Florian Wickelmaier
Forschungsmethoden und Mathematische Psychologie
Fachbereich Psychologie
Universität Tübingen

Vorlesung Computergestützte Methoden

Probeklausur

Vorbemerkungen:

1. Arbeitszeit: 90 Minuten
2. Erlaubte Hilfsmittel: R, Python, Texteditor
3. Es können maximal 30 Punkte erreicht werden; bei einer Punktzahl von 15 gilt die Klausur als bestanden.
4. Legen Sie Ihren Studentenausweis bereit.
5. Der Aufgabentext darf **nicht** ins Skript übertragen werden!

Aufgabe 1. (8 Punkte) In einer Erhebung von 1971 in Canada wurde das Ansehen von 102 verschiedenen typischen Berufen untersucht. Laden Sie die Datei `https://www.mathpsy.uni-tuebingen.de/Prestige.csv` herunter. Sie enthält u. a. die Variablen

- income: durchschnittliches Einkommen (Dollars) der Berufsinhaber im Jahr 1971;
- prestige: Pineo-Porter-Prestigescore für Berufe;
- type: Berufsgruppe mit den Stufen 'Blue collar' (`bc`), 'Professional' (`prof`) und 'White collar' (`wc`).

a) (3 Pkt.) Lesen Sie die Datei `Prestige.csv` in R ein und erzeugen Sie einen Dataframe. Entfernen Sie aus dem Dataframe alle Zeilen, die fehlende Werte (`NA`) enthalten. Erzeugen Sie einen Faktor aus `type` mit der Stufenreihenfolge `bc`, `wc`, `prof`.

b) (3 Pkt.) Erstellen Sie eine Abbildung des mittleren Prestigescores in den drei Berufsgruppen; zeichnen Sie auch die zugehörigen Standardfehler ein. Exportieren Sie die Abbildung als Pdf-Datei. Welche Berufsgruppe genießt deskriptiv das größte Ansehen?

c) (2 Pkt.) Testen Sie den Effekt der Berufsgruppe auf das Ansehen mit einer Varianzanalyse ($\alpha = 0.05$). Interpretieren Sie das Ergebnis.

Aufgabe 2. (12 Punkte) Cowles und Davis (1987) untersuchten bei 1421 Teilnehmern, ob bestimmte Persönlichkeitsmerkmale die Wahrscheinlichkeit bedingen, freiwillig an psychologischen Experimenten teilzunehmen. Laden Sie die Datei `https://www.mathpsy.uni-tuebingen.de/Cowles.txt` herunter. Sie enthält die Variablen

- neuroticism: Neurotizismus-Skala aus dem Eysenck-Persönlichkeits-Inventar (EPI);

- extraversion: Extraversions-Skala aus dem EPI;
- sex: Geschlecht des Teilnehmers;
- volunteer: freiwillige Teilnahme an psychologischen Experimenten (`no`, `yes`).

a) (2 Pkt.) Lesen Sie die Datei `Cowles.txt` in R ein und erzeugen Sie einen Dataframe. Erstellen Sie eine Kontingenztabelle mit den beiden Variablen `sex` und `volunteer`. Verwenden Sie die Funktion `mosaicplot()`, um die Häufigkeiten dieser Kontingenztabelle graphisch darzustellen. Was lässt sich über den Zusammenhang der beiden Variablen vermuten?

b) (2 Pkt.) Erzeugen Sie eine neue Variable in Ihrem Dataframe, die die Variable Extraversion in fünf gleichabständige Kategorien unterteilt. Verwenden Sie dafür die Funktion `cut(..., breaks = 5)`. Erzeugen Sie dann eine Kontingenztabelle mit dieser neuen Variable und `volunteer`. Erstellen Sie einen Mosaicplot und interpretieren Sie die Form des Zusammenhangs.

c) (2 Pkt.) Schätzen Sie mit `glm()` das Logit-Modell

$$\log \frac{P(\text{“Yes”})}{1 - P(\text{“Yes”})} = \beta_0 + \beta_1 \text{ neuroticism} + \beta_2 \text{ extraversion} + \beta_3 \text{sex},$$

verwenden Sie dabei `volunteer` als Zielgröße. Interpretieren Sie die Prädiktoren hinsichtlich Signifikanz ($\alpha = 0.05$) und Wirkungsrichtung.

d) (3 Pkt.) Erzeugen Sie den Dataframe `newdat`, der folgende Variablen enthält:

- `neuroticism` mit dem Mittelwert von Neurotizismus;
- `extraversion` mit zweimal 100 Werten von 0 bis 25;
- `sex` mit den Stufen `female` und `male` jeweils 100 mal.

