I. Population – (Universe, Target Population)

N: # of elementary units (or elements) in the population.

X_i , Y_i : the i^{th} elementary unit of the random variables X, Y, where i = 1, 2, 3, …, N.

A. **Population Parameters** – descriptive measures of a population.

1. Population Total of X

$$X = \sum_{i=1}^{N} X_i \tag{2.1}$$

2. Population Mean of X

$$\overline{X} = \frac{X}{N} = \frac{\sum_{i=1}^{N} X_i}{N} = \mu_X \tag{2.2}$$

3. Population Proportion (If X is dichotomous)

X_i = 1 if the attribute X is present in the i^{th} element, or

X_i = 0 if the attribute X is not present in the i^{th} element.

$X = \sum_{i=1}^{N} X_i$ (Sum of all 1's),

$$P_X = \frac{X}{N}. \tag{2.3}$$

4. Population Variance of X

$$\sigma_X^2 = \frac{\sum_{i=1}^{N}(X_i - \overline{X})^2}{N} = \frac{\sum_{i=1}^{N} X_i^2 - \frac{\left(\sum_{i=1}^{N} X_i\right)^2}{N}}{N} \tag{2.5}$$

$\sigma_X^2 = P_X(1 - P_X)$, for a dichotomous variable X. (2.6)

5. Population Standard Deviation of X

$$\sigma_X = \sqrt{\sigma_X^2} \tag{2.4}$$

6. Coefficient of Variation

$$V_X = \frac{\sigma_X}{\overline{X}} \tag{2.7}$$

$V_Y = \sqrt{\frac{1 - P_Y}{P_Y}}$, for a dichotomous variable Y. (2.8)

7. Relative Variance

$$V_X^2 = \frac{\sigma_X^2}{\overline{X}^2}$$

II. Sample

n: # of elementary units in the sample.

A. **Sample Statistics** – descriptive measures of a sample, used to estimate unknown population parameters.

1. Sample Total of X

$$x = \sum_{i=1}^{n} x_i \tag{2.9}$$

2. Sample Mean of X

$$\bar{x} = \frac{x}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{2.10}$$

3. Sample Proportion if X is dichotomous

x_i = 1 if the attribute X is present in the i^{th} element, or

x_i = 0 if the attribute X is not present in the i^{th} element.

$x = \sum_{i=1}^{n} x_i$ (Sum of all 1's)

$$p_x = \frac{x}{n}. \tag{2.11}$$

4. Sample Variance of X

$$s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left[\sum_{i=1}^{n} x_i\right]^2}{n}}{n-1} \tag{2.12}$$

5. Sample Variance for a dichotomous X

$$s_x^2 = \frac{np_x(1-p_x)}{n-1} \approx p_x(1-p_x) \text{ for n > 20.} \tag{2.13}$$

6. Sample Standard Deviation of X.

$$s_x = \sqrt{s_x^2} \tag{2.14}$$

III. Estimates of Population Characteristics

A. Estimates

1. Estimate of Population Mean:

$\hat{\mu}_X = \bar{x}$

2. Estimate of Population Proportion:

$\hat{P}_X = p_x$

3. Estimate of Population Total:

$$\hat{X} = N\hat{\mu}_X = N\bar{x} = \left(\frac{N}{n}\right)x = x' \quad (x \text{ is sample total}) \qquad (2.15)$$

4. Estimate of Population Variance:

$$\hat{\sigma}_X^2 = \left(\frac{N-1}{N}\right)s_x^2 \qquad (2.16)$$

For N large, $\hat{\sigma}_X^2 \approx s_x^2$

IV. Sampling Distribution - The relative frequency distribution of a sample statistic over all possible samples

A. In general, consider:
d: a population parameter
$\hat{d}$: a sample estimate of d.
T: possible samples of size n drawn from a population of size N.

1. Mean of the Sampling Distribution

$$E(\hat{d}) = \sum_{i=1}^{C} \hat{d}_i \pi_i, \qquad (2.17)$$

for C possible values of $\hat{d}$, and $\pi_i = \frac{f_i}{T}$,

where f_i is the number of times that a particular $\hat{d}_i$ occurs.

2. Variance of the Sampling Distribution

$$\text{Var}(\hat{d}) = \sum_{i=1}^{C} [\hat{d}_i - E(\hat{d})]^2 \pi_i \qquad (2.18)$$

$$= \sum_{i=1}^{C} \hat{d}_i^2 \pi_i - E^2(\hat{d}) \qquad (2.19)$$

3. Standard Error of $\hat{d}$

$$\text{SE}(\hat{d}) = \sqrt{Var(\hat{d})} \qquad (2.20)$$

V. Characteristics of Estimates of Population Parameters

A. Measures of Accuracy of an Estimate

1. **Bias**

$B(\hat{d}) = E(\hat{d}) – d.$ (2.27)

$\hat{d}$ is unbiased for d if $E(\hat{d}) = d$, ie. $B(\hat{d}) = 0$.