

# MATH 369/650 Fall 2020 Homework #3

## [INCOMPLETE]

Professor Adam Kapelner

Due by email noon Friday, October 2, 2020

(this document last updated Thursday 17<sup>th</sup> September, 2020 at 3:15pm)

### Instructions and Philosophy

The path to success in this class is to do many problems. Unlike other courses, exclusively doing reading(s) will not help. Coming to lecture is akin to watching workout videos; thinking about and solving problems on your own is the actual “working out.” Feel free to “work out” with others; **I want you to work on this in groups.**

Reading is still *required*. For this homework set, review Math 241 concerning the normal distribution. Then read about the class topics in the two recommended textbooks and online.

The problems below are color coded: **green** problems are considered *easy* and marked “[easy]”; **yellow** problems are considered *intermediate* and marked “[harder]”, **red** problems are considered *difficult* and marked “[difficult]” and **purple** problems are extra credit. The *easy* problems are intended to be “giveaways” if you went to class. Do as much as you can of the others; I expect you to at least attempt the *difficult* problems. “[MA]” are for those registered for the 600-level class and extra credit otherwise.

This homework is worth 100 points but the point distribution will not be determined until after the due date. See syllabus for the policy on late homework.

Up to 7 points are given as a bonus if the homework is typed using L<sup>A</sup>T<sub>E</sub>X. Links to installing L<sup>A</sup>T<sub>E</sub>X and program for compiling L<sup>A</sup>T<sub>E</sub>X is found on the syllabus. You are encouraged to use **overleaf.com**. If you are handing in homework this way, read the comments in the code; there are two lines to comment out and you should replace my name with yours and write your section. The easiest way to use overleaf is to copy the raw text from hwxx.tex and preamble.tex into two new overleaf tex files with the same name. If you are asked to make drawings, you can take a picture of your handwritten drawing and insert them as figures or leave space using the “\vspace” command and draw them in after printing or attach them stapled.

The document is available with spaces for you to write your answers. If not using L<sup>A</sup>T<sub>E</sub>X, print this document and write in your answers. I do not accept homeworks which are *not* on this printout. Keep this first page printed for your records.

NAME: \_\_\_\_\_

## Problem 1

In lecture 6-7, we did two-sided two-sample z and t tests. We will repeat these tests now but do them one sided. For extra practice, I will make them left-sided. To do this, I will switch the indexing of the two populations. The female population is now considered population #1 and the male population is now considered population #2. We assume the DGP for female height measurements is  $\overset{iid}{\sim} \mathcal{N}(\theta_1, \sigma_1^2)$  independent of the DGP for male height measurements assumed to be  $\overset{iid}{\sim} \mathcal{N}(\theta_2, \sigma_2^2)$ .

The sample sizes, point estimates for the mean and point estimates for the variance computed from the in-class student survey are:

$$\begin{aligned}n_1 &= 6 \\ \bar{x}_1 &= 62.3 \\ s_1^2 &= 2.25^2 \\ n_2 &= 10 \\ \bar{x}_2 &= 70.5 \\ s_2^2 &= 2.07^2\end{aligned}$$

The theory we wish to prove is that females are shorter than males. We will do so via hypothesis testing at size  $\alpha = 5\%$ . We will test this under many different assumption scenarios about the variances in the DGP. The results should be approximately the same.

- (a) [easy] Write the alternative and null hypotheses. Remember that  $\theta_1$  and  $\theta_2$  are now switched from class. We will use these hypotheses for the four scenarios in the rest of the problem.

---

We will first assume that the variances are known to be  $\sigma_1^2 = 3.5^2$  and  $\sigma_2^2 = 4^2$ .

- (b) [easy] Write the exact or approximate distribution of the standardized estimator under the null hypothesis which we denote  $(\hat{\theta}_1 - \hat{\theta}_2)/SE \mid H_0$ . You should first write the SE as a mathematical expression. Then compute SE to three decimal places. The answer is in lecture 6.

- (c) [easy] Illustrate the distribution from the previous problem. Label the x-axis and provide tick marks on the x-axis. Leave lots of room on the left side of the x-axis away from the bulk of the sampling distribution!
- (d) [harder] Will this test be an *exact test* or instead an *approximate test*? Explain.
- (e) [harder] Compute the retainment region and rejection region (remember  $\Theta = \mathbb{R}$ ) and denote these two regions in your illustration in (c). To compute these regions, I'll provide you with the following fact:  $\Phi(-1.645) = 5\%$  where  $\Phi$  is the CDF of the standard normal rv. (These are the kind of facts that will be provided to you on exams).
- (f) [easy] Why is this test named the *left-sided two-sample  $z$  test of unequal variances*?

- (g) [easy] Run the test and write your conclusion using an English sentence.
- (h) [easy] What type of error could you have made?
- (i) [harder] Find the p-value of our estimate as a function of  $\Phi$ . Illustrate the p-value in the illustration in (c).
- (j) [easy] Without computing the p-value explicitly, would it be above or below  $\alpha = 5\%$ ? Is the estimate *statistically significant*?
- (k) [difficult] [MA] In the general case of  $\theta_\Delta := \theta_1 - \theta_2$ ,  $\sigma_1^2$ ,  $\sigma_2^2$ ,  $n_1$ ,  $n_2$  and  $\alpha$ , find the power function  $POW(\theta_\Delta, \sigma_1^2, \sigma_2^2, n_1, n_2, \alpha)$ .

