

# PROC TTEST

zum Durchführen des t-Tests für zwei unverbundene Stichproben  
(für **quantitative normalverteilte** Merkmale)

Allgemeine Form:

```
PROC TTEST DATA=name ;  
  VAR variable ;  
  CLASS vergleichsvariable ;  
RUN ;
```

Beispiel und Beschreibung der Programm-Statements:

```
TITLE1 't-Test für zwei unverbundene Stichproben nach Rauchen' ;  
TITLE2 '-----' ;  
PROC TTEST DATA=beispiel ;  
  VAR groesse ;  
  CLASS rauchen ;  
RUN ;
```

Die Prozedur beginnt mit PROC endet mit RUN ;

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

TTEST 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.

**variable** wird mit dem Namen des zu verarbeitenden Merkmals überschrieben. **Geben Sie hier immer nur ein Merkmal an, da die berechneten Ergebnisse sonst sehr unübersichtlich ausgegeben werden.**

Mit dem CLASS-Statement definieren Sie **eine** Gruppenvariable. SAS vergleicht die quantitativen Merkmale in der Variablenliste nach den Ausprägungen dieses Merkmals. **Die Gruppenvariable muss qualitativ sein.**

**vergleichsvariable** ist der Name des **qualitativen** Merkmals, nach dessen Ausprägungen das quantitative Merkmal unterschieden werden sollen.

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.

SAS gibt neben dem p-Wert für den t-Test (unter  $Pr > |t|$ ) auch den p-Wert für den F-Test (unter  $PR > F$ ) aus und zeigt neben vielen anderen Werten die Mittelwerte und Standardabweichungen der Vergleichsgruppen und die mittlere Differenz mit zugehörigen Konfidenzintervallen.

Starten Sie die Programmzeilen aus dem Beispiel, zeigt SAS im Output-Fenster folgende Informationen:

t-Test fuer zwei unverbundene Stichproben nach Rauchen

-----  
The TTEST Procedure

Variable: **Groesse**

① Rauchen	N	Mean	Std Dev	Std Err	Minimum	Maximum
ja	13	173.5	9.2070	2.5536	157.0	193.0
nein	58	172.9	8.8200	1.1581	156.0	196.0
Diff (1-2)		0.6592	8.8885	2.7275		

② Rauchen	Method	Mean	95% CL Mean	Std Dev
Ja		173.5	168.0 179.1	9.2070
Nein		172.9	170.6 175.2	8.8200
Diff (1-2)	Pooled	0.6592	-4.7822 6.1005	8.8885
Diff (1-2)	Satterthwaite	0.6592	-5.2490 6.5673	

③ Rauchen	Method	95% CL	Std Dev
Ja		6.6022	15.1984
Nein		7.4565	10.7984
Diff (1-2)	Pooled	7.6212	10.6654
Diff (1-2)	Satterthwaite		

④ Method	Variances	DF	t Value	Pr >  t
Pooled	Equal	69	0.24	0.8098
Satterthwaite	Unequal	17.29	0.24	0.8169

Equality of Variances

⑤ Method	Num DF	Den DF	F Value	Pr > F
Folded F	12	57	1.09	0.7723

## Beschreibung des Outputs:

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

Teil ① zeigt den Namen des zu überprüfenden Merkmals (im Beispiel: **Groesse**) und Namen und Ausprägungen des Vergleichsmerkmals (im Beispiel: **Rauchen** mit den Ausprägungen **ja** und **nein**).

Zusätzlich gibt SAS Anzahl der Beobachtungen und statistische Maßzahlen getrennt nach den beiden Vergleichsgruppen aus:

**N = Anzahl der Beobachtungen:** 13 Raucher (Ausprägung **ja**), 58 Nichtraucher (Ausprägung **nein**).

**Mean = Mittelwerte bzw. mittlere Differenz:** Raucher 173,5 cm (Ausprägung **ja**), Nichtraucher 172,9 cm (Ausprägung **nein**), mittlere Differenz: 0,6592.

*Die mittlere Differenz ergibt sich aus der Differenz der beiden Mittelwerte. SAS stellt damit die bestehenden Mittelwertsunterschiede, d.h. die Effektgröße dar.*

**Std Dev = Standardabweichung bzw. Standardabweichung über beide Gruppen (bei Varianzhomogenität):** Raucher 9,207 cm (Ausprägung **ja**), Nichtraucher 8,82 cm (Ausprägung **nein**), Standardabweichung über beide Gruppen: 8,8885.

