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.

```r
library(mosaic)
```

7.11:

a)

```r
# Set parameters
n_female <- 4690
n_male <- 5120
samp_female_prop <- 0.352
samp_male_prop <- 0.595

# Compute sample standard error
samp_std_err <- sqrt(samp_female_prop * (1 - samp_female_prop) / n_female
                     + samp_male_prop * (1 - samp_male_prop) / n_male)
samp_std_err
```
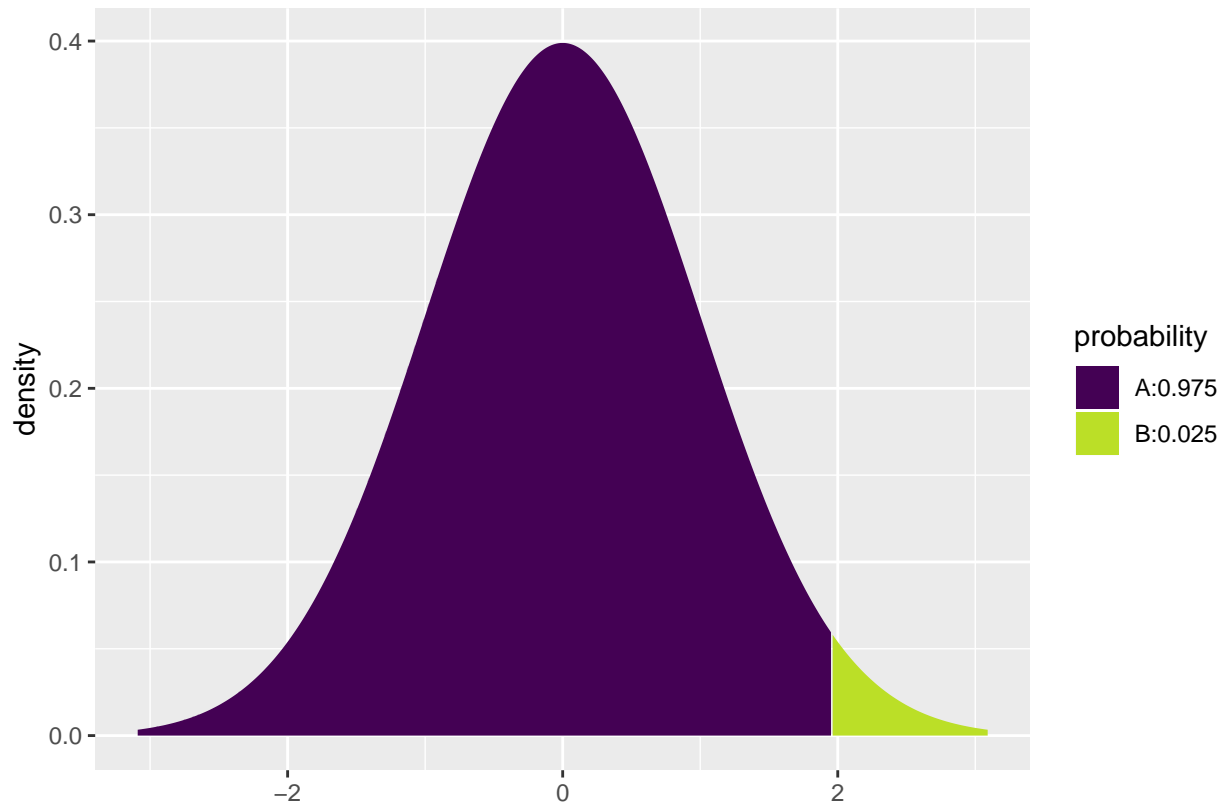
```
## [1] 0.009782636
```

b)

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

```
## [1] 0.2238264 0.2621736
```

The interval does not include 0, thus we conclude that there is a difference between genders. Moreover, we find that with 95% confidence between 22 and 26 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 males and females. Hence, the assumption is satisfied.

Hypothesis:

$H_0 : \pi_1 = \pi_2$

$H_a : \pi_1 \neq \pi_2$

Analysis:

```r
# Set parameters
n_female <- 1501
n_male <- 800
samp_prop_female <- 90 / n_female
samp_prop_male <- 44 / n_male
samp_prop_both <- (90+44)/(n_female+n_male)
```

```
# Compute the standard error
samp_std_err_0 <- sqrt(samp_prop_both * (1 - samp_prop_both)*(1 / n_female
                    +1 / n_male))

# Compute the test statistics
z <- (samp_prop_male - samp_prop_female) / samp_std_err_0
z
```

[1] -0.4838334

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

[1] 0.6285041

The p-value is large, so we do not reject the null hypothesis and thus we do not expect there to be a difference between genders regarding compulsive buying in the population.

