

# SAMPLE ASSIGNMENT – BUSINESS STATISTICS PROJECT HELP

---

Analysis of Secondary Data Using Minitab

This is a brief report based on the analysis of secondary data using Minitab. The report tries to analyze the performance of various Movie Chains and try to make recommendations for a new entrant.

## Contents

Introduction.....	2
Methodology .....	3
Context.....	3
Research design .....	3
Introduction of data.....	4
Introduction of Software .....	5
Statistical methods used.....	5
Statistical Analysis.....	5
Descriptive Statistics.....	6
Inferential Statistics .....	7
Hypothesis 1.....	7
Hypothesis 2.....	8
Establishing Empirical relations .....	9
Conclusion and Recommendation .....	11

## Introduction

Launching a different product line is a tricky affair. A lot of market insights have to be drawn before devising the product or service offerings. Some important insights include study of consumer behaviour, analysis of market potential, analysis of microenvironment, competitor analysis, and identification of key success factors. It should be noted that advanced data analysis plays an important role while drawing market insights. Data analysis also tests various hypothesis and gives strong statistical while accepting or rejecting a particular hypothesis.

In this case Film 2011 which is a large corporation specialising in the entertainment industry, wants to draw market insights about the entertainment industry as firm wants to build its own chain of cinemas UK wide. Therefore, a research has been conducted to assess the performance of other chains which are already running their business in United Kingdom. Therefore, the main of the report is summarized as follows

1. To use the statistical tools and framework for analysing data and thus drawing market insights
2. To take decisions based on the market insights drawn as a result of data analysis.

## Methodology

### Context

As already mentioned, Film 2011 is a large company in entertainment industry. However, it wants to expand its business and therefore interested in building a new chain of cinema. However, before starting a new business line, Film 2011, wants to conduct a market research and therefore data points related to 128 cinema halls has been collected. These cinema halls primarily belong to three major groups which is already operating in United Kingdom.

### Research design

The research is designed in such a way that objective of the research should be met (Uma Sekaran, and Roger Bougie, 2013). The aim of the research is to determine the market insights so that firm can launch chain of cinema in United Kingdom. Since, the nature of the research is descriptive in nature, therefore mixed method of research is used. There are primarily three types of research methods namely qualitative research method, quantitative research method, and mixed research method. Qualitative method is based on analysing a case with the help of logic and underlying theories. Therefore it is quite useful in cases when deeper analysis of a specific problem is required. However, qualitative analysis fails to establish any empirical relationships between dependent and explanatory variables; therefore result obtained by qualitative method could not be generalized. Moreover certain group of researcher do not even consider the Qualitative research as scientific (Saunders et al, 2012). On the other hand quantitative research is based on standard statistical framework, and therefore able

to test the research hypothesis with collected set of data, and produces empirical models which can be generalized (Creswell, 2014). However, there is a serious limitation of the quantitative method that once data is collected, it is very difficult to change the research question even slightly. Moreover the validity and reliability of research is limited to reliability of data. Mixed methodology of research uses the strengths of qualitative as well as quantitative method to formulate a formidable research methodology. However the major challenge faced by the researcher is to establish the synergy between the two (Creswell, 2014) i.e. how the research design has been formulated such that an optimal result can be derived.

In this case, though the data given for the purpose of the research is quite reliable but scope is quite broad i.e. to draw insights from market which includes the insights about competitors, insights about the consumers and so on. Therefore neither quantitative nor qualitative method of research suits to the needs. Therefore, mixed method of research is followed in which results of quantitative method of research is used to draw insights from the collected data. However, the results are validated against the theoretical frameworks and inferences are made. These inferences are used by management to take decisions regarding the launch of cinema chains in United Kingdom.

Therefore research design is formulated is described under the following points

- First of all collected data which is well coded is imported in the statistical package named MINITAB
- Descriptive analysis is conducted to determine basic insights about the data such as mean, median etc.
- Inferential statistics is conducted to test the research hypothesis and establish the empirical relations.
- The result of statics is validated against logical arguments and theoretical frameworks to make conclusions
- Finally, inferences are made which is used by management to take decision.

### **Introduction of data**

The survey data is coded in numerical forms and named under the following heads

NUMBER represents number of Cinema

CHAIN represents Cinema Chain where 1 for Chain A, 2 for Chain B, and 3 for Chain C

REGION represents Regional Location of the cinema where 1 for England, 2 for Scotland, and 3 for Wales

LOCATION represents Location of the cinema where 1 for In Town, and 2 for Out of Town

SIZE represents Size of the cinema where 1 for Small, 2 for Medium, and 3 for Large

GENDER represents gender of manager i.e. 1 for Male and 2 for Female

RENT represents Monthly Rent in (£00s)

DISTANCE represents distance of the cinema away from the High Street (miles).

ADVERT represents Monthly Advertising Expenditure in £.

WEEKEND represents Monthly weekend sales in £.

WEEKDAY represents Monthly weekday sales in £.

