Lecture 13

Confidence Intervals & Hypothesis Testing

Text: Chapter 5

STAT 8010 Statistical Methods I

October 6, 2020

Whitney Huang
Clemson University

Agenda

1. **Confidence Intervals**
2. **Hypothesis Testing**
3. **Type I & Type II Errors**
4. **Duality of Hypothesis Test with Confidence Interval**

Example: Average Height

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

CLEMSON UNIVERSITY

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:

$$\begin{aligned} &[69 - 1.96 \times 0.63, 69 + 1.96 \times 0.63] \\ &= [67.77, 70.23] \end{aligned}$$

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 $n = \left(\frac{1.96\times 4}{0.25}\right)^2 = 983.4496$

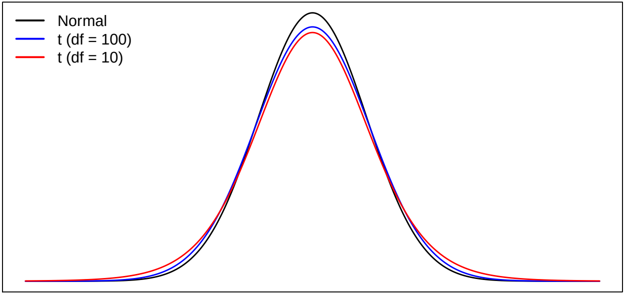
Therefore, the required sample size is 984

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 t Distribution (William Gosset, 1908)

Student 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 $\frac{\bar{X}_n - \mu}{\frac{s}{\sqrt{n}}} \sim \mathrm{t}_{df=n-1}$

CLEMSON UNIVERSITY

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 $\bar{X}_n$

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 $\bar{X}_{n=40} = \frac{s}{\sqrt{n}} = \frac{4.5}{\sqrt{40}} = 0.71$ 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 $\bar{X}_{n=40} = \frac{s}{\sqrt{n}} = \frac{4.5}{\sqrt{40}} = 0.71$ 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 $\bar{X}_{n=40} = \frac{s}{\sqrt{n}} = \frac{4.5}{\sqrt{40}} = 0.71$ 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:

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

Hypothesis Testing

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

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

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

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

Test Statistics

- In a hypothesis test, our “evidence” comes in the form of a test statistic
- A test statistic incorporates a number of aspects of the sample: the **sample size**, the **point estimate**, the **standard deviation**, and the **hypothesized value**
- If we’re conducting a hypothesis test about μ (assuming we don’t know σ) we would use the following test statistic:

$$t^* = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

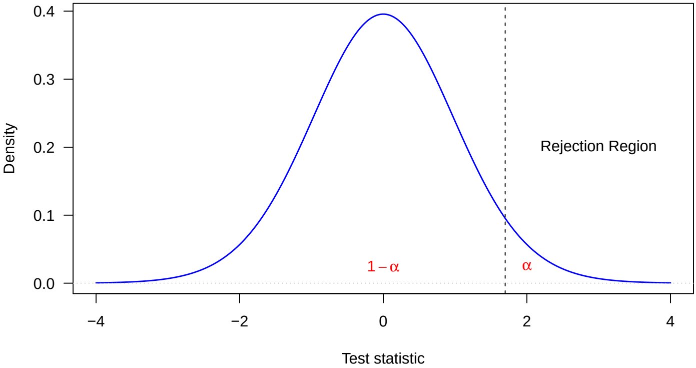
If $\mu = \mu_0$, we have $t^* \sim t_{df=n-1}$

Decision-Making: Rejection Region and P-Value Methods

- Decision based on t^*, H_a, and α, the **significant level**, that is pre-defined by the researcher
- Two approaches:
  - Rejection Region Method: reject H_0 if t^* is in the rejection region, otherwise fail to reject H_0
  - P-Value Method: reject H_0 if P-value is less than α, otherwise fail to reject H_0
- **Question:** How to determine the rejection region and how to compute P-value?

