Stochastik für Informatik 2023
Schätzer im linearen Model

Hanno Gottschalk

July 13, 2023

Inhaltsverzeichnis der Vorlesung

Transformationseigenschaften der Normalverteilung 3 / 34

Multivariate Standardnormalveteilung

Def: Die Multivatiate Standardnormalverteilung ist die Normalverteilung auf $\mathbb{R}^q$ mit $\Sigma = \mathbf{1}$ und $\mu = 0$, wobei $\mathbf{1}$ die Einheitsmatrix in $\mathbb{R}^q$ ist.

Für $\underline{x} = (x_1, \ldots, x_q)$ erhalten wir

$$f_{\underline{X}}(x) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_j^2} = \frac{1}{(2\pi)^{q/2}} \exp\left\{-\frac{1}{2}\|\underline{x}\|^2\right\}$$

Satz: Es sei $\underline{X} \sim N(0, \mathbf{1})$ und $\underline{\underline{U}}$ eine orthogonale $q \times q$-Matrix, $\underline{\underline{U}}\,\underline{\underline{U}}' = \mathbf{1}$ ($\underline{\underline{U}}' = \underline{\underline{U}}^{-1}$). Dann $\underline{\underline{U}}\,\underline{X} \sim N(0, \mathbf{1})$.

Beweis: Nach der Transformationsformel für Dichten:

$$f_{\underline{\underline{U}}\,\underline{X}}(\underline{x}) = \frac{1}{|\underline{\underline{U}}|} f_X(\underline{\underline{U}}'\underline{x}) = \frac{1}{(2\pi)^{q/2}} \exp\left\{-\frac{1}{2}\underbrace{\|\underline{\underline{U}}'\underline{x}\|^2}_{=\|x\|^2}\right\} \checkmark$$

Lineare Transformationen von Gauss-Zufallsvariablen

Satz: Es sei $\underline{X} \sim N(\mu, \sigma^2 \mathrm{Id})$ und $\underline{\underline{A}} : \mathbb{R}^d \to \mathbb{R}^q$ eine Matrix mit vollem Rang. Dann ist $\underline{Y} = \underline{\underline{A}}\,\underline{X} \sim N(\underline{\underline{A}}\,\underline{\mu}, \sigma^2 \underline{\underline{A}}\,\underline{\underline{A}}')$.

Beweis: OBdA: $\underline{\mu} = 0$. Setze $M = \underline{\underline{A}}'\mathbb{R}^q$ und wähle Orthonormalbasis so dass $\underline{u}_1, \ldots, \underline{u}_d$ so dass $\underline{u}_1, \ldots, \underline{u}_q$ eine ONB von M ist. $\underline{\underline{U}}$ zug. Matrix $\Rightarrow \underline{Y} = \underline{\underline{U}}\,\underline{X} \sim N(0, \sigma^2 \mathrm{Id})$ (Transformationssatz + Invarianz Gaußdichte)

Da $\underline{\underline{A}}\,\underline{u}_j = 0$ für $j = q+1, \ldots, n \Rightarrow$

$$\underline{\underline{A}}\,\underline{X} = \underline{\underline{A}}\,\underline{\underline{U}}'\underline{Y} = \underline{\underline{A}}(\underline{u}_1, \ldots, \underline{u}_q, 0, \ldots, 0)\underline{Y} = \underline{\underline{A}}(\underline{\underline{U}}_q, 0)\underline{Y}$$

Da $\underline{\underline{A}}\,\underline{\underline{U}}_q : \mathbb{R}^q \to \mathbb{R}^q$ bijektiv folgt nun die Aussage nach dem Transformationssatz und $\underline{\underline{U}}_q$ orthogonal + Invarianz $\Rightarrow$

$$\begin{aligned} \underline{\underline{A}}\,\underline{\underline{U}}_q(Y_1, \ldots, Y_q) &\sim N(0, \sigma^2 \underline{\underline{A}}\,\underline{\underline{U}}_q \mathrm{Id}_q \underline{\underline{U}}'_q \underline{\underline{A}}') \\ &= N(0, \sigma^2 \underline{\underline{A}}\,\underline{\underline{U}}\,\mathrm{Id}_d \underline{\underline{U}}'\underline{\underline{A}}') = N(0, \sigma^2 \underline{\underline{A}}\,\underline{\underline{A}}') \end{aligned}$$

Unkorreliert impliziert Unahängig

Satz: Zwei normalverteilte Zufallsvektoren $\underline{X}_1$, $\underline{X}_2$ seien unkorreliert $\Rightarrow \underline{X}_1, \underline{X}_2$ sind *unabhängig*.

Setze $\underline{X} = (\underline{X}_1, \underline{X}_2)$, dann

$$\Sigma_{\underline{X}} = \begin{pmatrix} \Sigma_{\underline{X}_1} & 0 \\ 0 & \Sigma_{\underline{X}_2} \end{pmatrix}$$

$$\begin{aligned} \left\langle (\underline{x} - \underline{\mu}), \Sigma_{\underline{X}}^{-1}(\underline{x} - \underline{\mu}) \right\rangle &= \left\langle (\underline{x}_1 - \underline{\mu}_1), \Sigma_{\underline{X}_1}^{-1}(\underline{x}_1 - \underline{\mu}_1) \right\rangle \\ &+ \left\langle (\underline{x}_2 - \underline{\mu}_2), \Sigma_{\underline{X}_2}^{-1}(\underline{x}_2 - \underline{\mu}_2) \right\rangle \end{aligned}$$

