

Lecture 10

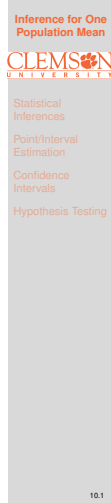
Inference for One Population Mean

Text: Chapter 5

STAT 8010 Statistical Methods I

February 18, 2020

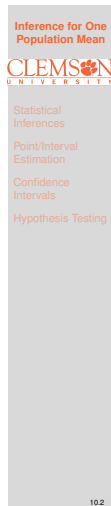
Whitney Huang
Clemson University



Notes

Agenda

- 1 Statistical Inferences
- 2 Point/Interval Estimation
- 3 Confidence Intervals
- 4 Hypothesis Testing

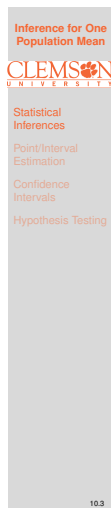


Notes

Statistical Inference

For the rest of the semester, we will focus on conducting **statistical inferences** for the following tasks:

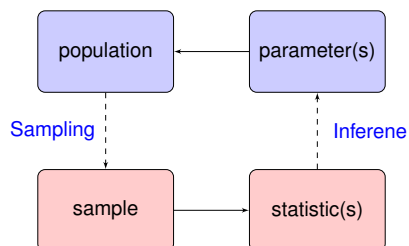
- Estimating one population mean
- Comparing two population means
- Comparing more than two population means
- Estimating population proportions
- Estimating relationship between two quantitative variables



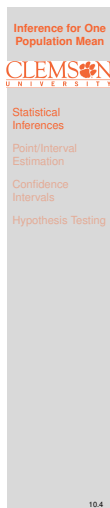
Notes

Statistical Inference Cont'd

- We use **parameters** to describe the population
Example: population mean (μ_X); population variance (σ_X^2)



- We use **statistics** of a sample to infer the population
Example: sample mean (\bar{X}); sample variance (s_X^2)

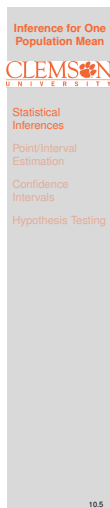


Notes

Estimating Population Mean μ

Goal: To estimate the population mean using a (representative) sample:

- The sample mean, $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$, is a reasonable **point estimate** of the population mean μ_X
- Need to quantify the level of uncertainty of the point estimate \Rightarrow **Interval estimation**
- Need to figure out the **sampling distribution** of \bar{X}_n in order to construct interval estimates \Rightarrow Central Limit Theorem (CLT)



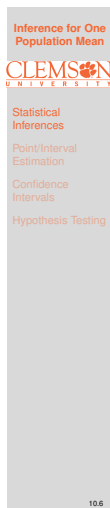
Notes

Central Limit Theorem (CLT)

CLT

The **sampling distribution** of \bar{X}_n will become approximately **normally distributed** as the **sample size** (n) becomes "large", **regardless of the shape of the population distribution!**

Let X_1, X_2, \dots, X_n be a random sample from a population X with $\mu = \mathbb{E}[X]$ and $\sigma^2 = \text{Var}[X]$. Then $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n} \xrightarrow{d} N(\mu, \frac{\sigma^2}{n})$ as $n \rightarrow \infty$.



Notes

CLT In Action

- ➊ Generate 100 (n) random numbers from an Exponential distribution (population distribution)
- ➋ Compute the **sample mean** of these 100 random numbers
- ➌ Repeat this process 120 times

Inference for One Population Mean

CLEMSON UNIVERSITY

Statistical Inferences

Point/Interval Estimation

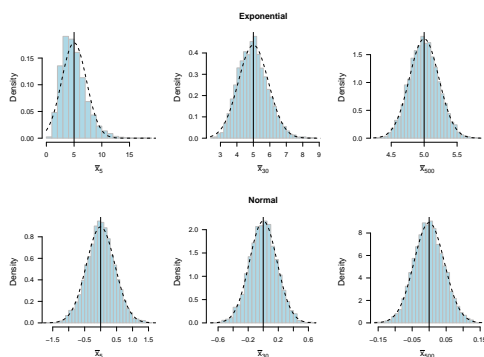
Confidence Intervals

Hypothesis Testing

10.7

Notes

CLT: Sample Size (n) and the Normal Approximation



Inference for One Population Mean

CLEMSON UNIVERSITY

Statistical Inferences

Point/Interval Estimation

Confidence Intervals

Hypothesis Testing

10.8

Notes

Why CLT is important?

