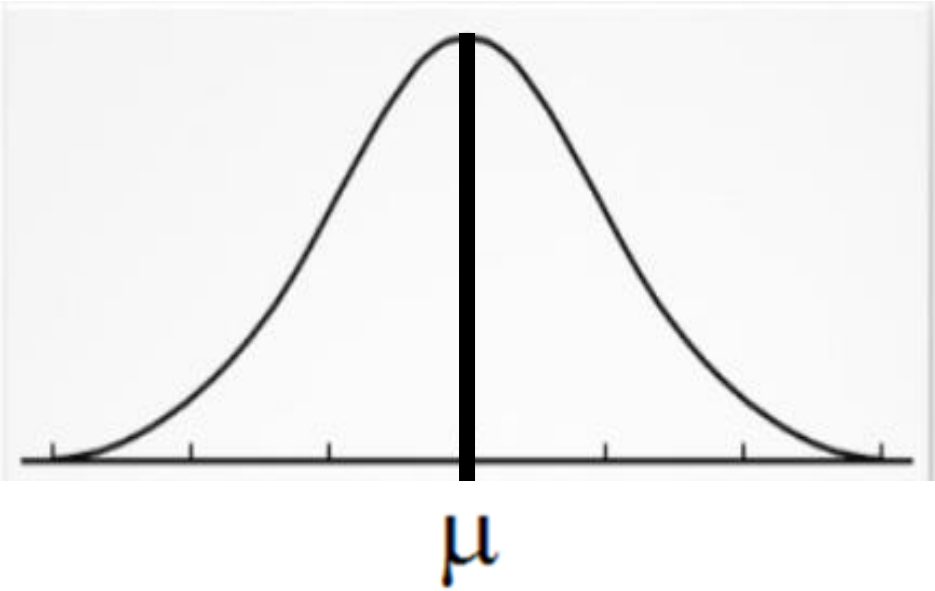
Normal distribution

We can check if a series is normally distributed

Rule: A serie of numbers is normal distributed if

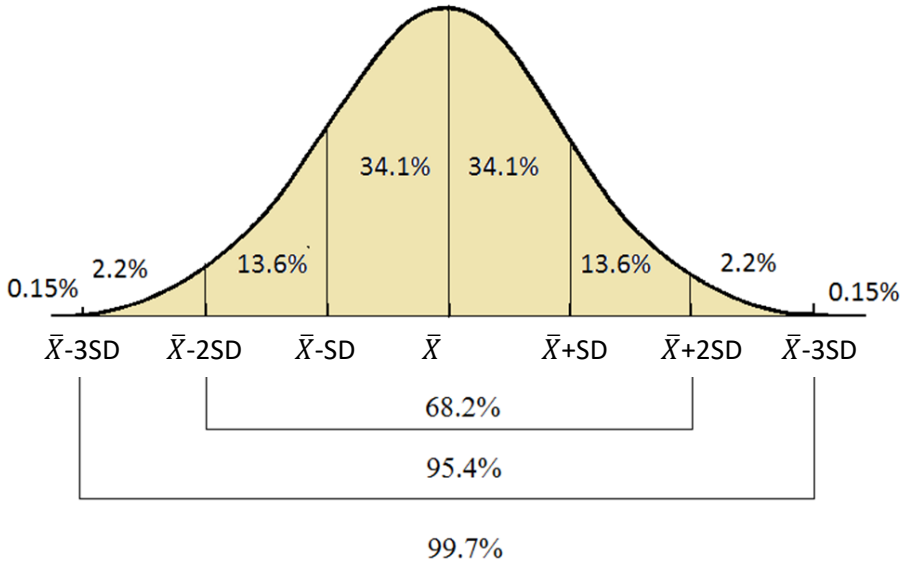
- ✓ Arithmetic mean = Median = Mode (or near equal)
- ✓ Quartile 1, Quartile 3 are simetrical with the mean (or near simetrical)
- ✓ Skewness ≈ 0 (between -1 to 1)
- ✓ Kurtosis ≈ 0 (between -1 to 1)



Other properties of normal distribution

- ✓ In the interval: mean ± st.dev. there are minimum 68.2% of data
- ✓ In the interval: mean ±2* st.dev. there are minimum 95.4% of data
- ✓ In the interval: mean ±3* st.dev. there are minimum 99.7% of data

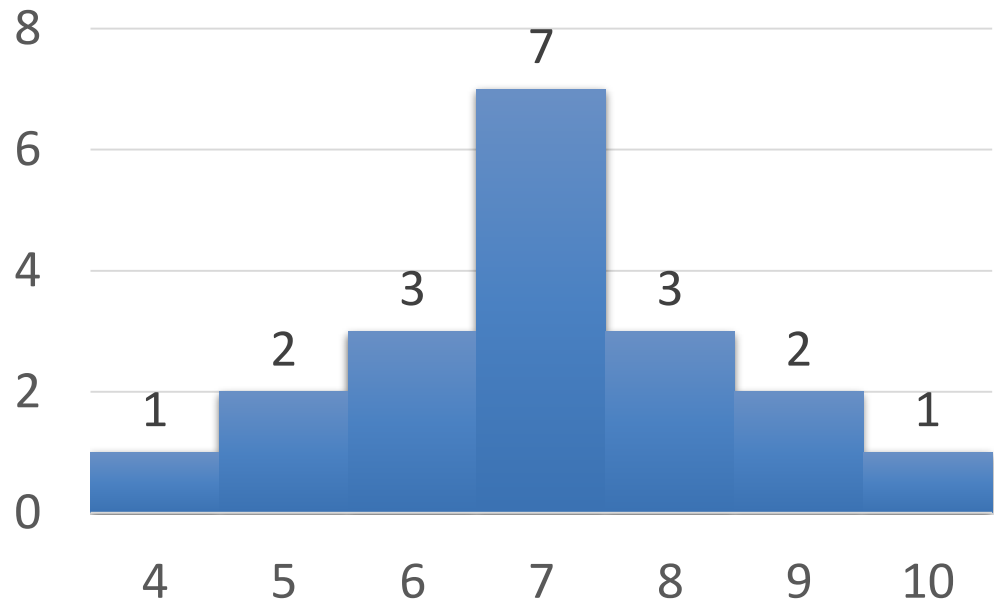
Where
st.dev. – standard deviation
mean – arithmetic mean
mean ± st.dev. – the interval between mean-st.dev and mean+st.dev.
μ- the population arithmetic mean
σ – the standard deviation



Example 1

Marks
4
5
5
6
6
6
7
7
7
7
7
7
7
8
8
8
9
9
10

Marks	Frequency
4	1
5	2
6	3
7	7
8	3
9	2
10	1



Mean	7
Median	7
Mode	7
Standard Deviation	1.49
Kurtosis	0.07
Skewness	0.00
Range	6
Minimum	4
Quartile 1	6
Quartile 3	8
Maximum	10
Count	19

normal distribution

	Marks
	4
	5
	5
1	6
2	6
3	6
4	7
5	7
6	7
7	7
8	7
9	7
10	7
11	8
12	8
13	8
	9
	9
	10

Mean	**7**
Standard Deviation (SD)	**1.49**

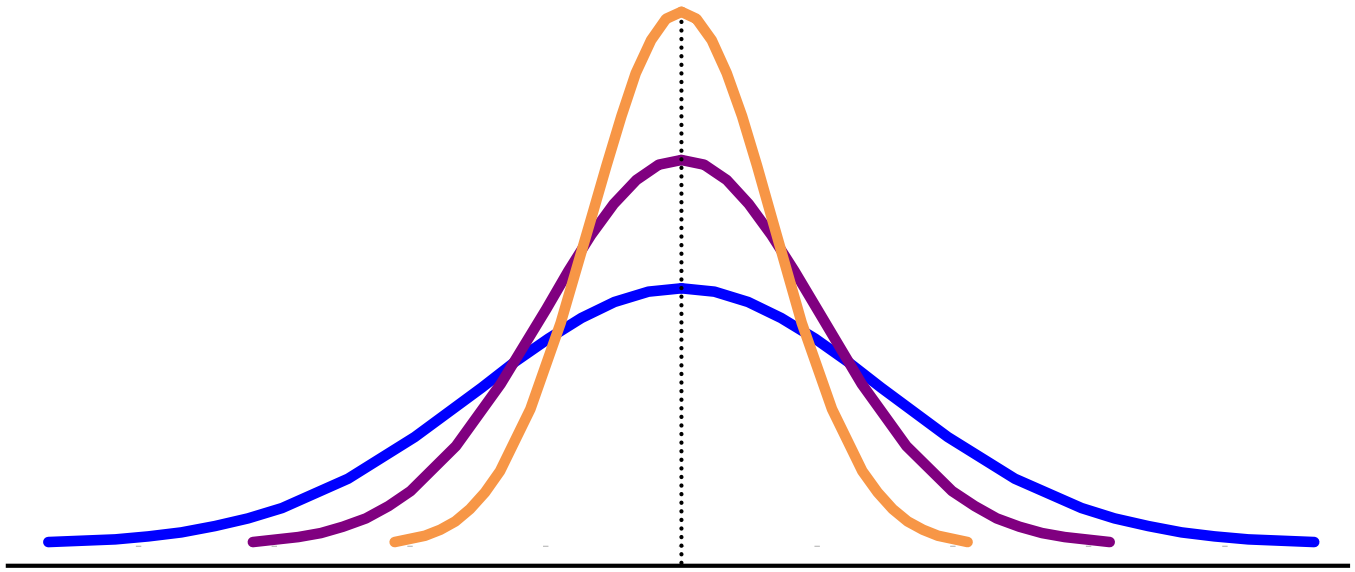
Mean-SD	7-1.49	5.51
Mean+SD	7+1.49	8.49

		Interval	No. of people	Procente
(Mean-SD; Mean+SD)	(7-1.49;7+1.49)	(5.51; 8.49)	13	68.4

in interval (Mean-SD; Mean+SD) there are 68.4% of values

If in the interval: mean ± st.dev. there are minimum **68.2%** of data
yes there are, we get normal distribution

normal distribution



Author: PhD. MsC. Bondor Cosmina-Ioana

Confidence intervals

A ALWAYS

S SEEK

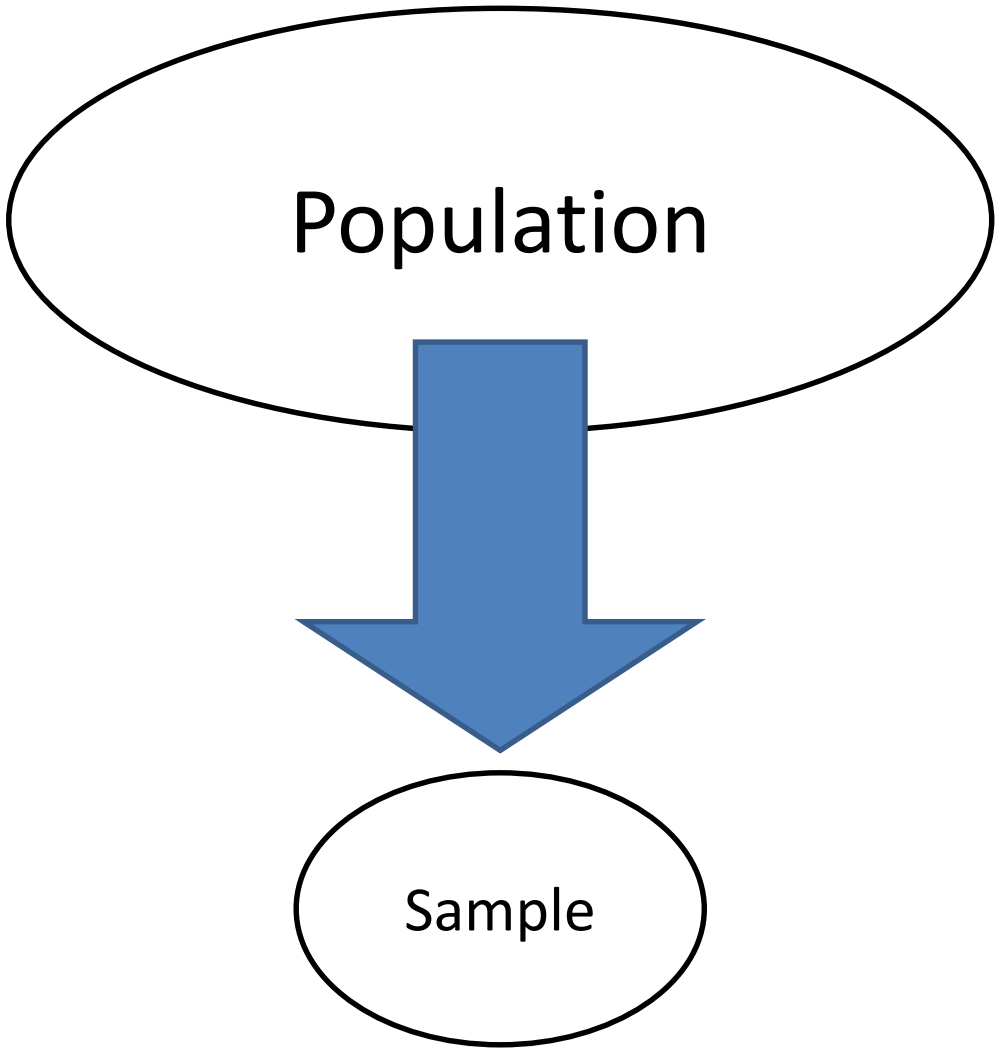
K KNOWLEDGE

Objectives

- Estimation
  - Point estimation
  - Confidence intervals estimation
- Sample size

Commonly used symbols in inference and statistics

Characteristic	**Population** Parameters Symbol	**Sample** Parameters Symbol
Mean	μ	$\bar{X}$
Standard deviation	σ	*SD (s)*
Proportion	φ	f



Unknown μ (inaccessible)

statistical inference = we want to estimate

We are in this point

Known $\bar{X}$ (accesibil)

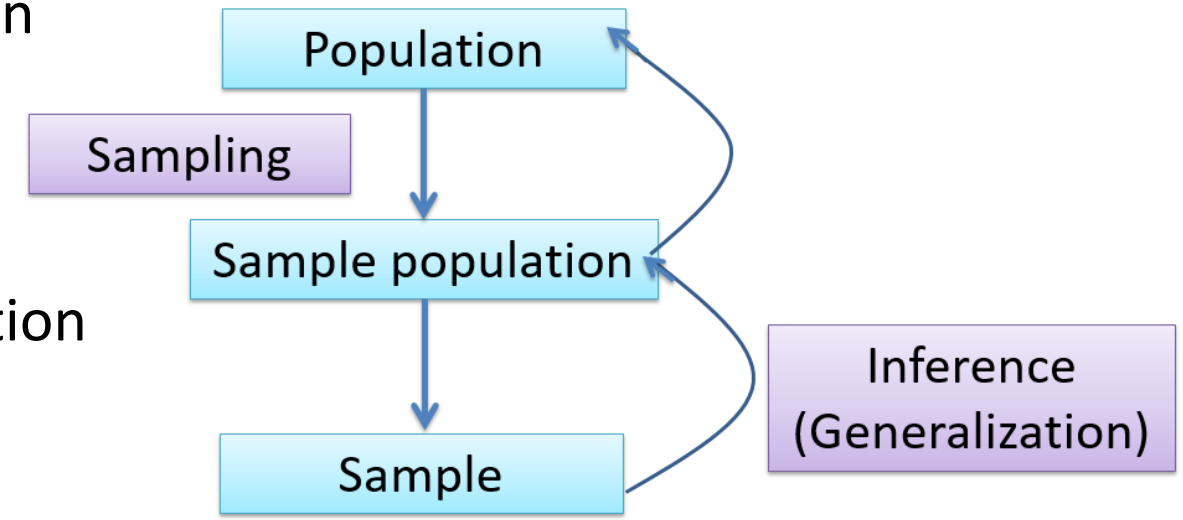
- **point estimation** - we can estimate the arithmetic mean μ of the population based on the known sample $\bar{X}$
- **confidence interval estimation** - we can construct an interval that in 95% of all possible studied samples, the population arithmetic mean μ will lie within this interval.

General principles for statistical inferences

population P

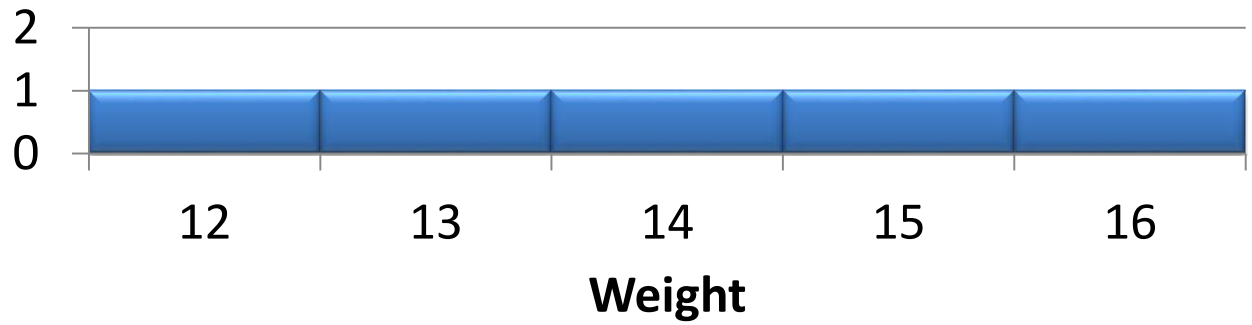
– a characteristic (quantitative or qualitative):

1. We select a random sample from the population
2. By means of descriptive statistics calculate:
   - Qualitative: observed frequency
   - Quantitative: mean and standard deviation
3. Results observed on the sample
4. By means of inferential statistics
   - results are extended to the entire population



Population - 5 boys two years old

- their weights: 12, 13, 14, 15, 16 kg
- This is the population



all possible samples of n subjects

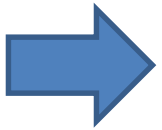
- Population 5 boys

- How many samples of 2 boys can be taken?

