



STECKBRIEF

t-Test für unabhängige Stichproben

WORUM GEHT'S?

Um Mittelwerts-Unterschiede bei zwei unabhängigen Gruppen.

IN WELCHER WELT DER STATISTIK SIND WIR?

Inferenzstatistik, Welt der Unterschiede

BEISPIELE

Männer weisen durchschnittlich höhere Ausprägungen in „Thrill & Sensation Seeking“ auf als Frauen. (gerichtet: rechtsseitig)

Die Risiko-Aversion ist bei Menschen im mittleren Erwachsenenalter niedriger ausgeprägt als bei Menschen über 60. (gerichtet: linksseitig)

Depressive Menschen, die dreimal pro Woche 60min Sport treiben, unterscheiden sich in ihren Depressions-Scores von Depressiven, die keinen Sport machen. (ungerichtet / zweiseitig)

HINWEISE:

- Es können sowohl gerichtete als auch ungerichtete Hypothesen getestet werden
- Das Ergebnis ist ein empirischer t -Wert, der in die t -Verteilung, die dazugehörige Testverteilung fällt
- **Bei der Ergebnis-Interpretation von gerichteten Hypothesen ist darauf zu achten, ob die Ergebnisse auch wirklich in die postulierte Richtung gehen!**
- Nur ein Messzeitpunkt
- Synonyme: Ungepaarter t -Test, t -Test für unverbundene Stichproben

VORAUSSETZUNGEN

- Unabhängige Stichproben
- Die UV ist eine kategoriale / diskrete Variable
- Die AV ist metrisch
- Die AV ist in beiden Gruppen normalverteilt
- Varianzhomogenität (= Varianzgleichheit) der AV in beiden Gruppen
- Der t -Test ist relativ robust gegenüber Voraussetzungsverletzungen, solange die Gruppen annähernd ähnlich groß und die Stichproben pro Gruppe nicht sehr klein sind (klein = unter 30)

ALTERNATIVES NON-PARAMETRISCHES VERFAHREN

Mann-Whitney- U -Test = U -Test = Wilcoxon-Mann-Whitney-Test = Wilcoxon-Rangsummen-Test

BERECHNUNG PER HAND

$$t = \frac{\bar{x}_A - \bar{x}_B}{\hat{\sigma}_{\bar{x}_A - \bar{x}_B}}$$

Im Nenner steht der Standardfehler der Mittelwerts-Differenz.

Hierfür gibt es zwei Formeln: einmal für ungleiche und einmal für gleich große Stichproben.



Formel für ungleiche Gruppengrößen:

$$\hat{\sigma}_{\bar{x}_A - \bar{x}_B} = \sqrt{\frac{(n_A - 1) \cdot \hat{\sigma}_A^2 + (n_B - 1) \cdot \hat{\sigma}_B^2}{(n_A - 1) + (n_B - 1)} \cdot \left(\frac{1}{n_A} + \frac{1}{n_B}\right)}$$

Formel für gleich große Gruppen:

$$\hat{\sigma}_{\bar{x}_A - \bar{x}_B} = \sqrt{\frac{\hat{\sigma}_A^2}{n} + \frac{\hat{\sigma}_B^2}{n}}$$

Bestimmung der Freiheitsgrade:

- Gesamt-Stichprobe minus 2 = $N - 2$
- Oder (äquivalent): Pro Gruppe von der Gruppengröße jeweils einen Freiheitsgrad abziehen

EFFEKTSTÄRKE COHENS d

Formel:

$$d = \frac{\mu_A - \mu_B}{\hat{\sigma}_{AB}} = \frac{\bar{x}_A - \bar{x}_B}{\hat{\sigma}_{AB}}$$

Im Nenner steht die gepoolte Standardabweichung.

Hierfür gibt es auch wieder zwei Formeln: einmal für ungleiche und einmal für gleich große Stichproben.

Formel für ungleiche Gruppengrößen:

$$\hat{\sigma}_{AB} = \sqrt{\frac{\hat{\sigma}_A^2 n_A + \hat{\sigma}_B^2 n_B}{n_A + n_B}}$$

Formel für gleich große Gruppen:

$$\hat{\sigma}_{AB} = \sqrt{\frac{\hat{\sigma}_A^2 + \hat{\sigma}_B^2}{2}}$$

Konventionen für die Interpretation von d (Cohen, 1988):

- Kleiner Effekt: $|d| = 0,2$
- Mittlerer Effekt: $|d| = 0,5$
- Großer Effekt: $|d| = 0,8$

ACHTUNG: jeweils **IM BETRAG!**



BERECHNUNG MIT SPSS

H_1 : Frauen weisen direkt nach dem Urlaub durchschnittlich ein höheres Stresslevel auf als Männer. (gerichtet: rechtsseitig)

