Sampling Variability

1 Quick recap of chapter two:

- In Chapter Two, we explored two case studies of American presidential elections, examining the reasons behind the pollsters' inaccurate predictions in the 1936 and 1948 presidential elections.
- We will continue with this theme and study about pre-election polls and exit polls in this chapter.

2 Pre-election polls and exit polls

- Pre-election polls is an attempt to predict the election before it happens. The media love these polls but statisticians don't love them because elections often don't go as forecasted, even when the polls use all the right statistical methods. The sample chosen can be voluntary response sample or convenient sample and biases such as response bias and non-response bias can arise in such samples.
- Exit polls, which interview voters as they leave the voting place, don't have these problems. The people in the sample have just voted. A good exit poll, based on a national sample of election precincts, can often call a presidential election correctly long before the polls close. But they aren't also infallible as historical evidence suggests.
- Two examples: Democrat Tom Bradley, the black mayor of Los Angeles, lost the 1982 California gubernatorial election to Republican George Deukmejian despite having been ahead in the polls, supposedly because voters were reluctant to tell interviewers that they were not going to vote for a black candidate.

  In 2016 elections, many of those who were polled simply were not honest about whom they intended to vote for. The idea of so-called "shy Trumpers" suggests that support for Trump was socially undesirable, and that his supporters were unwilling to admit their support to pollsters.

  The two cases above are the examples of response bias.
- Arguments against public pre-election polls charge that they influence voter behavior. Voters may decide to stay home if the polls predict a landslide–why bother to vote if the result is a foregone conclusion? Exit polls are particularly worrisome because, in effect, they report actual election results before the election is complete. The U.S. television networks agree not to release the results of their exit surveys in any state until the polls close in that state.

- For example: In 2004, early exit polls had Bush losing Ohio and North Carolina and dead even in Florida, Arizona, Colorado, South Carolina and Mississippi – all states where he prevailed. Again, the polls affected the coverage, with the media reluctant to award Ohio to Bush, though he won it by nearly $120,000$ votes.
- Pollsters conduct exit polls in different parts of the country to account for sampling variability before publishing their results.

3 Sampling variability

- A parameter is a number that describes the population. A parameter is a fixed number, but in practice we don't know the actual value of this number because we cannot access the entire population.
- A statistic is a number that describes a sample. The value of a statistic can be determined and is known once we have taken a sample, but its value can change from sample to sample. We often use a statistic to estimate an unknown parameter.
- Variability describes how the values of the sample statistic will vary when we take many samples. Large variability means that the result of sampling is not repeatable.
- A good sampling method has both small bias and small variability.
- To reduce the variability of an SRS, use a larger sample. You can make the variability as small as you want by taking a large enough sample.
- We don't know how close to the truth an estimate from this one sample is because we don't know what the truth about the population is. But large random samples almost always give an estimate that is close to the truth. Looking at the pattern of many samples shows how much we can trust the result of one sample.
- Thus, in exit polls and pre-election polls pollsters analyze the results from many different samples before drawing any conclusions.
- Example 1: If you want to study the sporting profile of athletes in any sports, it's better to watch a substantial number of games and analyze their sporting profiles to account for sampling variability.
- Example 2: If you want to analyze human behavior, it is better to gather a substantial number of samples to account for sampling variability.

Immigration poll. A June 2018 Gallup poll asked 1520 randomly selected American adults whether they felt immigration is a good thing or a bad thing for the United States. The poll found that 75% of the respondents said that immigration is a good thing for the United States. Explain to someone who knows no Statistics why we can't just say that 75% of all American adults think immigration is a good thing for the United States.

We have information only from one sample. Even if the sample was chosen randomly and doesn't contain any sampling biases, it is important to analyze the results from many samples to account for sampling variability on a sensitive political topic such as immigration.

4 Starburst activity to understand sampling variability

As seen from the starburst acitivity we did in class, the statistics for estimation of starbursts was closer to parameter in class sample compared to group sample. Individual samples were skewed due to small sample size.

Similarly, we saw large sampling variability in individual sample compared to the group sample. This illustrates the point that if you need to decrease sampling variability in simple random sample, you need to increase the sample size.