

p-Wert

Der **p-Wert** ist die Wahrscheinlichkeit, dass die Testgröße

- den beobachteten Wert oder einen noch extremeren Wert („weiter weg von H_0 “) annimmt
- unter der Bedingung, dass H_0 wahr ist.

Bemerkungen

- 1 Für die Berechnung der p -Werte benötigt man eine Statistik-Software oder Tabellen (**nicht** Thema dieser Veranstaltung).
- 2 Viele Statistik-Programme geben als Ergebnis eines statistischen Tests nur den p -Wert aus.

p-Wert und Signifikanzniveau

Die Nullhypothese wird genau dann abgelehnt, wenn der p -Wert kleiner oder gleich α ist.

p-Wert: Interpretation

- Wahrscheinlichkeit betrifft das Auftreten der Daten und nicht die Wahrscheinlichkeit von H_0
- p-Wert ist **kein Maß** für die Stärke des Effekts. Daher sollten Begriffe wie "hochsignifikant" eher vermieden werden.
- Angabe des p-Wertes immer mit Schätzung des Effekts und Konfidenzintervall



p-Wert: Ein- und zweiseitige Test

In vielen Fällen halbiert sich der p-Wert bei einseitiger Fragestellung.

Beispiel. Binomialtest 10 Experimente T Anzahl Treffer.

Ergebnis: $T = 9$

- zweiseitiger Test: $H_0 : \pi = 0.5$ gegen $H_1 : \pi \neq 0.5$.

Extrem oder gleich extreme Werte von T : 9, 10, 1, 0

p-Wert =

$$P_{H_0}(T \in \{9, 10, 0, 1\}) = 0.01 + 0.001 + 0.01 + 0.001 = 0.022$$

- einseitiger Test: $H_0 : \pi \leq 0.5$ gegen $H_1 : \pi > 0.5$

Extrem oder gleich extreme Werte von T : 9, 10

$$\text{p-Wert} = P_{H_0}(T \in \{9, 10\}) = 0.11$$

Test auf den Erwartungswert

- Wir interessieren uns für den Erwartungswert μ einer metrischen Zufallsgröße.
Beispiele: Alter, Einkommen, Körpergröße, Scorewert ...
- Wir können einseitige oder zweiseitige Hypothesen formulieren.
- Beispiele
 - "Vor 10 Jahren betrug die Durchschnittsgröße von Studienanfängern und -anfängerinnen 167 cm. Heute sind sie im Schnitt größer als 167 cm."
 - Die Differenz zwischen Gewicht vor und nach einer Diät ist 0.



Zweiseitiger Gauss-Test auf den Erwartungswert μ

- X Zufallsgröße mit Erwartungswert μ .
- Zweiseitige Hypothese über μ :

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

- Testgröße: Mittelwert in der Stichprobe X_1, \dots, X_n :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

- Stichprobenumfang n ist genügend groß (Faustregel $n > 30$)



Zweiseitiger Gauss-Test auf den Erwartungswert μ

Testentscheidung zum Signifikanzniveau α

Annahmebereich

$$\mu_0 \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}}$$

mit

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

H_0 wird abgelehnt, falls

$$\bar{X} < \mu_0 - z_{1-\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}} \quad \text{oder} \quad \bar{X} > \mu_0 + z_{1-\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}}$$

$z_{1-\frac{\alpha}{2}}$ ist das $(1 - \alpha/2)$ -Quantil der Standardnormalverteilung.

