

فصل ششم: ضریب همبستگی

مدرس: دکتر سمیه عباسی

عضو عیات علمی دانشگاه آزاد اسلامی اصفهان (خوراسگان)

دانشگاه آزاد اسلامی
واحد رصفهان خوراسگان

فصل اول

دانشگاه آزاد اسلامی
واحد رصفهان خوراسگان

همبستگی

وقتی شما صفتهای متعددی از یک جمعیت را مطالعه می‌کنید ممکن است بخواهید بدانید بین این صفتهای رابطه‌ای وجود دارد یا خیر؟ مثلاً می‌دانیم بطور معمول بین قد و وزن افراد در یک جامعه رابطه مستقیم وجود دارد. یعنی اشخاص بلند قد وزن بیشتری نسبت به افراد کوتاه قد دارند. در چنین حالتی می‌گویند بین قد و وزن همبستگی وجود دارد.

تحقیق همبستگی یکی از روش‌های تحقیق توصیفی (غیرآزمایشی) است که رابطه میان متغیرها را براساس هدف تحقیق بررسی می‌کند. همبستگی برای بررسی نوع و میزان رابطه متغیرها استفاده می‌شود.

ضریب همبستگی پیرسون، ضریب همبستگی اسپیرمن از مهمترین روش‌های محاسبه همبستگی میان متغیرها هستند. بطور کلی:

- 1- اگر هر دو متغیر با مقیاس نسبی و پیوسته (در صورت نرمال بودن توزیع داده‌ها) باشند از ضریب همبستگی پیرسون استفاده می‌شود.
- 2- اگر مقیاس کیفی ترتیبی و گسسته (در صورت غیر نرمال بودن داده‌ها) باشند از ضریب همبستگی اسپیرمن استفاده می‌شود.

ضریب همبستگی

ضریب همبستگی شاخصی است که میزان رابطه بین متغیرها را نشان می‌دهد. یک ملاک مناسب برای تعیین همبستگی دو متغیر کمی **ضریب همبستگی پیرسن** است که در جامعه آن را با ρ_{xy} و در نمونه آن را با r_{xy} نشان می‌دهیم. همچنین برای استفاده از این آزمون حتما باید توزیع داده‌ها نرمال باشد. این ضریب میزان همبستگی بین دو متغیر فاصله‌ای یا نسبی را محاسبه کرده مقدار آن بین $+1$ و -1 می‌باشد. اگر مقدار بدست آمده مثبت باشد به معنی این است که تغییرات دو متغیر به طور هم جهت اتفاق می‌افتد یعنی با افزایش در هر متغیر، متغیر دیگر نیز افزایش می‌یابد. اگر مقدار -1 منفی شد یعنی این که دو متغیر در جهت عکس هم عمل می‌کنند یعنی با افزایش مقدار یک متغیر مقادیر متغیر دیگر کاهش می‌یابد و برعکس. اگر مقدار بدست آمده صفر شد نشان می‌دهد که هیچ رابطه‌ای بین دو متغیر وجود ندارد و اگر $+1$ شد همبستگی مثبت کامل و اگر -1 شد همبستگی کامل و منفی است.

$$\rho_{xy} = \frac{\overline{E(xy)} - E(x)E(y)}{\sqrt{[E(x^2) - E^2(x)][E(y^2) - E^2(y)]}}$$

ضریب همبستگی
پیرسن در جامعه

$$r = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sqrt{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \sqrt{\sum_{i=1}^n y_i^2 - n\bar{y}^2}} = \frac{S_{xy}}{\sqrt{S_{xx}} \sqrt{S_{yy}}}$$

r_{xy} برآوردی نقطه‌ای از ρ_{xy} است.

ضریب همبستگی همواره بین منفی یک و یک است.

$$-1 \leq r_{xy} \leq 1$$

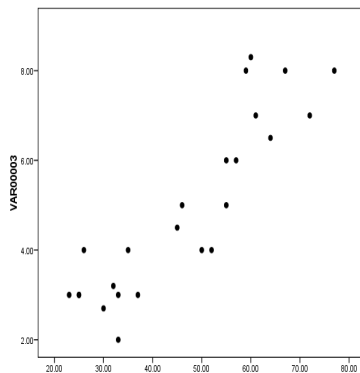
مثلا

همبستگی بین دو متغیر قد و وزن مثبت و مستقیم ولی برعکس همبستگی بین دو متغیر سن و ایمنی بدن، همبستگی معکوس یا منفی است.

