بسم الله الرحمن الرحيم

STAT 109

BIOSTATISTICS

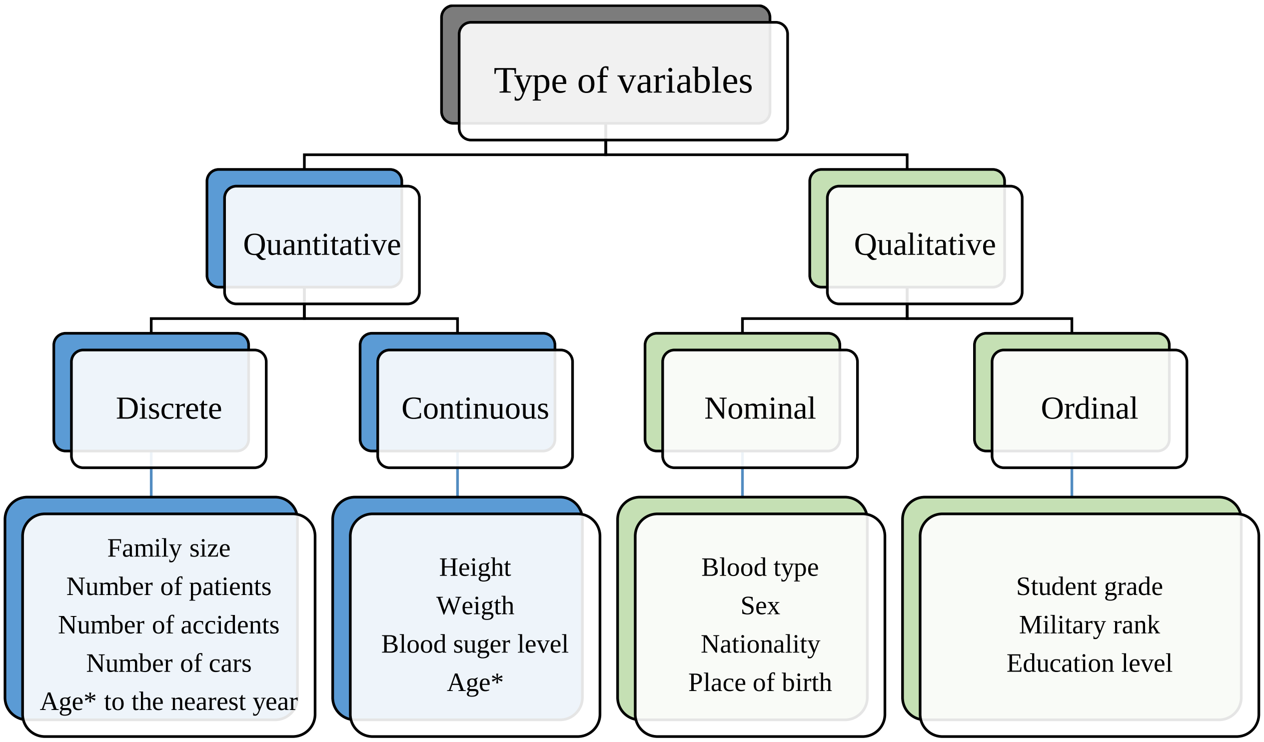
Prepared by
Abdulrahman Alfaifi
King Saud University
Department of Statistics and Operation Research
alfaifi@ksu.edu.sa
@AlfaifiStat
A Alfaifi

ملاحظة : ليس بالضرورة ان تكون المذكرة شاملة للمقرر

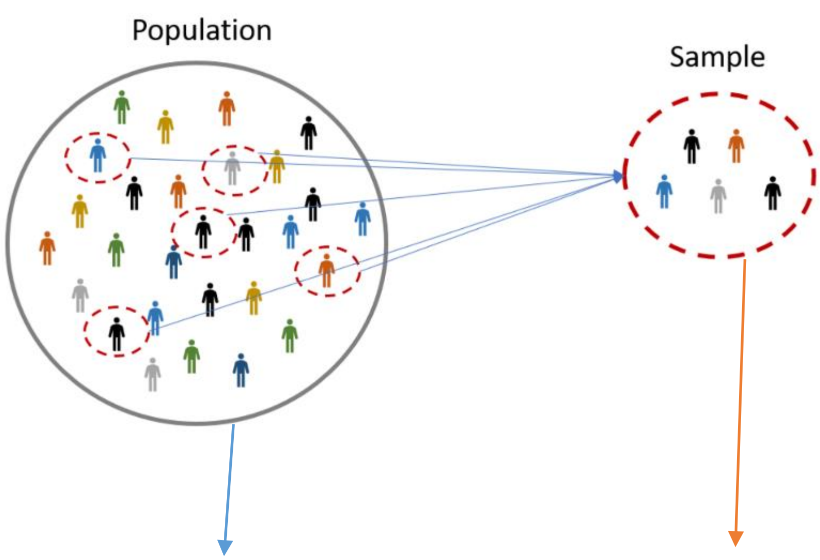
17 August 2023

Chapter 1 Introduction

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



- <u>Statistical Inference:</u>



Parameter

Population mean μ
Population variance σ^2

Statistic

Sample mean $\overline{X}$
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

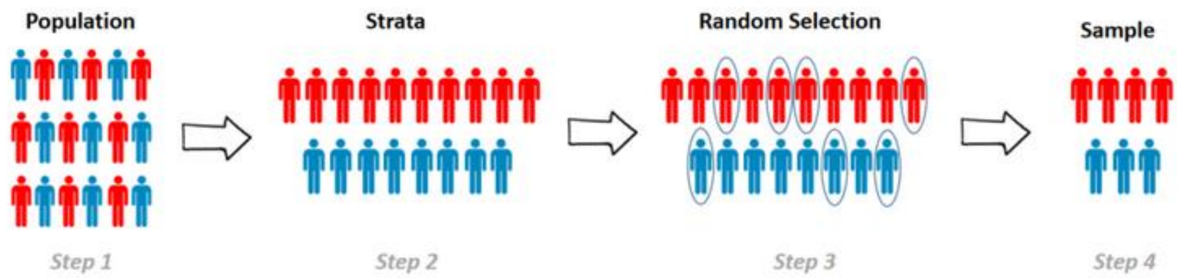
2. The population in this study is:

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.

- **<u>Stratified Random Sampling:</u>**



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:**

<u>Question 1:</u>

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

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: $A = 40 - (8 + 15 + 11) = 40 - 34 = 6$

2) The value of B is: $B = 8 + 15 + 11 = 34$

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

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

5) The true class interval for the first class is: $4.5 - 9.5$

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

Question 3:

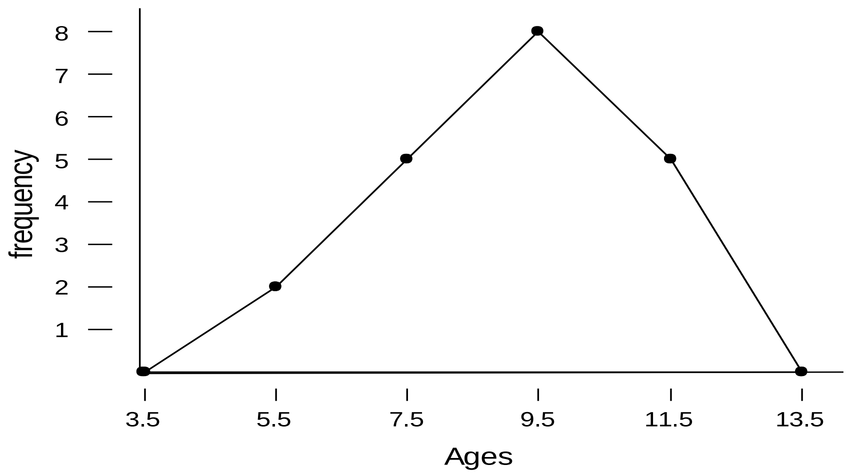
The table shows the weight loss (kg) of a sample of 40 healthy adults who fasted in Ramadan.

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 18
2) The value of the missing value C is 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)

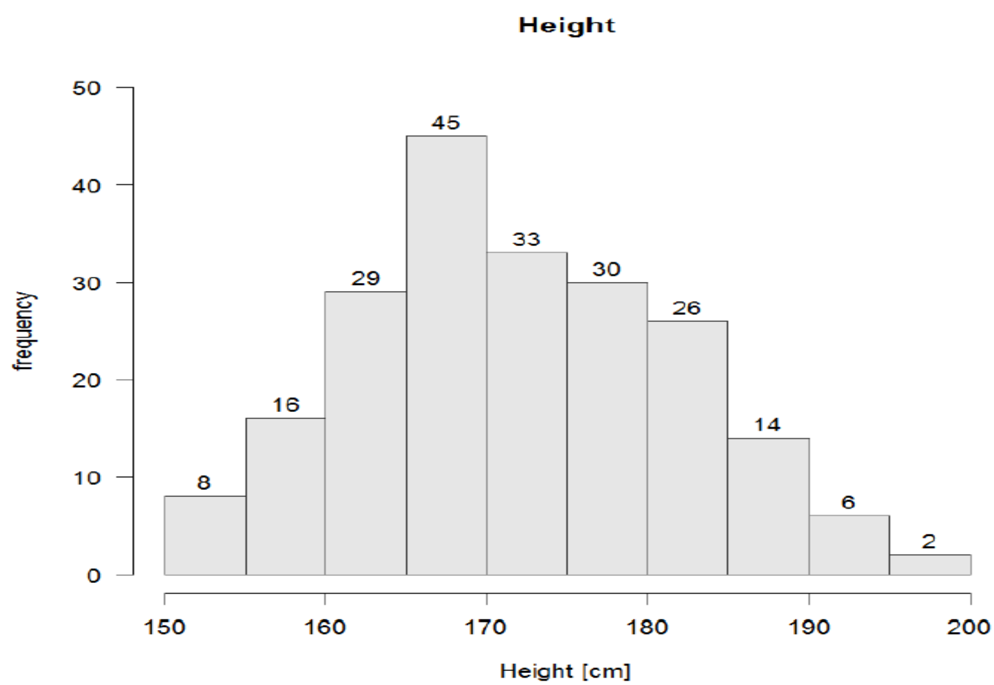
Class interval	5- 6	7-8	9- 10	11 -12
frequency	1	7	8	4

(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 5^{th} 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:

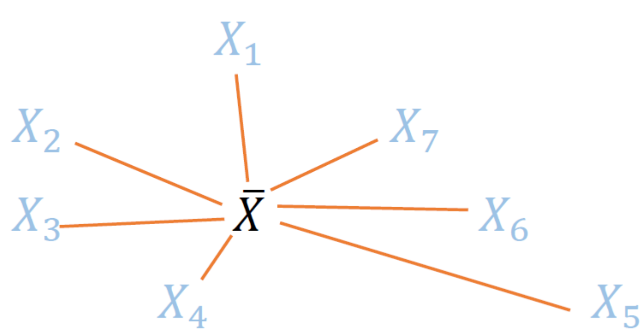
A	0.1556	B	7	C	45	D	15.56%

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



Question 10:

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:

   discrete

2. The sample mean is:

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

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:

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

5. The coefficient of variation is:

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

6. The range is:

   $$16 - 12 = 4$$

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:

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

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:

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

6. The coefficient of variation for this data is:

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

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

$$C.V_{Cont} = \frac{s_{Cont}}{\bar{X}_{Cont}} \times 100 = \frac{4.33}{6.5} \times 100 = 66.61\%$$

$$C.V_{Treat} = \frac{s_{Treat}}{\bar{X}_{Treat}} \times 100 = \frac{5.32}{7.6} \times 100 = 70\%$$

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:

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

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 $\bar{x} = 4$, 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

| 2 4 4 [5] 8 9 10 | $\Longrightarrow$ Median = 5

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

عدد المشاهدات زوجي

| 9 10 [11 13] 20 100 | $\Longrightarrow$ Median $= \frac{11+13}{2} = 12$

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

عدد المشاهدات فردي

| A B B [B] C C F | $\Longrightarrow$ Median = B

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

عدد المشاهدات زوجي

| A B B [B B] C C F | $\Longrightarrow$ Median = B

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

عدد المشاهدات زوجي

| A B B [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)	$C.V_1 = \frac{S}{\bar{\bar{X}}} \times 10$ $= \frac{10}{135} \times 100$ $= 7.41\%$	$C.V_2 = \frac{S}{\bar{\bar{X}}} \times 100$ $= \frac{10}{60} \times 100$ $= 16.67\%$

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)	$C.V_1 = \frac{S}{\bar{\bar{X}}} \times 10$ $= \frac{\sqrt{127.69}}{162.6} \times 100$ $= 6.95\%$	$C.V_2 = \frac{S}{\bar{\bar{X}}} \times 100$ $= \frac{\sqrt{23.14}}{52.36} \times 100$ $= 9.19\%$

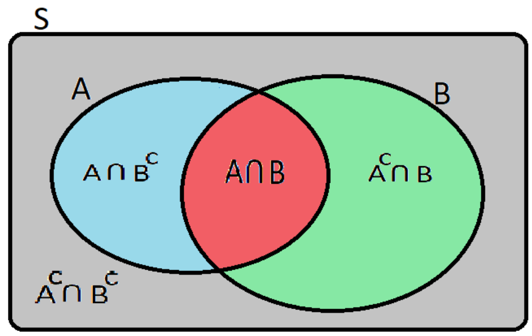
Since CV_2 greater than CV_1 , therefore we can say the weights show more variability than height

Chapter 3 Probability

Probability

Definitions and Theorems:

* $0 \leq P(A) \leq 1$
* $P(S) = 1$
* $P(\emptyset) = 0$



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:

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

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

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

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

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

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

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

5. The events A and B are:

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

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:

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

A	Disjoint	B	Dependent	C	Both are empties	D	Independent

2. The probability of A or B is:

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

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

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

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

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

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

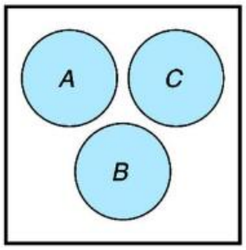
$$P(S) = 0.20 \quad P(S \cap C) = 0.15 \quad P(C|S) = ?$$

$$P(C|S) = \frac{P(C \cap S)}{P(S)} = \frac{0.15}{0.20} = 0.75$$

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)

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



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:

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

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

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

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

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

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

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

11. The events M and D are:

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

A	Mutually exclusive	B	Dependent	C	Equal	D	Independent

<u>Question 8:</u>

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

Fruits Eaten / Health Status	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 $P(\bar{T}\|\bar{D}) = \frac{d}{b+d}$

- The predictive value positive:

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

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

- The predictive value negative:

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

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

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: $\frac{10}{28} = 0.3571$
2. The probability of false negative of the test is: $\frac{5}{43} = 0.1163$
3. The sensitivity value of the test is: $\frac{38}{43} = 0.8837$
4. The specificity value of the test is: $\frac{18}{28} = 0.6429$

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

$1 - 0.113 = 0.887$

5. The predictive value positive of a symptom is:

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

6. The predictive value negative of a symptom is:

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

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: $\frac{288}{330} = 0.873$

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

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

4. The predictive probability positive for the disease is:

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

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: $\frac{500}{700} = 0.7143$

2. The specificity value of the test is: $\frac{1200}{1300} = 0.923$

3. The probability of false positive of the test is: $\frac{100}{1300} = 0.0769$

4. If the rate of the disease in the general population is 0.002, then the predictive value positive of the test is:

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

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

<u>Question 5:</u>

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: $\frac{40}{680} = 0.0588$
2. The probability of false negative of the test is: $\frac{25}{975} = 0.0256$
3. The sensitivity value of the test is: $\frac{950}{975} = 0.9744$
4. The specificity value of the test is: $\frac{640}{680} = 0.9412$

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: $\frac{10}{925} = 0.0108$
2. The probability of false negative of the test is: $\frac{25}{875} = 0.0286$
3. The sensitivity value of the test is: $\frac{850}{875} = 0.9714$
4. The specificity value of the test is: $\frac{915}{925} = 0.9892$
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 $P(C)$*:**

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

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

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$\Rightarrow \quad 0.9 = 0.5 \quad + \quad 0.4 - P(A \cap B)$$
$$P(A \cap B) = 0$$

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

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

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

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

Question 2:

Givens:

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

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

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

<u>Question 3:</u>
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*** $\boldsymbol{P(C \mid A \cap B)}$:

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

<u>Question 4:</u>
Givens:

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

(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: $P(F) \neq P(F \mid S)$)
(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

Exercise3:

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 = $P(T^+ \mid D^+) = 0.92$
specificity = $P(T^- \mid D^-) = 0.94$
$P(D) = 0.05$

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

2. The predictive value negative is:
(A) 0.7841
(B) 0.9955 ($P(D^- \mid T^-)$ = 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		
		(Present: D^+) Infected Patients	(Absent: D^-) Non-infected People	Total
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^-)?

