Lecture 12

Inference for One Population Mean

Readings: IntroStat Chapter 5; OpenIntro Chapter 7.1

STAT 8010 Statistical Methods I
June 1, 2023

Whitney Huang
Clemson University

Agenda

1. Statistical Inferences
2. Point/Interval Estimation
3. Confidence Intervals

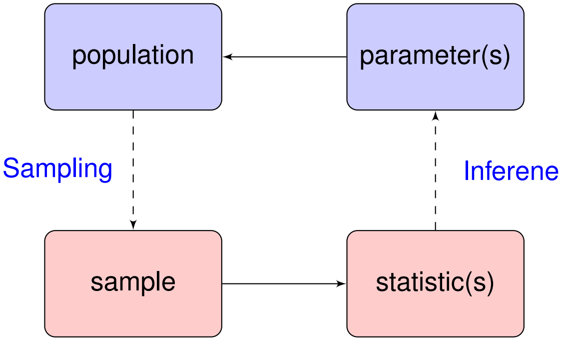
Statistical Inference

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

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)

Estimating Population Mean μ

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

- The sample mean, $\bar{X}_n = \frac{\sum_i^n X_i}{n}$, is a reasonable point estimate of the population mean μ_X

Estimating Population Mean μ

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

- The sample mean, $\bar{X}_n = \frac{\sum_i^n X_i}{n}$, 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

Estimating Population Mean μ

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

- The sample mean, $\bar{X}_n = \frac{\sum_i^n X_i}{n}$, 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)

Why CLT is important?

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

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

  - The distribution of $\bar{X}_n$ is center around the true mean μ
  - The variance of $\bar{X}_n$ is decrease with n
  - With normal approximation of the sampling distribution of $\bar{X}_n$, we can perform interval estimation about μ

- Applications: Confidence Interval, Hypothesis testing

Confidence Intervals (CIs) for μ

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

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 $Z \sim \mathrm{N}(0,1)$

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 $Z \sim \mathrm{N}(0,1)$

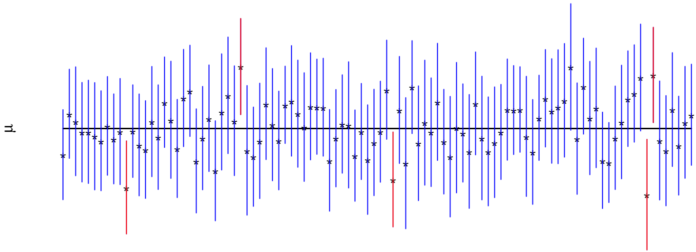
- $\frac{\sigma}{\sqrt{n}}$ is the standard error of $\bar{X}_n$, that is, the standard deviation of its sampling distribution

Making Sense of Confidence Intervals

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

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

Making Sense of Confidence Intervals Cont'd



Example: Average Height

We measure the heights of 40 randomly chosen men, and get a mean height of 5’9” (≈175cm). Suppose we know the standard deviation of men’s heights is 4” (≈10cm). Find the 95% confidence interval of the true mean height of ALL men.

Average Height Example Cont'd

1. Point estimate: $\bar{X}_{40} = \frac{\sum_{i=1}^{40} X_i}{40} = 69$ inches

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

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

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

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:

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

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

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

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

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

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: $1-\alpha$

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

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?"

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 $= z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

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

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

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

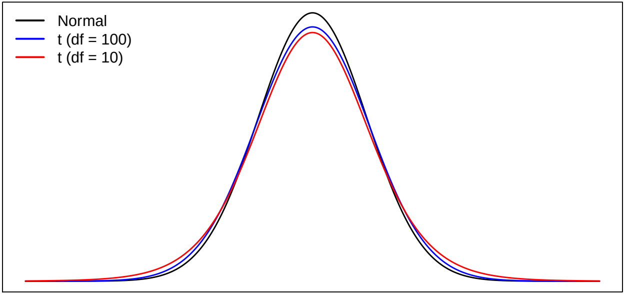
Therefore, the required sample size is 984

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

$\Rightarrow$ Student's t Distribution (William Gosset, 1908)

Student's t Distribution



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

Confidence Intervals (CIs) for μ When σ is Unknown

- $(1-\alpha)\times 100\%$ CI for μ:

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

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

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

Average Height Example Revisited

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

Average Height Example Cont'd

1. Point estimate: $\bar{X}_{40} = \frac{\sum_{i=1}^{40} X_i}{40} = 69$ inches

Average Height Example Cont'd

1. Point estimate: $\bar{X}_{40} = \frac{\sum_{i=1}^{40} X_i}{40} = 69$ inches
2. Sample standard deviation: $s = 4.5$ inches

Average Height Example Cont'd

1. Point estimate: $\bar{X}_{40} = \frac{\sum_{i=1}^{40} X_i}{40} = 69$ inches
2. Sample standard deviation: $s = 4.5$ inches
3. (Estimated) standard error of $\bar{X}_{n=40} = \frac{s_n}{\sqrt{n}} = \frac{4.5}{\sqrt{40}} = 0.71$ inches

Average Height Example Cont'd

1. Point estimate: $\bar{X}_{40} = \frac{\sum_{i=1}^{40} X_i}{40} = 69$ inches

2. Sample standard deviation: $s = 4.5$ inches

3. (Estimated) standard error of $\bar{X}_{n=40} = \frac{s_n}{\sqrt{n}} = \frac{4.5}{\sqrt{40}} = 0.71$ inches

4. 95%CI: Need to find $t_{0.05/2,39} = 2.02$ from a t-table (or using a statistical software)

Average Height Example Cont'd

1. Point estimate: $\bar{X}_{40} = \frac{\sum_{i=1}^{40} X_i}{40} = 69$ inches
2. Sample standard deviation: $s = 4.5$ inches
3. (Estimated) standard error of $\bar{X}_{n=40} = \frac{s_n}{\sqrt{n}} = \frac{4.5}{\sqrt{40}} = 0.71$ inches
4. 95%CI: Need to find $t_{0.05/2,39} = 2.02$ from a t-table (or using a statistical software)
5. 95% CI for μ_X is:

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

Summary

In this lecture, we learned

- Statistical Inferences
- Point and interval estimation
- Confidence Intervals