Beispiel: Verändert sich der Blutdruck nach einer Intervention

Beispiel zweiseitiger Test

- Nullhypothese: Die Blutdruckdifferenz ist 0.

$$H_0 : \mu = 0$$

$$H_1 : \mu \neq 0$$

- Testgröße: Durchschnittliche Bluddruck
- Faustregel $n = 62 > 30$ ist erfüllt
→ zweiseitiger **Gauß-Test**
- Falls wir annehmen, dass die Blutdruckdifferenz normalverteilt ist, können wir auch den zweiseitigen **t-Test** anwenden.
(Standardeinstellung in SPSS)

Dualität Test und Konfidenzintervall

- **Annahmebereich:** Wir behalten H_0 bei, falls die Testgröße T in der Nähe von μ_0 liegt:
- Äquivalente Formulierung über ein **Konfidenzintervall:** Wir behalten H_0 bei, falls μ_0 in der Nähe der Testgröße liegt
- Wir behalten H_0 bei, falls μ_0 im Konfidenzintervall für die Differenz liegt

- Dies gilt sehr allgemein für zweiseitige Test und Konfidenzintervalle
- Dies Prinzip kann zur Konstruktion von Konfidenzintervallen verwendet werden



Einseitiger Gauss-Test auf den Erwartungswert μ

- X Zufallsgröße mit Erwartungswert μ .
- Einseitige Hypothese über μ :

$$H_0 : \mu \leq \mu_0$$

$$H_1 : \mu > \mu_0$$

- Testgröße: Mittelwert in der Stichprobe X_1, \dots, X_n :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

- Stichprobenumfang n ist genügend groß (Faustregel $n > 30$)

Einseitiger Gauss-Test auf den Erwartungswert μ

Testentscheidung zum Signifikanzniveau α

Annahmebereich

$$\bar{X} \leq \mu_0 + z_{1-\alpha} \sqrt{\frac{S^2}{n}}$$

H_0 wird abgelehnt, falls

$$\bar{X} > \mu_0 + z_{1-\alpha} \sqrt{\frac{S^2}{n}}$$

$z_{1-\alpha}$ ist das $(1 - \alpha)$ -Quantil der Standardnormalverteilung.

Ein- bzw. zweiseitiger t-Test auf den Erwartungswert

Was tun wir, falls die Faustregel $n > 30$ nicht erfüllt ist?

Zusätzliche Voraussetzung

- Zufallsgröße X ist normalverteilt.

Wir stellen keine Bedingung an die Stichprobengröße n .

t-Test

- Der einseitige bzw. der zweiseitige t-Test auf den Erwartungswert μ hat die gleiche Form wie der einseitige bzw. zweiseitige Gauss-Test.
- Der t -Test unterscheidet sich vom Gauss-Test dadurch, dass wir das **Quantil z der Standardnormalverteilung** durch das **Quantil t der t -Verteilung mit $n - 1$ Freiheitsgraden** ersetzen.

Verbundene und unverbundene Stichprobe

Verbundene, abhängige Stichprobe

Zwei Stichproben heißen **verbunden**, falls an einem Merkmalsträger (z.B. einer Person) zwei vergleichbare Merkmale erhoben werden.

Man nennt verbundene Stichproben oft auch abhängige Stichproben.

Beispiel: Das Ziel einer medizinischen Studie ist es, die Wirkung eines cholesterin-senkenden Medikaments zu überprüfen.

- Unterteilung der Probanden und Probandinnen in 2 Gruppen: 1 Gruppe erhält das Medikament, 1 Gruppe erhält ein Placebo.
unverbundene Stichprobe
- Alle Probanden und Probandinnen erhalten das Medikament. Von allen Personen wird der Cholesterinspiegel am Anfang und am Ende der Studie erhoben.
verbundene Stichprobe

Approximativer Test auf Erwartungswert-Differenz bei unabhängigen Stichproben

Voraussetzungen:

- X und Y sind zwei Größen mit Erwartungswerten μ_X und μ_Y
- X_1, \dots, X_m und Y_1, \dots, Y_n **unabhängige** Stichproben
- Testgröße: standardisierte Differenz der Mittelwerte

$$\begin{aligned} T &= \bar{X} - \bar{Y} \\ &= \frac{1}{m} \sum_{i=1}^m X_i - \frac{1}{n} \sum_{i=1}^n Y_i \end{aligned}$$

