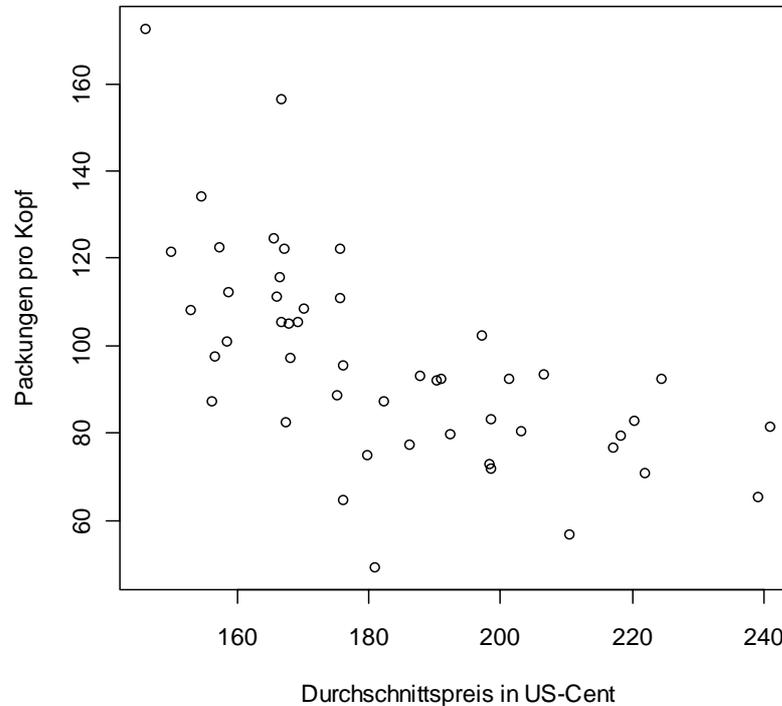


## Aufgabe 1

In einer US-Studie aus dem Jahr 1995 wurde der Zusammenhang zwischen dem Preis und dem Absatz von Zigaretten untersucht. Konkret wurde für  $n = 48$  Bundesstaaten der jeweilige Durchschnittsverkaufspreis einer Packung in Cent (Variable  $X$ ) und der jeweilige Absatz von Packungen je Einwohner (Variable  $Y$ ) im Jahr 1995 erhoben. Die Daten zeigt das Streudiagramm. Sie sollen im Folgenden eigene Berechnungen durchführen.



- Schätzen Sie die Parameter des linearen Regressionsmodells von  $Y$  bezüglich  $X$  mittels der Kleinste-Quadrate-Methode. Interpretieren Sie die geschätzten Parameter.
- Geben Sie den Korrelationskoeffizienten  $r_{XY}$  und das Bestimmtheitsmaß  $R^2$  der Regression an. Beurteilen Sie die beiden Kenngrößen.
- Ermitteln Sie für den Regressionsparameter  $\beta_1$  den Standardfehler und das 90%-Schätzintervall. Interpretieren Sie ihre Werte.
- Bestimmen Sie für  $x_0 = 200.00$  ein 90%-Prognoseintervall. Was sagt das Intervall aus?
- Zeichnen Sie die Regressionsgerade in das Streudiagramm ein.

