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.

Module 1.3

5.7:

a)

```
n <- 1845
sample_prop <- 0.17
sample_prop_se <- sqrt(sample_prop * (1 - sample_prop) / n)
sample_prop_se
```

```
## [1] 0.008745111
```

b)

```
sample_prop + c(-1, 1) * qnorm(1 - 0.05 / 2) * sample_prop_se
```

```
## [1] 0.1528599 0.1871401
```

The true proportion (population proportion) is with 95% confidence (**not** probability) included in the interval above. More precisely, after repeated trails (that is if we collect multiple samples and compute a confidence interval for each sample) 95% of the confidence intervals will include the true proportion. The interpretation is difficult and due to the fact that the true proportion (population proportion) is considered fixed while the confidence interval is stochastic.

5.17:

a)

```
sample_prop_A <- 0.45
n <- 500
sample_prop_A_se <- sqrt(sample_prop_A * (1 - sample_prop_A) / n)
sample_prop_A + c(-1, 1) * qnorm(1 - 0.005) * sample_prop_A_se
```

```
## [1] 0.3926914 0.5073086
```

We cannot (with 99% confidence) conclude anything regarding the outcome, since 0.5 (equal probability of both outcomes) is included in the confidence interval.

b)

```
n <- 50
sample_prop_A_se <- sqrt(sample_prop_A * (1 - sample_prop_A) / n)
sample_prop_A + c(-1, 1) * qnorm(0.995) * sample_prop_A_se
```

```
## [1] 0.2687743 0.6312257
```

For the same reasons as in the previous exercise we cannot draw any conclusion.

The uncertainty in b) is greater than in exercise a) since the sample size is smaller. This is due to the formula for the sample standard error where we divide by the number of observations. Hence, more observations implies greater certainty.

5.19:

a)

```
df <- 15 - 1
qt(1 - (1 - 0.95) / 2, df)
```

```
## [1] 2.144787
```

b)

```
df <- 25 - 1
qt(1 - (1 - 0.95) / 2, df)
```

```
## [1] 2.063899
```

c)

```
df <- 25
qt(1 - (1 - 0.95) / 2, df)
```

```
## [1] 2.059539
```

d)

```
df <- 25
qt(1 - (1 - 0.99) / 2, df)
```

```
## [1] 2.787436
```

5.31:

a)

```
mean <- 4.2
se <- 0.03
df <- 2500 - 1
mean + c(-1, 1) * qt(1 - (1 - 0.99) / 2, df) * se
```

```
## [1] 4.122666 4.277334
```

b.i)

Narrower, since less certainty is required. If we want to be very certain we need to include the possibilities
that happens with a low probability.

b.ii)

```
1.496 / sqrt(411)
```

```
## [1] 0.07379224
```

Wider, since the standart error is larger for this group of people.

c)

We assume that the data are on interval scale (numerical measurements where differences are sensible) and
that the spacing between the awnsers are equal.

5.37:

a)

```
prop_guess <- 0.1
sig_lvl <- 1 - 0.95
margin_of_error <- 0.03
ceiling(prop_guess * (1 - prop_guess) * (qnorm(1 - sig_lvl / 2) / margin_of_error)^2)
```

```
## [1] 385
```

b)

```
prop_guess <- 0.5
ceiling(prop_guess * (1 - prop_guess) * (qnorm(1 - sig_lvl / 2) / margin_of_error)^2)
```

```
## [1] 1068
```

5.41:

We Assume that 0 and 18 encompasses the mean plus and minus about 3 standard deviations, then the
standard deviation is approximately 3, and we need:

```
sig_lvl <- 1 - 0.95
margin_of_error <- 1
std <- 3
ceiling((std * qnorm(1 - sig_lvl / 2) / margin_of_error)^2)
```

```
## [1] 35
```