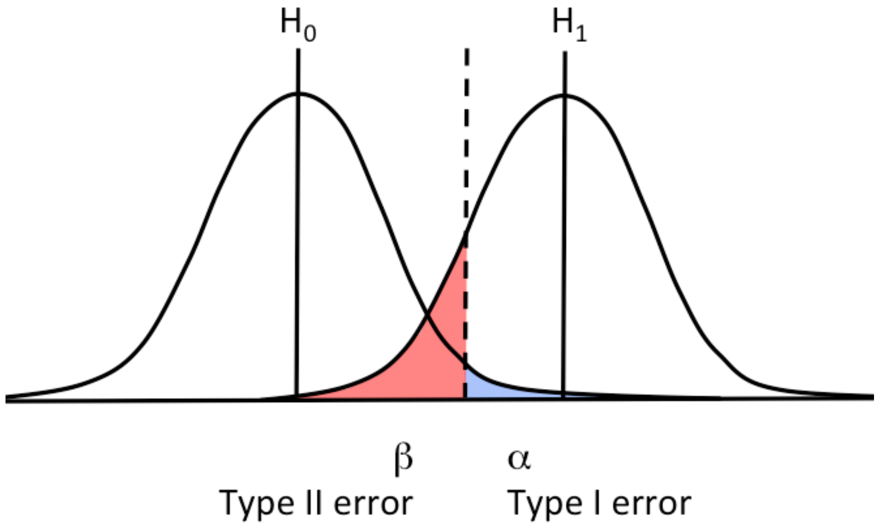
Chapter 7

Using Sample Statistics to Test Hypotheses about Population Parameters



Key Words:

- Null hypothesis H_0
- Alternative hypothesis H_A
- Testing hypothesis
- Type I error α
- Type II error β
- Power of the test $(1-\beta)$
- P-value
- Test statistic
- Testing hypotheses concerning one population mean
- Testing hypotheses concerning two population means
- Paired comparison

Learning Outcomes

- After studying this chapter, you will
- Understand how to correctly state a null and alternative hypothesis and carry out a structured hypothesis test.
- Understand the concepts of type I error, type II error, and the power of a test.
- Be able to calculate and interpret z and t test statistics for making statistical inferences.
- Understand how to calculate and interpret p values.

7.1 Introduction

Definition of a hypothesis

- It is a statement or claim or assertion about one or more populations. It is usually concerned with the parameters of the population.
- Examples:
  - The hospital administrator may want to test the hypothesis that the average length of stay of patients admitted to the hospital is 5 days.
  - A physician may hypothesize that a certain drug will be effective in 90% of cases for which it is used.

Types of Hypotheses

- They are two types of hypotheses:
- Research hypothesis is the conjecture or supposition that motivates the research.
- For example, a physician may recall numerous instants in which certain combination of therapeutic measures were more effective than any one of them alone.
- Statistical hypotheses are hypotheses that are stated in such a way that they may be evaluated by appropriate statistical techniques.

- There are two statistical hypotheses involved in hypothesis testing
  - **Null hypothesis** H_0: It is the hypothesis to be tested.
    - ✓ To conduct a hypothesis test about the mean of a population, we postulate a value for it, and call that value μ_0.
    - ✓ We write this $H_0: \mu = \mu_0$
  - **Alternative hypothesis** H_A: It is a statement of what we believe is true if our sample data cause us to reject the null hypothesis.
    - ✓ We set an alternative hypothesis for all other values of μ
    - ✓ We write this: $H_A: \mu \neq \mu_0$

	Condition of null hypothesis	
Result of the test	$\mu = \mu_0$ (H_0 is true)	$\mu \neq \mu_0$ (H_0 is False)
Reject H_0	**Type I error (α)**	**Correct Decision**
Do Not Reject H_0	**Correct Decision**	**Type II error (β)**

