Statistical estimation

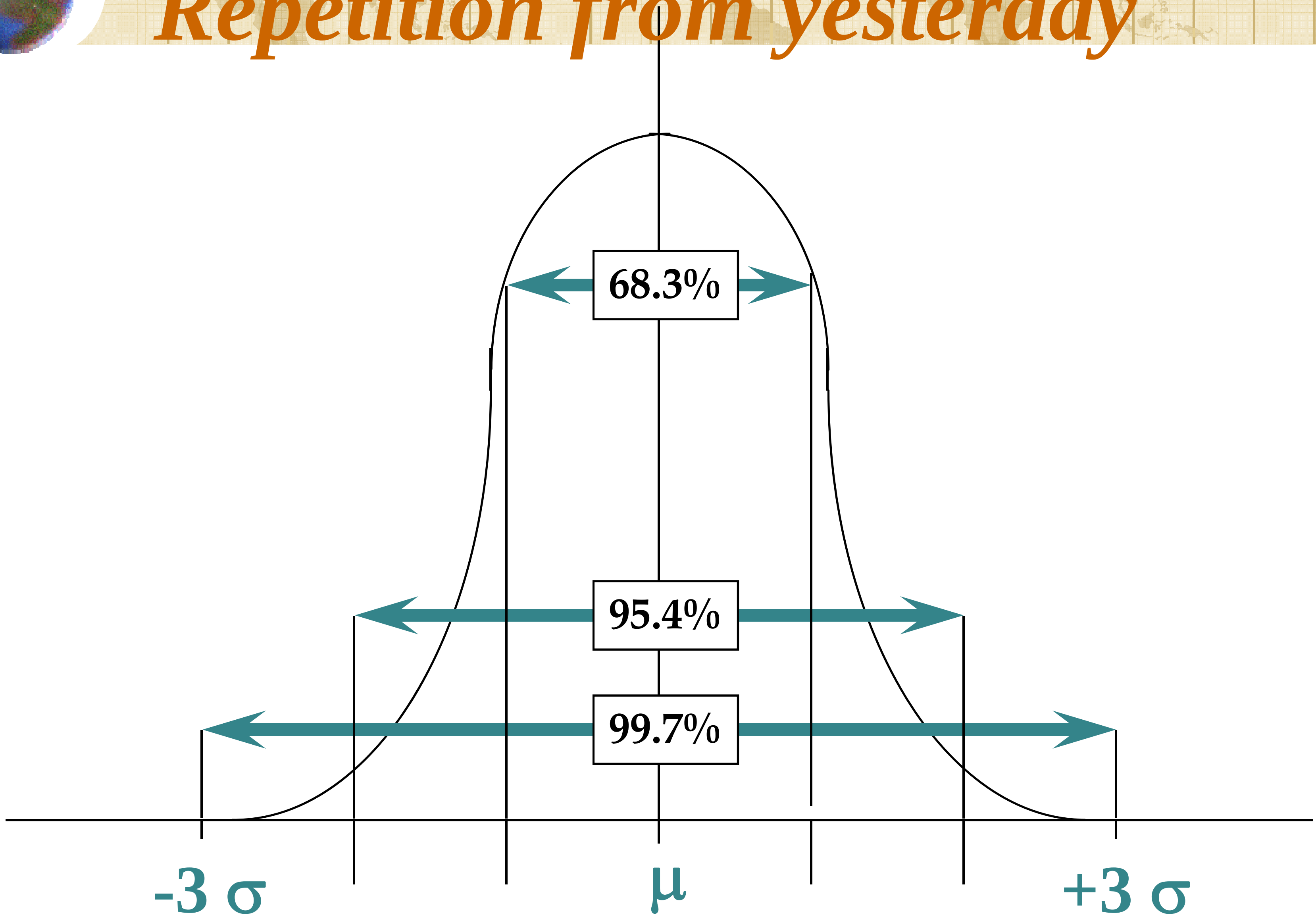
Vinjar Fønnebø

Repetition from yesterday

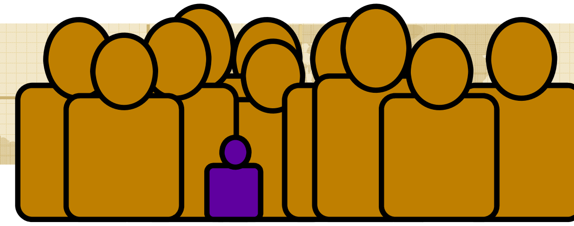
- Measure of centrality in a population: μ
- Measure of spread in a population:

$$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}}$$

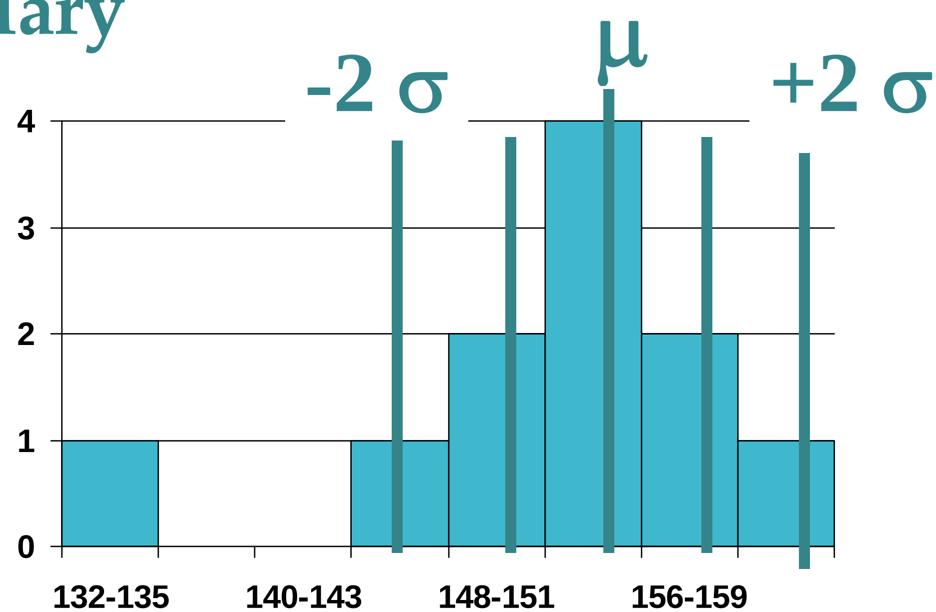
Repetition from yesterday



Repetition:



Histogram of the height of all Mary's peers and Mary



Most of the girls fall in the middle, a few further out. Mary is way out there: The statistics agree with our previous «eye-balling»

Definition

- Estimation
  - In medical statistics it means to draw conclusions about a population when you only have information from a sample

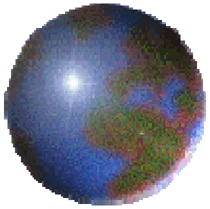
Examples of use of estimation

- MEDITERRANEAN DIET TRIAL:
  - Mediterranean diet was associated with a risk ratio for coronary events somewhere between 0.15 and 0.53
    - Circulation 1999
- MIGRAINE AND STROKE
  - Odds ratio for ischemic stroke in young women with migraine is somewhere between 1.3 and 9.6
    - BMJ 1999

Estimation

- In medicine we estimate on the basis of
  - Samples of one(individuals)
  - Samples of several

Estimation

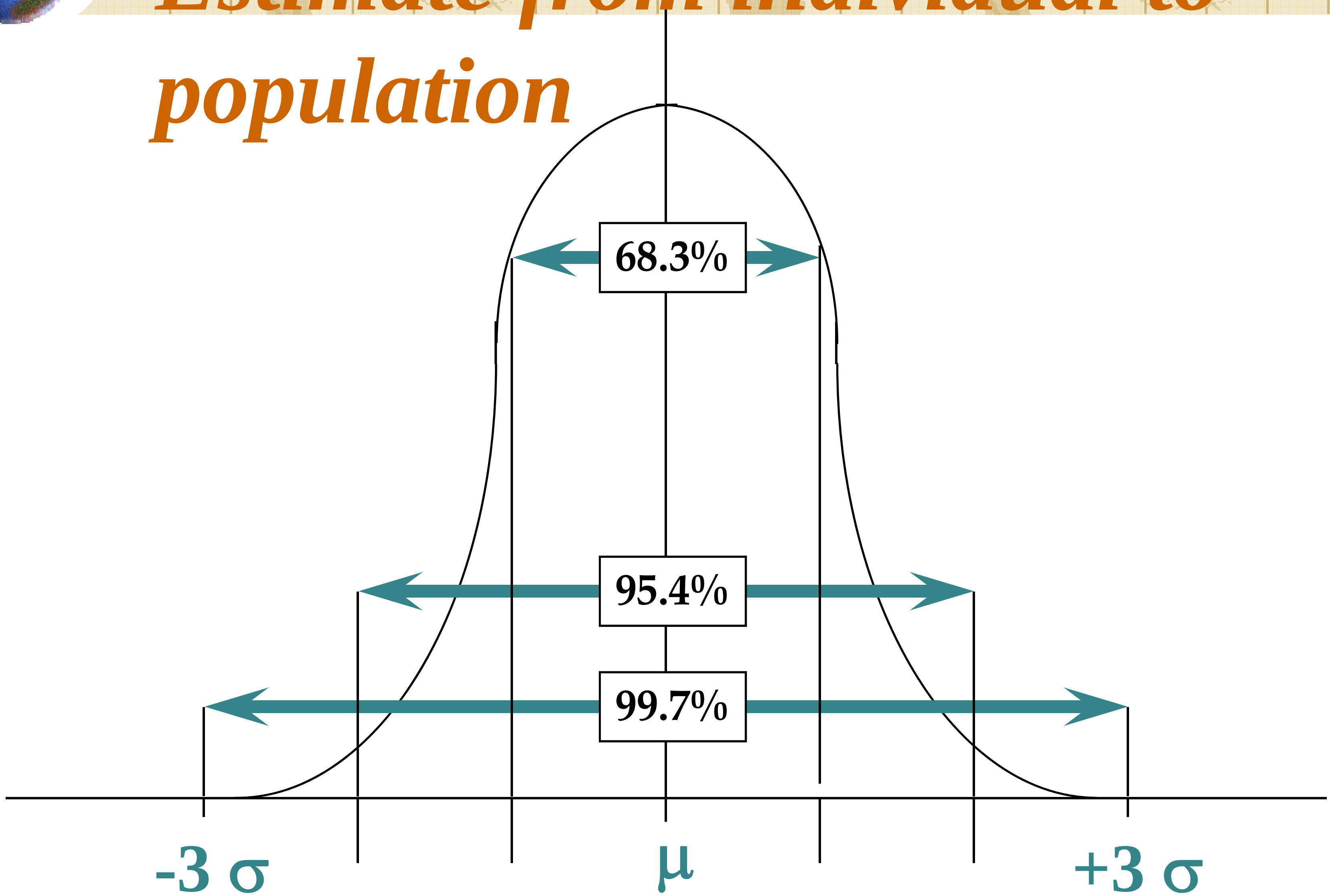


From a sample of one to a population

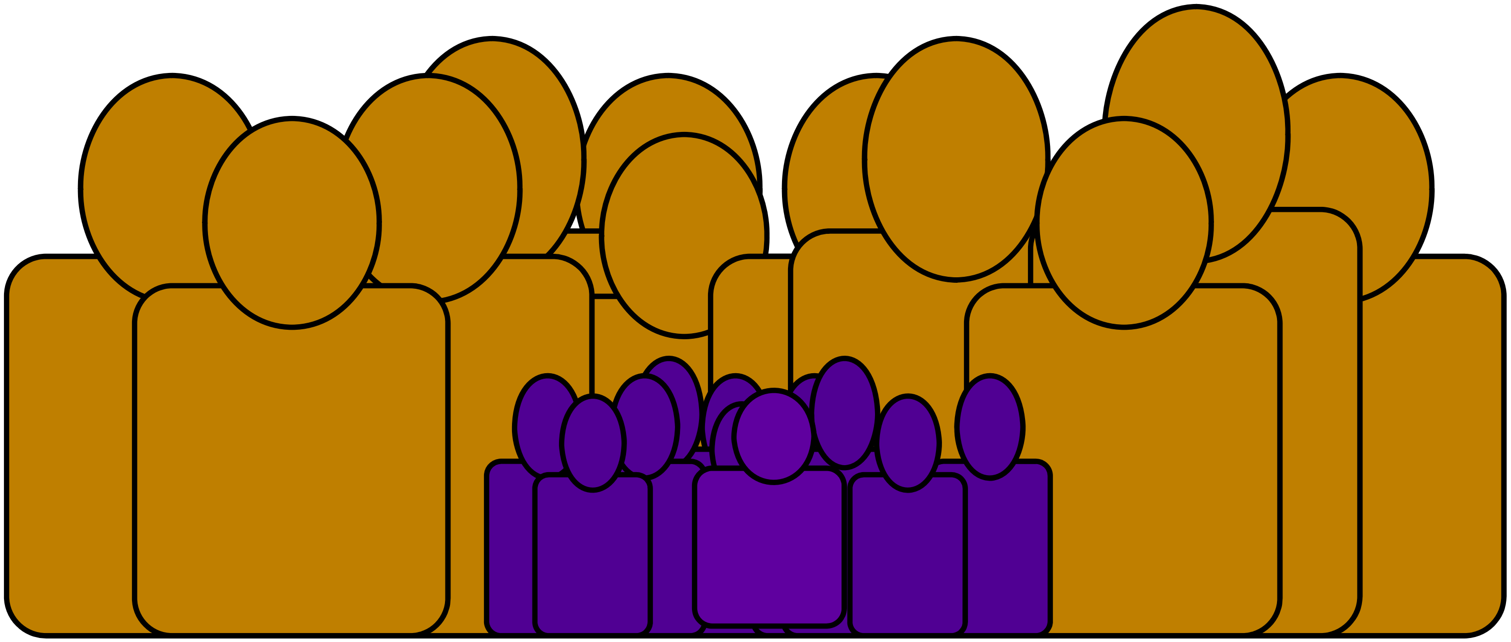
Estimate from individual to population

- We know:
  - 95% of individuals in a population have a value within ± 2 standard deviations(σ)from the mean(μ).
- That means:
  - An individual selected at random from the population will with 95% probability have a value not more than 2 standard deviations(σ)from the true population mean(μ).

Estimate from individual to population



Estimate from individual to population



Estimating from Mary to the population assuming she is randomly selected:

Estimate from individual to population

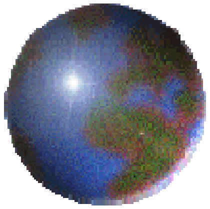
- Given a ***randomly*** selected individual
  - We can make a fairly good estimate about the mean(μ) of other persons belonging to the same statistical population

WE DO, HOWEVER, NEED TO KNOW THE VALUE OF THE STANDARD DEVIATION(σ)

Estimate from individual to population

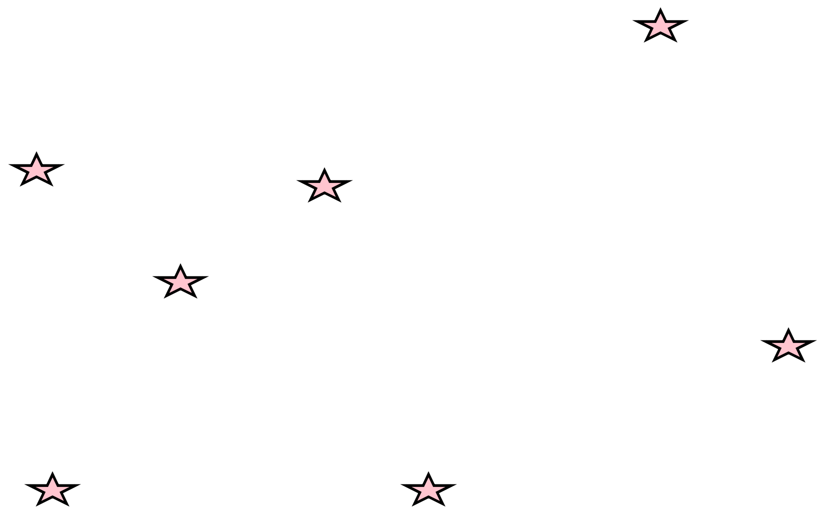
- Many prejudices in the world are based on conclusions drawn on the basis of a non-randomly selected individual and ignorance about σ
  - "All Norwegians are tall and blond"
  - "All elderly American women have blue hair"

