Testtheorie

Blockpraktikum zur Statistik mit R

29. März 2012

Sören Gröttrup

Institut für Mathematische Statistik
Universität Münster
SS 2012

Gliederung

Gliederung

Ausgangssituation

- 100 Geburten werden untersucht, von denen 54 Mädchen und 46 Jungen sind
- Aufstellen einer These:

  Die Wahrscheinlichkeit für eine Mädchengeburt ist höher als die für eine Jungengeburt.

$\leadsto$ Mit Hilfe der Testtheorie möchte man nun diese These bestätigen.

Vorgehen beim Testen

- Ziehen einer Stichprobe $(x_1, \ldots, x_n)$.
- Wir wollen von dieser Stichprobe auf die Grundgesamtheit schließen.
- Die Entscheidungsmöglichkeiten bezeichnet man als *(Null)Hypothese* H und *Alternative* K
- In unserem Falle:
  - Hypothese $\cong$ Geburt eines Jungen ist wahrscheinlicher
  - Alternative $\cong$ Geburt eines Mädchens ist wahrscheinlicher

Mögliche Fehler

- **Fehler 1. Art:** Es liegt die Hypothese vor, wir entscheiden uns aber für die Alternative
- **Fehler 2. Art:** Es liegt die Alternative vor, wir entscheiden uns aber für die Hypothese

Es gilt, im Hinblick auf die beiden möglichen Fehlentscheidungen eine möglichst "gute" Entscheidung zu treffen.

$\rightsquigarrow$ Die Fehler so klein wie möglich halten!

Achtung

Es ist i.A. nicht möglich beide Fehler zu minimieren!!

Mögliche Fehler

- **Fehler 1. Art:** Es liegt die Hypothese vor, wir entscheiden uns aber für die Alternative
- **Fehler 2. Art:** Es liegt die Alternative vor, wir entscheiden uns aber für die Hypothese

Es gilt, im Hinblick auf die beiden möglichen Fehlentscheidungen eine möglichst "gute" Entscheidung zu treffen.

$\rightsquigarrow$ Die Fehler so klein wie möglich halten!

Achtung

Es ist i.A. nicht möglich beide Fehler zu minimieren!!

Allgemeines Vorgehen in der Testtheorie

- Eine Stichprobe $x \in \mathfrak{X}$ ist gegeben
- Wir geben eine Schranke α für den Fehler 1. Art vor
- Wir wählen eine möglichst gute Testfunktion

  "Gut" bedeutet, dass die Testfunktion den "Verlust", bzw. die Wahrscheinlichkeit für die Fehler 1. und 2. Art in irgendeiner Form minimiert.

- Wir treffen mit Hilfe der gewählten Testfunktion abhängig von x eine Entscheidung für die Hypothese oder Alternative

  Können wir die Wahrscheinlichkeiten für beide Fehler zugleich minimieren?
  Nein!

Wahl der Testfunktion

Da man nicht beide Fehler gleichzeitig minimieren kann

- $\rightsquigarrow$ Absichern bzgl. des Fehlers 1. Art, d.h. diesen klein halten
- $\rightsquigarrow$ Fehler 2. Art klein machen

Wähle den Test nach folgender Optimalitätsregel:

Optimalitätsregeln

- Die Wahrscheinlichkeit für den Fehler 1. Art darf maximal $\alpha \in (0,1)$ betragen
- Die Wahrscheinlichkeit für den Fehler 2. Art soll möglichst gering sein

Wahl von Hypothese und Alternative

Nur bei Wahl der Alternative können wir davon ausgehen, mit geringer Wahrscheinlichkeit falsch zu liegen.

- $\rightsquigarrow$ Alternative ist daher die Aussage, die mit großer Sicherheit stimmen soll, wenn sie durch den Test bestätigt wird.
  - Die Hypothese wird dagegen so gewählt, dass ihr fälschliches Verwerfen (der Fehler 1. Art) der "schlimmere" Fehler ist.

Beispiel (Diagnose)

Ein Test gibt eine Indikation über eine Erkrankung.

Hypothese $\hat{=}$ der Patient ist krank

Obiges Beispiel: Alternative $\hat{=}$ "Es gibt mehr Mädchen- als Jungengeburten"

Stichprobenarten

Ein-Stichprobenfall: Testen eines Merkmals auf Grund einer einfachen Stichprobe $(x_1, \ldots, x_n)$ bzgl. der Kenngrößen seiner Verteilung (Mittelwert, Median, Vertfkt.)

