Einführung in Web- und Data-Science

Prof. Dr. Ralf Möller

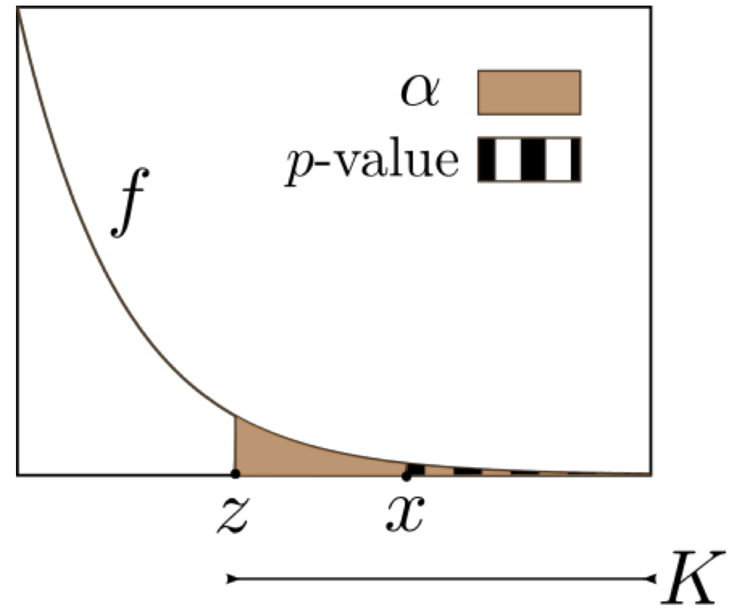
Universität zu Lübeck

Institut für Informationssysteme

Tanya Braun (Übungen)

P-Wert (einseitiger Ablehnungsbereich)

- Hypothesentest H_0 vs. H_1
- Wie extrem ist der auf Basis der erhobenen Daten berechnete Wert der Teststatistik?
- P-Wert = Wahrscheinlichkeit, bei Gültigkeit von H_0 den bestimmten oder einen extremeren Wert der Teststatistik zu erhalten



Für diese Realisation x im Ablehnbereich K ist der p-Wert kleiner als α, oder dazu äquivalent ist die Realisation der Teststatistik x größer als der kritische Wert z. Hier ist f die Wahrscheinlichkeitsdichte der Verteilung unter der Nullhypothese

In machen Veröffentlichungen wird leider α als p-Wert bezeichnet!

Reliabilität

- Zuverlässigkeit
  - Angegeben beispielsweise durch Konfidenzintervall

- Messgenauigkeit eines Tests mit mehreren Indikatoren bzw. Merkmalen (Beispiel: Fragebogen) und z.B. Mittelung der ermittelten Werte der Teilmerkmale
  - Interne (innere) Konsistenz
    - Wird von verschiedenen Merkmalen (z.B. an verschiedenen Stellen eines Fragebogens) dasselbe gemessen?
  - Zeitliche Stabilität
    - Wird zu verschiedenen Zeitpunkten (bei Testwiederholung) dasselbe gemessen?

Bestimmung der Reliabilität eines Tests

- Re-Test-Reliablität :
  - Bestimmung des statistischen Zusammenhangs (Korrelation) zwischen zwei aufeinanderfolgenden Messungen

- Ein Test misst dann genau, wenn er zu mehreren Zeitpunkten dasselbe Ergebnis liefert.

- Korrelation desselben Fragebogen-Gesamtwerts zu verschiedenen Zeitpunkten mit denselben Probanden (ungeeignet bei vorübergehenden Merkmalen, z.B. Stimmung)

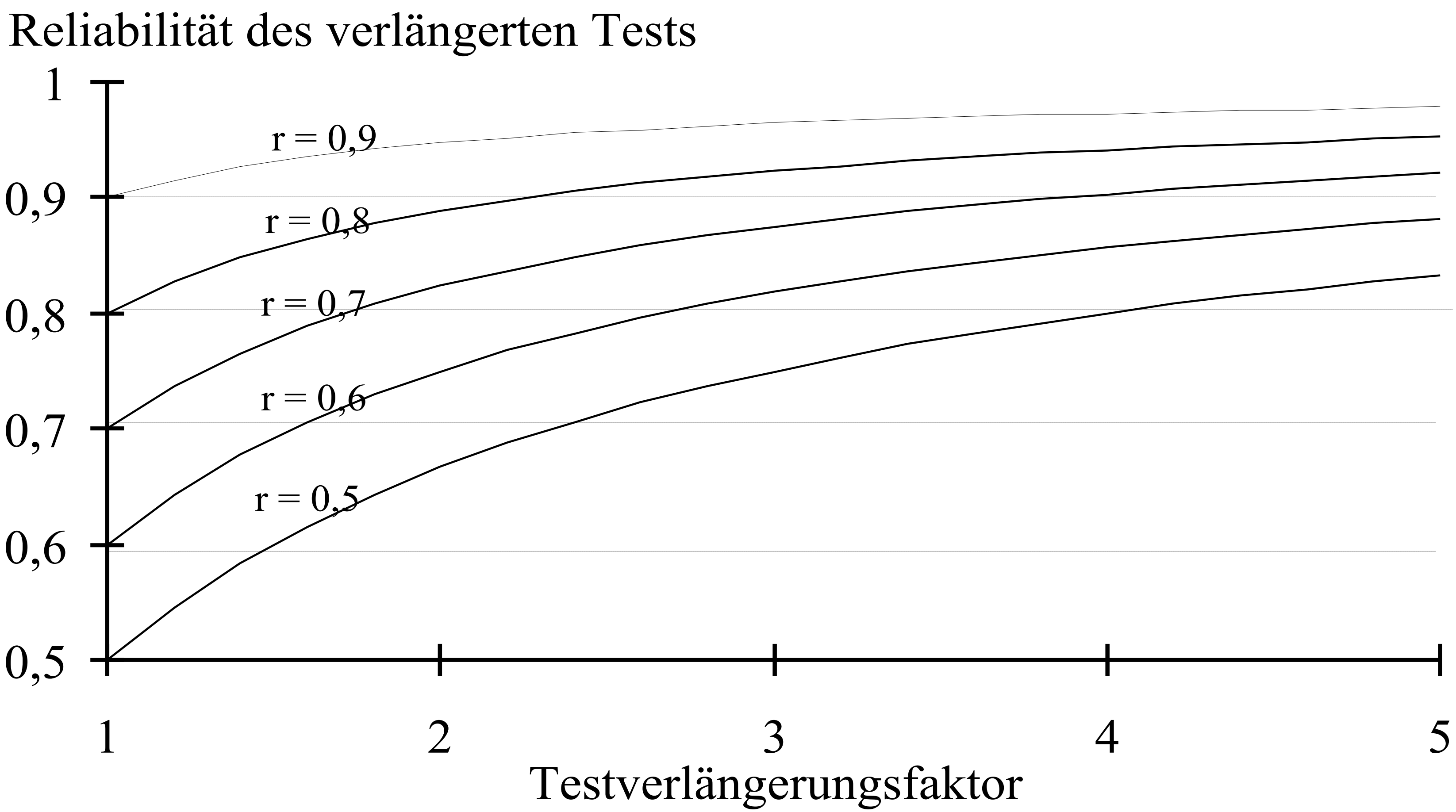
Bestimmung der Reliabilität eines Tests

- Split-Half-Reliabilität:
  - Korrelation zwischen zwei Hälften der Items eines Tests

