10. Confidence Interval for a Population Proportion

Recall: distribution of sample proportion

The parameter of interest for qualitative data is the proportion of times a particular outcome (a success) occurs.

To estimate population proportion p we use sample proportion

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

where X is the number of successes in the sample, and n is the sample size. Random variable X has a binomial distribution. However, if sample size n is large (specifically, $np(1-p) > 5$), then sample proportion $\hat{p}$ is approximately normally distributed:

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

This allows us to construct a z-CI for proportion p.

Large sample confidence interval for p

The $100(1-\alpha)\%$ confidence Interval for true proportion p is given by

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

This formula provides accurate results when the sample size is *large*, that is, $np(1-p) > 5$.

Since p is *unknown* population parameter, in practice, we check $n\hat{p}(1-\hat{p}) > 5$, because $p \approx \hat{p}$ by the Law of Large Numbers.

z-CI for proportion p: example

***Example 1*. A retail lumberyard routinely inspects incoming shipments of lumber from suppliers. For select grade 8-foot 2-by-4 pine shipments, the lumberyard supervisor chooses one gross (144 boards) randomly from a shipment of several tens of thousands of boards. In the sample, 18 boards are not salable as select grade. Calculate a 95% CI for the proportion in the entire shipment that is not salable as select grade.**

z-CI for proportion p: example

Solution.

The sample proportion

$$\hat{p} = \frac{18}{144} = .125$$

Since $n\hat{p}(1-\hat{p}) = 144 \times .125 \times .875 = 15.75 > 5$ the confidence interval is given by

$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} = .125 \pm 1.96\sqrt{\frac{.125 \times .875}{144}} = .125 \pm .054$$

Thus, we can claim that with confidence level 95% that the true proportion in the entire shipment that is not salable as select grade lies in the confidence interval $.125 \pm .054$

.

Sample size determination

The required sample size to produce an interval estimator $\hat{p} \pm W$ that estimates population proportion p with confidence level $1 - \alpha$ is given by

$$n \approx \left(\frac{z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})}}{W} \right)^2$$

Sample size determination

But the value of sample proportion $\hat{p}$ is unknown, because we only plan to collect observations. There are two ways how we can address this difficulty.

- (*historical value of* $\hat{p}$) We can use $\hat{p}$ from a prior sample.
- *(conservative choice of* $\hat{p}$) One can show that $\hat{p}(1-\hat{p}) \leq 1/4.$ Therefore, we can substitute $\hat{p}(1-\hat{p})$ by its largest possible value to protect ourselves against the worst case scenario. This leads to a more practical formula:

$$n \approx \left(\frac{z_{\alpha/2}}{2W}\right)^2$$

Sample size determination: example

Example 2. A manufacturer of boxes of candy is concerned about the proportion of imperfect boxes – those containing cracked, broken, or otherwise unappetizing candies. How large a sample is needed to get 95% CI for this proportion with a width no greater than .02?

Solution. Since no prior info on the sample proportion is available, we will use the conservative substitution:

$$n \approx \left(\frac{z_{\alpha/2}}{2W}\right)^2 \approx \left(\frac{1.96}{2 \times .01}\right)^2 \approx 9604.$$

CI for p: exercises

Exercise 1. A city council commissioned a statistician to estimate the proportion p of voters in favor of a proposal to build a new library. The statistician obtained a random sample of 200 voters, with 112 indicating approval of the proposal.

1. What is a point estimate for p?
2. Find a 98% confidence interval for p.
3. How large a sample is needed to get the common ± 3 percentage point margin of error with the same 98% confidence level?

CI for p: exercises

Exercise 2. To estimate the proportion p of voters favoring a nuclear freeze in your voting district, how large a random sample is needed to estimate p to within ± 2 percentage points with

1. 90% confidence?
2. 99% confidence?