

تحليل همبستگی

(EViews)

تدوین: مرکز تحلیل آماری خوارزمی

www.kharazmi-statistics.ir

مرکز آماری خوارزمی

مقدمه:

همبستگی بین دو متغیر، شدت و جهت رابطه خطی میان آنها را نشان می‌دهد. مقدار این کمیت، چگونگی برآزش معادله‌ای خطی یا غیرخطی که ارتباط بین متغیرها را توجیه می‌کند، بیان می‌نماید (اگر معادله ثابتی بین متغیرها برقرار باشد، همبستگی کامل خواهد بود). ضریب همبستگی (r) همواره بین -1 و $+1$ خواهد بود. عبارت دیگر از آنجایی که رسم نمودار پراکنش به تنهایی نمی‌تواند نوع رابطه بین دو متغیر را تعیین کند، معیاری عددی برای محاسبه قدرت و جهت رابطه خطی بین آنها، مورد نیاز است. تحلیل همبستگی معمولاً با کوواریانس یا ضریب همبستگی بیان می‌شود. در ادامه تلاش شده تا با بیانی مختصر موارد مورد نظر توضیح داده شود و چگونگی محاسبه در نرم افزار EViews آموزش داده شود.

مطالبی که در این متن بیان شده است عبارتند از:

کوواریانس

خواص کوواریانس

چگونگی محاسبه کوواریانس در نرم‌افزار EViews

ضریب همبستگی

خواص ضریب همبستگی

چگونگی محاسبه ضریب همبستگی در نرم‌افزار EViews

کوواریانس

برای بیان همبستگی خطی بین دو متغیر X و Y ، یکی از معیارهای مهم و کاربردی، کوواریانس است که به صورت زیر تعریف می‌شود.

$$Cov(X, Y) = \sigma_{xy} = E(X - \mu_x)(Y - \mu_y) = E(XY) - \mu_x\mu_y$$

که μ_x و μ_y به ترتیب امیدریاضی X و Y هستند. در اینجا کوواریانس X و Y را با $Cov(X, Y)$ یا σ_{xy} و تخمین آن را بر حسب داده‌های نمونه با $\hat{\sigma}_{xy}$ نشان می‌دهیم که برابر است با:

$$\hat{\sigma}_{xy} = \frac{\sum_{i=1}^n x_i y_i}{n} = \frac{\sum_{i=1}^n x_i Y_i}{n} = \frac{\sum_{i=1}^n X_i Y_i}{n}$$

$$= \frac{\sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n}}{n} = \frac{\sum X_i Y_i}{n} - \bar{X}\bar{Y}$$

مشکلی که در رابطه با کوواریانس وجود دارد این است که شدت و ضعف همبستگی خطی بین X و Y را بیان نمی‌کند. به عبارت دقیق‌تر مقدار آن کاملاً تحت تاثیر بزرگی و کوچکی مقادیر متغیرها قرار دارد.

خواص کوواریانس:

۱. اگر به X یا Y و یا به هر دو، مقدار ثابت a را اضافه یا کم کنیم، کوواریانس تغییری نمی‌کند.

$$Cov(X + a, Y) = \frac{\sum[(X_i + a) - (\bar{X} + a)](Y_i - \bar{Y})}{n}$$

$$= \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{n} = Cov(X, Y)$$

و در حالت کلی خواهیم داشت:

$$Cov(X \pm a, Y \pm b) = Cov(X, Y)$$

۲. اگر مقدار X ها یا Y ها را در عدد ثابت a ضرب کنیم، کوواریانس نیز در a ضرب می‌شود.

$$Cov(aX, Y) = \frac{\sum(aX_i - a\bar{X})(Y_i - \bar{Y})}{n}$$

$$= \frac{a \sum(X_i - \bar{X})(Y_i - \bar{Y})}{n} = aCov(X, Y)$$

$$Cov(X, aY) = \frac{\sum(X_i - \bar{X})(aY_i - a\bar{Y})}{n}$$

$$= \frac{a \sum(X_i - \bar{X})(Y_i - \bar{Y})}{n} = aCov(X, Y)$$

۳. اگر X ها در a و Y ها در b ضرب شوند، کوواریانس در ab ضرب خواهد شد.

$$Cov(aX, bY) = \frac{\sum(aX_i - a\bar{X})(bY_i - b\bar{Y})}{n}$$

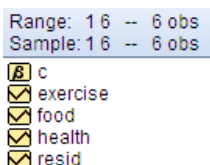
$$= \frac{ab \sum(X_i - \bar{X})(Y_i - \bar{Y})}{n} = abCov(X, Y)$$

چگونگی محاسبه کوواریانس در نرم افزار EViews

برای محاسبه کوواریانس می توان از گزینه های قرار گرفته شده در منوها استفاده کرد.

