DOI: 10.48305/jims.v42.i773.0560

مجله دانشکده پزشکی اصفهان — سال چهل و دوم/شماره‌ی ۷۷۲/هفته‌ی اول شهریور ۱۴۰۳

تاریخ دریافت: ۱۴۰۳/۰۱/۲۸ — تاریخ پذیرش: ۱۴۰۳/۰۵/۱۵ — تاریخ چاپ: ۱۴۰۳/۰۶/۰۸

انواع ضرایب همبستگی در پژوهش‌های علوم پزشکی با استفاده از نرم‌افزارهای SPSS و R

بهزاد مهکی[۲،۱]، سمیرا جعفری[۳]، منصور رضایی[۱]، لیلا سلوکی[۳،۲]

مقاله مروری

چکیده

مقدمه: ضریب همبستگی، یک معیار آماری است که میزان همبستگی بین دو متغیر را اندازه می‌گیرد و مقادیر بین ۱- تا ۱+ را اختیار می‌کند که نشان‌دهنده‌ی قدرت و جهت رابطه‌ی خطی بین دو متغیر است. ضرایب همبستگی، نقش مهمی در مطالعات علوم پزشکی ایفا می‌کنند، این ضرایب به محققان کمک می‌کنند تا رابطه‌ی بین متغیرهای مختلف بالینی و پیامدهای سلامت را شناسایی و اندازه‌گیری نمایند.

شرح مقاله: از این‌جهت با توجه به اهمیت محاسبه‌ی ضرایب همبستگی در پژوهش‌های علوم پزشکی، هدف از نوشتار حاضر، معرفی انواع ضرایب همبستگی، مفاهیم و روش‌های ساده و کاربردی آماری به‌منظور بررسی ارتباط بین متغیرها با توجه به ماهیت آن‌ها و نحوه‌ی محاسبه آن‌ها است. انواع ضرایب همبستگی پرکاربرد ازجمله ضریب همبستگی Pearson، Spearman، Kendall، Phi، V Kramer، Summers Delta، Gamma، ضریب توافق و Landa با استفاده از نرم‌افزارهای SPSS و R بررسی خواهد شد. همچنین شرایط استفاده هر یک از ضرایب همبستگی با توجه به پیش‌فرض‌های متفاوت و تفسیر منحصر به فرد هر یک از ضرایب بحث خواهد شد. در نهایت، مثال‌هایی از محاسبه و تفسیر هر یک از ضرایب همبستگی در حوزه‌ی علوم پزشکی ارائه شده است.

نتیجه‌گیری: همبستگی، پرکاربردترین معیار آماری برای ارزیابی روابط بین متغیرها است. با این‌حال باید با احتیاط به کار گرفته شود، در غیر این صورت، می‌تواند منجر به تفسیرها و نتایج اشتباه شود.

واژگان کلیدی: ضریب همبستگی؛ نرم‌افزار؛ تحقیقات زیست‌پزشکی

ارجاع: مهکی بهزاد، جعفری سمیرا، رضایی منصور، سلوکی لیلا. **انواع ضرایب همبستگی در پژوهش‌های علوم پزشکی با استفاده از نرم‌افزارهای SPSS و R.** مجله دانشکده پزشکی اصفهان ۱۴۰۳؛ ۴۲ (۷۷۳): ۵۶۰-۵۷۳.

مقدمه

در بسیاری از مطالعات علوم پزشکی، هدف مطالعه، یافتن روند تغییرات دو متغیر نسبت به یکدیگر است. معمولاً محققانی که مطالعات بالینی یا زیست‌پزشکی را انجام می‌دهند، علاقه‌مند به این موضوع هستند که آیا زمانی که مقدار یک متغیر افزایش یا کاهش می یا بد، ارزش متغیر دیگر ت.ما یل به افزایش/ یا برعکس کاهش تدریجی دارد؟ هدف پژوهشگران اغلب بر رسی وجود ارتباطی بین ۲ متغیر مشاهده شده و تخمین قدرت این رابطه است. روش‌های آماری مختلفی وجود دارد که شدت و جهت این رفتار متقابل متغیرها را بررسی می‌کنند که به ضرایب همبستگی معروف هستند (۱، ۲).

ضرایب همبستگی ابزاری قدرتمند برای بررسی روابط بین متغیرها در تحقیقات پزشکی هستند و به محققان این حوزه کمک می‌کنند تا درک بهتری از این روابط داشته باشند و مدل های مناسب‌تری را ارائه دهند، به عنوان مثال، Nishimura و همکاران ارزیابی کردند که آیا حجم مایع کریستالوئید تزریق شده با میزان نشت مایع بینابینی در طول جراحی مرتبط است یا خیر (۳).

ضریب همبستگی، یک معیار آماری است که برای اندازه‌گیری رابطه‌ی بین دو متغیر به کار می‌رود و نشان می‌دهد که آیا رابطه خطی بین متغیرها (X و Y) وجود دارد یا خیر، اما نشان نمی‌دهد که آیا X باعث Y می‌شود یا برعکس. این ضرایب معمولاً بین مقادیر ۱- تا ۱+

۱ –استاد، گروه آمار زیستی، دانشکده‌ی بهداشت، دانشگاه علوم پزشکی کرمانشاه، کرمانشاه، ایران.

۲ – مرکز تحقیقات علوم داده‌های سلامت، پژوهشکده‌ی سلامت، دانشگاه علوم پزشکی کرمانشاه، کرمانشاه، ایران

۳ - دانشجوی دکترا، گروه آمار زیستی، دانشکده‌ی بهداشت، دانشگاه علوم پزشکی کرمانشاه، کرمانشاه، ایران

نویسنده‌ی مسؤول: لیلا سلوکی: دانشجوی دکترا، گروه آمار زیستی، دانشکده‌ی بهداشت و مرکز تحقیقات علوم داده‌های سلامت، پژوهشکده‌ی سلامت، دانشگاه علوم پزشکی کرمانشاه، کرمانشاه، ایران

Email: l_soloki68@yahoo.com

قرار می‌گیرند. همبستگی ممکن است مثبت، منفی و یا صفر باشد. زمانی همبستگی منفی است که با افزایش یک متغیر، متغیر دیگر کاهش می‌یابد، در همبستگی مثبت وقتی یک متغیر افزایش می‌یابد، متغیر دیگر نیز افزایش می‌یابد و همبستگی صفر نشان می‌دهد که هیچ رابطه‌ی خطی بین دو متغیر ارزیابی‌شده وجود ندارد (۴).

در بسیاری از موارد، محققان با وجود تفاوت‌های عمده، ارتباط (Association) و همبستگی (Correlation) از این اصطلاحات به‌اشتباه به‌جای یکدیگر استفاده می‌کنند، درحالی‌که مفهوم ارتباط رابطه کلی بین متغیرها را بیان می‌کند که این رابطه می‌تواند هم خطی و هم غیرخطی باشد اما همبستگی، رابطه‌ی خطی بین دو متغیر را بیان می‌کند (۵، ۶).

اولین گام در تجزیه و تحلیل همبستگی بین دو متغیر کمی، توجه به نمودار پراکندگی است که قدرت و جهت همبستگی بین متغیرها را نشان می‌دهد، همچنین نشان می‌دهد آیا یک تغییر تدریجی بین مجموعه‌های متغیرها وجود دارد، آیا این تغییر یکنواخت (عمدتاً افزایشی یا کاهشی) است، آیا از یک گرایش متناسب (خطی) پیروی می‌کند، و آیا توزیع اساسی داده‌ها نرمال است (۷، ۸) (شکل ۱).

عدم برر سی این پیش‌نیازها منجر به نتیجه‌گیری ا شتباه می شود، حتی در هنگام کار با نمونه‌های بزرگ، از این جهت با توجه به اهمیت محاسبه‌ی ضرایب همبستگی در پژوهش‌های علوم پزشکی، هدف از نو شتار حا ضر، معرفی انواع ضرایب همبستگی، مفاهیم و روش‌های ساده و کاربردی آماری به‌منظور بررسی ارتباط بین متغیرها با توجه به ماهیت آن‌ها و نحوه‌ی محاسبه‌ی آن‌ها همچنین راهنمایی محققان و پزشکان در ا ستفاده و تفسیر مناسب از ضرایب همبستگی ا ست، لذا انواع ضرایب همبستگی پرکاربرد ازجمله ضریب همبستگی Pearson، Spearman، Kendall، Phi، V Kramer، Summers، Gamma، Delta، ضریب توافق و Landa با ا ستفاده از نرم‌افزارهای SPSS و R برر سی خواهد شد. همچنین شرایط ا ستفاده هر یک از ضرایب همبستگی با توجه به پیش‌فرض‌های متفاوت و تفسیر منحصر به فرد هریک از ضرایب بحث خواهد شد.

