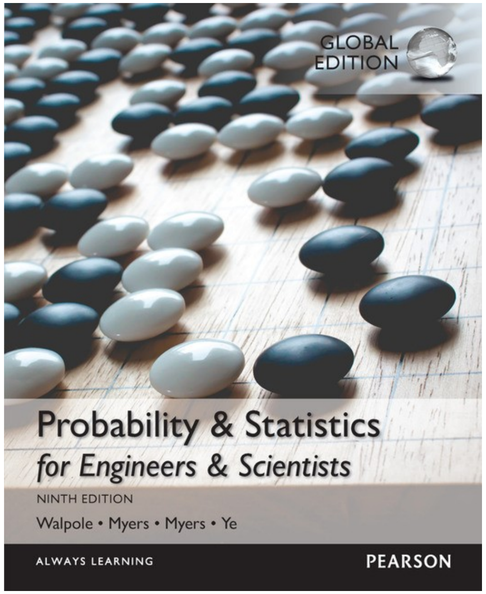
Chapter 10

One- and Two-Sample Tests of Hypotheses (2/2)

School of Computing, Gachon Univ.
Joon Yoo



가천대학교 AI·소프트웨어학부
Gachon University School of Computing

10.3 The Use of P-Values for Decision Making in Testing Hypotheses

- A scientist is testing the effect a drug on response time by injecting 100 rats. The scientist knows that the mean response time for rats not injected with the drug is 1.2 seconds. The mean response time of 100 injected rats is 1.05 seconds with a sample standard deviation of 0.5 seconds. Do you think that the drug has an effect of response time ?

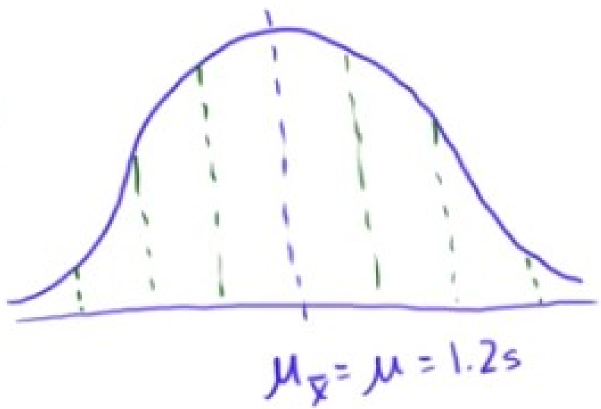
H_0: Drug has no effect $\Rightarrow$ $\mu = 1.2$ s (even w/ drug)

H_1: Drug has an effect $\Rightarrow$ $\mu \neq 1.2$ s when the drug is given

Assume H_0:

$$Z = \frac{1.2 - 1.05}{0.05}$$

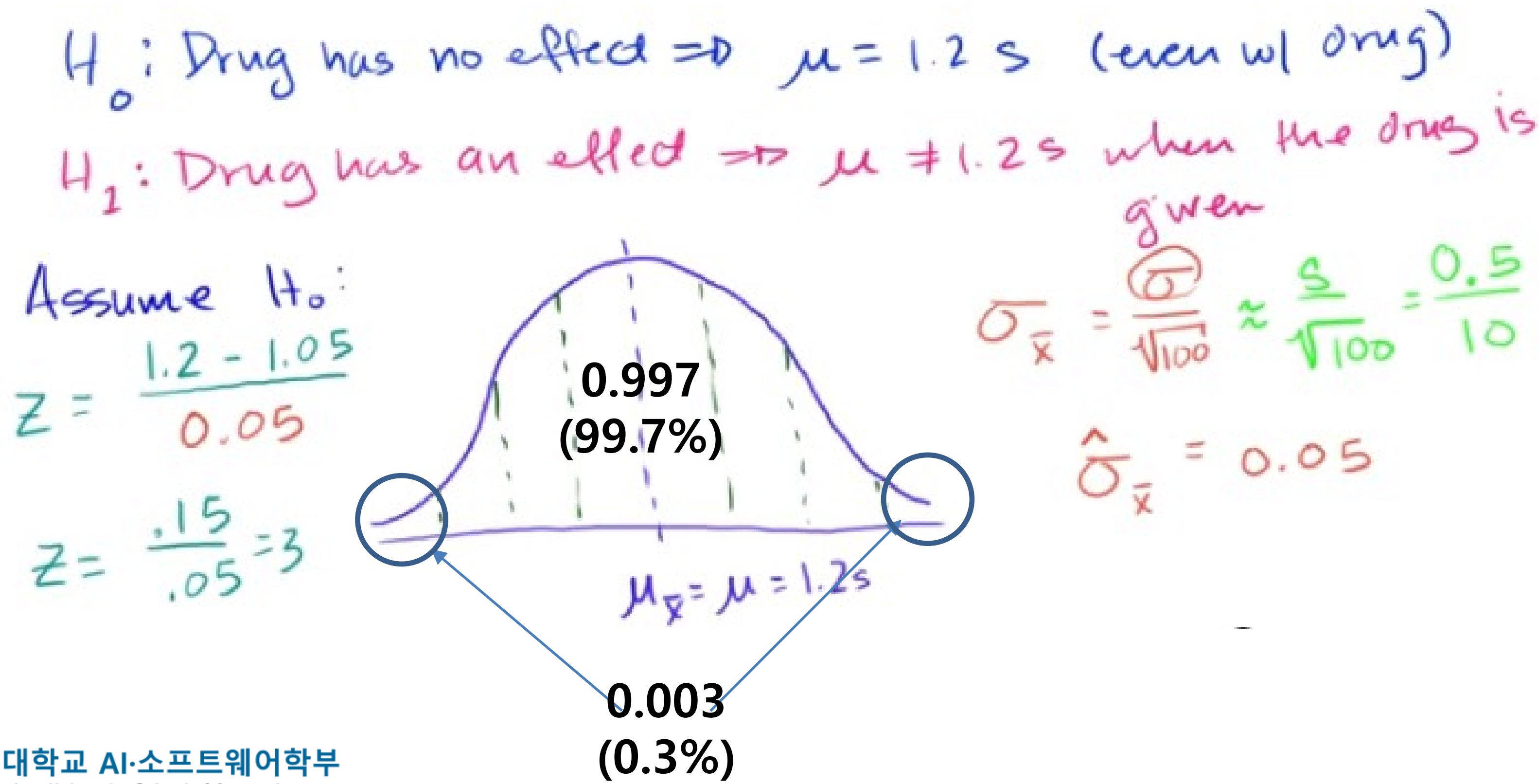
$$Z = \frac{.15}{.05} = 3$$



$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{100}} \approx \frac{s}{\sqrt{100}} = \frac{0.5}{10}$$

$$\hat{\sigma}_{\bar{x}} = 0.05$$

- Let's think about the probability of observing a result given that the null hypothesis(H_0) is true.
  - **p-value** = P (sample happens | H_0 is true) = 0.003 (or 0.3%)
  - If the probability is very small (e.g., less than **α=0.05**), what does it mean ?



P-Value

- In statistics, the p-value is the probability of obtaining the observed sample results given that the null hypothesis is true.
  - Example, p-value of .05
    - Indicates that you would have only a 5% chance of drawing the sample being tested if the null hypothesis was actually true.

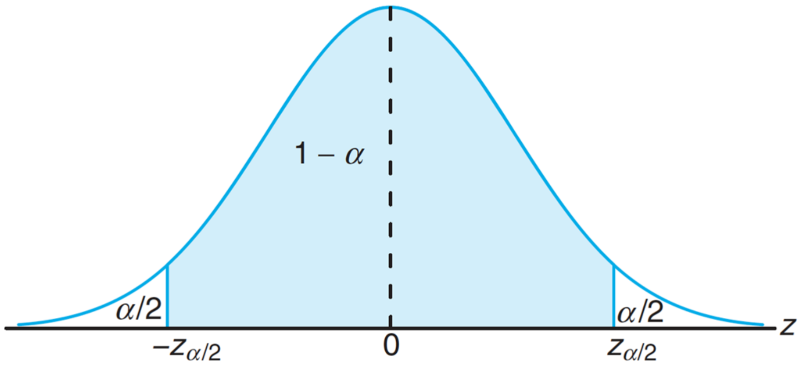
- We can use p-value as a score for decision making in testing hypothesis
  - This is a way of assessing the "believability" of the null hypothesis, given the evidence provided by a random sample.

Definition 10.5: A ***P*-value** is the lowest level (of significance) at which the observed value of the test statistic is significant.

Example

$H_0 : \mu = 10$ vs. $H_1 : \mu \neq 10$

- Suppose a sample value of $z = 1.87$ is observed. In fact, in a two-tailed scenario one can quantify this risk as
  - P-value = $P(Z \geq 1.87 \mid H_0 \text{ true})$
    $= P(Z \geq 1.87 \mid \mu = 10) = 0.0614$
- with $\boldsymbol{\alpha} = 0.05$ the value is not significant. (P-value > $\boldsymbol{\alpha}$)

