RM_CA3

by Shantanu Dileep Paranjpe

Submission date: 18-Jul-2020 11:57AM (UTC+0100)

Submission ID: 1358976120

File name: 43154_Shantanu_Dileep_Paranjpe_RM_CA3_66843_1552932598.pdf (1.8M)

Word count: 4157

Character count: 22151

Research Methods Project CA-3

On

Statistical Analysis on Household Waste and News Consumption of United Kingdom

MSc in Cloud Computing

Shantanu Paranjpe x19115644
Divya Thorat x18191878
Rahul Dhande x18182852

School of Computing

National College of Ireland

Supervisor: Dr. Vladimir Milosavljevic

Multiple Regression:

The multiple regression helps in identifying the level of variance in the dependent variable which can be illustrated by the independent variables. It also offers an indication of the comparative distribution of each of the independent variables. Multiple regression consists of multiple tests to demonstrate the relationship between one continuous and several independent variables.

Data Source:
Multiple regression has been performed on the total household waste generated in the UK nations from 2010 to 2018. The source for the data source from official government statistics in the UK are as follows:
https://data.gov.uk/dataset/882186e7-97b0-4ad0-b253-e28607252f42/uk-statistics-on-waste

Variables:
Year: From 2010 to 2018
Measure: 0=Non recycled,1=recycled,2=organic,3=general
UK: Total average values
England: Total average values
Northern Ireland: Total Average values
Scotland: Total average values
Wales: Total average values

Objectives:
The household waste in the United Kingdom varies from country to country. The prime objective of this report is to study the different types of household waste produced in four major countries in the UK. The other objectives of this report are as follows:
1. To survey the different types of household wastes
2. To understand the relationship between the dependent and independent variable
3. How the production of waste in different countries varies based on the type of waste

Research Question:
In our research question, the H0 variable exemplifies the null hypothesis and H1 variable signifies the alternative hypothesis.

Do the different measures vary the proportion of generated waste from UK nations?

H0: The volume of waste production is the same across all the UK nations for all types of household waste.

H1: The volume of generated waste varies across all the UK nations for all the types of household waste.

Multiple regression can be computed using the below equation:

$$Y=a+b1*x1+b2*x2+ \ldots\ldots + bp*Xp$$

where
Y indicates the Independent Variables as well as x1, x2 symbolizes the dependent variables.
a is the constant or intercept
Xp is the predictor variable (fixed, nonrandom)
b0, . . . , bp indicates regression coefficients

Descriptive Analysis

Descriptive Statistics

	N	Minimum	Maximum	Sum	Mean		Std. Deviation	Variance
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
Year	36	2010	2018	72504	2014.00	.436	2.619	6.857
Measure	36	0	3	54	1.50	.189	1.134	1.286
UK	36	10878	28932	638031	17723.08	1148.237	6889.425	47464175.62
England	36	4343	84394	595383	16538.42	2181.251	13087.506	171282814.4
NorthernIreland	36	123	976	19930	553.61	38.276	229.654	52740.873
Scotland	36	255	6784	63366	1760.17	258.889	1553.334	2412846.600
Wales	36	241	1943	35297	980.47	79.603	477.615	228116.371
Valid N (listwise)	36							

Descriptive Analysis of UK Waste Dataset

In this, we mainly focus on showing a detailed description of the whole dataset attributes statistics include mean, variance, range, standard deviation, sum, minimum and maximum values. The N indicates the total value of attributes. The purpose of this is to compare the mean value as well as determine the dataset reasonable for Multiple linear regression. For that, we need a Hypothesis.

Multiple Regression Analysis

Regression

Descriptive Statistics

	Mean	Std. Deviation	N
Measure	1.50	1.134	36
UK	17723.08	6889.425	36
England	16538.42	13087.506	36
NorthernIreland	553.61	229.654	36
Scotland	1760.17	1553.334	36
Wales	980.47	477.615	36

The above table gives information about the standard deviation and means of individual variables in the selected dataset. This table is helpful as a summary of individual data. Along with the statistics it is producing the correlation matrix as well.

Correlations

		Measure	UK	England	NorthernIreland	Scotland	Wales
Pearson Correlation	Measure	1.000	-.734	-.217	-.766	-.347	-.495
	UK	-.734	1.000	.504	.677	.132	.526
	England	-.217	.504	1.000	.008	.317	.099
	NorthernIreland	-.766	.677	.008	1.000	.235	.492
	Scotland	-.347	.132	.317	.235	1.000	.299
	Wales	-.495	.526	.099	.492	.299	1.000
Sig. (1-tailed)	Measure	.	.000	.102	.000	.019	.001
	UK	.000	.	.001	.000	.221	.000
	England	.102	.001	.	.482	.030	.282
	NorthernIreland	.000	.000	.482	.	.084	.001
	Scotland	.019	.221	.030	.084	.	.038
	Wales	.001	.000	.282	.001	.038	.
N	Measure	36	36	36	36	36	36
	UK	36	36	36	36	36	36
	England	36	36	36	36	36	36
	NorthernIreland	36	36	36	36	36	36
	Scotland	36	36	36	36	36	36
	Wales	36	36	36	36	36	36

The first row in this table shows the Pearson correlation coefficient among each pair of variables; also it is displaying the one-tailed significance of individual correlations and in the last row it is displaying the number of items contributing to individual correlation. In the Pearson correlation table if we see diagonally the values of the correlation coefficient are 1.00 that means it represents the correlation of individuals with itself so the ultimate outcome becomes 1. This matrix is useful for getting the idea of predictors and results and in the initial look, it is useful for multicollinearity. In this multicollinearity not be available between the data then the correlation between the predictors is not available.

