SOCY2200 Statistics **Instructor: Natasha Sarkisian**

Answers to Assignment 6. ANOVA

Use hypothesis testing to answer the questions below. ***For each problem (including the ones using Stata!), make sure to state your null and research hypotheses <u>in words</u> as well as using <u>formal notation</u>. After finishing the test, state your <u>formal conclusion</u> with regard to the null hypothesis as well as your <u>substantive answer</u> to the question.*** Please make sure to include your Stata output with this assignment (paste it into the same file).

1. In a study testing the effectiveness and safety of a new drug, volunteer participants were randomly divided into four groups: Group 1 took a placebo pill, while the other three groups took three different doses of the drug. You are interested in the number of side effects that people experience on this new drug. You would like to find out whether there are differences among the four groups in the number of different side effects that the participants reported.
Placebo: 1, 1, 0, 0, 3
Low dose of the drug: 2, 1, 2, 3, 2
Medium dose of the drug: 1, 4, 3, 2, 0
High dose of the drug: 3, 3, 4, 4, 1

a. What is the overall pattern of mean differences in the sample (i.e., do higher or lower doses of the experimental drug seem to have more side effects?).
b. Using 90% confidence level, test whether, in the population, the number of side effects would depend on the level of exposure to this experimental drug.
c. After completing the test, evaluate the probability of Type I and Type II error.
d. Calculate the effect size for the difference between placebo and high dose and discuss its practical significance.

a.
$\bar{X}_{\text{group1}}$=(1+1+0+0+3)/5=1; $\bar{X}_{\text{group2}}$=(2+1+2+3+2)/5=2; $\bar{X}_{\text{group3}}$=(1+4+3+2+0)/5=2;
$\bar{X}_{\text{group4}}$=(3+3+4+4+1)/5=3
In the sample, placebo group has the least side effects on average (1), and high dose of the drug group has the most side effects on average (3). Low and medium dose groups are in between, with 2 side effects on average in each of those groups.

b.
1. Our null hypothesis is that in the population, there would be no differences in the number of side effects regardless of whether one would take this drug at any dose or placebo; the research hypothesis is that at least one of the drug dose groups (none, low, medium, high), average number of side effects in the population would be different from that for some of the others.
H0: $\mu_1=\mu_2=\mu_3=\mu_4$ H1: $\mu_1\neq\mu_2\neq\mu_3\neq\mu_4$
We use non-directional, one-tailed test (always for ANOVA)
2. We choose α= .10 (1-.90=.10).
3. We will use F statistic.
4. Let's calculate grand mean, and then construct the variance decomposition table.

$\bar{X}_{\text{grand}}$=(1*5+2*5+2*5+3*5)/20=2

Group	X	$\bar{X}_{group}$	$\bar{X}_{grand}$	$X-\bar{X}_{group}$	$(X-\bar{X}_{group})^2$	$X-\bar{X}_{grand}$	$(X-\bar{X}_{grand})^2$	$\bar{X}_{group}-\bar{X}_{grand}$	$(\bar{X}_{group}-\bar{X}_{grand})^2$
1	1	1	2	0	0	-1	1	-1	1
1	1	1	2	0	0	-1	1	-1	1
1	0	1	2	-1	1	-2	4	-1	1
1	0	1	2	-1	1	-2	4	-1	1
1	3	1	2	2	4	1	1	-1	1
2	2	2	2	0	0	0	0	0	0
2	1	2	2	-1	1	-1	1	0	0
2	2	2	2	0	0	0	0	0	0
2	3	2	2	1	1	1	1	0	0
2	2	2	2	0	0	0	0	0	0
3	1	2	2	-1	1	-1	1	0	0
3	4	2	2	2	4	2	4	0	0
3	3	2	2	1	1	1	1	0	0
3	2	2	2	0	0	0	0	0	0
3	0	2	2	-2	4	-2	4	0	0
4	3	3	2	0	0	1	1	1	1
4	3	3	2	0	0	1	1	1	1
4	4	3	2	1	1	2	4	1	1
4	4	3	2	1	1	2	4	1	1
4	1	3	2	-2	4	-1	1	1	1
Σ				0	24	0	34	0	10

$df_{between}$ = k-1 = 4-1 = 3 df_{within}= 20-4 = 16 df_{total}= 20-1 = 19

Source	SS	df	Mean SS	F
Between groups	10	3	3.33	2.22
Within groups	24	16	1.5	
Total	34	19		

Computed test statistic F = 2.22
5. Critical value: Use table B3: df for numerator=3, df for denominator =16. Our critical value will be the value for these df and α=.10. Therefore, critical value = 2.46.
6. Test statistic is smaller than the critical value (2.22<2.46)
7. We fail to reject the null hypothesis.
8. Based on our sample of 20 individuals, we can say that there is not enough evidence to suggest that there are differences in side effects depending on experimental drug dose in the population. We can also report our finding as F(3, 16) = 2.22, p>.10

c.
Probability of Type I error is 0, probability of Type II error is likely large because the sample size is small.

d. To calculate effect size, we need variances for both placebo and high dose groups. We can use the values for X-X group squared for each of these groups from the table above to calculate these. Variance for placebo: 6/(5-1)=1.5 Variance for high dose: 6/(5-1)=1.5. Since variance is the same, we can use sqrt(1.5) for the average SD in the formula for effect size:

ES=(3-1)/sqrt(1.5)=1.63

The effect size is large, so it is practically significant even though it is not statistically significant; we may want to collect more data and evaluate the effect with a larger sample given its practical significance.

