Lennart Oelschläger (lennart.oelschlaeger@uni-bielefeld.de)
310214 Praktische Übung zu “Einführung in die Mikroökonometrie”
Wintersemester 2024/2025

Aufgabenblatt 0

Organisation

Die Praktische Übung zu “Einführung in die Mikroökonometrie” findet zweiwöchentlich statt. Die Termine und Räume finden Sie im eKVV. Ich lade vor den Treffen das aktuelle Aufgabenblatt im Moodle der Veranstaltung hoch. Im Anschluss finden Sie dort zudem Lösungen und R Code. Sie können in dieser Veranstaltung eine Studienleistung für 31-M23 oder 31-SW-StatM erwerben. Dazu analysieren Sie einen von mir bereitgestellten Datensatz mit den Methoden aus der Vorlesung. Bitte beachten Sie hierzu die Formalia im Moodle.

Aufgaben

Die folgenden Aufgaben sind eine Wiederholung einiger Konzepte aus Mathematik, Wahrscheinlichkeitstheorie und Statistik, die in der Vorlesung eingesetzt werden. Außerdem wiederholen wir das Programmieren in `R` und das Erstellen von Berichten in `R Markdown`.

1. Die Dokumentation *simpleR* von John Verzani liefert eine gute Einführung in `R`, Sie finden das Dokument unter http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf. Lesen Sie *Section 2: Data*, um zu lernen, wie Datensätze mit der `c()` Funktion eingegeben werden können. Erzeugen Sie anschließend folgende zwei Vektoren in `R`:

$$\begin{aligned} x^\top &= \begin{bmatrix} 10 & 8 & 13 & 9 & 11 & 14 & 6 & 4 & 12 & 7 & 5 \end{bmatrix} \\ y^\top &= \begin{bmatrix} 8.1 & 6.9 & 7.5 & 8.8 & 8.3 & 9.9 & 7.2 & 4.2 & 10.8 & 4.8 & 5.6 \end{bmatrix} \end{aligned}$$

2. Welche Operation wird durch `x + y` ausgeführt? Wenden Sie auch `-, *, /, %*%` an.

3. Die Einträge der beiden Vektoren können wir als Wertepaare $(x_1, y_1), \ldots, (x_{11}, y_{11})$ betrachten. Erzeugen Sie mit `plot()` ein Streudiagramm dieser Daten.

4. Lesen Sie *Section 5: Multivariate Data* und lernen Sie den Datentyp `data.frame` kennen. Erstellen Sie einen `data.frame` aus `x` und `y`.

5. Schätzen Sie die Koeffizienten β_0 und β_1 des linearen Modells $y = \beta_0 + \beta_1 x + u$ mithilfe der Funktion `lm()` (steht für *lineares Modell*). Erklärungen finden Sie in *Section 13: Regression Analysis*. Zeichnen Sie dann mithilfe von `abline()` die angepasste Regressionsgerade in Ihr Streudiagramm ein.

6. Erklären Sie den Output von `summary()`, angewendet auf Ihr geschätztes Modell.

7. `R Markdown` ist eine Kombination aus `R`, unserer Programmiersprache für Datenanalyse, und `Markdown`, einer einfachen Auszeichnungssprache für die Erstellung von Berichten. Durch die Kombination können wir `R` Code direkt in Textdokumente integrieren und mit minimalem Aufwand dynamische Analyseberichte generieren. Entwickelt wurde `R Markdown` 2014 von Yihui Xie, der einen einfachen Weg gesucht hat, seine `R` Hausaufgaben aufzuschreiben. Das

Buch *R for Data Science* von Hadley Wickham bietet unter anderem eine gute Einführung in `R Markdown`. Das Buch ist kostenlos online unter https://r4ds.had.co.nz/ verfügbar. Lesen Sie *Chapter 27: R Markdown*, um zu lernen, wie ein `R Markdown` Dokument erstellt werden kann. Verwenden Sie die Vorlage aus Abschnitt 27.2 und erstellen Sie selbst ein `R Markdown` Dokument im `.html`-Format.

8. Fügen Sie in das `R Markdown` Dokument Ihre Lösungen aus Aufgabe 1 ein.
9. Wie können Sie ein `.pdf`-Dokument erstellen?
10. Wie wird das multiple lineare Regressionsmodell definiert, und wie können die Modellparameter geschätzt werden?
11. Welche Eigenschaften hat der Kleinste-Quadrate Schätzer, und welche Voraussetzungen müssen dafür erfüllt sein?
12. Bitte prognostizieren Sie den y Wert für $x_1 = 6$ und $x_2 = -1$, gegeben die Daten
$$\begin{aligned} x_1 &= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 1 & 2 & 3 & 4 & 5 \end{pmatrix}, \\ x_2 &= \begin{pmatrix} 1 & 1 & 2 & 2 & 3 & 3 & 4 & 4 & 5 & 5 \end{pmatrix}, \\ y &= \begin{pmatrix} 1 & 3 & 1 & 5 & 2 & -3 & -3 & -1 & -2 & -2 \end{pmatrix}. \end{aligned}$$
13. Die Schlusskurse von Bitcoin und Ethereum aus 2022 haben eine Kovarianz von 7718820. Kann daraus ein starker, positiver Zusammenhang geschlossen werden?
14. Finden Sie ein Beispiel, dass im Allgemeinen $\mathrm{Var}(X+Y) \neq \mathrm{Var}(X) + \mathrm{Var}(Y)$ für zwei Zufallsvariablen X und Y gilt.
15. Finden Sie zwei Zufallsvariablen, die unkorreliert, aber nicht unabhängig sind.
16. Finden Sie zwei reelle Vektoren x und y, jeweils der Länge 10, sodass die empirische Korrelation $\widehat{\mathrm{Cor}}(x, y)$ exakt -1 bzw. 0 bzw. $+1$ beträgt.
17. Es seien X, Y zwei unabhängige Zufallsvariablen, die mit gleicher Wahrscheinlichkeit die Werte in $\{1, 2, 3\}$ annehmen. Berechnen Sie $\mathrm{E}(1 + 4X + 2Y \mid X = 2)$.
18. Ein Losverkäufer behauptet, dass mindestens 20% seiner Lose Gewinne seien. Die Käufer aber vermuten, dass der Anteil geringer ist. Es werden $n = 100$ Lose überprüft. Führen Sie einen statistischen Test zum Signifikanzniveau $\alpha = 5\%$ zur Streitschlichtung durch, wobei Sie die Aussage des Losverkäufers als Nullhypothese wählen.
19. Angenommen, der wahre Anteil an Gewinnlosen beträgt nur 10%. Berechnen Sie die Wahrscheinlichkeit, mit der Ihr Test dem Losverkäufer fälschlicherweise recht gibt.
20. Sie möchten die in b) berechnete Wahrscheinlichkeit auf unter 5% reduzieren. Auf welche Werte müssen Sie dafür entweder α oder n verändern?
21. Es seien X_1 und X_2 zwei unabhängige Zufallsvariablen mit Varianz 1. Wie lautet die Kovarianzmatrix von $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} X_1 + X_2 \\ X_2 \end{pmatrix}$?
22. Betrachten Sie die folgende Kontingenztabelle, die drei Alterskategorien und die Präferenz für Tee oder Kaffee in einer Gruppe von 30 Personen zeigt:

	Tee	Kaffee
jung	14	1
mittelalt	2	8
älter	3	2

- Erstellen Sie die Randverteilung der Alterskategorien.
- Finden Sie die bedingte Verteilung der Getränkepräferenz gegeben das Alter.
- Berechnen Sie die gemeinsame Verteilung von Alter und Getränkepräferenz.