7.20:

```
# Set parameters
n_junior <- 415
n_senior <- 560
samp_mean_junior <- 5.5
samp_mean_senior <- 7.5
samp_std_dev_junior <- 8.2
samp_std_dev_senior <- 9.8
```

a)

```
samp_mean_diff <- samp_mean_senior - samp_mean_junior
samp_mean_diff
```

[1] 2

b)

```
samp_std_err <- sqrt(samp_std_dev_junior^2 / n_junior
                    + samp_std_dev_senior^2 / n_senior)
```
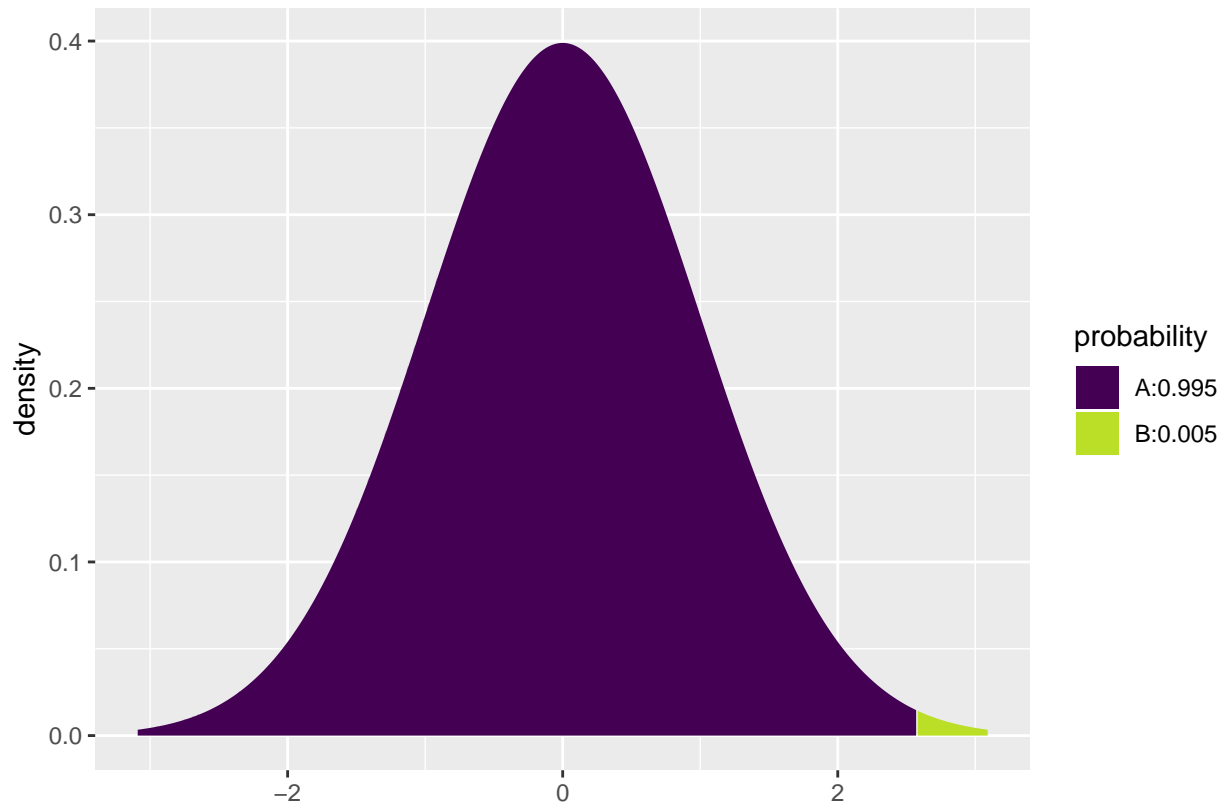
c) There are more than 30 observations in each sample, so we can use z-scores instead of t-scores.

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

```
## [1] 0.5124188 3.4875812
```

We can with 95% confidence conclude that senior researchers spend 0.51 to 3.49 hours more in the lab than junior researchers.