

STATISTIK – LINUS STUHLMANN

MODELWAHL UND MODELLANNAHMEN

Berücksichtigungen für eine Modellwahl sind:

- Verteilungsannahmen
- Unabhängigkeit
- Homogenität

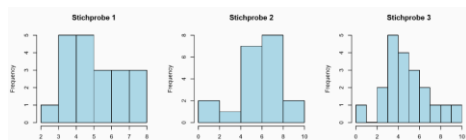
VERTEILUNGSANNAHMEN

Ob es sich um einer der bekannten parametrische Modelle wie Normalverteilung oder Exponentialverteilung handelt, kann bestimmt werden durch:

- Fachwissen
- Erfahrung mit ähnlichen Datensätzen
- Explorative Datenanalyse

Bei nicht parametrischen Modellen werden meist schwächere Annahmen getroffen, wie Symmetrie.

Histogramme können helfen die Verteilung besser zu verstehen und geben einen ersten Eindruck welche Verteilung vorliegen könnte. Es ist aber nicht immer leicht zu erkennen, wie ähnlich sie einer Verteilung sind



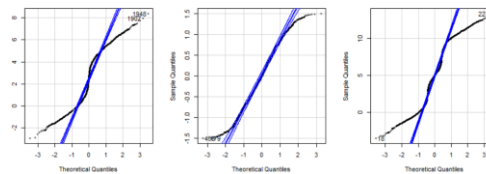
QQ-PLOTS

QQ-Plots helfen bei der bestimmung einer Verteilung, das sie die empirischen Quantile der Daten, mit jenen der Stichprobe vergleichen.

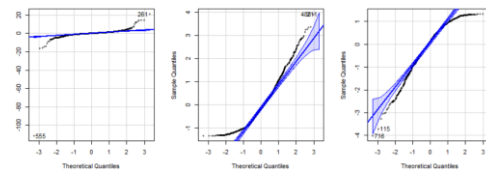
R – LIBRARY(CAR)

Library(car) - Die Datenpunkte müssen innerhalb der blauen Range (Konfidenzband) liegen.

- qqPlot(dat\$x, distribution="exp", rate=0.65)
 - Steigung 1 → Varianz übereinstimmend
 - $m = 2 \rightarrow s_n = \sigma + 1$
 - Gerade durch Ursprung (0,0) → Mittelwert übereinstimmend
 - $(0, 1) \rightarrow \bar{x} = \mu + 1$



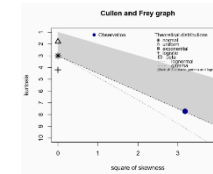
bimodal, kurzschwänzig, multimodal



langschwänzig, rechtsschief, linksschief

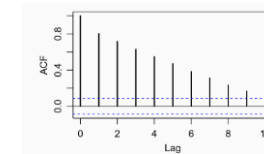
CULLEN-FREY DIAGRAMM

Das Cullen-Frey Diagramm ist eine Alternative zum QQ



UNABHÄNGIGKEIT

Können die Daten als stochastisch unabhängig angenommen werden? Dies kann durch die Auto-Korrelationsfunktion (ACF) bestimmt werden. R → acf(x).



Zeigt die empirische Korrelation $Cor(x_t, x_{t+\tau})$ gegen die Verschiebung τ (lag).

Wenn die Werte ausserhalb der blauen Range liegt, gibt es eine Abhängigkeit.

HOMOGENITÄT

Stammen die Daten alle aus derselben Verteilung und gibt es zeitliche Trends oder Ausreisser? Dies kann beispielsweise in QQ-Plots und ACF-Analysen ermittelt werden.

PUNKTSCHÄTZUNG

Meist gibt es in der Population eine bestimmte Grösse von Interesse, die aus einer Stichprobe geschätzt werden soll.

Parametrische Modelle:

- Ziel: Parameter der Verteilung bestimmen

Nichtparametrische Modelle:

- Ziel: bestimmte Kenngrössen schätzen

IID-STICHPROBE

Eine iid-Stichprobe (independent and identically distributed) ist definiert als Realisation von **unabhängig, identisch verteilten** Zufallsvariablen X_1, \dots, X_n deren Zufallsvariablen aus n Wiederholungen desselben Experiments stammen. Die Datenwerte, also jede Beobachtung x_i , ist eine konkrete Realisation der Stichprobenzufallsvariable.

- Unabhängig
- Identisch verteilt
- Unter gleichen Bedingungen erhoben

UNBEKANNTER PARAMETER θ

Parameter oder Kennzahlen, welche geschätzt werden sollen, werden als θ angegeben.

STATISTIK

Eine Statistik ist eine Funktion, die ausschließlich auf den Daten der Stichprobe basiert. Das bedeutet, sie wird berechnet oder abgeleitet aus den Daten, die man gesammelt hat.

- Stichprobenmittelwert
- Stichprobenvarianz
- Median

SCHÄTZER

Statistiken, die zur Schätzung des unbekanntem Parameters θ verwendet werden, nennt man **Schätzer**:

$$\hat{\theta} = T(X_1, \dots, X_n)$$

- Ziel
 - Schätzung für unbekanntem Parameter θ der Population
- Beispiel
 - Schätzer für Mittelwert einer Population
 - \rightarrow Mittelwert der Stichprobe

Der realisierte Wert t ist die **Schätzung**:

$$t = T(x_1, \dots, x_n)$$

Zufallsvariable	konstant
	Parameter θ
Stichprobe (X_1, \dots, X_n)	Realisierung/Daten (x_1, \dots, x_n)
Schätzer T	Schätzung t

SCHÄTZER FÜR ERWARTUNGSWERT

Schätzer für den Erwartungswert $E(X) = \mu$ der zugrundeliegenden Population.

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

- Erwartungswert von \bar{X}_n gleicht dem gesuchten Parameter: $E(\bar{X}_n) = \mu$
- Die Varianz der Stichproben wird kleiner mit grösserer Stichprobe: $Var(\bar{X}_n) = \frac{\sigma^2}{n}$
- Die genaue Verteilung von \bar{X}_n hängt von der Verteilung von X ab. Für **grosse n** gilt:
 - $\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$
 - **Zentraler Grenzwertsatz** \rightarrow Je mehr Mittelwerte gezogen werden umso näher an Normalverteilung

SCHÄTZER FÜR VARIANZ

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

Der Erwartungswert des Schätzers gleicht dem gesuchten Parameter:

$$E(S_n^2) = \sigma^2$$

GÜTEKRITERIEN FÜR SCHÄTZER

Es gibt mehrere Schätzer für den Parameter θ . Bspw. für den Erwartungswert einer Normalverteilung kann mittels der folgenden Statistiken geschätzt werden:

- **Arithmetisches Mittel**
- Median
- Mittelpunkt des Intervalls zwischen 25 - 75%

Wünschenswerte Eigenschaften für Schätzer sind:

- Erwartungstreue (keinen Bias)
- Konsistenz
- Effizienz

ERWARTUNGSTREUE

Schätzer T ist erwartungstreu (**unbiased**) für θ , falls er im Mittel den Parameter θ richtig schätzt.

$$E(T(X_1, \dots, X_n)) = \theta$$

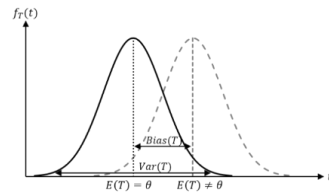
Asymptotisch erwartungstreu, falls für alle möglichen Werte von θ für steigenden Stichprobenumfang n gilt:

$$\lim_{n \rightarrow \infty} E(T(X_1, \dots, X_n)) = \theta$$

BIAS

Wenn ein Schätzer **systematisch daneben** liegt, spricht man von einer Verzerrung/ Bias.

$$Bias(T) = E(T) - \theta$$



EFFIZIENZ

Um zwischen zwei Schätzern T_1 & T_2 zu entscheiden, kann der MSE herangezogen werden. Wenn $MSE_{\theta}(T_1) \leq MSE_{\theta}(T_2)$ ist T_1 **MSE-effizienter** und somit der bessere Schätzer. Die Darstellung zeigt dieselbe Standardabweichung mit unterschiedlich grossen Stichproben.

