

# Midterm Exam Solutions

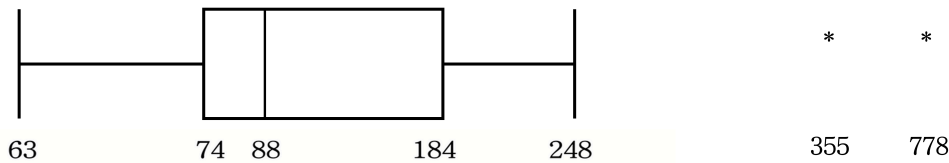
**Introduction to Statistics(통계의 이해 영어강의)**

**2025 2nd semester**

Section(교반): _____ Cadet Number(교번): _____ Name(성명): _____ Score: _____
---

- All solutions must include a detailed step-by-step explanation.
- If an answer has more than four decimal places, round to the **fourth decimal place**.

1. The following boxplot shows the population data (unit: thousand people) of 50 cities in a country in 1960. Fill in the blanks with the appropriate values. [10 points]



- (1) The median population is (                    ) thousand people.
- (2) The interquartile range (IQR) of the population is (                    ) thousand people.
- (3) The upper limit of whiskers( $Q3 + 1.5 \times IQR$ ) is (                    ) thousand people.
- (4) The number of outliers is (                    ).
- (5) The difference between the maximum population and the minimum population is (                    ) thousand people.

Solution: 88 / 110 / 349 / 2 / 715

## Reference Table

$z_{0.005} = 2.5758$	$z_{0.01} = 2.3263$	$z_{0.025} = 1.9600$	$z_{0.05} = 1.6449$
$z_{0.1} = 1.2816$	$z_{0.15} = 1.0364$	$z_{0.2} = 0.8416$	$z_{0.25} = 0.6745$
$P(Z < -2.0000) = 0.0228, Z \sim N(0, 1)$		$P(Z < -0.2273) = 0.4101, Z \sim N(0, 1)$	
$P(Z > 0.0190) = 0.4924, Z \sim N(0, 1)$		$P(Z > 0.1000) = 0.4602, Z \sim N(0, 1)$	
$P(Z > 0.5700) = 0.2843, Z \sim N(0, 1)$		$P(Z > 0.9000) = 0.1841, Z \sim N(0, 1)$	
$P(Z > 1.4000) = 0.0808, Z \sim N(0, 1)$		$P(Z > 1.5000) = 0.0668, Z \sim N(0, 1)$	

# Midterm Exam Solutions

Introduction to Statistics(통계의 이해 영어강의)

2025 2nd semester

Section(교반): \_\_\_\_\_ Cadet Number(교번): \_\_\_\_\_ Name(성명): \_\_\_\_\_ Score: \_\_\_\_\_

2. A tech company plans to launch a new project. To ensure its success, one of two external experts — A or B — will be appointed as the project director. The probabilities that A and B are appointed are 40% and 60%, respectively. If A is appointed, the probability that the project is successful is 70%; if B is appointed, the probability that the project is successful is 20%. Answer the following questions. [10 points]

(1) What is the probability that the new project is successful? (Define the relevant events before solving the problem.)

Solution:

$A$ : A is appointed as project director.

$B$ : B is appointed as project director.

$C$ : The new project is successful.

Given:  $P(A) = 0.40$ ,  $P(B) = 0.60$ ,  $P(C | A) = 0.70$ ,  $P(C | B) = 0.20$

The probability that the project is successful is

$$\begin{aligned} P(C) &= P(A)P(C | A) + P(B)P(C | B) \\ &= (0.40)(0.70) + (0.60)(0.20) = 0.28 + 0.12 = \boxed{0.40} \end{aligned}$$

Therefore, the probability that the new project will be successful is 0.4.

(2) If the project is successful, what is the probability that B was appointed as project director?

Solution:

By Bayes' Theorem:

$$\begin{aligned} P(B | C) &= \frac{P(B)P(C | B)}{P(A)P(C | A) + P(B)P(C | B)} \\ &= \frac{(0.60)(0.20)}{(0.40)(0.70) + (0.60)(0.20)} = \frac{0.12}{0.40} = \boxed{0.30} \end{aligned}$$

Thus, the probability that B was appointed given that the project is successful is 0.3.