Estimation

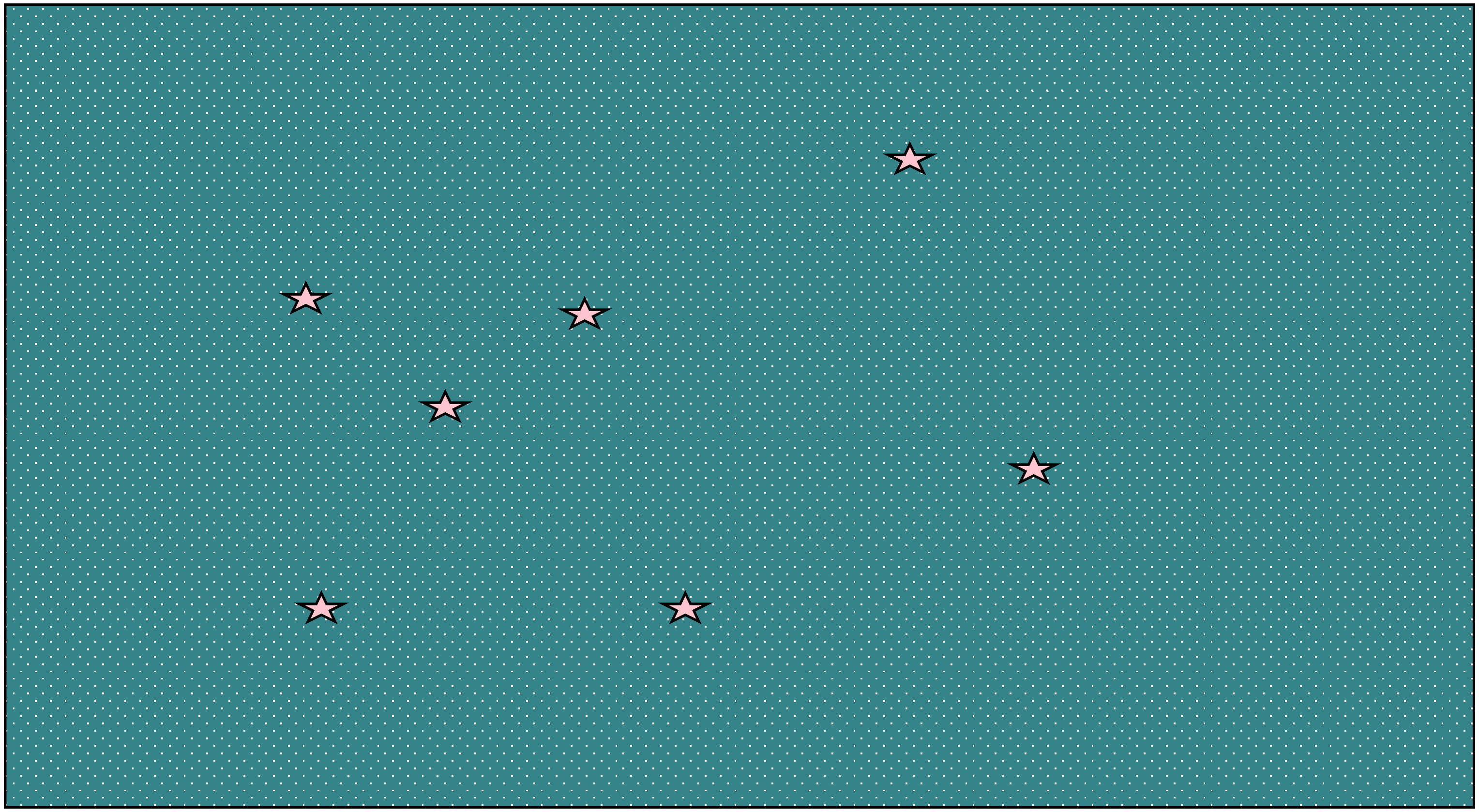


From a sample of several to a population

Estimate from a sample of several to population



Estimate from a sample of several to population



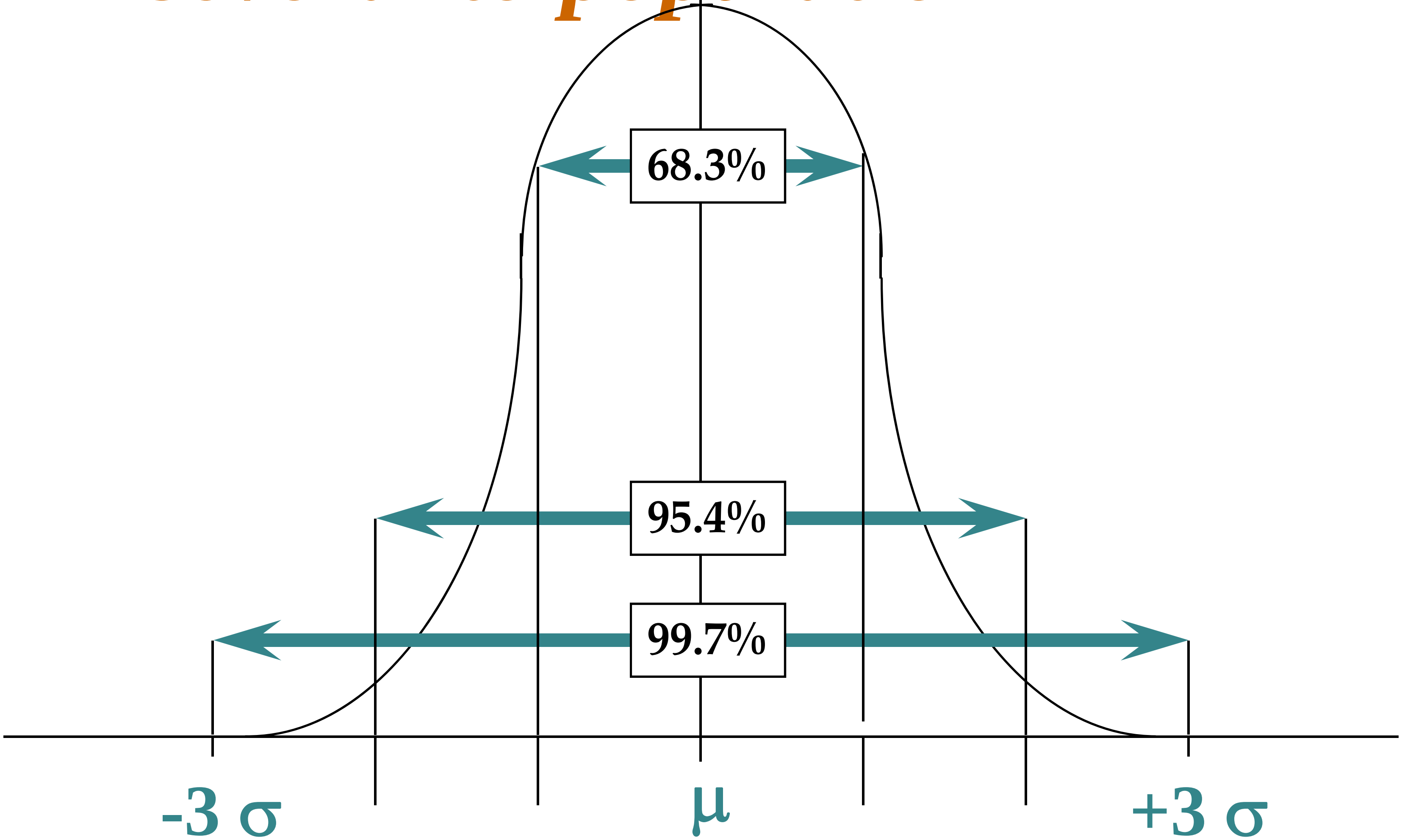
Estimate from a sample of several to population

- If you want to estimate anything about the population on the basis of a sample:
  - Your sample must be unbiased
  - The unbiased sample is a random sample where every individual in the population has the same chance of being included