*Die Standardabweichung über beide Gruppen berechnet über die Wurzel aus der Summe der gewichteten Gruppenvarianzen durch die Fallzahl-2, im Beispiel nach der Formel:*

$$\sqrt{((n_1 - 1) * VarGruppe1 + (n_2 - 1) * VarGruppe2) / (Fallzahl - 2)} =$$

$$\sqrt{(12 * 9,2070^2 + 57 * 8,8200^2) / (13 + 58 - 2)}$$

**Std Err = Standardfehler des Mittelwerts bzw. Standardfehler über beide Gruppen:** Raucher 2,5536 cm (Ausprägung **ja**), Nichtraucher 1,1581 cm (Ausprägung **nein**), Standardfehler über beide Gruppen: 2,7275.

**Minimum = kleinster Wert:** Raucher 157 cm (Ausprägung **ja**), Nichtraucher 156 cm (Ausprägung **nein**).

**Maximum = größter Wert:** Raucher 193 cm (Ausprägung **ja**), Nichtraucher 196 cm (Ausprägung **nein**).

In **Teil ②** finden Sie neben den nach den beiden Vergleichsgruppen aufgeschlüsselten **Mittelwerten** und die **Standardabweichungen**, die **mittleren Differenz** getrennt nach Auswertungsmethode, die **mittlere Standardabweichung** für Methode Pooled und die **95%-Konfidenzintervalle** (unter **95% CL Mean**) für die Mittelwerte bzw. die mittlere Differenz aufgeschlüsselt nach Auswertungsmethode.

Rauchen	Method	Mean	95% CL Mean		Std Dev
Ja		173.5	168.0	179.1	9.2070
Nein		172.9	170.6	175.2	8.8200
Diff (1-2)	Pooled	0.6592	-4.7822	6.1005	8.8885
Diff (1-2)	Satterthwaite	0.6592	-5.2490	6.5673	

*Überschneiden sich die Konfidenzintervalle der beiden Gruppen oder enthält das Konfidenzintervall für die mittlere Differenz die Null, kann man davon ausgehen, dass sich kein signifikantes Testergebnis ergibt.*

In **Teil ③** gibt SAS die **95%-Konfidenzintervalle (95% CL Std Dev)** für die Standardabweichungen der beiden Vergleichsgruppen und die Standardabweichung über beide Gruppen für Methode Pooled aus.

Rauchen	Method	95% CL Std Dev	
Ja		6.6022	15.1984
Nein		7.4565	10.7984
Diff (1-2)	Pooled	7.6212	10.6654
Diff (1-2)	Satterthwaite		

**Teil ④** entnehmen Sie **Teststatistik für den t-Test für zwei unverbundene Stichproben**. Je nachdem, ob der nachfolgende F-Test signifikant wird oder nicht, wählt man hier die Methode **Satterthwaite** (ungleiche Varianzen, modifizierter t-Test) bzw. die Methode **Pooled** (gleiche Varianzen werden angenommen). Die Zahlen unter **t Value** zeigen die

jeweiligen Testgrößen und die unter **Pr > |t|** den zugehörigen p-Wert. **DF** sind die Freiheitsgrade, die mit der Prüfverteilung zusammenhängen.

Im Beispiel zeigt der F-Test ein **nicht-signifikantes** Ergebnis, deshalb gelten die Angaben unter der Methode **Pooled** mit **Testgröße = 0.24** (unter **t Value**) und **p-Wert = 0,8098** (unter **Pr > |t|**).

Method	Variances	DF	t Value	Pr >  t
<b>Pooled</b>	<b>Equal</b>	<b>69</b>	<b>0.24</b>	<b>0.8098</b>
Satterthwaite	Unequal	17.29	0.24	0.8169

**Teil 5** zeigt die Ergebnisse für den **F-Test**, der jedem t-Test vorangesetzt wird. SAS überprüft mit dem **F-Test** die Gleichheit der Varianzen der beiden Vergleichsgruppen.

Die Zahl unter **F Value** zeigt die ermittelte Testgröße (im Beispiel: **1,09**), die unter **Num DF** (im Beispiel **12**) die Anzahl der kleineren Stichprobe - 1, die unter **Den DF** (im Beispiel **57**) die Anzahl der größeren Stichprobe - 1 und die Zahl unter **PR > F** den berechneten p-Wert (im Beispiel: **0,7723**).

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	12	57	1.09	0.7723

Im Beispiel liefert der **F-Test** **kein signifikantes Testergebnis**. Also sind die **Varianzen gleich** (*Equal*) daher müssen Sie unter der Statistik für den t-Test die zur Methode *Pooled* gemachten Angaben als Testergebnis für den t-Test verwenden.

Bei einem **signifikanten Testergebnis für den F-Test**, würde man davon ausgehen, dass die **Varianzen der beiden Vergleichsgruppen unterschiedlich** (*Unequal*) sind. Als p-Wert für den t-Test wäre dann der Wert in der Zeile mit den Stichworten *Satterthwaite* (unter *Method*) und *Unequal* (unter *Variances*) der richtige gewesen.