STAT475/675 TUT02

www.sfu.ca/~zza115/teaching

Zhiyang Zhou (zhiyang_zhou@sfu.ca)

2018-01-15

Setting

- n independent Bernoulli trials
- identical unknown $\pi = \Pr(\text{success})$ in each trial
- r.v. $W = \sum_{i=1}^{n} Y_i \sim B(n, \pi)$, number of successes out of the n trials, with $Y \sim B(1, \pi)$
- α and $1 - \alpha$, respectively, significance level and confidence level

Wald confidence interval for π

$$\hat{\pi} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

- Pros and cons
  - easy to derive and compute
  - When w is close to 0 or n, Calculated limits may be less than 0 or greater than 1.
  - When $w = 0$ or 1, lower and upper limits are exactly the same.

Wilson (score-type) confidence interval for π

$$\tilde{\pi} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\tilde{\pi}(1-\tilde{\pi})}{n + Z^2_{1-\alpha/2}}}$$

with

$$\tilde{\pi} = \frac{w + Z^2_{1-\alpha/2}/2}{n + Z^2_{1-\alpha/2}}$$

- Pros and cons
  - always has limits between 0 and 1

Agresti-Coull confidence interval for π

$$\tilde{\pi} \pm \frac{Z_{1-\alpha/2}\sqrt{n}}{n + Z^2_{1-\alpha/2}} \sqrt{\hat{\pi}(1-\hat{\pi}) + \frac{Z^2_{1-\alpha/2}}{4n}}$$

with

$$\tilde{\pi} = \frac{w + Z^2_{1-\alpha/2}/2}{n + Z^2_{1-\alpha/2}}$$

- Pros and cons
  - easier to calculate by hand than Wilson
  - resembles the popular Wald interval
  - may have limits less than 0 or greater than 1

Clopper-Pearson (exact) confidence interval for π

$$(\text{quantile.beta}(\alpha/2; w, n-w+1), \quad \text{quantile.beta}(1-\alpha/2; w+1, n-w))$$

- Pros and cons
  - always conservative: the true confidence level is always equal to or greater than the stated level
  - infeasible to compute by hand

Generic R function for computing confidence intervals

- binom::binom.confint()

Exercise (Q9 in Textbook pp.51)

79 out of 80 people oppose a new tax. Let π represent the probability that a randomly selected resident opposes the tax.

- Compute four different 95% CIs for π; interpret the results.
- Is it possible $\pi = 1$?
- Which interval is prefered?

True confidence levels (coverage probabilies) for confidence intervals

- We assert that the 95% (Wald, Wlison, Agresti-Coull or Clopper-Pearson) CI covers the true value with probability 95%. But what is the truth?
- The true confidence level at π, say $C(\pi)$, is the sum of the binomial probabilities for all intervals that do contain π:
$$C(\pi) = \sum_{w=0}^{n} I(w) \binom{n}{w} \pi^w (1-\pi)^{n-w},$$
where $I(w) = 1$ if the interval formed with w contains π and $I(w) = 0$ otherwise.

Exercise (continued)

What are the true coverage probabilities for the four methods with $n = 80$ and $\alpha = .05$?