Gruppenstatistiken					
	Geschlecht	N	Mittelwert	Std.-Abweichung	Standardfehler des Mittelwertes
Stresslevel direkt nach dem Urlaub	weiblich	51	46,7059	17,10239	2,39481
	männlich	54	40,9444	15,84407	2,15610

		Levene-Test der Varianzgleichheit		t-Test für die Mittelwertgleichheit							
		F	Sig.	T	df	Signifikanz Einseitiges p	Zweiseitiges p	Mittlere Differenz	Differenz für Standardfehler	95% Konfidenzintervall der Differenz	
										Unterer Wert	Oberer Wert
Stresslevel direkt nach dem Urlaub	Varianzen sind gleich	,209	,648	1,792	103	,038	,076	5,76144	3,21533	-,61540	12,13828
	Varianzen sind nicht gleich			1,788	101,188	,038	,077	5,76144	3,22241	-,63081	12,15369

Effektgrößen bei unabhängigen Stichproben					
		Standardisierte r^a	Punktschätzung	95% Konfidenzintervall	
			g	Unterer Wert	Oberer Wert
Stresslevel direkt nach dem Urlaub	Cohen's d	16,46692	,350	-,037	,735
	Hedges' Korrektur	16,58805	,347	-,036	,729
	Glass' Delta	15,84407	,364	-,027	,751

a. Der bei der Schätzung der Effektgrößen verwendete Nenner.
Cohen's d verwendet die zusammengefasste Standardabweichung.
Hedges' Korrektur verwendet die zusammengefasste Standardabweichung und einen Korrekturfaktor.
Glass' Delta verwendet die Standardabweichung einer Stichprobe von der Kontrollgruppe.

SPSS WIRFT STANDARDMÄSSIG DREI BEREICHE AUS:

- Zunächst die deskriptiven Statistiken, wovon uns besonders die Mittelwerte interessieren
- Dann folgt in der nächsten Ausgabe links der Levene-Test zur Überprüfung, ob die Voraussetzung der Varianzhomogenität gegeben ist, sowie direkt angrenzend rechts der „eigentliche“ t -Test.
In SPSS ist die Trennung beider Tests nicht so gut ersichtlich, daher habe ich einen grünen Trennstrich hinzugefügt.
- Als letztes folgen die verfügbaren Effektgrößen.

SO Liest DU DIE ERGEBNISSE AB:

1. Schritt: Mittelwerte ablesen

Bei den „Gruppenstatistiken“ sieht man in der Spalte **Mittelwert**, dass die Ergebnisse in die erwartete Richtung gehen: Der Mittelwert der Frauen (46,71) ist höher als derjenige der Männer (40,94). Wir können also fortfahren.

2. Schritt: Varianzhomogenität prüfen

Als nächstes liest man im mittleren Output auf der linken Seite den **Levene-Test der Varianzgleichheit** ab.

Dieser testet die Nullhypothese, dass die Varianzen in beiden Gruppen gleich sind.

Ein signifikantes Ergebnis bedeutet, dass KEINE Varianzhomogenität vorliegt!

Daher müsste man in diesem Falle das Ergebnis des t -Tests in der unteren Zeile ablesen: bei „Varianzen sind nicht gleich“.

Dort wird dann der Welch-Test durchgeführt. Dieser unterscheidet sich vom t -Test dadurch, dass eine Korrektur der Freiheitsgrade vorgenommen wird. Das kannst du in der Spalte df sehen: Beim t -Test sind es 1,792 Freiheitsgrade, beim Welch-Test 1,788.

Ist der Levene-Test nicht signifikant, wird das Ergebnis des t -Tests in der oberen Zeile „Varianzen sind gleich“ abgelesen.



3. Schritt: Ergebnis ablesen

Wegen des nicht signifikanten p -Werts von 0.648 des Levene-Tests liest man das **Ergebnis des t-Tests** in der oberen Zeile bei **Einseitiges p** ab (wir haben eine einseitige, rechtsseitige Hypothese).

Der p -Wert ist 0,038 und somit auf einem Signifikanzniveau von 0,05 signifikant.

HINWEIS:

Bei älteren Versionen von SPSS wird nur die zweiseitige Testung angezeigt.

Der p -Wert muss dann bei Ergebnissen, die in die erwartete Richtung gehen, halbiert werden!

4. Schritt: Bei signifikantem Ergebnis die Effektstärke bestimmen

Die Effektstärke Cohens d lässt sich in der ersten Zeile der Effektgrößen unter **Punktschätzung** ablesen.

Ein Cohens d von 0,350 ist ein schwacher bis mittelstarker Effekt.

