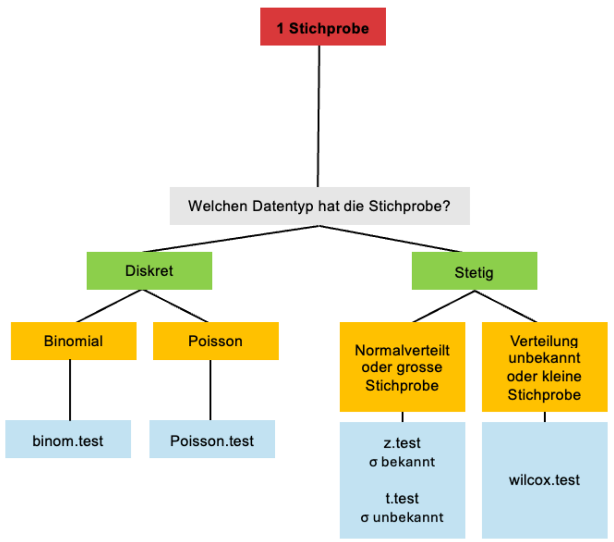
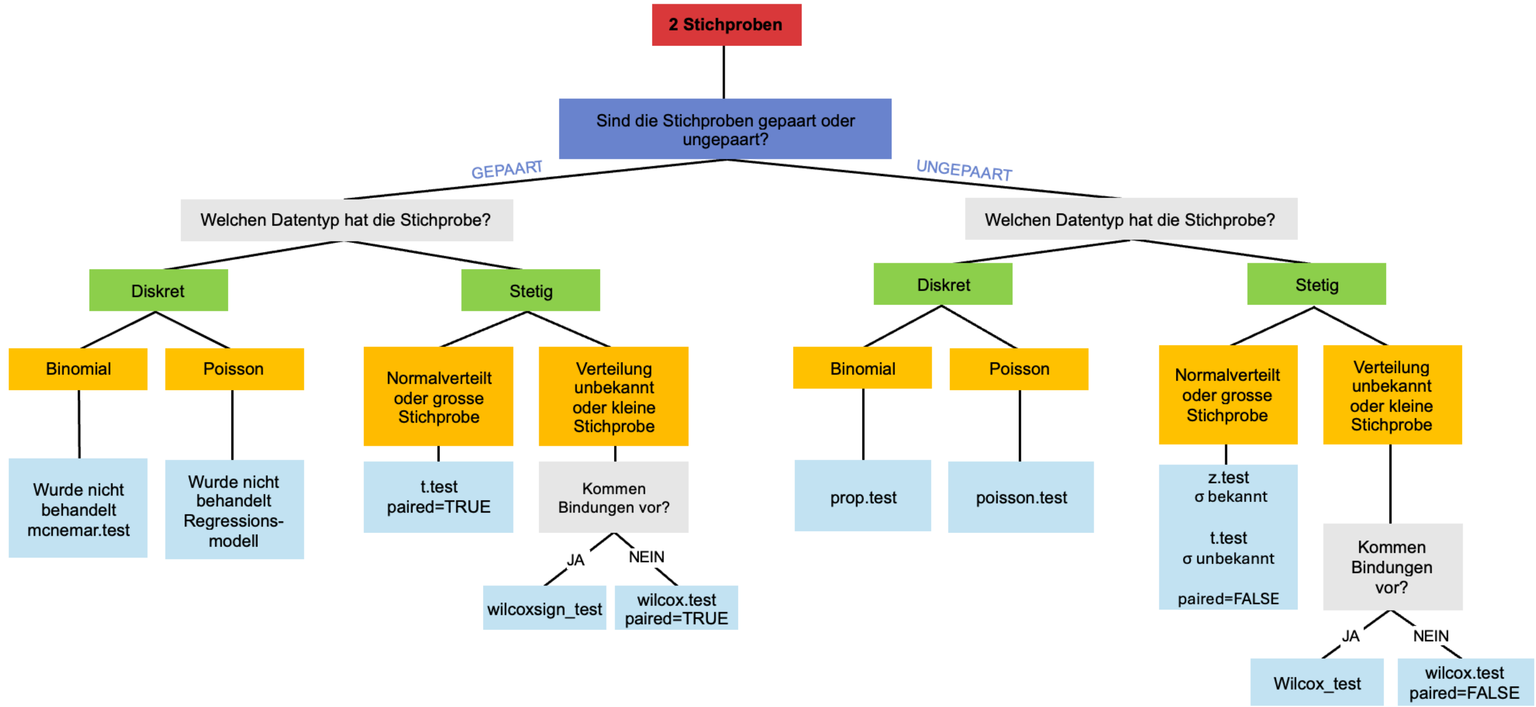
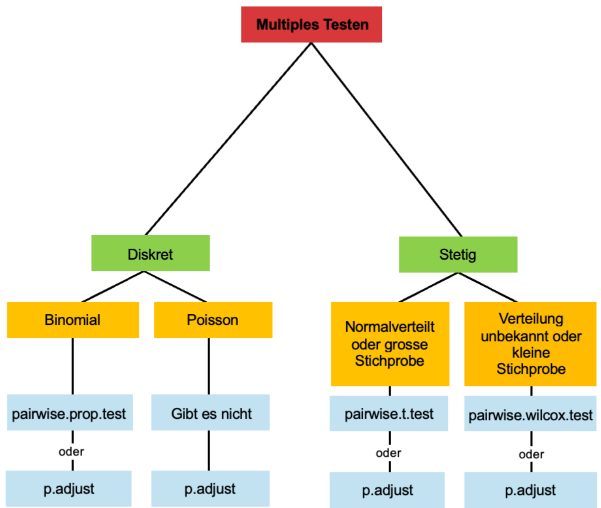
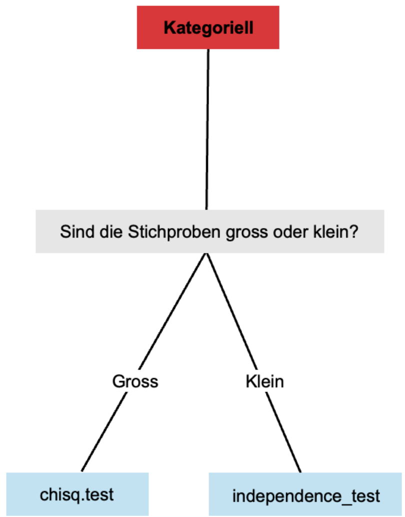
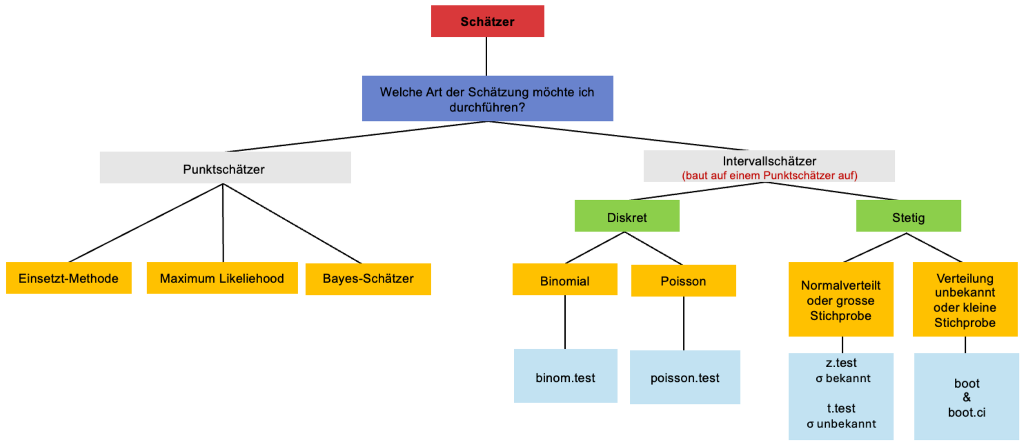

