

«با شناخت همبستگی بین پدیده‌ها، می‌توان
به رابطه علت و معلولی بسیاری از حوادث پی برد.»

فصل سوم

همبستگی متغیرها و ضریب همبستگی

هدفهای رفتاری: در پایان این فصل، از فراگیر انتظار می‌رود:

- ۱- همبستگی را تعریف کند.
- ۲- مفهوم کلی همبستگی را توضیح دهد.
- ۳- علل وجود همبستگی بین پدیده‌ها را برشمارد.
- ۴- انواع همبستگی را توضیح دهد.
- ۵- درجات و حالات همبستگی را نام برد.
- ۶- درجات و حالات همبستگی را تعریف کند.
- ۷- نمودارهای پراکندگی را رسم کرده، تفسیر کند.
- ۸- ضریب همبستگی را محاسبه کند.
- ۹- معنای ضریب همبستگی را بیان کند.
- ۱۰- ضریب تعیین را محاسبه کند.
- ۱۱- کاربرد ضریب تعیین را توضیح دهد.
- ۱۲- درجه معنی‌دار بودن ضریب همبستگی را تبیین کند.
- ۱۳- معادله خط رگرسیون را بنویسد و خط رگرسیون را رسم کند.
- ۱۴- کوواریانس «همپراش» را محاسبه و تفسیر کند.

مقدمه

حتماً دیده‌اید و یا شنیده‌اید که برخی از کوهنوردان، در ارتفاعات بالای کوهستان خون دماغ

می‌شوند. هیچ فکر کرده‌اید که دلیل این امر چیست؟ فیزیک‌دانان نشان داده‌اند که هرچه ارتفاع از سطح دریا افزایش پیدا کند، فشار هوا کاهش می‌یابد. و دقیقاً به همین دلیل است که گروهی، در ارتفاعات، خون دماغ می‌شوند. در حقیقت فشار داخلی بدن آنها، بر فشار بیرون غالب می‌شود. مثالی دیگر: معمولاً به چه کسانی چاق و به چه کسانی لاغر گفته می‌شود؟ آیا وزن یا قد معینی به عنوان معیار چاقی یا لاغری وجود دارد که تعیین‌کننده قضاوت ما درباره چاقی یا لاغری افراد باشد؟ طبیعی است که پاسخ این سؤال منفی باشد. اما نوع رابطه‌ای که باید بین وزن و قد افراد برقرار باشد، می‌تواند به عنوان یک معیار به کار برده شود. مثلاً در روان‌شناسی رشد می‌گویند، سانتیمترهای قد افراد در دوران رشد باید حدود ۱۰۰ واحد بیشتر از کیلوگرهای وزن آنها باشد. برای مثال کسی که ۱۶۰ سانتیمتر قد دارد، باید حدود ۶۰ کیلوگرم وزن داشته باشد تا نه چاق تلقی شود و نه لاغر. حال اگر فردی که ۱۶۰ سانتیمتر قد دارد، ۷۰ کیلوگرم وزن داشته باشد اصطلاح چاقی برای او به کار می‌رود و اگر فرضاً همین شخص ۵۰ کیلوگرم وزن داشته باشد؛ اصطلاحاً وی را لاغر می‌نامند. به‌طور دقیق‌تر برای تعیین چاقی و لاغری امروزه از شاخص توده بدن (B.M.I) که مخفف عبارت: Body Mass Index است و به‌صورت زیر محاسبه می‌شود، استفاده می‌کنند:

$$\text{وزن با مقیاس کیلوگرم} \\ \text{شاخص توده بدن} = \frac{\text{وزن با مقیاس کیلوگرم}}{(\text{قد به متر})^2}$$

که اگر این شاخص بین ۲۰ تا ۲۵ باشد قد و وزن انسان حکایت از سالم بودن دارد اگر بین ۲۵ تا ۳۰ باشد فرد را چاق می‌دانند و چنانچه این شاخص از ۳۰ تجاوز کند، خطرناک است و باید مواظبت بیشتری از وزن انجام شود.

به این ترتیب، می‌توان گفت که: بین فشار هوا و ارتفاع از سطح دریا و نیز بین وزن و قد افراد، نوعی رابطه (همبستگی) وجود دارد. در این زمینه می‌توان مثالهای مختلفی نظیر:

— رابطه بین کشیدن سیگار و ابتلا به بیماری سرطان ریه

— رابطه بین مقدار باران سالانه و مقدار تولید محصولات کشاورزی

— رابطه بین سرعت اتومبیل و طول خط ترمز آن

— رابطه بین هزینه‌های تولید و مقدار تولید و ... را ارائه داد.

در این فصل، به بررسی انواع همبستگیها، مقدار و شدت و ضعف همبستگیها، مدل ریاضی و رابطه‌های بین متغیرهای مختلف خواهیم پرداخت. توجه داشته باشید که برخلاف آمار توصیفی که فقط

یک متغیر را برای هریک از عناصر جامعه یا نمونه، اندازه‌گیری و بررسی می‌کردیم، در مبحث همبستگی، دو یا چند متغیر از جامعه یا نمونه آماری را در نظر گرفته، امکان وجود یا عدم وجود همبستگی بین آن دو یا چند متغیر را مورد مطالعه قرار خواهیم داد و بر دو مطلب زیر تأکید بیشتری خواهیم داشت :

۱- آیا دو متغیر به همدیگر مرتبط هستند یا خیر؟

۲- اگر دو متغیر مرتبط هستند، نوع ارتباط و شدت و ضعف ارتباط چگونه است؟

تعریف همبستگی

همبستگی را می‌توان نوعی رابطه کمی (مقداری) تعریف کرد که ممکن است بین متغیرهای مختلف وجود داشته باشد. به عبارت دیگر؛ همبستگی خصیصه‌ایست بین دو یا چند متغیر به نحوی که تغییر در یکی از متغیرها، تغییری قابل پیش‌بینی در متغیر یا متغیرهای دیگر را به دنبال داشته باشد. بنابراین، می‌توان با شناخت همبستگی در بسیاری از مسائل اقتصادی، مالی، فرهنگی، سیاسی، نظامی و... عمل تصمیم‌گیری را سهولت بخشید. مثلاً با محاسبه اندازه همبستگی بین «مقدار تولید» یک کارخانه و «مقدار ضایعات» آن کارخانه، مشکلات را بررسی کرد و یا با شناخت رابطه بین «مقدار کارآیی کارکنان» و «نتایج آزمایشهای استخدامی» در یک مؤسسه، به تنظیم و تدوین سؤالات و آزمونهای استخدامی پرداخت. ذکر این نکته ضروری است که هرگاه وجود همبستگی بین دو متغیر برای ما اثبات شد، نمی‌توان عوامل علت و معلول را از هم تشخیص داد، یعنی پدیده همبستگی را نباید با موضوع علت و معلول مساوی دانست، ضمن اینکه در برخی از انواع همبستگیها، علت و معلول نیز می‌تواند مطرح باشد.

علل وجود همبستگی بین متغیرها

در بررسی رابطه بین متغیرها علل مختلفی را می‌توان مشاهده کرد. از آن جمله :

تصادفی بودن همبستگی

گاهی همبستگی بین دو متغیر ممکن است جنبه کاملاً تصادفی داشته باشد. غالباً در این موارد

با تعویض اعضای نمونه یا با زیاد کردن حجم نمونه، نتیجه بررسی تغییر می کند.

تأثیر یک عامل شناخته شده

ممکن است وجود همبستگی بین دو متغیر بر اثر تأثیر یک عامل شناخته شده باشد. برای مثال بین هزینه های ضروری خانواده ها «مثلاً غذا» و هزینه های غیر ضروری (مثلاً تفریح و گردش)، نوعی همبستگی مشاهده می شود و این رابطه به دلیل پدیده «درآمد» است. به این معنی که، با بالا رفتن درآمد خانواده ها، به طور موازی، هم هزینه خوراک بالا می رود و هم هزینه تفریح و گردش. ضمن اینکه اگر درآمد این خانواده ها به دلیلی کاهش پیدا کند، فوراً هزینه های گردش را به صفر می رسانند، اما نمی توانند هزینه های خوراک را به صفر برسانند.

