

Test-Finder

Orientierung im Test-Chaos
Dipl.-Psych. Dr. Guido Strunk

Inhalt

1	Grundlagen	2
2	Maße der zentralen Tendenz vergleichen – Zwei Gruppen	3
2.1	T-Test für unabhängige Daten	3
2.2	T-Test für abhängige Daten	5
2.3	Wilcoxon-Rangsummen-Test	6
2.4	U-Test	7
2.5	Vorzeichen-Test	8
3	Mehr als zwei Gruppen	10
3.1	Einfaktoriell	10
3.1.1	Einfaktorielle Varianzanalyse	10
3.1.2	Einfaktorielle Varianzanalyse mit Messwiederholung	12
3.1.3	Kruskall-Wallis-H	12
3.1.4	Friedman-Test	14
3.2	Mehrfaktorielle Designs	15
4	Häufigkeiten	16
4.1	Häufigkeiten – 2 Gruppen	16
4.1.1	Fisher's exakter Test	16
4.1.2	McNemar-Test	17
4.2	Häufigkeiten – mehr als zwei Gruppen	19
4.2.1	Chi-Quadrat auf Gleichverteilung	19
4.2.2	Cochrans-Q	20
4.2.3	Chi-Quadrat, Fisher's exakter Test, Kendals-Tau-b	22

1 Grundlagen

- Skalenniveaus (Nominal, Ordinal, Intervall)
- uV, aV, einfaktoriell, mehrfaktoriell
- Zentraler Grenzwertsatz
- Normalverteilungsannahme
- Approximation von Verteilungen
- Exakte Tests
- Abhängige vs. Unabhängige Daten
- 1-seitige, 2-seitige Testung

	Unterschieds-Hypothesen		Zusammenhangs-Hypothesen	
2 Gruppen / Variablen / Objekte	Einfache Gruppenvergleiche		Einfache Korrelation	
Mehr als 2 Gruppen / Variablen / Objekte	Multiple Gruppenvergleiche		Multiple Korrelation / Regression	
Hypothesentyp	gerichtet (1-seitig)	ungerichtet (2-seitig)	gerichtet (1-seitig)	ungerichtet (2-seitig)

complexity-research.com

..\produkte\test.finder

Unterschiedstests

Prüfen auf Unterschiede

Maße der zentralen Tendenz vergleichen

Häufigkeiten, Anteile, Prozentsätze vergleichen

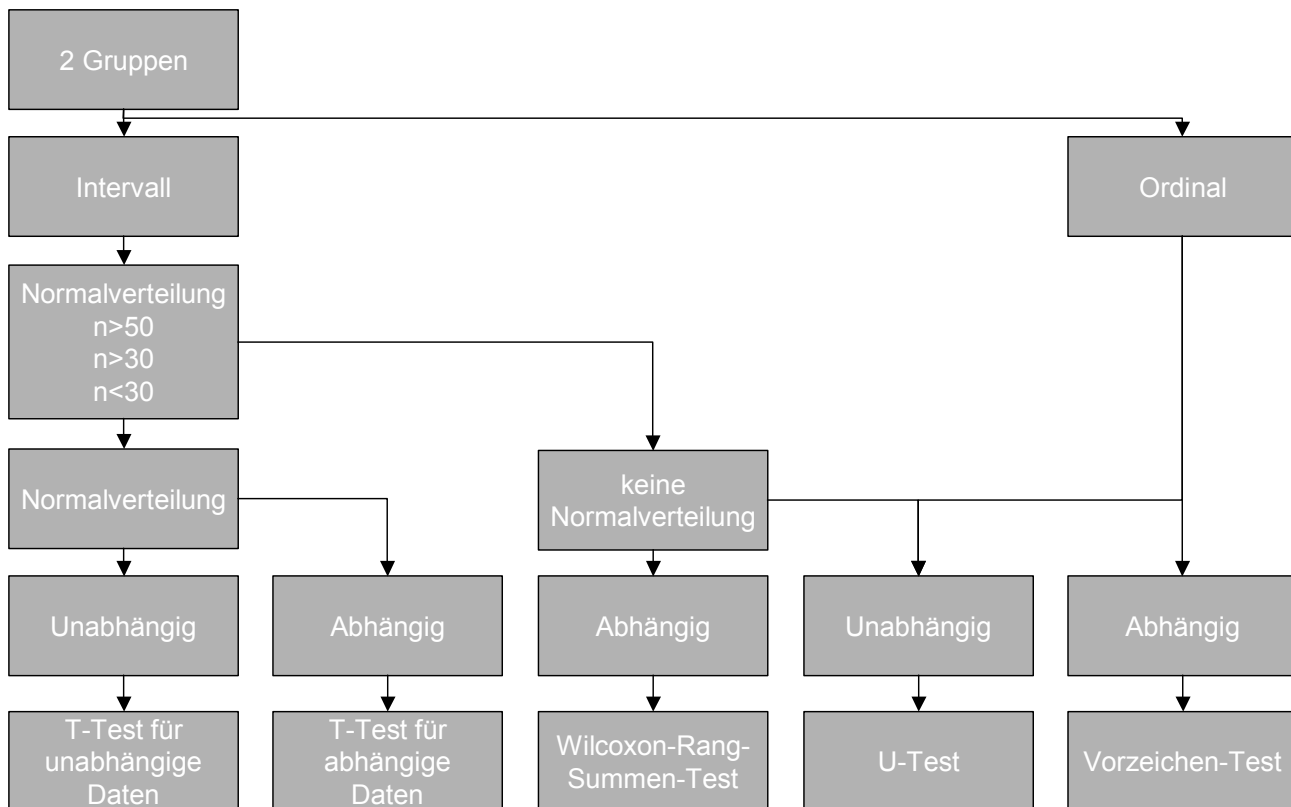
Varianzen vergleichen

Hintergrund

Mittelwert und Median sind so genannte Maße der zentralen Tendenz. Diese werden mit anderen statistischen Verfahren auf Unterschiede geprüft als Varianzen (auch Dispersionsmaße genannt).

Werden in einer Untersuchung Ereignisse oder Merkmale gezählt, so ergeben sich Häufigkeitsverteilungen, Anteile und Prozentsätze. Diese müssen ebenfalls mit speziellen Methoden untersucht werden.

2 Maße der zentralen Tendenz vergleichen – Zwei Gruppen



2.1 T-Test für unabhängige Daten

T-Test – unabhängige Daten

SPSS:

Statistik \ Mittelwerte vergleichen
T-Test bei unabhängigen ...

complexity-research.com

..\produkte\test.finder

T-Test

- **Normalverteilung:** Die Normalverteilung der Messwerte ist keine Voraussetzung für den T-Test. Vielmehr müssen die Differenzen der Mittelwerte normalverteilt sein. Das ist bei großen Stichproben automatisch der Fall. Je nach Misstrauen gegenüber dem zentralen Grenzwertsatz werden entweder 30 oder 50 Messungen pro Gruppe als Minimum empfohlen. Bei kleinen Stichproben müssen hingegen die Messwerte normalverteilt sein.
- **Varianzhomogenität:** Es gibt zwei verschiedene T-Tests, einen für und einen ohne Varianzhomogenität.

complexity-research.com
 ..\produkte\test.finder T-Test – unabhängige Stichproben

T-Test

Gruppenstatistiken					
	GSCHL	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
ITEM16	männlich	123	1,67	,70	6,29E-02
Gesamtzufriedenheit	weiblich	248	1,75	,74	4,72E-02

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit				
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz
ITEM16	Varianzen sind gleich	,206	,650	-1,037	369	,300	-8,33E-02	8,03E-02
Gesamtzufriedenheit	Varianzen sind nicht gleich			-1,060	257,815	,290	-8,33E-02	7,86E-02

Levene-Test: Ist die Signifikanz $\geq 0,20$?

nein: Varianzen sind nicht gleich: untere Zeile
 ja: Varianzen sind gleich: obere Zeile

T-Test

Gruppenstatistiken					
	GSCHL	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
ITEM16	männlich	123	1,67	,70	6,29E-02
Gesamtzufriedenheit	weiblich	248	1,75	,74	4,72E-02

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit				
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz
ITEM16	Varianzen sind gleich	,206	,650	-1,037	369	,300	-8,33E-02	8,03E-02
Gesamtzufriedenheit	Varianzen sind nicht gleich			-1,060	257,815	,290	-8,33E-02	7,86E-02

t-Test (korrekte Zeile): Ist die Signifikanz $\leq 0,05$?

