

محاسبه ضرایب همبستگی

(SAS)

تدوین: مرکز تحلیل آماری خوارزمی

www.kharazmi-statistics.ir

مرکز آماری خوارزمی

مقدمه:

در بسیاری از تحلیل‌های آماری یکی از مسائل مهم، بررسی همبستگی بین متغیرهاست. همبستگی پیرسون، اسپیرمن و محاسبه‌ی کوواریانس از جمله‌ی همبستگی‌هاست که در نرم‌افزار SAS قادر به محاسبه‌ی آن هستیم. در این متن به آموزش چگونگی محاسبه این ضرایب در نرم‌افزار SAS می‌پردازیم.

مطالبی که در این متن بیان شده است به شرح زیر است:

تحلیل ضرایب همبستگی

محاسبه‌ی ضرایب همبستگی (پیرسن، اسپیرمن و کوواریانس) در نرم‌افزار

تفسیر خروجی نرم‌افزار

تحلیل ضرایب همبستگی

در بحث رگرسیون و یا بحث تحلیل عوامل در محیط SAS از دستور `proc corr` برای تحلیل همبستگی استفاده می‌شود. فرم کلی این دستور بصورت زیر است:

```
proc corr option;
```

```
var variables name;
```

Options (دستورهای انتخابی) که می‌توانید همراه این دستور بکار ببرید عبارتند از:

- **Pearson**: برای تحلیل همبستگی پیرسن از این دستور استفاده می‌شود.
- **Spearman**: برای تحلیل همبستگی اسپیرمن از این دستور استفاده می‌شود.
- **Cov** (کوواریانس): برای اینکه بتوانید "ماتریس کوواریانس" را در خروجی ببینید؛ از این دستور استفاده کنید.

از دستور `var(iable)` (متغیر) برای تعیین متغیرهایی که باید روی آنها محاسبات صورت گیرد، استفاده می‌شود.

با اجرای دستور `var:`

- ۱- خلاصه آماری متغیرهای معرفی شده نمایش داده می شود.
- ۲- ضرایب همبستگی و مقادیر p-value برای آزمون محاسبه و گزارش می شود.

$$\begin{cases} H_0: \rho = \rho_0 = 0 \\ H_1: \rho \neq \rho_0 \neq 0 \end{cases}$$

زمانی می توان از ضریب همبستگی پیرسن و اسپیرمن استفاده کرد که هر دو متغیر کمیت پذیر هستند. دقت شود که ضریب همبستگی پیرسن در حالت پارامتریک و اسپیرمن در حالت ناپارامتریک استفاده می شود.

محاسبه ی ضرایب همبستگی (پیرسن، اسپیرمن و کوواریانس)

برای محاسبه ضرایب همبستگی مربوطه در نرم افزار به مثال های زیر توجه نمایید.

مثال ۱: داده های زیر در جدول را در نظر بگیرید. می خواهیم ضرایب همبستگی مربوط به متغیرهای A تا D را محاسبه نماییم.

دستورات را مطابق با نوشته پایین در نرم افزار بنویسید.

```

title'corr';
data kharazmi;
input x1 x2 x3 X4 X5;
datalines;
190.5 152.5 145.0 73.5 136.5
172.5 132.0 125.5 63.0 121.0
167.0 130.0 125.5 69.5 119.5
169.5 150.5 133.5 64.5 128.0
175.0 138.5 126.0 77.5 135.5
177.5 142.5 142.5 71.5 131.0
179.5 142.5 127.5 70.5 134.5
179.5 138.0 133.5 73.5 132.5
173.5 135.5 130.5 70.0 133.5
162.5 139.0 131.0 62.0 126.0
;
proc corr pearson spearman cov;
var x1-x5;
run;

```

	A	B	C	D	E
۱	190.5	152.5	145	73.5	136.5
2	172.5	132	125.5	63	121
3	167	130	125.5	69.5	119.5
4	169.5	150.5	133.5	64.5	128
5	175	138.5	126	77.5	135.5
6	177.5	142.5	142.5	71.5	131
7	179.5	142.5	127.5	70.5	134.5
8	179.5	138	133.5	73.5	132.5
9	173.5	135.5	130.5	70	133.5
10	162.5	139	131	62	126

توضیحات بیشتر را درباره‌ی چگونگی ورود داده در نرم‌افزار را می‌توانید در فایل‌های "[نحوه‌ی ورود داده](#)" و "[فراخوانی داده از منابع خارجی](#)" در صفحه مربوط به آموزش نرم‌افزار SAS در [سایت مرکز تحلیل آماری خوارزمی](#) مشاهده نمایید.

