Statistical Methods in Education Research
Vanderbilt University
Prof. Sean P. Corcoran

October 28, 2021
49 total points

Problem Set 7 Solutions

Instructions: Answer the following questions in their entirety in a separate document. Submit your completed problem set as a PDF document via email to `sean.corcoran@vanderbilt.edu`. Use your last name and problem set number as the filename (e.g., *Harris Problem Set 7.pdf*). Working together is encouraged, but it is expected that all submitted work be that of the individual student.

---

1. (**5 points**) A random sample of 172 new college graduates was asked to rate on a scale of 1 (not concerned) to 5 (very concerned) their level of concern about meeting future student loan debt obligations. The sample mean rating was 3.17 and the sample standard deviation was 0.70. Test at the 1% significance level the null hypothesis that the population mean rating is 3.0 against the alternative that it is bigger than 3.0. Write down the null and alternative hypotheses, and report your test statistic, p-value, and conclusion. (Note: this variable is ordinal-measured and arguably not appropriate for such a test. However, we will treat it as an interval measure here).

$$H_0 : \mu = 3.0$$
$$H_1 : \mu > 3.0$$

The test statistic is:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{3.17 - 3.00}{0.70/\sqrt{172}} = 3.185$$

In other words, the sample mean of 3.17 is 3.185 standard errors above the hypothesized population mean of 3.0. The probability of obtaining this test statistic (if the null were true) is $p = P(t > 3.185) =< 0.001$ (uses $df = 171$). Because $p < \alpha = 0.01$, we reject H_0 in favor of H_1.

In Stata: `display ttail(171, 3.185).`

2. (**5 points**) Children graduating from a particular high school in New York City score an average of 55 on a college-readiness test. A random sample of 36 students from this school is selected to participate in Project Advance, a special program to help high school students prepare for college. At the end of the program, the 36 students are given the college-readiness test. They obtain a sample mean of $\bar{x} = 58.5$ and a sample standard deviation of $s = 12$. Conduct a hypothesis test at significance level $\alpha = 0.05$

to determine whether students who participate in Project Advance score *higher* on average on the college readiness test than those in the school as a whole. State your null and alternative hypotheses, your test statistic, p-value, and conclusion.

$$H_0 : \mu = 55$$
$$H_1 : \mu > 55$$

The test statistic is $t = 1.75$ (below). Because s is used instead of σ, we know this test statistic has a t distribution with $n - 1 = 35$ degrees of freedom:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{58.5 - 55}{12/\sqrt{36}} = 1.75$$

From the t-distribution table (or Stata), the p-value is $p = P(t > 1.75) = 0.0446$. Because $p < \alpha$, we reject H_0 in favor of H_1.

Using Stata, the p-value would be found using `display ttail(35, 1.75)`.

3. (**4 points**) Same-sex marriage was legalized across Canada by the Civil Marriage Act enacted in 2005. Was this decision supported by a majority of the Canadian population? A poll conducted in July 2005 of 1,000 Canadians asked whether this bill should stand or be repealed. The responses were: 55% should stand, 39% should repeal, 6% don't know. Let π represent the population proportion of Canadians who believe it should stand. For testing $H_0 : \pi = 0.50$ against $H_1 : \pi > 0.50$:

(a) Find the standard error, and interpret in words.

The standard error is 0.0158 (below). This is a measure of variability in the sample proportion ($\hat{\pi}$) when the sample size is 1000. Note that the value of π assumed under the null hypothesis is used, because we are interested in the sampling distribution of $\hat{\pi}$ when H_0 is true.

$$\sqrt{\frac{\pi_0(1 - \pi_0)}{n}} = \sqrt{\frac{0.50(1 - 0.50)}{1000}} = \mathbf{0.0158}$$

(b) Find the test statistic, and interpret.

The test statistic is 3.16 (below). If H_0 is true, our sample proportion is 3.16 standard errors above the population mean of 0.50.

$$z = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0.55 - 0.50}{0.0158} = \mathbf{3.16}$$

(c) Find the p-value, and interpret in this context.