*Rechenhilfen:*

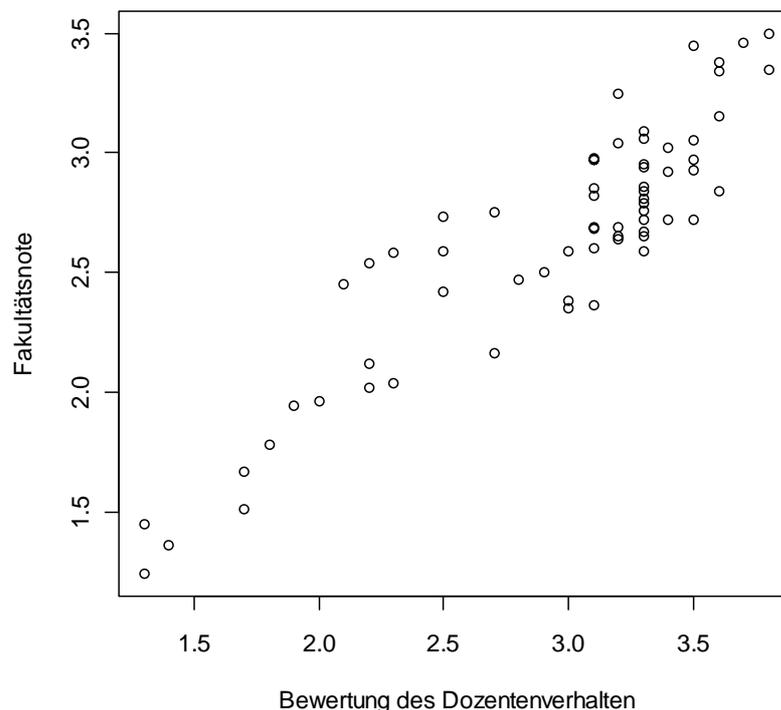
$$\sum_v x_v = 8795.964, \quad \sum_v y_v = 4623.749, \quad \sum_v x_v^2 = 1639037.827, \quad \sum_v y_v^2 = 471979.66,$$

$$\sum_v x_v y_v = 830386.2, \quad \sum_v \hat{\varepsilon}_v^2 = 16060.9$$

## Aufgabe 2

Der Spiegel veröffentlichte ein Ranking deutscher Universitäten, das auf der Befragung von Studierenden basierte. Gefragt wurde nach persönlichen Einschätzungen des Dozentenverhaltens, der Studieninhalte, der Studenten/Professoren-Relation, der Ausstattung und weiterer Aspekte der Studienbedingungen. Aus den persönlichen Einschätzungen der Studierenden wurden Durchschnittsnoten für Fakultäten und Universitäten abgeleitet.

Sie sollen den Einfluss des Dozentenverhaltens auf die Benotung untersuchen. Weiter unten sind in einem Streudiagramm die Durchschnittsnoten (Variable  $Y$ ) und die Durchschnittsbewertungen des Dozentenverhaltens (Variable  $X$ ) für die wirtschaftswissenschaftlichen Fakultäten von  $n = 66$  Universitäten graphisch dargestellt.



- Schätzen Sie die Parameter des linearen Regressionsmodells von  $Y$  bezüglich  $X$  mittels der Kleinste-Quadrate-Methode.
- Ermitteln Sie das Bestimmtheitsmaß  $R^2$  der Regression.
- Geben Sie für den Regressionsparameter  $b$  den Standardfehler und das 95%-Schätzintervall an.
- Bestimmen Sie für  $x_0 = 2.0$  ein 95%-Prognoseintervall.
- Interpretieren Sie Ihre Ergebnisse in Stichworten.

*Rechenhilfen:*

$$\sum_v x_v = 194.2 \quad , \quad \sum_v y_v = 174.32 \quad , \quad \sum_v x_v^2 = 597.66 \quad , \quad \sum_v y_v^2 = 477.1734 \quad ,$$
$$\sum_v x_v y_v = 531.894 \quad , \quad \sum_v \hat{\varepsilon}_v^2 = 3.0428$$

### Aufgabe 3

Einer kanadischen Studie über die Determinanten von Eigenheimpreisen aus dem Jahr 1987 lagen die Werte von 10 Variablen bei  $n = 546$  Verkäufen von Eigenheimen als Daten zugrunde (siehe Datei **eigenheimpreise.csv**). Die Variablen sind:

