

PROC MEANS

zum Berechnen statistischer Maßzahlen
(für **quantitative** Merkmale)

Allgemeine Form:

```
PROC MEANS DATA=name Optionen ;  
    VAR variablenliste ;  
RUN ;
```

Beispiel und Beschreibung der Programm-Statements:

```
TITLE1 'Statistische Maßzahlen' ;  
TITLE2 '-----' ;  
PROC MEANS DATA=beispiel MIN MAX MEAN MEDIAN STD MAXDEC=2 ;  
    VAR groesse gewicht ;  
RUN ;
```

Die Prozedur beginnt mit PROC endet mit RUN ;

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

MEANS ist der Name der Prozedur.

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

VAR ist das gehörende **Schlüsselwort** und steht für die Liste der Merkmalsnamen.

Ohne **VAR** - Statement berechnet SAS die statistischen Maßzahlen für **alle numerischen** Merkmale in der Datendatei.

Es sind beliebig viele **VAR**-Statements in einer PROC MEANS möglich.

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

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

SAS berechnet hier die statistischen Maßzahlen Minimum (**MIN**), Maximum (**MAX**), Mittelwert (**MEAN**), Median (**MEDIAN**) und Standardabweichung (**STD**) mit jeweils zwei Nachkommastellen (**MAXDEC=2**). Ohne Angabe ausgewählter Maßzahlen und einer Vorgabe von Dezimalstellen gibt SAS die statistischen Maßzahlen N (Anzahl Beobachtungen abzüglich fehlender Werte), Mittelwert, Standardabweichung, Minimum und Maximum mit jeweils sieben Nachkommastellen aus.

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 Beispiel, zeigt SAS im Output-Fenster folgende Informationen:

Statistische Masszahlen

The MEANS Procedure

Variable	Minimum	Maximum	Mean	Median	Std Dev
Groesse	156.00	196.00	173.00	173.00	8.83
Gewicht	46.00	106.00	66.23	64.00	12.46

Beschreibung des Outputs:

Die berechneten Maßzahlen beziehen sich auf die Gesamtzahl der Beobachtungen abzüglich der Datensätze mit fehlenden Werten.

Variable zeigt die Namen der verwendeten Variablen, im Beispiel: Groesse und Gewicht.

Minimum = kleinster Wert.

Merkmal Groesse: 156,00 (cm)

Merkmal Gewicht: 46,00 (kg)

Maximum = größter Wert.

Merkmal Groesse: 196,00 (cm)

Merkmal Gewicht: 106,00 (kg)

Mean = Mittelwert (arithmetisches Mittel).

Der Mittelwert ist ein Lagemaß, das nur bei quantitativen Merkmalen erlaubt ist. Er wird im Gegensatz zum Median stark von Ausreißern beeinflusst und sollte deshalb nur bei normalverteilten Merkmalen berechnet werden.

Merkmal Groesse: 173,00 (cm)

Merkmal Gewicht: 66,23 (kg)

Median = mittlerer Wert der sortierten Stichprobe.

Der empirische Median ist ein Lagemaß, das die nach ihrer Größe geordnete Stichprobe in 2 Hälften teilt. Bei ungeradem Stichprobenumfang ist der Median der mittlere Wert; bei geradem Stichprobenumfang der Durchschnitt aus den beiden mittleren Werten.

Merkmal Groesse: 173,00 (cm)

Merkmal Gewicht: 64,00 (kg)

Std Dev = Standardabweichung.

Wurzel aus der Varianz. Bevorzugtes Streuungsmaß bei quantitativ stetigen, symmetrisch verteilten Daten. Die Standardabweichung darf nur bei quantitativen Merkmalen berechnet werden. Bei normalverteilten Daten liegen etwa 2/3 der Messwerte innerhalb des Intervalls $\mu \pm \sigma$ und etwa 95 % innerhalb des Intervalls $\mu \pm 2\sigma$.

Merkmal Groesse: 8,83 (cm)

Merkmal Gewicht: 12,46 (kg)

Beschreibung nützlicher Optionen:

- **Option N:** SAS gibt die der Berechnung zugrunde liegende Anzahl aus:

```
TITLE1 'Statistische Maßzahlen' ;
TITLE2 '-----' ;
PROC MEANS DATA=beispiel N MIN MAX MEAN MEDIAN STD MAXDEC=2 ;
VAR groesse gewicht ;
RUN ;
```

- ⇒ SAS gibt zusätzlich zu den Maßzahlen aus, wie viele Beobachtungen der Berechnung zugrunde liegen:

```
Statistische Masszahlen
-----
```

The MEANS Procedure

Variable	N	Minimum	Maximum	Mean	Median	Std Dev
Groesse	71	156.00	196.00	173.00	173.00	8.83
Gewicht	70	46.00	106.00	66.23	64.00	12.46

Im Beispiel fehlt eine Gewichtsangabe.

- Option **MAXDEC=zahl:** SAS gibt statt sieben Nachkommastellen nur so viel aus wie **zahl** angibt, z.B. MAXDEC=2 ⇒ 2 Nachkommastellen:

Beispiel ohne Vorgabe für die Anzahl von Nachkommastellen:

```
TITLE1 'Statistische Maßzahlen' ;
TITLE2 '-----' ;
PROC MEANS DATA=beispiel ;
VAR groesse gewicht ;
RUN ;
```

- ⇒ SAS gibt sieben Nachkommastellen aus:

```
Statistische Masszahlen
-----
```

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
Groesse	71	173.0000000	8.8285252	156.0000000	196.0000000
Gewicht_vor	70	66.2285714	12.4634787	46.0000000	106.0000000

Beispiel mit Vorgabe für die Anzahl von Nachkommastellen:

```
TITLE1 'Statistische Maßzahlen' ;
TITLE2 '-----' ;
PROC MEANS DATA=beispiel MAXDEC=2 ;
VAR groesse gewicht ;
RUN ;
```

⇒ SAS gibt nur noch zwei Nachkommstellen aus:

Statistische Masszahlen

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
Groesse	71	173.00	8.83	156.00	196.00
Gewicht_vor	70	66.23	12.46	46.00	106.00

- Weitere statistische Kenngrößen, die man als Optionen angeben kann:

VAR = Varianz: mittlere quadratische Abweichung der Stichprobenwerte vom Mittelwert.

RANGE = Spannweite: Maximum - Minimum.

CV = Variationskoeffizienten: Quotient von Standardabweichung und arithmetischem Mittel.

SKEWNESS = Schiefe: Verschiebung der Merkmalsausprägungen nach links oder rechts vom Mittelwert aus gesehen.

Q1 = 1. Quartil und **Q3** = 3. Quartil, **P10** = 10. Perzentil und **P90** = 90. Perzentil.

CLM berechnet die obere bzw. untere Grenze des 95%-Konfidenzintervall bei $\alpha=5\%$ für den Mittelwert. Innerhalb dieses Bereichs liegt der Erwartungswert der Grundgesamtheit mit 95%-iger Wahrscheinlichkeit.