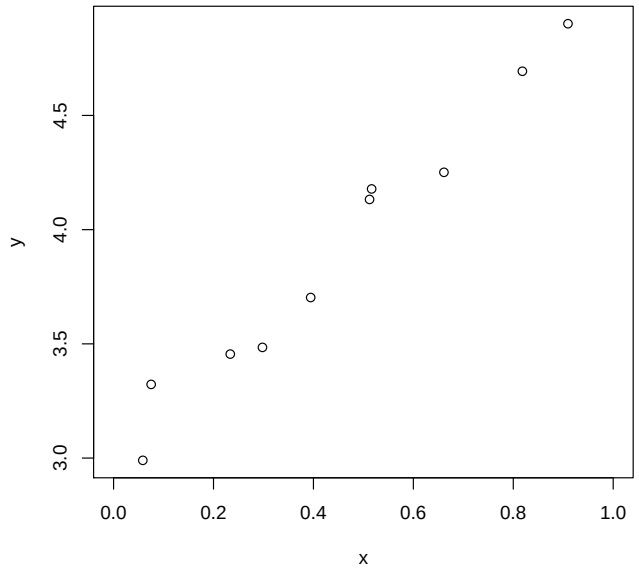
9. Lineare Regression



KQ-Schätzung

Es ist eine Gerade $y = \beta_1 + \beta_2 x$ gesucht, welche die Punktwolke in der Abbildung 'bestmöglichst' approximiert.
Dazu betrachten wir eine bivariate Zufallsvariable (Y, X) mit Beobachtungen $(y_i, x_i), i = 1, \ldots, n$ und definieren den Schätzer für die Parameter der Gerade als

$$(\hat{\beta}_1, \hat{\beta}_2) = \underset{\beta_1, \beta_2}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_i)^2.$$

KQ-Schätzung

Dieser Schätzer $(\hat{\beta}_1, \hat{\beta}_2)$ heißt (aus offensichtlichen Gründen) Kleinster-Quadrate-Schätzer und repräsentiert diejenige Gerade durch die Punktwolke, welche den quadratischen vertikalen Abstand jeder Beobachtung zur Geraden minimiert. Andere Kriterien sind denkbar, wie etwa

$$(\hat{\beta}_1, \hat{\beta}_2) = \underset{\beta_1, \beta_2}{\operatorname{argmin}} \sum_{i=1}^{n} |y_i - \beta_1 - \beta_2 x_i|.$$

Zielgröße

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \in \mathbb{R}^n$$

Einflussgrößen

$$X^j = \begin{pmatrix} X_1^j \\ X_2^j \\ \vdots \\ X_n^j \end{pmatrix}, j = 1, \ldots, k$$

welche wir in einer Matrix $\mathbf{X} = (X^1, X^2, \ldots, X^k) \in \mathbb{R}^{n,k}$ aggregieren.

Modellparameter

Die Parameter der Geraden ($k = 1$) bzw. Hyperebenen ($k > 2$) sind durch

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} \in \mathbb{R}^k$$

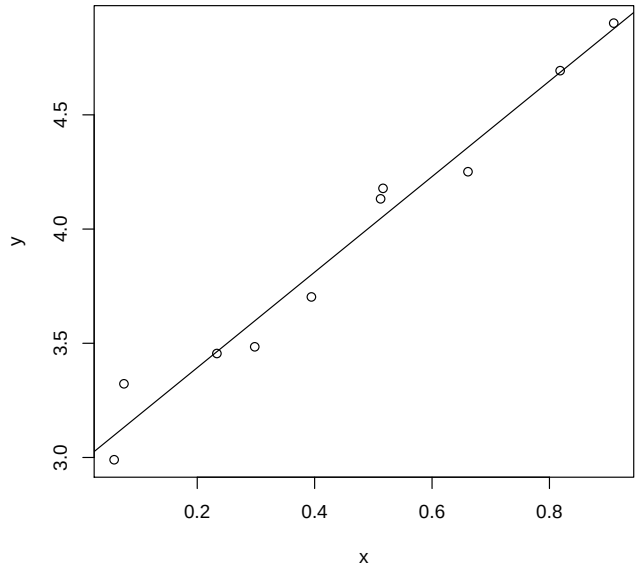
gegeben und wir betrachten das Modell $Y = \mathbf{X}\beta$.
Gesucht ist nach dem Kriterium der Kleinsten Quadrate ein Schätzer $\hat{\beta}$, sodass $||Y - \mathbf{X}\hat{\beta}||_2 \leq ||Y - \mathbf{X}\beta||_2 \forall \beta \in \mathbb{R}^k$.

KQ-Schätzung

Sei der Rang von $\mathbf{X}$ gleich k. Dann gilt:

$$\hat{\beta} = \underset{\beta}{\text{argmin}} \, ||Y - \mathbf{X}\beta|| = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top Y$$

KQ-Schätzung



Das Lineare Regressionsmodell

Das Modell

$$Y = \mathbf{X}\beta + U$$

heißt lineares Regressionsmodell. Dabei ist

$$U = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix} \in \mathbb{R}^n$$

ein n-dimensionaler Zufallsvektor.

Eigenschaften

- gegeben sind mehrere stetige Merkmale $Y, X^1, \ldots, X^k$
- $X^1, \ldots, X^k$ verursachen Y und *nicht* umgekehrt
- der Zusammenhang ist linear, also $Y_i = \sum_{j=1}^{k} \beta_j X_i^j + U_i$
- die X-Variablen heißen unabhängige Variable, Regressoren, exogene Variable oder Design-Variable
- die Y-Variable heißt abhängige Variable, Regressant, endogene Variable oder Response-Variable
- U sind nicht beobachtbare Störgrößen.

Annahmen

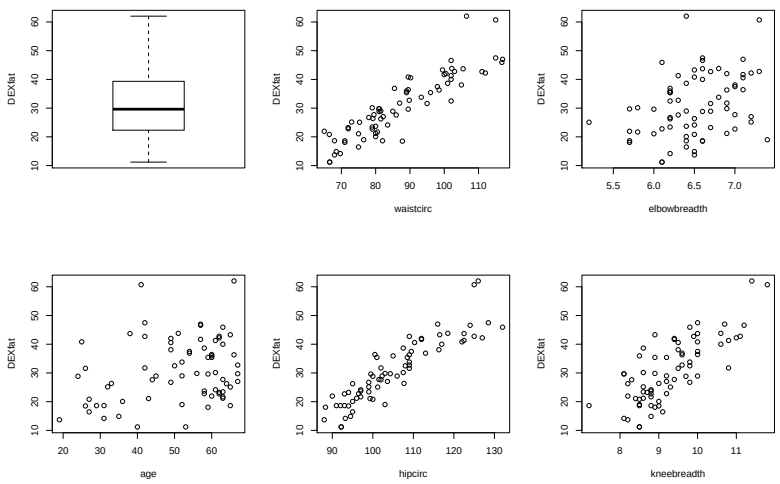
Zudem treffen wir vier Annahmen:
A1) $\mathbf{X}$ ist eine feste (nicht zufällige) $n \times k$ Matrix mit vollem Spaltenrang, also $\text{Rang}(\mathbf{X}) = k$.
A2) U ist ein Zufallsvektor mit $E(U) = (E(U_1), E(U_2), \ldots, E(U_n))^\top = 0$.
A3) Die Komponenten von U sind paarweise unkorreliert und haben alle die gleiche Varianz σ^2, formal: $\text{Cov}(U) = \sigma^2 \mathbf{I_n}$.
A4) $U \sim N(0, \sigma^2 \mathbf{I_n})$

Hinweis: Aus Annahme **A4)** folgt **A2)** und **A3)**.

Körperfettmessung

Garcia et al. (2005, Obesity Research) untersuchten $n = 71$ Frauen und erhoben (unter anderem) $k = 5$ Einflussgrößen (Alter, Bauchumfang, Hüftumfang, Ellenbogenbreite und Kniebreite), um deren Einfluss auf die Zielgröße, den Körperfettanteil gemessen mittels Dual Energy X-Ray Absorptiometry (DXA), zu untersuchen.
Es stellen sich folgende Fragen: Welche der unabhängigen Variablen haben tatsächlich einen Einfluss auf den Körperfettanteil? Welche haben einen positiven und welche einen negativen Einfluss? Kann man aus den unabhängigen Variablen auf den Körperfettanteil schließen? Diese Fragen können mittels eines linearen Regressionsmodells beantwortet werden.

Körperfettmessung



Geschätzte β-Koeffizienten

Model

```
(Intercept)         age   waistcirc     hipcirc elbowbreadth
   -59.5732      0.0638      0.3204      0.4340     -0.3012
kneebreadth
     1.6538
```

Interpretation der geschätzten Parameter:

- Intercept: Wenn alle anderen Kovariablen gleich 0 sind, beträgt der Körperfettanteil **durchschnittlich** $\beta_0 = -59.5732$ (oft nicht sinnvoll interpretierbar).

