})^2]$$

آخرین عبارت همان رابطهٔ ۱۴ـ۴ است.

مثال ۸ـ۴ مجدداً داده‌های طبقه‌بندی شده دستمزد ماهیانهٔ کارکنان شرکت ایران دارو (مثال ۱ـ۴) را در نظر می‌گیریم و انحراف متوسط از میانگین و واریانس آن را محاسبه می‌کنیم.

C-L	F_i	x_i	$F_i\ \lvert x_i-\mu_x \rvert$	$F_i\,(x_i-\mu_x)^۲$	$d_i^۲$	$F_i\,d_i^۲$
۱۰_۲۰	۱۵	۱۵	۲۷۰	۴۸۶۰	۴	۶۰
۲۰_۳۰	۳۰	۲۵	۲۴۰	۱۹۲۰	۱	۳۰
۳۰_۴۰	۲۵	۳۵	۵۰	۱۰۰	۰	۰
۴۰_۵۰	۲۰	۴۵	۲۴۰	۲۸۸۰	۱	۲۰
۵۰_۶۰	۱۰	۵۵	۲۲۰	۴۸۴۰	۴	۴۰
	$N=۱۰۰$		$\sum F_i\ \lvert x_i-\mu_x \rvert = ۱۰۲۰$	$\sum F_i\,(x_i-\mu_x)^۲ = ۱۴۶۰۰$		$\sum F_i\,d_i^۲ = ۱۵۰$

می‌دانیم که $\mu_x = ۳۳$ است؛ پس:

$$A.D_\mu \simeq \frac{۱۰۲۰}{۱۰۰} = ۱۰/۲۰$$

$$\sigma_x^۲ \simeq \frac{\sum F_i\,(x_i-\mu_x)^۲}{N} = \frac{۱۴۶۰۰}{۱۰۰} = ۱۴۶$$

$$\sigma_x^۲ \simeq I^۲\left[\frac{\sum F_i\,d_i^۲}{N} - \left(\frac{\sum F_i\,d_i}{N}\right)^۲\right]$$

$$= (۱۰)^۲\left[\frac{۱۵۰}{۱۰۰} - \left(\frac{-۲۰}{۱۰۰}\right)^۲\right] = ۱۴۶$$

بنابراین انحراف معیار برابر با $\sigma_x = \sqrt{۱۴۶} = ۱۲/۰۸۳$ خواهد شد.

۱-۲-۴ دامنهٔ میان چارکی و انحراف چارکی

در فصل دوم گفته شد که این دو پارامتر پراکندگی برای توزیعهایی مناسبند که در دنباله‌ها دارای مشاهدات اندکی هستند؛ به عبارت دیگر مشاهدات اندک در دنباله‌ها در این پارامترها نسبت به سایر پارامترهای پراکندگی مثل انحراف متوسط از میانگین و واریانس کمتر تأثیر می‌گذارد.

در برخی از توزیعهای فراوانی، خصوصیات جامعه و توزیعهای پراکندگی به گونه‌ای است که امکان محاسبهٔ پارامترهایی چون انحراف معیار نیست و این در صورتی است که دنباله‌های توزیع نامعین و باز باشد؛ مثل ۱۱۰ و کمتر، ۲۵۰ و بیشتر. از آنجا که مشاهدات این دسته از طبقات ناچیز است، چنانچه ملاک پراکندگی، چندکها

باشد مشاهدات به طور کامل در چهارچوب محاسبات قرار نمی‌گیرند؛ برای نمونه به مثال ۹ـ٤ توجه کنید:

مثال ۹ـ٤ محققی برای اندازه گیری نمرهٔ رهبری پرسشنامه‌ای طراحی کرده است. در این پرسشنامه نمرات رهبری مدیران از صفر تا ۲۰ طبقه‌بندی شده، سپس پرسشنامه تکثیر و در یکی از سازمانهای دولتی توزیع شده است. نمرات حاصل از پرسشنامه‌ها در این جدول خلاصه شده است:

فاصله طبقات C-L	F_i	F_{Ci}
۵>	۲	۲
۵ـ۸	۱۰	۱۲
۸ـ۱۱	۲۵	۳۷
۱۱ـ۱٤	۲۳	٦۰
۱٤ـ۱۷	۱۵	۷۵
۱۷≤	۵	۸۰
	N = ۸۰	

از آنجا که حد پایین طبقهٔ اول و حد بالای طبقهٔ آخر باز است، امکان محاسبهٔ میانگین نیست؛ بنابراین محاسبهٔ پارامتر مهمی چون انحراف معیار نیز غیرممکن است. در اینجا چاره‌ای نیست جز آنکه چندکها را برای توصیف مرکزیت و پراکندگی نمرات رهبری مدیران به کار بگیریم؛ از این رو به محاسبهٔ دامنه میان چارکی و انحراف چارکی می‌پردازیم و $Q_۱$ و $Q_۳$ را محاسبه می‌کنیم:

$$IQR = Q_۳ - Q_۱$$

$$SIQR = \frac{(Q_۳ - Q_۱)}{۲}$$

$$C_{Q_۱} = \frac{۱ \times ۸۰}{٤} = ۲۰$$

$$Q_۱ = ۸ + \frac{(۲۰ - ۱۲)}{۲۵} \times ۳ = ۸/۹٦$$

$$C_{Q_۳} = \frac{۳ \times ۸۰}{٤} = ٦۰$$

$$Q_{۳} \simeq ۱٤ + \frac{(٦۰ - ٦۰)}{۱٥} \times ۳ = ۱٤$$

$$IQR = (۱٤ - ۸/۹٦) = ٥/۰٤$$

$$SIQR = \frac{(۱٤ - ۸/۹٦)}{۲} = ۲/٥۲$$

مقدار ۲/٥۲ تقریبی مناسب از پراکندگی جامعه است. انحراف چارکی نشان می‌دهد که متوسط پراکندگی نمرات رهبری مدیران سازمان حدوداً ۲/٥۲ نمره است.

۲ـ۲ـ٤ واریانس و تصحیح شپارد[1]

در داده‌های طبقه‌بندی شده، برای محاسبهٔ میانگین و واریانس از نمایندهٔ طبقات استفاده می‌شود. نتیجه این عمل ممکن است دارای اختلاف با مقادیر واقعی داده‌ها باشد. در میانگین، اشتباه ناشی از این تقریب به علت مثبت و منفی بودن اشتباهات جبران می‌شود؛ از این رو از مجموع خطاها صرف‌نظر می‌گردد. در مورد واریانس چون خطاهای مثبت و منفی به توان دو می‌رسد، خطاها یکدیگر را خنثی نمی‌کنند؛ بنابراین مقداری که برای واریانس به دست می‌آید از مقدار واقعی واریانس بیشتر است. برای تصحیح این اشتباه، شپارد پیشنهاد کرده است از واریانس به دست آمده $\frac{I^{۲}}{۱۲}$ کم شود ـ $I^{۲}$ مربع فاصله طبقات است ـ که در آن صورت مقدار به دست آمده دقیقتر از واریانس خواهد بود. این واریانس را واریانس تصحیح شده گویند و آن را با $\sigma_c^{۲}$ نمایش می‌دهند و مقدارش برابر است با:

$$\sigma_c^{۲} = \sigma_x^{۲} - \frac{I^{۲}}{۱۲} \qquad (٤\text{ـ}۱٥)$$

باید توجه داشت که این تصحیح در مواردی به کار می‌رود که اولاً متغیر پیوسته باشد، ثانیاً تعداد N دست کم هزار تا باشد و ثالثاً تابع توزیع فراوانی از نوع متقارن یا اندکی متقارن باشد.

مثال ۱۰ـ٤ در اینجا واریانس اصلاح شدهٔ مثال ۸ـ٤ را محاسبه می‌کنیم.

---

1. Sheppard

واریانس اولیه مساوی ۱۴۶ است؛ بنابراین واریانس تصحیح شده عبارت خواهد بود از:

$$\sigma_c^2 = ۱۴۶ - \frac{(۱۰)^2}{۱۲} = ۱۳۷/۶۶۷$$

بنابراین انحراف معیار اصلاح شده مساوی خواهد بود با:

$$\sigma_c = \sqrt{\sigma_c^2} = \sqrt{۱۳۷/۶۶۷} = ۱۱/۷۳۳$$

با توجه به انحراف معیار اصلاح نشده که مساوی ۱۲/۰۸۳ بود، مقدار اشتباه حدوداً ۰/۳۵ است که به علت طبقه‌بندی مشاهدات ایجاد شده است.

۴ـ۳ عملیات جبری میانگین و واریانس

اگر K جامعه با تعداد مشاهدات N_1، N_2، ...، N_K با میانگینهای μ_1، μ_2، ...، μ_K و با واریانسهای σ_1^2، σ_2^2، ...، σ_K^2 به صورت جامعه‌ای واحد ترکیب شده و یک جامعهٔ کل را پدید آورده باشند، میانگین حسابی و واریانس آن جامعهٔ کل با استفاده از روابط ۴ـ۱۶ و ۴ـ۱۷ چنین به دست خواهد آمد:

$$\mu = \frac{N_1\mu_1 + N_2\mu_2 + \dots + N_K\mu_K}{N_1 + N_2 + \dots + N_K} \qquad (۴\text{ـ}۱۶)$$

$$= \frac{\sum N_i \mu_i}{N}$$

$$\sigma^2 = \frac{\sum N_i \sigma_i^2}{N} + \frac{\sum N_i (\mu_i - \mu)^2}{N} \qquad (۴\text{ـ}۱۷)$$

این روابط را می‌توان در قالب این قضیه بیان کرد که «روی میانگین و واریانسها می‌توان عمل جبری انجام داد». در این قضیه نکات نظری و کاربردی ویژه‌ای نهفته است که در بحثهای بعدی و مثال ۴ـ۱۱ به برخی از آنها پی خواهید برد.

مثال ۴ـ۱۱ اخیراً چهار طرح طبقه‌بندی مشاغل به سازمان امور اداری و استخدامی کشور پیشنهاد شده است که مشخصات میانگین و واریانس نمرهٔ ارزشیابی کارکنان، در

هریک از آنها در این جدول خلاصه شده است:

طرح طبقه‌بندی (i)	۱	۲	۳	۴	
تعداد کارکنان سازمان (N_i)	۲۰۰	۱۰۰	۳۰۰	۴۰۰	$N=1000$
میانگین (μ_i)	۲۵۰	۴۰۰	۳۰۰	۳۰۰	
واریانس (σ_i^2)	۱۰۰۰	۸۰۰	۶۰۰	۸۵۰	

کارشناسان سازمان امور اداری و استخدامی کشور درصدد هستند میانگین و واریانس کل این چهار طرح را به دست آورند؛ بنابراین با استفاده از روابط ۱۶ـ۴ و ۱۷ـ۴ پارامترهای مورد نظر خود را چنین محاسبه کرده‌اند:

$$\mu=\frac{(200\times 250)+(100\times 400)+(300\times 300)+(400\times 300)}{200+100+300+400}=\frac{300000}{1000}=300$$

$$\sigma^2=\frac{(200\times 1000)+(100\times 800)+(300\times 600)+(400\times 850)}{200+100+300+400}$$

$$+\frac{200(250-300)^2+100(400-300)^2+300(300-300)^2+400(300-300)^2}{1000}$$

$$=\frac{800000}{1000}+\frac{1500000}{1000}=2300$$

متوسط نمرهٔ ارزشیابی برای چهار طرح ۳۰۰ و انحراف معیار نمرات ۴۷/۹۶ نمره است.

از رابطهٔ ۱۷ـ۴ این قضیه استنتاج می‌شود: «هرگاه چند جامعه به صورت یک جامعهٔ کل ترکیب شوند، واریانس این جامعه بزرگتر از میانگین واریانس جوامع تشکیل‌دهندهٔ آن خواهد بود و فقط در صورتی واریانس جامعهٔ کل با میانگین واریانسهای جوامع تشکیل‌دهنده برابر خواهد بود که میانگینهای این جوامع یکسان باشد»؛ یعنی:

$$\sigma^2\geq\frac{\sum N_i\sigma_i^2}{N}$$

۴ـ۴ اهمیت کاربردی انحراف معیار

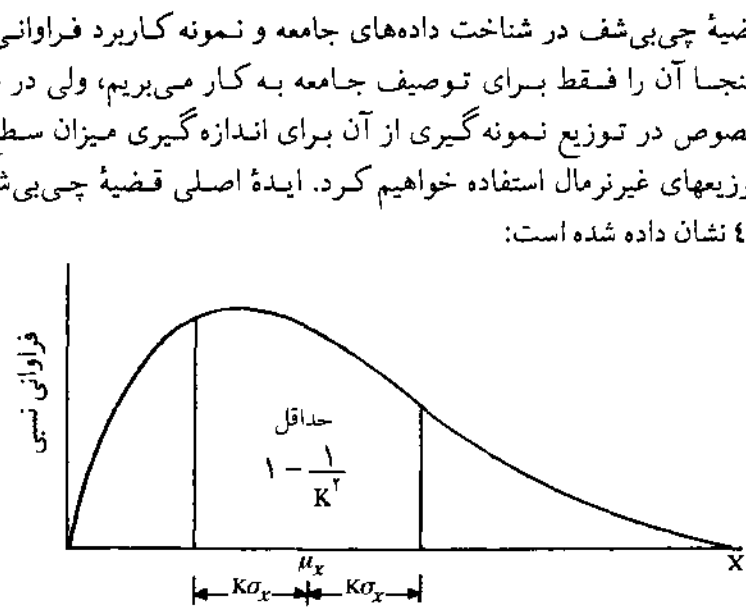
حال به ذکر یک قضیهٔ مفید و عملی در آمار که به وسیلهٔ چی‌بی‌شف [1] ریاضیدان معروف

1. Chebysheff

روسی ارائه شده است، می‌پردازیم. اثبات ریاضی قضیه را می‌توان در متون آمار ریاضی دید، ولی ما در اینجا آن را حذف کرده‌ایم.

قضیه: برای N مشاهده X_1، X_2، ...، X_N، حداقل $[1-\frac{1}{K^2}]$ درصد از مشاهدات در دامنهٔ K انحراف معیار از میانگین قرار می‌گیرند. K عددی مثبت است که مقادیر مساوی یا بزرگتر از یک را شامل می‌شود.

قضیهٔ چی‌بی‌شف در شناخت داده‌های جامعه و نمونه کاربرد فراوانی دارد. ما در اینجا آن را فقط برای توصیف جامعه به کار می‌بریم، ولی در فصول بعدی بخصوص در توزیع نمونه گیری از آن برای اندازه گیری میزان سطح زیر منحنی توزیعهای غیرنرمال استفاده خواهیم کرد. ایدهٔ اصلی قضیهٔ چی‌بی‌شف در شکل ۲ـ٤ نشان داده شده است:



شکل ۲.٤ نمایش هندسی قضیهٔ چی‌بی‌شف

به ازای هر مقدار بزرگتر یا مساوی یک می‌توان برای K حداقل سطح زیر منحنی یا درصد داده‌هایی را که حول میانگین متمرکز شده‌اند، پیدا نمود. ابتدا قضیه را به ازای مقادیر مختلف K آزمایش می‌کنیم. وقتی که ۱ = K است، مشاهده می‌شود که تمرکز داده‌ها از فاصلهٔ $\mu_x - K\sigma_x$ تا $\mu_x + K\sigma_x$ حداقل صفر $[1-\frac{1}{1}=0]$ است که البته در این صورت قضیه قابل استفاده نخواهد بود. چنانچه ۲ = K باشد، مشاهده خواهد شد که $[1-\frac{1}{2^2}=\frac{3}{4}]$ است؛ یعنی اینکه حداقل ۷۵ درصد مشاهدات جامعه در فاصلهٔ $\mu_x - 2\sigma_x$ تا $\mu_x + 2\sigma_x$ قرار دارند. هرچند اگر ۳ و ۲ = K باشد در عمل بیشتر مورد استفاده قرار می‌گیرد، ولی هیچ ضرورتی ندارد که K حتماً عدد صحیحی باشد؛ برای مثال اگر ۲/۵ = K باشد، باز قضیه کارآمد است و می‌توان نشان داد که تراکم داده‌ها حول میانگین حداقل $[1-\frac{1}{(2/5)^2}=0/84]$ است.

جدول ۱ـ٤ نشان‌دهندهٔ مقادیر مختلف K و حداقل پوشش مشاهدات در فاصلهٔ K برابر انحراف معیار از میانگین است.

جدول ۱ـ٤ نمایش مقادیر $\left[1-\frac{1}{K^2}\right]$

K	$1-\frac{1}{K^2}$
۱	۰
۲	$\frac{3}{4}$
۳	$\frac{8}{9}$

برای پی بردن به کاربرد قضیهٔ چی‌بی‌شف به این مثال توجه کنید.

مثال ۱۲ـ٤ بسیاری از سازمانها در طول خدمت کارکنان آنها را از نظر مسئولیت‌پذیری می‌آزمایند. در شرکت رادیاتور ایران نیز این آزمون انجام شده و میانگین و واریانس نمرات کارکنان به ترتیب ۷۵ و ۱۰۰ بوده است. از قضیهٔ چی‌بی‌شف برای توصیف توزیع نمرات مسئولیت‌پذیری شرکت استفاده می‌کنیم.

۱. حداقل ۷۵ درصد نمرات کارکنان شرکت درفاصلهٔ $(\mu_x \pm 2\sigma_x)=(75\pm 2\times 10)$ یا ۵۵ تا ۹۵ قرار دارد.

۲. حداقل $\frac{8}{9}$ نمرات مسئولیت‌پذیری کارکنان شرکت در فاصلهٔ $(\mu_x \pm 3\,\sigma_x) = (75 \pm 3 \times 10)$ یا ٤۵ تا ۱۰۵ قرار دارد.

تأکید بر قید «حداقل»[1] در قضیهٔ چی‌بی‌شف به دلیل قابلیت تعمیمش به انواع توزیعهای فراوانی است. حال برای بررسی صحت روابط مذکور، به صورت استقرایی حدود تراکم مشاهدات را به ازای K برابر σ_x حول میانگین در توزیع نرمال بیان می‌کنیم. اگر یک توزیع زنگی شکل[2] وجود داشته باشد، می‌توان دید که داده‌ها به ترتیب ذیل حول میانگین متمرکز شده‌اند:

(۱) $(\mu_x \pm \sigma_x)$: ٪۶۸/۲۶

(۲) $(\mu_x \pm 2\sigma_x)$: ٪۹۵/٤٤

(۳) $(\mu_x \pm 3\sigma_x)$: ٪۹۹/۷۳

---

1. at least
2. bell shaped

تأکید بر قید حداقل در قضیهٔ چی‌بی‌شف نشان می‌دهد که فراوانیهای نسبی حاصل در توزیع نرمال با قضیهٔ چی‌بی‌شف سازگاری دارد.

مثال ۱۳ـ٤ مدیر شرکت ماشین‌سازی، مدت زمان مورد انتظار برای تکمیل عملیات تولیدی را مورد مطالعه قرار داده‌است. حاصل مطالعه بر روی ٤۰ کارگر نشان می‌دهد که میانگین و پراکندگی زمان مورد نیاز به ترتیب ۱۲/۸ و ۱/۷ دقیقه است. داده‌ها را با استفاده از قاعده چی‌بی‌شف و توزیع نرمال توصیف می‌کنیم.

برای توصیف داده‌ها فواصل را محاسبه و با همدیگر مقایسه می‌کنیم:

$$(\mu_x \pm \sigma_x) = (۱۲/۸ \pm ۱/۷) \text{ یا } (۱۱/۱ - ۱٤/۵) \qquad (۱)$$

$$(\mu_x \pm ۲\sigma_x) = (۱۲/۸ \pm (۲\times۱/۷)) \text{ یا } (۹/٤ - ۱٦/۲) \qquad (۲)$$

$$(\mu_x \pm ۳\sigma_x) = (۱۲/۸ \pm (۳\times۱/۷)) \text{ یا } (۷/۷ - ۱۷/۹) \qquad (۳)$$

با پذیرفتن توزیع نرمال برای زمان تکمیل عملیات ٦۸/۲٦ درصد مشاهدات در فاصله ۱۱/۱ تا ۱٤/۵ دقیقه؛ ۹۵/٤٤ درصد داده‌ها در فاصله ۹/٤ تا ۱٦/۲ و ۹۹/۷۳ درصد مشاهدات در فاصله ۷/۷ تا ۱۷/۹ دقیقه قرار گرفته است. چنانچه در نرمال بودن زمان عملیات شک کنیم ناچاریم که حدود توزیع مشاهدات را با استفاده از قضیهٔ چی‌بی‌شف توصیف کنیم. قضیه نشان می‌دهد که حداقل ۷۵ درصد مشاهدات در فاصلهٔ ۹/٤ تا ۱٦/۲ و حداقل $\frac{۸}{۹}$ مشاهدات در فاصلهٔ ۷/۷ تا ۱۷/۹ دقیقه قرار دارد.

حال این سؤال را مطرح می‌کنیم که تا چه اندازه می‌توان سطوح توزیع نرمال را برای نرخ قیمت به درآمد ($\frac{P}{E}$) در جدول ۱ـ۳ به کار برد. محاسبات نشان می‌دهد که $\mu_x = ۱٦$ و $\sigma_x = ۵/٦$ است؛ پس می‌توانیم توزیع فراوانی مطلق و نسبی را به ازای $K = ۱, ۲, ۳$ با فرض نرمال بودن محاسبه کنیم.

جدول ۲ـ٤ توزیع فراوانی نرخ $\frac{P}{E}$

C-L	F_i	f_i
۵ـ۹	۳	۰/۱۲
۹ـ۱۳	۵	۰/۲۰
۱۳ـ۱۷	۷	۰/۲۸
۱۷ـ۲۱	٦	۰/۲٤
۲۱ـ۲۵	۳	۰/۱۲۰
۲۵ـ۲۹	۱	۰/۰٤

با در نظر گرفتن میانگین و پراکندگی جدول ۲ـ۴، جدول ۳ـ۴ را تکمیل می‌کنیم.

جدول ۳ـ۴ درصد فراوانی قرار گرفته در دامنه $\mu_x \pm ۲\sigma_x$ برای نرخ $\frac{P}{E}$

K	$(\mu_x \pm K\sigma_x)$	فراوانی مطلق	فراوانی نسبی	درصدهای توزیع نرمال	میزان اختلاف با توزیع نرمال (به درصد)
۱	۱۰/٤ـ۲۱/٦	۱٦	۰/٦٤	٦۸/۲۷	٤/۲۷
۲	٤/۸ـ۲۷/۲	۲٤	۰/۹٦	۹۵/٤٤	۰/۵٦
۳	−۰/۸۰ـ۳۲/۸	۲۵	۱	۹۹/۷۳	۰/۲۷

چنانکه مشخص است نتایج حاصل با قاعدهٔ فواصل برای توزیع نرمال تقریباً سازگاری دارد؛ در حالی که توزیع نرخ $\frac{P}{E}$ در ۲۵ شرکت کاملاً نرمال نیست ولی حداکثر اختلاف با توزیع نرمال ٤/۲۷ درصد است که تقریباً قابل اغماض است. خلاصه آنکه نتایج جدول ۳ـ٤ نشان می‌دهد که برای توصیف توزیع فراوانیها می‌توان از درصدهای توزیع نرمال استفاده کرد و حتی نتایج حاصل از قضیهٔ چی‌بی‌شف را امیدوارکننده دانست.

تمرین

۱. اندازه طول ۵۰ میلهٔ آهنی برحسب سانتیمتر مطابق این جدول توزیع شده است:

الف) پارامترهای مرکزی را محاسبه نمایید.

ب) پارامترهای پراکندگی را محاسبه نمایید.

حدود طبقات	فراوانی
۲/۳۵ـ۲/٤۵	۱
۲/٤۵ـ۲/۵۵	٤
۲/۵۵ـ۲/٦۵	۷
۲/٦۵ـ۲/۷۵	۱۵
۲/۷۵ـ۲/۸۵	۱۱
۲/۸۵ـ۲/۹۵	۱۰
۲/۹۵ـ۳/۰۵	۲
	N = ۵۰

۲. این جدول نشان‌دهنده میانگین و واریانس سه جامعه است:

الف) میانگین کل را محاسبه نمایید.

ب) واریانس کل را محاسبه نمایید.

N_i	۱۰۰	۱۲۰	۱۸۰
μ_i	۵	۷	۸
$\sigma_i^۲$	۱٤	۱٦	۱۰

۳. این جدول توزیع فراوانی را در نظر بگیرید:
پارامترهای مرکزی و پراکندگی مناسب را محاسبه نمایید.

فراوانی	حدود طبقات
۱۵	۱۰۰_۱۲۰
۳۰	۱۲۰_۱۴۰
۲۵	۱۴۰_۱۶۰
۴۰	۱۶۰_۱۸۰
۵۰	۱۸۰_۲۰۰
۴۰	≥ ۲۰۰

۴. در یک سیستم پرداخت دستمزد، امتیازات را بر روی منحنی نرمال برده‌اند؛ برای این کار ابتدا میانگین و انحراف معیار امتیازات کارکنان را محاسبه و سپس طبق این جدول عمل کرده‌اند.

امتیاز	دامنهٔ استاندارد
عالی	$X \geq \mu_x + ۲\sigma_x$
خوب	$\mu_x + \sigma_x \leq X < \mu_x + ۲\sigma_x$
متوسط	$\mu_x \leq X < \mu_x + \sigma_x$
بد	$\mu_x - \sigma_x \leq X < \mu_x$
بسیار بد	$X < \mu_x - \sigma_x$

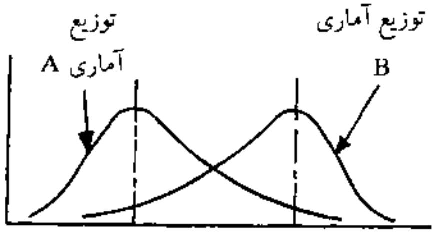
در یک سازمان ۱۰۰۰ نفری، هریک از این امتیازات تقریباً به چندکارمند تعلق خواهد گرفت؟

۵. اهمیت خواص جبری میانگین و واریانس را بیان نمایید.

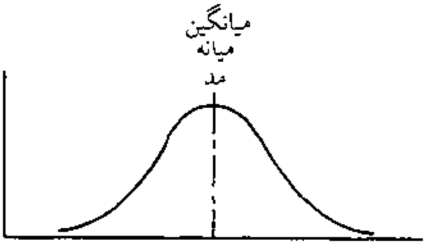
۴ـ۵ پارامترهای تعیین انحراف از قرینگی

در مقایسهٔ دو یا چند جامعه با همدیگر، ابتدا از پارامترهای مرکزی استفاده می‌شود. گاهی اوقات تصمیم‌گیرنده به‌علت مساوی بودن پارامترهای مرکزی (بخصوص میانگین) دچار مشکل می‌گردد، در این صورت اختلاف جوامع آماری به کمک شاخصهای پراکندگی مثل انحراف معیار مشخص می‌شود. پارامترهای پراکندگی نیز گاهی اوقات به علت مساوی بودن جوابگوی تصمیم‌گیرنده نیستند.

در شکل ۳ـ٤ دو توزیع آماری نشان داده خواهد شد که دارای میانگین و واریانس مساوی هستند، اما دو جامعه از توزیع یکسانی برخوردار نیستند. توزیع جامعهٔ A دارای تراکم در نزدیکی مبدأ مختصات است؛ در حالی‌که مد جامعهٔ B در نقطهٔ مقابل آن قرار دارد. این تفاوت را چولگی[1] یا انحراف از قرینگی گویند. چولگی توزیعها در مقایسه با توزیع متقارن معین می‌شود. شکل ٤ـ٤ نشان‌دهندهٔ یک توزیع متقارن است. توزیع متقارن توزیعی است که پارامترهای مرکزی آن (مد، میانگین و میانه) با همدیگر مساوی باشند. هرچه یک توزیع با شکل ٤ـ٤ بیشتر تفاوت داشته باشد، انحراف آن از قرینگی بیشتر خواهد بود.



شکل ۳ـ٤ توزیعهای نامتقارن A و B



شکل ٤ـ٤ توزیع زنگی شکل و متقارن

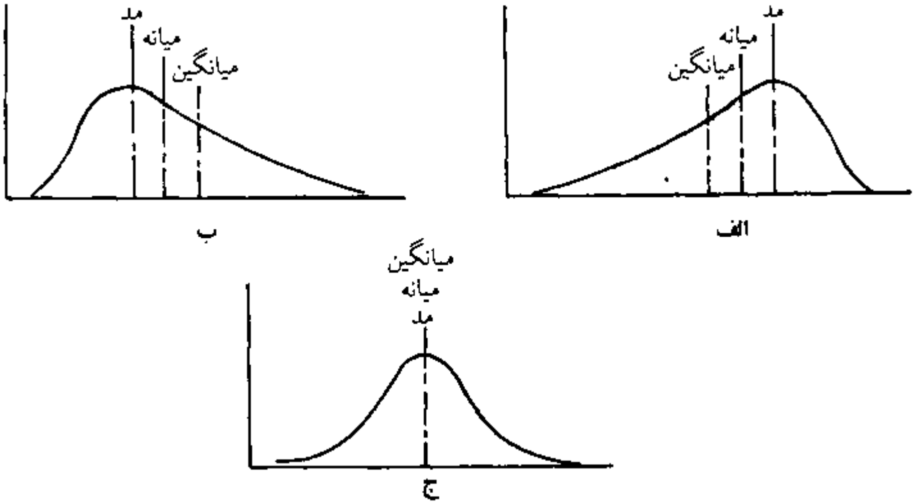
چوله به راست با توجه به تعریف میانه (ارزش عددی که از نظر مکان در وسط جامعه قرار می‌گیرد) وقتی است که مد جامعهٔ آماری پایین‌تر از میانه و افتادگی توزیع بالاتر از آن واقع شود. اگر مد جامعه بزرگتر از میانه باشد و افتادگی جامعه سمت چپ آن واقع گردد، جامعه دارای چوله به چپ خواهد بود. در شکل ۳ـ٤ توزیع آماری A دارای چوله به راست و توزیع آماری B دارای چوله به چپ است.

شاخص اندازه‌گیری پارامترهای تعیین انحراف از قرینگی، «ضریب چولگی» است که با SK نمایش داده می‌شود. در صورتی که جامعه دارای چوله به چپ باشد، ضریب چولگی منفی و در صورتی که دارای چوله به راست باشد ضریب چولگی مثبت و اگر جامعه از توزیع متقارن برخوردار باشد، ضریب چولگی مساوی صفر خواهد بود.

بررسی نمودارهای «الف»، «ب» و «ج» در شکل ۵ـ٤ نشان می‌دهد که در توزیع متقارن $Mo = Md = \mu_x$ است. در توزیعهایی که چولگی مثبت است $Mo < Md < \mu_x$ است؛ در حالی که اگر $Mo > Md > \mu_x$ باشد، جامعه دارای ضریب چولگی منفی خواهد بود.

---

1. skewness



شکل ٤.٥ نمایش ترتیب پارامترهای مرکزی در توزیعهای آماری

قدر مطلق ضریب چولگی نشان‌دهندهٔ میزان اختلاف جامعهٔ آماری با توزیع نرمال از نظر قرینگی است. بدیهی است هرچه $|SK|$ بزرگتر باشد، تفاوت جامعه از نظر قرینگی، با توزیع نرمال بیشتر خواهد بود به طوری که:

۱. اگر $|SK| \leq ۰/۱$ باشد، جامعه از نظر قرینگی تقریباً نرمال است.

۲. اگر $۰/۱ < SK \leq ۰/۵$ باشد، جامعه از نظر قرینگی دارای تفاوت اندکی با توزیع نرمال است.

۳. اگر $SK > ۰/۵$ باشد، جامعه از نظر قرینگی دارای تفاوت فاحشی با توزیع نرمال است.

برای محاسبهٔ ضریب چولگی، فرمولهایِ فراوانی ارائه شده است که هریک از آنها به شرح ذیل خواهد آمد.

رابطهٔ ٤ـ١٨ مهمترین فرمولی است که برای محاسبهٔ ضریب چولگی مورد استفاده قرار می‌گیرد:

$$SK = \frac{r_3}{\sigma_x^3} \qquad (٤ـ١٨)$$

در این رابطه، $r_3 = \frac{\sum F_i (x_i - \mu_x)^3}{N}$ و σ_x همان انحراف معیار است. از آنجا که در رابطهٔ

۱۸ـ٤ از گشتاور[1] مرتبه سوم به مبدأ میانگین حسابی (r_3) و از گشتاور مرتبهٔ دوم به مبدأ میانگین حسابی (σ_x^2) استفاده شده است، آن را ضریب چولگی به طریقهٔ گشتاورها نیز می‌خوانند.

روابط ۱۹ـ٤ و ۲۰ـ٤ به ترتیب به ضریب چولگی پیرسون شماره ۱ و ۲ معروفند. پیرسون به تجربه دریافته بود که:

$$(\mu_x - \text{Mo}) = ۳(\mu_x - \text{Md})$$

ما نیز از این روش استفاده کرده، برای محاسبهٔ ضریب چولگی این روابط را ذکر می‌کنیم:

$$\text{S.K}_۱ = \frac{(\mu_x - \text{Mo})}{\sigma_x} \qquad (٤\text{ـ}۱۹)$$

$$\text{S.K}_۲ = \frac{۳(\mu_x - \text{Md})}{\sigma_x} \qquad (٤\text{ـ}۲۰)$$

گفته شد برای برخی از توزیعها، محاسبهٔ میانگین و واریانس امکان‌پذیر نیست. در برخی نیز محاسبهٔ آنها منطقی به نظر نمی‌رسد. پراکندگی این‌گونه توزیعها به کمک چندکها بیان شد. واضح است که برای محاسبه چولگی آنها هم باید از

---

۱. مفهوم گشتاور همان است که در علم مکانیک مطرح است. در علم مکانیک حاصلضرب مقدار نیرو را در فاصلهٔ عمودی میان نیرو و نقطهٔ مورد نظر، گشتاور آن نیرو نسبت به این نقطه می‌نامند. اگر به جای یک نیرو، مجموعه‌ای از نیروها موجود باشد، در این صورت گشتاور مجموعه اثر نیروها نسبت به یک نقطه با مجموعهٔ گشتاورهای هریک از نیروها نسبت به همان نقطه مساوی خواهد بود. توزیع مشاهدات، با مجموعهٔ نیروهایی که بر یک جرم معین اثر می‌گذارند شباهت کامل دارد. این جرم معین همان مبدأ داده‌هاست.

دستهٔ اول گشتاورها نسبت به مبدأ مختصات تعریف می‌شوند که آنها را مرتبهٔ n ام به مبدأ صفر می‌گویند. این نوع از گشتاورها را با μ_n نشان می‌دهند و به این صورت تعریف می‌کنند:

$$\mu_n = \frac{\sum F_i x_i^n}{N} \quad , \quad n = ۱, ۲, \ldots$$

دستهٔ دوم آنهایی هستند که نسبت به میانگین حسابی تعریف می‌شوند و آنها را گشتاور مرتبهٔ n ام به مبدأ μ_x می‌گویند. این نوع از گشتاورها که با r_n نشان داده می‌شوند از این رابطه به‌دست می‌آیند:

$$r_n = \frac{\sum F_i (x_i - \mu_x)^n}{N} \quad , \quad n = ۱, ۲, \ldots$$

پارامترهای چندکی مانند ضریب چولگی چارکی (رابطه ۲۱ـ٤) و ضریب چولگی صدکی (رابطه ۲۲ـ٤) استفاده کرد.

$$SK_Q = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} \qquad (٤\text{ـ}۲۱)$$

$$SK_P = \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}} \qquad (٤\text{ـ}۲۲)$$

مثال ۱٤ـ٤ روابط ۱۹ـ٤ تا ۲۲ـ٤ را به کمک داده‌های جدول طبقه‌بندی مثال ۷ـ٤ محاسبه می‌کنیم. مشخص شد که میانگین، مد، میانه، چارک اول و سوم، صدک دهم و انحراف معیار، ضریب چولگی شماره ۱ و ۲ پیرسون و ضریب چولگی چارکی عبارتند از:

$$\mu_x = 55/7 \qquad Mo = 42/17 \qquad Md = 51/5$$

$$Q_1 = 23/5 + \left(\frac{12/5 - 6}{9}\right) \times 14 = 36/61$$

$$Q_3 = 65/5 + \left(\frac{37/5 - 33}{6}\right) \times 14 = 76$$

$$P_{10} = 21/17 \qquad P_{90} = 97 \qquad \sigma_x = 27/434$$

$$S.K_1 = \frac{(55/7 - 42/17)}{27/434} = 0/493$$

$$S.K_2 = \frac{3(55/7 - 51/5)}{27/434} = 0/459$$

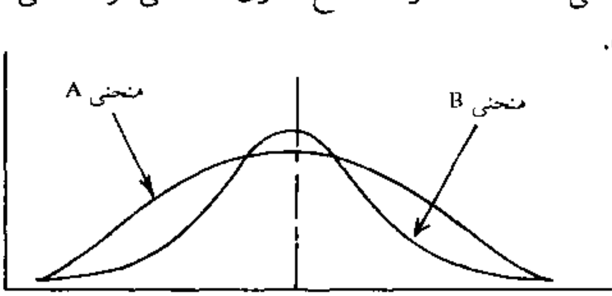
$$S.K_Q = \frac{76 - 2(51/5) + 36/61}{27/434} = 0/35$$

$$S.K_P = \frac{97 - 2(51/5) + 21/17}{27/434} = 0/55$$

چنانکه مشاهده می‌شود تمامی ضرایب محاسبه شده، چولگی مثبت را نشان می‌دهند؛ به عبارت دیگر تمایل سود سالانهٔ شرکتها به مقادیر کوچکتر است؛ یعنی $Mo < Md < \mu_x$. بعلاوه قدر مطلق ضرایب چولگی نشان می‌دهد که جامعه از نظر قرینگی دارای تفاوت اندکی با توزیع نرمال است.

٤ـ٦ پارامترهای تعیین انحراف ازکشیدگی

هرگاه از کشیدگی[1] توزیعها بحث می‌شود، در واقع مقدار اوج (بلندی) آنها موردنظر است؛ برای مثال نگاهی به منحنیهای A و B در شکل ٤ـ٦ نشان می‌دهد که تنها تفاوت آنها در بلندی آنهاست و علی‌رغم مساوی بودن پارامترهای مرکزی و چولگی، منحنی A کوتاهتر از منحنی B است که در اصطلاح آماری گفته می‌شود منحنی B کشیده‌تر از منحنی A است.



شکل ٤ـ٦ دو منحنی با پارامترهای مرکزی یکسان و کشیدگی متفاوت

در حالتهای خاص که تصمیم‌گیری به کمک پارامترهای مرکزی و چولگی امکانپذیر نیست، یکی از پارامترهای مناسب، استفاده از مقایسهٔ پراکندگی توزیع جامعه با توزیع نرمال است. شاخص سنجش پراکندگی جامعه نسبت به توزیع نرمال «ضریب کشیدگی» نام دارد که با E نشان داده می‌شود. با توجه به مقدار ضریب کشیدگی، درجات مختلفی برای کشیدگی پیدا خواهد شد که اندیشمندان آماری آنها را در سه گروه تقسیم کرده‌اند:

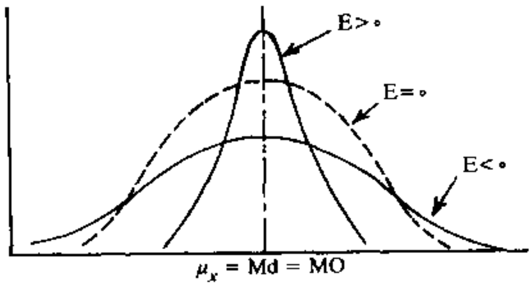
گروه *اول*. آن دسته از توزیعها که نسبت به توزیع نرمال از پراکندگی بیشتری برخوردارند، طبیعی است که پراکندگی بیشتر به دنبال تفرق داده حول میانگین پدید

1. kurtosis

می‌آید؛ یعنی منحنی توزیع نسبت به توزیع نرمال کوتاهتر[1] است. این دسته از توزیعها دارای ضریب کشیدگی منفی خواهد بود.

گروه دوم. توزیعهایی هستند که از توزیع نرمال بلندترند؛[2] یعنی از اوج بیشتری برخوردارند. علت اوج گرفتن توزیع آن است که داده‌ها حول میانگین متمرکزتر شده‌اند؛ به عبارت دیگر از پراکندگی داده‌ها کاسته شده است. ضریب کشیدگی این دسته از توزیعها مثبت خواهد بود.

گروه سوم. توزیعهایی هستند که کشیدگی آنها با کشیدگی توزیع نرمال کاملاً مساوی است. اینها توزیعهایی با کشیدگی متوسط[3] هستند. ضریب کشیدگی در این توزیعها مساوی صفر است. شکل ۷ـ٤ نشان‌دهندهٔ این سه دسته توزیع است.



شکل ۷ـ٤ مقایسهٔ انواع کشیدگی

علاوه بر علامت ضریب کشیدگی، قدر مطلق ($|E|$) آن نیز معنادار است. طبیعی است هرچه قدر مطلق ضریب کشیدگی بزرگتر باشد، تفاوت توزیع از نظر پراکندگی با توزیع نرمال بیشتر خواهد بود. در این زمینه استانداردهایی وجود دارد که عبارتند از:

۱. اگر $|E| \leq ۰/۱۰$ باشد، توزیع جامعه از نظر پراکندگی تقریباً نرمال است.

۲. اگر $۰/۱۰ < |E| \leq ۰/۵۰$ باشد، توزیع از نظر کشیدگی دارای تفاوت اندکی با توزیع نرمال است.

۳. اگر $|E| > ۰/۵۰$ باشد، تفاوت توزیع با توزیع نرمال به لحاظ پراکندگی فاحش است.

1. platy kurtic (Platy meaning Broad or flat)
2. lepty kurtic (Lepty meaning Slender)
3. meso kurtic (meso meaning: Intermediate)

برای محاسبهٔ ضریب کشیدگی روشهایی وجود دارد. چنانچه داده‌ها دارای طبقه‌بندی مشخص و کمّی باشند، بهترین روش برای محاسبهٔ ضریب کشیدگی استفاده از گشتاورهاست. ضریب کشیدگی گشتاوری عبارت است از:

$$E = \frac{r_4}{\sigma_x^4} - 3 \qquad (4\text{-}23)$$

$$r_4 = \frac{\sum F_i (x_i - \mu_x)^4}{N} \qquad (4\text{-}24)$$

در این رابطه، r_4 همان گشتاور مرتبهٔ چهارم به مبدأ μ_x و مقدار ثابت ۳ نشان‌دهنده کشیدگی گشتاوری توزیع نرمال است؛ بنابراین کسر $\frac{r_4}{\sigma_x^4}$ نشان‌دهنده کشیدگی هر توزیع دلخواهی است.

مثال ۴-۱۵ دراینجا ضریب کشیدگی گشتاوربرای درآمد کارکنان شرکت ایران دارو محاسبه می‌شود.

C-L	F_i	x_i	$F_i (x_i - \mu_x)^4$
۱۰-۲۰	۱۵	۱۵	۱۵۷۴۶۴۰
۲۰-۳۰	۳۰	۲۵	۱۲۲۸۸۰
۳۰-۴۰	۲۵	۳۵	۴۰۰
۴۰-۵۰	۲۰	۴۵	۴۱۴۷۲۰
۵۰-۶۰	۱۰	۵۵	۲۳۴۲۵۶۰
	N = ۱۰۰		$\sum F_i (x_i - \mu_x)^4$ = ۴۴۵۵۲۰۰

$$\mu_x = ۳۳ \qquad \sigma_x = ۱۲/۰۸۳$$

$$r_4 = \frac{۴۴۵۵۲۰۰}{۱۰۰} = ۴۴۵۵۲$$

$$E = \frac{۴۴۵۵۲}{(۱۲/۰۸۳)^4} - ۳ = -۰/۹۱$$

از آنجا که ضریب کشیدگی منفی است، مشخص می‌گردد که پراکندگی درآمد کارکنان شرکت از سطح نرمال بیشتر است. همچنین اینکه $۰/۵ < |-۰/۹۱|$ است نشان‌دهنده تفاوت فاحش توزیع جامعه با توزیع نرمال است.

برای آن دسته از توزیعهایی که با استفاده از چندکها توصیف می‌شوند نیز شاخص کشیدگی مناسبی وجود دارد که به ضریب کشیدگی چندکی معروف است.

$$E_p = \frac{SIQR}{P_{۹۰} - P_{۱۰}} - ۰/۲۶۳ \qquad (۴\_۲۵)$$

از این رابطه، برای تحلیل توزیعهایی که با استفاده از گشتاورها قابل توصیف نیستند، استفاده می‌شود.

مثال ۱۶ـ۴ ضریب کشیدگی صدکی را برای داده‌های طبقه‌بندی شده این جدول به این صورت محاسبه می‌کنیم:

C-L	F_i	Fc_i
۵>	۲	۲
۵ـ۸	۱۰	۱۲
۸ـ۱۱	۲۵	۳۷
۱۱ـ۱۴	۲۳	۶۰
۱۴ـ۱۷	۱۵	۷۵
۱۷≤	۵	۸۰
	N = ۸۰	

چندکهای توزیع برای این داده‌ها عبارتند از:

$$Q_1 = ۸/۹۶$$

$$Q_3 = ۱۴/۰۰$$

$$P_{۱۰} = ۶/۸۰$$

$$P_{۹۰} = ۱۶/۴۰$$

$$E_p = \frac{(۱۴/۰۰ - ۸/۹۶)/۲}{۱۶/۴۰ - ۶/۸۰} - ۰/۲۶۳ = -۰/۰۰۰۵$$

ضریب کشیدگی نشان می‌دهد که جامعه از پراکندگی مناسبی برخوردار است؛ به طوری که می‌توان گفت توزیع از نظر کشیدگی تقریباً نرمال است.

تمرین

۱. این جدول توزیع ساعات اضافه کاری سالیانه ۲۰۰ نفر ازکارکنان دانشگاه را نشان می‌دهد:

حدود طبقات	فراوانی
۵۰_۱۰۰	۲۰
۱۰۰_۱۵۰	۳۰
۱۵۰_۲۰۰	٤۰
۲۰۰_۲۵۰	۵۰
۲۵۰_۳۰۰	۲۵
۳۰۰_۳۵۰	۳۵
	N = ۲۰۰

الف) ضریب چولگی پیرسون شماره ۱ و ۲ را محاسبه و تحلیل کنید.
ب) ضریب کشیدگی گشتاوری را محاسبه و تحلیل کنید.

۲. مشاهدات مربوط به سر و صدای ترافیک بر حسب دسی‌بل به این صورت آمده است:

٦۷/۹	٦۵/۷	٦۲/۸	٦۳/۱	٦۱/۰	٦۰/۲	۵۹/٤	۵٦/۷	۵۵/۹	۵۲/۰
٦۸/۲	٦٦/۲	٦٤/۰	٦۲/٦	٦۱/٤	٦۰/۳	۵۹/۳	۵٦/۸	۵۵/۹	۵٤/٤
٦۸/۹	٦٦/۸	٦٤/٦	٦۲/۷	٦۱/۷	٦۰/۵	۵۹/۵	۵۷/۲	۵٦/۲	۵٤/۵
٦۹/٤	٦۷/۰	٦٤/۸	٦۳/۱	٦۱/۸	٦۰/٦	۵۹/۸	۵۷/٦	۵٦/٤	۵۵/۷
۷۷/۱	٦۷/۱	٦٤/۹	٦۳/٦	٦۲/۰	٦۰/۸	٦۰/۰	۵۸/۹	۵٦/٤	۵۵/۸

الف) مشاهدات را طبقه‌بندی کنید.
ب) پارامترهای مرکزی را محاسبه نمایید.
ج) پارامترهای پراکندگی را محاسبه نمایید.
د) قضیهٔ چی‌بی‌شف را به کمک مشاهدات بررسی کنید.
ه) پارامترهای کشیدگی را محاسبه کنید.
ز) ضریب چولگی گشتاوری را محاسبه نمایید.

۳. تفاوت پارامترهای تعیین انحراف از کشیدگی و پارامترهای تعیین انحراف از قرینگی را بنویسید.

٤_۷ خلاصه

در این فصل پس از بیان نحوهٔ محاسبهٔ پارامترهای مرکزی مثل میانگین و مد و برخی از چندکها مثل میانه، توضیح داده شد که این پارامترها به علت دستکاری کردن داده‌های جامعه با تقریبی از واقعیت محاسبه خواهند شد. واضح است که برای محاسبه چندکها و

مد لزوماً از حدود واقعی طبقات استفاده خواهد شد.

در قسمت محاسبهٔ پارامترهای پراکندگی در داده‌های طبقه‌بندی شده، نحوهٔ محاسبهٔ واریانس، انحراف متوسط از میانگین، نیمه واریانس و انحراف چارکی بیان گردید و در ادامه درباره تصحیح شپارد و اهمیت قضیهٔ چی‌بی‌شف توضیح داده شد. گفته شد که در بسیاری از تحلیلها به ابزارهای بیشتری برای شناخت جامعهٔ آماری نیاز هست، بعلاوه ممکن است پارامترهای مرکزی و پراکندگی، امکانات لازم را برای متمایز ساختن توزیعهای فراوانی از همدیگر فراهم نیاورند که در این موارد از ضریب چولگی استفاده خواهد شد. این پارامترها نشان‌دهندهٔ اختلاف جامعهٔ آماری با توزیع نرمال به لحاظ تمایل داده‌هاست. استفاده از گشتاورها و چندکها برای محاسبهٔ ضریب چولگی به تصمیم‌گیرنده فرصت می‌دهد که بخوبی جوامع آماری را از لحاظ مکان تمرکز مشاهدات با همدیگر مقایسه نماید.

دسته‌ای دیگر از پارامترهای توصیفی، پارامترهای تعیین انحراف از کشیدگی هستند. این پارامترها میزان پراکندگی جامعه مورد نظر را با پراکندگی هر جامعهٔ دیگر، بخصوص با توزیع نرمال مقایسه می‌کنند. چنانچه پراکندگی توزیع بیشتر از توزیع دیگر باشد، اصطلاحاً گفته می‌شود، توزیع کوتاهتر است و بالعکس در حالتی که پراکندگی کمتر باشد، گفته می‌شود توزیع بلندتر است و در حالتی که ضریب کشیدگی صفر باشد، گفته می‌شود که پراکندگی دو جامعه یکسان است.

٤ـ۸ سؤالات و مسائل

سؤالات دوگزینه‌ای

۱. پارامترهای مرکزی داده‌های طبقه‌بندی شده همان مقادیر (پارامترهای) واقعی هستند. ص ☐ غ ☐

۲. همیشه مقدار واریانس با روش مستقیم مساوی با مقدار واریانس به روش غیرمستقیم است. ص ☐ غ ☐

۳. میانگین اختیاری لزوماً از ستون متوسط طبقات انتخاب می‌شود. ص ☐ غ ☐

٤. پراکندگی داده‌ها در دامنه $\mu_x \pm 2\sigma_x$ مساوی با ۷۵ درصد است. ص ☐ غ ☐

۵. توزیع نرمال، توزیعی است که میانگین، میانه و مد آن باهم مساوی باشد. ص ☐ غ ☐

٦. میانگین حسابی هم یک گشتاور است. ص ☐ غ ☐

۷. ضریب چولگی، نشان‌دهندهٔ نوع تمایل داده‌هاست. ص ☐ غ ☐
۸. ضریب چولگی منفی همیشه مطلوبتر از ضریب چولگی مثبت است. ص ☐ غ ☐
۹. محاسبهٔ میانه لزوماً با حدود واقعی طبقات انجام می‌گیرد. ص ☐ غ ☐
۱۰. پارامترهای تعیین انحراف از کشیدگی، مقدار پراکندگی توزیع آماری را با توزیع نرمال مقایسه می‌کنند. ص ☐ غ ☐

سؤالات چهارگزینه‌ای

۱۱. میانه این جدول کدام است؟

C-L	۲_٤	٤_٦	٦_٨	٨_١٠
F_i	۵	۱۰	۵	۱۰

الف) ٤ ج) ٦
ب) ۵ د) ۷

۱۲. چارک اول این داده‌ها کدام است؟

C-L	۱۱۰_۱۱۹	۱۲۰_۱۲۹	۱۳۰_۱۳۹
F_i	۱۰	۲۰	۷۰

الف) ۱۲۹ ج) ۱۲۷
ب) ۱۳۰ د) ۱۳۲/۷۰

۱۳. مد این جدول کدام است؟

x_i	−۲	−۱	۰	۱	۲	۳
F_i	۱۰	۲۰	۱۰	۱۵	۱۵	۳۰

الف) ۰ ج) ۲/۵
ب) ۳ د) ۲

۱٤. میانه جدول سؤال ۱۳ کدام است؟

الف) $\frac{1}{2}$ ج) ۱
ب) ۰ د) ۱/۵

۱۵. فرض کنید $Q_1 = ۱۰۰$ ، $Md = ۱٤۰$ و $Q_3 = ۱۸۰$ باشد؛ نیم دامنه کدام است؟

الف) ٤۰ ج) ٦۰

ب) ۳۰ د) ۸۰

۱٦. فرض کنید $\mu_x = ۳۷$ ، $Mo = ٤۳/٦$ و واریانس ۱٤٤ باشد؛ ضریب چولگی پیرسون کدام است؟

الف) ۰/۹۵ ج) ۰/۰٤۵

ب) ۰/۵۵− د) ۰/۰٤۵−

۱۷. فرض کنید میانه ۵۰، دهک اول ۱۰ وصدک نودم ۹۰ باشد؛ ضریب چولگی کدام است؟

الف) $\frac{۱}{۲}$ ج) ۰

ب) ۰/٤٤ د) ۲

۱۸. میانه این جدول کدام است؟

C-L	۱۰_۲۰	۲۰_۳۰	۳۰_٤۰	٤۰_۵۰
f_i	۰/۱۰	۰/۲۰	۰/۳۰	۰/٤۰

الف) ۳٤/۷ ج) ۳۹/۰

ب) ٤۰/۲ د) ۳٦/٦۷

۱۹. مد جدول سؤال ۱۸ کدام است؟

الف) ٤۰ ج) ٤۲

ب) ۳۹ د) ٤۵

۲۰. فرض کنید $Q_1 = ۲۲۳/۵۵$ ، $Md = ۲٤٦$ و $Q_3 = ۲۷۱/۵۵$ باشد؛ ضریب چولگی کدام است؟

الف) ۰/٦۰ ج) ۰/۰٦۵

ب) ۰/۵۵۱ د) ۰/۰۸۹

۲۱. کشیدگی گشتاوری و صدکی توزیع نرمال کدام است؟

الف) ۳ و ۰/۲٦۳ ج) ۰/۲٦۳ و ۰/۲٦۳

ب) ۳ و ۳ د) ۰/۲٦۳ و ۳

۲۲. در یک توزیع متمایل به چپ، کدام یک از این گزینه‌ها صحیح است؟

الف) $\mu_x < Md < Mo$ ج) $Mo < Md < \mu_x$

ب) $Md < \mu_x < Mo$ د) $\mu_x < Mo < Md$

۲۳. با توجه به این اطلاعات میانگین کل کدام است؟

N_i	۱۰۰	۵۰	۱۵۰
μ_i	۲۰	۵۰	۳۰

الف) ۱۰۰ ج) ۵۰
ب) ۳۰ د) ۳۳/۳۳

۲۴. واریانس کل داده‌های این جدول کدام است؟

N_i	۱۰۰	۲۰۰	۷۰۰
μ_i	۸۰	۹۰	۱۰۰
σ_i^2	۱۶۰۰	۲۵۰۰	۲۵۰۰

الف) ۲۲۰۰ ج) ۲۴۱۰
ب) ۲۴۵۴ د) ۲۲۳۳/۳

۲۵. میانگین و میانه یک جامعهٔ آماری به ترتیب ۳۰ و ۵۰ است و توزیع جامعه از چولگی معقولی برخوردار است؛ مد کدام است؟

الف) ۹۰ ج) ۲۵
ب) ۴۰ د) مد ندارد.

۲۶. اگر ضریب کشیدگی توزیعی ۰/۷۱− باشد، کدام عبارت درباره پراکندگی این جامعه صادق است؟

الف) پراکندگی جامعه نسبت به توزیع نرمال بیشتر و تفاوت آن فاحش است.
ب) پراکندگی جامعه نسبت به توزیع نرمال کمتر و تفاوت آن فاحش است.
ج) پراکندگی جامعه نسبت به توزیع نرمال بیشتر و تفاوت آن اندک است.
د) پراکندگی جامعه نسبت به توزیع نرمال کمتر و تفاوت آن اندک است.

۲۷. اگر $N = ۵۰$، $\sum x_i^2 = ۳۲۵۰۰$، $\mu_x = ۷$ و $\sum (x_i - \mu_x)^3 = ۹۶$ باشد؛ ضریب چولگی جامعه کدام است؟

الف) ۳ درصد ج) ۱/۹۶ درصد
ب) ۶ درصد د) ۲/۳ درصد

۲۸. اگر $N = ۱۰$، $\sum (x_i - \mu_x)^4 = ۷۶۸۰$ و انحراف معیار جامعه برابر ۴ باشد؛ ضریب کشیدگی کدام است؟

الف) ۱ ج) ۴
ب) ۳ د) ۰

۲۹. میانهٔ توزیع آماری ٤۰ مشاهده ۳۲/۵ است. اگر ۵ = I و فراوانی طبقهٔ میانه‌دار ۱۰ و مجموع فراوانیهای ماقبل طبقهٔ میانه‌دار ۱٤ باشد، حدود کرانهٔ طبقهٔ میانه‌دار کدام است؟

الف) ۳٤/۵ ـ ۲۹/۵ ج) ٤۰ ـ ۳۰
ب) ۳۹ ـ ۲۹ د) ۳۹ ـ ۳۵

۳۰. اگر ضریب چولگی توزیع آماری ۰/٦٦- باشد، کدام عبارت درباره جامعه مورد مطالعه صحیح است؟

الف) نرمال است.
ب) با جامعهٔ نرمال تفاوت فاحش دارد.
ج) با جامعهٔ نرمال تفاوت مختصری دارد.
د) با اطلاعات داده شده نمی‌توان قضاوت کرد.

