

Klausur Biostatistik

Vitaly Belik (vitaly.belik@fu-berlin.de)

2021/03/30

Contents

1 Deskriptive Statistik	1
1.1 Maßzahlen	1
1.2 Häufigkeitstabelle	2
1.3 Histogramm	2
1.4 Boxplot	2
2 Allgemeines Lineares Model (Regression)	2
2.1 Einlesen von Daten	2
2.2 Graphische Darstellung	2
2.3 Statistisches Modell	2
3 Zusammenhänge in Daten	3
4 Zweifaktorielle ANOVA	3
5 Logistische Regression	3

Klausur dauert 90 Minuten. Bitte fügen Sie an den entsprechenden Stellen den fehlenden R-Code hinzu. Es ist empfohlen auch Ihre eigenen Bemerkungen und Kommentare in R-Code zu schreiben. Sie dürfen nur die im Voraus heruntergeladenen Kursmaterialien sowie Ihre eigenen Notizen benutzen. Außerdem vergessen Sie nicht die eingebaute Hilfe-Funktionalität von R. Viel Erfolg!

1 Deskriptive Statistik

Erzeugen Sie eine Reihe von Zahlen:

```
# erzeugt eine zufällige ganze Zahl zwischen 1 und 100  
N <- ceiling(runif(1, 1, 100))  
a <- sample(1:1000, N, replace=TRUE)
```

1.1 Maßzahlen

- Berechnen Sie den Median der Messwerte a .
- Bestimmen Sie den modalen Wert (Modus) der Messwerte a .
- Berechnen Sie die Variationsbreite (Range) der Messwerte a .
- Berechnen Sie die Standardabweichung der Messwerte a .
- Erstellen sie aus dem Vektor a eine *Dataframe* und benennen Sie die Spalte in y um.

1.2 Häufigkeitstabelle

Erstellen Sie die Häufigkeitstabelle der Messwerte a .

1.3 Histogramm

- Erstellen Sie das absolute Histogramm der Messwerte a . Benutzen Sie dabei das Paket **ggplot2**. Beachten Sie, dass
- Verändern Sie die Anzahl der **bins** (Klassen): 20 und 50.
- Erstellen Sie das relative Histogramm der Messwerte a .

1.4 Boxplot

Erstellen Sie das Boxplot-Diagramm der Messwerte a . Benutzen Sie dabei das Paket **ggplot2**.

2 Allgemeines Lineares Model (Regression)

2.1 Einlesen von Daten

Unter URL: [http://belik.userpage.fu-berlin.de/beuth/klausur/2/\[Teil Ihrer E-Mail-Adresse vor @\]/linreg.csv](http://belik.userpage.fu-berlin.de/beuth/klausur/2/[Teil Ihrer E-Mail-Adresse vor @]/linreg.csv) ist der Datensatz für die folgenden Aufgaben erstellt worden.

- Laden Sie den Satz herunter und lesen Sie ihn in das R-Studio ein.
- Lassen Sie R die Namen der Variablen (Spalten) ausgeben.

2.2 Graphische Darstellung

- Stellen Sie mit Hilfe von **ggplot2** die Zusammenhänge graphisch dar.

2.3 Statistisches Modell

Basierend auf den Daten aus dem vorherigen Abschnitt, erstellen Sie das lineare Regressionsmodell unter benutzung der vorhandenen Variablen.

- Erstellen sie vier Regressionsmodelle: mit jeweils einer Prädiktor-Variablen, mit beiden Prädiktoren, mit beiden Prädiktoren und deren Interaktion.
- Untersuchen sie die gewonnenen Modelle mit Hilfe von **summary()**-Befehl. Interpretieren Sie die Ergebnisse.
- Welche Bedeutung haben die Koeffizienten? Interpretieren Sie sie.
- Welches Modell und basierend auf welchen Kriterien würden Sie vorziehen?
- Untersuchen Sie die Residuen auf Normalität: i) mit Hilfe eines Tests und ii) graphisch.

3 Zusammenhänge in Daten

Unter URL: [http://belik.userpage.fu-berlin.de/beuth/klausur/2/\[Teil Ihrer E-Mail-Adresse vor @\]/chi2.csv](http://belik.userpage.fu-berlin.de/beuth/klausur/2/[Teil Ihrer E-Mail-Adresse vor @]/chi2.csv) ist der Datensatz für die folgenden Aufgaben erstellt worden. Dabei handelt es sich um den möglichen Zusammenhang zwischen dem Verzehr der *Spagetti* (1 - Spagetti, 0 - andere Nudelsorte) und *Rote Soße* (1 - Rote Soße, 0 - andere Soße).

- 1) Erstellen Sie sei eine entsprechende Kontingenztafel. Tipp: nutzen Sie den Befehl `table`.
- 2) Unter Verwendung vom Paket **gmodels** untersuchen die Tabelle und führen Sie den χ^2 -Test und exakten Fischer-Test durch.
- 3) Interpretieren Sie die Ergebnisse.

4 Zweifaktorielle ANOVA

Unter URL: [http://belik.userpage.fu-berlin.de/beuth/klausur/2/\[Teil Ihrer E-Mail-Adresse vor @\]/anova.csv](http://belik.userpage.fu-berlin.de/beuth/klausur/2/[Teil Ihrer E-Mail-Adresse vor @]/anova.csv) ist der Datensatz für die folgenden Aufgaben erstellt worden. **value** ist die abhängige Variable.

- a) Erstellen Sie Boxplots in Abhängigkeit von einzelnen Faktoren.
- b) Erstellen Sie Boxplots in Abhängigkeit von einem Faktor. Dabei werden die Niveaus vom anderen Faktor durch den Farbcode wiedergegeben.
- c) Verwandeln Sie die Daten in ein *langes Format* und erstellen sie Tafeln-Graphiken mit einzelnen Graphiken für verschiedene Faktoren
- d) Führen Sie den Test für Homogenität der Varianz
- e) Erstellen Sie die Graphiken, die Interaktionen zwischen den Faktoren illustrieren sollen.
- f) Führen Sie ANOVA-Untersuchung durch mit Hilfe eines linearen Modells und interpretieren Sie die Ergebnisse.
- g) Führen Sie Tukey post-hoc-Test durch und interpretieren Sie die Ergebnisse.

5 Logistische Regression

Unter URL: [http://belik.userpage.fu-berlin.de/beuth/klausur/2/\[Teil Ihrer E-Mail-Adresse vor @\]/logreg.csv](http://belik.userpage.fu-berlin.de/beuth/klausur/2/[Teil Ihrer E-Mail-Adresse vor @]/logreg.csv) ist der Datensatz für die folgenden Aufgaben erstellt worden. **y** ist die abhängige Variable.

- a) Umwandeln Sie falls nötig die kontinuierliche Variable in die binäre Variable
- b) Erstellen sie vier logistische Regressionsmodelle: mit jeweils einer Prädiktor-Variablen, mit beiden Prädiktoren, mit beiden Prädiktoren und deren Interaktion.
- c) Untersuchen sie die gewonnenen Modelle mit Hilfe von **summary()**-Befehl. Interpretieren Sie die Ergebnisse.
- d) Welche Bedeutung haben die Koeffizienten und Odds-Ratios? Interpretieren Sie sie. Benutzen sie den folgenden Befehl

```
exp(cbind(coef(model), confint(model)))
```

wobei **model** ihr logistisches Modell ist.