From the standard normal distribution, $P(z > 3.16) = 0.0008$. If H_0 is true, the probability of obtaining a random sample of 1000 with a sample proportion this far from the population mean would be 0.0008 (i.e., very unlikely). We can use the standard normal *(z)* distribution to find p, since the standard deviation is "known" under the null. (H_0 tells us that $\pi = 0.50$, and that piece of information alone determines the standard deviation of a proportion).

(d) What is your conclusion about H_0?

We can **reject** H_0 since $p < \alpha$. (We can assume an α of 0.05, but we would reject even under a more stringent α of 0.01).

4. (**6 points**) According to a union agreement, the mean income for all senior-level assembly-line workers in a large company is \$500 per week. A representative of a women's group decides to analyze whether the mean income μ for female employees matches this norm. From a random sample of nine female employees, the average weekly income is $\bar{x} = \$410$ and $s = \$90$.

(a) Test whether the mean income of female employees differs from \$500 per week. Report your hypotheses, test statistic, p-value, and conclusion. Use a significance level of $\alpha = 0.01$. (**4 points**)

The hypothesis test can be framed like this: how likely is it that a sample mean of \$410 could have occurred, if the true mean income of female employees was \$500?

$$H_0 : \mu = 500$$

$$H_1 : \mu \neq 500$$

The test statistic is:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{410 - 500}{90/\sqrt{3}} = -3$$

The sample mean is 3 standard errors below the hypothesized population mean of 500. The probability of obtaining a test statistic this far from the mean (if the null were true) is $p = 2 * P(t < -3) = 2 * 0.0085 = 0.017$ (uses $df = 8$). Because $p > \alpha = 0.01$, we cannot reject H_0 in favor of H_1. Despite strong suggestive evidence that the mean income for women differs from \$500, we cannot reject H_0 at the 0.01 level. (The main reason in this case is the small sample size).

In Stata: `display 2*(1-ttail(8, -3))`.

(b) Report the p-value for the one-sided alternative hypothesis $H_1 : \mu < \$500$. Would this change your conclusion? Again use $\alpha = 0.01$. (**2 points**)

The p-value for a one-sided alternative is just $P(t < -3) = 0.0085$, half of that found in part (a). In this case we *would* reject H_0 since $\alpha < 0.01$.

5. (**4 points**) The p-value from a statistical test for μ with n=25 was $p = 0.05$.

(a) Find the t statistic that has this p-value for the following alternative hypotheses: (i) $H_1 : \mu \neq 0$, (ii) $H_1 : \mu > 0$, (iii) $H_1 : \mu < 0$. (**2 points**)

(b) Does this p-value provide stronger, or weaker evidence against the null hypothesis than $p = 0.01$? Explain. (**2 points**)

The test statistic t with this p-value of 0.05 can be found as follows:

- H_1: $\mu \neq 0$. This is a two-sided test in which the p-value 0.05 is split across two tails (0.025 in each). We need to determine the t for which the probability of exceeding t is 0.025. (By symmetry, the probability of falling below $-t$ is also 0.025). Using Stata, `display invttail(24, 0.025)` yields **2.064**. The probability p of drawing a sample with a test statistic t larger than 2.064 or smaller than -2.064 is 0.05.
- H_1: $\mu > 0$. This is a one-sided test in which the p-value 0.05 is in the right tail. We need to determine the t for which the probability of exceeding t is 0.05. Using Stata, `display invttail(24, 0.05)` yields **1.711**. The probability p of drawing a sample with a test statistic t larger than 1.711 is 0.05.
- H_1: $\mu < 0$. This is a one-sided test in which the p-value 0.05 is in the left tail. We need to determine the t for which the probability of falling below t is 0.05. Using Stata, `display invttail(24, 0.95)` yields **-1.711**. The probability p of drawing a sample with a test statistic t smaller than -1.711

is 0.05. Notice I used 0.95 in `invttail` here since Stata gives you the t value with the probability *above* a t, not below. You could have also recognized that the t value is just be the negative of the one in scenario *ii*.

A p-value of 0.05 would provide **weaker** evidence against the null hypothesis than p=0.01. When p=0.01, one incorrectly rejects the null hypothesis in 1/100 random samples. This is a higher threshold than p=0.05, where an incorrect rejection occurs in 5/100 random samples.