Unkorreliert impliziert Unahängig II

$$|\Sigma_{\underline{X}}| = \left|\Sigma_{\underline{X}_1}\right| \left|\Sigma_{\underline{X}_2}\right|$$

$$\begin{aligned} & \frac{1}{(2\pi)^{q/2}|\Sigma_{\underline{X}}|^{1/2}} \exp\left\{-\frac{1}{2}\langle \underline{x}-\underline{\mu}, \Sigma_{\underline{X}}^{-1}(\underline{x}-\underline{\mu}\rangle\right\} \\ = & \frac{1}{(2\pi)^{q_1/2}|\Sigma_{\underline{X}_1}|^{1/2}} \exp\left\{-\frac{1}{2}\langle \underline{x}_1-\underline{\mu}_1, \Sigma_{\underline{X}_1}^{-1}(\underline{x}_1-\underline{\mu}_1\rangle\right\} \\ \times & \frac{1}{(2\pi)^{q_2/2}|\Sigma_{\underline{X}_2}|^{1/2}} \exp\left\{-\frac{1}{2}\langle \underline{x}_2-\underline{\mu}_2, \Sigma_{\underline{X}_2}^{-1}(\underline{x}_2-\underline{\mu}_2\rangle\right\} \end{aligned}$$

$$f_{\underline{X}}(\underline{x}) = f_{\underline{X}_1}(\underline{x}_1) f_{\underline{X}_1}(\underline{x}_2)$$

Faktorisierung der Dichten ist äquivalent zur Unabhängigkeit. **qed**

Verteilung von $\underline{\hat{\beta}}$

Wiederholung lin. Modell

$$Y_j = \beta_0 + \underline{x}_j' \underline{\beta} + \epsilon_j, \quad \epsilon_j \sim N(0, \sigma^2) \text{ i.i.d.}$$

Vektorwertige Schreibweise:

$$\underline{Y} = \underline{\underline{M}}\,\underline{\beta} + \underline{\epsilon} \quad \underline{\underline{M}} = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,q-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,q-1} \end{pmatrix}, \quad \underline{\epsilon} \sim N(0, \sigma^2 \mathbf{1})$$

Kleinste-Quadrate Schätzer:

$$\underline{\hat{\beta}} = \left(\underline{\underline{M}}'\underline{\underline{M}}\right)^{-1} \underline{\underline{M}}'\underline{Y}$$

Frage: Da Verteilung von $\underline{Y} \sim N(\underline{\underline{M}}\,\underline{\beta}, \sigma^2 \mathbf{1})$ bekannt ist, kann man dann die Verteilung von $\underline{\hat{\beta}}$ explizit berechnen?

Erwartungstreue im linearen Modell

Satz: $\underline{\hat{\beta}}$ ist im linearen Modell *erwartungstreu*, also

$$\mathbb{E}_{(\underline{\beta},\sigma^2)}[\underline{\hat{\beta}}] = \underline{\beta} \tag{1}$$

(Erwartungswert eines Zufallsvektors wird komponentenweise genommen).

Beweis: Dank Linearität von $\mathbb{E}[.]$ und $\mathbb{E}_{(\underline{\beta},\sigma^2)}[\underline{Y}] = \underline{\underline{M}}\,\underline{\beta}$

$$\begin{aligned}\mathbb{E}_{(\underline{\beta},\sigma^2)}[\underline{\hat{\beta}}] &= (\underline{\underline{M}}'\,\underline{\underline{M}})^{-1}\underline{\underline{M}}'\,\mathbb{E}_{(\underline{\beta},\sigma^2)}[\underline{Y}] \\ &= (\underline{\underline{M}}'\,\underline{\underline{M}})^{-1}\underline{\underline{M}}'\underline{\underline{M}}\,\underline{\beta} = \underline{\beta}\end{aligned}$$

qed.

Kovarianz und Verteilung des Schätzers im Linearen Modell

Satz: Die *Kovarianz* von $\underline{\hat{\beta}}$ ist

$$\Sigma_{\underline{\hat{\beta}}} = \sigma^2(\underline{\underline{M}}'\,\underline{\underline{M}})^{-1} \tag{2}$$

Beweis: Nach dem Kov-Transformationsgesetz

$$\begin{aligned}\Sigma_{\underline{\hat{\beta}}} &= \left((\underline{\underline{M}}'\,\underline{\underline{M}})^{-1}\underline{\underline{M}}'\right)\Sigma_{\underline{Y}}\left((\underline{\underline{M}}'\,\underline{\underline{M}})^{-1}\underline{\underline{M}}'\right)' \\ &= \sigma^2\left((\underline{\underline{M}}'\,\underline{\underline{M}})^{-1}\underline{\underline{M}}'\right)\mathrm{Id}\left((\underline{\underline{M}}'\,\underline{\underline{M}})^{-1}\underline{\underline{M}}'\right)' \\ &= \sigma^2\left((\underline{\underline{M}}'\,\underline{\underline{M}})^{-1}\underline{\underline{M}}'\,\underline{\underline{M}}\,(\underline{\underline{M}}'\,\underline{\underline{M}})^{-1}\right) \\ &= \sigma^2(\underline{\underline{M}}'\,\underline{\underline{M}})^{-1}\end{aligned}$$

qed.

