



Trainer: Bondor Cosmina-Ioana, PhD, MsC

Descriptive statistics

Quantitative Variables

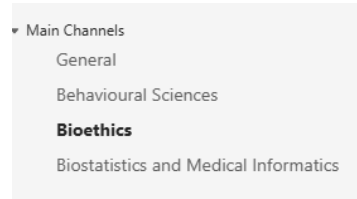
A ALWAYS

S SEEK

K KNOWLEDGE

Homework 1

- Download the requested file
 - Teams – channel Biostatistics and Medical Informatics - Files
- Resolve the exercises
- Complete the Form
 - Teams
- Deadline **3th November 2025**
 - if you made Homework 1 in time and obtain a mark higher or equal with 6 you will receive +0.2 for the exam mark



Biostatistics and Medical Informatics

Name	Modified
Homework_1_-_for_0.2pct_at_theoretical_e...	3 minutes ago

A screenshot of a Microsoft Teams post in the 'Biostatistics and Medical Informatics' channel. The post is from 'BONDOR COSMINA IOANA' and is titled 'Homework 1'. The text of the post reads: 'The Homework 1 file has been published in the Files section. Please download and print it. Complete all the exercises on the printed paper by hand. Once you finish, respond to the following quiz, which will consist of questions related to the exercises from Homework 1. Deadline to fill the quiz 11 November 24:00. If you obtain 6 at the quiz you will receive 0.2 points to the exam mark: https://forms.office.com/Pages/ResponsePage.aspx?id=Ja-p_IQ7bEyXNp1c3OIQxaw2FzieRu5DIi7g-wRMP2xUM0ZJSTNDU0tDOUM3R0hCSFgzVzdst1NEWC4u'. Below the text is a Microsoft Forms card with the text 'A post on Microsoft Forms provided by: forms.office.com' and the forms.office.com logo.

Summarizing Qualitative Variables

Organizing and presenting one qualitative variable



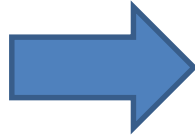
Variable type	Indicators	Charts
Qualitative		
• Nominal	Absolute and/or relative frequencies	 Pie or Column or Bar
• Dichotomous	Absolute and/or relative frequencies	 Pie or Column or Bar
• Ordinal		
– with text categories	Absolute and/or relative frequencies	 Column or Bar
– scores		
• with few categories ≤ 10	Absolute and/or relative frequencies minimum, maximum, median, quartiles, range, interquartile range, mode	 Column or Bar
• with many categories > 10	Absolute and/or relative frequencies minimum, maximum, median, quartiles, range, interquartile range, mode Absolute and/or relative frequencies after dividing the scores on intervals	<p>Column or Bar with the scores divided into intervals</p>

Summarizing Qualitative Variables

Organizing and presenting **two** qualitative variable

Two variables - both

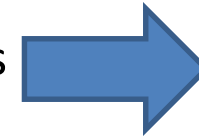
- **dichotomial /**
- **nominal /**
- **ordinal** (with text as categories)



Statistics

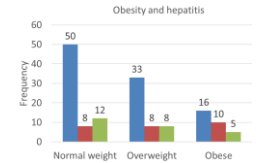
contingency table

- joint absolute frequencies
- frequencies relative to
 - totals on rows
 - totals on lines
 - grand total

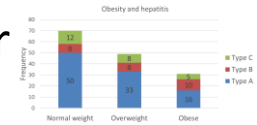


Chart

– Column/Bar



– Staked Column/Bar



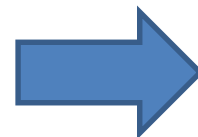
– 100% Staked Column/Bar



Two variables:

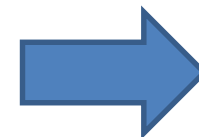
• **one**

- **dichotomial /**
- **nominal /**
- **ordinal** (with text as categories)

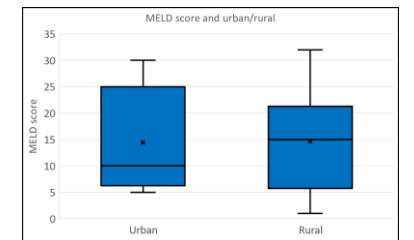


On each subgroup

- Range
- Median
- Quartile 0-4
- Interquartile range
- Mode



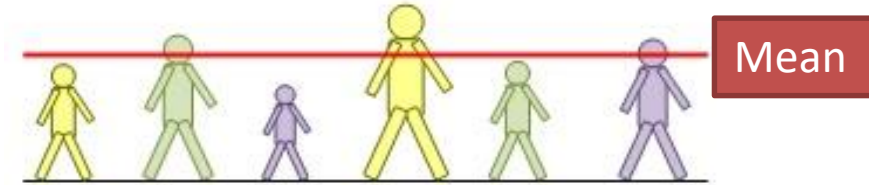
– Box-Whiskers Plot



Objectives

- Summarizing Quantitative (Numerical) Data (Continuous, Discrete)
 - Descriptive statistics
 - Charts

Measure of central tendency



Arithmetic mean \bar{X} is the average of the observations.

Calculation: add the observations to obtain the sum and then divide by the number of observations.

The **formula** for the arithmetic mean:
$$\bar{X} = \frac{1}{n} \sum X$$

Σ - to add, X represents the individual observations, n is the number of observations.

Example

- Ex. In 2011, 6 medical doctors earn in a month 2400, 2600, 2900, 2900, 3200, 3400 Euro.

$$\bar{X} = \frac{2400+2600+2900+2900+3200+3400}{6} = 2900 \text{ Euro}$$

The average income is 2900 Euro

Arithmetic mean is a measure of the middle

Median is a measure of the middle

Mean or median?



We can use **arithmetic mean** when data are **numerical without outliers**

Median can be use any time, but because the arithmetic mean is easy to understand ...

We use **median** only when the data are numerical with outliers or data are ordinal

Medical doctors' income in 2011

2400€

2600 €



2900 €



2900 €

3200 €

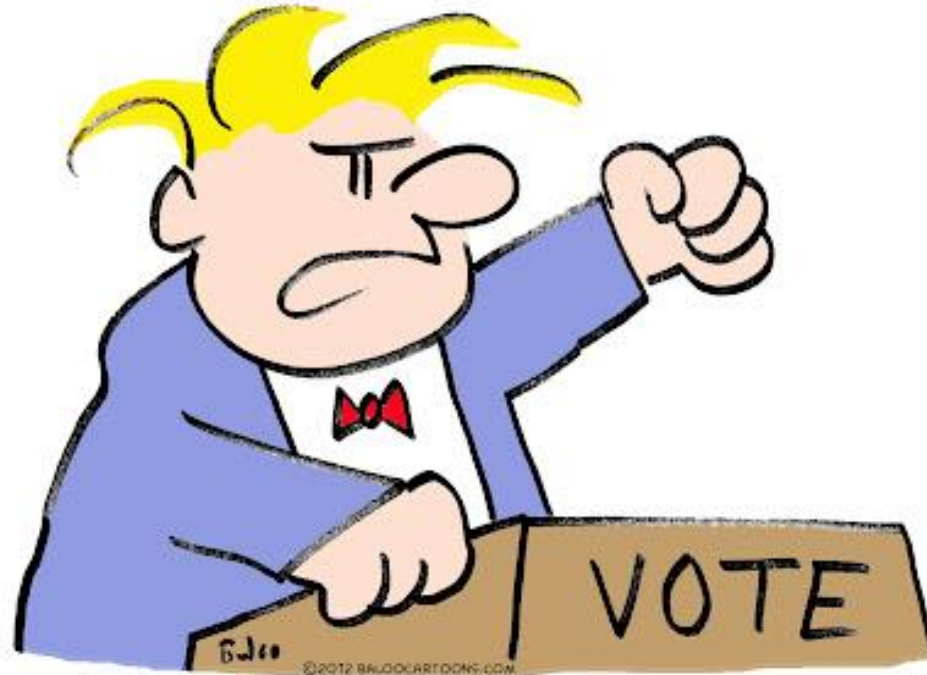
3400 €



Arithmetic mean = 2900 €

Average income of medical doctors is 2900 €

Some political changes



"Desperate times call for
desperate politicians!"

After 12 months politician announce that:

Medical doctors' income near **double** in 12 months

2400€ 2600 €



2900 € 2900 €



3200 €



19000 €



Arithmetic mean = **5500 €**

was 2900 €

- Conclusion: outliers **can** influence the mean

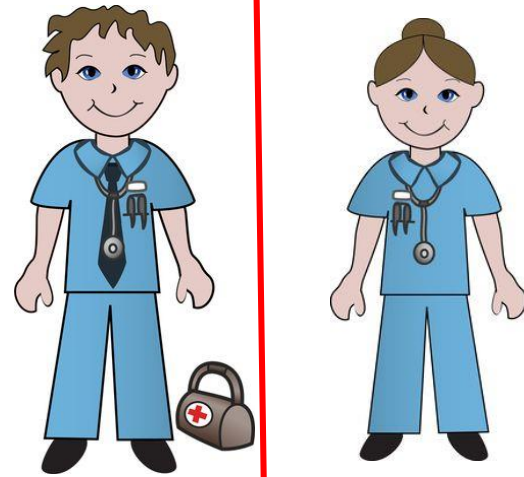
Outlier **19000 €**

Median

2400€ 2600 €



3200 € 3400 €



2900 € 2900 €

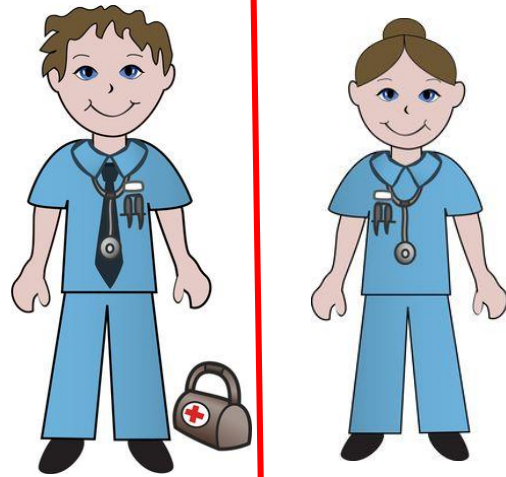
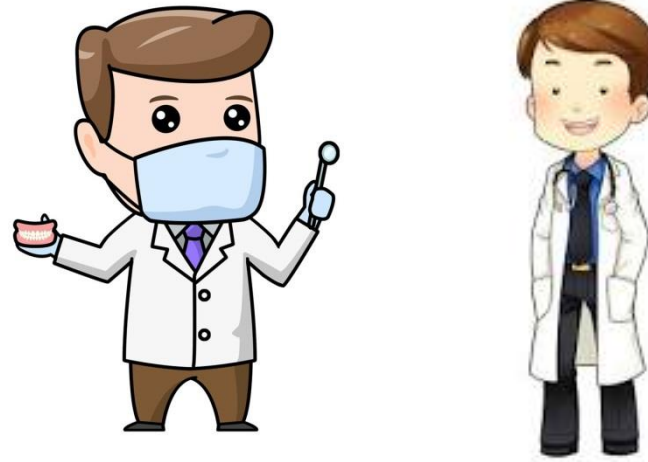
Median = 2900 €

Medical doctors income **after 12 months**

2400€ 2600 €



3200 € **19000 €**



2900 € 2900 €

Median = **2900 €**

was 2900 €

- Conclusion: outliers **cannot** influence the median

Outlier or extreme case

19000 €



Mean is a measure of the middle?

We use **mean** if data are **numerical without outliers**

We use **median** if data are **numerical with outliers** or **data are ordinal**



Summarizing Numerical Data (Continuous, Discrete)

- Measure of central tendency
 - Arithmetic mean
 - Median
 - Mode
- Measure of spread
 - Variances
 - Standard Deviation
 - Coefficient of variation
 - Standard error
- Other measures
 - Skewness
 - Kurtosis
 - Quartile
 - Percentile
- Charts
 - Histograms
 - Box-plots



Objectives

- Measure of dispersion
- Examples

Measures of Spread

Series 1	Series 2	Series 3
1	1	1
1	44	8
2	45	11
3	46	14
5	48	28
6	48	30
6	49	37
7	50	48
93	50	52
94	51	62
94	52	70
95	52	72
97	54	84
98	55	91
98	55	92
100	100	100
Sum800	Sum800	Sum800

Why we need them when we want to describe data?

Number of observations in each set = 16

Average of each set = 50

Median of each set = 50

Series 1

1

1

2

3

5

6

6

7

93

94

94

95

97

98

98

100

The distribution

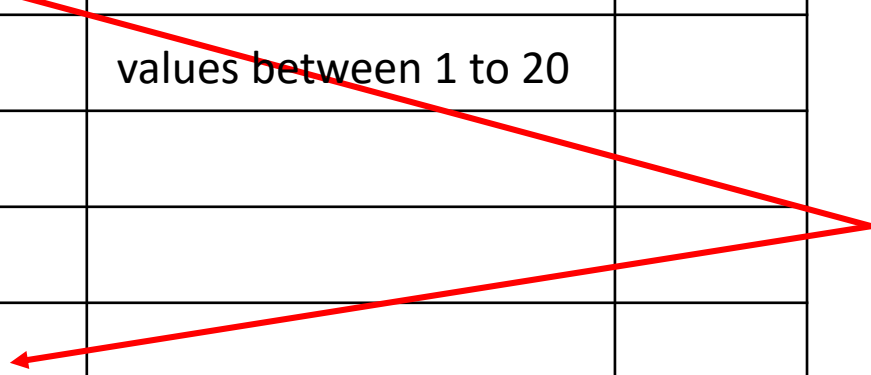
Frequency classes		
1-20	values between 1 to 20	
21-40		
41-60		
61-80		
81-100		
Total		

minimum = 1

maximum = 100

we divide the interval 1 - 100 in 5 equal subintervals

= 5 "classes" with the length of 20



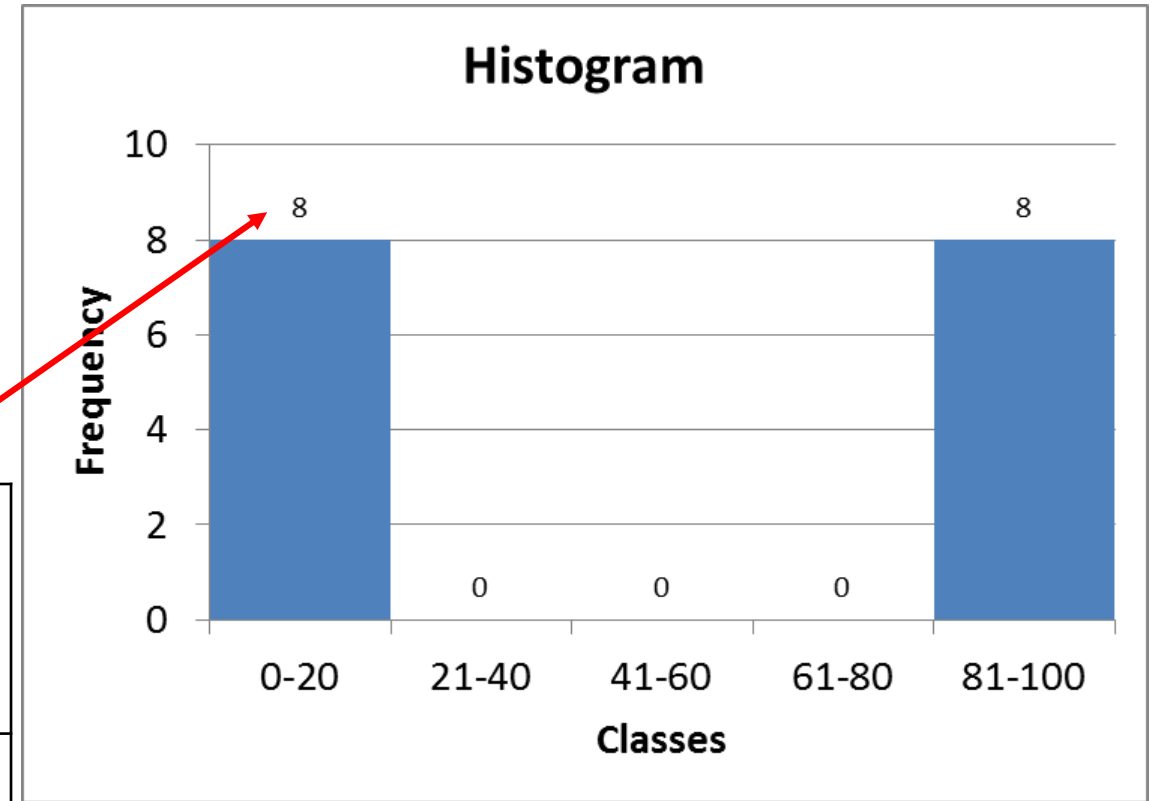
Series 1

1
1
2
3
5
6
6
7

The distribution

8 values between 0 to 20

Series 1 – Classes	Absolute Frequency	Relative Frequency %
0-20	8	50
21-40	0	0
41-60	0	0
61-80	0	0
81-100	8	50
Total	16	100



Series 1

93
94
94
95
97
98
98
100

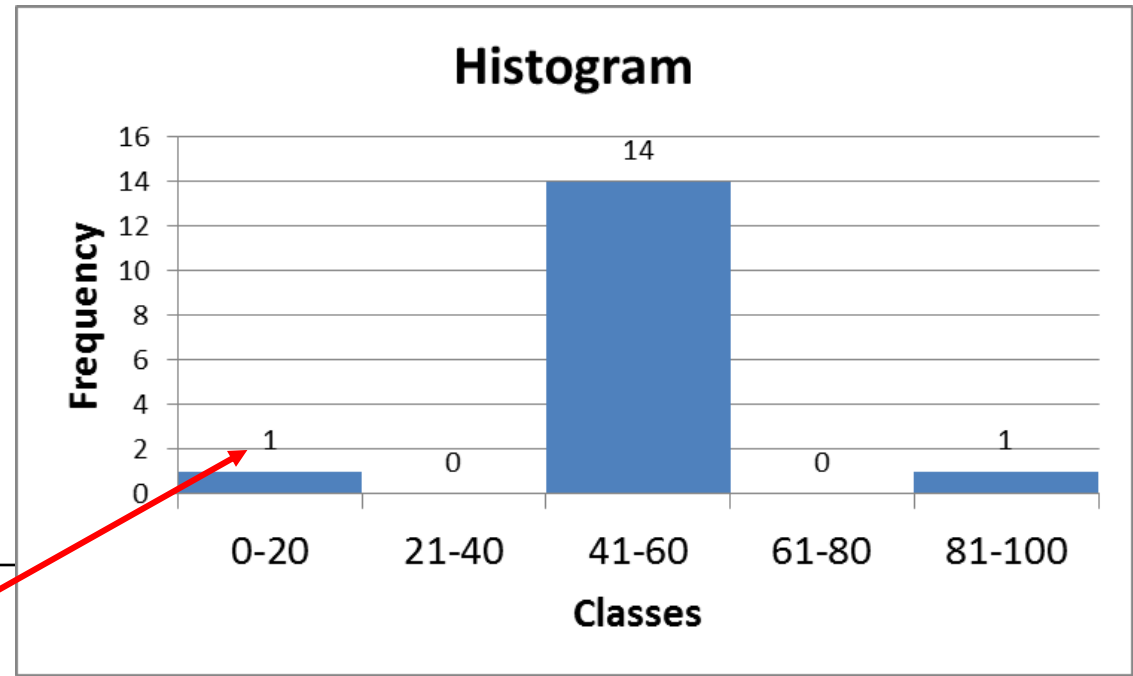
The distribution

Series 2

1
44 1 value between 0 to 20

45
46
48
48
49
50
50
51
52
52
54
55
55
100

Series 2 – Classes	Absolute Frequency	Relative Frequency %
0-20	1	6.25
21-40	0	0
41-60	14	87.50
61-80	0	0
81-100	1	6.25
Total	16	100



Series 2

Series 3

The distribution

1 }
11 } 2 values between 0 to 20

24

29

36

41

45

49

51

55

59

64

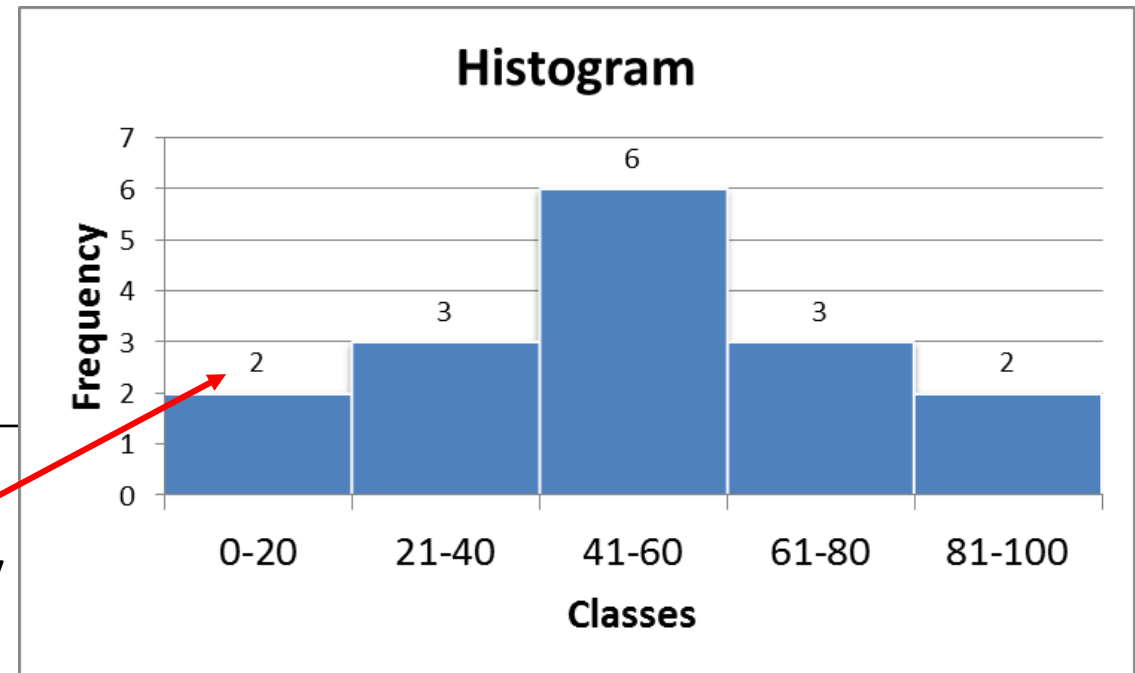
71

76

88

100

Series 3 – Classes	Absolute Frequency	Relative Frequency %
0-20	2	12.50
21-40	3	18.75
41-60	6	37.50
61-80	3	18.75
81-100	2	12.50
Total	16	100



Series 3

Histogram

- a graphical representation

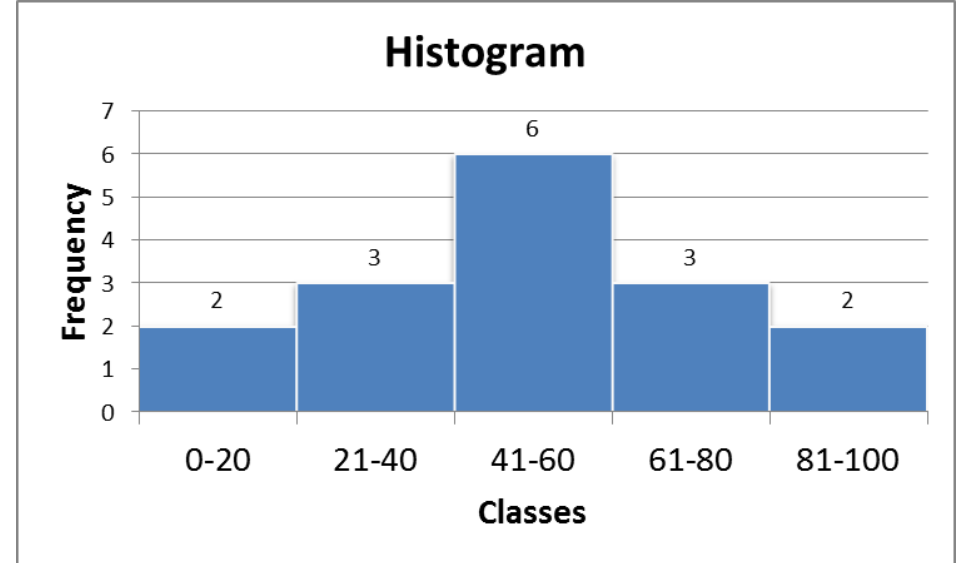
- the distribution of quantitative **continuous** variables

- bars

- show the frequency of values within ranges (bins)
- bar height = the number of values in that particular interval

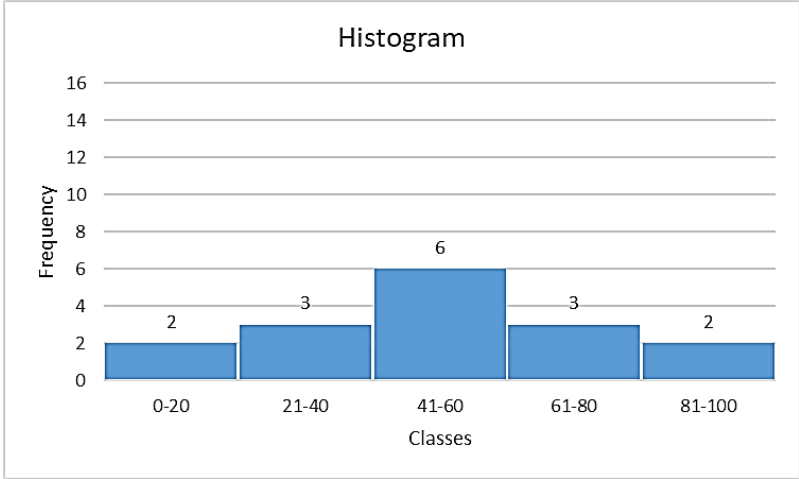
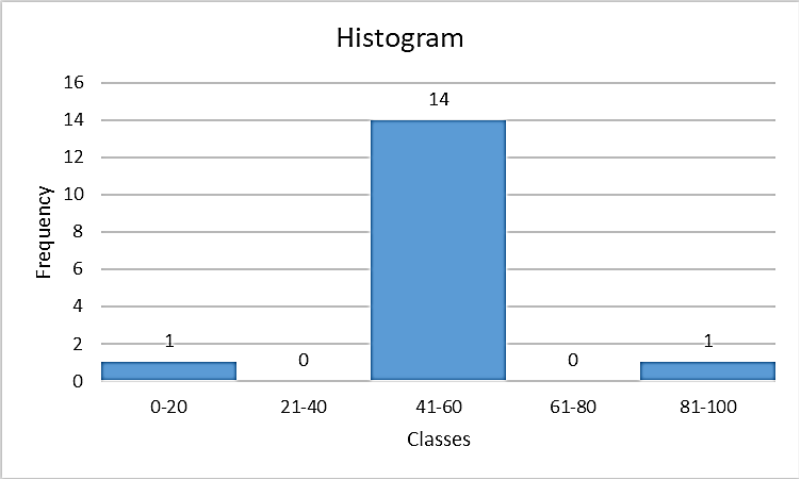
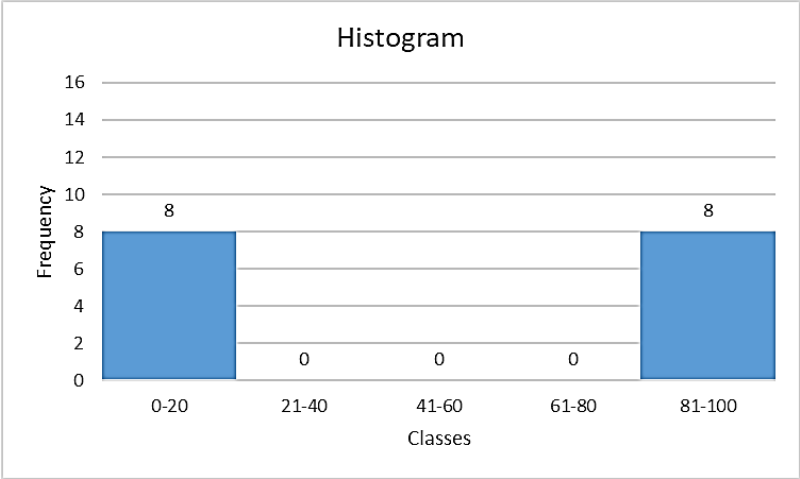
- **must have**

- **equal** bars width = equal intervals (to compare the frequencies)
- **no gap (space)** between the columns (intervals are adjacent and cover a continuous range of values)



Histogram

- useful to see the shape and the spread of data values



the same range on Oy
equal intervals on Ox



we can compare the distributions

Measures of Spread

The Standard Deviation – the average of the deviation (departure) from the mean

Formula for population (all observation):

$$\sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{N}}$$

Σ means "sum of,"

X = each value in the data set,

\bar{X} = mean of all values in the data set,

N = number of values in the data set.