مسأله علیّت

گاهی یک متغیر در متغیر دیگر تأثیر دارد و باعث نوعی همبستگی بین آن متغیرها می شود. مثلاً رشد درختان با مقدار آبی که به آنها داده می شود، نوعی رابطه دارد و این بیشتر مربوط به اصل علت و معلولی است. بسیاری از مباحث مربوط به همبستگی از نوع مسأله علت و معلولی هستند.

انواع همبستگی

اگر بین دو پدیده، همبستگی وجود داشته باشد، نوع آن از دو حالت زیر خارج نخواهد بود:

۱- همبستگی مستقیم (مثبت)

۲- همبستگی معکوس (غیر مستقیم یا منفی)

هرگاه در یک بررسی نمونه ای، تغییرات عددی دو پدیده مانند x و y در یک جهت باشد، به طوری که زیاد شدن x با زیاد شدن y و یا کم شدن x با کم شدن y همراه باشد، می گویند x و y همبستگی مستقیم دارند. مانند رابطه قد و وزن افراد در دوران رشد. اما در مواردی که تغییرات مقداری دو متغیر x و y در دو جهت مخالف باشد، یعنی زیاد شدن یک متغیر (مثل x) با کم شدن متغیر دیگر (مثلاً y) همراهی کند و بالعکس، مانند رابطه فشار هوا و ارتفاع از سطح دریا، همبستگی را غیر مستقیم یا معکوس می نامند، یکی از معیارهای تشخیص نوع همبستگی، ضریب همبستگی است

که در صفحات بعد به مطالعه آن خواهیم پرداخت. چون این ضریب در همبستگی مستقیم، مثبت و در همبستگی معکوس، منفی می شود، به همین دلیل، گروهی از آمارشناسان همبستگی مستقیم را مثبت و همبستگی معکوس را منفی نیز می نامند.

شدت و ضعف همبستگی (مقدار همبستگی)

مقدار همبستگی بین دو متغیر از دو حالت زیر خارج نیست.

۱- همبستگی کامل (وجود ۱۰۰٪ ارتباط بین دو متغیر)

۲- همبستگی ناقص (وجود کمتر از ۱۰۰٪ ارتباط بین دو متغیر)

هرگاه رابطه دو متغیر x و y به گونه ای باشد که هر تغییر در اندازه x تغییر متناسبی در اندازه y را به همراه داشته باشد، همبستگی را کامل (صد درصد) می نامند، مانند رابطه بین قطر دایره و اندازه محیط آن. اما اگر رابطه بین دو متغیر x و y کمتر از صد درصد باشد، همبستگی را ناقص می گویند، نظیر همبستگی بین قد و وزن دانش آموزان یک دبیرستان که می تواند تحت تأثیر عوامل اقتصادی، ارث، ژنتیک و نظایر آن برای افراد یکسان نباشد.

ابزارهای تشخیص همبستگی

برای تشخیص نوع همبستگی و مقدار شدت و ضعف آن، از ابزارهای مختلفی می توان استفاده کرد که در این کتاب به بررسی آنها خواهیم پرداخت.

نمودار پراکندگی (دیاگرام پراکنش)

ساده ترین روش بررسی وجود یا عدم وجود همبستگی بین دو متغیر، رسم دیاگرام پراکنش است. برای رسم چنین دیاگرامی، باید یک دستگاه مختصات را در نظر بگیریم و اندازه یکی از صفات را تحت عنوان صفت x روی محور طولها و اندازه های متناظر صفت مقابل آن را تحت عنوان صفت y ، روی محور عرضها قرار داده، برای هر مشاهده یک نقطه روی محورهای مختصات به دست

بیاوریم، به طوری که طول آن نقطه اندازه صفت x و عرض آن نقطه اندازه صفت y باشد. مجموعه نقاط حاصل را روی محورهای مختصات دیاگرام پراکنش گویند. با کمک شکل کلی نقاط دیاگرام پراکنش، می توان به وجود یا عدم وجود همبستگی و نیز به نوع آن پی برد.

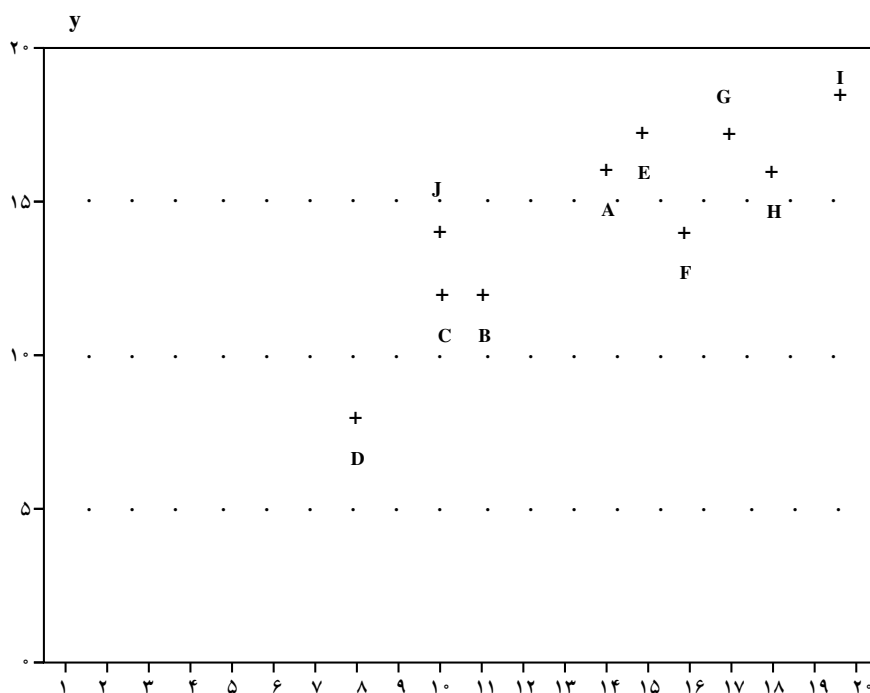
مثال ۱- فرض کنید نمرات ریاضی و نمرات آمار ۱۰ دانش آموز را طبق جدول زیر در اختیار دارید:

جدول ۱

نام دانش آموز	A	B	C	D	E	F	G	H	I	J
نمره ریاضی $x=$	۱۴	۱۱	۱۰	۸	۱۵	۱۶	۱۷	۱۸	۲۰	۱۰
نمره آمار $y=$	۱۶	۱۲	۱۲	۸	۱۷	۱۴	۱۷	۱۶	۱۸	۱۴

دیاگرام پراکنش اطلاعات فوق را رسم کنید.

حل: نمرات ریاضی را روی محور طولها و نمرات آمار را روی محور عرضها قرار می دهیم و برای هر دانش آموز یک نقطه روی دستگاه مختصات به دست می آوریم.



نمودار ۱

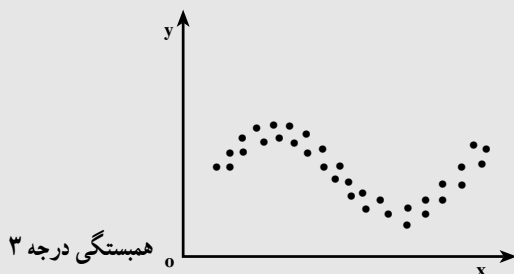
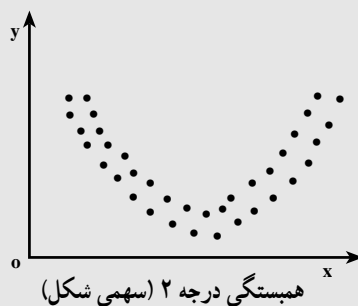
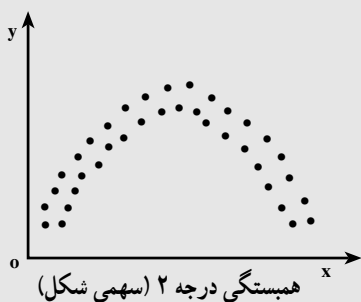
تفسیر دیاگرام پراکنش — برای اینکه مشخص شود آیا بین متغیرهای x_i و y_i همبستگی وجود دارد یا خیر؟ میانگین مقادیر $(\bar{x})x_i$ و میانگین مقادیر $(\bar{y})y_i$ را محاسبه کرده، روی دیاگرام پراکنش نقطه‌ای با طول \bar{x} و عرض \bar{y} ، به دست می‌آوریم و دو خط به موازات محورهای طولها و عرضها روی نقاط دیاگرام رسم می‌کنیم. به این ترتیب، چهار ناحیه ۱، ۲، ۳، ۴ به دست می‌آید. اکنون:

(۱) — اگر بیشتر نقاط دیاگرام در نواحی (۱) و (۳) باشند، همبستگی مستقیم و اگر این تجمع در نواحی (۲) و (۴) مشاهده شود، همبستگی معکوس خواهد بود.