1 Stichprobe
Welchen Datentyp hat die Stichprobe?
Diskret
Stetig
Binomial
Poisson
Normalverteilt oder grosse Stichprobe
Verteilung unbekannt oder kleine Stichprobe
binom.test
Poisson.test
z.test
σ bekannt
t.test
σ unbekannt
wilcox.test


2 Stichproben
Sind die Stichproben gepaart oder ungepaart?
GEPAART
UNGEPAART
Welchen Datentyp hat die Stichprobe?
Diskret
Stetig
Binomial
Poisson
Normalverteilt oder grosse Stichprobe
Verteilung unbekannt oder kleine Stichprobe
Wurde nicht behandelt mcnemar.test
Wurde nicht behandelt Regressions-modell
t.test paired=TRUE
Kommen Bindungen vor?
JA
NEIN
wilcoxsign_test
wilcox.test paired=TRUE
Welchen Datentyp hat die Stichprobe?
Diskret
Stetig
Binomial
Poisson
Normalverteilt oder grosse Stichprobe
Verteilung unbekannt oder kleine Stichprobe
prop.test
poisson.test
z.test
σ bekannt
t.test
σ unbekannt
paired=FALSE
Kommen Bindungen vor?
JA
NEIN
Wilcox_test
wilcox.test paired=FALSE


Multiples Testen
Diskret
Stetig
Binomial
Poisson
Normalverteilt oder grosse Stichprobe
Verteilung unbekannt oder kleine Stichprobe
pairwise.prop.test
oder
p.adjust
Gibt es nicht
p.adjust
pairwise.t.test
oder
p.adjust
pairwise.wilcox.test
oder
p.adjust


Kategoriell
Sind die Stichproben gross oder klein?
Gross
Klein
chisq.test
independence_test


Schätzer
Welche Art der Schätzung möchte ich durchführen?
Punktschätzer
Einsetzt-Methode
Maximum Likeliehood
Bayes-Schätzer
Intervallschätzer
(baut auf einem Punktschätzer auf)
Diskret
Stetig
Binomial
Poisson
Normalverteilt oder grosse Stichprobe
Verteilung unbekannt oder kleine Stichprobe
binom.test
poisson.test
z.test
σ bekannt
t.test
σ unbekannt
boot
&
boot.ci

X 1

STATISTIK – LINUS STUHLMANN

MODELWAHL UND MODELLANNAHMEN

Berücksichtigungen für eine Modellwahl sind:

- Verteilungsannahmen
- Unabhängigkeit
- Homogenität

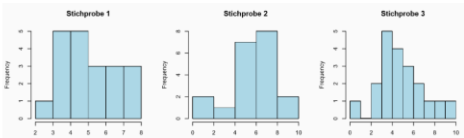
VERTEILUNGSANNAHMEN

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



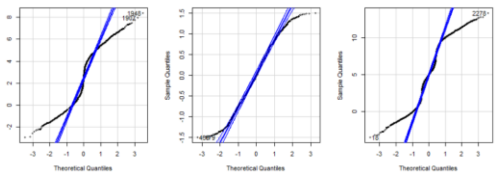
QQ-PLOTS

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

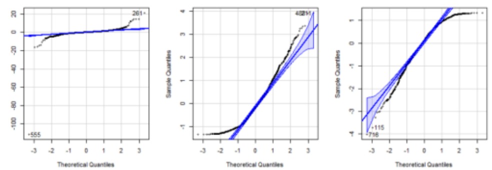
R – LIBRARY(CAR)

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



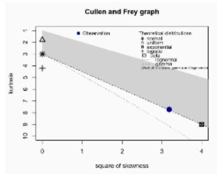
langschwänzig, rechtsschief, linksschief

CULLEN-FREY DIAGRAMM

Das Cullen-Frey Diagramm ist eine Alternative zum QQ

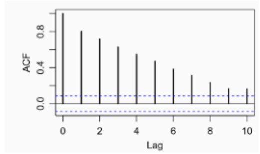
library(fitdistrplus)

descdist(data, print=FALSE)



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 $X_1, \ldots, X_n$ 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.

$\{\hat{\mu}, \hat{\sigma}^2 \ldots\} = \theta$ = Teta

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 unbekannten Parameters θ verwendet werden, nennt man **Schätzer**:

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

- Ziel
  - Schätzung für unbekannten Parameter θ der Population
- Beispiel
  - Schätzer für Mittelwert einer Population
  - → Mittelwert der Stichprobe

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

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

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

SCHÄTZER FÜR ERWARTUNGSWERT

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

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

- Erwartungswert von $\bar{X}_n$ gleicht dem gesuchten Parameter: $\boldsymbol{E(\bar{X}_n) = \mu}$
- 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 $\boldsymbol{n}$** gilt:
  - $\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$
  - **Zentraler Grenzwertsatz** → Je mehr Mittelwerte gezogen werden umso näher an Normalverteilung

SCHÄTZER FÜR VARIANZ Korrigiert

Unkorrigiert

$\frac{1}{n}$ = Verzerrt nach unten

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

Der Erwartungswert des Schätzers gleicht dem gesuchten Parameter:

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

Bernoulli

$\hat{p} = \bar{x} = \frac{1}{n} \cdot \sum x_i$

Binomial

$\hat{p} = \frac{\bar{x}}{n}$

Poisson

$\hat{\lambda} = \bar{x}$

Verzerrte Schätzer als Beispiel:

- Varianz mit $\frac{1}{n}$
- Exponential Parameter mit $\hat{\lambda} = \frac{1}{\bar{x}}$
- sd mit $\frac{1}{n}$
- nicht lineare Transformationen von $\bar{x}$

X 3

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
- Mittelpunk 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.

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

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

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

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.

- $\boxed{\text{kleiner } MSE_\theta(T) \rightarrow \text{effizient}}$ Kleiner Wert = Besser

Schätzer mit kleinster VARIANZ = effizientester Schätzer

MSE

Abwägung zwischen positiven Eigenschaften
Erwartungstreue vs. effizienz   Verzerrtheit vs. Varianz

Mittlere quadratische Abweichung (mean squared error)

$$E[(\hat{\theta} - \theta)^2] = Var(\hat{\theta}) + Bias(\hat{\theta})^2$$

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

mean((t_mean - mu_true)²)

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

$E[(\hat{\mu} - \mu)^2]$

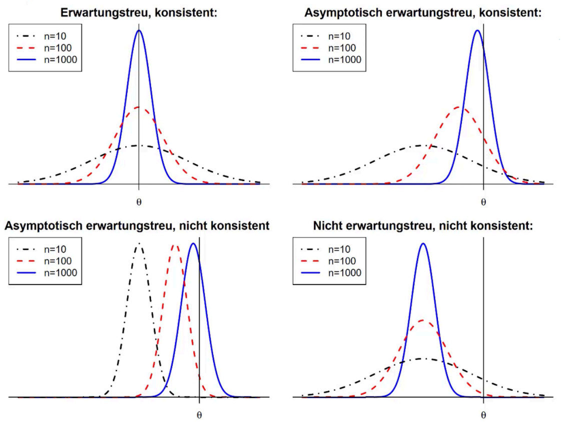
- dat: verschiedene Realisationen des Schätzers

KONSISTENZ

Konsistenz bedeutet, dass ein Schätzer mit wachsendem Stichprobenumfang n immer näher beim wahren Wert von θ liegt. Gesetz der grossen Zahlen!