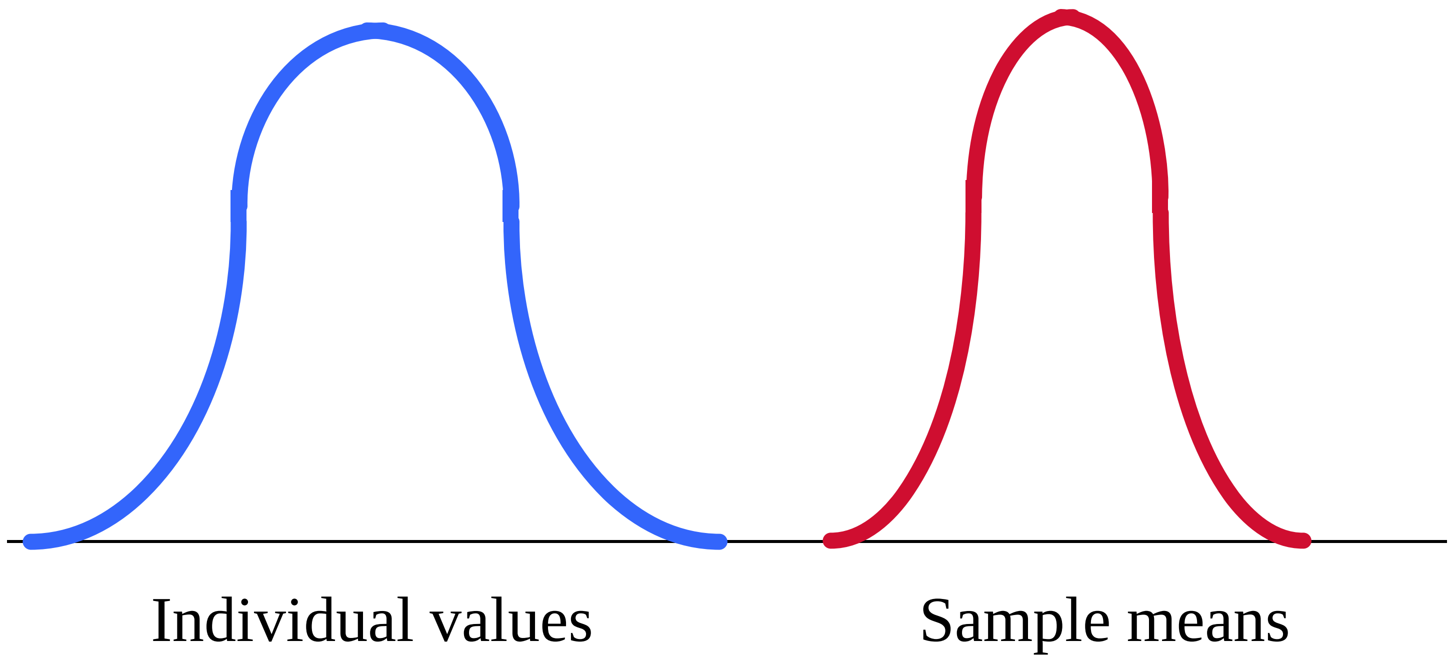
Estimate from a sample of several to population

- We know:
  - 95% of individuals in a population have a value within ± 2 σ from μ.
- Intuitively we can understand:
  - A sample selected at random from the population will have a mean value closer to the true population mean(μ) than an individual selected at random.

Estimate from a sample of several to population



Estimate from a sample of several to population



Estimate from a sample of several to population

- The variance of sample means drawn randomly from a population has been shown to be:

$$\frac{\sigma^2}{n}$$

n = number of individuals in the sample

Variance of individual values in the population

Estimate from a sample of several to population

- The variance of sample means:

$$\frac{\sigma^2}{n}$$

- and thus the standard deviation:

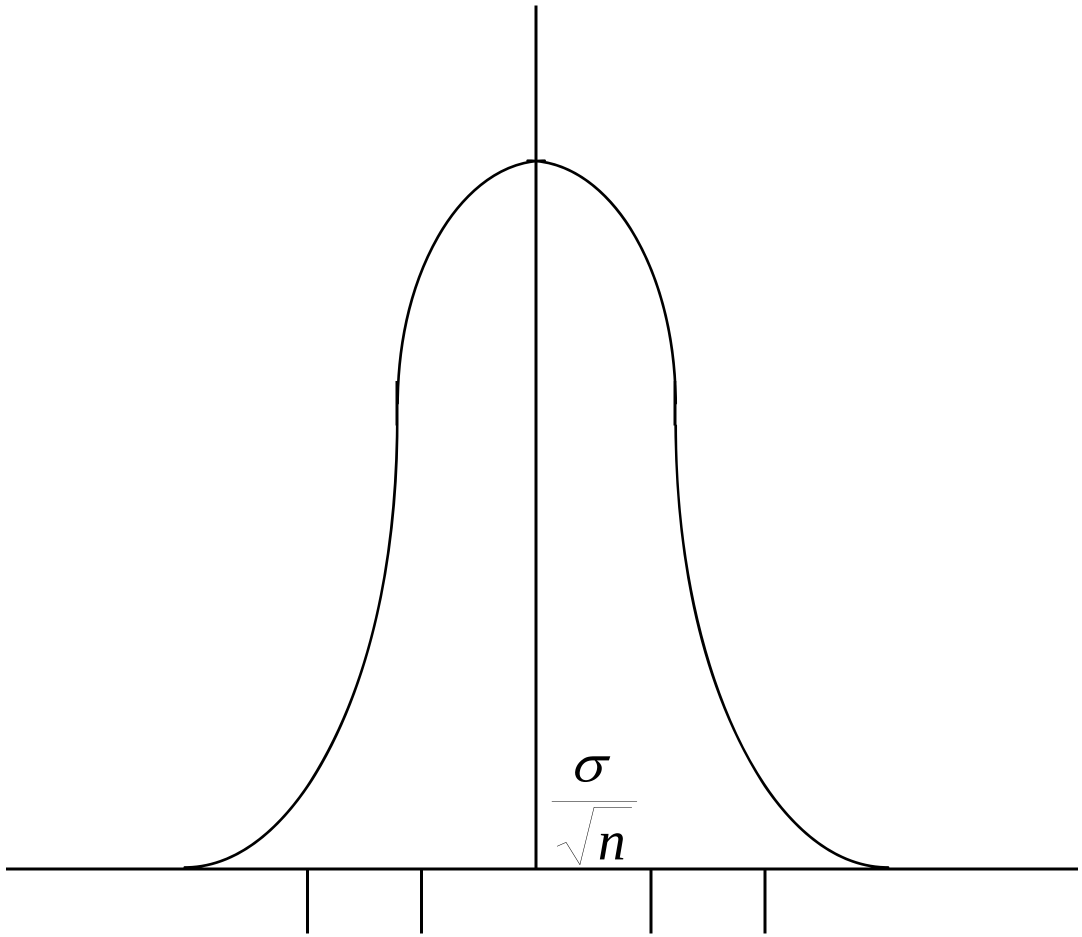
$$\sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

This is also called the standard error of the mean(SEM)

Estimate from a sample of several to population

- That sounds fine, we know
  - Our sample mean
  - How to work out the standard deviation of sample means (SEM)
- We can now:
  - First work out the numerical value of σ
  - Then work out where there is a 95% chance(probability) of finding the true mean(μ)

Estimate from a sample of several to population



Estimate from a sample of several to population

Let us start with working out σ: The formula for calculation of σ :

$$\sqrt{\frac{\sum (x-\mu)^2}{N}}$$

Estimate from a sample of several to population

Let us start with working out σ: The formula for calculation of σ :

$$\sqrt{\frac{\sum (x-\mu)^2}{N}}$$

BUT, WE ONLY HAVE A FEW VALUES IN A SAMPLE AND WE THEREFORE DON'T KNOW μ !!!

