Is There a Significant Difference in Transportation Choices Between Cork, Dublin, Galway and Limerick?

Vinicius Rech Verza

***Abstract* — The purpose of this report is to analyze if there are significant differences in the methods of transportation between Cork, Dublin, Galway and Limerick. It consists in two data-sets from CSO containing data from 2016 Census. This paper contains information about how the data was handled, prepared, cleaned, tested as well as conclusions. The statistical tests used in this analysis were Shapiro-Wilk for normality tests, Kruskal-Wallis for main hypothesis and Bonferroni tests for post-hoc analysis for deeper understanding. Following that, the interpretation of results as well as other relevant information.**

***Index Terms* — Methods of Transportation, Analysis, Statistical Tests, Cork, Dublin, Galway, Limerick, Ireland.**

I. Introduction

Ireland, a country member of the European Union, home of the famous Guinness beer, green countryside and happy people is also known as the Celtic Tiger [1]. This term refers to the rapid economic growth fueled by foreign investments, mostly on the Information Technology field. With this economic boom, people moved to Ireland attracted to job opportunities and the good quality of life. With that, cities became more populous and the transportation system was certainly affected. The finality of this analysis is to find out if there is a difference in the proportion of the methods of transportation chosen by people to go from home to work or school/college. The cities selected for this analysis were Cork, Dublin, Galway and Limerick. In order to get the data for this analysis, two data-sets were selected from The Central Statistics Office website, an official Irish government website. The data-sets were ‘Small Area Population Statistics’ [2], which provided the required information for commuting choices and ‘Small Areas Generalized Boundaries’ [3] data-set, in order to provide detailed information about small areas within the counties. The two datasets are from Census 2016. The idea of this report is to understand if there are significant differences in the transportation methods, comparing the cities by using statistical tests and descriptive analysis. The tools used for this report were Microsoft Excel, R, SPSS and JASP. The data will be used by separating only the necessary measures (fields) and the relevant cities. The tests were performed on a random sample basis, not with the whole data.

Below you will find detailed information about the preparation of the data for the analysis such as first preparations, review, merging, cleaning and tiding and coding (for some of the steps).

Manuscript received March 23, 2020, to be revised. The review of this paper is coordinated by Prof. Theo Mendonça.

Vinicius Rech Verza is with the National College of Ireland, Department of Computing Science, Dublin, Ireland (email: viniverza@gmail.com).

II. Data-set Preparation

The first data-set, named ‘Small Area Population Statistics’ contained information about population age, sex, martial status, place of birth, nationality, ethnic background, cultural background, religion, foreign languages, English and Irish proficiencies, family sizes, children ages, householding by type and size, renting, water supply, central heating, sewerage, population by social class, rank of education, means of commuting to work, school and college detailed by types of transportation used and lastly, disability population status. The second data-set, ‘Small Areas Generalized Boundaries’, contained much less information but one very important, the names of small areas of all the information collected on the 2016 Census.

With these two data-sets in hand, the first step was to use the field ‘GUID’ (which is the common field within the two data-sets) as a reference to merge the two. That was performed in order to have all the data from the first data-set now organized by small areas and county, from the second data-set. This procedure was performed on Microsoft Excel. Secondly, all the information that was not relevant for this analysis was deleted, remaining only the ‘Theme 11’, which is commute (from the first data-set) and counties and small areas from the second data-set.

Then, lastly, the names of the columns were renamed for best readability. This was followed by the glossary provided in the CSO Website. The columns were named previously as acronyms, for example ‘T11_1_BUW’ which stands for ‘bus, minivan or coach – work’.

The next following steps were then performed on the software named R Studio, using R programming language. The library package 'dplyr' [4] was also used. The first step was to get the previously edited and merged data-set loaded (with proper UTF-8-BOM encoder) and create subsets with the relevant cities: Cork, Dublin, Galway and Limerick. Secondly, getting rid of the irrelevant columns for the analysis while selecting only the ones that are going to be used: County, Area, Walktotal, Cycletotal, Bustotal, Traintotal, Drivertotal, Psgrtotal, Vantotal, Othertotal, Hometotal and Natotal. With that then selected, I was able to combine/create relevant columns into fewer ones using the mutate function, making the analysis easier. The column 'Walkcycle' is a sum of 'Walktotal' and 'Cycletotal', which are the total of people who walked or cycled from home to either work or school/college. The second column, named 'Public' is the sum of 'Bustotal', 'Traintotal' and 'Vantotal', combining methods of public transportation into only 1 variable. The third column was named 'Car', which is the sum of 'Drivertotal' and 'Psgrtotal'.

Following these steps, a re-selection of columns was performed and the final selected columns were: County, Area, Walk-cycle, Public and Car. With the data-set now almost ready for analysis, a sample of 100 rows from each county was selected with the function 'sample_n' from 'dplyr' library. Lastly, the file was saved and is ready for analysis on SPSS and JASP.

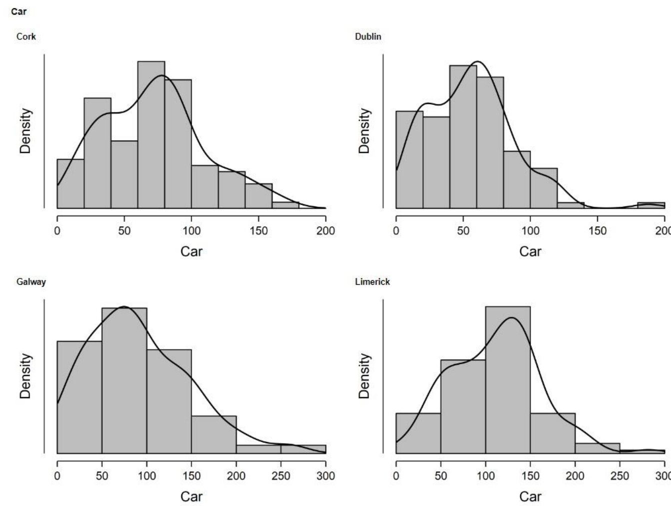
III. Descriptive Statistics and Normality Check

The first step in this analysis was to observe the data and interpret its normality. In other words, define if the data is parametric or non-parametric. This step is crucial because it defines the right path when choosing the correct type of statistical test that should be compatible with this specific data-set.

