

## Inhaltsverzeichnis

<b>1</b>	<b>Grundbegriffe</b>	<b>2</b>
1.1	Zufallsvariable	2
1.1.1	Standardabweichung/Varianz	2
1.1.2	Erwartungswert (Skript S. 14, S. 32)	2
1.1.3	Rechenregel (Skript S. 33)	2
1.1.4	Arithmetische Mittel	3
1.2	Mengenlehre (Skript S. 10)	3
1.3	Quantil/Median (Skript S. 28)	3
1.4	Normaler Wert	3
<b>2</b>	<b>Verteilungen</b>	<b>3</b>
2.1	Binominalverteilung ( $X \sim Bin(n, p)$ ) (diskret) (Skript S. 13f)	3
2.1.1	Erwartungswert	3
2.1.2	Varianz (Skript S. 16)	4
2.1.3	Binominalkoeffizienten („ $n$ Tief $k$ “) (FS: S. 28)	4
2.1.4	Vertrauensintervall (Skript S. 24)	4
2.2	Bernoulliverteilung ( $X \sim Bernoulli(p)$ ) (diskret) (Skript S. 16)	4
2.3	Poissonverteilung ( $X \sim Poisson(n, p)$ ) (diskret) (Skript S. 17)	4
2.3.1	Erwartungswert und Varianz	4
2.3.2	Vertrauensintervall	4
2.4	Stetige Zufallsvariablen (Skript S. 31)	4
2.4.1	Dichte (Skript S. 32)	5
2.5	Standardisieren (Skript S. 36f)	5
2.5.1	Kennzahlen von stetigen Verteilungen (Skript S. 32)	5
2.6	Uniforme Verteilung (stetig) (Skript S. 34)	5
2.7	Exponentialverteilung (Skript S. 34)	5
2.7.1	Erwartungswert und Varianz	5
2.8	Normalverteilung (stetig) (Skript S. 35)	6
2.8.1	Erwartungswert und Varianz	6
2.8.2	Standardisierung	6
2.8.3	Zentraler Grenzwertsatz (Skript S. 41)	6
2.9	t-Verteilung (Skript S. 43) (stetig)	6
<b>3</b>	<b>Statistischer Test</b>	<b>7</b>
3.1	0-/A-Hypothese (Skript S. 20f)	7
3.2	Fehler 1. Art (Skript S. 22)	7
3.3	Fehler 2. Art (Skript S. 22)	7
3.3.1	Macht	7
3.4	P-Wert (Skript S. 22f)	7
3.5	(Punkt-) Schätzung (Skript S. 41)	7
3.6	z-Test für $\mu$ (Skript S. 42)	7
3.6.1	Verwerfungsbereich	8
3.6.2	Vertrauensintervall (Skript S. 45)	8
3.7	t-Test für $\mu$ (Skript S. 43)	8
3.7.1	Verwerfungsbereich	8
3.7.2	Vertrauensintervall (Skript S. 45)	9
3.7.3	gepaart (Skript S. 48)	9
3.7.4	ungepaart (Skript S. 48, Serie 8 (1b))	9

3.7.5	Verwerfungsbereich . . . . .	9
3.8	Test für $\mu$ bei nicht-normalverteilten Daten (Skript S. 45) . . . . .	9
3.8.1	Vorzeichentest (Skript S. 46) . . . . .	9
3.8.2	Wilcoxon-Test (Skript S. 47) . . . . .	10
<b>4</b>	<b>Mehrere Variablen</b>	<b>10</b>
<b>5</b>	<b>R</b>	<b>10</b>
5.1	R-Output interpretieren (Vorlesungsnotizen 16.6.2005) . . . . .	10
5.1.1	Vertrauensintervall . . . . .	10
5.1.2	Residuenquadratsumme . . . . .	10
5.1.3	F-Test (F-statistic) (Skript S. 64) . . . . .	11
5.2	Grafisch . . . . .	11
	<b>Stichwortverzeichnis</b>	<b>12</b>

# 1 Grundbegriffe

## 1.1 Zufallsvariable

$$X(\omega) = x$$

Eine Zufallsvariable  $X$  ist eine Funktion welche einem Elementarereignis ( $\omega$ ) einen Zahlenwert  $x$  zuordnet.  $\omega$  ist zufällig, nicht die Funktion  $X$ .

