Unit 8

Stata for Analysis of One and Two$^+$ Samples

version 14

"Vive la difference!"

Statistical analysis often involves the fitting of sophisticated models (multiple predictor linear regression, logistic, survival, mixed models, etc). Among the limitations of these methods are: (1) it is difficult to appreciate the actual data; and (2) their validity rest on assumptions that may or may not hold.

Analyses of data should begin with simple approaches that are as close to the data and as ***"model-free"*** as possible. These have the advantage of being simple, relatively assumption free, and straightforward in their interpretation.

This unit describes the use of Stata for estimation and hypothesis tests of data in one, two and more than two samples.

Important! Be sure that you have already produced your data descriptions (See again, units 6 – ***Stata for Data Description*** and 7- ***Stata for Graphs***)!

Table of Contents

Learning Objectives

When you have finished this unit, you should be able to produce, using Stata:

- Confidence intervals and hypothesis tests for **one continuous variable** under the assumption of normality**;**
- A nonparametric hypothesis test for **one continuous or ordinal variable** in the small study setting where the assumption of normality is not appropriate;
- Confidence intervals and hypothesis tests for **one proportion** under the assumption of a binomial distribution;
- Confidence intervals and hypothesis for **paired continuous variables** under the assumption of normality;
- A nonparametric hypothesis test for **paired continuous or ordinal variables** in the small study setting where the assumption of normality is not appropriate;
- Confidence intervals and hypothesis tests for **two independent variables – continuous** under the assumption of normality;
- Confidence intervals and hypothesis tests for **two independent proportions** under the assumption of independent binomial distributions;
- A nonparametric hypothesis test for **two independent continuous or ordinal variables** in the small sample setting where the assumption of normality is not appropriate;
- A one way analysis of variance under the assumption of normality; ***and***
- A nonparametric hypothesis test for the comparison of three or more independent medians in the small sample setting where the assumption of normality is not appropriate.

How to follow along:

These notes utilize 3 data sets.

Download from the course website.

1. **sepsis.dta**

Access using the Stata command sysuse as indicated in the illustrations contained in these notes

2. **bpwide.dta**
3. **auto.dta**

Sample Session

Suggestion –follow along!
This sample session uses the data set sepsis.dta.

References to data set used:

Dupont WD Statistical Modeling for Biomedical Researchers, Second Edition. Cambridge University Press, 2008..

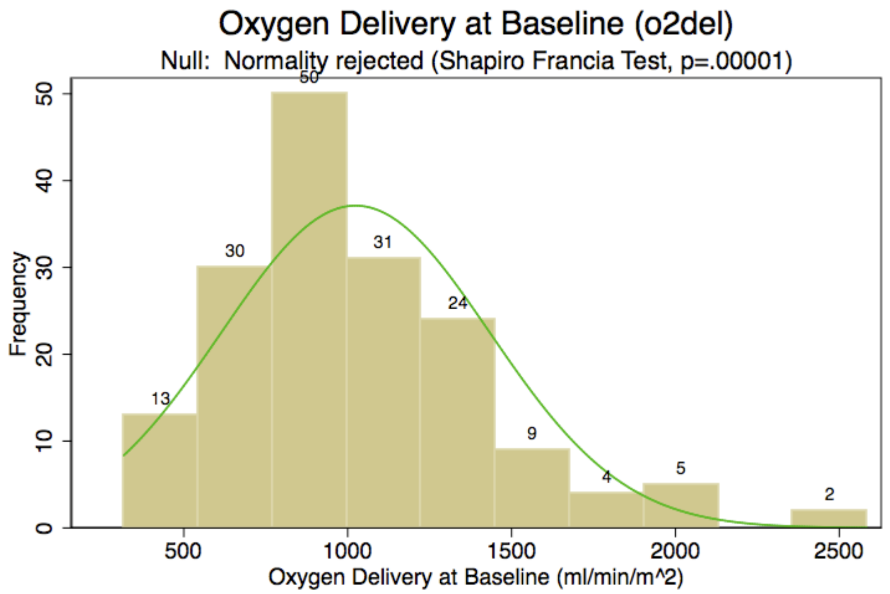
Benard GR, Wheeler AP et al (1997) The effects of ibuprofen on the physiology and survival of patients with sepsis. The Ibuprofen in Sepsis Study Group. NEJM 336: 912-8.

