JARAMOGI OGINGA ODINGA UNIVERSITY OF SCIENCE AND TECHNOLOGY

SCHOOL OF MATHEMATICS AND ACTURIAL SCIENCE

UNIVERSITY EXAMINATION FOR BACHELOR OF ACTUARIAL SCIENCE

AUGUST 2023/2024

MAIN SPECIAL RESIT

COURSE CODE: WAB 2103

COURSE TITLE: Linear Models and Forecasting

EXAM VENUE **STREAM: B.Sc. Actuarial Science**

DATE:…… **EXAM SESSION: APRIL 2024**

TIME: 2 HOURS

Instructions:

1. **Answer ONE and any other two questions only.**
2. **Candidates are advised not to write on the question paper.**
3. **Candidates must hand in their answer booklets to the invigilator while in the examination room.**

QUESTION ONE COMPULSORY (30 MARKS)

i) Briefly state five assumptions of simple linear regression model. (10Marks)

ii) State and explain the assumptions underlying one way Analysis of Variance(4Marks).

iii) For a simple linear regression model the sum of squares of the residuals is

$$\sum_{i=1}^{25} e_i^2 = 230 \ and \ R^2 = 0.64$$

Calculate the total sum of squares (TSS) for this model (6Marks)

iv) Sarah performs a regression of the return on a mutual fund (y) on four predictors plus an intercept. She uses monthly returns over 105 months. Her software calculates the F statistic for the regression as F = 20.0, but then it quits working before it calculates the value of R^2. While she waits on hold with the help desk, she tries to calculate R^2. from the F-statistic. Determine the attempted calculation. (5Marks)

v) Two actuaries are analyzing dental claims for a group of n = 100 participants. The predictor variable is sex, with 0 and 1 as possible values.

Actuary 1 uses the following regression model: $Y = \beta + \varepsilon$.

Actuary 2 uses the following regression model:$Y = \beta_0 + \beta_1 x \ Sex + \varepsilon$.

The residual sum of squares for the regression of Actuary 2 is 250,000 and the total sum of squares is 490,000.

Calculate the F-statistic to test whether the model of Actuary 2 is a significant improvement over the model of Actuary 1. (5Marks)

QUESTION TWO (20 MARKS)

Given a sample of $((X_i, Y_i), i = 1, \dots, 5$, we have the following summary statistics:

$$\sum_{i=1}^{n} X_i = 15, \ \sum_{i=1}^{n} Y_i = 5.6, \quad \sum_{i=1}^{n} X_I^2 = 55 \quad \sum_{i=1}^{n} Y_i^2 = 7.2$$

$$\sum_{i=1}^{n} (X_i - \bar{X})^2 = 10 \quad \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = 0.948 \quad \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) \quad = 3$$

$$\sum_{i=1}^{n} X_i \, Y_i = 19.8$$

Consider the model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$

i) Find the OLS estimates of β_0 and β_1 and their standard error (5mks)

ii) Find the coefficient of determination R^2 (5mks)

iii) Perform a hypothesis test for $H_0: \beta_1 = 0 \ vs \ H_1: \beta_1 \neq 0$ at a significant level of 5% (5mks)

iv) Perform a hypothesis test for $H_0: \beta_1 = 0 \ vs \ H_1: \beta_1 > 0$ at a significant level of 2.5% (5mks)

QUESTION THREE (20 MARKS)

Use the Method of Least Squares to Produce a full derivation for the result; $Y = \beta_1 + \beta_2 x + \mu$ by deriving the estimation of β_1 and β_2. (20 Marks)

QUESTION FOUR (20 MARKS)

Where y is expressed in thousands of pounds and x in hours, using fractions of an hour as units of low order decimal. The statistics of flight times and fuel consumption of airlines provides data on flight times and fuel consumption of24 different trips made by an aircraft of the company. From these data the following statistics were drawn;

$\sum x = 31.470$, $\sum x^2 = 51.075$, $\sum y = 219.719$, $\sum y^2 = 2396.504$, $\sum xy = 349.486$

(a). Estimate β_1 and β_2 (4 Marks)

(b). Decompose the variance of the variable y invariance explained by the regression and residual variance (6 Marks)

(c). Calculate the coefficient of determination. (5 Marks)

(d). Estimate total consumption, in thousands of pounds , for a flight program consisting of 100 half hour flights, 200 one hour flights and 100 two hour flights. (5 Marks)

QUESTION FIVE (20 MARKS)

Say you want to assess the importance of increasing the number of police officers on the crime rate. You have access to data at the municipality level and can estimate the equation:

$$Crime_{it} = \beta_0 + \beta_1 Police_{it} + \epsilon_{it}.$$

Both measures (number of crimes and number of police officers) are measured per 1000 inhabitants in each municipality.

(a). Write down the fixed effects regression model based on the regression model above. Based on this equation, describe what kind of problem you can potentially solve using fixed effects compared to the equation above? What kind of problems can a fixed effects model not solve? (5 Marks)

(b). What must be true (which assumption needs to hold) in order for us to interpret the coefficient from your panel data regression as a causal relationship?
Write down the equation (2 Marks)

(c). Using yearly panel data for several time periods, the estimated coefficient on *Police* is -0.1 in your fixed effects specification. How do you interpret this estimate? Do you believe that this estimate has a causal interpretation? Motivate your answer. (5 Marks)

(d). Suppose that around half of the municipalities in your data received central government support to increase police density at the same time period and the other half did not. Using this new information, how could you design a study to answer the question above in a causal way? Write down the equation, describe which parameter you are interested in and describe how you would show your readers that it is plausible that the most important assumption for estimating a causal effect holds. (8 Marks)