Testverfahren aus Statistik A, B und C

© Dorthe Lübbert, Dorthe.Luebbert@ruhr-uni-bochum.de

Meß-niveau	Kriterien	Hypothese	Testverfahren	Beispiel
nomi-nal	1 Stichprobe • dichotome Merkmalsausprägung	Die GG und die Stichprobe sind gleich verteilt: $F_0(x)=F(x)$	**Binominaltest**	Ein Hersteller von Geräten behauptet, daß höchstens 5 Prozent defekte Stücke produziert werden. Per Zufall werden 20 Stücke entnommen. 3 Stücke sind defekt. **Frage**: Kann man dem Hersteller bei 10% Signifikanzniveau trauen?
	1 Stichprobe • polytome Merkmalsausprägung	Die GG und die Stichprobe sind gleich verteilt: $F_0(x)=F(x)$	**χ^{2}-Anpassungstest**	200 Studenten wurden nach dem ihnen am sympathischsten Fach befragt. **Hypothese:** Geprüft werden soll, ob die Studenten alle fünf Fächer gleich sympathisch finden.
	2 abhängige Stichproben • dichotome Merkmale	Unterscheidet sich der Anteilswert einer Gruppe von Wechslern signifikant von 0,5?	**McNemar-Test** → Anzahl der Wechsler muß bekannt sein	Vor und nach einer Parteiveranstaltung werden die Besucher nach ihrem Wahlverhalten in Bezug auf diese Partei befragt.
	2 unabhängige Stichproben • dichotome Merkmale	Unterscheiden sich die Anteilswerte der Stichproben zufällig voneinander oder nicht?	**2x2 Kontigenztafel-Fisher-Test**	Während einer Industrieausstellung wurden 5 Vertreter inländischer Unternehmen und 6 Vertreter ausländischer Unternehmen über ihren Eindruck vom Geschäftsklima befragt: Geprüft werden soll: Sind die Unterschiede in den Antworten zufällig (H_0) oder nicht (H_a)?
	2 unabhängige Stichproben • polytome Merkmale	Das Merkmal A hat in den *r* GG jeweils dieselbe Verteilung, die Stichproben kommen aus einer GG.	χ^{2}-Tests	
	mehr als 2 unabhängige Stichproben		(χ^{2}-Homogenitätstest, χ^{2}-Anpassungstest)	
ordi-nal	1 Stichprobe 1 ordinal skalierte Beobachtungsvariable	H_0: Der Median der GG hat einen bestimmten Wert c	**a) Vorzeichentest (Mediantest)** → Informationen bzgl. Größe von Differenzen werden nicht berücksichtigt → bei steigenden n sinkt Testeffizienz → kann bei großem n als Schnelltest benutzt werden **b) Wilcoxon-Vorzeichen-Rangtest** → nutzt Info über Differenzen (Ränge drücken rangmäßigen Abstand aus) → ab n>25 kann W-V-Test durch eine Normalverteiltung approximiert werden → zählt zu den verteilungsfreien Tests mit der höchsten Testgüte	**a) Mediantest** 7 von 8 Werten einer Stichprobe liegen über dem Hypothesenwert der Grundgesamtheit, d.h. v (Anzahl der positiven Vorzeichen)=7. Frage: Ist der Unterschied zwischen vermutetem Grundgesamtheitsmedian und Ergebnis zufällig? **b) W-V-Rangtest** Ein Lehrer hatte in der Vergangenheit die Durchschnittsnote 3 vergeben. Seiner derzeitigen Schulklasse gab er Zensuren in der folgenden Häufigkeit: Geprüft werden soll, ob sich die Schulklasse von anderen Schulklassen im Notendurchschnitt unterscheidet. (5% Signifikanzniveau, beidseitig): H_0: $\mu_{0,5}=3$
		H_0: $F(x)=F_0(x)$ Stammt eine Stichprobe aus einer in einer bestimmten Weise verteilten Grundgesamtheit?	**a) Kolmogoroff/Smirnow-Ein-Stichproben-Anpassungstest** → einfacher als χ^2-Test, gleich effizient → Voraussetzung: H_0 ist voll spezifiziert **b) χ^2-Test** → anwenden, wenn Parameter der GG nicht expliziet vorgegeben sind	**a)** Die Altersverteilung von 50 Studentinnen zwischen 21 und 30 Jahren wurde gemessen. Die Hypothese lautet, daß das Alter in dem gemessenen Bereich von 21 bis 30 Jahren gleichverteilt sein soll.
	2 unabhängige Stichproben	Stammen die Mediane der Stichprobe aus zwei GG mit gleichem Median, d.h. sind die Verteilungen bis auf die Lageparameter gleich?	**a) Mann/Whitney-U-Test** → 95%-Effizienz **b) t-Test** → GG muß nicht normalverteilt sein **c) Vorzeichentest** als Schnelltest → Effizienz fällt bei steigendem n auf 63% gegen 95% für n1 +n2<6	**a)M-W-U-Test:** Zwei Leichtathletikgruppen mit fünf bzw. sechs Leuten ($n_1=5$, $n_2=6$) machen einen Fitnesstest, nachdem sie ein unterschiedliches Wintertraining durchge-führt haben. Unterscheiden sich die Gruppenmittel nur zufällig voneinander?
		Stammen zwei Stichproben aus einer GG?	**a) Komogoroff/Smirnow-Zwei-Stichproben-Anpassungstest** → Effizienz: 96 % **b) U-Test** → ist bei großen n effizienter **c)t-Test**	
	2 abhängige Stichproben	Stammen die Mediane zweier verbundener Stichproben aus einer GG?	**a) Wilcoxon-Vorzeichen-Rangtest** → Effizienz 96% **b) t-Test** **c) Vorzeichentest** als Schnelltest → geringere Effizienz	
	1 Stichprobe, 2	Zwischen den beiden		

	Untersuchungsvariablen	Merkmalen der Grundgesamtheit besteht kein Zusammenhang.		
me-trisch	1 unabhängige kleine Stichprobe	Unterscheidet sich das arithm. Mittel einer Stichprobe signifikant vom arithm. Mittel einer GG?	t-Test für eine kleine Stichprobe	
	2 unabhängige Stichproben	Unterscheidet sich das arithm. Mittel einer Stichprobe 1 signifikant vom arithm. Mittel einer zweiten Stichprobe?	t-Test für zwei unabhängige kleine Stichproben	In einer Goldmine werden zwei Gruppen gebildet, die mit unterschiedlichen Suchstrategien arbeitet. Die erste Gruppe erreicht im Durchschnitt 5 kg am Tag, die zweite Gruppe 7 kg. Unterscheiden sich die Gruppen signifikant voneinander?
	mehr als 2 unab. Stichproben	Stammen zwei oder mehr Mittelwerte aus der gleichen Grundgesamtheit?	einfache Varianzanalyse	Beeinflussen unterschiedliche Düngersorten (5 Sorten) den Getreideertrag?
	zwei verbundene Stichproben		t-Test für zwei verbundene Stichproben	
	mehr als 2 Stichproben		Varianzanalyse zweifacher Klassifikation	Beeinflussen verschiedene Anfangsgewichte (Faktor 1) von Schweinen sowie die verwendete Futterart (Faktor 2) die Gewchtszunahme (Untersuchungsvariable)?

1 Tests bei metrischem Meßniveau

1.1 Chi-Quadrat-Tests

1.1.1 χ^2-Anpassungstest

1.1.1.1 Testsituation

- eine einfach Zufallsstichprobe des Umfangs n
- die Untersuchungsvariable ist polytom → r mögliche Wauterealisationen
- es werden für die Werterealisationen die absoluten Häufigkeiten b_i i=1…,r, gezählt

1.1.1.2 Testidee

Der χ^2-Anpassungstest ist ein rechnerischer Test. Im Gegensatz zum Kolomogoroff-Smirnov-Test muß die H_0 nicht voll spezifiziert sein.
Geprüft wird, ob eine empirische Häufigkeitsverteilung einer theoretisch angenommenen ähnelt, ob sie sich ihr anpaßt:
H_0: $F(x)=F_0(x)$ H_a: $F(x)\neq F_0(x)$
Anders formuliert: Gegeben ist eine Stichprobe aus einer Grundgesamtheit. Man kennt die Verteilungsfunktion der Grundgesamtheit nicht, testet, ob sie einer theoretischen Verteilung folgt, z.B. der Normalverteilung, Binominalverteilung, Poissonverteilung.

Die Nullhypothese behauptet also, die Grundgesamtheit habe eine bestimmte Verteilung, wobei es sich um stetige wie diskrete Zufallsverteilungen handeln kann.

Mathematisch werden beim Chiquadrat-Anpassungstest die Abstände zwischen individuellen beobachteten und individuellen theoretischen Häufigkeiten in die Prüfvariable U transformiert, die (approximativ) eine Chiquadratvariable ist.

1.1.1.3 Die Prüfvariable

Die Prüfvariable ist die Pearsonsche Variable U:

$$U = \sum \left[\frac{(empirische - theoretische \quad Häufigkeiten)^2}{theoretische Häufigkeiten} \right] = \sum_{i=1}^{r} \frac{(B_i - e_i)^2}{e_i}$$

Es geht nicht um signifikanten Abweichungen einzelner Werte, sondern um signifikante Abweichungen der gesamten Verteilungsfunktion. Deshalb wird über 1 summiert.

Besteht ein großer Abstand zwischen e_i und $b_{i,}$ wird die H_0 eher abgelehnt.

Bei jedem Chiquadrat-Anpassungstest muß man die Zahl der **Freiheitsgrade** bestimmen. Die Zahl der Freiheitsgrade hängt von der Zahl der Klassen und der Anzahl der Parameter, die geschätzt werden müssen ab. Bei r Klassen und γ zu schätzenden Parametern beträgt die Zahl der Freiheitsgrade υ:
υ = r- γ - 1.
Soll beispielsweise getestet werden, ob eine Nominalverteilung vorliegt, so hat man zwei Parameter, nämlich μ(Mittelwert) und σ (Streuung), d.h. υ = 2. Bei einer Binominalverteilung muß π geschätzt werden (υ = 1), bei einer Poissonverteilung λ (υ=1).

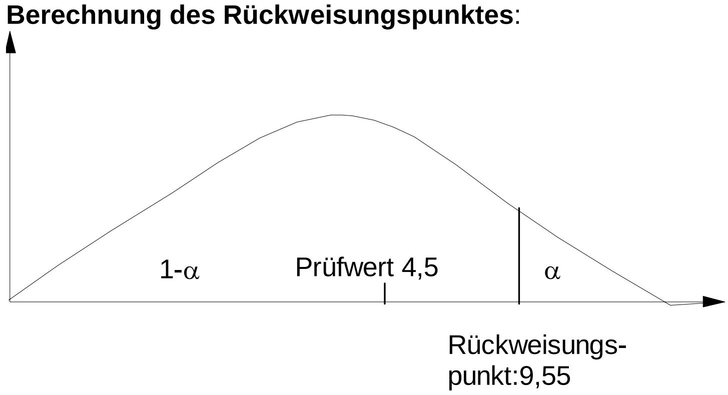
1.1.1.4 Beispiel (aus Tiede, S.42)

200 Studenten wurden nach dem ihnen am sympathischsten Fach befragt.

Fach	b_i	e_i	$(b_i-e_i)^2$	$\frac{(b_i - e_i)^2}{e_i}$
VWL	45	40	25	0,625
BWL	37	40	9	0,225
WiPo	48	40	64	1,600
Finanzwissenschaft	31	40	64	1,600
Statistik	39	40	1	0,025
	200	200		4,5

Hypothese: Geprüft werden soll, ob die Studenten alle fünf Fächer gleich sympathisch finden. Das heißt, wenn diese Hypothese zutrifft, müßten jeweils 1/5 der Studenten die fünf Fächer bevorzugen.
Die Nullhypothese lautet: $H_o = \pi = \pi_{0i} = 0{,}2; i = 1..5$ oder in Worten ausgedrückt: π ist genauso verteilt wie π_{0i}.
Das **Signifikanzniveau** sei 5 Prozent.

Berechnung des Rückweisungspunktes:



Da keine Parameter geschätzt werden sollen und wir fünf Klassen (=Fächer) haben, ist die Zahl der Freiheitsgrade=5-0-1=4. In der Tabelle (Formelsammlung S. 30) liest man nach, daß für 4 Freiheitsgrade und fünfprozentigem Signifikanzniveau der Rückweisungspunkt bei $\chi^2 = 9{,}49$ ist.

Nun wird die Pearsonschen Prüfvariable: $\sum_{i=1}^{r} \frac{(B_i - e_i)^2}{e_i}$ berechnet (s. Tabelle).