6. (**5 points**) Below are the results of a hypothesis test about the population mean self concept score in 8th grade (*slfcnc08*), using a random sample of 25 students. Some information from the table has been removed. Given what remains in the table, what is the standard error of the sample mean? Explain how you arrived at your answer.

```
. ttest slfcnc08=20

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
slfcnc08 |      25       21.96                            19.62959    24.29041
------------------------------------------------------------------------------
    mean = mean(slfcnc08)                                         t =
Ho: mean = 20                                    degrees of freedom =       24

    Ha: mean < 20                Ha: mean != 20                 Ha: mean > 20
 Pr(T < t) = 0.9523         Pr(|T| > |t|) = 0.0954          Pr(T > t) = 0.0477
```

There are two ways one can find this answer. One is to recognize that the 95% confidence interval is calculated as $\bar{x} \pm t * se(\bar{x})$. The t that would have been used in this CI is $t(df, \alpha/2) = t(24, 0.025) = 2.0639$. With an upper bound of 24.29041 in the CI, this would imply that: $se(\bar{x}) = (24.29041 - 21.96)/2.0639 = \mathbf{1.129}$.

Another way would be to start with the p-value of 0.0477 and to figure out what test statistic it is associated with. Using `display invttail(24,0.0477)`, this t-statistic is 1.7359. This value represents the number of standard errors $\bar{x}$ is from the null of 20. Since $1.7359 = (21.96 - 20)/se(\bar{x})$ then $se(x) = (21.96 - 20)/1.7359 = \mathbf{1.129}$.

For questions #7-8, use the Stata dataset *Grade4_classrooms.dta* on Github. This file represents a random sample of 4th grade classrooms located in urban school districts in Texas. Each observation is a classroom, and the variables either describe the teacher (e.g., total teaching experience, teacher race/ethnicity), or are an average for the classroom (e.g., math z-scores, % of students who are economically disadvantaged). All data are from 2006.

7. (**14 points**) For this problem you will be conducting two-sided hypothesis tests for the classroom mean math score (*mathz_class*) and mean reading score (*readz_class*) in urban Texas school districts.

   (a) Evaluate the tenability of the normality assumption for these variables or indicate why this assumption is not an issue for this analysis (i.e., that the results of a t-test would not be compromised by a failure to meet this assumption). (**2 points**)

   (b) For the state of Texas, the mean classroom math and reading (z-)scores are both 0. Write down the null and alternative hypotheses for t-tests to determine whether the mean urban 4th grade classroom average z-scores for math and reading are zero, or are different from zero. (**2 points**)

   (c) Using the appropriate Stata command, what are the test statistics and p-values associated with these tests? (**2 points**)

   (d) Use the p-values in each case to determine whether or not H_0 can be rejected in favor of the alternative. Use a significance level of $\alpha = 0.05$. (**2 points**)

   (e) Provide 95% confidence intervals for the mean math and reading z-scores in urban Texas classrooms. (**2 points**)

   (f) Use the confidence intervals found in part (e) to conduct the tests in parts (b)-(d). Are the results consistent? Why or why not? (**2 points**)

   (g) Would you be more or less likely to reject the null hypotheses if the sample size had been 1,000 classrooms, rather than the sample size(s) used above? Explain. (**2 points**)