(۲) — اگر ضمن رعایت بند (۱) نقاط دیاگرام درست روی یک خط مستقیم قرار بگیرند، همبستگی کامل است و چنانچه نقاط دیاگرام در اطراف یک خط مستقیم باشند، همبستگی ناقص خواهد بود.

(۳) — چنانچه نقاط دیاگرام پراکنش به موازات خطوط سازنده نواحی چهارگانه واقع شوند و یا هیچ نظام خاصی را تبعیت نکرده، و در مناطق چهارگانه پخش شده باشند، نشانه عدم همبستگی است.

تذکر: تفسیر فوق‌الذکر مربوط به همبستگیهای خطی است. توجه داشته باشید که ممکن است بین دو یا چند متغیر، همبستگی غیرخطی نظیر درجه دو (سهمی)، درجه ۳ و ... نیز وجود داشته باشد، مانند دیاگرامهای پراکنش زیر که بحث و بررسی آنها خارج از برنامه این کتاب است.

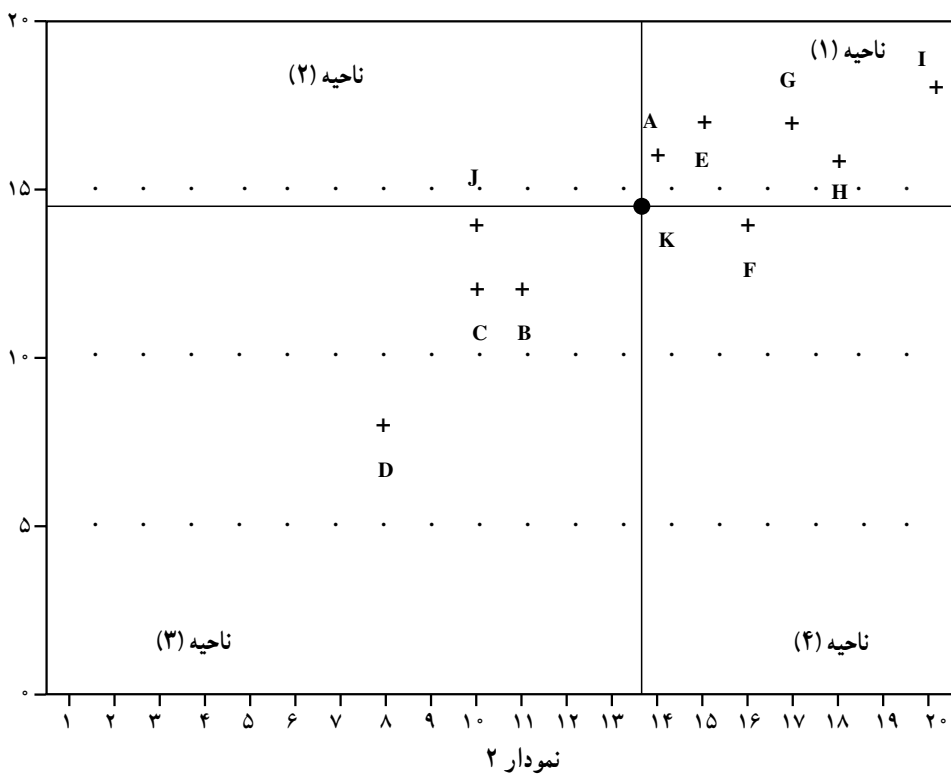


مثال ۲- در مثال ۱، \bar{x} و \bar{y} را محاسبه کرده، با کمک دیاگرام پراکنش در مورد وجود یا عدم وجود همبستگی، نوع و مقدار آن بحث کنید.

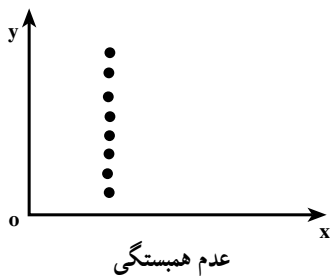
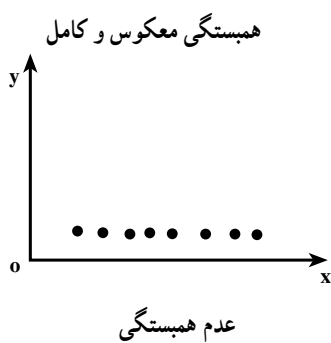
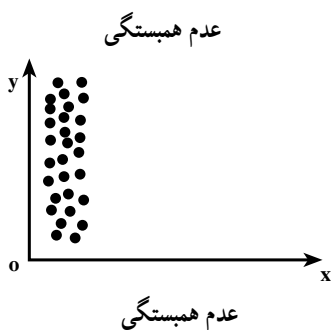
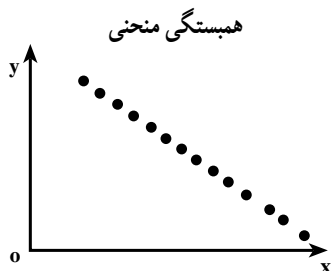
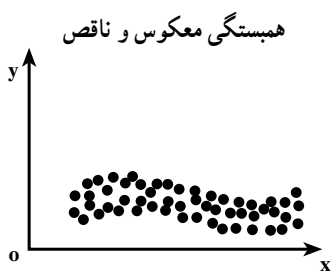
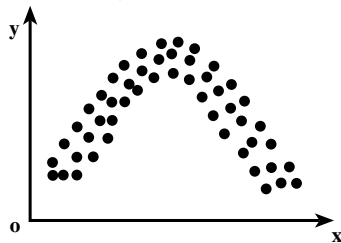
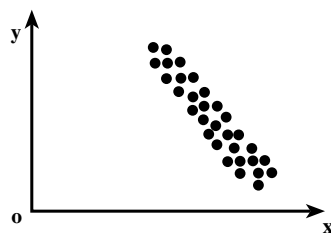
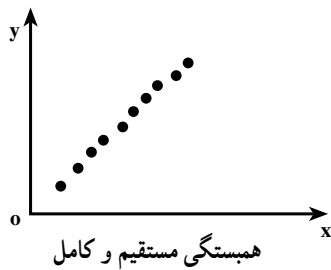
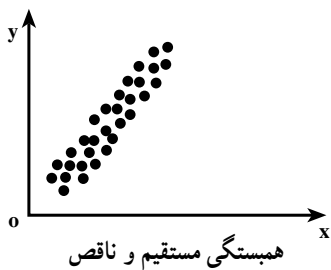
$$\bar{x} = \frac{14+11+10+8+15+16+17+18+20+10}{10} = 13/9$$

$$\bar{y} = \frac{16+12+12+8+17+14+17+16+18+14}{10} = 14/4$$

$$\bar{x} = 13/9 \quad \bar{y} = 14/4 \quad K \left| \begin{array}{l} 13/9 \\ 14/4 \end{array} \right.$$



ملاحظه می‌شود که بیشتر نقاط دیاگرام، در مناطق (۱) و (۳) در اطراف یک خط مستقیم (نه درست روی یک خط مستقیم) واقع شده‌اند و این می‌تواند نشانه همبستگی مستقیم و ناقص باشد. انواع دیاگرامهای پراکنش - در این قسمت، با تعدادی دیاگرام پراکنش، همراه با تفسیر آنها آشنا می‌شوید :



خودآزمایی – قد و وزن ۱۰ نفر از دانش‌آموزان کلاس خودتان را سؤال کنید و نمودار پراکنش آنها را رسم کنید.

