



Trainer: Bondor Cosmina-Ioana, PhD, MsC

Descriptive statistics

Variables type



ALWAYS



SEEK



KNOWLEDGE

Objectives for descriptive statistics I

Descriptive and inferential statistics

Population, sample

Variables, Measurements scale, Variable type

Data vs. informations

Qualitative variables: summarize method (absolute and relative frequencies), proportion, rate, ratio

Frequency and contingency tables

Graphical representations (Pie, Column, Bar).

General principles for table and charts

Statistics definition

Statistics is a field of science that deals with the study of the collectivities (people, objects, etc.) with respect to at least a common variable characteristic or attribute.

Statistical fields

Descriptive statistics - describe the characteristics of a community: averages, frequencies, and describes variability, the spread of those characteristics

Analytic (Inferential) Statistics – allow to generalize from some cases to all similar individuals, allows comparisons between two populations

Descriptive Statistics

Data collection

Summarize data

- Tables
- Charts

Calculating some synthetic **parameters**

- Arithmetic mean, median
- Variance, standard deviation.

Population - Sample

Population –all individuals who share the same characteristic

ex. population of diabetic people,

ex. females

Sample – a group of people



Glycemia – a characteristic

- blood glucose concentration

healthy subjects

tested in the morning before food or water – glycemia a Jeun (fasting glycemia)

subjects with diabetes

tested at any time during the day (e.g. postprandial if it's taken after a meal)

– possible values in an interval
... 20-650 ... mg/dl



What data can be memorized about - Glycemia

- Ex. 3 different possibilities:
 - **measuring – record the measured value**
 - possible values interval
 - ... 20-650 ... mg/dl
 - **measuring, divide into categories**
 - record the category
 - < 70 mg/dl - hypoglycemia
 - 70 – 120 mg/dl normal glycemia
 - 120 – 180 mg/dl prediabetes
 - ≥ 180 mg/dl diabetes
 - measuring, then divide in only 2 categories
 - record Diabetes (yes/no)
 - if glycemia is tested twice ≥ 180 mg/dl

	A	B	C	D
			Glycemia classification (hypoglycemia, normal glycemia, prediabetes, diabetes)	Diabetes (Yes/no)
1	Id_patient	Glycemia		
2	1	70	normal glycemia	No

Variables and data

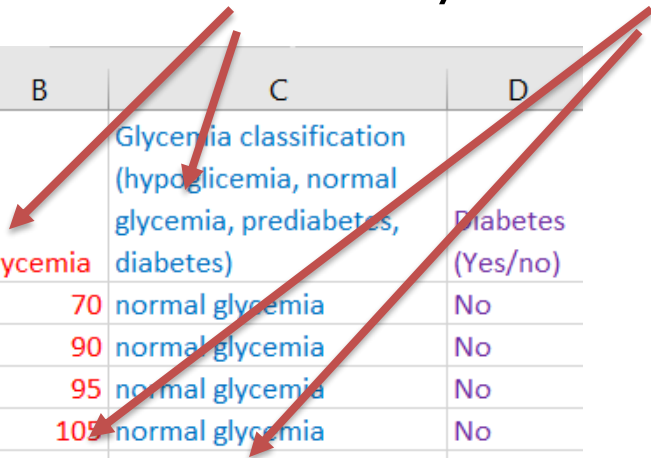
Variable = A characteristic of an individual – an attribute

It may take different values (**data**) depending on the individual from a **set of possible values**.

The opposite of a variable is a constant - a fixed value.

	A	B	C	D
			Glycemia classification (hypoglycemia, normal glycemia, prediabetes, diabetes)	Diabetes (Yes/no)
1	Id_patient	Glycemia		
2	1	70	normal glycemia	No
3	2	90	normal glycemia	No
4	3	95	normal glycemia	No
5	4	105	normal glycemia	No
6	5	120	prediabetes	No
7	6	200	diabetes	Yes
8	7	250	diabetes	Yes
9	8	150	prediabetes	No
10	9	100	normal glycemia	No
11	10	90	normal glycemia	No
12	11	70	normal glycemia	No
13	12	80	normal glycemia	No
14	13	90	normal glycemia	No
15	14	120	prediabetes	No
16	15	150	prediabetes	No
17	16	205	diabetes	Yes
18	17	95	normal glycemia	No
19	18	105	normal glycemia	No
20	19	75	normal glycemia	No
21	20	50	hypoglycemia	No

Variables / data examples



	A	B	C	D
		Glycemia	Glycemia classification (hypoglycemia, normal glycemia, prediabetes, diabetes)	Diabetes (Yes/no)
1	Id_patient			
2	1	70	normal glycemia	No
3	2	90	normal glycemia	No
4	3	95	normal glycemia	No
5	4	105	normal glycemia	No
6	5	120	prediabetes	No
7	6	200	diabetes	Yes
8	7	250	diabetes	Yes
9	8	150	prediabetes	No
10	9	100	normal glycemia	No
11	10	90	normal glycemia	No
12	11	70	normal glycemia	No
13	12	80	normal glycemia	No
14	13	90	normal glycemia	No
15	14	120	prediabetes	No
16	15	150	prediabetes	No
17	16	205	diabetes	Yes
18	17	95	normal glycemia	No
19	18	105	normal glycemia	No
20	19	75	normal glycemia	No
21	20	50	hypoglycemia	No

Variables

- Glycemia
- Glycemia classification
- Diabetes

Data

- 100, 90, 50, 205 etc.
- hypoglycemia, normal glycemia, prediabetes, diabetes
- Yes, No

When work with patient data: Ensure Ethical and Legal Framework

Obtain informed consent from patients or participants.

Adhere to regulations such as:

GDPR (General Data Protection Regulation - EU)

HIPAA (Health Insurance Portability and Accountability Act - US)

Institutional Review Board (IRB) approvals for research

Respect patient autonomy and confidentiality

Is there a difference between these variables/data?

B	C
Glycemia	Glycemia classification (hypoglycemia, normal glycemia, prediabetes, diabetes)
70	normal glycemia
90	normal glycemia
95	normal glycemia
105	normal glycemia
120	prediabetes
200	diabetes
250	diabetes
150	prediabetes
100	normal glycemia
90	normal glycemia
70	normal glycemia
80	normal glycemia
90	normal glycemia
120	prediabetes
150	prediabetes
205	diabetes
95	normal glycemia
105	normal glycemia
75	normal glycemia
50	hypoglycemia

Data can be

- a number → numerical, quantitative
- a quality → qualitative

B	C
	Glycemia classification (hypoglycemia, normal glycemia, prediabetes, diabetes)
Glycemia	
70	normal glycemia
90	normal glycemia
95	normal glycemia
105	normal glycemia
120	prediabetes
200	diabetes
250	diabetes
150	prediabetes
100	normal glycemia
90	normal glycemia
70	normal glycemia
80	normal glycemia
90	normal glycemia
120	prediabetes
150	prediabetes
205	diabetes
95	normal glycemia
105	normal glycemia
75	normal glycemia
50	hypoglycemia

Numerical scale of measurement

Individuals' data are evaluated on **scales of measurement**

Interval scale

- quantitative characteristics, measurable in real numbers in a range

Ratio scale

- quantitative characteristics, measurable in real numbers, which can be divided

Discrete scale

- quantitative characteristics, measurable in integer numbers, decimals are meaningless

Ex: Interval: glycemia Ratio: weight, age Discrete: number of children in a family

Usually - have a unit of measurement

Answer the question: How much/many?

The difference between any two units is the same.

If we make a measurement,
we obtained a

quantitative variable

Measurement, counting  number  quantitative variable

Quantitative variable (numerical)

continuous

discrete

Continuous quantitative variable

- It can be measured with decimals
 - ex. Weight (ex.13.5 kg), Age (time is continuous),
 - ex. Total cholesterol (204.3 mg/dL)

Discrete quantitative variable

- decimals are meaningless
- ex. Number of children (2.5 children???),
- ex. Number of pregnancies (1.3 ???)

G
Tooth number
12
13
12
14
12
13
14
13
12
14
9
12
10
10
12
14
10
15
14
14

B	J
Glycemia	Cranian perimeter (cm)
70	34
90	74
95	24
105	66
120	38
200	74
250	45
150	41
100	20
90	38
70	56
80	34
90	20
120	20
150	38
205	66
95	20
105	49
75	34
50	59
	31
	41

Example - Glycemia

– continuous quantitative variable

name of the variable: Glycemia

range of possible values

... ~50-650... mg/dl

we can measure it with a device

it have unit of measurement mg/dl

we cannot observe it

Decimals has meaning

Answer the question: How much is it?

The difference between any two units is the same.

e.g. between 50 and 51 the difference is the same as between 100 and 101

B
Glicemie
78
99
110
65
200
90
85
87
130
97
77
110
125
220

Example - Body height

– continuous quantitative variable

variable name: Body height

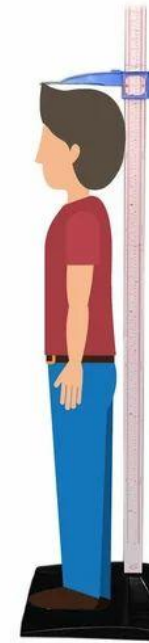
range of possible values: ...45-210... cm

We can measure it with a device

has unit of measure: cm

Decimals has meaning

Answer the question: How much is it?