Zwei-Stichprobenfall:

- Zwei unabhängige Stichproben $(x_1, \ldots, x_n)$, $(y_1, \ldots, y_n)$
- Ein Merkmal unter zwei Bedingungen am selben Objekt getestet. $\leadsto$ Verbundene Stichproben $(x_{1,1}, x_{1,2}), \ldots, (x_{n,1}, x_{n,2})$
- Zwei Merkmale am selben Merkmalsträger getestet. $\leadsto$ Verbundene Stichproben $(x_1, y_1), \ldots, (x_n, y_n)$

k-Stichprobenfall:

- Ein Merkmal unter k Bedingungen getestet
- k Merkmale am selben Merkmalsträger getestet

Testtheorie vs. Schätztheorie

- Ein *Punktschätzer* ordnet einer Stichprobe x einen Wert aus dem Parameterraum zu.
  - Beispiel: Maximum Likelihood-Schätzer
- Ein *Bereichsschätzer* ordnet einer Stichprobe x eine Teilmenge des Parameterraums zu.
  - Beispiel: Konfidenzintervalle
- Eine *Testfunktion* wählt anhand einer Stichprobe x zwischen zwei Parameterbereichen, der Hypothese und der Alternative.

Gliederung

Testproblem

Definition

Ein *Testproblem* ist ein 4-Tupel $\mathcal{E} = (\mathfrak{X}, \mathfrak{A}, \mathcal{W}, H)$ mit

- einer nichtleeren Menge $\mathfrak{X}$, dem *Stichprobenraum*,
- einer σ-Algebra $\mathfrak{A}$ über $\mathfrak{X}$ und
- einer nichtleeren Familie $\mathcal{W}$ von W-Maßen auf $(\mathfrak{X}, \mathfrak{A})$
- einer Hypothese $H \subset \mathcal{W}$

- Die Alternative ist dann $K = \mathcal{W} \setminus H$
- Lassen sich die W-Maße in $\mathcal{W}$ parametrisieren, d.h. $\mathcal{W} = (P_\theta^X)_{\theta \in \Theta}$ für ein $\Theta \subset \mathbb{R}^d$, so kann man $H \subset \Theta$ wählen. $\rightsquigarrow$ *parametrisches Testverfahren*.
- Andernfalls $\rightsquigarrow$ *nicht-parametrisches* Testverfahren.

Arten von Testproblemen

In der parametrischen Testtheorie gibt es unterschiedliche Arten von Testproblemen: *einseitige* und *zweiseitige Testprobleme*.

Einseitige Testprobleme:

- $H : \theta \leq \theta_0$ gegen $K : \theta > \theta_0$
- $H : \theta \geq \theta_0$ gegen $K : \theta < \theta_0$

Zweiseitige Testprobleme:

- $H : \theta = \theta_0$ gegen $K : \theta \neq \theta_0$
- $H : \theta \in [c_0, c_1]$ gegen $K : \theta \notin [c_0, c_1]$ (nicht standardmäßig in R)
- $H : \theta \notin (c_0, c_1)$ gegen $K : \theta \in (c_0, c_1)$ (nicht standardmäßig in R)

Modellierung anhand eines Beispiels

Beispiel der Mädchengeburten:

- $\mathfrak{X} = \{0, 1\}^n$, wobei n die Anzahl der Versuchsbeobachtungen ist,
- $\mathfrak{A} = \mathfrak{P}(\mathfrak{X})$,
- $P_\theta^X = \bigotimes_{i=1}^n \mathfrak{B}(1, \theta)$, $\theta \in \Theta = (0, 1)$,

wobei $0 \mathrel{\widehat{=}}$ Jungengeburt und $1 \mathrel{\widehat{=}}$ Mädchengeburt

- $H = [0, 0.5]$ und $K = (0.5, 1]$

Während die Wahl von $\mathfrak{X}$ und $\mathfrak{A}$ kanonisch ist, liegen der Wahl von P_θ^X *Modellierungsannahmen* zugrunde.

*In unserem Falle liegt ein **einseitiges** Testproblem mit linksseitiger Hypothese vor.*

Modellierung anhand eines Beispiels

Beispiel der Mädchengeburten:

- $\mathfrak{X} = \{0, 1\}^n$, wobei n die Anzahl der Versuchsbeobachtungen ist,
- $\mathfrak{A} = \mathfrak{P}(\mathfrak{X})$,
- $P_\theta^X = \bigotimes_{i=1}^n \mathfrak{B}(1, \theta)$, $\theta \in \Theta = (0, 1)$,

wobei $0 \mathrel{\widehat{=}}$ Jungengeburt und $1 \mathrel{\widehat{=}}$ Mädchengeburt

- $H = [0, 0.5]$ und $K = (0.5, 1]$

Während die Wahl von $\mathfrak{X}$ und $\mathfrak{A}$ kanonisch ist, liegen der Wahl von P_θ^X *Modellierungsannahmen* zugrunde.

