آمار و احتمالات مهندسی

توزیع‌های دوبعدی و ضریب همبستگی

دکتر بهناز بیگدلی

دانشکده مهندسی عمران

دانشگاه صنعتی شاهرود

ساختار درس

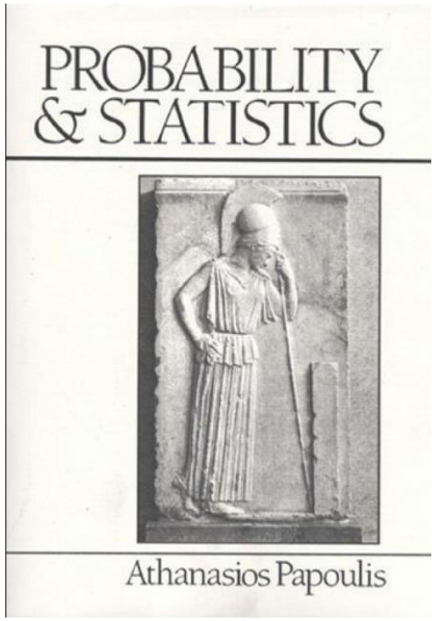
نحوه ارزیابی

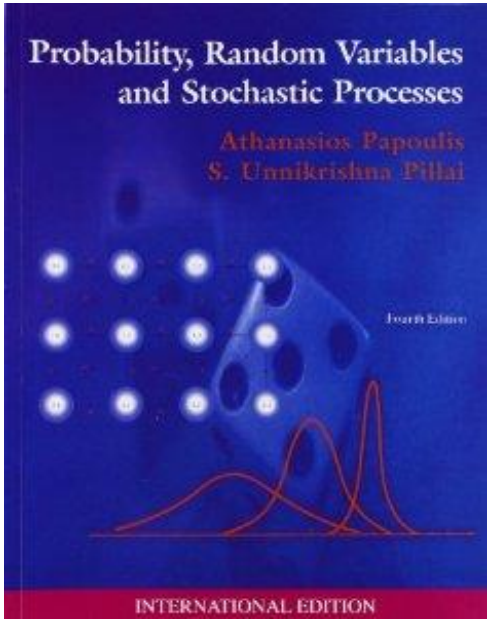
امتحان پایان ترم (۱۴)

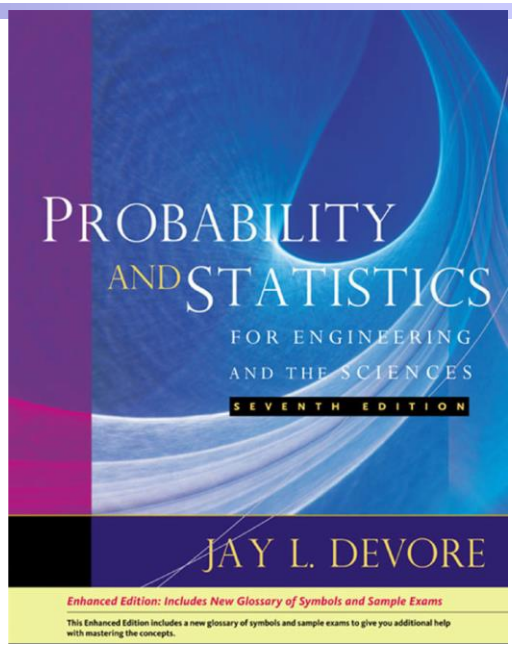
امتحان میان ترم (۴)

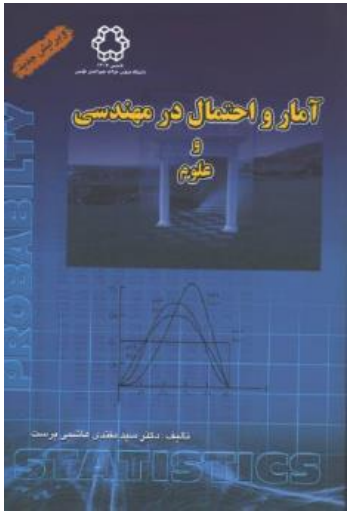
تمرینات و کوئیزها (۲)

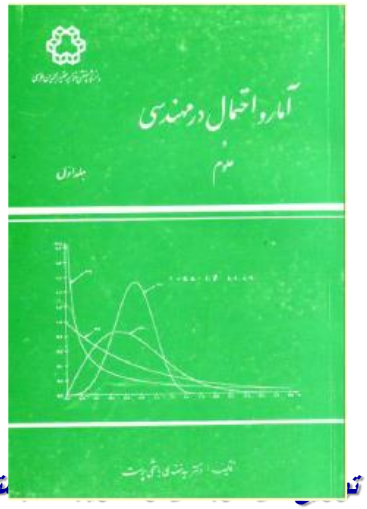
منابع


PROBABILITY
& STATISTICS
Athanasios Papoulis


Probability, Random Variables
and Stochastic Processes
Athanasios Papoulis
S. Unnikrishna Pillai
Fourth Edition
INTERNATIONAL EDITION


PROBABILITY
AND STATISTICS
FOR ENGINEERING
AND THE SCIENCES
SEVENTH EDITION
JAY L. DEVORE
Enhanced Edition: Includes New Glossary of Symbols and Sample Exams
This Enhanced Edition includes a new glossary of symbols and sample exams to give you additional help with mastering the concepts.


آمار و احتمال در مهندسی
و
علوم


آمار و احتمال در مهندسی

فهرست و عناوین درس

I. بخش اول: آمار مقدماتی (توصیفی)

1. آمار مقدماتی
2. اندازه‌گیری پراکندگی، گشاورها، چولگی و کشیدگی
3. برازش خط و منحنی بر داده‌ها
4. توزیع‌های دو بعدی و ضریب همبستگی

II. احتمال

1. احتمال و فضای نمونه
2. فضای نمونه با عناصر متعدد
3. احتمالات شرطی و نابستگی
4. تابع چگالی احتمال، تابع توزیع و امید ریاضی
5. توزیع‌های گسسته
6. توزیع‌های پیوسته مهم
7. نظریه برآورد
8. آزمون‌های فرض

توزیع‌های دوبعدی و ضریب همبستگی

فهرست مطالب این فصل:

۱- مقدمه-توزیع دوبعدی
۲- گشتاورهای دوبعدی
۳- کوواریانس (واریانس مشترک)
۴- نمودارهای پراکندگی
۵- ضریب همبستگی
۶- ضریب همبستگی رتبه‌ای اسپیرمن
۷- ضریب تعیین
۸- معادله خط رگرسیون
۹- خطای استاندار برآورد
۱۰- محاسبه زاویه بین دو خط رگرسیون
۱۱-ضرایب هبستگی داده‌های گروه‌بندی شده

مقدمه-توزیع دوبعدی

- ✓ حوادث متعددی در طبیعت اتفاق می‌افتد که بین آنها همبستگی یا رابطه وجود دارد. منظور از رابطه بین متغیرها وجود رابطه علت معلولی نیست.

