

فصل دهم:

همبستگی و رگرسیون

مطالب این فصل:

- کوواریانس
- ضریب همبستگی
- رگرسیون
- ضریب تعیین یا ضریب تشخیص (r^2)
- خطای معیار برآورد (S_{XY})
- انواع ضرایب همبستگی برای بررسی رابطه بین متغیرهای کمی و کیفی

- در بسیاری از بررسی‌های آماری به جای یک متغیر با دو متغیر سر و کار داریم و مطالعه مشترک آن دو متغیر مد نظر است.

- همبستگی و نمودار پراکنش (نمودار پراکنندگی) به عنوان ابزارهایی برای این که بفهمیم آیا ارتباطی خطی بین دو متغیر وجود دارد یا خیر، به کار می‌روند.

کواریانس

- کواریانس شاخصی است که شدت وابستگی خطی بین دو متغیر را نشان می‌دهد.

- کواریانس عبارت است از توزیع های مشترک

$$\text{Cov}(x,y) = S_{x,y} = \frac{SP}{n-1} = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{n-1} = \frac{\sum xy}{n-1}$$

• نکات

۱- اگر X و Y دو متغیر تصادفی و a, b, c, d مقادیر ثابت باشند، داریم:

$$\text{Cov}(aX+b, cY+d) = ac \text{Cov}(X, Y)$$

مثال) اگر $\text{Cov}(X, Y) = 4$ ، آن گاه $\text{Cov}(2+3X, 4+5Y)$ را معین کنید.

$$\text{Cov}(2+3X, 4+5Y) = 3 \times 5 \text{Cov}(X, Y) = 3 \times 5 \times 4 = 60$$

۲- اگر دو متغیر از هم مستقل باشند، کواریانس آن‌ها برابر صفر خواهد بود. یعنی دو متغیر هیچ وابستگی خطی با هم ندارند. ولی نمی‌توان عنوان کرد که چنانچه کواریانس دو متغیر صفر باشد، آن دو متغیر مستقل از یکدیگرند. ممکن است دو متغیر به صورت غیر خطی با یکدیگر وابسته و کواریانس بی آن‌ها صفر باشد.

۳- مقدار کواریانس تابع واحد اندازه گیری است.

۴- مقدار کواریانس به تغییر پذیری نمره های متغیرهای X و Y بستگی دارد. مقدار کواریانس تابع انحراف معیارهای متغیرهای X و Y است در نتیجه نمی‌توان کواریانس های حاصل از اندازه های مختلف را با هم مقایسه کرد.

ضریب همبستگی

- همانطور که عنوان شد، مقدار کواریانس به واحدهای اندازه گیری X و Y بستگی دارد در نتیجه تغییر در یکی از متغیرهای X و Y در محاسبه $Cov(X, Y)$ دخالت می کند.
- معیاری که تحت تأثیر واحدهای اندازه گیری X و Y نباشد، ضریب همبستگی است که به منظور تعیین شدت همبستگی بین دو متغیر استفاده می شود.
- برای اخذ تصمیم درباره اینکه آیا رابطه آماری معنادار بین دو متغیر وجود دارد یا خیر از ضریب همبستگی استفاده می کنیم.
- ضریب همبستگی، میزان تغییر در متغیر Y را به ازای هر واحد تغییر در متغیر X نشان می دهد.

• مفروضه های ضریب همبستگی پیرسون (بررسی رابطه بین دو متغیر فاصله ای)

- مقیاس فاصله ای یا نسبی
- رابطه خطی بین دو متغیر
- مشابه بودن اشکال توزیع
- یکسان بودن نمودار پراکندگی (یکسانی نقاط پراکندگی (واریانس))

• فرمول های محاسبه ضریب همبستگی پیرسون :

$$r_{X,Y} = \frac{Cov(X,Y)}{S_X S_Y}$$

$$r_{X,Y} = \frac{n \sum XY - \sum X \sum Y}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}}$$

$$r_{X,Y} = \frac{\sum XY - n\bar{X}\bar{Y}}{\sqrt{[\sum X^2 - n\bar{X}^2][\sum Y^2 - n\bar{Y}^2]}}$$

$$r_{X,Y} = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{(\sum(X - \bar{X})^2 \sum(Y - \bar{Y})^2)}}$$

نکته: بسته به اینکه صورت سؤال چه اطلاعاتی به شما می دهد می توانید از یکی از فرمول های بالا استفاده کنید. در محاسبه دقت نمائید برای مثال در محاسبه $\sum X^2$ باید هر عدد به توان دو رسیده و سپس مجموع اعدادی که به توان رسیده اند محاسبه شود ولی در محاسبه $(\sum X)^2$ باید ابتدا اعداد با هم جمع شده و مجموع اعداد به توان برسد.

