Worksheet 05 (Solutions)

1. We will start with the easier, though much less commonly used, task of comparing the variances between the samples before moving onto the confidence interval of the mean. Using what we know about the distributions of S_X^2 and S_Y^2, build a pivot based on the scaled ratio of these two quantities that has an F-distribution.

Solution: We know that the following hold:

$$\frac{(n-1)S_X^2}{\sigma_X^2} \sim \chi^2(n-1)$$
$$\frac{(m-1)S_Y^2}{\sigma_X^2} \sim \chi^2(m-1)$$

Given the definition of the F-statistic, then, we have:

$$\frac{\frac{(n-1)S_X^2}{\sigma_X^2} \cdot \frac{1}{n-1}}{\frac{(m-1)S_Y^2}{\sigma_Y^2} \cdot \frac{1}{m-1}} = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F(n-1, m-1)$$

2. Rearrange your previous result to get a confidence interval for the ratio σ_Y^2/σ_X^2. Note that the F distribution is not symmetric.

Solution: Starting with the previous, we have (I will avoid writing the degrees of freedom for the F-values as it's more confusing than helpful):

$$\mathbb{P}\left[f_{1-\alpha/2} \leq \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \leq f_{\alpha/2}\right] = 1-\alpha$$
$$\mathbb{P}\left[\frac{S_Y^2}{S_X^2} \cdot f_{1-\alpha/2} \leq \frac{\sigma_Y^2}{\sigma_X^2} \leq \frac{S_Y^2}{S_X^2} \cdot f_{\alpha/2}\right] = 1-\alpha$$

And that's it! Those are the upper and lower bounds of the confidence interval.

3. What is $\mathbb{E}[\bar{X} - \bar{Y}]$? Make use of the properties that we already know to make this relatively easy. You should see that this is an unbiased estimator of the difference in the means.

Solution: This should be easy based on the rules for expected values and the fact that we know $\mathbb{E}\bar{X} = \mu_X$ and $\mathbb{E}\bar{Y} = \mu_Y$. Namely:

$$\begin{aligned}\mathbb{E}[\bar{X} - \bar{Y}] &= \mathbb{E}\bar{X} - \mathbb{E}\bar{Y} \\ &= \mu_X - \mu_Y.\end{aligned}$$

4. What is $\mathrm{Var}[\bar{X} - \bar{Y}]$? Make use of the properties that we already know to make this relatively easy. The result should imply that the difference is a consistent estimator of the difference in sample means.

Solution: Same idea. We know that $\mathrm{Var}\bar{X} = \sigma_X^2/n$ and therefore that $\mathrm{Var}\bar{Y} = \sigma_Y^2/m$. So:

$$\begin{aligned}\mathrm{Var}[\bar{X} - \bar{Y}] &= \mathrm{Var}\bar{X} + \mathrm{Var}[-1 \cdot \bar{Y}] \\ &= \mathrm{Var}\bar{X} + \mathrm{Var}[\bar{Y}] \\ &= \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\end{aligned}$$

5. If $\mathcal{G}_X$ and $\mathcal{G}_Y$ are both normally distributed, then $\bar{X} - \bar{Y}$ also has a normal distribution. As we did in the one-sample case, construct a pivot Z that scales this difference to have a standard normal distribution.

Solution: You should know that if we subtract off the mean and divide by the standard deviation of a normal distribution we get a standard normal. So, from the previous results:

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim N(0, 1)$$

6. Assume that $\sigma_X^2 = \sigma_Y^2$, which we can write as just σ^2. Take the definition of the pooled sample variance and multiply both sides by $(n + m - 2)$ and divide by σ^2. If we assume that $\mathcal{G}_X$ and $\mathcal{G}_Y$ are both normally distributed, show that S_p^2 is a scaled version of a chi-squared distribution. What are its degrees of freedom?

Solution: Multiplying the definition throughout by $(n + m - 2)$ and dividing by σ^2 we have:

$$\frac{(n + m - 2)S_P^2}{\sigma^2} = \frac{(n-1)S_X^2}{\sigma^2} + \frac{(m-1)S_Y^2}{\sigma^2}$$

The second term has a $\chi^2(n-1)$ distribution and the third term has a $\chi^2(m-1)$ distribution. They are based on independent samples and are therefore independent. The sum independent chi-squared random variables is another chi-squared with the sum of the degrees of freedom. So:

$$\frac{(n+m-2)S_P^2}{\sigma^2} \sim \chi^2(n+m-2).$$

7. Put together the previous results to generate a pivot T that has a T-distribution. Rearrange the terms to get a confidence interval for the difference in means.

Solution: Now, we just divide the Z statistic above with the scaled version of S_P divided by its degrees of freedom:

$$\begin{aligned} T &= \frac{\frac{(\bar{X}-\bar{Y})-(\mu_X-\mu_y)}{\sqrt{\frac{\sigma^2}{n}+\frac{\sigma^2}{m}}}}{\frac{(n+m-2)S_P^2}{\sigma^2} \times \frac{1}{(n+m-2)}} \\ &= \frac{\frac{(\bar{X}-\bar{Y})-(\mu_X-\mu_y)}{\sqrt{\frac{1}{n}+\frac{1}{m}}}}{S_P^2} \\ &= \frac{(\bar{X}-\bar{Y})-(\mu_X-\mu_y)}{\sqrt{S_P^2\left[\frac{1}{n}+\frac{1}{m}\right]}} \end{aligned}$$

The statisic T abnove has a T distribution with $n+m-2$ degrees of freedom by its construction.