
Vorlesung: Programmieren in Anwendungen
Übungsblatt 4

http://softech.informatik.uni-kl.de/Homepage/PIA_SS13

Installation von R

Auf der Website <http://www.r-project.org> erhalten Sie Information und Material zur Statistiksoftware R.

1. Laden Sie die neueste Version für Ihr Betriebssystem herunter und installieren Sie es.
2. Installieren Sie die Library ggplot2 inklusive aller abhängigen Pakete.

Arbeiten mit qplot und ggplot2

Arbeitslosenstatistiken

- Schauen Sie sich die Struktur des Datensatzes economics an.
- Erstellen Sie ein Liniendiagramm mit Hilfe von qplot(), welches die Zahl der Arbeitslosen über die Jahre zeigt. Führen Sie eine Linienglättung durch, sodass ein Trend sichtbar wird.
- Lösen Sie die Aufgabe erneut unter Verwendung von ggplot().

Aus dem Tierreich

- Lesen Sie sich in die Struktur des Datensatzes msleep ein.
- Erstellen Sie einen Scatterplot mit ggplot(). Auf der X-Achse soll das Körpergewicht und auf der Y-Achse die Gesamtschlafdauer aufgetragen werden.
- Um das Diagramm anschaulicher zu machen, verwenden Sie für die X-Achse eine logarithmische Skala, färben Sie die Punkte nach Fressverhalten (Fleisch-, Pflanzen- oder Allesfresser) und beschriften Sie die Achsen sinnvoll.
- Angenommen, Sie wollen jetzt noch eine weitere Eigenschaft, wie z.B. die REM-Dauer in das Diagramm einfließen lassen. Ist dies möglich? Wenn ja, wie?
- Spalten Sie das Diagramm in mehrere Facetten, die die Tiere nach ihrem Erhaltungs-/Schutzstatus gruppieren.

Zufallsvariablen und Wahrscheinlichkeitsverteilungen

1. Erstellen Sie Plots für die Dichtefunktion der Normalverteilung mit verschiedenen Parametern. Wie verändern sich die Graphen bei verschiedenen Mittelwerten, wie bei verschiedenen Varianzen?
2. Erzeugen Sie Stichproben der Größe 10, 100, 1000, und 1000 für die Standardnormalverteilung. Berechnen Sie für den jeweiligen Stichprobenumfang Punktschätzer für den Mittelwert und die Varianz sowie Intervallschätzer für den Mittelwert. Wie verändern sich diese Schätzwerte bei steigender Stichprobengröße?

Hypothesentests

Der PISA-Test

Das *Programme for International Student Assessment*, kurz PISA, ist eine im dreijährigen Turnus durchgeführte, standardisierte Bewertung von (15-jährigen) Schülern der OECD Staaten. Auf der Homepage der Vorlesung finden Sie den Datensatz PISA.csv mit den Ergebnissen einiger ausgewählter Staaten, getrennt nach dem Geschlecht (Variable sex : 1 Female, 2 Male, Perc Sex gibt den Anteil an). Die weiteren Einträge sind:

R00 - R06 : Mittlerer Score zur Lesekompetenz im Jahr 2000 bzw. 2006

M00 - M06 : Mittlerer Score zur Kompetenz in der Mathematik im Jahr 2000 bzw. 2006

S00 - S06 : Mittlerer Score in den Naturwissenschaften (science) im Jahr 2000 bzw. 2006

- Lesen Sie die Daten in R ein und verschaffen Sie sich einen Überblick.
- Visualisieren Sie die Daten.
- Untersuchen Sie mit einem geeigneten Test, ob sich die PISA-Bewertungen im Laufe der Jahre signifikant verändert haben. Stellen Sie dazu zunächst die passende Hypothese auf.

Arzneimittelstudie

Gegeben seien die folgenden Daten aus einer klinischen Studie zur Wirksamkeit zweier Medikamente.

	Medikament A	Medikament B
Behandlungserfolg	6	4
kein Behandlungserfolg	12	9

1. Erstellen Sie eine vollständige Kontingenztabelle für die Medikamentenstudie.
2. Verfassen Sie eine formale Hypothese, um zu testen, ob Medikament A besser wirkt als Medikament B.
3. Warum ist der χ^2 -Unabhängigkeitstest nicht anwendbar? Welche Alternativen gibt es?
4. Verwenden Sie den Fisher-Test, um die Hypothese zu testen. Zu welchem Ergebnis kommen Sie?