Als Wert ergibt sich u=4,5.

Der Prüfwert u=4,5 ist kleiner als der Rückweisungspunkt von 9,45 d.h. die Nullhypothese wird nicht verworfen, die Studenten finden alle Fächer gleich sympathisch, die Variable ist in Wahrheit gleichverteilt.

1.1.1.5 Bemerkungen

1. Für den Fall, daß der Stichprobenumfang grenzenlos wächst, folgt U einer Chiquadratverteilung mit v=r-1 Freiheitsgeraden. Die Prüfgröße U ist nämlich approximativ eine χ^2-Variable.
2. Eigentlich gilt der χ^2-Anpassungstest nur für $n \rightarrow \infty$, man kann ihn aber auch für kleine Werte von n durchführen. Jede einzelne Klasse der theoretischen Verteilung muß mindestens mit 5 „theoretischen Beobachtungen" besetzt sein: $e_i \geq 5$
3. Die Stichprobe sollte nicht zu klein sein, die erwarteten Häufigkeiten nicht unter 1 liegen.
4. Voraussetzung für die Anwendbarkeit des χ^2-Anpassungstest (Tiede, S. 42):
   Die Modellvoraussetzungen für die **Multinominalverteilung** müssen gegeben sein:
   → bei jedem Einzelzug muß das Stichprobenelement mit der von Zug zu Zug unveränderlichen Wahrscheinlichkeit π_1 in die i-te Klasse fallen: Bei großem Auswahlsatz verwendet man das Modell mit Zurücklegen, bei kleinen wäre das nicht nötig. Führt man den Test mit Daten durch, deren Meßniveau über dem nominalen liegt, muß man die Daten gruppieren.
5. Der χ^2-Anpassungstest ist auch für den Fall von nur zwei Klassen verwendbar, dann entspricht er einem beidseitigem Binominaltest. Im Fall zweier Klassen entspricht der PEARSONsche Stichprobenumfang nämlich approximativ einer quadrierten Standardnormalvariable, also eine χ^2-Variable mit einem Freiheitsgrad. Beim χ^2-Test vergleicht man k^2 mit k_r^2, beim zweiseitigen Binominaltest k mit k_r.

1.1.2 χ^2-Homogenitätstest

Der χ^2-Homogenitätstest ist ein Test zur Prüfung der Unterschiede zwischen Anteilswerten aus mehr als zwei unabhängige Stichproben. Die Problemstellung entspricht der Problemstellung der Varianzanalyse.

Testsituation

r unabhängige Stichproben der Umfänge b_i

Vorgehensweise

n Objekte kommen in die Auswahl, in jeder Einzelstichprobe werden die *s* Ausprägungen des interessierenden polytomen Merkmals A ausgezählt. Man beobachtet die absoluten Häufigkeiten b_{ij} (bzw. die relativen Häufigkeiten $p_{ij} = \frac{b_{ij}}{n}$)

Zu prüfen ist die Hypothese: p_{ij} weicht nur zufällig von Grundgesamtheitsanteilswerten π_{ij} ab. Die Homogenitätshypothese dieses Tests behauptet also, daß ein Merkmal in den *r* Stichproben jeweils zugrundeliegenden Grundgesamtheiten jeweils die gleiche Verteilung hat. Zu prüfen ist also die Hypothese, daß die Stichproben aus einer Grundgesamtheit stammen: $H_0 : \pi_{ij} = \pi_j$ *für alle i; j=1…s*

Prüfvariable ist wie bei allen Chi-Quadrat-Tests die Pearsonsche Variable U:

$$U = \sum_{j=1}^{a} \sum_{i=1}^{r} \frac{(B_{ij} - e_{ij})^2}{e_{ij}}$$

Praktisch werden bei diesem Test die beobachteten Häufigkeiten b_{ij} mit den aufgrund der Homogenitätshypothese erwarteten Werten e_{ij} verglichen. Hierzu wird ein Chi-Quadrat-Test durchgeführt.

Beispiel

An der Ruhr-Uni Bochum wurden StudentInnen befragt, wie sie die Anforderungen ihres Studiums einschätzen. Es ergab sich folgender Befund:

Anforderungen Fakultät	zu hoch	gerade noch zu bewältigen	gut zu bewältigen bzw. gering	Summe
Philologie	10 (15,1)	58 (62,3)	47 (37,6)	115
Jura	6	69	45	120

	(15,8)	(65,0)	(39,3)	
Wirtschaftswissenschaft	18 (11,4)	53 (47,1)	16 (28,5)	87
Sozialwissenschaft	3 (7,5)	25 (30,9)	29 (18,6)	57
Maschinenbau	20 (7,2)	30 (29,8)	5 (18,0)	55
Summe	57	235	142	434

Entschieden werden soll, ob sich die Antworten über die Studienanforderungen zwischen den ausgewählten 5 Fakultäten wesentlich unterscheiden. Die Berechnung der erwarteten Häufigkeiten führt zu den eingeklammerten Werten: z.B. für „Philologie“ und „gerade noch zu bewältigen“ ergibt sich ein Wert von

$$e_{12} = \frac{Randsumme1 \cdot Randsumme2}{Gesamtsumme} = \frac{115 \cdot 235}{434} = 62{,}3$$

Für die Prüfvariable u ergibt sich ein Wert von 63,3 (Die Berechnung erfolgt analog zum Chi-Quadrat-Unabhängigkeitstest, S. 5).

Die χ^2-Verteilung besitzt *v=(r-1)(s-1)=4•2=8* Freiheitsgrade.

Der Rückweisungspunkt liegt für *v=8* und ein Signifikanzniveau von 10 Prozent bei 13,4 (Nachschlagen in Chi-Quadrat-Verteilung).

Der Stichprobenbefund u=63,3 liegt rechts vom entsprechenden Rückweisungspunkt χ_r^2 =13,4, die Hypothese muß zurückgewiesen werfen.

Ergebnis: Die Nullhypothese muß abgelehnt werden, die Stichproben stammen nicht aus einer GG, die Unterschiede zwischen den Einschätzungen der einzelnen Studiengänge sind signfikant.

1.1.3 χ^2-Unabhängigkeitstest

Testsituation:

- zwei dichotome oder polytome Merkmale X und Y
- eine Stichprobe mit *n* Objekten, die jeweils eine Ausprägung von X und Y haben

Vorgehensweise

Der χ^2-Unabhängigkeitstest prüft die Hypothese, daß die beiden dichotomen/polytomen Merkmale in der Grundgesamtheit voneinander unabhängig sind:

H_0: $\pi_{ij}=\pi_{i.}\pi_{.j}$

Das weitere Vorgehen (Berechnung der Prüfvariable etc.) entspricht dem Vorgehen beim χ^2-Anpassungstest (S. 4)

Beispiel

vergleiche das Beispiel zum Anpassungstest, S. 4.

Gezogen wird eine Stichprobe, wobei jede Person zwei Merkmale trägt: „Zugehörigkeit zu einer Fakultät“ und „Beurteilung der Studienanforderungen“. Geprüft werden soll, ob die These der **Unabhängigkeit** zwischen „Zugehörigkeit zu einer Fakultät“ und „Beurteilung der Studienanforderungen“ gestützt werden kann oder ob sie widerlegt werden muß. Der nun analog zu den anderen Chi-Quadrat-Tests durchzuführende Test wird als χ^2-Unabhängigkeitstest bezeichnet.

1.2 t-Tests

1.2.1 STUDENTisierung

Die Tatsache, daß die Stichprobenverteilung des arithmetischen Mittels bei normalverteilter Gesamtheit ebenfalls normalverteilt ist, läßt sich bei bekannter Grundgesamtheitsvarianz nutzen. Ist diese jedoch unbekannt, wäre eine entsprechende Schätzung über den Stichprobenbefund bei kleinem Stichprobenumfang zu ungenau. Eine Prüfvariable, die dies berücksichtigt, läßt sich mit Hilfe des StUDENTisierung-Verfahrens gewinnen:

Bei normalverteilter Grundgesamtheit gilt: $K = \dfrac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$.

Es läßt sich zeigen, daß $\frac{\sigma}{\sqrt{n}}$ wie $\sqrt{\frac{\chi^2}{v}}$ verteilt ist (S ist die zur Stichprobenstreuung gehörende Stichprobenvariable). Man kann nun die Definition der Student-t-Variablen

Varianz in der Grundgesamtheit / Verteilung der GG	bekannt	unbekannt
normalverteilt	$\frac{\overline{X}-\sigma}{\frac{\sigma}{\sqrt{n}}} = K$	$\frac{(\overline{X}-\mu)}{\frac{\sqrt{n}}{S}} = t, v = n-1$
nicht normalverteilt	Ausweichen auf verteilungsfreie Verfahren!	

einsetzen und erhält die Größe $t = \frac{K}{\sqrt{\frac{\chi^2}{v}}} = \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$

1.2.2 t-Test bei einer kleinen Stichprobe

Testsituation:

Der Mittelwertdifferenzentest ist ein t-Test bei einem kleinen Stichprobenumfang: n<30.
Voraussetzung ist, daß die Verteilung in der Grundgesamtheit entweder dichotom oder normalverteil ist, sonst sollte man auf andere verteilungsfreie (aber ebenso effiziente) Prüfverfahren ausweichen, z.B. den U-Test.)

Der t-Test testet, ob das arithmetische Mittel der Stichprobe sich signifikant vom arithmetischen Mittel der Grundgesamtheit unterscheidet. Der t-Test nutzt die Information, daß die Stichprobenverteilung des arithmetischen Mittels bei normalverteilter GG ebenfalls normalverteilt ist. Diese Beziehung läßt sich nutzen, wenn die Grundgesamtheitsvarianz ebenfalls 0 ist.

Es ergeben sich folgende Testsituationen:

S ist die zur Stichprobenstreuung gehörende Stichprobenvariable.

Vorgehensweise:

- Die Nullhypothese lautet: $H_0: \mu = \mu_0$
- Man bestimmt den Wert der t-Variablen unter der Verwendung von $\bar{x}, \mu_0, s$ und n .
- Dieser Wert wird mit t_r (ist abhängig vom SN α und dem Parameter *v=n-1)* verglichen (Formelsammlung S. 31)

1.2.3 Mittelwertdifferenzentest: t-Test bei zwei unabhängigen kleinen Stichproben

Testidee

Der t-Test für zwei unabhängige kleine Stichproben geht zurück auf das Studentisierungs-Verfahren (s. S.7).
Bei kleinen Stichprobenumfängen lassen sich Aussagen über die Verteilungen der bislang verwendeten Prüfgrößen nur machen, sofern die Grundgesamtheiten normalverteilt sind. Ist dies nicht der Fall, weicht man am besten auf verteilungsfreie Verfahren aus, z.B. den U-Test.

Testsituation

- zwei unabhängige kleine Stichproben (n<30)
- Voraussetzung: GG sind normalverteilt, Varianzen der Gesamtheiten sind gleich (Varianzquotiententest durchführen!)

Die Testsituation ergibt sich gemäß der Formelsammlung, S. 48.

Beispiel

Es werden zwei Gruppen von 4 bzw. 5 Schülern zufällig ausgewählt, die durch unterschiedliche Unterrichtsmethoden auf die Lösung von handwerklichen Problemen vorbereitet werden. Nach Abschluß des Unterrichts müssen die 9 Schüler jeweils 30 Probleme lösen. Man erhält für die beiden Gruppen folgendes Ergebnis (gelöste Probleme) und möchte wissen, ob die Differenz zwischen den beiden Grupenmitteln bei 5% Signifikanzniveau (einseitig) nur aus zufälligen Gründen aufgetreten ist.

Gruppe 1	13	15	17	18	
Gruppe 2	14	16	18	22	23

1. Die Nullhypothese lautet: $\pi_1 = \pi_2$
2. Man berechnet die arithmetischen Mittel: $\bar{x}_1 = 15{,}75$ $\bar{x}_2 = 18{,}6$
3. ... und die Varianzen $s_1^2 = 4{,}92$ und $s_2^2 = 14{,}8$ (für den t-Test *ist Voraussetzung, daß die GG normalverteilt sind und die Varianzen der Gesamtheiten gleich sind* → ggf mit Varianzquotiententest prüfen!)
4. Jetzt wird die Prüfgröße t_{1r} berechnet:
   $$\frac{D_{\bar{x}} - \delta_0}{\sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}}} = \frac{2{,}85 - 0}{\sqrt{\frac{4{,}92}{4} + \frac{14{,}8}{5}}} = \frac{2{,}85}{\sqrt{1{,}25 + 2{,}96}} = \frac{2{,}85}{2{,}05} = 1{,}39$$
   $D_{\bar{X}}$: Differenz der arithmetischen Mittel
   δ_0 : ?, Differenz der arithm. Mittel in der Grundgesamtheit (?), kann aufgrund der Homogenitätshypothese mit 0 angenommen werden
   $\sigma_1{}^2 = s_1^2$ =Varianz der Stichprobe 1,
5. Für $v=n_1+n_1-2=7$ Freiheitsgerade ergibt sich bei einem SN von 5% einseitig (also bei 10% nachschauen, Tabelle Formelsammlung S.31) ein Rückweisungspunkt von 1,81.
6. Die Prüfgröße liegt links vom Rückweisungspunkt (1,39<1,89), d.h. die Nullhypothese kann bestehen bleiben. Die Differenz zwischen den beiden Gruppenmitteln ist zufällig.