K
Height (cm)
54
55
60
55
45
60
59
51
60
62
58
58
57
61
62
63
61
58
59
55

The difference between any two units is the same.

e.g. between 50 and 51 cm the difference is the same as between 100 and 101 cm

Nominal scale

Nominal scale - qualitative characteristics, qualities of an individual

Ex. Diabetes, Profession

- not measurable
- observable/classified/transform
- the data cannot be ordered

Answer the questions:

- How is it?
- what is it
- ...

e.g. viruses or bacteria in a cell culture

D	C
Diabetes (Yes/no)	Profession
No	pupil
No	studenta
No	computer sci
No	economist
No	unemployed
Yes	banker
Yes	unemployed
No	worker
No	worker
No	salesman
No	student
No	teacher
No	weiter
No	salesman
No	worker
Yes	unemployed
No	computer sci
No	ospatar
No	weiter
No	student

Ordinal scale

Ordinal scale

- qualitative characteristics, qualities of an individual
- Nominal data **with an order between data categories**

Ex. Stages I-IV of cancer, rheumatoid arthritis class I to IV

- The difference between the categories **is not the same**
- e.g. Mark at informatics 0-10

Measured on the ordinal scale

- scores
- percentages
- proportions

Apgar Score
10
10
9
10
10
8
8
9
9
10
10
10
10
6
10
10
10
7
7
8

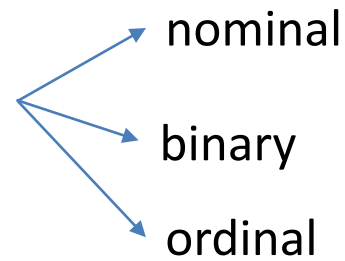
If we have observed/measured
and/or classified (divided) into
categories, then the values belong to
one

Qualitative variable

C	C	D
Glycemia classification (hypoglycemia, normal glycemia, prediabetes, diabetes)	Profession	Diabetes (Yes/no)
normal glycemia	pupil	No
normal glycemia	studenta	No
normal glycemia	computer sci	No
normal glycemia	economist	No
normal glycemia	unemployed	No
prediabetes	banker	No
diabetes	unemployed	Yes
diabetes	worker	Yes
prediabetes	worker	No
normal glycemia	salesman	No
normal glycemia	student	No
normal glycemia	teacher	No
normal glycemia	weiter	No
normal glycemia	salesman	No
normal glycemia	worker	No
prediabetes	unemployed	No
prediabetes	computer sci	No
diabetes	ospatar	Yes
normal glycemia	weiter	No
normal glycemia	student	No
normal glycemia	dr	No
normal glycemia	salesman	No
hypoglycemia		No

Observed / divide in categories  qualitative variable

Qualitative variables



Nominal variable

qualitative characteristics, qualities of an individual
with **more than two** categories, no order

Binary (Dichotomous) variable

qualitative characteristics, qualities of an individual
with **two categories**

Ordinal variable

qualitative characteristics, qualities of an individual
has categories that **can be ordered**

C	H	Apgar Score
Profession	Surgery	10
pupil	No	10
studenta	No	9
computer sci	No	10
economist	No	8
unemployed	No	8
banker	No	9
unemployed	No	9
worker	No	10
worker	No	10
salesman	No	10
student	No	10
teacher	No	10
weiter	No	10
salesman	No	6
worker	No	10
unemployed	No	10
computer sci	No	10
ospatar	No	10
weiter	Yes	7
student	Yes	7
dr	Yes	8
salesman		

Example - Maxillary

- Dichotomial qualitative variable

variable name: "Maxillary"

possible values (data): Superior/Inferior (two categories)

can't be measure (it has no unit of measure)

can be observe

The possible values cannot be ordered

D
Maxillary
superior
inferior
superior
inferior
superior
inferior
superior
inferior
inferior
superior
superior
inferior
inferior
inferior
inferior
inferior
superior
inferior
superior
superior

Example - Dentition

- Nominal qualitative variable

name of the variable “Dentition”

possible values (categories)

permanent, temporary, mixt, etc. (more than two categories)

can't be measure (it has no unit of measure)

can be observe

The possible values cannot be ordered

F
Dentition
mixt
permanent
permanent
permanent
mixt
temporary
permanent
permanent
mixt
permanent
mixt
permanent
temporary
temporary
mixt
mixt
temporary
permanent
permanent
permanent

Ordinal qualitative variable

Ex. Any score

- depression score etc.
- Any grade
 - grade in chemistry, physics, etc.

C	Apgar Score
Glycemia classification (hypoglycemia, normal glycemia, prediabetes, diabetes)	10
normal glycemia	10
normal glycemia	9
normal glycemia	10
normal glycemia	10
normal glycemia	8
prediabetes	8
diabetes	9
diabetes	9
prediabetes	10
normal glycemia	10
normal glycemia	10
normal glycemia	10
normal glycemia	10
normal glycemia	6
prediabetes	10
prediabetes	10
diabetes	10
normal glycemia	7
normal glycemia	7
hypoglycemia	8

Summary - Variables

Qualitative variables type are:

- Dichotomial

- Ordinal

- Nominal

Numerical/Quantitative:

- Continuous

- Discrete



supplementary learning: Best Practices in Medical Data Collection Prepared for Statistical Analysis

- Plan
 - The variables
 - The type of the variables
 - How missing data will be handled - Collect the reasons
 - e.g. "patient refused", "not applicable"
- Define
 - the variables. Use dictionaries
 - the allowable data
 - use uniform coding
 - ex. use "Yes/No" or "Y/N" for the same variable. Not both.
- Use table data forms
 - e.g. rows = observations, columns = variables
 - Avoid free-text fields
- collect sufficient metadata: when/how/device/method
 - e.g. "blood pressure measured seated after 5 min rest"
- maintain logs of data cleaning or transformations
- collect all potential factors
 - e.g. age, sex, other diseases, etc.
- collect basic data, not the transformed/calculated data
 - e.g. we prefer glycemia in mg/dl
 - we can always change it to diabetes/no diabetes
 - the reverse is not valid
 - e.g. collect weight and height, instead of body mass index (BMI)
 - $BMI = \text{weight (kg)} / \text{height (m}^2\text{)}$

Data versus Information

Data

- raw facts or values
- not processed
- not interpreted

Information

- data
- processed,
- organized,
- given meaning

Descriptive statistics for qualitative variables

Objectives

Summarizing qualitative variables

Frequencies

Tables

Charts

Gender	Dentition	Apgar Score
M	mixt	
M	permanent	10
F	permanent	10
F	permanent	9
F	mixt	10
F	temporary	10
M	permanent	8
F	permanent	8
M	mixt	9
M	permanent	9
F	mixt	10
F	permanent	10
F	temporary	10
F	temporary	10
F	mixt	6
F	mixt	10
M	mixt	10
F	temporary	10
M	permanent	7
F	permanent	7
F	permanent	8

Represent qualitative variables

- dichotomial variables
- nominal variables
- ordinal variables with few categories



Please describe these raw data

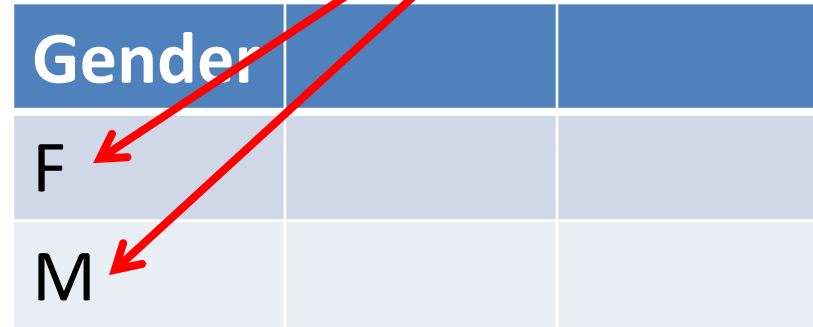
What can you say?

Gender

M
M
F
F
F
F
M
F
M
M
F
F
F
F
F
F
M
F
M
F
F

Frequency table: Dichotomial variable

Gender have two possible values in our data (two categories)



Gender		
F		
M		

we get 20 patients:
these are the data
about their gender

Gender
M
M
F
F
F
F
F
M
F
M
M
F
F
F
F
F
F
M
F
M
F
F

20
subjects

Frequency table: Dichotomial variable

Gender	Number = Absolute frequency	
F	7	
M	13	
Total	20	

How many female?

How many male?

Verify: $7+13=20$ – total number of subjects

Gender

M
M
F
F
F
F
F
M
F
M
M
F
F
F
F
F
M
F
M
F
F

Frequency table: Dichotomial variable

Female percentage = $7/20 \times 100 = 35\%$

Gender	Number = Absolute frequency	Percentage (%) = Relative frequency
F	7	35
M	13	65
Total	20	100

Male percentage = $13/20 \times 100 = 65\%$

Verify: $35 + 65 = 100\%$

Gender

M
M
F
F
F
F
F
M
F
M
M
M
F
F
F
F
F
F
M
F
M
F
F

Chart – Pie chart

Gender	Number = Absolute frequency	Percentage (%) = Relative frequency
F	7	35
M	13	65
Total	20	100

Gender

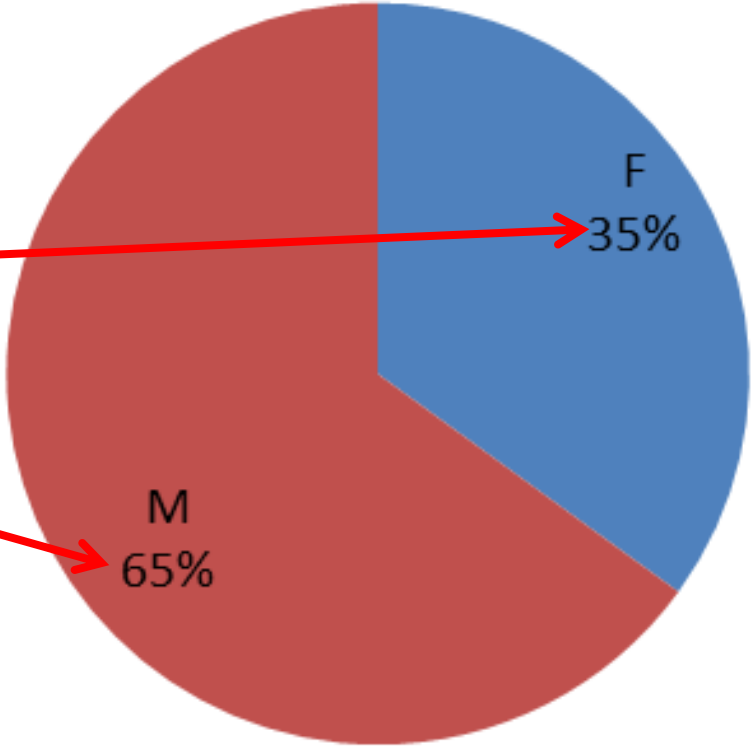


Figure. Gender percentages distribution

Gender

M
M
F
F
F
F
F
M
F
M
M
F
F
F
F
F
F
M
F
M
F
F

Chart – Pie chart

Gender	Number = Absolute frequency	Percentage (%) = Relative frequency
F	7	35
M	13	65
Total	20	100

