Übung 10

1. Die Zufallsvariable X habe folgende Dichte:

$$f(x) = \begin{cases} c(1 - \frac{1}{10}x) & \text{falls } 0 \leq x \leq 10 \\ 0 & \text{sonst} \end{cases}$$

- **a)** Skizziere $f(x)$.
- **b)** Bestimme die Konstante c in $f(x)$ so, dass f eine Dichte darstellt.
- **c)** Berechne die (kumulative) Verteilungsfunktion $F(x)$.
- **d)** Skizziere $F(x)$.
- **e)** Bestimme die Wahrscheinlichkeit, dass X einen Wert zwischen 7 und 9 annimmt. Zeichne diese Wahrscheinlichkeit in die Skizzen von a) und d) ein.
- **f)** Berechne den Erwartungswert, die Varianz und den Median von X.

2. 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.

- **a)** Berechne die Pearson'sche Produktmomenten-Korrelation und die Spearman'sche Rangkorrelation. Was besagen diese zwei Zahlen?
- **b)** Zeichne ein Streudiagramm der Temperatur in Abhängigkeit der Tiefe. Was fällt Dir bei diesen Daten auf? Gib (zwei) mögliche Interpretationen der Daten.
- **c)** Lasse den Ausreisser weg und passe eine Gerade an die übrigen Daten an. (Schätze den Achsenabschnitt $\widehat{\alpha}$ und die Steigung $\widehat{\beta}$ nach der Methode der kleinsten Quadrate und zeichne die Gerade in das Streudiagramm ein.)
  Könnte die Steigung auch -10 sein? Teste $H_0 : \beta = -10$ gegen $H_A : \beta \neq -10$. (Rechenhilfe: $\widehat{\sigma}(\widehat{\beta}) = 0.72$.)

3. Bei einer Untersuchung der Beziehung zwischen dem Verkehrsfluss x (in 1000 Autos pro 24 Stunden) und dem Bleigehalt y ($\mu g/(g$ 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.
- **b)** Schätze die Koeffizienten α und β und die Streuung σ der Messfehler.
- **c)** Ist die Steigung β signifikant von Null verschieden?

d) Beurteile die Anpassung, indem Du das Diagramm Residuen gegen die angepassten Werte (Tukey-Anscombe Plot genannt) betrachtest.

R-Anleitung:

```
> strasse <- read.table.url("http://stat.ethz.ch/Teaching/Datasets/
                                          strasse.dat", header=T)
> par(mfrow=c(1,2))                        # Zwei Grafiken ins gleche Fenster
> plot(strasse[,"verkehr"], strasse[,"blei"], main="Streudiagramm")
> strasse.fit <- lm(blei ~ verkehr, data=strasse)        # Anpassen eines linearen
Modelles
> abline(coef(strasse.fit))                # Gerade in das Diagramm einzeichnen
> summary(strasse.fit)                     # Zusammenfassung der Anpassung
> plot(fitted(strasse.fit), resid(strasse.fit), main="Tukey-Anscombe Plot"))
                                                          # Tukey-Anscombe Plot
> abline(h=0)                                             # horizontale Gerade y = 0
```

4. Im File `no2` sind meteorologische Daten und Messungen der Konzentration von Stickstoffdioxid $[\mu g/m^3]$ in Basel im Februar 1991 enthalten. (Der Tagesgrenzwert für Stickstoffdioxid ist 80 $\mu g/m^3$.) Die Variablen heissen `NO2`, `Wind`, `Temp` und `Tag`.

Teil 1 (1 Punkt):

a) Passe das Modell

$$NO2 = \alpha + \beta_1\,Wind + \varepsilon$$

und das Modell

$$NO2 = \alpha + \beta_2\,Temp + \varepsilon$$

mittels einfacher linearer Regression an die Daten an und gib die geschätzten Parameter und ihre geschätzten Standardabweichungen an.

b) Passe das Modell

$$NO2 = \alpha + \beta_1\,Wind + \beta_2\,Temp + \varepsilon$$

mittels linearer Regression an die Daten an. Vergleiche die geschätzten Parameter $\widehat{\alpha}$, $\widehat{\beta}_1$ und $\widehat{\beta}_2$ mit jenen in Teilaufgabe a). Vergleiche auch die jeweiligen Werte des Bestimmtheitsmasses R^2.

c) Unterscheiden sich die drei angepassten Modelle jeweils signifikant von einem konstanten Wert? Sind beide Variablen `Wind` und `Temp` für die Beschreibung der Stickstoffdioxid-Konzentration notwendig?

Teil 2, freiwillig (1 Punkt):

d) Im zweiten Teil dieser Aufgabe wollen wir die üblicherweise in der linearen Regression gemachten Voraussetzungen (Unabhängigkeit der Fehler (u.a. keine zeitliche Abhängigkeit), Konstanz der Fehlervarianz (u.a. keine Ausreisser und keinen Zusammenhang mit angepassten Werten und erklärenden Variablen) und Normalverteilung der Fehler) graphisch überprüfen (vergl. Stahel (1999), Abschnitt 13.7). Als erstes untersuche, ob die Residuen normalverteilt sein können (normal plot).

e) Lassen sich Strukturen in den Residuen erkennen, wenn Du sie gegen die angepassten („gefitteten") Werte (Tukey-Anscombe Plot), gegen die Variable `Wind` oder gegen die Variable `Temp` in einem Streudiagramm aufträgst? (Wir hoffen nicht.)

f) Trage die Residuen gegen ihre zeitliche Reihenfolge auf. Was ist dazu zu bemerken?

g) Wie beurteilst Du abschliessend die Anpassung?

R-Anleitung:

```
> no2 <- read.table.url("http://stat.ethz.ch/Teaching/Datasets/no2Basel.dat",
                                                                    header=T)
```

a)
```
> no2.fit1 <- lm(NO2 ~ Wind, data=no2)            # einfache lineare Regression
> summary(no2.fit1)                               # Überblick über die Anpassung
> no2.fit2 <- lm(NO2 ~ Temp, data=no2); summary(no2.fit2)
```

b)
```
> no2.fit <- lm(NO2 ~ Wind + Temp, data=no2)      # multiple lineare Regression
> summary(no2.fit)
```

d)
```
> par(mfrow=c(2,3))
> qqnorm(resid(no2.fit))                          # normal plot
```

e)
```
> plot(fitted(no2.fit), resid(no2.fit), main="Tukey-Anscombe Plot")
> plot(no2[,"Wind"], resid(no2.fit)); abline(h=0)
> plot(no2[,"Temp"], resid(no2.fit)); abline(h=0)
```

f)
```
> plot(no2[,"Tag"], resid(no2.fit),type="b"); abline(h=0)
```

Abgabe: Montag, 19. Juni, in der Pause der Vorlesungsstunde