- محاسبه ضریب همبستگی با استفاده از نمره استاندارد:

$$r_{X,Y} = \frac{\sum Z_x Z_y}{N} \quad Z_x = r_{X,Y} \times Z_y \quad Z_y = r_{X,Y} \times Z_x$$

مثال) از یک جامعه دو بعدی، پس از محاسبات لازم کمیت های زیر به دست آمده است:
 $n=5, \quad \sum X_i = 125, \quad \sum X_i^2 = 650, \quad \sum Y_i = 100, \quad \sum Y_i^2 = 436, \quad \sum X_i Y_i = 540$

ضریب همبستگی خطی بین دو صفت X و Y چقدر است؟

با توجه به اطلاعات داده شده در صورت سؤال بهترین فرمول دومی است:

$$r_{X,Y} = \frac{n \sum XY - \sum X \sum Y}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}} = \frac{(5)(540) - (125)(100)}{\sqrt{[(5)(650) - (125)^2][(5)(436) - (100)^2]}} = \frac{-9800}{9827/3}$$

$$r_{X,Y} = -0.99$$

• ضریب همبستگی رتبه ای اسپیرمن

ضریب همبستگی پیرسون، برای توصیف همبستگی بین دو متغیر که با استفاده از مقیاس فاصله ای یا نسبی اندازه گیری شده باشند، به کار برده می شود. در علوم انسانی متغیرهایی وجود دارد که نمی توان آن ها را با مقیاس فاصله ای یا نسبی اندازه گیری کرد. ضریب همبستگی اسپیرمن صورتی از ضریب همبستگی پیرسون است و زمانی به کار برده می شود که نمره ها رتبه بندی شده باشند یا به جای اعداد، رتبه های آن ها در دست باشد. برای محاسبه ضریب همبستگی اسپیرمن از فرمول زیر استفاده می شود:

$$r_s = 1 - \frac{6 \sum D^2}{n(n^2 - 1)}$$

$$D^2 = \text{مجدور تفاوت رتبه}$$

$$n = \text{تعداد آزمودنی ها یا رتبه ها}$$

نکته: برای تعیین رتبه های نمراتی که مساوی هستند، میانگین رتبه ها محاسبه می شود.

مثال) نمرات دروس آمار (X) و رشد (Y) ۵ نفر از دانشجویان درجدول زیر ارائه شده است. ضریب همبستگی اسپیرمن کدام است؟

دانشجویان	آمار (X)	رشد (Y)
۱	۱۸	۱۹
۲	۱۷	۱۸
۳	۱۵	۱۲
۴	۱۴	۱۲
۵	۱۳	۱۰

برای محاسبه ضریب اسپیرمن باید نمره ها را رتبه بندی کرده و مجذور تفاوت رتبه را محاسبه کنیم.

دانشجویان	آمار (X)	رشد (Y)	R_x (رتبه بندی نمرات آمار)	(رتبه بندی نمرات رشد)	D=	
۱	۱۸	۱۹	۱	۱	۰	۰
۲	۱۷	۱۸	۲	۲	۰	۰
۳	۱۵	۱۲	۳	۳/۵	۰/۵	۰/۲۵
۴	۱۴	۱۲	۴	۳/۵	۰/۵	۰/۲۵
۵	۱۳	۱۰	۵	۵	۰	۰
n=۵						

نکته: دو نمره ۳/۵ که در رتبه بندی نمرات رشد می بینید، به دلیل این است که دو نمره ۱۲ در نمره ها وجود دارد که رتبه های ۳ و ۴ متعلق به آنهاست، در نتیجه میانگین رتبه های ۳ و ۴ که ۳/۵ می شود را به هر دو نمره می دهیم.

حال اعداد را در فرمول جایگذاری می کنیم:

$$r_s = 1 - \frac{6 \sum D^2}{n(n^2 - 1)} = r_s = 1 - \frac{6 \times 0.5}{5(25 - 1)} = 0.24$$

• ویژگی های ضریب همبستگی

$$(1) \quad -1 \leq r \leq 1$$

(۲) ضریب همبستگی ماهیتاً قرینه است: $\Gamma_{xy} = \Gamma_{yx}$

(۳) ضریب همبستگی مستقل از مبدأ و مقیاس اندازه گیری است.

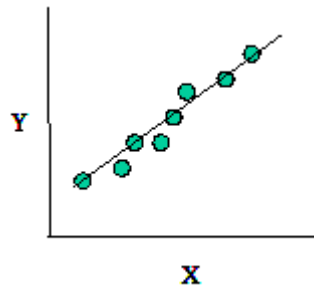
(۴) اگر X و Y مستقل باشند، ضریب همبستگی برابر صفر است اما اگر ضریب همبستگی برابر صفر باشد الزاماً دو متغیر مستقل نیستند.

(۵) ضریب همبستگی معیاری است جهت همبستگی خطی یا وابستگی خطی و برای توصیف ارتباطات غیرخطی به کار نمی رود.

(۶) ضریب همبستگی الزاماً بیانگر هیچ گونه رابطه علت و معلولی نیست.

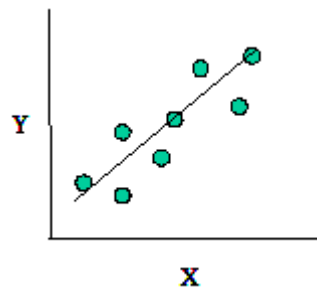
• انواع همبستگی

۱- همبستگی کامل و مثبت (مستقیم) (شیب مثبت)



$$r=1$$

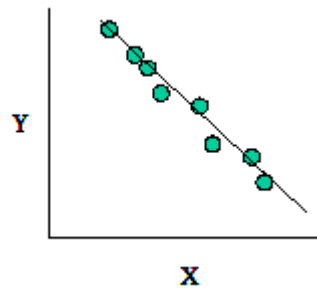
۲- همبستگی ناقص و مثبت (مستقیم)