Interpretation of Results for Collinearity:

The correlation among the independent variable and the dependent variables can be demonstrated using the Correlations table. The preferred value for displaying collinearity is 0.3. The quantity of the household wastes generated across the 5 different countries in the United Kingdom clearly illustrates collinearity with the dependent variable namely **Measure.** The collinearity for the household waste generated in the UK, England, Northern Ireland is -0.7, -0.21[illegible]0.766 whereas the collinearity for Scotland and Wales is -0.34 and -0.495. In the results it shows that all the independent variables significantly correlated with the dependent variable. In order to establish the collinearity among the independent variables, the ideal value for data to be collinear should be less than 0.7. In our case, all the values for the collinearity are less than 0.7, which implies that there is considerable collinearity between the independent variables.

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Wales, England, Scotland, Northernirela nd, UK[b]	.	Enter
2	.	Wales	Backward (criterion: Probability of F-to-remove >= .100).
3	.	England	Backward (criterion: Probability of F-to-remove >= .100).

a. Dependent Variable: Measure

b. All requested variables entered.

Model Summary[d]

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	.844[a]	.712	.664	.657	.712	14.819	5	30	.000	
2	.844[b]	.711	.674	.647	.000	.032	1	30	.860	
3	.840[c]	.705	.678	.644	-.006	.671	1	31	.419	2.537

a. Predictors: (Constant), Wales, England, Scotland, Northernireland, UK

b. Predictors: (Constant), England, Scotland, Northernireland, UK

c. Predictors: (Constant), Scotland, Northernireland, UK

d. Dependent Variable: Measure

In the summary model table, it consists of three models. The first model consists of all countries in Europe as the predictor; the second stage model consists of England, Scotland, Northern Ireland and the UK as a predictor, Third stage model consists of Scotland, Northern Ireland, and the UK as predictor and Measure as the dependent variable. This table tells us what the results are and which predictors are used in these three models. Column R tells us the values of the correlation coefficient between the predictor and results. When all the countries are considered as a predictor, then there is a correlation between the measures and all the countries, as shown in a table, its value is 0.844. R^2 values give values of variability in the results, which is considered by the predictors. In the first model, its values are 0.72, which means all countries consider 71.2 % variation in measures. When the other two models are considered, then its values decrease to 71.1 % and 70.5 %. The adjusted R^2 tells how the model is generalized and ideally, its value should be the same or close to R^2. In the third model, the difference between the R^2 and the adjusted R^2 is less (0.705 - 0.678 = 0.027).

The Durbin-Watson tests for statistics which help us to tell the assumption about the independent error are tenable. According to conservative rule, If the value is close to 2 then that means your chances of getting the error are below 5 %.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	32.031	5	6.406	14.819	.000[b]
	Residual	12.969	30	.432		
	Total	45.000	35			
2	Regression	32.017	4	8.004	19.113	.000[c]
	Residual	12.983	31	.419		
	Total	45.000	35			
3	Regression	31.736	3	10.579	25.523	.000[d]
	Residual	13.264	32	.414		
	Total	45.000	35			

a. Dependent Variable: Measure

b. Predictors: (Constant), Wales, England, Scotland, NorthernIreland, UK

c. Predictors: (Constant), England, Scotland, NorthernIreland, UK

d. Predictors: (Constant), Scotland, NorthernIreland, UK

ANOVA tests tell whether the model is better at forecasting the results than the other ways, such as using the mean. In this table, F-ratio tells that it is the ratio of the result to forecast from the model relating to errors that exist in a model. It consists of three models. In this table, it represents the column sum of squares, degree of freedom, and mean squares. Mean square calculated by dividing the sum of squares to the degree of freedom. F-ration represented by dividing the average enhancement in the forecast by the model by average dissimilarity between the model and your data.

<u>Output Interpretation for ANOVA:</u>

The ANOVA table compares three models representing the name of different countries in the UK. To verify the level of variance in the dependent variable using the predictor or factor variables, we give emphasis on the value of the F statistics and the significance value. As given in the table, the value of the F statistics is 14.819 for Model 1, the value of the F statistics is 19.113 for Model 2 and lastly, the Model 3 has the F statistics value as 25.523. The F statistics values suggest a significant improvement in the level of variance that the independent variables have on the dependent variables. We can also see that the significance value p for all the three models is 0.00 value is less than 0.05 and it indicates that there is a significant variance in the means values of the independent and the dependent variables.