Walk-Cycle	Value	Public	Value	Car	Value
Mean	45.5275	Mean	23.63	Mean	82.335
Standard Error	1.72810633	Standard Error	0.74566097	Standard Error	2.47637093
Median	39.5	Median	21	Median	74.5
Mode	28	Mode	12	Mode	39
Standard Deviation	34.5621266	Standard Deviation	14.9132193	Standard Deviation	49.5274185
Sample Variance	1194.5406	Sample Variance	222.40411	Sample Variance	2452.96519
Kurtosis	11.516614	Kurtosis	1.18480815	Kurtosis	0.72968164
Skewness	2.26307831	Skewness	1.12900044	Skewness	0.80941399
Range	313	Range	82	Range	282
Minimum	1	Minimum	0	Minimum	0
Maximum	314	Maximum	82	Maximum	282
Sum	18211	Sum	9452	Sum	32934
Count	400	Count	400	Count	400

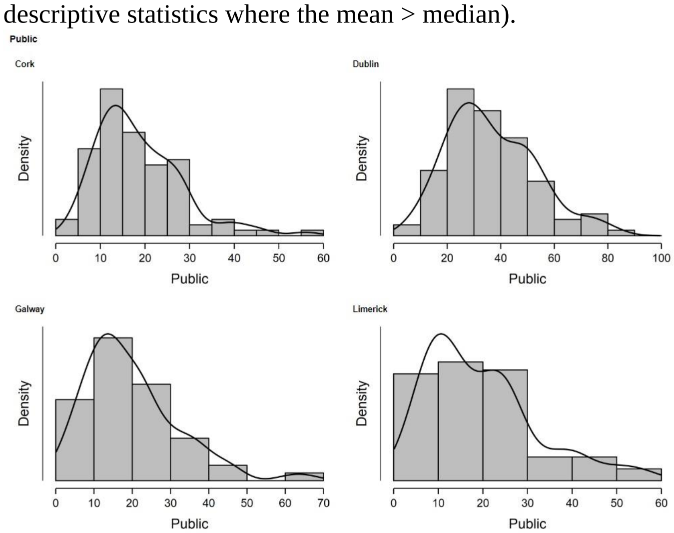
Higher resolution image can be found on Appendix A – Descriptive Statistics.

The above table contains the descriptive statistics for the three methods of transportation selected for this analysis. The mean values represent the average number of people who take this method of transportation. By comparing the mean values with median and mode (central tendency), we have the first evidence that this data is not normally distributed based on the fact the numbers differ reasonably within themselves. The standard deviation and variance are the measures of dispersion. The Walk-Cycle and Car groups have a higher spread of values compared to the Public group but yet slightly high. This also tells us the data is most likely non-parametric. Lastly, kurtosis and skewness tell us about the symmetry and if the data is tailed. The most symmetric data is from the Cars group, while the other two are less symmetric. The most peaked group (by a high value) is the Walk-Cycle while the other two groups have a much lower peak. Some other values on this descriptive statistics table are the count, the total number of values in each group (100 sampled for each city). Sum, which is the sum of the values of each group, minimum and maximum, the lowest and the highest number recorded on each group and range, which is basically the highest recorded value minus the lowest.



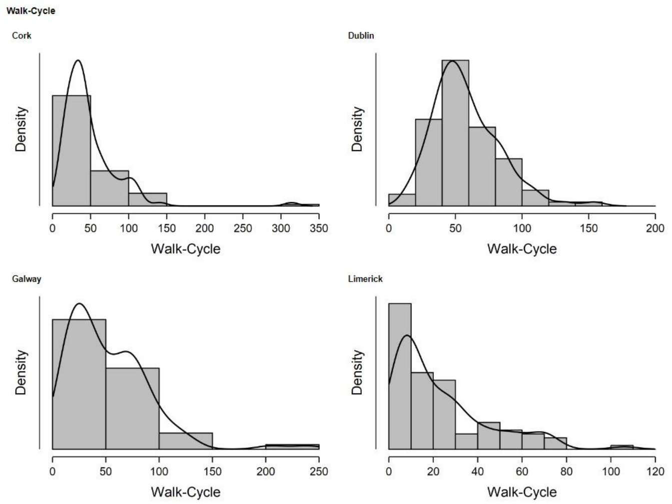
Higher resolution image can be found on Appendix B – Distribution Plots of Car group.

The four graphs above show the distribution of values from the Car group within the four cities (Cork, Dublin, Galway and Limerick). As we can observe, all the cities contain a distribution that is right skewed (also confirmed by the descriptive statistics where the mean > median).



Higher resolution image can be found on Appendix C – Distribution Plots of Public group.

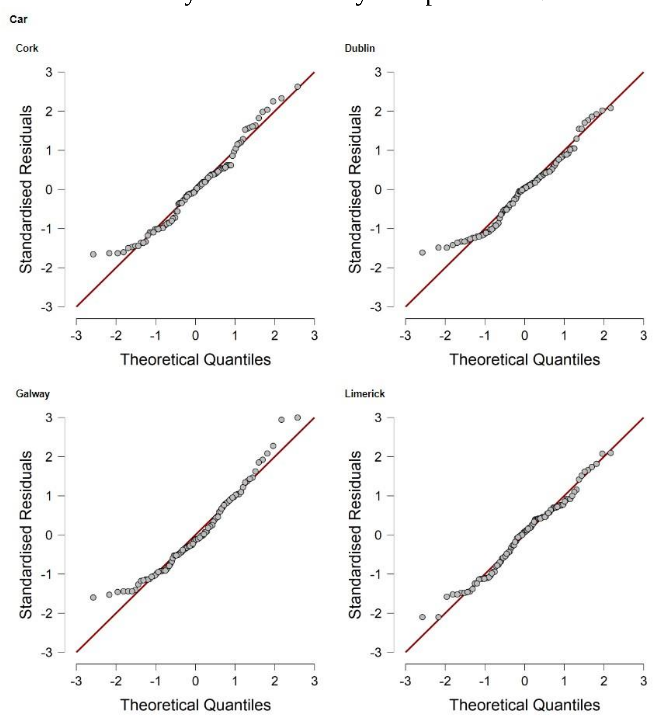
Above, plots of distribution for the Public group. Again, we can observe the same pattern as before: positive, or right, skewness. In these instances, the mean value is also higher than the median.



Higher resolution image can be found on Appendix D – Distribution Plots of Walk-Cycle group.

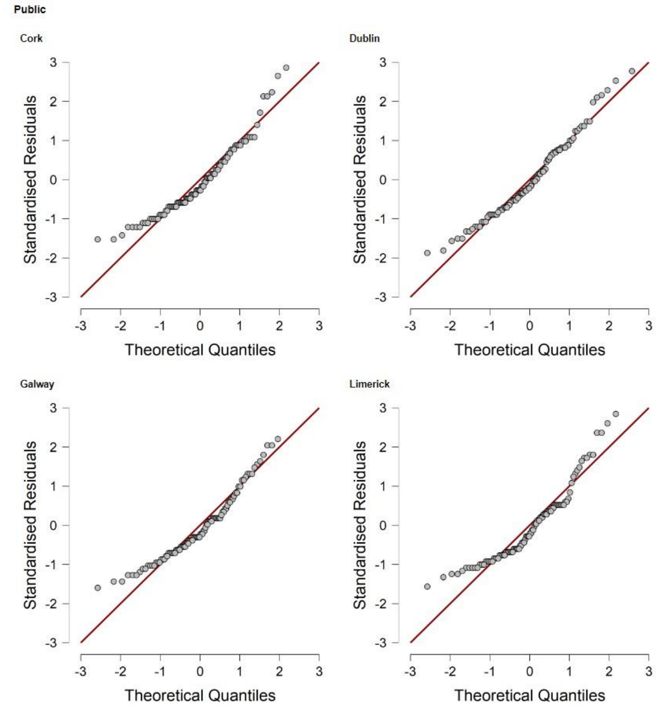
Now, the Walk-Cycle group, which shows us again the same right skewness pattern. Note that in this case the kurtosis (peak) is considerably higher than the other two groups.