|                   |  |
|-------------------|--|
| <i>price</i>      | Verkaufspreis des Hauser in Canadian Dollar,                       |
| <i>lotsize</i>    | Grundstücksgröße in Quadratmetern,                                 |
| <i>bedrooms</i>   | Anzahl der Schlafzimmer des Hauses,                                |
| <i>bathrooms</i>  | Anzahl der Badezimmer des Hauses,                                  |
| <i>stories</i>    | Anzahl der Stockwerke,   |
| <i>garage</i>     | Anzahl der Garagenplätze für PKW,                                  |
| <i>fullbase</i>   | Vollständig unterkellertes Haus (Dummy-Variable 0 = nein, 1 = ja), |
| <i>recreation</i> | Freizeitraum (Dummy-Variable 0 = nein, 1 = ja),                    |
| <i>gasheat</i>    | Gasheizung (Dummy-Variable 0 = nein, 1 = ja),                      |
| <i>aircon</i>     | Zentrale Klimaanlage (Dummy-Variable 0 = nein, 1 = ja).            |

Es soll mittels linearer Regressionsanalyse untersucht werden, inwiefern die Variablen 2 – 10 als Determinanten der Eigenheimpreise (Variable *price*) angesehen werden können. Gehen Sie insbesondere auf die folgende Fragen ein:

- Ist es im vorliegenden Beispiel sinnvoll, die Werte der Variablen *price* und *lotsize* zu logarithmieren? Begründen Sie Ihre Antwort ausführlich.
- Ist die Normalverteilungsannahme des klassischen linearen Regressionsmodells erfüllt?
- Interpretieren Sie das geschätzte Modell. Diskutieren Sie insbesondere die *marginalen* Effekte der Regressoren auf den Verkaufspreis.
- Unter welchen Voraussetzungen sind OLS-Schätzer BLUE und was bedeutet BLUE?
- Welche Konsequenzen hat die Verletzung der Homoskedastizitätsannahme im klassischen linearen Regressionsmodell?

### Aufgabe 4

Die Datei **einkommen.csv** enthält die Daten des Beispiels 2.1.2 aus Eckey, Kosfeld & Dreger: Ökonometrie, Grundlagen, Methoden, Beispiele; Gabler 2004, S. 35 ff. Schätzen Sie mit Hilfe des Programmpakets R eine Geldnachfragefunktion.

## Aufgabe 5

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.]

- Auf welchen Annahmen basiert das klassische lineare Regressionsmodell?
- 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?
- 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.]
- Ist die Normalverteilungsannahme des klassischen linearen Regressionsmodells erfüllt?
- 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  $\beta_i$  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.]

