STAT 2593

Lecture 026 - Large Sample Confidence Intervals for a Population Mean and Proportion

Dylan Spicker

Large Sample Confidence Intervals for a Population Mean and Proportion

Learning Objectives

1. Construct confidence intervals for normal population means in large samples with unknown variance.

2. Construct confidence intervals for the sample proportion in large samples.

A quick disclaimer on introductory statistics education.

Overcoming an Unknown Variance

- We have made the **unrealistic assumption** that σ^2 is known.

Overcoming an Unknown Variance

- We have made the **unrealistic assumption** that σ^2 is known.
- When it is not known, we will need to address this somehow.

Overcoming an Unknown Variance

- We have made the **unrealistic assumption** that σ^2 is known.
- When it is not known, we will need to address this somehow.
  - One possibility is to replace σ^2 with s^2, the sample variance.

Overcoming an Unknown Variance

- We have made the **unrealistic assumption** that σ^2 is known.
- When it is not known, we will need to address this somehow.
  - One possibility is to replace σ^2 with s^2, the sample variance.
  - Doing this renders the sampling distribution to be non-normal.

Overcoming an Unknown Variance

- We have made the **unrealistic assumption** that σ^2 is known.
- When it is not known, we will need to address this somehow.
  - One possibility is to replace σ^2 with s^2, the sample variance.
  - Doing this renders the sampling distribution to be non-normal.
- If n is sufficiently large we can appeal to the **central limit theorem** and say that

$$Z = \frac{\overline{X} - \mu}{s/\sqrt{n}} \dot{\sim} N(0,1).$$

Overcoming an Unknown Variance

- We have made the **unrealistic assumption** that σ^2 is known.
- When it is not known, we will need to address this somehow.
  - One possibility is to replace σ^2 with s^2, the sample variance.
  - Doing this renders the sampling distribution to be non-normal.
- If n is sufficiently large we can appeal to the **central limit theorem** and say that

$$Z = \frac{\overline{X} - \mu}{s/\sqrt{n}} \dot{\sim} N(0, 1).$$

  - Note, if n is not large enough, this argument will not hold.

Confidence Intervals for Large Sample Normal Population Means

- Working through the same argument as last time we find:

Confidence Intervals for Large Sample Normal Population Means

- Working through the same argument as last time we find:
  - $P(z_{\alpha/2} \leq Z \leq z_{1-\alpha/2}) \approx 1 - \alpha$.

Confidence Intervals for Large Sample Normal Population Means

- Working through the same argument as last time we find:
  - $P(z_{\alpha/2} \leq Z \leq z_{1-\alpha/2}) \approx 1 - \alpha$.
  - $\ell = \overline{X} - z_{1-\alpha/2}\frac{s}{\sqrt{n}}$.

Confidence Intervals for Large Sample Normal Population Means

- Working through the same argument as last time we find:
  - $P(z_{\alpha/2} \leq Z \leq z_{1-\alpha/2}) \approx 1 - \alpha$.
  - $\ell = \overline{X} - z_{1-\alpha/2}\frac{s}{\sqrt{n}}$.
  - $u = \overline{X} + z_{1-\alpha/2}\frac{s}{\sqrt{n}}$.

Confidence Intervals for Large Sample Normal Population Means

- Working through the same argument as last time we find:
  - $P(z_{\alpha/2} \leq Z \leq z_{1-\alpha/2}) \approx 1 - \alpha$.
  - $\ell = \overline{X} - z_{1-\alpha/2}\frac{s}{\sqrt{n}}$.
  - $u = \overline{X} + z_{1-\alpha/2}\frac{s}{\sqrt{n}}$.
  - The confidence interval is thus $\overline{X} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$.

Confidence Intervals for Large Sample Normal Population Means

- Working through the same argument as last time we find:
  - $P(z_{\alpha/2} \leq Z \leq z_{1-\alpha/2}) \approx 1 - \alpha$.
  - $\ell = \overline{X} - z_{1-\alpha/2}\frac{s}{\sqrt{n}}$.
  - $u = \overline{X} + z_{1-\alpha/2}\frac{s}{\sqrt{n}}$.
  - The confidence interval is thus $\overline{X} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$.
- This can still be used to determine sample sizes

Confidence Intervals for Large Sample Normal Population Means

- Working through the same argument as last time we find:
  - $P(z_{\alpha/2} \leq Z \leq z_{1-\alpha/2}) \approx 1 - \alpha$.
  - $\ell = \overline{X} - z_{1-\alpha/2}\frac{s}{\sqrt{n}}$.
  - $u = \overline{X} + z_{1-\alpha/2}\frac{s}{\sqrt{n}}$.
  - The confidence interval is thus $\overline{X} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$.
- This can still be used to determine sample sizes
  - Use an estimated s, or a conservative one if a range is known.

Confidence Intervals for Population Proportions in Large Samples

- We saw that, with large n and moderate p, binomial distributions are approximately normal.

Confidence Intervals for Population Proportions in Large Samples

- We saw that, with large n and moderate p, binomial distributions are approximately normal.
  - If $X \dot{\sim} N(np, np(1-p))$ then

$$\widehat{p} = \frac{X}{n} \dot{\sim} N\left(p, \frac{p(1-p)}{n}\right).$$

Confidence Intervals for Population Proportions in Large Samples

- We saw that, with large n and moderate p, binomial distributions are approximately normal.
  - If $X \dot{\sim} N(np, np(1-p))$ then
$$\widehat{p} = \frac{X}{n} \dot{\sim} N\left(p, \frac{p(1-p)}{n}\right).$$
  - As a result, by the **exact same argument** a confidence interval for $\widehat{p}$ is
$$\widehat{p} \pm z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}.$$

Confidence Intervals for Population Proportions in Large Samples

- We saw that, with large n and moderate p, binomial distributions are approximately normal.
  - If $X \dot{\sim} N(np, np(1-p))$ then
$$\widehat{p} = \frac{X}{n} \dot{\sim} N\left(p, \frac{p(1-p)}{n}\right).$$
  - As a result, by the **exact same argument** a confidence interval for $\widehat{p}$ is
$$\widehat{p} \pm z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}.$$
- We can use the same procedure for estimating the required sample size.

Confidence Intervals for Population Proportions in Large Samples

- We saw that, with large n and moderate p, binomial distributions are approximately normal.
  - If $X \dot{\sim} N(np, np(1-p))$ then

$$\widehat{p} = \frac{X}{n} \dot{\sim} N\left(p, \frac{p(1-p)}{n}\right).$$

  - As a result, by the **exact same argument** a confidence interval for $\widehat{p}$ is

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

- We can use the same procedure for estimating the required sample size.
  - If an estimate of p is not known, taking $p = 0.5$ remains conservative.

Summary

- When variance is unknown, but the population is still assumed to be normal, we can use the central limit theorem and large sample sizes to justify the same procedure for constructing confidence intervals.

- When a population proportion is estimated, if the normal approximation to the binomial applies, the same procedure can be applied.