These plots are a good way to visually-represent the data and to understand why it is most likely non-parametric.



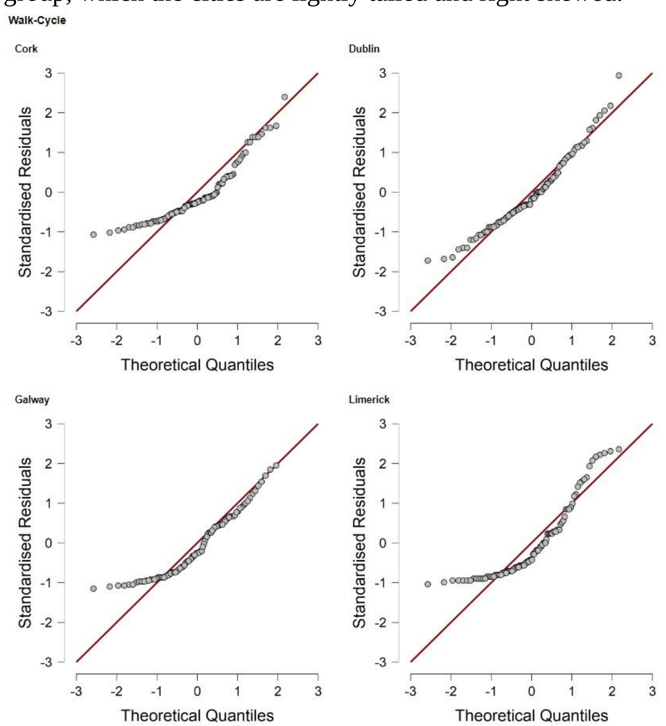
Higher resolution image can be found on Appendix E – QQ Plots of Car group.

The above Q-Q plots can also tell us about the distribution, considering skewness and if the data is normal or tailed. According to the plots above, they are lightly tailed and right skewed.



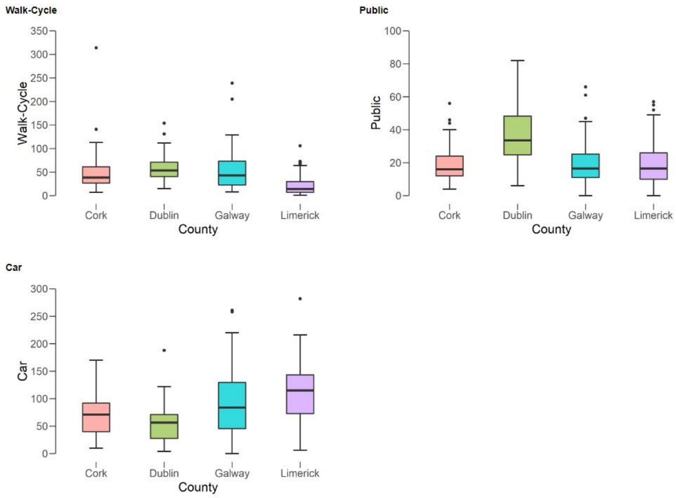
Higher resolution image can be found on Appendix F – QQ Plots of Public group.

Now, we can observe the same scenario with the Public group, which the cities are lightly tailed and right skewed.



Higher resolution image can be found on Appendix G – QQ Plots of Walk-Cycle group.

The group of people who walk/cycle has a slightly different scenario. This case, the cities are heavy-tailed and the right skewness is higher than the other groups.



Higher resolution image can be found on Appendix H – Box Plots.

Above we can observe some box plots representing the distribution of values or groups by cities, divided by color for ease of understanding. You can notice the outliers on some instances. This was then manually analyzed and defined that the values are in fact correct. They are featured from the others due to some areas of the cities that are more populous than others, therefore more people live there and use methods of transportation.

Finally, to be certain about this data-set being non-parametric, a hypothesis was created and conducted by using the Shapiro-Wilk test. The null and alternative hypothesis are as follow:

H_0: Normally Distributed Sample
H_1: Not Normally Distributed Sample

The null hypothesis states that the data is normally distributed between the groups. The alternative hypothesis states that the data is not normally distributed.

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	County	Statistic	df	Sig.	Statistic	df	Sig.
WalkCycle	Cork	.194	100	.000	.702	100	.000
	Dublin	.114	100	.003	.946	100	.000
	Galway	.130	100	.000	.843	100	.000
	Limerick	.175	100	.000	.840	100	.000
Public	Cork	.117	100	.002	.918	100	.000
	Dublin	.088	100	.054	.970	100	.022
	Galway	.128	100	.000	.922	100	.000
	Limerick	.127	100	.000	.912	100	.000
Car	Cork	.088	100	.052	.970	100	.020
	Dublin	.066	100	.200*	.947	100	.000
	Galway	.086	100	.068	.958	100	.003
	Limerick	.059	100	.200*	.980	100	.141

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

Higher resolution image can be found on Appendix I – Test of Normality.

A Shapiro-Wilk test was conducted to determine if there is a significant difference between the distribution of normality between three groups of methods of transportation of four cities. The alpha value of $\alpha = 0.05$ was selected. This means there is a 5% chance of rejecting the null hypothesis when we shouldn't (Type 1 error). The obtained T_{stats} and P values for all the instances are listed in the table above. With exception of T_{stat}: 0.980, P value = 0.141 for Car – Limerick, all other P values are < P = 0.05, therefore we can reject the null hypothesis in favor of the alternative one. It appears that the data is not normally distributed.

With this information, now it is possible to determine a compatible type of statistical test to verify the main hypothesis of the report.

