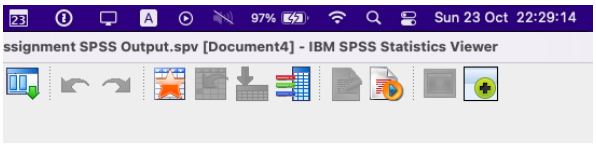
MBS659 Quantitative Research for Business Assignment - Ng Zhi Hui (34557319)

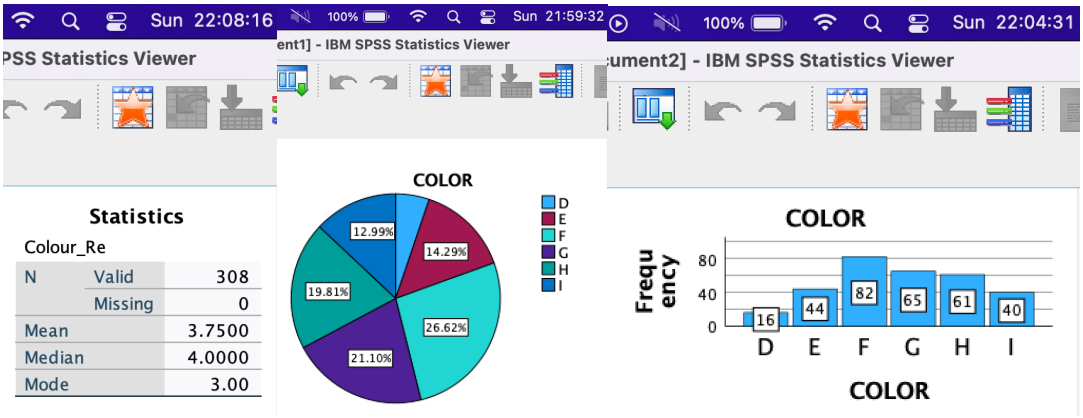
Question 1 (25 marks)



Statistics

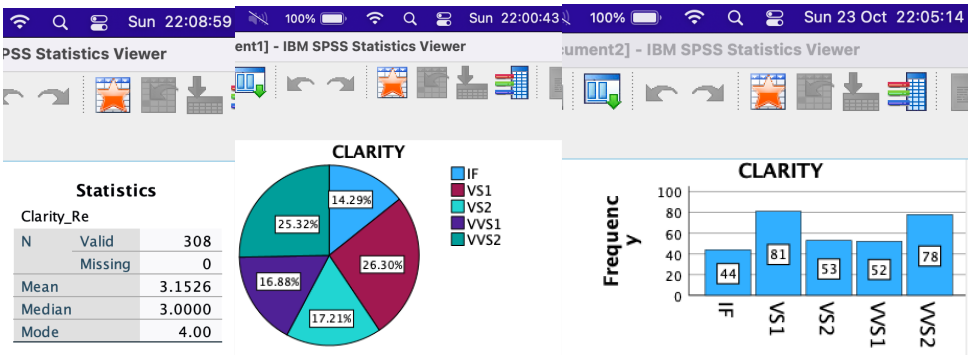
		WEIGHT	PRICE	Colour_Re	Clarity_Re	Rater_Re
N	Valid	308	308	308	308	308
	Missing	0	0	0	0	0
Mean		.6309	5019.48	3.7500	3.1526	1.7630
Median		.6200	4215.00	4.0000	3.0000	2.0000
Mode		1.00	5122	3.00	4.00	1.00
Std. Deviation		.27718	3403.116	1.40236	1.29372	.83040
Variance		.077	11581196.6	1.967	1.674	.690
Skewness		.015	.658	-.033	-.205	.467
Std. Error of Skewness		.139	.139	.139	.139	.139
Kurtosis		-1.241	-.325	-.858	-1.019	-1.395
Std. Error of Kurtosis		.277	.277	.277	.277	.277
Range		.92	15370	5.00	4.00	2.00
Minimum		.18	638	1.00	1.00	1.00
Maximum		1.10	16008	6.00	5.00	3.00