# Midterm Exam Solutions

Introduction to Statistics(통계의 이해 영어강의)

2025 2nd semester

Section(교반): \_\_\_\_\_ Cadet Number(교번): \_\_\_\_\_ Name(성명): \_\_\_\_\_ Score: \_\_\_\_\_

3. Alex and Ben each depart from home to a meeting point at the same time. The time taken (in minutes) for Alex is  $X$ , and for Ben is  $Y$ , and their joint probability density function is given below. [20 points]

$$f(x, y) = \begin{cases} \frac{1}{200}, & 35 \leq x \leq 55, 30 \leq y \leq 40, \\ 0, & \text{otherwise.} \end{cases}$$

- (1) Find the probability that Alex arrives earlier than Ben,  $P(X < Y)$ .

Solution:

$$\begin{aligned} P(X < Y) &= \int_{y=35}^{40} \int_{x=35}^y \frac{1}{200} dx dy \\ &= \int_{35}^{40} \frac{y-35}{200} dy \\ &= \frac{(y-35)^2}{400} \Big|_{35}^{40} = \frac{25}{400} = \boxed{0.0625}. \end{aligned}$$

- (2) Suppose Alex and Ben make this trip  $n = 30$  times. Out of  $n = 30$  trips, find the expected number of times that Alex arrives earlier than Ben. (Assume that each trip is independent.)

Solution:

Let  $W$  be the number of times Alex arrives earlier than Ben in 30 trips. Then

$$W \sim B(30, 0.0625),$$

which gives

$$E(W) = 30 \times 0.0625 = 1.825.$$

- (3) Determine whether the two random variables  $X$  and  $Y$  are independent.

Solution:

$$f_X(x) = \int_{30}^{40} f(x, y) dy = \frac{1}{20}, \quad f_Y(y) = \int_{35}^{55} f(x, y) dx = \frac{1}{10}.$$

Therefore, the marginal probability density functions of  $X$  and  $Y$  are

$$f_X(x) = \begin{cases} \frac{1}{20}, & 35 \leq x \leq 55, \\ 0, & \text{otherwise,} \end{cases} \quad f_Y(y) = \begin{cases} \frac{1}{10}, & 30 \leq y \leq 40, \\ 0, & \text{otherwise.} \end{cases}$$

Since  $f(x, y) = f_X(x)f_Y(y)$  holds, the random variables  $X$  and  $Y$  are **independent**.

- (4) Find the covariance  $\text{Cov}(X, Y)$  of the two random variables  $X$  and  $Y$ .

Solution:

Since the random variables  $X$  and  $Y$  are independent,

$$\text{Cov}(X, Y) = 0.$$

# Midterm Exam Solutions

Introduction to Statistics(통계의 이해 영어강의)

2025 2nd semester

Section(교반): \_\_\_\_\_ Cadet Number(교번): \_\_\_\_\_ Name(성명): \_\_\_\_\_ Score: \_\_\_\_\_

4. Boston Children's Hospital reports that  $p = 90\%$  of American adults had chickenpox during childhood. Given that  $n = 144$  American adults are randomly selected, let a random variable  $X$  denote the number of individuals who had chickenpox during childhood. Answer the following questions. [15 points]

(1) Calculate  $P(X = 2) \times 10^{142}$ .

Solution: Since  $X \sim B(144, 0.90)$ ,

$$P(X = 2) \times 10^{142} = \binom{144}{2} (0.90)^2 (0.10)^{142} \times 10^{142} \approx \boxed{8339.76}.$$

(2) Find  $E(X)$  and  $Var(X)$ .

The expectation and variance are:

$$E(X) = np = 144 \times 0.90 = \boxed{129.6},$$

$$Var(X) = np(1 - p) = 144 \times 0.90 \times 0.10 = \boxed{12.96} (= 3.6^2).$$

(3) Find the probability that 135 or more individuals had chickenpox during childhood. Assume that the conditions for the Central Limit Theorem are satisfied.

Solution:

Using the normal approximation:

$$X \sim N(129.6, 3.6^2).$$

Standardizing:

$$Z = \frac{X - 129.6}{3.6} \sim N(0, 1).$$

Then

$$P(X \geq 135) = P\left(Z \geq \frac{135 - 129.6}{3.6}\right) = P(Z \geq 1.5) \approx \boxed{0.0668}.$$

# Midterm Exam Solutions

Introduction to Statistics(통계의 이해 영어강의)

2025 2nd semester

Section(교반): \_\_\_\_\_ Cadet Number(교번): \_\_\_\_\_ Name(성명): \_\_\_\_\_ Score: \_\_\_\_\_

5. To test whether the birth rates of boys and girls are equal in a region,  $n = 100$  newborns were randomly sampled from the region. We conduct a **two-sided hypothesis test for a proportion** to check whether the boys' birth rate( $p$ ) is different from  $p_0 = 0.5$ , using a significance level of  $\alpha = 0.05$ . [15 points]

(1) State the null and alternative hypotheses.