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

Voraussetzungen:
- $\hat{\theta}$ ist asymptotisch erwartungstreu
- Die Varianz von $\hat{\theta}$ geht gegen 0 wenn $n \to \infty$



ÄQUIVARIANZ

Äquivarianz bedeutet, dass Schätzer bestimmte Transformationen zulassen, ohne dass sich der Schätzer verändert. C° → F°, 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 $\quad \boxed{\hat{\mu}_X = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i}$
  - Erwartungstreue
    - Wenn fair, unverzerrt
  - Konsistenz
    - Genug Grosses n.
  - Effizienz
- **Varianz** einer Stichprobe $\quad \boxed{\widehat{\sigma^2}_X = \hat{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2}$
  - Erwartungstreue
    - $\left(\frac{1}{n-1}\right)$ Korrektur
  - Konsistenz
    - Genug grosses n.
  - Effizienz
    - Wenn normalverteilt

Diese Schätzer sind:
- unverzerrt
- Effizient
- Konsistent

Gegebene Funktion in R definieren

$$f_\theta(x) = \begin{cases} \theta x^{\theta-1} & x \in (0,1) \\ 0 & \text{sonst} \end{cases}$$

```
f_theta <- function(x, theta) {
  ifelse(x > 0 & x < 1, theta * x^(theta - 1), 0)
}
```

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 Momenten-Methode

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

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

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

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 unbekannten Parameters θ die Plausibilität der beobachteten Stichprobe misst.

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

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

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

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 $-\frac{n}{\lambda^2} < 0$ → **Maximum**

S

1. Definiere Log Likelihood
2. Erzeuge ein Gitter für mu und sigma
3. Berechne Log Likelihood für jede Gitter Kombination
4. Wähle die Kombination mit maximaler Log Likelihood

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)

1. Du schreibst eine Funktion, die die negative Log Likelihood zurückgibt.
2. Du gibst Startwerte.
3. Du lässt `optim()` minimieren.
4. Du liest die geschätzten Parameter aus `res$par`.

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 → 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 → split
  mu <- params[1]
  sigma <- params[2]
  return(-sum(log(dnorm(dat, mean = mu,
sd = sigma))))}
```

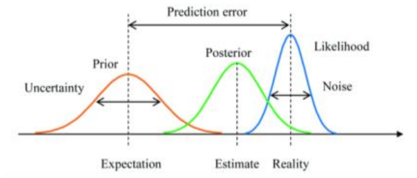
4.
```
# Optimierung durch Newtonverfahren
optim(par = c(mu, sigma), fn =
log_like, method = "L-BFGS-B", dat =
data, lower = 0, upper = Inf)
```

AUSGABE

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

BAYES-SCHÄTZER

Bayes Schätzer sind eine Methode der statistischen Inferenz, der 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 unbekannten Parameter (vorhandenes Vorwissen sollte mit einfliessen)
3. Man erhebt die Stichprobe

$$\underbrace{P(\Theta|\text{Daten})}_{\text{A-Posteriori-Verteilung}} = \frac{\overbrace{P(\text{Daten}|\Theta)}^{\text{Likelihood}}}{P(\text{Daten})} \cdot \underbrace{P(\Theta)}_{\text{A-Priori-Verteilung}}$$

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

$$h(\theta|x_1,\dots,x_n)$$

$$= \frac{\overbrace{f(x_1,\dots,x_n|\theta)}^{Likelihood} * \overbrace{g(\theta)}^{Priori}}{\int f(x_1,\dots,x_n|\theta) * g(\theta)d\theta}$$

$$= \frac{\prod_1^n f(x_i|\theta) * g(\theta)}{\int \prod_1^n f(x_i|\theta) * g(\theta)d\theta}$$

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

Regel	Formel
Definition	$\log_a(b) = x \iff a^x = b$
Produktregel	$\log_a(xy) = \log_a(x) + \log_a(y)$
Quotientenregel	$\log_a\left(\frac{x}{y}\right) = \log_a(x) - \log_a(y)$
Potenzregel	$\log_a(x^k) = k \cdot \log_a(x)$
Wurzelregel	$\log_a(\sqrt[n]{x}) = \frac{1}{n}\log_a(x)$
Basiswechsel	$\log_a(x) = \frac{\log_b(x)}{\log_b(a)}$
Logarithmus der Basis	$\log_a(a) = 1$
Logarithmus von 1	$\log_a(1) = 0$
Umkehrung zur Exponentialfunktion	$a^{\log_a(x)} = x$
Logarithmus eines Exponenten	$\log(e^x) = x$ (natürlicher Logarithmus)

Beispiel

$$\ell(\lambda) = \sum_{i=1}^{n} \log\left(\frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right)$$

$$\ell(\lambda) = \sum_{i=1}^{n}[\log(e^{-\lambda}) + \log(\lambda^{x_i}) - \log(x_i!)]$$

$$\ell(\lambda) = \sum_{i=1}^{n}[-\lambda + x_i\log(\lambda) - \log(x_i!)]$$

Regel	Formel
Definition	$\sum_{i=1}^{n} a_i = a_1 + a_2 + \cdots + a_n$
Konstante ausklammern	$\sum_{i=1}^{n} c\,a_i = c\sum_{i=1}^{n} a_i$
Summenregel	$\sum_{i=1}^{n}(a_i + b_i) = \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i$
Differenzregel	$\sum_{i=1}^{n}(a_i - b_i) = \sum_{i=1}^{n} a_i - \sum_{i=1}^{n} b_i$
Aufteilung einer Summe	$\sum_{i=1}^{n} a_i = \sum_{i=1}^{k} a_i + \sum_{i=k+1}^{n} a_i$
Indexverschiebung	$\sum_{i=1}^{n} a_{i+1} = \sum_{j=2}^{n+1} a_j$
Konstante Summe	$\sum_{i=1}^{n} c = n\,c$
Lineare Terme	$\sum_{i=1}^{n}(a\,i + b) = a\sum_{i=1}^{n} i + b\sum_{i=1}^{n} 1$
Einsersumme	$\sum_{i=1}^{n} 1 = n$
Natürliche Zahlen	$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$
Quadrate	$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$
Kuben	$\sum_{i=1}^{n} i^3 = \left(\frac{n(n+1)}{2}\right)^2$
Geometrische Reihe	$\sum_{i=0}^{n} a^i = \frac{a^{n+1}-1}{a-1}$ für $a \neq 1$
Exponential ab 1	$\sum_{i=1}^{n} a^i = \frac{a^{n+1}-a}{a-1}$ für $a \neq 1$
Log Summe	$\sum_{i=1}^{n} \log(x_i) = \log\left(\prod_{i=1}^{n} x_i\right)$

Je grösser die Stichprobe (n) desto genauer/schmaler wird das Konfidenzintervall

Länge eines KI's:

z = q

$$\boxed{L = 2\,z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}}$$ $$\boxed{n = \left(\frac{2z_{1-\alpha/2}\sigma}{L}\right)^2}$$ $$\boxed{\sigma = \frac{L\sqrt{n}}{2z_{1-\alpha/2}}}$$ $$\boxed{z_{1-\alpha/2} = \frac{L\sqrt{n}}{2\sigma}}$$

KONFIDENZINTERVALLE - KI $[\hat{\theta} \pm \hat{G}]$

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 z.test

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

Standardfehler

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

- $\bar{X}_n$: Stichprobenmittelwert = $\hat{\mu}$
- σ: bekannte Standardabweichung
- n: Stichprobenumfang
- $q_{1-\frac{\alpha}{2}}$: Quantilswert der Normalverteilung
  - z-transformiert (da $\frac{\sigma}{\sqrt{n}}$)
  - für 0.95 qnorm(0.975)
  - für 0.99 qnorm(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 t.test

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

- grosses n → annähernd Normalverteilt

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

- $\bar{X}_n$: Stichprobenmittelwert = $\hat{\mu}$
- $\hat{\sigma}$: geschätzte Standardabweichung
  - Stichproben Standardabweichung
- n: Stichprobenumfang
- $qt_{\left(1-\frac{\alpha}{2},n-1\right)}$: Quantilswert 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 t.test

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

- → t-Verteilung

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

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:

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

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

KONFIDENZINTERVALL FÜR PROPORTIONEN binom.test

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

$\hat{p} = \frac{\text{Erfolg}}{n}$

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

- `binom.test(x= successes, n=trials)$conf.int` } Intervall ist in %

Erfolg, Misserfolg ⇒ Bernoulli-Verteilung
binom-Verteilung (mehrere Bernoulli Versuche)
0:1

BOOTSTRAP KONFIDENZINTERVALLE

boot
boot ci

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!**

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

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

BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 500 bootstrap replicates
```

