

Aufgabenserie 3 zur Lehrveranstaltung "Computergestützte Datenanalyse"

1. Gegeben ist der Zufallsvektor $\vec{X} \sim \mathcal{N}_3(\vec{\mu}, \Sigma)$ mit

$$\vec{\mu} = \begin{pmatrix} -3 \\ 1 \\ 4 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

- a) Geben Sie die Verteilung des Vektors $\vec{Y} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ sowie die (Rand-)Verteilungen von X_1 und X_2 an! Wie groß ist die Korrelation zwischen X_1 und X_2 ? Formulieren Sie eine Aussage zur Abhängigkeit von X_1 und X_2 .
- b) Formulieren Sie eine Aussage zur Abhängigkeit von X_1 und X_3 bzw. X_2 und X_3 .

2. Gegeben ist der Zufallsvektor $\vec{X} \sim \mathcal{N}_2(\vec{\mu}, \Sigma)$ mit

$$\vec{\mu} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 16 & 11 \\ 11 & 9 \end{pmatrix}.$$

- a) Stellen Sie die Dichte von \vec{X} mit Hilfe von Mathematica dar. Wie groß ist die Korrelation zwischen X_1 und X_2 ?
- b) Ermitteln Sie die Verteilung des dreidimensionalen Zufallsvektors \vec{Z} mit

$$Z_1 = X_1 + 2, \quad Z_2 = -X_1 + 2X_2, \quad Z_3 = 4X_2 - 4.$$

Wie groß ist die Korrelation zwischen Z_1 und Z_2 ? Stellen Sie die Dichte von $(Z_1, Z_2)^T$ mit Hilfe von Mathematica dar.

3. Betrachten Sie die Datensätze "kohlefaserrohr", "stahl", "dampfturbine", "festigkeit". Bestimmen Sie die Mittelwerte, empirischen Varianzen, empirische Standardabweichungen und die empirischen Korrelationen der Variablen untereinander. Welche Korrelationen sind signifikant und welche nicht? Lassen Sie sich die entsprechenden Streubilder ausgeben. Untersuchen Sie ob die Merkmale normalverteilt sind. Welche Schlussfolgerungen können für die multivariate Verteilung gezogen werden?

4. Analysieren Sie die Datensätze "ernte" (Zielgröße: Ertrag), "abrieb" (Zielgröße: ma),

"basboeden" (Zielgröße: hoehe oder lsar), "cholestyramin" (Zielgröße: senkung), "tank" (Zielgröße: druck), "gussteile" (Zielgröße: g), "gewebe" (Zielgröße: abnutz), "pinien" (Zielgröße: hoehe) und "timber" (Zielgröße: festm oder elastm) mit Hilfe der Einfachregression. Wie gut passt sich das jeweilige Modell an die Daten an (Residuenplot)?

Aufgaben 3,4 → Belegthema 2