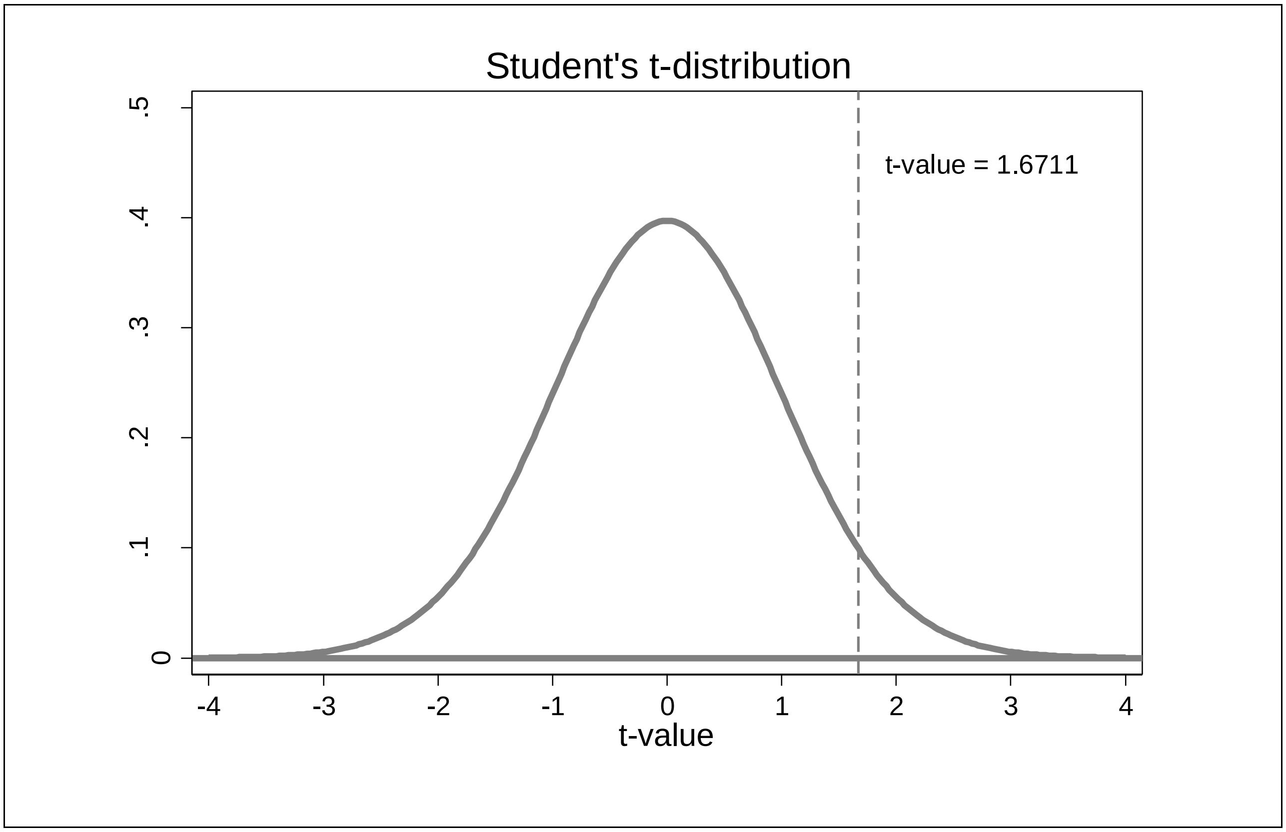
Stata Program Notes
Biostatistics: A Guide to Design, Analysis, and Discovery Second Edition
by Ronald N. Forthofer, Eun Sul Lee, Mike Hernandez
Chapter 7: Interval Estimation

