Hypothesentests

Hypothese

Behauptung eines Sachverhalts, dessen Überprüfung noch aussteht.

Wissenschaftliche Vorgehensweise beim Hypothesentest

Forscher formuliert eine ***Alternativhypothese*** $\boldsymbol{H_1}$ (die neue Erkenntnis, die gezeigt werden soll) gegenüber der ***Nullhypothese*** $\boldsymbol{H_0}$ (die bislang als gesichert geltende Erkenntnis). Sodann testet er mit statistischen Methoden, ob die tatsächlichen Daten <u>unter Gültigkeit der Nullhypothese</u> signifikant (und nicht nur als Folge von Zufallsstreuungen) abweichen. Erst bei ***überzufälligen*** Abweichungen ist die Nullhypothese abzulehnen, d.h. die Alternativhypothese ist verifiziert. Wissenschaftstheoretischer Hintergrund: Erkenntnisfortschritt durch Falsifikation, Verwerfen „falscher“ Hypothesen (*Popper, Lakatos*).

Ergebnissensibilität

- Da von ***Stichprobe(n)*** auf die Grundgesamtheit geschlossen wird (Modellannahme), sind Fehlentscheidungen nicht ausgeschlossen.
- Es werden willkürlich (und unabhängig von den Daten) eine Schranke (= ***Signifikanzniveau*** $\boldsymbol{\alpha}$), und das Verfahren (= ***Teststatistik***) gewählt.
- Ein statistischer Test kann nicht die Wahrheit der Nullhypothese zeigen

Ergo: ***Hypothesentests „beweisen“ nichts, sondern falsifizieren (vorläufig) eine Aussage!***

Systematisierung nach Inhalt der Hypothese

- ***Test für Erwartungswerte, Varianzen, Anteile*** von Grundgesamtheiten
- ***Anpassungstests***: Prüft, ob eine Zufallsvariable einer bestimmten Verteilung folgt
- ***Unabhängigkeitstests***: Prüft, ob Zufallsvariablen voneinander unabhängig sind

Systematisierung nach Teststatistik

- Z.B. Gauß-Test, t-Test, f-Test, Chi-Quadrat-Test
-

Eselsbrücke Testrichtung!

Systematisierung nach Hypotheseninhalt

- $H_0: x \leq x_0$ gegen $H_1: x > x_0$ einseitiger (rechtsseitiger) Test
- $H_0: x \geq x_0$ gegen $H_1: x < x_0$ einseitiger (linksseitiger) Test
- $H_0: x = x_0$ gegen $H_1: x \neq x_0$ zweiseitiger Test

Vorgehensweise (Gauß-Test)

Die Testentscheidung basiert auf der (beim parametrischen Test bekannten) Verteilung der Prüfstatistik unter $\boldsymbol{H_0}$, aus der man (vorab) ein (Konfidenz-)Intervall generiert, in das die zu prüfende Größe mit einer (hohen) Wahrscheinlichkeit $\boldsymbol{1 - \alpha}$ fällt. Die ***Irrtumswahrscheinlichkeit = Signifikanzniveau*** $\boldsymbol{\alpha}$ soll möglichst klein sein, üblich sind: ***0,05 („signifikant“), 0,01 („hoch signifikant“), 0,001 (höchst signifikant).***

> Wenn das Prüfergebnis außerhalb des Vertrauensintervalls $\boldsymbol{1 - \alpha}$, folglich innerhalb des Signifikanzniveaus (= Ablehnungsbereich) $\boldsymbol{\alpha}$ liegt, wird $\boldsymbol{H_0}$ abgelehnt.

<u>Schematischer Ablauf</u> (Beispiel Mittelwerttest bei normalverteilter Population, Varianz = 100)

1. Aufstellen und Formulierung der Hypothese, z.B. H_0: $\mu = 500$ gegen H_1: $x \neq 500$

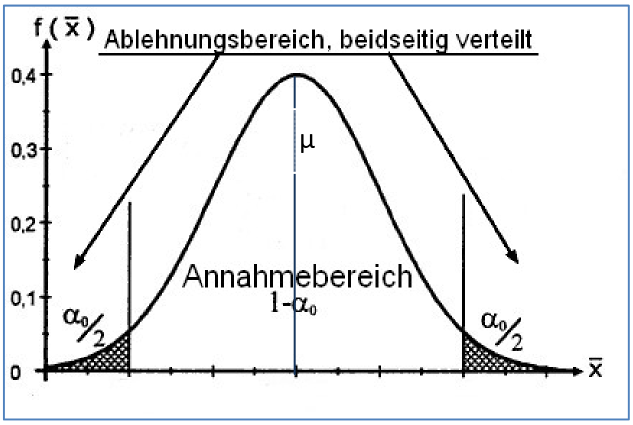
2. Definition des Signifikanzniveaus, z.B. $\alpha = 0{,}01$

