Problem Set 5 **Solutions**

EC 303: Empirical Economic Analysis

Due December 2, 2019 at 3:30pm.

1 Theoretical Problems

Problem 1.1: Performing hypothesis tests. This problem asks you to perform simple hypothesis tests for sample means and populations. For each test, make sure to (i) state the null and alternative hypotheses and the chosen level of significance, (ii) define the test statistic, (iii) calculate the value of the realized statistic with its corresponding p-value, and (iv) decide whether or not to reject the null hypothesis, with a sentence explaining what that decision means in the context of the problem.

a. Complete Exercise 9.25 from the textbook. (*Note:* For each of these problems, you may do a one- or two-sided test, but think about what would be the best in each context.)

b. Complete Exercise 9.40 from the textbook.

Problem 1.2: Getting comfortable with p-values. This problem is adapted from problem 9.47 in the textbook. For a fixed hypothesis test of $\mathcal{H}_0 : \mu = 5$ against $\mathcal{H}_1 : \mu > 5$, five test statistics T are listed below. For each, state the corresponding sampling distribution and compute the associated p-value.

a. $T = 1.42$, σ is known, $n = 100$

b. $T = 0.9$, σ is known, $n = 1{,}000{,}000$

c. $T = -1.96$, σ is unknown, $n = 26$

d. $T = 2.48$, σ is unknown, $n = 3$

e. $T = -0.11$, σ is unknown, $n = 800$

f. For which of these tests would we reject the null hypothesis when $\alpha = 0.05$? Does this always correspond to a large T? Why or why not?

Problem 1.3: Testing variance. This problem introduces you to testing variances, rather than sample means/populations. It is adapted from problems 9.88 and 9.89 in the textbook.

a. In Chapter 8, we formed the confidence interval for the variance σ^2 of a population. We relied on the fact that this statistic has a χ^2 distribution:

$$(n-1)\frac{S^2}{\sigma^2} \sim \chi^2(n-1). \tag{1}$$

Use this fact to write a test statistic for the test:

$$\begin{aligned} \mathcal{H}_0 &= \sigma^2 = \sigma_0^2 \\ \mathcal{H}_1 &= \sigma^2 > \sigma_0^2 \\ \alpha &= \alpha. \end{aligned}$$

That is, how would you convert Equation (1) into a test statistic?

b. What would change about this statistic if we want to perform a test on σ instead of σ^2?

c. The distribution of this test statistic is $\chi^2(n-1)$ and the associated critical value is $\chi^2_{\alpha,n-1}$ (recall that this means the value of the distribution with area α to its right (see p. 409 if this is rusty).

   A monopsonistic labor market is one where firms have wage-setting power, due to reduced competition for workers. One potential flag for monopsonies is a high degree of variation in wages within a market (see Webber, 2015). Suppose that a labor market is considered monopsonistic if the standard deviation of wages in that market is larger than \$10 an hour. If you interview 10 firms in an industry (e.g., the retail sector) and find that their wages have a standard deviation of 12, can you conclude that there is a significant degree of monopsony power in that market? Test the appropriate hypotheses using $\alpha = 0.05$.

d. Now suppose you interview $n = 21$ regional clinics that hire nurses in an area. For this data, and find a test statistic of 31.58.Use a computer to calculate the p-value for the same test in part (c). What does this tell you about monopsony power in the market for nurses? Does your test give you a sense of how "strongly" monopsonistic this area is?

Problem 1.4: Testing a difference in means. Frequently, we care about whether or not two groups have the same mean in a given outcome (this is the entire basis of estimating the effect of a treatment on a group relative to a control!). This problem will help extend the testing framework to that problem.

a. Consider two groups $\{X_1, ..., X_m\}$ and $\{Y_1, ..., Y_n\}$. We suppose that $X_i \sim_{\text{i.i.d.}} f(\mu_1, \sigma_1)$ and $Y_i \sim_{\text{i.i.d.}} f(\mu_2, \sigma_2)$. Additionally, we suppose that X and Y are independent samples.

