Chapter 5: Comparing one or two means

5.1 a) $n = 49$
$\bar{x} = 0.918$
$s = 0.071$
$\mu_0 = 0.9$

5.1 b) $H_0: \quad \mu_0 \leq 0.9$ $\qquad H_1: \quad \mu_0 > 0.9$

5.1 c) Lower bound: $\bar{x} - z_\alpha \times SE_\mu = 0.918 - 1.645 \times \frac{0.071}{\sqrt{49}} = 0.901$

5.1 d) The lower bound of the confidence interval for μ is higher than μ_0. H_0 is rejected with 95% confidence. The PFAS level in town is significantly higher than 0.9 microgram/kg dry soil. There is a risk of 5% for a type-I error.

5.1 e) z-score: $\frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{0.918-0.9}{0.071/\sqrt{49}} = 1.775$

5.1 f) The critical z-value for 95% confidence is 1.645. The p-value (the probability that H_0 is true) is 0.05. For any z-value higher than 1.645 the p-value is lower than 0.05. Since for this test the (absolute of the) calculated z-score = 1.775 and this is higher than 1.645, this means that the p-value is lower than 0.05.

Note: For right sided tests the z-score is negative, but since the standard normal distribution is symmetric we can simply use the positive value for comparing it with the critical z-value.

5.1 g) The calculated z-score is more extreme (higher) than the critical z-value of 1.645. H_0 is rejected with 95% confidence. The PFAS level in town is significantly higher than 0.9 microgram/kg dry soil. There is a risk of 5% for a type-I error. The two methods produce the same answer. It cannot be different; the methods are equivalent.

5.1 h) t-score: $\frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{0.930-0.9}{0.070/\sqrt{16}} = 1.714$

5.1 i) The calculated t-score of 1.714 is lower than the critical t-value of 1.753. H_0 is not rejected. The PFAS level in town is significantly below the norm. There is a risk type-II error.

5.1 j) Lower bound: $\bar{x} - t_{\alpha(df=15)} \times SE_\mu = 0.930 - 1.753 \times \frac{0.070}{\sqrt{16}} = 0.899$

5.1 k) Yes. The lower bound for the population mean μ is 0.899, it can therefore not be ruled out that the mean PFAS level is below 0.9.

5.1 l) The sample is so much smaller (16 vs. 49) that there is too much uncertainty in the result.

5.2 a) Result: 1.645

This is the z-value for a 95% one-sided confidence interval.

```
qnorm(p = 0.95, mean = 0, sd = 1, lower.tail = TRUE)
# Or because the standard normal distribution is the default simply use:
qnorm(0.95) # 1.645
```

5.2 b) Because for a two-sided interval you spread the risk over the two tails. There is 2.5% of risk on the left tail and 2.5% of risk on the right tail.

```
qnorm(p = 0.975) # 1.960
```

5.2 c) The `pnorm()` is the inverse of the `qnorm()` function: It returns the cumulative probability for a given value `q` in a specified normal distribution.

```
pnorm(q = 1.645, mean = 0, sd = 1) # 0.95
pnorm(1.960) # 0.975
```

5.2 d) The `qt()` function returns the t-value for a given cumulative probability `p` and given number of degrees of freedom `df`.

```
qt(p = 0.95, df = 15) # 1.753
```

5.2 e) The `pt()` is the inverse of the `qt()` function: It returns the cumulative probability for a given value `q` in a specified t-distribution.

```
pt(q = 1.753, df = 15) # 0.95
```

5.2 f) t_1: -1.06 t_2: -1.328
t_3: 1.328 t_4: 1.328

```
qt(p = 0.15, df = 19) # -1.06
qt(p = 0.10, df = 19) # -1.328
qt(p = 0.90, df = 19) # 1.328
qt(p = 0.10, df = 19, lower.tail = FALSE) # 1.328
```

5.2 g) p_1: 0.081 p_2: 0.929 p_2: 0.015

```r
pt(q = -1.5, df = 11) # 0.081
diff(pt(q = c(-2,2), df = 11)) # 0.929
pt(q = 2.5, df = 11, lower.tail = FALSE) # 0.015
```

