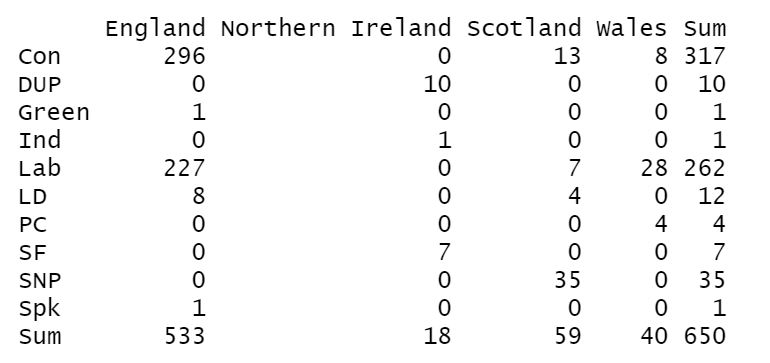
**Problem Set 2**

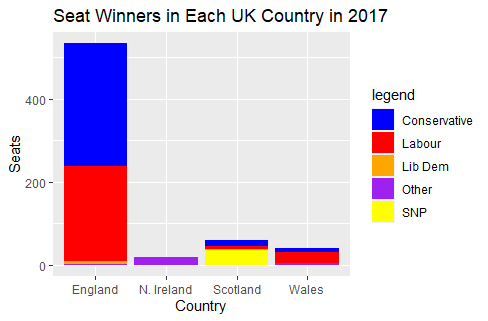
This problem set covers material from the first three classes including material from OIS Chapter 2 Section 2.2 to Chapter 5 Section 5.1. It makes use similar data to that we discussed in class. Partial credit may be given for answers that are correct in part, but not in full.

**Part I: The 2017 UK Election (16 points)**

Unlike the 2019 UK Election, the 2017 UK Election featured a close finish between the Conservatives (aka the Tories) and Labour. This section makes use of data from that election. (Note: Con=Conservatives, Green=Green, Lab=Labour, LD=Liberal Democrats, Spk=Speaker).



1. Below is a contingency table showing the number of seats won by each party by country. (8 points)
   1. Please calculate the row proportions for the Conservatives for each country. You may use a calculator to divide large numbers, but please show how you get the numbers (4 pts.)
   2. Please calculate the proportions for England for each party. The above rules about calculators apply; you can skip the calculation for the Speaker. (4 pts.)
2. Below is a stacked barplot I made in R that shows the results by country. I would like you to think about what the reader takes away from this visualization. (8 pts.)

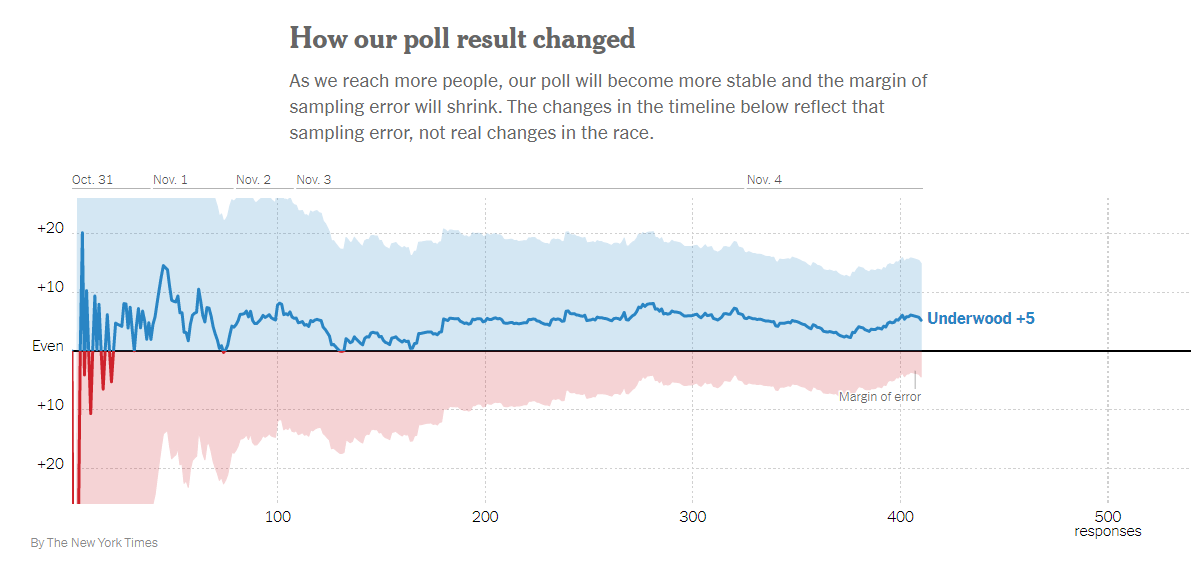


* 1. What is a benefit of using a stacked bar plot when considering the number of cases by country? What is a disadvantage when considering the relative number of seats won by a party within each country? (2 pts.)
  2. Let’s say you were most interested in having the reader know the relative number of seats won by party. What is an alternative type of bar plot you could make? Why would this type of bar plot do a better job of presenting this information to the reader? (2 pts.)
  3. Would a pie chart be a good way to display this data? Why or why not? (2 pts.)
  4. Imagine you were interested in having the reader know the percentage of cases won by a party within a country. What challenges might they face looking at the stacked bar plot? Are there any countries where this might particularly be a challenge? (2 pts.)

