Z5160339

UNSW

2/05/2022

Introduction

This research seeks to employ a multiple linear regression model to express the influence of profits, time, employment, level of wealth, and inventory on happiness, which directly relates to student marks. This investigation report aims to improve the Playconomics economy by employing multiple linear regression methodology on data. In a case where the researcher uses a linear regression model, it cannot yield out of bound predictions due to unbiased estimations. The model has to have at most two predictors and assume the form $Y= B_0+B_1X_1+ B_kX_{ki}+$ small error increment (e_i) for the results to be linear. When the value of p is equivalent to one, the equation above becomes a simple linear regression. In the case where $p> 1$, the model changes to become multiple due to many predictor variables. Linear regression analysis includes covariates to increase precision and statistical power (Gomila, 2021). The dataset for this investigation lacks such requirements, so the researcher excludes this method in the analysis.

This feature of discrete covariates enables the linear regression models to produce unbiased approximations of causal effects. Using many independent parameters makes the method a multiple regression approach that is also linear. The new model is assumed to be impartial and provide a better prediction than the linear technique. According to Gomila (2021), an excellent multiple linear regression whose aim is to estimate causal effects should take few values in the treatment of binary outcomes. Simple linear regression is not ideally suited to perform the estimation because of its unsaturation nature. It is standard for saturated models such as multiple linear regression estimates to be consistent, unbiased, and always within the bound. Student happiness has a strong association with their score marks. The model aims to estimate the relationships between economic parameters within the game using multiple linear regression and exclude various coefficients, which cause negative implications for dependent variables to a good equation.

Hypothesis

Null Hypothesis H_0= There is no association between students' marks and happiness

Alternative Hypothesis H_1= There is a strong association between student's marks and happiness

Body

The data, sampling, and sample size

This study employs multiple linear regression techniques in a discrete-continuous dataset. The sample size of 226 participants was obtained through the random sampling method on a population of students in 20 regions. Regions are the only categorical data. The dependent variable was happiness, while the Predictor variables were profits, time, wealthlow, wealthmedium, wealthhigh, employment, and inventory. These, together with dependent, are continuous variables. Time was used to measure the level of happiness in each region and at a point in time. The wealth level indicates the percentages of the poor, middle class, and wealthy students. Employment measures the number of jobs in a particular economy, with profits determining the variance in the producer surplus over time. Lastly, the inventory parameter tracks the number of units kept by firms in terms of stock. In table 1, there are no zero values in happiness, 14= wealthlow, 64=wealthmedium, 38=wealthhigh, 13=employment, 24=profit, and 48=inventory.

Methods

This research used multiple linear regression methods to estimate the relationship. Multiple linear regression has four assumptions that a statistician should not violate. Violation of such beliefs often leads to problems resulting in inconsistency, bias, and inefficiency in the least square estimation. According to Ernst & Albers (2017), such a dangerous move will lead to less accuracy in parameter estimations than other methods. Also, confidence intervals and P-values may become smaller even when the estimator is consistent, efficient, or unbiased (Ernst & Albers, 2017). These assumptions include:

1. Linearity
2. Homoscedasticity
3. Normality
4. Independence

In this study, the statistician will use multiple linear regression with few values and saturated. This method has few limitations other than the assumptions talked about in the previous literature.

Data analysis and results

Data analysis was conducted using SPSS.. First, descriptive statistics and the case summary were found. Then, data visualization such as bar graphs, scatterplots, histograms, boxes, and whisker plots occurred. This visual representation of information represents data found while conducting descriptive statistics. Other tests such as ANOVA, correlation, linear regression, and chi-square were undertaken to determine the relationship between dependent and predictor variables. Finally, there are the parametric tests.

Multiple linear regression was employed to test if wealthlow, wealthhigh, employment, profit, inventory, and wealthmedium significantly predicted happiness. Also, the aim was to investigate if there was any relationship between student marks and happiness. The fitted regression model was Happiness= 74851.832+0.121WealthLow-2274.908WealthMedium-1437.915WealthHigh-2284.235Eemployment+2.508Profit+0.669Inventory+10604.879+0.103+361.932+377.042+434.260+0.097+0.207. The overall regression was statistically significant with R^2=0.804, F=150.032, df1=6, df2=219, and sig=0.00. it was discovered that wealthhigh, employment, profit, inventory, and wealthmedium significantly predicted happiness with beta values -0.387, -0.131,-0.198,0.884, 0.186 at a p-value<0.05. It was found that wealthlow did not significantly predict happiness B=0.064, p-value=0.241>0.05.

The data had N=226 cases, implying that the research has enough statistical power to spot small or weak impacts. The researcher performed a one-tailed Pearson correlation. There was a -0.216 negative correlation between happiness and wealthlow with a significant value of <0.05 standard alpha. This outcome was a minimal negative relationship. Wealthmedium,

wealthhigh, employment, and inventory have -0.01, -0.205,-0.239, and -0.162 correlations. These were weak negative correlations with only the wealthMedium variable having a p-value of 0.441>0.05 hence not significant. Profit had a strong positive correlation of 0.872 with the response variable and p-value=0.000<0.05. The histograms for individual variables in figure 1 indicate visual aspects of the descriptive statistics and test the normality of data. Each figure displayed the mean, standard deviation, and the number of cases under investigation, as shown in Table 2. Except for happiness, the rest of the histograms shows that the predictor variable data are not normally distributed. The scatter plots in figure 2 show the visual distribution of the relationships between two data sets. The graphical actions indicate or confirm the results of the correlation analysis. Profit was the only variable with a strong positive correlation based on the graph spread of points, and as such, profit levels represented the most significant factor in determining students marks. The rest of the parameters had a weak negative correlation with happiness.

