



UNIVERSITY OF RUHUNA
FACULTY OF FISHERIES AND MARINE SCIENCES &
TECHNOLOGY

Academic Year 2022/2023

Bachelor of Science Honours in Marine and Freshwater Sciences Degree
Level III Semester I Examination – April/May 2024

OCG 3152: Statistics for Experimental Analysis II

Time: 2 hours

Answer **any four (04)** questions. Each question carries equal marks.

Calculators are allowed.

Use a significance level of 0.05.

Q1

- (a) Write down the formulas for Karl Pearson's correlation coefficient and the Spearman correlation coefficient. [20 marks]
- (b) Explain what is meant by a monotonic function. [15 marks]
- (c) A doctor is interested in the relationship between a person's Body Mass Index (BMI) and their level of fitness. She believes that a lower BMI leads to a greater level of fitness. She randomly selects 10 female 18-year-olds and calculates each individual's BMI. The females then run a race and the doctor records their finishing positions. The results are shown in the table.

Individual	A	B	C	D	E	F	G	H	I	J
BMI	17.4	21.4	18.9	24.4	19.4	20.1	22.6	18.4	25.8	28.1
Finishing position	3	5	1	9	6	4	10	2	7	8

- (i) Calculate Spearman's rank correlation coefficient for these data. [30 marks]
- (ii) Stating your hypotheses clearly and using a one tailed test with a 5% level of significance, interpret your rank correlation coefficient. [30 marks]
- (iii) Give a reason to support the use of the rank correlation coefficient rather than the product moment correlation coefficient with these data. [05 marks]

Q2

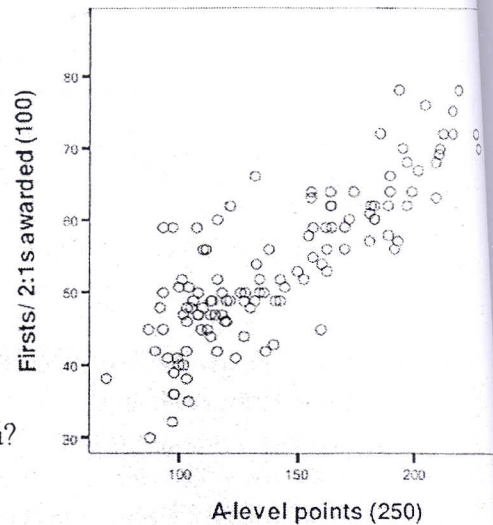
- (a) Briefly explain about the one-way and two-way analysis of variance. [30 marks]
- (b) Four brands of flashlight batteries are to be compared by testing each brand in five flashlights. Twenty flashlights are randomly selected and divided randomly into four groups of five flashlights each. Then each group of flashlights uses a different brand of battery. The lifetimes of the batteries, to the nearest hour, are as follows.

Brand A	Brand B	Brand C	Brand D
42	28	24	20
30	36	36	32
39	31	28	38
28	32	28	28
29	27	33	25

Preliminary data analysis indicate that the independent samples come from normal populations with equal standard deviations. At the 5% significance level, does there appear to be a difference in mean lifetime among the four brands of batteries? (You may use a one-way analysis of variance table). [70 marks]

Q3

- (a) The graph shows the proportion of Firsts or Upper Seconds as a measure of degree attainment, plotted against standards of A level passes obtained by new students, the data is from the Sunday times survey of higher education (HE). It is sighted as evidence that universities with an intake of students with "better" A-levels have an output of students with a higher percentage of "better" classed degrees.



- Is this an appropriate way to show the data?
- Is the graph labelled adequately?
- What does it show?
- Does this support the above argument?

[20 marks]

- (b) The following tables are the output of SPSS related to an analysis on total expenditure (HUF (Hungarian Forint)/month) of households.

Statistics		
Total expenditure (HUF/month)		
N	Valid	222
	Missing	0
Mean		99692.72
Median		85500.00
Mode		58500.00 ^a
Std. Deviation		57045.35
Skewness		1.02
Kurtosis		1.25
Minimum		5000.00
Maximum		337000.00
Percentiles	25	59875.00
	50	85500.00
	75	126700.00

a. Multiple modes exist. The smallest value is shown

Test Value = 0					
t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
				Lower	Upper
26.039	221	0.000	99692.71622	92147.4136	107238.0189

e from normal
oes there appear
may use a one-
[70 marks]

- (i) Interpret the statistics table. [10 marks]
(ii) What is the confidence level used in the? [05 marks]
(iii) Now test the hypothesis that the total expenditure of households equals \$100 000 with the significance level 5%. Then the SPSS results are shown as given in the following table.