- Cronbachs Alpha (Maß für sog. Interne Konsistenz):
  - Mittelwert der Korrelationen zwischen allen Einzelitems
  - Ausreichende Reliabilität: r: = 0.75
  - Gute Reliabilität: r = 0.90

- Probleme:
  - Die Messgenauigkeit kann nur für mehrere Items (Skala, Test, Subtest) bestimmt werden, nicht für Einzelitems
  - Daher liefert ein Test, der nicht vollständig durchgeführt wurde, keine zuverlässige Messung
  - Je mehr Items ein Test (Subtest, Skala) enthält, desto „genauer“ wird er

Reliabilitätssteigerung durch Testverlängerung



Validität

- Validität: Gültigkeit
- Misst ein Test das, was er messen soll?
  - Zusammenhang zwischen dem Testergebnis und anderen Kriterien für das Zielverhalten
  - Evaluation durch Bestimmung des Zusammenhangs (Korrelation) zwischen dem Testergebnis und anderen Kriterien für das messende Verhalten

Nächstes Thema: Unterschiedshypothesen

Danksagung

- Nachfolgende Materialen sind mit Änderungen übernommen aus:
- Vorlesung Statistik (WS08/09) aus dem
  Studiengang Psychologie and der Universität Freiburg

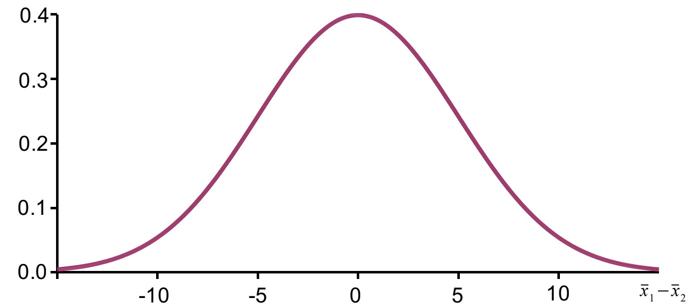
Unterschiedshypothesen

- Sind Frauen ängstlicher als Männer?
  - Unterscheiden sich die Mittelwerte von zwei Gruppen?
  - Unabhängige Stichproben
- Ist der Mittelwert der Ängstlichkeit nach einer Therapie größer als vor der Therapie?
  - Unterscheidet sich der Mittelwert einer Stichprobe zu zwei Messzeitpunkten?
  - Abhängige Stichproben
- Liegt der mittlere IQ einer Gruppe über 100?
  - Unterscheidet sich der Mittelwert einer Gruppe von einem vorgegeben Wert?
  - Test bzgl. Gruppe

Unterschiedshypothesen: Unabhängige Stichproben

Unterscheiden sich die Mittelwerte von zwei Gruppen?

- Differenz der Mittelwerte zweier Stichproben: $\Delta_x = \bar{x}_1 - \bar{x}_2$
- Schätze die Dichtefunktion für Δ_x wenn H_0 war ist
- Wenn $p < \alpha$, wird H_0 verworfen und H_1 angenommen
- Stichprobenkennwerteverteilung:
  Verteilung der Mittelwertsdifferenzen unter ***H_0***
- Wie verteilen sich empirische Mittelwertsdifferenzen, wenn man sehr oft Stichproben zieht?



- Verteilung von Mittelwerts-differenzen bei großen Stichproben normalverteilt

Standardfehler (Stichprobenfehler)

- Streuungsmaß einer Schätzfunktion
- Definiert über die Standardabweichung einer Schätzfunktion f
- Bei einem erwartungstreuen Schätzer f ist Standardfehler ein Maß für die durchschnittliche Abweichung des geschätzten Parameterwertes vom wahren Parameterwert
- Je kleiner der Standardfehler, desto genauer ist ein unbekannter Parameter mit Hilfe einer Schätzfunktion f bestimmbar
- Abhängig vom Stichprobenumfang und der Varianz σ^2 in der Grundgesamtheit:
  - Je größer der Stichprobenumfang, desto kleiner der Standardfehler; je kleiner die Varianz, desto kleiner der Standardfehler

Standardfehler

Der Standardfehler des arithmetischen Mittels ist gleich

$$\sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}},$$

wobei σ die Standardabweichung einer einzelnen Messung bezeichnet.

Korrigierte Stichprobenvarianz: Warum $n-1$?

$$
\begin{aligned}
\mathrm{E}(S_1^2) &= \frac{1}{n}\mathrm{E}\left(\sum_{i=1}^{n}(X_i-\bar{X})^2\right) = \frac{1}{n}\mathrm{E}\left(\sum_{i=1}^{n}(X_i-\mu+\mu-\bar{X})^2\right)\\
&= \frac{1}{n}\mathrm{E}\left(\sum_{i=1}^{n}(X_i-\mu)^2 - 2(X_i-\mu)(\bar{X}-\mu) + (\bar{X}-\mu)^2\right)\\
&= \frac{1}{n}\mathrm{E}\left(\sum_{i=1}^{n}(X_i-\mu)^2 - 2\sum_{i=1}^{n}(X_i-\mu)(\bar{X}-\mu) + n(\bar{X}-\mu)^2\right)\\
&= \frac{1}{n}\mathrm{E}\left(\sum_{i=1}^{n}(X_i-\mu)^2 - 2(\bar{X}-\mu)\sum_{i=1}^{n}(X_i-\mu) + n(\bar{X}-\mu)^2\right)\\
&= \frac{1}{n}\mathrm{E}\left(\sum_{i=1}^{n}(X_i-\mu)^2 - 2n(\bar{X}-\mu)(\bar{X}-\mu) + n(\bar{X}-\mu)^2\right)\\
&= \frac{1}{n}\mathrm{E}\left(\sum_{i=1}^{n}(X_i-\mu)^2 - n(\bar{X}-\mu)^2\right) = \frac{1}{n}\left(\sum_{i=1}^{n}\mathrm{E}\big((X_i-\mu)^2\big) - n\cdot\mathrm{E}\big((\bar{X}-\mu)^2\big)\right)\\
&= \frac{1}{n}\big(n\cdot\mathrm{Var}(X) - n\cdot\mathrm{Var}(\bar{X})\big) = \mathrm{Var}(X) - \mathrm{Var}(\bar{X}) = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2
\end{aligned}
$$

Standardfehler der Kennwerteverteilung

- Hängt von den Varianzen und den Größen der beiden Teilstichproben ab:

$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\hat{\sigma}_1^2}{N_1} + \frac{\hat{\sigma}_2^2}{N_2}}$$

