List of exercises

Note: *Please note when solving the problem set the conditions about the sample size to be met when the population is not Normal and/or when the population variance is unknown.*

** **Ex. 1 — Waiting at an ER, Part I.** A hospital administrator hoping to improve wait times decides to estimate the average emergency room waiting time at her hospital. He collects a simple random sample of 64 patients and determines the time (in minutes) between when they checked in to the ER until they were first seen by a doctor. A 95% confidence interval based on this sample is (128 minutes, 147 minutes), which is based on the normal model for the mean. Determine whether the following statements are true or false, and explain your reasoning for those statements you identify as false.

1. This confidence interval is not valid since we do not know if the population distribution of the ER wait times is nearly normal.
2. We are 95% confident that the average waiting time of these 64 emergency room patients is between 128 and 147 minutes.
3. We are 95% confident that the average waiting time of all patients at this hospital's emergency room is between 128 and 147 minutes.
4. 95% of such random samples would have a sample mean between 128 and 147 minutes.
5. A 99% confidence interval would be narrower than the 95% confidence interval since we need to be more sure of our estimate.
6. The margin of error is 9.5 and the sample mean is 137.5.
7. In order to decrease the margin of error of a 95% confidence interval to half of what it is now, we would need to double the sample size.

* **Ex. 2 — Stock Prices.** A random sample of the closing stock prices for the Oracle Corporation for a recent year (Source: Yahoo! Inc.) is shown below:

18.41	18.32	22.86	14.47	16.91	18.65
20.86	19.06	16.83	20.71	20.74	18.42
17.72	20.66	22.05	20.85	15.54	21.04
21.42	21.43	15.56	21.74	22.34	21.97
18.01	19.11	19.79	22.13	21.96	22.16
22.83	24.34	17.97	21.81		

(The data are also available in the Excel file `stock.xlsx`, under the tab `Stock`). Build 80%, 90%, and 95% confidence intervals for the population mean and interpret your findings.

* **Ex. 3 —** **Text Messaging.** A telecommunications company wants to estimate the mean length of time (in minutes) that 18-to 24-year-olds spend text messaging each day. In a random sample of twenty-seven 18-to 24-year-olds, the mean length of time spent text messaging was 29 minutes. From past studies, the company assumes that s is 4.5 minutes and that the population of times is normally distributed.
Determine a 95% confidence interval for the population mean.

** **Ex. 4 —** **Malaysian Airline.** A Malaysian Airline wanted to determine if customers would be interested in paying $10 flat fee for unlimited Internet access during long-haul flights. From a random sample of 200 customers, 125 indicated that they would be willing to pay this fee. Using this survey data, determine the 99% confidence interval estimate for the population proportion of the airline's customers who would be prepared to pay this fee for Internet use. Assume the sample is large.

** **Ex. 5 —** **Leak Rate.** LDS wants to be sure that the leak rate (in cubic centimeters per second) of transmission oil coolers (TOCs) meets the established specification limits. A random sample of 50 TOCs is tested, and the leak rates are recorded in the data file `tocs.xlsx`, under the tab `TOC`. Estimate the variance in leak rate with a 95% confidence level. Before you construct the confidence interval make sure that you check the *assumption of normality*.

** **Ex. 6 —** **Gender and Titanic.** Let's use the data set *titanic_train.xlsx*. In this dataset you will find a brief explanation of the variables. You can consider this dataset a large sample of a larger population. You can also assume that the conditions for inference are satisfied and that, therefore, the CLT holds. Use $\alpha = 0.05$ throughout your analysis. The variable 'Survived' takes value 1 if the passenger survived, and 0 otherwise. Calculate a confidence interval for the proportion of males who survived, and then repeat the analysis for females.

** **Ex. 7 —** **Class and Titanic.** Let's use the data set *titanic_train.xlsx* again. Calculate a confidence interval for the proportion of first class passengers who survived, and then repeat the analysis for the second and the third class.

