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

November 4, 2021
**50 total points**

## Problem Set 8 Solutions

1. You are up late studying, and hungry. You use your last $2 to buy a sandwich out of a vending machine in the basement of Payne Hall. As you unwrap your treat, you notice the inside of the sandwich is a funny color. You ponder whether you should eat the sandwich. In this scenario, what are the Type I and Type II errors associated with your decision to eat the sandwich? (Let $H_0$ be "the sandwich is safe"). (**3 points**)

   In this scenario the null hypothesis ($H_0$) is "the sandwich is safe to eat". The alternative ($H_1$) is that the sandwich is bad, and likely to make you sick. A Type I error is rejecting the null hypothesis when it is true: here, deciding *not* to eat the sandwich when it is actually safe to eat. A Type II error is *not* rejecting the null hypothesis when it is false: here, deciding to eat the sandwich when it is bad.

2. After recovering from a mysterious sandwich-based food illness, you set out to study whether charter school students perform *better* in mathematics, on average, than traditional public school students. You know that math scores in the population of public school students have a normal distribution with a mean of 420 and standard deviation of 84. You plan to administer the same test to a sample of charter school students to test your hypothesis. Use Stata or a web applet to answer the following questions. (**14 points**)

   (a) What are the null and alternative hypotheses in this problem? (**2 points**)

   Let $\mu$ be the mean test performance of charter school students in the population.

   $$H_0 : \mu = 420$$

   $$H_1 : \mu > 420$$

   (b) Describe what a Type I and Type II error would be in this study. (**2 points**)

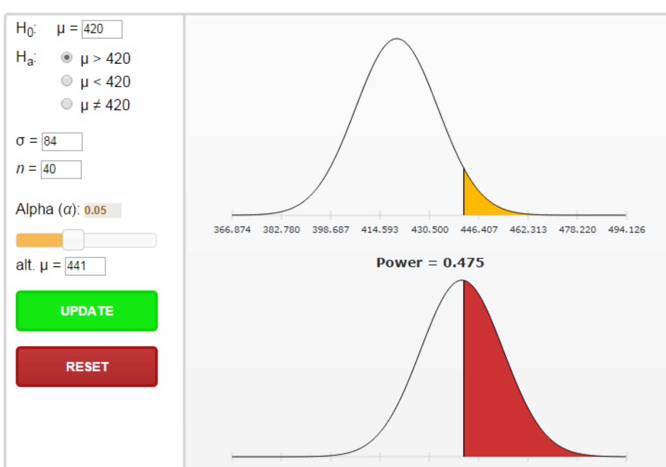
   A Type I error is rejecting $H_0$ when it is true. In this example, it would mean concluding that charter school students perform better, on average, than traditional public school students, when in fact they do not. A Type II error is *not* rejecting $H_0$ when it is false. In this example, it would mean concluding that charter school students perform no better, on average, than traditional public school students, when in fact they do.

(c) You consider a meaningful difference in test scores to be +21 points. What specific alternative $\mu$ would this imply? How large an effect size (Cohen's d) would this be? (**4 points**)

A 21 point difference in test scores would mean an alternative $\mu$ of $420 + 21 = \mathbf{441}$. In terms of an effect size (using Cohen's d), this would be: $(21/84) = \mathbf{0.25}$ sd units. In other words, charter school students would be performing 0.25 standard deviations better than non-charter students if their mean score was 441.

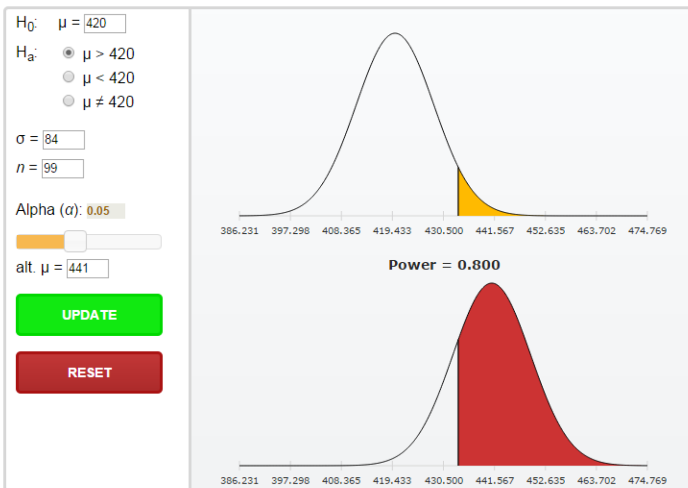
(d) If you plan to test 40 randomly selected charter school students, what is the power of the test? What is the probability of a Type II error? Use significance level $\alpha = 0.05$. (**3 points**)

