參考用

1. 請說明變異數(variance)、變異係數(coefficient of variation)和相關係數(coefficient of correlation)的意義與其應用時機。 (12%)

2. 常態母群體 $N(\mu,\sigma^2)$, $\sigma^2=9$。就假說檢定 $H_0: \mu=1, H_a: \mu>1$。
   a. 若 $n=16, \bar{x}=2.5$, 求 p-value。
   b. 若 $\alpha=0.05, n=16$ 和 $\bar{x}=2.5$ 時，其檢定結果如何? (12%)
   c. 若在 $\alpha=0.05$ 下，試求 power of test at $\mu=1.5$。

3. 某工廠擬估計其產品中不良品所佔比率，如果希望估計不良品比率的誤差在 ± 0.05 之內之可信度爲0.95。
   a. 試問在沒有任何資訊時，其樣本數應爲多少?(請分別用切氏(Chebyshev)不等式和中央極限定理求解)。
   b. 如果管理當局確信不良品比率不會超過 0.2，則其樣本數應爲多少?(請用切氏不等式求解)。 (10%)

4. Two independent random samples selected from normal populations $N(\mu_i, \sigma_i^2)$, $i=1,2$, produced the accompanying data summary (16%)
   sample 1 : $\bar{x}_1 = 22.1$, $s_1 = 4.8$, and $n_1 = 16$
   sample 2 : $\bar{x}_2 = 18.2$, $s_2 = 3.5$, and $n_2 = 12$
   a. Do the data contain sufficient evidence to conclude that the two population variances are different ? ($\alpha=0.05$)
   b. Suppose $\sigma_1 = \sigma_2$. Test the hypothesis $H_0: \mu_1 = \mu_2, H_a: \mu_1 > \mu_2$ at the α = 0.05 level of significance.

$F_{0.025,11,15} = 3.01$	$F_{0.025,12,16} = 2.89$	$F_{0.025,15,11} = 3.33$	$F_{0.025,16,12} = 3.16$
$F_{0.05,11,15} = 2.51$	$F_{0.05,12,16} = 2.42$	$F_{0.05,15,11} = 2.72$	$F_{0.05,16,12} = 2.61$
$z_{0.1683} = 0.96$	$z_{0.166} = 0.97$	$z_{0.1635} = 0.98$	$z_{0.1611} = 0.99$
$z_{0.05} = 1.645$	$z_{0.025} = 1.96$	$z_{0.0228} = 2$	$z_{0.0062} = 2.5$
$t_{0.025,26} = 2.056$	$t_{0.05,26} = 1.706$	$t_{0.025,27} = 2.052$	$t_{0.05,27} = 1.703$
$t_{0.025,28} = 2.048$	$t_{0.05,28} = 1.701$		

參考用

---

5. (30%) The normal error regression model is considered:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i. \qquad (1)$$

The least squares estimators, b_0 and b_1, are used to find a good fit of the linear regression function for Model (1):

$$\hat{Y} = b_0 + b_1 X,$$

where

$$b_1 = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sum(X_i - \overline{X})^2}, \quad \text{and} \quad b_0 = \overline{Y} - b_1\overline{X}.$$

The ith residual is the difference between the observed value Y_i and the corresponding fitted value $\hat{Y}_i$:

$$e_i = Y_i - \hat{Y}_i.$$

Consider the set of data below:

$x_i =$	1	1	2	2	3	3
$y_i =$	3	5	4	8	4	6

where $\overline{X} = 2$, $\overline{Y} = 5$, $\sum X_i = 12$, $\sum Y_i = 30$, $\sum X_i^2 = 28$, $\sum Y_i^2 = 166$, $\sum X_i Y_i = 62$, $\sum(X_i - \overline{X})^2 = 4$.

a) (15%) Construct the ANOVA table for simple linear regression.

b) (10%) Perform the lack-of-fit test at the level of significance $\alpha = .05$. State the alternatives, decision rule, and conclusion. ($F_{.95;3;3} = 9.28$, $F_{.95;2;3} = 9.55$, $F_{.95;1;3} = 10.1$, $F_{.95;2;4} = 6.94$, $F_{.95;1;4} = 7.71$, $F_{.95;1;5} = 6.61$).

c) (5%) Does the increase of SSLF (lack of fit sum of squares) imply the decrease of SSPE (pure error sum of squares)? Justify your answer.

---

6. (10%) Does a consistant estimator imply that it is also an unbiased estimator? Justify your answer. Note that an estimator $\hat{\theta}$ of the parameter θ is unbiased if

$$E(\hat{\theta}) = \theta,$$

and an estimator $\hat{\theta}$ is a consistent estimator of θ if

$$\lim_{n\to\infty} P(|\hat{\theta} - \theta| \geq \epsilon) = 0, \qquad \text{for any } \epsilon > 0.$$

---

7. (10%) Multiple choice (Choose all which apply).

A) Chi-square test may be used to test for independence.

B) Chi-square test may be used to test for homogeneity.

C) Chi-square distribution has only one parameter.

D) If chi-square is used to test for goodness-of-fit, it is an one-tailed test.

E) In the test for goodness-of-fit, if the sample size (n) is small (say $n < 10$), Komogorov-Smirnov method is "better" than chi-square.

F) Run test is frequently used to test for independence.

G) None of above.