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	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	4.114	.338		12.180	.000	3.424	4.804		
	UK	-8.845E-5	.000	-.537	-2.501	.018	.000	.000	.208	4.804
	England	1.130E-5	.000	.130	.823	.417	.000	.000	.383	2.614
	NorthernIreland	-.002	.001	-.358	-2.058	.048	-.004	.000	.318	3.145
	Scotland	.000	.000	-.240	-1.960	.059	.000	.000	.642	1.559
	Wales	5.382E-5	.000	.023	.178	.860	-.001	.001	.594	1.682
2	(Constant)	4.126	.326		12.660	.000	3.461	4.790		
	UK	-8.580E-5	.000	-.521	-2.716	.011	.000	.000	.253	3.960
	England	1.059E-5	.000	.122	.819	.419	.000	.000	.418	2.392
	NorthernIreland	-.002	.001	-.359	-2.101	.044	-.003	.000	.319	3.139
	Scotland	.000	.000	-.232	-2.058	.048	.000	.000	.730	1.369
3	(Constant)	4.148	.323		12.837	.000	3.489	4.806		
	UK	-6.691E-5	.000	-.407	-3.115	.004	.000	.000	.541	1.850
	NorthernIreland	-.002	.001	-.446	-3.353	.002	-.004	-.001	.520	1.923
	Scotland	.000	.000	-.188	-1.906	.066	.000	.000	.944	1.060

a. Dependent Variable: Measure

In a coefficient table, it's an option depends on the selection. The contribution of each variable to the multiple regression model found in the coefficient table. In the first part of the table, it gives the estimated b-values, and it represents the individual contribution of each forecast to the model. The b value represents the relationship between the measure and other forecasts. In this, each value of beta has an associated standard error, which tells the range of these values that may change across the samples. In a multiple regression model, t-tests used to tell if the forecast is making a significant contribution to the model. By looking at the sig. Value column we can tell where there is any contribution of each variable to obtain the results.

<u>Coefficient Table Output Interpretation:</u>

The Coefficients table consists of 3 models which are segregated as per the name of the countries. we mainly focus on the column named Beta which contains the Beta values for displaying the level of contribution of the factor variables in predicting the continuous variable namely Measure. As we can see in the Standardized coefficient column, the Beta value for the UK is -.537 and it is the highest contributor in forecasting the value of the continuous variable. While the Beta value for Wales as an independent variable is 0.023 is the lowest contributor in estimating the variability in the dependent variable. The significance value for all the independent variables in all the 3 models is below 0.05 and this clearly reveals that the factor variables significantly explain the variance in the mean of the dependent variable. The last two columns namely Tolerance and the Variance Inflation Factor. The value for the variance inflation factor should ideally be less than 10. In our case, all the predictor values are less than 10 and the probability of errors is low for these variables. To find the exact causes of the possible errors, we use the collinearity diagnostics table.

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions					
				(Constant)	UK	England	NorthernIreland	Scotland	Wales
1	1	5.144	1.000	.00	.00	.00	.00	.01	.00
	2	.362	3.768	.01	.00	.10	.01	.36	.02
	3	.303	4.123	.00	.00	.25	.00	.33	.01
	4	.096	7.305	.29	.00	.02	.04	.02	.73
	5	.077	8.170	.69	.04	.00	.18	.01	.09
	6	.017	17.399	.00	.95	.63	.75	.27	.14
2	1	4.270	1.000	.01	.00	.01	.00	.01	
	2	.344	3.526	.02	.01	.03	.02	.58	
	3	.288	3.850	.01	.00	.37	.02	.20	
	4	.079	7.357	.96	.05	.00	.11	.00	
	5	.019	14.923	.00	.94	.59	.85	.21	
3	1	3.539	1.000	.01	.01		.01	.02	
	2	.339	3.232	.02	.02		.01	.95	
	3	.079	6.678	.93	.07		.23	.01	
	4	.042	9.130	.05	.90		.75	.02	

a. Dependent Variable: Measure

In the collinearity diagnostics, Look for variables whose values have a high proportion to the dimension. The variance value may change between 0 and 1, and its forecast may be changed across different dimensions. The collinearity diagnostics table encompasses 3 models that are distributed as per the country names. The table contains a column named Dimensions and it consists of 6 dimensions with the corresponding Eigenvalue. The dimensions with high Eigenvalue display a large dependent contribution to the data. The dimensions with low Eigenvalue can lead to major issues as they do not provide a meaningful contribution to the sample data and is a sign of multicollinearity. The next column we need to look at is the Condition Index which is derived from the Eigenvalues. As given by IBM, the threshold value for the condition index 15 and when the value is above 15 it depicts problems with multicollinearity, and values above 15 are sharp indicators of multicollinearity

