King Saud University

College of Sciences

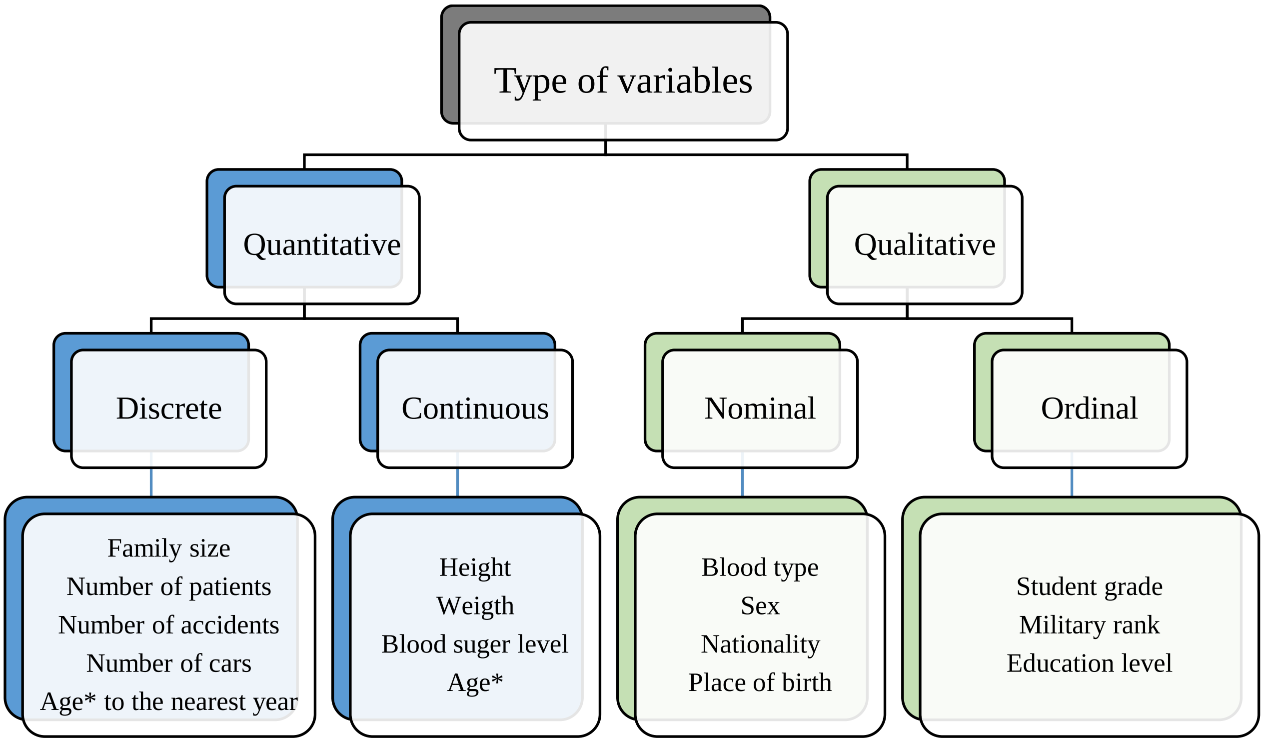
Department of Statistics & OR

STAT 145

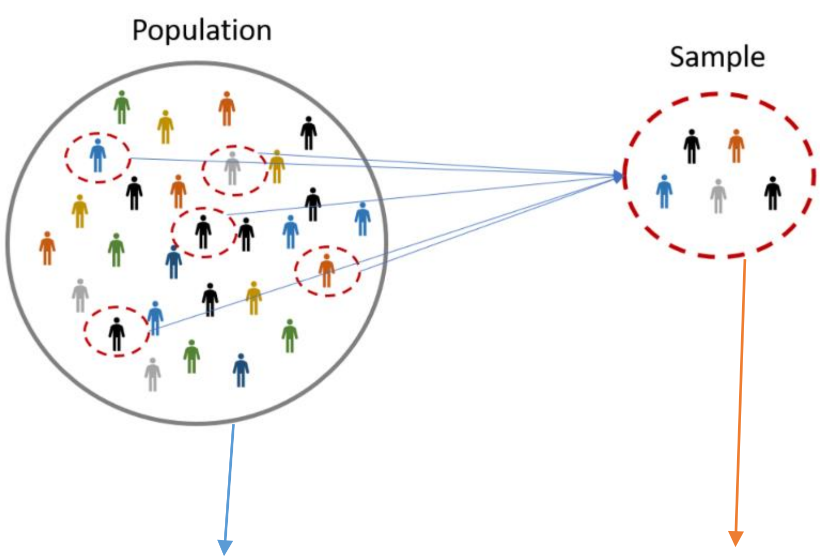
Biostatistics

Chapter 1 Introduction

- <u>Some Basic Concepts:</u>



- <u>Statistical Inference:</u>



Parameter	Statistic
Population mean μ	Sample mean $\overline{X}$
Population variance σ^2	Sample variance S^2

Question 1:

1. The number of students admitted in College of Medicine in King Saud University is a variable of type

A	Discrete	B	Qualitative	C	Continuous	D	Nominal

2. A mean of a population is called:

A	Parameter	B	Statistic	C	Median	D	Mode

3. The measure that obtained from the population is called

A	Parameter	B	Sample	C	Population	D	Statistic

4. The measure that obtained from the sample is called

A	Parameter	B	Sample	C	Population	D	Statistic

5. Which of the following is an example of a statistic?

A	Population variance	B	Sample median	C	Population mean	D	Population mode

6. A sample is defined as:

A	The entire population of values.
B	A measure of reliability of the population.
C	A subset of data selected from a population.
D	Inferential statistics.

7. The continuous variable is a

A	Variable takes on values within intervals.
B	Variable which can't be measured.
C	Variable with a specific number of values.
D	Variable with no mode.

8. The nominal variable is a

A	Qualitative variable which can't be ordered.
B	Quantitative variable.
C	Qualitative variable which can be ordered.
D	Variable with a specific number of values.

9. One of the following is an example of an ordinal variable:

A	Socio-economic level.
B	Blood type of a sample of patients.
C	The time of finish the exam.
D	The number of persons who are injured in accidents.

10. One of the following is an example of a statistic:

A	The sample mode.
B	The population median.
C	The population variance.
D	None of these.

11. One of the following is a part of a population:

A	Sample.
B	Statistic
C	Variable
D	None of these

12. The variable is a

A	Characteristic of the population to be measured.
B	Subset of the population.
C	Parameter of the population.
D	None of these.

Question 2:

From men with age more than 20 years living in Qaseem, we select 200 men. It was found that the average weight of the men was 76 kg.

1. The variable of interest is:

A	Age	B	Weight	C	200 men	D	76 kg

2. The sample size is:

A	76	B	20	C	200	D	1520

Question 3:

A study of 250 patients admitted to a hospital during the past year revealed that, on the average (mean), the patients lived 15 miles from the hospital.

1. The sample in the study is:

A	250 patients	B	250 hospitals	C	250 houses	D	15 miles

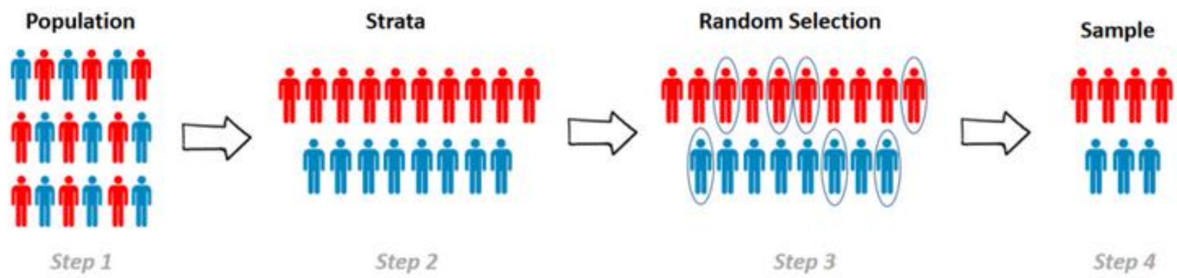
2. The population in this study is:

A	Some patients admitted to the hospital during the past year.
B	All patients admitted to the hospital during the past year.
C	250 patients admitted to the hospital during the past year.
D	500 patients admitted to the hospital during the past year.