Formula for subset of population (sample)

- The Standard Deviation of a **sample** (with correction for medical research)

$$S = \sqrt{\frac{\sum (X - \bar{X})^2}{N-1}} = \sqrt{\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_N - \bar{X})^2}{N-1}}$$

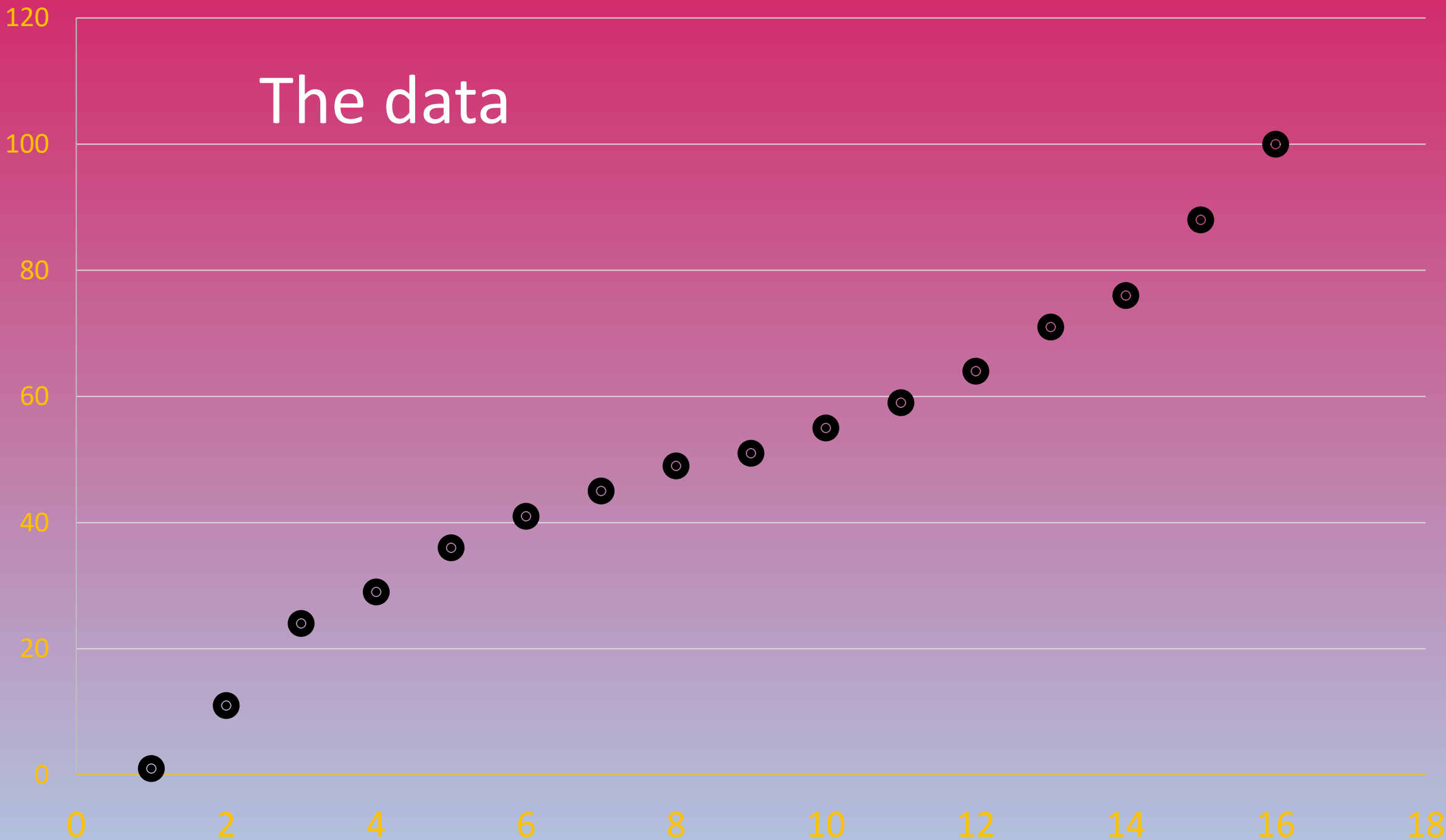
Where

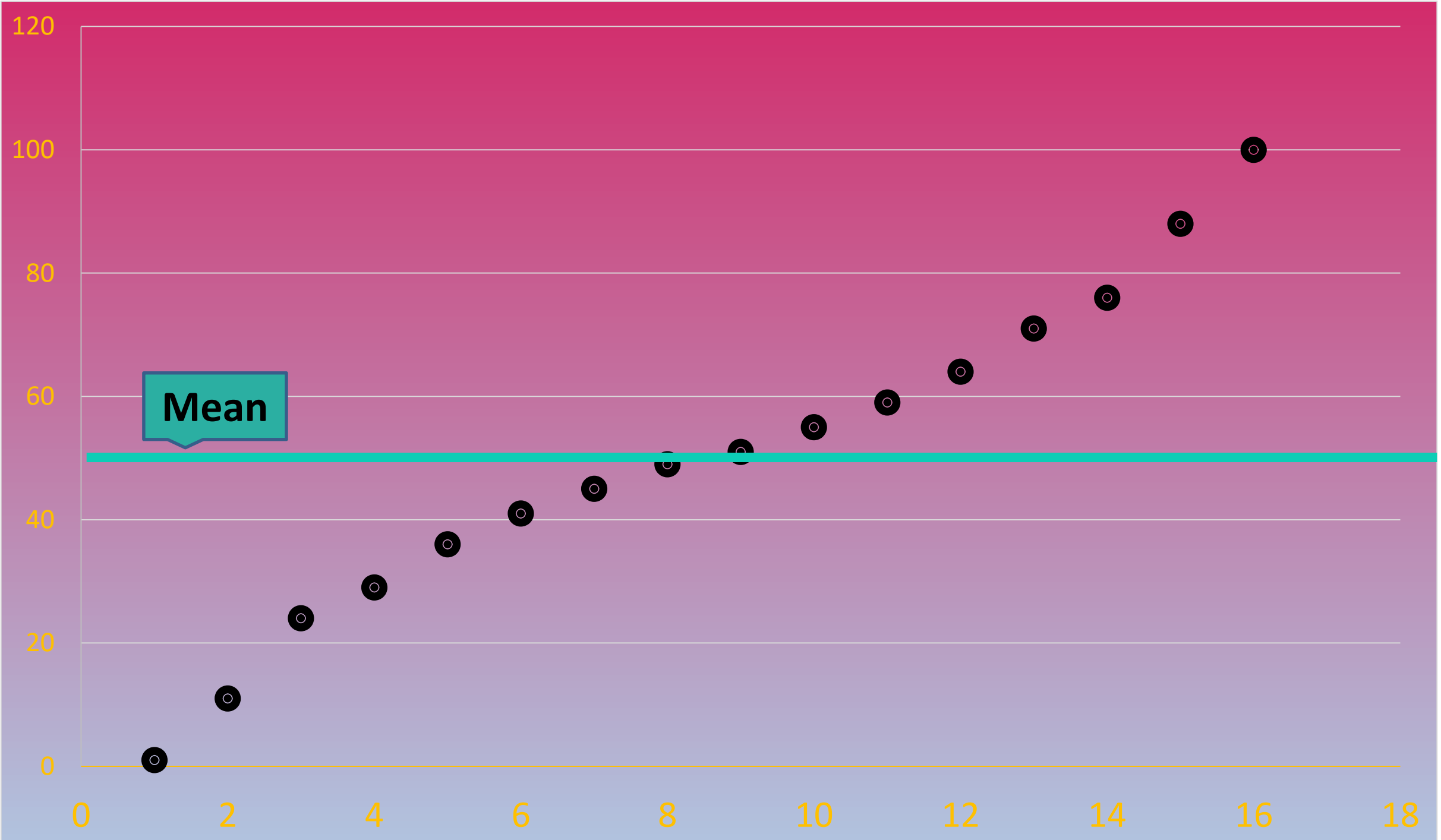
N - total number of observation

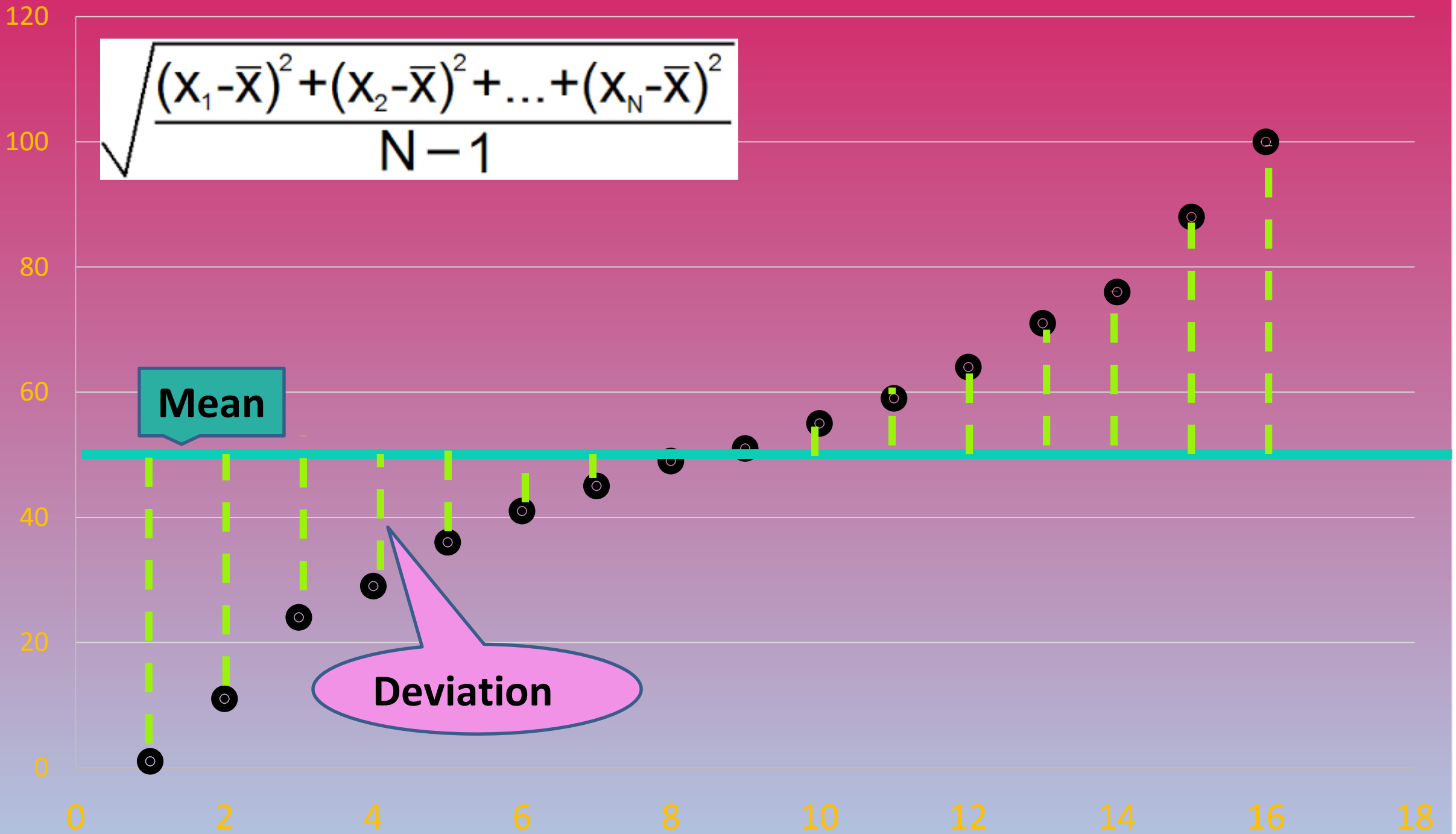
\bar{X} - arithmetic mean

X_1, \dots, X_N - observation

The data







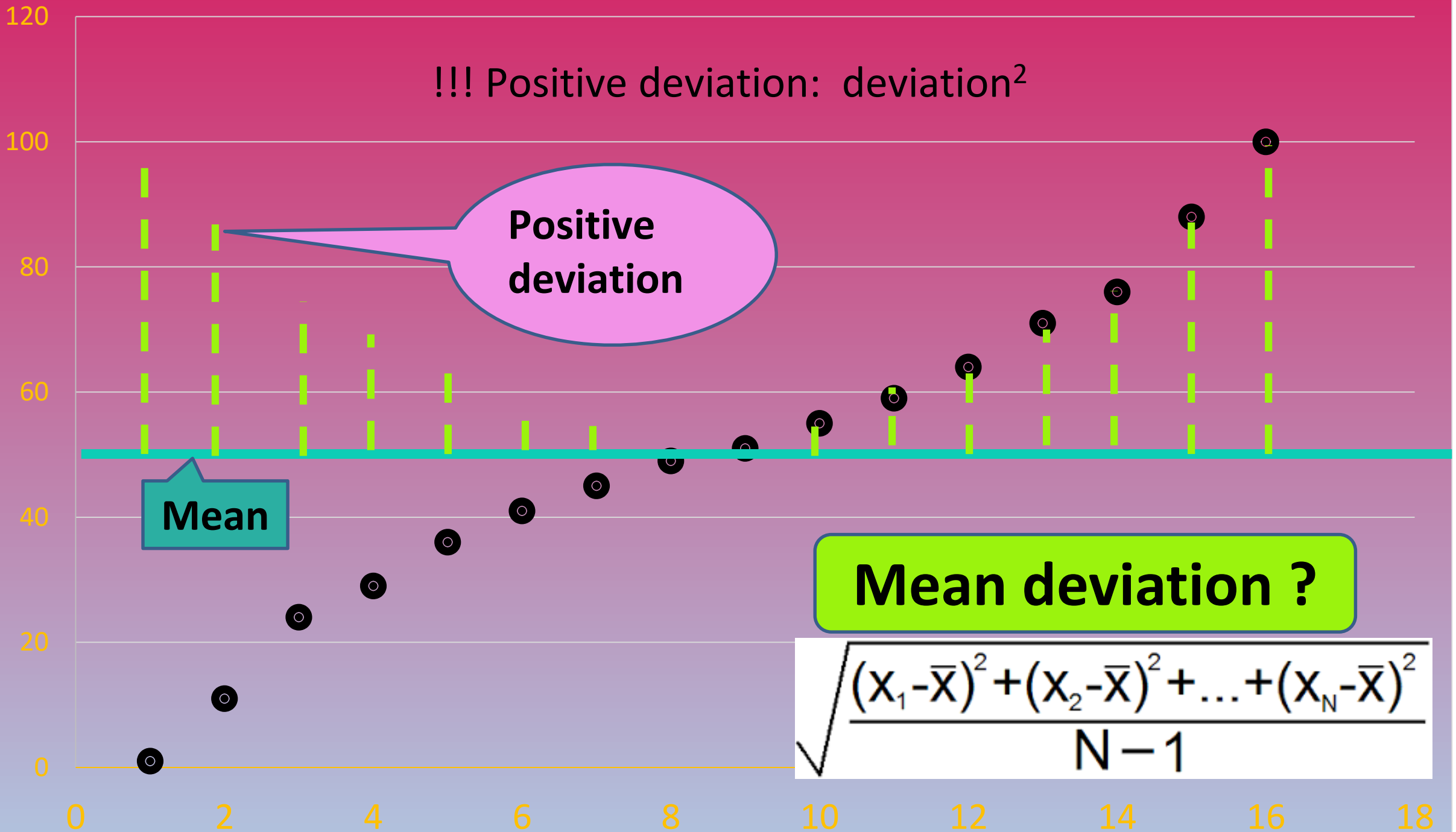
!!! Positive deviation: deviation²

Positive deviation

Mean

Mean deviation ?

$$\sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_N - \bar{x})^2}{N - 1}}$$





Mean deviation

$$\sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_N - \bar{x})^2}{N - 1}}$$

Actually, instead of lines are squares (deviation²)



Example

- No. of tooth for 6 patients: 20, 24, 26, 25, 25, 30
- Arithmetic mean $\bar{X} = (20+24+26+25+25+30)/6=25$
- Standard deviation:

$$\begin{aligned} S &= \sqrt{\frac{(20-25)^2 + (24-25)^2 + (26-25)^2 + (25-25)^2 + (25-25)^2 + (30-25)^2}{6-1}} = \\ &= \sqrt{\frac{(-5)^2 + (-1)^2 + 1^2 + 0^2 + 0^2 + 5^2}{5}} = \sqrt{\frac{25+1+1+25}{5}} = \sqrt{\frac{52}{5}} = \sqrt{10.4} = 3.22 \end{aligned}$$

10.4 is the variation

Mean deviation from the mean :-P

Standard Deviation - example

Ex. In a hepatitis A epidemic, a sample of 6 persons had clinical symptoms after 24, 25, 29, 29, 30, 31 days.

$X = 24, 25, 29, 29, 30, 31$ - the individual data, $\bar{X} = 28$

1. $X - \bar{X} = -4, -3, 1, 1, 2, 3$ (each data minus arithmetic mean)
2. $(X - \bar{X})^2 = 16, 9, 1, 1, 4, 9$ (square)
3. $\Sigma(X - \bar{X})^2 = 16+9+1+1+4+9 = 40$ (sum)
4. $\Sigma(X - \bar{X})^2 / (n-1) = 40/5 = 8$ (divide by $6-1=5$) = variance
5. $\sqrt{\Sigma(X - \bar{X})^2 / (n-1)} = 2.83$ This value is the standard deviation

Standard Deviation - example

Cholesterol values are given for 10 patients: 200, 180, 140, 160, 180, 150, 170, 110, 230, 170

	Cholesterol (X)	$X - \bar{X}$	$(X - \bar{X})^2$
	200	31	961
	180	11	121
	140	-29	841
	160	-9	81
	180	11	121
	150	-19	361
	170	1	1
	110	-59	3481
	230	61	3721
	170	1	1
Sum	1690	0	9690
\bar{X}	169	$(X - \bar{X})^2 / (n - 1)$	1076.6
		St.Dev.	32.8

**Deviation
from mean**

**Positive
deviation
from mean**

Characteristics of the study population

Fifty-three T2DM patients with normo or microalbuminuria were included in the study. Their median eGFR of 89.30 (67.86-99.50) mL/min was near the normal range. Clinical and biological characteristics of the studied subject are presented in [Table 1](#).

Table 1.

Descriptive clinical and biological characteristics of the studied subjects

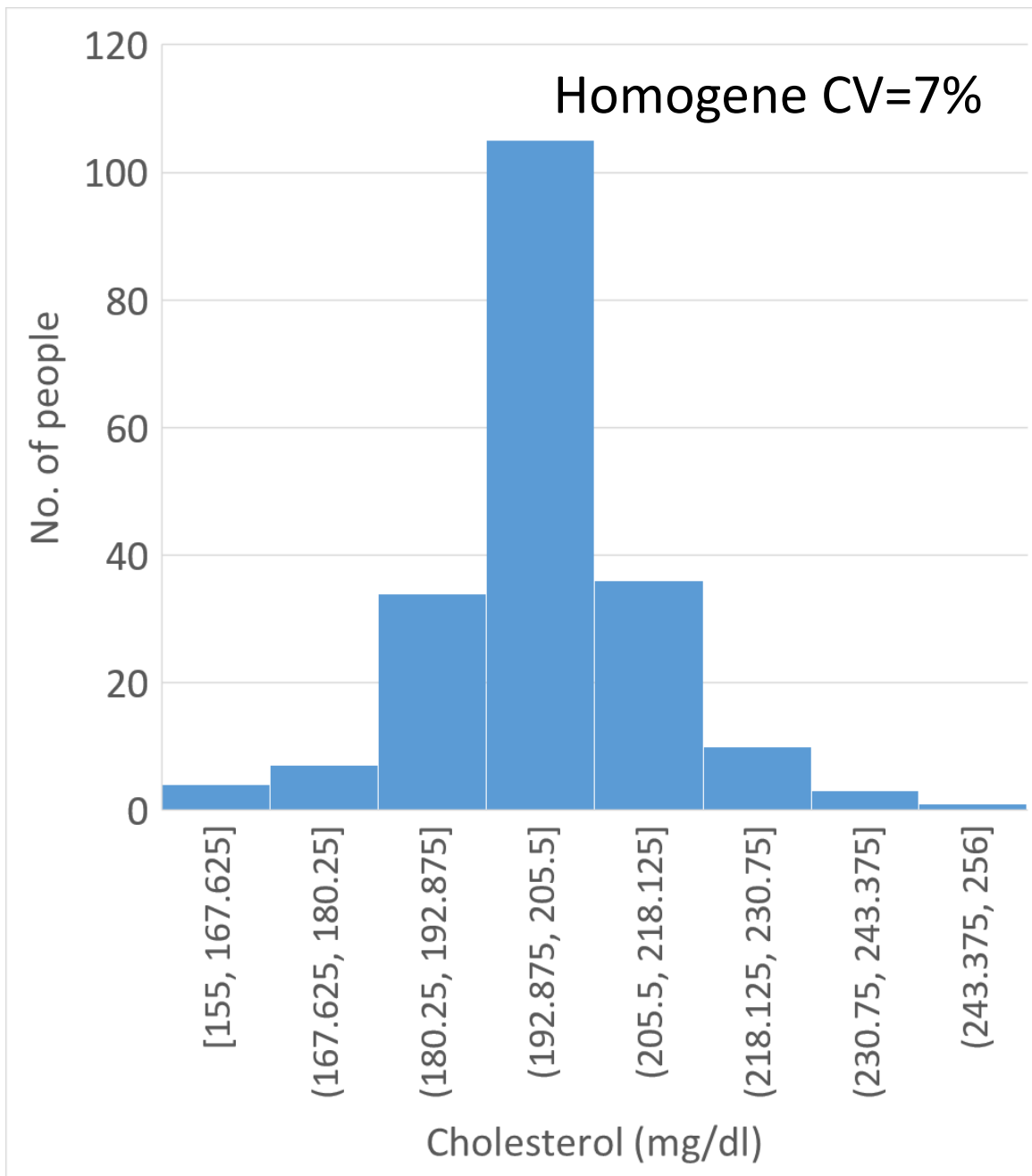
Parameters	Patients (n=53)
Male n (%)	34 (64.20)
Age (years)	64.98±10.62
Diabetes duration (years)	8.00 (6.00-15.00)

