



با سلام و احترام،

فصل بیستم و دوم: همبستگی دو متغیره (Bivariate Correlations)

دستورالعمل **Bivariate Correlations** رابطه‌های جفتی برای یک سری متغیرهاست و نتایج را در یک ماتریس نشان می‌دهد. این قابلیت برای تعیین توان و جهت رابطه بین متغیرهای کمی و ترتیبی (**Scale** و **Ordinal**) به کار می‌رود.

همبستگی‌های دو متغیره شامل دو نوع همبستگی فاصله‌ای (با استفاده از آزمون پیرسون) و ترتیبی (با استفاده از آزمون تاوکندال و اسپرمن) می‌باشد.

استفاده از همبستگی برای بررسی رابطه بین فروش خودروها و بازدهی سوخت

به منظور افزایش فروش، مهندسان طراح خودرو می‌خواهند توجه خود را در جنبه‌هایی از خودرو تمرکز دهند که برای مشتریان اهمیت دارد، مثلاً، نسبت بین بازدهی سوخت و فروش چیست؟ یک روش برای مقایسه این است که همبستگی بین فروش‌های گذشته و بازدهی سوخت محاسبه شود.

از همبستگی دو متغیره برای بررسی اهمیت بازدهی سوخت نسبت به میزان فروش خودرو استفاده نمایید.

شروع تحلیل

۱. مسیر **Analyze > Correlate > Bivariate** را برگزینید.
 ۲. **Sales in thousands** و **Fuel efficiency** را به عنوان متغیرهای تحلیل برگزینید.
 ۳. روی **Ok** کلیک کنید.
- این مراحل یک ماتریس همبستگی برای **Sales in thousands** و **Fuel efficiency** ایجاد می‌نماید.

ماتریس همبستگی

ضریب همبستگی پیرسون رابطه خطی بین دو متغیر کمی (Scale) را اندازه می‌گیرد. این ضریب برای محاسبه درجه و میزان ارتباط خطی بین دو متغیر در سطح فاصله‌ای و نسبی به کار می‌رود. مقادیر ضریب همبستگی پیرسون بین -۱ تا +۱ در نوسان است که علامت آن، نشانگر جهت رابطه (مثبت یا منفی) می‌باشد. منتهی در هنگام تفسیر نتایج، باید به مقادیر مطلق ضریب همبستگی توجه کنیم.



		Sales in thousands	Fuel efficiency
Sales in thousands	Pearson Correlation	1	-.017
	Sig. (2-tailed)		.837
	N	157	154
Fuel efficiency	Pearson Correlation	-.017	1
	Sig. (2-tailed)	.837	
	N	154	154

ضریب همبستگی پیرسون، یک آزمون پارامتری است. بنابراین لازم است که توزیع داده‌ها، از توزیع نرمال تبعیت کند. در واقع، بهترین زمان کاربرد این ضریب موقعی است که توزیع مقادیر متغیرها به صورت نرمال باشد. برای اینکه بتوانیم از نحوه رابطه خطی یا غیر خطی بین دو متغیر اطلاع یابیم، ترسیم یک نمودار پراکنش بین این دو متغیر می‌تواند کمک مفیدی به ما باشد.

ضریب همبستگی موجود در جدول منفی می‌باشد، اگر چه مقدار قابل توجهی با 0 اختلاف ندارد زیرا مقدار **p** (p-value) مربوط به **0.837** بزرگتر از **0.05** می‌باشد. این حکایت از آن دارد که طراحان نباید تلاش خود را روی بالا بردن بازدهی سوخت تمرکز دهند زیرا تأثیر محسوسی روی فروش ندارد.

با این وجود، وقتی متغیرها تقریباً توزیع نرمال دارند و مقادیر پرت ندارند ضریب همبستگی پیرسون بهتر کار می‌کند. یک نمودار پراکنش (Scatterplot) می‌تواند این مسائل ممکن را نشان دهد.

