Lecture 28 Worksheet

Chrysafis Vogiatzis

Every worksheet will work as follows.

1. You will be entered into a Zoom breakout session with other students in the class.
2. Read through the worksheet, discussing any questions with the other participants in your breakout session.
   - You can call me using the "Ask for help" button.
   - Keep in mind that I will be going through all rooms during the session so it might take me a while to get to you.
3. Answer each question (preferably in the order provided) to the best of your knowledge.
4. While collaboration between students in a breakout session is highly encouraged and expected, each student has to submit their own version.
5. You will have 24 hours (see Compass) to submit your work.

Worksheet 1: Comparing means

A pharmaceutical company is researching a new drug that has been cleared for human testing. This new drug (if cleared after testing) will replace a previously used drug that had a side effect: it impaired driving.

Two samples are selected and given the old drug (sample 1 of size $n_1 = 15$) and the new drug (sample 2 of size $n_2 = 10$). The reaction times while driving of people in the first sample ended up begin an average of 4.65 seconds with a standard deviation of 0.5 seconds. For the second sample, the same numbers were 4.36 seconds and 0.3 seconds, respectively.

You may assume that reaction times while driving are **normally distributed**.

Problem 1: Formulating the hypothesis

Formulate a suitable hypothesis to accept or reject that the second drug leads to different reaction times.

Answer to Problem 1.

H_0 :

H_1 :

Problem 2: Formulating the hypothesis (correctly)

What if we want to check whether the new drug *improves* the side effect? That is, we want to check whether the reaction times are better than the ones of the old drug.

Answer to Problem 2.

H_0 :

H_1 :

This is as good a time as any to observe one item of importance. When formulating a hypothesis, if we are interested in proving a claim, then we typically set it as the alternative hypothesis! The reason is that rejecting the null hypothesis is a stronger conclusion; rejecting the null in favor of the alternative essentially implies that we prove the alternative.

Hence, before you proceed, please verify that you have the following hypothesis setup:

$$\boxed{H_0 : \mu_1 = \mu_2 \ \text{ vs. } \ H_1 : \mu_1 > \mu_2.}$$

Problem 3: Known variances

Assume we know the variances of the reaction times when driving are known and equal to $\sigma_1^2 = 0.09$ seconds^2 for the first drug and $\sigma_2^2 = 0.16$ seconds^2 for the second drug. Using $\alpha = 0.05$, should you accept or reject the null hypothesis? Does the second drug lead to better side effects (=faster reaction times) or not? What is the P-value?

Answer to Problem 3.

Problem 4: β errors

It is not an easy feat to use the Student's T distribution (which appears alongside unknown variances) to calculate β errors and P-values. On the other hand, when the variances are known, and we have a standard normal distribution, calculations become easier.

With that in mind, assume you are formulating a hypothesis where the null hypothesis is that the two means are the same (i.e., $H_0 : \mu_1 - \mu_2 = 0$) vs. an alternative hypothesis that $\mu_1 < \mu_2$ (i.e., $H_1 : \mu_1 - \mu_2 < 0$). What is the β error of accepting the null hypothesis assuming that the true difference is $\mu_1 - \mu_2 = -0.3$? [1]

[1] This means that the second drug is actually worsening reaction times while driving by 0.3 seconds.

Answer to Problem 4.

Worksheet 2: Unknown variances

In this set of exercises, we will use the same data as in Worksheet 1. However, we will no longer assume that the variances are known.

Problem 5: Unknown (but equal!) variances

What if we have no idea what the true variances are, but we know they are supposed to be equal to one another? Then, using $\alpha = 0.05$, should you accept or reject the null hypothesis? [2]

[2] Recall that we have been told that the sample standard deviations for the two samples were $s_1 = 0.5$ seconds and $s_2 = 0.3$ seconds.

Answer to Problem 5.

Problem 6: Unknown (and not necessarily equal) variances

Continuing on the same line of logic, what if we have the most general case? Assume now that not only you do not know the true variances, but you also do not know whether or not they are equal to one another.

What would you deduce in this case? Under $\alpha = 0.05$, should you reject the null hypothesis?

Answer to Problem 6.

Worksheet 3: The "paired" t-test

The difference of two independent normally distributed random variables is also normally distributed. We have used this fact in many of our derivations.

Now, consider two independent and normally distributed populations with unknown variances σ_1^2 and σ_2^2. We get a random sample $X_1, X_2, \ldots, X_n$ from the first population and a random sample $Y_1, Y_2, \ldots, Y_n$ from the second population. Note how both samples are of equal size n. Now, consider $W_i = X_i - Y_i$. Clearly, if X and Y have the same mean, then we should expect each of the W_i to be small, no?

Problem 7: Devising a hypothesis testing procedure

Based on that, can you recommend a hypothesis test to accept or reject that the two populations have the same mean? Explain what the setup of the hypothesis test is, which statistic you would use, and when you would reject or fail to reject the null hypothesis. [3]

[3] $H_0 : \mu_1 - \mu_2 = 0$ vs. $H_1 : \mu_1 - \mu_2 \neq 0$.

Answer to Problem 7.