Statistik I

Paul Strimtu
3898312

Abgabe 1: 20 November 2023

1 Histogramm

1.1 a)

Anzahl der Personen aus der Stichprobe (abs. Häufigkeit) in den jeweiligen Intervallen (Intervall-Variablen passend zur Tabelle aus 1b)):

```
Entfernung <- read.csv("Aufgabe1.csv")
Ex <- sort(Entfernung$x)

#a)
a <- sum(Ex>00 & Ex<=05);a

## [1] 41

b <- sum(Ex>05 & Ex<=10);b

## [1] 31

c <- sum(Ex>10 & Ex<=20);c

## [1] 45

d <- sum(Ex>20 & Ex<=40);d

## [1] 42

e <- sum(Ex>40 & Ex<=80);e

## [1] 26
```

1.2 b)

1.2.1

Relative Häufigkeiten und n ausgerechnet:

```
#b)
n <- length(Entfernung$x);n

## [1] 185

a/n

## [1] 0.2216216

b/n

## [1] 0.1675676

c/n

## [1] 0.2432432

d/n

## [1] 0.227027

e/n

## [1] 0.1405405
```

1.2.2

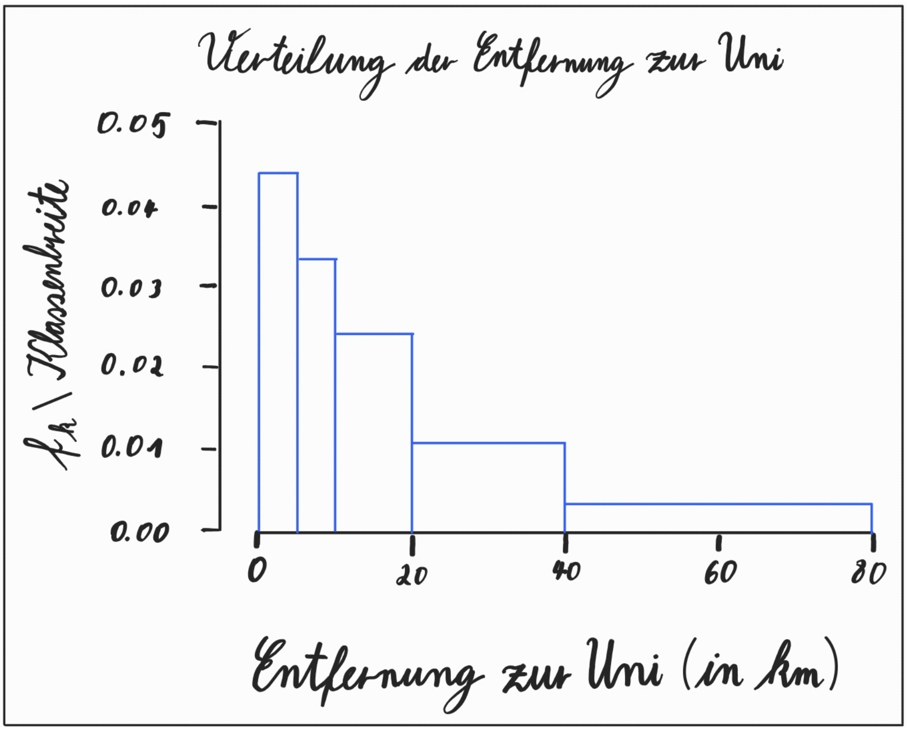
Säulenhöhe ausgerechnet:

```
SHa <- a/(n*05);SHa

## [1] 0.04432432

SHb <- b/(n*05);SHb

## [1] 0.03351351

SHc <- c/(n*10);SHc

## [1] 0.02432432

SHd <- d/(n*20);SHd

## [1] 0.01135135

SHe <- e/(n*40);SHe

## [1] 0.003513514
```

1.2.3

Klasse	abs. Häufigkeit h_k	rel. Häufigkeit f_k	Klassenbreite b	Säulenhöhe SH
$0 < \ldots \leq 5$ a	41	$\approx 0{,}22162$	5	$\frac{41}{(185 \cdot 5)} \approx 0{,}0443$
$5 < \ldots \leq 10$ b	31	$\approx 0{,}16757$	5	$\frac{31}{(185 \cdot 5)} \approx 0{,}0335$
$10 < \ldots \leq 20$ c	45	$\approx 0{,}24324$	10	$\frac{45}{(185 \cdot 10)} \approx 0{,}0243$
$20 < \ldots \leq 40$ d	42	$\approx 0{,}22703$	20	$\frac{42}{(185 \cdot 20)} \approx 0{,}0114$
$40 < \ldots \leq 80$ e	26	$\approx 0{,}14054$	40	$\frac{26}{(185 \cdot 40)} \approx 0{,}00351$

