SOCY2200 Statistics **Instructor: Natasha Sarkisian**

Assignment 5. Hypothesis Testing for Means of One Sample or Two Independent Samples: Answers

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. After that, evaluate the probability that you could be making Type I and Type II error.*** (For the problems where you do calculations by hand, I would recommend also clearly writing down your intermediate steps; that way, you can still receive partial credit even if something goes wrong in one of the steps.)

1. A study aims to assess whether those who regularly volunteer for a specific organization have higher average levels of social engagement than the overall population. We know that the average score on the social engagement scale in the population is 17.3. We take a random sample of 40 people who regularly volunteer for this organization and obtain a mean of 18.3 and a standard deviation of 2.0.

   a. Using 95% confidence level, can we conclude that those regularly volunteer for this organization have higher average social engagement levels than the general population?
   b. After the assessment, evaluate the probability that you may be making Type I and Type II error.

a.
1. Null hypothesis: The average social engagement score for people who regularly volunteer for this organization is the same as in the population (17.3).
Research hypothesis: The average social engagement score for people who regularly volunteer for this organization higher than in the general population (higher than 17.3).
H_0: $\mu = 17.3$
H_1: $\mu > 17.3$ -- since we expect higher scores for those volunteering → use a directional research hypothesis → one-tailed test
2. We want to be 95% confident – we need to use alpha = 0.05 (1-.95=.05).
3. Test statistic – Student's t
4. Compute:
t = (18.3-17.3)/ (2/sqrt(40)) = 3.16

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{N}}}$$

5. Use Table B2 to find the critical value: df=n-1=40-1=39, alpha=0.05, one-tailed test → t_{crit} = 1.684
6. Does computed statistic exceed critical value? 3.16 is larger than 1.684 so it does exceed the critical value.
7. Decision: We can reject the null hypothesis in favor of our research hypothesis.
8. Substantive conclusion: Based on these data from 40 volunteers, we can conclude with 95% confidence that the average social engagement score for people who regularly volunteer for this organization is higher than in the general population (higher than 17.3). [Optional: The difference is statistically significant at .05 level. We can write our finding as: t = 3.16, p<.05 (one-tailed).]

b. Probability of Type I error is less than .05, and the probability of Type II error is 0 (since we rejected the null hypothesis).

2. A new drug has been developed, and preliminary evidence indicates that it may reduce people's migraine pain. A study aims to assess the effectiveness of this new drug against migraines. The 30 patients with this condition who volunteered for this study are randomly assigned to one of two groups: The first group will take the new drug for three months, while the second group will take a placebo instead. The following table presents the average levels of migraine pain that each of these patients reported over this time period.

a. Using 95% confidence level, can we conclude that the new drug is more effective than placebo?

b. After the assessment, evaluate the probability that you may be making Type I and Type II error.

c. Calculate the effect size and discuss the practical significance of this drug's effect.

Patient ID#	Migraine score for those taking the drug	Patient ID#	Migraine score for those taking placebo	**Drug group: (X – mean)**	**Drug group: (X – mean)^2**	**Placebo group: (X-mean)**	**Placebo group: (X-mean)^2**
1	5	16	7	**0.4**	**0.16**	**1.07**	**1.15**
2	4	17	8	**-0.6**	**0.36**	**2.07**	**4.28**
3	6	18	6	**1.4**	**1.96**	**0.07**	**0.01**
4	8	19	2	**3.4**	**11.56**	**-3.93**	**15.44**
5	4	20	8	**-0.6**	**0.36**	**2.07**	**4.28**
6	3	21	5	**-1.6**	**2.56**	**-0.93**	**0.86**
7	5	22	7	**0.4**	**0.16**	**1.07**	**1.14**
8	2	23	6	**-2.6**	**6.76**	**0.07**	**0.01**
9	6	24	7	**1.4**	**1.96**	**1.07**	**1.14**
10	5	25	5	**0.4**	**0.16**	**-0.93**	**0.86**
11	3	26	4	**-1.6**	**2.56**	**-1.93**	**3.72**
12	7	27	8	**2.4**	**5.76**	**2.07**	**4.28**
13	4	28	5	**-0.6**	**0.36**	**-0.93**	**0.86**
14	4	29	7	**-0.6**	**0.36**	**1.07**	**1.14**
15	3	30	4	**-1.6**	**2.56**	**-1.93**	**3.72**
Sum	**69**		**89**	**0**	**37.6**	**0**	**42.93**
Means & variances	**4.6**		**5.93**		**2.69**		**3.07**

Mean for drug group = 4.6 (SD=1.64), mean for placebo group = 5.93 (SD=1.75)

a.

