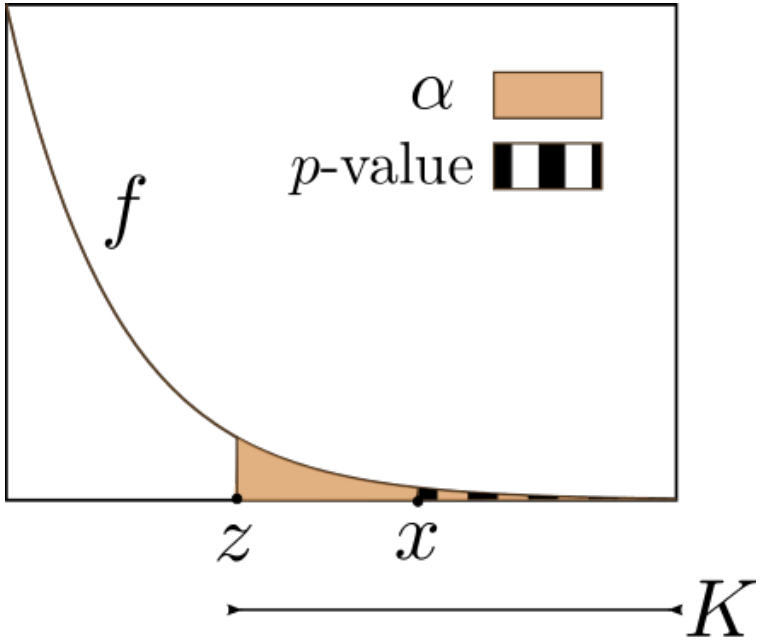
# Einführung in Web- und Data-Science

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?
- 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 $\alpha$, 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 manchen Veröffentlichungen wird leider $\alpha$ als p-Wert bezeichnet!

# Danksagung

Nachfolgende Materialen sind mit Änderungen übernommen aus:

Vorlesung Statistik (WS08/09) aus dem Studiengang Psychologie and der Universität Freiburg

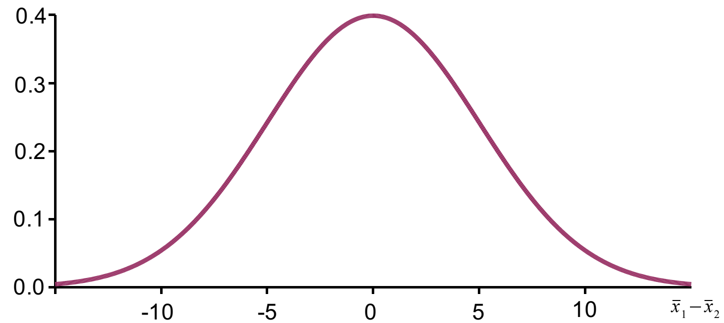
# Unterschiedshypothesen

- 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 bedingte Wahrscheinlichkeit : $p(\Delta_x \mid H_0)$
- Wenn $p < \alpha$, wird $H_0$ verworfen und $H_1$ angenommen
- Stichprobenkennwerteverteilung:
  Verteilung der Mittelwertsdifferenzen unter $\boldsymbol{H_0}$
- Wie verteilen sich empirische Mittelwertsdifferenzen, wenn man sehr oft Stichproben zieht?




- Verteilung von Mittelwerts-differenzen bei großen Stichproben normalverteilt

# Standardfehler der Kennwerteverteilung

- Hängt von den Standardabweichungen 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 Mittelwertsdifferenz
  durch Standardfehler dividiert ergibt sog. t-Verteilung