نکته: همبستگی رابطه بین دو متغیر در جامعه را توصیف می‌کند که در متغییرها، یکی را X و دیگری را Y می‌نامند.

مقدمه-توزیع دوبعدی

✓ نکته: ضریب همبستگی با r_{xy} نشان داده می‌شود. ضریب همبستگی هم جهت رابطه و هم شدت رابطه را بیان می‌کند.

نکته: جهت همبستگی با علامت -/+ نشان داده می‌شود و شدت همبستگی با قدر مطلق .

نکته: بالاترین ضریب همبستگی ± 1 می‌باشد.

مقدمه-توزیع دوبعدی

✓ اگر یک جامعه آماری nتایی، دو صفت X,Y را اندازه‌گیری کنیم و آنها را بر حسب X,Y دسته‌بندی کنیم، جدول زیر بدست می‌آید.

	$a'_1 - b'_1$	$a'_2 - b'_2$	.....	$a'_m - b'_m$	
$a_1 - b_1$	f_{11}	f_{12}	.....	f_{1m}	f_{1o}
$a_2 - b_2$	f_{21}	f_{22}	.....	f_{2m}	f_{2o}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$a_n - b_n$	f_{n1}	f_{n2}	.....	f_{nm}	f_{no}
	f_{o1}	f_{o2}	.....	f_{om}	

مقدمه-توزیع دوبعدی

	$a'_1 - b'_1$	$a'_2 - b'_2$	.....	$a'_m - b'_m$	
$a_1 - b_1$	f_{11}	f_{12}	.....	f_{1m}	f_{1o}
$a_2 - b_2$	f_{21}	f_{22}	.....	f_{2m}	f_{2o}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$a_n - b_n$	f_{n1}	f_{n2}	.....	f_{nm}	f_{no}
	f_{o1}	f_{o2}	.....	f_{om}	

$$f_{io} = \sum_{j=1}^{m} f_{ij} \qquad f_{oj} = \sum_{i=1}^{n} f_{ij} \qquad N = \sum_{j=1}^{m} \sum_{i=1}^{n} f_{ij}$$

مقدمه-توزیع دوبعدی

	$a'_1 - b'_1$	$a'_2 - b'_2$	.....	$a'_m - b'_m$	
$a_1 - b_1$	f_{11}	f_{12}	.....	f_{1m}	f_{1o}
$a_2 - b_2$	f_{21}	f_{22}	.....	f_{2m}	f_{2o}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$a_n - b_n$	f_{n1}	f_{n2}	.....	f_{nm}	f_{no}
	f_{o1}	f_{o2}	.....	f_{om}	

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{n} f_{io}x_i \qquad \sigma_X^2 = \frac{1}{N}\sum_{i=1}^{n} f_{io}(x_i - \bar{X})^2$$

$$\bar{Y} = \frac{1}{N}\sum_{j=1}^{m} f_{oj}y_j \qquad \sigma_Y^2 = \frac{1}{N}\sum_{i=1}^{m} f_{oj}(y_j - \bar{Y})^2$$

گشتاورهای دوبعدی

- اگر بخواهیم X,Y را همزمان و در کنار هم مورد مطالعه قرار دهیم، لازم است پارامترهای مناسبی تعریف کنیم. گشتاور $\mu'_{r_1 r_2}$ گشتاور مرتبه r_1ام برحسب X و گشتاور مرتبه r_2ام برحسب Y بصورت زیر تعریف می‌شود.

$$\mu'_{r_1 r_2} = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{m} x_i^{r_1} y_j^{r_2} f_{ij}$$

گشتاورهای دوبعدی

✓اگر $r_2 = 0, r_1 = r$ خواهیم داشت.

$$\mu'_{r,0} = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m} x_i^r f_{ij} = \frac{1}{N}\sum_{i=1}^{n} x_i^r \sum_{j=1}^{m} f_{ij} = \frac{1}{N}\sum_{i=1}^{n} x_i^r f_{io} = \mu'_{r(x)}$$

✓که $\mu'_{r(x)}$ گشتارو مرتبه rام حول مبدا x است و اگر r=1 باشد.

$$\mu'_{1,0} = \mu'_{1(x)} = \mu_x = \bar{X}$$

گشتاورهای دوبعدی

✓اگر $r_1 = 0, r_2 = r$ خواهیم داشت.

$$\mu'_{0,r} = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m} y_j^r f_{ij} = \frac{1}{N}\sum_{j=1}^{m} y_j^r f_{0j} = \mu'_{r(y)}$$

✓که $\mu'_{r(y)}$ گشتارو مرتبه rام حول مبدا y است و اگر r=1 باشد.

$$\mu'_{0,1} = \mu'_{1(y)} = \mu_y = \bar{Y}$$

گشتاورهای دوبعدی

✓ گشتاور $\mu'_{r_1 r_2}$ مرتبه r_1ام برحسب x و گشتاور مرتبه r_2ام برحسب y حول دو مقدار a,b بصورت زیر تعریف می‌شود.

$$\mu'_{r_1 r_2} = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m}(x_i - a)^{r_1}(y_j - b)^{r_2} f_{ij}$$

✓ اگر $b = \bar{Y}, a = \bar{X}$ باشد، گشتاور حول میانگین بدست خواهد آمد.

$$\mu_{r_1 r_2} = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m}(x_i - \bar{X})^{r_1}(y_j - \bar{Y})^{r_2} f_{ij}$$

گشتاورهای دوبعدی

✓ اگر $r_1 = r$, $r_2 = 0$ در گشتاور حول میانگین خواهیم داشت.

$$\mu_{r,0} = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m}(x_i - \bar{X})^r f_{ij} = \frac{1}{N}\sum_{i=1}^{n}(x_i - \bar{X})^r \sum_{j=1}^{m} f_{ij} = \frac{1}{N}\sum_{i=1}^{n}(x_i - \bar{X})^r f_{io} = \mu_{r(x)}$$

✓ در حالتی که $r=2$ وریانس x بدست خواهد آمد.