1.3 c)



2 Mittelwert, Median, Varianz & Quantile

2.1 a)

```
Tore <- read.csv("Aufgabe2.csv")
Tx <- sort(Tore$x)

#a)
freq <- table(Tx)

max_freq <- names(freq)[which.max(freq)]
cat("Die Toranzahl",max_freq, "kommt am häufigsten vor.")

## Die Toranzahl 2 kommt am häufigsten vor.
```

2.2 b)

Berechnung des Mittelwertes und des Medians:

```
#b)
mean(Tx)

## [1] 3.173203

median(Tx)

## [1] 3
```

2.3 c)

Brechnung der nicht korrigierten Varianz:

```
#c)
korVar <- var(Tx) * (length(Tx)-1) / length(Tx)
korVar

## [1] 3.136668
```

2.4 d)

Berechnung des 0.9-Quantilis mit der Formel aus dem Skript:

```
#d)
a <- 0.9
if (((length(Tx)*a) %% 1) != 0) { #case n*a is not integer
  l_1 <- Tx[ceiling(length(Tx)*a)]
  cat(l_1 ,"Tore") #6 at index 276
} else { #case n*a is integer
  l_2 <- (Tx[length(Tx)*a] + (Tx[(length(Tx)*a)+1])) / 2
  cat(l_2 ,"Tore")
}

## 6 Tore
```

3 ECDF & Boxplot

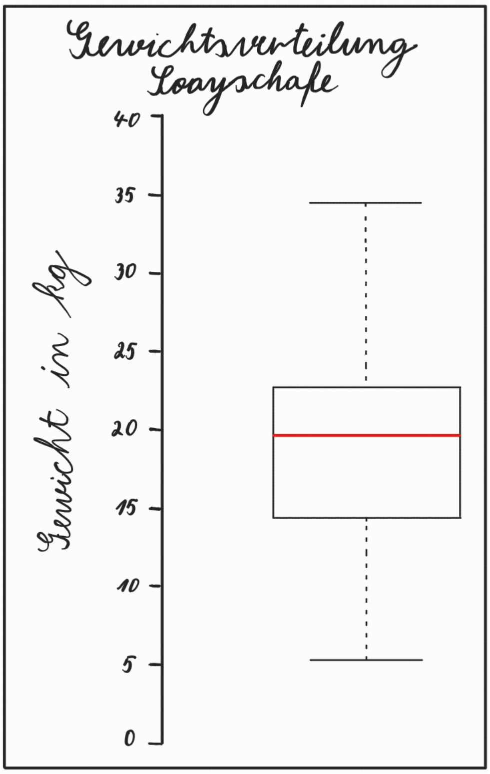
3.1 a)

Ca. 69% der Schafr sind schwerer als 15 kg.

3.2 b)

- $x_{min} \approx 5.4(kg)$,
- $x_{0.25} \approx 14.1(kg)$,
- $x_{0.50} \approx 19.4(kg)$,
- $x_{0.75} \approx 22.7(kg)$,
- $x_{max} \approx 33.9(kg)$

3.3 c)