Rejection Region Method

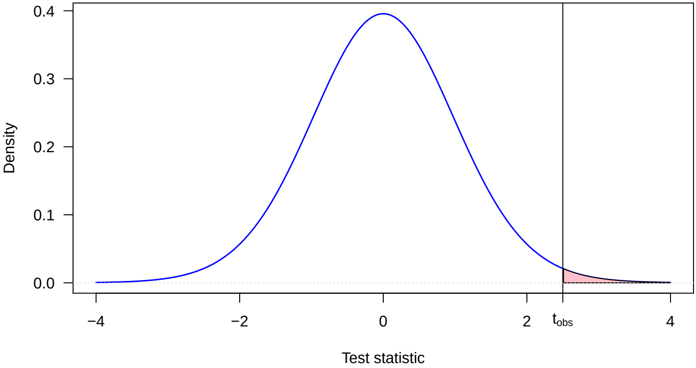
Let $H_0 : \mu = \mu_0$ vs. $H_a : \mu > \mu_0$ and $\alpha = 0.05$



Under the H_0, the test statistic $t^* = \frac{\bar{X}-\mu_0}{\frac{s}{\sqrt{n}}} \sim t_{df=n-1}$. The cutoff of the rejection region (=$t_{0.05,n-1}$) can be found from a t-table

P-Value Method

Let $H_0 : \mu = \mu_0$ vs. $H_a : \mu > \mu_0$



P-value: the probability of getting a test statistic that is at least as extreme as the one we actually observed **if the null hypothesis is true** $\Rightarrow \mathbb{P}(t^* \geq t_{obs})$

Draw a Conclusion

Use the following “generic” conclusion:

> “We (do/do not) have enough statistical evidence to conclude that (H_a in words) at α% significant level.”

- Reject $H_0 \Leftrightarrow$ do
- Fail to reject $H_0 \Leftrightarrow$ do not

CLEMSON UNIVERSITY

Example (taken from The Cartoon Guide To Statistics)

New Age Granola Inc claims that average weight of its cereal boxes is 16 oz. The Genuine Grocery Corporation will send back a shipment if the average weight is any less.

Suppose Genuine Grocery Corporation takes a random sample of 49 boxes, weight each one, and compute the sample mean $\bar{X} = 15.90$ oz and sample standard deviation $s = 0.35$ oz.

Perform a hypothesis test at 0.05 significant level to determine if they would reject H_0, and therefore, this shipment

Cereal Weight Example Cont'd

1. $H_0 : \mu = 16$ vs. $H_a : \mu < 16$

Cereal Weight Example Cont'd

1. $H_0 : \mu = 16$ vs. $H_a : \mu < 16$
2. Test Statistic: $t_{obs} = \frac{15.9-16}{0.35/\sqrt{49}} = -2$

Cereal Weight Example Cont'd

1. $H_0 : \mu = 16$ vs. $H_a : \mu < 16$
2. Test Statistic: $t_{obs} = \frac{15.9-16}{0.35/\sqrt{49}} = -2$
3. Rejection Region Method: $-t_{0.05,48} = -1.68 \Rightarrow$ Rejection Region is $(-\infty, -1.68]$. Since t_{obs} is in rejection region, we reject H_0

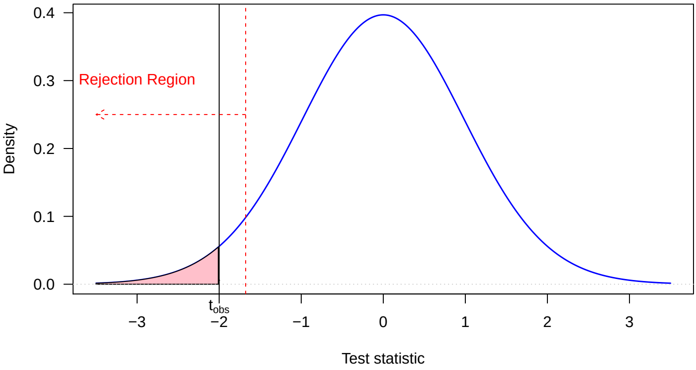
Cereal Weight Example Cont'd

1. $H_0 : \mu = 16$ vs. $H_a : \mu < 16$
2. Test Statistic: $t_{obs} = \frac{15.9-16}{0.35/\sqrt{49}} = -2$
3. Rejection Region Method: $-t_{0.05,48} = -1.68 \Rightarrow$ Rejection Region is $(-\infty, -1.68]$. Since t_{obs} is in rejection region, we reject H_0
4. P-Value Method: $\mathbb{P}(t^* \leq -2) = 0.0256 < \alpha = 0.05 \Rightarrow$ reject H_0

Cereal Weight Example Cont'd

1. $H_0 : \mu = 16$ vs. $H_a : \mu < 16$
2. Test Statistic: $t_{obs} = \frac{15.9-16}{0.35/\sqrt{49}} = -2$
3. Rejection Region Method: $-t_{0.05,48} = -1.68 \Rightarrow$ Rejection Region is $(-\infty, -1.68]$. Since t_{obs} is in rejection region, we reject H_0
4. P-Value Method: $\mathbb{P}(t^* \leq -2) = 0.0256 < \alpha = 0.05 \Rightarrow$ reject H_0
5. Draw a Conclusion: We do have enough statistical evidence to conclude that the average weight is less than 16 oz at 0.05 significant level

