

Signifikanztests zur Prüfung von Unterschieden in der zentralen Tendenz

-Teil 1-



t-Test bei einer Stichprobe - SPSS-Output

Der t-Test bei einer Stichprobe wird verwendet, um zu prüfen, ob der Mittelwert einer Stichprobe von einer bekannten Population abweicht. Beispiel hierfür wäre folgende Testsituation:

Fragestellung: Angenommen, das mittlere Examensalter von Studenten des Faches X beträgt im bundesweiten Durchschnitt 24. Nun soll untersucht werden, ob der Mittelwert der Studierenden einer bestimmten Uni Y davon abweicht.

Als erstes werden die deskriptiven Statistiken für die Variable „Alter“ ausgegeben.

	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Alter	93	28,61	4,409	,457

Dann folgt der eigentliche Test.

Test bei einer Stichprobe						
Testwert = 24						
	T	df	Sig. (2-seitig)	Mittlere Differenz	95% Konfidenzintervall der Differenz	
					Untere	Obere
Alter	10,091	92	,000	4,61	3,70	5,52

Anzahl der Freiheitsgrade

p-Wert (hier hochsignifikant, da $p < .001$)

empirische Prüfgröße t

Differenz zwischen Testwert (24) und Mittelwert der Stichprobe (28.61)

Interpretation des Tests:

Der Mittelwert der Studierenden dieser Hochschule unterscheidet sich signifikant vom bundesweiten Durchschnitt.

t-Test für unabhängige Stichproben - SPSS-Output

Fragestellung: Führen zwei verschiedene Lehrmethoden zu unterschiedlichem Klausurerfolg?

Zunächst werden deskriptive Statistiken für die beiden Gruppen ausgegeben:

Gruppenstatistiken					
	Art des Unterrichts	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Klausurpunkte	Computer gestützter Unterricht	36	112,78	21,948	3,658
	Buch gestützter Unterricht	30	105,17	32,787	5,986

Im Anschluss folgt der eigentliche t-Test:

Test bei unabhängigen Stichproben									
		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit					
		F	Signifikanz	Prüfgröße t					
				T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz
Klausurpunkte	Varianzen sind gleich	5,039	,028	1,124	64	,265	7,61	6,772	-5,918 21,141
	Varianzen sind nicht gleich			1,085	49,037	,283	7,61	7,015	-6,486 21,708

Mit dem **Levene-Test** wird automatisch die Voraussetzung der Varianzhomogenität überprüft. Geprüft wird die H_0 , die Varianzen seien gleich. Wird der Test signifikant ($p < 0.05$) spricht dies für die H_1 , dass die Varianzen nicht gleich sind. Ist dies der Fall, werden die Werte in der Zeile „Varianzen sind nicht gleich“ verwendet. Ansonsten nimmt man die erste Zeile und ignoriert die zweite.

Anzahl der Freiheitsgrade (für den Fall ungleicher Varianzen ist diese Anzahl korrigiert!)

p-Wert (nicht signifikant, da $p > 0.5$)

Interpretation des Tests:

Die beiden Lehrmethoden führen nicht zu unterschiedlich hohem Erfolg in der Klausur.

t-Test für abhängige Stichproben – SPSS-Output

Fragestellung: Unterscheidet sich die psychische Lage vor einer Intervention von der psychischen Lage nach der Intervention?

Zunächst werden für jede der beiden Variablen deskriptive Statistiken ausgegeben:

Statistik bei gepaarten Stichproben					
		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1	Psychische Lage vorher	3,22	107	1,525	,147
	Psychische Lage nachher	2,31	107	,840	,081

Daran anschließend wird die Korrelation der beiden Variablen ausgegeben.

Korrelieren die beiden Messwertreihen negativ miteinander, verliert der Test an Teststärke (d. h., die Wahrscheinlichkeit, dass tatsächlich vorhandene Unterschiede signifikant werden, ist geringer). In dem Fall sollte auf den Wilcoxon-Test zurückgegriffen werden.

Korrelationen bei gepaarten Stichproben				
		N	Korrelation	Signifikanz
Paaren 1	Psychische Lage vorher & Psychische Lage nachher	107	-,688	,000

Korrelation ist stark negativ und hoch-signifikant!

Dann folgt diese Tabelle:

Test bei gepaarten Stichproben									
		Gepaarte Differenzen							
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
					Untere	Obere	T	df	Sig. (2-seitig)
Paaren 1	Psychische Lage vorher - Psychische Lage nachher	,916	2,190	,212	,496	1,336	4,327	106	,000

p-Wert (hier hoch-signifikant, da $p < 0.001$)

Deskriptive Statistiken: Alle diese Angaben beziehen sich auf die **Differenzen der Messwert-Paare** (also Mittelwert dieser Differenzen etc.).

Prüfgröße t

Anzahl der Freiheitsgrade

☞ Trotz der hohen negativen Korrelation ist der Unterschied zwischen der Vorher- und der Nachher-Messung so hoch, dass die verminderte Teststärke keine Auswirkungen hat.

Interpretation des Tests:

Die psychische Lage nach der Intervention unterscheidet sich signifikant von der psychischen Lage vor der Intervention.

U-Test – SPSS-Output

Fragestellung: Unterscheiden sich Männer und Frauen in der Beurteilung ihrer Studienerfolge?

Wurden diese unter \Rightarrow Options angefordert, werden zunächst die deskriptiven Statistiken für alle verwendeten Variablen ausgegeben.

