NLNP Praktikum 2

Robin Baudisch, Merlin Kopfmann, Maximilian Neudert

Inhaltsverzeichnis

A1

```
load("applicants.RData")
library(tidyr)
library(ggplot2)
df = data.frame(scient, classic)
ggdata = gather(df, "group", "result", c(1, 2))
```

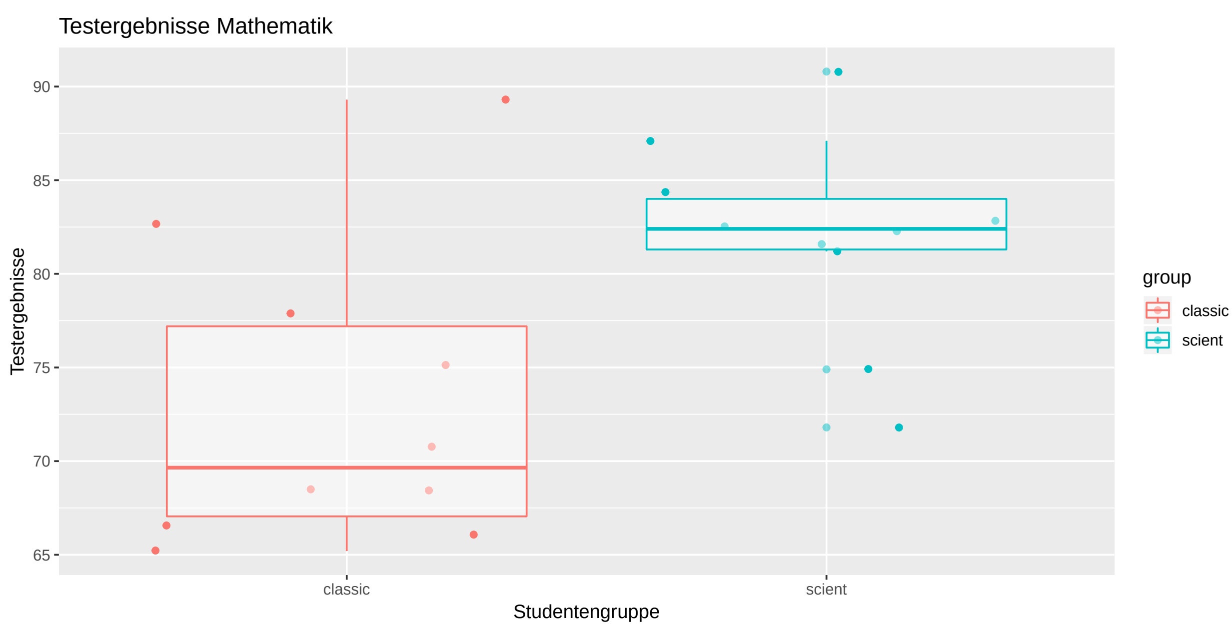
```
summary(scient)
```

```
#>    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#>   71.80   81.30   82.40   81.94   84.00   90.80
```

```
summary(classic)
```

```
#>    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#>   65.20   67.05   69.65   73.06   77.20   89.30
```

```
gg = ggplot(data = ggdata, mapping = aes(x = group, y = result, color = group))
gg = gg + geom_jitter()
gg = gg + geom_boxplot(alpha = 0.5)
gg = gg + labs(title = "Testergebnisse Mathematik", x = "Studentengruppe", y = "Testergebnis
gg
```



Hier eignet sich ein nichtparametrischer Test aus mehrerer Hinsicht:

1. Unsere Stichprobengröße ist mit $n = 10$ klein und gibt es eine geringere Trennschärfe, wodurch eine a priori Annahme der Verteilung der vorliegenden Daten nicht möglich ist.
2. Unsere Stichproben sind ordinalskaliert.

b)

```r
load("applicants.RData")
```

Seien $X_1, ..., X_{10}$ iid, grades der scient. Seien $Y_1, ..., Y_{10}$ iid, grades der classic.

Sei $F_X(x)$ die Verteilungsfunktion von X_i und $F_X(x) = F_Y(x + \theta)$.

Wir formulieren als Nullhypothese, dass die Verteilungen gleich sind

$H_0 : \{\theta \leq 0\}$ vs $H_1 : \{\theta > 0\}$

und Testen zu einem Signifikanzniveau von 5% und $n_1 = 10$, $n_2 = 10$ Daten der Stichproben. Nach Tabelle lehnen wir somit die Nullhypothese ab, wenn $\min(T_X, T_Y) < 82$.

```r
dfs = data.frame(score = scient, class = "s")
dfc = data.frame(score = classic, class = "c")
df = rbind(dfs, dfc)
df = df[order(df$score), ]
df$rk = 1:nrow(df)
df
```

	score	class	rk
18	65.2	c	1
11	66.1	c	2
20	66.6	c	3
14	68.4	c	4
17	68.5	c	5
16	70.8	c	6
9	71.8	s	7
3	74.9	s	8
13	75.1	c	9
15	77.9	c	10
2	81.2	s	11
6	81.6	s	12
1	82.3	s	13
10	82.5	s	14
19	82.7	c	15
8	82.8	s	16
5	84.4	s	17
4	87.1	s	18
12	89.3	c	19
7	90.8	s	20

```r
R1 = sum(df$rk[df$class == "c"])
R2 = sum(df$rk[df$class == "s"])
```

