Section 8.1 - Inference for a Single Proportion

Statistics 104

Autumn 2004

Inference for a Single Proportion

For most of what follows, we will be making two assumptions

1. The response variable is based on $X \sim Bin(n,p)$, i.e. we observe the number of successes X or the proportions of successes $\hat{p}$.

2. The sample size n is large, so the normal approximation to the binomial is valid. e.g.

$$\hat{p} \overset{approx.}{\sim} N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

Two problems of interest:

1. Confidence interval for p

2. Tests on the hypothesis $H_0 : p = p_0$

If the normal approximation for $\hat{p}$ is reasonable, we can use procedures similar to the ones for the mean of a single population.

Confidence Interval for p

The confidence interval is based on the Wilson estimate for p

$$\tilde{p} = \frac{X + 2}{n + 4}$$

This is similar to the sample proportion but it adds 2 successes and two failures to the data. This pulls the estimate of p towards $\frac{1}{2}$ slightly.

The standard error of $\tilde{p}$ is

$$SE_{\tilde{p}} = \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + 4}}$$

Then an approximate level C confidence interval for p is

$$\tilde{p} \pm z^* SE_{\tilde{p}} = \tilde{p} \pm z^* \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$$

where z^* is a normal critical value.

Note: This Wilson estimator is only used for calculating the confidence interval. You still want to use $\hat{p}$ as your best guess for p.

Example: Use of Aspirin to prevent strokes

155 patients who have had a previous stroke

- 78 received a daily aspirin dose

- 77 received a placebo

The response of interest was no stroke in a 6 month period after starting treatment.

Aspirin group: 63 or 78 with no strokes

Placebo group: 43 of 77 with no strokes

What are the 95% CI for have no addition strokes in the two groups.

Group	$\hat{p}$	$\tilde{p}$
Aspirin	$\frac{63}{78} = 0.808$	$\frac{63+2}{78+4} = 0.793$
Placebo	$\frac{43}{77} = 0.558$	$\frac{43+2}{77+4} = 0.556$

Group	$SE_{\tilde{p}}$
Aspirin	$\sqrt{\frac{0.793 \times 0.207}{78+4}} = 0.0447$
Placebo	$\sqrt{\frac{0.556 \times 0.444}{77+4}} = 0.0552$

95% confidence $\rightarrow z^* = 1.96$

Aspirin:

$$CI = 0.793 \pm 1.96 \times 0.0447 = 0.793 \pm 0.088 = (0.705, 0.881)$$

Placebo:

$$CI = 0.556 \pm 1.96 \times 0.0552 = 0.556 \pm 0.108 = (0.448, 0.664)$$

Since the confidence intervals don't overlap, it appears that giving an aspirin helps prevent strokes. (A hypothesis test for comparing two binomial proportions will be discussed later.)

Stata output:

```
. prtesti 78 63 0.5,count

One-sample test of proportion            x: Number of obs = 78
------------------------------------------------------------
    Variable |       Mean   Std. Err.    [95% Conf. Interval]
-------------+----------------------------------------------
           x |   .8076923   .0446246     .7202298    .8951548
------------------------------------------------------------
```

So the output doesn't match the suggested interval. However we can make it match by adding 2 successes and 2 failures to the data.

```
. prtesti 82 65 0.5,count

One-sample test of proportion            x: Number of obs = 82
------------------------------------------------------------
    Variable |       Mean   Std. Err.    [95% Conf. Interval]
-------------+----------------------------------------------
           x |   .7926829   .0447672     .7049407    .8804251
------------------------------------------------------------
```

This Wilson based interval (also known as the Agresti-Coull "Add Two Success and Two Failures" interval) is something that has started to become more popular recently since it has been shown to have better properties that the traditional interval, such as those calculated by Stata's `prtest` and `prtesti`.

The traditional approach for the binomial confidence interval is

$$\hat{p} \pm z^* SE_{\hat{p}} = \hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

The traditional approach gives intervals of (0.720,895) and (0.448, 0.669).

Note that the Wilson CI for p is not symmetric around $\hat{p}$, the standard estimate of p. Instead it is pulled towards $\frac{1}{2}$.

This is desirable, since the binomial distribution is not symmetric about its mean, unless $p = \frac{1}{2}$.

The asymmetry matches the skewness of the binomial distribution.

What happens if n is small?

Lets assume we did a study where there was 1 success in $n = 10$ trials.

95% Wilson CI for p in this case is (-0.00065, 0.429). The traditional interval is (-0.086, 0.286).

Both these intervals give values below 0, which is impossible for a probability. Stata (and other packages) have routines for calculating binomial confidence intervals for p that don't depend on the normal approximation. The Stata functions `ci` and `cii` base the confidence intervals with exact calculations with the binomial distribution.

```
. cii 10 1
                                                         -- Binomial Exact --
    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
             |         10          .1    .0948683        .0025286    .4450161
```

This interval avoids the problem of having values less than 0 or greater than 1.

