112-2 Statistics (II)

Midterm Solution

Spring 2024

1. (A)

2. (C)

3. In the matched sample design the two production methods are tested under similar conditions (i.e., with the same workers) ; hence this design often leads to a smaller **sampling error** than the independent sample design. The primary reason is that in a matched sample design, **variation** between workers is eliminated because the same workers are used for both production methods.

4. Whenever independent simple random samples of sizes n_1 and n_2 are selected from two **normal** populations with **equal variances**, the sampling distribution of s_1^2/s_2^2 is an $\mathcal{F}$ distribution with $n_1 - 1$ degrees of freedom for the numerator and $n_2 - 1$ degrees of freedom for the denominator; where s_1^2 (s_2^2) is the sample variance for the random sample of n_1 (n_2) items from population 1 (2).

5. (a) The interval estimation and hypothesis testing procedures are robust and can be used with **relatively small sample sizes**.

   (b) In most applications, equal or nearly equal sample sizes such that the total sample size $\boldsymbol{n_1 + n_2 \geq 20}$ can be expected to provide very good results even if the populations are not normal.

   (c) Larger sample sizes are recommended if the distributions of the populations are **highly skewed** or contain **outliers**.

   (d) Smaller sample sizes should only be used if the analyst is satisfied that the distributions of the populations are at least **approximately normal**.

6. (a) Denote σ_1^2 and σ_2^2 as population variances of women's and men's scores respectively.

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

(b) Significance Level: $\alpha = 0.1$
Test Statistic:

$$F = \frac{S_1^2}{S_2^2} \overset{H_0}{\sim} \mathcal{F}(n_1 - 1, n_2 - 1)$$

Rejection Region:

$$\{F \geq k\}$$

where

$$k = \mathcal{F}_{0.05}(20 - 1, 30 - 1) = 1.9581$$

(c) $\mathcal{F}_{0.2}(19, 29) = 1.4042$

$$F = \frac{2.4623^2}{2.2118^2} = 1.2393$$

The p-value is larger than 0.4 by using $\mathcal{F}$ table, so we do not reject H_0. Therefore, we can not conclude that there is a difference in the variability of golf scores for male and female golfers.

7. (a)

$$H_0 : \mu_D = 0$$

$$H_1 : \mu_D \neq 0$$

$$T = \frac{\bar{D} - \mu_D}{S_D/\sqrt{n}} = \frac{-1.05 - 0}{3.3162/\sqrt{20}} = -1.42$$

$t_{0.05}(19) = 1.729, \quad t_{0.1}(19) = 1.328, \quad \alpha = 0.1, \quad df = 19$
The p-value is between 0.05 and 0.1 for one-tail by using t table.
Therefore, the p-value is between 0.1 and 0.2.

Do not reject H_0. There is no significant difference between the mean scores for the first and fourth rounds.

(b) $\bar{D} = -1.05$, first round scores are lower than fourth round scores.

(c) Margin of error:

$$t_{0.05}\frac{S_D}{\sqrt{n}} = 1.729\frac{3.3162}{\sqrt{20}} = 1.28$$

The 90% confidence interval for μ_D is

$$\bar{D} \pm t_{0.05}\frac{S_D}{\sqrt{n}} = -1.05 \pm 1.28$$

Since the interval contains 0, the difference between the population means is not significant.

8. Let $X_1, X_2, ..., X_{n1}$ and $Y_1, Y_2, ..., Y_{n2}$ represent two independent random samples from the respective normal distributions $\mathcal{N}(\mu_x,\ \sigma_x^2)$ and $\mathcal{N}(\mu_y,\ \sigma_y^2)$

$$\bar{X} - \bar{Y}\ \sim\ \mathcal{N}\left(\mu_x - \mu_y,\ \ \frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}\right)$$

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}}}\ \sim\ \mathcal{N}(0,\ 1)$$

$$1 - \alpha = P\left(-z_{\frac{\alpha}{2}} < \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}}} < z_{\frac{\alpha}{2}}\right)$$

$$= P\left(\left(\bar{X} - \bar{Y}\right) - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}} < \mu_x - \mu_y < \left(\bar{X} - \bar{Y}\right) + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}}\right)$$

so the $(1 - \alpha)100\%$ confidence interval for $\mu_x - \mu_y$ is

$$\left(\bar{X} - \bar{Y}\right) \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}}$$