Satz (Folgerung) $\underline{\hat{\beta}} \sim N(\underline{\beta}, \sigma^2(\underline{\underline{M}}'\,\underline{\underline{M}})^{-1})$

Verteilung der Parameterschätzer

Haben in folgenden Satz abgeleitet:

$$\underline{\hat{\beta}} \sim N(\underline{\beta}, \sigma^2(\underline{\underline{M}}'\underline{\underline{M}})^{-1})$$

Wende dies an auf den Abnutzungsversuch

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	254.7500	6.1872	41.17	0.0000
run2	-2.2500	5.5340	-0.41	0.6984
run3	12.5000	5.5340	2.26	0.0647
run4	-9.2500	5.5340	-1.67	0.1457
position2	26.2500	5.5340	4.74	0.0032
position3	8.5000	5.5340	1.54	0.1755
position4	8.2500	5.5340	1.49	0.1866
materialB	-45.7500	5.5340	-8.27	0.0002
materialC	-24.0000	5.5340	-4.34	0.0049
materialD	-35.2500	5.5340	-6.37	0.0007

In diesem Falle ist der *Versuchsplan* (latin hypercube) so konstruiert, dass die Standardabweichung (hier: Std. Error) aller Effekte gleich ist.

$(\underline{\underline{M}}'\underline{\underline{M}})^{-1}$ und die Versuchsplanung

Wiederholung: Interpretierbarkeit der Effekte hängt eng mit *Orthogonalität* der Spaltenvektoren in $\underline{\underline{M}}$ zusammen

Beobachtung: Bis auf eine (unbekannte) Zahl σ^2, die auf alle Faktoren gleich wirkt, ist die Standardabweichung des Schätzers $\hat{\beta}_i$ allein durch die Modellmatrix bestimmt

$$\mathrm{Var}(\hat{\beta}_i) = \sigma^2(\underline{\underline{M}}'\underline{\underline{M}})^{-1}_{i,i}$$

Folgerung: Je nach Fragestellung kann der Versuchsplan schon im vorhinein optimiert werden, da $\underline{\underline{M}}$ nur vom Veruchsplan abhängt!!

Oft existiert sogar eine grobe Vorstellung von β_i und σ^2, so dass man oft *im Vorhinein* das zu erwartende *effect to error* $\hat{\beta}_i/\mathrm{sd}(\hat{\beta}_i)$ abschätzen kann

Grobe Orientierungshilfen zur Veruchsplanung

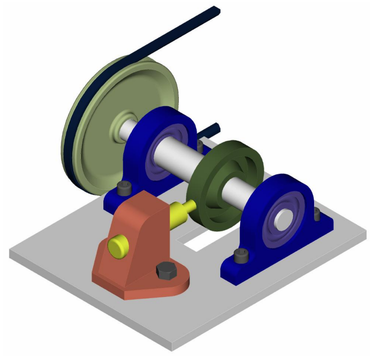
Bauernregeln für Nachweis eines Effektes

- Effect to Std.Error ratio $\gtrsim 3$
- Mehr Versuche $\Rightarrow$ weniger Std. Error ("noise")
- Störeffekte berücksichtigen (etwa Blockdesign) $\Rightarrow \sigma^2$ verringert, aber mehr Parameter
- Bei lin. Reg., möglichst große Parametervariation

Beispiel Blockdesign

Blockdesign für Abnutzungstest

- Zielgöße: Abnutzung
- Effekt: Materialeinfluss
- Störeffekt: Testmaschine
- Blockdesign: Jedes Material auf jeder Maschine



Std. Error für Materialeffekte: 10.84 ohne Berücksichtigung von 'position' und 8.67 mit! Das Verhältnis 1.25 kann one empirische Daten Vorhergesagt werden!

Effizienz von Schätzern

Def.: Zwei Schätzer S und T sollen denselben Parameter θ_i eines statistischen Modelles schätzen. S heißt *effizient* verglichen mit T, falls (für gegebenes n)

$$\mathrm{MSE}_\theta(S) \leq \mathrm{MSE}_\theta(T) \forall \theta \in \Theta \tag{3}$$

Def.: Ein erwartungstreuer Schätzer heißt bester Schätzer, wenn er gegenüber jedem anderen erwartungstreuen Schätzer effizient ist.
Ein Schätzer $\hat{\theta}$ heißt asymptotisch effizient, wenn für $n \to \infty$

$$\limsup \frac{MSE[\hat{\theta}_n](\theta)}{MSE[\hat{\theta}'_n](\theta)} \leq 1 \quad \forall \theta \in \Theta$$

für alle erwartungstreuen Schätzer $\hat{\theta}'$.

Wer ist BLUP?

Wir schränken die Vergleiche der Effizienz nun auf eine Unterklasse von Schätzern ein:

Def.: Ein Schätzer S für einen Parameter β_i in einem linearen Modell $\underline{Y} = \underline{\underline{M}}\,\underline{\beta} + \underline{\epsilon}$ heißt *linear*, falls $S = \langle b, Y \rangle$ für $b \in \mathbb{R}^n$.

Insbesondere ist $\hat{\beta}_i = \langle e_i, (\underline{\underline{M'}}\,\underline{\underline{M}})^{-1}\underline{\underline{M}}'\underline{Y} \rangle$ linear, $e_i = (0, 0, \ldots, 1, \ldots, 0)$ (i-te Stelle), $i = 1, \ldots, q$

Def.: Ein linearer, erwartungstreuer Schätzer S ist BLUP (Best Linear Unbiased Predictor), wenn S gegenüber jedem anderen linearen, erwartungstreuen Schätzer effizient ist.