		Sales in thousands	Fuel efficiency
Sales in thousands	Pearson Correlation	1	-.017
	Sig. (2-tailed)		.837
	N	157	154
Fuel efficiency	Pearson Correlation	-.017	1
	Sig. (2-tailed)	.837	
	N	154	154

۱. برای ایجاد یک نمودار پراکنشی از نسبت فروش با بازدهی سوخت، مسیر زیر را برگزینید:

Graphs > Legacy Dialogs > Scatter/ Dot

۲. دکمه **Define** را کلیک کنید.
 ۳. **Sales in thousands** را به عنوان متغیر **y** و **Fuel efficiency** را به عنوان متغیر **x** برگزینید.
 ۴. **Model** را به عنوان متغیر **Label Cases by** برگزینید.
 ۵. **Ok** را کلیک کنید.
- نمودار پراکنش نتیجه، دو مقدار پرت ممکن را نشان می‌دهد، یکی در پایین و سمت راست نمودار و یکی در بالا و سمت چپ نمودار.
۶. برای شناسایی این نقاط، نمودار را با دو بار کلیک کردن روی آن فعال نمایید.
 ۷. آیکن **Data Label Mode** را کلیک کنید.
 ۸. نقطه سمت راست و پایین را انتخاب نمایید. این نقطه با **Metro** نشان داده شده است.
 ۹. نقطه بالا و سمت چپ را برگزینید. این با عنوان **F- Series** نشان داده شده است.
- F- Series** بیانگر خودروهایی است که تیم طراحی روی آن کار می‌کنند، بنابراین بنا دارید تا آن را فعلاً در سری داده‌ها نگه دارید. این نقطه ممکن است به علت توزیع چوله **Sales in thousands (Skew)** به صورت مقدار پرت ظاهر شود، بنابراین در تحلیل‌های دیگر سعی کنید آن را با **Log – transformed sales** تعویض نمایید.
- Metro** بیانگر وسائط نقلیه‌ای که تیم طراحی روی آن کار می‌کنند نمی‌باشد، بنابراین به راحتی می‌توانید آن را در تحلیل‌های دیگر حذف نمایید.

بهبود برآوردهای همبستگی‌ها

۱. برای حذف **Metro** از محاسبات همبستگی، مسیر **Data > Select Cases** را از منوی اصلی برگزینید.
۲. گزینه **If Condition is Satisfied** را انتخاب کرده و دکمه **If** را کلیک کنید.
۳. عبارت **'Metro' = ~ model** را در کادر متنی تایپ نمایید.
۴. دکمه **Ok** را در کادر محاوره **Select Cases** کلیک کنید.
۵. یک متغیر جدید ایجاد می‌شود که تمام موارد را به جز **Metro** در محاسبات تکمیلی به کار می‌برد.
۶. برای تحلیل داده‌های فیلترگذاری شده، کادر محاوره **Bivariate Correlations** را بازخوانی نمایید.
۷. **Sales in thousands** را به عنوان متغیر تحلیل از حالت انتخاب خارج نمایید.
۸. **Log – transformed sales** را به عنوان متغیر تحلیل انتخاب نمایید.
۹. **Ok** را کلیک کنید.

بعد از حذف مقادیر پرت و بررسی فروش‌های **Log – transformed**، اکنون همبستگی مثبت می‌باشد اما خیلی با 0 تفاوت ندارد.

با این وجود، خصوصیات جمعیتی مشتریان کامیون‌ها و اتومبیل‌ها متفاوت است و دلایل خرید یک کامیون یا یک اتومبیل ممکن است یکی نباشند. به زحمتش می‌ارزد که به نمودار پراکنش نگاه کنید، در این لحظه کامیون‌ها و اتومبیل‌ها به طور جداگانه مشخص شده‌اند.