3. If a researcher interests to study the blood pressure level (High, Normal, Low) for 13 diabetics patients, what is the type of variables?

A	Qualitative nominal.
B	Quantitative nominal.
C	Qualitative ordinal.
D	Quantitative ordinal.

- **Stratified Random Sampling:**



Question 4:

A researcher was interested in estimating the mean of monthly salary of a certain city. There were 5000 employees in the city (2000 of which were female and 3000 of which were males). He selected a random sample of 40 female employees, and he independently selected a random sample of 60 male employees. Then, he combined these two random samples to obtain the random sample of his study.

1. The population size is:

A	100	B	3000	C	2000	D	5000

2. The sample size is:

A	40	B	60	C	100	D	1000

3. The variable of interested is:

A	Employee	B	City	C	Sex	D	Salary

4. The type of the random sample of this study is:

A	Stratified random sample	B	Simple random sample

5. If each element in the population has the same chance to be selected in the sample, then the sample called:

A	Simple random sample.
B	Sample space.
C	Stratified sample.
D	Complete sample.

Chapter 2 Strategies for Understanding the Meaning of Data

- **Frequency Tables:**

Question 1:

The "life" of 40 similar car batteries recorded to the nearest tenth of a year. The batteries are guaranteed to last 3 years.

Class Interval	True class Interval	Midpoint	Frequency	Relative Frequency
1.5–1.9	1.45–1.95	1.7	2	0.050
2.0–2.4	1.95–2.45	2.2	D	0.025
2.5–2.9	2.45–2.95	C	4	F
A	2.95–3.45	3.2	15	0.375
3.5–3.9	B	3.7	E	0.250
4.0–4.4	3.95–4.45	4.2	5	0.125
4.5–4.9	4.45–4.95	4.7	3	0.075

1. The value of A: $\boxed{3.0 - 3.4}$
2. The value of B: $\boxed{3.45 - 3.95}$
3. The value of C: $\boxed{C = \frac{2.45+2.95}{2} = 2.7}$
4. The value of D: $\boxed{\frac{D}{40} = 0.025 \Longrightarrow D = 40 \times 0.025 = 1}$
5. The value of E : $\boxed{\frac{E}{40} = 0.25 \Longrightarrow E = 40 \times 0.25 = 10}$
6. The value of F: $\boxed{F = \frac{4}{40} = 0.10}$

Question 2:

Fill in the table given below. Answer the following questions.

Class Interval	Frequency	Cumulative Frequency	Relative Frequency	Cumulative Relative Frequency
5 - 9	8			
10 - 14	15		C	
15 – 19	11	B		D
20 - 24	A	40	0.15	

1) The value of A is: $\boxed{A = 40 - (8 + 15 + 11) = 40 - 34 = 6}$

2) The value of B is: $\boxed{B = 8 + 15 + 11 = 34}$

3) The value of C is: $\boxed{C = \frac{15}{40} = 0.375}$

4) The value of D is: $\boxed{D = \frac{34}{40} = 0.85}$

5) The true class interval for the first class is: $\boxed{4.5 - 9.5}$

6) The number of observations less than 19.5 is: $\boxed{8 + 15 + 11 = 34}$

Question 3:

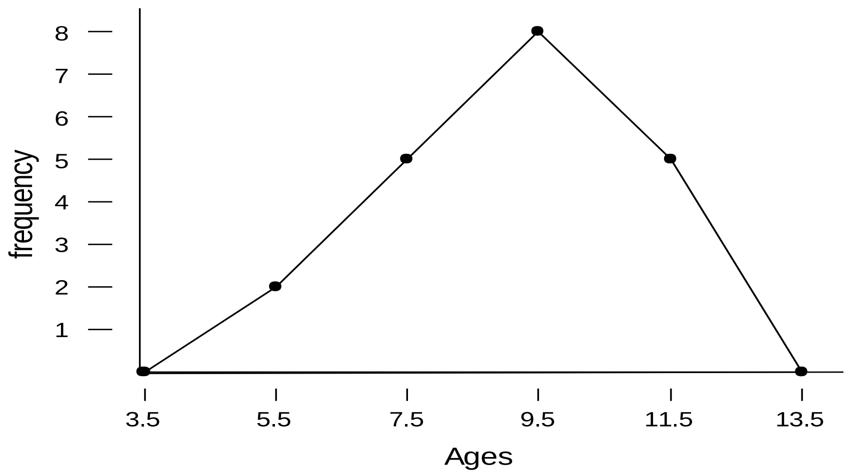
The table shows the weight loss (kg) of a sample of 40 healthy adults who fasted in Ramadan.

Class interval	Frequency	Cumulative Frequency
1.20 - 1.29	2	2
1.30 - 1.39	6	8
1.40 - 1.49	10	K
1.50 - 1.59	C	34
1.60 - 1.69	6	40

1) The value of the missing value K is $\boxed{18}$
2) The value of the missing value C is $\boxed{16}$

Question 4:

Consider the following frequency polygon of ages of 20 students in a certain school.



The frequency distribution of ages corresponding to above polygon is

(a)

True class limits	4.5- 6.5	6.5-8.5	8.5- 10.5	10.5 -12.5
frequency	2	5	8	5

(b)

True class limits	3.5- 5.5	5.5-7.5	7.5- 9.5	9.5 -11.5
frequency	2	5	8	4

(c)

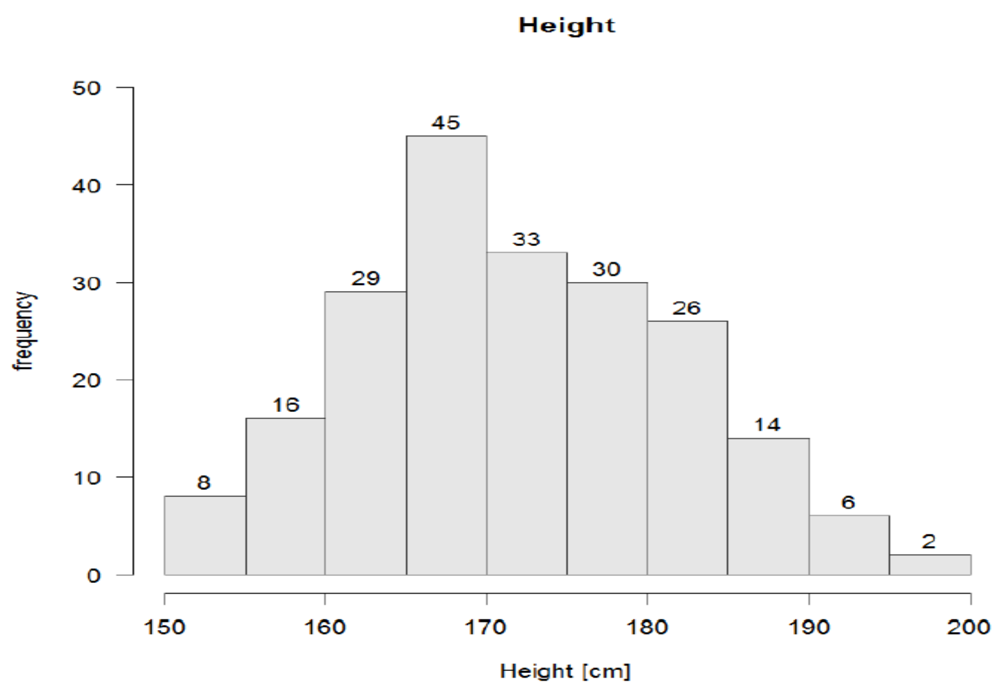
Class interval	5- 6	7-8	9- 10	11 -12
frequency	1	7	8	4

(d)

Class interval	5- 6	7-8	9- 10	11 -12
frequency	4	7	8	6

Question 5:

For a sample of students, we obtained the following graph for their height in (cm).



1. The variable under study is:

A	Patients	B	Graph	C	Height	D	Discrete

2. The type of variable:

A	Continuous	B	Discrete	C	Frequency	D	Height

3. The number of students with the lowest level height:

A	14	B	2	C	115	D	8

4. The sample size is:

A	28	B	209	C	156	D	130

5. The midpoint of the interval with highest frequency is:

A	182.5	B	130.5	C	167.5	D	30

6. The relative frequency of the interval with highest frequency is:

A	0.283	B	0.215	C	0.241	D	0.262

Question 6:

The following table gives the distribution of the ages of a sample of 50 patients who attend a dental clinic.