### **Introduction of Software**

The data has been analysed using standard statistical package named MINITAB. It has the provision of conducting both descriptive as well as inferential statistics. It has the facility to import data as well as data can be copy pasted in the worksheet. MINITAB work files get saved with an .MTW extension. It does not only gives text output but also has the capability to plot various graphs together, which makes representation quite easier.

### **Statistical methods used**

A set of statistical methods are used for the analysis of data. These methods include descriptive statistics as well as inferential statistics. Descriptive statistics include the basic analysis of data using mean, median, graphical analysis such as bar chart, pie charts etc. Inferential statistics include two sampled t-tests for unequal variances, and multivariate regression. It should be noted that that descriptive statistics introduces data and gives the basic property of data such as mean, median, normality, kurtosis etc. Moreover, descriptive statistics is used as input for inferential statistics. In this report two types of inferential statistics has been used. Two sampled t-test for unequal variances are used to test the hypothesis. Two sampled t-test is used to determine if two population means are equal. However t-tests fail to establish any empirical relation between dependent and explanatory variable. Therefore multivariate regression is used to determine empirical relation between dependent and explanatory variables. Explanatory variables are also known as independent variable.

### **Missing Data**

A set of values are missing, 2 missing values are found in gender, 1 missing value is in size and 2 missing values is in distance. While conducting the analysis, these values have been ignored. It is not possible to assign average values as there are only two set of values for available for gender and size. Distance is in absolute terms but putting average value does not make any sense. Therefore, all missing values have been removed from data set. Since only 5 data points have been removed so it will not affect the data much.

## Statistical Analysis

As described above both descriptive and inferential statistical analysis is used. Initially basic charts are used for describing the data.

### Descriptive Statistics

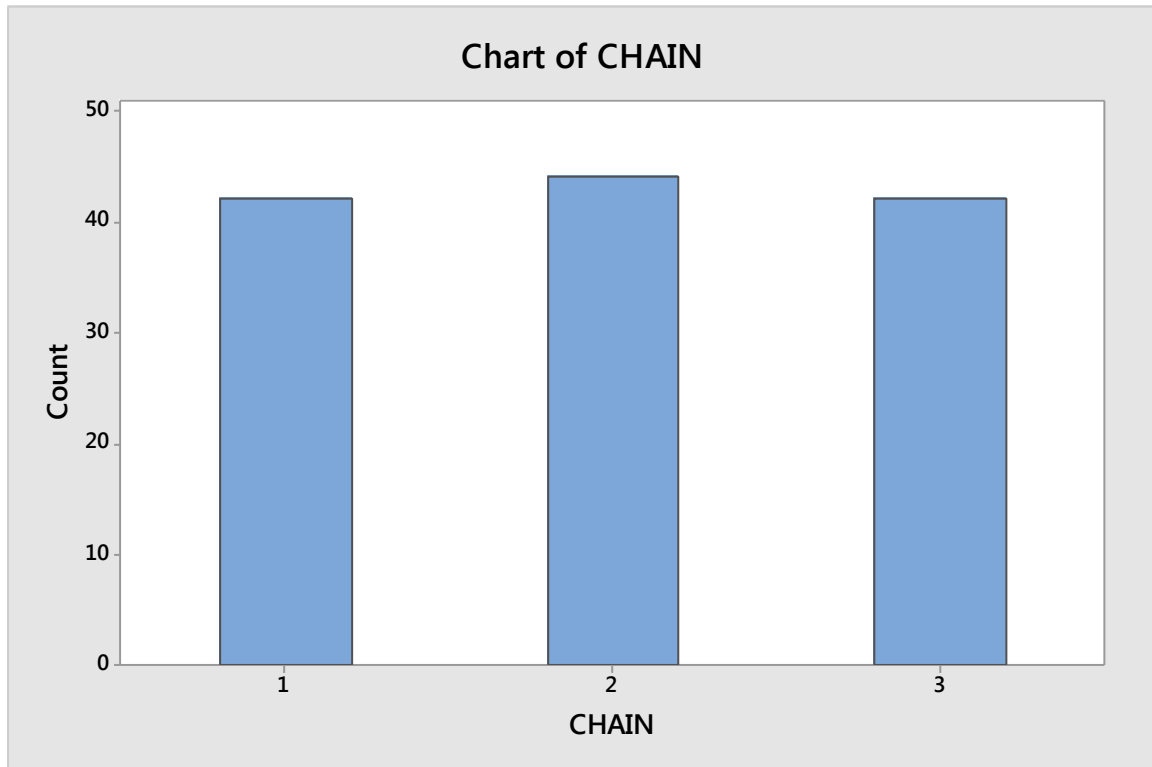
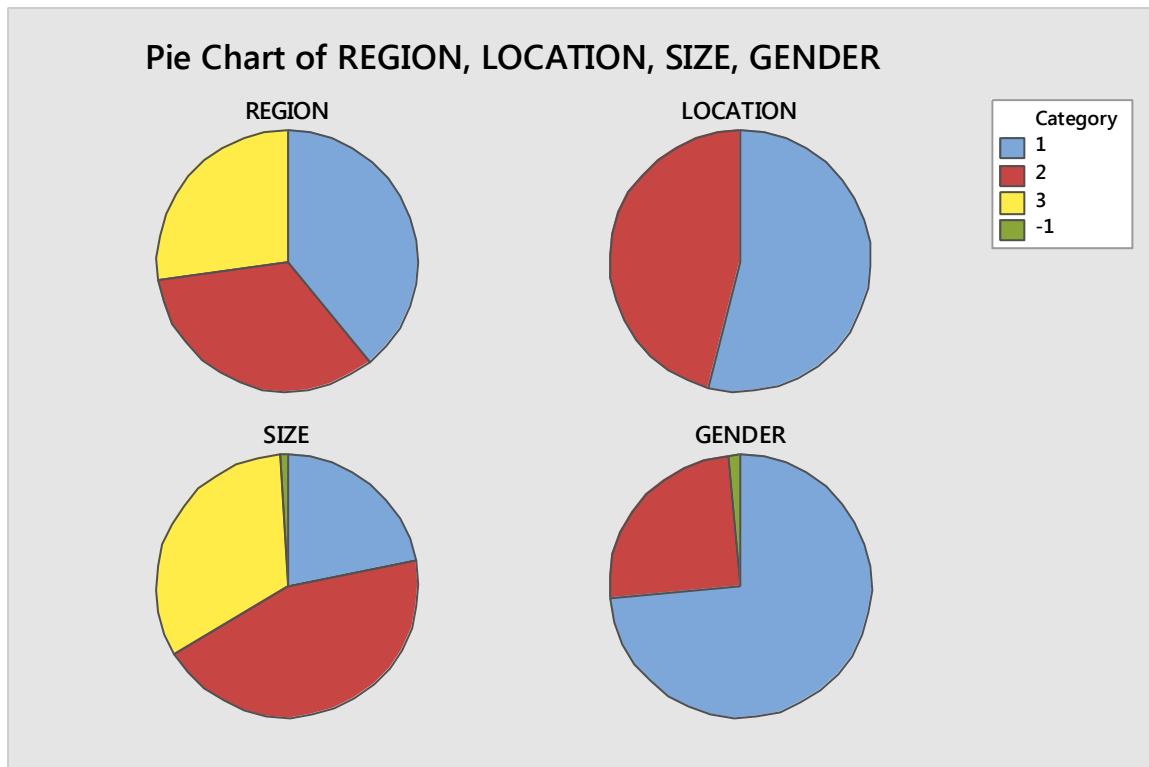


