POL 2504 PS 3

29 September 2020

Section 1

In a fictional academic job market, there is a total of 150 doctorate degree holders in political science who are seeking jobs. The following table shows the distribution of job applicants from different universities. Recently, a crisis hit this academic job market. The Khaalesi, the Mother of Dragons, burnt down multiple cities with her dragons, which resulted in a global economic crisis. Therefore, there are only 5 positions available to be filled this year only in one university, University of Dreamland. Suppose that everyone applies for these positions and no candidate rejects any job offer. Also assume that because the available call for job applications aim for a large pool of candidates, there is no constraint in terms of sub-fields, etc.

University	# of Graduates
University of Meereen	23
Braavos University	14
University of Yunkai	15
Astapor College	13
King's Landing College	22
Graduate School of Volantis	14
University of Casterly Rock	13
Institute of Dorne	17
University of Iron Islands	19

Question 1

Suppose that all candidates are well-published and well-equipped, therefore the probability of getting an offer is equal for each candidate. Assume that there is no spousal hire, nepotism, identity-based discrimination, or any special considerations. Offers are made all at once. What is the probability that at least one graduate of Braavos University would be employed?

Question 2

Repeat the question above using a Monte Carlo simulation. For reproducibility, use `set.seed(123)`.

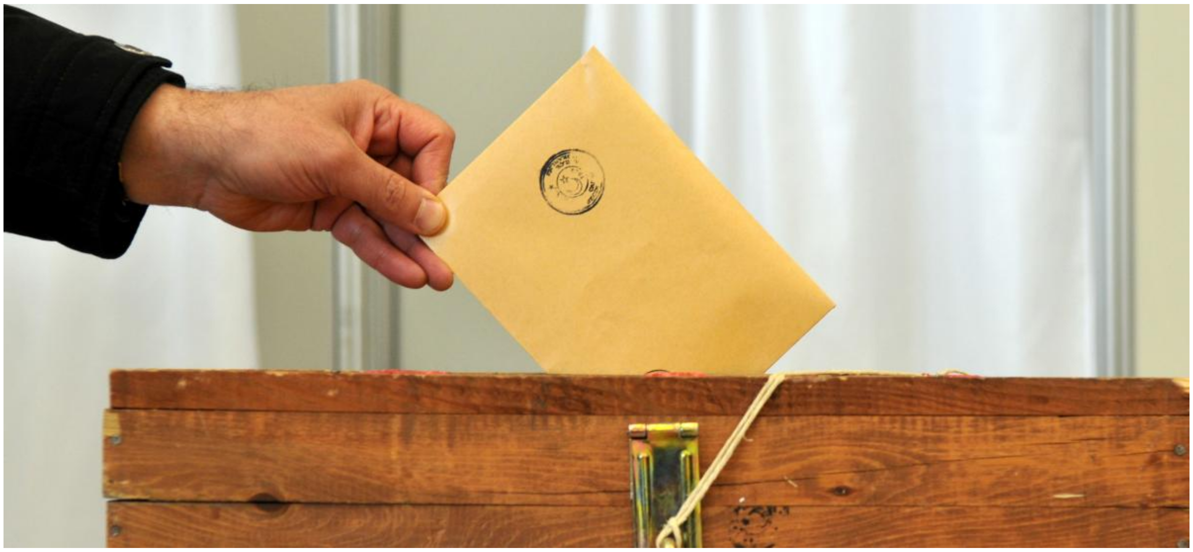
Question 3

What is the probability that only the graduates of Institute of Dorn would be employed? Would you suspect nepotism or special treatment?

Question 4

Now suppose that the hiring committee holds five consecutive meetings and at each round, they decide on who to make an offer. You know that Braavos University offers a special sequence of quantitative methods classes. Therefore, the probability of getting an offer for each graduate of Braavos University is twice the probability for who did not receive such training at each round. Calculate the probability that at least one graduate of Braavos University would be employed.

