

Modelling Voter Turnout in Dublin

Introduction

In this exercise we'll predict the variation in voter turnout for the 2002 General election in the 322 electoral divisions in Dublin. A spatial polygons data frame is available for you, with the variables of interest in its data slot.

The data

The data are described below. The response variable, *GenEl2004*, is to be modelled using a set of socio-economic indicators taken from the 2002 Census of Population. Voter turnout is of interest because it's an indicator of the electorates participation in the democratic process. Low turnout reflects a number of issues, including a sense of dis-enfranchisement.

The *DubVoter* data are available in the R *GWmodel* library.

The socio-economic indicators are thought to be potential drivers of turnout variability. They are described below.

DubVoter	Voter turnout data in Greater Dublin(SpatialPolygonsDataFrame)
----------	--

Description

Voter turnout and social characters data in Greater Dublin for the 2002 General election and the 2002 census. Note that this data set was originally thought to relate to 2004, so for continuity we have retained the associated variable names.

Usage

```
data(DubVoter)
```

Format

A SpatialPolygonsDataFrame with 322 electoral divisions on the following 11 variables.

DED_ID a vector of ID

X a numeric vector of x coordinates

Y a numeric vector of y coordinates

DiffAdd percentage of the population in each ED who are one-year migrants (i.e. moved to a different address 1 year ago)

LARent percentage of the population in each ED who are local authority renters

SC1 percentage of the population in each ED who are social class one (high social class)

Unempl percentage of the population in each ED who are unemployed

LowEduc percentage of the population in each ED who are with little formal education

Age18_24 percentage of the population in each ED who are age group 18-24

Age25_44 percentage of the population in each ED who are age group 25-44

Age45_64 percentage of the population in each ED who are age group 45-64

GenEl2004 percentage of population in each ED who voted in 2004 election

Details

Variables are from DubVoter.shp.

References

Kavanagh A (2006) Turnout or turned off? Electoral participation in Dublin in the early 21st Century. *Journal of Irish Urban Studies* 3(2):1-24

Harris P, Brunsdon C, Charlton M (2011) Geographically weighted principal components analysis. *International Journal of Geographical Information Science* 25 (10):1717-1736

Your task

The task is find a set of variables which best predicts turnout variation. If you are using OLS regression then you may need to examine the residuals for spatial structure. The *lm.morantest()* function is useful here: it's available in the *spdep* library.

Be aware that this is real data – the predictor variables were suggested by an expert in electoral geography who is not an expert in statistical modelling. There is suspicion that collinearity might be an issue here. What steps might you use to deal with it¹?

You might find geographically weighted techniques of use here.

Submission

A brief 1500 word report, with a suitable introduction and conclusion, and sections detailing (i) the task, (ii) your approach, (iii) the data, and (iv) your analysis will suffice. The submission date will be agreed in class, and I'll need hardcopy (I can't print colour).

Zafra
Task
Analysis
Conclusion

3 weeks?

¹ I found Belsey DA, Kuh E and Welsch RE, 1980, *Regression Diagnostics: identifying influential data and sources of collinearity*, Hoboken NJ: Wiley very useful. I'll put some R code up for you.