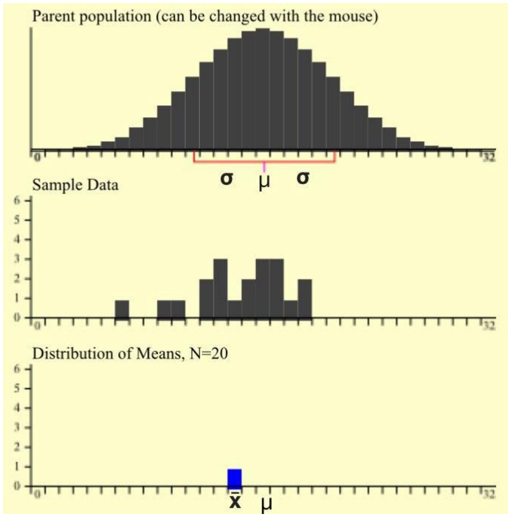
Population parameters:

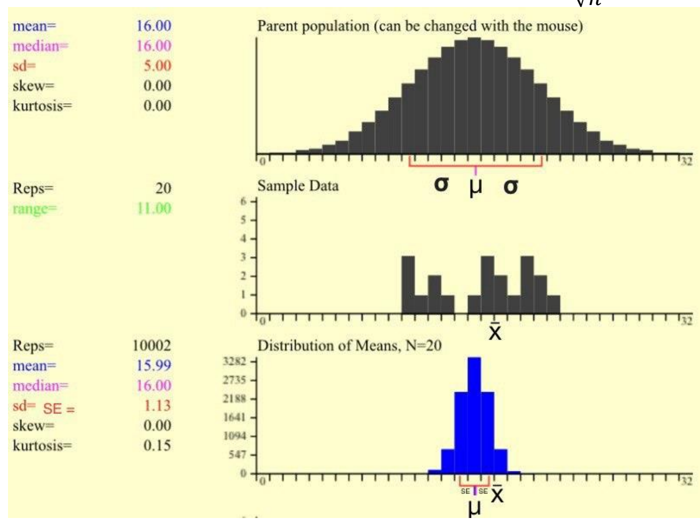
- μ is the population mean
- σ is the population standard deviation

According to **CLT** no matter **the underlying distribution of the population** as sample size n gets bigger:

(1) The distribution of the sample means tends to be normal.
(2) The mean of the distribution of the sample means will be equal to μ.
(3) The standard deviation (standard error) of the distribution of the sample means SE = $\frac{\sigma}{\sqrt{n}}$.



1 sample of size 20 is run



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

In hypothesis testing:

We first assume(H_0 - null hypothesis) population mean to be equal to some value **μ**.
Then we get some sample of size **n**, calculate its mean($\bar{x}$) and variance(s^2) and try to reject it with confidence level c or with significance level a = 1-c.

z_{score} calculates distance between $\bar{x}$(sample mean) and μ(the mean of the distribution of the sample means, not **population** mean) taking SE as a unit distance. It tells us where $\bar{x}$ is with respect to **μ**.

$$z_{score} = \frac{(\bar{x}-\mu)}{SE} = \frac{(\bar{x}-\mu)}{\frac{\sigma}{\sqrt{n}}}$$

But why do we need z_{score}, why do we need to know where $\bar{x}$ is with respect to μ?

According to **CLT** rule **(1).**

The distribution of the sample means is normal. **Normal distribution lets us to know what percentage of the data(sample means) is covered by [μ-n*SE, μ+n*SE] interval.**

z_{score}	Percentage of data covered by [μ-z*SE, μ+z*SE] interval	Minimum confidence level at which H_0 can't be rejected	P value(Maximum significance level at which H_0 can't be rejected) for 2-tailed example	P value for 1-tailed example
1	68%	68%, 0.68	0.32	0.16
2	95%	95%, 0.95	0.05	0.025
3	99.7%	99.7%, 0.997	0.003	0.0015

It means that if H_0 is true and we take a sample, then the probability of that sample mean($\bar{x}$) to be in [μ-SE, μ+SE] interval is 68%, to be in [μ-2*SE, μ+2*SE] interval is 95% and so on.

And if $\bar{x}$ is out of some interval, then we can reject the H_0 with the confidence level that is equal to the probability of $\bar{x}$ to be in that interval which is equal to the percentage of data(sample means) covered within that interval.

For example, if $\bar{x}$ is out of [μ-2*SE, μ+2*SE] interval then we can reject the H_0 with the confidence level 95%, because the probability of $\bar{x}$ to be in that interval is 95%, because 95% of the data(sample means) is covered by that interval.

And to reject the H_0 with the confidence level c%, $\bar{x}$ has to be out of such interval that covers more than c% of data.

For example, to reject H_0 with confidence level 95%, $\bar{x}$ has to be out of [μ-2*SE, μ+2*SE], in other words z_{score} has to be greater than 2 or less than -2, because [μ-2*SE, μ+2*SE] covers 95% of the data.

The bigger the confidence level, the bigger the distance between $\bar{x}$ and μ has to be(the more extreme $\bar{x}$ has to be) for us to be able to reject H_0.

When the population variance(σ^2) is unknown, sample variance(s^2) is used instead and z_{score} "becomes" t_{score}.

$$SE = \frac{s}{\sqrt{n}}$$

$$t_{score} = \frac{(\bar{x}-\mu)}{SE} = \frac{(\bar{x}-\mu)}{\frac{s}{\sqrt{n}}}$$

There are tables where for each confidence level their corresponding minimum z and t scores are shown to reject the H_0.

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.

H_0: $\mu = 2$

H_a: $\mu \neq 2$

a(significance level) = 0.01

c(confidence level) = 0.99

n(sample size) = 20

$\bar{x}$(sample mean) = 2.1

s(sample standard deviation) = 0.33

$$t_{score} = \frac{(x-\mu)}{SE} = \frac{(x-\mu)}{\frac{s}{\sqrt{n}}} = 0.4969$$

It's obvious that with t<1 we can't reject H_0 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 $\bar{x}$ isn't even out of that interval($\bar{x}$ is t_{score}*SE distant from μ) then we can't even reject H_0 with 68 % confidence level.

If it's not that obvious, then

1. 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 H_0. T-test table shows that t_{score} has to be not less than 2.861 for us to be able to reject H_0 with 99% confidence level. df = n-1=19

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	t
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.0
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.0
df										
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	
Confidence Level	0%	50%	60%	70%	80%	90%	95%	98%	99%	9

2. We can look at the p-value for two-tailed example corresponding to our t_{score} and compare it with a. If p > a, H_0 can't be rejected; otherwise, H_0 can be rejected.