

Datenanalyse

Statistische Methoden

PD Dr. Dr. Dipl.-Psych. Guido Strunk

Hidden Features, Making of, and Additional Resources at ...

<http://www.complexity-research.com/ProjekteLehre.htm>



Inhalt

1.	Software.....	3
1.1	Excel.....	3
1.2	GStat.....	3
1.3	SPSS.....	5
1.3.1	Struktur und Aufbau.....	5
1.3.2	Grundsätzliches.....	7
1.3.3	Erstellung von Datenfiles und Kodierung von Fragebögen.....	9
1.3.4	Grundsätzliches zu Datenauswertung mit SPSS.....	11
1.3.5	Deskriptive Statistiken.....	15
1.3.6	Testverfahren im SPSS.....	16
1.3.7	Hilfefunktionen.....	48
1.3.8	Test-Finder.....	49
2.	Glossar für einige wichtige statistische Begriffe.....	51
3.	Darstellung und Abkürzungen.....	56
3.1	Allgemein.....	56
3.2	Korrelationen.....	57
3.3	T-Test.....	58
3.4	Varianzanalyse.....	59
4.	Literatur.....	60

1. Software

1.1 Excel

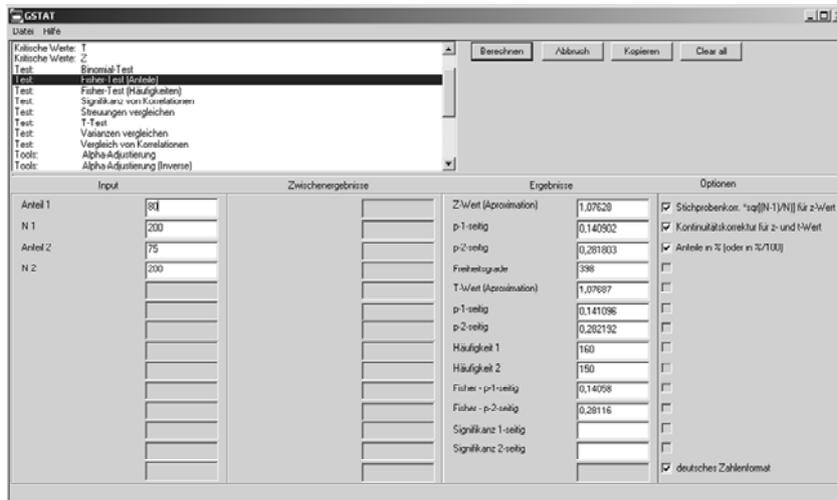
Was geht im Excel?

- Im Excel können Mittelwert, Median, Standardabweichung und Varianz (jeweils für Stichprobe oder Population) berechnet werden.
- Als Testverfahren sind der t-Test und der F-Test vorhanden.
 - Beim t-Test muss angegeben werden, ob von Varianzhomogenität ausgegangen werden kann.
 - Dazu muss zunächst der F-Test für die Varianzen gerechnet werden. Ist das Ergebnis (Wahrscheinlichkeit p) größer 0,2, so kann von Varianzhomogenität ausgegangen werden. Ist $p \leq 0,2$, dann sollte der t-Test für ungleiche Varianzen gerechnet werden.
- Auch die Chi-Verteilung ist im Excel vorhanden. Damit kann die Wahrscheinlichkeit $(1-\text{Alpha})$ für einen berechneten χ^2 -Wert bestimmt werden.
- Korrelationen: Eine einfache Produktmoment-Korrelation kann bestimmt werden.

1.2 GStat

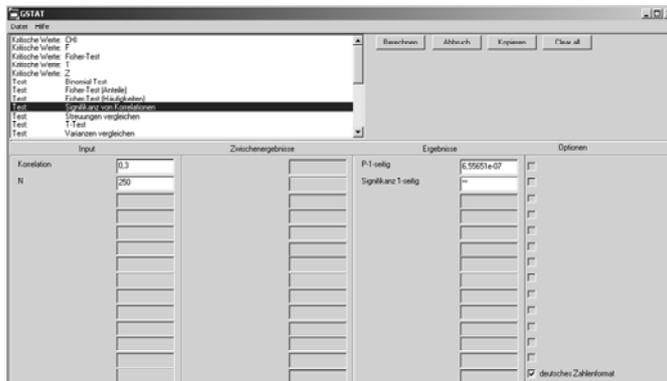
- **Unter www.complexity-research.com, Menü: Downloads, Software kann Gstat heruntergeladen werden.**
 - Sind Mittelwerte oder andere Kennwerte bekannt, kann mit Gstat die Signifikanz bestimmt werden.
 - Beispiel: Fisher-Test für den Vergleich von zwei Prozentwerten.
 - Wenn in zwei Gruppen jeweils 200 Personen befragt wurden und in der einen Gruppe 80% die Küche des Spitals für miserabel halten und in der anderen Stichprobe 75% stellt sich die Frage, ob der Unterschied hier signifikant ist.

GStat



Korrelationsrechnung Signifikanztestung in GStat

- Die Null-Hypothese geht davon aus, dass eine Null-Korrelation vorliegt.
 $H_0: r = 0$.
- Es kann 1-seitig (Vorhersage der Korrelationsrichtung) oder 2-seitig geprüft werden..



1.3 SPSS

Themenübersicht

- Struktur und Aufbau des Programmpaketes SPSS.
- Grundsätzliches zum Arbeiten mit SPSS.
- Erstellung von Datenfiles und Kodierung von Fragebögen.
- Grundsätzliches zur Datenauswertung mit SPSS.
- Deskriptive Statistiken in SPSS.
- T-Test, 1-faktorielle Varianzanalyse.
- Korrelationen.
- Nichtparametrische Verfahren.
- Hilfsfunktionen, Anleitungen und Literatur.

1.3.1 Struktur und Aufbau

Daten-Editor

Jede Zeile ist ein Fall (Case)

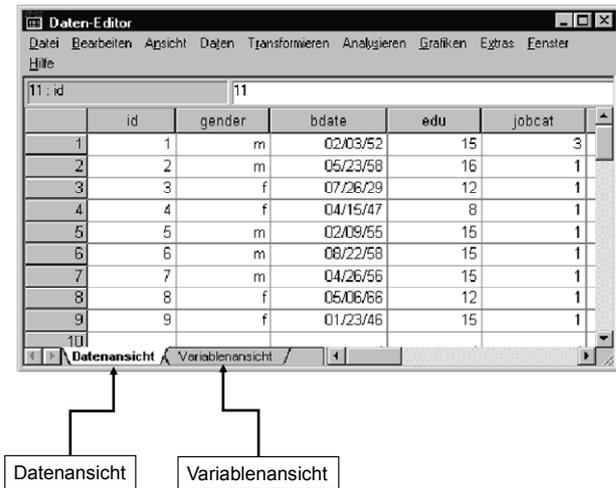
Jede Spalte ist eine Variable

Fehlende Angaben bleiben leer

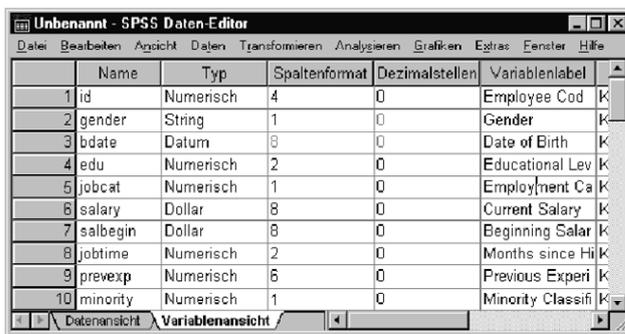
Case	auf	dname	pre_1	pre_5	pre_6	pre_7	pre_8	pre_9	ye_10	pre_11	pre_12	pre_13
1	.00	10000000011003447800	.00395	.00278	.00091	.00606	.00642	.00642	.00483	.00262	.00258	.
2	1.00	10000000011000000000	.00967	.01630	.00000	.00611	.00906	.06773	.01637	.01084	.03887	.
3	.00	10000000011003101000	.01106	.01038	.00406	.02484	.04674	.06214	.02962	.01981	.04281	.
4	.00	10000000011003121000	.01967	.01264	.01096	.02962	.04674	.01468	.10846	.02240	.01967	.
5	1.00	10000000011000000000	.00484	.05900	.05923	.00691	.01409	.05772	.00789	.03120	.01414	.03628
6	.00	10000000011000921801	.00395	.00278	.00091	.00606	.00642	.00642	.00483	.00262	.00258	.
7	.00	10000000011003101000	.00380	.02356	.01311	.00106	.04574	.06808	.06513	.03130	.08271	.04516
8	1.00	10000000011000000000	.00256	.06479	.00804	.03110	.04688	.12426	.02102	.02113	.01414	.04166
9	.00	10000000011000000000	.01142	.06347	.06999	.01411	.01121	.01488	.01611	.01112	.01414	.04916
10	1.00	10000000011000400000	.04577	.15021	.07516	.00917	.02469	.02451	.11865	.02625	.04888	.12286
11	.00	10000000011001200000	.00794	.01714	.02063	.03883	.04468	.06772	.00003	.02713	.01777	.
12	.00	10000000011002200000	.02109	.02101	.04064	.01264	.00091	.02401	.01589	.01166	.01110	.
13	.00	10000000011002600000	.06266	.01119	.00860	.01109	.01409	.06461	.04191	.02115	.01414	.
14	.00	10000000011002620000
15	.00	10000000011001950000	.00395	.00278	.00091	.00606	.00642	.00642	.00483	.00262	.00258	.
16	1.00	10000000011000160000	.06876	.11082	.03997	.06120	.01630	.02451	.00963	.04801	.04688	.06869
17	.00	10000000011000000000	.00312	.00700	.02011	.01177	.01460	.01400	.01710	.00789	.04332	.00571
18	.00	10000000011004500000
19	.00	10000000011004900004	.00482	.00700	.00784	.00020	.01409	.01372	.00963	.01899	.01360	.
20	1.00	10000000011000000000	.06171	.09247	.07249	.04495	.01409	.04451	.20040	.06719	.03478	.
21	.00	10000000011001950000	.00395	.00278	.00091	.00606	.00642	.00642	.00483	.00262	.00258	.

Eine Zelle enthält Werte (Values)

In der Regel wird mit Zahlen gearbeitet

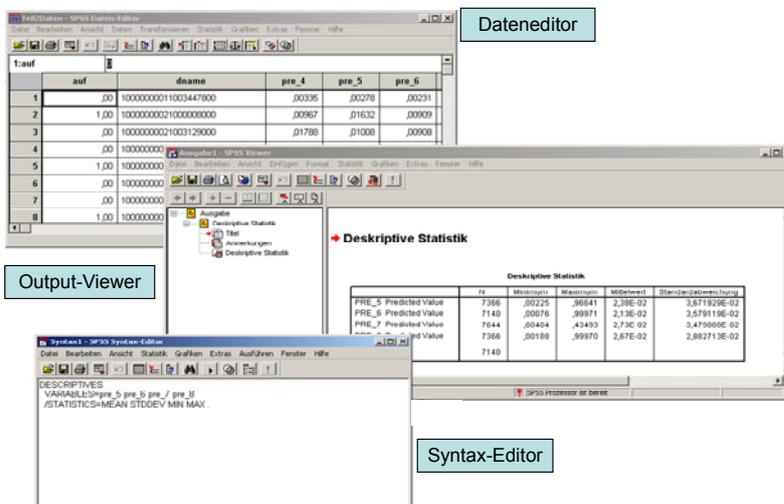


Variablenansicht



- Variablenname
- Datentyp
- Anzahl Ziffern oder Zeichen
- Anzahl Dezimalstellen
- Beschreibende Variablen- und Wertelabels
- Benutzerdefinierte fehlende Werte
- Spaltenbreite
- Messniveau

Programm-Struktur



Datei-Endungen

The screenshot displays three SPSS windows:

- Dateneditor: .sav**: Shows a data grid with columns 'auf', 'dname', 'pre_4', 'pre_5', and 'pre_6'. The data is as follows:

auf	dname	pre_4	pre_5	pre_6
.00	10000000011003447800	.00335	.00278	.00231
1.00	10000000021000008000	.00967	.01632	.00909
.00	10000000021003129000	.01788	.01008	.00908
.00	1000000000			
1.00	1000000000			
.00	1000000000			
1.00	1000000000			
- Output-Viewer: .spo**: Shows a 'Deskriptive Statistik' table with the following data:

	N	Minimum	Maximum	Mittelwert	Standardabweichung
PRE_5 Predicted Value	7366	.00225	.86641	2,39E-02	3,871929E-02
PRE_6 Predicted Value	7140	.00076	.89971	2,13E-02	3,578118E-02
PRE_7 Predicted Value	7644	.00404	.43493	2,73E-02	3,479900E-02
PRE_8 Predicted Value	7366	.00188	.89970	2,67E-02	2,882713E-02
	7140				
- Syntax-Editor: .sps**: Shows the following SPSS syntax code:


```
DESCRIPTIVES
  VARIABLES=pre_5 pre_6 pre_7 pre_8
  /STATISTICS=MEAN STDDEV MIN MAX.
```

1.3.2 Grundsätzliches

Statistisches Wissen

- SPSS ist ein Statistik-Programm für Profis.
- Es umfasst eine Fülle von Funktionen, statistischen Methoden, Algorithmen und Varianten von Methoden.
- Sie müssen wissen was Sie tun. Das Programm entscheidet nicht über die Sinnhaftigkeit einer Berechnung: Es rechnet auch den Mittelwert von Hausnummern aus.
- Die Ergebnisse der Berechnungen müssen von Ihnen interpretiert werden.
- Nicht immer benutzt SPSS die gleichen Methoden und Methodenbezeichnungen, wie gängige Statistik-Bücher.

