

Wahrscheinlichkeitsrechnung

①

Modell: Grundraum Ω aller möglichen Elementarereignisse (Bsp: gewürfelte Augenzahl bei einem Würfel). Ereignisse sind Teilmengen von Ω (Bsp: $A = \{\text{Wurf mit gerader Augenzahl}\}$). Den Ereignissen werden Wahrscheinlichkeiten $P(A)$ zugeordnet, die folgende Axiome erfüllen:

- $P(A) \geq 0$ für alle Ereignisse A
- $P(\Omega) = 1$ ("sicheres Ereignis Ω hat W'keit 1")
- $P(A \cup B) = P(A) + P(B)$ für $A \cap B = \emptyset$ (A und B schließen sich gegenseitig aus)

Daraus ergeben sich Rechenregeln:

$$\begin{aligned} & \bullet P(A^c) = 1 - P(A) \quad \text{[Diagram: A in a box, box shaded]} \\ & \bullet P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \text{[Diagram: A and B overlapping, intersection shaded]} \\ & \bullet P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ & \quad + P(A \cap B \cap C) \quad \text{[Diagram: A, B, C overlapping, all intersections shaded]} \end{aligned}$$

Def (Unabhängigkeit): A und B heißen unabhängig, falls gilt $P(A \cap B) = P(A) \cdot P(B)$

Bem: Δ Unterschied Unabhängigkeit ($P(A \cap B) = P(A)P(B)$) von "gegenseitig ausschließen" ($A \cap B = \emptyset$). Bsp: A = gerade Augenzahl, B = ungerade Augenzahl.

$P(A \cap B) = P(\emptyset) = 0 \neq P(A)P(B)$, da A und B i.A. nicht unabhängig.

Def (Bedingte W'keit): $P(A|B) = \frac{P(A \cap B)}{P(B)}$. "P von A gegeben B",
"W'keit von A, gegeben dass B eintritt" ($P(B) > 0$)

Satz von der totalen W'keit: $P(A) = \sum P(A|B_i) \cdot P(B_i)$ falls $B_i \cap B_j = \emptyset$,

$\cup B_i = \Omega$ und $P(B_i) > 0$ für alle i, j ($\{B_i\}$ eine Partition von Ω)

Bayes Theorem: $P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{\sum P(A|B_j) \cdot P(B_j)}$ falls $\{B_j\}$ eine Partition wie oben.

Zufallsvariablen

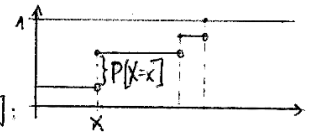
②

Def: Eine Abbildung $X: \Omega \rightarrow \mathbb{R}$ heisst Zufallsvariable

• X ist diskret, falls X nur endlich viele oder abzählbar unendlich viele Werte annimmt: $X(\Omega)$ ist diskret. Es gilt: $\sum_{x \in X(\Omega)} P(X=x) = 1$ für die W'keiten P

Verteilungsfunktion

$$F(b) := P(X \leq b) = \sum_{x \in X(\Omega), x \leq b} P(X=x)$$



F ist eine Treppenfunktion mit Sprüngen $P(X=x)$:

Erwartungswert

$$E[X] := \sum_{x \in X(\Omega)} x P(X=x) = \sum_{x \in X(\Omega)} x^2 P(X=x)$$

Varianz

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

Beispiele von diskreten Verteilungen:

- Gleichverteilung: X nimmt nur endlich viele Werte $x_1, \dots, x_n$ an mit $P(X=x_i) = \frac{1}{n}$ für alle $1 \leq i \leq n$.