11	21	31	41	51
12	22	32	42	52
13	23	33	43	53
14	24	34	44	45
15	25	35	45	55

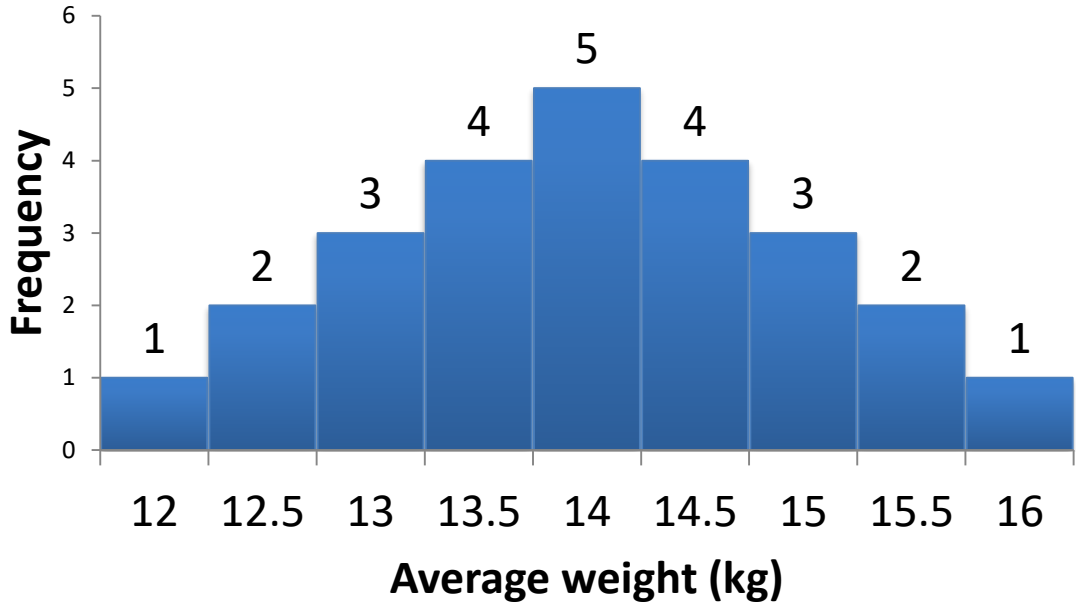
All possibilities of samples of two boys from 5 = 25

Sample no.	First boy	Second boy	Weight for the first boy	Weight for the second boy	Arithmetic Mean for the weights (kg)
1	1	1	12	12	12
2	1	2	12	13	12.5
3	1	3	12	14	13
4	1	4	12	15	13.5
5	1	5	12	16	14
6	2	1	13	12	12.5
7	2	2	13	13	13
8	2	3	13	14	13.5
9	2	4	13	15	14
10	2	5	13	16	14.5
11	3	1	14	12	13
12	3	2	14	13	13.5
13	3	3	14	14	14
14	3	4	14	15	14.5
15	3	5	14	16	15
16	4	1	15	12	13.5
17	4	2	15	13	14
18	4	3	15	14	14.5
19	4	4	15	15	15
20	4	5	15	16	15.5
21	5	1	16	12	14
22	5	2	16	13	14.5
23	5	3	16	14	15
24	5	4	16	15	15.5
25	5	5	16	16	16



Arithmetic Mean
12
12.5
12.5
13
13
13
13.5
13.5
13.5
13.5
14
14
14
14
14
14.5
14.5
14.5
14.5
15
15
15
15.5
15.5
16

	Frequency
12	1
12.5	2
13	3
13.5	4
14	5
14.5	4
15	3
15.5	2
16	1

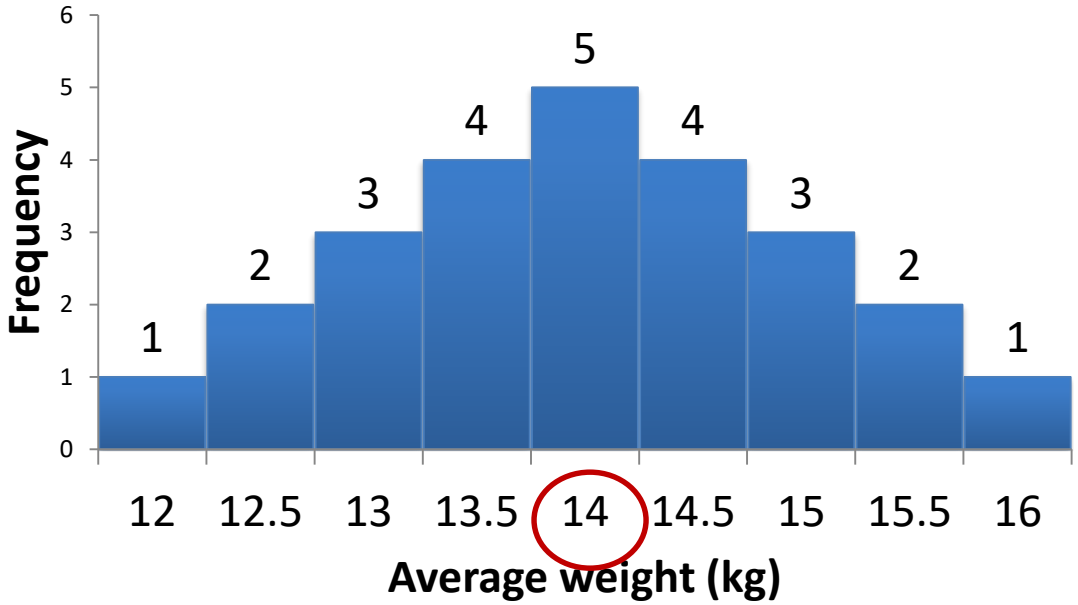


this is the **sampling distribution of the means**

(for all possible samples of 2 boys)

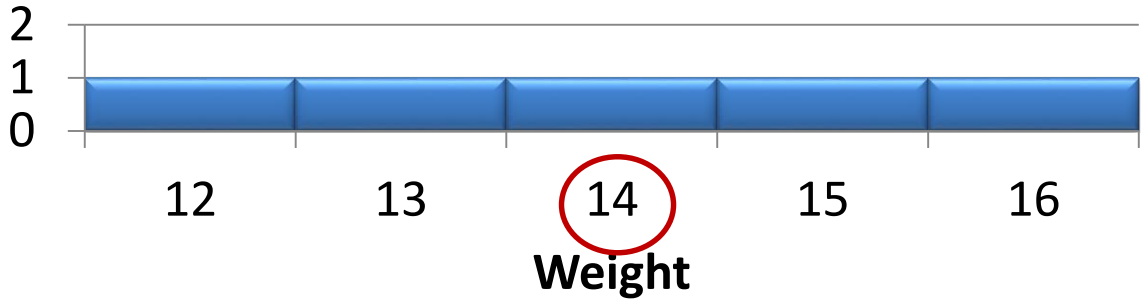
These are the means from the last two slides

Arithmetic Mean
12
12.5
12.5
13
13
13
13.5
13.5
13.5
13.5
14
14
14
14
14
14.5
14.5
14.5
14.5
15
15
15
15.5
15.5
16



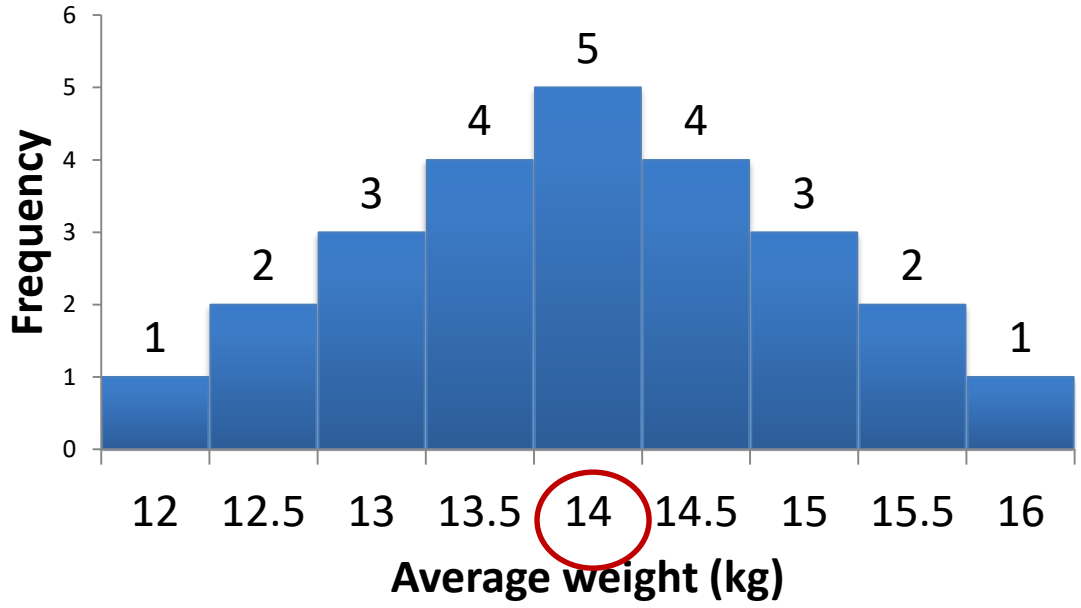
Mean (of the means) of the sampling distribution = 14

Arithmetic Mean
12
12.5
12.5
13
13
13
13.5
13.5
13.5
13.5
14
14
14
14
14
14.5
14.5
14.5
14.5
15
15
15
15.5
15.5
16



Population (all 5 boys) distribution

Mean of the population = 14

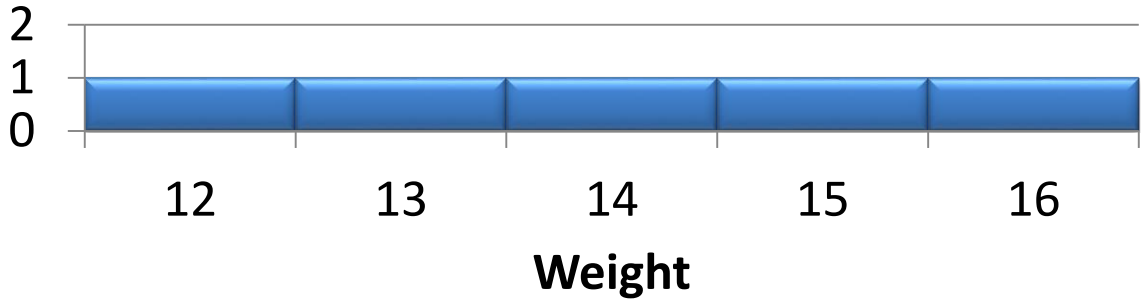


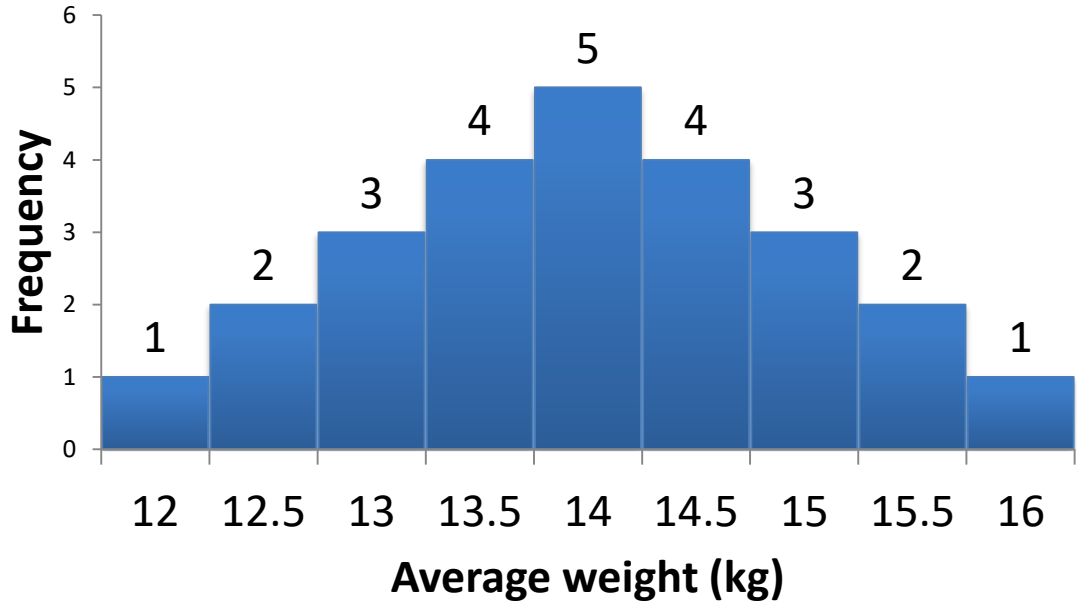
Sampling distribution of the means (all possible samples of 2 boys)

Mean (of the means) of the sampling distribution = 14

These are the means from the last two slides

Arithmetic Mean
12
12.5
12.5
13
13
13
13.5
13.5
13.5
13.5
14
14
14
14
14
14.5
14.5
14.5
14.5
15
15
15
15.5
15.5
16





These are the means from the last two slides

Population (all 5 boys) distribution

Mean of the population = 14

Standard deviation of the population = 1.58

Standard error of the population = 1.02

Sampling distribution of the means (all possible samples of 2 boys)

Mean (of the means) of the sampling distribution = 14

Standard deviation of the sampling distribution = 1.02

Central limit theorem

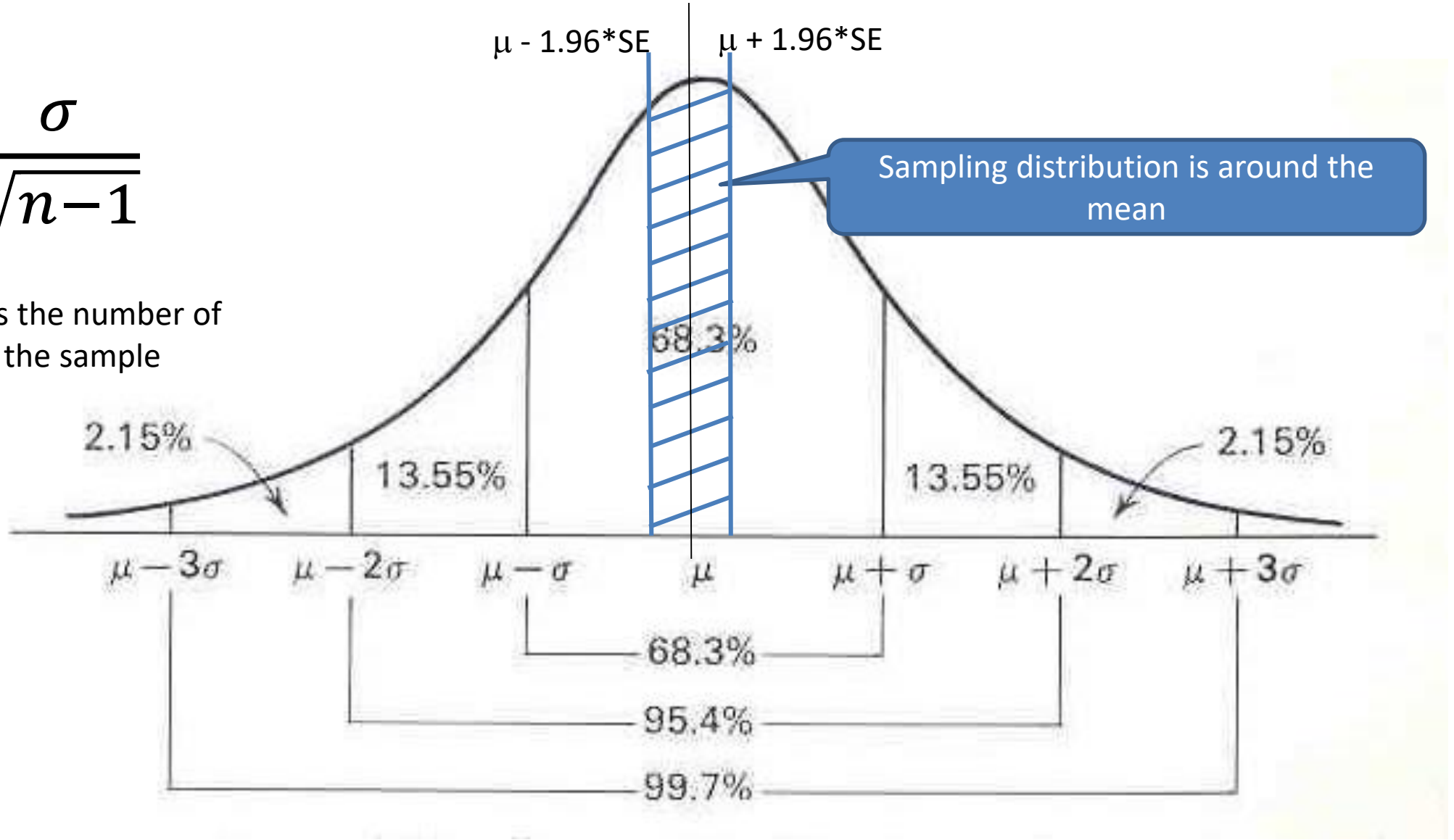
Given a population mean μ and standard deviation σ, a sampling distribution based on repeated sampling of size n has the properties:

- Mean of the sampling distribution is μ = average population
- Standard deviation of the sampling distribution = $\sigma / \sqrt{n}$, that is the standard error of the population
- If the population distribution is normal, then the sampling distribution is normal.
- If the sample is large enough, the sampling distribution approaches normal distribution.

σ and SE !!! **SE < σ** !!!