Begründung:

- Lineares Modell als (geschätzten) bedingten Erwartungswert betrachten:
  $E(Y_i|\mathbf{X_i} = \mathbf{x_i}) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k \Rightarrow E(Y_i|\mathbf{X_i} = \mathbf{0}) = \beta_0$

Interpretation der β-Koeffizienten - Fortsetzung

- Kovariablen: Erhöht sich die Kovariable (X^j) um eine Einheit, so erhöht sich die Zielvariable (Y) **durchschnittlich** um β_j Einheiten, wenn alle anderen Kovariablen gleich bleiben.
  z.B. Alter: Steigt das Alter um ein Jahr, so erhöht sich der durchschnittliche Körperfettanteil um $\beta_1 = 0.0638$, bei konstanthalten aller anderen Kovariablen.

Begründung:

- Erhöhung der Kovariable x_j um 1 bedeutet:
  $\beta_0 + \beta_1 x_1 + ... + \beta_j(x_j + 1) + ... + \beta_k x_k =$
  $\beta_0 + \beta_1 x_1 + ... + \beta_j x_j + \beta_j \cdot 1 + ... + \beta_k x_k =$
  $E(Y_i|\mathbf{X_i} = \mathbf{x_i}) + \beta_j = E(Y_i + \beta_j|\mathbf{X_i} = \mathbf{x_i})$

Eigenschaften der KQ-Methode

Unter A1, A2 und A3 ist $\hat{\beta}$ ein erwartungstreuer Schätzer für β mit Kovarianzmatrix $\text{Cov}(\hat{\beta}) = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}$.
Sei $Y \in \mathbb{R}^n$ ein beliebiger Zufallsvektor mit $E(Y) = (E(Y_1), \ldots, E(Y_n))^\top$ und

$$\text{Cov}(Y) = \begin{pmatrix} \text{Var}(Y_1) & \text{Cov}(Y_1, Y_2) & & \\ \text{Cov}(Y_2, Y_1) & \text{Var}(Y_2) & \text{Cov}(Y_2, Y_3) & \\ \vdots & \ddots & \ddots & \\ & & & \text{Var}(Y_n) \end{pmatrix}$$

mit $\text{Cov}(Y) = \text{Cov}(Y)^\top = E((Y - E(Y))(Y - E(Y))^\top)$.

Eigenschaften der KQ-Methode

Es gilt

1. $\text{Cov}(Y)$ is positiv semidefinit
2. $E(\mathbf{A}Y) = \mathbf{A}E(Y)$
3. $\text{Cov}(\mathbf{A}Y) = \mathbf{A}\text{Cov}(Y)\mathbf{A}^\top$

Lineare Funktionen

Unter Umständen sind wir an Linearkombinationen des Parametervektors β interessiert (welche auch 'Kontraste' genannt werden). Sei $c \in \mathbb{R}^k$ ein Vektor von Konstanten. Dann ist $c^\top\hat{\beta}$ eine erwartungstreue Schätzung von $c^\top\beta$ mit Kovarianzmatrix $\sigma^2 c^\top(\mathbf{X}^\top\mathbf{X})^{-1}c$.

Optimalität der KQ-Methode

Ein Schätzer $\tilde{\beta}$ heißt linear, wenn eine Matrix $\mathbf{C} \in \mathbb{R}^{k,n}$ existiert, sodass $\tilde{\beta} = \mathbf{C}Y$.

Gauß-Markov-Theorem:
Unter A1-A3 gilt:

1. $\hat{\beta}$ ist der beste lineare erwartungstreue Schätzer (BLUE) für β, d.h. $\text{Cov}(\hat{\beta}) \leq \text{Cov}(\tilde{\beta})$ im Sinne der Löwner-Halbordnung (d.h. $\text{Cov}(\tilde{\beta}) - \text{Cov}(\hat{\beta})$ psd).
2. BLUE ist eindeutig.

Desweiteren: Unter A1-A3 ist $c^\top\hat{\beta}$ der BLUE für $c^\top\beta$.

Prognose mit KQ

Gegeben sei Y und $X^1, \ldots, X^k$ mit Beobachtungen $(y_i, x_i = (x_i^1, \ldots, x_i^k)), i = 1, \ldots, n$ sowie x_{n+1}. Gesucht sei y_{n+1}. Bekannt ist, dass $Y_{n+1} = x_{n+1}^\top\beta + U_{n+1}$. Da die Störgrößen U nicht beobachtbar sind, jedoch per Annahme einen Erwartungswert gleich 0 haben, schätzen wir $\hat{Y}_{n+1} = x_{n+1}^\top\hat{\beta}$.
Es gilt: Unter A1-A3 ist $E(\hat{Y}_{n+1} - Y_{n+1}) = 0$.