- ✓ α = Probability of Type I Error =**P(rejecting H_0 given H_0 is true)**
- ✓ =**Chance to miss right answer**
- ✓ β = Probability of Type II Error = **P(Accepting H_0 given H_A is true)**
- ✓ **(1- β)** = Power of a test = **P(Rejecting H_0 given H_A is true)**
- ✓ = **Ability to detect right answer**

Simple and composite hypotheses

$H_0: \mu = 70$ this is simple hypothesis

$H_0: \mu \leq 70$ this is composite hypothesis

$H_0: \mu \geq 70$ this is composite hypothesis

One sided and two sided hypotheses

$H_0: \mu = 70$ vs $H_A: \mu > 70$ one sided hypothesis (or one tailed)

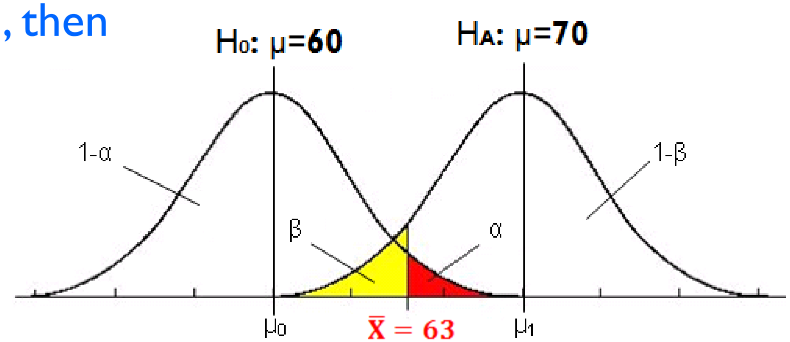
$H_0: \mu = 70$ vs $H_A: \mu < 70$ one sided hypothesis (or one tailed)

$H_0: \mu = 70$ vs $H_A: \mu \neq 70$ two sided hypothesis (or two tailed)

$H_0: \mu = 70$ vs $H_A: \mu > 70 \Leftrightarrow H_0: \mu \leq 70$ vs $H_A: \mu > 70$

$H_0: \mu = 50$ vs $H_A: \mu < 50 \Leftrightarrow H_0: \mu \geq 50$ vs $H_A: \mu < 50$

- Suppose we want to test
  - $H_0: \mu=60$ against $H_A: \mu=70$
  - If we decide to reject H_0 if $\bar{X} > 63$, then
  - $\alpha=P(\bar{X} > 63 | \mu=60)$
  - $\beta=P(\bar{X} < 63 | \mu=70)$
  - Power=1-β= β=P($\bar{X} > 63 | \mu=70)$



- **Type I and Type II errors**
  - Type I error: α is the specified significance level.
  - Type II: β generally **unspecified** and **unknown.**
  - **For a given n, α and β inversely related.**
  - Both types of errors may be reduced **simultaneously** by increasing n.
  - P-value is the smallest value of α for which we can reject a null hypothesis.
  - ***Reject H_0 if P-value less than α.***
    - If P-value<0.05, we say, that results can't be explained by chance alone, therefore we reject H_0 and accept H_A.
    - If P-value≥0.05, we say, that found difference can be due to chance alone, therefore we don't reject H_0.

Steps of Hypothesis Testing

1. State the statistical hypotheses, H_0 and H_A.
2. Select a level of significance, α.
3. Find the critical value(s).
4. Compute the test statistics.
5. Refer to a criterion for evaluating the sample evidence.
6. Make a decision to keep/reject the null.
7. Make your comments and conclusion.