<u>Output Interpretation of Correlation Diagnostics:</u>

In our table, we can observe that in Model 1, the condition index value for dimension 6 corresponding to the UK is 17.399 which is above the threshold value of 15 and it reveals problems associated with multicollinearity. All the other values are within the threshold value. We look for values in the Variance Proportion and the corresponding value for the dimension 6 and we search for values above 9.0. In our table, the variance proportion for the UK is 0.95 which is above the 9.0 threshold value. So this shows that the UK as a factor variable has a collinearity issue as opposed to all other values mentioned in the variance proportion column mentioned in the table

Excluded Variables[a]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance	VIF	Minimum Tolerance
2	Wales	.023[b]	.178	.860	.033	.594	1.682	.208
3	Wales	-.008[c]	-.065	.949	-.012	.650	1.539	.486
	England	.122[c]	.819	.419	.146	.418	2.392	.253

a. Dependent Variable: Measure

b. Predictors in the Model: (Constant), England, Scotland, NorthernIreland, UK

c. Predictors in the Model: (Constant), Scotland, NorthernIreland, UK

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	.13	2.61	1.50	.952	36
Residual	-2.460	1.031	.000	.616	36
Std. Predicted Value	-1.436	1.165	.000	1.000	36
Std. Residual	-3.820	1.602	.000	.956	36

a. Dependent Variable: Measure

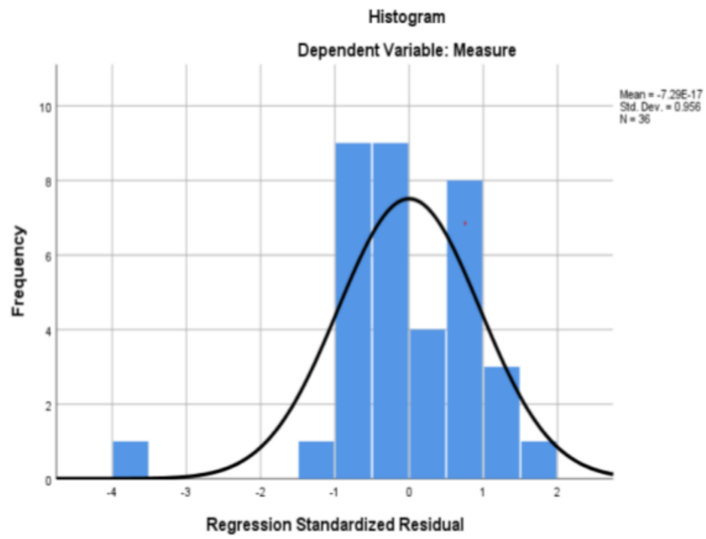
This table represents the summary of the variable that does not include in the model. This summary gives the details of the forecaster's beta value if it has entered into the problem and calculates its t-tests value for the same. The partial correlation gives the results that if the excluded predictor is included for the prediction, then what amount of contribution is given. VIF and tolerance tell whether there is collinearity between the data. If the VIF value higher than 10, then there is a problem. If the VIF is greater than 1, then the regression can be biased. A tolerance value below 0.1 shows a problem. The tolerance value below 0.2 shows the potential issue.

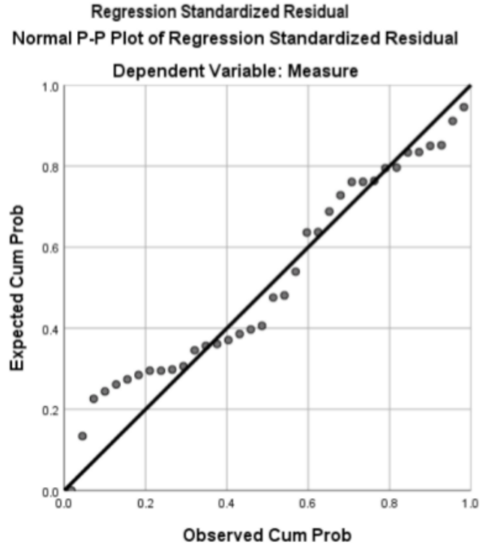
<u>Output Interpretation of Residual Statistics:</u>

The residual statistics table aims to focus on showing dependant variable mean difference of predicted value which is 1.50 and not equal to zero as well as the total value of N is 36 are fit for data. These all conclude the predicted value reasonable to show analysis through residual plots.

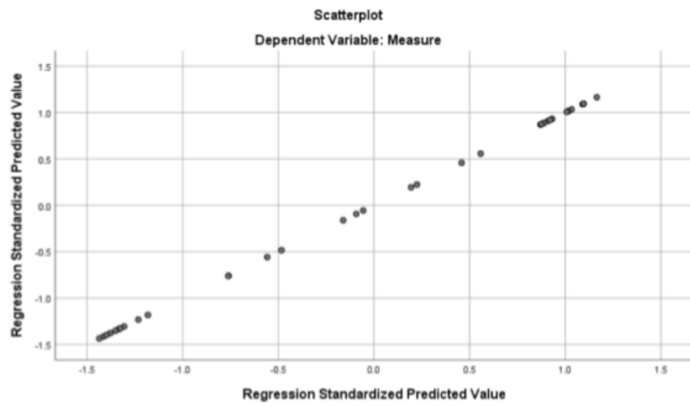
Regression diagnostics:

Regression Diagnostics is responsible for testing the assumptions of the regression model as well as to examine whether or not there are findings with a significant, undue effect on the study. We can check which of these assumptions are best fits in the following residual plots. The **purpose** of the following histogram in multiple regression is to verify the variance is typically distributed or not. It depicts an excellent normal distribution that is referred to as Multivariate Normality. The residuals on the left side predicted by the independent variables are normal.





