Advanced Business Data Analysis

CA2 – Report

ECO-FRIENDLY TRANSPORT IN IRELAND

A comparison of cities in terms of eco-friendly methods of transport to work, school and college in Ireland.

Student number – 17125286

Words count - 1478

Table of Contents

[Introduction 3](#_Toc511571869)

[1. Eco– friendly transport in 3 parts of Ireland (3 planning region). 3](#_Toc511571870)

[1.1 Descriptive Statistics 3](#_Toc511571871)

[1.2 Assessing the normality 4](#_Toc511571872)

[1.3 Comparison of South to West and North of Ireland in terms of eco-friendly transport to work, school or college - Wilcoxon Signed Rank Test 10](#_Toc511571873)

[2. Dublin’s attitude to environmentally friendly transport to get to work, school or college. 11](#_Toc511571874)

[2.1 Descriptive statistic 12](#_Toc511571875)

[2.2 Assessing the normality 14](#_Toc511571876)

[2.3 Mann-Whitney U Test and to compare Fingal and South Dublin in terms of environmental friendly transport to work, school or college. 20](#_Toc511571877)

[3. Comparison of Cork, Galway and Limerick City in terms of environmental friendly transport 21](#_Toc511571878)

[3.1 Descriptive statistics 21](#_Toc511571879)

[3.2 Assessing the normality 23](#_Toc511571881)

[3.3 Kruskal-Wallis H Test to compare Cork, Galway and Limerick City 32](#_Toc511571882)

[Conclusions 34](#_Toc511571883)

[Bibliography 34](#_Toc511571884)

## Introduction

Project is made for ‘Advanced Business Data Analysis’ module and it consist separate statistical tests:

* Is there a difference in using eco – friendly transport in 3 parts of Ireland.
* Dublin’s attitude to environmentally friendly transport to get to work, school or college.
* Comparison of Cork, Galway and Limerick City in terms of environmental friendly transport

Dataset is extracted from census 2011 made by Central Statistics Office and it is a "Small Area Population Statistics" downloaded from http://airo.maynoothuniversity.ie [[1]](#footnote-1)

By environmentally friendly we assume transport by: foot, bicycle, bus, couch, minibus, Luas, train and Dart.

Environment: Microsoft Excel, SPSS and R studio.

## Eco– friendly transport in 3 parts of Ireland (3 planning region).

In that test we will compare 3 parts of Ireland (planning region as per Census 2011). Comparison is based on sample of answers (randomly selected 3801 samples) of electorat division and their percentage of population using eco friendly transport.

* East and Midlands
* North and West
* South

### Descriptive Statistics



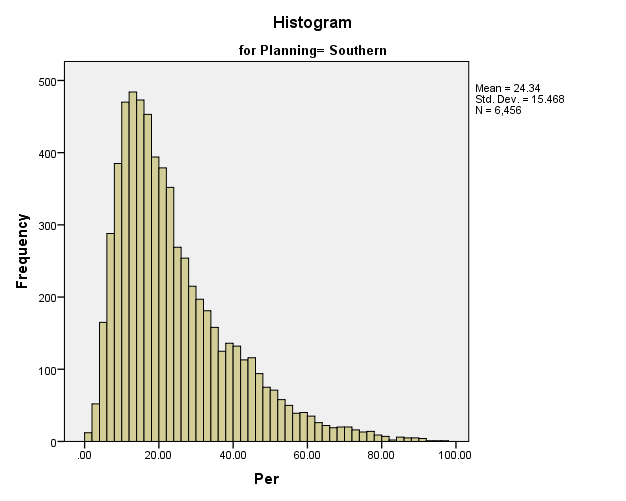
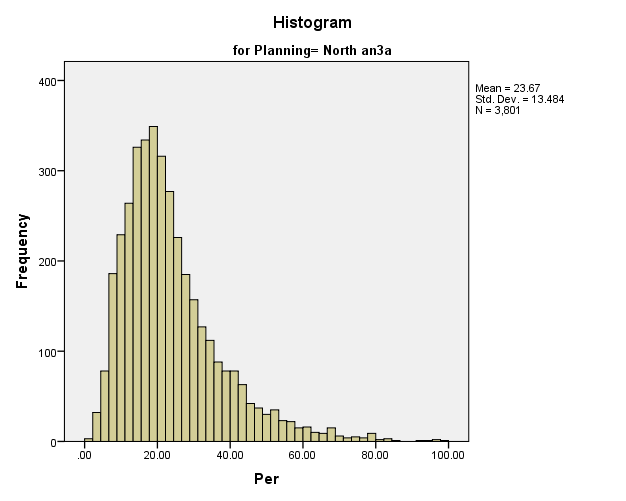
### Assessing the normality