$$\mu_{2,0} = \frac{1}{N}\sum_{i=1}^{n}(x_i - \bar{X})^2 f_{io} = \mu_{2(x)} = \sigma_x^2$$

گشتاورهای دوبعدی

✓ اگر $r_1 = 0, r_2 = r$ در گشتاور حول میانگین خواهیم داشت.

$$\mu_{0,r} = \frac{1}{N}\sum_{j=1}^{m}(y_j - \bar{Y})^r f_{oj} = \mu_{r(y)}$$

✓ در حالتی که $r=2$ وریانس y بدست خواهد آمد.

$$\mu_{0,2} = \frac{1}{N}\sum_{j=1}^{m}(y_j - \bar{Y})^2 f_{oj} = \mu_{2(y)} = \sigma_y^2$$

کوواریانس (واریانس مشترک)

✓ اگر $\mu'_{r_1 r_2}$ حول میانگین $\bar{X}$ و $\bar{Y}$ و $r_1 = r_2 = 1$ خواهیم داشت.

$$\mu_{1,1} = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m}(x_i - \bar{X})(y_j - \bar{Y})f_{ij}$$

✓ پراکندگی Xها را حول میانگین در کنار پراکندگی Yها حول میانگین خود را نشان می‌دهد. بنابراین نشان دهنده پراکندگی متقابل X,Y است.

$$\mu_{1,1} = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m}(x_i - \bar{X})(y_j - \bar{Y})f_{ij} = \sigma_{XY} = Cov(X,Y)$$

$$\sigma_{XY} = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m}x_i y_j f_{ij} - \bar{X}\bar{Y} = \overline{XY} - \bar{X}\bar{Y}$$

کوواریانس (واریانس مشترک)

- ✓ کوواریانس بیان کننده نحوه وابستگی X,Y است.
  - اگر کوواریانس مثبت باشد، یعنی تغییرات X,Y در یک جهت است.
  - اگر کوواریانس منفی باشد، یعنی تغییرات X,Y در خلاف جهت است.

خواص کوواریانس

✓ خواص کوواریانس

$$Cov(X,X) = \sigma_X^2$$

$$Cov(X,Y) = Cov(Y,X)$$

$$Cov(X,C) = 0$$

$$Cov(X,Y+Z) = Cov(X,Y) + Cov(X,Z)$$

$$Cov(X,\alpha Y) = \alpha Cov(X,Y)$$

$$Cov(X,\alpha Y + \beta Z + \gamma) = \alpha Cov(X,Y) + \beta Cov(X,Z)$$

$$[Cov(X,Y)]^2 \leq \sigma_X^2 \sigma_Y^2$$

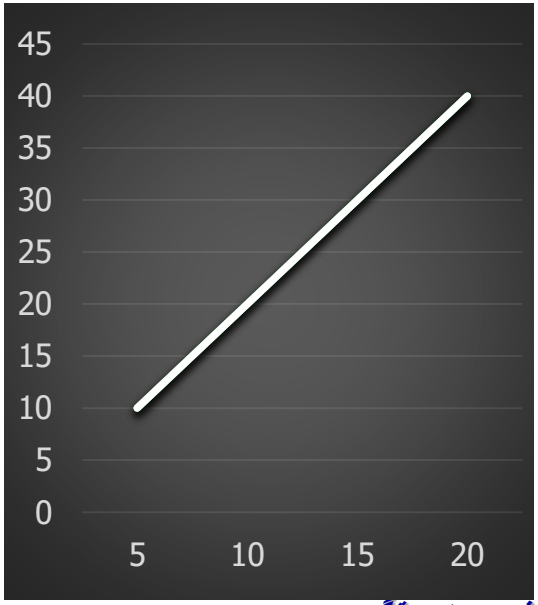
نمودارهای پراکندگی

- ✓ یکی از روش‌هایی که می‌توان به وسیله آن همبستگی بین دو متغیر را نشان داد، نمودار پراکندگی است.

اشکال مختلف نمودار پراکندگی:

- ✓ **الف:** زمانی که با افزایش متغیر X متغیر Y هم افزایش پیدا می‌کند.