Hypotheses Testing: A Single Population Mean

- A. **Sampling from normally distributed population with known σ (n small or large)**
- B. **Sampling from normally distributed population with unknown σ (n small)**
- C. **Sampling from nonormally distributed population (n large)**

7.2 Hypotheses Testing: A Single Population Mean

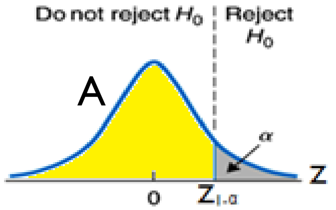
A: Sampling from normally distributed population with known σ (n small or large)

Null hypothesis $H_0: \mu = \mu_0$, α=level of significance

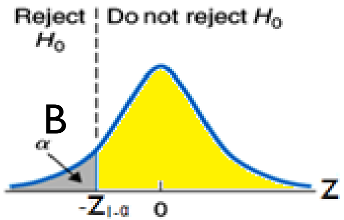
Test statistic:

$$Z_{cal} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

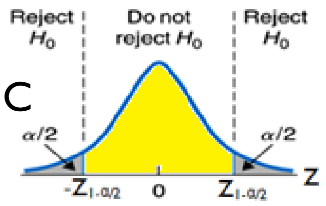
A. If Alternative $H_A: \mu > \mu_0$
   Reject H_0 if $Z_{cal} > Z_{1-\alpha}$
   P-Value= $P(Z > Z_{cal})$



B. If Alternative $H_A: \mu < \mu_0$,
   Reject H_0 if $Z_{cal} < -Z_{1-\alpha}$
   P-Value= $P(Z < Z_{cal})$

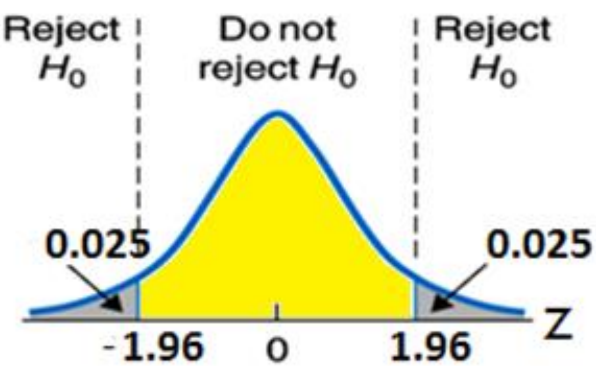


C. If Alternative $H_A: \mu \neq \mu_0$,
   Reject H_0 if $Z_{cal} < -Z_{1-\alpha/2}$ or $Z_{cal} > Z_{1-\alpha/2}$
   P-Value= $2P(Z > Z_{cal})$



Example: 7.2.1 Page 222

- Researchers are interested in the mean age of a certain population. A random sample of 10 individuals drawn from the population of interest has a mean of 27. Assuming that the population is approximately normally distributed with variance 20. Can we conclude that the mean is **different from 30** years? ($\alpha=0.05$) Find the p–value.



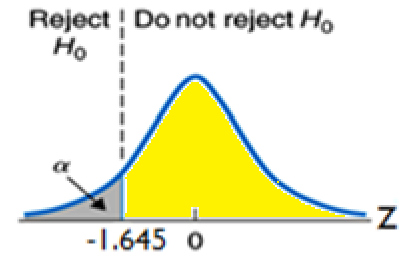
Solution

- ✓ $H_0: \mu=30$ vs $H_A: \mu \neq 30$
- ✓ Critical value$=Z_{1-\alpha/2}=Z_{0.975}=1.96$
- ✓ Test statistic: $Z_{cal}=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}=\frac{27-30}{\sqrt{20}/\sqrt{10}}=-2.12$
- ✓ **Decision: Reject H_0** *since z = |- 2.12| >1.96*
- ✓ P-value $=2P(Z > |-2.12|)=2(0.0174)=$**0.0348**
- ✓ **Conclusion:** We conclude that μ is not equal to 30.
- ✓ **95%confidence interval for μ is given by:** $\left[\bar{X}-z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X}+z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$
- ✓ **=[24.229, 29.771]**

Example: 7.2.2 Page 226

Referring to Example 7.2.1. Suppose that the researchers have asked: Can we conclude that μ<30.