**Part II: Probability and the 2018 Illinois 14th Congressional District Election (14 pts.)**

One of the final polls conducted by the Siena College for the Upshot was in Illinois 14th Congressional District. This [poll](https://www.nytimes.com/interactive/2018/upshot/elections-poll-il14-3.html) had Democrat Lauren Underwood ahead of incumbent Randy Hultgren by a 49-43% margin. (Note: the lead in the poll was actually 5 points due to rounding.) On [Election Day](https://ballotpedia.org/Illinois%27_14th_Congressional_District_election,_2018), Underwood defeated Hultgren by 5 percentage points, 52.5% to 47.5%. This section’s questions focus on that poll and the election results in this district.

1. This chart shows how the result in the poll changed as more people were polled. This question focuses on this chart. (4 pts.)



* 1. Why does Underwood’s lead in the poll stabilize close to her actual margin on Election Day as the number of poll respondents goes up? What principle does this illustrate? (2 pts.)
  2. Let’s say we were to add an additional 500 respondents to this poll. Would we expect the margin to stabilize or begin to oscillate more? Why? (2 pts.)

1. Now please consider the actual results on Election Day (Underwood 52.5%, Hultgren 47.5%) when calculating these probabilities. (10 pts.)
   1. What is the probability that a randomly selected voter is either an Underwood voter or a Hultgren voter? Please show your work. (2 pts.)
   2. What is the probability that a randomly selected voter is an Underwood voter and a Hultgren voter? Why is this the case? (2 pts.)
   3. Let’s say that you randomly selected 2 voters at random. What it is probability that they both would be Underwood supporters? What is the probability that they both would be Hultgren supporters? Please show your work. (4 pts.)
   4. If you randomly selected 5 voters and all of them were Underwood voters, are you “due” for a Hultgren supporter on your 6th draw? Thinking just about that draw, what is the probability of selecting a Hultgren supporter? (2 pts.)

**Part III: Probability, the Partisan Affiliation of Governors, and the Winning Party in 2016 Presidential Election (39 pts.)**

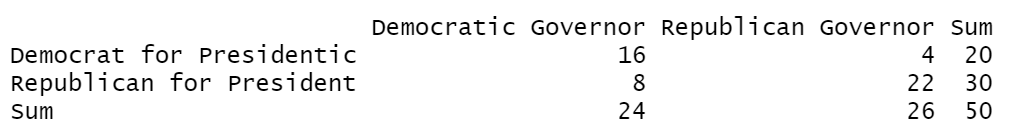
This section makes use of data on the current partisan affiliation of US Governors and the results of the 2016 presidential election at the state level. In 2016, Republican nominee Donald Trump won 30 states and Democratic nominee Hillary Clinton won 20 states (for the purposes of this exercise, we are counting Maine as a Clinton win even though Donald Trump won one of their electoral votes that they allocate by congressional district winner.) Currently 26 states have Republican governors and 24 states have a Democratic governor.

1. Here, I would like you to consider the probability of various outcomes when it comes to the partisan affiliation of Governors and a state’s vote in the 2016 presidential election. (12 pts.)
   1. The below probability distribution shows the four possible outcomes. Please fill in the two outcomes that are blank. Please show all of your work. (6 pts.)

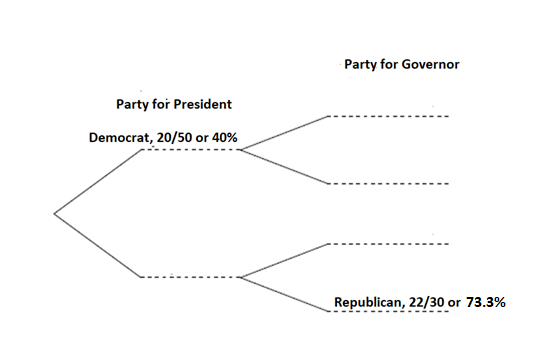
|  |  |  |  |
| --- | --- | --- | --- |
| DPres, DGov | DPres, RGov | RPres, RGov | RPres, D Gov |
|  | 8% | 44% |  |

* 1. Here I gave you the number of states that both presidential candidates won. Imagine that you only knew the total number of states and the number of states that Clinton won. How could you find the percentage of states that Trump won? What does Clintonc represent here? (3 pts.)
  2. Looking at the above table, if you knew the party of the presidential candidate that won a state in 2016, would that help you guess the party of the state’s governor? Please explain your reasoning. (3 pts.)