Answer of exercise 1

1. False. Provided the data distribution is not very strongly skewed ($n = 64$ in this sample, so we can be slightly lenient with the skew), the sample mean will be nearly normal, allowing for the method normal approximation described.
2. False. Inference is made on the population parameter, not the point estimate. The point estimate is always in the confidence interval.
3. True.
4. False. The confidence interval is not about a sample mean.
5. False. To be more confident that we capture the parameter, we need a wider interval. Think about needing a bigger net to be more sure of catching a fish in a murky lake.
6. True. Optional explanation: This is true since the normal model was used to model the sample mean. The margin of error is half the width of the interval, and the sample mean is the midpoint of the interval.
7. False. In the calculation of the standard error, we divide the standard deviation by the square root of the sample size. To cut the SE (or margin of error) in half, we would need to sample $2^2 = 4$ times the number of people in the initial sample.

Answer of exercise 2

The sample size is greater than 30 so we can assume normality. However, the sample size is less than 40 and the population variance is unknown, so we should use the t distribution. Let us first compute the statistics of the sample:

- $\overline{x} = 19.961$
- $s = 2.402$
- $n = 34$

Let us use the following formula for the computation of CI:

$$\overline{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Thus, let us define:

α	$t_{\alpha/2}$	Lower	Upper
0.2	1.3077	19.4221	20.4996
0.1	1.6924	19.2637	20.6581
0.05	2.0345	19.1227	20.7990

As expected, we can notice that the interval gets wider when the confidence level $1 - \alpha$ increases.

Answer of exercise 3

Let us compute the CI using:

$$\overline{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

where $\alpha = 0.05$. Thus, we get:

- $\overline{x} = 29$
- $s = 4.5$
- $t_{\alpha/2,26} = 2.0555$
- $s/\sqrt{n} = 4.5/\sqrt{27} = 0.866$ (standard error)

Thus, we get a 95% CI as:

$$[27.2199, 30.7801]$$

Answer of exercise 4

Let us define:

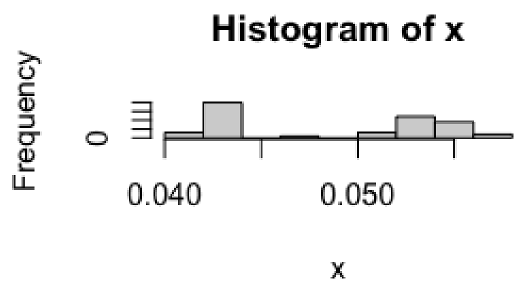
- $n = 200$
- $x = 125$
- $\hat{p} = x/n = 125/200 = 0.625$

In addition, we can easily verify that both $n\hat{p} \geq 15$ and $n\hat{q} \geq 15$ hold. To compute a 95% CI, we get:

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.625 \pm 2.58 \sqrt{\frac{0.625 \times 0.375}{200}} \rightarrow [0.5368, 0.7132]$$

Answer of exercise 5

1. Let us first check whether the data is normally distributed. We can use a histogram to visualize the distribution of the data.



It does not really seem to be normally distributed. One might wonder whether the approach of building confidence intervals using the CLT is valid in this case. For the sake of the exercise, let us continue here computing the interval for the variance.

2. From the sample we obtain the following sample statistics:
   - $n = 50$
   - $s^2 = 0.0000282$
   - $df = n - 1 = 49$

   We calculate the factors of reliability:
   $\chi^2_{n-1,\frac{\alpha}{2}} = \chi^2_{49,0.025} = 70.22$
   $\chi^2_{n-1,1-\frac{\alpha}{2}} = \chi^2_{49,0.975} = 31.55$
   Thus, a 95% CI can be obtained as:

$$\frac{0.0000282 \times 49}{70.22} \leq \sigma^2 \leq \frac{0.0000282 \times 49}{31.55}$$
$$0.0000197 \leq \sigma^2 \leq 0.0000438$$

Answer of exercise 6

The 95% CI for females is:

$$[0.5368, 0.7132]$$

The 95% CI for males is:

$$[0.157, 0.221]$$

Note that both confidence intervals do not overlap so they are different.

Answer of exercise 7

The 95% CI for first class passengers is:

$$[0.565, 0.694]$$

The 95% CI for second class passengers is:

$$[0.401, 0.545]$$

The 95% CI for third class passengers is:

$$[0.204, 0.280]$$

Note that the confidence intervals do not overlap so they are different.