Kleinste Quadrate Schätzer ist BLUP

Satz: (Markov)
$\underline{\hat{\beta}}$ ist der einzige BLUP im linearen Modell
D.h. $\forall \underline{c} \in \mathbb{R}^q$ ist $\langle \underline{c}, \underline{\hat{\beta}} \rangle$ BLUP für $\langle \underline{c}, \underline{\beta} \rangle$
Insbesondere ist $\hat{\beta}_i = \langle e_i, \underline{\hat{\beta}} \rangle$ BLUP für β_i

Beweis: Erwartungstreue ✓
zu zeigen: $\langle \underline{c}, \underline{\hat{\beta}} \rangle$ hat unter allen erwartungstreuen, linearen Schätzern $\langle \underline{b}, \underline{Y} \rangle$ für $\langle \underline{c}, \underline{\beta} \rangle$ die kleinste Varianz.

$\underline{a} := \underline{\underline{M}}\,(\underline{\underline{M}}'\underline{\underline{M}})^{-1}\underline{c} \Rightarrow \underline{\underline{M}}'\,\underline{a} = \underline{c}$.

$$\begin{aligned} \langle \underline{b}, \underline{\underline{M}}\,\underline{\beta} \rangle &= \mathbb{E}_{(\beta,\sigma^2)}[\langle \underline{b}, \underline{Y} \rangle] \quad (\langle \underline{b}, \underline{Y} \rangle \text{ erwartungstreu}) \\ &= \langle \underline{c}, \underline{\beta} \rangle = \langle \underline{a}, \underline{\underline{M}}\,\underline{\beta} \rangle \quad \forall \underline{\beta} \in \mathbb{R}^q \end{aligned}$$

$\Rightarrow \underline{b} - \underline{a} \in M^{\perp} \Rightarrow p_M \underline{b} = \underline{a} \Rightarrow |\underline{b}| \geq |\underline{a}|$ mit Gleichheit genau dann wenn $\underline{a} = \underline{b}$.

Beweis BLUP-Eigenschaft

Zeige nun $\langle \underline{a}, \underline{Y} \rangle$ ist BLUP:

$$\mathbb{E}_{(\underline{\beta},\sigma^2)}[\langle \underline{a}, \underline{Y} \rangle] = \langle \underline{a}, \underline{\underline{M}}\,\underline{\beta} \rangle = \langle \underline{\underline{M}}'\underline{a}, \underline{\beta} \rangle = \langle \underline{c}, \underline{\beta} \rangle$$

$$\begin{aligned} \mathrm{Var}_{(\underline{\beta},\sigma^2)}(\langle \underline{b}, \underline{Y} \rangle) - \mathrm{Var}_{(\underline{\beta},\sigma^2)}(\langle \underline{a}, \underline{Y} \rangle) &= \sigma^2(\underline{b}'\mathrm{Id}\underline{b} - \underline{a}'\mathrm{Id}\underline{a}) \\ &= \sigma^2(|b|^2 - |a|^2) \geq 0 \end{aligned}$$

Zeige $\langle \underline{a}, \underline{Y} \rangle = \langle \underline{c}, \underline{\hat{\beta}} \rangle$:

$$\begin{aligned} \langle \underline{a}, \underline{Y} \rangle &= \langle \underline{\underline{M}}\,(\underline{\underline{M}}'\underline{\underline{M}})^{-1}\underline{c}, \underline{Y} \rangle \\ &= \langle \underline{c}, (\underline{\underline{M}}'\underline{\underline{M}})^{-1}\underline{\underline{M}}'\underline{Y} \rangle = \langle \underline{c}, \underline{\hat{\beta}} \rangle \end{aligned}$$

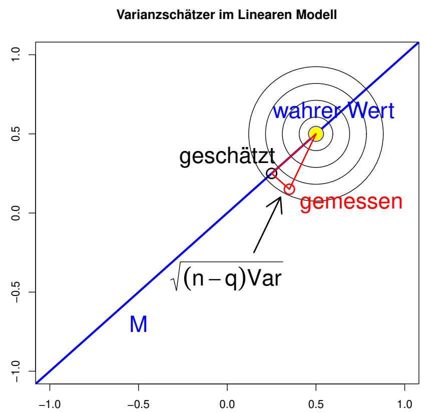
qed.

Schätzer der Residuenvarianz

Satz: Im Linearen Modell ist ein erwartungstreuer Schätzer der Varianz gegeben durch

$$\hat{\sigma}^2 = \frac{|\underline{Y} - p_M \underline{Y}|^2}{n-q} = \frac{|p_M^\perp \underline{Y}|^2}{n-q} \tag{4}$$

Motivation:



Beweis erwartungstreue Varianzschätzer

OBdA: $\underline{\beta} = 0$, sonst betrachte $\underline{Y} - \underline{\underline{M}}\,\underline{\beta}$.
Es sei $\underline{u}_1, \ldots, \underline{u}_n$ ONB so dass $\underline{u}_1, \ldots, \underline{u}_q$ Basis von M.

$$|p_M^\perp \underline{Y}|^2 = \sum_{j=q+1}^{n} |\langle \underline{u}_j, \underline{Y} \rangle|^2$$

Aufgrund der Invarianz der Verteilung von $\underline{Y}$ unter der orthogonalen Transformationen $\underline{\underline{u}}$ gilt

$$(\langle \underline{u}_{q+1}, \underline{Y} \rangle, \ldots, \langle \underline{u}_n, \underline{Y} \rangle)' \sim N(\underline{0}, \sigma^2 \mathrm{Id}_{(\mathrm{n-q})})$$

$\Rightarrow$

$$\mathbb{E}_{(\underline{0},\sigma^2)}\left[|\langle \underline{u}_j, \underline{Y} \rangle|^2\right] = \sigma^2, \quad j = q+1, \ldots, n$$

qed.