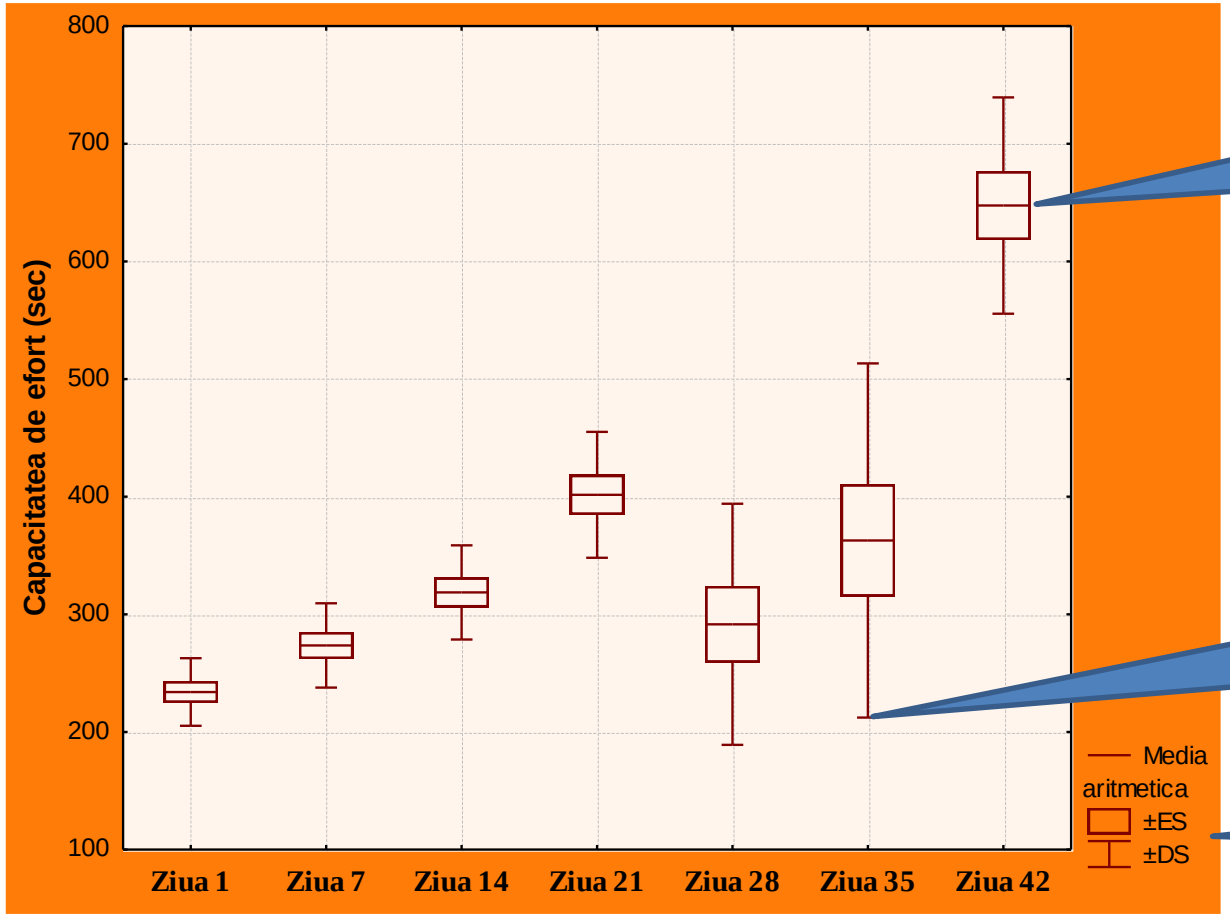
$$SE=\frac{\sigma}{\sqrt{n-1}}$$

where n is the number of people in the sample



Standard deviation vs. standard error

- σ = variation between individuals
- SE = $\sigma / \sqrt{n}$, variation between the means



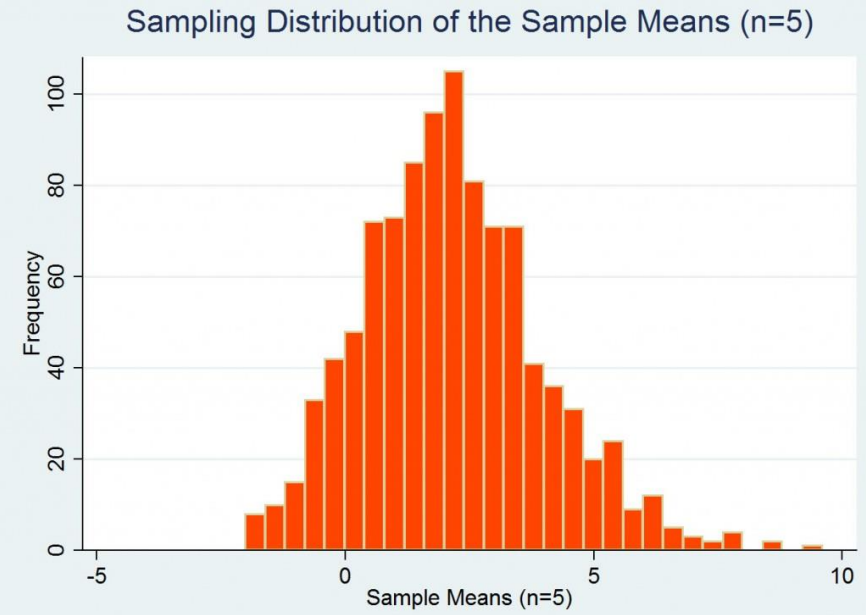
Box is arithmetic mean+/-SE = interval (mean-SE; mean+SE)

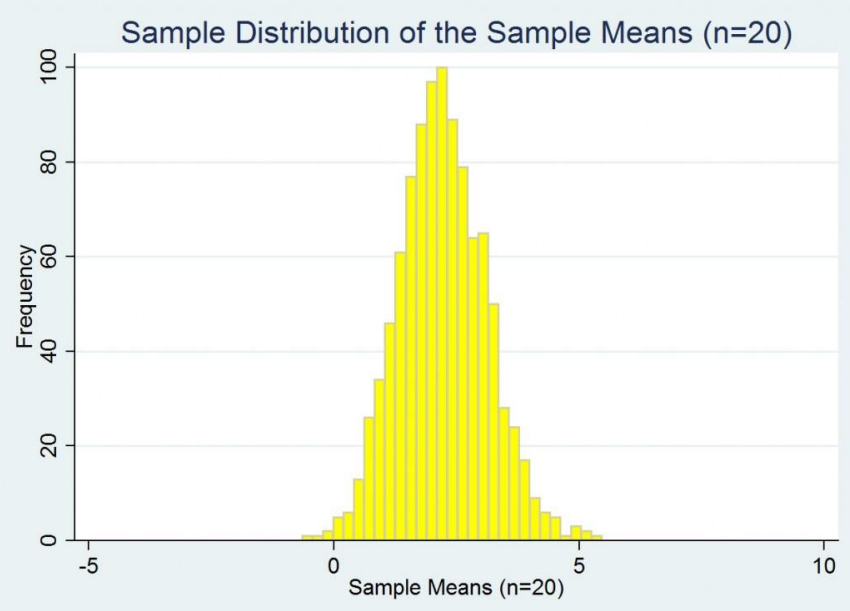
Whiskers are arithmetic mean+/- SD =interval (mean- SD; mean+ SD)

SD >SE

Law of big numbers

- bigger and bigger the sample is (more and more people taken into the study) = closer and closer to the parameter of the studied population





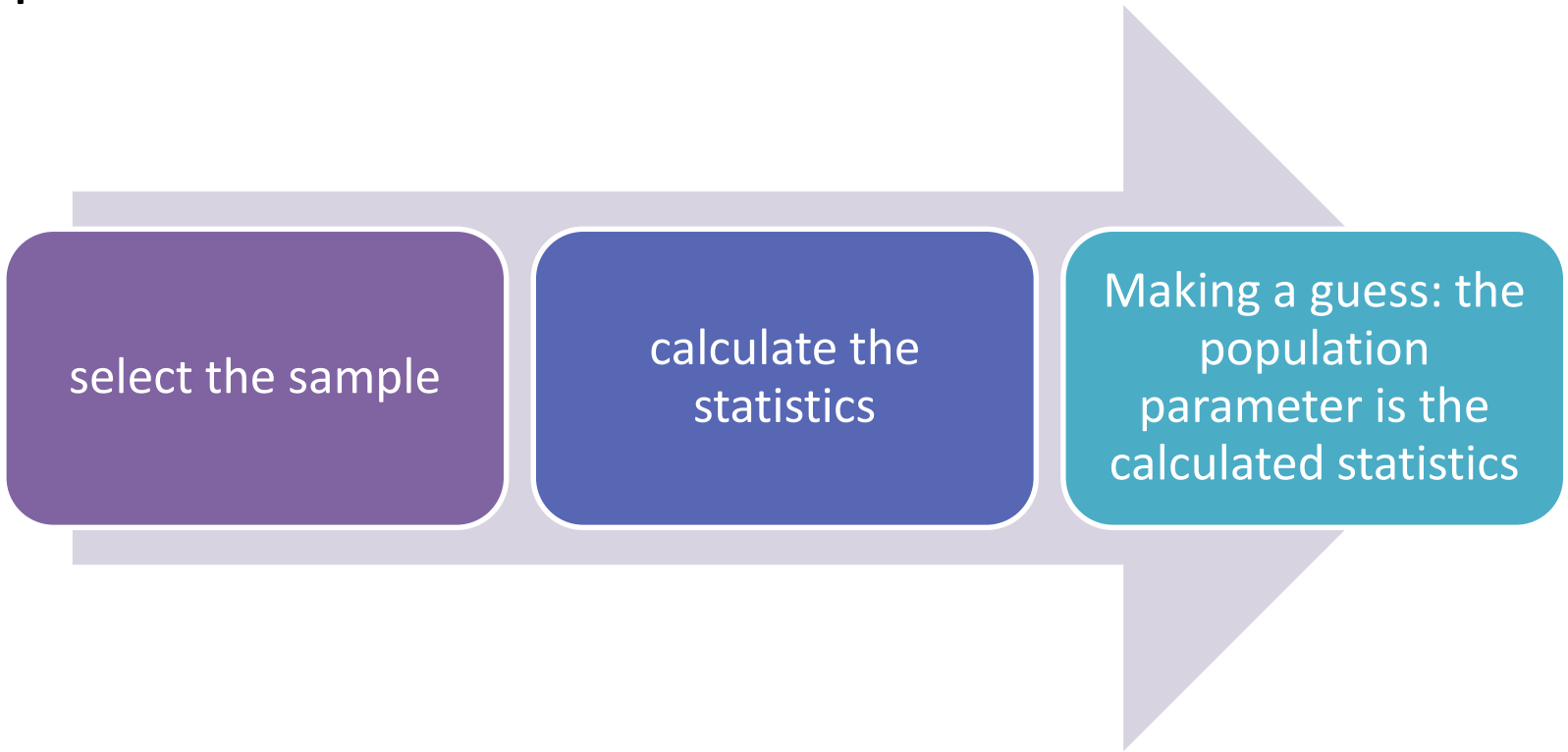
Albright EA. Sampling and the Central Limit Theorem. Nicholas School of the Environment, Duke University

- **point estimation** - we can estimate the arithmetic mean μ of the population based on the known sample $\bar{X}$
- **confidence interval estimation** - we can construct an interval that in 95% of all possible studied samples, the population arithmetic mean μ will lie within this interval.

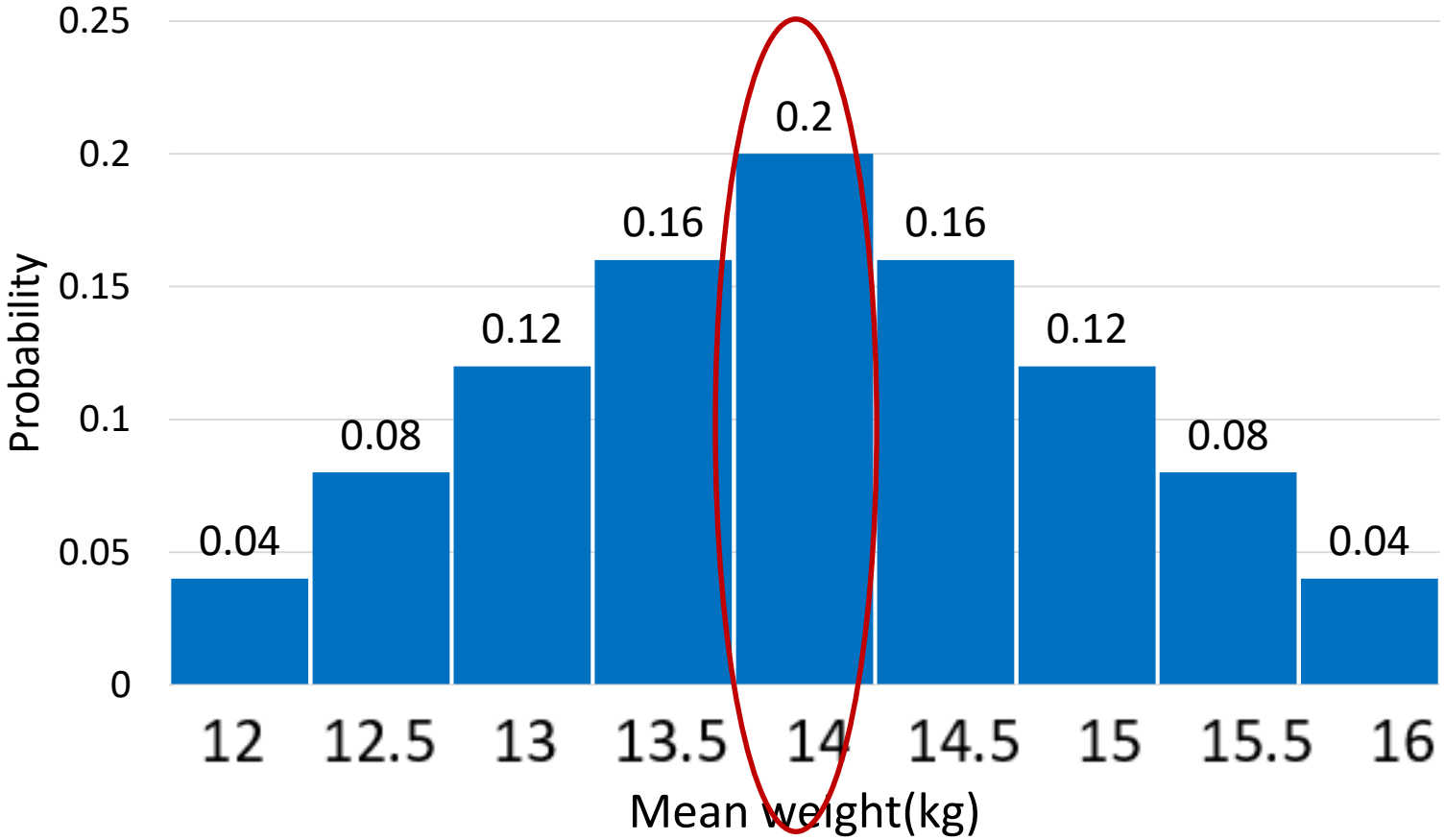
Point (punctual) estimation

estimate the arithmetic mean μ of the population
with the known sample $\bar{X}$

- we will estimate (approximate)
  - μ (mean in the population) by the sample average $\bar{X}$,
  - σ (standard deviation in the population) by the sample standard deviation s
  - ϕ (frequency in the population) by the sample frequency of the disease f



Let's pretend we select a random sample – one from the 25th possible (we estimate that the mean population is the sample mean)



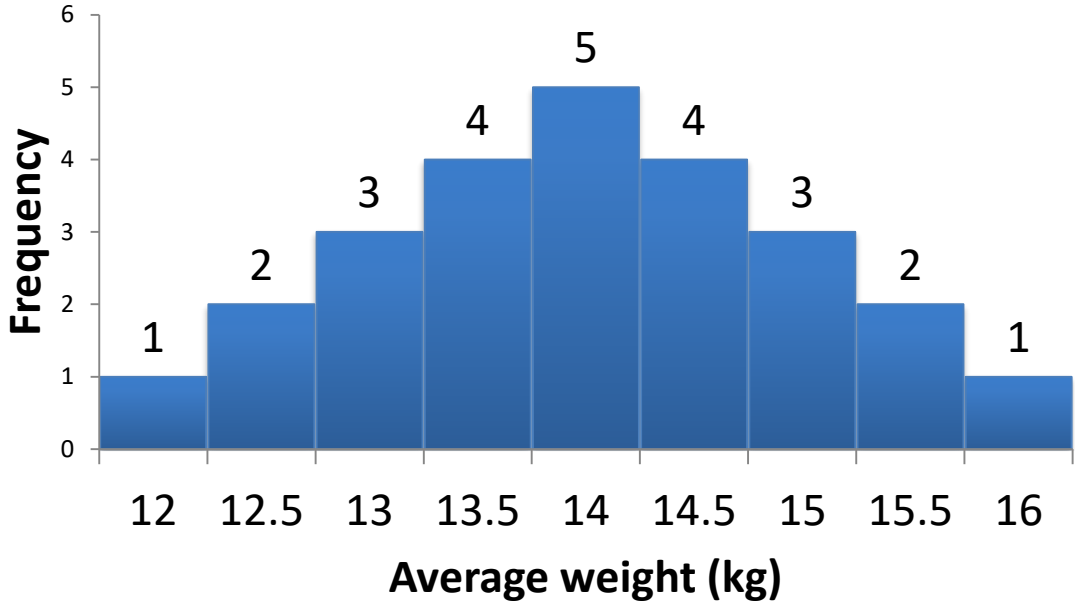
!!! In 20% of the possible selection we estimate correct μ = 14

we get 0.2 probability of correct estimation

In 16% we estimate μ = 14.5
In 16% we estimate μ = 13.5
...
in 4 % we estimate μ = 16
in 4% wee estimate μ = 12

95 % Confidence interval estimation

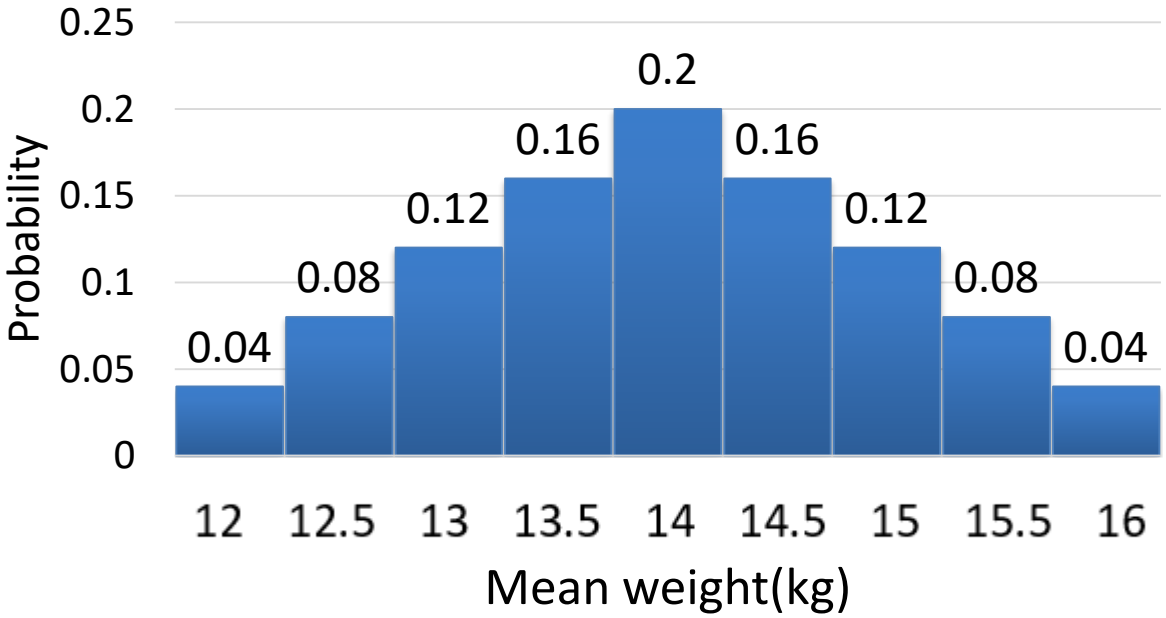
construct an interval that in 95% of all possible studied samples, the population arithmetic mean μ will lie within this interval.



