KENYATTA UNIVERSITY

CITY CAMPUS

UNIVERSITY EXAMINATIONS 2017/2018

FIRST SEMESTER EXAMINATION FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

EMP 923: ADVANCED EDUCATIONAL STATISTICS

DATE: Monday, 5th February, 2018 TIME: 2.00 p.m. - 5.00 p.m.

INSTRUCTIONS:

Question **ONE** is **Compulsory**. Answer any other **TWO** questions.

Question One (Compulsory) - (20 marks)

a) Explain the similarities and differences between simple regression and multiple regression. (1 mark)

b) A study was conducted and a significant relationship was found among the number of hours a teenager watches television (TV) per day (X_1) , the number of hours a teenager talks on cellphone per day (X_2), and the weight of the teenager (Y). The multiple regression equation generated was:

$Y = 98.7 + 3.82X_1 + 6.5\ X_2$

i) Interpret the multiple regression equation. (2 marks)

ii) Predict a teenagers weight of a student if she averages 3 hours watching TV and 1.5 hours talking on cellphone per day. (2 marks)

c) The following is a random sample of students' scores in Chemistry in a schools in Kiambu County.

Data set : 85, 25, 45, 30, 35, 40, 50, 50, 55, 60, 65, 70, 75, 20, 90.

Use the data to determine the following:

i) Median (1 mark)

ii) Arithmetic mean (1 mark)

iii) Quadratic mean (1 mark)

iv) Variance (1 mark)

v) Standard deviation (1 mark)

$$\frac{\Sigma X^2 - \frac{\Sigma(X)^2}{N}}{N.}$$

$Z = \frac{(X - \bar{X})^2}{S}$

vi) Coefficient of variation (1 mark)

vii) Coefficient skewness (1 mark)

viii) The z-score of a student who scored 78% in the examination (1 mark)

d) The following data represents the enrollment trend of secondary students in Kenya by gender in percentages since independence for selected years (from 1964 to 2014).

Year	1964	1974	1984	1994	2004	2014	Beyond 2014
Male (%)	80	75	65	55	58	50.5	49
Female (%)	20	25	35	45	42	49.5	51

i) Construct a compound time series for the data (1 mark)

ii) Interpret the policy implication of the observed trend. (1mark)

Question Two - [15 marks]

Ten competitors at a music competition held at *Kenyatta International Conference Centre in Nairobi Kenya* in March 2015 were ranked by three adjudicators in the following order.

Chi-square $\chi^2 = \frac{(O-E)^2}{E}$

$r_{1,2}$ $r_{1,3}$ r_{23}

Competitor	A	B	C	D	E	F	G	H	I	J
Adjudicator 1	1	6	5	10	3	2	4	9	7	8
Adjudicator 2	3	5	8	4	7	10	2	1	6	9
Adjudicator 3	6	4	9	7	1	2	3	10	5	8

a) Determine the pair of adjudicators who have the nearest approach to common tastes in music festival and therefore give the higher correlation between them. (8 marks)

b) The average correlation amongst the three adjudicators. (2 marks)

c) What is partial correlation coefficient? (1 mark)

d) The following data was obtained from three students randomly sampled students' performance in the three subjects:

subject/student	student one	student two	student three	subject mean	subject standard deviation
BIOLOGY	35	22	20	30	4
CHEMISTRY	23	14	28	21	3
PHYSICS	39	24	20	28	5

I) Find out whether three performed better in physics or biology (1 mark)

II) What is the mean z-score for student one in all the three subjects? (1 mark)

III) What is the mean z-score for student two in all the three subjects? (1mark)

IV) In which subject did one of the students excel? Explain your response. (1 mark)

Question Three [15 marks]

The dean *of Faculty of Education, University of Nairobi* wants to determine whether there is a connection between academic rank of members of academic staff and their opinions concerning a proposed curriculum change in which postgraduate students are to be taught research methods on-line. Interviewing a sample of 313 members of the academic staff constituting 104 Lecturers, 131 Senior Lecturers, and 78 Professors, the Dean obtained the results shown in the following table:

Response/Rank	Lecturer	Senior Lecturer	Professor
Against	47	34	14
Not Committed	41	49	29
In Support	16	48	35

a) Follow systematic steps to test the null hypothesis that there is no difference of opinion concerning the curriculum change among the three groups of academic staff members at the university. Use a level of significance of $\propto = .05$. (6 marks)

b) A sample of 200 students from the University of Cape Coast in Ghana were interviewed regarding their level of enjoying on-line learning at the university by gender. The data is displayed below:

Gender/Level of Enjoyment	**Enjoy on line learning**	**Do not enjoy on line learning**
Male	85	60
Female	15	40

Determine the following:

i) The phi correlation coefficient (2 marks)

ii) The chi-square $(\underline{\chi}^2)$ for the data (2 marks)

iii) Given that $\underline{\chi}^2 = (\text{phi})^2 \times (N)$, calculate the new chi-square value using the phi correlation value obtained in (i) above (1 mark)

iv) Cramer's V coefficient (2 marks)

v) Yates corrected value for chi-square (2 marks)

Question four [15 marks]

A sample of nine students provided the following data regarding their KCPE Mathematics scores and Physics scores.

KCPE Math (x)	97	90	86	80	75	70	66	55	57
Physics (y)	90	85	80	75	70	65	60	50	60

a) Fit a line of best fit (linear regression) of Physics scores against KCPE Mathematics. (4 marks)

b) Estimate the simple regression equation Y on X. (2 marks)

c) Estimate the Physics score of a student who obtained 78% in KCPE Mathematics. (1 mark)

d) A researcher wishes to establish whether a student's grade point average (GPA) as variable (X_1), age (X_2) and common Board Examination (Y). Five students were selected and data generated displayed in table below:-

$Y = 500 + 100x$

Student	GPA (X_1)	Age (X_2)	Common Exams (Y)
A	3.2	22	550
B	2.7	27	570
C	2.5	24	525
D	3.4	28	670
E	2.2	23	490

i) Determine $r_{yx.1}$; $r_{yx.2}$ r_{x1x2} (3 marks)

ii) Determine multiple coefficient R using data in (i) above. (2 marks)

iii) Determine the coefficient of multiple determination R^2. (1 mark)

iv) Compute the adjusted R^2 ($R^2_{adj.}$) . (1 mark)

v) Interpret the use of adjusted R^2 when reporting multiple correlation results in a research report. (1 mark)

Question Five

An educational researcher was interested in finding out the effectiveness of instructional technologies used in teaching Physics. A random sample of students were placed in four different groups and their relative performance after one semester as shown in the table.

Frequent practical	Class demonstration	Excursion	Lecture method
24	18	12	08
21	16	15	10
22	14	10	06
20	12	08	04

a) At alpha .05 level of confidence using ANOVA test can we conclude that there were significant differences in students' performance among these four groups? (10 marks)

b) If significant differences exist then use Scheffe test to find out where the differences exist. (5 marks)

Anova - more than 3 groups

→ State hypothesis

→ Determine Critical value DFN DFD.

→ Determine the computed value.

SS MS Fs. F

Between

Within

Scheffes $\frac{X_1 - X_2}{MSw\left(1-\frac{1}{n_1}\right)\left(1-\frac{1}{nR}\right)}$.

TABLE ONE

TABLE L Critical values for Spearman's r_s*

No. of pairs	Level of significance			
	One-tailed test		Two-tailed test	
N	.01	.05	.01	.05
4	—	1.000	—	—
5	1.000	.900	—	1.000
6	.943	.829	1.000	.886
7	.893	.714	.929	.786
8	.833	.643	.881	.738
9	.783	.600	.833	.700
10	.745	.564	.794	.648
11	.709	.536	.755	.618
12	.678	.503	.727	.587
13	.648	.484	.703	.560
14	.626	.464	.679	.538
15	.604	.446	.654	.521
16	.582	.429	.635	.503

* To be significant, the r_s obtained from the data must be equal to or larger than the value shown in the table.

