1 Sampling Distributions

1.1 Sampling Distributions of Sample Proportions

As we begin to use sample data to draw conclusions about a larger population, we must be clear about whether a number describes a sample or a population.

Parameter: is a number that describes some characteristic of a population

Statistic: is a number that describes some characteristic of a sample

Unbiased Estimator: a sample proportion or sample mean that is equal to the population proportion or population mean

 \hat{p} (sample proportion) estimates p (population proportion)

 \overline{x} (sample mean) estimates μ (population mean)

 s_x (sample standard deviation) estimates σ (population standard deviation)

Rather than showing real repeated samples, imagine what would happen if we were to actually draw many samples. Now imagine what would happen if we looked at the sample proportions for these samples. The histogram we'd get if we could see all the proportions from all possible samples is called the sampling distribution of the sample proportions.

- We would expect the histogram of the sample proportions to center at the true proportion, p, in the population.
- The spread is calculated as standard deviation based on the true proportion, *p*, and the sample size, *n*. As the sample size gets larger the standard deviation will get smaller.
- The shape of the histogram would be unimodal and symmetric.
- More specifically, a normal model is just the right one for the histogram of sample proportions.

Assumptions and Conditions

- Randomness: The sample should be a simple random sample of the population
- Independence (10% Condition): The sample size, n, must be no larger than 10\$ of the population
- Normality (Large Counts Condition): The sample size has to be big enough so that both number of successes and number of failures are at least 10. We also refer to this as the Success/Fail Condition.

Provided that the sampled values are independent and the sample size is large enough, the sampling distribution of p is modeled by a normal model with:

Sample proportions: $\hat{p}=\frac{\# {\rm successes}}{{\rm sample size}}$ Mean of sample proportions: $\mu_{\hat{p}}=p$

Standard deviation of sample proportions: $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

1.2 Sampling Distributions of Sample Means

The sampling mean distribution notation:

- Parameters: μ and σ
- Statistics: \bar{x} and s
- Sampling Distribution Mean: $\mu_{ar{x}}$

• Sampling Distribution Standard Deviation: $\sigma_{\bar{x}}$

Conditions for sample means:

- Random As long as the sampling method is random, our mean is an unbiased estimator.
- Independent When sampling, we have to make sure the 10% Condition is satisfied.
- Normal No longer checking Large Counts, instead we have the Central Limit Theorem.

Central Limit Theorem: The central limit theorem (CLT) states that when the sample size is sufficiently large, a sampling distribution of the mean of random variable will be approximately normally distributed.

The central limit theorem requires that the sample values are independent of each other and that n is sufficiently large.

Therefore, if the population distribution is normal, then so is the sampling distribution of \bar{x} . This is true no matter what the sample size of n is.

If the population distribution is not normal, the central limit theorem tells us that the sampling distribution of \bar{x} will be approximately normal in most cases of $n \ge 30$.

Summary:

- Shape: approximately normal
- Center: $\mu_{\bar{x}} = \mu$
- Spread: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

1.3 Combining Sample Proportions and Sample Means

Sampling distribution of $\hat{p}_1 - \hat{p}_2$, where \hat{p}_1 is the sample proportion from the first group and \hat{p}_2 is the sample proportion from the second group.

If both samples were randomly selected from the population, then $\mu_{\hat{p}_1-\hat{p}_2}=p_1-p_2.$

If both samples satisfy the 10% condition, then $\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\left(\frac{p_1(1-p_1)}{n_1}\right) + \left(\frac{p_2(1-p_2)}{n_2}\right)}$

If both samples satisfy the success/fail condition, then the shape is approximately normal.

Samplying distribution of $\bar{x}_1 - \bar{x}_2$ where \bar{x}_1 is the sample mean from the first group and \bar{x}_2 is the sample mean from the second group:

Same conditions apply for the center and spread, and remember that the central limit theorem rather than success/fail is used for shape.

The center is $\mu_{\bar{x}_1-\bar{x}_2} = \mu_1 - \mu_2$ and the spread is $\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.