- Faustregel: Stichprobenumfänge $m, n > 30$

Differenz von Erwartungswerten bei unabhängigen Stichproben

Annahmebereich

Für die beiden einseitigen Tests und den zweiseitigen Test auf die Differenz

$$H_0 : \mu_X - \mu_Y = d_0$$

$$H_1 : \mu_X - \mu_Y \neq d_0$$

$$H_0 : \mu_X - \mu_Y \leq d_0$$

$$H_1 : \mu_X - \mu_Y > d_0$$

$$H_0 : \mu_X - \mu_Y \geq d_0$$

$$H_1 : \mu_X - \mu_Y < d_0$$

ist der Annahmebereich

$$d_0 \pm z_{1-\frac{\alpha}{2}} \cdot s$$

$$]-\infty, d_0 + z_{1-\alpha} \cdot s]$$

$$[d_0 - z_{1-\alpha} \cdot s, \infty[$$

mit

$$s = \sqrt{\frac{s_X^2}{m} + \frac{s_Y^2}{n}}$$

Beispiel: Radio-Hördauer Ost-West

- Hören Personen in den alten Bundesländern im Schnitt mehr Radio?
 X : Hördauer in den alten Bundesländern, Y : Radiodauer in den neuen Bundesländern

$$H_0 : \mu_X - \mu_Y \leq 0$$

$$H_1 : \mu_X - \mu_Y > 0$$

- Befragung unter 253 Personen aus den alten Bundesländern und 932 Personen aus den neuen Bundesländern
 - unverbundene Stichproben X_1, \dots, X_{253} und Y_1, \dots, Y_{932}
 - Stichprobengrößen $m = 253, n = 932 > 30$
- Durchschnittliche Hördauer:
11,4 h (Standardabweichung 8,4 h) in den alten Bundesländern
9,5 h (Standardabweichung 8,4 h) in den neuen Bundesländern

Beispiel: Radio-Hördauer Ost-West

- Signifikanzniveau: $\alpha = 0,1$
- Differenz der Radio-Hördauer

$$\bar{X} - \bar{Y} = 11,4 - 9,5 = 1,9$$

- Annahmereich

$$\begin{aligned}\bar{X} - \bar{Y} &\leq z_{1-\alpha} \cdot \sqrt{\frac{s_X^2}{m} + \frac{s_Y^2}{n}} \\ &= z_{0,9} \cdot \sqrt{\frac{8,4^2}{932} + \frac{8,4^2}{253}} \\ &\approx 1,28 \cdot 0,65 \\ &\approx 0,83\end{aligned}$$

- H_0 wird abgelehnt, Personen aus den alten Bundesländern hören signifikant länger Radio.

Doppelter t -Test auf die Erwartungswertdifferenz bei unabhängigen Stichproben

Voraussetzungen:

- X und Y sind zwei Größen mit Erwartungswerten μ_X und μ_Y
- X_1, \dots, X_m und Y_1, \dots, Y_n **unabhängige** Stichproben
- Testgröße: Differenz der Mittelwerte

$$\begin{aligned} T &= \bar{X} - \bar{Y} \\ &= \frac{1}{m} \sum_{i=1}^m X_i - \frac{1}{n} \sum_{i=1}^n Y_i \end{aligned}$$

- X und Y sind **normalverteilt**.
- Die Varianzen sind gleich $\sigma_X^2 = \sigma_Y^2$

Doppelter t -Test auf die Erwartungswertdifferenz bei unabhängigen Stichproben

Annahmebereich

Für die beiden einseitigen t -Tests und den zweiseitigen t -Test auf die Differenz

$$H_0 : \mu_X - \mu_Y = d_0$$

$$H_1 : \mu_X - \mu_Y \neq d_0$$

$$H_0 : \mu_X - \mu_Y \leq d_0$$

$$H_1 : \mu_X - \mu_Y > d_0$$

$$H_0 : \mu_X - \mu_Y \geq d_0$$

$$H_1 : \mu_X - \mu_Y < d_0$$

ist der Annahmebereich

$d_0 \pm t_{(m+n-2; 1-\frac{\alpha}{2})} s_d$] $-\infty, d_0 + t_{(m+n-2; 1-\alpha)} s_d$] $[d_0 - t_{(m+n-2; 1-\alpha)} s_d, \infty[$
mit

$$s_d = s_{X,Y} \cdot \sqrt{\frac{1}{m} + \frac{1}{n}}$$
$$s_{X,Y}^2 = \frac{1}{m+n-2} ((m-1)S_X^2 + (n-1)S_Y^2)$$

