Achim Bühl

SPSS 20

Einführung in die moderne Datenanalyse

13., aktualisierte Auflage

ALWAYS LEARNING

PEARSON

Exakte Testmethoden

Der englische Statistiker Sir Ronald A. Fisher machte einst ein wissenschaftliches Experiment echt britischer Art. Eine Bekannte hatte behauptet, sie könne es einer Tasse Tee ansehen, ob zuerst der Tee oder zuerst die Milch eingegossen worden sei.

Fisher wollte dies überprüfen und setzte der Teetrinkerin nach einer Zufallsreihenfolge acht Tassen Tee mit Milch vor, von denen vier zuerst mit Tee und vier zuerst mit Milch gefüllt worden waren. Die Bekannte landete jeweils drei Treffer und lag jeweils einmal daneben.

- Laden Sie die Datei tee.sav, in der diese Ergebnisse gespeichert sind, und erstellen Sie mit Hilfe der Menüwahl

  Analysieren
  Deskriptive Statistiken
  Kreuztabellen...

  eine Kreuztabelle zwischen der Zeilenvariable gegossen und der Spaltenvariable geraten.

gegossen * geraten Kreuztabelle

Anzahl

		geraten		Gesamt
		Milch	Tee	
gegossen	Milch	3	1	4
	Tee	1	3	4
Gesamt		4	4	8

- Um zu überprüfen, ob die Erhöhung der Häufigkeit bei den Richtigwahlen signifikant ist, wünschen Sie sich in der Dialogbox *Kreuztabellen: Statistiken* die Durchführung des Chi-Quadrat-Tests. Dieser liefert nach der Pearson-Formel einen nicht signifikanten p-Wert von 0,157.

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)
Chi-Quadrat nach Pearson	2,000[a]	1	,157		
Kontinuitätskorrektur[b]	,500	1	,480		
Likelihood-Quotient	2,093	1	,148		
Exakter Test nach Fisher				,486	,243
Zusammenhang linear-mit-linear	1,750	1	,186		
Anzahl der gültigen Fälle	8				

a. 4 Zellen (100,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 2,00.
b. Wird nur für eine 2x2-Tabelle berechnet

Da die erwarteten Häufigkeiten sämtlich kleiner als 5 sind (nämlich 2), sind die Voraussetzungen zur Anwendung des Chi-Quadrat-Testes nicht gegeben. Fisher stellte also Überlegungen an, wie p exakt zu ermitteln ist.

Bezeichnet man die Häufigkeiten der Vierfeldertafel der Reihe nach mit a, b, c und d, so ist die exakte Wahrscheinlichkeit dafür, dass sich bei den gegebenen Randsummen die Häufigkeiten wie gegeben verteilen

$$p0 = \frac{\binom{a+c}{a} \cdot \binom{b+d}{b}}{\binom{a+b+c+d}{a+b}}$$

Mit Hilfe der Definition der Binomialkoeffizienten

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}$$

erhält man mit den gegebenen Werten

$$p0 = \frac{\binom{4}{3} \cdot \binom{4}{1}}{\binom{8}{4}} = \frac{4 \cdot 4}{70} = 0{,}22857$$

Bei den gegebenen Randsummen gibt es eine Verteilung, die noch unwahrscheinlicher ist, nämlich diejenige mit a = 4, b = 0, c = 0 und d = 4. Berechnet man hierzu in entsprechender Weise ebenfalls die Wahrscheinlichkeit und nennt diese p1, so erhält man p1 = 0,01429.

Die Gesamtwahrscheinlichkeit für den einseitigen Test ergibt sich als Summe von p0 und p1, also p = 0,22857 + 0,01429 = 0,24286. Dieses Ergebnis wird unter der Bezeichnung »Exakter Test nach Fisher« ausgegeben und zwar als einseitige Signifikanz.

Der Unterschied der zweiseitigen Anwendung zum Chi-Quadrat-Test ist also erheblich. Wegen des deutlich nicht signifikanten Ausfalls des Tests können also die erzielten Treffer bei der Beurteilung der Tassen auch zufällig zustande gekommen sein.

So kann man Fisher als den Vater der exakten Tests ansehen. Der nach ihm benannte Test ist Bestandteil des Basis-Moduls und kann vom Zusatzmodul Exakte Tests nicht weiter verbessert werden.

