CONFIDENCE INTERVALS

MAT 152 – Statistical Methods I
Lecture 3
Instructor: Dustin Roten
Fall 2020

Population Proportions

So far, we have used population samples to determine a population parameter and associated confidence interval. What if we wanted to ask a question that doesn't pertain to a parameter value?

- What percent of homes in the Winston-Salem area are equipped with solar panels?
- What percent of elementary school children have cell phones?
- What percent of college students smoke?

None of these questions involve a mean or standard deviation.

None of these questions can be calculated directly because the populations are too large.

How can we approximate a proportion?

Population Proportions

These questions can also be modeled by a binomial distribution. (Success/Failure; Yes/No; Win/Lose)

- What percent of homes in the Winston-Salem area are equipped with solar panels?
  - (Equipped/ Not Equipped)
  - X = the number of homes equipped with solar panels
- What percent of elementary school children have cell phones?
  - (Have / Does Not Have)
  - X = the number of school children that have cell phones
- What percent of college students smoke?
  - (Smokes / Does Not Smoke)
  - X = the number of college students that smoke

Recall that, for a binomial distribution, $X \sim B(n, p)$.

Binomial and Normal Distributions

The binomial distribution is discrete. $X \sim B(n, p)$

The normal distribution is continuous. $X \sim N(\mu, \sigma)$

If p is not close to 0 or 1, then a binomial distribution is approximately normal.

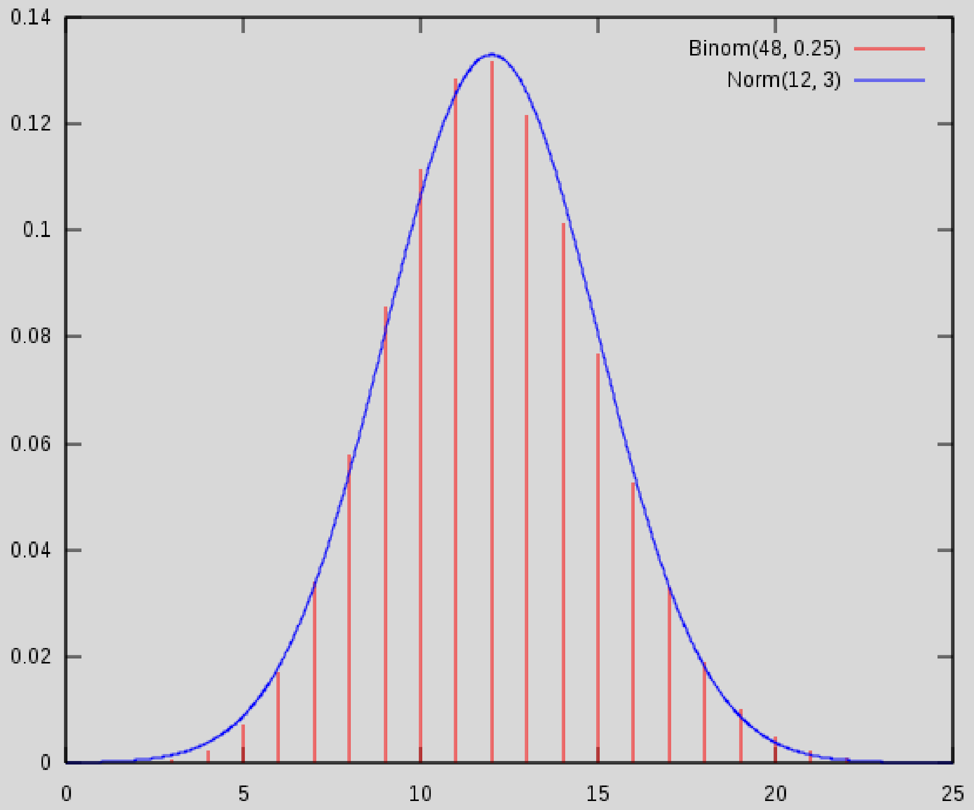
The mean and standard deviation of a binomial distribution can be calculated:

$\mu = np$

$\sigma = \sqrt{npq}$

Substituting into the normal distribution: $X \sim N\left(np, \sqrt{npq}\right)$

This only applies if the number of success and failures are greater than five.



A Normal Distribution for Proportions

Remember, X is the number of successes and $X \sim N\left(np, \sqrt{npq}\right)$.

If we want a distribution for proportions, we must "scale" our current distribution.

A proportion is the number of successes divided by the number of trials.

$$\frac{X}{n} \sim N\left(\frac{np}{n}, \frac{\sqrt{npq}}{n}\right) = N\left(p, \sqrt{\frac{pq}{n}}\right)$$

We do not know the true value of p for the population. However, we can get an estimate from a sample. An estimated p value is referred to as p'. (This is called "p prime".)

So, to calculate the **error bound for a proportion**: $EBP = \left(z_{\frac{\alpha}{2}}\right)\left(\sqrt{\frac{p'q'}{n}}\right)$

Example

For a class project, a political science student at a large university wants to estimate the percent of students who are registered voters. He surveys 500 students and finds that 300 hare registered to vote. Compute a 90% confidence interval for the true percent of students who are registered voters.