Test Value = 100000					
t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
				Lower	Upper
-.080	221	.936	-307.28378	-7852.5864	7238.0189

Can we accept the null Hypothesis? Explain your answer. [10 marks]

- (iv) Next test the hypothesis that the average expenditure on heating of households with gas and with solid fuels are equal ($\alpha = 0.05$) and following are the results.

	Levene's Test for Equality of Variances		t-test for Equality of Means				
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Equal variances assumed	5.012	0.026	-1.506	190	0.134	-5617.6	3729.59
Equal variances not assumed			-1.716	189.990	0.088	-5617.6	3273.48

Interpret the table. [10 marks]

- (v) Is there any relationship between the total expenditure and average income according to the following results? Explain your answer. [10 marks]

		Total expenditure (HUF/month)	Average income (HUF/month)
Total expenditure (HUF/month)	Pearson Correlation	1	0.539(**)
	Sig. (2-tailed)		0.000
	Covariance	3254171744.937	2147784135.965
	N	222	217
Average income (HUF/month)	Pearson Correlation	0.539(**)	1
	Sig. (2-tailed)	0.000	
	Covariance	2147784135.965	4814606330.816
	N	217	217

- (vi) Is there any relation between the average income, total expenditure and number of people living in the household? [10 marks]

Correlations				
		Average income	Total expenditure	Number of people living in the household
Average income	Pearson Correlation	1	0.539**	0.293**
	Sig. (2-tailed)		0.000	0.034
Total expenditure	Pearson Correlation	0.539**	1	0.373**
	Sig. (2-tailed)	0.000		0.000
Number of people living in the household	Pearson Correlation	.293**	0.373**	1
	Sig. (2-tailed)	0.034	0.000	

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

- (vii) Determine a linear relation between the average income, total expenditure and number of people living in the household. [10 marks]

Model Summary ^a

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	0.581 ^a	0.338	0.332	46 948.77	1.780

a. Predictors: (Constant), Average income (HUF/month), Number of people living in the household

b. Dependent Variable: Total expenditure (HUF/month)

- (c) The following are data about a population in 2020.

Month	Population (thousand persons)	Number of birth (persons)	Number of death (persons)
1	10 273	10 238	13 888
2	10 270	9 285	12 825
3	10 267	10 105	12 516
4	10 265	9 617	11 753
5	10 262	9 548	12 328
6	10 260	9 717	11 839
7	10 258	9 965	11 848
8	10 257	9 980	11 722
9	10 256	9 844	10 968
10	10 252	9 021	12 542
11	10 249	8 740	11 743
12	10 246	9 538	12 917

When creating a trend model for the birth using SPSS following results were found.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	536259,1	1	536259,084	3,261	,101
Residual	1644363	10	164436,258		
Total	2180622	11			

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Case Sequence	-61,238	33,910	-.496	-1,806	,101
(Constant)	10031,212	249,573		40,194	,000

- (i) Find the linear regression line for the above data. [10 marks]
(ii) Give estimation of number of births in January 2021. [05 marks]

expenditure and
[10 marks]

Durbin-Watson
1.780

people living in

with

found.

101
000

0 marks]
5 marks]

Q4

- (a) Explain the main differences between Hierarchical cluster analysis and K-Means cluster analysis. [10 marks]
- (b) Find the Euclidean, Manhattan and Maximum distance between two points $A = (x_1, y_1)$ and $B = (x_2, y_2)$. [15 marks]
- (c) Eight fish are receiving two different feeds as X and Y as follows.

Fish Name	A	B	C	D	E	F	G	H
Feeding type								
X	2	2	8	5	7	6	1	4
Y	10	5	4	8	5	4	2	9

Using the Euclidean distance and taking initial centers as A, D and G perform three iterations of K-means cluster analysis. [75 marks]

Q5

- (a) What is the practical importance of Time series analysis? [15 marks]
- (b) Write short notes on followings.
- (i) Trend. [10 marks]
- (ii) Seasonal Variations. [10 marks]
- (iii) Cyclic variations. [10 marks]

- (c) The following table relates to the number of Serendib Scopes Owls visited to a same location in several months.

Month	JAN	FEB	MAR	APR	MAY	JUN	JUL
Number of birds	18	20	23	25	24	28	30

Fit a straight-line trend by the method of least squares and estimate the number of Serendib Scopes Owls that would visit in August. [55 marks]