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

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^+)?

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

2. After-Test Questions:

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

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

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

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

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)

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

<u>2. After-Test Questions:</u>
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)

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

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

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

3. Efficiency of the Test:

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

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

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

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

$$= 0.9847$$

Chapter 4 Probability Distribution

Random Variables

- $0 \leq P(X = x) \leq 1$
- $\sum P(X = x) = 1$
- $E(X) \quad = \mu \quad = \sum x\, P(X = x)$
- $Var(X) = \sigma^2 \ = \sum (X - \mu)^2\, P(X = x)$
  $= E(X^2) - E(X)^2$

Question 1:

Given the following discrete probability distribution:

x	5	6	7	8
f(x)=P(X=x)	0.35	0.45	0.15	k

Find:

1. The value of k.

$$0.35 + 0.45 + 0.15 + k = 1 \implies \boxed{k = 0.05}$$

x	5	6	7	8
f(x)=P(X=x)	0.35	0.45	0.15	0.05

2. $P(X > 6) = 0.05 + 0.15 = 0.20$
3. $P(X \geq 6) = 0.45 + 0.15 + 0.05 = 0.65 \quad or \quad (1 - 0.35 = 0.65)$
4. $P(X < 4) = 0$
5. $P(X > 3) = 1$

Question 2:

Which of the following functions can be a probability distribution of a discrete random variable?

(a)

x	g(x)
0	0.6
1	-0.2
2	0.5
3	0.1
✘	

(b)

x	g(x)
0	0.4
1	0.1
2	0.5
3	0.2
✘	

(c)

x	g(x)
0	0.1
1	1.2
2	-0.6
3	0.3
✘	

(d)

x	g(x)
0	0.3
1	0.1
2	0.5
3	0.1
✓	

(e)

x	g(x)
0	0.2
1	0.4
2	0.3
3	0.4
✘	

(f)

x	g(x)
0	0.1
1	0.2
2	0.3
3	0.1
✘	

<u>Question 3</u>:
Which of the following is a probability distribution function:

$a. f(x) = \frac{x+1}{10}$; $x = 0,1,2,3,4$

$b. f(x) = \frac{x-1}{5}$; $x = 0,1,2,3,4$

$c. f(x) = \frac{1}{5}$; $x = 0,1,2,3,4$

$d. f(x) = \frac{5-x^2}{6}$; $x = 0,1,2,3$

a.

$$f(x) = \frac{x+1}{10};\ \ x = 0,1,2,3,4$$

x	0	1	2	3	4
$f(x)$	$^1/_{10}$	$^2/_{10}$	$^3/_{10}$	$^4/_{10}$	$^5/_{10}$

$f(x)$ *is not a P.D.F because* $\sum f(x) \neq 1$

b.

$$f(x) = \frac{x-1}{5};\ \ x = 0,1,2,3,4$$

x	0	1	2	3	4
$f(x)$	$^{-1}/_5$				

$f(x)$ *is not a P.D.F because every* $f(x)$ *shoud be* $0 \leq f(x) \leq 1$

c.

$$f(x) = \frac{1}{5};\ \ x = 0,1,2,3,4$$

x	0	1	2	3	4
$f(x)$	$^1/_5$	$^1/_5$	$^1/_5$	$^1/_5$	$^1/_5$

$f(x)$ *is a P.D.F*

d.

$$f(x) = \frac{5-x^2}{6};\ \ x = 0,1,2,3$$

x	0	1	2	3
$f(x)$				$^{-4}/_6$

$f(x)$ *is not a P.D.F because every* $f(x)$ *shoud be* $0 \leq f(x) \leq 1$

Question 4:

Given the following discrete probability distribution:

x	5	6	7	8
f(x)=P(X=x)	2k	3k	4k	k

Find the value of k.

$$2k + 3k + 4k + k = 1$$

$$10k = 1 \implies \boxed{k = 0.1}$$

x	5	6	7	8
f(x)=P(X=x)	0.2	0.3	0.4	0.1

Question 5:

Let X be a discrete random variable with probability mass function:
$f(x) = cx$; $x = 1,2,3,4$ What is the value of c?

x	1	2	3	4
$P(X = x)$	c	2c	3c	4c

$$c + 2c + 3c + 4c = 1 \quad \implies c = \frac{1}{10}$$

Then probability mass function:

x	1	2	3	4
$P(X = x)$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{4}{10}$

Question 6:

Let X be a discrete random variable with probability function given by:

$$f(x) = c(x^2 + 2) \quad ; \quad x = 0,1,2,3$$

$f(0) = c(0^2 + 2) = 2c$

$f(1) = c(1^2 + 2) = 3c$

$f(2) = c(2^2 + 2) = 6c$

$f(3) = c(3^2 + 2) = 11c$

x	0	1	2	3
f(x)	2c	3c	6c	11c

$$2c + 3c + 6c + 11c = 1 \qquad c = \frac{1}{22} = 0.04545$$

x	0	1	2	3
f(x)	$\frac{2}{22}$	$\frac{3}{22}$	$\frac{6}{22}$	$\frac{11}{22}$

Question 7:

Given the following discrete probability distribution:

x	5	6	7	8
f(x)=P(X=x)	0.2	0.4	0.3	0.1

Find:

1. Find the mean of the distribution $\mu = \mu_X = E(X)$.

$$E(X) = \mu = \mu_X = \sum_{x=5}^{8} x\, P(X = x)$$
$$= (5)(0.2) + (6)(0.4) + (7)(0.3) + (8)(0.1) = 6.3$$

2. Find the variance of the distribution $\sigma^2 = \sigma_X^2 = Var(X)$.

$$E(X^2) = (5^2 \times 0.2) + (6^2 \times 0.4) + (7^2 \times 0.3) + (8^2 \times 0.1) = 40.5$$

$$Var(X) = E(X^2) - E(X)^2$$
$$= 40.5 - 6.3^2 = 0.81$$

Or:

$$Var(X) = \sigma^2 = \sigma_X^2 = \sum (x - \mu)^2\, P(X = x)$$
$$= \sum_{x=5}^{8} (x - 6.3)^2\, P(X = x)$$

$$= (5 - 6.3)^2(0.2) + (6 - 6.3)^2(0.4) + (7 - 6.3)^2(0.3) + (8 - 6.3)^2(0.1) = 0.81$$

Question 8:

Given the following discrete distribution:

x	-1	0	1	2	3	4
P(X=x)	0.15	0.30	M	0.15	0.10	0.10

1. The value of M is equal to

$$M = 1 - (0.15 + 0.30 + 0.15 + 0.10 + 0.10) = 1 - 0.80 = 0.20$$

2. P (X ≤0.5) = $0.15 + 0.30 = 0.45$
3. P(X=0) = 0.30
4. The expected (mean) value E[X] is equal to

$$E(X) = (-1 \times 0.15) + (0 \times 0.30) + (1 \times 0.20) + (2 \times 0.15) + (3 \times 0.10) + (4 \times 0.10) = 1.05$$

<u>Question 9:</u>

The average length of stay in a hospital is useful for planning purposes. Suppose that the following is the probability distribution of the length of stay (X) in a hospital after a minor operation:

Length of stay (days)	3	4	5	6
Probability	0.4	0.2	0.1	k

(1) The value of k is

$$\boxed{k = 1 - (0.4 + 0.2 + 0.1) = 1 - 0.7 = 0.3}$$

(2) $P(X \leq 0) =$

$$\boxed{0}$$

(3) $P(0 < X \leq 5) =$

$$\boxed{0.4 + 0.2 + 0.1 = 0.7}$$

(4) $P(X \leq 5.5) =$

$$\boxed{0.4 + 0.2 + 0.1 = 0.7}$$

(5) The probability that the patient will stay at most 4 days in a hospital after a minor operation is equal to

$$\boxed{0.4 + 0.2 = 0.6}$$

(6) The average length of stay in a hospital is

$$\boxed{E(X) = (3 \times 0.4) + (4 \times 0.2) + (5 \times 0.1) + (6 \times 0.3) = 4.3}$$

<u>Question 10</u>:

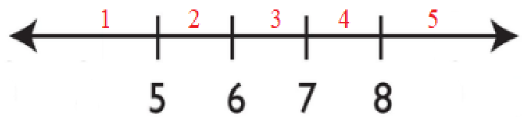
Given the following discrete probability distribution:

x	5	6	7	8
f(x)=P(X=x)	0.2	0.4	0.3	0.1

1. Find the cumulative distribution of X.

x	5	6	7	8
$F(x) = P(X \le x)$	0.2	0.6	0.9	1

$$F(x) = \begin{cases} 0 & X < 5 \\ 0.2 & 5 \le X < 6 \\ 0.6 & 6 \le X < 7 \\ 0.9 & 7 \le X < 8 \\ 1 & X \ge 8 \end{cases}$$



2. From the cumulative distribution of X, find:

a) $P(X \le 7) \quad = 0.9$

b) $P(X \le 6.5) = P(X \le 6)$ =0.6

c) $P(X > 6) \quad = 1 - P(X \le 6) = 1 - 0.6 = 0.4$

d) $P(X > 7) \quad = 1 - P(X \le 7) = 1 - 0.9 = 0.1$

<u>Question 11</u>:

Given that the cumulative distribution of random variable T, is:

$$F(t) = P(T \le t) = \begin{cases} 0 & t < 1 \\ 1/2 & 1 \le t < 3 \\ 8/12 & 3 \le t < 5 \\ 3/4 & 5 \le t < 7 \\ 1 & t \ge 7 \end{cases}$$

1. Find $P(T = 5)$

T	1	3	5	7
f(t)	$\frac{1}{2} - 0 = \mathbf{0.5}$	$\frac{8}{12} - \frac{1}{2} = \mathbf{0.167}$	$\frac{3}{4} - \frac{8}{12} = \mathbf{0.083}$	$1 - \frac{3}{4} = \mathbf{0.25}$

$$P(T = 5) = 0.083$$

2. Find $P(1.4 < T < 6) = 0.167 + 0.083 = 0.25$

Binomial Distribution:

$$P(X = x) = \binom{n}{x} p^x q^{n-x} \;\; ; \quad x = 0,1 \ldots.,n$$

$$* \, E(X) = np \qquad * \, Var(X) = npq$$

$$q = 1 - p$$

Question 1:

Suppose that 25% of the people in a certain large population have high blood pressure. A Sample of 7 people is selected at random from this population. Let X be the number of people in the sample who have high blood pressure, follows a binomial distribution then

1) The values of the parameters of the distribution are*:*

$$\boxed{p = 0.25 \quad , \quad n = 7}$$

A	7 , 0.75	B	7 , 0.25	C	0.25 , 0.75	D	25 , 7

2) The probability that we find exactly one person with high blood pressure, is:

X	0	1	2	3	4	5	6	7
$P(X=x)$		*						

$$\boxed{P(X = 1) = \binom{7}{1} (0.25)^1 (0.75)^6 = 0.31146}$$

3) The probability that there will be at most one person with high blood pressure, is:

X	0	1	2	3	4	5	6	7
$P(X=x)$	*	*						

$$\boxed{P(X \le 1) = \binom{7}{0} (0.25)^0 (0.75)^7 + \binom{7}{1} (0.25)^1 (0.75)^6 = 0.4449}$$

4) The probability that we find more than one person with high blood pressure, is:

X	0	1	2	3	4	5	6	7
$P(X=x)$			*	*	*	*	*	*

$$\boxed{P(X > 1) = 1 - P(X \le 1) = 1 - 0.4449 = 0.5551}$$

<u>Question 2:</u>

In some population it was found that the percentage of adults who have hypertension is 24 percent. Suppose we select a simple random sample of <u>five</u> adults from this population. Then the probability that the number of people who have hypertension in this sample, will be:

$$p = 0.24 \quad , \quad n = 5$$

1. Zero:

$$P(X = 0) = \binom{5}{0}(0.24)^0(0.76)^5 = 0.2536$$

2. Exactly one

$$P(X = 1) = \binom{5}{1}(0.24)^1(0.76)^4 = 0.4003$$

3. Between one and three, inclusive

$$P(1 \le X \le 3) = \binom{5}{1}(0.24)^1(0.76)^4 + \binom{5}{2}(0.24)^2(0.76)^3 + \binom{5}{3}(0.24)^3(0.76)^2 = 0.7330$$

4. Two or fewer (at most two):

$$P(X \le 2) = \binom{5}{0}(0.24)^0(0.76)^5 + \binom{5}{1}(0.24)^1(0.76)^4 + \binom{5}{2}(0.24)^2(0.76)^3 = 0.9067$$

5. Five:

$$P(X = 5) = \binom{5}{5}(0.24)^5(0.76)^0 = 0.0008$$

6. The mean of the number of people who have hypertension is equal to:

$$E(X) = np = 5 \times 0.24 = 1.2$$

7. The variance of the number of people who have hypertension is:

$$Var(X) = npq = 5 \times 0.24 \times 0.76 = 0.912$$

More Exercises

Exercise 1:

Find:

1. 6! $= 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$
2. ${}_8C_3 = \frac{8!}{3!\,(8-3)!} = \frac{8!}{3! \times 5!} = 56$
3. ${}_8C_{10} = 0$
4. ${}_8C_{-5} = 0$

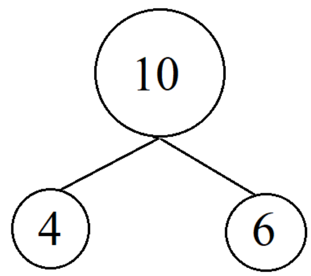
Exercise 2:

A box contains 10 cards numbered from 1 to 10. In how many ways can we select 4 cards out of this box?

Answer $= {}_{10}C_4 = \frac{10!}{4!\,(10-4)!} = \frac{10!}{4! \times 6!}$

$= \frac{10\times9\times8\times7\times6!}{(4\times3\times2\times1)\ 6!} = \frac{10\times9\times8\times7}{4\times3\times2\times1}$

$= 210$



Exercise 3:

The manager of a certain bank has recently examined the credit card account balances for the customers of his bank and found that 20% of the customers have excellent records. Suppose that the manager randomly selects a sample of 4 customers.

(A) Define the random variable X as:

X = The number of customers in the sample having excellent records.

Find the probability distribution of X.

