JARAMOGI OGINGA ODINGA UNIVERSITY OF SCIENCE & TECHNOLOGY

SPECIAL RESIT EXAMINATION 2020/2021

SAC 104: LINEAR MODELS AND FORECASTING

Question Compulsory (30 marks)

a) An expert on crime rates has collected the following information on five counties in one province.

County	Crime Rate (Y)	Poverty Rate (X)
1	10	5
2	19	7
3	20	11
4	16	8
5	15	9

Crime Rate (Y) stands for number of crimes per 10,000 people. Poverty Rate (X) is percent of families below poverty line. Sample statistics based on these data:

$$\sum X = 40 \quad \sum X^2 = 340 \quad \sum Y = 80 \quad \sum Y^2 = 1342 \quad \sum XY = 666$$

i) Find the regression line and interpret the intercept and slope coefficients. (5 marks)

ii) Compute the coefficient of determination, R-squared. What does it mean? Make an inference about the crime rate in a county with a poverty rate of 10. [Use α = 0.05 for the interval.] (3 marks)

iii) A governor claims that based on the estimated slope of the regression, a one percentage point increase in the poverty rate is associated with less than two additional crimes per 10,000 people. Test this claim at a 5% significance level. (3 marks)

iv) The provincial government does not agree with the results of this regression analysis arguing that distance from the provincial capital also impacts the crime rate. They have classified the counties in two general categories, far or close and hired you to refine this regression in a way that their concern can be statistically tested. Design a regression that allows for this concern to be tested for the following two cases: (3marks)

v) When the distance impacts the crime rate but not the slope of the poverty rate. (4 marks)

vi) When the distance impacts the crime rate because the impact of poverty for far and close counties is different. (2mks)

b) State four assumptions of Ordinary Least Square necessary to produce unbiased estimators (4mks)

c) What are the linear models, and why are they called `linear'? (2mks)

d) Prove that $\hat{\beta}$ is unbiased estimator of β (4mks)

Question 2(20 marks)

A researcher is studying the relationship between house prices Y_i (measured in *thousands* of dollars) and X_i house size (measured in square feet) for a sample of 14 homes in a particular neighborhood. Preliminary analysis of the sample data produces the following sample information:

$N = 14$ $\quad \sum_{i=1}^{N} Y_i = 2{,}834.9$ $\quad \sum_{i=1}^{N} X_i = 26{,}832.0$

$\sum_{i=1}^{N} Y_i^2 = 612{,}628.0$ $\quad \sum_{i=1}^{N} X_i^2 = 53{,}890{,}858.0$ $\quad \sum_{i=1}^{N} X_i Y_i = 5{,}664{,}696.0$

$\sum_{i=1}^{N} x_i y_i = 231{,}407.5$ $\quad \sum_{i=1}^{N} x_i^2 = 2{,}465{,}417.0$ $\quad \sum_{i=1}^{N} y_i^2 = 38{,}581.01$ $\quad \sum_{i=1}^{N} \hat{u}_i^2 = 16{,}860.79$

Where $x_i = X_i - \overline{X}$ **and** $y_i = Y_i - \overline{Y}$

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

(a) Use the above information to compute OLS estimates of the intercept coefficient β_1 and the slope coefficient β_2. **(5 marks)**

(c) Use the above information to calculate an estimate of σ^2, the error variance. **(5 marks)**

(d) Let $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$ (i = 1, …, N). What is the value of $\sum_{i=1}^{N} \hat{Y}_i$ for the sample regression equation you have estimated? Explain briefly how you obtained your answer.

(5marks)

© Use the above information to compute the value of R^2, the coefficient of determination for the estimated OLS sample regression equation. Briefly explain what the calculated value of R^2 means.

(5marks)

Question 3 (20 marks)

a) Given the following data for GNP and sales, use a simple OLS regression to forecast 2010 sales for Smedley Company:
(10marks)

X GNP ($000)	Y $ Sales	Year
18,000	10,000	2004
32,000	12,000	2005
47,000	16,000	2006
72,000	23,000	2007
61,000	19,000	2008
80,000	22,000	2009
90,000	?	2010

b) Prices and unit sales of a particular model of rat trap were recorded on five occasions with these results

Price	Units Sold
$10	3
8	5
12	2
6	8
4	12

i) Calculate r, the correlation coefficient

ii) Is there a linear relation between price and sales? Explain it (10mks)

Question 4 (20 marks)

a) Your company has recently introduced a new product. You have just received confidential information about the average daily sales (in dollars) for the first 46 days after the product launch. The following table summarizes this information:

Average Daily Sales	5,200
Standard Deviation	1,150
Observations	46

The accounting department of your company has estimated that it will not be profitable to produce the product if the average daily sales do not exceed $4,500. One of your colleagues claims that the product will not be profitable. In particular, she argues that future average daily sales are going to be $4,200 and that it is not in the company interest to keep the product on the market. Is her argument correct? If it is, use your data to support her argument. If it is not, use your data to show that it is wrong. (10 marks)

b) Consider the research question of whether woman exercise more regularly than men. A random sample of 200 women and 150 men yields these results:

	Men	Women
Exercise regularly	88	130
Do not exercise regularly	62	70
Total	150	200

i. Construct a 95% confidence interval estimate of the difference in the proportion of women and men who exercise regularly. Interpret the interval. (3marks)

ii. Conduct a hypothesis test to determine if women exercise more regularly than men. For a 5% significance level, find the standardized rejection region AND the pvalue. Make a conclusion for both the rejection region approach and p-value approach.

(5 marks)

iii. If 65% of the women and 55% of the men exercise regularly, what is the power of your test in Part (b) for $\alpha = 0.05$? (2 marks)

Question 5 (20 marks)

a) A random sample of 62 seventh grade students is selected to study the relationship between Grade Point Average (GPA) and IQ (Intelligence Quotient). The following descriptive statistics are obtained:

	GPA	IQ
Mean	7.5	105
Standard Deviation	2.1	16

The correlation coefficient between GPA and IQ is 0.64.

(i) Find the equation of a simple regression for predicting GPA from IQ.

What is the interpretation of the intercept? (5 marks)

(ii) Find the standard error of the slope estimate. Is the slope statistically significant at the 0.05 level? (5 marks)

b) A used car dealer offers a number of models for sale during a clearance event. The following data on price, Y (in \$1000's), and age of the car, X (in years), are obtained from past experience:

Y	39.9	32	25	20	16	20	13	13.7	11	12	20	9	9	12.5	7
X	1	2	4	5	6	6	10	11	11	12	12	12	12	13	15

(i) Estimate the regression equation for price as a function of age and interpret the parameter estimates. (3 marks)

(ii) Is there sufficient evidence to indicate a linear relationship between selling price and age? Test the appropriate hypotheses using 5% as the significance level.(2 marks)

(iii) The car dealer claims that a one year increase in the age of the car reduces the price by more than \$1500. Test this claim using $\alpha = 0.05$. (3 marks)

(iv) Suppose a car owner asks "What is the predicted selling price of my 5 year old car? Give me a range." Provide a 95% interval that can be used for such a prediction. (2 marks)