- Benötigt, um gefundene Mittelwertsdifferenz interpretieren zu können

t-Verteilung

- Empirische (gefundene) Mittelwertsdifferenz durch Standardfehler dividiert ergibt sog. t-Verteilung

$$t_{df} = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}}$$

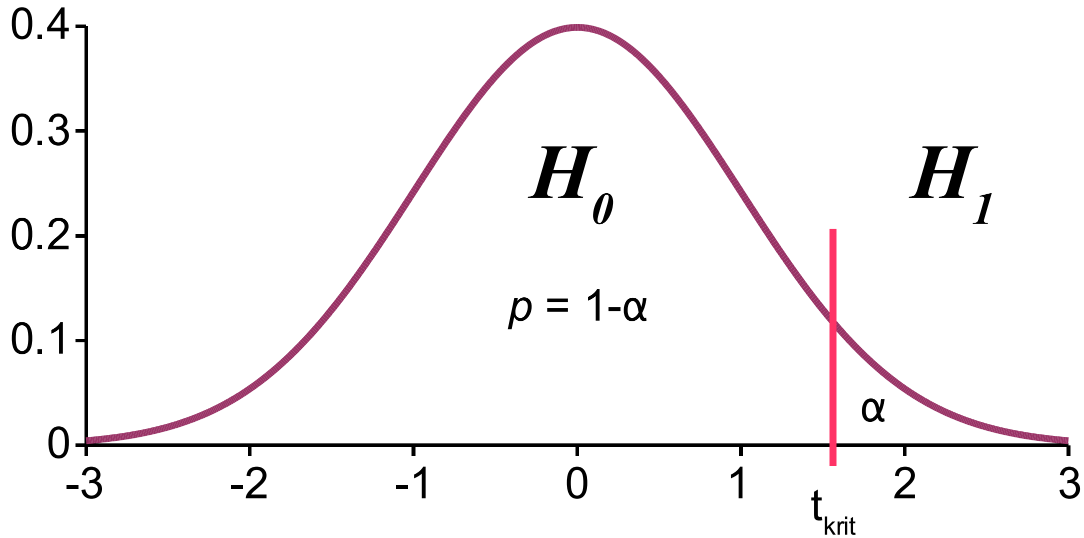
- Die genaue Form der t-Verteilung hängt von deren Freiheitsgraden (*df = degree of freedom*) ab

$$df = N_1 + N_2 - 2$$

  - Bei *df* >120 nahezu identisch mit *z*-Verteilung (St.Norm.V.)
  - Je kleiner df, desto schmalgipfliger die t-Verteilung
- Die Herleitung der Dichtefunktion und der kumulativen Funktion erfolgt später

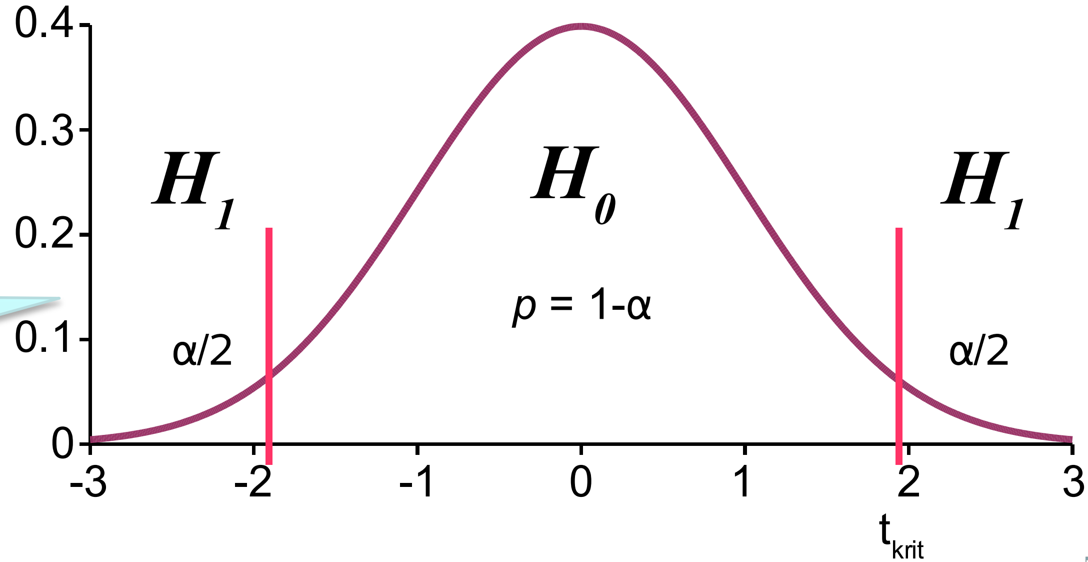
Der *t*-Test für unabhängige Stichproben

einseitiger Test
(gerichtete H_0)



zweiseitiger Test
(ungerichtete H_0)

Hier wird 1-α als
p bezeichnet!

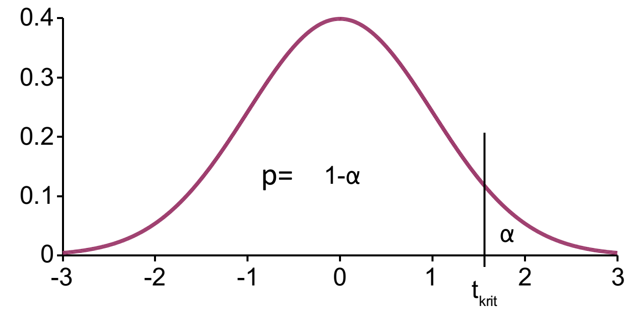


Entscheidung über die Nullhypothese

- Mittels einer *t*-Tabelle wird der empirische (gefundene) *t*-Wert interpretiert
- Dazu wird ein kritischer Wert aus der t-Tabelle entnommen
  - Der kritische *t*-Wert hängt dabei ab:
    - von den Freiheitsgraden,
    - von dem gewählten α-Niveau
    - von der Art des Tests (einseitig vs. zweiseitig)
  - Der kritische *t*-Wert definiert die Grenze des Bereichs für den empirischen *t*-Wert, ab dem H_0 verworfen wird

Die *t*-Verteilung

df	p=.800	p=.900	p=.950	p=.975	p=.990	p=.995
1	1,376	3,078	6,314	12,706	31,821	63,657
2	1,061	2,920	2,920	4,303	6,965	9,925
3	0,978	2,353	2,353	3,182	4,541	5,841
4	0,941	2,132	2,132	2,776	3,747	4,604
5	0,920	2,015	2,015	2,571	3,365	4,032
6	0,906	1,943	1,943	2,447	3,143	3,707
7	0,896	1,895	1,895	2,365	2,998	3,499
8	0,889	1,860	1,860	2,306	2,896	3,355
9	0,883	1,833	1,833	2,262	2,821	3,250
10	0,879	1,812	1,812	2,228	2,764	3,169
20	0,860	1,725	1,725	2,086	2,528	2,845
30	0,854	1,697	1,697	2,042	2,457	2,750
40	0,851	1,684	1,684	2,021	2,423	2,704
50	0,849	1,676	1,676	2,009	2,403	2,678
60	0,848	1,671	1,671	2,000	2,390	2,660
70	0,847	1,667	1,667	1,994	2,381	2,648
80	0,846	1,664	1,664	1,990	2,374	2,639
90	0,846	1,662	1,662	1,987	2,368	2,632
100	0,845	1,660	1,660	1,984	2,364	2,626
200	0,843	1,653	1,653	1,972	2,345	2,601
1000	0,842	1,646	1,646	1,962	2,330	2,581