Figure 4 indicates the interquartile range for each dependent and independent variable. Each box is divided into equal quartiles. A box plot aimed to show whether the game dataset was customarily distributed or skewed. According to figure 4, wealthLow, wealth-medium, and wealthHigh were positively skewed since the whisker was shorter on the lower end of the box. Any observations that were numerically far from the rest of the dataset were considered an outlier. For instance, from 449450.5 to 498733, 144732.5 and 229840.6 were outliers in happiness and profits**.** Figure 5 shows a bar graph for values of various regions visually represented. The values for pleasure and profits are high in Ashgate. In all areas, happiness correlates with the number of profits the region receives—high profits results in great satisfaction among people. Parts like sandy, spring folk, surmar, west chain, Bywoods, dry woods, equilibrium, lost isle, montio, and flowing plains have low happiness with no profits.

Conclusion

In conclusion, there is a positive correlation between most predictor variables and happiness. Except for the wealthmedium, which indicates students in the middle class, the rest of the variables show 95% confidence and alpha level=0.05 to be statistically significant. These factors such as profits, employment, poverty levels, availability of stocks in firms, and time affect the students' happiness. The research hypothesis was to check if there was any association between satisfaction and learners' marks at a standard alpha level of 5% and 95% C.I. four assumptions were made that were not to be violated while making the predictor model. A multiple linear regression that conducted estimations with 80.4% of sample data included in the analysis resulted in a general p-value=0.000<0.05. It led to the rejection of the null hypothesis and acceptance of the alternative. In that case, there is 95% confidence that there is a statistically significant relationship between happiness and student marks. As such it is recommended that students of ECON1101 focus on optimising the values of these statistically sinificant variables, and particularly should focus particularly on increasing profit, given the strength of its Pearson correlation, in order to maximise happiness, and therefore, marks.

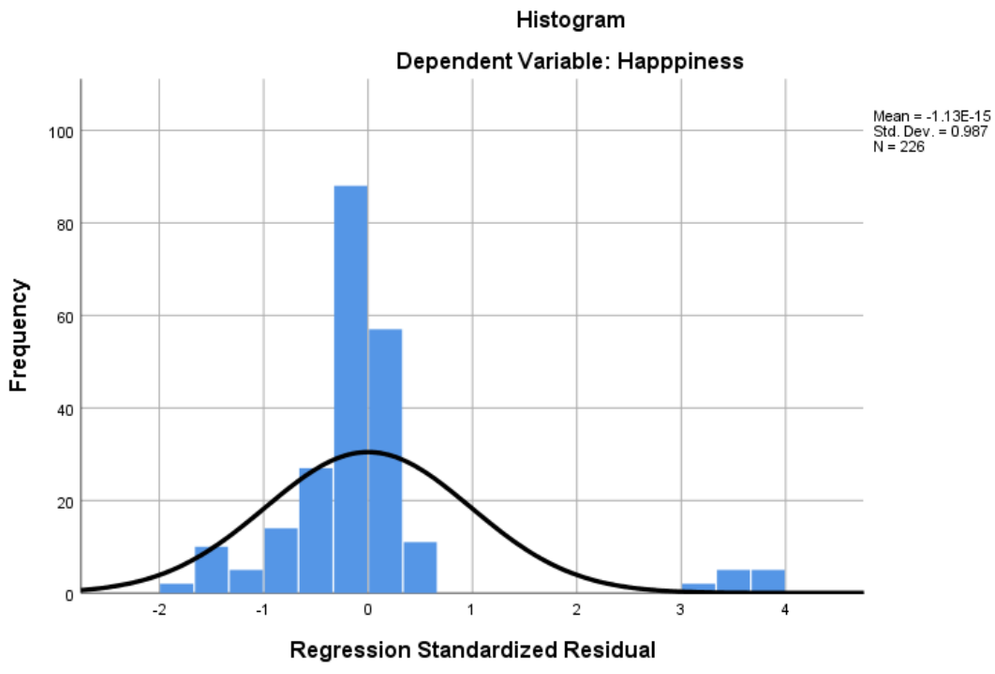
The main limitation of the multiple linear regression model was the assumption of linearity between the happiness and the rest of the predictor variables. It is rare to separate the linearity of data in the real world. A firm or weak correlation could not imply the effects and causes of associations. One could be particular about the relationships between variables but have no clue about the causal mechanism. Outliers were not removed in the box and whisker plots; hence, the R^2 value and Pearson correlation could have been affected negatively by these values. It’s recommended that those who conduct similar investigations in the future can the outliers and find ways to involve accurate data to solve the problem of linearity.