- Bernoulli(p): X ist Ergebnis eines 0-1-Experiments, d.h. $X(\Omega) = \{0, 1\}$
 $P(X=1) = p, P(X=0) = 1-p, E[X] = p, \text{Var}(X) = p(1-p)$

- Binomial(n, p): X ist die Anzahl Erfolge bei n unabhängigen Bern(p)-Exp.
 $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, k=0, \dots, n, E[X] = np, \text{Var}(X) = np(1-p)$
 $\left\{ \binom{n}{k} = \frac{n!}{(n-k)! k!} \right\}$

• $F(k)$ für np und $n(1-p)$ grösser als 5 mit Normalverteilungsapproximation:

$$F(k) \approx \Phi\left(\frac{k - E[X]}{\sqrt{\text{Var}(X)}}\right)$$

• Für $n \rightarrow \infty, p \rightarrow 0$ und $np = \lambda$ erhält man die Poisson-Verteilung:

- Poisson(λ): Anz. seltener Ereignisse in einem Zeitintervall
 $P(X=k) = e^{-\lambda} \frac{\lambda^k}{k!}$ für $k=0, 1, \dots, E[X] = \lambda, \text{Var}(X) = \lambda$

• Zufallsvariable X ist stetig, falls es eine Funktion (die Dichte) $f: \mathbb{R} \rightarrow \mathbb{R}$ gibt mit $F(b) = \int_{-\infty}^b f(x) dx$. Es gilt $\int_{-\infty}^{\infty} f(x) dx = 1$ ③

Verteilungsfunktion $F(b) = P[X \leq b] = \int_{-\infty}^b f(x) dx$ und damit $P[a < X \leq b] = F(b) - F(a)$

Die Dichte f ist die Ableitung der Verteilungsfunktion F : $f(x) = F'(x)$

Erwartungswert $E[X] = \int_{-\infty}^{\infty} x f(x) dx$

Varianz $Var(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2$

Beispiele von stetigen Verteilungen:

- Uniform (a, b) : "X ist gleichmässig im Intervall $(a, b]$ verteilt"

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in (a, b] \\ 0, & \text{sonst} \end{cases}, F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$$

$$E[X] = \frac{a+b}{2}, Var(X) = \frac{(b-a)^2}{12}$$

- Exponential (α) : X ist die Zeit bis zum Eintreffen eines Ereignisses, z.B. die Zeit bis zum Zerfall eines bestimmten radioaktiven Teilchens

$$f(t) = \begin{cases} \alpha e^{-\alpha t}, & t \geq 0 \\ 0, & t < 0 \end{cases}, F(t) = 1 - e^{-\alpha t}$$

$$E[T] = \frac{1}{\alpha}, Var(T) = \frac{1}{\alpha^2}$$

Bem: Die Exponentialverteilung ist die einzige gedächtnislose Verteilung:

$$P[T > s+t | T > t] = P[T > s] \quad \forall s, t \geq 0$$

$$\text{Bsp: } P[T > 40 | T > 30] = P[T > 50 | T > 40] = P[T > 10]$$

"Egal, wie lange man schon wartet, die W'keit, weitere 10 min zu warten ist immer gleich gross!"

- Normal (μ, σ^2) : Häufigste Verteilung für Messwerte X

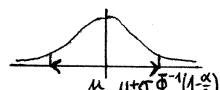
$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, E[X] = \mu, Var(X) = \sigma^2$$

Transformationsgesetz: $X \sim N(\mu, \sigma^2) \Rightarrow \frac{X-\mu}{\sigma} \sim N(0, 1)$

d.h. $F_{\mu, \sigma^2}(x) = F_{0,1}\left(\frac{x-\mu}{\sigma}\right) =: \Phi\left(\frac{x-\mu}{\sigma}\right)$, Φ die (tabellierte) Standardnormalvert.

Vertrauensintervalle: $X \sim N(\mu, \sigma^2)$, dann X mit 95% W'keit im Int. $[\mu \pm \sigma \Phi^{-1}(0.975)]$

$$\begin{array}{l} \frac{X}{\sigma} \sim N(0, 1) \\ \frac{X-\mu}{\sigma} \sim N(0, 1) \\ \frac{X-\mu}{\sigma/\sqrt{n}} \sim N(0, 1) \end{array} \quad \begin{array}{l} \text{---} \\ \text{---} \\ \text{---} \end{array} \quad \begin{array}{l} [\mu \pm \sigma \Phi^{-1}(0.975)] \\ [\mu \pm \frac{\sigma}{\sqrt{n}} \Phi^{-1}(0.975)] \\ [\mu \pm \frac{\sigma}{\sqrt{n}} \Phi^{-1}(0.975)] \end{array}$$



