M5S4 - Practice with CIs

Professor Jarad Niemi

STAT 226 - Iowa State University

October 18, 2018

Outline

- Constructing confidence intervals
    - Review
    - When to use z vs t
    - Practice
    - Proportions

Confidence Interval Review

Two methods of constructing confidence intervals for the population mean μ:

$$\overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad \text{and} \qquad \overline{x} \pm t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}$$

where

- $\overline{x}$ is the sample mean,
- s is the sample standard deviation,
- n is the sample size,
- σ is the known population standard deviation,
- $z_{\alpha/2}$ is the z critical value such that $P(Z > z_{\alpha/2}) = \alpha/2$,
- $t_{n-1,\alpha/2}$ is the t critical value such that $P(T_{n-1} > t_{n-1,\alpha/2})$ and $n-1$ is the degrees of freedom,
- α is the significance (error) level, and
- $100(1-\alpha)\%$ is the confidence level.

The interpretation of a $100(1-\alpha)\%$ confidence interval is that, on average, $100(1-\alpha)\%$ of the intervals constructed with this procedure will cover μ.

Deciding which method to use

Recall that all our confidence interval formulas require the observations be independent and identically distributed. We usually accomplish this by taking a random sample from the population.

Data	σ	Sample size	Interval
Normal	Known	any	z is exact
Normal	Unknown	any	t is exact
Not normal	Known	large	z is approximate
Not normal	Unknown	any	t is approximate

Estimator for a proportion

Let $X_i \stackrel{iid}{\sim} Ber(p)$, then $Y = \sum_{i=1}^n X_i \sim Bin(n,p)$. An estimator for p is

$$\hat{p} = \frac{Y}{n}$$

with

$$E[\hat{p}] = E\left[\frac{Y}{n}\right] = \frac{E[Y]}{n} = \frac{np}{p}$$

thus $\hat{p}$ is an unbiased estimator and

$$Var[\hat{p}] = Var\left[\frac{Y}{n}\right] = \frac{1}{n^2} Var[Y] = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

thus

$$SD[\hat{p}] = \sqrt{Var[\hat{p}]} = \sqrt{\frac{p(1-p)}{n}}.$$

Confidence interval for a proportion

To construct a $100(1-\alpha)\%$ confidence interval for p, we use the formula

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where $SE[\hat{p}] = \sqrt{\hat{p}(1-\hat{p})/n}$, i.e. our estimate of the SD.

It is common in polling to report $\hat{p}$ and the margin of error $z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$.

2018 Iowa Governor Poll

In the most recent Des Moines register poll of 555 likely voters

https://www.realclearpolitics.com/epolls/2018/governor/ia/iowa_governor_reynolds_vs_hubbell-6477.html

43% indicated they would vote for Fred Hubbell with a margin of error of 4.2.

Thus a 95% confidence interval for the actual proportion who say they would vote for Fred Hubbell is

$$0.43 \pm 0.042 = (0.388, 0.472) = (38.8\%, 47.2\%).$$

The margin of error calculation is

$$2 \cdot \sqrt{\frac{0.43(1 - 0.43)}{555}} = 0.042 = 4.2\%.$$

The best resource for combining all the information from polls is 538:

https://projects.fivethirtyeight.com/2018-midterm-election-forecast/governor/