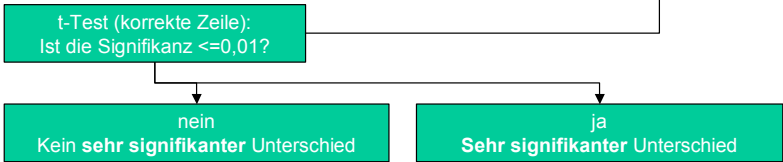
nein: Kein signifikanter Unterschied
 ja: Signifikanter Unterschied

Anmerkungen
 Gruppenstatistiken
 Test bei unabhängigen Stichp...

T-Test

Gruppenstatistiken					
	GSCHL	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
ITEM16	männlich	123	1,67	,70	6,29E-02
Gesamtzufriedenheit	weiblich	248	1,75	,74	4,72E-02

Test bei unabhängigen Stichproben								
		1-seitiger Test der Varianzgleichheit			T-Test für die Mittelwertgleichheit			
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz
ITEM16	Varianzen sind gleich	,206	,650	-1,037	369	,300	-0,33E-02	0,03E-02
Gesamtzufriedenheit	Varianzen sind nicht gleich			-1,060	297,915	,290	-8,33E-02	7,86E-02



- *Ungerichtete Hypothese:* Die von SPSS angegebene Wahrscheinlichkeit ist 2-seitig zu interpretieren.
- *Gerichtete Hypothese:* Für den 1-seitigen Wert gilt, dass er die Hälfte des 2-seitigen Wertes beträgt.

2.2 T-Test für abhängige Daten

T-Test – abhängige Daten

SPSS:

Statistik \ Mittelwerte vergleichen

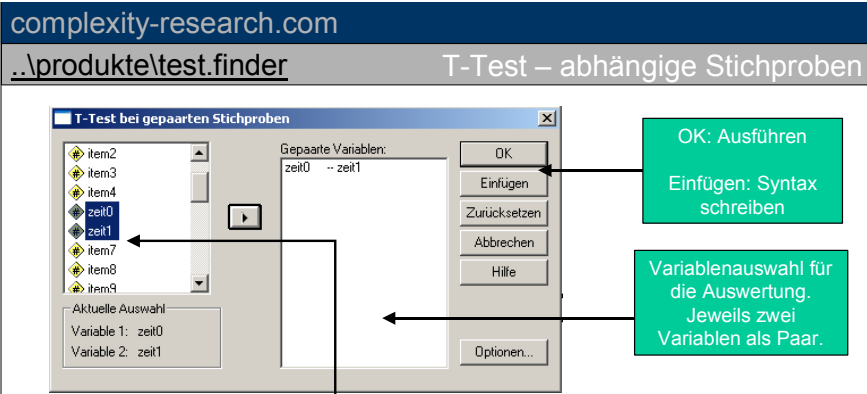
T-Test bei gepaarten ...

complexity-research.com

..\produkte\test.finder

T-Test

- **Normalverteilung:** Die Normalverteilung der Messwerte ist keine Voraussetzung für den T-Test. Vielmehr müssen die Differenzen der Mittelwerte normalverteilt sein. Das ist bei großen Stichproben automatisch der Fall. Je nach Misstrauen gegenüber dem zentralen Grenzwertsatz werden entweder 30 oder 50 Messungen pro Messzeitpunkt als Minimum empfohlen. Bei kleinen Stichproben müssen hingegen die paarweise berechneten Messwert-Differenzen normalverteilt sein.



Bei abhängigen Stichproben liegen pro Fall Messungen für verschiedene Zeitpunkte vor.

→ T-Test

Statistik bei gepaarten Stichproben

Paaren	Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
ZEIT0	1,90	358	,95	5,04E-02
ZEIT1	1,36	358	,61	3,24E-02

Korrelationen bei gepaarten Stichproben

Paaren	N	Korrelation	Signifikanz
ZEIT0 & ZEIT1	358	,456	,000

Ist die Korrelation negativ, sollte der Test nicht gerechnet werden. Als Alternative muss dann der Wilcoxon-Test berechnet werden.

Test bei gepaarten Stichproben

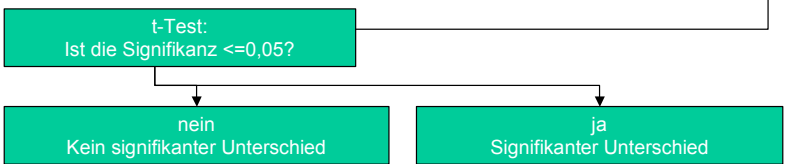
Paaren	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz		T	df	Sig. (2-seitig)
				Untere	Obere			
ZEIT0 - ZEIT1	,54	,87	4,59E-02	,45	,63	11,752	357	,000

Korrelationen bei gepaarten Stichproben

Paaren	N	Korrelation	Signifikanz
ZEIT0 & ZEIT1	358	,456	,000

Test bei gepaarten Stichproben

Paaren	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz		T	df	Sig. (2-seitig)
				Untere	Obere			
ZEIT0 - ZEIT1	,54	,87	4,59E-02	,45	,63	11,752	357	,000



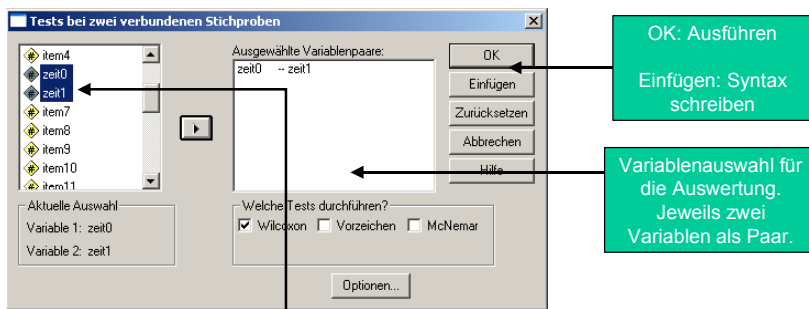
2.3 Wilcoxon-Rangsummen-Test

Wilcoxon-Rangsummen-Test

SPSS:

Statistik \ Nichtparametrische Tests
2 verbundene Stichproben ...

- Der Test gilt als Alternative zum T-Test für abhängige Stichproben, falls die Normalverteilungsannahme nicht erfüllt ist.
- Der Test zieht die Messwertpaare voneinander ab, wobei positive, negative und Nulldifferenzen unterschieden werden.
- Der Test berücksichtigt dabei auch die Höhe der Differenzen.



Je zwei Variablen anklicken, um ein Paar zu bilden.

Bei abhängigen Stichproben liegen pro Fall Messungen für verschiedene Zeitpunkte vor.

Wilcoxon-Test

Ränge		N	Mittlerer Rang	Rangsumme
ZEIT1 - ZEIT0	Negative Ränge	157 ^a	87,46	13731,50
	Positive Ränge	15 ^b	76,43	1146,50
	Dünnungen	186 ^c		
	Gesamt	358		

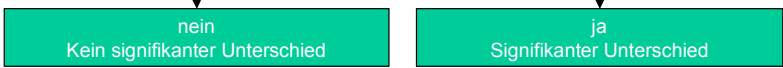
- a. ZEIT1 < ZEIT0
- b. ZEIT1 > ZEIT0
- c. ZEIT0 = ZEIT1

Statistik für Test^b

	ZEIT1 - ZEIT0
Z	-10,192 ^a
Asymptotische Signifikanz (2-seitig)	,000

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

z-Test für Wilcoxon-Test: Ist die Signifikanz <=0,05?



2.4 U-Test

U-Test

SPSS:

Statistik \ Nichtparametrische Tests
 2 unabhängige Stichproben ...