Zentraler Grenzwertsatz

$X_i, i=1, 2, \dots$ unabh. ZV mit $E[X_i] = \mu, Var(X_i) = \sigma^2$ (X_i nicht notwendigerw. $\sim N$) ④

$\Rightarrow$ für n "gross" gilt $S_n = \sum_{i=1}^n X_i$ ungefähr $N(n\mu, n\sigma^2)$

$\bar{X}_n = S_n/n$ ungefähr $N(\mu, \frac{\sigma^2}{n})$ verteilt

Momente: Def. und Rechenregeln

$E[X] = \sum x P[X=x]$ (X diskret), $E[X] = \int x f(x) dx$ (X stetig)

$E[aX+bY] = aE[X] + bE[Y]$ Linearität; gilt allg.

$Var(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2$, $\sigma(X) = \sqrt{Var(X)}$

$Var(aX+bY) = a^2 Var(X) + b^2 Var(Y)$ nur für X, Y unkorreliert

$Var(aX) = a^2 Var(X)$ (*)

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

$Cov(X, Y) = E[(X - E[X])(Y - E[Y])]$, $\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$

Anwendung von (*): $Var(\bar{X})$ und $\sigma(\bar{X})$ für X_i unabh mit gleicher Varianz:

$$Var(\bar{X}) = Var\left(\frac{1}{n} \sum X_i\right)$$

$$\stackrel{(*)}{=} \frac{1}{n^2} Var\left(\sum X_i\right)$$

$$= \frac{1}{n^2} \sum Var(X_i)$$

$$= \frac{1}{n^2} n \cdot Var(X_i)$$

$$= \frac{1}{n} Var(X_i)$$

$$\sigma(\bar{X}) = \sqrt{\frac{1}{n} Var(X_i)}$$

$$= \frac{1}{\sqrt{n}} \sigma(X_i)$$

Empirische Momente

$$E[X] \leftrightarrow \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, Var(X) \leftrightarrow s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Testen von Hypothesen

⑤

1. Nullhypothese H_0 aufstellen, z.B. $p = p_0 = \frac{1}{2}$ für $X \sim \text{Bin}(n, p)$
Alternativhypothese H_A aufst., z.B. $p \neq p_0$ (zweiseitig)
 $p \leq p_0$ (einseitig)
2. Wähle Prüfgrösse (sog. Teststatistik) T , die interpretierbar ist in dem Sinne, dass "T extrem (gross oder klein) $\Rightarrow H_0$ unwahrscheinlich"
 $T: \{\text{mögl. Beobachtungen}\} \rightarrow \mathbb{R}$ eine ZV mit bekannter Verteilung
3. Lege Niveau α fest. Unter H_0 soll gelten: T liegt mit W'keit α im Verwerfungsbereich $\rightarrow$ Grenzen für Verwerfungsbereich
4. Falls $T \in$ Verwerfungsbereich: verwirfe H_0 , sonst behalte H_0 bei

P-Wert

Def: Der P-Wert ist dasjenige Niveau, das gerade eben zur Verwerfung von H_0 führt

Bem: Die beobachtete Teststatistik liegt also gerade auf der Grenze des Verwerfungsbereichs bei einem Test zum Niveau $\alpha = \text{P-Wert} \rightarrow$ Interpretation des P-Werts:

$P[T \leq T(\text{Beob})]$ (einseitig), $P[|T| \geq |T(\text{Beob})|]$ (zweiseitig) unter H_0

"P-Wert = W'keit, dass die Teststat T unter H_0 mind. so extrem ist wie die Beob."