$t_{(m+n-2; 1-\frac{\alpha}{2})}$ und $t_{(m+n-2; 1-\alpha)}$ sind die Quantile der t -Verteilung mit $m+n-2$ Freiheitsgraden.

Tests auf Erwartungswertdifferenz bei abhängigen Stichproben

① Gegeben ist eine verbundene Stichprobe
 X_1, \dots, X_n und Y_1, \dots, Y_n

② Bilde die Differenz

$$W_i = X_i - Y_i \quad i = 1, \dots, n$$

③ Führe einen Test auf den Erwartungswert von W durch

- $n > 30 \rightarrow$ Gauß-Test
- W normalverteilt $\rightarrow t$ -Test



Differenz von Anteilen bei unabhängigen Stichproben

Voraussetzungen:

- X und Y sind zwei Bernoulli-Größen mit

$$p_X = P(X = 1)$$

$$p_Y = P(Y = 1)$$

- X_1, \dots, X_m und Y_1, \dots, Y_n **unabhängige** Stichproben
- Testgröße: Differenz der Anteile

$$\begin{aligned} T &= R_X - R_Y \\ &= \frac{1}{m} \sum_{i=1}^m X_i - \frac{1}{n} \sum_{i=1}^n Y_i \end{aligned}$$

- Faustregel: Stichprobenumfänge $m, n > 30$

Differenz von Anteilen bei unabhängigen Stichproben

Annahmebereich

Für die beiden einseitigen Tests und den zweiseitigen Test auf die Differenz

$$H_0 : p_X - p_Y = 0$$

$$H_1 : p_X - p_Y \neq 0$$

$$H_0 : p_X - p_Y \leq 0$$

$$H_1 : p_X - p_Y > 0$$

$$H_0 : p_X - p_Y \geq 0$$

$$H_1 : p_X - p_Y < 0$$

ist der Annahmebereich

$$\left[-z_{1-\frac{\alpha}{2}} \cdot s_r ; z_{1-\frac{\alpha}{2}} \cdot s_r \right] \quad] - \infty, z_{1-\alpha} \cdot s_r] \quad \left[-z_{1-\alpha} \cdot s_r, \infty[$$

mit

$$s_r = \sqrt{r(1-r) \left(\frac{1}{m} + \frac{1}{n} \right)} \quad \text{und} \quad r = \frac{m \cdot r_X + n \cdot r_Y}{m + n}$$

Beispiel: Ist Fernsehen informativ?

Weiterführung des Beispiels aus dem Thema „Schätzen“

- Beurteilen Personen aus den alten Bundesländern den Informationsgehalt im Fernsehen anders als Personen aus den neuen Bundesländern?

zweiseitiger Test

X : Person aus den alten Bundesländern hält Fernsehen für informativ

Y : Person aus den neuen Bundesländern hält Fernsehen für informativ

- Signifikanzniveau: $\alpha = 0,05$
- Umfrage: 253 Personen aus den alten Bundesländern, 932 Personen aus den neuen Bundesländern:
„Halten Sie Fernsehen für informativ? Ja/Nein“
 - **unverbundene Stichproben X_1, \dots, X_{253} und Y_1, \dots, Y_{932}**
 - **Stichprobengrößen $m = 253, n = 932 > 30$**

Beispiel: Ist Fernsehen informativ?

