Probability and Samples

Sampling

- We want the results from our sample to be true for the population and not just the sample.
- But our sample may or may not be representative of the population.
- *Sampling error* – the discrepancy, or amount of error, between a sample statistic and its corresponding population parameter.
- Collecting a second sample will likely lead to a different sample mean.

Point Estimates

- One of our fundamental questions is: “How well does our sample statistic estimate the value of the population parameter?”
- Equivalently, we may ask “Is our point estimate good?”
- A point estimate is a statistic (e.g., $\bar{X}$) that is calculated from sample data in order to estimate the value of the population parameter (e.g., $\mu$).
Point Estimates

- What makes a point estimate “good”?
  - A good estimate is one that is close to the actual value
  - A difference score, or deviate score, ($\bar{X} - \mu$) can tell us how close the sample statistic is to the population parameter
  - What statistic should we use to measure the average “goodness?”
    - Standard deviation

Sampling Distribution

- Draw a sample of a given size from the population
- Calculate the point estimate
- Repeat the previous two steps many times
- Draw a frequency distribution of the point estimates
- That distribution is called a distribution of sample means which is a type of sampling distribution

Central Limit Theorem

- The central limit theorem states that the shape of a sampling distribution will be normal as long as the sample size (n) is sufficiently large (> ~30)
- The mean of the sampling distribution will equal the mean of the population (µ)
- The standard deviation of the sampling distribution (the standard error of the mean) will equal the standard deviation of the population (σ) divided by the $\sqrt{n}$
Standard Error of the Mean

- The *standard error of the mean* is the standard deviation of the sampling distribution.
- \( \sigma_M \) or \( \sigma_x \) represents the standard error of the mean.
- Thus, it is a measure of how good our point estimate is likely to be.
- When \( \sigma_M \) is small, \( \bar{X} \) is likely to be close to \( \mu \).

Which Sampling Distribution Is Better?

Factors Influencing \( \sigma_M \)

- What influences the size of the standard error of the mean?
  - Sample size (n)
    - A sample mean based on a single observation will not be as accurate as a sample mean based on 10 or 100 observations.
    - law of large numbers
  - Population standard deviation (\( \sigma \))
    - When \( n = 1 \), \( \sigma_M = \sigma \)
Standard Error of the Mean

The standard error of the mean can be computed from the standard deviation of the population:

$$\sigma_M = \frac{\sigma}{\sqrt{n}}$$

Probability and the Distribution of Sample Means

The distribution of sample means can be used to find the probability associated with any given sample.

What is the probability that a randomly selected group of 20 people will have a mean IQ greater than 105?

- IQs are normally distributed in the population
- $\mu = 100$, $\sigma = 15$
- $n = 20$

Probability and the Distribution of Sample Means

- The sampling distribution is normal because IQs are normally distributed in the population
- The sampling distribution’s mean is $\mu = 100$ because the population $\mu = 100$
- The sampling distribution’s standard deviation is $\sigma_M = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{20}} \approx 3.35$
Probability and the Distribution of Sample Means

1. Find the z-score that corresponds to 105 in a normal distribution with a mean of 100 and a standard deviation of 3.35
2. \( z = \frac{105 - 100}{3.35} = 1.49 \)
3. Find area above \( z = 1.49 \) in a table
4. Probability = 0.0681