Tutorial: Methods for one-sample proportion inference

In this tutorial, we learn about Stata commands for one-sample proportion inference:

Confidence intervals:

`ci` and `cii` – calculate binomial confidence intervals

Hypothesis Tests:

`bitest` and `bitesti` – exact binomial one-sample proportion hypothesis test

`prtest` and `prtesti` – large-sample one-sample proportion hypothesis test

Recall that the extra 'i' at the end of a Stata command name denotes that the command is "immediate" and does not use the data in memory.

Exercises

1. Estimate the proportion of California residents who visit the doctor at least once in the previous year, denoted *p*.

```
. tabulate doctor
```

2. Construct a 95% confidence interval for *p* using three different methods (Can we use the normal approximation to the binomial distribution?). How do the widths of these three CI's compare?

```
. ci doctor, binomial
. ci doctor, binomial wald
. ci doctor, binomial Wilson
```

Exact: never has lower than expected coverage, but is sometimes too conservative
Wald: Large-sample, bad coverage, easy to calculate/flexible
Wilson: Large-sample, good coverage, less flexible

3. Using the 95% confidence level, is there evidence in the data that less than 80% of the population visits the doctor once per year? Repeat this analysis, stratifying by above/below poverty groups.

```
. bysort poverty: ci doctor, binomial
```

4. Let's formalize question 3 using a hypothesis test. Let p_1 denote the proportion of California residents below the federal poverty level who visited the doctor at least once in the past year. Test the hypothesis that $p_1 = 0.8$ versus the alternative that $p_1 \neq 0.8$ at the $\alpha = 0.05$ level.

(a) First, use the exact binomial test. What is the p-value?

```
bitest doctor == 0.8 if poverty == 1
```

(b) Next, use the normal approximation to the binomial distribution.

```
. prtest doctor == 0.8 if poverty==1
```

- Is the normal approximation appropriate?

$n_1 * p_1 > 5$; $n_1 * (1 - p_1) > 5$

Therefore, the normal approximation to the binomial is appropriate.

- What is the value of your test statistic?

  $Z = -2.02$

- What is the distribution of your test statistic under the null hypothesis?

  $Z \sim N(0,1)$

- What is the p-value of your test?

  $p = 0.044$

- Do you reject or not reject the null hypothesis?

  We reject the null hypothesis.

- What do you conclude?

We conclude that there is evidence in the data that p_1 is less than 0.8.

5. Given that you got different results using the exact and large sample hypothesis tests, what would you do if you were writing a paper?

   There are no meaningful differences between a p-value of 0.049 and 0.051 - try to include confidence intervals in practice, as p-values don't tell you anything about the magnitude of an effect.