✓ عبارت `proc corr pearson spearman cov` به فرمان مربوط به اجرای محاسبه‌ی ضرایب همبستگی پیرسن، اسپیرمن و کوواریانس است. و عبارت `var x1-x5` به معنای آن است که به نرم‌افزار این دستور داده می‌شود که ضرایب همبستگی متغیرهای `x1` تا `x5` را دو به دو محاسبه نماید. چنانچه مثلاً بخواهیم همبستگی دوه‌دو متغیرهای `x2` تا `x4` را محاسبه کنیم عبارت `var x2-x4` را تایپ می‌نماییم. برای محاسبه تک‌تک این ضرایب هم وان هر یک را به تنها تایپ نمود.

چنانچه موارد بالا به درستی در نرم‌افزار نوشته شود پاسخ‌ها به صورت زیر در پنجره‌ی `output` نمایش داده می‌شود.

corr 17:31 Friday, January 30, 2009 11

The CORR Procedure

5 Variables: x1 x2 x3 X4 X5

ماتریس کوواریانس و درجه آزادی

 →

Covariance Matrix, DF = 9

	x1	x2	x3	X4	X5
x1	60.56666667	29.33888889	32.12777778	25.15555556	32.93333333
x2	29.33888889	52.26666667	35.99444444	4.68888889	25.13333333
x3	32.12777778	35.99444444	47.52500000	7.55277778	18.67777778
X4	25.15555556	4.68888889	7.55277778	24.96944444	20.76111111
X5	32.93333333	25.13333333	18.67777778	20.76111111	35.90000000

خلاصه آماری

 →

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum
x1	10	174.70000	7.78246	174.25000	162.50000	190.50000
x2	10	140.10000	7.22957	138.75000	130.00000	152.50000
x3	10	132.05000	6.89384	130.75000	125.50000	145.00000
X4	10	69.55000	4.99694	70.25000	62.00000	77.50000
X5	10	129.80000	5.99166	131.75000	119.50000	136.50000

ضریب همبستگی
پیرسون

Pearson Correlation Coefficients, N = 10
Prob > |r| under H0: Rho=0

	x1	x2	x3	X4	X5
x1	1.00000	0.52145 0.1222	0.59883 0.0674	0.64686 0.0432	0.70627 0.0224
x2	0.52145 0.1222	1.00000	0.72221 0.0183	0.12979 0.7208	0.58022 0.0787
x3	0.59883 0.0674	0.72221 0.0183	1.00000	0.21925 0.5428	0.45219 0.1895
X4	0.64686 0.0432	0.12979 0.7208	0.21925 0.5428	1.00000	0.69342 0.0262
X5	0.70627 0.0224	0.58022 0.0787	0.45219 0.1895	0.69342 0.0262	1.00000

corr 17:31 Friday, January 30, 2009 12

ضریب همبستگی
اسپیرمن

The CORR Procedure
Spearman Correlation Coefficients, N = 10
Prob > |r| under H0: Rho=0

	x1	x2	x3	X4	X5
x1	1.00000	0.43598 0.2078	0.47401 0.1664	0.80183 0.0053	0.79636 0.0058
x2	0.43598 0.2078	1.00000	0.77065 0.0091	0.24390 0.4971	0.53496 0.1111
x3	0.47401 0.1664	0.77065 0.0091	1.00000	0.31499 0.3754	0.41464 0.2335
X4	0.80183 0.0053	0.24390 0.4971	0.31499 0.3754	1.00000	0.78420 0.0072
X5	0.79636 0.0058	0.53496 0.1111	0.41464 0.2335	0.78420 0.0072	1.00000

مقدار ضریب همبستگی

مقدار p-value

تفسیر خروجی نرم افزار:

در گزارش خروجی نرم افزار، ماتریس کوواریانس متغیرها، خلاصه‌ی آماری متغیرهای بیان شده، ضریب همبستگی پیرسون و اسپیرمن برای دوبه‌دو متغیرهای بیان شده نمایش داده شده است.