Arithmetic mean,
standard deviation

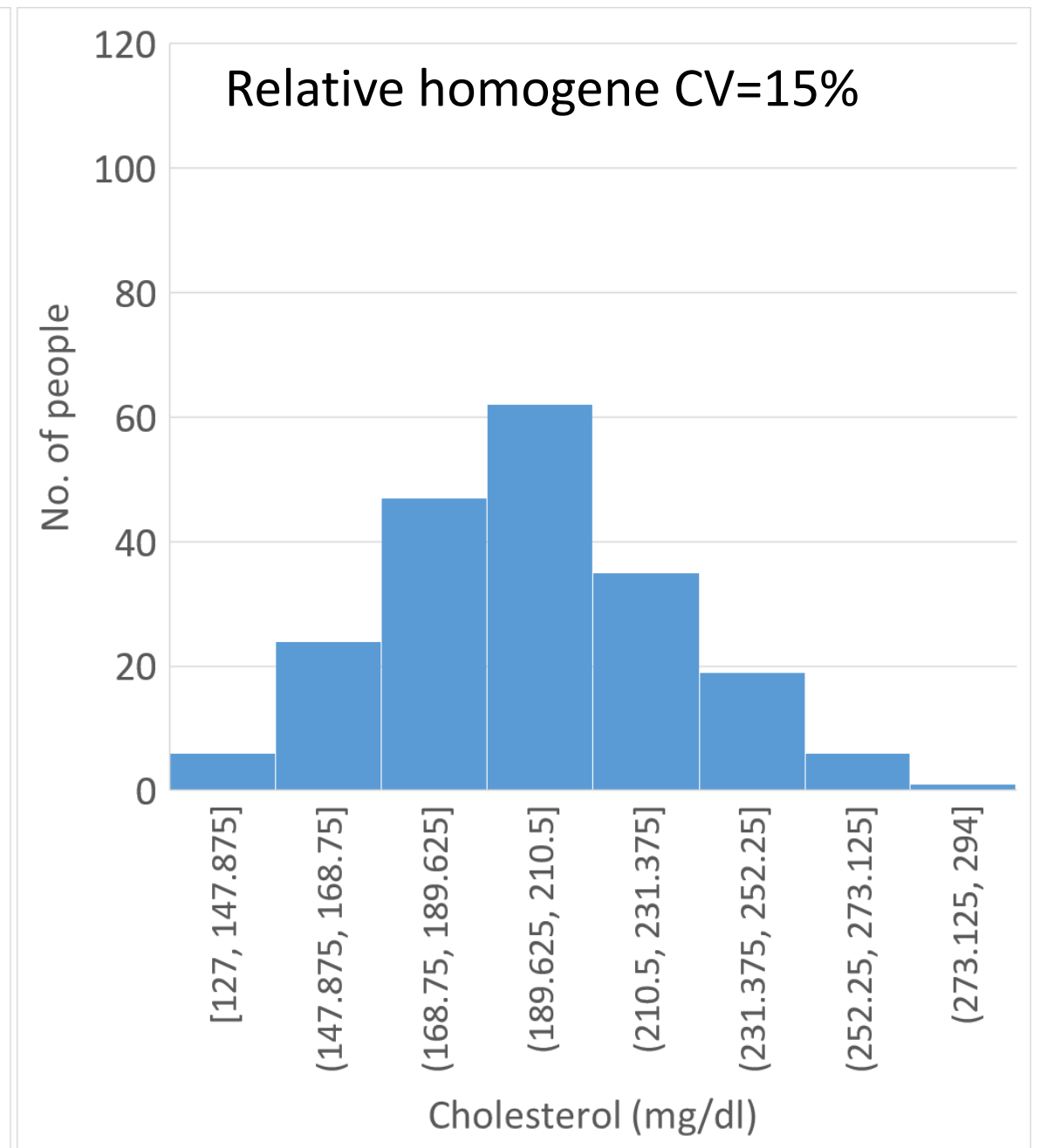
median, IQR

Other measures

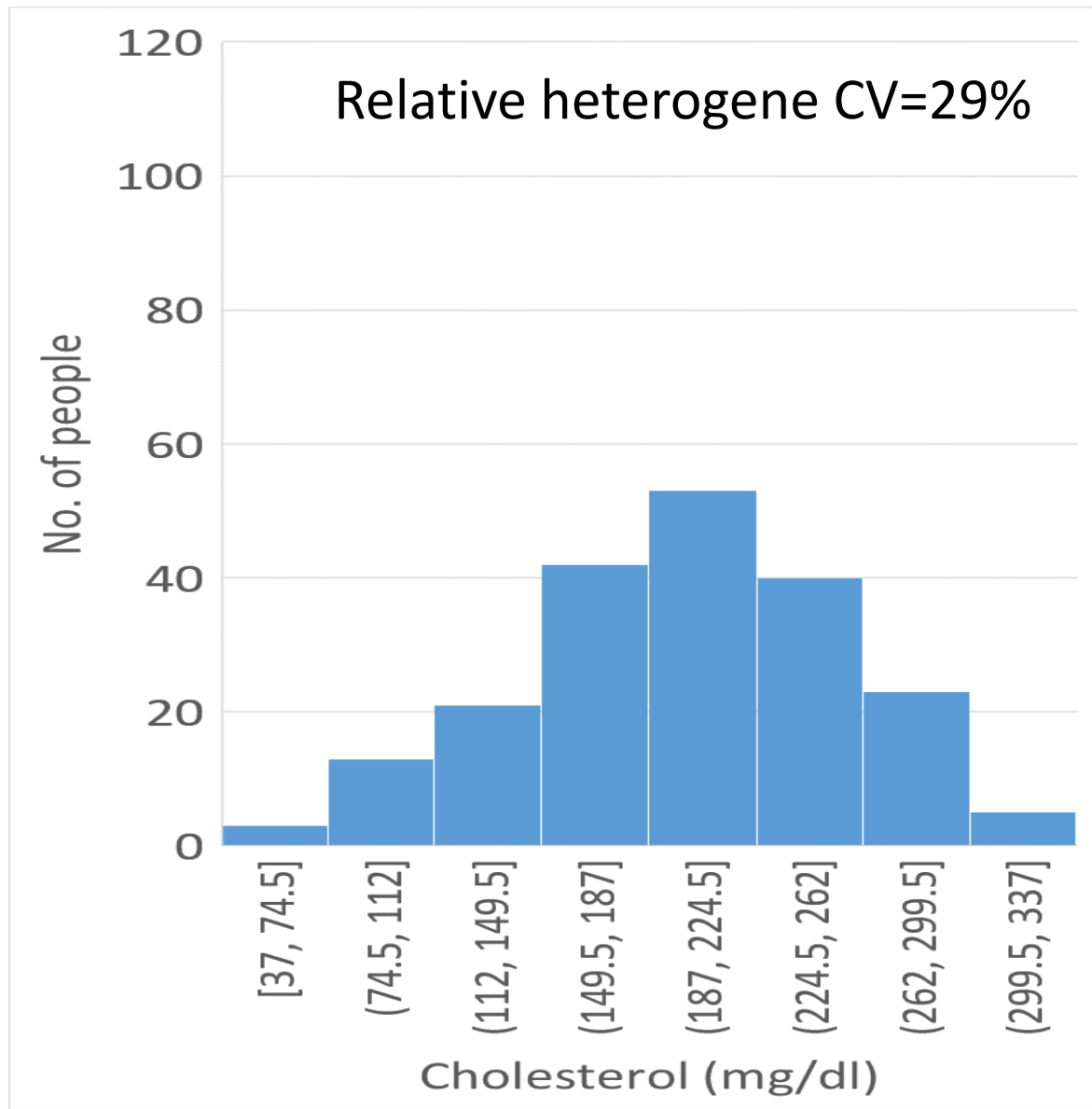
- **The coefficient of variation ($CV = s/\bar{X} * 100$)**
 - interpretation
 - $0\% \leq CV < 10\%$ data are homogeneous
 - $10\% \leq CV < 20\%$ data are relative homogeneous
 - $20\% \leq CV < 30\%$ data are relative heterogeneous
 - $30\% \leq CV$ data are heterogeneous
- where s =standard deviation, \bar{X} =arithmetic mean



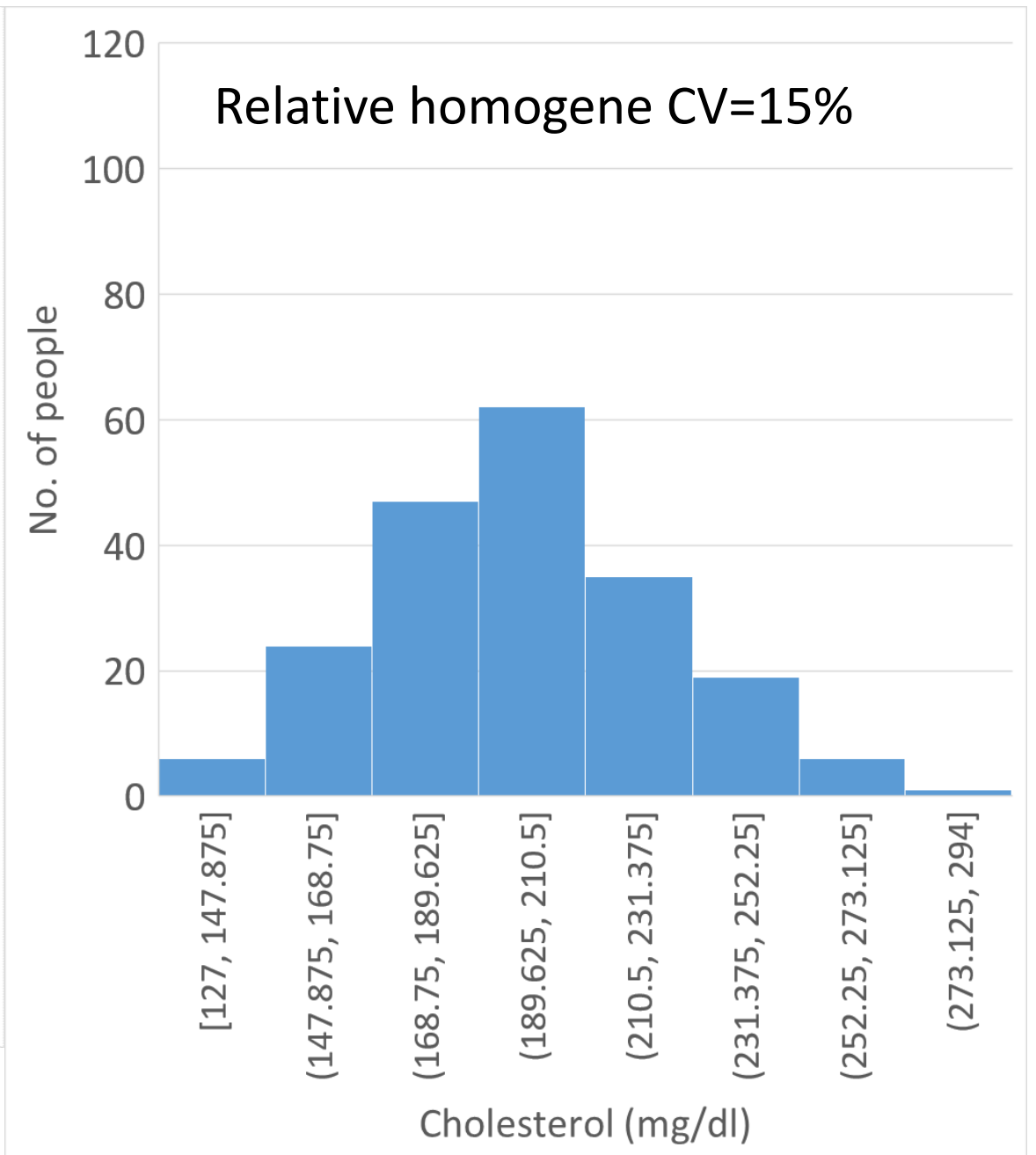
100 patients



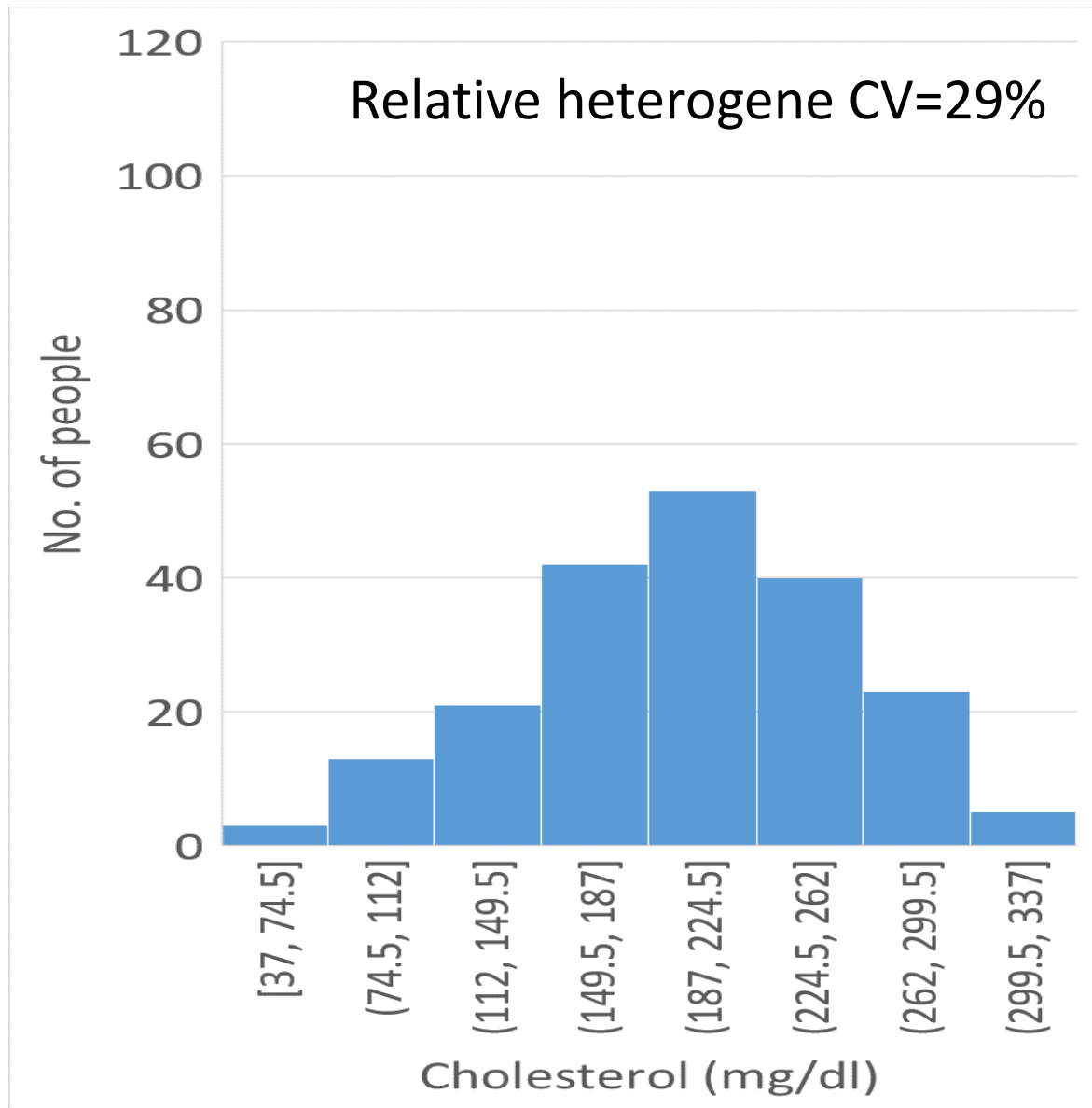
100 patients



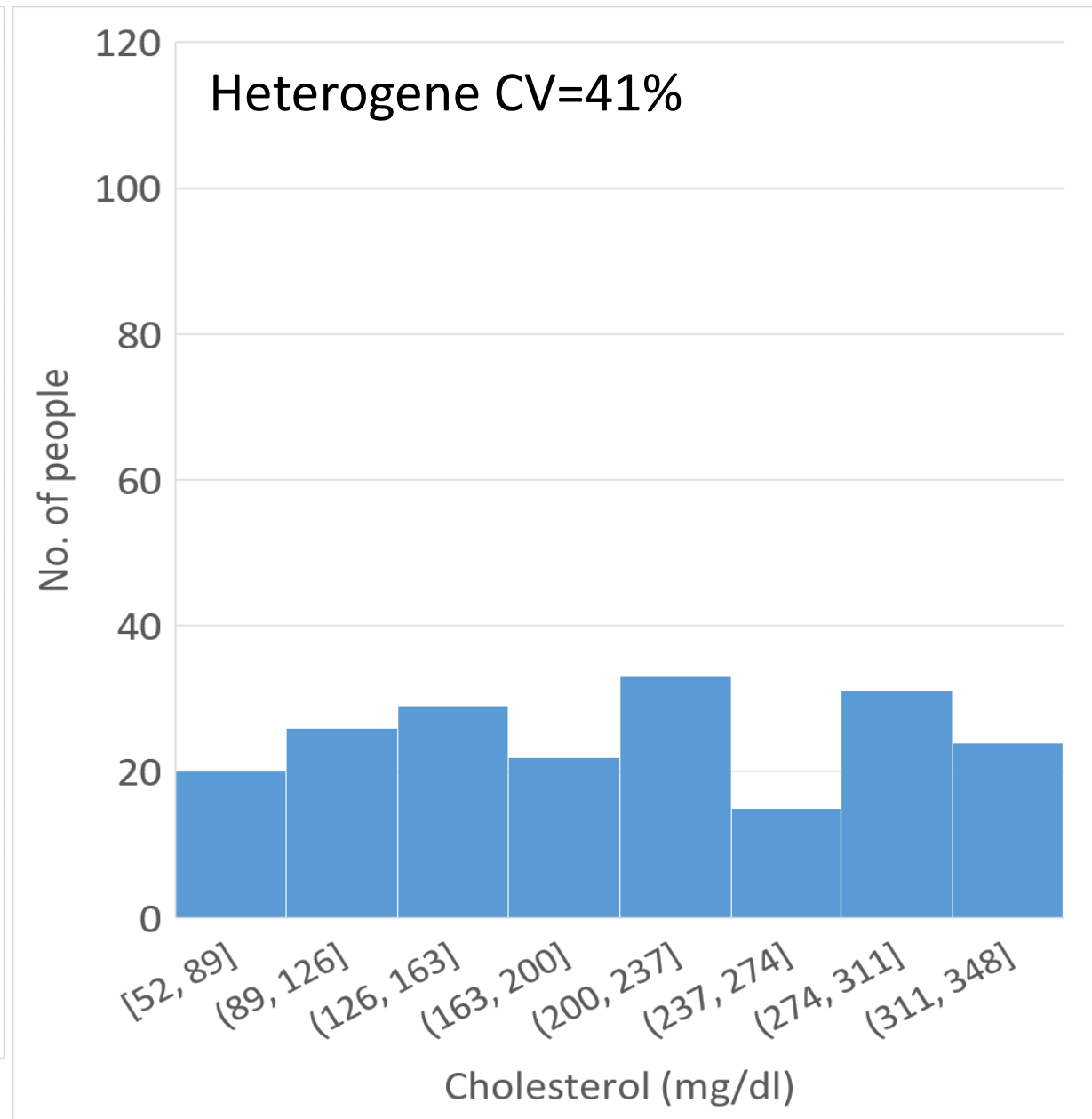
100 patients



100 patients

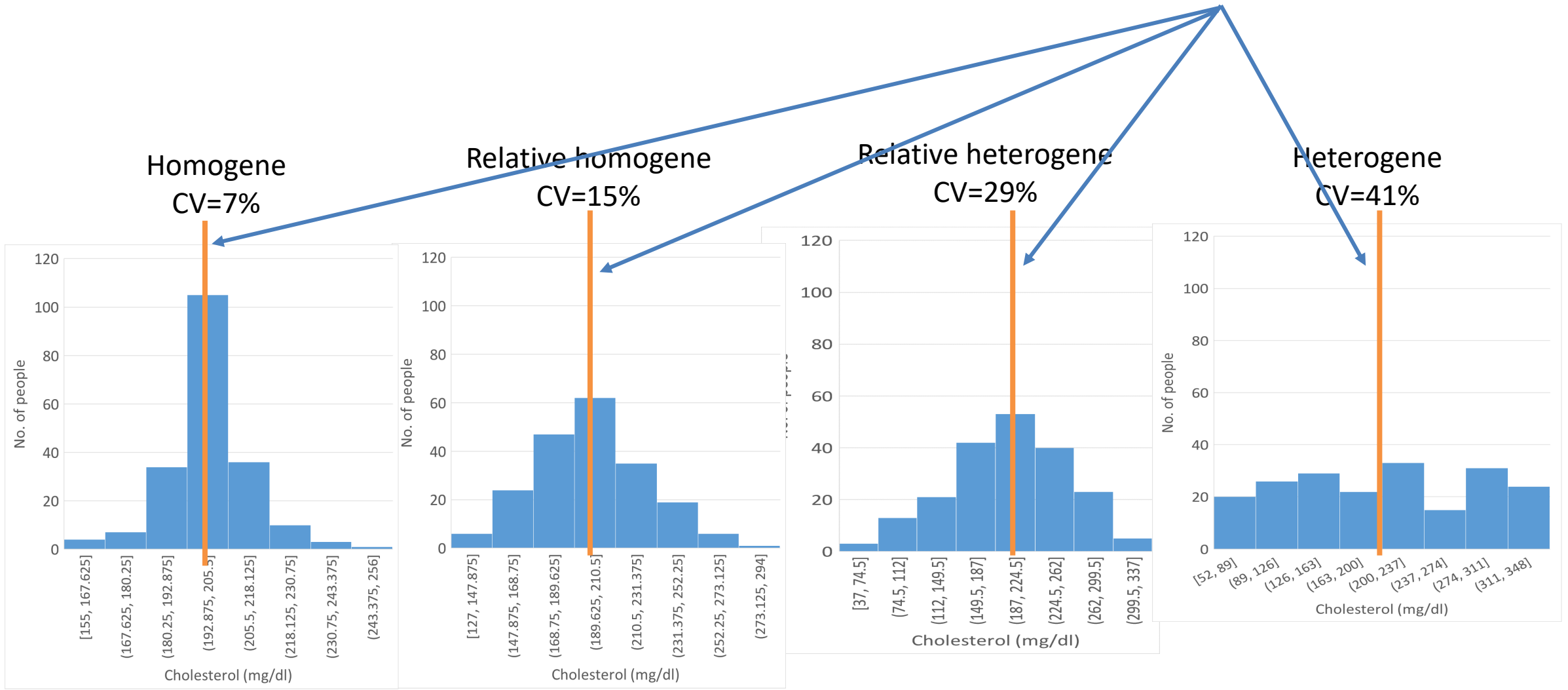


100 patients



100 patients

homogene = values are in vicinity of the arithmetic mean
heterogene = values are not in vicinity of the arithmetic mean



100 patients

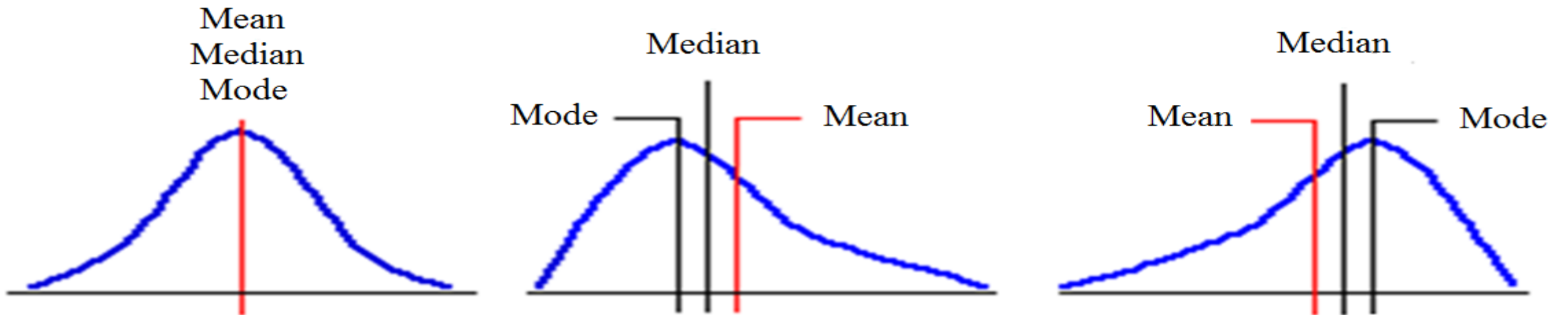
100 patients

Other measures

- **The standard error** ($E_s = s/\sqrt{n}$)
- where s =standard deviation, n =number of data

