

Übungszettel 2: Graphische Darstellung von Häufigkeiten

Onlineübung

Aufgabe 2.1 (Zettel & Stift Aufgabe)

Für eine Studie zur Lebenssituation von Studierenden der FH Dortmund wurden an 12 Studierenden u.a. folgende Merkmale abgefragt:

- Wohnform (alleine, WG/Studentenwohnheim, bei den Eltern, bei den Großeltern)
- Entfernung zur Fachhochschule vom Wohnung

| <i>Studierender</i> | <i>Wohnform</i> | <i>Entfernung zur FH</i> | <i>Studiensemester</i> |
|---------------------|-----------------|--------------------------|------------------------|
| 1 | Eltern | 19 | 1 |
| 2 | allein | 3 | 3 |
| 3 | WG | 14 | 1 |
| 4 | Grosseltern | 16 | 4 |
| 5 | Eltern | 10 | 2 |
| 6 | WG | 12 | 3 |
| 7 | WG | 3 | 4 |
| 8 | Eltern | 16 | 4 |
| 9 | allein | 5 | 4 |
| 10 | Eltern | 12 | 1 |
| 11 | Grosseltern | 14 | 5 |
| 12 | Eltern | 10 | 1 |

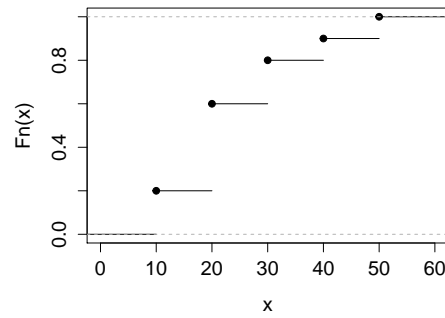
- Erstellen Sie für das Merkmal *Wohnform* eine Häufigkeitstabelle. Veranschaulichen Sie die erstellte Häufigkeitstabelle durch ein Kreisdiagramm und durch ein Säulendiagramm. Interpretieren Sie die Ergebnisse.
- Erstellen Sie ein Histogramm für das Merkmal *Entfernung zur FH*. Die Anzahl der Klassen entnehmen Sie dabei der $\lfloor \sqrt{n} \rfloor$ -Faustregel. Außerdem wählen Sie bitte äquidistante Klassengrenzen.

Aufgabe 2.2 (Zettel & Stift Aufgabe)

Nutzen Sie für diese Aufgaben die Daten aus der Studie zur Lebenssituation von Studierenden der FH Dortmund aus der vorherigen Aufgabe.

- Bestimmen Sie die empirische Verteilungsfunktion des Merkmals *Studiensemester* und zeichnen Sie diese.
- Bestimmen Sie mit Hilfe der empirischen Verteilungsfunktionen den Anteil der Studierender, die bereits länger als 3 Semester an der FH studieren.

Aufgabe 2.3 Für einen Datensatz mit $n = 100$ Beobachtungen eines Merkmals X wurde folgende empirische Verteilungsfunktion erhalten:



- (a) Welche verschiedenen Merkmalsausprägungen wurden für X beobachtet?
- (b) Bestimmen Sie die relative sowie die absolute Häufigkeitsverteilung von X .

Aufgabe 2.4 Die Datei *crashtest.txt* (ILIAS) enthält Informationen über Autos, die für Crash-Tests verwendet wurden. In den Spalten stehen die Automarke, das Modell, die Sicherheitsvorkehrung (z.B. airbag), die Anzahl Türen, das Baujahr und das Gewicht. Kopieren Sie bitte die Datei *crashtest.txt* aus ILIAS in Ihr Arbeitsverzeichnis.

- (a) Lesen Sie diese Daten in R ein, indem Sie in der R-Console den folgenden Befehl ausführen.

```
crash<- read.table( file = "//Ordnername/crashtest.txt",header= TRUE, sep= "\t")
```

Was ist die Bedeutung von `header = TRUE` in diesem Befehl? Schauen Sie sich dazu die Hilfe zu `read.table` an.

Führen Sie die Befehle `crash$Doors`, `crash[,1]` und `crash[10,]` aus und überlegen Sie, was diese bewirken.

- (b) Führen Sie die Befehle

```
abs.ergebnis <- table(crash$Doors)
abs.ergebnis
```

aus und erklären Sie welche Größen die so erzeugte Tabelle enthält.

Erstellen Sie einen Vektor der relativen Häufigkeiten und bezeichnen Sie diesen mit `rel.doors`.

Verwenden Sie dazu die Werte aus `abs.ergebnis` und den Befehl `length()`.

- (c) Führen Sie die folgenden Befehle aus und erklären Sie, was dargestellt wird:

```
namen <- c("keine Angabe", "2 Türen", "4 Türen")
farben <- c("blue", "red", "green")
pie(table(crash$Doors), labels=namen, col=farben)
```

- (d) Balkendiagramme können in R mit dem Befehl `barplot()` gezeichnet werden. Wenden Sie diese Funktion auf Ihren Vektor der relativen Häufigkeiten an.

