Chapter 18 - Sampling Distribution Models

Back in Chapter 12 we learned about **Sampling Variability** (The natural occurrence of the sample statistic to vary from sample to sample).

The **Sampling Distribution Model** shows the behavior of the sample statistic over all possible samples of the same size n.

Yesterday we simulated a Sampling Distribution Model

- The model was unimodal and symmetric (like a Normal curve)
- The model was centered around the population mean, μ.

Sampling Distribution Model for a Proportion

In order to find a model for the distribution of the Sample Proportion, two assumptions must be satisfied

- The sampled values must be independent of each other.
- The sample size, *n*, must be large enough.

Since assumptions are usually hard to check the reasonability of the assumptions can be checked by checking the following conditions:

- **10% Condition**: If sampling has not been made with replacement, then the sample size, n, must be no larger than 10% of the population.
- **Success/Failure Condition**: The sample size has to be big enough so that $np \geq 10$ and $nq \geq 10$ - i.e., there needs to be at least 10 successes and 10 failures.

If assumptions and conditions are met, then the sampling distribution of proportion is modeled by a Normal model with mean equal to the true proportion, $\mu_{\hat{p}} = p$, and standard deviation equal to $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$

- i.e. $N\left(p, \sqrt{\frac{pq}{n}}\right)$.

Sampling Distribution Model for a Mean

For the mean, the only assumption needed is that the observations must be independent. If the sample size is large enough then the population distribution does not matter.

Central Limit Theorem (The Fundamental Theorem of Statistics)

The mean of a random sample has a sampling distribution that is approximately normal with mean, $\mu_{\overline{x}} = \mu$, and standard deviation, $\sigma_{\overline{x}} = \dfrac{\sigma}{\sqrt{n}}$.

The larger the sample, the better the approximation.

Since population parameters are rarely known estimates called **Standard Error** will be used for the standard deviation:

- For the sample proportion: $SE_{\hat{p}} = \sqrt{\dfrac{\hat{p}\hat{q}}{n}}$
- For the sample mean: $SE_{\overline{x}} = \dfrac{s}{\sqrt{n}}$