

Aufgabenserie 4 zur Vorlesung "Computergestützte Datenanalyse"

1. Mit Hilfe der Mehrfachregression sind die Datensätze "ammoniak" (Zielgröße: loesl), "stahlherst" (Zielgröße: stahl), "speiseeis" (Zielgröße: eisverb), "asphalt" (Zielgröße: y, Variable v logarithmieren), "stahl" (Zielgröße nicht festgelegt), "kfett" (Zielgröße: brozek oder siri), "chemfaesser" (Zielgröße: entlad), "chemreakt" (Zielgröße: umsatz), "wohnungen" (Zielgröße: steuern), "fastfood" (Zielgröße: margin), "huehnereier" (Zielgröße: v), "loeslalkohol" und "festigkeit" zu analysieren. Wie gut passt sich das jeweilige Modell an die Daten an? Entwickeln Sie ein Modell, das nur mit relevanten Variablen auskommt. Erstellen Sie einen Residuenplot. Analysieren Sie damit, ob die Modellvoraussetzungen (Linearität, Homoskedastizität, normalverteilte Residuen) erfüllt sind.

2. Analysieren Sie die Datensätze "pinien" (Zielgröße hoehe, Gewichtsvariable alter), "alaskapipel" (Zielgröße tiefefeld), "rauschfilter" (Zielgröße HalfVarDelta, Regressor/Gewichtsvariable mean), "felsbloecke" (Zielgröße bruchkraft, Regressor-/Gewichtsvariable sunterst ≥ 1), "esterase" (Zielgröße Zählimpulse), "diamanten2" (Zielgröße preis) und "schokolwerb" (Zielgröße posantw, Gewichtsvariable aufw) unter Verwendung der gewichteten Regression.

zur Beachtung: Es werden nur ganze Werte der Gewichtsvariablen berücksichtigt, ggf. sind die Werte dieser Variablen mit einer Zehnerpotenz zu multiplizieren.

3. Mit Hilfe der polynomialen Regression sind die Datensätze "windmill", "thermoelektrik1" (eine Temperatur wählen: $v_2 = 150 | v_2 = 175 | v_2 = 200 \dots$), "anhalteweg", "bonions", "fichten", "bohne", "kupferausd", "siedepunkt" und "gewebe" zu analysieren. Welcher Grad des Polynoms führt zu einem passenden Modell für die Daten?

Aufgabe 1 \rightarrow Belegthema 2