



STECKBRIEF

t-Test für abhängige Stichproben

WORUM GEHT'S?

Um Mittelwerts-Unterschiede bei zwei Messungen bei denselben Personen oder eine einmalige Messung bei zwei miteinander verbundenen Personen.

IN WELCHER WELT DER STATISTIK SIND WIR?

Inferenzstatistik: Je nach Fragestellung in der Welt der Unterschiede oder der Welt der Veränderungen über die Zeit hinweg

BEISPIELE

Die soziale Kompetenz ist nach einem Soziale-Kompetenz-Training höher als zuvor. (gerichtet: rechtsseitig)

Die Aufmerksamkeitsspanne ist bei Nachtteulen morgens niedriger als abends. (gerichtet: linksseitig)

Die Konfliktfähigkeit unterscheidet sich bei Zwillingen. (ungerichtet / zweiseitig)

VORAUSSETZUNGEN

- Korrelierte Messungen auf der Messwertpaar-Ebene – aber keine Abhängigkeiten zwischen den einzelnen Messwertpaaren.
Um das zu illustrieren: Berts und Knuts Abenteuerlust-Werte sind korreliert, dürfen aber nicht mit Zennis und Annamirls Werten korreliert sein.
Diese Voraussetzung wird durch die Planung der Studie gewährleistet.
- Die **AV ist metrisch** und die **Varianz bzw. Standardabweichung unbekannt**
- Die Differenzen zwischen den Messungen (= die Unterschiede zwischen Berts und Knuts Messung oder zwischen morgens und abends) müssen normalverteilt sein. Bei einer Stichprobengröße von mindestens 30 Personen kann davon ausgegangen werden.
- Generell reagiert dieser t-Test relativ robust auf Voraussetzungs-Verletzungen.

ALTERNATIVES NONPARAMETRISCHES VERFAHREN

Wilcoxon-Test für gepaarte Stichproben = Wilcoxon-Vorzeichen-Rang-Test

BERECHNUNG PER HAND

$$t = \frac{\bar{x}_D}{\hat{\sigma}_{\bar{x}_D}}$$

Im Nenner steht der Standardfehler der Differenz:

$$\hat{\sigma}_{\bar{x}_D} = \frac{\hat{\sigma}_D}{\sqrt{n}}$$

Formel für die Standardabweichung der Differenzen im Zähler:

$$\hat{\sigma}_D = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{x}_D)^2}{n - 1}}$$



Bestimmung der Freiheitsgrade:

Anzahl an Messwertpaaren minus 1 = $N - 1$

EFFEKTSTÄRKE COHENS d_z

Formel:

$$d_z = \frac{\bar{x}_d}{\hat{\sigma}_d} = \frac{t}{\sqrt{n}}$$

Konventionen:

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

ACHTUNG: jeweils **IM BETRAG!**

ANMERKUNG:

Es existieren zudem weitere Konventionen für d_z , die jedoch nicht so gängig sind:

Kleiner Effekt: $|0,14|$

Mittlerer Effekt: $|0,35|$

Starker Effekt: $|0,57|$



BERECHNUNG MIT SPSS

H_1 : Im Dschungelcamp dauert es bei Konsum von Kuhurin länger, bis der Würgereiz einsetzt, als bei Konsum von Affenurin. (gerichtet: rechtsseitig)

Der Würgereiz wird in Sekunden gemessen und jede Celebrity im Dschungelcamp bekommt einmal den Kuh- und einmal den Affenurin vorgesetzt. Es wird also zweimal bei denselben Personen gemessen.

Statistik bei gepaarten Stichproben

		Mittelwert	N	Std.- Abweichung	Standardfehler des Mittelwertes
Paaren 1	Kuhurin	7,26	34	2,064	,354
	Affenurin	4,79	34	2,071	,355

Korrelationen bei gepaarten Stichproben

		N	Korrelation	Signifikanz	
				Einseitiges p	Zweiseitiges p
Paaren 1	Kuhurin & Affenurin	34	-,518	<,001	,002

Test bei gepaarten Stichproben

		Mittelwert	Std.- Abweichung	Gepaarte Differenzen			T	df	Signifikanz	
				Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				Einseitiges p	Zweiseitiges p
					Unterer Wert	Oberer Wert				
Paaren 1	Kuhurin - Affenurin	2,471	3,603	,618	1,213	3,728	3,998	33	<,001	<,001

Effektgrößen bei Stichproben mit paarigen Werten

			Standardisierte r^a	Punktschätzung	95% Konfidenzintervall	
					Unterer Wert	Oberer Wert
Paaren 1	Kuhurin - Affenurin	Cohen's d	3,603	,686	,307	1,056
		Hedges' Korrektur	3,688	,670	,300	1,032

a. Der bei der Schätzung der Effektgrößen verwendete Nenner. Cohen's d verwendet die Standardabweichung einer Stichprobe der Mittelwertdifferenz. Hedges' Korrektur verwendet die Standardabweichung einer Stichprobe der Mittelwertdifferenz und einen Korrekturfaktor.

