HANU Probability and Statistics

HYPOTHESIS TESTING

Lecture Contents

- Hypothesis Testing I

Hypothesis Testing

- Hypothesis testing is to check if a statement about the population is right or wrong (in probabilistic sense).

A statistical hypothesis is an assertion or conjecture concerning one or more populations.

- We could prove or reject the hypothesis from the evidence of data.

Hypothesis Testing

- What does this mean by “probabilistic sense”?

Since the conclusion is tested from finite sample -> probabilistic statement – i.e accept or reject the hypothesis with a probability (confident level) – p-value (probability of being wrong)

Hypothesis Testing

- **Hypothesis formulation for testing:**

H_0 – Null Hypothesis

H_1 – Alternative Hypothesis

Conclusions:

reject H_0 in favor of H_1 because of sufficient evidence in the data or

fail to reject H_0 because of insufficient evidence in the data.

H_0: defendant is innocent,

H_1: defendant is guilty.

Hypothesis Testing

- **Hypothesis formulation for testing:**

Example - certain type of cold vaccine is known to be only 25% effective after a period of 2 years. New vaccine has been developed and tested against 20 people, and shown to be effective on 8 of them. Does this mean that the new vaccine is more effective than the old?

Hypothesis Testing

- **Hypothesis formulation for testing:**

Hypotheses for testing:

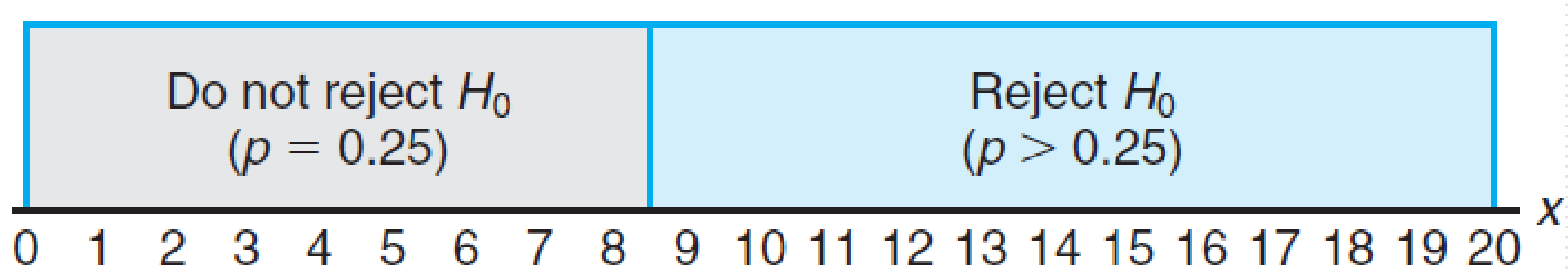
(Let p be the probability/proportion to be effective)

$$H_0:\ \ p = 0.25,$$
$$H_1:\ \ p > 0.25.$$

Hypothesis Testing

☐ Test Statistics:

We calculate the test statistics X an use X=8 as a boundary for our decision.



Hypothesis Testing

- **Probability of Errors:**

Type I error:

Rejection of the null hypothesis when it is true is called a type I error.

Type II error:

Nonrejection of the null hypothesis when it is false is called a type II error.

Hypothesis Testing

- **Probability of Errors:**

Table 10.1: Possible Situations for Testing a Statistical Hypothesis

	H_0 is true	H_0 is false
Do not reject H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

Hypothesis Testing

Probability of Errors:

Type I error:

$$\alpha = P(\text{type I error}) = P\left(X > 8 \text{ when } p = \frac{1}{4}\right) = \sum_{x=9}^{20} b\left(x; 20, \frac{1}{4}\right)$$

$$= 1 - \sum_{x=0}^{8} b\left(x; 20, \frac{1}{4}\right) = 1 - 0.9591 = 0.0409.$$

Hypothesis Testing

- **Probability of Errors:**

<u>Type II error</u>: Need an alternative for p.

$$\beta = P(\text{type II error}) = P\left(X \leq 8 \text{ when } p = \frac{1}{2}\right)$$

$$= \sum_{x=0}^{8} b\left(x; 20, \frac{1}{2}\right) = 0.2517.$$

Hypothesis Testing

- **Probability of Errors:**

Type II error: Lower type II error result in increase p.