1. Schätzer definieren
2. Schätzer als Funktion
3. Generieren von Bootstrap-Beispielen mit boot(data, statistic, R)
4. Bootstrap-Konfidenzintervall mit boot.ci(boot.out, conf, type)

Stichprobe
Schätzer
Anzahl Bootstrap Replikationen
Zurückgegebenes Objekt boot()
Konfidenz-intervall (z.B 0,95)
kann auch als Vektor übergeben werden
Optionen: "Norm", "Basic", "Stud", "Perc", "Bca", "All"

Ausgabe:

```
Bootstrap Statistics :
    original    bias    std. error
t1*  20833.5 47.34795     511.2575
```

original = Schätzer θ = 20833.5

bias = $\hat{\theta} - \theta$ = 47,34... überschätzt also!

Std.error = $\sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$ = 511.25

Schwankung von Stichprobe zu Stichprobe

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

```r
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	$(1-\alpha)$-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	$\left[\bar{X}_n - \frac{s_n}{\sqrt{n}} qt_{(1-\frac{\alpha}{2},(n-1))}, \bar{X}_n + \frac{s_n}{\sqrt{n}} qt_{(1-\frac{\alpha}{2},(n-1))}\right]$ `t.test(x,conf.level=1−α)$conf.int`
$E[X]$ für beliebige Verteilung (bei grossem n)	$\left[\bar{X}_n - \frac{s_n}{\sqrt{n}} qt_{(1-\frac{\alpha}{2},(n-1))}, \bar{X}_n + \frac{s_n}{\sqrt{n}} qt_{(1-\frac{\alpha}{2},(n-1))}\right]$ `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);schaetzer <- 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: $n * k^2$
  - ½*KI = $n *$ 4
- KI um Faktor k vergrössern: $n * 1/k^2$
  - 2*KI: $n *$ ¼

KONFIDENZNIVEAU BESTIMMEN

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

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

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

$$Konf.Niveau = 1 - \alpha$$

N BESTIMMEN

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

p-Wert $\leq \alpha$ $\Rightarrow$ H_0 verwerfen
p-Wert $> \alpha$ $\Rightarrow$ H_0 nicht verwerfen

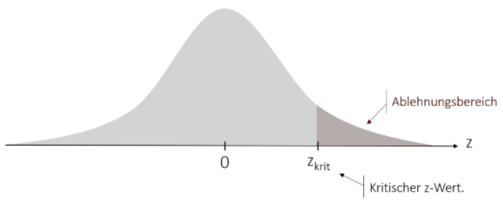
STATISTISCHES TESTEN

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

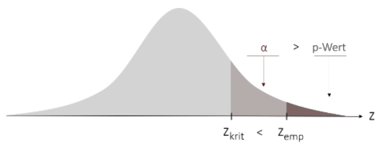
ANNAHMEBEREICH



- 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=teststatistik, df =n-1, 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
H_0 ist falsch	Fehler 2. Art	richtige Entscheidung

p-Wert < α durch puren Zufall

p-Wert nicht klein genug obwohl H_A wahr

$(1-\alpha)$ = Signifikantsniveau

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

$(1-\beta)$ = Macht / Power

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

$\boxed{1-\beta \geq \alpha}$

DURCHFÜHRUNG TEST

1. Formulierung der Nullhypothese H_0
   a. Meist keine Veränderung
2. Formulierung der Alternativhypothese H_A

   $H_A : p \neq p_0$ beidseitig "two.sided"
   $H_A : p < p_0$ nach unten (einseitig) "less"
   $H_A : p > p_0$ nach oben (einseitig) "greater"

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

```
A <- rpois(30, lambda = 2.5)
B <- rpois(30, lambda = 1.8)
```

⇒ `sum(A)` Ereignisse in `length(A)` Tagen
`sum(B)` Ereignisse in `length(B)` Tagen

Stichproben aus mehreren Tagen als Summe betrachten

T-TEST 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: $\boxed{t = \sqrt{n}\frac{\bar{X}-\mu_0}{\hat{\sigma}}}$ mit $\bar{x} = \hat{\mu}$
- p-Wert: pnorm(t, df=n-1, lower.tail)

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

        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
```

Stichprobe
H_0
H_A
p-Wert (falls <α wird H_0 verworfen
Konfidenzintervall
Schätzer

Z-TEST 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")

        One-sample z-Test

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
```

Package
Stichprobe muss existieren, kann künstlich generiert werden
H_A
H_0
bekanntes σ
p-Wert (falls <α wird H_0 verworfen
Konfidenzintervall
Schätzer

BINOM-TEST 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)

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

        Exact binomial test

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
```

Beobachtungen
Versuche
H_0
H_A
p-Wert (falls <α wird H_0 verworfen
Konfidenzintervall
Schätzer

POISSON TEST 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)

```
poisson.test(x = 4, T = 1, r = 9, alternative = "less")

        Exact Poisson test

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
```

Beobachtungen
Beobachtungsfenster (zeitlich/räumlich)
H_A
H_0
p-Wert (falls <α wird H_0 verworfen
Konfidenzintervall
Schätzer

2 STICHPROBENTEST paired= FALSE

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

2 STICHPROBEN BINOMIALTEST prop. test

Vergleich von zwei Anteilen p_1 und p_2.

- Nullhypothese: $H_0: p_1 = p_2$
  - $p_1 - p_2 = 0$
- Alternativhypothese: $H_A: p_1 \neq p_2$
  - $p_1 <\ngtr p_2$
- 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 p1-p2

Schätzung für p1 und p2

2 STICHPROBEN POISSON TEST 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 lambda1/lambda2

Schätzer für lambda1/lambda2

2 STICHPROBEN T-TEST 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

H_0 (d.h. mu1-mu2)

H_A

```
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 mu1 – mu2

Schätzer für mu1 und mu2

GEPAARTE 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

GEPAARTER TEST paired= TRUE

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

$$Differenz = D_1, \dots, D_n$$

- 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:

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

In R:

```
t.test(bs$Linux - bs$Windows, mu = 0, alternative = "two.sided")
```

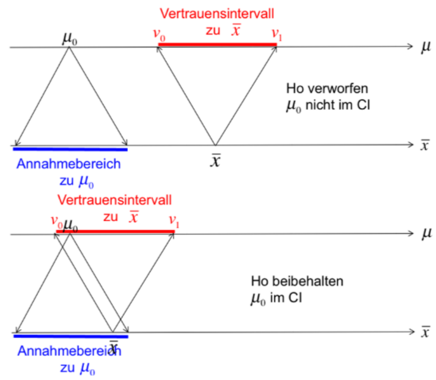
oder beide Stichproben an `t.test()` übergeben:

```
t.test(x=bs$Linux, y=bs$Windows, paired = TRUE, mu = 0, alternative = "two.sided")
```

DUALITÄT KI UND TEST

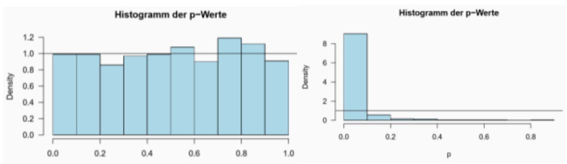
Welche Beobachtungen sind vereinbar mit H_0 bzw. mit einem bestimmten Sollwert für μ_0?

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:

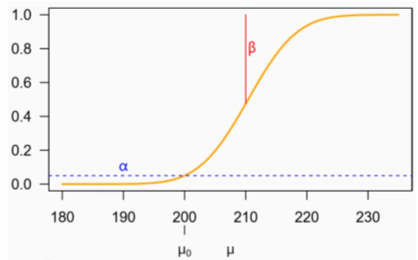
$$g(\mu) = P(H_0\ ablehnen|\mu)$$

$g(\mu)$ 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.

Wahrscheinlichkeit, einen echten Effekt zu finden

MACHT

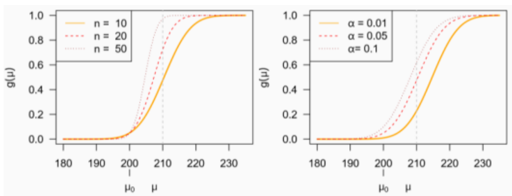
Die Macht (Power) bezeichnet die Wahrscheinlichkeit, eine Nullhypothese zu verwerfen, wenn sie wirklich falsch ist. Das heisst die Wahrscheinlichkeit einen Fehler 2. Art zu vermeiden.

$$Power = 1 - \beta$$

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

Die Macht ist abhängig von:

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

```
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.

$$\boxed{FWER = 1 - (1 - \alpha)^n}$$

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 p.adjust pairwise.t.test

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

```
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

Nachteil: Güte bzw. Macht wird kleiner

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

TESTENTSCHEIDE UND P-WERTE

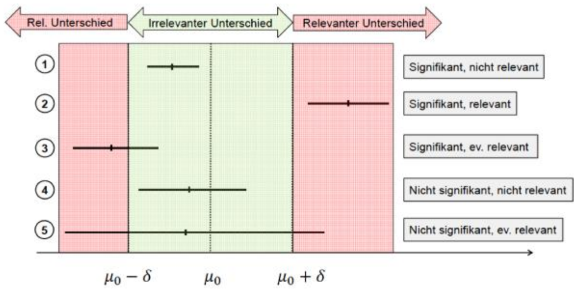
Testentscheidungen sind binär. Es gibt Situationen, die eine solch harte Schwelle erfordern (Medizin).

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

- p-Werte < 0.01: Effekt ziemlich sicher
- 0.01 < p-Werte ≤ 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.

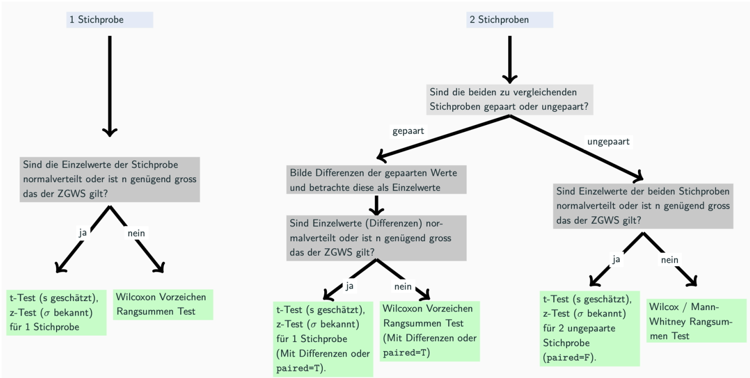


NICHT PARAMETRISCHES TESTEN

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



WILCOXON-RANGSUMMENTEST

wilcox.test
wilcox_test
wilcoxsign_test

`wilcox.test`
Berücksichtigt keine Bindungen

`wilcox_test`
Berücksichtigt Bindungen (ungepaart)

`wilcoxsign_test`
Berücksichtigt Bindungen (gepaart)

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

WILCOX-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

chisq.test

Der Test ziel darauf ab, zu prüfen, ob zwei kategorielle 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
  - $\chi_c^2 = \sum \left[\frac{(O-E)^2}{E}\right]$
  - $= \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{\left(n_{ij} - \left[\frac{n_{i\circ} n_{\circ j}}{n}\right]\right)^2}{\frac{n_{i\circ} n_{\circ j}}{n}}$

Berechnung von E:

$$E = \frac{Spaltensumme \, * \, Zeilensumme}{Stichprobengrösse \; n}$$

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 independence_test

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. $\frac{Anzahl\ von\ T_i^p > T_{obs}}{k}$