   We are trying to estimate the difference in means, $\mu_1 - \mu_2$. What sample estimator should we use? (You can use the method of moments in a pinch, but trust your gut.)

b. Prove that your estimator is unbiased, and that it has a standard deviation of

$$\sigma_T = \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$$

c. Now use your estimator to write a test statistic for the following test:

$$\begin{aligned} \mathcal{H}_0 &: \mu_1 - \mu_2 = \Delta_0 \\ \mathcal{H}_1 &: \mu_1 - \mu_2 > \Delta_0 \\ \alpha &= \alpha_0. \end{aligned}$$

   Recall that the general version of a test statistic is

$$T = \frac{\text{estimator} - \text{assumed value in } \mathcal{H}_0}{\text{s.d. of your estimator}}$$

d. This test statistic follows the same rules as other statistics—if we know σ^2 or have a large enough sample, $T \sim \mathcal{N}(0, 1)$. If instead, we use our estimate s^2, the test statistic uses a t distribution (whose degrees of freedom is a function of n and m).

   Suppose that we are trying to evaluate the effect of political information on voter's preferences. Specifically, we take 1,000 college students and divide them into a treated group of $n = 400$ students and a control group of $m = 600$. To the treated group, we show a series of economic articles discussing the Fundamental Welfare Theorems in economics, and then ask them what they think is the optimal wealth tax in the United States. We ask the control group the same question, but without giving them information on economic theory. We estimate that $\overline{X} = 6\%$ for the treated group, and $\overline{Y} = 4.5\%$ for the control group. The control group has a standard deviation of responses of 5%, but the treated group only has a standard deviation of 4% (these large standard deviations occur because of political leanings, which cause a large spread in answers).

   Do we have enough information to discern that the treatment had an effect (so that $\Delta > 0$?) Test this at the $\alpha = 0.05$ level.

- Note: Keep the reported averages in whole numbers (e.g., 6 and 3), not percentages (so don't write 0.06 and 0.03).

e. Are the results surprising to you? Would you argue that they are economically meaningful? Defend your answers.

Problem 1.5: Paired Data. A closely related problem to the issue raised in 1.4 is that of paired data, in which we have only one set of individuals that we treat over time. That is, instead of comparing two different groups where only one received treatment, we follow a group from a baseline outcome (before they are treated) to a post-treatment outcome.

a. In this setup, what is the relationship between n and m (if we observe every individual twice)? How does this simplify your test statistic?

b. We will simplify this even further by assuming that we have data $\{d_1, ..., d_n\}$ for our n individuals, where d is the difference in their treatment period and their baseline. From these data, we can directly calculate $\overline{d}$ and $\text{sd}(d)$. Adapt the test procedure for problem 1.4c to depend only on these two pieces of information (that is, write out the hypotheses and the test statistic).

c. Suppose that we perform an intervention in which college undergrads offer tutoring in statistics to disadvantaged high school students. We measure these students' performances before and after the tutoring as their scores on the midterm and the final exam:

Student	Midterm score	Final score	Difference
1	80	82	2
2	99	99	0
3	60	50	-10
4	70	72	2
5	88	80	-8
6	20	24	4
7	95	92	-3
8	100	94	-6
9	75	80	5
10	64	60	-4

Use your answers above to perform a **two-sided** test of the hypothesis that the tutoring had no effect on these students at the $\alpha = 0.05$ level (hint: note the small sample size). What do you conclude? Can you think of anything that might be confounding this experiment, or does it seem likely that your test results are correct?

Problem 1.6: Simple Bayesian Updating. We are interested in the fraction of low-income housing residents who would accept a government-sponsored voucher to move to a wealthier neighborhood. Within a given neighborhood, we suppose that the fraction m of movers comes from a uniform distribution $[0, \theta]$, but we don't know the end point of the interval. We will estimate this by Bayesian analysis.

a. What is a good upper bound to start with for θ? (That is, our prior should be defined for what values of θ?) What is θ telling us in this context (and why would we care about it)?

b. Suppose that we start with a sample of 5 neighborhoods, and observe $\{m_i\} = \{0.05, 0.1, 0.15, 0.2, 0.22\}$. Given this information, what values of θ can we rule out? Should our starting value of θ go up or down?

c. What is the general likelihood function for a data set of n neighborhoods, each drawn uniformly from $[0, \theta]$? What is the specific likelihood function for the 5 neighborhoods you've measured?

d. Now, let's define a suitable prior for this problem. If the likelihood function is uniform on $[0, \theta]$, a conjugate prior comes from the Pareto distribution with hyper-parameters x_{min} and k. Recall that the Pareto distribution is characterized by

$$f(\theta; x_{\text{min}}, k) = \begin{cases} \frac{kx_{\text{min}}^k}{\theta^{k+1}} & \theta \geq x_{\text{min}} \\ 0 & \text{otherwise.} \end{cases}$$

whenever $x > x_{\text{min}}$.

