

Q-Step Internship: The UK Data Service

Kun Yu, BA Linguistics & Sociology



In summer 2015 I undertook a Q-step internship with UK Data Service in Manchester. The UK Data Service is

funded by the Economic and Social Research Council and provides access and support for a wide range of key social and economic data.

Objectives

The UK Data Service creates teaching datasets for University lecturers (and other University teachers) to use with students. These teaching datasets are based on the quantitative crosssectional datasets held by the UK Data Service – for instance the British Social Attitudes Survey and the Crime Survey for England and Wales.

The aim of this internship is to:

- Create new teaching datasets and accompanying user guides
- Update current teaching datasets and accompanying user guides
- Create accompanying worksheets/exercises that use the teaching datasets to work through correlation and regression techniques (for use by lecturers/teachers and students)



Methodology

Used SPSS and STATA to do secondary analysis on UKDS teaching datasets.

Used correlation and regression techniques to create Health Survey for England relative worksheets.

Updated Crime Survey for England and Wales, English Housing Survey datasets with user guide and workbook

Outcome

English Housing Survey 2012-2013
Household Data
Teaching Dataset
ESDS Government
User Guide

Contact:
ESDS Government
The Centre for Census and Survey Research (CCSR)
University of Manchester, Manchester, M13 9PL
Tel: 0161 275 3980 (ESDS Government helpdesk)
Email: helpdesk@esds.ac.uk

Worksheet: Scatter Plots and Correlation

This worksheet shows you how to produce scatter plots and correlation to measure the association between two continuous variables. You will also consider a case where the data is ordinal. Following the worksheet, you will:

- generate histograms and create summary statistics to look at the distribution of each variable separately
- produce scatter plots to show the relationship between the dependent variable and each of the explanatory variables separately
- generate Correlation Coefficients for each pair of variables

The worksheet uses teaching data from the 2011 Health Survey for England from UK Data Service.

Task – Produce scatter plots and correlations

To produce the scatter plots and correlations about blood pressure. Specifically, we would like to find out:

1. Why blood pressure can reflect an individual's health condition?
2. What are the factors associated with an individual's blood pressure?
3. What relationship do the scatterplots suggest?
4. How to interpret the value of correlation for each pair of variables?

Follow instructions marked with a ➔ on your computer

1. Open the dataset

Before you start, you need to open the dataset in SPSS

- ➔ Go to <http://www.ukdataservice.ac.uk/>
- ➔ Download the Health Survey for England 2011 teaching data
- ➔ Open SPSS
- ➔ Open the dataset from within SPSS

1.1. Apply the weight

There you need to apply the weight. Survey data comes from a sample of a population. Weights help make the sample data representative of the population. The weights adjust a sample to make it more representative of the population. It was designed to reflect.

Worksheet: Simple Regression

This worksheet shows you how to use simple linear regression, that is to build a model that describes the linear relationship between an independent variable and a dependent variable. Following the worksheet, you will:

- generate statistics for each type of variable
- use simple regression to predict values of the dependent variable

The worksheet uses teaching data from the 2011 Health Survey for England from UK Data Service.

Task – Regression

To use simple regression and to build a model to predict systolic blood pressure (our dependent variable) using various different explanatory variables. Specifically, we would like to find out:

- Which explanatory variable shows the strongest correlation with the dependent variable (smoking)?
- How to interpret the regression equation and the value of regression for each pair of variables?

Follow instructions marked with a ➔ on your computer

1. Open the dataset

Before you start, you need to open the dataset in SPSS

- ➔ Go to <http://www.ukdataservice.ac.uk/>
- ➔ Download the Health Survey for England 2011 teaching data
- ➔ Open SPSS
- ➔ Open the dataset from within SPSS

1.1. Apply the weight

There you need to apply the weight. Survey data comes from a sample of a population. Weights help make the sample data representative of the population. The weights adjust a sample to make it more representative of the population. It was designed to reflect.

The analysis you will do today uses data collected in a special 'census' which formed part of the 2011 Census framework. The dataset comes with a special weight variable that corrects for non-response bias in the survey itself. Called 'wt_100'.

Session 2-4 Workbook
Exploring Housing Benefits with the English Housing Survey

In the first session, we came up with questions around housing benefits: e.g. who gets it, how much they get, do people in London get more on average than people in other parts of the country etc. Now we'll use the EHS 2010-2013 to examine the data!

Open SPSS 20 and open the data.

In the cluster we use for these sessions you should be able to SPSS as follows

Click on the **GET DATA** button in the bottom left corner of your screen

Select **All Programs** → **IBM SPSS Statistics** → **IBM SPSS Statistics 20**

Variables and datasets

In the list below are the variables we'll be using today. They are contained in two files. The first one is called 'interviews10.sav' and the second is called 'interviews13.sav'. Both contain variables about households from the English Housing Survey 2010-2013.

Name	Label	File
household	Household (HWP) - partner receives any housing benefit?	interviews10.sav
tenure1	Tenure Group 1	interviews10.sav
hsepg01	Household type - 6 categories	interviews10.sav
amb010	Weekly housing benefit	interviews10.sav
hsepg02	Number of persons in the household	interviews10.sav
agep01	Age of HWP - continuous	interviews10.sav
agep02	Age of household reference person - 6 band	interviews10.sav
HYEAR01	Household gross annual income	interviews10.sav
hsepg03	Total net income household actually has	interviews10.sav
excode	ESHS case number	interviews10.sav
intv010	MSD 2010/2013 decile ranking of areas (lower layer SGA)	interviews10.sav

Key Learning

During my time working for the UK Data Service, I updated Crime Survey for England and Wales and English Housing Survey teaching datasets and created the relevant user guide and workbook. I also used correlation and regression techniques to create Health Survey for England worksheets. From this I practiced my secondary analysis skill. I learned how to use STATA to create my dataset, how to analyse that data and how to present it in the worksheets. I believe this has given me experience of many transferable skills which I could apply to several jobs I may consider upon graduation. Apart from that, I provided students' feedback on how to improve the UK Data Service resources for students and advice for their new app. I was also able to attend the Census Conference (organised by the UK Data Service) in the University of Manchester, where I met with census preachers and have a better understanding about census data.

UK Data Service



Contact
kun.yu-2@student.manchester.ac.uk



A step-change in
quantitative social
science skills
Funded by the
Nuffield Foundation,
ESRC and HEFCE