$$0 < r < 1$$

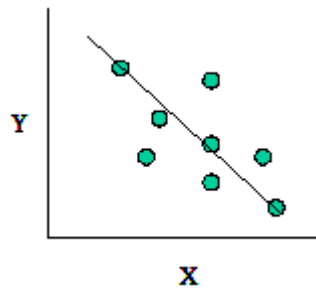
۳- همبستگی کامل و معکوس (شیب منفی)

$$r = -1$$



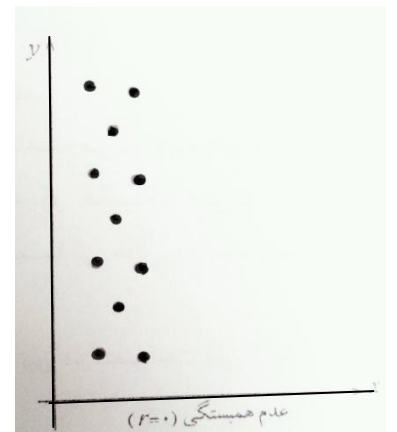
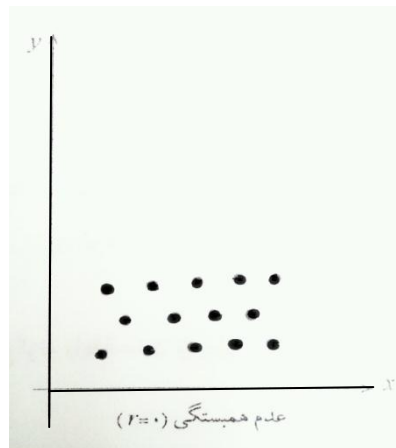
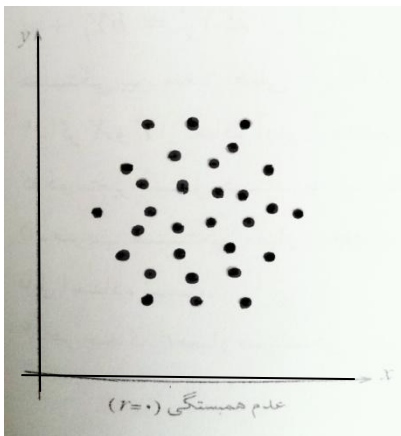
۴- همبستگی ناقص و مثبت

$$-1 < r < 0$$



۵- عدم همبستگی

در شرایط زیر همبستگی بین دو متغیر صفر است:



• نکاتی در مورد ضریب همبستگی

- وقتی از ضریب همبستگی صحبت می کنیم، نمی توان گفت که مثبت بهتر است یا منفی. به طور کلی هر چه قدر عدد به دست آمده صرفنظر از علامت آن، بزرگتر باشد، همبستگی بیشتر است و علامت مثبت و منفی فقط بر جهت همبستگی دلالت دارد.

- اگر $r = 0/3$ یا $r = 0/9$ باشد، این فقط بدان معنا است که ضریب همبستگی در دو حالت مثبت است و نمی توان گفت که $r = 0/9$ سه برابر همبستگی نسبت به $r = 0/3$ است.

- اگر همبستگی بین دو متغیر صفر باشد، این بدین معنی است که این دو متغیر با هم رابطه خطی ندارند. اما ممکن است به طور غیر خطی به یکدیگر مربوط باشند.

- اگر $Y = X + 1$ باشد، ضریب همبستگی برابر ۱ خواهد بود.

- ضریب ایستا خطی بودن یا نبودن رابطه بین دو متغیر را نشان می دهد.

- همبستگی در جامعه ای که از نظر متغیر مورد نظر ناهمگن باشد، بیشتر است.

- ضرب کردن متغیرها در اعداد مثبت تأثیری در ضریب همبستگی آنان ندارد.

- اگر تمام اعداد یا نمره ها با هم برابر باشند یعنی پراکندگی صفر است و ضریب همبستگی نیز صفر است.

- استفاده از گروه های انتهایی ضریب همبستگی را افزایش و محدودیت در دامنه ضریب همبستگی را کاهش می دهد.

- ضریب همبستگی جزئی یا تفکیکی: اثر متغیر سوم را از دو متغیر قبلی بر می داریم و رابطه اصلی بین دو متغیر را ضریب جزئی یا تفکیکی می گویند.

- ضریب همبستگی نیمه جزئی یا نیمه تفکیکی: اثر متغیر سوم را فقط از متغیر ملاک بر می داریم.

• آزمون معنی داری ضریب همبستگی

- می دانیم که r ضریب همبستگی خطی است که شدت همبستگی بین X و Y در نمونه را اندازه گیری می کند. پس r یک آماره نمونه است، اما ρ ضریب همبستگی خطی است که شدت همبستگی بین X و

- Y در جامعه را اندازه گیری می کند. پس ρ یک پارامتر جامعه است. در نتیجه ضریب همبستگی محاسبه شده از نمونه (r) بر آوردی از ضریب همبستگی جامعه (ρ) خواهد بود.
- معنادار بودن ضریب همبستگی بیش از همه به تعداد آزمودنی ها بستگی دارد.
 - گاهی ممکن است ضریب همبستگی دو متغیر در جامعه برابر صفر ولی در نمونه غیر صفر باشد برای روشن شدن این موضوع باید آزمون فرض را انجام دهیم:

(۱) نوشتن فرض صفر و فرض خلاف

در این حالت فرضیه های صفر و مقابل به صورت زیر خواهد بود

$$\begin{cases} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{cases}$$

(۲) انتخاب و محاسبه آماره آزمون

آماره مناسب برای بررسی معنادار بودن ضریب همبستگی آماره t است که از فرمول زیر محاسبه می شود:

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

درجه آزادی توزیع t برای معناداری ضریب همبستگی $n-2$ است.

(۳) نوشتن قاعده رد (مشخص کردن ناحیه بحرانی) و تصمیم گیری

$$|t| > t_{\frac{\alpha}{2}}$$

نکته: مقدار t را از طریق فرمول محاسبه می کنیم بحرانی را برای درجه آزادی $n-2$ باید از جداول انتهای کتاب استخراج کنیم.

رگرسیون

در تحلیل همبستگی هدف اولیه اندازه گیری میزان همبستگی خطی بین دو متغیر است اما اصولاً در تحلیل رگرسیون به دنبال چنین اندازه گیری نیستیم بلکه سعی داریم مقدار متوسط یک متغیر را بر اساس مقادیر ثابت متغیرهای دیگر تخمین بزنیم یا پیش بینی کنیم.

ما باید فقط وقتی از معادله خط رگرسیون استفاده کنیم که r یک همبستگی خطی معنادار را نشان می دهد. به طور کلی اگر r نزدیک به ۱ یا -۱ باشد، آنگاه خط رگرسیون بر داده ها برازنده است اما اگر r نزدیک صفر باشد، برازندگی خط رگرسیون کم خواهد بود.

• معادله رگرسیون

$$\hat{y} = \alpha + \beta x$$

در این فرمول α ، عرض از مبدا و ضریب پایه است و β شیب خط رگرسیون و ضریب زاویه (ضریب رگرسیون) است. به ازای هر واحد تغییر در متغیر x ، متغیر y به اندازه β تغییر می کند.

• نحوه پیدا کردن ضریب رگرسیون (β)

$$\beta = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2}$$

$$\beta = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum X^2 - n\bar{X}^2}$$

$$\beta = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2} = \frac{SPD_{X,Y}}{SSD_X}$$

$$\beta = \frac{Cov_{xy}}{S_x^2}$$

- رابطه ضریب همبستگی و ضریب رگرسیون

$$r = \beta \frac{S_x}{S_y} \quad \beta = r \frac{S_y}{S_x}$$

- در برخی از کتاب ها از علائم اختصاری زیر استفاده می شود:

$$S_{XX} = SSD_X = SS_X = \sum (X - \bar{X})^2$$

• نحوه پیدا کردن ضریب پایه یا عرض از مبدأ (α)

$$\alpha = \bar{y} - \beta \bar{x} \quad \bar{y} = \frac{\sum y}{n} \quad \bar{x} = \frac{\sum x}{n}$$

• نکاتی در مورد رگرسیون:

- در رگرسیون، اگر علامت β مثبت باشد، r مثبت و اگر علامت β منفی باشد، r منفی می باشد.
- خط رگرسیون، خط حداقل مجذورهاست: خطی که خطاهای پیش بینی را به حداقل می رساند.
- اگر $r=0$ باشد، زاویه خط های رگرسیون 90° درجه است.
- اگر $r=1$ باشد، زاویه بین خط های رگرسیون صفر است خطها بر هم منطبقند.
- دقت رگرسیون تابع ضریب همبستگی بین متغیرهاست. هر چه ضریب همبستگی بین متغیرها بیشتر باشد پیش بینی متغیر ملاک از روی پیش بینی قوی تر است.

مثال) اطلاعات زیر در دست است. شیب خط رگرسیون $y = \alpha + \beta x$ کدام است؟

$$\sum x = 13, \quad \sum y = 27, \quad \sum xy = 108, \quad \sum x^2 = 51$$

پاسخ:

$$\beta = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2} = \frac{5 \times 108 - 13 \times 27}{5 \times 51 - (13)^2} = 2/2$$

مثال) در مثال قبل، عرض از مبدأ خط رگرسیون و معادله رگرسیون را به دست آورید <

\bar{y}

$$\alpha = \bar{y} - \beta \bar{x} \qquad \alpha = \frac{27}{5} - \left(\frac{2}{2} \times \frac{13}{5} \right) \qquad \alpha = \frac{1}{5}$$

$$y = \alpha + \beta x \qquad y = \frac{1}{5} + \frac{2}{2}x$$

مثال) رابطه بین X و Y خطی است و داده های نمونه به شرح زیر است؟

$$\sum xy = 1150, \quad \sum x^2 = 550, \quad \bar{x} = 5, \quad \bar{y} = 10, \quad n = 20$$

مقدار پیش بینی شده برای Y به ازای $x=6$ کدام است؟

پاسخ: ابتدا باید معادله خط رگرسیون را پیدا کنیم:

$$\beta = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum X^2 - n\bar{X}^2} = \frac{1150 - 20 \times 5 \times 10}{550 - 20 \times 5^2} = 3$$

$$\alpha = \bar{y} - \beta \bar{x} = 10 - 3 \times 5 = -5$$

$$y = \alpha + \beta x \qquad y = -5 + 3x$$

$$y = -5 + 3 \times 6 = 13$$

• پدیده رگرسیون (بازگشت آماری)

- چنانچه همبستگی بین متغیرها در پیش بینی کامل نباشد، رگرسیون یا بازگشت به سمت میانگین رخ می دهد.

