Statistics (II) Quiz (1)

Version A Reference Solution

(I) Multiple Choice (7 pts for Q1 & Q2 ; 8 pts each for Q3 & Q4)

1. (A) Decreasing the standard error
2. (B) To combine information from both samples to improve variance estimation
3. (C) independent samples
4. (C) Because it arises from the sum of squared normal variables

(II) Fill in the Blanks (5 pts each)

5. Central Limit Theorem (CLT) ; normal
6. skewed ; outliers

(III) Short Answer (20 pts)

7. **Procedure for conducting a two-tailed hypothesis test for a population variance:**

   Step 1: Hypotheses

   $$H_0 : \sigma^2 = \sigma_0^2, \quad H_a : \sigma^2 \neq \sigma_0^2$$

   Step 2: Set a level of significance α

   Step 3: Test Statistic

   $$\text{Under } H_0 : \chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

   where s^2 is the sample variance

   Step 4: Rejection Rules
   (a) p-value approach: Reject H_0 if p-value $\leq \alpha$
   (b) critical value approach: Reject H_0 if $\chi^2 \leq \chi^2_{1-\alpha/2}$ or $\chi^2 \geq \chi^2_{\alpha/2}$

   Step 5: Draw conclusion

(IV) Computation (10 pts each)

8. **Given:** $n = 42,\ \bar{d} = 850,\ s_d = 1123$

   μ_1 = population mean grocery expenditures

   μ_2 = population mean dining-out expenditures

   (a) Hypotheses

$$H_0 : \mu_d = 0, \quad H_a : \mu_d \neq 0$$

   (b) Test Statistic

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{850}{1123/\sqrt{42}} \approx 4.91$$

   Given $t_{0.025,41} = 2.020$,

   Since $4.91 > 2.020$, we reject H_0.

   Conclusion: There is sufficient evidence that there is a difference between the annual population mean expenditures for groceries and for dining out.

   (c) Point Estimate & Convidence Interval

   The point estimate of the difference between the population means is:

$$\bar{d} = 850$$

   The 95% confidence interval for the difference is given by:

$$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$$

   Using $t_{0.025,41} = 2.020$:

$$850 \pm 2.020 \cdot \frac{1123}{\sqrt{42}}$$

$$= 850 \pm 350$$

$$= (500, 1200)$$

(V) Bonus Question (+20 pts)

9. **Derivation of** $(1-\alpha)100\%$ **Confidence Interval for Population Variance:**
   Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, s^2 is the sample variance
   we have:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

   Then,

$$P\left(\chi^2_{1-\alpha/2} \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi^2_{\alpha/2}\right) = 1 - \alpha$$

Take reciprocals (reverse inequalities):

$$P\left(\frac{1}{\chi^2_{1-\alpha/2}} \geq \frac{\sigma^2}{(n-1)s^2} \geq \frac{1}{\chi^2_{\alpha/2}}\right) = 1-\alpha$$

Rewrite:

$$P\left(\frac{1}{\chi^2_{\alpha/2}} \leq \frac{\sigma^2}{(n-1)s^2} \leq \frac{1}{\chi^2_{1-\alpha/2}}\right) = 1-\alpha$$

Multiply:

$$P\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right) = 1-\alpha$$

Hence, the $(1-\alpha)100\%$ confidence interval is:

$$\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right)$$