

# Gov 2002: Problem Set 8

## Problem Set Instructions

This problem set is due on **November 15, 11:59 pm** Eastern time. Please upload a PDF of your solutions to Gradescope. We will accept hand-written solutions but we strongly advise you to typeset your answers in Rmarkdown. Please list the names of other students you worked with on this problem set.

## Question 1 (25 points)

In the last week's problem set (Question 2), we derived the asymptotic distribution of the  $p$ -th sample quantile of a random variable  $X$ :

$$\sqrt{n}(Q_{X,p} - q_{X,p}) \xrightarrow{d} \mathcal{N}\left(0, \frac{p(1-p)}{\{f_X(q_{X,p})\}^2}\right)$$

Assume that  $X \sim \mathcal{N}(0, 1)$  and answer the following questions.

- You want to perform a two-sided hypothesis test with the null of  $H_0 : Q_{X,p} = c$ . Propose a test statistic to perform this hypothesis test.
- Regarding your answer in (a), briefly explain why it is a good test statistic.
- Suppose that you are testing whether the median is 0.1, using a two-sided hypothesis test. You have  $n = 1000$  samples and you obtained the sample median of 0.1785. Would you reject or accept the null at  $\alpha = 0.05$ ? Explain.

## Question 2: Power Analysis (25 points)

This question will give you some realistic practice conducting *power analysis*, a very useful procedure when planning data collection for research.

Recall the subprime data from previous problem sets. You have been given some money from IQSS to conduct a field experiment to replicate some of the findings in this observational dataset. Your field experiment will take the form of an audit study, where you test whether or not lenders are discriminating on the basis of gender. Such audit studies are a common way of studying discrimination in housing, wage, and many other markets. You will send out loan applications to a variety of lenders, and will randomly assign the gender of the applicant. This random assignment is important because it would allow you to make a causal claim about the relationship between gender and the amount of money applicants will be loaned.

IQSS's money pot is deep, but it has its limits, so you need to know what size experiment you can afford. You decide to conduct a power calculation, so that you can know the likelihood of “finding effects” (i.e. rejecting the null hypothesis) in your proposed study. For this problem, you will use the same `subprime.csv` dataset we have used in a previous problem set (uploaded on Ed).

- (a) You already know your estimator,  $\bar{X}_w - \bar{X}_m$ , your outcome measure, `loan.amount`, and the fact that you want to conduct a two-sided test. What are the factors do we need to consider that will affect the power of your experiment? (Hint: There are four of them.) In what way does each of them affect power, that is, what happens to power as each of these factors increases or decreases in size?
- (b) IQSS wants you to come up with an accurate estimate for the effect size you expect to see. Calculate a reasonable value for the effect size using difference in average loan amount between men and women in the population (i.e. the subprime data as a whole). In what ways might this value be an accurate estimate for the effect you expect to find in your experiment? In what ways might it be inaccurate? If this estimate shows a stronger effect than the real effect size, what will that imply about our power calculation?
- (c) Suppose that you have received a grant of \$3,000 for your experiment, and you estimate that your experiment will cost approximately \$5 per participant. You plan on assigning half of your applicants to be male and the other half to be female. Assuming an effect size equal to what you calculated in (b), and using a standard  $\alpha = 0.05$ , does this budget allow you to run an experiment where your power will be greater than or equal to 0.8? (Remember that you are conducting a two-sided test.)

**Hint:** For this part, assume that  $\sigma_w^2 = \sigma_m^2$ , and assume that the variances in the experiment will be the same as the population variances in the subprime dataset.

### Question 3 (25 points)

A welfare policy is applied to a group of  $n$  people to counter job loss caused by trade shocks, resulting in  $X_1, \dots, X_n$  which are i.i.d.  $\mathcal{N}(\theta, 1)$ , where  $\theta$  is the theoretical mean effect of the remedy (e.g., the logged wage difference between new and old jobs), defined so that  $\theta > 0$  if the remedy is helpful on average,  $\theta < 0$  if it is harmful on average, and  $\theta = 0$  if it does nothing. Let our Type I error rate be  $\alpha = 0.05$ . Consider testing  $H_0 : \theta = 0$  versus  $H_1 : \theta \neq 0$ . Let the rejection region be

$$\mathcal{R} = \{x : |\hat{\theta}_{PI}| > c\},$$

where  $\hat{\theta}$  is the plug-in estimator of  $\theta$ .

- (a) What is  $\hat{\theta}$ , the plug-in estimator of  $\theta$ ? Your answer should be a function of  $X_1, X_2, \dots, X_n$ . Also, under the null, what is the distribution of the estimator? (**Hint:** Under the null means that we assume that  $\theta = 0$ .)
- (b) Find  $c$  so that the test has Type I error rate (i.e., size) equal to  $\alpha$ . (Note that  $c$  will depend on the sample size,  $n$ .) (**Hint:** Type I error rate  $\alpha$  should be equal to

$$P(|\hat{\theta}| > c.)$$

- (c) Find the power of the test,  $\beta(\theta)$ , for  $\theta > 0$ . (**Hint:** This question follows almost the same procedure as parts (a) and (b), except for (1) we don't assume that the null is true and (2) the power function is defined as a conditional probability given  $\theta$ .)
- (d) Prove that the power  $\beta(\theta) \rightarrow 1$  as  $n \rightarrow \infty$  (**Hint:** Use the result from part (c).)
- (e) Suppose now that  $n = 10^4$  and we observe  $\bar{x} = 0.02$ . What is the p-value? Does the test say to reject or retain  $H_0$ ? (**Hint:** The critical value is the one obtained in part (b), i.e., under the null.)

### Question 4 (25 points)

In recent years it has become common in statistics to want to perform many simultaneous hypothesis tests. Let  $p_1 \dots p_m$  be independent p-values, corresponding to  $m$  hypothesis tests. Each of the  $m$  hypothesis tests has a simple null. Suppose that  $m_0$  of the  $m$  null hypotheses are true. We decide in advance to conduct these tests at level  $\alpha$  (i.e., we reject the null for tests where the p-value is less than  $\alpha$ ). The *familywise error rate* is the probability of making at least one Type I error.

- (a) Find the familywise error rate (FWER). What happens to the familywise error rate as  $m_0$  gets large? (**Hint:** Let  $V$  be the *number* of Type I error rates. Then FWER is  $P(V > 0)$ . You can simplify it further by using  $\alpha$  and  $m_0$ .)
- (b) One of the common procedure to deal with multiple hypothesis testings is called the *Bonferroni procedure*. Intuitively, this procedure allows us to deal with the increased likelihood of type I errors. This procedure is described as follows: instead of rejecting the null hypotheses with  $p_i < \alpha$ , we reject the null hypotheses with  $p_i < \frac{\alpha}{m}$  (i.e., correcting the cutoff by the number of hypotheses). Show that under this procedure, the familywise error rate is at most  $\alpha$ . (**Hint:** You might find Markov's Inequality helpful:  $P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$ . You want to bound the FWER using Markov's Inequality by setting the value of  $a$  appropriately.)
- (c) In (b), why not instead reject the null hypotheses with  $p_i < \frac{\alpha}{m_0}$ , considering that  $m_0 \leq m$  (and often in practice  $m_0$  is much smaller than  $m$ ), which would seem to result in rejecting more false nulls while still keeping the familywise error rate at most  $\alpha$ ?
- (d) Another procedure is to reject all null hypotheses with  $p_i < 1 - (1 - \alpha)^{1/m}$  (This is known as the Sidak Procedure). Show that under this procedure, the familywise error rate is again at most  $\alpha$ .