The **purpose** of the P-P plot of residuals is to compare standardized residuals of observed cumulative distribution function (CDF) on X-Axis with normal distribution 's expected CDF on Y-Axis. The above P-P plot shows the straight normality line with little circles from bottom left to bottom right. So the data are normal with no drastic deviations. It also shows Normality assumptions which are considerable.



The **objective** of the scatterplot of residuals is to determine whether a linear regression model is suitable for the data or not. The above scatterplot depicts the regular straight line from bottom left to the bottom right that is a good relationship between different measures like an organic, general waste with various UK nations like England, Scotland and Northern Ireland. Also, both X-Axis (Regression standardised predicted value) and Y-Axis show linearity and are approximately normally distributed with no outliers.

Result Analysis:

We successfully examined interpretations and performed all tests such as Durbin Watson and ANOVA as well as Correlations, Coefficients, Collinearity and Residuals Statistics. The Correlations show the multicollinearity between Independent and dependent variables, whereas Coefficients indicate the variance in significant values for all Independent variables.Also, correlation diagnosis shows the threshold value is high, so still facing collinearity issues. Additionally, Residuals plots help to understand the total ratio of waste from UK nations. Simultaneously, the **significant difference** between Dubin Watson tests depicts the independent variable error compared to ANOVA focus on mean value difference and F-statistics of Independent variables helps UK waste authorities for the decision-making process of waste management.

Conclusions and Decisions:

Household waste production is enormous and massively under-estimated in all UK nations. There is a requirement of proper management and decision-making process to control waste production. Multiple Linear Regression plays a significant role to show statistical analysis for UK household waste production yearly from 2010 to 2018. We successfully examined and significantly demonstrated the relationship between dependant (Measure) and several Independent variables (UK nations).It shows the variance between dependant and Independent variables.It includes results of correlations, collinearity, coefficients, residuals with Durbin Watson and ANOVA tests. Also, the Residuals plots graphical visualisation, including histogram, P-P plot scatterplot indicates the normal distributions with precise data. The results of the multiple regression statistical model will help on the enhancement of various kinds of waste and recycling management in United Kingdom nations for sustainable development as well as improve the accuracy and save the world. Besides, UK waste authorities will work on these results to assure optimum resource distribution and adequate service for society. The novel analysis will help to improve business revenue and sustainable development of UK nations.

Binary Logistic Regression

Binomial logistic regression (often simply called logistic regression) estimates the possibility that a sample dataset falls into either of two types of dichotomous dependent variable categories depending on one or more independent variables which can be either constant or categorical.

Data Source:

The binomial logistic regression has been demonstrated on the news consumption in the UK based on gender. The information about the dataset was gathered from official government statistics of the UK are as follows:
https://data.gov.uk/dataset/4166149f-f10f-4f60-8174-ad30ba99fa86/news-consumption-in-the-uk

Variables:

Gender: 0=male,1=female

Platforms:0=Magazine,1=Mobile,2=news,3=other Internet,4=Radio,5=Tablet,6=Television

TVRegions:0=BBC Midlands West,1=BBC Northern Ireland,2=BBC North-West,3=BBC South East,4=BBC South West

Objectives:

The modes of news consumption vary from country to country and are also dependent on gender. In the binomial regression analysis, the aim of this report is to cover the below objectives:

1. To survey different modes of news consumption
2. To understand the relationship between the dichotomous dependent variable and the categorical independent variables
3. To demonstrate how the consumption of news is dependent on gender

Descriptive Analysis

Descriptive Statistics

	N	Range	Minimum	Maximum	Sum	Mean		Std. Deviation	Variance
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
Gender	199	1	0	1	110	.55	.035	.498	.248
Platform	199	6	0	6	569	2.86	.155	2.188	4.788
TVRegion	199	4	0	4	423	2.13	.081	1.146	1.312
Valid N (listwise)	199								

Descriptive Analysis of News Consumption Survey in the UK dataset

In this, we mainly focus on showing a detailed description of the whole dataset attributes statistics include mean, variance, range, standard deviation, sum, minimum and maximum values. The N indicates the total value of attributes. The purpose of this is to compare the mean value as well as determine the dataset reasonable for binary logistic regression. For that, we need a Hypothesis.

Research Question

What is the variance for all types of peoples keen to watch the news on different platforms which fall into various tv regions?

