SAMPLING DISTRIBUTION (SD)

In statistics, a sampling distribution is the probability distribution, under repeated sampling of the population, of a given statistic (a numerical quantity calculated from the data values in a sample).

The formula for the sampling distribution depends on the distribution of the population, the statistic being considered, and the sample size used.

A more precise formulation would speak of the distribution of the statistic as that for all possible samples of a given size, not just "under repeated sampling".

For example, consider a very large normal population (one that follows the so-called bell curve).

Assume we repeatedly take samples of a given size from the population and calculate the sample mean (the arithmetic mean of the data values) for each sample.

Different samples will lead to different sample means.

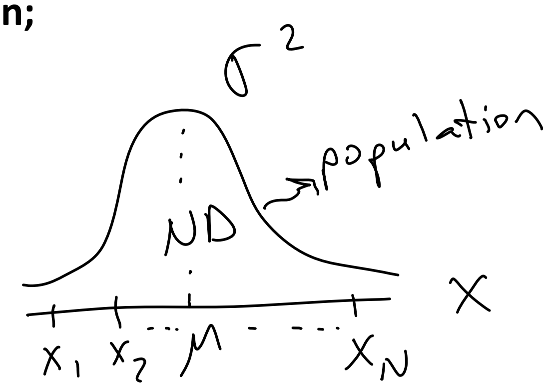
The distribution of these means is the "sampling distribution of the sample mean" (for the given sample size).

This distribution will be normal since the population is normal.

We are interested in the distribution of sample means because most of the common techniques are concerned with inferences about population means.

Consider all the scores (measurements) of a population;

Frequency (Y)	Scores (X)
2	5 → x_1
10	15 → x_2
⋮	⋮
40	60
10	70
⋮	⋮
1	100 → x_N

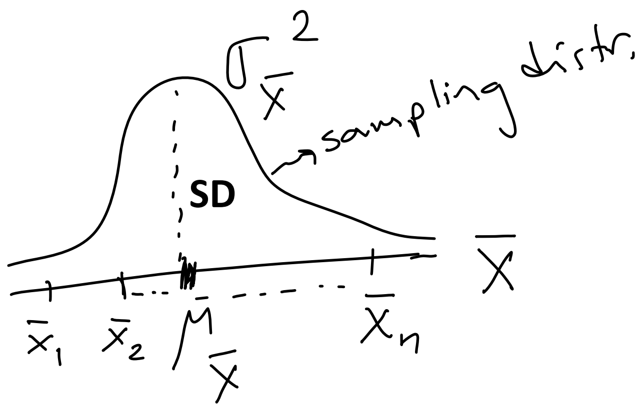


Suppose we are sampling from a population with mean μ and standard deviation σ.

Let $\overline{X}$ be a random variable representing the sample mean of n independently drawn observations from this population.

⊛ Now, draw several samples from this population and take the average of each ($\bar{x}$).

Sample	n	$\overline{X}$
1	10	$\bar{x}_1$
2	15	$\bar{x}_2$
3	10	$\bar{x}_3$
⋮	⋮	⋮
⋮	n	$\bar{x}_n$



1) As the sample size $\nearrow$ $\Rightarrow$ the better will be the normal approximation to the sampling distribution.

Then, the mean of population is equal the mean of the sampled population.

i.e., $\mu = \mu_{\bar{x}}$ $\rightsquigarrow$ mean of sampling distribution of means ($\bar{x}$).

mean of population which we are sampling

2) If we let $\sigma_{\bar{x}}$ denote the st. dev. of the sampling distribution of $\bar{x}$, then,

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, n = sample size

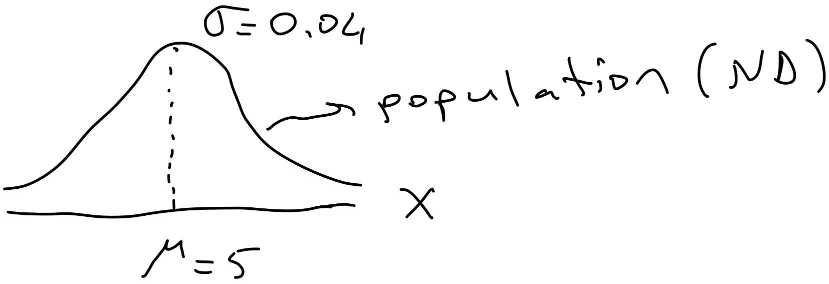
$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