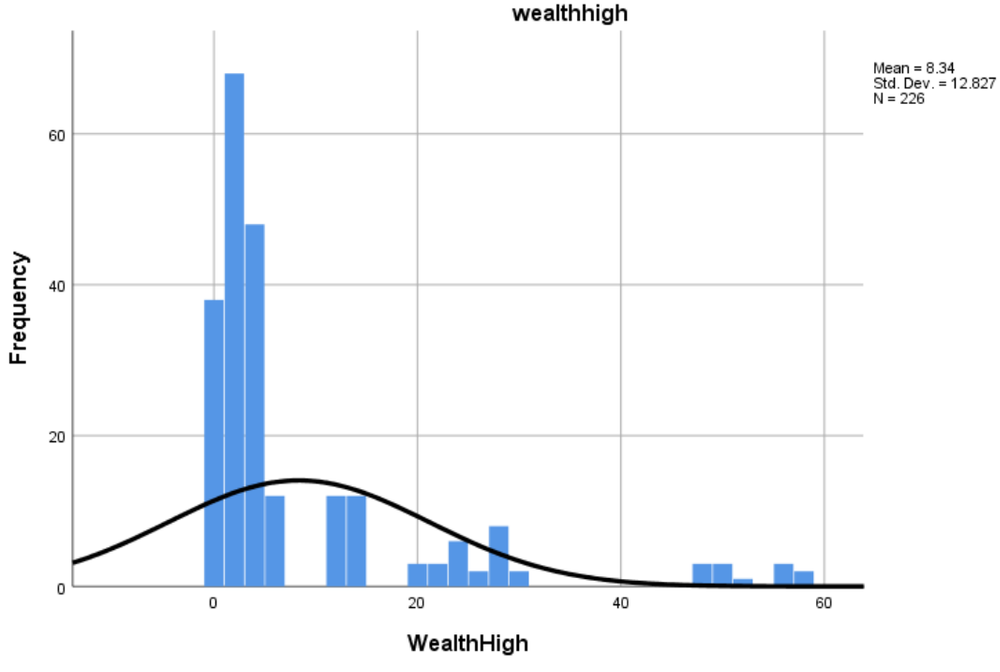
References

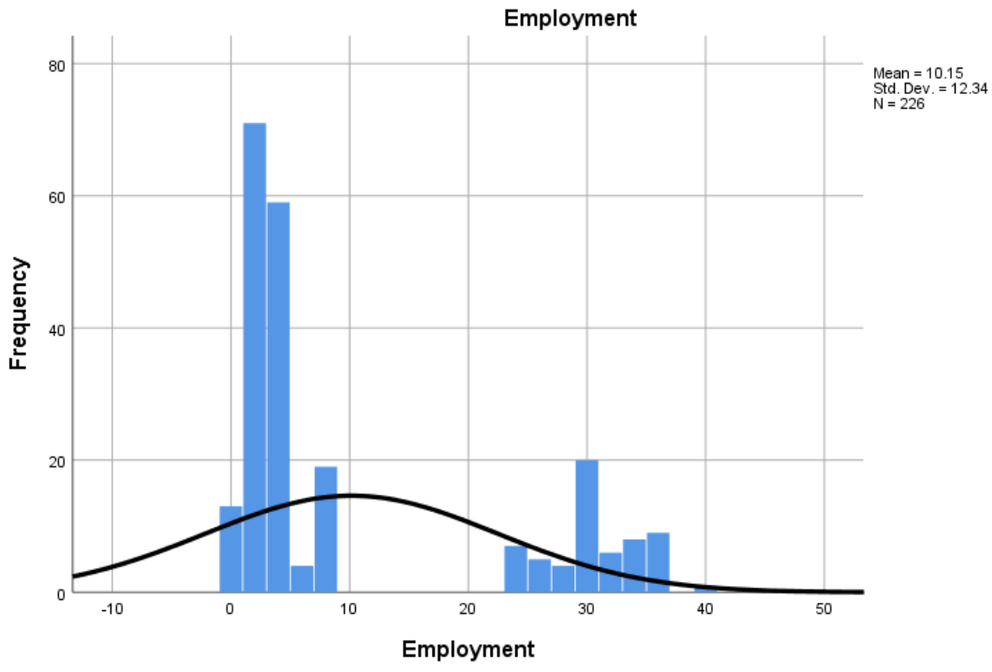
Ernst, A. F., & Albers, C. J. (2017). Regression assumptions in clinical psychology research practice—a systematic review of common misconceptions. *PeerJ*, *5*, e3323. https://peerj.com/articles/3323/

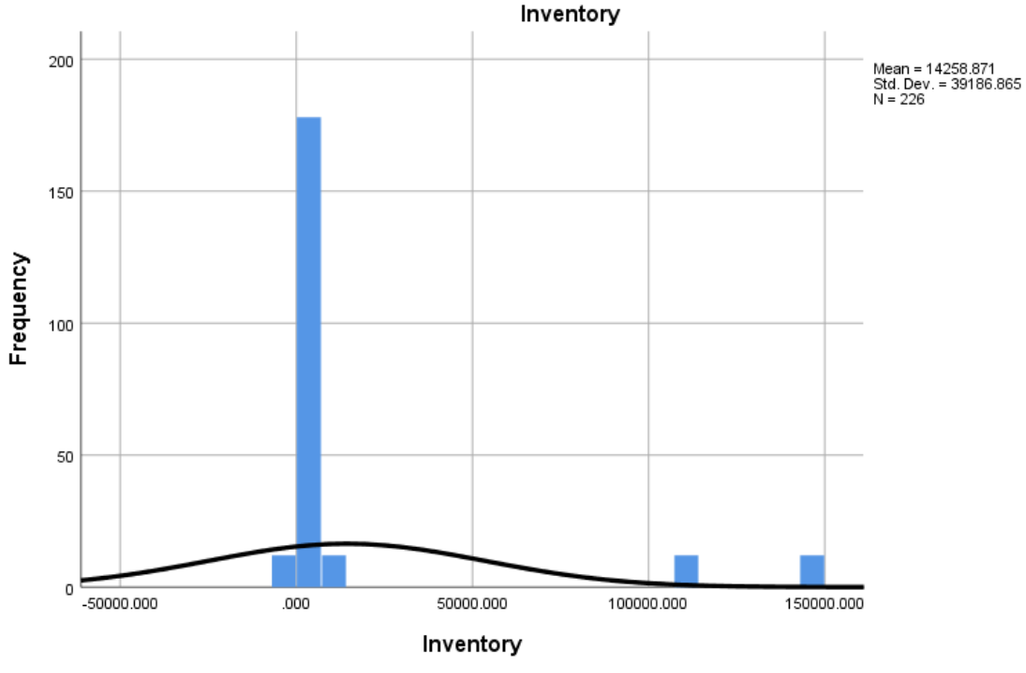
Gomila, R. (2021). Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. *Journal of Experimental Psychology: General*, *150*(4), 700. https://www.robingomila.com/files/publications_pdfs/Gomila_2020_Logistic_vs_Linear.pdf