Section 2



2018 was a turbulent year for Turkey. As per the electoral law, the standard procedure was to pre-stamp all ballot papers before the elections, and only stamped ballots were counted valid. The use of pre-stamped ballots was one of the procedures in place to prevent electoral fraud. However, just before the elections in 2018, the Supreme Election Board announced that counting unstamped ballots would be permitted. The incumbent electoral alliance of the Justice and Development Party (AKP) and the Nationalist Action Party (MHP) won 53.7% of the votes in the parliamentary election but faced many accusations of electoral fraud.

In this exercise, we use the rules of probability to detect electoral fraud by examining voting patterns in the 2018 Turkish election. This exercise is based on two readings:

- Klimek, Peter, et al. (2018). "Forensic analysis of Turkish elections in 2017–2018." *Plos One*, October.
- Rozenas, Arturas. (2017). "Detecting Election Fraud from Irregularities in Vote-Share Distributions." *Political Analysis* 25(1): 41-56.

We will use the official election results for the presidential election, contained in `fraud2018.xlsx` to investigate whether there is any evidence for electoral fraud. The data contain information on the total number of votes, the number of turnout, and the number of votes for President Erdogan for each polling station (in total, 180122) across 907 administrative units. The variables are described below.

Name	**Description**
`id`	The unique identifier for the administrative unit
`n_voters`	The total number of voters registered in the polling station
`n_turnout`	The total number of people who turned out
`n_winners`	The total number of people who voted for Erdogan

Question 1

To analyze the 2018 Turkish election results, first compute Erdogan's vote share as a proportion of the voters who turned out. Identify the 10 most frequently occurring fractions for the vote share. Then create a histogram that sets the number of bins to the number of unique fractions, with one bar created for each uniquely observed fraction, to differentiate between similar fractions like $\frac{1}{2}$ and $\frac{51}{100}$. This can be done by using the `breaks` argument in the `hist` function. The intuition behind this analysis is that if the frequency of coarse vote-shares is high (such as 0.5, 0.6, 0.75), then we might suspect that the results are fabricated.

Question 2

The mere existence of high frequencies at fractions with small numerators and denominators does not always imply electoral fraud, though. Under certain conditions, these high frequencies may stem from simple numeric laws. Indeed, more numbers are divisible by smaller integers like 2, 3, and 4 than by larger integers like 22, 23, and 24. To investigate the possibility that the low fractions arose by chance, assume the following probability model:

- Turnout for a polling station is binomially distributed, with size equal to the number of voters registered in the polling station and success probability equal to its observed turnout rate.
- Vote counts for the incumbent in a polling station is binomially distributed with size equal to the number of voters who simulated to turn out in the previous step and success probability equal to the polling station's observed vote share.

Conduct a Monte Carlo simulation under these assumptions. 500 simulated elections should be sufficient. (Note that this may be computationally intensive code.) What are the 10 most frequent vote share values? Create a histogram similar to the one in the previous question. Briefly comment on the results.

Question 3

To judge the Monte Carlo simulation results against the actual results of the 2018 Turkish election, we compare the observed fraction of observations within a bin of certain size with its simulated counterpart. To do this, create histograms showing the distribution of the following fractions $\frac{1}{2}$, $\frac{2}{3}$, $\frac{3}{4}$, and $\frac{3}{5}$. Then compare them with the corresponding fractions' proportion in the actual election. Briefly interpret the results.

Question 4

We now compare the relative frequency of observed fractions with the simulated ones beyond the four fractions examined in the previous question. To do this, we choose a bin size of 0.01 and compute the proportion of observations that fall into each bin. We then examine whether or not the observed proportion falls within the 2.5 and 97.5 percentiles of the corresponding simulated proportions. Plot the result with vote share bin on the horizontal axis and estimated vote share on the vertical axis. Now count the number of times an observed polling station vote share falls outside its simulated interval. Interpret the results.