3. Bestimmung der Testgröße mit geeigneter Teststatistik
(hier: Standardisierung der Testvariable, Stichprobenmittelwert = 477, n = 25)

$$Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X} - \mu_0}{\sigma} \cdot \sqrt{n} = \frac{477 - 500}{10} \cdot 5 = -11{,}5$$

Streuung, Standardfehler

4. Lokalisierung der kritischen Grenze und Entscheidung über Annahme/Ablehnung
(hier: zweiseitiger Test)

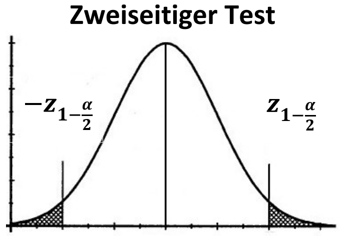
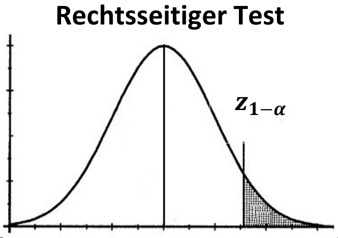
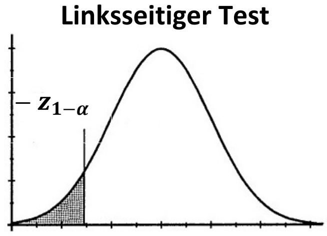


Ablesung in der Tabelle für Quantile der Standardnormalverteilung

$$z_{1-\frac{\alpha}{2}} = z_{1-\frac{0{,}01}{2}} = z_{1-\frac{0{,}01}{2}} = z_{0{,}995} = \mathbf{2{,}58}$$

Weil **-2,58 ≤ *Annahmebereich* ≤ 2,58** =|2,58|**,**
liegt (wegen **Z = - 11,5**) die gezogene Stichprobe außerhalb des Annahmebereichs, deswegen ist die Nullhypothese abzulehnen.

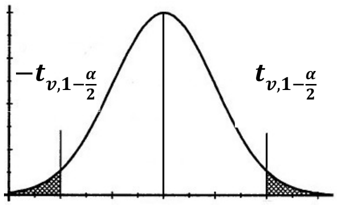
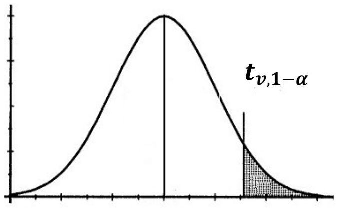
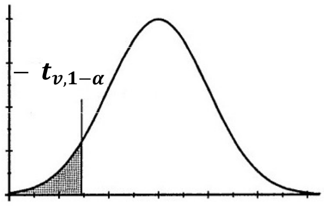
a.) Gaußtest für Mittelwert, σ bekannt, standardnormalverteilte Testgröße $Z = \frac{\bar{X} - \mu_0}{\sigma} \cdot \sqrt{n}$

Zweiseitiger Test	**Rechtsseitiger Test**	**Linksseitiger Test**
		
H_0: $x = x_0$ gegen H_1: $x \neq x_0$	H_0: $x \leq x_0$ gegen H_1: $x > x_0$	H_0: $x \geq x_0$ gegen H_1: $x < x_0$
z.B. Stichprobe aus Serienfertigung: H_0: Fertigung ist fehlerfrei H_1: F. weicht signifikant (-/+) ab	z.B. Stichprobe aus Serienfertigung: H_0: Fertigung ist fehlerfrei H_1: weicht signifikant nach oben ab	z.B. Stichprobe aus Serienfertigung: H_0: Fertigung ist fehlerfrei H_1: weicht signifikant nach unten ab
Kritischer Wert: $z_{1-\frac{\alpha}{2}}$	Kritischer Wert: $z_{1-\alpha}$	Kritischer Wert: $-z_{1-\alpha}$
Ablehnung H_0, wenn $\lvert Z \rvert > z_{1-\frac{\alpha}{2}}$	Ablehnung H_0, wenn $Z > z_{1-\alpha}$	Ablehnung H_0, wenn $Z < -z_{1-\alpha}$

b.) t-Test für Mittelwert, σ unbekannt, t-verteilte Testgröße

$$Z = \frac{\overline{X} - \mu_0}{S*} \cdot \sqrt{n}$$

Wegen unbekannter Varianz wird hier die *korrigierte Stichprobenstandardabweichung* ***S**** in der Testgröße verwendet. Die Ergebnisse sind t-verteilt, wodurch sich der Annahmebereich gegenüber der Normalverteilung verbreitert - wobei ab ***n ≥ 30*** die Quantile der Normalverteilung eine gute Näherung erreichen. Durch den zentralen Grenzwertsatz kann man mit diesem Verfahren bei großen Stichproben ***auch unbekannt verteilte Populationen*** testen, dann jedoch unter Verwendung der Quantile der Normalverteilung statt der der t-Verteilung.