Unterschiede zu Excel

- **SPSS ist bewusst unflexibel bei der Dateneingabe.**
- **SPSS ist für die Analyse großer Datensätze geeignet. Auch einfache Fragebögen führen schnell zu über 255 Variablen. Internationale Datensätze können schnell auch mal über 100.000 Cases enthalten.**
- **Berechnungen liefern im SPSS eine Fülle von Detail- und Zwischenergebnissen, die automatisch in Tabellen und Grafiken dargestellt werden. Im Excel wird nur eine Zahl berechnet.**
- **SPSS verfügt über keine automatische Aktualisierung von Ergebnissen oder von Zahlen im Dateneditor.**
- **(in älteren Versionen ist SPSS in den Variablennamen beschränkt auf 8 Zeichen – keine Umlaute – und kann nur jeweils ein Tabellenblatt geöffnet haben.)**

Sonstige Besonderheiten

- **Man kann nur rechnen, wenn die Daten in einem Datenblatt vorhanden sind. Der Vergleich zweier Datenblätter oder eines Datenblattes mit Werten aus der Literatur ist nur schwer oder gar nicht möglich.**
- **Tabellen und Grafiken sind nur selten direkt in wissenschaftlichen Arbeiten nutzbar. Sie enthalten in der Regel mehr Infos als man benötigt oder sind nur schwer zu formatieren. Ein Austausch der Ergebnisse über die Zwischenablage ist aber gut möglich.**
- **SPSS kostet zwischen 1.000 und 5.000 Euro. Studentenversionen und Demoversionen sind jedoch erhältlich.**

1.3.3 Erstellung von Datenfiles und Kodierung von Fragebögen

Definition der Variablen in SPSS

- **Variablenname (8 Zeichen, keine deutschen Sonderzeichen, jeden Namen nur ein mal vergeben), z.B.:**
 - lfnr
 - item1
 - item2
 - item3
 - item4_1
 - item4_2
 - item4_3
 - gschl
 - Alter
- **Datentyp (in der Regel „Numerisch“, aber auch „String“ für Texte und „Datum“ üblich)**

Definition der Variablen in SPSS

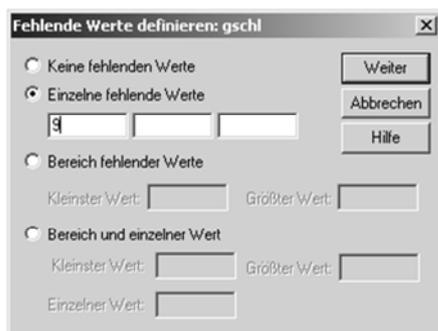
- **Anzahl Ziffern oder Zeichen (bei Text maximal 255)**
- **Anzahl Dezimalstellen (bei ganzen Zahlen sieht es schöner aus, wenn hier 0 eingegeben wird, ist aber egal)**
- **Beschreibende Variablen- und Wertelabels**
- **Benutzerdefinierte fehlende Werte**
- **Spaltenbreite (nur für die Anzeige relevant)**
- **Messniveau (Nominal, Ordinal, Metrisch). Dient als Gedächtnisstütze, hat aber sonst keine Bedeutung.**

Beschreibende Wertelabels



Zur Vergabe von Wertelabels den Wert und den Label eingeben und mit „Hinzufügen“ zur Liste hinzufügen. Am Schluss mit „Weiter“ den Dialog verlassen.

Benutzerdefinierte fehlende Werte



„weiß ich nicht“ ist eine Angabe der befragten Person, die kodiert werden sollte, aber bei Berechnungen ignoriert werden muss. Es handelt sich um einen „Definierten fehlenden Wert“ (Missing).
Fehlen Angaben im Fragebogen, kann bei der Dateneingabe einfach die Zelle im Dateneditor übersprungen werden. Man spricht hier von einem „System Missing“.

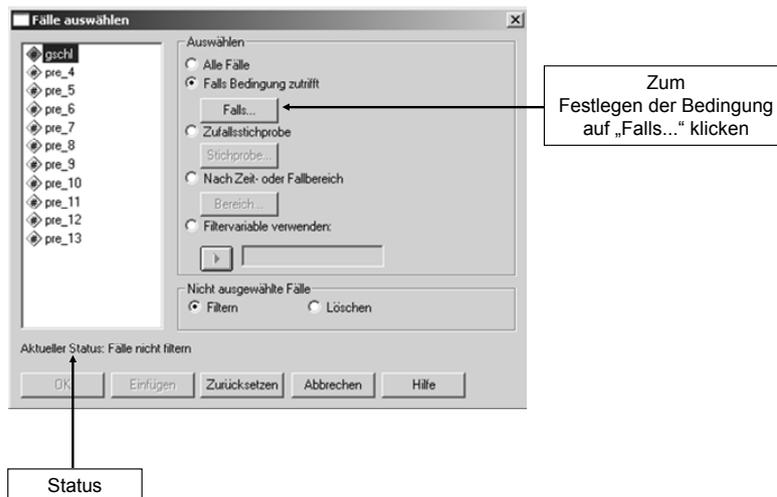
Eine der vier Optionen wählen und geforderte Angaben ausfüllen.
Am Schluss mit „Weiter“ den Dialog verlassen.

1.3.4 Grundsätzliches zu Datenauswertung mit SPSS

Vorbereitende Schritte

- Soll nur mit einem Teil der Daten gerechnet werden?
(Falls ja => Menü: Daten / Fälle auswählen ...)
- Müssen aus den Daten zunächst neue Variablen erzeugt werden?
(Falls ja => Menü: Transformieren / Berechnen ...)

Fälle auswählen ...



Fälle auswählen ... Falls...

Formel für die Bedingung

~ bedeutet „nicht“

Sobald die Formel steht, Kann der Dialog mit „Weiter“ geschlossen werden.

Variablenliste

Kopiert eine ausgewählte Variable in den Formel-Editor

Fälle auswählen ... Falls ... beendet

Bedingung wird angezeigt

Löschen?

„OK“ und „Einfügen“ wird verfügbar

Filter an

1:gschl	gschl	dname	pre_4	pre_5	pre_6	pre_7
1	0	10000000011003447800	,00335	,00278	,00231	,0060
2	1	10000000021000000000	,00967	,01632	,00909	,0291
3	0	10000000021003129000	,01788	,01008	,00908	,0246
4	0	10000000021003129002	,01369	,01284	,01036	,0258
5	1	10000000031000020800	,00484	,03309	,03323	,0063
6	0	10000000031000321801	,00335	,00278	,00231	,0060
7	0	10000000041003404059	,09380	,03256	,04314	,0919
8	1	10000000051000039000	,10255	,05478	,09004	,0371

Auswahl der Fälle

Filterstatus

Transformieren / Berechnen ...

Formel

Variablenliste

neue Zielvariable definieren

Kopiert eine ausgewählte Variable in den Formel-Editor

Transformieren / Berechnen ... Falls...

Gruppe=1

Die neue Variable „gruppe“ bekommt den Wert „1“ falls...

neue Zielvariable definieren

Transformieren / Berechnen ... Falls...

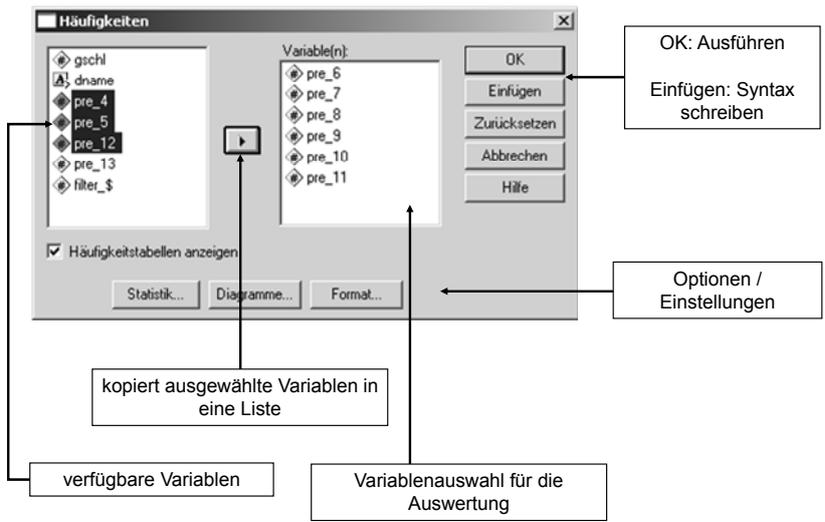
2. Bedingung

3. „Weiter“

1. Fall einschließen, wenn...

**Und-Verknüpfung: AND
Oder-Verknüpfung: OR**

Typischer SPSS-Dialog

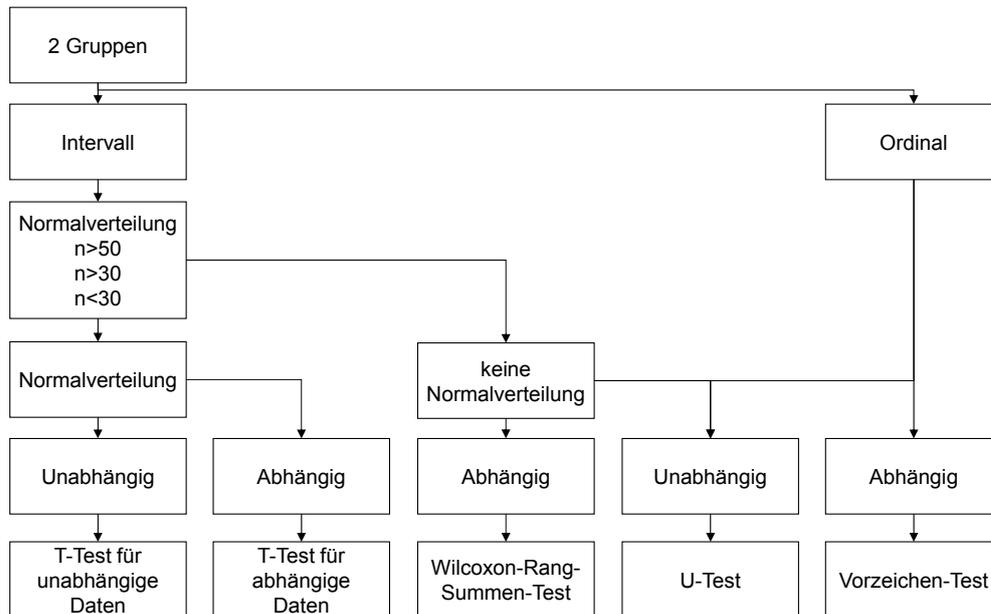


1.3.5 Deskriptive Statistiken

Statistik \ Zusammenfassen \ Häufigkeiten

Statistik \ Zusammenfassen \ Deskriptive Statistiken

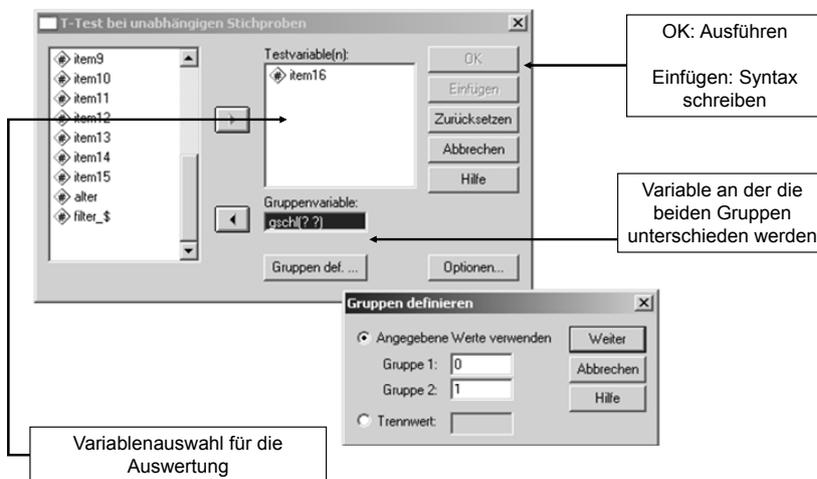
1.3.6 Testverfahren im SPSS



T-Test – unabhängige Daten

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

- **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.



