

PROC FREQ für Kontingenztafeln

zum Überprüfen von Zusammenhängen zweier qualitativer Merkmale

Allgemeine Form:

```
PROC FREQ DATA=name Optionen ;  
    TABLES (variablenliste) * (variablenliste) / Optionen ;  
RUN ;
```

Beispiel und Beschreibung der Programm-Statements:

```
TITLE1 'Haeufigkeitsvergleiche mit dem Chi-Quadrat-Test' ;  
TITLE2 '-----' ;  
PROC FREQ DATA=beispiel PAGE ;  
    TABLES rhesus*blutgruppe / CHISQ EXACT ;  
RUN ;
```

Die Prozedur beginnt mit PROC endet mit RUN ;

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

FREQ ist der Name der Prozedur.

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

TABLES ist das gehörende Schlüsselwort und steht für Tabellen.

Es sind beliebig viele TABLES-Statements in einer PROC FREQ möglich.

Die Aufzählung der Merkmalsnamen darf in einem TABLES-Statement auch über mehrere Zeilen gehen. Das Semikolon kommt dann nach der letzten Variablen.

variablenliste im TABLES-Statements wird mit den Namen der zu verarbeitenden Merkmale überschrieben. Trennen Sie die Merkmalsnamen mit Leerzeichen!!!

Die Ausprägungen der Merkmale **links vom * werden als Zeilen** und die Ausprägungen der Merkmale **rechts vom * als Spalten** in der Kontingenztafel aufgeführt. Setzen Sie zwei Merkmale und mehr in Klammern.

SAS erstellt für jeden Vergleich eine eigene Tabelle.

Option PAGE fest, dass SAS beim Drucken für jede Tabelle eine neue Seite beginnt, so dass die Kontingenztafel und die zugehörige Teststatistik auf einer Seite ausgegeben werden.

Option CHISQ führt den χ^2 -Test und Option EXACT den exakten Fisher-Test durch. Bei Vierfeldertafeln berechnet SAS den exakten Fisher-Test automatisch. Verwenden Sie die Option Exact nur dann zusätzlich, wenn SAS bei der Anwendung des χ^2 -Tests eine Fehlermeldung ausgibt.

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.

Eine Beschreibung der Ergebnisse der Prozedur, die SAS im Output-Fenster zeigt, finden Sie auf der Rückseite.

Die folgende Tabelle zeigt die Blutgruppenverteilung bei RH+ und Rh- in Form einer Kontingenztafel. Die Ausprägungen des Merkmals *Rhesus* den Ausprägungen des Merkmals *Blutgruppe* wurden gegenüber gestellt und die dabei entstandenen Kombinationen als Zellen gezeichnet. Im Anschluss an die Tabelle folgt die erstellte Teststatistik mit den p-Werten für den Chi-Quadrat-Test und den exakten Fisher-Test:

Häufigkeitsvergleiche mit dem Chi-Quadrat-Test

The FREQ Procedure

Table of rhesus by blutgruppe

Rhesus

Blutgruppe

| | A | AB | B | O | Total |
|--------------|-------------------------------|-----------------------------|-----------------------------|-------------------------------|-------|
| + | 27 38.03 45.00 87.10 | 3 4.23 5.00 100.00 | 7 9.86 11.67 77.78 | 23 32.39 38.33 82.14 | 60 |
| - | 4 5.63 36.36 12.90 | 0 0.00 0.00 0.00 | 2 2.82 18.18 22.22 | 5 7.04 45.45 17.86 | 11 |
| Total | 31 | 3 | 9 | 28 | 71 |

Statistics for Table of Rhesus by blugruppe

| Statistic | DF | Value | Prob |
|-----------------------------|----------|---------------|---------------|
| Chi-Square | 3 | 1.1396 | 0.7675 |
| Likelihood Ratio Chi-Square | 3 | 1.5728 | 0.6656 |
| Mantel-Haenszel Chi-Square | 1 | 0.4142 | 0.5198 |
| Phi Coefficient | | 0.1267 | |
| Contingency Coefficient | | 0.1257 | |
| Cramer's V | | 0.1267 | |

WARNING: 63% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Fisher's Exact Test

| | |
|-----------------------|---------------|
| Table Probability (P) | 0.0435 |
| Pr <= P | 0.8394 |

Sample Size = 71

Beschreibung des OUTPUT-Fensters am obigen Beispiel:

Frequency = absolute Häufigkeit:

Von insgesamt 60 Personen mit Rh+ hatten 27 Blutgruppe A, 3 Blutgruppe AB, 7 Blutgruppe B und 23 Blutgruppe 0.