```
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)
```

Diskret

Bernoulli Verteilung

$X \sim \text{Bernoulli}(p)$

$$P(X = x) = p^x(1-p)^{1-x}, \quad x \in \{0,1\}$$

Log Likelihood

$\hat{p} = \bar{x}$

$$\ell(p) = \sum_{i=1}^{n} [x_i \log p + (1-x_i)\log(1-p)]$$

Binomial Verteilung

$X \sim \text{Binom}(n,p)$

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Log Likelihood

$\hat{p} = \frac{\bar{x}}{n}$

$$\ell(p) = \sum_{i=1}^{n} \left[x_i \log p + (n-x_i)\log(1-p)\right] + \text{Konstante}$$

Poisson Verteilung

$X \sim \text{Pois}(\lambda)$

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Log Likelihood

$\hat{\lambda} = \bar{x}$

$$\ell(\lambda) = \sum_{i=1}^{n} [x_i \log \lambda - \lambda] + \text{Konstante}$$

Geometrische Verteilung

$X \sim \text{Geom}(p)$

$$P(X = x) = p(1-p)^x$$

Log Likelihood

$\hat{p} = \frac{1}{\bar{x}+1}$

$$\ell(p) = \sum_{i=1}^{n} [\log p + x_i \log(1-p)]$$

Negative Binomial Verteilung

$X \sim \text{NB}(r,p)$

$$P(X = x) = \binom{x+r-1}{x} p^r (1-p)^x$$

Log Likelihood

$\hat{p} = \frac{r}{r+\bar{x}}$

$$\ell(p) = \sum_{i=1}^{n} [r \log p + x_i \log(1-p)] + \text{Konstante}$$

Stetig

Gleichverteilung Uniform

$X \sim \text{Unif}[a,b]$

Dichte

$$f(x) = \frac{1}{b-a}, \quad a \le x \le b$$

Log Likelihood

$$\ell(a,b) = -n\log(b-a) \quad \text{falls alle } x_i \in [a,b]$$

Sonst

$$\ell = -\infty$$

Wichtig

Die Daten bestimmen die Schranken. ML ergibt

$\hat{a} = \min . x_i$, $\hat{b} = \max . x_i$

Exponentialverteilung

$X \sim \text{Exp}(\lambda)$

Dichte

$$f(x) = \lambda e^{-\lambda x}, \quad x \ge 0$$

Log Likelihood

$\hat{\lambda} = \frac{1}{\bar{x}}$

$$\ell(\lambda) = n \log \lambda - \lambda \sum x_i$$

Weibull Verteilung

$X \sim \text{Weibull}(\beta, \lambda)$

Dichte

$$f(x) = \lambda\beta(\lambda x)^{\beta-1} e^{-(\lambda x)^\beta}$$

Log Likelihood

$$\ell(\beta,\lambda) = n\log\beta + n\beta\log\lambda + (\beta-1)\sum \log x_i - \sum (\lambda x_i)^\beta$$

Lognormalverteilung

$X \sim \log\mathcal{N}(\mu, \sigma^2)$

Dichte

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right)$$

Log Likelihood

$$\ell(\mu,\sigma) = -\sum \log x_i - n\log\sigma - \frac{1}{2\sigma^2}\sum(\log x_i - \mu)^2 + \text{Konstante}$$

$\hat{\mu} = \log(\bar{x})$ $\quad \hat{\sigma}^2 = \frac{1}{n}\sum(\log(x_i) - \hat{\mu})^2$

Gamma Verteilung

$X \sim \Gamma(k,\lambda)$ $\quad \boxed{\Gamma(k) = (k-1)!}$

Dichte

$$f(x) = \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x}$$

Log Likelihood

$$\ell(k,\lambda) = nk\log\lambda - n\log\Gamma(k) + (k-1)\sum\log x_i - \lambda\sum x_i$$

Hier gibt es meist **keine geschlossene ML Lösung**

Numerische Optimierung nötig.

k gegeben: $\hat{\lambda} = \frac{k}{\bar{x}}$

Normalverteilung

Fall 1

µ unbekannt, σ bekannt

Dichte

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Log Likelihood

$\hat{\mu} = \bar{x}$

$$\ell(\mu) \propto -\sum(x_i - \mu)^2$$

Fall 2

µ und σ unbekannt

Log Likelihood

$$\ell(\mu,\sigma) = -n\log\sigma - \frac{1}{2\sigma^2}\sum(x_i-\mu)^2 + \text{Konstante}$$

$\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = \frac{1}{n}\sum(x_i - \bar{x})^2$

allgemein gilt:

$$\boxed{\frac{d}{d\lambda}\left(\sum_{i=1}^{n} f_i(\lambda)\right) = \sum_{i=1}^{n} \frac{d}{d\lambda} f_i(\lambda)}$$

Aussage	Bedeutung	R-Befehl
Höchstens x	$P(X \le x)$	pbinom(x, size, prob)
Mindestens x	$P(X \ge x)$	1 - pbinom(x - 1, size, prob)
Genau x	$P(X = x)$	dbinom(x, size, prob)
Mehr als x	$P(X > x)$	1 - pbinom(x, size, prob)
Weniger als x	$P(X < x)$	pbinom(x - 1, size, prob)

Es darf auch z.B. 0:20 angegeben werden für X
so wird angegeben x=0, x=1, x=2 ... x=20

Wichtig bei diskreten Verteilungen < oder ≤ ?
bei > oder ≥
$\geq \to -1$, $< \to -1$, $\le \to -0$, $> \to -0$

$P(X=3)$	dbinom(x=3, size=13, prob=1/3)
$P(X \le 3)$	pbinom(q=3, size=13, prob=1/3)
$P(X \ge 3) = 1 - P(X \le 2)$	1-pbinom(q=2, size=13, prob=1/3)
$P(2 \le X \le 9) = P(X \le 9) - P(X \le 1)$	(pbinom(q=9, size=13, prob=1/3)-pbinom(q=1, size=13, prob=1/3))

Dichtefunktion d... genau x

Kumultativ Plus alles darunter

Verteilungsfunktion p... höchstens $X = x \le$ rest siehe Tabelle

quantil → gibt den x-Wert zurück

Disktrete Verteilungen

Verteilung	W'keitsfunktion P(X=x)	Wertebereich	R (.. = d/p/q/r) random → n= Anzahl Zufallszahlen	Erwartungswert und Varianz	Anwendung
Diskrete Gleichverteilung Laplace	$P(X=x_i)=p$ → günstige Fälle / alle mögliche Fälle	$\{x_1,\ldots,x_k\}$	`werte <- c(x1,x2, ... xk)` `sample(werte, size = 1)` Anzahl Würfel die gezogen werden; z.B. Würfel 1:6 also Zahlen von 1-6	$E(X)=\sum_i x_i p$ $Var(X)=\sum_{i=1}(x_i-E[X])^2 p$	Alle Ereignisse sind gleich wahrscheinlich. Bsp: Würfeln
Bernoulli W' $X \sim$ Bernoulli(p)	$P(X=1)=pP(X=0)=1-p$	$\{0,1\}$ Erfolg(1) oder Misserfolg(0)	`sample(0:1, size=1,` → nur 1 Versuch `prob = c(1-p, p)`	$E(X)=p$ → Erfolg W' $Var(X)=p(1-p)$ → Misserfolg W'	Indikator, ob ein Ereignis eintrifft. Bsp: Münzwurf, Würfeln einer 6
Binomial $X \sim$ Binom(n,p) Anzahl Versuche, W'	$P(X=k)=\binom{n}{k}\cdot p^k\cdot(1-p)^{n-k}$ Kombinatorik, W' Erfolg, W' Misserfolg; k= wie oft Erfolg eingetreten ist	$\{0,1,\ldots,n\}$	`..binom(.., size=n, prob=p)` Anzahl Versuche, Erfolgs W'; dbinom x= 10 Anzahl Erfolge; rbinom( ) → gibt 0 oder 1 zurück	$E(X)=np$ → Anzahl Versuche · W' $Var(X)=np(1-p)$	Anzahl Erfolge in n unabhängigen Bernoulli-Versuchen. Bsp: Anzahl Studenten, die Prüfung bestehen
Poisson E(x) Durchschnitt $X \sim$ Pois(λ) Ereignis in einem festen Zeitraum	$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$	$\{0,1,\ldots,\infty\}$	`..pois(.., lambda = λ)` → Erwartungswert E(x)	$E(X)=\lambda$ $Var(X)=\lambda$ lamda = Ereignisrate = Rate	Anzahl Ereignisse k mit konstanter Eintreffrate λ. Ereignisse sind unabhängig! Bsp: Anzahl Kunden am Postschalter pro Stunde