- alte Bundesländer: 206 Personen halten Fernsehen für informativ
neue Bundesländer: 747 Personen halten Fernsehen für informativ
- Anteile

$$R_X = \frac{206}{253} \approx 0,81 \quad \text{und} \quad R_Y = \frac{747}{932} \approx 0,80$$

- Standardfehler

$$\begin{aligned} r &= \frac{m \cdot r_X + n \cdot r_Y}{m + n} = \frac{253 \cdot 0,81 + 932 \cdot 0,8}{932 + 253} \\ &= \frac{950,53}{1185} \approx 0,802 \end{aligned}$$

$$\begin{aligned} s_r &= \sqrt{r(1-r)\left(\frac{1}{m} + \frac{1}{n}\right)} = \sqrt{0,802 \cdot (1 - 0,802) \left(\frac{1}{932} + \frac{1}{253}\right)} \\ &\approx 0,03 \end{aligned}$$

Beispiel: Ist Fernsehen informativ?

- Annahmebereich

$$\begin{aligned}\pm z_{1-\frac{\alpha}{2}} \cdot s_r &\approx \pm z_{0,975} \cdot 0,03 \\ &\approx \pm 1,96 \cdot 0,03 \\ &\approx \pm 0.059\end{aligned}$$

- Differenz der Anteile in der Stichprobe

$$R_X - R_Y \approx 0,81 - 0,8 = 0,01$$

- H_0 wird beibehalten, der Unterschied zwischen alten und neuen Bundesländern ist nicht signifikant

Differenz von Anteilen bei abhängigen Stichproben

Voraussetzungen:

- X und Y sind zwei Bernoulli-Größen mit

$$p_X = P(X = 1)$$

$$p_Y = P(Y = 1)$$

- $(X_1, Y_1), \dots, (X_n, Y_n)$ **abhängige, verbundene** Stichproben
- Absolute Häufigkeiten werden in einer Kontingenztafel festgehalten

	Y=0	Y=1
X=0	n_{11}	n_{12}
X=1	n_{21}	n_{22}

- Verwende Test von McNemar

Zusammenhang zwischen 2 kategorialen Merkmalen

- Sind zwei kategoriale Merkmale unabhängig?
- **Beispiele**
 - Gibt es einen Zusammenhang zwischen besuchter Schule (Hauptschule, Realschule, Gymnasium) und Fernsehkonsum (hoch/niedrig)?
 - Gibt es einen Zusammenhang zwischen Geschlecht (m/w) und der Affinität zu Fußball (Fan/kein Fan)?
 - ...
- empirische Untersuchung mittels Kontingenztafeln und Kennzahlen



Randverteilungen

Die zu den einzelnen Merkmalen gehörigen empirischen Häufigkeitsverteilungen heißen **Randverteilungen** oder **marginale Verteilungen**.

<small>Fußball</small> Geschlecht	ja	nein	Summe
m	87	10	97
w	23	31	54
Summe	110	41	151

Randverteilung Geschlecht

m	w	
97	54	← Absolute Häufigkeiten
$\frac{97}{151}$	$\frac{54}{151}$	← Relative Häufigkeiten

Randverteilung Fußballfan

ja	nein	
110	41	← Absolute Häufigkeiten
$\frac{110}{151}$	$\frac{41}{151}$	← Relative Häufigkeiten

Bedingte Verteilung

Unter der **bedingten Verteilung** von X gegeben $Y = B_j$ versteht man die Verteilung von X in der Teilgesamtheit der Untersuchungseinheiten, die die Ausprägung B_j des Merkmals Y aufweisen.

Fußball Geschlecht	ja	nein	Summe
m	87	10	97
w	23	31	54
Summe	110	41	151

Verteilung „Fußballfan“ bei Männern

ja	nein	Ges.
87	10	97
90%	10%	100%

Verteilung “Geschlecht“ bei Fußballfans

m	w	Ges.
87	23	110
79%	21%	100%

Empirische Unabhängigkeit

- Falls die beiden Merkmale unabhängig voneinander sind, so sollte

„relative Häufigkeit Fußballfan“
 \approx „relative Häufigkeit Fußballfans unter Frauen“
 \approx „relative Häufigkeit Fußballfans unter Männern“

- Formel:

bedingte relative Häufigkeit $\frac{n_{ij}}{n_{\bullet j}} = \frac{n_{i\bullet}}{n}$ relative Randhäufigkeit

- Daraus folgt

$$e_{ij} := \frac{n_{i\bullet} n_{\bullet j}}{n}$$

ist die unter Unabhängigkeit erwartete Anzahl.