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

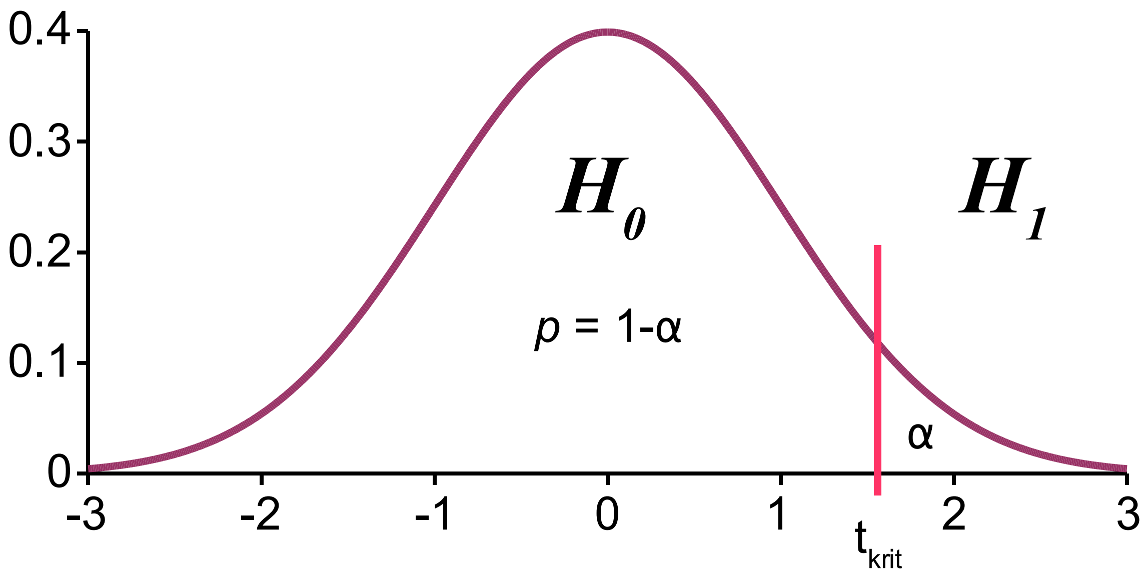
- Die genaue Form der t-Verteilung hängt von deren Freiheitsgraden (*df* = *degree of freedom*) ab

$$df = N_1 + N_2 - 2$$

  - Bei $df > 120$ nahezu identisch mit *z*-Verteilung
  - Je kleiner df, desto schmalgipfliger die t-Verteilung
- Die Herleitung der Dichtefunktion und der kumulativen Funktion erfolgt später

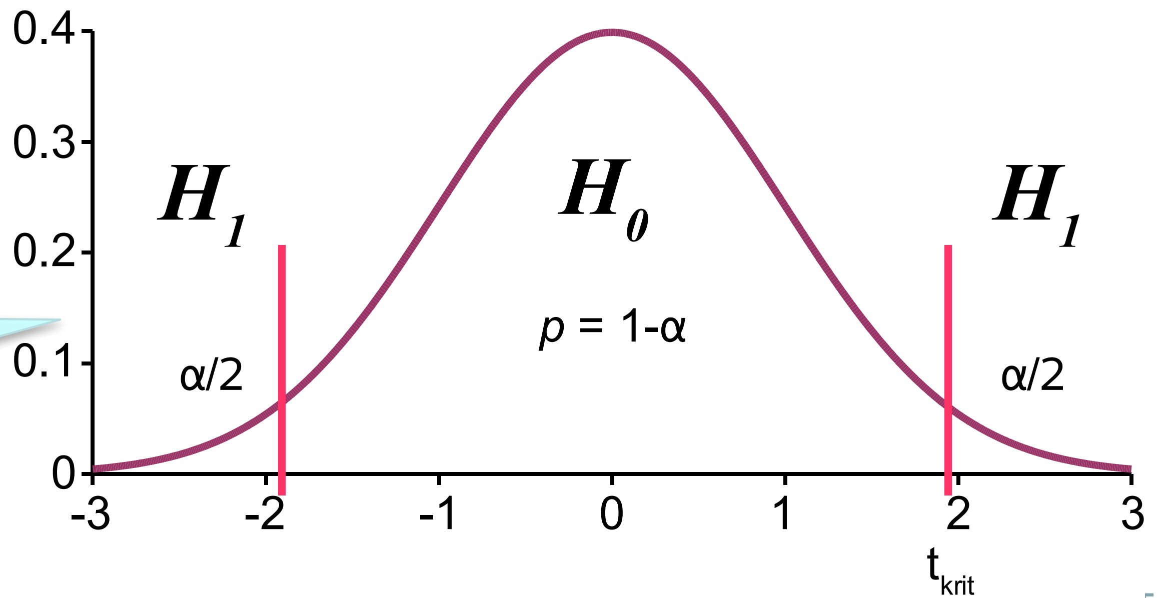
# Der *t*-Test für unabhängige Stichproben

einseitiger Test
(gerichtete $H_0$)




zweiseitiger Test
(ungerichtete $H_0$)

Hier wird 1-$\alpha$ als p-Wert bezeichnet!