Kritische t-Werte:

α = .05, einseitig, *df*=100:

$t_{krit}(100) = 1.66$

α = .05, zweiseitig, *df*=100:

$t_{krit}(100) = 1.98$

α = .01, einseitig, *df*=100:

$t_{krit}(100) = 2.36$

Der *t*-Test für unabhängige Stichproben

Entscheidungsregeln

- Einseitiger Test:
  - Wenn $t_{emp} > t_{krit}$ wird H_0 verworfen

- Zweiseitiger Test
  - Wenn $|t_{emp}| > t_{krit}$ wird H_0 verworfen

- In der *t*-Tabelle werden immer Werte für den einseitigen Test angegeben.
- Für einen 2-seitigen Test muss t_{krit} so gewählt werden, dass ein Bereich von $\alpha/2$ „von der Verteilung abgeschnitten wird“

Voraussetzungen

(1) Variable besitzt Intervallskala (arithm. Mittel ist definiert)

(2) Normalverteilung des Merkmals in der Grundgesamtheit

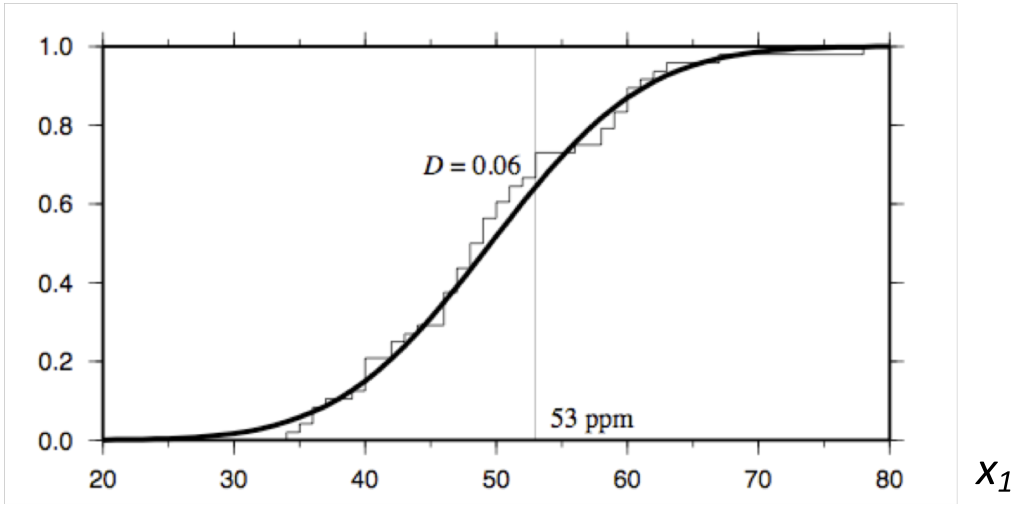
(3) „Varianzhomogenität“

- „Gleiche“ Varianzen des Merkmals in beiden Populationen
- „Varianz der Varianz“ klein

(4) Unabhängigkeit der Stichproben

Normalverteilung des Merkmals in Grundgesamtheit

- Normalverteilungsannahme für X statistisch überprüfbar
  - Kolmogorov-Smirnov-Test: $X \sim N(\mu,\sigma)$ ist H_0
    - Sortiere Stichprobendaten $x^T = [x_1, x_2]$ nach x_1 aufsteigend
    - Bestimme maximale Differenz D der kumulativen Verteilungen
    - Vergleiche ob D > KS-Wert bzgl. gewähltem Signifikanzniveau α (KS-Wert aus KS-Tabelle bestimmbar)



    - Herleitung der KS-Tabelle in höherem Semester

Varianzhomogenität

- Auch Varianzhomogenität kann statistisch überprüft werden (Levene-Test)

Nullhypothese: $H_0: \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$

Alternativhypothese: $H_1: \sigma_i^2 \neq \sigma_j^2$ für *mindestens ein* Gruppenpaar i, j mit $i \neq j$

Befindet sich der p-Wert des Tests unter einem zuvor bestimmten Niveau, so sind die Unterschiede in den Varianzen der Stichproben überzufällig (signifikant) und die Nullhypothese der Varianzgleichheit kann abgelehnt werden.[2]

Testverteilung für Levene-Test wird später hergeleitet

- Bei einem signifikanten Ergebnis (p < .05), werden die Freiheitsgrade des Tests „korrigiert“

$$df_{corr} = \frac{1}{\frac{c^2}{N_1 - 1} + \frac{(1-c^2)}{N_2 - 1}}$$

$$mit \quad c = \frac{\hat{\sigma}_{x_1}^2}{\hat{\sigma}_{x_1}^2 + \hat{\sigma}_{x2}^2}$$

Zusammenfassung: *t*-Test für unabhängige Stichproben

(1) Formulierung der (inhaltlichen und statistischen) Hypothesen
   - gerichtet oder ungerichtet?

(2) Erfassung des Merkmals in zwei unabhängigen Stichproben

(3) Berechnung der Mittelwerte in beiden Stichproben

(4) Schätzung der Populationsvarianz

(5) Berechnung des Standardfehlers der Mittelwertsdifferenz

(6) Berechnung des empirischen *t*-Werts

(7) Bestimmung des kritischen *t*-Werts
   - aus df, α, und Art des Tests

(8) Entscheidung für H_0 oder H_1

Unterschiedshypothesen: Abhängige Stichproben

- Ziehung eines Merkmalsträgers in die erste Stichprobe beeinflusst die Zugehörigkeit eines Merkmalsträgers zur zweiten Stichprobe
- Werte zweier Stichproben paarweise zugeordnet.
  - Beide Teilstichproben immer gleich groß!
- Messwiederholung
  - Gleiches Merkmal zweimal (oder mehrmals) bei den gleichen Personen erhoben
- Parallelisierung
  - Jeweils ähnliche 2 Personen einander zugeordnet
- Matching
  - Jeder Person der Stichprobe 1 ist einer Person der Stichprobe 2 zugeordnet

Abhängige Stichproben: Beispielrechnung

- Verändert sich die Einstellung zum Studienfach Informatik innerhalb der ersten 6 Wochen des Studiums?
- Abh. Variable: Einstellung zum Studium Informatik (Wertebereich 5 bis 25)
- Unabh. Variable: Messzeitpunkt (1. Woche vs. 6. Woche)

Versuchs-person	1. Woche	6. Woche
1	16	20
2	18	19
3	23	23
4	14	16
	…	…
mean	19.67	18.98

Beispielrechnung

- Für jede Person kann die Differenz der Messwerte berechnet werden (Einstellungsänderung)

Vp	1. Woche	6. Woche	$D=x_2-x_1$
1	16	20	4
2	18	19	1
3	23	23	0
4	16	14	-2
	…	…	…
mean	19.67	18.98	.68