Title of the chart

Percentages for all categories

Categories titles

Gender

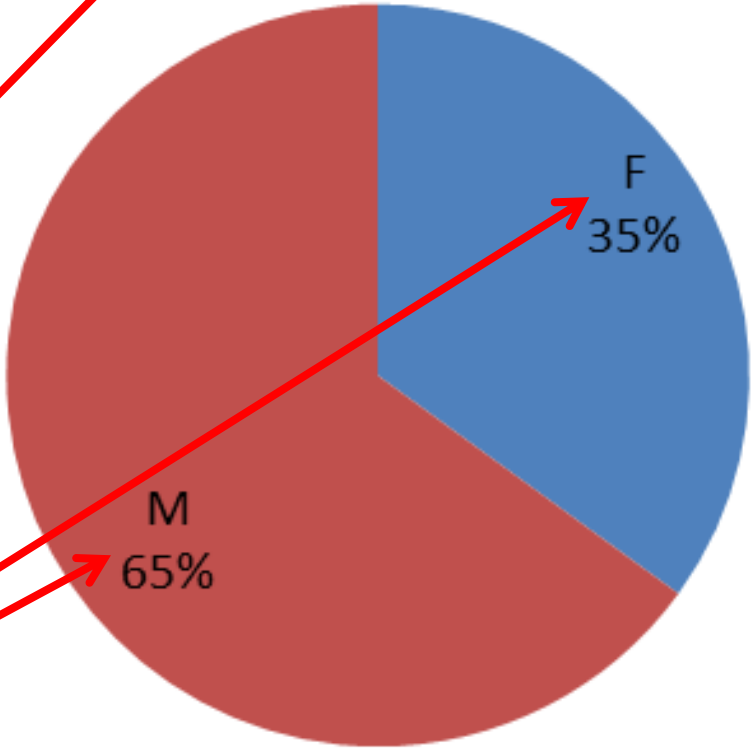


Figure. Gender percentages distribution

Gender

M
M
F
F
F
F
M
F
M
M
F
F
F
F
F
M
F
M
F
F

Chart – Column chart

Gender	Number = Absolute frequency	Percentage (%) = Relative frequency
F	7	35
M	13	65
Total	20	100

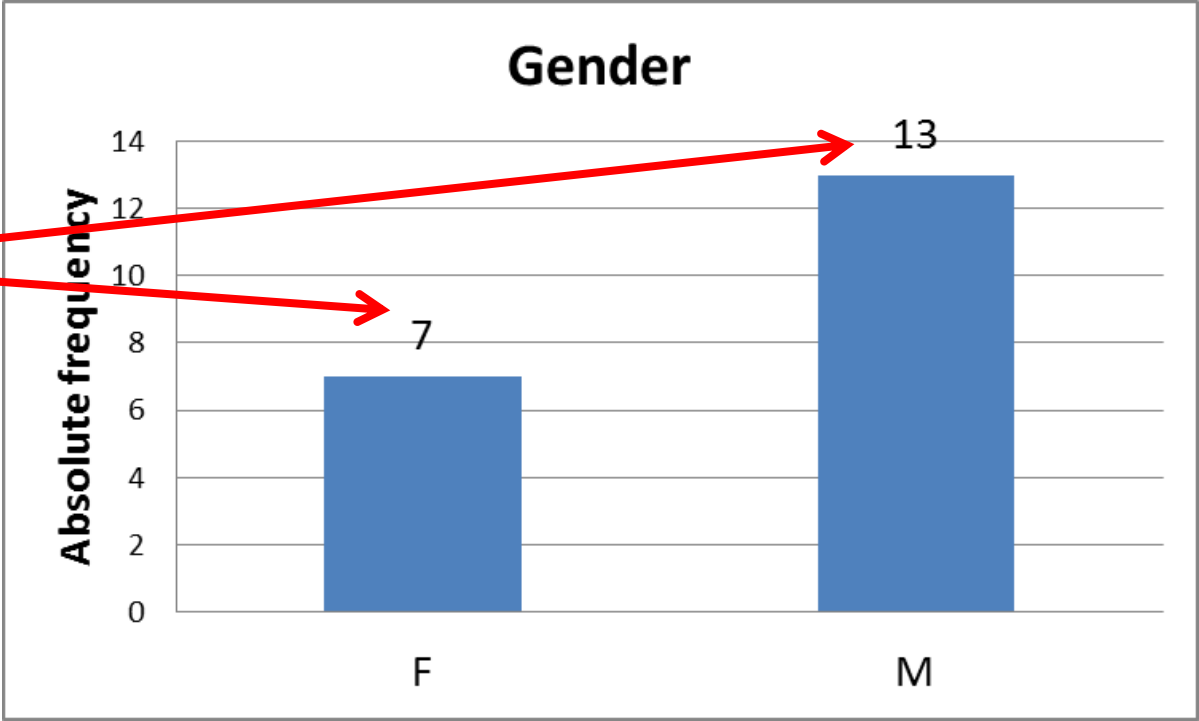


Figure. Gender distribution

Gender

M
M
F
F
F
F
F
M
F
M
M
F
F
F
F
F
M
F
M
F
F

Chart – Column chart

Gender	Number = Absolute frequency	Percentage (%) = Relative frequency
F	7	35
M	13	65
Total	20	100

Title of the chart

Axis title

Values

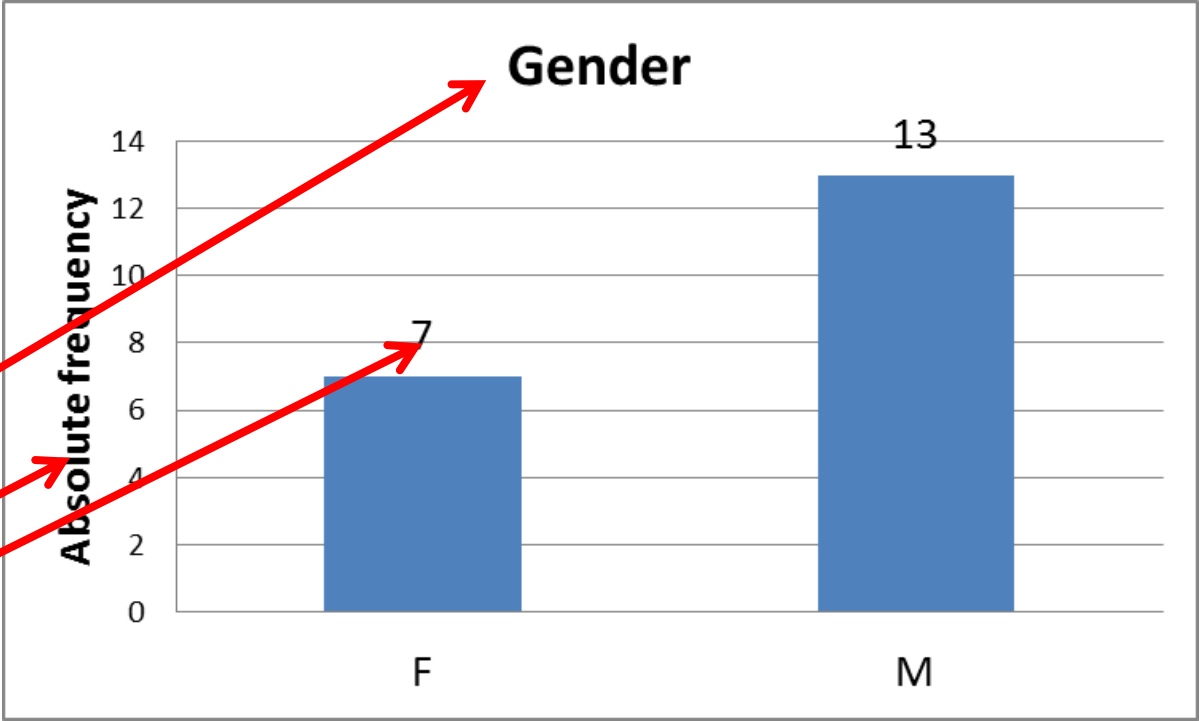


Figure. Gender distribution

Dentition

mixt
permanent
permanent
permanent
mixt
temporary
permanent
permanent
mixt
permanent
mixt
permanent
temporary
temporary
mixt
mixt
temporary
permanent
permanent
permanent

Frequency table: Nominal variable

How many people?

Dentition	Absolute frequency	Relative frequency
temporary	4	20
permanent	10	50
mixt	6	30
Total	20	100