Program Note 7.1 – Finding percentiles of a t-distribution

The Stata command **invttail(n, p)** returns percentiles from the t-distribution with **n** degrees of freedom where **p** is the area under the right-tail of the t-distribution's probability density function. For example, the 95th percentile of a t-distribution with 59 degrees of freedom can be found with the following commands.

```
Stata command:

invttail(59, 0.05)


Stata output:

1.6711

*-------------------------------------------------------------------------*

Stata command:

graph twoway (function y=tden(59,x), lcol(gray) range(-4 4) lw(thick)),   ///
  title("Student's t-distribution")                                       ///
  xtitle("t-value", size(medlarge))                                       ///
  ytitle("")                                                              ///
  xlabel(-4(1)4)                                                          ///
  ylab(0(0.1)0.5)                                                         ///
  xline(1.6711, lcolor(gray) lpattern(dash))                              ///
  yline(0, lcolor(gray) lwidth(thick))                                    ///
  scheme(s1color)                                                         ///
  text(0.45 2.75 "t-value = 1.6711")


Stata output:
```



Student's t-distribution
t-value = 1.6711
0
.1
.2
.3
.4
.5
-4
-3
-2
-1
0
1
2
3
4
t-value

Program Note 7.2 – Binomial confidence intervals

The **ci** command in Stata returns confidence intervals for means, proportions, and counts. The **level( )** option allows you to set the level of confidence for the confidence interval. For example, to find a 90 percent confidence interval for the proportion derived from twenty observations with four successes, we can use the Stata command **cii**. This command allows for immediate entry of the number of trials (20) and the number of successes (4): **cii 20 4**. The corresponding 90% confidence interval limits are (0.0714, 0.4010).

```
Stata command:

cii 20 4, level(90)


Stata output:

                                                         -- Binomial Exact --
    Variable |        Obs        Mean    Std. Err.       [90% Conf. Interval]
-------------+---------------------------------------------------------------
             |         20          .2    .0894427        .0713539    .4010281
```

Note that Stata has other options that can be used to derive confidence intervals:

```
exact        calculates  the exact confidence intervals by default
wald         calculates the Wald confidence intervals
wilson       calculates the Wilson confidence intervals
agresti      calculates the Agresti-Coull confidence intervals
jeffreys     calculates the Jeffreys confidence intervals
```

As another example, we illustrate the output that each of these options provide. Here, we consider 20 trials with 2 successes. An estimate of the success proportion is 0.1 or 10%, and the 95% confidence intervals are provided using each of the options listed above. One should notice that each option provides slightly different 95% confidence interval limits.

```
Stata command:
*Exact Option
cii 20 2, level(95)     exact

Stata output:

                        -- Binomial Exact --
Variable    Obs    Mean     Std. Err. [95% Conf. Interval]

            20     .1      .067082    .0123485    .3169827
*---------------------------------------------------------




Stata command:
*Wald Option
cii 20 2, level(95)     wald

Stata output:
```

```
                  -- Binomial Wald ---
Variable    Obs   Mean    Std. Err. [95% Conf. Interval]

            20    .1     .067082    0    .2314784

The Wald interval was   clipped at  the lower endpoint.
*--------------------------------------------------------
```

```
Stata command:
*Wilson Option
cii 20 2, level(95)     wilson

Stata output:

                  ------ Wilson ------
Variable    Obs   Mean    Std. Err. [95% Conf. Interval]

            20    .1     .067082    .0278665    .3010336
*--------------------------------------------------------
```

```
Stata command:
*Agresti Option
cii 20 2, level(95)     agresti

Stata output:

                  -- Agresti-Coull ---
Variable    Obs   Mean    Std. Err. [95% Conf. Interval]

            20    .1     .067082    .0156562    .3132439
*--------------------------------------------------------
```