1. Hypotheses

H0: The sample data are normally distributed

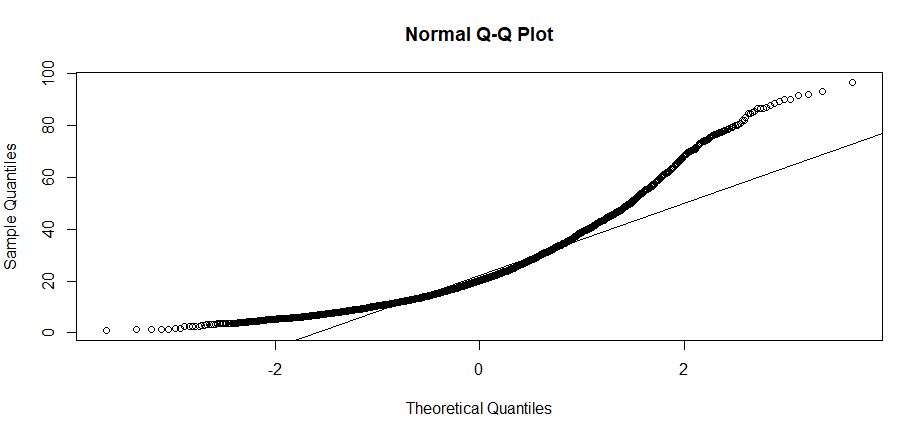
H1: The sample data are not normally distributed

1. Histograms

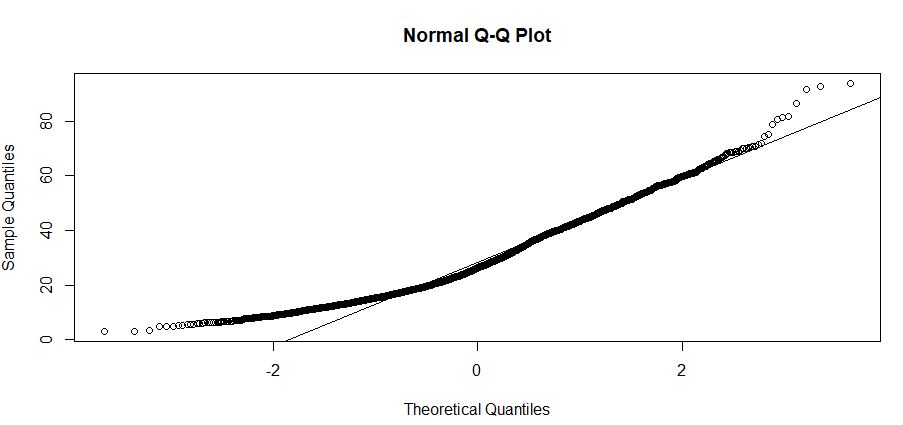


1. Q- Q plots

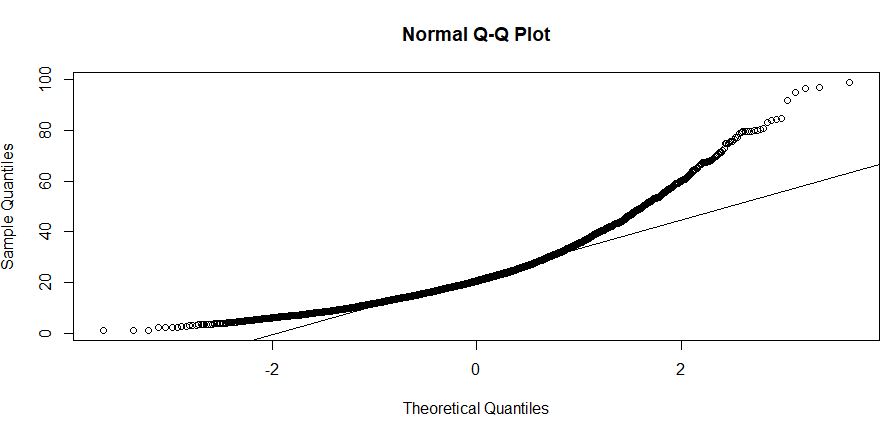
South:

