

Is there a difference in the number of people cycling between 2001 and 2011 in the East Midlands?

What is the data really telling us?

2nd April 2014

Author: dennis@lccg.org.uk

Analysis done on behalf of the Leicester Cycling Campaign group.

Contents

1	Summary	3
2	What does the data look like for the number of cyclists between 2001 and 2011?	3
2.1	Q: Is there a difference in the variance (spread) of the number of people cycling between 2001 and 2011?	4
2.2	Q: is there a difference in the median number of people cycling between 2001 and 2011?	4
2.2.1	Is there a difference between the 2 years?	4
2.2.2	Is it worse in 2011 compared with 2001?	5
2.3	Interim Conclusion	5
3	Is there a difference in the proportions of people (16-74) cycling between 2001 and 2011?	6
3.1	Is there a difference in the variance of proportions of cyclists between 2001 and 2011?	6
3.2	Is there a difference in the proportion of cyclists between 2001 and 2011?	7
3.2.1	Is there a difference?	7
3.2.2	Is 2011 worse than 2001?	7
3.3	Interim Conclusion	7
4	Final conclusion	8
5	Final note and disclaimer	8
6	Appendix	9
6.1	Source	9
6.2	Data set for East Midlands	9
6.3	graphical Summary of cycling 2001 and 2011	10
6.4	Median analysis	10
6.5	Graphical Summary for proportion of people cycling	11
6.6	Variance in proportion of cyclists	11
6.7	Median of Proportion of cyclists	12

1 Summary

We would like to measure the whole population of cyclist which commute to work. But this is impractical, and typically impossible. Thus the best we can do is to take samples and then make assumptions about the whole population. This is where you need statistics to help you on that journey.

When we are presented with 2 sets of sampled data, in this case for 2001 and 2011, you will always find differences between two values.

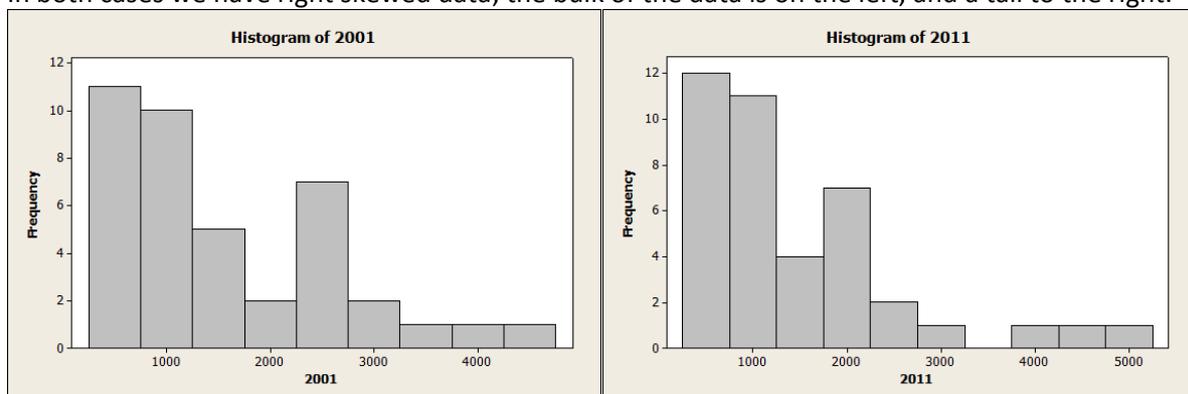
But the question we are always trying to ask ourselves, however, is how representative are these samples for the whole population of commuting cyclists?

As LCCG is based in the Midlands, I chose the Midlands data set for analysis purposes, see appendix.

The statistical analysis below may give us some surprising answers!

2 What does the data look like for the number of cyclists between 2001 and 2011?

In both cases we have right skewed data; the bulk of the data is on the left, and a tail to the right.