T-Test

Gruppenstatistiken				
	OSCHL	N	Mittelwert	Standardabweichung
ITEM16	1	123	1,67	,79
Gesamtfriedenheit	2	248	1,35	,74

		Levene-Test der Varianzhomogenität		T-Test für die Mittelwertgleichheit			
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz der Differenz
ITEM16	Varianzen sind gleich	,266	,607	-1,037	368	,300	-8,33E-02
Gesamtfriedenheit	Varianzen sind nicht gleich			-1,060	257,815	,290	-8,33E-02

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

nein
Varianzen sind nicht gleich: untere Zeile

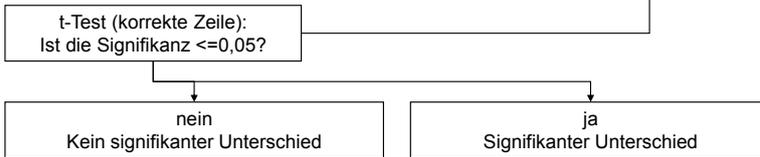
ja
Varianzen sind gleich: obere Zeile

- Titel
- Anmerkungen
- Gruppenstatistiken
- Test bei unabhängigen Stichproben

T-Test

Gruppenstatistiken					
	GESCHL	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
ITEM16 Gesamtbefriedenheit	männlich	123	1,87	,70	6,39E-02
	weiblich	248	1,75	,74	4,72E-02

		Levene-Test der Varianzhomogenität		T-Test für die Mittelwertgleichheit				
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz der Differenz	Standardfehler der Differenz
ITEM16 Gesamtbefriedenheit	Varianzen sind gleich	,206	,650	-1,037	389	,300	-6,39E-02	6,03E-02
	Varianzen sind nicht gleich			-1,080	257,815	,290	-6,39E-02	7,86E-02

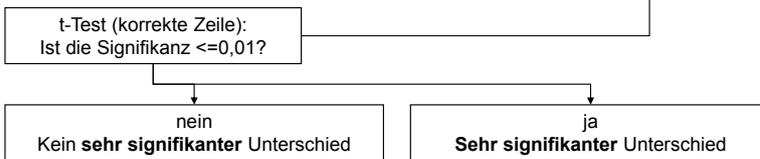


- Titel
- Anmerkungen
- Gruppenstatistiken
- Test bei unabhängigen Stichproben

T-Test

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

		Levene-Test der Varianzhomogenität		T-Test für die Mittelwertgleichheit				
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz der Differenz	Standardfehler der Differenz
ITEM16 Gesamtbefriedenheit	Varianzen sind gleich	,206	,650	-1,037	389	,300	-6,39E-02	6,03E-02
	Varianzen sind nicht gleich			1,080	267,816	,290	6,39E-02	7,86E-02

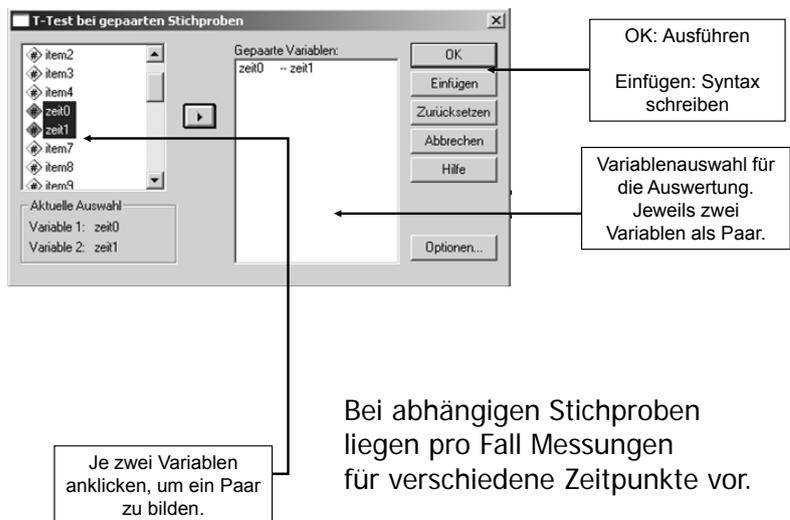


1- oder 2-seitig

- *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.

T-Test – abhängige Daten

- SPSS:
- Statistik \ Mittelwerte vergleichen
T-Test bei gepaarten ...
- **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.



Statistik bei gepaarten Stichproben

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren	ZEIT0	1,90	358	,95	5,04E-02
	ZEIT1	1,36	358	,81	3,24E-02

Korrelationen bei gepaarten Stichproben

		N	Korrelation	Signifikanz
Paaren	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

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

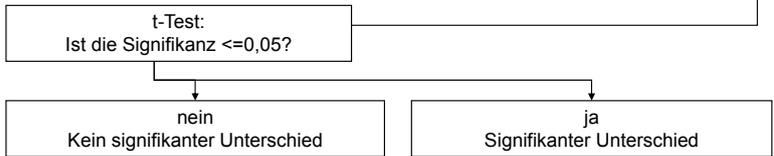
		Mittelwert	N	Standardabweichung	des Mittelwertes
Paaren	ZEIT0	1,90	358	,95	5,04E-02
	ZEIT1	1,36	358	,81	3,24E-02

Korrelationen bei gepaarten Stichproben

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

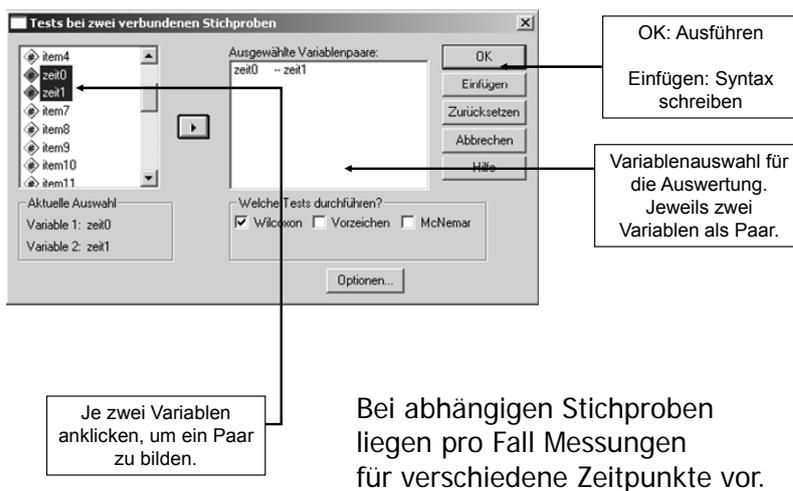
Test bei gepaarten Stichproben

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



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.



Wilcoxon-Test

Range

		N	Mittlerer Rang	Rangsumme
ZEIT1 - ZEIT0	Negative Ränge	157 ^a	87,46	13731,50
	Positive Ränge	15 ^b	76,43	1146,50
	Bindungen	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

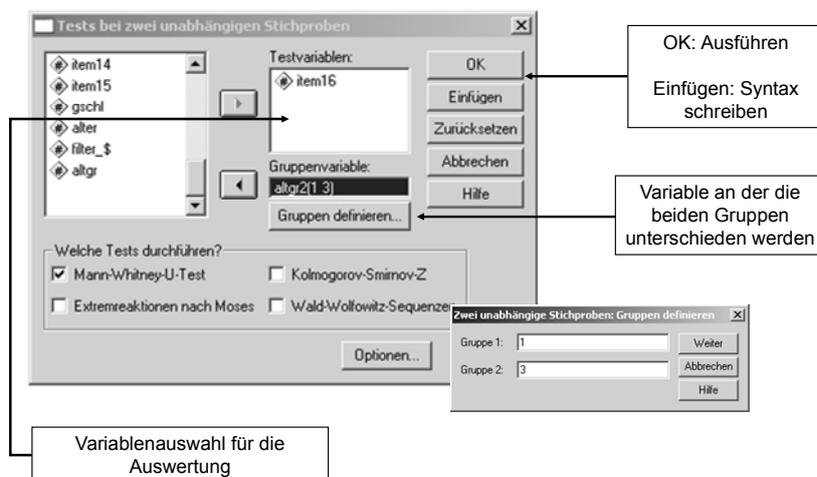
- a. Basiert auf positiven Rängen.
- b. Wilcoxon-Test

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

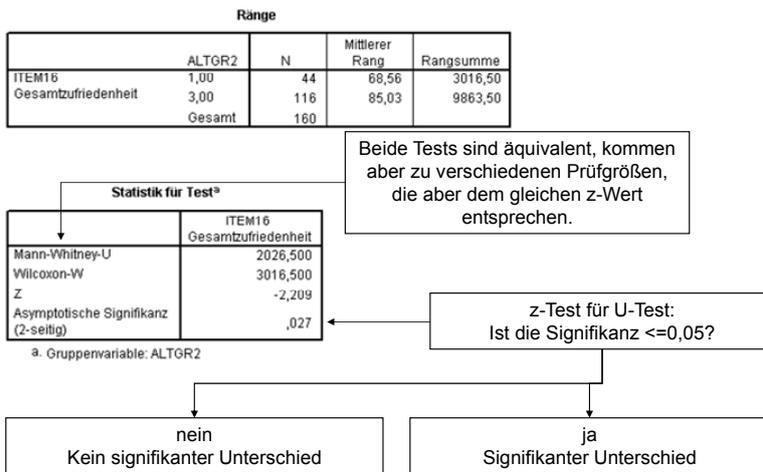


U-Test

- 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.

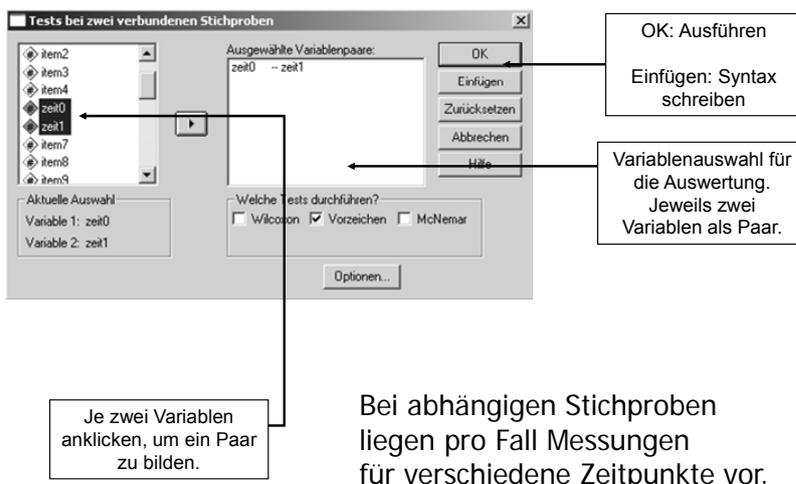


Mann-Whitney-Test



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.