- Falls die Merkmale unabhängig sind, sollte gelten:

$$e_{ij} \approx n_{ij}$$

χ^2 -Unabhängigkeitstest

- zwei Zufallsgrößen X und Y mit k bzw. l Ausprägungen

$$p_{ij} = P(X = i, Y = j)$$
$$p_{i\bullet} = P(X = i) \quad p_{\bullet j} = P(Y = j)$$

- Hypothesen

H_0 : X und Y sind stochastisch unabhängig

$$p_{ij} = p_{i\bullet} \cdot p_{\bullet j} \text{ für alle } i = 1, \dots, k, j = 1, \dots, l$$

H_1 : X und Y sind stochastisch abhängig

$$p_{ij} \neq p_{i\bullet} \cdot p_{\bullet j} \text{ für mindestens eine } ij\text{-Kombination}$$

- Prüfgröße

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

- Faustregel $n_{ij} \geq 5$ für alle i, j

Annahmebereich

Für den χ^2 -Unabhängigkeitstest ist der Annahmebereich

$$\chi^2 \leq q_{1-\alpha, (k-1)(l-1)}$$

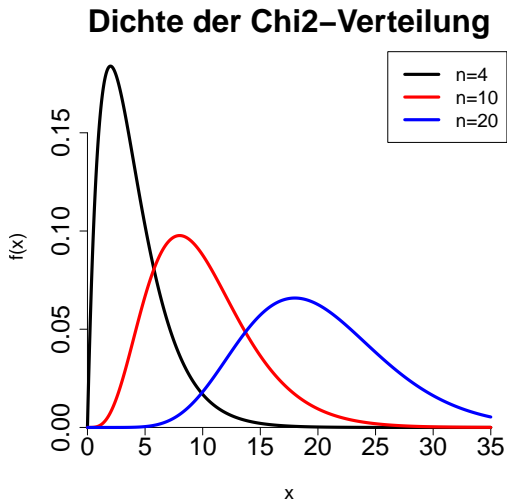
Die Nullhypothese wird abgelehnt, falls

$$\chi^2 > q_{1-\alpha, (k-1)(l-1)}$$

Dabei ist $q_{1-\alpha, (k-1)(l-1)}$ das

- $(1 - \alpha)$ -Quantil
- der χ^2 -Verteilung
- mit $(k - 1) \cdot (l - 1)$ Freiheitsgraden.

Erinnerung: Dichte der χ^2 -Verteilung mit n Freiheitsgraden



Beispiel

$$e_{ij} = \frac{n_{i\bullet} n_{\bullet j}}{n}$$

Fußball Geschlecht	ja	nein	Summe
m	87	10	97
w	23	31	54
Summe	110	41	151

Erwartete Besetzungszahlen bei Unabhängigkeit

	ja (j=1)	nein (j=2)
m (i=1)	$\frac{97 \cdot 110}{151} \approx 71$	$\frac{97 \cdot 41}{151} \approx 26$
w (i=2)	$\frac{54 \cdot 110}{151} \approx 39$	$\frac{54 \cdot 41}{151} \approx 15$

$$\begin{aligned}\chi^2 &= \sum_{i=1}^k \sum_{j=1}^l \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \\ &\approx \frac{(87 - 71)^2}{71} + \frac{(10 - 26)^2}{26} + \frac{(23 - 39)^2}{39} + \frac{(31 - 15)^2}{15} \\ &\approx 37,09\end{aligned}$$



Beispiel

- Signifikanzniveau: $\alpha = 0.01$
- Faustregel gültig?
Besetzungszahlen $n_{ij} \geq 5$
- Bestimmung der Freiheitsgrade: $k = l = 2$

$$\text{Freiheitsgrade} = (k - 1) \cdot (l - 1) = (2 - 1) \cdot (2 - 1) = 1$$

$$q_{1-0.01; (2-1)(2-1)} = q_{0.09; 1} \approx 6,63$$

- H_0 wird abgelehnt

Unabhängigkeit und Differenz von Anteilen

Die beide Fragen:

- Gibt es Unterschiede in den Anteilen von $Y=1$ zweier Gruppen ?
- Gibt es einen Zusammenhang zwischen Gruppen-Zugehörigkeit und einem binären Binären Merkmal Y ?

sind äquivalent.