Das Modul Exakte Tests stellt exakte Tests zur Verfügung im Rahmen der nichtparametrischen Tests und für Kreuztabellen-Statistiken. Exakte Tests sind empfehlenswert bei kleinen Fallzahlen, bei einem großen Anteil gebundener Werte und bei stark unterschiedlichen Fallzahlen in den einzelnen Stichproben.

10.1 Exakte p-Werte

Exakte Werte für die Irrtumswahrscheinlichkeit p sind natürlich bei allen statistischen Tests wünschenswert. In der Praxis ist es aber nicht immer möglich, solche exakten Werte zu berechnen, so dass man sich mit asymptotischen Werten begnügen muss.

Beim eingangs geschilderten Beispiel des exakten Tests nach Fisher für Vierfeldertafeln stellt die Berechnung des exakten p-Werts kein besonderes Problem dar, da eine entsprechende Formel existiert. Leider aber ist die Existenz einer solchen Formel nicht immer gegeben.

Um aufzuzeigen, wie in einem solchen Fall vorgegangen wird, wollen wir den Wilcoxon-Test betrachten, den nichtparametrischen Test zum Vergleich zweier verbundener Stichproben. Als Beispiel soll eine medizinische Untersuchung dienen, bei der zwanzig Probanden mit erhöhten Cholesterinwerten eine entsprechende medikamentöse Behandlung erfuhren, die den Cholesterinspiegel senken sollte.

- Laden Sie die Datei chol.sav.

Die Datei enthält vier Variablen: cholvor und cholnach für die Cholesterinwerte vor und nach der Behandlung, d für die Differenzen beider Werte und rang für die Rangplätze der absoluten Differenzen, also für den Betrag der Differenzen ohne Berücksichtigung des Vorzeichens. Eine Auflistung der Variablenwerte ist im Folgenden wiedergegeben.

CHOLVOR	CHOLNACH	d	RANG
350	320	30,00	13,000
325	320	5,00	2,000
290	271	19,00	7,000
334	340	–6,00	3,000
276	301	–25,00	11,000
370	312	58,00	15,000
400	367	33,00	14,000
289	300	–11,00	4,000
340	358	–18,00	6,000
288	271	17,00	5,000
345	369	–24,00	10,000
360	290	70,00	18,000
350	272	78,00	20,000
320	324	–4,00	1,000
412	345	67,00	17,000
299	270	29,00	12,000
320	340	–20,00	8,000

CHOLVOR	CHOLNACH	d	RANG
339	316	23,00	9,000
356	290	66,00	16,000
320	246	74,00	19,000

Beim Wilcoxon-Test werden nun die Summe der Rangplätze bei positiven Differenzen (T1) und die Summe der Rangplätze bei negativen Differenzen (T2) gebildet. Der kleinere der beiden Werte T1 und T2 ist die Prüfgröße T. Im gegebenen Beispiel ergibt sich T1 = 167, T2 = 43 und damit T = 43.

- Wir wollen die Berechnung des Wilcoxon-Tests mit SPSS nachvollziehen. Wählen Sie daher aus dem Menü

  Analysieren
  Nichtparametrische Tests
  Alte Dialogfelder
  Zwei verbundene Stichproben…

Voreingestellt ist der Wilcoxon-Test.

- Verschieben Sie die beiden Variablen cholvor und cholnach in das Test-Paar-Feld, und starten Sie die Berechnung mit *OK*.

Die folgenden Ergebnisse werden im Viewer angezeigt.

Ränge

		N	Mittlerer Rang	Rangsumme
cholvor - cholnach	Negative Ränge	7[a]	6,14	43,00
	Positive Ränge	13[b]	12,85	167,00
	Bindungen	0[c]		
	Gesamt	20		

a. cholvor < cholnach
b. cholvor > cholnach
c. cholvor = cholnach

Statistik für Test[a]

	cholvor - cholnach
Z	-2,315[b]
Asymptotische Signifikanz (2-seitig)	,021

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

Die Verteilung der Prüfgröße T geht für größere Stichprobenumfänge in eine Normalverteilung über, was zur Bildung eines z-Wertes führt, dem dann der p-Wert zugeordnet werden kann. Zumindest für kleinere Stichprobenumfänge ist dieser Wert also als asymptotisch zu betrachten.

