

Einführung in R für SozialwissenschaftlerInnen

Übungsblatt Woche 5

Vorgehensweise

Füge, wenn möglich, nur die Befehle in ihre Abgabe ein, die gefordert sind. Alle anderen Befehle können Dein Ergebnis verfälschen. Achte unbedingt darauf, keinen Plot-Befehl einzufügen!

Hintergrund

In dieser Hausaufgabe wirst du Lückentexte füllen. Probier erst den Code in einem Skript aus und fülle die Lücken. Anschließend musst du ein neues Skript erstellen, in welchem du für jede Lücke eine Variable erstellst und diese mit dem Wert belegst, der in der Lücke stehen sollten. Damit das System deine Lösungen auswerten kann, ist es wichtig, dass du folgende Regeln beachtest:

- Schreibe alle Werte in Anführungszeichen (auch wenn sie im Originalcode keine bräuchten)
- Achte auf die richtige Groß- und Klein-Schreibung
- Füge keine zusätzlichen Leerzeichen ein
- Benutze keine Sonderzeichen (auch das Dollarzeichen wird in den Aufgaben nicht benötigt)

Hier ist ein Lückentext als Beispiel:

```
#Runde 0,974 auf 1 Nachkommastelle  
a ( b , c )
```

Deine Lösung sollte dann so aussehen:

```
a<-"round"  
b<-"0.974"  
c<-"1"
```

Aufgabe 1 (Maximale Punktzahl: 7 Punkte)

Ziel: Daten gruppieren und visualisieren Mara arbeitet im Einkauf für einen großen Agrarbetrieb. Um zu entscheiden, welches Futtermittel sie für die Hühner kaufen soll, untersucht sie, welche Futtermittel sich besonders positiv auf das Wachstum von Hühnern auswirken. Hierbei benutzt sie den in R gespeicherten Datensatz chickwts. Dieser enthält das Gewicht von Küken aus einem Experiment. Neu geschlüpfte Küken wurden einer Versuchsgruppe zugewiesen, jede Gruppe erhielt 6 Wochen lang einen bestimmten Futterzusatz. In der Spalte weight ist das Gewicht in Gramm der 6 Wochen alten Küken angegeben, in der Spalte feed der jeweilige Futtertyp.

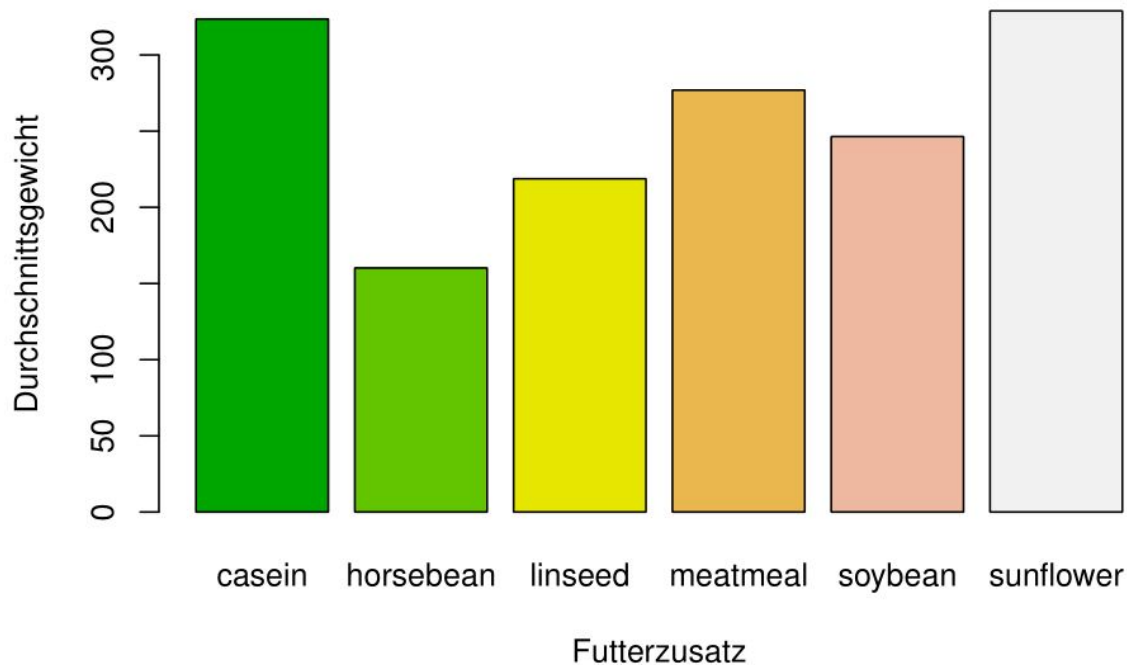
Zuerst berechnet Mara für jede Futtermittel-Gruppe das Durchschnittsgewicht der Küken. Anschließend visualisiert sie dieses Gewicht. Im Folgenden siehst du einen Lückentext von Maras Code und das Ergebnis ihrer Visualisierung.