Darstellung des Projektors

Satz: Der Projektor p_M auf $M = \mathrm{Bild}(\underline{\underline{M}})$ hat folgende Darstellung

$$p_M = \underline{\underline{M}}\,(\underline{\underline{M}}'\underline{\underline{M}})^{-1}\underline{\underline{M}}' \tag{5}$$

Beweis: Zu zeigen $p_M\underline{x} \in M$ ✓, $p_M\underline{x} = \underline{x}$ für $\underline{x} \in M$, und $\underline{x} - p_M\underline{x} \in M^\perp$ mit p_M gleich rechte Seite von (1).

Beweis Projektordarstellung

zu zeigen: $p_M\underline{x} = \underline{x}$ für $\underline{x} \in M$:

$\underline{x} \in M \to \exists \underline{\gamma} \in \mathbb{R}^q, \underline{x} = \underline{\underline{M}}\,\underline{\gamma}$

$\Rightarrow$

$$\begin{aligned} p_M\underline{x} &= \underline{\underline{M}}\,(\underline{\underline{M}}'\,\underline{\underline{M}})^{-1}\underline{\underline{M}}'\,\underline{x} \\ &= \underline{\underline{M}}\,(\underline{\underline{M}}'\,\underline{\underline{M}})^{-1}\underline{\underline{M}}'\underline{\underline{M}}\,\underline{\gamma} = \underline{\underline{M}}\,\underline{\gamma} = \underline{x} \end{aligned}$$

zu zeigen: $\underline{x} - p_M\underline{x} \in M^\perp = \mathrm{Kern}(\underline{\underline{M}}')$

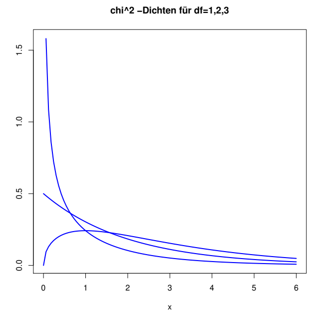
$$\begin{aligned} \underline{\underline{M}}'(\underline{x} - p_M\underline{x}) &= \underline{\underline{M}}'\underline{x} - \underline{\underline{M}}\,\underline{\underline{M}}'\,(\underline{\underline{M}}'\underline{\underline{M}})^{-1}\underline{\underline{M}}'\underline{x} \\ &= \underline{\underline{M}}'\underline{x} - \underline{\underline{M}}'\underline{x} = 0 \end{aligned}$$

qed.

χ^2-Verteilung

Def.: Es seien $X_1, \dots, X_n$ standardnormalverteilte Zufallsvariablen. Dann ist die $\chi^2(n)$-Verteilung – *χ^2-Verteilung mit n Freiheitsgraden* – definiert durch

$$\sum_{j=1}^{n} X_j^2 \sim \chi^2(n), \quad f_{\chi^2(n)}(x) = \frac{x^{n/2-1}}{\Gamma(n/2)\sqrt{2}^n} e^{-x/2}, \quad x > 0 \tag{6}$$



Verteilung des Varianzschätzers

Satz: (i) $\frac{(n-q)}{\sigma^2}\hat{\sigma}^2 \sim \chi^2(n-q)$

(ii) $\hat{\sigma}^2$ ist unabhängig von $\underline{\hat{\beta}}$

Beweis: (i) ✓
(ii)

$$\underline{\underline{M}}\,\underline{\hat{\beta}} = p_M \underline{Y} = p_M(\underline{\underline{M}}\,\underline{\beta} + \underline{\epsilon}) = \underline{\underline{M}}\,\underline{\beta} + p_M \underline{\epsilon}$$

Da $\underline{\underline{M}}$ auf M invertierbar $\Rightarrow$ $\underline{\hat{\beta}}$ hängt nur von $p_M\underline{\epsilon}$ ab.

$\hat{\sigma}^2$ hängt nur von $p_M^{\perp}\underline{\epsilon}$ ab

Da $p_M\underline{\epsilon}$ und $p_M^{\perp}\underline{\epsilon}$ normalvert. + unkorreliert $\Rightarrow$ unabh.

qed.

Explorative Validierung der Modellannahmen

Vorbemerkungen

Frage: Existieren lineare Modelle in der wirklichen Welt?

Wenn etwas zu *schön* ist um wahr zu sein... dann *ist* es nicht wahr!

Das lineare Modell *ist* zu schön um wahr zu sein...

Neue Frage: Kann mich jemand dran kriegen, wenn ich es trotzdem mache?

Mache die Tests, die andere machen würden, und wenn ich mich selbst nicht drankriege, dann können mich auch andere nicht drankriegen

Zeige: Mein Modell ist **'state of the art'** (aber nicht besser)!