Entscheidungsregel: H_0 verwerfen, falls $\text{P-Wert} > \alpha$

Fehlerentscheidungen beim Testen

1. Art: H_0 stimmt, aber man verwirft
 $P[\text{Fehler 1. Art}] = \alpha$ (unter H_0)



2. Art: (Ein) H_A stimmt, aber man behält H_0 bei

$\beta = P[\text{Fehler 2. Art}]$ hängt von H_A ab



(β gross, falls sich H_0 und H_A "nur wenig unterscheiden"). $1 - \beta =$ Macht des Tests

Folgerung: H_0 kann nur verwerfen, nicht beweisen

t-Test oder Wilcoxon-Test?

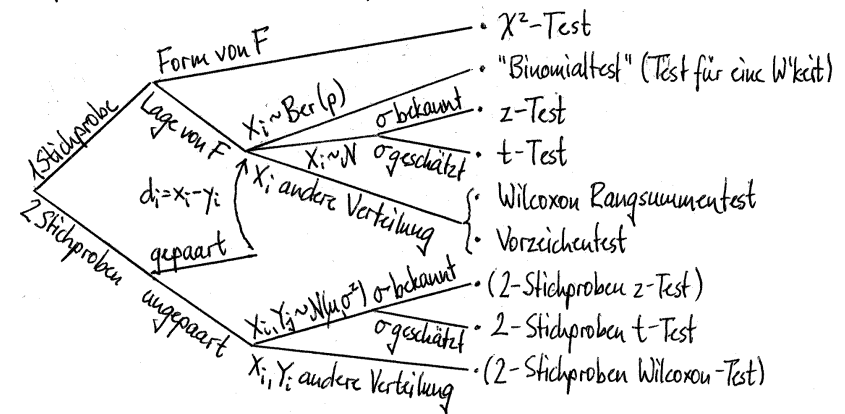
Frage: Sind die Modellvoraussetzungen ($X_i \sim N$) des t-Tests erfüllt?

- Niveau stimmt für Wilcoxon exakt, für t-Test genähert ohne Normalverteilung
- Macht des Wilcoxon-Tests wesentlich grösser ohne N als beim t-Test, mit N nur leicht kleiner

Testverfahren

⑥

Zufallsvariablen $X_1, \dots, X_n \sim F, Y_1, \dots, Y_m \sim G$



χ^2 -Test

Testet, ob Daten von einer bestimmten Verteilung (sfamilie) stammen. $T = \sum \frac{(\text{beobachtet} - \text{erwartet})^2}{\text{erwartet}}$

Vor: n Zufallsvar $X_1, \dots, X_n \sim F$, Fdiskret

Bsp1: Modell: $X_i \sim \text{Pois}(\lambda)$. Schätze λ durch ein $\hat{\lambda}$ und möchte wissen, ob $\text{Pois}(\hat{\lambda})$ eine plausible Verteilung für X_i . ($\hat{\lambda} = \bar{x}$)

Zähle mit N_k die Anz. Beob. x_i mit $x_i = k$ und vergleiche mit $E[N_k] = n P[X=k] = n e^{-\lambda} \frac{\lambda^k}{k!}$

Teststatistik $T = \sum_{k \geq 0} \frac{(N_k - E[N_k])^2}{E[N_k]} \sim \chi^2$ mit Anz. Summanden $- 2$ Freiheitsgraden

Bem: Damit T approximiert χ^2 -verteilt ist, müssen links und rechts Klassen N_k zusammengefasst werden: $N_R = \sum_{k \geq k_R} N_k$ und $N_L = \sum_{k \leq k_L} N_k$ s.d. $E[N_k] \geq 1$ für alle Klassen und $E[N_k] \geq 4$ für die Mehrzahl der Klassen

Bsp2: (Kontingenztafeln, Kreuztabellen)

2 Merkmale (Geschlecht, Schreibhand), 2 Ausprägungen. H_0 : Merkmale unabhängig

z.B. diese Daten:

	Männer	Frauen	total
rechth.	2780	3281	6061
linksh.	311	300	611
total	3091	3581	6672

unter H_0 (lasse Randtotalen fest) z.B. $\frac{3091}{6672}$

	2808	3253	6061
rechth.	283*	328	611
total	3091	3581	6672

Teststatistik $T = \frac{(2780-2808)^2}{2808} + \frac{(3281-3253)^2}{3253} + \frac{(311-283)^2}{283} + \frac{(300-328)^2}{328} \sim \chi^2$ mit 1 FG (restliche 3 Felder durch Spalte-/Zeilen totale bestimmt)

"Binomialtest", Test für eine W'keit

⑦

Modellannahmen: $X_1, \dots, X_n$ unabh. $\sim \text{Bern}(p)$ -verteilt. Testet, ob $p = p_0$

$H_0: p = p_0, H_A: p \neq p_0$. $T = \sum_{i=1}^n X_i \sim \text{Bin}(n, p)$ -verteilt unter H_0 .

Verwerfe H_0 , falls $T \notin [np_0 \pm \Phi^{-1}(0.975) \sqrt{np_0(1-p_0)}]$ (zweiseitig, genähert)
 $T > np_0 + \Phi^{-1}(0.95) \sqrt{np_0(1-p_0)}$ (einseitig rechts, genähert)

z-Test, t-Test

Modellannahmen: $X_1, \dots, X_n$ unabh. $\sim N(\mu, \sigma^2)$ -verteilt. Testet, ob $\mu = \mu_0$

$H_0: \mu = \mu_0, H_A: \mu \neq \mu_0$.

z-Test: $\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ ist unter H_0 $N(0,1)$ -verteilt. Verwerfe H_0 , falls $|T| > \Phi^{-1}(1-\frac{\alpha}{2})$ (zweiseitig)
 $|T| > \Phi^{-1}(1-\alpha)$ (einseitig)

t-Test: $\frac{\bar{X} - \mu_0}{s/\sqrt{n}}$ ist unter H_0 t_{n-1} -verteilt. Verwerfe H_0 , falls $|T| > t_{n-1, 1-\frac{\alpha}{2}}$ (zweiseitig)
 $|T| > t_{n-1, 1-\alpha}$ (einseitig)

Wilcoxon Rangsummentest

$X_1, \dots, X_n$ nicht notwendigerweise normalverteilt. $E[X_i] = \mu$. Testet, ob $\mu = \mu_0$

$H_0: \mu = \mu_0, H_A: \mu \neq \mu_0$. Teststatistik: Berechne $d_i = x_i - \mu_0$ und ordne betragsmässig.

Vertikale Ränge. $T = \sum_{d_i > 0} \text{Rang}(d_i)$. Tabelle für Verwerfungsbereich

Vorzeichenstest

$X_1, \dots, X_n$ nicht notwendigerweise normalverteilt. $\text{median}(X_i) = \tilde{\mu}$. Testet, ob $\tilde{\mu} = m_0$

$H_0: \tilde{\mu} = m_0, H_A: \tilde{\mu} \neq m_0$. $T = \text{Anz. } x_i \text{ mit } x_i - m_0 > 0 \sim \text{Bin}(n, \frac{1}{2})$ unter H_0

→ Test für W'keit mit $p = \frac{1}{2}$

2-Stichproben t-Test

Modellannahmen: $X_1, \dots, X_n \sim N(\mu_X, \sigma^2), Y_1, \dots, Y_m \sim N(\mu_Y, \sigma^2)$. Testet, ob $\mu_X = \mu_Y$

$H_0: \mu_X = \mu_Y, H_A: \mu_X \neq \mu_Y$

• gepaart: betrachte Differenzen $d_i = x_i - y_i$ und führe 1-Stichproben Test durch

$$T = \frac{\bar{d} - \mu_0}{s_{d-y}/\sqrt{n}} = \frac{\bar{x} - \bar{y}}{s_{x-y}/\sqrt{n}} \text{ unter } H_0 \text{ } t_{n-1} \text{-verteilt}$$

• ungepaart: $T = \frac{\bar{x} - \bar{y}}{s_{\text{pool}} \sqrt{\frac{1}{n} + \frac{1}{m}}}$, $s_{\text{pool}}^2 = \frac{1}{n+m-2} \left[\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{j=1}^m (y_j - \bar{y})^2 \right]$ unter H_0 t_{n+m-2} -verteilt

Lineare Regression

⑧

Modell: Zufallsvariablen $Y_i = \underbrace{\beta_0 + \beta_1 x_i^{(1)} + \dots + \beta_m x_i^{(m)}}_{\text{Modellwert}} + \underbrace{E_i}_{\text{zufälliger Fehler, } E_i \sim N(0, \sigma^2) \text{ unabh.}}$

Habe Daten $y_1, \dots, y_n$ (Realisierungen der Y_i).