مسائل

۳۱. مدیریت یک سازمان برنامه‌ای را برای مشارکت کارکنان در تصمیم‌گیریها فراهم کرده است. هدف مدیریت بهبود کیفیت زندگی کاری[1] با استفاده از این برنامه است. پس از اجرای این برنامه، مدیر عمومی تعداد شکایت رسیده از کارکنان را طی ۲٤ ماه ثبت کرده است. نتایج حاصل عبارتند از:

۳	۱٤	۱۰	۳۰	۳۵	۱۷
۷	٦	۸	۱٤	۲۸	۲۱
۱۸	۳۵	۱۱	۱۹	۵۷	٤۵
۱٤	۱۲	۲۳	۱۷	٦۱	۳۲

الف) توزیع فراوانی را تهیه کرده، نمودار بافت نگار را رسم کنید.
ب) آیا می‌توان گفت برنامه بهبود کیفیت زندگی کاری موفقیت‌آمیز بوده است؛ پارامترهای مناسب را محاسبه نمایید.
ج) درصدهای توزیع نرمال را در این مثال بررسی کنید.

1. quality of work life

۳۲. در جاده‌ای به طول ۱۰۰ کیلومتر، ۲۰۰ گاراژ وجود دارد. فاصلهٔ گاراژها برحسب کیلومتر به این شرح است:

فاصله از ابتدای جاده تا گاراژ	۷	۲۶	۲۸	۳۷	٤۰	٤٦	٦۰	۷۸	۸٦	۹۲
تعداد گاراژ	۱۰	۱۵	۵	۲۰	۵	۲۵	۱۵	۳۰	۱۰	٦۵

برای احداث پمپ بنزین در این جاده این طرحها وجود دارد:
طرح ۱. پمپ بنزین در وسط جاده (۵۰ کیلومتری) احداث شود.
طرح ۲. پمپ بنزین در محل میانگین فاصله گاراژها احداث شود.
طرح ۳. پمپ بنزین در محل میانه فاصلهٔ گاراژها احداث شود.
حاصل جمع قدر مطلق انحرافات را برای هریک از طرحها محاسبه کرده، بنویسید که ضمن مقایسه به چه نکتهٔ مفیدی می‌رسید.

۳۳. این جدول طبقه‌بندی را در نظر بگیرید:

C-L	$-10 \leq X < 0$	$0 \leq X < 10$	$10 \leq X < 20$	$20 \leq X < 30$	$30 \leq X < 40$	$40 \leq X < 50$	$50 \leq X < 60$
F_i	۳	۸	۱۲	۱٦	۹	٤	۲

الف) میانگین، میانه و مد را محاسبه کنید.
ب) واریانس را با استفاده از روش مستقیم و غیرمستقیم محاسبه کرده، مقایسه کنید.
ج) نمودارهای مناسب را برای توصیف داده‌ها به کار گیرید.

۳٤. این جدول طبقه‌بندی نشان‌دهنده مبالغ بیمه اخذ شده از تعدادی شرکت در خرداد ماه ۱۳۷٤ است:

فراوانی	مبالغ بیمه (به ۱۰ ریال)
۱	۱۰۰۰ـ۱۱۵۰
۳	۱۱۵۰ـ۱۳۰۰
٦	۱۳۰۰ـ۱٤۵۰
٤	۱٤۵۰ـ۱٦۰۰
۸	۱٦۰۰ـ۱۷۵۰
۹	۱۷۵۰ـ۱۹۰۰
۳	۱۹۰۰ـ۲۰۵۰
۲	۲۰۵۰ـ۲۲۰۰

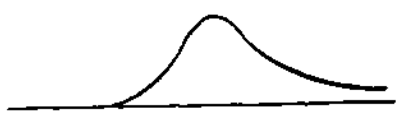
الف) جدول توزیع فراوانی نسبی بسازید.
ب) نمودار بافت نگار و نمودار تجمعی را ترسیم کنید.
ج) نتایج حاصل از پارامترهای تعیین انحراف از قرینگی را با تحلیل نمودار بافت نگار مقایسه کنید.
د) ضریب کشیدگی چندکی را محاسبه کرده، تحلیل نمایید.

۳۵. این جدول نشان‌دهنده جمعیت ۱۰۰ خانواده در روستای هفت چشمه است:

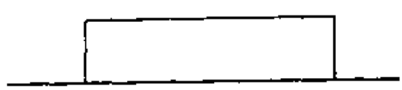
جمعیت خانواده‌ها	۱	۲	۳	٤	٥	٦	۷
تعداد خانواده‌ها	۱۰	۱۲	۱۳	۲۰	۲٥	۱٥	٥

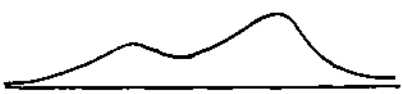
الف) نمودار مناسب را برای توصیف توزیع به کار ببرید.
ب) میانگین، مد و میانه را محاسبه کنید.
ج) انحراف متوسط از میانگین و واریانس را محاسبه کنید.
د) فرض کنید مأمور سرشماری اشتباهاً سرپرست خانواده‌ها را شمارش نکرده است؛ محاسبه کنید تأثیر این اشتباه بر میانگین و پراکندگی جمعیت چیست.

۳٦. کدام یک از شاخصهای مرکزی، پراکندگی، چولگی و کشیدگی را برای توصیف این توزیعها مناسب می‌دانید؟


الف


ب


ج


د

۳۷. مدت زمانی که بازدیدکنندگان یک غرفهٔ کتاب صرف کرده‌اند به این شرح طبقه‌بندی شده است:

C-L	فراوانی
< ۲	۳۰
۲_۳	٤۰
٤_٥	٤۰
٦_۷	۹۰
۸_۹	۷۰
۱۰_۱۱	٥۰
۱۲_۱۳	٥۰
۱٤_۱٥	۳۰

الف) نمودار مناسب برای توصیف داده‌ها را به کار بگیرید.
ب) کدام یک از پارامترهای مرکزی یا پراکندگی را برای توصیف مناسب می‌دانید؟ محاسبه کنید.
ج) کدام یک از پارامترهای چولگی را مناسب می‌دانید؟ محاسبه کنید.
د) کدام یک از پارامترهای کشیدگی را مناسب می‌دانید؟ محاسبه کنید.

۳۸. این داده‌ها نشان‌دهندهٔ سن افراد مراجعه کننده به یک درمانگاه ویژه در یک روز بخصوص است:

٤٠	٤٣	٦٦	٧٥	٨٥
٦٧	٥٦	٥٦	٨٠	٨٨
٧٥	٥٣	٦٥	٨٣	٨٩
٤٨	٤٤	٥٢	٨٣	٨٧

الف) یک توزیع فراوانی با طبقات ٤٠-٤٩ ، ٥٠-٥٩ و ... بسازید.
ب) میانگین و واریانس توزیع فراوانی را محاسبه نمایید.
ج) میانگین و واریانس داده‌های طبقه‌بندی نشده را محاسبه کنید.
د) نتایج بند ب و ج را با یکدیگر مقایسه و تحلیل نمایید.

۳۹. قیمت سهام ٦٠ شرکت در بازار بورس و اوراق بهادار تهران طی فروردین ماه ١٣٧٤ به این شرح به دست آمده است (داده‌ها برحسب ١٠ ریال است):

١٠٠٠	١٢٤٥	١٩٠٠	١٨٧٥	٢٠١٢	٢٢٠٠	١٩٧٥	٢١٠٠	١٢٠٠	٨٠٠
١٢٠٠	١٣٧٥	١٦٥٠	١٤٥٠	٢٠١٠	٢٠٨٠	٨٦٧	٩٥٠	١١٠٠	٧٥٠
٢١٠٠	١٤٨٥	١٣٧٠	١٣٢٠	١٩٧٥	٨٠٠	٩٥٠	٩٦٠	١٠٩٠	٩٠٠
٢٠٥٠	١٥٠٠	١٥٨٠	١٢١٠	١٦٠٠	٩٦٧	٩٥٥	٢٠٢٠	١١١٠	٩٥٠
١٨٠٠	١٦٥٠	١٧٨٠	١٢٢٥	١٦٥٠	٩١٠	٩٨٥	٢٠١٠	٩٥٠	٩٢٠
١٩٢٠	١٥٩٠	١٨٥٠	١٧٨٥	١٤٠٠	١٠١٢	٢١٠٠	٢٠١١	٨٧٠	١٠٠٠

الف) داده‌ها را طبقه‌بندی کنید.
ب) پارامترهای مرکزی را محاسبه کنید.
ج) پارامترهای پراکندگی را محاسبه نمایید.
د) ضریب چولگی گشتاوری را محاسبه و با ضرایب چولگی چندکی مقایسه کنید.
هـ) ضریب کشیدگی گشتاوری و چندکی را محاسبه و تحلیل نمایید.
و) صحت درصدهای توزیع نرمال را بررسی کنید.

٤٠. این داده‌ها مربوط به عمر لاستیکهای تولیده شده درکارخانهٔ ایران تایر است (داده‌ها بر حسب کیلومتر است):

۹۰۰	۸۰۰	۲۱۰۰	۱۷۴۰	۱۴۰۰	۹۱۰	۱۰۵۰	۹۷۰	۸۵۰	۱۰۰۰
۹۵۰	۸۷۰	۲۰۵۰	۱۷۸۰	۱۵۵۰	۱۰۰۰	۱۱۱۵	۲۱۰۰	۱۷۵۰	۱۱۰۰
۹۶۰	۸۱۰	۲۱۲۰	۱۵۷۰	۱۴۵۰	۱۰۱۰	۱۲۰۰	۲۲۰۰	۱۹۰۰	۱۰۵۰
۹۴۰	۸۲۰	۱۷۵۰	۱۶۰۰	۱۴۸۰	۱۷۵۰	۱۷۰۰	۸۷۵	۱۹۲۰	۲۰۰۰
۹۱۰	۸۸۰	۱۸۰۰	۱۸۰۰	۱۶۰۰	۱۸۹۰	۱۸۰۰	۹۷۸	۱۴۰۰	۲۱۰۰
۹۹۰	۸۹۰	۲۰۷۰	۱۴۷۰	۱۶۶۰	۱۹۰۰	۱۸۵۰	۱۰۰۰	۹۰۰	۲۱۵۰

الف) نمودار بافت‌نگار را تهیه نمایید.

ب) یک نمودار شاخه و برگ تهیه کنید به طوری که شاخه‌ها از صفر شروع شود.

ج) پارامترهای مرکزی را با استفاده از نمودارهای بند الف و ب محاسبه و مقایسه کنید.

د) آیا درصدهای قضیهٔ چی‌بی‌شف در این مثال مصداق دارد؟ چرا؟

۴۱. این جدول طبقه‌بندی میزان پرداخت ماهیانه به مدیران را در بخش عمومی و خصوصی نشان می‌دهد:

فراوانی بخش خصوصی	فراوانی بخش عمومی	درآمد (به ۱۰ هزار ریال)
۱۰	۳۰	۵۰-۶۰
۸	۳۴	۶۰-۷۰
۱۲	۳۷	۷۰-۸۰
۱۵	۵۰	۸۰-۹۰
۳۰	۴۰	۹۰-۱۰۰
۴۵	۴۵	۱۰۰-۱۱۰
۵۰	۳۰	۱۱۰-۱۲۰
۶۰	۲۰	۱۲۰-۱۳۰
۴۰	۱۰	۱۳۰-۱۴۰
۳۰	۴	۱۴۰-۱۵۰

الف) نمودار چند ضلعی دو جامعه را ترسیم و با یکدیگر مقایسه نمایید؛ بنویسید چه نتایجی حاصل می‌شود.

ب) ضریب چولگی گشتاوری دو جامعه را با یکدیگر مقایسه و تحلیل نمایید.

ج) ضریب کشیدگی دو جامعه را محاسبه و تحلیل نمایید.

پاسخنامهٔ سؤالات

۱) غ	۲) ص	۳) غ	۴) غ
۵) ص	۶) ص	۷) ص	۸) غ
۹) ص	۱۰) ص	۱۱) ج	۱۲) ج
۱۳) ب	۱۴) ج	۱۵) د	۱۶) ب
۱۷) ج	۱۸) د	۱۹) ج	۲۰) ج
۲۱) الف	۲۲) ج	۲۳) ب	۲۴) ب
۲۵) الف	۲۶) الف	۲۷) الف	۲۸) د
۲۹) الف	۳۰) ب		

فصل پنجم

مبادی احتمال

۵ـ۱ نظریهٔ احتمال

واژه «احتمال» دالّ بر «عدم اطمینان» نسبت به آینده است. ما در بسیاری از موارد از پیش‌بینی دقیق آینده ناتوان هستیم و برای اندازه گیری این عدم اطمینان از نظریهٔ احتمال استفاده می‌کنیم؛ مثلاً شما نمی‌دانید که وضع هوای فردا دقیقاً چگونه خواهد بود، نمی‌توانید به طور قطع نتیجهٔ بازی فوتبال بین تیم ملی جمهوری اسلامی ایران و یک کشور دیگر را تعیین کنید یا مدیری نمی‌داند امکان آنکه تقاضای محصولاتش بین ۶۰۰ تا ۷۰۰ واحد باشد چقدر است؛ در اینجا از نظریهٔ احتمال بدین صورت استفاده می‌شود که: احتمال اینکه فردا هوا آفتابی باشد ۴۰ درصد است، احتمال اینکه تیم ملی جمهوری اسلامی ایران برنده شود ۶۵ درصد است یا احتمال اینکه تقاضای محصولات بین ۶۰۰ تا ۷۰۰ واحد باشد ۲۸ درصد است.

ما غالباً در مورد حوادثی که در اطرافمان می‌گذرد اطلاعاتی، هرچند به صورت پراکنده، داریم یا می‌توانیم اطلاعاتی کسب کنیم. ما با سازماندهی کردن این اطلاعات به صورت منظم قادر خواهیم بود در تصمیم‌گیریهای خود، به جای اینکه از روشهای سرانگشتی استفاده کنیم و یا تیری در تاریکی بیندازیم، از نظریهٔ احتمال برای رسیدن به نتیجهٔ بهتر استفاده کنیم. مدیران نیز می‌توانند در پیش‌بینی و تصمیم‌گیریهای خود از نظریهٔ احتمال بهرهٔ زیادی ببرند.

افرادی که در ایجاد و توسعهٔ نظریهٔ احتمال نقش بسزایی داشتند عبارتند از: برنولی، دوموآور، بیز، لاگرانژ و لاپلاس.

تمرین

۱. نظریهٔ احتمال چه کاربردهایی برای مدیران دارد؟

۲. شرکتهای بیمه به دانستن نظریهٔ احتمال نیازمند هستند. ضمن ارائه چند سؤال در این زمینه، توضیح دهید که این شرکتها چگونه از نظریهٔ احتمال استفاده می‌کنند؟
۳. «استفاده از این کالا ممکن است برای سلامتی شما مضر باشد. این کالا حاوی ساکارین است که معلوم شده است موجب سرطان در حیوانات آزمایشگاهی می‌شود.» نظریهٔ احتمال چه نقشی در این عبارت ایفا می‌کند؟
٤. چهار نفر از کسانی را که در توسعهٔ نظریهٔ احتمال نقش عمده‌ای داشته‌اند، نام ببرید.

۵ـ۲ برخی از مفاهیم اساسی احتمال

قبل از هرچیز، مفاهیم اساسی‌ای را که در نظریهٔ احتمال به کار می‌رود، توضیح می‌دهیم و برای درک بهتر، برای هرکدام مثالی می‌آوریم.

۵ـ۲ـ۱ مفهوم احتمال

به طور کلی می‌توان احتمال را شانس وقوع پیشامد خاصی تعریف کرد؛ به تعبیری دیگر، احتمال وقوع یک پیشامد برابر نسبت دفعاتی است که پیشامد خاصی در تکرارهای زیاد رخ خواهد داد (این مفهوم به فراوانی نسبی که در فصل سوم مطرح شد نزدیک است)؛ مثلاً در پرتاب سکه، در تکرارهای زیاد در نیمی از موارد شیر ظاهر می‌شود؛ بنابراین احتمال ظاهر شدن شیر در پرتاب یک سکه ۵۰ درصد است. در پرتاب یک تاس، احتمال اینکه عدد ٤ ظاهر شود $\frac{1}{6}$ است. از این به بعد احتمال را با حرف P نشان می‌دهیم.

۵ـ۲ـ۲ احتمال عینی و ذهنی

برای درک مفهوم احتمال عینی و ذهنی به این دو مثال توجه کنید. در مثال اول فرض کنید می‌خواهید از کیسه‌ای که دارای ۵ مهره و ۳ عدد آنها سفید است، مهره‌ای را به تصادف بیرون بیاورید و احتمال بیرون آمدن یک مهره سفید برایتان مهم باشد. در مثال دوم فرض کنید یکی از مسافران پروازهای اصفهان ـ تهران را برحسب تصادف انتخاب کرده، می‌پرسید احتمال اینکه هواپیمای اصفهان ـ تهران در زمان مقرر به مقصد برسد چقدر است و او به شما می‌گوید که احتمال آن ۷۸ درصد است. در مثال اول، شما از قبل احتمال بیرون آمدن یک مهره سفید را می‌دانید (احتمال آن $\frac{3}{5} = 0/6$ است). این احتمال همواره ثابت است؛ ولی در مثال دوم، پاسخ این فرد صرفاً بیانگر

نظر این شخص است. این فرد پاسخ خود را با توجه به روند گذشته و پروازهای قبلی خود بیان می‌کند. ممکن است اگر همین سؤال را از چند نفر دیگر بپرسید پاسخهای مختلفی بشنوید که هریک بیانگر نظر شخصی و ذهنیت آنها از پروازهای اصفهان ـ تهران است.

احتمال در مثال اول عینی و در مثال دوم ذهنی است؛ بنابراین احتمال عینی، به نظر اشخاص مختلف وابسته نیست و احتمالِ وقوع از قبل مشخص است، ولی احتمال ذهنی به عقاید اشخاصی وابسته است که آن را ارزیابی می‌کنند. در واقع، احتمال ذهنی را می‌توان احتمال تخصیص داده شده به وسیلهٔ یک فرد به یک پیشامد تعریف کرد.

۵ـ۲ـ۳ آزمایش

در نظریهٔ احتمال فعالیتی که نتیجه آن از قبل مشخص نباشد به «آزمایش» معروف است؛ مثلاً شما می‌خواهید ببینید که با پرتاب سکه‌ای که در دست دارید شیر ظاهر می‌شود یا خط. انجام این کار یک آزمایش است.

۵ـ۲ـ۴ فضای نمونه

مجموعه پیامدهای ممکن یک آزمایش را فضای نمونه آن آزمایش گویند. فضای نمونه را با حرف S نشان می‌دهیم.

مثال ۵ـ۱ می‌خواهیم فضای نمونهٔ پرتاب یک سکه را مشخص کنیم. اگر ظاهر شدن شیر را با H و خط را با T نشان دهیم، در این صورت:

$$S = \{H , T\}$$

مثال ۵ـ۲ فضای نمونهٔ پرتاب یک تاس عبارت است از:

$$S = \{۱ , ۲ , ۳ , ۴ , ۵ , ۶\}$$

مثال ۵ـ۳ فضای نمونهٔ پرتاب دو سکه عبارت است از:

$$S = \{HH , HT , TH , TT\}$$

مثال ٥ـ٤ فضای نمونهٔ پرتاب دو تاس عبارت است از:

$$S = \{(۱,۱),(۱,۲),(۱,۳),(۱,٤),(۱,٥),(۱,٦),(۲,۱),\ldots,(٦,۱),(٦,۲),(٦,۳),(٦,٤),(٦,٥),(٦,٦)\}$$

این فضای نمونه دارای ۳٦ عضو است.

٥ـ۲ـ٥ فضای نمونهٔ محدود و نامحدود

در مثالهای قبل، فضاهای نمونهٔ تعداد محدودی عضو داشتند، ولی فضای نمونهٔ برخی آزمایشها نامحدود است؛ به عبارت دیگر تعداد اعضای آنها نامتناهی است.

مثال ٥ـ٥ فرض کنید شرکتی پیچ و مهره‌های صنعتی تولید می‌کند. مأمور کنترل کیفیت این شرکت می‌خواهد آنقدر پیچ آزمایش کند تا به اولین پیچ معیوب برسد. پیشامد موردنظر تعداد پیچهای انتخاب شده تا اولین پیچ معیوب است. فضای نمونه عبارت است از:

$$S = \{۱, ۲, ۳, ٤, ٥, \ldots\}$$

فضای نمونه در این مثال نامتناهی (نامحدود) است.

٥ـ۲ـ٦ فضای نمونهٔ گسسته و پیوسته

اگر فضای نمونه شامل تعداد متناهی یا تعداد نامتناهی ولی شمارش‌پذیر باشد، آن را «فضای نمونهٔ گسسته» گویند؛ مثلاً فضای نمونهٔ پرتاب یک سکه، یک تاس، یا تعداد پیچهای آزمایش شده تا اولین پیچ معیوب دارای فضای نمونهٔ گسسته هستند؛ چون تعداد عناصر فضای نمونهٔ پرتاب یک سکه یا یک تاس متناهی است و فضای نمونهٔ تعداد پیچهای انتخاب شده تا اولین پیچ معیوب نامتناهی ولی شمارش‌پذیر است.

فضای نمونهٔ برخی از آزمایشها که گسسته نباشد و در طول یک پاره خط تعریف شود اصطلاحاً آن را «فضای نمونهٔ پیوسته» گویند؛ مانند مدت زمانی که کارگری برای تراش یک یاطاقان صرف می‌کند. اگر بپذیریم زمان، متغیری است که می‌توان آن را بدقت اندازه گیری کرد، تعداد نامتناهی از زمانهای ممکن وجود دارد که نمی‌توان آنها را یک به یک مشخص کرد.

مثال ٦ـ٥ فضای نمونه برای نوعی لامپ که حداکثر عمر آن ۱۷۸۰ ساعت است، عبارت است از:

$$S = \{0 \leq x \leq 1780\}$$

در این عبارت x نشان‌دهنده عمر لامپ است.

مثال ۷ـ٥ فضای نمونهٔ مثال ٦ـ٥ را مجدداً در نظر می‌گیریم. می‌خواهیم بدانیم این فضا پیوسته است یا گسسته. چرا.

فضای نمونه پیوسته است؛ زیرا تعداد نامتناهی زمان ممکن برای عمر لامپ وجود دارد که نمی‌توان تک‌تک آنها را مشخص کرد.

٥ـ۲ـ۷ پیشامد

در نظریهٔ احتمال، پیشامد، یکی از زیرمجموعه‌های فضای نمونه است. در پرتاب یک سکه، خط آمدن یک پیشامد است و شیر آمدن پیشامد دیگری است. در پرتاب یک تاس، ظاهر شدن عدد ٤ یک پیشامد است و ظاهر شدن عدد ۲ یا ٦ نیز پیشامدهای دیگری است. پیشامدها را با حروف بزرگ انگلیسی مانند A ، B ، C ، ... نشان می‌دهیم.

مثال ۸ـ٥ پرتاب یک سکه را در نظر بگیرید. اگر A را پیشامد ظاهر شدن شیر (H) تعریف کنیم اعضای پیشامد A عبارتند از:

$$A = \{H\}$$

مثال ۹ـ٥ اگر B را پیشامد ظاهر شدن عدد زوج در پرتاب یک تاس تعریف کنیم، اعضای پیشامد B عبارتند از:

$$B = \{2, 4, 6\}$$

٥ـ۲ـ۸ پیامدهای مقدماتی هم شانس

اگر در آزمایش نوعی تقارن وجود داشته باشد که مطمئن باشیم وقوع یک پیامد همان قدر امکان دارد که وقوع هر پیامد مقدماتی دیگر، می‌گوییم فضای نمونه دارای

پیشامدهای اولیه یا «پیامدهای مقدماتی» هم‌شانس است؛ مثلاً در پرتاب یک تاس ۶ پیامد مقدماتی مختلف (۶ ، ۵ ، ٤ ، ۳ ، ۲ ، ۱) وجود دارد که امکان وقوع هریک با دیگری برابر است؛ یعنی امکان وقوع هرکدام $\frac{1}{6}$ است.

مثال ۱۰ـ۵ فضای نمونهٔ پرتاب دو سکه چنین است: $S = \{HH, HT, TH, TT\}$ می‌خواهیم بدانیم آیا هریک از پیامدهای مقدماتی فوق هم شانس هستند یا نه.

چون شانس وقوع هریک $\frac{1}{4}$ است، پس شانس وقوع هریک با دیگری برابر است.

تمرین

۱. فضای نمونهٔ پرتاب همزمان سه سکه را مشخص کنید.

۲. فضای نمونهٔ پرتاب همزمان یک سکه و یک تاس را مشخص کنید.

۳. در تمرین ۲:

الف) آیا پیامدهای مقدماتی هم شانس هستند.

ب) آیا فضای نمونه نامحدود است؟

٤. در پرتاب یک تاس، این پیشامدها را تعیین کنید:

الف) پیشامد آنکه عدد بزرگتر از ۳ ظاهر شود (A).

ب) پیشامد آنکه عدد ظاهر شده بر ۳ قابل قسمت باشد (B).

ج) پیشامد آنکه عدد ظاهر شده فرد یا عدد اول باشد (C).

د) پیشامد آنکه عدد ظاهر شده مضربی از ۷ باشد (D).

ه) پیشامد آنکه عدد ظاهر شده هم زوج و هم کوچکتر از ۵ باشد (E)

۵. سکه‌ای را سه بار پرتاب می‌کنیم و تعداد شیرهای ظاهر شده را ملاحظه می‌کنیم.

الف) فضای نمونه این آزمایش را بنویسید؛

ب) آیا هریک از پیامدهای مقدماتی فوق دارای شانس مساوی هستند؟

۶. به طور تصادفی از یکی از چندین متصدی ماشین فرز در یک کارخانه سؤال می‌شود که «احتمال اینکه ماشین فرز امروز خراب شود چقدر است؟» و او براساس تجربیات قبلی خود جواب می‌دهد: «به نظر من ۳۰ درصد». این احتمال عینی است یا ذهنی؟ چرا؟

۳ـ۵ احتمال یک پیشامد

پیش از این احتمال را شانس وقوع پیشامد خاصی تعریف کردیم. فرض کنید فضای نمونه‌ای داریم که در آن همهٔ پیامدهای مقدماتی، شانس مساوی برای انتخاب شدن، دارند.

در این صورت احتمال وقوع پیشامد خاصی مانند A عبارت است از تعداد عضوهای پیشامد A به تعداد عضوهای فضای نمونه. اگر تعداد عضوهای فضای نمونه را با n (S) و تعداد عضوهای پیشامد A را با n (A) نشان دهیم، احتمال A یعنی P (A) به این صورت خواهد بود:

$$P(A) = \frac{n(A)}{n(S)}$$

مثال ۱۱ـ۵ می‌خواهیم در پرتاب دو سکه این احتمالات را محاسبه کنیم:

الف) دقیقاً دو خط ظاهر شود.

ب) حداقل یک خط ظاهر شود.

ج) هر دو یک چیز را نشان دهند.

ابتدا فضای نمونهٔ پرتاب دو سکه را می‌نویسیم:

$$S = \{HH, HT, TH, TT\}$$

الف) پیشامد ظاهر شدن دقیقاً دو خط (A) به این صورت است:

$$A = \{TT\}$$

پس:

$$P(A) = \frac{n(A)}{n(S)} = \frac{۱}{۴}$$

ب) پیشامد ظاهر شدن حداقل یک خط (B) به این صورت است:

$$B = \{HT, TH, TT\}$$

پس:

$$P(B) = \frac{n(B)}{n(S)} = \frac{۳}{۴}$$

ج) پیشامد آنکه هر دو یک چیز را نشان دهند (C) به این صورت است:

$$C = \{HH, TT\}$$

پس:

$$P(C) = \frac{n(C)}{n(S)} = \frac{۲}{۴}$$

مثال ۱۲ـ۵ می‌خواهیم احتمال هریک از این پیشامدها را تعیین کنیم:

الف) در پرتاب تاسی عدد زوج ظاهر شود.

ب) در بیرون آوردن مهره‌ای از ظرفی که محتوی ۵ مهره قرمز، ٤ مهره سفید و ۳ مهره سیاه است، رنگ مهره سفید باشد.

ج) در انتخاب لامپی از بین ۸ لامپ که یک عدد آن سوخته است، لامپ انتخابی سالم باشد.

الف) در پرتاب تاس شش حالت هم شانس وجود دارد که سه حالت آن زوج است. اگر پیشامد مورد نظر را با حرف A نشان دهیم؛ آنگاه:

$$P(A) = \frac{۳}{٦} = \frac{۱}{۲}$$

ب) تعداد ۱۲ = ۳ + ٤ + ۵ مهره در ظرف وجود دارد که ٤ عدد آن سفید است. اگر پیشامد مورد نظر را با حرف B نشان دهیم؛ آنگاه:

$$P(B) = \frac{٤}{۱۲} = \frac{۱}{۳}$$

ج) تعداد کل لامپها ۸ عدد و تعداد لامپهای سالم ۷ عدد است. اگر پیشامد مورد نظر را با حرف C نشان دهیم؛ آنگاه:

$$P(C) = \frac{۷}{۸}$$

۱-۳-۵ احتمال و فراوانی نسبی

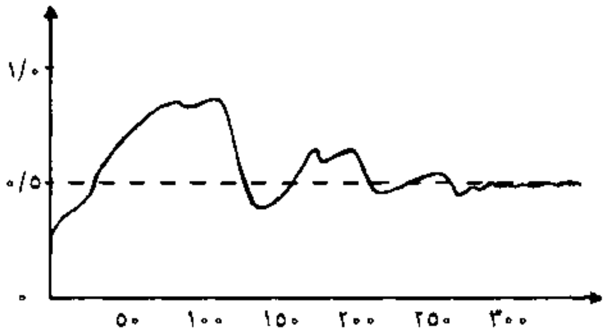
در بسیاری از آزمایشها، پیامدهای مقدماتی دارای شانس مساوی برای انتخاب شدن نیستند. در این صورت تعریف احتمال به صورت تعداد عناصر پیشامد مورد نظر به تعداد عناصر فضای نمونه نامناسب است؛ مثلاً وقتی که تاس ناسالمی را پرتاب می‌کنیم، وجوه مختلف را نمی‌توان هم شانس دانست. یا مثلاً تعداد ضایعات تولیدی در ابتدا و انتهای نوبت کاری با هم مساوی نیستند، همچنین تعداد تصادفات در ساعات مختلف روز یکسان نیست. در چنین مواردی اگر بخواهیم احتمال وقوع پیشامدی را تعیین کنیم، باید فراوانی وقوع پیشامد را در صورتی که آزمایش تحت شرایط یکسان مکرراً انجام شده باشد، در نظر بگیریم که در این صورت از فراوانی نسبی کمک گرفته‌ایم؛ بنابراین فراوانی نسبی پیشامد A در N بار تکرار آزمایش چنین تعریف می‌شود:

$$\text{فراوانی نسبی پیشامد A} = \frac{\text{تعداد دفعاتی که A در N تکرار آزمایش روی می‌دهد}}{N}$$

در صورتی می‌توان از فراوانی نسبی به عنوان مبنای احتمال استفاده کرد که تعداد تکرارهای آزمایش (N) به سمت بی‌نهایت میل کند که به زبان ریاضی چنین تعریف می‌شود:

$$P(A) = \lim_{N \to \infty} (\text{فراوانی نسبی پیشامد A در N تکرار})$$

مثلاً اگر احتمال شیر آمدن در پرتاب یک سکهٔ سالم مورد نظر باشد نمی‌توان فرضاً با سه بار پرتاب در مورد احتمال شیر آمدن قضاوت کرد؛ زیرا ممکن است در سه بار پرتاب دو بار شیر ظاهر شود و از این رو احتمال $\frac{2}{3}$ غیرمعقول است. با توجه به شکل ۱ـ۵، بعد از تقریباً ۳۰۰ پرتاب سکه به حالت پایدار، که درآن نوسانها تقریباً به صفر می‌رسد، می‌رسیم.



شکل ۱ـ۵ فراوانی نسبی تعداد شیرهای ظاهر شده در N پرتاب

شانس وقوع ضایعات یک کالا را در ساعت خاص نیز باید با نمونه‌های بسیار تعیین کرد.

مثال ۱۳ـ۵ در نمونه‌ای وسیع که قبلاً از جمعیت ایران گرفته شده است، تعداد فرزندان هر خانواده همراه با نسبت افرادی که دارای این تعداد فرزند هستند، نشان داده شده است. اطلاعات مربوط به این سرشماری به این شرح است:

تعداد فرزندان	۰	۱	۲	۳	٤	۵ و بیشتر
نسبت خانواده‌هایی که این تعداد فرزند دارند	۰/۰۵	۰/۱۰	۰/۳۵	۰/۲۵	۰/۱۵	۰/۱۰

اگر خانواده‌ای را به طور تصادفی انتخاب کنیم، می‌خواهیم احتمال هریک از این پیشامدها را تعیین کنیم.

الف) کمتر از دو فرزند داشته باشند.

ب) بین دو تا چهار فرزند داشته باشند.

ج) چهار فرزند یا بیشتر داشته باشند.

الف) احتمال داشتن کمتر از دو فرزند:

$$P(A) = ۰/۰۵ + ۰/۱۰ = ۰/۱۵$$

ب) احتمال داشتن دو تا چهار فرزند:

$$P(B) = ۰/۳۵ + ۰/۲۵ + ۰/۱۵ = ۰/۷۵$$

ج) احتمال داشتن چهار فرزند یا بیشتر:

$$P(C) = ۰/۱۵ + ۰/۱۰ = ۰/۲۵$$

۵ـ۳ـ۲ خواص مقدماتی احتمال

در اینجا خواص مقدماتی احتمال را مطرح می‌کنیم. خواص مقدماتی بدیهی بوده، نیاز به برهان ندارد. این خواص که برای هر پیشامدی، چه عضوهای فضای نمونه هم شانس باشند و چه نباشند، صادق است عبارتند از:

۱. احتمال پیشامدی همچون A متعلق به فضای نمونه S، همواره بزرگتر یا مساوی صفر و کوچکتر یا مساوی یک است؛ یعنی $۰ \le P(A) \le ۱$.

۲. احتمال وقوع فضای نمونه (S) برابر یک است؛ یعنی $P(S) = ۱$.

مثال ۵ـ۱۴ فضای نمونه ۴ عنصری $S = \{e_۱, e_۲, e_۳, e_۴\}$ را در نظر بگیرید. می‌خواهیم بدانیم کدام یک از این توابع فضای احتمال S را تعریف نمی‌کند:

الف) $P(e_۱) = \frac{۱}{۲}$ ، $P(e_۲) = \frac{۱}{۳}$ ، $P(e_۳) = \frac{۱}{۴}$ ، $P(e_۴) = \frac{۱}{۵}$.

ب) $P(e_۱) = \frac{۱}{۲}$ ، $P(e_۲) = -\frac{۱}{۴}$ ، $P(e_۳) = \frac{۱}{۴}$ ، $P(e_۴) = \frac{۱}{۲}$.

ج) $P(e_1) = \frac{1}{2}$ ، $P(e_2) = \frac{1}{4}$ ، $P(e_3) = \frac{1}{8}$ ، $P(e_4) = \frac{1}{8}$.

د) $P(e_1) = \frac{1}{2}$ ، $P(e_2) = \frac{1}{4}$ ، $P(e_3) = \frac{1}{4}$ ، $P(e_4) = 0$.

الف) چون مجموع احتمالات فضای نمونه ($\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}=\frac{77}{60}$) بزرگتر از یک است، این تابع فضای نمونه S را تعریف نمی‌کند.

ب) چون $P(e_3) = -\frac{1}{4}$ و عددی منفی است؛ بنابراین تابع، احتمال مربوط به S را تعریف نمی‌کند.

ج) از آنجا که تمام مقادیر غیرمنفی و مجموع آنها ($\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{8}=1$) برابر یک است، می‌تواند معرف فضای احتمال S باشد.

د) از آنجا که تمام مقادیر غیرمنفی و مجموع آنها ($\frac{1}{2}+\frac{1}{4}+\frac{1}{4}+0=1$) برابر یک است، می‌تواند فضای احتمال S را تعریف کند.

تمرین

۱. فضای نمونهٔ $S = \{e_1, e_2, e_3, e_4\}$ را در نظر بگیرید و فرض کنید P تابع احتمال روی S باشد.

الف) اگر $P(e_2) = \frac{1}{3}$ ، $P(e_3) = \frac{1}{6}$ و $P(e_4) = \frac{1}{9}$ باشد، $P(e_1)$ را پیدا کنید.

ب) اگر $P(e_1) = 2\,P(e_2)$ و $P(e_3) = P(e_4) = \frac{1}{4}$ باشد $P(e_1)$ و $P(e_2)$ را پیدا کنید.

ج) اگر $P(\{e_2, e_3\}) = \frac{2}{3}$ ، $P(\{e_2, e_4\}) = \frac{1}{2}$ و $P(e_2) = \frac{1}{3}$ باشد، $P(e_1)$ را پیدا کنید.

۲. تاسی ناسالم بوده به نحوی که احتمال آمدن هر شماره‌ای با عدد آن متناسب است (مثلاً احتمال آمدن عدد ٦ دوبرابر عدد ۳ است). فرض‌کنید که $A = \{$عدد زوج$\}$ ، $B = \{$عدد اول$\}$ ، $C = \{$عدد فرد$\}$ است.

الف) احتمال آمدن هر عدد را مشخص کنید.

ب) $P(A)$ ، $P(B)$ و $P(C)$ را پیدا کنید.

٤ـ٥ قواعد شمارش

گاهی اوقات با وضعیتهایی مواجه می‌شویم که یا باید تمام حالتهای ممکنِ آزمایشی

را فهرست کنیم یا دست کم مشخص کنیم که چند حالت ممکن و مختلف وجود دارد؛ مثلاً سازمانی که برای ماشینها پلاک صادر می‌کند باید بداند اگر از ۲ حرف و ٤ رقم استفاده کند، چند شماره می‌تواند صادر کند تا اگر این تعداد کم بود شماره‌ها را ۵ رقمی کند یا یک دانشگاه بزرگ باید مشخص کند که شمارهٔ دانشجویی باید چندرقمی باشد تا در عین حال که تعداد ارقام آن کم است، بتوان به تمام دانشجویان شماره داد.

۵ـ٤ـ۱ اصل اساسی شمارش

اساسی‌ترین اصل در شمارش «قاعدهٔ ضرب» است که بدین صورت تعریف می‌شود: اگر عملی مستلزم K مرحله باشد که مرحله اول به n_1 طریق، مرحله دوم به n_2 طریق، ... و مرحله k ام به n_K طریق انجام پذیرد؛ آنگاه عمل مزبور به $n_1 \times n_2 \times \ldots \times n_K$ طریق ممکن انجام می‌شود.

مثال ۵ـ۱۵ فرض کنید شخصی می‌تواند تعطیلات نوروزی خود را با اتوبوس یا قطار یا هواپیما به یکی از ٤ شهر اصفهان ، تبریز، شیراز یا مشهد برود. می‌خواهیم بدانیم این شخص به چند طریق می‌تواند این مسافرت را انجام دهد.

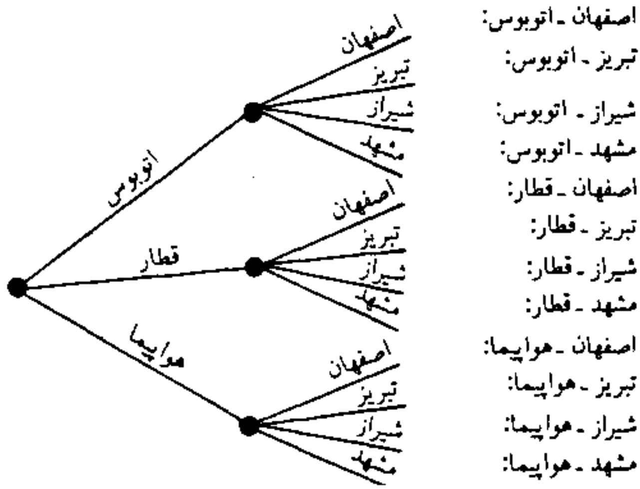
وسیلهٔ سفر را به $n_1 = ۳$ طریق و شهر مورد نظر را به $n_2 = ٤$ طریق می‌توان انتخاب کرد؛ بنابراین تعداد طرقی که می‌تواند مسافرت انجام شود از ضرب این دو به دست می‌آید:

$$n_1 \times n_2 = ۳ \times ٤ = ۱۲$$

در همین مثال، اگر بخواهیم تمام حالات ممکن را فهرست کنیم از «نمودار درختی» استفاده می‌کنیم. نمودار درختی روش منظمی برای نشان دادن این حالات است.

شکل ۵ـ۲ نشان می‌دهد که ۳ شاخه برای تعداد وسایل مسافرت و ٤ شاخه برای تعداد شهرها وجود دارد که بدین ترتیب ۱۲ گزینه برای انتخاب وسیله و شهر وجود خواهد داشت.

مثال ۵ـ۱۶ می‌خواهیم بدانیم اگر قرار باشد پلاک اتومبیلها را با استفاده از نام یک شهر، یک حرف فارسی و ۵ رقم مشخص کنیم با این شرط که فقط استفاده از نام ۲۵ شهر مجاز باشد، چند ماشین مختلف را می‌توانیم شماره گذاری کنیم (رقم اول شماره ماشین نباید صفر باشد).



شکل ۵.۲ نمودار درختی برای مسافرت با ۳ وسیله به ٤ شهر

چون در فارسی ۳۲ حرف وجود دارد و تنها می‌توان از نام ۲۵ شهر استفاده کرد و شماره ماشینها ۵ رقمی است، تعداد ماشینهایی که می‌توان شماره گذاری کرد عبارتند از:

$$۲۵ \times ۳۲ \times ۹ \times ۱۰ \times ۱۰ \times ۱۰ \times ۱۰ = ۷۲۰۰۰۰۰$$

۲ـ٤ـ۵ جایگشت (ترتیب)

گاه وضعیتهایی مورد توجه ماست که برای آنها، پیامدها، ترتیبها یا آرایشهای مختلفی وجود دارد که فقط برای گروهی از اشیا امکانپذیر است؛ مثلاً اگر بخواهیم ببینیم چند آرایش مختلف برای سخنرانی یک رئیس، یک معاون و یک مدیر کل متصور است از جایگشت استفاده می‌کنیم.

مثال ۱۷ـ۵ می‌خواهیم بدانیم برای سه حرف a، b، c چند جایگشت (بدون تکرار حروف) وجود دارد.

جایگشتهای آن عبارتند از: abc، acb، bac، bca، cab، cba. می‌توان بدون برشمردن یکایک جایگشتها، تعداد آنها را تعیین کرد برای مکان اول سه انتخاب، برای مکان دوم دو انتخاب و برای مکان سوم یک انتخاب ممکن وجود دارد؛ بنابراین تعداد کل جایگشتها (آرایشها) $۶ = ۱ \times ۲ \times ۳$ است. n شیء متمایز را می‌توان به این طریق مرتب کرد:

$$n(n-۱) \ldots \times ۳ \times ۲ \times ۱$$

این حاصلضرب را با نماد "$n!$" نمایش می‌دهیم و آن را n فاکتوریل می‌خوانیم. بنابرآنچه گفتیم: $1! = 1$، $2! = 2 \times 1 = 2$، $3! = 3 \times 2 \times 1 = 6$، $4! = 4 \times 3 \times 2 \times 1 = 24$ و ... است. بنا به تعریف $0! = 1$ است.

مثال ۵ـ۱۸ می‌خواهیم بدانیم به چند طریق می‌توان ۵ کارمند نمونه را در حضور تمام اعضای شرکت معرفی کرد.

$$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$$

مثال ۵ـ۱۹ تعداد جایگشتهای متشکل از ٤ حرف a ، b ، c و d برابر ۲٤ است. می‌خواهیم بدانیم اگر قرار باشد فقط از ۲ حرف در هر آرایش استفاده کنیم، تعداد جایگشتهای آن چند تا است.

برای اولین حرف ٤ انتخاب و برای دومین حرف ۳ انتخاب وجود دارد؛ بنابراین $4 \times 3 = 12$ طرق ممکن وجود خواهد داشت که عبارتند از:

ab , ac , ad , ba , bc , bd , ca , cb , cd , da , db , dc

با تعمیم این استدلال درمی‌یابیم که اگر تعیین تعداد جایگشتهای r شیء از n شیء متمایز مورد نظر باشد، می‌توان تعداد آنها را به کمک این دو فرمول به دست آورد:

$$P_r^n = n(n-1)\dots(n-r+1)$$

و

$$P_r^n = \frac{n!}{(n-r)!}$$

برای به دست آوردن فرمول دوم می‌توان $n!$ را به این صورت نشان داد:

$$n(n-1)\dots(n-r+1)(n-r)!$$

مثال ۵ـ۲۰ می‌خواهیم از بین ۱۵ عضو شرکت‌کننده در یک جلسه، یک رئیس، یک معاون و یک سخنگو انتخاب کنیم. ترتیب انتخاب نیز مهم است. این عمل به طرق مختلف امکانپذیر است که به این صورت محاسبه می‌شود:

$$P_3^{15} = \frac{n!}{(n-r)!} = \frac{15!}{(15-3)!} = \frac{15!}{12!} = 2730$$

اگر جایگشتهای اشیا روی دایره‌ای مرتب شده باشند «جایگشتهای دوری» نامیده می‌شوند. درجایگشت دوری اگر تغییر در جهت خلاف عقربه‌های ساعت وبه گونه‌ای باشد که اشیای قبل و بعد هر شیء یکسان باشد، در واقع دو آرایش مختلف نخواهیم داشت؛ مثلاً اگر ٤ نفر دور میز گردی نشسته باشند و همگی محل خود را یک صندلی درجهت عقربه‌های ساعت تغییر دهند،جایگشت جدیدی به وجود نخواهد آمد.

مثال ۵-۲۱ جایگشتهای دوری از ٤ نفر که دور یک میز نشسته‌اند، چنین محاسبه می‌شود:

اگر یک نفر را در مکان ثابتی در نظر بگیریم و ۳ نفر دیگر را به $3!$ طریق مرتب کنیم، متوجه می‌شویم که شش آرایش مختلف (جایگشت دوری) از این ٤ نفر به وجود خواهد آمد.

تعداد جایگشتهای n شیء متمایز که دور یک دایره مرتب شده‌اند، برابراست با $(n-1)!$.

تا به حال از بین n شیء متمایز r شیء را انتخاب می‌کردیم تا جایگشتهای آن را مشخص کنیم؛ اما اگر n شیء از هم متمایز نباشند؛ یعنی چند تا از آنها یکسان باشند، تعداد جایگشتها تغییر می‌یابد.

مثال ۵-۲۲ جایگشتهایی که با حروف کلمهٔ «حسابدار» می‌توان نوشت چنین محاسبه می‌شود:

موقتاً فرض می‌کنیم که دو حرف «ا» از هم متمایز هستند و آنها را با $ا_1$ و $ا_2$ نشان می‌دهیم. برای نمونه دو جایگشت ح س $ا_1$ ب د $ا_2$ ر و ح س $ا_2$ ب د $ا_1$ ر را در نظر بگیرید. این دو جایگشت با هم یکسان است؛ بنابراین هر دو جایگشت با اندیس فقط یک جایگشت بدون اندیس حساب می‌شود؛ بنابراین تعداد کل جایگشتهای ممکن از حروف کلمهٔ «حسابدار» $\frac{7!}{2!} = 2520$ خواهد بود.

مثال ۵-۲۳ جایگشتهایی که با حروف کلمهٔ «حسابداران» می‌توان نوشت چنین محاسبه می‌شود:

اگر فرض کنیم سه حرف «ا» مختلف هستند و آنها را با $ا_1$ ، $ا_2$ ، $ا_3$ نشان دهیم، تعداد $9!$ جایگشت مختلف از کلمهٔ «حسابداران» ایجاد می‌شود؛ اما چون تعداد جایگشتهای سه حرف $ا_1$ ، $ا_2$ ، $ا_3$ که به یک آرایش از کلمهٔ «حسابداران» منجر می‌شود برابر $3!$است، فقط $\frac{9!}{3!} = 60480$ آرایش مختلف از حروف این کلمه وجود خواهد داشت.

با تعمیم این استدلال نتیجه می‌گیریم که:

تعداد جایگشتهای n شیء که n_1 تای آنها از نوع اول، n_2 تای آنها از نوع دوم، ... و n_k تای آنها از نوع Kام و $n = n_1 + n_2 + ... + n_k$ است، برابر است با $\frac{n!}{n_1! n_2! ... n_k!}$.

مثال ۵ـ۲۴ برای چراغانی کردن سر در یک شرکت تولیدی یک لامپ قرمز، ۳ لامپ سبز، ٤ لامپ آبی و ۲ لامپ زرد داریم. می‌خواهیم بدانیم به چند طریق مختلف می‌توان آنها را در یک ردیف قرار داد.

تعداد کل لامپها برابر است با ۱۰ = ۲ + ٤ + ۳ + ۱؛ بنابراین:

$$\frac{10!}{1!\,3!\,4!\,2!} = 12600$$

۵ـ٤ـ۳ ترکیب

در ترکیب، تعداد راههای انتخاب r شیء از بین n شیء مهم ولی ترتیب (جایگشت) آنها بی‌اهمیت است؛ مثلاً با ٤ حرف a، b، c و d می‌توان ۲٤ جایگشت ۳ حرفی نوشت، ولی از میان آنها هر یک از گروههای «٦ جایگشتی» که حروف به کار رفته در آنها مشابه است، مانند abc، acb، bac، bca، cab و cba یک ترکیب محسوب می‌شوند.

ترکیبات	جایگشتها
abc	abc , acb , bac , bca , cab , cba
abd	abd , adb , bad , bda , dab , dba
acd	acd , adc , cad , cda , dac , dca
bcd	bcd , bdc , cbd , cdb , dbc , dcb

بنابراین تعداد ترکیبهای ۳ حرفی از ٤ حرف ٤ خواهد بود.

مثال ۵ـ۲۵ قرار است از بین ۸ نفر که داوطلب استخدام در سازمانی هستند ۳ نفر انتخاب شوند. طرق مختلفی که می‌توان این ۳ نفر را انتخاب کرد، چنین محاسبه می‌شود:

اگر به ترتیب انتخاب افراد توجه کنیم جواب $P_3^8 = 336$ خواهد بود و اگر ترتیب انتخاب افراد را در نظر نگیریم تعداد $3! = 6$ از هر مجموعه حذف خواهد شد که در این صورت به $\frac{336}{6} = 56$ طریق می‌توان ۳ نفر را از بین ۸ نفر انتخاب کرد.

با تعمیم این استدلال نتیجه می‌گیریم:

تعداد ترکیبهای r شیء از n شیء متمایز برابر است با:

$$\binom{n}{r}=\frac{n(n-1)\dots(n-r+1)}{r!}$$

و یا

$$\binom{n}{r}=\frac{n!}{(n-r)!\,r!}$$

می‌توان نشان داد که همواره روابط زیر برقرار است:

$$1)\ \binom{n}{0}=1 \qquad 2)\ \binom{n}{n}=1$$

$$3)\ \binom{n}{n-1}=n \qquad 4)\ \binom{n}{1}=n$$

مثال ۵ـ۲۶ قرار است از بین ۱۰ مشتری عمدهٔ یک شرکت، ۳ مشتری را انتخاب کنیم و نظر آنها را درباره کیفیت محصولات تولیدی جویا شویم. طرق مختلف انتخاب چنین محاسبه می‌شود:

چون ترتیب انتخاب مهم نیست، با استفاده از فرمول ترکیب، خواهیم داشت:

$$\binom{10}{3}=\frac{10!}{(10-3)!3!}=\frac{10!}{7!\,3!}=120$$

مثال ۵ـ۲۷ می‌خواهیم از بین ۷ سرکارگر مونتاژ و ٤ سرکارگر بسته‌بندی کمیته‌ای با ٤ سرکارگر مونتاژ و ۲ سرکارگر بسته‌بندی تشکیل دهیم. طرق مختلف انتخاب اعضا چنین محاسبه می‌شود:

براساس قاعدهٔ ضرب، اگر بتوان عمل اول را به n_1 طریق و عمل دوم را به n_2 طریق انجام داد؛ آنگاه کل کار را می‌توان به $n_1 \times n_2$ طریق انجام داد؛ بنابراین به $\binom{7}{4}=35$ طریق می‌توان ٤ سرکارگر مونتاژ را از بین ۷ نفر برگزید و به $\binom{4}{2}=6$ طریق می‌توان ۲ سرکارگر بسته‌بندی را از بین ٤ نفر انتخاب کرد و تعداد طرق انتخاب به طور کلی برابر است با:

$$\binom{7}{4}\binom{4}{2}=35\times 6=210$$

٥ ـ ٤ ـ ٤ افرازهای مرتب

گاهی مایلیم ترکیب r شیء از n شیء متمایز را به گونه‌ای خاص افراز (تفکیک) کنیم. در

واقع در این گونهٔ خاص هرترکیب در حکم یک مجموعه خواهد بود که هر مفروز آن یک زیرمجموعه است که در آن زیر مجموعه، ترتیب قرار گرفتن اشیا مهم نیست.

مثال ۵ـ۲۸ می‌خواهیم بدانیم به چند طریق مختلف می توان مجموعه‌ای از ٤ شیء را به سه زیر مجموعه که به ترتیب ۲ ، ۱ و ۱ عضو داشته باشد افراز کرد.

چهار شیء را با a، b، c و d نشان می‌دهیم. با شمارش درمی‌یابیم که ۱۲ حالت ممکن بدین شرح وجود دارد:

ab | c | d	ab | d | c	ac | b | d	ac | d | b
ad | b | c	ad | c | b	bc | a | d	bc | d | a
bd | a | c	bd | c | a	cd | a | b	cd | b | a

تعداد افرازها را برای این مثال با نماد $\binom{٤}{۲,۱,۱} = ۱۲$ نشان می‌دهیم که در آن عدد ٤ معرف تعداد کل اشیا و اعداد ۲ ، ۱ و ۱ معرف تعداد اشیایی است که به ترتیب در زیر مجموعه‌ها قرار می‌گیرند.

با تعمیم این استدلال نتیجه می‌گیریم که:

تعداد طرقی که می توان مجموعهٔ n شیئی را به K زیرمجموعه با $n_۱$ شیء در مجموعهٔ اول، $n_۲$ شیء در مجموعهٔ دوم، ... و n_k شیء را در مجموعهٔ K ام افراز کرده برابر است با:

$$\binom{n}{n_۱, n_۲, ..., n_k} = \frac{n!}{n_۱!\, n_۲! \dots n_k!}$$

که در آن $n_۱ + n_۲ + ... + n_k = n$ است.

مثال ۵ـ۲۹ می‌خواهیم بدانیم به چند طریق می توان ۸ کارمند را در دو اتاق ۳ نفره و یک اتاق ۲ نفره جای داد.

چون ۸ = ۲ + ۳ + ۳ است داریم:

$$\binom{۸}{۳, ۳, ۲} = \frac{۸!}{۳!\, ۳!\, ۲!} = ٥٦٠$$

تمرین

۱. اگر دایرهٔ بازاریابی و فروش شرکتی بخواهد یکی از ٥ متن تهیه شده را با یکی از ٤ وسیلهٔ تبلیغاتی (رادیو، تلویزیون، مجله، روزنامه) آگهی کند، این کار به چند طریق میسراست؟

۲. اگر شرکت بیمه‌ای بخواهد برای بیمه‌شدگان کدی انتخاب کند که از ۲ حرف فارسی و یک شماره ۶ رقمی تشکیل شده باشد، به چند بیمه شده می‌تواند کد بدهد؟
۳. در تمرین ۲ ، فرض کنید که شمارهٔ کد نباید از ۲۵۰ هزار کوچکتر باشد؛ با این قید به چند نفر می‌توان کد داد؟
۴. به چند طریق می‌توان ۲ مهره را از ظرفی که حاوی ۶ مهره است، بیرون آورد؟
الف) با جایگذاری
ب) بدون جایگذاری
۵. به چند طریق می‌توان ۶ کتاب مختلف را در قفسه‌ای در کنار هم قرار داد؟
۶. تعداد جایگشتهای ۳ حرفی از ۵ حرف a ، b ، c ، d و e را پیدا کرده، بنویسید.
۷. به چند طریق مختلف می‌توان ۵ شمع به رنگهای مختلف را در اطراف یک سینی گرد قرار داد؟
۸. با حروف کلمهٔ «دینار» چند کلمهٔ ۵ حرفی مختلف می‌توان نوشت؟
۹. با حروف اسم «ابوعلی سینا» چند کلمهٔ ۱۰ حرفی می‌توان نوشت؟
۱۰. به چند طریق می‌توان از بین ۶ مارک تجاری مختلف ۲ عدد آنها را انتخاب کرد؟
۱۱. به چند طریق می‌توان از بین ۶ مارک تجاری مختلف دست کم ۲ عدد آنها را انتخاب کرد؟
۱۲. از بین ۹ کالای موجود در یک کارتن ۳ عدد آنها معیوب است؛ به چند طریق می‌توان ۴ کالا انتخاب کرد به طوری که ۲ تای آنها سالم و ۲ تای آنها معیوب باشد؟
۱۳. به چند طریق می‌توان ۹ نفر را در یک اتاق ۴ نفره، دو اتاق ۲ نفره و یک اتاق یکنفره جای داد؟

۵ـ۵ عملیات روی پیشامدها و قواعد احتمال

در قسمت ۳ـ۵ دو قانون بدیهی و مهم در مورد احتمالات را مطرح کردیم که عبارت بودند از:

۱. احتمال وقوع پیشامدی همانند A همواره بزرگتر یا مساوی صفر و کوچکتر یا مساوی یک است ($0 \le P(A) \le 1$).

