

Final Assessment

POLS0010 Data Analysis Term II

2021

University College London
Department of Political Science

Guidelines for Completing and Submitting POLS0010 Term II Essay

- Read the guidelines below to avoid losing unnecessary marks.
- The assessment is due on Tuesday 20th April 2021 2.00pm. **It has two parts (I and II), both of which need to be submitted together. Part I of the essay is worth 60 marks. Part II is worth 40 marks.**
- Please follow all designated Department of Political Science submission guidelines. These may be different to those of your home department. You must submit one copy of your essay via Turnitin.
- **The word limit for both Parts I and II is 3,000 words**, excluding your R script appendix (see below). You can divide the word limit as you like between the two parts.
- The data-sets for the essay can be found in the folder ‘Datasets’ in the ‘Term 2 Assessment’ folder on Moodle, as well as on the course website under the section ‘Assessment’.
- This is an assessed piece of coursework for the POLS0010 module; collaboration and/or discussion with anyone is strictly prohibited. The rules for plagiarism apply and any cases of suspected plagiarism of published work or the work of classmates will be taken very seriously.
- Please read the guidelines in the ‘Essay Information’ folder in the ‘Term 2 Assessment’ section of Moodle if you require further guidance on logistics or plagiarism
- You may open up the data-sets and work on the essay questions anytime up until the submission date. There is no limit on the number of times you may open the data files. Be sure to save your data files and R script file.
- You should include a copy of your R script as an appendix to your essay. **Failure to include the R script will incur a 10 point penalty.** Note that your R script file should be neatly presented and easy to follow, including comments indicating the question being addressed.
- Include in your write-up all charts/plots and tables that you produce.
- Answers should be written in complete sentences; no bulleting or outlining.
- You may assume the methods you have used (e.g. logit regressions, etc) are understood by the reader and do not need definitions, but you do need to say which techniques you have used and why.
- As this is an assessed piece of work, **you may not email/ask the course tutors for help with the essay questions.**

Part I: Advanced Regression Techniques

This part of the final essay contains two questions. You must answer both of them. Question A is worth 25 points and Question B is worth 30 points.

Up to an additional five points will be awarded for clarity of presentation, especially tables and figures. See the week 5 Q&A slides for guidelines on presentation.

Both questions require you to write a brief report. It is up to you how you structure the reports, but it is advisable to keep introductory material to a minimum, given the word limit. Your reports should discuss your methods, your results and the conclusions that you draw from them.

Question A: Switzerland's Gun Control Referendum [25 points]

In Switzerland in 2011, a legally binding referendum was held that would have banned people from keeping guns at home, as well as introducing stricter background checks for those wishing to purchase them. The referendum failed, with 56% of voters opposing it. For this question, suppose that the referendum is going to be repeated next year, and the pro-gun-control campaign asks for your advice.

Specifically, the campaign group want to run an advertising campaign targeted at groups who are most likely to support the new referendum, to persuade them to turn out and vote. Your job is to tell them which types of people are most supportive of gun control. To help measure the likely effectiveness of their advertising, they also want to know how much each characteristic matters in explaining support. You'll use a survey of voters taken after the first referendum that asked about support for gun control. You need to:

- i. Choose a logit model that predicts support for gun control, carefully justifying your selection of variables for the model. You must use a minimum of three independent variables.
- ii. Present the model's findings in ways that clearly explain how much the variables matter in explaining support for gun control.
- iii. Provide information about how well your model's predictions perform.

You should present your approach and your findings in the form of a brief report. It should conclude by explaining which types of people you think the campaign should target. The dataset is called `s` and is contained in the file `swiss.Rda`. It contains the following variables for each individual in the survey:

Variable Name	Variable Description
<code>VoteYes</code>	Dependent variable: = 1 if respondent voted for gun control, 0 otherwise
<code>female</code>	=1 if female, 0 otherwise
<code>age</code>	In years
<code>LeftRight</code>	Individual's own assessment of how right-wing they are on a scale from 0-9, where higher values mean more right-wing [treat this as a continuous variable]
<code>trust</code>	Respondent's trust in government. =1 if trusts government, 0 otherwise
<code>university</code>	=1 if respondent has a university degree, 0 otherwise
<code>urban</code>	=1 if respondent lives in an urban area, 0 otherwise
<code>suburb</code>	=1 if respondent lives in a suburban area, 0 otherwise

Question B: A Second Brexit Referendum? [30 points]

In the 2016 UK referendum on leaving the EU, a majority of voters voted “Leave” and this started the multi-year Brexit process. At the end of 2018, after many rounds of negotiations with the European Union, then-Prime Minister Theresa May presented the draft agreement on a future relationship with Europe to parliament. At the same time, activist campaigns (most prominently People’s Choice) were trying to garner support for a second referendum.

After the Withdrawal Agreement failed to secure a majority of votes in two separate so-called “meaningful votes” in January and March 2019, the UK parliament decided to vote on a series of “indicative votes” between the 27th of March and 1st of April. These were not legally binding, but were supposed to help finding out whether there was a parliamentary majority for *any* other alternative. One of the motions that were voted on was for a “Confirmatory Public Vote”, which proposed a new referendum on the Brexit withdrawal agreement, and would include the option of voting for staying in the EU.