Um den p-Wert exakt zu berechnen, kann man wie folgt vorgehen. Im gegebenen Beispiel, das aus Gründen der Übersichtlichkeit so gewählt wurde, dass keine geteilten Rangplätze

(Bindungen) auftreten, gibt es die Rangplätze 1 bis 20; diesen zugeordnet ist entweder ein positives oder ein negatives Vorzeichen der entsprechenden Differenz. Alle theoretisch möglichen T-Werte erhält man, wenn man alle denkbaren Kombinationen der Vorzeichenverteilung auf die zwanzig Rangplätze betrachtet, jeweils dazu den T-Wert bestimmt und auszählt, wie oft diese theoretisch denkbaren T-Werte den tatsächlich errechneten T-Wert (43) unterschreiten. Bezeichnet man diese Anzahl mit n, so ist der dem errechneten T-Wert zuzuordnende p-Wert offensichtlich

$$p = \frac{n}{\text{Anzahl der möglichen Kombinationen}}$$

Die Anzahl der möglichen Vorzeichen-Kombinationen ist bei zwanzig Wertepaaren 2^{20} = 1048576. Um auf diese Weise einen exakten p-Wert zu berechnen, sind also sehr viele Rechenoperationen auszuführen; bei dreißig Wertepaaren würde ihre Anzahl auf über eine Milliarde steigen.

Da bis zu einer Fallzahl von 30 die Berechnung der exakten p-Werte bei SPSS sehr schnell vonstattengeht, werden offensichtlich intelligente Algorithmen eingesetzt, um das Problem zu lösen. Bei großen Fallzahlen (bei denen aber in der Regel auf die Berechnung exakter p-Werte verzichtet wird) kann die Rechenzeit dennoch enorm ansteigen. In diesem Fall kann man auf die so genannte Monte-Carlo-Methode zurückgreifen.

10.2 Monte-Carlo-Methode

Die Monte-Carlo-Methode besteht in der Nachbildung des Eintretens von Zufallsereignissen, indem der Ereigniseintritt aus vorgegebenen Wahrscheinlichkeitsverteilungen durch die Verwendung von Zufallszahlen bestimmt wird. Dabei werden die Zufallszahlen bei Verwendung eines Computers durch einen Zufallszahlengenerator erzeugt.

Nehmen Sie an, dass Sie Ihr Taschengeld mit Hilfe des Roulettespiels aufbessern möchten. Sie wissen, dass es beim Roulette 37 Zahlen gibt, nämlich 18 rote, 18 schwarze und die Null. Bevor Sie selbst zu spielen beginnen, möchten Sie erst einmal eintausend Spiele beobachten, um sich mit den Eigenheiten des Spiels vertraut zu machen.

Sie könnten nach Monte Carlo fahren und im dortigen Spielcasino über einen längeren Zeitraum hinweg eintausend aufeinanderfolgende Spiele beobachten und deren Ergebnis notieren; Sie können dies aber auch zu Hause in Sekundenschnelle an Ihrem Computer simulieren.

Jede Programmiersprache stellt einen Zufallszahlengenerator zur Verfügung, der gleichverteilte Zufallszahlen im Intervall zwischen 0 und 1 liefert. Ganz zufällig sind diese Zufallszahlen allerdings nicht, denn sie entstehen als Ergebnis einer Rechenprozedur, so dass sie sich periodisch wiederholen; die Periodenlänge ist allerdings außerordentlich groß und für praktische Anwendungen in den allermeisten Fällen ausreichend. Um gegebenenfalls reproduzierbare Ergebnisse zu erhalten, kann die Zufallszahlenfolge durch Angabe eines immer gleichen Startwerts initialisiert werden.

In der Programmiersprache BASIC stehen die parameterlose Funktion RND zur Erzeugung der Zufallszahlen und die Prozedur RANDOMIZE zur Initialisierung zur Verfügung. Das folgende BASIC-Programm erzeugt über 1000 Spiele hinweg eine Zufallsfolge der Ergebnisse rouge, noir und zero.

```
10 REM Simulation des Roulettespiels mit der Monte-Carlo-Methode
20 RANDOMIZE 10
30 FOR I=1 TO 1000
40 X=RND
50 IF X <= 1/37 THEN 90
60 IF X <= 19/37 THEN 110
70 PRINT "noir, "
80 GOTO 120
90 PRINT "zero, "
100 GOTO 120
110 PRINT "rouge, "
120 NEXT I
130 END
```