- CLT tells us the **distribution** of our estimator

$$\bar{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

- The distribution of \bar{X}_n is center around the true mean μ
- The variance of \bar{X}_n is decrease with n
- With normality approximation of the sampling distribution of \bar{X}_n , we can perform interval estimation about μ
- Applications: **Confidence Interval**, **Hypothesis testing**

Inference for One Population Mean

CLEMSON UNIVERSITY

Statistical Inferences

Point/Interval Estimation

Confidence Intervals

Hypothesis Testing

10.9

Notes

Confidence Intervals (CIs) for μ

- Let's assume we know the population variance σ^2 (will relax this assumption later on)

- $(1 - \alpha) \times 100\%$ CI for μ :

$$\left[\bar{X}_n - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right],$$

where $z_{\frac{\alpha}{2}}$ is the $1 - \frac{\alpha}{2}$ percentile of $Z \sim N(0, 1)$

- $\frac{\sigma}{\sqrt{n}}$ is the **standard error** of \bar{X}_n , that is, the standard deviation of its sampling distribution

Inference for One Population Mean

CLEMSON UNIVERSITY

Statistical Inferences

Point/Interval Estimation

Confidence Intervals

Hypothesis Testing

10.10

Notes

Making Sense of Confidence Intervals

For any $\alpha \in (0, 1)$:

$$\begin{aligned} &\mathbb{P}\left(\bar{X}_n - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) \\ &= \mathbb{P}\left(-z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \bar{X}_n - \mu \leq z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) \\ &= \mathbb{P}\left(-z_{\frac{\alpha}{2}} \leq \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\frac{\alpha}{2}}\right) \\ &= \mathbb{P}\left(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}\right) \\ &= \Phi(z_{\frac{\alpha}{2}}) - \Phi(-z_{\frac{\alpha}{2}}) \\ &= 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha \end{aligned}$$

Inference for One Population Mean

CLEMSON UNIVERSITY

Statistical Inferences

Point/Interval Estimation

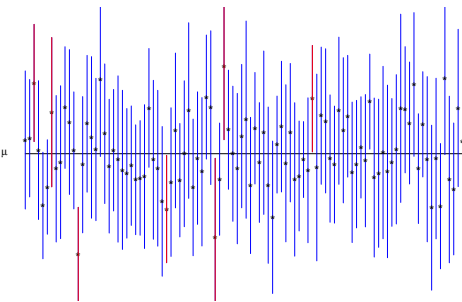
Confidence Intervals

Hypothesis Testing

10.11

Notes

Making Sense of Confidence Intervals Cont'd



Inference for One Population Mean

CLEMSON UNIVERSITY

Statistical Inferences

Point/Interval Estimation

Confidence Intervals

Hypothesis Testing

10.12

Notes

Example: Average Height

We measure the heights of 40 randomly chosen men, and get a mean height of 5'9" ($\approx 175\text{cm}$). Suppose we know the standard deviation of men's heights is 4" ($\approx 10\text{cm}$). Find the 95% confidence interval of the true mean height of ALL men.