Sample session green-comments black-commands blue-results

```
. *-------------------------------------------------------------------
. *  PubHlth 691f - Data Management & Statistical Computing 2015
. *
. *   prog:       Carol Bigelow
. *   date:       November 8,2015
. *  input:       sepsis.dta
. * output:       none
. *  title:       Illustration of One and Two Plus Sample Inference
. *------------------------------------------------------------------
. * -----------  Preliminaries -------------------------------------
. cd "/Users/cbigelow/Desktop/"
  /Users/cbigelow/Desktop/
 . set more off
 . set scheme lean1

. *____________  Read in sepsis.dta ____________________________________*
. use "http://people.umass.edu/biep691f/data/sepsis.dta"
. keep temp0 temp7 treat fate apache o2del id
. codebook, compact

Variable   Obs Unique      Mean     Min      Max  Label
-------------------------------------------------------------------------------------------
id         455    455       228       1      455  Patient ID
treat      455      2  .4923077       0        1  Treatment
apache     454     38   15.3304       0       41  Baseline APACHE Score
o2del      168    168  1023.817  316.88  2584.34  Oxygen Delivery at Baseline (ml/min/m^2)
fate       455      2  .3868132       0        1  Mortal Status at 30 Days
temp0      455    122  100.4269   91.58      107  Baseline Temperature (deg. F)
temp7      413    105  99.19448    88.7   104.18  Temperature after 24 hours
-------------------------------------------------------------------------------------------
```

Sample session, continued: green-comments black-commands blue-results

```
. *---------  Command LABEL LIST to review discrete variable value labels ------*
. label list
race:
           0 White
           1 Black
           2 Other
fate:
           0 Alive
           1 Dead
treatmnt:
           0 Placebo
           1 Ibuprofen

. * __________________  ONE SAMPLE INFERENCE - Continuous Variable  ______________*


. *------------------ Oxygen Delivery at Baseline (o2del) -------------------*
. * Command TABSTAT with option STAT( ) to obtain descriptives*
. tabstat o2del, stat(n mean sd sem med min max) longstub

    variable |         N      mean        sd  se(mean)       p50       min       max
-------------+--------------------------------------------------------------------
       o2del |       168  1023.817  409.4426  31.58918     947.2    316.88   2584.34
----------------------------------------------------------------------------------

. * Command SFRANCIA for test of Assumption of Normality (null: normal)*
. sfrancia o2del

                  Shapiro-Francia W' test for normal data

    Variable |    Obs       W'          V'        z       Prob>z
-------------+--------------------------------------------------
       o2del |    168    0.93575      8.926     4.411    0.00001

. * Command HISTOGRAM with option NORMAL for Histogram with overlay Normal*
. histogram o2del, bin(10) frequency addlabels normal title("Oxygen Delivery at
Baseline (o2del)") subtitle("Null:  Normality rejected (Shapiro Francia Test,
p=.00001)") caption("histogram_o2del.png")
(bin=10, start=316.88, width=226.74601)
```



histogram_o2del.png

Sample session, continued: green-comments black-commands blue-results

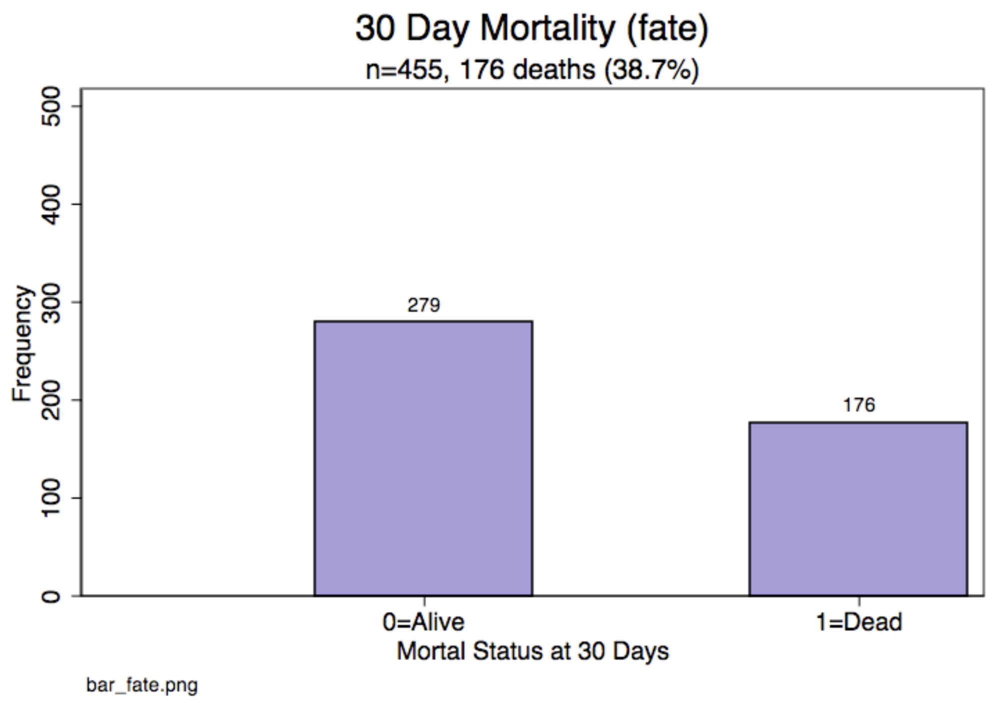
```
. * Command TTEST for One Sample t-test that mean = 950.   *
. ttest o2del=950

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
   o2del |     168    1023.817    31.58918    409.4426    961.4515    1086.183
------------------------------------------------------------------------------
    mean = mean(o2del)                                            t =   2.3368
Ho: mean = 950                                   degrees of freedom =      167

   Ha: mean < 950               Ha: mean != 950                 Ha: mean > 950
 Pr(T < t) = 0.9897         Pr(|T| > |t|) = 0.0206          Pr(T > t) = 0.0103

.*  Command CI with option LEVEL( ) for  confidence interval for mean.
. ci o2del, level(99)

    Variable |        Obs        Mean    Std. Err.       [99% Conf. Interval]
-------------+---------------------------------------------------------------
       o2del |        168    1023.817    31.58918        941.5086    1106.126
```