BERECHNUNG MIT R

SO LIEST DU DIE ERGEBNISSE AB:

1. Schritt: Mittelwerte ablesen

```
Descriptive statistics by group
group: 1
  vars n  mean  sd median trimmed  mad min max range
X1    1 51 46.71 17.1   47    47 19.27  10 88   78
      skew kurtosis  se
X1 -0.02   -0.65 2.39
-----
group: 2
  vars n  mean  sd median trimmed  mad min max range
X1    1 54 40.94 15.84   43   40.48 19.27  14 74   60
      skew kurtosis  se
X1  0.1   -0.96 2.16
```

Group 1 = Frauen, Group 2 = Männer

Bei den deskriptiven Statistiken überprüfst du bei **mean** (= Mittelwert), ob die Ergebnisse in die erwartete Richtung gehen. Wenn nicht, endet hier deine Auswertung – dies gilt wohlgermerkt nur für gerichtete Hypothesen.

Hier sieht man, dass die Ergebnisse in die erwartete Richtung gehen: Der Mittelwert der Frauen (46,71) ist höher als derjenige der Männer (40,94). Wir können also fortfahren.

2. Schritt: Varianzhomogenität prüfen

```
Levene's Test for Homogeneity of Variance (center = mean)
      Df F value Pr(>F)
group 1  0.2092 0.6484
      103
```

Als nächstes überprüfst du mit dem **Levene-Test der Varianzgleichheit**, ob die Voraussetzung der Varianzhomogenität vorliegt.

Dieser testet die Nullhypothese, dass die Varianzen in beiden Gruppen gleich sind.

Ein signifikantes Ergebnis ($p < 0,05$) bedeutet, dass KEINE Varianzhomogenität vorliegt!



In diesem Falle müsstest du den **Welch-Test** verwenden, der sich vom t-Test dadurch unterscheidet, dass eine Korrektur der Freiheitsgrade vorgenommen wird.

Ist der Levene-Test NICHT signifikant, führst du den regulären „Two Sample t-test“ durch.

Der p -Wert des Levene-Tests findet sich bei R unterhalb von „Pr(>F)“.

Ein $p = 0,6484$ ist nicht signifikant. Daher liegt Varianzgleichheit vor und wir benötigen keinen Welch-Test.

3. Schritt: Ergebnis ablesen

Two Sample t-test

```
data: urlaub_buch_is[urlaub_buch_is$gender == 1, "stressnach"] and urlaub_buch_i
s[urlaub_buch_is$gender == 2, "stressnach"]
t = 1.7919, df = 103, p-value = 0.03805
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.4246965      Inf
sample estimates:
mean of x mean of y
 46.70588  40.94444
```

Der **p-Wert** befindet sich in der Zeile, die mit dem t -Wert beginnt. Ein $p = 0,038$ ist ein signifikanter Wert.

Wir können festhalten: Frauen weisen direkt nach dem Urlaub durchschnittlich ein höheres Stresslevel auf als Männer.

ÜBRIGENS:

Beim Testen mit Statistik-Programmen brauchst du keinen kritischen Wert, sondern liest nur den p -Wert ab!

Die Angaben zum empirischen t -Wert (1,7919) sowie zu den Freiheitsgraden ($df = 103$) findest du links vom p -Wert.

4. Schritt: Bei signifikantem Ergebnis die Effektstärke bestimmen

Cohen's d

```
d estimate: 0.3498795 (small)
95 percent confidence interval:
  lower      upper
-0.0403213  0.7400803
```

Die Effektstärke Cohens d lässt sich in der ersten Zeile ablesen. Praktischerweise steht auch schon der Interpretations-Hinweis gemäß Cohens (1988) Konventionen dabei.

Ein Cohens d von gerundet 0,35 ist ein schwacher bis mittelstarker Effekt.



ERGEBNIS KORREKT BERICHTEN

Die Angabe von Mittelwerten (= M) und Standardabweichungen (= SD) wird nicht immer gefordert. Halte dich daher bitte im Zweifelsfall an das, was an deiner Hochschule Usus ist!

So sieht der korrekte Bericht aus:

Frauen ($M = 46,71$, $SD = 17,10$) weisen direkt nach dem Urlaub durchschnittlich ein signifikant höheres Stresslevel auf als Männer ($M = 40,94$, $SD = 15,84$), $t(103) = 1,79$, $p = 0,038$, $d = 0,35$.

INTERPRETATION FÜR TANTE ERNA



Wir haben untersucht, ob Frauen direkt nach dem Urlaub gestresster sind als Männer.

Tatsächlich haben wir einen signifikanten Unterschied gefunden:
Die Damen sind auf einer Skala von 0 bis 100 durchschnittlich um 6 Punkte gestresster als die Herren – und das ist ein praktisch durchaus relevanter Effekt.