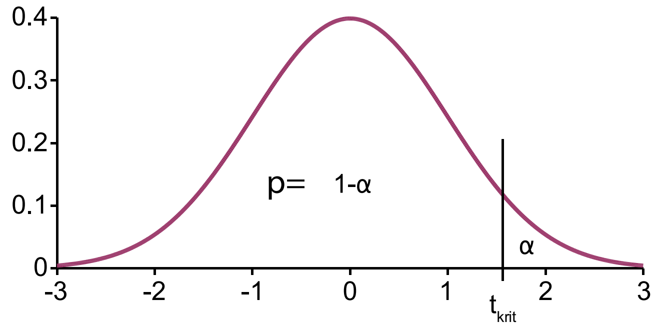


# Entscheidung über die Nullhypothese

- Mittels einer *t*-Tabelle wird der empirische *t*-Wert interpretiert
- Dazu wird ein kritischer t-Wert aus der t-Tabelle entnommen
  - Der kritische *t*-Wert hängt dabei ab:
    - von den Freiheitsgraden,
    - von dem gewählten $\alpha$-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:

$\alpha = .05$, einseitig, $df$=100:

$t_{krit}(100) = 1.66$

$\alpha = .05$, zweiseitig, $df$=100:

$t_{krit}(100) = 1.98$

$\alpha = .01$, 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 α/2 „von der Verteilung abgeschnitten wird"

# Voraussetzungen

(1) Intervallskalenniveau der Variable

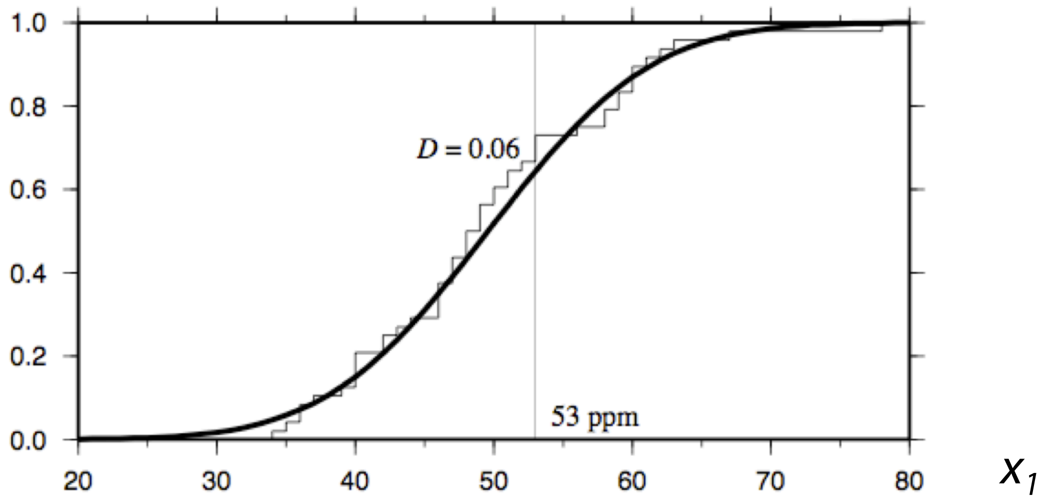
(2) Normalverteilung des Merkmals in der Grundgesamtheit

(3) „Varianzhomogenität“
(Gleiche Varianzen des Merkmals in beiden Populationen)

(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 $\alpha$ (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 $i \neq j$

Testverteilung für Levene-Test wird später hergeleitet

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]

- 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 statistische) 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 Teil 2

# 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):
  - $H_0$: $\mu_d \geq 0$
  - $H_1$: $\mu_d < 0$

# 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 *t*-Wert berechnet werden:

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

# 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:
  - Es wir überprüft, ob eine bestimmte Personengruppe sich in ihrer Intelligenz vom Populationsmittelwert (100) unterscheidet.
  - Es wird überprüft, ob sich die tatsächliche Studiendauer von der Regelstudienzeit unterscheidet.
  - Es wird überprüft, ob sich die Differenz von Reaktionszeiten unter zwei Bedingungen von Null unterscheidet.