5.2 h) Two tailed inequality test: 2.467
One-tailed right sided test: 2.153
One-tailed left sided test: -2.153

```r
# Two-tailed inequality test
qt(p = 0.99, df = 28) # 2.467

# One-tailed right sided test
qt(p = 0.98, df = 28) # 2.153

# One-tailed left sided test
qt(p = 0.98, df = 28, lower.tail = FALSE) # -2.153
```

5.2 i) Two-tailed inequality test: H_0 rejected (2.6 > 2.476)
One-tailed right sided test: H_0 rejected (-2.3 < -2.153)
One-tailed left sided test: H_0 not rejected (1.6 < 2.153)

5.3 a) Because different men get the caffeine and the placebo. There are 18 unique test subjects. Everyone gets tested once and every measurement is therefore independent.

5.3 b) H_0: $\mu_1 \leq \mu_2$ H_1: $\mu_1 > \mu_2$

5.3 c) ${}^2_p = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2)} = \frac{(9-1)\times 5.61^2+(9-1)\times 7.70^2}{9+9-2} = 45.38$

5.3 d) $t = \frac{(x_1-x_2)-D_0}{\sqrt{s_p^2(\frac{1}{n_1}+\frac{1}{n_2})}} = \frac{(94.22-100.56)}{\sqrt{45.38\times(\frac{1}{9}+\frac{1}{9})}} = 1.995$

5.3 e) The calculated t-score of 1.995 is more extreme (higher) than the critical value of 1.746. H_0 is rejected with 95% confidence. The mean placebo RER level is significantly higher than the mean caffeine RER level. There is a risk of 5% for type-I error.

5.3 f) **placebo**
Mean: 100.55
Standard deviation: 7.699

caffeine
Mean: 94.22
Standard deviation: 5.607

```
# These are the values for the RER test data set
placebo <- c(105, 119, 100, 97, 96, 101, 94, 95, 98)
caffeine <- c(96, 99, 94, 89, 96, 93, 88, 105, 88)


mean(placebo) # Mean: 100.55
sd(placebo) # Standard deviation: 7.699

mean(caffeine) # Mean: 94.22
sd(caffeine) # Standard deviation: 5.607
```

5.3 g) The code runs an independent (not paired) samples t-test with a confidence of 95% for `placebo` and `caffeine`, with the alternative hypothesis H_1 that the difference in means is greater than 0. The variances are assumed equal.

```
t.test(x = placebo, y = caffeine, alternative = 'greater', mu = 0,
       paired = FALSE, var.equal = TRUE, conf.level = 0.95)
```

5.3 h) The Welch Two Sample t-test leads to a slightly different p-value of 0.03252, but the same conclusion: rejection of H_0.

```
t.test(x = placebo, y = caffeine, alternative = 'greater', mu = 0,
       paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

5.3 i) You can use Hartley's F or Levene's test.

5.4 a) The observations are not independent because the same twelve people are tested twice. The two blood pressure measurements for one person are connected/dependent: a person with high blood pressure will have higher values in both experiments. That is why in a dependent t-test you look at the difference between the two measurements.

5.4 b) Mean standing: 140.83
Mean lying: 143.33
Mean difference: 2.5

```r
# These are the values for the blood pressure data set
standing <- c(132, 146, 135, 141, 139, 162, 128, 137, 145, 151, 131, 143)
lying <- c(136, 145, 140, 147, 142, 160, 137, 136, 149, 158, 120, 150)

mean(standing) # Mean standing: 140.83
mean(lying) # Mean lying: 143.33
differences <- lying - standing
mean(differences) # Mean difference: 2.5
```

5.4 c) H_0: $\mu_D \leq 0$ H_1: $\mu_D > 0$

5.4 d) The test is done with `alternative = 'greater'`, which means that now R will test for standing greater than lying, which is quite improbable given the sample results.

```r
t.test(x = standing, y = lying, alternative = 'greater', mu = 0,
       paired = TRUE, conf.level = 0.925)
```

5.4 e) The p-value for this sample outcome is 0.07189, which is below the limit of 0.075 (92.5% confidence). H_0 is rejected with 92.5% confidence. The blood pressure is significantly higher lying down than standing up. There is a risk of 7.5% for type-I error.

```r
# Correct: alternative = 'less'
t.test(x = standing, y = lying, alternative = 'less', mu = 0,
       paired = TRUE, conf.level = 0.925)
```