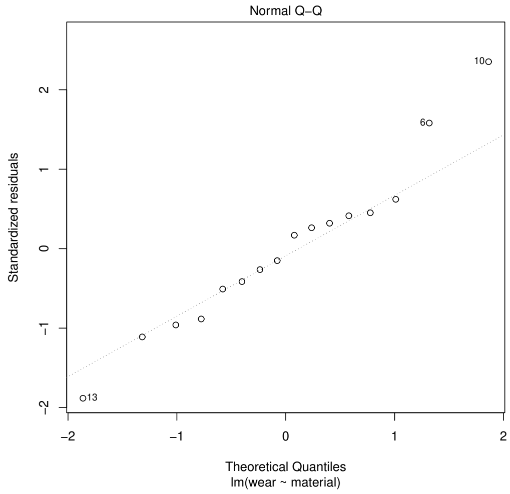
Explorative Validierung von Modellhypothesen im linearen Modell

Dem Statistischen Modell 'lineares Modell' liegen drei Hypothesen Zugrunde

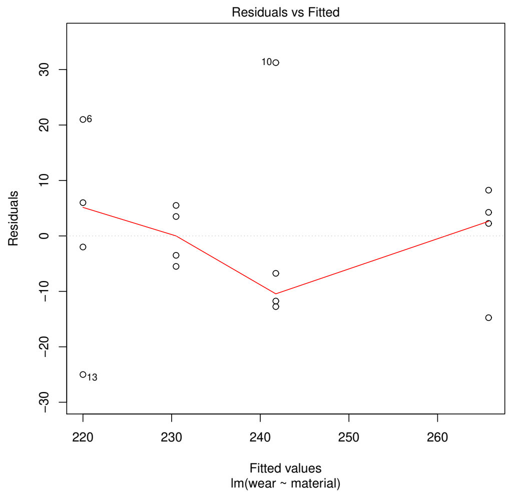
- Die Residuen ϵ_j sind *normalverteilt* $\longrightarrow$ Überprüfe mit QQ-plot (s.u.)
- Die Residuen haben alle dieselbe Varianz σ^2 $\longrightarrow$ Überprüfe mit 'Residuals over Fitted' plot
- Die Residuen sind unabhängig (unkorreliert) $\longrightarrow$ Überprüfe mit time-series-plot (s.u.)

Def.: QQ-Plot: *Falls* die Residuen einer Normalverteilung $N(0, \sigma^2)$ entstammen, dann sollte für die *empirischen Quantile* q_α der empirischen Residuen ϵ_j gelten $q_\alpha \approx \sigma z_\alpha$ mit z_α dem α-Quantil der Standardnormalverteilung.

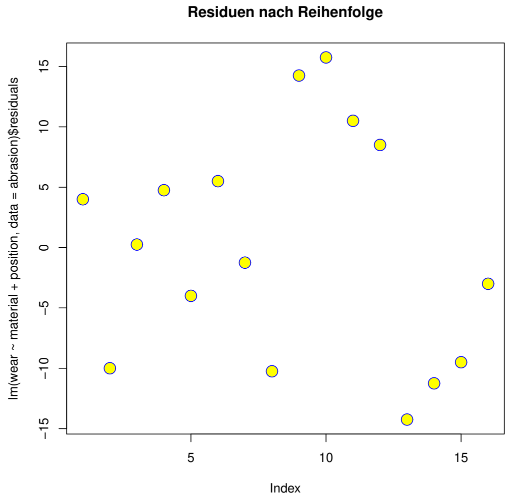
QQ-Plot



Residuals over Fitted Plot



TimeSeries Plot



Konfidenzbereiche für Parameter im linearen Modell 35 / 34

Verteilung der standardisierten Parameter

Haben gesehen

$$\underline{\hat{\beta}} \sim N\left(\underline{\beta}, \sigma^2(\underline{\underline{M}}'\underline{\underline{M}})^{-1}\right)$$

Außerdem

$$(n-q)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-q)$$

Satz: Es gilt

$$\frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}\sqrt{(\underline{\underline{M}}'\underline{\underline{M}})^{-1}_{i,i}}} \sim t(n-q) \qquad (7)$$

Denn:

$$\sqrt{n-q}\frac{\frac{\hat{\beta}_i - \beta_i}{\sigma\sqrt{(\underline{\underline{M}}'\underline{\underline{M}})^{-1}_{i,i}}}}{\sqrt{n-q}\frac{\hat{\sigma}}{\sigma}} \sim t(n-q)$$

Konfidenzintevalle für die Parameter im lin. Modell

Folgerung: Aus dem vorangehenden Satz ergeben sich folgende Konfidenzintervalle für die Parameter im lin. Modell:

Beidseitiges Konfidenzintervall für $\hat{\beta}_i$

$$\hat{\beta}_i \pm t_{1-\alpha/2}(n-q)\hat{\sigma}\sqrt{(\underline{\underline{M}}'\underline{\underline{M}})^{-1}_{i,i}} \tag{8}$$

Einseitiges Konfidenzintervall: Linksoffen

$$\left[-\infty, \hat{\beta}_i + t_{1-\alpha}(n-q)\hat{\sigma}\sqrt{(\underline{\underline{M}}'\underline{\underline{M}})^{-1}_{i,i}}\right) \tag{9}$$

Einseitiges Konfidenzintervall: rechtsoffen

$$\left[\hat{\beta}_i - t_{1-\alpha}(n-q)\hat{\sigma}\sqrt{(\underline{\underline{M}}'\underline{\underline{M}})^{-1}_{i,i}}, \infty\right) \tag{10}$$

Die ’summary’ Tabelle und Sternchen-Code im lin. Modell

In der Summary Tabelle finden wir Schätzwert, Wert der T-Statistik (p-Wert) und Sternchen-Code.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	255.0000	8.1129	31.43	0.0000	***
materialB	-45.7500	8.6731	-5.27	0.0005	***
materialC	-24.0000	8.6731	-2.77	0.0219	*
materialD	-35.2500	8.6731	-4.06	0.0028	**
position2	26.2500	8.6731	3.03	0.0143	*
position3	8.5000	8.6731	0.98	0.3527	
position4	8.2500	8.6731	0.95	0.3663	

- Kein Stern: 0 ist im 2 seitigen 90% Konfiintervall
- ’.’ : 0 ist nicht im 2 seitigen 90% Konfiintervall
- ’*’: 0 ist nicht im 2 seitigen 95% Konfiintervall
- ’**’: 0 ist nicht im 2 seitigen 99% Konfiintervall
- ’***’: 0 ist nicht im 2 seitigen 99.9% Konfiintervall

Interpretation des Sternchencodes

$\beta_i = 0 \Rightarrow$ den β_i-Effekt gibt es nicht!