ضریب همبستگی^۱

بعد از اینکه با کمک دیاگرام پراکنش مشخص شد که بین دو صفت x و y ، همبستگی خطی وجود دارد، برای تعیین درجه همبستگی، از پدیده‌ای به نام ضریب همبستگی استفاده خواهیم کرد، که با علامت $\rho_{(x,y)}$ (RHO)^۲ یا برآوردکننده آن r نشان داده می‌شود.

ضریب همبستگی شاخصی است که می‌تواند در فاصله -1 تا $+1$ تغییرات را بپذیرد. علامت آن نشانه نوع همبستگی (مستقیم بودن یا معکوس بودن) و قدر مطلق آن تعیین‌کننده درجه همبستگی است.

تفسیر ضریب همبستگی – ضریب همبستگی در دامنه $[-1, +1]$ تغییر می‌کند و به صورت زیر تفسیر می‌شود:

اگر $r = +1$ باشد، همبستگی مستقیم و کامل است.

اگر $r = -1$ باشد، همبستگی معکوس و کامل است.

اگر $r = 0$ باشد، نشانه عدم همبستگی است.

اگر $|r| > 0.1$ باشد، در صورت معنی‌دار بودن، (به اندازه کافی بزرگ و قابل قبول بودن) همبستگی ناقص اعلام می‌شود. در مورد معنی‌دار بودن ضریب همبستگی، بعداً بیشتر، توضیح خواهیم داد.

چگونگی محاسبه ضریب همبستگی – برای محاسبه ضریب همبستگی، روشهای مختلفی وجود دارد که معروفترین آنها را مورد بحث و بررسی قرار خواهیم داد.

ضریب همبستگی گشتاوری پیرسن – اگر اندازه‌های n مشاهده از زوجهای مرتب (x, y) را به صورت تصادفی در اختیار بگیریم، پس از کسب اطمینان از وجود همبستگی خطی بین x و y ، می‌توان ضریب همبستگی گشتاوری پیرسن را (برای تعیین درجه همبستگی خطی) به صورت زیر محاسبه کرد:

^۱ – Coefficient of correlation

^۲ – علامت ρ برای ضریب همبستگی در کل جامعه و علامت r برای ضریب همبستگی در نمونه‌های مورد آزمایش به کار برده می‌شود.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\sum (x_i - \bar{x})^2\right] \left[\sum (y_i - \bar{y})^2\right]}} \quad (\text{فرمول ۱})$$

فرمول ۱ را به صورت زیر نیز می‌توان بیان کرد :

$$\text{ضرب همبستگی} = \frac{\text{مجموع حاصلضرب انحرافات } x \text{ ها از میانگینشان در انحرافات } y \text{ ها از میانگینشان}}{\sqrt{(\text{مجموع مجزورات انحرافات } x \text{ از میانگینشان}) \times (\text{مجموع مجزورات انحرافات } y \text{ از میانگینشان})}}$$

در این فرمول (فرمول ۱) اگر به جای اصطلاح فارسی «مجموع حاصل ...» از علامت SP_{xy} (Sum Product ...) و به جای اصطلاح فارسی «مجموع مجزورات ...» از علامت SS (Sum Squares ...) استفاده کنیم، فرمول ۱ را به صورت زیر نیز می‌توان نوشت :

$$r = \frac{SP_{xy}}{\sqrt{SS_x \cdot SS_y}} \quad (\text{فرمول ۲ کوتاه شده فرمول ۱})$$

تذکر: اگر انحراف معیار مقادیر متغیرهای x و y در اختیار باشند، ضریب همبستگی را به صورت زیر نیز می‌توان محاسبه کرد :

$$r = \frac{SP_{xy}}{n \cdot \sigma_x \cdot \sigma_y} \quad (\text{فرمول ۳})$$

و چنانچه میانگینهای دو متغیر x و y (یا میانگین یکی از آنها) اعشاری باشد، بهتر است ضریب همبستگی را از فرمول ۴ که با فرمولهای ۱ و ۲ و ۳ هم‌ارز است، محاسبه کنیم تا با سهولت بیشتری به نتیجه برسیم.

$$r = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sqrt{\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right] \left[\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right]}} \quad (\text{فرمول ۴})$$

مثال ۳- در جدول زیر مقدار ضریب همبستگی را با کمک :

x_i	۴	۶	۸	۱۰	۱۲
y_i	۲۵	۲۰	۱۵	۱۰	۵

(۱) فرمول ۲،

۲) فرمول ۳،

۳) فرمول ۴، محاسبه کنید :

حل :

۱) با کمک فرمول ۲

$$\bar{x} = \frac{40}{5} = 8, \quad \bar{y} = \frac{75}{5} = 15$$

جدول ۲

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
۴	۲۵	-۴	۱۰	-۴۰	۱۶	۱۰۰
۶	۲۰	-۲	۵	-۱۰	۴	۲۵
۸	۱۵	۰	۰	۰	۰	۰
۱۰	۱۰	+۲	-۵	-۱۰	۴	۲۵
۱۲	۵	+۴	-۱۰	-۴۰	۱۶	۱۰۰
۴۰	۷۵			$SP_{xy} = -100$	$SS_x = 40$	$SS_y = 250$

$$\bar{x} = \frac{40}{5} = 8, \quad \bar{y} = \frac{75}{5} = 15$$

$$r = \frac{SP_{xy}}{\sqrt{SS_x \cdot SS_y}} = \frac{-100}{\sqrt{40 \times 250}} = \frac{-100}{\sqrt{10000}} = \frac{-100}{100} = -1$$

(۲) با کمک فرمول ۳

جدول ۳

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
۴	۲۵	-۴	۱۰	۱۶	۱۰۰
۶	۲۰	-۲	۵	۴	۲۵
۸	۱۵	۰	۰	۰	۰
۱۰	۱۰	+۲	-۵	۴	۲۵
۱۲	۵	+۴	-۱۰	۱۶	۱۰۰
۴۰	۷۵			$\sum (x_i - \bar{x})^2 = ۴۰$	$\sum (y_i - \bar{y})^2 = ۲۵۰$

$$\bar{x} = \frac{۴۰}{۵} = ۸$$

$$\bar{y} = \frac{۷۵}{۵} = ۱۵$$

$$SP_{xy} = -۱۰۰$$

$$\sigma_x = \sqrt{\frac{۴۰}{۵}} = \sqrt{۸}$$

$$\sigma_y = \sqrt{\frac{۲۵۰}{۵}} = \sqrt{۵۰}$$

$$r = \frac{SP_{xy}}{n \cdot \sigma_x \cdot \sigma_y} = \frac{-۱۰۰}{۵ \times \sqrt{۸} \times \sqrt{۵۰}} = \frac{-۱۰۰}{۵ \times \sqrt{۴۰۰}} = \frac{-۱۰۰}{+۱۰۰} = -۱$$

(۳) با کمک فرمول ۴

جدول ۴

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
۴	۲۵	۱۰۰	۱۶	۶۲۵
۶	۲۰	۱۲۰	۳۶	۴۰۰
۸	۱۵	۱۲۰	۶۴	۲۲۵
۱۰	۱۰	۱۰۰	۱۰۰	۱۰۰
۱۲	۵	۶۰	۱۴۴	۲۵
۴۰	۷۵	۵۰۰	۳۶۰	۱۳۷۵

$$\sum x_i = ۴۰ \text{ و } \sum y_i = ۷۵$$

$$\sum x_i y_i = ۵۰۰$$

$$\sum x_i^2 = ۳۶۰ \text{ و } \sum y_i^2 = ۱۳۷۵$$

$$r = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sqrt{\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n} \right] \left[\sum y_i^2 - \frac{(\sum y_i)^2}{n} \right]}} = \frac{۵۰۰ - \frac{۴۰ \times ۷۵}{۵}}{\sqrt{\left[۳۶۰ - \frac{(۴۰)^2}{۵} \right] \left[۱۳۷۵ - \frac{(۷۵)^2}{۵} \right]}}$$

$$= -۱$$

۱- فرمول انحراف معیار همانگونه که در آمار (۱) خوانده‌اید به صورت $\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}}$ می‌باشد.