This interval approach can be used for larger samples as well.

For the aspirin group

Exact interval:	(0.7027, 0.8882)
Wilson interval:	(0.7049, 0.8804)
Traditional interval:	(0.7202, 0.8952)

This illustrates the point suggesting the the Wilson interval has better coverage properties.

Significance Test for p

The test statistic for examining $H_0 : p = p_0$ is

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

If n is large, z is approximately normal. A rule thumb for n being large is $np_0 \geq 10$ and $n(1 - p_0) \geq 10$.

To examine how this test works, lets look at whether $H_0 : p = \frac{1}{2}$ holds for either the aspirin or the placebo group.

Aspirin:

$$z = \frac{0.808 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{78}}} = 5.435$$

Placebo:

$$z = \frac{0.558 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{77}}} = 1.026$$

The asymptotic p-values for the z test are given by

$$\begin{aligned} H_A : p > p_0 &\quad p\text{-value} = P[Z \geq z_{obs}] \\ H_A : p < p_0 &\quad p\text{-value} = P[Z \leq z_{obs}] \\ H_A : p \neq p_0 &\quad p\text{-value} = 2 \times P[Z \geq |z_{obs}|] \end{aligned}$$

So the two-sided p-values for these tests are

Aspirin: 5.48e-08

Placebo: 0.305

Stata output (Placebo group)

```
. prtesti 77 43 0.5,count

One-sample test of proportion              x: Number of obs = 77

------------------------------------------------------------------
    Variable |       Mean   Std. Err.        [95% Conf. Interval]
-------------+----------------------------------------------------
           x |   .5584416   .0565897         .4475277    .6693554
------------------------------------------------------------------

                          Ho: proportion(x) = .5

       Ha: x < .5              Ha: x != .5              Ha: x > .5
        z =  1.026               z =  1.026               z =  1.026
    P < z =  0.8475        P > |z| =  0.3051          P > z =  0.1525
```

Exact test

When n is small, this asymptotic test may give poor answers as the normal approximation to the binomial breaks down.

Instead of using the normal distribution to get p-values, we can use the binomial distribution to calculate them exactly.

Exact p-values:

$$\begin{array}{ll} H_A : p > p_0 & p\text{-value} = P[X \geq x_{obs}] = p_u \\ H_A : p < p_0 & p\text{-value} = P[X \leq x_{obs}] = p_l \\ H_A : p \neq p_0 & p\text{-value} = \text{complicated but approximately} \\ & \qquad\qquad 2 \times \min(p_u, p_l) \end{array}$$

This exact test can be done in Stata with `bitest` and `bitesti`

```
. bitesti 77 43 0.5

        N   Observed k   Expected k   Assumed p   Observed p
------------------------------------------------------------
       77         43          38.5       0.50000      0.55844

  Pr(k >= 43)             = 0.181016  (one-sided test)
  Pr(k <= 43)             = 0.872848  (one-sided test)
  Pr(k <= 34 or k >= 43) = 0.362032  (two-sided test)
```

As with the one-sample t test, testing a single proportion usually isn't particularly interesting.

Usually of more interest is a test comparing whether two proportions are the same. For example, is the stroke rate in the aspirin group the same as the rate in the placebo group. (To come next class)

However, this binomial test can be useful as a non-parametric test in the paired sample comparison when the normality assumption is not reasonable.

Instead of looking at $A - B$, you can look at whether A is greater or less than B.

If the distribution of A and B is the same, then $P[A > B] = P[A < B] = \frac{1}{2}$.

So instead of doing a t test on $A - B$, you can do a test on whether the sample proportion of times that $A > B$ differs from $\frac{1}{2}$. (For this test, you usually toss out the observations where $A = B$.

This test is often referred to as the Sign test.

Example: Shoe soles

In the 10 pairs of shoes, material A has greater wear than material B 8 times.

$H_0 : p = \frac{1}{2}$ vs $H_A : p \neq \frac{1}{2}$

$$\begin{aligned} p - \text{value} &= 2 \times P[\hat{p} \geq \hat{p}_{obs}] = 2 \times P[X \geq x_{obs}] \\ &= 2 \times P[X \geq x_{obs}] = 0.109 \end{aligned}$$

```
. bitesti 10 8 0.5

        N   Observed k   Expected k   Assumed p   Observed p
------------------------------------------------------------
       10          8            5       0.50000      0.80000

  Pr(k >= 8)            = 0.054688  (one-sided test)
  Pr(k <= 8)            = 0.989258  (one-sided test)
  Pr(k <= 2 or k >= 8) = 0.109375  (two-sided test)
```

The Sign test is usually less powerful than the paired t test. You usually one want to use it if the assumptions for the t test are strongly violated since the t test is fairly robust.