In das Programm ist eingeflossen, dass die Wahrscheinlichkeit für das Eintreffen der Null gleich 1/37 und für das Eintreffen von Rot bzw. Schwarz jeweils gleich 18/37 ist. Diese Vorgabe von Wahrscheinlichkeiten und ihre entsprechende Einbringung in das Programm ist typisch für die Monte-Carlo-Technik. Die ersten zwanzig Ergebnisse seien im Folgenden ausgegeben.

```
rouge, noir, rouge, rouge, rouge, rouge, noir, noir, noir, rouge, rouge, noir, rouge,
rouge, noir, zero, noir, rouge, rouge, noir
```

Die Monte-Carlo-Technik kann eingesetzt werden, wenn die Berechnung der exakten p-Werte bei großen Fallzahlen zu lange dauert. Bezogen auf das geschilderte Beispiel des Willcoxon-Tests bedeutet das, dass nicht alle theoretisch möglichen Vorzeichenkombinationen analysiert werden, sondern nur eine bestimmte vorgegebene Anzahl (z. B. 10000). Diese werden dann mit Hilfe der Monte-Carlo-Methode mit einem Computerprogramm erzeugt. So kann das folgende BASIC-Programm dazu benutzt werden, um zu einer vorgegebenen Fallzahl und einer vorgegebenen Testgröße mit einer vorgegebenen Zahl von Simulationsschritten dem exakten p-Wert im Wilcoxon-Test nahezukommen.

```
10 REM Monte-Carlo-Methode (Abschätzung des exakten p beim Wilcoxon-Test)
20 INPUT "Anzahl der Wertepaare";N
30 INPUT "Testgröße";TKRIT
40 INPUT "Anzahl der Wiederholungen";K
50 NT=0
60 REM Durchführung von k Eperimenten
70 FOR J=1 TO K
80 T1=0
90 T2=0
100 REM Ermittlung der Rangsummen T1 und T2
```

```
110 FOR I=1 TO N
120 REM Ziehung einer gleichverteilten Zufallszahl x zwischen 0 und 1
130 X=RND
140 IF X <= .5 THEN 190
150 REM negative Differenzen f_r x > 0.5
160 T2=T2+I
170 GOTO 200
180 REM positive Differenzen f_r x <= 0.5
190 T1=T1+I
200 NEXT I
210 REM Testgröße T als Minimum von T1 und T2
220 IF T1 < T2 THEN 250
230 T=T2
240 GOTO 270
250 T=T1
260 REM Auszählung, wie oft T kleiner ist als die gegebene Testgröße TKRIT
270 IF T >= TKRIT THEN 290
280 NT=NT+1
290 NEXT J
300 REM p als Quotient von Anzahl der Werte T < TKRIT und der Zahl
310 REM der Experimente
320 P=NT/J
330 PRINT USING "#.####";P
340 END
```

Die Monte-Carlo-Technik schlägt sich entscheidend in Zeile 140 nieder, wo für das Auftreten positiver und negativer Vorzeichen jeweils die Wahrscheinlichkeit 0,5 vorgegeben wurde.

Im Falle von 20 Wertepaaren und einer Testgröße T = 43 ergeben sich nach diesem Programm in Abhängigkeit von der Anzahl der Wiederholungen die folgenden Ergebnisse.

Wiederholungen	p-Wert
1000	0,0160
2000	0,0195
3000	0,0200
4000	0,0193
5000	0,0198
6000	0,0193
7000	0,0197
8000	0,0206
9000	0,0202
10000	0,0198

Die Monte-Carlo-Methode zur Bestimmung des exakten p-Werts ist auch in SPSS verfügbar. Sie wird entweder dann benutzt, wenn bei der Anwendung der exakten Methode ein vorgegebenes Zeitlimit überschritten wird, oder wenn Sie diese Methode ausdrücklich gewünscht haben. Wie im nächsten Kapitel gezeigt wird, berechnet SPSS mit der exakten Methode für p den Wert 0,019 und mit der Monte-Carlo-Methode bei 10000 Schritten den Wert 0,020.

10.3 Integration in das SPSS-Basis-Modul

Das Modul Exakte Tests tritt an zwei Stellen in Erscheinung. Zum einen ist es die Dialogbox *Kreuztabellen*, die nach der Menüwahl

Analysieren
Deskriptive Statistiken
Kreuztabellen...

geöffnet wird. Diese Dialogbox ist im folgenden Bild dargestellt, wobei wieder die Datei tee.sav geladen ist.