۲. احتمال فضای نمونه همواره برابر یک است ($P(S) = 1$).

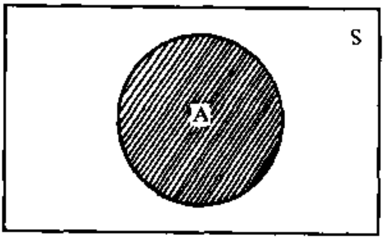
قبل از بیان قواعد دیگر، به تشریح مفاهیمی چون «نمودار ون»[1] و پیشامدهای ناسازگار و سازگار می‌پردازیم.

۱ـ۵ـ۵ نمودار ون، پیشامدهای ناسازگار و سازگار

برای نشان دادن یک یا چند پیشامد در فضای نمونه می‌توان از نمودار معروفی به نام

1. Venn diagram

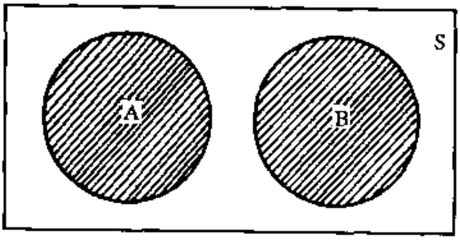
نمودار ون استفاده کرد. جان ون، ریاضیدان انگلیسی اواخر قرن نوزدهم و اوایل قرن بیستم، مبتکر این نمودار است. در این نمودار، کل فضای نمونه با مستطیلی نشان داده می‌شود و هر پیشامد قسمتی از این مستطیل را به خود اختصاص می‌دهد. نمودار شکل ۳ـ۵ را در نظر بگیرید:



شکل ۳ـ۵ نمونه‌ای از نمودار ون

چنانکه گفته شد، احتمال فضای نمونه برابر یک است (P (S) = ۱) ؛ بنابراین احتمال پیشامد A یا برابر با سطحی است که پیشامد A از فضای نمونه اشغال کرده یا برابر با نسبت تعداد اعضای A به S است. حال دو پیشامد ناسازگار را توضیح می‌دهیم.

دو پیشامد را در صورتی «ناسازگار» گویند که در یک لحظه، فقط و فقط یکی از آنها بتواند واقع شود؛ به عبارت دیگر امکان وقوع همزمان دو پیشامد ناسازگار وجود ندارد. اگر یکی از آنها واقع شود دیگری قطعاً واقع نمی‌شود ولی امکان دارد که هیچ کدام از آنها به وقوع نپیوندد. نمودار ون برای دو پیشامد ناسازگار در شکل ٤ـ۵ رسم شده است.



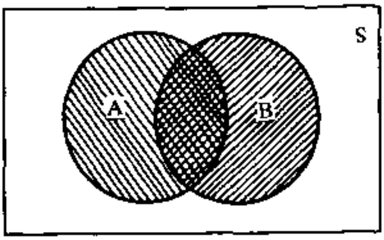
شکل ٤ـ۵ نمودار ون برای دو پیشامد ناسازگار

چنانکه ملاحظه می‌کنید دو پیشامد ناسازگارِ A و B هیچ وجه اشتراکی با هم ندارند.

مثال ۳۰ـ۵ می‌خواهیم بدانیم در پرتاب یک تاس، دو پیشامد ظاهر شدن عدد کمتراز ۳ وظاهرشدن عدد ٦ ناسازگارند یا نه.

پیشامد ظاهر شدن عدد کمتر از ۳ را با A و پیشامد ظاهر شدن عدد ٦ را با B نشان می‌دهیم؛ بنابراین {۱ , ۲} = A و {٦} = B خواهد بود. روشن است که اگر عدد یک یا ۲ ظاهر شود حتماً عدد ٦ ظاهر نمی‌شود و به عکس؛ پس این دو پیشامد ناسازگارند.

دو پیشامد را در صورتی «سازگار» گویند که وقوع یک پیشامد مستلزم عدم وقوع دیگری نباشد؛ به عبارت دیگر دو پیشامد سازگار دارای دست کم یک عضو مشترکند. نمودار ون برای دو پیشامد سازگار A و B در شکل ۵ـ۵ نمایش داده شده است.



شکل ۵.۵ نمودار ون برای دو پیشامد سازگار

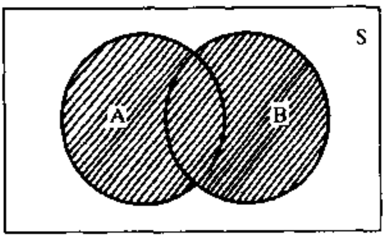
چنانکه ملاحظه می‌کنید دو پیشامد سازگارِ A و B دارای وجوه مشترکی هستند.

مثال ۵.۳۱ در پرتاب یک تاس A پیشامد ظاهرشدن عدد بزرگتر از ٤ و B پیشامد ظاهرشدن عددی زوج است. می‌خواهیم بدانیم آیا این دو پیشامد سازگارند یا نه.

چون پیشامد {٦ , ٥} = A و پیشامد {٦ , ٤ , ٢} = B دارای یک عضو مشترک (٦) هستند، این دو پیشامد سازگارند؛ یعنی اگر عدد ٦ ظاهر شود هر دو پیشامد واقع شده‌اند.

۵.۵.۲ اجتماع، اشتراک و متمم پیشامدها

دو پیشامد A و B را در نظر بگیرید. «اجتماع» دو پیشامد A و B مجموعهٔ تمام عضوهایی است که در A، یا در B، یا هم در A و هم در B قرار دارند. اجتماع دو پیشامد A و B را با AUB نشان می‌دهیم. وقوع AUB بدین معنی است که دست کم یکی ازدو پیشامد A یا B رخ داده است. نمودار ون برای اجتماع دو پیشامد A و B در شکل ۵.٦ نشان داده شده است.



شکل ۵.٦ نمودار ون برای اجتماع دو پیشامد

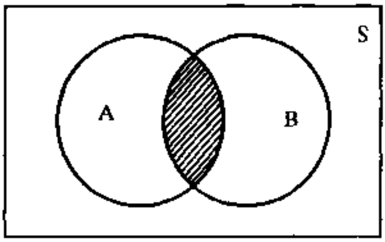
مثال ۵.۳۲ در پرتاب یک تاس A پیشامد ظاهرشدن عدد فرد و B پیشامد ظاهرشدن عددی است که بر ۳ بخش‌پذیر باشد؛ اجتماع این دو پیشامد چنین خواهد بود:

{۵ , ۳ , ۱} = A و {٦ , ۳} = .B است؛

بنابراین {٦ , ۵ , ۳ , ۱} = AUB.

دو پیشامد A و B را در نظر بگیرید. «اشتراک» دو پیشامد A و B را با $A \cap B$ نشان می‌دهیم. وقوع $A \cap B$ بدین معنی است که هر دو پیشامد A و B رخ داده است. نمودار ون برای اشتراک دو پیشامد A و B در شکل ۵.۷ نشان داده می‌شود.

مثال ۵ـ۳۳ با توجه به اطلاعات مثال ۵ـ۳۲، اشتراک دو پیشامد A و B چنین خواهد بود:
A = {۱ , ۳ , ۵} و B = {۳ , ۶} است؛ بنابراین A∩B = {۳}.

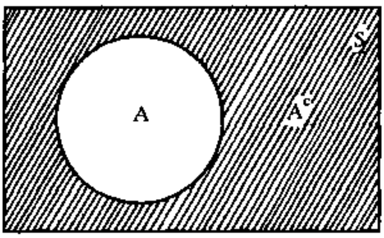


شکل ۵.۷ نمودار ون برای اشتراک دو پیشامد

پیشامد A را در نظر بگیرید. «متمم» پیشامد A مجموعهٔ تمام عضوهایی است که در A نیستند. متمم پیشامد A را با A^c نشان می‌دهیم. وقوع A^c بدین معنی است که A رخ نداده است. نمودار ون برای متمم A در شکل ۵ـ۸ نشان داده شده است.

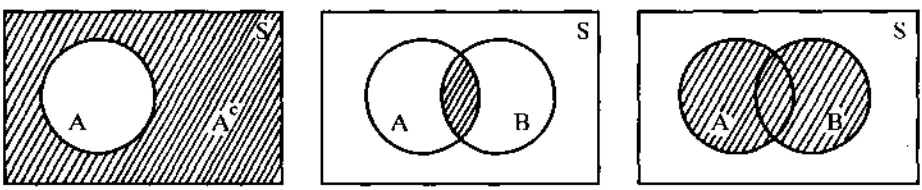
مثال ۵ـ۳۴ در پرتاب یک تاس، A پیشامد ظاهرشدن عدد کوچکتر از ۳ است؛ متمم پیشامد A چنین خواهد بود:

متمم پیشامد A شامل عضوهایی از فضای نمونه است که در A نباشد؛ بنابراین A^c = {۳ , ۴ , ۵ , ۶}.



شکل ۵ـ۸ نمودار ون برای متمم پیشامد A

حال برای مقایسهٔ بهتر، نمودارهای ون برای اجتماع، اشتراک و متمم در شکل ۵ـ۹ نمایش داده می‌شوند.



شکل ۵ـ۹ نمودارهای ون برای اجتماع دو پیشامد (نمودار سمت راست)، اشتراک دو پیشامد (نمودار وسط)، و متمم یک پیشامد (نمودار سمت چپ)

نباید پنداشت که اجتماع و اشتراک تنها برای دو پیشامد مطرح می‌شوند. می‌توانیم آنها را برای بیش از دو پیشامد نیز به کار ببریم؛ مثلاً A∪B∪C مرکب از عضوهایی از فضای نمونه است که دست کم در یکی از سه پیشامد A، B و C هستند و یا A∩B∩C مرکب از عضوهایی از فضای نمونه است که هم در A، هم در B و هم در C هستند.

مثال ۵ـ۳۵ با داشتن فضای نمونه $S = \{a, b, c, d, e, f, g, h\}$ و پیشامدهای $A = \{a, e, f, g\}$، $B = \{d, e, g\}$، $C = \{a, b\}$ و $D = \{f, g, h\}$ می‌خواهیم این موارد را پیدا کنیم:

الف) $A \cup B$ د) $(A \cap B)^c$

ب) $A \cap C$ هـ) $(A \cup B)^c \cup C$

ج) A^c و) $D^c \cap (A \cup B)$

الف) $A \cup B = \{a, d, e, f, g\}$

ب) $A \cap C = \{a\}$

ج) $A^c = \{b, c, d, h\}$

د) چون $A \cap B = \{e, g\}$ است، پس $(A \cap B)^c = \{a, b, c, d, f, h\}$.

هـ) $A \cup B$ را در بند الف به دست آوردیم، پس $(A \cup B)^c = \{b, c, h\}$ است؛ بنابراین: $(A \cup B)^c \cup C = \{a, b, c, h\}$.

و) چون $D^c = \{a, b, c, d, e\}$ است، پس $D^c \cap (A \cup B) = \{a, d, e\}$.

گاهی محاسبهٔ احتمال یک پیشامد مستلزم دانستن احتمال یک یا چند پیشامد دیگر است. روابطی وجود دارد که می‌توان با آنها احتمال پیشامدهای وابسته را پیدا کرد.

۵ـ۵ـ۳ برخی از قواعد احتمالات

فرض کنید A و B دو پیشامد مربوط به فضای یک آزمایش بوده، A زیرمجموعهٔ B نیز باشد (A زیر مجموعهٔ B را به صورت $A \subset B$ نشان می‌دهیم)، چون مجموعهٔ A شامل تعداد عضوهای کمتر و یا مساوی مجموعهٔ B است، در این صورت احتمال وقوع پیشامد A همواره کوچکتر یا مساوی احتمال پیشامد B خواهد بود؛ به عبارت دیگر خواهیم داشت:

اگر A و B دو پیشامد مربوط به فضای یک آزمایش باشند به طوری که $A \subset B$ باشد، در این صورت:

$$P(A) \le P(B)$$

در تعریف متمم A گفتیم که A^c آن قسمت از فضای نمونه است که در A نیست؛ پس $A \cup A^c = S$ و چون $P(S) = 1$؛ بنابراین $P(A) + P(A^c) = 1$. این قاعده به قاعدهٔ متمم‌گیری معروف است که می‌توان آن را به این صورت نوشت:

$$P(A^c) = 1 - P(A)$$

دو پیشامد A و B را در نظر بگیرید. اجتماع این دو پیشامد شامل عضوهایی است که یا در A، یا در B و یا در هردوی آنهاست. اگر بخواهیم عضوهای $A \cup B$ را مشخص کنیم، عضوهای مشترک را تنها یک بار محسوب می‌کنیم. از آنجا که $A \cap B$ عضوهای مشترک بین A و B است، برای محاسبهٔ $P(A \cup B)$ باید $P(A \cap B)$ را از $P(A) + P(B)$ کم کنیم. به عبارت دیگر باید از «قاعده جمع» استفاده کرد؛ یعنی:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

در دو پیشامد ناسازگار، $A \cap B$ مجموعه‌ای تهی است و بنابراین $P(A \cap B) = ۰$ است؛ در نتیجه قاعدهٔ جمع برای دو پیشامد ناسازگار به این صورت خلاصه می‌شود:

$$P(A \cup B) = P(A) + P(B)$$

مثال ۵ـ۳۶ احتمال اینکه خانواده‌ای اتومبیل، موتور سیکلت و یا هر دوی آنها را داشته باشد به ترتیب ۰/۶۱ ، ۰/۲۵ و ۰/۰۸ است. اگر خانواده‌ای به صورت تصادفی انتخاب شود می‌خواهیم احتمال این موارد را پیدا کنیم:

الف) اتومبیل نداشته باشد.

ب) دست کم یکی از این دو را داشته باشد.

اگر A را پیشامد داشتن اتومبیل و B را پیشامد داشتن موتور سیکلت در نظر بگیریم؛ آنگاه $P(A) = ۰/۶۱$ ، $P(B) = ۰/۲۵$ و $P(A \cap B) = ۰/۰۸$ است؛ بنابراین خواهیم داشت:

الف) $P(A^c) = ۱ - P(A) = ۱ - ۰/۶۱ = ۰/۳۹$

ب) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = ۰/۶۱ + ۰/۲۵ - ۰/۰۸ = ۰/۷۸$

مثال ۵ـ۳۷ استادی در هر جلسه فقط از یکی از دانشجویان کلاس می‌خواهد که تمرینهای درس قبل را حل کند. احتمال اینکه از سعید بخواهد تمرینها را حل کند ۰/۰۵ و احتمال اینکه از رضا بخواهد ۰/۰۷ است. احتمال اینکه از یکی از این دو خواسته شود که تمرینها را حل کنند، چقدر است.

اگر پیشامد صدا زدن سعید یا رضا را به ترتیب با S و R نشان دهیم، $P(S) = ۰/۰۵$ و $P(R) = ۰/۰۷$ خواهد بود. چون امکان صدا زدن هر دو وجود ندارد؛ این دو پیشامد ناسازگارند؛ پس $P(S \cap R) = ۰$ و بنابراین $P(S \cup R) = ۰/۰۵ + ۰/۰۷ = ۰/۱۲$ است.

تمرین

۱. در هریک از این عبارات اشتباهی وجود دارد. دلیل این اشتباه را توضیح دهید.

الف) احتمال آنکه کارمندی غیبت کند ۰/۰۴ و احتمال آنکه در محل کار خود حاضر شود ۰/۹۰ است.

ب) احتمال آنکه حساب مشکوک‌الوصولی سوخت شود ۰/۲۰ ، احتمال آنکه پرداخت شود ۰/۶۵ و احتمال آنکه هم سوخت و هم پرداخت شود ۰/۰۸ است.

ج) احتمال آنکه از بین ۸ تیر شلیک شده ۰ ، ۱ ، ۲ ، ۳، ٤ ، ۵ تیر یا بیشتر به هدف بخورد به ترتیب برابر است با $\frac{۱}{٤}$ ، $\frac{۳}{۱۶}$ ، $\frac{۵}{۳۲}$ ، $\frac{۷}{٦٤}$ ، $\frac{۲۳}{۱۲۸}$.

۲. دو پیشامد A و B را با $P(A \cup B) = \frac{۳}{٤}$ ، $P(A^c) = \frac{۲}{۳}$ و $P(A \cap B) = \frac{۱}{٤}$ درنظر بگیرید و $P(A)$ ، $P(B)$ و $P(A \cap B^c)$ را پیدا کنید.

۳. با فرض اینکه $P(A) = ۰/٤۵$ ، $P(B) = ۰/۳۲$ و $P(A \cap B^c) = ۰/۱۳$ باشد، احتمالات زیر را پیدا کنید:

الف) $P(A \cup B)$ ب) $P[(A \cup B)^c]$

ج) $P(A^c \cap B)$ د) $P(A^c \cup B)$

هـ) $P(A^c \cup B^c)$ و) $P(A^c \cap B^c)$

٤. چهار نفر داوطلب پُست معاونت اداری سازمانی هستند. اگر احتمال انتخاب شدن فرد A دو برابر احتمال انتخاب شدن فرد B باشد و B و C تقریباً شانس برابری برای انتخاب شدن داشته باشند و احتمال انتخاب شدن فرد C دو برابر فرد D باشد، احتمالات زیر را پید اکنید:

الف) احتمال اینکه فرد C انتخاب شود.

ب) احتمال اینکه فرد A انتخاب نشود.

ج) احتمال اینکه فرد A یا B یا C انتخاب شوند.

۵. دانشگاهی در یکی از استانها واقع شده است. $\frac{۱}{۳}$ دانشجویان آن دانشگاه خارج از خوابگاه دانشجویی زندگی می‌کنند. $\frac{۵}{۹}$ دانشجویان اهل آن استانند و $\frac{۳}{٤}$ دانشجویان یا اهل آن استان نیستند یا در خوابگاه دانشجویی زندگی می‌کنند. احتمال اینکه دانشجویی که به صورت تصادفی از این دانشگاه انتخاب می‌شود اهل آن استان نباشد و در خوابگاه زندگی کند چقدر است؟

٦. بازرسی قصد دارد قابلیت اطمینان دو پمپ بنزین را مقایسه کند. هر پمپ بنزینی در معرض دو نوع نقص قرار دارد؛ نقص فنی پمپ و نشت بنزین. هنگامی که یکی از این دو نقص (یا هر دو)

روی دهد، پمپ بنزین تعطیل می‌شود. داده‌های زیر وضعیت دو پمپ بنزین را به این صورت نشان می‌دهد:

پمپ بنزین	احتمال نقص فنی	احتمال نشت بنزین	احتمال هر دو
۱	۰/۰۷	۰/۱۰	۰
۲	۰/۰۹	۰/۱۲	۰/۰۶

کدام یک از این دو پمپ بنزین دارای قابلیت اطمینان کمتری هستند؟

۵ـ۶ احتمال شرطی

اگر بدون توصیف فضای نمونه صحبت از احتمال شود، ممکن است مشکلاتی پیش بیاید؛ مثلاً اگر از مدیری سؤال شود احتمال اینکه کارمندی را به تصادف انتخاب کنیم و حقوقش بالای ۴۰۰ هزار ریال باشد چقدر است، وی ممکن است چندین جواب مختلف بدهد که همهٔ آنها درست باشد. یکی از جوابها ممکن است شامل حال کارمندانی باشد که تحصیلاتشان در حد دیپلم یا زیر دیپلم است. جواب دیگر ممکن است شامل حال کارمندانی باشد که مدرک کاردانی یا کارشناسی دارند. سومین جواب ممکن است شامل حال افرادی باشد که کارشناس ارشد یا بالاتر هستند. برای پرهیز از چنین مشکلاتی و برای گرفتن پاسخ مورد نظر، باید فضای نمونه را مشخص و محدود کنیم.

می‌توان احتمال شرطی را به این صورت تعریف کرد:

اگر پیشامدی همانند A به پیشامد دیگری همانند B مربوط باشد و بدانیم پیشامد B به وقوع پیوسته است، در این صورت احتمال وقوع A ، به احتمال وقوعِ A به شرط B ((P (A / B)) تغییر می‌یابد که آن را «احتمال شرطی» گوییم.

نماد «/» در (P (A / B)) نشان دهندهٔ «کسر» نیست بلکه نشان‌دهندهٔ احتمال شرطی است.

مثال ۵ـ۳۸ یک سازمان تحقیقاتیِ حمایت از مصرف‌کننده درباره ۵۰ تعمیرگاه تلویزیون بررسیهایی انجام داده است. اطلاعات به دست آمده در این جدول خلاصه شده است.

	تعمیرگاه مجاز	تعمیرگاه غیرمجاز
با سابقه ۱۰ سال یا بیشتر	۱۶	۴
با سابقه کمتر از ۱۰ سال	۱۰	۲۰

می‌خواهیم هر یک از این احتمالات را در صورتی که فردی برحسب تصادف یکی‌از این تعمیرگاهها را انتخاب کند، محاسبه کنیم:

الف) احتمال اینکه این تعمیرگاه مجاز باشد.

ب) احتمال اینکه تعمیرگاهی که ۱۰ سال یا بیشتر سابقه دارد، مجاز باشد.

منظور ما از «انتخاب برحسب تصادف» این است که در هر حالت، تمام انتخابهای ممکن هم‌شانس باشند.

الف) اگر L انتخاب مرکز تعمیر مجاز را نشان دهد، n (L) تعداد عناصر L را و n (S) تعداد عناصر فضای نمونه را نشان می‌دهد؛ در این صورت خواهیم داشت:

$$P(L) = \frac{n(L)}{n(S)} = \frac{۱۶ + ۱۰}{۵۰} = ۰/۵۲$$

ب) برای سؤال دوم، فضای نمونه محدودتر می‌شود (۲۰ = ٤ + ۱۶)؛ زیرا حالا می‌دانیم که تعمیرگاه دارای ۱۰ سال یا بیشتر سابقه شغلی است پس سطر دوم جدول حذف می‌شود. از این ۲۰ تعمیرگاه، ۱۶ تای آن مجازند. اگر T تعمیرگاهی را نشان دهد که ۱۰ سال یا بیشتر سابقه دارد؛ در این صورت خواهیم داشت:

$$P(L/T) = \frac{۱۶}{۲۰} = ۰/۸$$

همان‌طور که انتظار می‌رفت، P (L / T) به طور قابل ملاحظه‌ای از P (L) بزرگتر است.

در مثال اخیر به نحوهٔ محاسبهٔ P (L / T) توجه کنید. می‌توانستیم این احتمال را با تقسیم P (L ∩ T) بر P (T) تعریف کنیم؛ یعنی:

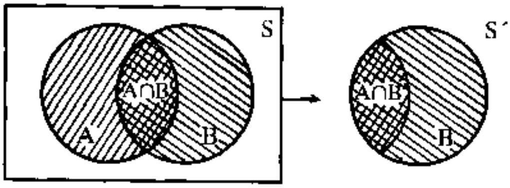
$$P(L/T) = \frac{P(L \cap T)}{P(T)}$$

با تعمیم این مثال، احتمال شرطی به این صورت تعریف می‌شود:

اگر A و B دو پیشامدِ دلخواه از فضای نمونه S باشد و $P(B) \neq ۰$، احتمال وقوع A به شرط B برابر است با:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

در شکل ۱۰ـ۵ نمودار ون مربوط به احتمال وقوع A به شرط B، نشان داده شده است.



شکل ۵.۱۰ احتمال وقوع A به شرط B

چنانکه مشاهده می‌شود در احتمال شرطی فضای نمونه محدودتر شده است.

مثال ۵.۳۹ اطلاعات مربوط به سود خالص یک سال ۱۵۰ شرکت که در ٤ صنعت مختلف فعالیت می‌کنند، به این شرح به دست آمده است:

میزان سود / صنعت	کمتر یا مساوی ۵۰ میلیون ریال E	بیشتر از ۵۰ میلیون ریال F	جمع
صنعت نساجی (A)	۱۷	۱۵	۳۲
صنعت آلومینیم (B)	۳۵	۳۰	٦۵
صنعت مواد غذایی (C)	۲۸	۵	۳۳
صنعت چوب و کاغذ (D)	۱۰	۱۰	۲۰
جمع	۹۰	٦۰	۱۵۰

می‌خواهیم این احتمالات را در صورتی که یکی از این شرکتها برحسب تصادف انتخاب شده باشد، محاسبه کنیم:

الف) احتمال اینکه شرکت انتخابی سودی کمتر یا مساوی ۵۰ میلیون ریال داشته باشد.

ب) احتمال اینکه شرکت انتخابی در صنعت آلومینیم مشغول فعالیت باشد.

ج) احتمال اینکه شرکت انتخابی هم از صنایع نساجی باشد و هم سودی بیشتر از ۵۰ میلیون ریال داشته باشد.

د) احتمال اینکه شرکت انتخابی یا در صنعت نساجی باشد یا سودی بیشتر از ۵۰ میلیون ریال داشته باشد.

ه) احتمال اینکه شرکت انتخابی در صنعت نساجی مشغول فعالیت باشد، در صورتی که بدانیم سودی بیشتر از ۵۰ میلیون ریال دارد.

و) احتمال اینکه شرکت انتخابی سودی کمتر یا مساوی ۵۰ میلیون ریال داشته باشد، در صورتی که بدانیم در صنعت نساجی فعالیت ندارد.

الف) $P(E)=\frac{۹۰}{۱۵۰}=۰/۶$

ب) $P(B)=\frac{۶۵}{۱۵۰}=۰/۴۳$

ج) $P(A\cap F)=\frac{۱۵}{۱۵۰}=۰/۱$

د) $P(A\cup F)=P(A)+P(F)-P(A\cap F)=\frac{۳۲}{۱۵۰}+\frac{۶۰}{۱۵۰}-\frac{۱۵}{۱۵۰}=۰/۵۱$

ه) $P(A/F)=\frac{P(A\cap F)}{P(F)}=\frac{۱۵/۱۵۰}{۶۰/۱۵۰}=\frac{۱۵}{۶۰}=۰/۲۵$

و) $P(E/A^c)=\frac{P(E\cap A^c)}{P(A^c)}=\frac{(۳۵+۲۸+۱۰)/۱۵۰}{(۶۵+۳۳+۲۰)/۱۵۰}=۰/۶۲$

مثال ۵ـ۴۰ در خانواده‌ای که دو فرزند دارند ـ فرض کنید برای داشتن یک فرزند احتمال پسر و دختر شدن برابر است ـ می‌خواهیم این احتمالات را محاسبه کنیم:

الف) احتمال اینکه هر دو فرزند پسر باشند.

ب) احتمال اینکه هر دو پسر باشند، در صورتی که بدانیم حداقل یکی از فرزندها پسر است.

اگر پسر را با B و دختر را با G نشان دهیم، فضای نمونه ما عبارت خواهد بود از: $S=\{GG, GB, BG, BB\}$.

الف) پیشامد داشتن دو پسر را با A نشان می‌دهیم. چون عضوهای فضای نمونه هم شانسند $A=\{BB\}$ است و $P(A)=\frac{۱}{۴}$.

ب) پیشامد داشتن حداقل یک پسر را با C نشان می‌دهیم؛ در این صورت $C=\{GB, BG, BB\}$ و $P(A/C)=\frac{P(A\cap C)}{P(C)}=\frac{۱/۴}{۳/۴}=\frac{۱}{۳}$ است.

۵ـ۶ـ۱ قانون ضرب احتمالات

گفته شد که احتمال A به شرط B برابر است با:

$$P(A/B)=\frac{P(A\cap B)}{P(B)}$$

حال اگر دو طرف این رابطه را در $P(B)$ ضرب کنیم، خواهیم داشت: $P(A\cap B)=P(B)P(A/B)$ که به «قانون ضرب احتمالات» معروف است.

اگر همین عمل را برای $P(B/A)$ انجام دهیم، خواهیم داشت $P(A\cap B)=P(A)P(B/A)$.

قانون ضرب احتمالات ما را در تعیین احتمال اشتراک دو پیشامد که محاسبهٔ آنها به راحتی امکانپذیر نباشد، کمک می‌کند.

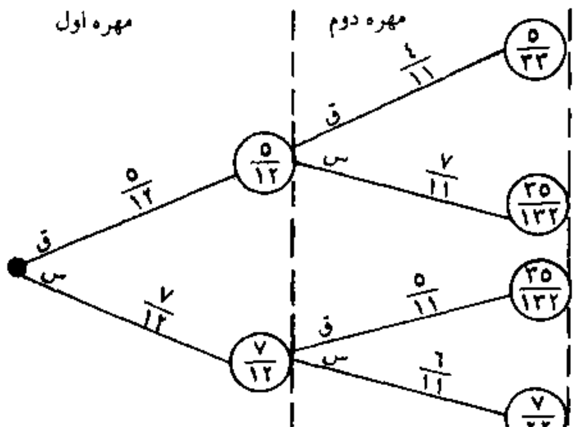
مثال ۵.۴۱ فرض کنید ظرفی حاوی ۱۲ مهره است که ۵ مهرهٔ آن قرمز و بقیه سبز هستند. می‌خواهیم بدانیم اگر دو مهره را بدون جایگزینی بیرون بیاوریم، احتمال آنکه هر دو قرمز باشند چقدر است.

P(قرمز بودن هر دو مهره) = P(قرمز بودن مهره اول) P(قرمز بودن مهره دوم / قرمز بودن مهره اول)

$$= \frac{۵}{۱۲} \times \frac{۴}{۱۱}$$

$$= \frac{۵}{۳۳}$$

برای این مثال، می‌توان از نمودار درختی احتمالات، به این صورت استفاده کرد:



قانون ضرب احتمالات را می‌توان به راحتی برای بیش از دو پیشامد نیز به کار برد. فرمول آن برای سه پیشامد این گونه خواهد بود:

$$P(A\cap B\cap C) = P(A)\,P(B/A)\,P(C/A\cap B)$$

مثال ۴۲ـ۵ با توجه به داده‌های مثال ۴۱ـ۵، اگر ۳ مهره از ظرف بیرون آورده شود می‌خواهیم این احتمالات را محاسبه کنیم:

الف) احتمال اینکه هر ۳ مهره قرمز باشد.

ب) احتمال اینکه اولی و سومی قرمز باشد.

الف) $P(\text{هر سه قرمز باشد}) = \frac{۵}{۱۲} \times \frac{۴}{۱۱} \times \frac{۳}{۱۰} = \frac{۱}{۲۲}$

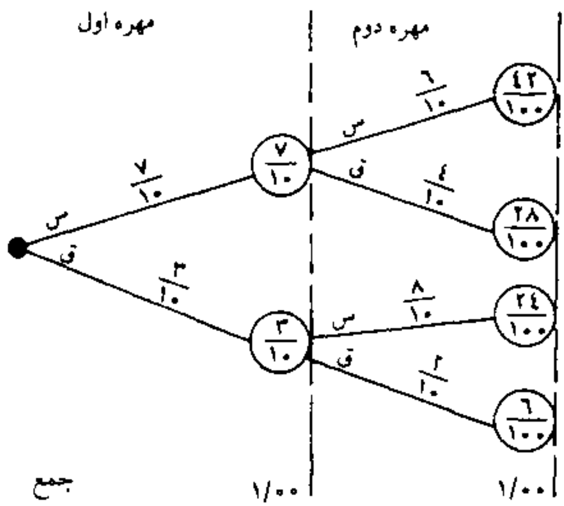
ب) $P(\text{مهره اول و سوم قرمز باشد}) = \frac{۵}{۱۲} \times \frac{۷}{۱۱} \times \frac{۴}{۱۰} = \frac{۷}{۶۶}$

مثال ۴۳ـ۵ ظرفی حاوی ۳ مهره قرمز و ۷ مهره سبز است. به‌طور تصادفی مهره‌ای از ظرف بیرون می‌آوریم و به جای آن مهره‌ای به رنگ دیگر داخل ظرف می‌اندازیم و سپس مهره دوم را از ظرف بیرون می‌آوریم، می‌خواهیم بدانیم هریک از این احتمالات چقدراست.

الف) احتمال آنکه هر دو مهره سبز باشد.

ب) احتمال آنکه مهره دوم سبز باشد.

بهتر است نمودار درختی احتمالات را برای این مثال رسم کنیم.



الف) $P(\text{هر دو سبز باشند}) = \frac{۷}{۱۰} \times \frac{۶}{۱۰} = \frac{۴۲}{۱۰۰} = ۰/۴۲$

ب) $P(\text{مهره دوم سبز باشند}) = P(\text{اولی سبز و دومی سبز باشد}) + P(\text{اولی قرمز و دومی سبز باشد})$

$$= \frac{۷}{۱۰} \times \frac{۶}{۱۰} + \frac{۳}{۱۰} \times \frac{۸}{۱۰} = \frac{۴۲}{۱۰۰} + \frac{۲۴}{۱۰۰} = ۰/۶۶$$

۲ـ۶ـ۵ دو پیشامد مستقل

دو پیشامد را «مستقل» گوییم، در صورتی که وقوع یا عدم وقوع یکی در وقوع و یا عدم وقوع دیگری هیچ تأثیری نداشته باشد؛ مثلاً احتمال اینکه در ساعت هشت و نیم در خیابان زند شیراز تصادفی روی دهد در وقوع و یا عدم وقوع غرق شدن یک قایق در بندر انزلی درهمین ساعت هیچ تأثیری نخواهد داشت.

در قانون ضرب احتمالات داشتیم $P(A\cap B) = P(B)P(A/B)$. اگر A و B دو پیشامد مستقل باشند، وقوع پیشامد B در احتمال وقوع پیشامد A هیچ تأثیری ندارد؛ یعنی $P(A/B) = P(A)$. قانون ضرب احتمالات برای دو پیشامد مستقلِ A و B به این صورت ساده می‌شود:

$$P(A\cap B) = P(A)\,P(B)$$

بنابراین می‌توان وجود این رابطه را «شرط استقلال» دو پیشامد نیز دانست.

مثال ۵ـ۴۴ سکه‌ای را دوبار پرتاب می‌کنیم. می‌دانیم در بار اول شیر آمده است، می‌خواهیم احتمال اینکه در بار دوم نیز شیر بیاید را محاسبه کنیم.

خط و شیر آمدن سکه در بار دوم هیچ ارتباطی به شیر آمدن در دفعهٔ اول ندارد؛ چرا که این دو از هم مستقلند، بنابراین احتمال شیرآمدن دردفعه دوم ۰/۵ است؛ به عبارت دیگر:

$$P(H/H) = P(H) = ٠/٥$$

گاهی دو پیشامد مستقل و ناسازگار از هم تمیز داده نمی‌شوند. دو پیشامد ناسازگار را دو پیشامدی تعریف کردیم که اگر یکی رخ دهد دیگری قطعاً رخ نخواهد داد. مجموعهٔ اشتراک دو پیشامد ناسازگار تهی است. دو پیشامد مستقل را دو پیشامدی تعریف کردیم که هیچ تأثیری روی هم ندارند؛ به عبارت دیگر اگر A و B دو پیشامد از فضای نمونه یک آزمایش باشند، خواهیم داشت:

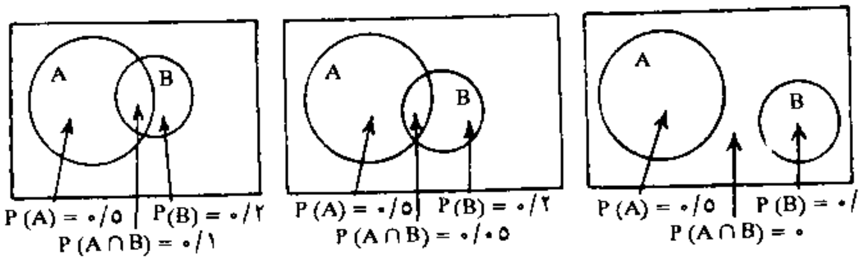
در دو پیشامد ناسازگار: $P(A\cap B) = ٠$

در دو پیشامد مستقل: $P(A\cap B) = P(A)\,P(B)$

همچنین دقت نمایید که مستقل بودن و ناسازگار بودن پشت و روی یک سکه

نیستند؛ به عبارت دیگر اگر دو پیشامد مستقل نبودند نمی‌توان گفت که قطعاً ناسازگارند بعلاوه ممکن است دو پیشامد غیرمستقل سازگار باشند.

مثال ٥.٤٥ این سه نمودار را در نظر بگیرید. می‌خواهیم بدانیم کدام یک ناسازگار، کدام یک مستقل و کدام یک نه ناسازگار و نه مستقلند.



نمودار ون سمت راست نشان‌دهندهٔ دو پیشامد ناسازگار است؛ زیرا $P(A \cap B) =$ ۰ و نمودار ون سمت چپ نشان‌دهندهٔ دو پیشامد مستقل است؛ زیرا ۰/۱ = $P(A \cap B) = P(A)\,P(B) =$ ۰/۵ × ۰/۲ = ۰/۱ و نمودار ون میانی نشان‌دهندهٔ دو پیشامدی است که نه مستقلند و نه ناسازگار؛ زیرا ۰ ≠ $P(A \cap B) =$ ۰/۰۵ و ۰/۰۵ = $P(A \cap B) \neq P(A)\,P(B) =$ ۰/۵×۰/۲ = ۰/۱.

تمرین

۱. فرض کنید اگر خانواده‌ای بیش از ۵۰ میلیون ریال در سال درآمد داشته باشند، احتمال اینکه دو اتومبیل داشته باشند ۷/۵ درصد است. ٦ درصد خانواده‌ها درآمد سالیانه‌شان بیش از ۵۰ میلیون ریال است و ۵/۲ درصد دو اتومبیل دارند؛ احتمال اینکه خانواده‌ای دو اتومبیل داشته و درآمدش نیز در سال بیش از ۵۰ میلیون ریال باشد، چقدر است؟

۲. اگر $P(A) = \frac{۳}{۱٤}$، $P(B) = \frac{۱}{٦}$، $P(C) = \frac{۱}{۳}$، $P(A \cap C) = \frac{۱}{۷}$ و $P(B/C) = \frac{۵}{۲۱}$ باشد؛ احتمالات $P(A/C)$، $P(C/A)$، $P(B \cap C)$، $P(C/B)$ را حساب کنید.

۳. دو پیشامد A و B و $P(A) =$ ۰/٦۵، $P(B) =$ ۰/۸۰، $P(A/B) = P(A)$ و $P(B/A) =$ ۰/۸۵ را در نظر بگیرید؛ آیا این احتمالات با هم همخوانی دارند؟ چرا؟

٤. فرض کنید A و B پیشامدهای مستقلی باشند به طوری که $P(A) =$ ۰/٦ و $P(B) =$ ۰/۲. مطلوب است محاسبهٔ $P(A \cup B^c)$.

۵. فرض کنید این اطلاعات را اخیراً از آموزشِ دانشکدهٔ زبانهای خارجی دانشگاه تهران گرفته‌ایم:

رشته / دانشجوی سال	زبان انگلیسی	زبان روسی	زبان آلمانی
اول	۲۰۰	۱۸۵	٤۰
دوم	۱۸۰	۱۲۰	۳۰
سوم	۱۲۰	۸۵	۳۵
چهارم	۱۰۵	۹۰	۲۰

می‌دانیم تعداد دانشجویان این دانشکده ۱۲۱۰ نفر است، اگر برحسب تصادف دانشجویی را انتخاب کنیم، هریک از این احتمالات چقدر است؟

الف) احتمال اینکه دانشجوی زبان روسی نباشد.

ب) احتمال اینکه دانشجوی زبان آلمانی نباشد، در صورتی که بدانیم این فرد دانشجوی سال چهارم است.

ج) احتمال اینکه هم دانشجوی سال دوم و هم زبان روسی باشد.

ه) آیا دانشجوی رشتهٔ زبان انگلیسی بودن مستقل از دانشجوی سال اول بودن است؟

٦. جعبه‌ای محتوی ۳ سکه است. یکی از آنها سالم، یکی دیگر دو شیره و دیگری نیز ناسالم است؛ به طوری که احتمال شیرآمدن سکه ناسالم $\frac{۱}{۳}$ است. سکه‌ای را به طور تصادفی انتخاب و پرتاب می‌کنیم. نمودار درختی آن را رسم کرده، مشخص نمایید احتمال آنکه شیر ظاهر شود، چقدر است؟

۷. در کلاسی ۱۰ پسر و ۵ دختر شرکت دارند. ۳ دانشجو به طور تصادفی یکی پس از دیگری انتخاب می‌شوند؛ هریک از این احتمالات را محاسبه کنید.

الف) احتمال اینکه دو دانشجوی اول پسر و سومی دختر باشد.

ب) احتمال اینکه اولی و سومی دختر باشد.

ج) احتمال اینکه اولی و سومی همجنس و دومی از جنس مخالف باشد.

۷ـ۵ قضیهٔ بیز (احتمالات پیشین و پسین)

ممکن است شما احتمال وقوع پیشامدی را در شرایط معمولی بدانید، ولی اگر اطلاعات جدیدی به دست آورید در احتمال وقوع پیشامد اولیه تجدید نظر کنید. به احتمال وقوع پیشامدی قبل از کسب اطلاعات جدید «احتمال پیشین[1]» و به احتمال وقوع آن پیشامد

1. prior probability

بعد از کسب اطلاعات جدید «احتمال پسین[1]» می‌گوییم. این دو گونه احتمال در نظریه تصمیم‌گیری کاربرد فراوان دارد و قضیهٔ بیز ما را در تجدید نظر احتمالات، در صورت دسترسی به اطلاعات جدید، کمک می‌کند؛ مثلاً مدیر شرکتی که تولید کنندهٔ پوشاک گرم است فکر می‌کند به احتمال ۷۰ درصد دمای زمستان امسال بین ۲– تا ۱۵– درجه سانتیگراد است و به فکر می‌افتد تعداد مشخصی پوشاک گرم تولید کند. حال اگر از رادیو بشنود که به دلایلی دمای زمستان امسال خیلی پایین‌تر از سالهای قبل خواهد بود، ممکن‌است احتمالش به ۸۵ درصد افزایش یافته و بر تعداد تولیدش بیفزاید. احتمال اول (۷۰ درصد) را احتمال پیشین و احتمال دوم (۸۵ درصد) را احتمال پسین می‌گوییم.

در مبحث احتمال شرطی داشتیم:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \, , \; P(A \cap B) = P(A)\,P(B/A)$$

حال با جایگزین کردن $P(A)\,P(B/A)$ به جای $P(A \cap B)$ خواهیم داشت:

$$P(A/B) = \frac{P(A)\,P(B/A)}{P(B)}$$

این فرمول شکل ساده‌ای از قضیهٔ بیز است.

مثال ۴۶-۵ در شهری ماشینها را از نظر دودزایی آزمایش و در صورت تشخیص دودزایی بیش از حد، ماشین را متوقف می‌کنند. ۲۵ درصد از ماشینهای این شهر بیش از حد دودزا هستند. در آزمایش، ۹۸ درصد از ماشینهایی که بیش از حد دودزا باشند، دودزای بیش از حد تشخیص داده می‌شوند، اما تنها ۷ درصد از ماشینهایی که بیش از حد دودزا نیستند، دودزای بیش از حد تشخیص داده می‌شوند. می‌خواهیم بدانیم احتمال اینکه ماشینی که دودزای بیش از حد تشخیص داده شده واقعاً بیش از حد دودزا باشد چقدر است.

اگر A را پیشامد تشخیص دودزایی بیش از حد و B را پیشامد دودزایی بیش از حد ماشین تعریف کنیم، طبق اطلاعات مثال داریم: $P(B) = ۰/۲۵$ ، $P(A/B) = ۰/۹۸$ و $P(A/B^c) = ۰/۰۷$.

قبل از اینکه $P(B/A)$ را حساب کنیم، ناچاریم احتمال $P(A)$ را حساب کنیم. $P(A)$ یا احتمال تشخیص دودزایی بیش از حد برابر است با احتمال اینکه ماشین دودزای بیش از حد باشد و تشخیص هم این گونه باشد به اضافه احتمال اینکه ماشین دودزای بیش از

1. posterior probability

حد نباشد، ولی تشخیص داده شود که دودزای بیش از حد است؛ یعنی:

$$P(A) = (۰/۲۵)(۰/۹۸) + (۰/۷۵)(۰/۰۷) = ۰/۲۹۷۵$$

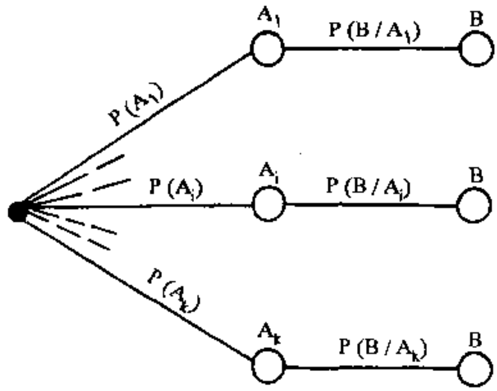
بنابراین احتمال اینکه ماشینی که دودزای بیش از حد تشخیص داده شده واقعاً هم بیش از حد دودزا باشد برابر است با:

$$P(B/A) = \frac{P(B)\,P(A/B)}{P(A)} = \frac{(۰/۲۵)(۰/۹۸)}{۰/۲۹۷۵} = ۰/۸۲$$

دربسیاری از موارد، معمولاً عواملی که نقش تعیین‌کننده‌ای در وقوع پیشامدی دارند بیش از دو عامل هستند. اگر A_1 ، A_2 ، ... و A_K نشان دهندهٔ K حادثهٔ ناسازگار باشند که می‌توانند حادثهٔ B را باعث شوند؛ آنگاه احتمال اینکه A_i عامل وقوع باشد، برابر است با:

$$P(A_i/B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)\,P(B/A_i)}{P(A_1)P(B/A_1) + \ldots + P(A_i)P(B/A_i) + \ldots + P(A_K)P(B/A_K)}$$

می‌توان این فرمول را به صورت نمودار درختی شکل ۱۱ـ۵ نمایش داد.



شکل ۱۱ـ۵ نمودار درختی احتمال برای فرمول بیز

مثال ۴۷ـ۵ سه ماشین A ، B و C به ترتیب ۶۰ درصد، ۳۰ درصد، و ۱۰ درصد کل محصولات کارخانه‌ای را تولید می‌کنند. درصد محصولات معیوب این ماشینها به ترتیب برابر ۲ درصد، ۳ درصد و ۴ درصد است. از میان محصولات این کارخانه محصولی به صورت تصادفی انتخاب می‌شود، می‌خواهیم هریک از این احتمالات را محاسبه کنیم:

الف) احتمال اینکه معیوب باشد.

ب) احتمال اینکه با ماشین C تولید شده باشد، در صورتی که بدانیم کالا معیوب است.
اگر پیشامد معیوب بودن را با X نشان دهیم:

الف)
$$P(X) = P(A)\,P(X/A) + P(B)\,P(X/B) + P(C)\,P(X/C)$$

$$= ۰/۶۰ \times ۰/۰۲ + ۰/۳۰ \times ۰/۰۳ + ۰/۱۰ \times ۰/۰۴ = ۰/۰۲۵$$

ب)
$$P(C/X) = \frac{P(C)\,P(X/C)}{P(A)\,P(X/A) + P(B)\,P(X/B) + P(C)\,P(X/C)}$$

$$= \frac{۰/۱۰ \times ۰/۰۴}{۰/۶۰ \times ۰/۰۲ + ۰/۳۰ \times ۰/۰۳ + ۰/۱۰ \times ۰/۰۴} = ۰/۱۶$$

مثال ۴۸ـ۵ ظرفی حاوی ۳ مهرهٔ قرمز و ۷ مهرهٔ سبز است. مهره‌ای از ظرف بیرون می‌آوریم و سپس مهره‌ای به رنگ دیگر به داخل ظرف می‌اندازیم و سپس مهرهٔ دومی را از ظرف بیرون می‌آوریم. اگر دو مهرهٔ انتخابی همرنگ باشند، می‌خواهیم بدانیم احتمال قرمز بودن آنها چقدر است.

اگر پیشامد قرمز بودن دو مهره را با RR، همرنگ بودن آنها را با S و سبز بودن آنها را با GG نشان دهیم، خواهیم داشت:

$$P(RR/S) = \frac{P(RR)\,P(S/RR)}{P(RR)\,P(S/RR) + P(GG)\,P(S/GG)}$$

$$= \frac{\frac{۶}{۱۰۰} \times ۱}{\frac{۶}{۱۰۰} \times ۱ + \frac{۴۲}{۱۰۰} \times ۱} = ۰/۱۲۵$$

نمودار درختی این احتمال در مثال ۴۳ـ۵ آورده شده است.

تمرین

۱. احتمال وقوع ۳ پیشامد A، B و C به ترتیب برابر است با ۰/۳۵، ۰/۴۵ و ۰/۲۰. احتمال وقوع پیشامد X به شرط وقوع هریک عبارت است از: $P(X/A) = ۰/۸$، $P(X/B) = ۰/۳$ و $P(X/C) = ۰/۶۵$. موارد ذیل را پیدا کنید:

الف) $P(A/X)$،

ب) $P(B/X)$،

ج) $P(C/X)$.

۲. پزشک یک تیم فوتبال می‌داند که در زمستان، ۴۰ درصد از بازیهای تیم روی چمن مصنوعی انجام می‌شود. او همچنین می‌داند که خطر مصدوم شدن یک بازیکن در چمن مصنوعی ۵۰ درصد بیشتر از چمن طبیعی است. اگر احتمال مصدوم شدن یک بازیکن در چمن مصنوعی ۰/۴۲ باشد، هریک از احتمالات زیر را محاسبه کنید:

الف) احتمال اینکه بازیکنی مصدوم شود.

ب) اگر بازیکنی مصدوم شده باشد، احتمال آنکه روی چمن مصنوعی مصدوم شده باشد.

۳. تمرین ۲ را در نظر بگیرید. پزشک تیم علاقه‌مند به بررسی رابطه بین مصدوم شدن و نوع بازیکن نیز هست. اطلاعاتش را در طول ۳ سال جمع‌آوری و در این جدول خلاصه کرده است:

نوع بازیکن / شرح	خط حمله	خط دفاعی	خط میانی	دروازه‌بان
تعداد بازیکنان	۴۵	۵۶	۲۴	۸
تعداد مصدومین	۳۲	۳۸	۱۱	۳

اگر بازیکن مصدومی را برحسب تصادف انتخاب کنیم، احتمال اینکه وی به ترتیب بازیکن خط حمله، خط دفاع، خط میانی، یا دروازه‌بان باشد، چقدر است؟

۴. ظرفی حاوی ۵ مهرهٔ قرمز و ۷ مهرهٔ سفید است. مهره‌ای را به صورت تصادفی از ظرف خارج می‌کنیم و به جای آن ۲ مهره به رنگ دیگر در داخل ظرف می‌اندازیم و سپس مهرهٔ دومی را از ظرف به صورت تصادفی بیرون می‌آوریم، هریک از این احتمالات را محاسبه کنید:

الف) احتمال اینکه هر دو مهره همرنگ باشند.

ب) احتمال اینکه هر دو مهره سفید باشند، در صورتی که بدانیم هر دو همرنگ هستند.

۸ـ۵ خلاصه

در اغلب موارد به دلیل فقدان اطلاعات کافی و کامل در مورد پدیده‌ها، برای اظهارنظر در مورد آنها، از «احتمال» استفاده می‌شود؛ به همین دلیل، مدیران معمولاً با استفاده از نظریه احتمال، مبادرت به پیش‌بینی و تصمیم‌گیری می‌کنند.

در این فصل، ضمن مطرح کردن مفاهیم اولیه احتمال (مفهوم احتمال، احتمال عینی و ذهنی، آزمایش، فضای نمونه و انواع آن، پیشامد و پیامدهای مقدماتی هم شانس)، نحوهٔ محاسبه احتمال یک پیشامد مشخص شد. همچنین گفته شد که در بعضی از شرایط نیازی به فهرست کردن همهٔ عناصر فضای نمونه نیست؛ بدین جهت با بیان

کردن قواعد احتمال و بررسی پیشامدها، مفهوم احتمال شرطی و نحوه تبدیل احتمالات پیشین به احتمالات پسین (با استفاده از قضیه بیز) توضیح داده شد.