SOURCE: From "Testing the significance of Kendall's τ and Spearman's r_s," by M. Nijse, *Psychological Bulletin*, 1988, *103*, 235–237. Copyright 1988 by the American Psychological Association. Reprinted by permission.

TABLE TWO

TABLE A Critical values for Pearson product-moment correlation coefficients, r^*

df	α Levels (two-tailed test) .1	.05	.02	.01	.001
(df = N − 2)	α Levels (one-tailed test) .05	.025	.01	.005	.0005
1	.98769	.99692	.999507	.999877	.9999988
2	.90000	.95000	.98000	.990000	.99900
3	.8054	.8783	.93433	.95873	.99116
4	.7293	.8114	.8822	.91720	.97406
5	.6694	.7545	.8329	.8745	.95074
6	.6215	.7067	.7887	.8343	.92493
7	.5822	.6664	.7498	.7977	.8982
8	.5494	.6319	.7155	.7646	.8721
9	.5214	.6021	.6851	.7348	.8371
10	.4973	.5760	.6581	.7079	.8233
11	.4762	.5529	.6339	.6835	.8010
12	.4575	.5324	.6120	.6614	.7800
13	.4409	.5139	.5923	.6411	.7603
14	.4259	.4973	.5742	.6226	.7420
15	.4124	.4821	.5577	.6055	.7246
16	.4000	.4683	.5425	.5897	.7084
17	.3887	.4555	.5285	.5751	.6932
18	.3783	.4438	.5155	.5614	.6787
19	.3687	.4329	.5034	.5487	.6652
20	.3598	.4227	.4921	.5368	.6524
25	.3233	.3809	.4451	.4869	.5974
30	.2960	.3494	.4093	.4487	.5541
35	.2746	.3246	.3810	.4182	.5189
40	.2573	.3044	.3576	.3932	.4896
45	.2428	.2875	.3384	.3721	.4648
50	.2306	.2732	.3218	.3541	.4433
60	.2108	.2500	.2948	.3248	.4078
70	.1954	.2319	.2737	.3017	.3799
80	.1829	.2172	.2565	.2830	.3568
90	.1726	.2050	.2422	.2673	.3375
100	.1638	.1946	.2301	.2540	.3211

* To be significant the r obtained from the data must be equal to or larger than the value shown in the table.

SOURCE: Table VII of Fisher and Yates (1963), *Statistical Tables for Biological, Agricultural and Medical Research*, published by Longman Group Ltd., London (previously published by Oliver & Boyd, Edinburgh), and by permission of the authors and publishers.

TABLE THREE

TABLE F Chi square distribution*

df	α levels .10	.05	.02	.01	.001
1	2.71	3.84	5.41	6.64	10.83
2	4.60	5.99	7.82	9.21	13.82
3	6.25	7.82	9.84	11.34	16.27
4	7.78	9.49	11.67	13.28	18.46
5	9.24	11.07	13.39	15.09	20.52
6	10.64	12.59	15.03	16.81	22.46
7	12.02	14.07	16.62	18.48	24.32
8	13.36	15.51	18.17	20.09	26.12
9	14.68	16.92	19.68	21.67	27.88
10	15.99	18.31	21.16	23.21	29.59
11	17.28	19.68	22.62	24.72	31.26
12	18.55	21.03	24.05	26.22	32.91
13	19.81	22.36	25.47	27.69	34.53
14	21.06	23.68	26.87	29.14	36.12
15	22.31	25.00	28.26	30.58	37.70
16	23.54	26.30	29.63	32.00	39.25
17	24.77	27.59	31.00	33.41	40.79
18	25.99	28.87	32.35	34.80	42.31
19	27.20	30.14	33.69	36.19	43.82
20	28.41	31.41	35.02	37.57	45.32
21	29.62	32.67	36.34	38.93	46.80
22	30.81	33.92	37.66	40.29	48.27
23	32.01	35.17	38.97	41.64	49.73
24	33.20	36.42	40.27	42.98	51.18
25	34.38	37.65	41.57	44.31	52.62
26	35.56	38.88	42.86	45.64	54.05
27	36.74	40.11	44.14	46.96	55.48
28	37.92	41.34	45.42	48.28	56.89
29	39.09	42.56	46.69	49.59	58.30
30	40.26	43.77	47.96	50.89	59.70

* To be significant the χ^2 obtained from the data must be equal to or larger than the value shown in the table.