Fig1 : Representation of types of chain

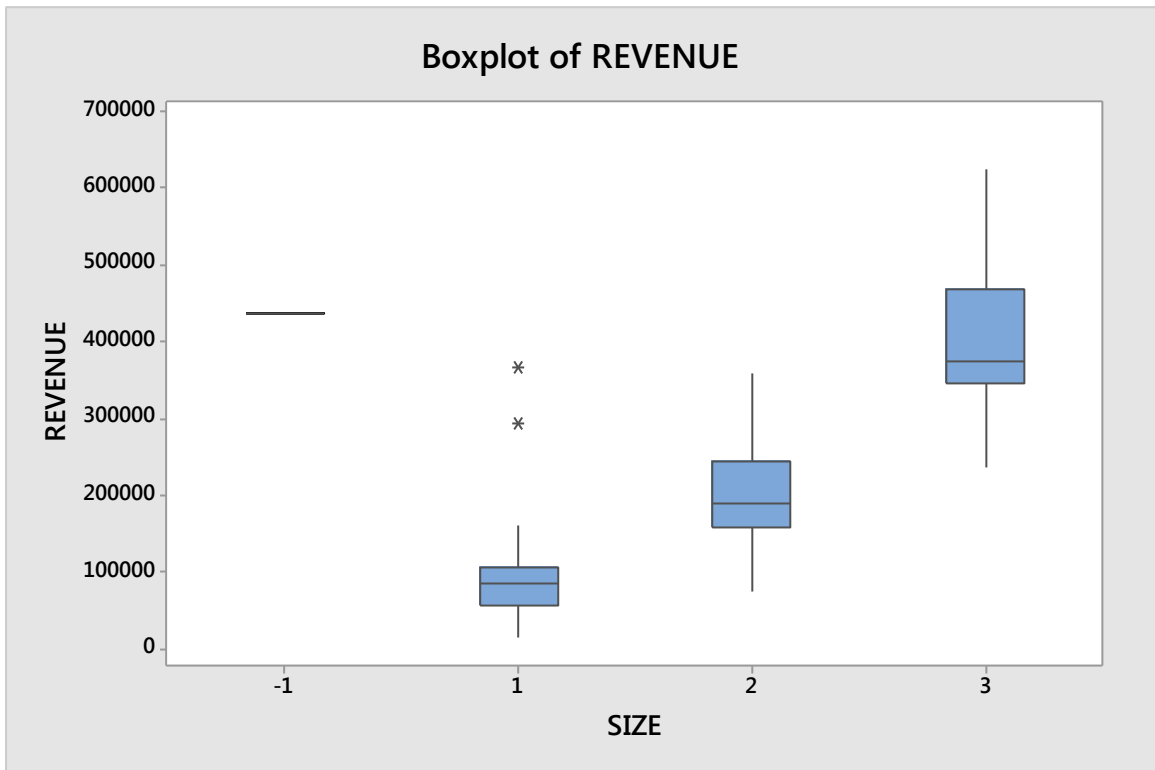
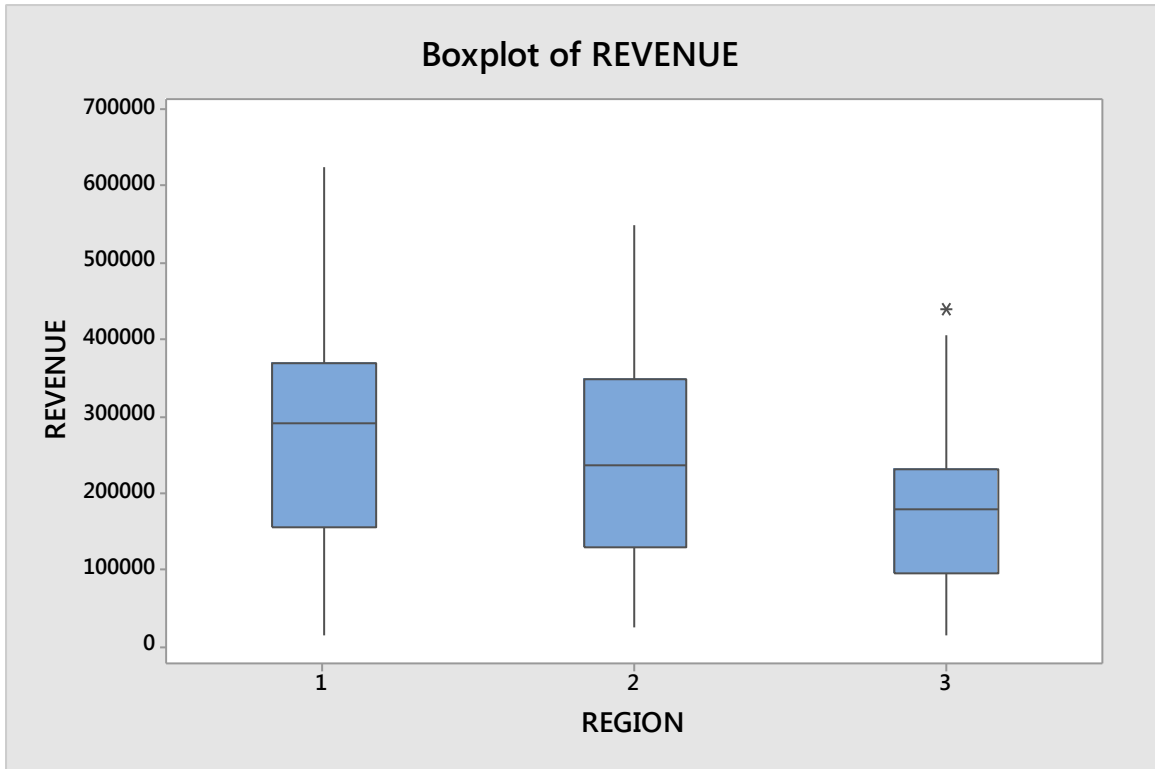
Bar chart suggests that sample contains more than 40 for all three types of chain. Therefore, the sample is quite balanced as far types of movie chain are concerned.



Four pie charts have been drawn. First graph is based on the number of regions; second graph is based on location, third graph is based on size, and fourth graph is based on Gender. In the section introduction of data, it has been shown that -1 does not represent anything, therefore -1 in the data file is considered as mistake and hence considered as 1.