⊛ The st. deviation of the sampling distribution is also called the standard error of the statistics.

i.e., $\boxed{\sigma_{\bar{x}} = \text{standard error of statistics}}$

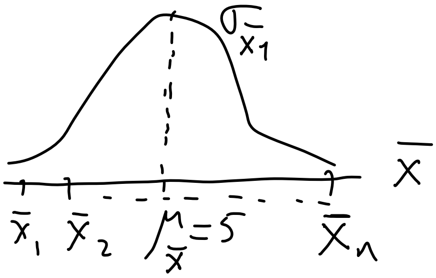
OR $\boxed{\sigma_{\bar{x}} = \text{standard error of the mean } (\bar{x})}$

e.g.;



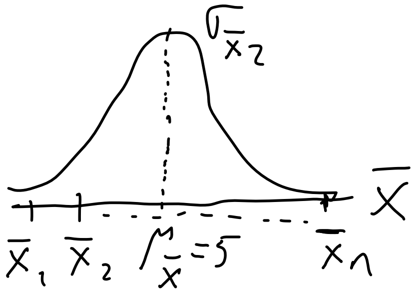
⊛ Let's take several samples from population (ND).

1) Several samples (each contains 10 data ⇒ $n = 10$)



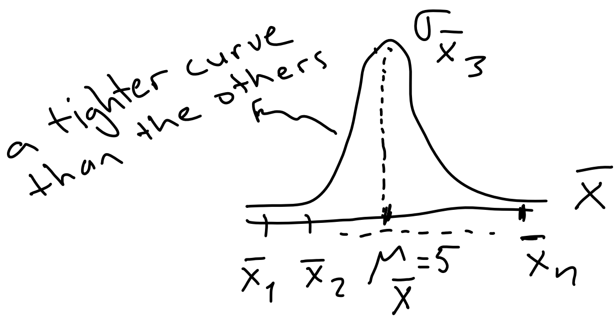
$$\sigma_{\bar{x}_1} = \frac{\sigma}{\sqrt{n}} = \frac{0.04}{\sqrt{10}} = 0.0126$$

2) Several samples (each contains 20 data ⇒ $n = 20$)



$$\sigma_{\bar{x}_2} = \frac{0.04}{\sqrt{20}} = 0.0089$$

3) Several samples (each contains 30 data ⇒ $n = 30$)



$$\sigma_{\bar{x}_3} = \frac{0.04}{\sqrt{30}} = 0.004$$

$$\Rightarrow \sigma > \sigma_{\bar{x}_1} > \sigma_{\bar{x}_2} > \sigma_{\bar{x}_3}$$

i.e., the larger n, the smaller the standard deviation of sampling distribution.

Theorem: If all possible random samples of size n are drawn with replacement from a finite population of size N, mean μ and Standard deviation σ, then,

Test statistics value : z_{calc}

$z_{calculated} = z_{statistics} =$ Test statistics,

$$z_{calc} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$ $\rightsquigarrow$ valid for the mean $(\bar{x})$ of n observations from any population when σ is known.

$$z_{calc} = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$ $\rightsquigarrow$ valid for the mean $(\bar{x})$ of n observations from any population when σ is unknown and $n \geq 30$.

$$z_{calc} = \frac{x - \mu}{\sigma}$$ $\rightsquigarrow$ valid for a single value (x) from a population.

Example: Given the population 1, 1, 1, 3, 4, 5, 6, 6, 6, 7. Find the P that a random sample of size 36, selected with replacement, will yield a sample mean greater than 3.8 but less than 4.5.

Solution:

$\mu = ?$, σ^2 or $\sigma = ?$ $\Rightarrow$ First, calculate them.

