<u>Midterm Exam Solutions</u>

<u>Introduction to Statistics(통계의 이해 영어강의)</u> **<u>2025 1st semester</u>**

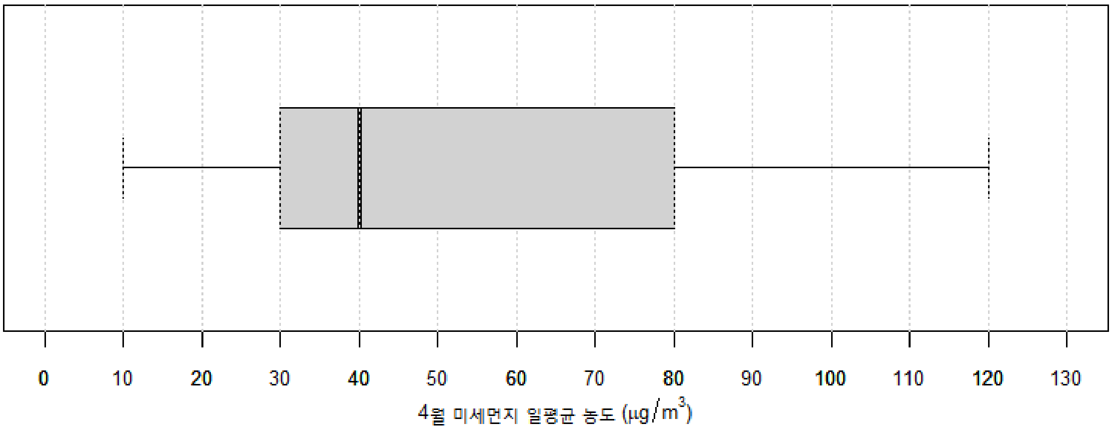
Section(교반): ______	Cadet Number(교번): ___________	Name(성명): _______________	Score: _____________

- 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**.
- Reference table is provided on the last page of the exam.

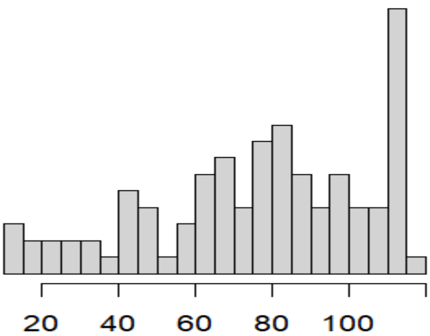
1. Daily air quality index(AQI) was reported for a sample of 91 days in 2024 in a certain city. Summary statistics of daily AQI is provided below. Answer the following questions. [10 points]

Minimum	Q1	Median	Mean	Q3	Maximum
10	30	40	50	80	120

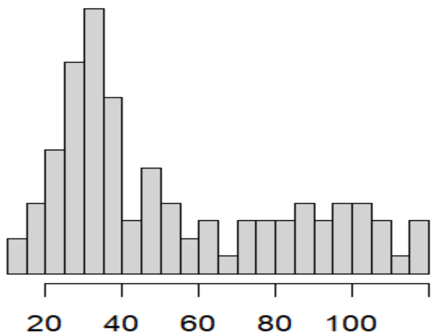
(1) Draw a boxplot for the daily AQI of this sample. (There is no outlier in the sample.)



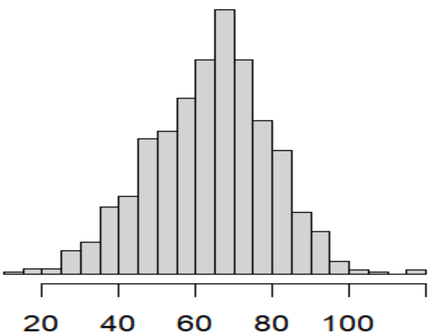
(2) Identify which of the following corresponds to the histogram of the daily AQI in the sample. Answer: B



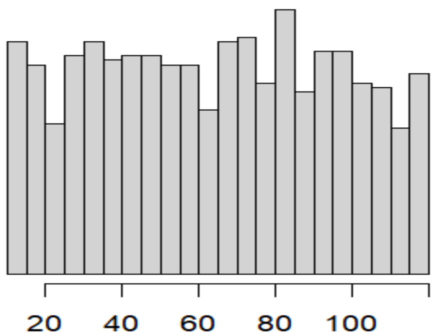
(A)