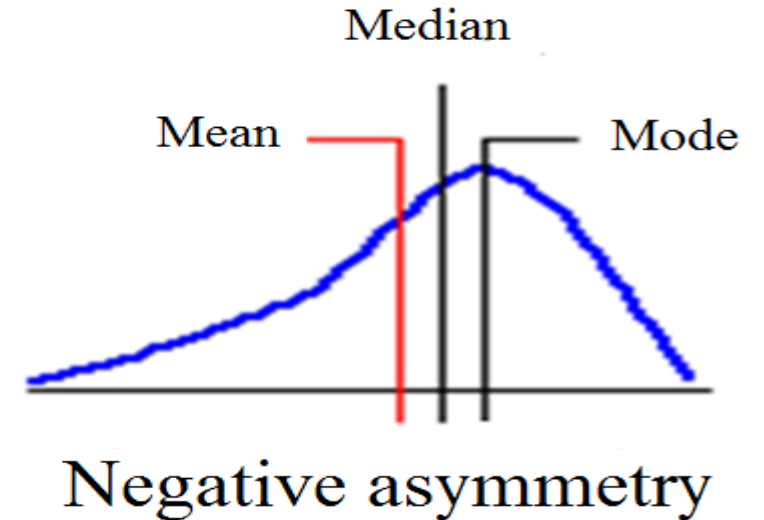
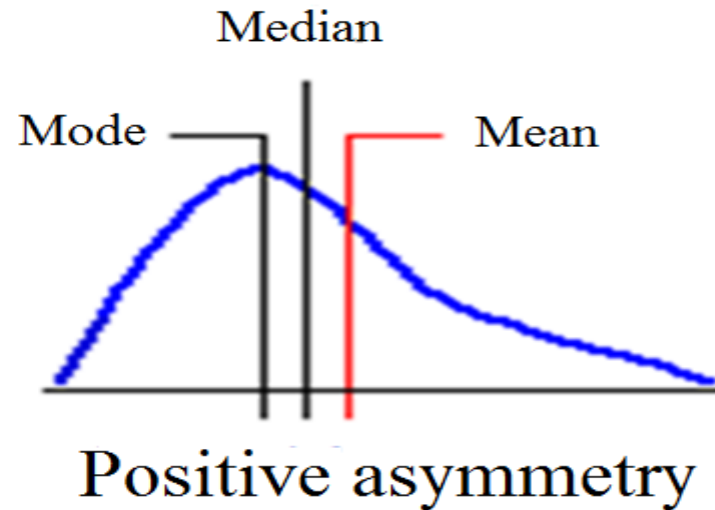
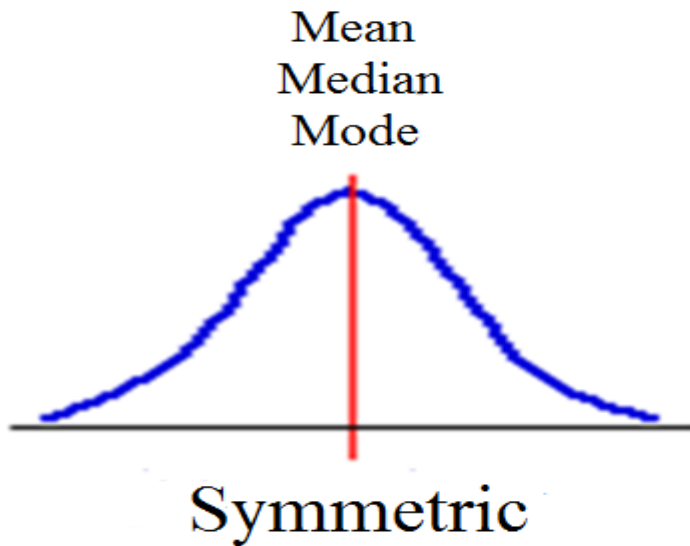
Measure of centrality - applicability

How to interpret the mean, median and mode:



Measure of centrality - applicability

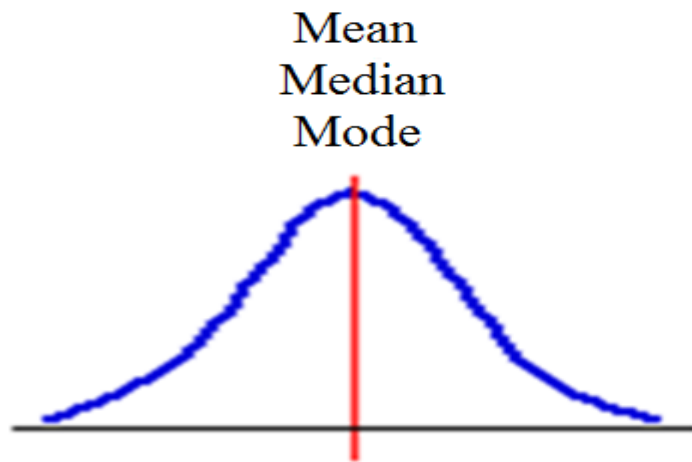
How to interpret the mean, median and mode:



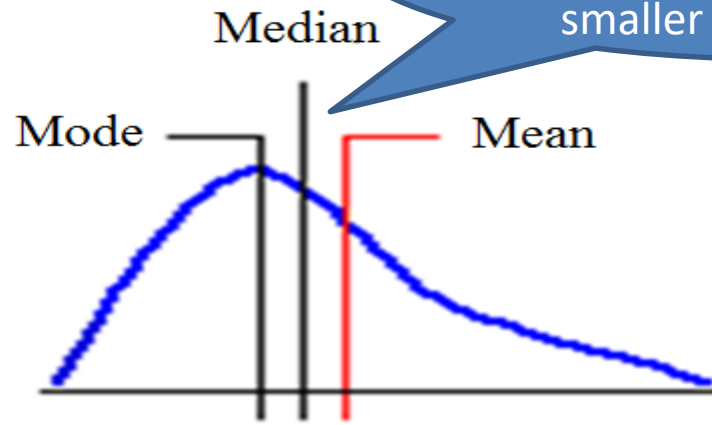
Measure of centrality - applicability



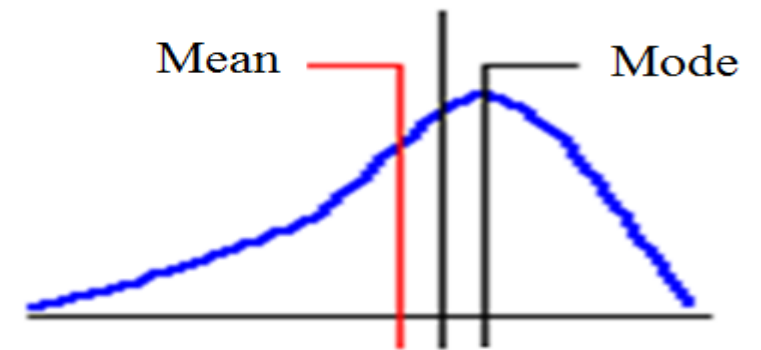
How to interpret the mean, median and mode:



Symmetric



Positive asymmetry



Negative asymmetry

Mean higher than median and mode, outliers are high values, but most of the values are smaller than the mean

Mean lower than median and mode, outliers are smaller values, most of the values are higher than the mean

Skewness

- Skewness – Asymmetry measure

$$S_K \approx \frac{1}{n} \frac{\sum_{i=1}^n (X_i - \bar{X})^3}{s^3}$$

where:

s = sample standard deviation , \bar{X} = arithmetic mean

Median = 2900 €

Arithmetic mean = 5500 €

2400€ 2600 €



2900 €

2900 €

5 values

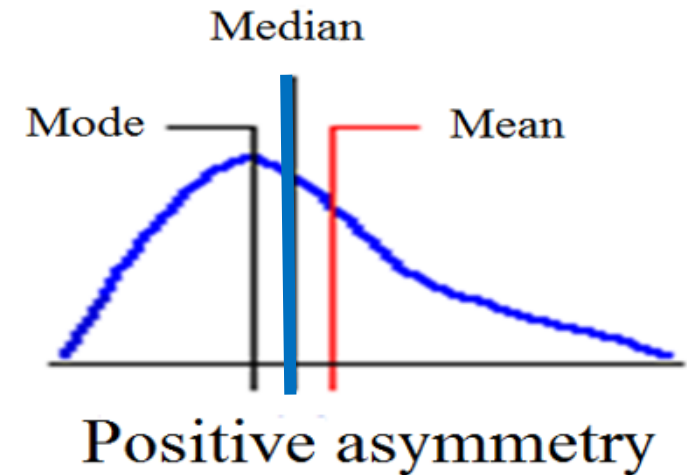
3200 €



19000 €

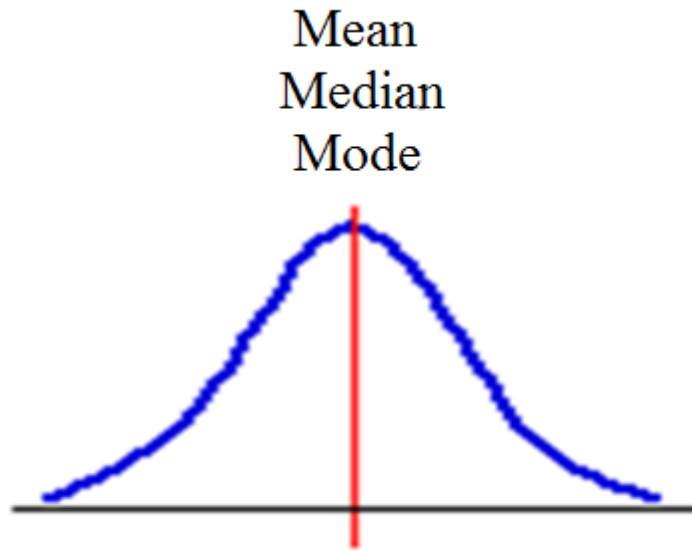


1 value



Skewness – Asymmetry measure

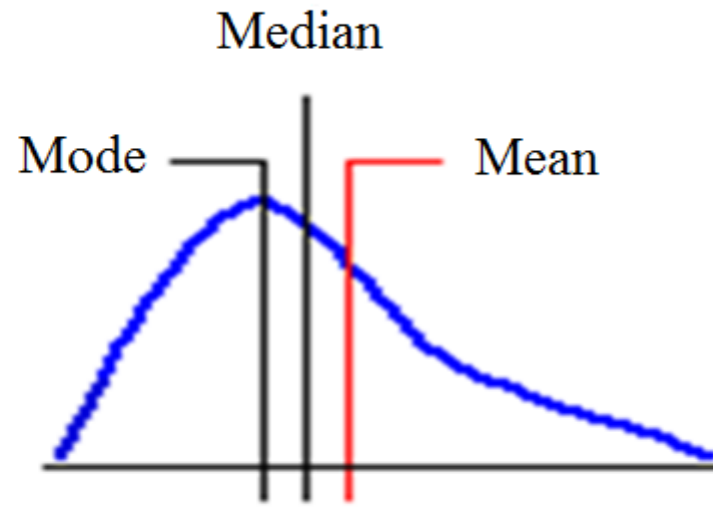
- **Skewness** is a measure of the asymmetry of the distribution of a quantitative variable.



Symmetric

Skewness=0

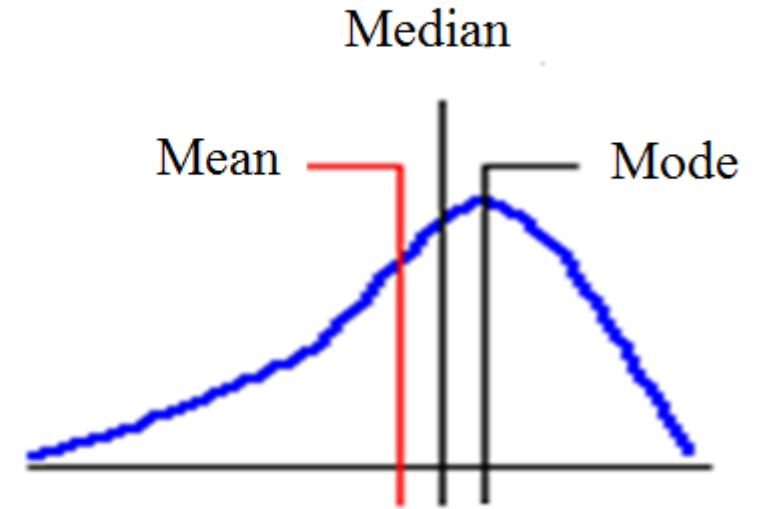
symmetrical distribution



Positive asymmetry

Skewness>0

asymmetry to the right



Negative asymmetry

Skewness<0

asymmetry to the left



Kurtosis

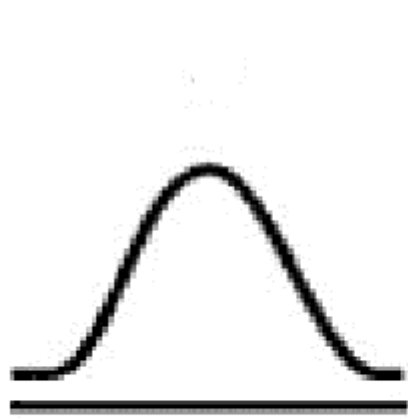
- Kurtosis – Peakedness, flatness, tailedness measure

$$K_E \approx \frac{1}{n} \frac{\sum_{i=1}^n (X_i - \bar{X})^4}{s^4} - 3$$

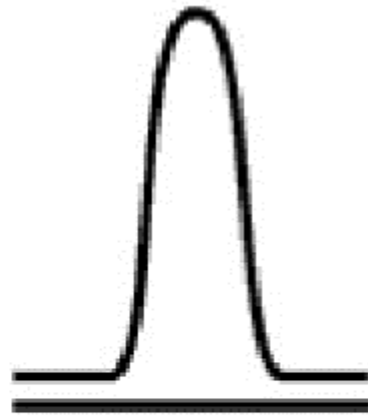
s = sample standard deviation , \bar{X} = arithmetic mean

Kurtosis - Peakedness, flatness, tailedness measure

- Kurtosis is a measure of the "tailedness" of the distribution of a variable



Mesokurtic
kurtosis=0

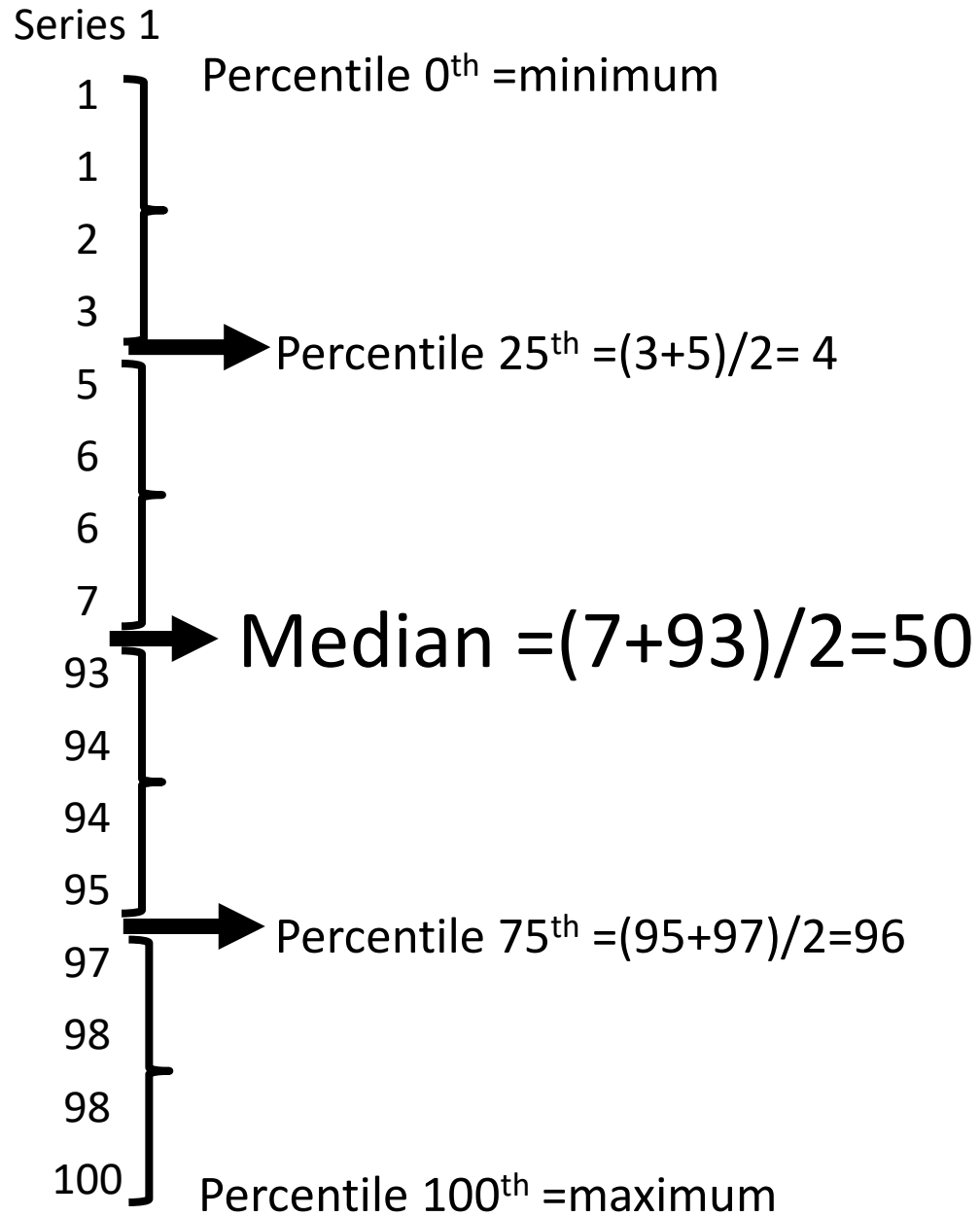


Leptokurtic
kurtosis>0

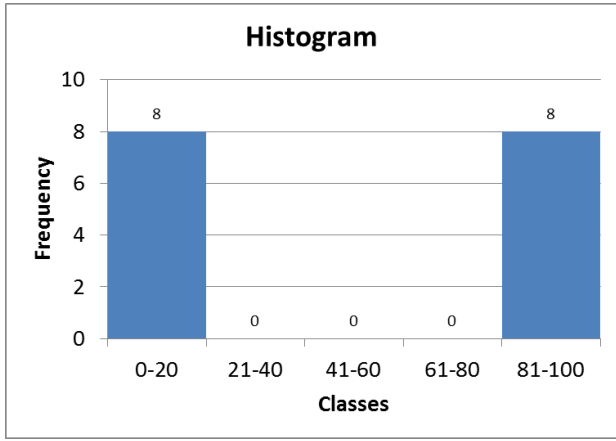
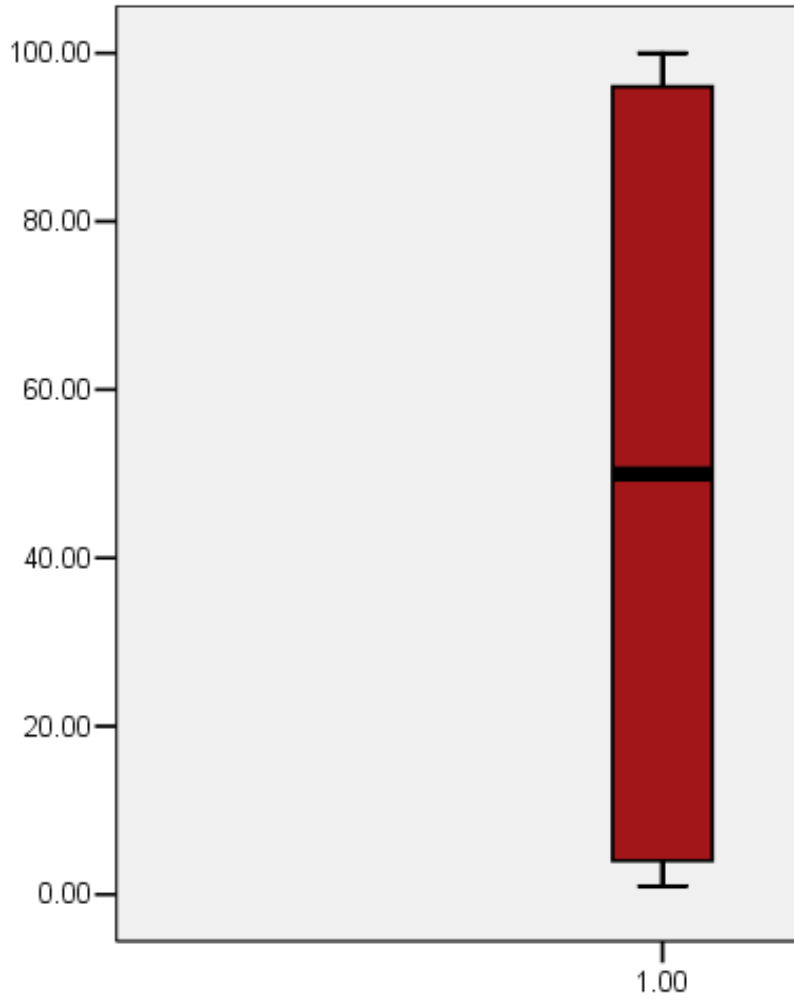


Platykurtic
kurtosis<0





Between minimum and percentile 25th is a small distance, but 25% of data are in this interval



Series 1

1 } Percentile 0th = minimum

1

2

3

5 } Percentile 25th = $(3+5)/2 = 4$

6

6

7

93 } Median = $(7+93)/2 = 50$

94

94

95

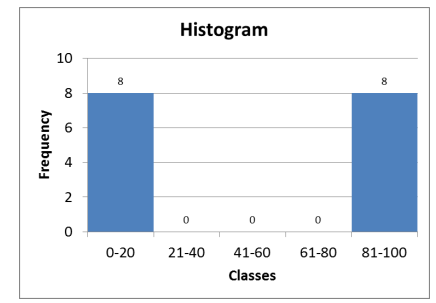
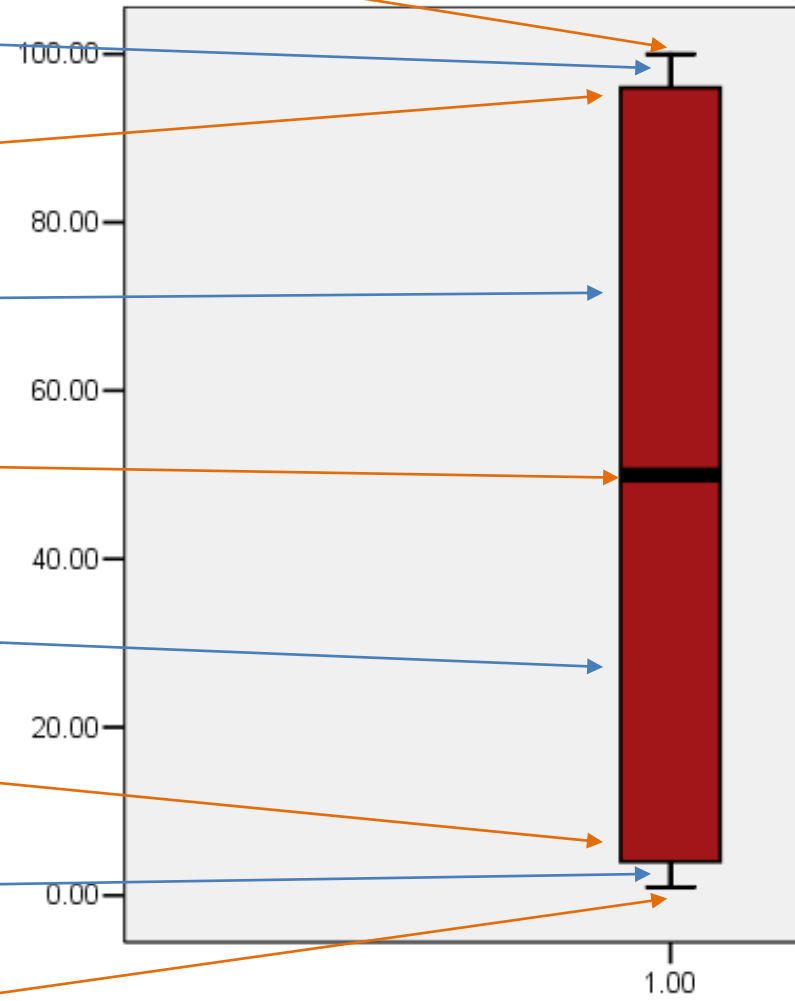
97 } Percentile 75th = $(95+97)/2 = 96$

98

98

100 } Percentile 100th = maximum

Between minimum and percentile 25th is a small distance, but 25% of data are in the interval