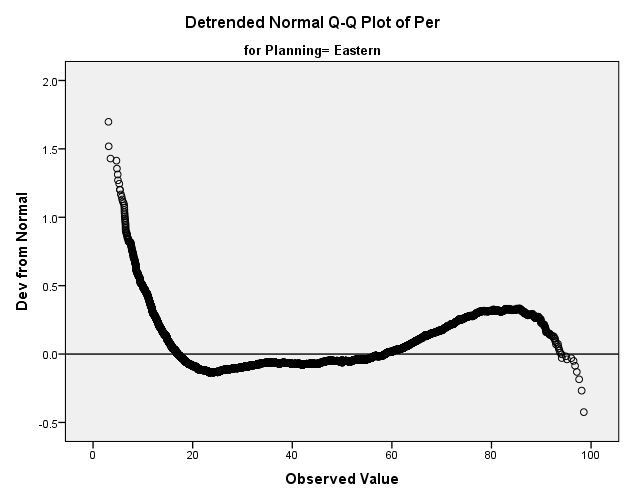
East 

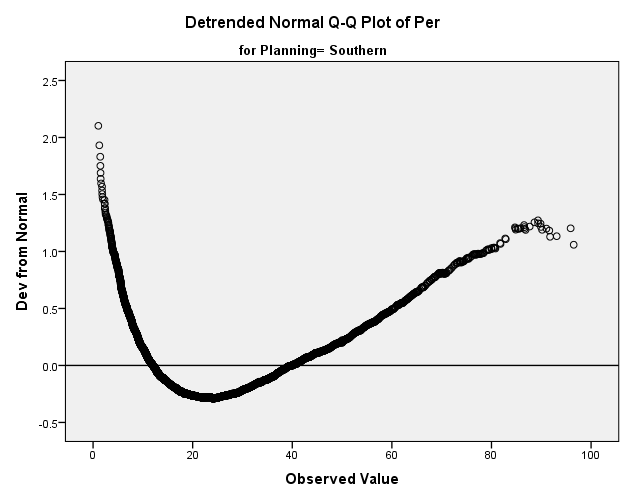
East and Midlands

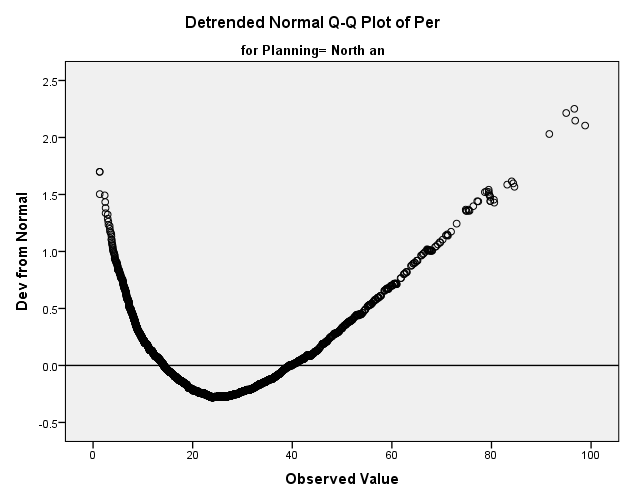


West and North:

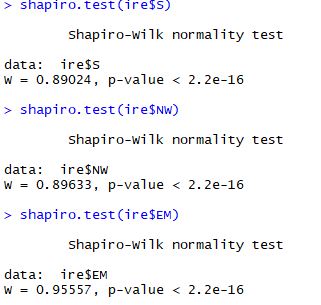








1. Kolmogorov-Smirnov and Shapiro-Wilk



1. Determine p value

p=0.05, p values less than 0.05 mean data are NOT normal

p values greater than 0.05 mean data are normal

1. Report

S – South P<0.05 – distribution of data are not normal

NW – North West P<0.05 – distribution of data are not normal P<0.05

EM – East and Midlands P<0.05 – distribution of data are not normal

1. Conclusions

Mean – Median – Mode and box plot comparison – We can see that East and Midland have totally different mean and median when compare to other parts of Ireland. In all cases we have big difference between mean and median that indicates not normal distribution of data.

Histogram and Q=Q plots confirmed what we could see from Mean median mode comparison. Data is not normal,

Last Shapiro test confirmed also that data is not normal in all cases as P was less than 0.05.

### Comparison of South to West and North of Ireland in terms of eco-friendly transport to work, school or college - Wilcoxon Signed Rank Test

Justification: Wilcoxon Signed Rank Test because as data distribution is not normal we need to use non parametric test. We excluded East and Midlands as per examination of descriptive statistics we can see that mean and median is drastically higher then in the other two parts of Ireland. We already know that there is a difference in that part, but is there a difference between other parts of Ireland?

1. Hypotheses

Null Hypothesis

H0: M South = M North+West

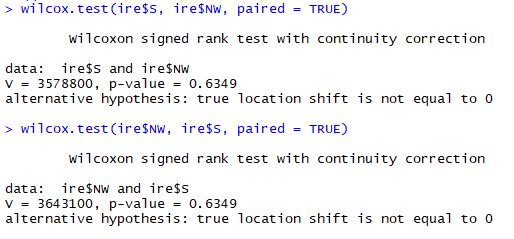
The median difference between percentage population using eco-friendly transport in observed parts of Ireland is zero

Alternate Hypothesis

H1: M South ≠ M North+West

The median difference between percentage population using eco-friendly transport in observed parts of Ireland is not zero

1. Test



1. Determine p value

p=0.05, p values less than 0.05 mean data are NOT normal

p values greater than 0.05 mean data are normal

1. Critical value (α = 0.05) – As 95 % sure that we wont make type I error is enough when comparing eco friendly transport.
2. Report

Report: (V = 36, p = 0.6349)

Wilcoxon Rank Test was run to determine if there were differences between percentage population using eco-friendly transport in South and West plus North part of Ireland.

Distributions of the percentage population using eco-friendly transport were similar both parts of Ireland, as assessed by visual inspection.

Percentage of population using desired transport type was not statistically significantly different between South(Mdn = 20.19) and West plus North (Mdn = 20.92),   
V = 36, p = .6349

1. Conclusion

We cannot reject H0, as p = 0.6349, p>0.05. so median difference between percentage population using eco-friendly transport in South and West plus North part of Ireland is zero. That’s mean that there is no difference between these two parts of Ireland.