(B)



(C)



(D)

Midterm Exam Solutions

Introduction to Statistics(통계의 이해 영어강의) **2025 1st semester**

Section(교반): _____ Cadet Number(교번): __________ Name(성명): ______________ Score: ____________

2. According to an analysis by a national intelligence agency, a recent cyberattack was carried out by either Hacker Group A or Hacker Group B. The probabilities that the attack was conducted by Group A and Group B are 40% and 60%, respectively. After analyzing the attack logs, investigators found signs of unauthorized access in some cases. The probability of such signs appearing is 80% if Group A was responsible, and 30% if it was Group B. Answer the following questions. [10 points]

(1) What is the probability that signs of unauthorized access appear in the attack logs? (Define the relevent events before solving the problem.)

Solution:

A is the event that the attack was conducted by Group A.

B is the event that the attack was conducted by Group B.

C is the event that signs of unauthorized access appear in the logs.

Given: $P(A) = 0.40$, $P(B) = 0.60$, $P(C \mid A) = 0.80$, $P(C \mid B) = 0.30$

The probability that signs of unauthorized access appear in the attack logs is

$$P(C) = P(C \mid A) \cdot P(A) + P(C \mid B) \cdot P(B) = (0.80)(0.40) + (0.30)(0.60) = 0.32 + 0.18 = \boxed{0.50}$$

(2) Given that signs of unauthorized access were found, what is the probability that the actual attacker was Group A?

Solution:

By Bayes' Theorem:

$$P(A \mid C) = \frac{P(C \mid A) \cdot P(A)}{P(C)} = \frac{(0.80)(0.40)}{0.50} = \frac{0.32}{0.50} = \boxed{0.64}$$

national intelligence agency: 국가정보기관, cyberattack: 사이버 공격, signs of unauthorized access: 무단접근흔적

Midterm Exam Solutions

Introduction to Statistics(통계의 이해 영어강의) **2025 1st semester**

Section(교반): _____ Cadet Number(교번): __________ Name(성명): ______________ Score: ____________

3. The joint probability mass function (pmf) of two random variables X and Y is given by:

$$f(x,y) = \begin{cases} \dfrac{x^2 y}{42}, & x = 1,2,3, \quad y = 1,2 \\ 0, & \text{otherwise} \end{cases}$$

Answer the following questions. [15 points]

(1) Find the marginal probability mass function of X, $f_X(x)$, and the marginal probability mass function of Y, $f_Y(y)$.

Solution:

$$f_X(x) = \sum_{y=1}^{2} \frac{x^2 y}{42} = \frac{1}{14}x^2, \quad x = 1,2,3$$

$$f_Y(y) = \sum_{x=1}^{3} \frac{x^2 y}{42} = \frac{1}{3}y, \quad y = 1,2$$

(2) Determine whether the two random variables X and Y are independent.

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

(3) Find the expectation of Y, $E(Y)$.

Solution:

$$E(Y) = \sum_{y=1}^{2} y f_Y(y) = \sum_{y=1}^{2} \frac{y^2}{3}$$
$$= \frac{1+4}{3} = \frac{5}{3}$$

Midterm Exam Solutions

Introduction to Statistics(통계의 이해 영어강의) **2025 1st semester**

Section(교반): _____	Cadet Number(교번): __________	Name(성명): ______________	Score: ____________

4. The monthly income of residents in a certain region follows a normal distribution $N(200, 200)$ with a mean of $\mu = 200$ and variance of $\sigma^2 = 200$. Two residents are randomly selected, and their monthly incomes X_1 and X_2 are observed. Answer the following questions. (Assume that X_1 and X_2 are independent.) [10 points]

(1) Find the expectation and variance of the average income $\bar{X} = \dfrac{X_1 + X_2}{2}$, $E(\bar{X})$ and $\mathrm{Var}(\bar{X})$.