$$\beta = P(\text{type II error}) = P(X \leq 8 \text{ when } p = 0.7)$$

$$= \sum_{x=0}^{8} b(x; 20, 0.7) = 0.0051.$$

Hypothesis Testing

- **Continuous Random Variable:**

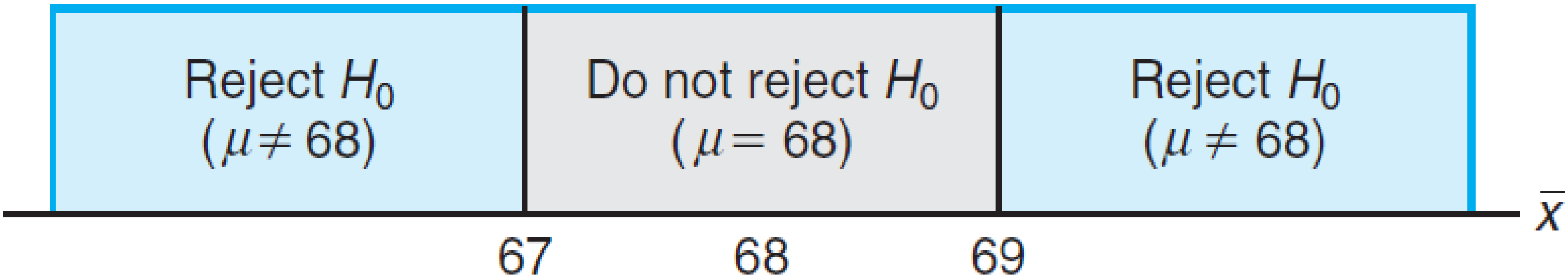
Consider the null hypothesis that the average weight of male students in a certain college is 68 kilograms against the alternative hypothesis that it is unequal to 68. That is, we wish to test

$$H_0: \quad \mu = 68,$$
$$H_1: \quad \mu \neq 68.$$

Hypothesis Testing

- **Continuous Random Variable:**

We calculate the sample mean and define the critical region for making our decision on hypothesis testing.



Hypothesis Testing

- **Continuous Random Variable:**

Assume the standard deviation of the population of weights to be σ = 3.6. For large samples, we may substitute s for σ if no other estimate of σ is available. Our decision statistic, based on a random sample of size n = 36, will be the sample mean , the most efficient estimator of μ.

Hypothesis Testing

- **Continuous Random Variable:**

The central limit theorem says that the distribution of $\bar{X}$ is approximately normal with standard deviation:

$$\sigma_{\bar{X}} = \sigma/\sqrt{n} = 3.6/6 = 0.6.$$

Hypothesis Testing

Continuous Random Variable:

Type I error:

$$\alpha = P(\bar{X} < 67 \text{ when } \mu = 68) + P(\bar{X} > 69 \text{ when } \mu = 68).$$

The z-values corresponding to $\bar{x}_1 = 67$ and $\bar{x}_2 = 69$ when H_0 is true are

$$z_1 = \frac{67 - 68}{0.6} = -1.67 \quad \text{and} \quad z_2 = \frac{69 - 68}{0.6} = 1.67.$$