In this analysis, we have one dependant variable(i.e.Genders) and two independent variables (i.e. Platforms and TV regions). The question address 1*2 factorial of Binary Regression Model, which has a **purpose** is to find out the relationship between dependent variables on several Independent variables. We individually consider Genders as a dependent variable to examine the mean value variance and Hypothesis.

Hypothesis:

Null-H0: The Genders mean value is more than Platforms and TV regions
Alternative-H1: TV Regions and Platforms mean values not more than Genders.

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	199	100.0
	Missing Cases	0	.0
	Total	199	100.0
Unselected Cases		0	.0
Total		199	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
0	0
1	1

The above **case processing summary** table indicates and aims to show no unweighted cases, and total cases are 199 in this analysis. Also, it shows missing cases 0 that concludes no missing data.
Additionally, the **Dependant variable encoding** table aims to focus on recording to the dependent variable for the analysis. The result will be marked 0 and 1 in binary logistic regression. In this table, there no outcome variable coded.so, predicting membership in the value is 1 as compared to membership in the value is 0

Categorical Variables Codings

		Frequency	Parameter coding (1)	(2)	(3)	(4)	(5)	(6)
Platform	0	38	.000	.000	.000	.000	.000	.000
	1	36	1.000	.000	.000	.000	.000	.000
	2	16	.000	1.000	.000	.000	.000	.000
	3	35	.000	.000	1.000	.000	.000	.000
	4	12	.000	.000	.000	1.000	.000	.000
	5	24	.000	.000	.000	.000	1.000	.000
	6	38	.000	.000	.000	.000	.000	1.000

In the above table, platform section 0 represents a magazine, 1 represents mobile, 2 represents news, 3 represents another internet, 4 represents radio, 5 represents tablet, and 6 represents television. The **purpose** of parameter coding is essential for calculating the probability of the result variable.

Block 0: Beginning Block

Iteration History[a,b,c]

Iteration		-2 Log likelihood	Coefficients Constant
Step 0	1	273.652	.211
	2	273.652	.212

a. Constant is included in the model.

b. Initial -2 Log Likelihood: 273.652

c. Estimation terminated at iteration number 2 because parameter estimates changed by less than .001.

Classification Table[a,b]

Observed			Predicted Gender 0	Gender 1	Percentage Correct
Step 0	Gender	0	0	89	.0
		1	0	110	100.0
	Overall Percentage				55.3

a. Constant is included in the model.

b. The cut value is .500

This is the baseline model and gives the output of data when the dependent variable is included. It also gives the log-likelihood output, i.e., 273.65. When the single dependent variable included the computer calculates its value by assigning each user to one category of the results variable.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	.212	.143	2.208	1	.137	1.236

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Platform	14.602	6	.024
		Platform(1)	.001	1	.970
		Platform(2)	.007	1	.935
		Platform(3)	.383	1	.536
		Platform(4)	10.332	1	.001
		Platform(5)	2.045	1	.153
		Platform(6)	.130	1	.718
		TVRegion	5.727	1	.017
	Overall Statistics		20.734	7	.004

The variable in the equation tells what the value of the constant is. In our case, it is 0.212. The table variable not in the equation defines the residual chi-square statistics. So it defines that the coefficient of the variable those not in the model usually is different than the zero.

Block 1: Method = Enter

Iteration History[a,b,c,d]

Iteration		-2 Log likelihood	Coefficients							
			Constant	Platform(1)	Platform(2)	Platform(3)	Platform(4)	Platform(5)	Platform(6)	TVRegion
Step 1	1	251.023	-1.148	.782	1.031	.868	2.306	.149	.692	.326
	2	248.703	-1.290	.844	1.133	.934	3.502	.184	.737	.375
	3	248.005	-1.301	.848	1.141	.937	4.560	.187	.739	.380
	4	247.762	-1.302	.848	1.141	.938	5.579	.187	.739	.380
	5	247.674	-1.302	.848	1.141	.938	6.586	.187	.739	.380
	6	247.642	-1.302	.848	1.141	.938	7.588	.187	.739	.380
	7	247.630	-1.302	.848	1.141	.938	8.589	.187	.739	.380
	8	247.626	-1.302	.848	1.141	.938	9.590	.187	.739	.380
	9	247.624	-1.302	.848	1.141	.938	10.590	.187	.739	.380
	10	247.623	-1.302	.848	1.141	.938	11.590	.187	.739	.380
	11	247.623	-1.302	.848	1.141	.938	12.590	.187	.739	.380
	12	247.623	-1.302	.848	1.141	.938	13.590	.187	.739	.380
	13	247.623	-1.302	.848	1.141	.938	14.590	.187	.739	.380
	14	247.623	-1.302	.848	1.141	.938	15.590	.187	.739	.380
	15	247.623	-1.302	.848	1.141	.938	16.590	.187	.739	.380
	16	247.623	-1.302	.848	1.141	.938	17.590	.187	.739	.380
	17	247.623	-1.302	.848	1.141	.938	18.590	.187	.739	.380
	18	247.623	-1.302	.848	1.141	.938	19.590	.187	.739	.380
	19	247.623	-1.302	.848	1.141	.938	20.590	.187	.739	.380
	20	247.623	-1.302	.848	1.141	.938	21.590	.187	.739	.380

a. Method: Enter

b. Constant is included in the model.

