

1. For this question, use the Franchises data file. The file has data on several variables explained below. We would like to understand what predicts financial growth in 2011. Run the required analysis and articulate the variables that are important. Run the required analysis and articulate the findings.

In order to understand what predicts financial growth in 2011, we've run a multi-regression model, using `oyrsact`, `distancetoHQ`, `officesincounty`, and `ownmgrexp`. Looking at our Analysis of Variance table (Appendix A.1), we can conclude that the overall model is significant with the F value of 21.83 and the p-value (<0.0001) for the F statistic being less than $\alpha(0.05)$. The adjusted R-squared is 0.0995, indicating the model explains 9.95% of the variance in Y. Therefore, the model might not be very effective in predicting the variation with relatively low adj R-squared.

We need to examine the influence of each of the variables by checking the Parameter Estimates table (Appendix A.2). We can observe that the intercept is significant. In this case, we can logically interpret the intercept as, when other variables are held to zero the financial growth in 2011 is 0.30481. The coefficient for `oyrsact`(years in the network) is -0.00920, indicating financial growth decreases as the number of years in the network increases. This variable is significant with a t-value of -6.92 and p value (<0.0001). `DistancetoHQ` also indicates a decrease in financial growth as distance from headquarters increases. With a t-value of -2.48 and a p-value of 0.0133, this variable is also significant. The number of offices in the county is not significant with a low-t and a high p-value(0.5895). However, we can conclude that experience in leadership also has a negative impact on financial growth with a coefficient of -0.00331. This is also significant with a t-value of -2.72 and a p-value of 0.0066. Note that none of the variance inflation scores are high enough to consider any issues between variables.

In order to assess the normality and equal variance assumption of the regression model, we can check the distribution of residuals histogram and residuals by predicted graph (Appendix A.3). The errors follow a normal distribution and there is a slight megaphone effect, we can conclude that assumption of normality is met by the model and equal variance might need to be checked further.

Finally, by checking the Cook's D graph and Outlier and Leverage Diagnostics chart (Appendix A.4), we can observe that there are multiple outliers and leverage points. There are also three observations that may qualify under the influential observations label. We may have to go into the data and analyze why some of the outliers and influential variables behave the way they do. If these observations are of concern, then the regression analysis needs to be redone after dropping these observations.

2. Using the world happiness dataset on sas named "happy" and explore if the following factors (Region, Freedom, Ladder, Social Support, Healthy Life Expectancy and GDP per capita) impact happiness. What can you conclude about the relative importance of these factors in influencing happiness? Are the results surprising? Why/Why not?

In search of determining whether given factors impact happiness, we've run a multi-regression model, using the categorical value Region, and continuous values freedom, ladder, social support, Healthy Life Expectancy, and GDP per capita. Looking at our Analysis of Variance table (Appendix B.1), we can conclude that the overall model is significant with the F value of 16.06 and the p-value ($<.0001$) for the F statistic being less than $\alpha(.05)$. The adjusted R-squared is 0.5763, indicating the model explains 57.63% of the variance in Y and overall is a good one.

We need to examine the influence of each of the variables by checking the Parameter Estimates table (Appendix B.2). We can observe that the intercept is significant. In this particular case, it doesn't make logical sense to interpret the intercept. By looking at each of the variables and checking their p-values associated with their T-values, we can conclude that Ladder has a significant impact on happiness with a T-value of 2.48 and a p-value of 0.0145. Freedom also has a substantial positive impact on happiness with a T-value of 4.29 and a p-value of less than 0.0001. We need to interpret regions compared to the comparison group (region 9). We can observe that regions 2, 3, 4, and 6 have a significant impact on happiness (region 2 having a negative and others having a positive impact). Other regions don't yield significant impact with p-values greater than nominal alpha.