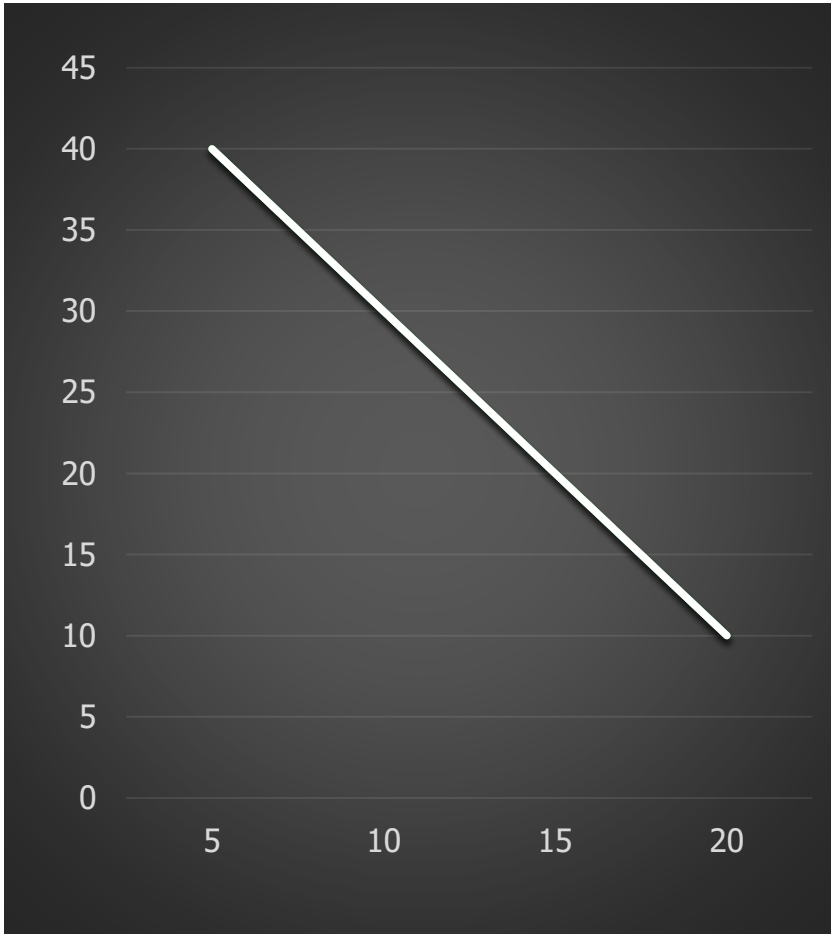
$$r_{xy} = +1$$



نمودارهای پراکندگی

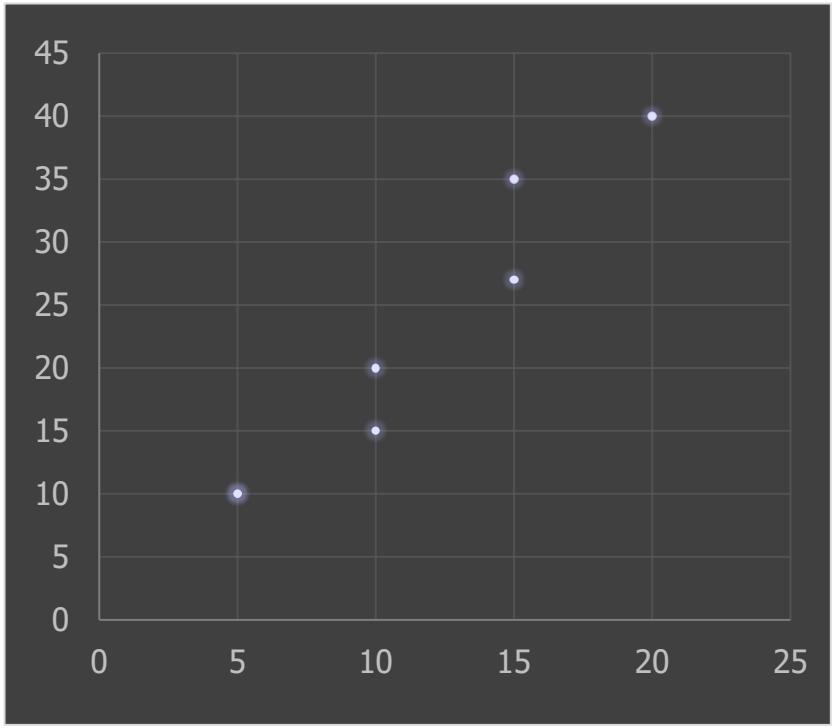
✓ب: اگر با افزایش یک متغیر X متغیر دیگر Y کاهش پیدا کند

$$r_{xy} = -1$$



نمودارهای پراکندگی

✓ ج: زمانی که با افزایش یک متغیر، متغیر دیگر افزایش پیدا کند ولی مقادیر افزایش یکسان نباشد رابطه مستقیم خواهد بود ولی کامل نیست.

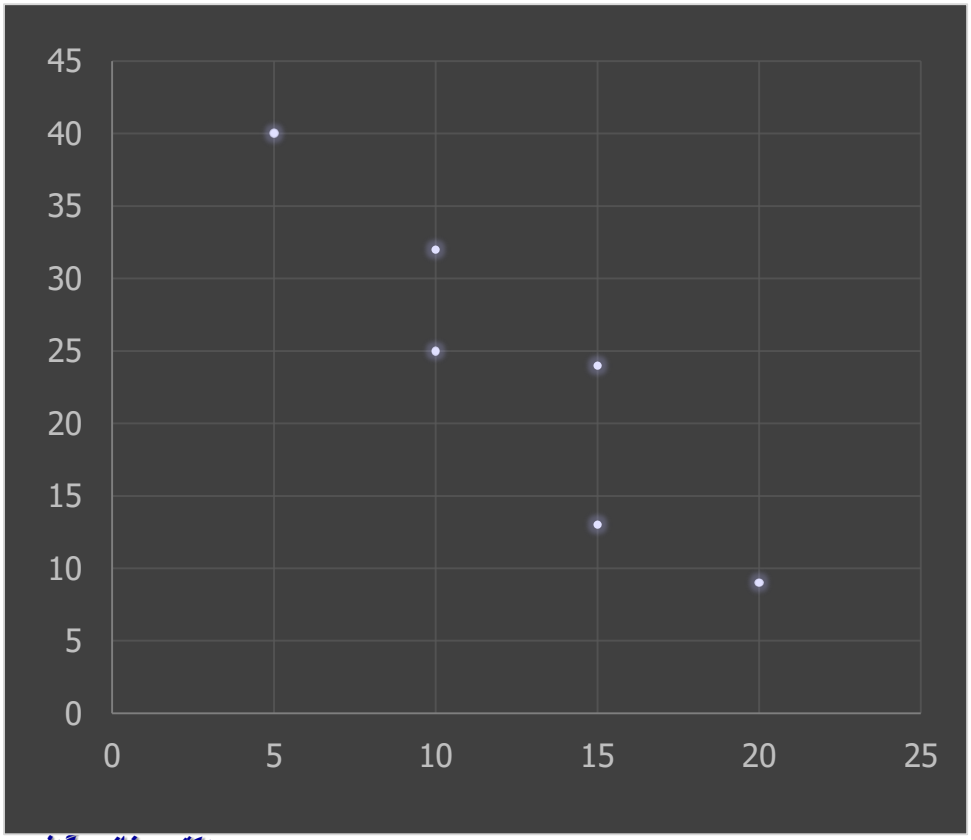


مثلاً

$$r_{xy} = +0.655$$

نمودارهای پراکندگی

✓ د: زمانی که با افزایش یک متغیر مقدار متغیر دیگر کاهش پیدا می‌کند، اما مقادیر یکسان نیست.

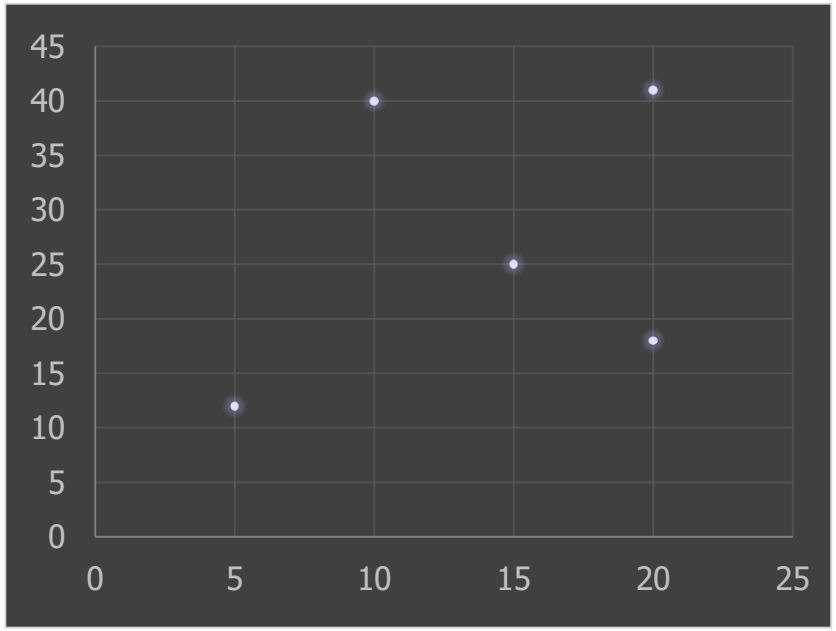


مثلاً

$$r_{xy} = -0.655$$

نمودارهای پراکندگی

✓ هـ: زمانی که رابطه ای بین متغیرها وجود نداشته باشد.



