National University of Computer and Emerging Sciences, Lahore Campus				
	Course Name:	Advance Statistics	Course Code:	DS2003
	Degree Program:	BS (DS)	Semester:	Spring 2023
	Exam Duration:	60 Minutes	Total Marks:	30
	Paper Date:	11-April-2023	Weight	15%
	Sections:	ALL	No of Page(s):	3
	Exam Type:	Rubric Midterm II		

Question No. 1 [CLO 2] **(Marks 8)**

FAST University conducted a Job fairs for their graduate to attain the job or internship in different software companies. A hiring manager of Software Company interviews candidates, one by one, to fill a vacancy. The manager is interested to know the number of candidates interviewed until one candidate receives an offer. He conducted the experiment 250 times; the results are presented in table.

X	**1**	**2**	**3**	**4**	**5**	**6**
F	**140**	**70**	**30**	**8**	**1**	**1**

Test the hypothesis at 5 percent level of significance that the observed distribution of *x* may be fitted by the geometric distribution.

Solution:

Mean	Estimate of p	Calculation of prob (pi)	Calculation of expected (ei)	Merging of ei's for less than 5	Calculation of $(oi-ei)^{2}/ei$	Critical Value	Hypothesis	Decision
1	**1**	**1**	**1**	**1**	**2**	**1**	**1**	**1**

x	**f**	**fx**	**pi**	**ei=pi*250**	**ei (merge)**	**(oi-eo)2/ei**
1	140	140	0.6053	151.3317	151.3317	0.8485
2	70	140	0.2389	59.7266	59.7266	1.7671
3	30	90	0.0943	23.5725	23.5725	1.7526
4	8	32	0.0372	9.3034	9.3034	0.1826
5	1	5	0.0147	3.6718	5.1210	1.9021
6	1	6	0.0058	1.4492		
Sum	**250**	**413**		**249.0551**	**249.0551**	**6.4529**

Mean= **1.652**

Chi-sqr (claculated) = 6.45

Chi-sqr (5-1-1)(0.05) **7.83**

Pmf (1-p)^(k-1)*p

Decision: **Do not reject Ho.**

p=1/mean **0.605327**

Question No. 2 [CLO 3 & 5] **(Marks 10)**

A freelancer wants to determine whether the average earning of the different services in the fiverr account are the same or not. He chooses the three services Graphics & Designs, Digital Marketing and Writing & Translation. From these services, he selects the 5 projects and records the earning of these projects. The data are presented in Table

Graphics & Designs	Digital Marketing	Writing & Translation
44.58	11.80	2.972
59.44	47.40	14.86
74.30	62.41	44.58
71.90	13.37	14.86
77.50	89.17	14.86

The model adequacy is examined by plotting the residuals. The residuals plots are given as

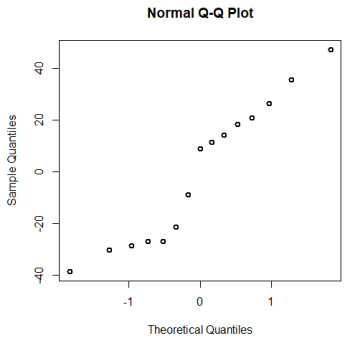


Fig. 1(a)

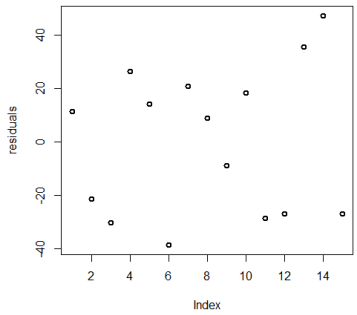


Fig. 1 (b)

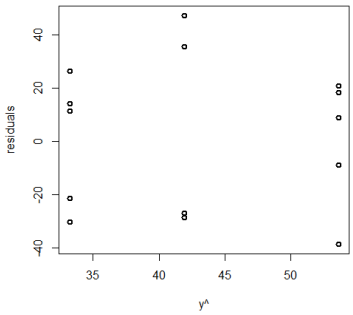
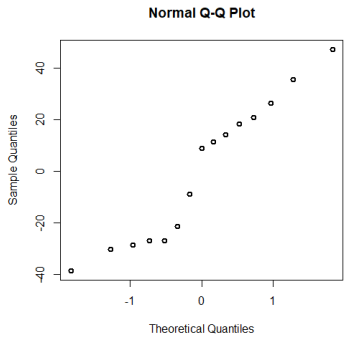
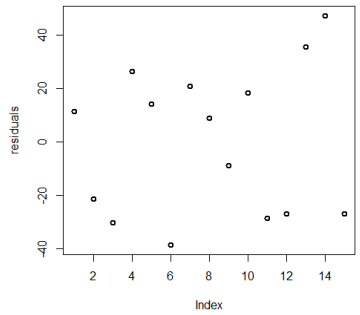
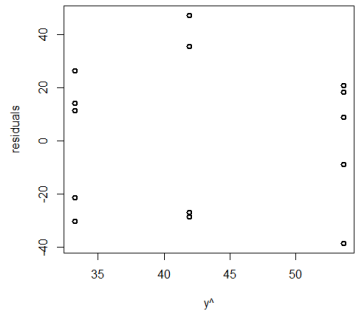