Colour (Recoded Values: 1 = D, 2 = E, 3 = F, 4 = G, 5 = H, 6 = I)



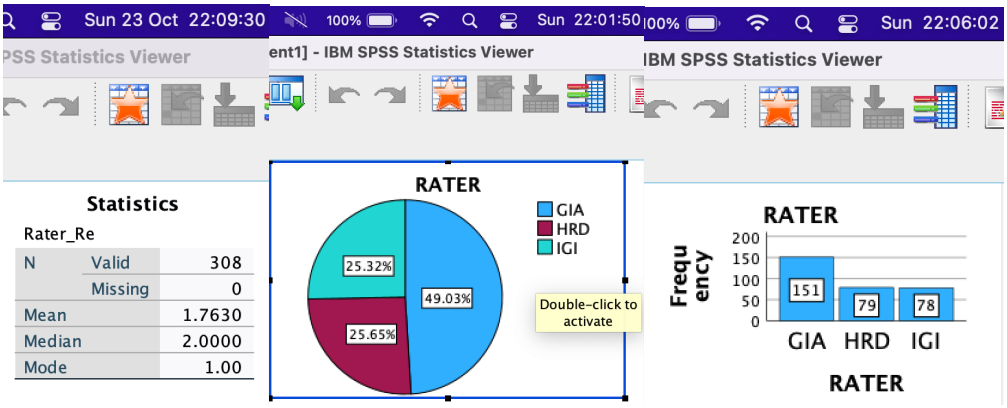
Colour is an ordinal data type as it is ranked by the degree of purity. The mode indicates that most are graded as 4(G) at 26.62%, while the median indicates that the mid diamond lies in the 3(F). The kurtosis of -0.857 is shown to be on the low end and since the skewness of -0.033 is within +/- 0.5, and can be said to be symmetrical.

Clarity (Recoded Values: 1 = IF, 2 = VVS1, 3 = VVS2, 4 = VS1, 5 = VS2)



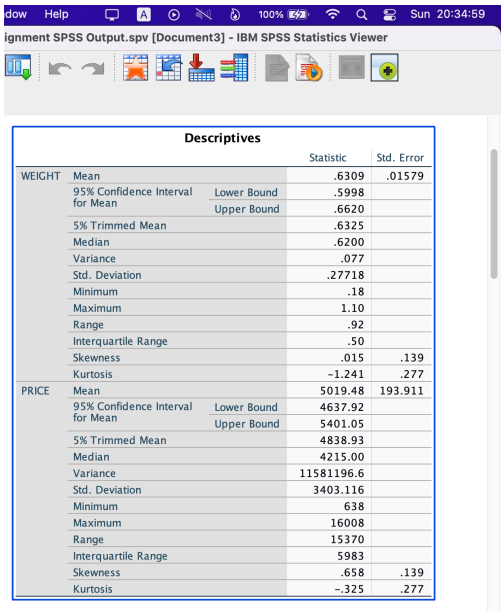
Clarity is an ordinal data type as it is ranked by the presence of flaw. The mode indicates that most are graded as 4(VS1) at 26.3%, while the median indicates that the mid diamond lies in the 3(VVS2). It possesses a kurtosis of -1.019 and skewness of -0.205 within +/- 0.5, it can be said to be symmetrical.

Rater (Recoded Values: 1 = GIA, 2 = HRD, 3 = IAI)