- kleiner $MSE_{\theta}(T) \rightarrow$ effizient

MSE

Mittlere quadratische Abweichung (mean squared error)

$$MSE = E((T - \theta)^2) = Var(T) + Bias(T)^2$$

```
mse <- function(dat, wahr) {mean((dat - wahr)^2)}
```

- dat: verschiedene Realisationen des Schätzers

KONSISTENZ

Konsistenz bedeutet, dass ein Schätzer mit wachsendem Stichprobenumfang n immer näher beim wahren Wert von θ liegt.

$$\lim_{n \rightarrow \infty} MSE_{\theta}(T) = 0$$

ÄQUIVARIANZ

Äquivarianz bedeutet, dass Schätzer bestimmte Transformationen zulassen, ohne dass sich der Schätzer verändert. $C^{\circ} \rightarrow F^{\circ}$, sollte gleiche Schätzer ergeben. Quantile sind nicht äquivariant, alle anderen Schätzer erfüllen diese Eigenschaft.

ROBUSTHEIT

Robustheit bedeutet, wie anfällig Daten auf Ausreisser sind. Ein gängiges Mass dafür ist der **Bruchpunkt** ε_n^* . Der Bruchpunkt $[0, 1]$ bestimmt prozentual, wie viele Ausreisser das Datenset «verträgt» ohne Verzerrung der Schätzer.

- Arithmetisches Mittel: $\varepsilon_n^* = \frac{1}{n}$

GÜTEKRITERIEN VON PUNKTSCHÄTZERN

- **Mittelwert** einer Stichprobe
 - Erwartungstreue
 - Wenn fair, unverzerrt
 - Konsistenz
 - Genug Grosses n .
 - Effizienz
- **Varianz** einer Stichprobe
 - Erwartungstreue
 - $\frac{1}{n-1}$
 - Konsistenz
 - Genug grosses n .
 - Effizienz
 - Wenn normalverteilt

KONSTRUKTION VON SCHÄTZERN

Beim Konstruieren von Schätzern geht es darum, Methoden zu entwickeln, mit denen man unbekannte Parameter einer zugrunde liegenden Wahrscheinlichkeitsverteilung schätzt, basierend auf beobachteten Daten einer Stichprobe.

- **Ziel:** Methode entwickeln, die unbekannte Parameter einer Population mithilfe von Stichprobe schätzt.

EINSETZPRINZIP

Mit Punktschätzern einer Stichprobe, kann durch Einsetzen (z.B. Erwartungswert für Exponentialverteilung) der gesuchte Verteilungsparameter ermittelt werden.

$$E(x) = \frac{1}{\lambda} \Rightarrow \hat{\lambda} = \frac{1}{E(x)}$$

Verteilung	Stichprobe
Erwartungswert $E(X)$	arithmetisches Mittel \bar{X}_n
p -tes Moment $E(X^p)$	empirisches p -tes Moment $\frac{1}{n} \sum_{i=1}^n X_i^p$
Varianz $Var(X)$	Stichprobenvarianz S_n^2
Standardabweichung	Stichprobenstandardabweichung S_n
Kovarianz $Cov(X, Y)$	empirische Kovarianz

- Vorteile
 - Einfacher Schätzer
- Nachteile
 - Nicht robust gegen Ausreisser
 - Verteilungsannahmen

MAXIMUM-LIKELIHOOD METHODE - ML

Die Maximum-Likelihood-Methode (ML-Methode) wird verwendet, um Parameter einer **Wahrscheinlichkeitsverteilung** zu schätzen.

- **Grundidee**
 - Findet unter allen möglichen Parameterwerten, den mit der höchsten Wahrscheinlichkeit die beobachteten Daten reproduziert.
 - $L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta)$
- Vorteile
 - Annähernd erwartungstreu
 - Konsistenz & Effizienz
- Nachteile
 - kleine Stichproben stark verzerrt
 - Nicht robust gegen Ausreisser

ANALYTISCHE ML-SCHÄTZUNG

$$L(\theta) = \prod_{i=1}^n f(x_i | \theta) = f(x_1, \dots, x_n | \theta)$$

Verteilung	E(X) und Var(X)	Parameter-Schätzer
$X \sim \mathcal{N}(\mu, \sigma^2)$	$E(X) = \mu$ $Var(X) = \sigma^2$	$\hat{\mu}_{ML} = \bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$ $\hat{\sigma}_{ML}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2$
$X \sim \text{Pois}(\lambda)$	$E(X) = \lambda$ $Var(X) = \lambda$	$\hat{\lambda}_{ML} = \bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$
$X \sim \text{Exp}(\lambda)$	$E(X) = \frac{1}{\lambda}$ $Var(X) = \frac{1}{\lambda^2}$	$\hat{\lambda}_{ML} = \frac{1}{\bar{x}_n} = \frac{n}{\sum_{i=1}^n x_i}$
$X \sim \text{Bin}(m, p)$	$E(X) = m \cdot p$ $Var(X) = m \cdot p \cdot (1 - p)$	$\hat{p}_{ML} = \frac{\bar{x}_n}{m} = \frac{1}{m \cdot n} \sum_{i=1}^n x_i$

Beispiel mit Dichtefunktion $\lambda e^{-\lambda x}$ der Exponentialverteilung für einer Stichprobe: $x_i = (x_1 = 1, x_2 = 2, x_3 = 3) \rightarrow n = 3$

1. **Aufstellen der Likelihood-Funktion** $L(\theta)$, die in Abhängigkeit des unbekanntes Parameters θ die Plausibilität der beobachteten Stichprobe misst.

$$L(\lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i}$$

2. **Logarithmieren** der Likelihood-Funktion $L(\theta)$

$$\log(L(\lambda)) = \sum_{i=1}^n \log(\lambda) + \log(e^{-\lambda x_i})$$

3. **partiell nach θ ableiten** $\frac{\partial \ell(\theta)}{\partial \theta}$

$$\frac{\partial \ell(\lambda)}{\partial \lambda} \log(L(\lambda)) = \sum_{i=1}^n \frac{1}{\lambda} - x_i$$

4. **Nullsetzen** $\frac{\partial \ell(\theta)}{\partial \theta} = 0$

$$\sum_{i=1}^n \frac{1}{\lambda} - x_i = \frac{n}{\lambda} + \sum_{i=1}^n -x_i = 0$$

5. **Auflösen** der Gleichung nach θ ergibt $\hat{\theta}$ (einsetzen der x_i Werte gibt Wert)

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i} = \frac{3}{1 + 2 + 3} = \frac{1}{2}$$

6. **Überprüfen** anhand des Vorzeichens der 2. Ableitung **ob Maximum**. $\frac{\partial^2 \ell(\theta)}{(\partial \theta)^2} < 0 \rightarrow \hat{\theta}$

$$\frac{\partial^2 \ell(\lambda)}{(\partial \lambda)^2} = \frac{\partial}{\partial \lambda} \left(\frac{n}{\lambda} + \sum_{i=1}^n -x_i \right) = -\frac{n}{\lambda^2}$$

Da $\lambda \geq 0$ und $n \in \mathbb{N}^+$ ist $-\frac{n}{\lambda^2} < 0 \rightarrow$ **Maximum**

ITERATIVE VERFAHREN FÜR ML-SCHÄTZER

1. Startschätzung. θ_0
2. $i = 1$: Ausgehend von θ_{i-1} wird eine neue Parameterkombination θ_i gefunden, für welche die log-Likelihood grösser ist $\ell(\theta_i) \geq \ell(\theta_{i-1})$.
3. Iterieren bis Konvergenz.

Achtung es kann kein globales Maximum garantiert werden.