1.2.4 Mittelwertdifferenzentest: t-Test bei zwei verbundenen Stichproben

Der t-Test für zwei verbundene Stichproben läßt sich als Problem des t-Test-Ein-Stichproben-Falls (S.8) auffassen. Ausgegangen wird von der Differenz der $D=X_1-X_2$. Aus den Paaren der Stichprobenvariablen (X_{i1},X_{2i}) und berechnet daraus den Durchschnitt:

$\bar{D} = \frac{1}{n}\sum D_i = \frac{1}{n}(X_{1i} - X_{2i}) = \bar{X}_1 - \bar{X}_2$ (weiteres siehe Formelsammlung)

1.3 F-Test: Varianzquotiententest

Der F-Test heißt auch Varianzquotiententest. Er ist ein Test für ***metrisches Meßniveau*** [→ Formelsammlung S. 48f]

Der F-Test bewertet die Varianzen zweier Grundgesamtheiten. Er testet, ob zwei Varianzen zufällig voneinander verschieden sind oder nicht.

Bsp.: Eine Klasse wird am Anfang des Schuljahres in zwei Gruppen geteilt. Der gleiche Lehrer benutzt unterschiedliche Lehrmethoden bei den Gruppen. Am Ende des Schuljahres wird ein Test geschrieben. Die Durchschnittspunktzahl beträgt 4.8 und 5.3. Frage: Ist der Unterschied zufällig?

Der F-Test ist auch für kleine Stichprobenumfänge geeignet. Er spielt bei der Varianzanalyse eine bedeutende Rolle, ist aber auch von Bedeutung, wenn man prüfen will, ob die Voraussetzungen für die Anwendung des t-Differenzentests für den Vergleich zweier Mittelwerte unabhängiger Stichproben erfüllt sind.

1.3.1 Testsituation

Aus zwei unabhängigen Stichproben der Umfänge n_1 und n_2 wird die metrisch skalierte Untersuchungsvariable X *aus normalverteilten Grundgesamtheiten* erhoben. Da die Grundgesamtheiten normalverteilt sind, genügt die Prüfvariable $\frac{S_1^{\,2}}{S_2^{\,2}}$ einer F-Verteilung mit

$v_1=n_1-1$ und $v_2=n_2-1$.

Die Nullhypothese lautet:

$$H_0 = \frac{\sigma_1^2}{\sigma_2^2} = 1$$

σ_1^2 Grundgesamtheitswert für $s_2^{\,2}$

$\sigma_2^{\,2}$ Grundgesamtheitswert für $s_2^{\,2}$

H_o: Die beiden Varianzen sind gleich.
Ist $s_2^{\,2}$ wesentlich größer als $s_1^{\,2}$, kann man sich die genaue Hypothesenprüfung sparen. Ansonsten wird die eingangsformulierte H_o mit einem F-Test geprüft.

1.3.2 Die Prüfvariable

Die Prüfvariable ist gegeben durch: $\frac{s_1^{\,2}}{s^{\,2}} = \frac{\text{Streuung zwischen den Gruppen}}{\text{Streuung innerhalb der Gruppen}}$

Sie muß 1 ergeben, wenn die Varianzen aus einer Grundgesamtheit stammen.
Die Rückweisungspunkte ergeben sich durch

$v_1 = r - 1$ r=Anzahl der Gruppen, *n*: Anzahl der Personen insgesamt

$v_2 = r(n-1)$

1.3.3 Lösungsansatz

1. Signifikanzniveau festlegen
2. den Wert für die Prüfvariable berechnen
3. die Rückweisungspunkte berechnen
4. In der Formelsammlung auf S. 32 in der Tabelle nachschauen, ob der Wert für die Prüfvariable größer oder kleiner als der dort angegebene Wert ist.
   Wenn die Prüfvariable größer als der tabellierte Wert ist, ist die Hypothese bestätigt: Es bestehen signifikante Unterschiede.
   Wenn die Prüfvariable kleiner als der tabellierte Wert ist, muß die Hypothese verworfen werden: Es bestehen keine signifikante Unterschiede.

1.3.4 Beispiel

Es werden zwei Gruppen von 4 bzw. 5 Schülern zufällig ausgewählt, die durch unterschiedliche Unterrichtsmethoden auf die Lösung von handwerklichen Problemen vorbereitet werden. Nach Abschluß des Unterrichts müssen die 9 Schüler jeweils 30 Probleme lösen. Unterscheiden sich die Varianzen signifikant voneinander?

Gruppe 1	13	15	17	18	
Gruppe 2	14	16	18	22	23

1. H_0: $\frac{s_1^{\,2}}{s^{\,2}}$
2. Die Varianzen ergeben sich als $s_1^2 = 4{,}92$ und $s_2^2 = 14{,}8$
3. Der Prüfwert ergibt sich aus $\frac{s_1^{\,2}}{s^{\,2}}$ (die größere Varianz steht im Zähler!)= $\frac{14{,}8}{4{,}92} \approx 3$
4. Der Rückweisungspunkt liegt für $v_1=n_1-1=3$ und $v_2=v_2=n_2-1=4$ bei 9,12 (SN=5%)
5. 3<9,12, die Nullhypothese (hier: Homogenitätshypothese) wird angenommen, die Varianzen unterscheiden sich nur zufällig voneinander.

1.4 Varianzanalyse (engl.: ANOVA)

1.4.1 Allgemeines

Die Varianzanalyse gehört zu den *Signifikanztests*. Sie (engl.: ANOVA=analysis of variance) ist ein statistisches Verfahren, durch das im allgemeinen geprüft wird, ob die Mittelwerte μ zweier oder mehrerer Stichproben aus Grundgesamtheiten gezogen wurden, die denselben Mittelwert besitzen. Das Verfahren ist eine Erweiterung der Zweistichprobentests, z.B. des t-Tests.

Vorraussetzung der Varianzanalyse ist die Annahme, daß alle Stichproben aus normalverteilten Grundgesamtheiten stanmmen.

Unabhängige Variablen werden im Zusammenhang mit der Varianzanalyse als **Faktoren** bezeichnet. Es handelt sich dabei stets um qualitative, nominalskalierte Variablen. Die einzelnen qualitativen Ausprägungen eines Faktors werden als Faktorstufen bezeichnet.

Im Gegensatz zu den Faktoren, handelt es sich bei den in einer Varianzanalyse betrachteten **abhängigen Variablen** immer um quantitative, intervallskalierte Variablen. Wird genau eine abhängige Variable betrachtet, so spricht man von einer univariaten Varianzanalyse. Werden mehr als eine abhänige Variable untersucht, so spricht man von einer multivariaten Varianzanalyse.

Wird lediglich ein Faktor betrachtet, so spricht man von einer einfaktoriellen Varianzanalyse. Werden mehr als ein Faktor untersucht, so spricht man von einer mehrfaktoriellen Varianzanalyse.

Je nachdem, *wieviele* Einflußgrößen [Faktoren] auf die Untersuchungsvariable gerichtet sind und die berücksichtigt werden sollen, trennt man Varianzanalysen in Varianzanalysen *einfacher* und *mehrfacher Klassifikation.*

Die verschiedenen, in der Varianzanalyse möglichen Variablenkonstellationen sind in der folgenden Tabelle nochmals zusammengefaßt:

	Anzahl der unabhängigen Variablen	
Anzahl der abhängige Variablen	1	>1
1	einfaktorielle univariate Varianzanalyse	mehrfaktorielle univariate Varianzanalyse
>1	einfaktorielle multivariate Varianzanalyse	mehrfaktorielle multivariate Varianzanalyse

1.4.2 Einfache Varianzanalyse = Varianzanalyse einfacher Klassifikation

1.4.2.1 Praktisches Anwendungsbeispiel

Für eine bestimmte Getreidesorte stehen *n* verschiedene Dünger zur Verfügung. Es soll geprüft werden, ob die verschiedenen Dünger auf den Ernteertrag den gleichen Einfluß ausüben oder nicht.

1.4.2.2 Testidee

Das Testproblem: Gegeben sind *r* unabhängige Stichproben der Umfänge n_i, deren Varianzen zwar unbekannt sein dürfen, aber **gleich** sein müssen.

Aus den Meßwerten $x_{ij}, i = 1..r;\ j = 1....n_i$ bildet man pro Stichprobe das arithmetische Mittel $x_i = \frac{1}{n_i}\sum_j x_{ij}$

Geprüft werden sollen die Unterschiede zwischen diesen Mittelwerten.
Die Hypothese lautet also: $H_0: \mu_1 = \mu_2 = ... = \mu_r$

Die Nullhypothese lautet ausformuliert: Die Stichproben stammen aus normalverteilten Grundgesamtheiten mit gleichen Mittelwerten und gleichen Varianzen, also aus einer einzigen Gesamtheit.

H_a besagt: nicht alle Mittelwerte sind gleich, in mindestens einem Fall unterscheiden sie sich
Für das Prüfverfahren benötigt man eine *Quadratsummenzerlegung*.

1.4.2.3 Das Prinzip der Quadratsummenzerlegung

Die Varianzanalyse beruht auf einer rein arithmetischen Zerlegung der „Quadratsumme“ (=Summe der Quadrate der Abweichungen der Stichprobenwerte vom Mittelwert). Die Quadratsumme wird in eine Summe von Bestandteilen zerlegt, jeder Bestandteil entspricht einer bestimmten Variationsursache eines Sachverhalts.
Bei einer einfachen Variationsanalyse wird in zwei Teile zerlegt. Der eine Teil entspricht dem **systematischen Teil**, dem Einfluß, den man untersuchen will . Der andere Teil entspricht der **Restgröße**, dem Zufallseinfluß.

$$q = \sum_{1=1}^{r}\sum_{j=1}^{m}(y_{ij} - \bar{y})^2 \qquad \mathrm{q}_1 = \sum_{\mathrm{i}=1}^{\mathrm{r}} m(\bar{y}_1 - \bar{y})^2 \qquad \mathrm{q}_2 = \sum_{\mathrm{i}=1}^{\mathrm{r}}\sum_{\mathrm{j}=1}^{\mathrm{m}}(Y_{ij} - \bar{y}_i)^2$$

q: Summe der Abstandsquadrate aller Beobachtungen
q_1: „erklärte Quadratsumme“, Summe der Abstandsquadrate der Gruppen-Arithmetische-Mittel umd das Gesamt-Arithmetische-Mittel.
q2: „unerklärte“ Quadratsumme, Summe der Quadrate innerhalb der Gruppen

1.4.2.4 Testdurchführung

1. Signifikanzniveau festlegen
2. Aus den Messwerten Yij bildet man pro Stichprobe das arithmetische Mittel
3. Quadratsumme der Mittelwerte der Gruppen berechnen
4. Quotienten bilden:

   $s_1^2 = \frac{q_1}{r-1} \qquad s_{21}^2 = \frac{q_2}{n-r}$

   Trifft die Ho zu, dürften sich s_1^2 und s_2^2 nur zufällig unterscheiden.
   Falls sie sich doch unterscheiden, müßte mittels eines F-Testes überptüft werden, ob die Unterschiede signifikant sind.
5. ggf. F-Test durchführen

1.4.3 Zweifache Varianzanalyse = Varianzanalyse zweifacher Klassifikation

Hier wird ein zweiter systematischer Faktor berücksichtigt, dieser kann mit dem ersten Faktor korrellieren, muß aber nicht.
Systematische Faktoren sind Variablen, die Stichproben trennen.
Ob die beiden Faktoren voneinander abhängen oder nicht, wird vor Aufstellen der Nullhypothese unabhängig festgelegt. Daher können zwei Fälle auftreten: Korrelation und Nicht-Korrelation.

Beispiel: Ernteertrag beeinflußt von den Faktoren Dünger und Boden. Einmal hängen diese beiden Faktoren voneinander ab, einmal nicht!