c. Initial -2 Log Likelihood: 273.652

d. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

The above Block 1 table aims to focus on the value of the coefficient of different platforms where Genders rely on to watch the news. We can see the Step 1 coefficient constant value and 2-Log likelihood 247.62 is different as compared to block 0 Step 0. The purpose is to determine the impact of Independent variables. It concludes that it highly impacted on dependent variable Genders.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	26.029	7	.000
	Block	26.029	7	.000
	Model	26.029	7	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	247.623[a]	.123	.164

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

In the above output, it gives info about the new model. It also tells the value of the chi-square goodness of the fittest results. It is used to check the new model with the improvement in the basic model. By using the chi-square test, it tells us that there is a significant difference between the log likelihoods of the basic model to the new model. In this model, the block is important as it compares the new model with the basic model. Step and block models are only important when one is going to add an explanatory variable to the model in such a stepwise manner. In a model summary table, Cox & snell square and Nagelkerke R square values are present. Its likelihood value is compared with the likelihood value of the previous model. Nagelkerke R square value is used to tell approximately 16% variation in the output is present.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	11.382	8	.181

Contingency Table for Hosmer and Lemeshow Test

		Gender = 0		Gender = 1		
		Observed	Expected	Observed	Expected	Total
Step 1	1	19	17.787	8	9.213	27
	2	12	11.171	7	7.829	19
	3	11	8.697	5	7.303	16
	4	9	8.814	8	8.186	17
	5	7	6.450	6	6.550	13
	6	5	7.214	11	8.786	16
	7	5	8.023	13	9.977	18
	8	6	10.299	19	14.701	25
	9	10	7.913	15	17.087	25
	10	5	2.633	18	20.367	23

Classification Table[a]

			Predicted		
			Gender		Percentage Correct
Observed			0	1	
Step 1	Gender	0	51	38	57.3
		1	28	82	74.5
	Overall Percentage				66.8

a. The cut value is .500

The Hosmer test is used for testing an overall model. It consists of three steps: the first one is probability in terms of ordering sequence, secondly group then into a decile and the third step is comparison between the observed and expected frequencies relevant to chi-square tests. It is also used to specify the goodness of fit to the regression model. The non-significant test results show a better fit model. As the sig value is 0.177 that means the model is a good fit.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	Platform			6.278	6	.393			
	Platform(1)	.848	.495	2.934	1	.087	2.336	.885	6.168
	Platform(2)	1.141	.658	3.013	1	.083	3.131	.863	11.359
	Platform(3)	.938	.497	3.565	1	.059	2.554	.965	6.761
	Platform(4)	21.590	11356.272	.000	1	.998	2379040974	.000	.
	Platform(5)	.187	.543	.118	1	.731	1.205	.416	3.491
	Platform(6)	.739	.479	2.387	1	.122	2.095	.820	5.351
	TVRegion	.380	.148	6.586	1	.010	1.462	1.094	1.954
	Constant	-1.302	.514	6.428	1	.011	.272		

a. Variable(s) entered on step 1: Platform, TVRegion.

Correlations

		Gender	Platform	TVRegion
Gender	Pearson Correlation	1	.076	.170*
	Sig. (2-tailed)		.284	.017
	N	199	199	199
Platform	Pearson Correlation	.076	1	.039
	Sig. (2-tailed)	.284		.581
	N	199	199	199
TVRegion	Pearson Correlation	.170*	.039	1
	Sig. (2-tailed)	.017	.581	
	N	199	199	199

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

The regression slope for the platform is positive and statistically non-significant(b = 0.058, p= 0.391), probability of watching different media in different regions is different. The regression slope for the TV region is positive and statistically non-significant. The regression slope for gender is negative and statistically non-significant.(b = -0.810, p = 0.356).

<u>Correlation Matrix:</u>

A correlation matrix involves computing all the possible pairs of correlation between the dependent and the independent variables. In this table, we calculate the correlation between the Platform independent variables along with all its categories and compare it with the dependent variable TV Region. We will be comparing the independent variables Platform and TV Region with the dependent variable Gender. In order to determine the correlation, we make use of the significance value and the value for which should be less than 0.05 for both the independent variables As we can observe, the significance value for the Platform variable when compared with the Gender variable is 0.284 value is higher than 0.05, so we can say that both the variables are not significantly correlated. On the contrary, the Gender and the

TVRegion variable are significantly correlated as the significance value is 0.017 value is lower than 0.05. Lastly, the significance value for TVRegion and Platform variables is 0.581 which is greater than 0.05, so it reveals that both the variables are not significantly correlated.

