

**Instructions: Show all work. Answers without necessary formulas and steps will not receive any credit.**

1. (35 points) A random sample of 22 residential properties was used in a regression of price on number of bedrooms and age (in years).
  - (a) (10) Suppose the variable “number of bedrooms” has 4 values: 1 bedroom, 2 bedrooms, 3 bedrooms and 4 bedrooms. Define 3 dummy variables  $(x_1, x_2, x_3)$  based on this variable, using “4 bedrooms” as the comparison group.
  - (b) (10) Let  $x_4 = \text{age}$ ,  $x_{14} = x_1 \times x_4$ ,  $x_{24} = x_2 \times x_4$ , and  $x_{34} = x_3 \times x_4$ . Complete the ANOVA tables.
  - (c) (15) Test to determine whether the linear relationship between price and age differs depending on the number of bedrooms. Use  $\alpha = 0.05$ .

-----  
 Analysis of Variance (x1,x2,x3,x4)

Source	DF	SS
Regression	--	4573.1
Residual Error	--	-----
Total	--	10867.6

Analysis of Variance (x1,x2,x3,x4,x14,x24,x34)

Source	DF	SS
Regression	--	4800.9
Residual Error	--	-----
Total	--	10867.6

Name: \_\_\_\_\_

2. (30 points) The Summer Olympic Games have been held from 1896 until 1996. Over this 100 year period, there have been a total of 40 times that the 100 meter dash has been run: 16 times for women and 24 times for men. For each event, the winning (gold medal) times have been recorded. For example, the winning men's time in 1996 was 9.84 seconds, while the winning women's time was 10.94 seconds.

An example of the data is given below:

Time (seconds)	Gender	Year
12.20	Women	1928
11.90	Women	1932
.....		
11.00	Men	1904
11.20	Men	1906

A linear regression model with the winning times ( $y$ ) as the response and the year ( $x$ ) and gender ( $G = 1$  for men and 0 for women) as the predictors was fit to the data. The model can be written as  $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i \times G_i + \varepsilon_i$ ,  $i = 1, \dots, 40$ . Note that this model imposes a common intercept  $\beta_0$  and separate slope parameters for men and women.

- (a) (15 points) What are the slopes for men and women? Write down your answers in symbols.
- (b) (15 points) The following diagnostic values (Cook's distance, studentized deleted residuals, and leverages) were computed for the earliest men's 100 meter events (from 1896 to 1906), based on the multiple regression model:

Year	Cook's Dist.	Stud. Del. Residual	Leverage
1896	0.579	4.2807	0.1222
1900	0.022	-0.7273	0.1097
1904	0.008	-0.4494	0.0984
1906	0.011	0.5725	0.0931

Based on these three diagnostic measures, comment on whether any of the above 4 observations are influential.

Name: \_\_\_\_\_

3. (35 points) A rehabilitation center researcher was interested in examining the relationship between physical fitness prior to surgery of persons undergoing corrective knee surgery and time required in physical therapy until successful rehabilitation. Patient records in the rehabilitation center were examined, and 24 male subjects who had undergone similar corrective knee surgery during the past year were selected for the study. The number of days required for successful completion of physical therapy and the prior physical fitness status (below average, average, above average) for each patient follow.

Fitness status	Time (days)								Sample mean	Sample variance
Below average	29	42	38	40	43	40	30	42	38	30
Average	30	35	39	28	31	31	29	35	32.25	13.93
Above average	26	32	21	20	23	22	29	23	24.5	17.43

- (a) (10 points) Write down an appropriate ANOVA model. Estimate the effects due to fitness status. Comment on the nature of the relationship between fitness status and time required in physical therapy.
- (b) (5 points) State the assumptions for the model in (a). List the names of two graphs that can most effectively show whether the residuals are normally distributed. [Hint: Boxplot is not one of them.]
- (c) (10 points) Obtain the analysis of variance table.
- (d) (10 points) Test whether the mean number of days required for successful rehabilitation is the same for the three fitness groups. Include all steps of a statistical test and give a p-value. Interpret your results in the context of the setting using  $\alpha = 0.01$ .