		Fuel efficiency	Log-transformed sales
Fuel efficiency	Pearson Correlation	1	.136
	Sig. (2-tailed)		.093
	N	153	153
Log-transformed sales	Pearson Correlation	.136	1
	Sig. (2-tailed)	.093	
	N	153	153

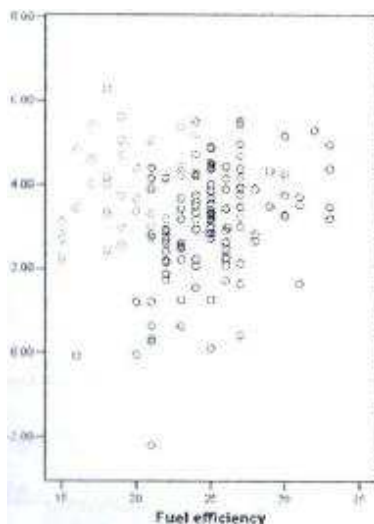
۱. برای ایجاد یک نمودار پراکنش از نسبت فروش با بازدهی سوخت، به منظور کنترل نوع وسیله نقلیه، کادر محاوره **Simple Scatterplot** را بازخوانی نمایید.

۲. **Sales in thousands** را از حالت انتخاب خارج کرده و **Log – transformed sales** را به عنوان متغیر **y** انتخاب کنید.

۳. **Vehicle type** را به عنوان متغیر کادر **Set Markers by** انتخاب نمایید.

۴. **Ok** را کلیک کنید.

نمودار پراکنش نشان می‌دهد که کامیون‌ها و اتومبیل‌ها، به وضوح گروه‌های مختلف را تشکیل می‌دهند. با جداسازی فایل داده‌ها بر طبق نوع وسیله نقلیه (**Vehicle type**)، ممکن است دید دقیق‌تری نسبت به رابطه پیدا نمایید. همچنین دقت نمایید که به دنبال فروش‌های **Log – transformed**، مقادیر پرت موجود در بالا و سمت چپ نمودار پنهان شده است.



بهبود برآوردهای همبستگی: جداسازی فایل

۱. برای جداسازی فایل داده‌ها بر طبق نوع وسیله نقلیه (**Vehicle type**)، مسیر **Data > Split File** را از منوی اصلی برگزینید.
۲. گزینه **Compare groups** را انتخاب کنید.
۳. **Vehicle type** را به عنوان متغیری که گروه‌ها باید بر اساس آن تقسیم بندی شوند (**Groups Based on**) انتخاب نمایید.
۴. **Ok** را کلیک کنید.
۵. برای تحلیل فایل جداسازی، کادر محاوره **Bivariate Correlations** را بازخوانی نمایید.
۶. **Ok** را کلیک کنید.

همبستگی‌ها

جداسازی فایل در **Vehicle type** رابطه بین فروش و بازدهی سوخت را واضح‌تر می‌نماید. یک همبستگی کاملاً مثبت بین فروش و بازدهی سوخت اتومبیل‌ها وجود دارد (a). در مورد کامیون‌ها، همبستگی مثبت است اما به طور معنی‌داری از 0 اختلاف ندارد (b).

Vehicle type			Fuel efficiency	Log-transformed sales
Automobile	Fuel efficiency	Pearson Correlation	1	1
		Sig. (2-tailed)		.451 ^a
	Log-transformed sales	Pearson Correlation	.451 ^a	1
		Sig. (2-tailed)	.000	
Truck	Fuel efficiency	Pearson Correlation	1	1
		Sig. (2-tailed)		.203 ^b
	Log-transformed sales	Pearson Correlation	.203 ^b	1
		Sig. (2-tailed)	.040	

برآوردهای همبستگی ناپارامتریک

آماره‌های **Spearman's rho** و **kendall's** رابطه مرتبه‌ای بین دو متغیر کمی و رتبه‌ای را اندازه می‌گیرد. آنها بی‌توجه به توزیع متغیرها عمل می‌کنند.

