Statistical Models and Regression

Exam 2 for Modules 2, 3, 4, 5, 6, 7, 8, 9, 10

Do all the problems below. Total points: 100

Provide necessary intermediate steps for all your work.

1. 100 people participated in a medical study. Each participant contributed three measurements (labeled y). Thus the study had a total of 300 measurements on y from the 100 participants. It is suspected that the three measurements from the same person are correlated. Without sufficient knowledge of the study background, a data analyst performed a multiple linear regression analysis using a model with three regressors, x_1, x_2, x_3, and assuming that all 300 measurements are statistically independent (i.e., no correlation between participants and between the three measurements of each participant). The model can be expressed as $E(y \mid x_1, x_2, x_3) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$. Thus the data analyst performed OLS (i.e., ordinary least-squares) estimation for all regression coefficients, using a commonly used statistical software.

   a) The data analyst assumed that all 300 measurements are statistically independent and used the OLS estimator to generate t test for testing significance of x_1 at the target significance level α. Is the type I error probability of this t test no larger than the targeted α level? Provide mathematical proof for your answer. [10 points]

   b) Later the data analyst was told that the statistical independence assumption does not hold. The data analyst was given the variance-covariance matrix $V = \sigma^2 \begin{pmatrix} 1 & .40 & .20 \\ .40 & 1 & .25 \\ .20 & .25 & 1 \end{pmatrix}$ for the three measurements of each participant, where σ^2 is unknown and needs to be estimated.

   (b1) Should the analyst adjust the OLS estimator and the statistical testing based on OLS method for the regression coefficient vector $B = (\beta_0, \beta_1, \beta_2, \beta_3)'$? Provide mathematical proof for your answer. [10 points]

   (b2) Construct an unbiased estimator for σ^2 and show unbiasedness. [10 points]

   (b3) Construct the best linear unbiased estimator for β_1 and its 95% confidence interval with mathematical proof of unbiasedness and correct confidence coverage probability. [10 points] (b4) Construct a statistically valid test for testing H_0: $\beta_1 = 0$ with mathematical proof for its validity. [10 points]

2. A controlled experiment was conducted to study the effect of treatment A and the effect of treatment B on the expected value of the response variable y which is normally distributed. Treatment A has two levels, labeled a1, a2. Treatment B has three levels, labeled b1, b2, b3. The y-responses within and across the six cells (i.e., combination treatment levels) are statistically independent with constant variance σ^2 whose value is unknown.

   Use any math/stat software (e.g., **www.**numbergenerator**.org/**randomnumbergenerator) of your choice to find a random number generator to randomly select from the dataset "Exam 2 Problem 2 data.txt" **one** row of **each** AB-combination level, e.g., A = a1, B = b1, etc. Place the six selected rows into the six cells, respectively, in Table 1, and N = 20 for each cell.

   Note: the raw data is not given; thus, it is not possible to use any statistical or mathematical software for regression analyses or analysis of variance. Show every step of your calculations for problems 2c and 2d.

Table 1

y: sample mean, y: sample SD, N	B = b1	B = b2	B = b3
A = a1			
A = a2			

SD = standard deviation

Perform two-way analysis of variance and the corresponding linear regression analysis, using the numbers you obtain in Table 1.

a) Write down the two-way analysis of variance model and the corresponding linear regression model. Make sure that constraints are specified. [10 points]
b) For the linear regression analysis with the model in a), will changing the indicator variables change the table of sums of squares? Provide mathematical proof of your answer. [10 points]
c) Suppose that treatment A and treatment B are known not to interact with each other. Let the linear contrast parameters be given by

$\theta_k = (-2) * E(y \mid A = ak, B = b1) + E(y \mid A = ak, B = b2) + E(y \mid A = ak, B = b3), \quad k = 1,2$

Construct the best linear unbiased estimator and 95% confidence interval for the common value of $\theta's$ with mathematical proof. Calculate the resulting estimate and the numerical values of the confidence interval [10 points]

d) In contrast to c), suppose that it is not known whether treatment A and treatment B interact. For the same $\theta's$ defined in c),

d1) Construct the best linear unbiased estimator and 95% confidence interval for the $\theta's$ with mathematical proof. Calculate the resulting estimates and the numerical values of the confidence intervals [10 points]

d2) Test the null hypothesis $\mathrm{H}_0: \theta_1 = \theta_2$ at the 0.05 level of statistical significance with mathematical proof of the validity of the constructed test. [10 points]