Körperfettmessung

Für die 45jährige Emma mit Bauchumfang 90cm, Hüftumfang 110cm, Ellenbogenbreite 7cm und Kniebreite 10cm ist der vorhergesagte Körperfettanteil:
$E(Y_{Emma}|\mathbf{X_i} = \mathbf{x_{Emma}}) = \beta_0 + \beta_1 \cdot 45 + ... + \beta_5 \cdot 10 =$

```
          [,1]
[1,] 34.30292
```

mit den folgenden verwendeten β-Koeffizienten:

```
   (Intercept)           age     waistcirc       hipcirc  elbowbreadth
      -59.5732        0.0638        0.3204        0.4340       -0.3012
  kneebreadth
       1.6538
```


Schätzung von Varianz und Kovarianz

Es fehlen noch Schätzer für σ^2 und $\text{Cov}(\hat{\beta})$. Dazu betrachten wir die Residuen

$$\hat{U} = Y - \mathbf{X}\hat{\beta}$$

als Ersatz für die nicht beobachtbaren Störgrößen U.

1. $\hat{U} = \mathbf{M}Y = \mathbf{M}U$ mit $\mathbf{M} = \text{diag}(n) - \mathbf{H}$, wobei die sogenannte Hat-Matrix $\mathbf{H}$ gegeben ist durch $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$ und $\hat{Y} = \mathbf{H}Y$ ($\mathbf{H}$ setzt dem Y den Hut auf).
2. $\mathbf{M}$ ist orthogonaler Projektor mit Rang (gleich Spur) $n - k$.

Unter A1-A3 gilt

$$\hat{\sigma}^2 = \frac{\hat{U}^\top\hat{U}}{n-k}$$

ist eine erwartungstreue Schätzung für σ^2.

Kovarianzschätzung

Damit können wir also auch die Kovarianzmatrix $\text{Cov}(\hat{\beta})$ schätzen, und zwar als

$$\hat{\sigma}^2(\mathbf{X}^\top\mathbf{X})^{-1}.$$

Desweiteren ist es möglich, die geschätzten Koeffizienten zu standardisieren, um sie miteinander vergleichen zu können:

$$\frac{\hat{\beta}_j}{\hat{\sigma}\sqrt{\text{diag}((\mathbf{X}^\top\mathbf{X})^{-1})}}$$

Im Beispiel Körperfettmessung ergeben sich die folgenden standardisierten Regressionskoeffizienten:

```
(Intercept)           age     waistcirc       hipcirc  elbowbreadth
     -7.0471        1.7061        4.3469        4.5365       -0.2474
kneebreadth
     1.9178
```


Annahme normalverteilter Fehler (**A4**)

Eigenschaften der (multivariaten) Normalverteilung:

Eine n-dimensionale Zufallsvariable Z folgt einer multivariaten Normalverteilung mit Erwartungswertvektor $\mu \in \mathbb{R}^n$ und Kovarianzmatrix $\Sigma \in \mathbb{R}^{n,n}$ (symmetrisch und pd), symbolisch

$$Z \sim \mathcal{N}(\mu, \Sigma).$$

Es gilt

1. $Z \sim \mathcal{N}(\mu, \Sigma) \Rightarrow \text{E}(Z) = \mu, \text{Cov}(Z) = \Sigma$ und $Z_i \sim \mathcal{N}(\mu_i, \Sigma_{ii})$.
2. Sei $\mathbf{A} \in \mathbb{R}^{p,n}$ mit Rang gleich p und $b \in \mathbb{R}^p$, dann $\mathbf{A}Z + b \sim \mathcal{N}(\mathbf{A}\mu + b, \mathbf{A}\Sigma\mathbf{A}^\top)$.
3. Die Komponenten von Z sind stochastisch unabhängig $\iff$ $\Sigma = \text{diag}(\sigma_{ii}^2)$.
4. $\mathbf{A} \in \mathbb{R}^{p,n}, \mathbf{B} \in \mathbb{R}^{q,n}, \mathbf{A}\Sigma\mathbf{B}^\top = 0 \Rightarrow \mathbf{A}Z, \mathbf{B}Z$ sind stochastisch unabhängig.

Annahme normalverteilter Fehler (**A4**) II

KQ- und Varianzschätzung:

Es gilt

- $Y = \mathbf{X}\beta + U \sim \mathcal{N}(\mathbf{X}\beta, \sigma^2 \text{diag}(n))$
- $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top Y \sim \mathcal{N}(\beta, \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1})$

Unter A1 - A4 sind $\hat{\beta}$ und $\hat{\sigma}^2$ stochastisch unabhängig.
Unter A1 - A4 ist $\hat{\beta}$ die ML-Schätzung für β.
Unter A1 - A4 ist $\hat{\sigma}^2_{ML} = \hat{U}^\top \hat{U}/n$ die ML-Schätzung für σ^2.