ضریب همبستگی اسپیرمن بر اساس رتبه (نمرات یا داده‌ها) است که میزان همبستگی بین دو متغیر در سطح ترتیبی را اندازه‌گیری می‌کند. این ضریب که معادل ناپارامتری ضریب همبستگی پیرسون است، صرفاً همبستگی میان رتبه‌های دو متغیر (همبستگی دو جانبه) در جامعه‌های با حجم نمونه کم را نشان می‌دهد.

بنابراین در همبستگی رتبه‌ای اسپیرمن، جفت متغیرها، به جای اندازه‌های فاصله‌ای یا نسبی، به صورت اندازه‌های ترتیبی (رتبه‌ای) بیان می‌شوند. برای محاسبه ضریب اسپیرمن نیز، افراد جامعه را بر حسب یکی از دو متغیر مورد مطالعه مرتب می‌کنیم. آنگاه به هر دو متغیر رتبه می‌دهیم. در نهایت نیز، میزان همبستگی بین این رتبه‌ها را حساب می‌کنیم.

همانند ضریب همبستگی پیرسون، دامنه ضریب همبستگی اسپیرمن از -1 تا $+1$ تغییر می‌کند. هنگامی که تمامی افراد، رتبه یکسانی در هر دو متغیر داشته باشند همبستگی رتبه‌ای $+1$ خواهد شد و هنگامی که رتبه‌های آنان در یک متغیر، دقیقاً مخالف رتبه‌شان در متغیر دیگر باشد، در آن صورت مقدار ضریب اسپیرمن -1 خواهد شد. اما اگر هیچ رابطه‌ای بین رتبه‌بندی‌ها وجود نداشته باشد، مقدار ضریب اسپیرمن صفر خواهد شد. مقدار ضریب رتبه‌ای اسپیرمن همانند **r** پیرسون تفسیر می‌شود.

۱. برای به دست آوردن یک تحلیل با استفاده از **Spearman's rho**، کادر محاوره **Bivariate Correlations** را بازخوانی نمایید.


۲. **Sales in thousands** را به عنوان متغیر تحلیل برگزینید.

۳. گزینه **Pearson** را غیر فعال و **Spearman** را فعال کنید.

۴. Ok را کلیک کنید.

همبستگی‌ها

ضریب همبستگی اسپیرمن به طور جداگانه برای اتومبیل‌ها و کامیون‌ها محاسبه می‌شود. همانند ضریب همبستگی پیرسون، رابطه بین **Log – transformed sales** و **Fuel efficiency** کاملاً قوی می‌باشد (a). با این وجود ضریب، اسپیرمن همبستگی مشابهی برای فروش‌های غیر تبدیلی (**Untransformed**) گزارش می‌کند. این به خاطر آن است که رو (**rho**) بر مرتبه رتبه‌هایی بنا شده است، که با تبدیل لگاریتم تغییر نمی‌کنند. به علاوه، مقادیر پرت تأثیر کمی روی ضریب اسپیرمن دارند، بنابراین با استفاده از آن به عنوان اندازه رابطه ممکن است در وقت و تلاش صرفه‌جویی شود (b).



Vehicle type				Fuel efficiency	Log-transformed sales	Sales in thousands
Automobile	Spearman's rho	Fuel efficiency	Correlation Coefficient	1.000	.418**	.418**
			Sig. (2-tailed)	.000	.000	.000
			N	114	114	114
	Log-transformed sales		Correlation Coefficient	.418**	1.000	1.000**
			Sig. (2-tailed)	.000		.000
			N	114	114	114
	Sales in thousands		Correlation Coefficient	.418**	1.000**	1.000
			Sig. (2-tailed)	.000		.000
			N	114	114	114
Truck	Spearman's rho	Fuel efficiency	Correlation Coefficient	1.000	.141	.141
			Sig. (2-tailed)	.40	.40	.40
			N	40	40	40
	Log-transformed sales		Correlation Coefficient	.237	1.000	1.000*
			Sig. (2-tailed)	.141		.41
			N	40	40	40
	Sales in thousands		Correlation Coefficient	.237	1.000**	1.000
			Sig. (2-tailed)	.141		.41
			N	40	40	40

