Strathmore
UNIVERSITY

STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES

MASTER IN STATISTICAL SCIENCE

END SEMESTER EXAMINATION

STA 8306 LONGITUDINAL DATA ANALYSIS

Date: 21st August 2024 **Time: 2hours 30 Minutes**

Instruction: Answer Question 1 and any other Two Questions

Question 1 (20 Marks)

(a) Differentiate between longitudinal studies and cross sectional studies (2 Marks)

(b) Describe the two source of correlation in Longitudinal Data Analysis (LDA) (2 Marks)

(c) The table below presents cross sectional data collected on 7 children on height, age and Response Variable; Volume of air exhaled in the first second of spirometry, FEV1.

Subject ID	Height	Age (years)	Log (FEV1)
20	1.46	11.29	0.63
17	1.81	6.9	0.3
176	1.23	8.34	0.14
81	1.62	13.32	0.57
3	1.36	9.87	0.50
214	1.35	18.23	0.94
34	1.67	16.23	0.98

(i) Use 'R' to fit a multiple linear regression model describing how the value of log(FEV1) varies linearly with age and log(height). (attach all the 'R' syntax.) (3 Marks)

(ii) Explain the regression coefficients (3 Marks)

(iii) Describe the fit of the model (2 Marks)

(d) Explain the four potential consequences of not accounting for correlation among longitudinal data in the analysis (4 Marks)

(e) The following matrix shows estimated correlation matrix for the blood lead levels at baseline, week 1, week 4, and week 6 for 50 children in the placebo group of a clinical trial

Correlation Matrix:

$$\begin{pmatrix} 1 & 0.83 & 0.84 & 0.59 \\ 0.83 & 1 & 0.86 & 0.6 \\ 0.84 & 0.86 & 1 & 0.74 \\ 0.59 & 0.60 & 0.74 & 1 \end{pmatrix}$$

Explain the observations about Correlation in Longitudinal data above (4 Marks)

Question 2 (20 Marks)

Two groups of elderly males; Males identified without senile factor and Males identified with a senile factor were studied. The variable scores on intelligence test namely, Information, Similarities, and Arithmetic and picture completion are listed below

	Group	
	no senile factor	senile factor
Subtest	$n_1 = 37$	$n_2=12$
Information	12.57	8.75
Arithmetic	9.57	5.33
Similarities	11.49	8.5
Picture Completion	7.97	4.75

Suppose the variance –covariance matrix S is given by;

$$\mathbf{S} = \begin{bmatrix} 11.2624 & 9.406 & 7.155 & 3.3791 \\ & 13.5265 & 7.34784 & 2.5014 \\ & & 11.5796 & 2.6167 \\ & & & 5.83133 \end{bmatrix}$$

(a) Determine whether the two profiles are parallel (4 Marks)
(b) Assuming the profiles are parallel determine whether the profiles are coincident (8 Marks)
(c) Assuming that the profiles are coincident test whether the profiles are level (8 Marks)

Question 3 (20 Marks)

(a) State and explain five assumptions that underpin the one-way repeated measures ANOVA. (5 Marks)

(b) Current research shows that long-term, low-level inflammation can be a cause and predictor of heart disease, which is the leading cause of premature death in the Western world. One measure of long-term, low-level inflammation that can be used to assess the risk of heart disease is called C-Reactive Protein (CRP). It can be measured in the blood and research shows that higher levels of CRP are associated with a higher risk of heart disease. Simultaneously, it is known that if you are overweight or obese you are at an increased risk of heart disease and that dieting (i.e., reducing your body fat percentage) can lead to reductions in traditional markers of heart disease, such as cholesterol concentration.

A researcher wanted to know whether dieting might also reduce low-level inflammation as assessed by CRP concentration in the blood. In order to investigate this idea, the researcher recruited 10 overweight participants who underwent a four-month dietary programme to reduce their body fat levels. CRP concentration was measured at the beginning of the dietary programme (i.e., at zero months), at the midpoint (i.e., two months into dieting), and immediately after the dietary intervention (i.e., at four months). Essentially, the researcher wants to discover whether CRP concentrations decrease over the period of the dietary programme (i.e., over the three time points). The data collected are in excel file labelled 'Question 3'

(a) Present
(i) the boxplots you used to check if there were any significant outliers and comment on your findings (3 Marks)
(ii) carry out a test for normality and interpret the results. (3 Marks)
(b) using 'R' present the one-way repeated measures ANOVA analysis (5 Marks)
(c) Discus the output of the results and draw your conclusions (4 Marks)

Question 4 (20 Marks)

A researcher published the results of an efficacy study on estrogen patches in treating postnatal depression. 17 women with major depression, which began within 3 months of childbirth and persisted for up to 18 months postnatally, participated in a double-blind, placebo-controlled study. Women were randomly assigned to active treatment (n=9) or placebo (n=8). Participants attended clinics monthly and at each visit self-ratings of depressive symptoms on the postnatal depression scale (PDS) were measured. PDS is a standardized, validated, self-rating scale consisting of 10 items, each rated on a 4-point scale of 0–3. The data is presented on the excel data file (sheet 2) labelled **'Question 4'.**

Depression scores are assessed across m = 7 months for the n = 17 subjects in the study and Depression scores for visit j are the longitudinal components measured on subject i . Using 'R' syntax,
(a) Summarize the depression scores by visit and treatment group and comment on the results (4 Marks)
(b) Examine within-person correlations among depression scores,graphically and numerically. (4 Marks)
(c) Graph depression scores over time, by treatment group. Include a smoother line for each group to summarize trends. (4 Marks)
(d) Plot individual trajectories by treatment group. (3 Marks)
(e) Reshape the data into long form and evaluate longitudinal associations between depression scores and treatment using. Use visit as a linear variable. (5 Marks)