

Research Report

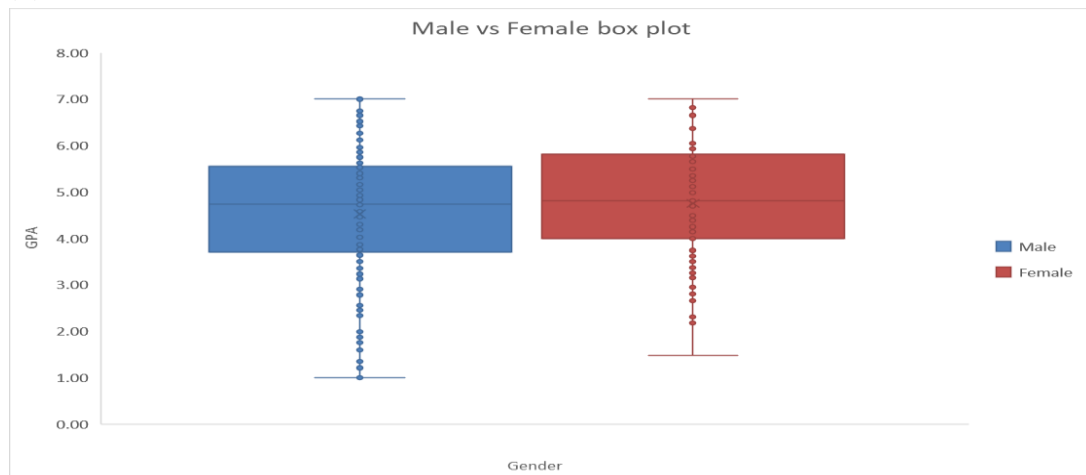
ASSESSMENT ITEM 2

BRENDAN DONCASTER

N9819363

Task 1 (Boxplot and t-tests)

1. (a)



Females GPA both has a higher central location than Male GPA and is skewed upwards towards higher GPA whereas males are skewed downwards. Females also have less of a spread compared to males.

(b)

Since the test is for any difference, the t-test will be two-tailed, to test if higher or lower average GPA.

$H_0: \text{MAverage} = \text{FAverage}$, $H_1: \text{MAverage} \neq \text{FAverage}$

Therefore, if $P(T \leq t) > 0.05$, there is significant difference.

Variance is assumed equal as the standard deviations of males and females are similar.

t-Test: Two-Sample Assuming Equal Variances

	Male	Female
Mean	4.522965517	4.752025316
Variance	1.96737795	1.403093281
Observations	145	79
Pooled Variance	1.769115769	
Hypothesized Mean Difference	0	
df	222	
t Stat	1.231527989	
P(T<=t) one-tail	0.109714208	
t Critical one-tail	1.651746359	
P(T<=t) two-tail	0.219428416	
t Critical two-tail	1.970707395	

The $P(T \leq t)$ two-tail result is 0.219, which is greater than the significance level of 0.05 therefore there is not strong evidence to accept the base hypothesis that male and female students have no difference in average GPA.

2.

(a)

Since the test is for specifically higher GPA, the t-test will be one-tailed.

$H_0: \text{PAverage} = \text{UAverage}$, $H_1: \text{PAverage} > \text{UAverage}$

Postgraduate parents' students have a higher average sample GPA. Therefore, if $P(T \leq t) < 0.05$, there is significant proof of achieving a higher GPA.

Variance is assumed equal as the standard deviations of postgraduates and undergraduates are similar.

t-Test: Two-Sample Assuming Equal Variances

	<i>Undergrad</i>	<i>Postgrad</i>
Mean	4.88647619	5.1040625
Variance	1.56784804	1.238366835
Observations	105	32
Pooled Variance	1.492189393	
Hypothesized Mean Difference	0	
df	135	
t Stat	-0.882122596	
P(T<=t) one-tail	0.189639238	
t Critical one-tail	1.656219133	
P(T<=t) two-tail	0.379278477	
t Critical two-tail	1.977692277	

$P(T \leq t)$ one-tail result is 0.189, which is greater than the significance level of 0.05, therefore there is not strong evidence to reject the hypothesis that postgraduate and undergraduate parents' students have no difference in average GPA, therefore there is strong evidence that postgraduate parents' students do not achieve a higher GPA.