- 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)
```

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

## Aufgabe 6

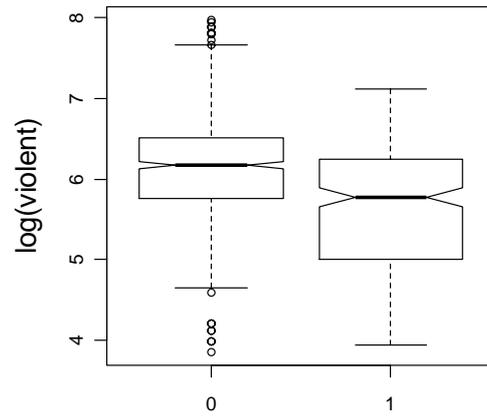
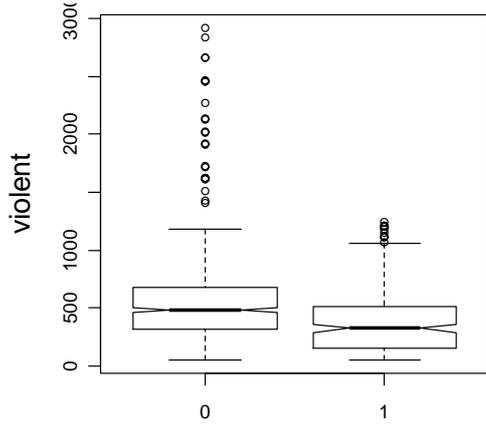
Die amerikanischen Juristen I. Ayres und J.J. Donohue veröffentlichten im Jahr 2003 eine empirische Studie mit dem Titel „*Shooting down the 'More Guns Less Crime' Hypothesis*“. Als Datenbasis dienten in der Studie Jahreswerte von 8 Merkmalsvariablen, die an 50 Staaten und einem Distrikt der USA in den Jahren 1997 bis 1999 erhoben wurden (Datenumfang  $n = 1173$ ). Die Variablen sind

- violent*    Gewaltkriminalitätsrate (Gewaltstraftaten auf 100000 Einwohner des Staates).
- law*        Recht zum Tragen von Schusswaffen in der Öffentlichkeit. Nominale Variable mit den Ausprägungen 1 = *ja* und 0 = *nein*.