> *In unserem Falle liegt ein* ***einseitiges*** *Testproblem mit linksseitiger Hypothese vor.*

Tests

Definition

Jede messbare Funktion $\varphi : \mathfrak{X} \to [0,1]$ heißt *Test* oder *Testfunktion*.

Bemerkung

Interpretation eines Testwerts $\varphi(x)$:

$$\varphi(x) = \begin{cases} 1, & \varphi \text{ rät, die Alternative zu wählen,} \\ \gamma \in (0,1) & \varphi \text{ rät, ein unabh. Zufallsexp. durchzu-} \\ & \text{führen, das mit W'keit } \gamma \text{ zur Wahl} \\ & \text{der Alternative führt.} \\ 0, & \varphi \text{ rät, die Hypothese zu wählen,} \end{cases}$$

Fehlerwahrscheinlichkeiten

Bei einem gegebenen Test gelten:

$$\begin{aligned} E_\theta \varphi(X) &= \text{W'keit für den Fehler 1. Art, falls } \theta \in H, \\ 1 - E_\theta \varphi(X) &= \text{W'keit für den Fehler 2. Art, falls } \theta \in K. \end{aligned}$$

Eine optimale Testfunktion sähe so aus:

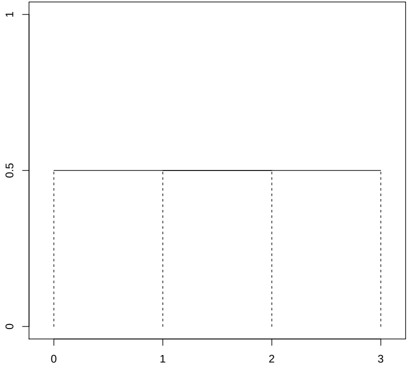
- Liegt die Hypothese vor, so liefert φ stets 0, ansonsten stets 1. Beide Fehler treten mit Wahrscheinlichkeit 0 auf.
- Formal:

$$\varphi(x) = \begin{cases} 1, & \text{falls } \theta \in K, \\ 0, & \text{falls } \theta \in H \end{cases}$$

Gibt es einen optimalen Test?

Heuristisches Beispiel gegen den optimalen Test

$$\mathfrak{X} = (0,3),\ \mathfrak{A} = \mathfrak{B}(0,3),\ H = \{R(0,2)\} \text{ gegen } K = \{R(1,3)\}$$

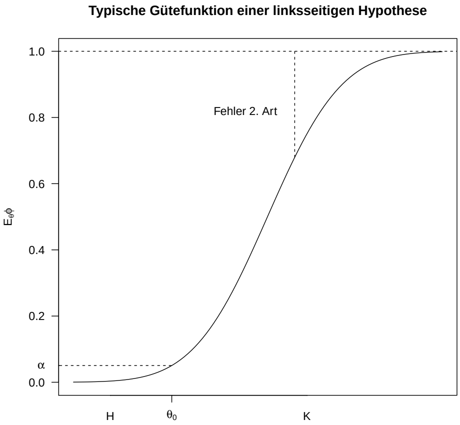


In diesem Fall gilt sogar: $P(Fehler 1.Art) + P(Fehler 2.Art) = 1$. (i.A gilt dies nicht)

Die Gütefunktion

Definition

Für einen Test φ heißt die Funktion $\theta \mapsto E_\theta\varphi(X)$ die *Gütefunktion* von φ.



Tests zum Niveau α

Definition

Sei $\alpha \in [0,1]$ ein vorgegebenes *Irrtums-* oder *Signifikanzniveau*. Dann heißt φ Test zum Niveau α, falls

$$E_\theta \varphi(X) \leq \alpha \quad \text{für alle } \theta \in H$$

gilt. Wir definieren Φ_α als die Menge der Tests zum Niveau α.