(b)

Since the test is for specifically higher GPA, the t-test will be one-tailed.

H0: UAverage = SAverage , H1: UAverage > SAverage

Undergraduate parents' students have a higher average sample GPA. Therefore, if $P(T \leq t) < 0.05$, there is significant proof of achieving a higher GPA.

Variance is assumed equal as the standard deviations of undergraduates and secondary or lower are similar.

t-Test: Two-Sample Assuming Equal Variances

	<i>Secondary</i>	<i>Undergrad</i>
Mean	4.078505747	4.88647619
Variance	1.785773323	1.56784804
Observations	87	105
Pooled Variance	1.666487905	
Hypothesized Mean Difference	0	
df	190	
t Stat	-4.317159085	
P(T<=t) one-tail	1.27085E-05	
t Critical one-tail	1.652912949	
P(T<=t) two-tail	2.54169E-05	
t Critical two-tail	1.972528182	

$P(T \leq t)$ one-tail result is 0.000012, which is less than the significance level of 0.05, therefore there is strong evidence to reject the hypothesis that undergraduate and secondary or lower parents' students have no difference in average GPA, therefore there is strong evidence that undergraduate parents' students achieve a higher GPA.

Task 2 (Regression Analysis)

3.

	<i>GPA</i>	<i>HS_SCI</i>	<i>HS_ENG</i>	<i>HS_MATH</i>	<i>ATAR</i>
GPA	1				
HS_SCI	0.344282	1			
HS_ENG	0.304046	0.579375	1		
HS_MATH	0.444387	0.575686	0.446887	1	
ATAR	0.423808	0.852447	0.763957	0.797436	1

All quantitative variables have a positive association towards GPA. Science and English have a moderate association towards GPA with 0.344 and 0.304, whereas Math and ATAR have a strong association with 0.444 and 0.423 respectively.

4. Based on the correlation matrix obtained in Question 3, would you say that:

- (i) HS_SCI is a predictor of GPA?

Based on the correlation coefficient 0.344, HS_SCI is a moderate predictor of GPA.

When performing simple regression for HS_SCI onto GPA, the results are:

SUMMARY
OUTPUT

Regression Statistics	
Multiple R	0.344281789
R Square	0.11852995
Adjusted R Square	0.114559365
Standard Error	1.253026831
Observations	224

ANOVA					
	df	SS	MS	F	Significance F
Regression	1	46.86992495	46.86992495	29.85200578	1.24741E-07
Residual	222	348.556925	1.570076239		
Total	223	395.42685			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	2.42181339	0.408032996	5.93533713	1.11773E-08	1.61769975	3.225927038	1.61769975	3.225927038
HS_SCI	0.26973167	0.049367962	5.46369891	1.24741E-07	0.172441871	0.367021484	0.172441871	0.367021484

$H_0: \beta_1 = 0$ (no relationship)

$H_0: \beta_1 \neq 0$ (some relationship)

$\alpha = 5\% \rightarrow Pvalue < \alpha \rightarrow \text{reject } H_0$

$\rightarrow Pvalue > \alpha \rightarrow \text{accept } H_0$

Since checking for $\beta_1 \neq 0$, Pvalue is two tailed

$\therefore Pvalue = 1.2474E^{-7}$

$Pvalue < \alpha \therefore \text{reject } H_0$

Therefore, simple regression supports the claim that HS_SCI is a predictor of GPA.

(ii) HS_ENG is a predictor of GPA?

Based on the correlation coefficient 0.304, HS_ENG is a moderate predictor of GPA.

When performing simple regression for HS_ENG onto GPA, the results are:

SUMMARY
OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.304045999
R Square	0.092443969
Adjusted R Square	0.088355879
Standard Error	1.271432515
Observations	224

<i>ANOVA</i>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	36.55482764	36.55482764	22.61299636	3.56432E-06
Residual	222	358.8720224	1.616540641		
Total	223	395.42685			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	2.430524	0.464838	5.228749	3.9359	1.514463	3.346585	1.514463	3.346585
HS_ENG	0.268506	0.056464	4.755312	3.5643	0.157231	0.379781	0.157231	0.379781

$H_0: \beta_1 = 0$ (no relationship)

$H_0: \beta_1 \neq 0$ (some relationship)

$\alpha = 5\% \rightarrow P\text{value} < \alpha \rightarrow \text{reject } H_0$

$\rightarrow P\text{value} > \alpha \rightarrow \text{accept } H_0$

Since checking for $\beta_1 \neq 0$, $P\text{value}$ is two tailed

$\therefore P\text{value} = 3.5643E^{-6}$

$P\text{value} < \alpha \therefore \text{reject } H_0$

Therefore, simple regression supports the claim that HS_ENG is a predictor of GPA.