Solution:

Since $X_1, X_2 \sim N(200, 200)$ and are independent:

$$E(\bar{X}) = E\left(\frac{X_1 + X_2}{2}\right) = \frac{E(X_1) + E(X_2)}{2} = \frac{200 + 200}{2} = \boxed{200}$$

$$\mathrm{Var}(\bar{X}) = \mathrm{Var}\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{4}(\mathrm{Var}(X_1) + \mathrm{Var}(X_2)) = \frac{1}{4}(200 + 200) = \boxed{100}$$

(2) Compute the probability that the average income $\bar{X} = \dfrac{X_1 + X_2}{2}$. is less than 180, $P(\bar{X} < 180)$.

Solution:

We found that $\bar{X} \sim N(200, 100)$, so the standard deviation is:

$$SD(\bar{X}) = \sqrt{100} = 10$$

We standardize to compute the probability:

$$\begin{aligned} P(\bar{X} < 180) &= P(\frac{X - 200}{10} < \frac{180 - 200}{10}) \\ &= P(Z < -2) = \boxed{0.0228} \end{aligned}$$

Midterm Exam Solutions

Introduction to Statistics(통계의 이해 영어강의) **2025 1st semester**

Section(교반): _____ Cadet Number(교번): __________ Name(성명): ______________ Score: ____________

5. A factory produces aircraft parts with a known defect rate of $p = 0.02$. Let a random variable X denote the number of defective aircraft parts in a random sample of $n = 1000$. Answer the following questions. [15 points]

(1) Calculate $P(X = 0) \times 10^9$

Solution: We have $X \sim B(1000, 0.02)$.

$$P(X = 0) = \binom{1000}{0}(0.02)^0(0.98)^{1000} = 1.6830 \times 10^{-9}.$$

Therefore, $P(X = 0) \times 10^9 = 1.683$

(2) Let $\hat{p} = \dfrac{X}{1000}$ denote the sample proportion of defective parts. Find the expectation and standard error of $\hat{p}$, $E(\hat{p})$ and $SE(\hat{p})$.

Solution:

$$E(\hat{p}) = p = \boxed{0.02}$$

$$SE(\hat{p}) = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.02(1-0.02)}{1000}} = \boxed{0.0044}$$

(3) Find $P(\hat{p} < 0.019)$ using normal approxiamtion of $\hat{p}$. (Check normality condition if necessary.)

Solution:

Independence: The sample is random. ⇒ satisfied.

Success-failure condition: $1000 \cdot 0.02 = 20 \geq 10$, $1000 \cdot (1 - 0.02) = 980 \geq 10$ ⇒ satisfied.

Therefore, the sampling distribution of $\hat{p}$ is:

$$\hat{p} \dot{\sim} N\left(0.02, \frac{0.02(1-0.02)}{1000} = 0.0044^2\right)$$

Therefore,

$$\begin{aligned} P(\hat{p} < 0.019) &= P\left(\frac{\hat{p} - 0.02}{0.0044} < \frac{0.019 - 0.02}{0.0044}\right) \\ &\approx P\left(Z < -0.2273\right) = 0.4101. \end{aligned}$$

Midterm Exam Solutions

Introduction to Statistics(통계의 이해 영어강의) **2025 1st semester**

Section(교반): _____ Cadet Number(교번): __________ Name(성명): ______________ Score: ____________

6. To test whether the birth rates of boys and girls are equal in a certain region A, 100 newborns were randomly sampled from the region A. We conduct **a 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]

(a) State the null and alternative hypothesis.

Solution:

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

(b) Find the null distribution of the test statistic. Check any conditions required.

Solution:

Independence: The sample is random. ⇒ satisfied.

Success-failure condition: $100 \cdot 0.5 = 50 \geq 10$, $100 \cdot (1 - 0.5) = 50 \geq 10$ ⇒ condition met.

Under H_0, the sampling distribution of $\hat{p}$ is:

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

or equivalently,

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

(c) In the random sample of 100 newborns, 57 were boys. Compute the observed test statistic.

Solution:

$$\hat{p} = \frac{57}{100} = 0.57$$

$$z = \frac{0.57 - 0.5}{0.05} = 1.4$$

(d) 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 1st semester**

Section(교반): _____ Cadet Number(교번): __________ Name(성명): ______________ Score: ____________

7. To test whether the birth rates of boys and girls are equal in a certain region A, 100 newborns were randomly sampled from the region A. We conduct **a hypothesis test for a goodness of fit** to check whether the boys' birth rate(p_1) is different from girls' birth rate(p_2), using a significance level of $\alpha = 0.05$. [15 points]

	Boys	**Girls**	**Total**
Observed	O_1	O_2	100
Expected	E_1	E_2	100

(a) State the null and alternative hypothesis.