Schreibe ein R-Skript, in welchem du die Variablen a-g erstellst und mit den Werten belegst, die in Maras Code stehen sollten. Schreibe die Werte jeweils in Anführungszeichen.

```
avgweights <- aggregate( a ~ b , data=chickwts, FUN= c )

barplot(
  avgweights$d ,
  names = avgweights$e ,
  main = "Durchschnitt. Gewicht von Hühnern in Abhängigkeit des Futters",
  f = "Durchschnittsgewicht",
  g = "Futterzusatz",
  col=terrain.colors(6))
```

Durchschnitt. Gewicht von Hühnern in Abhängigkeit des Futters



Aufgabe 2 (Maximale Punktzahl: 8 Punkte)

Ziel: Daten gruppieren und visualisieren Theo arbeitet für die Gesundheitsbehörde. Um zu entscheiden, ob abschreckende Bilder nicht nur auf Zigarettenschachteln, sondern auch auf Flaschen mit Alkohol notwendig wäre, untersucht er die Auswirkungen von Alkohol und Tabak auf Speiseröhrenkrebs. Hierzu verwendet er den in R gespeicherten Datensatz esoph. Dieser stammt aus einer Studie und beinhaltet 88 Gruppen. Jede Gruppe bildet eine andere Kombination aus Altersgruppe, Menge an täglichem Alkoholkonsum und Menge an täglichem Tabakkonsum ab. Für jede Gruppe ist die Menge an Krebskranken und Gesunden in dieser Gruppe angegeben.

Theo speichert den Datensatz in einer Variable e und fügt eine neue Spalte "cancerrate" hinzu, die für jede Gruppen den gerundeten Anteil der Krebskranken an der Menge der Versuchspersonen beschreibt.

Anschließend visualisiert er die Höhe der Krebsrate in Abhängigkeit vom Tabakkonsum und visualisiert die Höhe der Krebsrate in Abhängigkeit vom Alkoholkonsum, um diese beiden Diagramme zu vergleichen.

Zuletzt sucht er nach Gruppen, die ein besonders hohes Risiko für Speiseröhrenkrebs aufweisen. Hierzu berechnet er für jede Kombination aus Altersgruppe und Alkoholkonsum die durchschnittliche Krebsrate und speichert das Ergebnis in der Variablen a. Mit diesem Ergebnis berechnet er die Gruppen, in denen mindestens die Hälfte der Versuchspersonen an Krebs erkrankt sind.

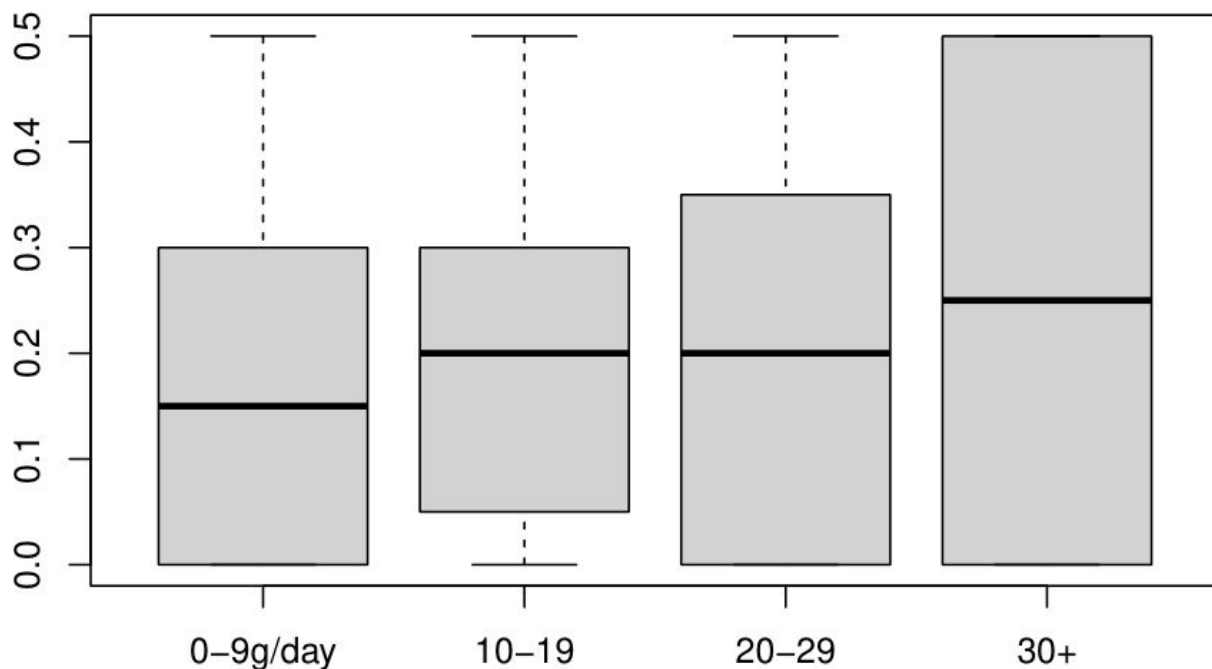
Im Folgenden siehst du einen Lückentext von Theos Code und das Ergebnis seiner Visualisierung. Schreibe ein R-Skript, in welchem du die Variablen a-g erstellst und mit den Werten belegst, die in Theos Code stehen sollten. Bei den Variablen f_1 und f_2 wären theoretisch zwei Lösungen möglich. Nutze hierbei für f_1 den Wert, der alphabetisch zuerst kommt und für f_2 den Wert, der alphabetisch danach kommt. Schreibe die Werte jeweils in Anführungszeichen.

```
e<-esoph
e$cancerrate<-round(e$ncases/(e$ncases+e$ncontrols),1)

a ( b ~ c , data=e , main="Krebs und Tabakkonsum")
a ( b ~ d , data=e , main="Krebs und Alkoholkonsum")

a<-aggregate( e ~ f_1 + f_2 , data=e , FUN=mean)
a[a$g >=0.5,]
```

Krebs und Tabakkonsum



Krebs und Alkoholkonsum