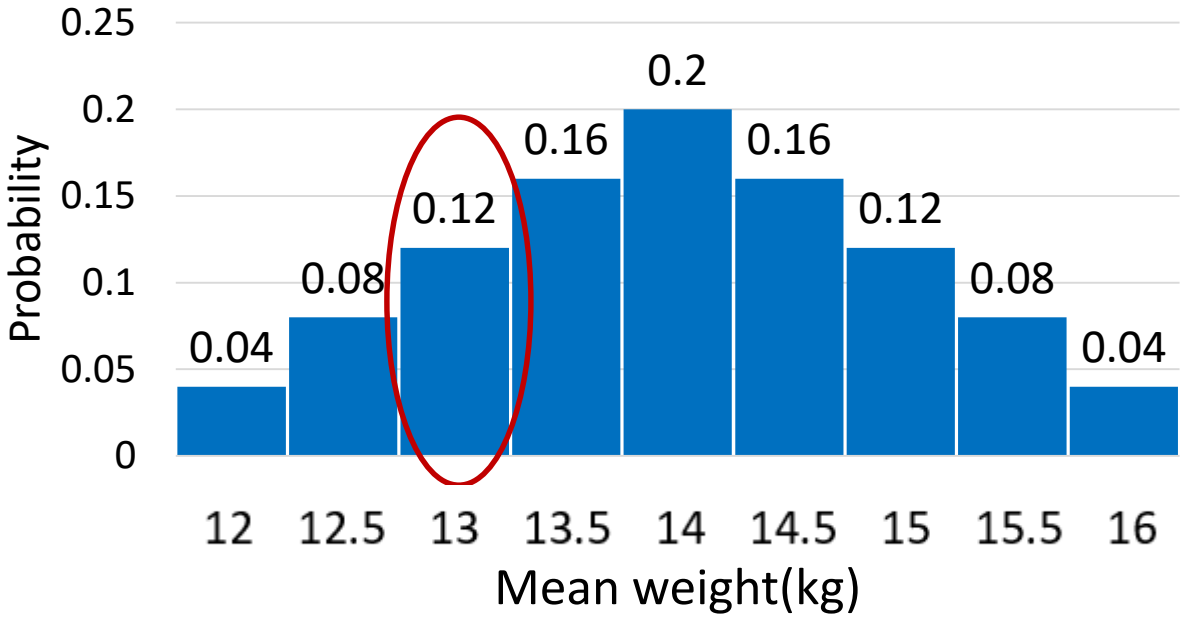
- Frequencies

How probably is to select a sample with $\bar{X}$ = 13?

How probably is it if we make a study to select a sample with the arithmetic mean < 13?

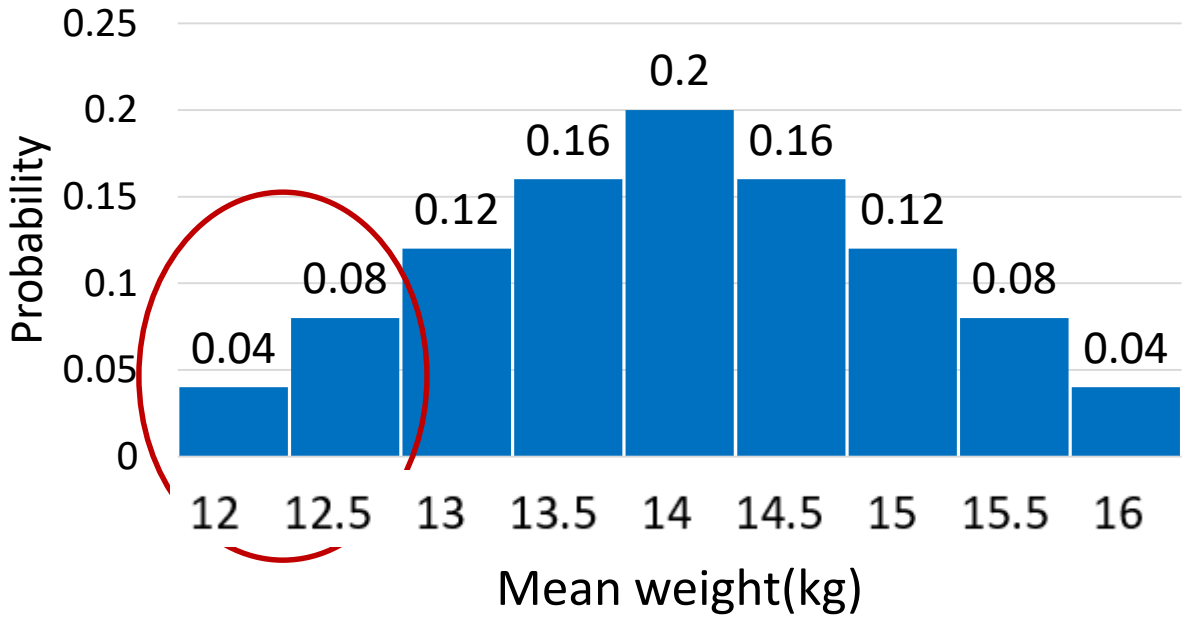


- Probabilities

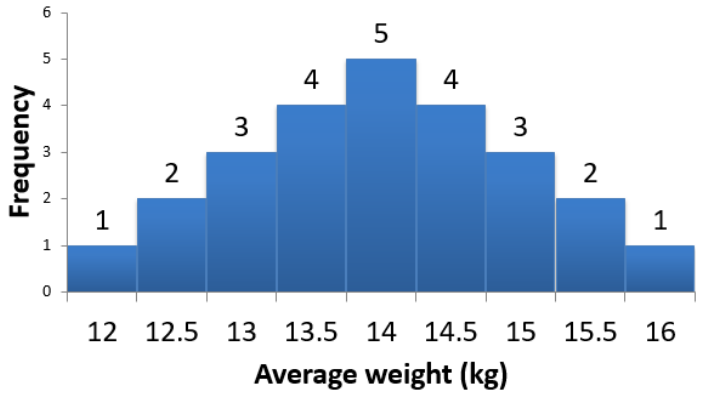


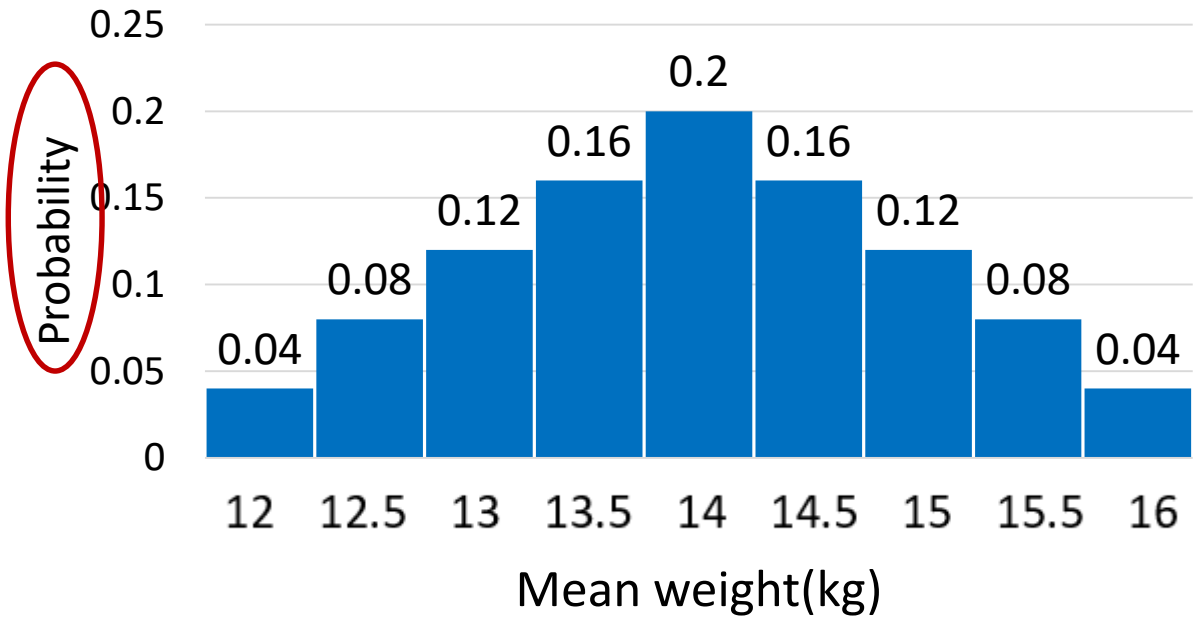
- How probably is to select a sample with $\bar{X}$ = 13?

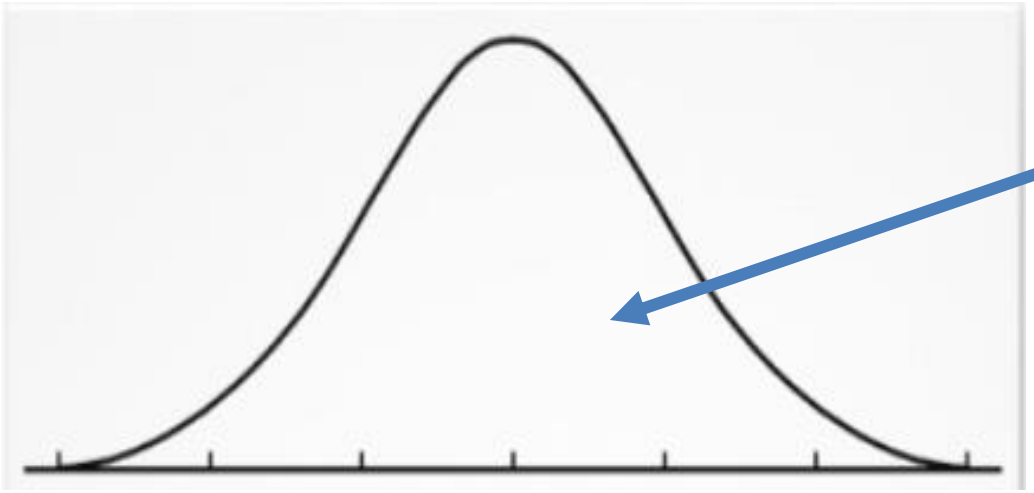
- 0.12



- How probably is it if we make a study to select a sample with the arithmetic mean < 13?
- 0.12







Area under the curve =1

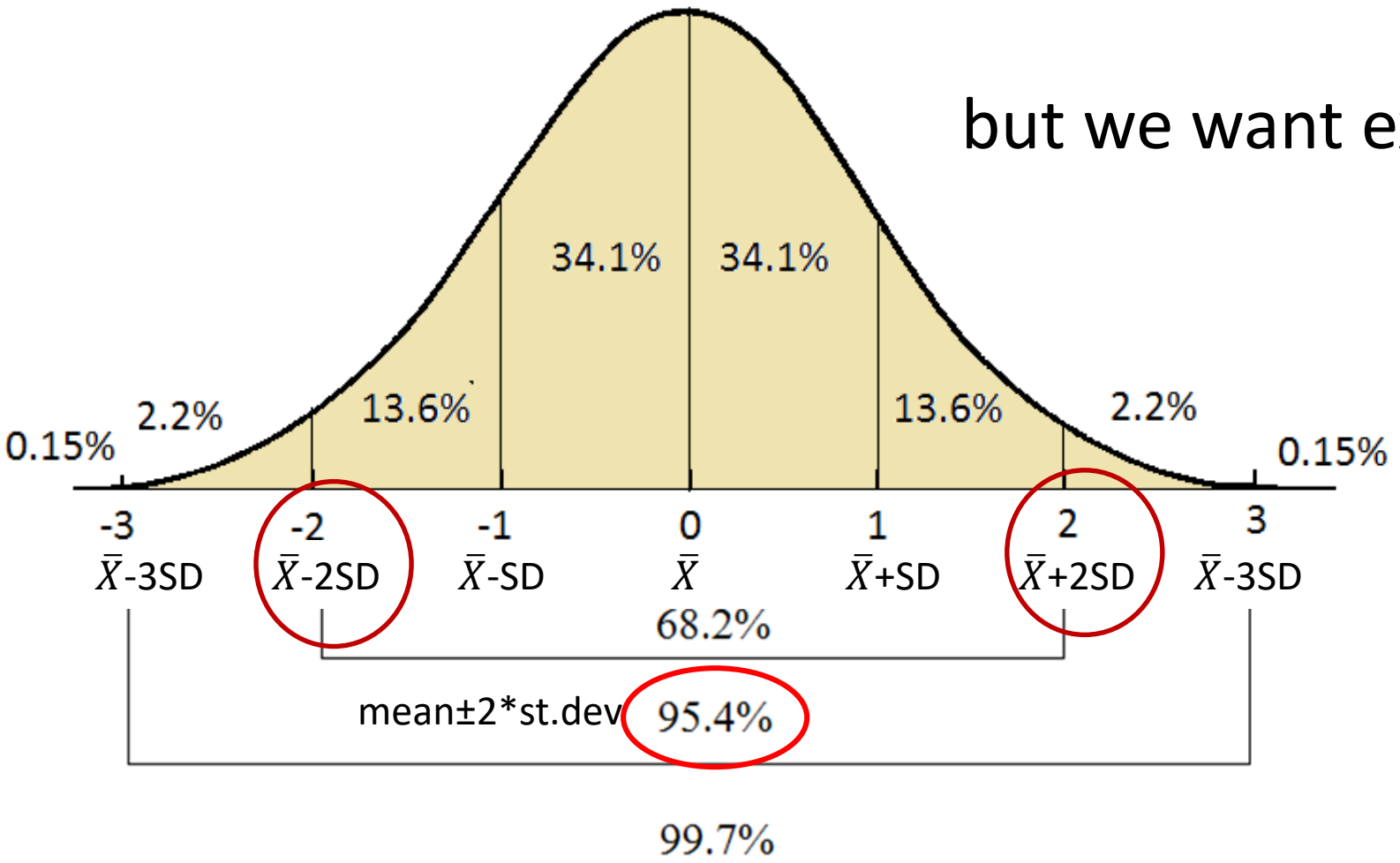
Remember the properties of Standard Normal Distribution:

mean=0, standard deviation=1

(-2;2) = ($\bar{X}$-2SD; $\bar{X}$+2SD)

between -2 and 2 there are minimum 95.4% of the data

(means in the case of the sampling distribution)

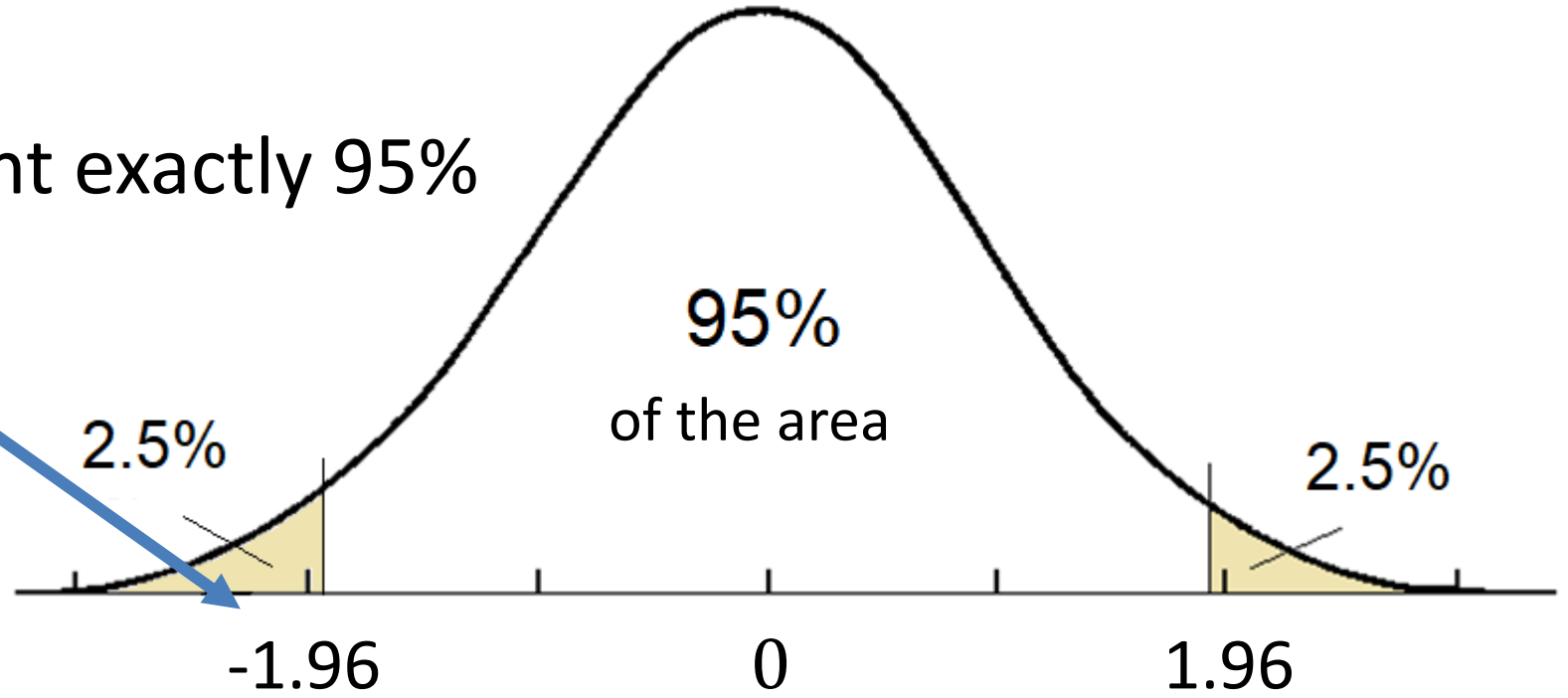


but we want exactly 95%

Standard Normal Distribution

mean=0, standard deviation=1

but we want exactly 95%



What value z divide the area in 97.5% and 2.5%?