Age intervals (in years)	Frequency	Relative frequency
10 - 15	4	-
16 - 21	8	-
22 - 27	z	0.32
28 - 33	-	-
34 - 39	10	-

Less than	Cumulative Frequency
10	0
16	4
22	y
28	--
34	--
40	x

1. The class width is:

A	6	B	10	C	150	D	19

2. The value of x is:

A	22	B	28	C	50	D	10

3. The value of y is:

A	4	B	12	C	19	D	150

4. The value of z is:

A	14	B	12	C	50	D	16

5. Percent of the patients with age between 16 and 21 is:

A	16%	B	8%	C	20%	D	32%

6. The 5th interval midpoint is:

A	38	B	52	C	27	D	36.5

Question 7:

Consider the following Table showing a frequency distribution of weights in a sample of 20 cans of fruits:

Class interval	True Class Limits	Midpoint	Frequency	Relative Frequency	Cumulative Frequency
19.2 – 19.4			1		
19.5 – 19.7				0.10	
19.8 – 20.0			8		
			4		

1. The fifth-class interval is:

A	20.2 - 20.4	B	20.1-20.3	C	21.0 - 21.2	D	20.4 - 20.6

2. The second true class interval is:

A	19.45 - 19.75	B	19.5 – 19.7	C	19.25 -19.35	D	20.2 - 20.4

3. The midpoint of the fourth-class interval is:

A	20.5	B	20.2	C	19.9	D	20.1

4. The frequency of the second-class interval is:

A	10	B	4	C	2	D	3

5. The relative frequency of the fourth-class interval is:

A	0.20	B	0.15	C	0.13	D	0.40

6. The cumulative frequency of the final class interval is:

A	13	B	4	C	20	D	100

Question 8:

Consider the following table showing a frequency distribution of blood test of 52 diabetes patients.

Class interval	Frequency	Cumulative frequency	Relative frequency	Cumulative relative frequency
101 – 120	--	--	0.4423	--
121 – 140	--	--	--	D
B	--	C	0.2115	--
161 – 180	--	--	0.0577	--
Total	A	--	1	--

[1] The value of A is

A	1	B	3	C	52	D	80

[2] The class interval B is

A	122-140	B	161-180	C	131-140	D	141-160

[3] The value of C is

A	49	B	15	C	34	D	52

[4] The value of D is

A	0.5308	B	0.7308	C	0.4308	D	0.8308

[5] The true class intervals are

A	100 –120 120.5 –139.5 141 –160 161 –180	B	99.5 – 119.5 120.5 –140.5 140.5 –159.5 160.5 –179.5	C	100.5 – 120.5 120.5 –140.5 140.5 – 160.5 160.5 – 180.5	D	100.5 – 120.5 121.5 – 140.5 141.5 –160.5 161.5 – 180.5

[6] The midpoint of the first-class interval is

A	110.5	B	20	C	220	D	19

[7] Histogram of the frequency distribution is built based on

A	Frequency and cumulative.
B	Midpoints and cumulative.
C	True class interval and frequency.
D	None of them.

Question 9:

1. To group a set of observations in a frequency table, we should not do one of the following:

A	The intervals are overlapping
B	The intervals are ordering from the smallest to the largest
C	The minimum value of the observation belongs to the first interval
D	The number of intervals should be no fewer than five class intervals

2. If the lower limit of a class interval is 25 and the upper limit of this class interval is 30, the midpoint is equal to

A	27.5	B	2.5	C	27	D	5

3. If 7 out of 45 CEOs have a master's degree, then the relative frequency is equal to:

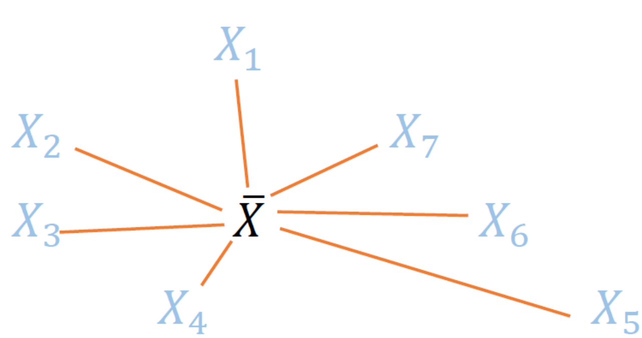
A	0.1556	B	7	C	45	D	15.56%

3. If 7 out of 45 CEOs have a master's degree, then the percentage of the relative frequency is equal to:

A	15.56%	B	0.1556	C	7	D	45

- **Measures of Central tendency and Dispersion**

Measures of central tendency (Location)		
Mean	$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$	Unit
Median		Unit
Mode	The value with the highest frequency	Unit
Measures of dispersions (Shape)		
Range	$R = max - min$	Unit
Variance	$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$ $S^2 = \frac{\sum_{i=1}^{n} X_i^{\,2} - n\bar{X}^2}{n-1}$	Unit^2
Standard deviation	$S = \sqrt{S^2}$	Unit
Coefficient of variation	$C.V = \frac{S}{\bar{X}}$	Unit less



<u>Question 10:</u>

If the number of visits to the clinic made by 8 pregnant women in their pregnancy period is:

12 15 16 12 15 16 12 14

1. The type of the variable is:

$\boxed{\text{discrete}}$

2. The sample mean is:

$$\boxed{\overline{X} = \frac{12+15+16+12+15+16+12+14}{8} = 14}$$

3. The sample standard deviation is:

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$$

$$= \frac{(12-14)^2+(15-14)^2+(16-14)^2+(12-14)^2+(15-14)^2+(16-14)^2+(12-14)^2+(14-14)^2}{8-1}$$

$$= 3.14 \Longrightarrow S = 1.77$$

4. The sample median is:

$$\boxed{12 \;\; 12 \;\; 12 \;\; \boxed{14 \;\; 15} \;\; 15 \;\; 16 \;\; 16} \Longrightarrow \frac{14+15}{2} = 14.5$$

5. The coefficient of variation is:

$$\boxed{C.V = \frac{s}{\overline{X}} = \frac{1.77}{14} = 0.1266}$$

6. The range is:

$$\boxed{16 - 12 = 4}$$

Question 11:

Consider the following marks for a sample of students carried out on 10 quizzes:

6, 7, 6, 8, 5, 7, 6, 9, 10, 6

1. The mean mark is:

$$\boxed{\overline{X} = \frac{6+7+6+8+5+7+6+9+10+6}{10} = 7}$$

2. The median mark is:

$$\boxed{5 \;\; 6 \;\; 6 \;\; 6 \;\; \boxed{6 \;\; 7} \;\; 7 \;\; 8 \;\; 9 \;\; 10} \Longrightarrow \frac{6+7}{2} = 6.5$$

3. The mode for this data is: $\boxed{6}$
4. The range for this data is: $\boxed{5}$
5. The standard deviation for this data is:

$$\boxed{S^2 = \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1} = \frac{(5-7)^2+(6-7)^2+\cdots+(10-7)^2}{10-1} = 2.434 \Longrightarrow S = 1.56}$$

6. The coefficient of variation for this data is:

$$\boxed{C.V = \frac{s}{\overline{X}} = \frac{1.56}{7} = 0.223}$$

Question 12:

Twenty adult males between the ages of 30 and 40 participated in a study to evaluate the effect of a specific health regimen involving diet and exercise on the blood cholesterol. Ten were randomly selected to be a control group, and ten others were assigned to take part in the regimen as the treatment group for a period of 6 months. The following data show the mean and the standard deviation of reduction in cholesterol experienced for the time period for the 20 subjects:

Control group: mean= 6.5, standard deviation=4.33

Treatment group: mean= 7.6, standard deviation=5.32.

By comparing the variability of the two data sets, we get

$$\boxed{\begin{aligned} C.V_{\text{Cont}} &= \frac{s_{\text{Cont}}}{\overline{X}_{\text{Cont}}} \times 100 = \frac{4.33}{6.5} \times 100 = 66.61\% \\ C.V_{\text{Treat}} &= \frac{s_{\text{Treat}}}{\overline{X}_{\text{Treat}}} \times 100 = \frac{5.32}{7.6} \times 100 = 70\% \end{aligned}}$$

The relative variability of the control group is less than relative variability of the treatment group.

