Assignment 1

GV903 Advanced Methods – 2020/2021
University of Essex, Department of Government

Instructions

This assignment is due on Wednesday, 18 November at 9:45 am. Please submit on FASER. Instructions can be found in the module handbook on Moodle.

Please create a new document with your answers. Use of LaTeX in conjunction with `knitr` via RStudio is strongly encouraged for creating the answer document. Up to five bonus points can be given for proper use of this technology (and still some if you demonstrate competent use of Markdown or other relevant technology) and proper formatting. Please submit a PDF document on FASER along with the source document before compilation if applicable.

The number of points you can obtain for each question or task is given in square brackets. The points add up to 100 and determine your final mark for this assignment. The bonus points come on top of these points, but the final mark is capped at 100. The final mark you earn for this assignment is given by the equation

$$m = \min \left\{ 100, b + \sum_{i=1}^{n} p_i \right\}, \tag{1}$$

where p_i denotes the number of points you earn for question or task i, n is the number of tasks in this assignment, and b is the number of bonus points you earn for use of technology and formatting. All answers are evaluated on three criteria: how clearly the results are presented; how correct the results are; and how elegant, computationally efficient, or sophisticated the solution is (where applicable). Good luck!

1 The Legislative Process in a Federal System

Imagine a country with a federal system, which has existed for 500 years. In this country, new legislation has to pass two consecutive votes with at least four weeks between them in the first chamber of the parliament. Additionally, if the new law affects the competences of the 12 sub-national districts (or a subset of them; we call it a "district-level" legislative proposal in this case), the law also has to pass a third vote in the second chamber of the parliament, which is composed of sub-national representatives, in order to become effective. Two thirds of all laws affect the competences of the districts. The probability of any law passing the first vote in the first chamber is 0.62. The probability of a positive vote in the second

round is relatively high at 0.93 because laws that make it to this stage have already demonstrated their acceptability in the first round. The odds of a successful (= positive) vote in the second chamber are 1.20. Each year, an average of 183 laws enter the parliament. The parliament (i. e., both chambers) is elected for five-year terms in simultaneous district and national elections.

Answer the following questions. Provide a short text each time that explains your reasoning and explains which laws or rules are applied. Show the equations, insert the numbers, and add one sentence that summarizes the result.

1. Over a parliamentary term, how many legislative proposals can be expected to be processed by the parliament that affect only the national level? [3 points.]

2. What is the standard deviation of the average number of legislative proposals in a given year? [3 points.]

3. What is the probability of any national-level legislative proposal to be successful? [3 points.]

4. What is the probability that a proposal is successful and a national-level proposal? [3 points.]

5. What is the probability of any district-level legislative proposal to be successful? [3 points.]

6. What is the probability that a proposal is successful and a district-level proposal? [3 points.]

7. Given that a district-level legislative proposal was successful in the first chamber in both rounds, what is the probability that it will fail in the second chamber? [3 points.]

8. What is the probability that any legislative proposal (without knowing a priori if it is national- or district-level) is successful? [3 points.]

9. Provide an estimate and confidence interval (at a significance level of 95 per cent) of how many adopted national-level laws you expect per term. [3 points.]

10. What is the probability to see exactly 50 legislative proposals enter the parliament in a given year? [3 points.]

11. What is the probability that we will see fewer than 160 legislative proposals in a given year? [3 points.]

12. What is the probability that more than 115 district-level laws will be adopted in a given legislative term? [3 points.]

2 Implementation of the binomial distribution

1. Implement an `R` function that computes the probability of observing exactly y successes in a series of n Bernoulli experiments with success probability p. Do not use the `*binom` family of functions. (You can use them to verify your results, though.) Show an example run with 50 trials, a probability of one third and 20 successes to demonstrate that it works. Show the `R` code of your function. Hint: You may want to use the `choose` function in your code for $\binom{n}{k}$. [9 points.]

2. Implement the cumulative distribution function of the binomial distribution in `R`. The function should return the probability of up to y successes with a given number of trials and a given probability per trial. Do not use the `*binom` family of functions. (You can use them to verify your results, though.) Show an example run with 50 trials, a probability of one third and up to 20 successes to demonstrate that it works. Show the `R` code of your function. [9 points.]

3. Implement and show a function that plots the density function of the binomial distribution for a value range defined by the user as well as a success probability provided by the user. Plot the probability density for $p = 0.27$ and a range of $[0; 10]$. [9 points.]

3 Modelling corruption

Load the CSV file `assignment1.csv` into `R`. It contains a number of variables with mostly cryptic column names for all countries in the world in 2006 (and some other territories). The variable `cpi_score2006_PL` contains the *Corruptions Perception Index* by Transparency International for each country or territory on a nearly continuous scale from 1 (high corruption) to 10 (little corruption). Transparency International describes the index as follows:

> *"The CPI scores and ranks countries/territories based on how corrupt a country's public sector is perceived to be by experts and business executives. It is a composite index, a combination of 13 surveys and assessments of corruption, collected by a variety of reputable institutions. The CPI is the most widely used indicator of corruption worldwide."*

In this task, you will model the Corruptions Perception Index graphically and statistically. To do so, focus on the following additional variables: Vanhanen's index of political participation (`vh_part2000_PL`); Vanhanen's index of electoral competition (`vh_comp2000_PL`); OECD membership (`oecd_PL`); population size (`pt_pop_PL`); population density as reported in the Wold Development Indicators by the World Bank (`POPDENS_9605_WDI`); and GDP per capita (`un_gdppc_PL`). Make any transformations as needed for the analysis. Complete the following tasks:

1. Visualise the corruption index in different ways in `R` to convey its distribution and its relationship with the other variables. Limit your presentation to a few plots that convey the main messages or insights. Use the variables mentioned

above for this purpose (and possibly the names of the countries if helpful). Show the `R` code and the resulting plots. [7 points.]

2. Explain corruption scores as a function of the other variables mentioned above using a linear model. Use the `texreg` package to show the regression results with a good layout. [6 points.]

3. Show the linear equation for the empirical model here, including the β parameters and variable names. [4 points.]

4. Discuss if the Gauss-Markov assumptions hold in this application. You can discuss each assumption theoretically in relation to the case or use `R` to support your answer as you see fit. [8 points.]

5. Interpret the summary of the model output, including an interpretation of the effect size and significance of Vanhanen's political participation and electoral competition indices and the goodness of fit of the model. [6 points.]

6. Recompute R^2 for the model in `R` using your knowledge of the mechanics of the linear model. Also show the relevant equations for each step, and briefly explain them. [6 points.]