APPENDICES


Histogram
Dependent Variable: Happpiness
Mean = -1.13E-15
Std. Dev. = 0.987
N = 226
Frequency
Regression Standardized Residual


wealthhigh
Mean = 8.34
Std. Dev. = 12.827
N = 226
Frequency
WealthHigh


Employment
Mean = 10.15
Std. Dev. = 12.34
N = 226
Frequency
80
60
40
20
0
-10
0
10
20
30
40
50
Employment


Inventory
Mean = 14258.871
Std. Dev. = 39186.865
N = 226
Frequency
200
150
100
50
0
-50000.000
.000
50000.000
100000.000
150000.000
Inventory

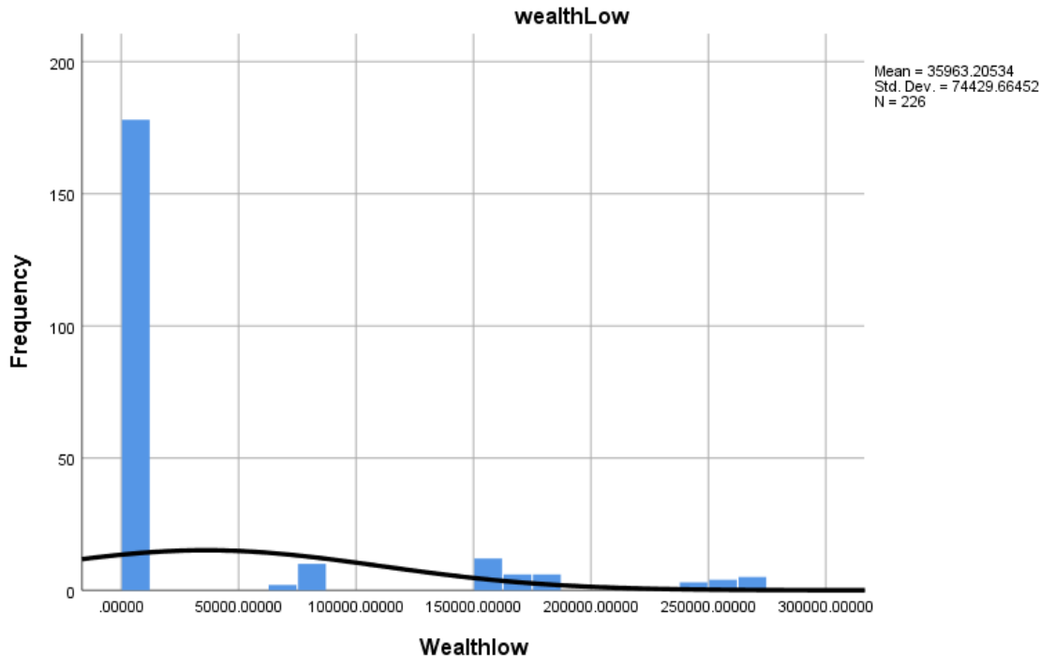
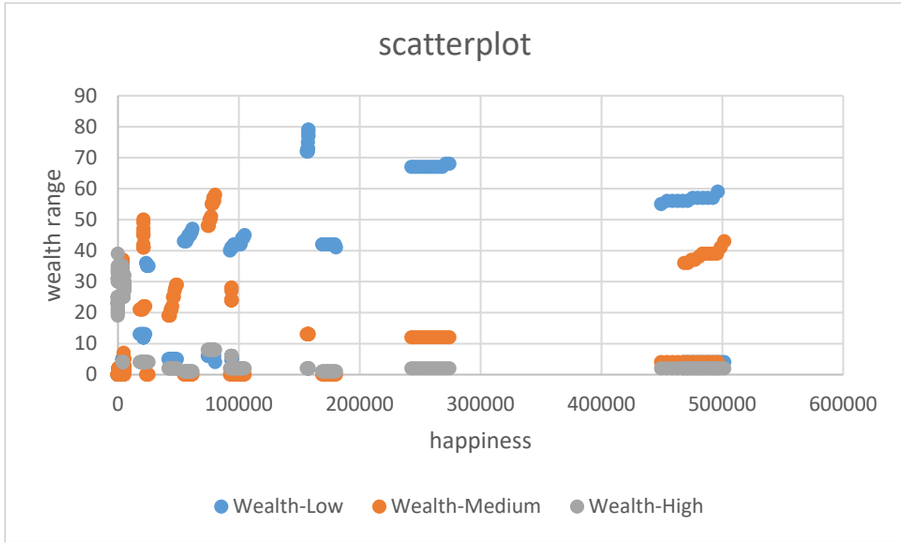
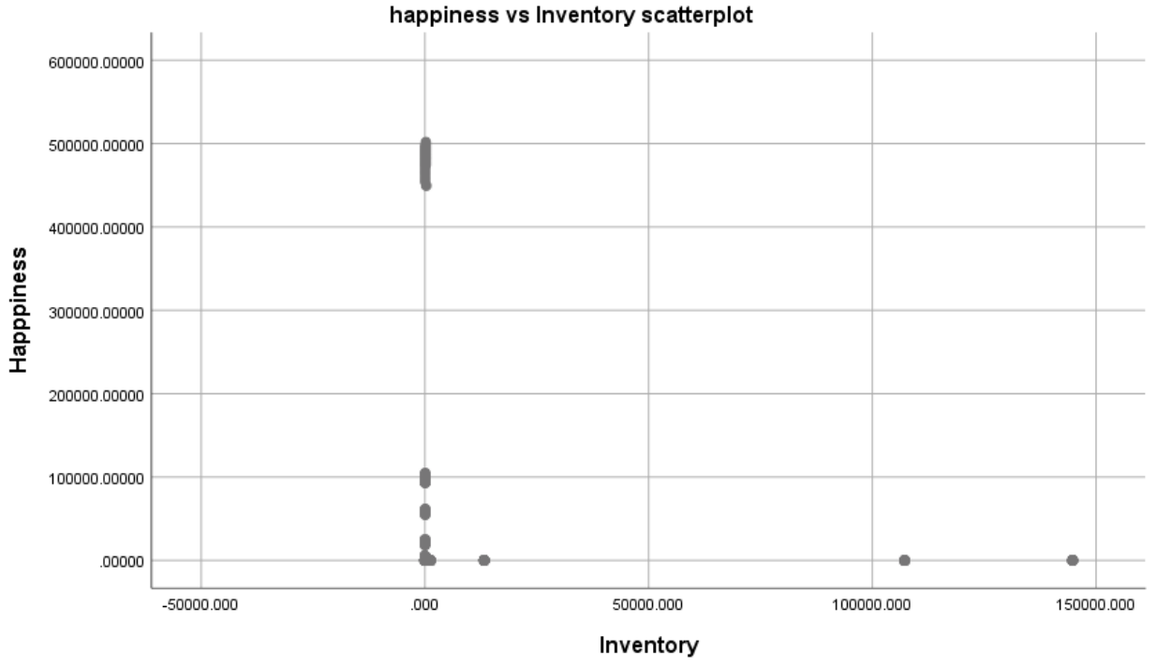
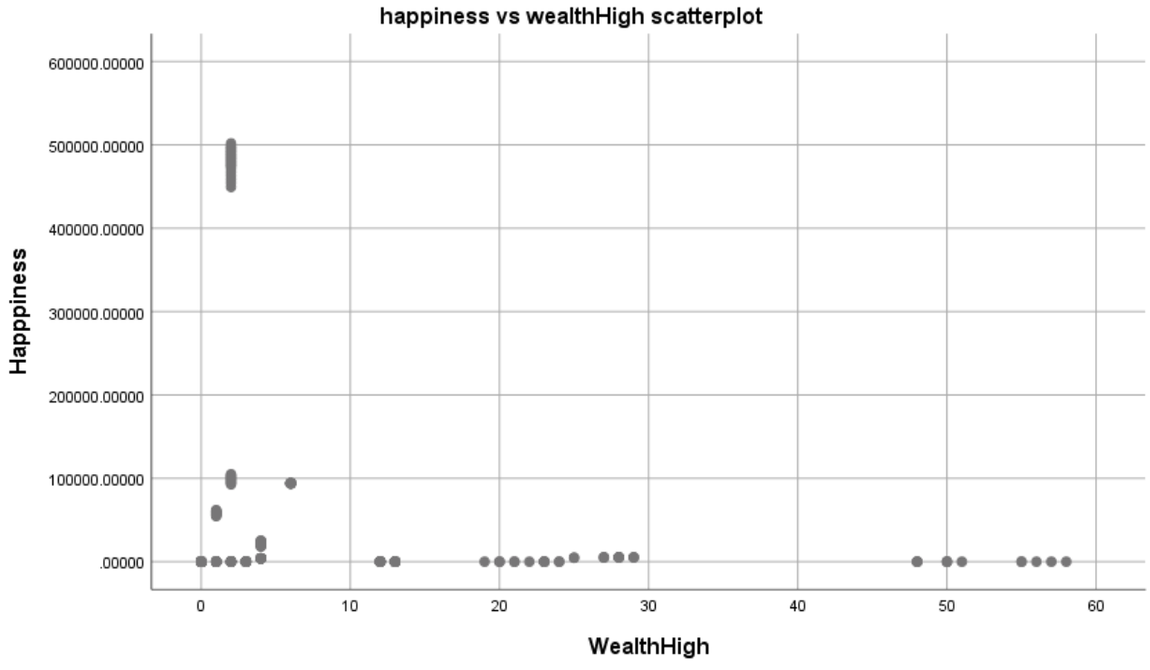