Question 13:

The data for measurements of the left ischia tuberosity (in mm Hg) for the SCI and control groups are shown below.

Control	131	115	124	131	122
SCI	60	150	130	180	163

1. The mean for the control group is:

$$\boxed{\overline{X} = \frac{131+115+124+131+122}{5} = 124.60}$$

2. The variance of the SCI group is:

$$\overline{X} = \frac{60+150+130+180+163}{5} = 136.6$$

$$S^2 = \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1} = \frac{(60-136.6)^2+(150-136.6)^2+(130-136.6)^2+(180-136.6)^2+(163-136.6)^2}{5-1} = 2167.8$$

3. The unit of coefficient of variation for SCI group is

A	mm Hg	B	Hg	C	mm	D	Unit-less

4. Which group has more variation:

$$S^2 = \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}$$

$$= \frac{(131-124.6)^2+(115-124.6)^2+(124-124.6)^2+(131-124.6)^2+(112-124.6)^2}{5-1} = 45.3 \Longrightarrow S = 6.7305$$

$$C.V_{Cont} = \frac{S_{Cont}}{\overline{X}_{Cont}} \times 100 = \frac{6.7305}{124.6} \times 100 = 5.4\%$$

$$C.V_{SCI} = \frac{S_{SCI}}{\overline{X}_{SCI}} \times 100 = \frac{\sqrt{2167.8}}{136.6} \times 100 = 34.08\%$$

A	Control group.
B	SCI group.
C	Both groups have the same variation.
D	Cannot compare between their variations.

Question 14:

Temperature (in Faraheniet) recorded at 2 am in London on 8 days randomly chosen in a year were as follows: 40 -21 38 -9 26 -21 -49 44

1) The average temperature for the sample is:

A	248	B	1	C	6	D	48

2) The median temperature for the sample is:

A	8.5	B	-21	C	-8.5	D	17

3) The mode of temperature for the sample is:

A	-21	B	44	C	2	D	-49

4) The standard deviation for the sample data is:

A	35.319	B	30.904	C	1247.43	D	4

5) The coefficient of variation for the sample is:

A	17%	B	49%	C	4%	D	588.7%

6) The range of the sample is:

A	4	B	8	C	40	D	93

Question 15:

Consider the following weights for a sample of 6 babies: 5, 3, 5, 2, 5, 4

[1] The sample mean is

A	4	B	5	C	3	D	6

[2] The sample median is

A	4	B	5	C	4.5	D	3

[3] The sample mode is

A	4	B	3	C	4.5	D	5

[4] The sample standard deviation is

A	3.2649	B	8.2649	C	1.2649	D	2.2649

[5] The coefficient of variation for this sample is

A	40.00%	B	31.62%	C	200%	D	12.50%

Question 16:

Some families were selected and the number of children in each family were considered as follows: 5, 8, 0, 8, 3, 7, 8, 9 Then,

1) The sample size is:

A	9	B	6	C	8	D	5

2) The sample mode is:

A	9	B	0	C	8	D	No mode

3) The sample mean is:

A	48	B	6	C	8	D	0

4) The sample variance is:

A	2.915	B	8.5	C	9.714	D	3.117

5) The sample median is:

A	5.5	B	7.5	C	8	D	7

6) The range of data is:

A	8	B	0	C	3	D	9

7) The sample coefficient of variation is:

A	5.5	B	8	C	0.52	D	7

Question 17:

1.Which of the following measures is not affected by the extreme values?

A	Median	B	Mean	C	Variance	D	Range

2.Which of the following location (central tendency) measures is affected by extreme values?

A	Range	B	Mean	C	Median	D	Mode

3.Which of the following measures can be used for the blood type in a given sample?

A	Median	B	Mean	C	Variance	D	Mode

Question 18:

The frequency table for daily number of car accidents during a month is:

Number of car accidents	Frequency
3	2
4	3
5	1
6	2
7	2
Total	10

1. The type of variable:

A	Nominal	B	Discrete	C	Ordinal	D	Continuous

2. The mean for the number of accidents is:

A	4.07	B	4.90	C	3.75	D	2.98

3. The median is:

A	5.5	B	5	C	4.5	D	4

4. The mode is:

A	4	B	5	C	6	D	3

5. The variance for the number of accidents is:

A	8.45	B	6.43	C	2.32	D	1.05

6. The coefficient of the variation is:

A	2%	B	31%	C	22%	D	12%

Question 19:

1. The biggest advantage of the standard deviation over the variance is:

A	The standard deviation is always greater than the variance.
B	The standard deviation is calculated with the median instead of the mean.
C	The standard deviation is better for describing the qualitative data.
D	The standard deviation has the same units as the original data.

2. Parameters and statistics:

A	Describe the same group of individuals.
B	Describe the population and the sample, respectively.
C	Describe the sample and the population, respectively.
D	None of these.

3. Which of the following location (central tendency) measures is affected by extreme values?

A	Median
B	Mean
C	Variance
D	Range

4. Which of the following measures can be used for the blood type in a given sample?

A	Mode
B	Mean
C	Variance
D	Range

5. If x_1,x_2 *and* x_3 has mean $\overline{x} = 4$, then x_1,x_2, x_3 and $x_4 = 4$ has mean:

A	equal 4
B	less than 4
C	greater than 4
D	None of this

6. The sample mean is a measure of

A	Relative position.
B	Dispersion.
C	Central tendency.
D	all of the above

7. The sample standard deviation is a measure of

A	Relative position.
B	Central tendency.
C	Dispersion.
D	all of the above.

8. Which of the following are examples of measures of dispersion?

A	The median and the mode.
B	The range and the variance.
C	The parameter and the statistic.
D	The mean and the variance.

9. If a researcher interests to study the blood pressure level (High, Normal, Low) for 13 diabetics patients, he may use:

A	Median and / or mode
B	Mean
C	Variance
D	Range

Question 20:

Find the mean and the variance for: 6 , 5 , 9 , 6 , 7 , 3

- $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{6+5+9+6+7+3}{6} = 6$

- $S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$

$$= \frac{(6-6)^2+(5-6)^2+(9-6)^2+(6-6)^2+(7-6)^2+(3-6)^2}{6-1} = 4$$

Question 21:

Find the mean and the variance: If $\sum_{i=1}^{6} X_i = 36$ and $\sum_{i=1}^{6} X_i^2 = 236.$

- $\overline{X} = \frac{\sum_{i=1}^{6} X_i}{6} = \frac{36}{6} = 6$

- $S^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}{n-1} = \frac{236-6\times 6^2}{6-1} = \frac{236-216}{5} = 4$

- **Find the median:**

Student's ages: 4, 5, 2, 9, 10, 8, 4

عدد المشاهدات فردي

$\boxed{2\ \ 4\ \ 4\ \ \boxed{5}\ \ 8\ \ 9\ \ 10} \Longrightarrow$ Median = 5

Student's ages: 10, 13, 9, 20, 11, 100

عدد المشاهدات زوجي

$\boxed{9\ \ 10\ \ \boxed{11\ \ 13}\ \ 20\ \ 100} \Longrightarrow \text{Median} = \frac{11+13}{2} = 12$

Student's grades: A, C, B, C, F, B, B

عدد المشاهدات فردي

$\boxed{A\ \ B\ \ B\ \ \boxed{B}\ \ C\ \ C\ \ F} \Longrightarrow$ Median = B

Student's grades: A, C, B, C, F, B, B, B

عدد المشاهدات زوجي

$\boxed{A\ \ B\ \ B\ \ \boxed{B\ \ B}\ \ C\ \ C\ \ F} \Longrightarrow$ Median = B

Student's grades: A, C, B, C, F, B, C, B

عدد المشاهدات زوجي

$\boxed{A\ \ B\ \ B\ \ \boxed{B\ \ C}\ \ C\ \ C\ \ F} \Longrightarrow$ No median

Question 22:

Suppose two samples of human males yield the following data (which is more variation)

	Sample 1 25 year	Sample 2 11 year
Mean weight	135 pound	60 pound
Standard deviation	10 pound	10 pound
Coefficient of variation (C.V)	$C.V_1 = \frac{S}{\overline{X}} \times 10$ $= \frac{10}{135} \times 100$ $= 7.41\%$	$C.V_2 = \frac{S}{\overline{X}} \times 100$ $= \frac{10}{60} \times 100$ $= 16.67\%$

Sample 2 has more variation tan sample 1

Question 23:

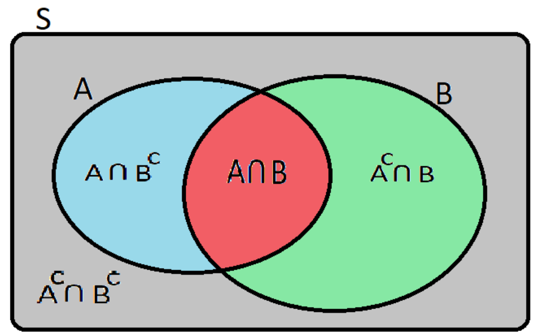
The following values are calculated in respect of heights and weights for sample of students, can we say that the weights shoe greater variation than the heights.