### 1.1.1 Standardabweichung/Varianz

Die Varianz beschreibt die Variabilität der Verteilung, die Standardabweichung deren Streuung (Skript S. 14)

$$\underbrace{\sigma(X)}_{\text{Standardabweichung}} = \sqrt{\text{Var}(X)}$$

$$\sigma(X)^2 = \underbrace{\text{Var}(X)}_{\text{Varianz}}$$

$$\text{Var}(X) = \mathcal{E}((X - \mathcal{E}(X))^2)$$

### 1.1.2 Erwartungswert (Skript S. 14, S. 32)

$$\mathcal{E}[X] = \sum_{x \in W_x} xP(X = x) = \int_{-\infty}^{\infty} xf(x) dx \quad \text{mit } W_x = \text{Wertebereich von } X$$

### 1.1.3 Rechenregel (Skript S. 33)

$$\begin{aligned} \mathcal{E}[a + bX] &= a + b\mathcal{E}[X] \\ \text{Var}(X) &= \mathcal{E}[X^2] - (\mathcal{E}[X])^2 \quad \text{stupides Einsetzen} \\ \text{Var}(a + bX) &= b^2\text{Var}(X) \end{aligned}$$

### 1.1.4 Arithmetische Mittel

diskret:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

### 1.2 Mengenlehre (Skript S. 10)

Vereinigt:  $\cup$   
 Durchschnitt/Schnittmenge:  $\cap$

### 1.3 Quantil/Median (Skript S. 28)

empirische  $\alpha$ -Quantil ist der Wert, bei dem  $\alpha \cdot 100\%$  der Datenpunkte kleiner und  $(1 - \alpha) \cdot 100\%$  der Punkte grösser.

Werte ordnen:

$$x_1 \leq x_2 \leq \dots \leq x_n$$

$$\alpha\text{-Quantil} = \begin{cases} \frac{1}{2}(x_{\alpha n} + x_{\alpha n + 1}) & : \alpha n \text{ ganze Zahl} \\ x_k & : \text{sonst, mit } k = \text{nächst grössere ganze Zahl von } \alpha n \end{cases}$$

Median (=50%-Quantil) ist der Wert  $x_{n/2+1/2}$  ( $n$  ungerade) bzw.  $(x_{n/2} + x_{n/2+1})/2$  ( $n$  gerade).

### 1.4 Normaler Wert

Definition: Ein normaler Wert ist höchstens 1.5 mal die Quartilsdifferenz von einem der beiden Quartile entfernt.

## 2 Verteilungen

### 2.1 Binominalverteilung ( $X \sim \text{Bin}(n, p)$ ) (diskret) (Skript S. 13f)

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

Mit  $Y = 2$  Werte  $(a, b)$   $\mathcal{E}[Y] = aP(x = a) + bP(x = b)$

$n$  = Anzahl Ereignisse

$p$  = Wahrscheinlichkeit für das Eintreffen des Ereignisses

$W$  = 0,1

$$p_0[X = k] = \binom{n}{k} p^k (1 - p)^{n-k}$$

$$p_0[X \leq c] = \sum_{k=0}^c \binom{n}{k} p^k (1 - p)^{n-k} \leq \alpha \stackrel{\text{i. a.}}{=} 0.05$$

Wertebereich ist beschränkt:  $W = \{0, 1, 2, \dots, n\}$ , falls nicht beschränkt Poissonverteilung (siehe 2.3) nehmen.

