ECO 5350-702 T. Fomby
Intro. Econometrics Spring 2021
2020

EXERCISE 3
KEY

Purpose: To work a homework problem that reviews the basics of a test between means using the parametric approach and does some diagnostic testing on the data provided. **This exercise is due on Thursday, February 11 at 5:00 pm CST.**

Consider the following hypothetical design of experiments with a control group and treatment group. Suppose the data refers to the antibody counts on patients in the control and treatment groups.

Control X1	Treatment X2
8.66	24.56
6.62	18.59
8.79	23.61
1.99	16.15
10.14	17.52
5.74	23.93
11.20	20.54
4.44	14.69
13.90	22.76
11.24	21.37
10.41	23.40
14.66	18.28
11.49	19.18
6.04	19.61
11.17	

For all of the EXCEL calculations see the EXCEL file Exercise 3 Key.xlsx posted on Canvas.

a) Let us calculate the following basic statistics on the above data.

$\bar{X}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{i1}$ = __**9.099333**______.

$\bar{X}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} X_{i2}$ = __**20.29929**______.

$s_1^2 = \frac{\sum_{i=1}^{n_1}(X_{i1}-\bar{X}_1)^2}{(n_1-1)}$ = __**12.5070638**______.

$s_2^2 = \frac{\sum_{i=1}^{n_2}(X_{i2}-\bar{X}_2)^2}{(n_2-1)}$ = ___**9.561746**______.

Skewness: $sk = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^3}{s^3}$ = ___**-0.32374**______. (for control group)

Skewness: $sk = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^3}{s^3}$ = ____**-0.18579**______. (for treatment group)

Kurtosis: $kur = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^4}{s^4}$ = ____**2.071911**______. (for control group)

Kurtosis: $kur = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^4}{s^4}$ = ____**1.661675**______. (for treatment group)

b) Now let us use the Jarque-Bera test for the normality of the control and treatment group scores.

$$JB = \frac{n}{6}\left[sk^2 + \frac{(kur-3)^2}{4}\right]$$

JB for control group = _**0.800368**__. P-value for observed chi-square(2) = _**0.670197**_.

JB for treatment group = _**1.125358**_. P-value for observed chi-square(2) = _**0.569681**_.

The null hypothesis for the JB test is __**population is normally distributed**_____.

The alternative hypothesis for the JB test is _**population is not normally distributed**__.

What is your conclusion concerning the normality of the control and treatment group observations? **Both groups appear to be normally distributed.**

c) Two assumptions are important for the conventional t-test for the difference in means. The first assumption is that the populations of the two groups are approximately normally distributed. The second assumption is that the two populations have equal variances. The following F-statistic is used to test the equal variance assumption, namely, $H_0: \sigma_1^2 = \sigma_2^2$ versus $H_1: \sigma_1^2 \neq \sigma_2^2$. If $s_1^2 > s_2^2$ then calculate the F-statistic, F = $\frac{s_1^2}{s_2^2}$ and observe the right-tail p-value. On the other if $s_1^2 < s_2^2$ then calculate the F-statistic, F = $\frac{s_2^2}{s_1^2}$ and observe the right-tail p-value. In both cases, the numerator and denominator degrees of freedom of the F statistic are equal to the degrees of freedom of the corresponding sample variances ($n_1 - 1$ for s_1^2 and $n_2 - 1$ for s_2^2). In the former case above, the numerator degrees of freedom of the F statistic is $n_1 - 1$ while the denominator degrees of freedom is $n_2 - 1$. The degrees of freedom are reverse in the second case. If the p-value of the F statistic is greater than 0.05 then accept the null hypothesis of equal variances. Otherwise, reject the null hypothesis and accept the alternative hypothesis of unequal variances. Do the variances of the two populations appear to be equal to each other?

The F statistic for the current data is __**1.308031**_____ and its p-value is __**0.317167**___. What is your conclusion? **The population variances of the two groups are equal to each other.**

d) In the case that non-normality and/or unequal variances hold, we have to apply one or both of the resampling techniques of (i) Bootstrapping or (ii) Randomization. On the one hand, if the populations appear to be both normal and have equal variances, we can use the conventional t-statistic. Let us assume that the null hypothesis of interest is $H_0: \mu_1 = \mu_2$ and the alternative hypothesis is $H_1: \mu_1 < \mu_2$, in other words, the treatment is effective in increasing the antibody counts of treated patients. The appropriate t statistic is

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

where $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$ is the pooled estimate of the equal same variance. This t-statistic has $n_1 + n_2 - 2$ degrees of freedom. On the other hand, if the populations appear to be normally distributed but with unequal variances, the Fisher/Behrens t statistic should be used.

$$t^* = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

If $n_1 > 30$ and $n_2 > 30$ then the t^*statistic is approximately distributed as a N(0,1) standard normal random variable. If the sample sizes of the groups are not sufficient for the Fisher/Behrens test, then one can always revert to the Bootstrapping and Randomization tests.

Given the results you have derived in parts a), b), and c) above, tell me which statistic should be used to test the difference in means in the present data. If the JB tests for the two groups indicate that one or both of the populations appear to be non-normally distributed then you can just say that, given additional information, you would use either a Bootstrapping or Randomization test. On the other hand, if both populations appear to be normally distributed then you need to decide whether to use the traditional t statistic. For the small data set here we unfortunately would be uncomfortable in using the Fisher/Behrens t statistic and would have to revert to either the Bootstrapping test or the Randomization test. What is the final conclusion you draw on testing the difference in means of the above data. Be sure and show me which statistic or approach you used, the corresponding p-value of your test and explain the outcome of your test.

Answer: Since both of the two populations appear to be normally distributed and their variances appear to be equal, we can use the traditional t-test for equal means. The t-statistic we get is -9.0506893. Its left-tail p-value is 5.7765E-10 which is very significant. We reject the null hypothesis of equal means and accept the alternative hypothesis that the population mean of the treatment group is greater than the population mean of the control group.