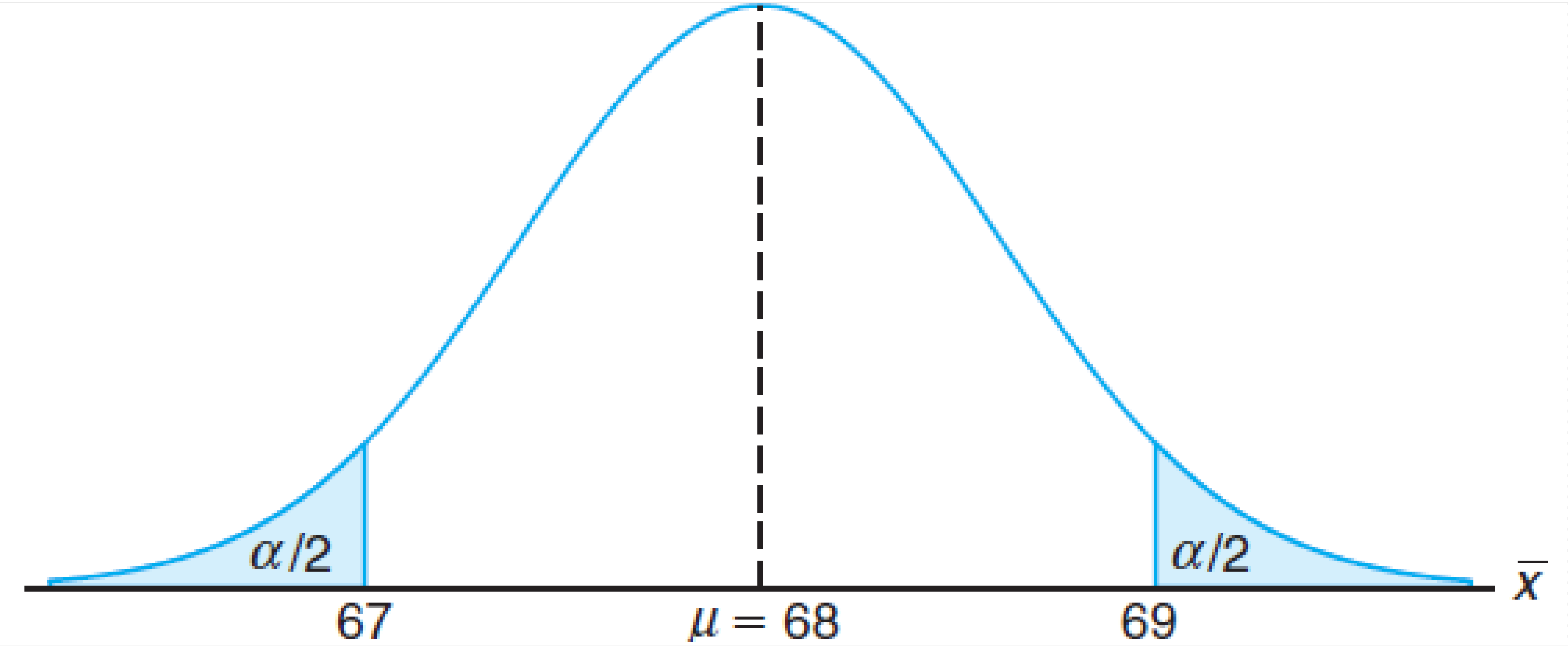
Therefore,

$$\alpha = P(Z < -1.67) + P(Z > 1.67) = 2P(Z < -1.67) = 0.0950.$$

Hypothesis Testing

- Continuous Random Variable:

Type I error:



Hypothesis Testing

Continuous Random Variable:

Type II error: testing over alternatives

$$\beta = P(67 \le \bar{X} \le 69 \text{ when } \mu = 70).$$

The z-values corresponding to $\bar{x}_1 = 67$ and $\bar{x}_2 = 69$ when H_1 is true are

$$z_1 = \frac{67 - 70}{0.45} = -6.67 \quad \text{and} \quad z_2 = \frac{69 - 70}{0.45} = -2.22.$$

