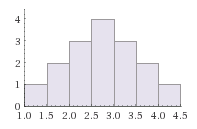
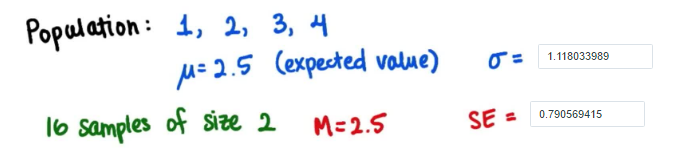
***Udacity Data Analyst Track***

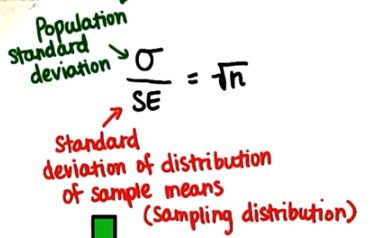
**I. Into to Descriptive Stats**

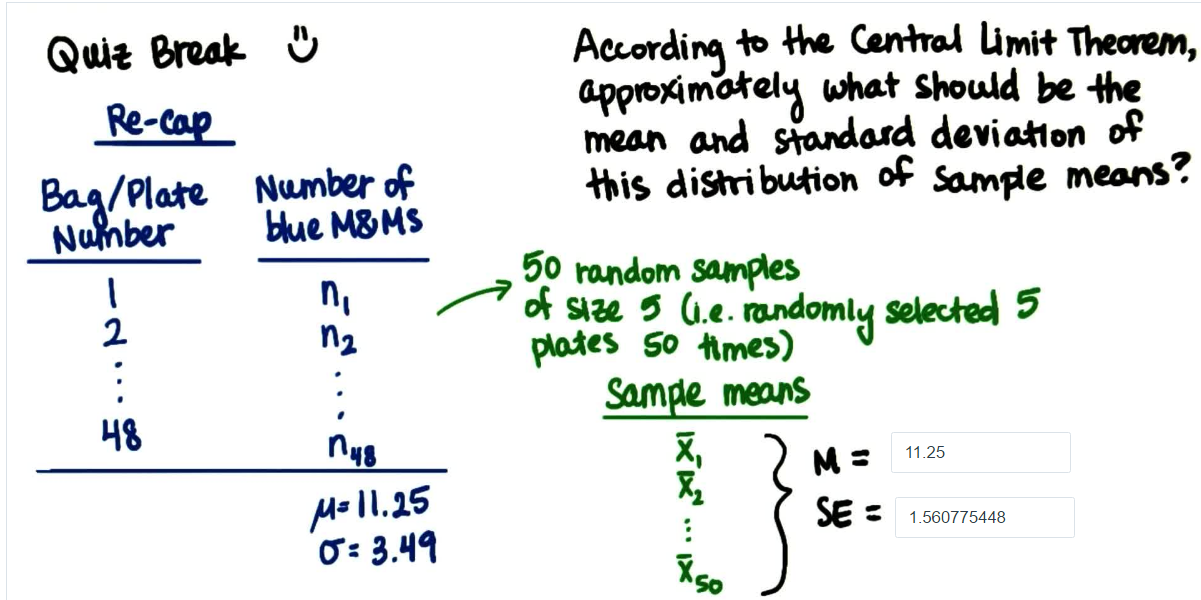
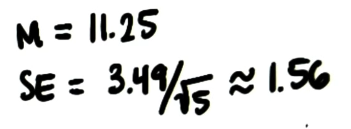
6. Sampling Distributions