$X \sim Binomial\ (n\ , p\ )$

$n = 4 \quad (Number\ of\ trials)$

$p = \frac{20}{100} = 0.2 \quad (Probability\ of\ success)$

$q = 1 - p = 1 - 0.2 = 0.8 \quad (Probability\ of\ failure)$

$x = 0, 1, 2, 3, 4 \quad (Possible\ values\ of\ X)$

(a) The probability function in a mathematical formula:

$$P(X=x)=\begin{cases}\dfrac{n!}{x!\,(n-x)!}\;p^x\,q^{n-x} & ;\quad x=0,1,2,\dots,n\\ 0 & ;Otherwise\end{cases}$$

$$P(X=x)=\begin{cases}\dfrac{4!}{x!\,(4-x)!}\;(0.2)^x\,(0.8)^{4-x} & ;\quad x=0,1,2,3,4\\ 0 & ;Otherwise\end{cases}$$

(b) The probability function in a table:

x	$P(X=x)$
0	$\dfrac{4!}{0!\,(4-0)!}\;(0.2)^0\,(0.8)^{4-0}=(1)(0.2)^0(0.8)^4=0.4096$
1	$\dfrac{4!}{1!\,(4-1)!}\;(0.2)^1\,(0.8)^{4-1}=(4)(0.2)^1(0.8)^3=0.4096$
2	$\dfrac{4!}{2!\,(4-2)!}\;(0.2)^2\,(0.8)^{4-2}=(6)(0.2)^2\,(0.8)^2=0.1536$
3	$\dfrac{4!}{3!\,(4-3)!}\;(0.2)^3\,(0.8)^{4-3}=(4)(0.2)^3\,(0.8)^1=0.0256$
4	$\dfrac{4!}{4!\,(4-4)!}\;(0.2)^4\,(0.8)^{4-4}=(1)(0.2)^4\,(0.8)^0=0.0016$
	Total = 1

x	$P(X=x)$
0	0.4096
1	0.4096
2	0.1536
3	0.0256
4	0.0016

(B) Find:

1. The probability that there will be 3 customers in the sample having excellent records.

$$P(X = 3) = 0.0256$$

2. The probability that there will be no customers in the sample having excellent records.

$$P(X = 0) = 0.4096$$

3. The probability that there will be at least 3 customers in the sample having excellent records.

$$P(X \geq 3) = P(x = 3) + P(X = 4) = 0.0256 + 0.0016$$
$$= 0.0272$$

4. The probability that there will be at most 2 customers in the sample having excellent records.

$$P(X \leq 2) = P(x = 0) + P(X = 1) + P(X = 2)$$
$$= 0.4096 + 0.4096 + 0.1536$$
$$= 0.9728$$

5. The expected number of customers having excellent records in the sample.

$$E(X) = \mu = \mu_X = np = 4 \times 0.2 = 0.8$$

6. The variance of the number of customers having excellent records in the sample.

$$Var(X) = \sigma^2 = \sigma_X^2 = npq = 4 \times 0.2 \times 0.8 = 0.64$$

<u>Exercise 4</u>: (Do it at home for yourself)

In a certain hospital, the medical records show that the percentage of lung cancer patients who smoke is 75%. Suppose that a doctor randomly selects a sample of 5 records of lung cancer patients from this hospital.

(A) Define the random variable X as:

X = The number of smokers in the sample.

Find the probability distribution of X.

(B) Find:

1. The probability that there will be 4 smokers in the sample.
2. The probability that there will be no smoker in the sample.
3. The probability that there will be at least 2 smokers in the sample.
4. The probability that there will be at most 3 smokers in the sample.
5. The expected number of smokers in the sample.
6. The variance of the number of smokers in the sample.

Poisson distribution:

$$P(X = x) = \frac{e^{-\lambda}\ \lambda^{x}}{x!} \quad ; \quad x = 0,1,2,\ldots$$

$$E(X) = Var(X) = \lambda$$

Question 1:

The number of serious cases coming to a hospital during a night follows a Poisson distribution with an average of 10 persons per night, then:

1) The probability that 12 serious cases coming in the next night, is:

$$\lambda_{one\ night} = 10$$

$$P(X = 12) = \frac{e^{-10}\ 10^{12}}{12!} = 0.09478$$

2) The average number of serious cases in a two nights' period is:

$$\lambda_{two\ nights} = 20$$

3) The probability that 20 serious cases coming in next two nights is:

$$\lambda_{two\ nights} = 20$$

$$P(X = 20) = \frac{e^{-20}\ 20^{20}}{20!} = 0.0888$$

<u>Question 2:</u>

Given the mean number of serious accidents per year in a large factory is five. If the number of accidents follows a Poisson distribution, then the probability that in the next year there will be:

1. Exactly seven accidents:

$$\lambda_{one\ year} = 5$$

$$P(X = 7) = \frac{e^{-5}\ 5^{7}}{7!} = 0.1044$$

2. No accidents

$$P(X = 0) = \frac{e^{-5}\ 5^{0}}{0!} = 0.0067$$

3. one or more accidents

X	0	1	2	3	4	5	6	...
$P(X = x)$		*	*	*	*	*	*	*

$$\begin{aligned} P(X \geq 1) &= 1 - P(X < 1) \\ &= 1 - P(X = 0) \\ &= 1 - 0.0067 = 0.9933 \end{aligned}$$

4. The expected number (mean) of serious accidents in the next two years is equal to

$$\lambda_{two\ years} = 10$$

5. The probability that in the next two years there will be three accidents

$$\lambda_{two\ years} = 10$$

$$P(X = 3) = \frac{e^{-10}\ 10^{3}}{3!} = 0.0076$$

More Exercise

Exercise 1:

Suppose that in a certain city, the weekly number of infected cases with Corona virus (COVID-19) has a Poisson distribution with an average (mean) of 5 cases per week.

(A) Find:

1. The probability distribution of the weekly number of infected cases (X).

$$P(X=x)=\begin{cases}\frac{e^{-\lambda}\,\lambda^{x}}{x!} & ;\quad x=0,1,2,3,\dots\\ 0 & ;Otherwise\end{cases}$$

$$\lambda=5$$

$$P(X=x)=\begin{cases}\frac{e^{-5}\,5^{x}}{x!} & ;\quad x=0,1,2,3,\dots\\ 0 & ;Otherwise\end{cases}$$

2. The probability that there will be 2 infected cases this week.

$$P(X=2)=\frac{e^{-5}\,5^{2}}{2!}=0.0842$$

3. The probability that there will be 1 infected case this week.

$$P(X=1)=\frac{e^{-5}\,5^{1}}{1!}=0.0337$$

4. The probability that there will be no infected cases this week.

$$P(X=0)=\frac{e^{-5}\,5^{0}}{0!}=0.0067$$

5. The probability that there will be at least 3 infected cases this week.

$$P(X\geq 3)=1-P(X<3)=1-P(X\leq 2)$$
$$=1-[P(X=0)+P(X=1)+P(X=2)]$$
$$=1-[\,0.0067+0.0337+0.0842\,]$$
$$=1-0.1246=0.8754$$

6. The probability that there will be at most 2 infected cases this week.

$$P(X\leq 2)=P(X=0)+P(X=1)+P(X=2)$$
$$=0.0067+0.0337+0.0842$$
$$=0.1246$$

7. The expected number (mean/average) of infected cases this week.

$$E(X) = \mu = \mu_X = \lambda = 5$$

8. The variance of the number of infected cases this week.

$$Var(X) = \sigma^2 = \sigma_X^2 = \lambda = 5$$

(B): Find:

1. The average (mean) of the number infected cases in a day.

$$\lambda = \frac{5}{7} = 0.7143$$

2. The probability distribution of the daily number of infected cases (X).

$$P(X = x) = \begin{cases} \dfrac{e^{-\lambda}\, \lambda^x}{x!} & ; \quad x = 0, 1, 2, 3, \dots \\ 0 & ; Otherwise \end{cases}$$

$$\lambda = \frac{5}{7}$$

$$P(X = x) = \begin{cases} \dfrac{e^{-\frac{5}{7}} \left(\frac{5}{7}\right)^x}{x!} & ; \quad x = 0, 1, 2, 3, \dots \\ 0 & ; Otherwise \end{cases}$$

3. The probability that there will be 2 infected cases tomorrow.

$$P(X = 2) = \frac{e^{-\frac{5}{7}} \left(\frac{5}{7}\right)^2}{2!} = 0.1249$$

4. The probability that there will be 1 infected case tomorrow.

$$P(X = 1) = \frac{e^{-\frac{5}{7}} \left(\frac{5}{7}\right)^1}{1!} = 0.3497$$

5. The probability that there will be no infected cases tomorrow.

$$P(X = 0) = \frac{e^{-\frac{5}{7}} \left(\frac{5}{7}\right)^0}{0!} = 0.4895$$

6. The probability that there will be at most 2 infected cases tomorrow.

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$
$$= 0.4895 + 0.3497 + 0.1249$$
$$= 0.9641$$

7. The probability that there will be at least 2 infected cases tomorrow.

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X \leq 1)$$
$$= 1 - [P(X = 0) + P(X = 1)]$$
$$= 1 - [0.4895 + 0.3497]$$
$$= 1 - 0.8392 = 0.1608$$

8. The expected number (mean/average) of infected cases tomorrow.

$$E(X) = \mu = \mu_X = \lambda = \frac{5}{7} = 0.7143$$

9. The variance of the number of infected cases tomorrow.

$$Var(X) = \sigma^2 = \sigma_X^2 = \lambda = \frac{5}{7} = 0.7143$$

(C): Assuming that 4 weeks are in a month, find:

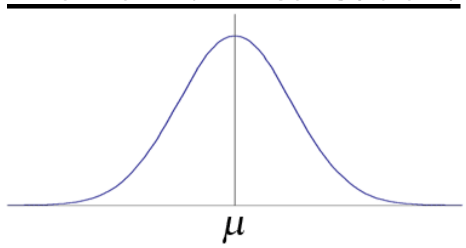
1. The average (mean) of the number infected cases per month.

$$E(X) = \mu = \mu_X = \lambda = 5 \times 4 = 20$$

2. The variance of the number of infected cases per month.

$$Var(X) = \sigma^2 = \sigma_X^2 = \lambda = 5 \times 4 = 20$$

The Normal Distribution:



Mean = Median = Mode

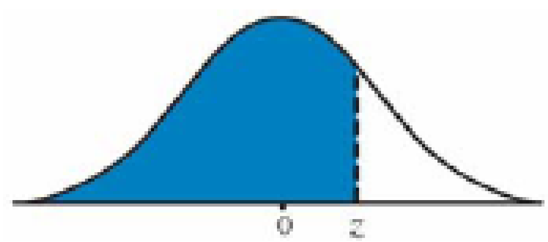
Normal distribution $X \sim N(\mu, \sigma^2)$
Standard normal $Z \sim N(0, 1)$

Question 1:

Given the standard normal distribution, Z~N (0, 1), find:

1. $P(Z < 1.43) = 0.92364$

z	0.00	0.01	0.02	0.03	0.04
0.00	0.50000	0.50399	0.50798	0.5[illegible]197	0.51595
0.10	0.53983	0.54380	0.54776	0.5[illegible]172	0.55567
0.20	0.57926	0.58317	0.58706	0.5[illegible]095	0.59483
0.30	0.61791	0.62172	0.62552	0.6[illegible]930	0.63307
0.40	0.65542	0.65910	0.66276	0.6[illegible]640	0.67003
0.50	0.69146	0.69497	0.69847	0.7[illegible]194	0.70540
0.60	0.72575	0.72907	0.73237	0.7[illegible]565	0.73891
0.70	0.75804	0.76115	0.76424	0.7[illegible]730	0.77035
0.80	0.78814	0.79103	0.79389	0.7[illegible]673	0.79955
0.90	0.81594	0.81859	0.82121	0.8[illegible]381	0.82639
1.00	0.84134	0.84375	0.84614	0.8[illegible]849	0.85083
1.10	0.86433	0.86650	0.86864	0.8[illegible]076	0.87286
1.20	0.88493	0.88686	0.88877	0.8[illegible]065	0.89251
1.30	0.90320	0.90490	0.90658	0.9[illegible]824	0.90988
1.40	[illegible]	[illegible]	[illegible]	0.92364	0.92507
1.50	0.93319	0.93448	0.93574	0.93699	0.93822
1.60	0.94520	0.94630	0.94738	0.94845	0.94950
1.70	0.95543	0.95637	0.95728	0.95818	0.95907
1.80	0.96407	0.96485	0.96562	0.96638	0.96712
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]



$P\big(Z <$ أطراف الجدول $\big) =$ داخل الجدول

2. $P(Z > 1.67) = 1 - P(Z < 1.67) = 1 - 0.95254 = 0.04746$

3. $P(-2.16 < Z < -0.65)$

$$= P(Z < -0.65) - P(Z < -2.16)$$
$$= 0.25785 - 0.01539 = 0.24246$$

Question 2:

Given the standard normal distribution, Z~N(0,1), find:

1. $P(Z > 2.71) = 1 - P(Z < 2.71) = 1 - 0.99664 = 0.00336$

2. $P(-1.96 < Z < 1.96)$

$$= P(Z < 1.96) - P(Z < -1.96)$$
$$= \quad 0.9750 \quad - \quad 0.0250 \quad = 0.9500$$

3. $P(Z = 1.33) = 0$
4. $P(Z = 0.67) = 0$

5. If $P(Z < a) = 0.99290$, then the value of $a = 2.45$
6. If $P(Z < a) = 0.62930$, then the value of $a = 0.33$

7. If $P(Z > a) = 0.63307 \Longrightarrow P(Z < a) = 1 - 0.63307$

$$\Longrightarrow P(Z < a) = 0.36693 \Longrightarrow a = -0.34$$

8. If $P(Z > a) = 0.02500 \Longrightarrow P(Z < a) = 1 - 0.02500$

$$\Longrightarrow P(Z < a) = 0.97500 \Longrightarrow a = 1.96$$

9. $Z_{0.9750} = 1.96$
10. $Z_{0.0392} = -1.76$
11. $Z_{0.01130} = -2.28$
12. $Z_{0.99940} = 3.24$
13. If $Z_{0.08} = -1.40$ then the value of $Z_{0.92}$ equals to:

A	-1.954	B	1	C	1.40	D	-1.40

14.If $P(-k < Z < k) = 0.8132$, then the value of k =

$$\Rightarrow 2 \times P(0 < Z < k) = 0.8132$$

$$\Rightarrow \quad P(0 < Z < k) = 0.4066$$

$$\Rightarrow P(Z < k) - P(Z < 0) = 0.4066$$

$$\Rightarrow P(Z < k) - \quad 0.5 \quad = 0.4066$$

$$\Rightarrow P(Z < k) = 0.9066$$

$$\Rightarrow k = 1.32$$

<u>Question 3:</u>

Given the standard normal distribution, then:

1) $P(-1.1 < Z < 1.1)$ *is:*

$$\Rightarrow P(Z < 1.1) - P(Z < -1.1)$$

$$0.86433 - 0.13567 = 0.72866$$

2) $P(Z > -0.15)$ *is:*

$$= 1 - P(Z < -0.15)$$

$$= 1 - 0.44038 = 0.55962$$

3) The k value that has an area of 0.883 to its right, is:

Left	*Right*
<	>

$$P(Z > k) = 0.883$$
$$P(Z < k) = 1 - 0.883$$
$$P(Z < k) = 0.117$$
$$k = -1.19$$

Question 4:

The finished inside diameter of a piston ring is normally distributed with a mean 12 cm and standard deviation of 0.03 cm. Then,

1. The proportion of rings that will have inside dimeter less than 12.05.

$$X \sim N(\mu \ , \sigma^2)$$
$$X \sim N(12 \ , 0.03^2)$$

$$P(X < 12.05) = P\left(Z < \frac{12.05 - \mu}{\sigma}\right)$$

$$= P\left(Z < \frac{12.05 - 12}{0.03}\right) = P(Z < 1.67) = 0.9525$$

2. The proportion of rings that will have inside dimeter exceeding 11.97.

$$P(X > 11.97) = P\left(Z > \frac{11.97 - \mu}{\sigma}\right)$$

$$= P\left(Z > \frac{11.97 - 12}{0.03}\right) = P(Z > -1)$$
$$= 1 - P(Z < -1)$$
$$= 1 - 0.1587 = 0.8413$$

3. The proportion of rings that will have inside dimeter between 11.95 and 12.05.

$$P(11.95 < X < 12.05) = P\left(\frac{11.95 - 12}{0.03} < Z < \frac{12.05 - 12}{0.03}\right)$$

$$= P(\ -1.67 \ < Z < \ 1.67)$$

$$= P(Z < 1.67) - P(Z < -1.67)$$

$$= \ 0.9525 \ - \ 0.0475 = 0.905$$

Question 5:

The weight of a large number of fat persons is nicely modeled with a normal distribution with mean of 128kg and a standard deviation of 9 kg

1. The probability of fat persons with weight at most 110 kg is:

$$X \sim N(\mu, \sigma^2)$$
$$X \sim N(128, 9^2)$$

$$P(X \leq 110) = P\left(Z < \frac{110 - 128}{9}\right) = P(Z < -2) = 0.0228$$

2. The probability of fat persons with weight more than 149 kg is:

$$P(X > 149) = P\left(Z > \frac{149 - 128}{9}\right) = 1 - P(Z < 2.33) = 1 - 0.9901 = 0.0099$$

3. The weight x above which 86% of those persons will be:

$$P(X > x) = 0.86 \Rightarrow P(X < x) = 0.14 \Rightarrow P\left(Z < \frac{x - 128}{9}\right) = 0.14$$

by searching inside the table for 0.14, and transforming X to Z, we got:

$$\frac{x - 128}{9} = -1.08$$

$$x - 128 = -1.08 \times 9$$

$$x = (-1.08 \times 9) + 128$$

$$x = 118.28$$

4. The weight x below which 50% of those persons will be:

$P(X < x) = 0.5$, *by searching inside the table for 0.5, and transforming X to Z*

$$\frac{x - 128}{9} = 0 \Rightarrow x = 128$$

Question 6:

If the random variable X has a normal distribution with the mean μ and the variance σ^2, then $P(X < \mu + 2\sigma)$ equal to:

$$P(X < \mu + 2\sigma) = P\left(Z < \frac{(\mu + 2\sigma) - \mu}{\sigma}\right) = P(Z < 2) = 0.9772$$

Question 7:

If the random variable X has a normal distribution with the mean μ and the variance 1, and if then $P(X < 3) = 0.877$ then μ equal to

$$\text{Given that } \sigma = 1$$

$$P(X < 3) = 0.877 \Rightarrow P\left(Z < \frac{3-\mu}{1}\right) = 0.877$$

$$3 - \mu = 1.16 \Rightarrow \mu = 1.84$$

Question 8:

Suppose that the marks of students in a certain course are distributed according to a normal distribution with the mean 70 and the variance 25. If it is known that 33% of the student failed the exam, then the passing mark is:

$$X \sim N(70, 25)$$

$$P(X < x) = 0.33 \Rightarrow P\left(Z < \frac{x - 70}{5}\right) = 0.33$$

by searching inside the table for 0.33, and transforming X to Z, we got:

$$\frac{x - 70}{5} = -0.44 \Rightarrow x = 67.8$$

Question 9:

What k value corresponds to 17% of the area between the mean and the z value?
$P(\mu < Z < k) = 0.17$

$P(\mu < Z < k) = 0.17$
$P(Z < k) - P(Z < \mu) = 0.17$
$P(Z < k) - \quad 0.5 \quad = 0.17$
$P(Z < k) = 0.67$
$k = 0.44$

<u>Question 10:</u>

A nurse supervisor has found that staff nurses complete a certain task in 10 minutes on average. If the times required to complete the task are approximately normally distributed with a standard deviation of 3 minutes, then:

1) The probability that a nurse will complete the task in less than 8 minutes is:

$$X \sim N(10, 3^2)$$

$$P(X < 8) = P\left(Z < \frac{8-10}{3}\right) = P(Z < -0.67) = 0.2514$$

2) The probability that a nurse will complete the task in more than 4 minutes is:

$$P(X > 4) = 1 - P\left(Z < \frac{4-10}{3}\right) = 1 - P(Z < -2) = 1 - 0.0228 = 0.9772$$

3) If eight nurses were assigned the task, the expected number of them who will complete it within 8 minutes is approximately equal to:

$$n \times P(0 < X < 8) = 8 \times P\left(\frac{0-10}{3} < Z < \frac{8-10}{3}\right)$$
$$= 8 \times P(-3.33 < Z < -0.67)$$
$$= 8 \times [P(Z < -0.67) - P(Z < -3.33)]$$
$$= 8 \times [0.2514 - 0.0004] = 2$$

4) If a certain nurse completes the task within k minutes with probability 0.6293; then k equals approximately:

$$P(0 < X < k) = 0.6293$$
$$\Rightarrow P\left(\frac{0-10}{3} < Z < \frac{k-10}{3}\right) = 0.6293$$
$$\Rightarrow P\left(-3.33 < Z < \frac{k-10}{3}\right) = 0.6293$$
$$\Rightarrow P\left(Z < \frac{k-10}{3}\right) - P(Z < -3.33) = 0.6293$$
$$\Rightarrow P\left(Z < \frac{k-10}{3}\right) - 0.0004 = 0.6293$$
$$\Rightarrow P\left(Z < \frac{k-10}{3}\right) = 0.6297$$
$$\Rightarrow \frac{k-10}{3} = 0.33 \Rightarrow k = 11$$

Question 11:

Given the normally distributed random variable X with mean 491 and standard deviation 119,

1. If P (X < k) = 0.9082, the value of k is equal to

A	649.27	B	390.58	C	128.90	D	132.65

2. If P (292< X <M) = 0.8607, the value of M is equal to

A	766	B	649	C	108	D	136

Question 12:

The IQ (Intelligent Quotient) of individuals admitted to a state school for the mentally retarded are approximately normally distributed with a mean of 60 and a standard deviation of 10, then:

1) The probability that an individual picked at random will have an IQ greater than 75 is:

A	0.9332	B	0.8691	C	0.7286	D	0.0668

2) The probability that an individual picked at random will have an IQ between 55 and 75 is:

A	0.3085	B	0.6915	C	0.6247	D	0.9332

3) If the probability that an individual picked at random will have an IQ less than k is 0.1587. Then the value of k

A	50	B	45	C	51	D	40

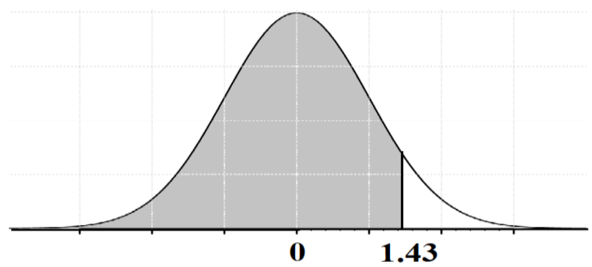
More Exercises:

Exercise 1:

Suppose that the random variable Z has a standard normal distribution

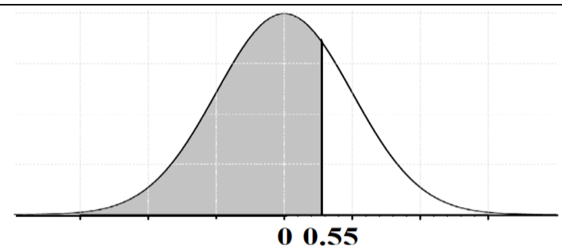
(a) Find the area to the left of Z =1.43.

$P(Z < 1.43) = 0.92364$



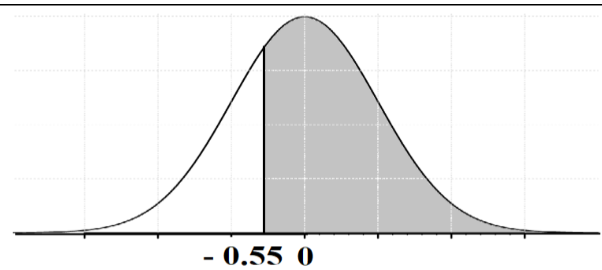
(b) Find P (Z < 0.55).

$P(Z < 0.55) = 0.70884$



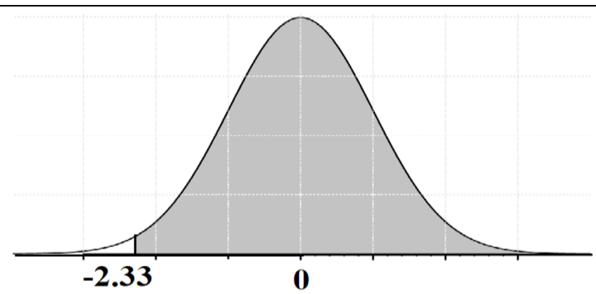
(c) Find P (Z > - 0.55).

$P(Z > -0.55)$
$=1-P(Z<-0.55)$
$=1-0.29116$
$= 0.70884$



(d) Find P (Z > - 2.33).

$P(Z > -2.33)$
$= 1-P(Z<-2.33) = 1- 0.00990$
$= 0.9901$

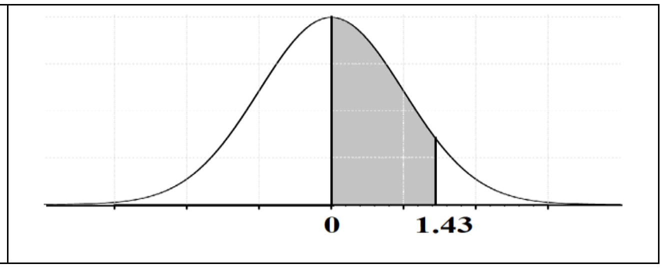


(e) Find the area to the right of z = 2.

$P(Z > 2.00) = 1 - P(Z < 2.00)$
$= 1 - 0.97725$
$= 0.02275$

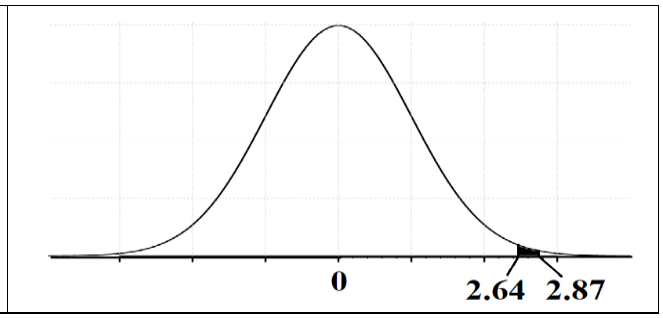
(f) Find the area under the curve between z = 0 and z = 1.43.

$P(0 < Z < 1.43)$
$= P(Z < 1.43) – P(Z < 0.00)$
$= 0.92364 - 0.5$
$= 0.42364$



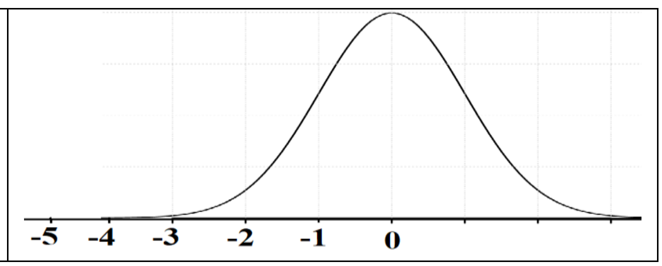
(g) Find the probability that Z will take a value between z = 2.64 and z = 2.87.

$P(2.64 < Z < 2.87)$
$= P(Z < 2.87) – P(Z < 2.64)$
$= 0.99795 - 0.99585$
$= 0.0021$



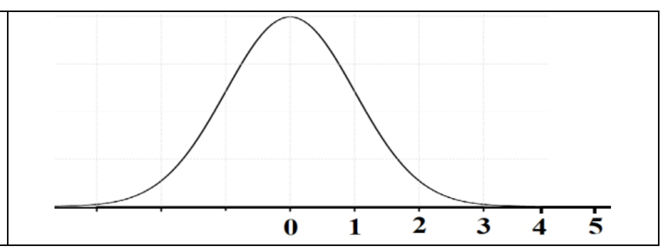
(h) Find P(Z < - 5).

$P(Z < -5) \approx 0$



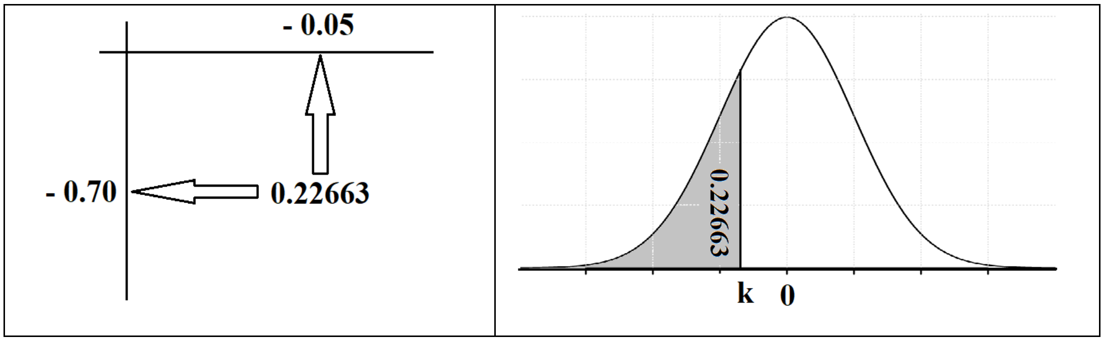
(i) Find P(Z > 5).

$P(Z > 5) = 1 - P(Z < 5)$
$\approx 1 - 1$
$= 0$



(j) If $P(Z \leq k) = 0.22663$, then find the value of k.

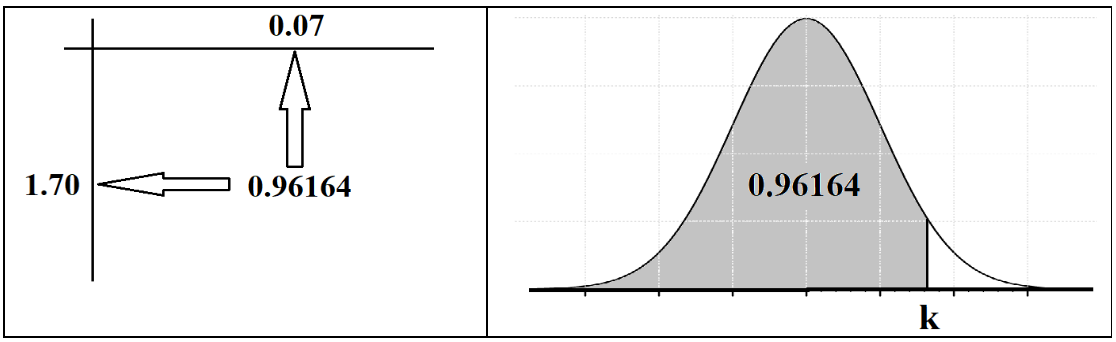
$k = -0.75$



(k) If $P(Z \geq k) = 0.03836$, then find the value of k.

$P(Z<k)=1- P(Z \geq k) =1-0.03836 = 0.96164$

$k = 1.77$



l) If $P(-2.67 < Z \leq k) = 0.97179$, then find the value of k.

$0.97179 = P(-2.67 < Z \leq k)$

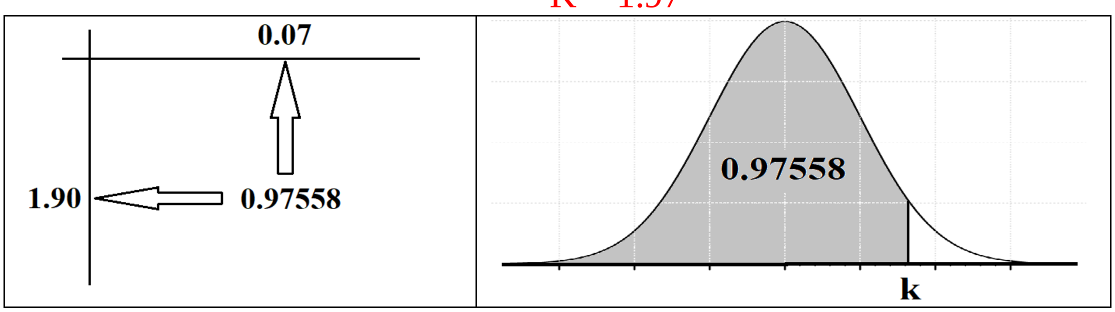
$= P(Z<K) – P(Z < -2.67)$

$P(Z<K) = 0.97179 + P(Z < -2.67)$

$P(Z<K) =0.97179 + 0.00379$

$= 0.97558$

$K = 1.97$



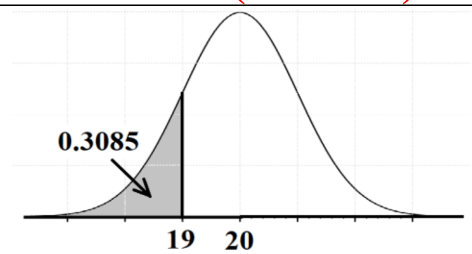
Exercise 2:

Suppose that the time for a person to be tested for corona virus (in minutes) has a normal distribution with mean $\mu = 20$ and variance $\sigma^2 = 4$.

(1) If we select a person at random, what is the probability that his examination period will be less than 19 minutes?