Series 2

1 Percentile 0th =minimum

44

45

46

→ Percentile 25th = $(46+48)/2=47$

48

48

49

50

→ Median = $(50+50)/2=50$

50

51

52

52

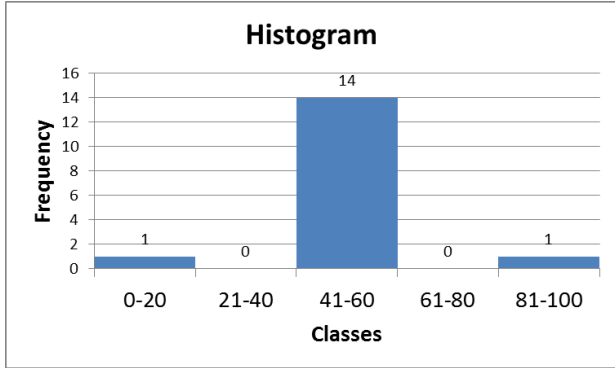
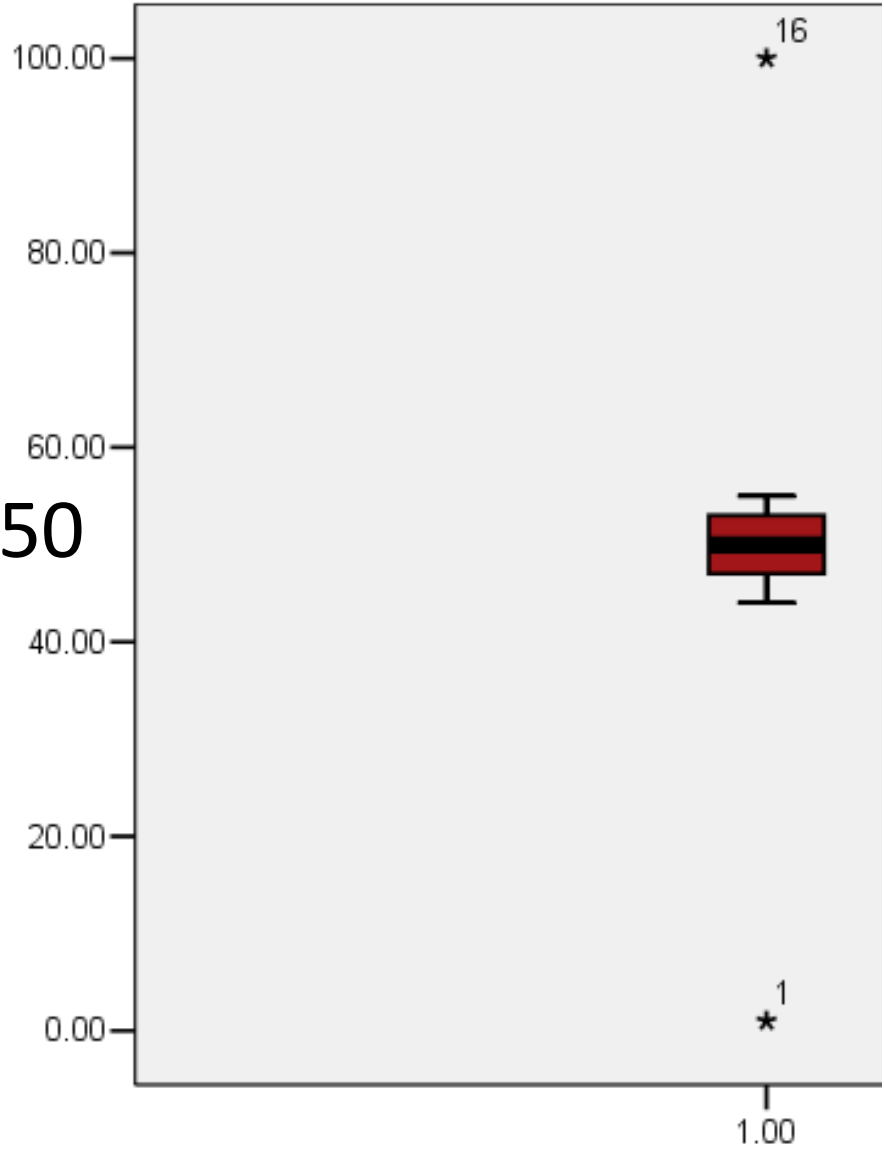
→ Percentile 75th = $(52+54)/2=53$

54

55

55

100 Percentile 100th =maximum



Series 3

Percentile 0th =minimum

1

11

24

29 → Percentile 25th = $(29+36)/2= 32.5$

36

41

45

49 → Median = $(49+51)/2=50$

51

55

59

64 → Percentile 75th = $(64+71)/2=67.5$

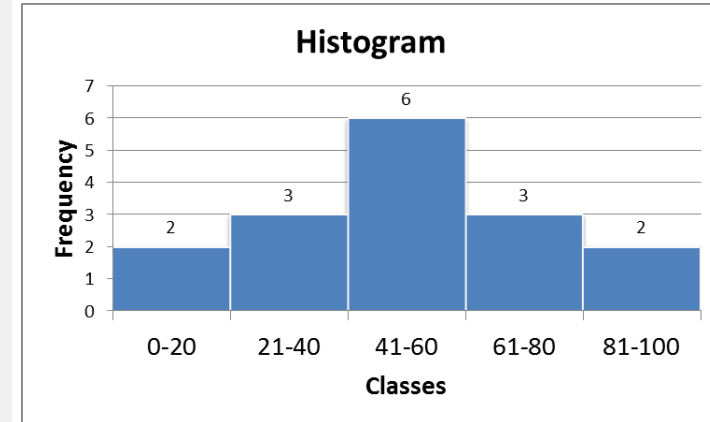
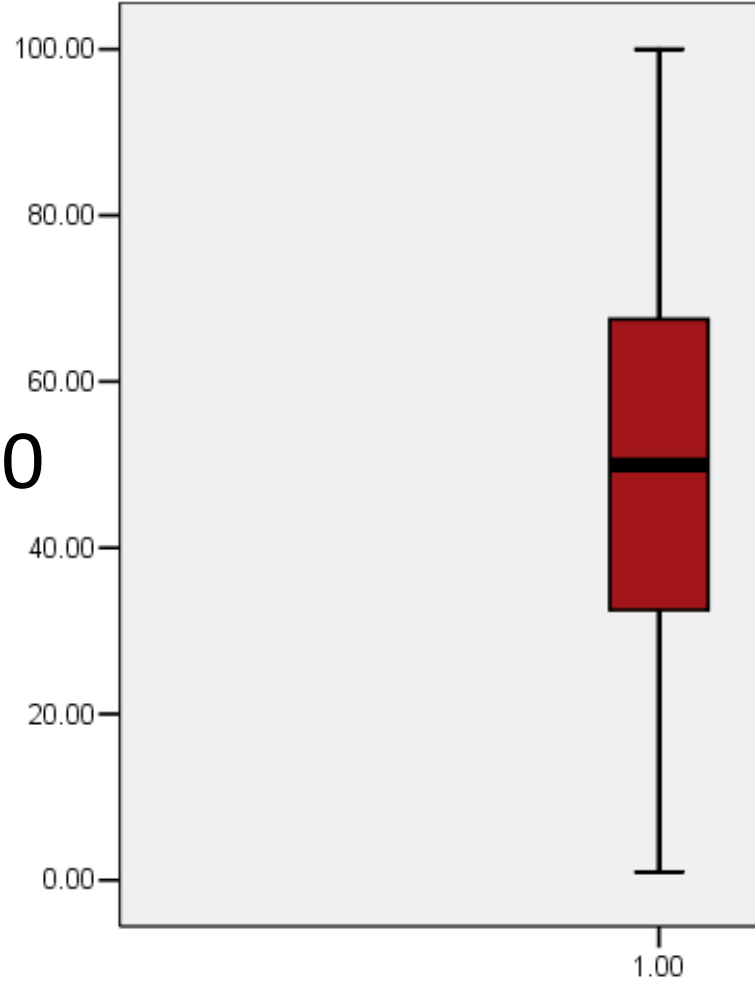
71

76

88

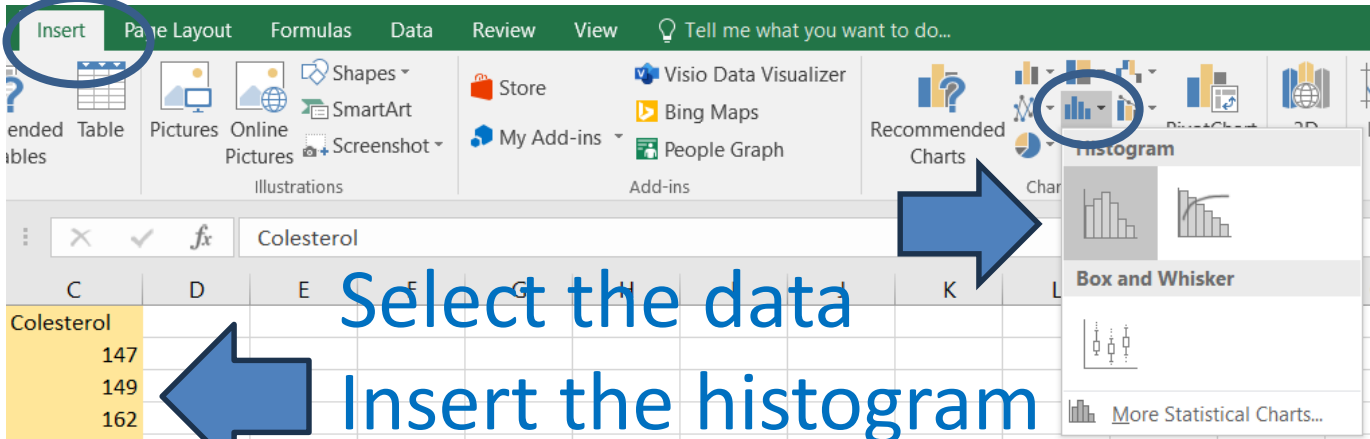
Percentile 100th =maximum

100



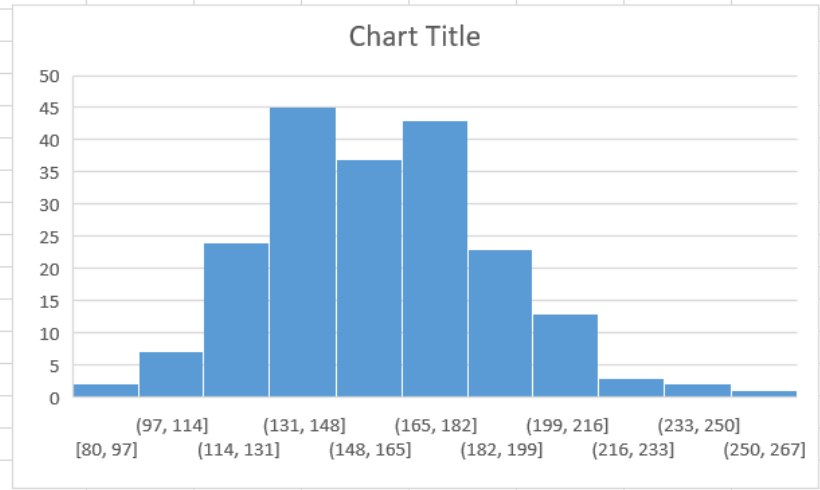
Centrality measurements	Functions	Formulas
Arithmetic mean	=AVERAGE(ARRAY)	
Median	=MEDIAN(ARRAY)	
Mode	=MODE.MULT(ARRAY) =MODE.SINGL(ARRAY)	
Central value		=(MAX(ARRAY)+MIN(ARRAY))/2
Dispersion		
Range		=MAX(ARRAY)-MIN(ARRAY)
Variation	=VAR.S(ARRAY)	
Standard deviation	=STDEV.S(ARRAY) =SQRT(VAR.S(ARRAY))	
Standard error		=STDEV.S(ARRAY)/SQRT(COUNT(ARRAY))
Coeficient of variation		CV =STDEV.S(ARRAY)/AVERAGE(ARRAY) CQV =(QUARTILE.INC(ARRAY,3)- QUARTILE.INC(ARRAY,1))/(QUARTILE.INC(ARRAY,3) +QUARTILE.INC(ARRAY,1))
Asymmetry		
Asymmetry	=SKEW(ARRAY)	
Kurtosis	=KURT(ARRAY)	

where ARRAY is the selection of cells containing the values of the variable of interest.

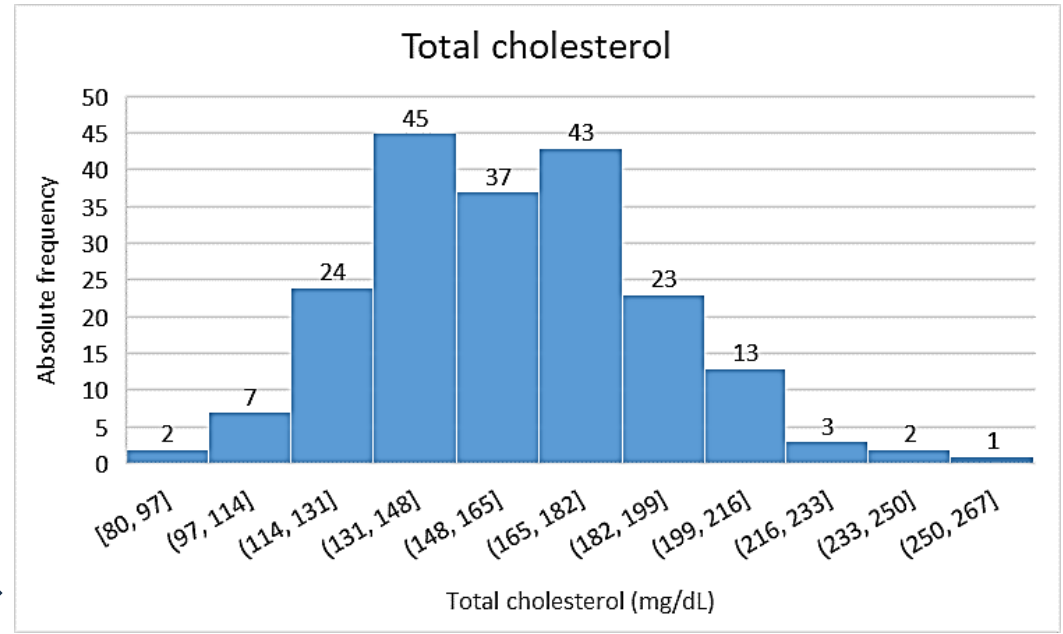


Select the data
Insert the histogram

Colesterol
147
149
162
147
160
134
142
173
168
213
146
172
180
175
129
168
140
128
183
206
204
131
157



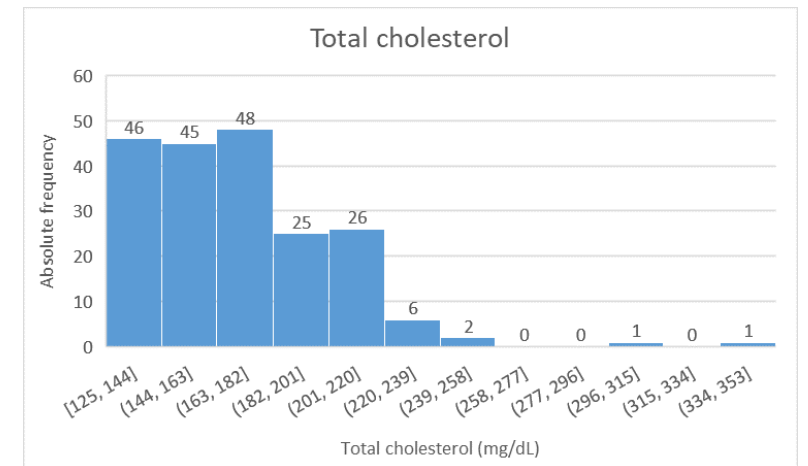
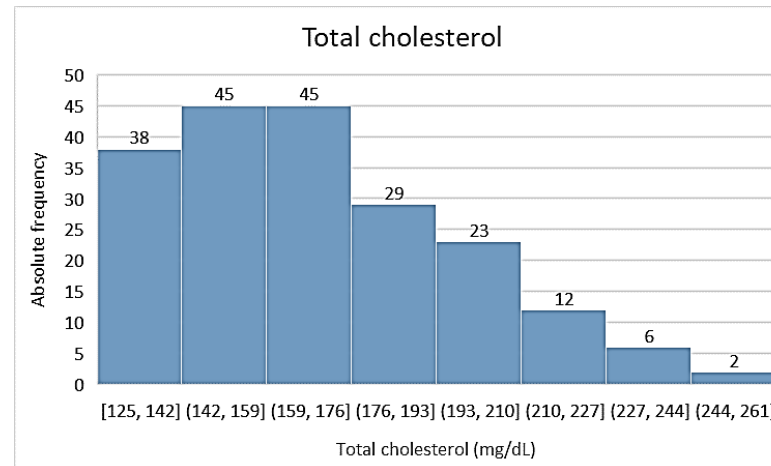
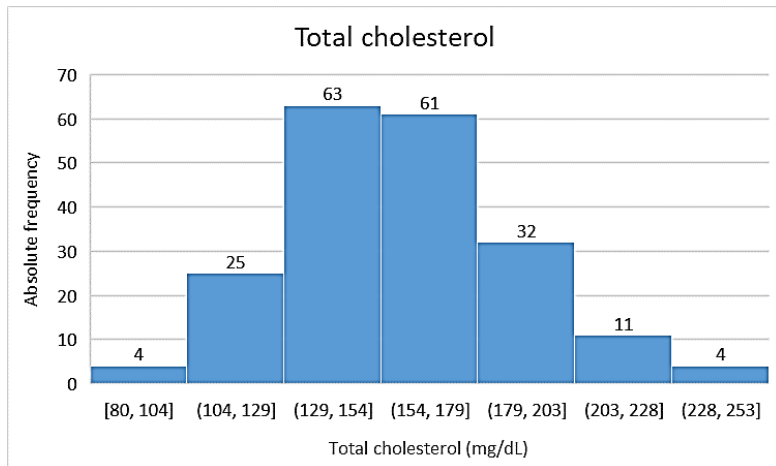
add title and axis titles



Variable name → Ox axis, “No. of subjects” or “Absolute frequency” → Oy axis.
Axis Ox must have unity of measurements!!!

Interpretations (see also Lecture about distributions)

- asymmetry
- peakedness, flatness, tailedness
- outliers, extreme cases
- distribution shape



Summarizing Numerical Data (Continuous, Discrete)

- Measure of central tendency
 - Arithmetic mean
 - Median
 - Mode
- Measure of spread
 - Variances
 - Standard Deviation
 - Coefficient of variation
 - Standard error
- Other measures
 - Skewness
 - Kurtosis
 - Quartile
 - Percentile
- Charts
 - Histograms (continuous)
 - Columns /bar (discrete)

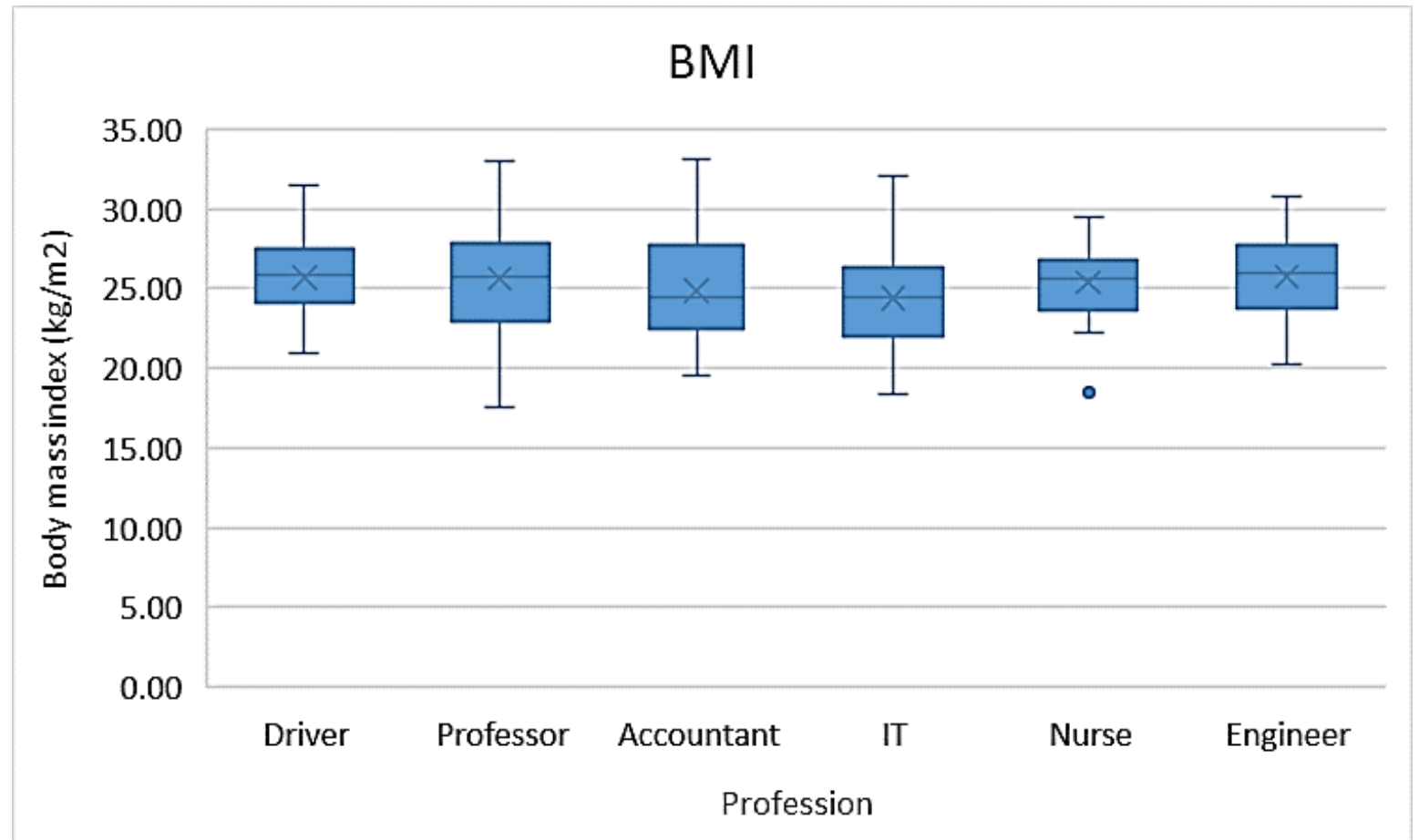


The relationship between two variables (from which one is quantitative or both are quantitative)

	A	B
1	Profession	Body mass index
2	Driver	25.66
3	Driver	27.45
4	Driver	22.33
5	Driver	26.27
6	Driver	25.96
7	Driver	25.24
8	Driver	25.05
9	Driver	28.19
10	Driver	31.52
11	Driver	28.07
12	Driver	23.36
13	Driver	21.99
14	Driver	25.80
15	Driver	25.81
16	Driver	24.39
17	Driver	20.90
18	Driver	26.76
19	Driver	27.68
20	Professor	32.98
21	Professor	31.10
22	Professor	23.72
23	Professor	28.09
24	Professor	22.92
25	Professor	25.76
26	Professor	21.38
27	Professor	19.13
28	Professor	25.28
29	Professor	27.47
30	Professor	24.89

The relationship between two variables (one quantitative, one qualitative)

- Box-whisker plot

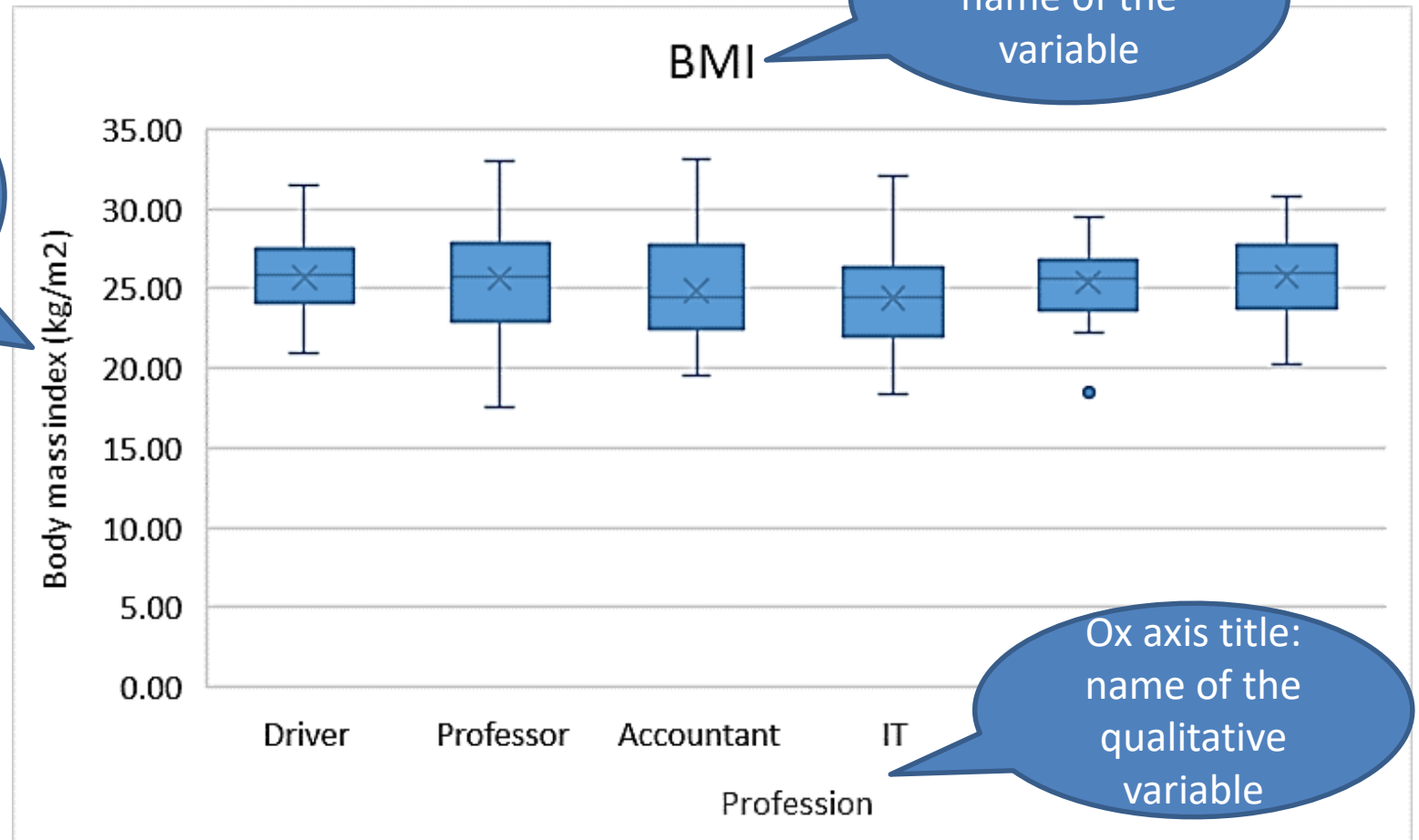


	A	B
1	Profession	Body mass index
2	Driver	25.66
3	Driver	27.45
4	Driver	22.33
5	Driver	26.27
6	Driver	25.96
7	Driver	25.24
8	Driver	25.05
9	Driver	28.19
10	Driver	
11	Driver	
12	Driver	
13	Driver	
14	Driver	
15	Driver	
16	Driver	
17	Driver	
18	Driver	26.76
19	Driver	27.68
20	Professor	32.98
21	Professor	31.10
22	Professor	23.72
23	Professor	28.09
24	Professor	22.92
25	Professor	25.76
26	Professor	21.38
27	Professor	19.13
28	Professor	25.28
29	Professor	27.47
30	Professor	24.89

