

1. Given a sample data,  $y = \{8, 6, 5, 10, 9\}$ , please answer the following questions (25 分):
  - a. Calculate the arithmetic mean of  $y$ . (4 分)
  - b. Calculate the geometric mean of  $y$ . (4 分)
  - c. Calculate the harmonic mean of  $y$ . (2 分)
  - d. Calculate the standard deviation of  $y$ . (4 分)
  - e. Calculate the standard error of  $y$ . (3 分)
  - f. Calculate the median of  $y$ . (3 分)
  - g. Calculate the 95% confidence interval of  $y$ . (5 分)
  
2.  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$  is a multiple linear regression model with  $Y$  being the response variable,  $\beta_k$  the  $k$ -th regression coefficient, and  $X_k$  the  $k$ -th predictor. Please answer the following questions (25 分):
  - a. Write down the solution for  $\beta_k$  in matrix notation. (5 分)
  - b. Write down the solution for  $\beta_k$  as a set of normal equation. (5 分)
  - c. When the above multiple linear regression model is fitted to a dataset, what type of test will you use to test for model overall goodness of fit? Please name the test, the test statistics, and the probability density function associated with the test. (5 分)
  - d. When the above multiple linear regression model is fitted to a dataset, what type of test will you use to test whether an estimated coefficient is significant different than zero? Please name the test, the test statistics, and the probability density function associated with test. (5 分)
  - e. When the above multiple linear regression model is fitted to a dataset, one of the outputs is multiple-R<sup>2</sup>. What does multiple-R<sup>2</sup> explain? What is the equation for calculating multiple-R<sup>2</sup>? (5 分)
  
3. Please answer the following questions above hypothesis testing (25 分):
  - a. What is a non-parametric hypothesis test for one-sample data? (2 分)
  - b. What is a non-parametric hypothesis test for unpaired two-sample data? (2 分)
  - c. What is a non-parametric hypothesis test for paired two-sample data? (2 分)
  - d. What is a non-parametric hypothesis test for testing Normality assumption? (2 分)
  - e. The following is an ANOVA table, please fill in the blank (17 分)

Source of Variation	Degree of Freedom	Sum of Squares	Mean Square	F-Ratio
Treatment	1		2528.95	
Days	2		6124.48	
Treatment × Days		11747.21		
Residual		45335.10		
Total	43			

4. Please answer the following questions (25 分):

- a. Prove sample variance can be written in two ways, i.e., prove  $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2 / n}{n-1}$  (10 分)
- b. Prove that  $\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$  is a biased estimator for sample variance (10 分)
- c. What is the sampling distribution of an estimated mean? Please state the theorem and the distribution. (5 分)