- Der Test gilt als Alternative zum T-Test für unabhängige Stichproben, falls die Normalverteilungsannahme bzw. die Intervallskalierung der Daten nicht erfüllt sind.
- Der Test sortiert die Messdaten und vergibt Ränge.
- Er zählt, wie häufig Rangplatzunterschiede vorkommen.

Mann-Whitney-Test

Ränge

	ALTGR2	N	Mittlerer Rang	Rangsumme
ITEM16	1,00	44	88,56	3016,50
Gesamtzufriedenheit	3,00	116	05,03	9063,50
Gesamt		160		

Beide Tests sind äquivalent, kommen aber zu verschiedenen Prüfgrößen, die aber dem gleichen z-Wert entsprechen.

Statistik für Test ^a		ITEM16 Gesamtzufriedenheit
Mann-Whitney-U		2026,500
Wilcoxon-W		3016,500
Z		-2,209
Asymptotische Signifikanz (2-seitig)		,027

a. Gruppenvariable: ALTGR2

z-Test für U-Test:
Ist die Signifikanz $\leq 0,05$?

nein
Kein signifikanter Unterschied

ja
Signifikanter Unterschied

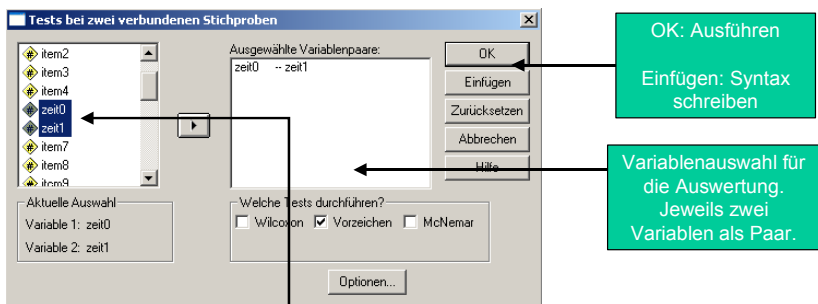
2.5 Vorzeichen-Test

Vorzeichen-Test

SPSS:

Statistik \ Nichtparametrische Tests
 2 verbundene Stichproben ...

- Der Test zieht die Messwertpaare voneinander ab, wobei positive, negative und Nulldifferenzen unterschieden werden.
- Der Test berücksichtigt nicht die Höhe der Differenzen, sondern zählt diese. Deswegen ist bei intervallskalierten Daten, für die der T-Test nicht gerechnet werden kann, der Wilcoxon-Rang-Summen-Test vorzuziehen.



Je zwei Variablen anklicken, um ein Paar zu bilden.

Bei abhängigen Stichproben liegen pro Fall Messungen für verschiedene Zeitpunkte vor.

Vorzeichentest

Häufigkeiten		
		N
ZEIT1 - ZEIT0	Negative Differenzen ^a	157
	Positive Differenzen ^b	15
	Bindungen ^c	186
	Gesamt	358

- a. ZEIT1 < ZEIT0
- b. ZEIT1 > ZEIT0
- c. ZEIT0 = ZEIT1

Statistik für Test ^a	
	ZEIT1 - ZEIT0
Z	-10,751
Asymptotische Signifikanz (2-seitig)	,000

a. Vorzeichentest

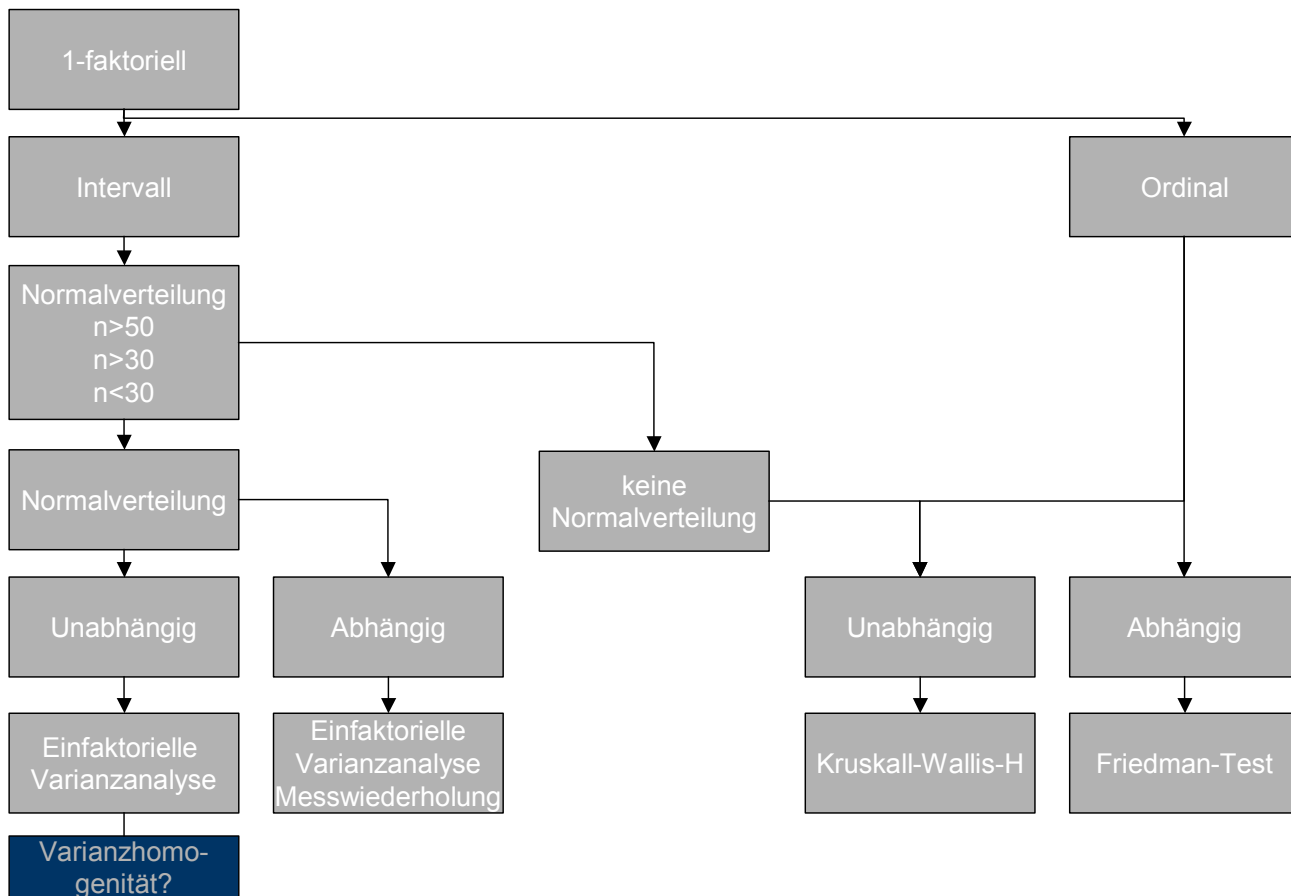
z-Test für Vorzeichen-Test:
Ist die Signifikanz <=0,05?

nein
Kein signifikanter Unterschied

ja
Signifikanter Unterschied

3 Mehr als zwei Gruppen

3.1 Einfaktoriell



3.1.1 Einfaktorielle Varianzanalyse

Einfaktorielle Varianzanalyse (ANOVA)

SPSS:

Statistik \ Mittelwerte vergleichen
einfaktorielle ANOVA ...

