NAME: JEREMIAH MAINA

ASSESSMENT ITEM 2 RESEARCH REPORT (2019 S1)

TASK 1

a) i) Test if there is any difference in Results on average between male and female students at the 5% level of significance.

To test if there is any difference in results on average between males and females we use the two tailed independent sample T test to carry out mean comparison. Our hypothesis for the test will be;

H_0: Average results of males is equal to the average results of females

H_1: Average results of males is equal to the average results of females

Findings

Independent Samples Test

	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	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	.003	.957	-3.321	647	.001	-4.249	1.280	-6.762	-1.737
Equal variances not assumed			-3.285	547.464	.001	-4.249	1.294	-6.790	-1.708

Conclusion

Since the t-values are in between the 95% Confidence Intervals of the Difference, we do not have enough evidence to reject the null hypothesis that the average results of males is equal to the average results of females. We conclude that the average results of both genders are not significantly different.

ii) Test at the level of significance of 5% if students who are involved in a romantic relationship have a lower result than the students who are not.

To test if students who are involved in a romantic relationship have a lower result than the students who are not. We carry out an independent sample T test for mean comparison between the two groups in question. Our hypothesis for the test will be;

H_0: The average results of students involved in a romantic relationship is less than the average results of students who are not

H_1: The average results of students involved in a romantic relationship is greater than or equal to the average results of students who are not.

Findings

Independent Samples Test

	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	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	3.825	.051	-2.323	647	.021	-3.043	1.310	-5.616	-.471
Equal variances not assumed			-2.222	433.076	.027	-3.043	1.370	-5.736	-.351

Conclusion

Since the t-values are less than the upper tail values in the 95% Confidence Intervals of the Difference, we fail to reject the null hypothesis that the average results of students involved in a romantic relationship is less than the average results of students who are not.

TASK 2(Regression Analysis)

(a) i) Construct a correlation matrix of all the quantitative variables in the dataset comment briefly on the linear associations between Result and other quantitative variables (viz. Age, Lectures and Tutorials) and whether these variables are a predictor of result.

Correlations

		Age	Result	LECTURES NOT ATTENDED	TUTORIALS NOT ATTENDED
Age	Pearson Correlation	1	-.107**	.152**	.147**
	Sig. (2-tailed)		.006	.000	.000
	N	649	649	649	649
Result	Pearson Correlation	-.107**	1	-.092*	-.084*
	Sig. (2-tailed)	.006		.019	.032
	N	649	649	649	649
LECTURES NOT ATTENDED	Pearson Correlation	.152**	-.092*	1	.946**
	Sig. (2-tailed)	.000	.019		.000
	N	649	649	649	649
TUTORIALS NOT ATTENDED	Pearson Correlation	.147**	-.084*	.946**	1
	Sig. (2-tailed)	.000	.032	.000	
	N	649	649	649	649

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

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

The variables Age, lectures not attended and tutorials not attended are all predictor variables of results. This is because there is a significant correlation between the variables and results at the 0.01 or 0.05 level of significance. Also, there is a negative linear relationship between the variables and the results.

(b) Conduct a simple regression on:

(i) Age is a predictor of Result?

Coefficients[a]

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		

1	(Constant)	86.145	9.751		8.835	.000
	Age	-1.419	.519	-.107	-2.735	.006

a. Dependent Variable: Result

(ii) Lectures is a predictor of Result?

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	60.704	.804		75.488	.000
	LECTURES NOT ATTENDED	-.319	.136	-.092	-2.344	.019

a. Dependent Variable: Result

(iii) Tutorials is a predictor of Result?

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	61.227	1.009		60.697	.000
	TUTORIALS NOT ATTENDED	-.545	.254	-.084	-2.149	.032

a. Dependent Variable: Result

Does this support your answers to Task 2 (a)?

Yes, this is because all the variables when regressed against results, they have a negative relationship with the dependent variable (results).

(c). For each of the independent variables contained in the regression model in Step 1, test their statistical significance. In testing statistical significance of a regression coefficient.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.084[a]	.007	.006	16.110	.007	4.618	1	647	.032

a. Predictors: (Constant), TUTORIALS NOT ATTENDED

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.092[a]	.008	.007	16.099	.008	5.496	1	647	.019

a. Predictors: (Constant), LECTURES NOT ATTENDED

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.107[a]	.011	.010	16.075	.011	7.478	1	647	.006

a. Predictors: (Constant), Age

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.230[a]	.053	.051	15.736	.053	35.985	1	647	.000

a. Predictors: (Constant), Mothers highest level of education

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.129[a]	.017	.015	16.031	.017	11.029	1	647	.001

a. Predictors: (Constant), Gender

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.091[a]	.008	.007	16.100	.008	5.395	1	647	.021

a. Predictors: (Constant), Relationship

TASK 3(Summary report);

According to the research analysis, there is no gender difference in academic achievement of the students. This is because the t-values from the test are in between the 95% Confidence Intervals of the Difference therefore we do not have enough evidence to reject the null hypothesis that the average results of males is equal to the average results of females. Additionally, being in a romantic relationship seems to have an effect on the academic performance of the students. This is because the t-values from the test are less than the upper tail values in the 95% Confidence Intervals of the Difference therefore we fail to reject the null hypothesis that the average results of students involved in a romantic relationship is less than the average results of students who are not and conclude that the average results of students involved in a romantic relationship is less than the average results of students who are not.

A regression analysis of different variables on results was further conducted to investigate the relationship of the variables to results. From the findings we can conclude that the variables Age, lectures not attended and tutorials not attended are all predictor variables of results. This is because there is a significant correlation between the variables and results at the 0.01 or 0.05 level of significance. Also, there is a negative linear relationship between the variables and the results, this means that as the variable values increases, the results will decrease.

For this study, I would recommend the use of a multiple linear regression model which is a versatile model for prediction when both categorical and continuous variables are in play. To check for the significance of the model we look for the r-square to see which percentage of the data has been plotted by the model, the final model for this research would be;

$$Results = -0.067(age) + 0.238(medu) + 0.85(relationship) - 0.056(tutorials\ missed) + 0.169(gender) - 0.075(lectures\ missed) + 55.314$$

For instance to predict the Result of a female student who is 18, whose mother has post-graduate qualifications, is not involved in a romantic relationship and attended all classes. We input the variable values to the above model to predict her results;

$$Results = -0.067(18) + 0.238(3) + 0.85(2) - 0.056(0) + 0.169(2) - 0.075(0) + 55.314$$

$$Results = 56.86$$

Although this study takes into consideration some variables that influence academic performance, it leaves out some of the vital variables that may impact the results too. Such variables include, the student's health, the coursework difficulty, study habits, the learning environment etc.