Figure 1: Histograms for Dependent and Independent Variables


scatterplot
90
80
70
60
50
40
30
20
10
0
wealth range
0
100000
200000
300000
400000
500000
600000
happiness
Wealth-Low
Wealth-Medium
Wealth-High


happiness vs Inventory scatterplot
600000.00000
500000.00000
400000.00000
300000.00000
200000.00000
100000.00000
.00000
Happpiness
-50000.000
.000
50000.000
100000.000
150000.000
Inventory



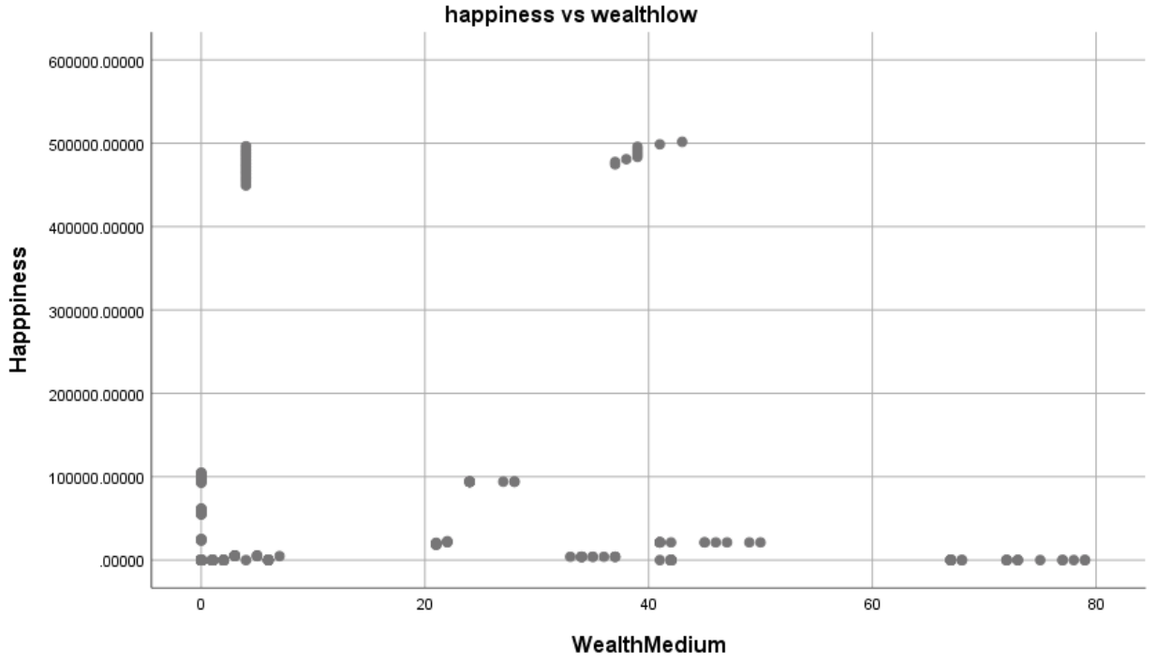
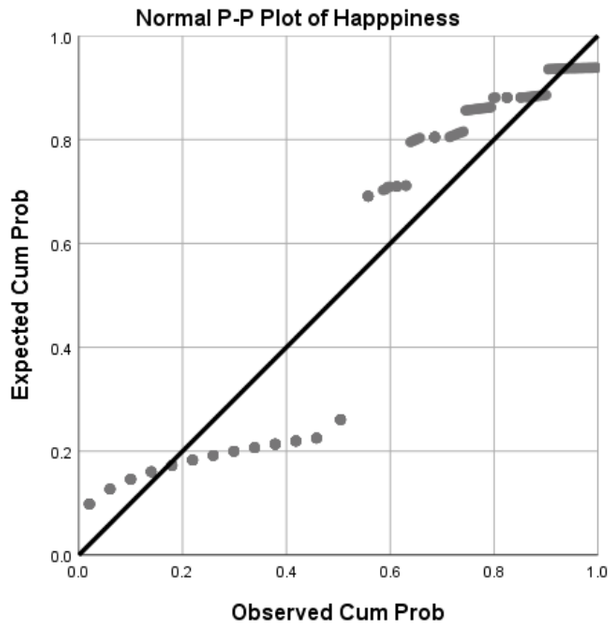
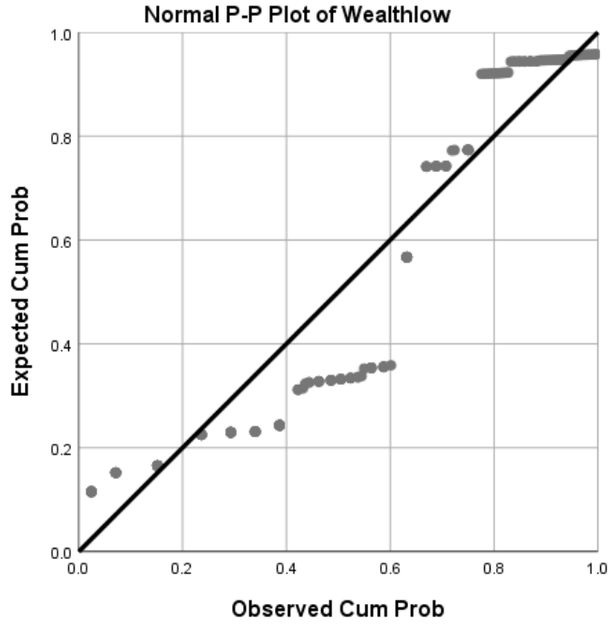


