**Sampling and Sampling Distribution:**

**1. Point Estimators (Inferential Statistics):** The largest part of statistics is about taking a sample from the population and using those as an estimation of the overall population. The idea of point estimation is to calculate a single value(statistic) by using sample data from the population. For example, it can be inference about the quality of some products based on sample data observations.

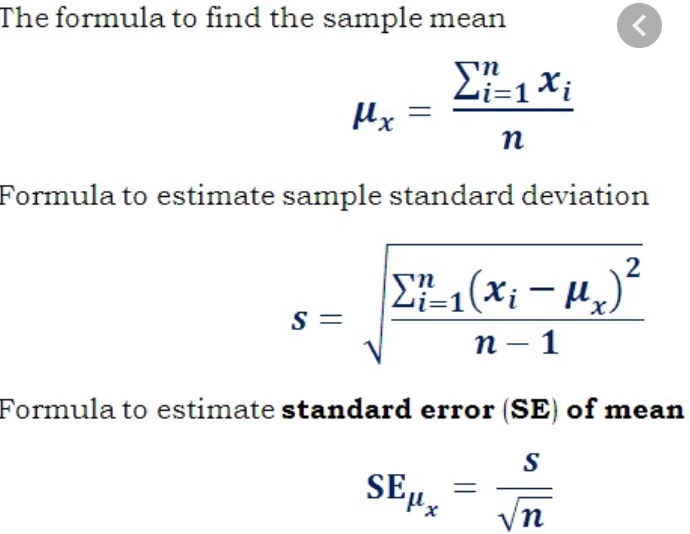
* When we don’t know parameters of the distribution, such as mean or standard deviation, we can make an estimation about these parameters by calculation **point estimators, such as the sample mean or sample standard deviation**. And we should remember that point estimates are never perfect since we are working only with a relatively small sample of the population. This is commonly referred to as the **margin of error**. And this error component is expressed as a **confidence interval (function of sample size and degree of “confidence”).**
* **So we use the point estimators to estimate the population parameters for instance sample mean, std, and proportion for estimating the population mean, std, and proportion.**

**2. Sampling and Sampling Distribution(Inferential Statistics):**

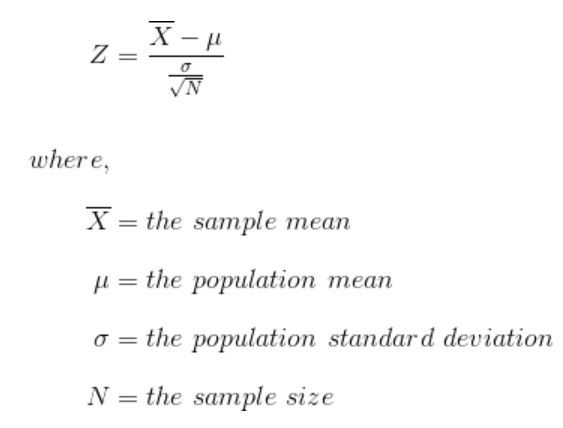
* If we take many random samples from the population each with its own sample mean and then create a distribution based on all of those samples mean, the mean of that sampling distribution based on all of those sample means, the mean of that sampling distribution is equal to the mean of population.
  + The ***expected value*** of the sampling distribution of xbar (sample mean) is at best going to be an estimate of µ (population mean).
  + The best we are going to do is find an interval (function of sample size and degree of “confidence”) estimate for µ.

**2.1. Sampling distribution:** the first step is to draw a histogram of samples mean to find the sampling distribution.

**2.2 Standard (margin) Error of the Mean: the second step**

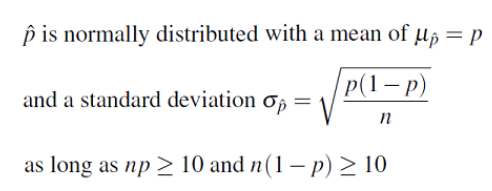
****

* First of all because of all samples have the same size by calculating the samples’ means we are calculating the mean of all observations because each sample mean is a weighted average of its observation.
* As ***n(sample size)*** increases the **SE** becomes smaller which means the standard error of the mean of the population moves toward zero. And hence the mean of sampling distribution will be closer to population mean. Yet increasing the sample size from some point doesn’t help us anymore.
* The **SE** is the ***same*** for all samples of the same size.
* The standard error is another name for standard deviation. It is just the standard deviation of the sampling distribution.
* Standard deviation allows us to calculate z-scores and hence the area (probability) under the curve for certain regions.
* Most often we don’t know the population Std therefore we have to estimate it and make necessary adjustment.
* Keep in mind that SE = Standard deviation of sampling distribution. Hence based on this fact we can draw a normal dist using z-score (if we know δ; µ; or n>30. Otherwise we use t-dist).

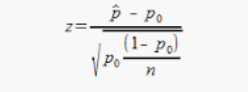


**3. Sample Proportions (point estimator):** The ***sample proportion (p bar)*** is similar to ***sample mean (x bar)***. It is just one of the numerous samples that could be taken from population. If someone else went out and did the sampling again, it is likely there would be different result. However, sample proportions are different than sample mean ***(x bar)*** in important ways:

1. **P bar** represent what is essentially a nominal (usually binary) variables; Yes/No, … .
2. It is a count/frequency of a nominal variable of interest (x/n)
3. In that sense it is related to *binomial distribution*.



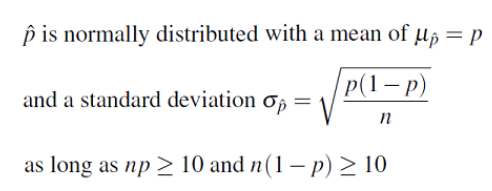
* Expected value of the (pbar) is equal to the mean of all potential values of (pbar) is equal to the population proportion (p). **E(pbar) = pbar = x/n**. where **x** is the number of observations that have what you are interested in. **n** is total sample observations.
* We refer to the Std of the sample proportion as the standard error of the sample proportion.
* However, the standard error of pbar is dependent on having finite or infinite population of interest. Yet in almost all cases, we are assuming the population is “infinite” or very large in relation to the sample size (np >= 10 and n(1-p) >= 10) or (n/N =< 0.05).
* For probability calculation; first calculate **z-value** using below formula, and then follow =NORM.DIST function in Excel. Where p^ is the probability we looking for, p0=x/n is the expected value or mean of the sample proportion.



* Using z-score cumulative probabilities we are able to find probability intervals ***BUT*** note this is not a confidence interval per se but we are simply trying to find slices of the sampling distribution.
* V.imp note: the number of sample doesn’t have anything to do with Std because **p** and **n** are the same for any number of samples. But increasing the number of samples reduces ***the standard error of the mean***, which means as much as we increase the number of sample, we move closer to population mean.

**4. Sampling Distribution for Proportions:**

* We use the same formula of a sample proportion for sampling distribution for proportions as well. Note that in below formulas p represent the mean of all samples. And n represent the sample size.



* Increasing the ***sample size*** reduces the ***standard deviation*** of sampling distribution or ***standard error.*** But if we increase the ***number of samples with the same size*** we become closer and closer to the ***expected standard error*** that we calculate from the above formula
* All above notes for normal sampling distribution apply here as well.