SECI 2143 PROBABILITY AND STATISTICAL DATA ANALYSIS

PROJECT 2: STUDENT PERFORMANCE

LECTURER: DR SHARIN HAZLIN BINTI HUSPI

SECTION : 01

NO	NAME	MATRIC NO
1.	GUI YU XUAN	A20EC0039
2.	PHANG CHENG YI	A20EC0131
3.	NG YEN THONG	A20EC0107
4.	GOH YITIAN	A20EC0038

1.0 INTRODUCTION

The student performance has played an important role in the development of the country. This is because a good student performance can train high-quality students and become human resources for country and social development. (Ali et.al, 2009). Therefore, many schools will work hard in all aspects to improve student performance. Student performance can be evaluated from many aspects such as classroom participation and test score. According to the PISA result done by OECD (2019), most of the students in Malaysia scored below OECD average in reading, mathematics and science. This PISA result increases our interest in investigating the student performance in Malaysia. Our aim of this project is to determine gender and test preparation courses in affecting the score of mathematics, reading and writing. However, our hypothesis is there is no difference between male and female in student performance and there are significant differences in the preparation for tests in terms of affecting scores.

2.0 METHODOLOGY

The dataset used in this research is obtained from a website called Kaggle (https://www.kaggle.com/spscientist/students-performance-in-exams). The dataset is secondary data. The targeted population is high school students in the United Kingdom. Inferential statistics are carried out by using hypothesis testing two samples, correlation, regression, and chi square test of independence.

3.0 DATA SET

This data set consists of eight variables which are gender, race or ethnicity, parental level of education, lunch, test preparation course, math score, reading score and writing score. The variables that we choose for our statistical analysis are gender, test preparation course, math score, reading score and writing score. Besides, we select the first 100 students as the sample for our statistical analysis from the origin dataset.

Statistical Test	Description	Variables
Hypothesis Test 2 Sample	Mean of Math score for both gender is different	Gender, Math score
Correlation Test	Determine whether there is a linear relationship between the reading score and the writing score	Reading score, Writing score
Regression Test	Determine whether there is a linear relationship between math score and reading score	Math score, Reading score
Chi-square Test of Independence	Student performance in writing is dependent on test preparation course	Test preparation course, Writing score

4.0 DATAANALYSIS

4.1 Hypothesis Testing Two- Sample

We wish to determine whether there is any difference between male's mathematics score and female's mathematics score under t-test 0.05 significance level. However, population variance is unknown. Hence, we assume that the sample is normally distributed since the sample size is large. Thus, Z-test is used instead of t-test.

Let μ_1 = sample mean of male's mathematics score

μ_2 = sample mean of female's mathematics score

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

Z-value is obtained from R. We fail to reject the null hypothesis if the Z-value calculated is in the range of critical value, $-1.959964 < Z_{0.025} < 1.959964$. Since Z-value = -0.056 is in the range of critical value, hence we fail to reject H_0. We can conclude that at the 0.05 significance level, there is insufficient evidence to conclude that the sample mean of male's mathematics score and the sample mean of female's mathematics score is different. In conclusion, there is no significant difference between genders in mathematics performance.

