

# Problem Set 4 Solution

Lars Lien Ankile

March 26, 2020

## Problem 1.

- a) I found the 95% confidence interval to be (95.6878, 96.5122) mg of caffeine. By the 95% confidence interval we mean the interval we with 95% confidence believe the true population mean to lie within. See below for R-calculations.

```
#use r as a calculator
x_bar <- 96.1
s <- 1.2
n <- 35

t_star <- qt(p = 0.975, df = n - 1)
m <- t_star * (s / sqrt(n))

sprintf("95% confidence interval: (%.4f, %.4f) mg caffeine.",
        x_bar - m, x_bar + m)
```

```
## [1] "95% confidence interval: (95.6878, 96.5122) mg caffeine."
```

- b) *State the hypotheses.*

The null hypothesis is that the mean is according to production standard,  $H_0 : \mu = 95$  mg. The alternative hypothesis is that the mean is higher than the standard,  $H : \mu > 95$  mg.

*Specify  $\alpha$ .*

I've chosen to use  $\alpha = 0.05$  as that is a pretty common choice.

*Calculate the t-statistic and p-value.*

The t-statistic can be calculated as follows.

$$\begin{aligned} t &= \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{96.1 - 95}{1.2/\sqrt{35}} \\ &= 5.423 \end{aligned}$$

This gives a p-value of  $p = P(T \geq 5.423) = 2.423e - 06$ . See below for calculations.

```
#use r as a calculator
mu_0 <- 95

# Calculate t-statistic
t <- (x_bar - mu_0) / (s / sqrt(n))
t

## [1] 5.423073

# Calculate p-value
p <- pt(t, df = n - 1, lower.tail = FALSE)
p
```

```
## [1] 2.422591e-06
```

*State the conclusion.*

Based on the above calculations, we see that we have a very small p-value  $p < 0.001$ , which is much smaller than our chosen alpha  $\alpha = 0.05$ , which means that it is highly unlikely that we could have observed what we did, given that our null-hypothesis, i.e. that the standards are met, is true. Therefore, we conclude that the production standards are likely not met.

- c) From the below calculation, we see that, if we consider this second case in isolation, the argument from above looks weaker. This is because the population mean of 95 mg is in the 95% confidence interval. Therefore, it's more likely based on this evidence that the observations are a product of pure chance and that the true mean is in fact 95 mg.

```
#use r as a calculator
x_bar <- 95.3
s <- 1.1
n <- 35

t_star <- qt(p = 0.975, df = n - 1)
m <- t_star * (s / sqrt(n))

sprintf("95%% confidence interval: (%.4f, %.4f) mg caffeine.",
        x_bar - m, x_bar + m)

## [1] "95% confidence interval: (94.9221, 95.6779) mg caffeine."
```

## Problem 2.

- a) **False.** The opposite, smaller margin of error, smaller confidence interval.
- b) **False.** We can be 100% confident that the mean of the sampled population will be in the confidence interval because the interval is centered on this exact mean.
- c) **False.** All we know is that in 95% of cases we would expect the true mean to lie in the confidence interval. We do not know anything for certain in this specific case.
- d) **False.** We only expect the full interval to contain the true mean 95% of the time and that doesn't say anything about the likelihood of any of the interior points of the interval being the true mean.
- e) **False.** We expect the true mean to be outside our confidence interval in 5% of cases, but that doesn't assign any probability of the true mean being outside any of the given confidence intervals we come up with. In any specific case the true mean is an actual number that is either inside or outside.
- f) **True.** Since 3.9 is outside our 95% confidence interval, that corresponds to a two-sided hypothesis test where  $p < \alpha$ , i.e. statistically significant according to that significance level  $\alpha$ .

**Problem 3.**

- a) If we choose  $\alpha = 0.05$  instead of  $\alpha = 0.10$  we make it more likely that we make type II errors compared to type I errors, i.e. it's more likely that we fail to identify a real effect than that we actually find evidence of an effect where there is none. In this specific case, I think that the potential reward for the company if the treatment is effective is huge. It's therefore more risky for them to miss out on a working drug than it is for them to incorrectly find a drug to be effective when it's not. It would therefore probably be advisable for them to use  $\alpha = 0.10$ .
- b) If we assume that the null hypothesis is that the new pills would cause no change in sleep hours, we see that the result is statistically significant if using  $\alpha = 0.05$  because 0 is outside the 95% confidence interval. The new evidence therefore suggests that the new formulation is a substantial improvement.