Some of the headline numbers for the samples taken

	2001	2011	Difference 2011-2001
Number of areas	40	40	0
Total number of cyclists in the sample	62644	57689	-4,955
Mean of the sample	1566	1442	-124
Median of the sample	1135	964	-171

As you can see there is a difference between the numbers in the samples, but remember, these are samples from 2 populations; we need to use statistics to make meaningful comparisons between the two populations.

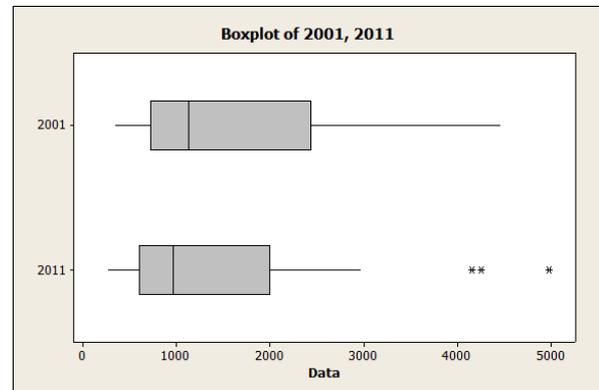
Leicester cycling campaign group

2.1 Q: Is there a difference in the variance (spread) of the number of people cycling between 2001 and 2011?

Variance is another word for the spread, the difference between the highest and lowest numbers.

As the data is right-skewed, in other words it deviates significantly from a normal distribution, we use Levene's test.

The picture to the right shows 2 box-plots of the sampled data. The box represents the 50% of the data in the middle, and the whiskers the remaining 50% either side. As the right hand side whisker is larger than the left, this is another indication of the skewness of the data. The asterisks indicate outliers within the sampled data.



Hypothesis:	Is there a difference in the variance of the populations from which these samples came of the number of people cycling between 2001 and 2011?
H₀:	there is no difference...
H₁:	there is a difference...
Alpha-value:	5% (industry standard; the risk we are willing to take to get the wrong answer [type I error])
Analysis:	The statistical analysis provides us with a p-value of 0.853
Conclusion:	We fail to reject the H ₀ → <u>there is no difference</u> in the variance of the populations between 2001 and 2011.

2.2 Q: is there a difference in the median number of people cycling between 2001 and 2011?

2.2.1 Is there a difference between the 2 years?

The most common method of comparing two population is by looking at the mean for normally distributed data, and for non-normal data at the median, which is a more useful number for where the majority of data lies.

As this is non-normal data, we look at the median, and we use a Mann-Whitney test.

Hypothesis:	Is there a difference in the populations from which these samples came of median number of people cycling to work between 2001 and 2011?
H₀:	there is no difference...
H₁:	there is a difference...
Alpha-value:	5% (industry standard; the risk we are willing to take to get the wrong answer [type I error])
Analysis:	The statistical analysis provides us with a p-value of 0.3998
Conclusion:	We fail to reject the H ₀ → <u>there is no difference</u> in the median of the populations between 2001 and 2011.

Leicester cycling campaign group

2.2.2 Is it worse in 2011 compared with 2001?

However, when you look at the samples, we have a theory that we have less people cycling in 2011 compared with 2001.

Thus we can make the Mann-Whitney test a single sided test, thus increasing the opportunity to detect whether there is a difference or not.

Hypothesis:	Is the population of cyclists for 2011 less than the population of cyclists in 2001?
H₀:	there is no difference, it is either the same or greater than...
H₁:	there is a difference, the population of 2011 is less than 2001
Alpha-value:	Single sided 5% (industry standard; the risk we are willing to take to get the wrong answer [type I error])
Analysis:	The statistical analysis provides us with a p-value of 0.1999
Conclusion:	We fail to reject the H ₀ → there is still no difference in the median between 2001 and 2011, as a matter of fact, 2001 is not greater than 2011, it is either same (or possibly even greater) than 2001.