WORLD HEIGHT CHART(MALE)



Notes

Average Height Example Cont'd

- 1 Point estimate: $\bar{X}_{40} = \frac{\sum_{i=1}^{40} X_i}{40} = 69$ inches
- 2 Population standard deviation: $\sigma = 4$ inches
- 3 Standard error of $\bar{X}_{n=40} = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{40}} = 0.63$ inches
- 4 95%CI: Need to find $z_{0.05/2} = 1.96$ from the Z-table
- 5 95% CI for μ_X is:

$$[69 - 1.96 \times 0.63, 69 + 1.96 \times 0.63] \\ = [67.77, 70.23]$$

Notes

Properties of Confidence Intervals

- In contrast with the point estimate, \bar{X}_n , a $(1 - \alpha)\%$ CI is an **interval estimate**, where the **length of CI** reflects our estimation uncertainty
- Typical α values: $0.01, 0.05, 0.1 \Rightarrow 99\%, 95\%, 90\%$ confidence intervals. **Interpretation:** If we were to take random samples over and over again, then $(1 - \alpha)\%$ of these confidence intervals will contain the true μ
- The length of a CI depends on
 - Population Standard Deviation: σ
 - Confidence Level: $1 - \alpha$
 - Sample Size: n

Notes

Sample Size Calculation

- We may want to estimate μ with a confidence interval with a predetermined margin of error (i.e. $z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$)
- For example, in estimating the true mean height of All men we may want our CI to be just 0.5 inches in width
- The question is then, “**how many observations do we need to take** so that we have the desired margin of error?”

Inference for One Population Mean

CLEMSON UNIVERSITY

Statistical Inferences

Point/Interval Estimation

Confidence Intervals

Hypothesis Testing

10.16

Notes

Sample Size Calculation Cont'd

To compute the sample size needed to get a CI for μ with a specified margin of error, we use the formula below

$$n = \left(\frac{z_{\alpha/2} \times \sigma}{\text{Margin of error}} \right)^2$$

Exercise: Derive this formula using margin of error $= z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

Inference for One Population Mean

CLEMSON UNIVERSITY

Statistical Inferences

Point/Interval Estimation

Confidence Intervals

Hypothesis Testing

10.17

Notes

Average Height Example Revisited

Compute the sample size needed in order to estimate the true mean height of All men such that the 95% CI to be 0.5 inches in width

- ➊ Length of CI: $2 \times z_{0.025} \frac{\sigma}{\sqrt{n}} = 2 \times \text{margin of error}$
- ➋ Want to find n s.t. $z_{0.025} \frac{\sigma}{\sqrt{n}} = 0.25$
- ➌ We have $n = \left(\frac{1.96 \times 4}{0.25} \right)^2 = 983.4496$

Therefore, the required sample size is 984

Inference for One Population Mean

CLEMSON UNIVERSITY

Statistical Inferences

Point/Interval Estimation

Confidence Intervals

Hypothesis Testing

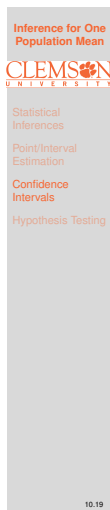
10.18

Notes

Confidence Intervals When σ Unknown

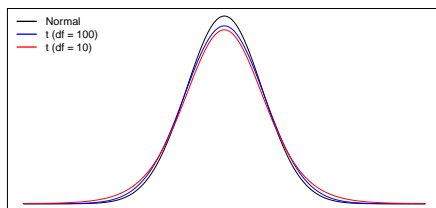
- In practice, it is unlikely that σ is available to us
- One reasonable option is to replace σ with s , the sample standard deviation
- We need to account for this added uncertainty with a (slightly) different sampling distribution that has fatter tails

⇒ Student t Distribution (William Gosset, 1908)

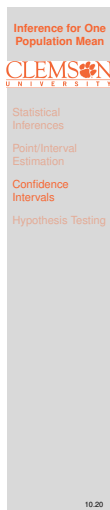


Notes

Student t Distribution



- Recall the standardize sampling distribution
 $\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$
- Similarly, the studentized sampling distribution
 $\frac{\bar{X}_n - \mu}{\frac{s}{\sqrt{n}}} \sim t_{df=n-1}$



Notes

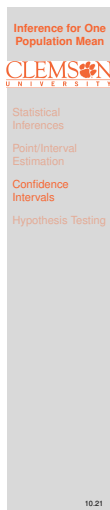
Confidence Intervals (CIs) for μ When σ is Unknown

- $(1 - \alpha) \times 100\%$ CI for μ :

$$\left[\bar{X}_n - t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}, \bar{X}_n + t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}} \right],$$

where $t_{\frac{\alpha}{2}, n-1}$ is the $1 - \frac{\alpha}{2}$ percentile of a student t distribution with the degrees of freedom = $n - 1$

- $\frac{s}{\sqrt{n}}$ is an estimate of the standard error of \bar{X}_n



Notes

Average Height Example Revisited

We measure the heights of 40 randomly chosen men, and get a mean height of 5'9" ($\approx 175\text{cm}$), and a standard deviation of 4.5" ($\approx 11.4\text{cm}$). Find the 95% confidence interval of the true mean height of ALL men.