Solution:

$$H_0 : p = 0.5, \quad H_A : p \neq 0.5$$

(2) Find the null distribution of the test statistic. Assume that conditions for the Central Limit Theorem are satisfied.

Solution:

$$\hat{p} \sim N\left(0.5, \frac{0.5(1-0.5)}{100} = 0.05^2\right)$$

or equivalently,

$$Z = \frac{\hat{p} - 0.5}{0.05} \sim N(0, 1)$$

(3) In the random sample of 100 newborns, 43 were boys. Compute the observed test statistic.

Solution:

$$\hat{p} = \frac{43}{100} = 0.43$$
$$z = \frac{0.43 - 0.5}{0.05} = -1.4$$

(4) Compute the p-value and complete the hypothesis test. State the conclusion in the context of data.

Solution:

$$p\text{-value} = 2 \cdot P(Z < -1.4) \approx 0.1616$$

Since the p-value is larger than  $\alpha = 0.05$ , we fail to reject the null hypothesis.

**Conclusion:** There is not enough evidence to conclude that the birth rate of boys differs significantly from 50% in region A.

# Midterm Exam Solutions

**Introduction to Statistics(통계의 이해 영어강의)**

**2025 2nd semester**

Section(교반): \_\_\_\_\_ Cadet Number(교번): \_\_\_\_\_ Name(성명): \_\_\_\_\_ Score: \_\_\_\_\_

6. From a population having the following probability distribution, a random sample  $X_1$  and  $X_2$  of size  $n = 2$  is taken.

$x$	1	2	3
$P(X = x)$	0.6	0.2	0.2

The joint probability distribution table of the two random variables  $X_1$  and  $X_2$ ,  $P(X_1 = x_1, X_2 = x_2)$  is given below.

Answer the following questions. [15 points]

$x_1 \backslash x_2$	1	2	3
1	$0.6 \times 0.6 = 0.36$	$0.6 \times 0.2 = 0.12$	$0.6 \times 0.2 = 0.12$
2	$0.2 \times 0.6 = 0.12$	$0.2 \times 0.2 = 0.04$	$0.2 \times 0.2 = 0.04$
3	$0.2 \times 0.6 = 0.12$	$0.2 \times 0.2 = 0.04$	$0.2 \times 0.2 = 0.04$

(1) Fill in the probability distribution table of the sample mean  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ .

$\bar{x}$					Otherwise
$P(\bar{X} = \bar{x})$					0

Solution:  $P(\bar{X} = 1) = 0.36$ ,  $P(\bar{X} = 1.5) = 0.24$ ,  $P(\bar{X} = 2) = 0.28$ ,  $P(\bar{X} = 2.5) = 0.08$ ,  $P(\bar{X} = 3) = 0.04$

(2) For the sample variance  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ , show that  $S^2 = \frac{1}{2} (X_1 - X_2)^2$ .

Solution:

$$\begin{aligned}
 S^2 &= \frac{1}{2-1} \left[ (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 \right] \\
 &= (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 \\
 &= \left( X_1 - \frac{X_1 + X_2}{2} \right)^2 + \left( X_2 - \frac{X_1 + X_2}{2} \right)^2 \\
 &= \frac{1}{4} \left[ (X_1 - X_2)^2 + (X_2 - X_1)^2 \right] \\
 &= \frac{1}{2} (X_1 - X_2)^2.
 \end{aligned}$$

(3) The probability distribution table of the sample variance  $S^2$  is given below.

$s^2$	0	0.5	2	Otherwise
$P(S^2 = s^2)$	0.44	0.32	0.24	0

Determine whether the sample mean  $\bar{X}$  and the sample variance  $S^2$  are independent.

Solution: Note that

$$P(\bar{X} = 1, S^2 = 0) = P(X_1 = 1, X_2 = 1) = 0.36.$$

but  $P(\bar{X} = 1) = 0.36$ ,  $P(S^2 = 0) = 0.44$ . Since  $P(\bar{X} = 1, S^2 = 0) \neq P(\bar{X} = 1) \times P(S^2 = 0)$ ,  $\bar{X}$  and  $S^2$  are **not independent**.