۹ـ۵ سؤالات و مسائل

سؤالات دوگزینه‌ای

۱. در نظریهٔ احتمال، به مجموعهٔ پیامدهای ممکن یک آزمایش، فعالیت می‌گویند. ص ☐ غ ☐

۲. با کسب اطلاعات جدید و با استفاده از قضیهٔ بیز می توان احتمالات پیشین را به احتمالات پسین تبدیل کرد. ص ☐ غ ☐

۳. احتمال ذهنی چیزی بیش از حدسی زیرکانه نیست. ص ☐ غ ☐

٤. در هنگام استفاده از فراوانی نسبی به عنوان احتمال، هر چه تعداد آزمایشها بیشتر شود احتمال کم اهمیت‌تر می‌گردد. ص ☐ غ ☐

۵. دو پیشامد لزوماً یا مستقلند یا ناسازگار. ص ☐ غ ☐

٦. قضیهٔ بیز، فرمولی برای تجدیدنظر در احتمالات پیشین است. ص ☐ غ ☐

۷. استفاده از فراوانی نسبی به عنوان احتمال نیازمند به حداکثر ۱۰۰ آزمایش است. ص ☐ غ ☐

۸. دو پیشامد لزوماً یا سازگارند یا ناسازگار. ص ☐ غ ☐

۹. اگر وقوع یک پیشامد بر احتمال وقوع پیشامد دیگر هیچ تأثیری نداشته باشد، این دو پیشامد را از نظر آماری مستقل می‌گوییم. ص ☐ غ ☐

۱۰. $P(A \cup B)$ یعنی احتمال وقوع هم پیشامد A و هم پیشامد B. ص ☐ غ ☐

۱۱. دو پیشامد A و B مستقلند، اگر $P(A/B) = P(A)$. ص ☐ غ ☐

۱۲. اگر دو پیشامد A و B مستقل باشند، آنگاه احتمال وقوع هم A و هم B برابراست با $P(A) \times P(B)$. ص ☐ غ ☐

۱۳. تعداد عضوهای فضای نمونهٔ نامحدود، لزوماً نامتناهی نیست. ص ☐ غ ☐

۱٤. احتمال وقوع یک پیشامد همواره کوچکتر از یک است. ص ☐ غ ☐

۱۵. در ترکیب، نحوهٔ آرایش اشیا مهم است. ص ☐ غ ☐

۱٦. تعداد جایگشتهای دوری n شیء برابر است با $(n-1)!$. ص ☐ غ ☐

۱۷. تعداد ترکیبهای r شیء از n شیء را با p_r^n نشان می‌دهیم. ص ☐ غ ☐

۱۸. همواره رابطهٔ $P(A \cap B) \geq P(A \cup B)$ برقرار است. ص ☐ غ ☐

۱۹. همواره رابطهٔ $P(A \cup B) \leq P(A) + P(B)$ برقرار است. ص ☐ غ ☐

۲۰. همواره رابطهٔ $P(A) + P(A^c) = 1$ برقرار است. ص ☐ غ ☐

۲۱. به پیشامدهایی که نمی‌توانند همزمان رخ دهند، پیشامدهای مستقل می‌گویند.
ص ☐ غ ☐

۲۲. شما معتقدید فردا احتمال باریدن برف ۰/۰٤ است، این احتمال عینی است. ص ☐ غ ☐

۲۳. احتمال A را با علم به وقوع پیشامد B، احتمال شرطی B به شرط A می‌گوییم. ص ☐ غ ☐

۲٤. سکه‌ای را قبلاً پرتاب کرده‌ایم، خط آمده است. حال در پرتاب مجدد آن انتظار داریم احتمال شیر آمدن بالاتر رود. ص ☐ غ ☐

۲۵. در احتمال شرطی، فضای نمونه کل آزمایش به فضای نمونهٔ کوچکتری تقلیل می‌یابد. ص ☐ غ ☐

سؤالات چهارگزینه‌ای

۲٦. احتمال آنکه مقداری که به صورت تصادفی از جامعه‌ای انتخاب می‌شود بیشتر از میانگین آن جامعه باشد چقدر است؟

الف) ۰/۲۵ ج) ۱

ب) ۰/۵ د) ۰/٦۷

۲۷. فرض کنید دو تاس به همراه سکه‌ای پرتاب می‌شود و از شما خواسته می‌شود درخت احتمال آنها را رسم کنید؛ این درخت چند شاخه خواهد داشت؟

الف) ۳۸ ج) ۱٤

ب) ۷۲ د) ۷٤

۲۸. اگر A و B دو پیشامد مستقل و $P(A) = ۰/٤$ و $P(A \cap B) = ۰/۲$ باشد، احتمال اجتماع این دو پیشامد چقدر است؟

الف) ۰/٦ ج) ۰/۸

ب) ۰/۷ د) ۰/۹

۲۹. اگر A و B دو پیشامد ناسازگار و $P(A) = ۰/٤$ و $P(B^c) = ۰/۵$ باشد، کدام‌یک از موارد صحیح نیست؟

الف) $P(A \cup B) = ۰/۹$ ج) $P(A^c \cup B) = ۰/٦$

ب) $P(A^c) = ۰/٦$ د) $P(A^c \cap B^c) = ۰$

۳۰. ٦۰ درصد محصولات کارخانه‌ای به وسیلهٔ ماشین شماره یک و ٤۰ درصد به وسیلهٔ ماشین شماره دو تولید می‌شوند. ۰/۰۲ محصولات ماشین شماره یک و ۰/۰۱ محصولات ماشین شمارهٔ دو

معیوبند. اگر یک کالا از محصولات کارخانه انتخاب شود، احتمال سالم بودن آن چقدر است؟

الف) ۰/۰۱۶ ج) ۰/۹۹۶

ب) ۰/۰۱۲ د) ۰/۹۸٤

۳۱. در سؤال ۳۰ ، اگر بدانیم کالای انتخابی معیوب است، احتمال اینکه به وسیلهٔ ماشین شماره دو تولید شده باشد چقدر است؟

الف) ۰/۲۰ ج) ۰/۷۰

ب) ۰/۲۵ د) ۰/۷۵

۳۲. مأمور کنترل کیفی کارخانه‌ای از بین دو انبار به طور تصادفی یک انبار و سپس کالایی را انتخاب و بازرسی می‌کند. انبار شماره یک دارای ۳۰ واحد کالاست که ۳ واحد آنها معیوب است و انبار شمارهٔ دو دارای ۱۰۰ واحد کالاست که ۱۰ واحد آنها معیوب است. احتمال معیوب بودن کالای انتخابی چقدر است؟

الف) ۰/۰۱ ج) ۰/۲۰

ب) ۰/۱۰ د) ۰/۰۰۲۵

۳۳. سیستمی دارای دو جزء است که احتمال کار نکردن هرکدام از آنها ۰/۲۰ است، اگر اجزا به صورت سری قرار گرفته باشند و مستقل از هم کار کنند، احتمال کارکردن سیستم چقدر است؟

الف) ۰/۹٦ ج) ۰/٤۰

ب) ۰/۰٤ د) ۰/٦٤

۳٤. سؤال ۳۳ را در نظر بگیرید. اگر اجزا به صورت موازی قرار گرفته باشند، احتمال کارکردن سیستم چقدر است؟

الف) ۰/۹٦ ج) ۰/٤۰

ب) ۰/۰٤ د) ۰/٦٤

۳۵. به چند طریق می‌توان با اعداد صفر تا ۹، شماره تلفن ٦ رقمی ساخت؟

الف) $۱۰^{٦}$ ج) ۹×۱۰^{۵}

ب) $۹^{٦}$ د) ٦×۱۰^{۵}

۳٦. اگر $P(A) = \frac{۱}{۳}$ ، $P(B) = \frac{۱}{۵}$ و $P(A/B) = \frac{۱}{۳}$ باشد، $P(A \cup B)$ چقدر است؟

الف) $\frac{۸}{۱۵}$ ج) $\frac{٦}{۱۵}$

ب) $\frac{۷}{۱۵}$ د) $\frac{۹}{۱۵}$

۳۷. فرض کنید احتمال داشتن فرزند پسر و دختر مساوی باشد. احتمال اینکه خانواده‌ای که ۳

فرزند دارد، دو پسر داشته باشد، چقدر است؟

الف) $\frac{۲}{۸}$ ج) $\frac{۱}{۸}$

ب) $\frac{۳}{۸}$ د) $\frac{٤}{۸}$

۳۸. در سؤال پیش، احتمال اینکه حداقل یک فرزند پسر داشته باشد، چقدر است؟

الف) $\frac{۵}{۸}$ ج) $\frac{۷}{۸}$

ب) $\frac{٦}{۸}$ د) ۱

۳۹. به چند طریق می‌توان ۳ کتاب آمار و ۲ کتاب ریاضی را در یک قفسه قرار داد؟

الف) ۱۲۰ ج) ۲٤

ب) ۱۲ د) ۹٦

٤۰. در سؤال پیش، اگر قرار باشد کتابهای آمار پهلوی هم قرار گیرند، به چند طریق می‌توان این ۵ کتاب را کنار هم قرار داد؟

الف) ۱۲ ج) ۳٦

ب) ۲٤ د) ٤۸

٤۱. در سؤال ۳۹، اگر قرار باشد کتابهای آمار پهلوی هم و کتابهای ریاضی نیز پهلوی هم قرار گیرند، به چند طریق می‌توان آنها را کنار هم قرار داد؟

الف) ۱۲ ج) ۳٦

ب) ۲٤ د) ٤۸

٤۲. به چند طریق می‌توان از ۱۲ کتاب، ۳ کتاب را به عنوان کتاب سال برگزید؟

الف) ۳٦ ج) ۱۳۲۰

ب) ۷۲ د) ۲۲۰

٤۳. به چند طریق می‌توان از ۱۲ کتاب که ۵ تای آن آمار و بقیه ریاضی هستند، یک کتاب آمار و ۲ کتاب ریاضی را به عنوان کتاب سال برگزید؟

الف) ۲۲۰ ج) ۱۱۰

ب) ۲۰۵ د) ۱۰۵

٤٤. در کلاسی ۵ دانشجوی دختر و ۱۰ دانشجوی پسر وجود دارد. اگر ۳ دانشجو به طور تصادفی انتخاب شود، احتمال اینکه هر سه پسر باشند، چقدر است؟

الف) $\frac{۸}{۲۷}$ ج) $\frac{۳}{۸}$

ب) $\frac{۲٤}{۹۱}$ د) $\frac{۵}{۱٦}$

٤۵. به چند طریق می‌توان دور یک میدان، ۱۰ پرچم مختلف را برافراشت؟

الف) !۱۰ ج) ۱ - !۱۰

ب) !۹ د) ۱ - !۹

٤٦. دانشجویی موظف است از ۵ سؤال اول به ۳ سؤال، و از ۱۵ سؤال بعد به ۱۲ سؤال جواب دهد، به چند طریق می‌تواند به سؤالات جواب دهد؟

الف) ۵۰۰٤ ج) ٤۵۵۰

ب) ۵۵٤۰ د) ۵٤۵۰

٤۷. ظرفی حاوی ۵ مهره قرمز و ۷ مهرهٔ سفید است. مهره‌ای را به طور تصادفی از ظرف خارج می‌کنیم، اگر قرمز بود به همراه آن یک مهرهٔ قرمز دیگر، و اگر سفید بود به همراه آن ۲ مهرهٔ سفید دیگر به داخل ظرف می‌اندازیم و سپس مهره دوم را بر حسب تصادف خارج می‌کنیم. احتمال آنکه مهرهٔ دوم قرمز باشد چقدر است؟

الف) $\frac{۱۲۵}{۳۱۲}$ ج) $\frac{۳۷}{۱٦۱}$

ب) $\frac{۵}{۲٤}$ د) $\frac{۷}{۳۲}$

٤۸. در سؤال پیش، اگر بدانیم مهرهٔ خارج شده در بار اول سفید بوده است، احتمال آنکه مهرهٔ دوم قرمز باشد چقدر است؟

الف) $\frac{۹}{۱٤}$ ج) $\frac{۹}{۱۲}$

ب) $\frac{۵}{۱۲}$ د) $\frac{۵}{۱٤}$

٤۹. در سؤال ٤۷، احتمال همرنگ نبودن ۲ مهره چقدر است؟

الف) $\frac{٤۵}{۱۰٤}$ ج) $\frac{۱۲۵}{۳۱۲}$

ب) $\frac{۵۹}{۱۰٤}$ د) $\frac{۱۸۷}{۳۱۲}$

۵۰. به چند طریق می‌توان ۹ اسباب‌بازی را بین ٤ بچه تقسیم کرد به شرط آنکه به کوچکترین بچه ۳ اسباب‌بازی و به هرکدام از بچه‌های دیگر ۲ اسباب بازی برسد؟

الف) ۷۵٦۰ ج) ۵٦۷۰

ب) ۵۷٦۰ د) ۷٦۵۰

مسائل

۵۱. ٤ داوطلب برای پُست مدیریت مالی، ۵ داوطلب برای پست مدیریت اداری، و ۲ داوطلب برای ریاست سازمانی وجود دارند.

الف) یک رأی دهنده به چند طریق می‌تواند رأی خود را به این سه نفر (برای هر پست یک نفر) بدهد؟

ب) یک رأی دهنده به چند طریق می‌تواند رأی دهد، هرگاه از حق رأی ندادن خود به هریک از این داوطلبها استفاده کند؟

۵۲. چند جایگشت مختلف می‌توان با کلمهٔ «برنامه‌ریزی» ساخت؟

۵۳. اگر ۸ نفر شام را با هم صرف کنند، به چند طریق می‌توانند ۳ غذای مرغ، ٤ غذای ماهی و یک غذای میگو سفارش دهند؟

۵٤. دو قفل بر در وجود دارد و کلیدهای آنها در بین ٦ کلیدی است که معمولاً در جیب خود دارید. به دلیل شتابزدگی یکی از آنها را گم کرده‌اید، احتمال اینکه هنوز هم بتوانید در را باز کنید چقدر است؟

۵۵. تاسی را ۳ بار پرتاب می‌کنیم. احتمال اینکه:

الف) هر بار عدد بزرگتری نسبت به بار پیش بیاید چقدر است؟

ب) هر بار یک واحد به شمارهٔ تاس اضافه شود چقدر است؟

۵٦. به چند طریق ۵ نفر می‌توانند برای سوار شدن به اتوبوس صف ببندند؟ اگر ۲ نفر از ۵ نفر از کنار هم ایستادن ابا کنند، به چند طریق صف‌بندی میسر است؟

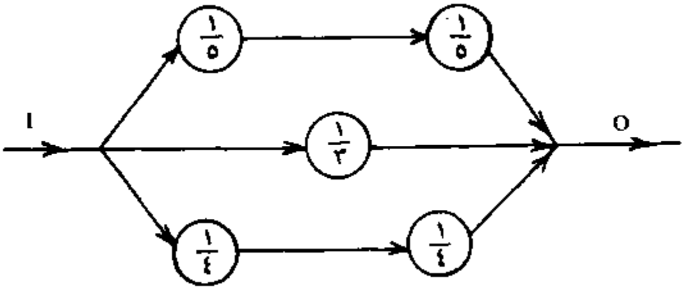
۵۷. فرض کنید علاقه‌مند به تکمیل پروژه‌ای هستیم که ممکن است به علت اعتصاب به تأخیر افتد. بعلاوه فرض کنید احتمال اینکه اعتصابی رخ دهد ۰/٦۰، احتمال اینکه اگر اعتصابی نباشد کار بموقع انجام شود ۰/۸۵ و احتمال اینکه اگر اعتصابی باشد کار بموقع انجام شود ۰/۳۵ باشد. احتمال اینکه کار بموقع انجام شود چقدر است؟

۵۸. در طول یک ماه ۸۰ درصد سهام صنعتی خاص، که در آن ۱۰ شرکت سهام خود را عرضه کرده‌اند، افزایش یافته است. اگر سهامداری سهام دو شرکت را خریده باشد، احتمال اینکه هر دو افزایش یافته باشد، چقدر است؟

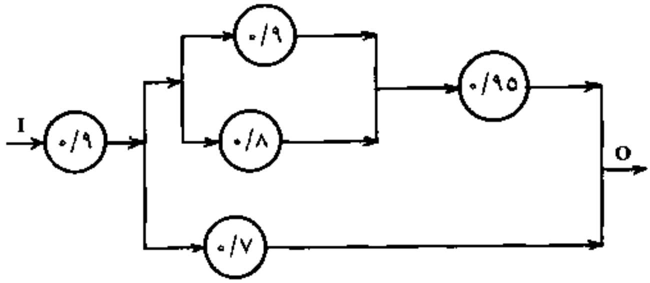
۵۹. احتمال اینکه شیوهٔ جدید بازاریابی موفقیت آمیز باشد ٦۰ درصد و احتمال اینکه هزینه‌های این شیوه از میزان بودجه‌بندی شده، تجاوز نکند ۵۰ درصد است. احتمال تحقق هر دو هدف ۳۰ درصد است. احتمال اینکه دست کم به یکی از این اهداف برسیم، چقدر است؟

۶۰. اگر $P(A^c) = \%۷۰$، $P(B) = \%۵۰$ و $P(A/B) = \%۶۰$ باشد، $P(A \cup B)$ چقدر است؟

۶۱. این مدار الکتریکی را در نظر بگیرید. در آن اعداد، احتمالهای از کار افتادن اتصالهای مختلفی راکه از هم مستقلند، نشان می‌دهد. احتمال اینکه جریان برق برقرار باشد، چقدر است؟



۶۲. در مدار ذیل، اگر اعداد نشان‌دهندهٔ احتمال کارکردن هر جزء باشد و کارکرد هر جزء مستقل از دیگری باشد، احتمال برقراری جریان در مدار چقدر است؟



۶۳. شخصی در آزمونی شرکت می‌کند و یکی از نمرات الف، ب، ج، د و هـ را دریافت می‌کند. چنانچه نمرهٔ الف بگیرد، برنده اعلام می‌شود و دیگر در آزمون شرکت نمی‌کند. چنانچه نمرهٔ او هـ باشد حق شرکت مجدد در آزمون را ندارد؛ در غیر این دو حال آنقدر در آزمون شرکت می‌کند که یکی از دونمرهٔ فوق را دریافت کند. فرض کنید که نتیجهٔ آزمونها مستقل از یکدیگر و احتمال گرفتن نمره‌های الف، ب، ج، د و هـ به ترتیب P_1، P_2، P_3، P_4 و P_5 باشد. احتمال توقف شرکت در آزمون به علت گرفتن نمرهٔ الف چقدر است؟

۶۴. دفتر انتشارات دانشگاهی دارای ۳ دستگاه چاپ است. دفتر انتشارات بابت تأخیر در انجام دادن کارها باید جریمه‌ای را پرداخت کند. داده‌های این جدول نشان‌دهندهٔ سوابق کاری این دستگاههاست.

دستگاه چاپ شمارهٔ i	نسبت قراردادهای مربوط به دستگاه شمارهٔ i	نسبت زمانهای تأخیر در تحویل بیش از یک ماه
۱	۰/۲۰	۰/۲
۲	۰/۳۰	۰/۵
۳	۰/۵۰	۰/۳

تکثیر جزوهٔ یک دانشکده بیش از یک ماه تأخیر داشته است. احتمال اینکه تأخیر ناشی از دستگاه شمارهٔ ۲ باشد چقدر است؟

۶۵. در کمیتهٔ کارشناسی تشکیلات و روشها، ۱۲ کارشناس بهبود روشها و ٤ کارشناس تشکیلات حضور دارند. اگر ۳ کارشناس به طور تصادفی انتخاب شوند، هریک از احتمالات زیر را محاسبه کنید:

الف) احتمال اینکه کلیه آنها کارشناس بهبود روشها باشند.

ب) احتمال اینکه یکی از آنها کارشناس تشکیلات باشد.

٦٦. احتمال زنده بودن یک زن و شوهر در ۲۰ سال آینده به ترتیب $\frac{۳}{۵}$ و $\frac{۱}{۲}$ است. احتمال اینکه در این مدت دست کم یکی از آنها زنده بماند، چقدر است؟

٦٧. فرض کنید ٦ کارت وجود دارد که شماره‌های ۱ تا ٦ روی آنها نوشته شده است. با قراردادن این کارتها به ترتیبهای مختلف، شماره‌های ٦ رقمی ساخته می‌شود. شماره‌ای به کمک آنها ساخته شده است، هریک از این احتمالات را محاسبه کنید:

الف) احتمال اینکه زوج باشد.

ب) احتمال اینکه بزرگتر از ۳۰۰ هزار باشد.

ج) احتمال اینکه بزرگتر از ۳۰۰ هزار یا کوچکتر از ۲۰۰ هزار باشد.

د) احتمال اینکه بر ۳ قابل قسمت باشد.

٦٨. یک فروشنده الماس در ظرفی ۲ الماس و یک شیشه، و در ظرف دیگر یک الماس و ۲ شیشه گذاشته است و برای اینکه به فرزند خود پاداش دهد به او می‌گوید یکی از دو ظرف را برگزیند. روشن است اگر فرزند برحسب تصادف ظرفی را برگزیند، احتمال به دست آوردن ۲ الماس $\frac{۱}{۲}$ است. اما پدر به فرزند خود می‌گوید قبل از برگزیدن ظرف می‌تواند از یکی از ۲ ظرف، قطعه‌ای را بردارد و ببیند آیا آن قطعه واقعاً الماس است یا نه و سپس ظرف را برگزیند. فرزند چنین فکر می‌کند که یک قطعه را از ظرفی برمی‌دارم، اگر الماس بود همان

ظرف و اگر شیشه بود ظرف دیگر را برمی‌دارم. احتمال اینکه فرزند هر ۲ الماس را به دست آورده باشد چقدر است؟

٦۹. شخصی پشت سرهم در ٤ امتحان شرکت می‌کند. احتمال قبول شدن او در امتحان اول P است و احتمال قبول شدن او در هریک از امتحانات بعدی، بسته به اینکه وی در امتحان قبلی قبول یا رد شود، به ترتیب P یا $\frac{P}{2}$ است. در صورتی که وی حداقل در ۳ امتحان قبول شود، واجد شرایط شناخته می‌شود. شانس وی برای اینکه واجد شرایط شناخته شود، چقدر است؟

۷۰. انفجار در یک کارگاه ساختمانی ممکن است در نتیجهٔ الکتریستهٔ ساکن، نقص در تجهیزات، بی‌احتیاطی یا خرابکاری رخ دهد. مهندسان ساختمان که مخاطرات موجود در هر مورد را تحلیل کرده‌اند، پی‌برده‌اند که چنین انفجاری ممکن است با احتمال ۲٥ درصد در نتیجهٔ الکتریستهٔ ساکن، ۲۰ درصد در نتیجهٔ نقص در تجهیزات، ٤۰ درصد در نتیجهٔ بی‌احتیاطی و ۱٥ درصد در نتیجه خرابکاری رخ دهد. احتمالهای پیشین این چهار علت به ترتیب ۲۰٪، ٤۰٪، ۲٥٪ و ۱٥٪ بوده‌اند. اگر انفجاری صورت گرفته باشد، هریک از این احتمالات را محاسبه کنید:

الف) محتملترین علت انفجار.

ب) احتمال اینکه علت، بی‌احتیاطی باشد.

پاسخنامهٔ سؤالات

۱) غ	۲) ص	۳) ص	٤) غ	٥) غ	٦) ص	۷) غ	۸) ص
۹) ص	۱۰) غ	۱۱) ص	۱۲) ص	۱۳) غ	۱٤) غ	۱٥) غ	۱٦) ص
۱۷) غ	۱۸) غ	۱۹) ص	۲۰) ص	۲۱) غ	۲۲) غ	۲۳) غ	۲٤) غ
۲٥) ص	۲٦) ب	۲۷) ب	۲۸) ب	۲۹) د	۳۰) د	۳۱) ب	۳۲) ب
۳۳) د	۳٤) الف	۳٥) ج	۳٦) ب	۳۷) ب	۳۸) ج	۳۹) الف	٤۰) ج
٤۱) ب	٤۲) د	٤۳) د	٤٤) ب	٤٥) ب	٤٦) ج	٤۷) الف	٤۸) د
٤۹) الف	٥۰) الف						

فصل ششم

توابع احتمال گسسته

۶-۱ متغیر تصادفی گسسته، تابع احتمال و تابع توزیع

۶-۱-۱ متغیر تصادفی گسسته

برای توصیف متغیر تصادفی ابتدا چند مثال می‌آوریم و سپس آن را تعریف می‌کنیم:

مثال ۶-۱ فرض می‌کنیم قرار است یک گروه بازیگر برای اجرای برنامهٔ خاصی به شهری برود. اگر هوای شهر آفتابی باشد، ۱۰ میلیون ریال سود می‌کند؛ اگر ابری باشد، سودش صفر است و اگر بارانی یا برفی باشد، ۵ میلیون ریال زیان می‌کند. ابتدا فضای نمونه را نشان می‌دهیم و سپس ارتباط آن را با سود این گروه مشخص خواهیم کرد.

فضای نمونه: {برفی، بارانی، ابری، آفتابی} = S

فضای نمونه	میزان سود
آفتابی ⟵	۱۰,۰۰۰,۰۰۰
ابری ⟵	۰
بارانی ⟵	−۵۰۰۰,۰۰۰
برفی ⟵	−۵۰۰۰,۰۰۰

مثال ۶-۲ فرض کنید سکهٔ سالمی را دو بار پرتاب می‌کنیم. ابتدا فضای نمونه را می‌نویسیم و سپس ارتباط آن را با تعداد شیرهای ظاهر شده مشخص می‌کنیم.

فضای نمونه: $S = \{TT, TH, HT, HH\}$

فضای نمونه	تعداد شیرها
TT ⟵	۰
TH ⟵	۱
HT ⟵	۱
HH ⟵	۲

مثال ۶-۳ فروش روزانهٔ دو بنگاه اتومبیل را در نظر می‌گیریم. تعداد اتومبیلهای

فروخته شده در بنگاه اول را با x (حداکثر دو اتومبیل) و تعداد اتومبیلهای فروخته شده در بنگاه دوم را با y (حداکثر یک اتومبیل) نشان می‌دهیم. ابتدا فضای نمونه را می‌نویسیم و سپس ارتباط آن را با تعداد اتومبیلهای فروخته شده نشان می‌دهیم.

هر عضوی از فضای نمونه را با (x , y) نشان داده‌ایم؛ بنابراین خواهیم داشت:

فضای نمونه: S = {(۰ , ۰)(۰ , ۱)(۱ , ۰)(۱ , ۱)(۲ , ۰)(۲ , ۱)}

(۰ , ۰) ⟶ ۰

(۰ , ۱) ⟶ ۱
(۱ , ۰) ⟶ ۱

(۱ , ۱) ⟶ ۲
(۲ , ۰) ⟶ ۲

(۲ , ۱) ⟶ ۳

فضای نمونه — تعداد اتومبیلهای فروخته شده در دوبنگاه

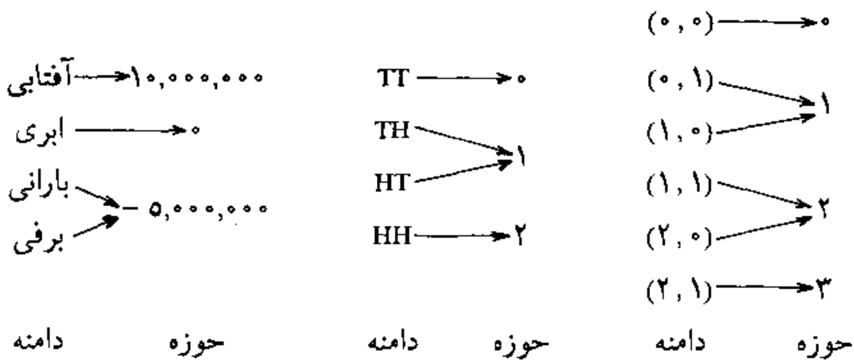
با توجه به مثالهای ذکر شده به این نتیجه می‌رسیم که ارتباط بین فضای نمونه و اعداد به وسیلهٔ متغیر تصادفی برقرار می‌شود.

متغیر تصادفی مقدار خود را به طور تصادفی می‌گیرد و این مقدار تحت کنترل مشاهده کننده نیست؛ مثلاً میزان سود گروه بازیگر به وضع آب و هوا بستگی دارد و یا تعداد شیرهای ظاهر شده در پرتاب دو تاس امری تصادفی است و همین طور تعداد اتومبیلهای فروخته شده به میزان تقاضا بستگی دارد.

مثالهای بسیار دیگری می‌توان برای متغیر تصادفی ذکر کرد؛ مثلاً تعداد ضایعات روزانهٔ یک کارگاه تولید کفش، تعداد مشتریانی که بین ساعت ۸ تا ۱۰ شب به رستورانی مراجعه می‌کنند، تعداد شماره‌های اشتباهی که یک تلفنچی در یک نوبت کاری می‌گیرد، تعداد نسخه‌هایی که از یک کتاب به فروش می‌رسد. بنابراین متغیر تصادفی در واقع، برخلاف اسمش، متغیر نیست بلکه تابعی است که روی فضای نمونه تعریف می‌شود و هریک از مقادیر آن، متناظر با یک یا چند عضو از اعضای فضای نمونه است. می‌دانیم که هر تابع دارای دامنه و حوزه است. بنابراین می‌توان متغیر تصادفی را این گونه تعریف

کرد: «متغیر تصادفی، تابعی است که دامنهٔ آن فضای نمونه و حوزهٔ آن مجموعه‌ای از اعداد حقیقی است».

دامنه و حوزهٔ مثالهای ۱ـ۶، ۲ـ۶ و ۳ـ۶ به این صورت است:



متغیر تصادفی را معمولاً برحسب تعداد مقادیری که می‌تواند اختیار کند به دو دستهٔ «متغیر تصادفی گسسته» و «متغیر تصادفی پیوسته» تقسیم می‌کنند. متغیر تصادفی گسسته متغیری است که تعداد مقادیری که می‌تواند اختیار کند، متناهی یا شمارش‌پذیر است؛ ولی تعداد مقادیر ممکنِ متغیر تصادفی پیوسته، نامتناهی است. در این فصل تنها متغیرهای تصادفی گسسته مطرح می‌شوند و در فصل بعد متغیرهای تصادفی پیوسته مطرح خواهند شد.

متغیرهای تصادفی با حروف بزرگ مانند X و هریک از مقادیری که این متغیرها می‌توانند بگیرند با حروف کوچک مانند x، نشان داده می‌شوند.

۲ـ۱ـ۶ تابع اِحتمال

به تابعی که بتوان با استفاده از آن احتمال هریک از مقادیر ممکنِ متغیر تصادفی را مشخص کرد «تابع احتمال» یا «توزیع احتمال» گفته می‌شود. تابع احتمال را می‌توان چنین تعریف کرد: «تابع احتمال تابعی است که دامنهٔ آن مقادیر ممکن متغیر تصادفی و حوزهٔ آن احتمالات مربوط به هر مقدار متغیر تصادفی است». توجه داشته باشید که تابع احتمال غیرمنفی است و مجموع احتمالات باید برابر یک باشد.

مثال ۴ـ۶ باتوجه به داده‌های مثال ۱ـ۶ می‌خواهیم بدانیم اگر احتمال آفتابی، ابری،

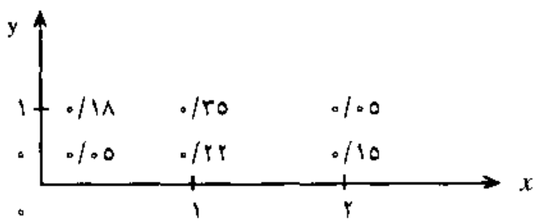
بارانی و برفی بودن به ترتیب ۰/٤۰، ۰/۳۰، ۰/۲۰ و ۰/۱۰ باشد، تابع احتمال میزان سود این گروه چقدر خواهد بود.

(میزان سود به ریال) x	۱۰,۰۰۰,۰۰۰	۰	۵۰۰۰,۰۰۰ −	جمع
$P(X=x)$ (احتمال)	۰/٤۰	۰/۳۰	۰/۳۰	۱

مثال ٦-۵ تابع احتمال مثال ۲-٦ را مشخص می‌کنیم:

x	۰	۱	۲	جمع
$P(X=x)$	۰/۲۵	۰/۵۰	۰/۲۵	۱

مثال ٦-٦ با توجه به داده‌های مثال ۳-٦ و احتمالات فضای نمونه که در این نمودار آمده است، می‌خواهیم تابع احتمال مربوط به تعداد اتومبیلهای فروخته شده در یک روز در دو بنگاه را مشخص کنیم.



اگر z را تعداد اتومبیلهای فروخته شده در دو بنگاه فرض کنیم، z متغیری تصادفی با این تابع احتمال خواهد بود:

z	۰	۱	۲	۳	جمع
$P(Z=z)$	۰/۰۵	۰/٤۰	۰/۵۰	۰/۰۵	۱

مثال ٦-۷ می‌خواهیم بدانیم آیا این تابع می‌تواند تابع احتمال باشد. چرا.

$$P(X=x)=\frac{\binom{٤}{x}}{۱٦} \quad , \quad x = ۰,۱,۲,۳,٤$$

بله. چون اولاً به ازای مقادیر ممکن x، $P(X=x)$ غیر منفی است و ثانیاً جمع احتمالات برابر یک است.

$$\sum_{x=0}^{4} P(X=x) = \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{16} + \frac{1}{16} = 1$$

۶-۱-۳ تابع توزیع (تابع احتمال تجمعی)

در مسائلی ممکن است دانستن احتمال اینکه متغیر تصادفی X مقداری کوچکتر یا مساوی x را انتخاب می‌کند، مورد توجه باشد، در این صورت تابع توزیع (تابع احتمال تجمعی) را به کار می‌بریم. تعریف این تابع چنین است: «تابع توزیع، تابعی است که به ازای جمیع مقادیر ممکن متغیر تصادفی X، احتمال وقوع مقداری کوچکتر یا مساوی با x را نشان می‌دهد». تابع توزیع را با $P(X \le x)$ یا $F(x)$ نشان می‌دهند. بدیهی است که تابع توزیع برای بیشترین مقداری که متغیر تصادفی می‌تواند اختیار کند برابر یک است.

مثال ۶-۸ این جدول تابع احتمال را در نظر بگیرید. می‌خواهیم تابع توزیع مربوط به آن را پیدا کنیم و مقدار $F(1)$ و $F(3)$ را به دست آوریم.

x	-2	0	3	5	10
$P(X=x)$	۰/۱۰	۰/۱۵	۰/۲۵	۰/۳۰	۰/۲۰

x	-2	0	3	5	10
$F(x)=P(X \le x)$	۰/۱۰	۰/۲۵	۰/۵۰	۰/۸۰	۱

بنابراین $F(1) = P(X \le 1) = 0/25$ و $F(3) = P(X \le 3) = 0/50$ است.

مثال ۶-۹ تعداد تلویزیونهای فروخته شدهٔ فروشگاهی در ۱۲۰ روز به این شرح است:

تعداد روزها	تعداد تلویزیونهای فروخته شده
۱۸	۲
۲۷	۳
۳۰	٤
۱۵	٥
۱۸	٦
۱۲	۷
۱۲۰	جمع

۱۰

الف) تابع احتمال را تعیین کنید.
ب) تابع توزیع را تعیین کنید.
ج) احتمال اینکه در یک روز کمتر و یا مساوی پنج تلویزیون فروخته شود، چقدر است؟

الف)

x (تعداد فروش)	۲	۳	٤	٥	٦	۷
$P(X=x)$	۰/۱۵۰	۰/۲۲٥	۰/۲٥۰	۰/۱۲٥	۰/۱٥۰	۰/۱۰۰

ب)

x	۲	۳	٤	٥	٦	۷
$F(x) = P(X \le x)$	۰/۱۵۰	۰/۳۷٥	۰/٦۲٥	۰/۷٥۰	۰/۹۰۰	۱

ج) $F(٥) = P(X \le ٥) = ۰/۷٥۰$

تمرین

۱. فرض کنید متغیر تصادفی X بتواند مقادیر ۲ ، ۳، ٤ و ٥ را بگیرد. آیا احتمالات ذیل می‌توانند تابع احتمال این متغیر تصادفی باشند؟ چرا؟

x	۲	۳	٤	٥
$P(X=x)$	۰/۲٥	۰	$-$۰/۱۰	۰/۸٥

۲. آیا این تابع می‌تواند به عنوان تابع احتمال یک متغیر تصادفی به کار رود؟ چرا؟

$$f(x) = \frac{x+1}{20} \quad و \quad x = 1,2,3,4,5$$

۳. در این تابع K را چنان تعیین کنید که بتوان آن را به عنوان تابع احتمال به کار برد.

$$f(x) = Kx^2 \quad و \quad x = 0\,,\,1,2,3,4,5$$

۴. از جعبه‌ای که محتوی ۱۲ کالاست و ۴ عدد آنها معیوب است ۳ تا را به صورت تصادفی انتخاب می‌کنیم. اگر X تعداد کالای معیوب خارج شده باشد، تابع احتمال X را پیدا کنید.

۵. شرکت بیمه این اطلاعات را که مربوط به تعداد تصادفات ۶۰ روز یکی از مناطق شهری است، جمع‌آوری کرده است:

تعداد روزها	تعداد تصادفات
۱۵	۰
۲۵	۱
۱۰	۲
۶	۳
۴	۴
۶۰	جمع

الف) تابع احتمال و تابع توزیع تعداد تصادفات را پیدا کنید.

ب) اگر روزی از این ۶۰ روز را بر حسب تصادف انتخاب کنیم، احتمال اینکه در این روز کمتر از ۴ تصادف رخ داده باشد، چقدر است؟

۶-۲ امید ریاضی و واریانس متغیر تصادفی

۱-۲-۶ امید ریاضی

امید ریاضی مفهومی اساسی در مطالعهٔ توابع احتمال است. امید ریاضی همان میانگین موزون است که احتمالات در آن، نقش وزنها (ضرایب) را بازی می‌کنند. امید ریاضی متغیر تصادفی X را با $E(X)$ نشان می‌دهیم. در میانگین موزون، داده‌ها را در ضرایب مربوط به آنها ضرب و سپس حاصل را بر مجموع ضرایب تقسیم می‌کنیم، ولی در امید ریاضی، چون مجموع احتمالات برابر یک است، پس از ضرب احتمالات (ضرایب) در متغیر تصادفی، دیگر نیازی به تقسیم آنها بر جمع احتمالات نیست. امید ریاضی متغیر

تصادفی گسستهٔ X به این صورت محاسبه می‌شود (($f(x)$ همان $P(X=x)$ است):

$$E(X)=\sum_{x} x.f(x)$$

مثال ۱۰ـ۶ شرکتی تولید کنندهٔ آبگرمکن گازی است. تقاضاهای ماهانه همراه با احتمالات مربوط در این جدول آورده شده است. می‌خواهیم امید ریاضی تعداد تقاضا را پیدا کرده، آن را تعبیر کنیم.

x (تقاضا)	۱۰۰	۲۰۰	۳۰۰	۴۰۰	۵۰۰
$f(x)$ (احتمال)	۰/۱۵	۰/۲۵	۰/۳۰	۰/۲۰	۰/۱۰

x	۱۰۰	۲۰۰	۳۰۰	۴۰۰	۵۰۰	جمع
$f(x)$	۰/۱۵	۰/۲۵	۰/۳۰	۰/۲۰	۰/۱۰	۱
$x.f(x)$	۱۵	۵۰	۹۰	۸۰	۵۰	۲۸۵

بنابراین $E(X)=285$ است؛ یعنی این شرکت انتظار دارد تقاضا برای کالایش ۲۸۵ واحد در ماه باشد.

با توجه به تعریف امید ریاضی متغیر تصادفی، اگر a و b مقادیر ثابتی فرض شوند، دارای این خواص خواهند بود:

$$E(b)=b$$

$$E(aX)=aE(X)$$

$$E(aX+b)=aE(X)+b$$

مثال ۱۱ـ۶ تابع احتمال ذیل مفروض است:

x	-2	۱	۳	۵
$f(x)$	۰/۲	۰/۴	۰/۳	۰/۱

می‌خواهیم این مقادیر را محاسبه کنیم:

الف) $E(X)$، ب) $E(2X)$، ج) $E(-3X+2)$، د) $E(X^2)$، هـ) $E[(X-5)^2]$

x	-۲	۱	۳	۵	جمع
$f(x)$	۰/۲	۰/٤	۰/۳	۰/۱	۱
$x f(x)$	$-۰/٤$	۰/٤	۰/۹	۰/۵	۱/٤
$x^{۲} f(x)$	۰/۸	۰/٤	۲/۷	۲/۵	٦/٤

بنابراین خواهیم داشت:

الف) $E(X) = \sum_x x.f(x) = ۱/٤$

ب) $E(۲X) = ۲E(X) = ۲ \times ۱/٤ = ۲/۸$

ج) $E(-۳X + ۲) = -۳E(X) + ۲ = -۳ \times ۱/٤ + ۲ = -۲/۲$.

د) $E(X^{۲}) = \sum_x x^{۲}.f(x) = ٦/٤$

ه) $E[(X-۵)^{۲}] = E(X^{۲} - ۱۰X + ۲۵) = E(X^{۲}) - ۱۰E(X) + ۲۵$

$$= ٦/٤ - ۱۰ \times ۱/٤ + ۲۵ = ۱۷/٤$$

۲-۲-٦ واریانس

واریانسِ متغیر تصادفی گسستهٔ X که میزان پراکندگی را حول میانگین (امید ریاضی) نشان می‌دهد و آن را با $V(X)$ نشان می‌دهیم، به این صورت محاسبه می‌شود:

$$V(X) = \sum_x (x-\mu)^{۲} f(x)$$

در این رابطه، $\mu = E(X)$است. می‌توان واریانس را به این صورت نیز محاسبه کرد:

$$V(X) = E(X^{۲}) - \mu^{۲}$$

در هر صورت، انحراف معیار به این صورت محاسبه می‌شود:

$$SD(X) = \sqrt{V(X)}$$

اگر a و b مقادیر ثابتی فرض شوند، این خواص برای واریانس متغیر تصادفی قابل اثبات است:

$$V(b) = ۰$$
$$V(aX) = a^{۲} V(X)$$
$$V(aX + b) = a^{۲} V(X)$$

مثال ۱۲ـ۱۶ اگر $E(X) = ۲/۵$ و $E(X^{۲}) = ۸$ باشد، می‌خواهیم این مقادیر را محاسبه کنیم:

الف) $V(۳)$؛ ب) $V(X)$؛ ج) $V(-۲X + ۳)$.

الف) $V(۳) = ۰$

ب) $V(X) = E(X^{۲}) - \mu^{۲} = ۸ - (۲/۵)^{۲} = ۱/۷۵$

ج) $V(-۲X + ۳) = (-۲)^{۲} V(X) = ٤ \times ۱/۷۵ = ۷$

تمرین

۱. این توزیع احتمال مفروض است:

x	۲	٤	٦	۸	۱۰
$f(x)$	۰/۱۰	۰/۲۵	۰/۳۰	۰/۲۵	۰/۱۰

این مقادیر را محاسبه کنید:

الف) $E(X)$

ب) $E(-۳X + ۷)$

ج) $E[(-۳x + ۷)^{۲}]$

د) $V(X)$

ه) $V(-٤x - ۳)$

۲. این جدول احتمال تعداد غیبتهای روزانهٔ کارگاهی را نشان می‌دهد که دارای ۵۰ کارگر است:

تعداد غیبتها	۰	۱	۲	۳	٤	۵	٦
احتمال	۰/۱۵	۰/۲۰	۰/۳۰	۰/۱۸	۰/۱۰	۰/۰۵	۰/۰۲

الف) میانگین و واریانس تعداد غیبتهای روزانه را پیدا کنید.
ب) احتمال اینکه در روزی بیشتر از ۲ نفر غیبت کنند چقدر است؟

۳. تولیدکنندهٔ تقویمی مایل است برای تعداد تیراژ خود در سال آینده تصمیم بگیرد. تولید هر تقویم ۹۰۰ ریال هزینه دربردارد و به بهای ۱۲۵۰ ریال فروخته می‌شود. هر تقویمی که در طول سال به فروش نرسد از آن به عنوان کاغذ باطله استفاده می‌شود که ارزش چندانی ندارد. این تولیدکننده، با استفاده از اطلاعات سالهای قبل، تابع احتمال ذیل را برای فروش خود تهیه کرده است:

تعداد فروش	۲۵,۰۰۰	٤۰,۰۰۰	۵۵,۰۰۰	۷۰,۰۰۰
احتمال	۰/۱۰	۰/۳۰	۰/٤۵	۰/۱۵

این تولیدکننده تصمیم دارد یکی از سطوح ۲۵، ٤۰ ، ۵۵ یا ۷۰ هزار واحدی را تولید کند. چه سطحی از تولید کل سود مورد انتظار را به حداکثر می‌رساند؟

٤. شرکت بیمه‌ای ۲ هزار کارخانه را در قبال آتش‌سوزی بیمه می‌کند. چنانچه این شرکت بابت هر کارخانه‌ای که دچار آتش‌سوزی شود مبلغ ۵۰ میلیون ریال بپردازد، به نقطهٔ سر به سر می‌رسد. اگر احتمال آتش‌سوزی کارخانه‌ای در مدت بیمه‌نامه، برابر ۰/۰٦ باشد، با این فرض که هر کارخانه فقط یک بار ممکن است دچار آتش‌سوزی شود، این موارد را محاسبه کنید:
الف) مبلغ بیمه باید چقدر باشد؟
ب) اگر احتمال آتش‌سوزی در این مسأله برابر ۰/۰٤ باشد، امید ریاضی سود شرکت با همان مبلغ تعیین شده در بند الف چقدر خواهد بود؟

۵. تولیدکننده‌ای کالای تولیدی خود را در بسته‌های ۱۰ تایی به بازار عرضه می‌کند. تابع احتمال تعداد کالاهای معیوب یک بسته به این صورت است (در یک بسته حداکثر ۳ کالای معیوب وجود دارد):

تعداد کالاهای معیوب	۰	۱	۲	۳
احتمال	۰/۷۵	۰/۱۵	۰/۰۷	۰/۰۳

این تولید کننده ۲ قیمت مختلف ۲۵ هزار ریالی و ۲٦ هزار ریالی را برای هر بسته پیشنهاد می‌کند؛ به این ترتیب که اگر مشتری از نرخ ۲۵ هزار ریالی استفاده کند چیزی در قبال ارائه کالای معیوب دریافت نمی‌کند، ولی اگر از نرخ ۲٦ هزار ریالی استفاده کند، در ازای ارائه هر کالای معیوب، ۲ هزار ریال دریافت می‌کند. شما به عنوان خریدار کدام پیشنهاد را قبول می‌کنید.

می‌خواهیم ابتدا جدول تابع احتمال توأم فروشهای روزانهٔ دو بنگاه اتومبیل را تنظیم و سپس هریک از این موارد را محاسبه کنیم:

الف) $P(X=۱)$

ب) $P(X>Y)$

ج) $P(Y \geq ۰$ و $X=۱)$

د) توزیع احتمال $Z=X+Y$

x \ y	۰	۱
۰	۰/۰۵	۰/۱۸
۱	۰/۲۲	۰/۳۵
۲	۰/۱۵	۰/۰۵

الف) $P(X=۱)=P(X=۱, Y=۰)+(X=۱, Y=۱)=۰/۲۲+۰/۳۵=۰/۵۷$

ب) $P(X>Y)=P(X=۱, Y=۰)+P(X=۲, Y=۰)+P(X=۲, Y=۱)$

$=۰/۲۲+۰/۱۵+۰/۰۵=۰/۴۲$

ج) $P(Y\geq ۰$ و $X=۱)=P(Y=۰, X=۱)+P(Y=۱, X=۱)=۰/۲۲+۰/۳۵=۰/۵۷$

د)

z	۰	۱	۲	۳
f(z)	۰/۰۵	۰/۴۰	۰/۵۰	۰/۰۵

با در دست داشتن تابع احتمال توأم، می‌توان تابع احتمال جداگانه هر یک از متغیرهای تصادفی را پیدا کرد؛ مثلاً در مثال ۱۳ـ۶ برای پیدا کردن تابع احتمال متغیر تصادفی X، می‌توان احتمالات هر سطر را جمع کرد. و آن را در حاشیهٔ سمت راست جدول نوشت و یا برای پیدا کردن تابع احتمال متغیر تصادفی Y، می‌توان احتمالات هر ستون را جمع کرد و آن را در حاشیهٔ پایین جدول نوشت. به این احتمالات، «احتمالات حاشیه‌ای»، به احتمالات حاشیهٔ راست جدول، احتمالات حاشیه‌ای X و به احتمالات حاشیهٔ پایین جدول، احتمالات حاشیه‌ای Y می‌گوییم. حتی می‌توان از ذکر کلمهٔ حاشیه‌ای نیز پرهیز کرد و آنها را تابع احتمال متغیر تصادفی X و تابع احتمال متغیر تصادفی Y نامید.

مثال ۱۴ـ۶ می‌خواهیم احتمالات حاشیه‌ای مثال ۱۳ـ۶ را پیدا کنیم و تابع احتمال X و تابع احتمال Y را در دو جدول مجزا بنویسیم.

x \ y	۰	۱	احتمالات حاشیه‌ای x
۰	۰/۰۵	۰/۱۸	۰/۲۳
۱	۰/۲۲	۰/۳۵	۰/۵۷
۲	۰/۱۵	۰/۰۵	۰/۲۰
احتمالات حاشیه‌ای y	۰/۴۲	۰/۵۸	۱

بنابراین خواهیم داشت:

x	۰	۱	۲
$f(x)$	۰/۲۳	۰/۵۷	۰/۲۰

و

y	۰	۱
$f(y)$	۰/۴۲	۰/۵۸

مثال ۱۵ـ۶ در جعبه‌ای ۸ بلبرینگ وجود دارد که ۲ تای آنها معیوب است. به طور تصادفی ۳ تا از این بلبرینگها را بیرون آورده‌ایم (X را تعداد بلبرینگهای سالم و Y را تعداد بلبرینگهای معیوب فرض می‌کنیم) می‌خواهیم تابع احتمال توأم آنها را پیدا کنیم.

X می‌تواند مقادیر ۳،۲،۱ و Y مقادیر ۰،۱،۲ را بگیرد. چون X و Y هریک می‌توانند ۳ مقدار انتخاب کنند؛ بنابراین ۹ زوج مختلف (۹=۳×۳) به وجود خواهد آمد که باید احتمال هرکدام را پیدا کنیم؛ برای نمونه $f(۱,۰) = P(X=۱, Y=۰) = ۰$ است؛ چون ما ۳ کالا انتخاب کرده‌ایم در صورتی که ۱=۱+۰ است؛ بنابراین مواردی را محاسبه می‌کنیم که مجموع مقادیر متغیرها در آنها برابر ۳ شود؛ برای مثال:

$$f(۲,۱) = P(X=۲, Y=۱) = \binom{۶}{۲}\binom{۲}{۱}/\binom{۸}{۳} = \frac{۳۰}{۵۶}$$

x \ y	۰	۱	۲
۱	۰	۰	$\frac{۶}{۵۶}$
۲	۰	$\frac{۳۰}{۵۶}$	۰
۳	$\frac{۲۰}{۵۶}$	۰	۰

تمرین

۱. این توزیع احتمال توأم را در نظر بگیرید:

x \ y	−۱	۰	۳
۲	۰/۱۰	۰/٤۰	۰
٤	۰/۱۵	۰/۲۰	۰/۱۵

این موارد را محاسبه کنید:

الف) $P(X = ۲)$

ب) $P(X = ٦)$

ج) $P(X > Y) + P(Y > X)$

د) $P(Z = X + Y)$

۲. فرض کنید از ۱۰ شرکت بزرگ مربوط به صنعتی خاص، تنها ۳ شرکت دارای سود خالصی بیش‌از یک میلیارد و پانصد میلیون ریال بوده‌اند. از این ۱۰ شرکت ۲ شرکت بر حسب تصادف انتخاب شده است. اگر متغیر تصادفی X نشان‌دهنده تعداد شرکتهایی که سودی بیش از یک میلیارد و ۵۰۰ میلیون ریال و Y نشان دهنده تعداد شرکتهایی باشد که سودی کمتر یا مساوی این مبلغ داشته‌اند ، هر یک از این موارد را محاسبه کنید:

الف) تابع احتمال توأم X و Y

ب) احتمـــال اینکه دو شـرکت انتخـاب شـده، سـودی بـیشتر از یک میلیـارد و ۵۰۰ میلیون ریال داشته باشند.

ج) اگر بدانیم کـمتر از ۲ شـرکت بـا سـود بـیش از یک میلیـارد و ۵۰۰ مـیلیون ریـال انتخاب شده‌اند، احتمال اینکه دقیقاً یک شرکت با سود کمتر یا مساوی این مـبلغ انتخـاب شده باشد.

د) احتمال اینکه تعداد شرکتهای با سود بیش از یک میلیارد و ۵۰۰ میلیون ریال بیشتر از تعداد شرکتهای با سود کمتر یا مساوی این مبلغ باشد.

ه) تابع احتمال متغیر تصادفی X

٦ـ٤ کوواریانس و استقلال دو متغیر تصادفی

٦ـ٤ـ١ کوواریانس

کوواریانس را امید ریاضی تغییرات دو متغیر برحسب میانگینشان تعریف می‌کنیم. در اینجا، کوواریانس معیاری عددی است که نوع و شدت رابطهٔ خطی بین دو متغیر تصادفی را نشان می‌دهد. رابطهٔ دو متغیر می‌تواند به یکی از این سه حالت باشد: ۱. با افزایش یک متغیر، دیگری افزایش یابد و با کاهش آن دیگری کاهش یابد (رابطهٔ مستقیم)، ۲. با افزایش یک متغیر، دیگری کاهش یابد و با کاهش آن افزایش یابد (رابطهٔ معکوس) ، ۳. افزایش و یا کاهش یک متغیر هیچ تأثیری در دیگری نداشته باشد (دو متغیر مستقل باشند). مقدار کوواریانس در حالت اول مثبت، در حالت دوم منفی و در حالت سوم صفر است.

کوواریانس بین دو متغیر تصادفی X و Y را با $\mathrm{Cov}(X, Y)$ نشان می‌دهیم و به این صورت تعریف می‌کنیم:

$$\mathrm{Cov}(X, Y) = E(X - \mu_x)(Y - \mu_y)$$

و یا

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

که در این رابطه، $\mu_x = E(X)$، $\mu_y = E(Y)$ و اگر X و Y گسسته باشند، $E(XY) = \sum_i \sum_j x_i y_j f(x_i, y_j)$ است.

مثال ٦ـ١٦ تابع احتمال توأم دو متغیر تصادفی X و Y به این صورت است. کوواریانس را حساب کرده، دربارهٔ ارتباط این دو متغیر توضیح می‌دهیم:

x \ y	۰	۱	۲
−۵	۰/۳۰	۰/۰۲	۰/۰۸
۱۰	۰/۱۰	۰	۰/۵۰

برای محاسبه $\mathrm{Cov}(X, Y)$ ابتدا باید $E(X)$، $E(Y)$ و $E(XY)$ را حساب کنیم.

y	۰	۱	۲	جمع
f(y)	۰/٤۰	۰/۰۲	۰/۵۸	۱
yf(y)	۰	۰/۰۲	۱/۱٦	۱/۱۸

x	−۵	۱۰	جمع
f(x)	۰/٤۰	۰/٦۰	۱
xf(x)	ـ۲	٦	٤

پس $E(X)=4$ و $E(Y)=1/18$ است.

$$E(XY)=\sum_i\sum_j x_i y_j f(x_i,y_j)=(-5)(0)(0/30)+(-5)(1)(0/02)+(-5)(2)(0/08)$$
$$+(10)(0)(0/10)+(10)(1)(0)+(10)(2)(0/50)=9/1$$

بنابراین:

$$Cov(X,Y)=E(XY)-E(X)E(Y)=9/1-4\times 1/18=4/38$$

چون مقدار کوواریانس مثبت است، از این رو با افزایش X، Y نیز افزایش می‌یابد.

۶-۴-۲ استقلال دو متغیر تصادفی

درفصل پنجم گفته‌شدکه دو پیشامد A وB درصورتی مستقلندکه‌رابطهٔ $P(A\cap B)=P(A)P(B)$ بین آنها برقرار باشد. به طریق همانند می‌توان گفت که دو متغیر تصادفی X و Y در صورتی مستقلند که به ازای تمام زوجهای (x_i,y_j) این رابطه: $f(x_i,y_j)=f(x_i).f(y_j)$، بین آنها برقرار باشد.

مثال ۶-۱۷ این تابع احتمال توأم را در نظر گرفته، بررسی می‌کنیم که آیا دو متغیر تصادفی X و Y مستقلند یا نه.

$x \backslash y$	۲	۴	۶
۰	۰/۲۰	۰/۰۸	۰/۱۲
۱	۰/۳۰	۰/۱۲	۰/۱۸

شرط استقلال $f(x_i,y_j)=f(x_i).f(y_j)$ را برای تمام i و jها بررسی می‌کنیم:

$x \backslash y$	۲	۴	۶	$f(x)$
۰	$0/20=0/5\times 0/4$	$0/08=0/2\times 0/4$	$0/12=0/3\times 0/4$	۰/۴
۱	$0/30=0/5\times 0/6$	$0/12=0/2\times 0/6$	$0/18=0/3\times 0/6$	۰/۶
$f(y)$	۰/۵	۰/۲	۰/۳	

چون تمام تساویهای مندرج در جدول برقرار است، پس دو متغیر تصادفی X و Y مستقلند.

پیش از این گفتیم که $\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)$. اگر دو متغیر تصادفی X و Y مستقل باشند، آنگاه $E(XY)=E(X)E(Y)$ است؛ زیرا:

$$
\begin{aligned}
E(XY) &= \sum_i \sum_j x_i y_j f(x_i, y_j) \\
&= \sum_i \sum_j x_i y_j f(x_i) f(y_j) \\
&= \sum_i x_i f(x_i) \sum_j y_j f(y_j) \\
&= E(X)E(Y)
\end{aligned}
$$

و بنابراین کوواریانس صفر می‌شود:

$$
\begin{aligned}
\mathrm{Cov}(X,Y) &= E(XY) - E(X)E(Y) \\
&= E(X)E(Y) - E(X)E(Y) \\
&= 0
\end{aligned}
$$

در مورد دو متغیر مستقل، کوواریانس لزوماً صفر است، ولی عکس این مطلب همواره درست نیست؛ یعنی دو متغیر تصادفی ممکن است مستقل نباشند، ولی کوواریانس آنها صفر باشد. به عبارت دیگر:

$$\mathrm{Cov}(X,Y)=0 \;\;\underset{\not\Longrightarrow}{\Longleftarrow}\;\; \text{X و Y مستقلند}$$

مثال ۱۸ـ۶ می‌خواهیم کوواریانس را برای تابع احتمال توأم مثال ۱۷ـ۶ حساب کنیم.

y	۲	۴	۶	جمع
$f(y)$	۰/۵	۰/۲	۰/۳	۱
$yf(y)$	۱	۰/۸	۱/۸	۳/۶

x	۰	۱	جمع
$f(x)$	۰/۴	۰/۶	۱
$xf(x)$	۰	۰/۶	۰/۶

پس $E(X) = 0/6$ و $E(Y) = 3/6$ است. حال $E(XY)$ را حساب می‌کنیم:

$$E(XY) = \sum_i \sum_j x_i y_j f(x_i, y_j) = (0)(2)(0/20) + (0)(4)(0/08) + (0)(6)(0/12)$$
$$+ (1)(2)(0/30) + (1)(4)(0/12) + (1)(6)(0/18)$$
$$= 2/16$$

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 2/16 - (0/6)(3/6) = 2/16 - 2/16 = 0$$

پیش از این در مثال ۶ـ۱۷ مشخص کردیم که X و Y مستقل هستند، حتی بدون محاسبهٔ کوواریانس می‌توانستیم ادعا کنیم که مقدار کوواریانس صفر است، چون اگر دو متغیر تصادفی، مستقل و دارای حوزهٔ محدود باشند، حتماً کوواریانس آنها صفر خواهد بود.

مثال ۶ـ۱۹ دو متغیر تصادفی X و Y را با توزیع توأم ذیل در نظر بگیرید. می‌خواهیم نشان دهیم با اینکه دو متغیر مستقل نیستند، کوواریانس آنها صفر است.

$x \backslash y$	-1	0	1
-1	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$
1	$\frac{1}{6}$	0	$\frac{1}{6}$

چون $(\frac{1}{3})(\frac{2}{3}) \neq \frac{1}{6}$ است پس X و Y مستقل نیستند، ولی طبق این محاسبات، کوواریانس آنها صفر است:

x	-1	1	جمع
$f(x)$	$\frac{2}{3}$	$\frac{1}{3}$	1
$xf(x)$	$-\frac{2}{3}$	$\frac{1}{3}$	$-\frac{1}{3}$

y	-1	0	1	جمع
$f(y)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1
$yf(y)$	$-\frac{1}{3}$	0	$\frac{1}{3}$	0

بنابراین $E(X) = -\frac{1}{3}$ و $E(Y) = 0$ خواهد بود. حال $E(XY)$ را محاسبه می‌کنیم:

$$E(XY)=(-۱)(-۱)(\frac{۱}{۶})+(-۱)(۰)(\frac{۱}{۳})+(-۱)(۱)(\frac{۱}{۶})+(۱)(-۱)(\frac{۱}{۶})$$
$$+(۱)(۰)(۰)+(۱)(۱)(\frac{۱}{۶})=۰$$

$$Cov(X,Y)=E(XY)-E(X)E(Y)=۰-(-\frac{۱}{۳})(۰)=۰$$

در اینجا چند قاعدهٔ مهم دربارهٔ امید ریاضی و واریانس مجموع دو متغیر تصادفی در حالت کلی و در حالتی که دو متغیر مستقل باشند، آورده می‌شود:

$$E(X+Y)=E(X)+E(Y)$$

$$V(X+Y)=V(X)+V(Y)+۲\,Cov(X,Y)$$

$$V(X-Y)=V(X)+V(Y)-۲\,Cov(X,Y)$$

اگر X و Y مستقل باشند، خواهیم داشت:

$$E(X+Y)=E(X)+E(Y)$$

$$V(X+Y)=V(X)+V(Y)$$

$$V(X-Y)=V(X)+V(Y)$$

این قواعد را می‌توان به راحتی اثبات کرد.

تمرین

۱. توزیع توأم ذیل را در نظر بگیرید:

x \ y	−٤	۲	۷
۱	۰	۰/۱	۰/٤
۵	۰/٤	۰/۱	۰

الف) کوواریانس را حساب کنید.

ب) رابطه X و Y را تحلیل کنید.

۲. در تمرین ۱ آیا X و Y مستقلند؟

۳. توزیع توأم ذیل را در نظر بگیرید:

$x \backslash y$	۳	٤	٥
۲−	۰/۲۰	۰/۰۸	۰/۱۲
٦	۰/۱٥	۰/۰٦	۰/۰۹
۸	۰/۱٥	۰/۰٦	۰/۰۹

الف) آیا دو متغیر X و Y مستقلند؟

ب) کوواریانس X و Y را حساب کنید.