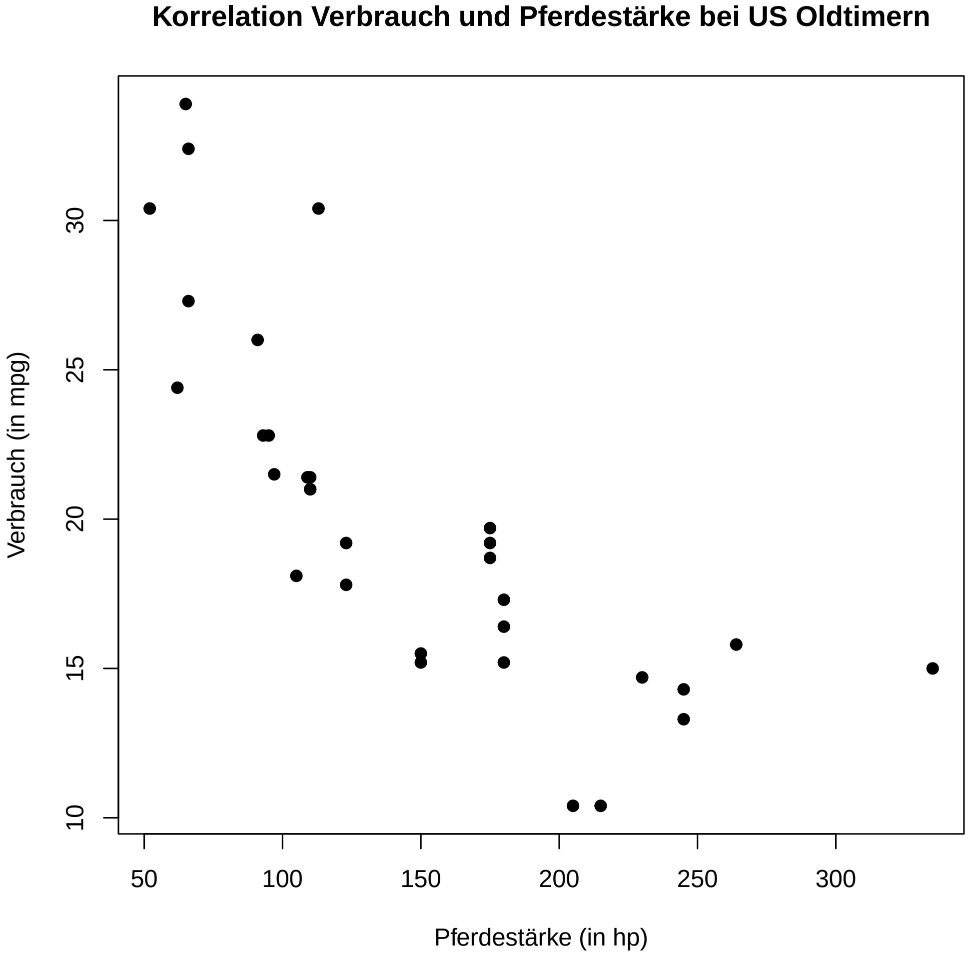
4 Streudiagramm & Korrelationskoesffizient

Das bekannte dataset mtcars beinhaltet unterschiedliche Daten, wie Kraftstoffverbrauch (mpg), hp (PS) usw., zu 32 US Oldtimern aus den 1970er Jahren. Im folgenden habe ich versucht eine Korrelation zwischen dem Kraftstoffverbrauch und den hp der Autos zu finden:

```
data <- read.csv("mtcars.csv")

head(data)

##             rownames  mpg cyl disp  hp drat    wt  qsec vs am gear carb
## 1          Mazda RX4 21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## 2      Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## 3         Datsun 710 22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## 4     Hornet 4 Drive 21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## 5 Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## 6            Valiant 18.1   6  225 105 2.76 3.460 20.22  1  0    3    1

x <- data$hp
y <- data$mpg

cor(x,y) #mpg to hp

## [1] -0.7761684
```

Das Ergebnis zeigt eine (starke) Korrelation von ca. -0.776, was für uns erstmal überraschend sein müsste, da wir ein PS-starkes Auto meist mit einem hohen Kraftstoffverbrauch in Beziehung setzen, was auch zu einer positiven Korrelation führen würde (je mehr desto mehr). Die oberen Daten stammen jedoch auch den USA, wo die Einhheit mpg (miles per gallon) und nicht l/100km verwendet wird. Der Wert der Einheit sinkt also bei einem hohen Kraftstoffverbrauch, da "weniger miles per gallon" gefahren werden können. Dies führt letzendlich zur negativen Korrelation, wie auch der folgede Scatterplot zeigt:

```
plot(x,y,
     main="Korrelation Verbrauch und Pferdestärke bei US Oldtimern",
     xlab="Pferdestärke (in hp)",
     ylab="Verbrauch (in mpg)",
     pch=19)
```


Korrelation Verbrauch und Pferdestärke bei US Oldtimern
Verbrauch (in mpg)
10
15
20
25
30
50
100
150
200
250
300
Pferdestärke (in hp)