	Sample 1 height	Sample 2 weight
Mean	162.6 cm	52.36 kg
variance	127.69 cm^2	23.14 kg^2
Coefficient of variation (C.V)	$C.V_1 = \frac{S}{\overline{X}} \times 10$ $= \frac{\sqrt{127.69}}{162.6} \times 100$ $= 6.95\%$	$C.V_2 = \frac{S}{\overline{X}} \times 100$ $= \frac{\sqrt{23.14}}{52.36} \times 100$ $= 9.19\%$

Since CV_2 greater than CV_1 , therefore we can say the weights show more variability than height

Chapter 3 Probability

Probability



Definitions and Theorems:

$*\ 0 \le P(A) \le 1$

$*\ P(S) = 1$

$*\ P(\emptyset) = 0$

1- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
2- $P(A \mid B) = P(A \cap B)/P(B)$
3- $P(A \cap B) = P(A) \times P(B)$ (if A & B are independent)
4- $P(A \cap B) = 0$ (if A & B are disjoint)
5- $P(A^c) = 1 - P(A)$; $P(A^c) = P(\overline{A})$

Question 1:

Suppose that we have: $P(A) = 0.4$, $P(B) = 0.5$ *and* $P(A \cap B) = 0.2$

1. The probability $P(A \cup B)$ is:

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.4 + 0.5 - 0.2 = 0.7}$$

2. The probability $P(A \cap B^c)$ is:

$$\boxed{P(A \cap B^c) = P(A) - P(A \cap B) = 0.4 - 0.2 = 0.2}$$

3. The probability $P(A^c \cap B)$ is:

$$\boxed{P(A^c \cap B) = P(B) - P(A \cap B) = 0.5 - 0.2 = 0.3}$$

4. The probability $P(A|B)$ is:

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.2}{0.5} = 0.4}$$

5. The events A and B are:

$$\boxed{P(A \cap B) \stackrel{?}{=} P(A) \times P(B) \Longrightarrow 0.2 = 0.4 \times 0.5}$$

A	Disjoint	B	Dependent	C	Equal	D	Independent

Question 2:

If the events A, B we have: P(A) = 0.2, P(B) = 0.5 and P (A ∩ B) = 0.1, then:

1. The events A, B are:

$$\boxed{P(A \cap B) \stackrel{?}{=} P(A) \times P(B) \Longrightarrow 0.1 = 0.2 \times 0.5}$$

A	Disjoint	B	Dependent	C	Both are empties	D	Independent

2. The probability of A or B is:

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.2 + 0.5 - 0.1 = 0.6}$$

3. If P(A) = 0.3, P(B) = 0.4 and that A and B are disjoint, then P(A∪B) =

$$\boxed{P(A \cup B) = P(A) + P(B) - 0 = 0.3 + 0.4 - 0 = 0.7}$$

4. If P(A) = 0.2 and P(B | A) =0.4, then P(A ∩ B) =

$$\boxed{P(B|A) = \frac{P(A \cap B)}{P(A)} \Longrightarrow 0.4 = \frac{P(A \cap B)}{0.2} \Longrightarrow P(A \cap B) = 0.2 \times 0.4 = 0.08}$$

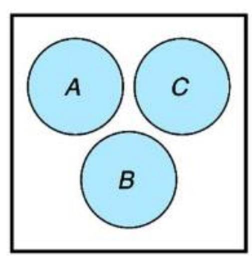
5. Suppose that the probability a patient smoke is 0.20. If the probability that the patient smokes and has a lung cancer is 0.15, then the probability that the patient has a lung cancer given that the patient smokes is

$$\boxed{\begin{array}{c} P(S) = 0.20 \quad P(S \cap C) = 0.15 \quad P(C|S) = ? \\ P(C|S) = \frac{P(C \cap S)}{P(S)} = \frac{0.15}{0.20} = 0.75 \end{array}}$$

Question 3:

The probability of three mutually exclusive events A, B and C are given by $1/3$, $1/4$ and $5/12$ then P(A ∪ B ∪ C)

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &= \frac{1}{3} + \frac{1}{4} + \frac{5}{12} = 1 \end{aligned}$$



A	0.57	B	0.43	C	0.58	D	1

Question 4:

Suppose that we have two events A and B such that,

$$P(A) = 0.4, P(B) = 0.5, P(A \cap B) = 0.2.$$

[1] $P(A \cup B)$: $\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.4 + 0.5 - 0.2 = 0.7}$

[2] $P(A^c \cap B)$: $\boxed{P(A^c \cap B) = P(B) - P(A \cap B) = 0.5 - 0.2 = 0.3}$

[3] $P(A^c \cap B^c)$: $\boxed{P(A^c \cap B^c) = 1 - P(A \cup B) = 1 - 0.7 = 0.3}$

[4] $P(A^c)$: $\boxed{P(A^c) = 1 - P(A) = 1 - 0.4 = 0.6}$

[5] $P(A^c \mid B)$: $\boxed{P(A^c \mid B) = \frac{P(A^c \cap B)}{P(B)} = \frac{0.3}{0.5} = 0.6}$

[6] $P(B \mid A)$: $\boxed{P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{0.2}{0.4} = 0.5}$

[7] The events A and B are …

$$\boxed{P(A \cap B) \stackrel{?}{=} P(A) \times P(B) \Longrightarrow 0.2 = 0.4 \times 0.5}$$

A	Exhaustive	B	Dependent	C	Equal	D	Independent

Question 5:

Let A and B two events defined on the same sample space.

If $P(A) = 0.7, P(B) = 0.3$

1. If the events A and B are mutually exclusive (disjoint) then, the value of $P(A \cup B)$.

$$P(A \cup B) = P(A) + P(B) - 0 = 0.7 + 0.3 - 0 = 1$$

A	0.21	B	0.52	C	0.79	D	1

2. If the events A and B are independent, then the value of $P(A|B)$.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A) = 0.7$$

A	0.3	B	0.5	C	0.7	D	0.9

3. If the events A and B are independent, then the value of $P(A \cap \bar{B})$.

$$P(A \cap \bar{B}) = P(A)P(\bar{B}) = (0.7)(0.7)$$

A	0.09	B	0.21	C	0.49	D	0.54

4. If the events A and B are independent, then the value of $P(\overline{A \cup B})$.

$$P(\overline{A \cup B}) = 1 - P(A \cup B)$$

$$= 1 - [0.7 + 0.3 - (0.7)(0.3)] = 0.21$$

A	0.21	B	0.39	C	0.49	D	0.54

Question 6:

Following table shows 80 patients classified by sex and blood group.