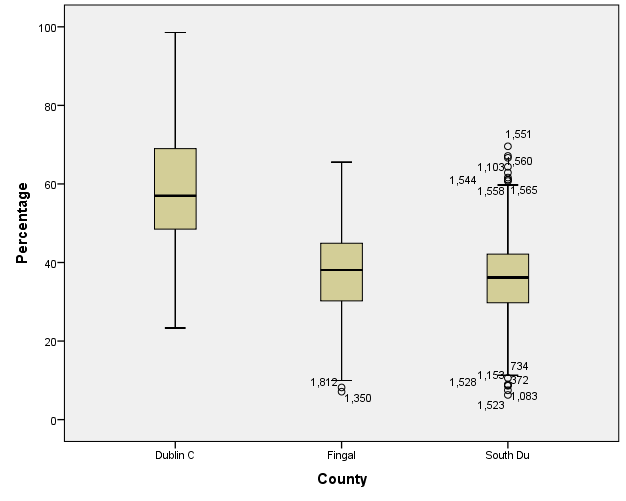
## Dublin’s attitude to environmentally friendly transport to get to work, school or college.

By Dublin we mean all Electorat Division answers in Census 2011 in Dublin City, South of Dublin and Fingal. We will compare answers of percentage of population in given Electorat Division as number of people living in every electorat division is different and simple number of people will be meaningless for this comparison. Our goal is compare the percentage of population using eco friendly transport.

### Descriptive statistic

Based on answers of percentage of population in given electorat division of Census 2011.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Descriptives** | | | | | |
|  | County | | | Statistic | Std. Error |
| Percentage | Dublin C | Mean | | 58.84 | .310 |
| 95% Confidence Interval for Mean | Lower Bound | 58.23 |  |
| Upper Bound | 59.45 |  |
| 5% Trimmed Mean | | 58.60 |  |
| Median | | 57.00 |  |
| Variance | | 212.030 |  |
| Std. Deviation | | 14.561 |  |
| Minimum | | 23 |  |
| Maximum | | 99 |  |
| Range | | 75 |  |
| Interquartile Range | | 20 |  |
| Skewness | | .295 | .052 |
| Kurtosis | | -.539 | .104 |
| Fingal | Mean | | 37.76 | .348 |
| 95% Confidence Interval for Mean | Lower Bound | 37.08 |  |
| Upper Bound | 38.44 |  |
| 5% Trimmed Mean | | 37.76 |  |
| Median | | 38.11 |  |
| Variance | | 113.441 |  |
| Std. Deviation | | 10.651 |  |
| Minimum | | 7 |  |
| Maximum | | 66 |  |
| Range | | 58 |  |
| Interquartile Range | | 15 |  |
| Skewness | | -.048 | .080 |
| Kurtosis | | -.373 | .160 |
| South Du | Mean | | 35.93 | .355 |
| 95% Confidence Interval for Mean | Lower Bound | 35.24 |  |
| Upper Bound | 36.63 |  |
| 5% Trimmed Mean | | 35.95 |  |
| Median | | 36.18 |  |
| Variance | | 114.062 |  |
| Std. Deviation | | 10.680 |  |
| Minimum | | 6 |  |
| Maximum | | 70 |  |
| Range | | 63 |  |
| Interquartile Range | | 12 |  |
| Skewness | | -.032 | .081 |
| Kurtosis | | .029 | .162 |



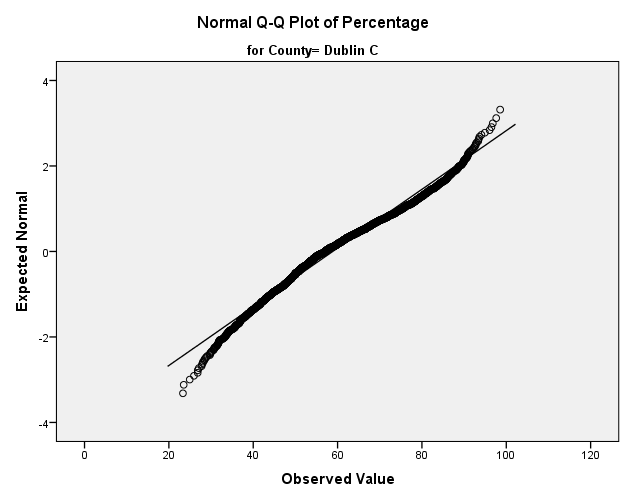
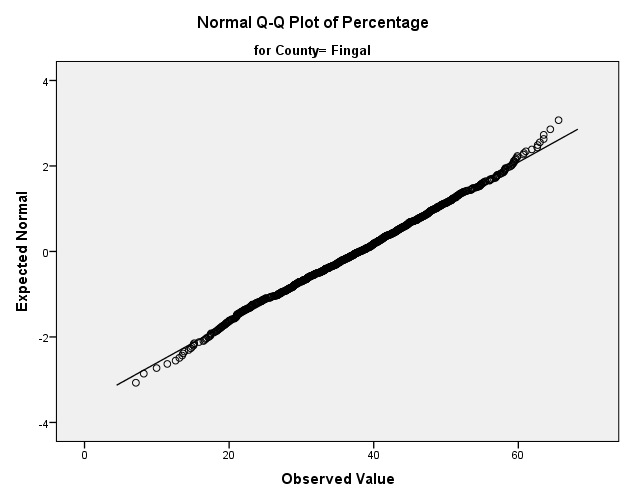
### Assessing the normality

