

Final Data Project - Data Management with R

Kris Best

15 December 2017

Introduction

This project seeks to identify factors that contributed to individuals being found guilty of and/or being executed for practicing witchcraft in early modern Scotland. It makes use of the *Scottish Witchcraft Database*, which was compiled by four researchers at the University of Edinburgh in 2003 and contains demographic and situational data on the cases of 3,219 individuals (both male and female) who were accused and tried for witchcraft in Scotland between 1563 and 1736. This paper uses logit and linear probability regression models to investigate the probability that an individual would be found guilty of and ultimately executed for witchcraft as a function of their sex, marital status and the specific motives for their accusation.

The public Github repository for this project can be found at: <https://github.com/krisbest/hertie-dataproject>

Data Preparation

Structure of the dataset

The *Scottish Witchcraft Database* is organised at three levels. The top ‘*accused*’ level contains the characteristics of the individual accused persons and their family members, including brief demographic information where available. The second ‘*case*’ level contains details regarding the case brought against the accused, e.g. the date and content of the accusations, number of accusers, and particular elements of witchcraft involved. The first and second level are linked by the key *accusedref*, which refers to the individual accused. The third ‘*trial*’ level of the database contains information related to the trial process, including the proceedings and outcome. As an individual accused may have multiple trials, each observation at the third level is assigned a *trialref* and is linked to the second level of the database through the key *caseref*.

Data preparation

The original data was provided in a zipped file that contained 38 separate tables in CSV format. These tables were first unzipped and read into R as dataframes. These 38 dataframes were reviewed, and any tables and columns that were not useful for the analysis were dropped from future steps. Three tables were used going forward that correspond to the ‘overview’ tables used at each level of the database: *accused*, *case*, and *trial*.

In addition to dropping unused tables and variables, some of the existing factor variables had to be recoded at this stage into simpler categories in order to make them more amenable for future stages of the analysis. For example, marital status and socioeconomic status were recoded from five and seven categories respectively into three categories each. More substantive recoding was undertaken for the indicator variables corresponding to the researchers’ assessment of the *primary motive or characterisation* of the accusation of witchcraft.

		Original motives (estimated by researchers)
Recoded motive	Description	
Association	Individual was primarily accused for their association with another witch	Consulting a witch, being implicated by another accused witch

Recoded motive	Description	Original motives (estimated by researchers)
Evil Magic	Individual was primarily accused for engaging in ‘evil’ or malevolent forms of magic	Demonic elements, demonic possession, maleficium
Good Magic	Individual was primarily accused for engaging in ‘good’ forms of magic	Fairies, folk healing, white magic
Non-Magic	Individual was primarily accused for suspected non-magic-related reasons	Neighbourhood dispute, property motive, political motive, treason, refused charity

The three ‘overview’ tables were then imported into a SQLite database. The three tables were joined by a database query using the relevant keys at each level (*accusedref,caseref*) and then pulled into a set of tibbles for local use in R.

Descriptive Statistics

The joined dataset contained 3,594 observations on 3,219 individuals accused of witchcraft, with each observation corresponding to an accused individual and a trial. Some accused individuals were tried multiple times and are therefore subject to multiple observations; in contrast, some accused individuals were not tried at all, and these observations relate only to the accusation. Note that despite the high number of observations, only a small number (often a few hundred or less) actually contained detailed data on the variables of interest in this analysis.

