

Robuste Statistik

Blatt 12

Aufgabe 12.1: (10 Punkte)

Betrachten Sie den Datensatz `mrdata`, der auf der Homepage gegeben ist. Dieser Datensatz enthält in den ersten vier Spalten die Regressoren \mathbf{x}_1 , \mathbf{x}_2 , \mathbf{x}_3 , \mathbf{x}_4 . Die in der fünften Spalte gegebene Zielgröße \mathbf{y} kann mit folgender Regressionsgleichung beschrieben werden

$$y_n = \beta_0 + \sum_{i=1}^4 \beta_i x_{i,n} + e_n, \text{ für } n = 1, \dots, N. \quad (1)$$

Dabei ist $x_{i,n}$ vom i -ten Regressor der n -te Datenpunkt, $\beta \in \mathbb{R}^5$ ist der unbekannte Modellparameter und e_n sind die Fehlerterme, die Realisationen von unabhängig, identisch verteilten Zufallsvariablen entsprechen. Das wahre Regressionsmodell lässt sich durch den Parameter $\beta = (1000, -20, 5, 15, 10^{-10})^\top$ bis auf die Fehlerterme e_n mit der Regressionsgleichung (1) korrekt beschreiben. Allerdings sind zwei Einträge der Zielvariable \mathbf{y} in dem Datensatz nachträglich verfälscht worden.

- Versuchen Sie mit elementaren grafischen Methoden (z.B. durch Plots der Regressoren gegen die Zielvariable) Vermutungen aufzustellen, welche beiden Einträge in \mathbf{y} verfälscht worden sind. Nehmen Sie dabei an, dass es zwischen den Regressoren und der Zielgröße einen linearen Zusammenhang gibt.
- Schätzen Sie β mit der Kleinsten-Quadrat-Summen-Schätzung und vergleichen Sie Ihre Schätzung mit dem wahren Parameter.
- Stellen Sie die Residuen, die sich aus der Schätzung in b) ergeben, grafisch dar. Welche Daten lassen sich anhand dieser Grafik als Ausreißer detektieren?
- Schätzen Sie β mit der Regressions-M-Schätzung mit der Score-Funktion nach Hampel und vergleichen Sie Ihre Schätzung mit dem wahren Parameter.
- Stellen Sie die Residuen, die sich aus der Schätzung in d) ergeben, grafisch dar. Welche Daten lassen sich anhand dieser Grafik als Ausreißer detektieren?
- Liefern die Aufgabenteile a), c) und e) unterschiedliche detektierte Ausreißer?
- Entfernen Sie aus dem Datensatz die in e) als Ausreißer detektierten Datenpunkte und führen Sie auf diesem Datensatz erneut eine Kleinste-Quadrat-Summen-Schätzung für β durch. Wie sehen die Ergebnisse nun aus?

Abgabe bis spätestens 24.06.2019, 10 Uhr.