Solutions to exercises

Listed below are the solutions to the exercises.

All solutions are found using RStudio, though **you should only do the exercises in RStudio if indicated in the list of exercises.** This may result in slight differences in numerical answers, which is due to rounding errors.

The solutions may often be computed in different ways and when two solutions are given it does not necessarily mean that more solutions does not exist. However, when two solutions are given we encourage you to think about why these two solutions are equivalent.

```
library(mosaic)
```

7.11:

a)

```
# Set parameters
n_female <- 817
n_male <- 336
samp_female_prop <- 0.218
samp_male_prop <- 0.467

# Compute sample standard errors
se_female=sqrt(samp_female_prop * (1 - samp_female_prop) / n_female)
se_female
```

```
## [1] 0.0144451
```

```
se_male=sqrt(samp_male_prop * (1 - samp_male_prop) / n_male)
se_male
```

```
## [1] 0.02721776
```

```
# Compute standard error of estimated difference
se_diff <- sqrt(se_female^2+se_male^2)
se_diff
```

```
## [1] 0.03081343
```

b)

```
# Compute the confidence interval
ci=(samp_male_prop - samp_female_prop) + c(-1, 1) * qdist("norm",0.975,mean=0,sd=1,plot=FALSE) * se_dif
ci
```

```
## [1] 0.1886068 0.3093932
```

The interval does not include 0, thus we conclude that there is a difference between genders. Moreover, we find that with 95% confidence between 19 and 31 percentage points more men agree with the statement compared to women.

7.17:

Assumption:

We have to assume that there is no dependence between the subjects of this survey since no information is given.

There are more than 10 subjects in each category, for both loneliness groups. Hence, this assumption is satisfied.

Hypothesis:

$H_0 : \pi_1 = \pi_2$

$H_a : \pi_1 \neq \pi_2$

Analysis:

```
# Set parameters
n_maladap <- 38
n_adap <- 81
samp_prop_maladap <- 25 / n_maladap
samp_prop_adap <- 17 / n_adap
samp_prop_both <- (17+25)/(n_maladap+n_adap)

# Compute the standard error
samp_std_err_0 <- sqrt(samp_prop_both * (1 - samp_prop_both)*(1 / n_maladap
                     +1 / n_adap))

# Compute the test statistics
samp_diff <- (samp_prop_maladap - samp_prop_adap)
z=samp_diff/ samp_std_err_0
z
```

```
## [1] 4.767969
```

```
# Compute the p-value
2 * pdist("norm",-abs(z),plot=F)
```

```
## [1] 1.860923e-06
```

We clearly reject the null hypothesis. So it is relevant to do a confidence interval.

```
# Set parameters
# Compute sample standard errors
se_adap=sqrt(samp_prop_adap * (1 - samp_prop_adap) / n_adap)
se_maladap=sqrt(samp_prop_maladap * (1 - samp_prop_maladap) / n_maladap)
# Compute standard error of estimated difference
se_diff <- sqrt(se_adap^2+se_maladap^2)
ci=(samp_diff) + c(-1, 1) * qdist("norm",0.975,mean=0,sd=1,plot=FALSE) * se_diff
ci
```

```
## [1] 0.2730412 0.6229952
```

The interval does not include 0, thus we conclude that there is a difference between the groups. Moreover, we find that with 95% confidence the maladaptive group are between 27 and 62 percentage points more exposed to COVID-19.

7.20:

```r
# Set parameters
n_cardio <- 258
n_lifting <- 321
samp_mean_cardio <- 294
samp_mean_lifting <- 252
samp_std_dev_cardio <- 23.1
samp_std_dev_lifting <- 20.2
```

a)

```r
samp_mean_diff <- samp_mean_cardio - samp_mean_lifting
samp_mean_diff
```

```
## [1] 42
```

b)

```r
samp_std_err <- sqrt(samp_std_dev_cardio^2 / n_cardio
                     + samp_std_dev_lifting^2 / n_lifting)
```

c) There are more than 30 observations in each sample, so we can use z-scores instead of t-scores.

```r
ci=samp_mean_diff + c(-1, 1) * qdist("norm",0.995,mean=0,sd=1,plot=FALSE) * samp_std_err
ci
```

```
## [1] 37.29292 46.70708
```

We can with 99% confidence conclude that cardio requires between 37 and 47 calories more than lifting.