

Practical Exercises for Day 2

Sonja Hartnack, Terence Odoch & Muriel Buri

October 2018

Exercise 2

- Create a data frame with 3 columns.

```
# Define single variables
a <- c(1, 20, 3, 40, 5, 60, 7, 8, 92)
b <- c("d", "h", "h", "d", "d", "h", "h", "d", "a")
c <- factor(c("male", "female", "male", "female", "male",
              "female", "male", "male", "female"),
            levels = c("female", "male"))

# Define data frame "dat"
dat <- data.frame(a, b, c)
dat

# For the advanced users:
# Generate dichotomized variable "a.dic" and add it to the data frame
mean.a <- mean(dat$a)
dat$a.dic <- ifelse(dat$a < mean.a, "smaller", "bigger")
dat
```

Exercise 3

- Install package MASS.

```
# install.packages("MASS")
library("MASS")
```

- Load data set bacteria.

```
data(bacteria)
head(bacteria)
str(bacteria)
summary(bacteria)
# ?bacteria
```

- Describe in your own words what the data set bacteria contains.
- Do summary statistic (numerically and graphically).

```
summary(bacteria)
table(bacteria$week)
barplot(table(bacteria$week))
barplot(table(bacteria$trt))
table(bacteria$trt, bacteria$ap)
table(bacteria$trt, bacteria$y)

prop.table(table(bacteria$trt, bacteria$y))
prop.table(table(bacteria$trt, bacteria$y), margin = 1)
prop.table(table(bacteria$trt, bacteria$y), margin = 2)

plot(prop.table(table(bacteria$trt, bacteria$y)))
mosaicplot(~trt + y, data = bacteria)
barplot(prop.table(table(bacteria$y, bacteria$trt), margin=1), beside=TRUE)
barplot(prop.table(table(bacteria$trt, bacteria$y), margin=1), beside=TRUE)
barplot(prop.table(table(bacteria$y, bacteria$trt), margin=1), beside=FALSE)
barplot(prop.table(table(bacteria$trt, bacteria$y), margin=1), beside=FALSE)
```

- Select only observations collected during the second week.

```
subset(bacteria, week == 2)
ss <- subset(bacteria, week == 2)
summary(ss)
# Check if we only have observations of week 2.
table(bacteria$week)
table(ss$week)
```