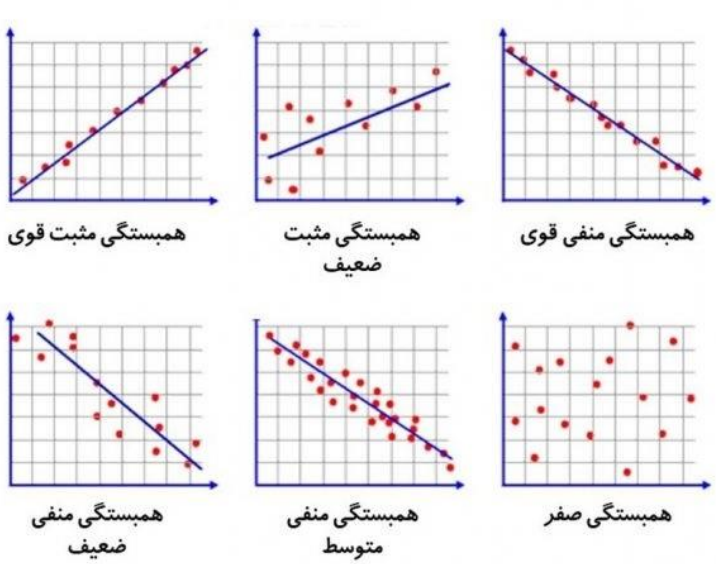
انواع ضرایب همبستگی:

ضریب همبستگی Pearson

ضریب همبستگی Pearson که با نام‌های ضریب همبستگی حاصل‌ضرب-گ شتاور و یا ضریب همبستگی مرتبه صفر Pearson شناخته می‌شود (۹)، میزان همبستگی خطی بین دو متغیر را می سنجد (۱۰). این ضریب تو سط Karl Pearson معرفی شده و برای استفاده از آن، که به‌عنوان رایج‌ترین ضریب همبستگی گزارش شده، نیاز به متغیرهای تصادفی پیو سته با توزیع نرمال ا ست (۱۱). این ضریب به‌منظور تعیین میزان رابطه، نوع و جهت رابطه بین دو متغیر فاصله‌ای یا نسبتی و یا یک متغیر فاصله‌ای و یک متغیر نسبتی به‌کاربرده می شود (۱۲). در مطالعه‌ای که تو سط تقی‌زاده و همکاران در سال ۱۳۹۸ انجام شد، از ضریب همبستگی Pearson برای بررسی ارتباط بین فشارخون با توده‌ی بدنی، سن و امتیاز الگوی تغذیه‌ای استفاده شده است (۱۳).

ضریب همبستگی Pearson برای یک نمونه آماری با n زوج داده (X_i ،Y_i) به‌صورت زیر تعریف می‌شود:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$



شکل ۱: انواع همبستگی و نمودار پراکندگی بین دو متغیر کمی

مقدار این ضـــریب بین ۱ و ۱- تغییر می‌کند که r = ۱ بیانگر رابطه مســـتقیم کامل بین دو متغیر بوده و r = ۱- وجود یک رابطه‌ی معکوس کامل بین دو متغیر را نشان می‌دهد. همچنین زمانی که ضریب همبستگی صفر باشد، نشان‌دهنده‌ی آن است که بین دو متغیر، رابطه‌ی خطی وجود ندارد. در علوم پزشکی، ضریب همبستگی Pearson بالاتر از مقدار ۰/۷ نشـان‌دهنده‌ی ارتباط خیلی قوی، مقادیر بین ۰/۶ تا ۰/۷ حاکی از ارتباط قوی، مقادیر بین ۰/۳ تا ۰/۶ ارتباط نسبتاً خوب و مقادیر پایین‌تر از ۰/۳ نیز نشان‌دهنده‌ی ارتباط ضعیف بین متغیرها است (۱۱).

مثال ۱: فرض کنید سطح قند خون ناشتا و میزان کلسترول HDL را برای ۱۰ بیمار مبتلا به دیابت اندازه گرفته‌ایم. داده‌ها در جدول ۱ آورده شده است.

جدول ۱: میزان میزان کلسترول HDL بر اساس سطح قند خون ناشتا

سطح قند خون ناشتا	میزان کلسترول HDL
۱۱۰	۴۵
۱۲۶	۴۰
۹۰	۳۸
۱۵۰	۳۵
۸۹	۴۰
۱۸۰	۳۰
۱۰۰	۳۹
۹۵	۴۴
۱۰۵	۴۲
۱۴۰	۳۶

High-density Lipoprotein :HDL

هدف محاسبه‌ی ضریب همبستگی Pearson برای این دو متغیر است. بعد از ورود داده‌ها در نرم‌افزار SPSS با استفاده از مسیر زیر می‌توان این ضریب همبستگی را برای این دو متغیر محاسبه کرد:

Analyze--->Correlate--->Bivariate

در پنجره‌ی ظاهر شده، متغیرها را از کادر سمت چپ وارد کادر متغیرها (Variables) کرده و در بخش ضـــرایب همبســـتگی (Correlation Coefficients) ضـــریب Pearson را فعال کرده و در نهایت گزینه‌ی OK را انتخاب می‌کنیم.

در نرم‌افزار R نیز با استفاده از دستورات زیر می‌توان در ابتدا متغیرها را تعریف کرده و مقدار ضریب همبستگی Pearson را محاسبه کرد:

x = c(۱۱۰،۱۲۶،۹۰،۱۵۰،۸۹،۱۸۰،۱۰۰،۹۵،۱۰۵،۱۴۰)

y = c (۴۵،۴۰،۳۸،۳۵،۴۰،۳۰،۳۹،۴۴،۴۲،۳۶)

cor.test (x, y, method = "pearson")

مقدار ضـــریب همبســـتگی Pearson برای دو متغیر X و Y برابر ۰/۷۸- بوده که نشـــان‌دهنده‌ی یک همبســـتگی خیلی قوی و معکوس بین X و Y است.

ضریب همبستگی *Spearman*

ضریب همبستگی Spearman، که به ضریب همبستگی رتبه‌ای (رو) Spearman معروف است، یک ضریب همبستگی بر اساس رتبه است که میزان همبستگی بین دو متغیر در سطح ترتیبی را اندازه‌گیری می‌کند.ضـــریب همبســـتگی Spearman معادل ناپارامتری ضـــریب همبستگی Pearson است، همچنین همبستگی میان رتبه‌های دو متغیر در جامعه‌هایی با حجم نمونه‌ی کم و متغیرهای کمی غیرنرمال را نشـــان می‌دهد. دامنه‌ی ضریب همبستگی Spearman، نظیر ضریب همبستگی Pearson، از ۱- تا ۱+ تغییر می‌کند. هنگامی که تمامی افراد، رتبه‌ی یکسانی در هر دو متغیر داشته باشند، همبستگی رتبه‌ای ۱+ خواهد شد و هنگامی‌که رتبه‌های آنان در یک متغیر، دقیقاً مخالف رتبه‌شان در متغیر دیگر باشد، در این صورت مقدار ضریب Spearman۱- خواهد شد. اما اگر هیچ رابطه‌ای بین رتبه‌بندی‌ها وجود نداشته باشد، مقدار ضریب همبســـتگی Spearman صـــفر خواهد شـــد. در مطالعه‌ای که توســـط Ghaderi و همکاران در ســـال ۲۰۲۳ انجام شـــد، ارتباط رضـــایت از اســـتفاده از ســـمعک و کیفیت زندگی در ســـالمندان کم شـــنوا با ضـــریب همبستگی Spearman سنجیده شد (۱۴).

اگر n زوج داده به صورت (Y_i،X_i) داده شده با شند و رتبه‌ی هر داده را به‌صـــورت (y_i ،x_i) تعریف کنیم، ضـــریب Spearman از طریق فرمول زیر محاسبه می‌گردد.

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

به‌طوری‌که مقدار d_i بیانگر فاصـــله‌ی بین دو رتبه در مشـــاهدات است که از طریق فرمول $d_i = R(x_i) - R(y_i)$ محاسبه می‌گردد و n بیانگر تعداد مشاهدات می‌باشد (۱۵).

همچنین، این ضـــریب را می‌توان به‌صـــورت ضـــریب همبســـتگی Spearman بین داده‌های رتبه‌بندی شـــده تعریف کرد. به‌عنوان مثال، اگر n زوج داده به‌صورت (X_i, Y_i) داده شده باشند، ابتدا رتبه‌ی هر داده را به صورت (x_i, y_i) محاسبه کرده و سپس ضریب همبستگی Spearman را به‌صورت زیر محاسبه می‌کنیم:

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

و به داده‌های تکراری مقدار میانگین رتبه‌ها را اختصاص می‌دهیم.