complexity-research.com

..\produkte\test.finder

Einfaktorielle ANOVA

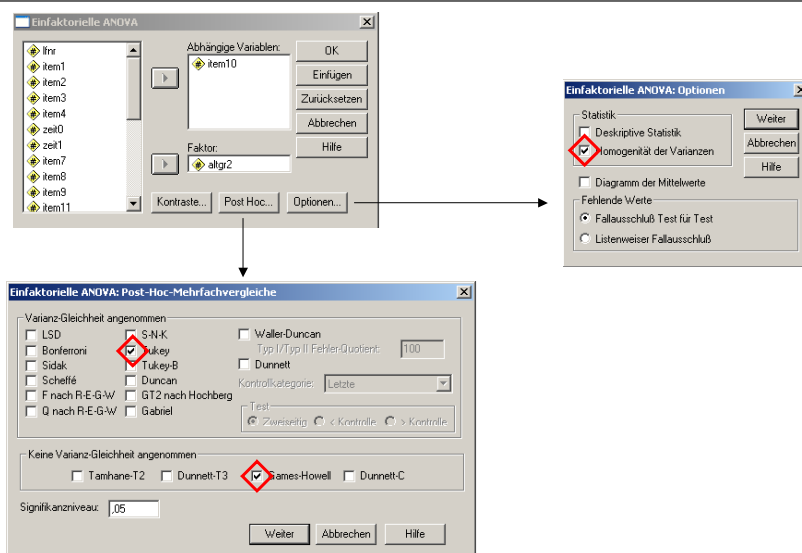
- **Varianzhomogenität:** Bei der Durchführung muss unbedingt die Varianzhomogenität geprüft werden. "Im Fall normalverteilter Populationen mit gleichen Stichprobenumfängen ist die Varianzanalyse relativ robust gegenüber mäßigen Unterschieden zwischen den Populationsvarianzen. Sie reagiert jedoch sehr empfindlich auf Varianzheterogenität, wenn die Stichproben nicht gleich groß sind." (Diehl & Arbinger, 1990, S. 214)
- **Bei fehlender Varianzhomogenität** sollte der Brown-Forsythe-JM- oder der Welch-JM-Test gewählt werden (beide in SPSS nicht enthalten). Im SPSS kann der H-Test von Kruskal-Wallis gerechnet werden, der aber in diesem Fall weniger robust ist, als die beiden zuerst genannten Verfahren.

- **Feste vs. zufällige Effekte:** Die klassische einfaktorielle Varianzanalyse geht davon aus, dass feste Effekte untersucht werden. Es wird also angenommen, dass man genau die Gruppen untersucht, die man auch untersuchen möchte und über die man Aussagen treffen möchte. Zufällige Effekte würden hingegen vorliegen, wenn aus einer Menge möglicher Gruppen, per Zufall eine begrenzte Anzahl gezogen wurde. Das Ziel ist dann in der Regel über diese Gruppen hinweg auf die anderen möglichen aber nicht untersuchten Gruppen zu generalisieren. In diesem Falle wäre eine klassische, einfaktorielle Varianzanalyse nicht angemessen.
- Die Varianzanalyse sagt nichts darüber aus, zwischen welchen Gruppen Unterschiede bestehen, sondern nur dass/ob überhaupt Unterschiede vorliegen:
 - **Tukey-Test:** Prüft auf Einzelunterschiede bei varianzhomogenen Stichproben.
 - **Games-Howell-Test:** Prüft auf Einzelunterschiede bei varianzheterogenen Stichproben.
- Für beide Testverfahren bedeuten Verletzungen der Normalität kein sehr großes Problem. Nur bei sehr schiefen Verteilungen und kleinen Gruppengrößen ist mit erheblichen Problemen zu rechnen.

complexity-research.com

..\produkte\test.finder

Einfaktorielle ANOVA



Einfaktoriell

Test der Homogenität der Varianzen

Levene-Statistik	df1	df2	Signifikanz
,219	2	338	,803

Levene-Test:
Ist die Signifikanz $\geq 0,20$?

ja
ANOVA und Tukey
interpretieren

Nein:
ANOVA abbrechen bei
kleinem n ($n_i < 10$) und
ungleichen n.

Games-Howell
interpretieren

ANOVA

ITEM10	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	,160	2	8,017E-02	,062	,940
Innerhalb der Gruppen	435,734	338	1,289		
Gesamt	435,894	340			

Post-Hoc-Tests

Die Quadratsumme der Abweichungen innerhalb und zwischen den Gruppen wird verglichen.
Ist die Signifikanz $\leq 0,05$ bzw. $0,01$ liegt ein signifikanter Unterschied vor.

Post-Hoc-Tests

Wenn der Levene-Test:
Signifikanz $\geq 0,20$?

Wenn der Levene-Test:
Signifikanz $\leq 0,20$?

Mehrfachvergleiche

Abhängige Variable: ITEM10

	(I) ALTGR2	(J) ALTGR2	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall	
						Untergrenze	Obergrenze
Tukey-HSD	1,00	2,00	5,77E-02	,195	,953	-,40	,52
	3,00	2,00	7,32E-02	,209	,935	-,42	,56
	1,00	3,00	-5,77E-02	,195	,953	-,52	,40
	3,00	3,00	1,55E-02	,137	,993	-,31	,34
	1,00	2,00	-7,32E-02	,209	,935	-,56	,42
	2,00	2,00	-1,55E-02	,137	,993	-,34	,31
Games-Howell	1,00	2,00	5,77E-02	,195	,947	-,38	,50
	3,00	2,00	7,32E-02	,209	,930	-,41	,55
	1,00	3,00	-5,77E-02	,195	,947	-,50	,38
	3,00	3,00	1,55E-02	,137	,993	-,31	,35
	1,00	2,00	-7,32E-02	,209	,930	-,55	,41
	2,00	2,00	-1,55E-02	,137	,993	-,35	,31

Paarweise wird jede Gruppe mit jeder anderen verglichen.
Ist die Signifikanz $\leq 0,05$ bzw. $0,01$ liegt ein signifikanter Unterschied vor.

3.1.2 Einfaktorielle Varianzanalyse mit Messwiederholung

Varianzanalyse mit Messwiederholung

SPSS:

Statistik \ Allgemeines lineares Modell
GLM-Messwiederholung ...

3.1.3 Kruskal-Wallis-H

Kruskal-Wallis-H-Test

SPSS:

Statistik \ Nichtparametrische Tests
k unabhängige Stichproben ...

- **Varianzhomogenität und Symmetrie der Verteilung.**
 Der Kruskal-Wallis H-Test gilt als Alternative für die einfaktorielle Varianzanalyse bei Verletzung der Normalverteilung und der Varianzhomogenität. Es gibt jedoch Hinweise darauf, dass auch der Kruskal-Wallis H-Test bei heterogenen Varianzen und schiefen Verteilungen ähnliche Probleme aufweist wie die Varianzanalyse (Diehl & Arbinger, 1990, S. 216).
- **Chi-Quadrat-Verteilung.** Der Test kann exakt berechnet werden und setzt dann keine bestimmte Verteilung voraus. In der Regel wird er jedoch über die Chi-Quadrat-Verteilung approximiert. Hier gilt, dass für 3 Gruppen $n_i > 8$ und für 5 Gruppen $n_i > 3$ die Untergrenze bilden für grobe „Routineentscheidungen“ (vgl. Bortz & Lienert, 2000, S. 225).

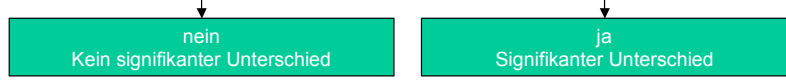
Kruskal-Wallis-Test

Ränge			
	ALTOR2	N	Mittlerer Rang
ITEM16	1,00	44	166,81
Gesamtzufriedenheit	2,00	207	176,80
	3,00	116	203,74
Gesamt		367	

Statistik für Test ^{a,b}	
	ITEM16 Gesamtzufriedenheit
Chi-Quadrat	7,643
df	2
Asymptotische Signifikanz	,022

a. Kruskal-Wallis-Test
 b. Gruppenvariable: ALTOR2

Chi-Quadrat-Test für Kruskal-Wallis-Test:
 Ist die Signifikanz $\leq 0,05$?