Zweiseitiger Test	**Rechtsseitiger Test**	**Linksseitiger Test**
		
H_0: $x = x_0$ gegen H_1: $x \neq x_0$	H_0: $x \leq x_0$ gegen H_1: $x > x_0$	H_0: $x \geq x_0$ gegen H_1: $x < x_0$
Kritischer Wert: $t_{v(=n-1);\,1-\frac{\alpha}{2}}$	Kritischer Wert: $t_{v;1-\alpha}$	Kritischer Wert: $-t_{v;1-\alpha}$
Ablehnung H_0, wenn $\lvert Z\rvert > t_{v;\,1-\frac{\alpha}{2}}$	Ablehnung H_0, wenn $Z > t_{v;\,1-\alpha}$	Ablehnung H_0, wenn $Z < -t_{v;\,1-\alpha}$

c.) Varianztest, χ^2-verteilte Testgröße

$$Z = \frac{n \cdot S^2}{\sigma_0^2} = \frac{(n-1) \cdot S^{*2}}{\sigma_0^2}$$

Da die Testgröße ein χ^2-verteiltes Ergebnis liefert, sind die ***Intervallgrenzen nicht mehr symmetrisch***. Die Kenntnis des Mittelwertes ***μ*** der Population ist nicht erforderlich.

Zweiseitiger Test	**Rechtsseitiger Test**	**Linksseitiger Test**
H_0: $\sigma = \sigma_0$ gegen H_1: $\sigma \neq \sigma_0$	H_0: $\sigma \leq \sigma_0$ gegen H_1: $\sigma > \sigma_0$	H_0: $\sigma \geq \sigma_0$ gegen H_1: $\sigma < \sigma_0$
z.B.	z.B.	z.B.
Kritischer Wert: ${\chi^2}_{v(=n-1);\,\frac{\alpha}{2}}$ und ${\chi^2}_{v(=n-1);\,1-\frac{\alpha}{2}}$	Kritischer Wert: ${\chi^2}_{v;\,\alpha}$	Kritischer Wert: ${\chi^2}_{v;\,1-\alpha}$
Ablehnung H_0, wenn $Z < {\chi^2}_{v;\,\frac{\alpha}{2}}$ oder wenn $Z > {\chi^2}_{v;\,1-\frac{\alpha}{2}}$	Ablehnung H_0, wenn $Z > {\chi^2}_{v;\,1-\alpha}$	Ablehnung H_0, wenn $Z < {\chi^2}_{v;\,\alpha}$

d.) Zweistichproben-Gaußtest für Mittelwerte,

σ bekannt, $\sigma_1 = \sigma_2$, **σ unbekannt**

standardnormalverteilte Testgröße

$$Z = \frac{\overline{X_1} - \overline{X_2}}{\sigma} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

$$Z = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{(n_1-1)\cdot S_1^{*2} + (n_2-1)\cdot S_2^{*2}}{n_1+n_2-2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Im Gegensatz zu den oben gezeigten Einstichproben-Tests liegen hier ***zwei unabhängige*** (***!***) Zufallsstichproben aus zwei normalverteilten Teilmengen einer Population vor, die hinsichtlich eines Lageunterschieds beider Mittelwerte getestet werden.

Zweiseitiger Test
H_0: $\mu_1 = \mu_2$ gegen H_1: $\mu_1 \neq \mu_2$
z.B. Versuchsgruppe gegen Kontrollgruppe
Kritischer Wert: : $t_{v(=n-1);\,1-\frac{\alpha}{2}}$
Ablehnung H_0, wenn $\lvert Z\rvert > t_{v;\,1-\frac{\alpha}{2}}$

e.) Unabhängigkeitstest

Testet auf Unabhängigkeit (bzw. einen möglichen Zusammenhang) zweier Merkmale X u. Y.

Ansatz *k* x *m* -Kontingenztabelle

Merkmal ***X*** hat Ausprägungen $a_1 \dots a_k$ und Merkmal ***Y*** hat die Ausprägungen $b_1 \dots b_m$.
Es gelten ***i*** Zeilen und ***j*** Spalten.