We can also observe that variance inflation scores for Ladder and Halthy_Life_Expectancy are fairly high at 5.23 and 6.64. These scores suggest that these variables are correlated with each other, therefore, it would be a better idea to re-run the model after dropping one of the two highly related independent variables. Note that GDP_Per_Capita might also be considered high with 3.94 depending on the threshold limit.

In order to assess the normality and equal variance assumption of the regression model, we can check the distribution of residuals histogram, QQ plot, and residuals by predicted graph (Appendix B.3). The errors follow almost a normal distribution and there is not a significant megaphone effect, we can conclude that both assumptions of normality and equal variance are met by the model.

However, by checking the Cook's D graph and Outlier and Leverage Diagnostics chart (Appendix B.4), we can observe that there are multiple outliers and leverage points. There is also one observation that may qualify under the influential observations label. We may have to go into the data and analyze why some of the outliers and influential variables behave the way they do. If these observations are of concern, then the regression analysis needs to be redone after dropping these observations, and after dropping one of the highly correlated variables.

3. Using the HBAT_200 dataset posted on blackboard, explore whether the following variables (X3 Firm Size, X6 Product Quality, X8 Tech Support, X9 Complaint Resolution, X17 Price Flexibility, and X18 Delivery Speed) impact X23(Purchase Level).

In order to understand whether the listed variables impact Purchase Level, we've run a multi-regression model, using the categorical value X3 Firm Size, and continuous values X6 Product Quality, X8 Tech Support, X9 Complaint Resolution, X17 Price Flexibility, and X18 Delivery Speed. Looking at our Analysis of Variance table (Appendix C.1), we can conclude that the overall model is significant with the F value of 104.40 and the p-value ($<.0001$) for the F statistic being less than $\alpha(.05)$. The adjusted R-squared is 0.7571, indicating the model explains 75.71% of the variance in Y. Therefore, the model can effectively predict the variation with relatively high adj R-squared.

We need to examine the influence of each of the variables by checking the Parameter Estimates table (Appendix C.2). We can observe that the intercept is not significant. In this case, it doesn't make logical sense to interpret the intercept. The coefficient for X6(product quality) is 3.27481, indicating product quality has a positive impact on purchase level. This variable is significant with a t-value of 11.44 and p value (<0.0001). X18(Delivery Speed) also shows a positive impact on purchase level. With a t-value of -8.24 and a p-value of <0.0001 , this variable is also significant. By comparing Firm_Size_0 to the comparison group (Firm_size_1). We can conclude that firm size has an impact on purchase level. Other variables do not have an impact on purchase level with low t-values and associated p-values greater than nominal α . We can also observe that variance inflation scores for X9(Complaint Resolution) and X18(Delivery Speed) can be considered high at 5.23 and 6.64.

In order to assess the normality and equal variance assumption of the regression model, we can check the distribution of residuals histogram and residuals by predicted graph (Appendix C.3). The errors follow a normal distribution and there is not a significant megaphone effect, we can conclude that both assumptions of normality and equal variance are met by the model.

By checking the Cook's D graph and Outlier and Leverage Diagnostics chart (Appendix C.4), we can observe there are 6 outliers and 3 leverage points. There is also one observation that may qualify under the influential observations label. We may have to go into the data and analyze why some of the outliers and influential variables behave the way they do. If these observations are of concern, then the regression analysis needs to be redone after dropping these observations and one of the highly correlated variables.