SPSS WIRFT STANDARDMÄSSIG VIER BEREICHE AUS:

- Zunächst erscheinen die deskriptiven Statistiken, wovon uns besonders die Mittelwerte interessieren.
- In der nächsten Ausgabe folgen die Korrelationen, die wir jedoch gewöhnlich nicht in die Analyse einbeziehen.
- Danach siehst du die Ergebnisse des t-Tests.
- Als letztes folgen die verfügbaren Effektgrößen.

SO LIEST DU DIE ERGEBNISSE AB

1. Schritt: Mittelwerte ansehen

Hier überprüfst du in der Spalte **Mittelwert**, ob die Ergebnisse in die erwartete Richtung gehen. Wenn nicht, endet hier deine Auswertung – dies gilt wohlgermerkt nur für gerichtete Hypothesen.

Im Beispiel siehst du, dass der Mittelwert bei Konsum von Kuhurin (7,26) deutlich größer als derjenige bei Konsum von Affenurin ist (4,79) und somit in die erwartete Richtung geht.

2. Schritt: Ergebnis ablesen

Zu Beginn bei „t-Test bei gepaarten Stichproben“ siehst du den **Mittelwert der Differenz**, der sich einfach aus der Differenz der beiden obigen Mittelwerte ergibt, und liest dann ganz rechts bei **Einseitiges p** den p-Wert ab (einseitiges p, weil wir eine einseitige, rechtsseitige Hypothese haben).

ÜBRIGENS:

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



Das $p < 0,001$ zeigt einen signifikanten Mittelwerts-Unterschied zwischen der Dauer, bis der Würgereiz bei Kuh- und bei Affenurin eintritt.

HINWEIS:

Bei älteren Versionen von SPSS wird nur die zweiseitige Testung angezeigt. Der p -Wert muss dann bei einseitigen Hypothesen und Ergebnissen, die in die erwartete Richtung gehen, halbiert werden!

3. Schritt: Bei signifikantem Ergebnis die Effektstärke bestimmen

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

Ein Cohens d_z von 0,686 ist ein starker Effekt. Dies zeigte sich bereits im großen Unterschied der beiden Mittelwerte.

BERECHNUNG MIT R

SO LIEST DU DIE ERGEBNISSE AB

1. Schritt: Mittelwert ansehen bei den deskriptiven Statistiken

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
kuhurin	1	34	7.26	2.06	7	7.11	1.48	4	13	9	0.70	0.35	0.35
affenurin	2	34	4.79	2.07	5	4.79	1.48	1	9	8	0.05	-0.83	0.36

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.

Im Beispiel siehst du, dass der Mittelwert bei Konsum von Kuhurin (7,26) deutlich höher ist als derjenige bei Konsum von Affenurin (4,79) und somit in die erwartete Richtung geht. Daher können wir fortfahren.

2. Schritt: Ergebnis des t-Tests ablesen

Paired t-test

```
data: dschungelcamp$kuhurin and dschungelcamp$affenurin
t = 3.9979, df = 33, p-value = 0.0001692
alternative hypothesis: true mean difference is greater than 0
95 percent confidence interval:
 1.424768      Inf
sample estimates:
mean difference
 2.470588
```

Der **p-Wert** befindet sich in der Zeile, die mit dem t -Wert beginnt. Ein p von 0,0001692 ist ein sehr kleiner und somit signifikanter Wert.

Die Angaben zum empirischen t -Wert (3,9979) sowie zu den Freiheitsgraden ($df = 33$) findest du links vom p -Wert.



Wir können festhalten: Beim Trinken von Kuhurin dauert es signifikant länger, bis der Würgereiz einsetzt, als beim Trinken von Affenurin.

Weiter unten bei **mean difference** siehst du den **Mittelwert der Differenz**, der sich einfach aus der Differenz der beiden obigen Mittelwerte aus den deskriptiven Statistiken ergibt.

ÜBRIGENS:

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

3. Schritt: Bei signifikantem Ergebnis die Effektstärke bestimmen

```
> cohensD(dschungelcamp$kuhurin, dschungelcamp$affenurin, method = "paired")  
[1] 0.6856411
```

Ein Cohens d_z von 0,686 ist ein starker Effekt.
Dies zeigte sich bereits im großen Unterschied der beiden Mittelwerte.

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:

Das durchschnittliche Dauer, bis der Würgereiz eintritt, ist nach Konsum von Kuhurin signifikant länger ($M = 7,26$, $SD = 2,06$) als bei Konsum von Affenurin ($M = 4,79$, $SD = 2,07$), $t(33) = 3,998$, $p < 0,001$, $d_z = 0,686$.

INTERPRETATION FÜR TANTE ERNA



In einem Dschungelcamp müssen Prominente oft ziemlich eklige Dinge zu sich nehmen. Daher haben wir in unserer Studie untersucht, ob es beim Trinken von Kuhurin länger dauert, bis der Würgereiz eintritt, als beim Trinken von Affenurin.

Es hat sich gezeigt, dass der Würgereiz bei Kuhurin tatsächlich signifikant später eintritt als bei Affenurin – und das ist ein praktisch relevanter, starker Unterschied.