Übungsblatt 11 zur Vorlesung

"Statistische Methoden"

Regression

Herausgabe des Übungsblattes: Woche 24, Abgabe der Lösungen: Woche 26 (bis Dienstag, 16.15 Uhr), Besprechung: Woche 26, Donnerstag, 13-15 od 16-18 Uhr

---

Must

Aufgabe 55 [K I; Erwartungswerte und Varianzen im multivariaten Fall]

Sei X ein n-dimensionaler Zufallsvektor mit Kovarianzmatrix Σ. Zeigen Sie (ausführliches Nachrechnen!), dass mit einer $k \times n$-Matrix A und konstantem Vektor $a \in \mathbb{R}^n$ gilt:

$$E[AX] = AE[X]$$

und

$$V[a^t X] = a^t \Sigma a.$$

Aufgabe 56 [K I; $\hat{\beta}$ ist erwartungstreu]

Zeigen Sie: der OLS-Schätzer für β ist in Modell (7.4) erwartungstreu.

Standard

Aufgabe 57 [K I; A wenn $k \in \{1, 2\}$] [1+2 Punkte]

a) Sei $A = \mathbf{1_n}$ eine Spalte aus lauter 1-ern. Untersuchen Sie unsere allgemeinen Schätzverfahren zu (7.4) und die Teststatistiken in diesem Spezialfall.

b) ["einfache Regression"] Sei $k = 2$ und $A_{.1} = \mathbf{1_n}$ eine Spalte aus lauter 1-ern. Untersuchen Sie unsere allgemeinen Schätzverfahren zu (7.4) und die Teststatistik für $\beta_2 = 0$ in diesem Spezialfall.

Aufgabe 58 [K II & K III; Finde die richtigen Koeffizienten] [6+6 Punkte]

a) Generieren Sie zuerst eine Realisation eines Vektors Y aus folgendem Modell: $n = 100$, $k = 4$ (inkl. Intercept), der erste nichttriviale Regressor ist $x_{i2} := i$, dann $x_{i3} := i^2$ und $x_{i4} := \sin(i)$, $1 \leq i \leq 100$. Die Parameter sind

$$Y_i = 33 + \mathrm{PN} x_{i2} + (\mathrm{PN} - 2) x_{i3} + e^{\mathrm{PN}} x_{i4} + \epsilon_i,$$

wo Sie $\sigma^2 = 23$ nehmen. Machen Sie (in Kenntnis der richtigen Werte von β und σ^2) eine Regressionsanalyse mit dem Top-Down-Verfahren und einem jeweiligen $\alpha = 0.05$. Nehmen Sie höchstens einen Regressor pro Mal weg.

b) Auf www.luchsinger-mathematics.ch/weissnichts.txt finden Sie einen Vektor Y, welcher mit den gleichen Regressoren erzeugt wurde wie in a). Sie kennen jetzt aber weder β noch σ^2. Machen Sie auch hier eine Regressionsanalyse mit dem Top-Down-Verfahren und einem jeweiligen $\alpha = 0.05$. Nehmen Sie höchstens einen Regressor pro Mal weg.

Honours

<u>Aufgabe 59 [K IV; Beispiel Versicherungswirtschaft]</u> [4+4+4+4 Punkte]

Auf www.luchsinger-mathematics.ch/hausrat.txt finden Sie aggregierte reale Daten mehrerer schweizerischer Versicherungsgesellschaften zur Hausratversicherung (ohne Grossereignisse wie zum Beispiel Sturm Lothar). Dabei ist V_i die versicherte Summe in CHF 1000 im Jahr i und Z_i der angefallene finanzielle Schaden in Jahr i.

a) Machen Sie eine gewöhnliche Regression in R, wobei Sie den sogenannten Schadensatz $X_i := Z_i/V_i$ mit der Variablen "Jahr" erklären:

$$X_i = \alpha + \beta i + \epsilon_i.$$

b) Machen Sie auch eine sogenannte gewichtete Regression (in der Vorlesung nicht behandelt). Suchen Sie in R mit dem Befehl "?lm" nach der Methode, mit der Sie Gewichte eingeben können. Dabei wollen wir eine gewichtete Regression derart machen, dass die Varianzen der Fehlerterme nicht als konstant angesehen werden. Wir werden im Gegenteil die Varianzen der Fehlerterme derart gewichten, dass

$$V[X_i] = \sigma^2/V_i.$$

Damit erhalten diejenigen Daten mehr Gewicht, welche mehr Volumen haben. Vergleichen Sie die Resultate aus a) und b).

c) Überlegen Sie sich theoretisch, wie die gewichtete OLS-Schätzung funktioniert. Benutzen Sie dazu Gleichung (7.4). Womit müssen Sie diese Gleichung multiplizieren, damit Sie dann wieder eine Gleichung haben, bei der die Varianz konstant ist? Was hat das für Auswirkungen auf die Schätzung von β?

d) Die Grösse X_i, welche wir in a) und b) zu erklären versuchen, hat eine grosse Bedeutung. Welche? Was muss wohl in der Praxis noch berücksichtigt werden?