

# MATH 341/641 Fall 2023 Homework #2

Professor Adam Kapelner

Due by email 11:59PM September 20, 2023

(this document last updated Monday 18<sup>th</sup> September, 2023 at 10:45pm)

## 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 340 concepts: random variables, PMF’s, PDF’s, CDF’s, binomial, and review MATH 241 concepts: the normal distribution.

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 621 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 5 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

Here we will do a binomial exact test using the survey data from our class. We want to demonstrate that the iphone users in our class is *greater* than the national average (which is 52.4%). Recall that our data was as follows: for  $n = 21$ , the  $\hat{\theta} = 0.476$  where the estimator we chose was the sample proportion.

- (a) [easy] Write down  $H_a$  then  $H_0$ .
- (b) [easy] Declare your  $\alpha$  level desired for this test. You do not need to justify it. It is what you are comfortable with.
- (c) [harder] Because we want to show something is greater than a point value, it is called a right-tailed test. In any test, we need to find the distribution (or approximate the distribution of) the estimator under the null hypothesis. Because we will reject on the right, why is the most conservative value of  $\theta$  to choose when deriving the null sampling distribution to be largest value in the null hypothesis region (in this case  $\theta = \theta_0 = 0.524$ )?
- (d) [easy] Regardless of if you understood the previous question or not, what is the exact null sampling distribution (write the PMF)? Marked easy because you can copy from MATH 340 class notes.
- (e) [easy] Draw the PMF of the null sampling distribution. Label all axes carefully and provide sufficient tick marks. Marked easy because you can copy from class. Probabilities are below rounded to the nearest 3 digits. Leave space below its horizontal axis.

0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	0	0	0.001	0.003	0.009	0.025	0.054	0.097	0.142	0.172	0.172	0.142	0.096
14	15	16	17	18	19	20	21						
0.053	0.023	0.008	0.002	0	0	0	0						

- (f) [easy] Indicate the RET and the rejection region in the above illustration. Use your level  $\alpha$ . Everyone's answer may be different!
- (g) [easy] What is  $\mathbb{P}\left(\hat{\theta} \notin RET\right)$  in this test with your level  $\alpha$ ?
- (h) [harder] Were you able to create a rejection region at your exact level of  $\alpha$ ? Yes / no and why?
- (i) [easy] What are all the possible sizes of this test (rounded to the number of digits in the PMF table).
- (j) [easy] Run the test. Write your conclusion in English.

(k) [easy] What is  $\mathbb{P}$ (Type I error) in this test with your level  $\alpha$ ?

(l) [easy] What is the definition of the pval for this 2-sided test?

(m) [easy] What is the p-value of the estimate in this test?

(n) [easy] Assuming  $\theta = 0.7$ , what is the true sampling distribution? Write its PMF below.

(o) [harder] Calculate the  $\mathbb{P}$ (Type II error) in this test.

(p) [easy] Easy only if you have the answer to the previous question: calculate the power in this test.

## Problem 2

Here we will review theory testing from a conceptual point of view. For each question, state whether the theory under consideration should become a null hypothesis or alternative hypothesis. If null, also write the alternative; if alternative also write the null.

(a) [easy] A new grand unified theory of physics.

(b) [easy] The latest conspiracy theory about the president.

(c) [easy] You are a shareholder in a pharmaceutical company. Your new drug cures cancer.

### Problem 3

We will revisit the concept of a degenerate point estimator as in the first problem but this time let the DGP be  $X_1, \dots, X_n \stackrel{iid}{\sim} \mathcal{N}(\theta, \sigma^2)$  where  $\sigma^2$  is considered known. We are focused on point estimation for  $\theta$  and the range of possible values is all numbers i.e.  $\Theta = \mathbb{R}$ . Of course we use  $\hat{\theta} = \bar{X}$ . But we also consider the “bad” point estimator  $\hat{\theta}_{\text{bad}} = 1$ .

(a) [harder] Graph the bias of  $\hat{\theta}_{\text{bad}}$  over all  $\theta$ . Label your axes.

(b) [harder] Graph the risk of  $\hat{\theta}$  and  $\hat{\theta}_{\text{bad}}$  under squared error loss. Label your axes and provide appropriate tick marks on the axes.

(c) [difficult] Compare  $\hat{\theta}_{\text{bad}}$  to  $\hat{\theta}$  using the sup risk under squared error loss. How much better is  $\bar{X}$ ?

#### Problem 4

We will now do a *one-sided one-sample exact Z test* using the male student height data:  $n = 13$  and  $\bar{x} = 68.85''$ . We want to test at  $\alpha = 1\%$  if the population that this sample was drawn from has a *greater* mean than the American female height mean of  $65''$ . According to this article, female American height is  $\stackrel{iid}{\sim} \mathcal{N}(65, 3.5^2)$ .