Using the applet, the power of the test $(1 - \beta)$ would be **0.475**. The probability of a Type II error ($\beta$) would be **0.525**. In words, there is a 0.525 probability of not rejecting $H_0$ when $\mu$ is actually 441. This is a fairly high probability of not detecting a rather large effect size. If the true difference between charter and non-charter students were 21 points (0.25 sd), one would hope that a study design would be able to identify such effects. To put the power of the test in words, there is a 0.475 probability of rejecting $H_0$ when $\mu$ is actually 441 (again, not a very high probability).



(e) Holding everything else constant, how large of a sample would you need in order to increase the power to 0.80? (**3 points**)

> One way to determine this is through trial and error with the power applet. Keep increasing $n$ until the power rises to 0.80. The required sample size is **99** or more students (see below). You could also use Stata's `power` command to find the minimum sample size (also shown below).




```
. power onemean 420 441, sd(84) knownsd onesided

Estimated sample size for a one-sample mean test
z test
Ho: m = m0  versus  Ha: m > m0

Study parameters:

        alpha =    0.0500
        power =    0.8000
        delta =    0.2500
           m0 =  420.0000
           ma =  441.0000
           sd =   84.0000

Estimated sample size:

            N =        99
```

3. A decision is planned in a test of $H_0 : \mu = 0$ against the alternative $H_a : \mu > 0$, using $\alpha = 0.05$. If the actual $\mu$ is 5, the probability of a Type II error ($\beta$) is 0.17. (**6 points**)

   (a) Explain the meaning of the last sentence in words. (**2 points**)

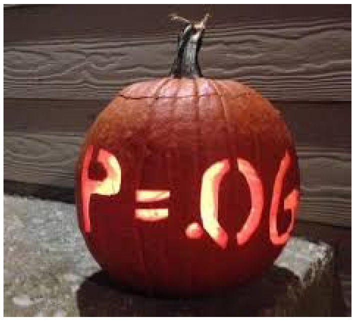
   > The probability of a Type II error ($\beta$) is the probability one does not reject $H_0$ when it is false. Put another way, in 17% of random samples this test will not reject $H_0$ if $\mu$ is 5 rather than 0.

   (b) If the test used $\alpha = 0.01$, would the probability of a Type II error be less than, equal to, or greater than 0.17? Explain. (**2 points**)

   > With an $\alpha = 0.01$ rather than $\alpha = 0.05$, you reject $H_0$ less often. Since $\beta$ is the probability one does not reject $H_0$ when it is false, a lower $\alpha$ <u>increases</u> $\beta$.

   (c) If $\mu$ were actually 10, would the probability of a Type II error be less than, equal to, or greater than 0.17? Explain. (**2 points**)

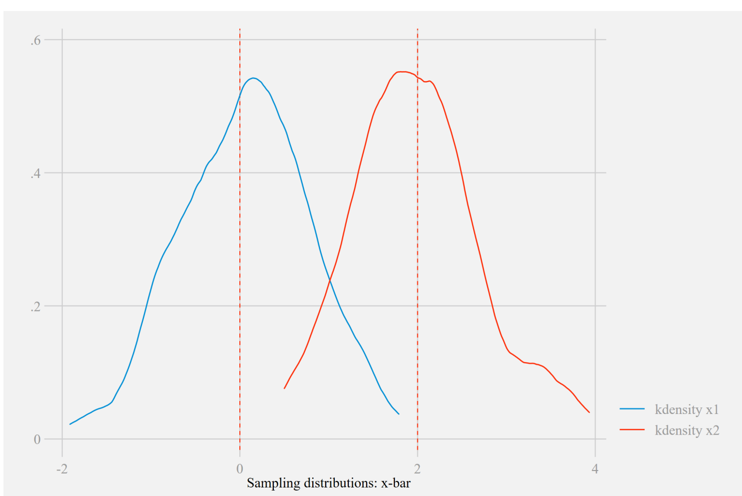
   > If the true $\mu = 10$, it is much less likely that the test will fail to reject $H_0 : \mu = 0$. This means a <u>lower</u> Type II error rate.