نمودار پراکنش

اگر هر يك از متغیرهاي x و y را روی یکی از محورهاي مختصات قرار دهیم و نقاط توأم آنها را در صفحه مشخص نماییم می‌توان تصویر بهتری از رابطه آنها را مشاهده کرد. به این نمودار، نمودار پراکنش می‌گویند.

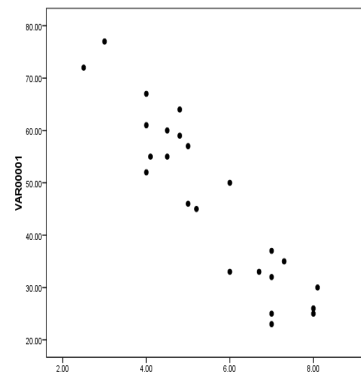
$$r_{xy} \approx +1$$



همبستگی مستقیم

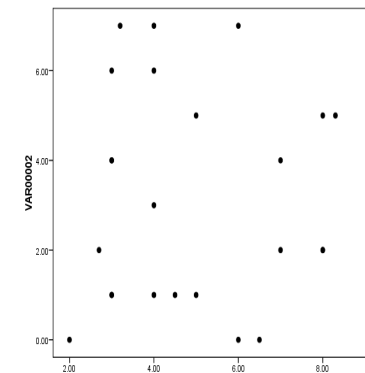
مدرس: دکتر سمیه عباسی

$$r_{xy} \approx -1$$



همبستگی معکوس





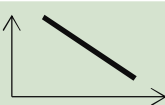
$$r_{xy} \approx 0$$



عدم همبستگی

ضریب همبستگی

بر حسب اینکه مقدار ضریب همبستگی مثبت، منفی یا صفر باشد می توان نتیجه را به صورت زیر بیان کرد:

ضریب همبستگی	نمودار پراکنش	تفسیر
$\rho_{xy} = 1$		همبستگی دو متغیر کامل و مستقیم است
$0 < \rho_{xy} < 1$		همبستگی دو متغیر ناقص و مستقیم است
$\rho_{xy} = 0$		دو متغیر همبستگی ندارند. (ناهمبسته)
$-1 < \rho_{xy} < 0$		همبستگی دو متغیر ناقص و معکوس است
$\rho_{xy} = -1$		همبستگی دو متغیر کامل و معکوس است

واحد رصفهان خوراکان

- تفسیر نتایج ضریب همبستگی

بر اساس یک قاعده کلی بر اساس مقادیر زیر می‌توان درباره میزان همبستگی متغیرها قضاوت کرد. بخاطر داشته باشید همین تفسیر برای مقادیر منفی نیز قابل استفاده است:

ضریب همبستگی	تفسیر
0.00 - 0.19	خیلی اندک و قابل چشم پوشی
0.20 - 0.39	خیلی اندک تا اندک
0.40 - 0.69	متوسط
0.70 - 0.89	زیاد
0.90 - 1.00	خیلی زیاد

با توجه به داده های جدول زیر بررسی کنید آیا بین تعداد سیگار و ساعات خواب افراد سیگاری رابطه وجود دارد؟
(x = تعداد سیگار و y = ساعات خواب است.)

x	y
۱	۸
۲	۶
۳	۷
۴	۶
۵	۳



تمرین

داده های زیر مربوط به هزینه تبلیغات و تعداد فروش می باشد. ضریب همبستگی بین هزینه تبلیغات (x) و تعداد فروش (y) را محاسبه کنید.

ردیف	x	y	x^2	y^2	xy
۱	۰	۳			
۲	۲	۵			
۳	۳	۶			
۴	۴	۸			
۵	۵	۸			
۶	۶	۱۰			
جمع	۲۰	۴۰			

ضریب همبستگی رتبه ای اسپیرمن

اگر محقق، بتواند بپذیرد که در جامعه ای که نمونه ها از آن انتخاب شده اند داده ها دارای توزیع نرمال هستند می تواند از ضریب همبستگی پیرسون استفاده کند وگرنه ضریب همبستگی پیرسون با اشتباه و گمراه کننده همراه خواهد بود.