1.4.3.1 Eine Beobachtung pro Zelle

Faktor A darf in *r* Stufen, Faktor B in *t* Stufen auftreten. Gezogen werden r*t Stichproben, jeweils vom Umfang 1 (Eine Beobachtung pro Zelle). Damit schließt man aus, daß man

eventuelle Wechselwirkungen (Interaktionen) beobachtet. Die Faktoren A und B müssen also unkorreliert sein.
Geprüft werden muß die Hypothese: $H_o = \mu_{ik} = \mu$ für alle i = 1...r und k = 1.... t
Die Quadratsummenzerlegung heißt in diesem Fall:
$q=q_1+q_2+q_3$ (ausführlich, siehe Formelsammlung S.52)
Die Stichprobenrealisationen s_1^2, s_2^2 und s_3^2 dürfen nur durch den Zufall akzeptierbaren Grenzen voneinander abweichen.
Wenn die Nullhypothese nicht zutrifft, ist $s_1^2 > s_3^2$ oder/und $s_2^2 > s_3^2$.
Man führt also zwei F-Tests durch, einen für $\frac{s_1^2}{s_3^2}$ und für $\frac{s_2^2}{s_3^2}$ (wobei Abwesenheit einer Wechselwirkung unterstellt wird).

1.4.4 Der Begriff der „Wechselwirkung“ (Interaktion)

Die Wechselwirkung bzw. Interaktion ist die Bezeichnung für die gemeinsame Wirkung zweier (z.B. bei der zweifaktoriellen Varianzanalyse) oder mehrere Variablen (mehrfaktorielle VA) auf eine dritte abhängige Variable, die nicht aus der Addition der einzelnen Einflüsse resultiert. Wechselwirkungen treten nicht auf, wenn die Faktoren unabhängig voneinander sind.

Bei einer zweifaktoriellen Varianzanalyse mit nur einer Beobachtung pro Zelle lassen sich die Auswirkungen von der Interaktion des Faktors A und B nicht von den Auswirkungen der Störvariablen trennen, weil sich dann keine zufallabhängige Variation innerhalb der Zellen bilden läßt. Hier besteht nur die Möglichkeit, die Variation zwischen den Stufenkombinationen der beiden Faktoren dem Zufall zuzuschreiben und eine Wechselwirkung auszuschließen. Es empfiehlt sich daher, stets mehrere Beobachtungen pro Zelle durchzuführen, um eine vierte Quadratsumme, die auch die Variation innerhalb der Zellen erfaßt, als Bezugsgröße zu erhalten.

1.4.5 Was hat die bivariate Regressionsrechnung mit der Varianzanalyse zu tun?

Die Variation auf der Regressionshyperebenen entspricht der Variation zwischen den Stufen der Varianzanlyse: Bei der einfaktoriellen Varianzanalyse stützt man sich zur Prüfung des Einflusses einer mehrmals gestuften unabhängigen Variable auf eine abhängige Variable auf eine Quadratsummenzerlegung. Die Gesamtvarianz der Meßwerte q wird in eine Variation der Gruppenmittelwerte bezüglich des Gruppenmittelwertes (Variation zwischen den Stufen) q_1 und in eine Variation der Einzelwerte bezüglich der Gruppenmittelwerte (Variation innerhalb der Stichprobe) q_2 aufgeteilt. Im allgemeinen sind q_1 und q_2 größer Null, bei q_1 wegen der Faktorwirkung und aus Zufallsgründen, bei q_2 ausschließlich aus Zufallsgründen. Bei der **multiplen Regressionshyperebene** (aber auch der einfachen Regressionsanalyse) wird ebenso die gesamte Varianz in einen deterministischen Teil (Varianz auf der Regressionshyperebene) und einen stochastischen Teil (Varianz um die Regressionshyperebene) aufgeteilt. Unter der Voraussetzung, daß die Gruppenmittelwerte der Varianzanalyse $\bar{y}_i$ den $\hat{y}_i$-Werten afu der Regressionshyperebene entsprechen, gilt sogar: Die Variation zwischen den Stufen der Varianzanalyse $q_1 = \sum_{i=!}^{r} m(\bar{y}_i - \bar{y})^2$ entspricht der Variation auf der Regressionshyperebene bei der bivariaten (und multiplen) Regressionsanalyse $q_1 = \sum \left(\bar{y}i - \bar{y}\right)^2$.

Diese Beziehung gilt nicht immer, sondern nur dann, wenn die exogene Variable klassifiziert vorliegt, z.B. 0-1-Kodierung.

2 Tests bei nominalem Meßniveau

2.1 Vorbemerkung

Bei nominalskalierten Daten ist der Anteilswert die statistische Maßzahl beim Testen.
Bei einem großen Stichprobenumfang dient die Normalverteilung als Stichprobenverteilung.

2.2 Binominaltest (= Anteilswerttest)

2.2.1 Testsituation

eine diskrete dichotome (→ zwei Ausprägungen) Zufallsvariable (z.B. Geschlecht mit den Ausprägungen männlich und weiblich)

2.2.2 Testidee

Geprüft werden soll, ob die Verteilungsfunktion F(x) der Zufallsvariablen X der durch die Hypothese festgelegten Verteilungsfunktion $F_0(x)$ entspricht, d.h.

H_0: $F(x)=F_0(x)$. Da die Variable dichotom ist, kann man vereinfachen zu H_0: $\pi = \pi_0$
Man prüft also, ob die Verteilung der Variablen X in der Stichproben der in der Grundgesamtheit entspricht.
Der Rückweisungspunkt beim Binominaltest ist durch das Signifikanzniveau gegeben. Da ein Signifikanzniveau von 0,05 oder 0,10 etc. nicht exakt tabelliert ist, muß man den Punkt ungefähr bestimmen und modifiziert die Nullhypothese H_o: $\pi \leq \pi_o$. Jeder Wert ≤ dem Rückweisungspunkt ist Rückweisungspunkt.

2.2.3 Methode

Da wir nur zwei Klassen vorliegen haben,
1. zählt man die Besetzungszahlen der beiden Klassen aus und
2. berechnet man die relativen Häufigkeiten p bzw. 1-p

In der Regel ergibt sich ein Unterschied zwischen der Verteilung der Grundgesamtheit und der Stichprobe. Geprüft wird nun, ob der Unterschied zufällig oder signifikant ist. (Signifikanzniveau beispielsweise 5 %).

2.2.4 Beispiel I (Vorlesung)

Ein Hersteller von Geräten behauptet, daß höchstens 5 Prozent defekte Stücke produziert werden. Per Zufall werden 20 Stücke entnommen. 3 Stücke sind defekt.
Frage: Kann man dem Hersteller bei 10% Signifikanzniveau trauen?
Folgende Werte sind gegeben:
- n = 20
- x = 3 (Anzahl der defekten Stücke)
- $p = \frac{x}{n} = \frac{3}{20}$ (Anteilswert)

Es gibt nun zwei Möglichkeiten, den Test durchzuführen:
1. über den Anteilswert p
2. über die absolute Häufigkeit

Berechnung über den Anteilswert:
- H_0: $\pi_0 \leq 0{,}05$ (Es sind höchstens 5 Prozent kaputt sind, 5 %=0,05; π entspricht der Erfolgswahrscheinlichkeit, ein fehlerhaftes Stück aufzufinden)
- H_a: $\pi_a > 0{,}05$ (in Wirklichkeit ist die Zahl der defekten Stücke größer als 5 Prozent)
- $p = \frac{x}{n} = \frac{3}{20} = 0{,}15$ (der Anteil der defekten Stücke beträgt 15 Prozent)
- P folgt $\frac{1}{20} B(20;0{,}05)$

- Der **Rückweisungspunkt** ergibt sich aus dem Signifikanzniveau. Das Signifikanzniveau beträgt 10%, d.h. $P(X \geq x_r)=0,10$.
- Da 0,10 nicht exakt in der Binominalverteilungstabelle tabelliert ist, schlägt man die beiden Werte nach, nie als nächstes an 0,1 herankommen (dabei benutzt man die Erfolgswahrscheinlichkeit des defekten Stückes von π=0,05 oder 5%)

Binominalverteilung B(n, π) (Auszug aus Formelsammlung B-Verteilung, Tafel 9):

n	X	0,050
20	0	0,358
	1	0,377
	2	0,189
	3	0,060
	4	0,013
	5	0,002
	6	0,000

Die Wahrscheinlichkeit, daß mehr als 3 defekte Stücke auftreten, $P(X \geq 3)$ ist 0,073 ($\rightarrow$man addiert die Werte ab 3 bis 20). 0,073 $\leq$0,10, d.h. der Rückweisungspunkt würde unterschritten.

Die Wahrscheinlichkeit, daß mehr als 2 defekte Stücke auftreten, $P(X \geq 2)$ ist 0,262 ($\rightarrow$man addiert die Werte ab 3 bis 20). 0,262 $\geq$0,10, d.h. der Rückweisungspunkt würde überschritten.

Der Wert für den Rückweisungspunkt läge zwischen 2 und 3, man wählt den Punkt, bei dem H_0 höchstwahrscheinlich nicht verworfen wird, d.h. der Rückweisungspunkt lautet nicht mehr p_r=0,10, sondern $p_r \leq$0,10, formal korrekt geschrieben: $P(X \geq x_r) \leq 0,10$. Diese Verschiebung des Signifikanzniveaus bezeichnet man als „Konservativer Test“:
Der Punkt, der für die Beibehaltung der H_0 tendenziell günstig liegt (Beispiel: genau 10% Signifikanzniveau) ist nicht möglich. Man wählt daher ein kleineres Signifikanzniveau, um die H_0 beibehalten zu können. Beim konservativen Testen werden die Parameter abgeschätzt, die H_0 im Zweifel eher angenommen.

Ergebnis: Der Stichprobenbefund von drei oder mehr defekten Stücken führt zur Rückweisung der H_0. Maximal zwei defekte Stücke sind erlaubt.

2.2.5 Beispiel II: klassische Aufgabenstellung für den Binomialtest

Zu testen ist die Hypothese, daß der Anteil weiblicher Studierender an der Ruhr-Uni-Bochum bei 50% liegt (Signifikanzniveau: 10 % einseitig). In einer Zufallsstichprobe werden n=16, x=4 Frauen beobachtet.

Frage: Sind die Frauen unterrepräsentiert?

Begründung des Prüverfahrens
Geprüft werden soll hier eine Hypothese über die Verteilung einer Zufallsvariablen in der Grundgesamtheit. Eine solche Hypothese läßt sich mit einem Anpassungstest prüfen. Da es sich hier um eine dichotome Untersuchungsvariable bei nominalem Meßniveau handelt, ist der Binomialtest das geeignete Prüfinstrument.

Prüfung der Voraussetzungen für ein Bernoulli-Experiment
1. Zwei Ergebnisalternativen: a-weiblich: P(a)=0.5 b-männlich: $P(\overline{A})$=1-0.5=0.5
2. Die Einzelwahrscheinlichkeiten sind mit π=0.5 konstant.
3. Die Wahrscheinlichkeit bei den 16 Experimenten ist gleich groß.
4. Die Zufallsstichprobe gewährleistet die Unabhängigkeit der Ergebnisse.

Es liegt eine Versuchsanordnung nach Bernoulli vor, der Binomialtest kann durchgeführt werden.

Die Nullhypothese lautet: $H_0\!:\pi_0 = 0.5$ (Die Frauen sind nicht unterrepräsentiert)

Bei einer Bernoulli-Verteilung folgt die Zufallsvariable X einer Binomialverteilung *B(n,π)*, x folgt hier *B(16, 0.5)*. Untersucht werden muß, ob der Stichprobenbefund (4 Frauen) oder ein noch weiter von der Hypothese abweichender Punkt (3, 2, 1, keine Frau) realisiert wird.

$P(p \leq 0.25 | H_0)$

→ falls $p > \alpha$, wird die H_0 angenommen

→ falls $p < \alpha$, wird die H_0 abgelehnt

Aus der Tabelllierung der Binomialverteilung kann man die Einzelwahrscheinlichkeiten für das Auftreten von keiner, 1, 2, 3, 4 Frauen ablesen:

P(4)=0.028	Die Wahrscheinlichkeit, daß der Stichprobenbefund oder ein noch weiter entfernter
P(3)=0.009	Befund realisiert wird, beträgt 0.039. Das entspricht einer prozentualen
P(2)=0.002	Wahrscheinlichkeit von 3,9 Prozent.
P(1)=0.0	
P(0)=0.0	Der Rückweisungspunt p_r ergibt sich durch das Signifikanzniveau von 10%, also
Σ 0.039	einem Anteilswert von 0.1

0.039<0.1 ($p<p_r$), daraus folgt, daß die H_0 verworfen werden muß. Die Frauen sind in der Stichprobe unterrepräsentiert.