5-Step Process:

1. Compute the **estimated proportion** of successes
2. Determine a z-score for the desired confidence level
3. Calculate the **EBP** (Error bound for the proportion)
4. Construct the confidence interval
5. Interpret the result

Example (Step 1)

For a class project, a political science student at a large university wants to estimate the percent of students who are registered voters. He surveys 500 students and finds that 300 hare registered to vote. Compute a 90% confidence interval for the true percent of students who are registered voters.

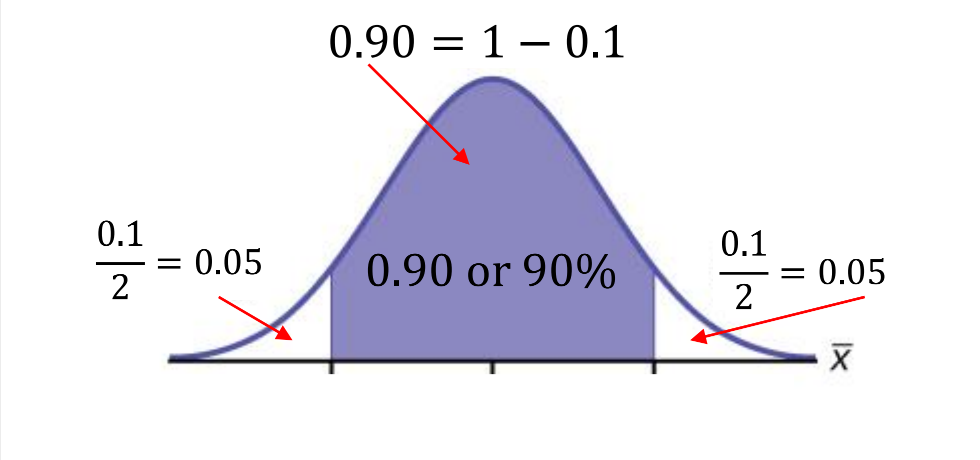
Compute the **estimated proportion** of successes

Remember: We do not know the TRUE proportion of successes for our population. We can only approximate it from the sample.

$$p' = \frac{x}{n} = \frac{300}{500} = 0.6; q' = 1 - 0.6 = 0.4$$

Example (Step 2)

Determine a z-score for the desired confidence level (90%)



Confidence Level	Area in One Tail	Exact Z-score
50%	0.25	0.674
80%	0.1	1.282
90%	0.05	1.645
95%	0.025	1.96
98%	0.01	2.326
99%	0.005	2.576

$$z_{\frac{\alpha}{2}} = z_{0.05} = 1.645$$

Example (Step 3 and Step 4)

Calculate the **EBP** (Error bound for the proportion)

Construct the confidence interval

$$EBP = \left(z_{\frac{\alpha}{2}}\right)\sqrt{\frac{p'q'}{n}} = 1.645\sqrt{\frac{0.6 \cdot 0.4}{500}} = 0.036$$

Lower Bound: $p' - EBP = 0.6 - 0.036 = 0.564$

Lower Bound: $p' + EBP = 0.6 + 0.036 = 0.636$

The confidence interval is (0.564, 0.636)

Example (Step 5)

For a class project, a political science student at a large university wants to estimate the percent of students who are registered voters. He surveys 500 students and finds that 300 hare registered to vote. Compute a 90% confidence interval for the true percent of students who are registered voters.

5-Step Process:

1. Compute the **estimated proportion** of successes $p' = 0.6$
2. Determine a z-score for the desired confidence level $z_{\frac{\alpha}{2}} = z_{0.05} = 1.645$
3. Calculate the **EBP** (Error bound for the proportion) $\text{EBP} = 0.036$
4. Construct the confidence interval (0.564, 0.636)
5. Interpret the result

We estimate with 90% confidence that the true percent of all students that are registered voters is between 56.4% and 63.6%

"Plus Four" Confidence Interval for p

Since point estimates are used to calculate the standard deviation of the sampling distribution, there is a certain amount of error introduced into the process of calculating a confidence interval for a proportion.

There is a simple rule that adjusts confidence intervals.

Four trials are added: two successes and two failures.

$$p' = \frac{\text{success}}{\text{trials}} = \frac{x}{n} \text{ becomes } p' = \frac{x+2}{n+4}$$

Quick Example

A random sample of 25 statistics students was asked: “Have you smoked a cigarette in the past week?” Six students reported smoking within the past week. Use the “plus four” method to find a 95% confidence interval for the true proportion of statistics students who smoke.

$p = \frac{x}{n} = \frac{6+2}{25+4} = \frac{8}{29} = 0.276$

$q = 1 - 0.276 = 0.724$

$CL = 0.95$ so $\frac{\alpha}{2} = 0.025$ and $z_{0.025} = 1.96$

$EBP = \left(z_{\frac{\alpha}{2}}\right)\sqrt{\frac{p'q'}{n}} = 1.96\sqrt{\frac{0.276 \cdot 0.724}{29}} \approx 0.163$

Confidence interval: (0.113, 0.439)

We are 95% confidence that the true proportion of all statistics students who smoke cigarettes is between 0.113 and 0.439.

A Quick Review

- If p isn't close to 0 or 1 and the number of successes and failures is at least five, a binomial distribution can be approximated by a normal distribution.
- $X \sim N\left(np, \sqrt{npq}\right)$
- An error bound for a proportion (EBP) can be calculated.
- Sample proportions p' and q' are estimates of the unknown population proportions p and q.
- The "plus four" adjustment can be used to reduce error in proportion estimates.