

Lösung zu Kapitel 8: Beispiel 3

- Beschreiben Sie die Verteilung der für zwei Runden im US-Masters in Augusta 2009 insgesamt benötigten Schläge (die Variablen `R1` und `R2` im Datenfile `augusta2009.csv` enthalten die zwei ersten Rundenergebnisse).
- In welchem Intervall würde man aufgrund dieser Stichprobe die notwendigen Schläge für die zwei ersten Runden annehmen?
- Ist die Anzahl Schläge für die ersten zwei Runden normalverteilt? Welche Ergebnisse zeigen der Shapiro-Wilk-Test und der Kolmogorov-Smirnov-Test? Weist der Q-Q-Plot Auffälligkeiten auf?

R

```
> augusta2009 <- read.csv2("augusta2009.csv")  
> attach(augusta2009)
```

Wirft man einen Blick auf die deskriptiv Statistiken der ersten beiden Rundenergebnisse so sieht man dass der Mittelwert in Runde 2 mit $\bar{x}_2 = 73.74$ etwas höher liegt als jener von Runde 1 mit $\bar{x}_1 = 72.25$. Auch die Standardabweichung und der Interquartilabstand sind in Runde 2 größer. Letzteres ist auch darauf zurückzuführen, dass das Maximum hier bei 84 liegt (verglichen mit Runde 1, hier liegt das Maximum nur bei 80).

R

```
> summary(R1)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
65.00	70.00	72.00	72.25	74.00	80.00

R

```
> sd(R1)
```

```
[1] 3.165605
```

R

```
> range(R1)
```

```
[1] 65 80
```

R

```
> IQR(R1)
```

[1] 4

R

```
> summary(R2)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
65.00	71.00	74.00	73.74	76.00	84.00

R

```
> sd(R2)
```

[1] 3.682804

R

```
> range(R2)
```

[1] 65 84

R

```
> IQR(R2)
```

[1] 5

Das Histogramm (s. ► Abbildung 2) lässt zumindest für die Schläge der ersten Runde auf eine Normalverteilung schliessen. Für Runde 2 liegt eine sehr stark rechtschiefe Verteilung vor. Dies ist wohl auf den moderaten Ausreisser der im Boxplot (s. ► Abbildung 1) erkennbar ist, zurückzuführen.

R

```
> par(mfrow = c(2, 1))
> hist(R1, freq = FALSE, main = "Runde 1")
> xx <- seq(min(R1), max(R1), 0.01)
> lines(xx, dnorm(xx, mean = mean(R1), sd = sd(R1)), col = "red")
> hist(R2, freq = FALSE, main = "Runde 2")
> xx <- seq(min(R2), max(R2), 0.01)
> lines(xx, dnorm(xx, mean = mean(R2), sd = sd(R2)), col = "red")
> par(mfrow = c(1, 1))
```

Wie bereits die Ergebnisse der deskriptiven Analyse zeigten, liegt das 95%-Konfidenzintervall für Runde 2 höher als jenes für Runde 1.

R

```
> boxplot(R1, R2, names = c("Runde 1", "Runde 2"))
```

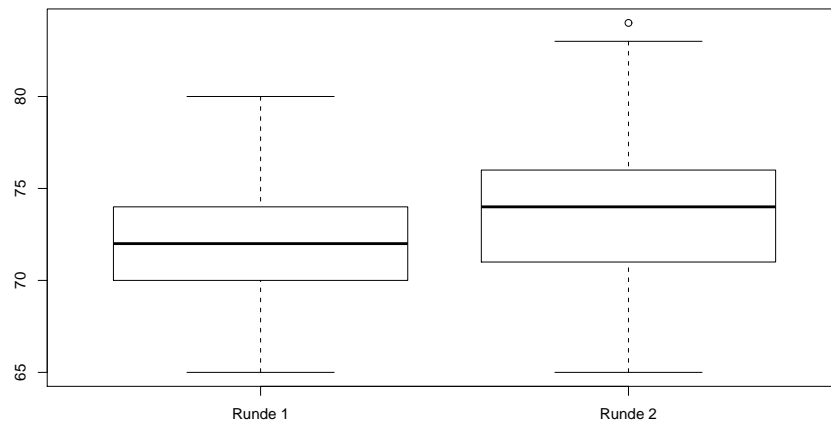


Abbildung 1: Boxplots der benötigten Schläge in den ersten beiden Rundne der US-Masters in Augusta 2009