Demographic statistics

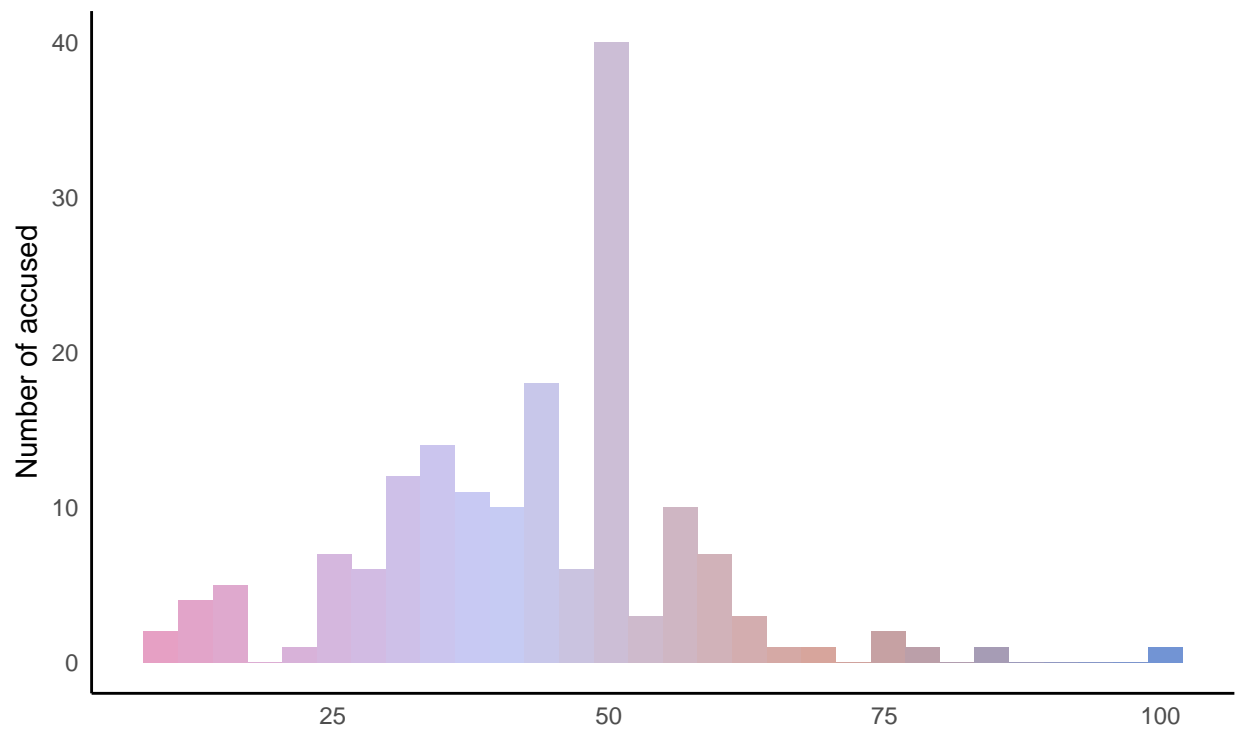
As expected from common stereotypes of witchhunts in early modern Europe, the gender balance of the accused individuals is highly skewed: 85% of the accused were female, with the counts by sex shown in the table below.

```
##
## Female   Male
##    2702    468
```

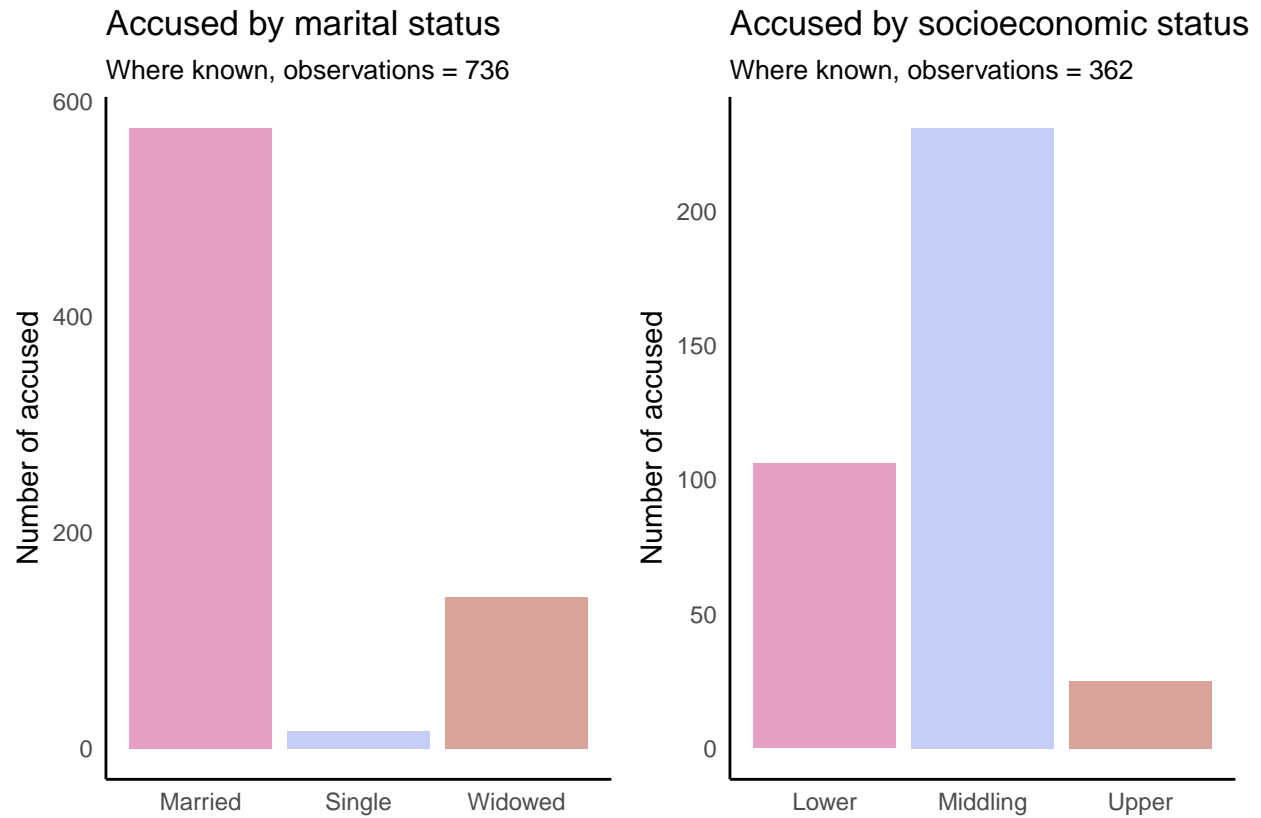
Age statistics were available from only 166 of the accused individuals and ranged from a minimum of 9 to a maximum of 100, with a mean age of 43. As the following figure shows, witchcraft accusations primarily affected those in their mid-20s or older, with a distinct peak in middle age. Note that the spike at precisely age 50 reflects the fact that many of the age statistics were estimations on the part of the researchers.

