Lecture 24

Confidence Intervals

Manju M. Johny

STAT 330 - Iowa State University

Confidence Intervals

Confidence Intervals

- MLE gives us a "point estimate" of the unknown parameter.
- But $\hat{\theta}$ probably won't *exactly* equal θ due to sampling error.
  $\rightarrow P(\theta = \hat{\theta}) = 0$
- Create a confidence interval to give range of reasonable values for the unknown parameter θ.

Example 1: Polling

Today's poll shows 58% of people favor the new bill. The margin of error is $\pm 3\%$.

The confidence interval for the proportion of people that favor the bill is $[0.55, 0.61]$.

Confidence Interval

Definition

A random interval $[a, b]$ is a $(1-\alpha)100\%$ *confidence interval* for the parameter θ if it contains θ with probability $(1-\alpha)$

$$P(a \leq \theta \leq b) = 1 - \alpha$$

- $(1-\alpha)$ is called the confidence level
- When you estimate an unknown parameter θ, it should be accompanied by a confidence interval

- Interpretation: We are [$(1-\alpha)\%$] confident that the [insert population parameter + context] is between [insert interval + units].

Constructing Confidence Intervals

In this class, we will construct normal distribution based intervals.

Suppose we have an estimator $\hat{\theta}$ for unknown parameter θ.

1. $\hat{\theta}$ is unbiased: $E(\hat{\theta}) = \theta$
2. $\hat{\theta}$ follows a normal distribution.

We can standardize $\hat{\theta}$ to get

$$Z = \frac{\hat{\theta} - \theta}{SE(\hat{\theta})} \sim N(0, 1)$$

where $SE(\hat{\theta}) = \sqrt{Var(\hat{\theta})}$ = standard deviation of $\hat{\theta}$

Constructing Confidence Intervals

Let $z_{\alpha/2}$ be the $1 - \frac{\alpha}{2}$ quantile of the standard normal distribution.

$$P\left(- z_{\alpha/2} \leq \frac{\hat{\theta} - \theta}{SE(\hat{\theta})} \leq z_{\alpha/2} \right) = 1 - \alpha$$

Constructing Confidence Intervals

Isolating θ in the middle, we get

$$P\left(\hat{\theta} - z_{\alpha/2}SE(\hat{\theta}) \leq \theta \leq \hat{\theta} + z_{\alpha/2}SE(\hat{\theta})\right) = 1 - \alpha$$

Thus, a $(1-\alpha)100\%$ confidence interval for θ is

$$\hat{\theta} \pm z_{\alpha/2}SE(\hat{\theta})$$

Common choices for α are $0.01, 0.05,$ and 0.1

$(1-\alpha)100\%$	80	90	95	98	99
$z_{\alpha/2}$	1.282	1.645	1.96	2.326	2.576

Constructing Confidence Intervals Cont.

We will make confidence intervals for four cases:

1. μ (population mean)
2. p (population proportion)
3. $\mu_1 - \mu_2$ (difference in population means)
4. $p_1 - p_2$ (difference in population proportions)

Confidence intervals for all 4 of the above cases can be constructed using normal distribution based inference.

Follow the same general procedure to construct these intervals.

Confidence Interval for Mean

Confidence Interval for μ

Confidence interval for the population mean

$X_1, \ldots, X_n \overset{iid}{\sim} f_X(x)$ with $E(X_i) = \mu$ and $Var(X_i) = \sigma$

First, we estimate μ using the *statistic* $\bar{X}$. From CLT, we know

- $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$
- $SE(\bar{X}) = \sqrt{Var(\bar{X})} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$

A $(1-\alpha)100\%$ confidence interval for μ is

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

In most cases, the population standard deviation σ will be unknown. Replace σ with the sample standard deviation s.

$$\bar{X} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$$

Confidence Interval for μ Cont.

If we want a 95% confidence interval, then

$$1 - \alpha = 0.95$$
$$\rightarrow \alpha = 0.05$$
$$\rightarrow \alpha/2 = 0.025$$

$z_{\alpha/2} = z_{0.025}$ is the 0.975^{th} quantile of the $N(0, 1)$ distribution.
$\rightarrow$ Using the $z - table$, we get $z_{0.025} = 1.96$

The 95% confidence interval for μ is

$$\bar{X} \pm 1.96 \frac{\sigma}{\sqrt{n}} \qquad \text{when } \sigma \text{ is known}$$
$$\bar{X} \pm 1.96 \frac{s}{\sqrt{n}} \qquad \text{when } \sigma \text{ is unknown}$$

Example

Example 2: A random sample of 50 batteries were taken for a particular brand. For the sample, the mean lifetime is 72.5 hours and variance is 19.3 hours^2. Find a 95% confidence interval for the true mean lifetime of batteries from that particular brand.

Example Cont.

Selecting Sample Size for Means

How to decide sample size?

- Can choose the sample size n to obtain a desired level of confidence & width for our confidence interval.
- Margin or error (Δ) is half the width of the confidence interval
  margin of error $= \Delta = z_{\alpha/2} SE(\bar{X}) = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$
- The bigger the sample size, the smaller the standard error of the estimator, and smaller the size of our interval

To attain a particular margin of error Δ, we need a sample size

$$n \geq \left(\frac{z_{\alpha/2} \sigma}{\Delta} \right)^2$$

Confidence Interval for Proportion

Confidence Interval for p

Confidence interval for the population proportion

- In this scenario, we want to estimate the proportion of population belonging to a particular category.
- Any individual in the population either belongs to the category of interest ("1"), or they don't ("0").
- Thus, we can think of each random variable X as a Bernoulli distribution with unknown parameter p
- We ultimately want to estimate and find a confidence interval for p.

Confidence Interval for p Cont.

$X_1, \ldots, X_n \overset{iid}{\sim} Bern(p)$

First, estimate p using the *statistic* $\hat{p} = \frac{\sum X_i}{n}$ = sample proportion.

- $E(\hat{p}) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n} np = p$ (unbiased)
- $Var(\hat{p}) = Var\left(\frac{\sum X_i}{n}\right) = \frac{1}{n^2} Var\left(\sum X_i\right) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$

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

Since $\hat{p}$ is the mean of the Bernoulli $X's$, CLT for means applies

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

Thus a $(1-\alpha)100\%$ confidence interval for p is

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

Example

Example 3: In a random sample of 1000 U.S. adults, 38.8% stated they believed in the existence of ghosts. Find a 90% confidence interval for the population proportion of all U.S. adults who believe in the existence of ghosts.

Selecting Sample Size for Proportions

How to decide sample size?

- Just as before, we can select the sample size based on how large we want our margin or error to be
  margin or error $= \Delta = z_{\alpha/2} SE(\hat{p}) = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
- Issue: We haven't taken the sample, so we don't know $\hat{p}$
- Solution: Use $\hat{p} = 0.5$ for most conservative sample size
  $\Delta = z_{\alpha/2}\sqrt{\frac{0.5 \cdot 0.5}{n}} = z_{\alpha/2}\sqrt{\frac{0.5^2}{n}}$

To attain a particular margin of error Δ, we need a sample size

$$n \geq \left(\frac{z_{\alpha/2} \cdot 0.5}{\Delta}\right)^2$$

Sample Size Calculation Cont.

Example 4: Political polls typically use 95% confidence and report margin of errors of 3%: $\hat{p} \pm 0.03$.

What sample size do we need to for such a poll?