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

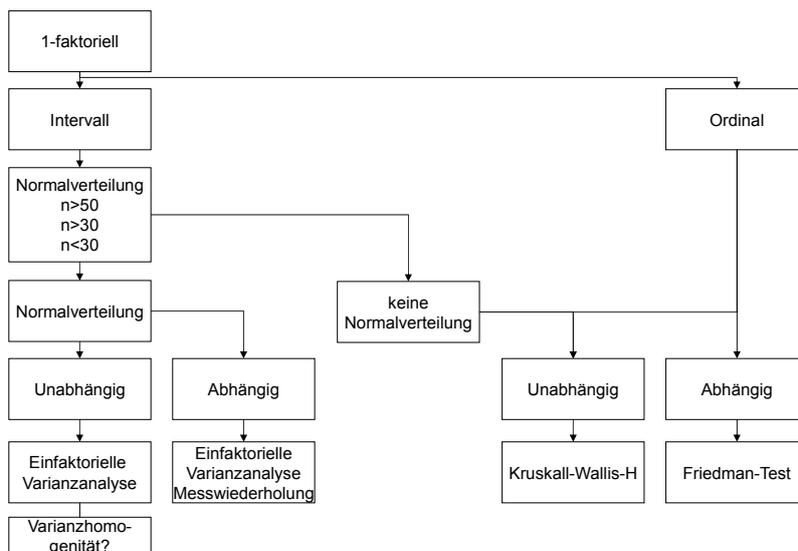
Statistik für Test ^a	
	ZEIT1 - ZEIT0
Z	-10,751
Asymptotische Signifikanz (2-seitig)	,000

a. Vorzeichentest

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



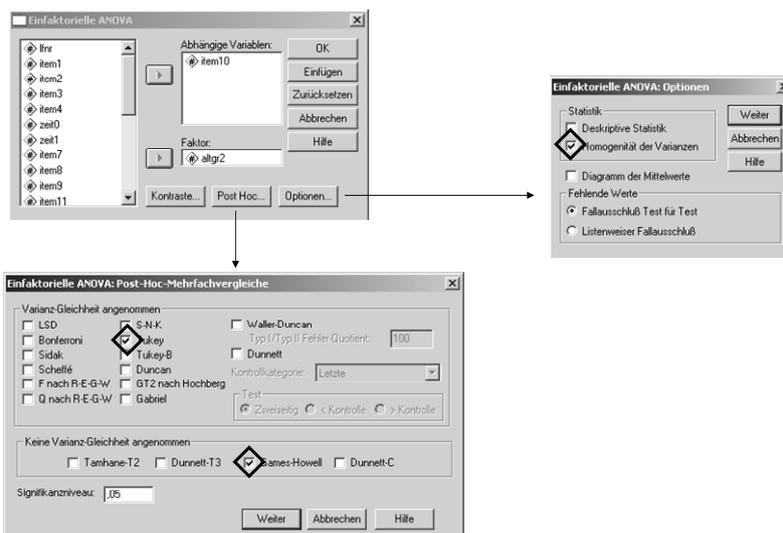
Mehr als zwei Gruppen / 1 Faktor



Einfaktorielle Varianzanalyse (ANOVA)

- SPSS:
- Statistik \ Mittelwerte vergleichen
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.



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

	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.

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	7,32E-02	,209	,935	-,42	,56
	2,00	1,00	-5,77E-02	,195	,953	-,52	,40
		3,00	1,55E-02	,137	,993	-,31	,34

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	7,32E-02	,209	,935	-,42	,56
	2,00	1,00	-5,77E-02	,195	,953	-,52	,40
		3,00	1,55E-02	,137	,993	-,31	,34
Games-Howell	1,00	2,00	5,77E-02	,195	,947	-,38	,50
		3,00	7,32E-02	,209	,930	-,41	,55
	2,00	1,00	-5,77E-02	,195	,947	-,50	,38
		3,00	1,55E-02	,137	,993	-,31	,35

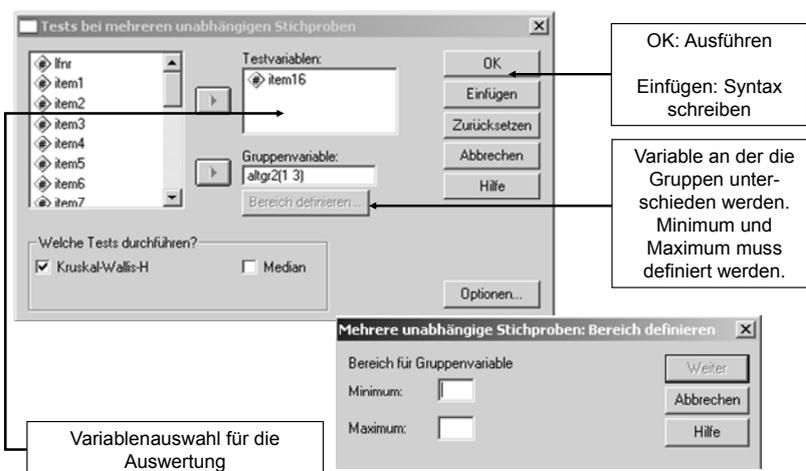
Paarweise wird jede Gruppe mit jeder anderen verglichen. Ist die Signifikanz $\leq 0,05$ bzw. $0,01$ liegt ein signifikanter Unterschied vor.

Varianzanalyse mit Messwiederholung

- SPSS:
- Statistik \ Allgemeines lineares Modell
GLM-Messwiederholung ...

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).



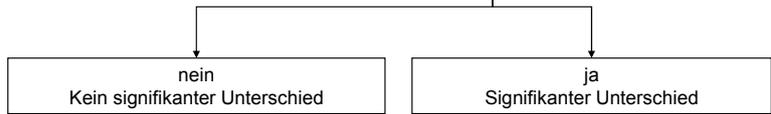
Kruskal-Wallis-Test

Ränge			
	ALTGR2	N	Mittlerer Rang
ITEM16	1,00	44	165,81
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

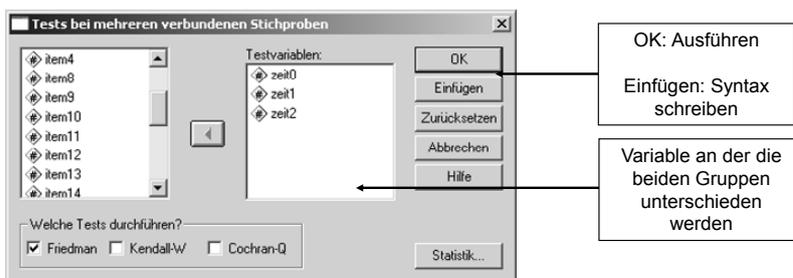
a. Kruskal-Wallis-Test
b. Gruppenvariable: ALTGR2

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



Friedman-Test

- SPSS:
- Statistik \ Nichtparametrische Tests
k verbundene Stichproben ...
- 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.



Friedman-Test

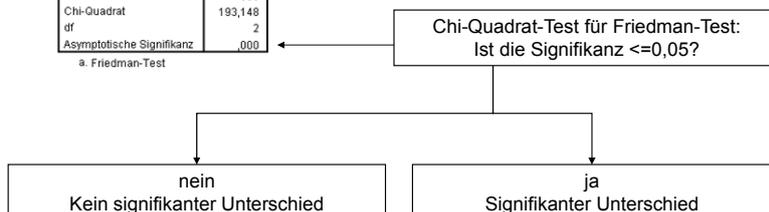
Ränge

	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

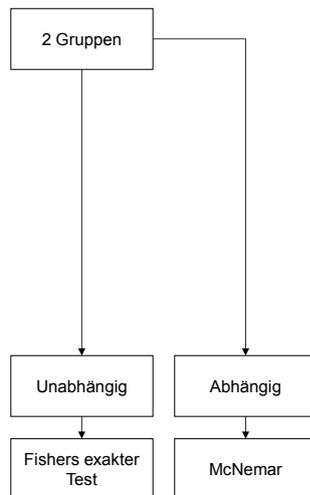
a. Friedman-Test



Mehrfaktorielle Designs

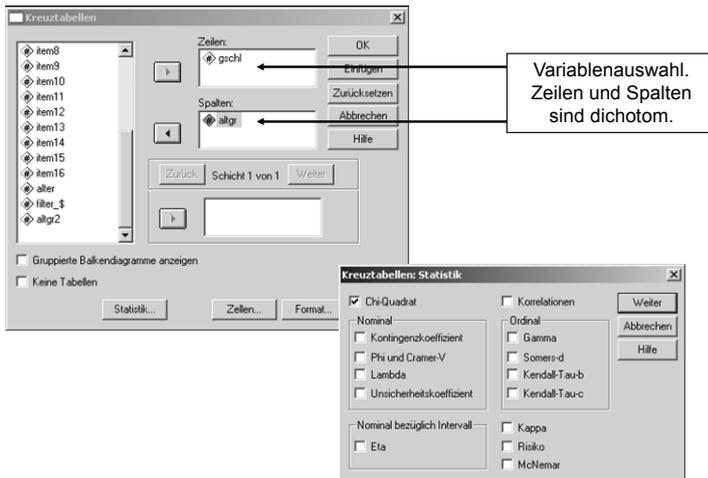
- **Chi-Quadrat:** Man kann versuchen die Daten in Häufigkeiten umzuwandeln (z.B: Median-Split).
- **Mehrfaktorieller Friedman-Test:** Es gibt Verallgemeinerungen des Friedman-Testes für mehr als einen Faktor. Diese sind in SPSS nicht implementiert.
- **GLM:** Mehrfaktorielle Designs finden sich in SPSS 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.

2 Gruppen / Häufigkeiten



Fishers exakter Test

- SPSS:
 - Statistik \ Zusammenfassen
 - Kreuztabellen
-
- **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 Fisher-Test auch für große Stichproben und für beliebig große Kontingenztafeln bestimmt werden.



Variablenauswahl.
Zeilen und Spalten
sind dichotom.

GSCHL * ALTGR Kreuztabelle

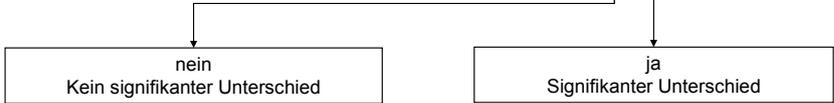
Anzahl		ALTGR		Gesamt
		1,00	2,00	
GSCHL	0 männlich	25	45	70
	1 weiblich	72	78	150
Gesamt		97	123	220

Exakter Wert für Fisher-Test:
Ist die Signifikanz $\leq 0,05$?

Chi-Quadrat-Tests

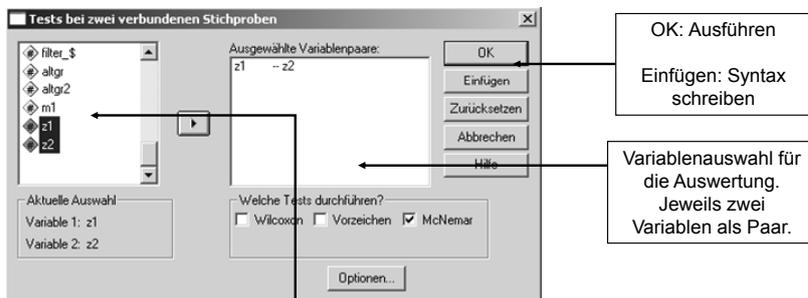
	Wert	df	Asymptotische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)
Chi-Quadrat nach Pearson	2,922 ^a	1	,087		
Kontinuitätskorrektur ^a	2,445	1	,118		
Likelihood-Quotient	2,955	1	,086		
Exakter Test nach Fisher				,109	,058
Zusammenhang linear-mit-linear	2,909	1	,088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



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.



