7.4: The Central Limit Theorem for Proportions

Recall: A *parameter* is a numerical summary of a population; a *statistic* is a numerical summary of a sample. (For example, the population mean and population standard deviation are parameters; the sample mean and sample standard deviation are statistics.)

The *sampling distribution of a statistic* is the probability distribution of all possible values for that statistic computed from all possible samples of fixed size *n*.

The sampling distribution of the sample proportion:

In this section, the parameter we are interested in is the *population proportion*, usually denoted *p*.

The proportion is the percentage *p* (in decimal form) of the population that possesses some characteristic of interest.

For example, we may be interested in the proportion of children who have a certain medical condition, the proportion of U.S. citizens who received a tax refund, the proportion of students at a certain high school that decide to go to college, or the proportion of nurse candidates who pass the nursing licensure exam.

The *sampling distribution of the sample proportion* is the probability distribution of all possible values for the sample proportion, denoted $\hat{p}$, computed from all possible samples of fixed size *n*.

If *x* is the number of data points in a sample of size *n* that have the characteristic of interest, then the sample proportion is

$$\hat{p} = \frac{x}{n}.$$

In the same manner as for the sample mean, we use the sample proportion $\hat{p}$ to make inferences about the population proportion *p*.

Shape, mean and standard deviation of the sampling distribution of the sample proportion:

Sampling distribution of the sample proportion:

Suppose random samples of size n are taken from a population with population proportion p.

Also suppose that the sample size is small compared to the size of the population.
(Rule of thumb: The sample must be less than 5% of the population size; otherwise we must use a finite population correction factor, which is beyond the scope of this class.)

Then:

The shape of the sampling distribution of $\hat{p}$ is approximately normal, provided that the sample sizes are sufficiently large.

Rule of thumb: to assume the sample proportion is normally distributed, we need both $np \geq 10$ and $n(1-p) \geq 10$.

The mean of the sampling distribution of $\hat{p}$ is $\mu_{\hat{p}} = p$.

The standard deviation of the sampling distribution of : $\hat{p}$ is $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$.

The z-score for a sample proportion $\hat{p}$ is $z = \dfrac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}}$.

Example 1: According to the U.S. Census Bureau, about 35% of U.S. adults age 25 or older have earned a bachelor's degree. Using this information, if a sample of 50 U.S. adults is selected, what is the probability that at least 20 of them have a bachelor's degree?
https://www.census.gov/content/dam/Census/library/visualizations/time-series/demo/fig2.jpg
https://www.census.gov/content/dam/Census/library/publications/2016/demo/p20-578.pdf
https://www.statista.com/statistics/184272/educational-attainment-of-college-diploma-or-higher-by-gender/