برای محاســـبه‌ی ضـــریب همبســـتگی Spearman در SPSS، می‌توانید از مراحل زیر استفاده کنید:

Analyze--->Correlate--->Bivariate

در پنجره ظاهر شـــده، متغیرها را از کادر ســـمت چپ وارد کادر متغیرهـا (Variables) کرده و در بخش ضـــرایب همبســـتگی (Correlation Coefficients) گزینه "Spearman"را فعال می‌کنیم.

در صورت نیاز، تنظیمات دیگری مانند محاسبه مقادیر P-value و محاسبه‌ی ماتریس همبستگی نیز اعمال کنید. درنهایت گزینه‌ی OK را انتخاب می‌کنیم.

مثال ۲: فرض کنید که می‌خواهیم ضریب همبستگی Spearman بین دو متغیر شدت درد (بر اساس مقیاس رتبه‌بندی درد) و میزان مصرف مسکن (بر اساس مقیاس رتبه‌بندی) را در ۱۰ بیمار را محاسبه کنیم. داده‌ها در جدول ۲ آورده شده است.

جدول ۲: میزان مصرف مسکن بر اساس شدت درد

شدت درد	میزان مصرف مسکن
۵	۴
۷	۸
۴	۳
۶	۵
۸	۹
۳	۲
۹	۱۰
۶	۷
۷	۶
۴	۵

مقدار ضریب همبستگی Spearman بین دو متغیر «شدت درد» و «میزان مصرف مسکن» به‌صورت زیر می‌باشد:

Spearman's rho: ۰/۹۳

این نتیجه نشان می‌دهد که بین دو متغیر «شدت درد» و «میزان مصرف مسکن» همبستگی مثبت وخیلی قوی وجود دارد. همچنین، مقدار P-value نیز در خروجی نمایش داده شده و می‌توانیم با استفاده از آن، بررسی کنیم که آیا این همبستگی معنی‌دار است یا خیر. برای مثال، ممکن است مقدار P-value برابر با ۰/۰۵ باشد، که نشان می‌دهد همبستگی معنادار است ولی در صورتی‌که این مقدار کمتر از ۰/۰۵ باشد، می‌توانیم با اطمینان بیشتری بگوییم که همبستگی معنادار است.

در نرم‌افزارR، برای محاسبه‌ی ضریب همبستگی Spearman، می‌توان از تابع cor.test() استفاده کرد. برای این کار، داده‌های دو متغیر را به‌عنوان آرایه ها به تابع cor.test() بدهید و در پارامتر method آن را برابر با "Spearman" قرار دهید. به‌عنوان‌مثال:

```
x <- c(۵،۷،۴،۶،۸،۳،۹،۶،۷،۴)
y <- c(۴،۸،۳،۵،۹،۲،۱۰،۷،۶،۵)
result <- cor.test(x, y, method = "spearman")
```

ضریب همبستگی Kendall

روش‌های دیگری مانند تای کندال را می‌توان بجای ضریب همبستگی Spearman استفاده نمود که همانند روش Spearman یک روش ناپارامتری محسوب می‌شود. ضریب همبستگی رتبه‌ای Kendall که به تای کندال مشهور است و با حرف یونانی تای نمایش داده می‌شود. یک آماره‌ی ناپارامتری است که برای سنجش همبستگی آماری میان دو متغیر تصادفی به کار می‌رود.

تاو کندال بسطی از rho Spearman است که شدت همبستگی بین دو متغیر ترتیبی و یا یک متغیر ترتیبی و دیگری متغیر فاصله‌ای در جامعه‌ای با حجم نمونه بزرگ را می‌سنجد. همبستگی Spearman نیز همه‌ی ویژگی‌های Kendall را داراست، با این تفاوت که در جامعه‌هایی با حجم نمونه‌ی پایین به کار می‌رود. تفاوت دیگر این است که در همبستگی Spearman، رتبه را به‌صورت نمره نشان می‌دهیم درحالی‌که در همبستگی Kendall از تعداد توافق‌ها و عدم توافق‌ها در رتبه‌بندی‌ها برای محاسبه‌ی ضریب همبستگی استفاده می‌کنیم. از کاربرد این ضریب همبستگی در علوم پزشکی می‌توان به مقاله‌ای که در سال ۲۰۱۹ توسط He و همکاران انجام شد، اشاره کرد. در این مقاله، از ضریب همبستگی Kendall برای ارتباط بین داده‌های تصویربرداری پرتو مغناطیسی (MRI) و نتایج بیوپسی در بیماران مبتلا به سرطان گردن رحم استفاده شده است. این تحقیق نشان داد که ضریب همبستگی Kendall بین داده‌های MRI از نواحی گردن رحم و نتایج بیوپسی، به خوبی می‌تواند اطلاعات بیشتری را در مورد تشخیص سرطان گردن رحم فراهم کند و به بهبود دقت تشخیص و تصمیم‌گیری برای بیماران کمک نماید (۱۶).

برای محاسبه‌ی تای کندال به‌این‌ترتیب عمل می‌شود: فرض کنید $(x_1, y_1), (x_2, y_2), (x_n, y_n)$ مقادیری از دو متغیر با توزیع توأم X و Y باشند. دو زوج (x_i, y_i) و (x_j, y_j) را سازگار(Concordant) می‌نامیم اگر رتبه‌های دو مشاهده باهم برابر باشند، یعنی هرگاه $x_i > x_j$ و $y_i > y_j$ یا اگر هردوی $x_i < x_j$ و $y_i < y_j$ برقرار باشند. دو زوج ناسازگار (Discordant) نامیده می‌شوند هرگاه $x_i > x_j$ و $y_i < y_j$ یا اگر $x_i < x_j$ و $y_i > y_j$ اگر $x_i = x_j$ یا $y_i = y_j$ زوج نه سازگار و نه ناسازگار نامیده می‌شوند. آنگاه تای کندال به‌صورت زیر محاسبه می‌شود:

$$r = \frac{(Concordant\ pairs) - ((\ Discordant\ pairs))}{\frac{1}{2}n(n-1)}$$

تای کندال همیشه بین ۱+ و ۱- است که این دو مقدار متناظر با تطابق و عدم تطابق کامل است. انتظار می‌رود تای کندال برای دو متغیر مستقل نزدیک به صفر باشد (۱۷).

ضریب تاوی- بی کندال (Kendall's tau-b)

زمانی که سطح سنجش متغیرها ترتیبی- ترتیبی یا ترتیبی- فاصله‌ای باشد از ضریب همبستگی تاوی- بی کندال استفاده می‌کنیم. در مطالعه ای که توسط Vincze و همکاران در سال ۲۰۲۳ انجام شد،

از این ضریب برای برای بررسی ارتباط بین مولکول‌های پروریتوژنیک (مولکول‌هایی که باعث حساسیت پوست و احساس خارش می‌شوند) در پوست بیماران با دارماتومیوزیتیس استفاده شده است (۱۸).

ضریب همبستگی کندال (Kendall's Tau coefficient) در SPSS با دستور زیر قابل‌محاسبه است:

Analyze--->Correlate--->Bivariate

در پنجره‌ی ظاهرشده، متغیرها را از کادر سمت چپ وارد کادر متغیرها (Variables) کرده و در بخش ضرایب همبستگی (Correlation Coefficients) گزینه‌ی "Kendall's tau-b"را فعال می‌کنیم. در صورت نیاز، تنظیمات دیگری مانند محاسبه‌ی مقادیر p-value و محاسبه‌ی ماتریس همبستگی نیز اعمال کنید، در نهایت گزینه‌ی OK را انتخاب می‌کنیم.

مثال ۳: فرض کنید که می‌خواهیم ضریب همبستگی Kendall بین دو متغیر شدت درد بعد از عمل (بر اساس مقیاس رتبه‌بندی از ۱ تا ۱۰) و مقدار مورفین مصرفی (میلی‌گرم) در ۱۵ بیمار را محاسبه کنیم. داده‌ها در جدول ۳ آورده شده است.

جدول ۳: شدت درد بعد از عمل بر اساس میزان مورفین مصرفی شرکت‌کنندگان

شدت درد بعد از عمل	میزان مورفین مصرفی
۷	۱۲
۵	۵
۳	۲
۴	۱۰
۸	۱۵
۶	۸
۲	۴
۹	۱۸
۵	۹
۶	۴
۸	۱۴
۶	۲۰
۳	۳
۷	۱۰
۵	۸

مقدار ضریب همبستگی Kendall بین دو متغیر «شدت درد» و «میزان مورفین مصرفی» به‌صورت زیر می‌باشد:

Kendall's tau-b: ۰/۶۹

این نتیجه نشان می‌دهد که بین دو متغیر «شدت درد» و «میزان مورفین مصرفی» یک همبستگی مثبت و قوی وجود دارد. همچنین، مقدار احتمال یا P-value نیز در خروجی نمایش داده شده و می‌توانیم با استفاده از آن، بررسی کنیم که آیا این همبستگی معنی‌دار است یا خیر. برای مثال، ممکن است مقدار P-value برابر با ۰/۰۵ باشد، که نشان می‌دهد همبستگی معنی‌دار است ولی در صورتی‌که این مقدار کمتر از ۰/۰۵ باشد، می‌توانیم با اطمینان بیشتری بگوییم که همبستگی معنی‌دار است.