Geometrisch $X \sim$ Geom(p)	$P(X=k)=p(1-p)^k$ Anzahl Fehlversuche, Erfolgs W'; je grösser p, desto schneller fällt die W' ab (weniger Fehlversuche bis zum Erfolg)	$\{0,1,\ldots,\infty\}$ Nur Fehlversuche zählen exklusive ersten Erfolg!	`..geom(.., prob = p)` dgeom (k, p)	$E(X)=(1-p)/p$ $Var(X)=(1-p)/p^2$	Anzahl Misserfolge k bis der erste Erfolg auftritt. Bsp: Anzahl Fehlwürfe, bis man das Haus verlassen bei Eile mit Weile kann
Negativ Binomial $X \sim$ NBinom(r,p)	$P(X=k)=\binom{k+r-1}{k}p^r(1-p)^k$ r= Anzahl Erfolge, k= Anzahl Misserfolge, Erfolgs W'	$\{0,1,\ldots,\infty\}$	`..nbinom(.., size=r, prob=p)` dnbinom (k, r, p); (r= 1) = geometrisch	$E(X)=r\cdot(1-p)/p$ $Var(X)=r\cdot(1-p)/p^2$	Anzahl Misserfolge bis zum r-ten Erfolg. Bsp: Anzahl erfolgloser Angelversuche bis man 2 Fische gefangen hat
Hypergeometrisch $X \sim Hyp(m,n,k)$	$P(X=j)=\frac{\binom{m}{j}\binom{n}{k-j}}{\binom{n+m}{k}}$ m= Anzahl Erfolgselemente, n= Anzahl nicht Erfolgselemente, k= Stichprobengrösse, j= Anzahl Erfolge in Stichprobengrösse, n+m= Grundgesamtheit	$\{0,1,\ldots,min(k,m)\}$	`..hyper(.., m=m, n=n , k=k)` dhyper (j, m, n, k); m= Anzahl Erfolgselemente, n= Anzahl nicht Erfolgselemente, k= Stichprobengrösse, j= Anzahl Erfolge in Stichprobengrösse	$E(X)=km/(m+n)$ $Var(X)=k\cdot(\frac{m}{m+n})\cdot(1-\frac{m}{m+n})\cdot(\frac{n+m-k}{m+n-1})$	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

Glockenkurve wenn n gross ist = Normalverteilung (stetig)

Kleines λ = stark rechtsschief; grosses λ>10 wird symmetrisch

ohne Zurücklegen

Diskret und Stetig

Dichte mittels Verteilung

	Diskrete Zufallsvariable	**Stetige** Zufallsvariable
Wertebereich	endlich oder unendlich abzählbar viele Werte an $W=\{x_1,x_2,\ldots\}$ 3	kann jeden beliebigen Wert eines Intervalls annehmen 3,141
Wahrscheinlichkeitsverteilung Dichte	Zu jedem möglichen Wert x_i gibt die Funktion die entsprechende Wahrscheinlichkeit p_i an: $P(X=x_i)=p_i$ Es gilt: $\sum_i p_i=1$	Funtion $f(x)\ge 0$ mit $\int_{-\infty}^{\infty}f(x)dx=1$ Achtung: f(x) gibt keine W'keiten an, $f(x)>1$ möglich
Kumulative Verteilungsfunktion	$F(x)=P(X\le x)=\sum_{x_i\le x}p_i$	$F(x)=P(X\le x)=\int_{-\infty}^{x}f(u)du$
Wahrscheinlichkeiten	$P(X=x_k)=p_k$ p_i = W', x_i = Wert der Eintrifft	$P(X\in[a,b])=\int_a^b f(u)du=F(b)-F(a)$
Erwartungswert (Lagemass, was man "im Schnitt" bei unendlich vielen Realisierungen von X erhält)	$E(X)=\sum_{i=1}^{\infty}x_iP(X=x_i)=\sum_{i=1}^{\infty}x_ip_i$	$E(X)=\int_{-\infty}^{\infty}xf(x)dx$
Varianz (Streuungsmass, "Breite" der Verteilung)	$Var(X)=E[(X-E[X])^2]=\sum_{i=1}^{\infty}(x_i-E[X])^2p_i$ Erwartungswert zuerst Rechnen	$Var(X)=\int_{-\infty}^{\infty}(x-E(X))^2f(x)dx$

$Var(X) = E(X^2) - [E(X)]^2$ | $E(X^2) = \sum x^2 \cdot P(X=x)$

$Cov(X,Y)=E((X-E(X))(Y-E(Y)))$

Rechenregeln Erwartungswert und Varianz

$E(aX+b)=a\cdot E(X)+b$ $E(a\cdot X) = a\,E(X)$

$E(X+Y)=E(X)+E(Y)$

$Var(aX+b)=a^2\cdot Var(X)$ $V(a\cdot X) = a^2\cdot V(X)$

$Var(X)=E((X-E(X))^2)=E(X^2)-(E(X))^2$

$Var(X+Y)=Var(X)+Var(Y)+2Cov(X,Y)$

Falls X und Y unabhängig, dann ist $Cov(X,Y)=0$

$\mathrm{Cov}(X,Y)=\mathrm{Corr}(X,Y)\cdot\sigma(X)\cdot\sigma(Y)$

Stetige Verteilungen

mit p... schaut man auf x-Achse wo man bei y liegt (W') Integral von 0 bis x (q)

mit q... schaut man auf y-Achse (W') was mein x ist

Verteilung	Dichte f(x) Keine W' wie stark bei x konzentriert	Verteilung F(X) ist eine W'	Kennzahlen und Zufallszahlen	Anwendung
Stetige Gleichverteilung **Uniform** $X \sim \text{Unif}([a,b])$ Intervall	$f(x) = \begin{cases} 0 & \text{falls } x < a \\ 1/(b-a) & \text{falls } a \le x \le b \\ 0 & \text{falls } x > b \end{cases}$ R: `dunif(x, min=a, max=b)` Höhe an Stelle x	$F(x) = \begin{cases} 0 & \text{falls } x < a \\ \frac{x-a}{b-a} & \text{falls } a \le x \le b \\ 1 & \text{falls } x > b \end{cases}$ R: `punif(q, min=a, max=b)` ↳ Quantile → x-Achse gegeben	→ liegt genau in der Mitte des Intervalls $E(X) = (a+b)/2$ $Var(X) = (b-a)^2/12$ Y-Achse gegeben Quantile: p = Aufsummierung W' R: `qunif(p, min=a, max=b)` Zufallszahlen: min, max: für den Intervall R: `runif(n, min=a, max=b)` z.B. 0-12 min	Wenn alle Werte in einem Bereich gleich wahrscheinlich sind. *Bsp:* Wartezeit auf Bus, der exakt alle 8 Minuten fährt
Exponential $X \sim \text{Exp}(\lambda)$	$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{falls } x \ge 0 \\ 0 & \text{falls } x < 0 \end{cases}$ R: `dexp(x, rate =`λ`)` → E(x) X-Achsen Wert Integralrechnung `pexp(0.5, rate=6) - pexp(0.25, rate=6)`	$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{falls } x \ge 0 \\ 0 & \text{falls } x < 0 \end{cases}$ R: `pexp(q, rate =`λ`)` W' bis x Wie gross ist die W', dass das Ereignis vor oder bei x eintrifft	$E(X) = \frac{1}{\lambda}$ $Var(X) = \frac{1}{\lambda^2}$ p = W' zwischen 0 & 1 Quantile: Resultat = x-Achsen Wert R: `qexp(p, rate =`λ`)` Zufallszahlen: R: `rexp(n, rate =`λ`)` gibt Lebensdauer zurück	Dauer von zufälligen Zeitintervallen (zwischen 2 Poisson-Ereignissen) Besonderheit: Gedächtnislosigkeit, d.h. $P(X > s+t \mid X > s) = P(X > t)$ *Bsp:* Wartezeit für Kunden
Weibull $X \sim \text{Weibull}(\beta, \lambda)$	$f(x) = \begin{cases} \lambda\beta(\lambda x)^{(\beta-1)} e^{-(\lambda x)^{\beta}} & \text{falls } x \ge 0 \\ 0 & \text{falls } x < 0 \end{cases}$ R: `dweibull(x, shape=`β`, scale=1/`λ`)`	$F(x) = \begin{cases} 1 - e^{-(\lambda x)^{\beta}} & \text{falls } x \ge 0 \\ 0 & \text{falls } x < 0 \end{cases}$ R: `pweibull(q, shape=`β`, scale=1/`λ`)` Form $\beta = 1$ Exponentialverteilung Streckung / Stauchung	(lambda <- gamma(1+1/2)/6) $E(X) = \frac{1}{\lambda}\Gamma(1+1/\beta)$ E(x) $Var(X) = \frac{1}{\lambda^2}[\Gamma(1+2/\beta) - \Gamma^2(1+1/\beta)]$ Quantile: R: `qweibull(p, shape=`β`, scale=1/`λ`)` Zufallszahlen: R: `rweibull(n, shape=`β`, scale=1/`λ`)`	Dauer von zufälligen Zeitintervallen mit Berücksichtigung der Lebensphase ($\beta < 1$ Fehlerrate abnehmend, $\beta = 1$ Fehlerrate konstant und $\beta > 1$ Fehlerrate zunehmend. *Bsp:* Lebensdauer von Bauteilen und elektrischen Komponenten