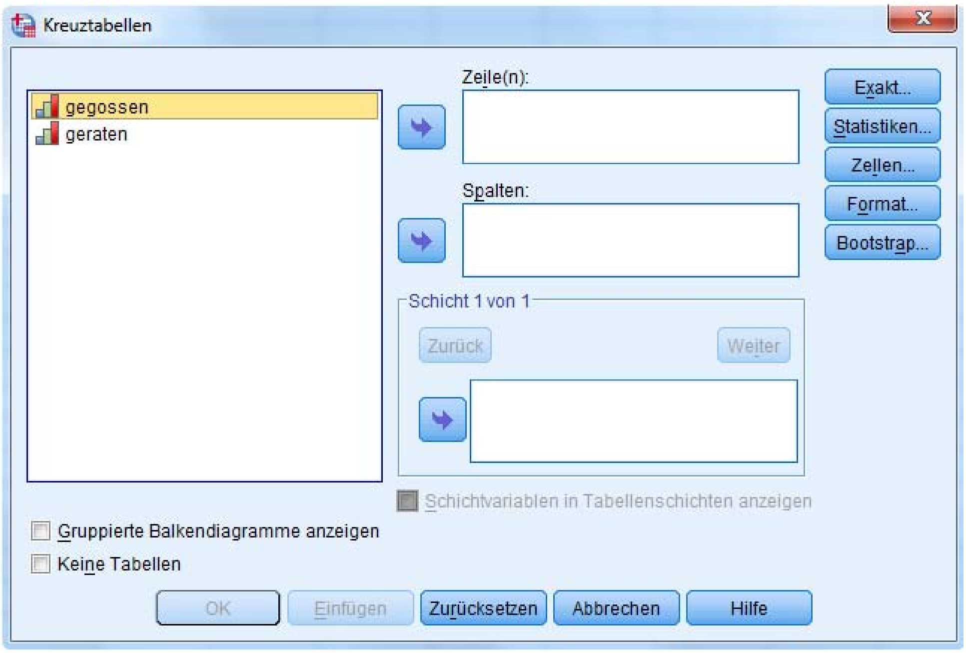

Bild 10.1: Dialogbox Kreuztabellen mit exakten Tests

Zum anderen können exakte Tests nach der Menüwahl

Analysieren
Nichtparametrische Tests

gerechnet werden.

- Laden Sie als Beispiel die Datei chol.sav und wählen Sie aus dem Menü

  Analysieren
    Nichtparametrische Tests
      Alte Dialogfelder
        Zwei verbundene Stichproben...

Es öffnet sich die Dialogbox *Tests bei zwei verbundenen Stichproben*, in welcher der Schalter Exakt... eingebaut ist.

Bild 10.2: Dialogbox Tests bei zwei verbundenen Stichproben

- Bringen Sie die beiden Variablen cholvor und cholnach in das Test-Paar-Feld, belassen Sie es beim voreingestellten Wilcoxon-Test, und betätigen Sie den Schalter *Exakt...*

Es öffnet sich die Dialogbox *Exakte Tests*.

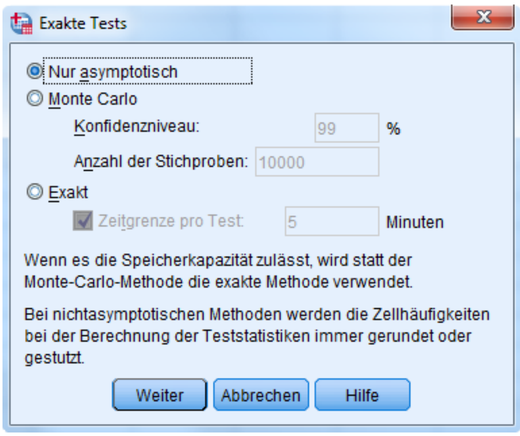

Bild 10.3: Dialogbox Exakte Tests

Sie haben die Wahl zwischen drei Möglichkeiten der Behandlung der exakten Tests.

- Nur asymptotisch

  Diese Option ist voreingestellt und liefert die herkömmlichen asymptotischen Resultate, die Sie auch ohne die Installation des Moduls *Exakte Tests* erhalten.

- Monte Carlo

  Diese Option können Sie wählen, wenn die Berechnung der exakten Werte zu lange Zeit in Anspruch nimmt. Stellt allerdings der Computer fest, dass die exakte Methode ebenfalls schnell zum Ziel führt, wird diese benutzt.

