

## PROC CORR

zum Überprüfung des Zusammenhanges zweier quantitativer Merkmale mit  
Korrelationskoeffizienten nach Pearson bzw. Spearman

### Was Sie wissen sollten:

Der Korrelationskoeffizient misst das Zusammenwirken zweier oder mehrerer Merkmale und gibt Auskunft über die Form des Zusammenhanges.

SAS verwendet standardmäßig den **Produktmomenten-Korrelationskoeffizienten nach Pearson**, der Werte zwischen -1 und +1 annehmen kann. Berechnet wird er, indem man die beiden Standardabweichungen durch die Kovarianz dividiert.

Liegt der errechnete Wert um Null, gibt es keinen Zusammenhang, ist er genau -1 oder +1, befinden sich die Wertepaare auf einer Geraden und zwischen den Merkmalen besteht ein starker Zusammenhang. Weist er Werte größer als Null auf, besteht ein gleichsinniger (konkordanter) Zusammenhang, was bedeutet, dass beim Ansteigen des einen Merkmals auch das andere ansteigt. Nimmt er dagegen einen Wert kleiner als Null an, geht man von einem gegensinnigen (diskordanten) Zusammenhang aus, was heißt, dass das eine Merkmal beim Ansteigen des anderen abfällt.

Ab wann man einen Zusammenhang als groß oder bedeutend bezeichnet, hängt vom Sachverhalt ab. Bei Messwerten legt man strengere Maßstäbe an als bei "weichen" Daten aus der Psychologie.

Bei nicht-normalverteilten metrischen Daten setzt man den **Rang-Korrelationskoeffizienten nach Spearman** ein. SAS bildet für die Berechnung dieses Korrelationskoeffizienten die Ränge der beobachteten Werte und berechnet dann daraus den Korrelationskoeffizienten.

### Allgemeine Form:

```
TITLE1 'Korrelationskoeffizient nach Pearson' ;  
TITLE2 '-----' ;  
PROC CORR DATA=name Optionen ;  
    VAR var1 var2 ;  
RUN ;
```

```
TITLE1 'Korrelationskoeffizient nach Pearson' ;  
TITLE2 '-----' ;  
PROC CORR DATA=name Optionen ;  
    VAR var2 var3 ;  
    WITH var1 ;  
RUN ;
```

### Beispiele und Beschreibung der Programm-Statements:

```
TITLE1 'Korrelationskoeffizient nach Pearson' ;  
TITLE2 '-----' ;  
PROC CORR DATA=beispiel ;  
    VAR groesse gewicht ;  
RUN ;
```

```
TITLE1 'Korrelationskoeffizient nach Pearson' ;  
TITLE2 '-----' ;  
PROC CORR DATA=beispiel ;  
    VAR groesse gewicht ;  
    WITH alter ;  
RUN ;
```

Die Prozedur beginnt mit PROC endet mit RUN ;

Das ; am Ende eines Befehls beendet eine Anweisung innerhalb der Prozedur.

**CORR** ist der Name der Prozedur.

**DATA = name** legt das zu verwendende Datenfile fest (im Beispiel **beispiel**).

**VAR** ist das gehörende **Schlüsselwort** und steht für Variablen.

Ohne **VAR** - Statement berechnet SAS Häufigkeiten für **alle** numerischen Merkmale in der Datendatei, was für Merkmale, die mit Zahlen verschlüsselt eingegeben wurden, unsinnig ist, z.B. beim Merkmal *Geschlecht: 1 für männlich, 2 für weiblich*.

Es sind beliebig viele **VAR**-Statements in einer PROC **CORR** möglich, wobei Sie für jeden Vergleich ein eigenes **VAR**-Statement angeben müssen.

**Var?** wird mit den Namen der zu verarbeitenden Merkmale überschrieben. Trennen Sie die Merkmalsnamen mit Leerzeichen!!!

**WITH** kommt zum Einsatz, wenn mehrere Merkmale mit dem gleichen Merkmal in Beziehung gesetzt werden sollen - verkürzte Schreibweise:

**anstelle von:**

```
TITLE1 ' Korrelationskoeffizient nach Pearson' ;
TITLE2 '-----' ;
PROC CORR DATA=name ;
    VAR alter groesse ;
    VAR alter gewicht ;
RUN ;
```

**schreiben Sie:**

```
TITLE1 ' Korrelationskoeffizient nach Pearson' ;
TITLE2 '-----' ;
PROC CORR DATA=name ;
    VAR groesse gewicht ;
    WITH alter ;
RUN ;
```

Mit dem **TITLE**-Befehl stellen Sie den Ergebnissen Überschriften voran. Der Befehl ist nicht zwingend notwendig, bringt aber Struktur in die Auswertung. **Achtung** - SAS übernimmt Titel in nachfolgende Prozeduren, wenn dort das **TITLE**-Statement fehlt.

