

DESCRIPTIVE STATISTICS FOR QUALITATIVE VARIABLES

Tudor Călinici

2026

STATISTICS



Collect data, present data, compute parameters in order to obtain information about groups

VARIABLES

Qualitative variables

- Dichotomous
- Nominals
- Ordinals

Quantitative variables

- Continuous
- Discrete

DATA COLLECTION

Collect the data of interest from the sample



QUALITATIVE VARIABLES

Usually are not numbers

Statistical parameter: Frequency of one value– How many

DESCRIPTIVE STATISTICS FOR ONE DICHOTOMOUS VARIABLE

Dichotomous variable – could take ONLY two values

We we'll report the DISTRIBUTION OF THE VARIABLE using FREQUENCY TABLE and DISTRIBUTION CHART (Pie or Column/Bar)

Distribution of the variable = each of the values and their frequency

EXAMPLE



Sample size 10; Define Gender variable, {F / M}

Collecting data for Gender = (F, M, F, M, F, F, F, M, F, M)

$$F(F) = 6$$

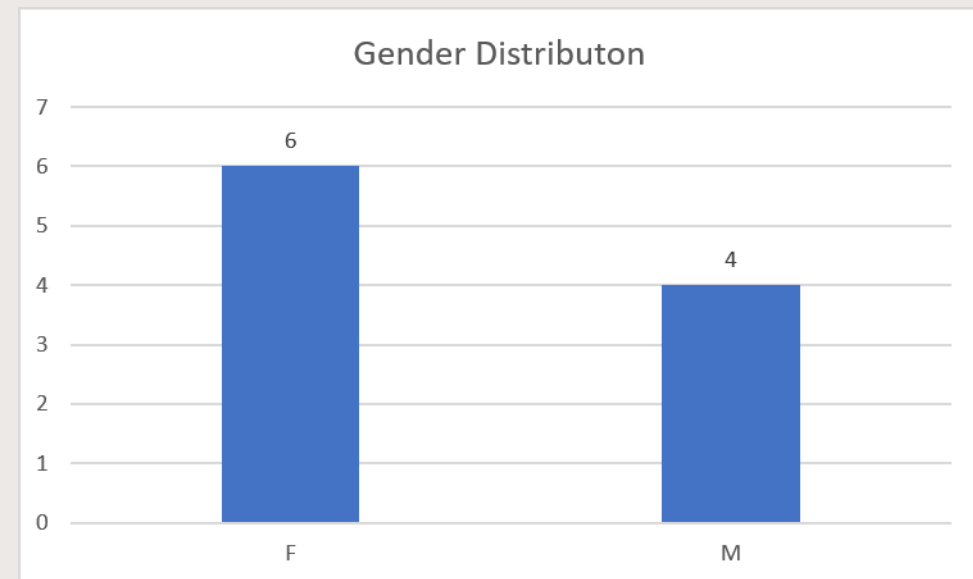
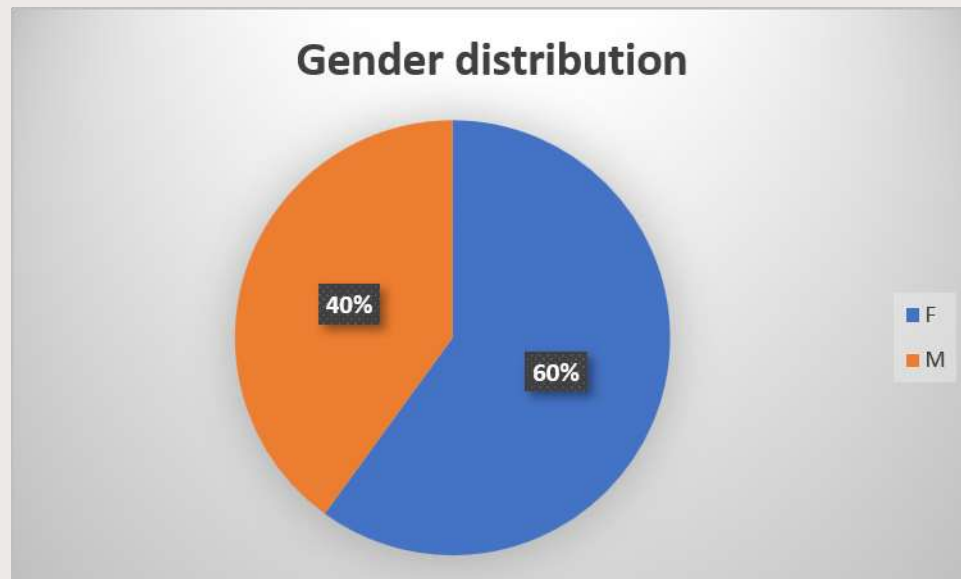
$$F(M) = 4$$