معنی دار بودن ضریب همبستگی — گفتیم که هرگاه $-1 < r < +1$ و $r \neq 0$ باشد، اگر همبستگی معنی دار بود، همبستگی را ناقص می گویند. حال سؤال اینست که با توجه به اینکه بین صفر تا -1 و بین صفر تا $+1$ بی نهایت عدد وجود دارد، مرز قابل قبول بودن و قابل قبول نبودن ضریب همبستگی کجاست؟ جواب سؤال را باید در جدولی به نام جدول معنی دار بودن r جستجو کنیم. این جدول را در انتهای کتاب ملاحظه می کنید که شامل پنج ستون است. در ستون اول از سمت چپ، درجات آزادی ($n-2$) نوشته شده و چهار ستون بعدی بترتیب به $0/01$, $0/05$, $0/02$, $0/01$ و $0/01$ اختصاص داده شده اند. این اعداد درجات عدم اطمینان آزمون هستند که اختلاف آنها با عدد (۱) میزان اطمینان قضاوت ما را در آزمون مشخص خواهد ساخت که اگر میزان اطمینان (احتمال درستی قضاوت) را با P نشان دهیم، می توان نوشت :

$$P = 1 - \alpha$$

به یک مدل ساده از جدول معنی دار بودن r نگاه کنید :

درجه آزادی	۰/۱	۰/۰۵	۰/۰۲	۰/۰۱
۱				
۲				
۳				
۴				

برای استفاده از جدول معنی دار بودن r ، ابتدا درجه آزادی را از رابطه $(D.F. = n - 2)$ به دست می آوریم (که در آن n تعداد زوجهای مرتب « x, y » است.) و سپس به جدول معنی دار بودن r مراجعه می کنیم. اگر ضریب همبستگی محاسبه شده مساوی یا بزرگتر از عدد جدول (نقطه بحرانی) باشد، (که در محل تلاقی درجه آزادی و α نوشته شده) ضریب همبستگی معنی دار و قابل قبول خواهد بود و در غیر این صورت، ضریب همبستگی معنی دار نیست و غیر قابل قبول اعلام خواهد شد. اگر $r = 0$ باشد، دقیقاً می توان گفت که بین x و y ارتباط خطی وجود ندارد. (ولی ممکن است بین دو متغیر مزبور، ارتباط غیر خطی وجود داشته باشد.)

مثال ۴- فرض کنید در محاسبه همبستگی بین هفت زوج مرتب از x و y ، $r = 0/8$ شده باشد و در جدول معنی دار بودن، نقاط بحرانی برای درجه آزادی پنج ($D.F = 7 - 2 = 5$) طبق جدول زیر مشاهده شود :

درجه آزادی	0/1	0/05	0/02	0/01
۱				
۲				
۳				
۴				
۵	0/6694	0/7545	0/8329	0/8745

معلوم کنید x و y چه رابطه ای دارند؟

حل: با توجه به اینکه $0/8 > 0/7545$ و $0/8 > 0/6694$ است، ضریب همبستگی در سطوح ۱٪ و ۵٪ معنی دار است (اما در سطوح ۲٪ و ۱٪ معنی دار نیست. زیرا $0/8$ از اعداد $0/8329$ و $0/8745$ کوچکتر است.) و به این ترتیب می توان ادعا کرد که با ۹۵٪ اطمینان، C و y با همدیگر همبستگی دارند. لذا بین x و C همبستگی مستقیم (چون r مثبت است) و ناقص وجود دارد. خودآزمایی: وزن و قد ده نفر از همکلاسیهای خود را که به طور تصادفی انتخاب می کنید (مثلاً با قرعه کشی) به عنوان دو متغیر x_i و y_i در نظر بگیرید. دیاگرام پراکنش آنها را رسم کرده، ضریب همبستگی را بین وزن و قد محاسبه نمایید و ببینید آیا معنی دار هست یا خیر؟

چند نکته ضروری در ضریب همبستگی

نکته اول- اگر در یک بررسی بین دو متغیر x و y (مثلاً نمرات آمار و ریاضی دانش آموزان کلاس الف) ضریب همبستگی فرضاً مساوی $0/8$ و در یک بررسی دیگر بین دو متغیر x' و y' (مثلاً نمرات آمار و ریاضی دانش آموزان کلاس ب) ضریب همبستگی مساوی $0/4$ شده باشد، نمی توان ادعا کرد که ارتباط دو متغیر x و y در بررسی اول دو برابر شدیدتر از ارتباط دو متغیر x' و y' در بررسی دوم می باشد. برای این مقایسه از ضریب تعیین، که توان دوم ضریب همبستگی است، استفاده خواهیم کرد. اگر ضریب تعیین را با علامت K نشان دهیم، می توان نوشت: $K = r^2$

مثلاً در مثال بالا : ضریب تعیین در بررسی اول $K_1 = r_1^2 - (0/8)^2 = 0/64$

ضریب تعیین در بررسی دوم $K_2 = r_2^2 - (0/4)^2 = 0/16$

همبستگی در بررسی اول، چهار بار شدیدتر از همبستگی در بررسی دوم است.

$$\frac{K_1}{K_2} = 0/64 \div 0/16 = 4$$

نکته دوم — مشاهده ضریب همبستگی بالا بین دو متغیر x و y ، لزوماً رابطه علیّت بین x و y را اثبات نمی کند بلکه ممکن است عامل یا عوامل دیگری باعث و بانی ارتباط و همبستگی x و y شده باشند.

نکته سوم — اگر مقادیر متغیر x را به صورت $ax \pm b$ و مقادیر y را به صورت $cy \pm d$ تغییر دهیم، r تغییری نخواهد کرد. (البته به شرط آنکه a و c هم علامت باشند.)

۳- معادله خط رگرسیون

اگر بین دو متغیر تصادفی x و y همبستگی مشاهده شد، می توان اندازه و مقدار یکی از این دو متغیر را بر حسب دیگری برآورد کرد. مطالعه این پدیده، موضوع بحث رگرسیون است. واژه رگرسیون^۱ به معنای بازگشت و رجعت است و چون اولین بار دانشمندی به نام گالتون در بررسی رابطه قد فرزندان پسر و قد والدین آنها، متوجه شد که غالباً قد فرزندان پسر به متوسط قد والدین گرایش پیدا می کند، یعنی قد فرزند برمی گردد به متوسط قد والدین. لذا او واژه رگرسیون را در این زمینه به کاربرد و از آن زمان تا به حال کاربرد این کلمه متداول شده است.

برای مثال اگر بین وزن و قد دانش آموزان یک دبیرستان همبستگی وجود داشته باشد، می توان با دانستن وزن فردی، اندازه قد او را و نیز با آگاهی از قد شخص، وزن او را تخمین زد. برای این منظور کافی است روی نقاط دیاگرام پراکنش، خطی رسم کنیم که در مقایسه با سایر خطوط ممکن، کمترین فاصله را از همه نقاط دیاگرام پراکنش داشته باشد. (مناسب ترین خطی باشد که می توان بر داده ها برازنده کرد)^۲ چنین خطی را خط رگرسیون می نامند. مثلاً اگر اندازه وزن (x) و اندازه قد (y) ده دانش آموز را به صورت جدول صفحه بعد در اختیار داشته باشیم :