1. Null hypothesis: There is no difference in migraine levels between those taking the experimental drug and those taking placebo in the population. Research hypothesis: Those

taking the experimental drug have lower migraine levels than those taking placebo in the population.
H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 < \mu_2$-- since we expect lower scores for those taking the experimental drug $\rightarrow$ use a directional research hypothesis $\rightarrow$ one-tailed test
2. We want to be 95% confident if we conclude that this program works – we need to use alpha = 0.05 (1-.95=.05).
3. Test statistic – Student's t
4. Compute:
Numerator: 4.6-5.93=-1.33
Denominator: sqrt[((15-1)*1.64*1.64 + (15-1)*1.75*1.75)/(15+15-2)* (15+15)/(15*15)] = 0.619
t = - 1.33 / 0.619 = -2.149
5. Use Table B2 to find the critical value: df=n1+n2-2=30-2=28, alpha=0.05, one-tailed test $\rightarrow$ $t_{crit} = 1.701$
6. Does computed statistic exceed critical value? 2.149 is larger than 1.701 so it does exceed the critical value
7. Conclusion about the null: We can reject the null hypothesis H_0: $\mu_1 = \mu_{2.}$
8. Substantive conclusion: Based on the data from 30 patients, we are 95% confident that, in the population, those treated with the experimental drug would have significantly lower migraine scores than those taking placebo (that is, we are 95% confident our drug is more effective against migraines than placebo).

b. The probability of type I error is based on our alpha, so less than 5%. The probability of type II error is 0 since we rejected the null hypothesis.

c. Calculating effect size:
Mean for drug group = 4.6 (SD=1.64), mean for placebo group = 5.93 (SD=1.75)

$$ES = \frac{Mean_{t1} - Mean_{t2}}{\sqrt{(SD_{t1}^2 + SD_{t2}^2)/2}}$$

Effect size (Cohen's D) = (5.93-4.6)/sqrt((1.64^2+1.75^2)/2)=.78
That is a medium size effect so it is practically significant as well as statistically significant.

3. Let's find out whether American women and men differ from each other in their TV watching habits using variables *tvhours* and *sex* from GSS2012 data.

a. Conduct the significance test using 90% confidence level.
b. After the assessment, evaluate the probability that you may be making Type I and Type II error.
c. Calculate the effect size and discuss the practical significance of this difference.
d. Construct a bar graph illustrating average hours of TV that American men and women watch.

a.
1. State your null and research hypotheses:
H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$
(where : μ_1 is average TV hours for women, and μ_2 is average TV hours for men)

Null hypothesis: American women and men do not differ from each other in their average hours of TV watched.
Research hypothesis: American women and men differ from each other in their average hours of TV watched.
2. 90% confidence level means alpha of .10.
3. Test statistic: Student's t .
4. We compute it using Stata:

```
. ttest tvhours, by(sex)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    male |     591    3.128596    .1289519    3.134885    2.875335    3.381856
  female |     707    3.055163    .0985412    2.620157    2.861694    3.248631
---------+--------------------------------------------------------------------
combined |    1298    3.088598    .0795248      2.8651    2.932587    3.244609
---------+--------------------------------------------------------------------
    diff |            .0734329    .1597371               -.2399387    .3868045
------------------------------------------------------------------------------
    diff = mean(male) - mean(female)                              t =   0.4597
Ho: diff = 0                                     degrees of freedom =     1296

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.6771         Pr(|T| > |t|) = 0.6458          Pr(T > t) = 0.3229
```

5. P-value associated for our chosen research hypothesis: `Pr(|T| > |t|) = 0.6458`
6. Compare p-value and alpha: 0.651>.10
7. Decision about H0: We fail to reject the null hypothesis. t = .4597, p>.1
8. Substantive conclusion. Based on this national representative sample of 1298 people, we do not have evidence that American women and men differ in the amount of TV they watch per day on average.

b. We failed to reject the null hypothesis, so the probability of Type I error is 0. The chances of Type II error are low because the sample size is so large (1298 observations).

c. Effect size:

$$ES = \frac{Mean_{t1} - Mean_{t2}}{\sqrt{(SD_{t1}^2 + SD_{t2}^2)/2}}$$

(3.13-3.06)/sqrt((3.13^2+2.62^2)/2)=.02

The effect size is very tiny, so this difference is neither statistically nor practically significant.

d. Bar graph:

```
. graph bar tvhours, over(sex) blabel(bar) ytitle("Average TV Hours")
```