Solution:

$$H_0 : p_1 = p_2 (= 0.5), \qquad H_A : p_1 \neq p_2$$

(b) Find the null distribution of the test statistic: $\sum_{i=1}^{2} \frac{(O_i - E_i)^2}{E_i}$

Solution:

Under H_0, the sampling distribution of the test statistic is:

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \dot{\sim} \chi^2(1)$$

(c) In the random sample of 100 newborns, $57 (= O_1)$ were boys $43 (= O_2)$ were girls. Compute the **observed** test statistic of $\sum_{i=1}^{2} \frac{(O_i - E_i)^2}{E_i}$.

Solution:

$$\chi^2 = \frac{(57 - 50)^2}{50} + \frac{(43 - 50)^2}{50} = 1.96$$

(d) Compute the p-value and complete the hypothesis test. Compare this with the p-value in Problem 6(d).

Solution:

$$p\text{-value} = P(\chi^2 > 1.96) \approx 0.1616$$

Since the p-value is larger than $\alpha = 0.05$, we fail to reject the null hypothesis. The p-value is exactly same as the p-value in 5(d).

(e) Find the relationship between the null distribution in 6(b) and 7(b). Verify that the observed test statistics in 6(c) and 7(c) follow the same relationship.

Solution:

$$Z^2 = \chi^2, \qquad \chi^2 \sim \chi^2(1).$$

The observed test statistic in 5(c) is 1.4 whereas the observed test statistic in 6(c) is 1.96. It can be seen that

$$1.4^2 = 1.96.$$

Midterm Exam Solutions

Introduction to Statistics(통계의 이해 영어강의) **2025 1st semester**

Section(교반): _____ Cadet Number(교번): __________ Name(성명): ______________ Score: ____________

8. During a physical fitness test, a random sample of $n = 10$ cadets was selected. The number of push-ups completed in 2 minutes were reported; the sample mean $\bar{x}$ was 58 and the sample standard deviation s was 8. Construct a 95% confidence interval for the mean number of push-ups completed by all cadets in 2 minutes. Assume that the normality condition is met. [10 points]

Solution:

We are given:

$$n = 10, \quad \bar{x} = 58, \quad s = 8.$$

To construct a 95% confidence interval for the mean, we use the t-distribution.

The degrees of freedom is $df = n - 1 = 9$.

Therefore, the 95% confidence interval is:

$$\begin{aligned}\bar{x} \pm t_{0.025}(9) \cdot \frac{s}{\sqrt{n}} &= 58 \pm 2.2622 \cdot \frac{8}{\sqrt{10}} \\ &\approx 58 \pm 5.7230 = (52.277, 63.723)\end{aligned}$$

Reference Table

$z_{0.005} = 2.5758$	$z_{0.01} = 2.3263$	$z_{0.025} = 1.9600$	$z_{0.05} = 1.6449$
$t_{0.005}(9) = 3.2498$	$t_{0.01}(9) = 2.8214$	$t_{0.025}(9) = 2.2622$	$t_{0.05}(9) = 1.8331$
$t_{0.005}(10) = 3.1693$	$t_{0.01}(10) = 2.7638$	$t_{0.025}(10) = 2.2281$	$t_{0.05}(10) = 1.8125$
pnorm(-2.0000, lower.tail = TRUE) = 0.0228		pnorm(-1.4142, lower.tail = TRUE) = 0.0787	
pnorm(-0.2273, lower.tail = TRUE) = 0.4101		pnorm(0.0190, lower.tail = TRUE) = 0.5076	
pnorm(0.5700, lower.tail = FALSE) = 0.2843		pnorm(1.4000, lower.tail = FALSE) = 0.0808	
pchisq(0.98, 1 , lower.tail = FALSE) = 0.3221		pchisq(1.96, 1 , lower.tail = FALSE) = 0.1616	
pchisq(0.98, 2 , lower.tail = FALSE) = 0.6126		pchisq(1.96, 2 , lower.tail = FALSE) = 0.3753	