Sample session, continued: green-comments black-commands blue-results

```
. * __________________  ONE SAMPLE INFERENCE - Discrete Variable  ______________*
. *------------------ 30 Day Mortality (fate) -------------------*
 * Command TAB1 to obtain one way descriptives*
. tab1 fate
-> tabulation of fate

     Mortal |
  Status at |
    30 Days |      Freq.     Percent        Cum.
------------+-----------------------------------
      Alive |        279       61.32       61.32
       Dead |        176       38.68      100.00
------------+-----------------------------------
      Total |        455      100.00


. * Command FRE to obtain one way descriptives (my preferred approach – cb)
. fre fate
fate -- Mortal Status at 30 Days
-------------------------------------------------------------
                |      Freq.    Percent      Valid       Cum.
----------------+--------------------------------------------
Valid   0 Alive |        279      61.32      61.32      61.32
        1 Dead  |        176      38.68      38.68     100.00
        Total   |        455     100.00     100.00
-------------------------------------------------------------

.* Command CI for a confidence Interval estimate of the probability of death
.* normal approximation method
. ci fate, level(95)

    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
        fate |        455    .3868132     .022857        .3418946    .4317318

.* Command CI with option BINOMIAL for a confidence interval using exact binomial method
. ci fate, binomial level(95)

                                                         -- Binomial Exact --
    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
        fate |        455    .3868132    .0228319        .3418278    .4332801

.* Command PRTEST for One sample test of proportion = 0.30
.* normal approximation method
. prtest fate=.30, level(95)

One-sample test of proportion                   fate: Number of obs =      455
------------------------------------------------------------------------------
    Variable |       Mean   Std. Err.                     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        fate |   .3868132   .0228319                      .3420636    .4315628
------------------------------------------------------------------------------
    p = proportion(fate)                                          z =   4.0409
Ho: p = 0.3

     Ha: p < 0.3                 Ha: p != 0.3                 Ha: p > 0.3
 Pr(Z < z) = 1.0000         Pr(|Z| > |z|) = 0.0001          Pr(Z > z) = 0.0000
```

Sample session, continued: green-comments black-commands blue-results

```
. * Command BITEST for one sample test of proportion, exact binomial method
. bitest fate=.30

    Variable |        N   Observed k   Expected k   Assumed p   Observed p
-------------+------------------------------------------------------------
        fate |      455        176        136.5       0.30000      0.38681

  Pr(k >= 176)           = 0.000047  (one-sided test)
  Pr(k <= 176)           = 0.999969  (one-sided test)
  Pr(k <= 98 or k >= 176) = 0.000079  (two-sided test)

. * Command HISTOGRAM with option DISCRETE for Discrete Variable Bar Chart
. histogram fate, discrete frequency barwidth(.5) addlabels title("30 Day Mortality (fate)")
   subtitle("n=455, 176 deaths (38.7%)") xlabel(0 "0=Alive" 1 "1=Dead") ylabel(0 (100)500)
   fcolor(lavender) lcolor(black)  note("bar_fate.png")
(start=0, width=1)
```

