

STAT 4352 - Mathematical Statistics Notes

JaimeGoB

February 28, 2021

1 Chapter 11 - Interval Estimation

Point Estimators

θ is a unknown parameter (feature of a population)

- Ex: population mean μ
- **Fixed.**

$\hat{\theta}$ is a point estimator of θ (it is a numerical value)

- Ex: sample mean \bar{x}
- **Varies from sample to sample.**
- No guarantee of accuracy
- Must be *supplemented by* $\text{Var}(\theta)$
Standard Error $\text{SE}(\hat{\theta})$ measures how much $\hat{\theta}$ varies from sample to sample.
small SE \implies low variance thus a more reliable estimate of θ

Interval Estimators

Def: Interval Estimate

Provides a range of values that best describe the population.

Let $L = L(x)$ be the Lower Limit

$U = U(x)$ be the Upper Limit

Both L, U are Random Variables because they are functions of sample data.

Def: Confidence Level / Confidence Coefficient

Is the probability that the **interval estimate** will include population parameter θ .

- Sample means will follow the normal probability distribution for large sample sizes ($n \geq 30$)
- For small sample forces us to use the t-distribution probability distribution ($n < 30$)
- A confidence level of 95% implies that **95% of all samples would give an interval that includes θ , and only 5% of all samples would yield an erroneous interval.**
- The most frequently used confidence levels are 90%, 95%, and 99% with corresponding Z-scores 1.645, 1.96, 2.576.
- The higher the confidence level, the more strongly we believe that the value of the parameter lies within the interval.

Def: Confidence Interval

Gives plausible values for the parameter θ being estimated where degree of plausibility specified by a confidence level.

To construct an interval estimator of unknown parameter θ . We must find two statistics **L** and **U** such that:

$$P\{\mathbf{L} \leq \theta \leq \mathbf{U}\} = 1 - \alpha$$

- $P\{\mathbf{L} \leq \theta \leq \mathbf{U}\}$ **Coverage Probability**, in repeated sampling, what percent of samples or Confidence Intervals capture true θ .
- $100(1 - \alpha)$ **Confidence Interval** - for unknown fixed parameter θ .
- **L, U - Lower and Upper Bounds** - RVs because they are functions of sample data. Vary from sample to sample.
- $1 - \alpha$ **Confidence Level** (Probability) estimate will include population parameter θ .
- α **Level of Significance** Percent chance Confidence Interval will not contain population parameter θ .

Def: Coverage Probability

$P\{\mathbf{L} \leq \theta \leq \mathbf{U}\}$ Gives what % of samples or Confidence Intervals capture true θ .

Ex: Coverage Probability = 95%

Will capture θ , 95% of the time.

Will NOT capture θ , 5% of the time.

Properties of Confidence Intervals

- Confidence Intervals are not unique.
- Desirable to have $E[\text{Length of CI}]$ to be small.
- A one-sided $100(1 - \alpha)$ lower-confidence interval on θ : $L = -\infty \implies P\{L \leq \theta\} = 1 - \alpha$
- A one-sided $100(1 - \alpha)$ upper-confidence interval on θ : $U = \infty \implies P\{\theta \leq U\} = 1 - \alpha$
- If **L, U** are both finite, then we have a two sided interval.

Correctly Interpreting Confidence Intervals**Not Correct**

There is 90% probability that the true population mean is within the interval.

Correct

There is a 90% probability that any given Confidence Interval from a random sample will contain the true population mean.

Theorem 11.1: Confidence Interval on the Mean of a Normal Distribution with known Variance

Let X be normal random variable with:

Unknown mean μ

Known variance σ^2

Suppose a random sample n , (X_1, X_2, \dots, X_n) is taken.

A $100(1-\alpha)\%$ confidence interval on μ can be obtained by considering sampling distribution of the sample mean \bar{X} .

Central Limit Theorem:

$E(\bar{X}) = \mu$ and $SD(\bar{X}) = \frac{\sigma}{\sqrt{n}}$, so $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$ as $n \rightarrow \infty$

Standardizing \bar{X} , Z will be a Standard Normal Distribution

Let $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$