WAHL DER STARTWERTE

- Grid search
 - Probieren durch eine grobe Gittersuche nach dem besten Startwert.
- Einsetzprinzip oder Momentenmethode
- Zufallsprinzip
 - Mehrmaliges Versuchen von zufälligen Werten.

NEWTON-RAPHSON METHODE

Diese Methode bestimmt numerisch die Nullstellen einer Funktion. Das Maximum der log-Likelihood Funktion entspricht der Nullstelle der 1. Ableitung.

$$\theta_{i+1} = \theta_i - \frac{\ell'(\theta_i)}{\ell''(\theta_i)}$$

Voraussetzung $\ell^{(\theta_i)}$ ist eine glatte Funktion (beliebig oft differenzierbar)

OPTIMIERUNGSMETHODEN IN R

`optim()` funktioniert für Funktionen mit mehreren Variablen und bietet verschiedene Optimierungsmethoden an:

- BFGS (Newton-Raphson Methode)
- Nelder-Mead (Simplex)
- SANN (stochastisch)

BEISPIEL MIT NORMALVERTEILUNG

- `optim()` minimiert \rightarrow negative Log-Like!
- *Log kann nur positive Zahlen verarbeiten*
 - Wertebereich von Verteilung
 - Grenzen setzen: "L-BFGS-B"
 - `upper=...`, `lower=...`

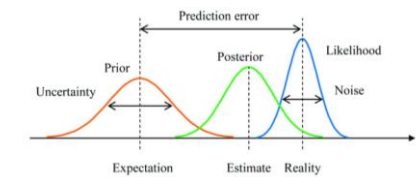
1. # Erzeugen von Beispieldaten
`data <- rnorm(500, 0, 1)`
2. # Startpunkt durch Einsetzprinzip
`m = mean(data)`
`s = sqrt(var(data)*3/pi^2)`
3. # Erzeugen von negativer Log-Likelihood
`log_like <- function(params, dat){`
> 1 Parameter \rightarrow split
 `mu <- params[1]`
 `sigma <- params[2]`
 `return(-sum(log(dnorm(dat, mean = mu,`
 `sd = sigma))))}`
4. # Optimierung durch Newtonverfahren
`optim(par = c(mu, sigma), fn =`
`log_like, method = "L-BFGS-B", dat =`
`data, lower = 0, upper = Inf)`

AUSGABE

- **\$par**: Schätzer der Parameter
- **\$value**: der minimierte Wert der Log-Likelihood
- **\$counts**: wie oft Funktion aufgerufen, Gradient berechnet
- **#convergence**: 0 = ordnungsgemäss konvergiert

BAYES-SCHÄTZER

Bayes Schätzer sind eine Methode der statistischen Inferenz, der das Prinzip der Bayes'schen Wahrscheinlichkeitsrechnung verwendet wird, um Schätzungen über unbekannte Parameter zu erhalten. Dabei wird bereits bekanntes Wissen für die Schätzung berücksichtigt.



1. Man wählt ein plausibles Modell für die zu erhebenden Daten aus
2. Man definierte die **A-Priori-Verteilung** $g(\theta)$ für die unbekannt Parameter (vorhandenes Vorwissen sollte mit einfließen)
3. Man erhebt die Stichprobe

4. Man berechnet aus der Stichprobe und der A-Priori-Verteilung die **A-Posteriori-Verteilung h** .

$$\begin{aligned} h(\theta|x_1, \dots, x_n) &= \frac{\overbrace{f(x_1, \dots, x_n|\theta)}^{\text{Likelihood}} * \overbrace{g(\theta)}^{\text{Priori}}}{\int f(x_1, \dots, x_n|\theta) * g(\theta) d\theta} \\ &= \frac{\prod_1^n f(x_i|\theta) * g(\theta)}{\int \prod_1^n f(x_i|\theta) * g(\theta) d\theta} \end{aligned}$$

```
norm <- integrate(f = function(t){1 * exp(-4.47 * t)}, lower = 0, upper = 1)
(nenner <- norm$value)
```

KONFIDENZINTERVALLE - KI

Das Konfidenzintervall $[G_u, G_o]$, zeigt in welchem Intervall ein der wahre Parameterwert mit einer gegebenen Wahrscheinlichkeit (z.B. 0.95) liegt.

- Konfidenzintervall: $P(G_u \leq \theta \leq G_o) \geq 1 - \alpha$
- Konfidenzniveau: $1 - \alpha$

Führt man ein Experiment mit einem 95% KI sehr oft durch, wird der Schätzer in 95% der Fälle in diesem Bereich liegen.

KONFIDENZINTERVALL NORMALVERTEILUNG

Bei einer **bekanntem Varianz** einer Normalverteilung kann das KI wie folgt angenommen werden:

$$\left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}, \bar{X}_n + \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} \right]$$

- \bar{X}_n : Stichprobenmittelwert
- σ : bekannte Standardabweichung
- n : Stichprobenumfang
- $q_{1-\frac{\alpha}{2}}$: Quantilwert der Normalverteilung
 - z-transformiert (da $\frac{\sigma}{\sqrt{n}}$)
 - für 0.95 $q_{norm}(0.975)$
 - für 0.99 $q_{norm}(0.995)$
- In R
 - `mean(data)+c(-1,1)*qnorm(0.975)*sd(dat)/sqrt(n)`
 - `z.test(dat, mu = mean(dat), sigma.x = sd(dat))conf.int`
 - `library(BSDA)`

KONFIDENZINTERVALL T-VERTEILUNG

Bei **unbekanntem Varianz** (geschätzt), wird die t-Verteilung angenommen (t-Verteilung ist immer standardisiert). Das KI sieht wie folgt aus:

- grosses $n \rightarrow$ annähernd Normalverteilt

$$\left[\bar{X}_n - \frac{S_n}{\sqrt{n}} qt_{(1-\frac{\alpha}{2},(n-1))}, \bar{X}_n + \frac{S_n}{\sqrt{n}} qt_{(1-\frac{\alpha}{2},(n-1))} \right]$$

- \bar{X}_n : Stichprobenmittelwert
- S_n : geschätzte Standardabweichung
 - Stichproben Standardabweichung
- n : Stichprobenumfang
- $qt_{(1-\frac{\alpha}{2},n-1)}$: Quantilwert der t-Verteilung
 - für 0.95 `qt(0.975)`
 - für 0.99 `qt(0.995)`
- In R
 - `mean(dat)+c(-1,1)*qt(0.975,df=n-1)*sd(dat)/sqrt(n)`
 - `t.test(dat, mu = mean(dat), sigma.x = sd(dat))conf.int`

KONFIDENZINTERVALLE BELIEBIGE VERTEILUNG

Ist **n gross genug**, kann für jede beliebige Verteilung die Normalverteilung angenommen werden. (Zentraler Grenzwert Satz)

- \rightarrow t-Verteilung

$$\left[\bar{X}_n - \frac{S_n}{\sqrt{n}} qt_{(1-\frac{\alpha}{2},(n-1))}, \bar{X}_n + \frac{S_n}{\sqrt{n}} qt_{(1-\frac{\alpha}{2},(n-1))} \right]$$

ASYMPTOTISCHE KONFIDENZINTERVALLE

Konfidenzintervalle können neben dem Erwartungswert auch **für alle anderen Parameter θ** bestimmt werden, falls er asymptotisch normalverteilt und konsistent ist (die meisten ML-Schätzer). Bzw. bei geschätztem $Var(\hat{\theta})$ (rechts)

Konfidenzintervall eines geschätzten Parameters $\hat{\theta}$, mit ebenfalls geschätzter Varianz. Das Konfidenzintervall ist wie folgt definiert:

$$\left[\hat{\theta} - q_{1-\frac{\alpha}{2}} \sqrt{var(\hat{\theta})}, \hat{\theta} + q_{1-\frac{\alpha}{2}} \sqrt{var(\hat{\theta})} \right]$$

- $\hat{\theta}$: Geschätzter Parameter
- $q_{1-\frac{\alpha}{2}}$: Quantil der Normalverteilung
 - für 0.95 `qnorm(0.975)`
- $\sqrt{var(\hat{\theta})}$: Geschätzte Varianz des Schätzers

KONFIDENZINTERVALL FÜR PROPORTIONEN

Das Konfidenzintervall für Verteilungen von proportionalen Schätzern (Binomialverteilung), wird typischerweise schlecht approximiert. Aus diesem Grund sollte das Clopper-Pearson-Intervall verwendet werden.

$$\left[\hat{p} - q_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + q_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

- `binom.test(x= successes, n=trials)$conf.int`

BOOTSTRAP KONFIDENZINTERVALLE

Bootstrap-Verfahren basiert auf dem Wiederholen der Stichprobendaten, um eine **Verteilung der Schätzwerte zu erzeugen**. Anstatt auf Verteilungsannahmen zurückzugreifen, werden die vorhandenen Daten genutzt, um die Unsicherheit der Schätzung zu modellieren. Aus der ursprünglichen Stichprobe werden neue Stichproben gezogen (mit zurücklegen). Dieser Prozess wird so oft wiederholt, bis eine Verteilung der Schätzung vorliegt, obwohl nur eine Stichprobe erhoben wurde.