- Typische Werte für α: 0.01, 0.05, 0.1

Gleichmäßig bester Test zum Niveau α

Definition

φ heißt *gleichmäßig bester Test* zum Niveau α, falls φ ein Test zum Niveau α ist, der die W'keit für einen Fehler 2. Art unter allen Test zum Niveau α gleichmäßig minimiert. D.h.

$$E_\theta \varphi(X) \;=\; max_{\psi \in \Phi_\alpha} E_\theta \psi(X), \quad \forall\ \theta \in K.$$

Problem: Manchmal existiert kein gleichm. bester Test z.N. α (z.B. bei zweiseitigen Testproblemen oder t-Test). $\rightsquigarrow$ Übergang zu einer anderen Teilklasse von Tests.

Weitere Teilklassen:

- Unverfälschten Tests zum Niveau α.
- J-ähnlichen Tests z. N. α.

Unverfälschte Tests zum Niveau α

Definition

Ein Test φ heißt *unverfälscht zum Niveau α (für H vs. K)*, falls $E_\theta\varphi(X) \leq \alpha$ für alle $\theta \in H$ und $E_\theta\varphi(X) \geq \alpha$ für alle $\theta \in K$ gilt.

Definition

Ein Test φ heißt *glm. bester unverfälschter Test zum Niveau α*, falls φ ein unverfälschter Test z. N. α ist und die Wahrscheinlichkeit für einen Fehler 2. Art unter allen unverfälschten Tests z. N. α glm. minimiert.

Warum macht man nicht Nägel mit Köpfen und gibt gleichzeitig "$1 - E_\theta\varphi(X) \leq \alpha$ für alle $\theta \in K$" als Schranke für den Fehler 2. Art vor?

α-Fraktil und $(1-\alpha)$-Quantil

Definition

Für $\alpha \in (0,1)$ und ein W'Maß Q heißt

$$fr_\alpha(Q) = \inf\{y \in \mathbb{R} : Q\big((y,\infty)\big) \leq \alpha\}$$

α-Fraktil von Q.

- $(1-\alpha)$-Quantil und α-Fraktil stimmen überein, d.h.

$$fr_\alpha(Q) = F^{-1}(1-\alpha)$$

mit F die Verteilungsfunktion einer Verteilung Q und F^{-1} die Quantilsfunktion.

Gliederung

Ein-Stichproben-Fall

Es wird ein einziges Merkmal X auf der Basis einer einfachen Zufallsstichprobe $(X_1, \ldots, X_n)$ bzgl. interessierender Fragestellungen getestet, z. B. auf

- die Lage von Mittelwert oder Median im Vergleich zu vermuteten Werten - hierbei wird unterschieden zwischen
  - parametrischen Verfahren
  - verteilungsfreien Verfahren
- die Klasse der zugrundeliegenden Verteilung.

Mittelwertvergleich - Der t-Test

Annahme: $X_1, \dots, X_n$ u. i. v. mit $X_1 \sim \mathcal{N}(\mu, \sigma^2)$ bzw. beliebig verteilt mit ex. Varianz und großem n ($n \geq 30$)

Glm. beste unverfällschte Tests z.N. α:

- $H: \mu \leq \mu_0$ gegen $K: \mu > \mu_0$

$$\varphi(X) = \mathbf{1}_{\{T(X) > fr_\alpha(t_{n-1})\}}$$

- $H: \mu = \mu_0$ gegen $K: \mu \neq \mu_0$

$$\varphi(X) = \mathbf{1}_{\{|T(X)| > fr_{\alpha/2}(t_{n-1})\}}$$

mit Prüfgröße (**Teststatistik**)

$$T(X) = \sqrt{n}\frac{\overline{X} - \mu_0}{S(X)}.$$

t-Test in R

In R führt man mit folgendem Befehl einen t-Test aus:

```
t.test(x, alternative=.., mu=0).
```

- `x` sind die Argumente der Stichprobe
- `alternative=c("two.sided", "less", "greater")` Art der Alternative.
- `mu` kritischer Parameter

Ausgabe der Funktion `t.test`

- `mean(LakeHuron)` $\rightsquigarrow$ Ausgabe=579.0041
- `t.test(LakeHuron, mu=580, alternative="less")`

```
        One Sample t-test

data:  LakeHuron
t = -7.4786, df = 97, p-value = 1.691e-11
alternative hypothesis: true mean is less than 580
95 percent confidence interval:
-Inf 579.2252
sample estimates:
mean of x
579.0041
```

- Was ist der *p*-Wert?
- Wann entscheiden wir uns für die Alternative?

Der p-Wert

p-Wert

Der *p-Wert* ist das kleinste Niveau $\tilde{\alpha}$, zu dem man bei vorliegen der Stichprobe $x \in \mathfrak{X}$ und Prüfgröße $T(x)$ die Hypothese noch ablehnen kann.