- prisoners*   Inhaftierungsrate (Strafgefangene auf 100000 Einwohner des Staates).
- density*     Bevölkerungsdichte des Staates (Einwohner pro Quadratmeile).
- income*     Pro-Kopf-Einkommen des Staates in US Dollar.
- afam*        Anteil der afroamerikanischen Bevölkerung an der Gesamtbevölkerung des Staates.
- cauc*        Anteil der kaukasischen Bevölkerung an der Gesamtbevölkerung des Staates.
- male*        Anteil der Männer im Alter von 10 bis 29 Jahren an der Gesamtbevölkerung des Staates.

Eine lineare Regressionsanalyse mit *violent* als abhängiger Variable ergab die weiter unten folgenden Ergebnisse. Kommentieren Sie die Ergebnisse. Gehen Sie dabei insbesondere auch auf die folgenden Punkte ein:

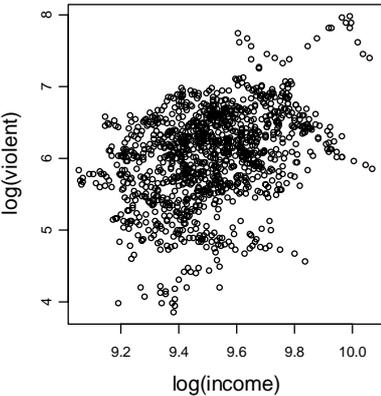