#### 2.1.1 Erwartungswert

$$\mathcal{E}[X] = np$$

### 2.1.2 Varianz (Skript S. 16)

$$\text{Var}(X) = np(1-p) \quad \Rightarrow \quad \sigma(X) = \sqrt{np(1-p)}$$

### 2.1.3 Binominalkoeffizienten („n Tief k“) (FS: S. 28)

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

### 2.1.4 Vertrauensintervall (Skript S. 24)

$$I[x, n] \approx \frac{x}{n} \pm 1.96 \sqrt{\frac{x}{n} \left(1 - \frac{x}{n}\right) \frac{1}{n}} \quad \Rightarrow \quad \left[ \frac{x}{n} - \dots, \frac{x}{n} + \dots \right]$$

Approximation mit Normalverteilung (oder auch Poissonverteilung siehe Skript S. 17)

## 2.2 Bernoulliverteilung ( $X \sim \text{Bernoulli}(p)$ ) (diskret) (Skript S. 16)

$X \sim \text{Bernoulli}(p)$ , Spezialfall der Binominalverteilung (nur 1 Ereignis).  
 $\mathcal{E} = 0 \cdot P[X = 0] + 1 \cdot P[X = 1] = p$

## 2.3 Poissonverteilung ( $X \sim \text{Poisson}(n, p)$ ) (diskret) (Skript S. 17)

$$P[X = x] = e^{-\lambda} \frac{\lambda^x}{x!}$$

$\lambda$ : Wert der Bekannt ist.

Falls der Wertebereich nicht wie bei der Binomial( $n, p$ )-Verteilung (siehe 2.1) beschränkt ist, bietet sich für Zähldaten die Poissonverteilung an.

Plotten von  $P(x)$  ergibt die Wahrscheinlichkeitsfunktion (Glockenkurve),  $\sum_{x=0}^{\infty} P(x)$  (bzw. bis  $2np$  da  $P(> 2np) \approx 0$ ) die kumulative Verteilungsfunktion.

Mit  $\lambda = np$  (falls  $n$  gross) kann die Binominalverteilung durch die Poissonverteilung approximiert werden.

### 2.3.1 Erwartungswert und Varianz

$$\mathcal{E}[X] = \lambda \quad \text{Var}(X) = \lambda \quad \sigma(X) = \sqrt{\lambda}$$

### 2.3.2 Vertrauensintervall

Approximation für zweiseitiges Vertrauensintervall (Konfidenzintervall) zum Niveau  $1 - \alpha = 0.95$ :

$$I[x] \approx x \pm 1.96 \sqrt{x} \quad \text{mit } x \text{ Anzahl Ereignisse}$$

$$I_{\text{Einzelmessung}}[x] \approx I[x]/n \quad \text{Vertrauensintervall für eine Messung}$$

Z. B. werden in  $n = 5$  Messungen 1, 0, 2, 1, 3 Asbestfasern gefunden  $\Rightarrow$  7 Ergebnisse.

## 2.4 Stetige Zufallsvariablen (Skript S. 31)

Eine Zufallsvariable  $X$  heisst stetig, wenn deren Wertebereich  $W_X$  kontinuierlich ist. Wahrscheinlichkeitsverteilung kann nur noch für alle Intervalle angegeben werden, mit der kumulativen Verteilungsfunktion  $F(x) = P(X \leq x)$ :

$$P(X \in (a, b]) = P(a < X \leq b) = F(b) - F(a)$$

### 2.4.1 Dichte (Skript S. 32)

Die Idee einer „Punkt“-Wahrscheinlichkeit  $P(X = x)$  kann im Infinitesimalen auch für stetige Zufallsvariablen formuliert werden. Die (Wahrscheinlichkeits-)Dichte  $f(\cdot)$  ist die Ableitung der kumulativen Verteilungsfunktion  $F(\cdot)$  (Übung 5).

Es gilt:

1.  $f(x) \geq 0 \quad \forall x$  (da  $F(\cdot)$  monoton wachsend ist)
2.  $P(a < X \leq b) = \int_a^b f(x) dx$
3.  $\int_{-\infty}^{\infty} f(x) dx = 1$  (wegen 2.)