Die beiden Testprozeduren Gauss Test für Differenzen von Anteilen und χ^2 -Unabhängigkeitstest führen zum gleichen Ergebnis.



Nichtparametrische Tests

- Bis auf den χ^2 -Unabhängigkeits-Test bauen alle Tests auf der (zumindestens approximativen Gültigkeit der) Normalverteilungsannahme auf.
- Problematisch, z.B.
 - bei kleinen Stichprobenumfängen (z.B. in der Vorbereitung von strukturierten Beobachtungen, bei nicht reaktiven Verfahren oder in der Psychologie und Medizin)
 - oder bei ordinalen Daten mit wenigen unterschiedlichen Ausprägungen.
- Hier kann die unreflektierte Anwendung der Standardtests zu Fehlergebnissen führen.
- Ein wichtiger Ausweg: nichtparametrische Tests = Verteilungsfreie Verfahren
- Hier wird die Information in den Beobachtungen auf Ränge, bzw. größer/kleiner Vergleiche reduziert.
- Bekannteste Beispiele: Wilcoxon-Test, Vorzeichentest.



Der Wilcoxon test für unabhängige Stichproben

Voraussetzungen:

- X und Y sind zwei Größen mit Medianen med_X und med_Y
- X_1, \dots, X_m und Y_1, \dots, Y_n **unabhängige** Stichproben
- $H_0 : med_X = med_Y$ vs. $H_1 : med_X \neq med_Y$

Grundidee: Betrachte die Ränge aus allen Beobachtungen X_i und Y_j und bezeichne diese mit $rg(X_i)$ und $rg(Y_j)$, z.B.

$X_1 = 3, X_2 = 5, Y_1 = 6, Y_2 = 1, Y_3 = 4 \Rightarrow$

$rg(X_1) = 2, rg(X_2) = 4, rg(Y_1) = 5, rg(Y_2) = 1, rg(Y_3) = 3$

Testgröße:

$$T = \sum_{i=1} X_i$$

Die exakte Verteilung von T kann berechnet werden. Für hinreichend große n und m kann sie durch eine NV approximiert werden. Ablehnung von H_0 falls große und kleine T .

Multiple Testprobleme:

- Gegeben sei ein rein zufälliger Datensatz mit 50 Variablen ohne irgendeinen Zusammenhang.
- Man testet alle Variablenpaare auf einen Zusammenhang 1225 Tests.

Bei vorgegebener Irrtumswahrscheinlichkeit von 5% gilt für die Anzahl fälschlich verworfener Nullhypothesen $X \sim B(1225, 0.05)$ und somit $E(X) = 61,25$.

Im Durchschnitt wird also mehr als 61 mal die Nullhypothese, dass kein Zusammenhang besteht, verworfen.

⇒ wenige, sinnvolle Hypothesen *vorher inhaltlich* überlegen. In den Daten entdeckte „Zusammenhänge“ als statistisch signifikant nachzuweisen, ist (fast) zirkulär.

- Es gibt Ansätze, wie man bei großen Hypothesensystemen diesem Problem entkommt.

⇒ Theorie des multiplen Testens.

Z.B. Adjustierung des Irrtumswahrscheinlichkeit: Statt α betrachte man $\alpha/(\text{Anzahl der Tests})$. Diese spezielle Korrektur ist aber meist überkonservativ und kann durch bessere Korrekturen ersetzt werden.