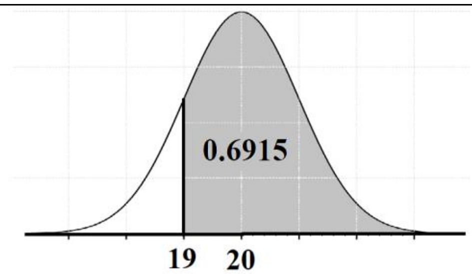
Let X = person's period of examination (in minutes)

$$P(X < 19) = P\left(\frac{X-\mu}{\sigma} < \frac{19-\mu}{\sigma}\right) = P\left(Z < \frac{19-20}{2}\right) = P(Z < -0.5) = 0.3085$$



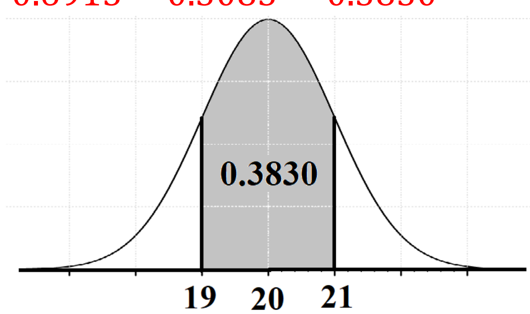
(2) If we select a person at random, what is the probability that his examination period will be more than 19 minutes?

$$P(X > 19) = 1 - P(X < 19) = 1 - 0.3085 = 0.6915$$



(3) If we select a person at random, what is the probability that his examination period will be between 19 and 21 minutes?

$$P(19 < X < 21) = P(X < 21) - P(X < 19)$$
$$= P\left(\frac{X-\mu}{\sigma} < \frac{21-\mu}{\sigma}\right) - P\left(\frac{X-\mu}{\sigma} < \frac{19-\mu}{\sigma}\right)$$
$$= P\left(Z < \frac{21-20}{2}\right) - P\left(Z < \frac{19-20}{2}\right)$$
$$= P(Z < 0.5) - P(Z < -0.5)$$
$$= 0.6915 - 0.3085 = 0.3830$$



(4) What is the percentage of persons whose examination period are less than 19 minutes?

$$\% = P(X < 19) * 100\% = 0.3085 * 100\%$$
$$= 30.85\%$$

(5) If we select a sample of 2000 persons, how many persons would be expected to have examination periods that are less than 19 minutes?

$$\text{Expected number} = 2000 \times P(X < 19)$$
$$= 2000 \times 0.3085$$
$$= 617$$

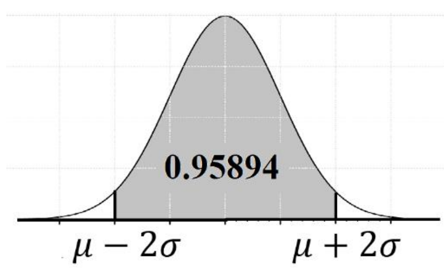
Exercise 3:

Suppose that we have a normal population with mean μ and standard deviation σ.

(1) Find the percentage of values which are between $\mu - 2\sigma$ and $\mu + 2\sigma$.

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = P(X < \mu + 2\sigma) - P(X < \mu - 2\sigma)$$

$$= P\left(\frac{X-\mu}{\sigma} < \frac{(\mu+2\sigma)-\mu}{\sigma}\right) - P\left(\frac{X-\mu}{\sigma} < \frac{(\mu-2\sigma)-\mu}{\sigma}\right)$$
$$= P\left(Z < \frac{(\mu+2\sigma)-\mu}{\sigma}\right) - P\left(Z < \frac{(\mu-2\sigma)-\mu}{\sigma}\right)$$
$$= P\left(Z < \frac{2\sigma}{\sigma}\right) - P\left(Z < \frac{-2\sigma}{\sigma}\right)$$
$$= P(Z < 2.00) - P(Z < -2.00)$$
$$= 0.97725 - 0.02275$$
$$= 0.9545$$



(2) Find the percentage of values which are between $\mu - \sigma$ and $\mu + \sigma$.

Dot it yourself

(3) Find the percentage of values which are between $\mu - 3\sigma$ and $\mu + 3\sigma$.

Dot it yourself

Exercise 4: (Read it yourself)
In a study of fingerprints, an important quantitative characteristic is the total ridge count for the 10 fingers of an individual. Suppose that the total ridge counts of individuals in a certain population are approximately normally distributed with a mean of 140 and a standard deviation of 50. Then:
(1) The probability that an individual picked at random from this population will have a ridge count of 200 or more is:

$$\begin{aligned} P(X > 200) &= 1 - P(X < 200) \\ &= 1 - P\left(Z < \tfrac{200-\mu}{\sigma}\right) \\ &= 1 - P\left(Z < \tfrac{200-140}{50}\right) \\ &= 1 - P(Z < 1.2) \\ &= 1 - 0.88493 = 0.11507. \end{aligned}$$

(2) The probability that an individual picked at random from this population will have a ridge count of less than 100 is:

$$\begin{aligned} P(X < 100) &= P\left(Z < \frac{100-\mu}{\sigma}\right) \\ &= P\left(Z < \tfrac{100-140}{50}\right) \\ &= P(Z < -0.80) = 0.18673 \end{aligned}$$

(3) The probability that an individual picked at random from this population will have a ridge count between 100 and 200 is:

$$\begin{aligned} P(100 < X < 200) &= P(X < 200) - P(X < 100) \\ &= P(X < 200) - P(X < 100) \\ &= P\left(Z < \tfrac{200-140}{50}\right) - \left(Z < \tfrac{100-140}{50}\right) \\ &= P(Z < 1.20) - (Z < -0.80) \\ &= 0.88493 - 0.18673 = 0.6982 \end{aligned}$$

(4) The percentage of individuals whose ridge counts are between 100 and 200 is:

$$\begin{aligned} P(100 < X < 200) * 100\% &= 0.6982 * 100\% \\ &= 69.82\% \end{aligned}$$

(4) If we select a sample of 5,000 individuals from this population, how many individuals would be expected to have ridge counts that are between 100 and 200?

$$\begin{aligned} \text{Expected number} &= 5000 \times P(100 < X < 200) \\ &= 5000 \times 0.6982 = 3491 \end{aligned}$$

Chapter 5 Sampling Distribution

Sampling Distribution

Single Mean	Two Means
$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ $E(\bar{X}) = \bar{X} = \mu$ $Var(\bar{X}) = \frac{\sigma^2}{n}$	$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2 , \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$ $E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2$ $Var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$
Single Proportion	**Two Proportions**
For large sample size ($n \geq 30,\ np > 5,\ nq > 5$) $\hat{p} \sim N\left(p, \frac{pq}{n}\right)$ $E(\hat{p}) = p$ $Var(\hat{p}) = \frac{pq}{n}$	For large sample size ($n_1 \geq 30,\ n_1p_1 > 5,\ n_1q_1 > 5$) ($n_2 \geq 30,\ n_2p_2 > 5,\ n_2q_2 > 5$) $\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}\right)$ $E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2$ $Var(\hat{p}_1 - \hat{p}_2) = \frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}$

	population normal or not normal n large ($n \geq 30$)		population normal n small ($n < 30$)	
	σ known	σ unknown	σ known	σ unknown
Sampling Distribution	$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	$Z = \frac{\bar{X} - \mu}{s/\sqrt{n}}$	$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$

Question 1:

The average life of a certain battery is 5 years, with a standard deviation of 1 year. Assume that the live of the battery approximately follows a normal distribution.

1. The sample mean $\bar{X}$ of a random sample of 5 batteries selected from this product has mean $E(\bar{X}) = \mu_{\bar{X}}$.

$$\boxed{\mu = 5 \;\; ; \;\; \sigma = 1 \;\; ; \;\; n = 5}$$

$$E(\bar{X}) = \mu = 5$$

2. The variance $Var(\bar{X}) = \sigma_{\bar{X}}^2$ of the sample mean $\bar{X}$ of a random sample of 5 batteries selected from this product is equal to:

$$Var(\bar{X}) = \frac{\sigma^2}{n} = \frac{1}{5} = 0.2$$

3. The probability that the average life of a random sample of size 16 of such batteries will be between 4.5 and 5.4.

$$n = 16 \rightarrow \frac{\sigma}{\sqrt{n}} = \frac{1}{4}$$

$$P(4.5 < \bar{X} < 5.4) = P\left(\frac{4.5-\mu}{\frac{\sigma}{\sqrt{n}}} < Z < \frac{5.4-\mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$= P\left(\frac{4.5-5}{\frac{1}{4}} < Z < \frac{5.4-5}{\frac{1}{4}}\right) = \quad P(-2 < Z < 1.6)$$

$$= P(Z < 1.6) - P(Z < -2)$$

$$= \quad 0.9452 \;\; - 0.0228 \quad = 0.9224$$

4. The probability that the average life of a random sample of size 16 of such batteries will be less than 5.5 years is:

$$P(\bar{X} < 5.5) = P\left(Z < \frac{5.5-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(Z < \frac{5.5-5}{1/4}\right) = P(Z < 2) = 0.9772$$

5. The probability that the average life of a random sample of size 16 of such batteries will be more than 4.75 years is:

$$P(\bar{X} > 4.75) = P\left(Z > \frac{4.75-\mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$= P\left(Z > \frac{4.75-5}{\frac{1}{4}}\right) = \; P(Z > -1)$$

$$= 1 - P(Z < -1) = 1 - 0.1587 = 0.841$$

6. If $P(\bar{X} > a) = 0.1492$ where $\bar{X}$ represent the sample mean for a random sample of size 9 of such batteries, then the numerical value of a is:

$P(\bar{X} > a) = 0.1492 \quad ; \quad n = 9$

$$P\left(Z > \frac{a-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = 0.1492$$

$$\Rightarrow P\left(Z < \frac{a-5}{\frac{1}{3}}\right) = 1 - 0.1492$$

$$\Rightarrow P\left(Z < \frac{a-5}{\frac{1}{3}}\right) = 0.8508$$

$$\frac{a-5}{\frac{1}{3}} = 1.04 \Longrightarrow a = 5.347$$

Question 2:

Suppose that you take a random sample of size $n = 64$ from a distribution with mean $\mu = 55$ and standard deviation $\sigma = 10$. Let $\bar{X} = \frac{1}{n}\sum x$ be the sample mean.

1. What is the approximate sampling distribution of $\bar{X}$.

$$\boxed{\mu = 55 \quad ; \quad \sigma = 10 \quad ; \quad n = 64}$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) = \bar{X} \sim N\left(55, \frac{100}{64}\right)$$

2. What is the mean of $\bar{X}$?

$$E(\bar{X}) = \mu = 55$$

3. What is the standard error (standard deviation) of $\bar{X}$?

$$S.D(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{64}} = \frac{10}{8}$$

4. Find the probability that the sample mean $\bar{X}$ exceeds 52.

$$P(\bar{X} > 52) = P\left(Z > \frac{52-55}{\frac{10}{8}}\right) = P(Z > -2.4)$$

$$= 1 - P(Z < -2.4)$$

$$= 1 - 0.0082 = 0.9918$$

Question 3:

Suppose that the hemoglobin levels (in g/dl) of healthy Saudi females are approximately normally distributed with mean of 13.5 and a standard deviation of 0.7. If 15 healthy adult Saudi female is randomly chosen, then:

1. The mean of $\overline{X}$ ($E(\overline{X})$ or $\mu_{\overline{X}}$)

A	0.7	B	13.5	C	15	D	3.48

2. The standard error of $\overline{X}$ ($\sigma_{\overline{X}}$)

A	0.181	B	0.0327	C	0.7	D	13.5

3. $P(\overline{X} < 14) =$

A	0.99720	B	0.99440	C	0.76115	D	0.9971

4. $P(\overline{X} > 13.5) =$

A	0.99	B	0.50	C	0.761	D	0.622

5. $P(13 < \overline{X} < 14) =$

A	0.9972	B	0.9944	C	0.7615	D	0.5231

Question 4:

If the uric acid value in normal adult males is approximately normally distributed with a mean and standard derivation of 5.7 and 1 mg percent, respectively, find the probability that a sample of size 9 will yield a mean

1. Greater than 6 is:

A	0.2109	B	0.1841	C	0.8001	D	0.8159

2. At most 5.2 is:

A	0.6915	B	0.9331	C	0.8251	D	0.0668

3. Between 5 and 6 is:

A	0.1662	B	0.7981	C	0.8791	D	0.9812

Question 5:

Medical research has concluded that people experience a common cold roughly two times per year. Assume that the time between colds is normally distributed with a mean 165 days and a standard deviation of 45 days. Consider the sampling distribution of the sample mean based on samples of size 36 drown from the population:

1. The mean of sampling distribution $\bar{X}$ is:

A	210	B	36	C	45	D	165

2. The distribution if the mean of $\bar{X}$ is:

A	$N(165,2025)$	B	$N(165,45)$	C	$T, with\ df = 30$	D	$N(165,7.5)$

3. $\mathrm{P}(\bar{\mathrm{X}} > 178) =$

A	0.0415	B	0.615	C	0.958	D	0.386

Sampling Distribution: Two Means:

$$*\ \bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2\ , \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$*\ E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2 \qquad *\ Var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Question 6:

A random sample of size $n_1 = 36$ is taken from normal population with a mean $\mu_1 = 70$ and a standard deviation $\sigma_1 = 4$. A second independent random sample of size $n_2 = 49$ is taken from a normal population with a mean $\mu_2 = 85$ and a standard deviation $\sigma_2 = 5$. Let $\bar{X}_1$ and $\bar{X}_2$ be the average of the first and second sample, respectively.

1. Find $E(\bar{X}_1 - \bar{X}_2)$ and $Var(\bar{X}_1 - \bar{X}_2)$.

$$n_1 = 36\ , \mu_1 = 70\ , \sigma_1 = 4$$
$$n_2 = 49\ , \mu_2 = 85\ , \sigma_2 = 5$$

$$E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2 = 70 - 85 = -15$$

$$Var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{16}{36} + \frac{25}{49} = 0.955$$

2. Find $P(\bar{X}_1 - \bar{X}_2 > -16)$.

$$P(\bar{X}_1 - \bar{X}_2 > -16) = P\left(Z > \frac{-16-(-15)}{\sqrt{0.955}}\right) = 1 - P\left(Z < \frac{-16-(-15)}{\sqrt{0.955}}\right)$$

$$= 1 - P(Z < -1.02) = 0.8461$$

Question 7:

A random sample of size 25 is taken from a normal population (1st population) having a mean of 100 and a standard of 6. A second random sample of size 36 is taken from a different normal population (2nd population) having a mean of 97 and a standard deviation of 5. Assume that these two samples are independent.

1. The probability that the sample mean of the first sample will exceed the sample mean of the second sample by at least 6 is:

$$n_1 = 25\,,\ \mu_1 = 100\,, \sigma_1 = 6$$
$$n_2 = 36\,,\ \mu_2 = 97\ \ , \sigma_2 = 5$$

$$E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2 = 100 - 97 = 3 \qquad Var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{36}{25} + \frac{25}{36} = 2.134$$

$$P(\bar{X}_1 > \bar{X}_2 + 6) = P(\bar{X}_1 - \bar{X}_2 > 6)$$

$$= P\left(Z > \frac{6-(3)}{\sqrt{2.134}}\right) = P(Z > 2.05)$$
$$= 1 - P(Z < 2.05)$$
$$= 1 - 0.9798 = 0.0202$$

2. The probability that the difference between the two-sample means will be less than 2 is:

$$P(\bar{X}_1 - \bar{X}_2 < 2) = P\left(Z < \frac{2-(3)}{\sqrt{2.134}}\right)$$

$$= P(Z < -0.68) = 0.2483$$

Question 8:

Given two normally distributed population with equal means and variances $\sigma_1^2 = 100$, $\sigma_2^2 = 350$. Two random samples of size $n_1 = 40$, $n_2 = 35$ are drown and sample means $\bar{X}_1$ and $\bar{X}_2$ are calculated, respectively, then

$\mu_1 - \mu_2 = 0$

1. $P(\bar{X}_1 - \bar{X}_2 > 12)$ is

A	0.1499	B	0.8501	C	0.9997	D	0.0003

2. $P(5 < \bar{X}_1 - \bar{X}_2 < 12)$ is

A	0.0789	B	0.9217	C	0.8002	D	None of these

Sampling Distribution: Single Proportion:

For large sample size ($n \geq 30,\ np > 5,\ nq > 5$)

$$* \hat{p} \sim N\left(p, \frac{pq}{n}\right)$$

$$* E(\hat{p}) = p \qquad * Var(\hat{p}) = \frac{pq}{n}$$

Question 9:

Suppose that 10% of the students in a certain university smoke cigarette. A random sample of 30 student is taken from this university. Let $\hat{p}$ be the proportion of smokers in the sample.