3.1.4 Friedman-Test

Friedman-Test

SPSS:

Statistik \ Nichtparametrische Tests
k verbundene Stichproben ...

complexity-research.com

..\produkte\test.finder

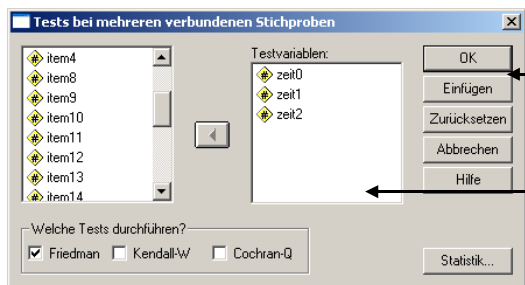
Friedman-Test

- Der Friedman-Test prüft rangskalierte abhängige Daten auf Unterschiede. Ein solcher Fall liegt z.B. vor, wenn n Beurteiler J verschiedene Elemente in eine Rangordnung bringen. Zum Beispiel könnten n=10 Statistiker gebeten werden J=5 gegebene statistische Testverfahren zu bewerten. Der beste Test bekommt den Rangplatz 1 der zweitbeste den Rangplatz 2 und so weiter. Die Bewertungen der Tests sind nun voneinander abhängig und zudem rangskaliert.
- Ein anderes typisches Beispiel für den Einsatz des Friedman-Tests stellt ein Design dar, in dem an n Personen J-mal wiederholt eine Messung einer Variablen vorgenommen wird. Es handelt sich hier um eine klassische Messwiederholung. Werden für jede Person die ermittelten Messwerte in Ränge umgewandelt (der geringste Messwert der Person erhält eine 1, der Nächsthöhere eine 2 und so weiter), so ist der Friedman-Test der passende Test.

complexity-research.com

..\produkte\test.finder

Friedman-Test



- OK: Ausführen
- Einfügen: Syntax schreiben
- Variable an der die beiden Gruppen unterschieden werden

Friedman-Test

Range

	Mittlerer Rang
ZEIT0	2,40
ZEIT1	1,80
ZEIT2	1,80

Statistik für Test^a

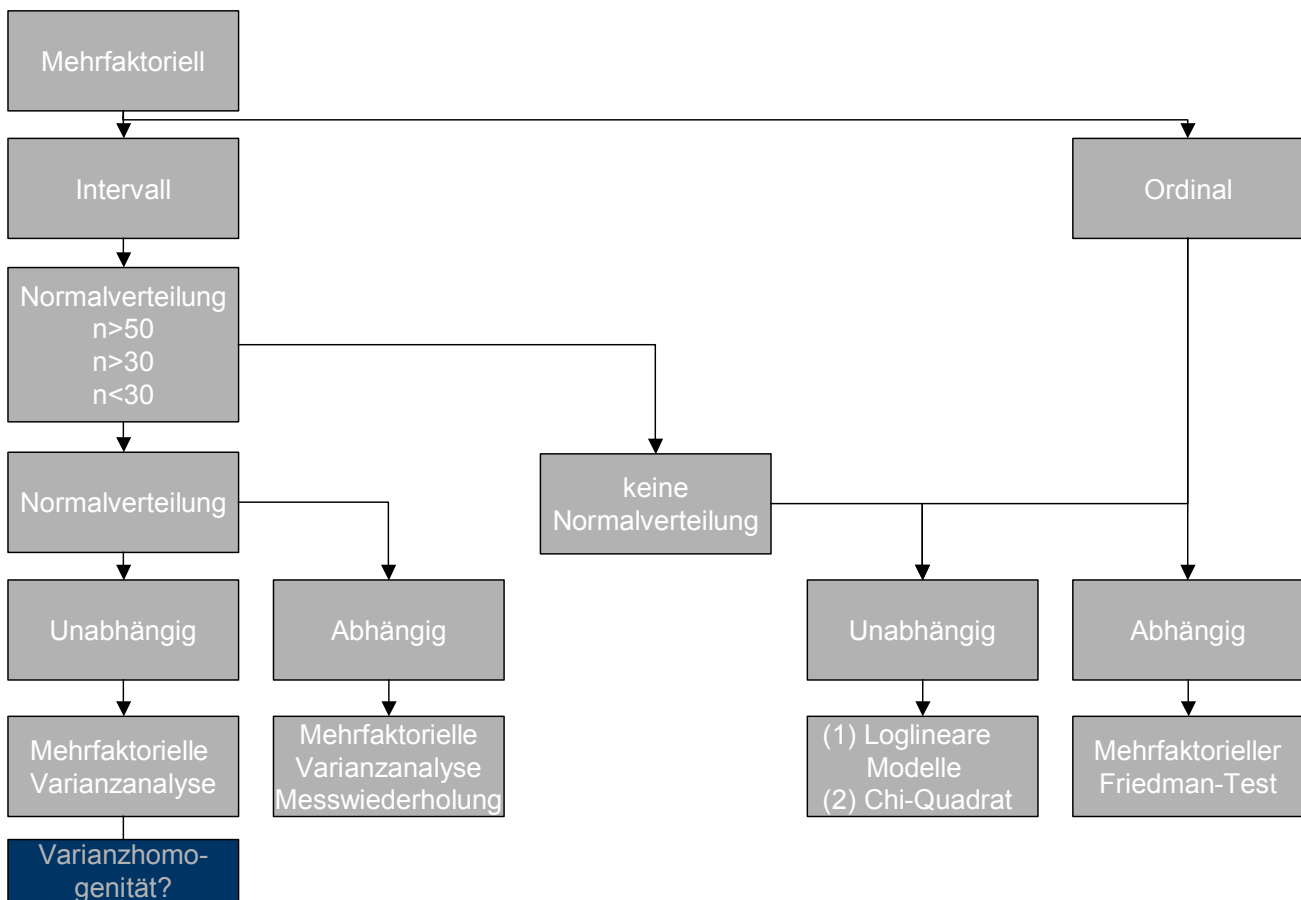
N	355
Chi-Quadrat	193,148
df	2
Asymptotische Signifikanz	,000

a. Friedman-Test

Chi-Quadrat-Test für Friedman-Test:
Ist die Signifikanz $\leq 0,05$?

- nein
Kein signifikanter Unterschied
- ja
Signifikanter Unterschied

3.2 Mehrfaktorielle Designs



complexity-research.com

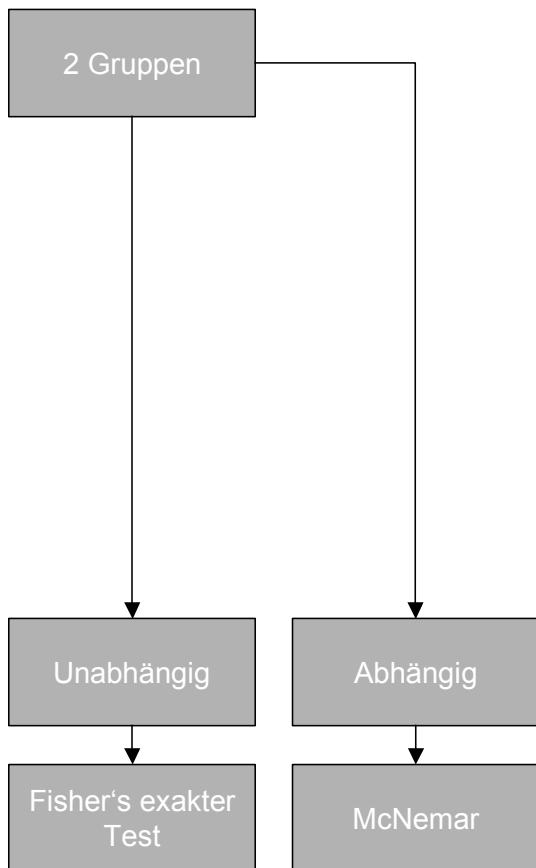
..\produkte\test.finder

Mehrfaktorielle Designs

- **Chi-Quadrat:** Man kann versuchen die Daten in Häufigkeiten umzuwandeln (z.B: Median-Split).
- **Mehrfaktorieller Friedman-Test:** Es gibt Verallgemeinerungen des Friedman-Testes für mehr als einen Faktor. Diese sind in SPSS nicht implementiert.
- **GLM:** Mehrfaktorielle Designs finden sich in SPSS unter dem Namen Allgemeines Lineares Modell.
- **Regressions-Modelle:** Mehrfaktorielle Varianzanalysen sind äquivalent zu Regressions-Modellen (Zusammenhangshypothesen).
 - **Lineare Regression:** Metrische abhängige Variable.
 - **Logistische Regression:** Dichotome abhängige Variable.
 - **Multimodale Logistische Regression:** Mehrfach gestufte abhängige Variable.
 - **Cox Regression:** Vorhersage von Wartezeiten bis ein Ereignis eintritt.