Table of critical Z for standard normal distribution

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.00	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554			0.8621
1.10	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770			0.8830
1.20	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962			0.9015
1.30	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.40	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.50	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.60			0.9474	0.9484			0.9515	0.9525	0.9535	0.9545
1.70			0.9573	0.9582			0.9608	0.9616	0.9625	0.9633
1.80			0.9656	0.9664			0.9686	0.9693	0.9699	0.9706
1.90	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.00	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.10	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.20	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.30	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.40	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936

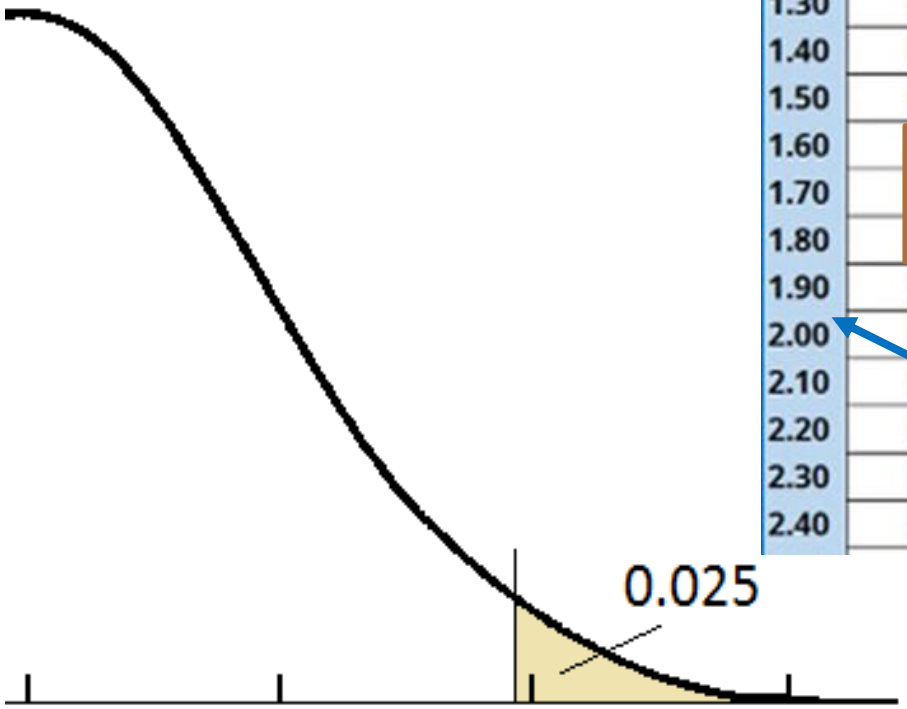
0.06

1.90

0.9750

First decimal

Second decimal



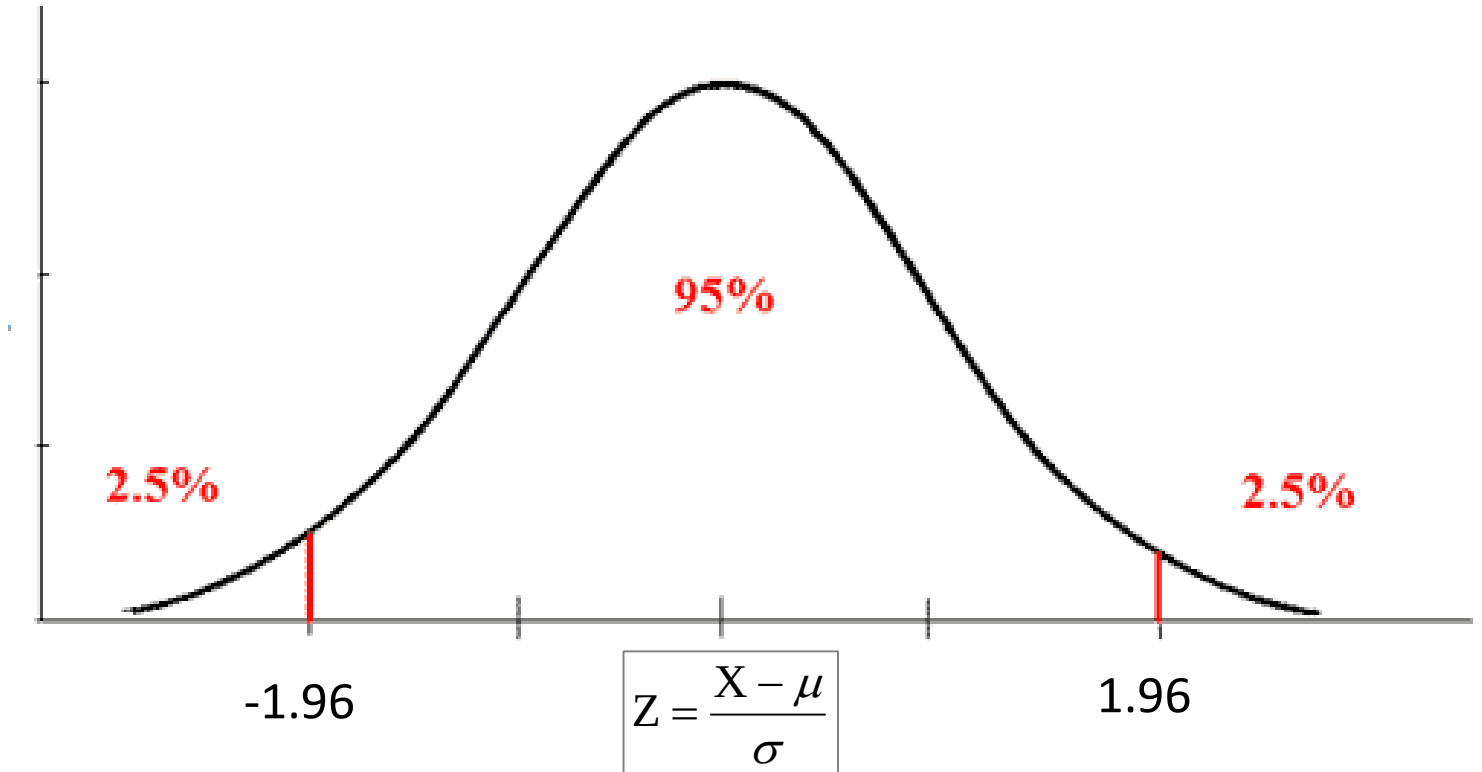
$$Z_{1-\frac{0.05}{2}} = 1.96$$

critical Z for α=0.05 is 1.96

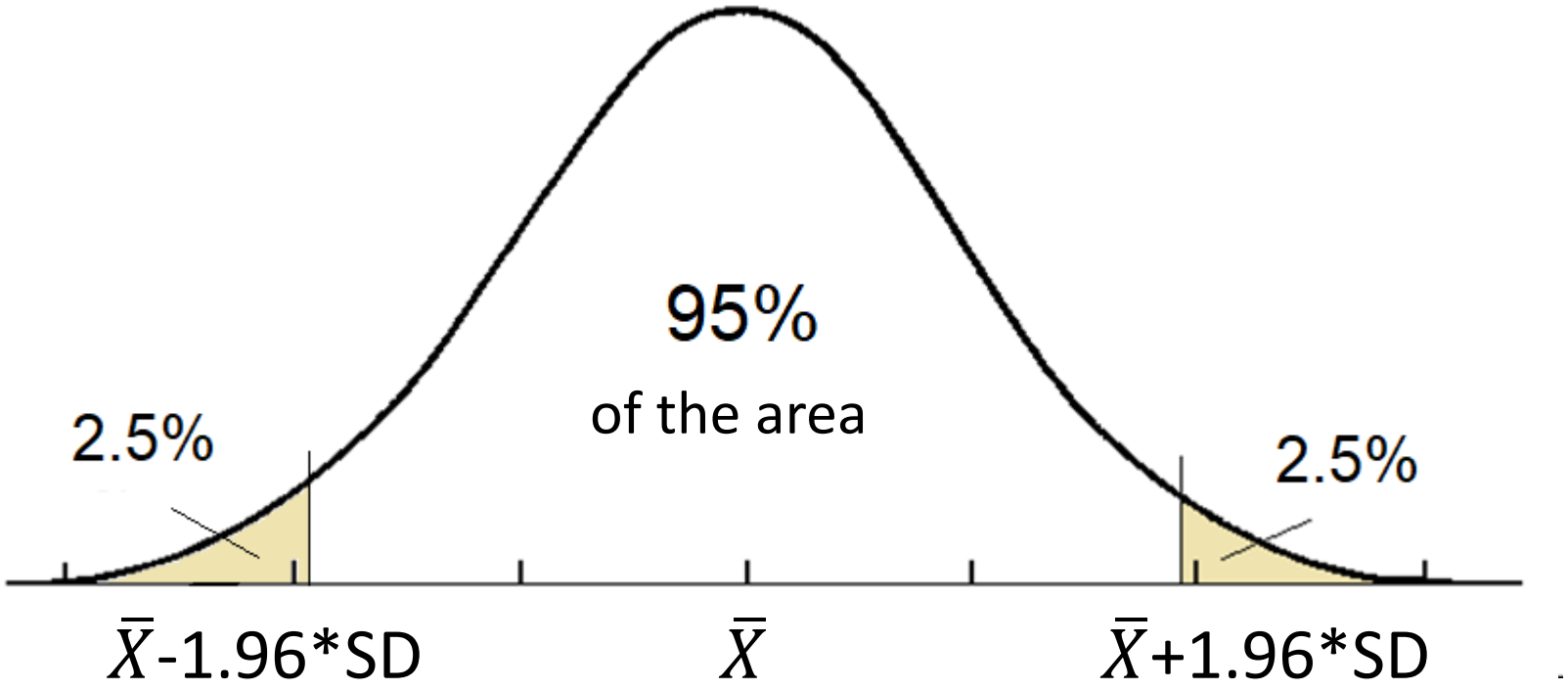
Confidence interval for the mean

We will find the confidence interval of 95%

Interval [-1.96, +1.96] had 95% of the z values of normal distribution

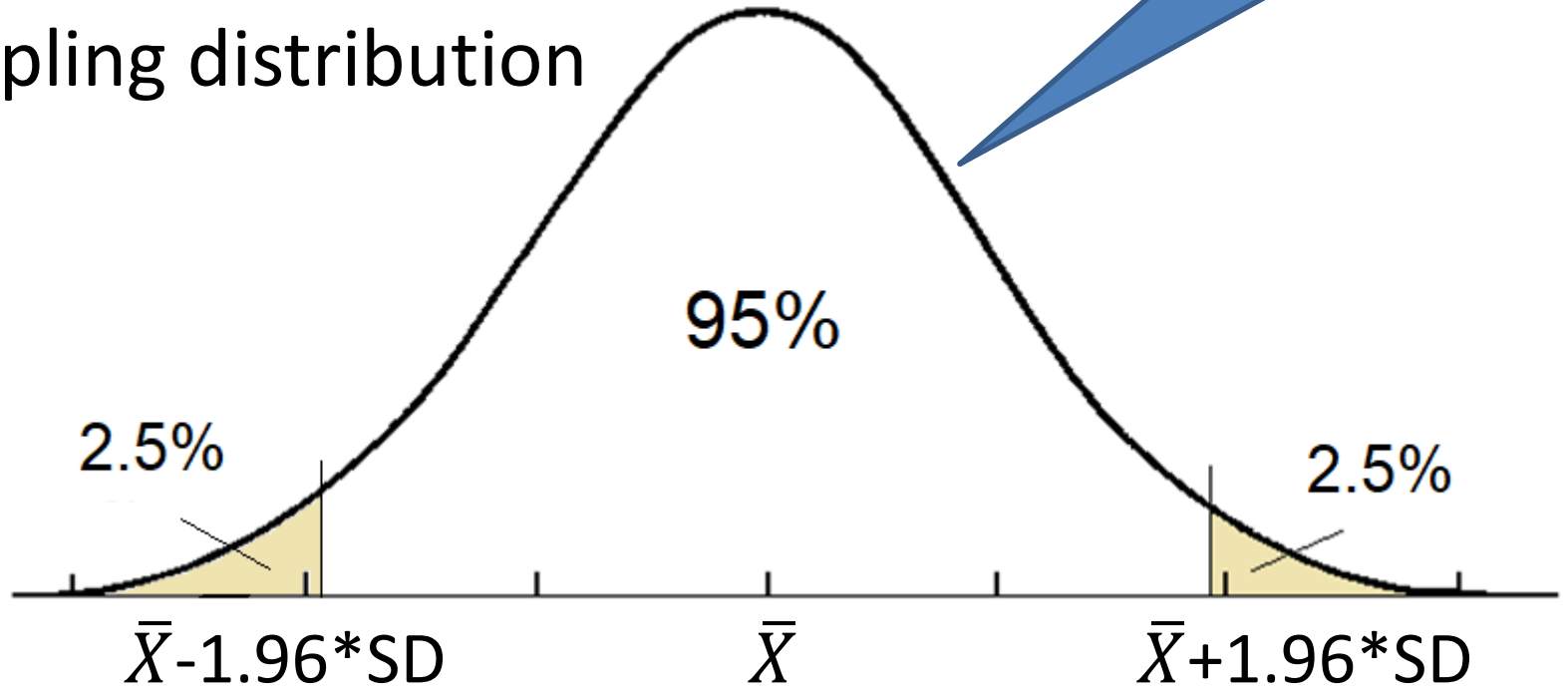


Normal Distribution

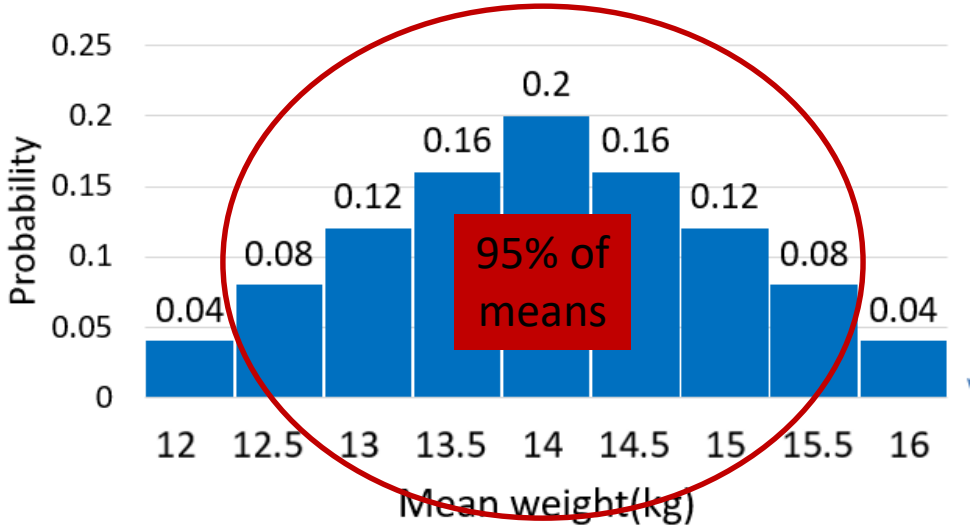


this is the distribution of the means if we repeat the study

sampling distribution



SD for sampling distribution = SE-standard error of the population



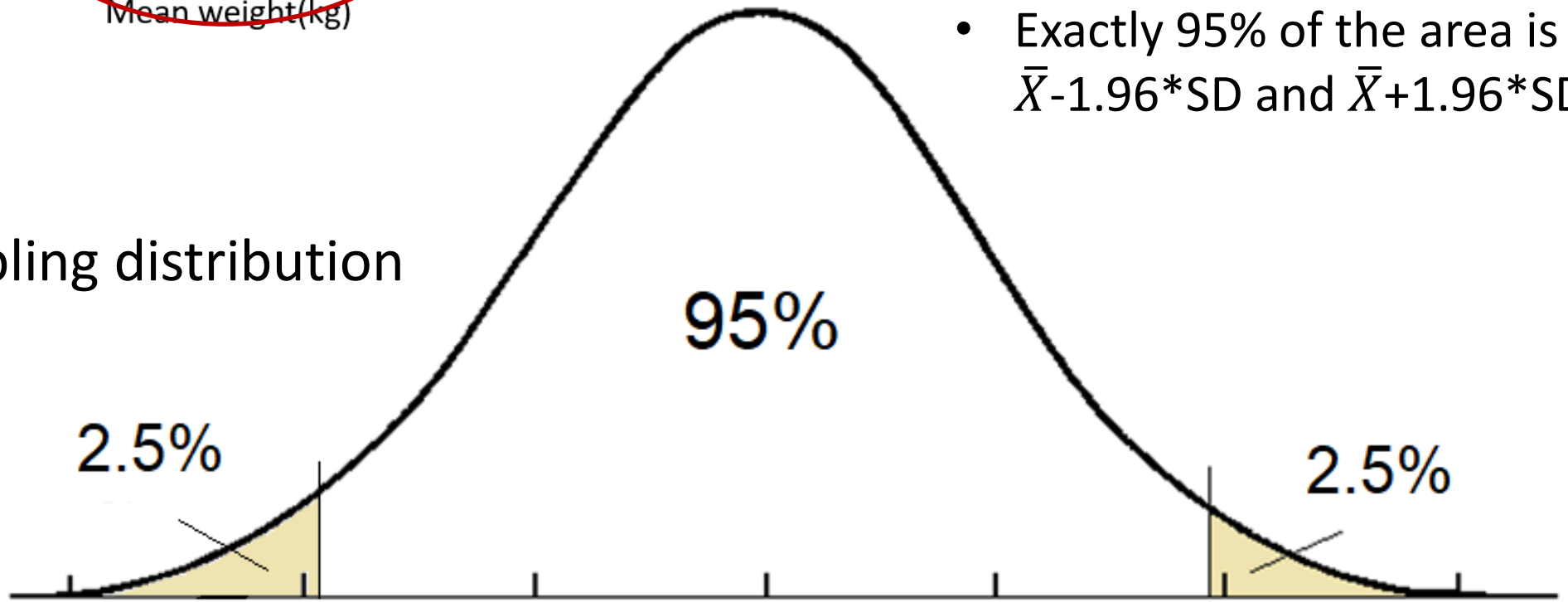
Central limit theorem

Sampling distribution

- Mean of the sampling distribution is μ = average population
- Standard deviation of the sampling distribution = $\sigma / \sqrt{n}$, that is the standard error of the population

