

Deskriptive Statistik und Abbildungen (fortgesetzt)

1. Boxplots und kontinuierliche Daten (siehe auch vorige Woche)

Abhängige Variable: ein kontinuierliche Variable

Unabhängige Variablen: ein oder mehrere Faktoren

```
lok = read.table(file.path(pfad, "lok.txt"))
head(lok); names(lok)
```

```
# Verteilung von slopes in Abhängigkeit von Kons (Artikulationsstelle) und P (Position)
boxplot(slopes ~ Kons * P, data = lok)
```

```
farben = c(rep("red", 3), rep("blue", 3), rep("yellow", 3))
boxplot(slopes ~ Kons * P, data = lok, col=farben, ylab = "Slopes")
legend(locator(1), legend=c("/b/", "/d/", "/g/"), fill=c("red", "blue", "yellow"))
```

2. Barplots und Tabellarische Daten (count data)

Abhängige Variable: eine kategoriale Variable (meistens als Faktor kodiert) und typischerweise binär (ja/nein).

Unabhängige Variablen: ein oder mehrere Faktoren

Kategorial

wie oft wird /r/ im Vergleich zu /R/ in Bayern verwendet? (Kategoriale Daten, weil die Wahl nur zwischen /r/ oder /R/ besteht).

Kontinuierlich: Die F2-Werte von /I/ im Vergleich zu /E/.

Die Zungenposition und -konfiguration in der Erzeugung von /r/-Lauten.

Nützliche Funktionen in der deskriptiven Statistik von kategorialen Daten

`table()` und `barplot()` sind nützliche Funktionen in der deskriptiven Statistik kategorialer Variablen, die auf Faktoren angewandt werden können, z.B.:

```
with(lok, table(G))
with(lok, table(Kons, P))
with(lok, table(Kons, P, G))
```

```
kj = read.delim(file.path(pfad, "kj.txt"))1
```

Wie ist Verteilung /str/ vs. /Str/ nach Geschlecht?

Abhängige Variable: `str` (ob "str" oder "Str")

Unabhängige Variable: `gender` (ob "m" oder "w")

```
with(kj, table(str, gender))
```

Barplot

```
tab = with(kj, table(str, gender))
barplot(tab, beside=T)
```

```
col = c("red", "green")
```

¹ Für diesen Befehl muss ausnahmsweise `read.delim()` statt `read.table()` verwendet werden, weil es in einigen Faktoren und Stufen Leerstellen gibt (z.B. wenn ein Faktor "Age range" statt "Agerange" genannt wird usw.).

```
barplot(tab, beside=T, col=col, ylab="Anzahl", main = "Verteilung von /s, S/")
legend(locator(1), c("b", "d"), fill=col, cex=2)
```

Wie ist die str/Str Verteilung nach Gender **und** Alter?

```
tab = with(kj, table(str, paste(gender, age)))
barplot(tab, beside=T, col=col)
```

```
space = rep(c(1, 0), ncol(tab))
space[7] = 4
barplot(tab, beside=T, col=col, space=space)
```

3. Interaktionen zwischen Faktoren

Zwei Faktoren interagieren miteinander, wenn der Einfluss von einem Faktor auf eine abhängige Variable von einem zweiten Faktor beeinflusst wird.

z.B. Ich erhebe Dauerwerte (abhängige Variable) für Männer und Frauen (Faktor: Geschlecht), die in Bayern und Hessen (Faktor: Dialekt) wohnen.

- Wenn ich feststelle, dass die Vokale 30% länger sind für Männer als Frauen sowohl in Bayern als auch in Hessen, dann gibt es keine Interaktion zwischen den Faktoren Geschlecht und Dialekt.
- Wenn ich feststelle, dass die Vokale 30% länger sind für Männer als Frauen aus Bayern, und 20% länger für Frauen als Männer aus Hessen, dann interagieren die Faktoren sehr stark (= die Dauerverteilung nach Geschlecht ist nicht dieselbe in Bayern und Hessen).

Ob Interaktionen vorliegen, kann durch `boxplot()` und `barplot()` Abbildungen eingeschätzt werden, zusammen mit der Funktion `interaction.plot()`

```
interaction.plot(Faktor1, Faktor2, abhängige Variable)
```

3.1 Abhängige Variable ist kontinuierlich

Inwiefern ist der Einfluss der Artikulationsstelle auf `slopes` unterschiedlich für Männer und Frauen?

```
par(mfrow=c(1,2))
boxplot(slopes ~ Kons * G, data=lok)
with(lok, interaction.plot(Kons, G, slopes))
```

Inwiefern unterscheidet sich der Einfluss von Geschlecht auf `slopes` nach den Artikulationsstellen?

```
with(lok, interaction.plot(G, Kons, slopes))
boxplot(slopes ~ G * Kons, data=lok)
```

Was ist hier die Frage?

```
with(lok, interaction.plot(P, G, slopes))
```

Die Frage?

```
with(lok, interaction.plot(P, Kons, slopes))
```

3.2 Abhängige Variable ist kategorial

```
par(mfrow=c(1,2))
tab = with(kj, table(str, paste(gender, age)))
space = rep(c(1, 0), ncol(tab))
space[7] = 4
barplot(tab, beside=T, col=col, space=space)
```

Für den Interaktion-Plot sollten binäre abhängige kategoriale Variablen in 0 und 1 umgesetzt werden, um Proportionen abzubilden

```
abh = unclass(with(kj, str))
abh = abh-1
with(kj, interaction.plot(gender, age, abh))
```

Fragen

(Sie müssen ggf. die zip-Datei in 4.1 der Webseite noch einmal herunterladen)

1. dip.txt

Inwiefern unterscheiden sich die Diphthongen in F2 und inwiefern fällt dieser Unterschied (falls vorhanden) ähnlich für die beiden Sprecher aus?

2. dr.txt

Inwiefern unterscheiden sich (a) die Dialekte und (b) die Silbenposition in F1. Ist der Unterschied in (a) von der Silbenposition abhängig?

3. mp.txt

Inwiefern unterscheidet sich der Parameter durv1 (Vokaldauer) je nach Quantität? Fällt dieser Unterschied für die verschiedenen Sprecher ähnlich aus?

4. kj.txt

Inwiefern beeinflussen Sozialklasse (Faktor `class`) und Alter die Wahl von str/Str?