

Arbeitsblatt 5

1) Statistik

Ein wahrscheinlichkeitstheoretisches Modell der linearen Regression ist

$$Y =_d \alpha + X_1\beta_1 + X_2\beta_2 + \dots + X_p\beta_p + \epsilon$$

$$\epsilon \perp\!\!\!\perp (X_1, X_2, \dots, X_p), \quad \mathbb{E}(\epsilon) = 0, \quad \mathbb{V}(\epsilon) =: \sigma^2 < \infty$$

oder kurz:

$$Y =_d X\beta + \epsilon, \quad \epsilon \perp\!\!\!\perp X, \quad \mathbb{V}(\epsilon) =: \sigma^2 < \infty$$

$(\beta_1, \dots, \beta_p)$ und σ^2 sind *Parameter* des Modells. Anstelle von Zufallsvariablen kann auch die bedingte Verteilungsfunktion benutzt werden, um das Modell zu beschreiben:

$$\Pr(Y \leq y | X = x) = \Pr(Y - x\beta \leq y - x\beta | X = x)$$

$$= \Pr(\epsilon \leq y - x\beta) = F_\epsilon(y - x\beta)$$

Man unterstellt dann n unabhängige und identisch verteilte Kopien dieses Modells:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} =_d \begin{pmatrix} X_1\beta \\ X_2\beta \\ \vdots \\ X_n\beta \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Bisher ist nur die bedingte Verteilung von Y gegeben X beschrieben. Um das Modell vollständig zu beschreiben (etwa um es zu simulieren), muss noch die Verteilung des Vektors X angegeben werden.

2) Simulation

Simulieren Sie 50 Beobachtungen des Modells

$$Y := 0.2 - 0.5X_1 + 0.2X_2 + \epsilon$$

mit ϵ ist normalverteilt mit Erwartungswert 0 und Varianz 1 (`rnorm(50,0,1)`), X_1 ist gleichverteilt auf dem Intervall $(0,2)$ (`runif(50,0,2)`) und X_2 ist normalverteilt mit Erwartungswert 1 und Varianz 1. X_1 , X_2 und ϵ sollen paarweise unabhängig sein.

a) Berechnen Sie den Erwartungswert $\mathbb{E}(Y)$ und vergleichen Sie mit den simulierten Ergebnissen.

b) Berechnen Sie den bedingten Erwartungswert $\mathbb{E}(Y | X_1 = 0.5, X_2 = 0)$ und vergleichen Sie mit den simulierten Ergebnissen.

c) Berechnen Sie den bedingten Erwartungswert $\mathbb{E}(Y | X_1 = 0.5)$ und vergleichen Sie mit den simulierten Ergebnissen.

d) Berechnen Sie die bedingte Varianz $\mathbb{V}(Y | X_1 = 0.5, X_2 = 0)$.

e) Berechnen Sie die unbedingte Varianz $\mathbb{V}(Y)$.

f) Berechnen Sie die bedingte Varianz $\mathbb{V}(Y | X_1 = 0.5)$.

3) Statistiken, Schätzer und Gütekriterien

Jede Funktion von $((Y_1, X_1), (Y_2, X_2), \dots, (Y_n, X_n))$ heißt *Statistik*. Genauer soll noch gelten, dass es eine Rechenvorschrift gibt, um den Wert der Funktion für alle n (möglicherweise ab einer Mindestgröße n) auszurechnen. Insbesondere sind $g_0(x_1, \dots, x_n) := \mathbb{3}$, $g_1(x_1, \dots, x_n) := x_1$, $g_2(x_1, \dots, x_n) := \sum_{i=1}^n x_i/n$, $g_3(x_1, \dots, x_n) := \min(x_1, \dots, x_n)$ Statistiken. Offenbar sind Statistiken wie $g_0(\cdot)$ oder $g_1(\cdot)$ schlecht, weil sie alle weitere Information ignorieren. Ein *Schätzer* $\hat{\beta}$ für einen Parameter β ist eine Statistik (also eine Rechenvorschrift $\hat{\beta}(x_1, \dots, x_n)$, die auf alle Daten angewandt werden kann), der möglichst „nah“ an den Parametern β liegt. Da man β nicht kennt, soll das für alle möglichen β des Modells gelten. Ein mögliches Gütekriterium ist das des kleinsten erwarteten quadratischen Abstands:

$$L(\beta, \hat{\beta}) := \mathbb{E}_\beta \left((\beta - \hat{\beta}(X_1, \dots, X_n))^2 \right)$$

a) Zeigen Sie: Es gibt keine gleichmäßig besten Schätzer, also Schätzer, die $L(\beta, \hat{\beta})$ für *alle* β minimieren. *Hinweis:* Benutzen Sie eine Variante von $g_0(\cdot)$ als $\hat{\beta}(\cdot)$.