- **Stichprobe sollte repräsentativ sein!**

```
library(boot)
set.seed(2017)
schaetzer <- function(data, ind) 1/mean(data[ind])
boot.erg <- boot(data = x, statistic = schaetzer, R = 500)
```

```
boot.ci(boot.erg, conf = 0.95, type = c("perc", "bca"))
```

BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 500 bootstrap replicates

LIBRARY(BOOT)

- library(boot)
- schaezer <- function (data, ind) mean(data[ind])
- bs <- boot(data, schaezer, R = Anzahl Boot-Straps)
 - braucht eine Schätzerfunktion mit den Parametern «data» und «ind».
 - Funktion muss abhängig sein von den Daten und dem Index.
 - mean(data[ind])
 - median(data[ind]) ...
- boot.ci(boot.out= bs, conf = 0.95, type = «bca»)
- bs\$t
 - Liefert alle Schätzer pro Boot
 - hist(bs\$t)

MANUELL

```
bs <- replicate(200, sample(data, replace = T))
```

```
bs_mean <- apply(bs, MARGIN = 2, FUN = mean)
```

```
hist(bs_mean)
```

```
quantile(bs_mean, c(0.025, 0.975))
```

ZUSAMMENFASSUNG FORMELN

Modell/Situation	(1-α)-Konfidenzintervall
μ einer Normalverteilung σ ² bekannt	$\left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} q_{(1-\frac{\alpha}{2})}, \bar{X}_n + \frac{\sigma}{\sqrt{n}} q_{(1-\frac{\alpha}{2})} \right]$ library(BSDA) z.test(x, sigma.x=σ, conf.level=1-α)\$conf.int
μ einer Normalverteilung σ ² unbekannt	$\left[\bar{X}_n - \frac{s_n}{\sqrt{n}} q_{(1-\frac{\alpha}{2}, (n-1))}, \bar{X}_n + \frac{s_n}{\sqrt{n}} q_{(1-\frac{\alpha}{2}, (n-1))} \right]$ t.test(x, conf.level=1-α)\$conf.int
E[X] für beliebige Verteilung (bei grossem n)	$\left[\bar{X}_n - \frac{s_n}{\sqrt{n}} q_{(1-\frac{\alpha}{2}, (n-1))}, \bar{X}_n + \frac{s_n}{\sqrt{n}} q_{(1-\frac{\alpha}{2}, (n-1))} \right]$ t.test(x, conf.level=1-α)\$conf.int
p einer Binomialverteilung	binom.test(x, n, conf.level=1-α)\$conf.int
λ einer Poisson-Verteilung	poisson.test(x, T, conf.level=1-α)\$conf.int
θ geschätzt mit asympt. normalverteiltem Schätzer	$\left[\hat{\theta} - q_{(1-\frac{\alpha}{2})} \sqrt{\widehat{Var}(\hat{\theta})}, \hat{\theta} + q_{(1-\frac{\alpha}{2})} \sqrt{\widehat{Var}(\hat{\theta})} \right]$
θ geschätzt keine Annahmen	library(boot); schaezter <- function(data, ind) ... bs <- boot(data=x, statistic=schaetzer, R=500) boot.ci(bs, conf = 0.95, type = c("perc", "bca"))

- KI um Faktor $\frac{1}{k}$ verkleinern: $n * k^2$
 - $\frac{1}{2} * KI = n * 4$
- KI um Faktor k vergrössern: $n * 1/k^2$
 - $2 * KI: n * \frac{1}{4}$

KONFIDENZNIVEAU BESTIMMEN

$$[\text{untere Grenze}, \text{obere Grenze}] = \bar{x} \pm Z * \frac{\sigma}{\sqrt{n}}$$

$$Z = \frac{\bar{x} - \text{untere Grenze}}{\frac{\sigma}{\sqrt{n}}} = \frac{\text{obere Grenze} - \bar{x}}{\frac{\sigma}{\sqrt{n}}}$$

$$\alpha = 2 * (1 - pnorm(Z))$$

$$\text{Konf. Niveau} = 1 - \alpha$$

N BESTIMMEN

$$Z = q\left(1 - \frac{\alpha}{2}\right)$$

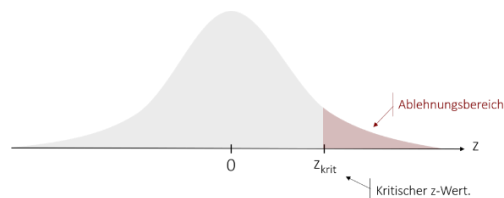
STATISTISCHES TESTEN

Statistisches Testen ist ein zentraler Bestandteil der Statistik. Es wird verwendet, um anhand von Stichprobendaten zu entscheiden, ob eine bestimmte Annahme (Hypothese) über eine Population zutreffend ist oder nicht. Die Grundidee ist dabei, die Wahrscheinlichkeit zu bewerten, mit der beobachtete (oder extremere) Daten auftreten würden, wenn die angenommene Hypothese wahr wäre.

Es wird jeweils eine Nullhypothese H_0 , sowie eine Alternativhypothese H_1 definiert. Die Nullhypothese ist in der Regel **eine Aussage über keine Veränderung**, keinen Effekt oder keinen Unterschied.

- Beobachtungen müssen aus **iid-Stichprobe** sein
- Beobachtungen haben bestimmte Verteilung

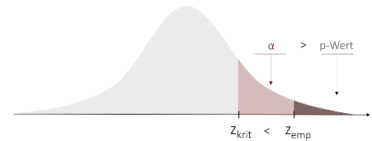
ANNAHMEBEREICH



- Annahmebereich ist der Bereich, indem die **Nullhypothese** «angenommen» wird
- Ablehnungsbereich: alles ausserhalb des Annahmebereichs.

P-WERT

Definition: Der p-Wert ist die Wahrscheinlichkeit, unter der Annahme, dass die Nullhypothese wahr ist, eine Teststatistik zu beobachten, die gleich oder extremer ist als der tatsächlich beobachtete Wert.



- Ein kleiner p-Wert (kleiner als ein Signifikanzniveau, z.B. 0.05)
 - Daten ungewöhnlich (aus anderer Population)
 - Nullhypothese verwerfen
- $pt(q=\text{teststatistik}, df = n-1, \text{lower.tail}=T/F)$
 - F für H_A : grösser
- Bei zweiseitig: $pt(t, df) * 2$

FEHLER BEI STATISTISCHEN TESTS

	H_0 beibehalten	H_0 verworfen	
H_0 ist wahr	richtige Entscheidung	Fehler 1. Art	p-Wert < α durch puren Zufall
H_0 ist falsch	Fehler 2. Art	richtige Entscheidung	
	p-Wert nicht klein genug obwohl H_0 wahr		

Fehler 1. Art (α -Fehler): Nullhypothese wird verworfen, obwohl sie korrekt ist: (**schlimmer**)

Fehler 2. Art (β -Fehler): Nullhypothese wird behalten, obwohl sie falsch ist.