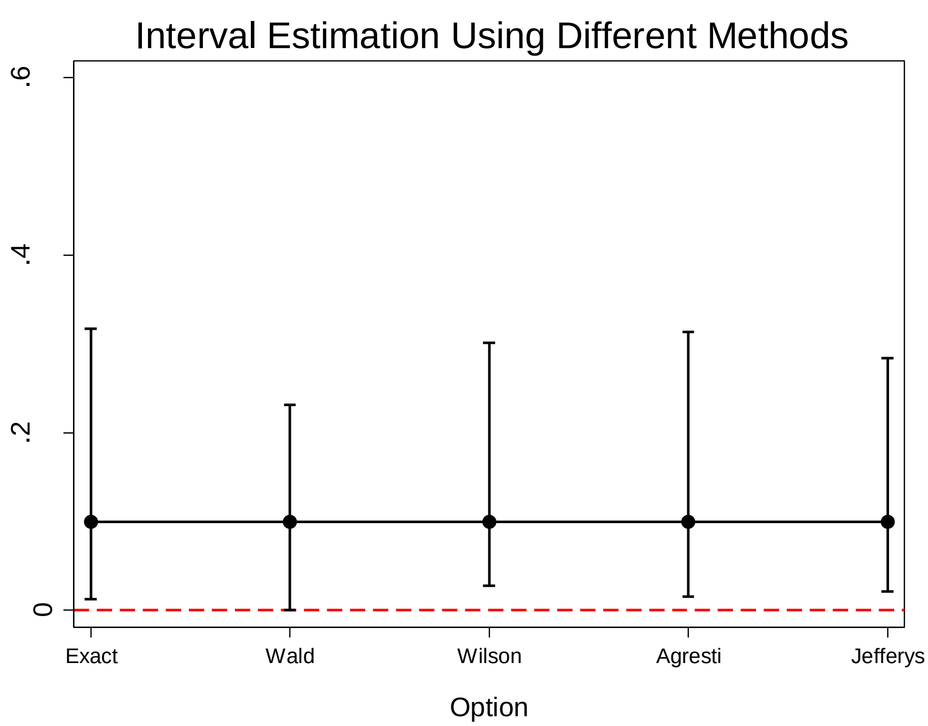
```
Stata command:
*Jeffreys Option
cii 20 2, level(95)     jeffreys

Stata output:

                  ----- Jeffreys -----
Variable    Obs   Mean    Std. Err. [95% Conf. Interval]

            20    .1     .067082    .0213725    .2838533
```

There are several ways of visualizing the results provided from each option. One method of visualizing this information is provided by the use of a scatter plot with standard deviation bars.

Stata command:

```
clear
input group  Option  prop   lower95    upper95
      1      1       0.1    0.0123485  0.3169827
      1      2       0.1    0.0000000  0.2314784
      1      3       0.1    0.0278665  0.3010336
      1      4       0.1    0.0156562  0.3132439
      1      5       0.1    0.0213725  0.2838533
end

* The upper and lower limits of the 95%CI

label define optdef  1 "Exact"   2 "Wald"   3 "Wilson"  4 "Agresti" 5 "Jefferys"
label values Option optdef

twoway (rcap upper95 lower95 Option, lcolor(black))                             ///
       (connected prop Option, mcolor(black) lcolor(black)  lpattern(solid)),  ///
       scheme(s1color)                                                          ///
       xtitle(" ")                                                              ///
       ytitle("{bf: Estimated Proportion}" "95% Confidence Interval")           ///
       ylab(0(0.2)0.6)                                                          ///
       xlab(1(1)5, valuelabel labsize(small))                                   ///
       legend(off)                                                              ///
       yscale(titlegap(4))                                                      ///
       xscale(titlegap(4))                                                      ///
       graphregion(margin(r+5))  yline(0, lpattern(dash) lcolor(red))
```

Stata output:

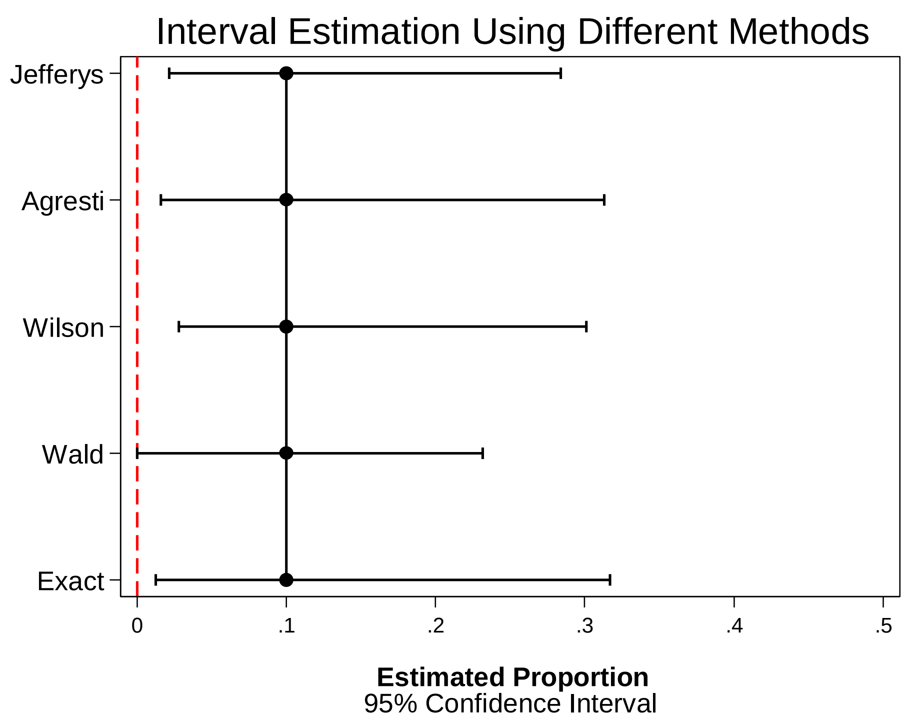


Another approach for visualizing the 95% CIs would be to present them horizontally.

Stata command:

```
twoway (rcap upper95 lower95 Option, lcolor(black))                          ///
       (connected prop Option, mcolor(black) lcolor(black)  lpattern(solid)), ///
       scheme(s1color)                                                        ///
       xtitle(" ")                                                            ///
       ytitle("{bf: Estimated Proportion}" "95% Confidence Interval")         ///
       ylab(0(0.2)0.6)                                                        ///
       xlab(1(1)5, valuelabel labsize(small))                                 ///
       legend(off)                                                            ///
       yscale(titlegap(4))                                                    ///
       xscale(titlegap(4))                                                    ///
       graphregion(margin(r+5))  yline(0, lpattern(dash) lcolor(red))
```

Stata output:



Some References:

Clopper CJ & Pearson ES (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, **26**, 404–413.

Wald A & Wolfowitz J (1939). Confidence limits for continuous distribution functions. *The Annals of Mathematical Statistics*, **10**, 105–118.

Wilson EB (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, **22**, 209–212.

Agresti A & Caffo B (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *The American Statistician*, **54**, 280–288.

Blyth CR & Still HA (1983). Binomial confidence intervals. *Journal of the American Statistical Association*, **78**, 108–116.

Program Note 7.3 – Confidence intervals using the t-distribution

Although hypothesis testing is not introduced until Chapter 8, the **ttest** command in Stata calculates confidence intervals by default. In Example 7.7, the mean age between AML and ALL patients can be compared using the Stata commands in the example below. From the output shown below, a 99 percent confidence interval for the difference between mean ages is (1.41, 25.02).

```
Stata command:

ttest age, by(dx_type) level(99)


Stata output:

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [99% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |      51    49.86275     2.31195    16.51063    43.67182    56.05367
       1 |      20       36.65     3.99096    17.84812    25.23212    48.06788
---------+--------------------------------------------------------------------
combined |      71    46.14085    2.113033    17.80473    40.54574    51.73595
---------+--------------------------------------------------------------------
    diff |            13.21275    4.456005                1.408892     25.0166
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   2.9652
Ho: diff = 0                                     degrees of freedom =       69

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9979         Pr(|T| > |t|) = 0.0042          Pr(T > t) = 0.0021
```

Program Note 7.4 – Confidence intervals for the Pearson correlation coefficient

By using the **display** command, Stata can be used as a calculator. As an example, we can calculate the lower and upper limits for the correlation coefficient. The limits can be found using the Stata commands in the example below.

```
Stata commands:

display (exp(2*(0.7882)) - 1)/(exp(2*(0.7882))+1)

* which provides 0.657, and the upper limit can be found using


display (exp(2*(2.0948)) - 1)/(exp(2*(2.0948))+1)

* which provides 0.970
```