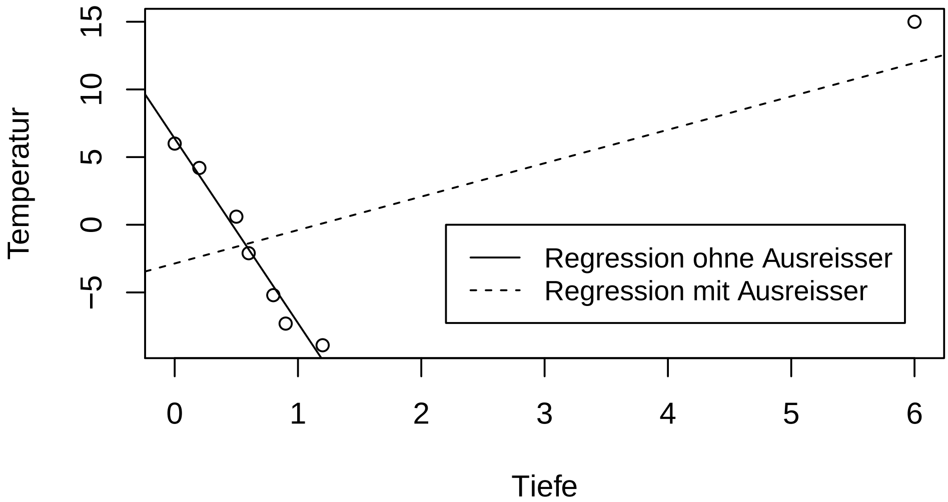
Musterlösung zu Serie 11

1. a) Aus dem Streudiagrammm sieht man, dass die ersten 7 Punkte sehr gut durch eine Gerade beschrieben werden. Der letzte Punkt liegt hingegen völlig ausserhalb der Geraden.



Es kann sich um einen groben Fehler handeln; es ist aber auch möglich, dass das lineare Modell zur Beschreibung des Zusammenhangs zwischen Tiefe und Temperatur nicht geeignet ist. (Zum Beispiel könnte der Zusammenhang quadratisch sein, oder stückweise linear. Weitere Möglichkeiten sind denkbar, z.B. Hinweis auf heisse Quelle, etc.)

b) Sei X die Tiefe und Y die Temperatur. Empirische Korrelation:

$$\rho_{\mathrm{X,Y}} = \frac{\sum_{i=1}^{7}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{7}(x_i - \overline{x})^2 \cdot \sum_{i=1}^{7}(y_i - \overline{y})^2}} = -0.99$$

wobei $\overline{x}$ und $\overline{y}$ auch ohne den Ausreisser zu berechnen sind.
Ohne Ausreisser sind Tiefe und Temperatur sehr stark negativ korreliert, mit dem Ausreisser aber positiv (0.6)!

c) In der obigen Figur sind die beiden Regressionsgeraden exakt eingetragen: der Ausreisser bewirkt, dass sich die Gerade fast um 90 dreht, d.h. die kleinste Quadrate Regressionsgerade ist sehr anfällig auf Ausreisser!
Die Schätzungen der Koeffizienten lauten ohne Ausreisser:

$$\begin{aligned}\widehat{\beta_1} &= \frac{\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{(6 + 1.81)(0 - 0.6) + \ldots + (-8.9 + 1.81)(1.2 - 0.6)}{(0 - 0.6)^2 + \ldots + (1.2 - 0.6)^2} \\ &= -13.64\end{aligned}$$

und

$$\widehat{\beta_0} = \overline{y} - \widehat{\beta_1}\overline{x} = -1.81 - (-13.64) * 0.6 = 6.37.$$

