CA660

ASSINGNMET-1

1.NAME-YASHASWI VERMA

S.NO-19211007

2.BHARGAVA RAMA RAJU DANDDU

S.NO-19210189

Statistical Data Analysis On Census Data of Ireland 1981-2011

Abstract- The report discusses the population of Ireland, along with the survey result of some year. The report is discussing various types of statistical tools and techniques to find out the research objectives. The objective of the research is to analyze the population change in the different regions of Ireland and how it will impact on different places in Ireland. To represent the descriptive analysis, Excel statistical tool is used along with its statistical summary functions. By using this report has been shown the mean, median, mode, along with other statistical calculations for each year survey result. In addition, regression analysis is used to find out the relationship between populations of different places over the year. The overall findings and results are represented using some graphical representation techniques such as bar graph, scatter plot. A pivot table is used to show a summarise of the total population for each place. Finally, the report has been critically analyzed the change in population along with its impact by using statistical methods.

Keywords- Descriptive method, regression analysis, population of Ireland, statistical tools.

I.INTRODUCTION

1.1 Background

The report is about to discuss the census of Ireland based on some respective year. One dataset is chosen over here in order to do some analysis on the population of different places in Ireland. The dataset contains some numerical data along with some categorical data. The numerical data indicates the number of population of each place within Ireland, and categorical data indicates the name of the place and code number of that particular place. The dataset includes the census of the year 1981, year 1986, year 1991, year 1996, year 2002, year 2006 and year 2011. It is observed in the dataset that the population or census is varied on the basis of places. The report is also discussing some analysis report using different types of statistical methods such as regression analysis, some inferential statistical and descriptive methods. After analysis, the findings and results of the analysis are represented using some graphical representation methods like graphs, charts, and more.

1.2 Aim

The aim of the report is to assess the change in population at different places in Ireland by using some statistical methods.

1.3.Objective

The objectives of the report are as follows:

- To analyze the change of population in Ireland
- To analyze the impact of the population in different places in Ireland

1.4 Related Work

In this section, the report will discuss various types of analysis by using two types of statistical methods, such as a descriptive statistical method along with an inferential statistical method[1]. The descriptive statistical method represents the detailed description of the overall population; on the other hand, the inferential method is used to show details of a particular sample. In this case, the

sample size is 42, and the overall analysis is done based on this sample. The descriptive analysis consists of mean, median, mode, standard deviation, and other statistical summaries that are represented using excel along with statistical tools and techniques[2]. In the case of inferential method, various types of regression analysis, t-test analysis, and other analysis. In the analysis section, the detailed description of every analysis report is described along with its elaborate description.

2.Dataset and exploratory analysis

2.1 Descriptive analysis

Census_2011	
Mean	6501.214286
Standard Error	947.9292595
Median	5366.5
Mode	#N/A
Standard Deviation	6143.283731
Sample Variance	37739935
Kurtosis	12.41741715
Skewness	2.94944933
Range	35304
Minimum	666
Maximum	35970
Sum	273051
Count	42
Confidence Level(95.0%)	1914.381956

Table 1: summary of Census_2011

(Source: created by learner)

	Census-2006
Mean	5714.095238
Standard Error	849.996713
Median	4880.5
Mode	#N/A
Standard Deviation	5508.608291
Sample Variance	30344765.31
Kurtosis	12.83636314
Skewness	3.01209763
Range	31663
Minimum	625
Maximum	32288
Sum	239992
Count	42
Confidence Level(95.0%)	1716.603168

Table 2: summary of Census_2006

The above-mentioned two descriptive analyses are showing the different types of statistical calculations for Census-2011 and census-2006. Similarly, by using descriptive statistics, it is possible to find out the descriptive statistics of other Census for different places.

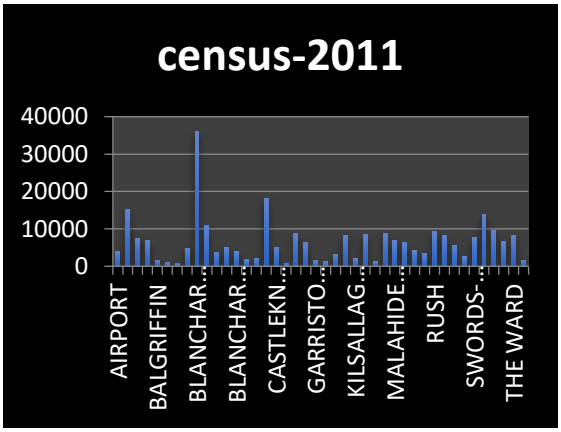

Figure 1: bar graph to represent the total population

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.962564
R Square	0.92653
Adjusted R Square	0.924693
Standard Error	1685.845
Observations	42

Table 3: regression analysis between census 2011 and census 2006

Regression analysis output: Summary Output

The above analysis shows the practical value which we achieved after analysis. The first value is **Multiple R**, which is 0.9625 this means the linear relationship strength is strong as the value close to 1. The Second value is **R square**, which is 0.92653 this indicates the indicator of goodness of fit. It means 92% of the dependent variable y(2011) is explained by the independent variable x(2006). The third value is **adjusted R**, which is 0.9246 this value can be used instead of R square if we do multiple regression. The fourth value is **Standard error**, which shows the precision of your regression analysis.