مثلاً

$$r_{xy} = 0.0$$

محاسبه ضریب همبستگی

- ✓ برای محاسبه ضریب همبستگی شاخص‌های متعددی تدوین شده است. معروف‌ترین و در عین حال پر مصرف‌ترین آنها زمانی به کار برده می‌شود که متغیرهای مورد مطالعه با استفاده از مقیاس فاصله‌ای یا نسبی اندازه‌گیری شده باشند.
- ✓ این روش توسط کارل پیرسون تهیه و تنظیم گردیده است و ضریب همبستگی گشتاوری پیرسون نامیده می شود.

محاسبه ضریب همبستگی پیرسون با سه روش

۱ – استفاده از اعداد خام

$$r_{xy} = \frac{N(\Sigma XY) - (\Sigma X)(\Sigma Y)}{\sqrt{\left(N\Sigma X^2 - (\Sigma X)^2\right)\left(N\Sigma Y^2 - (\Sigma Y)^2\right)}}$$

مثال

X	Y	X^2	Y^2	XY
10	15	100	225	150
15	14	225	196	210
5	11	25	121	55
8	13	64	169	104
12	12	144	144	144
50	65	558	855	663

$$r_{xy} = \frac{5(663)-(50)(65)}{\sqrt{\left[5(558)-(50)^2\right]\left[5(855)-(65)^2\right]}} = 0/54$$

محاسبه ضریب همبستگی پیرسون با سه روش

۲- محاسبه ضریب همبستگی از راه انحراف از میانگین :

$$r_{xy} = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma(X-\overline{X})^2 \quad \Sigma(Y-\overline{Y})^2}} = \frac{\operatorname{cov}(X,Y)}{\sigma_x \sigma_y}$$

$$r_{XY} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

مثال:

X	Y	X'	Y'	X'2	Y'2	X'Y'
10	15	0	2	0	4	0
15	14	5	1	25	1	5
5	11	-5	-2	25	4	10
8	13	-2	0	4	0	0
12	12	2	-1	4	1	-2
50	65			58	10	13

$$r_{xy} = \frac{13}{\sqrt{58 \times 10}} = 0/54$$

محاسبه ضریب همبستگی پیرسون با سه روش

۳- محاسبه ضریب همبستگی با استفاده از داده‌های استاندارد شده

$$r_{xy} = \frac{\Sigma Z_x Z_y}{N}$$

$$Z_x = \frac{x - \mu_x}{\sigma_x}$$

$$Z_y = \frac{y - \mu_y}{\sigma_y}$$

مثال:

X	Y	Z_x	Z_y	Z_xZ_y
10	15	0	1/41	0
15	14	1/47	0/71	1/04
5	11	-1/47	-1/41	2/1
8	13	-0/59	0	0
12	12	+0/59	-0/71	-0/42
50	65			2/72

$$r_{xy} = \frac{2/72}{5} = 0/54$$

ضریب همبستگی اسپیرمن

ضریب همبستگی اسپیرمن صورتی از ضریب همبستگی پیرسون است و زمانی به کار برده می‌شود که داده‌ها رتبه‌بندی شده باشند یا به جای اعداد رتبه‌های آنها در دست باشد.

$$P = 1 - \frac{6\Sigma D^2}{N(N^2 - 1)}$$

P = ضریب همبستگی رتبه‌ای اسپیرمن

D^2 = مجذور تفاوت رتبه‌ها

N = تعداد داده‌ها

مراحل محاسبه ضریب همبستگی رتبه‌ای اسپیرمن

۱- اعداد متغیر **X** به ترتیب از بزرگ به کوچک مرتب می کنیم (R_x) و بعد از آن اعداد **Y** مربوط به هرعدد **X** را در ستون **Y** ها قرار می دهیم و به آنها رتبه می‌دهیم (R_y)

۲- R_y را از R_x کم می‌کنیم و اختلاف آنها را **D** می‌نامیم .

۳- **D** ها را به توان دو می‌رسانیم و مجموع آنها را محاسبه می‌کنیم.

۴- مقادیر به دست آمده را در فرمول جایگزین می‌کنیم.

ضریب همبستگی رتبه‌ای اسپیرمن

✓ نکته:

بعضی مواقع یک عدد چند بار تکرار شده است در این مواقع رتبه‌های اعداد را با هم جمع کرده و بر تعداد آنها تقسیم می کنیم. نتیجه به دست آمده رتبه اعداد تکرار شده می‌باشد.

ضریب همبستگی اسپیرمن داده‌های زیر را محاسبه کنید.

X	46	45	47	45	43	41	45
Y	51	32	38	39	50	48	41

X	Y	R_x	R_y	D	D^2
47	38	1	6	-5	25
46	51	2	1	1	1
45	32	4	7	-3	9
45	39	4	5	-1	1
45	41	4	4	0	0
43	50	6	2	4	16
41	48	7	3	4	16
					68

$$P = 1 - \frac{6(68)}{7(7^2 - 1)} = -0/21$$

عواملی که بر ضریب همبستگی تأثیر می گذارند:

۱- اساس رابطه از جامعه‌ای به جامعه دیگر فرق می‌کند.

مثلاً در افراد بشر در سنین ۱۶-۱۰ سالگی بین سن تقویمی و توانایی فیزیولوژکی همبستگی بالایی وجود دارد. ولی بین این دو متغیر در سنین ۲۶-۲۰ سالگی همبستگی وجود ندارد.

عواملی که بر ضریب همبستگی تأثیر می گذارند:

۲- پراکندگی متغیرها در جوامع مختلف متفاوت است بدین معنی که هر چه تجانس بیشتر باشد همبستگی کمتر است. به عنوان مثال، در یک دانشکده (جامعه) بین قد و موفقیت در بازی بسکتبال همبستگی مثبت و بالای وجود دارد. اما در تیم بسکتبال یک کشور چنین رابطه ای وجود ندارد.

