Understanding Type I Errors, Type II Errors, and P-values

Prerequisites:

- Know that a one-sample z-test uses a random sample to investigate whether the population mean is different from a specified value.
- Understand the concepts of null and alternative hypotheses, and know how to formulate them in the context of a one-sample z-test.
- Be familiar with the concept of significance level, and its role in carrying out a hypothesis test.
- Know that a z-test is carried out by first calculating a test statistic, then using it to compute a P-value, and finally using the P-value to decide whether to reject the null hypothesis or not.

Learning Outcomes:

- Given a null and alternative hypotheses, identify how a type I and a type II error could occur in the context of the hypothesis test.
- Describe the relationship between the significance level and the probability of a type I error occurring.
- Know how to use the correct hypothesis testing terminology (i.e. reject or fail to reject the null hypothesis) and how to interpret this terminology in context.
- Be able to compute the P-value of a z-test and explain how to use it to reject (or fail to reject) the null hypothesis at a given significance level.

In this activity, we will learn about type I and type II errors, the relationship between type I errors and the significance level, and how to calculate P-values in the context of a one-sample z-test. Throughout the activity we will use two online resources, which are described in more detail in the sections which require their use.

Recall that a z-test is a hypothesis test in which the test statistic follows a standard Normal distribution. In this activity, you will be testing whether the mean of a Normally-distributed population is equal to a specific value μ_0 or different from it (that is, the test is two-sided). The value μ_0 is determined by the experimenter before collecting data. For this test, the test statistic is defined as

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}},$$

where $\bar{X}$ is the sample mean, μ_0 is the null population mean, σ is the population standard deviation, and n is the sample size. When the null hypothesis is true, the test statistic Z follows a standard Normal distribution, $\mathrm{N}(0,1)$.

The P-value of a z-test is calculated as the probability of observing data more extreme than what was actually observed, assuming that the null hypothesis is true.

In practice, we can use the standard Normal distribution tables to determine how likely it is to have observed a value larger than $|Z|$ or smaller than $-|Z|$; assuming the null hypothesis is true, this value is called the P-value of the test. Before carrying out this calculation, however, the practitioner must determine the *significance level* α at which they wish to carry out the test.

If the P-value is smaller than α, then the null hypothesis is *rejected*. Else, the practitioner *fails to reject* the null hypothesis.

Caution: neither the null nor the alternative hypotheses are ever "accepted"—the null is either rejected or not. Once the test has been carried out, there are four possible outcomes. These depend on whether H_0 is true or false and whether H_0 was rejected or not. If we erroneously reject the null but the null is true, a *type I error* occurred. If we erroneously fail to reject the null even when it is false, a *type II error* occurred. The table below summarises all the options:

	Reality	
Test conclusion	H_0 true	H_0 false
Reject H_0	Type I Error	No error
Fail to reject H_0	No error	Type II Error

In practice, it is not possible to know whether a type I or type II error occurred because we don't know the true value of the parameter. However, practitioners can take measures to make one of these errors less likely to occur.

1 Understanding type I and type II errors in practice

For this part of the activity, you will be presented with different examples of hypothesis tests. Answer the questions in the context of each scenario. Recall that the null hypothesis is specific and corresponds with to status quo—you will obtain data and determine if there is evidence to reject the null hypothesis.

Scenario A: Jury trials refer to a legal proceeding in which a set of jurors—the jury—decide whether an accused person is innocent or not based on evidence presented to the jury. Suppose that you are part of a jury tasked with determining whether an accused person is innocent or guilty of a crime.

A1) Formulate, without using technical terms, the null and alternative hypotheses. (*Hint: the accused is always assumed to be innocent unless there is evidence to reject their innocence.*)

A2) A type I error would refer to erroneously concluding the accused is not innocent when, in fact, they are. Based on your answer to question A1), briefly explain using statistical terminology why this is a type I error.

A3) A type II error would refer to erroneously concluding the accused is innocent when, in fact, they are not. Based on your answer to question A1), briefly explain using statistical terminology why this is a type II error.

A4) In this scenario, do you think it is more serious when a type I error occurs or when a type II error occurs? Explain briefly.