در چنین مواردی باید ضریب همبستگی رتبه ای اسپیرمن را محاسبه و تفسیر نمود زیرا در محاسبه این ضریب همبستگی به جای مقادیر واقعی داده ها رتبه هر داده در مقایسه با داده های دیگر به کار می رود. به عبارت دیگر داده ها را به ترتیب غیر صعودی مرتب کرده به بزرگترین داده رتبه 1 به داده بلافاصله بعد آن رتبه 2 و و بالاخره به کوچکترین داده بالاترین رتبه نسبت می دهیم.

این ضریب میزان همبستگی رابطه ی میان دو **متغیر ترتیبی** را نشان می دهد و به عبارت دیگر متناظر ناپارامتری ضریب همبستگی پیرسون می باشد.



رابطه ی مربوط به ضریب همبستگی رتبه ای اسپیرمن به صورت زیر تعریف می شود.

$$r_s = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n(n^2 - 1)}$$

در این فرمول r شدت همبستگی n تعداد نمونه و d تفاضل رتبه های داده ها می باشد.

تفاوت ضریب همبستگی رتبه ای اسپیرمن و ضریب همبستگی پیرسون

ضریب همبستگی پیرسون برای محاسبه ی همبستگی دو متغیر فاصله ای یا نسبی به کار برده می شود، ولی ضریب اسپیرمن ، همبستگی موجود بین دو متغیر ترتیبی را نشان می دهد.

به کمک ضریب همبستگی اسپیرمن روابط غیرخطی بررسی می شود در حالیکه ضریب همبستگی پیرسون به منظور بررسی یک رابطه ی خطی بکار برده می شود.

کارایی ضریب همبستگی رتبه ای اسپیرمن کمتر از ضریب همبستگی پیرسون است.

محاسبه ی ضریب همبستگی اسپیرمن ساده تر بوده و نیاز به پیش فرض های کمتری نسبت به ضریب پیرسون دارد.

مثال

نمرات آزمون های زبان انگلیسی و ادبیات فارسی یک گروه از دانشجویان دانشکده ای به شرح زیر است

همبستگی رتبه ای اسپیرمن را محاسبه کنید.

دانشجو	A	B	C	D	E	F	G
نمره زبان انگلیسی	10	8	12	14	4	10	5
نمره ادبیات فارسی	20	15	12	10	15	15	16



• تمرین) در متغیرهای زیر ضریب همبستگی اسپیرمن را محاسبه نمایید؟

X	Y
۲	۲
۴	۴
۵	۱
۲	۳

تمرین) در متغیرهای زیر ضریب همبستگی اسپیرمن را محاسبه نمایید؟

X	Y
15	80
16	69
17	41
18	78
19	53



آزمون ضریب همبستگی

ابتدا فرضیه های آماری را به شرح زیر تشکیل می دهیم:

$$H_0: \rho = 0 \quad ; \quad H_1: \rho \neq 0$$

توجه: فرض صفر " بین متغیر X و متغیر Y همبستگی وجود ندارد."

فرض می کنیم که (X, Y) متغیر تصادفی دو بعدی با توزیع دو متغیر نرمال باشد آماره t را به صورت زیر محاسبه می کنیم.

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

اکنون از جدول t در سطح α و با درجه آزادی $df=n-2$ نقطه بحرانی t را به دست می آوریم اگر آماره آزمون در

ناحیه بحرانی قرار بگیرد فرضیه H_0 رد می شود ($|t| > t_{(\frac{\alpha}{2}, n-2)}$) یعنی در سطح α همبستگی وجود دارد و

اگر آماره آزمون در ناحیه بحرانی قرار نگیرد فرضیه H_0 رد نمی شود یعنی بین X و Y همبستگی وجود ندارد.

دانشگاه آزاد اسلامی

واحد رصفهان خورگان

مثال



بر اساس اطلاعات داده شده و با فرض نرمال بودن توزیع ها بررسی کنید که آیا رابطه همبستگی بین X و Y وجود دارد؟

$$\begin{aligned} S_{xx} &= 698.6 \\ S_{yy} &= 6105.9 \\ S_{xy} &= 2027.7 \\ n &= 25 \\ \alpha &= 0.05 \end{aligned}$$





از توحه سما
پاسگرارم

مدرس: دکتر سمیه عباسی

صرب همبستگی