Percentages

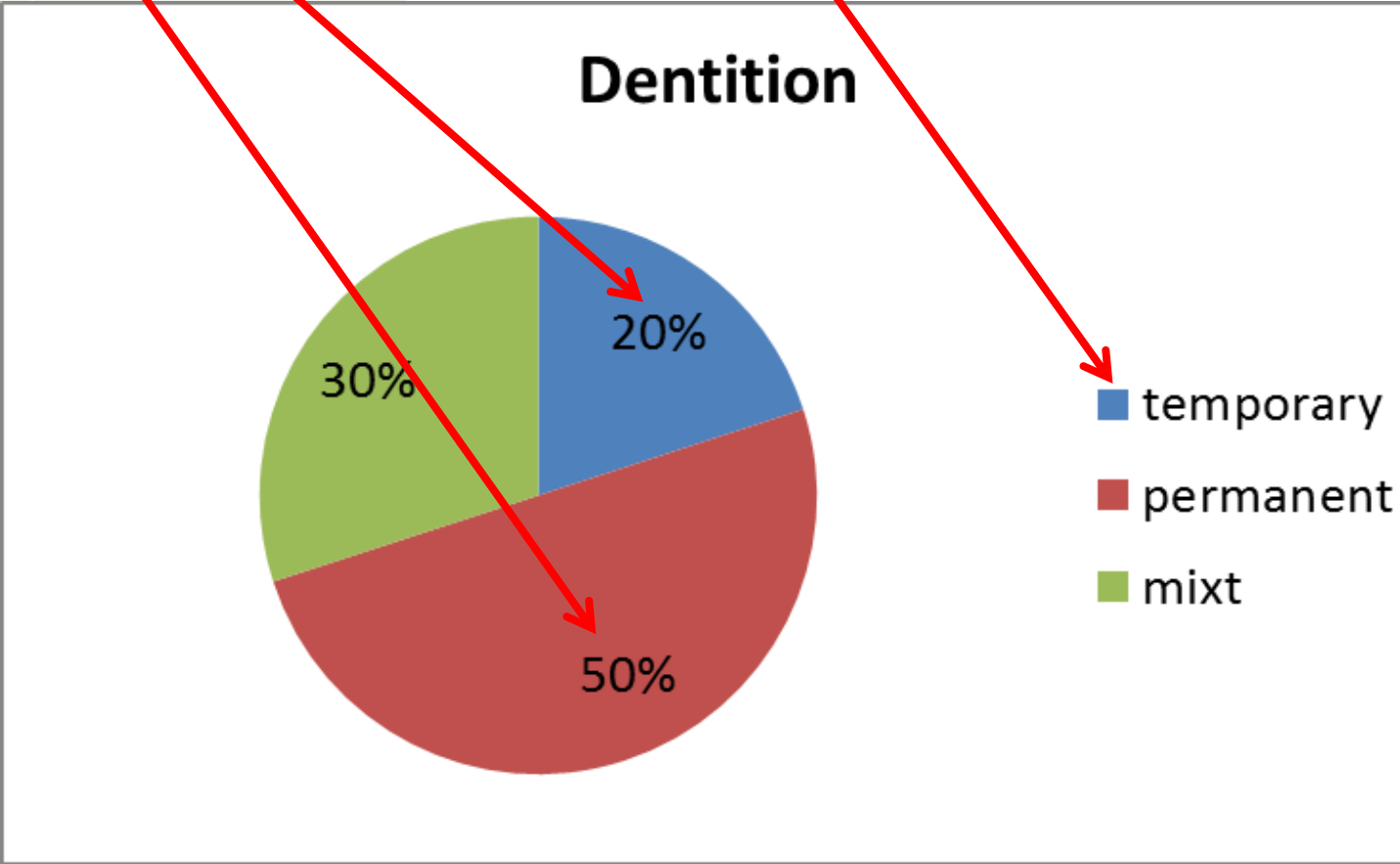
Dentition – 3 possible categories: temporary, permanent, mixt

Dentition
mixt
permanent
permanent
permanent
mixt
temporary
permanent
permanent
mixt
permanent
mixt
permanent
temporary
temporary
mixt
mixt
temporary
permanent
permanent
permanent

Dentition	Absolute frequency	Relative frequency
temporary	4	20
permanent	10	50
mixt	6	
Total	20	

Chart – Pie chart

Legend



Dentition
mixt
permanent
permanent
permanent
mixt
temporary
permanent
permanent
mixt
permanent
mixt
permanent
temporary
temporary
temporary
mixt
mixt
temporary
permanent
permanent
permanent

Dentition	Absolute frequency	Relative frequency
temporary	4	20
permanent	10	50
mixt	6	30
Total	20	100

Chart – Bar chart

- ✓ Title of the chart
- ✓ Axis title
- ✓ Values



Summarizing nominal & ordinal data

Absolute frequency of category “A” = a = number of apparition of category “A”

Relative frequency in percentages of category “A” = $a/n * 100$
(where n is the total number of subjects)

Relative frequency of category “A” = a/n

Table of frequencies – a structured table that displays frequencies for all categories of a qualitative variable

Summarizing nominal & ordinal data

Ratio - a comparison between two frequencies that shows how many persons there is compared with other

ex. 2 boys for 4 mothers

Rate - the frequency with which an characteristic occurs over a specified period of time

ex. 5 new cases of a disease per 1,000 people per year.

Proportion - the frequency of a particular characteristic divided by the total frequency of the population or group. It represents the part of the whole.

ex. If 20 out of 100 people have a disease, the proportion is $20/100 = 0.20$ (20%).

Two Qualitative Variables

How we understand and organize the data when we investigate the relationship between two qualitative characteristics?

Contingency Tables

— the way data are organized when we have two qualitative variables in some relationship

One independent= “the cause”

One dependent = “the effect”

What Is a Contingency Table?

a frequency table to organize data for two qualitative variables

Shows:

Joint frequencies

Totals on columns/rows

Grand total

What Is a Contingency Table?

a frequency table to organize data for two qualitative variables

Shows:

Joint frequencies

Totals on columns/rows

Grand total

	Variable 2 – category 1	Variable 2 – category 2	Total
Variable 1 – category 1	a	b	a+b
Variable 1 – category 2	c	d	c+d
Total	a+c	b+d	n=a+b+c+d

Contingency table for two dichotomial (binary) variable

Two Dichotomial Variables

Is Stress related to Sleeping (≥ 8 hours)

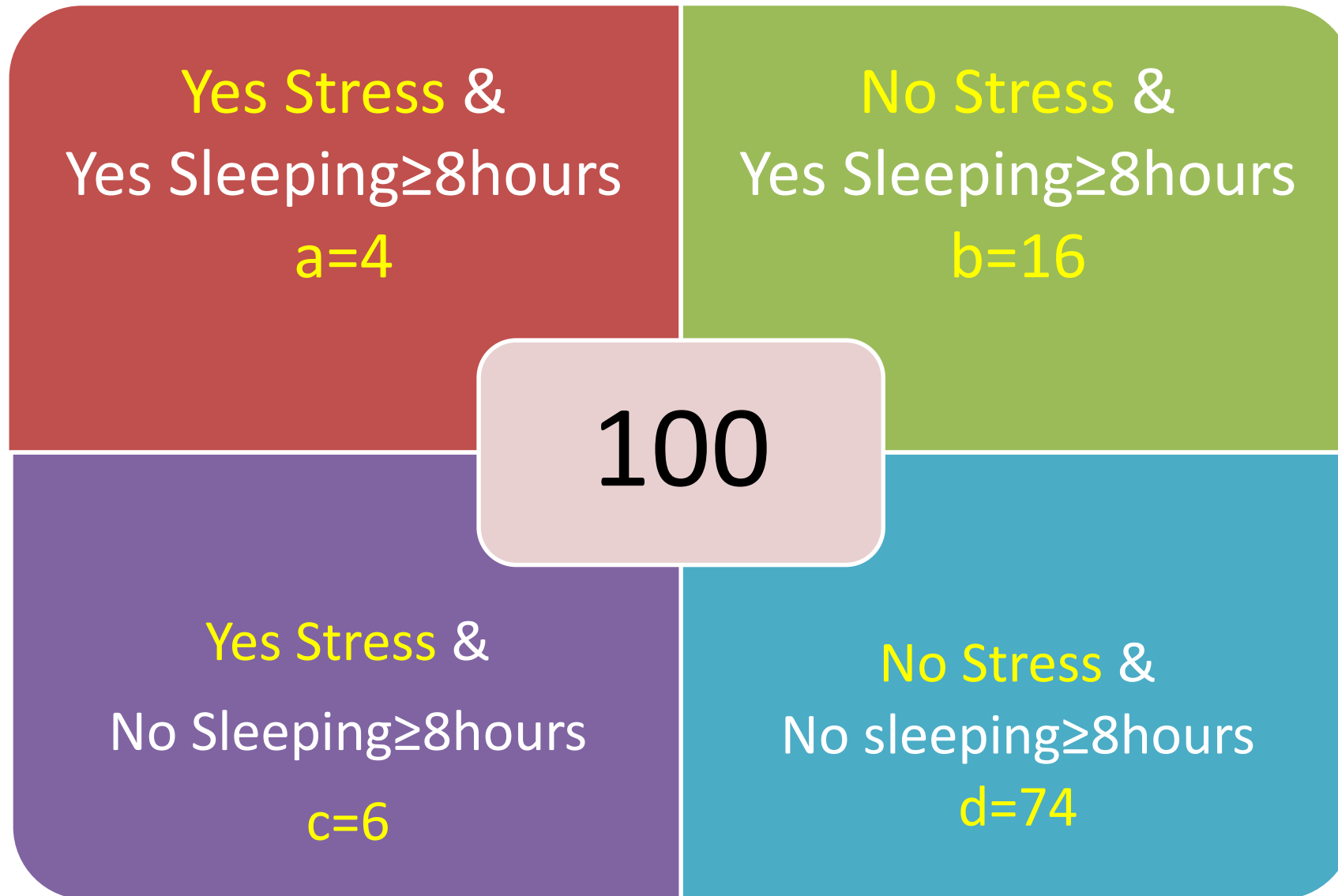
“Are you feeling stressed?”

“Did you sleep enough last two-three nights (≥ 8 hours)?”

	Stress Yes	Stress No	Total
Sleeping (≥ 8 hours) Yes	4	16	20
Sleeping (≥ 8 hours) No	6	74	80
Total	10	90	100

Contingency table for two dichotomial (binary) variable

2 dichotomial variables both with 2 categories → 4 possible combined categories



a, b, c, d –
absolute
frequencies

Two Dichotomial Variables

“Are you feeling stressed?”

Stress	Frequency
Yes	10
No	90
Total	100

“Did you sleep enough last two-three nights (≥ 8 hours)?”