DURCHFÜHRUNG TEST

1. Formulierung der Nullhypothese H_0
 - a. Meist keine Veränderung
2. Formulierung der Alternativhypothese H_A
 - a. $h < h_0$: Abw. nach oben, **einseitig**
 - b. $h > h_0$: Abw. nach unten, **einseitig**
 - c. $h \neq h_0$: Abw. nach unten und oben, **beidseitig**
3. Signifikanzniveau
 - a. $\alpha = 0.05$, oft auch 0.01
4. Teststatistik
 - a. T : hypothetische Annahme
 - b. $T = \sqrt{n} \frac{\bar{X} - \mu_0}{s_n}$ (für t-Test)
 - i. \bar{X} : Wert aus Stichprobe
 - ii. μ_0 : Nullhypothese
 - iii. s_n : Stichproben Sd
 - iv. n : Anzahl Daten
 - c. t : realisierte Grösse
5. P-Wert berechnen / Verwerfungsbereich definieren
6. Testentscheid
 - a. **Nullhypothese nicht verwerfen:**
p-Wert > α (Beobachtung liegt im Annahmebereich)
 - b. **Nullhypothese verwerfen:**
p-Wert < α (Messwert nicht im Annahmebereich)
 - c. Beibehaltung der H_0 bedeutet nicht, dass Nullhypothese bewiesen ist, sie bleibt lediglich plausibel

T-TEST

Für **unbekanntes σ** . Die t-Verteilung ist immer standardisiert.

- **Voraussetzung:** t-Test muss entweder
 - normalverteilt sein
 - oder $n > 25$ (ZGws)
- Nullhypothese: $H_0: \mu = \mu_0$
- Alternativhypothese: $H_A: \mu \neq, <, > \mu_0$
- Teststatistik: $t = \sqrt{n} \frac{\bar{x} - \mu_0}{s_n}$
- p-Wert: `pnorm(t, df=n-1, lower.tail)`

```
t.test(benzin, mu = 8.2, alternative = "greater")
```

Stichprobe \rightarrow H_0 \rightarrow H_A

One Sample t-test

```
data: benzin
t = 8.1649, df = 35, p-value = 6.413e-10
alternative hypothesis: true mean is greater than 8.2
95 percent confidence interval:
 8.56996      Inf
sample estimates:
mean of x
8.666492
```

p-Wert (falls α wird H_0 verworfen)

Konfidenzintervall

Schätzer

Z-TEST

Für einen z-Test muss die **Standardabweichung σ bekannt** sein.

- **Voraussetzung:** z-Test muss entweder
 - normalverteilt sein
 - oder $n > 25$ (ZGws)
- Nullhypothese: $H_0: \mu = \mu_0$
- Alternativhypothese: $H_A: \mu \neq, <, > \mu_0$
- Teststatistik: $t = \sqrt{n} \frac{\bar{x} - \mu_0}{\sigma_n}$
- p-Wert: `pnorm(t)`

```
library(BSDA)
s <- rep(193.1,30)
z.test(x = s, sigma.x = 20, mu = 200, alternative = "two.sided")
```

Package \rightarrow Stichprobe muss existieren, kann künstlich generiert werden $\rightarrow H_A$

One-sample z-Test \rightarrow bekanntes σ $\rightarrow H_0$

```
data: s
z = -1.8896, p-value = 0.05881
alternative hypothesis: true mean is not equal to 200
95 percent confidence interval:
185.9432 200.2568
sample estimates:
mean of x
193.1
```

p-Wert (falls α wird H_0 verworfen)

Konfidenzintervall

Schätzer

BINOM-TEST

Man möchte beweisen, dass die Erfolgsrate (Anzahl Bernoulli-Versuchen mit Erfolg) Wahrscheinlichkeit kleiner, grösser oder ungleich dem vorgegebenen p_0 ist.

- Nullhypothese $H_0: p = p_0$
- Alternativhypothese $H_A: p \neq, <, > p_0$
- Teststatistik: T
 - $T =$ Anzahl beobachtete Erfolge
- p-Wert: `pbinom(teststatistik)`

```
# Beobachtungen # Versuche
binom.test(x = 26, n = 112, p = 0.32, alternative = "less")
```

Exact binomial test $\rightarrow H_0$ $\rightarrow H_A$

```
data: 26 and 112
number of successes = 26, number of trials = 112, p-value = 0.0268
alternative hypothesis: true probability of success is less than 0.32
95 percent confidence interval:
0.0000000 0.3072063
sample estimates:
probability of success
0.2321429
```

p-Wert (falls α wird H_0 verworfen)

Konfidenzintervall

Schätzer

POISSON TEST

Der Poisson-Test überprüft, ob die Anzahl Ereignisse während einer Zeit/ an einem Ort mit einer mittleren Ereignisrate λ_0 vereinbar sind.

Man möchte beweisen, dass die Ereignisrate kleiner, grösser oder ungleich dem vorgegebenen λ_0 ist.

- Nullhypothese: $H_0: \lambda = \lambda_0$
- Alternativhypothese: $H_A: \lambda \neq, <, > \lambda_0$
- Teststatistik: $t = \lambda$
- p-Wert: `ppois(t)`

```
# Beobachtungen # Beobachtungsfenster (zeitlich/räumlich)
poisson.test(x = 4, T = 1, r = 9, alternative = "less")
```

Exact Poisson test $\rightarrow H_0$ $\rightarrow H_A$

```
data: 4 time base: 1
number of events = 4, time base = 1, p-value = 0.05496
alternative hypothesis: true event rate is less than 9
95 percent confidence interval:
0.000000 9.153519
sample estimates:
event rate
4
```

p-Wert (falls α wird H_0 verworfen)

Konfidenzintervall

Schätzer

2 STICHPROBENTEST

2 Stichprobentests dienen dazu, zwei **unabhängige** (ungepaarte) Stichproben miteinander zu vergleichen, um zu beurteilen, ob es signifikante Unterschiede zwischen ihnen gibt.

2 STICHPROBEN BINOMIALTEST

Vergleich von zwei Anteilen p_1 und p_2 .

- Nullhypothese: $H_0: p_1 = p_2$
 - $p_1 - p_2 = 0$
- Alternativhypothese: $H_A: p_1 \neq p_2$
 - $p_1 < \neq p_2$
- Voraussetzung:
 - $n_1 * p_1(1 - p_1) \geq 5$
 - $n_1 * p_2(1 - p_2) \geq 5$
- Teststatistik
 - $$T = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

In R:

```
prop.test(x=c(4,4),n=c(10,20),alternative="two.sided")
```

- x=Anzahl Erfolge beider Stichproben
- n=Anzahl Versuche beider Stichproben

```
data: c(4, 4) out of c(10, 20)
X-squared = 0.53267, df = 1, p-value = 0.4655
alternative hypothesis: two.sided
95 percent confidence interval:
-0.225609 0.625609
sample estimates:
prop 1 prop 2
0.4 0.2
```

p-Wert (falls α wird H_0 verworfen)

Konfidenzintervall für $p_1 - p_2$

Schätzung für p_1 und p_2

2 STICHPROBEN POISSON TEST

Vergleich von zwei Anzahlen λ_1 und λ_2 .

- Nullhypothese: $H_0: \lambda_1 = \lambda_2$
 - $\lambda_1 - \lambda_2 = 0$
- Alternativhypothese: $H_A: \lambda_1 \neq, <, > \lambda_2$

```
poisson.test(c(16,31), T=c(1,1), r=1, alternative = "two.sided")
```

- x = Beobachtete Anzahl
- T = Vektor der Zeitintervalle
 - r = Verhältnis
 - nur nötig, wenn kein Verhältnis in T

```
Comparison of Poisson rates
data: c(16, 31) time base: c(1, 1)
count1 = 16, expected count1 = 23.5, p-value = 0.03999
alternative hypothesis: true rate ratio is not equal to 1
95 percent confidence interval:
0.2636488 0.9729021
sample estimates:
rate ratio
0.516129
```

p-Wert (falls α wird H_0 verworfen)

Konfidenzintervall für λ_1/λ_2

Schätzer für λ_1/λ_2

2 STICHPROBEN T-TEST

Vergleich von zwei **iid** verteilten Stichprobenmittelwerten μ_1 und μ_2 .