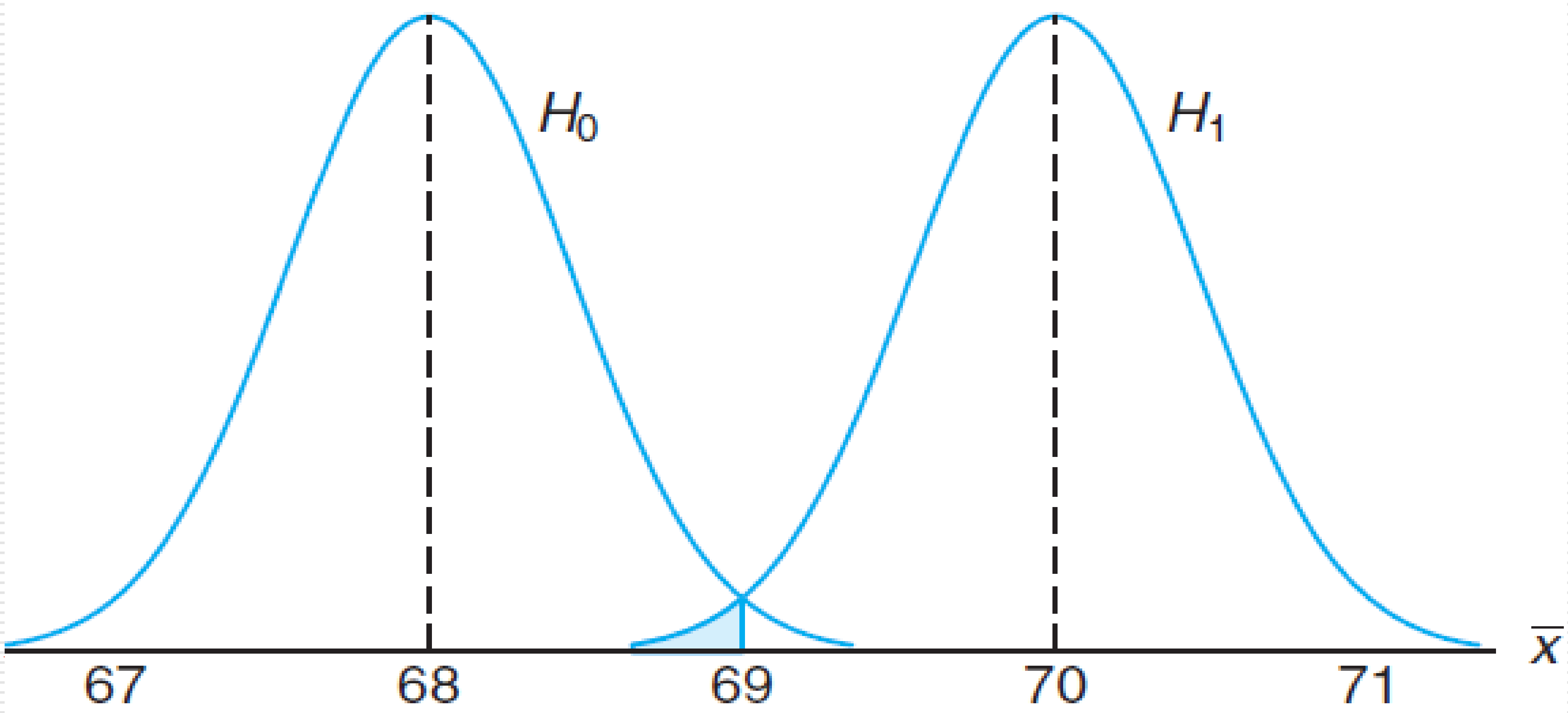
Therefore,

$$\begin{aligned}\beta = P(-6.67 < Z < -2.22) &= P(Z < -2.22) - P(Z < -6.67)\\ &= 0.0132 - 0.0000 = 0.0132.\end{aligned}$$

Hypothesis Testing

- **Continuous Random Variable:**

Type II error: testing over alternatives



Hypothesis Testing

Test Errors: properties

1. The type I error and type II error are related. A decrease in the probability of one generally results in an increase in the probability of the other.

2. The size of the critical region, and therefore the probability of committing a type I error, can always be reduced by adjusting the critical value(s).

3. An increase in the sample size n will reduce α and β simultaneously.

4. If the null hypothesis is false, β is a maximum when the true value of a parameter approaches the hypothesized value. The greater the distance between the true value and the hypothesized value, the smaller β will be.

Hypothesis Testing

One vs. Two-tailed Tests:

One-sided tests:

$$H_0\colon\ \theta = \theta_0, \qquad H_0\colon\ \theta = \theta_0,$$
$$H_1\colon\ \theta > \theta_0 \qquad H_1\colon\ \theta < \theta_0,$$

Two-sided tests:

$$H_0\colon\ \theta = \theta_0,$$
$$H_1\colon\ \theta \neq \theta_0,$$

Hypothesis Testing

How to choose Null and Alternative Hypotheses?

- Choose the Null hypothesis with equality (often to reject it).
- Choose the Alternative (H_1) (one-tailed, two-tailed) depending on what do we need when H_0 is rejected.

Hypothesis Testing

Significant Level

- To reject H_0 we need to compute the $P(H_0)$ (P-value) if this value is smaller than a threshold (significant level) α, then we could reject H_0
- It is customary to choose $\alpha=0.05$ or 0.01 ($z_{\alpha/2}$ –values for 0.05 and 0.01 are 1.96 and 2.57 respectively)

Hypothesis Testing

Hypothesis Testing Steps:

1. State the null and alternative hypotheses.
2. Choose a fixed significance level α.
3. Choose an appropriate test statistic and establish the critical region based on α.
4. Reject H_0 if the computed test statistic is in the critical region. Otherwise, do not reject.
5. Draw scientific or engineering conclusions.

Hypothesis Testing

Tests concerning mean: Single sample

When variance is known:

$$H_0\colon \mu = \mu_0,$$
$$H_1\colon \mu \neq \mu_0.$$

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}.$$

$$P\left(-z_{\alpha/2} < \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1 - \alpha$$

Hypothesis Testing

Tests concerning mean: Single sample

When variance is known:

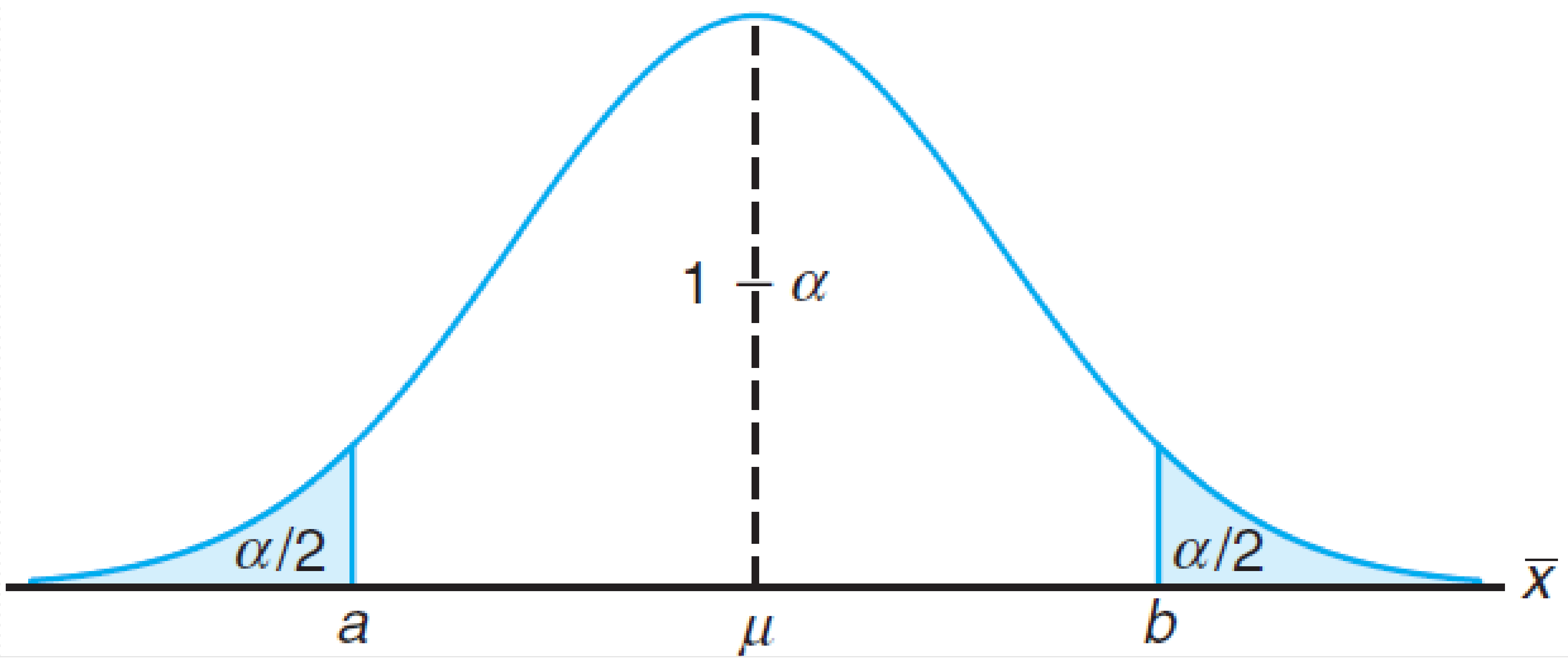
$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_{\alpha/2} \quad \text{or} \quad z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -z_{\alpha/2}$$

If $-z_{\alpha/2} < z < z_{\alpha/2}$, do not reject H_0. Rejection of H_0, of course, implies acceptance of the alternative hypothesis $\mu \neq \mu_0$. With this definition of the critical region, it should be clear that there will be probability α of rejecting H_0 (falling into the critical region) when, indeed, $\mu = \mu_0$.

Hypothesis Testing

Tests concerning mean: Single sample

When variance is known:



Hypothesis Testing

Tests concerning mean: Single sample

When variance is known: One-tailed tests

$$H_0:\ \mu = \mu_0,$$
$$H_1:\ \mu > \mu_0.$$

The critical region is when $Z>Z_\alpha$ ($Z<-Z_\alpha$)

Hypothesis Testing

Tests concerning mean: Single sample

When variance is known: Example

A random sample of 100 recorded deaths in the United States during the past year showed an average life span of 71.8 years. Assuming a population standard deviation of 8.9 years, does this seem to indicate that the mean life span today is greater than 70 years? Use a 0.05 level of significance.