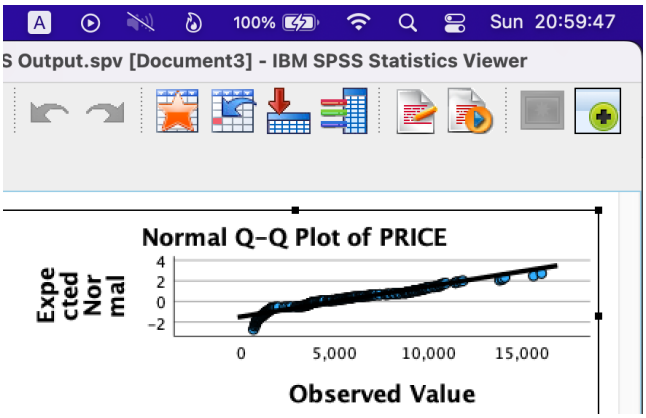
Rate is a nominal data type consisting of companies. The mode indicates that most of the grading was done by 1 (GIA) at 49.03%, while the median indicates that the mid diamond was graded by 2 (HRD). The kurtosis of -1.395 and since the skewness of 0.467 is within +/- 0.5, it is said to be symmetrical.

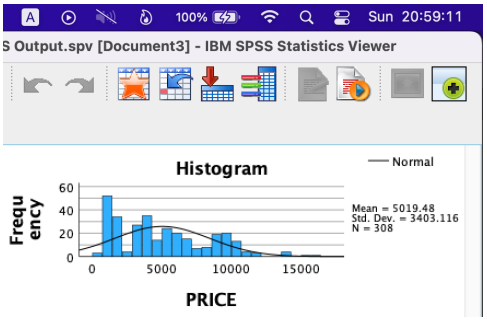
Price

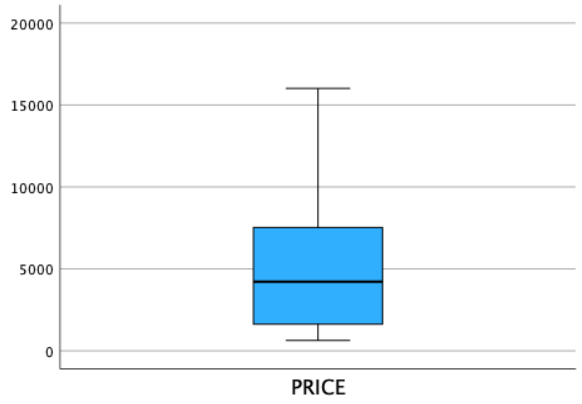


Descriptives

			Statistic	Std. Error
WEIGHT	Mean		.6309	.01579
	95% Confidence Interval for Mean	Lower Bound	.5998	
		Upper Bound	.6620	
	5% Trimmed Mean		.6325	
	Median		.6200	
	Variance		.077	
	Std. Deviation		.27718	
	Minimum		.18	
	Maximum		1.10	
	Range		.92	
	Interquartile Range		.50	
	Skewness		.015	.139
	Kurtosis		-1.241	.277
PRICE	Mean		5019.48	193.911
	95% Confidence Interval for Mean	Lower Bound	4637.92	
		Upper Bound	5401.05	
	5% Trimmed Mean		4838.93	
	Median		4215.00	
	Variance		11581196.6	
	Std. Deviation		3403.116	
	Minimum		638	
	Maximum		16008	
	Range		15370	
	Interquartile Range		5983	
	Skewness		.658	.139
	Kurtosis		-.325	.277