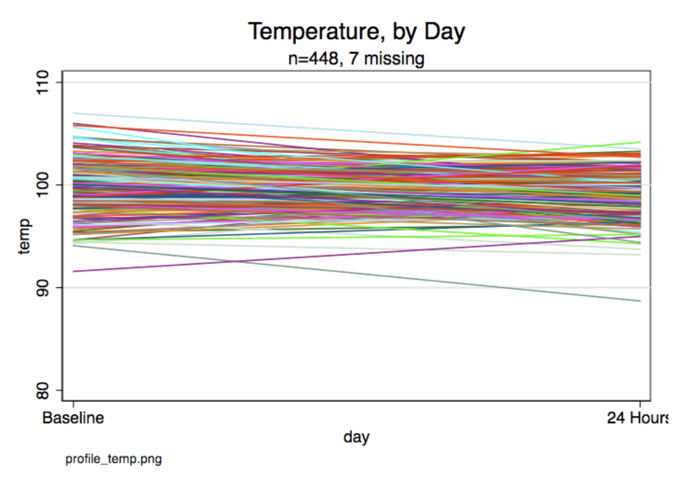


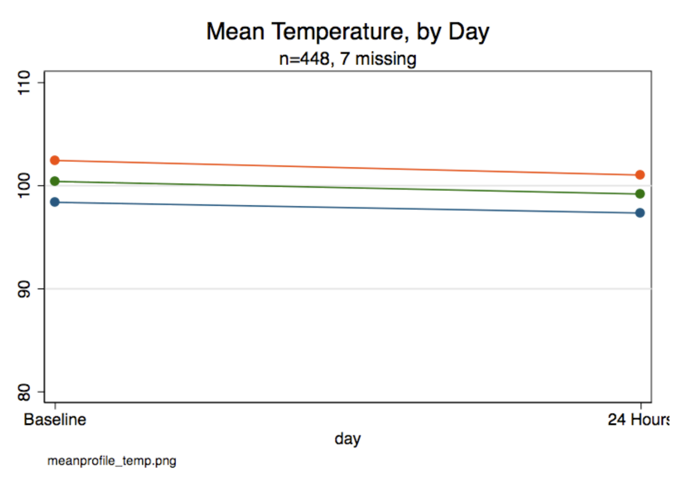
Sample session, continued: green-comments black-commands blue-results

```
. *______________ Paired Sample Inference __________________*
. *-------------- Repeated Measurement of temperature (temp0 and temp7) -------*
. generate chg_24hrs=temp0-temp7
(42 missing values generated)
. label variable chg_24hrs "Baseline - 24 Hour Change"
. tabstat temp0 temp7 chg_24hrs, col(stat) stat(n mean sd sem med min max) longstub

    variable |         N      mean        sd  se(mean)       p50       min       max
-------------+----------------------------------------------------------------------
       temp0 |       455  100.4269  2.026105  .0949853     100.7     91.58       107
       temp7 |       413  99.19448  1.842151  .0906463     99.14      88.7    104.18
   chg_24hrs |       413  1.285957  1.988315  .0978386  1.220001 -5.400002  8.299995
------------------------------------------------------------------------------------

. *------- Command TTEST for paired t test of equality of temp0 and temp 7 ------ *
. ttest temp0=temp7

Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
   temp0 |     413    100.4804    .0956495    1.943828    100.2924    100.6685
   temp7 |     413    99.19448    .0906463    1.842151    99.01629    99.37267
---------+--------------------------------------------------------------------
    diff |     413    1.285957    .0978386    1.988315    1.093632    1.478282
------------------------------------------------------------------------------
     mean(diff) = mean(temp0 - temp7)                             t =  13.1437
 Ho: mean(diff) = 0                              degrees of freedom =      412

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

Sample session, continued: green-comments **black-commands** blue-results

```
. *-------------- individual line plots -----------*
. *------- must re-shape data to long version prior to plot ----*
. * ---- IMPORTANT PRELMINARY:  Use preserve to set aside the original data for later use --*
. preserve
. reshape long temp, i(id) j(day)
(note: j = 0 7)

       ----- some output omitted ---

. *---------- individual profiles of change in temperature -----------*
. * SUGGESTION: Watch your screen after issuing the command below and see the lines appear!
. xtline temp, i(id) t(day) ylabel(80 (10)110, grid) overlay title("Temperature, by Day")
  subtitle("n=448, 7 missing") xlabel(0 "Baseline" 7 "24 Hours") legend(off)
  note("profile_temp.png")

. *---------- mean profile of change in temperature -----------*
. sort day
. collapse (mean) temp (sd) sdtemp=temp, by(day)
. generate high=temp + sdtemp
. generate low=temp - sdtemp