Wenn 0 *nicht* im 2-seitigen Konfibereich zu $1-\alpha$ Konfidenz liegt, dann können wir uns zu $(1-\alpha) \times 100\%$ sicher sein, dass es β_i doch gibt!

Diese Art zu schließen wird später in der Testtheorie noch formalisiert. . .

Schätzung und Vorhersage - verschiedene Fragen 40 / 34

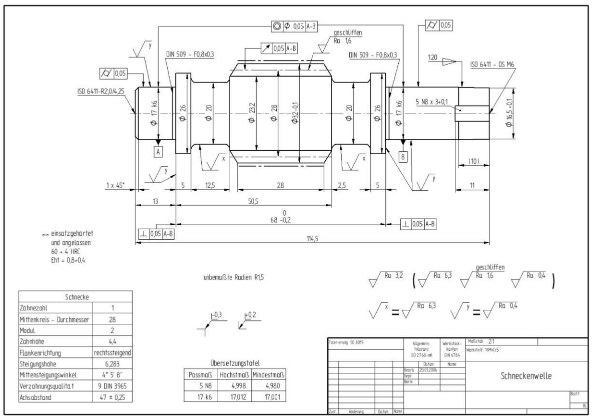
Fragen an lineare Modelle

Lineare Modelle können verschiedenen Zwecken dienen

- Herausfinden, welche Effekte wichtig sind (Screening)
- Quantifizierung der Effekte (Schätzen)
- Bei einem noch nicht beobachteten Fall mit bekannten Einflußgrößen $x_1^{\mathrm{neu}}, \dots, x_d^{\mathrm{neu}}$ die response y^{neu} *vorhersagen*
- Reststreuung quantifizieren (Streuung, die in dem Modell nicht reduziert werden kann)

Technische Zeichnungen

Die Kommunikation zwischen Designer/innen und Produzent/innen erfolgt über technische Zeichnungen.



Toleranzen nach ISO 2768

Toleranzen werden entweder explizit, oder durch Verweis auf ISO 2768 vorgegeben

Grenzmaße für Längenmaße entsprechend DIN ISO 2768-1				
	Toleranzklassen			
	f (fein)	m (mittel)	c (grob)	v (sehr grob)
Nennmaßbereich in mm	Toleranzen in mm			
0,5 bis 3	± 0,05	± 0,10	± 0,15	-
über 3 bis 6	± 0,05	± 0,10	± 0,20	± 0,50
über 6 bis 30	± 0,10	± 0,20	± 0,50	± 1,00
über 30 bis 120	± 0,15	± 0,30	± 0,80	± 1,50
über 120 bis 400	± 0,20	± 0,50	± 1,20	± 2,50
über 400 bis 1000	± 0,30	± 0,80	± 2,00	± 4,00
über 1000 bis 2000	± 0,50	± 1,20	± 3,00	± 6,00
über 2000 bis 4000	-	± 2,00	± 4,00	± 8,00

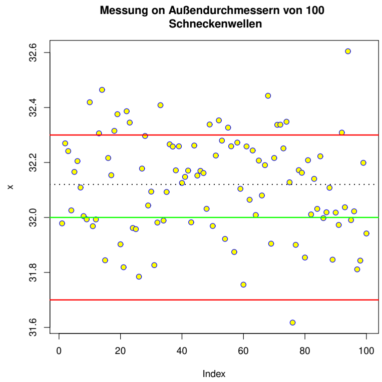
Qualitätskontrolle

Produktion von Schneckenwellen

- 32mm Außendurchmesser
- Toleranz (Stufe m) $\pm$ 0.3mm

Frage: Wo liegt Prozessmitte?
$\longrightarrow$ Konfidenzintervall Mittelwert

Frage: 90% Sicherheit, dass neues Teil unter Toleranzobergrenze liegt?
$\longrightarrow$ Konfidenzintervall Vorhersage neues Teil



Simulierte Konfidenzintervalle bei linearer Regression 45 / 34

Ein ’Gedankenexperiment’ bei lin. Reg.

Gedankenexperiment: Gegeben sei ein ’wahres’ einfaches lineares Modell mit $\alpha = \beta = \sigma^2 = 1$

$$y_i = 1 + x_i + \epsilon_i$$

Messe Y jeweils 11 mal für $x_1 = 0, x_2 = 0.1, \ldots, x_{11} = 1$

Erzeuge hierfür jeweils 11 Standardnormalverteilte Residuenwerte ϵ_i

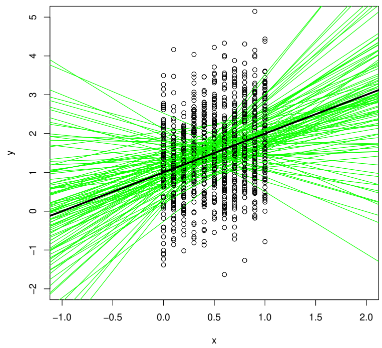
Fitte gerade und wiederhole diese Simulation 100 mal!