Im Falle linksseitiger Hypothesen bedeutet dies formal
$\tilde{\alpha} = P_{\theta_0}(T(X) \geq T(x))$.

Bemerkung

Der Ablehnungsbereich der Hypothese lautet "$p \leq \alpha$".

$\rightsquigarrow$ Annahme der Alternative im obigen Beispiel, falls $\alpha \geq 1.691e - 11$.

Vorzeichentest - Test auf Wert des Medians

Überlegung: 50% der Beobachtungen sind größer und kleiner als der Median m_0. $\rightsquigarrow$ Anzahl der Beobachtungen größer m_0 sind $B(n, 0.5)$-verteilt.

Dies gilt nur, denn m_0 nicht in der Stichprobe auftritt. Ansonsten $\rightsquigarrow$ Einschränkung auf die von m_0 verschiedenen Beobachtungen.

Teststatistik:

$$T(X) = \sum_{i=1}^{n} \mathbf{1}_{(0,\infty)}(X_i - m_0)$$

Ablehnungsbereich:

$$H: m \leq m_0 \text{ vs } K: m > m_0 \qquad T(X) > fr_\alpha(B(n, 0.5))$$

$$H: m = m_0 \text{ vs } K: m \neq m_0 \qquad \max(T(X), n - T(X)) > fr_{\alpha/2}(B(n, 0.5))$$

Exakter Binomialtest

Allgemeiner: Testen auf die Wahrscheinlichkeit für das Eintreten eines Ereignisses A.

Annahme: $X_1, \ldots, X_n$ u. i. v. mit $X_1 = \mathbf{1}_A \sim B(1, \pi)$

Teststatistik:

$$T(X) = \sum_{i=1}^{n} \mathbf{1}_A(X_i) \sim B(n, \pi)$$

Ablehnungsbereich:

$$H : \pi \leq \pi_0 \text{ vs } K : \pi > \pi_0 \qquad T(X) > fr_\alpha(B(n, \pi_0))$$

$$H : \pi = \pi_0 \text{ vs } K : \pi \neq \pi_0 \qquad \max(T(X), n - T(X)) > fr_{\alpha/2}(B(n, \pi_0))$$

Binomialtest in R

In R führt man mit folgendem Befehl einen Vorzeichentest aus:

```
binom.test(x, n, p=0.5, alternative=..).
```

- `x` Wert der Prüfgröße. Nicht die Stichprobe.
- `n`, `p` Parameter der $B(n, p)$-verteilung.
- `alternative=c("two.sided", "less", "greater")` Art der Alternative.

Binomialtest - Geburtenbeispiel

- 100 Geburten werden untersucht, von denen 54 Mädchen und 46 Jungen sind.
- Alternative $\hat{=}$ Geburt eines Mädchens ist wahrscheinlicher $\hat{=}$ $(0.5, 1]$
- `binom.test(54, n=100, p=0.5, alternative="greater")`

```
        Exact binomial test

data:  54 and 100
number of successes = 54, number of trials = 100, p-value = 0.2421
alternative hypothesis: true probability of success is greater than 0.5
95 percent confidence interval:
0.4529712 1.0000000
sample estimates:
probability of success
  0.54
```

- Der p-Wert ist sehr hoch $\rightsquigarrow$ Die Hypothese kann nicht verworfen werden.

Wilcoxon-Vorzeichen-Rangtest

Beim Vorzeichentest werden sehr wenig Informationen benutzt (Wert größer oder kleiner als m_0). Zugrunde liegende Verteilung symmetrisch $\rightsquigarrow$ Ziehe Ränge zur Auswärtung heran.

Annahme: $X_1, \ldots, X_n$ u. i. v. mit symmetrischer, (stetiger) Verteilung und m_0 ist Stichproben-Median.

Teststatistik:

$$W^+ = \sum_{i=1}^{n} \mathrm{rg}\,|X_i - m_0|\, \mathbf{1}_{(0,\infty)}(X_i - m_0)$$

Für große n ($n \geq 30$) gilt W^+ approximativ $\mathcal{N}\left(\frac{n(n+1)}{4}, \frac{n(n+1)(2n+1)}{24}\right)$.