2. Using variables *degree* and *tvhours* in GSS2012 data, find out whether, in the U.S. population, the hours spent watching TV vary depending on people's levels of education.

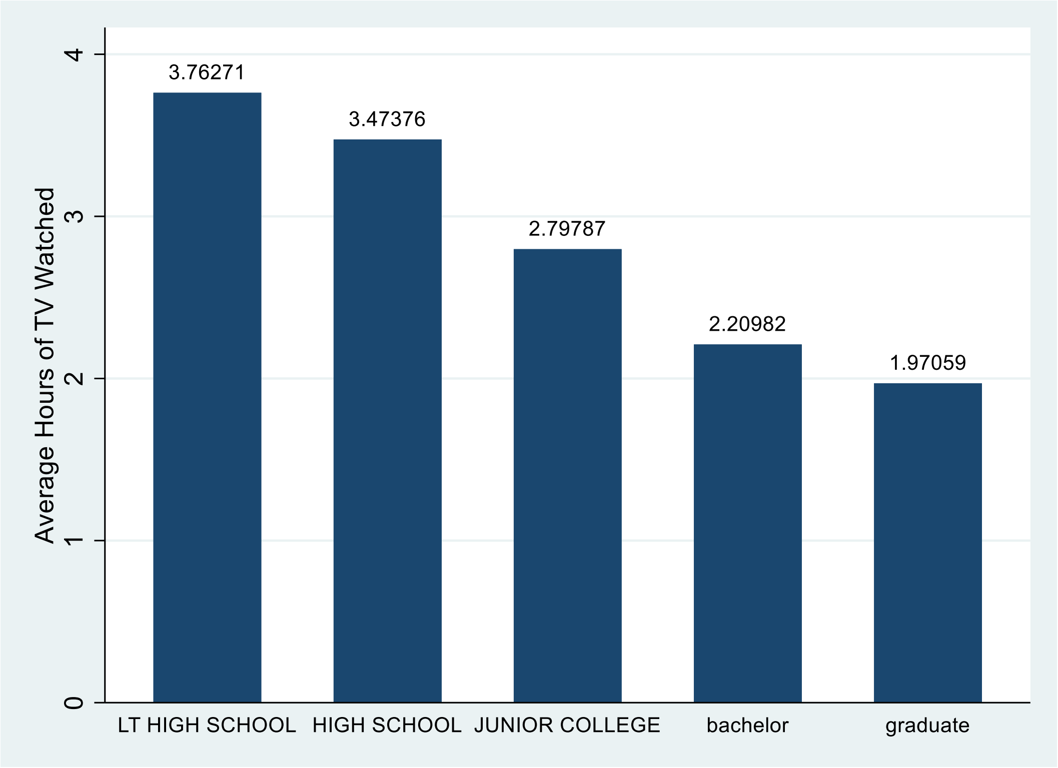
a. Use 99% confidence level for this assessment.
b. After completing the test, evaluate the probability of Type I and Type II error.
c. Construct a bar graph illustrating the means of these groups.
d. Conduct a post-hoc assessment using Stata and determine which pairs of educational groups are significantly different from each other. Again, use 99% confidence level for this assessment as well. List all the pairs that ARE significantly different and explain how they are different (i.e., for each pair, which educational group watches more TV?).
e. For those pairs that ARE significantly different, calculate the effect size and evaluate the practical significance.
f. How many pairwise comparisons in total were you making in (d)? Show how to use a formula to calculate that number.
g. In (d), when you requested post-hoc comparisons using Bonferroni correction, how exactly did Stata adjust the p-values in the output (that is, how was the adjustment calculated)?
h. Take the p-value for Less Than High School vs. Junior College comparison from the output in (d) and calculate what that p-value value would be prior to Bonferroni adjustment.

a. Our null hypothesis is that there are no differences in the number of hours spent watching TV depending on the educational degree one obtained. The research hypothesis is that at least one of the educational degree groups is different from some of the others in terms of the average hours of TV watched per day.
H0: $\mu_1=\mu_2=\mu_3=\mu_4=\mu_5$ H1: $\mu_1\neq\mu_2\neq\mu_3\neq\mu_4\neq\mu_5$
The Stata output gives us the result: F(4, 1293) = 16.945, p<.001.
Since the p-value is smaller than the alpha value (based on the confidence level, 99%, alpha=.01), we can reject the null hypothesis.
Conclusion: at least one of the educational degree groups is different from some of the others in terms of the average hours of TV watched per day.

b. Probability of Type II error is 0, probability of Type I error is smaller than 0.0001.

c. Construct a bar graph illustrating the means of these groups.

```
. graph bar tvhours, over(degree, label(labsize(small))) blabel(bar) ytitle("Average
Hours of TV Watched")
. *note: I added label(labsize(small)) option to make labels for education fit on the
line); we didn't cover that so that's ok if yours didn't fit well on the line
. graph export tvhours.png, replace
```



d. Conduct a post-hoc assessment using Stata and determine which pairs of educational groups are significantly different from each other. Again, use 99% confidence level for this assessment as well. List all the pairs that ARE significantly different and explain how they are different (i.e., for each pair, which educational group watches more TV).