. graph twoway (connected temp day) (connected high day) (connected low day), ylabel(80
(10)110, grid) xlabel(0 "Baseline" 7 "24 Hours") legend(off) note("meanprofile_temp.png")
title("Mean Temperature, by Day") subtitle("n=448, 7 missing")
```



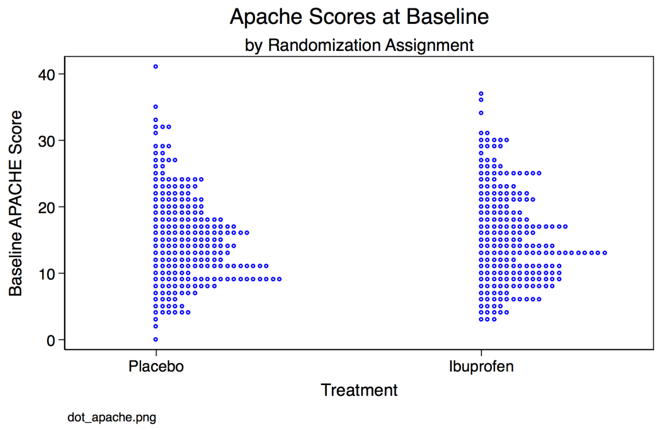


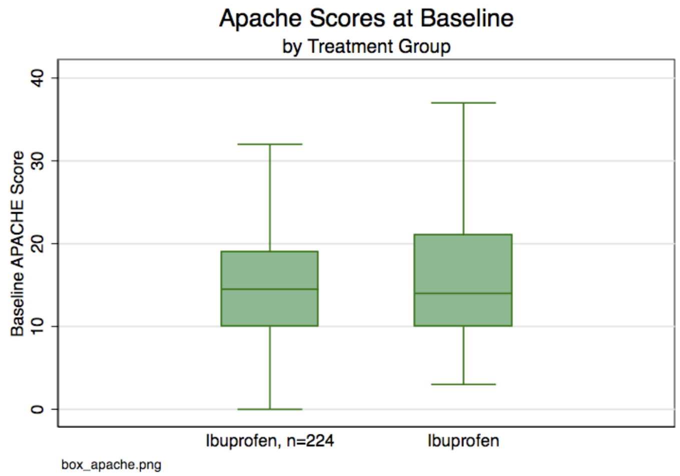
Sample session, continued: green-comments black-commands blue-results

```
. *________________  Two Independent Samples Inference _________________*
. * Continuous outcome (apache) in independent groups (treat)
. *---- IMPORTANT:  Issue the command restore to recover back the original data ------*
. clear
. restore
. tab1 treat
-> tabulation of treat

  Treatment |      Freq.     Percent        Cum.
------------+-----------------------------------
    Placebo |        231       50.77       50.77
  Ibuprofen |        224       49.23      100.00
------------+-----------------------------------
      Total |        455      100.00



.* Descriptives of outcome (apache) by group (treat)
. sort treat
. tabstat apache, by(treat) col(stat) stat(n mean sd sem med min max) longstub

treat         variable |         N      mean        sd  se(mean)       p50       min       max
-----------------------+----------------------------------------------------------------------
Placebo         apache |       230  15.18696  6.922831   .456478      14.5         0        41
Ibuprofen       apache |       224  15.47768  7.261882  .4852049        14         3        37
-----------------------+----------------------------------------------------------------------
Total           apache |       454   15.3304  7.085794  .3325528        14         0        41
----------------------------------------------------------------------------------------------




. *------ test of equality of variances -----*
. sort treat
. sdtest apache, by(treat)
Variance ratio test
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
 Placebo |     230    15.18696     .456478    6.922831    14.28752    16.08639
Ibuprofe |     224    15.47768    .4852049    7.261882    14.52151    16.43385
---------+--------------------------------------------------------------------
combined |     454     15.3304    .3325528    7.085794    14.67686    15.98393
------------------------------------------------------------------------------
    ratio = sd(Placebo) / sd(Ibuprofe)                            f =   0.9088
Ho: ratio = 1                                  degrees of freedom = 229, 223

    Ha: ratio < 1               Ha: ratio != 1                 Ha: ratio > 1
  Pr(F < f) = 0.2362         2*Pr(F < f) = 0.4724           Pr(F > f) = 0.7638
```

Sample session, continued: green-comments black-commands blue-results

```
. *------- Command TTEST with option BY( ) for 2 sample t test for independent groups -----*
. ttest apache, by(treat)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
 Placebo |     230    15.18696     .456478    6.922831    14.28752    16.08639
Ibuprofe |     224    15.47768    .4852049    7.261882    14.52151    16.43385
---------+--------------------------------------------------------------------
combined |     454     15.3304    .3325528    7.085794    14.67686    15.98393
---------+--------------------------------------------------------------------
    diff |            -.290722    .6657587               -1.599088    1.017644
------------------------------------------------------------------------------
    diff = mean(Placebo) - mean(Ibuprofe)                         t =  -0.4367
Ho: diff = 0                                     degrees of freedom =      452

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.3313         Pr(|T| > |t|) = 0.6626          Pr(T > t) = 0.6687

. *------- Command DOTPLOT with option OVER( ) for side by side dot plot ---------*
. dotplot apache, over(treat) nx(50) msymbol(oh) msize(small) mcolor(blue) title("Apache Scores
at Baseline") subtitle("by Randomization Assignment") note("dot_apache.png")