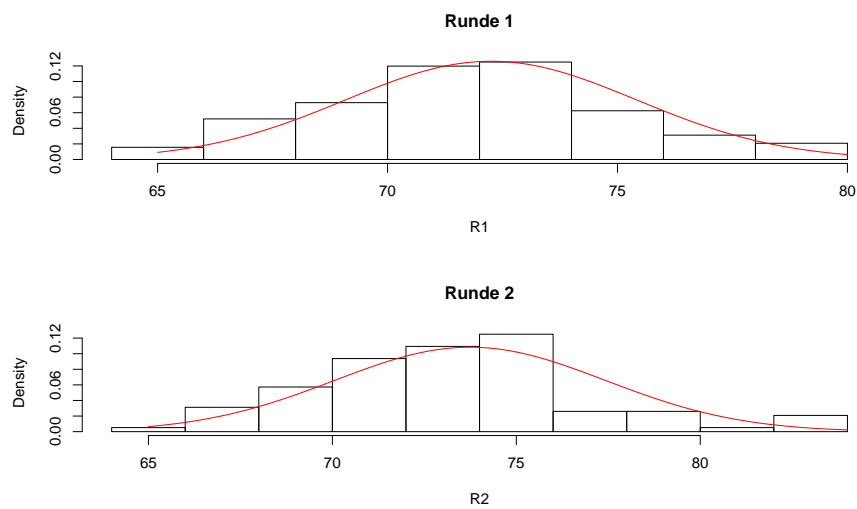


Abbildung 2: Histogramme der benötigten Schläge in den ersten beiden Rundne der US-Masters in Augusta 2009

R

```
> t.test(R1)$conf.int
```

```
[1] 71.60859 72.89141  
attr(,"conf.level")  
[1] 0.95
```

R

```
> t.test(R2)$conf.int
```

```
[1] 72.99338 74.48579  
attr(,"conf.level")  
[1] 0.95
```

Der Shapiro-Wilk Test bestätigt nochmals, was sich bereits im Histogramm in ► Abbildung 2 zeigte, die Schläge sind in der zweiten Runde nicht normalverteilt, in der ersten Runde jedoch schon. Der Kolmogorov-Smirnov Test liefert jedoch ein komplett gegensätzliches Ergebnis (Vorsicht: hier konnte auch kein 'exakter' p-Wert bestimmt werden):

R

```
> ks.test(R1, "pnorm", mean = mean(R1), sd = sd(R1))
```

One-sample Kolmogorov-Smirnov test

```
data: R1  
D = 0.1251, p-value = 0.09904  
alternative hypothesis: two-sided
```

R

```
> shapiro.test(R1)
```

Shapiro-Wilk normality test

```
data: R1  
W = 0.98, p-value = 0.1514
```

R

```
> ks.test(R2, "pnorm", mean = mean(R2), sd = sd(R2))
```

One-sample Kolmogorov-Smirnov test

```
data: R2
D = 0.1134, p-value = 0.169
alternative hypothesis: two-sided
```

R

```
> shapiro.test(R2)
```

Shapiro-Wilk normality test

```
data: R2
W = 0.9703, p-value = 0.02799
```

Auch eine Begutachtung der Daten mittels Q-Q-Plot zeigt lediglich Auffälligkeiten an den Enden der Verteilung der Daten für die Ergebnisse der zweiten Runde.

R

```
> par(mfrow = c(1, 2))
> qqnorm(R1, main = "Runde 1")
> qqline(R1)
> qqnorm(R2, main = "Runde 2")
> qqline(R2)
> par(mfrow = c(1, 1))
```

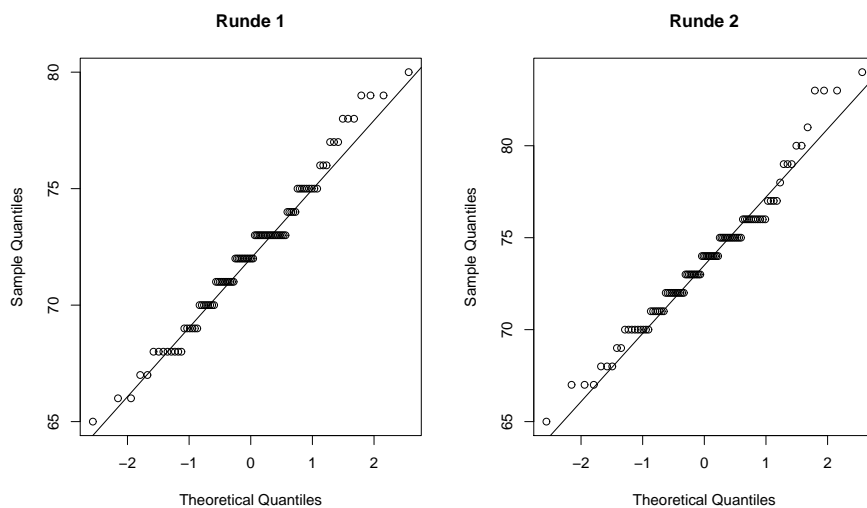


Abbildung 3: Q-Q-Plots der benötigten Schläge in den ersten beiden Runden der US-Masters in Augusta 2009

R

```
> detach(augusta2009)
```