- بنابراین تا زمانی که دو متغیر، به صورت کامل همبسته نباشند، این گرایش وجود دارد که نمره های گروهی در اولین متغیر به میانگین دومین متغیر نزدیک باشد. این اثر در نمره ها تأثیر رگرسیون نامیده می شود و غالباً چون رگرسیون به طرف میانگین دومین متغیر است، آن را رگرسیون در اطراف میانگین می نامند.

- وقتی $r = \pm 1$ است، اصلاً پدیده بازگشت آماری نداریم.
- اگر ضریب همبستگی صفر باشد، بازگشت آماری به طور کامل رخ می دهد و γ که پیش بینی می شود، دقیقاً معادل میانگین است.

ضریب تعیین یا ضریب تشخیص (r^2)

r^2 نسبت یا درصدی از تغییر کل در متغیر وابسته γ را که توسط متغیر X توضیح داده می شود، به دست می دهد.

r^2 نشان می دهد چند درصد از کل واریانس γ ناشی از واریانس X است.

r^2 برآوردی از واریانس مشترک بین دو متغیر را نشان می دهد.

r^2 میزان اشتراک تغییرات است.

r^2 درصد یا نسبتی است که تغییرات متغیر وابسته γ میتواند توسط تغییرات در متغیر X توضیح داده شود و $1-r^2$ درصد یا نسبتی است که تغییرات γ توسط تغییرات X توضیح داده نمی شود.

r^2 کمیتی غیر منفی است و حدودش قطعاً کوچکتر مساوی ۱ و بزرگتر مساوی صفر است. یعنی:

$$0 \leq r^2 \leq 1$$

در مسائل رگرسیون، ضریب همبستگی با جذر گرفتن از ضریب تعیین به دست می آید:

$$r = \pm \sqrt{r^2}$$

نکته مهم: ضریب تعیین برای مقایسه ضریب های همبستگی مفید است. وقتی فردی یک r به مقدار $0/8$ را با r دیگری به مقدار $0/4$ مقایسه می کند، این تمایل وجود دارد که $0/8$ را دو برابر $0/4$ بدانند اما این مقایسه صحیح نیست. ضریب های همبستگی را باید بر حسب تغییر مشترک دو متغیر (ضریب تشخیص) مقایسه کرد:

$$r_1 = 0/8$$

$$r_2 = 0/64$$

$$\longrightarrow \frac{0/64}{0/16} = 4 \text{ است. } r_2 \text{ چهار برابر } r_1$$

$$r_y = 0.4 \quad r_y^2 = 0.16$$

(مثال) اگر معادله خط رگرسیون به صورت $y = 2/4 - 0.8x$ باشد و ضریب تعیین 0.49 باشد، ضریب همبستگی کدام است؟

$$r^2 = 0.49 \quad r = \pm 0.7$$

از آنجا که در معادله رگرسیون β منفی است پس r نیز منفی است و ضریب همبستگی -0.7 است.

نکته مهم: اگر معادلات رگرسیون خطی y بر حسب x و x بر حسب y به صورت زیر باشد:

$$y = \alpha_1 + \beta_1 x \quad x = \alpha_2 + \beta_2 y$$

در این صورت: $r^2 = \beta_1 \times \beta_2$

(مثال) در یک جامعه نرمال دو بعدی معادلات رگرسیون y بر حسب x و x بر حسب y به صورت زیر است:

$$y = -2 + 0.55x \quad x = 1/6 + 1/8 y$$

ضریب تعیین بین دو صفت متغیر x و y کدام است؟

$$r^2 = \beta_1 \times \beta_2 = 0.55 \times 1/8 = 0.99$$

خطای معیار برآورد (S_{xy})

- وقتی همبستگی کامل نیست رگرسیون خطا دارد که به خطای استاندارد برآورد یا خطای معیار برآورد معروف است.
- خطای معیار برآورد مقدار تغییرپذیری نقاط را حول خط رگرسیون فراهم می کند.
- خطای معیار برآورد در توزیع مشترک، مشابه انحراف معیار در توزیع یک متغیری است.
- خطای معیار برآورد را می توان از فرمول های زیر محاسبه کرد:

$$S_{XY} = S_Y \sqrt{1 - r_{xy}^2}$$

$$S_{XY} = \sqrt{\frac{\sum (y - \hat{y})^2}{n-2}}$$

$$S_{XY} = \sqrt{\frac{\sum y^2 - \alpha \sum y - \beta \sum xy}{n-2}}$$

- رابطه ضریب همبستگی و خطای معیار برآورد معکوس است.