٤. توزیع توأم X و Y ذیل را در نظر بگیرید:

$x \backslash y$	۲−	۱۰
۰	۰/۳۳	۰/۱۷
٤	۰/۲۷	۰/۲۳

کوواریانس X و Y را حساب کرده و راجع به استقلال دو متغیر توضیح دهید.

٦ـ٥ توزیع برنولی

آزمایشهایی را در نظر بگیرید که دارای دو پیامد ممکن باشند و احتمال «موفقیت» و «شکست» هر پیامدی (در اینجا احتمال موفقیت یعنی احتمال وقوع پیشامد مورد نظر و احتمال شکست یعنی احتمال عدم وقوع پیشامد مورد نظر) از آزمایشی به آزمایش دیگر «ثابت» باشد و ضمناً آزمایشها «مستقل» از یکدیگر انجام شوند. در این صورت به هریک از آزمایشهای مزبور یک «آزمایش برنولی» و توزیع تعداد موفقیتها (۰ یا ۱) را توزیع برنولی گویند. احتمال موفقیت در یک آزمایش را با P و احتمال عدم موفقیت را با $1 - P$ یا q نشان می‌دهیم. واضح است که چون این دواحتمال مکمل هم هستند $P + q = 1$ خواهد بود.

مثال ۲۰ـ٦ در جعبه‌ای ۲۰ کالا وجود دارد که ٥ تای آنها نامرغوب است. اگر

بخواهیم با جایگذاری، چند کالا را انتخاب کنیم، می‌خواهیم بدانیم احتمال خارج کردن یک کالای مرغوب (احتمال موفقیت) در هر بار چقدر است.

چون آزمایشها مستقل از هم صورت می‌گیرد و در هر بار، تعداد کالاهای مرغوب و نامرغوب ثابت است (۵ به ۲۰)؛ از این رو احتمال موفقیت در هر بار $\frac{۱۵}{۲۰}$ یا ۰/۷۵ است؛ یعنی:

در هر آزمایش	
احتمال موفقیت (p)	۰/۷۵
احتمال عدم موفقیت $(q = ۱ - p)$	۰/۲۵

مثال ۶-۲۱ براساس آمار متولدین، مشخص شده است که از هر ۱۰۰ نوزاد به طور متوسط ۴۸ نفر دختر و بقیه پسرند. می‌خواهیم بدانیم آیا این آزمایش، آزمایش برنولی محسوب می‌شود.

بله، اگر احتمال دختر بودن نوزاد را احتمال موفقیت بنامیم، آنگاه برای هر نوزاد احتمال موفقیت ۰/۴۸ و احتمال شکست ۰/۵۲ است.

مثال ۶-۲۲ مثال ۶-۲۰ را در نظر می‌گیریم. اگر بخواهیم کالایی را بدون جایگزینی انتخاب کنیم، می‌خواهیم بدانیم آیا این آزمایش، آزمایش برنولی محسوب می‌شود.

خیر، چون در بار اول احتمال موفقیت (انتخاب کالای مرغوب) $\frac{۱۵}{۲۰}$ است، ولی در دفعهٔ دوم بسته به اینکه بار اول کالای مرغوب انتخاب شده یا نشده باشد، احتمال موفقیت به ترتیب برابر $\frac{۱۴}{۱۹}$ و $\frac{۱۵}{۱۹}$ است و به همین دلیل که احتمال موفقیت از آزمایشی به آزمایش دیگر ثابت نیست، این آزمایش، آزمایش برنولی محسوب نمی‌شود.

۶-۵-۱ نمونه گیری بدون جایگزینی از جامعهٔ بزرگ

در جامعه‌ای که تعداد اعضای آن و نیز تعداد افرادی که واجد شرایط خاصی هستند نیز بسیار زیاد است، احتمال موفقیت (احتمال انتخاب فردی که واجد شرایط باشد) در نمونه گیریهای بدون جایگزینی نسبتاً ثابت است ـ گرچه دقیقاً ثابت نیست ـ و می‌توان فرض کرد که این آزمایشها، آزمایش برنولی هستند.

مثال ۶-۲۳ از بین ۵ هزار مشتری بانکی ۲ هزار نفر در حسابهای کوتاه مدت

سرمایه‌گذاری کرده‌اند. اگر برحسب تصادف چند نفر از مشتریان این بانک را انتخاب کنیم، می‌خواهیم بدانیم احتمال اینکه هریک در حسابهای کوتاه مدت سرمایه‌گذاری کرده‌باشند، چقدر است و اینکه آیا این آزمایش، آزمایش برنولی محسوب می‌شود یا نه.

احتمال موفقیت در آزمایش اول $\frac{۲۰۰۰}{۵۰۰۰}$ و در آزمایش دوم مشروط به اینکه اولی در حسابهای کوتاه‌مدت سرمایه‌گذاری کرده باشد $\frac{۱۹۹۹}{۴۹۹۹}$ و مشروط به اینکه اولی در حسابهای کوتاه مدت سرمایه‌گذاری نکرده باشد $\frac{۲۰۰۰}{۴۹۹۹}$ است و ... . در هر صورت اختلاف $\frac{۲۰۰۰}{۵۰۰۰}$ (احتمال موفقیت در آزمایش اول) با $\frac{۱۹۹۹}{۴۹۹۹}$ یا $\frac{۲۰۰۰}{۴۹۹۹}$ (احتمال موفقیت در آزمایش دوم) و ... آنقدر ناچیز است که می‌توان احتمال موفقیت را از آزمایشی به آزمایش دیگر ثابت فرض کرد؛ بنابراین می‌توان گفت که این آزمایش، یک آزمایش برنولی است.

تمرین

۱. اگر از آزمایشهای ذیل ۲ نمونه گرفته شود، کدام یک آزمایش برنولی است؟

الف) در جعبه‌ای ۳ کالای ناسالم و ۲۷ کالای سالم وجود دارد. کالاها را بدون جایگزینی انتخاب می‌کنیم. موفقیت را خارج شدن کالای سالم در نظر می‌گیریم.

ب) از بین ۱۳۲۴ پیچ تولید شده در یک هفته، ۴۵۳ عدد آنها استاندارد نیستند. موفقیت را خارج شدن پیچ غیراستاندارد در نظر می‌گیریم.

ج) در یک کارگاه جوراب بافی که دارای ۲۰ دستگاه است، احتمال خراب شدن هر دستگاه ۰/۱۵ است. موفقیت را خراب شدن دستگاه در نظر می‌گیریم.

د) یک میلیون نفر می‌خواهند فردی را از بین ۵ نماینده انتخاب کنند. برآوردشده است که ۲۸۰ هزار نفر به نمایندهٔ سوم رأی می‌دهند. موفقیت را انتخاب فردی در نظر می‌گیریم که به این نماینده رأی دهد.

۶ـ۶ توزیع دو جمله‌ای

در n آزمایش برنولی که در آن احتمال موفقیت p است، متغیر تصادفی X را تعداد موفقیتها در نظر می‌گیریم. توزیع احتمال X را «توزیع دو جمله‌ای» با

احتمال موفقیت p می‌نامیم که در آن متغیر تصادفی X می‌تواند مقادیر ۰ ، ۱ ، ... و n را انتخاب کند.

فرض کنید آزمایشی ۳ بار تکرار شود و احتمال موفقیت در هر بار ثابت و برابر p باشد. حالتهای ممکن این آزمایش (موفقیتها را با S و شکست را با F نشان می‌دهیم)همراه با احتمال ۰ ، ۱ ، ۲ و ۳ بار موفقیت در این آزمایش به این صورت است:

تعداد موفقیتها	۰	۱	۲	۳
حالتهای ممکن	FFF	SFF FSF FFS	SSF SFS FSS	SSS
احتمال	$q \times q \times q$ $= q^3$	$p \times q \times q +$ $q \times p \times q +$ $q \times q \times p$ $= 3pq^2$	$p \times p \times q +$ $p \times q \times p +$ $q \times p \times p$ $= 3p^2q$	$p \times p \times p$ $= p^3$

بنابراین احتمال ۰ ، ۱ ، ۲ و ۳ موفقیت به ترتیب برابر است با:

$$P(X=0) = q \times q \times q = q^3$$
$$P(X=1) = p \times q \times q + q \times p \times q + q \times q \times p = 3pq^2$$
$$P(X=2) = p \times p \times q + p \times q \times p + q \times p \times p = 3p^2q$$
$$P(X=3) = p \times p \times p = p^3$$

حالتهای با ۲ موفقیت را در نظر بگیرید (SSF ، SFS ، FSS). تعداد این حالتها برابر است با تعداد جایگشتهای سه حرفی $3 = \binom{3}{2,1}$ و چون در این آزمایش تنها دو پیامد F و S وجود دارد، پس می‌توان به صورت ساده‌تر نوشت $\binom{3}{2} = 3$. از آنجا که احتمال کلیهٔ جایگشتهای مختلف یکسان است، می‌توانیم برای محاسبهٔ $P(X=2)$ ، حاصل $\binom{3}{2}$ را در احتمال مربوط به یکی از جایگشتها (p^2q) ضرب کنیم؛ یعنی:

$$P(X=2) = \binom{3}{2} p^2 q$$

با تعمیم مثال مذکور فرمول توزیع احتمال دو جمله‌ای را می‌توان به این صورت نوشت:

$$P(X=x)=\binom{n}{x}p^x q^{n-x} \quad , \quad x=۱,۲,\dots,n$$

در این فرمول، x تعداد موفقیتهای مورد نظر، n تعداد آزمایشها، p احتمال موفقیت در هر آزمایش و q یا $(۱-p)$ احتمال عدم موفقیت در هر آزمایش است.

مثال ۲۴-۶ دانشجویی می‌خواهد به ۵ سؤال دو گزینه‌ای (صحیح ـ غلط) پاسخ دهد. احتمال دادن پاسخ درست به هر سؤال ۰/۷ است. می‌خواهیم بدانیم احتمال اینکه دقیقاً به ۲ سؤال پاسخ درست بدهد، چقدر است.

چون سؤالات مستقل از یکدیگر و احتمال دادن پاسخ درست به هر سؤال با سؤال دیگر برابر است (۰/۷) ، پس توزیع تعداد پاسخهای درست دارای توزیع دو جمله‌ای است که در آن n = ۵ ، p = ۰/۷ و x = ۰ , ۱ , ۲ , ۳ , ٤ , ۵ است؛ بنابراین:

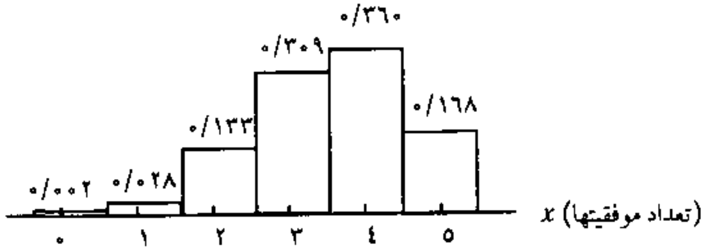
$$P(X=۲)=\binom{۵}{۲}(۰/۷)^۲(۰/۳)^{۵-۲}=۰/۱۳۲۳$$

مثال ۲۵-۶ می‌خواهیم تابع احتمال مثال ۲٤-۶ را پیدا کرده، نمودار احتمال آن را رسم کنیم.

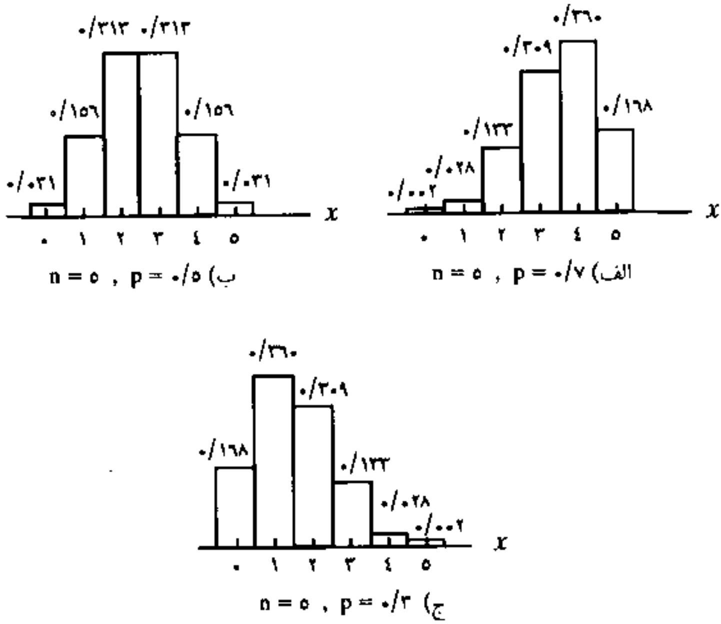
تابع احتمال مربوط به n=۵ و p=۰/۷

تعداد جوابهای درست x	احتمال $P(X=x)=\binom{n}{x}p^x q^{n-x}$
۰	۰/۰۰۲
۱	۰/۰۲۸
۲	۰/۱۳۳
۳	۰/۳۰۹
٤	۰/۳۶۰
۵	۰/۱۶۸

نمودار احتمال به این صورت خواهد بود:



حال تأثیر احتمال موفقیت توزیع دو جمله‌ای (p) را بر نمودار احتمال آن بررسی می‌کنیم. برای n = ۵ و ۳ مقدار متفاوت p = ۰/۳ ، p = ۰/۵ ، p = ۰/۷ ، ۳ نمودار احتمال در شکل ۱ـ۶ رسم شده است. با توجه به شکل ۱ـ۶ ملاحظه می‌شود که اگر احتمال موفقیت، بیشتر از ۰/۵ باشد (نمودار الف) ، نمودار احتمال مربوطه چوله به چپ است. اگر احتمال موفقیت کوچکتر از ۰/۵ باشد (نمودار ج) نمودار احتمال مربوطه چوله به راست است و اگر احتمال موفقیت مساوی ۰/۵ باشد (نمودار ب) نمودار احتمال، متقارن است.



شکل ۱ـ۶ تأثیر pهای مختلف روی نمودار احتمال

مثال ۲۶-۶ از بین شرکتهایی که سال گذشته حسابرسی شده‌اند، ۰/۲۵ آنها حسابهایشان رد شده است. اگر ۴ شرکت بر حسب تصادف انتخاب شده باشند، می‌خواهیم بدانیم احتمال هریک از این موارد چقدر است.

الف) حسابهای یک شرکت رد شده باشد.

ب) حداکثر حسابهای دو شرکت رد شده باشد.

در این مثال $n = 4$ است و اگر احتمال موفقیت را ردشدن حسابهای شرکتی در نظر بگیریم، $p = 0/25$ خواهد بود، پس خواهیم داشت:

الف)
$$P(X = 1) = \binom{4}{1}(0/25)^1(0/75)^{4-1} = 0/42$$

ب)
$$P(X \leq 2) = p(X = 0) + p(X = 1) + p(X = 2)$$
$$= \binom{4}{0}(0/25)^0(0/75)^{4-0} + \binom{4}{1}(0/25)^1(0/75)^{4-1} + \binom{4}{2}(0/25)^2(0/75)^{4-2}$$
$$= 0/32 + 0/42 + 0/21 = 0/95$$

۱-۶-۶ استفاده از جدولهای توزیع دو جمله‌ای

وقتی n نسبتاً بزرگ است، محاسبهٔ احتمالهای دو جمله‌ای کار خسته کننده‌ای است؛ زیرا در این محاسبات باید p و q به توانهای بزرگی برسند. برای حل این مشکل، جدولهایی تهیه شده و در انتهای کتاب (جدول ۱ پیوست) آورده شده است. در این جدولها احتمال تجمعی $P(X \leq c) = \sum_{x=0}^{c} \binom{n}{x} p^x q^{n-x}$ برای مقادیر مختلف n، p و c آورده شده است.

مثال ۲۷-۶ فرض کنید ۰/۲۰ از مصرف‌کنندگان پودرهای شوینده، مصرف‌کنندهٔ مارک تجاری ما هستند. اگر ۱۷ مصرف‌کننده انتخاب شوند، می‌خواهیم بدانیم هریک از این احتمالات چقدر است.

الف) ۳ مصرف کننده مشتری ما باشند.

ب) حداقل ۴ مصرف کننده مشتری ما باشند.

ج) حداقل ۲ و حداکثر ۵ مصرف کننده مشتری ما باشند.

الف)
$$P(X = 3) = P(X \leq 3) - P(X \leq 2) = 0/549 - 0/310 = 0/239$$

ب) $P(X \geq ۴) = ۱ - P(X < ۴) = ۱ - P(X \leq ۳) = ۱ - ۰/۵۴۹ = ۰/۴۵۱$

ج) $P(۲ \leq X \leq ۵) = P(X \leq ۵) - P(X < ۲) = P(X \leq ۵) - P(X \leq ۱)$

$$= ۰/۸۹۴ - ۰/۱۱۸ = ۰/۷۷۶$$

۲-۶-۶ میانگین و واریانس توزیع دو جمله‌ای

می‌دانیم که توزیع دو جمله‌ای شامل n آزمایش مستقل برنولی است. اگر فرض کنیم که ۱ = n است، توزیع دو جمله‌ای به توزیع برنولی‌ای تبدیل می‌شود که میانگین و واریانس آن به این صورت است:

y	۰	۱
P(Y = y)	q	p

$$E(Y) = ۰ \times q + ۱ \times p = p$$

$$E(Y^۲) = (۰)^۲ q + (۱)^۲ p = p$$

$$V(Y) = E(Y^۲) - [E(Y)]^۲$$

$$= p - p^۲ = p(۱ - p) = pq$$

اگر X را تعداد موفقیتها در توزیع دو جمله‌ای بدانیم، خواهیم داشت:

$$X = Y_۱ + Y_۲ + ... + Y_n$$

در این رابطه، $Y_۱$ تعداد موفقیتها در آزمایش اول (۰ یا ۱)، $Y_۲$ تعداد موفقیتها در آزمایش دوم (۰ یا ۱) و ... است و چون آزمایشها مستقل هستند، خواهیم داشت:

$$E(X) = E(Y_۱) + ... + E(Y_n) = p + ... + p = np$$

$$V(X) = V(Y_۱) + ... + V(Y_n) = pq + ... + pq = npq$$

تمرین

۱. توزیعی دو جمله‌ای با $n = ۷$ و $p = ۰/۲$ را در نظر گرفته، این موارد را پیدا کنید.

الف) $P(X = ۵)$

ب) $P(X > ۲)$

ج) $P(X < ۸)$

د) $P(X \geq ۴)$

ه) $P(۲ < X \leq ۵)$

و) $P(۲ \leq X \leq ۵)$

۲. احتمال اینکه مشتری‌ای که وارد فروشگاهی می‌شود چیزی بخرد ۰/۷ است. اگر ۷ مشتری وارد فروشگاهی شده باشند، این احتمالات را محاسبه کنید:

الف) همه مشتریان خرید کنند.

ب) حداقل ٤ مشتری خرید کند.

۳. در تمرین ۲، فرض کنید ۲۰ مشتری وارد فروشگاه شده باشند، این موارد را محاسبه کنید:

الف) ۱۰ مشتری خرید کند.

ب) حداکثر ۱۵ مشتری خرید کند.

ج) حداقل ۵ مشتری و حداکثر ۱۵ مشتری خرید کند.

د) انتظار دارید از این ۲۰ مشتری چند نفر خرید کنند؟

٤. در تمرین ۲، امید ریاضی و واریانس تعداد مشتریان خرید کرده، چقدر است؟

٦ـ۷ توزیعهای دو جمله‌ای منفی و هندسی

٦ـ۷ـ۱ توزیع دو جمله‌ای منفی

در آزمایشهای برنولی، گاهی می‌خواهیم احتمال x موفقیت از n آزمایش را بدانیم که به توزیع دو جمله‌ای مربوط می‌شود. گاهی نیز به دانستن تعداد آزمایشهایی که در آنها K موفقیت رخ می‌دهد، علاقه‌مند هستیم که به توزیع دو جمله‌ای منفی مربوط می‌شود؛ مثلاً احتمال اینکه پنجمین فردی که شایعه‌ای را شنیده، دومین فردی باشد که آن را باور می‌کند یا احتمال اینکه ششمین کالای معیوب در یک کارتن، پنجمین کالای معیوبی باشد که مأمورین کنترل کیفیت متوجه آن شده‌اند یا احتمال اینکه دزدی که در دفعهٔ هشتم دست به دزدی زده، برای سومین بار دستگیر شود.

فرمول توزیع دو جمله‌ای منفی به این صورت است:

$$P(X=x)=\binom{x-1}{K-1}p^k q^{x-k} \quad , \quad x=K, K+1, K+2, \ldots$$

در این فرمول، K تعداد موفقیتها در x آزمایش برنولی و p احتمال موفقیت در هر آزمایش برنولی است. امید ریاضی و واریانس توزیع دو جمله‌ای منفی به ترتیب برابر است با $E(X)=\frac{K}{P}$ و $V(X)=\frac{Kq}{P^2}$. به توزیع دو جمله‌ای منفی، «توزیع پاسکال» یا «توزیع زمان انتظار» نیز می‌گویند.

مثال ۲۸-۶ احتمال اینکه دزدی در حین دزدی دستگیر شود ۰/۶ است. می‌خواهیم بدانیم احتمال اینکه در هشتمین دزدی خود برای چهارمین بار دستگیر شود، چقدر است.

در این مثال $x=8$، $K=4$ و $P=0/6$ است، پس خواهیم داشت:

$$P(X=8)=\binom{8-1}{4-1}(0/6)^4(0/4)^{8-4}=0/12$$

مثال ۲۹-۶ اگر کالایی معیوب باشد مأمور کنترل کیفیت با احتمال ۰/۸۰ متوجه آن می‌شود. احتمال اینکه ششمین کالای معیوب، پنجمین کالای معیوبی باشد که وی متوجه آن شده است، چقدر است.

در این مثال $x=6$، $K=5$ و $p=0/8$ است، پس خواهیم داشت:

$$P(X=6)=\binom{6-1}{5-1}(0/8)^5(0/2)^{6-5}=0/39$$

۲-۷-۶ توزیع هندسی

اگر در توزیع دو جمله‌ای منفی $K=1$ باشد، یعنی بخواهیم اولین موفقیت را در x آزمایش بدانیم، باز هم می‌توان از این توزیع استفاده کرد، ولی این حالت، حالت خاصی از توزیع دو جمله‌ای منفی است که آن را توزیع هندسی می‌نامند؛ مثلاً احتمال اینکه پنجمین فردی که شایعه‌ای را شنیده، اولین فردی باشد که آن را باور می‌کند یا احتمال اینکه ششمین کالای معیوب در یک کارتن اولین کالای معیوبی باشد که مأموران کنترل کیفیت متوجه آن می‌شوند یا احتمال اینکه دزدی که در دفعهٔ هشتم دست به دزدی زده برای اولین بار دستگیر شود.

فرمول توزیع هندسی به این صورت است:

$$P(X=x)=pq^{x-1} \ , \ x=1,2,...$$

در این فرمول، p احتمال موفقیت در یک آزمایش و x شمارهٔ آزمایشی است که اولین موفقیت در آن رخ داده است.

امید ریاضی و واریانس توزیع هندسی به ترتیب برابر است با $E(X)=\frac{1}{p}$ و $V(X)=\frac{q}{p^2}$.

مثال ۶ـ۳۰ تیراندازی ۰/۷۷ از تیرهای خود را به هدف می‌زند. می‌خواهیم بدانیم احتمال اینکه سومین تیر وی اولین تیری باشد که به هدف اصابت می‌کند، چقدر است.

در این مثال $x=3$ و $p=0/77$ است، پس خواهیم داشت:

$$P(X=3)=0/77\times 0/23^{3-1}=0/04$$

مثال ۶ـ۳۱ فردی متقاضی گواهینامهٔ رانندگی است و با احتمال ۰/۶۵ در آزمون رد می‌شود. می‌خواهیم بدانیم احتمال اینکه در دومین آزمون قبول شود، چقدر است.

در این مثال $x=2$ و $p=0/35$ است، پس خواهیم داشت:

$$P(X=2)=0/35\times 0/65^{2-1}=0/23$$

تمرین

۱. اگر احتمالهای داشتن فرزند پسر و دختر مساوی باشد، این احتمالات را حساب کنید:

الف) چهارمین فرزند خانواده‌ای اولین پسر آنها باشد.

ب) هفتمین فرزند خانواده‌ای دومین دختر آنها باشد.

ج) دهمین فرزند خانواده‌ای چهارمین یا پنجمین پسر آنها باشد.

۲. احتمال اینکه فردی از چراغ قرمز عبور کند و پلیس متوجه نشود ۰/۳۵ است. احتمال اینکه در حین عبور از چهارمین چراغ قرمز جریمه شود، چقدر است؟

۳. ۵ تا از حسابهای دریافتی شرکتی دارای اشتباه است. احتمال اینکه حسابرسی داخلی متوجه هر حساب اشتباه شود ۰/۵۵ است. این احتمالات را محاسبه کنید.

الف) احتمال اینکه چهارمین حساب اشتباه، دومین حسابی باشد که حسابرس داخلی متوجه آن شده است.

ب) احتمال اینکه چهارمین حساب اشتباه، اولین حسابی باشد که حسابرس داخلی متوجه آن شده است.

٤. احتمال مطلع شدن هر بار مشتری از آگهی مربوط به یک شرکت که از تلویزیون پخش می‌شود ۰/۶۷ است. احتمال اینکه در سومین آگهی شرکت، مشتری مطلع شود، چقدر است؟

۶ـ۸ توزیع چند جمله‌ای

اگر آزمایشی شامل بیش از دو پیامد ممکن باشد و احتمال هر پیامد در آزمایشهای مختلف ثابت و آزمایشها مستقل از یکدیگر باشند، توزیع مربوط به آن، توزیع چند جمله‌ای است. تنها تفاوت این توزیع با توزیع دو جمله‌ای این است که در توزیع دو جمله‌ای دو پیامدِ ممکن وجود دارد (موفقیت و شکست)، ولی در این توزیع بیش از دو پیامد ممکن وجود دارد؛ مثلاً نتیجهٔ مسابقه‌ای ممکن است برد، مساوی یا باخت باشد یا تولیدات کارخانه‌ای ممکن است به عالی، خوب، متوسط و ضعیف طبقه‌بندی شود یا ممکن است نظر افراد راجع به موضوعی موافق، ممتنع یا مخالف باشد.

با تعمیم مثالهای مذکور می‌توان نوشت:

اگر آزمایشی n بار به صورت مستقل انجام گیرد و هر آزمایش شامل K پیامد مجزا با احتمالهای ثابت p_1 ، p_2 ، ... ، p_K باشد به طوری که $p_1 + p_2 + ... + p_k = 1$ باشد، آنگاه احتمال وقوع x_1 بار از پیامد ۱ ، x_2 بار از پیامد ۲ ، ... و x_K بار از پیامد K دارای چنین توزیعی است:

$$P(X_1 = x_1 , X_2 = x_2 , ... , X_K = x_K) = \binom{n}{x_1 , x_2 , ... , x_K} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$$

این توزیع چند جمله‌ای است و در آن $x_1 + x_2 + ... + x_K = n$ است.

مثال ۶ـ۳۲ ۰/۲۵ از افراد شهری به نامزد اول، ۰/۳۵ به نامزد دوم و ۰/۴۰ به نامزد سوم رأی می‌دهند. فرض می‌کنیم ۱۰ نفر هم اکنون پای صندوق منتظر رأی دادن هستند، می‌خواهیم بدانیم احتمال اینکه ۲ نفر به نامزد اول، ۳ نفر به نامزد دوم و ۵ نفر به نامزد سوم رأی دهند، چقدر است.

در این مثال $p_1 = ۲۵\%$ ، $p_2 = ۳۵\%$ ، $p_3 = ۴۰\%$ و $x_1 = ۲$ ، $x_2 = ۳$ ، $x_3 = ۵$ و $n = ۱۰$ است، پس خواهیم داشت:

$$P(x_1 = ۲ , x_2 = ۳ , x_3 = ۵) = \binom{۱۰}{۲ , ۳ , ۵} (۰/۲۵)^۲ (۰/۳۵)^۳ (۰/۴۰)^۵ = ۰/۰۷$$

تمرین

۱. فرض کنید تیمی ۵ مسابقه در پیش دارد و احتمال برد، باخت، و مساوی در هر بازی به ترتیب ۰/۵، ۰/۴ و ۰/۱ است. احتمال اینکه در این ۵ مسابقه دو برد، دو مساوی و یک باخت داشته باشد، چقدر است؟

۲. بنا به اطلاعات گذشتهٔ مدیریت نیروی انسانی سازمانی، ۰/۳۲ از افرادی که این سازمان استخدام می‌کند قبل از دو سال کار سازمان را ترک می‌کنند. ۰/۲۷ در سازمان باقی می‌مانند، ولی از کار خود ناراضی هستند. ۰/۱۶ در سازمان باقی می‌مانند و نسبتاً راضی هستند و ۰/۲۵ باقی می‌مانند و از کار خود کاملاً راضی هستند. احتمال اینکه از بین ۷ نفری که شرکت جدیداً استخدام کرده است ۲ نفر سازمان را ترک کنند، یک نفر باقی بماند و ناراضی باشد، ۲ نفر باقی بمانند و نسبتاً راضی باشند و ۲ نفر نیز باقی بمانند و کاملاً راضی باشند، چقدر است؟

۶ـ۹ توزیع فوق هندسی

فرض کنید می‌خواهیم بدانیم اگر از بین ۱۰ عدد کالا که ۳ عدد آنها معیوب است، ۵ کالا را انتخاب کنیم احتمال اینکه ٤ عدد آنها سالم باشد، چقدر است. در فصل پنجم گفتیم که احتمال وقوع چنین پیشامدی برابر است با نسبت تعداد حالات مساعد به حالات ممکن؛ یعنی:

$$\text{احتمال} = \frac{\text{تعداد حالات مساعد}}{\text{تعداد حالات ممکن}}$$

تعداد حالات ممکن، یعنی انتخاب ۵ کالا از بین ۱۰ کالا (همان ترکیبهای ۵ تایی از ۱۰ تا یعنی $\binom{10}{5}$)، ولی برای محاسبهٔ تعداد حالات مساعد، می‌خواهیم از بین ۷ کالای سالم ٤ کالا و از بین ۳ کالای معیوب یک کالا انتخاب کنیم. تعداد طرقی که می‌توان از بین ۷ کالا ٤ تا را انتخاب کرد همان ترکیبهای ٤ تایی از ۷ تاست $\binom{7}{4}$ و تعداد طرقی که می‌توان از بین ۳ کالای معیوب ۱ عدد را انتخاب کرد نیز همان ترکیبهای ۱ تایی از ۳ تاست $\binom{3}{1}$.
طبق قانون ضرب، اگر بتوانیم تعداد موردنظر از کالاهای سالم را به $\binom{7}{4}$ طریق و تعداد موردنظر از کالاهای معیوب را به $\binom{3}{1}$ طریق انتخاب کنیم، تعداد موردنظر از کالاهای سالم و معیوب را می‌توان به $\binom{7}{4}\binom{3}{1}$ طریق انتخاب کرد؛ یعنی در این مثال احتمال

مورد نظر مساوی است با:

$$\frac{\binom{۷}{۴}\binom{۳}{۱}}{\binom{۱۰}{۵}}=\frac{۳۵\times ۳}{۲۵۲}=\frac{۳۵}{۸۴}$$

با تعمیم این مثال می‌توان گفت هرگاه بخواهیم این احتمال را پیدا کنیم که از بین N شیء مورد نظر که k تای آنها واجد شرایط است، n شیء را برگزینیم به طوری که x تای آن واجد شرایط باشد، از این فرمول که فرمول توزیع فوق هندسی نامیده می‌شود، استفاده می‌کنیم:

$$P(X=x)=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}\ ,\quad x=۰,۱,\ldots,k$$

امید ریاضی و واریانس توزیع فوق هندسی به این صورت است:

$$E(X)=\frac{nk}{N}\ ,\ V(X)=\frac{nK(N-k)(N-n)}{N^{۲}(N-۱)}$$

مثال ۶ـ۳۳ از بین ۸ مدیری که‌به جلسهٔ امروز دعوت شده‌اند، ۳ نفرشان رابطه مدار و بقیه کار مدارند. می‌خواهیم بدانیم اگر برحسب تصادف ٤ نفر را انتخاب کنیم، احتمال اینکه ۲ مدیر رابطه مدار انتخاب شود، چقدر است.

در این مثال $N=۸$ ، $n=۴$ ، $k=۳$ و $x=۲$ است؛ پس خواهیم داشت:

$$P(X=۲)=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}=\frac{\binom{۳}{۲}\binom{۵}{۲}}{\binom{۸}{۴}}=\frac{۳\times ۱۰}{۷۰}=\frac{۹}{۱۴}$$

مثال ٦ـ۳٤ در مثال ۶ـ۳۳ ، امید ریاضی و واریانس تعداد مدیران رابطه مدار را محاسبه می‌کنیم:

$$E(X)=\frac{nK}{N}=\frac{۴\times ۳}{۸}=۱/۵$$

یعنی از ٤ مدیر انتخابی، به طور متوسط ۱/۵ نفر آنها رابطه مدارند.

$$V(X) = \frac{nK(N-k)(N-n)}{N^۲(N-۱)} = \frac{٤ \times ۳(۸-۳)(۸-٤)}{۸^۲(۸-۱)} = ۰/۵۳٦$$

اگر N بزرگ و n در مقایسه با N، کوچک باشد (با محاسبهٔ سرانگشتی اگر n کمتر از ۵ درصد N باشد)، تفاوت چندانی بین نمونه‌گیری با جایگزینی و بدون جایگزینی وجود نخواهد داشت و می‌توان از توزیع دو جمله‌ای با پارامترهای n و $P = \frac{k}{N}$ به عنوان تقریب توزیع فوق هندسی استفاده کرد.

مثال ٦ـ۳۵ از بین ۲۰۰ متقاضی شغلی، تنها ۷۰ نفر واجد شرایط هستند. اگر ٦ نفر را برحسب تصادف انتخاب کنیم، می‌خواهیم احتمال اینکه ۳ نفر آنها واجد شرایط باشند را محاسبه کنیم.

برای حل این مثال که مثالی از نمونه‌گیری بدون جایگزینی است، باید از توزیع فوق هندسی استفاده شود ($N = ۲۰۰$، $n = ٦$، $k = ۷۰$ و $x = ۳$)، ولی چون $n < ۰/۰۵N$ است ($٦ < ۰/۰۵ \times ۲۰۰$) می‌توان از توزیع دو جمله‌ای به عنوان تقریب توزیع فوق هندسی استفاده کرد؛ یعنی:

$$n = ٦ \ , \ p = \frac{k}{N} = \frac{۷۰}{۲۰۰} = ۰/۳۵$$

$$P(X = ۳) = \binom{٦}{۳}(۰/۳۵)^۳(۰/٦۵)^{٦-۳} = ۰/۲٤$$

تمرین

۱. از بین ۱۲ نفری که متقاضی استخدام درکاری هستند، ۳ نفر قادربه انجام دادن آن هستند. قرار است به صورت تصادفی ۲ نفر انتخاب شوند. این احتمالات را محاسبه کنید:

الف) هیچ کدام قادر به انجام کار نباشند.

ب) تنها یک نفر قادر به انجام کار باشد.

ج) هر دو قادر به انجام کار باشند.

۲. ظرفی شامل ۱۰ کالاست که ۳ عدد آنها خراب است. نمونه‌ای ۲ تایی انتخاب شده‌است، احتمال اینکه حداکثر یک واحد آن خراب باشد، چقدر است؟

۳. انجمنی شامل ۲ زن و ۷ مرد است. تابع احتمال را در یک نمونه دو تایی به گونه‌ای بنویسید که X

نشان‌دهندهٔ زنان انتخاب شده، باشد. امید ریاضی (میانگین) X را محاسبه کنید.
٤. طی تحقیقاتی که اخیراً درباره ۱۲۰ مشتری یک شرکت انجام گرفته، مشخص شده است که ۹۰ نفر آنها از کیفیت تولیدات و خدمات بعد از فروش شرکت راضی هستند. اگر از این ۱۲۰ مشتری برحسب تصادف ۵ نفر انتخاب شوند، احتمال اینکه ٤ نفر از تولیدات و خدمات شرکت راضی باشند، چقدر است؟

۶-۱۰ توزیع پواسون

در توزیع دو جمله‌ای، وقتی n بزرگ شود محاسبات کار ساده‌ای نخواهد بود؛ مثلاً اگر بخواهیم بدانیم از بین ۲۰۰۰ واحد کالای موجود که هریک با احتمال ۰/۰۰۱۵ معیوبند، با چه احتمالی ۵ عددآنها معیوبند، ناچاریم $\binom{۲۰۰۰}{۵}$ و همینطور $(۰/۹۹۸۵)^{۱۹۹۵}(۰/۰۰۱۵)^{۵}$ را محاسبه و درهم ضرب کنیم که این محاسبات کار دشواری خواهد بود. در این گونه مواقع یعنی هنگامی که n به سمت بی‌نهایت و p به سمت صفر میل کند و در عین حال مقدار np ثابت بماند، استفاده از توزیع پواسون با این تابع احتمال تقریب مناسبی برای توزیع دو جمله‌ای خواهد بود.

$$P(X=x)=\frac{e^{-\lambda}\lambda^{x}}{x!}\ ,\ \ x=۰,۱,۲,...$$

که در این رابطه، λ پارامتر توزیع و برابر با np، $(\lambda=np)$ و $e \simeq ۲/۷۱۸$ است. جدول ۲ پیوست، احتمالات تجمعی پواسون را به ازای مقادیر مختلف λ و x نشان می‌دهد.

مثال ۶-۳۶ می‌خواهیم بدانیم احتمال اینکه از بین ۲۰۰۰ واحد کالای موجود شرکتی که هرکدام با احتمال ۰/۰۰۱۵ معیوبند، ۵ عدد آنها معیوب باشد، چقدر است.

چون n بزرگ و P خیلی کوچک است، می‌توان از توزیع پواسون استفاده کرد؛ یعنی:

$$\lambda=np=۲۰۰۰\times ۰/۰۰۱۵=۳$$

بنابراین:

$$P(X=۵)=\frac{e^{-۳}(۳)^{۵}}{۵!}=۰/۱۰۰۸$$

به طور کلی وقتی $n \geq ۲۰$ و $p \leq ۰/۰۵$ باشد، توزیع پواسون تقریب خوبی برای توزیع دو جمله‌ای و وقتی $n \geq ۱۰۰$ و $np \leq ۱۰$ باشد، تقریبی بسیار عالی برای آن محسوب می‌شود.

امید ریاضی و واریانس توزیع پواسون λ است. به عبارت دیگر $E(X)=\lambda$ و

$\lambda = V(X)$. از بین توزیعهای رایج، چه پیوسته و چه گسسته، توزیعی که میانگین (امید ریاضی) و واریانس آن با هم برابر است توزیع پواسون است و این یکی از خصوصیات جالب توجه این توزیع است.

مثال ۶ـ۳۷ براساس تجربه مشخص شده است که یک تلفنچی ۳ درصد از تلفنها را اشتباه وصل می‌کند. اگر امروز ۱۵۰ تلفن وصل کرده باشد، می‌خواهیم این موارد را محاسبه کنیم:

الف) امید ریاضی (میانگین) تلفنهایی که اشتباه وصل شده است.

ب) احتمال اینکه ۳ شماره را اشتباه وصل کرده باشد.

ج) احتمال اینکه بیش از یک شماره را اشتباه وصل کرده باشد.

چون $n = ۱۵۰ > ۲۰$ و $p = ۰/۰۳ \leq ۰/۰۵$ است، پس می‌توان از توزیع پواسون برای تقریب توزیع دو جمله‌ای استفاده کرد.

الف)
$$\lambda = np = ۱۵۰ \times ۰/۰۳ = ۴/۵$$

ب)
$$P(X = ۳) = \frac{e^{-۴/۵}(۴/۵)^{۳}}{۳!} = ۰/۱۶۸۷$$

ج)
$$P(X > ۱) = ۱ - P(X \leq ۱) = ۱ - [P(X = ۰) + P(X = ۱)]$$

$$= ۱ - \left(\frac{e^{-۴/۵}(۴/۵)^{۰}}{۰!} + \frac{e^{-۴/۵}(۴/۵)^{۱}}{۱!}\right) = ۰/۹۳۸۹$$

۶ـ۱۰ـ۱ توزیع پواسون برای تعداد مراجعات

معمولاً تعداد مراجعاتی که به سیستمی می‌شود، از توزیع پواسون برخوردار است که در این صورت λ متوسط تعداد مراجعات در واحد زمان است. کاربرد توزیع پواسون در این زمینه خیلی بیشتر از تقریب آن برای توزیع دو جمله‌ای است؛ مثلاً تعداد مشتریانی که در هر ساعت به یک سیستم صف مراجع می‌کنند، تعداد اتومبیلهایی که در هر دقیقه برای زدن بنزین به پمپ بنزینی مراجعه می‌کنند، تعداد مشتریانی که در هر ساعت به رستورانی مراجعه می‌کنند، ... می‌تواند دارای توزیع پواسون باشد. در این حالت فرمول توزیع پواسون به صورت زیر است:

$$P(X = x) = \frac{e^{-\lambda t}(\lambda t)^{x}}{x!}\ ,\quad x = ۰, ۱, ۲, \ldots$$

در این رابطه، t واحد زمانی است که متوسط مراجعات (λ) با توجه به آن تعریف می‌شود.

مثال ۶.۳۸ـ٦ تعداد مشتریانی که به بانکی مراجعه می‌کنند دارای توزیع پواسون با میانگین ۲ مشتری در هر دقیقه هستند؛ می‌خواهیم این احتمالات را محاسبه کنیم:

الف) در یک دقیقه اول هیچ مشتری‌ای مراجعه نکند.

ب) در ۱۰۵ ثانیه اول کمتر از ۳ مشتری مراجعه کند.

ج) در ۱/۵ دقیقه اول یک مشتری مراجعه کند.

الف)

$$\lambda t = ۲ \times ۱ = ۲$$

$$P(X=۰) = \frac{e^{-۲}(۲)^۰}{۰!} = ۰/۱۳۵۳$$

ب)

$$\lambda t = ۲ \times \frac{۱۰۵}{٦۰} = ۳/۵$$

$$P(X<۳) = P(X=۰) + P(X=۱) + P(X=۲)$$

$$= \frac{e^{-۳/۵}(۳/۵)^۰}{۰!} + \frac{e^{-۳/۵}(۳/۵)^۱}{۱!} + \frac{e^{-۳/۵}(۳/۵)^۲}{۲!}$$

$$= ۰/۰۳۰۲ + ۰/۱۰۵۷ + ۰/۱۸۵۰ = ۰/۳۲۰۹$$

ج)

$$\lambda t = ۲ \times ۱/۵ = ۳$$

$$P(X=۱) = \frac{e^{-۳}(۳)^۱}{۱!} = ۰/۱٤۹٤$$

تمرین

۱. توزیع پواسونی با $\lambda = ۲$ را در نظر گرفته $P(X=۰)$، $P(X \le ۳)$ و $P(X \ge ۲)$ را حساب کنید.

۲. تعداد اتومبیلهای سواری‌ای که به رستورانی مراجعه می‌کنند دارای توزیع پواسون با میانگین ۲/۵ اتومبیل در هر ۱۰ دقیقه هستند، این احتمالات را حساب کنید.

الف) در ۱۰ دقیقه اول بیش از یک اتومبیل مراجعه کند.

ب) در ۲۰ دقیقه بیش از ۳ اتومبیل و کمتر یا مساوی ٦ اتومبیل مراجعه کند.

ج) در ۵ دقیقه اول اتومبیلی مراجعه نکند.

۳. طبق تجارب گذشته مشخص شده است که ۰/۰۲۵ از مسافرین قطار بلیت خود را باز پس می‌دهند. اگر امروز ۲۵۰ نفر بلیت گرفته باشند، این احتمالات را محاسبه کنید.

الف) ۵ نفر بلیت خود را بازپس دهند.

ب) کمتر یا مساوی ۳ نفر بلیت خود را باز پس دهند.

۴. توزیع پواسونی با $\lambda = 1$ را در نظر بگیرید و برای $x = 0, 1, 2, 3, 4$، $P(X=x)$ را محاسبه و نمودار احتمال آن را رسم کنید. آیا این نمودار چولگی دارد؟

۵. توزیع پواسونی با $\lambda = 4$ را در نظر بگیرید و برای $x = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9$، $P(X=x)$ را محاسبه و نمودار احتمال آن را رسم کنید. آیا این نمودار چولگی دارد؟

۶. از مقایسهٔ نمودارهای احتمال دو تمرین ۴ و ۵ چه نتیجه‌ای می‌گیرید؟

۶ـ۱۱ خلاصه

در این فصل ابتدا متغیر تصادفی و سپس تابع احتمال و تابع تجمعی احتمال (تابع توزیع) معرفی شدند. همان طور که ذکر شد، هر متغیر تصادفی، دارای میانگین (امید ریاضی) و واریانس است؛ به طوری که می‌توان میانگین و واریانس متغیرهای تصادفی گسسته را محاسبه کرد. گاهی لازم است که رفتار همزمان دو متغیر تصادفی مورد مطالعه قرار گیرد که بدین منظور تابع احتمال توأم بیان شد. با معرفی کوواریانس نوع و شدت رابطهٔ خطی بین دو متغیر تصادفی قابل محاسبه است.

از بین توزیعهای رایج گسسته، توزیعهای برنولی، دو جمله‌ای، دو جمله‌ای منفی، هندسی، چند جمله‌ای، فوق هندسی، و توزیع پواسون معرفی شدند و شرایط استفاده و موارد کاربرد آنها با ارائه مثالهای متعدد مطرح گردید. معرفی و بررسی توزیع متغیرهای تصادفی پیوسته به فصل بعد موکول شد.

۶ـ۱۲ سؤالات و مسائل

سؤالات دوگزینه‌ای

۱. تابع احتمال و تابع توزیع را می‌توان به صورت مترادف به کار برد. ص ☐ غ ☐

۲. امید ریاضی همان میانگین موزون است که احتمالات در آن نقش وزنها یا ضرایب را در میانگین موزون ایفا می‌کنند. ص ☐ غ ☐

۳. تابع احتمال توأم به بررسی رفتار همزمان دو یا چند متغیر تصادفی می‌پردازد. ص ☐ غ ☐

۴. در توزیع احتمال توأم، احتمالات حاشیه‌ای X همان توزیع احتمال X است. ص ☐ غ ☐

۵. اگر مقدار کوواریانس ۲۱- شود می‌توان گفت که رابطهٔ مستقیمی بین دو متغیر تصادفی وجود دارد. ص ☐ غ ☐

۶. اگر کوواریانس دو متغیر تصادفی X و Y، صفر باشد، این دومتغیر لزوماً مستقلند. ص ☐ غ ☐

۷. اگر دومتغیر تصادفی X و Y مستقل باشند، کوواریانس آنهالزوماً صفر است. ص ☐ غ ☐

۸. اگر دو متغیر تصادفی X و Y مستقل باشند، آنگاه $E(XY) = E(X) \times E(Y)$. ص ☐ غ ☐

۹. اگردو متغیرتصادفی X و Y مستقل باشند، آنگاه $V(X - Y) = V(X) + V(Y)$. ص ☐ غ ☐

۱۰. نام دیگر توزیع دو جمله‌ای منفی، توزیع هندسی است. ص ☐ غ ☐

۱۱. مقدار یک متغیر تصادفی را معمولاً می‌توان قبل از وقوع آن تعیین کرد. ص ☐ غ ☐

۱۲. همواره می‌توان توزیع دو جمله‌ای را با توزیعهای دیگر تقریب زد. ص ☐ غ ☐

۱۳. امید ریاضی (میانگین) توزیع دو جمله‌ای برابر با np است. ص ☐ غ ☐

۱٤. وقتی احتمال موفقیت در توزیع دو جمله‌ای برابر ۰/۵ باشد، نمودار احتمال آن متقارن است. ص ☐ غ ☐

۱۵. وقتی احتمال موفقیت در توزیع دو جمله‌ای بزرگتر از ۰/۵ باشد، نمودار احتمال آن چوله به راست است. ص ☐ غ ☐

۱٦. توزیع دو جمله‌ای حالت خاصی از توزیع چند جمله‌ای است. ص ☐ غ ☐

۱۷. توزیع فوق هندسی برای نمونه گیریهای بدون جایگذاری استفاده می‌شود. ص ☐ غ ☐

۱۸. اگر در توزیع فوق هندسی $n < 0/05 N$ باشد، می‌توان از توزیع دو جمله‌ای برای تقریب توزیع فوق هندسی استفاده کرد. ص ☐ غ ☐

۱۹. از بین توزیعهای رایج، توزیعی که میانگین (امید ریاضی) و واریانس آن با هم برابر است، توزیع پواسون است. ص ☐ غ ☐

۲۰. کاربرد توزیع پواسون برای تعداد مراجعات، به مراتب بیشتر از کاربرد آن برای تقریب توزیع دو جمله‌ای است. ص ☐ غ ☐

۲۱. تعداد آزمایشها (n) و احتمال موفقیت در هر آزمایش (p)، دو پارامتر لازم برای بیان توزیع دو جمله‌ای هستند. ص ☐ غ ☐

۲۲. تنها پارامتر توزیع هندسی، احتمال موفقیت (p) در یک آزمایش است. ص ☐ غ ☐

۲۳. اگر متوسط تعداد مراجعه کنندگان به اداره‌ای ۵ نفر درهر دقیقه با توزیع پواسون باشد، با اطمینان می‌توان گفت که در هر ۲ دقیقه ۱۰ مراجعه کننده به این اداره وارد می‌شوند. ص ☐ غ ☐

۲٤. واریانس توزیع هندسی برابر است با $\frac{1}{p}$. ص ☐ غ ☐

۲۵. $F(x) = P(X \leq x)$ را تابع توزیع متغیر تصادفی X می‌گوییم. ص ☐ غ ☐

سؤالات چهارگزینه‌ای

۲۶. اگر امید ریاضی سود روزانهٔ فروشگاهی ۲۰ هزار ریال باشد، آنگاه کدام یک از این موارد صحیح است؟

الف) فردا ۲۰ هزار ریال سود می‌برد.
ب) سود فردای این فروشگاه کمتر از ۲۰ هزار ریال است.
ج) سود فردای این فروشگاه حداقل ۲۰ هزار ریال است.
د) هیچ کدام.

۲۷. در کدام یک از این موارد توزیع پواسون تقریب خوبی برای توزیع دو جمله‌ای است؟

الف) $P = ۰/۳۲ , n = ٤۰$
ب) $P = ۰/٦۸ , n = ٤۰$
ج) $q = ۰/۹۸ , n = ۲۰۰$
د) $P = ۰/۰۳ , n = ۲۰$

۲۸. اگر درتوزیع دو جمله‌ای $p = ۰/۲۵$ باشد، نمودار احتمال آن چه شکلی خواهد داشت؟

الف) متقارن
ب) چوله به چپ
ج) چوله به راست
د) به n بستگی دارد.

۲۹. در کدام یک از این موارد می‌توان از توزیع دو جمله‌ای برای تقریب توزیع فوق هندسی استفاده کرد؟

الف) K و N هر دو بزرگ باشند.
ب) K بزرگ ولی N کوچک باشد.
ج) N بزرگ و n کوچک باشد.
د) K و N هر دو کوچک باشند.

۳۰. در یک توزیع دو جمله‌ای خاص با $p = ۰/٤$ ، مقدار $\frac{۷!}{۳! \times ٤!}(۰/٤)^۳(۰/٦)^٤$ نشان‌دهنده احتمال کدام یک از این حالات است؟

الف) دقیقاً ٤ موفقیت در ۷ آزمایش
ب) دقیقاً ۳ موفقیت در ۷ آزمایش
ج) ۳ موفقیت یا بیشتر در ۷ آزمایش
د) ٤ موفقیت یا کمتر در ۷ آزمایش

۳۱. ۰/۸۰ از محصولات کارخانه‌ای سالم است. احتمال اینکه از ٤ کالای خریداری شده از این کارخانه یک کالا سالم باشد، چقدر است؟

الف) $\frac{۱۲}{٦۲۵}$
ب) $\frac{۲٤}{٦۲۵}$
ج) $\frac{۱٦}{٦۲۵}$
د) $\frac{۵}{۱٦}$

۳۲. از ۱۰ محصول تولیدی به وسیلهٔ ماشینی ۳ واحد آنها معیوب است. نمونه ۲ تایی ازمحصولات

این ماشین انتخاب شده است. احتمال اینکه هیچ کدام سالم نباشد، چقدر است؟

الف) $\frac{۳}{٤٥}$ ج) $\frac{۷}{٤٥}$

ب) $\frac{۲۱}{٤٥}$ د) صفر

۳۳. تعداد سرکشیهای اضطراری به یک خط تولید دارای توزیع پواسون با متوسط ۳ سرکشی در روز است. احتمال اینکه در یک روز هیچ سرکشی‌ای صورت نگرفته باشد، چقدر است؟

الف) $e^{-۳}$ ج) $۲\,e^{-۳}$

ب) $۳\,e^{-۳}$ د) $\frac{۱}{۳}e^{-۳}$

۳٤. به طور متوسط با توزیع پواسون در هر دقیقه دو اتومبیل برای زدن بنزین بدون سرب به پمپ‌بنزینی مراجعه می‌کنند. احتمال اینکه در پنج دقیقه دو اتومبیل مراجعه کنند، چقدر است؟

الف) $۵۰\,e^{-۱۰}$ ج) $٤\,e^{-٤}$

ب) $۲\,e^{-۵}$ د) $۱۰\,e^{-۱۰}$

۳۵. به طور متوسط با توزیع پواسون در هر ساعت دوازده اتومبیل برای زدن بنزین مراجعه می‌کنند. احتمال اینکه در ۱۵ دقیقه ۳اتومبیل مراجعه کنند، چقدر است؟

الف) $e^{-۳}$ ج) $٤/۵\,e^{-۳}$

ب) $e^{-۱۲}$ د) $۱۲\,e^{-۳}$

۳٦. احتمال اینکه هر پرتاب بازیکنی به هدف بخورد ۰/۸۰ است. احتمال اینکه سومین پرتابی که به هدف می‌خورد، پنجمین پرتاب وی باشد، چقدر است؟

الف) ۰/۱۳۲ ج) ۰/۲۳۱

ب) ۰/۳۲۱ د) ۰/۱۲۳

۳۷. احتمال اصابت موشکی به یک جنگنده ۰/۳۰ است. با اصابت یک موشک، جنگنده سقوط می‌کند. احتمال اینکه در پرتاب پنجمین موشک، جنگنده سقوط کند، چقدر است؟

الف) ۰/۰۰۵ ج) ۰/۰۵۵

ب) ۰/۰۷۲ د) ۰/۰۵۰

۳۸. میانگین و واریانس توزیع پواسون به ترتیب (از راست به چپ) کدام است؟

الف) λ , λ ج) $\frac{۱}{\lambda^{۲}}, \lambda$

ب) $\sqrt{\lambda} , \lambda$ د) $\frac{۱}{\lambda^{۲}}, \frac{۱}{\lambda}$

۳۹. اگر $Cov(X, Y) = 0$ باشد، کدام عبارت درباره رابطه X و Y صحیح است؟

الف) رابطه‌ای وجود ندارد. ج) دو متغیر مستقلند.

ب) رابطهٔ غیرخطی وجود دارد. د) یا رابطهٔ غیرخطی یا استقلال وجود دارد.

۴۰. اگر $E(X) = 2$، $E(Y) = 1$ و X و Y مستقل باشند، آنگاه کدام عبارت صحیح است؟

الف) $E(XY) = 2$ ج) $Cov(X, Y) = 0$

ب) $E(X + Y) = 3$ د) هر سه صحیح است.

۴۱. اگر $V(X) = 4$ و $V(Y) = 6$ و X و Y مستقل نباشند و رابطهٔ آنها مستقیم باشد، آنگاه کدام مورد صحیح است؟

الف) $V(X + Y) = 10$ ج) $V(X + Y) > 10$

ب) $V(X + Y) < 10$ د) هیچ‌کدام

۴۲. اگر $V(X) = \frac{1}{2}$ و $V(Y) = \frac{2}{3}$ و $V(X + Y) = \frac{5}{6}$ باشد، آنگاه کدام عبارت صحیح است؟

الف) $Cov(X, Y) = \frac{1}{3}$ ج) $Cov(X, Y) = -\frac{1}{6}$

ب) $Cov(X, Y) = \frac{1}{2}$ د) $Cov(X, Y) = -\frac{1}{4}$

۴۳. این تابع احتمال توأم را در نظر بگیرید:

$x \backslash y$	0	10
-2	$\frac{1}{6}$	$\frac{1}{3}$
2	$\frac{1}{3}$	$\frac{1}{6}$

احتمالات حاشیه‌ای Y کدام است؟

الف)

y	0	10
f(y)	$\frac{1}{6}$	$\frac{1}{3}$

ج)

y	0	10
f(y)	$\frac{1}{2}$	$\frac{1}{2}$

ب)

Y	0	10
f(y)	$\frac{1}{3}$	$\frac{1}{6}$

د) هیچ کدام

۴۴. در سؤال ۴۳، $P(Y > X)$ با کدام یک از این موارد برابر است؟

الف) $\frac{1}{6}$ ب) $\frac{1}{3}$ ج) $\frac{1}{2}$ د) $\frac{2}{3}$

۴۵. در سؤال ۴۳، $P(Y = 10 / X = 2)$ با کدام یک از این موارد برابر است؟

الف) $\frac{1}{6}$ ب) $\frac{1}{3}$ ج) $\frac{1}{2}$ د) ۱

۴۶. در سؤال ۴۳ ، $E(XY)$ با کدام یک از این موارد برابر است؟

الف) $\frac{3}{10}$ ب) صفر ج) ۱۰ د) $-\frac{10}{3}$

۴۷. در سؤال ۴۳، کوواریانس با کدام یک از این موارد برابر است؟

الف) $\frac{3}{10}$ ب) $\frac{10}{3}$ ج) $-\frac{10}{3}$ د) صفر

۴۸. درسؤال ۴۳، کدام یک از این روابط صادق است؟

الف) $E(X+Y) = E(X)$ ج) $E(X+Y) = E(XY)$

ب) $E(X+Y) = E(Y)$ د) $E(X+Y) = E(X) - E(Y)$

۴۹. در سؤال ۴۳، واریانس X و واریانس Y به ترتیب (از راست به چپ) کدام است؟

الف) ۵ , ۰ ج) ۵ , ۴

ب) ۲۵ , ۰ د) ۲۵ , ۴

۵۰. درسؤال ۴۳، $V(X - Y)$ با کدام یک از این موارد برابر است؟

الف) ۲۹ ج) $\frac{77}{3}$

ب) ۲۱ د) $\frac{107}{3}$

مسائل

۵۱. یک قرعه کشی با یک میلیون امتیاز قابل کسب دارای یک جایزهٔ اول به مبلغ ۵ میلیون ریال، ۹ جایزهٔ دوم هریک به مبلغ ۲۵۰ هزار ریال، ۹۰ جایزهٔ سوم هریک به مبلغ ۲۵ هزار ریال و ۹۰۰ جایزهٔ چهارم هریک به مبلغ ۲۵۰۰ ریال است. اگر در ازای هر ۱۵ ریال خرید، یک امتیاز به خریدار تعلق گیرد، این موارد را محاسبه کنید.

الف) امید ریاضی سود هر امتیاز از نظر خریدار.

ب) فرض کنید تنها ۸۰ درصد از امتیازها کسب شده باشند. امید ریاضی سود فروشگاه مجری این قرعه کشی.

۵۲. مدیر یک شیرینی‌پزی بنا به تجربه می‌داند که تعداد کیکهای شکلاتی‌ای که ممکن است در روزی معین بفروشد، متغیری تصادفی‌است که دارای توزیع احتمال $P(X = x) = \frac{1}{6}$ به ازای $x = 0, 1, 2, 3, 4, 5$ است. ضمناً می‌داند که هر کیکی که می‌فروشد هزار ریال سود دارد و هر کیکی که فروش نمی‌رود در اثر فاسد شدن ضرری برابر ۴۰۰ ریال خواهد داشت. با فرض اینکه

هر کیک را فقط همان روزی که پخته می‌شود می‌توان فروخت، سود مورد انتظار شیرینی‌پز را برای هریک از این روزها تعیین کنید.

الف) ۳کیک بپزد.

ب) ٤کیک بپزد.

ج) ٥کیک بپزد.

۵۳. تعداد فروش کت و شلوار فروشگاه لباسی در هر روز، همراه با احتمال آن، در این جدول آورده شده است.

تعداد فروش	۱	۲	۳	٤	٥	٦	۷	۸
احتمال	۰/۱۰	۰/۱٥	۰/۱٦	۰/۲۰	۰/۱۹	۰/۰۸	۰/۰۷	۰/۰٥

الف) این فروشگاه به طور متوسط در هر روز چند کت و شلوار می‌فروشد؟

ب) واریانس تعداد فروش چقدر است؟

٥٤. امتحانی مشتمل بر ٤ سؤال ۳گزینه‌ای است. فردی به صورت تصادفی به تمام سؤالات پاسخ می‌دهد. اگر X تعداد جوابهای درست باشد، این موارد را محاسبه کنید.

الف) تابع احتمال X

ب) امید ریاضی و واریانس X

ج) اگر معلمی مقیاس نمره را طبق تبدیل $Y = ۲۲/٥ X + ۱۰$ عوض کند، در این صورت امید ریاضی و واریانس Y را با دو روش حساب کنید.

٥٥. فرض کنید X متغیری تصادفی با میانگین μ و انحراف معیار σ باشد. میانگین و انحراف معیار Z را در هریک از این حالات حساب کنید.

الف) $Z = ۳X - ۲$

ب) $Z = \frac{X - ۲}{٤}$

ج) $Z = \frac{X - \mu}{\sigma}$

٥٦. آیا این تابع می‌تواند معرف یک تابع احتمال باشد؟ چرا؟

$$f(x) = \frac{x - ۳}{۸} \quad , \quad x = ۳, ٤, ٦, ۸$$

٥۷. فرض کنید احتمال اینکه یک ناو جنگی در هر شلیک به هدف بزند $\frac{۱}{٥}$ است. این احتمالات را در ٦ شلیک این ناو محاسبه کنید.

الف) دقیقاً ۲ بار به هدف بزنند.
ب) حداقل ۲ بار به هدف بزنند.
ج) حداکثر ۲ بار به هدف بزنند.

۵۸. شورای مدیران شرکتی مرکب از مدیر عامل و ٦ مدیر اداری، تدارکات، پرسنلی، تولید، بازاریابی و مالی است. هر موضوعی که طرح شود با اکثریت آرای این ۷ نفر درباره آن تصمیم گرفته می‌شود. فرض کنید موضوعی در این شورا طرح می‌شود و هریک از اعضا مستقلاً با احتمال ۰/٤ به موضوع مذکور رأی مثبت می‌دهند. چقدر احتمال دارد که موضوع تصویب شود؟

۵۹. با توجه به داده‌های سؤال ۵۸، فرض کنید مدیر عامل، طالب تصویب این موضوع است و از قبل با مدیر مالی صحبت کرده و مطمئن شده است که او هم به این موضوع رأی مثبت می‌دهد. در این صورت احتمال تصویب موضوع چقدر است؟

٦۰. اظهارنظر حسابرسان راجع به حسابهای شرکتی ممکن است قبول، مردود، عدم‌اظهارنظر یا اظهارنظر مشروط باشد. در سال پیش ۰/۲۰، ۰/۱۵، ۰/٤۰ و ۰/۲۵ نظرها به ترتیب قبول، مردود، عدم اظهارنظر و اظهارنظر مشروط بوده است. اگر فرض کنیم ٦ شرکت به صورت تصادفی انتخاب شده‌اند، هریک از این احتمالات چقدر است؟

الف) احتمال اینکه حسابهای ۲ شرکت قبول شده باشد.
ب) احتمال اینکه حسابهای ۲ شرکت قبول و ۲ شرکت نیز مردود شده باشد.
ج) احتمال اینکه حسابهای ۲ شرکت قبول، ۲ شرکت مردود، یک شرکت مشروط شده و راجع به حسابهای یک شرکت اظهار نظری نشده باشد.

٦۱. توزیع احتمال توأم دو متغیر تصادفی X و Y به این صورت است:

x \ y	۲۰	٤۰
۳	۰/٤	۰/۲
۵	۰/۲	۰/۱
۷	۰/۱	۰

الف) $E(X)$ و $E(Y)$ را حساب کنید.
ب) $E(XY)$ را حساب کنید.
ج) $\mathrm{Cov}(X, Y)$ را پیدا کرده، راجع به ارتباط این دو متغیر توضیح دهید.

۶۲. فرض کنید نسخهٔ دستنویس کتابی که ۵۰۰ صفحه دارد، دارای ۵۰ غلط است که به صورت تصادفی در سراسر کتاب پخش هستند. هر یک از این احتمالات را محاسبه کنید:
الف) فصلی از کتاب که ۳۰ صفحه دارد، دارای ۲ غلط یا بیشتر باشد.
ب) فصلی از کتاب که ۵۰ صفحه دارد، دارای ۲ غلط یا بیشتر باشد.
ج) یک صفحه انتخابی غلط نداشته باشد.

۶۳. فرض کنید ۵ مهندس و ۹ تکنیسین روی پروژه‌ای کار می‌کنند. اگر ۵ نفر را به صورت تصادفی انتخاب کنیم، احتمال اینکه دقیقاً دو مهندس انتخاب شود چقدر است؟

۶۴. فرض کنید که در مزرعهٔ بزرگی ۴۰ درصد کارگران ساعتی طالب رسمی شدن هستند. این موارد را در صورتی که ۱۰ نفر از این کارگران انتخاب شوند، محاسبه کنید:
الف) احتمال اینکه آنها طالب رسمی شدن باشند.
ب) میانگین و واریانس تعداد افرادی که طالب رسمی شدن هستند.

۶۵. در ظرفی ۵ مهرهٔ سفید، ۳ مهره قرمز و ۲ مهرهٔ سبز وجود دارد. اگر چهار مهره به طور تصادف بیرون آورده شود، احتمال اینکه ۲ مهره قرمز و حداقل یک مهرهٔ سفید بیرون بیاید، چقدر است؟

۶۶. هنگام تهیه یک فیلم سینمایی، احتمال اینکه بازیگری نقش خود را در هر دور فیلمبرداری درست بازی کند ۰/۳۵ است. احتمال اینکه در چهارمین دور فیلمبرداری برای اولین بار نقش خود را درست‌بازی کند، چقدر است؟

۶۷. فرض کنید احتمال اینکه یک اظهار نامهٔ مالیاتی به طور صحیح پر شود، فقط شامل خطاهایی به نفع مالیات‌دهنده، فقط شامل خطاهایی به نفع دولت، و یا شامل خطاهایی از هر دو نوع باشد، به ترتیب ۰/۶۰ ، ۰/۲۰ ، ۰/۱۰ و ۰/۱۰ است. احتمال اینکه بین ۱۲ تا از چنین اظهارنامه‌هایی که برحسب تصادف برای حسابرسی انتخاب شده‌اند، ۵ تا صحیح پرشده باشند، ۴ تا فقط شامل خطاهایی به نفع مالیات‌دهنده، ۲ تا شامل فقط خطاهایی به نفع دولت و یکی شامل خطاهایی از هر دو نوع باشد، چقدر است؟

۶۸. فرض کنید برای ساختن عمارتی.جدید، دو مرحلهٔ متوالی نقشه‌برداری و بنا کردن لازم است. زمان مورد نیاز بر حسب سال، برای تکمیل نقشه‌برداری (S) و بنا کردن (C)، دو متغیر تصادفی مستقل با این توزیعهای احتمال هستند:

$$P(S) = (۰/۸)(۰/۲)^{S-۱} \ , \ S = ۱, ۲, ۳, \ldots$$

$$P(C) = (۰/۵)(۰/۵)^{C-۲} \ , \ C = ۲, ۳, ۴, \ldots$$

بدیهی است که کل زمان ساختن (T) عبارت است از $T = S + C$. اگر انجام کار ساختمان

بیش از ٤ = T سال طول بکشد، پیمانکار جریمه می‌شود. احتمال جریمه شدن پیمانکار چقدر است؟

٦۹. مدیر یک مجموعهٔ آپارتمانی، ۳ یخچال جدید که فروشنده سالم بودن آنها را تضمین می‌کند، سفارش می‌دهد. طبق ضمانتنامه، در صورت معیوب بودن یخچال فروشنده باید آن را تعمیر کند. احتمال معیوب بودن هر یخچال ۲۰ درصد است.

الف) توزیع احتمال تعداد یخچالهای معیوب (X) را پیدا کرده، در جدولی بنویسید.

ب) میانگین و واریانس X چقدر است.

ج) فرض کنید هزینهٔ تعمیر، شامل کار مزدی ثابت (۵۰ هزار ریال) به اضافهٔ هزینه متغیر برای هر یخچال (۲۵ هزار ریال) باشد؛ یعنی:

$$C(x) = \begin{cases} ۰ & ، \quad x = ۰ \\ ۵۰۰۰۰ + ۲۵۰۰۰\,x & ، \quad x > ۰ \end{cases}$$

متوسط هزینهٔ تعمیر را پیدا کنید.

۷۰. طبق آمار سالیانه‌ای که اداره راهنمایی و رانندگی منتشر کرده است، از هر ۱۰۰ هزار نفر به طور متوسط ۳ نفر در اثر حوادث رانندگی کشته می‌شوند. این احتمالات را در شهری با ۲۰۰ هزار نفر جمعیت محاسبه کنید.

الف) ۵ نفر کشته شوند.

ب) کمتر از ۳ نفر کشته شوند.

ج) بین ٤ تا ۸ نفر کشته شوند (خود ٤ و ۸ را در نظر بگیرید).

پاسخنامهٔ سؤالات

۱) غ	۲) ص	۳) ص	٤) ص	۵) غ	٦) غ	۷) ص
۸) ص	۹) ص	۱۰) غ	۱۱) غ	۱۲) غ	۱۳) ص	۱٤) ص
۱۵) غ	۱٦) ص	۱۷) ص	۱۸) ص	۱۹) ص	۲۰) ص	۲۱) ص
۲۲) ص	۲۳) غ	۲٤) غ	۲۵) ص	۲٦) د	۲۷) د	۲۸) ج
۲۹) ج	۳۰) ب	۳۱) ج	۳۲) الف	۳۳) الف	۳٤) الف	۳۵) ج
۳٦) د	۳۷) ب	۳۸) الف	۳۹) د	٤۰) د	٤۱) ج	٤۲) ج
٤۳) ج	٤٤) د	٤۵) ب	٤٦) د	٤۷) ج	٤۸) ب	٤۹) د
۵۰) د						

فصل هفتم

توابع احتمال پیوسته

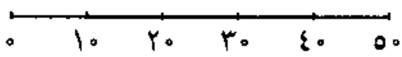
۷ـ۱ تابع چگالی احتمال و تابع توزیع متغیر تصادفی پیوسته

۷ـ۱ـ۱ متغیر تصادفی پیوسته

در فصل پیش درباره متغیرهای تصادفی گسسته و توابع احتمال خاص آن توضیح دادیم. در این فصل درباره متغیرهای تصادفی پیوسته و چند تا از توابع مهم[1] آن توضیح می‌دهیم.

فرض کنید علاقه‌مند به دانستن احتمال وقوع تصادف در یک اتوبان ۵۰ کیلومتری یا احتمال وقوع تصادف در نقطه یا فاصله‌ای مشخص در این اتوبان باشیم. ضمناً فرض کنید که احتمال وقوع تصادف در هر نقطه‌ای از این اتوبان با نقطهٔ دیگر برابر است. در این صورت فضای نمونه این آزمایش پیوستاری از نقاط است که روی ۰ تا ۵۰ کیلومتر واقع شده‌اند.

احتمال اینکه درهر فاصله‌ای به اندازهٔ x تصادفی رخ دهد برابر با $\frac{x}{۵۰}$ است؛ مثلاً احتمال اینکه تصادف در جایی بین کیلومتر دهم تا چهلم رخ دهد برابر $\frac{۳}{۵}$ است؛ زیرا فاصلهٔ این دو، برابر $۴۰ - ۱۰ = ۳۰$ است و کل فاصله ۵۰ کیلومتر است، پس احتمال آن $\frac{۳}{۵} = \frac{۳۰}{۵۰}$ خواهد بود؛ به عبارت دیگر خواهیم داشت:



$$P(۱۰ \le X \le ۴۰) = \frac{۳۰}{۵۰} = \frac{۳}{۵}$$

بدیهی است هرچه این فاصله کوتاهتر شود، احتمال مربوط به آن نیز کمتر می‌شود؛ مثلاً

---

[1]. در این فصل صرفاً به معرفی توابع یکنواخت و نمایی بسنده می‌کنیم. در فصول بعدی با توابع نرمال، t استیودنت، کای ـ مربع، و فیشر آشنا خواهید شد. برای اطلاع از سایر توزیعهای پیوسته به منبع شماره ۱۵ انتهای کتاب مراجعه شود.

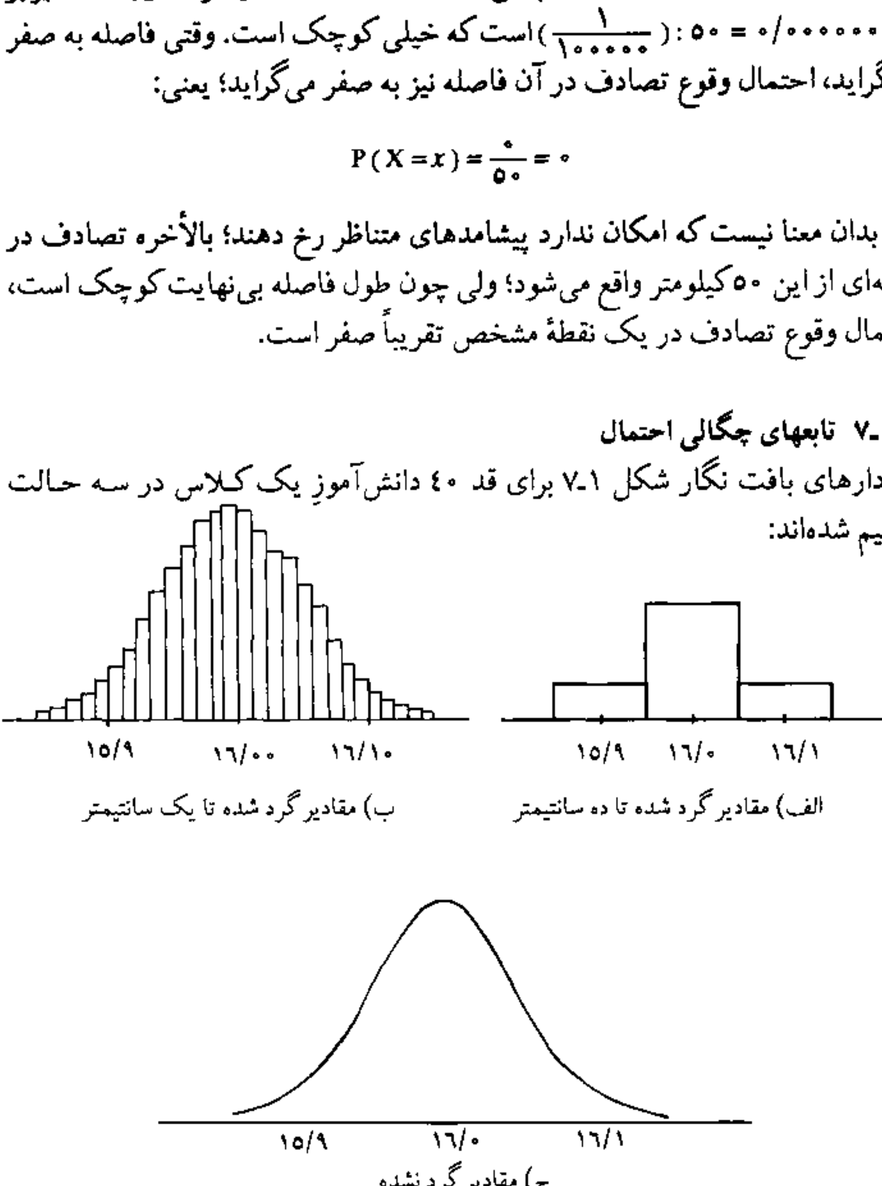
احتمال اینکه تصادف در فاصلهٔ کوچکی (مانند فاصله یک سانتیمتری) روی دهد برابر با ۰/۰۰۰۰۰۰۲ = ۵۰ : $\left(\frac{۱}{۱۰۰۰۰۰۰}\right)$ است که خیلی کوچک است. وقتی فاصله به صفر می‌گراید، احتمال وقوع تصادف در آن فاصله نیز به صفر می‌گراید؛ یعنی:

$$P(X=x)=\frac{۰}{۵۰}=۰$$

این بدان معنا نیست که امکان ندارد پیشامدهای متناظر رخ دهند؛ بالأخره تصادف در نقطه‌ای از این ۵۰ کیلومتر واقع می‌شود؛ ولی چون طول فاصله بی‌نهایت کوچک است، احتمال وقوع تصادف در یک نقطهٔ مشخص تقریباً صفر است.

۲ـ۱ـ۷ تابعهای چگالی احتمال

نمودارهای بافت نگار شکل ۱ـ۷ برای قد ۴۰ دانش‌آموزِ یک کلاس در سه حالت ترسیم شده‌اند:



شکل ۱ـ۷ تعریف احتمال برای متغیرهای پیوسته

واضح است که اگر مقادیر گرد نشوند یا به مقادیر بسیار کوچکی همچون یک صدم میلیمتر گرد شوند، بافت نگار توزیع احتمال از حالت گسسته به پیوسته تبدیل

می‌شود. تابعی که معرف این متغیر تصادفی پیوسته است تابع چگالی احتمال خوانده می‌شود که مساحت زیر منحنی آن برابر یک است. تابع چگالی احتمال را با f(x) نشان می‌دهیم. تابع چگالی احتمال را «تابع چگالی» و «چگالی احتمال» نیز می‌نامند. احتمال اینکه متغیر تصادفی پیوستهٔ X مقداری بین دو نقطهٔ a و b را بگیرد، برابر است با سطح زیر منحنی بین این دو نقطه؛ یعنی:

$$P(a \le X \le b) = \int_a^b f(x)\,dx$$

همانطور که پیش از این گفتیم، در متغیرهای پیوسته، احتمال اینکه متغیر تصادفی X دقیقاً یک مقدار مشخص را بگیرد، برابر صفر است؛ زیرا:

$$P(X = a) = P(a \le X \le a) = \int_a^a f(x)\,dx = ۰$$

با توجه به مطلب مذکور، روشن است که:

$$P(a < X < b) = P(a \le X < b) = P(a < X \le b) = P(a \le X \le b)$$

باید توجه داشت که اولاً احتمال، مقداری غیرمنفی است؛ یعنی $P(a \le X \le b) \ge ۰$ و ثانیاً سطح زیر منحنی برابر یک است؛ یعنی:

$$\int_{-\infty}^{+\infty} f(x)\,dx = ۱$$

بنابراین تابعی با مقادیر $f(x)$ را که روی مجموعهٔ اعداد حقیقی تعریف شده باشد، تابع چگالی احتمال می‌گوییم اگر:

$$f(x) \ge ۰$$

$$\int_{-\infty}^{+\infty} f(x)\,dx = ۱$$

مثال ۱ـ۷ متغیر تصادفی X با این تابع چگالی احتمال داده شده است:

$$f(x) = \begin{cases} \frac{۲}{۳}x + \frac{۱}{۳}K & ، \quad ۰ \le x \le ۱ \\ ۰ & ، \quad \text{در غیر این صورت} \end{cases}$$

می‌خواهیم این مقادیر را محاسبه کنیم: الف) مقدار K، ب) $P(0 \le X \le 0/5)$ و ج) $P(0/5 \le X \le 1)$.

الف) می‌دانیم که $\int_{-\infty}^{+\infty} f(x)\,dx = 1$؛ بنابراین خواهیم داشت:

$$P(+\infty < X < +\infty) = \int_{-\infty}^{+\infty} f(x)\,dx = \int_0^1 \left(\frac{2}{3}x + \frac{1}{3}K\right)dx$$

$$= \left[\frac{1}{3}x^2 + \frac{1}{3}kx\right]_0^1 = 1 \Rightarrow k = 2$$

ب) $$P(0 \le X \le 0/5) = \int_0^{0/5}\left(\frac{2}{3}x + \frac{2}{3}\right)dx = \left[\frac{1}{3}x^2 + \frac{2}{3}x\right]_0^{0/5} = \frac{5}{12}$$

ج) $$P(0/5 \le X \le 1) = \int_{0/5}^{1}\left(\frac{2}{3}x + \frac{2}{3}\right)dx = \left[\frac{1}{3}x^2 + \frac{2}{3}x\right]_{0/5}^{1} = \frac{7}{12}$$

مثال ۲ـ۷ در یک شرکت مواد غذایی میزان ربّی که در قوطیهای ۳۳۰ گرمی ریخته می‌شود، دارای این تابع چگالی است:

$$f(x) = \begin{cases} \frac{1}{12} & ، 324 < x < 336 \\ 0 & \text{، در غیر این صورت} \end{cases}$$

فرض کنید یک قوطی به طور تصادفی انتخاب می‌شود، می‌خواهیم این احتمالات را محاسبه کنیم:

الف) حداکثر ۳۳۲ گرم رب داشته باشد.
ب) حداقل ۳۲۵ گرم رب داشته باشد.
ج) بین ۳۳۰/۵ تا ۳۳۵/۵ گرم رب داشته باشد.
د) دقیقاً ۳۳۰/۱۲ گرم رب داشته باشد.

الف) $$P(X \le 332) = \int_{-\infty}^{332} f(x)\,dx = \int_{324}^{332}\frac{1}{12}dx = \left[\frac{1}{12}x\right]_{324}^{332} = \frac{8}{12} = \frac{2}{3}$$

ب) $$P(X \ge 325) = \int_{325}^{\infty} f(x)\,dx = \int_{325}^{336}\frac{1}{12}dx = \left[\frac{1}{12}x\right]_{325}^{336} = \frac{11}{12}$$

ج) $P(330/5 \le X \le 335/5) = \int_{330/5}^{335/5} \frac{1}{12} dx = [\frac{1}{12}x]_{330/5}^{335/5} = \frac{5}{12}$

د) $P(X = 330/12) = \int_{330/12}^{330/12} \frac{1}{12} dx = [\frac{1}{12}x]_{330/12}^{330/12} = \frac{0}{12} = 0$

مثال ۳ـ۷ متغیر تصادفی X با این تابع چگالی احتمال مفروض است:

$$f(x) = \begin{cases} 2e^{-2x} & ، \ x > 0 \\ 0 & ، \ \text{در غیر این صورت} \end{cases}$$

می‌خواهیم این مقادیر را محاسبه کنیم:

الف) $P(1 \le X \le 5)$

ب) $P(-5 \le X \le 2/75)$

الف) $P(1 \le X \le 5) = \int_1^5 2e^{-2x} dx = [-e^{-2x}]_1^5 = -e^{-10} + e^{-2} = 0/135$

ب) $P(-5 \le X \le 2/75) = \int_0^{2/75} 2e^{-2x} dx = [-e^{-2x}]_0^{2/75} = -e^{-5/5} + e^0 = 0/996$

۳ـ۱ـ۷ تابع توزیع متغیر تصادفی پیوسته

در بسیاری از موارد مایلیم احتمال اینکه مقدار متغیر تصادفی پیوسته از عدد حقیقی x کوچکتر یا مساوی آن است را بدانیم. همان‌طور که در متغیرهای تصادفی گسسته مطرح کردیم، در این موارد باید از تابع توزیع که گاهی توزیع تجمعی گفته می‌شود و آن را با $F(x)$ نشان می‌دهیم، استفاده کنیم؛ یعنی:

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\, dx$$

بدیهی است که $F(-\infty) = 0$ و $F(+\infty) = 1$ است.

پیش از این گفتیم احتمال اینکه متغیر تصادفی پیوستهٔ X مقداری بین دو عدد حقیقی a و b را بگیرد $(a \le b)$ به این صورت محاسبه می‌شود:

$$P(a \le X \le b) = \int_a^b f(x)\, dx$$

بدیهی است این احتمال را به این صورت نیز می‌توان محاسبه کرد:

$$P(a \le X \le b) = F(b) - F(a)$$

مثال ۴ـ۷ با در نظرگرفتن چگالی احتمال مثال ۳ـ۷ می‌خواهیم از تابع توزیع آن برای محاسبهٔ مجدد قسمت الف و ب استفاده کنیم.

ابتدا تابع توزیع را به دست می‌آوریم

$$F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{0}^{x} 2e^{-2x}\,dx = \left[-e^{-2x}\right]_{0}^{x} = 1 - e^{-2x}$$

بنابراین می‌توان نوشت:

$$F(x) = \begin{cases} 0 & , \; x \le 0 \\ 1 - e^{-2x} & , \; x > 0 \end{cases}$$

الف) $P(1 \le X \le 5) = F(5) - F(1) = 0/999 - 0/864 = 0/135$

ب) $P(-5 \le X \le 2/75) = F(2/75) - F(-5) = 0/996 - 0 = 0/996$

تمرین

۱. اگر تابع چگالی متغیر تصادفی X به این صورت باشد:

$$f(x) = \begin{cases} 6x(1-x) & , \; 0 < x < 1 \\ 0 & , \; \text{در غیر این صورت} \end{cases}$$

این موارد را محاسبه کنید.

الف) $P(X > \frac{1}{3})$ ج) $P(X = \frac{1}{3})$

ب) $P(\frac{1}{3} < X < \frac{3}{4})$ د) $P(\frac{1}{3} \le X \le \frac{1}{2})$

۲. اگر تابع توزیع متغیر تصادفی X به این صورت باشد:

$$F(x) = \begin{cases} 1 - \frac{4}{x^2} & , \; x > 2 \\ 0 & , \; \text{در غیر این صورت} \end{cases}$$

این موارد را محاسبه کنید:

الف) $P(X < ۵)$

ب) $P(۱ < X < ۱/۵)$

۳. تابع چگالی احتمال متغیر تصادفی X به این صورت است:

$$f(x) = \begin{cases} \frac{K}{\sqrt{x}} & ، \quad ۰ < x < ۴ \\ ۰ & \text{، در غیر این صورت} \end{cases}$$

این موارد را محاسبه کنید:

الف) مقدار K

ب) تابع توزیع این متغیر تصادفی

ج) $P(X < ۲)$

۲ـ۷ امید ریاضی و واریانس متغیر تصادفی پیوسته

امید ریاضی (میانگین) و واریانس متغیر تصادفی گسستهٔ X را در فصل پیش به این صورت تعریف کردیم:

$$E(X) = \sum_{x} x \,.\, f(x)$$

$$V(X) = \sum_{x} [x - E(X)]^{۲} f(x)$$

در مورد متغیرهای تصادفی پیوسته، برای محاسبهٔ میانگین باید متغیر تصادفی را در تابع چگالی خود ضرب و سپس به ازای مقادیر ممکن متغیر، انتگرال‌گیری کرد؛ یعنی:

$$E(X) = \int_{-\infty}^{+\infty} x \,.\, f(x)\, dx$$

و برای محاسبهٔ واریانس، باید متغیر تصادفی را از میانگین خود کم کرد و حاصل را به توان دوم رساند و سپس در تابع چگالی خود ضرب و نهایتاً به ازای مقادیر ممکن متغیر انتگرال‌گیری کرد؛ یعنی:

$$V(X) = \int_{-\infty}^{+\infty} [x - E(X)]^{۲} f(x)\, dx$$

مثال ۵ـ۷ این تابع چگالی مفروض است:

$$f(x) = \begin{cases} \frac{۱}{۳} & ، \quad ۱ < x < ۴ \\ ۰ & \text{، در غیر این صورت} \end{cases}$$

می‌خواهیم امید ریاضی و واریانس را پیدا کنیم.

$$E(X) = \int_{-\infty}^{+\infty} x \, . \, f(x) \, dx = \int_{۱}^{۴} x \left(\frac{۱}{۳}\right) dx = \left[\frac{۱}{۶} x^{۲}\right]_{۱}^{۴} = \frac{۱۵}{۶} = ۲/۵$$

$$V(X) = \int_{-\infty}^{+\infty} [x - E(X)]^{۲} f(x) dx = \int_{۱}^{۴} (x - ۲/۵)^{۲} \left(\frac{۱}{۳}\right) dx = \frac{۱}{۳}\left[\frac{۱}{۳} x^{۳} - \frac{۵}{۲} x^{۲} + ۶/۲۵ x\right]_{۱}^{۴} = ۰/۷۵$$

مثال ۶ـ۷ اگر سود شرکتی را متغیری پیوسته در نظر بگیریم که دارای این چگالی احتمال است:

$$f(x) = \begin{cases} \frac{۲}{۶۳}(x + ۳) & ، \quad -۲ < x < ۵ \\ ۰ & \text{، در غیر این صورت} \end{cases}$$

می‌خواهیم محاسبه کنیم سود مورد انتظار این شرکت چقدر است.

$$E(X) = \int_{-\infty}^{+\infty} x \, . \, f(x) \, dx = \int_{-۲}^{۵} x \, . \, \frac{۲}{۶۳}(x + ۳) \, dx = \frac{۲}{۶۳}\left[\frac{۱}{۳} x^{۳} + \frac{۳}{۲} x^{۲}\right]_{-۲}^{۵} = ۲/۴۰۷$$

به طور کلی اگر $Y = g(X)$ باشد، $E(Y)$ به این صورت محاسبه می‌شود:

$$E(Y) = \int_{-\infty}^{+\infty} g(x) \, . \, f(x) \, dx$$

اگر $Y = X$ باشد، $E(Y)$ میانگین متغیر تصادفی X و اگر $Y = [x - E(X)]^{۲}$ باشد، $E(Y)$ واریانس X محسوب می‌شود.

مثال ۷ـ۷ تابع چگالی احتمال متغیر تصادفی X به این صورت است:

$$f(x) = \begin{cases} \frac{۲x}{۲۵} & ، \quad ۰ < x < ۵ \\ ۰ & \text{، در غیر این صورت} \end{cases}$$

می‌خواهیم این موارد را محاسبه کنیم.

الف) $E(X)$، گشتاور مرتبهٔ اول نسبت به صفر

ب) $E(X^{2})$، گشتاور مرتبهٔ دوم نسبت به صفر

ج) $E(X^{3})$، گشتاور مرتبهٔ سوم نسبت به صفر

د) $E(X^{4})$، گشتاور مرتبهٔ چهارم نسبت به صفر

الف) $$E(X)=\int_{-\infty}^{+\infty} x.f(x)\,dx=\int_{0}^{5} x.\left(\frac{2x}{25}\right)dx=\left[\frac{2x^{3}}{75}\right]_{0}^{5}=\frac{10}{3}$$

ب) $$E(X^{2})=\int_{-\infty}^{+\infty} x^{2}.f(x)\,dx=\int_{0}^{5} x^{2}.\left(\frac{2x}{25}\right)dx=\left[\frac{x^{4}}{50}\right]_{0}^{5}=\frac{25}{2}$$

ج) $$E(X^{3})=\int_{-\infty}^{+\infty} x^{3}.f(x)\,dx=\int_{0}^{5} x^{3}.\left(\frac{2x}{25}\right)dx=\left[\frac{2x^{5}}{125}\right]_{0}^{5}=50$$

د) $$E(X^{4})=\int_{-\infty}^{+\infty} x^{4}.f(x)\,dx=\int_{0}^{5} x^{4}.\left(\frac{2x}{25}\right)dx=\left[\frac{x^{6}}{75}\right]_{0}^{5}=\frac{625}{3}$$

تمرین

۱. میانگین و واریانس متغیر تصادفی X با این تابع چگالی احتمال را پیدا کنید.

$$f(x)=\begin{cases} 6x(1-x) & ، \quad 0<x<1 \\ 0 & ، \quad \text{در غیر این صورت} \end{cases}$$

۲. میانگین متغیر تصادفی X با این تابع چگالی احتمال را پیدا کنید (به انتگرال‌گیری جزء به جزء نیاز است).

$$f(x)=\begin{cases} 2e^{-2x} & ، \quad x>0 \\ 0 & ، \quad \text{در غیر این صورت} \end{cases}$$

۳. اگر تابع چگالی متغیر تصادفی X به این صورت باشد:

$$f(x)=\begin{cases} -Kx & ، \quad -2<x<0 \\ Kx & ، \quad 0\le x<4 \\ 0 & ، \quad \text{در غیر این صورت} \end{cases}$$

این مقادیر را محاسبه کنید:

الف) مقدار K
ج) تابع توزیع X

ب) امید ریاضی X
د) $P(-۱ < X < ۲/۵)$

۷ـ۳ توزیع یکنواخت

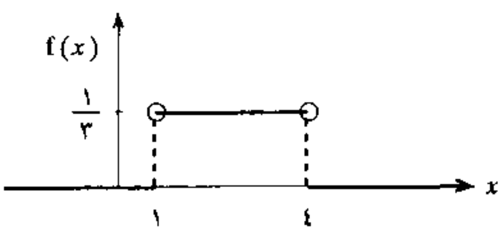
یکی از ساده‌ترین و در عین حال مهمترین توزیعها، توزیع احتمال یکنواخت است. در ابتدای همین فصل برای بیان متغیر تصادفی پیوسته مثالی آورده شد ـ وقوع تصادف در اتوبان ۵۰ کیلومتری ـ که توزیع آن توزیع یکنواخت بود. در اینجا به تعریف این چگالی احتمال و میانگین و واریانس آن می‌پردازیم.

متغیر پیوستهٔ X را در نظر بگیرید که مقادیر خود را بین دو نقطه α و β انتخاب می‌کند و $\alpha < \beta$ است. اگر احتمال وقوع X در فاصله‌های هم‌اندازه، در فاصلهٔ α و β برابر باشد، چگالی احتمال مربوط به آن یکنواخت خواهد بود.

مثال ۷ـ۸ این تابع چگالی احتمال یکنواخت مفروض است:

$$f(x) = \begin{cases} \frac{۱}{۳} & ، \quad ۱ < x < ٤ \\ ۰ & \text{، در غیر این صورت} \end{cases}$$

الف) نمودار تابع چگالی را رسم کنید.
ب) تابع توزیع را پیدا کنید.
ج) نمودار تابع توزیع را رسم کنید.
الف)



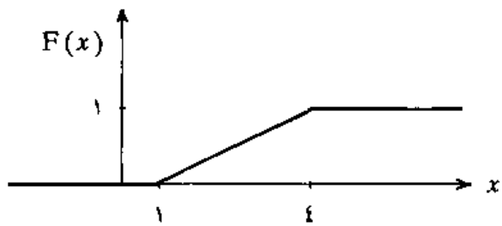
ب) برای پیدا کردن تابع توزیع، باید از تابع چگالی آن انتگرال بگیریم.

$$F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{۱}^{x} \frac{۱}{۳}\,dx = \left[\frac{۱}{۳}x\right]_{۱}^{x} = \frac{۱}{۳}(x-۱)$$

بنابراین:

$$F(x)=\begin{cases}0 & ، x\le 1\\ \frac{1}{3}(x-1) & ، 1<x<4\\ 1 & ، x\ge 4\end{cases}$$

ج)



برای پیدا کردن میانگین و واریانس چگالی یکنواخت طبق معمول از انتگرال استفاده می‌کنیم:

$$E(X)=\int_{-\infty}^{+\infty} x.f(x)\,dx=\int_{\alpha}^{\beta} x(\frac{1}{\beta-\alpha})\,dx=\frac{1}{2}(\alpha+\beta)$$

$$V(X)=\int_{-\infty}^{+\infty}[x-E(X)]^2 f(x)\,dx=\int_{\alpha}^{\beta}(x-\frac{\alpha+\beta}{2})^2(\frac{1}{\beta-\alpha})\,dx=\frac{1}{12}(\beta-\alpha)^2$$

متغیر تصادفی پیوستهٔ X در صورتی دارای چگالی یکنواخت است که چگالی احتمال آن به این صورت باشد:

$$f(x)=\begin{cases}\frac{1}{\beta-\alpha} & ، \alpha<x<\beta\\ 0 & \text{، در غیر این صورت}\end{cases}$$

میانگین و واریانس چگالی یکنواخت به این صورت است:

$$E(X)=\frac{1}{2}(\alpha+\beta)\quad,\quad V(X)=\frac{1}{12}(\beta-\alpha)^2$$

مثال ۹ـ۷ سود شرکتی دارای توزیع یکنواخت بین ۳− و ۵ (به ۱۰ میلیون ریال)

است. می‌خواهیم موارد ذیل را محاسبه کنیم.

الف) تابع چگالی احتمال

ب) احتمال آنکه سود شرکت بین ۰ تا ۳/۵ باشد.

ج) متوسط سود مورد انتظار شرکت

د) واریانس سود شرکت

الف)
$$f(x)=\begin{cases}\frac{۱}{۸} & ، \ -۳<x<۵\\ ۰ & ، \ \text{در غیر این صورت}\end{cases}$$

ب)
$$P(۰\le X\le ۳/۵)=\int_{۰}^{۳/۵}\frac{۱}{۸}dx=\left[\frac{۱}{۸}x\right]_{۰}^{۳/۵}=۰/٤۳۷۵$$

ج)
$$E(X)=\frac{۱}{۲}(\alpha+\beta)=\frac{۱}{۲}(-۳+۵)=۱$$

د)
$$V(X)=\frac{۱}{۱۲}(\beta-\alpha)^{۲}=\frac{۱}{۱۲}(۵+۳)^{۲}=۵/۳۳$$

تمرین

۱. این توزیع یکنواخت را در نظر بگیرید:

$$f(x)=\begin{cases}\frac{۱}{\beta-\alpha} & ، \ \alpha<x<\beta\\ ۰ & ، \ \text{در غیر این صورت}\end{cases}$$

اگر $\alpha=-۱۰$ و $E(X)=۲$ باشد، این مقادیر را محاسبه کنید.

الف) مقدار β

ب) $V(X)$

ج) احتمال اینکه متغیر تصادفی X مقداری بین ٤ و ۵− را انتخاب کند.

د) احتمال اینکه X مقداری بزرگتر از ۵− را انتخاب کند.

۲. سود مورد انتظار شرکتی دارای توزیع یکنواخت با میانگین ٤۰ میلیون ریال و واریانس ۱۲۰ میلیون ریال است. احتمال اینکه سود آن بیشتر از ٤/۲۵ میلیون ریال باشد، چقدر است؟

۳. این چگالی احتمال یکنواخت را در نظر بگیرید:

$$f(x) = \begin{cases} \frac{1}{6} & ، \; -2 < x < 4 \\ 0 & ، \text{در غیر این صورت} \end{cases}$$

الف) نمودار تابع چگالی را رسم کنید.
ب) تابع توزیع را پیدا کنید.
ج) نمودار تابع توزیع را رسم کنید.

۷ـ۴ توزیع نمایی

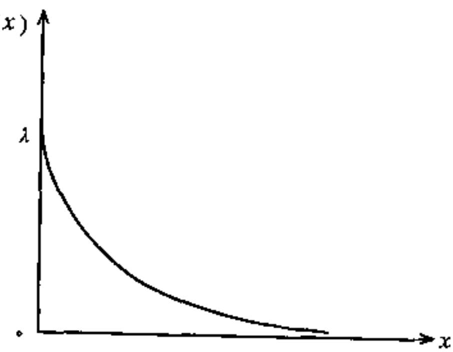
اگر تعداد موفقیتها یا ورودیها دارای توزیع پواسون باشد، زمان بین موفقیتها یا ورودیهای متوالی دارای «توزیعی نمایی منفی» است. چون زمان پیوسته است، توزیع نمایی منفی نیز توزیعی پیوسته است (از این به بعد توزیع نمایی‌منفی را صرفاً توزیع‌نمایی می‌نامیم).

تابع چگالی توزیع‌نمایی به این صورت است:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & ، \; x > 0 \\ 0 & ، \text{در غیر این صورت} \end{cases}$$

در این تابع، λ پارامتر توزیع است و متوسط تعداد موفقیتها یا ورودیها را در واحد زمان (که λ در آن تعریف شده) نشان می‌دهد.

نمودار تابع چگالی توزیع نمایی در شکل ۲ـ۷ رسم شده است.



شکل ۷ـ۲ نمودار تابع چگالی توزیع نمایی

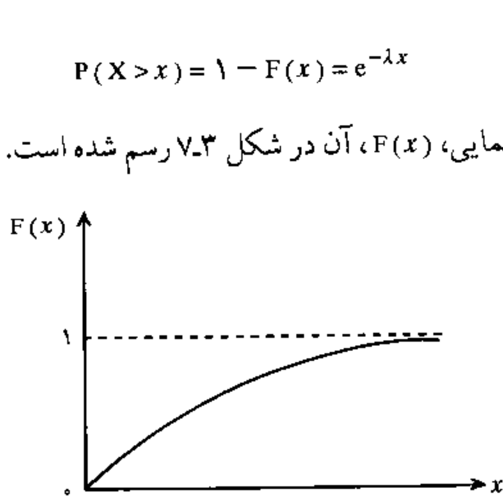
توجه کنید که تابع چگالی همواره نزولی است. تابع توزیع این متغیر تصادفی به این صورت محاسبه می‌شود:

$$F(x) = P(X \leq x) = \int_{\circ}^{x} \lambda e^{-\lambda x} dx = [-e^{-\lambda x}]_{\circ}^{x} = 1 - e^{-\lambda x}$$

به همین ترتیب، احتمال اینکه زمان بین دو موفقیت یا دو ورود بیش از x طول بکشد، برابر است با:

$$P(X > x) = 1 - F(x) = e^{-\lambda x}$$

نمودار تابع توزیع‌نمایی، $F(x)$، آن در شکل ۳-۷ رسم شده است.



شکل ۷.۳ نمودار تابع توزیع نمایی

میانگین و واریانس توزیع نمایی به این صورت محاسبه می‌شود:

$$E(X) = \int_{-\infty}^{+\infty} x . f(x) dx = \int_{\circ}^{+\infty} (x) \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

$$V(X) = \int_{-\infty}^{+\infty} [x - E(X)]^{2} f(x) dx = \int_{\circ}^{+\infty} (x - \frac{1}{\lambda})^{2} \lambda e^{-\lambda x} dx = \frac{1}{\lambda^{2}}$$

مثال ۱۰-۷ به مدیری به طور متوسط در هر ساعت ۵ بار تلفن، با توزیع پواسون، زده می‌شود. می‌خواهیم این مقادیر را محاسبه کنیم:

الف) به طور متوسط فاصلهٔ زمانی بین دو تلفن زدن متوالی

ب) احتمال اینکه فاصلهٔ زمانی بین دو تلفن زدن متوالی کمتر از ۱۰ دقیقه طول بکشد

ج) احتمال اینکه فاصلهٔ دو تلفن متوالی بیش از ۳ دقیقه طول بکشد

از آنجا که تعداد تلفنها دارای توزیع پواسون است، فاصلهٔ زمانی بین دو تلفن زدن متوالی دارای توزیع نمایی است و چون به طور متوسط ۵ تلفن در هر ساعت زده می‌شود، فاصلهٔ بین دو تلفن زدن متوالی $\frac{۱}{۵}$ ساعت است یعنی:

$$E(X)=\frac{۱}{\lambda}=\frac{۱}{۵}=۰/۲$$

ب) چون واحد زمانی ساعت است، ابتدا ۱۰ دقیقه را به ساعت تبدیل می‌کنیم:

$$\frac{۱۰}{۶۰}=\frac{۱}{۶}=۰/۱۶۷$$

بنابراین:

$$P(X\leq ۰/۱۶۷)=F(۰/۱۶۷)=۱-e^{-\lambda x}=۱-e^{-۵(۰/۱۶۷)}=۱-e^{-۰/۸۳۳}=۰/۵۶۵$$

ج) ابتدا ۳ دقیقه را به ساعت تبدیل می‌کنیم:

$$\frac{۳}{۶۰}=\frac{۱}{۲۰}=۰/۰۵$$

بنابراین:

$$P(X > ۰/۰۵) = ۱ - F(۰/۰۵) = e^{-\lambda x} = e^{-۵(۰/۰۵)} = ۰/۷۷۹$$

مثال ۱۱ـ۷ به طور متوسط یک روز در میان یک کشتی با توزیع نمایی به اسکله‌ای می‌رسد. اگر امروز یک کشتی رسیده باشد، می‌خواهیم بدانیم احتمال اینکه ۴ روز طول بکشد تا کشتی دیگری برسد، چقدر است.

در این مثال $\lambda=\frac{۱}{۲}$ است؛ بنابراین خواهیم داشت:

$$P(X>۴)=۱-F(۴)=e^{-\lambda x}=e^{-\frac{۱}{۲}(۴)}=e^{-۲}=۰/۱۳۵$$

تمرین

۱۰. مدت تعمیر ماشینی بر اساس توزیع‌نمایی با میانگین ۱/۵ ساعت است.

الف) احتمال اینکه مدت تعمیر کمتر از یک ساعت طول بکشد، چقدر است؟

ب) احتمال اینکه مدت تعمیر بین ۱ تا ۲ ساعت طول بکشد، چقدر است؟

۱۱. تعداد اشتباهات یک حروفچین در حروفچینی یک متن ۵۰۰ سطری ۲ عدد با توزیع پواسون است. احتمال اینکه در ۱۰۰ سطر اول اشتباهی نباشد، چقدر است؟

۱۲. بــه طــور مــتوسط هــر ۰/۵ دقـیقه ۱/۵ مشتری با توزیع پواسون به گیشهٔ پرداخت بـانکی مــراجــعه مــی‌کند. احتمــال ایــنکه اولیــن مشــتری بعد از ۲ دقیقه وارد شـود، چقدر است؟

۷ـ۵ خلاصه

همان‌طور که ذکر شد، اگر متغیر تصادفی پیوسته باشد، تابع معرف این متغیر، مشروط بر آنکه سطح زیر منحنی برابر یک باشد، تابع چگالی احتمال نامیده می‌شود. در این فصل نحوهٔ محاسبهٔ احتمال در یک توزیع پیوسته، میانگین و واریانس متغیر تصادفی پیوسته بیان شد.

از بین توزیعهای رایج پیوسته، توزیعهای یکنواخت و نمایی با تأکید بر شرایط استفاده از هر کدام مطرح شدند و مثالهایی در مورد کاربرد آنها ذکر شد ولی آشنایی با مهمترین توزیع پیوسته، یعنی توزیع نرمال، با توجه به اهمیت آن، به فصل بعد موکول شد.

۷ـ۶ سؤالات و مسائل

سؤالات دوگزینه‌ای

۱. مساحت زیر منحنی تابع چگالی احتمال برابر یک است. ص ☐ غ ☐

۲. احتمال اینکه متغیر تصادفی پیوسته‌ای دقیقاً مقدار مشخصی را بگیرد، بزرگتر از صفر است. ص ☐ غ ☐

۳. همواره برای هر متغیر پیوسته‌ای همچون X رابطهٔ $P(X < x) = P(X \le x)$ برقرار است. ص ☐ غ ☐

۴. در توابع پیوسته، احتمال اینکه متغیر X مقداری بین a و b $(a \le b)$ را انتخاب کند، برابر است با $\int_a^b f(x)\,dx$. ص ☐ غ ☐

۵. اگر $F(x)$ تابع توزیع باشد، همواره رابطهٔ $P(a \le X \le b) = F(b) - F(a)$ برقرار است. ص ☐ غ ☐

۶. میانگین (امید ریاضی) متغیر تصادفی پیوستهٔ X برابراست با $\int_{-\infty}^{+\infty} x \cdot f(x)\,dx$. ص ☐ غ ☐

۷. اگر α و β فاصله‌ای باشند که متغیر پیوستهٔ X با چگالی یکنواخت مقدار خود را در آن فاصله انتخاب کند، آنگاه میانگین (امید ریاضی) متغیر X برابر است با $\frac{\alpha + \beta}{2}$. ص ☐ غ ☐

۸. اگر تعداد موفقیتها (ورودیها) دارای توزیع پواسون باشد، زمان بین موفقیتهای متوالی دارای توزیع یکنواخت است. ص □ غ □

۹. تابع چگالی احتمال توزیع نمایی به این صورت است:

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & ، \ x>\circ \\ \circ & \text{، در غیر این صورت}\end{cases}$$

ص □ غ □

۱۰. اگر λ پارامتر توزیع نمایی باشد، واریانس توزیع نمایی برابر است با $(\frac{۱}{\lambda})^۲$. ص □ غ □

سؤالات چهارگزینه‌ای

۱۱. به ازای چه مقداری از K این تابع می‌تواند تابع چگالی احتمال باشد؟

$$f(x)=\begin{cases}Kx^۲ & ، \ \circ<x<۴ \\ \circ & \text{، در غیر این صورت}\end{cases}$$

الف) $\frac{۱}{۴}$ ج) $\frac{۳}{۶۴}$

ب) $\frac{۳}{۱۶}$ د) $\frac{۱}{۱۶}$

۱۲. اگر متغیر پیوسته X تنها مقادیر غیرمنفی را با تابع چگالی احتمال $f(x)=e^{-x}$ اختیار کند، احتمال اینکه X مقداری بین ۱ تا ۳ را بگیرد، برابر است با:

الف) ۰/۲۳۲۵ ج) ۰/۱۳۵۳

ب) ۰/۳۱۸۱ د) ۰/۴۶۵۰

۱۳. میانگین متغیر تصادفی X با این تابع چگالی چقدر است؟

$$f(x)=\begin{cases}\frac{۲}{۳} & ، \ -۱<x<\frac{۱}{۲} \\ \circ & \text{، در غیر این صورت}\end{cases}$$

الف) $\frac{۳}{۴}$ ج) $-\frac{۳}{۴}$

ب) $\frac{۱}{۴}$ د) $-\frac{۱}{۴}$

۱٤. واریانس متغیر تصادفی X در سؤال ۱۳ چقدر است؟

الف) $\frac{۱}{۱٦}$ ج) $\frac{۹}{٤}$

ب) $\frac{۳}{۱٦}$ د) $\frac{٤}{۹}$

۱۵. اگر $P(-۲ \leq X \leq ۸) = ۰/۷۱$ و $F(-۲) = ۰/۱۷$ باشد، $F(۸)$ با کدام یک از این موارد برابر است؟

الف) ۰/۸۸ ج) ۰/۷۷

ب) ۰/۵٤ د) ۰/٤۵

۱٦. کدام یک از این مقادیر همواره در یک چگالی احتمال نادرست است؟

الف) $F(۵) = ۱, F(۰) = ۰/۵$ ج) $F(۵) = ۰/۵, F(۰) = ۱$

ب) $F(۵) = ۱, F(۰) = ۱$ د) $F(۵) = ۰/۲۵, F(۰) = ۰$

۱۷. متغیر تصادفی X با توزیع نمایی با پارامتر $\lambda = ۳$ مفروض است. واریانس متغیر تصادفی X با کدام یک از این موارد برابر است؟

الف) ۳ ج) $\frac{۱}{۳}$

ب) ۹ د) $\frac{۱}{۹}$

۱۸. به طور متوسط با توزیع نمایی هر ۵ دقیقه یک مشتری به اتو بانکی مراجعه می‌کند. امید ریاضی تعداد افرادی که در هر دقیقه مراجعه می‌کنند با کدام یک از این موارد برابر است؟

الف) ۵ ج) ۱۲

ب) $\frac{۱}{۵}$ د) $\frac{۱}{۱۲}$

۱۹. در سؤال ۱۸، واریانس تعداد افرادی که در هر ۵ دقیقه مراجعه می‌کنند با کدام یک از این موارد برابر است؟

الف) $\frac{۱}{۵}$ ج) $\frac{۱}{۱۲}$

ب) $\frac{۱}{۲۵}$ د) $\frac{۱}{۱٤٤}$

۲۰. متغیر تصادفی X با این چگالی یکنواخت را در نظر بگیرید:

$$f(x) = \begin{cases} \frac{۱}{۷} & \text{،} \quad \alpha < x < \beta \\ ۰ & \text{، در غیر این صورت} \end{cases}$$

اگر امید ریاضی متغیر تصادفی X برابر ۱۰۰ باشد، α و β با کدام یک از این موارد برابرند؟

الف) $\alpha = ۹۳ , \beta = ۱۰۷$ ج) $\alpha = ۹۷ , \beta = ۱۰۳$

ب) $\alpha = ۹۳/۵ , \beta = ۱۰۷/۵$ د) $\beta = ۱۰۳/۵$ و $\alpha = ۹٦/۵$

مسائل

۲۱. تابع چگالی احتمال متغیر تصادفی X به این صورت است:

$$f(x) = \begin{cases} \frac{k}{\sqrt{x}} & ، \; ۰ < x < ۹ \\ ۰ & ، \text{در غیر این صورت} \end{cases}$$

این مقادیر را محاسبه کنید:

الف) مقدار k ج) امید ریاضی X

ب) $P (۱ < X < ٤)$

۲۲. تابع توزیع متغیر تصادفی X به این صورت است:

$$f(x) = \begin{cases} \frac{۱}{۱۰} & ، \; ۵ < x < ۱۵ \\ ۰ & ، \text{در غیر این صورت} \end{cases}$$

این مقادیر را محاسبه کنید:

الف) $P (۰ \leq X \leq ۱۰)$

ب) امید ریاضی و واریانس X

۲۳. این تابع چگالی را که به چگالی مثلثی معروف است در نظر بگیرید:

$$f(x) = \begin{cases} C(x-۱) & ، \; ۱ < x < ۲ \\ -\frac{۱}{۲}C(x-٤) & ، \; ۲ \leq x < ٤ \\ ۰ & ، \text{در غیر این صورت} \end{cases}$$

الف) مقدار C را پیدا کنید.

ب) مقدار $P (۱ \leq X \leq ۳)$

ج) تابع توزیع X را پیدا کنید.

د) میانگین و واریانس X را پیدا کنید.

ه) نمودار چگالی و تابع توزیع را در شکلی واحد رسم کنید.

۲٤. این چگالی یکنواخت را در نظر بگیرید و امید ریاضی و واریانس را حساب کنید.

$$f(x)=\begin{cases}\frac{۱}{٤} & ، \; -۲<x<۲ \\ \\ ۰ & ، \text{در غیر این صورت}\end{cases}$$

۲۵. این چگالی نمایی را در نظر بگیرید:

$$f(x)=\begin{cases}\frac{۱}{۲}\,e^{-x/۲} & ، \; x>۰ \\ \\ ۰ & ، \text{در غیر این صورت}\end{cases}$$

الف) تابع توزیع X را به دست آورید. ج) میانگین و واریانس X چقدر است؟

ب) احتمال اینکه X مقداری بین $\frac{۱}{۲}$ و $\frac{۳}{۲}$ را اختیار کند چقدر است؟

۲٦. مشخص شده است که تعداد خرابیهای ماشینی دارای توزیع پواسون با میانگین ۳ خرابی در ماه است (هر ماه را ۳۰ روز در نظر بگیرید).