APPENDIX A

A.1

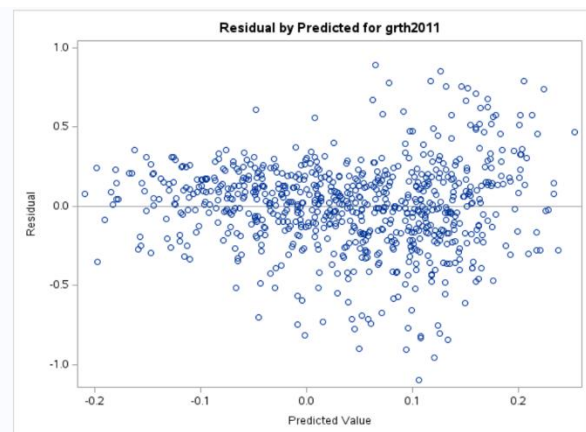
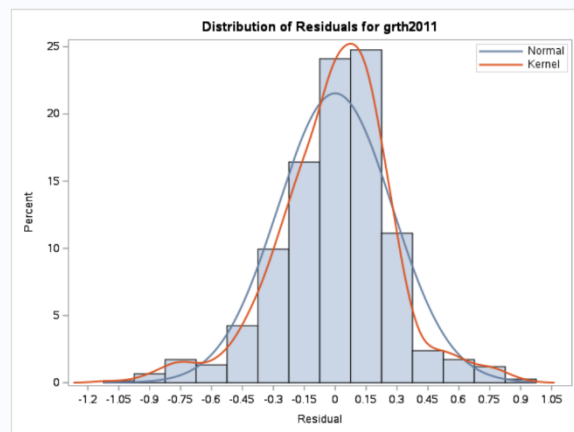
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	6.77816	1.69454	21.83	<.0001
Error	750	58.21569	0.07762		
Corrected Total	754	64.99385			

Root MSE	0.27861	R-Square	0.1043
Dependent Mean	0.04075	Adj R-Sq	0.0995
Coeff Var	683.73367		

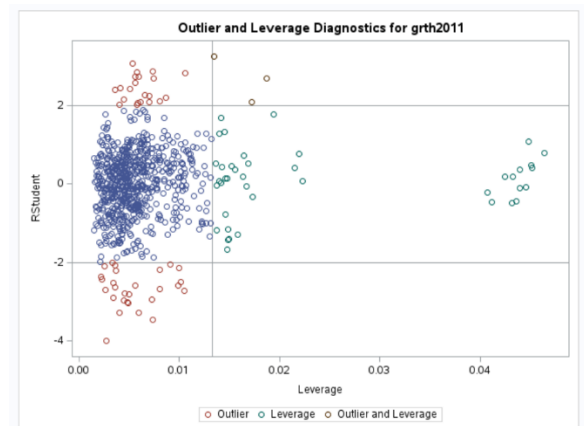
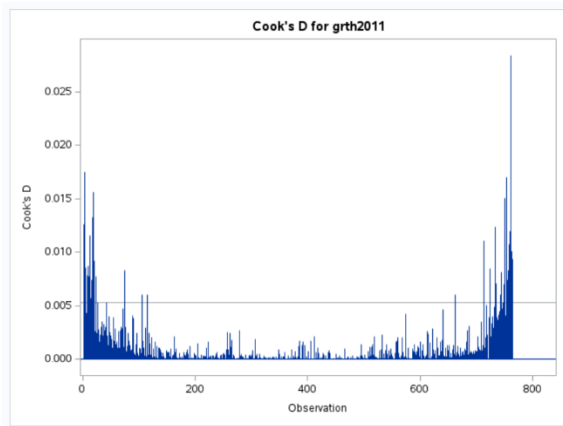
A.2

Parameter Estimates								
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate	Variance Inflation
Intercept	Intercept	1	0.30481	0.03810	8.00	<.0001	0	0
oyrsact	oyrsact	1	-0.00920	0.00133	-6.92	<.0001	-0.26408	1.22061
distancetoHQ	distancetoHQ	1	-0.00003647	0.00001469	-2.48	0.0133	-0.08672	1.02202
officesincounty	officesincounty	1	-0.00045199	0.00083727	-0.54	0.5895	-0.01869	1.00367
ownmgrexp	ownmgrexp	1	-0.00331	0.00122	-2.72	0.0066	-0.10312	1.20208