. *------- Command GRAPH BOX with option OVER( ) for by side box and whisker plot ---------*
. graph box apache, nooutsides over(treat,relabel(0 "Placebo, n=230" 1 "Ibuprofen, n=224"))
outergap(150) title("Apache Scores at Baseline") subtitle("by Treatment Group")
note("box_apache.png")
```





Introduction to "Immediate" Commands in Stata

Stata has a number of what are called "immediate commands".

The typical command in STATA instructs STATA to perform a calculation using a dataset opened by the user and stored in memory.

Example

. ci mpg, level(90)	This produces a 90% confidence interval estimate of the mean of the variable mpg using the data in memory

An "immediate" command instructs STATA to perform a calculation using numbers provided in the command.

Example

. cii 74 21.29 5.78, level(90) **n xbar s**	This produces a 90% confidence interval estimate of the mean of a variable for which we do not have the actual data but for which we do know that n = 74 xbar = 21.29 s = 5.78

Immediate Commands in Stata

They end in "i"
and you provide the numbers….

Examples: cii, ttesti, sdtesti, bitesti.

1. One Sample Inference

How to follow along;

```
. clear
. sysuse auto
```

1.1 Continuous Outcome: Mean of a Normal Distribution

Command	Example
Confidence Interval for Mean ci *variable,* level*(#)*	**.ci mpg, level(90)** This produces a 90% confidence interval estimate of the mean of the variable mpg
Confidence Interval for Mean, *"immediate"* cii *n xbar s,* level*(#)*	**.cii 74 21.2973 5.785503, level(90)** This produces a 90% confidence interval estimate of the mean of an UNNAMED variable for which n=74, xbar=21.2973 and the sample s=5.785503
t-test for Mean ttest *variable=nullmean,* level*(#)*	**.ttest mpg=20, level(90)** This produces a one sample t-test of the null hypothesis that the mean of mpg is μ = 20 for the . The output includes a 90% confidence interval
t-test for Mean, *"immediate"* ttesti *n xbar sigma nullmean,* level*(#)*	**.ttesti 74 21.2973 5.785503 20, level(90)** This produces a one sample t-test of the null hypothesis that the mean of an UNNAMED variable is μ = 20 in the setting where n=74, xbar=21.2973 and the sample s=5.785503

1.2 Continuous Outcome – Nonparametric Test: The Signed Rank Test

Command	Example
<u>**One Sample Signed Rank Test of Median**</u> **signrank *variable=nullmedian,*** **The option level( ) is NOT allowed**	**.signrank mpg=20** **This produces a one sample Wilcoxon Signed Rank test of the median of mpg is = 21**

1.3 Continuous Outcome – Variance of a Normal Distribution

Command	Example
<u>**One Sample Test of Variance**</u> **sdtest *variable=nullsigma,*** **NOTE! You supply the null standard deviation, NOT the null variance**	**.sdtest mpg=5** **This produces a one sample test of the null hypothesis that the variance of mpg is 5^2 = 25**
<u>**One Sample Test of Variance, *“immediate”***</u> **sdtesti *variable n . sigma nullsigma*** **Notes - (1) The period that you type is in place of the sample mean. You could supply this if you have it, but it is not necessary for the test of variance. (2) You specify the null standard deviation, NOT the null variance.**	**.sdtesti 74 . 5.78 6** **Note the period -** **Take care to provide the period in place of the sample mean. Otherwise you will get an uninterpretable error message!**

1.4 Discrete Outcome – Binomial Proportion

Command	Example
Exact Confidence Interval for Binomial π **ci *variable,* binomial level*(#)*** **This produces Clopper-Pearson "exact" confidence interval** **Confidence Interval for Binomial π, *"immediate"*** **cii *n observedproportion,* binomial level*(#)***	**.ci foreign, binomial level(90)** **This produces an exact 90% confidence interval estimate of the binomial parameter π for the variable foreign** **.cii 74 .2973, level(90)** **This produces an exact 90% confidence interval estimate of the binomial parameter π for an UNNAMED variable**
Exact test for Binomial π **bitest *variable=nullpi*** **The option level( ) is NOT allowed** **Exact test for Binomial π, *"immediate"*** **bitesti *n #successes nullpi*** **The option level( ) is NOT allowed**	**.bitest foreign=.28** **This produces an exact test of significance of the null hypothesis that the binomial parameter π = .28 for the variable foreign** **.bitesti 74 22 .28** **This produces an exact test of significance of the null hypothesis that the binomial parameter π = .28 for an UNNAMED variable in the setting where N=74, # successes = 22 and the null hypothesis that π = .28**