Sex	Blood Group		
	A	B	O
Male (M)	25	17	15
Female (F)	11	9	3

1) The probability that a patient selected randomly is a male and has blood group A is

A	25/36	B	25/80	C	25/57	D	52/80

2) The probability that a patient selected randomly is a female is

A	6/80	B	40/80	C	23/80	D	None

3) In a certain population, 4% have cancer, 20% are smokers and 2% are both smokers and have cancer. If a person is chosen at random from the population, find the probability that the person chosen is a smoker or has cancer.

$$P(C) = 0.04 \quad P(S) = 0.20 \quad P(S \cap C) = 0.02 \quad P(S \cup C) = ?$$
$$P(S \cup C) = P(C) + P(S) - P(S \cap C)$$
$$P(S \cup C) = 0.04 + 0.20 - 0.02 = 0.22$$

Question 7:

Gender	Diabetics (D)	Not Diabetic (D^c)	TOTAL
Male (M)	72	288	360
Female (F)	48	192	240
Total	120	480	600

Consider the information given in the table above. A person is selected randomly

7. The probability that the person found is male and diabetic is:

$$\boxed{P(M \cap D) = \frac{72}{600} = 0.12}$$

8. The probability that the person found is male or diabetic is:

$$\boxed{P(M \cup D) = P(M) + P(D) - P(M \cap D) = \frac{360}{600} + \frac{120}{600} - \frac{72}{600} = \frac{408}{600}}$$

9. The probability that the person found is female is:

$$\boxed{P(F) = \frac{240}{600} = 0.4}$$

10. Suppose we know the person found is a male, the probability that he is diabetic, is:

$$\boxed{P(D|M) = \frac{P(M \cap D)}{P(M)} = \frac{72/600}{360/600} = \frac{72}{360} = 0.2}$$

11. The events M and D are:

$$\boxed{P(M \cap D) = P(M) \times P(D) \Longrightarrow \frac{72}{600} = \frac{360}{600} \times \frac{120}{600}}$$

A	Mutually exclusive	B	Dependent	C	Equal	D	Independent

Question 8:

A group of people is classified by the number of fruits eaten and the health status:

Health Status \ Fruits Eaten	Few (F)	Some (S)	Many (M)	Total
Poor (B)	80	35	20	135
Good (G)	25	110	45	180
Excellent (E)	15	95	75	185
Total	120	240	140	500

If one of these people is randomly chosen give:

1. The event "(eats few fruits) and (has good health) ", is defined as.

A	$F \cup G^c$	B	$F \cap G$	C	$F \cup E$	D	$S \cup E$

2. $P(B \cup M) =$

A	0.51	B	0.28	C	0.27	D	0.04

3. $P(G \cap S) =$

A	0.48	B	0.36	C	0.22	D	0.62

4. $P(E^c) =$

A	0.63	B	0.37	C	0.50	D	1

5. $P(G|S) =$

A	0.6111	B	0.2200	C	0.4583	D	0.36

6. $P(M|E) =$

A	0.6111	B	0.2200	C	0.405	D	0.36

Question 9:

The following table classifies a sample of individuals according to gender and period (in years) attendance in the college:

College Attended	Gender		
	Male	Female	Total
None	12	41	53
Two Years	14	63	77
Three Years	9	49	58
Four Years	7	50	57
Total	42	203	245

Suppose we select an individual at random, then:

1. The probability that the individual is male is:

A	0.8286	B	0.1714	C	0.0490	D	0.2857

2. The probability that the individual did not attend college (None) and female is:

A	0.0241	B	0.0490	C	0.1673	D	0.2163

3. The probability that the individual has three year or two-year college attendance is:

A	0.551	B	0.0939	C	0.4571	D	0

4. If we pick an individual at random and found that he had three-year college attendance, the probability that the individual is male is:

A	0.0367	B	0.2143	C	0.1552	D	0.1714

5. The probability that the individual is not a four-year college attendance is:

A	0.7673	B	0.2327	C	0.0286	D	0.1429

6. The probability that the individual is a two-year college attendance or male is:

A	0.0571	B	0.8858	C	0.2571	D	0.4286

7. The events: the individual is a four-year college attendance and male are:

A	Mutually exclusive	B	Independent	C	Dependent	D	None of these

<u>Question 10:</u>

	Blood pressure		
	Low (L)	Medium (M)	High (H)
Has obesity (B)	50	150	300
Does not have obesity ($\bar{B}$)	250	240	110

If an individual is selected at random from this group, then the probability that he/she

1. has obesity or has medium blood pressure is equal to

A	0.442	B	0.50	C	0.725	D	0.673

2. has low blood pressure given that he/she has obesity is equal to

A	0.90	B	0.1	C	0.66	D	0.44

- Bayes' Theorem, Screening Tests, Sensitivity, Specificity, and Predictive Value Positive and Negative

	Disease		
Test Result	Present (D)	Absent ($\bar{D}$)	Total
Positive (T)	a	b	
Negative ($\bar{T}$)	c	d	
Total	a+c	b+d	n

Sensitivity $P(T\|D)=\frac{a}{a+c}$	Probability of false positive (f +) $P(T\|\bar{D})=\frac{b}{b+d}$
Probability of false negative (f -) $P(\bar{T}\|D)=\frac{c}{a+c}$	Specificity $P(\bar{T}\|\bar{D})=\frac{d}{b+d}$

- <u>The predictive value positive:</u>

$$P(D|T)=\frac{P(D\cap T)}{P(T)}=\frac{P(T|D)P(D)}{P(T|D)P(D)+P(T|\bar{D})P(\bar{D})}$$

$$=\frac{(Sen)\times P(D_{given})}{(Sen)\times P(D_{given})+(f+)\times P(\bar{D}_{given})}$$

- <u>The predictive value negative:</u>

$$P(\bar{D}|\bar{T})=\frac{P(\bar{T}\cap\bar{D})}{P(\bar{T})}=\frac{P(\bar{T}|\bar{D})P(\bar{D})}{P(\bar{T}|\bar{D})P(\bar{D})+P(\bar{T}|D)P(D)}$$

$$=\frac{(Spe)\times P(\bar{D}_{given})}{(Spe)\times P(\bar{D}_{given})+(f-)\times P(D_{given})}$$

Question 1:

The following table shows the results of a screening test:

	Disease confirmed (D)	Disease not confirmed ($\overline{D}$)
Positive test (T)	38	10
Negative test ($\overline{T}$)	5	18

1. The probability of false positive of the test is: $\boxed{\frac{10}{28} = 0.3571}$

2. The probability of false negative of the test is: $\boxed{\frac{5}{43} = 0.1163}$

3. The sensitivity value of the test is: $\boxed{\frac{38}{43} = 0.8837}$

4. The specificity value of the test is: $\boxed{\frac{18}{28} = 0.6429}$

Suppose it is known that the rate of the disease is 0.113,

$\boxed{1 - 0.113 = 0.887}$

5. The predictive value positive of a symptom is:

$$= \frac{(\text{Sen})\times P(D_{\text{given}})}{(\text{Sen})\times P(D_{\text{given}})+(f+)\times P(\overline{D}_{\text{given}})} = \frac{0.8837 \times 0.113}{0.8837\times 0.113+0.3571\times 0.887} = 0.2397$$

6. The predictive value negative of a symptom is:

$$= \frac{(\text{Spe})\times P(\overline{D}_{\text{given}})}{(\text{Spe})\times P(\overline{D}_{\text{given}})+(f-)\times P(D_{\text{given}})} = \frac{0.6429 \times 0.887}{0.6429 \times 0.887+0.1163 \times 0.113} = 0.9772$$

Question 2:

It is known that 40% of the population is diabetic. 330 persons who were diabetics went through a test where the test confirmed the disease for 288 persons. Among 270 healthy persons, test showed high sugar level for 72 persons. The information obtained is given in the table below.

Test	Diabetics (D)	Not Diabetic (D^c)	TOTAL
Positive (T)	288	72	360
Negative ($\bar{T}$)	42	198	240
TOTAL	330	270	600

1. The sensitivity of the test is: $\boxed{\frac{288}{330} = 0.873}$

2. The specificity of the test is: $\boxed{\frac{198}{270} = 0.733}$

3. The probability of false positive is: $\boxed{\frac{72}{270} = 0.267}$

4. The predictive probability positive for the disease is:

$$\boxed{= \frac{(\text{Sen})\times \text{P}(\text{D}_{\text{given}})}{(\text{Sen})\times \text{P}(\text{D}_{\text{given}})+(\text{f}+)\times\text{P}(\bar{\text{D}}_{\text{given}})} = \frac{0.873 \times 0.40}{0.873 \times 0.40 + 0.267 \times 0.60} = 0.686}$$

Question 3:

The following table shows the results of a screening test evaluation in which a random sample of 700 subjects with the disease and an independent random sample of 1300 subjects without the disease participated:

Disease Test result	Present	Absent
Positive	500	100
Negative	200	1200

1. The sensitivity value of the test is: $\boxed{\frac{500}{700} = 0.7143}$
2. The specificity value of the test is: $\boxed{\frac{1200}{1300} = 0.923}$
3. The probability of false positive of the test is: $\boxed{\frac{100}{1300} = 0.0769}$
4. If the rate of the disease in the general population is 0.002, then the predictive value positive of the test is:

$$\boxed{= \frac{(\text{Sen})\times \text{P}(\text{D}_{\text{given}})}{(\text{Sen})\times \text{P}(\text{D}_{\text{given}})+(\text{f}+)\times\text{P}(\bar{\text{D}}_{\text{given}})} = \frac{0.7143 \times 0.002}{0.7143 \times 0.002 + 0.0769 \times 0.998} = 0.01827}$$