  Für den ermittelten p-Wert wird ein Konfidenzbereich ausgegeben, wobei Sie als Konfidenzniveau eine Zahl zwischen 0.01 und 99.9 eingeben können. Voreingestellt ist 99. Für die Anzahl der simulierten Stichproben ist 10000 voreingestellt. Die maximal mögliche Zahl ist 1 000 000 000. Die erzielte höhere Genauigkeit bei größeren Stichprobenzahlen geht natürlich zulasten der Rechendauer.

- Exakt

  Bei dieser Option werden die p-Werte exakt berechnet. Das voreingestellte Zeitlimit von 5 Minuten können Sie in eine andere Minutenzahl abändern.

- Aktivieren Sie zunächst die exakte Methode, und verlassen Sie die Dialogboxen über *Weiter* bzw. *OK*.

Im Viewer werden die folgenden Ergebnisse angezeigt.

Ränge

		N	Mittlerer Rang	Rangsumme
cholnach - cholvor	Negative Ränge	13[a]	12,85	167,00
	Positive Ränge	7[b]	6,14	43,00
	Bindungen	0[c]		
	Gesamt	20		

a. cholnach < cholvor
b. cholnach > cholvor
c. cholnach = cholvor

Statistik für Test[b]

	cholnach - cholvor
Z	-2,315[a]
Asymptotische Signifikanz (2-seitig)	,021
Exakte Signifikanz (2-seitig)	,019
Exakte Signifikanz (1-seitig)	,010
Punkt-Wahrscheinlichkeit	,001

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

Es werden die exakten p-Werte für den ein- und zweiseitigen Test ausgegeben, ferner nach wie vor der asymptotische Wert.

- Wiederholen Sie die Rechnung, indem Sie in der Dialogbox *Exakte Tests* die Monte-Carlo-Methode aktivieren.

Es werden die folgenden Ergebnisse angezeigt.

Ränge

		N	Mittlerer Rang	Rangsumme
cholnach - cholvor	Negative Ränge	13[a]	12,85	167,00
	Positive Ränge	7[b]	6,14	43,00
	Bindungen	0[c]		
	Gesamt	20		

a. cholnach < cholvor

b. cholnach > cholvor

c. cholnach = cholvor

Statistik für Test[a,c]

			cholnach - cholvor
Z			-2,315[b]
Asymptotische Signifikanz (2-seitig)			,021
Monte-Carlo-Signifikanz (2-seitig)	Signifikanz		,018
	99%-Konfidenzintervall	Untergrenze	,014
		Obergrenze	,021
Monte-Carlo-Signifikanz (1-seitig)	Signifikanz		,008
	99%-Konfidenzintervall	Untergrenze	,005
		Obergrenze	,010

a. Wilcoxon-Test

b. Basiert auf positiven Rängen.

c. Basiert auf 10000 Stichprobentabellen mit einem Startwert von 299883525.

Benutzen Sie die Monte-Carlo-Methode, so können Sie wählen, ob Sie mit immer gleichen Zufallszahlenfolgen (initiiert durch denselben Startwert) oder jeweils anderen rechnen wollen. Im ersten Fall haben Sie den Vorteil, dass Sie reproduzierbare Ergebnisse erhalten.

- Sie können dies steuern, indem Sie über die Menüwahl

  Transformieren
  Zufallszahlengeneratoren...

  die Dialogbox *Zufallszahlengenerator* öffnen.

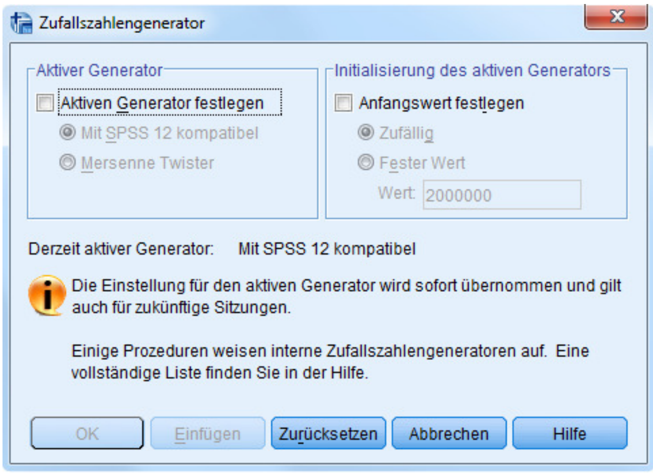

Bild 10.4: Dialogbox Zufallszahlengenerator