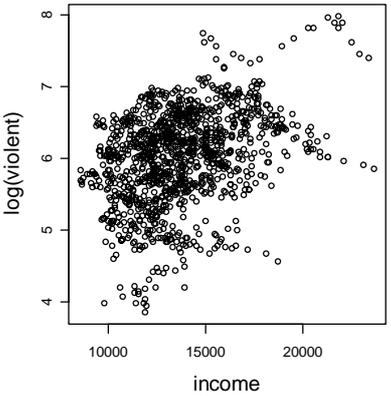
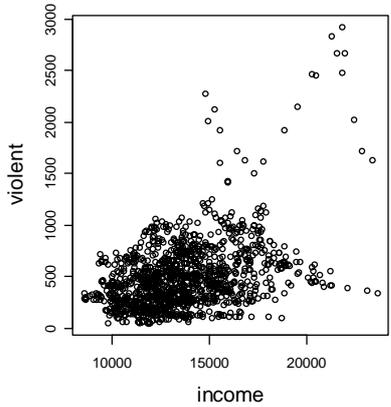
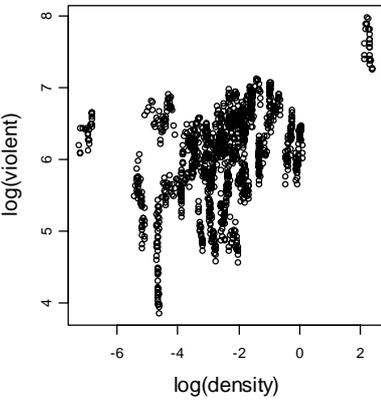
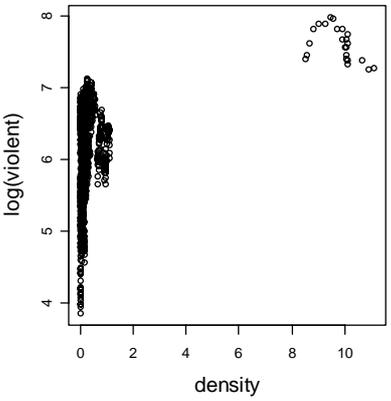
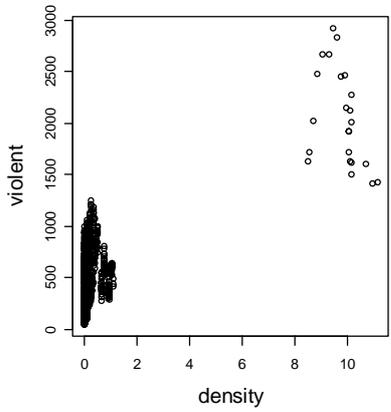
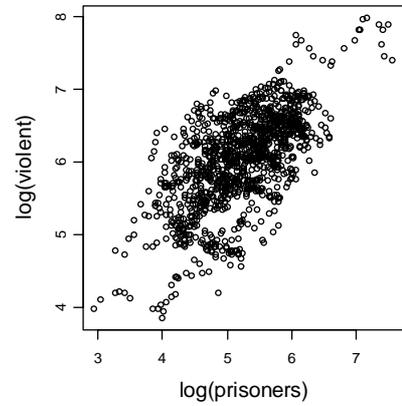
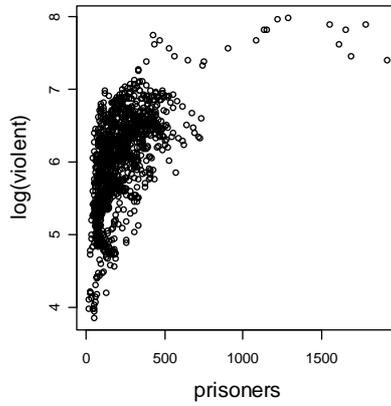
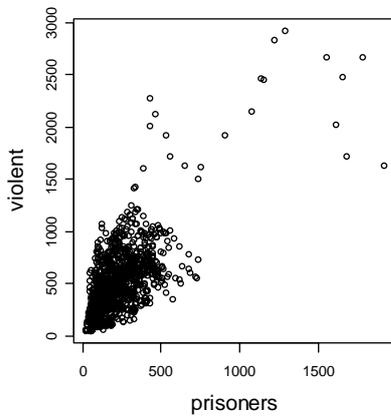
- (a) Annahmen des klassischen linearen Regressionsmodells und die Konsequenzen von Annahmeverletzungen.
- (b) Begründung der Logarithmustransformationen.
- (c) Interpretieren Sie das geschätzte Modell und insbesondere die geschätzten Regressionskoeffizienten. Wie wirkt sich das Recht zum Tragen von Schusswaffen in der Öffentlichkeit aus?

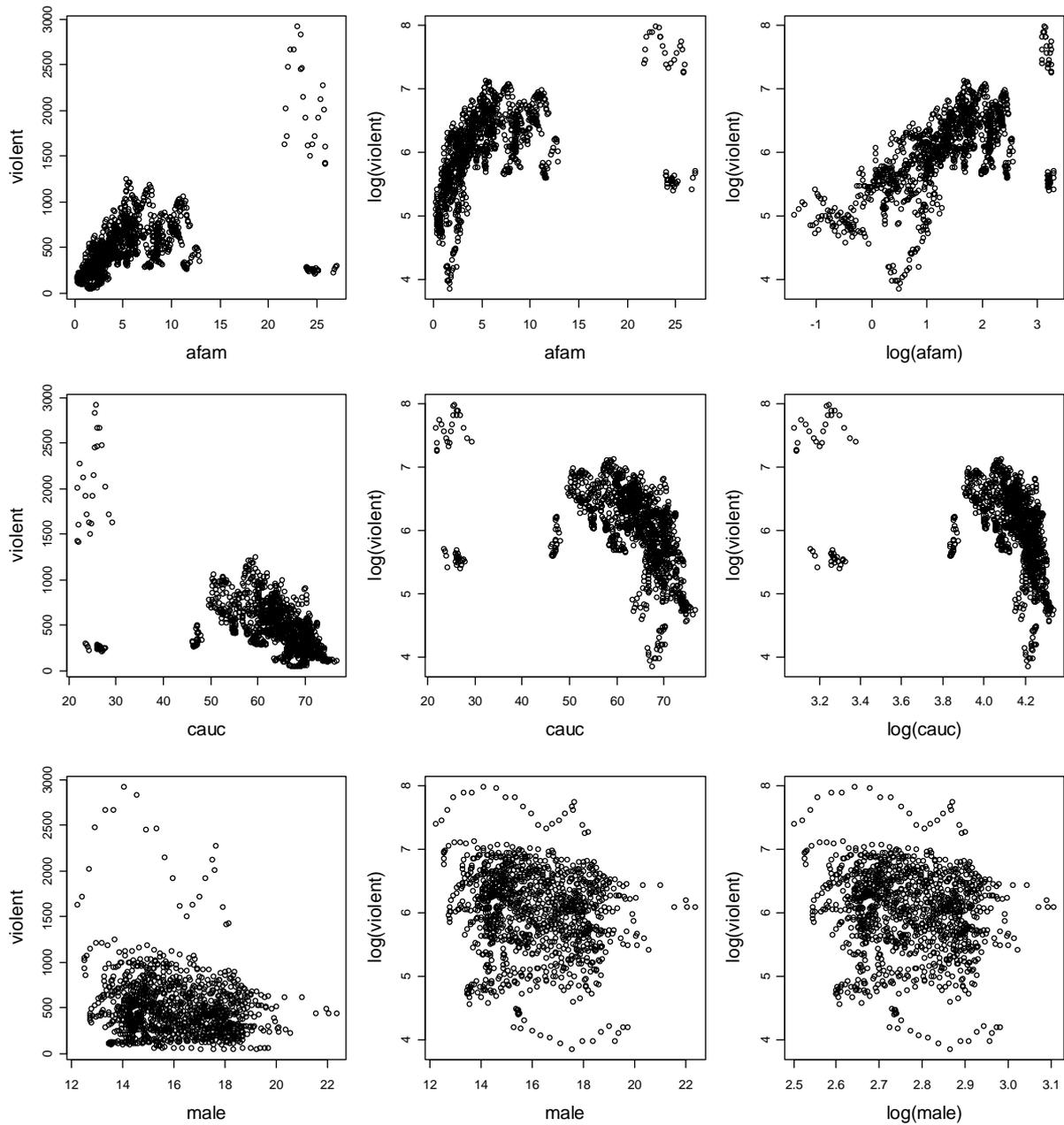
# # Graphische Abbildungen



law

law





### # Modellschätzung