---

We will now assume that the variances are equal and known to be  $\sigma_1^2 = \sigma_2^2 = \sigma^2 = 3.76^2$ .

- (l) [easy] Write the exact or approximate distribution of the standardized estimator under the null hypothesis which we denote  $(\hat{\theta}_1 - \hat{\theta}_2)/SE \mid H_0$ . You should first write the SE as a mathematical expression. Then compute SE to three decimal places. The answer is in lecture 6.

- (m) [easy] Illustrate the distribution from the previous problem. Label the x-axis and provide tick marks on the x-axis. Leave lots of room on the left side of the x-axis away from the bulk of the sampling distribution!

- (n) [harder] Will this test be an *exact test* or instead an *approximate test*? Explain.

- (o) [harder] Compute the retainment region and rejection region (remember  $\Theta = \mathbb{R}$ ) and denote these two regions in your illustration in (m). To compute these regions, I'll provide you with the following fact:  $\Phi(-1.645) = 5\%$  where  $\Phi$  is the CDF of the standard normal rv. (These are the kind of facts that will be provided to you on exams).
- (p) [easy] Why is this test named the *left-sided two-sample z test of equal variances*?
- (q) [easy] Run the test and write your conclusion using an English sentence.
- (r) [easy] What type of error could you have made?
- (s) [harder] Find the p-value of our estimate as a function of  $\Phi$ . Illustrate the p-value in the illustration in (m).
- (t) [easy] Without computing the p-value explicitly, would it be above or below  $\alpha = 5\%$ ? Is the estimate *statistically significant*?

---

We will now assume that the variances are equal i.e.  $\sigma_1^2 = \sigma_2^2 = \sigma^2$  but its value is *unknown*.

- (u) [easy] Write the exact or approximate distribution of the standardized estimator under the null hypothesis which we denote  $(\hat{\theta}_1 - \hat{\theta}_2)/SE \mid H_0$ . You should first write the SE as a mathematical expression. Then compute SE to three decimal places. The answer is in lecture 6.

- (v) [easy] Illustrate the distribution from the previous problem. Label the x-axis and provide tick marks on the x-axis. Leave lots of room on the left side of the x-axis away from the bulk of the sampling distribution!

- (w) [harder] Will this test be an *exact test* or instead an *approximate test*? Explain.

(x) [harder] Compute the retainment region and rejection region (remember  $\Theta = \mathbb{R}$ ) and denote these two regions in your illustration in (v). To compute these regions, I'll provide you with the following fact:  $\mathbb{P}(T_{14} \leq -1.76) = 5\%$  where  $T_{14}$  denotes a standard Student's t rv with 14 degrees of freedom. (These are the kind of facts that will be provided to you on exams).

(y) [easy] Why is this test named the *left-sided two-sample t test of equal variances*?

(z) [easy] Run the test and write your conclusion using an English sentence.

(aa) [easy] What type of error could you have made?

(bb) [harder] Find the p-value of our estimate by writing a statement like  $\mathbb{P}(T_{df} < t)$  or  $\mathbb{P}(T_{df} > t)$ . You need to solve for  $df$ ,  $t$ . Illustrate the p-val in the illustration in (v).

(cc) [easy] Without computing the p-value explicitly, would it be above or below  $\alpha = 5\%$ ? Is the estimate *statistically significant*?



---

We will now assume that the variances are unequal i.e.  $\sigma_1^2 \neq \sigma_2^2$  and both values are *unknown*. This is known as the Behrens-Fisher problem.

- (dd) [easy] Write the exact or approximate distribution of the standardized estimator under the null hypothesis which we denote  $(\hat{\theta}_1 - \hat{\theta}_2)/SE \mid H_0$ . You should first write the SE as a mathematical expression. Then compute SE to three decimal places. The answer is in lecture 7.

- (ee) [easy] Illustrate the distribution from the previous problem. Label the x-axis and provide tick marks on the x-axis. Leave lots of room on the left side of the x-axis away from the bulk of the sampling distribution!

- (ff) [harder] Will this test be an *exact test* or instead an *approximate test*? Explain. (Hint: the exact distribution was solved in 2018 and can be found in this paper).

(gg) [harder] Compute the retainment region and rejection region (remember  $\Theta = \mathbb{R}$ ) and denote these two regions in your illustration in (ee). To compute these regions, I'll provide you with the following fact:  $\mathbb{P}(T_{9.94} \leq -1.81) = 5\%$  where  $T_{9.94}$  denotes a standard Student's t rv with 9.94 degrees of freedom. (These are the kind of facts that will be provided to you on exams).

(hh) [easy] Why is this test named the *left-sided two-sample t test of unequal variances*?

(ii) [easy] Run the test and write your conclusion using an English sentence.

(jj) [easy] What type of error could you have made?

(kk) [harder] Find the p-value of our estimate by writing a statement like  $\mathbb{P}(T_{df} < t)$  or  $\mathbb{P}(T_{df} > t)$ . You need to solve for  $df$ ,  $t$ . Illustrate the p-val in the illustration in (ee).

(ll) [easy] Without computing the p-value explicitly, would it be above or below  $\alpha = 5\%$ ? Is the estimate *statistically significant*?