انواع ضرایب همبستگی برای بررسی رابطه بین متغیرهای کمی و کیفی

فاصله ای	رتبه ای	طبقه ای اسمی چند ارزشی	طبقه ای اسمی دو ارزشی حقیقی	طبقه ای اسمی دو ارزشی ساختگی	
r_{bis} دورشته ای (با فرض نرمال)			Φ_{bis} فای بای سریال یا دورشته ای	r_{tet} (تتراکوریک)	طبقه ای اسمی دو ارزشی ساختگی
r_{bis} دورشته ای نقطه ای (بدون فرض نرمال)			ضریب فای (ϕ)		طبقه ای اسمی دو ارزشی حقیقی
		C توافق پیرسون			طبقه ای اسمی چند ارزشی
	- ضریب اسپیرمن (بین دو متغیر) T کندال (بین بیش از دو متغیر)				رتبه ای
r^2 پیرسون ضریب ای					فاصله ای

نمونه سؤالات تستی فصل ۱۰

۱- مفهوم ضریب همبستگی ۱/۵- کدام است؟

(۱) به ازای هر واحد تغییر در متغیر ۱/۵،Y واحد تغییر در X رخ می دهد.

(۲) به ازای هر واحد تغییر در متغیر ۱/۵،X واحد تغییر در Y رخ می دهد.

(۳) تغییرات Y، ۱/۵ برابر تغییرات X است.

(۴) تغییرات X، ۱/۵ برابر تغییرات Y است.

۲- درجه آزادی برای آزمون معنادار بودن ضریب همبستگی کدام است؟

(۱) n-۱ (۲) n-۲ (۳) n (۴) (n-۲)(n-۱)

۳- اگر ضریب همبستگی بین X و Y مساوی ۰/۲ باشد، ضریب همبستگی بین X و ۲Y چقدر است؟

(۱) ۰/۸ (۲) ۰/۴ (۳) ۰/۲ (۴) ۰/۱

۴- اگر ضریب همبستگی بین دو متغیر ۰/۹ و دو متغیر دیگر ۰/۳ باشد، می توان گفت همبستگی دو متغیر اول چند برابر قوی تر از دو متغیر دوم است؟

(۱) ۳ برابر (۲) ۲ برابر (۳) ۹ برابر (۴) ۶ برابر

۵- همبستگی نمره های X و Y برابر با ۰/۶ است. در صورتی که $\bar{x} = 15$ ، $s_x = 3$ ، $\bar{y} = 20$ ، $s_y = 4$ باشد، در معادله پیش بینی

Y از طریق X، میانگین نمرات پیش بینی شده \hat{y} چقدر است؟ (دکتری سراسری، ۱۳۹۴)

(۱) ۱۸ (۲) ۲۴ (۳) ۱۵ (۴) ۲۰

۶- بر اساس اطلاعات زیر پارامترهای معادله رگرسیون را حساب کنید (دکتری سراسری، ۱۳۹۴)

$$\bar{x} = 5, \bar{y} = 10, s_x = 4, s_y = 3, Cov_{xy} = 10$$

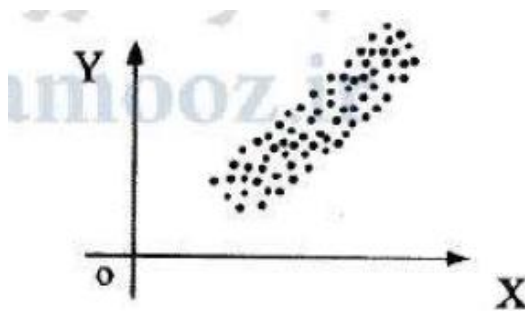
$$\hat{y} = 6/87 + 0/62x \quad (1)$$

$$\hat{y} = 4/25 + 0/33x \quad (2)$$

$$\hat{y} = -6 + 1/11x \quad (3)$$

$$\hat{y} = -5 + 2x \quad (4)$$

۷) کدام جمله در مورد شکل مقابل درست است؟ (دکتری سراسری، ۱۳۹۳)



- (۱) مدل رگرسیون خطی با همبستگی ۱ می تواند برای برازش شود.
 (۲) مدل رگرسیون خطی با شیب ۱ می تواند برای داده ها برازش شود.
 (۳) مدل رگرسیون غیر خطی با شیب ۱ می تواند برای داده ها برازش شود.
 (۴) مدل رگرسیون غیر خطی با همبستگی ۱ می تواند برای برازش شود.

۸- دقت پیش بینی متغیر Y از روی متغیر X بیشتر تابع کدام شاخص است؟ (دکتری سراسری، ۱۳۹۳)

- (۱) ضریب همبستگی (۲) اندازه ی پارامتر شیب (۳) میانگین متغیرهای X و Y (۴) مقدار عرض از مبدأ

۹- بر اساس داده های زیر، بهترین معادله خطی برازش شده به داده ها کدام است؟ (دکتری سراسری، ۱۳۹۲)

x	۱	۲	۳	۴
y	۱	۱	۲	۴

- (۱) $\hat{y} = -\frac{1}{2}x - 1$ (۲) $\hat{y} = -x + \frac{1}{2}$ (۳) $\hat{y} = x - \frac{1}{2}$ (۴) $\hat{y} = \frac{1}{2}x + 1$

۱۰- در یک مدل رگرسیون خطی ساده $y = \alpha + \beta x + \varepsilon$ بر اساس یک نمونه ۲۵ تایی خلاصه اطلاعات زیر حاصل شده است.

