Introduction to Statistics - Quiz #4(70 minutes)

December 23, 2025 (Tuesday)

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. A health researcher wants to test whether people who exercise regularly sleep **more** on average than people who do not exercise regularly. Group 1 consists of people who exercise regularly, while Group 2 consists of people who do not exercise regularly. To test this at the significance level $\alpha = 0.05$, the researcher samples $n_1 = 8$ people from Group 1 and $n_2 = 18$ people from Group 2. Assume the daily sleep durations (in hours) follow:

$$X_{11}, \cdots, X_{1,8} \sim N(\mu_1, 1), \qquad X_{21}, \cdots, X_{2,18} \sim N(\mu_2, 2.25),$$

where μ_1 and μ_2 denote the mean daily sleep duration for Group 1 and Group 2, respectively. [40 points]

(1) State the null and alternative hypotheses. (Use a **one-sided test**.)

Solution: The claim is that Group 1 sleeps more on average, so the alternative is $\mu_1 - \mu_2 > 0$.

$$H_0 : \mu_1 - \mu_2 = 0, \qquad H_A : \mu_1 - \mu_2 > 0$$

(2) Find the test statistic and its null distribution.

Solution: The sampling distributions are $\bar{X}_1 \sim N(\mu_1, 1/8)$, $\bar{X}_2 \sim N(\mu_2, 2.25/18)$. Thus,

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2,\ \frac{1}{8} + \frac{2.25}{18}\right).$$

Since $1/8 + 2.25/18 = 0.25$, under $H_0 : \mu_1 - \mu_2 = 0$,

$$\bar{X}_1 - \bar{X}_2 \sim N(0,\ 0.5^2).$$

(3) Find the rejection region.

Solution: For a right-sided test at $\alpha = 0.05$, the critical value is $z_{0.05} = 1.645$. Reject H_0 if

$$\bar{x}_1 - \bar{x}_2 > z_{0.05} \cdot SE = (1.645)(0.5) = 0.8225.$$

Thus, the rejection region is $\bar{x}_1 - \bar{x}_2 > 0.8225$.

(4) Given $\bar{x}_1 = 8.2$ and $\bar{x}_2 = 7.0$, complete the hypothesis test. State the conclusion in the context of the data.

Solution:

$$\bar{x}_1 - \bar{x}_2 = 8.2 - 7.0 = 1.2.$$

Since $1.2 > 0.8225$, the observed difference lies in the rejection region. Therefore, we reject H_0.

Conclusion: At the significance level $\alpha = 0.05$, there is sufficient evidence to conclude that people who exercise regularly sleep more (on average) than people who do not exercise regularly.

(5) Find the power of the test assuming $\mu_1 - \mu_2 = 1$.

Solution:

$$\begin{aligned} P(\text{Reject } H_0 \mid H_A \text{ is true}) &= P(\bar{X}_1 - \bar{X}_2 > 0.8225 \mid \mu_1 - \mu_2 = 1) \\ &= P\left(\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{0.5} > \frac{0.8225 - 1}{0.5}\right) \\ &= P\left(Z > -0.355\right) \approx \boxed{0.6387}. \end{aligned}$$

2. A sample of $n = 42$ adults was selected from an insurance study. For each individual, their age and annual medical charges (in hundreds of dollars) were recorded. We are interested in investigating whether age (x) is associated with annual medical charges (y). Conduct a hypothesis test on the correlation coefficient ρ to determine **if there is a correlation between age and annual medical charges.** (Assume that all the necessary conditions are satisfied.) [30 points]

(1) State the null and alternative hypotheses. (Use a **two-sided** test)

Solution:

$$H_0 : \rho = 0, \qquad H_A : \rho \neq 0$$

(2) Find the test statistic and its null distribution.

Solution:

Under H_0, the test statistic and its distribution are

$$T = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}} \sim t(n-2)$$

where $n = 42$.

(3) Compute the sample correlation coefficient r and the observed test statistic using the following summary statistics.

$s_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$	$s_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2$	$s_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$
361	576	319.2

Solution:

First, compute the sample correlation coefficient:

$$r = \frac{s_{xy}}{\sqrt{s_{xx} \cdot s_{yy}}} = \frac{319.2}{\sqrt{361 \cdot 576}} = \frac{319.2}{456} = 0.70$$

Then compute the observed test statistic:

$$t = \frac{0.70 \cdot \sqrt{40}}{\sqrt{1-(0.70)^2}} \approx \boxed{6.1993}$$

(4) Compute the p-value and complete the hypothesis test.

Solution:

The p-value for a two-sided test is

$$p\text{-value} = 2 \times P(T > 6.1993) \approx \boxed{2.4792 \times 10^{-7}}.$$

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

Conclusion: There is strong evidence of a correlation between age and annual medical charges.

3. A travel analyst collected data on $n = 30$ flights. For each flight, the analyst recorded the flight distance (x, in miles) and the flight time (y, in minutes). We will use linear regression to predict flight time from distance.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \cdots, n.$$

The summary statistics are as follows. [30 points]

$\bar{x}$	s_x	$\bar{y}$	s_y	r
1400	480	210	60	0.64

(1) Compute the least squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$.

Solution:

$$\hat{\beta}_1 = r \cdot \frac{s_y}{s_x} = 0.64 \cdot \frac{60}{480} = 0.08$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 210 - 0.08(1400) = 98$$

(2) A flight has distance $x_i = 1000$ miles and actual flight time $y_i = 190$ minutes. Use the model to predict the mean flight time $\hat{y}_i$ at $x_i = 1000$ and compute the residual e_i.

Solution:

$$\hat{y}_i = 98 + 0.08(1000) = 178 \text{ minutes}$$

$$e_i = y_i - \hat{y}_i = 190 - 178 = 12 \text{ minutes}$$

(3) Calculate the coefficient of determination R^2 and the sum of squared errors $SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.

Solution:

$$R^2 = r^2 = (0.64)^2 = 0.4096$$

$$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2 = (n-1)s_y^2 = 29(60^2) = 104400$$

$$SSE = SST(1 - R^2) = 104400(1 - 0.4096) = 61637.76$$

Reference Table

$z_{0.005} = 2.5758$	$z_{0.01} = 2.3263$	$z_{0.025} = 1.9600$	$z_{0.05} = 1.6450$
$P(Z > -0.0400) = 0.5160, \quad Z \sim N(0,1)$		$P(Z > -0.3550) = 0.6387, \quad Z \sim N(0,1)$	
$P(Z < -2.0400) = 0.9793, \quad Z \sim N(0,1)$		$P(Z < -2.3548) = 0.9907, \quad Z \sim N(0,1)$	
$P(T > 0.7000) = 0.2439, \quad T \sim t(40)$		$P(T > 0.7000) = 0.2439, \quad T \sim t(41)$	
$P(T > 6.1993) = 1.2396 \times 10^{-7}, \quad T \sim t(40)$		$P(T > 6.1993) = 1.1261 \times 10^{-7}, \quad T \sim t(41)$	
$P(T > 8.6808) = 4.8380 \times 10^{-11}, \quad T \sim t(40)$		$P(T > 8.6808) = 3.9286 \times 10^{-11}, \quad T \sim t(41)$	