Question 4:

In a study of high blood pressure, 188 persons found positive, of a sample of 200 persons with the disease subjected to a screening test. While, 27 persons found positive, of an independent sample of 300 persons without the disease subjected to the same screening test. That is,

	High Blood Pressure		
Test Result	Yes D	No $\bar{D}$	Total
Positive T	188	27	215
Negative $\bar{T}$	12	273	285
Total	200	300	500

[1] Given that a person has the disease, the probability of a positive test result, that is, the "sensitivity" of this test is:

A	0.49	**B**	**0.94**	**C**	0.35	**D**	0.55

[2] Given that a person does not have the disease, the probability of a negative test result, that is, the "specificity" of this test is:

A	**0.91**	**B**	0.75	**C**	0.63	**D**	0.49

[3] The "false negative" results when a test indicates a negative status given that the true status is positive is:

A	0.01	**B**	0.15	**C**	0.21	**D**	**0.06**

[4] The "false positive" results when a test indicates a positive status given that the true status is negative is:

A	0.16	**B**	0.31	**C**	**0.09**	**D**	0.02

Assuming that 15% of the population under study is known to be with high blood pressure.

[5] Given a positive screening test, what is the probability that the person has the disease? That is, the "predictive value positive" is:

A	0.22	**B**	**0.65**	**C**	0.93	**D**	0.70

[6] Given a negative screening test result, what is the probability that the person does not have the disease? That is, the "predictive value negative" is:

A	0.258	**B**	0.778	**C**	**0.988**	**D**	0.338

Question 5:

Suppose that the ministry of health intends to check the reliability of the central Diabetic Lab in Riyadh. A sample person with Diabetic disease (D) and another without the disease ($\bar{D}$) had the Lab tests and the results are given below:

	Present (D)	Absence ($\bar{D}$)
Positive (T)	950	40
Negative ($\bar{T}$)	25	640

Then:

1. The probability of false positive of the test is: $\boxed{\frac{40}{680} = 0.0588}$

2. The probability of false negative of the test is: $\boxed{\frac{25}{975} = 0.0256}$

3. The sensitivity value of the test is: $\boxed{\frac{950}{975} = 0.9744}$

4. The specificity value of the test is: $\boxed{\frac{640}{680} = 0.9412}$

Assume that the true percentage of Diabetic patients in Riyadh is 25%. Then

5. The predictive value positive of the test is:

$$= \frac{(\text{Sen})\times P(D_{\text{given}})}{(\text{Sen})\times P(D_{\text{given}}) + (f+)\times P(\bar{D}_{\text{given}})} = \frac{0.9744 \times 0.25}{0.9744 \times 0.25 + 0.0588 \times 0.75} = 0.8467$$

6. The predictive value negative of the test is:

$$= \frac{(\text{Spe})\times P(\bar{D}_{\text{given}})}{(\text{Spe})\times P(\bar{D}_{\text{given}}) + (f-)\times P(D_{\text{given}})} = \frac{0.9412 \times 0.75}{0.9412 \times 0.75 + 0.0256 \times 0.25} = 0.9910$$

Question 6:

A Fecal Occult Blood Screen Outcome Test is applied for 875 patients with bowel cancer. The same test was applied for another sample of 925 without bowel cancer. Obtained results are shown in the following table:

	Present Disease (D)	*Absent Disease* ($\bar{D}$)
Test Positive (T)	*850*	*10*
Test Negative ($\bar{T}$)	*25*	*915*

1. The probability of false positive of the test is: $\boxed{\frac{10}{925} = 0.0108}$

2. The probability of false negative of the test is: $\boxed{\frac{25}{875} = 0.0286}$

3. The sensitivity value of the test is: $\boxed{\frac{850}{875} = 0.9714}$

4. The specificity value of the test is: $\boxed{\frac{915}{925} = 0.9892}$

5. If the rate of the disease in the general population is equal to 15% then the predictive value positive of the test is

$$= \frac{(\text{Sen})\times P(D_{\text{given}})}{(\text{Sen})\times P(D_{\text{given}}) + (f+)\times P(\bar{D}_{\text{given}})} = \frac{0.9714 \times 0.15}{0.9714 \times 0.15 + 0.0108 \times 0.85} = 0.9407$$

More Exercises

Question 1:

Givens:

$$\boldsymbol{P(A) = 0.5, \quad P(B) = 0.4, \quad P(C \cap A^c) = 0.6,}$$
$$\boldsymbol{P(C \cap A) = 0.2, \quad P(A \cup B) = 0.9}$$

(a) ***What is the probability of*** $\boldsymbol{P(C)}$**:**

$$P(C) = P(C \cap A^c) + P(C \cap A) = 0.6 + 0.2 = 0.8$$

(b) ***What is the probability of*** $\boldsymbol{P(A \cap B)}$**:**

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ \Rightarrow \quad 0.9 &= 0.5 + 0.4 - P(A \cap B) \\ P(A \cap B) &= 0 \end{aligned}$$

(c) ***What is the probability of*** $\boldsymbol{P(C \mid A)}$**:**

$$P(C \mid A) = \frac{P(C \cap A)}{P(A)} = \frac{0.2}{0.5} = 0.4$$

(d) ***What is the probability of*** $\boldsymbol{P(B^c \cap A^c)}$:

$$P(B^c \cap A^c) = 1 - P(B \cup A) = 1 - 0.9 = 0.1$$

Question 2:

Givens:

$$\boldsymbol{P(B) = 0.3, \quad P(A \mid B) = 0.4}$$

$\boldsymbol{Then\ find\ P(A \cap B)} = ?$

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ \Rightarrow 0.4 &= \frac{P(A \cap B)}{0.3} \\ \Rightarrow P(A \cap B) &= 0.4 \times 0.3 = 0.12 \end{aligned}$$

Question 3:

Givens:

$$\boldsymbol{P(A) = 0.3,\quad P(B) = 0.4,\quad P(A \cap B \cap C) = 0.03,\quad P(\overline{A \cap B}) = 0.88}$$

(1) Are the event A and b independent?

$$P(A \cap B) = 1 - P(\overline{A \cap B}) = 1 - 0.88 = 0.12$$

$$P(A) \times P(B) = 0.3 \times 0.4 = 0.12$$
$$\Rightarrow P(A \cap B) = P(A) \times P(B)$$

Therefore, A and B are independent.

(2) ***What is the probability of $\boldsymbol{P(C \mid A \cap B)}$***:

$$P(C \mid A \cap B) = \frac{P(A \cap B \cap C)}{P(A \cap B)} = \frac{0.03}{0.12} = 0.25$$

Question 4:

Givens:

$$\boldsymbol{P(A_1) = 0.4,\quad P(A_1 \cap A_2) = 0.2,\quad P(A_3 \mid A_1 \cap A_2) = 0.75}$$

(1) ***Find the $\boldsymbol{P(A_2|A_1)}$***:

$$P(A_2|A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)} = \frac{0.2}{0.4} = 0.5$$

(2) ***Find the $\boldsymbol{P(A_1 \cap A_2 \cap A_3)}$***:

$$P(A_3 \mid A_1 \cap A_2) = \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)}$$

$$0.75 = \frac{P(A_1 \cap A_2 \cap A_3)}{0.2}$$

$$P(A_1 \cap A_2 \cap A_3) = 0.75 \times 0.2 = 0.15$$

Exercise 1:

A group of 400 people are classified according to their nationality as (250 Saudi and 150 non-Saudi), and they are classified according to their gender (100 Male and 300 female). The number of Saudi males is 60. Suppose that the experiment is to select a person at random from this group.