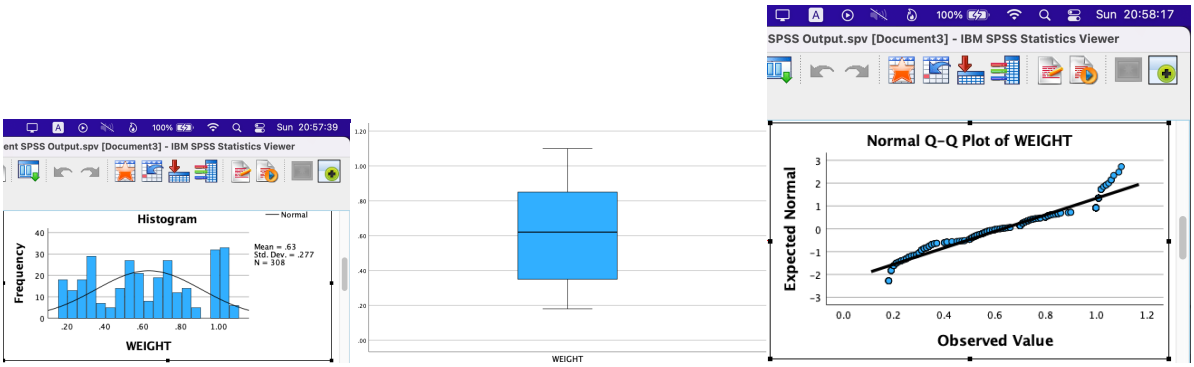






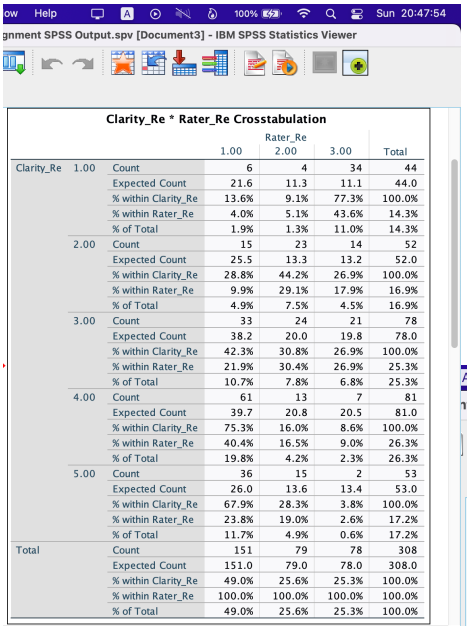
Descriptive statistics indicate that the scores of Price have a mean of 5019.48 with 193.911 standard error from the mean of the population. A confidence interval of 95% of the mean makes it located between the upper bound of 5401.05 and the lower bound of 4637.92. The trimmed mean value is lower than the mean value explaining the influence of the extreme values. With a median value of 4215 which is 804.48 less than the mean, the distribution of the scores is moderately positive skewed with a value of .658 as shown in the figure above. The standard deviation value is 3403.116 and the minimum and maximum scores of 638 and 16008 respectively. The results of Kolmogorov-Smirnov and Shapiro-Wilk tests are significant with a value of <.001. The histogram shows that the data is moderately positively skewed. Q-Q Plot indicates that the data tend to be deviate on one end, the difference between the mean and the median (804.48) is large when considering the minimum and the maximum scores. The skewness value is not small as well. Therefore the data is not normally distributed. In this regard, the median of 4215 explains the center of the data which 50% is spread within an interquartile range of 5983 as shown in the boxplot.

Weight



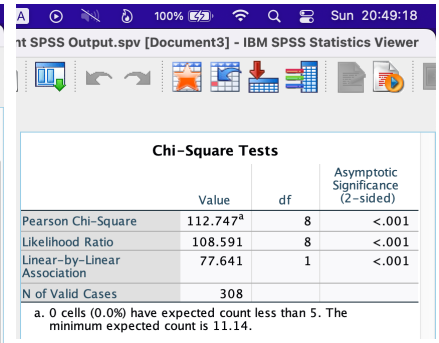
Descriptive statistics indicates that the scores of Weight have a mean of .6309 with .01579 standard error from the mean of the population. A confidence interval of 95% of the mean makes it located between the upper bound of .6620 and the lower bound of .5998. The trimmed mean value is slightly higher than the the mean value explaining the influence of the extreme values with a value of .6325. The standard deviation value is .27718 and the minimum and maximum scores of .18 and 1.10 respectively. The results of Kolmogrov-Samirnov and Shapiro-Wilk tests are significant; the histogram shows that the data is symmetrical and has a bell-shape. Q-Q Plot indicates that the data tend to be away from the line from the two ends, the difference between the mean and the median (0.0109) is large when considering the minimum and the maximum scores; the skewness values is relatively small. Therefore, the data is normally distributed.