Hypothesis Testing

Tests concerning mean: Single sample

When variance is known: Example

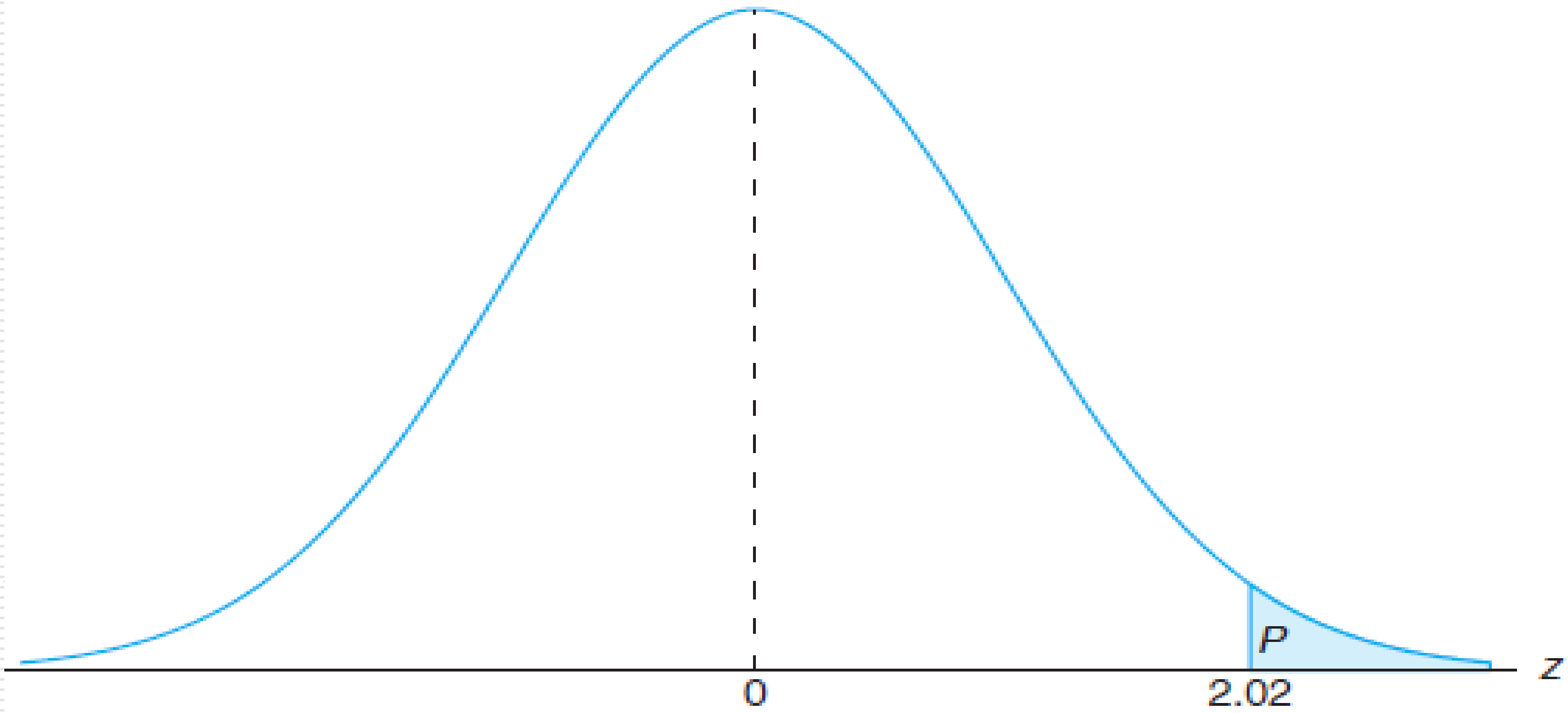
1. H_0: $\mu = 70$ years.
2. H_1: $\mu > 70$ years.
3. $\alpha = 0.05$.
4. Critical region: $z > 1.645$, where $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$.
5. Computations: $\bar{x} = 71.8$ years, $\sigma = 8.9$ years, and hence $z = \frac{71.8-70}{8.9/\sqrt{100}} = 2.02$.
6. Decision: Reject H_0 and conclude that the mean life span today is greater than 70 years.

Hypothesis Testing

Tests concerning mean: Single sample

When variance is known: Example

P-value(2.02)=P(Z>2.02)=0.0217



Hypothesis Testing

Tests concerning mean: Single sample

When variance is known: Example

A manufacturer of sports equipment has developed a new synthetic fishing line that the company claims has a mean breaking strength of 8 kilograms with a standard deviation of 0.5 kilogram. Test the hypothesis that $\mu = 8$ kilograms against the alternative that $\mu \neq 8$ kilograms if a random sample of 50 lines is tested and found to have a mean breaking strength of 7.8 kilograms. Use a 0.01 level of significance.

Hypothesis Testing

Tests concerning mean: Single sample

When variance is known: Example

1. H_0: $\mu = 8$ kilograms.

2. H_1: $\mu \neq 8$ kilograms.

3. $\alpha = 0.01$.