Ergänzen Sie in Ihrem Balkendiagramm einen vernünftigen Titel und Achsenbeschriftungen.

Versuchen Sie, die Grafik zu verschönern, indem Sie zusätzlich die Farbe und die Breite der Säulen verändern.

- (e) Erstellen Sie ein Histogramm für die Variable Gewicht in kg (`wt`). Zur Bestimmung der Anzahl Klassen soll die Sturges-Regel aus der Vorlesung benutzt werden. Achten Sie außerdem bitte darauf äquidistante Klassen zu bilden. Histogramme können in R mit dem Befehl `hist()` erstellt werden. Wenden Sie diese Funktion auf den Vektor der Beobachtungen an. Mit dem Argument `breaks` können die Klassengrenzen manuell als Vektor angegeben werden. Damit die Rechteckhöhe korrekt berechnet wird, muss außerdem `freq=FALSE` als Argument gesetzt werden.

- (f) Mit der Funktion `ecdf()` kann die empirische Verteilungsfunktion in R berechnet werden. Diese Funktion wird auf den Vektor der Beobachtungen angewendet. Eine Grafik kann aus dem zuvor erstellten Objekt (`ecdf(Beobachtungsvektor)`) anschließend mit `plot()` produziert werden. Erstellen Sie die empirische Verteilungsfunktion für das Baujahr und erzeugen Sie die zugehörige Grafik.

Weitere Aufgaben zur selbständigen Bearbeitung

Aufgabe 2.5 (Zettel & Stift Aufgabe)

An einem Mittwoch im Jahr 2019 wurde für 20 zufällig ausgewählte Kunden der KostBar die Wahl des Gerichtes erhoben.

| | | | | |
|--------------|--------------|------------|-------------|------------|
| Gyros | Gyros | Currywurst | Putenbraten | Currywurst |
| Paninitasche | Currywurst | Currywurst | Putenbraten | Currywurst |
| Currywurst | Paninitasche | Currywurst | Gyros | Currywurst |
| Putenbraten | Currywurst | Currywurst | Gyros | Currywurst |

- Erstellen Sie eine Häufigkeitstabelle, die die Merkmalsausprägungen gemeinsam mit einer Strichliste und den jeweiligen absoluten und relativen Häufigkeiten enthält.
- Stellen Sie das Ergebnis durch ein Kreisdiagramm und durch ein Säulendiagramm graphisch dar. Interpretieren Sie die Ergebnisse.

Aufgabe 2.6 Für ein Optimierungsproblem haben zwei Studierende unterschiedliche Algorithmen (Heuristiken) entwickelt. Um zu prüfen, welches Verfahren am schnellsten funktioniert, haben sie ihre Algorithmen das Optimum von 20 mathematischen Funktion bestimmen lassen. Anschließend haben sie die Zeitdifferenzen (in Sekunden) von *Algorithmus 1* zu *Algorithmus 2* berechnet:

| | | | | | | | | | | |
|---------------------------|----|----|-----|----|----|----|----|------|----|----|
| Differenz $Alg_1 - Alg_2$ | 50 | 61 | -71 | 11 | 32 | 37 | 1 | -139 | 4 | 37 |
| | 18 | 16 | 2 | 41 | 5 | 16 | -2 | 44 | 38 | 30 |

- Wenden Sie die drei in der Vorlesung erwähnten Faustregeln zur Bestimmung der Anzahl der Klassen an und vergleichen Sie die Zahlen.
- Bilden Sie äquidistante Klassen und eine Häufigkeitstabelle. Zur Festlegung der Anzahl an Klassen nutzen Sie bitte die Sturges-Faustregel. Zeichnen Sie ein Histogramm.
- Bilden Sie nun Klassen, die möglichst gleich stark besetzt sind (Anzahl der Klassen wie in Teil (b)). Geben Sie die zugehörigen Klassengrenzen mit an. Zeichnen Sie erneut ein Histogramm.
- Welche der beiden vorgenommenen Klasseneinteilungen erscheint Ihnen sinnvoller für diese Daten?

Aufgabe 2.7 (Zettel & Stift Aufgabe)

Die folgende Tabelle listet die Seitenzahlen einer wöchentlich erscheinenden Zeitschrift in einem Halbjahr auf (also 26 Ausgaben).

| | | | | | | | | | | | | | |
|------------|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Ausgabe | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| Seitenzahl | 55 | 55 | 71 | 55 | 59 | 71 | 83 | 71 | 71 | 83 | 83 | 83 | 71 |
| Ausgabe | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 |
| Seitenzahl | 71 | 83 | 71 | 95 | 83 | 83 | 95 | 95 | 91 | 99 | 83 | 83 | 83 |

- Bestimmen Sie die empirische Verteilungsfunktion und stellen Sie das Ergebnis grafisch dar.
- Bestimmen Sie mit Hilfe der empirischen Verteilungsfunktion den Anteil der Zeitschriftsausgaben, die eine Seitenzahl größer als 71, aber nicht größer als 91 besitzen.