Cereal Weight Example Cont'd

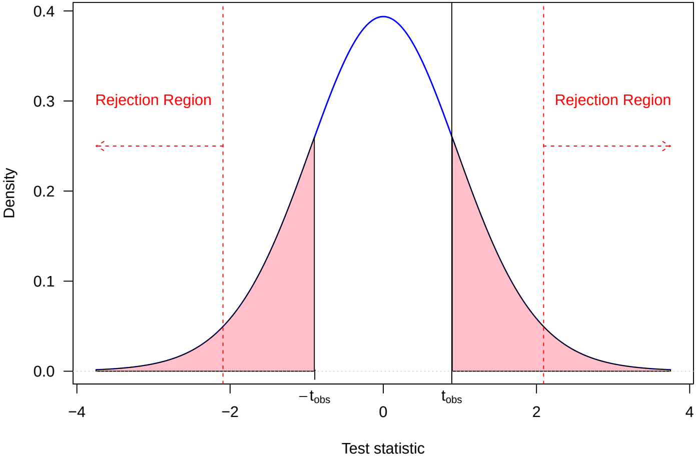


Example

A series of blood tests were run on a particular patient over five days. It is of interest to determine if the mean blood protein for this patient differs from 7.25, the value for healthy adults. Suppose the sample mean (n=20) is 7.35 and sample standard deviation is 0.5. Perform a hypothesis test using significance level of 0.05

1. $H_0 : \mu = 7.25$ vs. $H_a : \mu \neq 7.25$
2. $t_{obs} = \frac{7.35-7.25}{0.5/\sqrt{20}} = 0.8944$
3. P-value: $2 \times \mathbb{P}(t^* \geq 0.8944) = 0.3823 > 0.05$
4. We do not have enough statistical evidence to conclude that the mean blood protein is different from 7.25 at 5% significant level

Example Cont'd



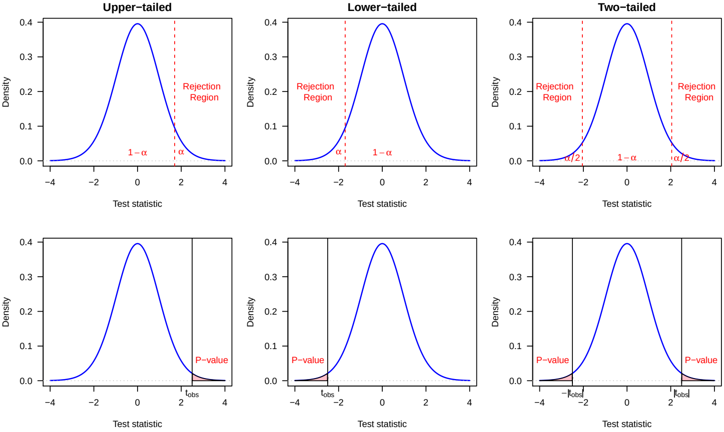
Recap: Hypothesis Testing

1. State the null H_0 and the alternative H_a hypotheses
   - $H_0 : \mu = \mu_0$ vs $H_a : \mu > \mu_0 \Rightarrow$ Upper-tailed
   - $H_0 : \mu = \mu_0$ vs $H_a : \mu < \mu_0 \Rightarrow$ Lower-tailed
   - $H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0 \Rightarrow$ Two-tailed
2. Compute the test statistic

   $t^* = \frac{\bar{X}_n - \mu_0}{s/\sqrt{n}}$ (σ unknown); $z^* = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}$ (σ known)
3. Identify the rejection region(s) (or compute the P-value)
4. Draw a conclusion

   We do/do not have enough statistical evidence to conclude H_a at α significant level

Region Region and P-Value Methods



The 2×2 Decision Paradigm for Hypothesis Testing

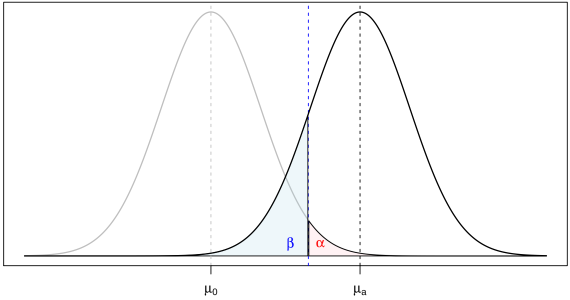
True State	Decision	
	Reject H_0	Fail to reject H_0
H_0 is true	Type I error	Correct
H_0 is false	Correct	Type II error