- Exactly 95% of the area is between $\bar{X}$-1.96*SD and $\bar{X}$+1.96*SD

Sampling distribution

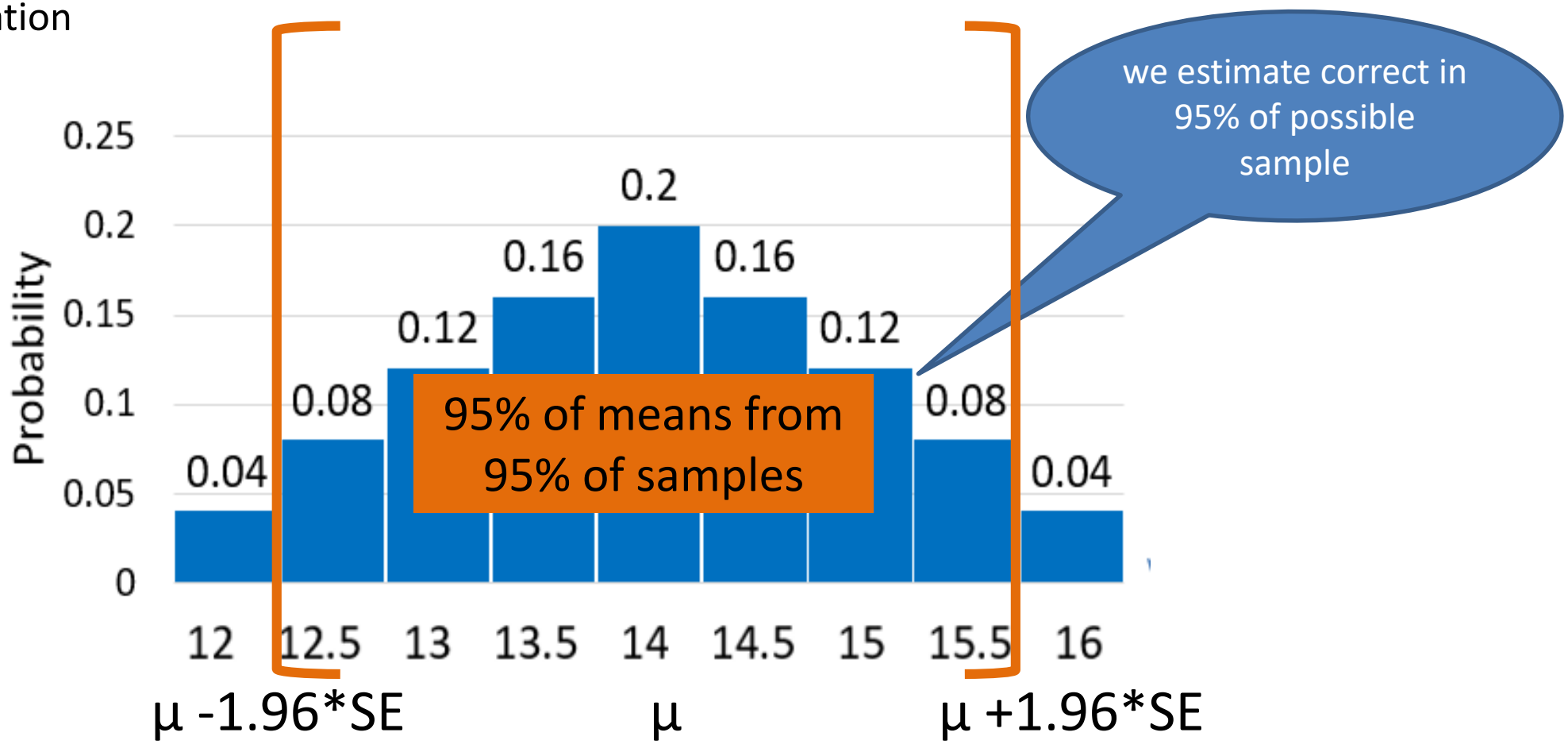


$\bar{X}$-1.96*SD
μ -1.96*SE

$\bar{X}$
μ

$\bar{X}$+1.96*SD
μ +1.96*SE

Exactly 95% of the weight means of the replicated studies on all possible samples of 5 boys will be between μ-1.96SE to μ+1.96SE, where μ is the arithmetic mean of the weight in the population and SE is the standard error in the population



Confidence interval estimation

95 % confidence interval for the mean μ

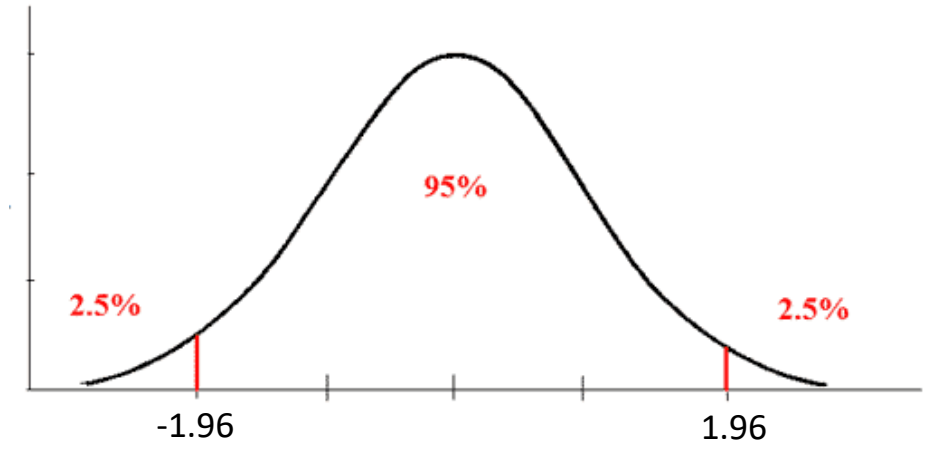
$P(-1.96 \leq z \leq 1.96) = 0.95$

$P(-1.96 \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq 1.96) = 0.95$ ← $Z = \frac{X-\mu}{\sigma}$

$P(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$

$[\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}]$ 95 % confidence interval for the mean μ

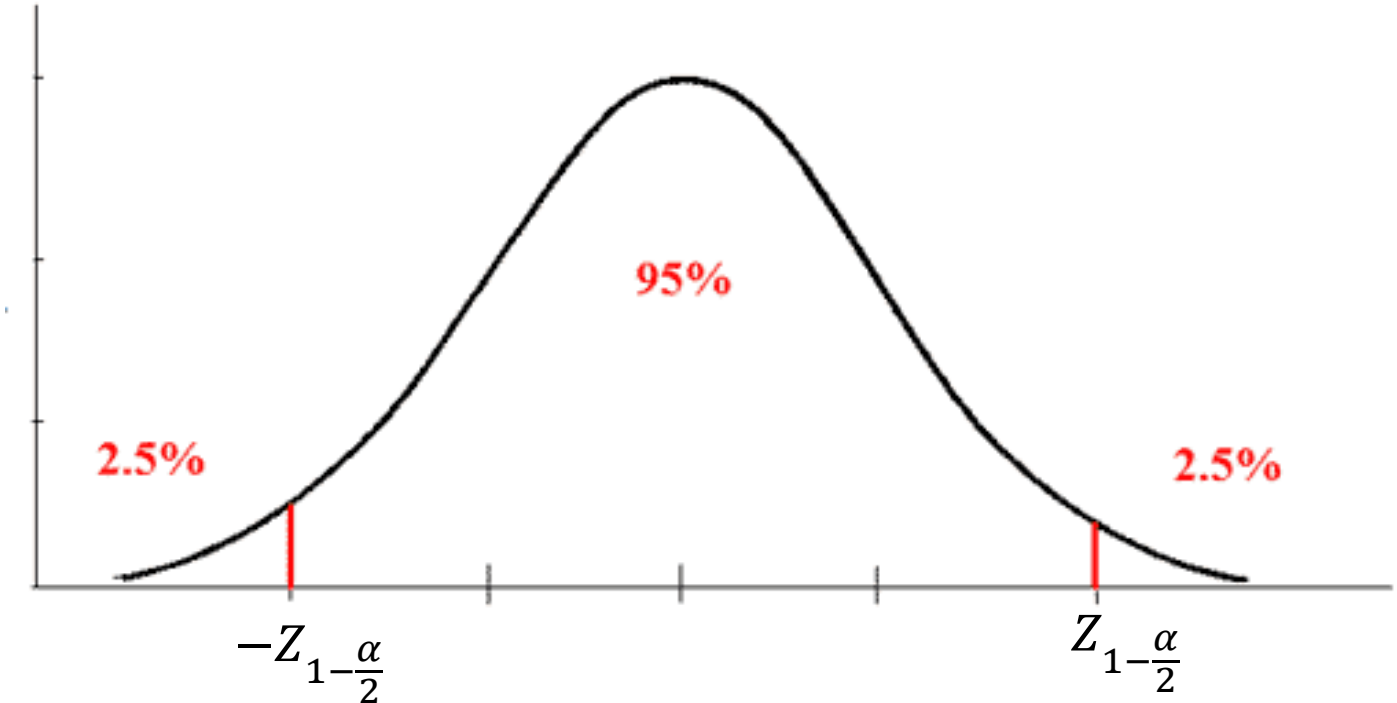
standard error SE = $\frac{\sigma}{\sqrt{n}}$ = standard deviation of sampling distribution



1-α Confidence interval for the arithmetic mean μ when **σ is known**

$$[\bar{X} - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}]$$

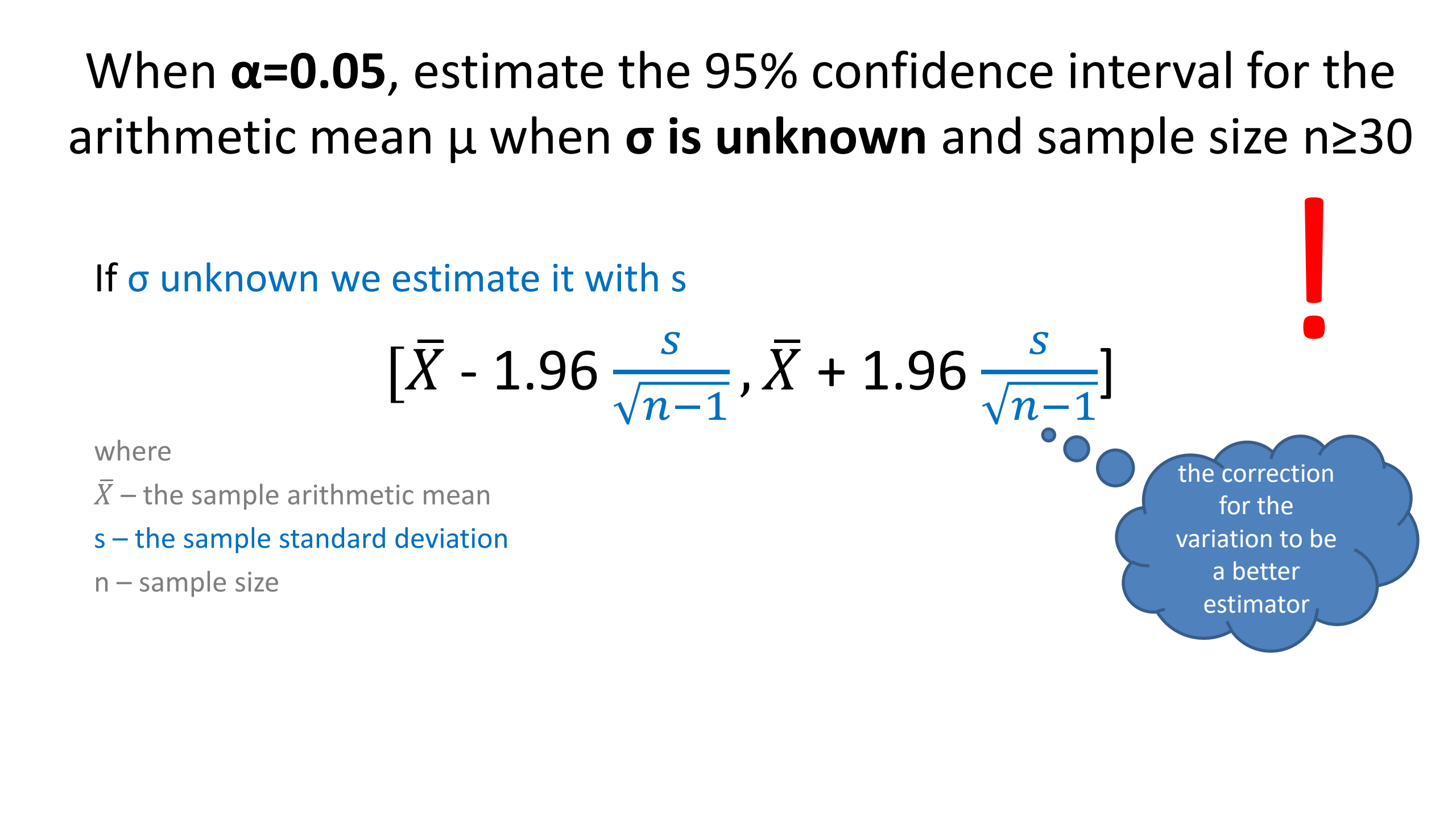
$Z_{1-\frac{\alpha}{2}}$ = Z critic



When **α=0.05**, estimate the 95% confidence interval for the arithmetic mean μ when **σ is unknown** and sample size n≥30

If σ unknown we estimate it with s

$$\left[\bar{X} - 1.96 \frac{s}{\sqrt{n-1}}, \bar{X} + 1.96 \frac{s}{\sqrt{n-1}}\right]$$



where

$\bar{X}$ – the sample arithmetic mean

s – the sample standard deviation

n – sample size

Second method of making estimation

The 95% confidence interval of the arithmetic mean in the case of n≥30 and σ unknown

$$(\bar{X} - 1.96 \cdot SE; \bar{X} + 1.96 \cdot SE)$$

where

$\bar{X}$ - arithmetic mean of the sample,

SE=$\frac{s}{\sqrt{n-1}}$ standard error of the sample,

s- standard deviation,

n – sample size (number of people)

σ – standard deviation of the population

$$(\bar{X} - 1.96 \frac{s}{\sqrt{n-1}}, \bar{X} + 1.96 \frac{s}{\sqrt{n-1}})$$

interpretation of the 95% confidence interval

= 95% of the calculated means of all possible sample of n subjects taken from the population are included (if we repeat the study)

- or

= we are 95% confident that the population (true) mean is between the lower and upper margins of the interval

= interval where the population (true) mean is included with a 5% level of error

we do not know exactly the value of the mean, but is in this interval with 95% probability

Example

The average incubation period in days of a random sample of 43 cases of hepatitis A from an epidemic source in city X in 1992 is $\bar{X}$ = 28.93 and standard deviation s=3.68

95% confidence interval of the mean incubation period:

$$(\bar{X} - 1.96 \frac{s}{\sqrt{n-1}}, \bar{X} + 1.96 \frac{s}{\sqrt{n-1}})$$

$$\left(28.93 - 1.96\frac{3.68}{\sqrt{43-1}}; 28.93 - 1.96\frac{3.68}{\sqrt{43-1}}\right)$$

$$(28.93 - 1.11; 28.93 + 1.11)$$

(27.82;30.04)

Interpretation: the population mean incubation period is between 27.82 and 30.04 with a 5% error

Example

Objectiv: to estimate the arithmetic mean μ of cholesterol in the population

- a sample **n=101**
- arithmetic mean in the sample

$$\bar{X} = 120 \text{ mg/dl}$$

- standard deviation in the sample

s=16

$$[\bar{X} - 1.96 \frac{s}{\sqrt{n-1}}, \bar{X} + 1.96 \frac{s}{\sqrt{n-1}}]$$

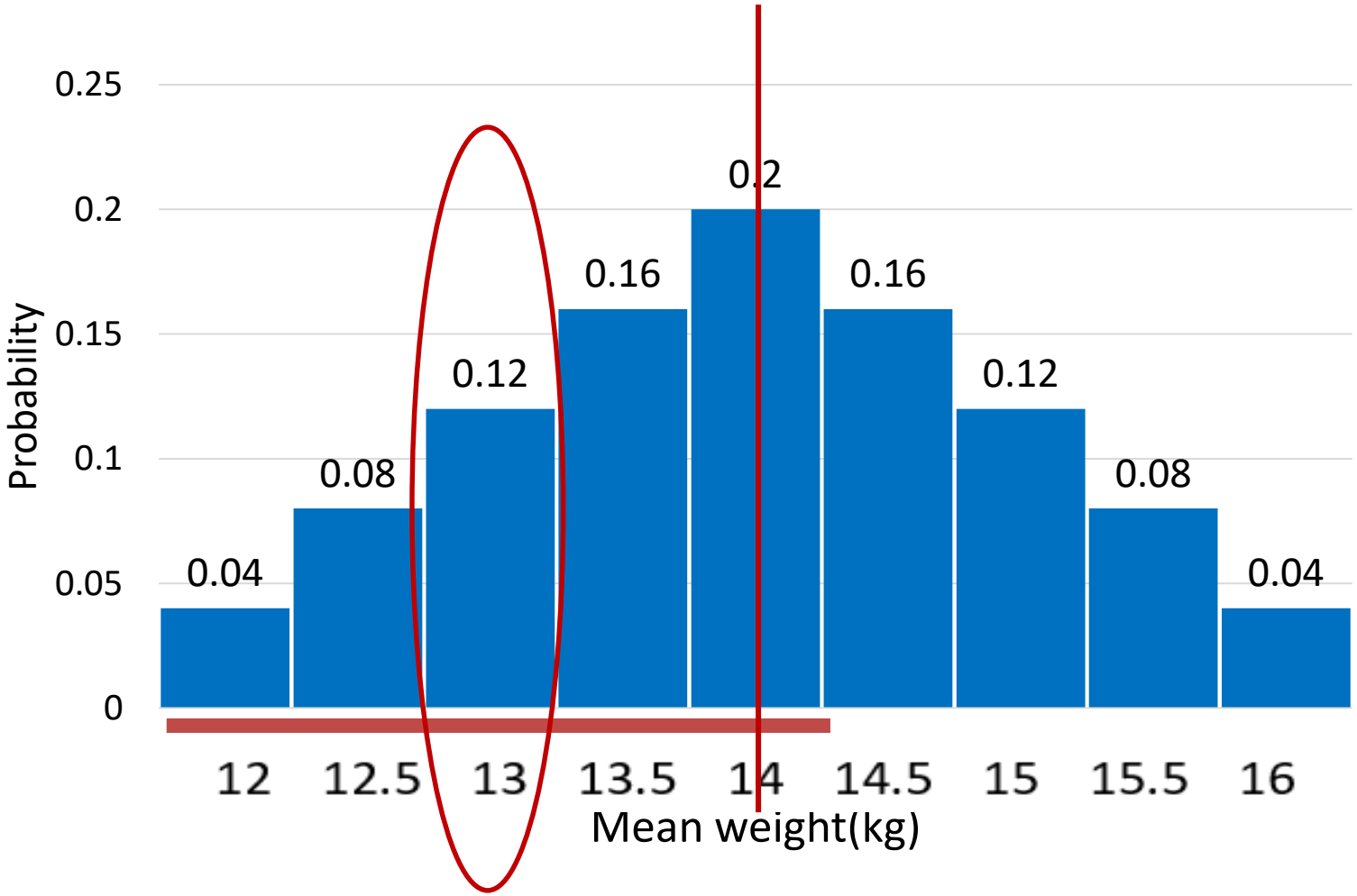
$$[120 - 1.96 \frac{16}{\sqrt{101-1}}; 120 - 1.96 \frac{16}{\sqrt{101-1}}]$$

$$[120 - 3.14; 120 + 3.14]$$

[116.86; 123.14] – the 95% confidence interval

Result: with a 5% error we estimate that the population arithmetic mean μ of cholesterol is between 116.86 and 123.14 mg/dl

