

Lösung zu Kapitel 9: Beispiel 5

- Haben die Teilnehmer am US-Masters 2009 (`augusta2009.csv`) die Runden 3 und 4 gleich gut absolviert (Variablen `R3` und `R4`)?

Wir lesen zunächst den Datensatz ein und erstellen ein Streudiagramm, in das wir zum Vergleich eine 45°-Gerade einzeichnen.

R

```
> golf <- read.csv2("augusta2009.csv", header = TRUE)
> attach(golf)
> plot(R3, R4)
> abline(c(0, 1))
```

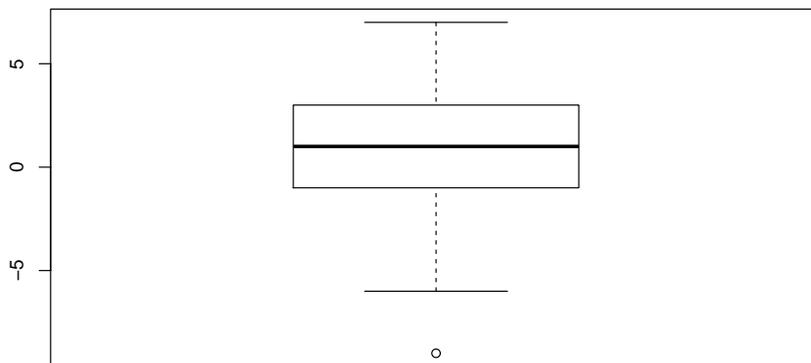


Abbildung 1: Streudiagramme von Runde 3 und Runde 4 in Augusta 2009.

Im Streudiagramm (► [Abbildung ??](#)) liegen etwas – aber nicht sehr viel – mehr Punkte unterhalb der Vergleichsgeraden. Das bedeutet, dass in der dritten Runde durchschnittlich etwas mehr Schläge benötigt wurden als in der vierten. Das zeigt sich auch, wenn wir einen Boxplot der Differenzen der benötigten Schläge bilden (► [Abbildung 2](#)). Es gibt mehr positive als negative Differenzen.

R

```
> boxplot(R3 - R4)
```

Um zu testen, ob signifikante Unterschiede in der Anzahl benötigter Schläge gibt, muss berücksichtigt werden, dass abhängige (verbundene oder gepaarte) Stichproben vorliegen. Daher darf nicht der übliche Zwei-Stichproben-t-Test eingesetzt werden, sondern der t-Test für abhängige Stichproben.

R

```
> t.test(R3, R4, paired = TRUE)
```

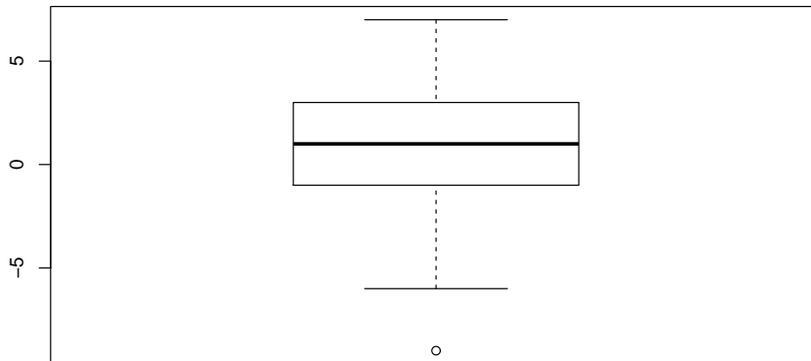


Abbildung 2: Boxplot der Differenzen der Schläge in Runde 3 und Runde 4.

Paired t-test

```
data: R3 and R4
t = 1.1154, df = 49, p-value = 0.2701
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.4168561  1.4568561
sample estimates:
mean of the differences
          0.52
```

Es liegen keine signifikanten Unterschiede in den Mittelwerten vor, der p-Wert liegt mit 0.27 über den üblichen Signifikanzniveaus.

Zum Schluss lösen wir die Verbindung zum Datensatz wieder auf:

R

```
> detach(golf)
```