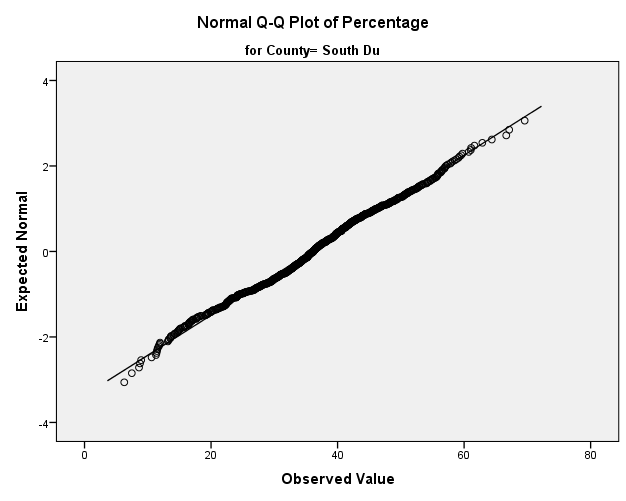
1. Hypotheses

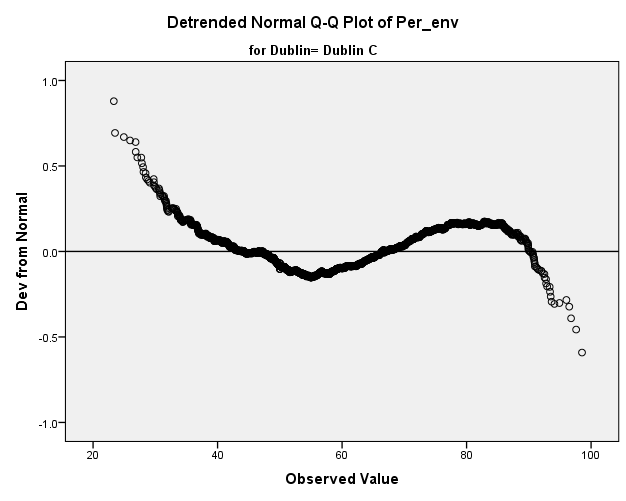
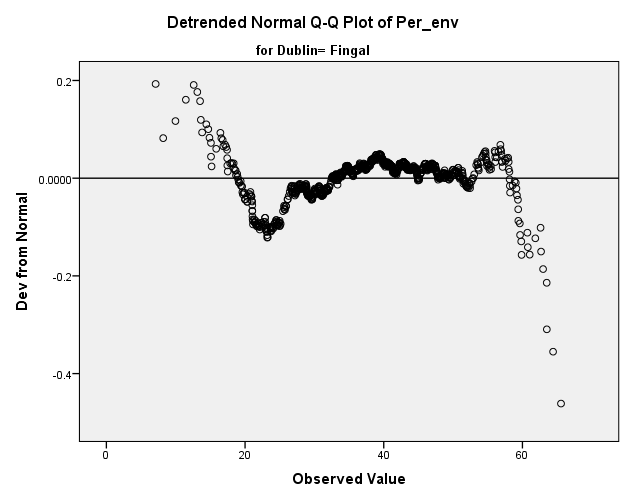
H0: The sample data are normally distributed

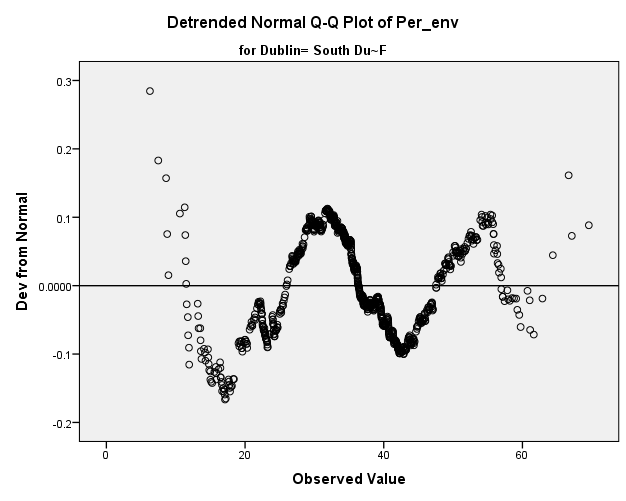
H1: The sample data are not normally distributed

1. Histograms
2. Dublin City
3. Fingal
4. South Dublin
5. Q-Q plots

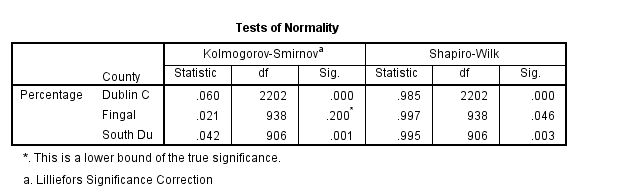








1. Kolmogorov-Smirnov and Shapiro-Wilk



1. Determine p value

p=0.05, p values less than 0.05 mean data are NOT normal

p values greater than 0.05 mean data are normal

1. Report

Dublin City – p<0.05 not normal distribution of data

Fingal – p<0.05 not normal distribution of data as per Shapiro Wilk test, but it is close to be normally distributed (0.046)

South Dublin - – p<0.05 not normal distribution of data

1. Conclusions

Mean – Median – Mode and box plot comparison - South Dublin and Fingal have similar mean but Dublin City have much higher. It’s expected, of course is somebody is leaving is city centre is much more probable that he is using foot or bus than a car. I would be surprised if mean or median would be lower than in other parts of Dublin like South or North. That could be alarming. There is no need to compare City Centre in that case with other parts of Dublin.