b) Wenn X normalverteilt mit Erwartungswert β und Varianz 1 ist, dann ist $L(\beta, g_2(X_1, \dots, X_n))$ konstant in β .

c) Was passiert, wenn X nicht normalverteilt ist?

d) Wenn $\mathbb{E}_\beta(\hat{\beta}(X_1, \dots, X_n)) = \beta$ für alle β ist, dann heißt $\hat{\beta}(\cdot)$ *erwartungstreu* für β . Zeigen Sie, dass der Mittelwert $g_2(\cdot)$ erwartungstreu für den Erwartungswert einer Normalverteilung ist.

e) Zeigen Sie: Unter allen erwartungstreuen Schätzern des Erwartungswerts einer Normalverteilung minimiert der Mittelwert $g_2(\cdot)$ das Gütekriterium $L(\beta, \hat{\beta})$ (das in diesem Fall die Varianz ist).

f) Was passiert, wenn an Stelle der Normalverteilung eine beliebige Verteilung (mit endlicher Varianz) verwendet wird? Was, wenn eine andere (parametrische) Verteilungsklasse verwendet wird?

4) Methode der kleinsten Quadrate (OLS) und Berechnung der linearen Regression

Für das (wahrscheinlichkeitstheoretisch formulierte) lineare Modell $Y =_d X\beta + \epsilon$ ist der erwartungstreu Schätzer kleinster Varianz (minimales $L(\cdot, \cdot)$) durch

$$\hat{\beta} := (X'X)^{-1}(X'Y)$$

gegeben. Dies ist der sogenannte Kleinst-Quadrate-Schätzer (Ordinary Least Squares Estimator). Denn zunächst ist diese Funktion erwartungstreu:

$$\begin{aligned} \mathbb{E}_\beta \left(\hat{\beta}((Y_1, X_1), \dots, (Y_n, X_n)) \right) &= \mathbb{E} \left(\mathbb{E} \left(\hat{\beta} \mid X_1, \dots, X_n \right) \right) \\ &= \mathbb{E} \left((X'X)^{-1} X' \mathbb{E}(Y \mid X) \right) \\ &= \mathbb{E} \left((X'X)^{-1} X' (X\beta) \right) = \beta \end{aligned}$$

Und da sie für jedes fixe $X = x$ die kleinste Varianz unter allen erwartungstreuen Schätzern hat, so auch insgesamt (im Durchschnitt über die Verteilung der X). Zudem ist $\hat{\beta}$ eine lineare Funktion der Y , nämlich gerade die Linearkombination $((X'X)^{-1}X')Y$. Der Kleinst-Quadrate-Schätzer ist daher in dem Fall der linearen Regression der beste Schätzer unter allen linearen und erwartungstreuen Schätzern (BLUE: best linear unbiased estimator).

`lm(y ~ x1 + x2)` führt die lineare Regression in R durch. Man kann das auch mit der Formel für den Kleinst-Quadrat-Schätzer in R berechnen. Dazu muss man zunächst die entsprechenden Matrizen bilden. Mit den Daten aus der ersten Aufgabe ergibt sich:

```
X <- cbind(1,x1,x2)
XX <- crossprod(X)
### oder: XX <- t(X) %*% X
A <- solve(XX)
```

```
B <- crossprod(X,Y)
### oder: B <- t(X) %*% Y
beta <- A %*% B
beta
```

Vergleichen Sie diese Ergebnisse mit den Ergebnissen von

```
erg <- lm( y ~ x1 + x2)
summary(erg)
### oder: erg$coeff
```

5) Bilder

Oft werden lineare Regressionen nur 2-dimensional veranschaulicht. Unser Beispiel hat drei Variable, aber wir koennen einige Tricks benutzen, uns ein entsprechendes Bild zu machen. Das Paket `rgl` erzeugt Grafiken, die wie 3D-Bilder wirken und auch so manipuliert werden können:

```
library(rgl)
```

```
open3d()
```

```
rgl.bg(color="white")
##points3d(y,x1,x2, col=rainbow(100), size=5) ### ohne Achsen
plot3d(x1,y,x2, col=rainbow(100), size=5) ## mit Achsen
```

```
### Das Koordinatensystem in rgl ist zunaechst ungewoehnlich, wird
### aber in fast allen 3D Programmen verwandt: x-Richtung: nach
### rechts, y-Richtung: nach oben, z-Richtung: nach hinten.
### Das ist ein linkshaendiges System.
```

```
### Modell:
yy <- outer(x1,x2, function(x1,x2) 0.2 - x1/2 + 0.2*x2)
##yy ist Matrix der y Werte
```

```
### Modell-Regression:
rgl.surface(x=x1,y=yy,z=x2, col="blue", alpha=0.1, ambient="gray50")
```