2.3 Interim Conclusion

Based on the data provided we conclude that statistically speaking there is no difference in the population from which these samples came in either the spread or the median number of cyclists between 2001 and 2011. Any differences that we can measure between the samples are simply within the margin of errors of the sampling.

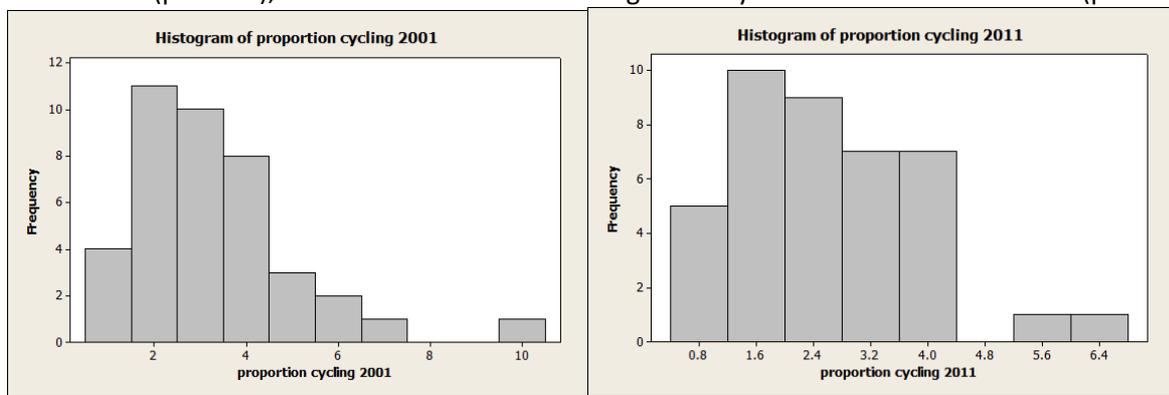
Leicester cycling campaign group

3 Is there a difference in the proportions of people (16-74) cycling between 2001 and 2011?

Rather than looking at the absolute values, which of course will be different from place to place, depending on the size of those places, it might be better to look at the proportion of people cycling. In this case they have assumed that the people are from 16-74 years.

What is not shown is how they knew either the age, or whether these were working people, rather than leisure, students, or other.

In this case when we do the tests for normality, 2001 deviates significantly from a normal distribution ($p=0.021$), and 2011 does not deviate significantly from a normal distribution ($p=0.068$).



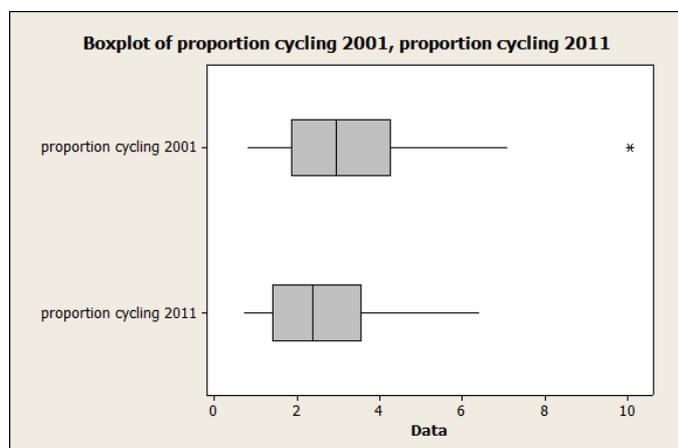
Some of the headline numbers of the samples taken

	2001	2011	Difference (2011-2001)
Number of areas	40	40	0
Mean Proportion of cyclists in the sample	3.25%	2.56%	-0.69%
Median proportion of cyclists in the sample	2.94 %	2.38%	-0.56%

3.1 Is there a difference in the variance of proportions of cyclists between 2001 and 2011?

As before, we do a test for equal variances, and do a Levene's test.