#### Problem 4.

- a) From the below R-code, I found the 95% confidence interval for the mean amount of minutes a patient has to wait to be (127.2581, 146.7419) minutes. The value of 127 therefore falls barely outside of this confidence interval and we would say that the change is statistically significant and the wait time has likely increased.

```
# Define some variables
x_bar <- 137
s <- 39
n <- 64

# Calculate t star
t_star <- qt(p = 0.975, df = n - 1)
# Find the margin
m <- t_star * (s / sqrt(n))

# Print the results
sprintf("95%% confidence interval: (%.4f, %.4f) minutes.",
        x_bar - m, x_bar + m)
```

```
## [1] "95% confidence interval: (127.2581, 146.7419) minutes."
```

- b) Since the old mean was right on the boundary of the previous confidence interval, 127 would be inside the new, broader 99% confidence interval, and the new measured wait time would not be judged to be statistically different from the old of 127 minutes.
- c) If the wait time actually changed, it's not certain that the administrator is at fault. It could e.g. just be that more people are getting hurt and needs the ER this year compared to last. Furthermore, it could be that the wait time did not increase in reality and it's just chance that makes it seem like that. By using  $\alpha = 0.05$ , we will get chance results posing as statistically significant 1 in 20 times, which can happen.

### Problem 5.

- a) From the blow calculations in R, i found the 95% confidence interval for babies' birth weight in North Carolina was (7.0074, 7.1946) pounds. This means that we're 95% confident that the true mean birth weight is in the confidence interval.

```
# Load the data
nc <- read.csv("datasets/nc.csv")
# Take out NA's of the data
nc_complete <- nc[complete.cases(nc$weight), ]

# Compute confidence interval
x_bar <- mean(nc_complete$weight)
s <- sd(nc_complete$weight)
n <- length(nc_complete$weight)

t_star <- qt(p = 0.975, df = n - 1)
m <- t_star * (s / sqrt(n))

# Print out the result
sprintf("95% confidence interval: (%.4f, %.4f) lb.", x_bar - m, x_bar + m)

## [1] "95% confidence interval: (7.0074, 7.1946) lb."
```

- b) *State the hypotheses.*

The null hypothesis is that the weight is not different from 7.5,  $H_0 : \mu = 7.5$ . The alternative hypothesis is that the weight is different from 7.5,  $H_A : \mu \neq 7.5$ .

*Specify  $\alpha$ .*

Using  $\alpha = 0.05$ .

*Calculate the t-statistic.*

The t-statistic is given by

$$\begin{aligned} t &= \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{5.128 - 7.5}{1.96958/\sqrt{152}} \\ &= -14.84515. \end{aligned}$$

```
# Get just the premature births
nc_premie <- nc_complete[nc_complete$premie == 'premie', ]
nc_premie <- nc_premie[complete.cases(nc_premie$weight), ]
x_bar <- mean(nc_premie$weight, na.rm = T)
s <- sd(nc_premie$weight)
n <- length(nc_premie$weight)
mu_0 <- 7.5
```

```
# Print some values for use in formulas above
sprintf("xbar: %.3f, s: %.5f, n: %d", x_bar, s, n)
```

```
## [1] "xbar: 5.128, s: 1.96958, n: 152"
```

```
# Calculate t-statistic
t <- (x_bar - mu_0) / (s / sqrt(n))
t
```

```
## [1] -14.84515
```

*Calculate the p-value.*

The p-value is  $p = P(T \geq 8.362264) + P(T \leq -8.362264) = 2P(|T| \leq -8.362264) = 2.598501e - 31$ .

```
# Calculate p-value
p <- 2 * pt(t, df = n - 1)
p
```

```
## [1] 2.598501e-31
```

*State the conclusion.*

From the above calculation, I found the p-value to be  $p < \alpha = 0.05$ , and in particular, our p-value is a lot smaller than our  $\alpha$ . This means that there's a very small chance that the null hypothesis is true which strengthens the alternative hypothesis. In this case, the birth weight of the prematurely born babies is statistically different from 7.5 lb.

c)

- i. In this case I would use a one-sided alternative hypothesis,  $H_A : \mu < 7.5$ . This is because it's only relevant to start interventions if the birth weight of these babies is below 7.5 pounds. If it's above 7.5 pounds that might also be a problem for all I know, but they for sure don't have problems with underweight babies in that case.
- ii. I think a significance level of  $\alpha = 0.10$  is advisable in this case because I think type II errors have more severe consequences than type I errors in this case. I.e. falsely determining that there's no problem with birth weight when there in fact is one, causing funding to not be given to babies and mothers in need is worse than potentially funding a program that isn't needed, in my opinion.



