|  |
| --- |
| Changing Places: Lessons using data skills |

|  |
| --- |
| **Lesson 4**: The geography of health: How does life expectancy vary across London?  |

**Lesson objective**

To test the strength of the relationship through the use of scatter graphs and Spearman’s rank correlation coefficient between life expectancy and distance across London.

**Setting the scene**

Maps can be a very powerful way of highlighting inequalities across space. In 2013, Danny Dorling wrote the book *The 32 Stops* to investigatehow life expectancy of the people living along the Central Tube Line varies across London. A correlation of life expectancy and distance from the city can provoke discussions about the quality of life for Londoners. Danny Dorling was not the first to investigate inequalities in the capital city. Charles Booth’s survey into life and labour in London between1886 and1903 can be accessed [here](http://booth.lse.ac.uk) and John Snow’s map of the 1854 cholera outbreak, recreated by ESRI, is [here](http://www.arcgis.com/apps/PublicInformation/index.html?appid=d7deb67f810d46dfacb80ff80ac224e9). The human outcome of deprivation and high levels of communicable diseases is a low life expectancy.

This lesson enables students to compare life expectancy across London at ward level, using a tube line as the transect from which data is selected. Danny Dorling followed the Central Line for this purpose in his book. This example follows the District Line from Upminster in the east to South Kensington. There are some limitations to collecting data using tube stations as sampling points and ward level data. The distance between tube stations is not regularly spaced therefore the samples may not be wholly representative of the area. Ward level data is an average so hides pockets of the least and most deprived neighbourhoods. A solution could be to use data from the smaller Lower-layer Super Output Areas (LSOAs), which represent around 1,500 residents and 650 households.



**The data**

Life expectancy for males and females from the *Ward Profiles* of London are mapped on the London Datastore [website](http://londondatastore-upload.s3.amazonaws.com/instant-atlas/ward-profiles-html/atlas.html) and Transport for London have a tube map for London [here](http://content.tfl.gov.uk/standard-tube-map.pdf). In the Microsoft Excel spreadsheet for Lesson 4 you can find the data for male and female life expectancy extracted from the Greater London Authority (GLA) calculations and based on Office for National Statistics data (ONS) Census data.

**Figure 1** Life expectancy across London



**How the data was collected**

1. Open the Ward Profiles interactive map on the London Datastore [website](http://londondatastore-upload.s3.amazonaws.com/instant-atlas/ward-profiles-html/atlas.html).

2. Click on the Male Life Expectancy 2009-2013 measure. This can be found under the HEALTH domain on the left-hand legend.

3. Start at the ward furthest to the east (Upminster), zoom in to find the tube station, then record the life expectancy as you travel west along the District Line. Any tube line could have been chosen. Data could have been collected all the way to Richmond (spatial sampling choice) or at a different scale, either the LSOA or borough level.

**Tasks**

1. **Describe how life expectancy changes as you travel towards Central London?**

Use the above graph and explore the data at ward level from the London Datastore [website](http://londondatastore-upload.s3.amazonaws.com/instant-atlas/ward-profiles-html/atlas.html).

1. **Use the Microsoft Excel data to see whether there is a correlation between life expectancy and place across London.**
2. Choose to focus on the male or female life expectancy data. Create a scatter graph to see if it looks like there is a relationship between life expectancy as you travel across London. Add a line of best-fit if it is appropriate.
3. State the null hypothesis for this test, for example there is no relationship between life expectancy and where you live in London. The alternative hypothesis is that there is a relationship. Rank both sets of data from highest to lowest, the stops are already in order. In the case of a joint rank find the average e.g. if two values occupy rank 4 and 5 they take rank 4.5, if three values occupy rank 6,7 and 8 they take rank 7. The first 3 are ranked.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Male life expectancy** | **District Line****(already ranked)** | **Ranked****male life ex** | **Difference** | **Difference (d2)** |
| 80.8 | 31 |  |  |  |
| 80.4 | 30 |  |  |  |
| 80.5The next number to be ranked will be 80.8. It is unique so will take the rank place 11 | 29 |  |  |  |
| 78.9 | 28 |  |  |  |
| 77.1 | 27 |  |  |  |
| 75.0 | 26 |  |  |  |
| 76.8 | 25 |  |  |  |
| 78.9 | 24 |  |  |  |
| 80.2 | 23 |  |  |  |
| 76.2 | 22 |  |  |  |
| 78.5 | 21 |  |  |  |
| 78.3 | 20 |  |  |  |
| 77.2 | 29 |  |  |  |
| 80.1 | 28 |  |  |  |
| 77.2 | 17 |  |  |  |
| 78.5 | 16 |  |  |  |
| 76.1 | 15 |  |  |  |
| 76.1 | 14 |  |  |  |
| 77.4 | 13 |  |  |  |
| 84.3 | 12 | 4 | 8 | 64 |
| 84.3 | 11 | 4 | 7 | 49 |
| 84.3 | 10 | 4 | 6 | 36 |
| 84.3 | 9 | 4 | 5 | 25 |
| 84.3 | 8 | 4 | 4 | 16 |
| 83.5 | 7 | 8.5 | 1.5 | 2.25Σ means sum. Sum all the d2 values and add the total to this box |
| 83.5 | 6 | 8.5 | 2.5 | 5.25 |
| 83.5 | 5 | 8.5 | 3.5 | 12.25 |
| 83.5 | 4 | 8.5 | 4.5 | 20.25 |
| 80.7 | 3 |  |  |  |
| 79.6 | 2 |  |  |  |
| 86.4 | 1 | 1 | 0 | 0 |
|  |  |  |  | Σd2 |