- Nullhypothese: $H_0: \mu_1 = \mu_2, \mu_1 - \mu_2 = 0$
- Alternativhypothese: $H_A: \mu_1 \neq, <, > \mu_2$
- Teststatistik: $T = \frac{\bar{X} - \bar{Y}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$
 - mit $s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$

- Voraussetzung:
 - **Müssen identische Varianzen σ^2 haben.** (dürfen unterschiedliche Erwartungswerte haben)
 - **Müssen approx. normalverteilt sein.**
 - nicht normalverteilten Stichproben mit grossen Werten für n_1 und n_2 ist der Test approximativ korrekt (Zentraler Grenzwertsatz)

```
Stichproben
H0 (d.h. mu1-mu2)
HA
t.test(x=self, y=manuell, mu=0, alternative = "two.sided",
var.equal = TRUE, paired = FALSE)
Gleiche Varianz
ungepaarter Test
Two Sample t-test
data: self and manuell
t = -1.0394, df = 16, p-value = 0.3141
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-428.0323 146.3823
sample estimates:
mean of x mean of y
2902.800 3043.625
```

p-Wert (falls α wird H_0 verworfen)

Konfidenzintervall für $\mu_1 - \mu_2$

Schätzer für μ_1 und μ_2

GEPAAARTE STICHPROBEN

	Gepaarte Stichproben	Ungepaarte Stichproben
Messung	Messgrösse wird für jede Beobachtungseinheit 2x unter verschiedenen Bedingungen gemessen.	Stichproben werden aus zwei unabhängigen Grundgesamtheiten gezogen.
Paarbildung	Jedem Wert der ersten Stichprobe kann genau ein Wert der zweiten Stichprobe zugeordnet werden.	keine Paarbildung möglich.
Stichprobengrösse	gleich	können verschieden sein

GEPAAARTER TEST

Vorgehen wie bis anhin, nur wird Differenz der beiden Tests verwendet.

$$\text{Differenz} = D_1, \dots, D_n$$

- Differenzen können annähernd als normalverteilt angenommen werden.
- Mit Differenz 1-Stichproben t-Test durchführen.
- Nullhypothese: $H_0: \mu_1 = \mu_2, \mu_1 - \mu_2 = 0$
- Alternativhypothese: $H_A: \mu \neq, <, > \mu_0$

Teststatistik:

$$T = \sqrt{n} \frac{(\bar{D} - \mu_0)}{\sigma_D}$$

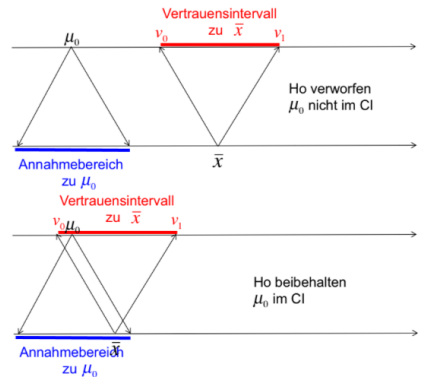
In R:

```
t.test(bs$Linux - bs$Windows, mu = 0, alternative = "two.sided")
**oder** beide Stichproben an t.test() übergeben:
t.test(x=bs$Linux, y=bs$Windows, paired = TRUE, mu = 0, alternative = "two.sided")
```

DUALITÄT KI UND TEST

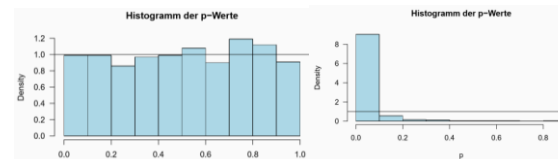
Welche Beobachtungen sind vereinbar mit H_0 bzw. mit einem bestimmten Sollwert für μ_0 ?

Wenn das Konfidenzintervall (Vertrauensintervall) des Mittelwerts der Stichprobe mit dem Annahmebereich von H_0 (μ_0) überlappt, wird die Nullhypothese nicht verworfen.



VERTEILUNG DES P-WERTS

p-Verteilung H_0 gilt, p-Verteilung H_0 gilt nicht:



Die Werte konzentrieren sich nahe bei 0, was bedeutet, dass die p-Werte zum Grossteil kleiner als α sind.

GÜTEFUNKTION

Die Fehler 1. Art (α) und 2. Art (β) lassen sich mittels der Gütefunktion beschreiben, die wie folgt definiert ist:

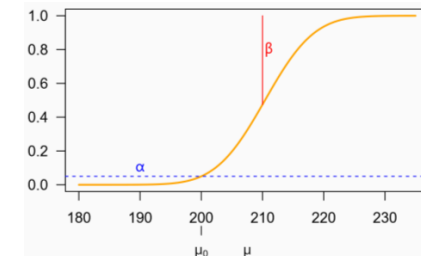
$$g(\mu) = P(H_0 \text{ ablehnen} | \mu)$$

$g(\mu)$ gibt für μ an, mit welcher Wahrscheinlichkeit H_0 verworfen wird.

Bei einem Signifikanzniveau von $\alpha = 0.05$ gilt:

$$g(\mu_0) = 0.05$$

$$\mu \leq \mu_0 \Leftrightarrow g(\mu) \leq \alpha$$



Bei einem $\mu = 210$ verwerfen wir H_0 mit einer Wahrscheinlichkeit von 0.475. D.h. wir machen einen Fehler 2. Art mit einer Wahrscheinlichkeit von:

$$\beta = 1 - 0.475 = 0.525$$

Je grösser der Abstand (Effektgrösse) zwischen μ_0 und μ ist, umso geringer ist die Wahrscheinlichkeit einen Fehler zweiter Art zu machen.

MACHT

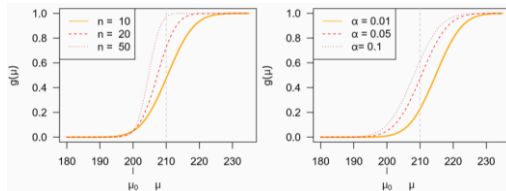
Die Macht (Power) bezeichnet die Wahrscheinlichkeit, eine Nullhypothese zu verwerfen, wenn sie wirklich falsch ist. Das heisst die Wahrscheinlichkeit einen Fehler 2. Art zu vermeiden.

$$Power = 1 - \beta$$

- β : **Fehler 2. Art** (oder β -Fehler): Nullhypothese wird behalten, obwohl sie falsch ist.