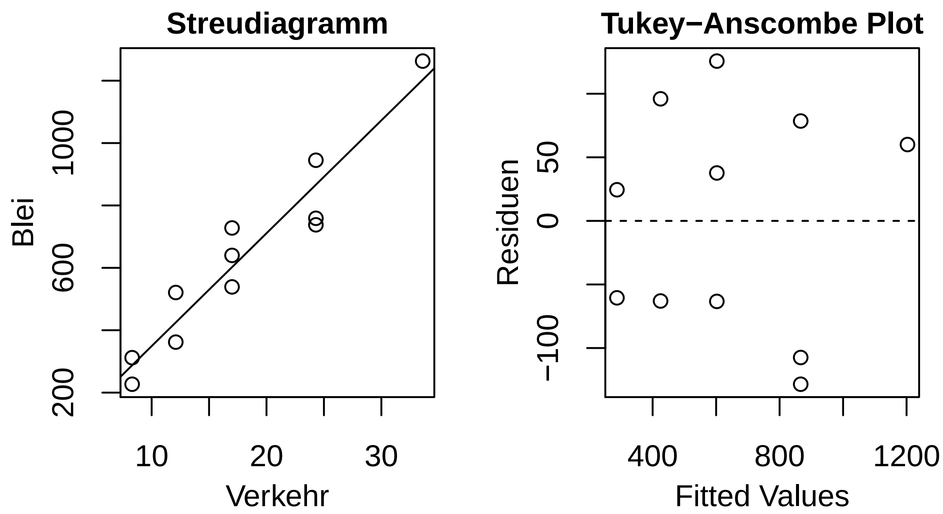
Mit Ausreisser ergibt sich: $\widehat{\beta_1} = 2.47$ und $\widehat{\beta_0} = -2.86$.
Diese Schätzungen können auch im **R**-output abgelesen werden.

d) ohne Ausreisser:
Nullhypothese H_0: $\beta_1 = 0$; Alternative H_A: $\beta_1 \neq 0$.
Teststatistik: $T = \frac{\widehat{\beta_1}-0}{\widehat{\sigma}(\widehat{\beta_1})} = \frac{-13.64}{1.01} = -13.5$.
5%-Verwerfungsbereich: $\mathcal{K} = \{T : |T| > t_{7-2,\,0.975}\} = \{T : |T| > 2.57\}$
Testergebnis: H_0 wird verworfen; die Steigung kann nicht 0 sein.

Mit Ausreisser ergibt sich $\widehat{\beta_1} = 2.47$ und $\widehat{\sigma}(\widehat{\beta_1}) = 1.33$. Dies entspricht einem p-Wert von 0.112 und H_0 kann nicht verworfen werden.
Die entsprechenden Outputs erhält man in **R** mit dem Befehl `summary`. Also z.B.

```
> summary(fit1)

Call:
lm(formula = Temperatur ~ Tiefe)

Residuals:
    Min      1Q  Median      3Q     Max
-9.0023 -4.9011  0.7531  3.9237  8.8616

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -2.862      2.950  -0.970    0.370
Tiefe          2.470      1.327   1.861    0.112

Residual standard error: 6.836 on 6 degrees of freedom
Multiple R-squared: 0.366,        Adjusted R-squared: 0.2603
F-statistic: 3.463 on 1 and 6 DF,  p-value: 0.1121
```

2. a)



b)

```
> summary(strasse.fit)

Call:
lm(formula = blei ~ verkehr, data = strasse)

Residuals:
    Min      1Q  Median      3Q     Max
-128.43  -63.13   24.52   69.32  125.72

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -12.842     72.143  -0.178    0.863
verkehr       36.184      3.693   9.798 4.24e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 92.19 on 9 degrees of freedom
```

```
Multiple R-squared: 0.9143,       Adjusted R-squared: 0.9048
F-statistic: 96.01 on 1 and 9 DF,  p-value: 4.239e-06
```

Wir erhalten die Schätzungen $\widehat{\beta}_0 = -12.842$, $\widehat{\beta}_1 = 36.184$ und $\widehat{\sigma}_e = 92.19$ (residual standard error).

$$\Longrightarrow Blei = -12.842 + 36.184 \cdot Verkehr$$

c) **t-Test** für β_1: Nullhypothese H_0: $\beta_1 = 0$, Alternative H_A: $\beta_1 \neq 0$

Teststatistik: $T = (\widehat{\beta}_1 - 0)/\widehat{\sigma}(\widehat{\beta}_1)$. Unter H_0 gilt: $T \sim t_{n-2}$, also hier $T \sim t_9$

Kritischer Bereich: Tabelle: $t_{9,0.975} = 2.26$ somit $\mathcal{K} = \{|T| > 2.26\}$.

Wert der Statistik: $T = 36.184/3.693 = 9.8 > 2.26$ (vgl. Computer-Output).
Die Steigung β_1 ist stark signifikant von Null verschieden (`P-Wert` $= 4.24e - 06$).

d) $x = 40 : y = -12.842 + 36.184 \cdot 40 = 1434.518$

3. 1) a
2) a
3) b
4) d
5) a
6) c
7) c