W1-2-60-1-6

JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY
UNIVERSITY EXAMINATIONS 2023/2024

END OF SEMESTER EXAMINATION FOR THE DEGREE OF MASTER OF SCIENCE IN EPIDEMIOLOGY & BIOSTATISTICS, PUBLIC HEALTH, GLOBAL HEALTH, MOLECULAR MEDICINE, MEDICAL MICROBIOLOGY, MEDICAL VIROLOGY

PEH 3102/TID 3104: BIOSTATISTICS

DATE: AUGUST 2023 **TIME: 3 HOURS**

INSTRUCTIONS: ANSWER ANY FOUR QUESTIONS. EACH QUESTION IS 25 MARKS

QUESTION ONE (25 MARKS)

Suppose a scientist wants to demonstrate the association between exam score (dependent variable) and the predictor variables (independent variables) hours studied and preparatory exams taken. Our scientist thinks that each independent variable has a linear relation with ~~health care~~ Exam S costs. He therefore decides to fit a multiple linear regression model. The SPSS output is as below:

Regression Statistics	
Multiple R	0.857
R Square	0.734
Adjusted R Square	0.703
Standard Error	5.366
Observations	20

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	1350.76	675.38	23.46	0.00
Residual	17	489.44	28.79		
Total	19	1840.20			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	67.67	2.82	24.03	0.00	61.73	73.61
hours	5.56	0.90	6.18	0.00	3.66	7.45
prep_exams	-0.60	0.91	-0.66	0.52	-2.53	1.33

a) Write the regression equation arising from this regression model **(5marks)**

b) Interpret the results with regard to:

i) The coefficients for hours and pre_exams **(5 marks)**

ii) Significance (p-value) **(5 marks)**

iii) The regression statistics **(5 marks)**

iv) The ANOVA table **(5 marks)**

QUESTION TWO (25 MARKS)

Describe the assumptions of simple linear regression and how to deal with any violations of these assumptions **(25 marks)**

QUESTION THREE (25 MARKS)

a) Describe the statistical requirements for calculation of a sample size in simple random sampling to determine the prevalence of a disease or medical condition **(15 marks)**

b) In calculating the sample size in a) above, the result is 276 samples. However, the target population is 5,500 individuals. Calculate the final sample size and give the rationale for this new sample size **(10 marks)**

QUESTION FOUR (25 MARKS)

a) Describe the six steps of hypothesis testing **(12 marks)**

b) In a welfare check at a school, the team is concerned that the children in the PP2 class appear to be generally of lower weight than that expected at their age. All 30 children are weighed and a two-tailed test carried out assuming a confidence level of 95%. (The t-table is provided).

i. Which t-test is being conducted **(2 marks)**

ii. State the hypothesis being tested **(2 marks)**

iii. Once the t-test statistic is obtained, what critical value should it be compared to? **(2 marks)**

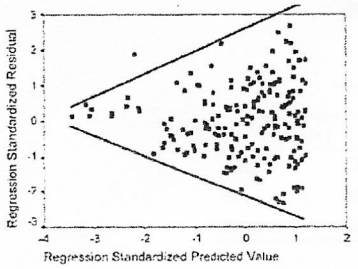
iv. Suppose that the test statistic obtained is 2.756, give the result interpretation including the associated p-value **(2 marks)**

v. How else might you be able to graphically bring out the weight problem what might still be present in the class **(5 marks)**

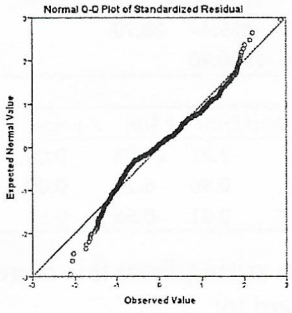
QUESTION FIVE (25 MARKS)

a) Regarding linear correlation, explain the instances in which the correlation coefficient **should not** be calculated **(5 marks)**

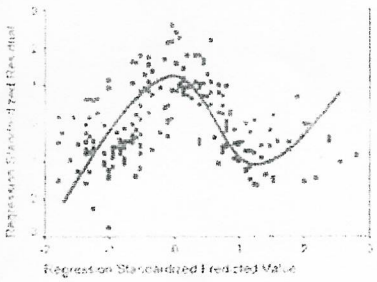
b) In each case in the diagrams below, indicate the assumption for linear regression that is being tested and the interpretation: **(5 marks for each)**



i)



ii)



iii)

	HHNo	sex	age	div	cros	water	graze	salt
HHNo	1.000	-.089	.005	.175	-.040	.085	.045	.140
sex	-.089	1.000	.086	.107	.061	.035	-.044	-.024
age	.005	.086	1.000	.036	.073	.129	.093	.096
division	.175	.107	.036	1.000	.041	.322	.260	.313
crosses	-.040	.061	.073	.041	1.000	.039	.013	.077
water	.085	.035	.129	.322	.039	1.000	.902	.786
graze	.045	-.044	.093	.260	.013	.902	1.000	.816
salt	.140	-.024	.096	.313	.077	.786	.816	1.000

iv)

QUESTION SIX (25 MARKS)

a) Explain the instances in which non-parametric tests would be more appropriate in data analysis in biomedical research **(6 marks)**

b) State the equivalent non-parametric test for the following parametric tests **(12 marks)**

Parametric test	Non-parametric test
One-sample t-test	
Paired t-test	
Two-sample t-test	
One-way ANOVA	
Two-way ANOVA	
Pearson correlation coefficient	

c) In a Kruskal Wallis test to test the differences in medians in a three-sample data set with 5, 4 and 3 data points, the calculated value H comes out to be 6.0778 and the critical value for a 5: 4: 3 sample at α=0.05 is 5.656. Give the conclusion about the medians **(7 marks)**

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										