4. The Stata syntax below will create a dataset with 200 sample means for $\bar{x}_1$ and $\bar{x}_2$, and 200 sample standard deviations $s_1$ and $s_2$. Each sample consists of $n = 50$ random draws from normal distributions in which $x_1 \sim N(0, 5)$ and $x_2 \sim N(2, 5)$. (That is, $x_1$ has a normal distribution with mean zero and standard deviation 5; $x_2$ has a normal distribution with mean 2 and standard deviation 5). (**18 points**)

   (a) Run the Stata syntax above, and choose your own starting seed value (#). Take care that your quotations are correct, as they may not copy over correctly from this document into a do-file. Use the `twoway` graphing function with `kdensity` to overlay the resulting distributions of $\bar{x}_1$ and $\bar{x}_2$ (see below). `kdensity` is a kernel density function, a kind of "smoothed" histogram. (**3 points**)

```
set seed 1989

forvalues j=1/200 {
   clear
   tempfile sample`j'
   set obs 50
   gen x1 = rnormal(0,5)
   gen x2 = rnormal(2,5)
   collapse (mean) x1 x2 (sd) sx1=x1 sx2=x2
   gen n= 50
   save `sample`j''
   }
use `sample1', clear
forvalues j=2/200 {
   append using `sample`j''
   }
twoway (kdensity x1) (kdensity x2), xtitle(Sampling distributions: x-bar) ///
   xline(0) xline(2)
```



(b) Create two new variables that contain the lower and upper bounds of a 95% confidence interval for $\mu$. (Each resulting $\bar{x}_1$—all 200 of them—will have a confidence interval associated with it). Be sure to use the sample standard deviation when creating your confidence interval, not the population standard deviation, and continue to do so in part (c). In what percentage of samples does your confidence interval contain the true population mean $\mu = 0$? In what percentage does it not? (Hint: you can create a third variable that flags whether the CI contains zero or doesn't). (**3 points**)

> Stata syntax and results below. In my case, 93% of random samples result in confidence intervals that contain the $\mu$ of 0 (and in 7% they do not). Notice the confidence interval is constructed using the $t$-distribution, since I am using $s$ rather than the known standard deviation $\sigma$.

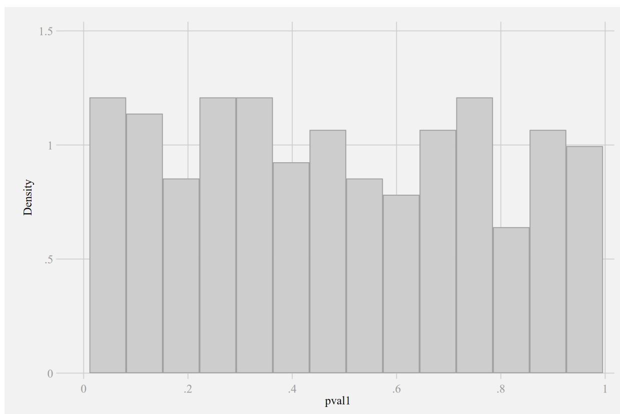
```
// confidence intervals
gen lb1 = x1 - invttail(49, 0.025)*(sx1/sqrt(50))
gen ub1 = x1 + invttail(49, 0.025)*(sx1/sqrt(50))

// contains mu=0
gen contains01 = (lb1<0 & ub1>0)
tabulate contains01

 contains01 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |         14        7.00        7.00
          1 |        186       93.00      100.00
------------+-----------------------------------
      Total |        200      100.00
```

(c) Create two new variables that equal the $t$-statistic and $p$-value from a test of $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$. (Each resulting $\bar{x}_1$ will have a $t$ and $p$-value for this test). Create a histogram to show the resulting distribution of $p$-values across the 200 random samples. In what percentage of these random samples is $p < 0.05$? Explain in words what this percentage represents. (**4 points**)