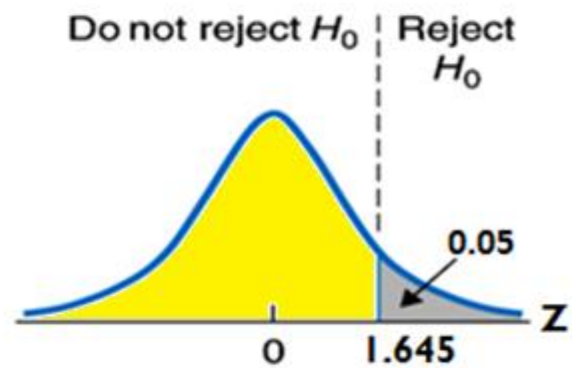
- H_0: μ≥30 vs H_A: μ<30
- Critical value=$Z_{1-\alpha}$=$Z_{0.95}$= 1.645
- Test statistic: $Z_{cal} = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} = \frac{27-30}{\sqrt{20}/\sqrt{10}} = -2.12$
- **Decision**: **Reject *H_0*** *since z = - 2.121 < - 1.645*
- **Conclusion:** We conclude that μ is smaller than 30.
- P-value =P(Z < -2.121)=**0.0174**



Example: Exercise 7.2.11 Page 235

A random sample 16 emergency reports was selected from files of an ambulance service. The mean time (computed from the sample data) required for ambulance to reach their destinations was 13 minutes. Assume that the population of times is normally distributed with variance of 9. Can we conclude at 0.05 level of significance that the population mean is **greater than 10** minutes?

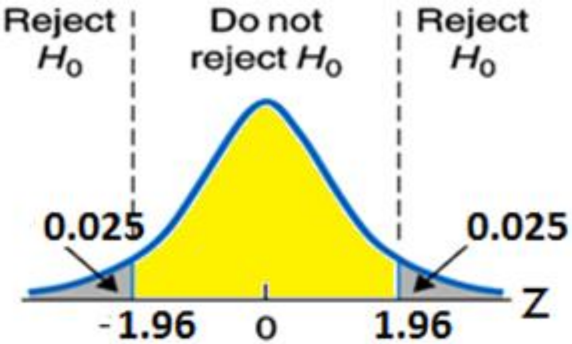
Solution



- H_0: $\mu=10$ vs H_A: $\mu>10$
- Critical value $=Z_{1-\alpha}=Z_{0.95}=1.645$
- Test statistic: $Z_{cal}=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}=\frac{13-10}{\sqrt{9}/\sqrt{16}}=4.0$
- **Decision**: **Reject H_0** since z = 4.0 > 1.645.
- **Conclusion:** We conclude that μ is greater than 10.
- P-value =P(Z > 4.0) ⟶ P-value **< 0.0001**

Example: Exercise 7.2.17 Page 236

Suppose it is known that the IQ scores of a certain population of adults are approximately normally distributed with a standard deviation of 15. A simple random sample of 25 adults drawn from this population had a mean IQ score of 105. On the basis of these data can we conclude that the mean IQ score is **not 100**? (α=0.05)



- **Solution**
  - ✓ H_0: μ=100 vs H_A: $\mu \neq 100$
  - ✓ Critical value=$Z_{1-\alpha/2}$=$Z_{0.975}$= 1.96
  - ✓ Test statistic: $Z_{cal}=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}=\frac{105-100}{15/\sqrt{25}}=1.67$
  - ✓ **Decision: Don't Reject H_0** *since z = |1.67| <1.96*
  - ✓ P-value =2P(Z > 1.67)=2(0.0475)=**0.095**
  - ✓ **Conclusion:** We conclude that μ is equal to 100.
  - ✓ **95%confidence interval for μ is given by:** $\left[\bar{X}-z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X}+z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$
  - ✓ **=[99.12, 110.88]**

B: Sampling from normally distributed population with unknown σ (n small)

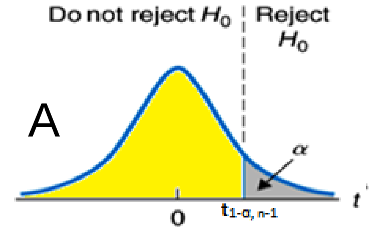
Null hypothesis $H_0: \mu = \mu_0$, α=level of significance