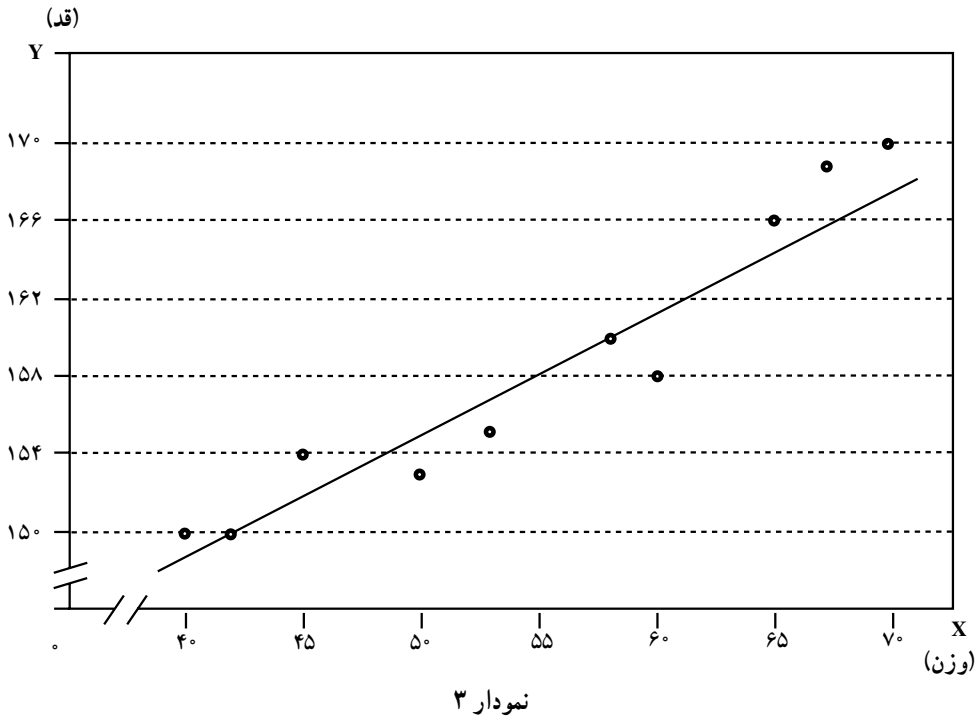
^۱ - Regression

^۲ - مناسب ترین خط، خطی است که مجموع توان های دوم انحرافات نقاط دیاگرام با خط، حداقل باشد. (از دیگر خطوط ممکن کمتر باشد.)

جدول ۵

x_i	۴۰	۴۲	۴۵	۵۰	۵۳	۵۸	۶۰	۶۵	۶۷	۷۰
y_i	۱۵۰	۱۵۰	۱۵۴	۱۵۳	۱۵۵	۱۶۰	۱۵۹	۱۶۶	۱۶۸	۱۷۰

شکل خط رگرسیون روی نقاط دیاگرام پراکنش، می تواند به صورت نمودار ۳ باشد :



برای تخمین مثلاً قد دانش آموزی که ۶۵ کیلوگرم وزن دارد، کافی است از نقطه ۶۵ روی محور x ها، عمودی به موازات محور y ها رسم کنید و از محل تلاقی آن با خط رگرسیون، خطی به موازات محور x ها آن گونه رسم کنید که محور y ها را قطع کند. محل برخورد خط اخیر با محور y ها، برآوردی برای قد دانش آموزی است که وزن او ۶۵ کیلوگرم است.

چون رسم این خط به شیوه بالا نمی تواند خیلی دقیق باشد، به کمک معادله دو مجهولی درجه اول که حالت کلی آن به صورت زیر است، خط رگرسیون را رسم می کنند : به این معادله اصطلاحاً معادله خط رگرسیون گفته می شود.

$$y = ax + b$$

(فرمول ۵)

در این معادله: a شیب خط رگرسیون را مشخص می‌کند و از رابطه $a = \frac{SP_{xy}}{SS_x}$ حاصل

می‌شود. b مقدار ثابتی است که از رابطه $b = \bar{y} - a\bar{x}$ به دست می‌آید و y مقدار برآوردی است که در ازای مقدار معینی از x حاصل می‌شود.

این روش را اصطلاحاً روش کمترین مربعات می‌نامند.

ضمناً برای تعیین مقدار a در معادله بالا می‌توانید از رابطه زیر نیز استفاده کنید.

$$a = r \frac{\sigma_y}{\sigma_x}$$

\downarrow شیب خط \downarrow ضریب همبستگی \times $\frac{\text{انحراف معیار مقادیر } y}{\text{انحراف معیار مقادیر } x}$

برای آزمایش فرمول ۵، مثال ۵ را با کمک فرمول مزبور حل کنید.

مثال ۵- اگر مقادیر دو صفت x و y (مثلاً مقدار تولید و مقدار ضایعات تولید) به صورت زیر

در پنج مشاهده، در اختیار باشد معادله رگرسیون را نوشته، خط رگرسیون را رسم کنید.

x_i	۲	۳	۱	۵	۴
y_i	۸	۱۰	۶	۱۶	۱۵

حل:

$$\bar{x} = \frac{۱۵}{۵} = ۳, \quad \bar{y} = \frac{۵۵}{۵} = ۱۱$$

$$a = \frac{SP_{xy}}{SS_x} = \frac{۲۷}{۱۰} = ۲/۷$$

$$b = \bar{y} - a\bar{x} = ۱۱ - (۲/۷) \times ۳ = ۱۱ - ۸/۷ = ۲/۹$$

$$y = ۲/۷x + ۲/۹ \quad \text{معادله خط رگرسیون}$$

جدول ۶

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
۲	۸	-۱	-۳	۳	۱
۳	۱۰	۰	-۱	۰	۰
۱	۶	-۲	-۵	۱۰	۴
۵	۱۶	+۲	+۵	۱۰	۴
۴	۱۵	+۱	+۴	۴	۱
$\sum x_i =$ ۱۵	$\sum y_i =$ ۵۵			$SP_{xy} = ۲۷$	$SS_x = ۱۰$

ابتدا دیاگرام پراکش را رسم می‌کنیم. سپس خط رگرسیون را روی نقاط دیاگرام برازنده می‌کنیم. برای این منظور، به x دو مقدار دلخواه را نسبت می‌دهیم (مثلاً مقادیر ۲ و ۴ را) و مقادیر متناظر y را به دست می‌آوریم (که در این مثال اعداد $۸/۳$ و $۱۳/۷$ می‌باشند). و آنگاه دو نقطه A و B را با مختصات $A|_{۸/۳}$ و $B|_{۱۳/۷}$ معلوم کرده، به هم وصل می‌کنیم تا خط رگرسیون حاصل شود.

توجه: مقدار a (شیب خط) را از رابطه $a = r \frac{\sigma_y}{\sigma_x}$ نیز می‌توان محاسبه کرد. مثلاً در مثال ۵ خواهیم داشت:

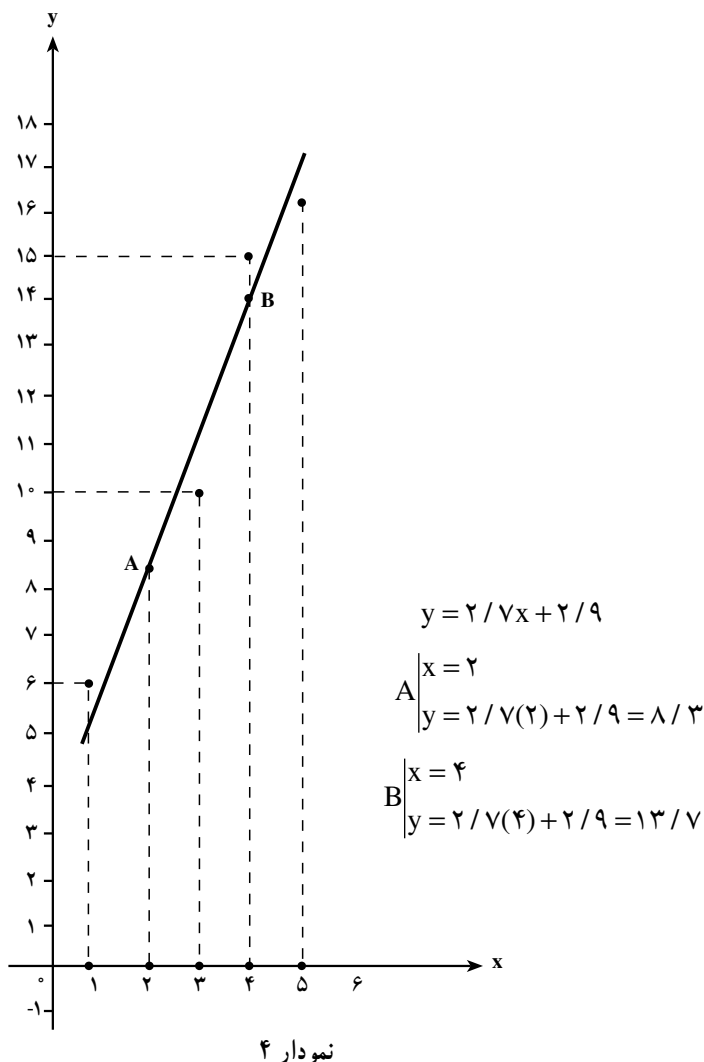
$$r = \frac{SP_{xy}}{\sqrt{SS_x \cdot SS_y}} = \frac{۲۷}{\sqrt{۱۰ \times ۷۶}} = \frac{۲۷}{\sqrt{۷۶۰}} = \frac{۲۷}{۲۷/۶} = ۰/۹۷۸$$

$$\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{\frac{۱۰}{۵}} = \sqrt{۲} = ۱/۴۱$$

$$\sigma_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n}} = \sqrt{\frac{۷۶}{۵}} = \sqrt{۱۵/۲} = ۳/۸۹$$

$$a = ۰/۹۷۸ \times \frac{۳/۸۹}{۱/۴۱} = ۰/۹۷۸ \times ۲/۷۵۸ = ۲/۶۹۷ \approx ۲/۷$$

ملاحظه می‌کنید که در هر دو روش، مقدار a مساوی $۲/۷$ به دست می‌آید.