> Stata syntax and results below. In my case, 7% of random samples have a $p < 0.05$, and 93% do not. In a two-tailed test this should produce the same conclusion as the 95% confidence interval test in part (b). Notice the $p$-value is found with reference to the $t$-distribution, since I am using $s$ rather than the known standard deviation $\sigma$. These 7% of random samples are instances in which one improperly rejects $H_0$—a Type I error.

```
// tstat and pvalue
gen tstat1 = (x1-0)/(sx1/sqrt(50))
gen pval1 = 2*ttail(49, abs(tstat1))
gen sig1 = pval1<=0.05
tabulate sig1
histogram pval1
```

```
       sig1 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        186       93.00       93.00
          1 |         14        7.00      100.00
------------+-----------------------------------
      Total |        200      100.00
```




(d) You are still interested in the null hypothesis $H_0 : \mu = 0$. However, suppose that—unbeknownst to you—the sample means $\bar{x}_2$ (not the $\bar{x}_1$) reflect draws from the true distribution of $x$. That is, $x$ has a mean of 2, not 0. As in part (b), create two new variables that contain the lower and upper bounds of a 95% confidence interval for $\mu$, based on $\bar{x}_2$ (and $s_2$). In what percentage of random samples does your confidence interval contain *zero*? Explain in words what this percentage represents. (**4 points**)

Stata syntax and results below. In my case, 23% of random samples result in confidence intervals that contain the hypothesized mean of 0 (and in 77% they do not). This 23% represents a Type II error: failing to reject $H_0 : \mu = 0$ when it is false.

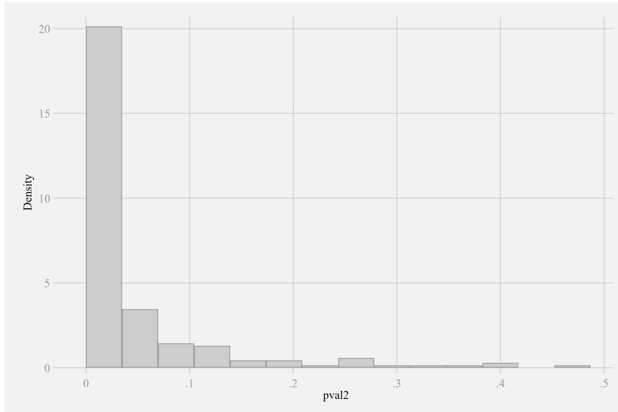
```
// now using x2

// confidence intervals
gen lb2 = x2 - invttail(49, 0.025)*(sx2/sqrt(50))
```

```
gen ub2 = x2 + invttail(49, 0.025)*(sx2/sqrt(50))

// contains zero
gen contains02 = (lb2<0 & ub2>0)
tabulate contains02

 contains02 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        154       77.00       77.00
          1 |         46       23.00      100.00
------------+-----------------------------------
      Total |        200      100.00
```

(e) Repeat part (c) using $\bar{x}_2$ and continue to test the hypothesis $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$. In what percentage of these random samples is $p < 0.05$? In what percentage of these random samples is $p > 0.05$? Explain in words what these percentages represent. (**4 points**)

> Stata syntax and results below. In my case, 77% of random samples have a $p < 0.05$, and 23% do not. Again, this 23% represents a Type II error: failing to reject $H_0 : \mu = 0$ when it is false.

```
// tstat and pvalue
gen tstat2 = (x2-0)/(sx2/sqrt(50))
gen pval2 = 2*ttail(49, abs(tstat2))
gen sig2 = pval2<=0.05
tabulate sig2
histogram pval2

       sig2 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |         46       23.00       23.00
          1 |        154       77.00      100.00
------------+-----------------------------------
      Total |        200      100.00
```




5. Run the same simulation as in Question #4, but draw `x1` from the exponential distribution, rather than the normal (see syntax below). You can omit the references to $x_2$ in this problem. Recall from Problem Set 5 that the exponential distribution is heavily skewed, and that the population mean of $x_1$ is 1/2 here. (**9 points**)