مثال: داده هایی به صورت زیر وجود دارد. داده ها مربوط به شاخص سلامتی برای افراد متفاوت است و شاخص مربوط به ورزش و غذا نیز بیان شده است. و بسته به میزان ورزش مناسب و غذای سالم مصرف شده از ۰ تا ۱۰ درجه بندی شده اند. داده ها از نرم افزار excel فراخوانی می کنیم. (برای آموزش چگونگی فراخوانی داده ها از نرم افزارهای خارجی به نرم افزار EViews به فایل "[معرفی نرم افزار و چگونگی ورود داده](#)" در صفحه آموزش

in eviews در پایگاه [مرکز تحلیل آماری خوارزمی](#) مراجعه نمایید).



پس از فراخوانی داده ها، نمایش داده ها به صورت تصویر روبرو خواهد بود.

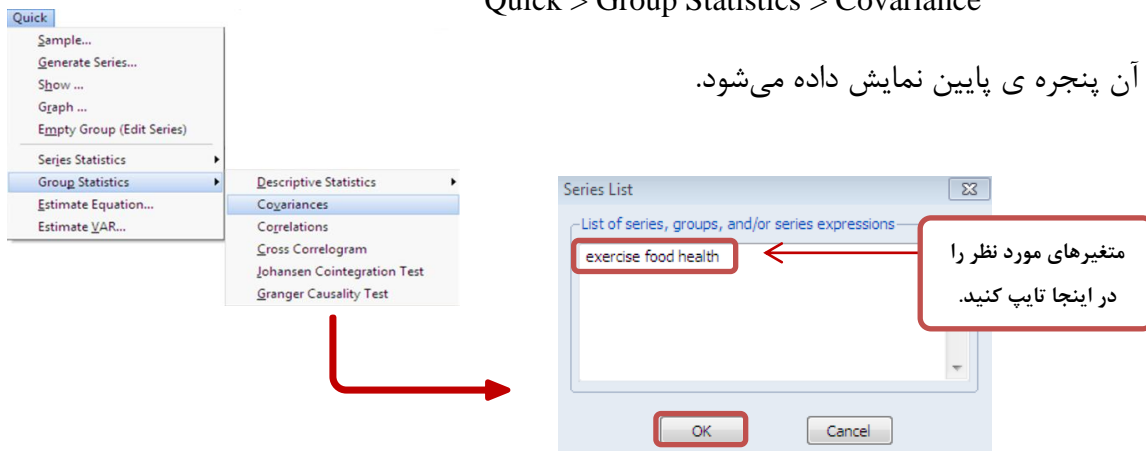
obs	EXERCISE	FOOD	HEALTH
1	9	7	8
2	4	7	5
3	3	8	6
4	8	7	9
5	4	7	5
6	6	2	3

در این مثال ۶ مشاهده که برای هر مشاهده ۳ متغیر ورزش، غذا و سلامتی آورده شده است.

برای محاسبه کوواریانس به صورت زیر عمل نمایید.

Quick > Group Statistics > Covariance

پس از آن پنجره ی پایین نمایش داده می شود.



در کادر list of series نام متغیرهایی که می‌خواهیم کوواریانس دوبه‌دو آنها را محاسبه کنیم تایپ کنید. (چنانچه در تصویر اول در صفحه قبل متغیرها را انتخاب کنیم و سپس مراحل را طی نمایید متغیرهای انتخاب شده به طور خودکار در پنجره‌ی series list تایپ شده ظاهر می‌شود.) پس از زدن دکمه ok نتایج نرم‌افزار نمایش داده می‌شود.

	EXERCISE	FOOD	HEALTH
EXERCISE	4.888889	-0.722222	2.666667
FOOD	-0.722222	3.888889	2.500000
HEALTH	2.666667	2.500000	4.000000

در خروجی گزارش شده ماتریسی نمایش داده شده که کوواریانس دوبه‌دو این متغیرها را محاسبه کرده است.

به طور مثال عدد درون کادر در تصویر کوواریانس مربوط به غذا و سلامتی را نشان می‌دهد که برابر ۲.۵ است. قطر اصلی این ماتریس واریانس هر یک از متغیرها را نشان می‌دهد. به عنوان مثال واریانس ورزش ۴.۸۸ است.

ضریب همبستگی

کوواریانس کاملاً متأثر از مقدار متغیرها است. بدیهی است که با تغییر مقیاس، شدت و ضعف همبستگی بین X و Y دچار تغییر نخواهد شد، در حالی که کوواریانس کوچک و بزرگ می‌شود.