با این ترتیب می‌توان گفت: خط رگرسیون، وسیله‌ای است که با کمک آن می‌توانیم مقدار یک متغیر را با توجه به مقدار متغیر دیگری که به آن وابسته است، تخمین بزنیم.

تذکر: توجه داشته باشید که نوشتن معادله خط رگرسیون و رسم این خط، زمانی صورت می‌پذیرد که ضریب همبستگی بین x و y معنی‌دار بوده باشد.

کوواریانس^۱ (همپراش)

معیار دیگری که مقدار و شدت و ضعف وابستگی بین دو متغیر را نشان می‌دهد، کوواریانس یا همپراش است. کوواریانس به گونه‌ای محاسبه می‌شود که اگر همبستگی بین x و y خطی مثبت باشد، علامت کوواریانس نیز مثبت است و اگر وابستگی خطی x و y منفی باشد، علامت کوواریانس نیز منفی می‌شود. علت این امر، آن است که اگر تغییرات x و y در یک جهت باشند، هر دو مقدار انحرافات $x_i - \bar{x}$ و $y_i - \bar{y}$ یا مثبت خواهند بود و یا منفی (هم علامت هستند). و اگر تغییرات x و y در دو جهت مخالف باشند، دو مقدار $x_i - \bar{x}$ و $y_i - \bar{y}$ هم علامت نخواهند شد و چون کوواریانس میانگین حاصلضرب این دو انحراف می‌باشد. یعنی:

$$\text{COV}_{(x,y)} = \frac{SP_{xy}}{n} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n} \quad (\text{فرمول ۶})$$

بنابراین، همواره علامت کوواریانس از علامت ضریب همبستگی تبعیت می‌کند (فراموش نکنیم که علامت ضریب همبستگی نیز با علامت SP_{xy} هماهنگی دارد).

مثال ۶- اگر در بررسی همبستگی بین پنج مشاهده از دو متغیر x و y ، معلوم شده باشد که $SP_{xy} = -100$ ، مقدار کوواریانس را معلوم کنید.

$$\text{COV}_{(x,y)} = \frac{SP_{xy}}{n} = \frac{-100}{5} = -20$$

تذکر: اگر x و y دو متغیر مستقل از همدیگر باشند، حتماً کوواریانس آنها مساوی صفر می‌شود، اما عکس این مطلب لزوماً نباید درست باشد، یعنی اگر کوواریانس دو متغیر، تصادفی مساوی صفر شد، نباید تصور کرد که حتماً آن دو متغیر مستقل از هم هستند. زیرا ممکن است دو متغیر مزبور به صورت غیرخطی به یکدیگر وابسته باشند.

اگر x و y دو متغیر تصادفی مستقل از یکدیگر باشند، کوواریانس آنها مساوی صفر خواهد بود. اما اگر کوواریانس x و y مساوی صفر باشد، نباید لزوماً x و y را مستقل از هم دانست.

مثال ۷- اندازه دو متغیر تصادفی x (تعداد کارمندان پنج شعبه بانک) و y (تعداد حسابهای جاری این پنج شعبه) را در جدول زیر با واحد $\frac{1}{1000}$ برای متغیر x و با واحد $\frac{1}{10000}$ برای متغیر y در اختیار دارید. مطلوب است محاسبه ضریب همبستگی با استفاده از فرمول ۴.

نام شعبه	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
شمال	۷	۵	۳۵	۴۹	۲۵
جنوب	۸	۷	۵۶	۶۴	۴۹
شرق	۵	۴	۲۰	۲۵	۱۶
غرب	۴	۲	۸	۱۶	۴
مرکز	۴	۳	۱۲	۱۶	۹
	$\sum x_i = ۲۸$	$\sum y_i = ۲۱$	$\sum x_i y_i = ۱۳۱$	$\sum x_i^2 = ۱۷۰$	$\sum y_i^2 = ۱۰۳$

$$r = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sqrt{\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n} \right] \left[\sum y_i^2 - \frac{(\sum y_i)^2}{n} \right]}} = \frac{۱۳۱ - \frac{۲۸ \times ۲۱}{۵}}{\sqrt{\left[۱۷۰ - \frac{۲۸ \times ۲۸}{۵} \right] \left[۱۰۳ - \frac{۲۱ \times ۲۱}{۵} \right]}}$$

$$r = \frac{۱۳ / ۴}{\sqrt{۱۹۵ / ۲۲۸}} = \frac{۱۳ / ۴}{۱۳ / ۹۷} = ۰ / ۹۵۹ \quad \text{ضریب همبستگی}$$

----- تمرینهای فصل سوم -----

- ۱- مفهوم کلی همبستگی را توضیح دهید.
- ۲- همبستگی را تعریف کنید.
- ۳- علل وجودی همبستگی را برشمارید.
- ۴- انواع همبستگی را نام ببرید و برای هر کدام ضمن ارائه مثال، تعریف مناسبی بیان کنید.
- ۵- درجات همبستگی را از نظر شدت و ضعف همبستگی، تعریف کنید و برای هر کدام مثالی ذکر کنید.
- ۶- اگر اندازه دو صفت x (مقدار تولید) و y (هزینه تولید) در ۶ ماه در یک کارخانه به صورت زیر باشد، با کمک دیاگرام پراکش، نوع و درجه همبستگی را معلوم کنید.

شهریور	مرداد	تیر	خرداد	اردیبهشت	فروردین	ماه
۷۰	۱۰۰	۸۰	۳۰	۵۰	۴۰	مقدار تولید
۸	۱۰	۷	۳	۶	۴	هزینه تولید

- ۷- اگر در پنج نمونه آماری در پنج کارخانه تولیدی، تعداد کارگران فنی (x) و مقدار تولید (y) را به صورت زیر در اختیار داشته باشید، ضریب همبستگی بین این دو متغیر را محاسبه کرده، آزمون معنی دار بودن ضریب همبستگی را انجام دهید.

۱۵	۸	۷	۶	۴	تعداد کارگر
۱۷	۱۷	۱۴	۱۲	۱۰	مقدار تولید

- ۸- تعداد کارمندان ۱۰ شعبه از بانکهای ملی (x) را همراه با تعداد حسابهای جاری این بانکها (y) در اختیار دارید. ضریب تعیین را محاسبه کنید.

نام شعبه	A	B	C	D	E	F	G	H	I	J
تعداد کارمند	۵	۶	۷	۱۰	۴	۱۵	۲۰	۱۲	۱۶	۸
تعداد حساب جاری با واحد $\frac{۱}{۱۰۰}$	۱۲	۱۴	۱۵	۱۸	۱۰	۲۵	۳۰	۲۰	۳۰	۱۶

۹- در جدول زیر، معادله خط رگرسیون y نسبت به x را نوشته، خط رگرسیون را رسم کنید.

(x)	حجم تولید با واحد $\frac{1}{30}$	۵۰	۷۰	۶۰	۴۰	۱۰۰	۹۰	۷۵	۴۵
(y)	حجم ضایعات	۱۲	۱۶	۱۴	۱۰	۲۰	۱۸	۱۷	۱۱

۱۰- اگر $SP_{xy} = -120$ ، $SS_x = 100$ و $SS_y = 144$ باشد، ضریب همبستگی را محاسبه کرده، معنای آن را بیان کنید.

