

The University of Chicago
Booth School of Business
36106 Managerial Decision Modeling
Kipp Martin

1 Simulating Elections Using @RISK

Simulation can play an important role in forecasting election outcomes. This discussion is based on work by Larry Robinson, a former Booth Ph.D. student. He is now a professor in the Johnson Graduate School of Management at Cornell University. See Larry's articles [Article 1](#) and [Article 2](#). These two articles describe the use of simulation to forecast the result of the November, 2014 United States Senate race. These articles lean heavily on the work of well-known prognosticator Nate Silver of [FiveThirtyEight](#).

The November, 2014 race generated considerable interest about whether or not the Democratic party could hold the majority in United States Senate. Going into the November election there were 36 senate seats up for grabs. Of the remaining 64 seats not subject to election, 34 were held by the Democratic party and 30 by the Republican party. In order to reach the majority of 51 seats the Republican party needed to win 21 of the 36 seats up for grabs. Stated another way, in order to maintain their majority the Democratic party needed to win 16 of the 36 seats up for grabs. The Democratic party maintains the "majority" in case of a 50-50 tie because Joe Biden, Vice President of United States, and pro tempore of the Senate will obviously break ties in favor of the Democratic party.

Here are the relevant [Excel files](#). See spreadsheet **Model** in the workbook **2014USSenate_538Independent.xlsx** for a list of the 36 seats that are up for grabs. (Color Coding: blue is used for Democratic states and senators, red for Republican states and senators)

There are three approaches to this forecasting problem.

1. Take a poll in each state and then simply count the states where each party was leading. The forecast is then based on this count. There is no simulation with this approach, and until recently, this was the most common approach.
2. Use a probability distribution to model the outcome in each state. Run a simulation and then build a distribution based on the simulation results. This

is what was done in the workbook **2014USSenate_538Independent.xlsx**. With this approach there is no correlation between states.

3. Use a probability distribution to model the outcome in each state. In addition, build a correlation matrix to account for correlated election results between the states. More on this later.

2 Some Details – Zero Correlation Case

First consider Method 2 described above. This methodology is implemented in the workbook **2014USSenate_538Independent.xlsx**. Crucial to this approach is selecting a probability distribution to model the outcome in each state. In this study the Bernoulli distribution was selected. The Bernoulli random variable is listed under Discrete Distribution in @RISK. This is a very simple random variable with two outcomes, either 1 or 0. The only parameter necessary to define the distribution is p the probability of winning or success. In the spreadsheet **Model** the range H7:H42 contains the Bernoulli random variables. Larry essentially used probabilities from the Nate Silver Website to determine the Bernoulli parameter. These probabilities are in range E7:E42. Note that this model is set up from the standpoint of the Democratic party, the probability of win or success is for the Democratic candidate.

The simulation output is in cell E46 and is a count of the number of seats held by the Democratic party. It is a sum of the Bernoulli trials in H7:H42 plus the 34 incumbent seats. The simulation results are in the **Summary** spreadsheet. According to the simulation, the probability of the Democratic party holding 50 seats or more is 34.8%.

3 Some Details – Nonzero Correlation Case

Now, what about correlation? Does it really make sense for there to be zero correlation *between all 50 states*? Here is what Larry did in his study. The correlated simulation is in the workbook **2014USSenate_538Simulation**. The model is in the spreadsheet **Model**. The correlation matrix is in the range N7:AV41 (for some reason Alabama is not in the range). Examine the entry in each cell in the matrix

= IF(ROW() = COLUMN()-7, 1, \$K\$5)

Note the use of the Excel Row() and Column() functions. The formula returns a value of 1 if the column and row indexes are identical; that is, a state has a 100%

correlation with itself. However, if the row and column indexes are not equal, the correlation used is the value in cell K5.

Sensitivity analysis is performed on the value in cell K5. Cell K5 contains the formula

```
= RiskSimtable(K7:K33)
```

and 27 simulations are done over the range 0% to 100% in the range K7:K33.

The correlation matrix is used to modify the outcomes in column H. They are now

```
= RiskBernoulli(E10, RiskCorrmat($N$7:$AV$41, ROW() - 6))
```

The correlation substantially affects the forecasted outcome. See the spreadsheet **Retain Graph**.

Discussion Questions:

1. What are your thoughts about how the correlation matrix was constructed?
2. How do you think the correlation matrix should be constructed?