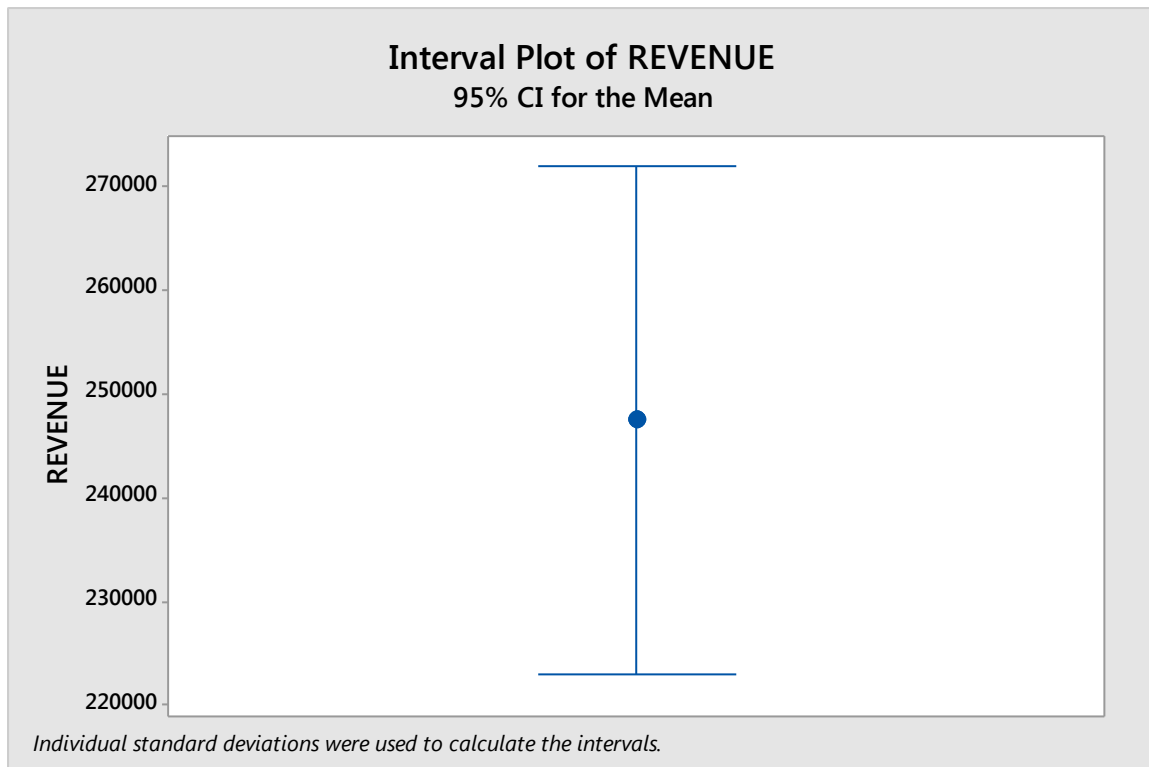
#### Tally for Location and Size

Location	Count	Size	Count
In town	69	Large	42
Out of Town	59	Medium	57
N=	128	Small	28
		N=	127
		*=	1



**Interval Plot of Revenue**





The above figure explains that there is 95% chance that the average total monthly sale is between £220,000 - £270,000

### Inferential Statistics

A set of hypothesis have been formulated for the research which has been mentioned below:

Hypothesis 1: There is no significant difference between profits of Cinema Halls run by male managers and Cinema Halls run by female managers

Hypothesis 2: There is no significant difference between profits of Cinema Halls located in the town and Cinema Halls located outside the town.

In addition to above hypothesis an empirical relation has been established i.e. how total revenue depends upon the expenses on advertisement and distance

Total Revenue = f (Expenses incurred on Advertisement, Distance)

However, two information needs to be drawn from existing data set. First is total revenue which is obtained by adding revenue obtained during weekdays and revenue obtained during weekends, second is net profit which is obtained by subtracting rent and advertising expenses from total revenue.

#### Hypothesis 1

H<sub>0</sub>:  $\mu$  (Profits of Cinema Halls run by male managers) =  $\mu$  (Profits of Cinema Halls run by female managers)

H<sub>a</sub>:  $\mu$  (Profits of Cinema Halls run by male managers)  $\neq$   $\mu$  (Profits of Cinema Halls run by male managers)

Independent sample test for unequal variances with significance level 0.05 is used for testing the hypothesis. Below table has been obtained after conducting independent sample t-test in Minitab

### Two-Sample T-Test and CI: PROFIT\_Male, PROFIT\_Female

Two-sample T for PROFIT\_Male vs PROFIT\_Female

	N	Mean	StDev	SE Mean
PROFIT_Male	96	266283	135826	13863
PROFIT_Female	32	173448	129417	22878

Difference =  $\mu$  (PROFIT\_Male) -  $\mu$  (PROFIT\_Female)

Estimate for difference: 92835

95% CI for difference: (39226, 146444)

T-Test of difference = 0 (vs  $\neq$ ): T-Value = 3.47 P-Value = 0.001 DF = 55

**Interpretation:** The p value for the test is 0.001 which is less than 0.05, therefore there is enough statistical evidence to reject null hypothesis. Therefore alternate hypothesis holds good i.e.  $\mu$  (Profits of Cinema Halls run by male managers)  $\neq$   $\mu$  (Profits of Cinema Halls run by female managers). Therefore there is significant statistical to reject the hypothesis that there is no significant difference between profits of Cinema Halls run by male managers and Cinema Halls run by female managers.

### Hypothesis 2

H0:  $\mu$  (Profits of Cinema Halls run in the town) =  $\mu$  (Profits of Cinema Halls run outside the town)

Ha:  $\mu$  (Profits of Cinema Halls run in the town)  $\neq$   $\mu$  (Profits of Cinema Halls run outside the town)

Independent sample test for unequal variances with significance level 0.05 is used for testing the hypothesis. Below table has been obtained after conducting independent sample t-test in Minitab

### Two-Sample T-Test and CI: PROFIT\_ITOWN, PROFIT\_OTOWN

Two-sample T for PROFIT\_ITOWN vs PROFIT\_OTOWN

	N	Mean	StDev	SE Mean
PROFIT_ITOWN	69	254273	141476	17032
PROFIT_OTOWN	59	229977	137645	17920

Difference =  $\mu$  (PROFIT\_ITOWN) -  $\mu$  (PROFIT\_OTOWN)