Let's pretend we select a random sample – one from the 25th possible (the arithmetic mean μ of the population is not known)

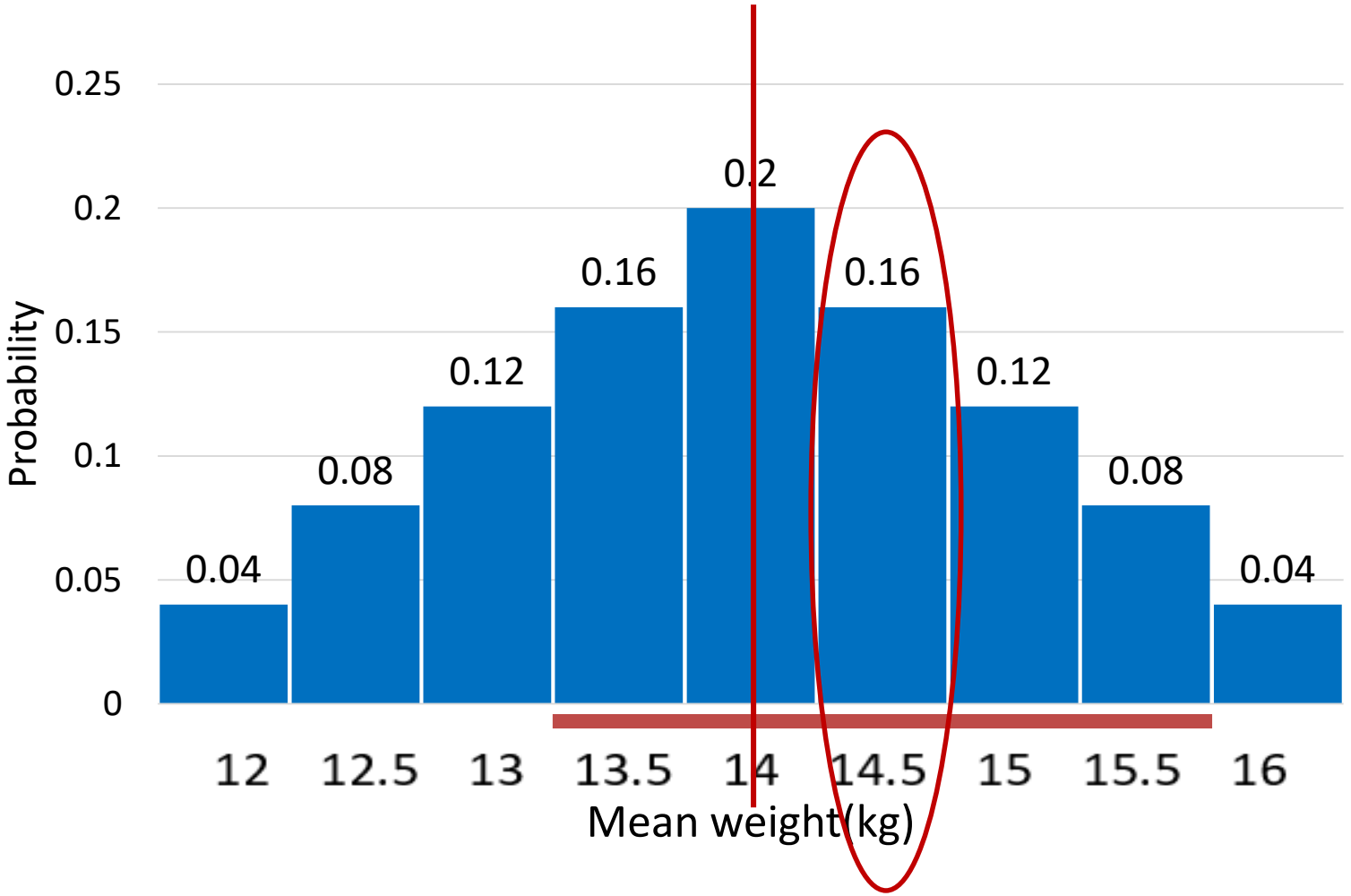


let's pretend that we select one with the arithmetic mean 13

we calculate a 95% confidence interval 12 - 14 where with 95% probability the population arithmetic mean is

!!! 14 is in the interval, we estimate correct

Let's pretend we select a random sample – one from the 25th possible (the arithmetic mean μ of the population is not known)



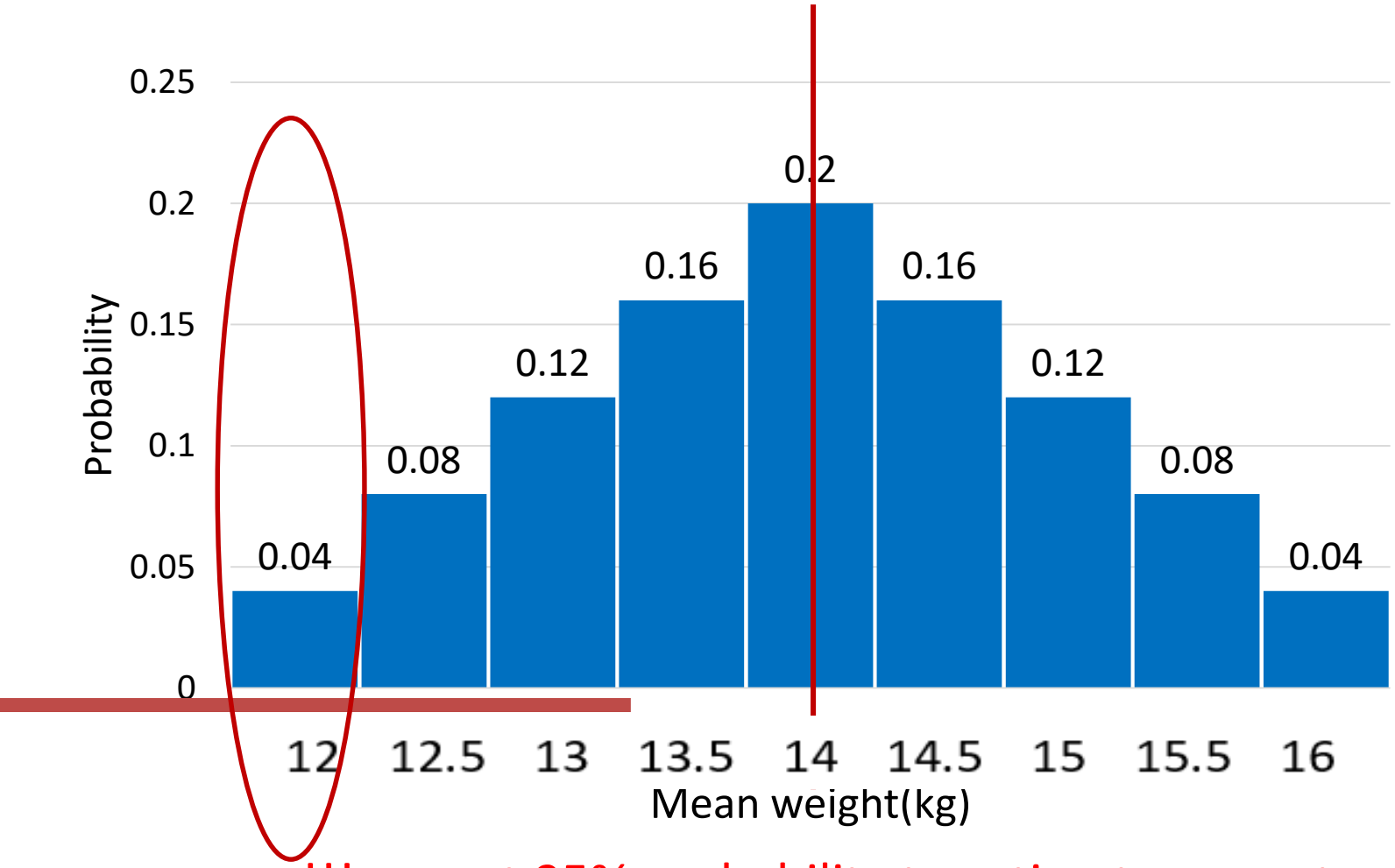
!!! we get 95% probability to estimate correct

let's pretend that we select one with the arithmetic mean 14.5

we calculate the 95% confidence interval 13.5 – 15.5 where with 95% probability the population arithmetic mean is

!!! 14 is in the interval, we estimate correct

Let's pretend we select a random sample – one from the 25th possible (the arithmetic mean μ of the population is not known)



!!! we get 95% probability to estimate correct

!!! we will have 5% probability of incorrect estimation

let's pretend that we select one with the arithmetic mean 12

we calculate the 95% confidence interval 8 – 13 where with 95% probability the population arithmetic mean is

!!! 14 is not in the interval, we estimate incorrect

1-α confidence intervals for a frequency

- The formula:

$$\left(f - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{f(1-f)}{n}} ; f + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{f(1-f)}{n}} \right)$$

Where f – frequency, f<1, n – total number of subjects

When **α=0.05**, estimate the 95% confidence intervals for a frequency

- The formula:

$$\left(f - 1.96\sqrt{\frac{f(1-f)}{n}}; f + 1.96\sqrt{\frac{f(1-f)}{n}} \right)$$

Where f – frequency, f<1, n – total number of subjects

Example

- We want to estimate the frequency of mouth cancer in population with age higher than 60 years old.
- In a study with 10000 participants, 300 had mouth cancer

$$f = \frac{300}{10000} = 0.03$$

95% confidence intervals for the frequency of mouth cancer

$$(f - 1.96\sqrt{\frac{f(1-f)}{n}}; f + 1.96\sqrt{\frac{f(1-f)}{n}})$$

$$\left(0.03 - 1.96\sqrt{\frac{0.03(1-0.03)}{10000}}; 0.03 + 1.96\sqrt{\frac{0.03(1-0.03)}{10000}}\right)$$

$$(0.03 - 0.003; 0.03 + 0.003)$$

(0.027;0.033)

Interpretation: with a 5% error we estimate that the population frequency of mouth cancer in the population is between 0.027 and 0.033

If we select a small sample size?

- We want to estimate the frequency of mouth cancer in population with age higher than 60 years old.
- In a study with 1000 participants, 300 had mouth cancer

$$f = \frac{30}{1000} = 0.03$$

$$[0.03 - 1.96\sqrt{\frac{0.03(1-0.03)}{1000}}; 0.03 + 1.96\sqrt{\frac{0.03(1-0.03)}{1000}}]$$

$$[0.03 - 0.011; 0.03 + 0.011]$$

[0.019; 0.041] – interval for n=1.000

frequency of mouth cancer between 1.9% și 4.1% with a 95% probability

[0.027; 0.033] – interval for n=10.000

frequency of mouth cancer between 2.7% și 3.3% with a 95% probability

Answer: smaller the sample → wider the interval

an increasing n at the denominator, an opposite effect

If the sample increase → it increases the measurement precision by decreasing the interval required for estimation

If the probability decrease 95% → 80%

- We want to estimate the frequency of mouth cancer in population with age higher than 60 years old.
- In a study with 1000 participants, 300 had mouth cancer

For 80% confidence, Z critic =1.29

$$f = \frac{30}{1000} = 0.03$$

$$[0.03 - 1.29\sqrt{\frac{0.03(1-0.03)}{1000}}; 0.03 + 1.29\sqrt{\frac{0.03(1-0.03)}{1000}}]$$

$$[0.03 - 0.007; 0.03 + 0.007]$$

[0.023; 0.037] – interval for 80%

frequency of mouth cancer between 2.3% și 3.7% with a 95% probability

[0.019; 0.041] – interval for 95%

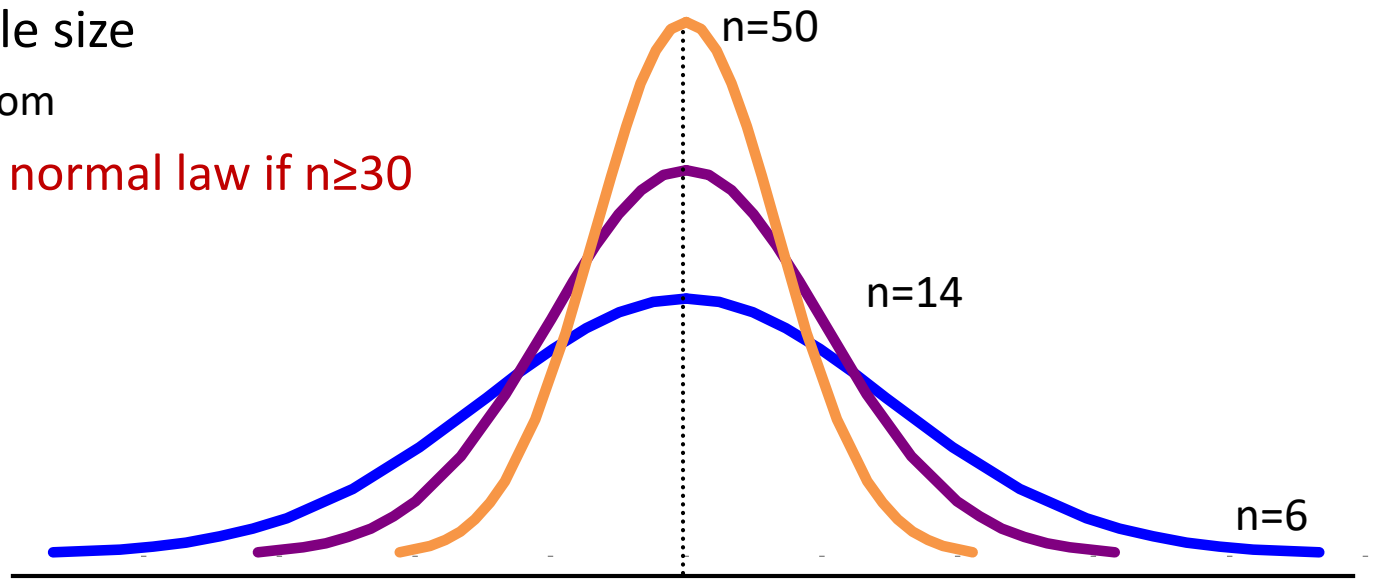
frequency of mouth cancer between 1.9% și 4.1% with a 95% probability

Answer: the smaller the probability – the narrower the interval

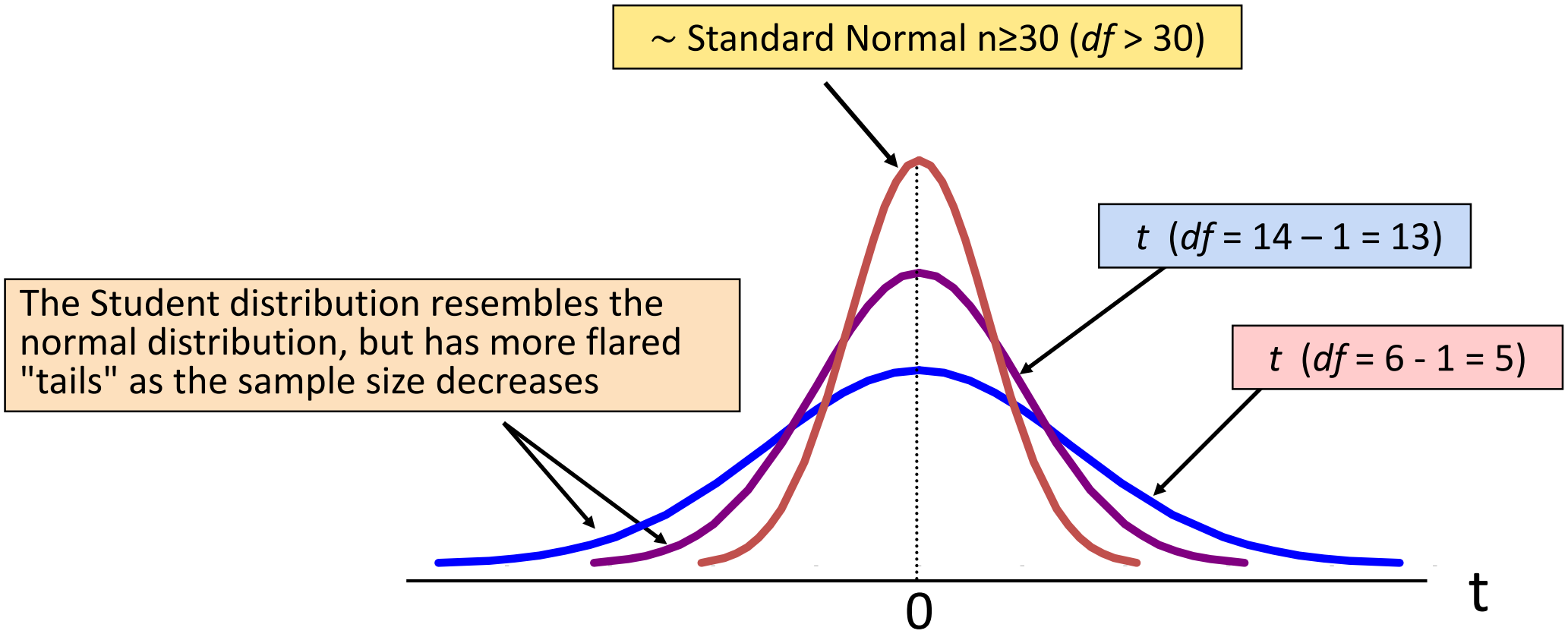
Z at numerator

Increasing precision → interval increase

- The sampling distribution of the mean is normally distributed when the population standard deviation σ is known
- Often σ is not known
  - it is estimated with the standard deviation obtained on the sample s
    - in this case the sampling distribution no longer follows the normal law
    - the sampling distribution follows Student's law – very similar
    - Student's law varies with sample size
      - depends on the degrees of freedom
      - Student's law is close to the normal law if $n \geq 30$