IV. Main Hypotheses

Considering the type of data and its non-parametricity, the Kruskal-Wallis test was selected for further analysis. Below, the null and alternative hypotheses.

$$H_0: \mu_{Cork\text{-}WC} = \mu_{Dublin\text{-}WC} = \mu_{Galway\text{-}WC} = \mu_{Limerick\text{-}WC}$$
$$H_1: \mu_{Cork\text{-}WC} \neq \mu_{Dublin\text{-}WC} \neq \mu_{Galway\text{-}WC} \neq \mu_{Limerick\text{-}WC}$$

$$H_0: \mu_{Cork\text{-}Public} = \mu_{Dublin\text{-}Public} = \mu_{Galway\text{-}Public} = \mu_{Limerick\text{-}Public}$$
$$H_1: \mu_{Cork\text{-}Public} \neq \mu_{Dublin\text{-}Public} \neq \mu_{Galway\text{-}Public} \neq \mu_{Limerick\text{-}Public}$$

$$H_0: \mu_{Cork\text{-}Car} = \mu_{Dublin\text{-}Car} = \mu_{Galway\text{-}Car} = \mu_{Limerick\text{-}Car}$$
$$H_1: \mu_{Cork\text{-}Car} \neq \mu_{Dublin\text{-}Car} \neq \mu_{Galway\text{-}Car} \neq \mu_{Limerick\text{-}Car}$$

The null hypotheses states that the population mean of all methods of transportation are the same across the cities. The alternative hypothesis states that at least one population mean of methods of transportation differs from the others.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of WalkCycle is the same across categories of County.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.
2	The distribution of Public is the same across categories of County.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.
3	The distribution of Car is the same across categories of County.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .050.

Higher resolution image can be found on Appendix J – Kruskal-Wallis Test.

A Kruskal-Wallis test was conducted to determine if there is a significant difference between the population mean of methods of transportation of four cities. The alpha value of $\alpha = 0.05$ was selected. The obtained P values of 0.000 < 0.05 significance level, therefore we can reject the null hypothesis in favor of the alternative one. It appears that at least one of the groups differs from the others.

After rejecting the null hypothesis and knowing that there are significant differences with some, and maybe all, of the groups, a Bonferroni post-hoc test was conducted to investigate which groups differ or do not differ between themselves.

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Limerick-Cork	101.960	16.349	6.237	.000	.000
Limerick-Galway	115.435	16.349	7.061	.000	.000
Limerick-Dublin	156.245	16.349	9.557	.000	.000
Cork-Galway	-13.475	16.349	-.824	.410	1.000
Cork-Dublin	-54.285	16.349	-3.320	.001	.005
Galway-Dublin	40.810	16.349	2.496	.013	.075

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Higher resolution image can be found on Appendix K – Bonferroni Walk-Cycle.

The obtained P value of 0.410 > 0.05 significance level, therefore we fail to reject the null hypothesis in favor of the alternative one. It appears that Cork and Galway have a population mean of people who choose to go work/college by walking or cycling that are not significantly different.

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Cork-Limerick	-.985	16.344	-.060	.952	1.000
Cork-Galway	-5.160	16.344	-.316	.752	1.000
Cork-Dublin	-129.315	16.344	-7.912	.000	.000
Limerick-Galway	4.175	16.344	.255	.798	1.000
Limerick-Dublin	128.330	16.344	7.852	.000	.000
Galway-Dublin	124.155	16.344	7.596	.000	.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Higher resolution image can be found on Appendix L – Bonferroni Public.

The obtained P values of 0.952 for Cork-Limerick, 0.752 for Cork-Galway and 0.798 for Limerick-Galway > 0.05 significance level, therefore we fail to reject the null hypothesis in favor of the alternative one. It appears that in these instances, the population mean of people choosing to go work/college by using the public transportation is not significantly different.

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Dublin-Cork	46.915	16.350	2.869	.004	.025
Dublin-Galway	-81.130	16.350	-4.962	.000	.000
Dublin-Limerick	-132.155	16.350	-8.083	.000	.000
Cork-Galway	-34.215	16.350	-2.093	.036	.218
Cork-Limerick	-85.240	16.350	-5.214	.000	.000
Galway-Limerick	-51.025	16.350	-3.121	.002	.011

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Higher resolution image can be found on Appendix M – Bonferroni Car.

The obtained P values < 0.05 significance level, therefore we can reject the null hypothesis in favor of the alternative one. It seems that there is no significant difference between the population mean of people who choose to go to work/college by car.

V. Conclusion

After all the statistical tests performed, it is possible to state that there are only a few significant differences within the methods of transportation chosen by people while commuting to work or school/college within the cities, while the majority of the choices statistically not different. The results are as follow:

DIFFERS	Cork	Dublin	Galway	Limerick
Cork			Walk-Cycle	
Dublin				
Galway	Public			
Limerick	Public		Public	

For public transportation, there is a significant difference of people from Cork compared to Galway and Limerick and from Limerick compared to Galway.

For people who walk or cycle, there is only a significant difference within Cork and Galway.

There are many possible causes for these differences such as population number and density, conditions of roads, distances from people's houses to their work, their social status and so on. Also, there are many other external factors that might affect these decisions, like weather, politics and so on. While this analysis only points out if there are differences between the groups, it is impossible to draw a conclusion of why such decisions differ or not.

References

[1] Celtic Tiger [Internet]. Wikipedia - Celtic Tiger. [cited 2020Mar23]. Available from: https://en.wikipedia.org/wiki/Celtic_Tiger

[2] Census 2016 Small Area Population Statistics [Internet]. CSO - Central Statistics Office. [cited 2020Mar23]. Available from: https://www.cso.ie/en/census/census2016reports/census2016smallareapopulationstatistics/

[3] Small Areas Generalised 100m - OSi National Statistical Boundaries - 2015 [Internet]. CSO - Central Statistics Office. [cited 2020Mar23]. Available from: https://data.gov.ie/dataset/small-areas-generalised-100m-osi-national-statistical-boundaries-2015

