

Statistik I — Aufgabenzettel 1

(Abgabe: 20.11.2023)

Hinweise zur Abgabe dieses Aufgabenzettels

- es handelt sich um eine Gruppenabgabe, wobei Gruppen mit 2-3 Mitgliedern zulässig sind — bei diesem ersten Aufgabenzettel akzeptieren wir zunächst auch noch Einzelabgaben
- Gruppen können tutorienübergreifend gebildet und bei späteren Aufgabenzetteln jederzeit neu zusammengesetzt werden
- jede Person kann in nur einer Gruppe mitwirken und jede Gruppe nur eine Lösung abgeben
- **Achtung:** eindeutig abgeschriebene Lösungen werden als Täuschungsversuch betrachtet!
- die Abgabe ist bis zum 20.11.2023, 23:59 Uhr über Moodle möglich und von nur *einer* (beliebigen) Person aus jeder Gruppe durchzuführen
- geben Sie in Ihrer Abgabedatei auf der ersten Seite der Lösungen klar und deutlich alle Mitglieder der Gruppe an (Vorname, Name, Matrikelnummer)
- wir empfehlen, dass die abgebende Person sowohl die finale Abgabedatei als auch einen Screenshot der Abgabe in Moodle an die anderen Gruppenmitglieder schickt
- Lösungen können u.a. aus folgenden Komponenten bestehen: R Code und R Output, Screenshots, abfotografierte oder eingescannte handschriftliche Lösungen, mit einem Editor wie MS Word erstellte Lösungen — wir sind hier flexibel, alles was lesbar ist wird akzeptiert
- die einzelnen Komponenten sind für die Abgabe jedoch zu einer *einzigsten PDF Datei* zusammenzufügen — hierfür gibt es kostenlose Online-Tools, z.B. Smallpdf
- der Lösungsweg muss immer klar ersichtlich und die Lösung vollständig sein — sollten beispielsweise Grafiken zu erstellen sein, so sollten diese auch Teil Ihrer Abgabe sein
- wir nutzen das folgende Punkteschema (pro Aufgabe):
 - gar nichts gemacht \rightsquigarrow 0 Punkte
 - sich an der Aufgabe versucht, aber sehr wenig richtig gemacht \rightsquigarrow 1 Punkt
 - teilweise richtige Lösungen/Lösungsansätze vorgelegt \rightsquigarrow 2 Punkte
 - die Aufgabe gut bearbeitet, mit kleinen Schönheitsfehlern \rightsquigarrow 3 Punkte
 - die Aufgabe vollständig zufriedenstellend bearbeitet \rightsquigarrow 4 Punkte

(bewertet wird nicht kleinteilig jede Teilaufgabe, sondern nach Gesamteindruck)

- nach Ende der Abgabefrist werden Musterlösungen bei Moodle bereitgestellt, eine Besprechung in den Tutorien wird es nicht geben
- falls während der Bearbeitungszeit Fragen auftreten sollten, dann melden Sie sich jederzeit gerne über die [Pinnwand](#) (eine Antwort kommt in der Regel innerhalb einiger Stunden)

Aufgabe 1: Histogramm

Laden Sie wie folgt die in der Umfrage angegebenen Entfernungen zur Uni (in Km) in R:

```
Entfernung <- read.csv("http://www.rolandlangrock.com//Daten//Aufgabe1.csv")$x
```

- a) Überprüfen Sie, für wie viele Personen in der Stichprobe die angegebene Entfernung zur Uni
- im Intervall $(0, 5]$ (Km) liegt;
 - im Intervall $(5, 10]$ (Km) liegt;
 - im Intervall $(10, 20]$ (Km) liegt;
 - im Intervall $(20, 40]$ (Km) liegt;
 - im Intervall $(40, 80]$ (Km) liegt.
- b) Nutzen Sie die Ergebnisse aus a), um nachfolgende Tabelle zur Erstellung eines Histogramms auszufüllen (der Rechenweg muss hierbei nicht angegeben werden):

Klasse	abs. Häufigkeit	rel. Häufigkeit	Klassenbreite	Säulenhöhe
$0 < \dots \leq 5$				
$5 < \dots \leq 10$				
$10 < \dots \leq 20$				
$20 < \dots \leq 40$				
$40 < \dots \leq 80$				

Hinweis: Bestimmen Sie die Säulenhöhen so, dass der Flächeninhalt des gesamten Histogramms 1 beträgt.

- c) Zeichnen Sie per Hand ein Histogramm mit den in der Tabelle in b) angegebenen Werten für die Säulenhöhen.

