Assignment 3

*Note:* Unless otherwise specified, data sets comes from Wooldridge’s econometrics textbook. To use them, use the R package “wooldridge”.

*Note 2:* These problems walk you through some of the math behind hypothesis testing. While important, mathematical derivations will not be on any exam in this course.

**Problem 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. All of this should be explained clearly in the context of the problem.



**Problem 2: Getting comfortable with** *p***-values.** This problem is adapted from problem 9.47 in the textbook. For a fixed hypothesis test of 0 : *µ* = 5 against 1 : *µ >* 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

1. *T* = −1*.*96, *σ* is unknown, *n* = 26
2. *T* = 2*.*48, *σ* is unknown, *n* = 3
3. *T* = −0*.*11, *σ* is unknown, *n* = 800
4. For which of these tests would we reject the null hypothesis when *α* = 0*.*05? Does this always correspond to a large *T* ? Why or why not?

*Proof.* Solution

1. When *σ* is known, the test statistic comes from a standard normal (0*,* 1) distribution. Hence, the p-value here is 0.07784. This distribution does *not* depend on the sample size.

N

1. Again, the test statistic’s distribution is (0*,* 1), and we ignore the (large) sample size. In this case,

N

*p* = 0*.*1846.

1. When *σ* is unknown, we draw from a *t*-distribution with *n* 1 degrees of freedom. In this case, that distribution is *t*(25), from which we get a *p*-value of 0*.*0306.

−

1. Now, the distribution is *t*(2), and the corresponding *p*-value is 0*.*066.
2. When the sample size is large enough, the *t* distribution approximates the standard normal, so we could either sample from *t*(799) here or N (0*,* 1). Either way, the *p*-value should be about 0*.*456.
3. We would reject the null hypothesis only for the third hypothesis test. Notice that part (d) has an even larger test statistic, but because its sample size is so small, that much variation is unreliable. A large *T* does not always mean a small enough *p* to reject a null hypothesis.

**Problem 3: Testing variance.** This problem introduces you to testing variances, rather than sample means/populations.

1. To construct the confidence interval for the variance *σ*2 of a population, we rely on the fact that this statistic has a *χ*2 distribution:

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

H0 = *σ*

2

H1 = *σ*

2

= *σ*2

*> σ*2

0

0

*α* = *α.*

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

1. What would change about this statistic if we want to perform a test on *σ* instead of *σ*2?
2. 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 *α* = 0*.*05. (The distribution of this test statistic is *χ*2(*n* − 1) and the associated critical value is ).

*α,n−*1

1. 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?

*Proof.* a. Equation 1 is a test statistic if we replace *σ*2 with the value specified in the null hypothesis, *σ*2.

0

Then, *S*2 comes from our sample, and we will have a statistic whose distribution is *χ*2(*n* − 1).

1. Nothing would change—since *σ* and *σ*2 have the same information (the s.d. cannot be negative, so there is a one-to-one mapping between the two), equation (1) can be used on tests for either variable.
2. The hypotheses of this test are:

H0 : *σ*0 = 10

H1 : *σ*0 *>* 10

*α* = 0*.*05*.*

For this data, the test statistic is calculated as:

*T* = 9

144

100

= 12*.*96*.*

By contrast, the critical value *χ*2

0*.*05*,*9

is 16*.*919, so we fail to reject the null hypothesis given our data.

There is not enough evidence of variation in wages for the retail market for us to conclude that firms have wage-setting power.

1. When we sample 21 firms, our test statistic comes from a *χ*2(20) distribution. Hence, we can find a *p*-value of 0*.*048, so we can classify this market for nurses as monopsonistic at *α* = 0*.*05. However, we only *barely* rejected the null hypothesis, so we may not be dealing with a lot of wage-setting power among these clinics.

**Problem 4: Testing a difference in means.** Frequently, we care about whether 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.

1. Consider two groups {*X*1*, ..., Xm*} and {*Y*1*, ..., Yn*}. We suppose that *Xi* ∼i.i.d. *f* (*µ*1*, σ*1) and *Yi* ∼i.i.d.

*f* (*µ*2*, σ*2). Additionally, we suppose that *X* and *Y* are independent samples.

