7.1 Point Estimates

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

Estimation is the assignment of value(s) to a population parameter based on a value of the corresponding sample statistic.
An estimator is a rule used to calculate an estimate.
An estimate is a specific value of an estimator.
Note: in this chapter, always assuming an SRS (single random sample).

Def'n: A point estimate is a *single number* that is our "best guess" for the parameter.
→ like a *statistic*, but more precise towards parameter estimation.
An interval estimate is an *interval of numbers* within which the parameter value is believed to fall.

- Notation:
- Let θ be a generic parameter.
- Let $\hat{\theta}$ be an estimator – a statistic calculated from a random sample
- Consequently, $\hat{\theta}$ is an r.v. with mean $E(\hat{\theta}) = \mu_{\hat{\theta}}$ and std. dev. $\sigma_{\hat{\theta}}$

Ex7.1) $\hat{p}$ = proportion of ppl with a specific characteristic in a random sample of size n
p = population proportion of people with a specific characteristic

7.3 Sampling Distributions

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.

5.3.2 Central Limit Theorem & 7.3.2 Sample Mean

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

Ex7.2) An epically gigantic jar contains a large number of balls, each labeled 1, 2, or 3, with the same proportion for each value.

Let X be the label on a randomly selected ball. Find μ_X and σ_X.

x	$P(X = x)$

$\mu_X = \sum x_i P(X = x_i) =$
$E(X^2) = \sum x_i^2 P(X = x_i) =$
$\sigma_X^2 = E(X^2) - [E(X)]^2 =$

Let $\{X_1, X_2\}$ be a random sample of size $n = 2$. Find the sampling distribution of the sample mean $\overline{X}$. Calculate $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$.

There are ___ possible samples:

$\bar{x}$					
$P(\overline{X} = \bar{x})$					

$\mu_{\bar{x}} = E(\overline{X}) = \sum \bar{x}_i P(\overline{X} = \bar{x}_i) =$
$E(\overline{X}^2) = \sum \bar{x}_i^2 P(\overline{X} = \bar{x}_i) =$
$\sigma_{\bar{x}}^2 = E(\overline{X}^2) - [E(\overline{X})]^2 =$

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

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

Rule 1: $\mu_{\bar{x}} = \mu$ *Rule 2*: $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

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

Ex7.3) 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) How large must n (sample size) 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 $\bar{x}$ is also normal for any sample size n.

Combining the 3 rules, if the population distribution is $N(\mu, \sigma^2)$, then $\bar{X}$ is $N(\mu, \sigma^2/n)$.

Rule 4 (**Central Limit Theorem**): When n is sufficiently large, the sampling distribution of $\bar{x}$ 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 exceeds 30.

Using all 4 rules, if n is large and/or the population is normal, then

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

Ex7.4) 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 $\bar{X}$ be the average length for a random sample of 100 episodes.

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

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

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

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

7.3.1 Sample Proportion

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

Let $\hat{p}$ and p be as defined earlier. 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: $\mu_{\hat{p}} = p$ *Rule 2*: $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

Ex7.5) Suppose the population proportion is 0.5.

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

b) How large must n (sample size) 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 (≥ 15), then it is safe to use a normal approximation.

Using all 3 rules, if $np \geq 15$ and $n(1-p) \geq 15$, then

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

Ex7.6) Suppose that the true proportion of people who have heard of Sidney Crosby is 0.87 and that a new sample consists of 158 people.

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

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

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