Aufgabe 2: Mittelwert, Median, Varianz & Quantile

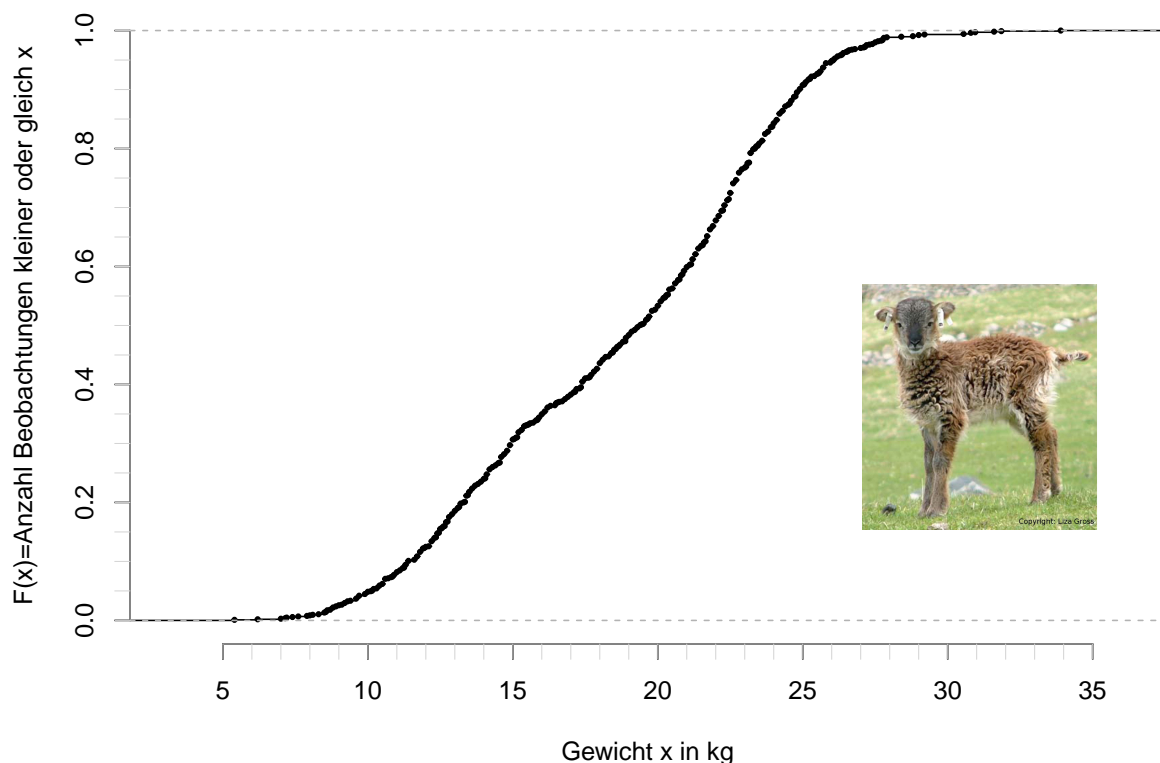
Wir betrachten die Anzahl pro Spiel geschossener Tore in den 306 Spielen (34 Spieltage mit je 9 Spielen) der Bundesligasaison 2022/23:

```
Tore <- read.csv("http://www.rolandlangrock.com//Daten//Aufgabe2.csv")$x
```

- a) Welche Toranzahl (d.h. Merkmalsausprägung) kommt am häufigsten vor?
- b) Bestimmen Sie Mittelwert und Median der Anzahl Tore pro Spiel, am besten einfach mit den entsprechenden R-Befehlen.
- c) Bestimmen Sie nun die (*nicht* korrigierte) empirische Varianz. Hierzu sollten R-Befehle genutzt werden — sonst ist es sehr aufwändig! — achten Sie aber darauf, dass der richtige Faktor verwendet wird.
- d) Bestimmen Sie zuletzt noch das 0.9-Quantil der Anzahl Tore pro Spiel, dabei der Definition aus der Vorlesung folgend (mit anderen Worten: hierzu sollten Sie *nicht* den R-Befehl `quantile(...)` nutzen, da dieser einer leicht anderen Definition folgt).

Aufgabe 3: ECDF & Boxplot

Die folgende Grafik zeigt die empirische Verteilungsfunktion (ECDF) der Körpergewichte (in kg) von 1051 Soayschafen.



- Vervollständigen Sie den folgenden Satz: ca. _____ % der Schafe sind schwerer als 15 kg.
- Lesen Sie aus der Grafik die Fünf-Punkte-Zusammenfassung der Verteilung ab — so genau wie das eben möglich ist.
- Zeichnen Sie (per Hand) den entsprechenden Boxplot der Körpergewichte.

Aufgabe 4: Streudiagramm & Korrelationskoeffizient

Suchen Sie sich online (oder erheben Sie selbst) einen Datensatz mit zwei metrischen Merkmalen X und Y , deren Zusammenhang (und vor allem der Effekt von X auf Y) potenziell für Sie von Interesse ist. Der Themenbereich ist dabei beliebig: Wirtschaft, Politik, Medizin, Arbeit, Schule, Sport, etc. sind z.B. denkbar. Allerdings sollte das Beispiel jenem aus dem Tutorium nicht allzu ähnlich sein.

Speichern Sie die gefundenen (oder erhobenen) Daten dann wie folgt in R als Objekte x und y :

```
x <- c(...hier die Werte von X eintragen, durch Kommas getrennt...)
y <- c(...hier die Werte von Y eintragen, durch Kommas getrennt...)
```

Falls Sie einen größeren Datensatz haben sollten, dann erfolgt das Einlesen in R am besten über die Befehle `read.csv(...)` oder `read.table(...)`. Wir können hierbei auch helfen — bei Problemen melden Sie sich einfach bei Ihrer Tutorin/Ihrem Tutor oder über die [Pinnwand](#).

- (a) Erstellen Sie ein Streudiagramm in **R**. Achten Sie dabei auf eine sinnvolle Achsenbeschriftung.
- (b) Bestimmen Sie in **R** den Korrelationskoeffizienten von X und Y und kommentieren Sie kurz das Ergebnis.

Die Gruppe mit dem besten Streudiagramm (inhaltlich interessant und gut dargestellt) gewinnt Schoko-Nikoläuse!