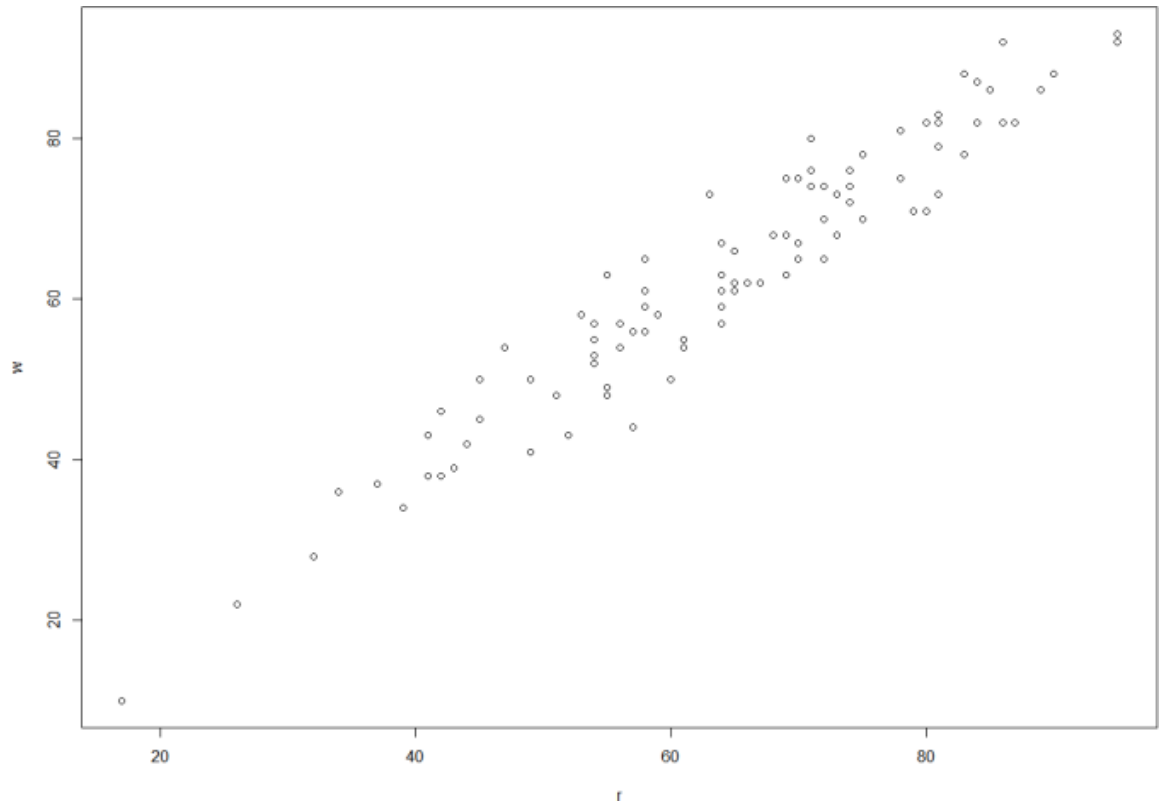
4.2 Correlation

In this test, a random sample of 100 students have been selected. Variables used in this test are reading score and the writing score. We wish to check if there is a linear relationship between the reading score and the writing score at the 0.05 level of significance.

H_0: There is no linear correlation between reading score and the writing score

H_1: There is a linear correlation exists between reading score and the writing score

Based on the result computed, $t > t_{0.025,98}$ （35.65 > 1.9845）.
Therefore, H_0 is rejected. There is sufficient evidence that there is a linear relationship between reading score and the writing score at the 5% level of significance.