1. Find $E(\hat{p}) = \mu_{\hat{p}}$ the mean of $\hat{p}$.

$$p = 0.1 \quad ; \quad n = 30 \quad ; \quad q = 1 - p = 0.9$$

$$E(\hat{p}) = p = 0.1$$

2. Find $Var(\hat{p}) = \sigma^2_{\hat{p}}$ the variance of $\hat{p}$.

$$Var(\hat{p}) = \frac{pq}{n} = \frac{0.1 \times 0.9}{30} = 0.003$$

3. Find an approximate distribution of $\hat{p}$

$$\hat{p} \sim N(0.1, 0.003)$$

4. Find $P(\hat{p} > 0.25)$.

$$P(\hat{p} > 0.25) = P\left(Z > \frac{0.25 - 0.1}{\sqrt{0.003}}\right) = P(Z > 2.74)$$

$$= 1 - P(Z < 2.74) = 1 - 0.99693 = 0.00307$$

Question 10:

Suppose that 15% of the patients visiting a certain clinic are females. If A random sample of 35 patients was selected, $\hat{p}$ represent the proportion of females in the sample. then find:

1. The expected value of $(\hat{p})$ is:

A	0.35	B	0.85	C	0.15	D	0.80

2. The standard deviation of $(\hat{p})$ is:

A	0.3214	B	0.0036	C	0.1275	D	0.0604

3. The approximate sampling distribution of $(\hat{p})$ is:

A	N(0.15,0.0604)	B	Binomial(0.15,35)	C	$N(0.15, 0.0604^2)$	D	$Binomial(0.15, 35^2)$

4. The $P(\hat{p} > 0.15)$ is:

A	0.35478	B	0.5	C	0.96242	D	0.46588

Question 11:

In a study, it was found that 31% of the adult population in a certain city has a diabetic disease. 100 people are randomly sampled from the population. Then

1. The mean for the sample proportion $\left(E(\hat{p}) \text{ or } \mu_{\hat{p}}\right)$ is:

A	0.40	B	0.31	C	0.69	D	0.10

2. $P(\hat{p} > 0.40) =$

A	0.02619	B	0.02442	C	0.0256	D	0.7054

Sampling Distribution: Two Proportions:

For large sample size $(n_1 \geq 30,\ n_1 p_1 > 5,\ n_1 q_1 > 5)$
$(n_2 \geq 30,\ n_2 p_2 > 5,\ n_2 q_2 > 5)$

$$* \hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right)$$

$$* E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2 \qquad * Var(\hat{p}_1 - \hat{p}_2) = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

Question 12:

Suppose that 25% of the male student and 20% of the female student in certain university smoke cigarettes. A random sample of 35 male students is taken. Another random sample of 30 female student is independently taken from this university. Let $\hat{p}_1$ and $\hat{p}_2$ be the proportions of smokers in the two sample, respectively.

1. Find $E(\hat{p}_1 - \hat{p}_2) = \mu_{\hat{p}_1 - \hat{p}_2}$, the mean of $\hat{p}_1 - \hat{p}_2$.

$$p_1 = 0.25 \ ; \quad n_1 = 35$$
$$p_2 = 0.20 \ ; \quad n_2 = 30$$

$$E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2 = 0.25 - 0.20 = 0.05$$

2. Find $Var(\hat{p}_1 - \hat{p}_2) = \sigma^2_{\hat{p}_1 - \hat{p}_2}$, the variance of $\hat{p}_1 - \hat{p}_2$.

$$Var(\hat{p}_1 - \hat{p}_2) = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2} = \frac{0.25 \times 0.75}{35} + \frac{0.2 \times 0.8}{30} = 0.01069$$

3. Find an approximate distribution of $\hat{p}_1 - \hat{p}_2$.

$$\hat{p}_1 - \hat{p}_2 \sim N(0.05, 0.01069)$$

4. Find $P(0.1 < \hat{p}_1 - \hat{p}_2 < 0.2)$

$$P(0.1 < \hat{p}_1 - \hat{p}_2 < 0.2) = \left(\frac{0.1 - 0.05}{\sqrt{0.01069}} < Z < \frac{0.2 - 0.05}{\sqrt{0.01069}}\right)$$

$$= \quad (0.48 < Z < 1.45)$$
$$= P(Z < 1.45) - P(Z < 0.48)$$
$$= 0.92647 - 0.68439 = 0.24208$$

Question 13:

Suppose that 7 % of the pieces from a production process A are defective while that proportion of defective for another production process B is 5 %. A random sample of size 400 pieces is taken from the production process A while the sample size taken from the production process B is 300 pieces. If $\hat{p}_1$ and $\hat{p}_2$ be the proportions of defective pieces in the two samples, respectively, then:

1. The sampling distribution of $(\hat{p}_1 - \hat{p}_2)$ is:

A	N(0,1)	B	Normal	C	T	D	Unknown

2. The value of the standard error of the difference $(\hat{p}_1 - \hat{p}_2)$ is:

A	0.02	B	0.10	C	0	D	0.22

Question 13:

In a study to make an inference between the proportion of houses heated by gas in city A and city B, the following information was collected:

	Proportion of houses heated by gas	Sample size
City A	43%	90
City B	51%	150

Suppose p_A proportion of city A houses which are heated by gas, p_B proportion of city B houses which are heated by gas. The two sample are independent.

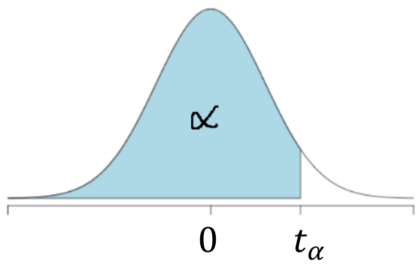
1. The sampling distribution for the sample proportion of city B which are heated by gas is:

A	$\hat{p}_B \sim N\left(p_B, \frac{p_B q_B}{n_B}\right)$	B	$\hat{p}_B \sim N\left(\hat{p}_B, \frac{\hat{p}_B \hat{q}_B}{n_B}\right)$	C	$\hat{p}_B \sim N(\hat{p}_B, \hat{p}_B \hat{q}_B)$	D	$\hat{p}_B \sim N(p_B, p_B q_B)$

2. The sampling distribution of $\hat{p}_A$ is (approximately) normal if:

A	$n_A \geq 30$ $n_A p_A > 5$	B	$n_A \geq 30$ $n_A p_A > 5$ $n_A q_A > 5$	C	$n_A p_A > 5$	D	$\frac{p_A}{n_A} > 5$

The student (t) Distribution:



يجب ان تكون اشارة الاحتمال أقل من (<) قبل البحث في جدول (t) :

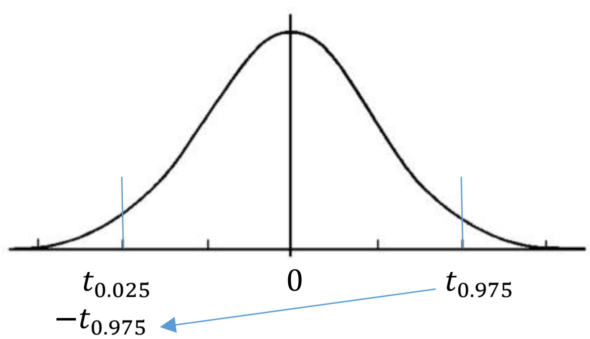
$$t_{\alpha,v} \Longrightarrow P(T < t_{\alpha,v})$$
$$\Longrightarrow P(T < -t_{1-\alpha,v}) \qquad ; v = n - 1$$

- $\mathit{If}\ P(T < t_{0.99,22}) \Longrightarrow t_{0.99,22} = 2.508$

- $\mathit{If}\ P(T > t_{0.975,18}) \Longrightarrow P(T < t_{0.025,18})$

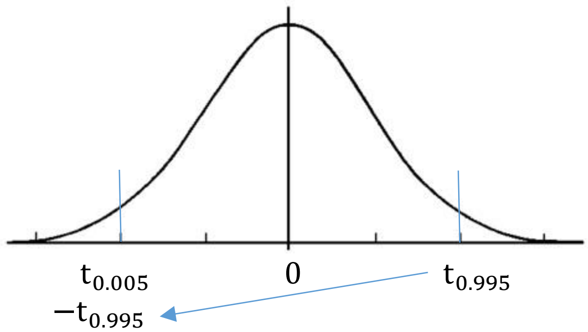
$$\Longrightarrow P(T < -t_{0.975,18})$$

$$\Longrightarrow -t_{0.975,18} = -2.101$$



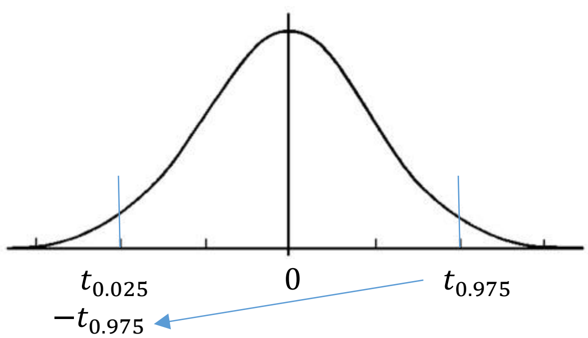
- If $P(T > t_{\alpha,v})$ where $v = 24, \alpha = 0.995$

$$\Rightarrow P(T > t_{0.995,24})$$
$$\Rightarrow P(T < t_{0.005,24})$$
$$\Rightarrow P(T < -t_{0.995,24})$$
$$\Rightarrow -t_{0.995,24} = -2.797$$



- If $P(T > t_{\alpha,v})$ where $v = 7, \alpha = 0.975$

$$\Rightarrow P(T > t_{0.975\ 7})$$
$$\Rightarrow P(T < t_{0.025\ 7})$$
$$\Rightarrow P(T < -t_{0.975,7})$$
$$\Rightarrow -t_{0.975,7} = -2.365$$



Question 1:

Let T follow the t distribution with 9 degrees of freedom, then
The probability $(T < 1.833)$ equal to:

بما ان الاشارة اقل من (<) اذن ننظر للجدول

ν=df	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055

$\alpha = 0.95$

- $The\ probabitity\ P(T < -1.833)\ equal\ to:$

ν=df	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055

$\alpha = 1 - 0.95 = 0.05$

<u>Question 2:</u>

Let T follow the t distribution with 9 degrees of freedom, then
The t-value that leaves an area of 0.10 to the right is:

$$P\left(T > t_{0.10,9}\right)$$
$$P\left(T < t_{0.90,9}\right) \implies t_{0.90,9} = 1.383$$

ν=df	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055

<u>Question 3:</u>

Given the t-distribution with 12 degrees of freedom, then
The t-value that leaves an area of 0.025 to the left is:

$$P\left(T < t_{0.025,12}\right)$$
$$P\left(T < -t_{0.975,12}\right)$$

ν=df	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055

$$-t_{0.975,12} = -2.179$$

Question 4:

Consider the student t distribution:

Find the t-value with $\boldsymbol{n = 17}$ the leaves an area of 0.01 to the left:

$\text{df} = \; n - 1$

$= 17 - 1 = 16$

$P(T < \; t_{0.01,16})$

$P(T < -t_{0.99,16})$

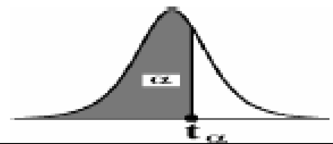
v=df	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861

$t_{0.99,16} = \; 2.583$

$-t_{0.99,16} = -2.583$

A	-2.58	B	2.567	C	2.58	D	-2.567

Critical Values of the t-distribution (t_α)



v=df	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
35	1.3062	1.6896	2.0301	2.4377	2.7238
40	1.3030	1.6840	2.0210	2.4230	2.7040
45	1.3006	1.6794	2.0141	2.4121	2.6896
50	1.2987	1.6759	2.0086	2.4033	2.6778
60	1.2958	1.6706	2.0003	2.3901	2.6603
70	1.2938	1.6669	1.9944	2.3808	2.6479
80	1.2922	1.6641	1.9901	2.3739	2.6387

- **Question from previous midterms and finals:**

<u>Question:</u>

Given two normally distributed populations with a mean $\mu_1 = 10$ and $\mu_2 = 20$, and variances of $\sigma_1^2 = 100$ and $\sigma_2^2 = 80$. If two samples are taken from the populations of size $n_1 = 25$ and $n_1 = 16$ are taken from the populations.
Let $\bar{X}_1$ and $\bar{X}_2$ be the average of the first and second sample, respectively.

$$n_1 = 25\,, \mu_1 = 10\,, \sigma_1^2 = 100$$
$$n_2 = 16\,, \mu_2 = 20\,, \sigma_2^2 = 80$$

1. **Find the sampling distribution for $\bar{X}_1$.**

$$\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$$

$$\bar{X}_1 \sim N\left(10, \frac{100}{25}\right)$$

$$\bar{X}_1 \sim N(10, \ 4)$$

2. **Find the sampling distribution for $(\bar{X}_1 - \bar{X}_2)$.**

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2\ , \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$\bar{X}_1 - \bar{X}_2 \sim N\left(10 - 20\ , \frac{100}{25} + \frac{80}{16}\right)$$

$$\bar{X}_1 - \bar{X}_2 \sim N(-10 \quad , \ 4 \quad + 5)$$

$$\bar{X}_1 - \bar{X}_2 \sim N(-10, 9)$$

Chapter 6 Estimation and Confidence Interval

Estimation and Confidence Interval

Single Mean	Two Means
$\bar{X} \pm Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ σ known	$(\bar{X}_1 - \bar{X}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ $\sigma_1$and σ_2 known
$\bar{X} \pm t_{1-\frac{\alpha}{2},(n-1)} \frac{S}{\sqrt{n}}$ σ unknown	$(\bar{X}_1 - \bar{X}_2) \pm t_{1-\frac{\alpha}{2},\ (n_1+n_2-2)}\ S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ σ_1 and σ_2 unknown
Single Proportion	**Two Proportions**
For large sample size $(n \geq 30,\ np > 5,\ nq > 5)$ $\hat{p} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}$	For large sample size $(n_1 \geq 30,\ n_1p_1 > 5,\ n_1q_1 > 5)$ $(n_2 \geq 30,\ n_2p_2 > 5,\ n_2q_2 > 5)$ $(\hat{p}_1 - \hat{p}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$

$$S_p^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

$$\bar{X} \ \pm \underbrace{\left(\overbrace{Z_{1-\frac{\alpha}{2}}}^{\text{Reliability coefficeint}} \quad \frac{\sigma}{\sqrt{n}} \right)}_{\text{margin of error} \atop {or \atop \text{precision of the estimate}}}$$

Question 1:

Suppose we are interested in making some statistical inference about the mean μ, of a normal population with standard deviation $\sigma = 2$. Suppose that a random sample of size $n = 49$ from this population gave a sample mean $\bar{X} = 4.5$.

a. Find the upper limit of 95% of the confident interval for μ

$\sigma = 2 \quad \bar{X} = 4.5 \quad n = 49$

$95\% \rightarrow \alpha = 0.05 \qquad Z_{1-\frac{\alpha}{2}} = Z_{0.975} = 1.96$

$$\bar{X} + \left(Z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}\right) = 4.5 + \left(1.96 \times \frac{2}{7}\right) = 5.06$$

b. Find the lower limit of 95% of the confident interval for μ

$$\bar{X} - \left(Z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}\right) = 4.5 - \left(1.96 \times \frac{2}{7}\right) = 3.94$$

Question 2:

A researcher wants to estimate the mean of a life span a certain bulb. Suppose that the distribution is normal with standard deviation 5 hours. Suppose that the researcher selected a random sample of 49 bulbs and found that the sample mean is 390 hours.

$\sigma = 5 \ , \ \bar{X} = 390 \ , \ n = 49$

a. find $Z_{0.975}$:

$Z_{0.975} = 1.96$

b. find a point estimate for μ

$E(\bar{X}) = \hat{\mu} = \bar{X} = 390$

c. Find the upper limit of 95% of the confident interval for μ

$$\bar{X} + \left(Z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}\right) = 390 + \left(1.96 \times \frac{5}{\sqrt{49}}\right) = 391.4$$

d. Find the lower limit of 95% of the confident interval for μ

$$\bar{X} - \left(Z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}\right) = 390 - \left(1.96 \times \frac{5}{\sqrt{49}}\right) = 388.6$$

Question 3:

A sample of 16 college students were asked about time they spent doing their homework. It was found that the average to be 4.5 hours. Assuming normal population with standard deviation 0.5 hours.