Hypothesen

- Die statistischen Hypothesen des t-Tests für abhängige Stichproben beziehen sich auf den Mittelwert der Differenzen aller Personen
  - Vorteil: Es ist nun unerheblich, ob innerhalb der Messzeitpunkte große Varianz gegeben ist.
- Ungerichtete Hypothese:
  - H_0: $\mu_d = 0$
  - H_1: $\mu_d \neq 0$
- Gerichtete Hypothese (1):
  - H_0: $\mu_d \leq 0$
  - H_1: $\mu_d > 0$
- Gerichtete Hypothese (2):
  - H_0: $\mu_d \geq 0$
  - H_1: $\mu_d < 0$

Wdh.: Standardfehler bei Kennwert = Mittelwert

- Diese „Verteilung der der Mittelwerte“ ist selbst wieder normalverteilt (wenn das Merkmal normalverteilt ist)
- Der Mittelwert der Stichprobenkennwerteverteilung entspricht dem Mittelwert in der Population
- Die Streuung der Stichprobenkennwerteverteilung wird als Standardfehler (des Mittelwerts) bezeichnet.
  - Der Standardfehler gibt an, wie nah ein empirischer Stichprobenmittelwert am wahren Populationsmittelwert liegt
  - Dieser Standardfehler des Mittelwertes kann auch aus einer einzigen Stichprobe geschätzt werden:

$$\hat{\sigma}_{\bar{x}} = \sqrt{\frac{\hat{\sigma}_x^{\;2}}{N}} = \frac{\hat{\sigma}_x}{\sqrt{N}}$$

Standardfehler und *t*-Wert

- Um die empirisch gefundene Differenz beurteilen zu können, wird der Standardfehler benötigt

$$\hat{\sigma}_{\bar{x}_d} = \frac{\hat{\sigma}_{x_d}}{\sqrt{N}} \quad mit \quad \hat{\sigma}_{x_d} = \sqrt{\frac{\sum_{i=1}^{N}(x_{di} - \bar{x}_d)^2}{N-1}}$$

Basierend auf korrigierter Stichprobenvarianz

- Mit dem Standardfehler kann nun ein empirischer normalisierter *t*-Wert berechnet werden
  - Normalisierung bzgl. Standardabweichung (vgl. z-Standardisierung)

$$z_i = \frac{x_i - \bar{x}}{\hat{\sigma}}$$

$$t_{df} = \frac{\bar{x}_d}{\hat{\sigma}_{\bar{x}_d}} \text{ mit } df = N-1$$

Standardfehler und *t*-Wert

Im Beispieldatensatz:

$$\bar{x}_d = 0.68$$

$$\hat{\sigma}_{x_d} = 2.78$$

$$N = 60$$

- Es ergibt sich :

$$\hat{\sigma}_{\bar{x}_d} = \frac{2.78}{\sqrt{60}} = 0.36$$

$$t_{59} = \frac{0.68}{0.36} = 1.89$$

Kritischer *t*-Wert & Interpretation

- $T_{emp,59} = 1.89$
- $T_{krit,59} = ?$
  - Offene Fragestellung ⇨ zweiseitiger Test
  - $\alpha = .05$

- Interpretation:
  - $t_{emp} < t_{krit}$
  - Also: Kein bedeutsamer Unterschied!

df	p=.800	p=.900	p=.950	p=.975	p=.990	p=.995
1	1,376	3,078	6,314	12,706	31,821	63,657
2	1,061	2,920	2,920	4,303	6,965	9,925
3	0,978	2,353	2,353	3,182	4,541	5,841
4	0,941	2,132	2,132	2,776	3,747	4,604
5	0,920	2,015	2,015	2,571	3,365	4,032
6	0,906	1,943	1,943	2,447	3,143	3,707
7	0,896	1,895	1,895	2,365	2,998	3,499
8	0,889	1,860	1,860	2,306	2,896	3,355
9	0,883	1,833	1,833	2,262	2,821	3,250
10	0,879	1,812	1,812	2,228	2,764	3,169
20	0,860	1,725	1,725	2,086	2,528	2,845
30	0,854	1,697	1,697	2,042	2,457	2,750
40	0,851	1,684	1,684	2,021	2,423	2,704
50	0,849	1,676	1,676	2,009	2,403	2,678
60	0,848	1,671	1,671	2,000	2,390	2,660
70	0,847	1,667	1,667	1,994	2,381	2,648
80	0,846	1,664	1,664	1,990	2,374	2,639
90	0,846	1,662	1,662	1,987	2,368	2,632
100	0,845	1,660	1,660	1,984	2,364	2,626
200	0,843	1,653	1,653	1,972	2,345	2,601
1000	0,842	1,646	1,646	1,962	2,330	2,581

Eingruppen *t*-Test

- Ziel: Vergleich des Mittelwerts einer Stichprobe mit einem vorgegebenen (konstanten) Wert
- Beispiele:
  - Prüfe, ob eine bestimmte Personengruppe sich in ihrer Intelligenz vom Populationsmittelwert (100) unterscheidet
  - Prüfe, ob sich die tatsächliche Studiendauer von der Regelstudienzeit unterscheidet
  - Prüfe, ob sich die Differenz von Reaktionszeiten unter zwei Bedingungen von Null unterscheidet

Eingruppen *t*-Test

Voraussetzungen

- *Normalverteilung* des Merkmals
- *Intervallskalenniveau* des Merkmals
- Es handelt sich um eine *Zufallsstichprobe*

Eingruppen *t*-Test

Statistische Hypothesen

- Ungerichtete Hypothese:
  - H_0: $\mu = c$
  - H_1: $\mu \neq c$
- Gerichtete Hypothese (1):
  - H_0: $\mu \leq c$
  - H_1: $\mu > c$
- Gerichtete Hypothese (2):
  - H_0: $\mu \geq c$
  - H_1: $\mu < c$

Standardfehler und *t*-Wert

- Berechnung des Standardfehlers

$$\hat{\sigma}_{\bar{x}} = \frac{\hat{\sigma}_x}{\sqrt{N}}$$

- Berechnung des t-Werts

$$t(df = N - 1) = \frac{\bar{x} - c}{\hat{\sigma}_{\bar{x}}}$$

Beispiel

- Liegt der IQ der Kinder, die als hochbegabten klassifiziert werden, wirklich über dem Populationsmittelwert (100)?
- Hypothesen:
  - H_0: $\mu \leq 100$
  - H_1: $\mu > 100$
- Stichprobenkennwerte bei *N*=10:
  - Mittelwert: 108.50
  - Standardabweichung: 14.35

$$\hat{\sigma}_{\bar{x}} = \frac{14.35}{\sqrt{10}} = 4.54 \qquad t(9) = \frac{108.5 - 100}{4.54} = 1.87$$