A.3



A.4



APPENDIX B

B.1

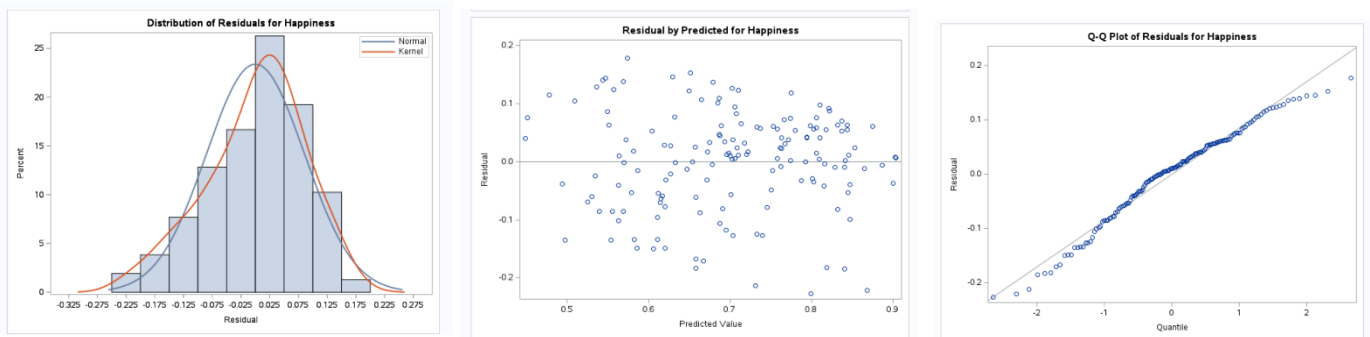
Least Squares Model (No Selection)					
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	14	1.79820	0.12844	16.06	<.0001
Error	141	1.12793	0.00800		
Corrected Total	155	2.92613			

Root MSE	0.08944
Dependent Mean	0.70066
R-Square	0.6145
Adj R-Sq	0.5763
AIC	-580.99715
AICC	-577.08348
SBC	-693.24931