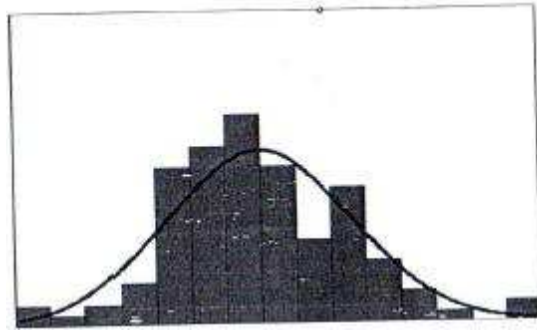
خروجی‌های همبستگی‌های دو متغیره

توصیف‌های همبستگی‌های دو متغیره

این جدول آماره‌های توصیفی هر متغیر را نمایش می‌دهد. ستون **Mean** مقدار میانگین، ستون **Std. Deviation** شاخص‌های تغییرپذیری مقادیر، و ستون **N** تعداد موارد با مقادیر معتبر می‌باشد.

همبستگی‌های دو متغیره

جدول همبستگی‌ها ضرائب همبستگی پیرسون، سطح معنی‌داری، و تعداد موارد با مقادیر معتبر را نشان می‌دهد. ضرائب همبستگی پیرسون فرض کرده که داده‌ها به صورت نرمال توزیع شده‌اند.



ضریب همبستگی پیرسون نشان دهنده همبستگی خطی بین دو متغیر می باشد. مقادیر ضریب همبستگی در دامنه ۱- تا ۱ می باشد. علامت ضریب همبستگی، جهت رابطه را نشان می دهد (مثبت یا منفی). مقدار مطلق ضریب همبستگی میزان استحکام را نشان می دهد، با مقادیر مطلق بزرگتر، روابط مستحکم تری وجود دارد.

ضرائب همبستگی در قطر اصلی همیشه 1 می باشد، چون هر متغیر یک رابطه کاملاً خطی با خودش دارد (a).

	Area	Price	Location
Pearson Correlation	Area	0.960	0.484
	Price	0.960	0.707
	Location	0.484	0.707
Sig. (2-tailed)	Area	0.000	0.019
	Price	0.000	0.000
	Location	0.019	0.000
N	Area	23	23
	Price	23	22
	Location	23	23

همبستگی های بالای قطر اصلی، تصویر قرینه همبستگی های پایین هستند در این مثال، ضریب همبستگی **price** و **Area** برابر 0.960 است. چون نسبتاً به 1 نزدیک است، نشان می دهد که **price** و **Area** به طور مثبت همبسته هستند. ضریب همبستگی **Area** و **Location** برابر 0.484 می باشد. از آنجا که 0.484 به 1 یا 1- نزدیک نیست. نشان می دهد که **Area** و **Location** به طور قوی همبسته نیستند.

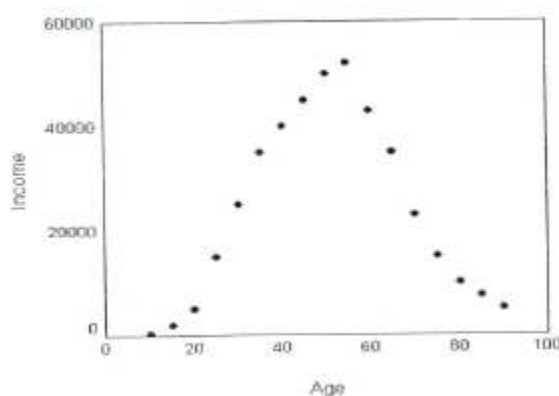
سطح معنی داری (Sig.) نیز در جدول همبستگی ظاهر شده است. سطح معنی دار (یا مقدار P) احتمال به دست آوردن نتایج به بزرگی مورد مشاهده شده می باشد. اگر سطح معنی داری خیلی کوچک باشد آنگاه همبستگی معنی دار است و دو متغیر به طور خطی وابسته می باشند و اگر این مقدار نسبتاً بزرگ باشد، آنگاه همبستگی معنی دار نیست و دو متغیر به طور خطی وابسته نیستند.