4 Häufigkeiten

4.1 Häufigkeiten – 2 Gruppen



4.1.1 Fisher's exakter Test

Fisher's exakter Test

SPSS:
Statistik \ Zusammenfassen
Kreuztabellen

complexity-research.com

..\produkte\test.finder

Fisher's exakter Test

- **Idee:** Wenn Häufigkeiten in zwei Gruppen ermittelt werden, so ergibt sich eine 4-Felder-Kreuztabelle. Das Merkmal für das die Häufigkeit bestimmt wird liegt entweder vor oder nicht, was zwei Bedingungen sind. Weitere zwei ergeben sich für die beiden Gruppen, die untersucht und verglichen werden.
- **Exakter Test:** Der Test bildet eine exakte Wahrscheinlichkeitsverteilung, die in jedem Fall korrekt ist und keinerlei Einschränkungen unterliegt, wie sie z.B. für den Chi-Quadrat-Test gelten. Aber, bei großen Stichprobenumfängen treten bei der Berechnung auch sehr große Zahlen auf. Einige Computer-Programme geben bei einem N größer 1000 auf. Hier sollte dann doch der Chi-Quadrat-Test herangezogen werden, oder eine Aproximation für den Fisher-Test bestimmt werden.
- **SPSS:** Verfügt SPSS über das Modul „Exakte Tests“ kann der Fischer-Test auch für große Stichproben und für beliebig große Kontingenztafeln bestimmt werden.

GSCHL * ALTGR Kreuztabelle

Anzahl		ALTGR		Gesamt
		1,00	2,00	
GSCHL	0 männlich	25	45	70
	1 weiblich	72	78	150
Gesamt		97	123	220

Exakter Wert für Fisher-Test:
Ist die Signifikanz <=0,05?

	Wert	df	Asymptotische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)
Chi-Quadrat nach Pearson	2,922 ^b	1	,087		
Kontinuitätskorrektur ^a	2,445	1	,118		
Likelihood-Quotient	2,955	1	,086		
Exakter Test nach Fisher				,109	,058
Zusammenhang linear-mit-linear	2,909	1	,088		
Anzahl der gültigen Fälle	220				

a. Wird nur für eine 2x2-Tabelle berechnet

b. 0 Zellen (.0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 30,86.

nein
Kein signifikanter Unterschied

ja
Signifikanter Unterschied

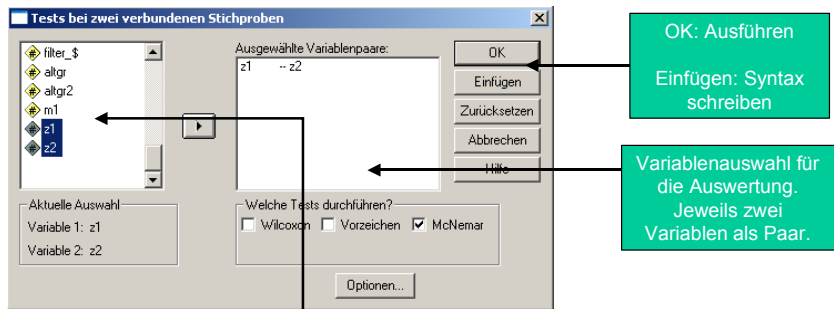
4.1.2 McNemar-Test

McNemar

SPSS:

Statistik \ Nichtparametrische Tests
Zwei verbundene Stichproben

- Der Test kann für kleine Stichproben exakt berechnet werden. Für große Stichproben erfolgt eine Approximation über die Normalverteilung.
- **SPSS:** Die beiden gepaarten Variablen müssen dichotom kodiert sein und beide müssen gleich kodiert sein.



OK: Ausführen
 Einfügen: Syntax schreiben

Variablenauswahl für die Auswertung. Jeweils zwei Variablen als Paar.

Je zwei Variablen anklicken, um ein Paar zu bilden.

Bei abhängigen Stichproben liegen pro Fall Messungen für verschiedene Zeitpunkte vor. Der Test erwartet gleich kodierte, dichotome Variablenpaare.

McNemar-Test

Kreuztabellen

1 & 2

	1 & 2	
	z1	z2
z1		
0	25	45
1	72	78

Statistik für Test^a

	1 & 2
N	220
Chi-Quadrat ^a	5,778
Asymptotische Signifikanz	,016

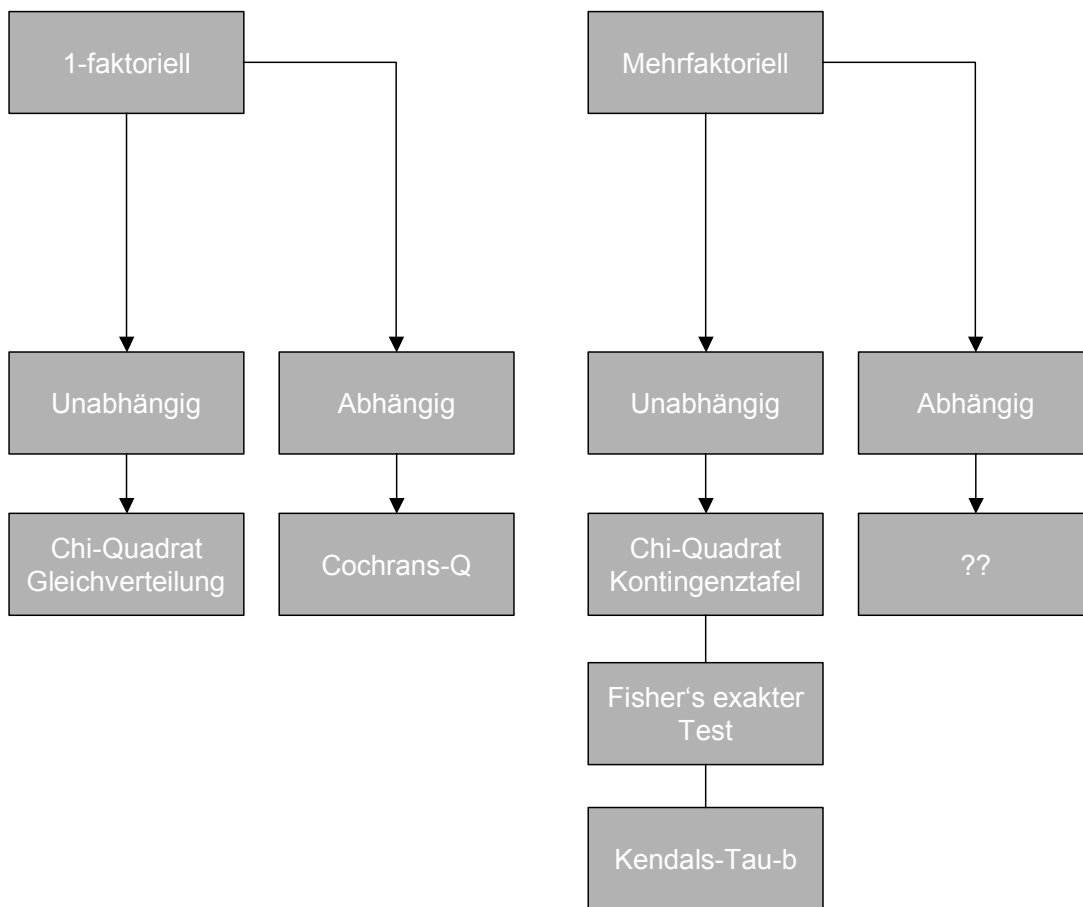
Chi-Quadrat-Test für McNemar-Test: Ist die Signifikanz <=0,05?

nein
Kein signifikanter Unterschied

ja
Signifikanter Unterschied

a. Kontinuität korrigiert
 b. McNemar-Test

4.2 Häufigkeiten – mehr als zwei Gruppen



4.2.1 Chi-Quadrat auf Gleichverteilung

Chi-Quadrat – Gleichverteilung

SPSS:

Statistik \ Nichtparametrische Tests\
 Chi-Quadrat - Erwartete Werte: Alle
 Kategorien gleich

complexity-research.com

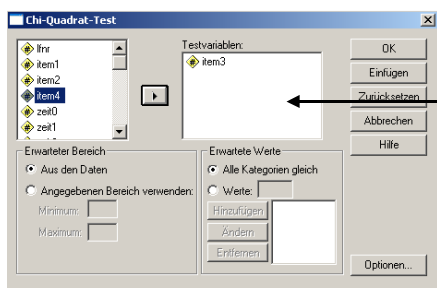
..\produkte\test.finder

Chi-Quadrat – Gleichverteilung

- **Post-Hoc:** Der Chi-Quadrat-Test zeigt nicht, zwischen welchen Gruppen Unterschiede bestehen, sondern nur dass/ob überhaupt Unterschiede vorliegen. Einzelunterschiede können mit mehrfachen Paarvergleichen getestet werden. Hier ist der Fisher-Test ein exaktes Verfahren. Der Alpha-Fehler muss für die Zahl der Paarvergleiche adjustiert werden. Zudem sind Gleichungen für die Bestimmung von Kontrasten verfügbar (vgl. Diehl & Arbinger, 1990, S. 455).