# Eingruppen *t*-Test

## ***Voraussetzungen***

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

# Eingruppen *t*-Test

## *Statistische Hypothesen*

- Ungerichtete Hypothese:
  - $H_0$: $\mu = c$
  - $H_1$: $\mu \neq c$
- Gerichtet Hypothese (1):
  - $H_0$: $\mu \leq c$
  - $H_1$: $\mu > c$
- Gerichtet 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 | | | |

# Vergleich der 3 Arten des *t*-Tests

| | **Unabhängige Stichproben** | **Abhängige Stichproben** | **Eingruppen *t*-Test** |
|---|---|---|---|
| Kritischer *t*-Wert hängt ab von … | | | |
| $H_0$ wird verworfen, wenn … | | | |
| $H_0$ wird verworfen, wenn … | | | |

# Nonparametrische Testverfahren

**Definition:**

- ***Nonparametrische*** (***verteilungsfreie) Verfahren***
  - Keine bestimmte Verteilungsformen des erfassten Merkmals vorausgesetzt (z.B. Normalverteilung)
- ***Nonparametrische Verfahren*** werden eingesetzt…
  - ⇨ Für die Analyse von ordinal- oder nominalskalierten Variablen
  - ⇨ Wenn die Normalverteilungsannahme verletzt ist
- ***Parametrische Verfahren*** dürfen nur verwendet werden, wenn die beteiligten Variablen die geforderte Verteilungsform ausweisen (z.B. Normalverteilung für den *t*-Test)
  - Dann aber meist mehr "Aussagekraft" (Power)

# Der $\chi^2$ -Test

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

***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 $\chi^2$ -Test

***Voraussetzung für den $\chi^2$ -Test (Faustregeln)***

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

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

Wenn diese Voraussetzungen nicht erfüllt sind, gibt es andere Tests (später behandelt)

## ***$\chi^2$ -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 $\chi^2$ -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 Männer und 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 $\chi^2$ -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 $\chi^2$ -Test

- ***Schritt 3***: Vergleich des empirischen $\chi^2$-Werts mit dem kritischen $\chi^2$-Wert.
- Der kritische $\chi^2$-Wert wird in Abhängigkeit von den Freiheitsgraden und dem gewählten *α-Niveau* aus einer Tabelle zur *$\chi^2$-Verteilung* abgelesen
- Für α=.05 ergibt sich bei *df=1*:

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

$$\chi^2_{krit} = 5.02$$

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

# Der $\chi^2$ -Test

### *$\chi^2$ -Test – Beispiel 2*

| Angst | Geschlecht: 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
  - $\boldsymbol{H_0}$: $\pi(\text{Angst} \mid \text{Frau}) = \pi(\text{Angst} \mid \text{Mann})$
  - $\boldsymbol{H_1}$: $\pi(\text{Angst} \mid \text{Frau}) \neq \pi(\text{Angst} \mid \text{Mann})$

# Der $\chi^2$ -Test

***Schritt 1***: Zunächst werden aus den Randsummen die nach der $\boldsymbol{H_0}$ 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 |

# Der $\chi^2$ -Test

***Schritt 2***: Nun wird der (empirische) $\chi^2$-Wert berechnet:

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

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 χ² -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 $\chi^2$ -Test

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

- Der kritische $\chi^2$-Wert wird in Abhängigkeit von den Freiheits-graden und dem gewählten *α-Niveau* aus einer Tabelle zur *$\chi^2$-Verteilung* abgelesen (Leonhart, S.448f).

- Für α=.05 ergibt sich bei *df=1*:

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

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

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

# Überblick weitere Verfahren:

| Stichproben | Nominalskalen | Ordinalskalen |
|---|---|---|
| Unabhängig | • $\chi^2$ Test<br>• Fisher-Yates-Test | • Mediantest<br>• U-Test (Mann-Whitney)<br>• H-Test (Kruskal & Wallis) |
| Abhängig | • McNemar-Test<br>• Cochran-Test | • Vorzeichen-Test<br>• Vorzeichen-Rang-Test (Wilkoxon)<br>• Friedman-Test |

# Zusammenfassung

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