Lecture 23 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 battery lives

When a new smartphone, tablet, laptop is released, one of the things that everyone wants to know is how good their batteries are. Assume a new phone has just come out and we want to see whether the new battery is better than the old one.

For the rest of Worksheet 1, we assume that the standard deviation of the previous model battery life is equal to $\sigma_1 = 1$ hour, and the standard deviation of the new model batter life is $\sigma_2 = 0.6$ hours.

Problem 1: Pooled standard deviation

What is the pooled standard deviation s_P, assuming that we have a sample of $n_1 = 12$ previous model phones, and $n_2 = 20$ new model phones?

Answer to Problem 1.

$\sigma_P =$

Problem 2: Computing a 95% confidence interval

What is the (two-sided) 95% confidence interval on the difference between the two means $\mu_2 - \mu_1$? You may assume that you calculated the averages $\overline{X}_1 = 22.5$ hours and $\overline{X}_2 = 23$ hours. [1]

[1] Do we need a z or a t critical value here? No matter which one you need, recall there are z and t tables in Lecture 20!

Answer to Problem 2.

Problem 3: Comparing battery lives

When comparing the two battery lives, can you make the claim (with 95% confidence) that the new model has better battery life than the older model? Why/Why not? Briefly explain in a sentence or two. [2]

[2] Does the confidence interval include cases where the new phone is better? How about cases where the old phone is better?

Answer to Problem 3.

This is a nice consequence of confidence intervals. When comparing two means, if the confidence interval contains both positive and negative values, this means that the believable range of values for the difference contains values where population 1 or population 2 are bigger. In our case, then, since the confidence interval contains negative values, this means that we cannot be sure (with 95% confidence) that the claim that the new phone has better battery life is true.

Worksheet 2: Unknown variances

In the previous worksheet, we assumed that the true standard deviations were known. What if this is not true?

Problem 4: Using the sample standard deviations

We use the same samples as before: we got $n_1 = 12$ previous model phones, and $n_2 = 20$ new model phones. These two samples led to averages $\overline{X}_1 = 22.5$ hours and $\overline{X}_2 = 23$ hours as well as sample standard deviations $s_1 = 1.6$ hours and $s_2 = 1.3$ hours. What is the estimated pooled standard deviation now?

Answer to Problem 4.

$s_p =$

Problem 5: A 95% confidence interval (again)

Construct a two-sided 95%-confidence interval on the difference between the two means $\mu_2 - \mu_1$. [3]

Answer to Problem 5.

[3] Do we need a z or a t critical value here? And, if we need a t value, what are the degrees of freedom?

Worksheet 3: The F distribution and ratios of variances

During Lecture 23, we saw the "weird" F distribution. As a reminder, it is useful for constructing confidence intervals for the ratio of two variances. Formally the F distribution describes the ratio of two χ^2 random variables. Before we put it to the use (to construct a confidence interval), let's find some values.

Problem 6: Values straight from the F table

Let us experiment reading the table. What are the f values for:

Answer to Problem 6.

- 12 degrees of freedom in the numerator, 15 degrees of freedom in the denominator, and probability p equal to 0.95?
- 5 degrees of freedom in the numerator, 5 degrees of freedom in the denominator, and probability p equal to 0.90?
- 15 degrees of freedom in the numerator, 12 degrees of freedom in the denominator, and probability p equal to 0.95?

Problem 7: Values that do not exist in the F table

What about the following values that do **not** exist in the F table? Recall that you can use the property that $f_{u,v,1-p} = \frac{1}{f_{v,u,p}}$.

Answer to Problem 7.

- 12 degrees of freedom in the numerator, 15 degrees of freedom in the denominator, and probability p equal to 0.05?
- 5 degrees of freedom in the numerator, 5 degrees of freedom in the denominator, and probability p equal to 0.10?
- 15 degrees of freedom in the numerator, 12 degrees of freedom in the denominator, and probability p equal to 0.05?

Problem 8: Confidence intervals on ratios of variances

Let us go back to the new cell phone vs. the old cell phone batteries. Another important characteristic is how variable the battery life itself is. We have now collected a sample of $n_1 = 16$ new phones and $n_2 = 13$ old ones, which resulted in sample variances equal to $s_1^2 = 2$ hours^2 and $s_2^2 = 3.81$ hours^2 respectively. What is the 90% confidence interval on the ratio of the true variances? You may assume that both battery lives (for new and older version models) are normally distributed.

Answer to Problem 8.

Worksheet 4: Election Day 2020

Tomorrow is election day! Recently, there has been quite the influx of new polls from some of the "battleground" states. One of them, our neighboring Iowa, had multiple polls from well-respected pollsters show up this past weekend. Let us try to analyze one of them with the help of a confidence interval.

Problem 9: Comparing proportions

The poll in question asked $n_1 = 444$ female likely voters and $n_2 = 409$ male likely voters [4]. Among the first sample, the observed population voting for a candidate was 188 (so, $\hat{p}_1 = \frac{188}{444} = 0.4$). In the second sample, the same population was found to be 233 (and hence, $\hat{p}_2 = \frac{233}{409} = 0.57$).

Can we make the claim (with 95% confidence) that the two populations view this election similarly? Construct the 95% confidence interval on $p_1 - p_2$ to help you answer the question.

[4] The poll separated voters only for these two identities. Had the poll included more options, the results would have definitely been more characteristic of the true population.

Answer to Problem 9.