7.3. Intervals Based on a Normal Population Distribution

MATH 365

Table of contents

Topics for Today

- Review: Large Sample Confidence Intervals
- The t Distribution: Why and When
- Confidence Interval for μ with Small Samples
- Prediction Intervals
- Tolerance Intervals (brief intro)
- Applications: Reaction times, Fuel efficiency , and more!

Review: Large Sample Intervals (Section 7.2)

> **Quick Discussion**
>
> - What assumptions underlie the large-sample z-based confidence intervals?
> - Why is the Central Limit Theorem so important?

Formula Recall:

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

Take a minute to sketch a rough diagram showing the sampling distribution of $\bar{X}$ for large n.

The t Distribution

Assumptions

- $X_1, X_2, \ldots, X_n$ are a random sample from a normal population.
- Population mean μ unknown.
- Population standard deviation σ unknown.

Theorem

The statistic

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

follows a t distribution with $n-1$ degrees of freedom ($\nu = n-1$).

Properties of t Distributions

- Symmetric, bell-shaped
- More spread out than z for small n
- As $n \to \infty$, t approaches z

> Reflection
>
> Think about why using S (instead of σ) introduces extra variability!

One-Sample t Confidence Interval for μ

If $\bar{x}$ and s are the sample mean and sample standard deviation:

$$\bar{x} \pm t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}$$

where $t_{\alpha/2,n-1}$ is the critical t value.

Example: Reaction Times at Kalamazoo College

Suppose we collect reaction times (in milliseconds) from 12 students taking 8:15 AM classes. Data are obtained from this site.

Sample Mean: __________

Sample Standard Deviation: __________

$n = 12$

Task:

- Construct a 95% confidence interval for the average reaction time.

Work below:

Example: Fuel Efficiency of 2012 Prius

User-reported MPG from 14 drivers:

- Sample mean = 53.3 MPG
- Sample standard deviation = 5.2 MPG

Part (a)

Is it reasonable to use this data to estimate the average MPG for all 2012 Prius drivers?

Write your reasoning:

Part (b)

Calculate a 95% confidence interval for the mean MPG.

Work below:

Prediction Interval for a Single Future Observation

To predict a future value:

$$\bar{x} \pm t_{\alpha/2,n-1} \cdot s\sqrt{1 + \frac{1}{n}}$$

Example: Predicting MPG for a Random New 2012 Prius Driver

Use the data above (Prius example)!

Work out a 95% prediction interval:

Discussion: Why is the prediction interval wider than the confidence interval?

Another Application: Material Strength

You are testing the tensile strength (in MPa) of a new batch of material samples:

- Sample mean = 42.5 MPa
- Sample standard deviation = 3.8 MPa
- $n = 10$

Task:

- Construct a 90% confidence interval for the true mean tensile strength.

Work here:

- Construct a 90% prediction interval for a single future sample.

Work here:

In-Class Investigation: Robustness of t

> **Tip**
>
> **Experiment in R:** Let's investigate how the t-interval behaves under non-normal distributions and different n.

```r
set.seed(1234)

# Function to simulate confidence interval coverage
simulate_t_coverage <- function(n, dist_fn, n_sim = 5000, alpha = 0.05) {
  coverages <- replicate(n_sim, {
    sample <- dist_fn(n)
    ci_lower <- mean(sample) - qt(1 - alpha/2, df=n-1) * sd(sample)/sqrt(n)
    ci_upper <- mean(sample) + qt(1 - alpha/2, df=n-1) * sd(sample)/sqrt(n)
    0 >= ci_lower & 0 <= ci_upper
  })
  mean(coverages)
}

# Normal distribution
simulate_t_coverage(10, function(n) rnorm(n, mean=0, sd=1))

# Exponential distribution
simulate_t_coverage(10, function(n) rexp(n, rate=1) - 1)

# t-distribution with 2 df (heavy tails)
simulate_t_coverage(10, function(n) rt(n, df=2))

# Try larger n (n=30)
simulate_t_coverage(30, function(n) rexp(n, rate=1) - 1)
```

Notes on robustness:

Summary Points

- The t distribution adjusts for uncertainty in estimating σ.
- Use t for small samples from normal populations.

- Confidence intervals estimate the mean.
- Prediction intervals predict new individual values.
- Wider intervals reflect greater uncertainty.

Closing Thought

In what types of real-world situations might normality be a poor assumption, and how would you check?

Write a few ideas here to share next class: