Convergence Informatics

Chapter 9 (Week 5)

Statistical hypothesis testing

Asst. Prof. Minseok Seo
mins@korea.ac.kr

Contents

Hypothesis testing

Statistical test

01

Hypothesis testing (Remind)

Basic concept of hypothesis testing

- Basic questions when given data
  - Difference
  - Equivalency

- What is a statistical hypothesis?
  - Null hypothesis
  - Alternative hypothesis

Hypothesis testing

Advanced concept of hypothesis testing (Today's topic)

- Why is the null hypothesis important for statistical test?
- What is the statistical(Hypothesis) test?
- What is statistical power?
- What is test statistic?
- What is P-value?

Decision making

Decision making under uncertainty

- We have to make decisions even when you are unsure. School, marriage, therapy, jobs, whatever.

- Statistics provides an approach to decision making under uncertainty. Sort of decision making by choosing the same way you would bet. Maximize expected utility (subjective value).

- In inferential statistics, the null hypothesis is a general statement or default position that there is nothing new happening, like there is no association among groups, or no relationship between two measured phenomena.

Main idea of Hypothesis testing

Statistical testing

- A statistical hypothesis, sometimes called confirmatory data analysis, is a hypothesis that is testable on the basis of observing a process that is modeled via a set of random variables.

*** Main idea ***

- It is difficult to prove that a fact is “right”.
- But it is easy to prove that it is “wrong”.
- The reason is that you only have to find one counter example.

Two types of hypothesis

Alternative and null hypothesis

Alternative hypothesis $\boldsymbol{H_1}$	**Null hypothesis** $\boldsymbol{H_0}$
H_1, H_A	H_0
Hypothesis as main purpose of the study	Hypothesis as opposed to purpose
Hypothesis that you want to claim to be right	Hypothesis that you want to claim to be wrong

- It is a way to induce contradiction after assuming null hypothesis is correct.

Main idea of Hypothesis testing

Statistical testing (Cont.)

*** Main idea ***

- It is difficult to prove that a fact is "right", but it is easy to prove that it is "wrong".
- The reason is that you only have to find one counter example.

H_1: All students in Convergence Informatics classes are statistical experts.

- Let's set up the opposite hypothesis

H_0: All students in the Convergence Informatics class do not know statistics.

Main idea of Hypothesis testing

Statistical testing (Cont.)

H_1: All students in Convergence Informatics classes are statistical experts.

Our goal is to prove that this proposition is “true”

H_0: All students in the Convergence Informatics class do not know statistics.

If we find evidence for "false" in this null hypothesis, the original proposition is "true."

Main idea of Hypothesis testing

Another example

H_1: The height of men and women is different

Our goal is to prove that this proposition is "true"

H_0: The height of men and women is same

- Assuming that men and women are the same height, let's examine the men's and women's heights, respectively, to find the mean and variance.
    - If we find numerical evidence that men and women's height are different in collected data?
    - Original proposition is "true"

Two types of errors

Statistical error

- Statistical test results reject null hypothesis or not.

	H_1: signal present	H_0: signal absent
Detection	True Positive	False Positive ***type I error***
Null result	False Negative ***type II error***	True Negative
	TPF = sensitivity = power = TP/(TP+FN)	**FPF = 1-specificity = 1-CL = FP/(FP+TN)**

Two types of errors

Statistical error (i.e. classification?)

- Statistical test results reject null hypothesis or not.

"Confusion Matrix" or Contingency Table for Binary Hypothesis Testing		
Truth / Declaration	True Hypothesis H_0 (Null)	True Hypothesis H_1
Declared Hypothesis H_0(Null)	$P(H_0 \mid H_0) = P_{Spec} = Specificity$ $= \frac{\# H_0 \ Samples \ Declared \ H_0}{\# H_0 \ Samples}$	$P(H_0 \mid H_1) = P_{Miss} = P(Miss)$ $= \frac{\# H_1 \ Samples \ Declared \ H_0}{\# H_1 \ Samples}$
Declared Hypothesis H_1	$P(H_1 \mid H_0) = P_{FA} = P(False\ Alarm)$ $= \frac{\# H_0 \ Samples \ Declared \ H_1}{\# H_0 \ Samples}$	$P(H_1 \mid H_1) = P_D = P(Detection)$ $= \frac{\# H_1 \ Samples \ Declared \ H_1}{\# H_1 \ Samples}$

$$P(H_0 \mid H_0) + P(H_1 \mid H_0) = 1 \qquad P(H_0 \mid H_1) + P(H_1 \mid H_1) = 1$$

$$P_{cc} = P(Correct\ Classification) = P(H_0 \mid H_0)P(H_0) + P(H_1 \mid H_1)P(H_1)$$

$$P_e = P(Error) = 1 - P_{cc} = P(H_0 \mid H_1)P(H_1) + P(H_1 \mid H_0)P(H_0)$$

$$\text{Assume : Correct classification is given zero cost} \Rightarrow C_{00} = C_{11} = 0$$

$$\text{Incorrect classification is given full cost} \Rightarrow C_{01} = C_{10} = 1$$

Several measures

Confusion matrix (Basic)

- TP: true positive
- TN: true negative
- FP: false positive (Type I error)
- FN: false negative (Type II error)

Several measures (extension)

Confusion matrix (Advanced)

- TPR (True Positive Rate, Power, Sensitivity, Hit rate, Recall)

$$= \frac{TP}{TP+FN} = 1 - \text{FPR}$$

- TNR (True Negative Rate, Specificity)

$$= \frac{TN}{TN+FP} = 1 - \text{FPR}$$

- Be sure to understand and memorize this concept because it is the most widely used measure in statistics & any data science field.

Several measures (extension)

Confusion matrix (Advanced)

- PPV (Positive Predictive Value, Precision)

$$= \frac{TP}{TP+FP} = 1 - \text{FDR}$$

- NPV (Negative Predictive Value)

$$= \frac{TN}{TN+FN} = 1 - \text{FOR}$$

- FNR (False Negative Rate, Miss rate)

$$= \frac{FN}{FN+TP} = 1 - \text{TPR}$$

- FPR (False Positive Rate, Fall-out)

$$= \frac{FP}{FP+TN} = 1 - \text{TNR}$$

Several measures (extension)

Confusion matrix (Advanced)

- FDR (False Discovery Rate)

$$= \frac{FP}{FP+TP} = 1 - \text{PPV}$$

- ACC (Accuracy)

$$= \frac{TP+TN}{TP+TN+FP+FN}$$

- ERR (Error Rate)

$$= \frac{FP+FN}{TP+TN+FP+FN} = 1 - \text{ACC}$$

Illustrative example

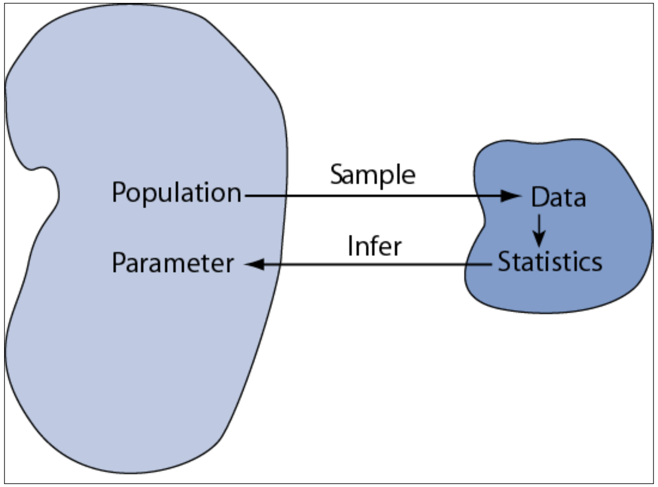
Body weight study

- **The problem:** In the 1970s, 20–29 year old men in the U.S. had a mean μ body weight of 170 pounds. Standard deviation σ was 40 pounds. We test whether mean body weight in the population now differs.

- **Null hypothesis** $H_{0:}\ \mu = 170$ ("no difference")

- The **alternative hypothesis** can be either $H_{a:}\ \mu > 170$ (**one-sided test**) or $H_{a:}\ \mu \neq 170$ (**two-sided test**)

- One side vs Two-side



Illustrative example

Body weight study

The **sampling distributions of a mean (SDM)** describes the behavior of a sampling mean

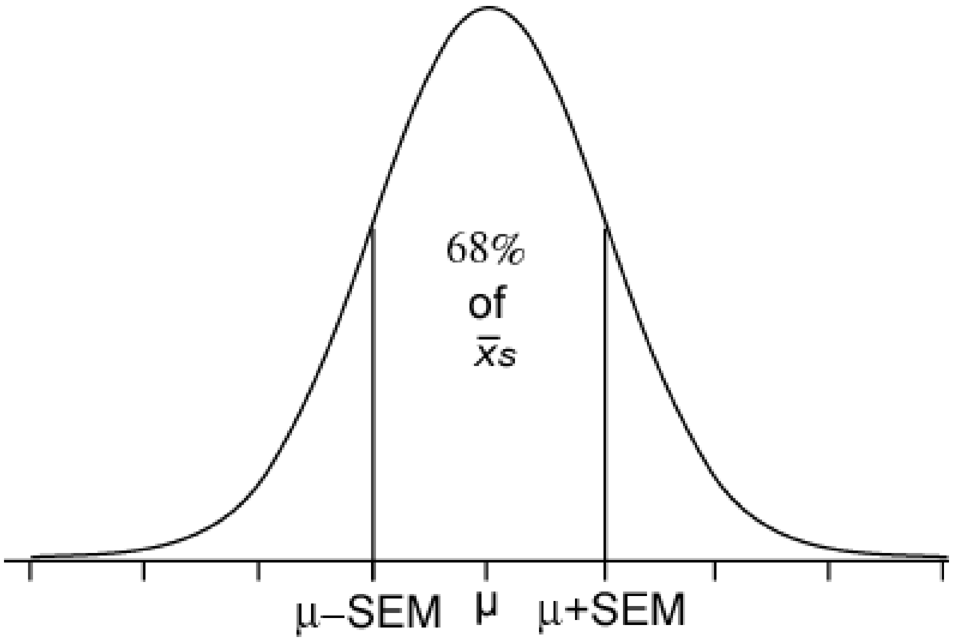


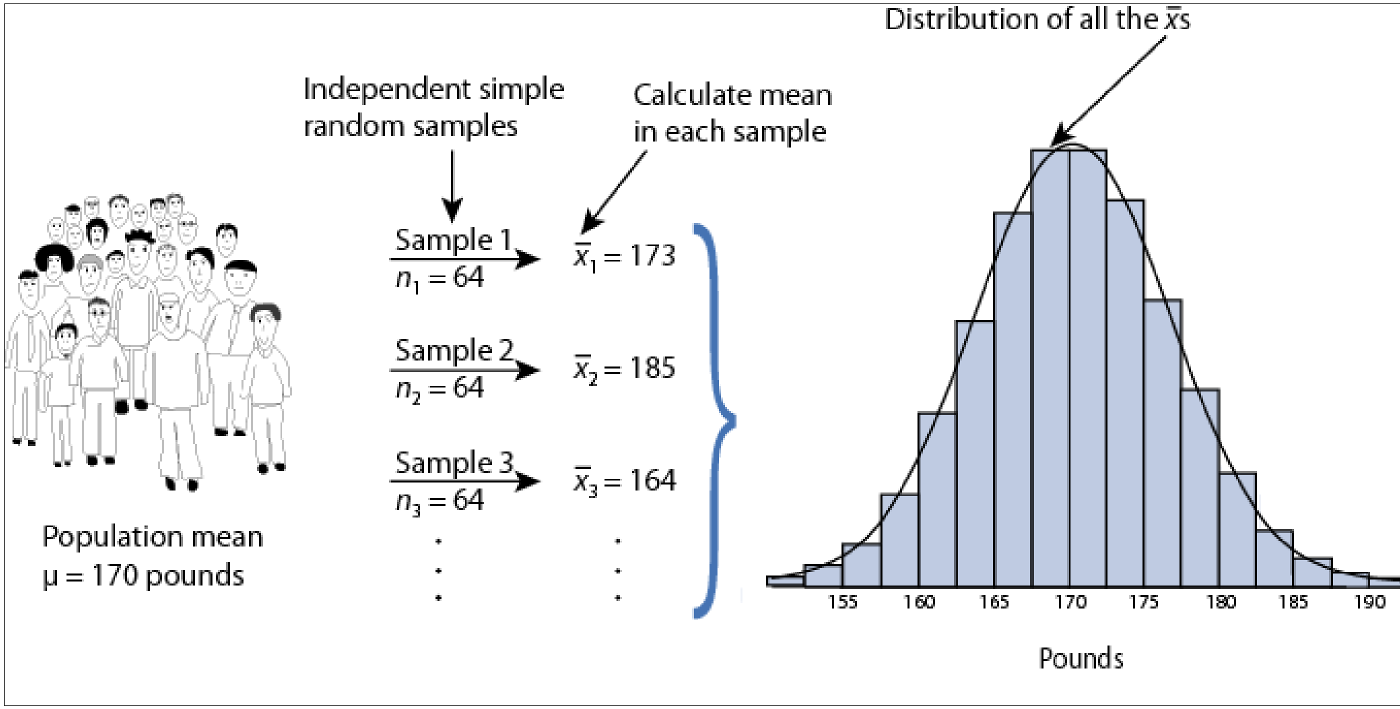
Fig. sdm&se68%.ai

$$\bar{x} \sim N\left(\mu, SE_{\bar{x}}\right)$$

$$\text{where } SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Illustrative example

Body weight study



$$\bar{x} \sim N\left(\mu, SE_{\bar{x}}\right)$$

$$\text{where } SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Illustrative example

Z-statistic (Test statistic)

This is an example of a one-sample test of a mean when σ is known. Use this statistic to test the problem:

$$z_{\text{stat}} = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}}$$

where $\mu_0 \equiv$ population mean assuming H_0 is true

$$\text{and } SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Illustrative example

Z-statistic (Test statistic)

- This is an example of a one-sample test of a mean when σ is known. Use this statistic to test the problem:

$$z_{\text{stat}} = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}}$$

$$\text{where } \mu_0 \equiv \text{population mean assuming } H_0 \text{ is true}$$

$$\text{and } SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- When will the test statistic value increase or decrease?

Illustrative example

Z-statistic (Test statistic)

- For the illustrative example, μ_0 = 170

- We know σ = 40 and *n* = 64. Therefore

$$SE_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{40}{\sqrt{64}} = 5$$

- If we found a sample mean of 173, then

$$z_{\text{stat}} = \frac{\overline{x} - \mu_0}{SE_{\overline{x}}} = \frac{173 - 170}{5} = 0.60$$

Illustrative example

Z-statistic (Test statistic)

- If we found a sample mean of 185, then

$$z_{\text{stat}} = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}} = \frac{185 - 170}{5} = 3.00$$

P-value

Probability value

- The *P*-value answer the question: What is the probability of the observed test statistic or one more extreme **when H_0 *is true*?**

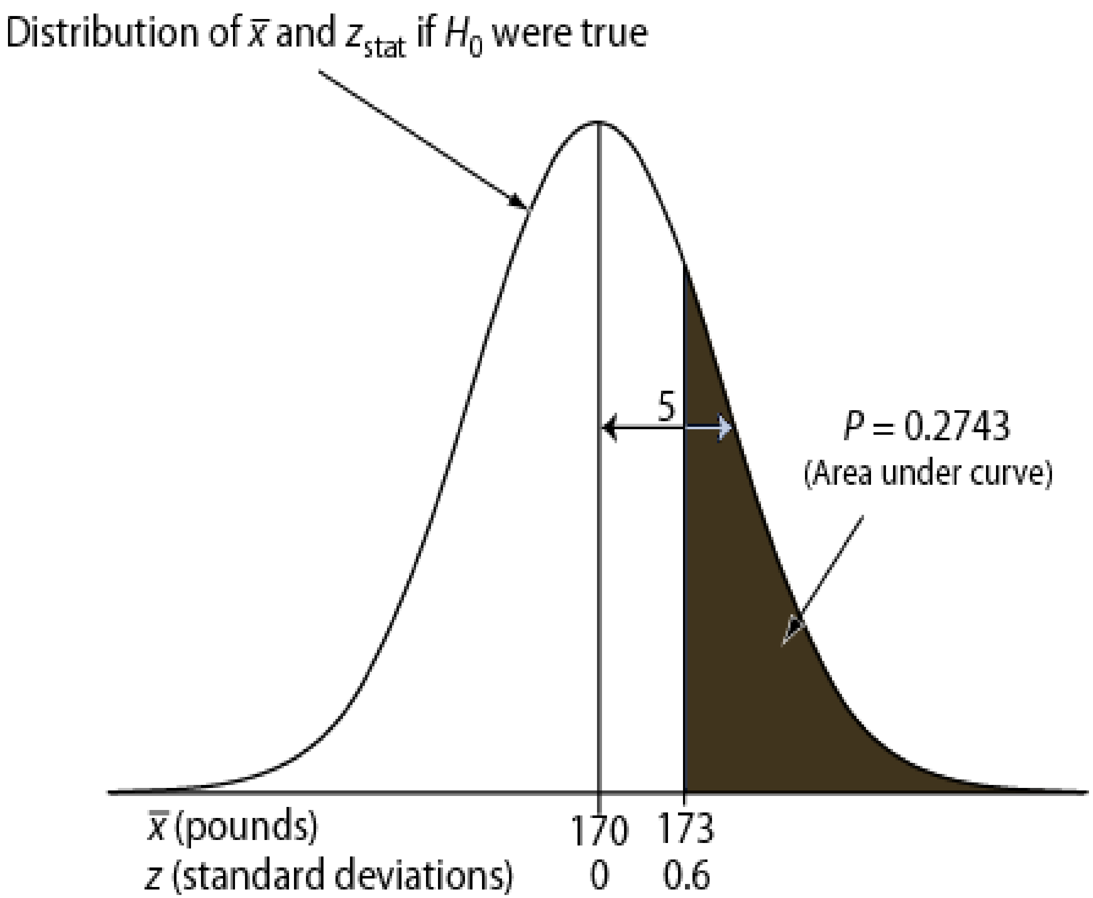
> **Definition of P-value**: Under the assumption that H0 is correct, probability that the observed test statistic is skewed towards H1.

- Convert *z* statistics to *P*-value :
  - For H_a: $\mu > \mu_0 \Rightarrow P = \Pr(Z > z_{stat})$ = right-tail beyond z_{stat}
  - For H_a: $\mu < \mu_0 \Rightarrow P = \Pr(Z < z_{stat})$ = left tail beyond z_{stat}
  - For H_a: $\mu {}^{1} \mu_0 \Rightarrow P = 2 \times$ one-tailed *P*-value

P-value

Probability value

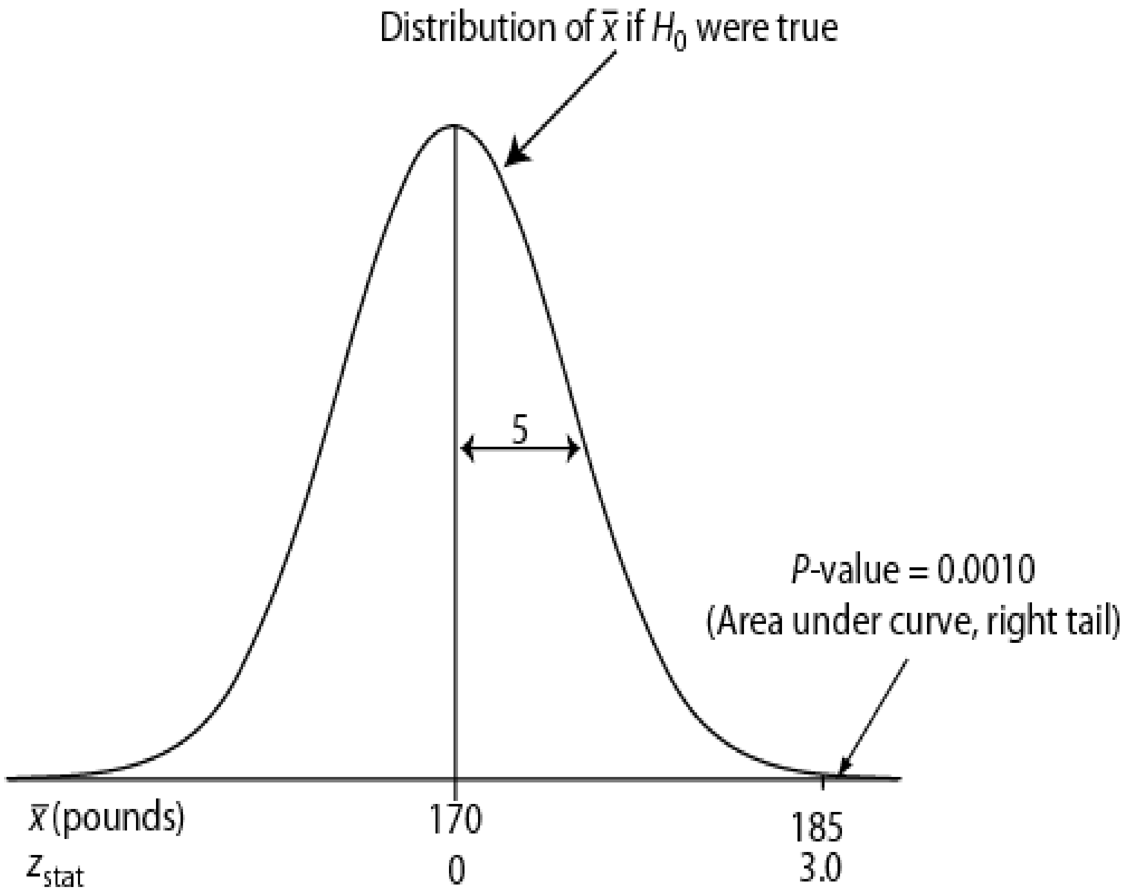
- One-sided P-value for z_{stat} of 0.6



P-value

Probability value

- One-sided *P*-value for z_{stat} of 0.6



Interpretation of P-value

Probability value

Conventions*

- $P > 0.10 \Rightarrow$ non-significant evidence against H_0
- $0.05 < P \leq 0.10 \Rightarrow$ marginally significant evidence
- $0.01 < P \leq 0.05 \Rightarrow$ significant evidence against H_0
- $P \leq 0.01 \Rightarrow$ highly significant evidence against H_0

Examples

- $P = .27 \Rightarrow$ non-significant evidence against H_0
- $P = .01 \Rightarrow$ highly significant evidence against H_0

*** It is *unwise* to draw firm borders for "significance"**

Significance Level

α-level

- Let α ≡ probability of erroneously rejecting H_0
- Set α threshold (e.g., let α = .10, .05, *or whatever*)
- Reject H_0 when $P \leq \alpha$
- Retain H_0 when $P > \alpha$
- Example: Set α = .10. Find $P = 0.27 \Rightarrow$ retain H_0
- Example: Set α = .01. Find $P = .001 \Rightarrow$ reject H_0

Summary

One sample Z-test

A. Hypothesis statements
H_0: $\mu = \mu_0$ vs.
H_a: $\mu \neq \mu_0$ (two-sided) or
H_a: $\mu < \mu_0$ (left-sided) or
H_a: $\mu > \mu_0$ (right-sided)

B. Test statistic

C. P-value: convert z_{stat} to P value

D. Significance statement (usually not necessary)

Practice in real example

Lake Wobegon

- Let X represent Weschler Adult Intelligence scores (WAIS)
- Typically, X ~ N(100, 15)
- Take n = 9 from Lake Wobegon population
- Data ⇒ {116, 128, 125, 119, 89, 99, 105, 116, 118}
- Calculate: x-bar = 112.8
- Does sample mean provide strong evidence that population mean μ > 100?

Practice in real example

Lake Wobegon

A. **Hypotheses:**

H_0: μ = 100 versus
H_a: μ > 100 (one-sided)
H_a: μ ≠ 100 (two-sided)

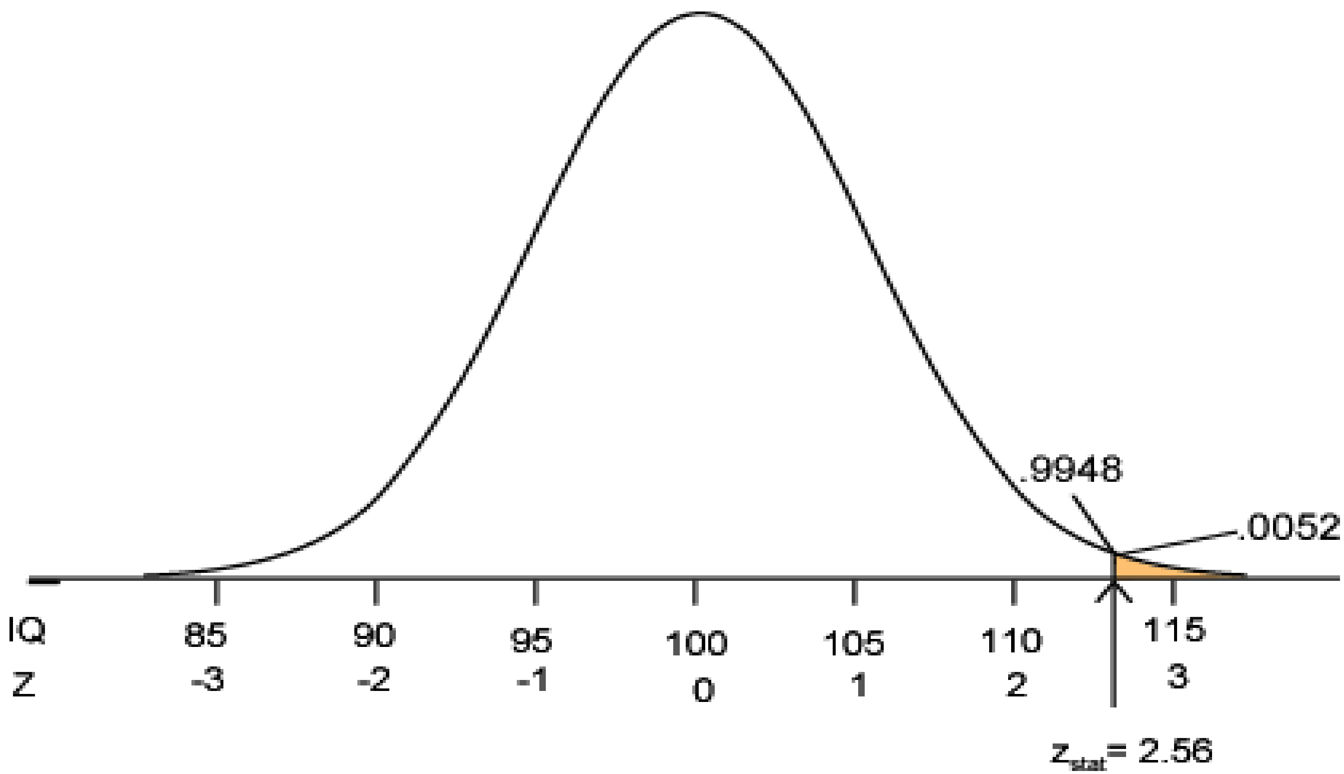
B. **Test statistic:**

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{9}} = 5$$

$$z_{\text{stat}} = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}} = \frac{112.8 - 100}{5} = 2.56$$

Practice in real example

Lake Wobegon



P =.0052 $\Rightarrow$ it is unlikely the sample came from this null distribution $\Rightarrow$ strong evidence against H_0

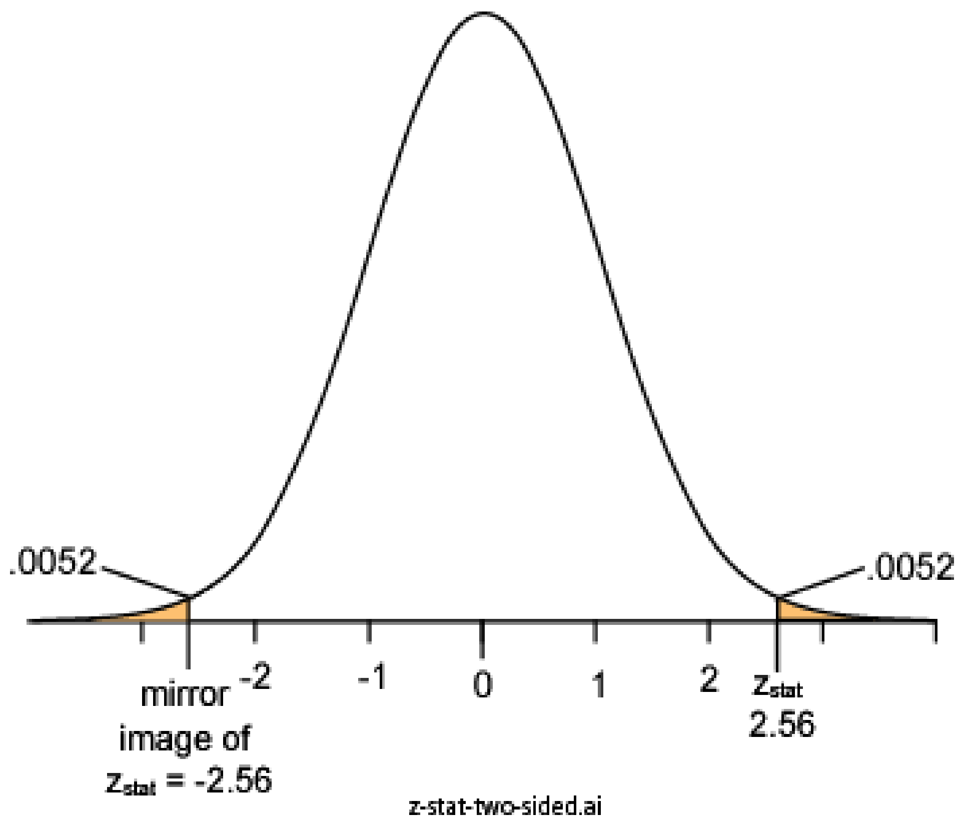
Practice in real example

Lake Wobegon

- H_a: μ ≠100

- Considers random deviations “up” and “down” from $\mu_0 \Rightarrow$ tails above and below $\pm z_{stat}$

- Thus, two-sided *P* = 2 × 0.0052 = 0.0104



Statistical Power & Sample size

Concept of "Power"

Two types of decision errors:

Type I error = erroneous rejection of true H_0

Type II error = erroneous retention of false H_0

Decision	Truth: H_0 true	Truth: H_0 false
Retain H_0	Correct retention	Type II error
Reject H_0	Type I error	Correct rejection

α ≡ probability of a Type I error

β ≡ Probability of a Type II error

Power of z-test

Concept of "Power"

$$1-\beta=\Phi\left(-z_{1-\frac{\alpha}{2}}+\frac{|\mu_0-\mu_a|\sqrt{n}}{\sigma}\right)$$

,where

- $\Phi(z)$ represent the cumulative probability of Standard Normal Z
- μ_0 represent the population mean under the null hypothesis
- μ_a represents the population mean under the alternative hypothesis

Calculation of Power

Example to calculate power in Z test

- A study of n = 16 retains H_0: μ = 170 at α = 0.05 (two-sided); σ is 40. What was the power of test's conditions to identify a population mean of 190?

$$1-\beta = \Phi\left(-z_{1-\frac{\alpha}{2}} + \frac{|\mu_0 - \mu_a|\sqrt{n}}{\sigma}\right)$$

$$= \Phi\left(-1.96 + \frac{|170-190|\sqrt{16}}{40}\right)$$

$$= \Phi(0.04)$$

$$= 0.5160$$

KOREA UNIVERSITY 1905

Calculation of Effective Sample size

Example to calculate effective sample size in Z test

- Sample size for one-sample z test:

$$n = \frac{\sigma^2 \left(z_{1-\beta} + z_{1-\frac{\alpha}{2}} \right)^2}{\Delta^2}$$

, where

$1 - \beta \equiv$ desired power

$\alpha \equiv$ desired significance level (two-sided)

$\sigma \equiv$ population standard deviation

$\Delta = \mu_0 - \mu_a \equiv$ the **difference worth detecting**

Example

Practice to calculate effective sample size

- How large a sample is needed for a one-sample *z* test with 90% power and $\alpha = 0.05$ (two-tailed) when $\sigma = 40$?

- Let H_0: $\mu = 170$ and H_a: $\mu = 190$ (thus, $\Delta = \mu_0 - \mu_a = 170 - 190 = -20$)

$$n = \frac{\sigma^2 \left(z_{1-\beta} + z_{1-\frac{\alpha}{2}}\right)^2}{\Delta^2} = \frac{40^2 (1.28 + 1.96)^2}{-20^2} = 41.99$$

- We can conclude that up to 42 samples should be required

End of Slide

Next class

- This Thursday is a national holiday, so there is no class.
- In next Monday, we will learn clustering analysis.
- If possible, we will learn additional advanced hypothesis tests at that time together.
- In next Thursday, we will do practice statistical hypothesis test and clustering analysis using R programming.