## Problem 6.

- a)
  - i. **Independent.** Here they're comparing the measurement of the salary of one group to the same measurement of the other group. There's no differences between two quantities for each individual involved.
  - ii. I think a study design using independent data is the most reasonable in this case because the question we're trying to answer is whether or not there exists differences between the groups. If it was possible to change the gender of people and measure their change in income, a study that produces paired data might be the best. To my knowledge, that is unfortunately not ethical nor practically feasible for such a study.
- b)
  - i. **Paired.** Here they're testing each individual twice, once before intervention and once after, and comparing the difference between the two measurements between individuals.
  - ii. The alternate study design would be to have a group of people who take vitamin E and a group of people who does not, and then compare the mean artery thickness between the groups. However, since the question we're trying to answer is whether vitamin E *increases* thickness, I think it makes more sense to use the paired data because that makes it easier to control for other effects and isolate the effect of starting to take the vitamin.
- c)
  - i. **Paired.** Again, they're taking two measurements for one individual and taking the difference between the two measurements to get the number that can be used for analysis.
  - ii. Here I also think the paired design is the most reasonable for answering whether one will *lose* weight by starting this diet. Its again easier to control and isolate if one can randomly select people to start and not start the diet and then measure the change in weight/body fat/etc.
- d)
  - i. **Independent.** Here, they're comparing each data point in one group to a corresponding data point in the other group.
  - ii. I think it makes much more sense to use paired data in this case. The absolute value of a stock doesn't mean much, but the relative change from day to day is more meaningful. One should therefore derive new data from the original by taking the difference between the price on day  $i + 1$  and day  $i$  and dividing by price day  $i$ . This gives the daily percentage change for each day for each stock. One can then calculate the mean daily return for each stock and compare.

## Problem 7.

a) *State the hypotheses.*

Let  $\mu_{ns}$  denote the mean expiratory volume for children with non-smoking parents and  $\mu_s$  the mean expiratory volume of children with parents who do smoke. The null hypothesis is that there's no difference between the means,  $H_0 : \mu_{ns} - \mu_s = 0$ . The alternative hypothesis is that there is a difference,  $H_A : \mu_{ns} - \mu_s \neq 0$ .

*Specify  $\alpha$ .*

I'm using  $\alpha = 0.05$  as that's a pretty standard significance level.

*Calculate the t-statistic.*

The t-statistic is calculated as follows.

$$t = \frac{(\bar{x}_{ns} - \bar{x}_s)}{\sqrt{\frac{s_{ns}^2}{n_{ns}} + \frac{s_s^2}{n_s}}} = \frac{(2.3 - 2.1)}{\sqrt{\frac{0.4^2}{20} + \frac{0.7^2}{23}}} = 1.168326$$

The value for the t-statistic came out to be  $t = 1.168326$ .

```
# Define relevant variables
x_bar_ns <- 2.3
x_bar_s <- 2.1
s_ns <- 0.4
s_s <- 0.7
n_ns <- 20
n_s <- 23

# Calculate t-statistic
t <- (x_bar_ns - x_bar_s) / sqrt(s_ns^2/n_ns + s_s^2/n_s)
t
```

```
## [1] 1.168326
```

*Calculate the p-value.*

We have a positive t-statistic, which means that we want to know the probability that the difference is to the right of the t-value, which we get by using `lower.tail = FALSE` with the `pt`-command in R. I multiply this number by two because we're interested in both tails of the t-distribution. The result is a p-value of  $p = 0.257$ .

```
# Calculate the p-value
p <- 2 * pt(t, df = min(n_ns, n_s) - 1, lower.tail = FALSE)
p
```

```
## [1] 0.2571266
```

*State the conclusion.*

Here the result is that our p-value is larger than our alpha,  $p = 0.2571266 > 0.05 = \alpha$ , which means that we don't have sufficient evidence to reject the null hypothesis, and the mean of the expiratory volume for each group is not statistically different.

- b) In this study, the expiratory volume in each group was so similar, while the variance between each measurement within each group was so large, that it seems very likely that the results we got was just the result of randomness, and not a result of an underlying difference in the expiratory volume between kids whose parents smoke or not.
- c) For one, more data generally means better results, and we might get a different result in the new study just because we have more random samples. However, I'm not sure whether household income would have any impact on childrens expiratory volume. One thought I have, though, is that lower-income parents probably have a higher probability of being smokers. Also, children living in "poor" households might be exposed to a wider range of things that could affect their expiratory volume, like mold, dust, etc. Therefore, I think one would see a smaller difference between the group if we control for income. That is because poor kids of non-smokers will presumably have the same amount of non-smoke things that are detrimental to their expiratory volume as children of smokers. This is controlled for in the latter study but not the former.