عواملی که بر ضریب همبستگی تأثیر می گذارند:

۳- همبستگی بین دو متغیر تحت تأثیر همبستگی آنها با متغیر ثالثی قرار دارد به عنوان مثال، همبستگی بین فیزیک و ریاضی ممکن است به دلیل همبستگی این متغیرها با هوش باشد.

نکته: تفسیر ضریب همبستگی نباید بر حسب درصد و نسبت باشد. مثلاً r_{xy} =۰/۷۰ هفتاد درصد رابطه بین متغیرها را تبیین نمی‌کند و r_{xy} =۰/۹۰ دقیقاً دو برابر r_{xy} =۰/۴۵ نیست.

ضریب تعیین

ضریب تعیین نشان دهنده میزان تأثیری است که متغیر **X** (مستقل) در متغیر **Y** (وابسته) ایجاد می‌کند یا به عبارتی با محاسبه این ضریب می‌توان تعیین کرد چند درصد از کل واریانس **X** ناشی از واریانس **Y** است.

$$V = (r_{xy})^2 (100)$$

نکته: ضریب تعیین هیچ وقت منفی نخواهد شد زیرا برای محاسبه آن ضریب همبستگی مجذور می‌شود.

پیش بینی

- ✓ زمانی که بین دو متغیر همبستگی وجود داشته باشد می توان از طریق رگرسیون مقدار یک متغیر Y را از روی یک متغیر دیگر X پیش بینی یا برآورد کرد، و هر چه همبستگی بین متغیرها بالاتر باشد، به همان اندازه پیش بینی دقیق‌تر است.

نکته: رابطه بین متغیر پیش بینی شونده Y و متغیر پیش بینی کننده X تابع علامت و شدت ضریب همبستگی است.

پیش بینی داده‌های استاندارد (Z)

مقدماتی ترین روشی که در استفاده ازضریب همبستگی پیرسون برای پیش بینی به کار برده می شود داده‌های استاندارد است.

$$Z_x = \frac{x - \mu_x}{\sigma_x}$$

$$Z_y = (Z_x)(r_{XY})$$

رگرسیون

زمانی که همبستگی بین دو متغیر پایین باشد، داده‌های پیش بینی شده نزدیک به میانگین داده پیش‌بینی شونده هستند تا داده واقعی، به این پدیده رگرسیون می‌گویند.

نکته: میزان همبستگی بین دو متغیر حدود یا مقدار اتفاق رگرسیون را تعیین می‌کند .

نکته: پدیده رگرسیون اولین بار بوسیله گالتن مورد استفاده قرار گرفت. بر اساس مطالعات گالتن، فرزندان والدین بلند قد، بلند قد هستند. اما نه به اندازه والدین خود. به همین ترتیب فرزندان والدین کوتاه قد، کوتاه قد هستند اما نه به کوتاهی والدین خود.

خط رگرسیون

خطی که در رابطه با داده‌های پیش بینی شده است **خط رگرسیون** نامیده می‌شود. به عبارتی خطی است که خطاهای پیش بینی را به حداقل می‌رساند (خط حداقل مجذورها).

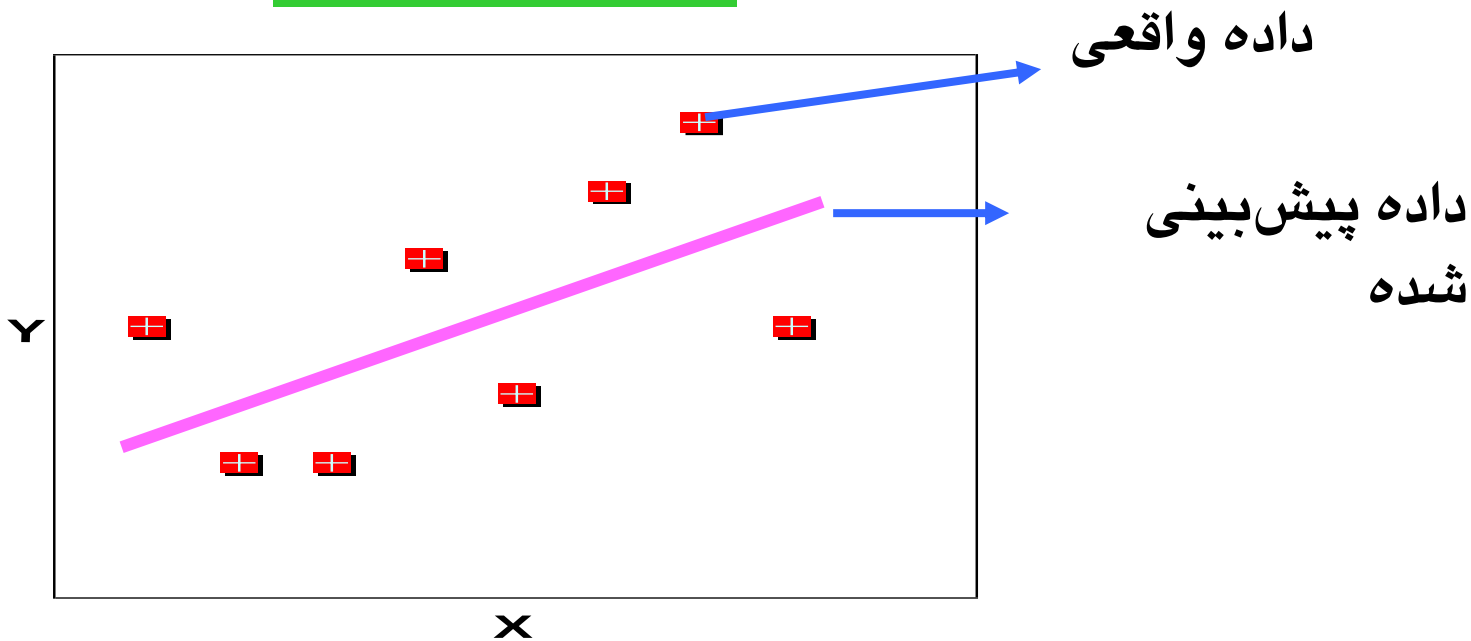
- ✓ یعنی اینکه مجموع مجذور فاصله **Y** ها از خط رگرسیون کوچکتر از فاصله هر خط دیگری تا محور **Y** ها می‌باشد.