Scenario B: You want to investigate whether a coin is fair or not, that is, whether when tossed it is equally likely to land on heads or on tails. For this purpose, you will flip the coin 10 times, record how many times you tossed heads, and use this information to determine whether there is evidence to conclude that the coin is not fair. Below, match each statement on the left with the correct description on the right, in the context of this scenario.

The coin is fair.	No error.
The coin is not fair.	Type II error.
Concluding the coin is fair when the coin is fair.	Alternative hypothesis.
Concluding the coin is fair when the coin is not fair.	Type I error.
Concluding the coin is not fair when the coin is fair.	No error.
Concluding the coin is not fair when the coin is not fair.	Null hypothesis.

Scenario C: You are assisting a team of engineers at a chemical plant to determine whether the temperature at which their chemical process runs is equal to (or different from) 55°C.

C1) Formulate the null and alternative hypotheses, and make sure to define any parameters that you use.
(*Hint: you will assume the process runs at 55°C unless there is evidence to the contrary.*)

C2) Explain, using non-technical words, when would a type I error occur in this scenario.

C3) Explain, using non-technical words, when would a type II error occur in this scenario.

C4) Fill in the table below using your answers to the three previous questions. Each of the four entries should be either *type I error*, *type II error*, or *no error*.

	Reality	
Test conclusion	Temperature = 55°C	Temperature ≠ 55°C
Conclude Temperature = 55°C		
Conclude Temperature ≠ 55°C		

2 Calculating P-values to carry out hypothesis tests

For this part of the activity, you will use an online interactive resource, which can be accessed at

`https://homepage.divms.uiowa.edu/~mbognar/applets/normal.html`

This resource allows you to specify the mean and standard deviation of a Normal distribution. You can also specify a value x, in which case the resource calculates the probability of the Normal distribution exceeding that value. Alternatively, you can specify a probability and the resource will calculate the upper quantile of that order. You can also calculate the probability of being below a given value, or the lower quantile of a given order, by selecting $P(X < x)$ instead of $P(X > x)$ in the dropdown menu.

We want to investigate the average concentration μ in moles per liter (mol/L) of a given chemical in the soil of a garden in the (fictional) City of Theed. Assume that the concentration of the chemical in a given sample of soil follows a Normal distribution with standard deviation $\sigma = 1.7$ mol/L. We want to test the null hypothesis that

$\mu = 14$ mol/L against $\mu \neq 14$ mol/L at a significance level of $\alpha = 0.05$. For this purpose, you will obtain $n = 25$ samples of soil from the garden.

1) Write down the null and alternative hypotheses.
2) Suppose you observe data with a sample mean of $\bar{X} = 14.5$ mol/L. Intuitively, without any calculation, do you consider this to be close enough to the value of the null population mean?
3) For these data, what is the value of the test statistic Z?
4) What is the distribution of the test statistic if the null hypothesis is true?
5) Does the distribution you answered in question 4) depend on the observed data? Explain briefly.
6) Modify the necessary settings in the resource to reflect the distribution of the test statistic under the null hypothesis. Calculate, using the resource, the probability of observing a value larger than $|Z|$.
7) Calculate, using the resource, the probability of observing a value smaller than $-|Z|$.
8) Comparing your answers to questions 6) and 7), the two probabilities are ______________.

   (*Fill in the blank with one of <u>equal</u>, <u>different</u>, or <u>not comparable</u>.*)
9) The P-value for this sample is equal to the sum of the probabilities you calculated in questions 6) and 7). Based on your answer to question 8), calculate the P-value using only your answer to question 6).
10) Can you calculate the P-value in question 9) using another option in the dropdown menu of the resource?
11) Based on your answer to question 9), do you reject the null hypothesis? State your conclusion in a complete sentence, in the context of the original research question, and including the significance level. Be careful to use the correct terminology, as mentioned in the preamble to the activity.
12) What type of error, if any, could occur as a result of your answer to question 11)?
13) Is your answer to question 11) in line with your assessment of question 2)?

14) Now suppose you observe a new data set with a sample mean of $\bar{X} = 12.8$ mol/L. Ignoring the original data set of question 2), calculate the P-value and report your conclusion in a complete sentence, as in question 11).

15) What type of error, if any, could occur as a result of your answer to question 14)?