Estimate for difference: 24296

95% CI for difference: (-24640, 73233)

T-Test of difference = 0 (vs  $\neq$ ): T-Value = 0.98 P-Value = 0.328 DF = 123

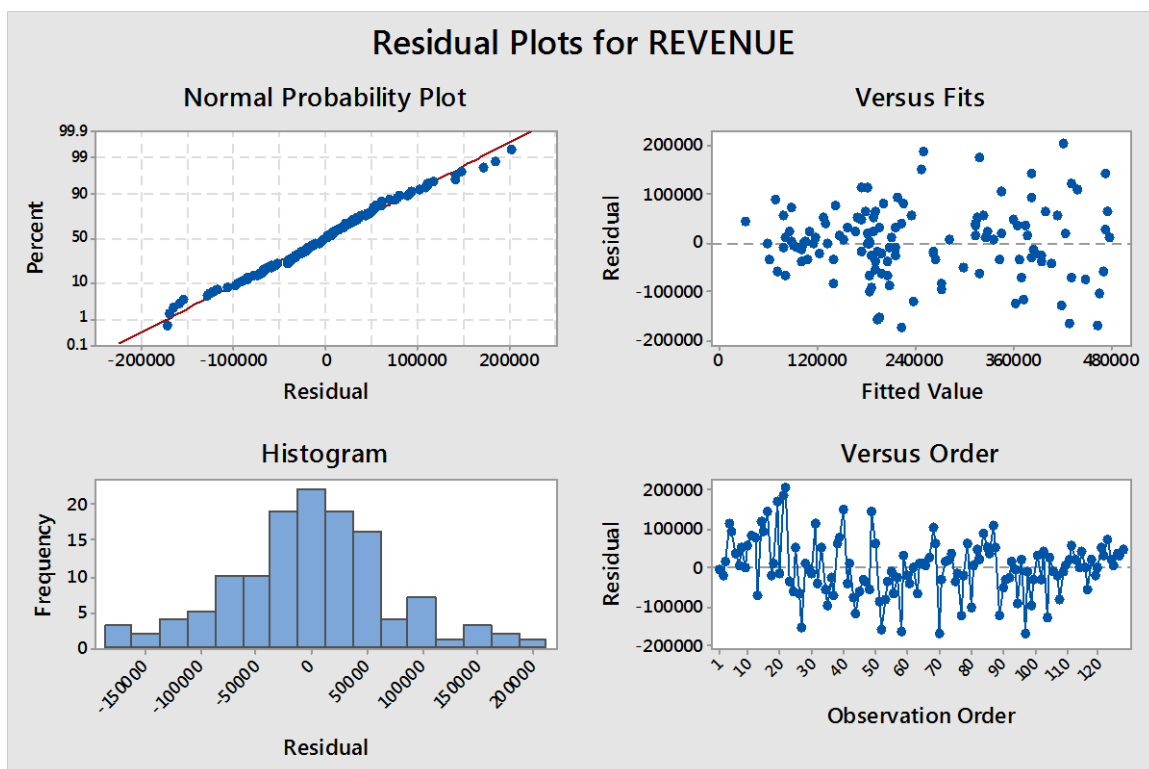
**Interpretation:** The p value for the test is 0.328 which is more than 0.05, therefore there is not enough statistical evidence to reject null hypothesis. Therefore alternate hypothesis does not hold good i.e.  $\mu$  (Profits of Cinema Halls run by male managers) =  $\mu$  (Profits of Cinema Halls run by female managers).

Therefore there is not significant statistical to reject the hypothesis that there is no significant difference between profits of Cinema Halls located in the town and Cinema Halls located outside the town

### Establishing Empirical relations

An empirical relationship has been established between the total revenue and expenses incurred on advertisement and distance in kilometers. Once the relationship has been established the empirical relationship was reestablished using location as control Variable.

### Empirical Relationships



### Regression Analysis: REVENUE versus DISTANCE, ADVERT

Analysis of Variance

Source	DF	Adj SS	AdjMS	F-Value	P-Value
Regression	2	1.83644E+12	9.18219E+11	171.44	0.000
DISTANCE	1	8904575778	8904575778	1.66	0.200
ADVERT	1	1.74646E+12	1.74646E+12	326.09	0.000
Error	125	6.69472E+11	5355775020		
Lack-of-Fit	121	6.54203E+11	5406634321	1.42	0.411
Pure Error	4	15269124652	3817281163		
Total	127	2.50591E+12			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
73183.2	73.28%	72.86%	72.06%

#### Coefficients

Term	Coef	SECoef	T-Value	P-Value	VIF
Constant	-143447	26345	-5.44	0.000	
DISTANCE	-5052	3918	-1.29	0.200	1.02
ADVERT	176.34	9.77	18.06	0.000	1.02