Regression analysis output: ANOVA

After the summary output , the next step is ANOVA, which showed important value to analyze the data. The **Significance F** value gives an idea of how reliable (statistically significant) the result is. As Significance F is less than 0.05 (5%) i.e., 0.000273, our model is OK.

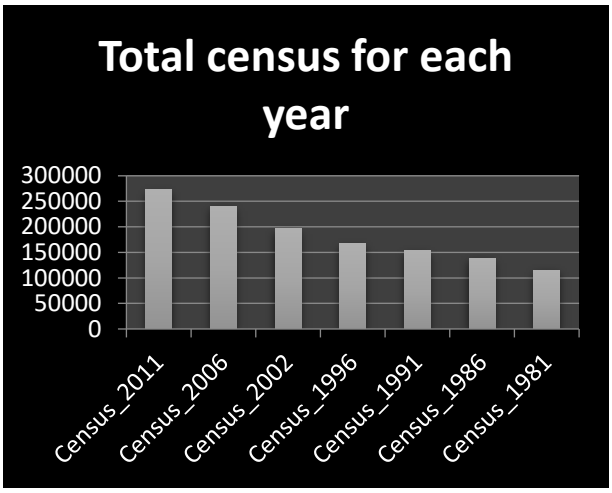

Figure 2: bar graph to represent the total population based on year

3.Research questions

Q1. What is the change of population in Ireland?

Q2. What is the impact of population in different places in Ireland?

4.Methods used and why

In this report, some descriptive and inferential statistics are used to analyze the two research questions and reach the objectives of the research. In this case, descriptive statistics are used to display the mean, median mode, and other statistical summaries of the collected dataset. In the case of inferential statistics, regression analysis and the pivot table are used. The main advantage of regression analysis is to find out the relationship between the selected variables from the sample, and it also shows how one variable depends on one or more variables. For doing this, ANOVA testing is used. The most popular statistical tool that is Excel is used in order

for this analysis part. Another advantage of ANOVA testing is to find out r-value along with the p-value and correlation coefficient of independent variables. Based on this value, it is possible to find out proper dependency among the variables. On the other hand, a pivot table is used to represent the summarize result, and based on this summarise result. The bar graph is designed for each place in Ireland.

5.Result and findings

From the above-mentioned analysis, it is observed that various types of analysis are used in order to meet the objection of this report. The report shows how the population is changed from one place to another place, and all these findings are represented using some statistical analysis [6]. Using descriptive statistics analysis is done for each year census. Here the sample size is 42 and based on sample size, and other analysis is done such as regression analysis, analysis using a pivot table showing total numbers population for each place [7]. Apart from this, a bar graph is used to show the graphical representation of how the Census is varied for each year according to the places in Ireland. Besides this, regression analysis is used to represent the relationship between two variables [3]. In this case, two variables are considered one is an independent variable, dependent variable. In addition, the main functionality of regression analysis is to describe how a dependent variable depends on independent variables. From the analysis report, it is also understood that how the population makes impacts on different places in Ireland. Table 5 is showing the overall analysis report that is done by ANOVA testing using EXCEL. In addition, the table represents the correlation coefficient of independent variables along with r-value and p-value. Those are used to determine the relationship between the independent variable and the dependent variable. Mean, median mode, and the standard deviation is represented based on this selected data sample [5]. Mean is considered as the average population of each year; the standard deviation is showing the difference between two groups of the mean for the selected dataset.

6.Conclusion

Considering the above-mentioned all those observations and analysis report, it is concluded that the population of Ireland is varied from one place to another place. In order to do this analysis, the report considered various types of aspects. Different types of statistical methods are used for finding the analysis report, and the findings are represented using some graphical tools of statistics. From this analysis report, it is also stated that how the population is changed according to different places in Ireland along with the impacts of population on Ireland along with different places, and this done based on collected samples for some respective years.

References

[1]. Bradley, Michael T., and Andrew Brand. "Accuracy when inferential statistics are used as measurement tools." *BMC research notes* 9, no. 1 (2016): 241.

[2]. Colorafi, Karen Jiggins, and Bronwynne Evans. "Qualitative descriptive methods in health science research." *HERD: Health Environments Research & Design Journal* 9, no. 4 (2016): 16-25.

[3]. Hoshmand, Reza. Statistical methods for environmental and agricultural sciences. CRC Press, 2017.

[4]. https://data.gov.ie/organization/central-statistics-office

[5]. Manly, Bryan FJ, and Jorge A. Navarro Alberto. *Multivariate statistical methods: a primer*. Chapman and Hall/CRC, 2016.

[6]. Schabenberger, Oliver, and Carol A. Gotway. *Statistical methods for spatial data analysis*. Chapman and Hall/CRC, 2017.

[7]. Trafimow, David, and Justin A. MacDonald. "Performing inferential statistics prior to data collection." *Educational and psychological measurement* 77, no. 2 (2017): 204-219.