Test statistic: $$t_{cal} = \frac{\overline{X} - \mu_0}{s/\sqrt{n}} \Rightarrow t_{n-1}$$

A. If Alternative $H_A: \mu > \mu_0$

Reject H_0 if $t_{cal} > t_{1-\alpha, n-1}$

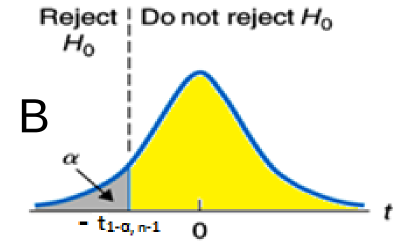
P-Value= $P(t_{n-1} > t_{cal})$



B. If Alternative $H_A: \mu < \mu_0$,

Reject H_0 if $t_{cal} < -t_{1-\alpha, n-1}$

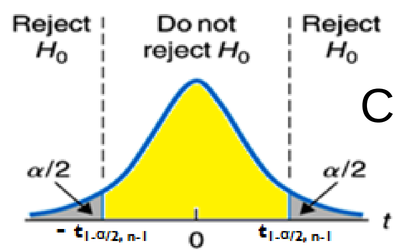
P-Value= $P(t_{n-1} < t_{cal})$



C. If Alternative $H_A: \mu \neq \mu_0$,

Reject H_0 if $t_{cal} < -t_{1-\alpha/2, n-1}$ or $t_{cal} > t_{1-\alpha/2, n-1}$

P-Value= $2P(t_{n-1} > t_{cal})$

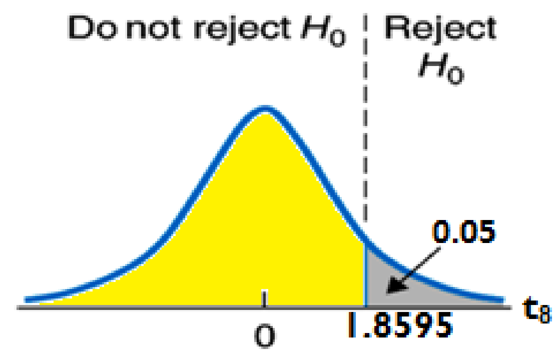


Example: Exercise 7.2.6 Page 234

Nine Laboratory animals were infected with a certain bacterium and then immunosuppressed (كبت المناعة). The mean number of organism later recovered from tissue specimens was 6.5 (coded data) with a standard deviation of 0.6. Can we conclude from theses data that the population mean is **greater than 6**? Let $\alpha=0.05$. What assumption are necessary?

- **Solution**

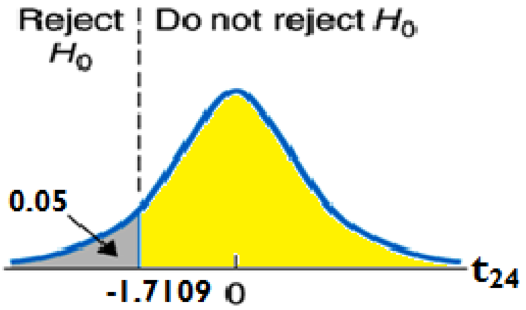
- ✓ Assumption: the population is normally distributed.
- ✓ $H_0: \mu \le 6$ vs $H_A: \mu > 6$
- ✓ Critical value= $t_{(1-\alpha,8)} = t_{(0.95,8)} = 1.8595$
- ✓ Test statistic: $t_{cal} = \frac{\bar{X}-\mu_0}{s/\sqrt{n}} = \frac{6.5-6}{0.6/\sqrt{9}} = 2.50$
- ✓ **Decision**: Reject H_0 since $t_{cal} = 2.50 > 1.8595$
- ✓ P-value = $P(t_8 > 2.50)$ → $0.01 < \text{p-value} < 0.025$
- ✓ **Conclusion:** we conclude from these data that the population mean is greater than 6



Example: Exercise 7.2.7 Page 234

A sample of **25** freshman nursing students made a mean score of **77** on a test designed to measure attitude toward the dying patient. The sample standard deviation was **10**. Do these data provide sufficient evidence to indicate at the **0.05** level of significance, that the population mean is less than 80? What assumptions are necessary?