Question 2.1:



Clarity_Re * Rater_Re Crosstabulation

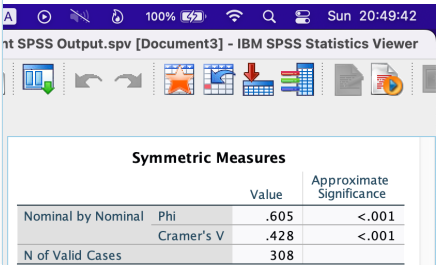
			Rater_Re 1.00	2.00	3.00	Total
Clarity_Re	1.00	Count	6	4	34	44
		Expected Count	21.6	11.3	11.1	44.0
		% within Clarity_Re	13.6%	9.1%	77.3%	100.0%
		% within Rater_Re	4.0%	5.1%	43.6%	14.3%
		% of Total	1.9%	1.3%	11.0%	14.3%
	2.00	Count	15	23	14	52
		Expected Count	25.5	13.3	13.2	52.0
		% within Clarity_Re	28.8%	44.2%	26.9%	100.0%
		% within Rater_Re	9.9%	29.1%	17.9%	16.9%
		% of Total	4.9%	7.5%	4.5%	16.9%
	3.00	Count	33	24	21	78
		Expected Count	38.2	20.0	19.8	78.0
		% within Clarity_Re	42.3%	30.8%	26.9%	100.0%
		% within Rater_Re	21.9%	30.4%	26.9%	25.3%
		% of Total	10.7%	7.8%	6.8%	25.3%
	4.00	Count	61	13	7	81
		Expected Count	39.7	20.8	20.5	81.0
		% within Clarity_Re	75.3%	16.0%	8.6%	100.0%
		% within Rater_Re	40.4%	16.5%	9.0%	26.3%
		% of Total	19.8%	4.2%	2.3%	26.3%
	5.00	Count	36	15	2	53
		Expected Count	26.0	13.6	13.4	53.0
		% within Clarity_Re	67.9%	28.3%	3.8%	100.0%
		% within Rater_Re	23.8%	19.0%	2.6%	17.2%
		% of Total	11.7%	4.9%	0.6%	17.2%
Total		Count	151	79	78	308
		Expected Count	151.0	79.0	78.0	308.0
		% within Clarity_Re	49.0%	25.6%	25.3%	100.0%
		% within Rater_Re	100.0%	100.0%	100.0%	100.0%
		% of Total	49.0%	25.6%	25.3%	100.0%



Chi-Square Tests

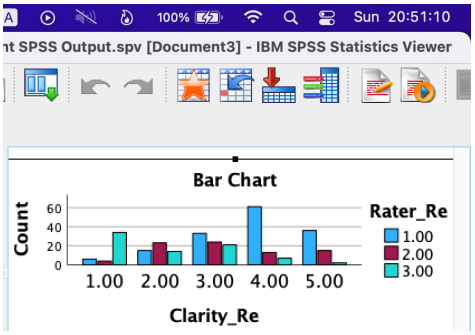
	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	112.747[a]	8	<.001
Likelihood Ratio	108.591	8	<.001
Linear-by-Linear Association	77.641	1	<.001
N of Valid Cases	308		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 11.14.