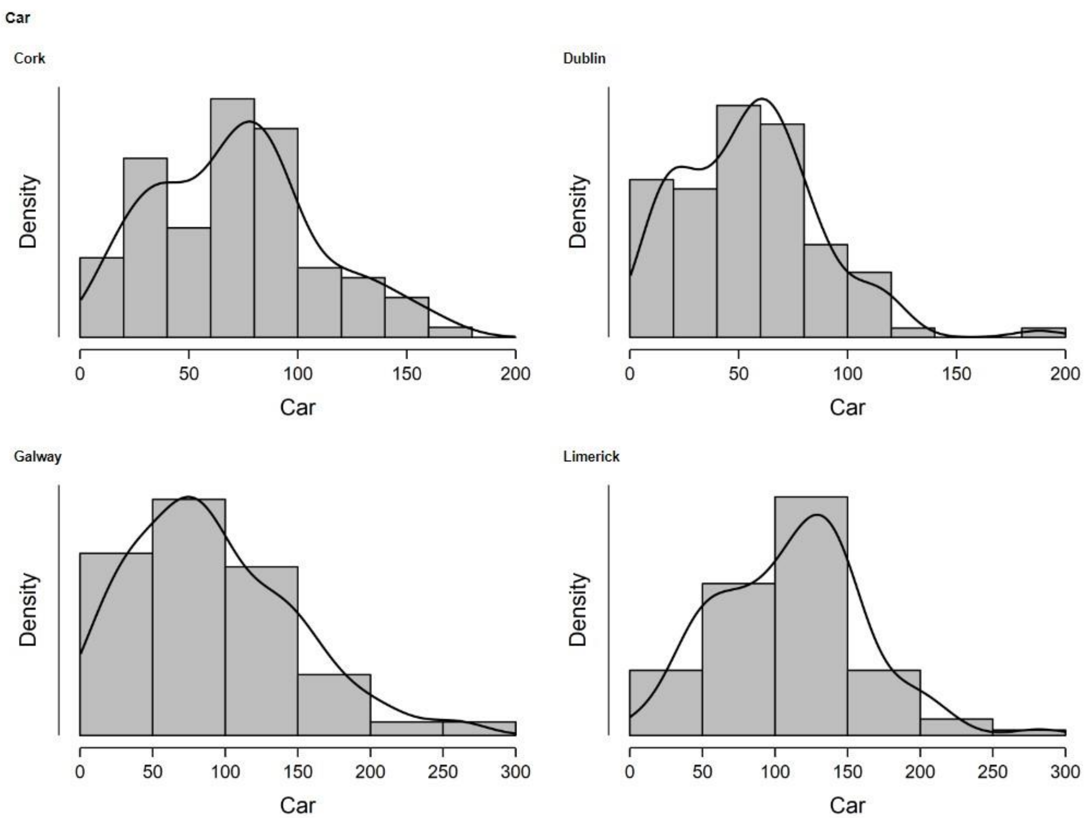
[4] dplyr [Internet]. dplyr tidyverse. [cited 2020Mar23]. Available from: https://dplyr.tidyverse.org/

APPENDIX

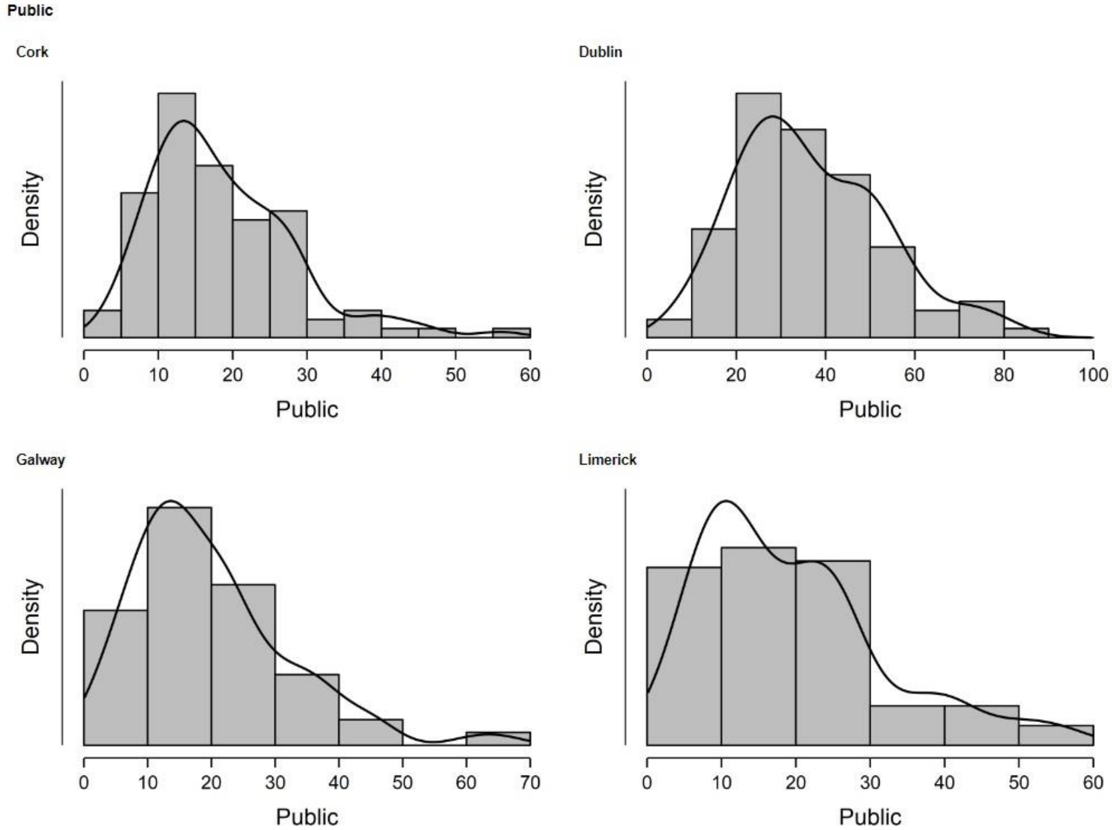
Appendix A – Descriptive Statistics

Walk-Cycle	Value	Public	Value	Car	Value
Mean	45.5275	Mean	23.63	Mean	82.335
Standard Error	1.72810633	Standard Error	0.74566097	Standard Error	2.47637093
Median	39.5	Median	21	Median	74.5
Mode	28	Mode	12	Mode	39
Standard Deviation	34.5621266	Standard Deviation	14.9132193	Standard Deviation	49.5274185
Sample Variance	1194.5406	Sample Variance	222.40411	Sample Variance	2452.96519
Kurtosis	11.516614	Kurtosis	1.18480815	Kurtosis	0.72968164
Skewness	2.26307831	Skewness	1.12900044	Skewness	0.80941399
Range	313	Range	82	Range	282
Minimum	1	Minimum	0	Minimum	0
Maximum	314	Maximum	82	Maximum	282
Sum	18211	Sum	9452	Sum	32934
Count	400	Count	400	Count	400