در نرم‌افزار R، می‌توانید از تابع cor.test() برای محاسبه‌ی ضریب همبستگی Kendall استفاده کنید. برای این منظور، ابتدا داده‌های خود را به‌عنوان دو بردار جداگانه به تابع می‌دهید. برای مثال، فرض کنید دو بردار XوY دارید که می‌خواهید ضریب همبستگی Kendall آن‌ها را محاسبه کنید. در این صورت، می‌توانید کد زیر را اجرا کنید.

```
X<-C(۷،۵،۳،۴،۸،۶،۲،۹،۵،۴،۸،۶،۳،۷،۵)
Y<-C(۱۲،۵،۲،۱۰،۱۵،۸،۴،۱۸،۹،۶،۱۴،۲۰،۳،۱۰،۸)
Cor.test(x,y,method=”kendall”)
```

ضریب تاوی- سی کندال (Kendall's tau-c)

تاوی- سی کندال یک اندازه‌گیری ناپارامتری ارتباط است که برای ارزیابی قدرت و جهت رابطه بین دو کمیت اندازه‌گیری شده، به‌ویژه زمانی که متغیرها ترتیبی هستند، استفاده می‌شود. موارد کاربرد این آزمون شبیه به تاوی- بی کندال است، با این تفاوت که این ضریب برای تجزیه و تحلیل داده‌ها بر اساس جداول غیر مربع (۳×۴،۴×۵ و ...) مناسب‌تر از tau-b است. ضریب tau-c بر اساس تعداد جفت‌های همخوان و ناسازگار نقاط داده تعریف می‌شود و از ۱- تا ۱ متغیر است که ۱ نشان‌دهنده‌ی توافق کامل، ۱- نشان‌دهنده‌ی عدم توافق کامل، و ۰ نشان‌دهنده عدم ارتباط بین رتبه‌بندی است. در مطالعه‌ای که در سال ۲۰۲۴ توسط Soydan و همکاران انجام شد، ضریب همبستگی تاوی- سی کندال به عنوان یکی از ابزارهای ارزیابی تصاویر MRI برای تشخیص و دسته‌بندی تخریب دیسک نشان داده شده است. استفاده از این ضریب همبستگی برای تعیین ارتباط بین پارامترهای مختلف تصاویر MRI و تخریب دیسک و تشخیص فرایند تخریبی دیسک مفید است. این ابزار به محققان کمک می‌کند تا به‌طور دقیق‌تر و قابل اعتمادتر تصاویر MRI را تحلیل کرده و دیسک‌هایی که تحت تأثیر فرایندهای تخریبی قرار گرفته‌اند را شناسایی کنند (۱۹).

ضریب همبستگی Phi:

اگر بخواهیم همبستگی بین دو متغیر طبقه‌بندی (اسمی) را که در قالب جدول توافقی ۲×۲ قرار می‌گیرند محاسبه نماییم، از آماره ناپارامتری فی (Phi) استفاده می‌کنیم (۲۰). ضریب Phi که آن را با Φ نشان می‌دهند، شدت همبستگی بین دو متغیر را بیان می‌کند. در مطالعه‌ای که توسط Sadiq و همکاران در سال ۲۰۱۹ انجام شد،

ضریب همبستگی Phi برای شناسایی عوامل مرتبط با عمل سزارین استفاده گردید. در این مطالعه، از ضریب همبستگی Phi برای ارتباط بین متغیرهای دسته‌ای استفاده شده تا ارتباطات بین متغیرها در مدل تا حد امکان با دقت بررسی شود (۲۱).

برای محاسبه ضریب Phi ابتدا جدولی را طراحی می‌نماییم که در این جدول تعداد افراد حاضر در گروه‌های مختلف مانند جدول ۴ قرار گیرند.

جدول۴ : جدول فراوانی برای توزیع دو متغیر اسمی دو وجهی

متغیرها	طبقه اول	طبقه دوم	جمع
طبقه اول	a	b	a+b
طبقه دوم	c	d	c+d
جمع	a+c	b+d	n

با توجه به جدول بالا، ضریب همبستگی Phi به‌صورت زیر محاسبه می‌شود:

$$\Phi = \frac{(bc) - (ad)}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$$

مقدار این ضریب بین ۰ تا ۱ تغییر کرده و شامل مقادیر منفی نمی‌باشد. مقادیر همبستگی بالای ۰/۲۵ نشان‌دهنده‌ی ارتباط خیلی قوی، مقادیر بین ۰/۱۵ و ۰/۲۵ ارتباط قوی، مقادیر بین ۰/۱ و ۰/۱۵ ارتباط متوسط و مقادیر بین ۰/۰۵ و ۰/۱ نشان‌دهنده‌ی ارتباط ضعیف بین متغیرها می‌باشد، همچنین مقادیر بین ۰ و ۰/۰۵ حاکی از وجود عدم ارتباط یا ارتباط خیلی کم بین متغیرها است (۱۱).

مثال ۴: فرض کنید هدف محققی محاسبه بررسی میزان همبستگی بین «سیگار کشیدن» و «بروز سرطان ریه» افراد می‌باشد، اطلاعات در جدول ۵ جمع‌آوری شده است.

جدول۵ : جدول توافقی ابتلا به سرطان ریه بر اساس وضعیت سیگاری بودن

متغیرها	سیگاری	غیر سیگاری	جمع
مبتلا به سرطان ریه	۱۵	۵	۲۰
سالم	۳۰	۵۰	۸
کل	۴۵	۵۵	۱۰۰

بعد از ورود اطلاعات افراد در نرم‌افزار SPSS با استفاده از مسیر زیر مقدار این ضریب همبستگی را برای این متغیرهای اسمی محاسبه کرد:

Analyze--->Descriptive Statistics--->Crosstabs

در پنجره ظاهرشده، یکی از متغیرها را به کادر Row(s) و متغیر دیگر را به کادر Column(s) وارد نمایید. سپس گزینه‌ی Statistics را انتخاب کرده و در کادر Nominal، گزینه Phi and Cramer's V را انتخاب نمایید. در ادامه گزینه Continue و سپس OK را انتخاب کنید.

همچنین در نرم‌افزار R با استفاده از دستورات زیر یک جدول ۲×۲ از فراوانی‌های فرض شده با نام data ایجاد کرده و سپس مقدار ضریب همبستگی Phi را با فراخوانی کتابخانه psych بین دو متغیر ابتلا به سرطان ریه و وضعیت سیگاری بودن را محاسبه می‌کنیم:

```
library(psych)
data<-matrix(c(۱۵, ۵, ۳۰, ۵۰), nrow = ۲)
phi(data, digits = ۳)
```

مقدار همبستگی Phi بین دو متغیر مورد نظر برابر با ۰/۳۲ بود که نشان‌دهنده‌ی یک ارتباط خیلی قوی بین دو متغیر ابتلا به سرطان ریه و وضعیت سیگاری بودن افراد می‌باشد.

ضریب همبستگی *V Kramer*

ضریب همبستگی V Kramer نیز همانند ضریب همبستگی Kendall و Spearman، یکی از روش‌های محاسبه‌ی همبستگی بین دو متغیر است. این ضریب برای متغیرهایی که داده‌های آن‌ها به‌صورت جدولی هستند، مناسب است و جایگزینی برای Phi در جداول بزرگ‌تر از جدول ۲×۲ است. این ضریب مانند ضریب Phi زمانی استفاده می‌شود که یکی از متغیرها در سطح اسمی و دیگری در سطح اسمی یا ترتیبی باشد.آزمون V Kramer، پرکاربردترین آزمون برای سنجش میزان رابطه‌ی بین متغیرهای اسمی چند وجهی و همچنین یک متغیر اسمی و یک متغیر ترتیبی است. برای مثال، در مطالعه‌ای که در سال ۲۰۲۰ توسط Shanbehzadeh و Nopour انجام شد، ضریب همبستگی V Kramer برای بررسی ارتباط بین معیارهای تشخیصی مختلف با بیماری کووید-۱۹ مورد استفاده قرار گرفت. این ضریب به عنوان یک معیار آماری برای اندازه‌گیری ارتباط بین دو متغیر دو گانه و دودویی، در این مورد بین معیارهای تشخیصی و بیماری کووید-۱۹، استفاده شد. این استفاده از ضریب همبستگی V Kramer به منظور تعیین مهم‌ترین معیارهای تشخیصی برای بیماری کووید-۱۹ و ایجاد یک سیستم پشتیبانی تصمیم بالینی هوشمند صورت گرفته است (۲۲).