The p-value is 0.240 → we fail to reject the H_0 : there is no difference in the variances of the populations from which these proportions came



Leicester cycling campaign group

3.2 Is there a difference in the proportion of cyclists between 2001 and 2011?

As one of the samples is non-normal, i.e. the data is right skewed, we look again at the median proportions, rather than the mean proportions.

3.2.1 Is there a difference?

We use again the Mann-Whitney test

Hypothesis:	Is there a difference between the populations from which these samples came of the proportion of cyclists (16-74) between 2001 and 2011?
H₀:	there is no difference, it is either the same or greater than...
H₁:	there is a difference, the population of 2011 is less than 2001
Alpha-value:	Single sided 5% (industry standard; the risk we are willing to take to get the wrong answer [type I error])
Analysis:	The statistical analysis provides us with a p-value of 0.0859
Conclusion:	We fail to reject the H ₀ → there is no difference in the median proportions between 2001 and 2011.

A note: as the p-value is getting close to 0.05 (the alpha-value) it may be worth to make this a more powerful test, see below.

3.2.2 Is 2011 worse than 2001?

As we did before, we can make the test more powerful, as we actually think that 2011 is worse than 2001.

Hypothesis:	Is the population from which we calculated the proportion of cyclists (16-74) for 2011 less than the proportion of the population of cyclists in 2001?
H₀:	there is no difference, it is either the same or greater than...
H₁:	there is a difference, the population of 2011 is less than 2001
Alpha-value:	Single sided 5% (industry standard; the risk we are willing to take to get the wrong answer [type I error])
Analysis:	The statistical analysis provides us with a p-value of 0.0429
Conclusion:	We reject the H ₀ → thus we accept the alternative, H ₁ : there is a difference in proportion of cyclists between 2001 and 2011, as a matter of fact, 2011 is indeed worse than 2001.

Note: as the p-value is close to 0.05, it may be prudent to collect more data before making bold conclusions.

3.3 Interim Conclusion

We have demonstrated that there is a difference in proportion of people cycling between 2001 and 2011, 2001 having the higher proportion.

4 Final conclusion

To summarise:

1. There is no difference in the variance nor median between the total number of people cycling between 2001 and 2011
2. There is a difference in the median of the proportion of people cycling between 2001 and 2011, namely worse in 2011, but only just. Thus it would be prudent to collect more data to be sure that there truly is a difference.

~~~~~End of the analysis~~~~~

## 5 Final note and disclaimer.

For further information, please contact the author.

This statistical analysis has not been peer reviewed, and is based on best assumptions of the data provided. The source and validity of the data nor the method of sampling are described, researched, nor verified.

The statistical analysis made herein do not constitute an offer and have been done in good faith. The author cannot bear any liability for decisions by a third party based on this analysis.

Copyright © 2014. All rights reserved. No part of this document may be copied or reproduced in any form or by any means without written permission from the author.

## Leicester cycling campaign group

### 6 Appendix

#### 6.1 Source

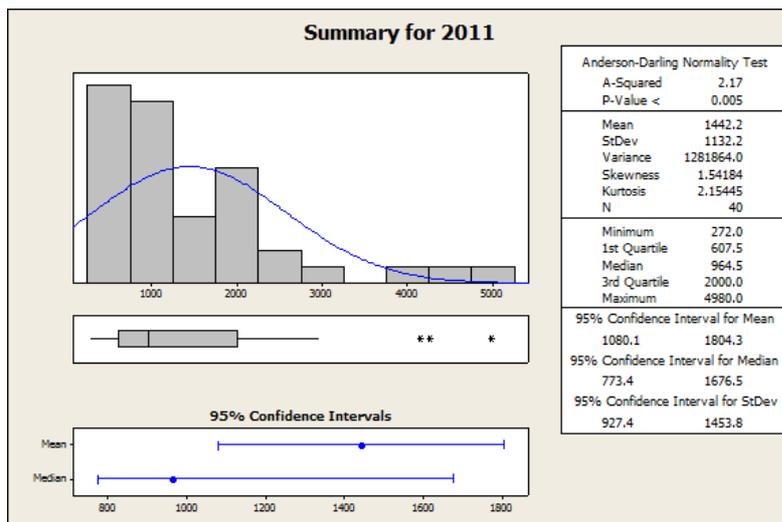
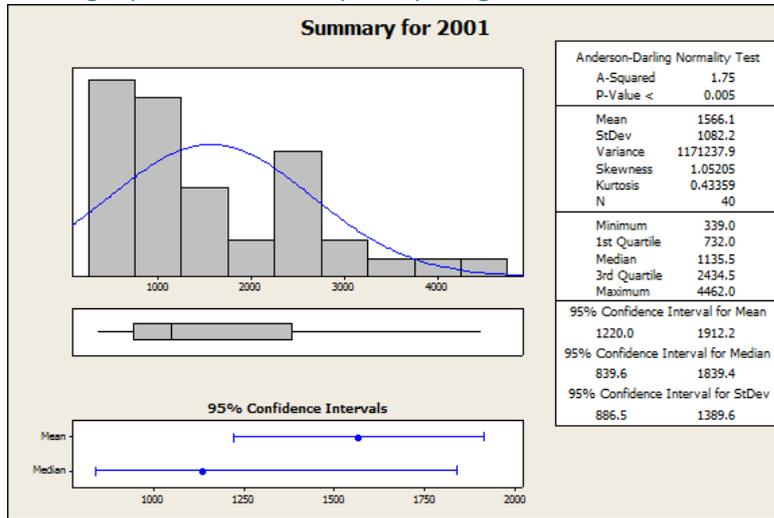
- This spreadsheet provides summary cycle to work data for local authorities comparing the 2001 and 2011 census.
- The 2001 data is from Census Table KS15 available within the 2001 census key statistics tables on the Neighbourhood Statistics Website
- <http://www.neighbourhood.statistics.gov.uk>
- The 2011 data is from Census Table CT0015EW
- <http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-295663>

#### 6.2 Data set for East Midlands

| code      | region        | city                      | Number of residents (16-74) cycling to work |        |                              |                    | Proportion of working residents (16-74) cycling to work |          |                             |
|-----------|---------------|---------------------------|---------------------------------------------|--------|------------------------------|--------------------|---------------------------------------------------------|----------|-----------------------------|
|           |               |                           | 2001                                        | 2011   | absolute change<br>2001-2011 | % change 2001-2011 | 2001 (%)                                                | 2011 (%) | % point change<br>2001-2011 |
| E12000004 | East Midlands |                           | 62,644                                      | 57,689 | -4,955                       | -8%                | 3.3                                                     | 2.7      | -0.6                        |
| E06000018 | East Midlands | Nottingham                | 3,731                                       | 4,257  | 526                          | 14%                | 3.7                                                     | 3.5      | -0.2                        |
| E06000016 | East Midlands | Leicester                 | 4,462                                       | 4,980  | 518                          | 12%                | 4.0                                                     | 3.6      | -0.4                        |
| E07000037 | East Midlands | High Peak                 | 549                                         | 612    | 63                           | 11%                | 1.3                                                     | 1.3      | 0.1                         |
| E07000176 | East Midlands | Rushcliffe                | 1,495                                       | 1,627  | 132                          | 9%                 | 2.9                                                     | 2.9      | 0.1                         |
| E07000035 | East Midlands | Derbyshire Dales          | 339                                         | 357    | 18                           | 5%                 | 1.0                                                     | 1.0      | 0.0                         |