### 2.5 Standardisieren (Skript S. 36f)

Eine stetige Zufallsvariable  $X$  kann immer linear transformiert werden, so dass  $\mathcal{E} = 0$  und  $Var = 1$ :

$$g(x) = -\frac{\mathcal{E}(X)}{\sigma_X} + \frac{1}{\sigma_X}x \Rightarrow Z = g(X) = \frac{X - \mathcal{E}(X)}{\sigma_X}$$

Falls  $X \sim \mathcal{N}(\mu, \sigma)$  so ist  $Z \sim \mathcal{N}(0, 1)$  (gültig bei Normalverteilung, im Allgemeinen nicht).

### 2.5.1 Kennzahlen von stetigen Verteilungen (Skript S. 32)

$$\mathcal{X} = \int_{-\infty}^{\infty} x f(x) dx$$

$$Var(X) = \int_{-\infty}^{\infty} (x - \mathcal{E}(X))^2 f(x) dx = \mathcal{E}((X - \mathcal{E}(X))^2)$$

### 2.6 Uniforme Verteilung (stetig) (Skript S. 34)

Gleichmässige Verteilung, jeder Wert ist gleich wahrscheinlich. Da stetig Verteilung, ist die Wahrscheinlichkeit genau einen bestimmten Wert zu erhalten = 0.

### 2.7 Exponentialverteilung (Skript S. 34)

Abstände (meist Zeitdauer) zwischen Ereignissen. Beachte Anzahl Ausfälle ist in gegebenen (Zeit-) Abstand ist Poissonverteilt (2.3)

Dichte:

$$f(x) = \begin{cases} \lambda \exp -\lambda x & : \text{ falls } x \geq 0 \\ 0 & : \text{ sonst} \end{cases}$$

kumulative Verteilungsfunktion:

$$F(x) = \begin{cases} 1 - \exp -\lambda x & : \text{ falls } x \geq 0 \\ 0 & : \text{ sonst} \end{cases}$$

#### 2.7.1 Erwartungswert und Varianz

$$\mathcal{E} = \frac{1}{\lambda} \text{ bei } X \sim Exp(\lambda)$$

$$Var(X) = \frac{1}{\lambda^2} \quad \sigma_X = \sqrt{Var(X)}$$

## 2.8 Normalverteilung (stetig) (Skript S. 35)

Normal-Verteilung ist die häufigste Verteilung für Messwerte.

### 2.8.1 Erwartungswert und Varianz

Für  $\mathcal{N}(\mu, \sigma_X^2)$ :

$$\begin{aligned}\mathcal{E}[X] &= \mu \\ \text{Var}(X) &= \sigma^2 \quad \sigma_X = \sqrt{\text{Var}(X)}\end{aligned}$$

### 2.8.2 Standardisierung

Falls  $X \sim \mathcal{N}(\mu, \sigma_X^2)$ , so ist die standardisierte Zufallsvariable

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

Berechnen von Wahrscheinlichkeiten bei  $\mathcal{N}(\mu, \sigma_X^2)$

$X \sim \mathcal{N}(\mu, \sigma_X^2)$  und gesucht ist  $P(X \leq c)$

$$P(X \leq c) = P\left(\frac{X - \mu}{\sigma} \leq \frac{c - \mu}{\sigma}\right) = P\left(Z \leq \Phi\left(\frac{c - \mu}{\sigma}\right)\right)$$

Quantil

$$\alpha\text{-Quantil} = (\Phi^{-1}(\alpha) + \mu) \frac{1}{\sigma}$$

ACHTUNG Widerspruch zu Beispiel auf Seite 37 im Skript.

### 2.8.3 Zentraler Grenzwertsatz (Skript S. 41)

$$\bar{X}_n \sim \mathcal{N}(\mu, \sigma_X^2/n) \text{ falls } X_1, \dots, X_n \text{ i.i.d. } \sim \mathcal{N}(\mu, \sigma_X^2)$$

$$\bar{X}_n \approx \mathcal{N}\left(\mu, \frac{\sigma_X^2}{n}\right)$$

Approximation wird mit grösserem  $n$  besser, zudem ist sie auch besser je näher sie bei  $\mathcal{N}(\mu, \sigma_X^2)$  liegt, ebenso gilt für die standardisierte Zufallsvariable:

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma_X} \approx \mathcal{N}(0, 1)$$

## 2.9 t-Verteilung (Skript S. 43) (stetig)

Langschwänziger als Normalverteilung.