Konfidenzintervalle und Tests für β

Wir möchten nun Hypothesen der Form

$$H_0 : d^\top \beta = 0 \text{ vs. } H_1 : d^\top \beta \neq 0$$

testen oder Konfidenzintervalle für den Parameter $d^\top \beta$ herleiten. Dabei ist $d \in \mathbb{R}^k$ beliebig.
Unter A1 - A4 gilt

$$\frac{d^\top \hat{\beta} - d^\top \beta}{\sqrt{\hat{\sigma}^2 d^\top (\mathbf{X}^\top \mathbf{X})^{-1} d}} \sim t_{n-k},$$

wobei t_{n-k} die t-Verteilung mit $n-k$ Freiheitsgraden bezeichnet.
Damit lautet die Testentscheidung: Lehne H_0 ab, wenn

$$T = \frac{|d^\top \hat{\beta}|}{\sqrt{\hat{\sigma}^2 d^\top (\mathbf{X}^\top \mathbf{X})^{-1} d}} > t_{n-k,1-\alpha/2}$$

und ein $(1-\alpha) \times 100\%$ Konfidenzintervall für $d^\top \beta$ ist

$$d^\top \hat{\beta} \pm t_{n-k,1-\alpha/2} \sqrt{\hat{\sigma}^2 d^\top (\mathbf{X}^\top \mathbf{X})^{-1} d}.$$

Körperfettmessung

Jetzt können wir für jede der Einflussgrößen die Teststatistik ausrechnen, dabei ist d der j-te Einheitsvektor, sodass $d^\top \beta = \beta_j$:

```
 (Intercept)           age     waistcirc
     -7.0471        1.7061        4.3469
    hipcirc elbowbreadth  kneebreadth
      4.5365       -0.2474        1.9178
```

und die zweiseitigen P-Werte aus der t-Verteilung ablesen

```
 (Intercept)           age     waistcirc
      0.0000        0.0928        0.0000
    hipcirc elbowbreadth  kneebreadth
      0.0000        0.8054        0.0595
```


Körperfettmessung

```
Call:
lm(formula = DEXfat ~ age + waistcirc + hipcirc + elbowbreadth +
    kneebreadth, data = bodyfat)

Residuals:
    Min      1Q  Median      3Q     Max
-9.1782 -2.4973  0.2089  2.5496 11.6504

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -59.57320    8.45359  -7.047 1.43e-09 ***
age           0.06381    0.03740   1.706   0.0928 .
waistcirc     0.32044    0.07372   4.347 4.96e-05 ***
hipcirc       0.43395    0.09566   4.536 2.53e-05 ***
elbowbreadth -0.30117    1.21731  -0.247   0.8054
kneebreadth   1.65381    0.86235   1.918   0.0595 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.988 on 65 degrees of freedom
Multiple R-squared:  0.8789,        Adjusted R-squared:  0.8696
F-statistic: 94.34 on 5 and 65 DF,  p-value: < 2.2e-16
```

Körperfettmessung

Und auch noch die Konfidenzintervalle

	2.5 %	97.5 %
(Intercept)	-76.45619185	-42.6902064
age	-0.01088410	0.1385129
waistcirc	0.17321558	0.4676638
hipcirc	0.24291126	0.6249985
elbowbreadth	-2.73231557	2.1299704
kneebreadth	-0.06842371	3.3760367

Wir sehen also, dass hauptsächlich der Bauch- und Hüftumfang informativ für den Körperfettanteil sind.

Kategoriale Kovariablen - Dummy-Codierung

- Bisher: Nur metrische Kovariablen in das Modell aufgenommen
- Jetzt: Auch Verwendung von Kategoriale Variablen

Verwendung der Dummy-Kodierung mit einer Referenzkategorie:
Aus einer c-kategorialen Kovariable $X_{kat} \in \{1, \ldots, c\}$ entstehen $c-1$ Dummy-Variablen:

$$X_{kat}^1 \begin{cases} 1, & \text{falls } X_{kat} = 1 \\ 0, & \text{sonst} \end{cases}, \ldots, X_{kat}^{c-1} \begin{cases} 1, & \text{falls } X_{kat} = c-1 \\ 0, & \text{sonst} \end{cases}$$

Kategorie c ($X_{kat} = c$) ist Referenzkategorie.