#### Regression Equation

REVENUE = -143447 - 5052 DISTANCE + 176.34 ADVERT

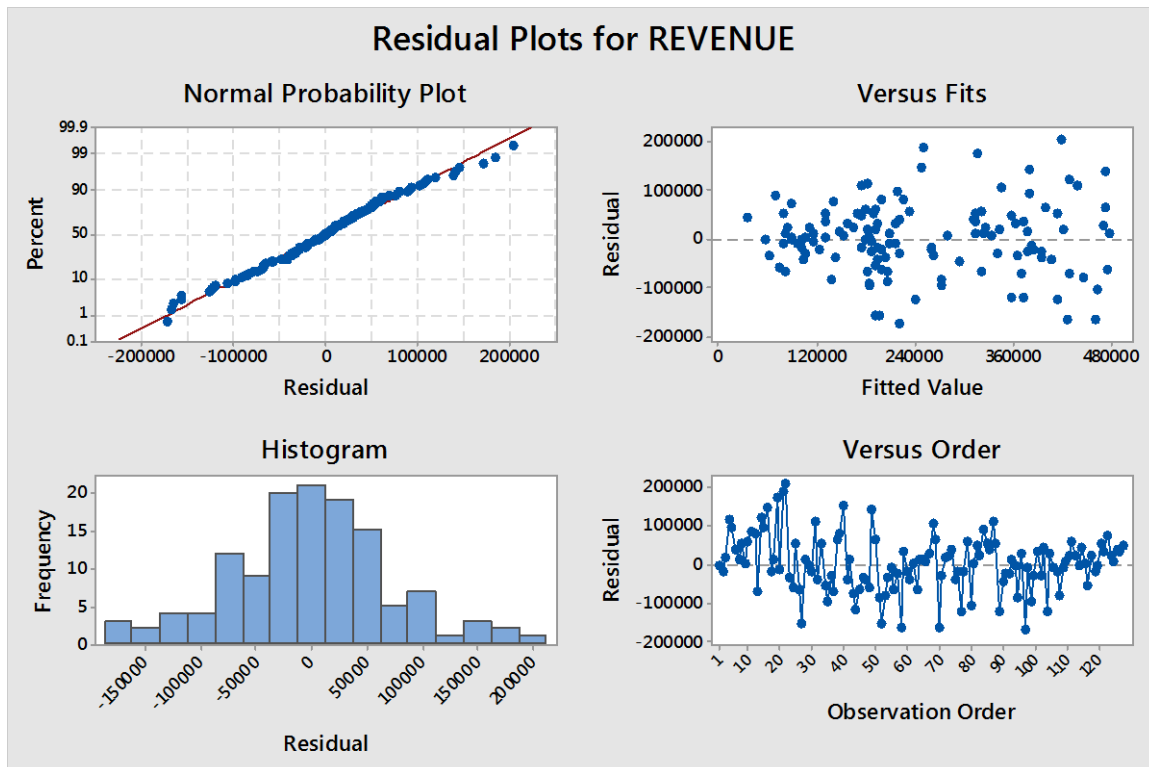
#### Fits and Diagnostics for Unusual Observations

Obs	REVENUE	Fit	ResidStdResid	
19	489649	318046	171603	2.36 R
21	436429	250798	185631	2.56 R
22	624896	421087	203809	2.82 R
27	39911	195845	-155934	-2.14 R
40	395664	247824	147840	2.04 R
52	35065	194082	-159017	-2.18 R
58	262801	429142	-166341	-2.30 R
70	293700	463147	-169447	-2.35 R
97	49424	223083	-173659	-2.39 R

**Interpretation:** The result of regression analysis is quite insightful. R-square of the regression is 73.28%. However it is a multivariate regression therefore, adjusted R-square is better measure to determine the quality of regression. It has been argued that as number of explanatory variables increases value of R-square also increases even if newly added variable might not have very high explanatory power. However, adjusted r-square consider these factors. The adjusted R-square is 72.86%. It indicates that two explanatory variable his actually trying to describe the changes in dependent variable. It also means that 72.86% changes in dependent variable is explained by explanatory variables.

Since the empirical relation is  $REVENUE = -143447 - 5052 DISTANCE + 176.34 ADVERT$  therefore it can be inferred that distance is negatively related to revenue while expenses incurred on advertisement is positively related with Revenue. P value of Ordinary Least Square regression is lower than 0.05 except for the coefficients of DISTANCE. It means explanatory power of distance is comparatively lower.

#### Use of Control variable



Categorical predictor coding (1, 0)

#### Analysis of Variance

Source	DF	Adj SS	AdjMS	F-Value	P-Value
Regression	3	1.83676E+12	6.12253E+11	113.46	0.000
DISTANCE	1	6016594465	6016594465	1.11	0.293
ADVERT	1	1.73419E+12	1.73419E+12	321.36	0.000
LOCATION	1	320377834	320377834	0.06	0.808
Error	124	6.69151E+11	5396383062		
Lack-of-Fit	120	6.53882E+11	5449019792	1.43	0.407
Pure Error	4	15269124652	3817281163		
Total	127	2.50591E+12			

#### Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
73460.1	73.30%	72.65%	71.62%

#### Coefficients

Term	Coef	SECoef	T-Value	P-Value	VIF
Constant	-142730	26608	-5.36	0.000	
DISTANCE	-6075	5754	-1.06	0.293	2.19
ADVERT	176.17	9.83	17.93	0.000	1.03
LOCATION					
2	4653	19095	0.24	0.808	2.15

