

Kaplan Chapter 2

Example 1: The Campaign Advisor States Confidently....

The political advisor to a fringe group's candidate for President has news just 10 days before the election. "I'm afraid the latest poll looks grim, sir. Only 8% of voters support us. We pledged to withdraw from the election if we appealed to fewer than 10% of voters-- perhaps now is the time to get out of the race."

Interrupting, the legal advisor opines, "But if we get more than 10% of the vote, we are guaranteed federal campaign money in the next election. I think we should stay in."

The candidate, **Ross Anderson**, summarizes, "So if we get out now we won't be criticized for splitting the vote and we remain viable for the next election. But if we stay in and get 10% of the vote, we'll get campaign money for next time. [Expletive deleted]. How confident are you in that 8% figure?"

"Quite confident, sir," the advisor responds. "The pollsters talked to 500 randomly selected registered voters. Only 40 of them supported us. That makes ...," he taps on his calculator, "exactly 8%."

The above fictional exchange is unrealistic: modern political advisors know enough about statistics to realize that the 8% figure isn't completely precise. It is based on a random sample. If a different group of 500 had been picked the result would have been somewhat different. More to the point, there is some probability that a random sample of 500 would show only 8% support even if the true support level were greater than 10%. The candidate's question about confidence in the results is really a question about whether there is any possibility that the true support level is more than 10%.

Here's one way to answer that question: collect more data. But the candidate needs to make a decision about withdrawing **now** and it takes time (and money!) to collect more data. The candidate needs to infer from the data at hand how precise is the 8% figure. So, rather than actually running several more polls, we use resampling to simulate the polls. Then we look at how much sampling variability there is in the simulated polls.

The following commands are in the script file **Kaplan_2_1.m**.

```
%Kaplan_2_1.m
clear
seedRNG0;    %zero seed the random number generator algorithm
%seedRNG;

pollsize = 500;
%voter population poll results: 92% against, 8% support candidate
voters = [zeros(92,1); ones(8,1)];
nIterations = 10000;
resamples = zeros(nIterations, 1);
%run the resampling simulation
for i=1:nIterations
    sample = randsample(voters, pollsize, true);
    support = sum(sample)/pollsize;
    resamples(i) = support;
end
```

```

%find values requested in original Kaplan exercise
%compute the most likely poll result
percent4 = mean(resamples);
%calculate the boundaries of the 95% confidence interval
[bounds] = prctile(resamples, [2.5 97.5]);

%now, let's go beyond Kaplan's simple 95% confidence interval computation:
%generate cumulative probability function for possible poll results
[cumprobs, pollvals] = ecdf(resamples);
%now, find the probability of obtaining 10% of votes or higher
temp = abs(pollvals - 0.10); %subtract 10%
index = find(temp == min(temp)); %find location of minimum difference from
10%
p10plus = 1.0 - cumprobs(index); %compute tail probability for 10% plus

%plot cumulative probability function for resampling simulation
figure, clf, plot(pollvals, cumprobs);
xlabel('Poll Results (Candidate's Percentage)');
ylabel('Cumulative Probability');
title('Resampled Political Poll Projections');
line([pollvals(index) pollvals(index)], [0 1]); %draw line delineating upper
tail
annostr = sprintf('Prob(10% plus election result) = %6.3f', p10plus);
text(0.045, 0.9, annostr); %annotate figure with critical result

```

This cumulative probability of the bootstrap distribution suggests there is a possibility that the candidate has more than the magic 10% level needed to qualify for federal funding – in this case the estimated probability of such an outcome is 4.4%.

“Those poll data are worthless!” the candidate shouts. “Get me some solid numbers now!” he demands. (We shall revisit this issue in a subsequent exercise.)

Notes on origin of this exercise:

Text is from Chapter 2 of online text “Resampling Stats in MATLAB” by Daniel T. Kaplan

URL: <http://www.resample.com/support/user-guides/resampling-stats-for-matlab-users-guide/>

The MATLAB script has been completely rewritten – using a different approach -- so that it does not require the Resampling Stats library (FS 6 January 2014)