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

Type I & Type II Errors

- Type I error: $\mathbb{P}(\text{Reject } H_0 | H_0 \text{ is true}) = \alpha$
- Type II error: $\mathbb{P}(\text{Fail to reject } H_0 | H_0 \text{ is false}) = \beta$



$\alpha \downarrow \beta \uparrow$ and vice versa

Type II Error and Power

- The type II error, β, depends upon the true value of μ (let's call it μ_a)
- We use the formula below to compute β

$$\beta(\mu_a) = \mathbb{P}(z^* \leq z_\alpha - \frac{|\mu_0 - \mu_a|}{\sigma/\sqrt{n}})$$

- The power (PWR): $\mathbb{P}(\text{Reject } H_0 | H_0 \text{ is false}) = 1 - \beta$. Therefore $\text{PWR}(\mu_a) = 1 - \beta(\mu_a)$

Question: What increases Power?

Sample Size Determination

Suppose that we wish to determine what sample size is required to detect the difference between a hypothesized mean and true mean $\mu_0 - \mu_a$, denoted by Δ, with a given power $1 - \beta$ and specified significance level α and known standard deviation σ. We can use the following formulas

$$n = \sigma^2 \frac{(z_\alpha + z_\beta)^2}{\Delta^2} \text{ for a one-tailed test}$$

$$n \approx \sigma^2 \frac{(z_{\alpha/2} + z_\beta)^2}{\Delta^2} \text{ for a two-tailed test}$$

Example

An existing manufacturing process produces, on average, 100 units of output per day. A pilot plant is used to evaluate a possible process change. Suppose the Company CEO wants to know if yield is increased. The CEO uses $\alpha = 0.05$ and the sample mean $(n = 25)$ is 103. Do we have sufficient evidence to conclude that the mean yield exceeds 100 if $\sigma = 10$?

1. $H_0 : \mu = 100$ vs. $H_a : \mu > 100$
2. $z_{obs} = \frac{103-100}{10/\sqrt{25}} = 1.5$
3. The cutoff value of the rejection region is $z_{0.05} = 1.645$. Therefore we do not have enough evidence to conclude that the new process mean yield exceeds 100

Example Cont'd

Suppose the true true mean yield is 104.

- What is the power of the test?

Example Cont'd

Suppose the true true mean yield is 104.

- What is the power of the test?

$$
\begin{aligned}
\beta(\mu = 104) &= \mathbb{P}\left(Z \leq z_{0.05} - \frac{|100 - 104|}{10/\sqrt{25}}\right) \\
&= \mathbb{P}(Z \leq 1.645 - 4/2) = \mathbb{P}(Z \leq -0.355) \\
&= \Phi(-0.355) = 0.3613
\end{aligned}
$$

Therefore, the power is $1 - 0.3613 = 0.6387$

- What sample size is required to yield a power of 0.8 with a significance level of 0.05?

Example Cont'd

Suppose the true true mean yield is 104.

- What is the power of the test?

$$\begin{aligned}\beta(\mu = 104) &= \mathbb{P}\left(Z \le z_{0.05} - \frac{|100 - 104|}{10/\sqrt{25}}\right) \\ &= \mathbb{P}(Z \le 1.645 - 4/2) = \mathbb{P}(Z \le -0.355) \\ &= \Phi(-0.355) = 0.3613\end{aligned}$$

Therefore, the power is $1 - 0.3613 = 0.6387$

- What sample size is required to yield a power of 0.8 with a significance level of 0.05?

$$n = \sigma^2 \frac{(z_{0.05} + z_{0.2})^2}{\Delta^2} = 10^2 \frac{(1.645 + 0.8416)^2}{4^2} = 38.6324$$

Therefore, the required sample size is 39

Duality of Hypothesis Test with Confidence Interval

There is an interesting relationship between CIs and hypothesis tests. If H_0 is rejected with significance level α then the corresponding confidence interval does not contain the value μ_0 targeted in the hypotheses with the confidence level $(1-\alpha)$, and vice versa

Hypothesis test at α level	$(1-\alpha)\times$ 100% CI
$H_0 : \mu = \mu_0$ vs. $H_a : \mu \neq \mu_0$	$\bar{X} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}$
$H_0 : \mu = \mu_0$ vs. $H_a : \mu > \mu_0$	$\left(\bar{X} - t_{\alpha,n-1} s/\sqrt{n}, \infty\right)$
$H_0 : \mu = \mu_0$ vs. $H_a : \mu < \mu_0$	$\left(-\infty, \bar{X} + t_{\alpha,n-1)s/\sqrt{n}}\right)$