Methoden Verkehrsökonometrie für Master-Studierende

Winter semester 2021/22, Tutorial No. 4

Aufgabe 4.1: Über das Fitten von Elefanten

Gegeben ist die Messreihe (schwarze Punkte in den Abbildungen)

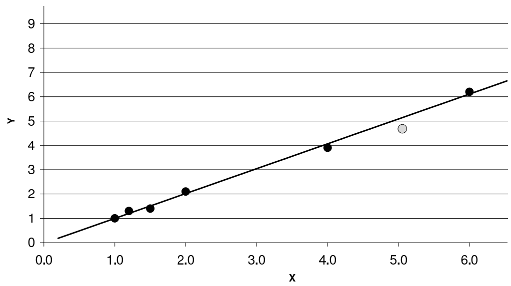
x_i	1.0	1.2	1.5	2.0	4.0	6.0
Y_i	1.0	1.3	1.4	2.1	3.9	6.2

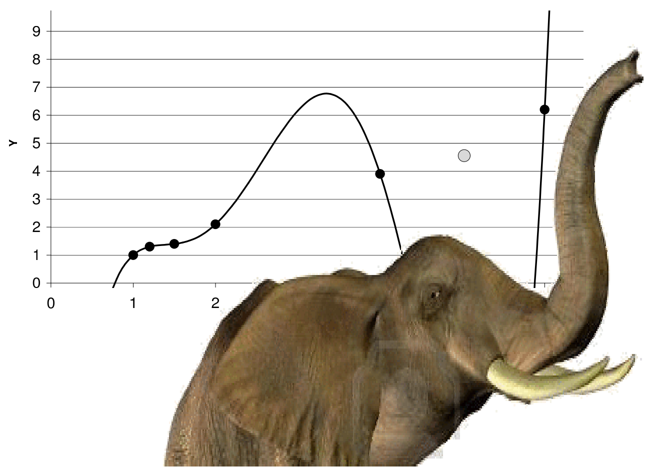
Zur Analse der Messreihe gibt es zwei Modellkandidaten: (i), einfache lineare Regression $Y = a + bx + \epsilon$, (ii) polynomische Regression $Y = \sum_{j=0}^{5} b_j x^j + \epsilon$.

(a) Sind die beiden Modelle jeweils linear, parameterlinear oder nichtlinear?

(b) Sind sie wohlspezifiziert? Falls nein, welche Spezifikationsbedingung ist verletzt?

(c) Die LSE-Schätzer $\hat{Y}_i$ der jeweiligen Modelle haben folgenden Verlauf:





Welches Modell hat die geringere Fehlerquadratsumme? Welches Modell beschreibt den Zusammenhang zwischen x und y besser?

Aufgabe 4.2: Regression mit nominalskalierten Variablen

Um zu analysieren ob und wieviel die Eisenbahn in den letzten Jahrzehnten schneller geworden ist, werden die gemessenen Durchschnittsgeschwindigkeiten für die langsameren Zugtypen (S-Bahnen, Regional-, Vorortszüge etc.) und die schnelleren Züge (IC, EC, ICE etc.) analysiert:

Jahr x_1 (seit 1 900)	0	30	60	60	80	80	100	100
Zugkategorie x_2	0	0	0	1	0	1	0	1
Geschw (km/h) y	45	60	65	90	86	105	83	130

Die Geschwindigkeitsentwicklung soll mit dem linearen Regressionsmodell

$$y(\vec{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

beschrieben werden. Dabei kann x_2 die Werte 0 und 1 (langsame bzw. schnelle Züge) annehmen. Vor 1960 gibt es diese Unterteilung allerdings nicht und die Züge wurden der langsameren Klasse zugeordnet.

(a) Betrachten Sie zunächst die funktionale Spezifikation: Wie sind die exogenen und endogenen Variablen skaliert? Enthält das Modell überflüssige exogene Variable? Fehlen exogene Faktoren? Kann Linearität angenommen werden? Kann Homogenität in den Daten angenommen werden oder gibt es einen Strukturbruch?

(b) Zeigen Sie anhand des Sachverhalts und der Daten der obigen Tabelle, dass das Modell bezüglich der Daten korrekt spezifiziert ist, also die exogenen Variablen keine Zufallsvariablen sind, keine Multi-Kolinearität vorliegt und der Stichprobenumfang aureicht.

(c) Geben Sie anschauliche Bedeutungen für die drei Parameter an. Welche Vorzeichen erwarten Sie für die geschätzen Werte von β_1 und β_2? (Begründung!)

(d) Nehmen Sie in dieser und der folgenden Teilaufgabe folgenden Schätzvektor und die zugehörige Varianz-Kovarianzmatrix an:

$$\hat{\vec{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 42.2 \\ 0.474 \\ 28.2 \end{pmatrix}, \quad \underline{\underline{\hat{V}}} = \begin{pmatrix} 45 & -0.56 & -0.21 \\ -0.56 & 0.0104 & -0.27 \\ -0.21 & -0.27 & 46 \end{pmatrix}.$$

Prognostizieren Sie (ohne Fehlerabschätzung) die erwarteten mittleren Geschwindigkeiten der langsameren und schnelleren Zugklassen im Jahr 2020.

(e) Bestimmen Sie nun die Standardabweichung des Fehlers der obigen Prognose. Berücksichtigen Sie dabei allgemeine Formeln über die Varianz von Linearkombinationen von Zufallsvariablen und die Varianz-Kovarianzmatrix $\underline{\underline{\hat{V}}}$.

(f) Berechnen Sie zu einer Fehlerwahrscheinlichkeit von 5 % das Konfidenzintervall von β_1. *Hinweis:* Verwenden Sie die obige Varianz-Kovarianzmatrix!

(g) Das Modell berücksichtigt nicht, dass sich die Geschwindigkeiten der langsameren und der schnelleren Züge mit unterschiedlichen Raten über die Zeit ändern können. Erweitern Sie das Modell (also die funktionale Modellspezifikation) so, dass Unterschiede erfasst werden können.

(h) Es sollen nun wieder gemeinsame Zeitabhängigkeitern, aber drei Zugtypen betrachtet werden, indem man bei den schnellen Zügen zwischen ICE und sonstigen schnellen Zügen unterscheidet. Wie berücksichtigt man das in der funktionalen Spezifikation?