(iii) HS_MATH is a predictor of GPA?

Based on the correlation coefficient 0.444, HS_MATH is a strong predictor of GPA.

When performing simple regression for HS_MATH onto GPA, the results are:

SUMMARY
OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.444387078
R Square	0.197479875
Adjusted R Square	0.193864919
Standard Error	1.195596284
Observations	224

<i>ANOVA</i>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	78.08884482	78.08884482	54.62857668	2.93473E-12
Residual	222	317.3380052	1.429450474		
Total	223	395.42685			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	1.598844	0.414330	3.858865	0.000149	0.782321	2.415368	0.782321	2.415368
HS_MAT	0.361104	0.048856	7.391114	2.93473E	0.264822	0.457386	0.264822	0.457386
H	485	566	712	-12	489	481	489	481

$H_0: \beta_1 = 0$ (no relationship)

$H_0: \beta_1 \neq 0$ (some relationship)

$\alpha = 5\% \rightarrow Pvalue < \alpha \rightarrow \text{reject } H_0$

$\rightarrow Pvalue > \alpha \rightarrow \text{accept } H_0$

Since checking for $\beta_1 \neq 0$, Pvalue is two tailed

$\therefore Pvalue = 2.9347E^{-12}$

$Pvalue < \alpha \therefore \text{reject } H_0$

Therefore, simple regression supports the claim that HS_MATH is a predictor of GPA.

(iv) ATAR is a predictor of GPA?

Based on the correlation coefficient 0.4238, ATAR is a strong predictor of GPA.

When performing simple regression for ATAR onto GPA, the results are:

SUMMARY
OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.423807694
R Square	0.179612961
Adjusted R Square	0.175917524
Standard Error	1.208832104
Observations	224

<i>ANOVA</i>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	71.02378743	71.02378743	48.6039826	3.53788E-11
Residual	222	324.4030626	1.461275057		
Total	223	395.42685			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	1.007992 343	0.522053 917	1.930820 382	0.05477 886	0.020823 171	2.036807 857	0.020823 171	2.036807 857
ATAR	0.045888 818	0.006582 198	6.971655 657	3.53788 E-11	0.032917 232	0.058860 405	0.032917 232	0.058860 405

$H_0: \beta_1 = 0$ (no relationship)

$H_0: \beta_1 \neq 0$ (some relationship)

$\alpha = 5\% \rightarrow Pvalue < \alpha \rightarrow reject H_0$

$\rightarrow Pvalue > \alpha \rightarrow accept H_0$

Since checking for $\beta_1 \neq 0$, Pvalue is two tailed

$\therefore Pvalue = 3.53788E^{-11}$

$Pvalue < \alpha \therefore reject H_0$

Therefore, simple regression supports the claim that ATAR is a predictor of GPA.

5.

Step 1: HS_SCI only

SUMMARY
OUTPUT

Regression Statistics								
Multiple R	0.344281789							
R Square	0.11852995							
Adjusted R Square	0.114559365							
Standard Error	1.253026831							
Observations	224							

ANOVA						Significance		
	df	SS	MS	F		F		
Regression	1	46.86992495	46.86992495	29.85200578		1.24741E-07		
Residual	222	348.556925	1.570076239					
Total	223	395.42685						

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	2.421813	0.408032	5.93533	1.1177	1.617699	3.225927	1.617699	3.225927
	394	996	713	3E-08	75	038	75	038
HS_SCI	0.269731	0.049367	5.46369	1.2474	0.172441	0.367021	0.172441	0.367021
	678	962	891	1E-07	871	484	871	484