2.2.5.1 Bemerkungen

Der Binominaltest ist ein konservativer Test. Er begünstigt die Annahme der Nullhypothese.

2.3 McNEMAR-Test

2.3.1 Testsituation

Der McNemar-Test ist ein Homogenitätstest, der sich eignet für

- zwei abhängige (=verbundene) Stichproben
- nominalskalierte, *dichotome* Variablen

! Wenn die Variablen mehrstufig sind, z.B. die Religionszugehörigkeit, muß man dichotomisieren, z.B. Katholiken gegen Nichtkatholiken oder christliche Religionen gegen andere.

2.3.2 Testidee

Der McNemar-Test gehört zu den binominalen Testverfahren. Er kann angewendet werden, wenn eine Vierfeldertafel vorliegt und wenn untersucht werden soll, ob sich bestimmte Merkmalskombinationen in bedeutsamer Weiser verändert haben.
Gegeben sind zwei Alternativmerkmale, die jeweils zwei Variationen aufweisen (dichotome Variable).
Es werden nun die Anzahl der Wechsler betrachtet, d.h. der Elemente, die in der zweiten Stichprobe eine andere Merkmalsausprägung angenommen haben. Man prüft, ob sich der Anteilswert von den Wechslern (siehe Beispiel unten) signifikant von 0,5 unterscheidet. 0,5 ist deshalb der Entscheidungswert, weil es sich um eine dichotome Variable handelt, deren Erfolgswahrscheinlichkeit 0,5 beträgt. Letztendlich wird ein Binominaltest durchgeführt, der zur Abweisung oder Annahme der Variablen führt (Details siehe Beispiel).
Der McNemar-Test ist ein konservativer Test, der oft bei vorher-nachher-Erhebungen oder mit-ohne Erhebungen durchgeführt wird.

2.3.3 Beispiel (Tiede S. 54)

Vor und nach einer Parteiveranstaltung werden die Besucher nach ihrem Wahlverhalten in Bezug auf diese Partei befragt. Für den McNemar-Test interessiert nun die Anzahl der Personen, die ihre Wahlabsichten geändert haben (→ Wechsler).

vorher↓	nachher→	vorher	nachher	Summe
Partei wird gewählt		30	35	65
Partei wird nicht gewählt		30	25	55
Summe		60	60	120

Zu diesem Zweck stellt man die Tabelle folgendermaßen um, um die Differenz zwischen vorher und nachher deutlich machen zu können:

		nachher		Summe
		Partei wird nicht gewählt	Partei wird gewählt	
	Partei wird gewählt	25 a	5 b	30
vorher	Partei wird nicht gewählt	10 c	20 d	30
Summe		35	25	60

In den vier Feldern stehen die Anzahl der Leute, die ihre Meinung geändert haben (b=5, c=10) Personen, sowie der Anzahl der Leute, die ihre Meinung beibehalten haben (a=25, d=20).
Falls die Nullhypothese $H_0 = \pi_1 = \pi_2$ zutrifft, wird man erwarten, daß sich c und b nur zufällig voneinander unterscheiden. Als Prüfvariable dient hier beispielsweise (!) die Zahl der Wechsler b.
Sie ist, wenn die Nullhypothese zutrifft, binominalverteilt nach B(b+c; 0,5). Für c gilt das gleiche!
Das heißt, man kann hier einen Binominaltest (vgl. S. 14) durchführen, der testet, ob sich der Anteilswert $\frac{b}{b+c}$ signifikant von 0,5 unterscheidet.

Auf das Beispiel angewendet ergibt sich: $p = \frac{b}{b+c} = \frac{5}{5+10} = 0{,}33$

Die Rückweisungspunkte der vorliegenden Binominalverteilung B(15;0,5) liest man nun in Tiedes Buch, S. 153, Tabelle 3, oder in der Formelsammlung (Tabellierung für Median-Vorzeichen-Test) ab. Die Rückweisungspunkte lauten 3 und 12 bei einem Signifikanzniveau von 5 %.

$p_r = \frac{3}{15} = 0{,}2$ und $p_r = \frac{12}{15} = 0{,}8$

Da 0,2 (→ 1. Rückweisungspunkt) ≤ 0,33 (Prüfwert) ≤ 0,8 (→ 2. Rückweisungspunkt) ist, wird die Nullhypothese angenommen. D.h. der Unterschied zwischen vorher und nachher ist zufällig. Das heißt ebenfalls: De Anteilswerte $p_1 = \frac{a+b}{n} = \frac{30}{60} = 0{,}5$ und $p_2 = \frac{a+c}{n} = \frac{35}{60} = 0{,}583$ unterscheiden sich nicht signifikant voneinander.

Auf den praktischen Inhalt der Aufgabe bezogen heißt das, daß die Wahlveranstaltung die Chancen der Partei nicht wesentlich verändert hat.

Der McNemar-Test nutzt die Tabellierung für den Median-(Vorzeichen)-Test.

2.4 Differenzentest für zwei Anteilswerte aus zwei großen unabhängigen Stichproben

Tiede faßt die Ergebnisse des ersten Bandes, S. 160 und S. 182, zusammen, in dem die Stichprobenverteilung der Differenz zweier Anteilswerte bereits erörtert wurde.

2.4.1 Testsituation

- zwei große unabhängige Stichproben
- die dichotome Variable ist nominal skaliert
- die Auszählung der Kategorien ist gegeben, p_1 und p_2 sind bekannt

2.4.2 Testidee

Gegeben ist eine Stichprobe mit den Anteilswert π_1. Über die Grundgesamtheit vermutet man jedoch einen Grundgesamtheitsanteilswert von π_2. Ist der Unterschied zufällig oder signifikant?
Bei großem Stichprobenumfang und unbekannten Grundgesamtheitsangaben kann man die Stichprobenverteilung der Differenz zweier Anteilswerte durch eine Normalverteilung approximieren.
Die Rückweisungspunkte für die Hypothese H_0:π_1- π_2=δ lauten:

$$\delta_{p_r} = \delta_0 + k_r \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Da man π_1 und π_2 nicht kennt, wird π_1 und π_2=0 geschätzt. Der Fischer-Test ist hier der bessere Lösungsweg.

Voraussetzung ist allerdings, daß die Anteilswerte nicht zu nahe bei Null liegen dürfen
Falls die Stichprobe klein ist, greift man zum Fisher-Test (siehe S. 18)

2.5 FISHER-Test

2.5.1 Testsituation

- zwei unabhängige Stichproben mit zwei dichotomen Gesamtheiten
- die unbekannten Verteilungsfunktionen heißen: $F_1(X)$ und $F_2(X)$.

2.5.2 Testidee

Der Fishertest ist eine Sonderform des χ2-Test, insofern also auch ein Anpassungstest.
Mit dem Fisher-Test läßt sich eine Vierfelderverteilung prüfen, auch wenn die Stichprobe und damit die erwarteten und beobachteten Häufigkeiten klein sind.

Mit Hilfe der hypergeometrischen Verteilung, die für **diskrete** Variablen definiert ist, kann man exakt die Wahrscheinlichkeit bestimmen, mit der bei gegebenen univariaten Randverteilungen eine bestimmte Häufigkeitverteilung in den vier Feldern der Tabelle auftritt.

Geprüft werden soll: H_0: $F_1(X)=F_2(X)$ für alle x, d.h. die Verteilung der ersten Stichprobe entspricht der der zweiten Stichprobe, beide Stichproben stammen aus der gleichen Grundgesamtheit.
Eine Vierfeldertafel sei folgendermaßen notiert:

Ausprägung des Merkmals → Nr. der Stichprobe ↓	a	$\bar{a}$	Summe
1	x_1	n_1-x_1	n_1
2	x_2	n_2-x_2	n_2
Summe	x	n-x	n

Falls die Nullhypothese zutrifft, unterscheiden sich die Anteilswerte $p_1 = \frac{x_1}{n_1}, p_2 = \frac{x_2}{n_2}$ nur zufällig voneinander.
Gesucht ist die Wahrscheinlichkeit für $X_1 \mid x$ unter der Bedingung, daß H_0 in Wahrheit zutrifft.
Um den Unterschied zwischen p_1 und p_2 zu beurteilen, genügt es, sich mit x_1 (oder x_2) zu befassen, da die anderen Werte ($x=\pi_1 n=\pi_2 n=\pi n$ und n_1 und n_2) bekannt sind, bzw. sich errechnen lassen (→ Nur ein Feld in der 2x2-Tafel ist frei variierbar).
Man numeriert die Stichproben so um, daß $x_1 \leq$ die anderen drei Häufigkeiten ist!
Die Prüfgröße X_1 folgt bei korrekter Nullhypothese X_1 einer hypergeometrischen Verteilung mit:

$$P(x_1) = \frac{\binom{n_1}{n_2}\binom{n_2}{x-x_1}}{\binom{n}{x}} = \frac{n_1!n_2(n-x)!}{x_1x_2!(n_1-x_1)!(n_2-x_2)!n!}$$

Eigenschaften von $x1$:

- relativ größere Werte von x_1 führen dazu, daß die Differenz zwischen p_1 und p2 größer wird (relativ oberer Bereich).
- relativ große Werte von x_1 führen eher zur Ablehnung der H_o.

2.5.3 Beispiel

Während einer Industrieausstellung wurden 5 Vertreter inländischer Unternehmen und 6 Vertreter ausländischer Unternehmen über ihren Eindruck vom Geschäftsklima befragt:

Änderung des Geschäftsklimas → Unternehmen ↓	Verschlechterung	Verbesserun g	Summe
Inländisch	1	4	5
Ausländisch	5	1	6
Summe	6	5	11

- Geprüft werden soll: Sind die Unterschiede in den Antworten zufällig (H_o) oder nicht (H_a)?

  H_o (umformuliert): $p_1 = \frac{1}{5} = 0{,}2$ und $p_2 = \frac{5}{6} = 0{,}83$ unterscheiden sich nur zufällig voneinander.
- Gegeben sei ein 5%iges Signifikanzniveau.
- Man berechnet die Prüfgröße X_1: $P(X_1) = \frac{\binom{5}{x_1}\binom{6}{6-x_1}}{\binom{11}{6}} = \frac{5!6!6!5!}{x_1!x_2!(5-x_1)!(6-x_2)!11!}$
- Man prüft nun für x_1=1, ob der Rückweisungspunkt x_1r rechts oder links von x_1 liegt.
- P(1)=0,0649.
- Bei einem Signifikanzniveau von 5%, d.h. einem P(0,05), wird die Hypothese nicht verworfen, da 0,0649>0,05.
- Ergebnis: die Unterschiede in beiden Stichproben sind zufällig!

3 Tests bei ordinalem Meßniveau

Der χ^2-Anpassungstest ist prinzipiell auch für ordinalskalierte Daten anwendbar. Er berücksichtigt jedoch nicht die Größer-Kleiner-Relation, die zwischen den Stichproben besteht.

3.1 Mediantest (=Vorzeichentest)

Literatur: Tiede: S 74 ff

3.1.1 Testsituation

- gegeben sei ein stetiges Merkmal X, das mindestens ordinal skaliert ist
- es wird eine Zufallsstichprobe gezogen mit x_i, i=1 .. n Merkmalswerten
- $x_{0,5}$ ist der Median der Stichprobe, d.h. der Wert, der in der Mitte der nach Größe geordneten Werten liegt (→ Zentralwert)

3.1.2 Testidee

Die Hypothese lautet: der Median der Grundgesamtheit hat einen bestimmten Wert c.

Hypothese: H_0: $\gamma_{0,5}$=c (c ist irgendein Wert), $\gamma_{0,5}$ ist der Median der Grundgesamtheit

Lösungsansatz:
Man vergleicht jeden Stichprobenwert x_i mit dem Hypothesenwert c.

- Falls x_i >c ist, notiert man ein positives Vorzeichen, a_i=1
- Falls x_i <c ist, notiert man ein negatives Vorzeichen, a_i=0
- falls x_i=c ist, wird x_i entfernt, n verringert sich
  Werden viele Werte aussortiert und die H_0 trotzdem angenommen, spricht das Ergebnis <u>sehr</u> für die Annahme der H_0.

$$V = \sum_i^n A_i$$

Die Prüfvariable V (= „Anzahl der positiven Vorzeichen") ist die Summe aller a_i:

Falls die Nullhypothese zutrifft, folgt V der Binominalverteilung B(n, 0,5), d.h. bei n Vergleichen wird die Zahl v positiver Vorzeichen durch B(n,0,5) angegeben. Die Wahrscheinlichkeit eines positiven Vorzeichens ist nämlich bei zutreffender Nullhypothese bei jedem Vergleich zwischen x_i und c gleich 0,5.