Show that this prior is a conjugate prior by proving that the product of the likelihood function and the prior distribution is proportional to another Pareto distribution. This is most easily shown by ignoring the constant kx_{min}^k, since we're showing *proportionality* instead of *equality*; hence, show that the product of your likelihood function and $\frac{1}{\theta^k}$ follows the Pareto pattern $\frac{1}{\theta^{k'}}$.

  - Use the general likelihood function from part (c), not the one specific to your 5 data points.

e. What should the posterior distribution of θ look like, based on your answer to (d)? That is, what are the two new hyper-parameters of the posterior distribution?

  - To get the change in x_{min}, consider your answer to (b)—for what values of θ will the prior and the likelihood *both* be non-zero (as these are the values for which the posterior will be non-zero).

f. The second parameter of the Pareto distribution defines the lower bound for the interval over which θ is likely to occur, and the first parameter of the Pareto controls how *concentrated* probability is around the lower bound (higher k = higher likelihood of smaller θ). Based on this, and your answer to (e), how does data affect the posterior distribution of θ? Overall, what does this tell us about how collecting data among various neighborhoods will lead to an accurate estimate of θ?

Problem 1.7: Connecting HDIs to the Real World. Suppose an election is approaching, and you are interested in knowing whether the general population prefers candidate A or candidate B. There is a just published poll in the newspaper, which states that of 100 randomly sampled people, 58 preferred candidate A and the remainder preferred candidate B.

a. Suppose that before the newspaper poll, your prior belief on the fraction of candidates preferring A was a uniform distribution. What is the 95% HDI on your beliefs after learning of the newspaper poll results? Who do you think will win? Can you say this will with 95% confidence?

  - *Hint:* Once you have specified the posterior, use this website to calculate the critical values.

b. You want to conduct a follow-up poll to narrow down your estimate of the population's preference. In your follow-up poll, you randomly sample 100 other people and find that 57 prefer candidate A and the remainder prefer candidate B. Assuming that peoples' opinions have not changed between polls, what is the 95% HDI on the posterior? Now what is your conclusion about the election—can you declare victory for one candidate (with 95% confidence)?

2 Stata Exercises

Problem 2.1: Hypothesis Testing in Stata. This problem introduces you to hypothesis testing in Stata. For this problem, use the jtrain2 data set (*Note*: Don't use the jtrain data set here.).

a. This data contains information on workers, some of whom attended a job training program. The variables of interest are train, which tells us if the person received training, and re78, which tells us their real monthly earnings in 1978 (measured in 1000's of dollars). First, perform a simple test using the ttest command that the mean real wage in 1978 was equal to $5,000. What do you conclude? Use $\alpha = 0.05$.

b. Repeat this test separately for the groups that got the job training, and those that didn't. Use $\alpha = 0.05$. What do you conclude?

c. Construct 95% confidence intervals for re78 for both groups (trained and untrained). How do these confidence intervals relate to your answers in (b)? What information do they add?

d. One way to visualize hypothesis testing is to visualize confidence intervals. Use the procedure outlined below to collapse your data into different groups of individuals who received job training. Make 95% confidence intervals for re78 for each of the groups, and plot them all on a graph. Include a dashed line at \$5,000 so we can easily see which confidence intervals contain the null hypothesis value. Which groups can we say have wages different from $\$5,000$?

   - Don't forget to drop those without the job training
   - Make one variable that has 3 levels for the 3 buckets of age: < 30, 30–39, and 40+. Make another that has 3 levels of education: less than high school, HS diploma, and some/all college (remember that a HS degree is 12 years of education).
   - Use the command `local group black hispanic age_grp ed_grp` to define a grouping mechanism
   - Use the collapse command to get means and standard errors for each group. I suggest typing `collapse (mean) y = re78 (semean) se_y = re78, by(group)`, but there are multiple ways to do this
   - Drop all groups without standard errors (they have small sample sizes). You should be left with 13 groups.
   - For graphing, check out the rcap command—it will help the CI's look nice in your twoway plot.

Problem 2.2: Hypothesis Testing and Simulation. This problem exposes you to the competing notions of Type I and II errors in hypothesis testing, as well as the concept of a test's **power**. For this problem, you will simulate your own data and treatment effect.

a. Think of a context where you would like to perform a hypothesis test. This could be related to any of the examples that we have discussed in class, but should reflect something that you might be interested in researching one day. State your problem's context, and the null/alternative hypotheses you are seeking to test. Use $\alpha = 0.05$.

b. Simulate data according to this context. Bake a *true rejection of* $\mathcal{H}_0$ into your simulation—that is, if my null hypothesis is that $\mu = 0$, I may want to draw my data from a normal distribution with a true mean of $\mu = 2$ instead.