Ablehnungsbereich:

$H: m \leq m_0$ vs $K: m > m_0$	$W^+ > fr_\alpha(w^+)$
$H: m = m_0$ vs $K: m \neq m_0$	$W^+ > fr_{\alpha/2}(w^+)$ oder $W^+ < fr_{1-\alpha/2}(w^+)$

Wilcoxon-Vorzeichen-Rangtest in R

In R führt man mit folgendem Befehl einen Wilcoxon-Vorzeichen-Rangtest aus:

```
wilcox.test(x, alternative=.., mu=0, exact=.., correct=..).
```

- `x` Stichprobe.
- `alternative=c("two.sided", "less", "greater")` Art der Alternative.
- `exact=c(TRUE, FALSE)` Exakte oder aprroximative Verteilung der Teststatistik.
- `correct=c(TRUE, FALSE)` Stetigkeitskorrektur

χ^2-Anpassungstest

Testen auf eine bestimmte Verteilung.

Testproblem: $H : P = P_0$ gegen $K : P \neq P_0$

Vorbereitung: Zerlegen der Stichprobe $(x_1, \ldots, x_n)$ in k disjunkte Klassen $K_1, \ldots, K_k$ und zählen der abs. Häufigkeiten.

Klasse	K_1	K_2	...	K_k	Σ
abs. Häufigkeiten	n_1	n_2	...	n_k	n

$$\pi_i = P_0(K_i) \quad \text{und} \quad e_i = n_i \pi_i$$

- Die Klassen müssen vorher durch Histogramme, ... ermittelt werden.
- Sie sollten nicht zu klein sein, d.h. $e_i \geq 1$ und ≥ 5 bei mind. bei 80%.

χ^2-Anpassungstest - Teststatistik und R-Befehl

Teststatistik:

$$T(X) = \sum_{i=1}^{k} \frac{(n_i - e_i)^2}{e_i} \overset{d}{\underset{n\to\infty}{\simeq}} \chi^2_{k-1-r}$$

mit r die Anzahl an Parametern, um die Verteilung P_0 vollständig zu charakterisieren.

Ablehnungsbereich: $T(X) \geq fr_\alpha(\chi^2_{k-1-r})$

R-Befehl:

```
chisq.test(x, p=rep(1/length(x),length(x)), rescale.p=F)
```

- `x` abs. Häufigkeiten der Klassen.
- `p` Vektor der Wahrscheinlichkeiten $(\pi_1, ..., \pi_k)$ der Klassen oder Vektor der erwarteten abs. Häufigkeiten $(e_1, ..., e_k)$.
- `rescale.p=c(TRUE, FALSE)` Macht aus `p` einen W'keitsvektor

Kolmogorov-Smirnov-Test

Testen der Verteilung über die Verteilungsfunktion.

Testprobleme:

- $H : F = F_0$ gegen $K : \exists t \in \mathbb{R} : F(t) \neq F(t)_0$
- $H : F \leq F_0$ gegen $K : \exists t \in \mathbb{R} : F(t) > F(t)_0$

Voraussetzungen:

- Das Merkmal muss metrisch skaliert sein.
- F_0 ist stetig.

Teststatistiken:

- $T(X) \ = \ \sup_{t\in\mathbb{R}} |F_n(t|x) - F_0(t)|$
- $T(X) \ = \ \sup_{t\in\mathbb{R}} F_n(t|x) - F_0(t)$

mit $F_n(t|x)$ empirische Verteilungsfunktion der Stichprobe $x = (x_1, \dots, x_n)$.

Ablehnungsbereich: Falls die Teststatistiken zu groß werden. Kritische Werte liegen tabellarisch vor. Für $n \geq 40$ ist krit. Wert $\approx (\ln(2/\alpha)/2n)^{1/2}$

Kolmogorov-Smirnov-Test in R

In R führt man mit folgendem Befehl einen Kolmogorov-Smirnov-Test aus:

```
ks.test(x, y, ..., alternative=..).
```

- `x` Stichprobe.
- `y` Verteilungsfunktion, z.B. `pnorm`
- `...` Parameter der Verteilung `y`, z.B. `mu, sd`
- `alternative=c("two.sided", "less", "greater")` Art der Alternative.

Shapiro-Wilk-Test

Testen auf Normalverteilung.

Annahmen: $X_1, \ldots, X_n$ u. i. v. mit stetiger Verteilungsfunktion F

Testproblem: H: F ist eine Normalverteilung vs K: F ist keine Normalverteilung

Teststatistik: $W = \frac{(\sum_{i=1}^{n} a_i X_{(i)})^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$ mit $\mathbf{a}^{\mathrm{t}} = (a_1, \ldots, a_n)$ durch $\mathbf{a}^{\mathrm{t}} = \frac{\mathbf{m}^{\mathrm{t}}\mathbf{V}^{-1}}{(\mathbf{m}^{\mathrm{t}}\mathbf{V}^{-2}\mathbf{m})^{1/2}}$, wobei $\mathbf{m}$ und $\mathbf{V}$ Erwartungswertvektor bzw. Kovarianzmatrix eines geordneten Vektors von n u. i. v. $\mathcal{N}(0, 1)$-Variablen.