Step 2: HS_SCI and HS_ENG

SUMMARY
OUTPUT

Regression Statistics								
Multiple R	0.367413355							
R Square	0.134992574							
Adjusted R Square	0.127164452							
Standard Error	1.244075847							
Observations	224							

ANOVA						Significance		
	df	SS	MS	F		F		
Regression	2	53.37968815	26.68984408	17.244568		1.09823E-07		
Residual	221	342.0471618	1.547724714					
Total	223	395.42685						

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	1.874651	0.485078	3.864635	0.000146	0.918680	2.830622	0.918680	2.830622
	543	473	619	249	101	984	101	984
HS_SCI	0.198276	0.060136	3.297075	0.001138	0.079760	0.316791	0.079760	0.316791
	036	944	388	29	778	295	778	295
HS_ENG	0.139019	0.067785	2.050858	0.041460	0.005429	0.272608	0.005429	0.272608
	237	868	687	405	812	662	812	662

Step 3: HS_SCI, HS_ENG and HS_MATH

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.464008313
R Square	0.215303715
Adjusted R Square	0.204603311
Standard Error	1.187606417
Observations	224

ANOVA					
	df	SS	MS	F	Significance F
Regression	3	85.13686975	28.37895658	20.12108301	1.46316E-11
Residual	220	310.2899802	1.410409001		
Total	223	395.42685			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.988217781	0.49932212	1.979118771	0.049049736	0.004150954	1.972284608	0.004150954	1.972284608
HS_SCI	0.066884563	0.063736315	1.049394886	0.295147826	-0.058727323	0.19249645	-0.058727323	0.19249645
HS_ENG	0.086024345	0.065665727	1.31003416	0.19155049	-0.043390036	0.215438727	-0.043390036	0.215438727
HS_MATH	0.285795199	0.060229116	4.745133559	3.7495E-06	0.16709532	0.404495077	0.16709532	0.404495077

Step 4: HS_SCI, HS_ENG, HS_MATH and PARENT EDUCSUMMARY
OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.483235669
R Square	0.233516712
Adjusted R Square	0.21593682
Standard Error	1.179115031
Observations	224

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	5	92.33877792	18.46775558	13.28317109	2.64537E-11
Residual	218	303.0880721	1.390312257		
Total	223	395.42685			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	1.645405529	0.58260841	2.824204903	0.005179169	0.49713935	2.793671709	0.49713935	2.793671709
HS_SCI	0.073707482	0.063370002	1.163128927	0.246048774	-0.051188809	0.198603774	-0.051188809	0.198603774
HS_ENG	0.053236241	0.066787964	0.797093334	0.42626411	-0.078396532	0.184869014	-0.078396532	0.184869014
HS_MATH	0.246371369	0.0623664	3.950386288	0.00010537	0.123453083	0.369289656	0.123453083	0.369289656
PARENT EDUC (1=P)	0.152716051	0.239075283	0.638778081	0.523637856	-0.318478764	0.623910865	-0.318478764	0.623910865
PARENT EDUC (1=S)	-0.362405148	0.187469737	-1.933139472	0.054514243	-0.731890302	0.007080006	-0.731890302	0.007080006

Step 5: HS_SCI, HS_ENG and HS_MATH, PARENT EDUC AND GENDER

SUMMARY
OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.48457333
R Square	0.234811312
Adjusted R Square	0.213654021
Standard Error	1.18083028
Observations	224

<i>ANOVA</i>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	6	92.85069735	15.47511623	11.0983638	8.66761E-11
Residual	217	302.5761526	1.39436015		
Total	223	395.42685			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	1.785415 287	0.627546 42	2.845072 857	0.004865 164	0.548548 716	3.022281 857	0.548548 716	3.022281 857
HS_SCI	0.083460 714	0.065471 756	1.274759 06	0.203757 801	-0.045 581256	0.212502 684	-0.045 581256	0.212502 684
HS_ENG	0.036872 164	0.072131 848	0.511177 301	0.609746 615	-0.105 296558	0.179040 885	-0.105 296558	0.179040 885
HS_MATH	0.244315 461	0.062549 221	3.905971 284	0.000125 351	0.121033 68	0.367597 243	0.121033 68	0.367597 243
PARENT EDUC (1=P)	0.162986 456	0.240022 316	0.679047 095	0.497831 628	-0.310 087039	0.636059 952	-0.310 087039	0.636059 952
PARENT EDUC (1=S)	-0.363 53318	0.187751 678	-1.936 244641	0.054136 555	-0.733 583531	0.006517 172	-0.733 583531	0.006517 172
GENDER (1=M)	-0.108 726003	0.179440 344	-0.605 917265	0.545202 949	-0.462 39508	0.244943 075	-0.462 39508	0.244943 075

