Practical Exercises for **Monday, July 22, 2019, Day 1**

Sonja Hartnack, Terence Odoch & Muriel Buri

July 2019

Exercise 1: Statistical terminologies

Group the following terminology items into the three categories:

(1) sample & variables

(2) hypothesis testing & statistical modelling

(3) descriptive statistics

- alternative hypothesis
- anova
- barplot
- binary
- binomial
- Bonferroni
- boxplot
- categorical
- Chisquare test
- confounding
- contingency table
- continuous
- correlation coefficient
- count
- data format
- data point
- data type
- degree of freedom
- dependent variable
- effect size
- error
- explanatory variable
- factor
- Fisher’s exact test
- histogram
- hypothesis testing
- hypothesis tests
- independent variable
- integer
- interaction
- intercept
- IQR
- linear model
- linear regression
- logistic regression
- mean
- median
- multiple comparison
- nominal
- normal
- null hypothesis
- numeric
- observation
- odds ratio
- ordinal
- outcome
- paired samples
- poisson
- population
- predictor
- proportion
- p-value
- QQ-plot
- quantile
- range
- regression coefficient
- residuals
- response
- sample
- sampling variation
- scatter plot
- significance
- single-sided test
- skewed data
- slope
- standard deviation
- standard error
- student t-distribution
- treatment effect
- t-test
- two-sided test
- unpaired samples
- variable
- variance
- vector

Exercise 2: Getting to know R and `chickwts`

(a) Open R Studio.

(b) Open a new R-Script.

(c) Load data set `chickwts`.

```r
# ?chickwts
data("chickwts")
head(chickwts)
```

Exercise 3: Summary statistics for the `chickwts` data set

(a) Do summary statistics (numerically and graphically).

```r
### Numerical Statistics
summary(chickwts)
mean(chickwts$weight)
median(chickwts$weight)
sd(chickwts$weight)
# tapply(chickwts$weight, chickwts$feed, mean)
# tapply(chickwts$weight, chickwts$feed, median)
# tapply(chickwts$weight, chickwts$feed, sd)

### Graphics
table(chickwts$feed)
barplot(table(chickwts$feed))
boxplot(chickwts$weight ~ chickwts$feed)
boxplot(weight ~ feed, data = chickwts)
hist(chickwts$weight)
hist(chickwts$weight, freq = FALSE)
lines(density(chickwts$weight), col = "red", lwd = 3)
boxplot(weight ~ feed, data = chickwts, col = "lightgray",
        varwidth = TRUE, main = "chickwt data",
        ylab = "Weight at six weeks (gm)")
barplot(table(chickwts$feed))
```

(b) For advanced R users: Try an anova (are the assumptions fulfilled?) and a Tukey-Anscombe plot.

Try a histogram with a density line on top. ...

```
lm.mod <- lm(weight ~ feed, data = chickwts)
summary(lm.mod)
anova <- aov(weight ~ feed, data = chickwts)
TukeyHSD(anova)
summary(anova)
par(mfrow=c(2,2))
plot(lm.mod)
```

Exercise 4: Data import to R and summary statistics `perulung_ems.csv`

(a) Import the data set `perulung_ems.csv` (taken from Kirkwood and Sterne, 2nd edition) into R.

Data from a study of lung function among children living in a deprived suburb of Lima, Peru.

Variables:

- `fev1`: in liter, ”Forced Expiratory Volume in 1 second” measured by a spirometer. This is the maximum volume of air which the children could breath out in 1 second
- `age`: in years
- `height`: in cm
- `sex`: 0 = girl, 1 = boy
- `respsymp`: respiratory symptoms experienced by the child over the previous 12 months

(b) What *delimiter* do you need to choose?

```
perulung_ems <- read.csv("data/perulung_ems.csv", sep = ";")
lung <- perulung_ems
head(lung)
str(lung)
```

(c) Do summary statistics (numerically and graphically).

```
# summary(lung)
# lung$sex <- factor(lung$sex, levels = c("0", "1"))
# levels(lung$sex) <- c("female", "male")
# lung$respsymptoms <- factor(lung$respsymptoms, levels = c("0", "1"))
# Continuous and factor
tapply(lung$height, lung$sex, mean)
tapply(lung$height, lung$respsymptoms, mean)
# Factor and factor
table(lung$respsymptoms, lung$sex)
```

```r
prop.table(table(lung$respsymptoms, lung$sex))
# Continuous and factor
tapply(lung$age, lung$sex, mean)
tapply(lung$age, lung$respsymptoms, mean)
# Continuous and factor
tapply(lung$fev1, lung$sex, mean)
tapply(lung$fev1, lung$respsymptoms, mean)
```

(d) Plot a boxplot.

```r
boxplot(lung$fev1 ~ lung$sex)
boxplot(lung$fev1)
boxplot(lung$age)
boxplot(lung$height)
```