We are testing 10 sets of null and research hypotheses here, e.g. for the first pair:
H0: Those with less than high school and those with a high school degree watch the same number of hours of TV per day.
H1: Those with less than high school and those with a high school degree differ in terms of the number of hours of TV watched per day.

The pairs that are significantly different on .01 level (p<.01) are:
Less than high school vs Bachelor’s degree
Less than high school vs Graduate degree
High school vs Bachelor’s degree
High school vs Graduate degree
How they are different: those with a high school degree or less than high school education watch more hours of TV per day than either those with a bachelor’s or a graduate degree.

e. For those pairs that ARE significantly different, calculate the effect size and evaluate the practical significance.

Let’s calculate effect size for four pairs listed above. We will use these means and standard deviations:

```
  LT HIGH S |   3.7627119    2.785957         177
  HIGH SCHO |   3.4737631   3.1897666         667
  JUNIOR CO |   2.7978723   2.5968541          94
   bachelor |   2.2098214   2.1854989         224
   graduate |   1.9705882   1.4449974         136
```

Less than high school vs Bachelor’s degree: ES=(3.76-2.21)/sqrt((2.79^2+2.19^2)/2)=.62 This is a medium size difference → practically significant.

Less than high school vs Graduate degree: ES=(3.76-1.97)/sqrt((2.79^2+1.44^2)/2)=.81
This is a large difference → practically significant.

High school vs Bachelor’s degree: ES=(3.47-2.21)/sqrt((3.19^2+2.19^2)/2)=.46
This difference is small, but close to a cutoff of .5 so we would still consider it practically significant.

High school vs Graduate degree: ES=(3.47-1.97)/sqrt((3.19^2+1.44^2)/2)=.61
This is a medium size difference → practically significant.

f. How many pairwise comparisons in total are you making in (d)? Show how to use a formula to calculate that number.

Number of pairwise comparisons = k*(k-1)/2=5*4/2=10 comparisons

g. In (d), when you are requesting post-hoc comparisons using Bonferroni correction, how exactly does Stata adjust the p-values in the output (that is, how is the adjustment calculated)?

Stata calculates Bonferroni correction by multiplying each p-value by the number of comparisons – here, by 10.

h. Take the p-value for Less Than High School vs. Junior College comparison from the output in (d) and calculate what that p-value value would be <u>prior</u> to Bonferroni adjustment:

The p-value for less than high school vs junior college comparison is 0.07 after the Bonferroni adjustment, and it would be 0.07/10 = 0.007 prior to Bonferroni correction.

3. Does the Bonferroni adjustment make it easier or more difficult to reject the null hypothesis? Please circle your answer and explain.

a. Easier
[✓ b. More difficult]

Explanation:
Bonferroni adjustment makes it more difficult to reject the null hypothesis because p-values become higher after the adjustment, so it is more difficult to find p-values below our chosen alpha cutoff (which is exactly what we want, since we are trying to avoid inflated type I error that results from making multiple comparisons).

Stata output:

```
. oneway tvhours degree, means st obs bonferroni
            |  Summary of HOURS PER DAY WATCHING
 RS HIGHEST |               TV
     DEGREE |        Mean   Std. Dev.        Obs.
------------+------------------------------------
  LT HIGH S |   3.7627119    2.785957         177
  HIGH SCHO |   3.4737631   3.1897666         667
  JUNIOR CO |   2.7978723   2.5968541          94
   bachelor |   2.2098214   2.1854989         224
```

```
   graduate |   1.9705882   1.4449974         136
------------+------------------------------------
      Total |   3.0885978      2.8651        1298

                        Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      530.306175      4   132.576544     16.94     0.0000
 Within groups      10116.5051   1293   7.82405651
------------------------------------------------------------------------
    Total           10646.8112   1297   8.20879819

Bartlett's test for equal variances:  chi2(4) = 128.3549  Prob>chi2 = 0.000

         Comparison of HOURS PER DAY WATCHING TV by RS HIGHEST DEGREE
                                (Bonferroni)
Row Mean-|
Col Mean |   LT HIGH    HIGH SCH   JUNIOR C   bachelor
---------+--------------------------------------------
HIGH SCH |   -.288949
         |      1.000
         |
JUNIOR C |    -.96484   -.675891
         |      0.070      0.285
         |
bachelor |   -1.55289   -1.26394   -.588051
         |      0.000      0.000      0.874
         |
graduate |   -1.79212   -1.50317   -.827284   -.239233
         |      0.000      0.000      0.276      1.000
```