# Midterm Exam Solutions

**Introduction to Statistics(통계의 이해 영어강의)**

**2025 2nd semester**

Section(교반): \_\_\_\_\_ Cadet Number(교번): \_\_\_\_\_ Name(성명): \_\_\_\_\_ Score: \_\_\_\_\_

7. Suppose we obtain a random sample  $X_1, X_2, \dots, X_n$  from the population with probability density function

$$f(x) = \begin{cases} \frac{3x^2}{\theta^3}, & 0 < x < \theta \quad (\theta > 0), \\ 0, & \text{otherwise.} \end{cases}$$

For the estimator  $\hat{\theta} = \frac{4}{3}\bar{X}_n = \frac{4}{3n} \sum_{i=1}^n X_i$ , answer the following questions. [15 points]

(1) Show that the estimator  $\hat{\theta}$  is an unbiased estimator of  $\theta$ .

Solution:

$$E(\hat{\theta}) = \frac{4}{3}E(\bar{X}_n) = \frac{4}{3}E(X_1) = \frac{4}{3} \int_0^\theta x \frac{3x^2}{\theta^3} dx = \frac{4}{3} \times \frac{3}{4} \theta = \theta.$$

Therefore,  $\hat{\theta}$  is an unbiased estimator of  $\theta$ .

(2) Show that the estimator  $\hat{\theta}$  is a consistent estimator of  $\theta$ .

Solution: From (1), since  $E(\hat{\theta}) = \theta$ ,  $\lim_{n \rightarrow \infty} E(\hat{\theta}) = \lim_{n \rightarrow \infty} \theta = \theta$ . Also,

$$E(X_1^2) = \int_0^\theta x^2 \frac{3x^2}{\theta^3} dx = \frac{3}{5} \theta^2, \quad \text{Var}(X_1) = E(X_1^2) - [E(X_1)]^2 = \frac{3}{5} \theta^2 - \left(\frac{3}{4} \theta\right)^2 = \frac{3}{80} \theta^2 < \infty.$$

$$\text{Var}(\hat{\theta}) = \frac{16}{9} \text{Var}(\bar{X}_n) = \frac{16}{9n} \text{Var}(X_1) = \frac{16}{9n} \cdot \frac{3}{80} \theta^2 = \frac{\theta^2}{15n} \implies \lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}) = \lim_{n \rightarrow \infty} \frac{\theta^2}{15n} = 0.$$

Therefore,  $\hat{\theta}$  is a consistent estimator of  $\theta$ .

(3) From a random sample of size  $n = 60$ , the observed sample mean is  $\bar{x} = 3$ . Construct an approximate 95% confidence interval for  $\theta$ . (Hint: Consider the sampling distribution of  $\hat{\theta}$  using the Central Limit Theorem.)

Solution: Since  $n = 60$  is large, by the Central Limit Theorem we have  $\hat{\theta} \sim N(\theta, \theta^2/(15n))$ , or

$$Z = \frac{\hat{\theta} - \theta}{\theta/\sqrt{15n}} = \frac{\hat{\theta} - \theta}{\theta/30} \sim N(0, 1)$$

$$\begin{aligned} \therefore 1 - \alpha &\approx P\left(-z_{\alpha/2} \leq \frac{\hat{\theta} - \theta}{\frac{\theta}{30}} \leq z_{\alpha/2}\right) = P\left(\theta - \frac{\theta}{30} z_{\alpha/2} \leq \hat{\theta} \leq \theta + \frac{\theta}{30} z_{\alpha/2}\right) \\ &= P\left(\frac{\hat{\theta}}{1 + z_{\alpha/2}/30} \leq \theta \leq \frac{\hat{\theta}}{1 - z_{\alpha/2}/30}\right). \end{aligned}$$

Putting  $\hat{\theta} = 4\bar{x}/3 = 4$  and  $z_{\alpha/2} = z_{0.025} = 1.96$ , a 95% confidence interval for  $\theta$  is  $\boxed{[3.7547, 4.2796]}$

Using  $\frac{\hat{\theta} - \theta}{\theta/30} \sim N(0, 1)$ , one can instead get  $\left[\hat{\theta} - \frac{\hat{\theta}}{30} z_{\alpha/2}, \hat{\theta} + \frac{\hat{\theta}}{30} z_{\alpha/2}\right] = \boxed{[3.7387, 4.2613]}$ .