Für  $T \sim t_\nu$  ( $\nu = \text{degrees of freedom}$ ) gilt:

$$\mathcal{E}(T) = 0$$

$$\text{Var}(T) = \frac{\nu}{\nu - 2}$$

### 3 Statistischer Test

#### 3.1 0-/A-Hypothese (Skript S. 20f)

$$H_0: \mu = \mu_0$$

Im Allgemeinen möchte man zeigen, dass man  $H_0$  verwerfen kann (gewünschtes Ergebnis), kann  $H_0$  nicht verworfen werden, kann nichts gezeigt/statistisch belegt werden.

$$H_A: \mu \neq \mu_0 \quad (\text{zweiseitig})$$

$$H_A: \mu > \mu_0 \quad (\text{einseitig nach oben})$$

$$H_A: \mu < \mu_0 \quad (\text{einseitig nach unten})$$

$\mu_0$  = Wert der Bekannt ist z. B. Herstellerangabe, Wert der Erreicht werden soll,...

$\mu$  = Effektiver Wert, kennt nur das Orakel

#### 3.2 Fehler 1. Art (Skript S. 22)

Fälschliches Verwerfen von  $H_0$ , obwohl  $H_0$  richtig ist. Wird direkt mittels Signifikanzniveau  $\alpha$  kontrolliert.

#### 3.3 Fehler 2. Art (Skript S. 22)

Fälschliches Beibehalten von  $H_0$ , obwohl  $H_A$  richtig wäre. Die Wahrscheinlichkeit für einen Fehler 2. Art wird grösser falls  $\alpha$  kleiner gewählt wird.

##### 3.3.1 Macht

1 – die Wahrscheinlichkeit für einen Fehler 2. Art.

#### 3.4 P-Wert (Skript S. 22f)

Der P-Wert ist definiert als das kleinste Signifikanzniveau, bei dem  $H_0$  (gerade noch) verworfen wird.

#### 3.5 (Punkt-) Schätzung (Skript S. 41)

Die (Punkt-) Schätzung für den Erwartungswert:

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

und die Varianz (empirische Standardabweichung  $\hat{\sigma}_X^2$ ):

$$\sigma_X^2 = \text{Var}(X) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

#### 3.6 z-Test für $\mu$ (Skript S. 42)

Modell:  $X_1, \dots, X_n$  i.i.d. (unabhängig, identisch verteilt) mit Verteilung  $\mathcal{N}(\mu, \sigma_X^2)$ .

Daten:  $x_1, \dots, x_n$  Realisierungen von  $X_1, \dots, X_n$

**Beachte meist nur eine Realisierung (1 Versuch)  $\Rightarrow n = 1$**

Varianz:  $\sigma_X^2$  bekannt.

$$H_0: \mu = \mu_0$$

$$H_A: \mu \neq \mu_0 \quad (\text{oder } < \text{ bzw. } >)$$

Teststatistik:

$$z = \frac{\sqrt{n}(\bar{x}_n - \mu_0)}{\sigma_X}$$

### 3.6.1 Verwerfungsbereich

$$K = (-\infty, \mu_0 - \Phi^{-1}(1 - \alpha/2)\sigma_X/\sqrt{n}] \cup [\mu_0 + \Phi^{-1}(1 - \alpha/2)\sigma_X/\sqrt{n}, \infty)$$

Verwerfe  $H_0$ , falls arithmetisches Mittel der Daten  $\bar{x}_n \in K$

Zusammenfassend: verwerfe  $H_0$  falls:

$$\begin{aligned} \left| \frac{\sqrt{n}(\bar{x}_n - \mu_0)}{\sigma_X} \right| &> \Phi^{-1}(1 - \alpha/2) && \text{bei } H_A : \mu \neq \mu_0 \\ \frac{\sqrt{n}(\bar{x}_n - \mu_0)}{\sigma_X} &< -\Phi^{-1}(1 - \alpha) && \text{bei } H_A : \mu < \mu_0 \\ \frac{\sqrt{n}(\bar{x}_n - \mu_0)}{\sigma_X} &> \Phi^{-1}(1 - \alpha) && \text{bei } H_A : \mu > \mu_0 \end{aligned}$$

