GPPI PPOL 501–02 & -06: Fall 2014
Course Notes #13: Inference for One Population Proportion
Professor Carolyn Hill

I. CONFIDENCE INTERVALS FOR ONE PROPORTION

$$\text{CI} = \hat{\pi} \pm \text{Z}\sqrt{\frac{\pi(1-\pi)}{n}}$$

Where: $\hat{\pi}$ = the sample proportion (or percentage)
π = the population proportion (or percentage)
Z = the Z-score, determined by the alpha level that you select
n = the sample size.
$\pm \text{Z}\sqrt{\frac{\pi(1-\pi)}{n}}$ = the margin of error

→ This formula involves something that we don't know – the population proportion. There are 3 options here: (1) use the sample proportion; (2) use the hypothesized population proportion (if there is one); (3) use 0.5. Different textbooks and different programs use different values.

For this class, we'll use the sample proportion. It is also o.k. to use 0.5 which will result in the largest possible numerator for that fraction, and thus the largest (and most conservative) possible interval.

Example: Random sample of 1000 patients treated in a program for alcohol and drug dependency over the last 10 years. 530 patients had been readmitted at least once. Construct a 95% confidence interval of the population proportion.

$$\text{CI} = \hat{\pi} \pm 1.96\sqrt{\frac{(0.53)(0.47}{1000}}$$

$= 0.53 \pm (0.03093)$ CI: [0.499, 0.561]

So, we can say with 95% confidence that between 49.9% and 56.1% of the patients had been readmitted at least once.

Note: we could have used 0.50 in the numerator:

$$\text{CI} = \hat{\pi} \pm 1.96\sqrt{\frac{(0.50)(0.50}{n}}$$ $= 0.53 \pm (0.03099)$ CI: [0.499, 0.561]

II. HYPOTHESIS TESTING FOR ONE POPULATION PROPORTION

The test statistic for a single population proportion is a Z-statistic:

$$Z_{calc} = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$$

• NOTE: As above for the CI, different programs and different books estimate the s.e. in this hypothesis test differently – some use the sample proportion; some use the hypothesized proportion (as Weiss does for the single proportion test); and some use the value that will provide the larger possible s.e. for a proportion: 0.50 (see Weiss p. 55).

I emphasize the sample proportion here, to be consistent with how Weiss presents the standard error for the difference of two proportions, later.

- Example: A survey shows that 10% of the population is victimized by property crime each year. A random sample of 527 older citizens (65 years or older) shows a victimization rate of 14%. Are older people victimized to the same degree as the population as a whole?

(1*) Informal expectation…..

(2) State the null and alternative hypotheses:

H_0: $\pi_{elderlyvic} = 0.10$
H_1: $\pi_{elderlyvic} \neq 0.10$

(3) Identify confidence level and critical values:

-- $\alpha = 0.05$, two-tailed test, Z (critical): $= \pm 1.96$
-- $\alpha = 0.01$, two-tailed test, Z (critical): $= \pm 2.576$

(4) Calculate the test statistic: $Z = \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} = \frac{0.14 - 0.10}{\sqrt{(0.14)(0.86)/527}} = \frac{0.04}{0.01512} = 2.646$

(5) Make a decision. Compare the absolute value of the calculated Z-statistic to the Z-critical values: 2.646 > 1.96>1.96

- Draw a picture:

- Conclude that it is unlikely that we would have drawn a sample where 14% of the elderly persons were victimized by property crime, if in fact they were drawn from a population of elderly where 10% of the elderly were victimized by property crime ($p<0.01$)

⇒ **Decision: Reject the Null Hypothesis** that 10% of the elderly are victimized by property crime ($p<0.01$)

NOTE1: If we would have used 0.5 instead of the observed sample proportion above, the calculated test statistic would have been 1.8365.

NOTE2: you can use the Normal curve (and thus the Z-score) for testing sample proportions when the sample size is "sufficiently large." What does this mean?

From Weiss, *Introductory Statistics, 9th Ed.* (p. 547):

-- depends on the sample size (*n)* and the sample proportion $\hat{\pi}$.

-- "If $\hat{\pi}$ is close to 0.5, the approximation is quite accurate, even for moderate *n.* The farther $\hat{\pi}$ is from 0.5, the larger *n* must be for the approximation to be accurate. As a rule of thumb, we use the normal approximation when $n\hat{\pi}$ *and* $n(1-\hat{\pi})$ *are both 5 or greater."*

STATA EXAMPLE : Back to the NLS data used in the earlier examples.

- Consider the proportion of persons in this sample who are married: the codebook indicates that the variable MARRIED is equal to 1 if a sample member was married in 1976. It is implied that this variable is equal to 0 if this condition is not met

- I looked at the Census Bureau website, and used data from 1970 and 1980 to interpolate the percentage of the population that were married in 1976 – about 63%.

- Does the NLS sample represent a population where 63% of all persons are married?

  H_0: $\pi_{marr} = 0.63$
  H_1: $\pi_{marr} \neq 0.63$

- Indicate levels of statistical significance and corresponding critical values.
  -- $\alpha = 0.05$, two-tailed test, Z (critical): $= \pm 1.96$
  -- $\alpha = 0.01$, two-tailed test, Z (critical): $= \pm 2.576$

.