3 The relationship between significance and type I errors

For this part of the activity, you will use an online interactive resource, which can be accessed at

http://digitalfirst.bfwpub.com/stats_applet/stats_applet_14_signif.html

This resource allows you to specify the values of the null population mean μ_0, the population standard deviation, the significance level, and the sample size. You can also specify the true value of the population mean, which can equal μ_0 or not. For this setting, you can simulate a sample from the true population and the resource calculates the P-value. Alternatively, if you have your own data, you can input the observed value of the sample mean and the resource calculates the P-value.

For a Normally-distributed population, we want to test the null hypothesis that $\mu = 0$ against the alternative hypothesis that $\mu \neq 0$ at a significance level of $\alpha = 0.2$. The population standard deviation is known to be $\sigma = 1$, and you will obtain a sample of size $n = 8$. Use the controls on the left side of the resource to reflect this and check the option that says *"The truth about the population is $\mu = 0$."*

1) For these settings, the null hypothesis is ________.
(*Fill in the blank with one of <u>true</u> or <u>false</u>.*)

2) For these settings, can a type I error occur? Can a type II error occur? Briefly explain.

3) For what range of P-values would a type I error occur?

4) If you click on the NEW SAMPLE button, the resource will generate a sample from the null distribution and calculate the corresponding P-value.[1]

[1] CAUTION: since the sample is simulated, it will differ from a classmate's, or even from what you might see if you repeated the task. Try this! Click on the NEW SAMPLE button a couple of times and keep track of how the plot is modified each time a new sample is simulated.

4.1) In the next question, you will generate five samples. Before doing that, in how many of those five would you expect a type I error to occur? To what proportion does this number correspond?

4.2) Generate five samples and count in how many of them a type I error occurs. What proportion of the five samples led to a type I error?

(*Hint: what is the relationship between a type I error occurring and the vertical blue line being in the yellow region of the plot?*)

5) Repeat the experiment of question 4) 19 more times. Use the table at the end of the activity to write down how many samples led to a type I error in each 5-sample batch. Include the results from question 4) as well.
(*Using sample batches makes the process faster!*)

6) From the total of 100 samples that you generated, what proportion of them led to a type I error?

7) Compare the proportion you obtained in question 6) with the significance level α. Are these two values close?

8) Suppose that you now want to test at a significance level of $\alpha = 0.1$. Repeat the experiment from questions 4) to 6) with this new value of α.

9) Compare the new value of α with the proportion you obtained in question 8). Are these two values close?

10) Based on your answer to questions 7) and 9), the probability of a type I error occurring and the significance level are ______________.
(*Fill in the blank with one of equal, different, or not comparable.*)

4 Applied exercise

For this part of the activity, you will study the age at which the players of Naboo's[2] national sports teams specialized in their respective sports—that is, stopped playing other organized sports to focus solely on their sport. For this purpose, you will obtain a sample of 15 athletes and ask them the age at which they specialized in the sport they currently compete in. A census from 10 years ago concluded that the average age was 12.1 years old, with a standard deviation of 1.4 years. You will assume that the standard deviation is still 1.4. Your objective is to determine if the average age of specialization is still 12.1 years, or if it has changed. For this purpose, you will conduct a z-test at a significance level of 0.1.

[2] Naboo is a fictional planet in a galaxy far, far away.

1) State the null and alternative hypotheses. Define any variables, parameters, and quantities you use.

2) Suppose that you obtained the following data:

   11.1, 13.6, 14.3, 14.5, 13.7, 13.8, 12.2, 12.3, 15.1, 10.4, 11.7, 14.4, 14.9, 11.2, 14.8.

   What is the sample mean?

3) What is the value of the test statistic?

4) Using an online resource, calculate the P-value.

5) Based on the P-value, do you reject the null hypothesis? State your conclusion in a complete sentence, including the significance level. Be careful to use the correct terminology, as mentioned in the preamble to the activity.

6) What type of error, if any, could occur as a result of your answer to question 5)?

7) What is the probability that the error you answered in question 6) occurs?

8) Summarise your conclusion to the original research question in a couple of sentences.

Batch no.	Rejections for $\alpha = 0.2$	Rejections for $\alpha = 0.1$
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		