

Arbeitsblatt 3

Einfache lineare Regression (Diagnostik & Erweiterungen)

Aufgabe 1 (Windmühle)

Ein Entwicklungsingenieur untersucht den Einsatz von einer Windmühle zur Elektrizitätserzeugung. Dabei sammelte er unter anderem auch Daten über die Stromgewinnung (in Ampère) durch die Windmühle bei entsprechenden Windgeschwindigkeiten (in Metern pro Sekunde). Die Daten sind im File `Windmuehle.dat`.

(a)

Wir betrachten zuerst das Modell

$$\text{Strom}_i = \alpha + \beta \cdot \text{Windgeschwindigkeit}_i + E_i$$

Passen Sie ein entsprechendes Regressionsmodell an und prüfen Sie mittels grafischer Residuenanalyse, ob die Voraussetzungen erfüllt sind.

(b)

Wenden Sie die First-Aid Transformationen auf die erklärende Variable und die Zielvariable an, passen Sie das Regressionsmodell mit den transformierten Grössen an und beurteilen Sie grafisch, ob die Voraussetzungen nun erfüllt sind.

(c)

Aus der Theorie ist folgender funktionaler Zusammenhang bekannt:

$$\text{Strom} \approx \alpha + \beta \cdot \frac{1}{\text{Windgeschwindigkeit}}$$

Passen Sie dieses Modell aus der Fachtheorie mittels einer Regression an (d.h. **Strom** als Zielgrösse und $x = 1/\text{Windgeschwindigkeit}$ als erklärende Variable) und stellen Sie das Regressionsmodell in einem Streudiagramm dar.

(d)

Was bedeuten die beiden Parameter α und β im Modell aus der Fachtheorie?

(e)

Prüfen Sie mittels Residuenanalyse, ob die Voraussetzung für das Modell aus der Fachtheorie erfüllt sind.

(f)

Gibt es einen signifikanten Zusammenhang zwischen Windgeschwindigkeit und der Stromgewinnung aufgrund des Modell aus der Fachtheorie? Führen Sie einen geeigneten Test auf dem 5% Niveau durch.

(g)

Was sind plausible Werte für die maximale Stromgewinnung der Windmühle? Geben Sie ein 95% Vertrauensintervall an.

(h)

Machen Sie eine Vorhersage, wie gross die erwartete Stromproduktion bei einer Windgeschwindigkeit von 15 m/s ist. Geben Sie ein 95%-Prognoseintervall an.

Aufgabe 2 - Highway in Texas

Auf der Strasseoberfläche akkumulieren sich in trockenen Zeiten sehr viele Schmutzstoffe. Diese werden während dem Regen zum grössten Teil abgewaschen. In Austin (Texas) wurde an einer bestimmten Stelle am Highway die Abflussmenge bei Regen bestimmt. Die Forscher wollen nun den Zusammenhang zwischen dem Regenfallvolumen und der Abflussmenge bestimmen. Daraus können sie dann vorhersagen, wie gross die Menge an Schadstoffen ist, die in die Natur gelangt. Die gemessenen Daten befinden sich im File `highway.csv`.

(a)

Passen Sie eine Regressionsgerade (`runoff` als Zielgrösse und `rain` als erklärende Grösse). Visualisieren Sie die geschätzte Gerade.

(b)

Welcher Anteil der beobachteten Variation in der Abflussmenge kann mit dem einfachen linearen Regressionsmodell erklärt werden?

(c)

Besteht ein auf 5% signifikanter linearer Zusammenhang zwischen Abflussmenge und Regenfallvolumen? Geben Sie auch die anschauliche Interpretation des Regressionskoeffizienten an.

(d)

Überprüfen Sie die für das Regressionsmodell getroffenen Annahmen. Können wir Ihren Schlussfolgerungen aus (c) vertrauen?

(e)

Warum könnte hier eine Logarithmus Transformation der beiden Variablen weiterhelfen?

(f)

Passen Sie eine Log-Log Regression an. Zeichnen Sie die neue Regressionsbeziehung in ein Streudiagramm auf der Originalskala ein. Prüfen Sie erneut die Stärke des linearen Zusammenhangs und die Signifikanz zwischen Abflussmenge und Regenfallvolumen auf dem 5% Niveau.

(g)

Berechnen Sie eine Vorhersage, wie gross die erwartete Abflussmenge bei einer Regenfallmenge von 50 ist. Erzeugen Sie ein 95%-Prognoseintervall für beliebige Regenfallmengen und zeichnen dieses als sogenanntes Prognoseband in das Streudiagramm ein.

(i)

Führen Sie eine Residuenanalyse für das log-log Modell durch. Beantworten Sie anschliessend, welches der beiden Modelle Ihnen besser geeignet scheint, und begründen Sie Ihre Antwort.