The relationship between two variables (one quantitative, one qualitative)

- Box-whisker plot

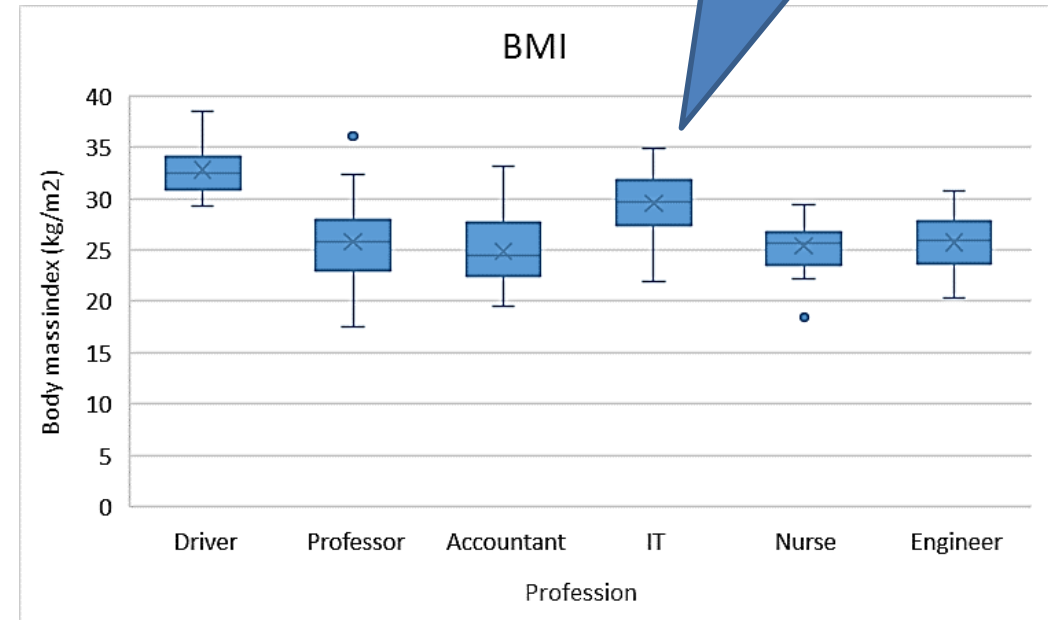
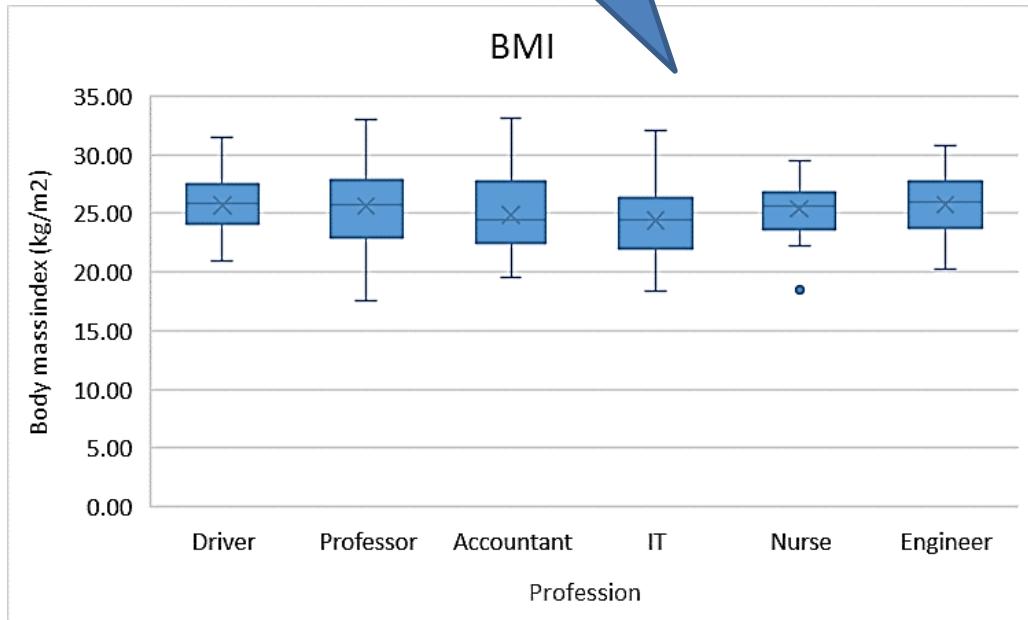
Oy Axis title: name of the quantitative variable and **unity of measurements!!!**



Interpretation (see also lecture about statistical tests)

No major differences between professions about the distribution of BMI

Drivers and IT had higher BMI than the other professions



Statistics - two variables (one quantitative, one qualitative)

Professions

BMI (kg/m ²)	IT	Driver	Professor	Accountant	Nurse	Engineer
Mean	29.82	32.79	25.77	24.83	25.38	25.72
Standard Error	0.50	0.60	0.77	0.75	0.46	0.49
Median	29.78	32.54	25.73	24.50	25.67	26.00
Mode	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
Standard Deviation	2.55	2.53	3.95	3.53	2.44	2.63
Sample Variance	6.53	6.38	15.58	12.44	5.95	6.90
Kurtosis	-0.57	0.38	1.12	0.05	1.03	-0.52
Skewness	-0.06	0.66	0.35	0.54	-0.51	-0.25
Range	10.51	9.16	18.56	13.63	11.01	10.51
Minimum	24.43	29.28	17.51	19.51	18.46	20.27
Maximum	34.93	38.44	36.07	33.13	29.47	30.78
Sum	775	590	670	546	710	745
Count	26	18	26	22	28	29

there is not one value with the highest frequency

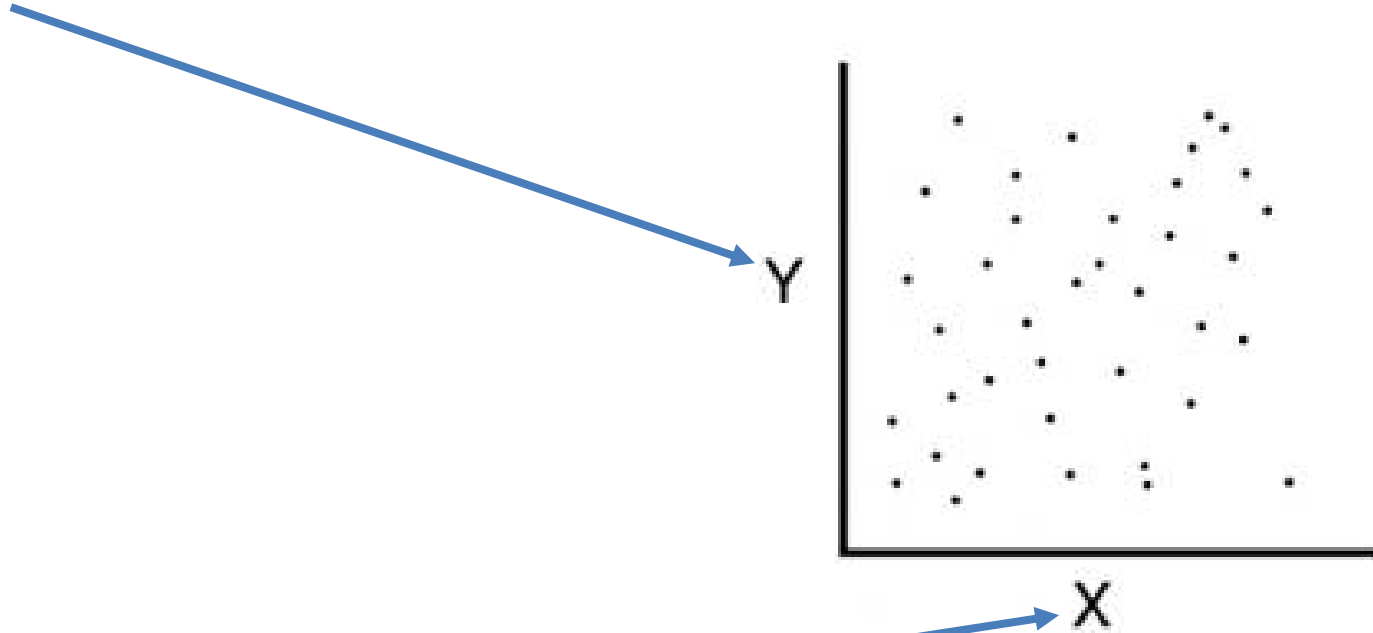
no. of people in each group

XY Scatter chart

- The relationship between two quantitative or ordinal (scores) characteristics
- How is the relationship?

XY scatter chart

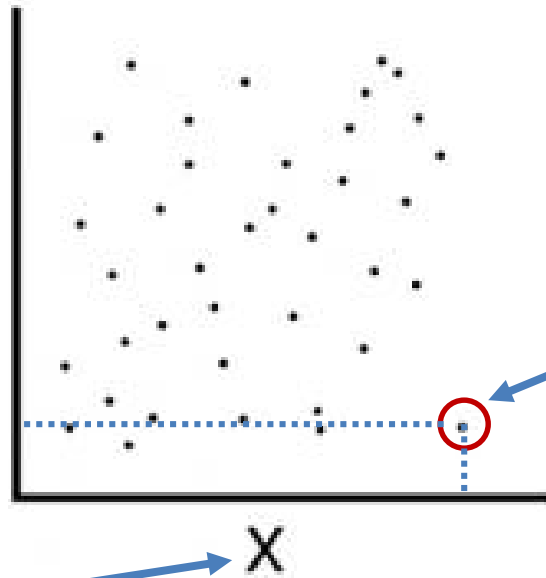
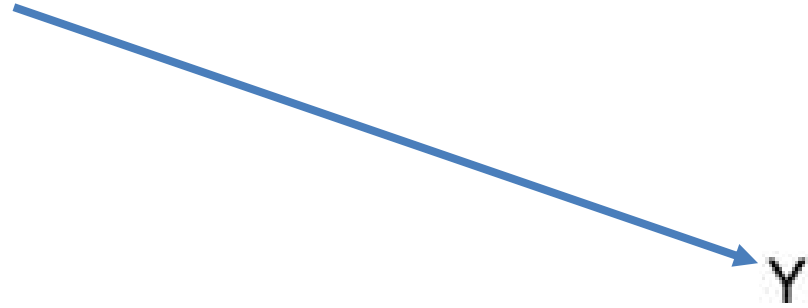
Y – quantitative variable or ordinal (score)



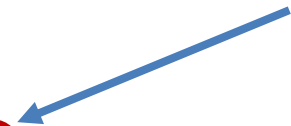
X – quantitative variable or ordinal (score)

XY scatter chart

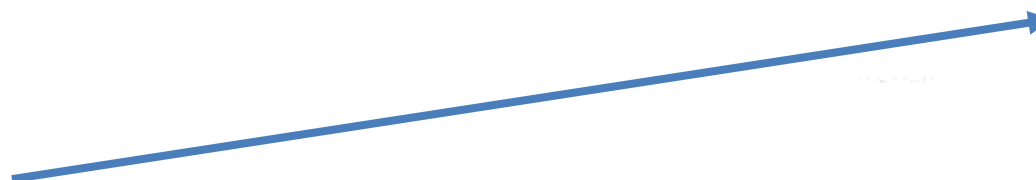
Y – quantitative variable or ordinal (score)



A subject

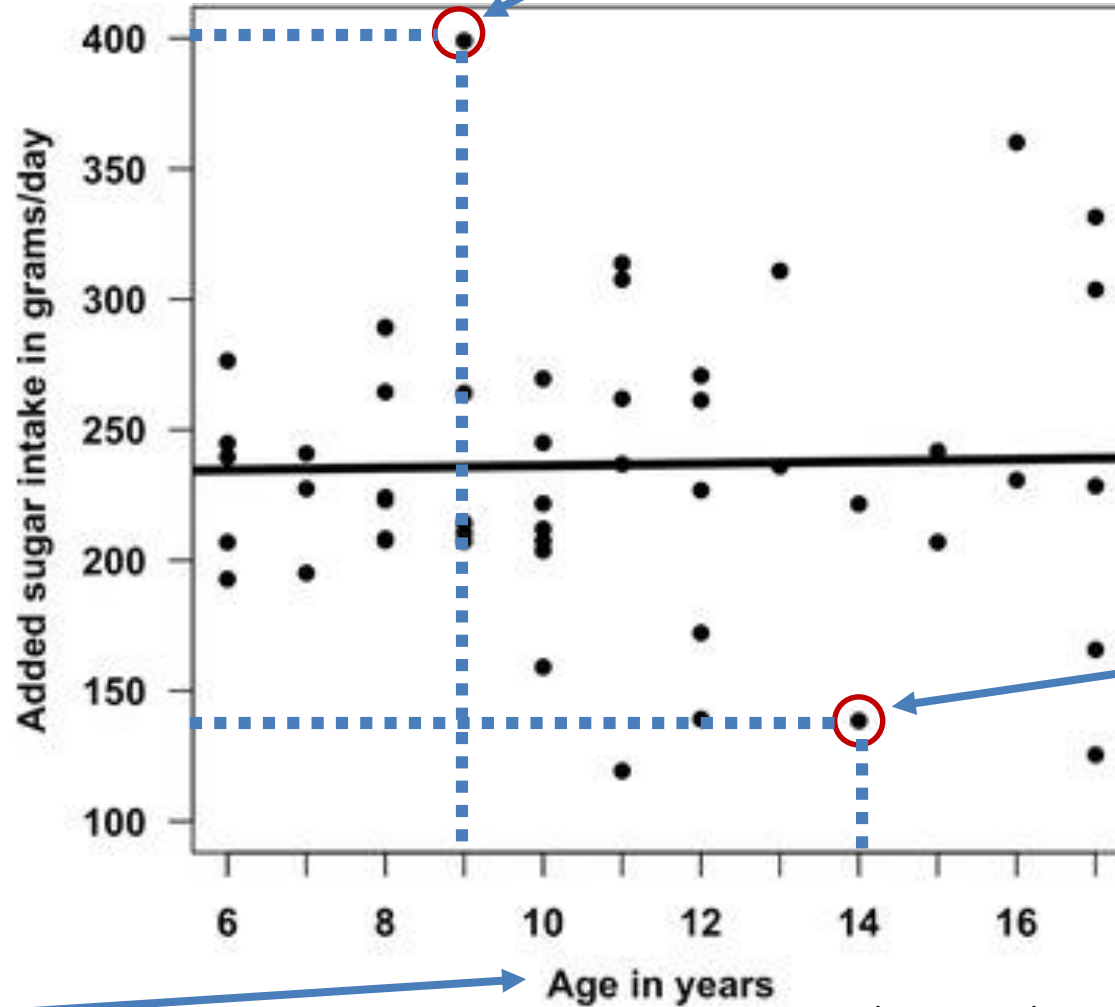


X – quantitative variable or ordinal (score)



Y – quantitative variable - Sugar intake

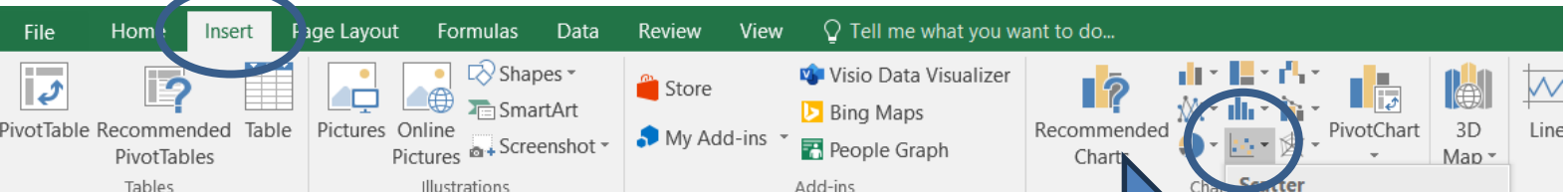
A subject with age=9 years and 400 grams/days sugar intake



A subject with age=14 years and 140 grams/days sugar intake

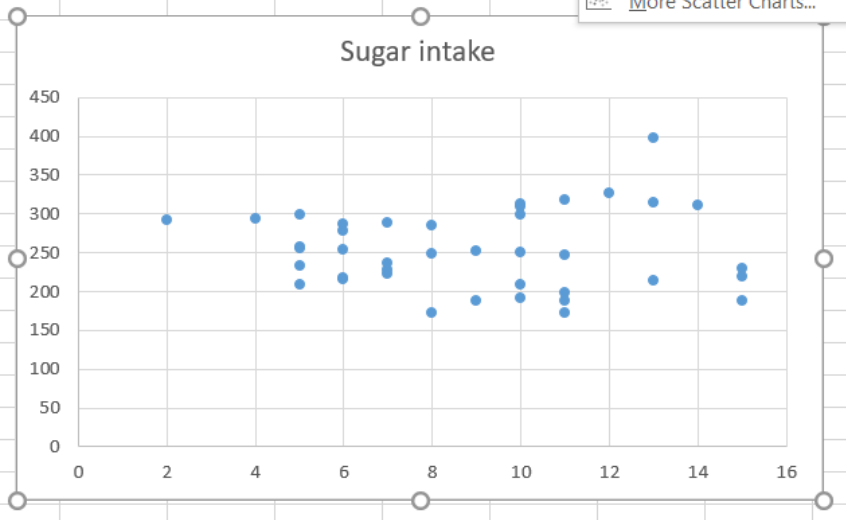
X – quantitative variable - Vârsta

Chi DL, Hopkins S, O'Brien D, Mancl L, Orr E, Lenaker D. Association between added sugar intake and dental caries in Yup'ik children using a novel hair biomarker. BMC Oral Health. 2015 Oct 9;15(1):121.

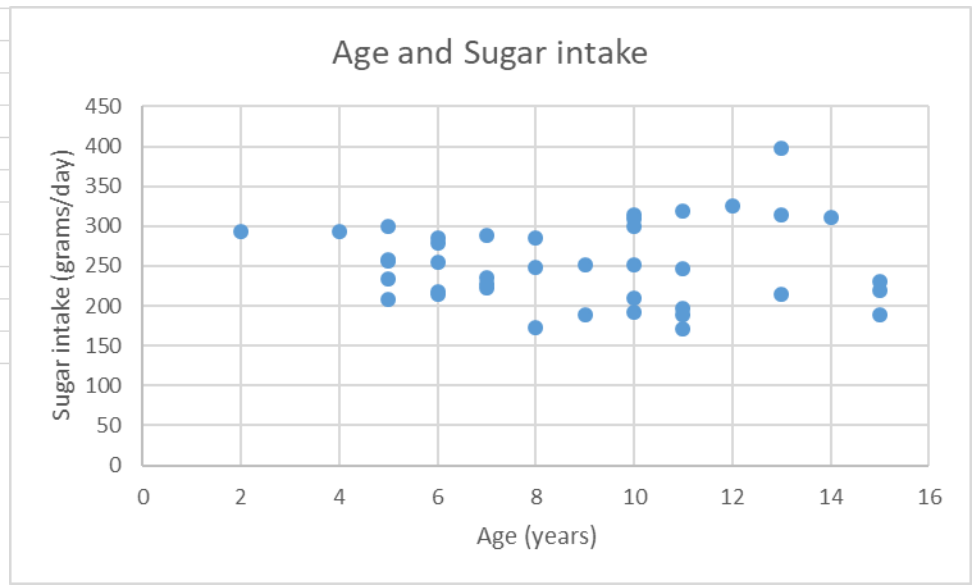


Select the data
Insert the Scatter
chart

	Age	Sugar intake
1		
2	14	311
3	10	251
4	11	318
5	5	233
6	11	172
7	15	189
8	6	215
9	5	257
10	7	222
11	9	188
12	12	326
13	10	192
14	8	248
15	7	228
16	7	236
17	2	293
18	10	209
19	15	220
20	8	173
21	4	294
22	9	251
23	5	209
24	15	230
25	13	314



add title and axis titles



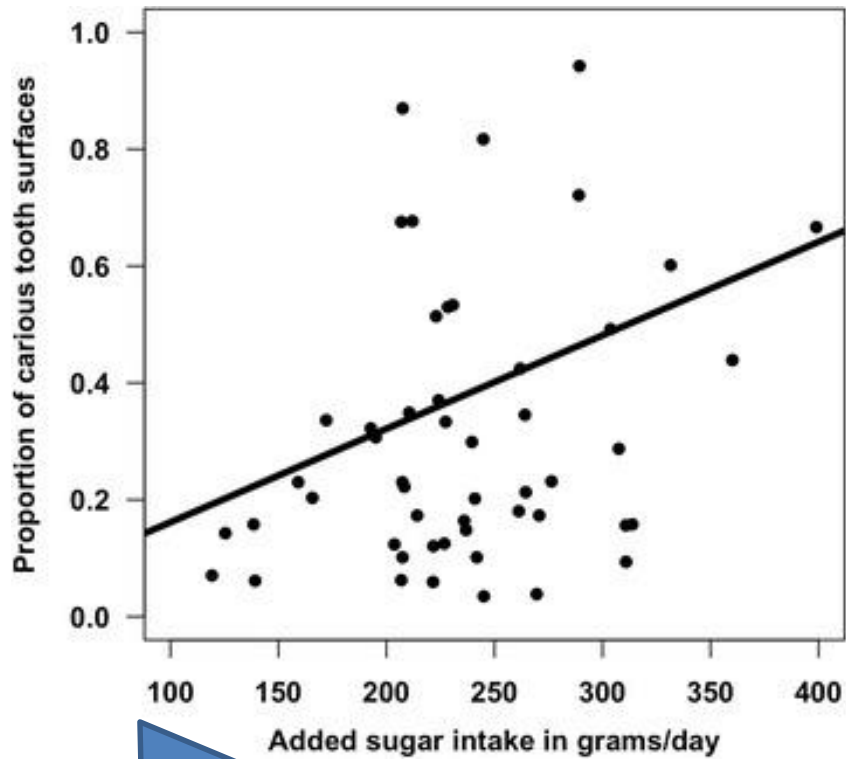
First variable → Ox axis, second variable → Oy axis.
Each axis must have unity of measurements!!!

Scenario

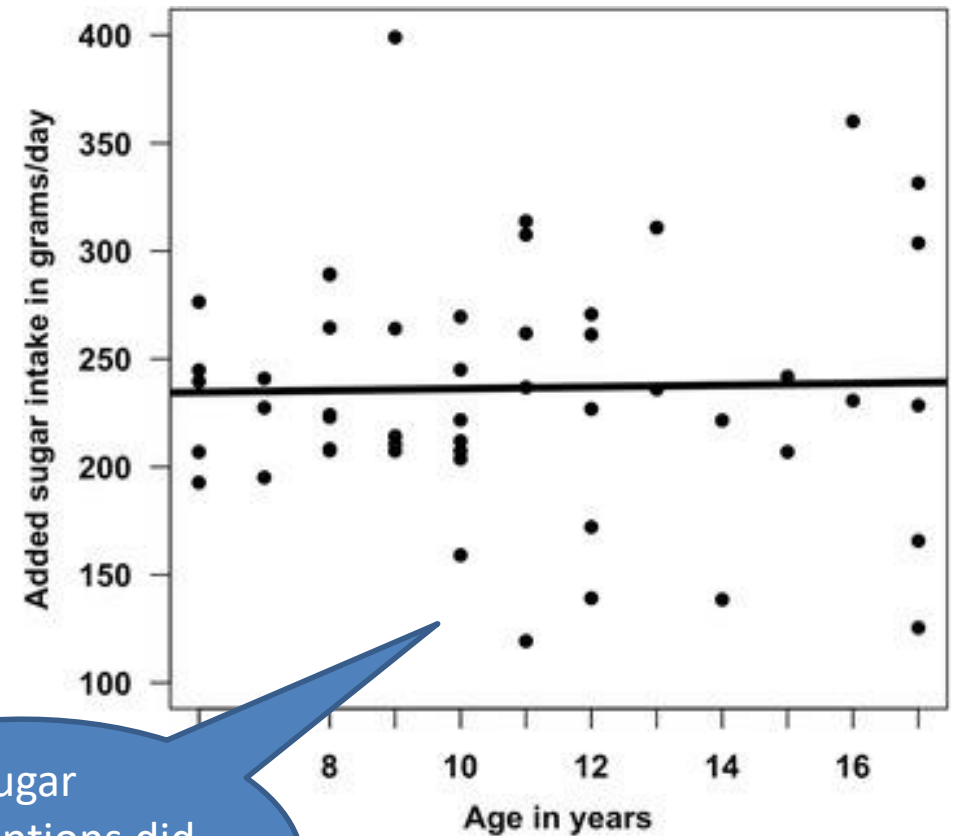
- In this study, they used a validated hair-based biomarker to measure added sugar intake in Alaska Native Yup'ik children ages 6–17 years (N = 51). They want to demonstrate “that added sugar intake would be positively associated with tooth decay”.

[Chi DL, Hopkins S, O'Brien D, Mancl L, Orr E, Lenaker D. Association between added sugar intake and dental caries in Yup'ik children using a novel hair biomarker. BMC Oral Health. 2015 Oct 9;15(1):121.]