Normal Approximation test for Binomial π **prtest *variable=nullpi,* level*(#)*** **Normal Approximation test for Binomial π, *"immediate"*** **prtesti *n #successes nullpi,* count level*(#)***	**.prtest foreign=.28, level(95)** **This produces a normal aproximation test of significance of the null hypothesis that the binomial parameter π = .28 for the variable foreign. The output includes a 95% confidence interval estimate of π.** **.prtesti 74 22 .28, count level(95)** **This produces a normal approximation test of significance of the null hypothesis that the binomial parameter π = .28 for an UNNAMED variable in the setting where N=74, # successes = 22 and the null hypothesis that π = .28. The output includes a 95% confidence interval estimate of π.**

1.5 Continuous Outcome – Tests of Assumption of Normality

Review - Many statistical methods (especially linear regression) assume that the distribution of a variable (for example the dependent or Y-variable) is normal. Thus, it is useful to test this assumption. Stata offers two tests of normality: **Shapiro-Wilks** and **Shapiro-Francia**. Each is a test of the null hypothesis that the data are distributed normal.

What to look for -

	Data are Normal	Data are NOT Normal
Null hypothesis ("normality")	NOT rejected	rejected
p-value*	large	small

* **Note – In Stata the p-value appears the value listed under "Prob > z"**

Violations of the assumption of normality, if modest, are sometimes not a serious problem:

- Estimation and hypothesis tests of regression parameters are fairly robust to modest violations of normality;
- ***When to worry***: Predictions are sensitive to violations of normality
- ***Beware***: Sometimes the cure for violations of normality is worse than the problem.

Command	Example
Shapiro-Wilk Test **swilk *variable***	**.swilk mpg** **The null hypothesis is normality. Thus, the assumption of normality is reasonable when the test returns a p-value that is NOT statistically significant.**
Shapiro-Francia Test **sfrancia *variable***	**.sfrancia mpg** **The null hypothesis is again normality. Thus, the assumption of normality is reasonable when the test returns a p-value that is NOT statistically significant.**
Skewness-Kurtosis Test **sktest *variable***	**.sktest mpg** **The null hypothesis is again normality. Thus, the assumption of normality is reasonable when the test returns a p-value that is NOT statistically significant.**

2. Paired Sample Inference

How to follow along:
. clear
. sysuse bpwide

2.1 Continuous Outcome – Paired Means Under Normality

Command	Example
Paired t-test for Mean **ttest** *var1==var2*, **level***(#)* **Tip –** Note the requirement of TWO equal signs.	**.ttest bp_before==bp_after, level(99)** **This produces a paired t-test of the null hypothesis that the mean of bp_before equals the mean of bp_after. The output includes three 99% confidence intervals: (1) for bp_before (2) bp_after (3) difference**

2.2 Nonparametric Tests of Paired Medians

Tip – Two tests are provided here.

(1) **signrank** - Use for paired outcomes measured on an ordinal scale.
(2) **signtest** - Use for paired outcomes measured on a nominal scale.

Command	Example
Ordinal data … **Paired Data Wilcoxon Signed Rank Test of Equal Medians** **signrank** *var1=var2* **The option level( ) is NOT allowed**	**.signrank bp_before=bp_after** **This produces a paired data Wilcoxon Signed Rank test of equality of medians**
Nominal data … **Paired Data Sign Test of Equal Medians** **signtest** *var1=var2* **The option level( ) is NOT allowed**	

2.3 Continuous Outcome – Paired Variances Under Normality

Command	**Example**
Paired Data Test of Equal Variances **sdtest** ***var1=var2*** **NOTE –This will produce an Unpaired comparison of the variances using Levene's test, thus disregarding the paired-ness of the data. Stata does have a test of equality of variances for paired data. The command is sdpair and must be installed from the internet**	**.sdtest bp_before=bp_after** **This tests the equality of variances of bp_before and bp_after, as if the data were UNpaired**

3. Two Independent Samples Inference

How to follow along.