Sleeping (≥ 8 hours)	Frequency
Yes	20
No	80
Total	100

	Stress Yes	Stress No	Total
Sleeping (≥ 8 hours) Yes	4	16	20
Sleeping (≥ 8 hours) No	6	74	80
Total	10	90	100

Contingency table for two dichotomial (binary) variable

Chart to show the **absolute frequencies** (column chart)

	Stress Yes	Stress No	Total
Sleeping (≥ 8 hours) Yes	4	16	20
Sleeping (≥ 8 hours) No	6	74	80
Total	10	90	100

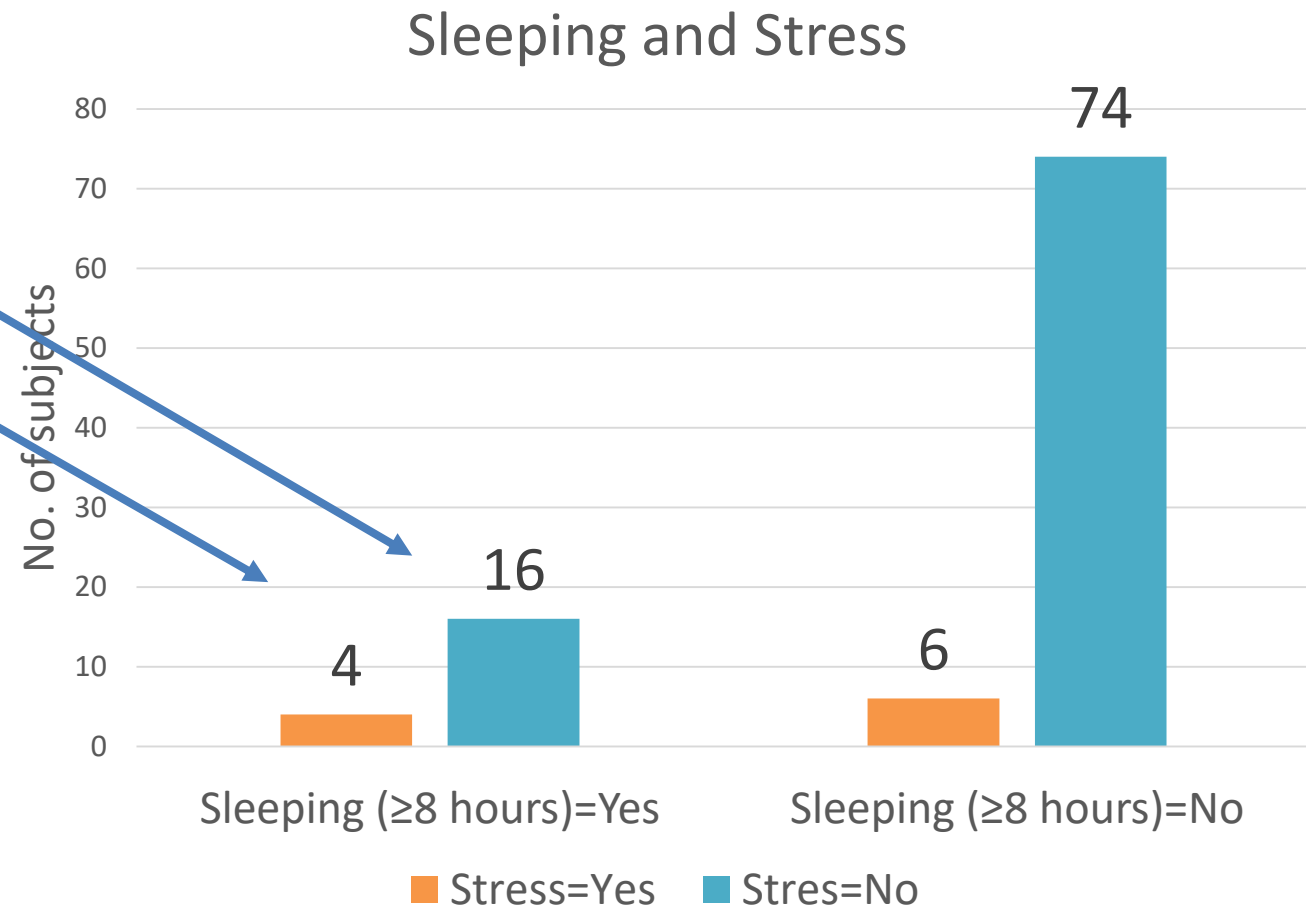
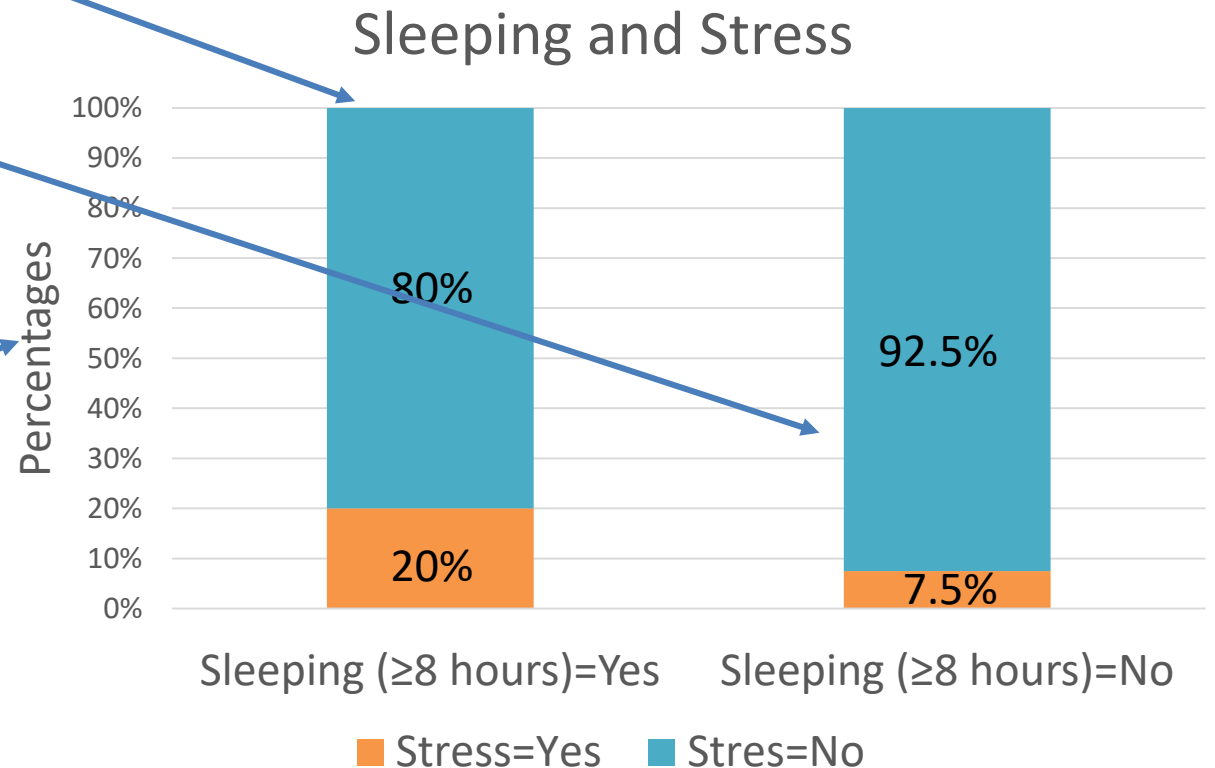


Chart to show the **percentages** (relative frequencies – 100% stacked column chart)

	Stress Yes	Stress No	Total
Sleeping (≥ 8 hours) Yes	4	16	20
Sleeping (≥ 8 hours) No	6	74	80
Total	10	90	100

- each row from the contingency table is a column representing 100%
- 20 is the total for the first column
- 80 is the total for the second column