Kategoriale Variablen im linaren Modell

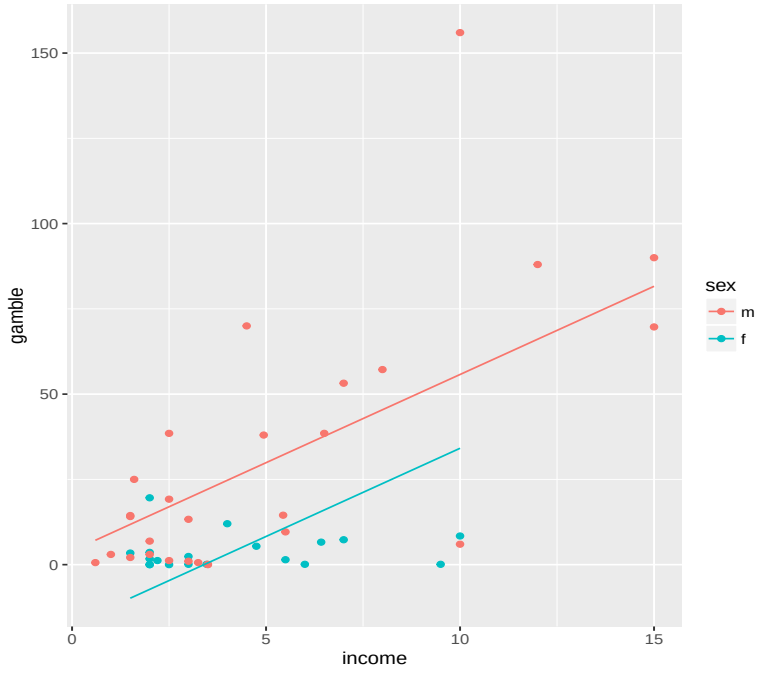
Die Dummy-Variablen aller κ kategorialer Variablen werden als Kovariablen in das Regressionsmodell aufgenommen:

$$Y_i = \underbrace{\beta_0}_{\text{Intercept}} + \underbrace{\sum_{j=1}^{k} \beta_j X_i^j}_{\text{Metrische Kovariablen}} + \underbrace{\sum_{q=1}^{\kappa} \overbrace{\sum_{m=1}^{c_q-1}}^{\text{Dummy-Variablen}} \beta_{mq} X_{kat,i}^{mq}}_{\text{Kategoriale Kovariablen}} + \underbrace{U_i}_{\text{Störterm}}$$

$\Rightarrow$ Jede Dummy-Variable besitzt ihren eigenen β-Koeffizienten.
Interpretation:

- Intercept: β_0 ist der **durchschnittliche** Wert der abhängigen Variablen in der Referenzkategorie c wenn alle metrischen Variablen 0 sind.
- Koeffizienten der kategorialen Variablen: β_{mq} beschreibt den Effekt von Kategorie m der kategorialen Variable q im Vergleich zur ihrer Referenzkategorie c_q, dies entspricht einer Parallelverschiebung der Regressionsgeraden um β_{mp}.

Kategoriale Variablen im linearen Modell II



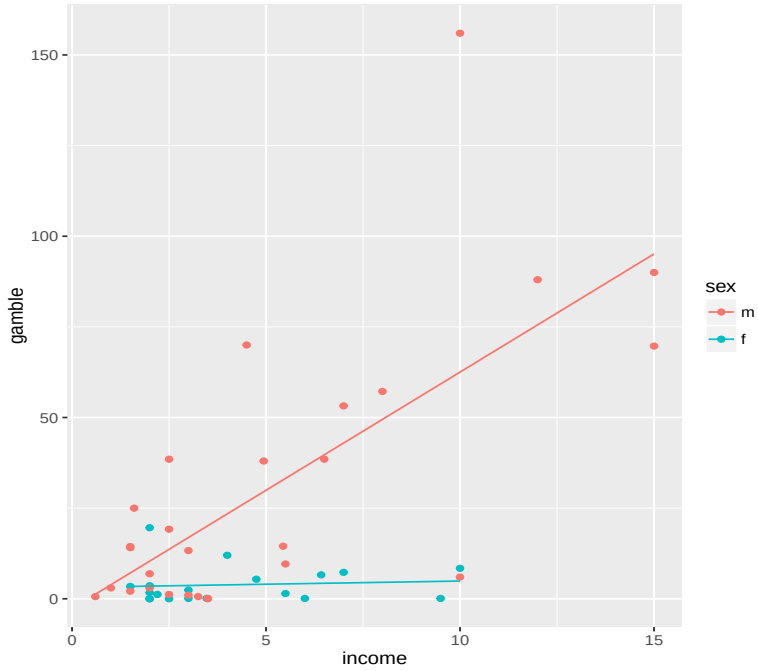
Interaktion von Variablen

- Problem: Parallelverschiebungen bietet oft keine gute Anpassung an die Daten (Siehe Graphik auf vorheriger Folie).
- Lösung: Mehr Flexibilität durch gruppenspezifische Steigungen zulassen (Entspricht Interaktion von Kovariablen)