Beispiel

- $t_{emp}(9) = 1.87$
- $t_{krit}(9) = ?$
  - Gerichtete Fragestellung ⇨ einseitiger Test
  - $\alpha = .05$

df	p=.800	p=.900	p=.950	p=.975	p=.990	p=.995
1	1,376	3,078	6,314	12,706	31,821	63,657
2	1,061	2,920	2,920	4,303	6,965	9,925
3	0,978	2,353	2,353	3,182	4,541	5,841
4	0,941	2,132	2,132	2,776	3,747	4,604
5	0,920	2,015	2,015	2,571	3,365	4,032
6	0,906	1,943	1,943	2,447	3,143	3,707
7	0,896	1,895	1,895	2,365	2,998	3,499
8	0,889	1,860	1,860	2,306	2,896	3,355
9	0,883	1,833	1,833	2,262	2,821	3,250
10	0,879	1,812	1,812	2,228	2,764	3,169
20	0,860	1,725	1,725	2,086	2,528	2,845
30	0,854	1,697	1,697	2,042	2,457	2,750
40	0,851	1,684	1,684	2,021	2,423	2,704
50	0,849	1,676	1,676	2,009	2,403	2,678
60	0,848	1,671	1,671	2,000	2,390	2,660
70	0,847	1,667	1,667	1,994	2,381	2,648
80	0,846	1,664	1,664	1,990	2,374	2,639
90	0,846	1,662	1,662	1,987	2,368	2,632
100	0,845	1,660	1,660	1,984	2,364	2,626
200	0,843	1,653	1,653	1,972	2,345	2,601
1000	0,842	1,646	1,646	1,962	2,330	2,581

- Interpretation:
  - $t_{emp} > t_{krit}$
  - H_0 wird verworfen

Vergleich der 3 Arten des *t*-Tests

	unabhängige Stichproben	**abhängige Stichproben**	**Eingruppen *t*-Test**
Fragestellung			
Voraus-setzungen			

Vergleich der 3 Arten des *t*-Tests

	Unabhängige Stichproben	Abhängige Stichproben	Eingruppen *t*-Test
Ungerichtete Hypothese			
Gerichtete Hypothese			

Vergleich der 3 Arten des *t*-Tests

	Unabhängige Stichproben	Abhängige Stichproben	Eingruppen *t*-Test
Kennwert des Tests			
Standardfehler des Kennwerts			
t-Wert			
Freiheitsgrade			

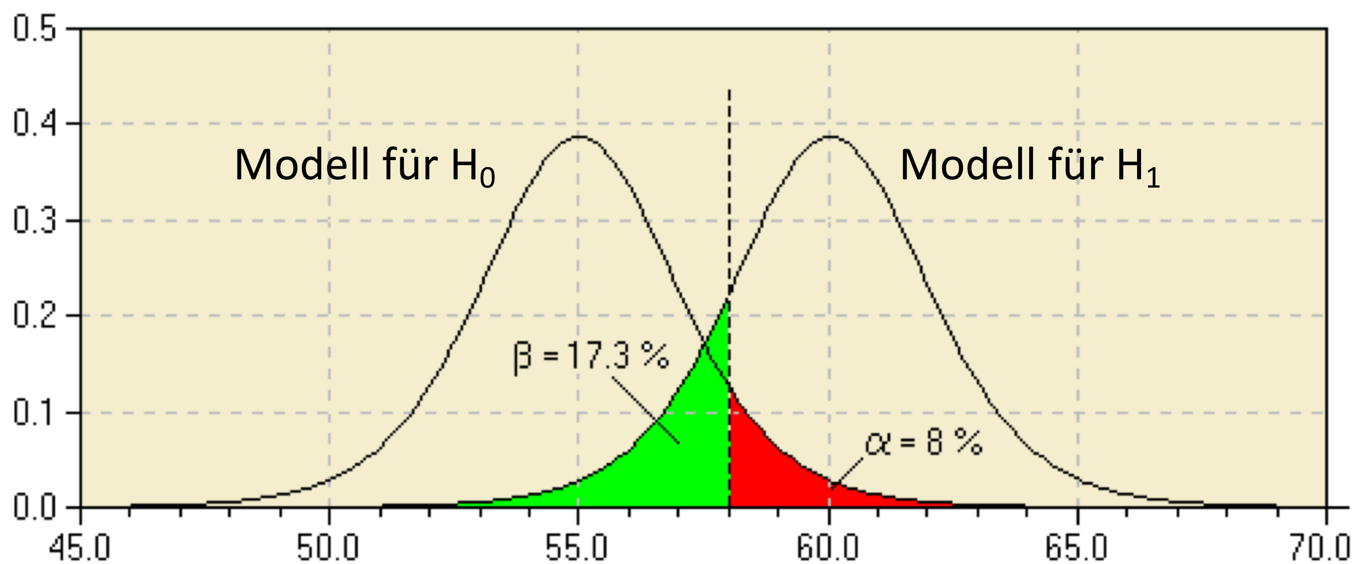
Trennschärfe (Power) eines Tests

- Mit welcher Wahrscheinlichkeit weist ein statistischer Test die abzulehnende Nullhypothese H_0 („Es gibt keinen Unterschied“) korrekt zurück, wenn die Alternativhypothese H_1 („Es gibt einen Unterschied“) wahr ist
- Interpretiert als „Ablehnungskraft“ des Tests
  - Hohe Trennschärfe des Tests spricht gegen, niedrige Trennschärfe für die Nullhypothese H_0
- Ziel: den Ablehnungsbereich A so bestimmen, dass die Wahrscheinlichkeit für die Ablehnung einer „falschen Nullhypothese“ H_0, d. h. für Annahme der Alternativhypothese H_1 unter der Bedingung, dass H_1 wahr ist, möglichst groß ist

Trennschärfe

	H_0 ist wahr	H_1 ist wahr
Durch einen statistischen Test fällt eine **Entscheidung für** H_0	Richtige Entscheidung (Spezifität) Wahrscheinlichkeit: 1 - α	**Fehler 2. Art** Wahrscheinlichkeit: β
Durch einen statistischen Test fällt eine **Entscheidung für** H_1	**Fehler 1. Art** Wahrscheinlichkeit: α	richtige Entscheidung Wahrscheinlichkeit: 1-β (Trennschärfe des Tests, Sensitivität)

- Trennschärfe hat den Wert 1-β
- Wahl des β-Niveaus?
- Modell für H_1 benötigt



Determinanten der Trennschärfe

Die Trennschärfe (1-β) wird größer:[6]

- mit wachsender Differenz von $(\mu_0 - \mu_1)$ (das bedeutet: ein großer Unterschied zwischen zwei Teilpopulationen wird seltener übersehen als ein kleiner Unterschied)
- mit kleiner werdender Merkmalsstreuung σ
- mit größer werdendem Signifikanzniveau α (sofern β nicht festgelegt ist)
- mit wachsendem Stichprobenumfang, da der Standardfehler dann kleiner wird: $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$
- bei einseitigen Tests im Vergleich zu zweiseitigen Tests: Für den zweiseitigen Test braucht man einen etwa um 25 % größeren Stichprobenumfang, um dieselbe Trennschärfe wie für den einseitigen Test zu erreichen.