SOURCE: Table IV of Fisher and Yates (1963), *Statistical Tables for Biological, Agricultural and Medical Research*, published by Longman Group Ltd., London (previously published by Oliver & Boyd, Edinburgh), and by permission of the authors and publishers.

TABLE FOUR

TABLE D The *t* distribution*

	Confidence interval percents (two-tailed)					
	80%	90%	95%	98%	99%	99.9%
	Significance level for two-tailed test					
	.20	.10	.05	.02	.01	.001
	Significance level for one-tailed test					
df	.10	.05	.025	.01	.005	.0005
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.924
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.869
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.408
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.474	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

* To be significant the *t* obtained from the data must be equal to or larger than the value shown in the table.

SOURCE: Table III of Fisher and Yates (1963), *Statistical Tables for Biological, Agricultural and Medical Research*, published by Longman Group Ltd., London (previously published by Oliver & Boyd, Edinburgh), and by permission of the authors and publishers.

TABLE FIVE

$\chi^2 = \frac{(O-E)^2}{E}$ Chisquare

$E = \frac{C_T \times R_T}{N = \text{Grand total}}$

~~Table 8~~: The percentage of points for the F-Distribution
Degrees of freedom ($\alpha = .05$)

Degrees of freedom denominator (DFD)	*Degrees of freedom numerator (DFN)*							
		1	2	3	4	5	6	7
	1	161.4	199.5	215.7	224.6	230.2	234.0	236.8
	2	18.51	19.00	19.16	19.25	19.30	19.30	19.35
	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89
	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09
	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88
	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21
	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79
	8	5.32	4.46	4.07	3.84	3.69	3.58	3.50
	9	5.12	4.26	3.86	3.63	3.48	3.37	3.29
	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14
	11	4.84	3.98	3.59	3.36	3.20	3.09	3.01
	12	4.75	3.89	3.49	3.26	3.11	3.00	2.91
	13	4.67	3.81	3.41	3.18	3.03	2.92	2.83
	14	4.60	3.74	3.34	3.11	2.96	2.85	2.76
	15	4.54	3.68	3.29	3.06	2.90	2.79	2.71
	16	4.49	3.63	3.24	3.01	2.85	2.74	2.66
	17	4.45	3.59	3.20	2.96	2.81	2.70	2.61
	18	4.41	3.55	3.16	2.93	2.77	2.66	2.58
	19	4.38	3.52	3.13	2.90	2.74	2.63	2.54
	20	4.35	3.49	3.10	2.87	2.71	2.60	2.51
	21	4.32	3.47	3.07	2.84	2.68	2.57	2.49
	22	4.30	3.44	3.05	2.82	2.66	2.55	2.46
	23	4.28	3.42	3.03	2.80	2.64	2.53	2.44
	24	4.26	3.40	3.01	2.78	2.62	2.51	2.42
	25	4.24	3.39	2.99	2.76	2.60	2.49	2.40
	26	4.23	3.37	2.98	2.74	2.59	2.47	2.39
	27	4.21	3.35	2.96	2.73	2.57	2.46	2.37
	28	4.20	3.34	2.95	2.71	2.56	2.45	2.36
	29	4.18	3.33	2.93	2.70	2.55	2.43	2.35
	30	4.17	3.32	2.92	2.69	2.53	2.42	2.33
	40	4.08	3.23	2.84	2.61	2.45	2.34	2.25
	60	4.00	3.15	2.76	2.53	2.37	2.25	2.17
	120	3.92	3.07	2.68	2.45	2.29	2.17	2.09
		3.84	3.00	2.60	2.37	2.21	2.10	2.01

mean
mode — frequently
median — centre