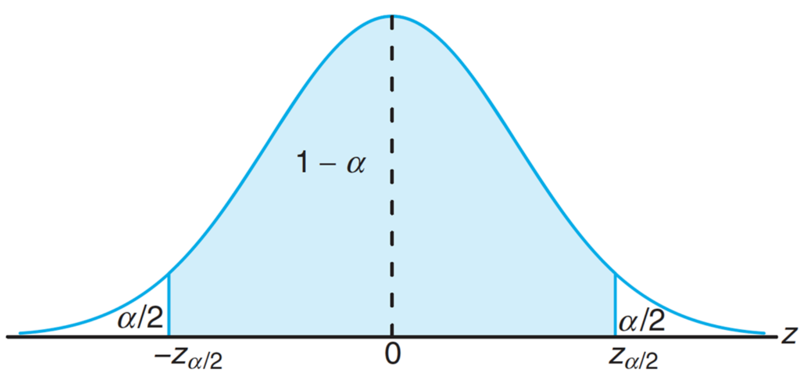


$\alpha = 0.05$
$z_{0.025} = 1.96$
from table A.3

Example (cont.)

- For example, if z is 2.73, it is informative for the user to observe that
  - P-value = $P(Z \geq 2.73 \mid H_0 \text{ true})$
    $= 2P(Z > 2.73 \mid \mu = 10) = 0.0064$

- It is important to know that under the condition of H_0, a value of $z = 2.73$ is an extremely rare event. Namely, a value at least that large in magnitude would only occur 64 times in 10,000 experiments.



$\alpha = 0.05$
$z_{0.025} = 1.96$
from table A.3

Two Approaches

- **Approach 1 : Hypothesis Testing (using critical value/region) with a fixed Significance Level α**

- **Approach 2 :** Significance testing based on the calculated ***P*-value**

Approach 1 : Classic Hypothesis Testing

1. State the null and alternative hypotheses.
2. Choose a fixed significance level α.
3. Specify the appropriate test statistic (e.g., mean) and establish the critical region based on α.
4. Make a decision to reject H_0 or fail to reject H_0, based on the location of the test statistic.
5. Make an engineering or scientific conclusion.

Approach 2 : P-value Approach

- Approach 2 - Significance testing based on the calculated *P*-value

1. State the null and alternative hypotheses.
2. Choose an appropriate test statistic.
3. Calculate value of test statistic and compute *P*-value.
4. Make a decision to reject H_0 or fail to reject H_0, based on the *P*-value.
5. Make an engineering or scientific conclusion.

10.4. Single Sample: Tests Concerning a Single Mean (Variance Known)

Tests on a Single Mean (Variance Known)

- The model for the underlying situation centers around an experiment with $X_1, X_2, \ldots, X_n$ representing a random sample from a distribution with mean μ and variance $\sigma^2 > 0$. Consider first the hypothesis

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

- The appropriate test statistic should be based on the random variable $\bar{X}$.
- In Chapter 8, the Central Limit Theorem was introduced, which essentially states that despite the distribution of X, the random variable $\bar{X}$ has approximately a normal distribution with mean μ and variance σ^2/n for reasonably large sample sizes (n≥30). So, $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}}^2 = \sigma^2/n$.
- We can then determine a critical region based on the computed sample average, $\bar{x}$.

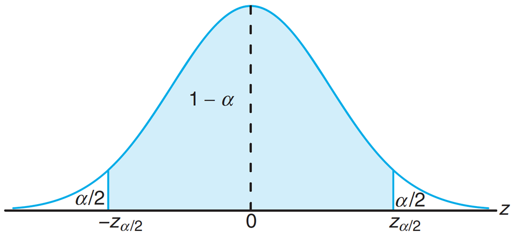
$$\boxed{Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.}$$

Approach 1.

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

- Non-rejection region:

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



- Rejection region:

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

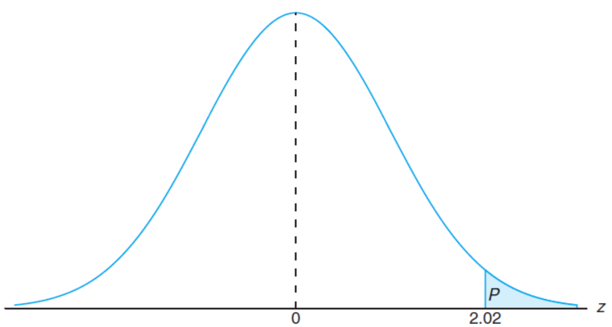
Rejection of H_0, of course, implies acceptance of the alternative hypothesis $\mu = \mu_0$. With this definition of the critical region, it should be clear that there will be probability α of rejecting H_0 (falling into the critical region) when, indeed, $\mu = \mu_0$.

Example 10.3

Example 10.3: 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.