4. Critical region: $z < -2.575$ and $z > 2.575$, where $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$.

5. Computations: $\bar{x} = 7.8$ kilograms, $n = 50$, and hence $z = \frac{7.8-8}{0.5/\sqrt{50}} = -2.83$.

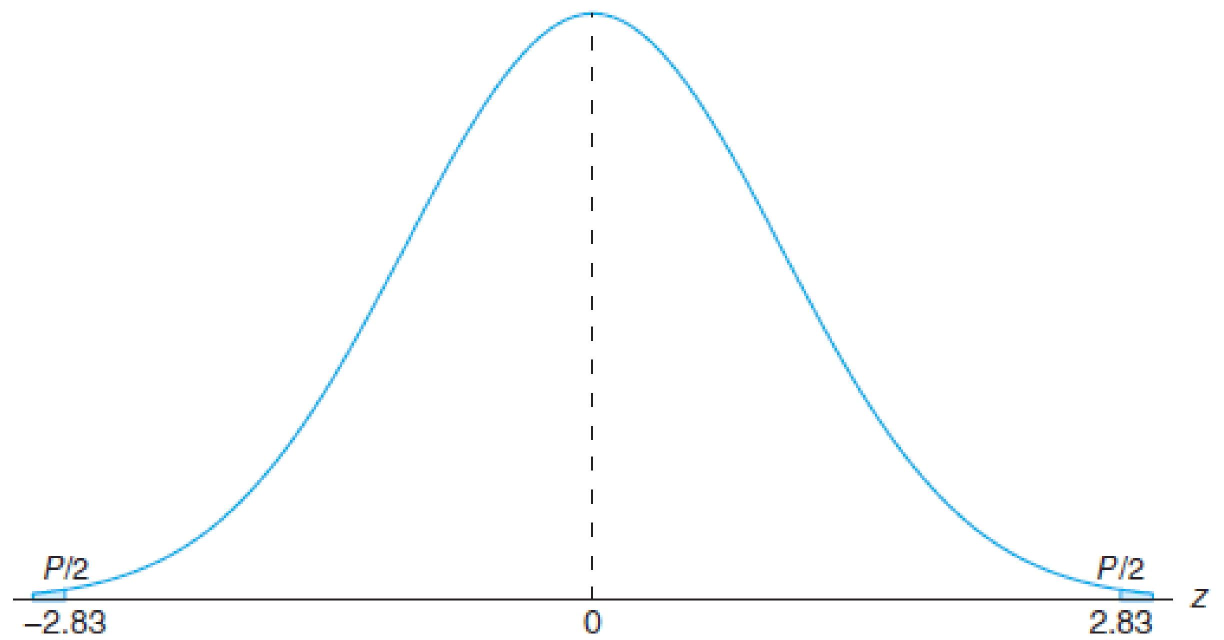
6. Decision: Reject H_0 and conclude that the average breaking strength is not equal to 8 but is, in fact, less than 8 kilograms.

Hypothesis Testing

Tests concerning mean: Single sample

When variance is known: Example

$$P = P(|Z| > 2.83) = 2P(Z < -2.83) = 0.0046,$$



Hypothesis Testing

Tests concerning mean: Single sample

When variance is unknown:

For the two-sided hypothesis

$$H_0\colon\ \mu = \mu_0,$$
$$H_1\colon\ \mu \neq \mu_0,$$

we reject H_0 at significance level α when the computed t-statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

exceeds $t_{\alpha/2,n-1}$ or is less than $-t_{\alpha/2,n-1}$.

Hypothesis Testing

Tests concerning mean: Single samples

When variance is unknown: Example

The Edison Electric Institute has published figures on the number of kilowatt hours used annually by various home appliances. It is claimed that a vacuum cleaner uses an average of 46 kilowatt hours per year. If a random sample of 12 homes included in a planned study indicates that vacuum cleaners use an average of 42 kilowatt hours per year with a standard deviation of 11.9 kilowatt hours, does this suggest at the 0.05 level of significance that vacuum cleaners use, on average, less than 46 kilowatt hours annually? Assume the population of kilowatt hours to be normal.

Hypothesis Testing

Tests concerning mean: Single samples

When variance is unknown: Example

1. H_0: $\mu = 46$ kilowatt hours.
2. H_1: $\mu < 46$ kilowatt hours.
3. $\alpha = 0.05$.
4. Critical region: $t < -1.796$, where $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ with 11 degrees of freedom.
5. Computations: $\bar{x} = 42$ kilowatt hours, $s = 11.9$ kilowatt hours, and $n = 12$. Hence,

$$t = \frac{42 - 46}{11.9/\sqrt{12}} = -1.16, \qquad P = P(T < -1.16) \approx 0.135.$$

6. Decision: Do not reject H_0 and conclude that the average number of kilowatt hours used annually by home vacuum cleaners is not significantly less than 46.

Hypothesis Testing

Tests concerning mean: Two samples

When the variances are known:

Compute statistics. The test is similar to single sample

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

Hypothesis Testing

Tests concerning mean: Two samples

When variances are unknown but equal:

For the two-sided hypothesis

$$H_0\colon\ \mu_1 = \mu_2,$$
$$H_1\colon\ \mu_1 \neq \mu_2,$$

we reject H_0 at significance level α when the computed t-statistic

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p\sqrt{1/n_1 + 1/n_2}},$$

where

$$s_p^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

exceeds $t_{\alpha/2, n_1+n_2-2}$ or is less than $-t_{\alpha/2, n_1+n_2-2}$.

Hypothesis Testing

Tests concerning mean: Two samples

Example -

An experiment was performed to compare the abrasive wear of two different laminated materials. Twelve pieces of material 1 were tested by exposing each piece to a machine measuring wear. Ten pieces of material 2 were similarly tested. In each case, the depth of wear was observed. The samples of material 1 gave an average (coded) wear of 85 units with a sample standard deviation of 4, while the samples of material 2 gave an average of 81 with a sample standard deviation of 5. Can we conclude at the 0.05 level of significance that the abrasive wear of material 1 exceeds that of material 2 by more than 2 units? Assume the populations to be approximately normal with equal variances.