Conclusions and Decisions:

The news consumption is a growing sector every day from various TV regions in the UK. There is an urgent requirement of an analytical model to understand how news consumption platforms fit into individual human life. In this, we use Binary logistic analysis using news consumption dataset using IBM SPSS version 26 and successfully examined and demonstrated results and show the relationship between Independent, dependent and Categorical variables.It shows the variance between dependant and Independent variables. Binary logistic regression provides significant statistical results, and interpretations indicate a holistic image of news networks with TV areas. Results of Hosmer and Lemeshow tests, correlations will boost the accuracy of news platforms. Also, it will allow us to define the types of channels frequently used together and the different demographic features associated with similar media consumption trends. Besides, It will help UK news channel authorities to add enhancement in the pattern of news consumption and boost business revenue. In the future, the analysis will offer new ways to access reliable news as compared to social media and change people's attitude towards news consumption.

Contribution of Team member for the project:

Project work is distributed among the team members equally; all team members are punctual and worked together with productive work morality. To complete the work within the time, we follow deadlines strictly and complete the individual's work.

Contribution::

Name	Contribution
Rahul Dhande	34%
Divya Thorat	33%
Shantanu Paranjape	33%

References:

Jmp.com. 2020. *Multiple Regression Residual Analysis And Outliers*. [online] Available at: https://www.jmp.com/en_us/statistics-knowledge-portal/what-is-multiple-regression/mlr-residual-analysis-and-outliers.html

Ezeudu, O., Ozoegwu, C. and Madu, C., 2019. A Statistical Regression Method for Characterization of Household Solid Waste: A Case Study of Awka Municipality in Nigeria. Recycling, 4(1), p.1.

Statistics.laerd.com. 2020. *How To Perform A Multiple Regression Analysis In SPSS Statistics I Laerd Statistics.* [online] Available at: https://statistics.laerd.com/spss-tutorials/multiple-regression-using-spss-statistics.php

1
Restore.ac.uk. 2020. *4.12 The SPSS Logistic Regression Output.* [online]Available at: https://www.restore.ac.uk/srme/www/fac/soc/wie/research-new/srme/modules/mod4/12/index.html#:~:text=The%20Omnibus%20Tests%20of%20Model,model%20and%20the%20new%20model.

7
Statistics Solutions. 2020. *The Logistic Regression Analysis In SPSS - Statistics Solutions.* [online]Available at: <https://www.statisticssolutions.com/the-logistic-regression-analysis-in-spss/> [Accessed 17 July 2020].

18
Field, A., 2018. *Discovering Statistics Using IBM SPSS Statistics*. London [etc.]: Sage.

RM_CA3

ORIGINALITY REPORT

13%	7%	3%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to National College of Ireland Student Paper	1%
2	trap.ncirl.ie Internet Source	1%
3	Submitted to International Islamic University Malaysia Student Paper	1%
4	Submitted to University College London Student Paper	1%
5	Submitted to OTHM Qualifications Student Paper	1%
6	digscholarship.unco.edu Internet Source	1%
7	Submitted to Swinburne University of Technology Student Paper	1%
8	Submitted to Brunel University Student Paper	1%

9	Submitted to iGroup Student Paper	<1%
10	Submitted to University of Surrey Student Paper	<1%
11	Submitted to University of Southampton Student Paper	<1%
12	Submitted to Brandenburgische Technische Universität Cottbus Student Paper	<1%
13	Submitted to Birkbeck College Student Paper	<1%
14	Submitted to Loughborough University Student Paper	<1%
15	Submitted to Christ College of Law Student Paper	<1%
16	www.mei.org.uk Internet Source	<1%
17	Alejandro Romero Perez de Tudela, Colin M. Rose, Julia A. Stegemann. "Quantification of material stocks in existing buildings using secondary data—A case study for timber in a London Borough", Resources, Conservation & Recycling: X, 2020 Publication	<1%

18	hdl.handle.net Internet Source	<1%
19	www.restore.ac.uk Internet Source	<1%
20	epdf.pub Internet Source	<1%
21	computerresearch.org Internet Source	<1%
22	Submitted to City University Student Paper	<1%
23	repository.up.ac.za Internet Source	<1%
24	statistics.laerd.com Internet Source	<1%
25	Submitted to University of Nottingham Student Paper	<1%
26	othes.univie.ac.at Internet Source	<1%
27	www.theseus.fi Internet Source	<1%
28	Submitted to University of Huddersfield Student Paper	<1%
29	Submitted to Iqra Uninversity, Gulshan Student Paper	

		<1%
30	Submitted to Eiffel Corporation Student Paper	<1%
31	Submitted to Institut Teknologi Brunei Student Paper	<1%
32	jcmel.swk.cuhk.edu.hk Internet Source	<1%
33	Submitted to Higher Education Commission Pakistan Student Paper	<1%
34	TETSUO MITSUHASHI. "Evaluation of stereoscopic picture quality with CFF", Ergonomics, 1996 Publication	<1%
35	Submitted to Middlesex University Student Paper	<1%
36	Submitted to University of Leicester Student Paper	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		