

---

## Arbeitsblatt 4

**A 1** Berechnen Sie in Ihrer Simulationsstudie aus Blatt 3 (A1) in jedem Simulationsschritt auch das AIC aus dem linearen und dem polynomialen Modell. Berechnen Sie dann den durchschnittlichen AIC-Wert für jedes der beiden Modelle jeweils für  $n=100$  und  $n=20$ . Wie würde Ihre Modellwahl anhand des AIC-Kriteriums ausfallen? Passt dieses Ergebnis zu Ihren geschätzten quadratischen Vorhersagefehlern?

**A 2** Laden Sie den Datensatz Prestige aus dem Paket car und betrachten Sie mittels geeigneter Plots die paarweisen Zusammenhänge der Variablen (die Variable census wird nicht weiter betrachtet und kann unberücksichtigt bleiben). Entfernen Sie alle Beobachtungen mit fehlenden Werten in einer der Variablen.

- a) Berechnen Sie anhand linearer Regressionsmodelle, inwieweit sich das Ansehen eines Berufs (prestige)
1. durch das Einkommen (income)
  2. durch die Länge der Ausbildung (education)
  3. durch die Berufsklasse (type)
  4. durch den prozentualen Frauenanteil Berufsklasse (women)
  5. durch Ausbildungslänge und Einkommen
  6. durch Ausbildungslänge und Berufsklasse
  7. durch Ausbildungslänge und Frauenanteil
  8. durch Ausbildungslänge, Einkommen und Berufsklasse
  9. durch Ausbildungslänge, Einkommen, Berufsklasse und Frauenanteil
- erklären lässt. Berechnen Sie für jedes der Regressionsmodelle den Trainingsfehler, den Optimismusterm, den erwarteten (in-sample) Testfehler und das AIC. Erläutern Sie Ihre Ergebnisse.
- b) Warum ist es für einen sinnvollen Vergleich der AICs notwendig, in allen Modellen alle Beobachtungen mit fehlenden Werten zu entfernen? Da fehlende Werte nur in der Variable type auftreten, hätte man diese Beobachtungen theoretisch auch nur bei der Berechnung der Modelle 3., 6., 8. und 9. entfernen können.