مقدار ضریب همبستگی V Kramer بین صفر تا یک متغیر است و برای مقادیر نزدیک به ۰ به معنی عدم وجود همبستگی و برای مقادیر نزدیک به ۱ به معنی وجود همبستگی کامل است. با این‌حال، مقدار بزرگ‌تر از ۰/۲۵ به‌عنوان یک رابطه‌ی بسیار قوی برای Cramer's V نام‌گذاری شده است. تفسیر Phi و V Kramer براساس مقادیر ضریب همبستگی در جدول ۶ نشان داده شده است.

$$V = \sqrt{\frac{\chi^2}{n(\min(r,c)-1}}$$

جدول ۶: تفسیر Phi و V Kramer بر اساس مقادیر ضریب همبستگی

مقدار Phi و V Kramer	شدت ارتباط
۰/۲۵>	بسیار قوی
۰/۱۵ – ۰/۲۵	قوی
۱۵/۱۰ – ۰/۰	متوسط
۱۰/۰۵ – ۰/۰	ضعیف
<۰/۰۵	بسیار ضعیف یا ندارد

که χ^2 مقدار آماره مربع‌کای بین دو متغیر است، n تعداد کل مشاهدات است و r تعداد دسته‌های متغیر اول و c تعداد دسته‌های متغیر دوم هستند (۲۳).

برای محاسبه ضریب همبستگی V Kramer در SPSS، می‌توانید از مراحل زیر استفاده کنید:

Analyze--->"Descriptive Statistics"--->"Crosstabs"

۲-در پنجره‌ی باز شده، متغیرهای مربوط به دو متغیر مورد نظر را به قسمت "Row(s)" و "Column(s)" اضافه کنید. در قسمت "Statistics"، گزینه‌ی "Chi-square" و "Cramer's V" را انتخاب کنید، در صورت نیاز، تنظیمات دیگری مانند محاسبه‌ی مقادیر P-value و محاسبه‌ی ماتریس همبستگی نیز اعمال کنید و در نهایت گزینه‌ی OK را انتخاب می‌کنیم.

مثال ۵: فرض کنید می‌خواهیم ارتباط بین میزان فعالیت فیزیکی (کم، متوسط و زیاد) و سطح کلسترول خون (مطلوب، مرزی و نامطلوب) را در ۳۰۰ نفر بررسی کنیم. داده‌ها در جدول ۷ نشان داده شده است.

جدول ۷: میزان سطح کلسترول بر اساس فعالیت فیزیکی

میزان فعالیت فیزیکی	سطح کلسترول خون		
	مطلوب	مرزی	نامطلوب
کم	۵۰	۳۰	۱۰
متوسط	۸۰	۶۰	۱۵
زیاد	۴۰	۲۰	۵

بعد از اجرای دستور، SPSS ضریب همبستگی V Kramer بین دو متغیر «سطح کلسترول» و «فعالیت فیزیکی» به‌صورت زیر باشد:

Cramer's V: ۰/۲۱

این نتیجه نشان می‌دهد که بین دو متغیر «سطح کلسترول» و «فعالیت فیزیکی» همبستگی متوسطی وجود دارد.

در نرم‌افزار R با استفاده از دستورات زیر یک جدول ۴×۳ از فراوانی‌های فرض شده بانام data ایجاد کرده و سپس مقدار ضریب همبستگی V Kramerرا با فراخوانی کتابخانه rcompanion بین دو متغیر «سطح کلسترول» و «فعالیت فیزیکی» را محاسبه می‌کنیم:

```
library(rcompanion)
data=matrix(c(۵۰،۸۰،۴۰،۳۰،۶۰،۲۰،۱۰،۱۵, ۵),nrow=۳)
cramerV(data)
```

ضریب همبستگی *Summers Delta*

ضریب همبستگی Summers Delta که به‌اختصار Samers' d خوانده می شود در سال ۱۹۶۲ تو سط Somers معرفی و پی شنهاد شد (۲۴). این ضریب همبستگی، معیاری برای سنجش میزان توافق بین دو متغیر ترتیبی بوده که مقدار آن بین ۱ تا ۱- تغییر می‌کند. چنانچه که همه‌ی جفت‌ها باهم توافق داشته باشند مقدار این ضریب همبستگی برابر ۱ و اگر هیچ‌کدام از جفت‌ها تطابق یا توافقی نداشته باشند، مقدار برابر ۱- می‌شود (۲۵). در مطالعه‌ای که توسط Ciuhu و همکاران در سال ۲۰۱۶ انجام شده، از ضریب همبستگی دلتای سامر برای بررسی ارتباط بین اعلام خبر بد، تطبیق با بیماری سرطان و اختلالات خلق روانی مرتبط استفاده شده است. این ضریب همبستگی برای محاسبه میزان رابطه بین این متغیرها به کار گرفته شده است (۲۶).

فرض کنید $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ یک مجموعه از مشاهدات بردارهای مستقل X و Y با شند. برای محا سبه‌ی ضریب همبستگی Summers Delta برای این بردارها، ابتدا باید ضریب همبستگی رتبه‌ای Kendall را با استفاده از فرمول زیر محاسبه کرد:

$$\tau = \frac{N_C - N_D}{n(n-1)/2}$$

که NC نشان‌دهنده‌ی تعداد جفت های ساز گار و ND نشان‌دهنده‌ی تعداد جفت‌های ناسازگار است. ضریب همبستگی Summers Delta با فرمول زیر به دست می‌آید:

$$D_{XY} = \tau(X،Y)/\tau(Y،X)$$

در مواردی که متغیر X یک متغیر مستقل گسسته و Y متغیر پیامد دوحالته باشد ضریب همبستگی Summers Delta با فرمول زیر محاسبه می‌شود:

$$D_{XY} = \frac{N_C - N_D}{N_C + N_D + N_T}$$

که در فرمول بالا NT نشان‌دهنده‌ی تعداد جفت‌های گره‌خورده روی متغیر X می‌باشد (۲۵).

مثال ۶: به‌عنوان مثال فرض کنید که محققی می‌خواهد ارتباط بین سرطان ریه (مرحله‌ی۱ تا ۴) و وضعیت عملکرد ریوی (بد، متوسط، خوب) را در ۱۰۰ بیمار مورد بررسی قرار دهد. متغیر سرطان ریه به‌صورت یک متغیر ترتیبی با مرحله‌ی ۱ تا ۴ کدگذاری شده و متغیر وضعیت عملکرد ریوی نیز به صورت بد (۱)، متو سط (۲)، خوب (۳) رتبه‌بندی شده است. جدول فراوانی آن‌ها در جدول ۸ نشان داده شده است:

جدول ۸: میزان وضعیت عملکرد ریوی بر اساس مرحله‌ی سرطان ریه

مرحله سرطان ریه	وضعیت عملکردی ریوی		
	بد	متوسط	خوب
۱	۰	۱۵	۱۵
۲	۵	۲۰	۱۰
۳	۱۰	۱۰	۵
۴	۷	۳	۰

در نرم‌افزار SPSS بعد از واردکردن داده‌ها، با مسیر زیر می‌توان مقدار ضریب همبستگی دلتای سامر را برای متغیرهای موردنظر محاسبه کرد:

Analyze--->Descriptive Statistics--->Crosstabs

در پنجره‌ی ظاهر شده، یکی از متغیرها را به کادر Row(s) و متغیر دیگر را به کادر Column(s) وارد نمایید. سپس گزینه‌ی Statistics را انتخاب کرده و در کادر Ordinal، گزینه‌ی Samers' d را انتخاب نمایید. در ادامه گزینه‌ی Continue و سپس OK را انتخاب کنید.

در نرم‌افزار R ابتدا دو متغیر را تعریف کرده (به‌عنوان‌مثال X و Y)، سپس با استفاده از کتابخانه DescTools و دستور زیر مقدار ضریب همبستگی دلتای سامر قابل‌محاسبه خواهد بود:

Somers Delta(x, y = NULL, direction = c("row", "column"))

در این مثال مقدار ضریب همبستگی برابر ۰/۵۱ به دست می‌آید که نشان‌دهنده‌ی یک همبستگی مثبت بین میزان وضعیت عملکرد ریوی و مرحله سرطان ریه است.