Wir erhalten als kritischer Wert $74 < 78$ und somit lehnen wir die Nullhypothese ab und folgern, dass die Stichproben nicht gleich verteilt sind.

c)

```r
wilcox.test(scient, classic, alternative = "greater", paired = FALSE)
```

```
#>
#>  Wilcoxon rank sum test
#>
#> data:  scient and classic
#> W = 81, p-value = 0.009272
#> alternative hypothesis: true location shift is greater than 0
```

Mit einem $\alpha = 5\%$ wird im obigen Test die Nullhyothese ($\theta < 0$) abgelehnt (p-value $= 0.009272 < 0.05$). Dies bedeutet, die Scores der Naturwissenschaftsstudenten sind signifikant höher als die Scores der Nicht-Naturwissenschaftsstudenten.

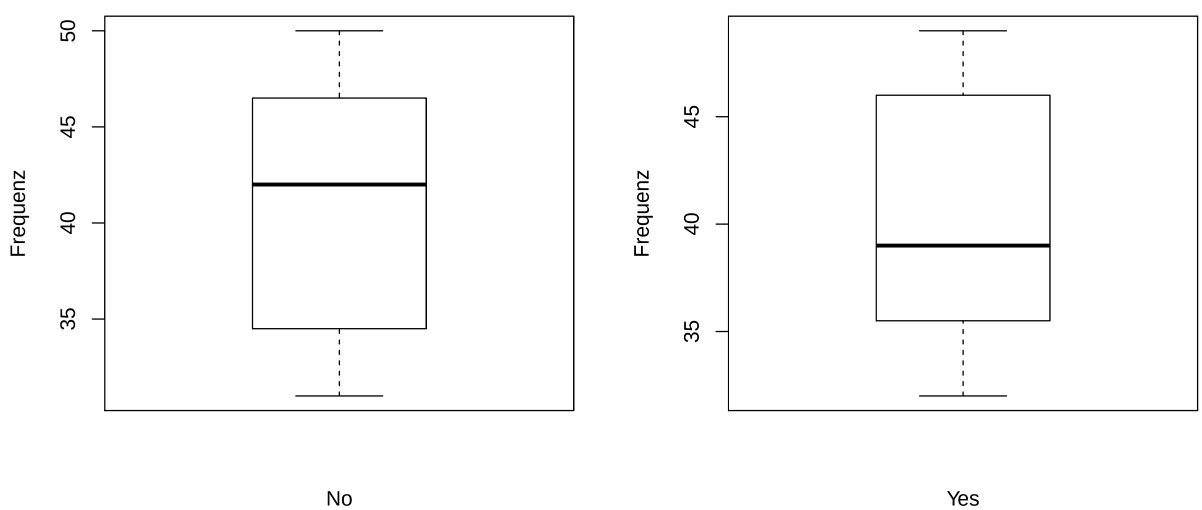
Wir erhalten das gleiche Testergebnis wie bei der Berechnung per Hand.

A2

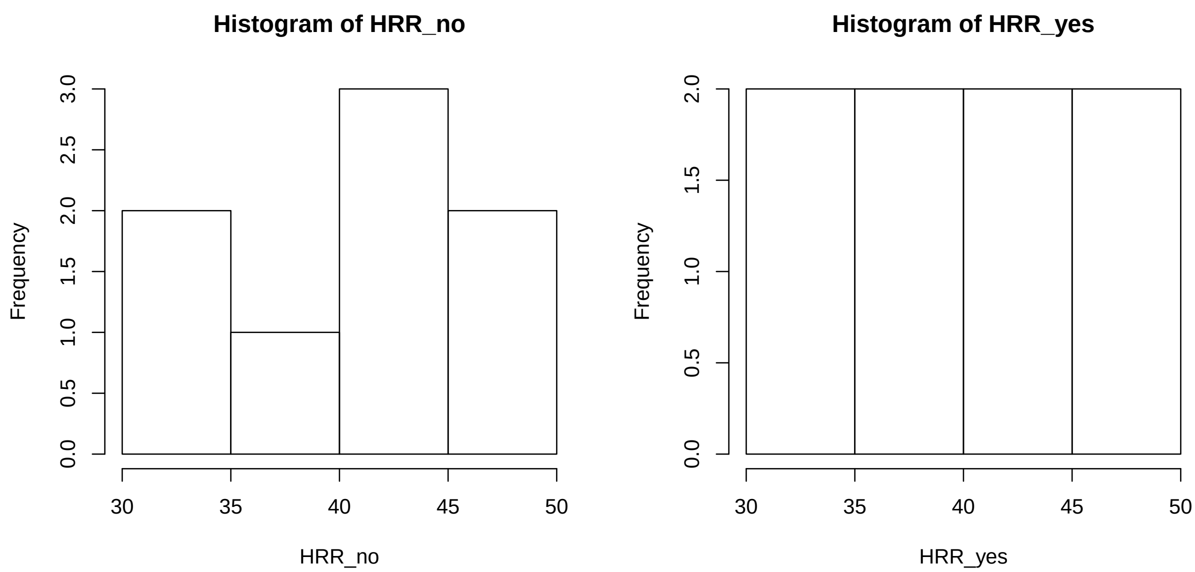
a)