If n<30: Student vs Normal – we estimate with Student distribution



Student's law varies with sample size

depends on the degrees of freedom

Degree of freedom = df

The number of components that are "free" to vary in a data set

- df = n – 1

Ex: 5 measurements, the average = 6

- we know the first four values: 8, 9, 10, 11
- the fifth can be calculated

--> the data set has only 4 degrees of freedom

1-α% confidence interval for the average μ in the case of small samples n<30 with σ unknown

- using Student t distribution

$$\left(\bar{X} - t_{n-1,1-\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}}, \bar{X} + t_{n-1,1-\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}}\right)$$

where

$\bar{X}$ - sample arithmetic mean,

s – sample standard deviation,

n – sample size,

$t_{n-1,1-\frac{\alpha}{2}}$ critical t for n-1 degree of freedom

$1-\alpha$ level of confidence

- df = n-1
- df = degree of freedom

VALUES OF *t* FOR 90%, 95%, AND 99% CONFIDENCE INTERVALS

df	90%	95%	99%
5	2.015	2.571	4.032
6	1.943	2.447	3.707
7	1.895	2.365	3.499
8	1.860	2.306	3.355
9	1.833	2.262	3.250
10	1.812	2.228	3.169
11	1.796	2.201	3.106
12	1.782	2.179	3.055
13	1.771	2.160	3.012
14	1.761	2.145	2.977
15	1.753	2.131	2.947
16	1.746	2.120	2.921
17	1.740	2.110	2.898
18	1.734	2.101	2.878
19	1.729	2.093	2.861
20	1.725	2.086	2.845
30	1.697	2.042	2.750
40	1.684	2.021	2.704
60	1.671	2.000	2.660
120	1.658	1.980	2.617
∞	1.645	1.960	2.576

$t_{n-1,1-\frac{\alpha}{2}}$ variation on df

- n=6 → df =n-1= 5
  - confidence interval 95% → t= ±2.571
- n=10 → df = 9
  - confidence interval 95% → t= ±2.262
- n=30 → df = 29
  - confidence interval 95% → t= ±2.042
- The increase of n causes the value of t to approach 1.96, →the curve tends towards the normal distribution.
- The differences in calculations can be considered negligible for samples of size n ≥ 30

VALUES OF *t* FOR 90%, 95%, AND 99% CONFIDENCE INTERVALS

df	90%	95%	99%
5	2.015	2.571	4.032
6	1.943	2.447	3.707
7	1.895	2.365	3.499
8	1.860	2.306	3.355
9	1.833	2.262	3.250
10	1.812	2.228	3.169
11	1.796	2.201	3.106
12	1.782	2.179	3.055
13	1.771	2.160	3.012
14	1.761	2.145	2.977
15	1.753	2.131	2.947
16	1.746	2.120	2.921
17	1.740	2.110	2.898
18	1.734	2.101	2.878
19	1.729	2.093	2.861
20	1.725	2.086	2.845
30	1.697	2.042	2.750
40	1.684	2.021	2.704
60	1.671	2.000	2.660
120	1.658	1.980	2.617
∞	1.645	1.960	2.576

Example

Study of mobility by extension of the lumbar spine in individuals aged between 30 and 39 years

n=17, mean=40° şi t_α=2.11, s=2.36°

$$[40 - 2.11 * \sqrt{\frac{2.36}{17-1}}; 40 - 2.11 * \sqrt{\frac{2.36}{17-1}}]$$

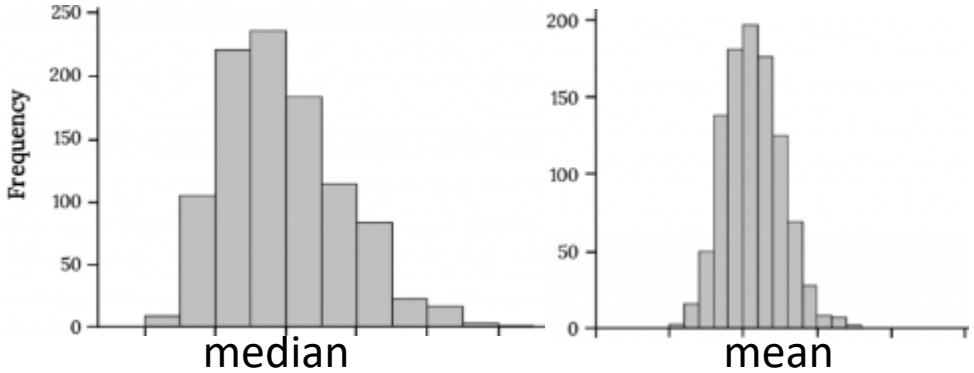
$$[40 - 1.24; 40 + 1.24]$$

[38.76°; 41.24°]

Answer: mobility of the lumbar spine in young people is between 38.76° and 41.24° with a 5% error

Quality of the estimator

- to be without errors
  - sometimes corrections are needed for a better estimate
  - e.g. if we use the standard deviation s as an estimator of σ we need to divide by n-1 instead of n otherwise the estimate produces a value too small
- have high stability (the variance of the sampling distribution should not be large)
  - the median is not a stable estimator compared to the mean
  - if possible, we will use the mean as the estimator, not the median



Kindred Grey via Virginia Tech (2021). "Sampling Distributions of the Sample Mean from a Non-Normal Population." CC BY-SA 4.0. Retrieved from https://commons.wikimedia.org/wiki/File:Sampling_Distributions_of_the_Sample_Mean_from_a_Non-Normal_Population.png

The precision of the estimation of a population parameter

depend on

- the estimator
  - ex. mean is more precise then median
- the sample size
  - law of big numbers – bigger sample size – bigger precision
- variation of the data
  - less variation – bigger precision
- needed precision
  - bigger precision – wider interval

Possible questions at the theoretical exam

*The average of tooth decay of a random sample of 101 clients of a dental office was 10 and standard deviation s=3 (Z_{α}=1.96, α=0.05). Compute the 95% confidence interval for the mean.

A. 10.6

B. 9.41-10.59

C. 9.4

D. 9.39-10.51

E. 9.5-10.7

Possible questions at the theoretical exam

*The average of tooth decay of a random sample of 101 clients of a dental office was 10 and standard deviation s=3 (Z_{α}=1.96, α=0.05). Compute the 95% confidence interval for the mean.

A. 10.6

B. 9.41-10.59

C. 9.4

D. 9.39-10.51

E. 9.5-10.7

Possible questions at the theoretical exam

*The frequency of tooth gangrene in a random sample of 100 subjects was 50 (Z_α=1.96, α=0.05). Compute the 95% confidence interval for the frequency of tooth gangrene.

A. 49.41-50.59

B. 0.402-0.598

C. 0.200-0.800

D. 45-55

E. It cannot be computed from these data

Possible questions at the theoretical exam

*The frequency of tooth gangrene in a random sample of 100 subjects was 50 (Z_α=1.96, α=0.05). Compute the 95% confidence interval for the frequency of tooth gangrene.

A. 49.41-50.59

B. 0.402-0.598

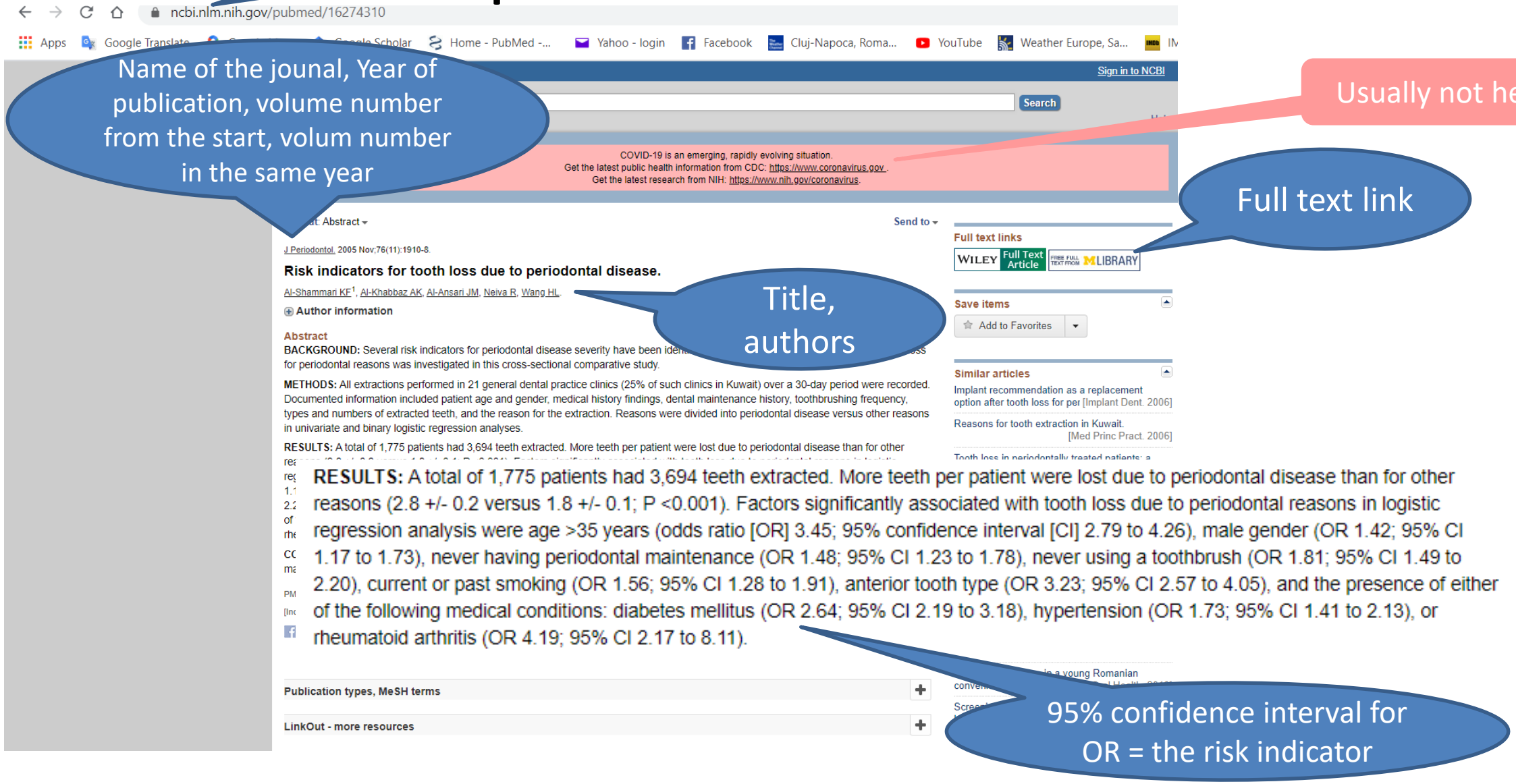
C. 0.200-0.800

D. 45-55

E. It cannot be computed from these data

Example from scientific literature

Address of PubMed- the greatest data base of medical scientific articles index



Name of the jounal, Year of publication, volume number from the start, volum number in the same year

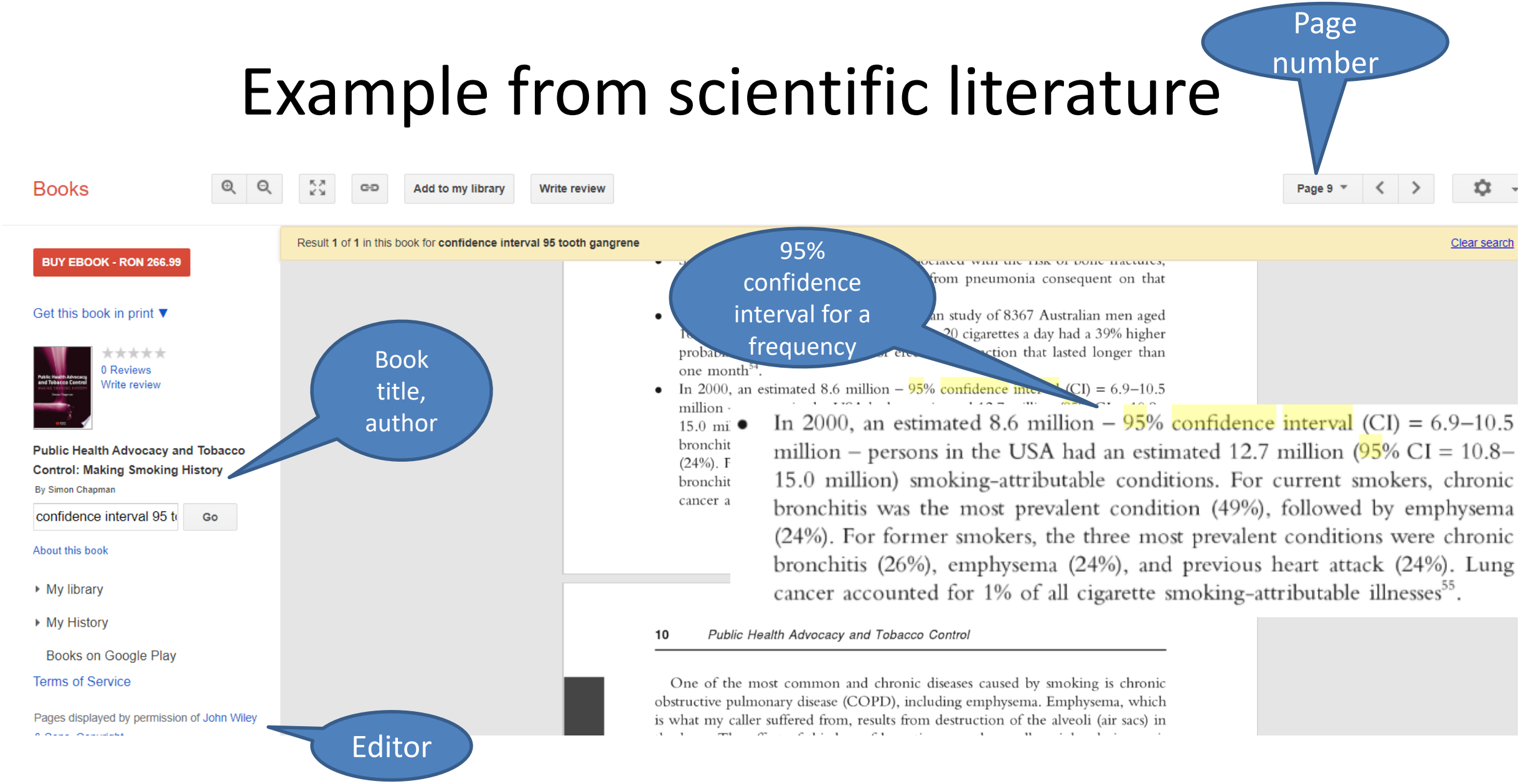
Usually not here!!!

Full text link

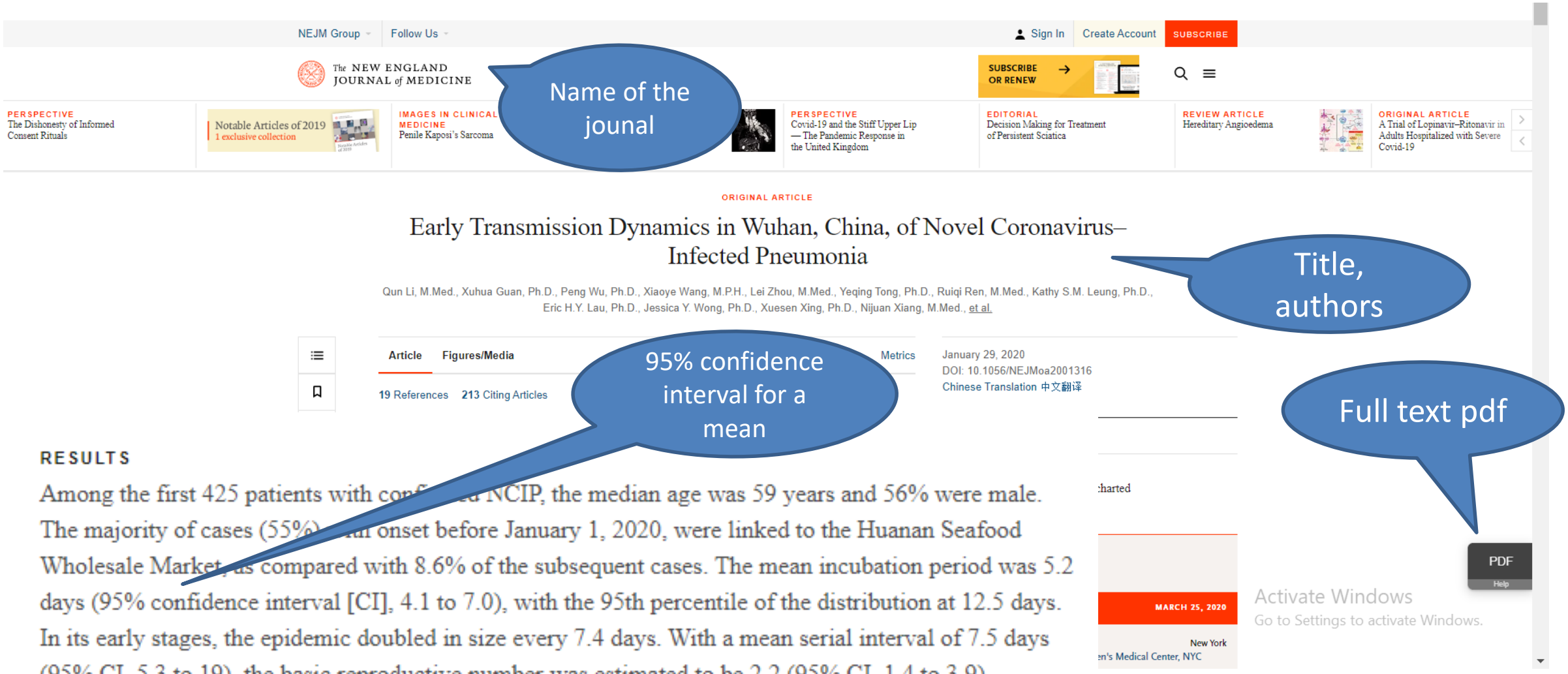
Title, authors

95% confidence interval for OR = the risk indicator

Example from scientific literature



Example from scientific literature



- Thank you !!!