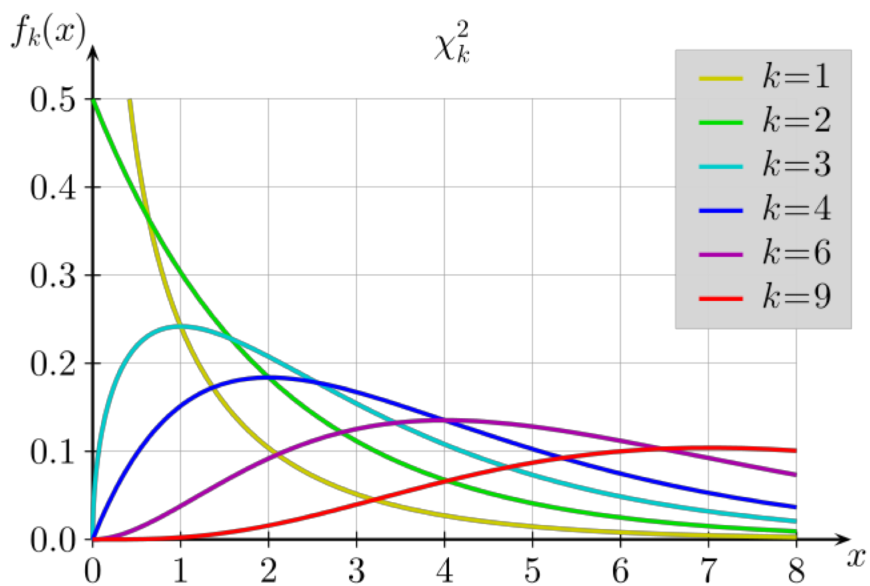
[Wikipedia]

Testverfahren

- ***Parametrische Verfahren***
  - Beteiligte Variablen müssen geforderte Verteilungsform aufweisen (z.B. Normalverteilung für den *t*-Test)
  - Dann aber gute "Aussagekraft“ (siehe den Begriff der Trennschärfe)
- ***Nonparametrische Verfahren*** werden eingesetzt…
  - ⇨ Für die Analyse von ordinal- oder nominalskalierten Variablen
  - ⇨ Wenn die Normalverteilungsannahme des Gesamtmerkmals verletzt ist
    - ⇨ Beispiel: Summe der Quadrate von k normalverteilten Zufallsvariablen ist nicht normalverteilt

χ^2-Verteilung

χ^2 ist eine der Verteilungen, die aus der Normalverteilung $\mathcal{N}\left(\mu, \sigma^2\right)$ abgeleitet werden kann: Hat man k Zufallsvariablen Z_i, die unabhängig und standardnormalverteilt sind, so ist die Chi-Quadrat-Verteilung mit k Freiheitsgraden definiert als die Verteilung der Summe der quadrierten Zufallsvariablen $Z_1^2 + \cdots + Z_k^2$



Der χ² -Test

- Der ***χ²-Test*** („Chi-Quadrat-Test“) dient dem Vergleich von *beobachteten* und *erwarteten* Häufigkeiten. Er kann eingesetzt werden, wenn 1 oder 2 nominalskalierte unabhängige Variablen vorliegen.

	Merkmal		
	Auspr. 1	...	Auspr. k
Beobachtet	$f_{b,1}$		$f_{b,k}$
Erwartet	$f_{e.1}$		$f_{e,k}$

Beispiele:

- Leiden Männer und Frauen gleich häufig an einer bestimmten Erkrankung?
- Leisten hoch-ängstlich und gering-ängstliche Personen gleich häufig Hilfe in einer Notsituation?

Der χ^2 -Test

Voraussetzung für den χ^2 -Test (Faustregeln)

(1) Weniger als 1/5 aller Zellen hat eine *erwartete Häufigkeit* kleiner als 5.

(2) Keine Zelle weist eine *erwartete Häufigkeit* kleiner als 1 auf.

Wenn Voraussetzungen nicht erfüllt → andere Tests

Der χ^2 -Test

χ^2 -Test – Beispiel 1

- Es soll geprüft werden, ob die Verteilung von Männern und Frauen in einer Gruppe signifikant von einer Gleichverteilung abweicht
- $N = 76$ (Frauen: 56; Männer: 20)
- Statistische Hypothesen
  - $\boldsymbol{H_0}$: $\pi(\text{Frau}) = \pi(\text{Mann})$
  - $\boldsymbol{H_1}$: $\pi(\text{Frau}) \neq \pi(\text{Mann})$

Der χ^2 -Test

Schritt 1:

- Zunächst werden die nach der $\boldsymbol{H_0}$ zu erwartenden Häufigkeiten berechnet:
- *Beobachtet:* $N_F = 56$; $N_M = 20$
- *Erwartet*: ???
  - Gesamtzahl: 76
  - Bei einer Gleichverteilung wären also 3 Männer und 3 Frauen zu erwarten.

Der χ² -Test

Schritt 2:

- Nun wird der (empirische) χ²-Wert berechnet:

$$\chi^2_{df=k-1} = \sum_{i=1}^{k} \frac{\left(f_{b,i} - f_{e,i}\right)^2}{f_{e,i}}$$

	Merkmal		
	Auspr. 1	...	Auspr. k
Beobachtet	$f_{b,1}$		$f_{b,k}$
Erwartet	$f_{e.1}$		$f_{e,k}$

mit:

- k: Anzahl der Stufen der beiden Variablen
- $f_{b,i}$: Beobachtete Häufigkeit in der Zelle (i)
- $f_{e,i}$: Erwartete Häufigkeit in der Zelle (i)

Der χ² -Test

	Geschlecht		
	Frau	Mann	
Beobachtet	56	20	76
Erwartet	38	38	76

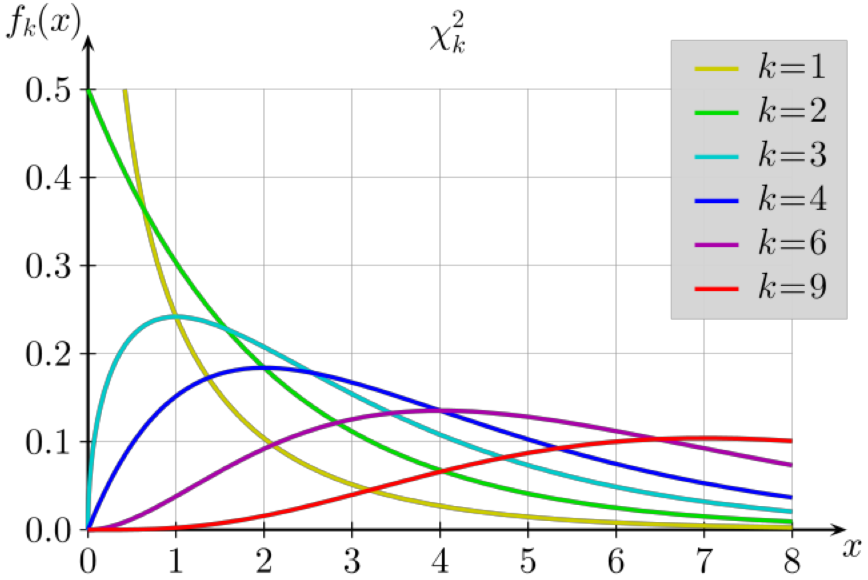
$$\chi^2_{df=k-1} = \sum_{i=1}^{k} \frac{\left(f_{b,i} - f_{e,i}\right)^2}{f_{e,i}}$$

$$\chi^2_{df=1} = \frac{(56-38)^2}{38} + \frac{(20-38)^2}{38} = \frac{18^2}{38} + \frac{(-18)^2}{38} = 8.53 + 8.53 = 17.05$$

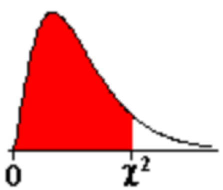
Der χ² -Test

- ***Schritt 3***: Vergleich des empirischen χ^2-Werts mit dem kritischen χ^2-Wert.