Der Dataframe soll folgende Struktur haben:

```
> str(newdat)

'data.frame':        200 obs. of  3 variables:
 $ neuroticism : num  11.5 11.5 11.5 11.5 11.5 ...
 $ extraversion: num  0 0.253 0.505 0.758 1.01 ...
 $ sex         : Factor w/ 2 levels "female","male": 1 1 1 1 1 1 1 1 1 1 ...
```

e) (3 Pkt.) Erzeugen Sie mit der Funktion `predict()` vorhergesagte Wahrscheinlichkeiten des Logit-Modells für die Werte in `newdat`. Plotten Sie die vorhergesagten Wahrscheinlichkeiten in Abhängigkeit von der Extraversion. Zeichnen Sie dabei für Männer und Frauen getrennte Kurven. Inwiefern haben Extraversion und Geschlecht Einfluss auf die Wahrscheinlichkeit, freiwillig an psychologischen Experimenten teilzunehmen?

Aufgabe 3. (10 Punkte) Robey, Shea, Rutstein und Morris (1992) untersuchten das Bevölkerungswachstum und die Verbreitung von Verhütungsmitteln in 50 Entwicklungsländern. Laden Sie die Datei
`https://www.mathpsy.uni-tuebingen.de/Robey.txt` herunter. Sie enthält die Variablen

- region: Region mit den Stufen Afrika (`Africa`), Asien und Pazifik (`Asia`), Lateinamerika und Karibik (`Latin.Amer`) sowie Naher Osten und Nordafrika (`Near.East`);
- tfr: Fruchtbarkeitsrate (total fertility rate) gemessen in Anzahl der Kinder pro Frau;
- contraceptors: Verhütungsmittelgebrauch gemessen durch den Anteil der verhütenden Frauen an allen verheirateten Frauen im gebärfähigen Alter.

a) (2 Pkt.) Lesen Sie die Datei `Robey.txt` in R ein und erzeugen Sie einen Dataframe. Erstellen Sie mit `lattice::xyplot()` je ein Streudiagramm des Verhütungsmittelgebrauchs in Abhängigkeit von der Fruchtbarkeitsrate für jede Region. Zeichnen Sie jeweils die Regressionsgerade ein (`type = c("p", "r")`).

b) (4 Pkt.) Passen Sie folgende Regressionsmodelle an die Daten an:

$$\text{contraceptors}_i = \alpha + \beta\ \text{tfr}_i + \varepsilon_i \tag{1}$$
$$\text{contraceptors}_i = \alpha + \beta\ \text{tfr}_i + \gamma_{\text{region}} + \varepsilon_i, \tag{2}$$

mit $\varepsilon_i \sim N(0, \sigma^2)$ i.i.d. Testen Sie den bedingten Effekt der Region auf den Verhütungsmittelgebrauch gegeben die Fruchtbarkeitsrate. Bestimmen Sie dafür die Likelihoodquotienten-Statistik der beiden Modelle. Zu welchem Schluss kommen Sie unter Annahme, dass die Statistik F-verteilt ist ($\alpha = 0.05$)?

c) (4 Pkt.) Testen Sie den bedingten Effekt der Region mit einem parametrischen Bootstrap-Verfahren (500 Bootstrap-Replikationen). Stellen Sie die Bootstrap-Stichprobenverteilung des Likelihoodquotienten graphisch dar. Zeichnen Sie die Teststatistik ein. Geben Sie den Bootstrap-p-Wert an und interpretieren Sie das Ergebnis.