Im Modell mit einer metrischen Variable und einer kategorialen Variable mit $c = 2$ Kategorien:

$$Y_i = \underbrace{\beta_0}_{\text{Intercept}} + \underbrace{\beta_1 X_i^1}_{\text{Metrische Kovariable}} + \underbrace{\beta_{kat} X_{kat,i}^1}_{\text{Kategoriale Kovariable}} + \underbrace{\beta_{int} X_i^1 \cdot X_{kat,i}}_{\text{Interaktionsterm}} + \underbrace{U_i}_{\text{Störterm}}$$

Interaktion von Variablen II



Modelldiagnose

Fragestellung:

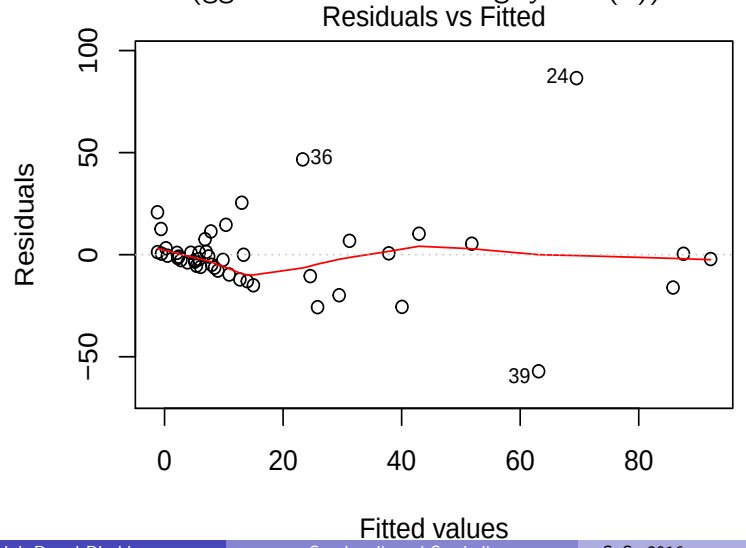
- Sind die Modellannahmen erfüllt?
- Was tun bei verletzten Annahmen?

Diagnose:

- Möglicher Einfluss von Ausreißern
- Normalverteilung des Fehlers
- Implizit auch Annahme konstanter Varianz
- Betrachten von Diagnoseplots
- *Achtung*: Bei echten Daten werden immer Abweichungen vom "Ideal" beobachtet. Es wird eher nach groben Verletzungen als nach perfekter Übereinstimmung mit der Theorie gesucht.

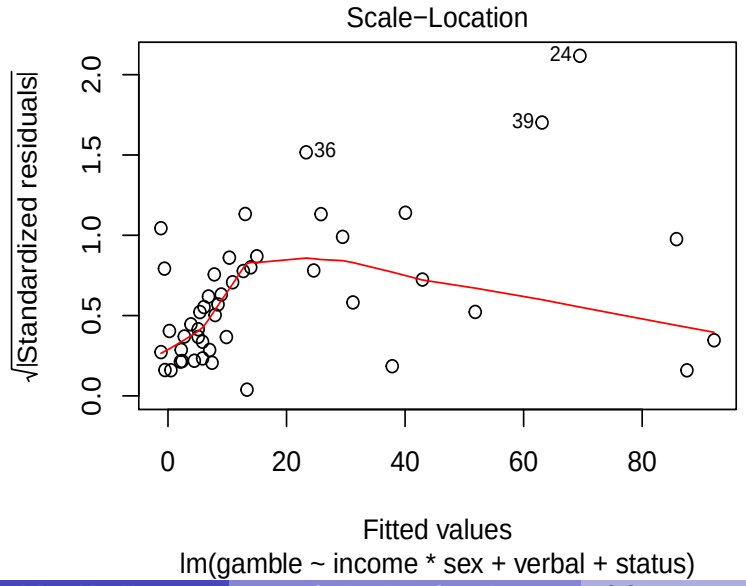
Diagnoseplots - Residuals vs. Fitted

Residuen $\hat{U}_i = y_i - \hat{y}_i$ (y-Achse) vs. Fitted values $\hat{y}_i$ (x-Achse). Die Punkte sollten gleichmäßig um 0 streuen. Insbesondere ist auf *Trichterform* zu achten (Verletzung der Annahme konstanter Varianz) und Struktur im Verlauf der Residuen (ggf. nicht-linearer Zshg. $y = f(x)$)