Gamma $X \sim G(k, \lambda)$	$f(x) = \begin{cases} e^{-\lambda x} \cdot x^{k-1} \cdot \frac{\lambda^k}{\Gamma(k)} & \text{falls } x \ge 0 \\ 0 & \text{falls } x < 0 \end{cases}$ R: `dgamma(x, shape=`k`, rate=`λ`)` wie viele Ereignisse werden abgewartet } k = 1 Exponentialverteilung	$F(x) = \int_0^x e^{-\lambda z} \cdot z^{k-1} \cdot \frac{\lambda^k}{\Gamma(k)} dz$ (keine geschlossene Form möglich) R: `pgamma(q, shape=`k`, rate=`λ`)`	$E(X) = \frac{k}{\lambda}$ $Var(X) = \frac{k}{\lambda^2}$ Quantile: R: `qgamma(p, shape=`k`, rate=`λ`)` Zufallszahlen: R: `rgamma(n, shape=`k`, rate=`λ`)`	Summe von exponentialverteilten Zufallsvariablen *Bsp:* Wartezeit an einer Schalterschlange, Kosten Wartezeit bis zum Eintreten von k Ereignissen
Glockenkurve **Normal** / Gauss $X \sim \mathcal{N}(\mu, \sigma^2)$ E(x) Mittelwert, Varianz Standardnormalverteilung $\mu = 0$, $\sigma^2 = 1$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$ R: `dnorm(x, mean=`μ`, sd=`σ`)` E(x); var = σ^2, $\sqrt{var} = \sqrt{\sigma^2}$, $\sqrt{var} = \sigma$	$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{1}{2}\frac{(z-\mu)^2}{\sigma^2}\right) dz$ (keine geschlossene Form möglich) R: `pnorm(q, mean=`μ`, sd=`σ`)`	$E(X) = \mu$ $Var(X) = \sigma^2$ Standard Normalverteilung = 0, = 1 Quantile: R: `qnorm(p, mean=`μ`, sd=`σ`)` Zufallszahlen: R: `rnorm(n, mean=`μ`, sd=`σ`)`	Abweichungen von (Mess-) Werten, Näherung bei grossen Stichproben *Bsp:* Körpergrösse, Messfehler, Abweichungen vom Soll einer der Produktion, Füllmenge von Packungen, Intelligenz
Lognormal $X \sim \log\mathcal{N}(\mu, \sigma^2)$ Eng verwandt Normalverteilung	$f(x) = \frac{1}{\sigma x\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(\log(x)-\mu)^2}{\sigma^2}\right)$ R: `dlnorm(x, mean=`μ`, sd=`σ`)` für x kann auch eine Sequenz angegeben werden z.B. seq(0, 2000, length.out = 1000) Erzeugt genau 1000 Punkte	$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma z\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(\log(z)-\mu)^2}{\sigma^2}\right) dz$ (keine geschlossene Form möglich) R: `plnorm(q, mean=`μ`, sd=`σ`)`	$E(X) = e^{(\mu + \frac{\sigma^2}{2})}$ $Var(X) = e^{(2\mu+\sigma^2)}(e^{\sigma^2} - 1)$ Quantile: R: `qlnorm(p, mean=`μ`, sd=`σ`)` Zufallszahlen: R: `rlnorm(n, mean=`μ`, sd=`σ`)`	rechtsschiefe Messdaten mit nur positiven Werten *Bsp:* Geflogene Meilen, Lohn

Zielparameter	Verteilung	Schätzer	R Funktion	Unverzerrt	Bemerkung
Erwartungswert μ	beliebig mit E(X)	Stichprobenmittel	mean(x)	Ja	linearer Schätzer
Varianz σ²	Normal	korrigierte Varianz	var(x)	Ja	teilt durch n minus 1
Standardabweichung σ	Normal	Wurzel der Varianz	sd(x)	Nein	Wurzel erzeugt Bias
Bernoulli p	Bernoulli	Stichprobenanteil	mean(x)	Ja	x enthält nur 0 und 1
Binomial p	Binomial	y geteilt durch n	y / n	Ja	y Anzahl Erfolge
Poisson λ	Poisson	Stichprobenmittel	mean(x)	Ja	Erwartungswert gleich Parameter
Exponential E(X)	Exponential	Stichprobenmittel	mean(x)	Ja	schätzt 1 durch λ
Exponential λ	Exponential	Kehrwert des Mittels	1 / mean(x)	Nein	nichtlinearer Schätzer
Median	allgemein	Stichprobenmedian	median(x)	Nein	robust aber verzerrt
Minimum	allgemein	Minimum	min(x)	Nein	Extremwert
Maximum	allgemein	Maximum	max(x)	Nein	Extremwert

p = Echter Parameter

1. Zufallsvariabel definieren

   X = Anzahl richtig getippten Gläser

2. Modellannahme notieren

   $X \sim Bin(n, p)$

3. Nullhypothese (H_0) definieren

   Meist keine Veränderung!

   $H_0: \quad p = 0{,}5$

   $p = p_0$

4. Alternativhypothese (H_A) definieren

   Zweiseitig: $p \neq p_0$ "two. sides"

   Einseitig: $p < p_0$ "less"

   Einseitig: $p > p_0$ "greater"

   $H_A: \quad p > 0{,}5$

   $p > p_0$

Aspekt des Histogramms	Was sehe ich	Was bedeutet das	Typische Prüfungsformulierung
Inhalt	Histogramm der Bootstrap Werte des Schätzers	Darstellung der empirischen Stichprobenverteilung des Schätzers	Das Histogramm zeigt die Bootstrap Verteilung des Schätzers
Horizontale Achse	Werte des Schätzers	Mögliche realisierte Schätzerwerte bei Wiederholung des Experiments	Die x Achse enthält mögliche Werte des Schätzers
Vertikale Achse	Häufigkeit	Wie oft ein Schätzerwert in den Bootstrap Stichproben vorkommt	Die Höhe der Balken beschreibt die Häufigkeit
Lage Zentrum	Mittelpunkt der Verteilung	Typischer Schätzwert und Plausibilitätskontrolle	Die Verteilung ist um den ursprünglichen Schätzwert zentriert
Breite	Streuung der Verteilung	Unsicherheit des Schätzers Standardfehler	Die Streuung der Bootstrap Verteilung reflektiert den Standardfehler
Form	Symmetrisch oder schief	Eignung der Normalapproximation	Die Verteilung ist annähernd symmetrisch
Schiefe	Links oder rechts schief	Hinweis auf Median Quantile oder kleine Stichprobe	Die Verteilung ist schief daher ist bca geeigneter
Mehrere Peaks	Mehrgipflige Struktur	Instabiler Schätzer oder kleine Stichprobe	Die Verteilung weist mehrere Modi auf
Ausreisser	Extreme Werte	Hohe Sensitivität des Schätzers	Einzelne extreme Bootstrap Werte sind sichtbar
Bezug zu KI	Lage der Ränder	Herleitung der Konfidenzintervalle	Konfidenzintervalle basieren auf der Bootstrap Verteilung

Sollwert eingehalten oder unterschritten wird

Es handelt sich um einen Test auf einen Sollwert. Falls wir davon ausgehen, dass die Standardabweichung unbekannt ist, so ist ein 1-Stichproben-t-Test passend. Die Hypothesen sind
H_0: „Die Schreibdauer beträgt mindestens 10 Stunden " $\mu \geq \mu_0$
H_A: „Die Schreibdauer beträgt weniger als 10 Stunden" $\mu < \mu_0$.

d) Wie sieht der Test und die Entscheidung aus, falls Sie sich nur dafür interessieren, ob sich die Schreibdauer vom Sollwert unterscheidet und die Teststatistik $t = -2.7$ und $\alpha = 0.01$ ist?

H_0: „Die Schreibdauer beträgt 10 Stunden" $\mu = \mu_0$ und
H_A: „Die Schreibdauer unterscheidet sich von 10 Stunden" $\mu \neq \mu_0$.