Ch. 15 – Sampling Distributions

Expanded def'n: A parameter is: - a numerical value describing some aspect of a pop'n
- usually regarded as constant
- usually unknown

A statistic is: - a numerical value describing some aspect of a sample
- regarded as random before sample is selected
- observed after sample is selected

The observed value depends on the particular sample selected from the population; typically, it varies from sample to sample. This variability is called sampling variability. The distribution of all the values of a statistic is called its sampling distribution.

Def'n: $\hat{p}$ = proportion of ppl with a specific characteristic in a random sample of size n
p = population proportion of ppl with a specific characteristic

The estimate of the standard deviation of a sampling distribution is called a standard error.

General Properties of the Sampling Distribution of $\hat{p}$:

Let $\hat{p}$ and p be as above. Also, $\mu_{\hat{p}}$ and $\sigma_{\hat{p}}$ are the mean and standard deviation for the distribution of $\hat{p}$. Then the following rules hold:

Rule 1: $\mu_{\hat{p}} = p$. (Textbook uses $\mu(\hat{p})$)

Rule 2: $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}} = \sqrt{\dfrac{pq}{n}}$. (standard error → $\hat{\sigma}_{\hat{p}}$)

Ex15.1) Suppose the population proportion is 0.5.

a) What is the standard deviation of $\hat{p}$ for a sample size of 4?

b) What is the smallest that n can be so that the sample proportion has a standard deviation of at most 0.125?

Rule 3: When n is large and p is not too near 0 or 1, the sampling distribution of $\hat{p}$ is approximately normal. The farther from p = 0.5, the larger n must be for accurate normal approximation of $\hat{p}$. Thus, if np and $n(1-p)$ are both sufficiently large ($\geq$ 15), then it is safe to use a normal approximation.

Further assumptions: the sample should always be random and, if sampling without replacement, the sample should be less than 10% of the population.

Using all 3 rules, the distribution of $\hat{p}$ is approximately normal.

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

Ex15.2) Suppose that the true proportion of people who have heard of Sidney Crosby is 0.87 and that a new random sample consists of 158 people.

a) Find the mean and standard deviation of $\hat{p}$.

b) What can you say about the shape of the distribution of $\hat{p}$?

c) What is the probability of getting a sample proportion greater than 0.94?

d) What is the probability of less than 140 people hearing of Sidney Crosby in the sample?

Sampling Distribution of Mean
How does the sampling distribution of the sample mean compare with the distribution of a single observation (which comes from a population)?

Ex15.3) An epically gigantic jar contains a large number of balls, each labeled 1, 2, or 3, with the same proportion for each value.

Let Y be the label on a randomly selected ball. Find μ_Y and σ_Y.

Let $\{Y_1, Y_2\}$ be a random sample of size $n = 2$. Find the sampling distribution of the sample mean $\bar{Y}$. Calculate $\mu_{\bar{y}}$ and $\sigma_{\bar{y}}$.

There are ___ possible samples:

$\bar{y}$					
$P(\bar{Y} = \bar{y})$					

Progressing further with inference, we can now discuss the following properties.

General Properties of the Sampling Distribution of $\bar{y}$ (or $\bar{x}$):
Let $\bar{y}$ denote the mean of the observations in a random sample of size n from a population having mean μ and standard deviation σ. Also, $\mu_{\bar{y}}$ and $\sigma_{\bar{y}}$ are the mean and standard deviation for the distribution of $\bar{y}$. Then the following rules hold:

Rule 1: $\mu_{\bar{y}} = \mu$.

Rule 2: $\sigma_{\bar{y}} = \dfrac{\sigma}{\sqrt{n}}$.

Note also that:
1. The spread of the sampling dist'n of $\bar{y}$ is smaller than the spread of the pop'n dist'n.
2. As n increases, $\sigma_{\bar{y}}$ decreases.

Ex15.4) Suppose the population standard deviation is 10.

a) What is the std. dev. of the sample mean for some of the following sample sizes?
n = 1, 2, 4, 9, 16, 25, 100

b) What is the smallest that n can be so that the sample mean has a standard deviation of at most 2?

Rule 3: When the population distribution is normal, the sampling distribution of $\bar{y}$ is also normal for any sample size n.

Combining the 3 rules, if the population distribution is $N(\mu, \sigma)$, then $\bar{Y}$ is $N(\mu, \sigma/\sqrt{n})$.

Rule 4 (**Central Limit Theorem**): When n is sufficiently large, the sampling distribution of $\bar{y}$ is well approximated by a normal curve, even when the population distribution is not itself normal. The Central Limit Theorem can safely be applied if n is at least 30.

Using all 4 rules, if n is large and/or the population is normal, then the sampling distribution of $\bar{Y}$ is approximately normal.

$$Z = \frac{\bar{Y} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

Ex15.5) Suppose the mean length of all episodes of a (formerly) hilarious series is 20.834 minutes, whereas the standard deviation is 0.593 minutes. Let $\bar{Y}$ be the average length for a random sample of 100 episodes.

a) Find the mean and standard deviation of $\bar{Y}$.

b) What can you say about the shape of the distribution of $\bar{Y}$?

c) What is the probability of getting a sample mean between 20.7 and 21 minutes?

d) Can you find $P(20.7 \leq Y \leq 21)$, where Y is the length of a single randomly selected episode? How would this value compare with the one in part c)?