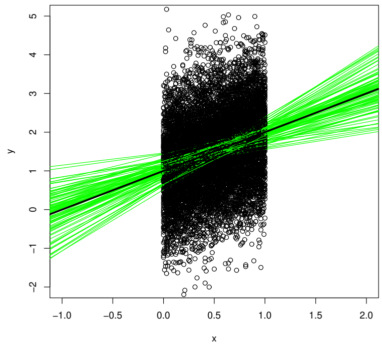
Ergebnis Simulation

Grüne Zone: Unsicherheit Vorhersage lin.Mod

Schwarze Zone: Residuenstreuung

Gesamtunsicherheit Vorhersage neuer Punkt: 'Grün+Schwarz'





Die grüne Zone Lässt sich mit mehr Messpunkten verkleinern, die schwarze nicht!

Konfidenzbereich für die Modellvorhersage

Fragestellung Modellvorhersage

Gegeben ein Lineares Modell $\underline{Y} = \underline{\underline{M}}\,\underline{\beta} + \underline{\epsilon}$, wie groß ist die Unsicherheit für die Vorhersage des *Mittelwertes* der Response Y für neue Daten $\underline{x}^{\text{new}} = (x_1^{\text{new}}, \dots, x_d^{\text{new}})$?

Aus den Daten erhalte ich $\underline{m}' = (g_1(\underline{x}^{\text{new}}), \dots, g_q(\underline{x}^{\text{new}}))$

Def: Schätzung des Mittelwertes

$$\hat{y} = \underline{m}'\underline{\hat{\beta}} \tag{11}$$

Satz: Nach dem Transformationsgesetz für normalvert. Z.V. und dem Satz über die Verteilung von $\underline{\hat{\beta}}$ gilt

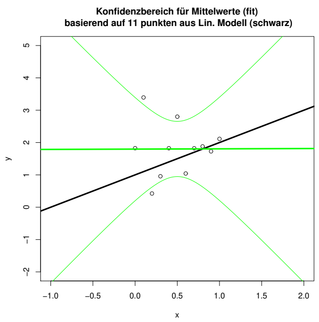
$$\hat{y} \sim N(\underline{m}'\underline{\beta}, \sigma^2 \underline{m}'(\underline{\underline{M}}'\underline{\underline{M}})^{-1}\underline{m}) \tag{12}$$

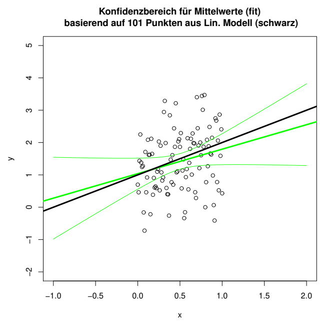
Konfidenzintervalle Vorhersage Mittelwerte

Satz: Ein 2-seitiges Konfidenzbereich zum Konfidenzniveau $1-\alpha$ für die Modellvorhersage des Mittelwerts bei $\underline{x}^{\text{new}}$ ist

$$\underline{m}'\underline{\hat{\beta}} \pm t_{1-\alpha/2}(n-q)\hat{\sigma}\sqrt{\underline{m}'(\underline{\underline{M}}'\underline{\underline{M}})^{-1}\underline{m}} \tag{13}$$

Die Aussage $\underline{m}'\underline{\beta}$ liegt in diesem Intervall hat Konfidenzniveau $1-\alpha$ (Analog für einseitige Kofi-Intervalle)





Konfidenzbereich für Vorhersage neuer Wert 51 / 34

Modell für Vorhersage neuer Wert

Man stelle sich vor, dass eine Zeile 'new' aus dem linearen Modell gelöscht worden wäre...

$$Y^{\text{new}} = \underline{m}'\underline{\beta} + \epsilon^{\text{new}} \tag{14}$$

Die Unsicherheit der Vorhersage $\underline{m}'\underline{\hat{\beta}}$ von Y^{new} ergibt sich

- 1) Aus der Unsicherheit von $\underline{\hat{\beta}}$
- 2) Aus der Streuung von ϵ^{new}

$\epsilon^{\text{new}} \sim N(0,\sigma^2)$ gleichverteilt zu und unabhängig von ϵ_i

Konfidenzbereich für die Vorhersage neuer Werte

Satz: Die Verteilung von Y^{new} ist im linearen Modell

$$Y^{\text{new}} \sim N\left(\underline{m}'\underline{\beta}, \sigma^2\left(\underline{m}'(\underline{\underline{M}}'\underline{\underline{M}})^{-1}\underline{m} + 1\right)\right) \tag{15}$$

Satz: Im lin. Modell ist die Vorhersage mit W-keit $1-\alpha$ wahr, dass der neue, unbeobachtete Wert zu $\underline{x}^{\text{new}}$ mit zugeöriger Modell-Matrix-Zeile $\underline{m}'$ im folgenden Bereich liegt

$$\underline{m}'\underline{\hat{\beta}} \pm t_{1-\alpha/2}(n-q)\hat{\sigma}\sqrt{\underline{m}'(\underline{\underline{M}}'\underline{\underline{M}})^{-1}\underline{m} + 1} \tag{16}$$

Analog für halboffene Konfi-Intervalle

Konfidenz Vorhersage lin. Reg. neuer Wert

In Blau die Konfidenzbereiche für neue Werte, in grün die für den Mittelwert