```
. clear
. sysuse auto
```

3.1 Continuous Outcome – Comparison of Two Normal Means

Command	Example
Assuming Equal Variances … **2 Sample t-test for Equality of Means** **sort *groupvariable*** **ttest *variable,* by(*groupvariable*) level*(#)***	**.sort foreign** **. ttest mpg, by(foreign) level(99)** **This produces a two sample t-test of the equality of means of the variable mpg, across the two groups of the variable foreign. The output includes a 99% confidence interval. Variances are assumed equal.**
Assuming UNequal Variances … **2 Sample t-test for Equality of Means** **sort *groupvariable*** **ttest *variable,* by(*groupvariable*) unequal level*(#)***	**.sort foreign** **. ttest mpg, by(foreign) unequal level(99)** **This produces a two sample t-test of the equality of means of the variable mpg, across the two groups of the variable foreign. The output includes a 99% confidence interval. Variances are assumed UNequal.**

3.2 Nonparametric Test of Two Independent Medians: Rank Sum Test

Tip - The nonparametric test of equality of two independent medians goes by multiple names. All are referring to the same thing:

- Mann Whitney
- Wilcoxon Rank Sum
- Rank Sum

The Stata command to use is the same one: **ranksum**

Command	**Example**
2 Sample Rank Sum Test for Equality of Medians **sort *groupvariable*** **ranksum *variable,* by(*groupvariable*)** **The option level( ) is NOT allowed**	**.sort foreign** **. ranksum mpg, by(foreign)** **This produces a Wilcoxon Rank Sum test of the equality of medians of the variable mpg, across the two groups of the variable foreign.**

3.3 Continuous Outcome: Comparison of Two Independent Variances

Command	**Example**
2 Sample Test for Equality of Variances **sdtest *variable,* by(*groupvariable*)** **The option level( ) is NOT allowed**	**. sdtest mpg, by(foreign)** **This produces a test of the equality of variances of the variable mpg, across the two groups of the variable foreign.**
2 Sample Test for Equality of Variance, *immediate* **sdtesti *n1 . sigma1 n2 . sigma2*** **The option level( ) is NOT allowed**	**. sdtesti 75 . 6.5 65 . 7.5** **Again, take care to provide two periods, this time as placeholders for the two sample mean values.**

3.4 Discrete Outcome: Comparison of Two Binomial Proportions

Recall - The normal approximation two sample test of equality of independent proportions and the chi square test of association in a 2x2 table are equivalent.

(a) Two Sample Normal Approximation Test of Equality of Independent Proportions

Command	Example
Normal Approximation Test for Equality of Two Indpendent Binomial π **sort** ***groupingvar*** **prtest** ***0/1variable,*** **by(*groupingvar*) level*(#)***	**.sort sex** **. prtest cure, by(sex) level(95)** **This produces a normal approximation test of significance of the null hypothesis equality of probability of cure in the two groups defined by sex. The output includes a 95% confidence interval estimate of the difference in the two binomial proportions π.**
"immediate with n's and observed proportions" **Normal Approximation test for Binomial π,** **prtesti** ***n1 proportion1 n2 proportion2***	**.prtesti 30 .4 45 .67** **In the 1st group: n = 30 % event = .40** **In the 2nd group: n=45 % event = .67**
"immediate with all counts" **Normal Approximation test for Binomial π,** **prtesti** ***n1 eventcount1 n2 eventcount2,*** **count**	**.prtesti 30 12 45 30, count**

(b) Two Sample Chi Square Test of Association for a 2x2 Table

Command	Example
Chi Square Test of Zero Association **tabulate *rowvar colvar,* chi2** ***OR*** **tab *rowvar colvar,* chi2** **Chi Square Test, *immediate*** **tabi *#11 #12* ..*#21 #22…* , chi2**	**. tab drug died, chi2** **. tabi 1 19\\8 6\\8 6, chi nolog**
All possible Two Way Tests of Zero Association **tab2 *var1 var2 var3* …., exact** ***OR*** **tab2 *var1 var2 var3* …., chi2** **Use the command tab2 to obtain tests of associations for all pairwise combinations of discrete variables.**	

3.5 Fisher's Exact Test of Association for a 2x2 Table

Command	Example
Fisher's Exact Test of Zero Association **tabulate *rowvar colvar,* exact nolog** ***OR*** **tab *rowvar colvar,* exact nolog**	**. tab drug died, exact nolog** **Tip! The option nolog suppresses the printing of the enumeration log for Fisher's exact test.**
Fisher's Exact Test as an "immediate" command **tabi *#11 #12* ..*#21 #22* ..., exact**	**. tabi 1 19\\8 6\\8 6, exact nolog**

4. K Independent Samples Inference

How to follow along.

```
. clear
. sysuse auto
```

4.1 Continuous Outcome: One Way Analysis of Variance

Command	Example
K Sample One Way Anova for Equality of Means **sort *groupvariable*** **oneway *variable groupvariable,* tabulate level*(#)***	**.sort foreign** **. oneway mpg foreign, level(99)** **This produces a one way anova of the equality of the means of the variable mpg, across the k=2 groups of the variable foreign. Tip! The option tabulate produces some nice descriptive statistics The output includes a 99% confidence interval**

4.2 Nonparametric Test of K Medians – The Kruskal Wallis Test

Command	Example
K Sample Kruskal Wallis Test for Equality of Medians **sort *groupvariable*** **kwallis *variable,* by(*groupvariable*)** **The option level( ) is NOT allowed**	**.** **.sort foreign** **. kwallis mpg, by(foreign)** **This produces a kruskal wallis test of the equality of medians of the variable mpg, across the K groups of the variable foreign. When the number of groups K=2, the results are identical to those obtained with the ranksum command.**