$\sigma = 0.5 \quad \bar{X} = 4.5 \quad n = 16$

1. The point estimate for μ is:

A	0 hours	B	10 hours	C	0.5 hours	D	4.5 hours

2. The standard error of $\bar{X}$ is:

$$S.E(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{0.5}{\sqrt{16}}$$

A	0.125 hours	B	0.266 hours	C	0.206 hours	D	0.245hours

3. The correct formula for calculating $100\,(1-\alpha)\%$ confidence interval for μ is:

A	$\bar{X} \pm t_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$	B	$\bar{X} \pm Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$	C	$\bar{X} \pm Z_{1-\frac{\alpha}{2}} \frac{\sigma^2}{n}$	D	$\bar{X} \pm t_{1-\frac{\alpha}{2}} \frac{\sigma^2}{n}$

4. The upper limit of 95% confidence interval for μ is:

A	4.745	B	4.531	C	4.832	D	4.891

5. The lower limit of 95% confidence interval for μ is:

A	5.531	B	7.469	C	3.632	D	4.255

6. The length of the 95% confidence interval for μ is:

$$\text{Length} = 4.745 - 4.255 = 0.49$$

A	4.74	B	0.49	C	0.83	D	0.89

Question 4:

Let us consider a hypothetical study on the height of women in their adulthood. A sample of 24 women is drown from a normal distribution with population mean μ and variance σ^2. The sample mean and variance of height of the selected women are 151 cm and 18.65 cm^2 respectively. Using given data, we want to constract a 99% confidentce interval for the mean height of the adult women in the populatopn from which the sample was drown randomly.

$$\overline{X} = 151 \quad ; \quad n = 24 \quad ; \quad S^2 = 18.65 \Longrightarrow S = 4.32$$

a. Point estimate for μ

$$\hat{\mu} = \overline{X} = 151$$

b. Find the upper limit of 99% of the confident interval for μ

$$\overline{X} + \left(t_{1-\frac{\alpha}{2},n-1} \times \frac{S}{\sqrt{n}}\right)$$

$$= 151 + \left(2.807 \times \frac{4.32}{\sqrt{24}}\right) = 153.4753$$

$$99\% \rightarrow \alpha = 0.01$$

$$t_{1-\frac{\alpha}{2},n-1} = t_{1-\frac{0.01}{2},24-1}$$

$$= t_{0.995,23} = 2.807$$

c. Find the lower limit of 99% of the confident interval for μ

$$\overline{X} - \left(t_{1-\frac{\alpha}{2},n-1} \times \frac{S}{\sqrt{n}}\right)$$

$$= 151 - \left(2.807 \times \frac{4.32}{\sqrt{24}}\right) = 148.5247$$

Estimation and Confidence Interval: Two Means

1- $(\bar{X}_1 - \bar{X}_2) \pm \left(Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$

2- $(\bar{X}_1 - \bar{X}_2) \pm \left(t_{1-\frac{\alpha}{2}, n_1+n_2-2}\, Sp \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$

$$S_p^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

Question 5:

The tensile strength of type I thread is approximately normally distributed with standard deviation of 6.8 kg. A sample of 20 pieces of the thread has an average tensile strength of 72.8 kg. Another type of thread (type II) is approximately followed normal distribution with standard deviation 6.8 kg. A sample of 25 pieces of the thread has an average tensile strength pf 64.4 kg. then for 98% confidence interval of the difference in tensile strength means between type I and type II, we have:

$$\text{Theard 1}: \; n_1 = 20, \bar{X}_1 = 72.8, \sigma_1 = 6.8$$
$$\text{Thread 2}: \; n_2 = 25, \bar{X}_2 = 64.4, \sigma_2 = 6.8$$

$$98\% \rightarrow \alpha = 0.02 \rightarrow \quad Z_{1-\frac{\alpha}{2}} = Z_{0.99} = 2.325$$

$$(\bar{X}_1 - \bar{X}_2) \pm \left(Z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$
$$(72.8 - 64.4) \pm \left(2.325 \times \sqrt{\frac{6.8^2}{20} + \frac{6.8^2}{25}}\right)$$

$$(3.657, 13.143)$$

(1): The lower limit $= 3.657$
(2): The upper limit $= 13.143$

Question 6:

	First sample	Second sample
Sample size (n)	12	14
Sample mean $(\overline{X})$	10.5	10
Sample variance (S^2)	4	5

1. Estimate the difference $\mu_1 - \mu_2$:

$$\hat{\mu}_1 - \hat{\mu}_2 = \overline{X}_1 - \overline{X}_2 = 10.5 - 10 = 0.5$$

2. Find the pooled standard deviation estimator Sp :

$$S_p^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2} = \frac{4(11) + 5(13)}{24} = 4.54 \Longrightarrow \boxed{Sp = 2.13}$$

3. The upper limit of 95% confidence interval for $(\mu_1 - \mu_2)$ is:

$$95\% \rightarrow \alpha = 0.05 \rightarrow t_{1-\frac{\alpha}{2}, n_1+n_2-2} = t_{0.975, 24} = 2.064,$$

$$(\overline{X}_1 - \overline{X}_2) + \left(t_{1-\frac{\alpha}{2}, n_1+n_2-2} \times Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

$$(0.5) + \left(2.064 \times 2.13\sqrt{\frac{1}{12} + \frac{1}{14}}\right) = 2.23$$

4. The lower limit of 95% confidence interval for $(\mu_1 - \mu_2)$ is:

$$(\overline{X}_1 - \overline{X}_2) - \left(t_{1-\frac{\alpha}{2}, n_1+n_2-2} \times Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

$$(0.5) - \left(2.064 \times 2.13\sqrt{\frac{1}{12} + \frac{1}{14}}\right) = -1.23$$

Question 7:

A researcher was interested in comparing the mean score of female students μ_1, with the mean score of male students μ_2 in a certain test. Assume the populations of score are normal with equal variances. Two independent samples gave the following results:

	Female	Male
Sample size	$n_1 = 5$	$n_2 = 7$
Mean	$\overline{X}_1 = 82.63$	$\overline{X}_2 = 80.04$
Variance	$S_1^2 = 15.05$	$S_2^2 = 20.79$

1. The point estimate of $\mu_1 - \mu_2$ is:

A	2.63	B	-2.37	C	2.59	D	0.59

2. The estimate of the pooled variance S_p^2 is:

A	17.994	B	18.494	C	17.794	D	18.094

3. The upper limit of the 95% confidence interval for $\mu_1 - \mu_2$ is :

A	26.717	B	7.525	C	7.153	D	8.2

4. The lower limit of the 95% confidence interval for $\mu_1 - \mu_2$ is :

A	-21.54	B	-2.345	C	-3.02	D	-1.973

Estimation and Confidence Interval: Single Proportion

For large sample size ($n \geq 30,\ np > 5,\ nq > 5$)

$$* \text{ Point estimate for } P \text{ is: } \frac{x}{n}$$

$$* \text{ Interval estimate for } P \text{ is: } \hat{p} \pm \left(Z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\hat{p}\hat{q}}{n}} \right)$$

Question 7:

A random sample of 200 students from a certain school showed that 15 student smoke. Let p be the proportion of smokers in the school.

1. Find a point estimate for p.

$$n = 200 \ \ \& \ \ x = 15$$

$$\hat{p} = \frac{x}{n} = \frac{15}{200} = 0.075 \rightarrow \hat{q} = 0.925$$

2. Find 95% confidence interval for p.

$$95\% \rightarrow \alpha = 0.05 \quad \rightarrow \quad Z_{1-\frac{\alpha}{2}} = Z_{0.975} = 1.96$$

$$\hat{p} \pm \left(Z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\hat{p}\hat{q}}{n}} \right) = 0.075 \pm \left(1.96 \times \sqrt{\frac{0.075 \times 0.925}{200}} \right)$$

The 95% confidence interval is: $(0.038,\ 0.112)$

Question 8:

A researcher's group has perfected a new treatment of a disease which they claim is very efficient. As evidence, they say that they have used the new treatment on 50 patients with the disease and cured 25 of them. To calculate a 95% confidence interval for the proportion of the cured.

1. The point estimate of p is equal to:

A	0.25	B	0.50	C	0.01	D	0.33

2. The reliability coefficient $\left(z_{1-\frac{\alpha}{2}}\right)$ is equal is:

A	1.96	B	1.645	C	2.02	D	1.35

3. The 95% confidence interval is equal to:

A	(0.1114,0.3886)	B	(0.3837,0.6163)	C	(0.1614,0.6386)	D	(0.3614,0.6386)

Estimation and Confidence Interval: Two Proportions

For large sample size $(n_1 \geq 30,\ n_1 p_1 > 5,\ n_1 q_1 > 5)$
$(n_2 \geq 30,\ n_2 p_2 > 5,\ n_2 q_2 > 5)$

$$* \text{ Point estimate for } P_1 - P_2 = \hat{p}_1 - \hat{p}_2 = \frac{x_1}{n_1} - \frac{x_2}{n_2}$$

$$* \text{ Interval estimate for } P_1 - P_2 \text{ is: } (\hat{p}_1 - \hat{p}_2) \pm \left(Z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} \right)$$

Question 9:

A random sample of 100 students from school "A" showed that 15 students smoke. Another independent random sample of 200students from school "B" showed that 20 students smoke. Let p_1 be the proportion of smoker in school "A" and let p_2 be the proportion of smoker in school "B".

1. Find a point estimate for $P_1 - P_2$.

$$n_1 = 100, x_1 = 15 \rightarrow \hat{p}_1 = \frac{15}{100} = \boxed{0.15} \Longrightarrow \hat{q}_1 = 1 - 0.15 = \boxed{0.85}$$

$$n_2 = 200, x_2 = 20 \rightarrow \hat{p}_2 = \frac{20}{200} = \boxed{0.10} \Longrightarrow \hat{q}_2 = 1 - 0.10 = \boxed{0.90}$$

$$\boxed{\hat{p}_1 - \hat{p}_2 = 0.15 - 0.1 = 0.05}$$

2. Find 95% confidence interval for $P_1 - P_2$.

$$95\% \rightarrow \alpha = 0.05 \rightarrow Z_{1-\frac{\alpha}{2}} = Z_{0.975} = 1.96$$

$$(\hat{p}_1 - \hat{p}_2) \pm \left(Z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} \right)$$

$$= (0.05) \pm \left(1.96 \times \sqrt{\frac{(0.15)(0.85)}{100} + \frac{(0.1)(0.9)}{200}} \right)$$

$$= 0.05 \pm \left(1.96 \times \sqrt{0.001725} \right)$$

The 95% confidence interval is: $(-0.031,\ 0.131)$

Question 10:

a first sample of 100 store customers, 43 used a MasterCard. In a second sample of 100 store customers, 58 used a Visa card. To find the 95% confidence interval for difference in the proportion $(P_1 - P_2)$ of people who use each type of credit card?

1. The value of α is:

A	0.95	B	0.50	C	0.05	D	0.025

2. The upper limit of 95% confidence interval for the proportion difference is:

$$n_1 = 100\ ,\ x_1 = 43 \ \rightarrow\ \hat{p}_1 = \frac{43}{100} = \boxed{0.43} \ \Longrightarrow \hat{q}_1 = 1 - 0.43 = \boxed{0.57}$$
$$n_2 = 100\ ,\ x_2 = 58 \ \rightarrow\ \hat{p}_2 = \frac{58}{100} = \boxed{0.58} \ \Longrightarrow \hat{q}_2 = 1 - 0.58 = \boxed{0.42}$$

$$(\hat{p}_1 - \hat{p}_2) + \left(Z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}\right)$$

$$= (0.43 - 0.58) + \left(1.96 \times \sqrt{\frac{(0.43)(0.57)}{100} + \frac{(0.58)(0.42)}{100}}\right) = -0.013$$

3. The lower limit of 95% confidence interval for the proportion difference is:

$$(\hat{p}_1 - \hat{p}_2) - \left(Z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}\right)$$

$$= (0.43 - 0.58) - \left(1.96 \times \sqrt{\frac{(0.43)(0.57)}{100} + \frac{(0.58)(0.42)}{100}}\right) = -0.287$$

Question from previous midterms and finals:

- In procedure of construction $(1-\alpha)100\%$ confidence interval for the population mean (μ) of a normal population with a known standard deviation (σ) based on a random sample of size n.

1. The width of $(1-\alpha)100\%$ confidence interval for (μ) is:

A	$2\,Z_{1-\alpha}\frac{\sigma^2}{n}$	B	$2\,Z_{1-\alpha}\frac{\sigma}{\sqrt{n}}$	C	$2\,Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$	D	$2\,Z_{1-\alpha}\frac{\sigma^2}{\sqrt{n}}$

2. For $n=70$ and $\sigma=4$ the width of a 95% confidence interval for (μ) is:

A	3.1458	B	1.5153	C	6.1601	D	1.8741

3. For $\overline{X}=60$ and a 95% confidence interval for μ is $(57\,,\ k)$, then the value of the upper confidence limit k is:

A	64.5	B	66	C	61.5	D	63

4. When comparing the width of the 95% confidence interval (C.I.) for μ with that of 90% C.I., we found that:

	95% C.I. is shorter	B	95% C.I. is wider	C	They have the same width	D	We can't decide

5. When the sample size n increase, the width of the C.I. will:

A	Decrease	B	Increase	C	Not be change	D	We can't decide

6. The most typical form of a calculated confidence interval is:

A	Point estimate $\pm$ standard error
B	Population parameter $\pm$ margin of error
C	Population parameter $\pm$ standard error
D	Point estimate $\pm$ margin of error

7. Confidence intervals are useful when trying to estimate .... parameter:

A	Sample	B	Statistics	C	Population	D	None of these

8. The following C.I. is obtained for a population proportion $(0.505, 0.545)$, then the margin of error equals (let $\hat{p}=0.525$)

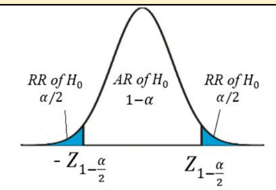
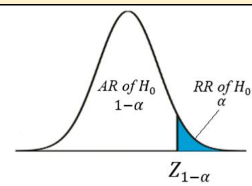
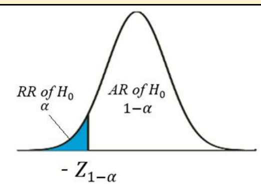
A	0.01	B	0.04	C	0.03	D	0.02

Chapter 7 Hypotheses Testing

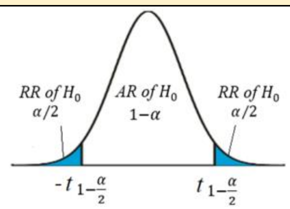
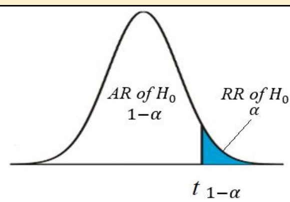
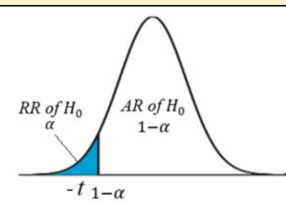
Hypotheses Testing

1-Single Mean

($if\ \sigma\ known$):

Hypothesis *Null H_0* *Alternative (Research) H_A*	$H_0: \mu = \mu_o$ $H_A: \mu \neq \mu_o$	$H_0: \mu \leq \mu_o$ $H_A: \mu > \mu_o$	$H_0: \mu \geq \mu_o$ $H_A: \mu < \mu_o$
Test Statistics (TS)	$Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$		
Rejection Region (RR) of H_0 *Acceptance Region (AR) of H_0*			
Decision	We reject H_0 at the significance level α if		
	$Z < -Z_{1-(\alpha/2)}$ *or* $Z > Z_{1-(\alpha/2)}$ *Two sides test*	$Z > Z_{1-\alpha}$ *One side test*	$Z < -Z_{1-\alpha}$ *One side test*

($if\ \sigma\ unknown$):