Starten Sie die Programmzeilen aus dem **oberen Beispiel auf der Seite 1**, zeigt SAS im Output-Fenster folgende Informationen:

Korrelationskoeffizient nach Pearson

The CORR Procedure

2 Variables: Groesse Gewicht

Variable	Simple Statistics			
	N	Mean	Std Dev	Sum
Groesse	71	173.00000	8.82853	12283
Gewicht_vor	70	66.22857	12.46348	4636

Variable	Simple Statistics	
	Minimum	Maximum
Groesse	156.00000	196.00000
Gewicht_vor	46.00000	106.00000

### Pearson Correlation Coefficients

Prob > |r| under H0: Rho=0

Number of Observations

	Groesse	Gewicht
Groesse	1.00000	0.76247
		<.0001
	71	70
Gewicht	0.76247	1.00000
	<.0001	
	70	70

Starten Sie die Programmzeilen aus dem unteren Beispiel auf der Seite 1, zeigt SAS im Output-Fenster folgende Informationen:

### Korrelationskoeffizient nach Pearson

-----

#### The CORR Procedure

```
1 With Variables: alter
2 Variables: Groesse Gewicht
```

Variable	N	Simple Statistics		Sum
		Mean	Std Dev	
alter	71	24.63380	2.32526	1749
Groesse	71	173.00000	8.82853	12283
Gewicht	70	66.22857	12.46348	4636

Variable	Simple Statistics	
	Minimum	Maximum
alter	21.00000	27.00000
Groesse	156.00000	196.00000
Gewicht	46.00000	106.00000

### Pearson Correlation Coefficients

Prob > |r| under H0: Rho=0

Number of Observations

	Groesse	Gewicht
alter	0.12317	-0.05018
	0.3061	0.6799
	71	70

### Beschreibung des OUTPUT-Fensters am obigen Beispiel:

#### Simple Statistics.

Hier zeigt SAS die Namen der verwendeten Merkmale und die zugehörigen einfachen statistischen Maßzahlen: N=Anzahl, MEAN=Mittelwert, STD DEV=Standardabweichung, SUM=Summe der Beobachtungswerte, MINIMUM und MAXIMUM (vgl. PROC MEANS).

Unter den Maßzahlen gibt SAS Überschriften für die berechneten Korrelationskoeffizienten, p-Werte und die Anzahl der der Berechnung zugrunde liegenden Beobachtungen aus. Die Reihenfolge der Überschriften entspricht der der Werten darunter:

### Pearson Correlation Coefficients.

Hier gibt SAS den berechneten Korrelationskoeffizienten nach Pearson aus. – Der Zusammenhang eines Merkmals mit sich selbst ist logischerweise immer gleich 1 und der zugehörige p-Wert wird daher nicht ausgegeben (siehe Beispiel auf Seite 2).

Im ersten Beispiel besteht zwischen den Merkmalen **groesse** und **gewicht** ein recht starker Zusammenhang ( $r=0.76247$ ). Die Form dieses Zusammenhangs ist nicht bekannt und man kann auch keine Kausalbeziehung zwischen den Variablen damit begründen.

Im zweiten Beispiel besteht zwischen Merkmal **alter** und dem Merkmal **groesse** bzw. **gewicht** nur ein geringer Zusammenhang ( $r=0.12317$  bzw.  $r=-0.05018$ ), aber das liegt an den Beispielsdaten, die in Biomathematik-Vorlesung erhoben wurden: Es liegen die Daten von Studenten zugrunde, die zwischen 21 und 27 Jahre alt sind.

### Prob > |r| under H0: Rho=0

Diese Zeile zeigt den p-Wert für den Test des Korrelationskoeffizienten auf Null an. – Die statistische Signifikanz dieses Hypothesentests hat keine Bedeutung für die Aussage über die Stärke des Zusammenhangs zwischen den beiden Merkmale, dafür ist der Koeffizienten selbst von Wichtigkeit.