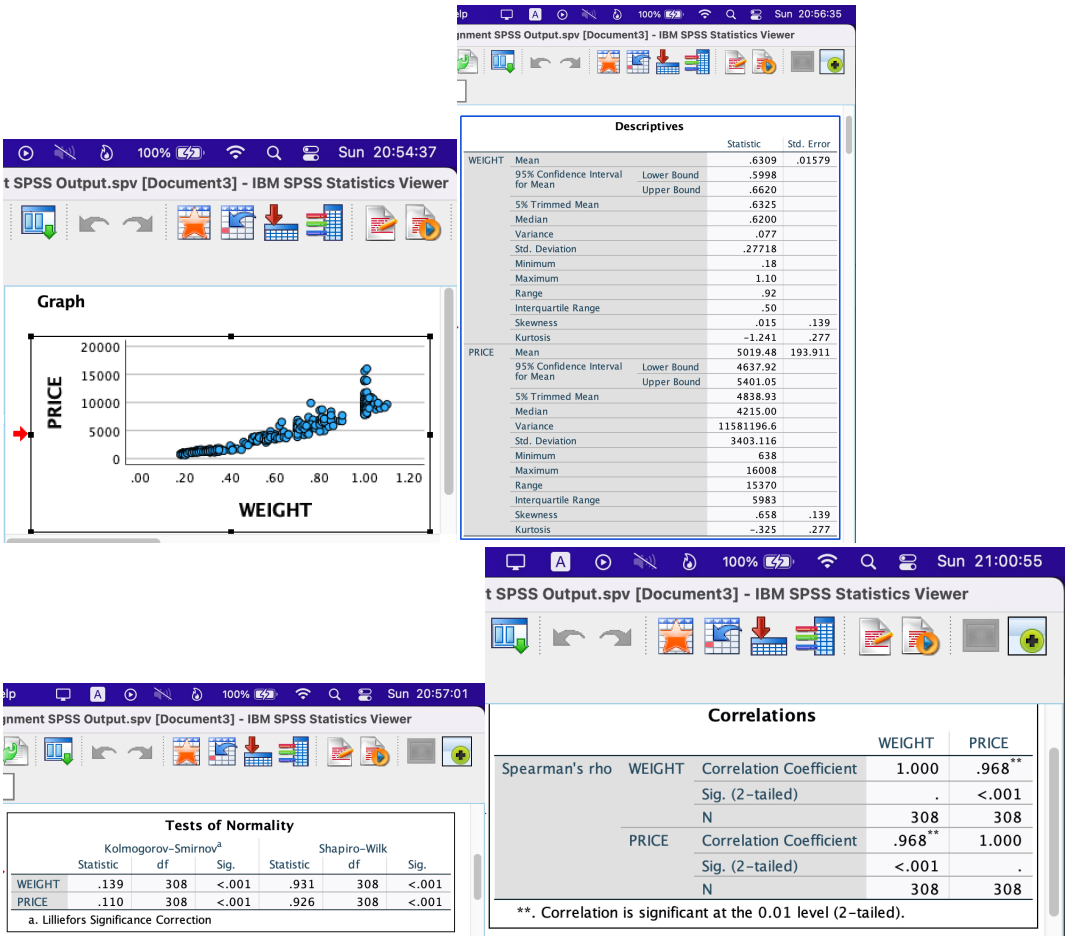
Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.605	<.001
	Cramer's V	.428	<.001
N of Valid Cases		308	



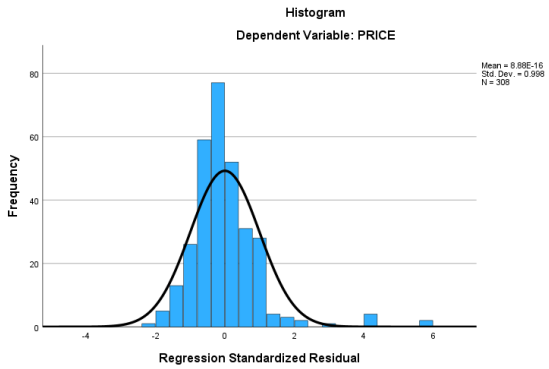
Cramer's V test indicates that the contingency between the raters and clarity was significant with a value of .428, and an approximate significance of <.001 which could be described as medium. This can be observed from the table and bar chart pattern that GIA (1) is more likely to rate VS1 (4) which relates to the overall statistics with GIA.