ضریب همبستگی *Gamma*

ضریب همبستگی Gamma که با نام گامای گودمن و کروسکال نیز شناخته می شود برای بررسی میزان ارتباط بین دو متغیر ترتیبی به کار می‌رود و هنگامی که هر دو متغیر در سطح ترتیبی اندازه‌گیری می‌شوند، قدرت ارتباط داده های جدول‌بندی شده متقاطع را اندازه‌گیری می‌کند (۸). این ضریب همبستگی به افتخار Goodman و Kruskal که آن را در یک سری مقالات از سال ۱۹۵۴ تا ۱۹۷۲ ارائه کردند، نام‌گذاری شده است (۳۰-۲۷).

در مطالعه‌ای که توسط Fang و همکاران در سال ۲۰۱۹ انجام شده، محققان از ضریب همبستگی Gamma برای بررسی ارتباط بین عوامل مختلف و رضایت بیماران استفاده کردند. در این مطالعه، پژوهشگران ارتباط بین عوامل مختلف مانند زمان انتظار، کیفیت خدمات و راحتی بیماران با رضایت بیماران را با استفاده از ضریب همبستگی گاما بررسی کرده‌اند. این روش آماری به آن‌ها کمک کرد تا ارتباطات میان عوامل مختلف را درک کرده و مهم‌ترین عامل مؤثر در رضایت بیماران را شناسایی کنند (۳۱).

مقدار Gamma بین ۱- و ۱ تغییر می‌کند (۳۲) و با استفاده از رابطه‌ی زیر محاسبه می‌شود:

$$G = \frac{N_C - N_D}{N_C + N_D}$$

همان‌گونه که در رابطه بالا مشخص است این ضریب بسیار به ضریب همبستگی دلتای سامر شبیه بوده با این تفاوت که هیچ تعدیلی برای جفت‌های گره خورده اعمال نمی‌شود.

اگر مقدار ضریب Gamma مثبت باشد، نشان‌دهنده‌ی آن است که تعداد جفت‌های سازگار بیش از جفت‌های ناسازگار است و بنابراین بین دو متغیر رابطه مستقیم وجود دارد. به همین ترتیب گامای منفی نشان‌دهنده‌ی وجود رابطه‌ی معکوس بین دو متغیر است. اگر تعداد زوج‌های هماهنگ و ناهماهنگ برابر باشند، ضریب Gamma برابر صفر خواهد بود. در جداول توافقی دوطرفه‌ی Gamma، صفر نشان‌دهنده‌ی استقلال متغیرها می‌باشد ولی در جداول بیش از ۲×۲ اگر مقدار Gamma صفر شود لزوماً به معنای مستقل بودن مشاهدات نیست.

مثال ۷: محققی معتقد است که بین مرحله‌ی سرطان پروستات و شدت درد رابطه وجود دارد. متغیر سرطان پروستات به صورت یک متغیر ترتیبی مرحله‌ی ۱ تا ۳ کدگذاری شده و متغیر شدت درد نیز به صورت خفیف (۱)، متوسط (۲)، شدید (۳) رتبه‌بندی شده است. نتایج در جدول ۹ نشان داده شده است.

جدول ۹: شدت درد بر اساس مرحله‌ی سرطان پروستات

مرحله سرطان پروستات	شدت درد		
	خفیف	متوسط	شدید
۱	۳۰	۷۰	۱۰۰
۲	۷۰	۷۰	۷۰
۳	۱۰۰	۵۰	۵۰

در نرم‌افزار SPSS بعد از واردکردن داده‌ها، با مسیر زیر می‌توان مقدار ضریب همبستگی Gamma را برای متغیرهای مورد نظر محاسبه کرد:

Analyze--->Descriptive Statistics--->Crosstabs

در پنجره‌ی ظاهر شده، یکی از متغیرها را به کادر Row(s) و متغیر دیگر را به کادر Column(s) وارد نمایید. سپس گزینه‌ی Statistics را انتخاب کرده و در کادر Ordinal، گزینه‌ی Gamma را انتخاب نمایید. در ادامه، گزینه‌ی Continue و سپس OK را انتخاب نمایید.

در نرم‌افزار R ابتدا دو متغیر را تعریف کرده (به‌عنوان‌مثال X و Y)، سپس با استفاده از کتابخانه GoodmanKruskalGamma و دستور زیر مقدار ضریب همبستگی Gamma قابل‌محاسبه خواهد بود:

Goodman KruskalGamma(x, y)

در این مثال مقدار ضریب همبستگی برابر ۰/۳۸ به دست می‌آید که نشان‌دهنده‌ی یک همبستگی ضعیف بین شدت درد و مرحله‌ی سرطان پروستات است.

ضریب همبستگی توافق:

ضریب توافق (C) که به ضریب توافق Pearson معروف است، نوعی تقریب اسمی برای ضریب همبستگی Pearson به‌حساب می‌آید. این ضریب برای اندازه‌گیری میزان هم‌خوانی بین دو متغیر کیفی، که هریک از آن‌ها ممکن است دارای بیش از دوطبقه باشند، به کار می‌روند این ضریب مبتنی بر مجذور کای می‌باشد. در مطالعه‌ای که توسط Batchelder و همکاران در سال ۲۰۲۱ انجام شده، ضریب همبستگی توافقی برای بررسی ارتباط بین تشخیص‌های سلامت روان و مصرف مواد مخدر با جنسیت و گرایش جنسی در نمونه‌ای از مراکز بهداشت جامعه استفاده شده است. این ابزار آماری به محققان کمک کرد تا ارتباطات میان متغیرها را در نظر گرفته و به دقت تر بررسی شود که آیا این ارتباطات تصادفی هستند یا واقعاً بر اساس جنسیت و گرایش جنسی وجود دارند (۳۳).

این معیار بین ۰ تا ۱ قرار می‌گیرد، که مقدار ۰ به معنی عدم وجود هرگونه رابطه بین دو متغیر و مقدار ۱ به معنی وجود رابطه‌ی کامل بین دو متغیر است. مقدار بیشتر از ۰/۵ نشان‌دهنده‌ی وجود رابطه‌ی معنی‌دار بین دو متغیر است. هرچه مقدار این ضریب بیشتر باشد، نشان می‌دهد که درجه‌ی وابستگی صفات نسبت به هم بیشتر است.

$$\text{Contingency Coefficient} = \text{sqrt}\ (\chi^2 / (n + \chi^2))$$

که در این فرمول، χ^2 برابر با آماره مربع‌کای است و n برابر با تعداد کل داده‌ها است.

در نرم‌افزار SPSS بعد از واردکردن داده‌ها، با مسیر زیر می‌توان مقدار ضریب همبستگی توافق را برای متغیرهای موردنظر محاسبه کرد:

Analyze--->Descriptive Statistics--->Crosstabs

در پنجره ظاهرشده، یکی از متغیرها را به کادر Row(s) و متغیر دیگر را به کادر Column(s) وارد نمایید. سپس گزینه‌ی Statistics را انتخاب کرده و در قسمت Statistics، گزینه‌ی Contingency Coefficient را انتخاب نمایید. در ادامه گزینه‌ی Continue و سپس OK را انتخاب کنید.

مثال ۸: فرض کنید که در یک مطالعه، تعداد ۳۰۰ نفر، برای بررسی ارتباط بین سیگاری بودن و ابتلا به عفونت ریه به دو دسته عفونت ریه دارد و ندارد تقسیم شده‌اند. در این صورت، می‌توان جدول متقابل را مطابق جدول ۱۰ ایجاد کرد.

جدول ۱۰: ابتلا به عفونت ریه بر اساس وضعیت سیگاری بودن

عفونت ریه	وضعیت سیگاری بودن		
	سیگاری نیست	سیگاری هست	جمع
ندارد	۵۰	۱۰۰	۱۵۰
دارد	۸۰	۷۰	۱۵۰
جمع	۱۳۰	۱۷۰	۳۰۰

سپس، با محاسبه‌ی آماره‌ی مربع‌کای و تعداد مشاهدات، می‌توان ضریب همبستگی توافق را محاسبه کرد:

$$\chi^2 = ۱۱/۴۶$$

$$n = ۳۰۰$$

بنابراین، ضریب همبستگی توافق بین وضعیت سیگاری بودن و عفونت ریه برابر با ۰/۱۹ است. این نتیجه نشان می‌دهد که وجود ارتباط ضعیفی بین وضعیت سیگاری بودن و عفونت ریه وجود دارد.

