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.)

```stata
. 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).)

```stata
. 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).

```stata
. * 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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>.2272234</td>
<td>.0035646</td>
<td>.2202368 .2342099</td>
</tr>
</tbody>
</table>

$p = \text{proportion}(x)$

$z = -64.1321$
Ho: p = 0.5

Ha: p < 0.5

Ha: p != 0.5

Ha: p > 0.5

Pr(Z < z) = 0.0000

Pr(|Z| > |z|) = 0.0000

Pr(Z > z) = 1.0000

* Work Stress example p. 494

prtesti 100 68 0.75, level(95) count

One-sample test of proportion

x: Number of obs = 100

Variable | Mean Std. Err. [95% Conf. Interval]
-----------|--------+----------------------------------
       x | .68 .0466476 .5885724 .7714276
-----------+----------------------------------

p = proportion(x)

z = -1.6166

Ho: p = 0.75

Ha: p < 0.75

Ha: p != 0.75

Ha: p > 0.75

Pr(Z < z) = 0.0530

Pr(|Z| > |z|) = 0.1060

Pr(Z > z) = 0.8470

* Gallup Poll 22 Oct 2008, n=2788, 51% favor Obama (43% McCain, 6% Undecided)

prtesti 2788 0.51 0.5, level(95)

One-sample test of proportion

x: Number of obs = 2788

Variable | Mean Std. Err. [95% Conf. Interval]
-----------|--------+----------------------------------
       x | .51 .0094675 .491444 .528556
-----------+----------------------------------

p = proportion(x)

z = 1.0560

Ho: p = 0.5

Ha: p < 0.5

Ha: p != 0.5

Ha: p > 0.5

Pr(Z < z) = 0.8545

Pr(|Z| > |z|) = 0.2910

Pr(Z > z) = 0.1455

3 Two Sample Test for Proportions in Stata

The syntax for a two sample test, when inputting the data in summary form, is either prtesti n1 p1 n2 p2 where Stata expects that p1 and p2 are proportions, or prtesti n1 X1 n2 X2 where Stata expects that X1 and X2 are integer.

* sample of middle-ages men taking experimental drug or placebo

* n1=2051 taking drug, n2=2030 placebo

* X1=56, X2=84 are numbers with heart attack for drug and placebo, resp.

prtesti 2051 56 2030 84, count level(90)

Two-sample test of proportion

x: Number of obs = 2051

y: Number of obs = 2030

Variable | Mean Std. Err. z P>|z| [90% Conf. Interval]
-----------|--------+----------------------------------
       x | .0273038 .0035985 .0213848 .0332227
       y | .0413793 .0044205 .0341083 .0486503
-----------+----------------------------------

diff | -.0140756 .0056988 -.0234511 -.0047

| under Ho: .0056984 -2.47 0.014
diff = prop(x) - prop(y)  
Ho: diff = 0

Ha: diff < 0  
Pr(Z < z) = 0.0068

Ha: diff != 0  
Pr(|Z| < |z|) = 0.0135

Ha: diff > 0  
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        | 0.0277642 | 0.0036261 | 0.0217999 | 0.0337286 |
| y        | 0.0418307 | 0.0044413 | 0.0345255 | 0.0491359 |

| diff     | -0.0140665 | 0.0057335 | -0.0234972 | -0.0046357 |
| z under Ho: | 0.005732 | -2.45 | 0.014 |

diff = prop(x) - prop(y)  
Ho: diff = 0

Ha: diff < 0  
Pr(Z < z) = 0.0071

Ha: diff != 0  
Pr(|Z| < |z|) = 0.0141

Ha: diff > 0  
Pr(Z > z) = 0.9929

When we have the individual data the syntax is prtest var, by(groupvar)

```
. * 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        | 0.8461538 | 0.0353795 | 0.7768114 | 0.9154963 |
| 1        | 0.7631579 | 0.0344838 | 0.6955709 | 0.8307449 |

| diff     | 0.082996 | 0.0494048 | -.0138358 | .1798277 |
| z under Ho: | 0.0511986 | 1.62 | 0.105 |

diff = prop(0) - prop(1)  
Ho: diff = 0

Ha: diff < 0  
Pr(Z < z) = 0.9475

Ha: diff != 0  
Pr(|Z| < |z|) = 0.1050

Ha: diff > 0  
Pr(Z > z) = 0.0525