Figure 2: Scatterplots for Dependent and Independent variables



Transforms: natural logarithm



Transforms: natural logarithm

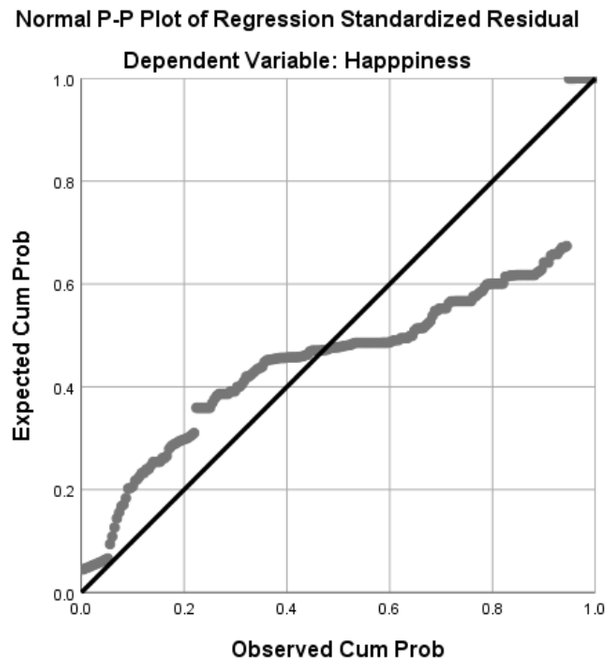


Figure 3: P-P plot for regression Model

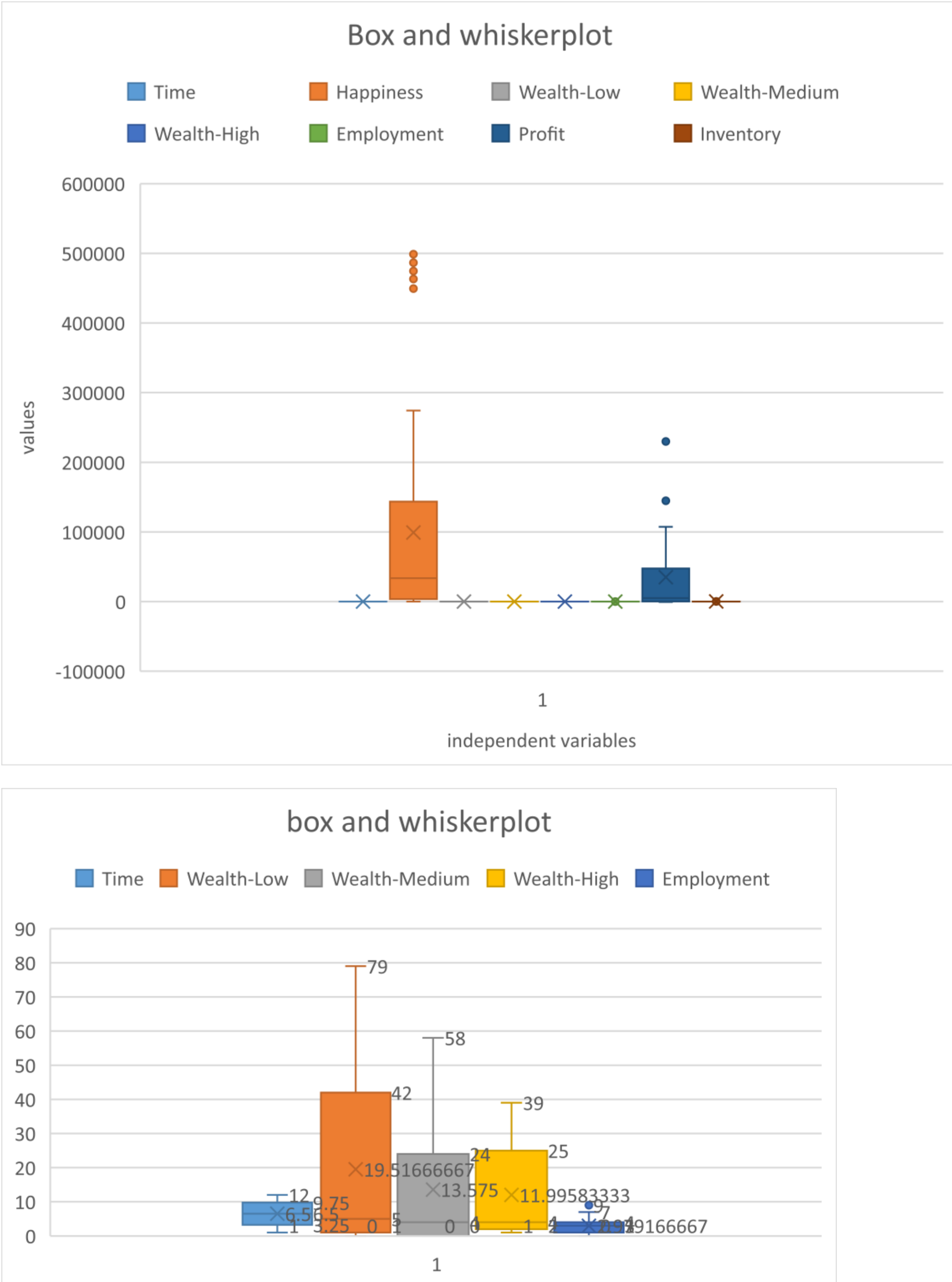


Figure 4: Box and Whisker plots

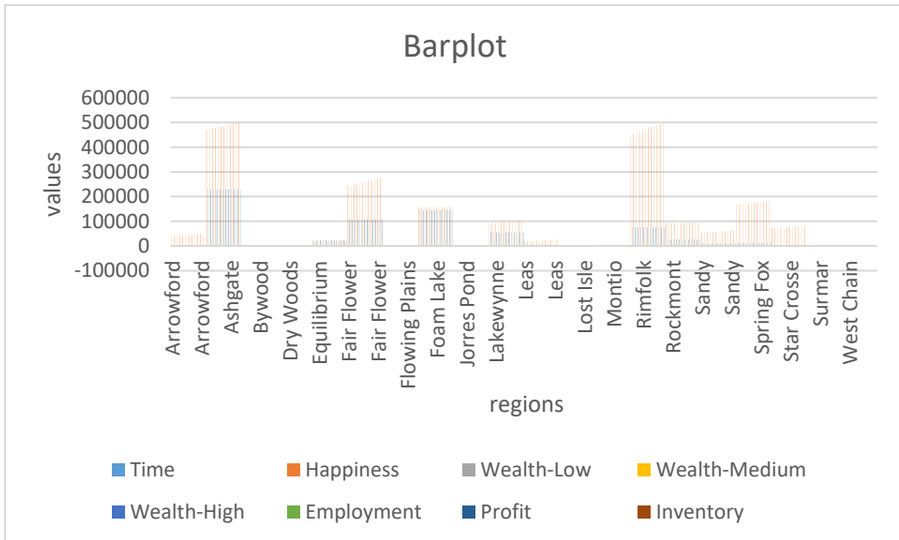


Figure 5: Bar Graph for Dependent and Independent Variables

Case Processing Summary

		Happpiness	Wealthlow	WealthMedium	WealthHigh	Employment	Profit	Inventory
Series or Sequence Length		226	226	226	226	226	226	226
Number of Missing Values in the Plot	Negative or Zero Before Log Transform	0	14[a]	64[a]	38[a]	13[a]	24[b]	48[c]
	User-Missing	0	0	0	0	0	0	0
	System-Missing	0	0	0	0	0	0	0

The cases are unweighted.

a. The minimum value is .000.

b. The minimum value is -636.000.

c. The minimum value is -52.000.