| E07000173 | East Midlands | Gedling                   | 1,032                                       | 1,043  | 11                           | 1%                 | 1.9                                                     | 1.9      | 0.0                         |
| E07000172 | East Midlands | Broxtowe                  | 2,210                                       | 2,220  | 10                           | 0%                 | 4.3                                                     | 4.2      | -0.1                        |
| E07000039 | East Midlands | South Derbyshire          | 755                                         | 754    | -1                           | 0%                 | 1.9                                                     | 1.6      | -0.3                        |
| E07000138 | East Midlands | Lincoln                   | 2,648                                       | 2,640  | -8                           | 0%                 | 7.1                                                     | 5.9      | -1.2                        |
| E07000154 | East Midlands | Northampton               | 2,750                                       | 2,732  | -18                          | -1%                | 2.9                                                     | 2.6      | -0.3                        |
| E07000130 | East Midlands | Charnwood                 | 2,999                                       | 2,969  | -30                          | -1%                | 4.1                                                     | 3.7      | -0.4                        |
| E06000015 | East Midlands | Derby                     | 4,233                                       | 4,158  | -75                          | -2%                | 4.4                                                     | 3.7      | -0.7                        |
| E07000150 | East Midlands | Corby                     | 880                                         | 857    | -23                          | -3%                | 3.6                                                     | 2.8      | -0.8                        |
| E07000153 | East Midlands | Kettering                 | 876                                         | 846    | -30                          | -3%                | 2.1                                                     | 1.8      | -0.3                        |
| E07000129 | East Midlands | Blaby                     | 1,363                                       | 1,306  | -57                          | -4%                | 2.9                                                     | 2.7      | -0.2                        |
| E07000032 | East Midlands | Amber Valley              | 833                                         | 796    | -37                          | -4%                | 1.5                                                     | 1.3      | -0.2                        |
| E07000151 | East Midlands | Daventry                  | 559                                         | 532    | -27                          | -5%                | 1.5                                                     | 1.3      | -0.2                        |
| E07000174 | East Midlands | Mansfield                 | 702                                         | 668    | -34                          | -5%                | 1.7                                                     | 1.4      | -0.3                        |
| E07000135 | East Midlands | Oadby and Wigston         | 849                                         | 794    | -55                          | -6%                | 3.2                                                     | 3.0      | -0.2                        |
| E07000038 | East Midlands | North East Derbyshire     | 349                                         | 326    | -23                          | -7%                | 0.8                                                     | 0.7      | -0.1                        |
