

Lösung zu Kapitel 10: Beispiel 2

In `teenagework.csv` wird die Hausarbeit von Teenagern (`stunden`) inklusive der Berufstätigkeit der Mutter (`mutter`) und dem Geschlecht (`sex`) erfasst. Es soll geprüft werden, ob sich die Anzahl der Stunden zwischen Männern und Frauen – nach Arbeitstätigkeit der Mutter – unterscheidet.

- Die Datei `teenagework.csv` wird in das Objekt `work` eingelesen und mit `attach` kann man direkt auf die Variablen zugreifen. Mit `factor` erstellen wir ein Objekt `sx`, in dem das Geschlecht der Teenager mit Labels erfasst ist. Dasselbe machen wir mit der Variable `mutter` und speichern die Ergebnisse im Objekt `beruf`.

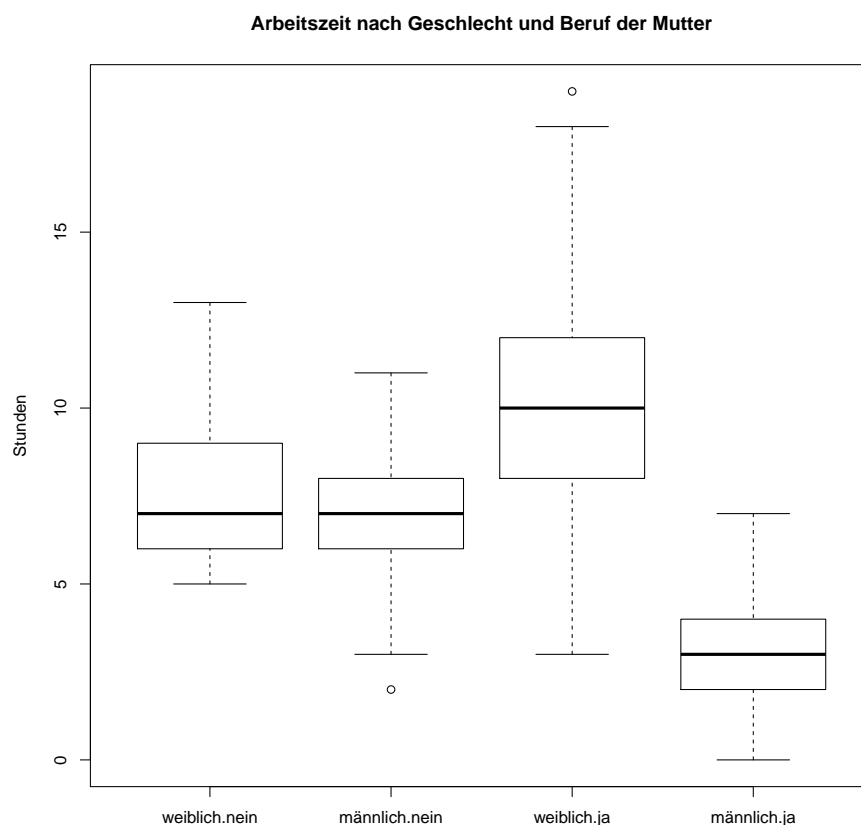
R

```
> work <- read.csv2("teenagework.csv", header = TRUE)
> attach(work)
> sx <- factor(sex, labels = c("weiblich", "männlich"))
> beruf <- factor(mutter, labels = c("nein", "ja"))
```

- Um die Unterschiede zu visualisieren können wir einen Boxplot verwenden.

R

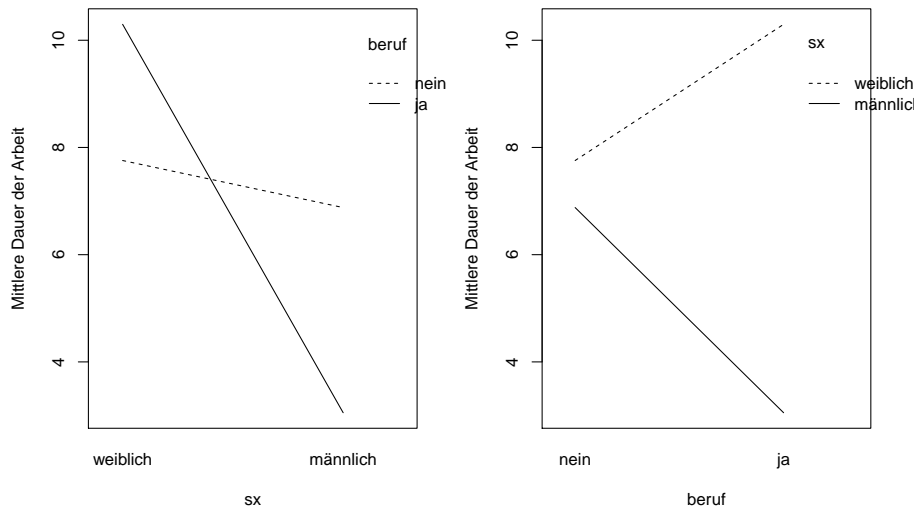
```
> boxplot(stunden ~ sx + beruf, ylab = "Stunden", main = "Arbeitszeit nach Geschlecht und Beruf der Mutter")
```



Für diese Konstellation ist auch ein sog. Interaktionsplot möglich, der die Mittelwerte der Gruppen abbildet. Hier sehen wir, dass die Linien nicht parallel verlaufen, was für die Existenz einer Interaktion spricht.

R

```
> par(mfrow = c(1, 2))
> interaction.plot(sx, beruf, stunden, ylab = "Mittlere Dauer der Arbeit")
> interaction.plot(beruf, sx, stunden, ylab = "Mittlere Dauer der Arbeit")
> par(mfrow = c(1, 1))
```



- Die deskriptive Analyse der Mittelwerte kann man mit **tapply** abkürzen, indem man für jede Kombination von **sx** und **beruf** den Mittelwert der Stunden berechnet. Die Ergebnisse geben wir auf zwei Dezimalstellen gerundet aus. Wir berechnen ein Modell **mod.int** mit **lm**, das die Variable **stunden** durch **sx*beruf** vorhersagt, wodurch wir beide Haupteffekte (**sx** und **beruf**) sowie deren Interaktion (**sx:beruf**) spezifizieren. Wenden wir **anova** auf dieses Objekt an, sehen wir, dass die Interaktion mit $p < .001$ signifikant ist, wodurch wir darauf schließen, dass die Unterschiede der gearbeiteten Stunden abhängig von der Berufstätigkeit der Mutter ist.

R

```
> mittelwerte <- tapply(stunden, list(sx, beruf), mean)
> round(mittelwerte, digits = 2)
```

```
      nein  ja
weiblich 7.76 10.30
männlich 6.88  3.05
```

R

```
> mod.int <- lm(stunden ~ sx * beruf)
> anova(mod.int)
```

Analysis of Variance Table

Response: stunden

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
sx	1	1161.95	1161.95	173.849	< 2.2e-16 ***
beruf	1	9.86	9.86	1.475	0.2261

```
sx:beruf      1  449.47  449.47  67.250 3.663e-14 ***
```

```
Residuals 188 1256.53      6.68
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

R

```
> detach(work)
```