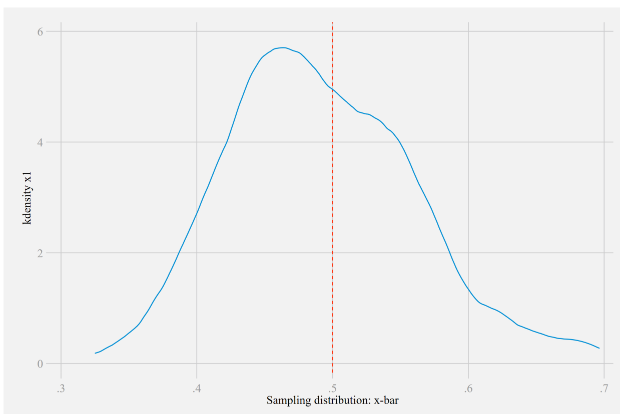
```
gen x1=rexponential(0.5)
```

(a) Create two new variables that contain the lower and upper bounds of a 95% confidence interval for $\mu$. In what percentage of samples does your confidence interval contain the true population mean $\mu = 0.5$? In what percentage does it not? (**3 points**)

> Stata syntax and results below. In my case, 93% of random samples result in a confidence interval that contains the $\mu$ of 0.5 (and in 7% they do not).

```
clear all
set seed 1989
set more off

forvalues j=1/200 {
   clear
   tempfile sample`j'
   set obs 50
   gen x1 = rexponential(0.5)
   collapse (mean) x1 (sd) sx1=x1
   gen n= 50
   save `sample`j''
   }
use `sample1', clear
forvalues j=2/200 {
   append using `sample`j''
```

```
   }

twoway (kdensity x1), xtitle(Sampling distribution: x-bar) ///
   xline(0) xline(0.5)

// confidence intervals
gen lb1 = x1 - invttail(49, 0.025)*(sx1/sqrt(50))
gen ub1 = x1 + invttail(49, 0.025)*(sx1/sqrt(50))

// contains mu=1/2
gen contains01 = (lb1<0.5 & ub1>0.5)
tabulate contains01

 contains01 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |         14        7.00        7.00
          1 |        186       93.00      100.00
------------+-----------------------------------
      Total |        200      100.00
```

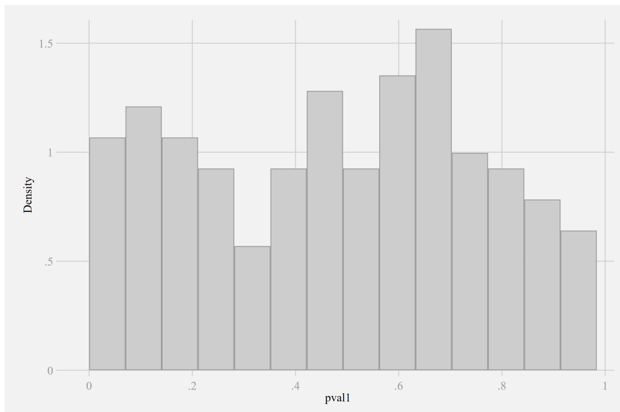



(b) Create two new variables that contain the $t$-statistic and $p$-value from a test of $H_0 : \mu = 0.5$ against $H_1 : \mu \neq 0.5$. In what percentage of these random samples is $p < 0.05$? (**3 points**)

> Stata syntax and results below. In my case, 7% of random sample result in a rejection of $H_0$, and 93% do not.

```
// tstats and pvalues
gen tstat1 = (x1-0.5)/(sx1/sqrt(50))
gen pval1 = 2*ttail(49, abs(tstat1))
gen sig1 = pval1<0.05
tabulate sig1
histogram pval1
```

| sig1 | Freq. | Percent | Cum. |
|---|---|---|---|
| 0 | 186 | 93.00 | 93.00 |
| 1 | 14 | 7.00 | 100.00 |
| Total | 200 | 100.00 | |




(c) The expected coverage rate of the confidence interval (part a) and the hypothesis test $p$-values rely on (approximate) normality of the sampling distribution. Your original variable $x_1$ here was heavily skewed. How did the CI and hypothesis test perform in this case, with the skewed $x_1$? Explain. (**3 points**)

The confidence intervals and hypothesis tests performed quite well, even with a heavily skewed $x_1$. With normality, we expect a Type I error in 5% of random samples. In this example, a Type I error would occur in 7% of random samples, which is no different (in my case) than in Question 4 when the original variable had a normal distribution.