Step 6: HS_SCI, HS_ENG and HS_MATH, PARENT EDUC, GENDER and ATAR

SUMMARY
OUTPUT

<i>Regression Statistics</i>								
Multiple R	0.4846456							
R Square	0.234881358							
Adjusted R Square	0.210085846							
Standard Error	1.183506354							
Observations	224							

<i>ANOVA</i>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	7	92.87839541	13.2683422	9.472736919	3.03317E-10			
Residual	216	302.5484546	1.40068729					
Total	223	395.42685						

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	1.805044 186	0.644271 494	2.801682 525	0.005545 411	0.535180 258	3.074908 115	0.535180 258	3.074908 115
HS_SCI	0.094815 062	0.104045 807	0.911281 908	0.363162 722	-0.110 260002	0.299890 125	-0.110 260002	0.299890 125
HS_ENG	0.047577 887	0.104988 341	0.453173 048	0.650879 037	-0.159 354919	0.254510 692	-0.159 354919	0.254510 692
HS_MATH	0.255711 15	0.102456 076	2.495812 457	0.013314 863	0.053769 459	0.457652 84	0.053769 459	0.457652 84
PARENT EDUC (1=P)	0.160789 746	0.241072 929	0.666975 536	0.505499 616	-0.314 366798	0.635946 29	-0.314 366798	0.635946 29
PARENT EDUC (1=S)	-0.363 417157	0.188178 981	-1.931 231397	0.054763 147	-0.734 319335	0.007485 021	-0.734 319335	0.007485 021
GENDER (1=M)	-0.106 870835	0.180330 221	-0.592 639627	0.554042 273	-0.462 303047	0.248561 377	-0.462 303047	0.248561 377
ATAR	-0.003 750572	0.026671 252	-0.140 622271	0.888299 403	-0.056 319808	0.048818 664	-0.056 319808	0.048818 664

6.

HS_SCI:

High school Science score has a regression coefficient of 0.083460714, which means that each increase in score from 1 to 10 will increase the expected GPA by about 0.08. Overall this means that by achieving the maximum of 10, you would increase your expected GPA by 0.72 compared to someone who achieved the absolute minimum science score of 1. The p-value of HS_SCI is 0.203757801, yet since we are only interested in if HS_SCI is increasing the expected GPA, the t-test would only be 1 tailed and therefore the p-value must be halved. This gives a new p-value of about 0.1019, which is almost double that of the usual significance level cut-off of 0.05. Overall, both the p-value and regression coefficient have shown that HS_SCI has only a very small impact on the expected GPA and is therefore not statistically significant.

HS_ENG:

High school English score has a regression coefficient of 0.036872164, which means that each increase in score from 1 to 10 will increase the expected GPA by about 0.037. Overall this means that by achieving the maximum of 10, you would increase your expected GPA by 0.333 compared to someone who achieved the absolute minimum science score of 1. The p-value of HS_ENG is 0.609746615, yet since we are only interested in if HS_ENG is increasing the expected GPA, the t-test would only be 1 tailed and therefore the p-value must be halved. This gives a new p-value of about 0.30487, which is over 6 times that of the usual 0.05. Overall, based on both the p-value and regression coefficient, HS_ENG has much less impact on the expected GPA than HS_SCI did and therefore is also not statistically significant.

HS_MATH:

High school Maths score has a regression coefficient of 0.244315461, which means that each increase in score from 1 to 10 will increase the expected GPA by about 0.244. Overall this means that by achieving the maximum of 10, you would increase your expected GPA by 2.196 compared to someone who achieved the absolute minimum science score of 1. The p-value of HS_MATH is 0.000125351, yet since we are only interested in if HS_MATH is increasing the expected GPA, the t-test would only be 1 tailed and therefore the p-value must be halved. This gives a new p-value of about 0.00006, which is considerably less than the usual significance level cut-off of 0.05. Overall, both the p-value and regression coefficient have shown HS_MATH to have a large impact on the expected GPA and therefore is statistically significant.