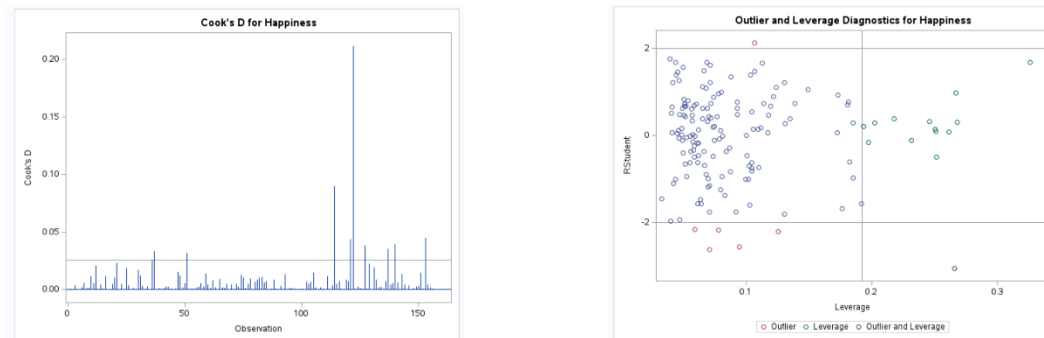
B.2

Model: MODEL1 Dependent Variable: Happiness Happiness								
Parameter Estimates								
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate	Variance Inflation
Intercept	Intercept	B	0.27659	0.09228	3.00	0.0032	0	0
Ladder	Ladder	1	0.03692	0.01491	2.48	0.0145	0.29606	5.22921
Social_Support	Social_Support	1	0.17534	0.09322	1.88	0.0621	0.15502	2.48477
GDP_Per_Capita	GDP_Per_Capita	1	-9.47958E-7	0.00000101	-0.93	0.3519	-0.09693	3.93994
Healthy_Life_Expecta	Healthy_Life_Expecta	1	-0.00270	0.00165	-1.64	0.1039	-0.22060	6.64366
Freedom	Freedom	1	0.31529	0.07344	4.29	<.0001	0.32125	2.04839
Region 0	Region 0	B	0.06596	0.05080	1.30	0.1963	0.16438	5.86253
Region 1	Region 1	B	-0.04628	0.04392	-1.05	0.2937	-0.10530	3.65180
Region 2	Region 2	B	-0.08672	0.03934	-2.20	0.0291	-0.16872	2.14346
Region 3	Region 3	B	0.10151	0.04172	2.43	0.0162	0.17281	1.84547
Region 4	Region 4	B	0.12163	0.04377	2.78	0.0062	0.17079	1.38175
Region 5	Region 5	B	0.08899	0.05328	1.67	0.0971	0.12495	2.04699
Region 6	Region 6	B	0.08837	0.04039	2.19	0.0303	0.22458	3.85447
Region 7	Region 7	B	0.11948	0.06521	1.83	0.0690	0.13789	2.07193
Region 8	Region 8	B	-0.00286	0.03854	-0.07	0.9410	-0.00697	3.23798
Region 9	Region 9	0	0

B.3



B.4



APPENDIX C

C.1

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	12230	2038.30449	104.40	<.0001
Error	193	3768.17307	19.52421		
Corrected Total	199	15998			

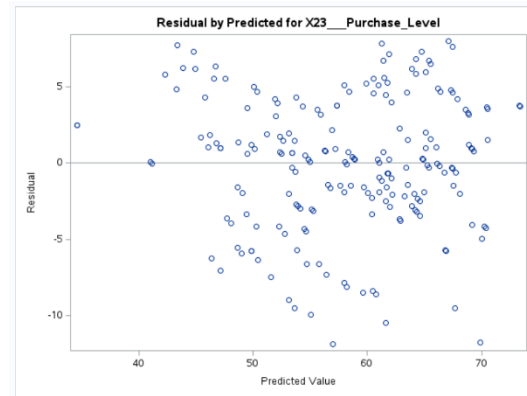
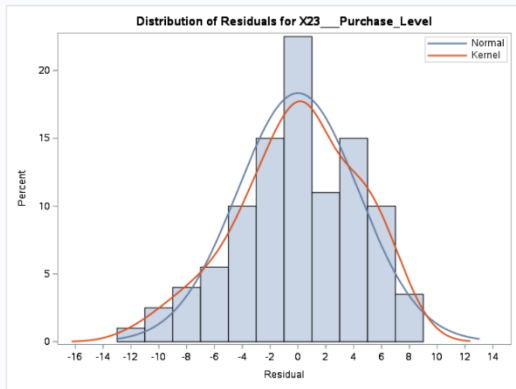
Root MSE	4.41862
Dependent Mean	58.20000
R-Square	0.7645
Adj R-Sq	0.7571
AIC	803.20564
AICC	803.95957
SBC	624.29386

C.2

Model: MODEL1
Dependent Variable: X23__Purchase_Level X23 - Purchase Level

Parameter Estimates								
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate	Variance Inflation
Intercept	Intercept	B	4.93659	3.16787	1.56	0.1208	0	0
X6__Product_Quality	X6__Product_Quality	1	3.27481	0.28619	11.44	<.0001	0.50513	1.59679
X8__Technical_Suppo	X8__Technical_Suppo	1	-0.20246	0.19276	-1.05	0.2949	-0.03738	1.03756
X9__Complaint_Resol	X9__Complaint_Resol	1	0.25044	0.54441	0.46	0.6460	0.03380	4.42251
X17__Price_Flexibil	X17__Price_Flexibil	1	-0.44252	0.40148	-1.10	0.2717	-0.05886	2.33682
X18__Delivery_Speed	X18__Delivery_Speed	1	7.85068	0.95242	8.24	<.0001	0.65614	5.19192
X3__Firm_Size 0	X3__Firm_Size 0	B	-1.74186	0.67128	-2.59	0.0102	-0.09736	1.15352
X3__Firm_Size 1	X3__Firm_Size 1	0	0

C.3



C.4

