

Statistische Verfahren WS 2015

Projekt 6 - Buche

Problemstellung:

Herleitung einer generischen Biomassefunktion für die Buche aus analysierten Biomassedaten aus unterschiedlichen Beständen.

Datensätze: (buche.csv)

Der Datensatz „buche.csv“ enthält Daten zur oberirdischen Biomasse von Buchen aus unterschiedlichen, über Mitteleuropa verteilten Studien (zusammengetragen von Th. Wutzler, MPI Biogeochemie Jena). Geeignete lineare Modelle zur Vorhersage der Biomasse können durch Logarithmieren sogenannter allometrischer Funktionen gewonnen werden.

$$y = \beta_1 x_2^{\beta_2} \dots x_k^{\beta_k}$$

Die im Datensatz enthaltenen Variablen sind:

- author – Autor der entsprechenden Studie
- hsl – Höhe über NN (height above sea level)
- age – Alter des Baumes
- dbh – Brusthöhendurchmesser (diameter at breast height)
- height – Höhe des Baumes
- biom – gesamte oberirdische Biomasse des Baumes

Aufgaben:

- Leiten Sie zunächst für die Teildatensätze jedes einzelnen Autors separate Biomassefunktionen her!
- Leiten Sie dann ein gemeinsames lineares Modell zur Vorhersage der oberirdischen Biomasse der Buche mit dem Autor als kategorieller Einflussgröße her! Diskutieren Sie vor allem das Vorliegen von Wechselwirkungen mit der Einflussgröße Autor!
- Eine Alternative zur Verwendung eines linearen Modells nach logarithmischer Transformation wäre die Verwendung eines verallgemeinerten linearen Modells mit der Annahme einer Gammaverteilung der Zielgröße und logarithmischer Link-Funktion. Vergleichen Sie für den vom Autor Joosten stammenden Teildatensatz die Genauigkeit der Schätzwerte für die erwartete Biomasse der beiden Modelle:

$$(M1) \quad E \log(Y_i) = \mu_i = \beta_1 + \beta_2 \log(dbh_i), \quad Y_i \sim N(\mu_i, \sigma^2)$$

$$(M2) \quad \log(EY_i) = \log \mu_i = \beta_1 + \beta_2 \log(dbh_i), \quad Y_i \text{ gammaverteilt mit Erw.-Wert } \mu_i$$

Simulieren Sie Pseudobeobachtungen der Zielgröße, die dem Modell M2 entsprechen. Als Designmatrix wählen Sie dabei zufällig ausgewählte Zeilen der ursprünglichen Designmatrix (evtl. mit Wiederholung). Vergleichen Sie die Genauigkeit der Schätzung von β_2 basierend auf Modell M2 (dem „wahren“ Modell) und basierend auf Modell M1 (einem „falschen“ Modell) für unterschiedliche Stichprobenumfänge.