ضریب همبستگی **price** و **Area** برابر 0.960 است. سطح معنی داری یا مقدار P برابر 0.000 است. سطح معنی داری کوچک نشان می دهد که **price** و **Area** به طور معنی داری همبستگی مثبت دارند.

با افزایش **Area** مورد **price** نیز افزایش می یابد. و با کاهش **Area** مورد **price** نیز کاهش می یابد.

همانطور که گفته شد **N** تعداد موارد با مقادیر معتبر برای هر دوی **price** و **Area** برابر 23 می باشد. در این جدول، تعداد موارد به مقادیر معتبر برای همه متغیرها در این تحلیل (**price**، **Area** و **Location**) برابر 22 می باشد.

حتی اگر همبستگی بین دو متغیر معنی دار نباشد، متغیرها ممکن است همبسته باشند اما رابطه خطی نباشد. قبل از محاسبه همبستگی ها، یک نمودار پراکنش از دو متغیر ایجاد کنید تا نحوه وابستگی آنها را ببینید.



مادامی که همبستگی بین **Age** و **Income** قابل توجه نباشد، این متغیرها وابسته ظاهر می شوند.

همبستگی های ناپارامتریک دو متغیره

جدول همبستگی های ناپارامتریک شامل ضرائب همبستگی، سطح معنی داری، و تعداد موارد با مقادیر معتبر می باشد.

در جدول همبستگی های ناپارامتریک می توانید اطلاعات **kendall's tau- b** و / یا **Spearman's rho** را ببینید.

			Area	Price	Location
Kendall's tau- b	Correlation Coefficient	Area	1.000	.858	.372
		Price	.858	1.000	.560
		Location	.372	.560	1.000
	Sig. (2-tailed)	Area		.000	.029
		Price	.000		.001
		Location	.029	.001	
	N	Area	24	23	23
		Price	23	23	22
		Location	23	22	23
Spearman's rho	Correlation Coefficient	Area	1.000	.963	.469
		Price	.963	1.000	.663
		Location	.469	.663	1.000
	Sig. (2-tailed)	Area		.000	.024
		Price	.000		.001
		Location	.024	.001	
	N	Area	24	23	23
		Price	23	23	22
		Location	23	22	23

هم $kendall's\ tau - b$ و هم $Spearman's\ rho$ از رتبه‌های داده‌ها برای محاسبه ضرائب همبستگی استفاده می‌کنند.



Area	Rank of Area
1582	7
1870	11
2180	14
2100	13
1455	5
1630	8
1210	1
1857	10
2000	12
2256	15
1290	2
1725	9
1400	4
1535	6
1333	3

$kendall's\ tau - b$ استحکام رابطه بین متغیرهای اندازه گرفته شده در سطح رتبه‌ای را اندازه می‌گیرد. $Spearman's\ rho$ یک ضریب همبستگی رتبه‌ای است که رابطه را در سطح ترتیبی اندازه می‌گیرد. این یک نمونه ناپارامتریک از همبستگی پیرسون است که بر رتبه‌های داده‌ها به جای مقادیر واقعی بنا شده است. مقادیر ضریب همبستگی در محدوده -1 تا 1 می‌باشد. همه مشخصاتی که درباره داده‌های روی قطر اصلی ضریب همبستگی و مقادیر بالا و پایینی آن گفته شد و مطالبی که درباره کوچک یا بزرگ بودن سطح معنی‌داری بیان شد در اینجا نیز صادق است. درضمن کاربرد آن در جوامع با حجم نمونه بزرگ می‌باشد بر خلاف اسپیرمن که در جوامع با حجم کوچک به کار می‌رفت.

مبحث را در مرجع کاربردی SPSS 20 (36) دنبال نمایید.



Telegram.me/iepnu
 کانال تخصصی مهندسی صنایع دانشگاه پیام نور