| E07000175 | East Midlands | Newark and Sherwood       | 2,430                                       | 2,103  | -327                         | -13%               | 5.1                                                     | 3.9      | -1.3                        |
| E07000134 | East Midlands | North West Leicestershire | 884                                         | 759    | -125                         | -14%               | 2.1                                                     | 1.6      | -0.5                        |
| E07000140 | East Midlands | South Holland             | 2,000                                       | 1,711  | -289                         | -14%               | 5.7                                                     | 4.1      | -1.6                        |
| E07000139 | East Midlands | North Kesteven            | 2,436                                       | 2,008  | -428                         | -18%               | 5.5                                                     | 3.8      | -1.7                        |
| E07000171 | East Midlands | Bassetlaw                 | 1,609                                       | 1,320  | -289                         | -18%               | 3.5                                                     | 2.5      | -1.0                        |
| E07000156 | East Midlands | Wellingborough            | 644                                         | 527    | -117                         | -18%               | 1.8                                                     | 1.4      | -0.4                        |
| E07000131 | East Midlands | Harborough                | 1,097                                       | 886    | -211                         | -19%               | 2.8                                                     | 2.0      | -0.7                        |
| E07000170 | East Midlands | Ashfield                  | 1,381                                       | 1,109  | -272                         | -20%               | 2.8                                                     | 2.0      | -0.8                        |
| E07000137 | East Midlands | East Lindsey              | 2,519                                       | 2,019  | -500                         | -20%               | 4.7                                                     | 3.6      | -1.1                        |
| E07000036 | East Midlands | Erewash                   | 2,330                                       | 1,866  | -464                         | -20%               | 4.4                                                     | 3.4      | -1.0                        |
| E07000132 | East Midlands | Hinckley and Bosworth     | 1,538                                       | 1,212  | -326                         | -21%               | 3.0                                                     | 2.3      | -0.7                        |
| E07000034 | East Midlands | Chesterfield              | 657                                         | 515    | -142                         | -22%               | 1.5                                                     | 1.1      | -0.5                        |
| E07000136 | East Midlands | Boston                    | 2,523                                       | 1,976  | -547                         | -22%               | 10.1                                                    | 6.4      | -3.7                        |
| E07000133 | East Midlands | Melton                    | 741                                         | 570    | -171                         | -23%               | 3.1                                                     | 2.2      | -0.9                        |
