PUBH 501 Biostatistics

T-Tests

What we will learn today

- Check normality
- Paired t-test
- Check variance
- Two-sample t-test

Examine your data first!

Load your dataset and check its contents using Data Editor

Use the **describe** and **codebook** commands to investigate your variables

Label variables:

label define educstatus 0 "HS/GED" 1 "Did not finish HS"
label val educnew educstatus
tab educnew

label var cesd0 "CES-D at first interview (third trimester)"
label var cesd1 "CES-D at second interview (1 month post delivery)"

Summary Statistics

- From last week, we can use the –summ- and –tabstat- commands to look a the summary statistics of our variables
- Use wildcard “*” to look at all variables that share the same text that prefaces the *
- For example: summ h* //will summarize all variables that start with “h”

```
summ h*
    Variable |       Obs        Mean    Std. dev.       Min        Max
-------------+--------------------------------------------------------
hours_zoom |       154    7.837662    7.677929         0         47
hw_min |             202    68.86881    46.85687         0        255
hw_minutes~3 |     163    92.66871    102.9772         1       1110
hotdog_san~h |      0
```

- tabstat var*, stats(n, mean, var, p25, p50, p75)

Standard Error of the Mean (SEM)

- Use the –mean- command to find the standard error of the mean
- Use the –summ- command to see the standard deviation…
- `mean cesd*`
- `summ cesd*`
- `tabstat cesd*, stats(n, mean, var, p25, p50, p75)`
- `tabstat cesd0 cesd1, stats(n, mean, var, p25, p50, p75, SEM)`

Standard Error of the Mean (SEM)

- Look at summary statistics and histogram
- Add a normal curve to the histogram to help with visualization
- Mean, median, and mode close in value?
- Sample size >= ~ 30

- summ cesd0, detail
- tabstat cesd0 cesd1, stats(n, mean, var, p25, p50, p75)
- histogram cesd0, freq normal

Paired T-Test

Tests if the means of dependent observations are different

- One individual at different times
- Matched samples
- Here we will examine CES-D score(depression measure) during pregnancy, and one month after delivery
- Each participant has two values of CES-D

list STUDYID cesd0 cesd1 in 1/10

ttest cesd0 = cesd1

ttest cesd0 = cesd1

```
Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
   cesd0 |     351    13.31339    .5777698    10.82451    12.17705    14.44973
   cesd1 |     351    11.63023    .5767115    10.80469    10.49597    12.76449
---------+--------------------------------------------------------------------
    diff |     351    1.683161    .7274479    13.62873    .2524417     3.11388
------------------------------------------------------------------------------
     mean(diff) = mean(cesd0 - cesd1)                              t =   2.3138
 Ho: mean(diff) = 0                     degrees of freedom =      350

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.9894         Pr(|T| > |t|) = 0.0213          Pr(T > t) = 0.0106
```

Mean difference between CES-D pre and post delivery is 1.68

With p=0.02; significant at the p

Two-Sample T-Test

- Using the same data, we can compare the CES-D scores of different groups
- Our question: does the CES-D score vary post-delivery (CESD1) for women based on education level?

- `summary stats… use bysort`
- `bysort educnew: summ cesd1`
- `bysort educnew: tabstat cesd1, stats(n, mean, var, p25, p50, p75, SEM)`

Checking Assumptions

- Normality in each group
  - Summary statistics
  - Histogram
  - Sample size
- Independent samples
- Equality of variance

Assumption of Equal Variance

Can check if the variances are equal. If not, we need to run our t-test accordingly
using –sdtest- command

- If a two-tailed t-test is significant, the variance is not equal
- null hypothesis is that the variance is equal between groups

- sdtest cesd1, by(educnew) // significant at p<0.05: variance unequal, reject null variance equal
- sdtest variable1=variable2, unpaired //for data set differently

graph box cesd1, over(educnew)

Two-sample t-test with unequal variance

```
ttest cesd1, by(educnew) unequal
```

highlighted two-tailed t-test

```
Two-sample t test with unequal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      no |     209    9.477965    .6709951    9.700464    8.155142    10.80079
     yes |     142      14.798    .9714168    11.57577    12.87757    16.71842
---------+--------------------------------------------------------------------
combined |     351    11.63023    .5767115    10.80469    10.49597    12.76449
---------+--------------------------------------------------------------------
    diff |           -5.320033    1.180629               -7.64458   -2.995487
------------------------------------------------------------------------------
    diff = mean(no) - mean(yes)                                   t =  -4.5061
Ho: diff = 0                     Satterthwaite's degrees of freedom =  266.518

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

Mean difference between CES-D post delivery by education level is 5.32

•With $p<0.001$; significant at the $p<0.05$ (and 0.001) level, with a two-tailed t-test
• The mean was significantly lower in the group who completed HS than in the group that did not

• Mean for HS graduates = 9.48
• Mean for non-graduates = 14.80