- (a) [easy] Write the alternative and null hypotheses.
- (b) [easy] As we discussed in class, which value of  $\theta_0$  do we use to generate the null sampling distribution?
- (c) [difficult] Why do we use that value of  $\theta_0$  when there are potentially infinite values to choose from?
- (d) [easy] Write the null sampling distribution on both the original scale (inches) and the standardized scale.
- (e) [easy] Write the RET region as a set on both the original scale (inches) and the standardized scale.
- (f) [easy] What is the  $\mathbb{P}(\text{Type I error})$  in this test?
- (g) [easy] Is it possible to calculate the  $\mathbb{P}(\text{Type II error})$  of this test given the information you have? Yes / no
- (h) [easy] Is it possible to calculate the power of this test given the information you have? Yes / no

- (i) [easy] Calculate the test statistic on the standardized scale.
- (j) [easy] Run the test using both RET regions (the original scale and standardized scale). Show both answers are the same.
- (k) [easy] Calculate the p-val for this test. Use a z-calculator on your graphing calculator or find one on the Internet.
- (l) [easy] Was this an exact test? Yes / no
- (m) [easy] Write a conclusion of this test in English.
- (n) [easy] Was this conclusion expected? Did you expect such a low/high p-val? Discuss.
- (o) [harder] Find the power function for this test.
- (p) [easy] If  $n \rightarrow \infty$ , what does power converge to?

### Problem 5

Assume male American height is  $\overset{iid}{\sim} \mathcal{N}(70, 4^2)$  and female American height is  $\overset{iid}{\sim} \mathcal{N}(65, 3.5^2)$ . Using the male and female height data from class, we will now prove that the QC STEM 300-level population mean male height is greater than the QC STEM 300-level population mean female height.

- (a) [harder] Write the hypotheses, declare an  $\alpha$  of your choosing, find the sampling distribution under  $H_0$ , find the RET region, run the test, calculate the p-val and provide a conclusion sentence.

- (b) [easy] Was this an exact test? Yes / no

### Problem 6

These questions will be about the Method of Moments (MM) procedure for generating estimators/estimates.

- (a) [easy] Define the  $k$ th moment of a rv.
- (b) [easy] For sample size  $n$ , define the of the  $k$ th sample moment *estimator* of a rv.
- (c) [easy] For sample size  $n$ , define the of the  $k$ th sample moment *estimate* of a rv.



- (d) [easy] Give an example of a DGP with two parameters.
- (e) [easy] For a DGP with  $K = 3$  parameters, write the system of equations that relates each moment to parameters. There should be 3 equations with 3 unknowns. Then write the system of equations that relates each parameter to moments. There should be another 3 equations with 3 unknowns.
- (f) [easy] For any iid DGP with finite mean, find an MM estimator for the mean.
- (g) [easy] For any iid DGP with finite variance, find an MM estimator for the variance.
- (h) [easy] Consider the  $X_1, \dots, X_n \stackrel{iid}{\sim} \text{Binomial}(\theta_1, \theta_2)$  DGP and derive the MM estimators  $\hat{\theta}_1^{\text{MM}}$  and  $\hat{\theta}_2^{\text{MM}}$  for  $\theta_1$  and  $\theta_2$  and express them in terms of  $\bar{X}$  and  $\hat{\sigma}^2$ .

- (i) [easy] Provide an example dataset (different from the one in class) where the MM estimates  $\hat{\theta}_1^{\text{MM}}$  and  $\hat{\theta}_2^{\text{MM}}$  in (h) are illegal and explain why they're illegal.
- (j) [harder] Imagine you are a NYPD officer at precinct 100 in Queens. You want to estimates of the number of crimes in your precinct and people's propensity to phone in crimes. The number of daily phone reports for two weeks are: 13, 21, 25, 21, 15, 19, 15, 15, 17, 23, 16, 15, 19, 15. Estimate the true mean number of daily total crimes in the precinct and probability of the crime being phoned in.
- (k) [difficult] What exactly what you need to know if you wanted to test if the true mean number of crimes daily exceeds 20? This is conceptual and should be answered with a sentence or two.
- (l) [harder] Derive the MM estimator  $\hat{\theta}^{\text{MM}}$  for  $\theta$  in the  $X_1, \dots, X_n \stackrel{iid}{\sim} \text{U}(\theta, 17)$  DGP.
- (m) [easy] Provide an example dataset where the MM estimate  $\hat{\theta}^{\text{MM}}$  in the previous question is illegal and explain why it is illegal.

- (n) [difficult] [MA] Derive the MM estimators  $\hat{\theta}_1^{\text{MM}}$  and  $\hat{\theta}_2^{\text{MM}}$  for the  $\overset{iid}{\sim} \text{U}(\theta_1, \theta_2)$  DGP. Then estimate  $\theta_1$  and  $\theta_2$  given the dataset 9.8, 3.1, 1.2, -0.1, 12.1, 15.9, -9, 3.4, 14.9, -3.6, 16.5, -9.6, 11.2, 11.6, -3.9, -9.3, -1 using these new MM estimators. Remember that  $\theta_1$  is defined to be  $< \theta_2$ !