* By knowing the mean and SD of a normal distribution, we can compare any value in the population to the rest of that population by finding the % of values greater than or less than that value
* To compare *samples* of a population, we can find the mean *of a sample* + of *other samples* and compare them
* Ex: tetrahedral die 🡪 sides of 1, 2, 3, 4, w/ mean = 2.5 = **expected value** if we take a sample of this population
* Ex: Roll tetrahedral die twice + take average of the 2 rolls 🡪 chances of getting >= 3
* Sample size = 2
* Samples of Size 2 = 16 🡪1,1 1,2 1,3 1,4 2,1 2,2 2,3 2,4 3,1 3,2 3,3 3,4 4,1 4,2 4,3 4,4
* Sample means 🡪 1, 1.5, 2, 2.5, 1.5, 2, 2.5, 3, 2, 2.5, 3, 3.5, 2.5, 3, 3.5, 4



* Probability of having sample mean >= 3 🡪 6 / 16 = **0.375**
* Can easily calculate probability for discrete samples in a discrete population (like above),
* Can’t do this in huge, real-life populations + don’t want to 🡪 16 samples from a population of only 4
* See in this example that the distribution of sample means is normal and the sample mean here = the population mean here (2.5)
* Take random sample 🡪 find mean of it 🡪 to compare this mean w/ others in the sample distribution, we need **sample SD**
* 
* The Ratio of sigma (population) to the SE (sample) 🡪 **1.414213562 = Sqrt(2**) 🡪 *Square root of our sample size*



* So, SE (sample SD) = Sigma / Sqrt(n)
* In a distribution of sample means, where every sample is size n, the SD of the distribution of samples is population SD / sqrt(n) = **central limit theorem** (holds true for *any* population of *any* shape)
* Draw samples from a population and find the mean and distribute those means, assuming the sample size is large enough, it should be normal w/ SE = sigma/Sqrt(n)
* Ex: roll 1 die 100 times, distribution ends up being uniform, since n = 1, so SE = sigma / sqrt(1) = sigma
* Ex: roll 2 die 100 times, distribution ends up being normal 🡪 SE = sigma / sqrt(2)
* Wider sample size = skinner distribution 🡪 denominator is bigger 🡪 SE is smaller
* As n increases, SE decreases, as does our interval in which we are sure the population mean lies (should be equal to the sample distribution means’ SD)
* Need to quadruple n to get ½ the sampling error 🡪 Sigma/sqrt(4n) = sigma/sqrt(4)\*sqrt(n) = sigma/2sqrt(n)
*  
* **The Central Limit Theorem** = used to help understand the following facts regardless of whether a population distribution is normal or not:
* *the mean of the sample means is the same as the population mean*
* the SD of the sample means is always equal to the **standard error** (i.e. SE = sigma/Sqrt(n) )
* the distribution of sample means will become *more normal* as sample size, n, increases
* **Sampling Distribution** **of a statistic** = the distribution of that statistic.
* May be considered as the distribution of a statistic for *all* possible samples from the *same* population of a given size.
* Example: We are interested in average height of trees in a particular forest.
* 5 students measure a sample of 20 trees + return w/ the average tree height from their samples.
* Sample results: 35.23 , 36.71, 33.21, 38.2, 35.54
* *If it is known the population average = 36 feet w/ SD = 2 feet, how many Standard errors is the students average away from the population mean?*
* To solve this problem 1st need to find the average of the students averages so
* x¯ = 35.23+36.71+33.21+38.2+35.54 / 5 = **35.78**
* Now find Standard error of the sample 🡪 SE = SD/Sqrt(n) = 2 / 5 = **0.4**
* To get the # of SE’s away from the mean our observation is, use the z-score formula 🡪 x¯ - mu / SE 🡪 (35.7836 – 46) / 0.4 = **0.55**
* So our sample distribution is relatively close to the population distribution!
* Known average time to deliver a pizza = 22.5 minutes w/ SD = 2 minutes. I ordered pizza every week for the last 10 weeks w/ average time = 18.5 minutes. What is the probability that get this average?
* N = 10, sample Mean = 18.5, SE = 22.5/sqrt(10) 🡪 z-score = (18.5 – 22.5) / SE 🡪 **0.2887**
* If I continue to order pizzas for eternity what could I expect this average to get close to? 🡪 22.5