- همبستگی مثبت نشان می‌دهد که با افزایش مقدار یکی از متغیرها، مقدار حاصل از متغیر دیگر افزایش می‌یابد.
- همبستگی منفی نشان می‌دهد که با افزایش مقدار یکی از متغیرها، مقدار حاصل از متغیر دیگر کاهش می‌یابد.
- لازم به یادآوری است که در گزارش مربوط به جدول کوواریانس مقادیر قطر اصلی واریانس متغیر و مابقی درایه‌ها، کوواریانس دوبه‌دو متغیرها است.

عدد اول نوشته شده در خروجی‌های هر یک از ضرایب محاسبه شده ضریب همبستگی مورد نظر است و عدد پایینی مقدار p محاسبه شده (p -value) است.

آزمون فرض : (به طور مثال در اینجا خروجی مربوط به آزمون اسپیرمن در نظر گرفته شده است).

I. آزمون‌های مورد نظر در این مثال عبارتند از:

$$\begin{cases} H_0: \rho_{x1,x4} = 0 \\ H_0: \rho_{x1,x4} \neq 0 \end{cases} \quad \text{و} \quad \begin{cases} H_0: \rho_{x2,x5} = 0 \\ H_0: \rho_{x2,x5} \neq 0 \end{cases}$$

II. آماره‌ی آزمون:

برای متغیرهای دوم و پنجم $p - value = 0,1111$

برای متغیرهای اول و چهارم $p - value = 0,0053$

III. ناحیه بحرانی:

$$\alpha = 0,05$$

IV. نتیجه‌گیری:

$$\alpha = 0,05 < p_{X2,x5} - value = 0,1111$$

$$\alpha = 0,05 < p_{X1,x4} - value = 0,0053$$

در نتیجه فرض H_0 برای متغیرهای ۲ و ۵ و متغیرهای ۱ و ۴ رد نمی‌شود. ولی در صورتی که مقدار p از ۰,۰۵ بزرگتر می‌شد فرض صفر رد می‌شود و دلیلی بر پذیرش فرض صفر وجود ندارد.

- همانطور که در مقادیر همبستگی‌ها مشاهده می‌شود تمامی مقادیر مثبت هستند در نتیجه می‌توان گفت با افزایش در مقدار یکی از متغیرها، متغیر دیگر نیز افزایش می‌یابند.

مثال ۲: داده‌های مربوط به مثال قبل را در نظر بگیرید. می‌خواهیم ضرایب همبستگی مربوط به متغیرهای A تا D را محاسبه نماییم. به طوری که در این حالت خلاصه آماری و مقدار p-value چاپ نشود.

```
title'corr';
data kharazmi;
input x1 x2 x3 X4 X5;
datalines;
190.5 152.5 145.0 73.5 136.5
172.5 132.0 125.5 63.0 121.0
167.0 130.0 125.5 69.5 119.5
169.5 150.5 133.5 64.5 128.0
175.0 138.5 126.0 77.5 135.5
177.5 142.5 142.5 71.5 131.0
179.5 142.5 127.5 70.5 134.5
179.5 138.0 133.5 73.5 132.5
173.5 135.5 130.5 70.0 133.5
162.5 139.0 131.0 62.0 126.0
;
proc corr pearson spearman cov nosimple;
var x1-x5;
run;
```

با اضافه کردن عبارت فوق (داخل کادر) در این حالت خروجی نرم‌افزار خلاصه آماری چاپ نمی‌شود.

چنانچه مایل باشیم که مقدار p-value نیز چاپ نشود، می‌توان از دستور `prob corr` در دستور `proc corr` استفاده کرد. (به صورت روبرو)

```
proc corr pearson spearman cov noprob;
```

منبع:

- آموزش گام به گام sas، رسول نصیری، انتشارات مرکز فرهنگی نشر گستر، ۱۳۸۴

- ضرایب همبستگی و محاسبه آن، www.iranpde.ir