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

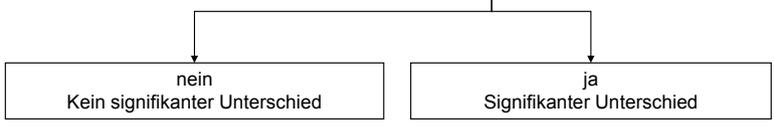
	Z2	
Z1	0	1
0	25	45
1	72	78

Statistik für Test^a

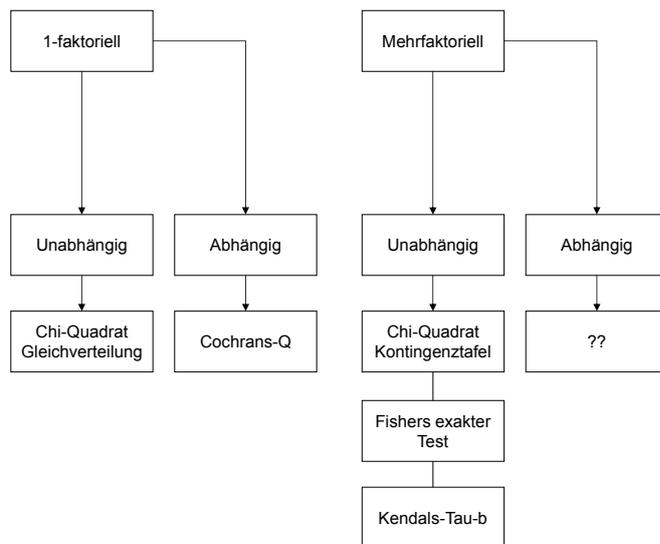
	1 & 2
N	220
Chi-Quadrat ^b	5,778
Asymptotische Signifikanz	,016

a. Kontinuität korrigiert
b. McNemar-Test

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



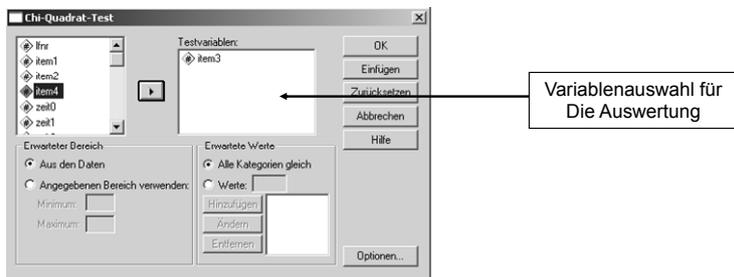
>2 Gruppen / Häufigkeiten



Chi-Quadrat – Gleichverteilung

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



ITEM3 Institut

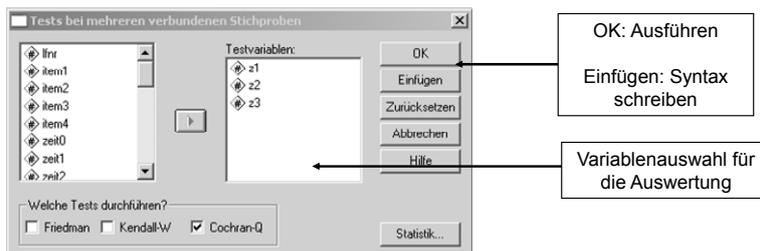
	Beobachtetes N	Erwartete Anzahl	Residuum
1 Institut A	3	37,2	-34,2
2 Institut B	18	37,2	-19,2
3 Institut C	83	37,2	45,8
4 Institut D	29	37,2	-8,2
5 Institut E	76	37,2	38,8
6 Institut F	43	37,2	5,8
7 Institut G	5	37,2	-32,2
8 Institut H	58	37,2	20,8
9 Institut I	41	37,2	3,8
10 Institut J	16	37,2	-21,2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



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

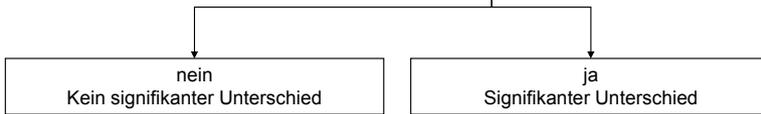
	Wert	
	0	1
Z1	70	150
Z2	97	123
Z3	70	150

Statistik für Test

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

^a. 1 wird als Erfolg behandelt.

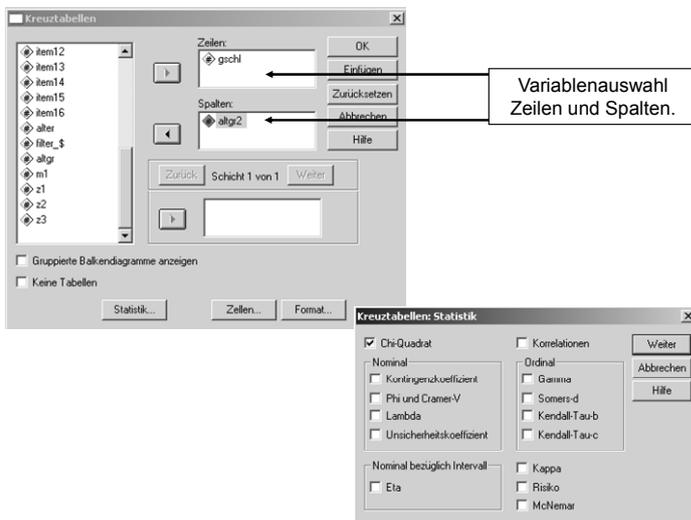
Asymptotische Signifikanz für Cochrans-Q-Test:
Ist die Signifikanz $\leq 0,05$?



Chi-Quadrat – Kontingenztafel

- SPSS:
 - Statistik \ Zusammenfassen
 - Kreuztabellen
-
- **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		

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

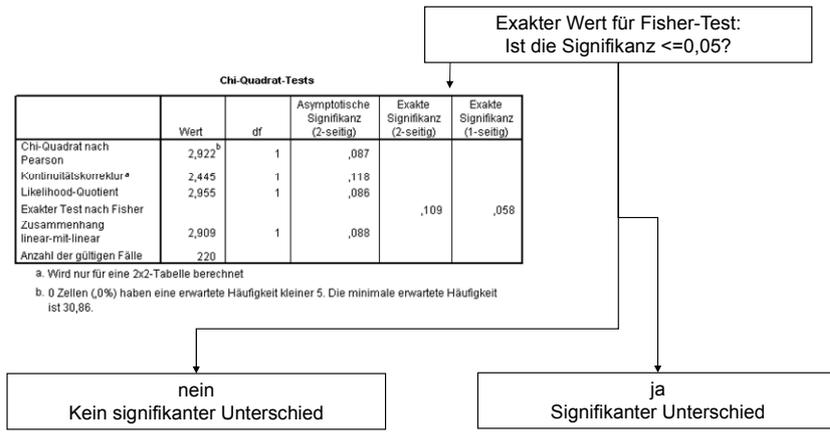
Chi-Quadrat-Test nach Pearson:
Ist die Signifikanz $\leq 0,05$?

nein
Kein signifikanter Unterschied

ja
Signifikanter Unterschied

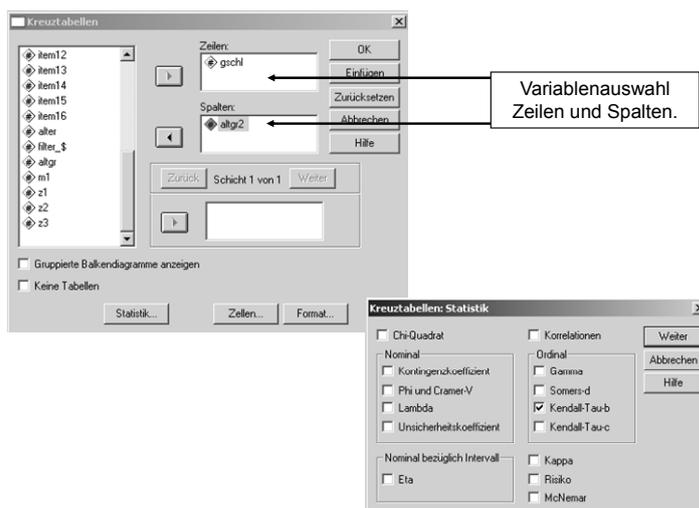
Fishers exakter Test

- 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 Fisher-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 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.



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 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.
 - 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

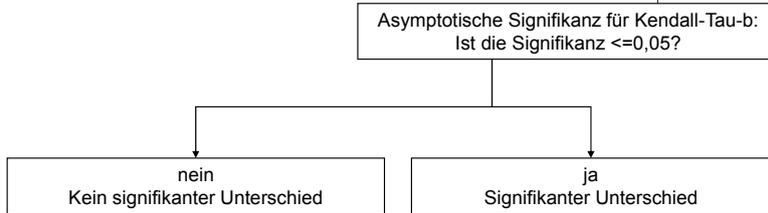
Anzahl		ALTGR2			Gesamt
		1,00	2,00	3,00	
GSCHL	0 männlich	10	69	41	120
	1 weiblich	34	138	75	247
Gesamt		44	207	116	367

Symmetrische Maße

Ordinal- bzgl. Ordinalmaß	Kendall-Tau-b	Wert	Asymptotischer Standardfehler ^a	Näherungsweise T^b	Näherungsweise Signifikanz
Ordinal- bzgl. Ordinalmaß	Kendall-Tau-b	-,062	,049	-1,263	,207
Anzahl der gültigen Fälle		367			

a. Die Null-Hypothese wird nicht angenommen.

b. Unter Annahme der Null-Hypothese wird der asymptotische Standardfehler verwendet.



1.3.7 Hilfefunktionen

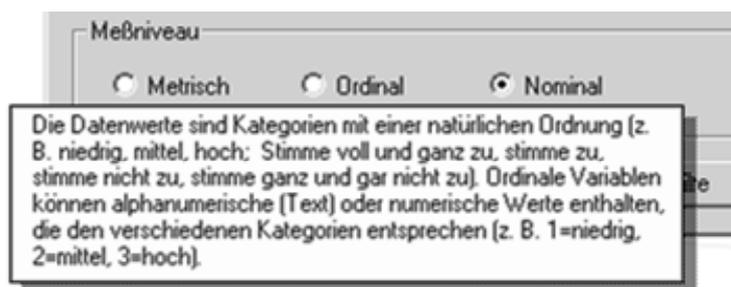
Hilfen...

- Die Hilfe von SPSS ist durchaus komfortabel. Es gibt einen Statistik-Assistenten und für alle Teile des Programmpaketes gibt es Online-PDF-Handbücher.
- Die Pop-Up-Hilfe bietet eine gute Unterstützung während der Arbeit.
- Zahlreiche Handbücher zum Arbeiten mit SPSS sind im Buchhandel zu haben. Auch zahlreiche SPSS-Handbücher zu aufwändigen Verfahren sind erschienen. Da ständig neue SPSS-Versionen erscheinen wird hier keine Empfehlung gegeben.
- Einige Statistikbücher geben auch einen kurze Einführung zu Rechnen in SPSS, so z.B. in: Bortz, J. (1989). Statistik für Sozialwissenschaften. Berlin, Heidelberg, New York, Springer.

Hilfen...

- Viel Hilfe findet sich im Internet, so z.B. auf: <http://www.uni-koeln.de/themen/statistik/software/spss/index.html>
- Besonders hilfreich sind die SPSS-PDF-Dokumente des Universitäts-Rechenzentrum Trier. Suchworte „SPSS Universität Trier“ in google bringt gute Treffer.