IT IS THEREFORE IMPOSSIBLE TO CALCULATE σ

Estimate from a sample of several to population

- We first have to make a best “guess” of μ
  - This guess can only be obtained by the mean value of our (hopefully) random sample of the population

μ can be ”guessed” by the sample mean $\bar{x}$

Estimate from a sample of several to population

$$\sqrt{\frac{\sum (x-\mu)^2}{N}}$$

What happens to the calculation of σ if we use $\bar{x}$ instead of μ?

Example:

		μ = 3,2	μ = 3,2	
x		$x-\mu$	$(x-\mu)^2$	
1		-2,2	4,84	
2		-1,2	1,44	
3		-0,2	0,04	
4		0,8	0,64	
5		1,8	3,24	
Sum		-1	10,2	

Estimate from a sample of several to population

$$\sqrt{\frac{\sum (x-\mu)^2}{N}}$$

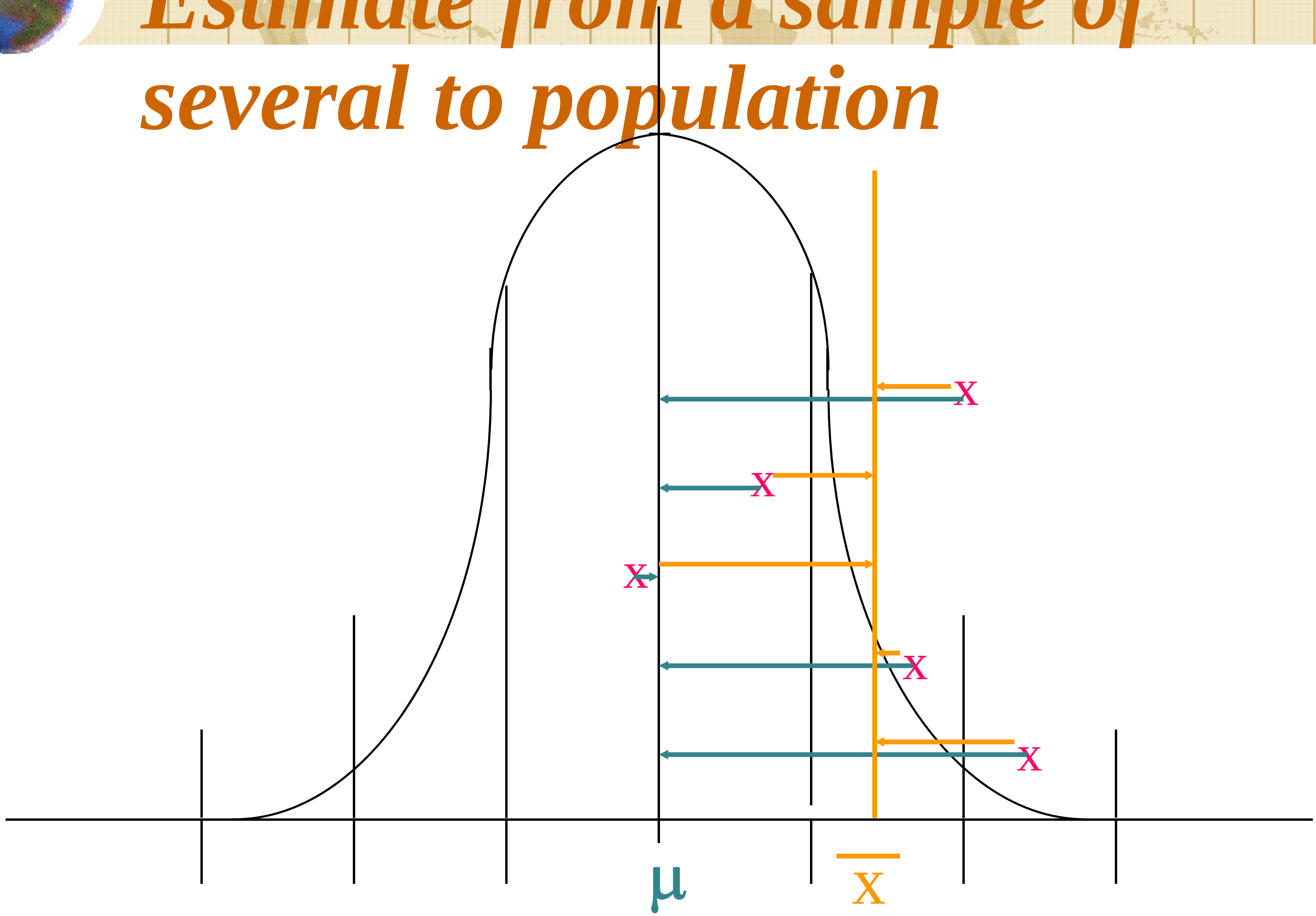
What happens to the calculation of σ if we use $\bar{x}$ instead of μ?

Example:

	$\bar{x} = 3$	$\bar{x} = 3$	$\mu = 3,2$	$\mu = 3,2$	
x	$x - \bar{x}$	$(x-\bar{x})^2$	$x - \mu$	$(x-\mu)^2$	
1	-2	4	-2,2	4,84	
2	-1	1	-1,2	1,44	
3	0	0	-0,2	0,04	
4	1	1	0,8	0,64	
5	2	4	1,8	3,24	
Sum	0	10	-1	10,2	

The numerator expression $\sum (x-\bar{x})^2$ **decreases!!**

Estimate from a sample of several to population



Estimate from a sample of several to population

- The estimate we make of σ will therefore be too small
- When the numerator decreases we have to adjust the denominator down as well:

Best guess of σ

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

The standard deviation is called s when it is an estimate calculated from a sample

