APPLIED REGRESSION **TEST-II** **SPRING 2022**

Name: _____ **Instructor: D. Kushary**

Information and Instruction:

(1) Test is open book and open notes and you can use your computer.

(2) Time is 1 hour 30 minutes (strictly).

(3) YOUR ANSWER PAPER SHOULD BE A PDF FILE AND FILE-NAME SHOULD BE TEST2_LASTNAME_FIRSTNAME.PDF

(4) YOUR R-PROGRAM NEEDS TO BE SUBMITTED IN CANVAS - FILE NAME SHOULD BE TEST2_LASTNAME_FIRSTNAME.R)

1. A study was conducted to understand the relationship of amount of body fat (Y) to several possible predictor variables, based on a sample of 20 healthy females 25- 34 years old. The possible predictor variables are triceps- skinfold thickness (X1), thigh circumference (X2), and midarm-circumference (X3).

   The attached data set has 3 independent variables (X1-X3) and one dependent variable (Y). , backward($\alpha = 0.05$)

   (a) Use forward selection procedure (use $\alpha = 0.05$) and write down the final estimated model.

   (b) Use backward selection procedure (use $\alpha = 0.05$) and write down the final estimated model.

   (c) Run the model with all three independent variables and write down the ANOVA table and use the ANOVA table to test whether X2 is significant or not at 5%. (You can run only one model (ie. ONLY ONE LM-STATEMENT IN R-program) and you can only use the ANOVA TABLE TO ANSWER THE QUESTION)

2. A hospital administrator wished to study the relationship between patient satisfaction (Y) and patient's age (X1, in years), severity of illness (X2, an index), and anxiety level (X3 , an index). The ANOVA tables is given below

ANOVA TABLE

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
X1	1	8275.39	8275.39	81.80	0.0000
X3	1	763.42	763.42	7.55	0.0088
X2	1	81.66	81.66	0.81	0.3741
Residuals	42	4248.84	101.16		

   (a) Is the above model significant at 5% level? (Graph Needed)

   (b) Now if the administrator decides to drop the variable X2 and rerun the model. Will the new model have higher adjusted-R^2? (Compute and compare)

   (c) Is X3 significant in the new model (after dropping X2) at 1% level?

3. (Continuation of Question 2) Residual analysis was performed on the full model (with independent variables as X1, X2 and X3) and following data came up (Observation number is 17):

   $SSE = 4248.8, \quad \bar{y} = 61.56 \quad y_{17} = 79, \quad e_{17} = 16.61, \quad h_{17} = 0.1195$

   (a) Find R-student for the 17th observation and judge whether it should be considered an outlier or not using the test at 5% level.

   (b) Find the Cook's D for the 17th point ?

   (c) Find the DFFITS for the 17th point ?

   EXTRA CREDIT

   (d) Regardless of your own decision in part (a), the administrator dropped the 17th observation. Can you find the R^2 for the new model after you dropped the 17th observation?