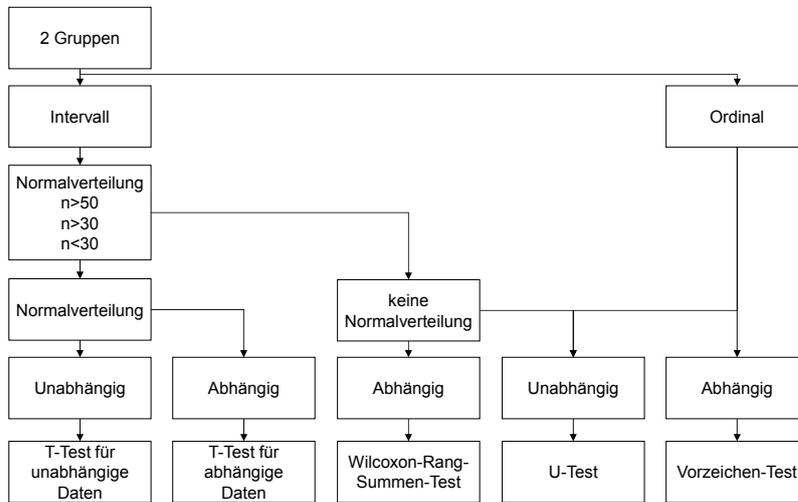
Popup-Hilfe



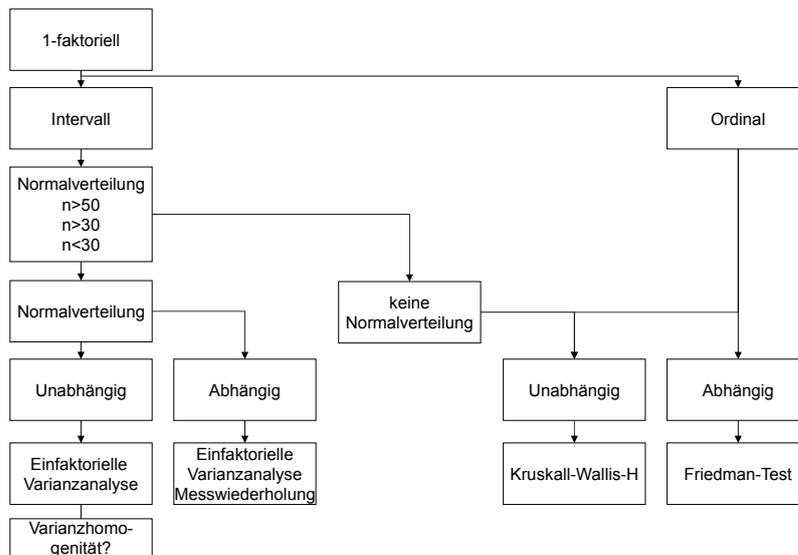
Ein Klick mit der rechten Maustaste auf ein Dialogelement aktiviert die Popup-Hilfe für dieses Dialogelement.

1.3.8 Test-Finder

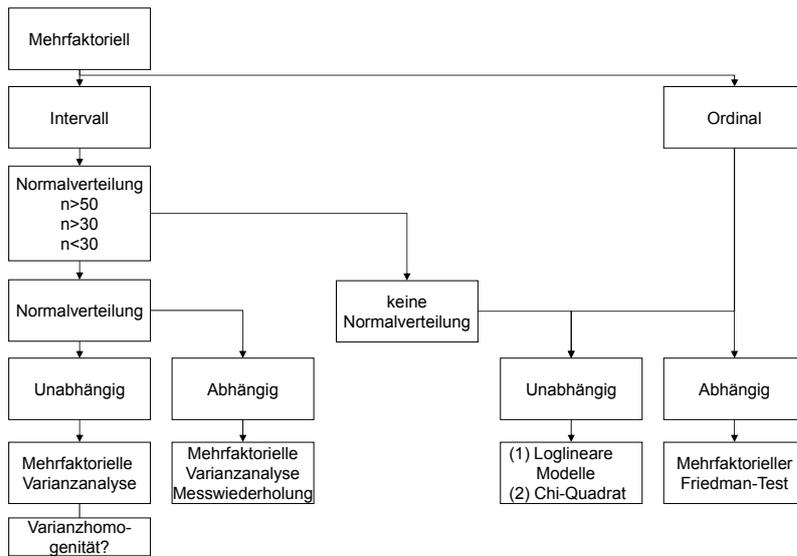
1 Faktor / 2 Gruppen / zentrale Tendenz



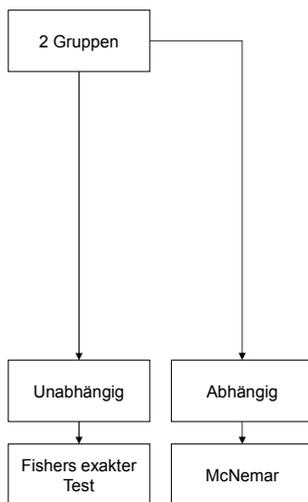
1 Faktor / > 2 Gruppen / zentrale Tendenz



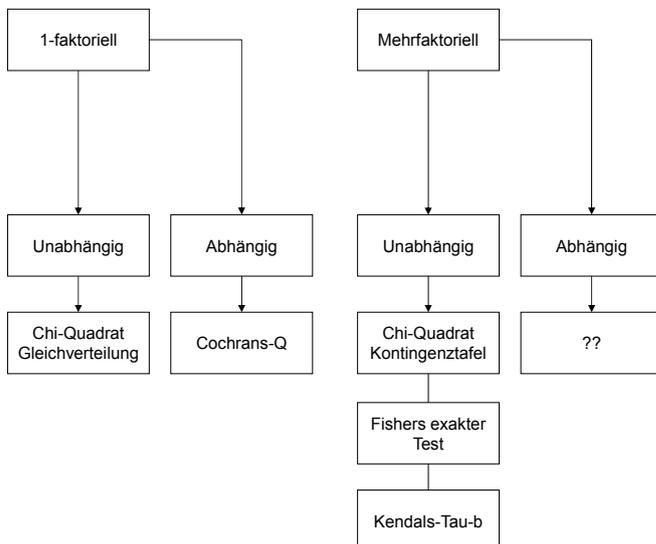
>1 Faktor / > 2 Gruppen / zentrale Tendenz



2 Gruppen / Häufigkeiten



>2 Gruppen / Häufigkeiten



2. Glossar für einige wichtige statistische Begriffe

Alpha-Fehler. Ein Signifikanztest (\nearrow statistische Signifikanz) befindet den Unterschied zwischen zwei Messwerten dann als signifikant, wenn der Unterschied so groß ist, dass es nach den Gesetzen der Wahrscheinlichkeitsrechnung als extrem unwahrscheinlich angesehen werden kann, dass kein Unterschied besteht. Nun ist es jedoch relativ offen, welche Wahrscheinlichkeit als klein genug gelten kann. Es handelt sich daher um eine Übereinkunft, dass gemeinhin bei einer Wahrscheinlichkeit von 5% (und darunter) von Signifikanz gesprochen wird. Nun heißt dies jedoch, dass ein Signifikanztest, der zwei Messwerte nur mit 5% Wahrscheinlichkeit für ähnlich hält, dazu verleitet, die beiden Messwerte eben für unterschiedlich zu halten. Dennoch besteht laut Test aber eine Wahrscheinlichkeit von 5%, dass sie doch ähnlich sind und sich nicht unterscheiden. Wenn man aufgrund des Tests also davon ausgeht, dass sie sich unterscheiden, macht man mit eben jener 5%-tigen Wahrscheinlichkeit einen Fehler. Dieser Fehler wird Alpha-Fehler genannt.

Alpha-Fehler-Adjustierung. In der Regel sind Signifikanztests in der Lage, nur zwei Messwerte miteinander zu vergleichen. Einige Fragestellungen machen daher mehrere Vergleiche zwischen jeweils zwei Messwerten nötig, um die Frage insgesamt beantworten zu können. Beantworten drei Personengruppen einen Fragebogen (Gruppe A, B, C), so kommt man auf insgesamt drei paarweise Vergleiche (A mit B; A mit C und B mit C). Allgemein gilt Anzahl der Vergleiche = [Anzahl der Gruppen mal [Anzahl der Gruppen minus Eins]] geteilt durch 2. So ergeben sich für vier Gruppen bereits: $(4 \times 3)/2 = 6$ Vergleiche. Wenn die Fragestellung relativ offen formuliert ist und generell nach Unterschieden zwischen den Gruppen gefragt wird, so wächst die Wahrscheinlichkeit, einen Unterschied zu finden, je mehr Vergleiche möglich werden. Da man ja bei jedem Paarvergleich einen Alpha-Fehler von 5% begeht, summieren sich die Fehler von Paarvergleich zu Paarvergleich. Bei drei Vergleichen macht man also einen viel höheren Fehler als bei nur einem. Höhere Fehler als 5% sind jedoch nach der oben angesprochenen Vereinbarung nicht signifikant. Um insgesamt nur auf einen Fehler von 5% zu kommen, müssen für jeden Einzelvergleich strengere Alpha-Fehler-Grenzwerte festgelegt werden. Für 3 Vergleiche ergibt sich z.B. ein Wert von 1,7%, bei vier Vergleichen sind es 1,3%, bei 10 Vergleichen 0,5%, usw. Eine Alternative für die Berechnung vieler Signifikanztests, die nur jeweils zwei Messwerte vergleichen können ist die sogenannte Varianzanalyse (\nearrow Varianzanalyse, ANOVA).

Chi-Quadrat-Test. Der Chi-Quadrat-Test ermöglicht den Vergleich von erwarteten Häufigkeiten mit tatsächlich beobachteten Häufigkeiten. Erwartet man aufgrund von Vorerfahrungen oder aus der Literatur zum Beispiel, dass jeder vierte männliche Österreicher Raucher ist, so würde man bei 100 befragten Personen 25 Raucher erwarten. Der Chi-Quadrat-Test vergleicht die erwarteten 25 Raucher dann mit den tatsächlich im Rahmen einer Befragung vorgefundenen Rauchern. Im Rahmen eines Chi-Quadrat-Tests können beliebig viele verschiedene Häufigkeiten miteinander verglichen werden. So ergibt sich beim Chi-Quadrat-Test auf eine Gleichverteilung hin die erwartete Häufigkeit als Mittelwert der beobachteten. Aufgrund geringer Voraussetzungen kann der Chi-Quadrat-Test immer berechnet werden, wenn es um Häufigkeiten geht und eine bestimmte oder mehrere bestimmte Häufigkeiten erwartet werden können. Der Chi-Quadrat-Test ermittelt einen Chi-Quadrat-Wert, für den zusammen mit den sog. Freiheitsgraden (in der Regel Zahl der Messwerte minus eins) die Wahrscheinlichkeit bekannt ist. Die Wahrscheinlichkeit ist das Ergebnis des Tests. Man spricht von einer \nearrow statisti-

schen Signifikanz, wenn diese Wahrscheinlichkeit kleiner als der vorher festgelegte α -Fehler ist.

Fishers exakter Test. Ein besonders *sicherer* Test ist Fishers exakter Test, da er kaum an Voraussetzungen gebunden ist und immer berechnet werden kann, wenn es um den Vergleich zweier Prozentzahlen geht. Eine Berechnung durch einen Computer setzt jedoch meist voraus, dass insgesamt nicht mehr als 1000 Personen befragt wurden, da bei der Berechnung extrem hohe Zahlen als Zwischenergebnisse auftreten. Neben der exakten Variante dieses Tests gibt es für große Stichproben daher auch Näherungsformeln über den α -T-Test, die jedoch mit Vorsicht zu genießen sind. Fishers exakter Test liefert ohne weitere Kennwerte die Wahrscheinlichkeit für die Übereinstimmung der beiden Prozentzahlen. Die Wahrscheinlichkeit ist das Ergebnis des Tests. Man spricht von einer α -statistischen Signifikanz, wenn diese Wahrscheinlichkeit kleiner als der vorher festgelegte α -Fehler ist.