Interpretations (see also Lecture about correlation)



Proportions of carious tooth surface depend on the quantity of sugar intake, quantity increase, surface increase too



Sugar consumptions did not depend by age

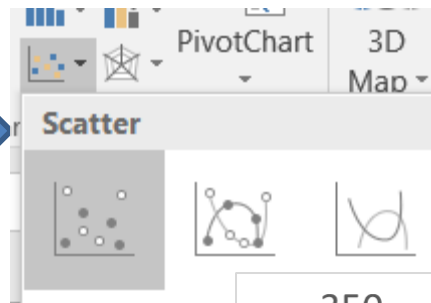
XY Scatter chart

- Two variables – can have different unities of measurements

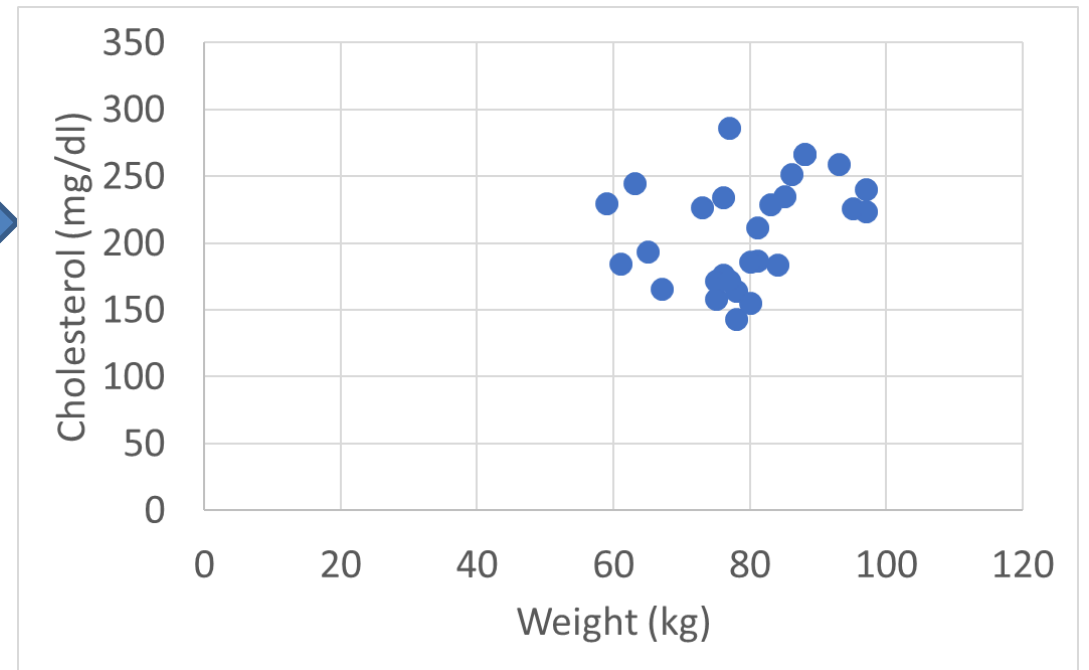
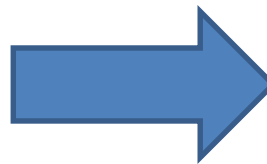


B	C
Weight (kg)	Cholesterol (mg/dL)
81	212
76	176
97	224
78	164
76	234
86	252
75	172
84	184
78	143
67	166
73	227
93	259
80	155
83	229
77	286
88	267
75	158
81	187
61	185
77	172
95	226
59	230
65	194

B	C
Weight (kg)	Cholesterol (mg/dL)
81	212
76	176
97	224
78	164
76	234
86	252
75	172
84	184
78	143
67	166
73	227
93	259
80	155
83	229
77	286
88	267
75	158
81	187
61	185
77	172
95	226
59	230
65	194
85	225

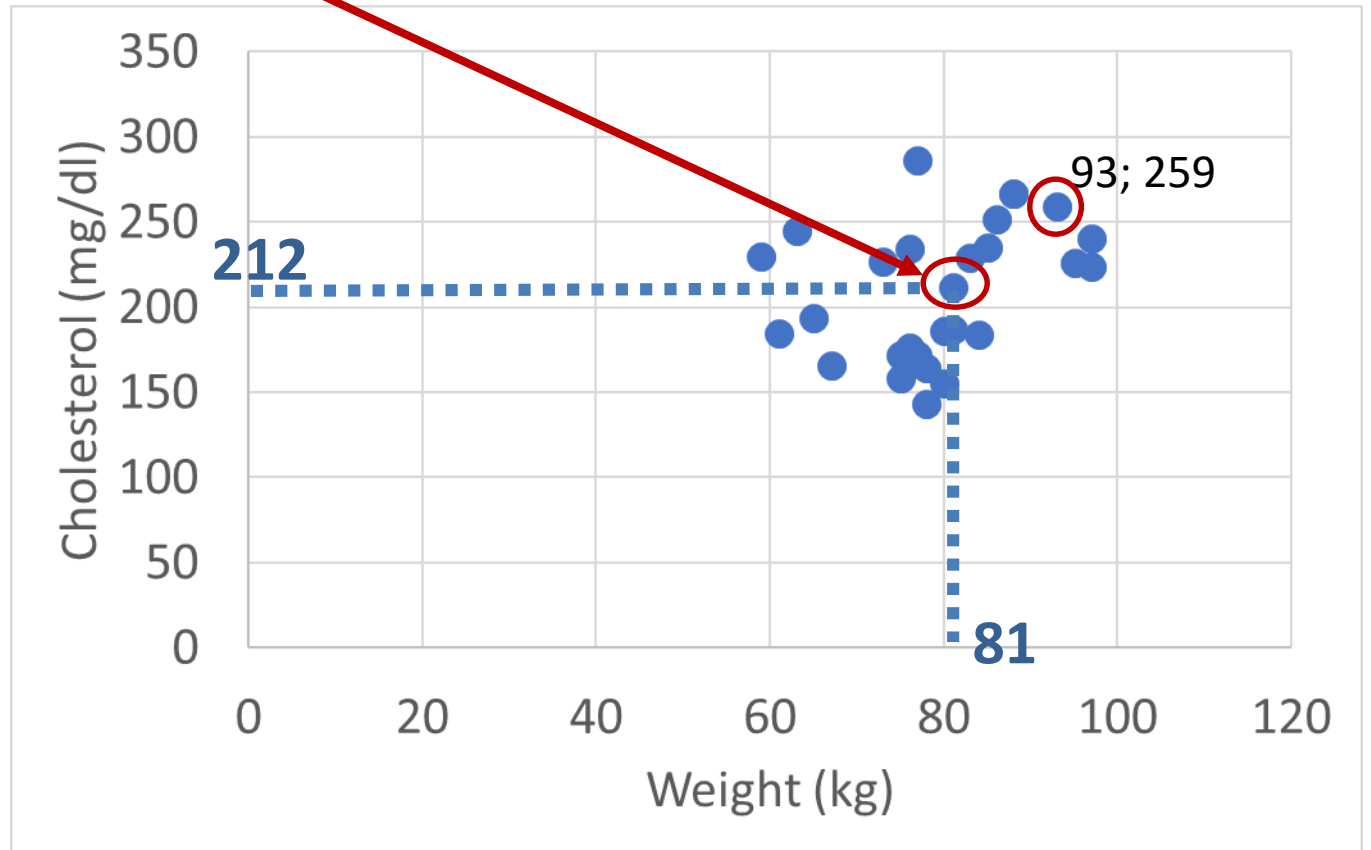


Insert XY Scatter chart



Select both column

B	C
Weight (kg)	Cholesterol (mg/dL)
81	212
76	176
97	224
78	164
76	234
86	252
75	172
84	184
78	143
67	166
73	227
93	259
80	155
83	229
77	286
88	267
75	158
81	187
61	185
77	172
95	226
59	230
65	194
63	245
80	186
88	267
97	240
85	225



Reccomendations

- A table should be self-explanatory.
 - Title
 - Definitions of columns
 - Definitions of rows
 - Units of measurements in the titles of each column / row (if it is applied)

Reccomendations

- A table should be self-explanatory.
 - Abbreviations and symbols must be explained at the bottom of the table (in the legend)
 - Brief descriptive heading: what? when? where?
 - Total on rows and columns (if it is applied)
- Make it simple
 - 2 or 3 small tables instead of one big table

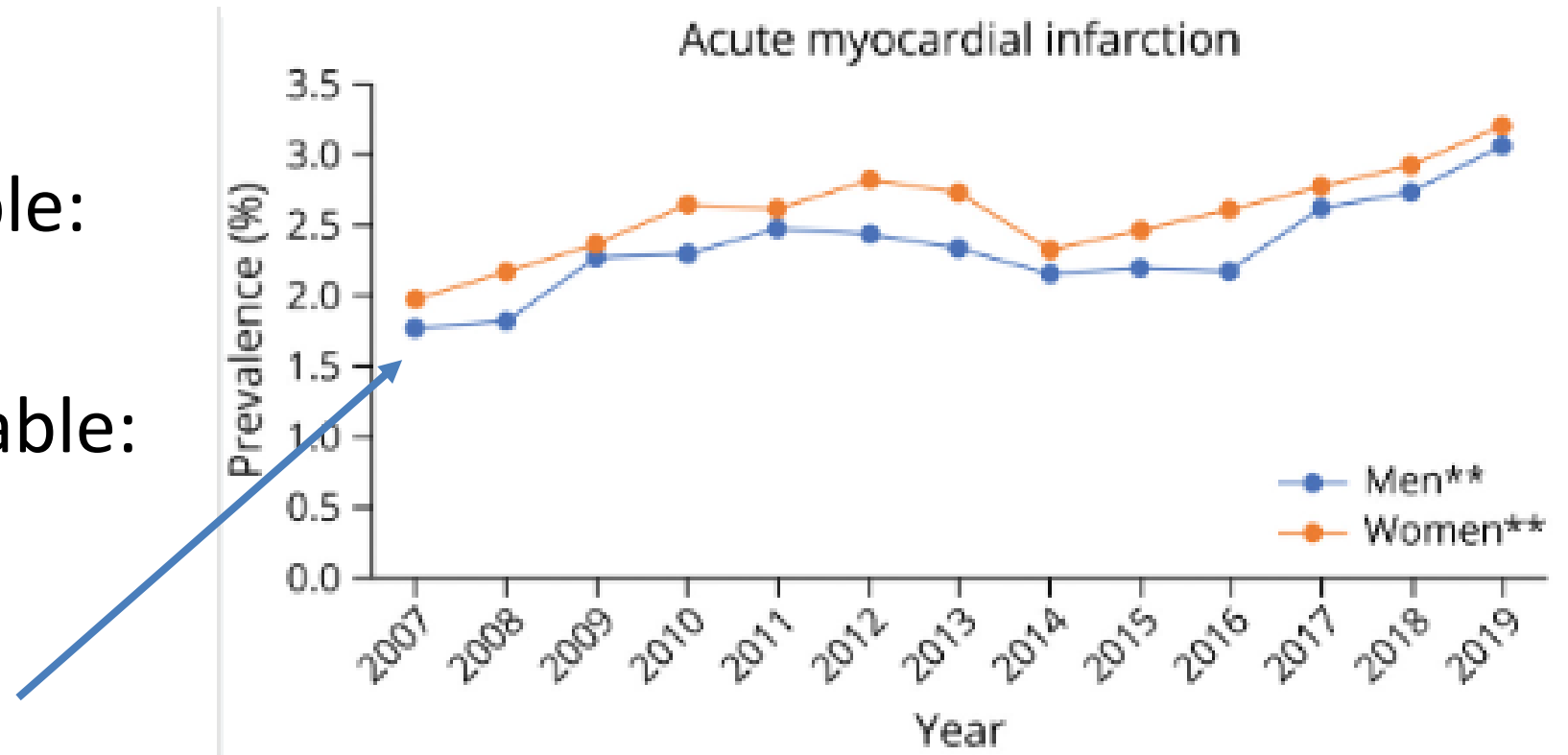
Reccomendations

- A chart should be self-explanatory.
 - Title
 - Definitions of axes (if it is applied)
 - Units of measurements for each axis (if it is applied)
 - Legend (if it is applied)

Other charts

Line charts

- qualitative variable:
 - hypertension
- quantitative variable:
 - time

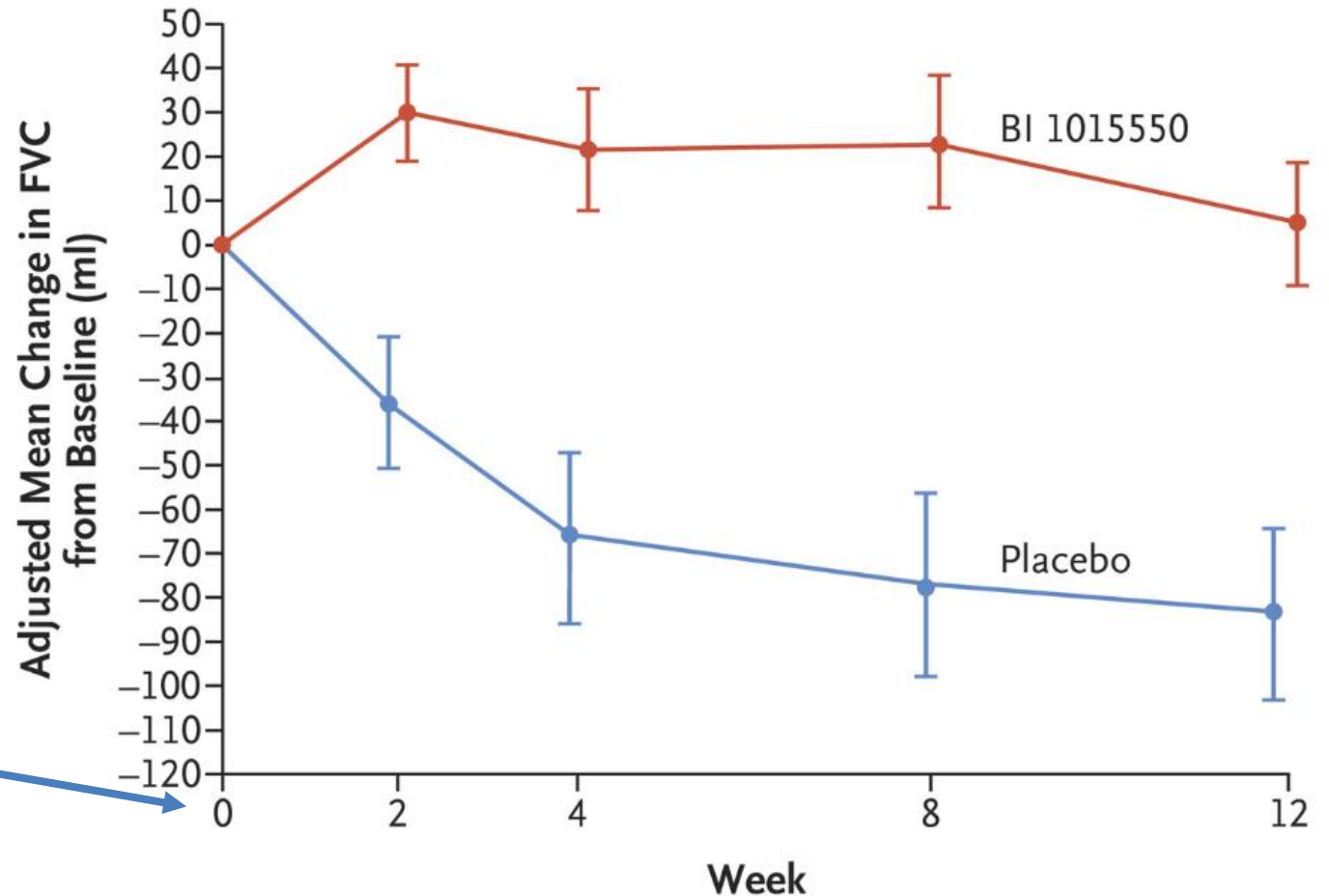


Trends in the frequency of noninfectious complications after acute ischemic stroke admission in the United States according to sex

Other charts

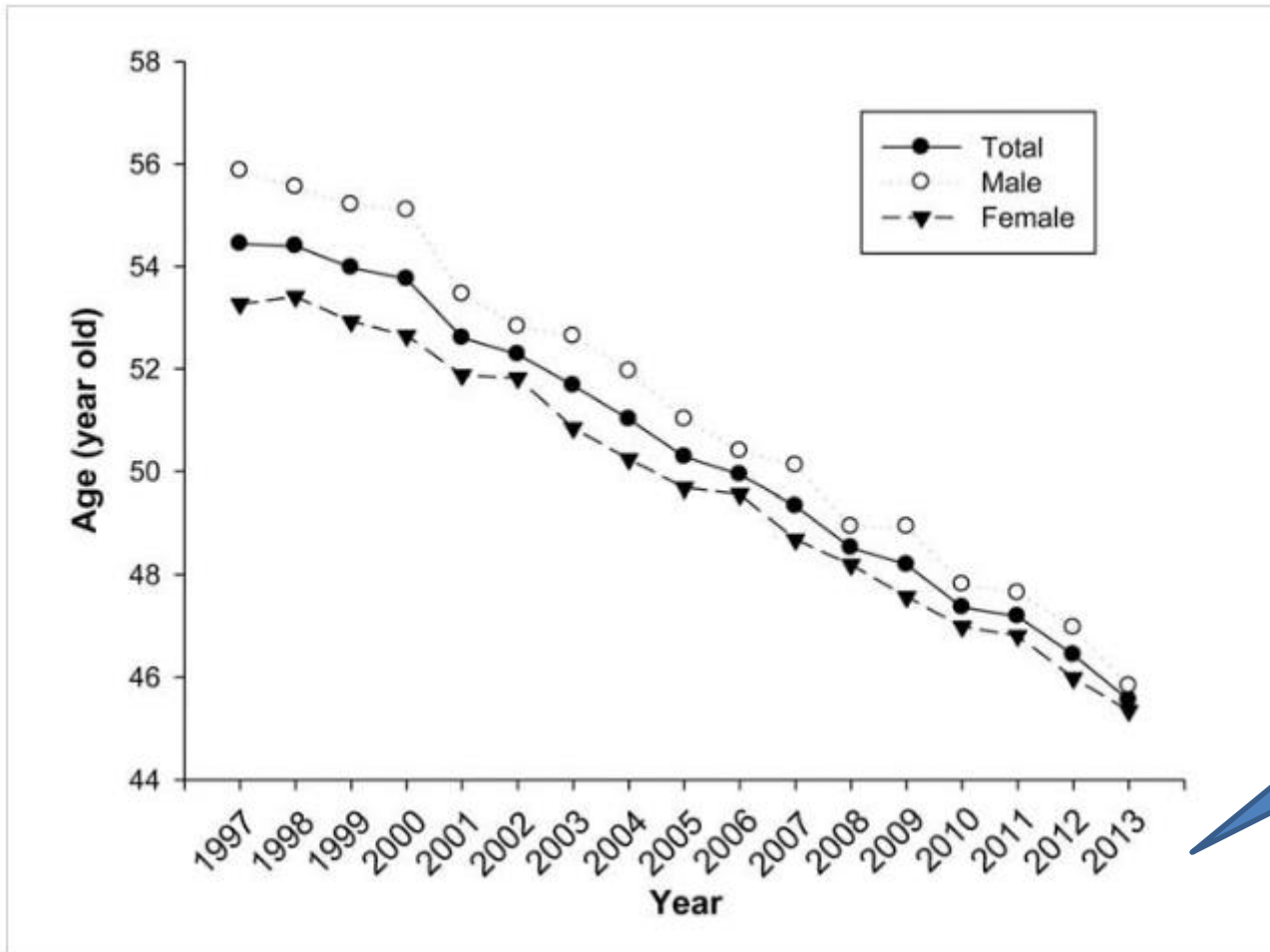
Line charts

- quantitative variable:
 - FVC (forced vital capacity)
 - mean, standard error
- quantitative variable:
 - time



Change in FVC over time for subjects with BI 1015550 versus placebo

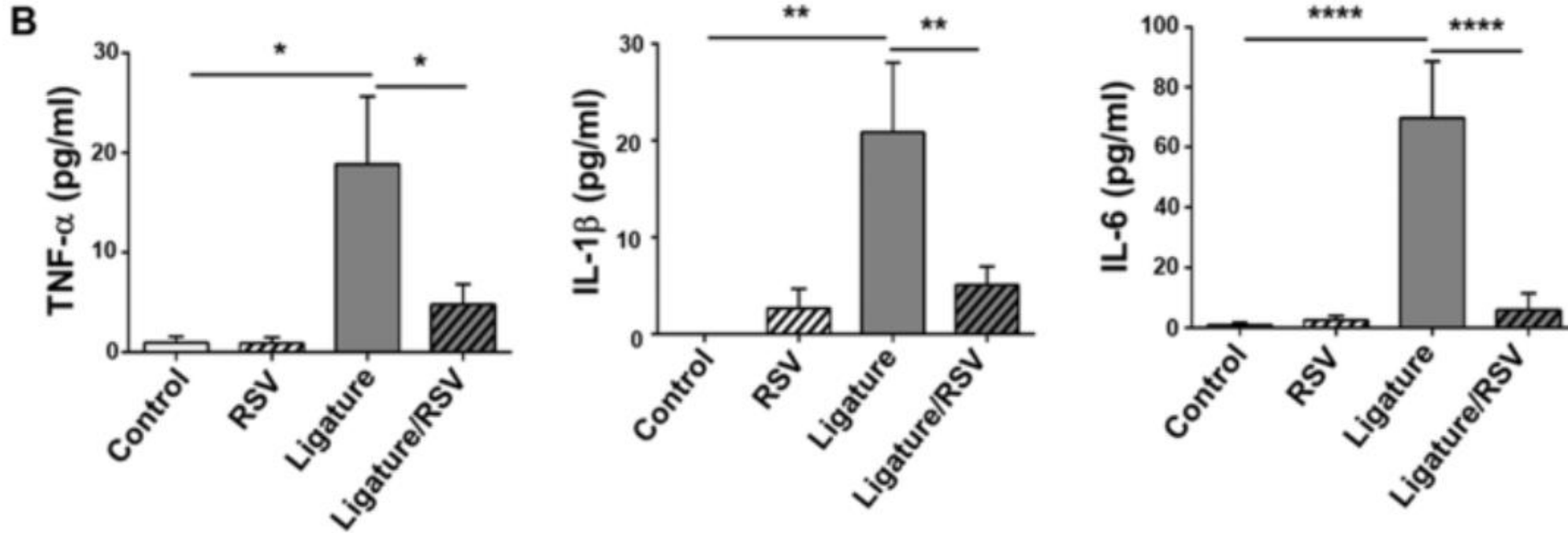
Line charts with arithmetic means



Mean age of patients with periodontitis in Taiwan. The mean age for periodontitis was shown a decrease pattern from 1997 to 2013.

