

Single Proportion Inference

Proportion definition

- ▶ A proportion is a mean of 0s and 1s.
- ▶ For example, 29% of Americans live with a cat. There are 330 million Americans. Thus, if we categorized each American as either 0 (for not living with cat) or 1 (for living with cat), there would be 243 million 0s and 96 million 1s.
- ▶ We use p for the population parameter (analogous to μ).
- ▶ We use \hat{p} for the sample statistic (analogous to \bar{x}).
- ▶ We often measure \hat{p} to infer about p . We say \hat{p} is a point estimate of p .

Measuring a sample proportion

- ▶ Measuring a sample proportion can be modelled as a sampling distribution.
- ▶ Let W represent a Bernoulli random variable with probability of success p .

outcome	probability
0	$1 - p$
1	p

- ▶ A sample proportion \hat{P} can represent an average of many instances of W . For example, maybe the sample size is 7:

$$\hat{P} = \frac{W + W + W + W + W + W + W}{7}$$

Formulas

- ▶ The expected value of \hat{P} is p .

$$E(\hat{P}) = p$$

- ▶ So, \hat{p} is a point estimate of the parameter of interest, p .

$$\hat{p} \approx p$$

- ▶ The standard deviation of \hat{P} is the standard error.

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

- ▶ If $pn > 10$ and $(1-p)n > 10$, then \hat{P} is approximately normal.

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

Confidence interval

- If $\hat{p}n > 10$ and $(1 - \hat{p})n > 10$, then we can find confidence intervals using the margin of error.

$$p \approx \hat{p} \pm ME$$

$$ME = z^* SE$$

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

$$z^* \text{ satisfies } \Pr(|Z| < z^*) = CL$$

Example problem

1. Determine the 95% confidence interval of the proportion when a sample of size 100 has 34 successes.

Example problem

2. How large of a sample is needed to have a 95% confidence interval with a margin of error less than 0.01 if we can assume $\hat{p} \approx 0.34$?

Example problem

3. How large of a sample is needed to have a 95% confidence interval with a margin of error less than 0.01?

Hypothesis testing (two-tailed)

When testing a null hypothesis, use the null proportion in the standard error calculation.

$$H_0 : p = p_0$$

$$H_A : p \neq p_0$$

You first determine the standard error.

$$SE = \sqrt{\frac{p_0(1 - p_0)}{n}}$$

Then, you determine a z score.

$$z = \frac{\hat{p} - p_0}{SE}$$

Then, you determine a p -value (which you compare to α).

$$p\text{-value} = 2 \cdot \Phi(-|z|)$$

Example problem

4. A 6-sided die, with one side painted green, is rolled 600 times. The green side showed up 80 times. Is the die fair?