

Serie 11

1. Wir bohren ein Loch in einen Permafrostboden. In den Tiefen von 0, 0.2, 0.5, 0.6, 0.8, 0.9, 1.2 und 6 m messen wir die Bodentemperatur. Diese könnte im Sommer folgende Werte haben: 6, 4.2, 0.6, -2.1, -5.2, -7.3, -8.9 und 15°C.

Als Hilfe sind bei den Unteraufgaben die R Befehle angegeben.

R-Hinweise: jeweils .. durch korrekte Werte ersetzen!

- a) Zeichne ein Streudiagramm der Temperatur in Abhängigkeit der Tiefe. Was fällt Dir bei diesen Daten auf? Gib (zwei) mögliche Interpretationen der Daten.

```
> tiefe <- c(..); temp <- c(..)
> plot(tiefe, temp, main="Streudiagramm")
```

- b) Berechne die empirische Korrelation der Daten ohne den Ausreisser und vergleiche sie mit derjenigen aller Daten ($\rho = 0.6$).

```
> tiefeoA <- tiefe[-..]; tempoA <- temp[-..] # 1 Beobachtung weglassen
> cor(..,..)
```

- c) Passe eine Gerade an die Daten an. (Schätze den Achsenabschnitt β_0 und die Steigung β_1 nach der Methode der kleinsten Quadrate und zeichne die Gerade in das Streudiagramm ein.) Lasse dann den Ausreisser weg und schätze die Gerade mit den übrigen Daten.

```
> fit1 <- lm(temp ~ tiefe) # Lineares Modell anpassen
> fit2 <- lm(tempoA ~ tiefeoA) # Lineares Modell anpassen ohne Ausreisser
> ## Regressionsgerade fuer fit1 zeichnen mit gestrichelter Linie
> abline(fit1,lty=2)
> abline(fit2) # Regressionsgerade fuer fit2 zeichnen
```

- d) Könnte die Steigung auch 0 sein? Teste $H_0 : \beta_1 = 0$ gegen $H_A : \beta_1 \neq 0$ mit und ohne Ausreisser.

```
> summary(fit1) # Regressions-Output fuer fit1 ausgeben
> summary(fit2)
```

2. Bei einer Untersuchung der Beziehung zwischen dem Verkehrsfluss x (in 1000 Autos pro 24 Stunden) und dem Bleigehalt y ($\mu\text{g}/(\text{g Trockengewicht})$) von Baumrinden in der Nähe der Schnellstrassen wurden folgende Daten erhoben:

x_i	8.3	8.3	12.1	12.1	17.0	17.0	17.0	24.3	24.3	24.3	33.6
y_i	227	312	362	521	640	539	728	945	738	759	1263

Die Daten stehen im File `strasse.dat` mit der erklärenden Variable `verkehr` und mit der Zielvariable `blei` zur Verfügung.

- a) Zeichne ein Streudiagramm dieser beiden Grössen.

R-Anleitung:

```
> strasse <- read.table("http://stat.ethz.ch/Teaching/Datasets/strasse.dat",
  header=TRUE)
> plot(strasse[, "verkehr"], strasse[, "blei"], main="Streudiagramm")
```

- b) Schätze die Koeffizienten β_0 und β_1 und die Streuung σ der Messfehler.

```
> strasse.fit <- lm(blei ~ verkehr, data=strasse) # Lineares Modell anpassen
> summary(strasse.fit) # Lineares Modell anzeigen
> abline(strasse.fit) # Regressionsgerade in das Diagramm einzeichnen
```

- c) Ist die Steigung β_1 signifikant von Null verschieden?

d) Mache eine Vorhersage an der Stelle $x = 40$.

3. Die untenstehenden Daten geben das Einkommen auf amerikanischen Höfen sowie die Anzahl Kühe und die Hoffläche (in acres) an.

Einkommen (Dollars)	960	830	1260	610	590	900	820	880	860	760
Anzahl Kühe (cows)	18	0	14	6	1	9	6	12	7	2
Hofgrösse (acres)	60	220	180	80	120	100	170	110	160	230
Einkommen (Dollars)	1020	1080	960	700	800	1130	760	740	980	800
Anzahl Kühe (cows)	17	15	7	0	12	16	2	6	12	15
Hofgrösse (acres)	70	120	240	160	90	110	220	110	160	80

An diese Daten wurde ein lineares Regressionsmodell der Form

$$Dollar_i = \beta_0 + \beta_1 cows_i + \beta_2 acres_i + E_i$$

mit $E_i \sim \mathcal{N}(0, \sigma^2)$ i.i.d. angepasst

Hier sehen Sie einen Teil des R-Outputs:

Coefficients:

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept)  285.457     81.379   3.508  0.0027 **
cows          32.569      3.728    ??? 1.08e-07 ***
acres         2.138       0.394   5.434 4.47e-05 ***

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 76.45 on ??? degrees of freedom

Multiple R-squared: 0.8179, Adjusted R-squared: 0.7965

F-statistic: 38.17 on ??? and ??? DF, p-value: 5.165e-07

- Die Hofgrösse hat einen statistisch signifikanten Einfluss auf das Einkommen.
 - stimmt
 - stimmt nicht
- Die Anzahl Kühe hat einen statistisch signifikanten Einfluss auf das Einkommen.
 - stimmt
 - stimmt nicht
- Wie entscheidet der Test der Nullhypothesen $H_0 : \beta_2 = 0$ gegen $H_A : \beta_2 \neq 0$?
 - H_0 beibehalten
 - H_0 verwerfen
- Wie gross ist die Anzahl Freiheitsgrade (degrees of freedom)
 - ∞
 - 20
 - 18
 - 17
 - 3
- Welches der folgenden Intervalle ist ein exaktes 95%-Vertrauensintervall für β_1 ?
 - $32.569 \pm 2.11 \cdot 3.7276$
 - $32.569 \pm 1.96 \cdot 3.7276$
 - $32.569 \pm \frac{2.11}{\sqrt{17}} \cdot 5.45$
 - keines der Intervalle
- Wie gross ist das prognostizierte Einkommen für einen Hof ohne Kühe und 100 acres Land?
 - 213
 - 285
 - 499
 - 325
- In einer einfachen linearen Regression mit nur noch der Hoffläche als erklärender Variablen, wäre der Einfluss der Hoffläche auf das Hofeinkommen signifikant.
 - auf jeden Fall
 - auf keinen Fall
 - nicht ohne weiteres klar

Besprechung: 30., 31. Mai.

Ferienpräsenz: Donnerstag 26.07., 02.08. und 16.08.2012 jeweils 14:00 – 15:00 im HG G 26.1.

Prüfungseinsicht: Donnerstag 27.09.2012 12:00 – 12:30 im HG G26.5.