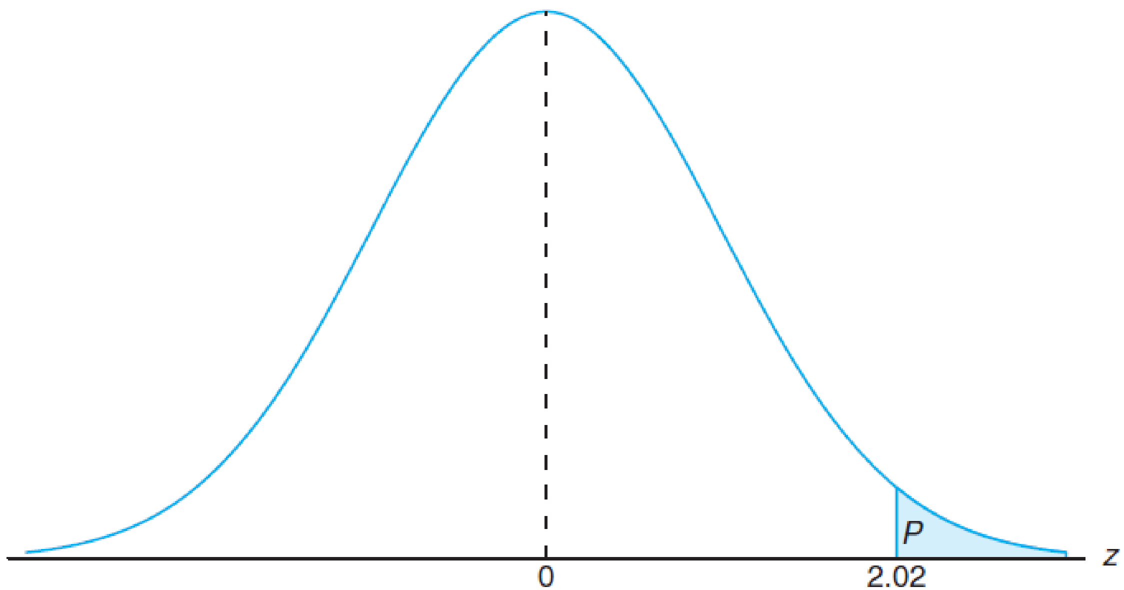
Solution:

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.



Cont.

- The P-value corresponding to $z = 2.02$ is given by the area of shaded region in Figure. $P(Z > 2.02) = 0.0217 < (\alpha = 0.05)$. As a result, the evidence in favor of H_1 is even stronger than suggested by a 0.05 level of significance.



Example 10.4

Example 10.4: 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.

***Solution*:**

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.

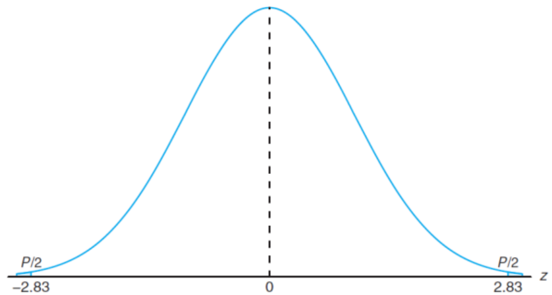


Figure 10.11: P-value for Example 10.4.

Cont.

Since the test in this example is two tailed, the desired P-value is twice the area of the shaded region in Figure 10.11 to the left of $z = -2.83$. Therefore, using Table A.3, we have

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

which allows us to reject the null hypothesis that $\mu = 8$ kilograms at a level of significance smaller than 0.01.

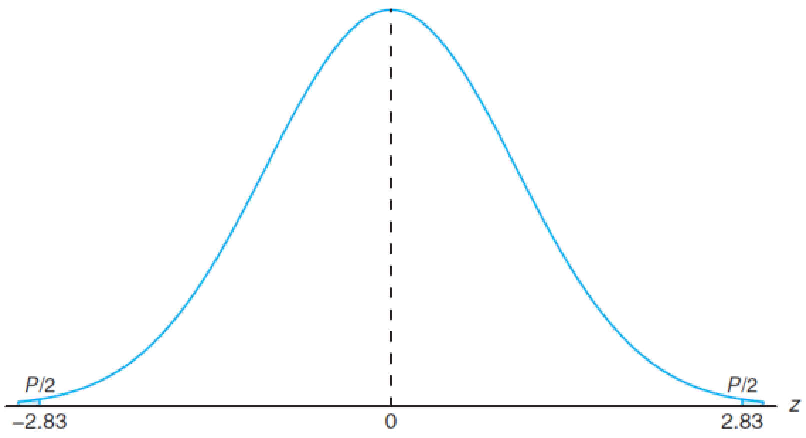


Figure 10.11: P-value for Example 10.4.

Tests on a Single Mean (Variance unknown)

The t-Statistic for a Test on a Single Mean (Variance 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 $-t_{\alpha/2,n-1}$.

Example 10.5

Example 10.5: 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.

Solution:

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.

https://www.psycom.net/bipolar-questions-answers