EXAMPLE - REPORTING THE DISTRIBUTION

Value	n (%)
F	6 (60%)
M	4 (40%)

Value	Freq
F	6
M	4

Frequency table



DESCRIPTIVE STATISTICS FOR ONE NOMINAL VARIABLE

Nominal variables could take more than 2 values but there is no inherit order or ranking

We we'll report the DISTRIBUTION OF THE VARIABLE using FREQUENCY TABLE and DISTRIBUTION CHART (Pie or Column/Bar)

MODE - most frequent value (could be more than one)

EXAMPLE



Sample size: 20, Define Continent = {Africa, America, Asia, Europe}

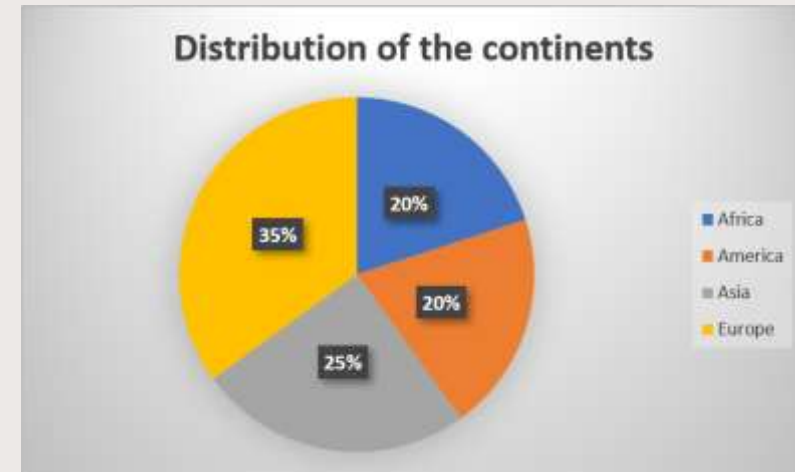
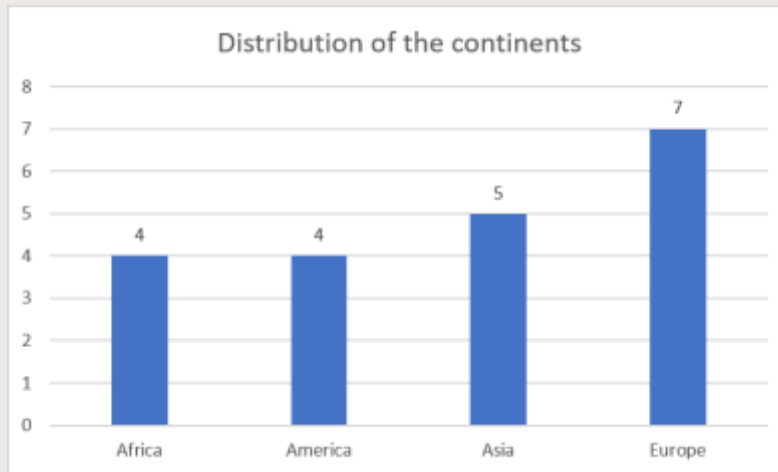
Collecting data = (Europe, Europe, Asia, Africa, Europe, America, Asia, America, Africa, Asia, Europe, Europe, Asia, Africa, Europe, America, Asia, America, Africa, Europe)

$F(\text{Africa}) = 4$; $F(\text{America})=4$; $F(\text{Asia}) =5$ $F(\text{Europe})=7$

EXAMPLE – REPORTING THE DISTRIBUTION

Value	Freq
Africa	4
America	4
Asia	5
Europe	7

Value	n(%)
Africa	4 (20%)
America	4 (20%)
Asia	5 (25%)
Europe	7 (35%)