Distribution of the accused by age

Where known, observations = 166



The following figures show the distribution of the accused individuals by marital status and socioeconomic status, where these statistics were available. Individuals accused of witchcraft were overwhelmingly married, with most hailing from the middle or lower classes.

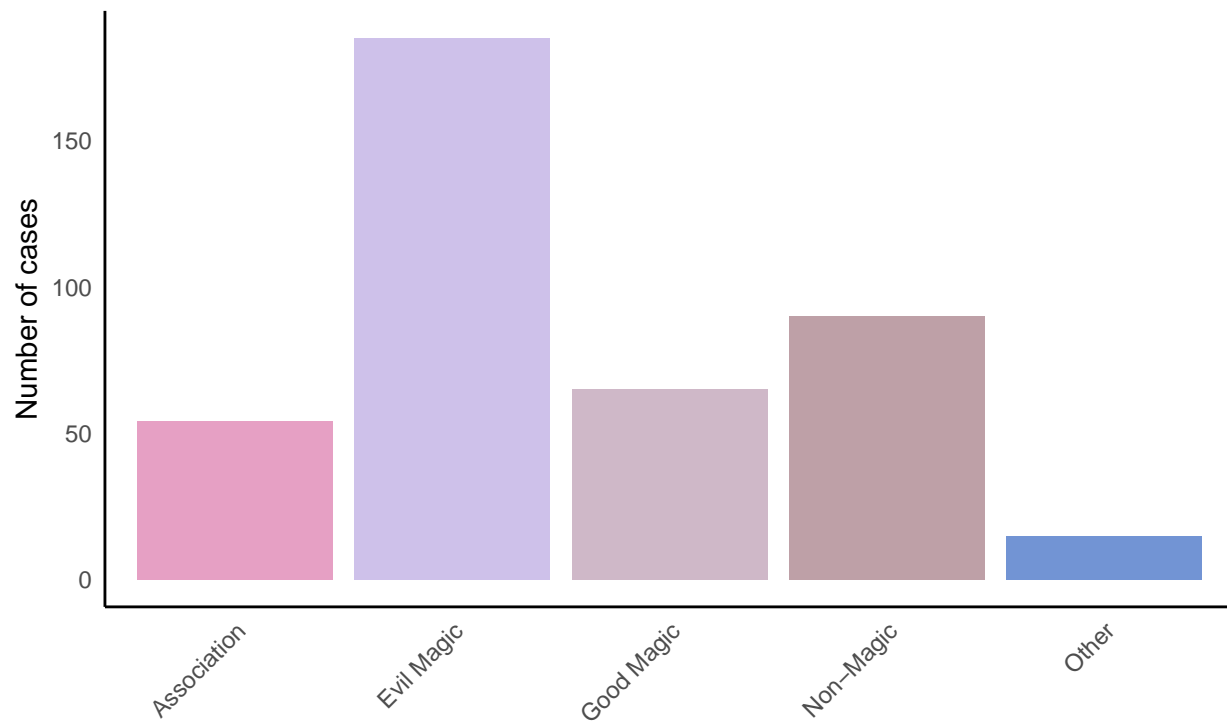


Descriptive statistics of the case and outcome

Individuals were accused of witchcraft for a variety of different reasons. The following figure shows the distribution of primary motives or characteristics of the accusations where this was known to or could be accurately estimated by the researchers (409 cases).