الف) متوسط بین دو خرابی چند روز است؟

ب) احتمال اینکه در ۱۰ روز اول پس از سرویس، ماشین خراب شود چقدر است؟

۲۷. متوسط عمر مفید نوعی تایر اتومبیل دارای توزیع‌نمایی با ۷۵ هزار کیلومتر است. این احتمالات را محاسبه کنید:

الف) بین ۲۰ تا ٦۰ هزار کیلومتر بتوان با آن طی کرد.

ب) حداکثر ۵۰ هزار کیلومتر بتوان با آن طی کرد.

۲۸. تعداد دقایقی که ممکن است پروازی زودتر یا دیرتر انجام شود، متغیری تصادفی است که چگالی آن به این صورت است:

$$f(x)=\begin{cases}\frac{۱}{۲۸۸}(۳٦-x^{۲}) & ، \; -٦<x<٦ \\ \\ ۰ & ، \text{در غیر این صورت}\end{cases}$$

در این رابطه، مقادیر منفی، معرف زود انجام شدن پرواز و مقادیر مثبت، معرف تأخیر در پرواز است. این احتمالات را در مورد یکی از این پروازها محاسبه کنید:

الف) حداقل ۲ دقیقه زود انجام شود. ج) مدتی از یک تا سه دقیقه زود انجام شود.

ب) حداقل یک دقیقه دیر انجام شود. د) دقیقاً ۵ دقیقه دیر صورت گیرد.

۲۹. تخمین زده می‌شود که میانگین زمان منتهی به خرابی یک لامپ تلویزیون طبق توزیع‌نمایی ۳ سال است. شرکتی این لامپها را در طول اولین سال استفاده بیمه می‌کند. برای چند درصد از بیمه نامه‌ها خسارت خواهد پرداخت؟

۳۰. یکی از خاصیتهای جالب و منحصر به فرد توزیع‌نمایی خاصیت بی‌حافظه بودن است؛ یعنی $P(X > x + S / X > x) = P(X > S)$، مثلاً اگر لامپی دارای توزیع‌نمایی برای زمان منتهی به خرابی آن باشد و در زمان x ببینیم هنوز در حال کار کردن است، در این صورت عمر باقیمانده نیز دارای همان توزیع‌نمایی است که لامپ در زمان صفر داشته است. با این توضیح اگر عمر ترانزیستوری دارای توزیع‌نمایی با میانگین ۱۰ هزار ساعت باشد و این ترانزیستور قبلاً ۱۰ هزار ساعت کار کرده باشد، احتمال اینکه بعد از ۱۵ هزار ساعت خراب شود، چقدر است؟

پاسخنامهٔ سؤالات

۱) ص	۲) غ	۳) ص	٤) ص	۵) ص	٦) ص
۷) ص	۸) غ	۹) ص	۱۰) ص	۱۱) ج	۱۲) ب
۱۳) د	۱٤) ب	۱۵) الف	۱٦) ج	۱۷) د	۱۸) ب
۱۹) ب	۲۰) د				

فصل هشتم

توزیع نرمال

۸ـ۱ توزیع نرمال چیست؟

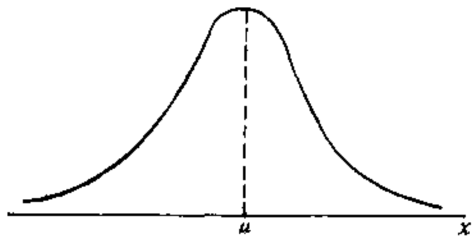
مهمترین توزیع پیوسته، توزیع نرمال است و بدین جهت مطالب این فصل صرفاً به این توزیع اختصاص داده می‌شود. توزیع نرمال، توزیعی زنگی شکل است که اولین بار در قرن هجدهم مورد بررسی و مطالعه قرار گرفت. لاپلاس و دومواور نقش چشمگیری در کشف آن داشتند و گوس[1] توزیع نرمال را به عنوان توزیع احتمال خطای اندازه گیریها با روش ریاضی به دست آورد. زمانی معتقد بودند که پدیده‌های واقعی باید دارای این توزیع باشد وگرنه در داده‌ها و روش جمع‌آوری آن باید تردید کرد؛ به همین دلیل آن را توزیع نرمال نام نهادند. بتدریج با بررسیهای بیشتر نادرستی این فکر مشخص شد و توزیعهای دیگری که در فصل پیش بعضی از آنها را بررسی کردیم، مطرح شد، با این حال این توزیع نقشی اساسی در آمار دارد و دارای کاربردهای وسیعی است؛ چرا که اولاً خیلی از پدیده‌های طبیعی دارای این توزیع هستند و ثانیاً شکل حدّی بسیاری از توزیعهای دیگر نیز نرمال است.

توزیع نرمال را می‌توان چنین تعریف کرد: «متغیر تصادفی پیوستهٔ X با میانگین μ و انحراف معیار σ دارای توزیع نرمال است اگر تابع چگالی آن به این صورت باشد:

$$f(x) = \frac{۱}{\sqrt{۲\pi\sigma^{۲}}}\ e^{-\frac{۱}{۲}[(x-\mu)/\sigma]^{۲}}$$

در این رابطه ...۳/۱۴۱۵۹ = π و ...۲/۷۱۸۲۸ = e است». شکل ۸ـ۱ نشان‌دهنده منحنی نرمال است.

1. Gauss

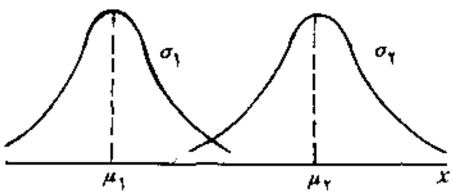


شکل ۸.۱ منحنی نرمال

به دلیل استفادهٔ زیاد از این توزیع، از این به بعد متغیر تصادفی X را که دارای توزیع نرمال با میانگین μ و انحراف معیار σ است به این صورت: $X \sim N(\mu,\sigma)$ نشان می‌دهیم و به این صورت می‌خوانیم: X دارای توزیع نرمال با میانگین μ و انحراف معیار σ است. دو پارامتر توزیع نرمال μ و σ است که با مشخص بودن آنها، توزیع دقیقاً مشخص و منحنی آن قابل ترسیم می‌شود. حال به بررسی تأثیر این دو پارامتر روی منحنی آن می‌پردازیم.

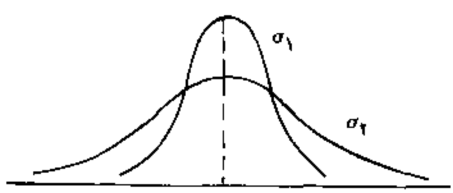
۸ـ۱ـ۱ تأثیر میانگین و انحراف معیار روی منحنی نرمال

شکل ۸ـ۲ دو منحنی نرمال را نشان می‌دهد که انحراف معیارشان برابر ($\sigma_1 = \sigma_2$) ولی میانگین اولی کوچکتر از میانگین دومی است ($\mu_1 < \mu_2$).



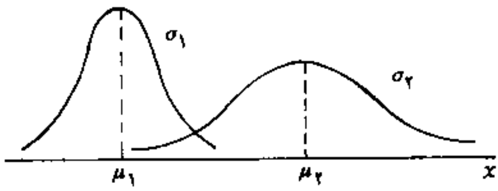
شکل ۸ـ۲ منحنیهای نرمال با $\sigma_1 = \sigma_2$ و $\mu_1 < \mu_2$

شکل ۸ـ۳ دو منحنی نرمال را نشان می‌دهد که میانگینشان یکسان ($\mu_1 = \mu_2$) ولی انحراف معیار دومی بزرگتر از اولی است ($\sigma_2 > \sigma_1$).



شکل ۸ـ۳ منحنیهای نرمال با $\mu_1 = \mu_2$ و $\sigma_2 > \sigma_1$

شکل ٤ـ۸ دو منحنی نرمال را نشان می‌دهد که نه میانگینشان با هم برابر است و نه انحراف معیارشان ($\mu_1 < \mu_2$ و $\sigma_1 < \sigma_2$).



شکل ٤.۸ منحنیهای نرمال با $\mu_1 < \mu_2$ و $\sigma_1 < \sigma_2$

بنابراین هر چقدر میانگین افزایش یابد، منحنی به سمت راست انتقال می‌یابد و هر چقدر انحراف معیار افزایش یابد منحنی کوتاهتر می‌شود.

۸ـ۱ـ۲ خصوصیات توزیع نرمال

توزیع نرمال دارای خصوصیات مهمی به شرح ذیل است (توجه داشته باشید که دو خصوصیت اول مربوط به تمام چگالیهای احتمال است).

۱. سطح زیرمنحنی بالای محور X ها برابر ۱ است؛ به بیان دیگر $\int_{-\infty}^{+\infty} f(x)\,dx = 1$ است.

۲. به ازای تمام مقادیر X، مقدار $f(x)$ بزرگتر یا مساوی صفر است؛ به بیان دیگر به ازای تمام X ها، $f(x) \geq 0$ خواهد بود.

۳. حداکثر مقدار تابع در $X = \mu$ حاصل می‌شود؛ به بیان دیگر در $X = \mu$ اولاً $f'(x) = 0$ ثانیاً $f''(x) < 0$ خواهد بود.

٤. تابع حول میانگین، μ، متقارن است؛ به بیان دیگر $f(x+\mu) = f(-x+\mu)$ است.

۵. امید ریاضی و واریانس X به ترتیب μ و σ^2 است؛ به بیان دیگر $E(X) = \int_{-\infty}^{+\infty} x\,f(x)\,dx = \mu$ و $V(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,dx = \sigma^2$ است.

٦. با دورتر شدن از μ، چه در سمت راست و چه در سمت چپ، منحنی به محور X ها نزدیکتر می‌شود، ولی هیچ‌گاه به صفر نمی‌رسد؛ به بیان دیگر $\lim_{x \to +\infty} f(x) = 0$ و $\lim_{x \to -\infty} f(x) = 0$ است.

۷. در این توزیع میانگین، میانه و مد با هم برابرند.

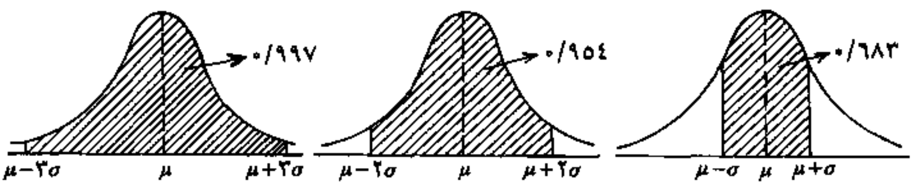
۸. احتمال فاصله‌ای به اندازهٔ «۱ انحراف معیار در هر طرف میانگین»، برابر ۰/۶۸۳ ، «۲ انحراف معیار در هر طرف میانگین» برابر ۰/۹۵٤؛ و «۳ انحراف معیار در هر طرف میانگین»، برابر ۰/۹۹۷ است؛ به بیان ریاضی:

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = ۰/۶۸۳$$

$$P(\mu - ۲\sigma \leq X \leq \mu + ۲\sigma) = ۰/۹۵٤$$

$$P(\mu - ۳\sigma \leq X \leq \mu + ۳\sigma) = ۰/۹۹۷$$

این مفاهیم در شکل ۵ـ۸ نمایش داده می‌شود.



شکل ۵ ـ۸ احتمال فاصله‌ای به اندازهٔ ۱ انحراف معیار، ۲ انحراف معیار و ۳ انحراف معیار در هر دو طرف میانگین

پیش از این گفتیم که منحنی به ازای هیچ مقداری از X به صفر نمی‌رسد، ولی از آنجا که مساحت سطوح انتهایی خارج از فاصلهٔ $(\mu - ۳\sigma, \mu + ۳\sigma)$ اندک است، معمولاً نمایش هندسی آن در دو سر این نقاط پایان می‌پذیرد.

۸ـ۱ـ۳ توزیع نرمال استاندارد

تابع چگالی احتمال نرمال به گونه‌ای است که محاسبهٔ احتمال مورد نظر از آن، به این دلیل که نمی‌توان از این تابع به سادگی انتگرال گرفت، کار مشکلی است؛ بنابراین جدولی تهیه شده که فقط برای توزیع نرمال با میانگین صفر و واریانس یک قابل استفاده است. حال اگر بتوانیم به جای متغیر $X \sim N(\mu,\sigma)$ از متغیر نرمالی استفاده می‌کنیم که میانگین آن صفر و واریانس آن یک باشد، می‌توانیم برای پیدا کردن احتمال موردنظر به جدول ۳ پیوست مراجعه کنیم.

گفتیم که تابع چگالی منحنی نرمال چنین است:

$$f(x) = \frac{۱}{\sqrt{۲\pi\sigma^۲}}\, e^{-\frac{۱}{۲}[(x-\mu)/\sigma]^۲} \quad \text{برای} \quad -\infty < x < \infty$$

در این تابع، متغیر تصادفی X دارای میانگین μ و واریانس $\sigma^۲$ است. حال اگر متغیری همچون Z را به این صورت تعریف کنیم: $Z = \frac{X-\mu}{\sigma}$؛ آنگاه تابع چگالی به این صورت خلاصه می‌شود:

$$\varphi(z) = \frac{۱}{\sqrt{۲\pi}} e^{-\frac{۱}{۲}z^۲}$$

این متغیر دارای توزیع نرمال با میانگین صفر و واریانس یک به شرح ذیل است:

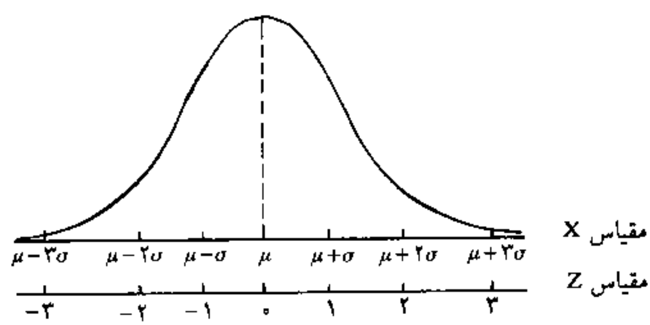
$$E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{۱}{\sigma}E(X-\mu) = \frac{۱}{\sigma}[E(X) - E(\mu)] = \frac{۱}{\sigma}(\mu - \mu) = ۰$$

$$V(Z) = V\left(\frac{X-\mu}{\sigma}\right) = \frac{۱}{\sigma^۲}V(X-\mu) = \frac{۱}{\sigma^۲}[V(X) + V(\mu)] = \frac{۱}{\sigma^۲}(\sigma^۲ + ۰) = ۱$$

به این توزیع، توزیع نرمال استاندارد می‌گوییم؛ بنابراین می‌توان گفت هر توزیع متغیر تصادفی نرمالی همانند Z که دارای میانگین صفر و واریانس یک باشد، توزیع نرمال استاندارد است که به زبان ریاضی چنین نوشته می‌شود: $Z \sim N(۰, ۱)$.

با داشتن میانگین و واریانس هر متغیری ـ در صورتی که توزیع آن نرمال باشد ـ می‌توان ابتدا آن را به توزیع نرمال استاندارد تبدیل کرد و سپس با مراجعه به جدول ۳ پیوست احتمال آن را پیدا کرد.

با تبدیل متغیر تصادفی X به Z مقیاس تغییر می‌کند. در شکل ٦ـ۸ این دو مقیاس نشان داده شده‌اند.



شکل ٦ـ۸ مقایسه مقیاس متغیر تصادفی نرمال X با متغیر تصادفی نرمال استاندارد Z

۲ـ۸ استفادهٔ مستقیم از جدول توزیع نرمال استاندارد

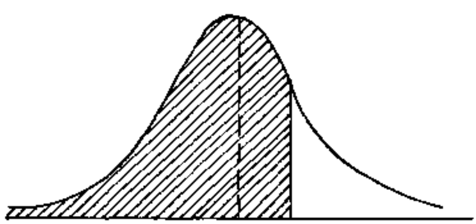
در استفادهٔ مستقیم از جدول توزیع نرمال استاندارد، هدف پیدا کردن احتمال پیشامدی مشخص است. در این‌گونه استفاده، ابتدا متغیر تصادفی نرمال را به متغیر تصادفی نرمال استاندارد تبدیل و سپس احتمال مورد نظر را پیدا می‌کنیم.

جدول ۳ پیوست، احتمال تجمعی را (احتمال اینکه متغیر نرمال استاندارد مقداری بین $-\infty$ تا Z را بگیرد) نشان می‌دهد. شایان توجه است که احتمال اینکه متغیر تصادفی Z مقداری کمتر از ۳/۵۹- را بگیرد، نزدیک به صفر است (۰/۰۰۰۲)؛ از این رو مقادیر کمتر از ۳/۵۹- در آن آورده نشده و احتمال اینکه مقداری کمتر از ۳/۵۹+ را بگیرد، نزدیک به یک است (۰/۹۹۹۸)؛ بنابراین مقادیر بزرگتر از ۳/۵۹ نیز در آن آورده نشده است. متغیر تصادفی نرمال استانداردِ Z، متغیری پیوسته است و بین این دو مقدار (۳/۵۹+ ، ۳/۵۹-) بی‌نهایت عدد حقیقی وجود دارد که نمایش همهٔ آنها امکانپذیر نیست؛ بنابراین واحد افزایش آن یک صدم است یعنی در جدول، احتمال مقادیر کوچکتر از ۳/۵۹- ، ۳/۵۸- ، ... ، ۳/۵۹ آورده شده است.

مثال ۱ـ۸ می‌خواهیم مقادیر هریک از این احتمالات راپیدا کنیم:
الف) $P(Z \leq ۱/۲۵)$ ، ب) $P(Z \leq -۰/۴۰)$ ، ج) $P(Z \geq ۱/۵۹)$ ، د) $P(-۱/۱۰ < Z < ۲/۷٦)$

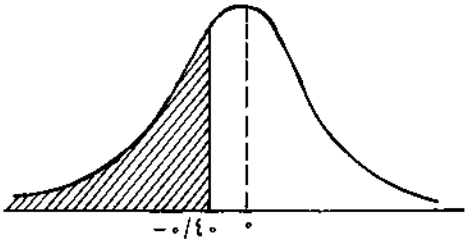
در محاسبهٔ هر احتمالی در توزیع نرمال، بهتر است منحنی نرمالی برای آن رسم کنیم و سطح مورد نظر را به صورت تقریبی نشان دهیم. این کار، پیدا کردن احتمال مورد نظر را ساده‌تر و به درک بهتر مسأله کمک می‌کند.

الف) هدف پیدا کردن سطح هاشور خورده‌است.



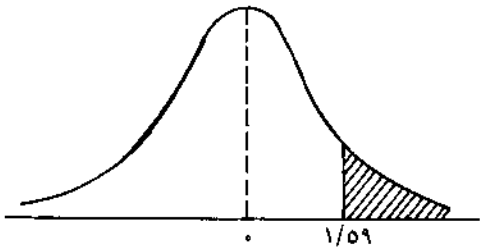
با مراجعه به جدول ۳ پیوست، در ستون اول از سمت چپ مقدار ۱/۲ را پیدا می‌کنیم و در تقاطع این سطر با ستونی که بالای آن مقدار ۰/۰۵ نوشته شده است، احتمال مورد نظر را پیدا می‌کنیم (مقدار ۰/۸۹٤٤)؛ بنابراین $P(Z \leq ۱/۲۵) = ۰/۸۹٤٤$ خواهد بود.

ب)



$P(Z \leq -۰/٤۰) = ۰/۳٤٤٦$

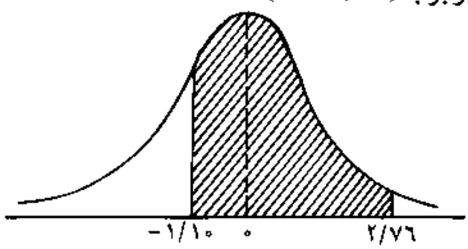
ج)



جدول نرمال استاندارد، احتمال مقادیر کوچکتر یا مساوی Z را نشان می‌دهد. برای محاسبهٔ احتمال مقادیر بزرگتر یا مساوی Z می‌توان از قانون متمم استفاده کرد؛ یعنی:

$$\begin{aligned} P(Z \geq ۱/۵۹) &= ۱ - P(Z \leq ۱/۵۹) \\ &= ۱ - ۰/۹٤٤۱ \\ &= ۰/۰۵۵۹ \end{aligned}$$

د) برای محاسبهٔ احتمال مورد نظر، می‌توان $P(Z \leq ۲/۷٦)$ را محاسبه و سپس مقدار $P(Z \leq -۱/۱۰)$ را از آن کم کرد. توجه کنید که چون توزیع نرمال Z، پیوسته است، $P(Z \leq ۲/۷٦)$ برابر با $P(Z < ۲/۷٦)$ است.



$$\begin{aligned} P(-۱/۱۰ < Z < ۲/۷٦) &= P(Z < ۲/۷٦) - P(Z < -۱/۱۰) \\ &= ۰/۹۹۷۱ - ۰/۱۳۵۷ \\ &= ۰/۸٦۱٤ \end{aligned}$$

مثال ۲ـ۸ توزیع نرمالی با $\mu = ۳۰$ و $\sigma = ۹$ را در نظر می‌گیریم، می‌خواهیم بدانیم احتمال اینکه متغیر تصادفی X مقداری بین ۲۴ تا ۴۳ را بگیرد، چقدر است.

با تبدیل متغیر X به Z (با استفاده از رابطهٔ $Z = \frac{X-\mu}{\sigma}$) می‌توان به جای $P(x_1 \leq X \leq x_2)$ مقدار $P(z_1 \leq Z \leq z_2)$ را محاسبه و از جدول پیدا کرد؛ یعنی:

$$P(۲۴ \leq X \leq ۴۳) = P(\frac{۲۴-۳۰}{۹} \leq Z \leq \frac{۴۳-۳۰}{۹})$$

$$= P(-۰/۶۷ \leq Z \leq ۱/۴۴)$$

$$= P(Z \leq ۱/۴۴) - P(Z \leq -۰/۶۷)$$

$$= ۰/۹۲۵۱ - ۰/۲۵۱۴$$

$$= ۰/۶۷۳۷$$

مثال ۳ـ۸ دستگاه پرکننده شیشه‌های آبلیمو طوری تنظیم شده است که فقط ۳۳۰ گرم آبلیمو را داخل هر شیشه بریزد، با وجود این میزان آبلیمویی که وارد هر شیشه می‌شود دارای توزیع نرمال با میانگین ۳۳۰ گرم و انحراف معیار ۵ گرم است. می‌خواهیم این احتمالات را محاسبه کنیم:

الف) شیشه‌ای بین ۳۲۲ تا ۳۲۸ گرم آبلیمو داشته باشد.

ب) شیشه‌ای بیش از ۳۳۵ گرم آبلیمو داشته باشد.

ج) دایرهٔ کنترل کیفیت، میزان آبلیموی ۷۰ شیشه را به صورت تصادفی وزن می‌کند. انتظار می‌رود چند شیشه بیش از ۳۳۵ گرم آبلیمو داشته باشد.

در این مثال $\mu = ۳۳۰$ و $\sigma = ۵$ است؛ بنابراین

الف)
$$P(۳۲۲ \leq X \leq ۳۲۸) = P(\frac{۳۲۲-۳۳۰}{۵} \leq Z \leq \frac{۳۲۸-۳۳۰}{۵})$$

$$= P(-۱/۶ \leq Z \leq -۰/۴)$$

$$= P(Z \leq -۰/۴) - P(Z \leq -۱/۶)$$

$$= ۰/۳۴۴۶ - ۰/۰۵۴۸$$

$$= ۰/۲۸۹۸$$

ب)
$$P(X \geq ۳۳۵) = P(Z \geq \frac{۳۳۵ - ۳۳۰}{۵})$$
$$= P(Z \geq ۱)$$
$$= ۱ - P(Z \leq ۱)$$
$$= ۱ - ۰/۸۴۱۳$$
$$= ۰/۱۵۸۷$$

ج) تعداد شیشه‌هایی که انتظار می‌رود بیش از ۳۳۵ گرم آبلیمو داشته باشند:

$$۰/۱۵۸۷ \times ۷۰ = ۱۱/۱۰۹ \simeq ۱۱$$

مثال ۴ـ ۸ زمان لازم برای تعمیر یک ماشین برچسب‌زنی دارای توزیع نرمال با میانگین ۵۰ دقیقه و انحراف معیار ۲۰ دقیقه است. هزینه هر بار تعمیر ۵ هزار ریال است. اگر تعمیر این ماشین بیش از ۸۵ دقیقه طول بکشد، به علت توقف خط تولید، ضرری معادل ۱۰۰ هزار ریال به بار می‌آید؛ می‌خواهیم محاسبه کنیم امید ریاضی هزینه هر بار خرابی این دستگاه چقدر است.

ابتدا باید محاسبه کنیم با چه احتمالی مدت تعمیر بیش از ۸۵ دقیقه طول می‌کشد؛ یعنی:

$$P(X \geq ۸۵) = P(Z \geq \frac{۸۵ - ۵۰}{۲۰})$$
$$= P(Z \geq ۱/۷۵)$$
$$= ۱ - P(Z \leq ۱/۷۵)$$
$$= ۱ - ۰/۹۵۹۹$$
$$= ۰/۰۴۰۱$$

پس شرکت با احتمال ۰/۰۴۰۱ هم هزینهٔ تعمیر و هم هزینهٔ توقف خط تولید و با احتمال ۰/۹۵۹۹ تنها هزینهٔ تعمیر را متحمل می‌شود؛ بنابراین امید ریاضی هزینه هر تعمیر، $E(C)$، چنین خواهد بود:

$$E(C) = ۰/۹۵۹۹(۵۰۰۰) + ۰/۰۴۰۱(۵۰۰۰ + ۱۰۰,۰۰۰) = ۹۰۱۰$$

تمرین

۱. اگر Z متغیر تصادفی نرمال استاندارد باشد، این احتمالات را با استفاده از جدول پیدا کنید.

الف) $P(Z \leq ۲/۰۶)$

ب) $P(Z > -۰/۲۵)$

ج) $P(-۱/۱۵ < Z \leq ۰/۵۹)$

د) $P(-۱/۹۶ \leq Z \leq ۱/۹۶)$

ه) $P(۲ \leq Z \leq ۴/۲۵)$

۲. فرض کنید $X \sim N(۱۵ , ۳)$ است؛ هریک از این احتمالات را پیدا کنید.

الف) $P(X < ۱۲/۵)$

ب) $P(X \geq ۱۸)$

ج) $P(۱۲/۵ \leq X < ۱۸)$

۳. دستگاه تراشی در هفتهٔ گذشته نوعی پیستون را با میانگین قطر خارجی ۱۱/۲ سانتیمتر و واریانس ۰/۰۰۱۴ با توزیع نرمال، تراش داده است. محاسبه کنید قطر خارجی چند درصد از پیستونها:

الف) کمتر از ۱۱/۲۳۵۲ است.

ب) بیشتر از ۱۱/۲۳۵۲ است.

ج) بین ۱۱/۲۳۵۲ و ۱۱/۲۳۴۹ است.

۴. سن کارگران کارخانه‌ای دارای توزیع نرمال با میانگین ۳۵ سال و انحراف معیار ۱۲ سال است؛ یعنی: $X \sim N(۳۵ , ۱۲)$. اگر خط مشی شرکت بازنشسته کردن تمام افرادی باشد که بیش از ۵۵ سال سن دارند، چند درصد از کارگران بازنشسته می‌شوند؟

۵. مدیریت پرسنلی شرکتی برای تعدادی از متقاضیان استخدام آزمونی برگزار کرده است که حداقل نمرهٔ قبولی در آن ۷۵ است. مشخص شده است که میانگین نمرات متقاضیان ۶۳ با انحراف معیار ۱۵ است. محاسبه کنید چند درصد از متقاضیان پذیرفته می‌شوند.

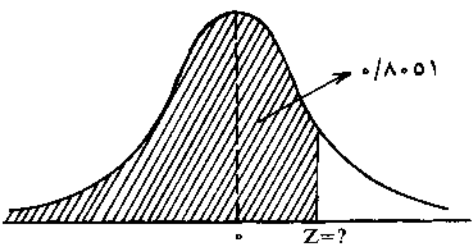
۶. قطر یک بلبرینگ، متغیر تصادفی نرمال با میانگین ۷ سانتیمتر و انحراف معیار ۰/۱ است. استاندارد فنی قطر این کالا عبارت است از $۶/۹۱ \leq X \leq ۷/۰۹$ و تولید یک بلبرینگ استاندارد ۱۲۵۰ ریال سود دارد. اگر قطر بلبرینگ تولیدی کمتر از ۶/۹۱ باشد غیرقابل استفاده است و ۱۰۵۰ ریال زیان دارد و در صورتی که قطر آن بیش از ۷/۰۹ باشد با انجام دادن کار اضافی و صرف هزینه‌ای معادل ۲۰۰ ریال می‌توان آن را به بلبرینگی تبدیل کرد که استاندارد باشد. سود مورد انتظار تولید هر بلبرینگ را حساب کنید.

۳ـ۸ استفادهٔ معکوس از جدول توزیع نرمال استاندارد

در استفادهٔ مستقیم از توزیع نرمال، ابتدا Z را مشخص و سپس احتمال آن را از جدول پیدا می‌کردیم، در استفادهٔ معکوس، مقدار Z برای ما مشخص نیست و تنها احتمال آن مشخص است، احتمال را در جدول پیدا کرده، سپس Z متناظر با آن را مشخص می‌کنیم.

مثال ۵ـ۸ می‌خواهیم مقدار Z را در صورتی که $P(Z \leq z) = ۰/۸۰۵۱$، محاسبه کنیم.

عدد ۰/۸۰۵۱ را در جدول پیدا می‌کنیم و سپس z متناظر با آن را با توجه به سطر و ستون مربوط پیدا می‌کنیم. ۰/۸٦ پس $P(Z \leq ۰/۸٦) = ۰/۸۰۵۱$ خواهد بود.



در روش مستقیم برای حل مسائل، ابتدا متغیر تصادفی X را به Z تبدیل و سپس به جدول مراجعه می‌کردیم ، ولی در روش معکوس، پس از پیدا کردن Z مقدار X را با استفاده از این رابطه پیدا می‌کنیم:

$$Z = \frac{X - \mu}{\sigma} \ , \ X = \mu + \sigma Z$$

مثال ٦ـ۸ فرض کنید $X \sim N(۱۰ , ۲)$ است. می‌خواهیم بدانیم به ازای چه مقداری از X رابطهٔ $P(X \leq x) = ۰/٦۷۰۰$ برقرار است.

در جدول نرمال استاندارد، احتمال ۰/٦۷۰۰ متناظر با $Z = ۰/٤٤$ است؛ یعنی $P(Z \leq ۰/٤٤) = ۰/٦۷۰۰$ است؛ پس

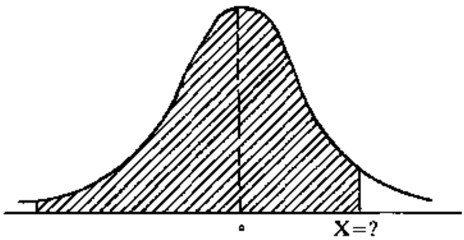
$$X = \mu + \sigma Z = ۱۰ + ۲(۰/٤٤) = ۱۰/۸۸$$

و یا

$$P(X \leq ۱۰/۸۸) = P(Z \leq ۰/٤٤) = ۰/٦۷۰۰$$

مثال ۷ـ۸ میزان مصرف مواد اولیهٔ ماهانه در یک شرکت تولیدی، دارای توزیع نرمال با میانگین ۷۵۲ تن و انحراف معیار ۸٦ است. این شرکت مواد اولیهٔ مورد نیاز خود

را در ابتدای هر ماه تهیه می‌کند. می‌خواهیم حساب کنیم شرکت باید چند تن مواد اولیه برای ماه بعدی تهیه کند تا با ۹۵ درصد اطمینان بداند از نظر مواد اولیه کمبودی نخواهد داشت.



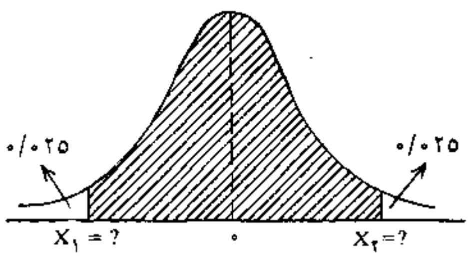
با مراجعه به جدول داریم: $P(Z \leq ۱/۶۴۵) = ۰/۹۵۰۰$

پس

$$X = \mu + \sigma Z = ۷۵۲ + ۸۶(۱/۶۴۵) = ۸۹۳/۴۷$$

یعنی شرکت با خرید ۸۹۳/۴۷ تن مواد اولیه با ۹۵ درصد اطمینان کمبودی نخواهد داشت.

مثال ۸ ـ ۸ دستگاهی طوری تنظیم شده است که به صورت خودکار و به طور متوسط پیچهایی به قطر ۳۸ میلیمتر و با انحراف معیار ۰/۲ تولید می‌کند. می‌خواهیم بدانیم ۹۵ درصد پیچهای تولیدی در چه دامنه‌ای در دو طرف میانگین قرار می‌گیرند.



هدف پیدا کردن مقدار X_1 و X_2 است. جمع مساحت هاشور نخورده در سمت چپ و راست منحنی برابر ۰/۰۵ = ۰/۹۵ − ۱ است. به دلیل متقارن بودن توزیع نرمال حول میانگین، آن دو با هم برابرند و مساحت هر یک ۰/۰۲۵ = ۲ : ۰/۰۵ است؛ پس $P(Z \geq ۱/۹۶) = ۰/۰۲۵$ و $P(Z \leq -۱/۹۶) = ۰/۰۲۵$ است و بنابراین:

$$X_1 = ۳۸ - ۰/۲(۱/۹۶) = ۳۷/۶۰۸$$

$$X_2 = ۳۸ + ۰/۲(۱/۹۶) = ۳۸/۳۹۲$$

به عبارت دیگر ۹۵ درصد پیچهای تولیدی دارای قطری بین ۳۷/۶۰۸ و ۳۸/۳۹۲ میلیمتر هستند.

تمرین

۱. در هریک از این موارد مقدار z را پیدا کنید.

الف) $P(Z \geq z) = ۰/۰۴۸۵$

ب) $P(Z \geq z) = ۰/۸۵۰۸$

ج) $P(|Z| \leq z) = ۰/۸۷۴۰$

۲. فرض کنید $X \sim N(۱/۱ , ۰/۱۵)$ باشد. به ازای چه مقدار X، روابط ذیل برقرار است؟

الف) $P(X \leq x) = ۰/۲۵۱۴$

ب) $P(X \geq x) = ۰/۵۴۵۰$

۳. زمان لازم برای تعمیر سیستم نیروی محرکهٔ یک نوع ماشین دارای توزیع نرمال با میانگین ۴۸ دقیقه و واریانس ۲۵ است. تعمیرکار به رانندهٔ آن می‌گوید که زمان تعمیر حداکثر یک ساعت طول می‌کشد، محاسبه کنید:

الف) احتمال غلط بودن این ادعا چقدر است؟

ب) اگر این تعمیرکار با ۹۸ درصد اطمینان بخواهد مدت تعمیری را مشخص کند، آن مدت چقدر خواهد بود؟

۴. زمان لازم برای انجام دادن کارهای بانکی یک مشتری به طور متوسط ۱۱۰ ثانیه با انحراف معیار ۳۵ ثانیه است که به صورت نرمال توزیع شده است، محاسبه کنید:

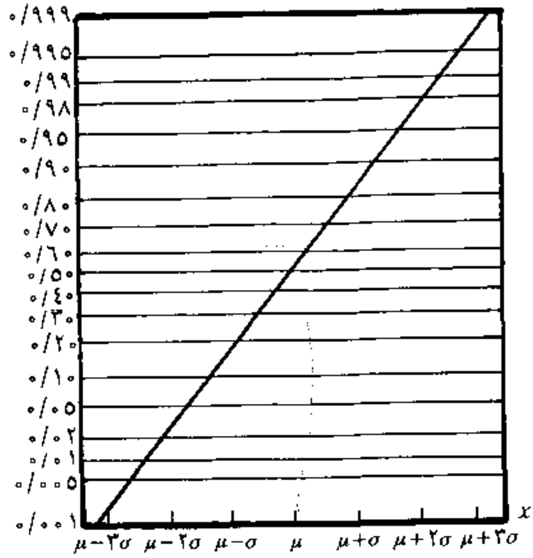
الف) کار ۸۵ درصد از مراجعه کنندگان به این بانک در چه دامنه‌ای در دو طرف میانگین زمان مراجعه قرار می‌گیرد؟

ب) کار ۵٪ از افرادی که بیشترین زمان را به خود اختصاص می‌دهد، حداقل چقدر طول می‌کشد؟

۸ـ۴ روشی برای تشخیص نرمال بودن توزیع

روشهای مختلفی برای تشخیص و آزمون نرمال بودن مشاهدات وجود دارد. در اینجا از روشی نسبتاً ساده و تا حدی ذهنی، برای تشخیص نرمال بودنِ توزیع

مشاهدات استفاده می‌کنیم. صفحهٔ مخصوصی به نام «صفحهٔ احتمال نرمال» وجود دارد که نمونه آن در شکل ۷ـ ۸ آورده شده است. احتمالهای تجمعی طوری روی محور عمودی آن درج شده که نمایش هندسی $P(X \leq x)$ به خط مستقیمی تبدیل شود و در محور افقی نیز باید مقیاس طوری انتخاب شود که تمام مشاهدات ما را در برگیرد.



شکل ۷ـ ۸ صفحهٔ احتمال نرمال

بنا به تجربه مشخص شده‌است که دست کم ۱۵ تا ۲۰ مشاهده برای قضاوت در مورد توزیع لازم است.

برای رسم نمودار احتمال نرمال داده‌های طبقه‌بندی نشده در صفحهٔ احتمال نرمال، باید این مراحل را انجام دهید:

۱. n مشاهده را از کوچک به بزرگ مرتب کنید.

۲. روی محور افقی برای نمایش همهٔ داده‌ها مقیاسی برگزینید.

۳. براساس مشاهدهٔ مرتب‌شدهٔ iام در محور افقی $(i = 1, 2, \ldots, n)$ و فراوانی نسبی اصلاح شدهٔ $\frac{i - 0/5}{n}$ در محور عمودی نقاطی را روی کاغذ پیدا کنید (نقاطی با مختصات $\frac{i - 0/5}{n}$ و مشاهدهٔ مرتب شده iام).

۴. نمودار حاصل را با یک خط مستقیم (خط نرمال) مقایسه کنید. وجود

انحرافهای منظم نشان‌دهندهٔ نرمال نبودن توزیع مشاهدات است.

البته می‌توان به جای فراوانی نسبی اصلاح‌شدهٔ $\frac{i - ۰/۵}{n}$، از فراوانی نسبی تجمعی معمولی $\frac{i}{n}$ استفاده کرد؛ در این صورت بزرگترین مشاهده که فراوانی نسبی تجمعی آن یک است در نمودار قرار نمی‌گیرد.

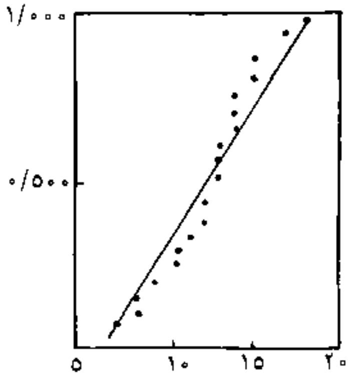
اگر داده‌ها طبقه‌بندی شده باشند، می‌توان از حدود بالای طبقات و فراوانی نسبی تجمعی آنها استفاده کرد؛ به عبارت دیگر می‌توان از نقاطی با مختصات فراوانی نسبی تجمعی طبقهٔ iام و حد بالای طبقهٔ iام استفاده کرد.

مثال ۹ـ۸ برای مشخص شدن توزیع تعداد ضایعات روزانهٔ یک دستگاه تولیدی، ۲۰ روز را به صورت تصادفی انتخاب و ضایعات آن را یادداشت کرده‌ایم که بدین قرار است: ۱۰، ۱۴، ۱۳، ۱۸، ۱۲، ۱۵، ۱۱، ۱۷، ۹، ۱۳، ۱۴، ۸، ۱۲، ۵، ۷، ۱۴، ۱۳، ۱۰، ۱۵، ۸. می‌خواهیم بدانیم آیا توزیع تعداد ضایعات نرمال است یا نه.

ابتدا داده‌ها را مرتب می‌کنیم.

۵, ۷, ۸, ۸, ۹, ۱۰, ۱۰, ۱۱, ۱۲, ۱۲, ۱۳, ۱۳, ۱۳, ۱۴, ۱۴, ۱۴, ۱۵, ۱۵, ۱۷, ۱۸

سپس نقاط $(۵, ۰/۰۲۵) = (۵, \frac{۱ - ۰/۵}{۲۰})$، $(۷, ۰/۰۷۵) = (۷, \frac{۲ - ۰/۵}{۲۰})$، ... و $(۱۸, ۰/۹۷۵) = (۱۸, \frac{۲۰ - ۰/۵}{۲۰})$ را در صفحهٔ احتمال نرمال پیدا می‌کنیم.



شکل ۸ـ۸ رسم نمودار احتمال برای مثال ۹ـ۸

چون نمودار احتمال در این شکل به خط نرمال نزدیک است، می‌توان گفت که

توزیع مشاهدات ممکن است نرمال باشد.

مثال ۱۰ـ۸ این جدول، توزیع زمان لازم برای پرکردن فرمهای درخواست کار به وسیلهٔ ۲۰۰ متقاضی را نشان می‌دهد.

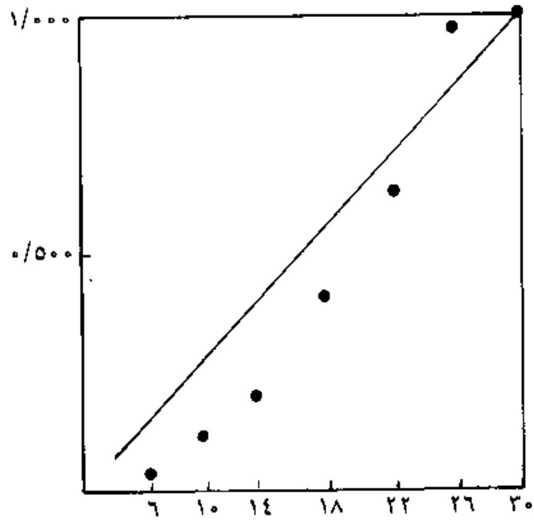
فراوانی	زمان (به دقیقه)
۵	۳ـ۶
۱۷	۷ـ۱۰
۲۰	۱۱ـ۱۴
۴۱	۱۵ـ۱۸
۵۰	۱۹ـ۲۲
۶۰	۲۳ـ۲۶
۷	۲۷ـ۳۰

می‌خواهیم با استفاده از صفحهٔ احتمال نرمال مشخص کنیم آیا توزیع نرمال است یا نه.

برای رسم نمودار احتمال از حدود بالای طبقات و فراوانی نسبی تجمعی هر طبقه استفاده می‌کنیم.

فراوانی نسبی تجمعی	فراوانی تجمعی	حدود بالای طبقات
۰/۰۲۵	۵	۶
۰/۱۱۰	۲۲	۱۰
۰/۲۱۰	۴۲	۱۴
۰/۴۱۵	۸۳	۱۸
۰/۶۶۵	۱۳۳	۲۲
۰/۹۶۵	۱۹۳	۲۶
۱/۰۰۰	۲۰۰	۳۰

نقاطی با مختصات فراوانی نسبی تجمعی طبقهٔ iام و حد بالای طبقهٔ iام را روی صفحهٔ احتمال نرمال نشان می‌دهیم:



شکل ۹ـ۸ رسم نمودار احتمال برای مثال ۱۰ـ۸

همان طور که در نمودار احتمال شکل ۹ـ۸ ملاحظه می‌شود اکثر نقاط در زیر خط واقع شده‌اند؛ بنابراین نمی‌توان گفت که توزیع داده‌ها نرمال است.

تمرین

۱. اطلاعات مربوط به توزیع تعداد فروش شرکتی در ۵۰۰ روز جمع‌آوری و در این جدول خلاصه شده است:

تعداد فروش	روز
≤۲٤	۳۸
۲۵ـ۲۹	۵۲
۳۰ـ۳٤	۱٦۵
۳۵ـ۳۹	۱۵۰
٤۰ـ٤٤	٦۰
≥٤۵	۳۵

با استفاده از صفحهٔ احتمال نرمال، مشخص کنید آیا این توزیع می‌تواند نرمال باشد.

۲. مقادیر واقعی وزن ۳۰ کرهٔ ۱۰۰ گرمی که برحسب گرم گرد شده‌اند و به طور تصادفی از میان

تولیدات یک هفتهٔ یک شرکت انتخاب شده‌اند، بدین قرار است: ۹۸، ۹۶، ۱۰۴، ۱۰۶، ۹۷، ۹۸، ۱۰۷، ۱۰۲، ۱۰۶، ۱۰۰، ۱۰۱، ۹۷، ۹۸، ۱۰۶، ۱۰۴، ۱۰۰، ۱۰۱، ۹۵، ۹۷، ۱۰۵، ۱۰۷، ۹۹، ۹۹، ۱۰۰، ۱۰۲، ۱۰۵، ۱۰۳، ۱۰۶، ۹۴، ۱۰۷. آیا توزیع وزن کره‌ها نرمال است؟

۸ـ۵ تقریب توزیع دو جمله‌ای به وسیلهٔ توزیع نرمال

در فصل ششم درباره توزیع گسستهٔ دو جمله‌ای ـ تعداد موفقیتها در n آزمایش مستقل که احتمال موفقیت (p) در هر آزمایش ثابت است ـ توضیح دادیم. جداول مربوط به توزیع دو جمله‌ای، احتمالات مربوط به n های کوچک را نشان می‌دهد. اگر n بزرگ شود، محاسبهٔ احتمال کار دشواری خواهد بود. در تقریب دو جمله‌ای به وسیله پواسون، قسمت ۱۰ـ۶، گفتیم اگر n بزرگ و p نزدیک به صفر یا یک باشد، توزیع پواسون با $\lambda = np$ تقریب خوبی برای دو جمله‌ای خواهد بود.

در اینجا قضیهٔ دیگری را مطرح می‌کنیم و آن اینکه اگر متغیر تصادفی X دارای توزیع دو جمله‌ای با پارامترهای n و p باشد، در صورتی که n بزرگ باشد و p به صفر یا یک زیاد نزدیک نباشد، تقریب نرمال با پارامترهای $\mu = np$ و $\sigma = \sqrt{npq}$ تقریب خوبی‌برای توزیع دو جمله‌ای خواهد بود؛ به عبارت ریاضی: $X \sim N(np, \sqrt{npq})$. بنا به تجربه، اگر np و nq هر دو بزرگتر از ۵ باشند، تقریب نرمال تقریب خوبی خواهد بود. در مواردی که p به ۰/۵ نزدیک باشد، تقریب نرمال برای n های کوچک نیز خوب است.

تقریب نرمال به ما این امکان را می‌دهد که احتمالهای مربوط به متغیر تصادفی X با توزیع دوجمله‌ای را چنان محاسبه کنیم که گویی X متغیر تصادفی با توزیع نرمال است.

چنانچه پیش از این نیز گفتیم، توزیع دو جمله‌ای توزیعی گسسته و توزیع نرمال توزیعی پیوسته است؛ مثلاً در توزیع دو جمله‌ای با ۱۰ = n و ۰/۴ = p، $P(X = ۵)$ مقداری مثبت است، ولی در توزیع نرمال مقدار $P(X = ۵)$ برابر صفر است؛ بنابراین وقتی تقریب نرمال را برای دو جمله‌ای به کار می‌بریم، باید از تصحیح پیوستگی استفاده کنیم؛ یعنی به جای محاسبهٔ $P(X = ۵)$ باید $P(۴/۵ \le X \le ۵/۵)$ محاسبه شود. تصحیحهای پیوستگی مختلف در جدول ۱ـ۸ آورده شده‌است.

جدول ۸.۱ تصحیح‌های پیوستگی توزیع دو جمله‌ای به منظور تقریب نرمال

احتمال مورد نظر از توزیع دو جمله‌ای	احتمال مورد نظر از توزیع نرمال
$P(X = x)$	$P(x - ۰/۵ \le X \le x + ۰/۵)$
$P(X \le x)$	$P(X \le x + ۰/۵)$
$P(X < x)$	$P(X \le x - ۱ + ۰/۵) = P(X \le x - ۰/۵)$
$P(X \ge x)$	$P(X \ge x - ۰/۵)$
$P(X > x) = P(X \ge x + ۱)$	$P(X \ge x + ۱ - ۰/۵) = P(X \ge x + ۰/۵)$
$P(x_۱ \le X \le x_۲)$	$P(x_۱ - ۰/۵ \le X \le x_۲ + ۰/۵)$

مثال ۸-۱۱ متغیر تصادفی X دارای توزیع دو جمله‌ای با n = ۱۵ و P = ۰/۴ است.

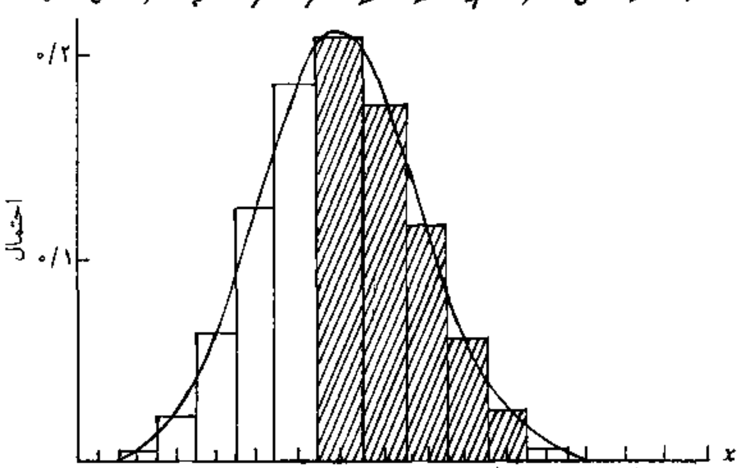
الف) می‌خواهیم احتمالات $P(X = x)$ را به ازای $x = ۰, ۱, ..., ۱۵$ از جدول پیدا کرده، نمودار مستطیلی احتمال را رسم کنیم.

ب) می‌خواهیم تعیین کنیم آیا تقریب نرمال می‌تواند تقریب خوبی برای این توزیع باشد.

ج) می‌خواهیم مقدار $P(۷ \le X \le ۱۰)$ را با توزیع دو جمله‌ای و تقریب نرمال (با تصحیح پیوستگی) محاسبه و آن دو را با یکدیگر مقایسه کنیم.

الف) با مراجعه به جدول ۱ پیوست، توزیع احتمال را به ازای مقادیر مختلف به دست می‌آوریم:

x	۰	۱	۲	۳	٤	۵	٦	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱٤	۱۵
$P(X=x)$	۰/۰۰۰	۰/۰۰۵	۰/۰۲۲	۰/۰٦۳	۰/۱۲۷	۰/۱۸٦	۰/۲۰۷	۰/۱۷۷	۰/۱۱۸	۰/۰٦۱	۰/۰۲۵	۰/۰۰۷	۰/۰۰۲	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰



شکل ۸.۱۰ نمودار مستطیلی احتمال توزیع دوجمله‌ای با n=۱۵ و p=۰/٤ و منحنی نرمال تقریبی آن

ب) تقریب نرمال تقریب خوبی برای این توزیع است زیرا $np = ۱۵ \times ۰/۴ = ۶ > ۵$ و $nq = ۱۵ \times ۰/۶ = ۹ > ۵$ است.

ج) با استفاده از توزیعهای دو جمله‌ای خواهیم داشت:

$$P(۷ \leq X \leq ۱۰) = ۰/۳۸۱$$

ابتدا با استفاده از توزیع نرمال پارامترهای آن را پیدا می‌کنیم و سپس تصحیح پیوستگی را انجام می‌دهیم.

$$\mu = np = ۶ \qquad \sigma = \sqrt{npq} = ۱/۸۹۷$$

$$P(۶/۵ \leq X \leq ۱۰/۵) = P(۰/۲۶ \leq Z \leq ۲/۳۷)$$
$$= P(Z \leq ۲/۳۷) - P(Z \leq ۰/۲۶)$$
$$= ۰/۳۸۸$$

تفاوت احتمالات به دست آمده از دو روش ناچیز است. ملاحظه می‌شود که تقریب نرمال حتی برای n هایی به کوچکی ۱۵ نیز قابل استفاده است. هرچقدر n افزایش یابد، تقریب دقیقتر می‌شود.

مثال ۱۲ـ۸ شصت و پنج درصد از کل افرادی که به فروشگاهی مراجعه می‌کنند خرید می‌کنند. اگر در یک روز ۳۰ نفر به این فروشگاه مراجعه کنند، می‌خواهیم این احتمالات را پیدا کنیم:

الف) حداقل ۲۲ نفر خرید کنند.

ب) کمتر از ۱۵ نفر خرید کنند.

ج) دقیقاً ۲۰ نفر خرید کنند.

در این مثال $np = ۳۰ \times ۰/۶۵ = ۱۹/۵ > ۵$ و $nq = ۳۰ \times ۰/۳۵ = ۱۰/۵ > ۵$ است؛ بنابراین می‌توان از تقریب نرمال استفاده کرد.

$$\mu = np = ۱۹/۵ \qquad \sigma = \sqrt{npq} = ۲/۶۱$$

بنابراین:

الف)
$$P_{\text{دو جمله‌ای}}(X \geq ۲۲) = P_{\text{نرمال}}(X \geq ۲۱/۵)$$
$$= P(Z \geq ۰/۷۷)$$
$$= ۱ - P(Z \leq ۰/۷۷)$$
$$= ۰/۲۲۰۶$$

ب) $P_{\text{دو جمله‌ای}}(X < ۱۵) = P_{\text{نرمال}}(X \leq ۱۴/۵)$

$= P(Z \leq -۱/۹۲)$

$= ۰/۰۲۷۴$

ج) $P_{\text{دو جمله‌ای}}(X = ۲۰) = P_{\text{نرمال}}(۱۹/۵ \leq X \leq ۲۰/۵)$

$= P(۰ \leq Z \leq ۰/۳۸)$

$= P(Z \leq ۰/۳۸) - P(Z \leq ۰)$

$= ۰/۱۴۸۰$

تمرین

۱. فرض کنید حسابهای ۳۰ درصد از شرکتهایی که یک سازمان آنها را حسابرسی می‌کند قبول می‌شوند. اگر از شرکتهایی که این سازمان حسابرسی کرده‌است، ۵۰ شرکت را به صورت تصادفی برگزینیم، احتمال قبول شدن حسابها را در این موارد محاسبه کنید:
الف) بین ۲۵ تا ۴۵ شرکت
ب) حداکثر ۴۰ شرکت
ج) دقیقاً ۳۵ شرکت

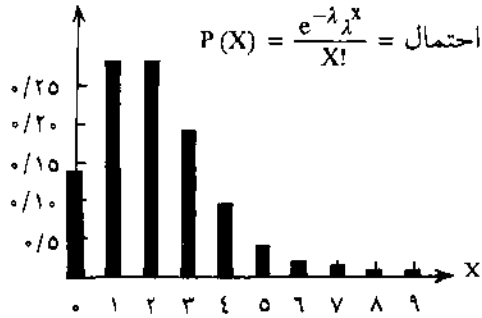
۲. استادی به تجربه‌دریافته است که ۴۵ درصد از دانشجویانش در درس خاصی نمرهٔ «ب» می‌گیرند. اگر در ترم جاری ۲۵ نفر این درس را با او گرفته باشند، احتمال گرفتن نمره «ب» را برای این تعداد محاسبه کنید: الف) حداقل ۱۰ نفر، ب) بین ۸ تا ۲۰ نفر، ج) دقیقاً ۱۰ نفر

۳. یک شبکهٔ تلویزیونی ادعا می‌کند که ۷۰ درصد از کل بینندگان تلویزیون وقت خود را به دیدن برنامهٔ خاصی در شبهای سه شنبه اختصاص می‌دهند. به فرض درست بودن این ادعا، احتمال آنکه از بین ۴۰۰ بیننده، حداقل ۱۵۰ نفر از آنان برنامه را تماشا کرده باشند، چقدر است؟

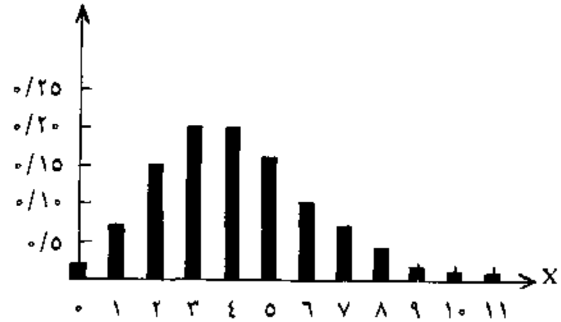
۸ـ۶ تقریب توزیع پواسون به وسیلهٔ توزیع نرمال

وقتی میانگین توزیع پواسون (λ) نسبتاً بزرگ شود می‌توان تقریب نرمال را برای توزیع پواسون به کار برد. از اثبات ریاضی این موضوع خودداری می‌کنیم. همان‌طور که در شکل ۱۱ـ ۸ نشان داده شده است هر چقدر λ افزایش یابد توزیع به نرمال نزدیکتر می‌شود. به طور کلی اگر $\lambda \geq ۱۰$ باشد، تقریب نرمال تقریب خوبی برای پواسون

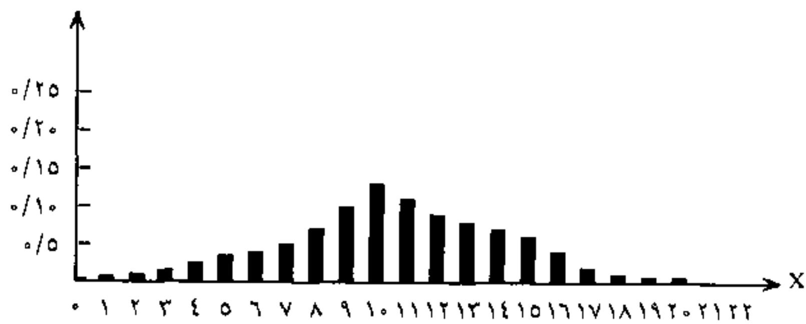
است. در این صورت میانگین و انحراف معیار توزیع نرمال برابر $\lambda = \mu$ و $\sigma = \sqrt{\lambda}$ خواهد بود.