Estimate from a sample of several to population

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

Best guess of σ

A large n gives a more certain estimate of μ which gives a smaller adjustment

A small n gives an umcertain estimate of μ which gives a larger adjustment

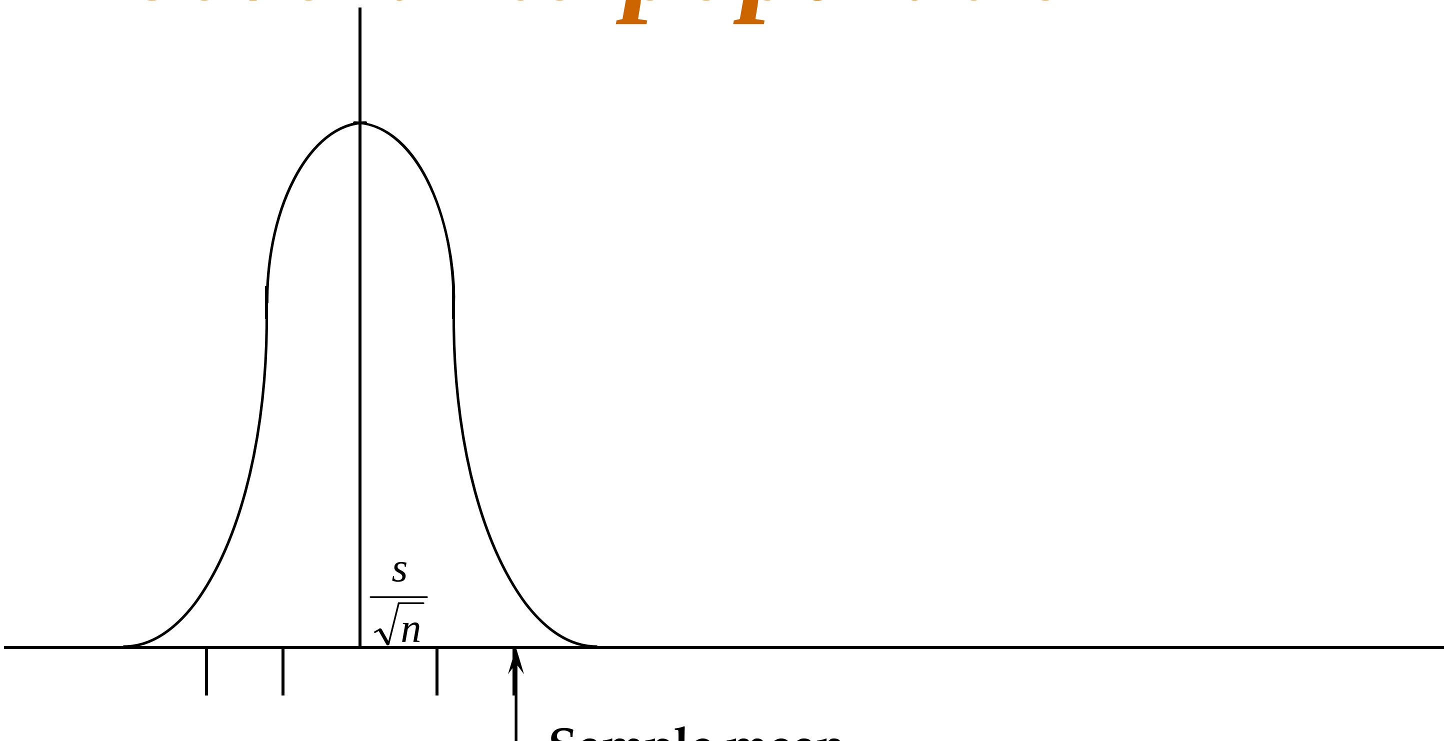
Estimate from a sample of several to population

- What do we now know?
  - The sample mean
  - The sample's estimate of σ
- Since we know that the standard deviation of sample means (SEM) is:

$$\frac{\sigma}{\sqrt{n}}$$

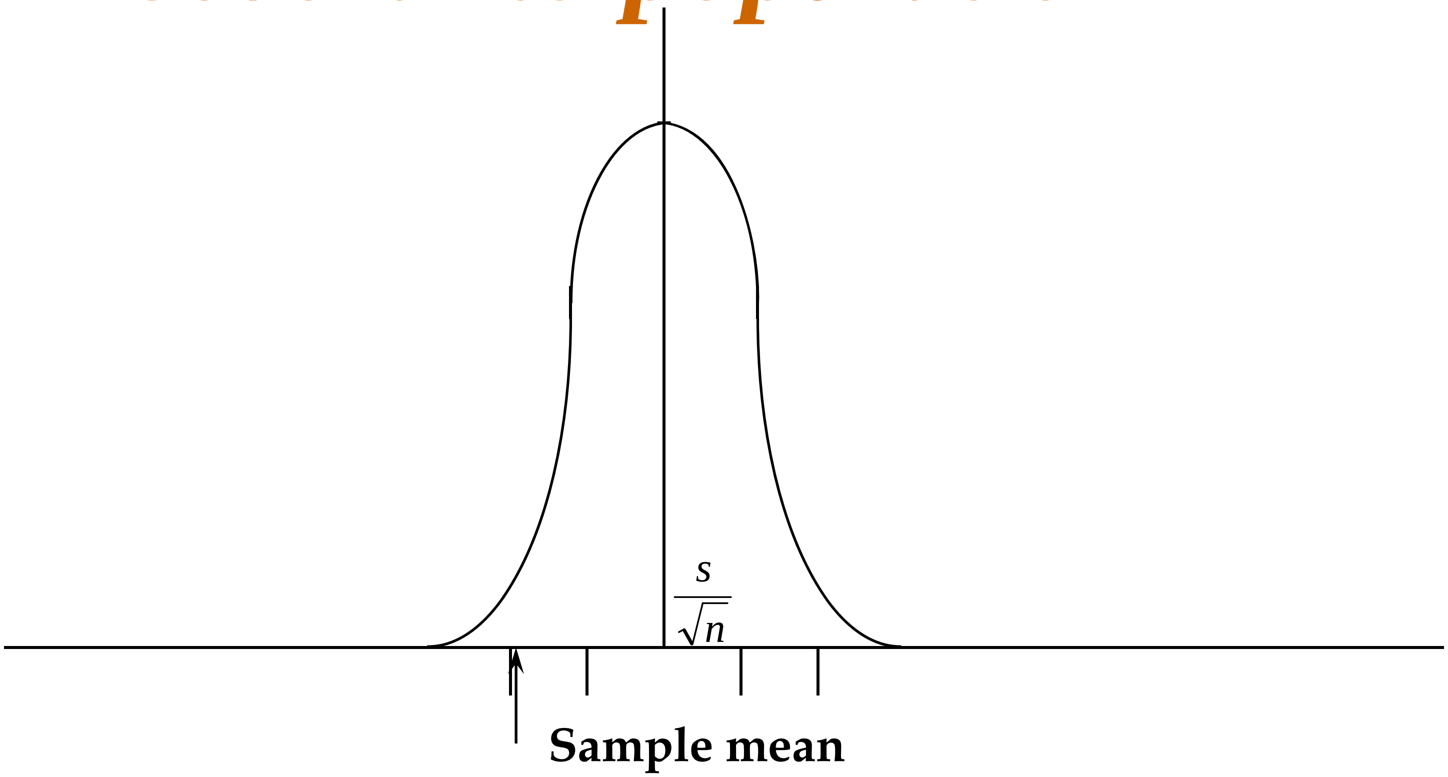
- we can substitute s for σ and calculate the standard deviation of sample means (SEM)