### 3.6.2 Vertrauensintervall (Skript S. 45)

Diejenigen Werte  $\mu$  bei denen der Test nicht verwirft:

$$\bar{x}_n \pm \Phi^{-1}(1 - \alpha/2) \frac{\sigma_x}{\sqrt{n}} \quad \text{falls } \sigma_x \text{ bekannt, sonst t-Test}$$

### 3.7 t-Test für $\mu$ (Skript S. 43)

Modell:  $X_1, \dots, X_n$  i.i.d. (unabhängig, identisch verteilt) mit Verteilung  $\mathcal{N}(\mu, \sigma_X^2)$ .

Daten:  $x_1, \dots, x_n$  Realisierungen von  $X_1, \dots, X_n$

Varianz (Schätzung):

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

$$H_0 : \mu = \mu_0$$

$$H_A : \mu \neq \mu_0 \text{ (oder } < \text{ bzw. } > \text{)}$$

Teststatistik:

$$t = \frac{\sqrt{n}(\bar{x}_n - \mu_0)}{\hat{\sigma}_X}$$

und deren Verteilung unter der Nullhypothese  $H_0$  ist:  $T = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\hat{\sigma}_X} \sim t_{n-1}$ .

### 3.7.1 Verwerfungsbereich

$$K = (-\infty, \mu_0 - t_{n-1, 1-\alpha/2} \hat{\sigma}_X / \sqrt{n}] \cup [\mu_0 + t_{n-1, 1-\alpha/2} \hat{\sigma}_X / \sqrt{n}, \infty)$$

Verwerfe  $H_0$ , falls  $t \in K$

Zusammenfassend: verwerfe  $H_0$  falls:

$$\begin{aligned} |t| = \left| \frac{\sqrt{n}(\bar{x}_n - \mu_0)}{\hat{\sigma}_X} \right| &> t_{n-1, 1-\alpha/2} && \text{bei } H_A : \mu \neq \mu_0 \\ t = \frac{\sqrt{n}(\bar{x}_n - \mu_0)}{\hat{\sigma}_X} &< -t_{n-1, 1-\alpha} && \text{bei } H_A : \mu < \mu_0 \\ t = \frac{\sqrt{n}(\bar{x}_n - \mu_0)}{\hat{\sigma}_X} &> t_{n-1, 1-\alpha} && \text{bei } H_A : \mu > \mu_0 \end{aligned}$$

### 3.7.2 Vertrauensintervall (Skript S. 45)

$$\bar{x}_n \pm t_{n-1, 1-\alpha/2} \frac{\hat{\sigma}_x}{\sqrt{n}} \quad \text{falls } \sigma_x \text{ unbekannt, sonst z-Test}$$

### 3.7.3 gepaart (Skript S. 48)

Bei der Analyse von gepaarten Vergleichen arbeitet man mit den Differenzen innerhalb der Paare:

$$u_i = x_i - y_i \quad (i = 1, \dots, n)$$

Realisierungen von i.i.d. Zufallsvariablen  $U_1, \dots, U_n$ . Kein Unterschied:  $\mathcal{E}[U_i] = 0$ .