Ablehnungsbereich: Kleines W

R-Befehl: `shapiro.test(x)`

Gliederung

Zwei-Stichprobenfälle

In diesem Fall wird ein Merkmal unter zwei Bedingungen untersucht oder man betrachtet zwei Merkmale, die am selben Merkmalsträger erhoben werden:

1. Zwei unabhängige Zufallsstichproben $(X_{1,1}, \ldots, X_{1,n_1}), (X_{2,1}, \ldots, X_{2,n_2})$, $n_1, n_2 \in \mathbb{N}$, wobei sich die Randbedingungen bei der Entnahme der Stichproben in genau einer Randbedingung unterscheiden.
2. Ein Merkmal unter zwei verschiedenen Bedingungen am selben Merkmalsträger: $(X_{1,1}, X_{1,2}), \ldots, (X_{n,1}, X_{n,2})$ (verbundene Stichproben, *matched pairs*).
3. Zwei Merkmale X und Y am selben Merkmalsträger (unter jeweils gleichen Bedingungen): $(X_1, Y_1), \ldots, (X_n, Y_n)$ (verbundene Stichproben).

Reduktion auf den Ein-Stichprobenfall

- Das Problem (2) der Messung eines Merkmals unter verschiedenen Bedingungen am selben Merkmalsträger wird im Falle intervallskalierter Merkmale häufig durch Differenzbildung auf das Ein-Stichprobenproblem zurückgeführt.
- Dies wird in `R` in den Befehlen `t.test` und `wilcox.test` über den Parameter `paired` (=`TRUE` / `FALSE`) gesteuert.

t-Test unabhängiger Stichproben

Vergleichen der Mittelwerten.

Annahme:

- $X_1, \ldots, X_n$ u. i. v. mit $X_1 \sim \mathcal{N}(\mu_x, \sigma_x^2)$
- $Y_1, \ldots, Y_m$ u. i. v. mit $Y_1 \sim \mathcal{N}(\mu_y, \sigma_y^2)$

bzw. beliebig verteilt mit ex. Varianz und großem n ($n \geq 30$)

Teststatistik:

$$T(X,Y) = \sqrt{\frac{nm}{n+m}} \frac{\overline{X} - \overline{Y}}{\frac{n-1}{n+m-2} S(X)^2 + \frac{m-1}{n+m-2} S(Y)^2}$$

Ablehnungsbereich:

$$H : \mu_x \leq \mu_y \text{ gegen } K : \mu_x > \mu_y \qquad T(X,Y) > fr_\alpha(t_{n+m-2})$$

$$H : \mu_x = \mu_y \text{ gegen } K : \mu_x \neq \mu_y \qquad |T(X,Y)| > fr_{\alpha/2}(t_{n+m-2})$$

t-Test in R

In R führt man mit folgendem Befehl einen t-Test aus:

```
t.test(x, y, alternative=.., mu=0, paired=T, var.equal=F)
```

- `x` Argumente der ersten Stichprobe
- `y` Argumente der zweiten Stichprobe
- `alternative=c("two.sided", "less", "greater")` Art der Alternative.
- `mu` gegen die zu testende Differenz der Erwartungswerte
- `paired=c(FALSE, TRUE)` verbundene Stichprobe (standard `FALSE`)
- `var.equal=c(FALSE, TRUE)` gleiche Varianz der Stichproben

Pearson-Korrelationstest

Testen einer verbundenen, normalverteilten Stichprobe auf (linearen) Zusammenhang

Annahme: $(X_1, Y_1), \ldots, (X_n, Y_n)$ u. i. v. mit $(X_1, Y_1) \sim \mathcal{N}_2(\mu, \Sigma)$

Testprobleme:

(a) H: X_1, Y_1 unkorreliert vs K: X_1, Y_1 korreliert

(b) H: X_1, Y_1 negativ-korreliert vs K: X_1, Y_1 nicht negativ-korreliert

Teststatistik:

$$T(X, Y) = \sqrt{n-2}\frac{r_{X,Y}}{\sqrt{1 - r_{X,Y}^2}} \sim t_{n-2}$$

Ablehnungsbereich:

(a) $|T(X, Y)| > fr_{\alpha/2}(t_{n-2})$

(b) $T(X, Y) > fr_{\alpha}(t_{n-2})$

Spearman-Rang-Korrelationstest

Fall (X_1, Y_1) nicht normalverteilt sind, muss man auf den Spearman-Rang-Korrelationstest zurückgreifen.