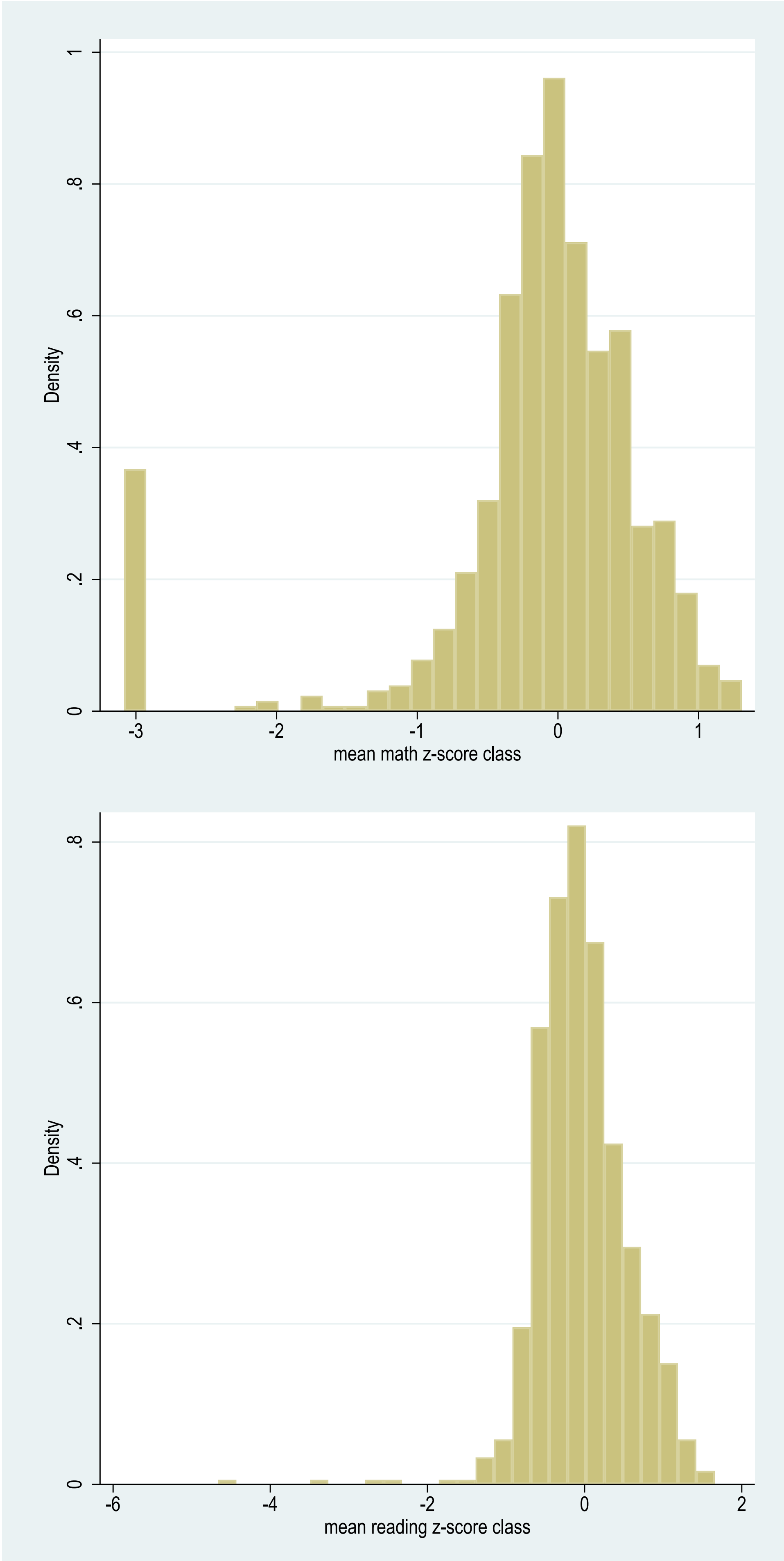
> See attached Stata log.

8. (**6 points**) Use the same data from Question #7 to test whether urban Texas classrooms on average exclude more than 14 percent of their students from state testing because of special education accommodations. (The relevant variable is *excl_sped*). Follow the same procedure as parts (a)-(d) in Question #8 to carry out this test.

> See attached Stata log. Since *excl_sped* is a proportion, I'll refer to it as π. The hypothesis being tested is:
>
> $H_0 : \pi = 0.14$ and $H_1 : \pi > 0.14$
>
> The t-statistic is 2.8522 with a p-value for this alternative hypothesis of 0.0022. Since $p < \alpha = 0.05$, we can reject at the 95% significance level the null hypothesis

in favor of the alternative that the exclusion rate is higher than 14%.

```
.
. // ***********************************************************************
. // LPO.8800 Problem Set 7 - Solutions to Questions 7-8
. // Last updated: October 22, 2021
. // ***********************************************************************
.
. cd "$db\_TEACHING\Statistics I - PhD\Problem sets\Problem set 7"
C:\Users\corcorsp\Dropbox\_TEACHING\Statistics I - PhD\Problem sets\Problem set 7
.
. // *************
. // Question 7
. // *************
.
. use https://github.com/spcorcor18/LPO-8800/raw/main/data/Grade4_classrooms.dta, clear
.
. // ********
. // Part a
. // ********
. histogram mathz_class, name(hist1, replace) nodraw
(bin=28, start=-3.0838358, width=.15680618)
. histogram readz_class, name(hist2, replace) nodraw
(bin=27, start=-4.6651998, width=.23412534)
.
. graph combine hist1 hist2 , xsize(3) ysize(6) col(1)
. graph export hists.pdf, as(pdf) replace
(file hists.pdf written in PDF format)
```