| E07000141 | East Midlands | South Kesteven            | 2,489                                       | 1,878  | -611                         | -25%               | 4.1                                                     | 2.8      | -1.3                        |
| E07000152 | East Midlands | East Northamptonshire     | 729                                         | 548    | -181                         | -25%               | 1.9                                                     | 1.3      | -0.6                        |
| E06000017 | East Midlands | Rutland                   | 811                                         | 606    | -205                         | -25%               | 4.9                                                     | 3.4      | -1.5                        |
| E07000142 | East Midlands | West Lindsey              | 1,174                                       | 860    | -314                         | -27%               | 3.3                                                     | 2.0      | -1.3                        |
| E07000155 | East Midlands | South Northamptonshire    | 651                                         | 470    | -181                         | -28%               | 1.5                                                     | 1.0      | -0.5                        |
| E07000033 | East Midlands | Bolsover                  | 387                                         | 272    | -115                         | -30%               | 1.3                                                     | 0.8      | -0.5                        |

## Leicester cycling campaign group

### 6.3 graphical Summary of cycling 2001 and 2011



### 6.4 Median analysis

#### Mann-Whitney Test and CI: 2001, 2011

|      | N  | Median |
|------|----|--------|
| 2001 | 40 | 1135.5 |
| 2011 | 40 | 964.5  |

Point estimate for ETA1-ETA2 is 124.5

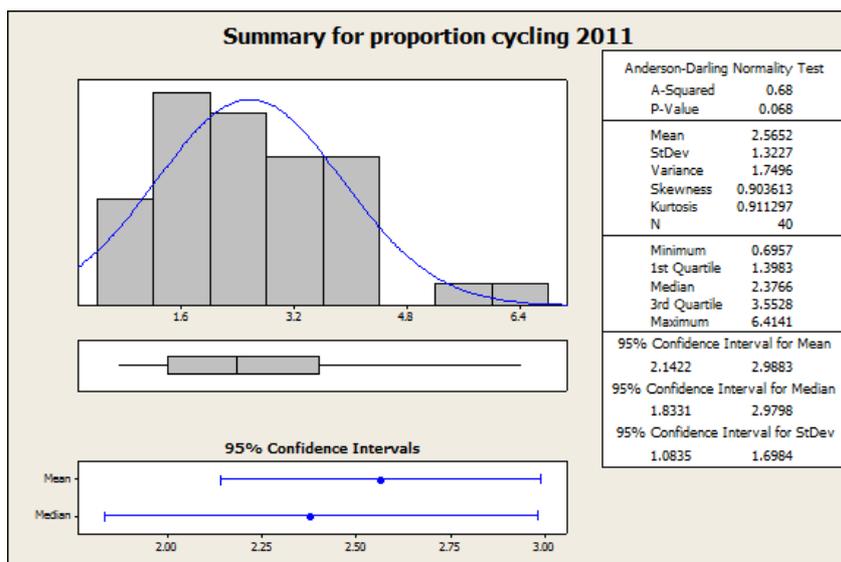
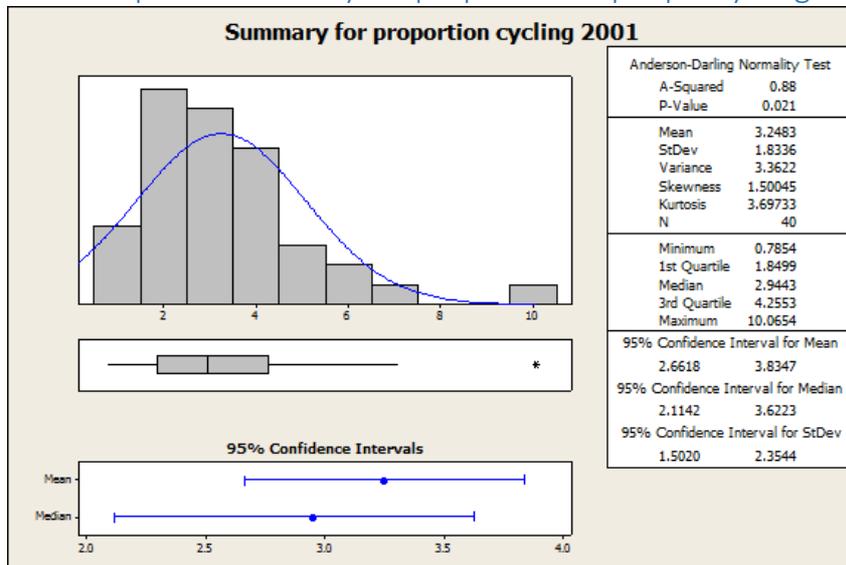
95.1 Percent CI for ETA1-ETA2 is (-201.8,475.9)

W = 1708.0

Test of ETA1 = ETA2 vs ETA1 > ETA2 is significant at 0.1999

Leicester cycling campaign group

6.5 Graphical Summary for proportion of people cycling



6.6 Variance in proportion of cyclists

Test and CI for Two Variances: proportion cycling 2001, proportion cycling 2011

Method

Null hypothesis  $\sigma(\text{proportion cycling 2001}) / \sigma(\text{proportion cycling 2011}) = 1$   
 Alternative hypothesis  $\sigma(\text{proportion cycling 2001}) / \sigma(\text{proportion cycling 2011}) \text{ not } = 1$   
 Significance level  $\alpha = 0.05$

Statistics

| Variable                | N  | StDev | Variance |
|-------------------------|----|-------|----------|
| proportion cycling 2001 | 40 | 1.834 | 3.362    |
| proportion cycling 2011 | 40 | 1.323 | 1.750    |

Ratio of standard deviations = 1.386  
 Ratio of variances = 1.922

## Leicester cycling campaign group

---

95% Confidence Intervals

| Distribution of Data | CI for StDev Ratio | CI for Variance Ratio |
|----------------------|--------------------|-----------------------|
| Normal               | (1.008, 1.906)     | (1.016, 3.633)        |
| Continuous           | (0.838, 1.842)     | (0.703, 3.391)        |

Tests

| Method                                | DF1      | DF2       | Test Statistic | P-Value      |
|---------------------------------------|----------|-----------|----------------|--------------|
| F Test (normal)                       | 39       | 39        | 1.92           | 0.045        |
| <b>Levene's Test (any continuous)</b> | <b>1</b> | <b>78</b> | <b>1.40</b>    | <b>0.240</b> |

### 6.7 Median of Proportion of cyclists

#### Mann-Whitney Test and CI: proportion cycling 2001, proportion cycling 2011

|                         | N  | Median |
|-------------------------|----|--------|
| proportion cycling 2001 | 40 | 2.944  |
| proportion cycling 2011 | 40 | 2.377  |

Point estimate for ETA1-ETA2 is 0.543

95.1 Percent CI for ETA1-ETA2 is (-0.067,1.199)

W = 1799.0

Test of ETA1 = ETA2 vs ETA1 > ETA2 is significant at 0.0429