Die Macht ist abhängig von:

- Signifikanzniveau α
 - Grösseres $\alpha \rightarrow$ mehr Macht
 - höheres Risiko auf α -Fehler
- Stichprobengrösse n
 - grosses $n \rightarrow$ mehr Macht
- Effektgrösse (Abstand zur Nullhypothese)
 - grösserer Abstand \rightarrow mehr Macht



In R:

```
power.t.test(n = 10, delta = NULL, sd = 0.3, sig.level = 0.05,
             power = 0.8, type = "two.sample",
             alternative = "one.sided")
```

- Immer gesuchter Wert = NULL

MULTIPLES TESTEN

```
pairwise.t.test(x = pulp$bright, g = pulp$operator,
               p.adjust.method = "none", pool.sd = F)
```

- x: response vector
- g: grouping vector or factor
- pool.sd: same variance

FAMILY-WISE ERROR RATE

Mit steigender Anzahl Tests n steigt die Wahrscheinlichkeit einen Fehler 1. Art (H_0 verworfen, obwohl wahr) zu erhalten.

$$FWER = 1 - (1 - \alpha)^n$$

Bsp. die Wahrscheinlichkeit bei 10 Tests mit Signifikanzniveau 0.05 einen Fehler zweiter Art zu machen, liegt bei 40%: $1 - (1 - 0.05)^{10} = 0.4$

BONFERRONI-KORREKTUR

Um die steigende Wahrscheinlichkeit bei grossem n zu vermeiden, wird die Bonferroni-Korrektur ($\frac{\alpha}{n}$) verwendet. $FWER_{korrigiert} = 1 - \left(1 - \frac{\alpha}{n}\right)^n$

Option 1: `p.adjust(p, method = "bonferroni")`

Option 2: `pairwise.t.test(x, g, p.adjust.method = "bonf")`

- `pairwise.t.test()`: führt multiple Tests durch, mit dem Parameter: `method` kann eine Korrektur gewählt werden

- `p.adjust(method="«bonf»)`: bezieht sich auf bereits unkorrigiert berechnete p-Werte. (sinnvoll, wenn pairwise.... nicht verfügbar)

TESTENTSCHEIDE UND P-WERTE

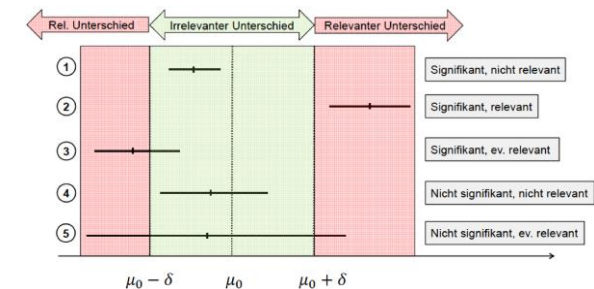
Testentscheidungen sind binär. Es gibt Situationen, die eine solch harte Schwelle erfordern (Medizin).

Wenn von einem 0.95 Signifikanzniveau ausgegangen wird, kann man allgemein davon ausgehen:

- p-Werte < 0.01 : Effekt ziemlich sicher
- $0.01 < p\text{-Werte} \leq 0.1$: erfordert weitere Untersuchungen
- P-Wert > 0.1 : ziemlich sicher kein Effekt

SIGNIFIKANZ VS. PRAKTISCHE RELEVANZ

Nicht immer ist ein signifikanter Unterschied auch ein relevanter Unterschied. In der Praxis ist davon kann es vorkommen, dass ein Effekt zwar signifikant aber irrelevant ist.



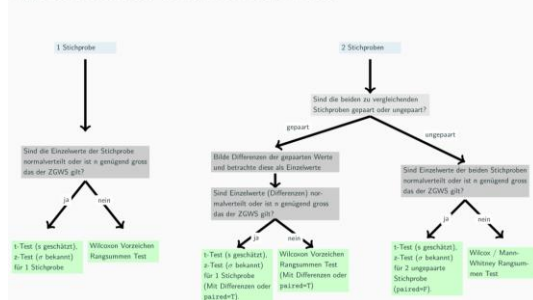
NICHT PARAMETRISCHES TESTEN

Für Verteilungen deren Verteilungen unbekannt ist, können nicht parametrische Tests verwendet werden. Der Wilcoxon-Rangsummentest ist eine nicht-parametrisierte Alternative zu herkömmlichen parametrisierten Test, wenn die Verteilungsannahmen für herkömmliche t-Tests nicht erfüllt sind.

ANWENDUNGEN

- **Kleine Stichproben** (kein ZGWS)
 - $n < 30$
- Nicht bekannte Verteilung
- Ordinal/Rangdaten
- Mangel an Homoskedastizität

Hypothesentest für den Mittelwert/Median



WILCOXON-RANGSUMMENTEST



Nicht parametrischer Test für 2 unabhängige Stichproben X und Y . Voraussetzung für diesen Test sind:

- Stetigkeit der Daten
- X und Y haben dieselbe Verteilung
- bis auf eine Verschiebung um δ identisch

```
wilcox.test(x=x, y=y, mu=0, alternative="greater", paired=FALSE)
```

```
library(coin)
```

```
# Kombinieren der Daten in einem Dataframe mit Faktor
df <- data.frame(Data = c(Sand1, Sand2),
                  Sand = factor(c(rep("Sand1", length(Sand1)), rep("Sand2", length(Sand2))))
)
```

```
# Durchführung des Wilcoxon-Tests mit der Formelschnittstelle
wilcox_test(Data ~ Sand, data = df,
             alternative = "two.sided", distribution = "exact")
```

Exact: kleine Stichproben

WILCOXON-VORZEICHEN-RANGSUMMENTEST

Für gepaarte Stichproben.

```
wilcox.test(reifen$Profil.A.sA.in.m, reifen$Profil.B.sB.in.m, paired = TRUE)
```

```
library(coin)
wilcoxsign_test(data$A - data$B,
                alternative = "two.sided", distribution = "exact")
```

Exact Wilcoxon-Pratt Signed-Rank Test

```
data: y by
      x (pos, neg)
stratified by block
Z = -2.5005, p-value = 0.007812
alternative hypothesis: true mu is not equal to 0
```

CHI QUADRATTEST

Der Test zielt darauf ab, zu prüfen, ob zwei kategoriale Merkmale unabhängig sind oder nicht. **Sagt nichts über Art des Zusammenhangs aus!**

- Nullhypothese: $H_0: O = E$
- Alternativhypothese: $H_A: O \neq E$
 - O = die beobachteten Häufigkeiten in den Zellen der Kreuztabelle
 - E = die erwarteten Häufigkeiten in den Zellen der Kreuztabelle
- Teststatistik T
 - $\chi_c^2 = \sum \left[\frac{(O-E)^2}{E} \right]$
 - $= \sum_{i=1}^r \sum_{j=1}^s \frac{\left(n_{ij} - \frac{n_{i0} n_{0j}}{n} \right)^2}{\frac{n_{i0} n_{0j}}{n}}$

Berechnung von E:

$$E = \frac{\text{Spaltensumme} * \text{Zeilensumme}}{\text{Stichprobengröße } n}$$

IN R

- `daten <- matrix(c(28, 44, 60, 58), nrow=2, byrow=T)`
- `rownames(daten) <- c("ländlich", "städtisch")`
- `colnames(daten) <- c("Margarine", "Butter")`
- `chisq.test(daten)`
- `chisq.test(var1, var2)$observed`
- `chisq.test(var1, var2)$expected`
- `chisq.test(var1, var2)$residuals`

PERMUTATIONSTEST

Wie Bootstrap-Verfahren wird mit Hilfe von simulierten Daten die empirische Verteilung der Teststatistik unter H_0 erzeugt.

1. Daten werden zufällig permutiert
 - a. Ziehen **ohne** Zurücklegen
2. Berechnung der Teststatistik T_i^p für permutierte Daten
 - a. Z.B. Mittelwert
3. K-fache Wiederholung von Schritt 1 und 2
4. Berechnung des p-Werts mit Hilfe der empirischen Verteilung
 - a. Anzahl extremere Werte der Teststatistiken als der beobachtete Wert, geteilt durch Anzahl der permutierten Teststatistiken
 - b. $\frac{\text{Anzahl von } T_i^p > T_{\text{obs}}}{k}$

```
library(coin)
independence_test(Rainfall ~ Treatment,
                 data = case0301,
                 distribution = "exact",
                 alternative = c("two.sided"))
```

Exact General Independence Test

```
data: Rainfall by Treatment (Seeded, Unseeded)
Z = 1.9421, p-value = 0.04419
alternative hypothesis: two.sided
```

