

Aufgabe 1 (Statistische Test)

Bearbeiten Sie die Probleme

- „Schlafmittel“ (siehe: doc_schlafmittel.pdf, schlafmittel.r)
- „Werbeanzeigen“ (siehe: doc_werbeanzeigen.pdf, werbeanzeigen.r)

Aufgabe 2 (Regressionsanalyse – Lineare Einfachregression)

Bearbeiten Sie das Problem

- „Bayernwahl“ (siehe: doc_bayernwahl.pdf, bayernwahl.r)

Aufgabe 3 (Regressionsanalyse – Lineare Mehrfachregression)

Die Datei ps1985.csv enthält eine Teilmenge der im Rahmen der Bevölkerungsstichprobe des US Bureau of Census im Jahr 1985 erhobenen Daten. Der Datensatz umfasst Angaben von 534 Arbeitnehmern aus Kalifornien bezüglich 11 Variablen. Die Variablen sind:

<i>wage</i>	Arbeitnehmerlohn (US-Dollar pro Stunde),
<i>education</i>	Länge der Ausbildung in Jahren,
<i>experience</i>	Potentielle Arbeitserfahrung in Jahren (= Alter – Ausbildung – 6),
<i>ethnicity</i>	Ethnische Herkunft (kategoriale Variable mit den Stufen „cauc“ und „other“),
<i>region</i>	Region (kategoriale Variable mit den Stufen „south“ und „other“),
<i>gender</i>	Geschlecht (kategoriale Variable mit den Stufen „female“ und „male“),
<i>occupation</i>	Beruf (kategoriale Variable mit den Stufen „worker“, „technical“, „services“, „office“, „sales“, and „management“),
<i>union</i>	Gewerkschaftsmitglied (kategoriale Variable mit den Stufen „yes“ und „no“),
<i>married</i>	Verheiratet (kategoriale Variable mit den Stufen „yes“ und „no“).

Es soll mittels Regressionsanalyse untersucht werden, inwiefern die Variablen 2 – 9 als Determinanten des Arbeitnehmerlohns (Variable *wage*) angesehen werden können. [Hinweis: Es ist nicht das Ziel, den Arbeitnehmerlohn mittels der Regressoren vorherzusagen. Es soll lediglich untersucht werden, ob und ggf. wie die Regressoren den Lohn beeinflussen.]

- a) Auf welchen Annahmen basiert das klassische lineare Regressionsmodell?
- b) Untersuchen Sie zunächst den Zusammenhang von *wage* und den beiden Regressoren *education* und *experience*. Gehen Sie dabei insbesondere auf die folgenden Fragen ein:
 - Ist es sinnvoll, die Werte des Regressanden zu logarithmieren?
 - Verbessert die Wurzeltransformation der Variable *experience* den Fit der Regression?
- c) Nutzen Sie nun zusätzlich die kategorialen Variablen als potentielle Regressoren. Üben alle Variablen einen signifikanten Einfluss auf $\log(wage)$ aus? [Hinweis: Eine Umkodierung der kategorialen Variablen ist nicht erforderlich. Die R-Funktion *lm* erledigt dies automatisch.]

- d) Ist die Normalverteilungsannahme des klassischen linearen Regressionsmodells erfüllt?
- e) Auf der nachfolgenden Seite finden Sie einige regressionsanalytische Ergebnisse. Interpretieren Sie die Ergebnisse. Diskutieren Sie insbesondere den *marginalen* Effekt der Regressoren auf den Arbeitnehmerlohn. [Hinweis: Im linearen Modell

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + U$$

gibt β_i den (marginalen) Effekt bezüglich der Zielvariable Y an, wenn der Wert x_i des Regressors X_i auf $x_i + 1$ erhöht wird. Beachten Sie, dass die Zielvariable im vorliegenden Beispiel logarithmiert ist.]

- f) In zahlreichen empirischen Modellen mit einer Preisvariable als Zielvariable wird eine Logarithmustransformation der Zielvariable vorgenommen. Gibt es hierfür (neben möglichen datenanalytischen Überlegungen) auch theoretische Gründe?

Call:

```
lm(formula = log(wage) ~ education + I(experience^0.5) + region +
    gender + occupation + union, data = ps1985)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.361138	-0.282871	0.008493	0.293674	1.876056

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.845492	0.177104	4.774	2.35e-06	***
education	0.069946	0.009674	7.230	1.73e-12	***
I(experience^0.5)	0.094694	0.012665	7.477	3.23e-13	***
regionsouth	-0.099971	0.041201	-2.426	0.015587	*
gendermale	0.212388	0.041266	5.147	3.76e-07	***
occupationoffice	-0.208955	0.075367	-2.773	0.005761	**
occupationsales	-0.341585	0.090653	-3.768	0.000183	***
occupationservices	-0.389272	0.079947	-4.869	1.49e-06	***
occupationtechnical	-0.040809	0.072069	-0.566	0.571467	
occupationworker	-0.199778	0.074829	-2.670	0.007826	**
unionyes	0.197705	0.050410	3.922	9.96e-05	***

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

Residual standard error: 0.4247 on 522 degrees of freedom

Multiple R-squared: 0.3651, Adjusted R-squared: 0.3529

F-statistic: 30.01 on 10 and 522 DF, p-value: < 2.2e-16