- Solution:
  - ✓ Assumption: the population is normally distributed.
  - ✓ $H_0: \mu \geq 80$ **vs** $H_A: \mu < 80$
  - ✓ Critical value= $t_{(1-\alpha,24)} = t_{(0.95,24)} = 1.7109$
  - ✓ Test statistic: $t_{cal} = \frac{\bar{X}-\mu_0}{s/\sqrt{n}} = \frac{77-80}{10/\sqrt{25}} = -1.5$
  - ✓ **Decision: Don't Reject H_0** since $t_{cal} = -1.5 > -1.7109$
  - ✓ P-value = $P(t_{24} < -1.5)$ ⟶ $0.05 < \text{p-value} < 0.10$
  - ✓ **Conclusion:** we conclude from these data that the population mean is not less than 80.



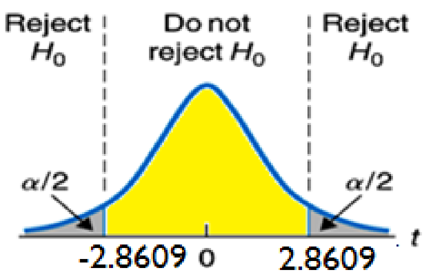
Example: Exercise 7.2.13 Page 235

Can we conclude that the mean maximum voluntary ventilation (التهوية الطوعية) value for apparently health college seniors is not 110 liters per minutes? A sample of 20 yielded the following values:

132, 33, 91, 108, 67, 169, 54, 203, 190, 133, 96, 30, 187, 21, 63, 166, 84, 110, 157, 138

Let $\alpha=0.01$. What assumptions are necessary?

- Solution:
  - ✓ **Mean=111.6 , S=56.303**
  - ✓ Assumption: the population is normally distributed.
  - ✓ **$H_0: \mu=110$ vs $H_A: \mu\neq 110$**
  - ✓ Critical value= $t_{(1-\alpha,19)} = t_{(0.99,19)} = 2.8609$
  - ✓ Test statistic: $t_{cal} = \frac{\bar{X}-\mu_0}{s/\sqrt{n}} = \frac{111.6-110}{56.303/\sqrt{20}} = 0.127$
  - ✓ **Decision: Don't Reject H_0** since $t_{cal} = |0.1271| < 2.8609$
  - ✓ P-value = $2P(t_{19} > 0.1271)$ → p-value > 0.20
  - ✓ **Conclusion:** we conclude from these data that the population mean is equal to 110. **A 99% C.I for μ is (75.582, 147.618).**



C: Sampling from nonnormally distributed population (n large)

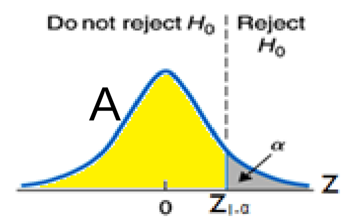
Null hypothesis $H_0: \mu = \mu_0$, α=level of significance

Test statistic:

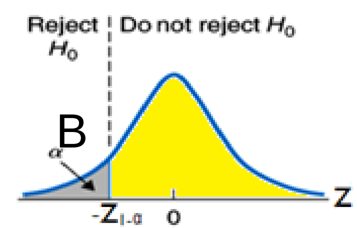
$$Z_{cal} = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}, \text{ if } \sigma \text{ known}$$

$$Z_{cal} = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}, \text{ if } \sigma \text{ unknown}$$