Von 11 Personen mit Rh- 4 Blutgruppe A, 0 Blutgruppe AB, 2 Blutgruppe B und 5 Blutgruppe 0.

Percent = relative Häufigkeiten (ausgehend von der Gesamtzahl der Befragten, im Beispiel n=71):

38,03% der Personen gaben an Blutgruppe A zu haben, 4,23% Blutgruppe B, 9,86% Blutgruppe B und 32,39% Blutgruppe 0.

Von den Personen mit Rh- hatten 5,63% Blutgruppe B, 0% Blutgruppe AB, 2,82% Blutgruppe B und 7,04% Blutgruppe 0.

Row Pct = Zeilenprozentzahl, Prozentzahlen bezogen auf die Zeilensumme (ausgehend vom Rhesusfaktor, n=60 Rh+ und n=11 Rh-):

45,00% der Personen mit Rh+ haben Blutgruppe A, 5% Blutgruppe AB, 11,67% Blutgruppe B und 38,33% Blutgruppe 0.

36,36% der Personen mit Rh- gaben an Blutgruppe A zu haben, 0% Blutgruppe AB, 18,18% Blutgruppe B und 45,45% Blutgruppe 0.

Col Pct = Spaltenprozentzahl, Prozentzahlen bezogen auf die Spaltensumme (ausgehend von der Blutgruppe, n=31 Blutgruppe A, n=3 Blutgruppe AB, n=9 Blutgruppe B und n=28 viel Blutgruppe 0):

87,1% der Personen mit Blutgruppe A gaben an Rh+ zu sein und 12,9% Rh-.

Alle Personen mit Blutgruppe AB waren Rh+.

77,78% der Personen mit Blutgruppe B gaben an Rh+ zu sein und 22,22% Rh-.

82,14% der Personen mit Blutgruppe 0 gaben an Rh+ zu sein und 17,86% Rh-.

Total = Zeilen- bzw. Spaltensummen:

Die Zahlen am unteren und rechten Rand der Tabelle sind die absoluten Häufigkeiten für die Zeilen und Spalten und die Gesamtzahlen (71) für die Tabelle.

Chi-Quadrat-Test:

Der Wert **Prob=0,7675** ist der p-Wert für den durchgeführten Chi-Quadrat-Test, die Zahl **Value=1,1396** gibt die Testgröße (Summe der χ^2 -Anteile) an und **DF=3** die Freiheitsgrade (degrees of freedom). Der berechnete p-Wert zeigt ein nicht-signifikantes Testergebnis an, denn er ist größer als 0,05.

Mit **WARNING: 63% of the cells...** weist SAS darauf hin, dass der durchgeführte Test nicht verwendet werden sollte, da bei fast zwei Dritteln der Zellen (63%) die erwarteten Häufigkeiten zu klein sind. Um ein korrektes Testergebnis zu erhalten, berechnet man daher zusätzlich den **exakten Fischer-Test** (wird bei Vierfelder-Tafeln automatisch berechnet).

Der **Fisher-Test** ist ein "exakter" Test, weil die Prüfgröße im Gegensatz zum Chi-Quadrat-Test direkt berechnet wird. Im Beispiel wäre das Testergebnis mit **p=0,8394** nicht-signifikant, weil der p-Wert größer als 5% ist.

Beschreibung weiterer nützlicher Optionen und Statements:

- Option **NOPERCENT**: SAS unterlässt die Ausgabe der relativen Häufigkeiten und der relativen Summenhäufigkeiten:

```
PROC FREQ DATA=beispiel ;
  TABLES alkohol * rauchen / NOPERCENT ;
RUN ;
```

- Statement **WEIGHT variablenname** dient zur Angabe einer Variablen, die Angabe für die Häufigkeiten enthält. Die Ausprägungen des Merkmals mit den Häufigkeiten (im Beispiel *anzahl*) werden als absolute Häufigkeiten eingesetzt, als Häufigkeiten ausgegeben und die gewählten Tests durchgeführt:

| brille | sitzplatz | anzahl |
|--------|-----------|--------|
| ja | vorne | 5 |
| ja | hinten | 6 |
| nein | vorne | 11 |
| nein | hinten | 12 |

```
TITLE1 'Vergleich zweier qualitativer Merkmale' ;
TITLE2 '-----' ;
PROC FREQ DATA=beispiel PAGE ;
  TABLES brille * sitzplatz / CHISQ EXACT EXPECTED NOPERCENT ;
  WEIGHT anzahl ;
RUN ;
```

- ⇒ SAS erstellt eine Kontingenztafel mit den Merkmalen *brille* und *sitzplatz*. Merkmal *anzahl* enthält die Häufigkeiten für das Vorkommen der Kombinationen der Ausprägungen der Variablen *brille* und *sitzplatz* (siehe Datentabelle oben):