Mode = Europe

DESCRIPTIVE STATISTICS FOR ONE ORDINAL VARIABLE

Nominal variables could take more than 2 values but there is inherit order or ranking

We we'll report the DISTRIBUTION OF THE VARIABLE using FREQUENCY TABLE and DISTRIBUTION CHART (Pie or Column/Bar)

MODE - most frequent value (could be more than one)

EXAMPLE



Sample size = 23 Define Age category = {Child, Adult less than 60, Adult more than 60}

Collecting data = {Adult less than 60, Adult less than 60, Adult less than 60, Adult less than 60, Child, Child, Child, Adult less than 60, Adult more than 60, Adult less than 60, Adult more than 60, Adult less than 60, Adult less than 60, Adult less than 60, Adult less than 60, Adult less than 60, Adult less than 60, Adult less than 60, Adult less than 60, Adult less than 60, Adult less than 60, Adult less than 60, Adult less than 60}

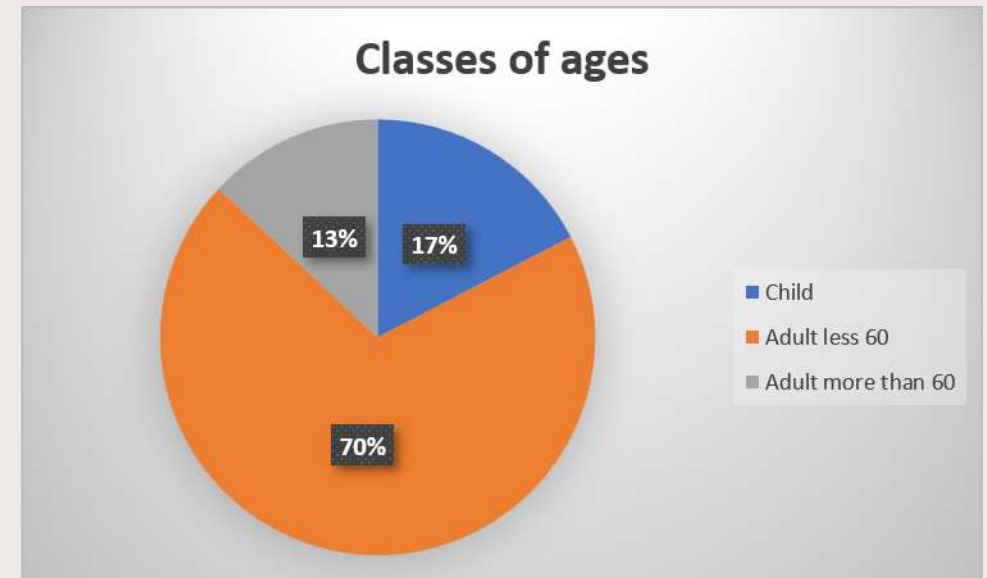
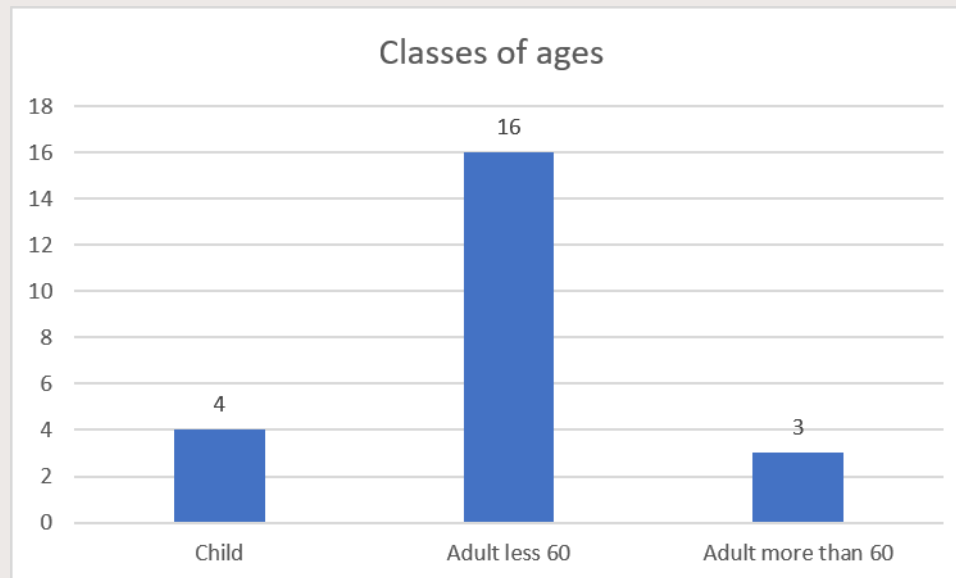
$F(\text{Child}) = 4$; $F(\text{Adult less than 60}) = 16$ $F(\text{Adult more than 60}) = 3$