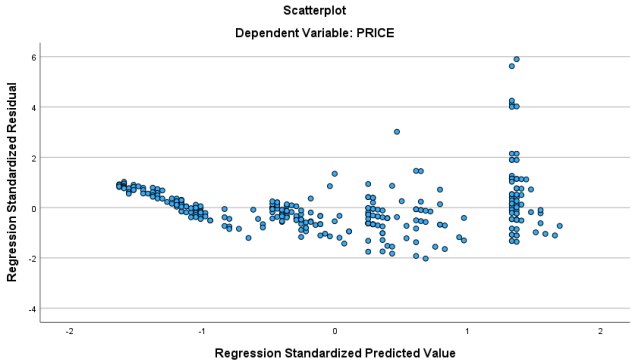
Question 2.2



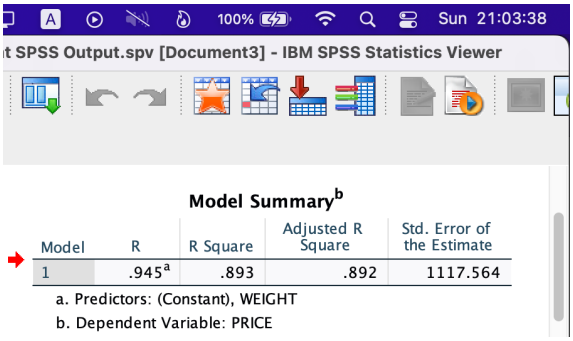
The relationship between Total diamond weight and Total diamond price was investigated using Spearman's rho correlation after the assumptions of linearity and normality had been violated. There was a small, positive correlation between the two variables [r=.968, n=308, p<.001], with high levels of diamond weight associated with highest levels of diamond price.

Question 3.1:

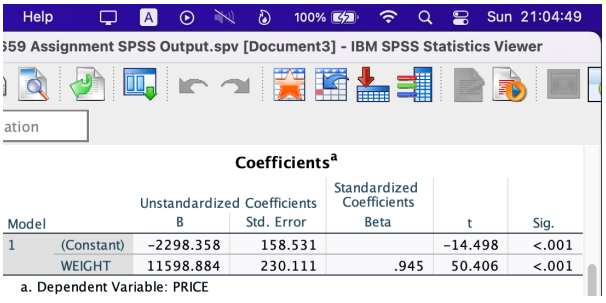




The Normality and Linearity assumption was passed successfully, but there was an assumption violation for homoscedasticity which has been breached.

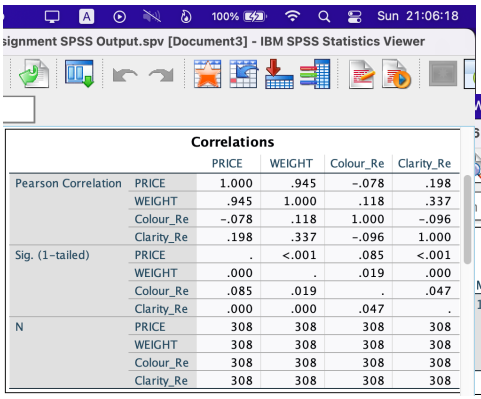


Adjusted R Square = .892 shown is very high, this suggests that 89.2% of the variance in Diamond Price is explained by the variance of Diamond Weight.



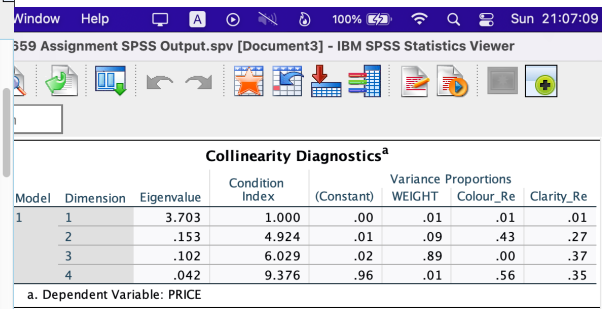
b_0 = -2298.358, b_1 = 11598.884, $p < .001$, indicates that the result was significant. The t value for B was significant and implies that Diamond Weight is a significant predictor. Additionally, b_1 means for every one unit rise (1 carat increase in Diamond Weight) in B, Diamond Price rises by 11598.884.