Falls v „zu groß" oder „zu klein" ist, wird die Nullhypothese verworfen.
Ist v relativ klein oder groß, spricht das eher gegen die H_0.

Ist der Wertebereich von V eher in der Mitte, könnte die H_0 angenommen werden.

Die Rückweisungspunkte ergeben sich aus der Zahl der Elemente n. Man sieht in der Binominaltabelle unter der Anzahl der Elemente nach und liest die Rückweisungspunkte einfach ab.
Geprüft wird nun, ob der Stichprobenwert v im Annahmebereich liegt.

3.1.3 Bemerkungen

Der Mediantest wird häufig als Schnelltest vorgeschoben: Wenn die H_0 im Median-Test verworfen wird, gilt generell, daß ein besserer Test auch zur Verwerfung führt. Umgekehrt gilt bei einer Annahme der H_0 nicht unbedingt, daß auch der bessere Test zu diesem Ergebnis führen würde.
Ein Vorteil des Mediantestes ist es, daß er auch bei kleinen Stichproben angewendet werden kann. Der größte Nachteil besteht in der Tatsache, daß er das ordinale Meßniveau nicht nutzt, da er nur auf Größer/Kleiner-Relationen zu einem bestimmten Wert c (Wert des Medians) Informationen gibt → die Rangfolgen bei ordinalem Meßniveau werden nicht genutzt.

Liegt eine metrische Skalierung vor und ist die Grundgesamtheit symmetrisch verteilt, dann ist der Mediantest ein Konkurrent zum t-Test.

Den Mediantest kann man mit oder ohne störende Werte (Ausreißer?) durchführen.

3.1.4 Beispiel (Tiede S. 74 unten)

7 von 8 Werten einer Stichprobe liegen über dem Hypothesenwert der Grundgesamtheit, d.h. v (Anzahl der positiven Vorzeichen)=7.
Frage: Ist der Unterschied zwischen vermutetem Grundgesamtheitsmedian und Ergebnis

- zufällig (→ Annahme der H_0, daß sich Grundgesamtheits- und Stichprobenmedian nur zufällig unterscheiden)

oder

- signifikant (→ Ablehnung der H_0 → es existiert ein wirklicher Unterschied zwischen Grundgesamtheit und Stichprobe)

Gegeben sei ein 5%iges Signifikanzniveau!

Aus der Tabelle entnimmt man die **Rückweisungspunkte** für n=8, π=0,5 (V folgt ja B(8;0,5)): v_{r1}=0, v_{r2}=8.

Der Stichprobenbefund von v=7 liegt im Annahmebereich, das heißt, der Unterschied zwischen Grundgesamtheit und Stichprobe ist nur zufällig aufgetreten.
Ist die Schlußfolgerung richtig?

3.2 WILCOXON-Vorzeichen-Rangtest für den Median

3.2.1 Testsituation

- Gegeben ist eine einfache Stichprobe.
- Das Untersuchungsmerkmal X ist in der Grundgesamtheit symmetrisch verteilt.

3.2.2 Testidee

Im Gegensatz zum Mediantest berücksichtigt der Wilcoxon-Vorzeichen-Rangtest auch die Information über die in Rängen ausgedrückte Größe des Unterschieds zwischen jedem Stichprobenwert und dem Hypothesenwert.

Geprüft werden soll: Der Grundgesamtheitsmedian der stetigen Untersuchungsvariablen hat einen bestimmten Wert c.
Die Hypothese lautet also: $H_0 = \mu_{0,5} = c$
Jeder Stichprobenwert x_i wird mit dem Hypothesenwert c verglichen.

1. Man ordnet die berechneten Differenzen $|di = x_i - c|$ der absoluten Größe nach, zunächst <u>ohne</u> Beachtung des Vorzeichens.
2. $|di|$ Rangplätze werden vergeben: der kleinste Wert $|di|$ erhält den Rang 1, der zweitkleinste den Rang 2. (Sind einige $|di|$ gleich groß, bekommen sie ihren Durchschnittsrang zugewiesen)
3. Man ordnet den Rangzahlen die entsprechenden d_i zu.
4. Man summiert die Rangzahlen getrennt für positive und negative Vorzeichen.
5. Falls die Summe der Ränge für negative oder für positive Differenzen „zu klein“ oder „zu groß“ ist, wird die Hypothese verworfen (siehe Mediantest, S.19).

! Bei einem Wilcoxon-Test bezieht man sich stets auf die **absolut kleinere**, nicht die größere Summe der Ränge! (Vereinbarung unter dem Volk der Statistiker)

Die Prüfgröße für den Wilcoxon-Test W ist formal definiert als:

$W = \sum A_i Rg(|d_i|)$, a_i=0, falls d_i negativ ist, a_i=1, falls d_i positiv ist (oder umgekehrt).

3.2.3 Wertebereich und Verteilung von W

1. W kann 0 sein, wenn in der Stichprobe alle Werte kleiner als der Median sind.
2. W nimmt den größtmöglichen Wert ein, wenn alle Werte größer als der Median sind. Der maximale Wert ergibt sich aus $w_{\max} = \dfrac{n(n+1)}{2}$
3. W folgt einer diskreten Verteilung.
4. W ist symmetrisch.

3.2.4 Bemerkungen

1. Der Wilcoxon-Vorzeigen-Rangtest ist effizienter als der Mediantest, da er auch die Informationen über die Ränge verwendet
2. Bei großem Stichprobenumfang (n>25) und $\alpha \geq 0{,}05$ läßt sich W durch die folgende Normalverteilung approximieren: $N\left(\dfrac{1}{4}n(n+1); \sqrt{\dfrac{1}{24}n(n+1)(2n+1)}\right)$
3. Der Wilcoxon-Test setzt Symmetrie in der Grundgesamtheit voraus, die Prüfung der Symmetrie wird in der Praxis oft vernachlässigt.

4. Zwei Hypothesen können getestet werden:
   a) H_o: der Median der Grundgesamtheit hat einen bestimmten Wert c.
   b) H_o: Die Grundgesamtheit ist symmetrisch.
5. Die Rückweisungspunkte in der Formelsammlung auf S. 41 sind **linksseitig** tabelliert.
6. **Problem des mehrfachen Auftretens von Testwerten:** Wenn beim Mediantest mehrere Testwerte gleichgroß sind, ist das kein Problem, da allein das Faktum der Abweichung vom Median beim Median zählt. Beim Wilcoxon-Test wird durch das Zuweisen von Durchschnittsrängen das Vorliegen der sogenannten „Bindungen" überwunden. Das stellt ein Problem dar, das jedoch ab n>10 vernachlässigbar ist.
7. Der Wilcoxon-Test gehört mit einer Effizienz von 95 Prozent zu den trennschärfsten parameterfreien Verfahren.

3.2.5 Beispiel (Tiede S. 77)

Ein Lehrer hatte in der Vergangenheit die Durchschnittsnote 3 vergeben. Seiner derzeitigen Schulklasse gab er Zensuren in der folgenden Häufigkeit:

Note	1	2	3	4	5
Häufigkeit	0	6	6	4	6

Geprüft werden soll, ob sich die Schulklasse von anderen Schulklassen im Notendurchschnitt unterscheidet. (5% Signifikanzniveau, beidseitig): H_o: $\mu_{0,5}=3$

Berechnung nach dem Mediantest:

Zensur	d_i	Rang für $\lvert d_i \rvert$
2	-1	5,5
2	-1	5,5
2	-1	5,5
2	-1	5,5
2	-1	5,5
2	-1	5,5
4	1	5,5
4	1	5,5
4	1	5,5
4	1	5,5
5	2	13,5
5	2	13,5
5	2	13,5
5	2	13,5
5	2	13,5
5	2	13,5

Nach dem Mediantest ergibt sich: für n=16 (die 6 Schüler, die eine drei erreicht haben, fallen raus): v = 10 positive Vorzeichen.
Die Rückweisungspunkte $r_1=3$ und $r_2=13$ liest man in der Formelsammlung auf S. 40 ab (n=15, α=5%).

Da $P(V\leq3)\leq0{,}025$, $P(V\geq13)\leq0{,}025$ wird die Hypothese H_o: $\mu_{0,5}=3$ angenommen.

Berechnung nach dem Wilcoxon-Vorzeichen-Rangtest
Die Summe der „negativen Ränge" beträgt w=33, d.h. der Stichprobenwert ist 33.
Der Rückweisungspunkt w_r ist in diesem Beispiel bei einem beidseitigem Signifikanzniveau von 5 Prozent $w_r=30$. (Formelsammlung; S. 41)

Da also $w>w_r$, ist die Hypothese nicht zu verwerfen, der Unterschied ist zufällig.

3.2.6 Wilcoxon-Vorzeichen-Rangtest für verbundene Paare

3.2.6.1 Testsituation

- Das Merkmal X wird in zwei verbundenen (abhängigen) Stichproben erhoben
- Es werden die Differenzen zwischen Wert *i* der ersten Stichprobe und Wert *i* der zweiten Stichprobe berechnet:
  $d_i=x_{1i}-x_{2i}$, *1…n*
  - die Differenzen sind unabhängig voneinander
  - die Differenzen sind symmetrisch verteilt
  - ${}_1x_{0,5}$ ist der Median der Stichprobe Nr.1
  - ${}_2x_{0,5}$ ist der Median der Stichprobe Nr. 2

3.2.6.2 Testidee

Der Wilcoxon-Vorzeichen-Rangtest für verbundene Paare nutzt die gegebenen Informationen über die Differenz eines Meßwertpaares aus:

1. Die Differenz hat ein Vorzeichen.
2. Die Differenz hat einen Betrag.

Getestet wird, ob zwei verbundene Stichproben aus einer Gesamtheit stammen.
Das heißt, getestet wird auch, b sich die beiden Stichprobenmediane signifikant voneinander unterscheiden:

$$H_0 : {}_a\mu_{0,5} - {}_b\mu_{0,5} = c; \quad {}_a\mu_{0,5} - {}_b\mu_{0,5}$$

Wie beim Wilcoxon-Vorzeichen-Rangtest wird die Prüfvariable $W = \sum_{i=1}^{n} A_i \cdot Rg(|d_i|)$ zugrunde gelegt. a_i ist 0, falls $(x_{1i} - x_{2i})>c$

Mediantest	verbundene Paare
$\mu_{0,5}$	${}_1\mu_{0,5} - {}_2\mu_{0,5}$
$x_{0,5}$	${}_1x_{0,5} - {}_2x_{0,5}$
$\lvert d_i \rvert = \lvert x_i - c \rvert$	$\lvert d_i \rvert = \lvert (x_{1i} - x_{2i}) - c \rvert$

3.2.6.3 Beispiel (Tiede S. 97)

Eine neue winterharte Weizensorte (Sorte 2) wird mit einer gängigen Standardsorte bezüglich des Ernteertrags verglichen. Man baut jeweils beide Sorten in etwa gleichen, ansonsten zufällig gewählten Standorten an und ermittelt die Erträge in kg pro Flächeneinheit. Es werden Ränge verteilt, die folgende Tabelle ergibt sich:

d_i	-4	28	-5	20	17	3	9	13
Rg ($\lvert d_i \rvert$)	2	8	3	7	6	1	4	5
a_i	1	0	1	0	0	0	0	0

Der Wilcoxon-Test prüft hier nicht nur die Frage des Mittelwertunterschiedes, sondern prüft auch, ob beide Stichproben aus der gleichen Grundgesamtheit stammen könnten.
Es gibt sich w=5. Dieser Wert liegt bei 5% Signifikanzniveau im Annahmebereich der Nullhypothese, wie auch beim Vorzeichentest. Der Durchschnittsertrag der Neuzüchtung ist nicht anders als der der Standardsorte. Der beobachtete geringere Durchschnittsertrag hat sich aus zufälligen Gründen ergeben.

3.3 Kolmogoroff/Smirnov-Test (Ein-Stichproben-Anpassungstest)

Der Kolmogoroff/Smirnov-Test verwendet im Gegensatz zum χ^2-Anpassungstest auch die Größer-Kleiner-Relation, die zwischen den Stichprobenwerten besteht.