```
model <- lm(log(violent) ~ law + log(prisoners) + log(density) +
log(income)+ log(afam) + log(cauc) + log(male), data=guns)
```

```
summary(model)
```

Coefficients:

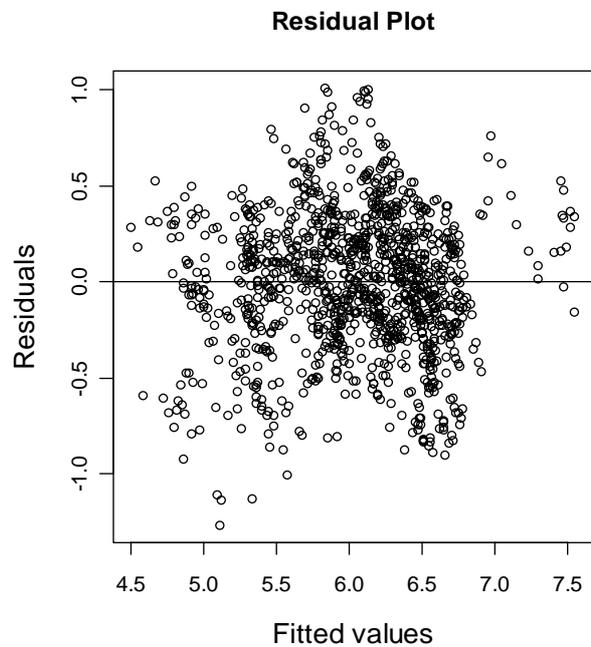
|                | Estimate  | Std. Error | t value | Pr(> t ) |     |
|----------------|-----------|------------|---------|----------|-----|
| (Intercept)    | -9.825639 | 1.072637   | -9.160  | < 2e-16  | *** |
| law            | -0.222554 | 0.027829   | -7.997  | 3.05e-15 | *** |
| log(prisoners) | 0.560291  | 0.028190   | 19.875  | < 2e-16  | *** |
| log(density)   | 0.116401  | 0.008241   | 14.124  | < 2e-16  | *** |
| log(income)    | 0.467496  | 0.075151   | 6.221   | 6.88e-10 | *** |
| log(afam)      | 0.271832  | 0.023972   | 11.340  | < 2e-16  | *** |
| log(cauc)      | 0.877869  | 0.082198   | 10.680  | < 2e-16  | *** |
| log(male)      | 1.754429  | 0.155040   | 11.316  | < 2e-16  | *** |

---

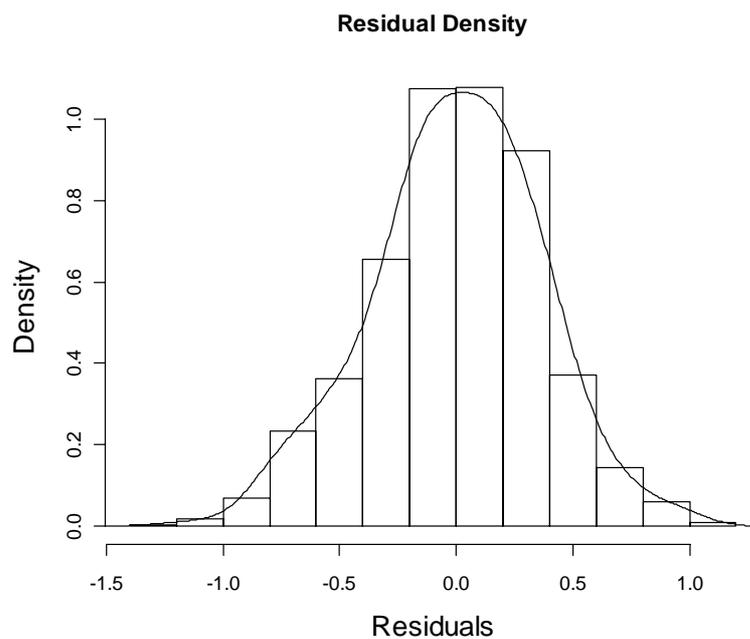
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3649 on 1165 degrees of freedom  
Multiple R-squared: 0.6825, Adjusted R-squared: 0.6806  
F-statistic: 357.8 on 7 and 1165 DF, p-value: < 2.2e-16