Y

X	b_1	b_2	b_m	Randverteilung X
a_1	h_{11}	h_{12}	h_{1m}	$h_{1.}$
a_2	h_{21}	h_{22}	h_{2m}	$h_{2.}$
a_k	h_{k1}	h_{k2}	h_{km}	$h_{k.}$
Randverteilung Y	$h_{.1}$	$h_{.2}$	$h_{.m}$	n

k-Richtung (X(a)) m-Richtung (Y(b))

z.B.:

	Bildung (Y)		
Fremdenhass (X)	***Schulabschluss***	***kein SA***	***Σ***
Nein	648	153	801
Ja	97	102	199
Σ	745	255	**1000**

> Empirische Unabhängigkeit besteht, wenn $\mathbf{h_{ij}} = \tilde{h}_{ij}$, d.h. die beobachtete Häufigkeit (Ablesung aus der betreffenden Zelle) mit der zu erwartenden Häufigkeit (Zugrundelegung der gegebenen Randsummen bei Gleichverteilung) durchgehend (!) übereinstimmt.

Berechnungsbeispiel (h_{11} ***aus Tabelle oben):***

$$\tilde{h}_{11} = \frac{h_{i.} \cdot h_{.j}}{n} = \frac{\text{Zeilensumme} \cdot \text{Spaltensumme}}{\text{Gesamtsumme}} = \frac{h_{i.} \cdot h_{.j}}{n} = \frac{(h11+h12)\cdot(h11+h21)}{n} = \frac{801 \cdot 745}{1000} = \mathbf{596}$$

h_{11} = Abgelesener Wert aus der betreffenden Zelle = **648**

Es ergibt sich somit die Aussage, dass für h_{11} keine Unabhängigkeit besteht.
Eine Aussage über den Grad der Abhängigkeit bzw. die Signifikanz ist zunächst nicht gegeben.
Bei Unabhängigkeit ($\mathbf{h_{ij}} = \tilde{h}_{ij}$) beinhalten die Randverteilungen alle erforderlichen Informationen.
Bei Abhängigkeit reichen die Randinformationen nicht aus, da sich eine Differenz $\mathbf{h_{ij}} - \tilde{h}_{ij} \neq 0$ ergibt.
Eine weitergehende Quantifizierung liefert der

χ^2- Unabhängigkeitstest, nichtparametrisch, Testgröße

$$\chi^2 = \sum_{i=1}^{k} \sum_{i=1}^{m} \frac{(h_{ij} - \tilde{h}_{ij})^2}{\tilde{h}_{ij}}$$

Vereinfachte Formel im Vierfelderfall:

$$\chi^2 = \frac{n \cdot (h_{11}h_{22} - h_{12}h_{21})^2}{h_{1.}\ h_{2.}\ h_{.1}\ h_{.2}}$$

Produkt der Randhäufigkeiten

Rechtsseitiger Test
H_0: X und Y sind unabhängig H_1: X und Y sind abhängig
z.B. Abhängigkeit von Fernsehkonsum und Fehlsichtigkeit
Kritischer Wert: : $\chi^2_{(k-1)\cdot(m-1);\,1-\alpha}$
Ablehnung H_0, wenn $\chi^2 > \chi^2_{(k-1)\cdot(m-1);\,1-\alpha}$

Der p-Wert

Stellt eine alternative Variante von Hypothesentests dar. Es wird hier die konkrete Wahrscheinlichkeit für einen $\boldsymbol{\alpha}$- Fehler angegeben. Die Werte werden meist von Computerprogrammen ausgegeben, so dass der Benutzer keine tabellierten kritischen Werte ablesen, sondern nur $\boldsymbol{\alpha}$ und $\boldsymbol{p}$ gegenüberstellen muss. **H_0 wird genau dann abgelehnt, wenn der p-Wert $\leq$ α ist**. Zudem kann unmittelbar auf den Grad der Signifikanz geschlossen werden.

Fehlerarten

- **Fehler 1. Art (α - Fehler):** Nullhypothese wird abgelehnt, obwohl sie richtig ist.
  Beispiel: Obwohl die Fertigung normgerecht ist, wird sie zurückgewiesen.
  Die Wahrscheinlichkeit dafür beträgt (beim einseitigen Test) $\boldsymbol{\alpha}$.
  Nur ein Fehler 1. Art kann über Testdesign kontrolliert werden.
- **Fehler 2. Art (β - Fehler)**: Nullhypothese wird beibehalten, obwohl sie falsch ist.
  Beispiel: Die Fertigung wird fortgesetzt, obwohl sie von der Norm abweicht.
  Die Wahrscheinlichkeit dafür ist schwer bzw. gar nicht bestimmbar. Je weiter μ_0 von μ_1 entfernt ist, desto kleiner wird der **$\boldsymbol{\beta}$ – Fehler.** Je größer der Stichprobenumfang, desto kleiner wird $\boldsymbol{\beta}$. Reduziert man (die Wahrscheinlichkeit für) $\boldsymbol{\alpha}$, erhöht sich die Wahrscheinlichkeit für $\boldsymbol{\beta}$ (und umgekehrt). Allerdings besteht keine Komplementärbeziehung!

Gütefunktion (für Tests für den Erwartungswert μ)

Dient zur Beurteilung der Leistungsfähigkeit eines Tests.