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

−

1. (“extra credit”): Prove that your estimator is unbiased, and that it has a standard deviation of
2. Now use your estimator to write a test statistic for the following test:

H0 : *µ*1 − *µ*2 = ∆0

H1 : *µ*1 − *µ*2 *>* ∆0

*α* = *α*0*.*

Recall that the general version of a test statistic is

*T* = estimator − assumed value in H0

s.d. of your estimator

1. This test statistic follows the same rules as other statistics—if we know *σ*2 or have a large enough sample, . 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*).

∼ N

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 *X* = 6% for the treated group, and *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 ∆ *>* 0?) Test this at the *α* = 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).

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

*Proof.* a. Since we are trying to estimate a linear combination of population means, we can just use the same combination of sample means (try proving this by MOM if you doubt me). Hence, our estimator can just be *X* − *Y* for *µ*1 − *µ*2.

1. To prove that the estimator is unbiased, notice that

E[*X* − *Y* ] = E[*X*] − E[*Y* ]

= *µ*1 − *µ*2*.*

Hence, the bias of the estimator is

*B*(*X* − *Y* ) = E[*X* − *Y* ] − (*µ*1 − *µ*2)

= (*µ*1 − *µ*2) − (*µ*1 − *µ*2) = 0*.*

To find the standard deviation, we find the variance of the estimator using the linear combination formula (since all of our data are independent, there are no covariance terms to worry about):

V[*X* − *Y* ] = V[*X*] + V[*Y* ]

*σ*2 *σ*2

= 1 + 2 *.*

*n m*

The standard deviation is just the square root of this variance, which is what we wanted to prove.

1. Given this information, our desired test statistic is just

(*X* − *Y* ) − ∆0

*Z* = *.*

* *σ* +

2 2

*σ*

1 2

*n m*

1. Our hypotheses for this test are:

H0 : *µ*1 − *µ*2 = 0

H1 : *µ*1 − *µ*2 *>* 0

*α* = 0*.*05

When we calculate the test statistic, we find

5*.*25*.*

When we calculate the associated *p*-value from a standard normal, we find that *p <* 0*.*00001, so we can reject the null hypothesis. This shows us that providing voters information about the usefulness of a welfare tax (in theory) should encourage them to support one more strongly, on average (it would be interesting to look at how these results differ by party affiliation).

1. A difference of 1.5% doesn’t appear very meaningful, especially in the face of large standard deviations in responses. However, given that this tax would be applied to the wealthy, this difference could correspond to a large difference in tax revenues collected, so I would argue that this is an economically meaningful difference.

**Problem 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.

1. In this setup, what is the relationship between *n* and *m* (if we observe every individual twice)? How does this simplify your test statistic?
2. We will simplify this even further by assuming that we have data *d*1*, ..., dn* for our *n* individuals, where *d* is the difference in their treatment period and their baseline. From these data, we can directly calculate *d* and 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).

{ }

1. 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 *α* = 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?

*Proof.* a. Since each individual appears in both *X* and *Y* , *n* and *m* should be equal to each other, so our test statistic’s standard deviation would simplify to

2 *σ*2

J *σ*

1 + 2 =

*m n*

2 2

J *σ*

*σ*

1 + 2

1. *n*

J *σ*2 + *σ*2

=

1

2

*n*

* + *σ*2 + *σ*2

1

=

√*n*

2

1. If we can compute the mean and standard deviation of the difference directly, the test becomes:

H0 : *d* = ∆0

H1 : *d >* ∆0

*α* = *α*0*,*

with the test statistic:

*d* ∆

*Z* = √ *.*

− 0

sd(*d*)*/ n*

1. In this case, we are testing the hypotheses for ∆0 = 0. We can compute the average difference to be

-1.8 and the standard deviation to be 5.18, so the test statistic becomes

1*.*8

−

*t* = 5*.*18*/*√10 = −1*.*099*.*

This test statistic suggests that the tutoring actually **lowered** test scores! When we calculate the *p*-value, we get about 0.30, so we cannot reject the null hypothesis of no effect of tutoring. However, one possible confounder to this experiment is that the final exam may have been more difficult than the midterm exam, so even if the tutoring had helped students prepare more, they still may have performed worse on the harder exam. This is why a control group is still important even though we could directly compare individuals to themselves!