$$\Rightarrow \mu = \frac{1+1+1 \cdots + 7}{10} = 4 \quad \text{and} \quad \sigma^2 = \frac{\sum (x_i - \mu)^2}{N} = 5$$

$$\Rightarrow \sigma = 2.24$$

Assume $\mu_{\bar{X}} \cong \mu = 4$

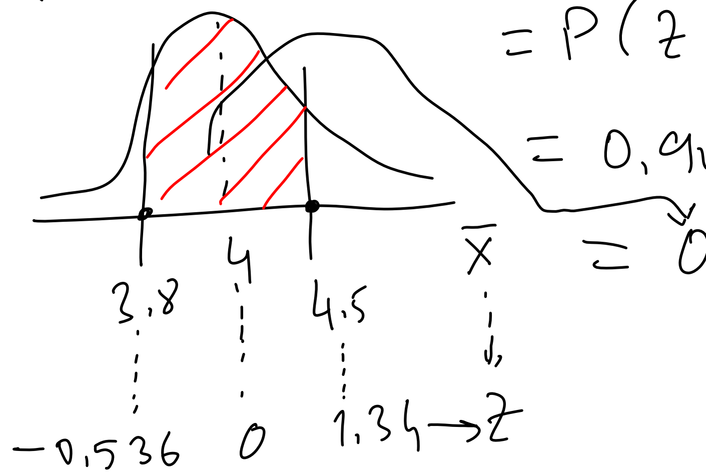
$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} = \frac{5}{36} \Rightarrow \sigma_{\bar{X}} = 0,373$$

The P of $\bar{X}$ that is between 3.8 and 4.5?

$\bar{X}_1 = 3.8$, $\bar{X}_2 = 4.5$

$$z_1 = \frac{3.8 - 4}{\underbrace{0,373}_{\frac{\sigma}{\sqrt{n}} = \sigma_{\bar{X}}}} = -0,536, \quad z_2 = \frac{4.5-4}{0,373} = 1,34$$

$$P(3.8 < \bar{X} < 4.5) \cong P(-0,536 < z < 1,34)$$
$$= P(z < 1,34) - P(z < -0,536)$$
$$= 0,9099 - 0.2963$$
$$= 0,6136 \text{ OR}$$



$\underline{\underline{61.36\%}}$

<u>Example</u>: The number of calories in a certain fast food restaurant is approximately normally distributed with a mean of 740 calories and a standard deviation of 20.

a) What is the probability of a randomly selected food has at least 760 calories ?

b) What is the probability of the mean of 9 randomly selected foods is at least 760 calories ?

Solution:

a) $x = 760$, $\mu = 740$, $\sigma = 20$

x is a single value drawn from population.

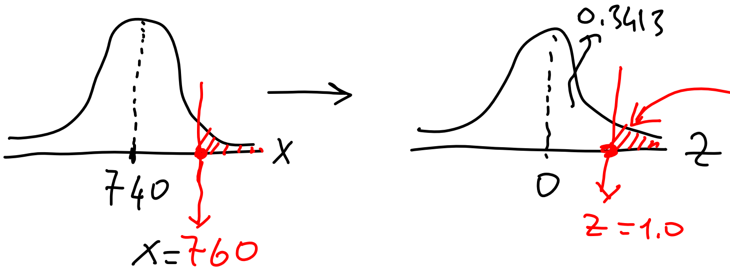
$$Z_{calc} = \frac{x-\mu}{\sigma} = \frac{760-740}{20} = 1.0$$

$$P(x \geq 760) \cong P(Z \geq 1)$$

$$= 0.5 - 0.3413 \quad \text{(Table 1)}$$

$$= 0.1587$$

OR 15.87 %



b) $\bar{x} = 760$, $n = 9$ (drawing sample from population).

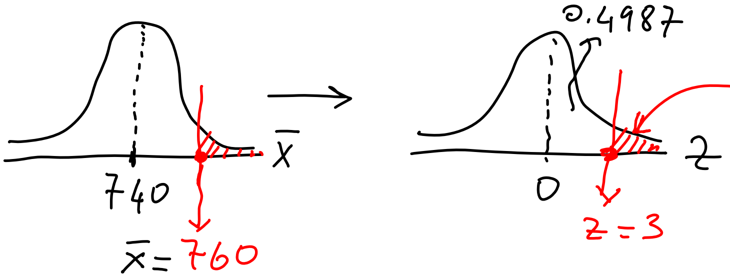
$$Z_{calc} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{760-740}{20/\sqrt{9}} = 3$$

$$P(\bar{x} \geq 760) \cong P(Z \geq 3)$$

$$= 0.5 - 0.4987 \quad \text{(Table 1)}$$

$$= 0.0013$$

OR 0.13 %



Summarizing this question ⇒

distribution of 9 foods

$\mu_{\bar{x}} = 740$, $\sigma_{\bar{x}} = \frac{20}{\sqrt{9}} = 6.66$