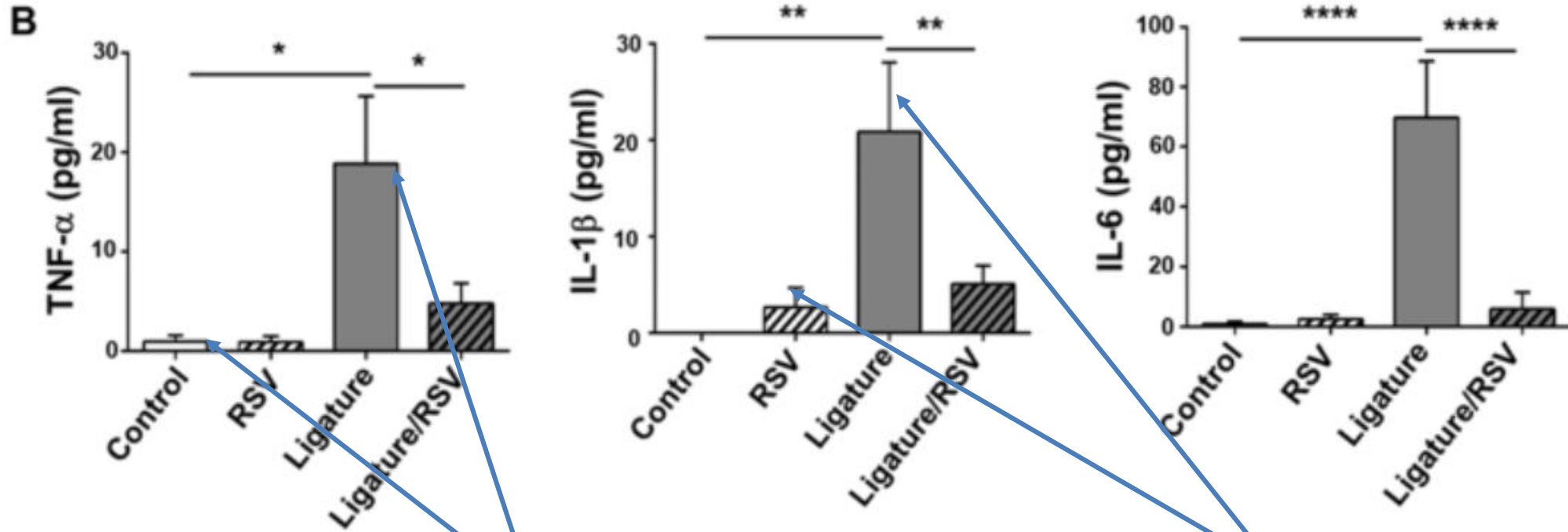
Time

Mean error plot



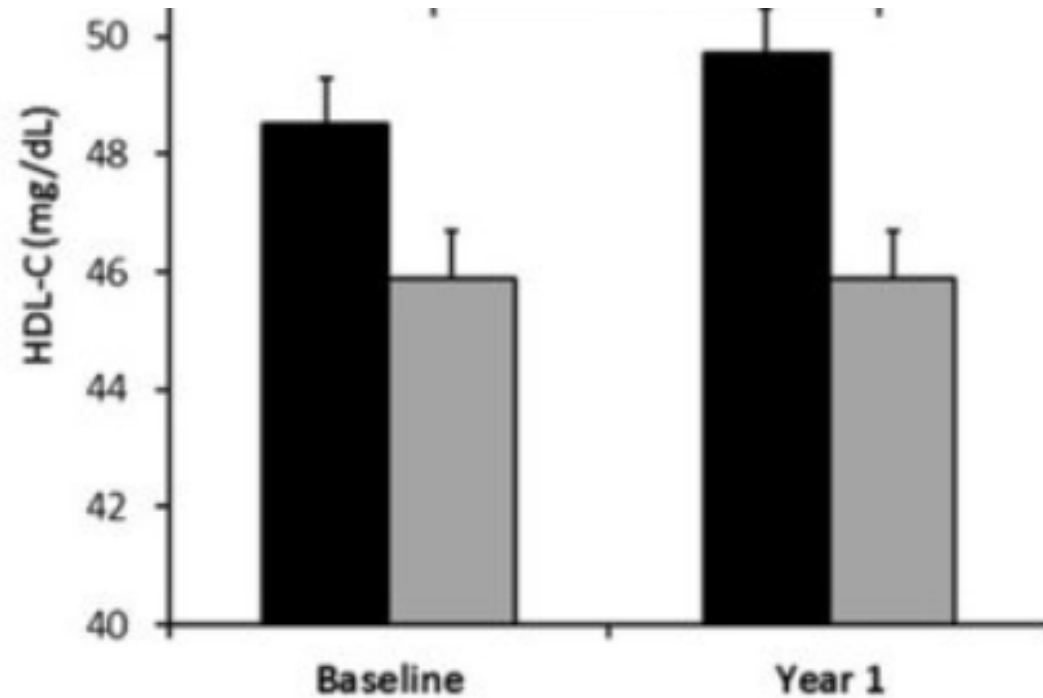
- Ligature placement (paradontosis disease induced in mice) notably increased the serum levels of pro-inflammatory cytokines, and rosovastatin (RSV) completely inverted such increases.
- **(B)** Levels of TNF- α , IL-1 β and IL-6 from the mice
- means + SE

Mean error plot



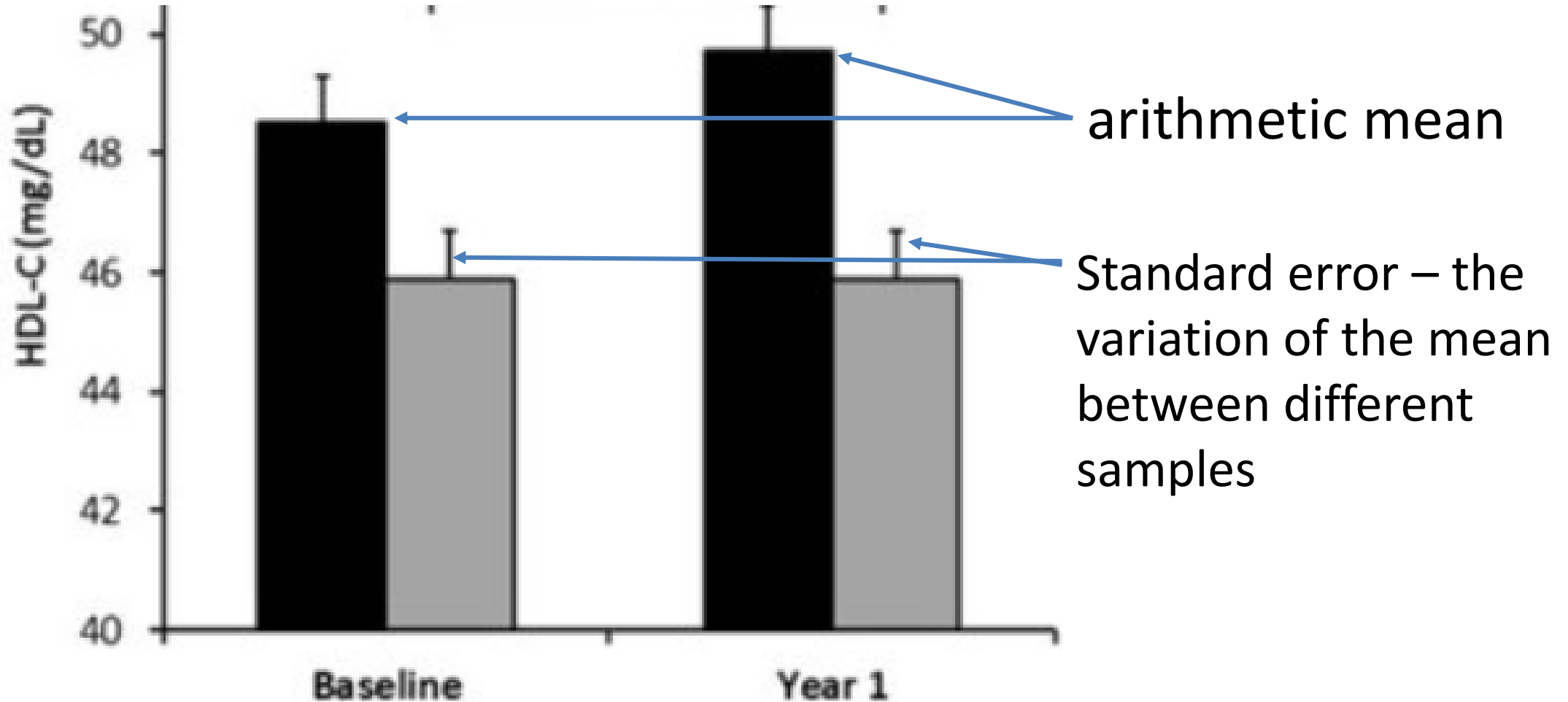
arithmetic mean

Standard error – the variation of the mean between different samples

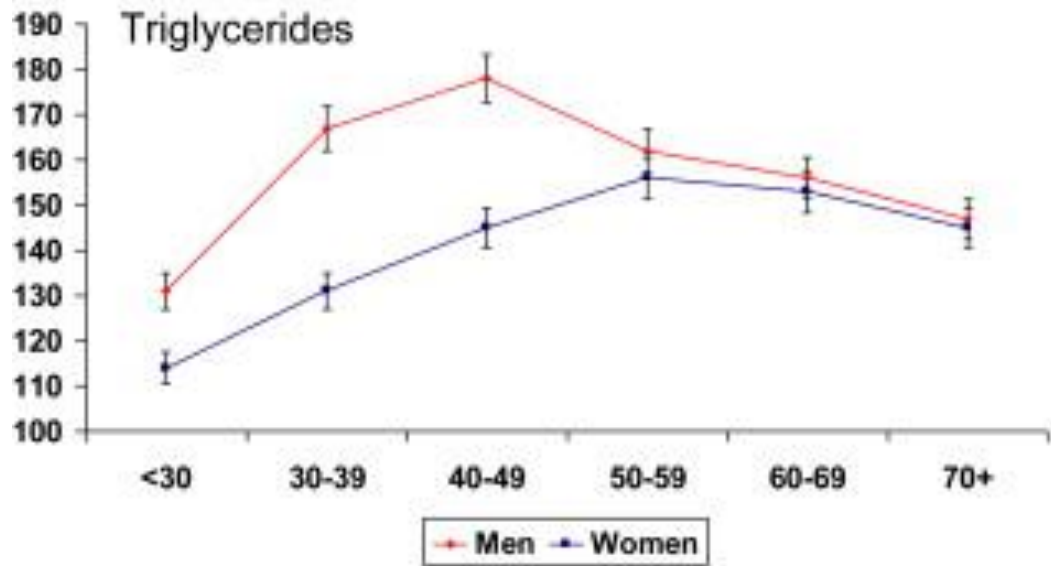
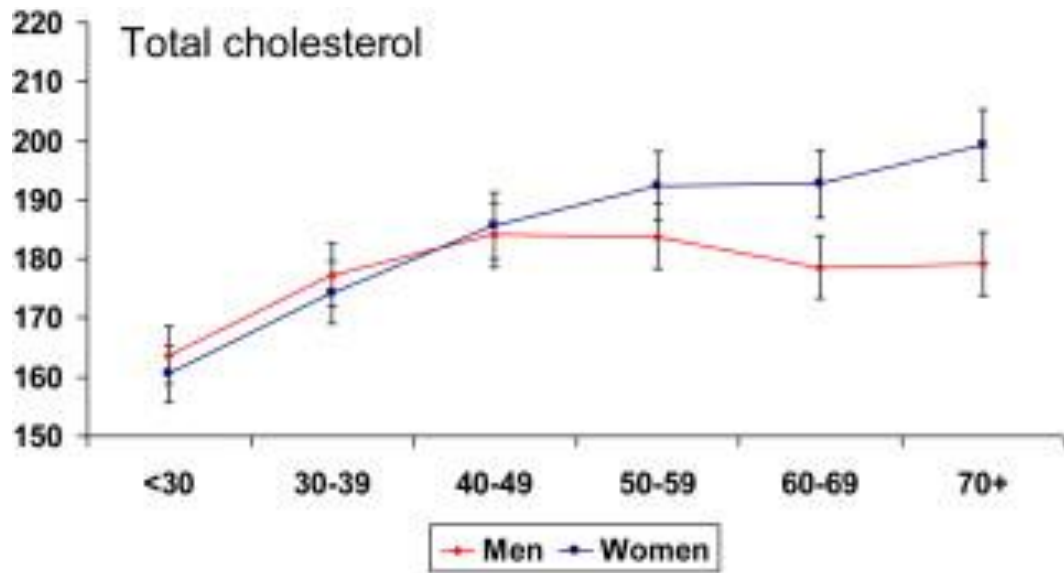


- Error-plot
- HDL-cholesterol in subjects treated with Metformin at baseline and after one year
- 221 African Americans

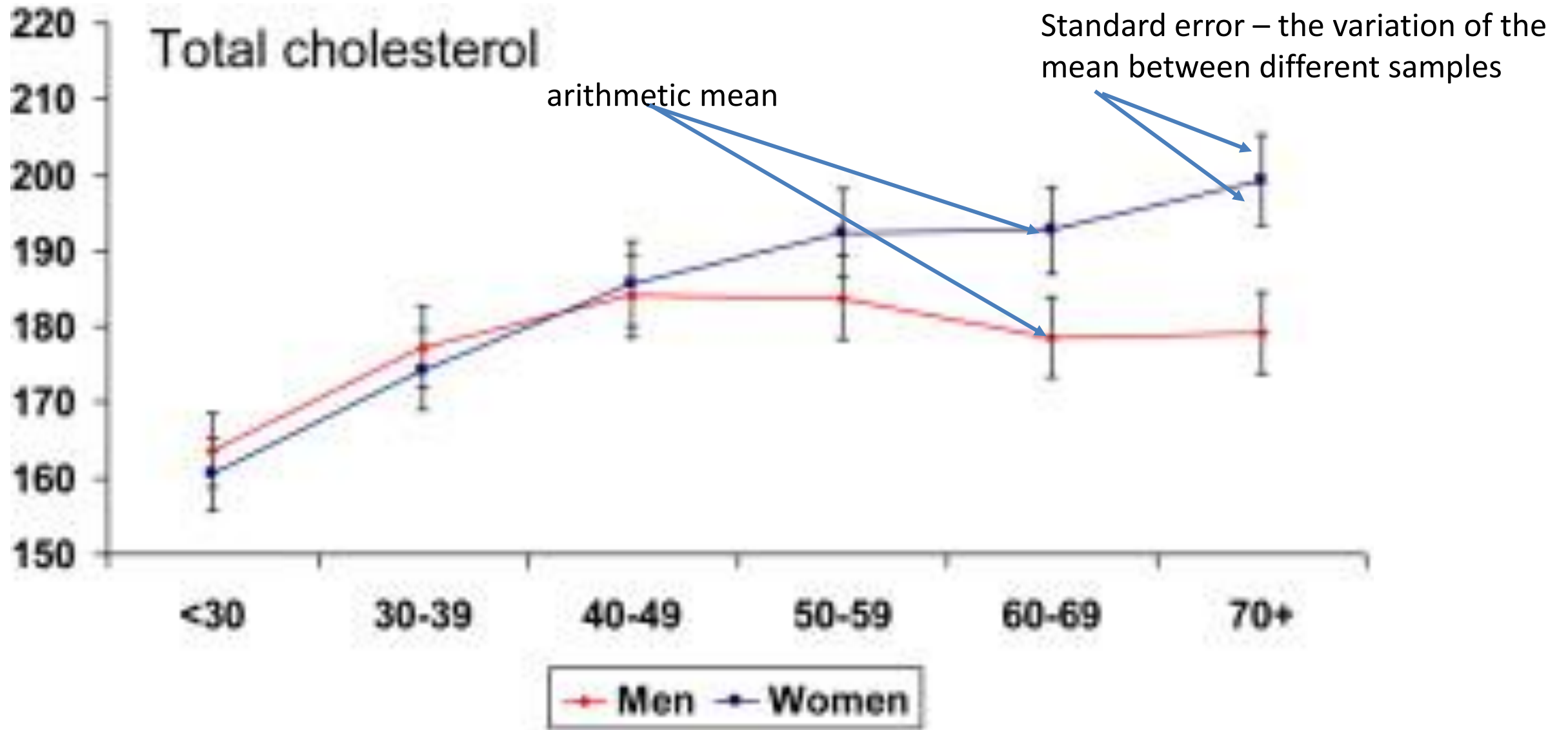
- Error-plot



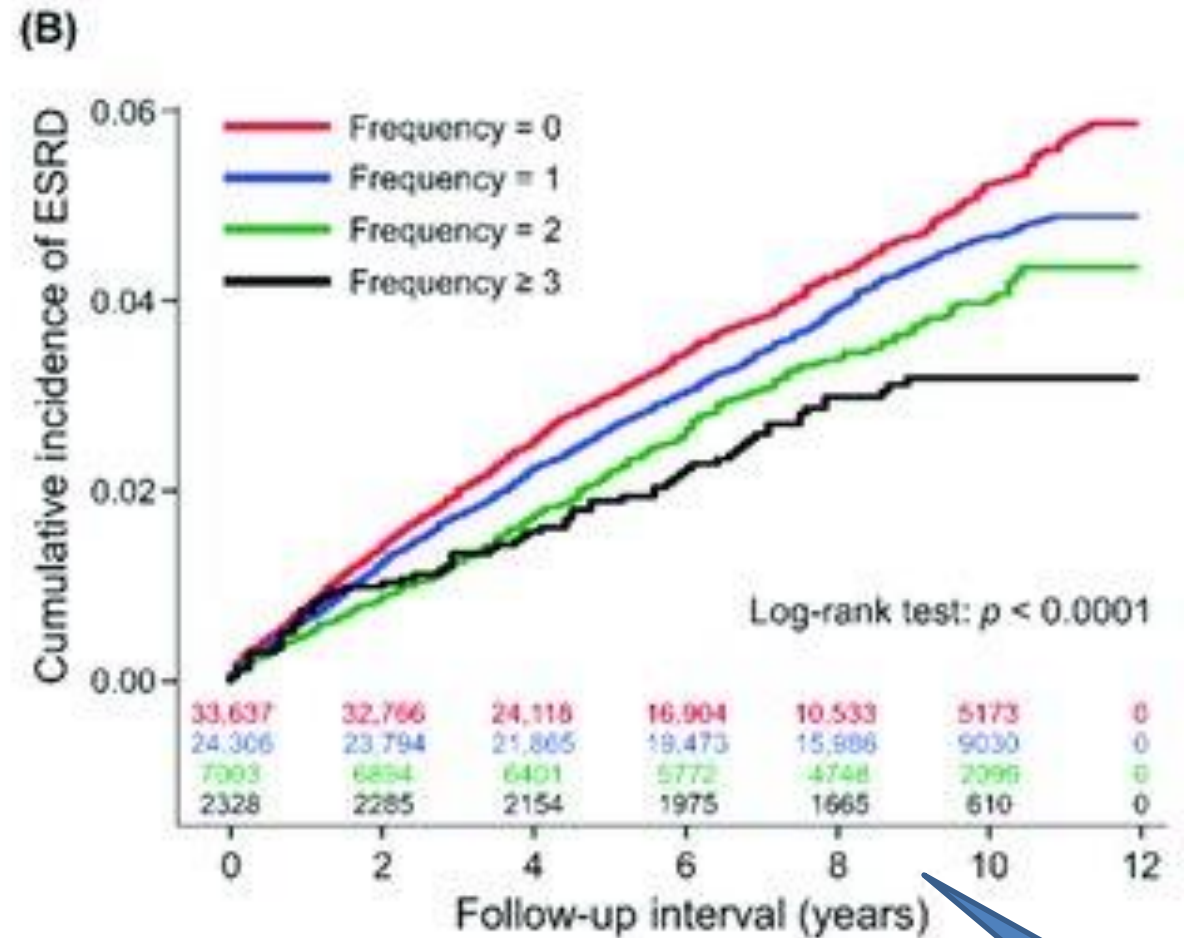
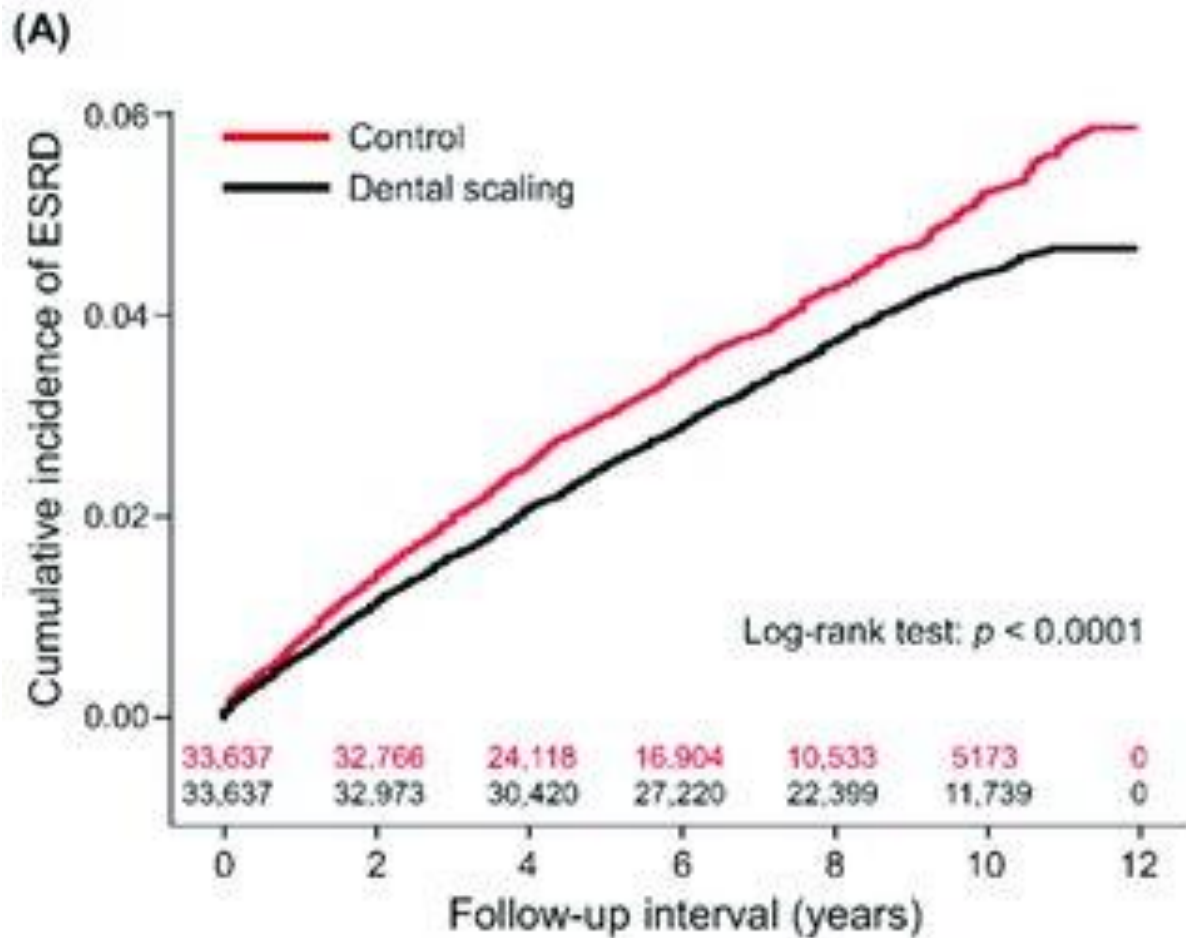
Zhang C, Gao F, Luo H, Zhang CT, Zhang R. Differential response in levels of high-density lipoprotein cholesterol to one-year metformin treatment in prediabetic patients by race/ethnicity. *Cardiovasc Diabetol.* 2015 Jun 12;14:79.



- 6123 subjects (men 3388)
- Age-group specific levels of total cholesterol, triglycerides (mg/dl) in women and men.



Guptha S, Gupta R, Deedwania P, Bhansali A, Maheshwari A, Gupta A, Gupta B, Saboo B, Singh J, Achari V, Sharma KK. Cholesterol lipoproteins and prevalence of dyslipidemias in urban Asian Indians: a cross sectional study. Indian Heart J. 2014 May-Jun;66(3):280-8.



- Cumulative incidences of end stage renal disease between people with periodontitis and without dental scaling treatment (A), and with a varying frequency of dental scaling (B), with the number of subjects at risk.

Time

Summarizing Numerical Data (Continuous, Discrete)



One quantitative variable

- Measure of centrality
 - Arithmetic mean
 - Median
 - Mode
- Measure of spread
 - Variances
 - Standard Deviation
 - Coefficient of variation
 - Standard error
- Measures of asymmetry
 - Skewness
 - Kurtosis
- Measures of location
 - Quartiles
 - Percentiles
- Charts
 - Histograms

Two quantitative variables

- Charts
 - XY Scatter
 - line charts

One qualitative, one quantitative variable

- Charts
 - Box-whiskers
 - Error plot
 - Line

Theoretical exam – example of questions

*In a study the researcher are interested in the oral hygiene of children. They ask 10 children (a sample) about the number of tooth brushing per day. These are their responses: 1, 2, 3, 2, 1, 0, 2, 0, 2, 2. Which of the following are the correct mean, standard deviation and coefficient of variation?

- A. 1.5, 0.97, 0.65
- B. 1.5, 0.92, 0.61
- C. 1.5, 0.97, 0.31
- D. No answer is correct
- E. 1.5, 0.85, 0.57

Theoretical exam – example of questions

*In a study the researcher are interested in the oral hygiene of children. They ask 10 children (a sample) about the number of tooth brushing per day. These are their responses: 1, 2, 3, 2, 1, 0, 2, 0, 2, 2. Which of the following are the correct mean, standard deviation and coefficient of variation?

- A. 1.5, 0.97, 0.65
- **B. 1.5, 0.92, 0.61**
- C. 1.5, 0.97, 0.31
- D. No answer is correct
- E. 1.5, 0.85, 0.57

Theoretical exam – example of questions

*In a study the researcher measure the degree of upper first right premolar inclination for 10 patients. They obtain 10 different values of the tooth inclination, one for each patient: 11, 13, 12, 17, 14, 9, 8, 11, 10, 7. Arithmetic mean is 11.2, standard deviation is 2.97. How many data are in the interval $\text{mean} \pm \text{standard deviation}$ (in percentages)? What is the minimum percentage of data in the interval $\text{mean} \pm \text{standard deviation}$ to have a normal distribution? Which of the following are the correct answers?

- A. 60, 68.3
- B. 100, 95.4
- C. No answer is correct
- D. 70, 68.3
- E. 80, 95.4

Theoretical exam – example of questions

*In a study the researcher measure the degree of upper first right premolar inclination for 10 patients. They obtain 10 different values of the tooth inclination, one for each patient: 11, 13, 12, 17, 14, 9, 8, 11, 10, 7. Arithmetic mean is 11.2, standard deviation is 2.97. How many data are in the interval $\text{mean} \pm \text{standard deviation}$ (in percentages)? What is the minimum percentage of data in the interval $\text{mean} \pm \text{standard deviation}$ to have a normal distribution? Which of the following are the correct answers?

- A. 60, 68.3
- B. 100, 95.4
- C. No answer is correct
- **D. 70, 68.3**
- E. 80, 95.4

Theoretical exam – example of questions

The following statements about median are TRUE:

- A. It is not affected by extreme values
- B. It is a useful parameter for nominal data
- C. Has a poor sample stability
- D. It is affected by the skewed distribution of data
- E. It is useful for qualitative ordinal data

Theoretical exam – example of questions

The following statements about median are TRUE:

- A. It is not affected by extreme values
- B. It is a useful parameter for nominal data
- C. Has a poor sample stability
- D. It is affected by the skewed distribution of data
- E. It is useful for qualitative ordinal data

Theoretical exam – example of questions

*In a study the researcher measure the degree of upper first right premolar inclination for 10 patients. They obtain 10 different values of the tooth inclination, one for each patient: 11, 13, 12, 17, 14, 9, 8, 11, 10, 7. Arithmetic mean is 11.2, standard deviation is 2.97. How many data are in the interval $\text{mean} \pm \text{standard deviation}$ (in percentages)? What is the minimum percentage of data in the interval $\text{mean} \pm \text{standard deviation}$ to have a normal distribution? Which of the following are the correct answers?

- A. 60, 68.3
- B. 100, 95.4
- C. No answer is correct
- D. 70, 68.3
- E. 80, 95.4

Theoretical exam – example of questions

*In a study the researcher measure the degree of upper first right premolar inclination for 10 patients. They obtain 10 different values of the tooth inclination, one for each patient: 11, 13, 12, 17, 14, 9, 8, 11, 10, 7. Arithmetic mean is 11.2, standard deviation is 2.97. How many data are in the interval $\text{mean} \pm \text{standard deviation}$ (in percentages)? What is the minimum percentage of data in the interval $\text{mean} \pm \text{standard deviation}$ to have a normal distribution? Which of the following are the correct answers?

- A. 60, 68.3
- B. 100, 95.4
- C. No answer is correct
- **D. 70, 68.3**
- E. 80, 95.4

Thank you!