Korrelationen. Eine Korrelation beschreibt den statistischen Zusammenhang zwischen zwei Merkmalen. Beide Merkmale müssen in unterschiedlichen Ausprägungen vorkommen können. Ist das nicht der Fall, so kann keine Korrelation berechnet werden. Wird z.B. die Frage danach gestellt, ob die Zahl der Geburten und die Zahl der Störche einen Zusammenhang (also eine Korrelation) aufweist, so muss sowohl die Zahl der Störche, als auch die Zahl der Geburten variieren können. Es bietet sich hier an, die Zahl der Geburten und die Zahl der Störche pro Monat zu erheben. Dadurch erhält man Zahlenpaare von Geburtenzahl und Storchpopulation für jeden Monat. Es stellt sich nun die Frage, ob sich die Zahl der Störche und die Zahl der Geburten über das Jahr hinweg in die gleiche Richtung entwickelt, also ob mit ansteigender Zahl der Geburten auch die Zahl der Störche wächst und ob mit sinkender Zahl der Geburten auch die Zahl der Störche abnimmt. Ist es so, dass die Zahl der Störche und die Zahl der Geburten sich jeweils in die gleiche Richtung entwickeln, so spricht man von einer positiven Korrelation. Steigt jedoch die Zahl der Geburten, immer wenn die Zahl der Störche abnimmt (und umgekehrt: die Zahl der Geburten sinkt und gleichzeitig nimmt die Zahl der Störche zu), so spricht man von einer negativen Korrelation. Korrelationen können Zahlenwerte zwischen -1 und $+1$ annehmen. Dabei zeigt das Vorzeichen nur an, ob es sich um einen positive oder um eine negative Korrelation handelt. Je näher die Zahlenwerte bei 1 (bzw. -1) liegen, desto „perfekter“ ist der Zusammenhang. Ist eine Korrelation jedoch 0 , dann liegt gar keine Korrelation, also auch kein Zusammenhang vor. Viele Zusammenhänge, die z.B. in der Psychologie beschrieben werden haben relativ kleine Werte um $0,3$ (bzw. $-0,3$), wohingegen z.B. in der Physik nicht selten Korrelationen um $0,9$ (bzw. $-0,9$) gefunden werden können. Ob eine kleine Korrelation nicht eventuell doch auf das Fehlen einer Korrelation (Null-Korrelation) hinweist, kann nur durch einen Signifikanztest (α -Statistische Signifikanz) entschieden werden. Es wird grundsätzlich 2-seitig getestet (α -P-2-seitig). Erst, wenn eine Korrelation sich als signifikant herausstellt, kann sie interpretiert werden. Ist sie nicht signifikant, so kann man nicht davon ausgehen, dass ein Zusammenhang beobachtet wurde. Ist sie jedoch signifikant, so bedeutet das noch nicht, dass der beobachtete Zusammenhang kausal zu interpretieren ist. Es gibt Studien, die zeigen, dass die Zahl der Störche mit der Zahl der Geburten in einigen Gegenden im Verlauf des Jahres tatsächlich korreliert. Das bedeutet jedoch nicht, dass die Störche die Kinder bringen.

Mann-Whitney-U-Test. Besteht der Verdacht, dass die Voraussetzungen für einen α -T-Test verletzt sein könnten, kann am besten der U-Test von Mann und Withney berechnet werden.

Mittelwert, Median, Modalwert. (Maße der zentralen Tendenz) Ergebnisse einer Untersuchung, wie z.B. einer schriftlichen Befragung werden häufig unter Rückgriff auf das sog. arithmetische Mittel, den Mittelwert, berichtet. Der Mittelwert, als Summe aller Antworten durch die Zahl der Antworter, trägt der Tatsache Rechnung, dass in der Regel von verschiedenen Personen auch verschiedene Antworten gegeben werden. Dennoch kann man als zentrale Tendenz aller Antworten den Mittelwert als gute Näherung für einen Großteil der Antworter ansehen. Allerdings ist der Mittelwert nicht das einzig sinnvolle Maß. So ist der Mittelwert empfindlich gegen extreme Antworten, auch dann, wenn diese nur von wenigen Personen gegeben werden.

Der Median weist eine solche Empfindlichkeit nicht auf. Der Median ist die Mitte der Messwerte, die sich dort finden lässt, wo exakt 50% der befragten Personen einen geringeren Messwert aufweisen und exakt 50% einen höheren (bzw. den gleichen). Der Modalwert ist relativ einfach definiert; er ist der Messwert, der insgesamt am häufigsten vorkommt. Bei einigen Fragestellungen ergibt es sich, dass Mittelwert, Median und Modalwert exakt den gleichen Wert aufweisen. Dies ist jedoch nicht immer der Fall. Aus der Anordnung der drei Werte gewinnen Statistiker wichtige Informationen über das Antwortverhalten der befragten Personen. Bei Merkmalen, die durch extreme Antworten verzerrt sein könnten (z.B. Altersangaben, bei denen einige wenige Personen sehr alt sind), sollte immer der Median angegeben werden. Bei einigen anderen Messwerten kann der Mittelwert zudem nicht sinnvoll berechnet werden.

P-2-seitig. Das Ergebnis eines Signifikanztests (\nearrow statistische Signifikanz) ist im wesentlichen die Wahrscheinlichkeit dafür, dass sich zwei Messwerte nicht voneinander unterscheiden. Da Wahrscheinlichkeit auf Englisch *Probability* heißt, wird sie mit dem Buchstaben „p“ abgekürzt. p kann jedoch grundsätzlich auf zwei verschiedene Arten berechnet werden. p kann 1-seitig oder auch 2-seitig bestimmt werden. Welche der beiden Berechnungen im Einzelfall anzugeben ist, entscheidet sich durch die Fragestellung, die mit dem Signifikanztest beantwortet werden soll. Eine zweiseitige Fragestellung prüft, ob zwischen zwei Messwerten ein Unterschied besteht, ohne genauer darauf einzugehen, welche Richtung der Unterschied hat (ob der eine Messwert größer als der andere ist oder ob das Umgekehrte zu erwarten ist, wird nicht berücksichtigt). Eine einseitige Fragestellung prüft nicht nur, ob allgemein ein Unterschied besteht, sondern zudem, ob er in die erwartete Richtung geht. Der 2-seitige Wert wird also bei ungerichteten Signifikanztests angegeben. Er ist immer exakt doppelt so hoch wie der entsprechende 1-seitige Wert. Der 1-seitige Wert hat es damit „leichter“ signifikant zu werden, erfordert aber die genauere Vorhersage.

Standardabweichung, Streuung, Varianz. Die Standardabweichung oder auch Streuung genannt, ist ein Wert für die mittlere Abweichung der Messwerte vom Mittelwert. Die Streuung gibt damit einen Eindruck von der Variationsbreite der Antworten und damit zum Teil auch über die Messgenauigkeit. Bei ideal verteilten Messwerten liegt der \nearrow Mittelwert exakt in der Mitte (zusammen mit dem \nearrow Median und dem \nearrow Modalwert) aller Antworten. Insgesamt 68% aller Antworten befinden sich dann in dem Messwertebereich zwischen dem Mittelwert minus der Streuung und dem Mittelwert plus der Streuung. Beispiel: Ein IQ-Test weist in der Regel einen Mittelwert von 100 und eine Streuung von 10 auf. Damit liegen 68% aller Menschen mit ihrem IQ zwischen einem IQ von 90 und 110. Die Varianz ist nichts anderes als das Quadrat der Streuung.

Statistische Signifikanz. (Statistische Bedeutsamkeit) Jeder im Rahmen einer Messung gewonnene Messwert ist mit einer gewissen Fehlertoleranz behaftet. Die Ergebnisse einer Befragung sind daher nie exakt. Die Genauigkeit einer Messung kann in vielen Fällen mit Hilfe der Wahrscheinlichkeitsrechnung angegeben werden. In diesem Sinne bezeichnet z.B. die σ Streuung eine bestimmte Schwankungsbreite für einen Messwert. Wenn nun zwei Messwerte verglichen werden sollen, z.B. die Messwerte Männer mit denen für Frauen, so muss immer auch mitbedacht werden, dass beide Messwerte ungenau sind. Sagen z.B. 40% der Männer auf eine bestimmte Frage "ja" und antworten auf dieselbe Frage nur 30% der Frauen mit "ja", so ist es nicht leicht zu entscheiden, ob dieser Unterschied durch Messungenauigkeiten zustande kommen konnte oder auf Unterschiede zwischen Männern und Frauen zurückgeführt werden kann. Ein statistischer Signifikanztest beantwortet nun die Frage danach, ob ein Unterschied zwischen zwei Messwerten durch Messungenauigkeiten erklärt werden kann. Erst wenn die Wahrscheinlichkeit für den vorgefundenen Unterschied so klein ist, dass nur noch mit einer Wahrscheinlichkeit von 5% (oder weniger) davon ausgegangen werden kann, dass kein Unterschied zwischen den Messwerten vorliegt, sagt man, dass die Unterschiede statistisch signifikant sind. D.h., dass ein statistischer Signifikanztest niemals behaupten würde, dass ein Unterschied zwischen den Männern und den Frauen besteht. Statistisch signifikant heißt nur, dass es unwahrscheinlich (aber nicht unmöglich) ist, dass kein Unterschied besteht.

Je nach erhobenen Daten müssen verschiedene Verfahren für die Signifikanzprüfung angewandt werden. Wichtige Testverfahren sind z.B.: \rightarrow T-Test, \rightarrow Fishers exakter Test, \rightarrow Chi-Quadrat-Test, \rightarrow Mann-Whitney-U-Test, \rightarrow Varianzanalyse. Das wichtigste Ergebnis eines Testes ist die Wahrscheinlichkeit (\rightarrow P-2-seitig) dafür, dass sich die Messwerte nicht unterscheiden. Diese Wahrscheinlichkeit wird mit einem vorher festgelegten Grenzwert, der Signifikanzgrenze (\rightarrow Alpha-Fehler-Adjustierung) verglichen.

T-Test. Ein besonders gebräuchlicher Signifikanztest für den Vergleich von zwei Mittelwerten ist der T-Test (\rightarrow statistische Signifikanz). Der T-Test besitzt jedoch einige Voraussetzungen, die erfüllt sein müssen, damit er berechnet werden kann. Diese Voraussetzungen sind allerdings nicht immer erfüllt. Zu den Grundvoraussetzungen gehört u.a., dass mit gutem Gewissen ein \rightarrow Mittelwert und die dazu gehörige \rightarrow Streuung berechnet werden können. Die Verteilung der Mittelwerte muss einer Normal- bzw. T-Verteilung folgen, was bei kleinen Stichproben Probleme machen kann. Bei Stichproben mit einer Gruppengröße von mindestens 25 bis 50 Personen pro Untersuchungsgruppe, liegt automatisch eine Normalverteilung der Mittelwerte vor (\rightarrow zentraler Grenzwertsatz), so dass dann kein Problem bei der Anwendung des T-Tests besteht. Der T-Test berechnet einen t-Wert, für den zusammen mit den sog. Freiheitsgraden (in der Regel Zahl der Messwerte minus eins) die Wahrscheinlichkeit bekannt ist. Die Wahrscheinlichkeit ist das Ergebnis des Tests. Man spricht von einer \rightarrow statistischen Signifikanz, wenn diese Wahrscheinlichkeit kleiner als der vorher festgelegte \rightarrow Alpha-Fehler ist.

Validität, prognostische. Wenn Verfahren zur Personalauswahl eingesetzt werden, so verspricht man sich davon Hinweise, die es tatsächlich erlauben, die geeignetsten Kandidatinnen bzw. die geeignetsten Kandidaten aus den Bewerberinnen und Bewerbern heraus zu suchen. Die Verfahren sollen also im weitesten Sinne die „Eignung“ feststellen. Ob ein Verfahren tatsächlich das misst, was es zu messen vorgibt, hier die „Eignung“, wird als Validität des Verfahrens bezeichnet. Zur Feststellung der Validität wird in der Regel eine \rightarrow Korrelation zwischen den Ergebnissen des eingesetzten Verfahrens und passender Außenkriterien (z.B. Leistungsbeurteilung durch einen Vorgesetzten) berechnet. Damit ist die Validität

quantifizierbar mit Werten zwischen Null und Eins, wobei hohe Werte einer hohen Validität entsprechen. Da es bei der Personalauswahl darum geht, die Eignung zu prognostizieren und als passende Außenkriterien Merkmale in Frage kommen, die in der Zukunft liegen, spricht man von einer prognostischen Validität, also von der Fähigkeit des eingesetzten Verfahrens, Vorhersagen über die Verwendbarkeit einer Bewerberin eines Bewerbers zu erstellen. Wie hoch die Validität im Idealfall sein soll, hängt vom Einsatzziel (z.B. von der Anzahl der wahrscheinlich ohnehin geeigneten Bewerberinnen und Bewerbern: sind wahrscheinlich ohnehin alle für die Stelle geeignet, kann die Auswahl einfach gehalten werden) und vom Aufwand (Kosten vs. Nutzen) ab. Eine hohe Validität wird Verfahren mit einem Wert über 0,3 zugesprochen. Hierzu gehört z.B. das Assessment Center, wohingegen Bewerbungsunterlagen, Schulnoten und graphologische Gutachten darunter liegen.