خط رگرسیون

نکته : خط رگرسیون را خط برازنده نیز می‌نامند.

نکته: اختلاف بین داده واقعی Y و داده پیش‌بینی شده Y' را خطای پیش بینی e می‌گویند.

$$e = y - y'$$



معادله خط رگرسیون

برای پیش بینی یک متغیر از روی متغیر دیگر از معادله خط رگرسیون استفاده می‌شود.

فرمول معادله خط :

$$y' = a + bx$$

b = شیب خط

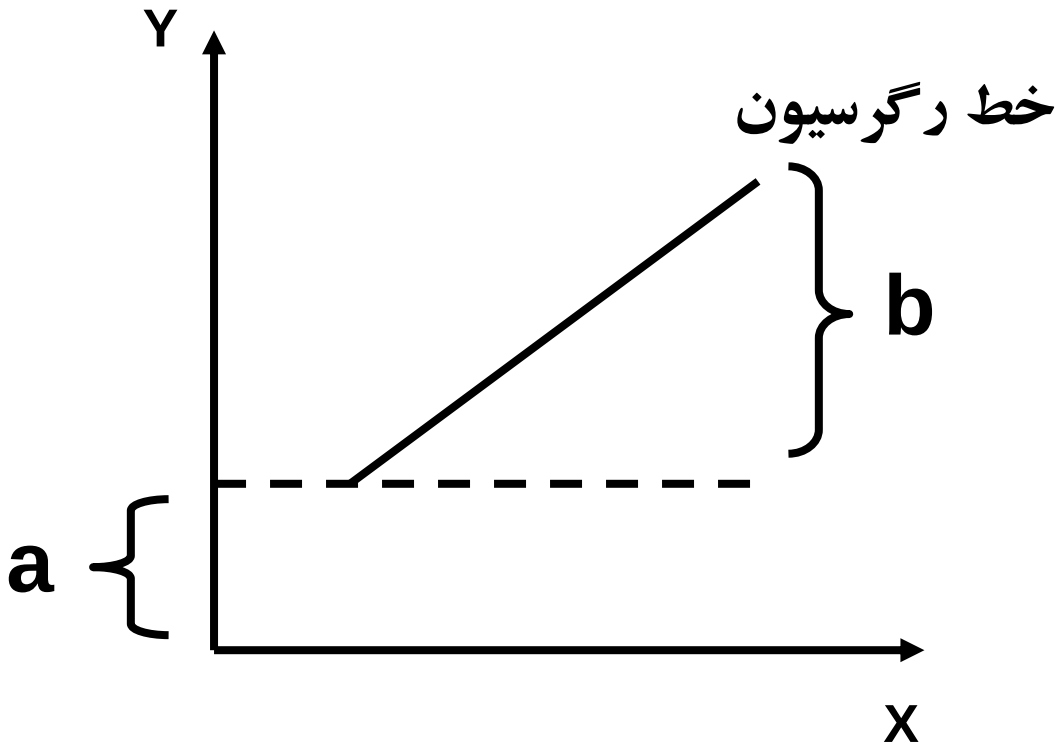
y` = متغیری که می خواهیم پیش بینی کنیم

a = عرض از مبدأ یا محل تلاقی خط رگرسیون با محورایگرگ

X = متغیری که مقدار آن را در اختیار داریم

معادله خط رگرسیون

✓ نکته: بعضی مواقع مقدار a و b در اختیار است در این صورت با جایگزینی مقادیر محاسبه خیلی ساده می‌باشد.



معادله خط رگرسیون

✓ نکته : بعضی مواقع مقدار a و b در اختیار است در این صورت با جایگزینی مقادیر محاسبه خیلی ساده می باشد.

مثال:

X = 0	Y = 4 + 2 (0) = 4
X = 1	Y = 4 + 2 (1) = 6
X = 4	Y = 4 + 2 (4) = 12
X = 8	Y = 4 + 2 (8) = 20

روش محاسبه ضرایب a و b

✓ a و b را به دو روش می توان محاسبه کرد:

الف: زمانی که داده‌های خام در اختیار نداشته باشیم و یک سری اطلاعات به ما داده باشند.

ب: زمانی که اطلاعات و داده‌های خام در اختیار داشته باشیم .

محاسبه a و b از روی اطلاعات ثانوی

$$b_{yx} = r_{xy} \frac{\sigma_Y}{\sigma_X}$$

$$b_{xy} = r_{xy} \frac{\sigma_X}{\sigma_Y}$$

b_{yx} و b_{xy} = شیب خط رگرسیون

r_{xy} = ضریب همبستگی

σ_y = y انحراف استاندارد

σ_x = x انحراف استاندارد

محاسبه a و b از روی اطلاعات ثانوی

$$a_{xy} = \bar{x} - b_{xy}(\bar{y})$$

$$a_{yx} = \bar{y} - b_{yx}(\bar{x})$$

$$a_{yx}, a_{xy} =$$ عرض از مبدأ

$$\bar{x}, \bar{y} =$$ میانگین

$$b_{xy}, b_{yx} =$$ شیب خط

محاسبه a و b از روی داده‌های خام

$$b_{yx} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N\Sigma X^2 - (\Sigma X)^2} = \frac{\Sigma XY - N\overline{X}\,\overline{Y}}{\Sigma X^2 - N\overline{X}^2}$$

$$a_{yx} = \frac{\Sigma Y - b_{yx}\Sigma X}{N} = \overline{Y} - b_{yx}\overline{X}$$

مثال: برای داده‌های زیر معادله خط رگرسیون را به دست آورید؟

X	Y	X^2	Y^2	XY
6	8	36	64	48
9	9	81	81	81
5	7	25	49	35
20	24	142	194	164

$$b_{yx} = \frac{3(164) - (20)(24)}{3(142) - (20)^2} = 0/46$$

$$a_{yx} = \frac{24 - (0/46)(20)}{3} = 4.93$$

$$y = 4.93 + 0/46(X)$$

خطای استاندار برآورد

اکثر مواقع بین داده‌های پیش‌بینی شده و داده‌های مشاهده تفاوت وجود دارد، به این اختلاف **خطای استاندارد برآورد** می‌گویند.

نکته: هر چه همبستگی بین متغیرها بیشتر باشد، خطای پیش‌بینی کمتر خواهد بود. این خطا به دو روش محاسبه می شود.