Deskriptive Statistiken								
	N	Mittelwert	Standardabweichung	Minimum	Maximum	Perzentile		
						25.	50. (Median)	75.
eigene Studienerfolge	45	3,89	,910	1	5	3,00	4,00	4,50
Geschlecht	46	,83	,383	0	1	1,00	1,00	1,00

Achtung, diese Tabelle auf keinen Fall ohne Überarbeitung weiterverwenden: Auch für die Gruppierungsvariable werden alle deskriptiven Statistiken berechnet, was bei nominalskalierten Variablen natürlich Unsinn ist!!!

Dann werden die Statistiken zu den gebildeten Rängen ausgegeben:

Ränge				
	Geschlecht	N	Mittlerer Rang	Rangsumme
eigene Studienerfolge	männlich	8	25,00	200,00
	weiblich	37	22,57	835,00
	Gesamt	45		

Anhand der Rangsummen wird der empirische U-Wert berechnet.

Die mittleren Ränge sind ähnlich wie Mittelwerte bei Intervalldaten lesbar: Höhere mittlere Ränge zeigen eine stärkere Ausprägung des Merkmals an.

☞ Wurde der U-Test mit ordinalskalierten Daten durchgeführt, können die mittleren Ränge oder die Mediane als deskriptive Maße verwendet werden (statt Mittelwerten). Wurde der U-Test mit eigentlich intervallskalierten Daten durchgeführt, die bestimmte Voraussetzungen für t-Test nicht erfüllten, können als deskriptive Maße natürlich die Mittelwerte verwendet werden.

Danach folgt der Test:

Statistik für Test ^a	
	eigene Studienerfolge
Mann-Whitney-U	132,000
Wilcoxon-W	835,000
Z	-,512
Asymptotische Signifikanz (2-seitig)	,609
Exakte Signifikanz [2*(1-seitig Sig.)]	,651 ^a

a. Nicht für Bindungen korrigiert.
b. Gruppenvariable: Geschlecht

empirischer U-Wert

Wilcoxon-W ist die Rangsumme, aufgrund derer der U-Wert berechnet wurde.

☞ Bei $n > 30$ wird die asymptotische Irrtumswahrscheinlichkeit für die Prüfgröße z berechnet (dabei wird davon ausgegangen, dass sich die Verteilung einer Normalverteilung annähert, der U-Wert wurde dafür in den z -Wert umgerechnet).

☞ Bei $n < 30$ wird zusätzlich die exakte Irrtumswahrscheinlichkeit berechnet!

p-Wert ist nicht signifikant, da $< .05$

Interpretation des Tests:

Männer und Frauen unterscheiden sich nicht signifikant in der Beurteilung ihres Studienerfolges.

Wilcoxon-Test - Output

Fragestellung: Führen zwei verschiedene Lehrmethoden zu unterschiedlichem Klausurerfolg?
(Hier wird noch einmal die gleiche Fragestellung wie beim t-Test für abhängige Stichproben untersucht.)

Zunächst werden die deskriptiven Statistiken für beide Variablen ausgegeben, wenn diese unter \Rightarrow Options angefordert wurden.

Deskriptive Statistiken								
	N	Mittelwert	Standardabweichung	Minimum	Maximum	Perzentile		
						25.	50. (Median)	75.
Psychische Lage vorher	108	3,23	1,520	1	6	2,00	3,00	5,00
Psychische Lage nachher	107	2,31	,840	1	4	2,00	2,00	3,00

Dann wird wieder eine Statistik über die gebildeten Ränge ausgegeben. Der Wilcoxon-Test verwendet die *Ränge der Messwertdifferenzen* der beiden Variablen (psyche1-psyche2). Dazu werden die Differenzen zunächst nach ihren *absoluten Beträgen* in eine *Rangreihe* gebracht. Diese Ränge werden dann nach dem Vorzeichen der ursprünglichen Differenzen in *negative*, *positive Ränge* (besser: Differenzen) und *Rangbindungen* (Nulldifferenzen) sortiert. Für die negativen und positiven Ränge werden dann *Rangsummen* und *mittlere Ränge* gebildet.

Ränge				
		N	Mittlerer Rang	Rangsumme
Psychische Lage nachher - Psychische Lage vorher	Negative Ränge	55 ^a	57,87	3183,00
	Positive Ränge	38 ^b	31,26	1188,00
	Bindungen	14 ^c		
	Gesamt	107		

mittlerer Rang und Rangsumme (T) sind in dem Fall nicht weiter verwendbar, sondern nur Zwischenergebnisse für den Wilcoxon-Test.

- a. Psychische Lage nachher < Psychische Lage vorher
b. Psychische Lage nachher > Psychische Lage vorher
c. Psychische Lage nachher = Psychische Lage vorher

offensichtlich liegen wesentlich mehr negative Ränge (Fälle, bei denen die psych. Lage hinterher schlechter ist!) als positive Ränge vor, was darauf hindeutet, dass sich die psychische Lage eher verschlechterte.

Danach folgen die Ergebnisse des Tests.

Statistik für Test ^b	
Psychische Lage nachher - Psychische Lage vorher	
Z	-3,147 ^a
Asymptotische Signifikanz (2-seitig)	,002

empirischer Wert z: Die Rangsummen verteilen sich angenähert normal, deshalb wird die kleinere Rangsumme z-standardisiert.

p-Wert (hier sehr signifikant!)

- a. Basiert auf positiven Rängen.
b. Wilcoxon-Test

Interpretation des Tests:

☞ Die psychische Lage war nach der Intervention signifikant verschieden von der psychischen Lage vorher.