PARENT EDUC:

Since Undergraduate education was used as the reference category for Parent Education level, both Postgraduate education and Secondary education have coefficients and p-values that are comparing the difference between those education levels and the Undergraduate education level.

Postgraduate education level has a regression coefficient of 0.162986456, meaning that a student with a parents' education level of postgraduate would have an expected GPA increase of about 0.16 compared to a student with a parents' education level of only undergraduate. Conversely, a student with a parents' education level of Secondary education or lower has a regression coefficient of -0.36353318, meaning that student would have an expected GPA decrease of about 0.36 compared to undergraduate ones.

Postgraduate has a p-value of 0.497831628 and Secondary or lower has a p-value of 0.054136555, however we are only interested in an increase of GPA and therefore the t-test would be 1 tailed and therefore the p-values must be halved. This gives new p-values of about 0.2489 for P and 0.027 for S. Using the same significance level cut-off, the difference between P and U has a no statistical significance, whereas the difference between S and U has a p-value less than the cut-off and therefore is shown to have statistical significance. Overall, since Secondary > Undergraduate > Postgraduate is a constant increase in education, the PARENT EDUC variable has shown that there is significant increase in expected GPA from those with a parental education higher than Secondary school, compared to those of Secondary or lower education.

GENDER:

Gender has a regression coefficient of -0.108726003 when males were set as 1 and females were set as 0, meaning that males have an expected GPA of about 0.109 less than that of females. Gender has a p-value of 0.545202949, however we are only interested in an increase of GPA based on gender and therefore the t-test would be 1 tailed and therefore the p-values must be halved. This gives new p-values of about 0.2726 which is much larger than the cut-off of 0.05 and therefore the gender variable is shown to not be of statistical significance.

7.

ATAR having a negative coefficient at first seems surprising, however it can also be observed that the three grades' coefficients have increased in turn, meaning that these two will largely counteract each other. This is because, by the way ATAR is calculated, an increase in ATAR would come with an increase of all classes that the student is taking, meaning both an increase and decrease of expected GPA at the same time. After the inclusion of ATAR, the model fit, or R^2 , goes from 0.234811312 in step 5 to 0.234881358. This is an increase of model fit by 0.00007004 which is an improvement upon step 5's fit by a measly 0.0298%, which can safely be described as no real improvement of overall model fit. Overall, this just shows that the inclusion of ATAR is just used as an extremely small adjustment that would only occur if a student does better on their subjects that are counted towards ATAR that are not Science, English or Maths.

Task 3 (Summary Report)

Based on the data analysis, there is an observable correlation between a student's GPA and their high school grades, most prominently that of their high school maths score, which during stepwise regression at step 3, improved the model fit from 13% to 21%. The students Science score is the second most important out of the three grades and was the first variable to be used during the stepwise regression. However, as English and Maths were added, the Science coefficient and p-value drop drastically, dropping out of significance range once Maths was added in. This shows that, while Science scores seemed to predict GPA somewhat on its own, it is not a good predictor overall once compared to Maths scores.

There has also been shown to be statistical significance in a students' social economic status affecting their academic achievement through the observation of a students' parent's level of education being above secondary education level having a significant impact on the student's GPA.

Gender and ATAR, however, proved to have an insignificant impact on a student's GPA during stepwise regression and shows that neither should be included in the regression model as both causes almost no increase to model fit yet including 2 extra variables would increase model complexity considerably. Despite this, ATAR itself is still a good predictor for GPA as is shown in the correlation matrix, it is just that most of its prediction functionality has already been taken by HS_SCI, HS_ENG and HS_MATH.

Overall, the best regression model came at step 4, before adding in the insignificant variables of gender and ATAR. This is the point where all the significant variables, namely HS_MATH and parental secondary education, has been added and further variables do not aid the model fit. Although, there is room for other factors which could influence academic performance, such as country of citizenship, class attendance and employment status which could all have a significant impact on a student's GPA.