

Soci708-001 Statistics for Sociologists

Fall 2009

Professor François Nielsen

Stata Commands for Module 8 – Inference for Proportions

For further information on any command in this handout, simply type `help` followed by the name of the command in Stata.

For confidence intervals, also see page 35 of the Stata and SAS Guide pdf (click on Documents in side bar; guide is linked under Software Documentation).

1 Statistical Functions in Stata

1.1 Normal Distribution Functions

The function `normal(z)` returns $P(Z \leq z)$, the area under the standard normal curve to the left of z . (Compare with Table A.)

```
. display normal(1.207)
.88628393
```

The function `invnormal(p)` returns z such that $P(Z \leq z) = p$, i.e. such that the area under the standard normal curve to the left of z is p . (Compare with Table A and Table D (bottom row).)

```
. display invnormal(0.975)
1.959964
```

2 One Sample t-Test for Proportion in Stata

When inputting the data in summary form, i.e. we don't have the individual data, the syntax is either `prtesti n p p0` where Stata expects that p is a proportion, or `prtesti n X p0, count` where Stata expects X is an integer (number of successes).

```
. * Binge Drinking example p. 490
. * Note hypothesis p0 = 0.5 makes no sense here; we still get correct CI
. prtesti 13819 3140 0.5, level(95) count
```

```
One-sample test of proportion                                x: Number of obs =   13819
-----+-----
Variable |          Mean   Std. Err.          [95% Conf. Interval]
-----+-----
      x |   .2272234   .0035646   .2202368   .2342099
-----+-----
p = proportion(x)                                         z = -64.1321
```



```
-----
      diff = prop(x) - prop(y)                                z = -2.4701
Ho: diff = 0

      Ha: diff < 0                Ha: diff != 0                Ha: diff > 0
Pr(Z < z) = 0.0068                Pr(|Z| < |z|) = 0.0135                Pr(Z > z) = 0.9932
```

Now the same example but using the “plus four” technique. (Add +2 to each n and +1 to each X .)

```
. prtesti 2053 57 2032 85, count level(90)
```

```
Two-sample test of proportion                                x: Number of obs =    2053
                                                            y: Number of obs =    2032
-----
Variable |      Mean   Std. Err.      z    P>|z|    [90% Conf. Interval]
-----+-----
      x |   .0277642   .0036261                .0217999   .0337286
      y |   .0418307   .0044413                .0345255   .0491359
-----+-----
      diff |  -.0140665   .0057335                -.0234972  -.0046357
          | under Ho:      .005732    -2.45    0.014
-----
      diff = prop(x) - prop(y)                                z = -2.4540
Ho: diff = 0

      Ha: diff < 0                Ha: diff != 0                Ha: diff > 0
Pr(Z < z) = 0.0071                Pr(|Z| < |z|) = 0.0141                Pr(Z > z) = 0.9929
```

When we have the individual data the syntax is `prtest var, by(groupvar)` as in the following example.

```
. * now use survey2b data
. use "C:\Users\nielsen\Documents\soci708\data\survey2b.dta", clear
. prtest drinks, by(female)
```

```
Two-sample test of proportion                                0: Number of obs =    104
                                                            1: Number of obs =    152
-----
Variable |      Mean   Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
      0 |   .8461538   .0353795                .7768114   .9154963
      1 |   .7631579   .0344838                .6955709   .8307449
-----+-----
      diff |   .082996   .0494048                -.0138358   .1798277
          | under Ho:      .0511986     1.62    0.105
-----
      diff = prop(0) - prop(1)                                z = 1.6211
Ho: diff = 0

      Ha: diff < 0                Ha: diff != 0                Ha: diff > 0
Pr(Z < z) = 0.9475                Pr(|Z| < |z|) = 0.1050                Pr(Z > z) = 0.0525
```