Fig. 1(c)

(1) Analyze each residual plot provided in Fig. 1 concerning the assumptions of ANOVA and comment on model adequacy. Specify which assumption is tested with Fig. 1: (a), (b) and (c).
(2) After examining part (1) Test the hypothesis that the three earning sources have the same earning capabilities by an appropriate technique.

Solution:

Assumption 1	Assumption 2	Assumption 3
The residuals are normally distributed	Observations of the residuals are independent	The residuals have the same variances.
		
This Fig indicates that the residuals are not normally distributed because that QQ plot does not show a straight line. Which indicates that the assumption of **normality is not met** (Violation of assumption leads us towards non-parametric test)	From the above fig, we can see some randomness in the observations. So this assumption is **fulfilled**	There is heteroscedasticity observed in this Fig. The assumption of equal variance is **also not met**.
1.5	1.5	1.5

Non-parametric Test

R1=56.5 R2= 40, R3=23.5 W=5.445 Table value= 5.99; Do not Reject Ho.

Question No. 3 [CLO 2 & 5] **(Marks 12)**

A software engineer wants to determine the effect of network speed and systems capacity on the downloading time. She choose the three different net packages (6Mbps, 8Mbps, 25Mbps) and 4 different systems of core i5(8^{th}, 10^{th},11^{th}, 12th) generation. The analysis of variance for these data are

S.O.V	df	SS	MS	F-ratio	P-value
Between Net packages	-	87.73	-	-	0.21
Between Systems	-	-	-	-	0.37
Error	-	-	14.70		
Total	**-**	**230.05**			

a) Fill in the missing entries
b) Conclude at a 5% level of significance whether
   (i) the different net packages have different downloading times.
   (ii) the different types of systems have different downloading times.
c) The result of the Tukey posthoc test for net packages is given in Figure 2. Explain the result if applicable.

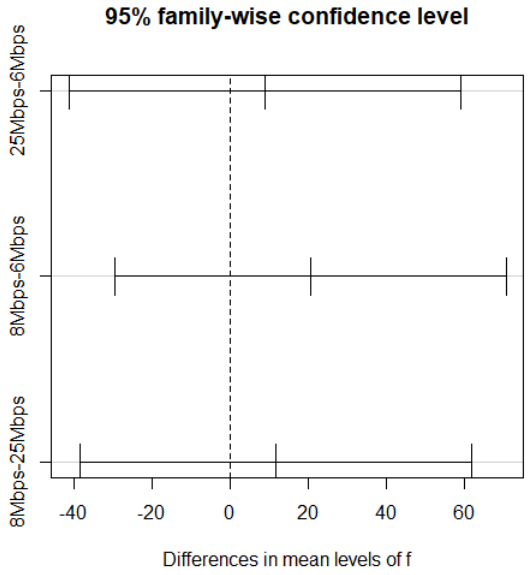


Figure 2: Difference in means of downloading time with respect to net packages

Solution:

(a) 1 mark for each missing value x 10 = 10

S.O.V	df	SS	MS	F-ratio	P-value
Between Net packages	**2**	87.73	**43.86**	**2.98**	0.21
Between Systems	**3**	**54.12**	**18.04**	**1.23**	0.37
Error	**6**	**88.2**	14.70		
Total	**11**	230.05			

(b)

i) Do not reject Ho. (No difference between network packages) 1+1.5=**2.5 Marks**

ii) Do not reject Ho. (No difference between Systems) 1+1.5=**2.5 Marks**

(c) Since we do not reject our null hypothesis, so Tukey Test is not applicable **2 Marks.**

If a student explains Tukey Test he will award **ZERO mark**, as Test is not applicable.

END OF RUBRIC