EXAMPLE – REPORTING THE DISTRIBUTION

Value	Freq
Child	4
Adult less 60	16
Adult more than 60	3

Value	n (%)
Child	4 (17.39%)
Adult less 60	16 (69.57%)
Adult more than 60	3 (13.04%)

Value	n	Cum (%)
Child	4	17.39%
Adult less 60	16	86.96%
Adult more than 60	3	100.00%



DESCRIPTIVE STATISTICS FOR ORDINAL (NUMERICAL) VARIABLES

Ordinal variables could be coded as numbers. For those variables the numerical order is inherit but the difference between two consecutive values is not the same (we can't do math)

Example Quality of life score 1=Very Poor, 15=Very Good

We we'll report the DISTRIBUTION OF THE VARIABLE using FREQUENCY TABLE and DISTRIBUTION CHART (Column/Bar or Box and Whisker)

Mode – most frequent

Median – half of the values are less than it (and half of the values are more than it)

1st Quartile (Q1) 25% of the data is less than it (and 75% of the data are more than it)

3rd Quartile (Q3) 75% of the data is less than it (and 25% of the data is more than it)

EXAMPLE

Sample size: 136

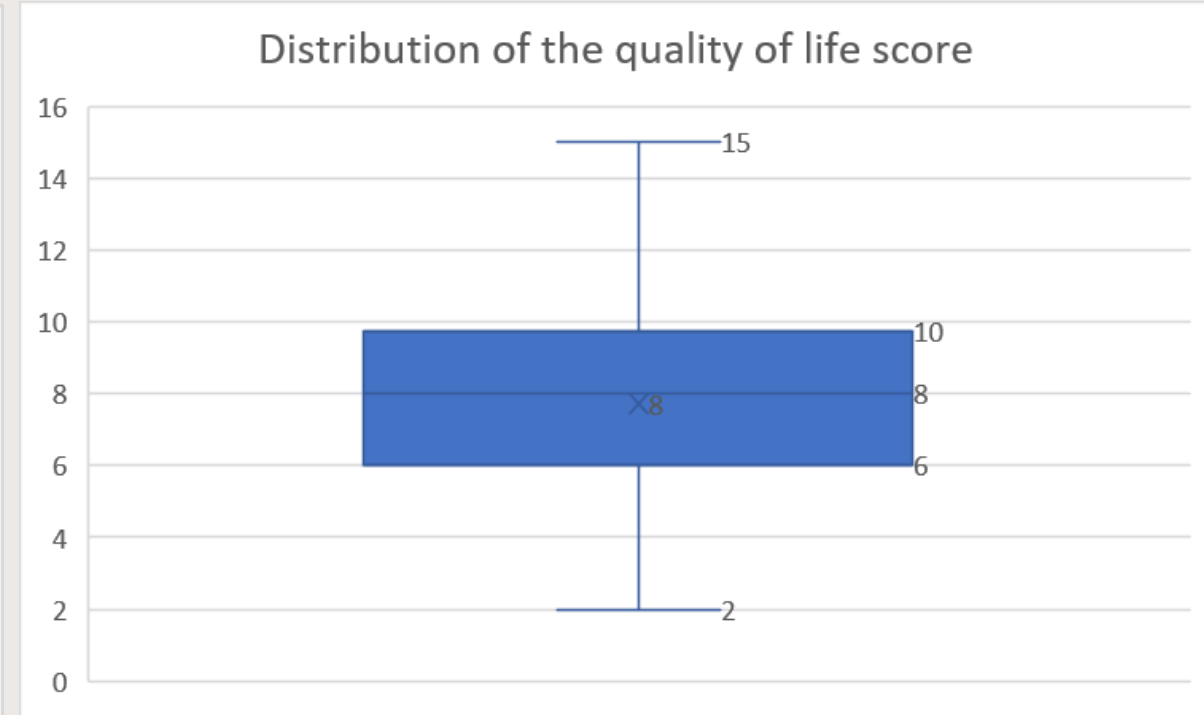
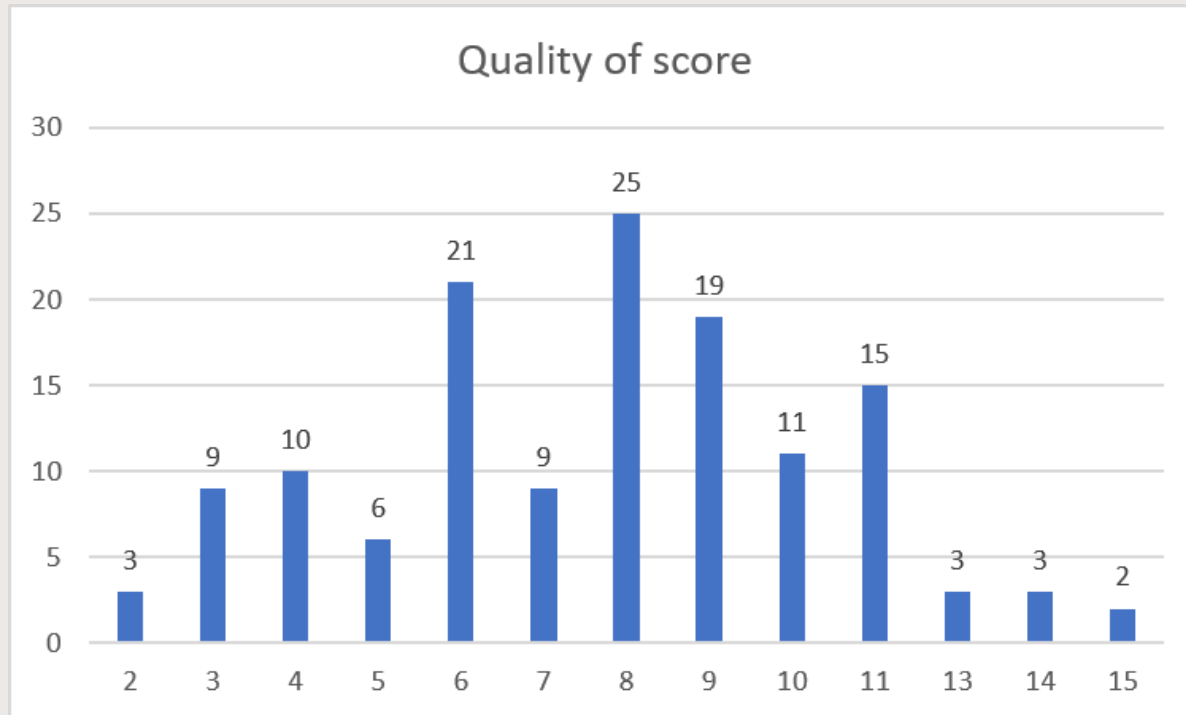
Quality of life score						
4	11	11	14	2	11	9
6	8	10	3	9	9	13
11	5	10	9	8	5	11
9	10	3	4	6	11	6
8	6	8	9	8	9	10
8	8	9	6	3	11	3
11	11	6	2	6	14	8
8	8	10	9	6	6	9
5	15	3	8	10	9	3
10	4	13	6	7	7	8
6	7	10	8	8	3	4
8	6	4	3	6	4	6
11	5	4	6	8	5	
8	8	9	6	7	10	
15	14	7	10	9	11	
4	3	2	7	13	4	
7	9	8	8	11	9	
6	4	11	6	6	6	
5	9	9	8	11	9	
8	6	7	7	10	8	

Value	Freq
2	3
3	9
4	10
5	6
6	21
7	9
8	25
9	19
10	11
11	15
13	3
14	3
15	2

Value	Freq	%
2	3	2.21%
3	9	6.62%
4	10	7.35%
5	6	4.41%
6	21	15.44%
7	9	6.62%
8	25	18.38%
9	19	13.97%
10	11	8.09%
11	15	11.03%
13	3	2.21%
14	3	2.21%
15	2	1.47%