Die Fragestellung ist verbunden, da unsere Daten aus der gleichen Stichprobe kommen und symmetrisch sind.

```r
load("HRR.RData")
par(mfrow = c(1, 2))
boxplot(HRR_no, ylab = "Frequenz", xlab = "No")
boxplot(HRR_yes, ylab = "Frequenz", xlab = "Yes")
```




```r
hist(HRR_no)
hist(HRR_yes)
```



Hier eignet sich, aus den gleichen Gründen wie in Aufgabe 1, ein nichtparametrischer Test zur Überprüfung der Hypothesen.

c)

```
load("HRR.RData")
D = HRR_no - HRR_yes
df = data.frame(D)
df$rank = seq.int(nrow(df))
wp = sum(df[df$D > 0, ]$rank)
wm = sum(df[df$D < 0, ]$rank)
z = (wp - 8 * 9/4)/sqrt(8 * 9 * 17/24)
```

Wir führen einen Wilcoxon-Vorzeichen-Rangtest auf den Differenzen der Stichproben durch.

Seien $X_1, \dots, X_8$ iid, Herzfrequenz ohne Monitoring. Seien $Y_1, \dots, Y_8$ iid, Herzfrequenz mit Monitoring. Sei $D_i := X_i - Y_i$.

Als Nullhypothese nehmen wir an, dass dass die zentrale Lage der Verteilung bei 0 liegt, also

$H_0 : \{D_{i\,\text{med}} = 0\}$ vs $H_1 : \{D_{i\,\text{med}} \neq 0\}$

Nun geben wir den Differenzen Ränge, angefangen bei 1 und das ohne Sortierung und bilden anschließen die Summen über die positiven und negativen Zahlen und erhalten:

$$\begin{aligned} w_+ &= 21 \\ w_- &= 15 \\ T = \min(w_+, w_-) = w_- &= 15 \end{aligned}$$

Da $n = 8$ und Test zweiseitig zu $\alpha = 0.1$ muss nach Tabelle 5 unterschritten werden, damit die Nullhypothese abgelehnt werden muss. Wir haben aber $T = 15 > 5$ somit lehnen wir die Nullhypothese nicht ab. Es gibt keinen signifikanten Unterschied zwischen den Daten.

d)

```r
wilcox.test(HRR_yes, HRR_no, paired = TRUE, alternative = "less")
```

```
#>
#>  Wilcoxon signed rank test with continuity correction
#>
#> data:  HRR_yes and HRR_no
#> V = 16, p-value = 0.4167
#> alternative hypothesis: true location shift is less than 0
```

Mit einem $\alpha = 5\%$ wird im obigen Test die Nullhyothese ($\theta < 0$) nicht widerlegt (p-value $= 0.4167 > 0.05$). Dies bedeutet, der Median von `HRR_yes` ist statistisch nicht signifikant kleiner als der Median von HRR_no.

A3

a)

Für den Permutationstest spielt die Reihnfolge welchen Wert man aus der jeweiligen Gruppe zieht keine Rolle. Folglich kann man die Anzahl an Permutationen berechnen, indem man die Anzahl Permutationen mit Wiederholung für 5 Elemente aus 2 Gruppen mit je 2 und 3 Elementen berechnet. Wir erhalten:

$\binom{2+3}{2} = 10$

b)

Wir haben nur 10 Permutationen, was bedeutet, dass jedes α mindestens 10% beträgt. Wir können somit die geforderten 5% nicht unterschreiten und lehnen die Nullhypothese somit nicht ab.

c)

```r
library(perm)
```

```
#> Error in library(perm): there is no package called 'perm'
```

```r
Braeburn = c(55, 104)
Golden_Delicious = c(89, 108, 98)
permTS(Braeburn, Golden_Delicious)
```

```
#> Error in permTS(Braeburn, Golden_Delicious): konnte Funktion "permTS" nicht finden
```

Mit einem $\alpha = 5\%$ wird im obigen Test die Nullhyothese ($Z \sim Z_\pi$) nicht widerlegt (p-value $= 0.6 > 0.05$). Dies bedeutet, dass es keine tendenziell besser bewertete Apfelsorte gibt.