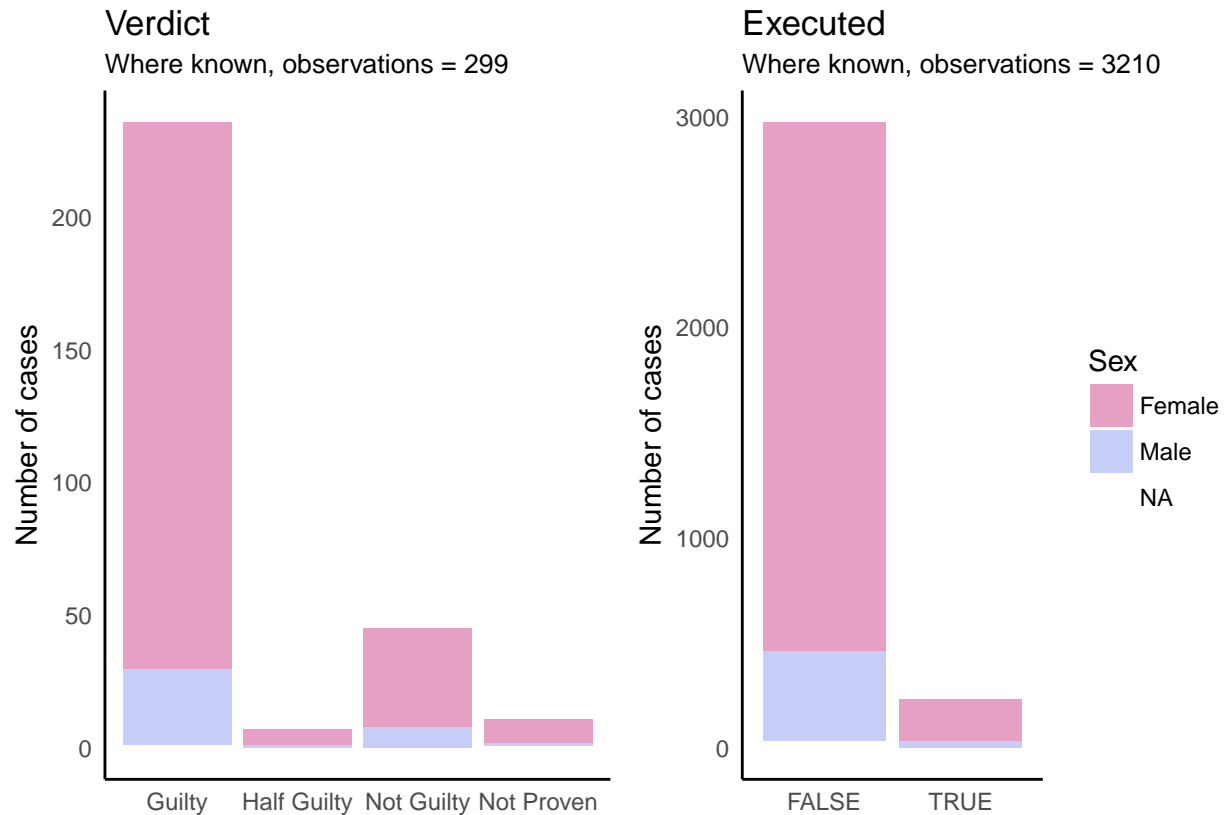
Motives for accusation (as estimated by researchers)

Where known, observations = 409



As the figure above indicates, the most common characteristics of a witchcraft accusation related to the practice of 'evil' magic, i.e. to demonic possession, demonic activity, or maleficium (malevolent sorcery). However, 90 accusations were categorised by the researchers as having non-magic-related motives. Of the non-magic-related accusations, the researchers estimated that 65 related to a neighbourhood dispute and 10 related to 'refused charity', e.g. that the individual refused to share their food, shelter or wealth. Among the few cases that listed 'other' motives, the researchers listed the commission of crimes (notably murder and theft) as well as 'sexual misconduct', incest and rape. In the majority of cases (more than 3000), the primary motivation of the accusation could not be determined.