1. Summarizing the information in a table:

		Gender		Total
		Male (M)	Female (F)	
Nationality	Saudi (S)	**60**	190	**250**
	Non-Saudi (N)	40	110	**150**
Total		**100**	**300**	**400**

2. The probability that the selected person is Saudi is:
(A) 0.6
(B) 0.15
(C) 0.3158
(D) 0.625 (P(S) = 250/400)

3. The probability that the selected person is female is:
(A) 0.375
(B) 0.75 (P(F) = 300/400)
(C) 0.3667
(D) 0.6333

4. The probability that the selected person is female given that the selected person is Saudi is:
(A) 0.6333
(B) 0.3667
(C) 0.76 (P(F | S) = 190/250)
(D) 0.475

5. The events "S"={Selecting a Saudi} and "F"={Selecting a female} are:
(A) Not independent events (Because: P(F) ≠ P(F | S))
(B) Complement of each other
(C) Independent events
(D) Disjoint (mutually exclusive) events

Exercise 2:

A new test is being considered for diagnosis of leukemia. To evaluate this test, the researcher has applied this test on 30 leukemia patient persons and 50 non-patient persons. The results of the screen test applied to those people are as follows: positive results for 27 of the patient persons, and positive results for 4 of the non-patient persons, and negative results for the rest of persons.

1. Summarizing the information in a table:

		Leukemia Disease		Total
		D^+	D^-	
Result of the Test	T^+	**27**	**4**	31
	T^-	3	46	49
Total		**30**	**50**	80

2. The sensitivity of the test is:
(A) 0.3375
(B) 0.9 (P(T^+ | D^+) = 27/30)
(C) 0.88
(D) 0.1111

3. The specificity of the test is:
(A) 0.92 (P(T^- | D^-) = 46/50)
(B) 0.575
(C) 0.087
(D) 0.9388

4. The probability of false positive result is: (FP)
(A) 0.05
(B) 0.1481
(C) 0.1290
(D) 0.08 (P(T^+ | D^-) = 4/50) { Note: P(FP) = 1 – Specificity }

5. The probability of false negative result is: (FN)
(A) 0.0652
(B) 0.1111
(C) 0.1 (P(T^- | D^+) = 3/30) { Note: P(FN) = 1 – Sensitivity }
(D) 0.0375

<u>Exercise3:</u>

A new test is being considered for diagnosis of leukemia. To evaluate this test, the researcher has applied this test on a group of people and found that the sensitivity of the test was 0.92 and the specificity of the test was 0.94. Based on another independent study, it is found that the percentage of infected people with leukemia in the population is 5% (the rate of prevalence of the disease).

Given information:

Sensitivity = P(T^+ | D^+) = 0.92
specificity = P(T^- | D^-) = 0.94
P(D) = 0.05

1. The predictive value positive is:
(A) 0.4466 (P(D^+ | T^+) = Bayes rule)
(B) 0.3987
(C) 0.9328
(D) 0.6692

2. The predictive value negative is:
(A) 0.7841
(B) 0.9955 (P(D^- | T^-) = Bayes rule)
(C) 0.8774
(D) 0.3496

<u>Exercise</u>: (Hypothetical Example)

A new proposed test is being considered for diagnosis of Corona (COVID-19) disease. To investigate the efficiency of this test, the researcher has applied this test on 80 infected patients and 900 non-infected persons. The results of the screen test are given in the following table:

		Nature of the Disease		Total
		(Present: D^+) Infected Patients	(Absent: D^-) Non-infected People	
Result of the Test	$+ve$ (T^+)	75 (TP)	10 (FP)	85
	$-ve$ (T^-)	5 (FN)	890 (TN)	895
Total		80	900	980

Based on another independent study, it is found that the percentage of infected people with Corona (COVID-19) in this city is 4% (the rate of prevalence of the disease).

<u>1. Before-Test Questions:</u>

a) If a person was infected (D^+), what is the probability that the result of the test will be $+ve$ (T^+)?

$$\boldsymbol{P(T^+|D^+) = \textbf{Sensitivity of the Test}}$$

b) If a person was infected (D^+), what is the probability that the result of the test will be $-ve$ (T^-)?

$$\boldsymbol{P(T^-|D^+) = \textbf{False Negative Result (FNR)}}$$
$$\boldsymbol{= 1 - P(T^+|D^+)}$$
$$\boldsymbol{= 1 - \textbf{Sensitivity of the Test}}$$

c) If a person was not infected (D^-), what is the probability that the result of the test will be $-ve$ (T^-)?

$$\boldsymbol{P(T^-|D^-) = \textbf{Specificity of the Test}}$$

d) If a person was not infected (D^-), what is the probability that the result of the test will be $+ve$ (T^+)?

$$\boldsymbol{P(T^+|D^-) = \textbf{False Positive Result (FPR)}}$$
$$\boldsymbol{= 1 - P(T^-|D^-)}$$
$$\boldsymbol{= 1 - \textbf{Specificity of the Test}}$$

2. After-Test Questions:

a) If the result of the test was $+ve$ (T^+), what is the probability that the person is infected (D^+)?

$$\boldsymbol{P(D^+|T^+) = \textbf{Predictive Value Positive (PVP)}}$$

b) If the result of the test was $-ve$ (T^-), what is the probability that the person is not infected (D^-)?

$$\boldsymbol{P(D^-|T^-) = \textbf{Predictive Value Negative (PVN)}}$$

3. Efficiency of the Test:

$$\textbf{Efficiency} = \frac{\boldsymbol{True\ Positives + True\ Negatives}}{\boldsymbol{Total}} = \frac{\boldsymbol{TP + TN}}{\boldsymbol{n}}$$

Solution:

1. Before-Test Questions:

(a) The probability that the result of the test will be $+ve$ given that the person was infected is: (Sensitivity of the Test)

$$P(T^+|D^+) = \frac{P(T^+ \cap D^+)}{P(D^+)} = \frac{n(T^+ \cap D^+)}{n(D^+)} = \frac{75}{80} = 0.9375$$

(b) The probability that the result of the test will be $-ve$ given that the person was infected is: (False Negative Result =FNR)

$$P(T^-|D^+) = 1 - P(T^+|D^+) = 1 - 0.9375 = 0.0625$$

(c) The probability that the result of the test will be $-ve$ given that the person was not infected is: (Specificity of the test)

$$P(T^-|D^-) = \frac{P(T^- \cap D^-)}{P(D^-)} = \frac{n(T^- \cap D^-)}{n(D^-)} = \frac{890}{900} = 0.9889$$

(d) The probability that the result of the test will be $+ve$ given that the person was not infected is: (False Positive Result = FPR)

$$P(T^+|D^-) = 1 - P(T^-|D^-) = 1 - 0.9889 = 0.0111$$

2. After-Test Questions:
Define the following events:

D = {A randomly chosen person from the city is infected} → 4%
$P(D) = \frac{4}{100} = 0.04$

$\overline{D}$ = {A randomly chosen person from the city is not infected}
$P(\overline{D}) = 1 - P(D) = 1 - 0.04 = 0.96$

(a) The probability that the person is infected (D), given that the result was $+ve$ (T^+) is: (Predictive Value Positive = PVP)

$$\begin{aligned} P(D|T^+) &= \frac{P(D \cap T^+)}{P(T^+)} \\ &= \frac{P(T^+|D)P(D)}{P(T^+|D)P(D) + P(T^+|\overline{D})P(\overline{D})} \\ &= \frac{0.9375 \times 0.04}{0.9375 \times 0.04 + 0.0111 \times 0.96} \\ &= \frac{0.0375}{0.0375 + 0.010656} \\ &= \frac{0.0375}{0.048156} \\ &= 0.7787 \end{aligned}$$

(b) The probability that the person is not infected ($\overline{D}$), given that the result was $-ve$ (T^-) is: (Predictive Values Negative = PVN)

$$\begin{aligned} P(\overline{D}|T^-) &= \frac{P(\overline{D} \cap T^-)}{P(T^-)} \\ &= \frac{P(T^-|\overline{D})P(\overline{D})}{P(T^-|\overline{D})P(\overline{D}) + P(T^-|D)P(D)} \\ &= \frac{0.9889 \times 0.96}{0.9889 \times 0.96 + 0.0625 \times 0.04} \\ &= \frac{0.949344}{0.949344 + 0.00254} \\ &= \frac{0.949344}{0.951884} = 0.9974 \end{aligned}$$

3. Efficiency of the Test:

$$\text{Efficiency} = \frac{True\ Positives + True\ Negatives}{Total}$$

$$= \frac{TP + TN}{n}$$

$$= \frac{75 + 890}{980}$$

$$= \frac{965}{980}$$

$$= 0.9847$$