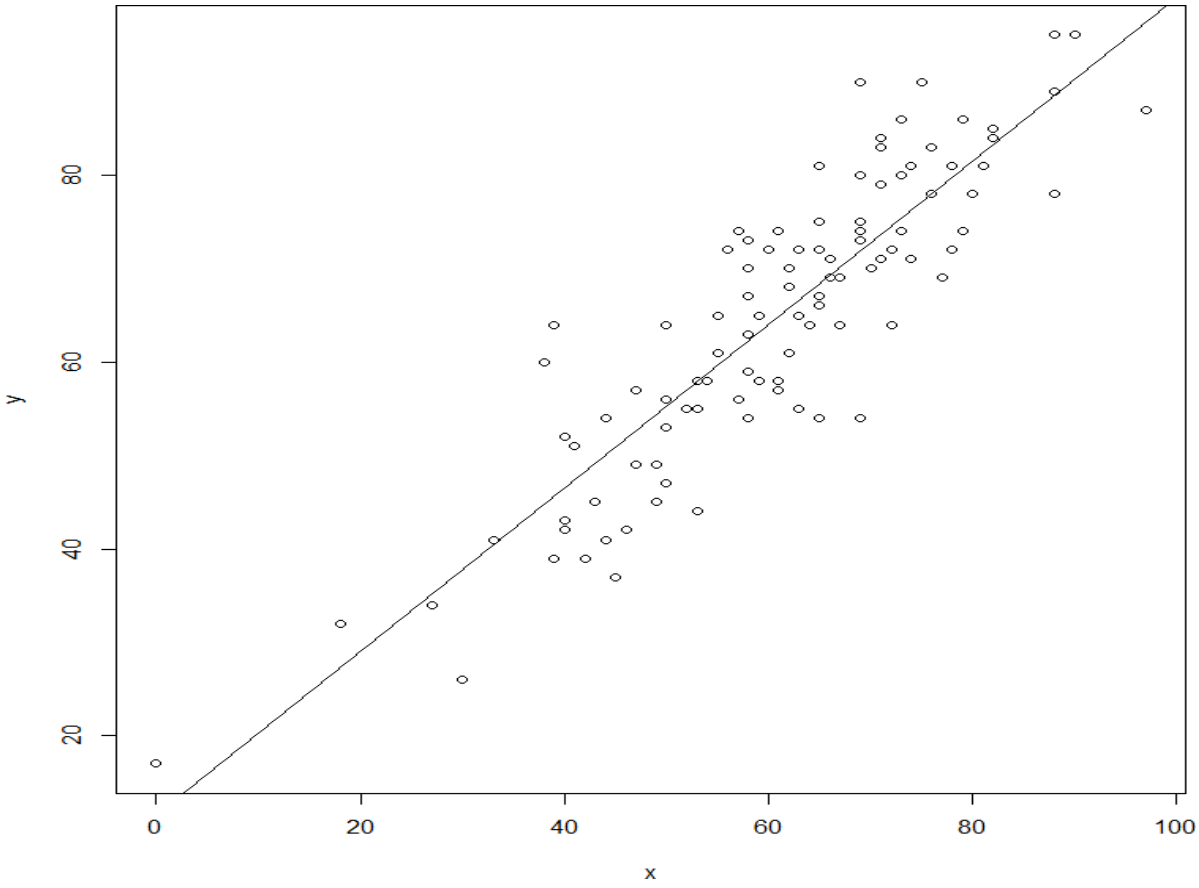


It can be seen from the graph that the writing score increases as the reading score increases. A scatter plot and correlation analysis of the data indicates that there is a positive relationship between reading score and writing score.

Based on the computed correlation value which is 0.9635, we can conclude that there is a relatively strong positive linear association between reading score and writing score of students.

4.3 Regression

We use the regression test to test the presence of a linear relationship between independent variable, x and dependent variable, y. The independent variable in this regression is math score while the dependent variable in this regression is reading score. A random sample of 100 students have been selected to determine whether there is a linear relationship between math score and reading score.



The scatter plot above shows that there is a positive linear relationship between math score (x) and reading score (y). Its least squares equation is $\hat{y}$=11.48582+0.87538*x. From the equation, we can know that the slope coefficient, b_1 is 0.87538 which tells us that the estimated changes in average reading score increased by 0.87538 on average.

Moreover, we can know that the value of y-intercept is 11.48582 which represents the intersection coefficient, b_0. This indicates that the estimated score for reading is 11.48582 when the math score is 0.

Since the value of coefficient of determination, $R^2 = 0.7803$ which is near to 1, we can consider that the linear relationship between x and y is quite strong.

4.4 Chi-square test of independence

A random sample of 100 students are selected, and we wish to test whether there is a relationship between the test preparation course status of students and their writing scores at 0.05 significance level.

The null hypothesis, H_0: The test preparation course status of students and their writing scores are independent.

The alternative hypothesis, H_1: The test preparation course status of students and their writing scores are related.

According to the result above, we can observe that the chi-square value (x-squared) =12.587 while the critical value $X^2_{(0.05,3)} = 7.815$. Since the x-squared value (12.587) is greater than (>) the critical value (7.815) means it falls in the critical region, we reject the null hypothesis. Therefore, there is insufficient evidence to conclude that the test preparation course status of students and their writing scores are independent which indicates that the test preparation course status of students does not influence their test writing scores.

5.0 CONCLUSION

Based on all the findings through analytical study, we conclude that students' performance is affected by a few aspects. As a result, we have analyzed all aspects that may influence the students' performance. Firstly, we found that there is no significant difference between genders of students in their mathematics performance. Secondly, we found that there is a relatively strong positive linear relationship between the students' reading score and writing score as the writing score increases when the reading score increases. Besides, we also figured out that the reading score is affected by the mathematics score as the linear relationship between them is quite strong. Finally, we observed that the test preparation course status of students does not influence their test writing scores. In addition, through these analyses, we have learnt to conduct inference statistical analysis using knowledge we learned and efficient tools like R Software. This study teaches us to get reasonable conclusions regarding real life's problems by the knowledge we have.

