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)
```

6.5:

a)

```
n <- 200
samp_mean <- 53
samp_std_dev <- 20
samp_std_err <- samp_std_dev / sqrt(n)
test_value <- 50

# Compute the test statistic
t <- (samp_mean - test_value) / samp_std_err
t
```

```
## [1] 2.12132
```

```
# Compute degrees of freedom
df <- n - 1

# Compute P-value
2 * pdist("t",-t, df=df,plot=F)
```

```
## [1] 0.03513239
```

We reject the null hypothesis at a 5% significance level since the p-value is smaller than 0.05. That is, the population mean is different from 50 with 95% confidence.

b)

```
n <- 800
samp_std_err <- samp_std_dev / sqrt(n)

# Compute the test statistic
t <- (samp_mean - test_value) / samp_std_err
t
```

```
## [1] 4.242641
```

```
# Compute degrees of freedom
df <- n - 1
```

```r
# Compute P-value
2 * pdist("t",-t, df=df,plot=F)
```

```
## [1] 2.467896e-05
```

As the sample size increases, the sample standard error decreases (the uncertainty decreases), in turn making the test statistic more extreme. Hence, the probability of observing a test statistic that is more extreme than the one observed given that the null hypothesis is true (the p-value) gets smaller.

6.7:

a)

Assumptions:

The exercise states that the sample is random.

The variable is definitely quantitative. It is salary measured in dollars.

Income is usually not normally distributed, and since we only have very few observations $n = 10$ we cannot by the central limit theorem claim that the sample mean in normally distributed and thus the assumptions are not met. However, for the sake of this exercise we assume that the income for this group of people is normally distributed.

Hypothesis:

$H_0 : \mu = 2000$

$H_a : \mu \neq 2000$

Testing:

```r
# Specify parameters
n <- 10
samp_mean <- 1500
samp_std_dev <- 100
samp_std_err <- samp_std_dev / sqrt(n)
test_value <- 2000

# Compute the test statistic
t <- (samp_mean - test_value) / samp_std_err
t
```

```
## [1] -15.81139
```

```r
# Compute degrees of freedom
df <- n - 1

# Compute P-value
2 * pdist("t",t, df=df,plot=F)
```

```
## [1] 7.133289e-08
```

The p-value is practically 0 and thus we reject the null hypothesis at all practical significance levels. That is, the mean income for graduate managers differs from the norm. Further, since the test statistic is negative we note that the mean income for graduate managers is lower than the norm.

b)

The p-value for the 1-sided test with $H_a : \mu < 2000$ is based on the same test statistic and is in this case the probability of observing a lower value than the observed test statistic (due to the alternative hypothesis).

Thus the p-value is half the p-value from Exercise a).

```
p_val <- pdist("t",t, df=df,plot=F)
p_val
```

```
## [1] 3.566644e-08
```

We thus again reject the null hypothesis for all reasonable choices of significance level. This is in correspondence to the observation we did in Exercise a): graduate managers has a lower mean income than the norm.

c)

As the exercise states, the total area under any distribution curve is equal to 1, which means that we may compute the p-value for the test with alternate hypothesis $H_a : \mu > 2000$ as follows:

```
1 - p_val
```

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

The conclusion is in correspondence with the conclusions from the previous exercises a) and b).

6.15:

a)

$H_0 : \pi = 0.5$

$H_a : \pi \neq 0.5$

b)

```
n <- 1155
samp_prop <- 0.453
test_value <- 0.5
samp_std_err <- sqrt(test_value * (1 - test_value) / n)
z <- (samp_prop - test_value) / samp_std_err
z
```

```
## [1] -3.194617
```

The sample proportion is 3.19 standard deviations below (because the test statistic is negative) the test value 0.5.

c)

```
p_val <- 2 * pdist("norm",z,0,1)
```

```
## Verbose output not yet implemented.
```

```
p_val
```

```
## [1] 0.001400163
```

If H_0 is true, the probability of getting a sample proportion at least 3.19 standard deviations from 0.5 is 0.001. Hence, there is strong evidence against the null hypothesis and thus the population proportion is most likely different (smaller even) from 0.5.

d)

It gives an interval of plausible values the population proportion may take. This is opposed to a point estimate that gives only a single value.