Bestimmung des Stichprobenumfangs

- Bestimmung mit Obergrenze für Fehler 2. Art (β -Fehler)
- $\beta = P(\text{Nullhypothese wird nicht abgelehnt} \mid H_1 \text{ wahr})$
- $1-\beta$ wird auch als Power bezeichnet
- üblicher Wert Power = 0.8 bzw $\beta = 0.2$



Fehler 2. Art im Binomialtest

Interessante Größen

$H_0 : \pi = 0.5$ versus $H_1 : \pi > 0.5$

β : P(H_0 wird beibehalten— H_1)

π_w : wahre Erfolgswahrscheinlichkeit

Numerische Werte für $n=10$ und $c=7$

$\pi_w = 0.6$ und $\beta = 0.95$

$\pi_w = 0.7$ und $\beta = 0.85$

$\pi_w = 0.8$ und $\beta = 0.62$

Numerische Werte für $n=20$ und $c=13$

$\pi_w = 0.6$ und $\beta = 0.87$

$\pi_w = 0.7$ und $\beta = 0.58$

$\pi_w = 0.8$ und $\beta = 0.02$

Numerische Werte für $n=100$ und $c=57$

$\pi_w = 0.6$ und $\beta = 0.37$

$\pi_w = 0.7$ und $\beta = 0.007$

$\pi_w = 0.8$ und $\beta < 0.0001$

Fehler 2. Art im Beispiel des Binomialtests

Fehler 2. Art sinkt hier mit:

- Wachsender wahrer Erfolgswahrscheinlichkeit ($\pi > 0.5$)
- Wachsendem Stichprobenumfang

Stichprobenumfangbestimmung

Festlegen von:

α , β , Effektgröße π_w

Beispiel:

$$\pi_w = 0.7$$

$$\Rightarrow n = 20 \quad \Rightarrow \beta = 0.58 \text{ zu klein}$$

$$\Rightarrow n = 100 \quad \Rightarrow \beta = 0.07 \text{ zu groß}$$

Wähle kleinstes n mit:

$$\beta \leq 0.2$$

Hier:

$$n = 37, \beta = 0.19$$

$$n = 36, \beta = 0.26$$

Bestimmung des Stichprobenumfangs beim t-Test

Vergleich zweier Gruppen

Zielgröße:	Y
Relevanter Unterschied:	d
Annahme zur Streuung	σ
Signifikanzniveau:	$\alpha = 0.05$
Fehler 2.Art:	$\beta = 0.2$

Stichprobenumfang beim t- Test

Einige Werte für Stichprobenumfang unter $\alpha = 0.05, \beta = 0.2$ (zweiseitige Fragestellung Zwei-Stichproben t-Test)

Beachte: Es kommt nur auf das Verhältnis d/σ (Effektstärke) an.

d/σ	$n_1 = n_2$
2	5
1.33	10
1	17
0.66	37
0.58	50
0.29	100

Statistischer Test

Ein statistischer Test ist eine Untersuchung, ob man eine Hypothese über die Grundgesamtheit mit Hilfe einer Stichprobe widerlegen kann

Wichtige Tests

- einseitige Tests \longleftrightarrow zweiseitiger Test
- Tests auf Anteile, Erwartungswerte und deren Differenzen
 - eine Stichprobe
 - zwei Stichproben \rightarrow verbunden oder unverbunden
- Test auf Unabhängigkeit: χ^2 -Unabhängigkeitstest

Interpretation eines Tests

- Statistische Signifikanz bedeutet: Das Ergebnis ist nicht durch Zufall erklärbar.
- Statistische Signifikanz bedeutet nicht unbedingt, dass der Unterschied relevant ist.
→ Daher immer die Größe des Effekts (Schätzung) angeben!
- Ein statistischer Test liefert aber allein noch keinen kausalen Zusammenhang.
- Bei nicht signifikantem Ergebnis immer Konfidenzintervalle der entsprechenden Parameter angeben. Sie geben eine Idee, wie groß der nicht nachgewiesene Effekt sein könnte.