Density
0
.2
.4
.6
.8
1
-3
-2
-1
0
1
mean math z-score class
Density
0
.2
.4
.6
.8
-6
-4
-2
0
2
mean reading z-score class

```
.
. // Both distributions are left skewed, and the math distribution has a mass of
. // observations that are especially low (around -3). Because our sample size is
. // large, the CLT predicts that means calculated from samples drawn of this size
. // will be approximately normally distributed. There is a greater risk of
. // departure from normality for math vs. reading, so the mass of low
. // observations are worth further investigation.
.
. // **********
. // Parts b-d
. // **********
. // H0: mu_m = 0 and H1: mu_m neq 0
. // H0: mu_r = 0 and H1: mu_r neq 0
.
. // Results shown below. The t-statistic for math is -5.2303. For reading, it is
. // -1.2237. The p-values for a two-sided hypothesis test are <0.0001 and 0.2215,
. // respectively.
.
. ttest mathz_class=0
One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
mathz_~s |     816   -.1590247    .0304047     .868533   -.2187056   -.0993439
------------------------------------------------------------------------------
    mean = mean(mathz_class)                                      t =  -5.2303
Ho: mean = 0                                     degrees of freedom =      815
    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
. ttest readz_class=0
One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
readz_~s |     765   -.0255273    .0208613    .5769953   -.0664796     .015425
------------------------------------------------------------------------------
    mean = mean(readz_class)                                      t =  -1.2237
Ho: mean = 0                                     degrees of freedom =      764
    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.1107         Pr(|T| > |t|) = 0.2215          Pr(T > t) = 0.8893
.
. // Because p<alpha for math, we reject the null hypothesis that the mean
. // classroom math score in urban Texas districts is the same as the statewide
. // average of zero. As p>alpha for reading, we do not reject the null hypothesis
. // that the mean classroom reading score in urban Texas districts is the same as
. // the statewide average of zero. (It is notable, however, that there are a
. // number of missing values for readz_class. One reason this may be is that
. // some classrooms are entirely English language learners who are exempted
. // from the reading test).
.
. // **********
. // Parts e-f
. // **********
. // 95% confidence intervals
.
. // The 95% confidence intervals are provided in the Stata output below:
. // (-0.219, -0.099) for math, and (-0.066, 0.0015) for reading.
.
. // The mean under the null hypothesis (0) is outside the 95% confidence
```

```
. // interval for math, so we can reject that null hypothesis. Zero is contained
. // within the 95% confidence interval for reading, so we cannot reject that
. // null hypothesis. These are the same conclusions reached using the p-value
. // method.
.
. mean mathz_class
Mean estimation                      Number of obs   =        816
--------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
 mathz_class |  -.1590247   .0304047     -.2187056   -.0993439
--------------------------------------------------------------
. mean readz_class
Mean estimation                      Number of obs   =        765
--------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
 readz_class |  -.0255273   .0208613     -.0664796     .015425
--------------------------------------------------------------
.
. // **********
. // Part g
. // **********
.
. // You would be more likely to reject the null hypothesis with a larger sample
. // size. All else equal, a larger n is associated with a smaller p-value.
. // Another way to think about this is that the confidence interval around x-bar
. // will be narrower, making it less likely that zero will be contained within it.
.
.
.
. // *************
. // Question 8
. // *************
.
. ttest excl_sped=0.14
One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
excl_s~d |     835    .1659154     .009086    .2625515    .1480813    .1837494
------------------------------------------------------------------------------
    mean = mean(excl_sped)                                        t =   2.8522
Ho: mean = 0.14                                  degrees of freedom =      834
   Ha: mean < 0.14             Ha: mean != 0.14              Ha: mean > 0.14
 Pr(T < t) = 0.9978         Pr(|T| > |t|) = 0.0044          Pr(T > t) = 0.0022
.
.
. capture log close
```