- Calculate the test statistic: $Z = \dfrac{\hat{\pi} - \pi}{\sqrt{\pi(1-\pi)/n}} = \dfrac{0.714 - 0.63}{\sqrt{(0.63)(0.37)/3003}} = \dfrac{0.084}{0.0088} = 9.53$

- Compare the absolute value of the calc Z-statistic to the Z-crit value:$|\ 9.53\ | > |\ \pm 2.576|$

- Conclude that it is highly unlikely that 71.4% of the persons in this sample were married, if in fact they were drawn from a population where 63% of the persons were married. Thus,

we **reject the null hypothesis** that the 63% of the persons are married in this sampled population ($p<0.01$)

- Look at the STATA output: Z= 9.53 here too: the particular syntax specified the null hypothesis, and that is what Stata used in calculating the S.E.

```
. describe, short

Contains data from J:\STATA datasets\card.dta
  obs:         3,010
 vars:            41
 size:       174,580 (98.3% of memory free)
Sorted by:

. describe, simple
id        fatheduc  step14    reg665    south66   wage      libcrd14  collgrd1  wagenew
nearc2    motheduc  reg661    reg666    black     enroll    exper     hsgrad2
nearc4    weight    reg662    reg667    smsa      kww       lwage     hsgrmore
educ      momdad14  reg663    reg668    south     iq        expersq   collgrd2
age       sinmom14  reg664    reg669    smsa66    married   hsgrad1   collmore

. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
          id |      3010    2581.749    1500.539          2       5225
      nearc2 |      3010    .4408638    .4965731          0          1
      nearc4 |      3010    .6820598    .4657535          0          1
        educ |      3010    13.26346    2.676913          1         18
         age |      3010     28.1196    3.137004         24         34
-------------+--------------------------------------------------------
    fatheduc |      2320    10.00345    3.720737          0         18
    motheduc |      2657    10.34814    3.179671          0         18
      weight |      3010    321185.3    170645.8      75607    1752340
    momdad14 |      3010    .7893688    .4078247          0          1
    sinmom14 |      3010    .1006645    .3009339          0          1
-------------+--------------------------------------------------------
      step14 |      3010    .0388704    .1933182          0          1
      reg661 |      3010    .0465116    .2106253          0          1
      reg662 |      3010    .1607973     .367405          0          1
      reg663 |      3010    .1956811      .39679          0          1
      reg664 |      3010    .0641196    .2450066          0          1
-------------+--------------------------------------------------------
      reg665 |      3010    .2083056     .406164          0          1
      reg666 |      3010    .0960133    .2946584          0          1
      reg667 |      3010    .1099668    .3129003          0          1
      reg668 |      3010    .0282392     .165683          0          1
      reg669 |      3010    .0903654    .2867522          0          1
-------------+--------------------------------------------------------
     south66 |      3010    .4142857    .4926801          0          1
       black |      3010    .2335548    .4231624          0          1
        smsa |      3010    .7129568    .4524571          0          1
       south |      3010    .4036545    .4907113          0          1
      smsa66 |      3010    .6495017    .4772053          0          1
-------------+--------------------------------------------------------
        wage |      3010    577.2824    262.9583        100       2404
```

```
      enroll |      3010    .0923588    .2895799          0          1
         kww |      2963    33.54067    8.611619          4         56
          iq |      2061    102.4498    15.42376         50        149
     married |      3003    2.271395    2.066823          1          6
-------------+--------------------------------------------------------
    libcrd14 |      2997     .674341    .4686987          0          1
       exper |      3010    8.856146    4.141672          0         23
       lwage |      3010    6.261832    .4437976    4.60517   7.784889
     expersq |      3010    95.57907    84.61831          0        529
     hsgrad1 |      3010    .8348837    .3713472          0          1
-------------+--------------------------------------------------------
    collgrd1 |      3010    .2714286    .4447705          0          1
     hsgrad2 |      3010    .3295681    .4701344          0          1
    hsgrmore |      3010     .233887    .4233715          0          1
    collgrd2 |      3010    .1524917    .3595566          0          1
    collmore |      3010    .1189369     .323768          0          1
-------------+--------------------------------------------------------
     wagenew |      3010    5.772824    2.629583          1      24.04

. summarize wagenew

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     wagenew |      3010    5.772824    2.629583          1      24.04

. tabulate married

    married |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      2,144       71.40       71.40
          2 |         14        0.47       71.86
          3 |          3        0.10       71.96
          4 |        155        5.16       77.12
          5 |        102        3.40       80.52
          6 |        585       19.48      100.00
------------+-----------------------------------
      Total |      3,003      100.00

. gen marr76=0

. replace marr76=1 if (married==1)
(2144 real changes made)

. replace marr76=. if (married==.)
(7 real changes made, 7 to missing)

. tab married marr76, missing

           |              marr76
   married |         0          1          . |     Total
-----------+---------------------------------+----------
         1 |         0      2,144          0 |     2,144
         2 |        14          0          0 |        14
         3 |         3          0          0 |         3
         4 |       155          0          0 |       155
```

```
         5 |       102          0          0 |       102
         6 |       585          0          0 |       585
         . |         0          0          7 |         7
-----------+---------------------------------+----------
     Total |       859      2,144          7 |     3,010
```

```
. prtest marr76=0.63

One-sample test of proportion                marr76: Number of obs =     3003
------------------------------------------------------------------------------
    Variable |       Mean   Std. Err.                     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      marr76 |   .7139527   .0082466                      .6977896    .7301158
------------------------------------------------------------------------------
    p = proportion(marr76)                                            z =   9.5289
Ho: p = 0.63

    Ha: p < 0.63                 Ha: p != 0.63                 Ha: p > 0.63
 Pr(Z < z) = 1.0000         Pr(|Z| > |z|) = 0.0000          Pr(Z > z) = 0.0000
.
end of do-file
```