Estimate from a sample of several to population



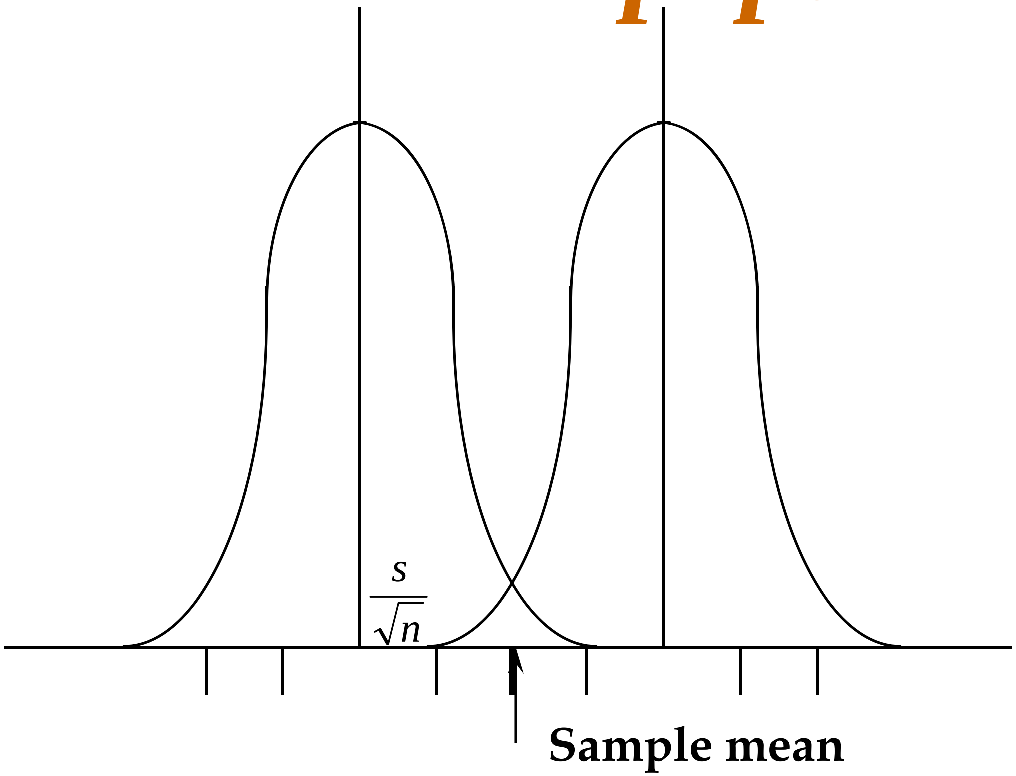
- Our sample mean could be exactly 2 SEM above μ

Estimate from a sample of several to population



- Our sample mean could be exactly 2 SEM below μ

Estimate from a sample of several to population

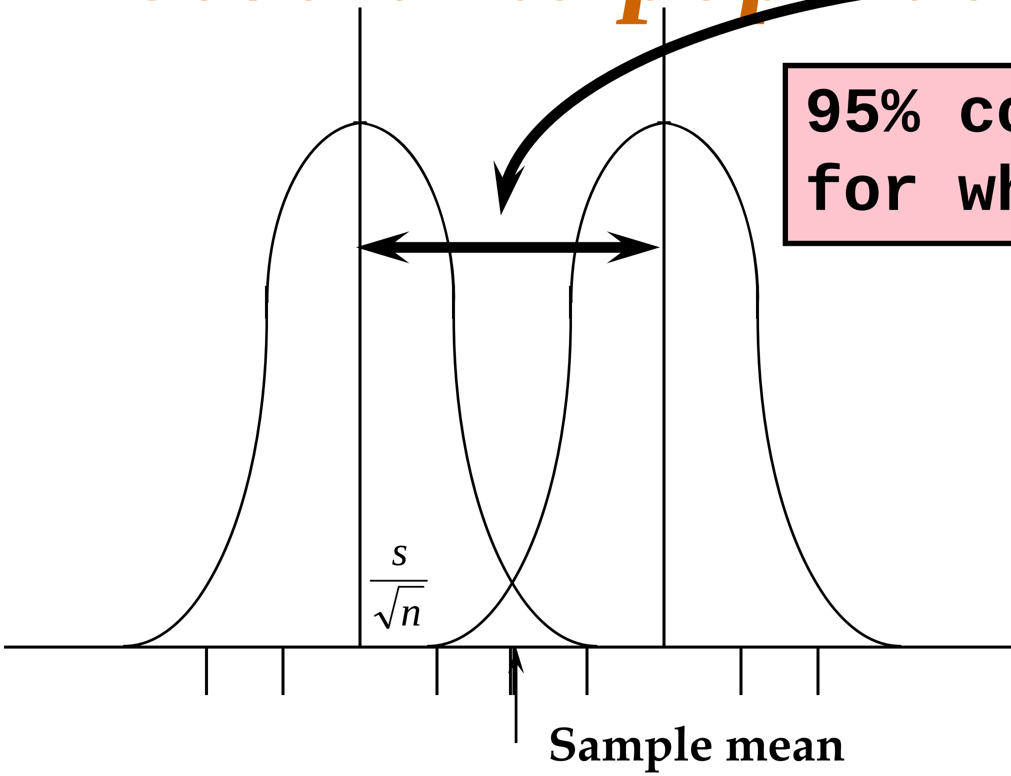


- The true μ must therefore with 95% probability lie within 2 SEMs above or below our sample mean

We call the distance between -2 SEM and +2 SEM from our sample mean our

95% confidence interval

Estimate from a sample of several to population



95% confidence interval for where μ can be found

The exact probability of μ being equal to the sample mean is close to 0