Methode der kleinsten Quadrate: Schätze die Parameter $\beta_0, \dots, \beta_m$ durch $\hat{\beta}_0, \dots, \hat{\beta}_m$ s.d.

$\sum_{i=1}^n r_i^2$ minimal wird mit den Residuen $r_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i^{(1)} + \dots + \hat{\beta}_m x_i^{(m)})$

(Für $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, x^{(0)} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, x^{(j)} = \begin{pmatrix} x_1^{(j)} \\ \vdots \\ x_n^{(j)} \end{pmatrix}$ ist $\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_m \end{pmatrix}$ die Lösung des linearen Gleichungssystems

$$(X^T X) \hat{\beta} = X^T y \text{ wobei } X \text{ die } n \times (m+1)\text{-Matrix mit Spaltenvekt. } x^{(j)}$$

Bem.: $\hat{\beta}_0 + \hat{\beta}_1 x_i^{(1)} + \dots + \hat{\beta}_m x_i^{(m)}$ ist eine Schätzung für Y_i , der angepasste Wert

• Das Residuum r_i ist eine Schätzung für den Fehler E_i . Damit kann man σ schätzen:

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n r_i^2$$

• Y_i unabh. und normalverteilt, $\hat{\beta}$ eine lineare Fkt von $Y_1, \dots, Y_n \Rightarrow \hat{\beta}_j$ normalverteilt

mit Erwartungswert β_j und Standardabweichung $\sigma_{\hat{\beta}_j} = \sigma \sqrt{(X^T X)^{-1}}_{jj}$. Damit:

Test für $H_0: \beta_j = 0, H_A: \beta_j \neq 0$: $T = \frac{\hat{\beta}_j}{\text{se}(\hat{\beta}_j)}$ ist t-verteilt mit $n-m-1$ FG unter H_0

Verwerfen, falls $|T| > t_{n-m-1, 1-\frac{\alpha}{2}}$ (zweiseitig)

Vertrauensintervall für β_j : $\hat{\beta}_j \pm t_{n-m-1, 1-\frac{\alpha}{2}} \cdot \text{se}(\hat{\beta}_j)$

Bestimmtheitsmass R^2

$$\sum_{i=1}^n r_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \dots = \sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad R^2 = \frac{\text{"durch Regr."}}{\text{"durch Daten"}} \in [0, 1]$$

"durch Daten erkl." "durch Regression erkl."

Interpretation: $R^2 = 1 \Leftrightarrow$ Daten werden perfekt angepasst $\Leftrightarrow \sum_{i=1}^n r_i^2 = 0$ (falls R^2 definiert)

Bem.: In der einfachen linearen Regression gilt R^2 ist Schätzung für die Korrelation zw x, Y

Residuenanalyse

Modellannahmen: $E_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$, d.h. 1. $E[E_i] = 0$, 2. E_i unabh., 3. $\text{Var}(E_i) = \sigma^2 \forall i$, 4. $E_i \sim N$

Betrachte r_i , da r_i Schätzung von E_i

zu 1.: Tukey-Auscombe (Residuals vs. Fitted): Residuen sind um 0 gestreut

zu 2.: falls Daten in einer natürlichen Reihenfolge vorliegen: Residuen in dieser Reihenfolge keine Struktur

zu 3.: Tukey-Auscombe: keine Zunahme/Abnahme der Streuung

zu 4.: Normal-Plot: Residuen sollten auf einer Gerade liegen