Last Time: When taking a sample from a finite population,
- The parameter is a numerical summary of the population (e.g., population mean, \( \mu \); population proportion, \( \pi \))
- The sampling method is unbiased if the center of the distribution of the statistic from many samples is centered at the population parameter
  - Convenience sampling is typically biased
- Random sampling is unbiased
  - We are willing to “generalize” the results from our sample back to the larger population
  - As long as the sampling frame contains every member of the population
- The distribution of sample means from random samples follows a predictable pattern
  - Standard deviation decreases as increase the sample size
  - If the sample size is “large,” the distribution is approximately normal

Example 1: Below are the sample proportions of e-words from the two sampling methods we used yesterday. In the population, 46.6% of the words contain at least one letter e.

(a) Is the convenience sampling method biased? Is the random sampling method unbiased? Explain how you are deciding.

(b) Below are 1000 samples from the Sampling Words applet for samples of size \( n = 5 \) and for samples of size \( n = 20 \). Which is which? How do you know?
(c) Below are 1000 samples (of size \( n = 20 \)) from the Sampling Words applet for population sizes of 268 and of 10,720. The third graph is 1000 samples from the One Proportion applet with \( \pi = 0.466 \). How do the three distributions compare? How do they compare to what our theory-based approach would have predicted?

**Key Result:** When taking random samples from a large population (more than 20 times the size of the sample), the sample proportions behave just like they are coming from a random process. So we will use the same exact methods as in Ch. 1 to determine whether a sample proportion is unlikely to happen by chance (random sampling variability) alone.

**To Turn In:** Founded in 1890, the *Literary Digest* was a popular magazine in the early 1900s. The magazine had correctly predicted the outcomes of the 1916, 1920, 1924, 1928, and 1932 presidential elections by conducting polls. With the Great Depression in full swing, the magazine ventured forth in 1936 to predict another presidential election outcome. Questionnaires were mailed out to more than 10 million adult Americans whose names and addresses were obtained from subscribers to the magazine and also from phone books and vehicle registration lists. More than 2.4 million responses were received, the largest survey that had ever been undertaken at that time, with 57% indicating that they planned to vote for Republican challenger Alf Landon over Democrat incumbent Franklin Roosevelt. Alf Landon actually lost in a landslide, earning only 36.5% of the votes cast in the election.

(a) Identify the population, parameter, sample, sample size, and statistic for this study. Is this a "large" population?

(b) Is "random chance" (from random sampling) a plausible explanation for the discrepancy between 0.57 and 0.365?

1. Calculate the theoretical standard deviation for the distribution of sample proportions using \( \pi = 0.365 \) and \( n = 2,400,000 \) using \( \sqrt{\pi(1-\pi)/n} \).

2. Calculate and evaluate the standardized statistic \((\text{obs-mean})/\text{SD}\) to answer this question.

(c) Suggest two reasons, other than random chance, for why the sample result (0.57) could have been so different from the actual voter percentage for Landon.