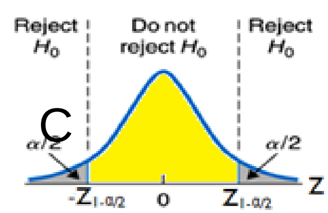
A. If Alternative $H_A: \mu>\mu_0$
   Reject H_0 if $Z_{cal} > Z_{1-\alpha}$
   P-Value= $P(Z > Z_{cal})$



B. If Alternative $H_A: \mu<\mu_0$,
   Reject H_0 if $Z_{cal} < -Z_{1-\alpha}$
   P-Value= $P(Z < Z_{cal})$



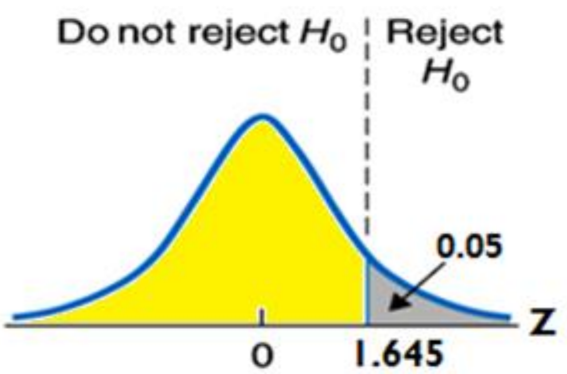
C. If Alternative $H_A: \mu\neq\mu_0$,
   Reject H_0 if $Z_{cal} < -Z_{1-\alpha/2}$ or $Z_{cal} > Z_{1-\alpha/2}$
   P-Value= $2P(Z > Z_{cal})$



Example: 7.2.4 Page 231

Among 157 African-American men, the mean systolic blood pressure was 146 mmHg with a standard deviation of 27. We wish to know if on the basis of these data, we may conclude that the mean systolic blood pressure for a population of African-American is greater than 140. Use α=0.05.

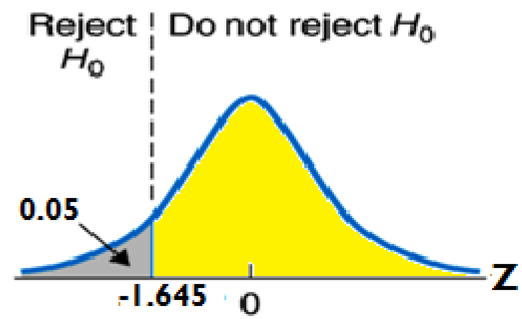
- Solution
  - ✓ $H_0: \mu \leq 140$ vs $H_A: \mu > 140$
  - ✓ Critical value=$Z_{1-\alpha}=Z_{0.95}=$ 1.645
  - ✓ Test statistic: $Z_{Cal}=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}=\frac{146-140}{27/\sqrt{157}}=2.78$
  - ✓ **Decision**: **Reject H_0** *since z = 2.78*>**1.645**
  - ✓ P-value =P(Z > 2.78)= **0.0027**
  - ✓ **Conclusion:** We conclude that the mean systolic blood pressure for the sampled population is greater than 140.



Example: Exercise 7.2.5 Page 234

In a sample of 49 adolescent who served as the subjects in an immunologic study, one variable of interest was the diameter of skin test reaction to an antigen. The sample mean and standard deviation were 21 and 11 mm erythematic, respectively. Can we conclude from these data that the population mean is less than 30? $\alpha=0.05$

Solution:



- H_0: $\mu \geq 30$ vs H_A: $\mu < 30$.
- Critical value=$Z_{1-\alpha}=Z_{0.95}= 1.645$.
- Test statistic: $Z_{Cal} = \frac{\bar{X}-\mu_0}{s/\sqrt{n}} = \frac{21-30}{11/\sqrt{49}} = -5.72$
- **Decision:** Yes, Reject H_0 *since* $z = -5.727 < -1.645$.
- **Conclusion:** We conclude that the population mean is not equal to 30.
- P-value $=P(Z < -5.727) \approx 0.000$. we write **p-value < 0.01.**