3.3.1 Testsituation

- eine einfache Stichprobe, bei ordinalem Meßnivau
- alle Paramter und die Verteilung müssen explizit spezifiziert sein und nicht aus der Stichrpobe heraus geschätzt werden

3.3.2 Testidee

Der Kolmogoroff/Smirnov-Test ist für zwei Problemfälle anwendbar, er beantwortet die folgenden Fragen:

1. Stammt eine Stichprobe aus einer in einer bestimmten Weise verteilten Grundgesamtheit? (Kolmogoroff/Smirnov-Einstichproben-Anpassungstest, s.u.)
2. Stammen zwei unabhängige Stichproben aus einer Grundgesamtheit? (siehe Kolmogoroff/Smirnov-Zweistichproben-Anpassungstest, Tiede S. 104ff)

Getestet werden soll, ob die vorliegende (empirische) Verteilung einigermaßen einer theoretischen Verteilung entspricht. Der Kolmogoroff/Smirnov-Test benutzt dabei nicht die einfache Häufigkeitsverteilung (wie beim Chiquadrat-Test).

Bei diesem Testverfahren wird die **kumulierte Häufigkeitsverteilung** mit der entsprechenden theoretischen Verteilungsfunktion verglichen. Die theoretische Verteilungsfunktion entspricht bei **stetigen** Variablen der kumulierten Häufigkeit.

Die Nullhypothese lautet: H_o: $F(x)=F_o(x)$, die Alternativhypothese lautet: H_a: $F(x)\neq F_o(x)$
Die Prüfvariable für den Kolmogoroff/Smirnov-Test lautet:

$$D = \max_{-\infty<x<\infty} \left|F_B(X) - F_e(X)\right|$$

D.h. man berechnet zusätzlich zu den vorliegenden tatsächlichen Wahrscheinlichkeiten $F_B(X)$die erwarteten theoretischen Wahrscheinlichkeiten $F_e(X)$. Die maximale Differenz d_{max} bildet den Stichprobenwert. Den Rückweisungspunkt d_r liest man für α% (zweiseitiges Signifikanzniveau) aus der Tabelle in der Formelsammlung S. 42 ab.
Ist $d_{max}\geq d_r$ wird die Hypothese abgelehnt.

3.3.3 Bemerkungen

1. Die Verteilung von *D* ist nur bei metrisch skalierten Untersuchungsvariablen exakt, man kann den Test aber auch bei klassifizierten (gruppierten) Meßreihen anwenden, wenn gilt: *n*>20 und *r* (→Anzahl der Klassen) >5
2. Will man eine spezifizierte Verteilungshypothese (siehe S. **Fehler! Textmarke nicht definiert.**) testen, sollte man den Test nur mit Vorsicht verwenden.

3.3.4 Beispiel (Graff, S. 34)

Die Altersverteilung von 50 Studentinnen zwischen 21 und 30 Jahren wurde gemessen. Die Hypothese lautet, daß das Alter in dem gemessenen Bereich von 21 bis 30 Jahren gleichverteilt sein soll. Die Prüfvariable *D* mißt die Differenz der kumulierten relativen Häufigkeiten zwischen empirischer und theoretischer Verteilung.

	Empirisch			theoretisch			
X	f	relativ	kumuliert	f	relativ	kumuliert	D
21	6	0,12	0,12	5	0,10	0,10	0,02
22	7	0,14	0,26	5	0,10	0,20	0,04
23	9	0,18	0,55	5	0,10	0,30	0,14
24	8	0,16	0,60	5	0,10	0,40	0,20
25	6	0,12	0,72	5	0,10	0,50	0,22
26	6	0,12	0,84	5	0,10	0,60	0,24
27	3	0,06	0,90	5	0,10	0,70	0,20
28	2	0,04	0,94	5	0,10	0,80	0,14
29	1	0,02	0,96	5	0,10	0,90	0,06
30	2	0,04	1,0	5	0,10	1,00	0,00
Summe	50	1,0		50	1,00		

d_{max} steht in der letzten Spalte (→ der größte Differenzwert): d_{max}=0,24.
Bei einem zweiseitigem Signifikanzniveau von α = 1 % schlägt man nun für n=50 in der Tabelle in der Formelsammlung auf S. 42 den Rückweisungspunkt d_r nach.
Dabei nutzt man aus, daß für große n approximiert werden kann nach: $\frac{1,52}{\sqrt{n}}$. Das angegebene Signifikanzniveau ist dann zwar 2 %, aber der Unterschied ist akzeptabel. Es ergibt sich für
dr = 0,2164.
Ist $d_{max}\geq d_r$ wird die Hypothese abgelehnt, hier ist 0,24≥0,2164, d.h. die Nullhypthese wird abgelehnt werden. Das Alter von den Studentinnen ist nicht gleich verteilt.

3.4 Kolmogoroff/Smirnov-Test (Zwei-Stichproben-Anpassungstest)

3.4.1 Testsituation

- zwei (u.U.) verschieden große unabhängige Stichproben
- mindestens metrisches Meßniveau (ordinales Meßniveau ist ok)
- stetige Untersuchungsvariable

3.4.2 Testidee

Stammen die beiden Stichproben aus GG, die die gleiche Verteilung aufweisen?

$H_0: F_1(x) = F_2(x)$ für alle x

$H_a: F_1(x) \neq F_2(x)$ [bei zweiseitigem Test]

3.4.3 Testdurchführung

Die Prüfvariable wird analog zum Kolmogoroff/Smirnov-Test (Ein-Stichproben-Anpassungstest, s. S.23) berechnet: $d = \max_{-\infty<x<\infty} |F_{1b}(x) - F_{2b}(x)|$. *d* ist Realisation der Stichprobenvariablen *D.*

3.4.4 Beispiel

Zwei Leichtathletikgruppen mit fünf bzw. sechs Leuten (n_1=5, n_2=6) machen einen Fitnesstest, nachdem sie ein unterschiedliches Wintertraining durchgeführt haben. Jeder Athlet kann bis zu 20 Punkte erreichen. (Dieses Beispiel ist eigentlich kein korrekter Anwendungsfall für den KS-2-Stichprobenanpassungsfall, da gegen die Faustregeln $n_1+n_2>35$ verstoßen wurde. Da er aber rechnerisch sehr aufwendig ist, habe ich dieses Beispiel mit kleinen *n* gewählt).

x_{1j}	5	10	15	17	12	
x_{2j}	6	6	7	9	9	13

H_0: Stammen die beiden Stichproben aus einer Gesamtheit? $F_1(x)=F_2(X)$
Ich berechne die kummulierten Einzelwahrscheinlichkeiten. Zunächst sortiere ich alle Werte in einer gemeinsamen Stichprobe der Größe nach, wobei doppelt vorkommende Werte herausgenommen werden.

x_{1j}	5	10	15	17	18	
x_{2j}	6	6	7	9	9	13

gesamt	5	6	7	9	10	13	15	17	18
Fb1	0.2	0.2	0.2	0.2	0.4	0.4	0.5	0.9	1
Fb2	0	0.33	0.5	0.83	0.83	1	1	1	1
d	0.2	0.1	0.3	0.63	0.43	0.6	0.4	0.4	0

<u>Werte für Fb1</u>

- Für jeden Wert der „gesamt"-Zeile schaue ich zunächst, ob dieser Wert aus der ersten Stichprobe stammt. Wenn er aus dieser Stichprobe stammt, zähle ich, wie oft er in der Stichprobe vorkommt. Diese Zahl teile ich durch die Anzahl der Elemente der Stichprobe1 (hier=5). Diesen Wert addiere ich zum vorherigen Wert, der sich für die vorherige Zahl ergab. Wenn ein Wert nicht aus der Stichprobe 1 stammt, ignoriere ich ihn und übernehme nur das Ergebnis aus der Zelle der vorherigen Zahl. (Bsp: 10 stammt aus Stichprobe 1, kommt einmal vor, also hat 10 den Anteilswert von 1/5=0.2 plus dem vorherigen Wert von 0.2=0.4. 7,9,6 würde ich einfach ignorieren

<u>Werte für Fb2</u>

- Die Werte für FB2 entstehen analog zum FB1.

<u>Absolute Differenzen</u>

- Jetzt berechnet man die absoluten Differenzen:
  z.B. 0.2-0.33=-0.13; Abs(-0.13)=0.13≈0.1
- d_{max} ist die maximale Differenz, hier 0.63

Rückweisungspunkt

- der Rückweisungspunkt $d_r = c(\alpha)\sqrt{\frac{n_1+n_2}{n_1 n_2}}$ (für $c(\alpha)$ siehe Formelsammlung)

$$d_r = 1.22 \cdot \sqrt{\frac{5+6}{5*7}} = 0{,}74$$

Da die maximale Differenz d_{max}=0,63 kleiner ist als der Rückweisungspunkt, wird die Nullhypothese angenommen: Es gibt keine signifikanten Unterschiede zwischen den Gruppen.

3.5 Mann-Whitney-U-Test für zwei unabhängige Stichproben

Der Mann/Whitney-U-Test gehört zu den stärksten Tests, um einen Zusammenhang zwischen einer zweistufigen nominalskalierten Variablen und einer ordinalskalierten Variablen zu testen. Man kann aber auch den Unterschied zwischen zwei unabhängigen Stichproben hinsichtlich der ordinal skalierten Variablen testen.

3.5.1 Testsituation

- zwei unabhängige Stichproben, wobei $n_1 \leq n_2$ sein *kann*
- stetige Untersuchungsvariable X
- die Verteilung in beiden Gesamtheiten ist bis auf den Lageparamter λ gleich, λ ist eine Konstante: $F_1(x) = F_2(x+\lambda)$

3.5.2 Testidee

Geprüft wird, ob ein Zusammenhang zwischen zwei unabhängigen Stichprobenverteilungen existiert, genau genommen, ob die Verteilungen fon $F_1(x)$ und $F_2(X)$ bis auf eine Konstante gleich sind.
Die **Hypothese** lautet gewöhnlich: H_0:λ=0, d.h. der Parameter ist 0 → es besteht ein Zusammenhang zwischen den beiden Stichprobenverteilungen.
Die **Alternativhypothese** lautet:

  - bei einem zweiseitigen Test: $\lambda \neq 0$
  - bei einem einseitigem Test: λ>0 oder λ<0

- Vereinbart wird: $n_1 \leq n_2$, man numieriert die Stichproben entsprechend
- Aus den Werten der beiden Stichproben bildet man nun **eine gemeinsame Reihe** in aufsteigender Folge.
- Man vergleicht jeden Wert der einen Stichprobe (x_{1i}=1...n) mit dem Wert der anderen Stichprobe (x_{2j}, j=1...n_2) ($n_1 n_2$-Vergleiche)

3.5.3 Die Konstruktion der Prüfvariablen

- Man bildet nun die Indikatorvariable U mit $U = \sum_{i}^{n_1}\sum_{j}^{n_2} A_{ij}$.

  A) Falls der Wert der ersten Stichprobe > als der Wert der zweiten Stichprobe ist, wird a_{ij}=1 gesetzt: Falls x_{ij}>x_{2j} → a_{ij}=1
  B) Falls der Wert der ersten Stichprobe < als der Wert der zweiten Stichprobe ist, wird a_{ij}=0 gesetzt:
  Falls x_{ij}<x_{2j} → a_{ij}=0
- Falls die obige (Homogenitäts-)Hypothese zutrifft, muß die Anzahl der Fälle von A so groß sein wie die Anzahl der Fälle von B - abgesehen von zufälligen Abweichungen.

3.5.4 Vorgehensweise

- Die Hypothese läßt sich umschreiben in: H_0:E(U):0,5$n_1 n_2$, d.h. die erwartete Häufigkeit der Prüfvariable U ist so groß wie die Hälfte der miteinander multiplizierten Stichprobenanzahlwerte.
- Der **Rückweisungspunkt** u_r sind in der Formelsammlung auf S. 43 tabelliert.

- Um den Stichprobenbefund u (Prüfwert) zu ermitteln, nutzt man die folgende Beziehung zwischen u und w aus: $u=w-0{,}5n_1(n_1+1)$

  → w ist vereinbarungsgemäß die Summe der Rangzahlen in der **kleineren Stichprobe**.
- weisen beide Stichproben zum Teil gleiche Werte auf, werden die Werte entfernt und n verringert (siehe auch Mediantest (=Vorzeichentest), S. 19).

3.5.5 Bemerkungen

1. Wenn man einen großen Stichprobenumfang vorliegen hat ($n \geq 20$), kann man U asymptotisch approximieren nach: $N\left(0{,}5n_1n_2;\sqrt{\frac{n_1n_2(n+1)}{12}}\right)$
2. Falls $u>0{,}5n_1n_2$ ist, nutzt man die Komplementärvariable von U U* aus. (Es gilt: $U=n_1n_2-U^*$). Man berechnet den transformierten Wert $u= n_1n_2-u^*$.
3. Ab $n_2 \geq 20$ darf man als Näherung die Normalverteilung benutzen, wobei man die Kontinuitätsberichtigung nach YATES berücksichtigen muß.
4. Der Mann/Whitney-U-Test hat wegen der geringeren Voraussetzungen bei Anwendungen eine relativ große Bedeutung, Er ist nur geringfügig schlechter (95 Prozent Effizienz) als der vergleichbare t-Differenzentest für zwei Mittelwerte.

