

Einführung in R

12. Aufgabenblatt

Präsenzaufgabe 1

Der Datensatz *survey* im Paket *MASS* enthält u.a. die Spannweite der Schreibhand *Wr.Hnd* und die Spannweite der Nicht-Schreibhand *NW.Hnd*. Gibt es einen statistischen signifikanten Zusammenhang zwischen *Wr.Hnd* und *NW.Hnd*?

Verwenden Sie $\alpha = 0.05$ für Ihren Test.

Veranschaulichen Sie den Zusammenhang der beiden Merkmale durch ein Streudiagramm.

Präsenzaufgabe 2

Die Daten aus der PISA-Studie-2001 haben wir mit der Varianzanalyse (ANOVA) ausgewertet. Zwei wichtige Voraussetzungen für die Anwendung der ANOVA sind 1) die Normalverteilung der Residuen ($y_i - \mu_i$) und 2) die Varianzhomogenität (der Residuen). Sind diese Voraussetzungen bei diesem Datensatz erfüllt?

Eine Alternative zur **parametrischen** Varianzanalyse ist der nicht-parametrische *Kruskal-Wallis-Test* (`kruskal.test()`), für den keine Normalverteilungsannahme für die Residuen getroffen werden muss. Beantworten Sie die Fragestellung (Unterscheidung der Gruppen) mit diesem Test.

Präsenzaufgabe 3

Zwei zwanzigseitige W20-Würfel (einer lila, einer rot) sind möglicherweise unfair. Zwei prokrastinierende Studierende werfen die Würfel im Laufe eines Abends und erhalten folgende Daten:

Würfel	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	N
Lila	15	13	12	9	11	12	12	11	14	19	18	17	13	15	11	12	12	11	13	4	254
Rot	32	16	12	3	13	11	19	19	10	13	1	9	11	16	14	8	6	10	10	21	254

- Stellen Sie die Ergebnisse als Barplot dar und überprüfen Sie mit dem χ^2 -Test, ob die Würfel fair sind.
- Führen Sie nun einen Zweistichproben- χ^2 -Test durch. Sind die Würfel signifikant verschieden voneinander?

Hausaufgabe 1 (6+6 Punkte)

Betrachten Sie im Datensatz *survey* aus dem Paket *MASS* die Variablen *Pulse* (Pulsrate, Schläge pro Minute) und *Smoke* (Rauchverhalten, verschiedene Kategorien).

- Prüfen Sie mittels einfaktorieller Varianzanalyse, ob es einen statistisch signifikanten Zusammenhang zwischen der Pulsrate und dem Rauchverhalten gibt! Verwenden Sie als Signifikanzniveau $\alpha = 0.05$. Veranschaulichen Sie den Sachverhalt zusätzlich mithilfe einer geeigneten Grafik.
- Wie in Präsenzaufgabe 2 untersuchen Sie, ob 1) die Residuen normal-verteilt sind und 2) Varianzhomogenität vorliegt. Beantworten Sie die Fragestellung der Aufgabe (Zusammenhang zwischen der Pulsrate und dem Rauchverhalten) mit dem *Kruskal-Wallis-Test* (Funktion: `kruskal.test()`).

Hausaufgabe 2 (8 Punkte)

Laden Sie den Datensatz *airquality*. Untersuchen Sie den Zusammenhang zwischen dem metrisch-skalierten Merkmal *Ozone* und dem hier als nominal-skaliert zu betrachtenden Merkmal *Month* in diesem Datensatz mit geeigneten Tests. Begründen Sie mithilfe geeigneter Methoden, warum der von Ihnen gewählte Test geeignet ist. Verwenden Sie $\alpha = 0.05$ für Ihre Tests. Welche Schlussfolgerung ziehen Sie aus dem Testergebnis? Veranschaulichen Sie den Sachverhalt mithilfe von gruppierten Boxplots.

Zusatzaufgabe (7 Punkte)

Betrachten Sie den Datensatz *survey* und ziehen Sie ggf. die Dokumentation zur Hilfe heran. Versuchen Sie Ihre Auswertungen übersichtlich (etwa mit Hilfe von Schleifen) auszuführen.

- Testen Sie (paarweise) mittels der (ein-faktoriellen) Varianzanalyse die Abhängigkeit der verschiedenen metrisch-skalierten Merkmale von den kategorialen Merkmalen mit dem Signifikanzniveau 5%. Die Tests sollen nur explorative Bedeutung haben; dabei ignorieren wir das Multiple-Test-Problem. Welche aufgedeckten signifikanten Zusammenhänge haben Ihrer Meinung nach eine Erklärung? Es gibt 5 numerische und 7 kategorielle Variablen, sodass insgesamt 35 Tests durchzuführen sind.
- Testen Sie jetzt, ob zwischen jeweils zwei kategorialen Merkmalen ein signifikanter Zusammenhang besteht. Verwenden Sie das Signifikanzniveau 5%. Ignorieren Sie auch hier das Problem des multiplen Testens. Welche aufgedeckten signifikanten Zusammenhänge haben Ihrer Meinung nach eine Erklärung?

Abgabe der Lösungen: bis **Montag 20.01.2020**,

`maendle@uni-bremen.de`