Annahmen: X_1 und Y_1 sind stetig verteilt.

Testprobleme:

(a) H: X_1, Y_1 kein monotoner Zshg. vs K: X_1, Y_1 monotoner Zshg.

(b) H: X_1, Y_1 gegensinniger monotoner Zshg. vs K: X_1, Y_1 nicht gegen. mon. Zshg.

Teststatistik:

$$T(X, Y) = \sqrt{n-2}\frac{r_{SP}}{\sqrt{1 - r_{SP}^2}} \underset{n \to \infty}{\overset{d}{\simeq}} t_{n-2}$$

Ablehnungsbereich: Wie beim Pearson-Korrelationstest.

Korrelationstest in R

In R führt man mit folgendem Befehl einen Korrelationstest aus:

```
cor.test(x, y, alternative=.., method=..)
```

- `x` Argumente der ersten Stichprobe
- `y` Argumente der zweiten Stichprobe
- `alternative=c("two.sided", "less", "greater")` Art der Alternative.
- `method=c("pearson", "spearman", "kendall")` Art des Korrelationstests

χ^2-Unabhängigkeitstest

Testen zweier Merkmale X und Y auf Unabhängigkeit.

Testproblem: $H : X, Y$ unabhängig gegen $K : X, Y$ abhängig

Vorbereitung: Zerlegen der Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$ in $k \cdot l$ disjunkte Klassen $A_i \times B_j$ und zählen der abs. Häufigkeiten.

$X \backslash Y$	B_1	B_2	$\cdots$	B_l	Σ
A_1	n_{11}	n_{12}	$\cdots$	n_{1l}	$n_{1\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_k	n_{k1}	n_{k2}	$\cdots$	n_{kl}	$n_{k\cdot}$
Σ	$n_{\cdot 1}$	$n_{\cdot 2}$	$\ldots$	$n_{\cdot l}$	n

$$\pi_{ij} = P(X \in A_i, Y \in B_j) \quad \text{und} \quad e_{ij} = n_{ij}\pi_{ij}$$

- Sollte gelten: $e_{ij} \geq 1$ und ≥ 5 bei mind. bei 80%.

χ^2-Unabhängigkeitstest - Teststatistik und R-Befehl

Teststatistik:

$$T(X,Y) = \sum_{i=1}^{k}\sum_{j=1}^{l}\frac{(n_{ij}-e_{ij})^2}{e_{ij}} \overset{d}{\simeq} \chi^2_{(k-1)(l-1)}$$

Ablehnungsbereich: $T(X,Y) \geq fr_\alpha(\chi^2_{(k-1)(l-1)})$

R-**Befehl:**

```
chisq.test(x, y)
```

- `x` Vector des ersten Merkmals oder Matrix einer Kontingenztafel
- `y` Vector des zweiten Merkmals, falls `x` auch Vektor

Im Falle zweier vektoren, wird die Kontingenztafel von R selber erstellt.

Kolmogorov-Smirnov-Test

Testen zweier Verteilungen über die Verteilungsfunktion F und G

Testprobleme:

- $H: F = G$ gegen $K: \exists t \in \mathbb{R}: F(t) \neq G(t)$
- $H: F \leq G$ gegen $K: \exists t \in \mathbb{R}: F(t) > G(t)$

Voraussetzungen:

- Das Merkmal muss metrisch skaliert sein.
- F und G stetige Verteilungsfunktionen.

Teststatistiken:

- $K(x, y) \;=\; \sup_{t \in \mathbb{R}} |F_n(t|x) - G_m(t|y)|$
- $K(x, y) \;=\; \sup_{t \in \mathbb{R}} F_n(t|x) - G_m(t|y)$

der Stichprobe $x = (x_1, \ldots, x_n)$ und $y = (y_1, \ldots, y_m)$.

Ablehnungsbereich: Falls die Teststatistiken zu groß werden.

Kolmogorov-Smirnov-Test in `R`

In `R` führt man mit folgendem Befehl einen Kolmogorov-Smirnov-Test aus:

```
ks.test(x, y, alternative=..).
```

- `x` erste Stichprobe.
- `y` zweite Stichprobe
- `alternative=c("two.sided", "less", "greater")` Art der Alternative.

Vielen Dank für Eure Aufmerksamkeit