#### Regression Equation

LOCATION

1 REVENUE = -142730 - 6075 DISTANCE + 176.17 ADVERT

2 REVENUE = -138077 - 6075 DISTANCE + 176.17 ADVERT

### Fits and Diagnostics for Unusual Observations

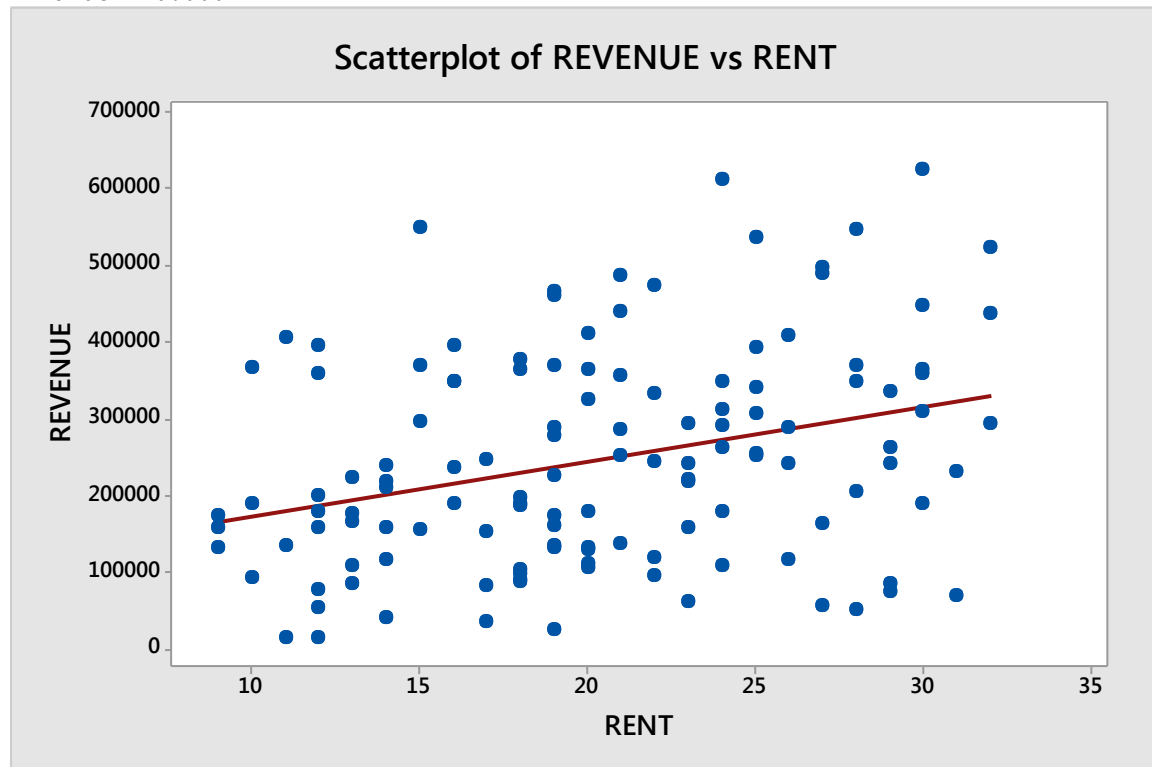
Obs	REVENUE	Fit	Resid	StdResid	
19	489649	316774	172875	2.37	R
21	436429	250619	185810	2.55	R
22	624896	420229	204667	2.83	R
27	39911	197801	-157890	-2.17	R
36	189632	222103	-32471	-0.48	X
40	395664	248465	147199	2.02	R
48	411056	475790	-64734	-0.96	X
52	35065	191387	-156322	-2.16	R
58	262801	427762	-164961	-2.28	R
70	293700	461477	-167777	-2.33	R
91	349759	377824	-28065	-0.40	X
97	49424	222674	-173250	-2.38	R

**Interpretation:** When location has been used as control variable, then coefficients of explanatory variables have not changed. However the constant has been lower if location is in the town.

### Total Revenue and Rent

A Pearson Correlation between total revenue and rent has been established.

Pearson correlation of RENT and REVENUE = 0.312  
P-Value = 0.000



**Interpretation:** It indicates that there exist a positive weak correlation between total revenue generated by cinema hall and rent it paid. It suggests if rent increases by 1£ revenue increases by 0.312£

### Conclusion and Recommendation

Based on the result of the data analysis following recommendations are made

- There is significant difference between profits of Cinema Halls run by male managers and Cinema Halls run by female managers
- There is no significant difference between profits of Cinema Halls located in the town and Cinema Halls located outside the town
- Distance is negatively related to revenue while expenses incurred on advertisement are positively related with Revenue.

Based on the results obtained as a result of data analysis, it should be recommended that outside the town or inside the town does not affect the profit generated from Cinema Hall, however distance from High Street affects the revenue. Moreover, increased expenses on advertising also increase the revenue. However, there is no logical arguments behind the finding that there is a difference in profit generated by male and female managers.