```
plot(model$fitted.values, model$residuals, xlab="Fitted values",  
ylab="Residuals", main="Residual Plot")
```



```
hist(model$residuals, freq=FALSE, xlab="Residuals", ylab="Density",  
main="Residual Density")  
lines(density(model$residuals, kernel="gaussian", adjust=1.5))
```



## Lösung zu Aufgabe 2

(a)

$$\bar{x} = \frac{194.2}{66} = 2.9424 \quad , \quad \bar{y} = \frac{174.32}{66} = 2.6412$$

$$\hat{\beta}_1 = \frac{\sum_{v=1}^n x_v y_v - n \bar{x} \bar{y}}{\sum_{v=1}^n x_v^2 - n \bar{x}^2} = \frac{531.894 - 66 \cdot 2.9424 \cdot 2.6412}{597.66 - 66 \cdot 2.9424^2} = 0.72293$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 2.6412 - 0.72293 \cdot 2.9424 = 0.51404$$

$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}}}{n-p-1} = \frac{1}{n-p-1} \sum_{v=1}^n \hat{\varepsilon}_v^2 = \frac{3.0428}{64} = 0.04754$$

(b)

$$TSS = \sum_{v=1}^n (y_v - \bar{y})^2 = \sum_{v=1}^n y_v^2 - n \bar{y}^2 = 477.1734 - 66 \cdot 2.6412^2 = 16.76153$$

$$RSS = \sum_{v=1}^n \hat{\varepsilon}_v^2 = 3.0428$$

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{3.0428}{16.76153} = 0.81847$$

(c)

$$\hat{\sigma}_{\hat{\beta}_1}^2 = \frac{\hat{\sigma}^2}{\sum_{v=1}^n (x_v - \bar{x})^2} = \frac{\hat{\sigma}^2}{\sum_{v=1}^n x_v^2 - n \bar{x}^2} = \frac{0.04754}{597.66 - 66 \cdot 2.9424^2} = \frac{0.04754}{26.25063} = 0.00181 \quad ,$$

$$\hat{\sigma}_{\hat{\beta}_1} = 0.04257$$

$$\left[ \hat{\beta}_1 - t_{\left[1-\frac{\alpha}{2}; n-p-1\right]} \cdot \hat{\sigma}_{\hat{\beta}_1} \quad , \quad \hat{\beta}_1 + t_{\left[1-\frac{\alpha}{2}; n-p-1\right]} \cdot \hat{\sigma}_{\hat{\beta}_1} \right]$$

$$\approx [0.72293 - 2.0086 \cdot 0.04257 \quad , \quad 0.72293 + 2.0086 \cdot 0.04257] = [0.6374 \quad , \quad 0.8084]$$

mit  $t_{\left[1-\frac{\alpha}{2}; n-p-1\right]} = t_{[0.975; 64]} \approx t_{[0.975; 50]} = 2.0086$  (vgl. Tabelle der  $t$ -Verteilung)

[Näherungslösung, da  $t$ -Verteilung nur bis 50 Freiheitsgrade tabelliert ist]

**(d)**

Punktprognose für  $Y_0$

$$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_0 = 0.51404 + 0.72293 \cdot 2.0 = 1.9599$$

Varianz des Prognosefehlers  $\hat{Y}_0 - Y_0$

$$\hat{\sigma}_{\hat{Y}_0 - Y_0}^2 = \hat{\sigma}^2 \cdot \left( 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{v=1}^n x_v^2 - n\bar{x}^2} \right) = 0.04754 \cdot \left( 1 + \frac{1}{66} + \frac{(2.0 - 2.9424)^2}{26.25063} \right) = 0.050$$

Standardabweichung des Prognosefehlers  $\hat{Y}_0 - Y_0$

$$\hat{\sigma}_{\hat{Y}_0 - Y_0} = 0.224$$

95% - Prognoseintervall

$$[\hat{y}_0 - t_{[1-\frac{\alpha}{2}; n-p-1]} \cdot \hat{\sigma}_{\hat{Y}_0 - Y_0}, \hat{y}_0 + t_{[1-\frac{\alpha}{2}; n-p-1]} \cdot \hat{\sigma}_{\hat{Y}_0 - Y_0}]$$

$$\approx [1.9599 - 2.0086 \cdot 0.224, 1.9599 + 2.0086 \cdot 0.224] = [1.510, 2.410]$$

mit  $t_{[1-\frac{\alpha}{2}; n-p-1]} = t_{[0.975; 64]} \approx t_{[0.975; 50]} = 2.0086$  (vgl. Tabelle der  $t$ -Verteilung)

[Näherungslösung, da  $t$ -Verteilung nur bis 50 Freiheitsgrade tabelliert ist]

**(e)**

Interpretation ... selbst

### **Lösung zu Aufgabe 3**

**(a-c)**

Selbst ...

**(d-e)**

Siehe Eckey, Kosfeld & Dreger, Ökonometrie, Gabler Verlag 2004, Seite 42ff und Seite 98 ff

### **Lösung zu Aufgabe 4**

Siehe Eckey, Kosfeld & Dreger, Ökonometrie, Gabler Verlag 2004, Seite 35 ff