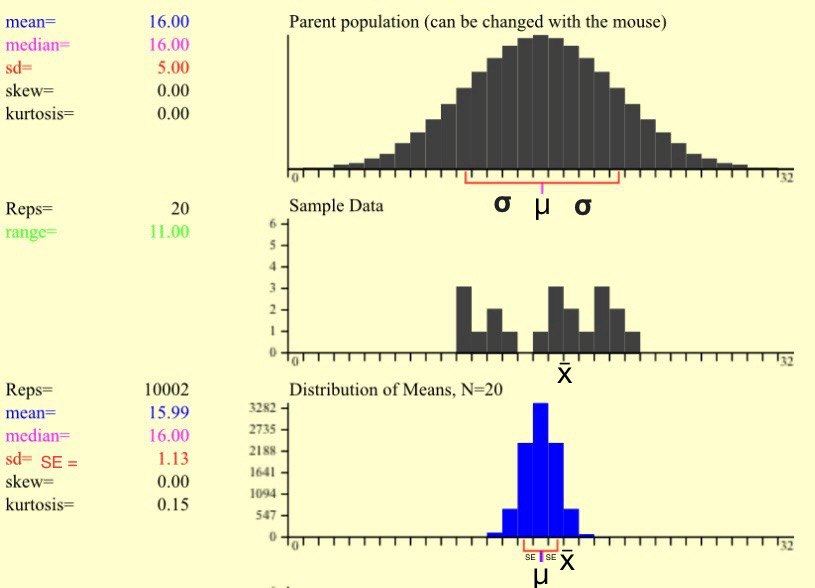
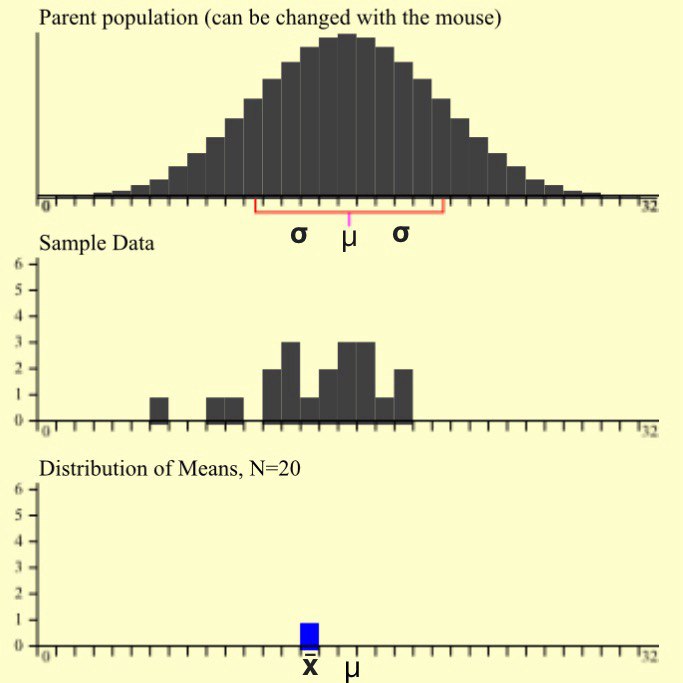
Population parameters:

* μ is the population mean
* σ is the population standard deviation

According to **CLT for normal distributions** for any sample size n:

1. The distribution of the sample mean will be normal.
2. The mean of the distribution of the sample mean will be equal to μ.
3. The standard deviation (standard error) of the distribution of the sample mean SE = .



10002 samples of size 20 are run and the results follow **(1), (2), (3)** rules

In hypothesis testing:

We first assume(H0 - null hypothesis) population mean to be equal to some value **μ**.

Then we get some sample of size **n**, calculate its mean(**x̄**) and variance(**s2**) and try to reject it with confidence interval c%.

zscore calculates distance between x̄(sample mean) and μ(the mean of the distribution of the sample mean, not population mean) taking SE as a unit distance.

zscore = =

**But why do we need zscore, why do we need distance between x̄ and μ?**

According to **CLT for normal distributions** the distribution of the sample mean follows **(1), (2), (3)** rules.

(2) The mean of the distribution of the sample mean is equal to μ.

(3) SE(Standard deviation of the distribution of the sample mean) = .

(1) The distribution of the sample mean is normal. Which lets us to know what percentage of the data is covered by [u-n\*SE, u+n\*SE] interval.

|  |  |
| --- | --- |
| n SE | Percentage of data covered within n SE of the mean |
| 1 SE | 68% |
| 2 SE | 95% |
| 3 SE | 99.7% |

It means that if H0 is true and we take a sample, sample mean(x̄) has 68% chance to be in [u-SE, u+SE] interval, 95% chance to be in [u-2\*SE, u+2\*SE] interval and etc.

**And if** x̄ **is out of [u-n\*SE, u+n\*SE] interval then we can reject the H0 with the confidence level that is equal to the percentage of data covered within that interval.**

For example, if x̄ is out of [u-2\*SE, u+2\*SE] interval then we can reject the H0 with the confidence level 95%, because 95% of the data that is covered by that interval.

**And to reject the H0 with the confidence level c%,** x̄ **has to be out of such interval that covers more than c% of data.**

For example, to reject H0 with confidence level 95%, x̄ has to be out of[u-2\*SE, u+2\*SE], in other words distance between x̄ and μ has to be more than 2\*SE, because [u-2\*SE, u+2\*SE] covers 95% of the data.

**The bigger the confidence level, the bigger the distance between x̄ and μ(zscore) has to be for us to be able to reject H0.**

When the population variance(σ2) is unknown, sample variance(s2) is used instead and zscore “becomes” tscore.

SE =

tscore = =

There are tables where for each confidence level their corresponding minimum z and t scores are written to reject the H0.

**An example problem**

At a water-bottling factory, a machine is supposed to put 2 liters of water into the bottles. After an overhaul, management thinks the machine is no longer putting the correct amount of water in. They sample 20 bottles and find an avg of 2.10 L of water with standard deviation of 0.33 L. Test the claim at 0.01 level of significance.

H0: μ = 2 n(sample size) = 20

Ha: μ != 2 x̄(sample mean) = 2.1

C = 0.99 s(sample standard deviation) = 0.9

t =  = = 0.4969

It’s obvious that with t<1 we can’t reject H0 with 99% confidence level. Because we know that only 68% of the data is covered within 1 SE of the μ in normal distributions. So if x̄ isn’t even out of that interval(x̄ is t\*SE distant from μ) then we can’t even reject H0 with 68 % confidence level.

If it’s not that obvious, then we can look at t-test table to see how much t at least has to be for us to be able to reject the H0. T-test table shows that t had to be not less than 2.845 for us to be able to reject H0 with 99% confidence level.