IN R

```
# Beobachtete Teststatistik
obs_stat <- abs(mean(group_a) -
                mean(group_b))

# Kombinierte Daten / Nullhypothese → keinen
Unterschied
combined <- c(group_a, group_b)

# Anzahl der Permutationen
n_perm <- 10000

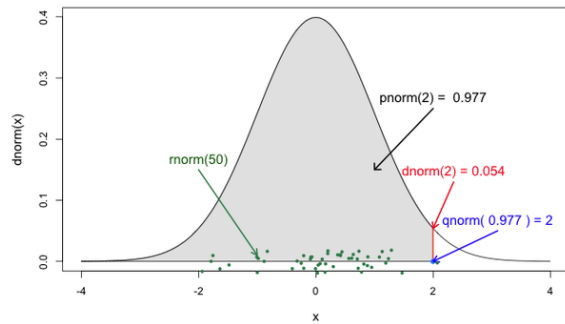
perm_stats <- numeric(n_perm)

# Permutationstest
for (i in 1:n_perm) {
  permuted <- sample(combined)

  perm_stats[i] <- abs(mean(permuted[1:10]) -
                    mean(permuted[11:20]))
}

# Berechnung des p-Werts
p_value <- mean(perm_stats >= obs_stat)
```

ANHANG



- d: Density function
- p: Probability distribution function
- q: Quantile function
- r: Random deviates function

Diskrete Verteilungen					
Verteilung	W'keitsfunktion P(X=x)	Wertebereich	R (.. = d/p/q/r)	Erwartungswert und Varianz	Anwendung
Diskrete Gleichverteilung	$P(X = x_i) = p$	$\{x_1, \dots, x_k\}$	<code>werte <- c(x1,x2, ... xk)</code> <code>sample(werte, size = 1)</code>	$E(X) = \sum_{i=1}^k x_i p$ $Var(X) = \sum_{i=1}^k (x_i - E[X])^2 p$	Alle Ereignisse sind gleich wahrscheinlich. Bsp: Würfeln
Bernoulli $X \sim \text{Bernoulli}(p)$	$P(X = 1) = p, P(X = 0) = 1 - p$	$\{0, 1\}$	<code>sample(0:1, size=1, probs = c(1-p, p))</code>	$E(X) = p$ $Var(X) = p(1-p)$	Indikator, ob ein Ereignis eintritt. Bsp: Münzwurf, Würfeln einer 6
Binomial $X \sim \text{Binom}(n, p)$	$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$	$\{0, 1, \dots, n\}$	<code>..binom(..., size=n, prob=p)</code>	$E(X) = np$ $Var(X) = np(1-p)$	Anzahl Erfolge in n unabhängigen Bernoulli-Versuchen. Bsp: Anzahl Studenten, die Prüfung bestehen
Poisson $X \sim \text{Pois}(\lambda)$	$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$	$\{0, 1, \dots, \infty\}$	<code>..pois(..., lambda = \lambda)</code>	$E(X) = \lambda$ $Var(X) = \lambda$	Anzahl Ereignisse k mit konstanter Eintreffrate λ . Bsp: Anzahl Kunden am Postschalter pro Stunde
Geometrisch $X \sim \text{Geom}(p)$	$P(X = k) = p(1-p)^k$	$\{0, 1, \dots, \infty\}$	<code>..geom(..., prob = p)</code>	$E(X) = (1-p)/p$ $Var(X) = (1-p)/p^2$	Anzahl Misserfolge k bis der erste Erfolg auftritt. Bsp: Anzahl Fehlwürfe, bis man das Haus verlassen bei Eile mit Weile kann
Negativ Binomial $X \sim \text{NBinom}(r, p)$	$P(X = k) = \binom{k+r-1}{k} p^r (1-p)^k$	$\{0, 1, \dots, \infty\}$	<code>..nbinom(..., size=r, prob=p)</code>	$E(X) = r \cdot (1-p)/p$ $Var(X) = r \cdot (1-p)/p^2$	Anzahl Misserfolge bis zum r-ten Erfolg. Bsp: Anzahl erfolgloser Angelversuche bis man 2 Fische gefangen hat
Hypergeometrisch $X \sim \text{Hyp}(m, n, k)$	$P(X = j) = \frac{\binom{m}{j} \binom{n-m}{k-j}}{\binom{n}{k}}$	$\{0, 1, \dots, \min(k, m)\}$	<code>..hyper(..., m=m, n=n, k=k)</code>	$E(X) = km/(m+n)$ $Var(X) = k \cdot \left(\frac{m}{m+n}\right) \cdot \left(1 - \frac{m}{m+n}\right) \cdot \left(\frac{n+1}{m+1}\right)$	Anzahl Erfolge j in einer Stichprobe der Größe k bei einer Grundgesamtheit von m (Erfolgselemente) und n (Nicht-Erfolgen). Bsp: Anzahl defekter Schrauben in einer Stichprobe

Stetige Verteilungen					
Verteilung	Dichte	Wertebereich	R (.. = d/p/q/r)	Erwartungswert und Varianz	Anwendung
Uniform $X \sim \text{Unif}(a, b)$	$f(x) = \begin{cases} 0 & a < 0 \\ \frac{1}{b-a} & (a \leq x \leq b) \\ 0 & b > 0 \end{cases}$	$[a, b]$	<code>..unif(..., min=a, max=b)</code>	$E(X) = (a+b)/2$ $Var(X) = (b-a)^2/12$	Wenn alle Werte in einem Bereich gleich wahrscheinlich sind. Bsp: Wartezeit auf Bus, der exakt alle 8 Minuten fährt
Exponential $X \sim \text{Exp}(\lambda)$	$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$	≥ 0	<code>..exp(..., rate = \lambda)</code> <i>λ auf Zeiteinheit bezogen wenn $E(X) = 100h \rightarrow \lambda = \frac{1}{100h}$</i>	$E(X) = \frac{1}{\lambda}$ $Var(X) = \frac{1}{\lambda^2}$	Dauer von zufälligen Zeitintervallen (zwischen 2 Poisson-Ereignissen) Bsp: Wartezeit für Kunden
Weibull $X \sim \text{Weibull}(\beta, \lambda)$	$f(x) = \lambda \beta (\lambda \cdot x)^{\beta-1} e^{-(\lambda x)^\beta}$ für $x \geq 0$	≥ 0	<code>..weibull(..., shape=\beta, scale=1/\lambda)</code> <i>rate = \lambda</i> <i>β = Anzahl Exp. verteilten Ereignisse</i>	$E(X) = \frac{1}{\lambda} \Gamma(1 + 1/\beta)$ $Var(X) = \frac{1}{\lambda^2} [\Gamma(1 + 2/\beta) - \Gamma^2(1 + 1/\beta)]$ <i>$\lambda = 1/E(X) \cdot \Gamma(1 + 1/\beta)$</i>	Dauer von zufälligen Zeitintervallen mit Berücksichtigung der Lebensphase Bsp: Lebensdauer von Bauteilen
Gamma $X \sim G(k, \lambda)$	$f(x) = e^{-\lambda x} \cdot x^{k-1} \cdot \frac{\lambda^k}{\Gamma(k)}$ für $x \geq 0$	≥ 0	<code>..gamma(..., shape=k, rate=\lambda)</code> <i>shape = Anzahl exp. verteilte Ereignisse rate = \lambda</i>	$E(X) = \frac{k}{\lambda}$ $Var(X) = \frac{k}{\lambda^2}$	Summe von exponentialverteilten Zufallsvariablen Bsp: Kosten
Normal $X \sim \mathcal{N}(\mu, \sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right)$	\mathbb{R}	<code>..norm(..., mean=\mu, sd=\sigma)</code>	$E(X) = \mu$ $Var(X) = \sigma^2$	Abweichungen von (Mess-) Werten, Näherung bei grossen Stichproben Bsp: Körpergröße
Lognormal $X \sim \text{log.}\mathcal{N}(\mu, \sigma^2)$	$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(\log(x)-\mu)^2}{\sigma^2}\right)$	> 0	<code>..lnorm(..., mean=\mu, sd=\sigma)</code>	$E(X) = e^{\mu + \frac{\sigma^2}{2}}$ $Var(X) = e^{2\mu + 2\sigma^2} (e^{\sigma^2} - 1)$	rechtsschiefe Messdaten mit nur positiven Werten Bsp: Geflogene Meilen