- **Chi-Quadrat-Verteilung.** Im Test werden erwartete und beobachtete Häufigkeiten verglichen. Dabei ergibt sich approximativ eine Chi-Quadrat-Verteilung, wenn die erwarteten Häufigkeiten nicht zu klein sind. Als Faustregel gilt heute, dass keine erwartete Häufigkeit kleiner 1 sein darf und höchstens 20% eine erwartete Häufigkeit kleiner 5 aufweisen dürfen. Alternativ können eventuell exakte Tests durchgeführt werden. Auch paarweise Vergleiche mit dem Fisher-Test und eine Alpha-Fehler-Adjustierung sind denkbar. Oder man kann inhaltlich passende Gruppen so zusammen legen, dass die erwarteten Häufigkeiten die Grenze überschreiten.

complexity-research.com
 ..\produkte\test.finder Chi-Quadrat – Gleichverteilung



ITEM3 Institut

	Beobachtetes N	Erwartete Anzahl	Residuum
1 Institut A	3	37,2	-34,2
2 Institut B	18	37,2	-19,2
3 Institut C	83	37,2	45,8
4 Institut D	29	37,2	-8,2
5 Institut E	76	37,2	38,8
6 Institut F	43	37,2	5,8
7 Institut G	5	37,2	-32,2
8 Institut H	58	37,2	20,8
9 Institut I	41	37,2	3,8
10 Institut J	16	37,2	-21,2
Gesamt	372		

Die kleinste erwartete Häufigkeit muss größer 1 sein. Höchstens 20% dürfen eine erwartete Häufigkeit kleiner 5 aufweisen.
 Für Alternativen siehe oben.

Statistik für Test

	ITEM3 Institut
Chi-Quadrat ^a	192,892
df	9
Asymptotische Signifikanz	,000

Chi-Quadrat-Test:
 Ist die Signifikanz <=0,05?

a. Bei 0 Zellen (.0%) werden weniger als 5 Häufigkeiten erwartet. Die kleinste erwartete Zellenhäufigkeit ist 37,2.

nein
 Kein signifikanter Unterschied

ja
 Signifikanter Unterschied

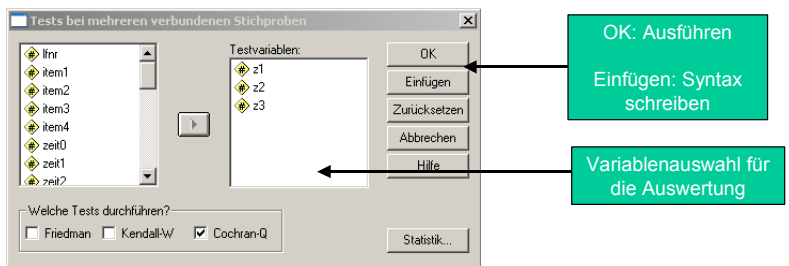
4.2.2 Cochrans-Q

Cochrans-Q

SPSS:

Statistik \ Nichtparametrische Tests
 k verbundene Stichproben

- **Post-Hoc:** Der Test zeigt nicht, zwischen welchen Gruppen Unterschiede bestehen, sondern nur dass/ob überhaupt Unterschiede vorliegen. Einzelunterschiede können mit mehrfachen Paarvergleichen getestet werden. Hier ist der McNemar-Test ein passendes Verfahren. Der Alpha-Fehler muss für die Zahl der Paarvergleiche adjustiert werden. Zudem sind Gleichungen für die Bestimmung von Kontrasten verfügbar (vgl. Diehl & Arbinger, 1990, S. 482).
- **Chi-Quadrat-Verteilung.** Der Test kann exakt berechnet werden und setzt dann keine Verteilung voraus. In der Regel wird er jedoch über die Chi-Quadrat-Verteilung approximiert. Hier gilt, dass für 3 Messzeitpunkte $n=10$ und für 5 Messzeitpunkte $n=8$ die Untergrenze bilden für grobe „Routineentscheidungen“.
- **SPSS:** Die verbundenen Variablen müssen dichotom kodiert sein und müssen dabei gleich kodiert sein.



Bei abhängigen Stichproben liegen pro Fall Messungen für verschiedene Zeitpunkte vor. Der Test erwartet gleich kodierte, dichotome Variablen.

Cochran-Test

Häufigkeiten

	Wert	
	0	1
Z1	70	150
Z2	97	123
Z3	70	150

Statistik für Test

N	220
Cochrans Q-Test	12,462 ^a
df	2
Asymptotische Signifikanz	,002

a. 1 wird als Erfolg behandelt.

Asymptotische Signifikanz für Cochrans-Q-Test:
 Ist die Signifikanz $\leq 0,05$?



4.2.3 Chi-Quadrat, Fisher's exakter Test, Kendals-Tau-b

Chi-Quadrat – Kontingenztafel

SPSS:

Statistik \ Zusammenfassen

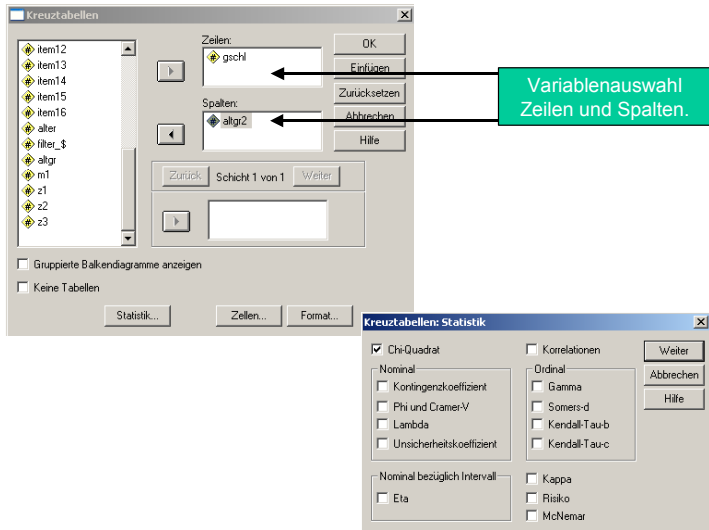
Kreuztabellen

complexity-research.com

..\produkte\test.finder

Chi-Quadrat – Kontingenztafel

- **Post-Hoc:** Der Chi-Quadrat-Test zeigt nicht, zwischen welchen Gruppen Unterschiede bestehen, sondern nur dass/ob überhaupt Unterschiede vorliegen. Einzelunterschiede können mit mehrfachen Paarvergleichen getestet werden. Hier ist der Fisher-Test ein exaktes Verfahren. Der Alpha-Fehler muss für die Zahl der Paarvergleiche adjustiert werden. Zudem sind Gleichungen für die Bestimmung von Kontrasten verfügbar (vgl. Diehl & Arbinger, 1990, S. 455).
- **Chi-Quadrat-Verteilung.** Im Test werden erwartete und beobachtete Häufigkeiten verglichen. Dabei ergibt sich approximativ eine Chi-Quadrat-Verteilung, wenn die erwarteten Häufigkeiten nicht zu klein sind. Als Faustregel gilt heute, dass keine erwartete Häufigkeit kleiner 1 sein darf und höchstens 20% eine erwartete Häufigkeit kleiner 5 aufweisen dürfen. Alternativ können eventuell exakte Tests durchgeführt werden. Auch paarweise Vergleiche mit dem Fisher-Test und eine Alpha-Fehler-Adjustierung sind denkbar. Oder man kann inhaltlich passende Gruppen so zusammen legen, dass die erwarteten Häufigkeiten die Grenze überschreiten.
- **Vorsicht bei geordneter Zellenstruktur:** Dem Test ist es egal, was die Zeilen und Spalten der Kontingenztafel bedeuten. Enthalten diese eine Ordnung, z.B. indem sie die Stufen eines Ratings abbilden, so ist es inhaltlich nicht mehr bedeutungslos, wenn man diese Kategorien vertauscht. Der Test merkt hier aber keinen Unterschied. Beschreiben die Zeilen oder Spalten Ordnungsrelationen, so ist Kendals-Tau (siehe unten) zu rechnen.



