

Aufgabe 41

Nehmen Sie an, Sie wollen nachweisen, dass es einen signifikanten Unterschied zwischen dem mittleren Nettoeinkommen (das als normalverteilt angenommen wird) von Männern und Frauen gibt. Sie planen dazu eine Stichprobe von n_F Frauen und n_M Männern zu erheben. Sie wollen einen Signifikanztest auf einem Niveau von 1% durchführen und bei einem wahren Unterschied der Einkommen von 100 Euro eine Power von mindestens 80% haben.

Bestimmen Sie die für einen solchen Test notwendige Stichprobengröße. Benutzen Sie dazu das unter <http://homepage.stat.uiowa.edu/~rlenth/Power/> zur Verfügung gestellte Java-Applet. Berechnen Sie n unter verschiedenen Annahmen für die Standardabweichungen der Einkommen der Männer ($s_M = 50, 100, 1000$) und Frauen ($s_F = 50, 100, 1000$).

Aufgabe 42

Betrachten Sie den SPSS-Output zur linearen Regression über die Nettomiete (nm) von Münchner Wohnungen in Abhängigkeit von der Variable Anzahl Zimmer (rooms) und Zentralheizung (zh0), die als Dummyvariable (Ausprägung 0 entspricht vorhandener Zentralheizung) aufgenommen wurde.

- Interpretieren Sie die Regressionskoeffizienten.
- Testen Sie, ob die Variablen rooms und zh0 jeweils einen signifikanten ($\alpha = 0.05$) Einfluß haben.
- Berechnen Sie ein Konfidenzintervall ($\gamma = 0.95$) für den Parameter der Einflußgröße rooms ($n = 2053$).

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	201,099	65,530		3,069	,003
	rooms	164,664	22,980	,625	7,166	,000
	zh0	-213,234	71,988	-,258	-2,962	,004

a. Abhängige Variable: nm

Aufgabe 43

In einer Studie zu Lesefähigkeiten von Schülern (Christa Kieferle, Pädagogik, LMU) wurden unter anderem die folgenden Variablen von 180 Schülern aus 8 Klassen (3. und 4. Klasse) erhoben:

- **Fehlerzahl**: Anzahl der Fehler bei einem Lesetest
- **Lesezeitmin**: Leseförderzeit in Minuten (Leseförderung in der Schule)
- **Jahrgang**: Dummyvariable mit Ausprägungen 3. Klasse = 1, 4. Klasse = 0
- **Geschlecht**: Dummyvariable mit Ausprägungen männlich = 1, weiblich = 0
- **Lesenoft, Lesenselten**: Dummyvariablen für die Variable zur Frage „Wie oft liest du außerhalb der Schule?“ mit Ausprägungen oft, mittel, selten

Der Datensatz steht auf der Veranstaltungshomepage zum Download bereit, direkter Link: http://www.statistik.lmu.de/institut/ag/statsoz_neu/lehre/2012_SoSe/Stat2Soz_12/uebung/lesen2.txt

Es soll eine Regressionsanalyse durchgeführt werden mit Fehleranzahl als Zielgröße und den anderen Variablen im Datensatz als Einflussgrößen.

a) Erzeugen Sie den Output auf der nächsten Seite selbst in **R**.

Laden Sie dazu die Datei herunter und starten Sie **R**. Mit dem Befehl

```
lesen2 <- read.table(file.choose(), header = T)
```

können Sie den Datensatz in **R** einlesen; mit `head(lesen2)` können Sie sich danach zur Überprüfung die ersten Zeilen des Datensatzes anschauen. Der Befehl `attach(lesen2)` führt dazu, dass Sie die Variablennamen im Datensatz `lesen2` wie oben aufgelistet direkt benutzen können. (Ansonsten müssten Sie z.B. statt `boxplot(Fehlerzahl)` den Befehl `boxplot(lesen2$Fehlerzahl)` benutzen, um einen Boxplot für die Anzahl Fehler im Lesetest zu erhalten.) Mit den Befehlen

```
mod1 <- lm(Fehlerzahl ~ Lesezeitmin + Jahrgang + Geschlecht + Lesenoft  
          + Lesenselten)  
summary(mod1)
```

erhalten Sie dann den Output.

- b) Interpretieren Sie die Schätzungen der Regressionskoeffizienten.
- c) Welche der Variablen haben einen signifikanten ($\alpha = 0.05$) Einfluss?
- d) Berechnen Sie ein Konfidenzintervall zum Sicherheitsgrad von 99% für den Parameter der Einflussgröße **Geschlecht**.

Call:
lm(formula = Fehlerzahl ~ Lesezeitmin + Jahrgang + Geschlecht +
Lesenoft + Lesenselten)

Residuals:
Min 1Q Median 3Q Max
-14.354 -5.574 -1.361 4.344 25.291

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 14.80650 1.76753 8.377 1.75e-14 ***
Lesezeitmin -0.08726 0.03272 -2.667 0.00838 **
Jahrgang 5.90591 1.17928 5.008 1.34e-06 ***
Geschlecht 3.16466 1.22758 2.578 0.01077 *
Lesenoft -2.93325 1.38026 -2.125 0.03499 *
Lesenselten 4.94820 1.90625 2.596 0.01024 *

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

Residual standard error: 7.859 on 174 degrees of freedom
Multiple R-squared: 0.3081, Adjusted R-squared: 0.2882
F-statistic: 15.5 on 5 and 174 DF, p-value: 1.34e-12