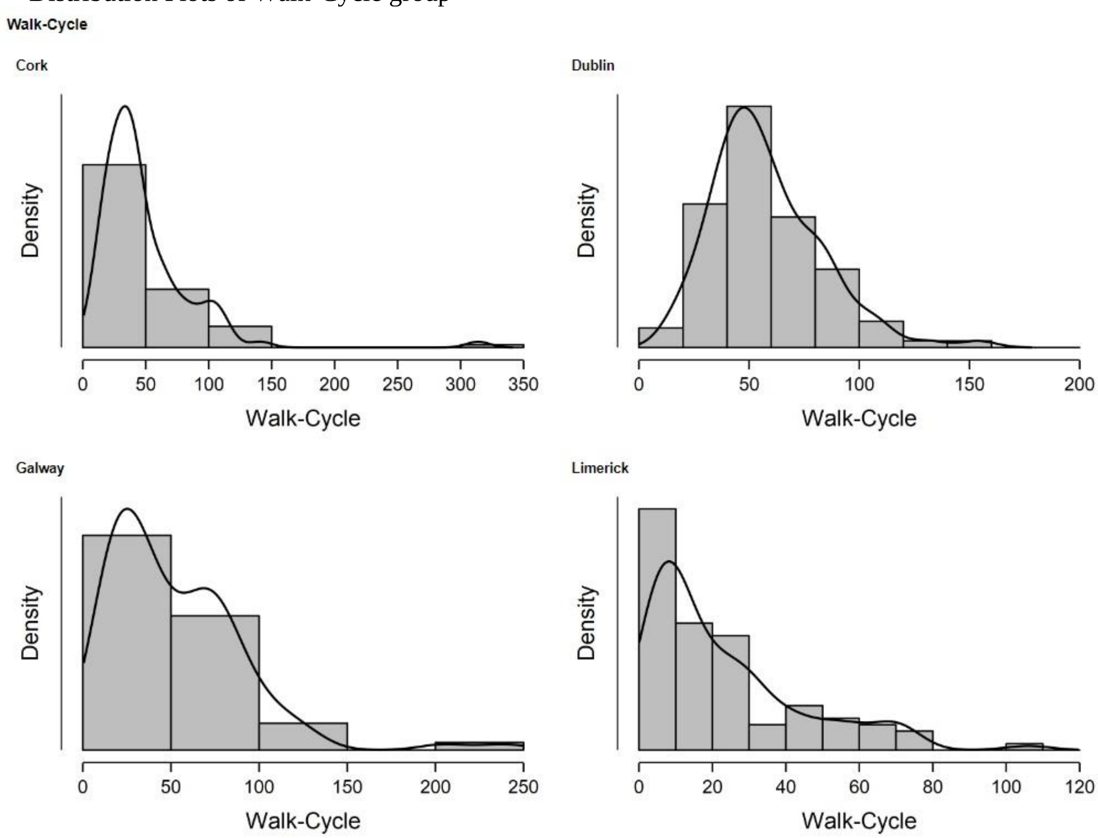
Appendix B – Distribution Plots of Car group



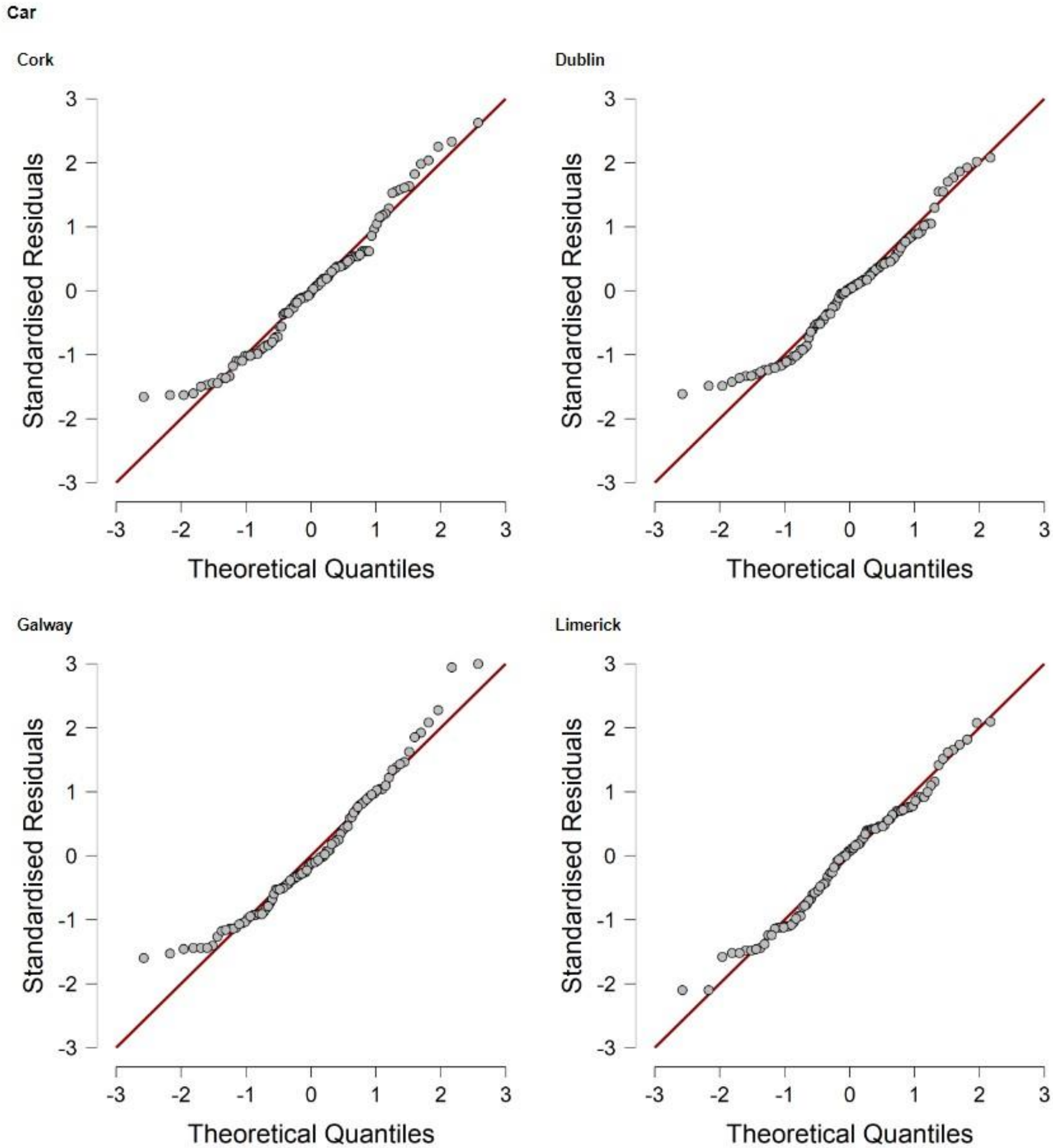
Appendix C – Distribution Plots of Public group