در نرم‌افزار R، از کتابخانه stats برای محاسبه ضریب همبستگی توافق استفاده می‌شود. برای محاسبه این ضریب، ابتدا باید جدول متقابل (Contingency table) دو متغیر دسته‌ای را ایجاد کرد. سپس با استفاده از تابع chisq.test()، می‌توان آماره‌ی کای دو و ضریب همبستگی توافق را مطابق فرمول بالا محاسبه کرد.

```
lung <- matrix(c(20, 30, 40, 60), nrow = 2, dimnames =
list(c("No", "Yes"), c("Non-smoker", "Smoker")))
result <- chisq.test(lung)
> result$statistic
X-squared
 11.41629
C<-sqrt(result$statistic / (result$statistic + 300))
X-squared
0.1914661
```

ضریب همبستگی Landa (لامبدا)

یکی دیگر از ضرایب همبستگی که برای دو متغیر اسمی و یا یکی اسمی و یکی رتبه‌ای استفاده می‌شود، ضریب همبستگی Landa می‌باشد که به کاهش نسبی خطا در پیش‌بینی متغیر وابسته برحسب متغیر مستقل اشاره دارد. در مطالعه‌ای که در سال ۲۰۱۸ توسط Dimitriu و همکاران انجام شد، از ضریب همبستگی Landa برای بررسی ارتباط بین متغیرهای مختلف استفاده گردید. این ابزار آماری برای اندازه‌گیری ارتباط خطی بین دو متغیر کمک کننده است و در این مقاله نیز برای بررسی ارتباط بین بارداری در دوران نوجوانی و مشکلات مرتبط با آن از این ابزار آماری استفاده شده است (۳۴).

ضریب Landa۱ نشان‌دهنده‌ی آن است که متغیر مستقل به‌طور کامل متغیر وابسته را پیش‌بینی می‌کند و ضریب Landa صفر نشان‌دهنده‌ی آن است که متغیر مستقل هیچ کمکی به‌پیش‌بینی متغیر وابسته نمی‌کند.

با استفاده از فرمول زیر می‌توان این ضریب همبستگی را محاسبه کرد:

$$\lambda = \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1}$$

که در آن ε_1 اشتباه گروه‌بندی در موقعیتی که متغیر اول متغیر مستقل باشد و ε_2 اشتباه گروه‌بندی در موقعیتی که متغیر دوم مستقل است می‌باشد. با توجه به اینکه کدام متغیر را متغیر مستقل در نظر گیریم دو مقدار Landa به دست می‌آید که در نرم‌افزار SPSS برای هر دو جای‌گشت ممکن این آماره مقدار

Landa را محاسبه می‌کند. اگر محاسبه‌ی دو Landa بسته به اینکه کدام متغیر را مستقل و کدام را وابسته در نظر بگیریم متفاوت باشد به آن لاندای نامتقارن می‌گویند و می‌توان از میانگین آن‌ها به‌عنوان معیاری مناسب استفاده کرد. اما در حالت متقارن بسته به اینکه کدام متغیر را وابسته و کدام را مستقل قرار دهیم تفاوتی نمی‌کند، زمانی که دو متغیر مستقل باشند مقدار ضریب همبستگی Landa برابر صفر است اما اگر ضریب همبستگی Landa صفر باشد نمی‌توان گفت دو متغیر مستقل هستند.

مثال ۹: فرض کنید می‌خواهیم رابطه‌ی بین شدت درد زایمان (خفیف، متوسط، شدید) و میزان رضایتمندی مادر از زایمان (کم، متوسط، زیاد) را در ۱۰۰ مادر مورد بررسی قرار دهیم. اطلاعات در جدول ۱۱ آورده شده است.

جدول ۱۱: میزان رضایتمندی مادر از زایمان بر اساس شدت درد زایمان

شدت درد زایمان	میزان رضایتمندی مادر			
	کم	متوسط	زیاد	جمع
خفیف	۰	۵	۳۰	۳۵
متوسط	۵	۲۵	۱۵	۴۵
شدید	۱۰	۱۰	۰	۲۰
جمع	۱۵	۴۰	۴۵	۱۰۰

بعد از ورود اطلاعات افراد در نرم‌افزار SPSS با استفاده از مسیر زیر مقدار این ضریب همبستگی را برای این متغیرها محاسبه کرد:

Analyze > Descriptive Statistics > Crosstabs

دو متغیر را در سطر و ستون وارد کرده و سپس در قسمت Statistics گزینه‌ی Lambda را فعال کنید. مقدار این ضریب همبستگی زمانی که متغیر میزان رضایتمندی مادر را به‌عنوان متغیر وابسته یا مستقل در نظر بگیریم برابر ۰/۳۶ به دست آمده است.

در نرم‌افزار R نیز با استفاده از کتابخانه DescTools و دستور

(Lambda(X$Y , direction="row", conf.level =.95

می‌توان ضریب همبستگی Landa را برای دو متغیر X و Y محاسبه کرد.

بحث و نتیجه‌گیری

همبستگی، پرکاربردترین معیار آماری برای ارزیابی روابط بین متغیرها است. با این‌حال باید با احتیاط به کار گرفته شود، در غیر این صورت، می‌تواند منجر به تفسیرها و نتایج اشتباه شود. ضریب همبستگی روابط میان دو متغیر را به ما نشان می‌دهد. یعنی با توجه به این ضریب می‌توانیم بفهمیم که آیا دو متغیر باهم ارتباط دارند یا نه. اگر ارتباطی دارند جهت این ارتباط چگونه است، یعنی با تغییر مقدار یک متغیر، متغیر دیگر تمایل به تغییر در چه جهتی دارد.

بحث ضریب همبستگی بحث بسیار گسترده‌ای بوده که در زمینه‌های مختلف ازجمله پزشکی، حوزه‌ی سلامت، دنیای کسب و کار و ... کاربرد زیادی دارد. همچنین می‌توان به کاربرد این ضریب در زندگی واقعی اشاره کرد، مانند: کاهش درآمد و مخارج، عرضه و تقاضا، غیبت و نمرات و غیره.

بسته به هدفی که دارید و جامعه‌ی آماری که مدنظر شما است، شیوه‌های متعددی برای محاسبه ضریب همبستگی وجود دارد که در این مقاله ۱۱ مورد از رایج‌ترین روش‌ها ارائه شد. جهت مراجعه‌ی آسان‌تر به مطالب به عنوان ضمیمه خلاصه‌ای از مطالب در جدول ۱۲ آمده است. برای پژوهش‌های آتی توصیه می‌شود از روش‌های جدیدتر استفاده شده و همانند این مقاله مفاهیم و کاربرد این روش‌ها توضیح داده شود. همچنین در این مقاله برای درک بهتر مفاهیم و کاربرد این روش‌ها در پایان هر قسمت یک مثال در نرم‌افزارهای آماری SPSS و R با توجه به سادگی محیط کاربری و عمومیت بیشتر این دو نرم‌افزار برای محققین حوزه‌ی علوم پزشکی ارائه گردید، پیشنهاد می شود در پژوهش‌های آینده از نرم‌افزارهای دیگری همانند پایتون به دلیل پشتیبانی از کتابخانه‌ها و زبان برنامه‌نویسی عمومی برای محاسبه ضرایب همبستگی استفاده شود.

جدول ۱۲: اطلاعات مربوط به انواع ضرایب همبستگی

مقیاس	اسمی		رتبه‌ای		فاصله‌ای	
	ضریب همبستگی	فرمول	ضریب همبستگی	فرمول	ضریب همبستگی	فرمول
اسمی	(φ)Phi	$\Phi = \frac{(bc)-(ad)}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$	V Kramer	$V=\sqrt{\frac{\chi^2}{n(\min(r,c)-1}}$		
	V Kramer	$V=\sqrt{\frac{\chi^2}{n(\min(r,c)-1}}$				
	Landa	$\lambda=\frac{\varepsilon_1-\varepsilon_2}{\varepsilon_1}$	Landa	$\lambda=\frac{\varepsilon_1-\varepsilon_2}{\varepsilon_1}$		
	ضریب توافق	$C\ C = \mathrm{sqrt}\ (\chi^2 / (n + \chi^2))$				
رتبه‌ای			Spearman	$r_s=1-\frac{6\sum d_i^2}{n(n^2-1)}$	Spearman	$r_s=1-\frac{6\sum d_i^2}{n(n^2-1)}$
			Kendall	$\tau=\frac{N_C-N_D}{n(n-1)/2}$	Kendall	$\tau=\frac{N_C-N_D}{n(n-1)/2}$
			Summers	$D_{XY}=\frac{N_C-N_D}{N_C+N_D+N_T}$		
			Delta			
			Gamma	$G=\frac{N_C-N_D}{N_C+N_D}$		
فاصله ای					Pearson	$r=\frac{\sum_{i=1}^n(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^n(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^n(Y_i-\bar{Y})^2}}$

تشکر و قدردانی

بدین وســـیله از تمامی کســـانی ک ما را در انجام این پژوهش یاری رساندند سپاسگزاری می‌گردد.