Ein kleiner p-Wert besagt hier, dass es unwahrscheinlich ist, einen größeren Korrelationskoeffizienten allein durch Zufall zu bekommen, wenn eine Stichprobe aus einer Grundgesamtheit gezogen worden ist, deren Korrelation gleich Null ist (Hypothesentest mit der Nullhypothese  $H_0: \text{Rho}=0$ ). Auch wenn die Nullhypothese abgelehnt werden kann, bedeutet dies nur, dass die Korrelation mit großer Wahrscheinlichkeit von Null verschieden ist, aber wie groß sie tatsächlich ist oder wie nahe sie bei 1 liegt, darüber ist keine Aussage möglich. Bei einer ausreichend großen Stichprobe, kann man einen beliebigen Korrelationskoeffizienten als statistisch signifikant von Null unterscheiden.

### Number of Observations

Kommen fehlende Werte vor, gibt SAS in einer dritten Zeile die Anzahl der Beobachtungen aus, die in die Auswertung eingegangen sind. Gibt es keine fehlenden Werte, erscheint die Anzahl der Beobachtungen hinter **N=** in der Tabellenbeschriftung:

#### Pearson Correlation Coefficients

Prob > |r| under H0: Rho=0

Number of Observations

	Groesse	Gewicht
alter	0.12317	-0.05018
	0.3061	0.6799
	71	70

#### Pearson Correlation Coefficients, N = 71

Prob > |r| under H0: Rho=0

	alter	Groesse
alter	1.00000	0.12317
		0.3061

## Beschreibung nützlicher Optionen:

- Option **NOSIMPLE**: SAS unterdrückt die Berechnung der Summenhäufigkeiten:

```
TITLE1 'Korrelationskoeffizient nach Pearson' ;
TITLE2 '-----' ;
PROC CORR DATA=beispiel NOSIMPLE;
VAR groesse gewicht;
RUN ;
```

⇒ Der Ausgabe-Teil **Simple Statistics** fehlt, SAS gibt nur noch die Koeffizienten und p-Werte aus:

```
Korrelationskoeffizient nach Pearson
-----
```

The CORR Procedure

```
1 With Variables:      alter
2   Variables:      Groesse      Gewicht
```

```
Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations
```

	Groesse	Gewicht
alter	0.12317	-0.05018
	0.3061	0.6799
	71	70

- Option **NOPROB**: SAS unterlässt die die Anzeige der p-Werte für die Prüfung des Korrelationskoeffizienten auf Null:

```
TITLE1 'Korrelationskoeffizient nach Pearson' ;
TITLE2 '-----' ;
PROC CORR DATA=beispiel NOPROB ;
VAR groesse gewicht;
RUN ;
```

⇒ SAS gibt nur die statistischen Maßzahlen und die Korrelationskoeffizienten aus, die Angabe der p-Werte (**Prob > |r| under H0: Rho=0**) fehlt:

```
Korrelationskoeffizient nach Pearson
-----
```

The CORR Procedure

```
1 With Variables:      alter
2   Variables:      Groesse      Gewicht
```

	Simple Statistics			
Variable	N	Mean	Std Dev	Sum
Groesse	71	173.00000	8.82853	12283
Gewicht_vor	70	66.22857	12.46348	4636

Variable	Simple Statistics	
	Minimum	Maximum
Groesse	156.00000	196.00000
Gewicht_vor	46.00000	106.00000

**Pearson Correlation Coefficients**  
Number of Observations

	Groesse	Gewicht
alter	0.12317 71	-0.05018 70

- Option **SPEARMAN**: berechnet den Rangkorrelationskoeffizienten nach Spearman

```
TITLE1 'Korrelationskoeffizient nach Spearman' ;
TITLE2 '-----' ;
PROC CORR DATA=beispiel SPEARMAN ;
VAR VD CRP_postop ;
RUN ;
```

⇒ SAS zeigt, wie beim Pearson, die statistischen Maßzahlen, den Korrelationskoeffizienten (dieses Mal nach Spearman), den zugehörigen p-Wert und die zugrunde liegende Fallzahl an:

Zusammenhang Verweildauer und CRP (nach SPEARMAN)

The CORR Procedure

```
1 With Variables: VD
2 Variables: CRP
```

Variable	N	Simple Statistics		
		Mean	Std Dev	Median
VD	15	8.40000	4.33919	7.00000
CRP	15	147.46667	98.00282	143.00000

Variable	Simple Statistics	
	Minimum	Maximum
VD	4.00000	21.00000
CRP_postop	20.00000	300.00000

**Spearman Correlation Coefficients, N=15**  
Prob > |r| under H0: Rho=0

	CRP_postop
VD	0.66249 0.0071