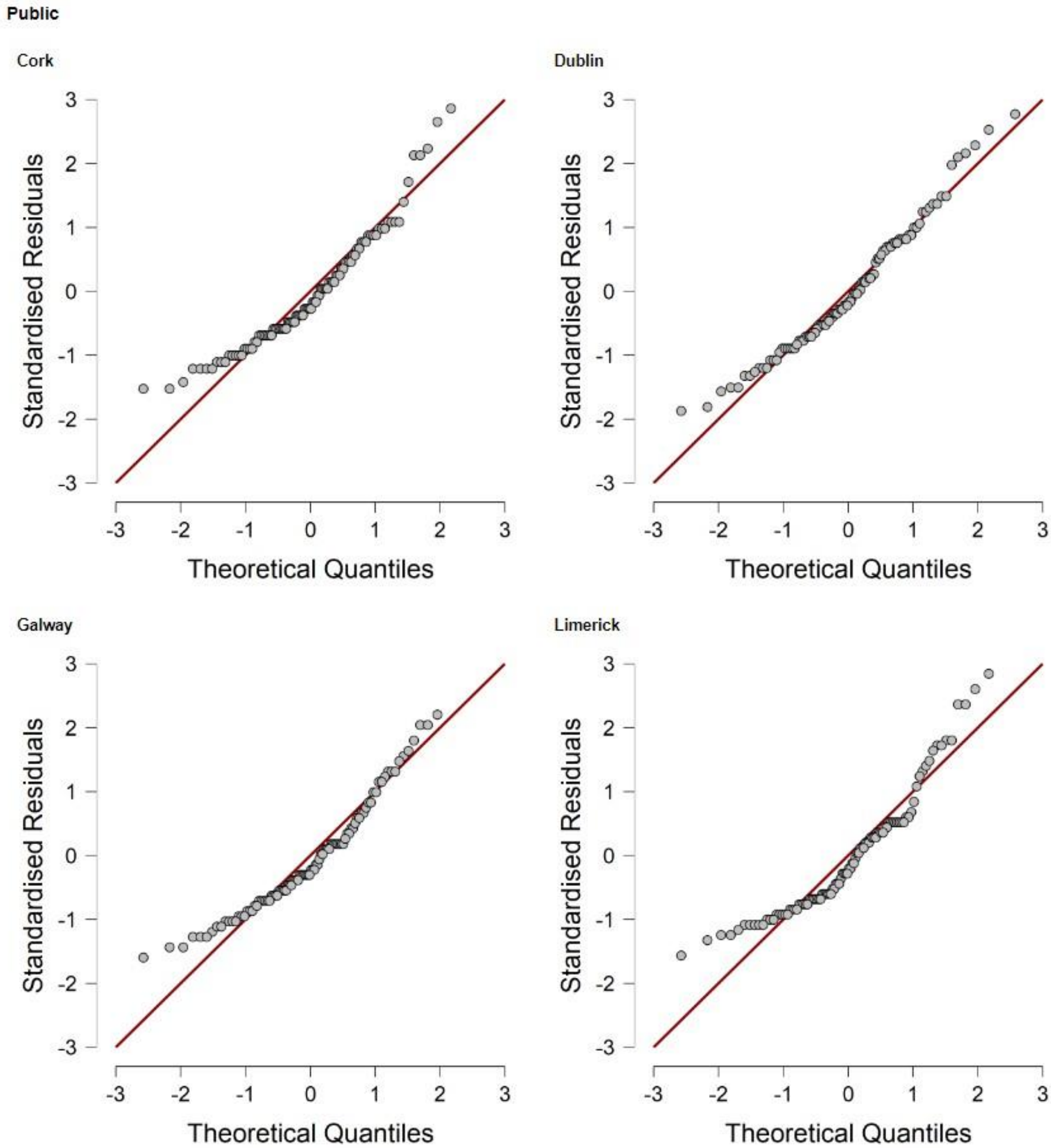
Appendix D – Distribution Plots of Walk-Cycle group