Question 3.2



Correlations

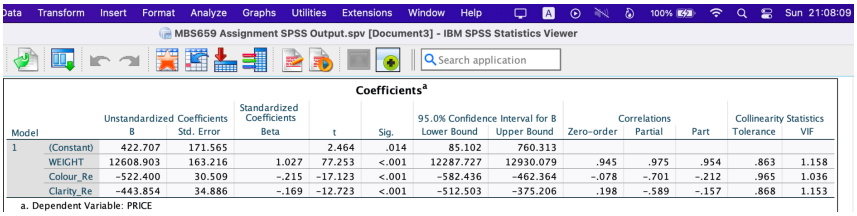
		PRICE	WEIGHT	Colour_Re	Clarity_Re
Pearson Correlation	PRICE	1.000	.945	-.078	.198
	WEIGHT	.945	1.000	.118	.337
	Colour_Re	-.078	.118	1.000	-.096
	Clarity_Re	.198	.337	-.096	1.000
Sig. (1-tailed)	PRICE	.	<.001	.085	<.001
	WEIGHT	.000	.	.019	.000
	Colour_Re	.085	.019	.	.047
	Clarity_Re	.000	.000	.047	.
N	PRICE	308	308	308	308
	WEIGHT	308	308	308	308
	Colour_Re	308	308	308	308
	Clarity_Re	308	308	308	308



Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions (Constant)	WEIGHT	Colour_Re	Clarity_Re
1	1	3.703	1.000	.00	.01	.01	.01
	2	.153	4.924	.01	.09	.43	.27
	3	.102	6.029	.02	.89	.00	.37
	4	.042	9.376	.96	.01	.56	.35

a. Dependent Variable: PRICE



Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound	Correlations Zero-order	Correlations Partial	Correlations Part	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	422.707	171.565		2.464	.014	85.102	760.313					
	WEIGHT	12608.903	163.216	1.027	77.253	<.001	12287.727	12930.079	.945	.975	.954	.863	1.158
	Colour_Re	-522.400	30.509	-.215	-17.123	<.001	-582.436	-462.364	-.078	-.701	-.212	.965	1.036
	Clarity_Re	-443.854	34.886	-.169	-12.723	<.001	-512.503	-375.206	.198	-.589	-.157	.868	1.153

a. Dependent Variable: PRICE

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.977[a]	.954	.953	736.596

a. Predictors: (Constant), Clarity_Re, Colour_Re, WEIGHT

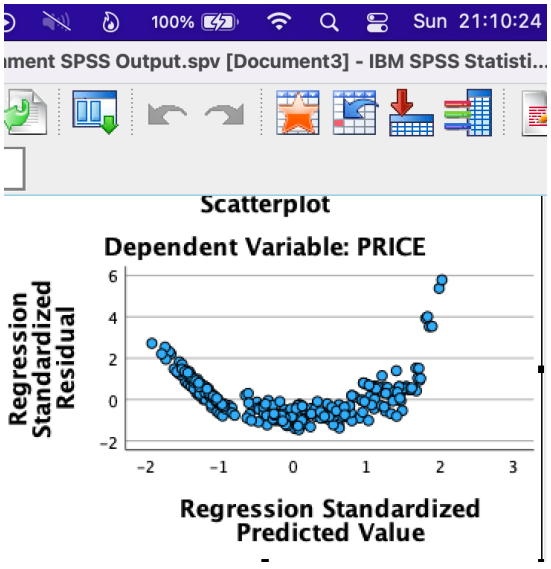
b. Dependent Variable: PRICE

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3390484772.0	3	1130161590.7	2082.962	<.001[b]
	Residual	164942574.89	304	542574.260		
	Total	3555427346.9	307			

a. Dependent Variable: PRICE

b. Predictors: (Constant), Clarity_Re, Colour_Re, WEIGHT



Multiple regression analysis was used to test if the three measures of control (Weight, Color, Clarity) significantly predicted the diamond's price. The results of the regression indicated the three predictors explained 95.4% of the variance ($R^2 = .954$, $F(3, 304) = 2082.962$, $p < .05$). It was found that Weight significantly predicted diamond price ($\beta = 1.027$, $p < .05$), as did both Color ($\beta = -.215$, $p < .05$), and Clarity ($\beta = -.169$, $p < .05$).