1. Use the formula for Spearman’s rank, below, to work out the correlation.

d is the difference between ranks



n is the number of observation

Compare the calculated rs with the critical values in the statistical table to see whether there is a relationship in the data.

|  |  |
| --- | --- |
| **N** | **Significance level** |
| 95 % | 99% |
| 31 | 0.30 | 0.42 |

If the relationship is strong and the calculated rs  exceeds the critical values, we can say that we are 95% or 99% sure that there is a relationship between the sets of data.

There are limitations with using Spearman’s rank. The sample size should be close to 30; seven is a minimum. Correlation does not necessarily imply causation e.g. there is a correlation between reading ability and shoe size, but this does not mean large shoes cause good reading skills!

**Plenary**

Go back to the key question and ask students to write a mini-saga, 50 words summarising what they have learnt during the lesson.

**Take it further**

Complete the process again, but this time use data for the opposite gender. Alternatively you could choose a different tube line to see if there is a correlation between life expectancy between different places, or collect data about deprivation, income or education following the District Line. You could then write up your results as a short report.

**Reference**

Danny Dorling (2013) *The 32 Stops.* London: Penguin Books.

Critical values for Spearman’s Rank are found widely on the Internet, [here](http://www.york.ac.uk/depts/maths/tables/spearman.pdf) is one from the University of York.

**Spearman’s Rank worked example - answer**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Male life expectancy** | **District Line****(already ranked)** | **Ranked****Male life ex** | **Difference** | **Difference (d2)** |
| 80.8 | 31 | 11 | 20 | 400 |
| 80.4 | 30 | 14 | 16 | 256 |
| 80.5The next number to be ranked will be 80.8. It is unique so will take the rank place 11 | 29 | 13 | 16 | 256 |
| 78.9 | 28 | 18.5 | 9.5 | 90.25 |
| 77.1 | 27 | 26 | 1 | 1 |
| 75.0 | 26 | 31 | 5 | 25 |
| 76.8 | 25 | 27 | 2 | 4 |
| 78.9 | 24 | 18.5 | 5.5 | 30.25 |
| 80.2 | 23 | 15 | 8 | 64 |
| 76.2 | 22 | 28 | 6 | 36 |
| 78.5 | 21 | 20.5 | 0.5 | 0.25 |
| 78.3 | 20 | 22 | 2 | 4 |
| 77.2 | 29 | 24.5 | 4.5 | 20.25 |
| 80.1 | 28 | 16 | 12 | 144 |
| 77.2 | 17 | 24.5 | 7.5 | 56.25 |
| 78.5 | 16 | 20.5 | 4.5 | 20.25 |
| 76.1 | 15 | 29.5 | 14.5 | 210.25 |
| 76.1 | 14 | 29.5 | 15.5 | 210.25 |
| 77.4 | 13 | 23 | 10 | 100 |
| 84.3 | 12 | 4 | 8 | 64 |
| 84.3 | 11 | 4 | 7 | 49 |
| 84.3 | 10 | 4 | 6 | 36 |
| 84.3 | 9 | 4 | 5 | 25 |
| 84.3 | 8 | 4 | 4 | 16 |
| 83.5 | 7 | 8.5 | 1.5 | 2.25Σ means sum. Sum all the d2 values and add the total to this box |
| 83.5 | 6 | 8.5 | 2.5 | 5.25 |
| 83.5 | 5 | 8.5 | 3.5 | 12.25 |
| 83.5 | 4 | 8.5 | 4.5 | 20.25 |
| 80.7 | 3 | 12 | 9 | 81 |
| 79.6 | 2 | 17 | 15 | 225 |
| 86.4 | 1 | 1 | 0 | 0 |
|  |  |  |  | Σd2 2464 |



 rs = 1 – 6 x 2464 rs = 1 – 14,784/29,760 rs = 0.50

 31(961–1)

|  |  |
| --- | --- |
| **N** | **Significance level** |
| 95 % | 99% |
| 31 | 0.30 | 0.42 |

In this example, we can see that there is a fairly strong correlation between life expectancy and distance along the tube line as the calculated rs  of 0.50 exceeds the critical values of 0.30 and 0.42, we can say that we are 95% or 99% sure that there is a relationship between the sets of data. Thus we can reject the null hypothesis that there is no relationship between life expectancy and where you live in London and accept the alternative hypothesis that there is a relationship.

The much harder task is to explain why there is inequality in life expectancies across London.