   - Use a sample size of 10,000 observations.
   - You may assume your data are normally distributed with a variance of $\sigma^2 = 4$, no matter your true context (although other simulations are encouraged, if you like).
   - Make the treatment effect about 10% higher than your null hypothesis value μ_0.

c. What is the p-value associated with a simple test of your hypothesis using the full data set? What would you conclude from the full data?

d. Now simulate 1,000 different tests. For each test, take a sample (with replacement; use the bsample command) of 100 observations from your population, and perform the t-test on that subset. For each test, record the p-value. Once you have finished, make a histogram of your p-values. What fraction of the time do you reject the null hypothesis? How often do you *fail* to reject the null hypothesis even though you should? What type of error is this?

e. Now suppose that your null hypothesis is true for your context. Re-simulate data from a distribution with mean μ_0 and repeat the procedure in (d). Re-report the histogram. What fraction of the time do you falsely reject the null hypothesis (and how does this compare to your set level of α)? What kind of error is this? What would the consequences of this error be in your context?

f. The **power** of the test is the probability of correctly rejecting the null hypothesis, and is written as $1 - \beta$. In the setting above, what was the power of the test? What do you think happens to a test's power as the difference between $\mathcal{H}_a$ and $\mathcal{H}_0$ decreases? What does this mean in your context?

Problem 2.3: Bayesian Analysis and MCMC Methods. This problem reviews the Bayesian analysis code we've discussed in class.

One conjugate pair that we can use easily in Stata (but less easily in computation) is the normal-normal pair. Suppose that we are environmental economists attempting to estimate the true impact of a new farming technique on emissions reductions. We know that the technology causes a shock to emissions that follows a normal distribution $\mathcal{N}(\mu, 10)$, so that the variance of this technology is known[1]. Based on lab results, we have a prior belief that μ itself is normally distributed with mean $\nu = -8$ and $\tau = 5$. All units here are *percentage points of carbon emissions.*

a. The first farm to implement the technology in our new country experiences a 10 percentage point decline in their carbon emissions. Graph the prior and the posterior distributions given this data. How does this observation change our beliefs about μ?

   – To obtain the posterior distribution, use the fact that if the original hyper-parameters of the prior are (ν, τ^2), the hyper-parameters of the posterior will be

$$\left(\frac{\sigma^2\nu + \tau^2 n\overline{x}}{n\tau^2 + \sigma^2}, \frac{\sigma^2\tau^2}{n\tau^2 + \sigma^2}\right)$$

b. Repeat this exercise for a fluke farm, that somehow experienced a 10 percentage point *increase* in their emissions after using the technology. How does the fluke affect the posterior?

   – Don't use the -10 from part (a) in your calculation of the posterior.

c. We deliver this technology to 20 farms sequentially, and observe the following reductions in each of their emissions (in order):

$$\begin{aligned} X = \{&-10, -8, -5, -3, 3, -4, -3, -2, -2, 0 \\ &-1, -1, 1, -2, 0, 2, 2, 3, 5, 6\} \end{aligned}$$

   Suppose that we receive government funding to implement this technology only as long as our posterior distribution has a mean at or below 2 percentage points (a sizeable enough reduction to justify the cost of the technology). At what point in this chain of data would our funding have been cut?

   – Notice that you only need the posterior mean after each iteration, not the full distribution. This should simplify things.

d. Do the data look to be independent? If we suppose that there is a month or more between each observation, can you think of something that might be happening to our data that caused us to lose funding?

e. The research team decides that measuring each individual farm's reduction in percentage points doesn't make sense given the wide variation in farm output. Instead, they recode the data as 1 if emissions are reduced, and 0 otherwise. Use the MCMC algorithm discussed in class (for this Binomial data) and a flat prior to estimate the posterior distribution of θ. How do we interpret θ? Should we continue to push the technology based on a flat prior if we want $\theta \geq 0.5$? How about with the prior distribution $\text{Beta}(0.1 * 20, 0.9 * 20)$ (where we're less sure there's an effect)?

   – Be sure to check the diagnostics to make sure your algorithm converged well.

3 References

- Webber, Douglas A., 2015. Firm market power and the earnings distribution. *Labor Economics.*

[1] This may not be a completely unreasonable assumption if, for example, we are adapting a technology for use in one developing country that has already been used more widely in other areas of the globe, and we hence have prior information on its variation.