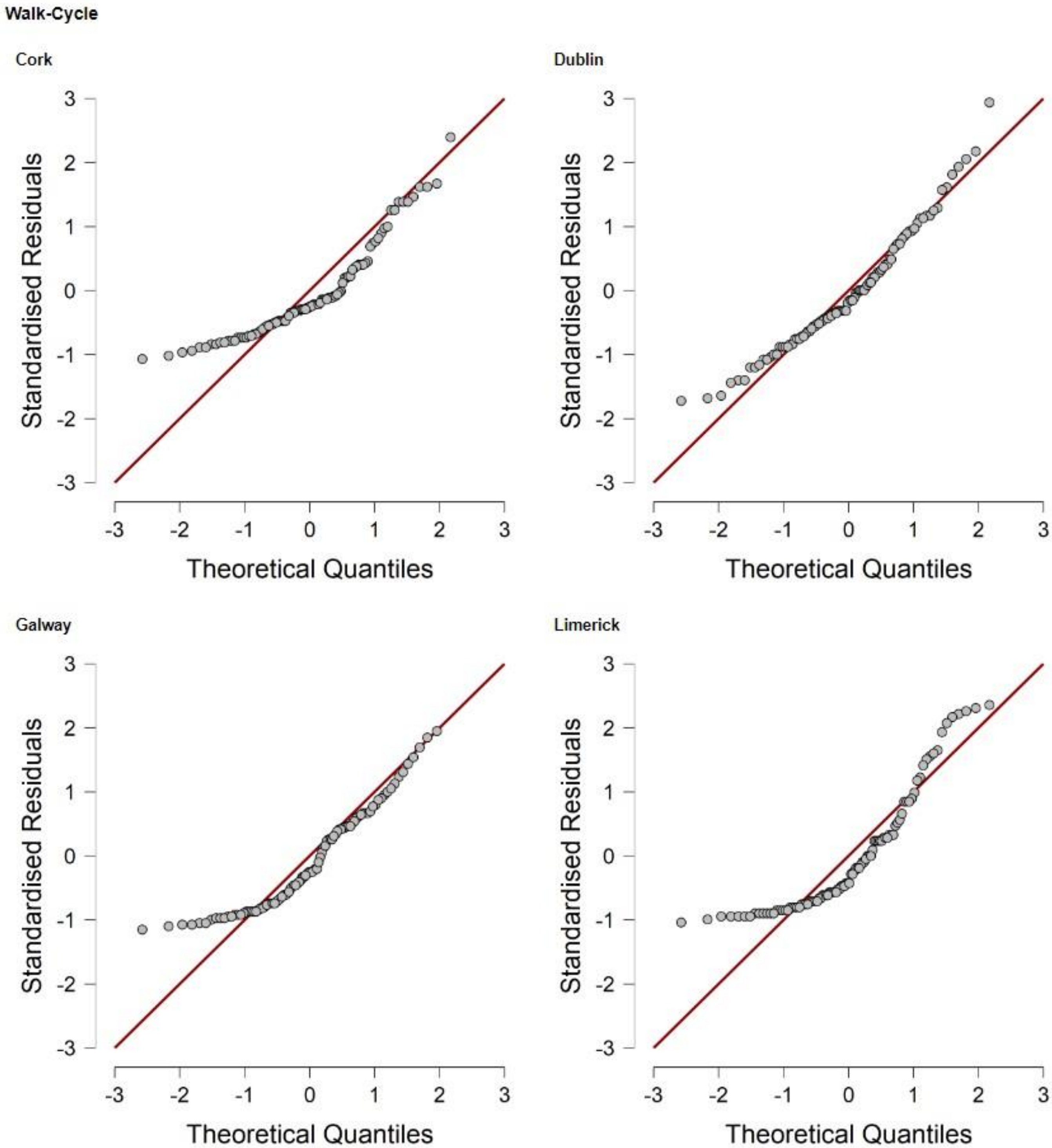
Appendix E – QQ Plots of Car group



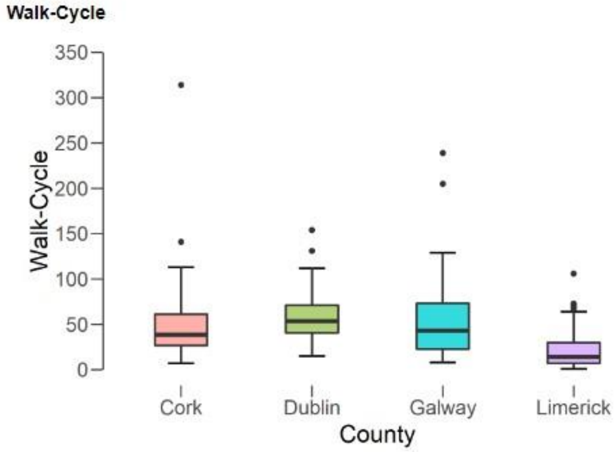
Appendix F – QQ Plots of Public group

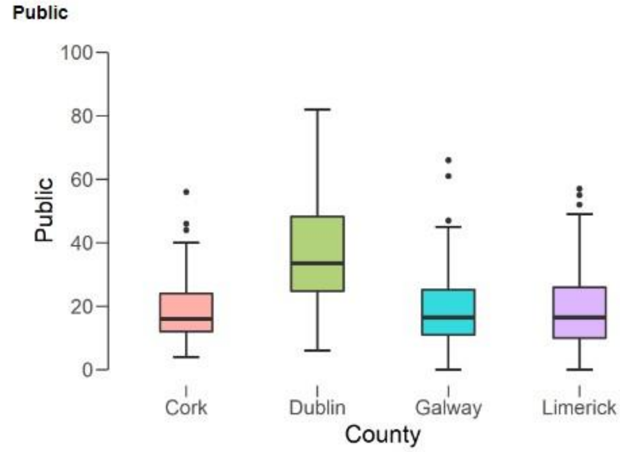


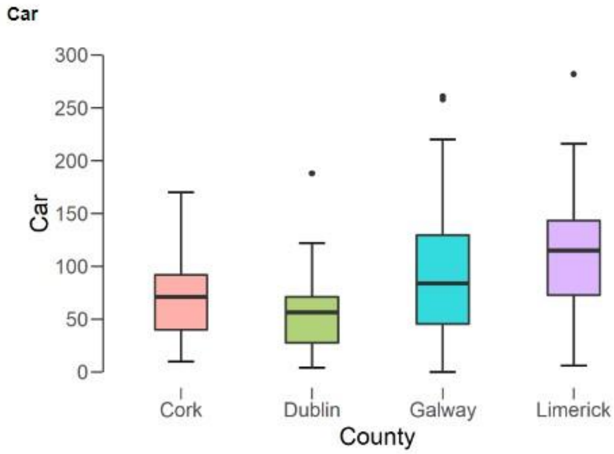
Appendix G – QQ Plots of Walk-Cycle group



Appendix H – Box Plots







Appendix I – Test of Normality

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	County	Statistic	df	Sig.	Statistic	df	Sig.
WalkCycle	Cork	.194	100	.000	.702	100	.000
	Dublin	.114	100	.003	.946	100	.000
	Galway	.130	100	.000	.843	100	.000
	Limerick	.175	100	.000	.840	100	.000
Public	Cork	.117	100	.002	.918	100	.000
	Dublin	.088	100	.054	.970	100	.022
	Galway	.128	100	.000	.922	100	.000
	Limerick	.127	100	.000	.912	100	.000
Car	Cork	.088	100	.052	.970	100	.020
	Dublin	.066	100	.200*	.947	100	.000
	Galway	.086	100	.068	.958	100	.003
	Limerick	.059	100	.200*	.980	100	.141

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Appendix J – Kruskal-Wallis Test

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of WalkCycle is the same across categories of County.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.
2	The distribution of Public is the same across categories of County.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.
3	The distribution of Car is the same across categories of County.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .050.

Appendix K – Bonferroni Walk-Cycle

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Limerick-Cork	101.960	16.349	6.237	.000	.000
Limerick-Galway	115.435	16.349	7.061	.000	.000
Limerick-Dublin	156.245	16.349	9.557	.000	.000
Cork-Galway	-13.475	16.349	-.824	.410	1.000
Cork-Dublin	-54.285	16.349	-3.320	.001	.005
Galway-Dublin	40.810	16.349	2.496	.013	.075

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Appendix L – Bonferroni Public

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Cork-Limerick	-.985	16.344	-.060	.952	1.000
Cork-Galway	-5.160	16.344	-.316	.752	1.000
Cork-Dublin	-129.315	16.344	-7.912	.000	.000
Limerick-Galway	4.175	16.344	.255	.798	1.000
Limerick-Dublin	128.330	16.344	7.852	.000	.000
Galway-Dublin	124.155	16.344	7.596	.000	.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Appendix M – Bonferroni Car

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Dublin-Cork	46.915	16.350	2.869	.004	.025
Dublin-Galway	-81.130	16.350	-4.962	.000	.000
Dublin-Limerick	-132.155	16.350	-8.083	.000	.000
Cork-Galway	-34.215	16.350	-2.093	.036	.218
Cork-Limerick	-85.240	16.350	-5.214	.000	.000
Galway-Limerick	-51.025	16.350	-3.121	.002	.011

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Appendix N – R Code

```
###### Required Libraries ######
library('dplyr')

###### Opening Dataset ######
dataset <- read.csv(file="combined.csv", head=TRUE, sep=",", fileEncoding="UTF-8-BOM")

###### Subsetting Cities ######
Big4 <- subset(dataset, COUNTY == 'Dublin City' | COUNTY == 'Cork City' | COUNTY == 'Limerick City and County' | COUNTY == 'Galway City')

###### Selecting First Relevant Columns ######
Big4 <- select (Big4, c(COUNTY, AREA, WALKTOTAL, CYCLETOTAL, BUSTOTAL, TRAINTOTAL, BIKETOTAL, DRIVERTOTAL, PSGRTOTAL, VANTOTAL, OTHERTOTAL, HOMETOTAL, NATOTAL))

###### Combining Columns ######
Big4 <- mutate(Big4,WALKCYCLE=WALKTOTAL+CYCLETOTAL)
Big4 <- mutate(Big4,PUBLIC=BUSTOTAL+TRAINTOTAL+VANTOTAL)
Big4 <- mutate(Big4,CAR=DRIVERTOTAL+PSGRTOTAL)

###### Selecting Final Relevant Columns ######
Big4 <- select (Big4, c(COUNTY, AREA, WALKCYCLE, PUBLIC, CAR))

###### Saving File ######
write.csv(Big4, file="Big4new.csv", row.names=FALSE)

###### Getting Samples ######
Big4 <- Big4 %>% group_by(COUNTY)
Big4_samples <- sample_n(Big4, 100)

###### Saving File ######
write.csv(Big4_samples, file="Big4_Samples.csv", row.names=FALSE)
```