GSCHL * ALTGR2 Kreuztabelle

Anzahl		ALTGR2			Gesamt
		1,00	2,00	3,00	
GSCHL	0 männlich	10	69	41	120
	1 weiblich	34	138	75	247
Gesamt		44	207	116	367

Die kleinste erwartete Häufigkeit muss größer 1 sein. Höchstens 20% dürfen eine erwartete Häufigkeit kleiner 5 aufweisen.
 Für Alternativen siehe oben.

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)
Chi-Quadrat nach Pearson	2,395 ^a	2	,302
Likelihood-Quotient	2,522	2	,283
Zusammenhang linear-mit-linear	1,728	1	,189
Anzahl der gültigen Fälle	367		

Chi-Quadrat-Test nach Pearson: Ist die Signifikanz $\leq 0,05$?

a. 0 Zellen (.0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 14,39.

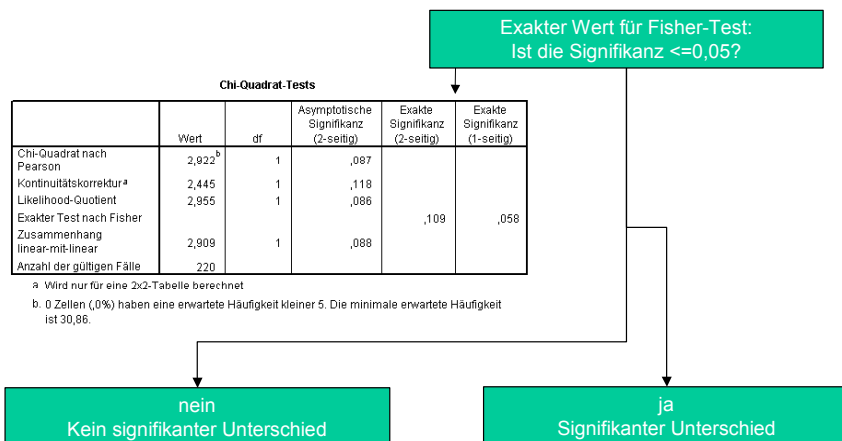
nein
Kein signifikanter Unterschied

ja
Signifikanter Unterschied

Fischer's exakter Test

SPSS:
 Statistik \ Zusammenfassen
 Kreuztabellen
 Exakte Tests

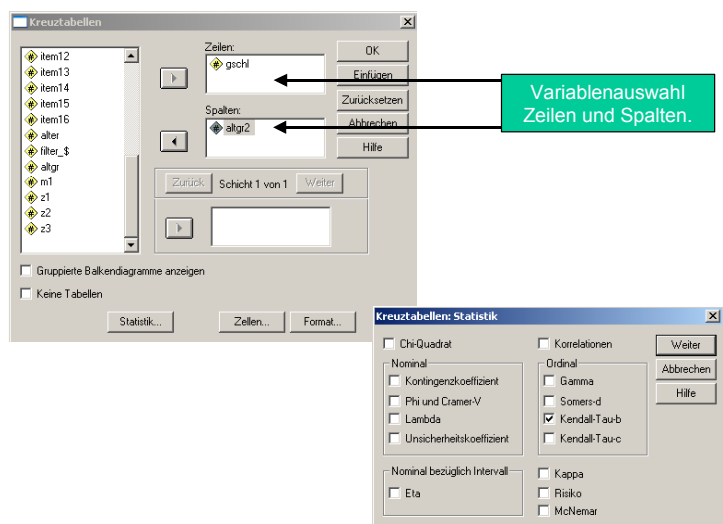
- **Idee:** Wenn Häufigkeiten in zwei Gruppen ermittelt werden, so ergibt sich eine 4-Felder-Kreuztabelle. Das Merkmal für das die Häufigkeit bestimmt wird liegt entweder vor oder nicht, was zwei Bedingungen sind. Weitere zwei ergeben sich für die beiden Gruppen, die untersucht und verglichen werden. Große Kontingenztafeln lassen sich in viele kleine 4-Felder-Tafeln zerlegen.
- **Exakter Test:** Der Test bildet eine exakte Wahrscheinlichkeitsverteilung, die in jedem Fall korrekt ist und keinerlei Einschränkungen unterliegt, wie sie z.B. für den Chi-Quadrat-Test gelten. Aber, bei großen Stichprobenumfängen treten bei der Berechnung auch sehr große Zahlen auf. Einige Computer-Programme geben bei einem N größer 1000 auf. Hier sollte dann doch der Chi-Quadrat-Test herangezogen werden, oder eine Aproximation für den Fisher-Test bestimmt werden.
- **SPSS:** Verfügt SPSS über das Modul Exakte Tests kann der Fischer-Test auch für große Stichproben und für beliebig große Kontingenztafeln bestimmt werden.
- **Vorsicht bei geordneter Zellenstruktur:** Dem Test ist es egal, was die Zeilen und Spalten der Kontingenztabelle bedeuten. Enthalten diese eine Ordnung, z.B. indem sie die Stufen eines Ratings abbilden, so ist es inhaltlich nicht mehr bedeutungslos, wenn man diese Kategorien vertauscht. Der Test merkt hier aber keinen Unterschied. Beschreiben die Zeilen oder Spalten Ordnungsrelationen, so ist Kendals-Tau (siehe unten) zu rechnen.



Kendals-Tau-b

SPSS:
Statistik \ Zusammenfassen
Kreuztabellen

- **Beim Chi-Quadrat-Test bzw. Fisher-Test gilt: Vorsicht bei geordneter Zellenstruktur:** Den Verfahren ist es egal, was die Zeilen und Spalten der Kontingenztafel bedeuten. Enthalten diese eine Ordnung, z.B. indem sie die Stufen eines Ratings abbilden, so ist es inhaltlich nicht mehr bedeutungslos, wenn man diese Kategorien vertauscht. Der Test merkt hier aber keinen Unterschied. Beschreiben die Zeilen oder Spalten Ordnungsrelationen, so ist Kendals-Tau (siehe unten) zu rechnen.
- Es handelt sich bei Kendals-Tau-b um ein Zusammenhangsmaß zwischen Rangordnungsdaten.
- **Rangbindungen.** Es wird mit Rängen gerechnet und so kann es vorkommen, dass mehrere Fälle den gleichen Rang aufweisen. Rangbindungen können zwar herausgerechnet werden, verschlechtern aber die Statistik.



GSCHL * ALTGR2 Kreuztabelle

Anzahl		ALTGR2			Gesamt
		1,00	2,00	3,00	
GSCHL	0 männlich	10	69	41	120
	1 weiblich	34	138	75	247
Gesamt		44	207	116	367

Symmetrische Maße

		Wert	Asymptotischer Standardfehler ^a	Näherungsweise T ^b	Näherungsweise Signifikanz
Ordinal- bzgl. Ordinalmaß	Kendall-Tau-b	-,062	,049	-1,263	,207
Anzahl der gültigen Fälle					

a. Die Null-Hyphothese wird nicht angenommen.
 b. Unter Annahme der Null-Hyphothese wird der asymptotische Standardfehler verwendet

Asymptotische Signifikanz für Kendall-Tau-b:
 Ist die Signifikanz <=0,05?