Hypothesis Testing

Tests concerning mean: Two samples

Example -

Let μ_1 and μ_2 represent the population means of the abrasive wear for material 1 and material 2, respectively.

1. H_0: $\mu_1 - \mu_2 = 2$.
2. H_1: $\mu_1 - \mu_2 > 2$.
3. $\alpha = 0.05$.
4. Critical region: $t > 1.725$, where $t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p\sqrt{1/n_1 + 1/n_2}}$ with $v = 20$ degrees of freedom.
5. Computations:

$$\bar{x}_1 = 85, \qquad s_1 = 4, \qquad n_1 = 12,$$
$$\bar{x}_2 = 81, \qquad s_2 = 5, \qquad n_2 = 10.$$

Hypothesis Testing

Tests concerning mean: Two samples

Example -

Hence

$$s_p = \sqrt{\frac{(11)(16) + (9)(25)}{12 + 10 - 2}} = 4.478,$$

$$t = \frac{(85 - 81) - 2}{4.478\sqrt{1/12 + 1/10}} = 1.04,$$

$$P = P(T > 1.04) \approx 0.16. \qquad \text{(See Table A.4.)}$$

6. Decision: Do not reject H_0. We are unable to conclude that the abrasive wear of material 1 exceeds that of material 2 by more than 2 units.

Hypothesis Testing

Tests concerning mean: Two samples
Unknown and unequal variances

$$T' = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

has an approximate t-distribution with approximate degrees of freedom

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1-1) + (s_2^2/n_2)^2/(n_2-1)}.$$

As a result, the test procedure is to *not reject* H_0 when

$$-t_{\alpha/2,v} < t' < t_{\alpha/2,v},$$

Hypothesis Testing

Tests concerning mean: Two samples

Pairwise t-tests: For “before” vs “after”

Blood Sample Data: In a study conducted in the Forestry and Wildlife Department at Virginia Tech, J. A. Wesson examined the influence of the drug succinylcholine on the circulation levels of androgens in the blood. Blood samples were taken from wild, free-ranging deer immediately after they had received an intramuscular injection of succinylcholine administered using darts and a capture gun. A second blood sample was **obtained from each deer 30 minutes after the**

Hypothesis Testing

Tests concerning mean: Two samples

Pairwise t-tests: For “before” vs “after”

first sample, after which the deer was released. The levels of androgens at time of capture and 30 minutes later, measured in nanograms per milliliter (ng/mL), for 15 deer are given in Table. Assuming that the populations of androgen levels at time of injection and 30 minutes later are normally distributed, test at the 0.05 level of significance whether the androgen concentrations are altered after 30 minutes.

Hypothesis Testing

Tests concerning mean: Two samples

Table 10.2: Data for Case Study 10.1

Deer	Androgen (ng/mL) At Time of Injection	Androgen (ng/mL) 30 Minutes after Injection	d_i
1	2.76	7.02	4.26
2	5.18	3.10	−2.08
3	2.68	5.44	2.76
4	3.05	3.99	0.94
5	4.10	5.21	1.11
6	7.05	10.26	3.21
7	6.60	13.91	7.31
8	4.79	18.53	13.74
9	7.39	7.91	0.52
10	7.30	4.85	−2.45
11	11.78	11.10	−0.68
12	3.90	3.74	−0.16
13	26.00	94.03	68.03
14	67.48	94.03	26.55
15	17.04	41.70	24.66

Hypothesis Testing

Tests concerning mean: Two samples

Let μ_1 and μ_2 be the average androgen concentration at the time of injection and 30 minutes later, respectively. We proceed as follows:

1. H_0: $\mu_1 = \mu_2$ or $\mu_D = \mu_1 - \mu_2 = 0$.
2. H_1: $\mu_1 \neq \mu_2$ or $\mu_D = \mu_1 - \mu_2 \neq 0$.
3. $\alpha = 0.05$.
4. Critical region: $t < -2.145$ and $t > 2.145$, where $t = \frac{\bar{d} - d_0}{s_D/\sqrt{n}}$ with $v = 14$ degrees of freedom.
5. Computations: The sample mean and standard deviation for the d_i are

$$\bar{d} = 9.848 \qquad \text{and} \qquad s_d = 18.474.$$

Hypothesis Testing

Tests concerning mean: Two samples

Therefore,

$$t = \frac{9.848 - 0}{18.474/\sqrt{15}} = 2.06.$$

6. Though the t-statistic is not significant at the 0.05 level, from Table A.4,

$$P = P(|T| > 2.06) \approx 0.06.$$

As a result, there is some evidence that there is a difference in mean circulating levels of androgen.

Further Readings

- Textbook 2 - Chapter 10.
- Textbook 1- Chapter 8-9.