TABLE 1: Case-Processing Summary

Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Happpiness	226	1.00000	501707.8000	62841.65816	140787.2703	1.982E+10	2.521	.162	4.790	.322
Wealthlow	226	.00000	274249.5000	35963.20534	74429.66452	5539774960	1.953	.162	2.488	.322
WealthMedium	226	0	79	18.73	23.948	573.496	1.078	.162	-.093	.322
WealthHigh	226	0	58	8.34	12.827	164.526	2.340	.162	5.144	.322
Employment	226	0	39	10.15	12.340	152.271	1.083	.162	-.611	.322
Profit	226	-636.000	230087.600	20616.00794	49645.51100	2464676763	3.405	.162	11.501	.322
Inventory	226	-52.000	144752.400	14258.87055	39186.86535	1535610416	2.665	.162	5.441	.322
Valid N (listwise)	226									

TABLE 2: Descriptive Statistics

Correlations

		Happpiness	Wealthlow	WealthMedium	WealthHigh	Employment	Profit	Inventory
Pearson Correlation	Happpiness	1.000	-.216	-.010	-.205	-.239	.872	-.162
	Wealthlow	-.216	1.000	.726	.198	-.288	-.201	.791
	WealthMedium	-.010	.726	1.000	-.025	-.466	.103	.773
	WealthHigh	-.205	.198	-.025	1.000	-.267	-.192	.106
	Employment	-.239	-.288	-.466	-.267	1.000	-.218	-.237
	Profit	.872	-.201	.103	-.192	-.218	1.000	-.150
	Inventory	-.162	.791	.773	.106	-.237	-.150	1.000
Sig. (1-tailed)	Happpiness	.	.001	.441	.001	.000	.000	.007
	Wealthlow	.001	.	.000	.001	.000	.001	.000
	WealthMedium	.441	.000	.	.353	.000	.062	.000
	WealthHigh	.001	.001	.353	.	.000	.002	.056
	Employment	.000	.000	.000	.000	.	.000	.000
	Profit	.000	.001	.062	.002	.000	.	.012
	Inventory	.007	.000	.000	.056	.000	.012	.
N	Happpiness	226	226	226	226	226	226	226
	Wealthlow	226	226	226	226	226	226	226
	WealthMedium	226	226	226	226	226	226	226
	WealthHigh	226	226	226	226	226	226	226
	Employment	226	226	226	226	226	226	226
	Profit	226	226	226	226	226	226	226
	Inventory	226	226	226	226	226	226	226

TABLE 3: Correlation coefficients.

Regression Analysis Approach

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.897[a]	.804	.799	63125.21887	.804	150.032	6	219	.000	.138

a. Predictors: (Constant), Inventory, WealthHigh, Profit, Employment, Wealthlow, WealthMedium

b. Dependent Variable: Happpiness

TABLE 4: Model Summary

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3587067762379.819	6	597844627063.303	150.032	.000[b]
	Residual	872669723416.084	219	3984793257.608		
	Total	4459737485795.903	225			

a. Dependent Variable: Happiness

b. Predictors: (Constant), Inventory, WealthHigh, Profit, Employment, Wealthlow, WealthMedium

TABLE 5: Analysis of Variance (ANOVA)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	74851.832	10604.879		7.058	.000	53951.150	95752.514
	Wealthlow	.121	.103	.064	1.177	.241	-.082	.325
	WealthMedium	-2274.908	361.932	-.387	-6.285	.000	-2988.223	-1561.594
	WealthHigh	-1437.915	377.042	-.131	-3.814	.000	-2181.011	-694.819
	Employment	-2264.235	434.260	-.198	-5.214	.000	-3120.098	-1408.372
	Profit	2.506	.097	.884	25.906	.000	2.315	2.697
	Inventory	.669	.207	.186	3.226	.001	.260	1.078

a. Dependent Variable: Happpiness

TABLE 6: Coefficients Results

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-28431.3477	557768.8750	62841.65816	126263.7057	226
Std. Predicted Value	-.723	3.920	.000	1.000	226
Standard Error of Predicted Value	6763.706	19067.051	10574.022	3415.329	226
Adjusted Predicted Value	-30835.3535	565700.5625	62987.51790	127159.3814	226
Residual	-107117.531	252176.8906	.00000000	62277.86207	226
Std. Residual	-1.697	3.995	.000	.987	226
Stud. Residual	-1.715	4.035	-.001	.999	226
Deleted Residual	-109372.836	257324.7813	-145.859740	63810.72381	226
Stud. Deleted Residual	-1.722	4.185	.004	1.021	226
Mahal. Distance	1.588	19.532	5.973	4.782	226
Cook's Distance	.000	.047	.004	.009	226
Centered Leverage Value	.007	.087	.027	.021	226

a. Dependent Variable: Happpiness

TABLE 7: Residuals Statistics

Linear Regression Method

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.326[a]	.106	.094	134003.3375	.106	8.786	3	222	.000	.169

a. Predictors: (Constant), WealthHigh, WealthMedium, Wealthlow

b. Dependent Variable: Happpiness

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4.733E+11	3	1.578E+11	8.786	.000[b]
	Residual	3.986E+12	222	1.796E+10		
	Total	4.460E+12	225			

a. Dependent Variable: Happpiness

b. Predictors: (Constant), WealthHigh, WealthMedium, Wealthlow

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	70821.234	13190.461		5.369	.000	44826.694	96815.774
	Wealthlow	-.732	.184	-.387	-3.977	.000	-1.094	-.369
	WealthMedium	1574.885	560.562	.268	2.809	.005	470.181	2679.590
	WealthHigh	-1339.596	734.086	-.122	-1.825	.069	-2786.265	107.073

TABLE 8: Table Above Indicates Model summary, ANOVA, and Coefficients Values for a Linear Regression Technique