Chapter 14 ML 4202

- The larger the sample is, the smaller the sampling error
- If sample size is quadrupled, data collection cost is almost quadrupled, but the level of sampling error is reduced by only 50%

Budget Available:

- The sample size for a project is often determined, at least indirectly, by the budget available.

Rule of Thumb:

- Potential clients may specify in the RFP (request for proposal) that they want a sample of 200, 400, 500, or some other size. Sometimes, this number is based on desired sampling error

Number of Subgroups Analyzed:

- Consideration must be given to the number and anticipated size of various subgroups of the total sample that must be analyzed and about which statistical inferences must be made. Ex: a researcher might decide that a sample of 400 is quite adequate overall. However, if male and female respondents must be analyzed separately and the sample is expected to be 50% male and 50% female, then the expected sample size for each subgroup is only 200.
- The larger the number of subgroups to be analyzed, the larger the required total sample size. It has been suggested that a sample should provide, at a minimum, 100 or more respondents in each major subgroup and 20 to 50 respondents in each of the less important subgroups.

General Properties:

- Normal Distribution is crucial to classical statistical inference. Several reasons:
    - 1. Many variables encountered by marketers have probability distributions that are close to the normal distribution. Ex: number of cans, bottles, or glasses of soft drinks consumed by soft drink users etc
    - 2. The normal distribution is useful for a number of theoretical reasons; one of the more important of these relates to **the Central Limit Theorem** (CLT): Idea that a distribution of a large number of sample means or sample proportions will approximate a normal distribution, regardless of the distribution of the people from which they were drawn.
    - 3. The normal distribution is a useful approximation of many other discrete probability distributions.
- **Normal distribution**: Continuous distribution that is bell-shaped and symmetric about the mean; the mean, median, and mode are equal.
    - Bell-shaped and only has one mode. The mode is a measure of central tendency and is the particular value that occurs most frequently.

- Symmetric about its mean… so it is not skewed and the (mean, median, and mode) are equal
    - Uniquely defined by its mean and SD
    - Total area under the curve is equal to 1, so it takes in all observations
    - 34.13% chance of selecting from the distribution
    - The area between the mean and +/- one SD is 68.26% area under the curve. **Proportional property of the normal distribution**: Feature that the number of observations falling between the mean and a given number of standard deviations from the mean is the same for all normal distributions.
    - Z values (1=68.26%) (2=85.44%) (3=99.74%)

- **Standard normal distribution**: Normal distribution with a mean of zero and a SD of one
- Z=
    - Z=
    - Where x= value of the variable
    - μ= mean of the variable
    - σ= standard deviation of the variable
    - **Standard deviation:** Measure of dispersion calculated by subtracting the mean of the series from each value in a series, squaring each result, summing the results, dividing the sum by the number of items, minus 1, and taking the square root of this value.

Population and Sample Distributions

- A sample is the subset of the population.
- **Population distribution**: Frequency distribution of all elements of the population. It has a mean, usually represented by the Greek letter (μ); and a standard deviation, usually represented by the Greek letter (σ)
- **Sample distribution**: A frequency distribution of all the elements of an individual (single) sample. In a sample distribution, the mean is usually represented by *X* and the standard deviation is usually represented by *S*.

Sampling Distribution of the Mean

- **Sampling distribution of the mean**: Theoretical frequency distribution of the means of all possible samples of a given size drawn from a particular population; it is normally distributed.
- If samples are sufficiently large and random, then the resulting distribution of sample means will approximate a normal distribution. This assertion is based on the central limit theorem, which states that as sample size increases, the distribution of the means of a large number of random samples taken from virtually any population approaches a normal distribution with a mean equal to μ and a standard deviation (referred to as *standard error*) Sx bar where n= sample size
    - S x bar=
- **Standard error of the mean**: Standard deviation of a distribution of sample means.

- It is important to note that the central limit theorem holds regardless of the shape of the population distribution from which the samples are selected.

Notation for Means and Standard Deviations box page 411

Basic Concepts:

- Ex: Consider a case in which a researcher takes 1,000 simple random samples of size 200 from the population of all consumers who have eaten as a fast-food restaurant at least once in the past 30 days. The purpose is to estimate the average number of times these individuals eat at a fast-food restaurant in an average month.
- The sampling distribution of the mean for simple random samples that are large $N \geq 30$ it has the following characteristics:
    - The distribution is a normal distribution
    - The distribution has a mean equal to the population mean
    - The distribution has a standard deviation (the standard error of the mean) equal to the population standard deviation divided by the square root of the sample size:
        - Page 413
        - This statistic is referred to as the standard error of the mean (instead of the SD) to indicate that it applies to a distribution of sample means rather than to the standard deviation of a sample or a population. This calculation applies ONLY to a simple random sample.
        - Formula doesn't account for any type of bias

Making Inferences on the Basis of a Single Sample

- What is the probability that any one simple random sample of a particular size will produce an estimate of the population mean that is within one standard error (plus or minus) of the true population mean? 68.26 (Z=1) 95.44 (Z=2) 99.74 (Z=3)

Point and Interval Estimates

- The sample mean is the best **point estimate** of the population mean. **Point estimate**: Particular estimate of a population value.
- The distance between the sample mean and the true population mean is the **sampling error.**
- **Interval estimate**: Interval or range of values within which the true population value is estimated to fall. In addition to stating the size of the interval, the researcher usually states the