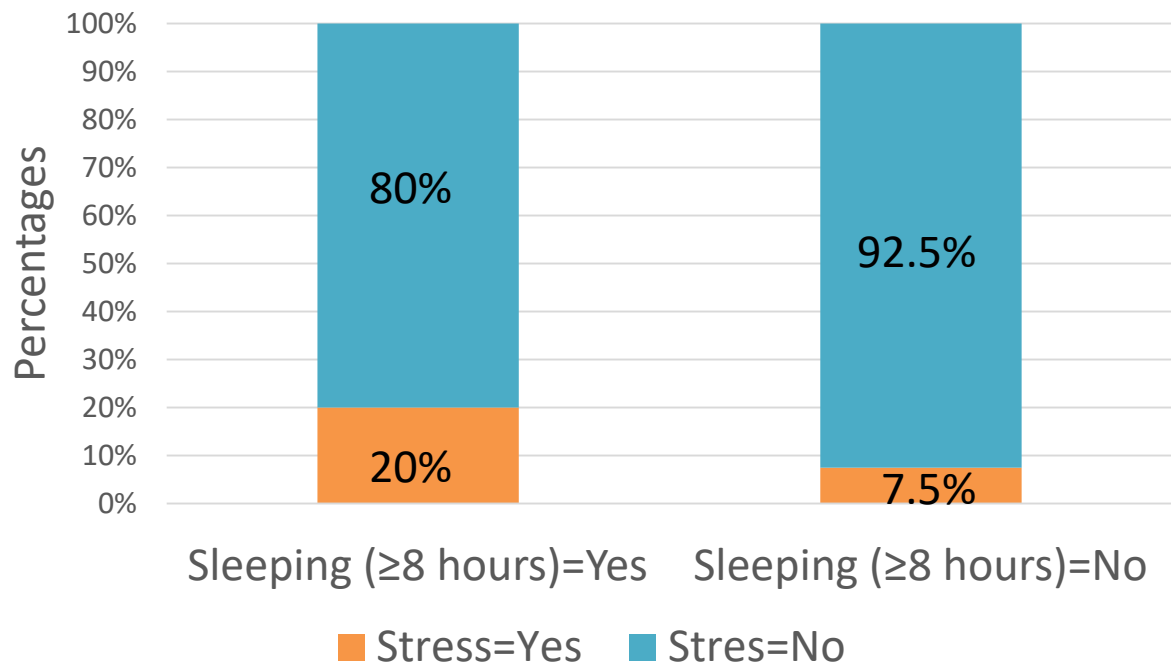


Common Mistake

The charts in % do not look the same

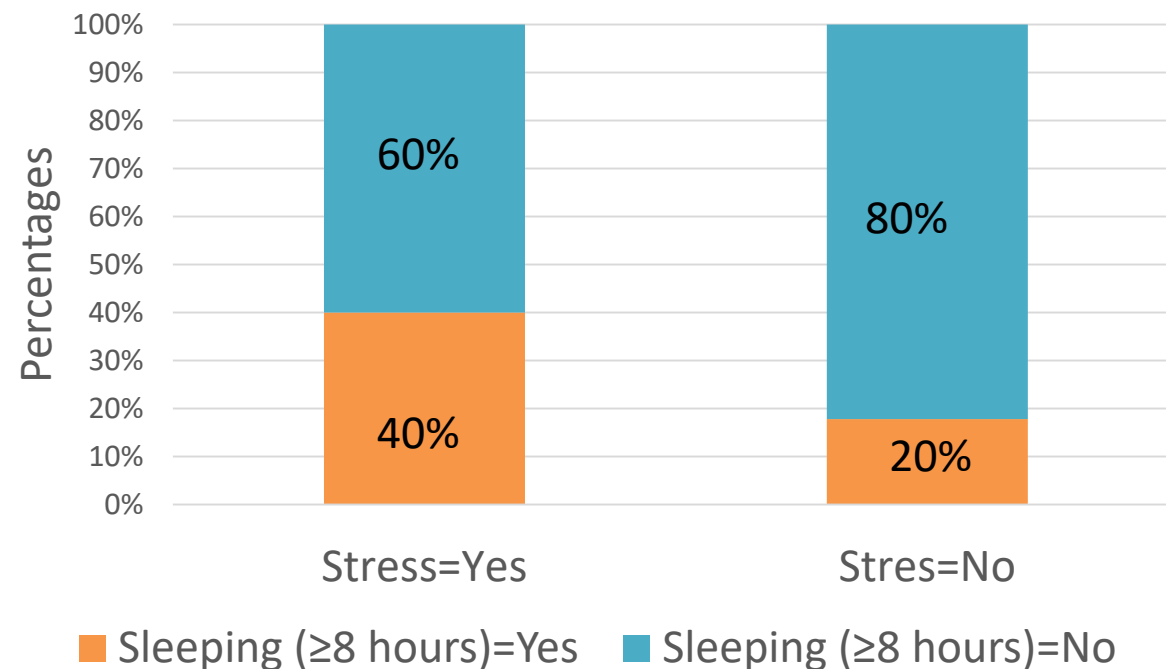
	Stress = Yes	Stress = No	Total
Sleeping (≥8hours) = Yes	4	16	20
Sleeping (≥8hours) = No	6	74	80
Total	10	90	100

Sleeping and Stress



	Sleeping (≥8hours) Yes	Sleeping (≥8hours) No	Total
Stress Yes	4	6	10
Stress No	16	74	90
Total	20	80	100

Sleeping and Stress



nominal variable with 3 categories

ordinal variable with 3 categories

B	C
Hepatitis type (Type A / Type B / Type C)	Obesity (normal weight/overweight /obese)
Type C	Normal weight
Type C	Normal weight
Type C	Normal weight
Type B	Overweight
Type C	Normal weight
Type C	Normal weight
Type C	Normal weight
Type B	Overweight
Type A	Overweight
Type A	Overweight
Type B	Obese
Type A	Obese

...

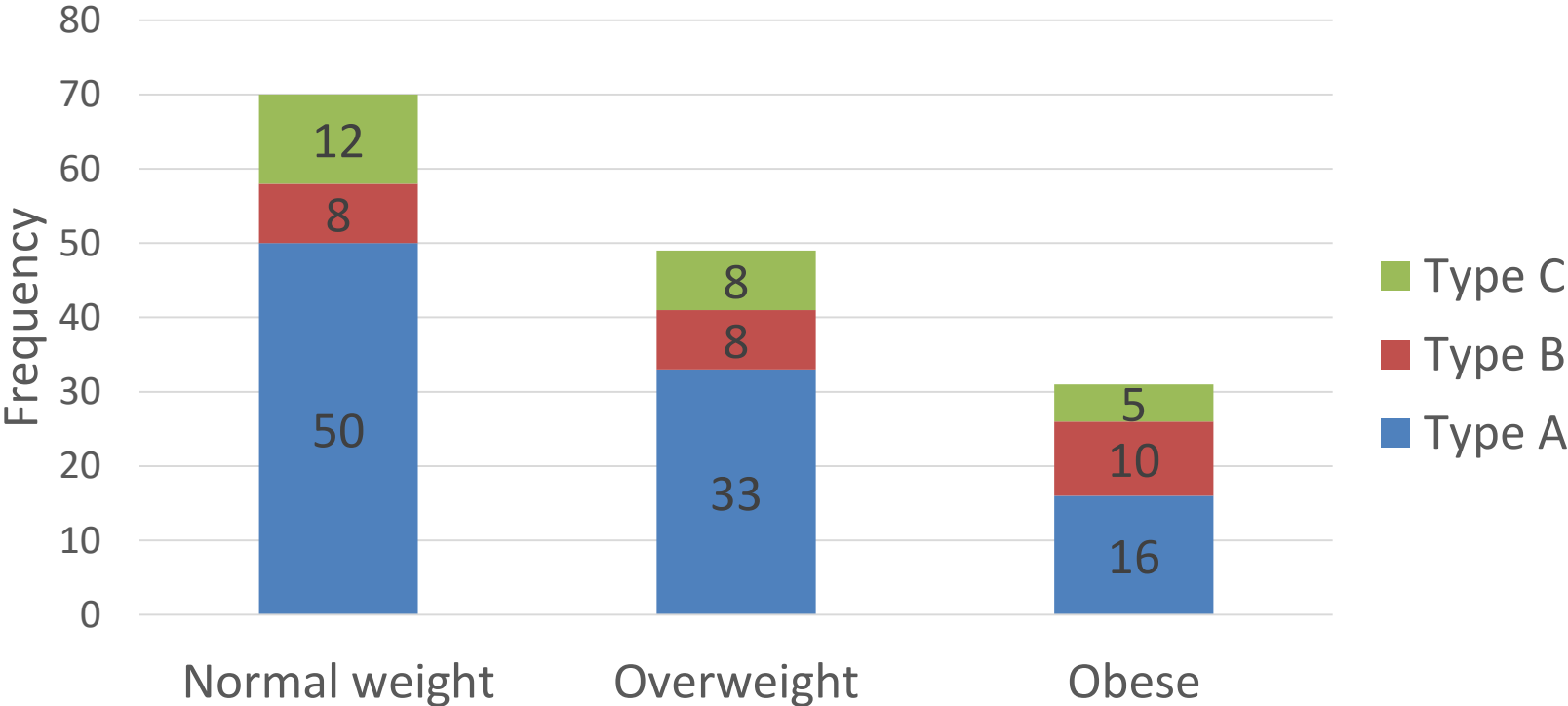
categories are in order

Contingency table

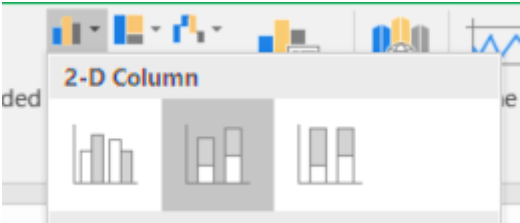
	Hepatitis type			
	Type A	Type B	Type C	Grand Total
Body weight				
Normal weight	50	8	12	70
Overweight	33	8	8	49
Obese	16	10	5	31
Grand Total	99	26	25	150

	Hepatitis type			
Body weight	Type A	Type B	Type C	Grand Total
Normal weight	50	8	12	70
Overweight	33	8	8	49
Obese	16	10	5	31
Grand Total	99	26	25	150

Obesity and hepatitis

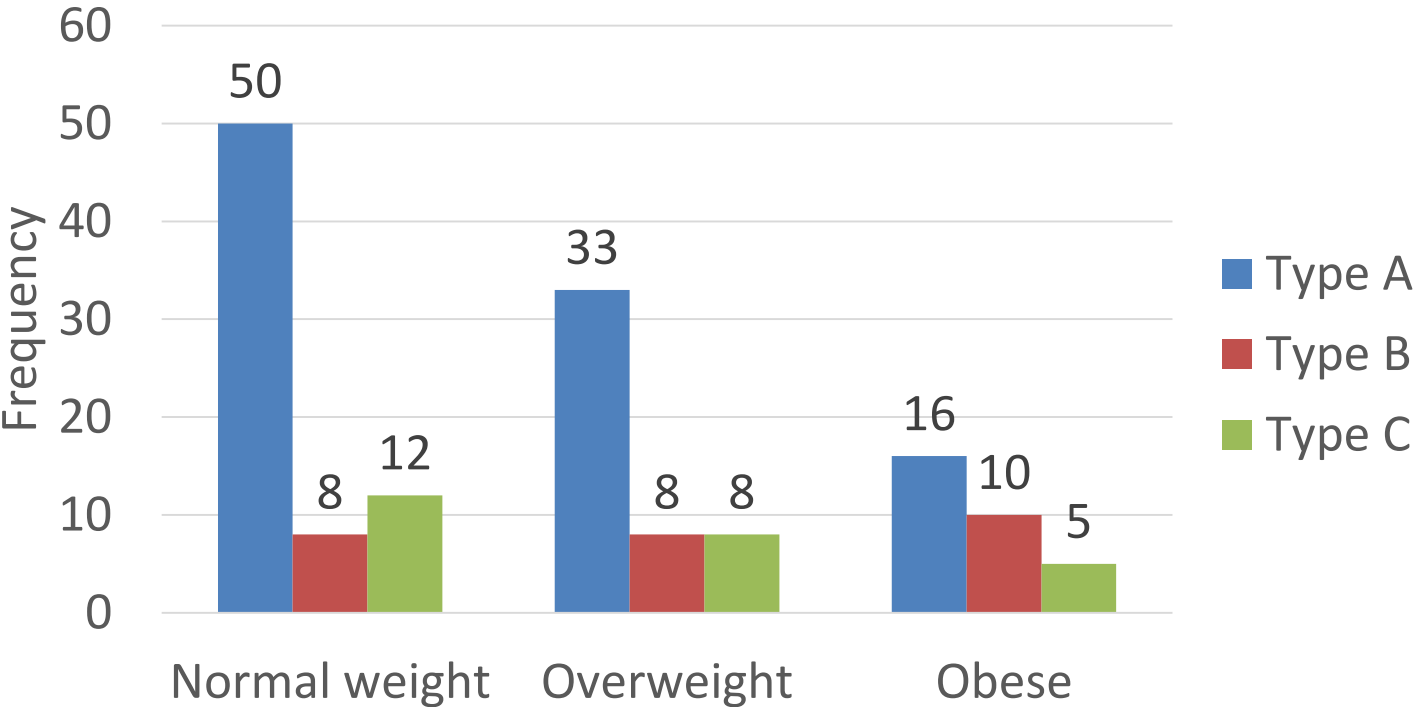


Stacked column chart

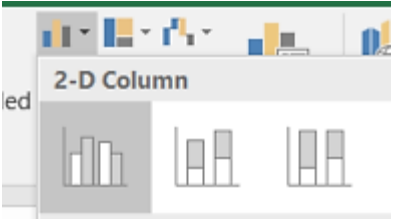


	Hepatitis type			
Body weight	Type A	Type B	Type C	Grand Total
Normal weight	50	8	12	70
Overweight	33	8	8	49
Obese	16	10	5	31
Grand Total	99	26	25	150