The following figures show the ultimate verdict of the trial (where known) as well as whether or not the accused individual was executed as a result. Both figures are further broken down according to the sex of the accused individual.



The accused individual was found guilty in 236 out of the 299 cases where verdict data exists, i.e. nearly 80% of the time. A total of 230 accused individuals were executed out of 3210 accused for which execution data exists. Note that this does not mean that almost all of the 236 individuals found guilty were executed: in many cases the data indicated that an execution took place where there was no corresponding data for the verdict.

Of the 230 individuals who were executed, details about the method of execution exist in 175 cases. The most common method of execution noted in 125 cases was 'strangle and burn', followed by burning alone in 46 cases. The remaining few execution cases are divided between beheadings (3) and hangings (1).

Model Specification

Logit, linear probability model

Had to recode variables due to lack of variation/small sample size Checked size of non-NAs for each variable

Model Results

Table with all four models

Coefplots for both sets

Table 2:

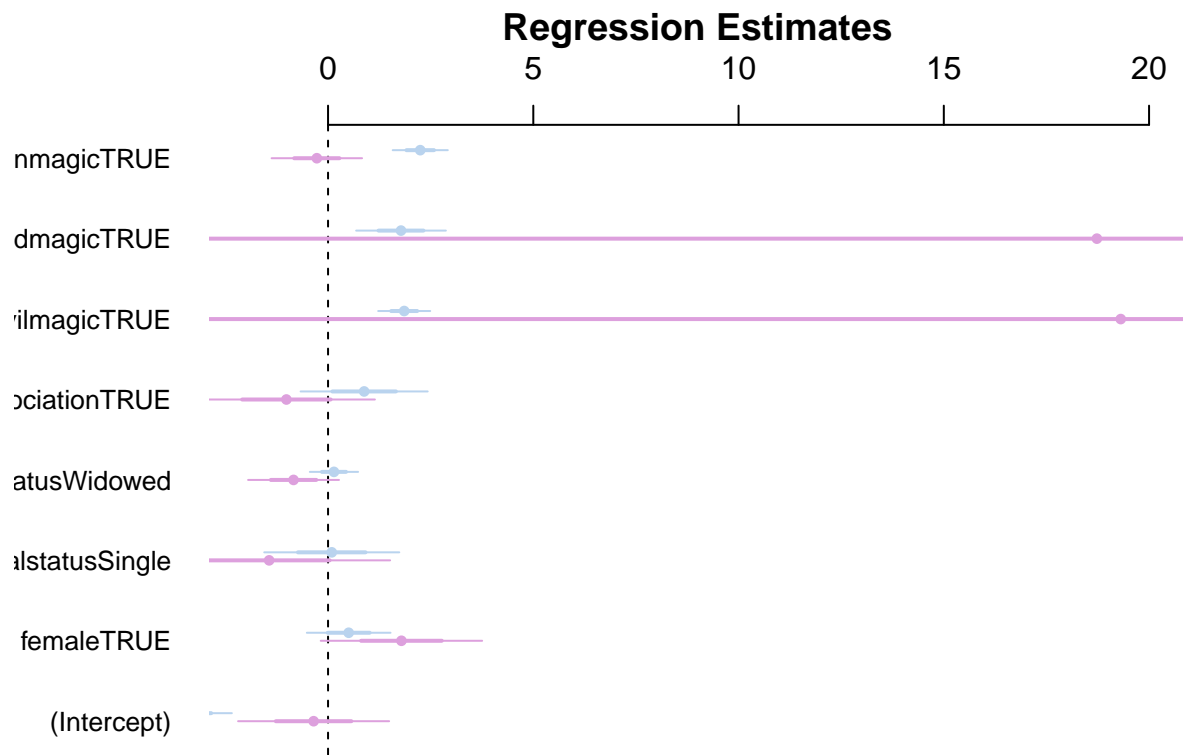
	<i>Dependent variable:</i>	
	Guilty	Executed
	(1)	(2)
Female	5.972** (2.673)	1.653 (1.665)
Single	0.238 (4.365)	1.094 (2.277)
Widowed	0.431 (1.740)	1.155 (1.342)
Association	0.362 (2.939)	2.408 (2.169)
Evil Magic	243,365,267.000 (Inf.000)	6.390*** (1.373)
Good Magic	136,051,281.000 (Inf.000)	5.911*** (1.728)
Non-Magic	0.760 (1.735)	9.447*** (1.399)
Constant	0.703 (2.510)	0.035 (1.653)
Observations	116	828
Log Likelihood	-48.507	-234.342
Akaike Inf. Crit.	113.013	484.685

