

Übungsblatt LE1

Lösungen

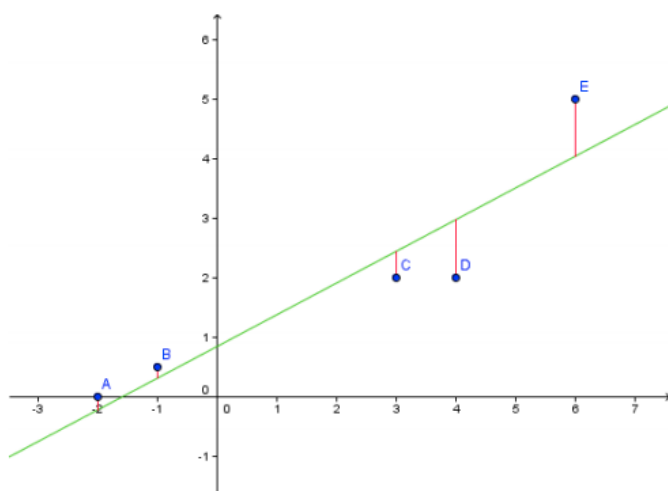
Lösung 1.

- (a) kategorial
- (b) kontinuierlich. Kann kategorial sein, wenn man nur Jahre oder Altersklassen berücksichtigt.
- (c) kontinuierlich. Kann kategorial sein, wenn man nur in Gewichtsklassen einteilt.

Lösung 2.

(a) $A = \begin{pmatrix} -2 & 1 \\ -1 & 1 \\ 3 & 1 \\ 4 & 1 \\ 6 & 1 \end{pmatrix}, x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, b = \begin{pmatrix} 0 \\ 0.5 \\ 2 \\ 2 \\ 5 \end{pmatrix}, .$

(b) $A^T A = \begin{pmatrix} 66 & 10 \\ 10 & 5 \end{pmatrix}, (A^T A)^{-1}(A^T b) = \begin{pmatrix} 0.53 \\ 0.83 \end{pmatrix}$



(c)

Lösung 3.

Die Lösung ist als jupyter notebook verfügbar, wobei es natürlich auch andere Lösungswege gibt.

- (a) Dies ist mittels `data.drop` lösbar.
- (b) Scatterplots zeigen klar, dass die Absenzen vom Alter abhängen. Die Jahre im Betrieb sind nicht relevant.
- (c) Ab einem Alter von ca. 25 Jahren erwartet man, dass Mitarbeiter fehlen. Zwischen 25-50 gibt es trotzdem Mitarbeiter, welche so gut wie nie fehlen. Dies verursacht die scharfe Kante, da die Vorhersage für ihr Fehlen deutlich höher ist. Ab einem Alter von ca. 50 fehlen alle, d.h. die Residuen werden kleiner.
- (d) Im Fit des Alters sind die Bedingungen einigermaßen erfüllt, wobei man je nach gewünschter Genauigkeit auch sagen könnte, sie seien nicht erfüllt. Die Residuen sind normalverteilt, aber die Tukey Ascombe Diagramme zeigen eine Streuung, die nicht ganz zufällig ist. Im Fit der Dienstjahre sind die Bedingungen nicht erfüllt, da die Residuen nicht normalverteilt sind.
- (e) Alter: $R^2 = 0.65$, was nicht perfekt ist, aber genügend. Dienstjahre: $R^2 = 0.0005$, was auf einen sehr schlechten Fit und ein nicht brauchbares Modell hindeutet.
- (f) Die Firma sollte genauer untersuchen, warum das Alter eine grosse Rolle bei den Absenzen spielt. Da Dienstjahre keinen Einfluss haben, ist es wahrscheinlich kein "Burnout", sondern es könnten zum Beispiel gesundheitliche Probleme sein.

Lösung 4.

Wir berechnen die Odds bei beiden Fachexperten: $45/16 = 2.8$ und $21/15 = 1.4$. Das Odds Ratio, also der Vergleich der Chancen ist $2.8/1.4 = 2.0$. Somit waren die Chancen beim Fachexperten A doppelt so hoch zu bestehen. Allerdings ist dies nur eine Stichprobe mit relativ wenigen Teilnehmern. Deswegen sollte man zusätzlich noch Konfidenzintervalle berechnen, die in LE3 vorgestellt werden.

Lösung 5.

Die Lösung ist als jupyter notebook verfügbar, wobei es natürlich auch andere Lösungswege gibt.