Introduction to Statistics - Quiz #3(50 minutes)

May 16, 2025 (Friday)

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. Read the following statements and choose the most appropriate word to complete the sentence. [15 points]

(1) As the sample size increases, the variance of the sample mean becomes (larger / smaller).

(2) The t distribution is a (left-skewed / symmetric / right-skewed) distribution.

(3) Compared to the standard normal(Z) distribution , the t distribution has (thicker / thinner) tail.

(4) As the degrees of freedom increase, the tails of the t distribution become (thicker / thinner).

(5) In a two-sample t-test, if the population variances are assumed to be (equal / unequal), the degrees of freedom are calculated as $n_1 + n_2 - 2$, where n_1 and n_2 represent the sample sizes of the two groups.

Solution: smaller / symmetric / thicker / thinner / equal

2. The average storage capacity of USB drives produced by Company A, denoted by μ, used to be reported as 16GB. Recently, some consumers have claimed that the average capacity of Company A's USB drives is **less** than 16GB. To verify the validity of this claim, a random sample of 9 USB drives was selected, and their capacities (in GB) were measured. We conduct **a hypothesis test for a population mean** at the significance level $\alpha = 0.05$. (Assume that the storage capacities of USB drives produced by Company A follow a normal distribution.) [45 points]

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

Solution:

$$H_0 : \mu = 16, \qquad H_A : \mu < 16$$

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

Solution:

Under H_0, the test statistic follows:

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{\bar{X} - 16}{S/\sqrt{9}} \overset{H_0}{\sim} t(8)$$

where $\bar{X}$ is the sample mean, and S is the sample standard deviation.

(c) The sample mean and the sample standard deviation of storage capacity from the 9 USB drives are $\bar{x} = 15.84$ and $s = 0.24$. Compute the observed test statistic.

Solution:

$$T = \frac{15.84 - 16}{0.24/\sqrt{9}} = \frac{-0.16}{0.08} = \boxed{-2}$$

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

Solution:

$$p\text{-value} = P(T < -2) \approx \boxed{0.0403}$$

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

Conclusion: There is statistically significant evidence that the average storage capacity of Company A's USB drives is less than 16GB.

3. In region G, corn yields from fertilizers A and B are normally distributed, with means μ_A and μ_B and variances σ_A^2 and σ_B^2, respectively. Recently, some farmers have claimed that the mean yields μ_A and μ_B are **different**. To investigate, researchers selected 12 locations with similar soil quality, randomly applying fertilizer A to 6 sites and fertilizer B to the other 6 sites. Based on the corn yields (unit: kg) from the treated soils, we conduct **a hypothesis test for two means** at the significance level $\alpha = 0.05$. [40 points]

(1) State the null and alternative hypotheses.

Solution:

$$H_0 : \mu_A - \mu_B = 0, \qquad H_A : \mu_A - \mu_B \neq 0$$

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

Solution:

Under H_0, the test statistic is given by

$$T = \frac{\bar{X}_A - \bar{X}_B - 0}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{S_A^2}{6} + \frac{S_B^2}{6}}} \overset{H_0}{\sim} t(\psi),$$

where $\bar{X}_A$ and $\bar{X}_B$ are the sample means of corn yields from fertilizers A and B, S_A and S_B are the sample standard deviations, and ψ is the Satterthwaite's degrees of freedom.

(3) 12 corn yields (kg) were measured after applying fertilizers A and B. Using the R code and R output provided, write down the p-value and complete the hypothesis test. State the conclusion in the context of the data.

```
fertA <- c(580, 600, 600, 550, 650, 620)
fertB <- c(510, 650, 570, 670, 630, 510)
t.test(fertA, fertB, alternative = "two.sided", mu = 0, var.equal = FALSE)
```

```
Welch Two Sample t-test

data:  fertA and fertB
t = 0.31311, df = 7.2174, p-value = 0.7631
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -65.06129  85.06129
sample estimates:
mean of x mean of y
      600       590
```

Solution:

The observed test statistic is $t = 0.31311$, and the p-value is 0.7631.

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

Conclusion: There is insufficient evidence to conclude a difference in mean corn yields between fertilizers A and B.

Reference Table

Satterthwaite's df : $\psi = \frac{(s_A^2/n_A + s_B^2/n_B)^2}{(s_A^2/n_A)^2/(n_A-1) + (s_B^2/n_B)^2/(n_B-1)}$, $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$	
pt(-2, lower.tail = TRUE, df = 8) = 0.0403	pt(-2, lower.tail = FALSE, df = 8) = 0.9597
pt(-2, lower.tail = TRUE, df = 9) = 0.0383	pt(-2, lower.tail = FALSE, df = 9) = 0.9617
pnorm(-2, lower.tail = TRUE) = 0.0228	pnorm(-2, lower.tail = FALSE) = 0.9772