Chapter 15 - Sampling Distribution Models

15.1 Sampling Distribution of a Proportion

Back in Chapter 10 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

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 $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$

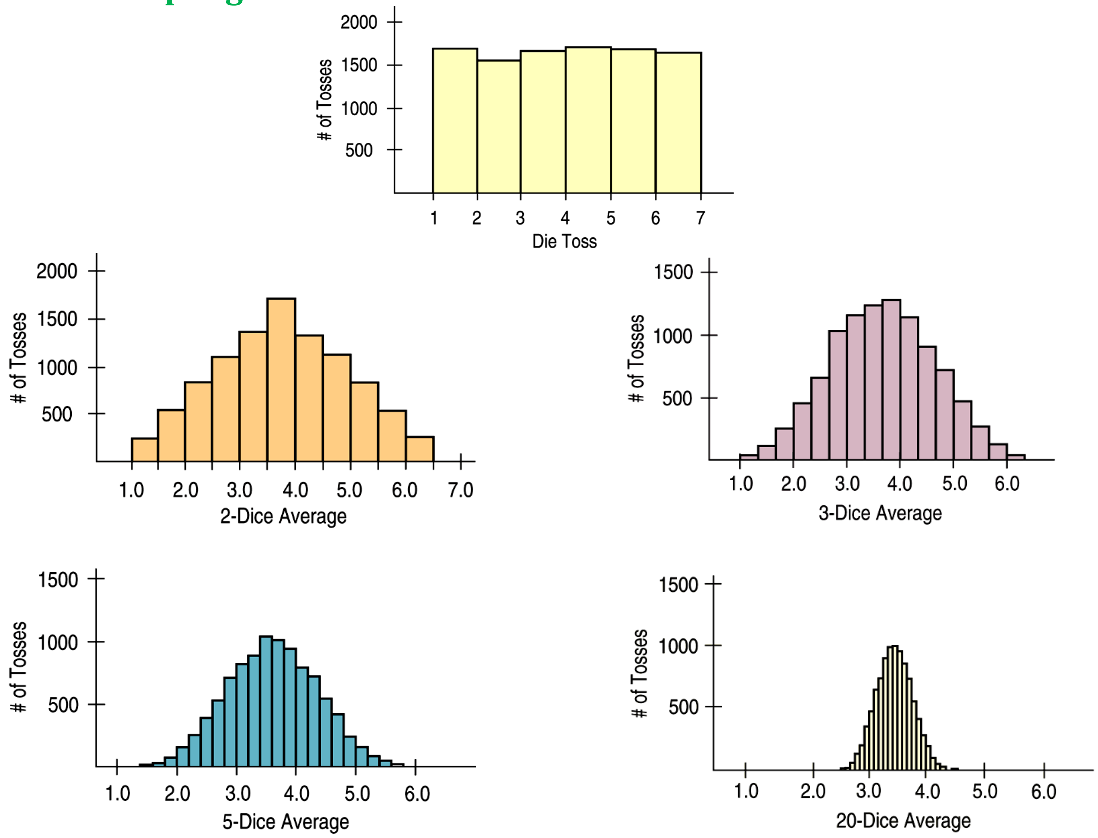
- i.e. $N\left(p, \sqrt{\frac{pq}{n}}\right)$.

15.2 When does the Normal Model Work? Assumptions and Conditions

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

- **Independence Assumption**: The sampled values must be independent of each other.
- **Randomization Condition**: Data should come from randomized source.
- **10% Condition**: The sample size, n, must be no larger than 10% of the population. If the sample size is larger than 10% of the population the remaining individuals are no longer independent of each other.
- **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.

15.3 The Sampling Distribution of Other Statistics



15.4 The Central Limit Theorem: The Fundamental Theorem of Statistics

Sampling Distribution Model for a Mean

For the mean, the only assumption needed is that the observations must be independent and random. Also want the sample size, *n*, to be no more than 10% of the population.

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, $\mu_{\bar{x}} = \mu$, and

standard deviation, $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$.

- i.e. $N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$.

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: $SE_{\overline{x}} = \dfrac{s}{\sqrt{n}}$