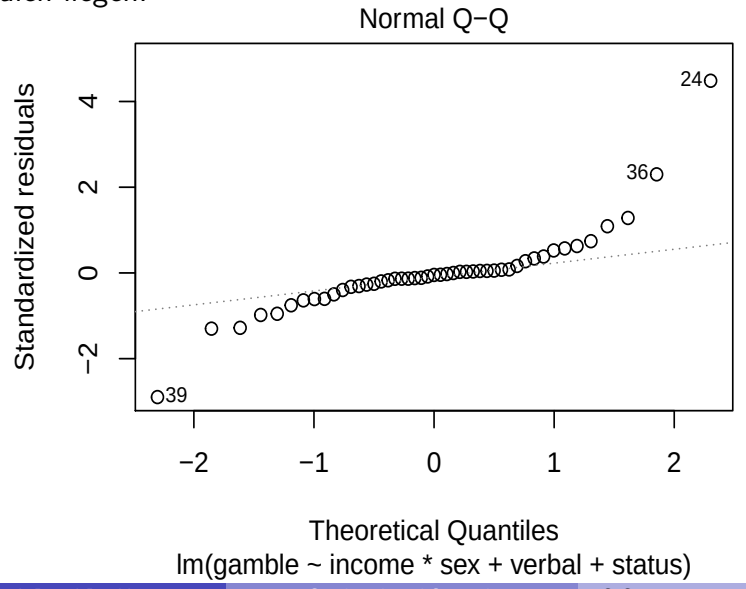
Diagnoseplots - Scale-Location Plot

Wurzel der absoluten, standardisierten Residuen vs. Fitted values. Es sollte kein Trend zu sehen sein (d.h. die eingezeichnete rote Linie sollte flach sein)



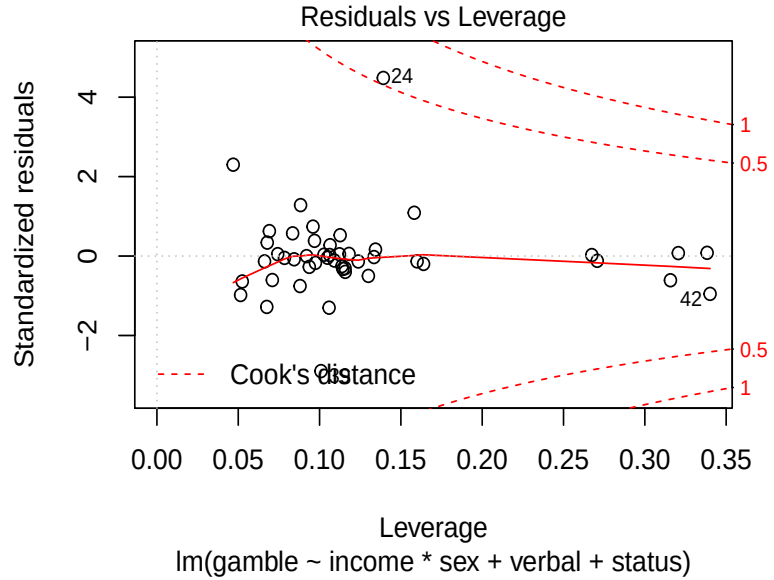
Diagnoseplots - Normal QQ-Plot

Wenn die Annahmen stimmen, sollten standardisierte Residuen einer Standardnormalverteilung folgen, d.h. die Punkte sollten auf der Diagonalen liegen.



Diagnoseplots - Residuals vs. Leverage

Liefert Hinweise auf mögliche Ausreißer. Punkte jenseits der 0.5 Höhenlinie sollten näher untersucht werden, Punkte jenseits von 1 sind kritisch.



Ausblick: Modellierung von dichotomen Y-Variablen - Logit-Modell

$Y \in \{0, 1\} \Rightarrow$ lineares Modell ungeeigenet weil:

- Wertebereich ist $\mathbb{R}$.
- Normalverteilungsannahme der Residuen ist verletzt.

Idee: Den linearen Prädiktor $\eta = \mathbf{X}\beta$ in den Wertebereich $[0, 1]$ transformieren. Eine mögliche Funktion ist die logistische Funktion:

$$f(\eta) = \frac{1}{1 + exp(-\eta)} = \frac{exp(\eta)}{1 + exp(\eta)} \in (0, 1)$$

Damit ergibt sich das logistische Regressionsmodell (Logit-Modell):

$$P(Y_i = 1 | \mathbf{X} = \mathbf{x_i}) = \frac{exp(\mathbf{x_i'}\beta)}{1 + exp(\mathbf{x_i'}\beta)}$$

Es wird also die Wahrscheinlichkeit für $Y = 1$, gegeben Kovariablen-Ausprägungen, modelliert.