6.0 REFERENCES

Ali, N., Jusof, K., Ali, S., Mokhtar, N., & Salamat, A. S. A. (2009). THE FACTORS INFLUENCING STUDENTS' PERFORMANCE AT UNIVERSITI TEKNOLOGI MARA KEDAH, MALAYSIA. *Management Science and Engineering*, *3*(4), 2. http://flr-journal.org/index.php/mse/article/viewFile/j.mse.1913035X20090304.010/820

OECD (2019), PISA 2018 Results (Volume I-III): What 15-year-old students in Malaysia know and can do. PISA, OECD Publishing, Paris. https://www.oecd.org/pisa/publications/PISA2018_CN_MYS.pdf

APPENDIX

	gender	test preparation course	math score	reading score	writing score
1	female	none	72	72	74
2	female	completed	69	90	88
3	female	none	90	95	93
4	male	none	47	57	44
5	male	none	76	78	75
6	female	none	71	83	78
7	female	completed	88	95	92
8	male	none	40	43	39
9	male	completed	64	64	67
10	female	none	38	60	50
11	male	none	58	54	52
12	male	none	40	52	43
13	female	none	65	81	73
14	male	completed	78	72	70
15	female	none	50	53	58
16	female	none	69	75	78
17	male	none	88	89	86
18	female	none	18	32	28
19	male	completed	46	42	46
20	female	none	54	58	61
21	male	none	66	69	63
22	female	completed	65	75	70
23	male	none	44	54	53
24	female	none	69	73	73
25	male	completed	74	71	80
26	male	none	73	74	72
27	male	none	69	54	55
28	female	none	67	69	75
29	male	none	70	70	65
30	female	none	62	70	75
31	female	none	69	74	74
32	female	none	63	65	61
33	female	none	56	72	65
34	male	none	40	42	38
35	male	none	97	87	82
36	male	completed	81	81	79
37	female	none	74	81	83
38	female	none	50	64	59
39	female	completed	75	90	88
40	male	none	57	56	57
41	male	none	55	61	54
42	female	none	58	73	68
43	female	none	53	58	65
44	male	completed	59	65	66
45	female	none	50	56	54
46	male	none	65	54	57
47	female	completed	55	65	62
48	female	none	66	71	76
49	female	completed	57	74	76
50	male	completed	82	84	82
51	male	none	53	55	48
52	male	completed	77	69	68
53	male	none	53	44	42
54	male	none	88	78	75
55	female	completed	71	84	87
56	female	none	33	41	43
57	female	completed	82	85	86
58	male	none	52	55	49
59	male	completed	58	59	58
60	female	none	0	17	10
61	male	completed	79	74	72
62	male	none	39	39	34
63	male	none	62	61	55
64	female	none	69	80	71
65	female	none	59	58	59
66	male	none	67	64	61
67	male	none	45	37	37
68	female	none	60	72	74
69	male	none	61	58	56
70	female	none	39	64	57
71	female	completed	58	63	73
72	male	completed	63	55	63
73	female	none	41	51	48
74	male	none	61	57	56
75	male	none	49	49	41
76	male	none	44	41	38
77	male	none	30	26	22
78	male	completed	80	78	81
79	female	completed	61	74	72
80	female	none	62	68	68

79	male	completed	80	78	81
80	female	completed	61	74	72
81	female	none	62	68	68
82	female	none	47	49	50
83	male	none	49	45	45
84	male	completed	50	47	54
85	male	none	72	64	63
86	male	none	42	39	34
87	female	none	73	80	82
88	female	none	76	83	88
89	female	none	71	71	74
90	female	none	58	70	67
91	female	none	73	86	82
92	female	none	65	72	74
93	male	none	27	34	36
94	male	none	71	79	71
95	male	completed	43	45	50
96	female	none	79	86	92
97	male	completed	78	81	82
98	male	completed	65	66	62
99	female	completed	63	72	70
100	female	none	58	67	62
101	female	none	65	67	62