Varianzanalyse. (heißt auch ANOVA) In der Regel sind Signifikanztests in der Lage nur zwei Messwerte miteinander zu vergleichen. Einige Fragestellungen machen daher mehrere Vergleiche zwischen jeweils zwei Messwerten nötig, um die Frage insgesamt beantworten zu können. Beantworten drei Personengruppen einen Fragebogen (Gruppe A, B, C), so kommt man auf insgesamt drei paarweise Vergleiche (A mit B; A mit C und B mit C). Obwohl es hier möglich ist, jede Kombination der Gruppen einzeln zu vergleichen und eine Alpha-Fehler-Adjustierung vorzunehmen (\nearrow Alpha-Fehler-Adjustierung), ist eine Varianzanalyse eleganter und weniger aufwändig zu rechnen. Die Varianzanalyse löst das Problem durch einen Trick: Es werden im wesentlichen zwei Varianzen (\nearrow Streuung) ermittelt und diese mit einem F-Test verglichen. Es werden also auch hier nur zwei Werte durch den Test verglichen. Die eine Varianz ist die innerhalb der Gruppen, die andere ist die zwischen den Gruppen. Sind die Unterschiede (also die Varianz) zwischen den Gruppen größer als die Unterschiede innerhalb der Gruppen, so unterscheiden sich die Gruppen. Allerdings ist dann noch nicht bekannt, welche Gruppen sich voneinander unterscheiden. Um dies heraus zu finden werden anschließend doch wieder paarweise Vergleiche durchgeführt.

Zentraler Grenzwertsatz. Ist eine Untersuchungsstichprobe groß, so ergibt sich unabhängig von der Verteilung der Rohdaten für den Mittelwert eine Normalverteilung. Diesen Zusammenhang kann man sich wie folgt vorstellen: Es wird aus einer größeren Stichprobe eine begrenzte Zufallsauswahl getroffen und für diese Zufallsauswahl ein Mittelwert berechnet. Dies wird mehrfach wiederholt. Jeder berechnete Mittelwert beruht dann nur auf einer Zufallsauswahl und stimmt damit mit dem echten Mittelwert nur mehr oder weniger gut überein. Es zeigt sich, dass die Mittelwerte der Zufallsauswahlen um den echten Mittelwert normalverteilt streuen und zwar unabhängig von der Verteilung der eigentlichen Rohwerte. Testverfahren wie der \nearrow T-Test oder die \nearrow Varianzanalyse benötigen solche normalverteilten Mittelwerte. Diese sind nach dem zentralen Grenzwertsatz für große Stichproben immer gegeben. Was groß ist und was nicht hängt vom jeweiligen Lehrbuch ab. Einige sagen, dass 25 Personen pro Untersuchungs-Gruppe genügen, andere fordern 30 und ganz strenge sogar 50 Personen. Da für kleine Stichproben der zentrale Grenzwertsatz nicht gilt, müssen hier dann auch die Rohwerte strikt normalverteilt sein, damit Testverfahren wie der \nearrow T-Test oder die \nearrow Varianzanalyse angewendet werden können.

3. Darstellung und Abkürzungen

3.1 Allgemein

AM oder *Mittelwert (arithmetisches Mittel; Mean)*

M oder

\bar{x}

SD oder *Standardabweichung (Standard Deviation; Streuung)*

s oder

Std.

df oder *Freiheitsgrade (degree of freedom)*

FG

N *Größe der erfassten Grundgesamtheit.*

n *Größe einer Stichprobe oder Gruppe aus der Grundgesamtheit.*

p *Wahrscheinlichkeit (kann Werte zwischen 0 und 1 annehmen. 0,6 bedeutet also eine Wahrscheinlichkeit von 60%).*

p-2-seitig *Wahrscheinlichkeit dafür, dass etwas nicht signifikant ist (2-seitig getestet).*

p-1-seitig *Wahrscheinlichkeit dafür, dass etwas nicht signifikant ist (1-seitig getestet).*

* *Der Unterschied ist signifikant bei einem Alphafehler von 5% ($p \leq 0,05$)*

** *Der Unterschied ist hoch signifikant bei einem Alphafehler von 1% ($p \leq 0,01$)*

3.2 Korrelationen

r Korrelation.

p-2-seitig Wahrscheinlichkeit dafür, dass eine Korrelation Null ist.

* Die Korrelation ist mit einem Alphafehler (einer Wahrscheinlichkeit) von 5% ($p \leq 0,05$) Null. Die Korrelation ist signifikant.

** Die Korrelation ist mit einem Alphafehler (einer Wahrscheinlichkeit) von 1% ($p \leq 0,01$) Null. Die Korrelation ist signifikant.

Im Text: „...Es besteht also ein hoch signifikanter Zusammenhang zwischen dem Verhalten der beobachteten Frauen und dem Verhalten der beobachteten Männer ($r = 0,52$; p -2-seitig = $0,001$)...“

In Tabellen:

Tabelle 1:
Interkorrelationsmatrix der Skalen

N = 506	Skala 1	Skala 2	Skala 3	Skala 4	Skala 5
Skala 2	0,758 **				
Skala 3	0,924 **	0,723 **			
Skala 4	0,815 **	0,589 **	0,292 *		
Skala 5	0,810 **	0,491 **	0,587 **	0,517 **	
Skala 6	0,849 **	0,599 **	0,062	0,706 **	0,562 **

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

- Skala 1: Neurotizismus
- Skala 2: Emotionale Stabilität
- Skala 3: Extraversion
- Skala 4: Selbstdarstellung
- Skala 5: Führungsmotivation

3.3 T-Test

- t *Prüfgröße für den T-Test.*
- df *Freiheitsgrade (degree of freedom).*
- p-2-seitig *Wahrscheinlichkeit dafür, dass zwei Mittelwerte sich nicht signifikant unterscheiden (2-seitig getestet).*
- p-1-seitig *Wahrscheinlichkeit dafür, dass zwei Mittelwerte sich nicht signifikant unterscheiden (1-seitig getestet).*
- * *Der Unterschied ist signifikant bei einem Alphafehler von 5% ($p \leq 0,05$)*
- ** *Der Unterschied ist hoch signifikant bei einem Alphafehler von 1% ($p \leq 0,01$)*

Im Text: „...Es besteht also ein hoch signifikanter Unterschied zwischen dem Verhalten der beobachteten Frauen und dem Verhalten der beobachteten Männer ($t = 3,52$; $df = 255$; p -2-seitig = $0,003$)...“

In Tabellen:

Tabelle 2:

Ergebnisse der Befragung: Männer vs. Frauen¹

	Männer			Frauen			t	df	p
	AM	SD	n	AM	SD	n			
Skala 1	5,25	1,32	500	6,01	1,12	420	2,57	918	0,004**
Skala 2	4,98	1,04	499	5,98	1,20	420	2,62	917	0,003**

** Die Unterschiede sind auf dem Niveau von 0,01 (2-seitig) signifikant.
 * Die Unterschiede sind auf dem Niveau von 0,05 (2-seitig) signifikant.

Skala 1: Neurotizismus
 Skala 2: Emotionale Stabilität

Tabelle 2 (alternative):

Ergebnisse der Befragung: Männer vs. Frauen²

	Männer			Frauen			t	df	p
	AM (SD)	n		AM	SD	n			
Skala 1	5,25 (1,32)	500		6,01	1,12	420	2,57	918	0,004**
Skala 2	4,98 (1,04)	499		5,98	1,20	420	2,62	917	0,003**

** Die Unterschiede sind auf dem Niveau von 0,01 (2-seitig) signifikant.
 * Die Unterschiede sind auf dem Niveau von 0,05 (2-seitig) signifikant.

Skala 1: Neurotizismus
 Skala 2: Emotionale Stabilität

¹ Alle Zahlenwerte (inklusive der Ergebnisse des Tests) sind frei erfunden
² Alle Zahlenwerte (inklusive der Ergebnisse des Tests) sind frei erfunden

3.4 Varianzanalyse

- F *Prüfgröße für den F-Test, der in der Varianzanalyse benutzt wird*
- df *Freiheitsgrade (degree of freedom).*
- p *Wahrscheinlichkeit dafür, dass sich die Gruppen nicht signifikant unterscheiden*
- * *Der Unterschied zwischen den Gruppen ist signifikant bei einem Alphafehler von 5% ($p \leq 0,05$)*
- ** *Der Unterschied zwischen den Gruppen ist hoch signifikant bei einem Alphafehler von 1% ($p \leq 0,01$)*

Im Text: „...Es besteht also ein hoch signifikanter Unterschied zwischen den beobachteten Gruppen ($F = 3,52$; $p = 0,003$)....“

In Tabellen:

Tabelle 3:
Ergebnisse der Beobachtung³

Skala 1	AM	SD	n	F	p
Männer	5,25	1,32	500	3,25	0,004**
Frauen	4,98	1,04	499		
Kinder	3,12	0,98	356		
Skala 2					
Männer	4,25	1,32	500	2,56	0,003**
Frauen	4,98	1,04	499		
Kinder	3,12	0,98	356		

** Die Unterschiede sind auf dem Niveau von 0,01 (2-seitig) signifikant.
 * Die Unterschiede sind auf dem Niveau von 0,05 (2-seitig) signifikant.
 Skala 1: Neurotizismus
 Skala 2: Emotionale Stabilität

Tabelle 4:
Signifikante Gruppenunterschiede nach Scheffé-Test⁴

Skala 1	Frauen	Kinder
Männer	*	**
Frauen		**
Skala 2		
Männer	*	
Frauen		**

** Die Unterschiede sind auf dem Niveau von 0,01 (2-seitig) signifikant.
 * Die Unterschiede sind auf dem Niveau von 0,05 (2-seitig) signifikant.
 Skala 1: Neurotizismus
 Skala 2: Emotionale Stabilität

Bei einer Varianzanalyse bieten sich zusätzlich auch Grafiken an.

³ Alle Zahlenwerte (inklusive der Ergebnisse des Tests) sind frei erfunden
⁴ Alle Zahlenwerte (inklusive der Ergebnisse des Tests) sind frei erfunden

4. Literatur

- Bortz, J. (1999) *Statistik für Sozialwissenschaften (5. Auflage)*. Berlin, Heidelberg, New York: Springer
- Bortz, J. & Döring, N. (2002) *Forschungsmethoden und Evaluation*. Berlin, Heidelberg: Springer
- Bühner, M. (2004) *Einführung in die Test- und Fragebogenkonstruktion*. München u. a.: Pearson Studium
- Churchill, G. A. & Iacobucci, D. (2002) *Marketing Research: Methodological Foundations*. Orlando, FL: Harcourt College Publishers
- Friedrichs, J. (1990) *Methoden empirischer Sozialforschung*. Opladen: Westdeutscher Verlag
- Langewitz, W., Keller, A. & Denz, M. (1995) Patientenzufriedenheits-Fragebogen (PZF): Ein taugliches Mittel zur Qualitätskontrolle der Arzt-Patient-Beziehung? *Zeitschrift für Psychotherapie Psychosomatik und medizinische Psychologie*, 45, 351-357
- Lienert, G., A. & Raatz, U. (1994) *Testaufbau und Testanalyse*. Weinheim: Beltz
- Mayring, P. (2003) *Qualitative Inhaltsanalyse*. Weinheim und Basel: Beltz Verlag
- Weick, K. E. (1985) *Der Prozeß des Organisierens*. Frankfurt am Main: Suhrkamp