Musterlösung zu Serie 11

1. **a)** Der Test ist gepaart, da zu jedem Läufer genau ein Wert mit jedem Schuhtyp vorliegt und die beiden Werte eines Läufers von dessen Lauffähigkeiten abhängen.

b) 1. **Modell:** $X_1 - Y_1, \ldots, X_n - Y_n$ i.i.d. $\mathcal{N}(\mu_{X-Y}, \sigma^2_{X-Y})$,
σ_{X-Y} wird durch $\widehat{\sigma_{x-y}}$ geschätzt.

2. **Nullhypothese:** $H_0: \ \mu_{X-Y} = \mu_0 = 0$.
Alternative: $H_A: \ \mu_{X-Y} < 0$.

3. **Teststatistik:**

$$T = \frac{\sqrt{n}(\bar{X}_n - \bar{Y}_n - \mu_0)}{\widehat{\sigma_{X-Y}}}.$$

Verteilung der Teststatistik unter H_0: $T \sim \ t_{n-1}$.

4. **Signifikanzniveau:** $\alpha = 5\%$.

5. **Verwerfungsbereich für die Teststatistik:**

$$K = (-\infty, -t_{9;0.95}) = (-\infty, -1.833)$$

6. **Testentscheid:**

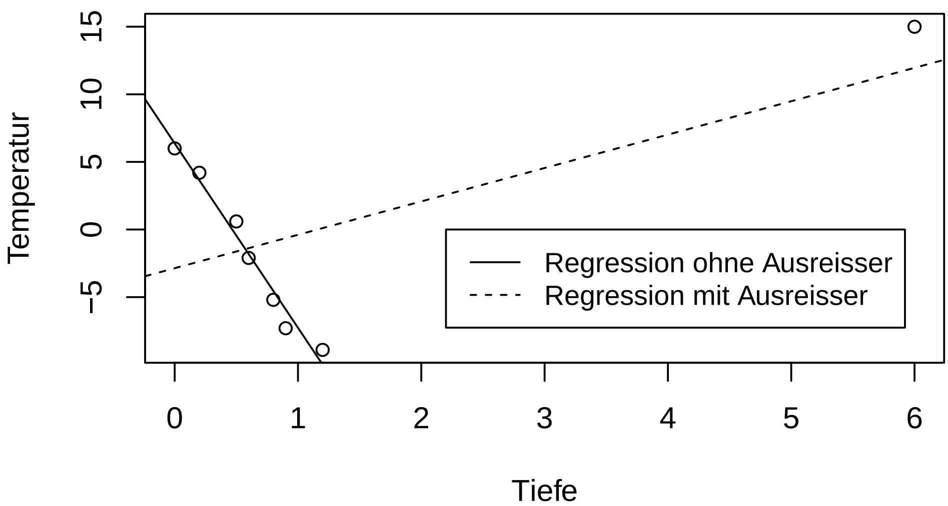
$$t = \frac{\sqrt{n}(\bar{x} - \bar{y})}{\hat{\sigma}_{x-y}} = \frac{\sqrt{10}(46.02 - 46.24)}{0.26} = -2.68.$$

Der Wert t der Teststatistik liegt im Verwefungsbereich, d.h. "SpeedShoe" verhilft zu besseren Leistungen.

c) Das Vertrauensintervall erstreckt sich bis $t_{9;0.95}/\sqrt{10}$-mal die Schätzung der Standardabweichung oberhalb der Schätzung des Durchschnitts $\mu_x - \mu_y$. D.h. es ist

$$\begin{aligned} & (-\infty, \bar{x} - \bar{y} + \hat{\sigma}_{x-y} \cdot t_{9;0.95}/\sqrt{10}) \\ = \ & (-\infty, -0.22 + 0.26 \cdot 1.833/\sqrt{10}) = (-\infty, -0.069) \ . \end{aligned}$$

2. **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.