The FREQ Procedure

Table of brille by sitzplatz

brille sitzplatz

| Frequency Expected Row Pct Col Pct | sitzplatz | | Total |
|---|--------------------------------|--------------------------------|-------|
| | vorne | hinten | |
| ja | 5 5.1765 45.45 31.25 | 6 5.8235 54.55 33.33 | 11 |
| nein | 11 10.824 47.83 68.75 | 12 12.176 52.17 66.67 | 23 |
| Total | 16 | 18 | 34 |

Statistics for Table of brille by sitzplatz

| Statistic | DF | Value | Prob |
|-----------------------------|----|--------|--------|
| Chi-Square | 1 | 0.0168 | 0.8969 |
| Likelihood Ratio Chi-Square | 1 | 0.0168 | 0.8968 |
| Continuity Adj. Chi-Square | 1 | 0.0000 | 1.0000 |
| Mantel-Haenszel Chi-Square | 1 | 0.0163 | 0.8984 |

```
Phi Coefficient          -0.0222
Contingency Coefficient  0.0222
Cramer's V              -0.0222
```

Fisher's Exact Test

```
Cell (1,1) Frequency (F)      5
Left-sided Pr <= F            0.5945
Right-sided Pr >= F           0.6890

Table Probability (P)         0.2834
Two-sided Pr <= P            1.0000
```

Sample Size = 34

- Statement **BY** **variablenliste** sortiert die Daten nach der/den hinter dem Statement angegebenen Variablen vor und ermöglicht so die Unterscheidung der Tabellen und der Berechnungen nach einer Gruppenvariablen. Die Variablen in der **BY**-Anweisung müssen qualitativ sein!!!

Bei Verwendung des Statements **BY** müssen die Daten über die Prozedur **SORT** vorsortiert werden. SAS zeigt die Datensätze sortiert nach der/den hinter **BY** aufgezählten Variablen an:

```
PROC SORT DATA=beispiel ; BY geschlecht ;
PROC FREQ DATA=beispiel ;
    TABLES rauchen*alkoholg / NOPERCENT ;
    BY geschlecht ;
RUN ;
```

- ⇒ SAS erstellt zwei Kontingenztafeln mit den Merkmalen rauchen und alkohol und berücksichtigt in der einen nur die Frauen (*Geschlecht=weiblich*) und in der anderen nur die Männer (*Geschlecht=männlich*) ohne Prozentangaben.

Geschlecht=weiblich

The FREQ Procedure

Table of alkoholg by RAUCHEN

alkoholg RAUCHEN(Rauchen)

| Frequency Row Pct Col Pct | nein | selten | maessig | stark | Total |
|---------------------------------|----------------------|---------------------|---------------------|---------------------|-------|
| kein | 32 71.11 74.42 | 5 11.11 62.50 | 7 15.56 70.00 | 1 2.22 100.00 | 45 |
| maessig | 10 62.50 23.26 | 3 18.75 37.50 | 3 18.75 30.00 | 0 0.00 0.00 | 16 |
| viel | 1 100.00 2.33 | 0 0.00 0.00 | 0 0.00 0.00 | 0 0.00 0.00 | 1 |
| Total | 43 | 8 | 10 | 1 | 62 |

Geschlecht=maennlich

The FREQ Procedure

Table of alkoholg by RAUCHEN

alkoholg RAUCHEN(Rauchen)

| Frequency Row Pct Col Pct | nein | selten | maessig | stark | Total |
|---------------------------------|----------------------|---------------------|---------------------|---------------------|-------|
| kein | 12 80.00 54.55 | 1 6.67 25.00 | 1 6.67 100.00 | 1 6.67 25.00 | 15 |
| maessig | 9 75.00 40.91 | 2 16.67 50.00 | 0 0.00 0.00 | 1 8.33 25.00 | 12 |
| viel | 1 25.00 4.55 | 1 25.00 25.00 | 0 0.00 0.00 | 2 50.00 50.00 | 4 |
| Total | 22 | 4 | 1 | 4 | 31 |

- Option **CMH** berechnet Mantel-Haenszel-Statistiken. SAS gibt zusammenfassende relative Risiken und Chancen mit Konfidenzintervallen aus und ermittelt die Homogenität der relativen Chancen mit dem Breslow-Test (mit einem Statistiker besprechen).
- Option **EXPECTED** druckt die Erwartungswerte (erwartete Zellhäufigkeiten) für den χ^2 -Test in die entsprechenden Zellen der Tabelle.
- Option **AGREE** wendet den McNemar-Test an und berechnet den Koeffizienten nach Kappa (wird separat erklärt).