مقدار $(\hat{\alpha}$ و $\hat{\beta})$ کدام است؟ (دکتری سراسری، ۱۳۹۲)

$$\sum x_i = 50, \sum y_i = 75, \sum x_i^2 = 1314, \sum y_i^2 = 3831, \sum x_i y_i = 1900$$

- (۱) $(\frac{71}{32}, \frac{25}{64})$ (۲) $(\frac{11}{64}, \frac{15}{64})$ (۳) $(\frac{11}{64}, \frac{15}{32})$ (۴) $(\frac{71}{64}, \frac{25}{64})$

(۱۱) وقتی $r_{yx_1} = 0$ است، احتمالاً $\hat{\beta}_1$ کدام مقدار است؟ (دکتری سراسری، ۱۳۹۱)

- (۱) مثبت (۲) صفر (۳) منفی (۴) همه ی موارد فوق می تواند باشد.

۱۲) اگر همه متغیرها در رگرسیون چند متغیری کمی باشد، X در رابطه با متغیر وابسته (Y) چه چیز را نشان می دهد؟
(دکتری سراسری، ۱۳۹۱)

- (۱) مقدار افزایش Y به ازای هر واحد X (۲) ضریب همبستگی X با Y
- (۳) مقدار افزایش X به ازای هر واحد Y (۴) مجذور ضریب همبستگی X با Y

۱۳) در کدام نوع همبستگی، تعیین درجه رابطه بین دو متغیر پس از کنترل متغیر سوم مطرح است؟ (ارشد سراسری، ۱۳۹۴)

- (۱) فی (۲) توافقی (۳) جزئی (۴) دو رشته ای

۱۴- در صورتی که اطلاعات زیر داده شده باشد، شیب معادله پیش بینی شده، چقدر است؟ (ارشد سراسری، ۱۳۹۴)

$$X = 7 - 5 - 4 - 3 - 1$$

$$Y = 2 - 10 - 6 - 8 - 16$$

$$r_{xy} = 0.7 \quad S_x^2 = 16 \quad S_y^2 = 4$$

$$0.35 \quad (4)$$

$$0.175 \quad (3)$$

$$2/8 \quad (2)$$

$$1/4 \quad (1)$$

۱۵- چه زمان تساوی $b_{xy} = r_{xy}$ برقرار است؟ (ارشد سراسری، ۱۳۹۴)

$$S_x^2 = S_y^2 \quad (1) \quad r_{xy}^2 = r_{yx}^2 \quad (2) \quad r_{xy} = \frac{S_x}{S_y} \quad (3) \quad S_x^2 \cdot S_y^2 = r_{xy}^2 \quad (4)$$

۱۶- همبستگی میان طول قد و امتیاز آوری در بازیکنان حرفه ای بسکتبال نسبت به افراد عادی جامعه احتمالاً
(ارشد سراسری، ۱۳۹۲)

- (۱) کمتر است. (۲) بیشتر است (۳) تفاوتی ندارد. (۴) یکسان است ولی جهت آن فرق دارد

۱۷- باتوجه به معادله رگرسیون $\hat{Y} = -2/5 + 0.75 X$ می توان گفت: (ارشد سراسری، ۱۳۹۱)

- (۱) ضریب همبستگی بین X و Y برابر 0.75 است.
(۲) خط رگرسیون در نقطه $Y = 0.75$ محور Y حاصل می شود.
(۳) به ازای هر واحد تغییر در X ، 0.75 واحد افزایش در Y حاصل می شود.
(۴) به ازای هر واحد تغییر در X به اندازه $2/5$ واحد کاهش در Y مشاهده می شود.

۱۸- اگر نتایج آزمون هوش دانش آموزان، آنها را در دو سطح هوشی بالا و پایین تقسیم کند، با کدام نوع ضریب همبستگی رابطه بین هوش و نمره درس ریاضی را می توان محاسبه کرد؟ (ارشد سراسری، ۱۳۹۰)

(۱) دو رشته ای (۲) فی (۳) پیرسون (۴) دو رشته ای نقطه ای

پاسخ ها

۱- گزینه ۲

ضریب همبستگی، میزان تغییر در متغیر Y را به ازای هر واحد تغییر در متغیر X نشان می دهد.

۱- گزینه ۲

درجه آزادی توزیع t برای معناداری ضریب همبستگی $n-2$ است.

۳- گزینه ۳

ضرب کردن متغیرها در اعداد مثبت تأثیری در ضریب همبستگی آنان ندارد.

۴- گزینه ۳

$$r_1 = 0/9 \quad r_1^2 = 0/81$$

$$\longrightarrow \frac{0/81}{0/9} = 9 \quad \text{نه برابر } r_2 \text{ است.}$$

$$r_2 = 0/3 \quad r_2^2 = 0/9$$

۵- گزینه ۴

ابتدا باید ضریب رگرسیون و سپس معادله رگرسیون را بنویسیم، سپس به جای X میانگین آن را که در صورت سؤال داریم بگذاریم تا میانگین Y را به دست آوریم. چون همبستگی و انحراف معیارها را داریم از فرمول زیر می توانیم برای محاسبه ضریب رگرسیون استفاده کنیم:

$$\beta = r \frac{S_y}{S_x} = 0/6 \times \frac{4}{3} = 0/8 \quad \alpha = \bar{y} - \beta \bar{x} = 20 - 0/8 \times 15 = 8$$

$$\hat{y} = \alpha + \beta x \quad \hat{y} = 8 + 0/8 \times 15 = 20$$

۶- گزینه ۱

ابتدا ضریب رگرسیون را محاسبه می کنیم.

$$\beta = \frac{Covxy}{S_x^2} = \frac{10}{16} = 0.62$$

احتیاجی به ادامه حل و به دست آوردن معادله نیست زیرا تنها گزینه ای که ضریب رگرسیون آن ۰/۶۲ است، گزینه ۱ است.