Mean – Median- Mode comparison also show that in all cases Mean and median are similar that could indicate normal distribution.

Histogram and Q=Q plots show that data is close to be normally distributed but its not. There is some distance from line of normal distribution on Q-Qplots.

Shapiro-Wilk test show us that as Dublin City and South Dublin do not have normally distributed data that Fingal is very close normal distribution but still its not. (p = 0.046)

### Mann-Whitney U Test and to compare Fingal and South Dublin in terms of environmental friendly transport to work, school or college.

Justification of choosing this test on this dataset- as I cannot use ANOVA to compare means (we cannot assume normal distribution in all our data) the decision is to use Mann-Whitney U Test to compare parts of Dublin. Is there a difference between this 2 groups?

1. Hypotheses

Null Hypothesis (H0)

The distribution of percentage of people using eco-friendly transport is the same in South and North of Dublin.

Alternate Hypothesis (H1)

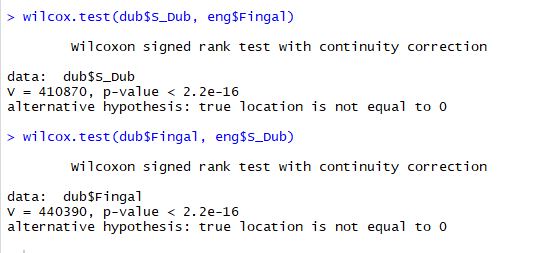
The distribution of percentage of people using eco-friendly transport is not the same in South and North of Dublin.

1. Critical value (α = 0.05) – As 95 % sure that we wont make type I error is enough when comparing eco friendly transport.
2. Determine p value :

p=0.05, p values less than 0.05 mean data are NOT normal

p values greater than 0.05 mean data are normal

1. Test



1. Report

p<0.05

A Mann-Whitney U test was run to determine if there were differences in percentage of people using eco-friendly transport in South and North of Dublin

Distribution of percentage of people using eco-friendly transport is not the same in South and North of Dublin, as assessed by visual inspection

Percentage of people using eco-friendly transport was statistically significantly different between parts of Dublin  
(Mdn = 36.18) and females (Mdn = 38.11),   
V = 41 and V =44, p <0.05

1. Conclusion

p < 0.05

Significant difference between parts of Dublin.

I reject H0 that distribution of percentage of people using eco-friendly transport is the same in South and North of Dublin. Its not this same.

## Comparison of Cork, Galway and Limerick City in terms of environmental friendly transport

### Descriptive statistics

### 





### Assessing the normality

1. Hypothesis

H0: The sample data are normally distributed

H1: The sample data are not normally distributed

1. Histogram

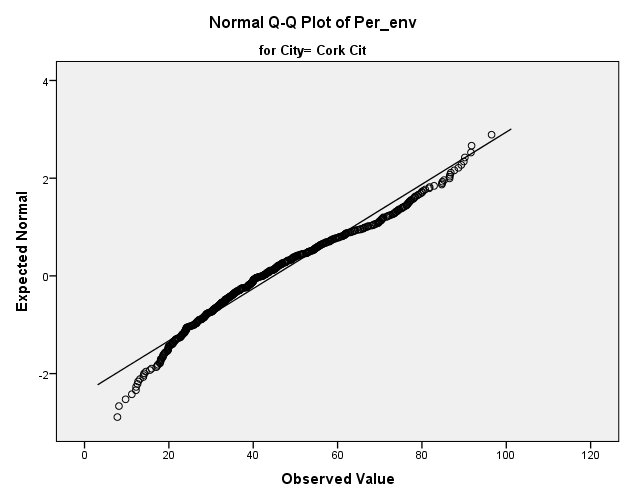
Cork City

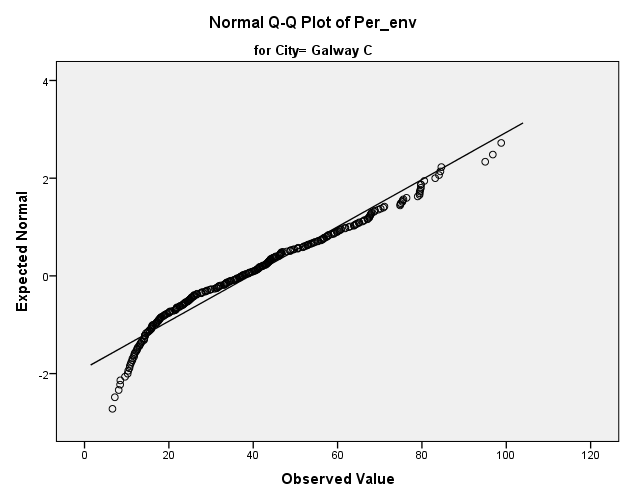
Galway City:

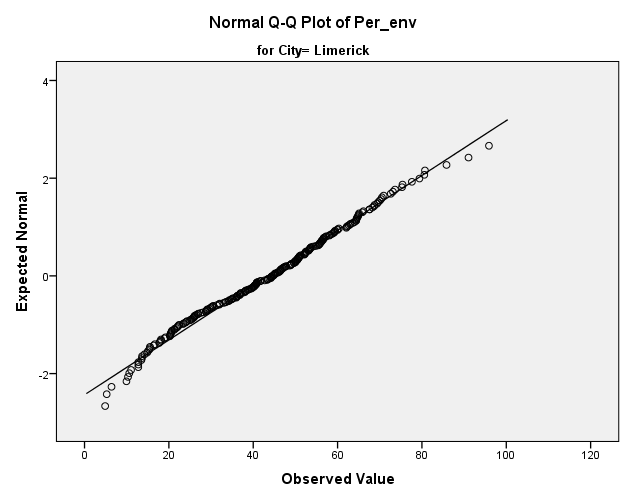
Limerick City:

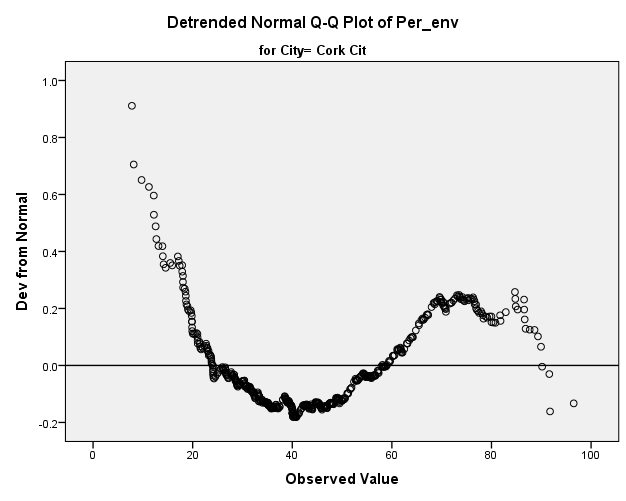
1. Q-Q plot

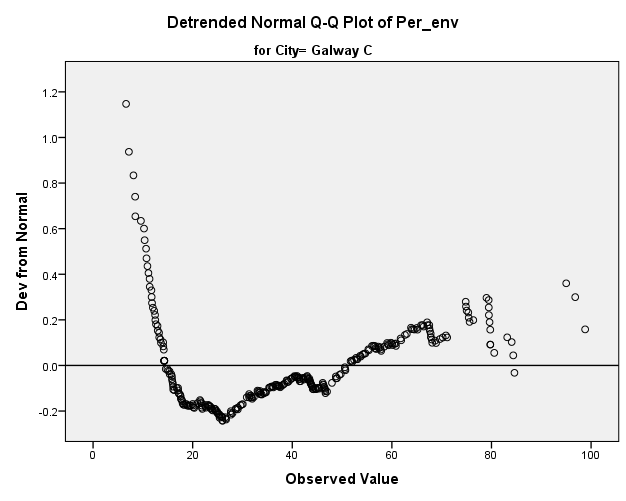
|  |
| --- |
|  |

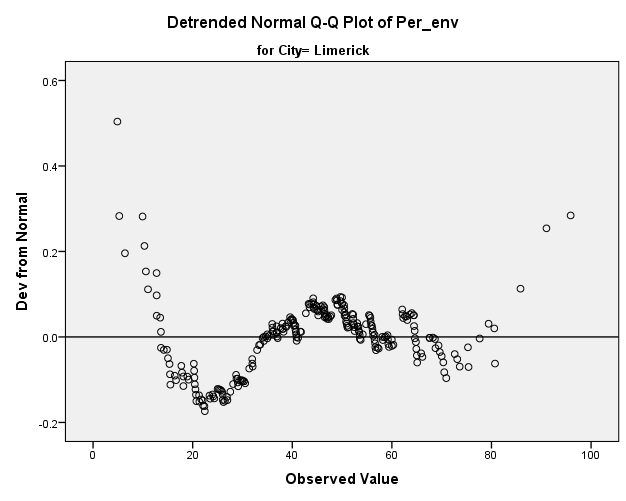


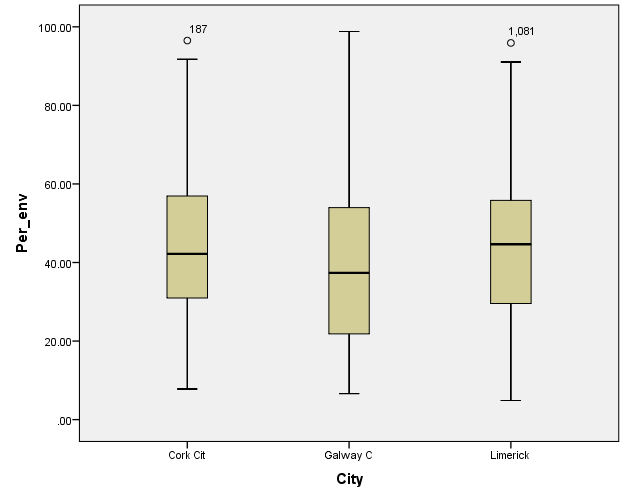












1. Kolmogorov-Smirnov and Shapiro-Wilk

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Tests of Normality** | | | | | | | |
|  | City | Kolmogorov-Smirnova | | | Shapiro-Wilk | | |
|  | Statistic | df | Sig. | Statistic | df | Sig. |
| Per\_env | Cork Cit | .073 | 519 | .000 | .971 | 519 | .000 |
| Galway C | .087 | 307 | .000 | .957 | 307 | .000 |
| Limerick | .042 | 258 | .200\* | .991 | 258 | .096 |
| \*. This is a lower bound of the true significance. | | | | | | | |
| a. Lilliefors Significance Correction | | | | | | | |