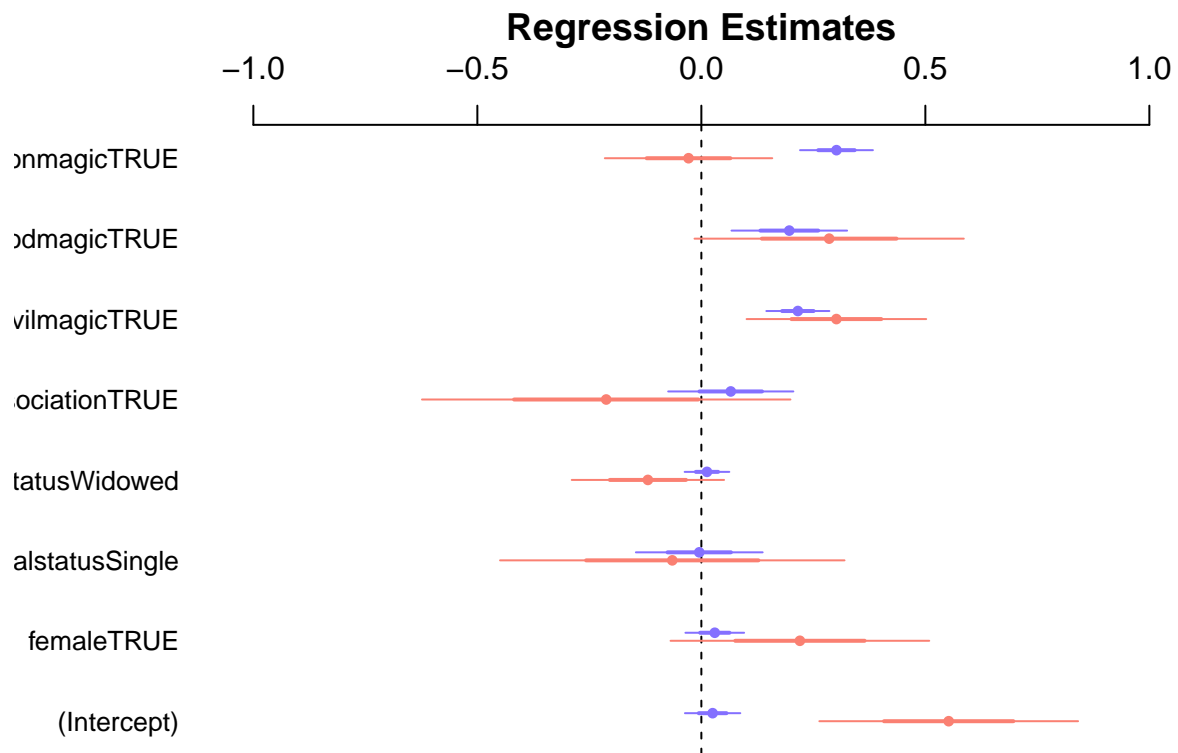
Note: *p<0.1; **p<0.05; ***p<0.01

Table 3:

	<i>Dependent variable:</i>	
	Guilty	Executed
	(1)	(2)
Female	0.220 (0.144)	0.030 (0.033)
Single	-0.065 (0.192)	-0.005 (0.071)
Widowed	-0.119 (0.085)	0.013 (0.025)
Association	-0.212 (0.206)	0.066 (0.070)
Evil Magic	0.301*** (0.100)	0.215*** (0.035)
Good Magic	0.285* (0.150)	0.196*** (0.065)
Non-Magic	-0.029 (0.093)	0.302*** (0.041)
Constant	0.552*** (0.144)	0.025 (0.031)
Observations	116	828
R ²	0.141	0.103
Adjusted R ²	0.085	0.096
Residual Std. Error	0.395 (df = 108)	0.284 (df = 820)
F Statistic	2.530** (df = 7; 108)	13.510*** (df = 7; 820)

Note: *p<0.1; **p<0.05; ***p<0.01





Discussion

Data Source

Julian Goodare, Lauren Martin, Joyce Miller and Louise Yeoman, *'The Survey of Scottish Witchcraft'*, www.arts.ed.ac.uk/witches/ (archived January 2003, accessed 13 December 2017).