Inference for One Population Mean

CLEMSON UNIVERSITY

Statistical Inferences

Point/Interval Estimation

Confidence Intervals

Hypothesis Testing

10.22

Notes

Average Height Example Cont'd

- ➊ Point estimate: $\bar{X}_{40} = \frac{\sum_{i=1}^{40} X_i}{40} = 69$ inches
- ➋ Sample standard deviation: $s = 4.5$ inches
- ➌ (Estimated) standard error of $\bar{X}_{n=40} = \frac{s}{\sqrt{n}} = \frac{4.5}{\sqrt{40}} = 0.71$ inches
- ➍ 95%CI: Need to find $t_{0.05/2,39} = 2.02$ from a t-table (or using a statistical software)
- ➎ 95% CI for μ_X is:

$$[69 - 2.02 \times 0.71, 69 + 2.02 \times 0.71]$$
$$= [67.57, 70.43]$$

Inference for One Population Mean

CLEMSON UNIVERSITY

Statistical Inferences

Point/Interval Estimation

Confidence Intervals

Hypothesis Testing

10.23

Notes

Hypothesis Testing

- ➊ **Hypothesis Testing:** A method for using sample data to decide between two competing claims (hypotheses) about a population characteristic (a parameter. e.g. μ)
- ➋ **Examples:**
 - ➊ The true mean starting salary for graduates of four-year business schools is \$4,500 per month $\Rightarrow \mu = 4,500$
 - ➋ The true mean monthly income for systems analysts is at least \$6,000 $\Rightarrow \mu \geq 6,000$

Inference for One Population Mean

CLEMSON UNIVERSITY

Statistical Inferences

Point/Interval Estimation

Confidence Intervals

Hypothesis Testing

10.24

Notes

Hypotheses

- **Null Hypothesis:** A claim about a parameter that is initially assumed to be true. We use H_0 to denote a null hypothesis
- **Alternative Hypothesis:** The competing claim, denoted by H_a
- In carrying out a test of H_0 versus H_a , the hypothesis H_0 will be rejected in favor of H_a only if sample evidence strongly suggests that H_0 is false. If the sample data does not contain such evidence, H_0 will not be rejected
- Therefore, the two possible decisions in a hypothesis test are:
 - Reject H_0 (and go with H_a)
 - Fail to Reject H_0

Inference for One Population Mean

CLEMSON UNIVERSITY

Statistical Inferences

Point/Interval Estimation

Confidence Intervals

Hypothesis Testing

10.25

Notes

Hypotheses

Be careful setting up hypotheses:

- A statistical hypothesis test is only capable of demonstrating strong support for the alternative hypothesis H_a (by rejecting the null hypothesis H_0)
- Failing to reject H_0 does not show strong support for the null hypothesis – **only a lack of strong evidence against H_0 , the null hypothesis**

Inference for One Population Mean

CLEMSON UNIVERSITY

Statistical Inferences

Point/Interval Estimation

Confidence Intervals

Hypothesis Testing

10.26

Notes

The 2 × 2 Decision Paradigm for Hypothesis Testing

True State	Decision	
	Reject H_0	Fail to reject H_0
H_0 is true	Type I error	Correct
H_0 is false	Correct	Type II error

Errors in Hypothesis Testing

- The probability of a **type I error** is denoted by α
- The probability of a **type II error** is denoted by β

Inference for One Population Mean

CLEMSON UNIVERSITY

Statistical Inferences

Point/Interval Estimation

Confidence Intervals

Hypothesis Testing

10.27

Notes
