

# Statistische Verfahren WS 2019

## Projekt 1 – Produktivität von Frischwiesen

### Problemstellung:

Einfluss von Bodenfaktoren und Artenzahl auf die Produktivität von Frischwiesen.

### Datensatz: (frischwiesen.csv)

Der Datensatz enthält Daten zur Produktivität von Frischwiesen im Saale- und Ilmtal (Christiane Roscher, FSU Jena)

- Probe – Bezeichnung der Probefläche
- Gebiet – Saaletal/Ilmtal
- P – Phosphorgehalt des Bodens
- K – Kaliumgehalt des Bodens
- pH – pH-Wert
- Cges – Gesamtkohlenstoffgehalt
- Corg – Gehalt an organischem Kohlenstoff
- Corg/N – Kohlenstoff-/Stickstoff-Verhältnis
- Artenzahl – Gesamtzahl vorkommender Arten
- biom – Frischmasse der geernteten Biomasse

### Aufgaben zur Datenanalyse:

- Leiten Sie zunächst getrennt für das Saaletal und das Ilmtal geeignete lineare Modelle zur Prognose der Biomasseproduktion her.
- Analysieren Sie dann beide Teildatensätze gemeinsam und untersuchen Sie insbesondere das Vorliegen von Wechselwirkungen, d.h. unterschiedliche quantitative Effekte der Einflussgrößen in den beiden Untersuchungsgebieten.
- Vergleichen Sie die Genauigkeit der Vorhersage der Biomasse für das Saaletal basierend auf dem separaten und dem gemeinsamen Modell. Verwenden Sie dabei den auf geeignete Art geschätzten erwarteten Prognosefehler SPSE.

### Simulationsaufgabe:

- Untersuchen Sie in einer Simulationsstudie den Einfluss des Stichprobenumfangs auf die Güte der Modellwahl basierend auf Mallows Cp-Kriterium. Untersuchen Sie dabei insbesondere:
  - die relative Häufigkeit, mit der die „richtigen“ Prädiktoren ausgewählt werden
  - die Anzahl der ausgewählten Prädiktoren.
- Wählen Sie dazu ein „wahres Modell“ in Anlehnung an die Ergebnisse des ersten Teils und eine Designmatrix, die zufällig ausgewählte Zeilen der realen Design-Matrix (mit Wiederholung) enthält. Simulieren Sie dann mehrfach Pseudo-Beobachtungen der Zielgröße und führen Sie für die so simulierten Pseudo-Datensätze die Modellwahl mit Hilfe von Mallows Cp-Kriterium durch.