1. It appears that the probability of a state voting for a party for president and voting for the same party for governor may be related. Based on this seeming relationship, please answer the following questions. I have also provided a contingency table that summarizes the distribution of this data. (21 pts.)



* 1. If you randomly select a state, what is the marginal probability of a state having a Democratic governor? Please show your work. (2 pts.)
  2. If you randomly select a state, what is the joint probability of the state having a Democratic governor and voting Democratic for President? Please show your work. (2 pts.)
  3. What is the conditional probability of having a Democratic governor given that a state voted Democratic for President? Please show your work. (2 pts.)
  4. Does having a Democratic governor and voting Democratic for President appear to be independent? (Hint: use your answer from question 5c). Based on this, use the general multiplication rule to find the probability of having a Democratic governor and voting Democratic for President. Please show your work. (3 pts.)
  5. Probability trees are a good way to organize outcomes. Please fill in this probability tree. (4 pts.)



* 1. Using the probabilities you filled in above, please calculate the conditional probabilities that apply to secondary branch. (Note: you can either fill them in above or put them here, but if you place them here, please note which secondary branch they apply to below.) Please show your work. (8 pts.)

1. In class, we also discussed whether a state is part of the “Blue Wall.” 18 states are part of the “Blue Wall.” (6 pts.)
   1. The probability of having a Democratic governor given that a state is in the Blue Wall is 83.33% (15 states). What is the probability of a state being in the Blue Wall given that it has a Democratic governor? (3 pts.)
   2. The probability of having a Democratic governor given that a state is *not* in the Blue Wall is 28.125% (9 states). What is the probability of a state being in the Blue wall given that it does *not* have a Democratic governor? (3 pts.)

**Part IV: Some Normal (or Non-Normal) Questions about Distributions (12 pts.)**

This section makes use of data from the 2012 Election at the state level that the [CQ Voting and Elections Collection](http://library.cqpress.com.proxy.lib.duke.edu/elections/download-data.php) collects and is available through. Duke libraries.

1. Below is a histogram of the percentage of the vote received by President Obama at the state level in 2012 (bin width= 2 percentage point) and a Q-Q plot of the distribution of the data.

|  |  |
| --- | --- |
|  |  |

1. Does this data look like it follows a normal distribution? Discussing both plots, what evidence is there that the data follows a normal distribution and what evidence is there that it is not distributed normally? (4 pts.)
2. Among the 50 states, the mean Obama two-party percentage is 49.15 and the standard deviation is 10.47. President Obama received 52.96% of the two-party percentage in Iowa. What is Iowa’s z-score? What is its percentile for Democratic two-party vote? Please show your work or explain how you got the value. (4 pts.)
3. In contrast, President Obama received 36.33% of the two-party vote in West Virginia. What is West Virginia’s z-score and percentile? Please show your work or explain how you got the value (in the case of the percentile). (4 pts.)

**Part V: Liddy Dole is 92 (or 93) (6 pts.)**

1. In 2008, the Democratic Senatorial Campaign Committee ran a famed [television ad](https://www.youtube.com/watch?v=LLkazmjpcIs) against Senator— and Duke alum— Elizabeth Dole (R-NC) accusing her of being either 92 (in terms of the percent of time she voted with George W. Bush) or 93 (in terms of her ran in effectiveness.
   1. Imagine that ten pieces of legislation came up for a vote. Taking the percentage of time the ad claimed she voted with George W. Bush, what is the probability she would vote for all ten? Please show your work. (3 pts.).
   2. Again imagine that ten pieces of legislation came up for a vote and take the percentage of time the ad claimed Senator Dole voted with George W. Bush. What is the probability that she would vote for two pieces of legislation that came up for a vote? (3 pts.)

**Part VI: A Lift in the Polls? (13 pts.)**

Imagine that a researcher is very interested in public opinion about Cherie Berry and he decided to conduct a series of polls in North Carolina on Election Day in 2016. The actual percentage of vote Cherie Berry received in North Carolina in 2016 was 55.25 percent (i.e, 0.5525) and the standard deviation was 0.497.

1. After conducting a series of polls, the researcher looks at the sampling distribution. These questions pertain to that sampling distribution.
   1. Imagine the researcher conducted a series of polls with a sample size of 40. Does this meet the success-failure threshold for the central limit theorem to apply? What would you expect the distribution to look like? What would the mean and standard error be for the sampling distribution? (7 pts.)
   2. Imagine the researcher now conducted a series of polls with a sample size of 1,000. What would you expect the distribution to look like? How would it compare to the distribution for part a? What would the mean and standard error be for the sampling distribution? (6 pts.)