PUBH 501 Biostatistics

STATA: NON-PARAMETRIC TESTS, WILCOXON SIGNED-RANK, RANK SUM TEST

Tip of the day: using the / in Stata

- The single / is used as part of an operator, like in recode where we show Stata a range

```
recode var1 (1/5=1) (6/10=2), gen(var2)
```

- The /// denotes a comment when added to the beginning of a line

```
///the next line of code runs a tabulation of my variable 1
```

- The // adds a comment at the end of a line of code

```
tab var1 //run a tabulation of my var1
```

- The /// at the end of the line of a code inserts a line break. If you insert a new line without a line break (you hit enter or enter code on the next line without a "///"), Stata thinks it's a unique line of code, not connected to the one above it. Using the line break may help with readability.

```
graph hbar, ///
over(sat) ///
ytitle("Percent of Respondents") ///
title("In general, how satisfied are you with your job?")
```

Overview

- Sign test

- Wilcoxon sign-rank test

- Wilcoxon rank sum test

Checking assumptions in Stata

- For parametric tests, we are looking for normality
  - Overall
  - Within groups

- Can use –tab-, -summ-, and –histogram- to look at normality

```
bysort groups: summ var1, detail
```

Sign test

Sign test

- Tests that the median of the differences is zero (null hypothesis)

- Account for change in direction, but not a magnitude of change
  - Would need to use Wilcoxon signed-rank test to look at magnitude of change

- The Stata command is –signtest-

```
signtest cesdT = cesdTE
```

Sign test

```
         sign |  observed      expected
--------------+-----------------------------
     positive |         6             5
     negative |         4             5
         zero |         0             0
--------------+-----------------------------
          all |        10            10
```

Two-sided test:
  Ho: median of cesdT - cesdTE = 0 vs.
  Ha: median of cesdT - cesdTE != 0
      Pr(#positive >= 6 or #negative >= 6) =
        min(1, 2*Binomial(n = 10, x >= 6, p = 0.5)) = 0.7539

Conclusions

- We do not reject the null hypothesis the median of the differences between the groups is zero

Signed-rank test

Wilcoxon signed-rank test

- Tests the equality of matched pairs of observations. Account for magnitude of change

- Null hypothesis is that median difference in variables =0

- The Stata command is –signrank-

```
signrank cesdT = cesdTE
```

Results

```
Wilcoxon signed-rank test


      sign |       obs     sum ranks      expected          unadjusted variance        96.25
-------------+---------------------------------------        adjustment for ties        -0.38
                                                             adjustment for zeros        0.00
   positive |        6          36.5          27.5                                  ----------
   negative |        4          18.5          27.5           adjusted variance          95.88
       zero |        0             0             0
-------------+---------------------------------------        Ho: cesdT = cesdTE
        all |       10            55            55
                                                                           z =     0.919
                                                              Prob > |z| =    0.3580
                                                              Exact Prob =    0.4043
```

Conclusions

- We cannot reject the null hypothesis that the median difference between cesdT and cesdTE = 0

- The median value of cesdT was 28, with a range of 15-35, and the median value of cesdTE was 26.5 with a range of 12-35.

- The exact p-value is best used in sample size <200. In our case, we use the exact p-value
  - the approximated p-value is based on a normal approximation of a randomized distribution and is better in larger samples. The exact p-value is computationally intense, so it is not recommended in n>200, and cannot be used at all in n>2000

Rank sum test

Wilcoxon rank sum / Mann-Whitney U test

- Tests the equality of non-matched, unpaired observations.

- Null hypothesis is that median difference between groups =0

- Tied ranks are averaged

- The Stata command is –ranksum dependent_var, by(independent_var)-

```
ranksum painscore, by(groups)
```

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

      groups |        obs      rank sum      expected
-----------------+---------------------------------------------
        TENS |         11           190           132
  Pain killer |         12            86           144
-----------------+---------------------------------------------
    combined |         23           276           276

unadjusted variance         264.00
adjustment for ties          -3.52
                         -----------------------
adjusted variance           260.48


Ho: painscore(groups==TENS) = painscore(groups==Pain killer)
          z =     3.594
   Prob > |z| =   0.0003
   Exact Prob =   0.0001
```

Conclusions

- We reject the null hypothesis that the median difference between groups in pain score is 0, there was a significant difference in median pain score between groups

- The median value of painscore for the TENS (nerve stimulation) group was 16, with a range of 12-17, and the median value of painscore for the pain killer group was 6.5 with a range of 3-16.

- The exact p-value is best used in sample size <200. In our case, we use the exact p-value

Multiple graphs

Storing and recalling graphs

- `set autotabgraphs on //allows more than one graph to be open, by using tab in the graph displayer`

- Name graphs to recall them
  - `histogram painscore if groups==1, name(histG1) freq`
  - `histogram painscore, name(histB) by(groups) freq`

  - `graph display histG1`
  - `graph drop histB //need to drop the graph name before using it again this session`

STATA DATA CUSTOMS

File >> open

Stata File (*.dta)

File >> import

Excel spreadsheet (*.xls;*.xlsx)
Text data (delimited, *.csv, ...)
SPSS data (*.sav)
SAS data (*.sas7bdat)
JDBC data source
ODBC data source
Text data in fixed format