References

1. Liu J, Tang W, Chen G, Lu Y, Feng C, Tu XM. Correlation and agreement: overview and clarification of competing concepts and measures. Shanghai Arch Psychiatry 2016; 28(2): 115-20.
2. Pal M, Bharati P, Pal M, Bharati P. Introduction to correlation and linear regression analysis.[online 2019] Available from: https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_multivariable/bs704_multivariable5.html
3. Nishimura A, Tabuchi Y, Kikuchi M, Masuda R, Goto K, Iijima T. The amount of fluid given during surgery that leaks into the interstitium correlates with infused fluid volume and varies widely between patients. Anesth Analg 2016; 123(4): 925-32.
4. Senthilnathan S. Usefulness of correlation analysis. [July 9 2019]. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3416918.
5. Altman N, Krzywinski M. Association, correlation and causation. Nat Methods 2015;12(10): 899-900.
6. Rezaee Z, Aliabadi S, Dorestani A, Rezaee NJ. Application of time series models in business research: correlation, association, causation. Sustainability 2020; 12(12): 4833.
7. Miot HA. Assessing normality of data in clinical and experimental trials.[Apr-Jun 2017]. Available from: https://www.scielo.br/j/jvb/a/FPW5hwZ6DTH4gvj5mJYpt6B/?lang=en.
8. Scott M, Flaherty D, Currall J. Statistics: are we related? Journal of Small Animal Practice. 2013; 54(3): 124-8.
9. Pearson K, Lee A. On the laws of inheritance in man: I. Inheritance of physical characters. Biometrika. 1903; 2(4): 357-462.
10. Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. Anesth Analg 2018; 126(5): 1763-8.
11. Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med 2018; 18(3): 91-3.
12. Sedgwick P. Pearson's correlation coefficient. BMJ2012; 345: e4483.
13. Taghizadeh S, Abbasalizad Farhangi M, Poorali F. Correlation between blood pressure, body mass index, life style and dietary habits in children and adolescents aged 6 to 18 years in Tabriz, Iran [in Persian]. J Shahid Sadoughi Univ Med Sci 2020; 28(2): 2399-410.
14. Ghaderi R, Mahmoodi-Shan GR, Taziki MH, Behnampour N. Association between satisfaction with hearing aid use and quality of life among elderly individuals with hearing loss. J Res Dev Nurs Midw 2023; 20(2): 30-3.
15. Hauke J, Kossowski T. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. Quaestiones geographicae 2011; 30(2): 87-93.
16. He YL, Li Y, Lin CY, Qi YF, Wang X, Zhou HL, et al. Three-dimensional turbo-spin-echo amide proton transfer-weighted MRI for cervical cancer: A preliminary study. J Magn Reson Imaging 2019; 50(4): 1318-25 .
17. Xiao C, Ye J, Esteves RM, Rong C. Using Spearman's correlation coefficients for exploratory data analysis on big dataset. Concurrency and Computation: Practice and Experience 2016; 28(14): 3866-78.
18. Vincze A, Herczeg-Lisztes E, Szabó K, Béldi TG, Nagy-Vincze M, Pór Á, et al. Pruritogenic molecules in the skin of patients with dermatomyositis. Front Med (Lausanne) 2023; 10: 1168359.
19. Soydan Z, Bayramoglu E, Urut DU, Iplikcioglu AC, Sen C. Tracing the disc: The novel qualitative morphometric MRI based disc degeneration classification system. JOR Spine 2024; 7(1): e1321.
20. Field A. Discovering statistics using IBM SPSS statistics. 4th Ed. New York, NY: SAGE Publications Ltd; 2013.
21. Sadiq M, Mehmood T, Aslam M. Identifying the factors associated with cesarean section modeled with categorical correlation coefficients in partial least squares. PLoS One 2019; 14(7): e0219427.
22. Shanbehzadeh M, Nopour R. Determination of the most important diagnostic criteria for COVID-19: A step forward to design an intelligent clinical decision support system. J Adv Med Biomed Res 2021; 29(134): 176-82.
23. Piaw CY. Mastering research statistics. 1st Ed. Malaysia: McGraw Hill Education; 2013.
24. Somers RH. A new asymmetric measure of association for ordinal variables. Am Sociol Rev 1962; 27(6): 799-811.
25. Metsämuuronen J. Directional nature of Goodman–Kruskal gamma and some consequences: identity of Goodman–Kruskal gamma and Somers delta, and their connection to Jonckheere–Terpstra test statistic. Behaviormetrika 2021; 48(2): 283-307.
26. Ciuhu AN, Spircu T, Padure L, Badiu C, Poroch V, Omer S, et al. Relationships between communicating bad news, cancer disease adaptation, and related mood disorders. Proc Rom Acad Series B 2016; 18(2): 79-88.
27. Goodman LA, Kruskal WH. Measures of association for cross classifications. II: Further discussion and references. Journal of the American Statistical Association 1959; 54(285): 123-63.
28. Goodman LA, Kruskal WH. Measures of association for cross classifications III: Approximate sampling theory. Journal of the American Statistical Association 1963; 58(302): 310-64.
29. Goodman LA, Kruskal WH. Measures of association for cross classifications, IV: Simplification of asymptotic variances. Journal of the American Statistical Association 1972; 67(338): 415-21.

30. Goodman LA, Kruskal WH, Goodman LA, Kruskal WH. Measures of association for cross classifications. New York, NY: Springer; 1979.
31. Fang J, Liu L, Fang P. What is the most important factor affecting patient satisfaction–a study based on gamma coefficient. Patient Prefer Adherence 2019; 13: 515-25.
32. Kvålseth TO. An alternative measure of ordinal association as a value-validity correction of the Goodman–Kruskal gamma. Communications in Statistics-Theory and Methods 2017; 46(21): 10582-9.
33. Batchelder AW, Stanton AM, Kirakosian N, King D, Grasso C, Potter J, et al. Mental health and substance use diagnoses and treatment disparities by sexual orientation and gender in a community health center sample. LGBT Health 2021; 8(4): 290-9.
34. Dimitriu M, Ionescu CA, Matei A, Viezuina R, Rosu G, Ilinca C, et al. The problems associated with adolescent pregnancy in Romania: A cross-sectional study. J Eval Clin Pract 2019; 25(1): 117-24.

DOI: *10.48305/jims.v42.i773.0560*

Journal of Isfahan Medical School **Vol. 42, No. 773, 1st Week, September 2024**

Received: 16.04.2024 Accepted: 05.08.2024 Published: 29.08.2024

Types of Correlation Coefficients in Medical Science Research Using SPSS and R Software: A Review Article

Behzad Mahaki[1,2], Samira Jafari[3], Mansour Rezaei[1], Leila Solouki[2,3]

Review Article

Abstract

Background: The correlation coefficient is a statistical measure that measures the degree of correlation between two variables and takes values between -1 and +1, which indicates the strength and direction of the linear relationship between two variables. Correlation coefficients play a significant role in medical science studies. These coefficients help researchers to identify and measure the relationship between different clinical variables and health outcomes .

Description of the article: Considering the importance of calculating correlation coefficients in medical research, the purpose of this paper is to introduce various types of correlation coefficients, simple and applied statistical concepts and, methods to investigate the relationship between variables according to their nature and how to calculate them. Types of commonly used correlation coefficients, including Pearson, Spearman, Kendall, Phi, V Kramer, Summers Delta, Gamma, agreement coefficient and, Landa correlation coefficient will be analyzed using SPSS and R software. Also, the conditions of using each of the correlation coefficients will be discussed according to the different assumptions and the unique interpretation of each of the coefficients. Finally, examples of the calculation and interpretation of each of the correlation coefficients in medical sciences are provided.

Conclusion: Correlation is the most widely used statistical measure to evaluate the relationships between variables. However, it should be used with caution. Otherwise, it can lead to erroneous interpretations and results.

Keywords: Correlation coefficient; Software; Biomedical research

Citation: Mahaki B, Jafari S, Rezaei M, Solouki L. **Types of Correlation Coefficients in Medical Science Research Using SPSS and R Software: A Review Article.** J Isfahan Med Sch 2024; 42(773): 560-73.

1- Professor, Department of Biostatistics, School of Health, Kermanshah University of Medical Sciences, Kermanshah, Iran.
2- Health Data Science Research Center, Health Research Institute, Kermanshah University of Medical Sciences, Kermanshah, Iran.
3- PhD Student, Department of Biostatistics, School of Health, Kermanshah University of Medical Sciences, Kermanshah, Iran.
Corresponding Author: Leila Solouki, PhD Student, Department of Biostatistics, School of Health, AND Health Data Science Research Center, Health Research Institute Kermanshah University of Medical Sciences, Kermanshah, Iran;
Email: L_soloki68@yahoo.com