### 3.7.4 ungepaart (Skript S. 48, Serie 8 (1b))

$$H_0 : \mu_X = \mu_Y$$

$$H_A : \mu_X \neq \mu_Y \text{ (oder } < \text{ bzw. } > \text{)}$$

Teststatistik:

$$T = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right) \frac{1}{n+m-2} \left((n-1)s_x^2 + (m-1)s_y^2\right)}} = \frac{\bar{X}_n - \bar{Y}_m}{S_{pool} \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

bzw. falls  $m = n$

$$T = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{\frac{1}{n}(s_x^2 + s_y^2)}}$$

$$\text{mit } S_{pool} = \frac{1}{n+m-2} \left(\sum_{i=1}^n (X_i - \bar{X}_n)^2 + \sum_{i=1}^m (Y_i - \bar{Y}_m)^2\right) \text{ bzw. } s_{pool} = \sqrt{\frac{\text{Var}(X) + \text{Var}(Y)}{2}}$$

### 3.7.5 Verwerfungsbereich

Zusammenfassend: verwerfe  $H_0$  falls:

$$\begin{aligned} |T| &= \frac{|\bar{X}_n - \bar{Y}_m|}{S_{pool} \sqrt{\frac{1}{n} + \frac{1}{m}}} > t_{n+m-2, 1-\alpha/2} && \text{bei } H_A : \mu_X \neq \mu_Y \\ T &= \frac{\bar{X}_n - \bar{Y}_m}{S_{pool} \sqrt{\frac{1}{n} + \frac{1}{m}}} < -t_{n+m-2, 1-\alpha} && \text{bei } H_A : \mu_X < \mu_Y \\ T &= \frac{\bar{X}_n - \bar{Y}_m}{S_{pool} \sqrt{\frac{1}{n} + \frac{1}{m}}} > t_{n+m-2, 1-\alpha} && \text{bei } H_A : \mu_X > \mu_Y \end{aligned}$$

## 3.8 Test für $\mu$ bei nicht-normalverteilten Daten (Skript S. 45)

### 3.8.1 Vorzeichentest (Skript S. 46)

Modell:  $(X_i)_{1 \leq i \leq n}$  i.i.d. (unabhängig, identisch verteilt) mit Verteilung  $X_i \sim F$  (F irgendeine Verteilung welche symmetrisch ist) und Median  $\tilde{m}$ .

$H_0 : \tilde{m} = m_0 = \text{bekannter Wert (z. B. Gerät misst richtig)}$

$H_A : \tilde{m} \neq m_0$  (z. B. Gerät misst falsch)

Teststatistik K: Wieviele Werte liegen unter dem Median:

$$K := \#\{i | X_i < m_0\}$$

Unter  $H_0$  ist  $K \sim \text{Bin}(n, \frac{1}{2})$ , wodurch mit der Fausregel:

$$I \approx \left[ \frac{n}{2} - \sqrt{n}, \frac{n}{2} + \sqrt{n} \right]$$

das 5%-Vertrauensintervall folgt und draus wiederum der Verwerfungsbereich:  $K = [0, n] \setminus I$ .  
Ohne Fausregel, die kumulative Verteilfunktion bis  $\leq \alpha$  berechnen:

$$P[V = 0] + P[V = 1] + \dots + P[V = k] \leq \alpha \text{ mit } P[V = k] = \binom{n}{k} \left(\frac{1}{2}\right)^n$$

ergibt den Verwerfungsbereich:  $K = [0, c] \cup [n - k, n]$ .

### 3.8.2 Wilcoxon-Test (Skript S. 47)

Kompromiss zwischen t-Test und Vorzeichentest. Setzt keine Normalverteilung voraus (wie t-Test) nützt die Daten jedoch besser aus wie ein Vorzeichentest (hat besser Macht als t-Test und Vorzeichentest).

## 4 Mehrere Variablen

Zur numerische Zusammenfassung der Abhängigkeit, der Daten von der Form:  $(x_1, y_1), \dots, (x_n, y_n)$ , ist die empirische Korrelation  $r$  am gebräuchlichsten (nie  $r$  berechnen ohne ein Streudiagramm zu plotten).  $r$  ist eine dimensionslose Zahl zwischen -1, 1 (Vorzeichen: Richtung, Betrag: Stärke des Zusammenhangs (0 = keiner)).

$$r = \frac{s_{xy}}{s_x s_y}, \quad s_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

## 5 R

### 5.1 R-Output interpretieren (Vorlesungsnotizen 16.6.2005)

Regressionsmodell:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_n x_{ji} + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Coefficients:

	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	$\hat{\beta}_0$	$\widehat{s.e.}(\hat{\beta}_0)$	$\hat{\beta}_0 / \widehat{s.e.}(\hat{\beta}_0)$	P-Wert für $H_0 : \beta_0 = 0; H_A : \beta_0 \neq 0$
x	$\hat{\beta}_1$	$\widehat{s.e.}(\hat{\beta}_1)$	$\hat{\beta}_1 / \widehat{s.e.}(\hat{\beta}_1)$	P-Wert für $H_0 : \beta_1 = 0; H_A : \beta_1 \neq 0$

Residual standard error:  $\hat{\sigma}$  on  $n - 2$  (bzw.  $n - (j + 1)$ :  $(j + 1) = \text{Anzahl Variablen}$ , hier  $\beta_0, \beta_1$ ) degrees of freedom (df)

Multiple R-squared: ... (je besser, je näher bei 1), Adjusted R-squared: ...

#### 5.1.1 Vertrauensintervall

$$\beta_j \pm t_{df, 1-\alpha} \cdot \text{Std. Error}$$

#### 5.1.2 Residuenquadratsumme

$$\hat{\sigma}^2 \cdot \text{degrees of freedom}$$

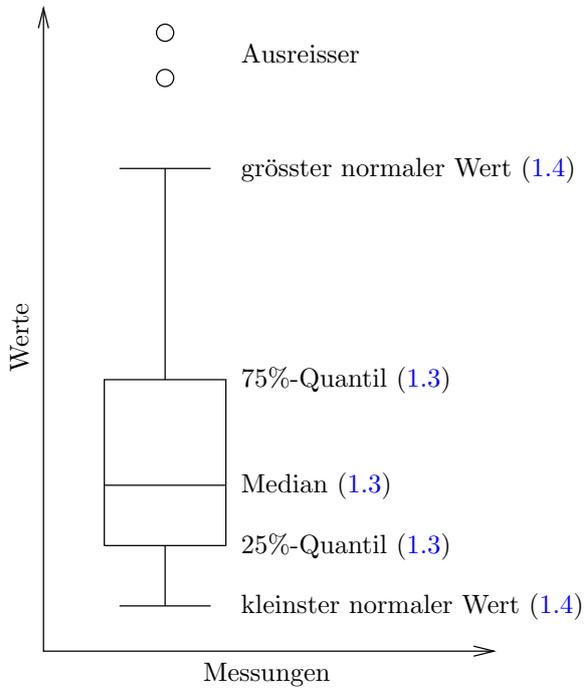
**5.1.3 F-Test (F-statistic) (Skript S. 64)**

P-Wert für

$$H_0 : \beta_0 = \beta_1 = \dots = \beta_{j+1} = 0$$

$H_A$  : mindestens ein  $\beta_i \neq 0$  mit  $i = \{0, \dots, j\}$

**5.2 Grafisch**



## Stichwortverzeichnis

0-Hypothese, 7  
A-Hypothese, 7  
Arithmetische Mittel, 3  
Bernoulliverteilung, 4  
Binominalkoeffizienten (Tief), 4  
Binominalverteilung, 3  
Dichte, 5  
empirische Standardabweichung, 7  
Exponentialverteilung, 5  
Fehler 1. Art, 7  
Fehler 2. Art, 7  
gepaart, 9  
Median, 3  
Mehrere Variablen, 10  
Mengenlehre, 3  
Normaler Wert, 3  
Normalverteilung, 6  
P-Wert, 7  
Plot, 11  
Poissonverteilung, 4  
Punktschätzung, 7  
Quantil, 3  
R-Output, 10  
Rechenregel Erwartungswert, Varianz, 2  
Schätzung, 7  
Standardabweichung, 2  
Standardisierung, 6  
Statistischer Test, 7  
stetige Verteilung, 5  
Stetige Zufallsvariable, 4  
t-Test, 8  
t-Verteilung, 6  
ungepaart, 9  
Varianz, 2  
Vorzeichentest, 9  
Wilcoxon-Test, 10  
z-Test, 7  
Zufallsvariable, 2