For this question, imagine we are just before these parliamentary votes will be held, and a pro-Second Referendum organisation has asked you to provide estimates for **constituency-level** support for such a second referendum. They want to use this information to lobby MPs to vote for the motion. The organisation has provided you with nationally representative survey data as well post-stratification data with demographic profiles in each constituency. Your tasks then are to:

- i. Estimate an appropriate logistic multilevel model predicting support for a second referendum by using the variables from the data-set `bes`. Briefly justify the inclusion of the chosen variables.¹
- ii. Present the multilevel results and discuss how the different variables affect preferences for a second referendum. (Note: you do **not** need to discuss statistical significance).
- iii. Produce post-stratified estimates of support for a second referendum for all 632 constituencies in England, Scotland and Wales based on the data-set `post`.
- iv. Compare these results to the constituency-level share of ‘remain’ voters² in 2016, which is included in the data-set `motionE`.

You should present and explain your approach and results in a brief report. The **survey data** you will use is Wave 15 of the British Election study (March 2019). It is called `bes` and is in the file `bes15.Rda`. Note that **all the group-level variables are already scaled**. It includes the following variables:

Variable Name	Variable Description
<code>ccode</code>	UK parliamentary constituency code
<code>cname</code>	UK parliamentary constituency name
<code>eurefdoOver</code>	Dependent variable: = 1 if respondent indicated they supported a second referendum, = 0 if they opposed it
<code>ageGroup</code>	8-point age scale (1 = 16-19, 2 = 20-24, 3 = 25-29, 4 = 30-44, 5 = 45-59, 6 = 60-64, 7 = 65-74, 8 = 75+) [treat this as a continuous variable]
<code>gender</code>	Binary factor variable indicating whether the respondent is male or female
<code>highed</code>	= 1 if respondent has a university degree, 0 otherwise
<code>lowed</code>	= 1 if the respondent has no qualifications or only a level 1 qualification, 0 otherwise
<code>housing</code>	Binary factor variable indicating if the respondents owns or rents the accommodation they live in
<code>socialGrade</code>	Factor variable indicating social grade (AB - upper and middle middle class, C1 - lower middle class, C2 - skilled working class, DE - working class and non-working)
<code>c_Con17_s</code>	Scaled vote share of tory party in 2017 General Election in the constituency
<code>c_leaveHanretty_s</code>	Scaled share of leave voters in the 2016 EU Referendum in the constituency
<code>c_whitebritish_s</code>	Scaled percent of constituency population who are white British
<code>c_deprived_s</code>	Scaled percent of constituency population living in poverty

¹When including constituency level variables, use the **scaled** version of the variable, which is included in the data-set and has the suffix “_s”. You do not need to do the scaling yourself.

²As the results of the vote were not released for individual electoral constituencies, these are themselves estimates of constituency-level ‘remain’ and ‘leave’ votes by Chris Hanretty, hence the name of the variable (see below).

The **post-stratification data** is called `post` and is in the file `psw_Hanretty.Rda`.³ Each row contains one particular demographic group in one constituency, for all 632 constituencies. The **scaled** group-level variables are also included. In addition to all of the variables in `bes`, it contains the variables:

Variable Name	Variable Description
<code>count</code>	Number of people in the demographic group
<code>population</code>	Number of people in the constituency
<code>weight</code>	Proportion of constituency represented by the demographic group

On the 1st of April, the last of these indicative votes were held, with no option achieving a majority. The motion to “hold a Confirmatory Public Vote” came closest with 280 votes in favour, 292 against and 62 abstentions. Information about each individual MP’s vote is to be found in the data-set `motionE`. Finally, therefore:

- v. Provide some figures comparing the estimates of constituency level support for a second referendum to the votes on the motion by their respective MP.

The data to use for the comparisons is called `motionE`, which is in the file `motionE.Rda`. This data has 632 rows (one for each constituency), and includes the votes by MP on the motion to hold a Confirmatory Public Vote on the Withdrawal Act (essentially a second referendum). Note that all Northern Irish MPs are omitted from the data and the speaker of the house and the deputy speakers are marked as voting `NA`, since they abstain from parliamentary votes. The data includes:

Variable Name	Variable Description
<code>ccode</code>	UK parliamentary constituency code
<code>mp_name</code>	The MP’s name
<code>mp_party</code>	The MP’s party affiliation
<code>voteOn2Ref</code>	The MP’s vote on motion E to hold a confirmatory public vote (three possibilities: "Aye" (=Yes), "No" and "Abstained")
<code>c_remainHanretty</code>	The remain vote share in the constituency

³This is the post-stratification data used by Chris Hanretty (2020). “An Introduction to Multilevel Regression and Post-Stratification for Estimating Constituency Opinion.” *Political Studies Review* 18(4): 630 –645.

Part II: Introduction to Data Mining

Question C

[Will be released on Wednesday 24th March]