Value	Freq	Cum %
2	3	2.21%
3	9	8.82%
4	10	16.18%
5	6	20.59%
6	21	36.03%
7	9	42.65%
8	25	61.03%
9	19	75.00%
10	11	83.09%
11	15	94.12%
13	3	96.32%
14	3	98.53%
15	2	100.00%

EXAMPLE



DISTRIBUTION OF ONE QUALITATIVE VARIABLE IN GROUPS

Two qualitative variables

- One variable of interest
- One variable for grouping

The distribution will be presented as a contingency table and 100% stacked column chart

EXAMPLE

To test the hypothesis that maternal alcohol consumption during pregnancy increases the risk of malformations in newborns, researchers interviewed two groups: 65 mothers whose children were born with birth defects and 210 mothers whose children were born without birth defects. In the first group, 50 mothers had consumed alcohol during pregnancy, compared with 90 mothers in the second group.

Variables:

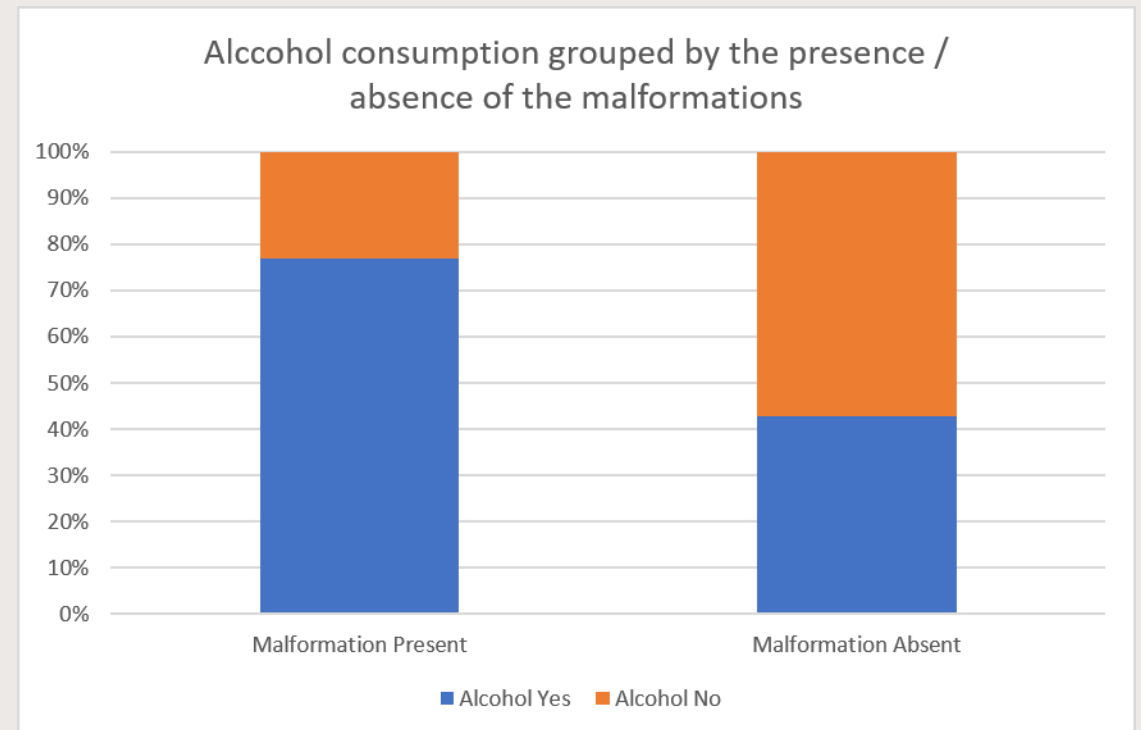
- Malformations {Present / Absent}
- Alcohol consumption {Yes / No}

REPORTING THE INFORMATION

	Malformation Present	Malformation Absent	Total
Alcohol Yes	50	90	140
Alcohol No	15	120	135
Total	65	210	275

Contingency table

The chart suggests that maternal alcohol consumption during pregnancy increases the risk of malformations in newborns



THANK YOU!