- Der kritische χ^2-Wert wird in Abhängigkeit von den Freiheitsgraden k und dem gewählten *α-Niveau* aus einer Tabelle zur *χ^2-Verteilung* abgelesen



χ^2 -Tabelle



Lesebeispiel: Gesucht sei der χ^2-Wert, unter dem bei df=10 Freiheitsgraden 95% aller möglichen Werte einer χ^2-verteilten Zufallsvariablen X^2 liegen. In der Zeile für df=10 finden Sie in der Spalte $1\text{-}\alpha$ =0,95 den gesuchten Wert χ^2= 18,31.

df	**(rote/dunkle) Fläche** $1\text{-}\alpha$								
	0,7	**0,75**	**0,8**	**0,85**	**0,9**	**0,95**	**0,975**	**0,99**	**0,995**
1	1,07	1,32	1,64	2,07	2,71	3,84	5,02	6,63	7,88
2	2,41	2,77	3,22	3,79	4,61	5,99	7,38	9,21	10,60
3	3,66	4,11	4,64	5,32	6,25	7,81	9,35	11,34	12,84
4	4,88	5,39	5,99	6,74	7,78	9,49	11,14	13,28	14,86
5	6,06	6,63	7,29	8,12	9,24	11,07	12,83	15,09	16,75
6	7,23	7,84	8,56	9,45	10,64	12,59	14,45	16,81	18,55
7	8,38	9,04	9,80	10,75	12,02	14,07	16,01	18,48	20,28
8	9,52	10,22	11,03	12,03	13,36	15,51	17,53	20,09	21,95
9	10,66	11,39	12,24	13,29	14,68	16,92	19,02	21,67	23,59
10	11,78	12,55	13,44	14,53	15,99	18,31	20,48	23,21	25,19

Der χ^2 -Test

- ***Schritt 3***: Vergleich des empirischen χ^2-Werts mit dem kritischen χ^2-Wert.

- Für α=.05 ergibt sich bei *df=1*:

$$\chi^2_{emp} = 17.05$$

$$\chi^2_{krit} = 3.84$$

- Die H_0 muss verworfen werden; folglich kann ein Unterschied nachgewiesen werden.

Der χ^2 -Test

χ^2 -Test – Beispiel 2

	Geschlecht		
Angst	Frau	Mann	
gering	25	14	39
hoch	33	6	39
	58	20	78

- Frage: Ist die (relative) Häufigkeit hoher bzw. geringer Ängstlichkeit bei Männern und Frauen gleich?
- Statistische Hypothesen
  - H_0: π(Angst | Frau) = π(Angst | Mann)
  - H_1: π(Angst | Frau) ≠ π(Angst | Mann)

Der χ^2 -Test

Schritt 1: Zunächst werden aus den Randsummen die nach der H_0 zu erwarteten Häufigkeiten geschätzt:

Beobachtet:

	Geschlecht		
Angst	Frau	Mann	
gering	25	14	39
hoch	33	6	39
	58	20	78

$$f_{e(i,j)} = \frac{f_{b(i.)}}{N} \cdot \frac{f_{b(.j)}}{N} \cdot N$$

$$= \frac{f_{b(i.)} \cdot f_{b(.j)}}{N}$$

Erwartet:

	Geschlecht		
Angst	Frau	Mann	
gering	29	10	39
hoch	29	10	39
	58	20	78

b(1 ,) = 39 b(, 1) = 58
N = 78
(39 · 58) / 78 = 29

Der χ² -Test

Schritt 2: Nun wird der (empirische) χ^2-Wert berechnet:

$$\chi^2_{df=(k-1)\cdot(l-1)} = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{\left(f_{b(i,j)} - f_{e(i,j)}\right)^2}{f_{e(i,j)}}$$

mit:

- *k, l:* Anzahl der Stufen der beiden Variablen
- $f_{b(i,j)}$: Beobachtete Häufigkeit in der Zelle (i,j)
- $f_{e(i,j)}$: Erwartete Häufigkeit in der Zelle (i,j)

Der χ^2 -Test

Beobachtet:

	Geschlecht		
Angst	Frau	Mann	
gering	25	14	39
hoch	33	6	39
	58	20	78

Erwartet:

	Geschlecht		
Angst	Frau	Mann	
gering	29	10	39
hoch	29	10	39
	58	20	78

$$\chi^2_{df=(k-1)\cdot(l-1)} = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{\left(f_{b(i,j)} - f_{e(i,j)}\right)^2}{f_{e(i,j)}}$$

$$\chi^2_{df=1} = \frac{(25-29)^2}{29} + \frac{(33-29)^2}{29} + \frac{(14-10)^2}{10} + \frac{(6-10)^2}{10}$$

$$= 0.55 + 0.55 + 1.60 + 1.60 = 4.30$$

Der χ^2 -Test

- ***Schritt 3***: Vergleich des empirischen χ^2-Werts mit dem kritischen χ^2-Wert.

df	(rote/dunkle) Fläche $1-\alpha$								
	0,7	**0,75**	**0,8**	**0,85**	**0,9**	**0,95**	**0,975**	**0,99**	**0,995**
1	1,07	1,32	1,64	2,07	2,71	3,84	5,02	6,63	7,88
2	2,41	2,77	3,22	3,79	4,61	5,99	7,38	9,21	10,60
3	3,66	4,11	4,64	5,32	6,25	7,81	9,35	11,34	12,84

- Für α=.05 ergibt sich bei *df=1*:

$$\chi^2_{emp} = 4.30$$

$$\chi^2_{krit} = 3.84$$

- Die $\boldsymbol{H_0}$ muss verworfen werden; folglich kann ein Unterschied nachgewiesen werden.

Überblick weitere Verfahren:

Stichproben	Nominalskalen	Ordinalskalen
Unabhängig	• χ^2 Test • Fisher-Yates-Test	• Mediantest • U-Test (Mann-Whitney) • H-Test (Kruskal & Wallis)
Abhängig	• McNemar-Test • Cochran-Test	• Vorzeichen-Test • Vorzeichen-Rang-Test (Wilkoxon) • Friedman-Test

Zusammenfassung

- *Nonparametrische Testverfahren* können verwendet werden, wenn
  a) die vorliegenden Daten kein Intervallskalenniveau aufweisen oder
  b) die Normalverteilungsannahme der parametrischen Tests verletzt ist.
- Der ***χ^2-Test*** überprüft, ob *beobachtete* und *erwartete* Häufigkeiten *signifikant* voneinander abweichen.