۱۱- در جدول زیر کوواریانس را محاسبه کرده، مفهوم آن را توضیح دهید.

x_i	۲	۳	۴	۵	۶
y_i	-۲	-۴	-۶	-۸	-۱۰

۱۲- برای مطالعه رابطه بین مدت دوام یک خودکار تولیدی و قیمت آن، یک نمونه پنج تایی از مدل‌های مختلف خودکار را انتخاب کرده، مدت دوام را با واحد ساعت کار (x_i) و قیمت آنها را با واحد ۱۰۰ تومان (y_i) به صورت زیر یادداشت کرده‌ایم:

x_i	مدت دوام به ساعت	۱۰	۱۲	۱۴	۲۰	۲۴
y_i	قیمت به صد تومان	۱	۱/۲	۱/۳	۱/۵	۲

مطلوب است:

الف) محاسبه ضریب همبستگی

ب) آزمون معنی دار بودن ضریب همبستگی و تفسیر آن

ج) محاسبه ضریب تعیین

د) در صورت معنی دار بودن r ، تعیین معادله خط رگرسیون

ه) با کمک معادله خط رگرسیون معلوم کنید خودکاری که ۳۰۰ تومان قیمت دارد، چند ساعت باید کار کند؟

۱۳- در یک کارخانه اتومبیل سازی، برای بررسی رابطه بین سرعت اتومبیل (x_i) و طول خط ترمز (y_i) را، در پنج آزمایش مختلف به صورت زیر اندازه گیری کرده‌ایم:

x_i	سرعت اتومبیل به کیلومتر در ساعت	۳۰	۴۰	۵۰	۶۰	۷۰
y_i	طول خط ترمز به متر	۳	۵	۵/۵	۷	۸/۵

معلوم کنید اتومبیلی که با سرعت 100 کیلومتر در ساعت ترمز می کند، چند متر بعد از لحظه ترمز کردن، متوقف خواهد شد؟ (از معادله خط رگرسیون کمک بگیرید.)

۱۴- نمرات دانش آموزان دو کلاس A و B را در دروس آمار و ریاضی طبق جدول زیر در اختیار دارید. ضریب تعیین را در این دو کلاس محاسبه نموده، معلوم کنید شدت ارتباط همبستگی بین نمرات آمار و ریاضی در این دو کلاس چگونه است؟

کلاس	نمرات آمار	۱۴	۱۶	۱۵	۱۷	۱۸
A	نمرات ریاضی	۱۲	۱۴	۱۳	۱۶	۱۵
کلاس	نمرات آمار	۱۷	۱۸	۲۰	۱۶	۱۹
B	نمرات ریاضی	۱۴	۱۴	۱۹	۱۶	۱۷

۱۵- در جدول زیر تعداد غیبتها و تعداد روزهای بیمار شدن ۱۰ دانش آموز کلاس را در طول ترم گذشته مشاهده می کنید.

نام دانش آموز	A	B	C	D	E	F	G	H	I	J
تعداد غیبتهای ترم پیش	۲	۳	۲	۱	۰	۱	۲	۳	۳	۳
تعداد روزهای بیماری	۱	۲	۲	۱	۰	۰	۱	۳	۲	۱

تعداد روزهای بیماری دانش آموزان را x و تعداد غیبتها را y در نظر بگیرید و ضریب همبستگی بین x و y را محاسبه و تفسیر نمایید. از این مثال چه نتیجه ای به دست آوردید؟

----- تستهای چهار گزینه‌ای -----

۱- بین x و y رابطه $y = 2x - 15$ برقرار است. اگر میانگین x برابر با ۱۴ باشد، میانگین y کدام است؟

- ۳۳ (۱) ۲۸ (۲) ۱۶ (۳) ۱۳ (۴)

۲- اگر $SP_{xy} = -9$ ، $SS_x = 225$ ، $SS_y = 144$ باشد، ضریب همبستگی کدام است؟

- ۰/۵ (۱) -۰/۷۵ (۲) ۰/۵ (۳) ۰/۷۵ (۴)

۳- مهندسی برای نظارت کارهای ساختمانی، ۲۵ درصد کل هزینه را دریافت می‌کند. اگر x کل هزینه و y سهم مهندس باشد، ضریب همبستگی x و y کدام است؟

- ۱ (۱) ۰ (۲) ۱ (۳) ۰/۲۵ (۴)

۴- اگر $SP_{xy} = 300$ ، $SS_x = 200$ ، $\bar{x} = 2$ و $\bar{y} = 3$ باشد، معادله خط رگرسیون کدام است؟

- $y = \frac{2}{3}x + 2$ (۱) $y = 1/5x + 2$ (۲) $y = \frac{2}{3}x$ (۳) $y = 1/5x$ (۴)

۵- از نمونه‌ای به حجم $n = 5$ ضریب همبستگی بین دو متغیر $0/8$ است. اگر در جدول معنی‌دار بودن r با احتمال 99% عدد $0/959$ مشاهده شود، در این صورت با احتمال 99% بین دو متغیر همبستگی معنی‌دار ...

- (۱) کامل وجود دارد (۲) معکوس وجود دارد
(۳) ناقص وجود دارد (۴) وجود ندارد

۶- اگر اندازه‌های دو صفت x و y به صورت زیر باشد، ضریب همبستگی کدام است؟

x_i	۱	۲	۳	۴	۵
y_i	-۵	-۴	-۳	-۲	-۱

۰/۵ (۴) -۱ (۳) +۱ (۲) صفر (۱)

۷- فرض کنید $y = \frac{1}{3}x + 3$ معادله خط رگرسیون y نسبت به x باشد در صورتی که $SS_y = 4$

و $SS_x = 1$ باشد، ضریب همبستگی کدام است؟

- $\frac{1}{3}$ (۱) $-\frac{1}{4}$ (۲) $\frac{1}{4}$ (۳) $-\frac{1}{3}$ (۴)

۸- در کدام حالت دو متغیر x و y ، دارای همبستگی مستقیم می‌باشند؟ (r ضریب همبستگی)

است.

$$(1) 1 \geq r \geq 0 \quad (2) r = 0 \quad (3) 1 \geq r \geq 1 \quad (4) 1 \geq r > 0$$

۹- در جدول زیر، تغییرات x و y نشان داده شده است. ضریب همبستگی x و y کدام است؟

x_i	۱	۳	۵	۷
y_i	۱	۱	۱	۱

$$(1) 1 \quad (2) \text{ صفر} \quad (3) -1 \quad (4) 0.97$$

۱۰- اگر x و y دارای همبستگی کامل و معکوس باشند و $SS_x = SS_y$ باشد، معادله خط

رگرسیون y نسبت به x کدام است؟

$$(1) y = -\frac{1}{4}x + b \quad (2) y = \frac{1}{4}x + b$$

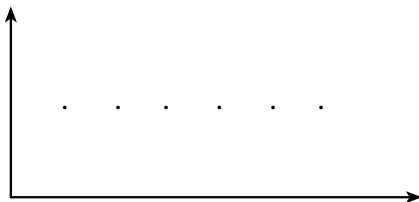
$$(3) y = -x + b \quad (4) y = x + b$$

۱۱- اگر x و y دو متغیر تصادفی مستقل از هم باشند کوواریانس آنها کدام است؟

$$(1) \text{ صفر} \quad (2) 0.5 \quad (3) 0.25 \quad (4) 1$$

۱۲- اگر دیاگرام پراکنش زیر برای دو متغیر x و y به دست آمده باشد، چه رابطه‌ای بین x و y

وجود دارد؟



(۱) همبستگی آنها کامل و مستقیم است.

(۲) همبستگی ندارند.

(۳) همبستگی آنها کامل و معکوس است.

(۴) همبستگی آنها ناقص است.

۱۳- اگر ضریب همبستگی بین x و y مساوی ۰/۲۵ باشد، ضریب تعیین کدام است؟

$$(1) 0.625 \quad (2) 0.5 \quad (3) 0.625 \quad (4) 0.25$$