$\lambda = ۲$



$\lambda = ٤$



$\lambda = ۱۰$

شکل ۸.۱۱ با بزرگ شدن λ توزیع پواسون به نرمال نزدیک می‌شود

در هنگام حل مسائل، در استفاده از توزیع نرمال به جای پواسون نیز، به دلیل گسسته بودن توزیع پواسون و پیوسته‌بودن توزیع نرمال، باید از تصحیح پیوستگی -

همانند تقریب دو جمله‌ای به وسیلهٔ نرمال ـ استفاده کرد.

مثال ۱۳ـ ۸ تعداد تلفنهایی که به روابط عمومی صدا و سیما زده می‌شود دارای توزیع پواسون با میانگین ۱۰ تلفن در هر دقیقه است. می‌خواهیم بدانیم احتمال اینکه کمتر از ۸ تلفن در یک دقیقه خاص زده شود، چقدر است.

در این مثال چون $\lambda = ۱۰$ است، می‌توان از تقریب نرمال با پارامترهای $\mu = ۱۰$ و $\sigma = \sqrt{۱۰} = ۳/۱۶$ استفاده کرد.

$$\begin{aligned} P_{\text{پواسون}}(X < ۸) &= P_{\text{نرمال}}(X \leq ۷/۵) \\ &= P(Z \leq -۰/۷۹) \\ &= ۰/۲۱۴۸ \end{aligned}$$

مثال ١٤ـ ۸ مثال ۱۳ـ۸ را در نظر بگیرید. می‌خواهیم بدانیم احتمال اینکه در ۲/۵ دقیقه بین ۲۰ تا ۳۰ تلفن زده شود، چقدر است.

در اینجا $\lambda = ۲/۵ \times ۱۰ \geq ۱۰$ است، پس می‌توان از تقریب نرمال با $\mu = ۲۵$ و $\sigma = \sqrt{۲۵} = ۵$ استفاده کرد.

$$\begin{aligned} P_{\text{پواسون}}(۲۰ \leq X \leq ۳۰) &= P_{\text{نرمال}}(۱۹/۵ \leq X \leq ۳۰/۵) \\ &= P(-۱/۱ \leq Z \leq ۱/۱) \\ &= P(Z \leq ۱/۱) - P(Z \leq -۱/۱) \\ &= ۰/۷۲۸۶ \end{aligned}$$

تمرین

۱. به طور متوسط در هر دقیقه ۰/۵ مشتری با توزیع پواسون به قسمت پرداخت فروشگاهی مراجعه می‌کند. احتمال اینکه بیش از ۲۰ مشتری در طی نیم ساعت مراجعه کنند، چقدر است؟

۲. تعداد کشتیهایی که به اسکله‌ای جهت تخلیهٔ بار می‌رسند، دارای توزیع پواسون با میانگین ۱/۷۵ کشتی در هر شبانه‌روز است. احتمالات زیر را برای رسیدن کشتی در یک هفته محاسبه کنید.

الف) بین ٤ تا ۱۱ کشتی برسد.

ب) حداقل ۸ کشتی برسد.

۳. در خلال ساعت ۴ تا ۷ بعد از ظهر، دورهٔ اوج کاری، به طور متوسط در هر ۵ دقیقه یک ماشین جهت تعمیر به تعمیرگاهی مراجعه می‌کند. احتمال اینکه بین ساعت ۵ تا $\frac{1}{4}$۶ بیش از ۱۲ ماشین به این تعمیرگاه مراجعه کند، چقدر است؟

۸.۷ قضیهٔ حدّ مرکزی

دو موضوع در قضیهٔ حدّ مرکزی مطرح می‌شود؛ موضوع اول جمع n متغیر تصادفی مستقل و موضوع دوم توزیعهای نمونه گیری است. در کتاب حاضر فقط درباره موضوع اول توضیح خواهیم داد و موضوع دوم در جلد دوم کتاب مطرح خواهد شد.

توزیع نرمال تقریب خوبی برای توزیع مجموع متغیرهای تصادفی است. اگر X_1، X_2، ...، X_n، n متغیر تصادفی مستقل باشند، مشروط بر آنکه n به اندازهٔ کافی بزرگ باشد، آنگاه متغیر تصادفی $Y = X_1+X_2+...+X_n$ دارای توزیع نرمال با $\mu_Y=\sum_{i=1}^{n}\mu_{x_i}$ و $\sigma_Y^2=\sum_{i=1}^{n}\sigma_{x_i}^2$ خواهد بود که μ_{x_i} و $\sigma_{x_i}^2$ به ترتیب میانگین و واریانس متغیر تصادفی X_i هستند. ممکن است این سؤال مطرح شود که اندازهٔ n چقدر باید باشد تا بتوان از توزیع نرمال برای جمع n متغیر تصادفی استفاده کرد. اگر توزیع هر جملهٔ X_i چولگی زیادی نداشته باشد و اگر محتمل نباشد که یک جمله سهم عمده و تعیین کننده‌ای از مجموع را داشته باشد، توزیع مجموع حداقل ۱۰ متغیر تصادفی $(n \geq 10)$ تقریباً توزیع نرمال خواهد بود.

این نکته که X_i ها می‌توانند هر توزیعی داشته باشند، اهمیت توزیع نرمال را نشان می‌دهد. مثالهای عملی این قضیه ممکن است زمانبندی پروژه‌ها که از فعالیتهای مختلفی تشکیل شده‌اند و یا وزن محموله‌هایی که وزن هر جزء از آنها دارای توزیع خاصی است، باشد.

مثال ۸ـ۱۵ برای تکمیل پروژه‌ای لازم است فعالیتهای مختلفی صورت گیرد. فعالیتهای مسیر بحرانی آن ـ مسیر بحرانی مسیری است که هرگونه تأخیری در انجام فعالیتهای آن باعث تأخیر در زمان اتمام کل پروژه می‌شود ـ همراه با میانگین و واریانس زمان انجام هر فعالیت در جدول ۲ـ۸ آورده شده است.

جدول ۲ـ۸ میانگین و واریانس زمان انجام فعالیتهای مسیر بحرانی (به هفته)

فعالیت	میانگین	واریانس	فعالیت	میانگین	واریانس
۱	۳	۱/۲	۸	٤	۱/۸
۲	۲/٥	۱/۱	۹	۲/٥	۱/۲
۳	٤	۱/٥	۱۰	۳/۱	۲
٤	۲/۷	۲	۱۱	۲/۲	۱/٦
٥	۱/۲	۰/٥	۱۲	٦	۲/۷
٦	۳	۱/۲	۱۳	۳/۲	۱/۷
۷	٥	۲/۲	۱٤	۱	۰/۲

الف) می‌خواهیم میانگین و واریانس زمان کل پروژه را به دست آوریم.
ب) می‌خواهیم بدانیم احتمال اینکه پروژه حداکثر ٤٥ هفته طول بکشد، چقدر است.

الف) با پارامترهای ذیل صرف‌نظر از توزیع هریک از فعالیتها، توزیع زمان مجموع فعالیتها تقریباً نرمال است.

$$\mu_Y = \mu_{x_1} + \mu_{x_2} + \dots + \mu_{x_{14}}$$
$$= ۳ + ۲/٥ + \dots + ۱$$
$$= ٤۳/٤$$

$$\sigma_Y^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2 + \dots + \sigma_{x_{14}}^2$$
$$= ۱/۲ + ۱/۱ + \dots + ۰/۲$$
$$= ۲۰/۹$$

ب) زمان اتمام کل پروژه (Y) متغیری تصادفی با $Y \sim N(٤۳/٤\ ,\ ٤/٥۷)$ است؛ بنابراین:

$$P(Y \le ٤٥) = P(Z \le ۰/۳٥)$$
$$= ۰/٦۳٦۸$$

مثال ۱٦ـ۸ با توجه به اطلاعات مثال ۱٥ـ ۸ می‌خواهیم مدتی را برای تکمیل پروژه در نظر بگیریم که با ۹٥ درصد اطمینان بدانیم پروژه طی آن مدت انجام می‌شود.

با مراجعه به جدول نرمال استاندارد داریم:

$$P(Z \leq ۱/۶۴۵) = ۰/۹۵$$

$$Y = \mu_Y + \sigma_Y Z$$

$$= ۴۳/۴ + ۴/۵۷(۱/۶۴۵)$$

$$= ۵۰/۹۲$$

حاصل عبارات نشان می‌دهد که برای تکمیل پروژه با ۹۵ درصد اطمینان به زمانی معادل ۵۰/۹۲ هفته نیازمندیم.

تمرین

۱. می‌خواهیم ۶۵ گونی سیب‌زمینی را که وزن هرگونی، متغیری تصادفی با میانگین ۷۵ کیلوگرم و واریانس ۱۵ است با کامیون از شهر «الف» به شهر «ب» حمل کنیم. ظرفیت این کامیون ۵ تن است. اگر بار کامیون بیش از ظرفیت باشد، راننده جریمه می‌شود. احتمال جریمه‌شدن راننده این کامیون چقدر است؟

۲. اتوبوسی بین شهر اول و یازدهم تردد می‌کند. بین این دو شهر ۹ شهر دیگر وجود دارد که میانگین زمان حرکت بین آنها (به ساعت) همراه با انحراف معیار آن در این جدول آمده است.

محدودهٔ سفر	میانگین	انحراف معیار	محدودهٔ سفر	میانگین	انحراف معیار
۱_۲	۱	۰/۲۰	۶_۷	۲	۰/۱۵
۲_۳	۱/۵	۰/۳۰	۷_۸	۳	۰/۴۰
۳_۴	۲	۰/۲۵	۸_۹	۱/۲	۰/۱۵
۴_۵	۰/۷۵	۰/۱	۹_۱۰	۰/۸	۰/۱۰
۵_۶	۱/۵	۰/۲۰	۱۰_۱۱	۱/۶	۰/۳۰

الف) میانگین و واریانس کل سفر را حساب کنید.

ب) اگر شرکت مسافربری بخواهد با ۹۰ درصد احتمال مدت زمانی را اعلام کند که اتوبوس به مقصد می‌رسد، آن زمان چند ساعت خواهد بود؟

۸ـ۸ خلاصه

توزیع نرمال؛ مهمترین توزیع پیوسته است. در این فصل ضمن معرفی این توزیع، نحوهٔ تأثیر میانگین و انحراف معیار هر توزیع بر منحنی نرمال آن توزیع بیان شد و ویژگیهای توزیع نرمال مطرح شد.

همچنین نحوهٔ تبدیل متغیر تصادفی نرمال به متغیر نرمال استاندارد و نحوهٔ استفاده از جدول آماری برای محاسبه مقادیر احتمالی ذکر شد و توضیح داده شد که اگر «مقدار احتمال» مشخص باشد، می‌توان با استفاده معکوس از جدول مذکور، مقدار متغیر تصادفی را پیدا کرد. همچنین با تقریب زدن توزیعهای دو جمله‌ای و پواسون به وسیلهٔ توزیع نرمال، می‌توان از محاسبات بیش از حد، اجتناب کرد. در انتهای فصل نیز یکی از ابعاد حایز اهمیت «قضیه حد مرکزی»، در مورد جمع n متغیر تصادفی مستقل مطرح شد.

۹ـ۸ سؤالات و مسائل

سؤالات دو گزینه‌ای

۱. مقدار z برای نقطه‌ای مانند x در توزیع نرمال عبارت است از سطح زیر منحنی بین x و میانگین توزیع. ص ☐ غ ☐

۲. در توزیع نرمال، میانگین همواره بین میانه و مد قرار می‌گیرد. ص ☐ غ ☐

۳. دنباله‌های راست و چپ توزیع نرمال تا بی‌نهایت امتداد داشته، هرگز محورافقی را قطع نمی‌کند. ص ☐ غ ☐

٤. سطح محصور بین $\mu + \sigma$ و $\mu - \sigma$ برابر با ۰/۹۵٤ است. ص ☐ غ ☐

۵. هرچقدر واریانس افزایش یابد، منحنی نرمال کشیده‌تر می‌شود. ص ☐ غ ☐

۶. اگر نمودار پس از رسم در صفحه احتمال نرمال، انحرافهای منظمی از خط نرمال داشت، توزیع داده‌ها نرمال است. ص ☐ غ ☐

۷. توزیع دو جمله‌ای را در صورتی که $np > 5$ و $nq > 5$ باشد، می‌توان با توزیع نرمال تقریب زد. ص ☐ غ ☐

۸. توزیع پواسون را درصورتی که $\lambda > 5$ باشد، می‌توان با توزیع نرمال تقریب زد. ص ☐ غ ☐

۹. سطح زیر منحنی نرمال بین میانگین و میانگین بعلاوه یک انحراف معیار برای توزیعی‌که میانگین آن ۱۰۰ است بیشتراز توزیعی است که میانگین آن ۱۰ است. ص ☐ غ ☐

سؤالات چهارگزینه‌ای

۱۰. میانگین و واریانس توزیع نرمال استاندارد به ترتیب (از چپ به راست) کدامند؟

الف) (۰ , ۱) ج) (۱ , ۱)

ب) (۱ , ۰) د) (۰ , ۰)

۱۱. فرض کنید Z متغیر استاندارد و $\int_0^2 \varphi(z)\,dz = ۰/۴۷۷۲$ است. $P(Z \geq ۲)$ کدام است؟

الف) ۰/۰۲۲۸ ج) ۰/۹۷۷۲

ب) ۰/۴۷۷۲ د) ۰/۵۲۲۸

۱۲. اگر متغیر تصادفی X دارای توزیع نرمال با میانگین ۵۰ باشد، $P(X \leq ۵۰)$ کدام است؟

الف) صفر ج) $\frac{۳}{۴}$

ب) $\frac{۱}{۲}$ د) ۱

۱۳. اگر اندازهٔ دو نفر از جامعهٔ نرمالی، ۱۳ و ۱۹ و اندازهٔ این دو برحسب متغیر استاندارد Z، صفر و ۳ باشد، میانگین و انحراف معیار (به ترتیب از چپ به راست) کدامند؟

الف) (۶ , ۳) ج) (۱۳ , ۲)

ب) (۱۹ , ۳) د) (۲ , ۱۹)

۱۴. اگر $P(۰ \leq Z \leq ۰/۵) = ۰/۱۹۱۵$ و $P(۰ \leq Z \leq ۱) = ۰/۳۴۱۳$ باشد، $P(-۱ \leq Z \leq ۰/۵)$ کدام است؟

الف) ۰/۱۴۹۸ ج) ۰/۵۳۲۸

ب) ۰/۴۶۷۲ د) ۰/۸۵۰۲

۱۵. اگر $P(Z \leq -۲) = ۰/۰۲۲۸$ و X نیز دارای توزیع نرمال با میانگین ۲۵ و $P(X \geq ۵) = ۰/۹۷۷۲$ باشد، انحراف معیار X کدام است؟

الف) صفر ج) ۱۰

ب) ۵ د) ۱۵

۱۶. اگر n و p دو پارامتر توزیع دو جمله‌ای باشند، کدام یک از این موارد را می‌توان با توزیع نرمال تقریب زد؟

الف) $n = ۵ , P = ۰/۳$ ج) $n = ۱۵ , P = ۰/۴۵$

ب) $n = ۱۰ , P = ۰/۴$ د) الف و ج

۱۷. اگر بخواهیم $P(X > ۹)$ را در توزیعی دو جمله‌ای با توزیع نرمال تقریب بزنیم ـ مشروط بر آنکه

تقریب مجاز باشد ـ کدام یک از این موارد را، با تصحیح پیوستگی، باید حساب کنیم؟

الف) $P(X \geq ۸/۵)$ ج) $P(X \geq ۹/۵)$

ب) $P(X \leq ۹/۵)$ د) $P(X \leq ۸/۵)$

۱۸. اگر محموله‌ای متشکل از ۱۰ قلم کالا باشد که وزن هر قلم کالا دارای توزیع یکنواخت با پارامترهای ۱۳ = α و ۲۵ = β تن باشد، کدام یک از عبارتهای ذیل در مورد محموله صحیح است؟

الف) ۱۹۰ = $\sigma^۲$، ۱۲۰ = μ، و توزیع یکنواخت

ب) ۱۹۰ = $\sigma^۲$، ۱۲۰ = μ، و توزیع نرمال

ج) ۱۲۰ = $\sigma^۲$، ۱۹۰ = μ، و توزیع نرمال

د) ۱۲۰ = $\sigma^۲$، ۱۹۰ = μ، و توزیع نامشخص

۱۹. توزیع پواسونی با ۳۶ = λ را در نظر بگیرید؛ پارامترهای تقریب نرمال آن کدام است؟

الف) $\sigma = ۳۶ , \mu = ۳۶$ ج) $\sigma = ۶ , \mu = ۳۶$

ب) $\sigma = ۳۶ , \mu = ۶$ د) $\sigma = ۶ , \mu = ۶$

۲۰. برای رسم نمودار احتمال در صفحهٔ احتمال نرمال برای داده‌های طبقه‌بندی شده، از کدام یک از موارد ذیل استفاده می‌شود؟

الف) محور افقی حدود طبقات، محور عمودی فراوانی نسبی

ب) محور افقی فراوانی نسبی، محور عمودی حدود طبقات

ج) محور افقی حدود بالای طبقات، محور عمودی فراوانی تجمعی یا نسبی

د) محور افقی حدود بالای طبقات، محور عمودی فراوانی نسبی تجمعی

مسائل

۲۱. فرض کنید میانگین و انحراف معیار نمرات نهایی درسی به ترتیب ۷۰ و ۱۴ باشد (توزیع نمرات نرمال است). اگر بخواهیم به ۱۰ درصد افراد کلاس نمرهٔ «الف» ، به ۲۰ درصد نمره «ب» ، به ۴۰ درصد نمره «ج» ، به ۲۰ درصد نمره «د»، و به ۱۰ درصد نمرهٔ «هـ» بدهیم، حداقل نمره برای قرارگرفتن در هریک از گروههای مذکور چقدر است؟

۲۲. دو سینما برای جلب هزار مشتری با هم رقابت می‌کنند. فرض کنید که هریک از مشتریان کاملاً مستقل از دیگران یکی از دو سینما را انتخاب می‌کند. هر سالن سینما باید چند صندلی

داشته باشد تا احتمال جواب کردن مشتری به دلیل کمبود صندلی، کمتر از یک درصد باشد؟

۲۳. وقتی را که شخصی برای رفتن از منزل تا محل کار صرف می‌کند دارای توزیع نرمال با میانگین ۵۰ دقیقه و انحراف معیار ۱۴ دقیقه است.

الف) احتمال اینکه این شخص فردا بیش از یک ساعت از وقتش را برای این کار صرف کند، چقدر است؟

ب) اگر این شخص بخواهد ریسک دیر رسیدن به محل کار را به ۳ درصد کاهش دهد، برای رفتن از منزل به محل کار باید حداقل چند دقیقه از وقت خود را تخصیص دهد؟

۲۴. احتمال اینکه از بین ۱۲ هزار رقم تصادفی، حداکثر ۱۰۵۰ بار رقم ۵ تکرار شده باشد، چقدر است؟

۲۵. زمان مورد انتظار (میانگین) و انحراف معیار زمان انجام دادن فعالیتهای مسیر بحرانی پروژه‌ای در این جدول آورده شده است (واحد زمان هفته است):

فعالیت	میانگین	واریانس	فعالیت	میانگین	واریانس
۱	۳	۱/۲	۶	۳	۱/۰
۲	۲	۰/۸	۷	٤/۲	۲/۱
۳	۱/۸	۰/۵	۸	۱	۰/۳
٤	۵	۲/۰	۹	۰/۵	۰/۲
۵	۳	۱/۵			

الف) احتمال تکمیل پروژه بین ۲۰ تا ۲٦ هفته چقدر است؟

ب) احتمال تکمیل پروژه طی حداکثر ۲۲ هفته چقدر است؟

ج) اگر بخواهیم باکارفرمای پروژه قراردادی منعقد کنیم، چند هفته برای تکمیل آن در نظر بگیریم به طوری که با ۹۵ درصد احتمال مطمئن باشیم پروژه بیشتر از آن مدت طول نخواهد کشید؟

د) کارفرمای پروژه گفته است علاوه بر مبلغ قرارداد، به شرط تکمیل پروژه در زمانی کمتر از ۲۰ هفته، مبلغ ۵ میلیون ریال و به شرط تکمیل پروژه بین ۲۰ تا ۳۰ هفته، مبلغ یک میلیون ریال به عنوان پاداش خواهد پرداخت، ولی اگر بیش از ۳۰ هفته طول بکشد مبلغ ۱۰ میلیون ریال خسارت خواهد گرفت. امید ریاضی دریافت پاداش برای شرکت بابت زمان تکمیل پروژه چقدر است؟

۲٦. کشتارگاهی ٤۷۰ گوسفند را بدون توزین آنها، یکجا می‌خرد. وزن هر گوسفند به طور متوسط

۴۵ کیلوگرم با انحراف معیار ۴ است. احتمال اینکه وزن این ۴۷۰ گوسفند بیش از ۲۲ تن باشد، چقدر است؟

۲۷. تعداد شکایات جنایی که ماهانه به دادگاهی می‌رسد دارای توزیع پواسون با میانگین ۵ شکایت است. احتمال اینکه در طی فصل پاییز بیش از ۱۷ شکایت جنایی به این دادگاه برسد، چقدر است (از تقریب نرمال استفاده کنید)؟

۲۸. توزیع درآمد ماهانهٔ ۱۲۰۰ خانوارِ انتخاب شده از شهری در این جدول آورده شده است:

طبقات (ارقام به ده هزار ریال)	فراوانی
۱۰_۱۵	۱۰۰
۱۵_۲۰	۲۵۰
۲۰_۲۵	۲۳۰
۲۵_۳۰	۲۲۰
۳۰_۳۵	۱۴۰
۳۵_۴۰	۸۰
۴۰_۴۵	۵۵
۴۵_۵۰	۷۰
۵۰_۵۵	۳۳
۵۵_۶۰	۲۲

با استفاده از صفحهٔ احتمال نرمال، تعیین کنید آیا توزیع درآمد این شهر می‌تواند نرمال باشد؟

۲۹. فرض کنید $X \sim N(۲۰, \sigma)$ باشد، اگر $P(X \leq ۲۲ + \sigma) = ۰/۹۷۷۲$ باشد، σ را پیدا کنید.

۳۰. به طور متوسط ۵ درصد از تولیدات روز و ۸ درصد از تولیدات شب کارخانه‌ای معیوب است. تولیدات روز دو برابر شب است. تولیدات روز و شب در یک جا و به صورت مخلوط نگهداری می‌شود. احتمال اینکه مأمور کنترل کیفیت پس از انتخاب ۱۵۰ کالا و آزمایش آنها متوجه وجودِ ۸ تا ۱۲ کالای معیوب شود، چقدر است؟

پاسخنامهٔ سؤالات

۱) غ	۲) غ	۳) ص	۴) غ	۵) غ
۶) غ	۷) ص	۸) غ	۹) غ	۱۰) الف
۱۱) الف	۱۲) ب	۱۳) ج	۱۴) ج	۱۵) ج
۱۶) ج	۱۷) ج	۱۸) ج	۱۹) ج	۲۰) د

پیوست

جدول ۱ احتمالهای تجمعی دو جمله‌ای $P[X \le c] = \sum_{x=0}^{c} \binom{n}{x} p^x (1-p)^{n-x}$

	c	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
$n=1$	0	.950	.900	.800	.700	.600	.500	.400	.300	.200	.100	.050
	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=2$	0	.902	.810	.640	.490	.360	.250	.160	.090	.040	.010	.002
	1	.997	.990	.960	.910	.840	.750	.640	.510	.360	.190	.097
	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=3$	0	.857	.729	.512	.343	.216	.125	.064	.027	.008	.001	.000
	1	.993	.972	.896	.784	.648	.500	.352	.216	.104	.028	.007
	2	1.000	.999	.992	.973	.936	.875	.784	.657	.488	.271	.143
	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=4$	0	.815	.656	.410	.240	.130	.063	.026	.008	.002	.000	.000
	1	.986	.948	.819	.652	.475	.313	.179	.084	.027	.004	.000
	2	1.000	.996	.973	.916	.821	.688	.525	.348	.181	.052	.014
	3	1.000	1.000	.998	.992	.974	.938	.870	.760	.590	.344	.185
	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=5$	0	.774	.590	.328	.168	.078	.031	.010	.002	.000	.000	.000
	1	.977	.919	.737	.528	.337	.188	.087	.031	.007	.000	.000
	2	.999	.991	.942	.837	.683	.500	.317	.163	.058	.009	.001
	3	1.000	1.000	.993	.969	.913	.813	.663	.472	.263	.081	.023
	4	1.000	1.000	1.000	.998	.990	.969	.922	.832	.672	.410	.226
	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=6$	0	.735	.531	.262	.118	.047	.016	.004	.001	.000	.000	.000
	1	.967	.886	.655	.420	.233	.109	.041	.011	.002	.000	.000
	2	.998	.984	.901	.744	.544	.344	.179	.070	.017	.001	.000
	3	1.000	.999	.983	.930	.821	.656	.456	.256	.099	.016	.002
	4	1.000	1.000	.998	.989	.959	.891	.767	.580	.345	.114	.033
	5	1.000	1.000	1.000	.999	.996	.984	.953	.882	.738	.469	.265
	6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=7$	0	.698	.478	.210	.082	.028	.008	.002	.000	.000	.000	.000
	1	.956	.850	.577	.329	.159	.063	.019	.004	.000	.000	.000
	2	.996	.974	.852	.647	.420	.227	.096	.029	.005	.000	.000
	3	1.000	.997	.967	.874	.710	.500	.290	.126	.033	.003	.000
	4	1.000	1.000	.995	.971	.904	.773	.580	.353	.148	.026	.004
	5	1.000	1.000	1.000	.996	.981	.938	.841	.671	.423	.150	.044
	6	1.000	1.000	1.000	1.000	.998	.992	.972	.918	.790	.522	.302
	7	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

		p										
	c	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
$n=8$	0	.663	.430	.168	.058	.017	.004	.001	.000	.000	.000	.000
	1	.943	.813	.503	.255	.106	.035	.009	.001	.000	.000	.000
	2	.994	.962	.797	.552	.315	.145	.050	.011	.001	.000	.000
	3	1.000	.995	.944	.806	.594	.363	.174	.058	.010	.000	.000
	4	1.000	1.000	.990	.942	.826	.637	.406	.194	.056	.005	.000
	5	1.000	1.000	.999	.989	.950	.855	.685	.448	.203	.038	.006
	6	1.000	1.000	1.000	.999	.991	.965	.894	.745	.497	.187	.057
	7	1.000	1.000	1.000	1.000	.999	.996	.983	.942	.832	.570	.337
	8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=9$	0	.630	.387	.134	.040	.010	.002	.000	.000	.000	.000	.000
	1	.929	.775	.436	.196	.071	.020	.004	.000	.000	.000	.000
	2	.992	.947	.738	.463	.232	.090	.025	.004	.000	.000	.000
	3	.999	.992	.914	.730	.483	.254	.099	.025	.003	.000	.000
	4	1.000	.999	.980	.901	.733	.500	.267	.099	.020	.001	.000
	5	1.000	1.000	.997	.975	.901	.746	.517	.270	.086	.008	.001
	6	1.000	1.000	1.000	.996	.975	.910	.768	.537	.262	.053	.008
	7	1.000	1.000	1.000	1.000	.996	.980	.929	.804	.564	.225	.071
	8	1.000	1.000	1.000	1.000	1.000	.998	.990	.960	.866	.613	.370
	9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=10$	0	.599	.349	.107	.028	.006	.001	.000	.000	.000	.000	.000
	1	.914	.736	.376	.149	.046	.011	.002	.000	.000	.000	.000
	2	.988	.930	.678	.383	.167	.055	.012	.002	.000	.000	.000
	3	.999	.987	.879	.650	.382	.172	.055	.011	.001	.000	.000
	4	1.000	.998	.967	.850	.633	.377	.166	.047	.006	.000	.000
	5	1.000	1.000	.994	.953	.834	.623	.367	.150	.033	.002	.000
	6	1.000	1.000	.999	.989	.945	.828	.618	.350	.121	.013	.001
	7	1.000	1.000	1.000	.998	.988	.945	.833	.617	.322	.070	.012
	8	1.000	1.000	1.000	1.000	.998	.989	.954	.851	.624	.264	.086
	9	1.000	1.000	1.000	1.000	1.000	.999	.994	.972	.893	.651	.401
	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=11$	0	.569	.314	.086	.020	.004	.000	.000	.000	.000	.000	.000
	1	.898	.697	.322	.113	.030	.006	.001	.000	.000	.000	.000
	2	.985	.910	.617	.313	.119	.033	.006	.001	.000	.000	.000
	3	.998	.981	.839	.570	.296	.113	.029	.004	.000	.000	.000
	4	1.000	.997	.950	.790	.533	.274	.099	.022	.002	.000	.000
	5	1.000	1.000	.988	.922	.753	.500	.247	.078	.012	.000	.000
	6	1.000	1.000	.998	.978	.901	.726	.467	.210	.050	.003	.000
	7	1.000	1.000	1.000	.996	.971	.887	.704	.430	.161	.019	.002
	8	1.000	1.000	1.000	.999	.994	.967	.881	.687	.383	.090	.015
	9	1.000	1.000	1.000	1.000	.999	.994	.970	.887	.678	.303	.102
	10	1.000	1.000	1.000	1.000	1.000	1.000	.996	.980	.914	.686	.431
	11	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

	c	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
							p					
$n=12$	0	.540	.282	.069	.014	.002	.000	.000	.000	.000	.000	.000
	1	.882	.659	.275	.085	.020	.003	.000	.000	.000	.000	.000
	2	.980	.889	.558	.253	.083	.019	.003	.000	.000	.000	.000
	3	.998	.974	.795	.493	.225	.073	.015	.002	.000	.000	.000
	4	1.000	.996	.927	.724	.438	.194	.057	.009	.001	.000	.000
	5	1.000	.999	.981	.882	.665	.387	.158	.039	.004	.000	.000
	6	1.000	1.000	.996	.961	.842	.613	.335	.118	.019	.001	.000
	7	1.000	1.000	.999	.991	.943	.806	.562	.276	.073	.004	.000
	8	1.000	1.000	1.000	.998	.985	.927	.775	.507	.205	.026	.002
	9	1.000	1.000	1.000	1.000	.997	.981	.917	.747	.442	.111	.020
	10	1.000	1.000	1.000	1.000	1.000	.997	.980	.915	.725	.341	.118
	11	1.000	1.000	1.000	1.000	1.000	1.000	.998	.986	.931	.718	.460
	12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=13$	0	.513	.254	.055	.010	.001	.000	.000	.000	.000	.000	.000
	1	.865	.621	.234	.064	.013	.002	.000	.000	.000	.000	.000
	2	.975	.866	.502	.202	.058	.011	.001	.000	.000	.000	.000
	3	.997	.966	.747	.421	.169	.046	.008	.001	.000	.000	.000
	4	1.000	.994	.901	.654	.353	.133	.032	.004	.000	.000	.000
	5	1.000	.999	.970	.835	.574	.291	.098	.018	.001	.000	.000
	6	1.000	1.000	.993	.938	.771	.500	.229	.062	.007	.000	.000
	7	1.000	1.000	.999	.982	.902	.709	.426	.165	.030	.001	.000
	8	1.000	1.000	1.000	.996	.968	.867	.647	.346	.099	.006	.000
	9	1.000	1.000	1.000	.999	.992	.954	.831	.579	.253	.034	.003
	10	1.000	1.000	1.000	1.000	.999	.989	.942	.798	.498	.134	.025
	11	1.000	1.000	1.000	1.000	1.000	.998	.987	.936	.766	.379	.135
	12	1.000	1.000	1.000	1.000	1.000	1.000	.999	.990	.945	.746	.487
	13	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=14$	0	.488	.229	.044	.007	.001	.000	.000	.000	.000	.000	.000
	1	.847	.585	.198	.047	.008	.001	.000	.000	.000	.000	.000
	2	.970	.842	.448	.161	.040	.006	.001	.000	.000	.000	.000
	3	.996	.956	.698	.355	.124	.029	.004	.000	.000	.000	.000
	4	1.000	.991	.870	.584	.279	.090	.018	.002	.000	.000	.000
	5	1.000	.999	.956	.781	.486	.212	.058	.008	.000	.000	.000
	6	1.000	1.000	.988	.907	.692	.395	.150	.031	.002	.000	.000
	7	1.000	1.000	.998	.969	.850	.605	.308	.093	.012	.000	.000
	8	1.000	1.000	1.000	.992	.942	.788	.514	.219	.044	.001	.000
	9	1.000	1.000	1.000	.998	.982	.910	.721	.416	.130	.009	.000
	10	1.000	1.000	1.000	1.000	.996	.971	.876	.645	.302	.044	.004
	11	1.000	1.000	1.000	1.000	.999	.994	.960	.839	.552	.158	.030
	12	1.000	1.000	1.000	1.000	1.000	.999	.992	.953	.802	.415	.153
	13	1.000	1.000	1.000	1.000	1.000	1.000	.999	.993	.956	.771	.512
	14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

	c	.05	.10	.20	.30	.40	p .50	.60	.70	.80	.90	.95
n = 15	0	.463	.206	.035	.005	.000	.000	.000	.000	.000	.000	.000
	1	.829	.549	.167	.035	.005	.000	.000	.000	.000	.000	.000
	2	.964	.816	.398	.127	.027	.004	.000	.000	.000	.000	.000
	3	.995	.944	.648	.297	.091	.018	.002	.000	.000	.000	.000
	4	.999	.987	.836	.515	.217	.059	.009	.001	.000	.000	.000
	5	1.000	.998	.939	.722	.403	.151	.034	.004	.000	.000	.000
	6	1.000	1.000	.982	.869	.610	.304	.095	.015	.001	.000	.000
	7	1.000	1.000	.996	.950	.787	.500	.213	.050	.004	.000	.000
	8	1.000	1.000	.999	.985	.905	.696	.390	.131	.018	.000	.000
	9	1.000	1.000	1.000	.996	.966	.849	.597	.278	.061	.002	.000
	10	1.000	1.000	1.000	.999	.991	.941	.783	.485	.164	.013	.001
	11	1.000	1.000	1.000	1.000	.998	.982	.909	.703	.352	.056	.005
	12	1.000	1.000	1.000	1.000	1.000	.996	.973	.873	.602	.184	.036
	13	1.000	1.000	1.000	1.000	1.000	1.000	.995	.965	.833	.451	.171
	14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.995	.965	.794	.537
	15	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
n = 16	0	.440	.185	.028	.003	.000	.000	.000	.000	.000	.000	.000
	1	.811	.515	.141	.026	.003	.000	.000	.000	.000	.000	.000
	2	.957	.789	.352	.099	.018	.002	.000	.000	.000	.000	.000
	3	.993	.932	.598	.246	.065	.011	.001	.000	.000	.000	.000
	4	.999	.983	.798	.450	.167	.038	.005	.000	.000	.000	.000
	5	1.000	.997	.918	.660	.329	.105	.019	.002	.000	.000	.000
	6	1.000	.999	.973	.825	.527	.227	.058	.007	.000	.000	.000
	7	1.000	1.000	.993	.926	.716	.402	.142	.026	.001	.000	.000
	8	1.000	1.000	.999	.974	.858	.598	.284	.074	.007	.000	.000
	9	1.000	1.000	1.000	.993	.942	.773	.473	.175	.027	.001	.000
	10	1.000	1.000	1.000	.998	.981	.895	.671	.340	.082	.003	.000
	11	1.000	1.000	1.000	1.000	.995	.962	.833	.550	.202	.017	.001
	12	1.000	1.000	1.000	1.000	.999	.989	.935	.754	.402	.068	007
	13	1.000	1.000	1.000	1.000	1.000	.998	.982	.901	.648	.211	.043
	14	1.000	1.000	1.000	1.000	1.000	1.000	.997	.974	.859	.485	.189
	15	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.997	.972	.815	.560
	16	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
n = 17	0	.418	.167	.023	.002	.000	.000	.000	.000	.000	.000	.000
	1	.792	.482	.118	.019	.002	.000	.000	.000	.000	.000	.000
	2	.950	.762	.310	.077	.012	.001	.000	.000	.000	.000	.000
	3	.991	.917	.549	.202	.046	.006	.000	.000	.000	.000	.000
	4	.999	.978	.758	.389	.126	.025	.003	.000	.000	.000	.000
	5	1.000	.995	.894	.597	.264	.072	.011	.001	.000	.000	.000
	6	1.000	.999	.962	.775	.448	.166	.035	.003	.000	.000	.000
	7	1.000	1.000	.989	.895	.641	.315	.092	.013	.000	.000	.000
	8	1.000	1.000	.997	.960	.801	500	.199	.040	.003	.000	.000
	9	1.000	1.000	1.000	.987	.908	.685	.359	.105	.011	.000	.000
	10	1.000	1.000	1.000	.997	.965	.834	.552	.225	.038	.001	.000

	c	.05	.10	.20	.30	.40	p .50	.60	.70	.80	.90	.95
	11	1.000	1.000	1.000	.999	.989	.928	.736	.403	.106	.005	.000
	12	1.000	1.000	1.000	1.000	.997	.975	.874	.611	.242	.022	.001
	13	1.000	1.000	1.000	1.000	1.000	.994	.954	.798	.451	.083	.009
	14	1.000	1.000	1.000	1.000	1.000	.999	.988	.923	.690	.238	.050
	15	1.000	1.000	1.000	1.000	1.000	1.000	.998	.981	.882	.518	.208
	16	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.998	.977	.833	.582
	17	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=18$	0	.397	.150	.018	.002	.000	.000	.000	.000	.000	.000	.000
	1	.774	.450	.099	.014	.001	.000	.000	.000	.000	.000	.000
	2	.942	.734	.271	.060	.008	.001	.000	.000	.000	.000	.000
	3	.989	.902	.501	.165	.033	.004	.000	.000	.000	.000	.000
	4	.998	.972	.716	.333	.094	.015	.001	.000	.000	.000	.000
	5	1.000	.994	.867	.534	.209	.048	.006	.000	.000	.000	.000
	6	1.000	.999	.949	.722	.374	.119	.020	.001	.000	.000	.000
	7	1.000	1.000	.984	.859	.563	.240	.058	.006	.000	.000	.000
	8	1.000	1.000	.996	.940	.737	.407	.135	.021	.001	.000	.000
	9	1.000	1.000	.999	.979	.865	.593	.263	.060	.004	.000	.000
	10	1.000	1.000	1.000	.994	.942	.760	.437	.141	.016	.000	.000
	11	1.000	1.000	1.000	.999	.980	.881	.626	.278	.051	.001	.000
	12	1.000	1.000	1.000	1.000	.994	.952	.791	.466	.133	.006	.000
	13	1.000	1.000	1.000	1.000	.999	.985	.906	.667	.284	.028	.002
	14	1.000	1.000	1.000	1.000	1.000	.996	.967	.835	.499	.098	.011
	15	1.000	1.000	1.000	1.000	1.000	.999	.992	.940	.729	.266	.058
	16	1.000	1.000	1.000	1.000	1.000	1.000	.999	.986	.901	.550	.226
	17	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.998	.982	.850	.603
	18	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=19$	0	.377	.135	.014	.001	.000	.000	.000	.000	.000	.000	.000
	1	.755	.420	.083	.010	.001	.000	.000	.000	.000	.000	.000
	2	.933	.705	.237	.046	.005	.000	.000	.000	.000	.000	.000
	3	.987	.885	.455	.133	.023	.002	.000	.000	.000	.000	.000
	4	.998	.965	.673	.282	.070	.010	.001	.000	.000	.000	.000
	5	1.000	.991	.837	.474	.163	.032	.003	.000	.000	.000	.000
	6	1.000	.998	.932	.666	.308	.084	.012	.001	.000	.000	.000
	7	1.000	1.000	.977	.818	.488	.180	.035	.003	.000	.000	.000
	8	1.000	1.000	.993	.916	.667	.324	.088	.011	.000	.000	.000
	9	1.000	1.000	.998	.967	.814	.500	.186	.033	.002	.000	.000
	10	1.000	1.000	1.000	.989	.912	.676	.333	.084	.007	.000	.000
	11	1.000	1.000	1.000	.997	.965	.820	.512	.182	.023	.000	.000
	12	1.000	1.000	1.000	.999	.988	.916	.692	.334	.068	.002	.000
	13	1.000	1.000	1.000	1.000	.997	.968	.837	.526	.163	.009	.000
	14	1.000	1.000	1.000	1.000	.999	.990	.930	.718	.327	.035	.002
	15	1.000	1.000	1.000	1.000	1.000	.998	.977	.867	.545	.115	.013
	16	1.000	1.000	1.000	1.000	1.000	1.000	.995	.954	.763	.295	.067

	c	.05	.10	.20	.30	.40	p .50	.60	.70	.80	.90	.95
	17	1.000	1.000	1.000	1.000	1.000	1.000	.999	.990	.917	.580	.245
	18	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.986	.865	.623
	19	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=20$	0	.358	.122	.012	.001	.000	.000	.000	.000	.000	.000	.000
	1	.736	.392	.069	.008	.001	.000	.000	.000	.000	.000	.000
	2	.925	.677	.206	.035	.004	.000	.000	.000	.000	.000	.000
	3	.984	.867	.411	.107	.016	.001	.000	.000	.000	.000	.000
	4	.997	.957	.630	.238	.051	.006	.000	.000	.000	.000	.000
	5	1.000	.989	.804	.416	.126	.021	.002	.000	.000	.000	.000
	6	1.000	.998	.913	.608	.250	.058	.006	.000	.000	.000	.000
	7	1.000	1.000	.968	.772	.416	.132	.021	.001	.000	.000	.000
	8	1.000	1.000	.990	.887	.596	.252	.057	.005	.000	.000	.000
	9	1.000	1.000	.997	.952	.755	.412	.128	.017	.001	.000	.000
	10	1.000	1.000	.999	.983	.872	.588	.245	.048	.003	.000	.000
	11	1.000	1.000	1.000	.995	.943	.748	.404	.113	.010	.000	.000
	12	1.000	1.000	1.000	.999	.979	.868	.584	.228	.032	.000	.000
	13	1.000	1.000	1.000	1.000	.994	.942	.750	.392	.087	.002	.000
	14	1.000	1.000	1.000	1.000	.998	.979	.874	.584	.196	.011	.000
	15	1.000	1.000	1.000	1.000	1.000	.994	.949	.762	.370	.043	.003
	16	1.000	1.000	1.000	1.000	1.000	.999	.984	.893	.589	.133	.016
	17	1.000	1.000	1.000	1.000	1.000	1.000	.996	.965	.794	.323	.075
	18	1.000	1.000	1.000	1.000	1.000	1.000	.999	.992	.931	.608	.264
	19	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.988	.878	.642
	20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n=25$	0	.277	.072	.004	.000	.000	.000	.000	.000	.000	.000	.000
	1	.642	.271	.027	.002	.000	.000	.000	.000	.000	.000	.000
	2	.873	.537	.098	.009	.000	.000	.000	.000	.000	.000	.000
	3	.966	.764	.234	.033	.002	.000	.000	.000	.000	.000	.000
	4	.993	.902	.421	.090	.009	.000	.000	.000	.000	.000	.000
	5	.999	.967	.617	.193	.029	.002	.000	.000	.000	.000	.000
	6	1.000	.991	.780	.341	.074	.007	.000	.000	.000	.000	.000
	7	1.000	.998	.891	.512	.154	.022	.001	.000	.000	.000	.000
	8	1.000	1.000	.953	.677	.274	.054	.004	.000	.000	.000	.000
	9	1.000	1.000	.983	.811	.425	.115	.013	.000	.000	.000	.000
	10	1.000	1.000	.994	.902	.586	.212	.034	.002	.000	.000	.000
	11	1.000	1.000	.998	.956	.732	.345	.078	.006	.000	.000	.000
	12	1.000	1.000	1.000	.983	.846	.500	.154	.017	.000	.000	.000
	13	1.000	1.000	1.000	.994	.922	.655	.268	.044	.002	.000	.000
	14	1.000	1.000	1.000	.998	.966	.788	.414	.098	.006	.000	.000
	15	1.000	1.000	1.000	1.000	.987	.885	.575	.189	.017	.000	.000
	16	1.000	1.000	1.000	1.000	.996	.946	.726	.323	.047	.000	.000
	17	1.000	1.000	1.000	1.000	.999	.978	.846	.488	.109	.002	.000

ادامه جدول ۱

	p										
c	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
18	1.000	1.000	1.000	1.000	1.000	.993	.926	.659	.220	.009	.000
19	1.000	1.000	1.000	1.000	1.000	.998	.971	.807	.383	.033	.001
20	1.000	1.000	1.000	1.000	1.000	1.000	.991	.910	.579	.098	.007
21	1.000	1.000	1.000	1.000	1.000	1.000	.998	.967	.766	.236	.034
22	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.991	.902	.463	.127
23	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.998	.973	.729	.358
24	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.996	.928	.723
25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

جدول ۲ احتمالهای تجمعی پواسون $P[X \le c] = \sum_{x=0}^{c} \frac{e^{-\lambda}\lambda^{x}}{x!}$

	λ									
c	.10	.20	.30	.40	.50	.60	.70	.80	.90	1.00
0	.905	.819	.741	.670	.607	.549	.497	.449	.407	.368
1	.995	.982	.963	.938	.910	.878	.844	.809	.772	.736
2	1.000	.999	.996	.992	.986	.977	.966	.953	.937	.920
3	1.000	1.000	1.000	.999	.998	.997	.994	.991	.987	.981
4	1.000	1.000	1.000	1.000	1.000	1.000	.999	.999	.998	.996
5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999
6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

	λ									
c	1.10	1.20	1.30	1.40	1.50	1.60	1.70	1.80	1.90	2.00
0	.333	.301	.273	.247	.223	.202	.183	.165	.150	.135
1	.699	.663	.627	.592	.558	.525	.493	.463	.434	.406
2	.900	.879	.857	.833	.809	.783	.757	.731	.704	.677
3	.974	.966	.957	.946	.934	.921	.907	.891	.875	.857
4	.995	.992	.989	.986	.981	.976	.970	.964	.956	.947
5	.999	.998	.998	.997	.996	.994	.992	.990	.987	.983
6	1.000	1.000	1.000	.999	.999	.999	.998	.997	.997	.995
7	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.999	.999
8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

	λ									
c	2.10	2.20	2.30	2.40	2.50	2.60	2.70	2.80	2.90	3.00
0	.122	.111	.100	.091	.082	.074	.067	.061	.055	.050
1	.380	.355	.331	.308	.287	.267	.249	.231	.215	.199
2	.650	.623	.596	.570	.544	.518	.494	.469	.446	.423
3	.839	.819	.799	.779	.758	.736	.714	.692	.670	.647
4	.938	.928	.916	.904	.891	.877	.863	.848	.832	.815
5	.980	.975	.970	.964	.958	.951	.943	.935	.926	.916
6	.994	.993	.991	.988	.986	.983	.979	.976	.971	.966
7	.999	.998	.997	.997	.996	.995	.993	.992	.990	.988
8	1.000	1.000	.999	.999	.999	.999	.998	.998	.997	.996
9	1.000	1.000	1.000	1.000	1.000	1.000	.999	.999	.999	.999
10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
11	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

c	λ 3.10	3.20	3.30	3.40	3.50	3.60	3.70	3.80	3.90	4.00
0	.045	.041	.037	.033	.030	.027	.025	.022	.020	.018
1	.185	.171	.159	.147	.136	.126	.116	.107	.099	.092
2	.401	.380	.359	.340	.321	.303	.285	.269	.253	.238
3	.625	.603	.580	.558	.537	.515	.494	.473	.453	.433
4	.798	.781	.763	.744	.725	.706	.687	.668	.648	.629
5	.906	.895	.883	.871	.858	.844	.830	.816	.801	.785
6	.961	.955	.949	.942	.935	.927	.918	.909	.899	.889
7	.986	.983	.980	.977	.973	.969	.965	.960	.955	.949
8	.995	.994	.993	.992	.990	.988	.986	.984	.981	.979
9	.999	.998	.998	.997	.997	.996	.995	.994	.993	.992
10	1.000	1.000	.999	.999	.999	.999	.998	.998	.998	.997
11	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.999	.999
12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
13	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

c	λ 4.50	5.00	5.50	6.00	6.50	7.00	7.50	8.00	8.50	9.00
0	.011	.007	.004	.002	.002	.001	.001	.000	.000	.000
1	.061	.040	.027	.017	.011	.007	.005	.003	.002	.001
2	.174	.125	.088	.062	.043	.030	.020	.014	.009	.006
3	.342	.265	.202	.151	.112	.082	.059	.042	.030	.021
4	.532	.440	.358	.285	.224	.173	.132	.100	.074	.055
5	.703	.616	.529	.446	.369	.301	.241	.191	.150	.116
6	.831	.762	.686	.606	.527	.450	.378	.313	.256	.207
7	.913	.867	.809	.744	.673	.599	.525	.453	.386	.324
8	.960	.932	.894	.847	.792	.729	.662	.593	.523	.456
9	.983	.968	.946	.916	.877	.830	.776	.717	.653	.587
10	.993	.986	.975	.957	.933	.901	.862	.816	.763	.706
11	.998	.995	.989	.980	.966	.947	.921	.888	.849	.803
12	.999	.998	.996	.991	.984	.973	.957	.936	.909	.876
13	1.000	.999	.998	.996	.993	.987	.978	.966	.949	.926
14	1.000	1.000	.999	.999	.997	.994	.990	.983	.973	.959
15	1.000	1.000	1.000	.999	.999	.998	.995	.992	.986	.978
16	1.000	1.000	1.000	1.000	1.000	.999	.998	.996	.993	.989
17	1.000	1.000	1.000	1.000	1.000	1.000	.999	.998	.997	.995
18	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.999	.998
19	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.999
20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
21	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
22	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

جدول ۳

احتمالهای نرمال استاندارد

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−.7	.2420	.2389	.2358	.2327	.2297	.2266	.2236	.2206	.2177	.2148
−.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

ادامه جدول ۳

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998

منابع و مآخذ

۱. آذر، عادل؛ «تبیین آماری فرضیات در پژوهشهای رفتاری ـ مدیریتی»، فصلنامه علمی پژوهشی دانش مدیریت؛ دانشکده علوم اداری و مدیریت بازرگانی دانشگاه تهران، شماره ۲۶، پاییز ۱۳۷۳.

۲. باتاچاریا، گوری کی. و ریچارد ای. جانسون؛ *مفاهیم و روشهای آماری*؛ ج ۱، ترجمه مرتضی ابن‌شهر آشوب و فتاح میکائیلی، تهران: مرکز نشر دانشگاهی ۱۳٦٤.

۳. رضائیان، علی؛ *مدیریت رفتار سازمانی*؛ ج ۱، انتشارات دانشکده علوم اداری و مدیریت بازرگانی دانشگاه تهران، خرداد ۱۳۷۲.

٤. مدنی، علی؛ *مفاهیم اساسی آمار*؛ ج ۱، تهران: انتشارات فروردین، ۱۳٦٦.

٥. نصفت، مرتضی؛ *اصول و روشهای آماری*؛ ج ۱، انتشارات دانشگاه تهران، ۱۳۷۲.

٦. هاف، تارل؛ *چگونه با آمار دروغ می‌گویند؟*؛ ترجمه مهدی تقوی، اسفند ۱۳۷۱.

7. Ben-Horim, M. and H. Levy; *Statistics*:Decision and Applications in Business; 2nd ed., New York: Random House, Inc., 1984.

8. Bhatacharyya, Gouri K. and Richard A. Johnson; *Statistical Concepts and Methods*; John Wiley and Sons' Pub., 1981.

9. Cangelosi, V. E., P. H. Taylor and P. F. Rice; *Basic Statistics: A Real World Approach*; 3rd ed., St. Paul' West, 1983.

10. Chambers, J. W. et al.; *Graphical Methods for Data Analysis*; Boston: Duxbury Press, 1983.

11. Chung, Kai Lai; *Elementary Probability Theory With Stochastic Processes*; 2nd ed., Springer-Verlang, 1989.

12. Croft, D. J. et al.; *Statistical Inference for Management and Economics*; 3rd ed., Boston: Massachusetts, Allyn and Bacon, Inc., 1992.

13. Everitt, B. S. and G. Dunn; *Advanced Methods of Data Exploration and Modelling*; London: Heinemann Education Books, Ltd; 1991.

14. Freund, John E.; *Modern Elementary Statistics*; 6th ed., Prentice-Hall, 1984.

15. Freund, John E. and Ronald E. Walpole; *Mathematical Statistics*; 3rd ed. Prentice-Hall, 1980.

16. Hines, William W. and Douglas C. Montgomery; *Probability and Statistics in Engineering and Management Science*; 3rd ed., John Wiley and Sons', 1990.

17. Huff, D. and I. Geis; *How to Lie With Statistics*; New York: Norton, 1970.

18. Kazmier, Leonard J.; *Business Statistics*; McGraw-Hill, 1976.

19. Koopman, L. H.; *An Introduction to Contemporary Statistics*; 2nd ed. Boston: Pws-Kent, 1988.

20. Levin, Richard I.; *Statistics for Management*; 4th ed., New York: Prentice-Hall Pub., 1990.

21. Levin, Richard I. and David S. Rubin; *Statistics for Management*; 5th ed., Prentice-Hall, 1991.

22. Lipschuts, Seymour; *Theory and Problems of Probability*; McGraw-Hill, 1977.

23. Mendenhall and Reinmuth and Beaver; *Statistics for Management and Economics*; 6th ed., New York: Pws-Kent Publishing Company, 1989.

24. Moscowitz, H. and P. Wright; *Statistics for Management and Economics*; Ohio: C. E. Merill Publishing Company, 1990.

25. New Bold, P.; *Statistics for Business and Economics*; New Jersey: Englewood Cliffs, Prentice-Hall, Inc. 1989.

26. Richman, W. J.; *Use and Abuse of Statistics*; London: Methuen, 1981.

27. Rouncefield, Mary and P. Holmes; *Practical Statistics*; London: Macmillan Education Ltd., 1991.

28. Shavelson, Richard J.; *Statistics Reasoning for Behavioral Sciences*; New York: Allyn and Bacon, Inc., 1988.

29. Ullmann, John E.; *Quantitative Methods in Management*; McGraw-Hill, 1987.

30. Velleman, P. E. and D. Hoaglin; *Applications, Basics and Computing of Exploratory Data Analysis*; Boston: Duxbury Press, 1981.

31. Wonnacott, Thomas H. and Ronald J. Wonnacott; *Introductory Statistics*; 3rd ed.; New York: John Wiley and Sons', 1977.

واژه‌نامه

addition rule	قانون جمع
arithmetic mean	میانگین حسابی
average	متوسط، میانگین
bar chart	نمودار میله‌ای (ستونی)
Bayes' theorem	قضیهٔ بیز
Bernoulli distribution	توزیع برنولی
bias	اریبی
biased	اریب
binominal distribution	توزیع دو جمله‌ای
central limit theorem	قضیهٔ حد مرکزی
central parameters	پارامترهای مرکزی
class	طبقه
class interval	فاصله طبقات
conditional probability	احتمال شرطی
continuous random variable	متغیر تصادفی پیوسته
counting techniques	روشهای شمارش
data resources	منابع داده‌ها
decision tree	درخت تصمیم
degree of freedom	درجه آزادی
density function	تابع چگالی
discrete random variable	متغیر تصادفی گسسته
dispersion	پراکندگی
event	پیشامد
expected value	امید ریاضی
experiment	آزمایش
exponential distribution	توزیع‌نمایی
finite population correction factor	ضریب تصحیح جامعه محدود
frame	چهارچوب
frequency polygon	چند ضلعی فراوانی
frequency distribution	توزیع فراوانی
geometric mean	میانگین هندسی
histogram	هیستوگرام، بافت‌نگار
hypergeometric distribution	توزیع فوق هندسی
independence	استقلال
independent events	پیشامدهای مستقل
independent variable	متغیر مستقل
infinite population	جامعه نامحدود
interquartile	بین چارکی
irregular variation	پراکندگی نامنظم
joint probability	احتمال توأم
marginal probability	احتمال حاشیه‌ای
mean absolute deviation	انحراف متوسط
median	میانه
mode	مد، نما

multiplication rule قانون ضرب
mutually exclusive events پیشامدهای ناسازگار
normal approximation تقریب نرمال
normal curve منحنی نرمال
normal distribution توزیع نرمال
outcome پیامد
permutation ترتیب
population جامعه
posterior probability احتمال پسین
prior probability احتمال پیشین
probability احتمال
probability distribution توزیع احتمال
qualitative variable متغیر کیفی
random variable متغیر تصادفی
range دامنه
relative frequency فراوانی نسبی
sample space فضای نمونه
skewed to left چوله به چپ
skewed to right چوله به راست
skewness چولگی
standard deviation انحراف معیار
subjective (personal) probability احتمال ذهنی (شخصی)
success موفقیت
symmetrical متقارن
table جدول
tree diagram نمودار درختی
uniform distribution توزیع یکنواخت
unimodal یک نمایی
weighted mean میانگین موزون
without replacement sampling نمونه گیری بدون جایگذاری