خطای استاندار برآورد

$$\sigma_{yx} = \sigma_y \sqrt{1 - r^2_{xy}}$$

$$\sigma_{yx} = \sqrt{\frac{\Sigma e^2}{N}}$$

σ_{yx} = خطای استاندارد پیش بینی

σ_y = y انحراف استاندارد

σ_x = x انحراف استاندارد

r_{xy} = ضریب همبستگی

e = خطای پیش بینی

محاسبه زاویه بین دو خط رگرسیون

- ✓ شیب‌های خطوط رگرسیون Y روی X و X روی Y به ترتیب عبارتند از:

$$\frac{1}{r} \times \frac{\sigma_x}{\sigma_y} \qquad\qquad r\frac{\sigma_y}{\sigma_x}$$

- ✓ در نتیجه زاویه بین دو خط رگرسیون یعنی θ برابر است با:

$$\tan(\theta) = \frac{\frac{1}{r} \times \frac{\sigma_x}{\sigma_y} - r\frac{\sigma_y}{\sigma_x}}{1 + \frac{1}{r} \times \frac{\sigma_x}{\sigma_y} \times r\frac{\sigma_y}{\sigma_x}} = \frac{1-r^2}{r} \times \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}$$

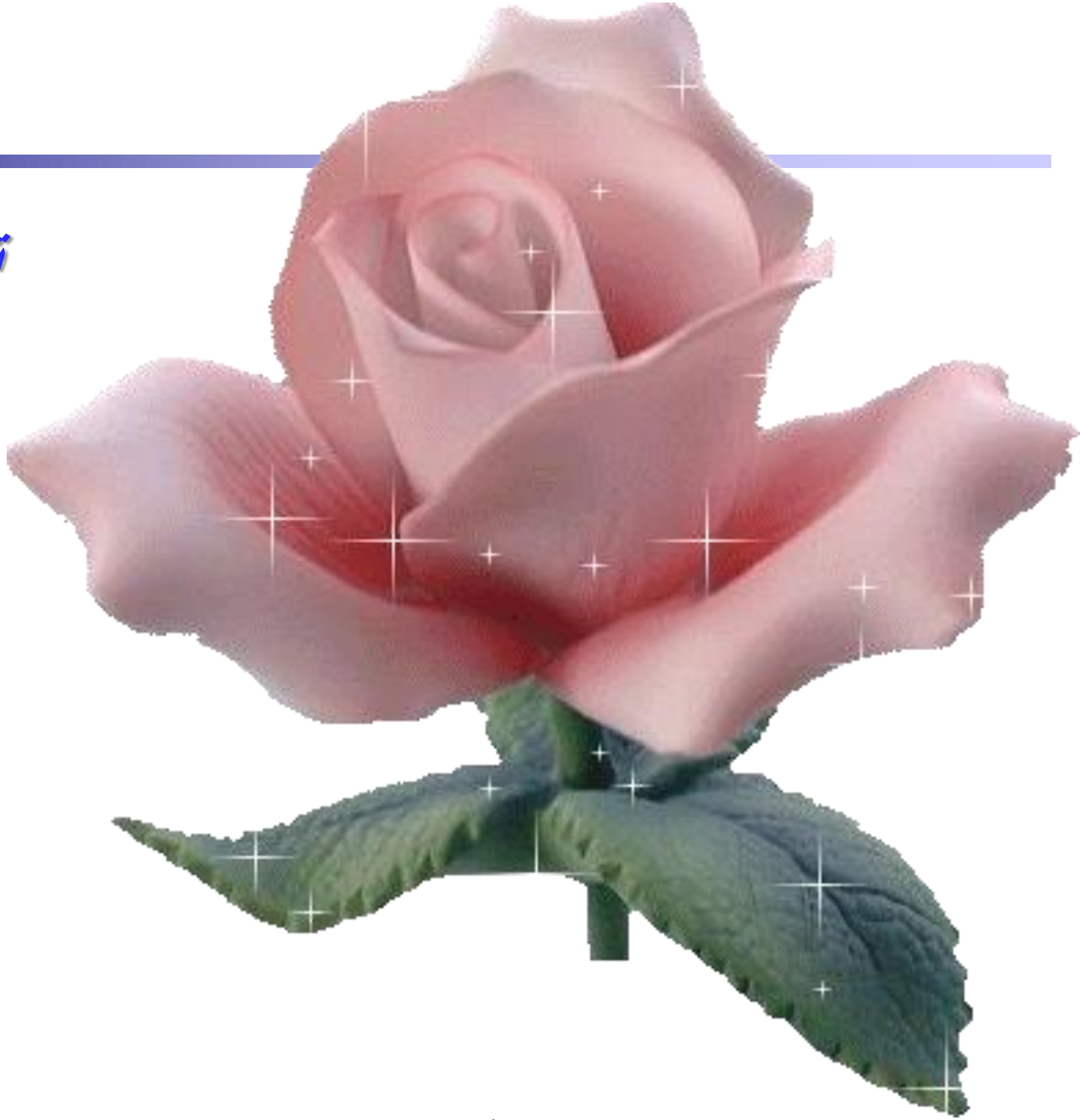
محاسبه زاویه بین دو خط رگرسیون

- ✓ دو حالت ممکن است اتفاق بیافتد:
  - ✓ همبستگی کامل $r = \pm 1$، در این حالت زاویه بین دو خط صفر است، همچنین چون دو خط از میانگین‌ها رد می‌شوند، پس بر هم منطبق‌اند.
  - ✓ همبستگی وجود ندارد $r = 0$، در این حالت زاویه بین دو خط ۹۰ است و در نتیجه دور خط بر هم عمودند.

$$\tan(\theta) = \frac{\frac{1}{r} \times \frac{\sigma_x}{\sigma_y} - r\frac{\sigma_y}{\sigma_x}}{1 + \frac{1}{r} \times \frac{\sigma_x}{\sigma_y} \times r\frac{\sigma_y}{\sigma_x}} = \frac{1-r^2}{r} \times \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}$$

ضریب همبستگی داده‌های گروه‌بندی شده

- ✓ یک جدول فراونی دو طرفه متشکل از ستون‌ها (X) و سطرها (Y) دانست که هرکدام فراوانی دارند، در اینگونه از جداول از رابطه زیر به محاسبه ضریب همبستگی می‌پردازیم.

$$r = \frac{\sum fxy - \frac{\sum fx \times \sum fy}{N}}{\left[\sum fx^2 - \frac{(\sum fx)^2}{N}\right]^{1/2} \times \left[\sum fy^2 - \frac{(\sum fy)^2}{N}\right]^{1/2}}$$

تشکر از توجه شما