3.5.6 Beispiel (Tiede S. 93)

Zwei Leichtathletikgruppen mit fünf bzw. sechs Leuten ($n_1=5$, $n_2=6$) machen einen Fitnesstest, nachdem sie ein unterschiedliches Wintertraining durchgeführt haben. Jeder Athlet kann bis zu 20 Punkte erreichen.

x_{1j}	5	10	15	17	12	
x_{2j}	6	6	7	9	9	13

Geprüft werden soll: Unterscheiden sich die Gruppenmittel von 15 und 8 nur zufällig voneinander? (5-Prozentiges Signifikanzniveau einseitig). Mathematisch gesprochen prüft man, ob beide Stichproben aus einer Gesamtheit stammen.

Die Hypothese lautet also : H_0: $\lambda=0$ (→ es gibt keinen signifikanten Unterschied zwischen Gruppe 1 und Gruppe 2), anders formuliert (s.o) H_0:E(U):$0{,}5n_1n_2$

1. Die beiden Gruppen werden zu einer Gruppe zusammengefaßt. Jeder Wert aus der ersten Stichprobe wird zur Kennzeichnung mit einer „0“ markiert, die Werte aus der zweiten Stichprobe bekommen eine „1“.:

Gesamtgruppe	5	10	15	17	12	6	6	7	9	9	13
stammt aus Gruppe:	0	0	0	0	0	1	1	1	1	1	1

Die Gesamtstichprobe wird der Übersicht halber sortiert, und es werden Ränge verteilt (Doppeltränge beachten!)

Gesamtgruppe	5	6	6	7	9	9	10	12	13	15	17
Kodierung	0	1	1	1	1	1	0	0	1	0	0
Ränge	1	2,5	2,5	4	5,5	5,5	7	8	9	10	11

- Man verteilt nun Rangzahlen (der schlechteste Athlet bekommt den Rang 1 usw):

x_{1j}	5	10	15	17	12	6	6	7	9	9	13
x_{2j}	1	7	9	10	11	2,5	2,5	4,	5,5	5,5	8

- Jetzt berechnet man *w*, die Summe der Rangzahlen der kleineren Stichprobe (sind beide gleich groß, nimmt man die erste Stichprobe): w=37
- Die Prüfvariable u berechnet sich aus $u = w - 0{,}5n_1(n_1+1) =$ *37-0,5•5•6=37-15=22*
- Die Verteilung für den Mann-Whitney-Test ist nur linksseitig definiert. Deshalb muß man nun prüfen, ob der Prüfwert u in die transformierte Variable u* umgeformt werden muß, die tabelliert ist:

- Geprüft werden muß, ob $u>0{,}5n_1n_2$ gilt. Hier gilt 22>15, d.h. man muß u* berechnen: $u^*=n_1n_2 - u=30-22=8$.
- Aus der Formelsammlung (S. 43) liest man für $n_1=5$ und $n_2=6$ den Wert für den Rückweisungspunkt $u^*_r=5$ ab.
- Da der Rückweisungspunkt $u_r^*=5$ rechts von Prüfwert 8 (5<8) liegt, kann die Nullhypothese angenommen werden: Es gibt keinen signifikanten Unterschied zwischen den beiden Gruppen.

3.6 Kruskal/Wallis-H-Test

Der Kruskal/Wallis-H-Test ist eine Verallgemeinerung des U-Tests von Mann/Whitney. Er verwertet zusätzlich Informationen über die rangmäßige Abstufung der Stichprobenwerte.

3.6.1 Testsituation

- zwei unabhängige Stichproben der Umfänge n, i=1…,r
- stetige Untersuchungsvariable X mit x_{ij}, i=1…r, j=1…n_i

3.6.2 Testidee

Es wird geprüft, ob r unabhängige Stichproben, in denen eine stetige Untersuchungsvariable betrachtet wird, aus r Grundgesamtheiten mit gleichem Median stammen. Vereinfacht gesagt prüft man: stammen die r Stichproben aus der gleichen oder identischen Grundgesamtheit. Die Alternativhypothese behauptet: nicht alle Gesamtheiten haben den gleichen Median, anders formuliert: die r Stichproben stammen aus verschiedenen Grundgesamtheiten, wobei sich diese Verschiedenheit auf die Lageparameter (Mittelwerte) bezieht.

1. Den Stichprobenwerten x_{ij} werden Rangzahlen $Rg(x_{ij})$ zugeordnet, so daß für die zusammengefaßte Stichprobe eine **Rangreihe** entsteht.
2. → der kleinste Wert x_{ij} bekommt den Rang 1
3. man bildet in jeder Stichprobe die entsprechende Rangsumme $rs_i = \sum(x_{ij})$
4. Der Mittelwert der Rangsumme lautet: $\overline{r_{si}} = \frac{r_{si}}{n_i}$ (→ Man dividiert für jede Stichprobe diese Summe durch die entsprechende Anzahl Elemente)
5. Falls die Nullypothese zutrifft, wird man für $\overline{r_{si}}$ nur zufällige Abweichungen ewarten.
6. Der **Prüfwert** ergibt sich aus der Formelvon Kruskal/Wallis:
$$H = \frac{12}{n(n+1)} \sum_{i=1}^{r} \frac{RS_1^2}{n_1} - 3 \cdot (n+1)$$

3.6.3 Bemerkungen

1. H kann für ni≥5 und r≥4 (Faustregel) einer χ^2-Verteilung mit r-1 Freiheitsgraden angenähert werden.

3.6.4 Beispiel (Tiede S. 99)

Für eine geplante Fragebogenation wird über zufällig ausgewählte Personen ausprobiert, wie zeitraubend die Bearbeitung einer Reihe von Fragebogentypen ist. Man erhält die folgenden Zeitwerte in Minuten:

Fragebogen-typ	Rangzahlen des Zeitaufwands										
1	11	14	16	19	27	29	30	31	40	41	42
2	15	17	18	25	26	32	33	33	42	43	
3	20	21	30	32	35	38	45	50	55	56	
4	8	10	12	13	22	28	28	28	32	33	

1. Man ordnet den Ursprungswerten Rangzahlen zu und erhält folgende Tabelle:

Fragebogen-typ	Rangzahlen des Zeitaufwands											Rang-summen
1	3	6	8	11	17	21	22,5	24	33	34	35,5	215
2	7	9	10	15	16	26	29	29	35,5	37		213,5
3	12	13	22,5	26	31	32	38	39	40	41		294,5
4	1	2	4	5	14	19	19	19	26	29		138
												861

2. Man berechnet H nach der obigen Formel und erhält: h=8,76.
3. Da r≥4 und ni≥5 ist, folgt H hinreichend genau einer χ^2-Verteilung mit (4-1)=3 Freiheitsgeraden.
4. Man liest aus der χ^2-Tabelle in der Formelsammlung auf Seite 30 den Rückweisungspunkt χ^2_r = 7,81 ab (→ 5 Prozent Signifikanzniveau).
5. 8,76>7,81, d.h. die Nullhypothese, daß die Unterschiede zwischen den Fragebögen nicht zufällig sind, wird abgelehnt: Es existieren Unterschiede bei dem Zeitaufwand für die Fragebögen.

Der Vorzeichentest (→ Schnelltest) kommt beim gleichen Beispiel zu einem anderen Ergebnis, der Kruskal-Wallis-Test hat eine Effizienz von 95%, seine Entscheidung gilt im Zweifelsfalle mehr.

3.7 Prüfung des Rangkorrelationskoeffizienten

Der SPEARMANsche Rangkorrelationskoeffizient dient zur Beurteilung des Zusammenhangs zweier stetiger Untersuchungsvariablen:

$$r_{sp} = 1 - \frac{6\sum_{i-1}^{n} d_i^2}{n^3 - n}$$ wobei i: d_i =Rg(y_i) ist

3.7.1 Testsituation

- X,Y, ordinal skaliert, n statistische Einheiten
- Gemessen wird das Wertepaar x_i, y_i.
- Man ordnet den Werten Rangzahlen zu, der kleinste Wert bekommt die Rangzahl 1.

3.7.2 Testidee

Geprüft wird, ob die Merkmale von X und Y in der Grundgesamtheit nicht korreliert sind, d.h. die Hypothese lautet: in der Grundgesamtheit hat der Rangkorrelationskoeffizient den Wert 0. Anders formuliert: X und Y sind statistisch unabhängig voneinander

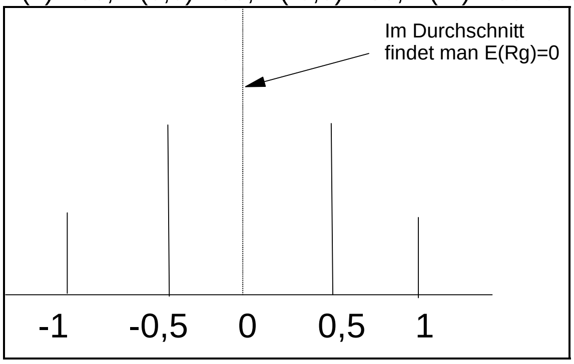
H_o: $\}_{Sp}=0$, $H_a=\}_{sp}\neq 0$

Nun wird es in der Stichprobe unter Umständen immer so sein, daß die gezogene Stichprobe einen Rangkorrelationskoeffizienten hat, der nicht Null ist. Mathematisch formuliert lautet das Problem:
δsp↔Ysp, rsp≠0
Lösungsidee: Man wählt eine geeignete Prüfvariable, hier den Rangkorrelationskoeffizienten selbt.
Um die Stichprobenverteilung zu ermitteln, ordnete man die Rangzahlen Rg(x_i) 1 2 3 ... n
Falls x und y unkorreliert sind, ist jede Beliebige Anordnung für Rg(yi) möglich, auch eine vollständige geordnete (absteigend/aufsteigend), die zur vollständigen Korrelation führen würde.
Bei n verschiedenen Elementen gibt es n! gleichmögliche verschiedene Möglichkeiten einer Ordnung für Rg(yi).

Beispiel: für n=3 ergeben sich folgende Kombinationsmöglichkeiten für $Rg(y_i)$, sowie die folgenden Rangkorrelationskoeffizienten r_{sp}. Daraus ergibt sich die Wahrscheinlichkeitsverteilung

$Rg(x_i)$	1	2	3	r_{sp}
$Rg(y_i)$	1	2	3	1
	1	3	2	0,5
	2	1	3	0,5
	2	3	1	-0,5
	3	1	2	0,5
	3	2	1	1

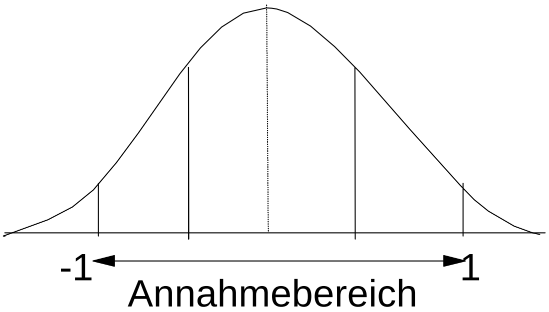
P(1)=1/6, P(0,5)=1/3, P(-0,5)=1/3, P(-1)=1/6. Graphisch ergibt sich folgendes Bild:



Die Stichprobenverteilung R_{sp} ist also diskret und symmetrisch zum Wert E(Rg)=0. Ihre Varianz ist VAR (R_{sp})= $\frac{1}{n-1}$. Die Rückweisungspunkte in Abhängigkeit von n und dem Signifikanzniveau sind in der Formelsammlung auf Seite 46 für n<11 tablliert. Ist n größer als 9, kann man approximativ die t-verteilte Prüfvariable

$t = R_{SP}\sqrt{\frac{n-2}{1-R_{SP}^{\ 2}}}$ mit v=n-2 Freiheitsgraden verwenden.

Ab n>20 kann $R_{SP}\sqrt{n-1}$ durch die Standardnormalverteilung N(0,1) approximiert werden.
Annahmebereich der Testvariablen



3.8 Literatur

Tiede, Manfred; Voß, Werner	Prüfverfahren in der Wirtschafts und Sozialstatistik	Studienverlag Brockmeyer, 1982	
Billeter, Ernst P.	Grundlagen der erforschenden Statistik	Springer-Verlag	UB AXB226
Graff, Jörg	Nichtparametrische Statistik in den Sozialwissenschaften	Centataurus-Verlagsgesellschaft	UB CSA2997