برای حل این مسئله ابتدا باید متغیرهای X و Y را استاندارد نمود تا مقیاس اندازه‌گیری متغیرها تاثیر خود را از دست بدهد و سپس کوواریانس آنها را محاسبه کنیم:

$$Z_x = \frac{X - \mu_x}{\sigma_x} \quad , \quad Z_y = \frac{Y - \mu_y}{\sigma_y}$$

که σ_x و σ_y به ترتیب واریانس X و Y هستند. Z_x و Z_y استاندارد شده‌ی X و Y می‌باشند و از مقیاس متغیرهای X و Y متأثر نمی‌شوند. به عبارت دیگر اگر X و Y بزرگ یا کوچک شوند، استاندارد شده‌ی آنها تغییر نمی‌کند. به عنوان مثال اگر X را دو برابر کنیم، میانگین و انحراف معیار X نیز دو برابر شده و لذا صورت و مخرج نیز دو برابر خواهند شد و Z_x ثابت می‌ماند. حال با توجه به اینکه امید ریاضی Z_x و Z_y برابر با صفر است کوواریانس Z_x و Z_y را که موسوم به ضریب همبستگی X و Y است بصورت زیر محاسبه می‌کنیم.

$$\begin{aligned} Cov(Z_x, Z_y) &= \sigma_{Z_x, Z_y} = E(Z_x Z_y) \\ &= E\left(\frac{X - \mu_x}{\sigma_x}\right)\left(\frac{Y - \mu_y}{\sigma_y}\right) \end{aligned}$$

$$= \frac{E(X - \mu_x)(Y - \mu_y)}{\sigma_x \sigma_y} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \rho_{xy}$$

اما تخمین ضریب همبستگی براساس داده‌های نمونه را با $\hat{\rho}_{xy}$ یا r_{xy} و یا به طور خلاصه با $\hat{\rho}$ یا r نشان می‌دهیم.

$$\begin{aligned} \hat{\rho} = r &= \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y} = \frac{\hat{\sigma}_{xy}}{\sqrt{\hat{\sigma}_x^2 \hat{\sigma}_y^2}} = \frac{\frac{\sum x_i y_i}{n}}{\sqrt{\frac{\sum x_i^2}{n} \times \frac{\sum y_i^2}{n}}} \\ &= \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\sqrt{(\sum X_i^2 - n \bar{X}^2)(\sum Y_i^2 - n \bar{Y}^2)}} \end{aligned}$$

به سادگی می‌توان ثابت نمود که اگر X و Y در مقدار ثابتی مانند a ضرب یا تقسیم شوند مقدار r هیچ تغییر نخواهد کرد.

خواص ضریب همبستگی

۱. اگر به X ها یا Y ها و یا هر دو آنها مقدار ثابتی را اضافه کنیم، ضریب همبستگی تغییری نخواهد کرد.

$$\frac{\sum [(X_i + a) - (\bar{X} + a)][(Y_i + b) - (\bar{Y} + b)]}{\sqrt{\sum [(X_i + a) - (\bar{X} + a)]^2 \sum [(Y_i + b) - (\bar{Y} + b)]^2}} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}} = r$$

۲. اگر X ها یا Y ها یا هر دو را در عدد ثابتی ضرب یا تقسیم کنیم، ضریب همبستگی تغییری نخواهد کرد.

$$\begin{aligned} \frac{\sum (aX_i - a\bar{X})(bY_i - b\bar{Y})}{\sqrt{\sum (aX_i - a\bar{X})^2 \sum (bY_i - b\bar{Y})^2}} &= \frac{ab \sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{a^2 \sum (X_i - \bar{X})^2 b^2 \sum (Y_i - \bar{Y})^2}} \\ &= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}} = r \end{aligned}$$

چگونگی محاسبه ضریب همبستگی در نرم افزار EViews

برای محاسبه ضریب همبستگی از داده‌های مثال قبل استفاده نمایید. برای این منظور مسیر زیر را طی نمایید.

Quick > Group Statistics > Correlations

در اینجا نیز همانند قبل پنجره‌ی series list باز می‌شود. پس از نوشتن متغیرهای مورد نظر برای محاسبه‌ی ضریب همبستگی نتایج به شکل زیر نمایش داده می‌شود.

	EXERCISE	FOOD	HEALTH
EXERCISE	1.000000	-0.165635	0.603023
FOOD	-0.165635	1.000000	0.633866
HEALTH	0.603023	0.633866	1.000000

قطر اصلی ماتریس ارائه شده ضریب همبستگی هر یک از متغیرها با خودش است که برابر ۱ است. سایر درایه‌ها ضریب همبستگی ما بین دو متغیر را نشان

می‌دهد. به طور مثال ضریب همبستگی سلامتی و ورزش ۰.۶۰۳ است. یعنی ارتباط بین ورزش و سلامتی رابطه‌ی مثبت است. هرچه شاخص ورزش افزایش یابد شاخص سلامتی نیز بالاتر می‌رود. و ضریب همبستگی غذا و سلامتی نیز برابر ۰.۶۳۳۸ است. این رابطه نیز مثبت است اما ارتباط و شدت تاثیر غذا نسبی به ورزش بالاتر است و اثرگذاری بیشتری (هرچند به مقدار نه چندان زیاد) نسبت به ورزش بر شاخص سلامتی دارد.

منبع:

- اقتصاد سنجی همراه با کاربرد نرم افزار eviews7، دکتر علی سوری، انتشارات فرهنگ شناسی، بهار ۱۳۹۰

- کاربرد نرم افزارهای تخصصی در اقتصاد کشاورزی، تالیف: دکتر یآوری، قوامی، امیر سرداری، انتشارات دانشگاه پیام نور