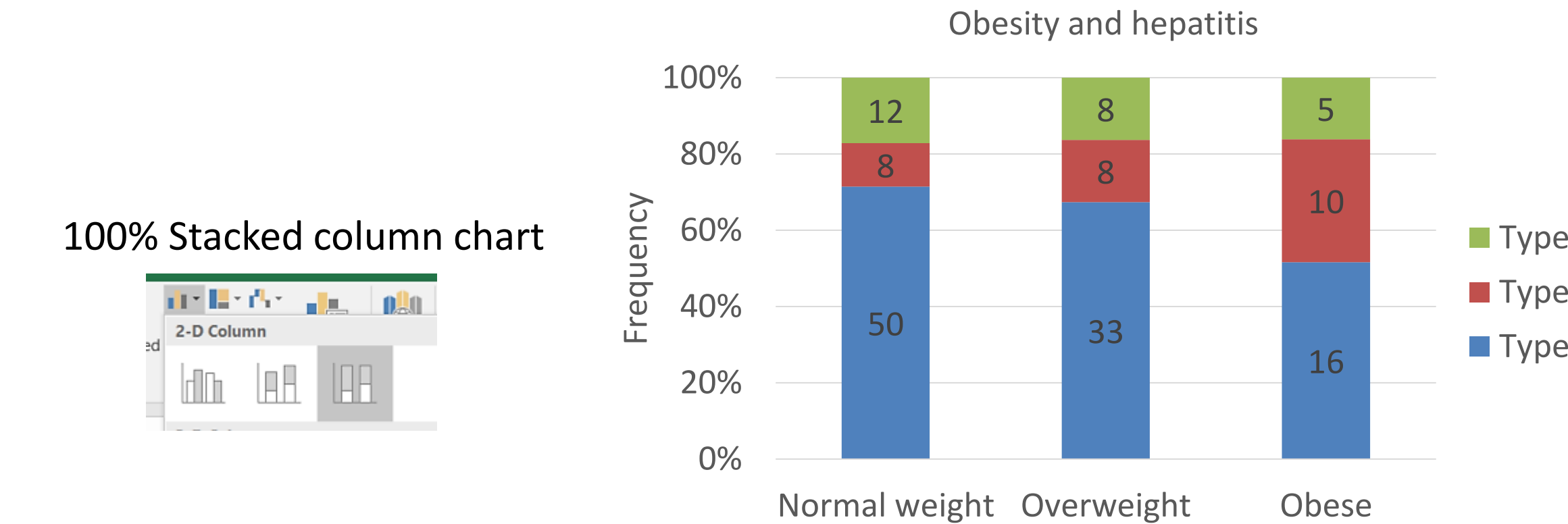
Obesity and hepatitis



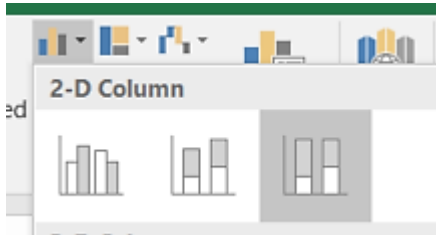
Column chart



	Hepatitis type			
Body weight	Type A	Type B	Type C	Grand Total
Normal weight	50	8	12	70
Overweight	33	8	8	49
Obese	16	10	5	31
Grand Total	99	26	25	150



100% Stacked column chart

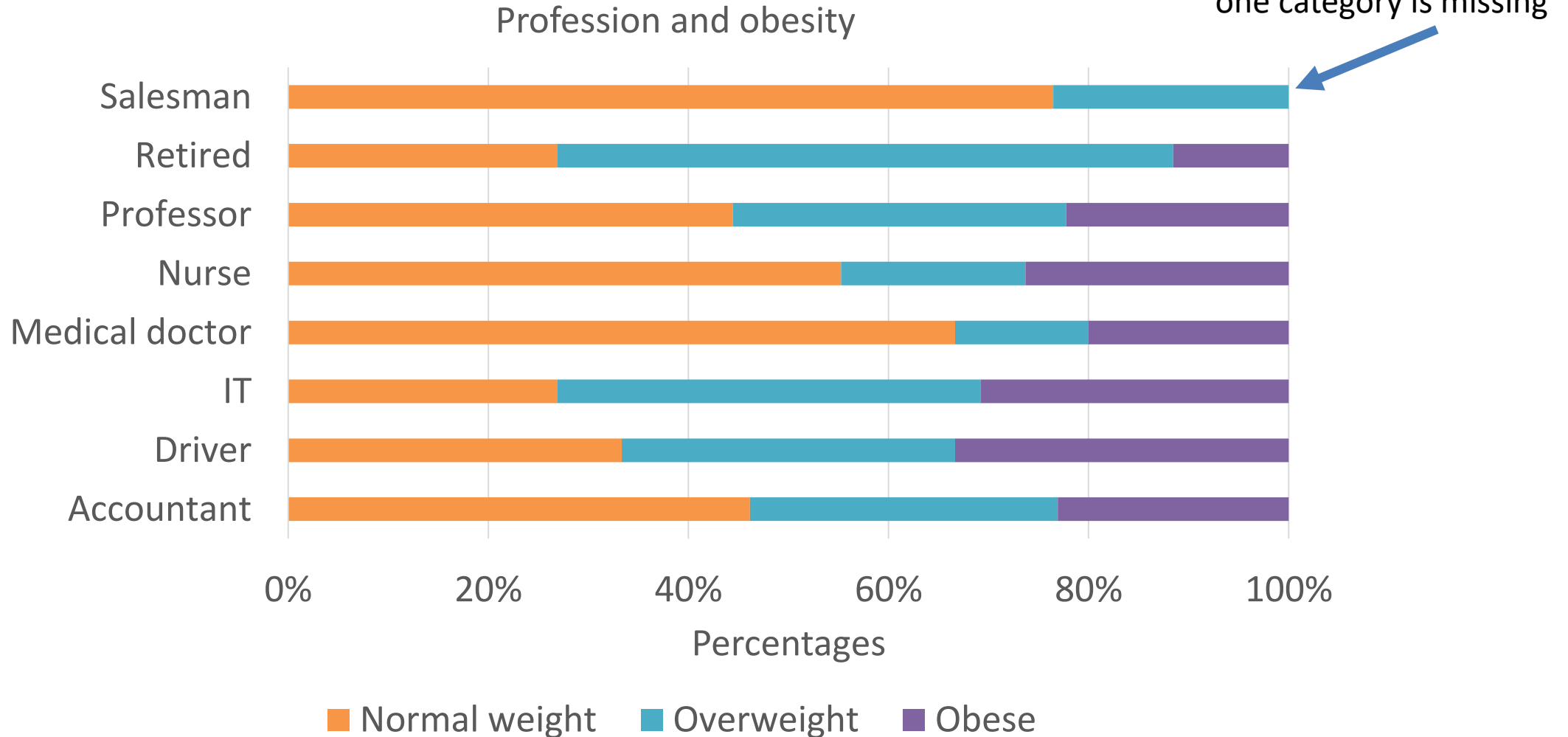
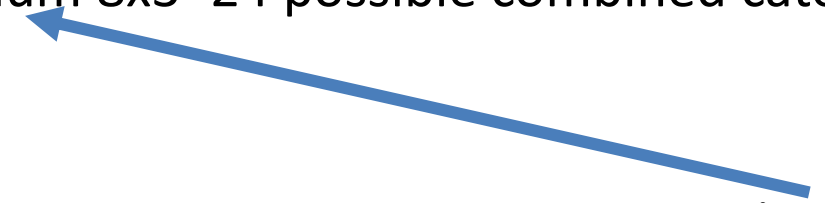


1 qualitative variable with 8 categories



maximum $8 \times 3 = 24$ possible combined categories

1 qualitative variable with 3 categories



!!! they will be summarized the same as a **nominal** variable!!!

Summarizing **ORDINAL** qualitative variables

!!! they will be summarized the same as a **discrete** variable!!!

Ordinal variables

1. ordinal variables with **less** categories

ordinal data with few categories

transformed from quantitative variables

2. ordinal variables which are semiquantitative

scores with few values

scores with many values

grades

Model of End-Stage Liver Disease (MELD) score – the need to liver transplant
>20 less than 90 days of survival

Apgar Score	C
10	
10	
9	
10	MELD score
10	5
8	1
8	6
9	9
9	11
9	8
10	25
10	30
10	32
10	14
10	8
10	7
6	6
10	5
10	22
10	25
10	29
7	12
7	16
8	19

summarize
like a discrete
variable

summarize
like a discrete
variable

Question 1	C
Somewhat disagree	Glycemia classification
Strongly disagree	(hypoglycemia, normal
Somewhat agree	glycemia, prediabetes,
Somewhat agree	diabetes)
Somewhat disagree	normal glycemia
Somewhat disagree	normal glycemia
Somewhat disagree	normal glycemia
Strongly agree	prediabetes
Strongly agree	diabetes
Strongly agree	diabetes
Neither agree or disagree	prediabetes
Somewhat agree	normal glycemia
Somewhat disagree	normal glycemia
Somewhat disagree	normal glycemia
Neither agree or disagree	normal glycemia
Neither agree or disagree	normal glycemia
Somewhat agree	prediabetes
Somewhat agree	prediabetes
Somewhat disagree	diabetes
Somewhat disagree	normal glycemia
Somewhat disagree	normal glycemia
Neither agree or disagree	normal glycemia
Neither agree or disagree	hypoglycemia

summarize
like a nominal
variable

summarize
like a nominal
variable

Summarizing qualitative variables representations

One qualitative variable

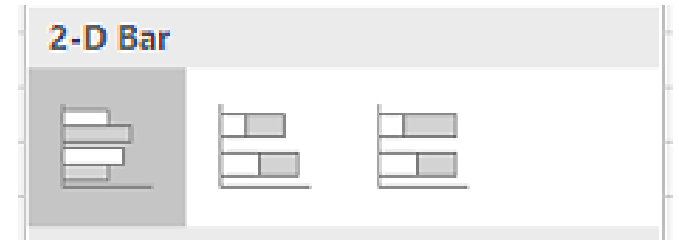
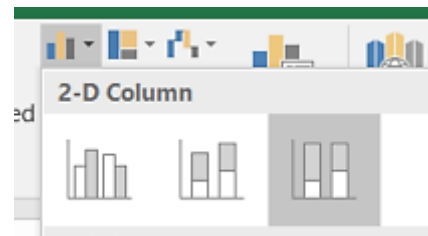
- Frequencies Tables
- Charts
 - Pie
 - Column
 - Bar

