



Übungsaufgabenblatt No. 4 zur Vorlesung Statistik für WIng

24. Mai 2024

Auf diesem Aufgabenblatt sind Präsenz- und Hausaufgaben gegeben. Die Präsenzaufgaben werden vom Übungsleiter vorgerechnet, die Hausaufgaben sind elektronisch am Tag der nächsten Übung (pünktlich !) abzugeben. Die Ergebnisse der Berechnungen sind als Excel-, die Dokumentation und Interpretation als PDF-Datei zuzusenden.

Anmerkung zur Bearbeitung:

Für die hier im folgenden zu bearbeitenden Aufgaben steht Ihnen die Datei 'Aufg3.xlsm' zur Verfügung. Bitte nehmen Sie vor Beginn der Bearbeitung die Individualisierung der Daten über Ihre Emailadresse vor.

Abgabe der H-Aufgaben dieses Aufgabenblattes: 30.05.2024

Präsenzaufgaben:

P1: Multiple Regressionsanalyse

In der Datei 'Aufg4H.xlsm', Tabelle 'H1 Originaldaten' finden Sie einen Datensatz mit mehreren stetigen Merkmalen. Bei den weiteren Analysen ist die Variable 'mpg' (Miles per Gallon) stets die Zielgröße.

Führen Sie folgende Untersuchungen durch:

- Berechnen Sie die deskriptive Statistik für die Zielgröße, sowie die anderen stetigen Merkmale

- Berechnen Sie den Korrelationskoeffizienten und die Kovarianz zwischen den stetigen Merkmalen und der Zielgröße. Existiert ein linearer Zusammenhang ?
- Sofern ein linearer Zusammenhang gegeben ist: ist dieser positiv oder negativ proportional ?
- Berechnen Sie für die Variablen mit linearem Zusammenhang zur Zielgröße jeweils ein einfaches Regressionsmodell
- Berechnen Sie für die Variablen mit linearem Zusammenhang zur Zielgröße das multiple Regressionsmodell
- Reduzieren Sie das Modell auf die signifikanten Terme und berechnen Sie für die verbliebenen Variablen das multiple Regressionsmodell

P2: Regressionsanalyse unter Transformation

Die zu betrachtenden Daten sind wiederum in der Tabelle 'H1 Originaldaten'.

Finden Sie die Variablen, für die ein nichtlineares Modell besser geeignet wäre.

Transformieren Sie die Daten so, dass eine lineare Regression sinnvoll wird und zeigen Sie dieses auf.

Hausaufgaben:

H1: Multiple Regressionsanalyse

In der Datei 'Aufg4H.xlsm', Tabelle 'H1' finden Sie einen Datensatz mit mehreren stetigen Merkmalen. Bei den weiteren Analysen ist die Variable 'mpg' (Miles per Gallon) stets die Zielgröße.

Führen Sie folgende Untersuchungen durch:

- Berechnen Sie die deskriptive Statistik für die Zielgröße, sowie die anderen stetigen Merkmale
- Berechnen Sie den Korrelationskoeffizienten und die Kovarianz zwischen den stetigen Merkmalen und der Zielgröße. Existiert ein linearer Zusammenhang ?
- Sofern ein linearer Zusammenhang gegeben ist: ist dieser positiv oder negativ proportional ?
- Berechnen Sie für die Variablen mit linearem Zusammenhang zur Zielgröße jeweils ein einfaches Regressionsmodell

- Berechnen Sie für alle Variablen mit linearem Zusammenhang zur Zielgröße das multiple Regressionsmodell
- Reduzieren Sie das Modell auf die signifikanten Terme und berechnen Sie für die verbliebenen Variablen das multiple Regressionsmodell

H2: Regressionsanalyse unter Transformation

Analysieren Sie zu den Daten in Arbeitsblatt 'H2' der Excel-Datei zum Übungsblatt, ob folgende Anpassungen sinnvoll sind:

- $income_i = \beta_0 + \beta_1 * prestige_i + \epsilon_i$
- $income_i = \beta_0 + \beta_1 * \ln(prestige_i) + \epsilon_i$
- $\ln(income_i) = \beta_0 + \beta_1 * \ln(prestige_i) + \epsilon_i$

Berechnen Sie zur Lösungsfindung die Korrelationskoeffizienten, den Anpassungsgrad und - falls sinnvoll - das lineare Modell.

Wiederholen Sie die Fragestellung für die Variablen 'education' und 'women'.

Analysieren Sie das multiple lineare Modell, das sich aus den bestpassenden Einzelmodellen zusammensetzt.