1. Determine p value

p=0.05, p values less than 0.05 mean data are NOT normal

p values greater than 0.05 mean data are normal

1. Critical value (α = 0.05) – As 95 % sure that we wont make type I error is enough when comparing eco friendly transport.
2. Report

Galway p<0.05 – not normal distribution

Limerick p>0.05 – normal distribution

Cork p<0.05 – not normal distribution

1. Conclusions

Result of Shapiro\_Wilk Test (p=0.096) and Q-Q plots shows us that Limerick data distribution is normal other cities distribution of data is not normal. Histograms were not so clear about that but just a simple look into descriptive statistics median and mean comparison confirm the test result.

Anyway, not all our data have normal distribution so we not able to use parametric test

### Kruskal-Wallis H Test to compare Cork, Galway and Limerick City

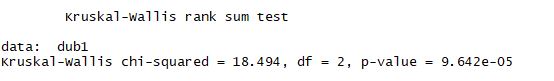
1. Hypothesis

H0: The three probabilities distributions are the same

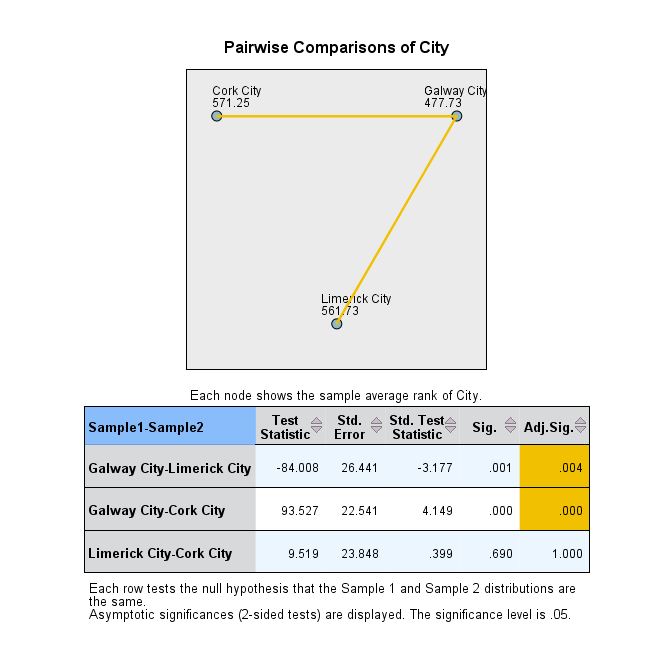
H1: The three probabilities distributions are NOT the same

1. Kruskal-Wallis H Test

Justification: ANOVA assume normality, it is better to use non-parametric test Kruskal-Wallis H Test when we do not clearly can state normality of data. As in our normality test we see that two of cities do not have normal distribution, then I decided to use this test.



1. Post – hoc test



1. Determine p value

p=0.05, p values less than 0.05 mean data are NOT normal

p values greater than 0.05 mean data are normal

1. Report

Kruskal-Wallis = 18.494, df =2, p<0.05

Kruskal-Wallis test was run to determine if there were differences in distribution of data of percentage of people using eco-friendly transport in 3 cities of Ireland.

Distributions were not similar, as assessed by visual inspection.

Percentage of people using eco-friendly transport was statistically significantly different between Cities. Further Post Hoc test show is that there is a difference only between Galway and Limerick (0.004) and Galway and Cork City (0.000)

1. Conclusions

Probabilities < 0.05

Reject the Null Hypotheses. There is a difference between cities in terms of percentage people using eco – friendly transport.

Significant difference between the Cities has been found (α = 0.05

Further Post hoc test show us that Limerick and Cork City are this same, but the Galway is the City which is different from other two. I would recommend making some action to promote eco-friendly transport in Galway.

## Conclusions

Our test to asses the normality and further to compare distribution of data showed that Ireland is not this same and in some parts we more likely to use eco – friendly methods of transports and in other we prefer to use a car. Like for example Galway city seems to be much more behind other cities. Or like comparing North and South of Dublin. In North they more likely using eco-friendly transport. It is worth to check why.

## Bibliography

1. <http://airo.maynoothuniversity.ie/files/dDATASTORE/small_areas/theme_11_small_areas.csv>. [Access date 10/04/2018]
2. [*https://statistics.laerd.com/spss-tutorials/kruskal-wallis-h-test-using-spss-statistics.php*](https://statistics.laerd.com/spss-tutorials/kruskal-wallis-h-test-using-spss-statistics.php) *[*Access date 10/04/2018]

1. <http://airo.maynoothuniversity.ie/files/dDATASTORE/small_areas/theme_11_small_areas.csv>. [Access date 10/04/2018] [↑](#footnote-ref-1)