# Using Stata for Confidence Intervals

All of the confidence interval problems we have discussed so far can be solved in Stata via either (a) statistical calculator functions, where you provide Stata with the necessary summary statistics for means, standard deviations, and sample sizes; these commands end with an i, where the i stands for "immediate" (but other commands also sometimes end with an i) (b) modules that directly analyze raw data; or (c) both. Some of these solutions require, or may be easier to solve, if you first add the Stataquest menus and commands; see

http://www.stata.com/support/faqs/res/quest7.html

The commands shown below can all be generated via Stata's pulldown menus if you prefer to use them. Also, as we will see, several other Stata commands produce confidence intervals as part of their output.

---

## A. Confidence Intervals Case I. Population normal, σ known.

Problem. $\overline{X}$ = 24.3, σ = 6, n = 16, X is distributed normally. Find the 90% confidence interval for the population mean, E(X).

Solution. I don't know of any Stata routine that will do this by directly analyzing raw data. Further, unlike the other 2 cases, I don't know of a standalone command for confidence intervals only. However, the `ztesti` command (which is installed with Stataquest) will do this when you have the summary statistics. Enter

```
ztesti 16 24.3 6 0, level(90)
```

where the 5 parameters are the sample size N, the sample mean, the known population standard deviation, the hypothesized mean under $H_0$ (for now you can just fill in any arbitrary value, the CI will not be affected), and the desired CI level (e.g. 95 for 95% confidence interval, 90 for 90% c.i.) The Stata results (which match up perfectly with our earlier analysis) are

```
. ztesti 16 24.3 6 0, level(90)

                                              Number of obs =       16

------------------------------------------------------------------------------
Variable |      Mean    Std. Err.       z     P>|z|       [90% Conf. Interval]
---------+--------------------------------------------------------------------
       x |      24.3         1.5      16.2    0.0000       21.83272    26.76728
------------------------------------------------------------------------------
```

`ztesti` also produces other output that will become meaningful when we talk about hypothesis testing.

## B. Confidence Intervals Case II. Binomial parameter p.

Problem. N = 100, $\hat{p}$ = .40. Construct a 95% c.i.

Solution. Use the `ci` or `cii` command. If you just have the summary statistics,

**cii 100 40, level(95) wilson**

The parameters are the sample size N, the # of successes, the desired confidence interval, and the formula to use (in this case we are telling it to use the Wilson confidence interval which we computed before.) The results, which are identical to what we got earlier, are

```
. cii 100 40, level(95) wilson

                                                         ------ Wilson ------
    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
             |        100          .4    .0489898        .3094013    .4979974
```

If you don't specify a formula, Stata will default to `exact`, i.e. Clopper-Pearson:

```
. cii 100 40, level(95)

                                                         -- Binomial Exact --
    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
             |        100          .4    .0489898        .3032948    .5027908
```

Other formula options are `agresti` and `jeffries`.

If you do have raw data, then the binomial variable should be coded 0/1. In this case, there would be 60 cases coded 0 and 40 cases coded 1. If "x" is the binomial variable, then the command is

**ci x, binomial wilson level(95)**

The `binomial` parameter is necessary so that `ci` knows the variable is binomial. The output is

```
. ci x, binomial wilson level(95)

                                                         ------ Wilson ------
    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
           x |        100          .4    .0489898        .3094013    .4979974
```

To use the default Clopper-Pearson (aka `exact`) formula,

```
. ci x, binomial level(95)

                                                         -- Binomial Exact --
    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
           x |        100          .4    .0489898        .3032948    .5027908
```

Approximate c.i.: If for some reason you want it (e.g. to double-check your hand calculations) you can also get the approximate confidence interval by using Stata's `bintesti` command that is installed with Stataquest. The command is

```
bintesti 100 40 .5, normal level(95)
```

The first parameter is the sample size N, the second is the number of observed successes, the third is the hypothesized value of p under $H_0$ (for now just use any value between 0 and 1), the `normal` parameter tells it to use a normal approximation, and `level` is the desired c.i.

```
. bintesti 100 40 .5, normal level(95)

  Variable |      Obs  Proportion   Std. Error
  ---------+--------------------------------
         x |      100          .4    .0489898

           Ho:     p = 
                   z = -2.00
            Pr > |z| = 0.0455
              95% CI = (0.3040,0.4960)
```

## C. Confidence Intervals Case III. Population normal, σ unknown.

Problem. X is distributed normally, n = 9, $\overline{X}$ = 4, $s^2$ = 9. Construct a 99% c.i. for the mean of the parent population.

Solution. Again we can use the `ci` and `cii` commands, with slightly different formats. If you are working with the summary statistics, the command is

```
cii 9 4 3, level(99)
```

where the parameters are the sample size N, the sample mean, the sample standard deviation (note that the variance, not the sd, is reported in the problem), and the desired confidence interval. The results, which are identical to what we calculated before, are

```
. cii 9 4 3, level(99)

    Variable |        Obs        Mean    Std. Err.       [99% Conf. Interval]
-------------+---------------------------------------------------------------
             |          9           4           1        .6446127    7.355387
```

If we had raw data and our variable was called "x", the command would simply be

```
ci x, level(99)
```

and the output would be

```
. ci x, level(99)

    Variable |        Obs        Mean    Std. Err.       [99% Conf. Interval]
-------------+---------------------------------------------------------------
           x |          9           4           1        .6446127    7.355387
```