۷- گزینه ۲

از شکل مشخص است که رگرسیون خطی است و همبستگی برای رگرسیون خطی مطرح می شود، پس گزینه ۳ و ۴ اشتباه است. از سوی دیگر در همبستگی کامل ۱ تمام نقاط روی یک خط قرار می گیرند، در صورتیکه در شکل نقاط حول یک خط راست هستند ولی همه بر روی خط قرار ندارند پس همبستگی ۱ نیست اما شیب خط مثبت و ۱ است.

۸- گزینه ۱

در پیش بینی متغیر ها، ضریب همبستگی بین متغیر در دقت پیش بینی مهم ترین نقش را دارد و هر چه ضریب همبستگی بین متغیرها بیشتر باشد پیش بینی متغیر ملاک از روی پیش بینی قوی تر است.

۹- گزینه ۳

ابتدا باید ضریب رگرسیون را پیدا کنیم. بهترین فرمول برای محاسبه ضریب رگرسیون از طریق داده های خام فرمول زیر است:

$$\beta = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2}$$

ابتدا جدول مقابل را تشکیل می دهیم:

x	y	xy	x ²	
۱	۱	۱	۱	
۲	۱	۲	۴	
۳	۲	۶	۹	
۴	۴	۱۶	۱۶	
$\sum x = 10$		$\sum xy = 25$	$\sum x^2 = 30$	

$$\beta = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2} = \frac{4 \times 25 - 10 \times 8}{4 \times 30 - 10^2} = 1$$

بدون نوشتن معادله رگرسیون می توانیم بفهمیم که گزینه ۱ صحیح است زیرا تنها در این گزینه ضریب رگرسیون ۱ است.

۱۰) گزینه صحیح وجود ندارد.

$$\beta = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2} = \frac{25 \times 1900 - 50 \times 75}{25 \times 1314 - 50^2} = \frac{43750}{30350} = \frac{875}{607}$$

$$\bar{y} = \frac{\sum y}{n} = \frac{75}{25} = 3 \quad \bar{x} = \frac{\sum x}{n} = \frac{50}{25} = 2 \quad \alpha = \bar{y} - \beta \bar{x} = 3 - \frac{875}{607} \times 2 = \frac{71}{607}$$

در سال ۹۲ شاهد سؤالاتی بودیم که بسیار وقت گیر با اعداد بسیار بزرگ بودیم و گزینه صحیح نداشتیم که نمونه اش سؤال حاضر است. احتمال تکرار نمونه این سؤالات بسیار بعید است پس نگران نباشید. سؤال بعدی همین سال خواسته بود ضریب همبستگی محاسبه شود که باز هم گزینه درست در بین جوابها وجود نداشت و از حل آن صرفنظر می کنیم.

۱۱- گزینه ۲

وقتی همبستگی صفر باشد، ضریب رگرسیون نیز صفر خواهد بود. این مطلب را می توان با فرمول زیر نیز توجیح کرد:

$$\beta = r \frac{S_y}{S_x} \quad \beta = 0 \times \frac{S_y}{S_x} = 0$$

۱۲) گزینه ۱

ضریب همبستگی، میزان تغییر در متغیر Y را به ازای هر واحد تغییر در متغیر X نشان می دهد.

۱۳) گزینه ۳

ضریب همبستگی جزئی یا تفکیکی: اثر متغیر سوم را از دو متغیر قبلی بر می داریم و رابطه اصلی بین دو متغیر را ضریب جزئی یا تفکیکی می گویند.

۱۴- گزینه ۴

سؤال شیب خط رگرسیون یعنی همان ضریب رگرسیون را می خواهد و با توجه به اینکه سؤال همبستگی را به ما داده پس از فرمول زیر استفاده می کنیم:

$$\beta = r \frac{S_y}{S_x} = 0.7 \times \frac{2}{4} = 0.35$$

۱۵- گزینه ۱

با توجه به فرمول رو به رو اگر $S_x = S_y$ باشد، با هم ساده می شوند و $b_{xy} = r_{xy}$ برقرار می شود:

$$\beta = r \frac{S_y}{S_x} \quad \beta = r \frac{S_y}{S_y}$$

۱۶- گزینه ۱

همبستگی در جامعه ای که از نظر متغیر مورد نظر ناهمگن باشد، بیشتر است. بازیکنان حرفه ای بسکتبال نسبت به افراد عادی جامعه همگن ترند و قدشان بیشتر بهم نزدیک است در نتیجه همبستگی در بین آنها کمتر است.

۱۷- گزینه ۳

به ازای هر واحد تغییر در متغیر x ، متغیر y به اندازه β تغییر می کند.

۱۸- گزینه ۱

وقتی هوش دانش آموزان را به دو سطح تقسیم می کنیم یعنی از آن یک متغیر کیفی دو ارزشی ساخته ایم. از طرفی نمره ریاضی یک متغیر فاصله ای است و ضریب همبستگی بین متغیر طبقه ای اسمی دو ارزشی ساختگی و متغیر فاصله ای ، با توجه به جدول ارائه شده در فصل، ضریب دو رشته ای است.