Two qualitative variables

- Contingency table

Charts

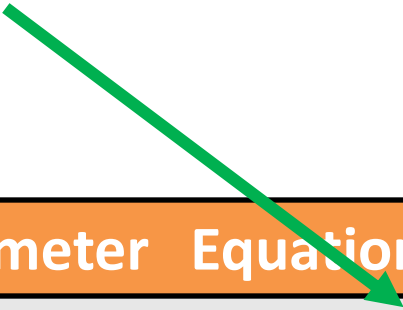
- Column / Bar
- Stacked Column / Bar
- 100% Stacked Column / Bar



Calculate the parameters in Excel

range C2:C21

	C
1	MELD score
2	5
3	1
4	6
5	9
6	11
7	8
8	25
9	30
10	32
11	14
12	8
13	7
14	6
15	5
16	22
17	25
18	29
19	12
20	16
21	19

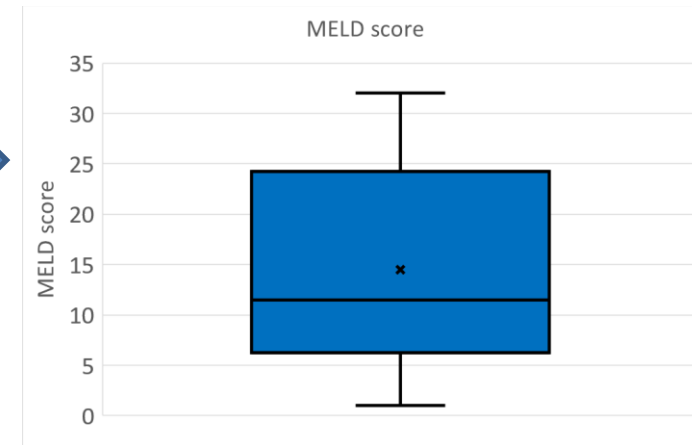
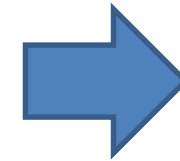
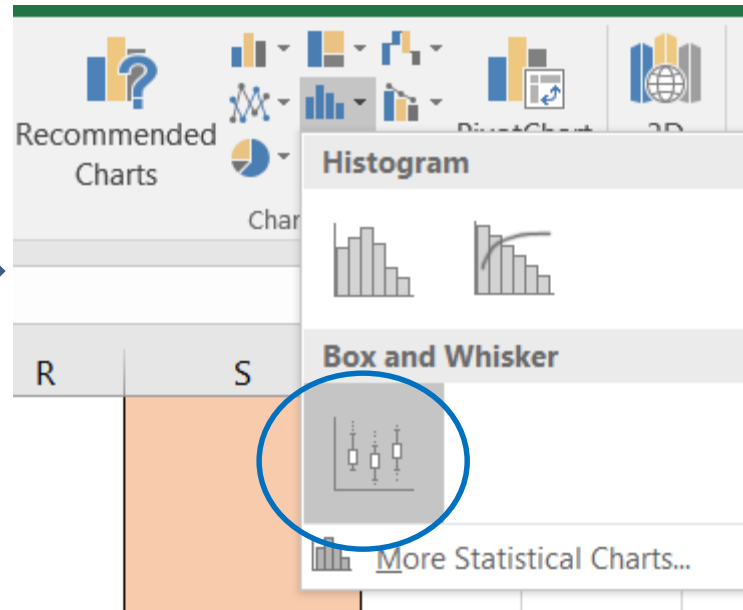
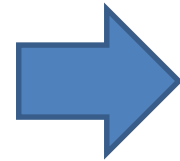


Parameter	Equation in Excel
Minimum	=min(C2:C21)
Maximum	=max(C2:C21)
Range	=max(C2:C21)-min(C2:C21)
Median	=median(C2:C21)
Cuartile 1	=quartile(C2:C21,1)
Cuartile 3	=quartile(C2:C21,3)
Mode	=mode(C2:C21)

Box-plot in Excel for one ordinal variable

Select the data

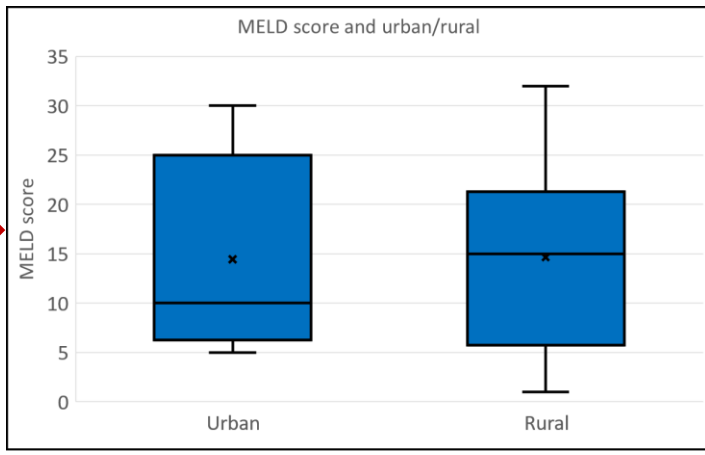
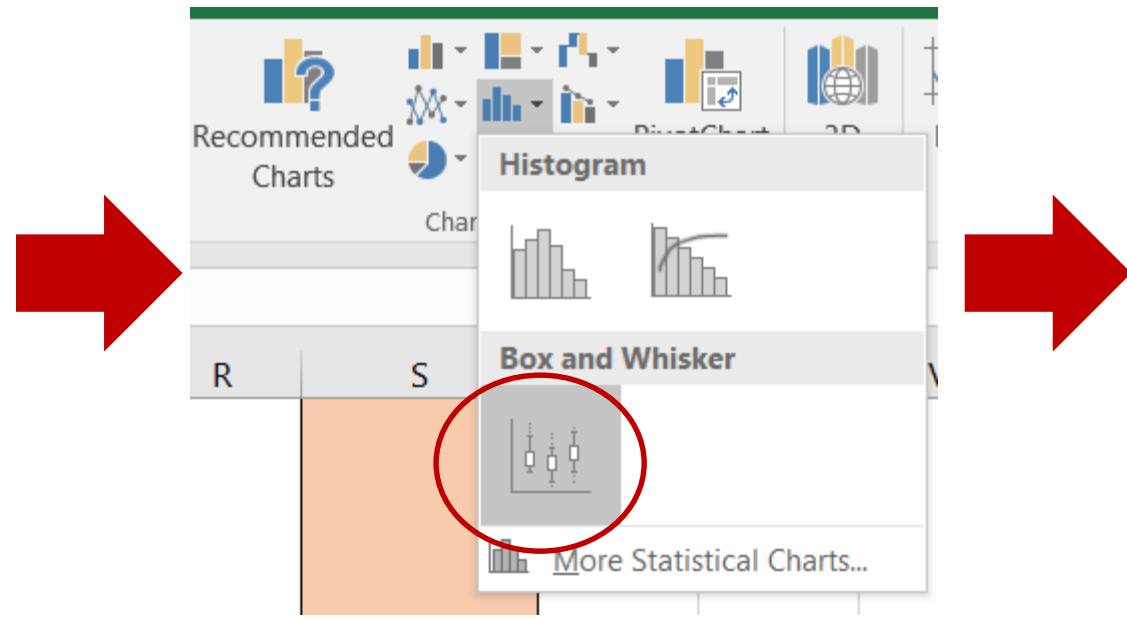
	C
1	MELD score
2	5
3	1
4	6
5	9
6	11
7	8
8	25
9	30
10	32
11	14
12	8
13	7
14	6
15	5
16	22
17	25
18	29
19	12
20	16
21	19



Select both columns

Box-plot in Excel for one ordinal variable

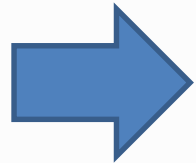
B	C
Urban/Rural	MELD score
Urban	5
Rural	1
Urban	6
Urban	9
Urban	11
Rural	8
Urban	25
Urban	30
Rural	32
Rural	14
Urban	8
Urban	7
Urban	6
Rural	5
Rural	22
Urban	25
Urban	29
Urban	12
Rural	16



Principles of Making Tables in Statistics

A statistical table should be

- clear,
- concise,
- accurate,
- logically arranged,
- properly labeled,
- easy to interpret.

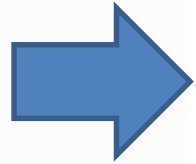


1. With a **clear title** - what the table shows, where, and when
2. **Appropriate row and column headings**
3. Keep the table **simple**.
4. **Organize** data logically
5. Include **totals** when appropriate
6. **Mention the data source**
7. **Number the table**

Principles of Making Tables in Statistics

A statistical chart should be

- clear,
- accurate,
- simple,
- properly labeled,
- appropriately scaled,
- suitable for the data type



1. A clear title
2. Appropriate type of chart
3. Label the axes
4. Use an appropriate scale
5. Keep the chart simple
6. Represent the data accurately
7. The size of bars, lines, or sectors should correspond correctly to the values
8. Use a suitable legend
9. If different colors or symbols are used, explain what they represent.
10. Make comparisons easy
11. Categories should be arranged logically
12. If the data comes from another source, indicate the source below the chart.
13. Make it easy to read

Thank you!