Hypothesis *Null H_0* *Alternative (Research) H_A*	$H_0: \mu = \mu_o$ $H_A: \mu \neq \mu_o$	$H_0: \mu \leq \mu_o$ $H_A: \mu > \mu_o$	$H_0: \mu \geq \mu_o$ $H_A: \mu < \mu_o$
Test Statistics (TS)	$t = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$		
Rejection Region (RR) of H_0 *Acceptance Region (AR) of H_0*			
Decision	We reject H_0 at the significance level α if		
	$t < -t_{1-(\alpha/2)}$ *or* $t > t_{1-(\alpha/2)}$ *Two sides test*	$t > t_{1-\alpha}$ *One side test*	$t < -t_{1-\alpha}$ *One side test*

Question 1:

Suppose that we are interested in estimating the true average time in seconds it takes an adult to open a new type of tamper-resistant aspirin bottle. It is known that the population standard deviation is $\sigma = 5.71$seconds. A random sample of 40 adults gave a mean of 20.6 seconds. Let μ be the population mean, then, to test if the meanμ is 21 seconds at level of significant 0.05 ($H_0: \mu = 21$ vs $H_A: \mu \neq 21$) then:

(1) The value of the test statistic is:

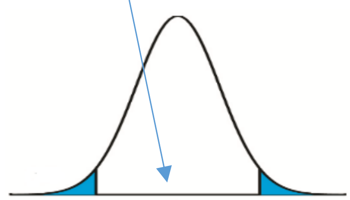
$$\sigma = 5.71 \quad n = 40 \quad \bar{X} = 20.6$$

$$Z = \frac{\bar{X}-\mu_o}{\sigma/\sqrt{n}} = \frac{20.6-21}{5.71/\sqrt{40}} = -0.443$$

A	0.443	B	-0.012	C	-0.443	D	0.012

(2) The acceptance area is:

$$Z_{1-\frac{\alpha}{2}} = Z_{1-\frac{0.05}{2}} = Z_{0.975} = 1.96$$



A	$(-1.96, 1.96)$	B	$(1.96, \infty)$	C	$(-\infty, 1.96)$	D	$(-\infty, 1.645)$

(3) The decision is:

A	Reject H_0	B	Accept H_0	C	No decision	D	None of these

$$P - value = 2 \times P(Z < -0.443) = 2 \times 0.32997 = 0.66 > 0.05$$

or

$$P - value = 2 \times P(Z > |-0.443|) = 2 \times P(Z > 0.443) = 0.66 > 0.05$$

Question 2:

If the hemoglobin level of pregnant women (امرأه حامل) is normally distributed, and if the mean and standard deviation of a sample of 25 pregnant women were $\bar{X} = 13$ (g/dl), $s = 2$ (g/dl). Using $\alpha = 0.05$, to test if the average hemoglobin level for the pregnant women is greater than 10 (g/dl) [$H_0: \mu \leq 10$, $H_A: \mu > 10$].

$$\boxed{s = 2 \ , \quad n = 25 \ , \quad \bar{X} = 13}$$

1. The test statistic is:

A	$Z = \frac{\bar{X}-10}{\sigma/\sqrt{n}}$	B	$Z = \frac{\bar{X}-10}{S/\sqrt{n}}$	C	$t = \frac{\bar{X}-10}{\sigma/\sqrt{n}}$	D	$t = \frac{\bar{X}-10}{S/\sqrt{n}}$

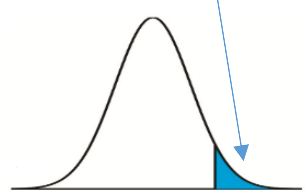
2. The value of the test statistic is:

$$t = \frac{\bar{X} - \mu_o}{S/\sqrt{n}} = \frac{13 - 10}{2/\sqrt{25}} = \boxed{7.5}$$

A	10	B	1.5	C	7.5	D	37.5

3. The rejection of H_0 is:

$$t_{1-\alpha,n-1} = t_{0.95,24} = 1.711$$



1.711

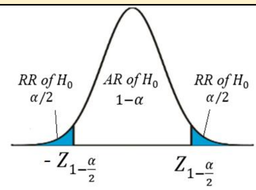
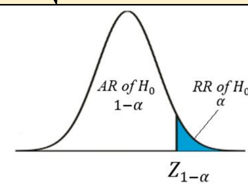
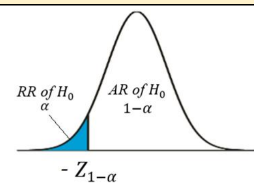
A	$Z < -1.645$	B	$Z > 1.645$	C	$t < -1.711$	D	$t > 1.711$

4. The decision is:

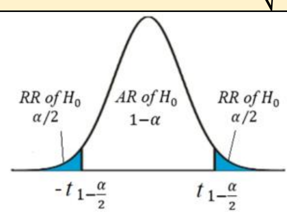
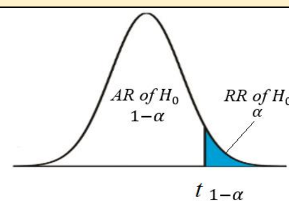
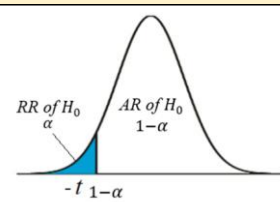
A	Reject H_0
B	Do not reject (Accept) H_0.
C	Accept both H_0 and H_A.
D	Reject both H_0 and H_A.

2-Two Means:

(*if* σ_1 *and* σ_2 *known*):

Hypothesis *Null* H_0 *Alternative (Research)* H_A	$H_0: \mu_1 - \mu_2 = d$ $H_A: \mu_1 - \mu_2 \neq d$	$H_0: \mu_1 - \mu_2 \leq d$ $H_A: \mu_1 - \mu_2 > d$	$H_0: \mu_1 - \mu_2 \geq d$ $H_A: \mu_1 - \mu_2 < d$
Test Statistics (TS)	$Z = \frac{(\bar{X}_1 - \bar{X}_2) - d}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$		
Rejection Region (RR) of H_0 *Acceptance Region (AR) of* H_0			
Decision	We reject H_0 at the significance level α if		
	$Z < -Z_{1-(\alpha/2)}$ *or* $Z > Z_{1-(\alpha/2)}$ *Two sides test*	$Z > Z_{1-\alpha}$ *One side test*	$Z < -Z_{1-\alpha}$ *One side test*

(*if* σ_1 *and* σ_2 *unknown*):

Hypothesis *Null* H_0 *Alternative (Research)* H_A	$H_0: \mu_1 - \mu_2 = d$ $H_A: \mu_1 - \mu_2 \neq d$	$H_0: \mu_1 - \mu_2 \leq d$ $H_A: \mu_1 - \mu_2 > d$	$H_0: \mu_1 - \mu_2 \geq d$ $H_A: \mu_1 - \mu_2 < d$
Test Statistics (TS)	$t = \frac{(\bar{X}_1 - \bar{X}_2) - d}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}} = \frac{(\bar{X}_1 - \bar{X}_2) - d}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$		
Rejection Region (RR) of H_0 *Acceptance Region (AR) of* H_0			
Decision	We reject H_0 at the significance level α if		
	$t < -t_{1-(\alpha/2)}$ *or* $t > t_{1-(\alpha/2)}$ *Two sides test*	$t > t_{1-\alpha}$ *One side test*	$t < -t_{1-\alpha}$ *One side test*

$$S_p^2 = \frac{S_1^2(n_1-1) + S_2^2(n_2-1)}{n_1+n_2-2}$$

Question 3:

A standardized chemistry test was given to 50 girls and 75 boys. The girls made an average of 84, while the boys made an average grade of 82. Assume the population standard deviations are 6 and 8 for girls and boys respectively. To test the null hypothesis

$$H_0: \mu_1 - \mu_2 \leq 0 \;\; vs\; H_A: \mu_1 - \mu_2 > 0 \quad use\; \alpha = 0.05$$

(1) The standard error of $(\bar{X}_1 - \bar{X}_2)$ is:

$$girls: \quad n_1 = 50,\; \bar{X}_1 = 84,\; \sigma_1 = 6$$
$$boys: \quad n_2 = 75,\; \bar{X}_2 = 82,\; \sigma_2 = 8$$

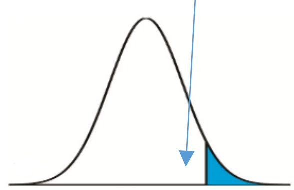
$$S.E(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{6^2}{50} + \frac{8^2}{75}} = 1.2543$$

(2) The value of the test statistic is:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(84-82)}{\sqrt{\frac{6^2}{50} + \frac{8^2}{75}}} = \frac{2}{1.2543} = \boxed{1.5945}$$

(3) The rejection region (RR) of H_0 is:

$$Z_{1-\alpha} = Z_{1-0.05} = Z_{0.95} = 1.645$$



A	$(1.645, \infty)$	B	$(-\infty, -1.645)$	C	$(1.96, \infty)$	D	$(-\infty, -1.96)$

(4) The decision is:

A	Reject H_0
B	Do not reject (Accept) H_0.
C	Accept both H_0 and H_A.
D	Reject both H_0 and H_A.

$$P - value = P(Z > 1.59) = 1 - P(Z < 1.59) = 0.056 > 0.05$$

Question 4:

Cortisol level determinations were made on two samples of women at childbirth. Group 1 subjects underwent emergency cesarean section (عملية قيصرية) following induced labor. Group 2 subjects natural childbirth route following spontaneous labor (الولادة الطبيعية). The random sample sizes, mean cortisol levels, and standard deviations were $(n_1 = 40, \bar{x}_1 = 575, \sigma_1 = 70)$, $(n_2 = 44, \bar{x}_2 = 610, \sigma_2 = 80)$.

If we are interested to test if the mean Cortisol level of group 1 (μ_1) is less than that of group 2 (μ_2) at level 0.05 (or $H_0: \mu_1 \geq \mu_2$ vs $H_1: \mu_1 < \mu_2$), then:

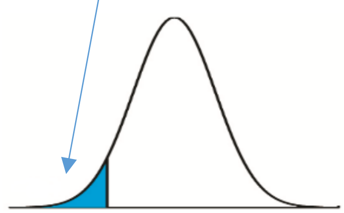
(1) The value of the test statistic is:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(575-610)}{\sqrt{\frac{70^2}{40} + \frac{80^2}{44}}} = \boxed{-2.138}$$

A	-1.326	B	-2.138	C	-2.576	D	-1.432

(2) Reject H_0 if :

$$Z_{1-\alpha} = Z_{0.95} = 1.645$$



-1.645

A	$Z > 1.645$	B	$T > 1.98$	C	$Z < -1.645$	D	$T < -1.98$

(3) The decision is:

A	Reject H_0	B	Accept H_0	C	No decision	D	None of these

$$P - value = P(Z < -2.138) = 0.01618 < 0.05$$

Question 5:

An experiment was conducted to compare time length (duration time in minutes) of two types of surgeries (A) and (B). 10 surgeries of type (A) and 8 surgeries of type (B) were performed. The data for both samples is shown below.

Surgery type	A	B
Sample size	10	8
Sample mean	14.2	12.8
Sample standard deviation	1.6	2.5

Assume that the two random samples were independently selected from two normal populations with equal variances. If μ_A and μ_B are the population means of the time length of surgeries of type (A) and type (B), then, to test if μ_A is greater than μ_B at level of significant 0.05 ($H_0: \mu_A \leq \mu_B$ vs $H_A: \mu_A > \mu_B$) then:

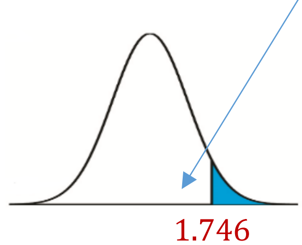
1. The value of the test statistic is:

$$S_p^2 = \frac{S_1^2(n_1-1)+S_2^2(n_2-1)}{n_1+n_2-2} = \frac{1.6^2(10-1)+2.5^2(8-1)}{10+8-2} = 4.174$$

$$t = \frac{(\bar{X}_1-\bar{X}_2)}{Sp\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{(14.2-12.8)}{\sqrt{4.174}\sqrt{\frac{1}{10}+\frac{1}{8}}} = \boxed{1.44}$$

2. Reject H_0 if:

$$t_{1-\alpha,n_1+n_2-2} = t_{0.95,10+8-2} = t_{0.95,16} = 1.746$$



A	$Z > 1.645$	B	$Z < -1.645$	C	$T > 1.746$	D	$T < -1.746$

3. The decision is:

A	Reject H_0	B	Accept H_0	C	No decision	D	None of these

Question 6:

A researcher was interested in comparing the mean score of female students μ_1 ,with the mean score of male students μ_2 in a certain test. Assume the populations of score are normal with equal variances. Two independent samples gave the following results:

	Female	Male
Sample size	$n_1 = 5$	$n_2 = 7$
Mean	$\bar{X}_1 = 82.63$	$\bar{X}_2 = 80.04$
Variance	$S_1^2 = 15.05$	$S_2^2 = 20.79$

Test that is there is a difference between the mean score of female students and the mean score of male students.

1. The hypotheses are:

A	$H_0: \mu_1 = \mu_2$ $H_A: \mu_1 \neq \mu_2$	B	$H_0: \mu_1 = \mu_2$ $H_A: \mu_1 < \mu_2$	C	$H_0: \mu_1 < \mu_2$ $H_A: \mu_1 > \mu_2$	D	$H_0: \mu_1 \leq \mu_2$ $H_A: \mu_1 > \mu_2$

2. The value of the test statistic is:

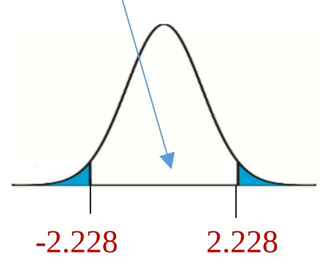
$$S_p^2 = \frac{S_1^2(n_1-1)+S_2^2(n_2-1)}{n_1+n_2-2} = \frac{15.05(4)+20.79(6)}{5+7-2} = 18.494$$

$$t = \frac{(\bar{X}_1-\bar{X}_2)}{Sp\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{82.63-80.04}{\sqrt{18.494}\sqrt{\frac{1}{5}+\frac{1}{7}}} = \boxed{1.029}$$

A	1.3	B	1.029	C	0.46	D	0.93

3. The acceptance region (AR) of H_0 is:

$$t_{1-\frac{\alpha}{2},n_1+n_2-2}$$
$$= t_{1-\frac{0.05}{2},5+7-2}$$
$$= t_{0.975,10}$$
$$= 2.228$$



A	$(2.228, \infty)$	B	$(-\infty, -2.228)$	C	(-2.228, 2.228)	D	(-1.96, 1.96)

Question 7:

A nurse researcher wished to know if graduates of baccalaureate (بكالوريس) nursing program and graduate of associate degree (الزمالة) nursing program differ with respect to mean scores on personality inventory at α = 0.02. A sample of 50 associate degree graduates (sample A) and a sample of 60 baccalaureate graduates (sample B) yielded the following means and standard deviations:

$$\bar{X}_A = 88.12,\ \ S_A = 10.5,\ \ n_A = 50$$
$$\bar{X}_B = 83.25,\ \ S_B = 11.2,\ \ n_B = 60$$

1) The hypothesis is:

A	$H_0: \mu_1 = \mu_2$ $H_A: \mu_1 \neq \mu_2$	B	$H_0: \mu_1 \leq \mu_2$ $H_A: \mu_1 > \mu_2$	C	$H_0: \mu_1 \geq \mu_2$ $H_A: \mu_1 < \mu_2$	D	None of these

2) The test statistic is:

A	Z	B	t	C	F	D	None of these

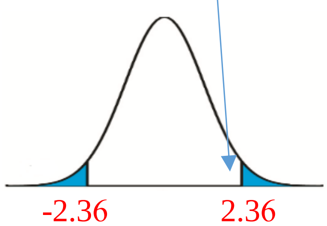
3) The computed value of the test statistic is:

$$S_p^2 = \frac{S_1^2(n_1-1)+S_2^2(n_2-1)}{n_1+n_2-2} = \frac{10.5^2(50-1)+11.2^2(60-1)}{50+60-2} = 118.55$$

$$t = \frac{(\bar{X}_1-\bar{X}_2)}{Sp\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{88.12-83.25}{\sqrt{118.55}\sqrt{\frac{1}{50}+\frac{1}{60}}} = \boxed{2.34}$$

4) The critical region (rejection area) is:

$$t_{1-\frac{\alpha}{2},n_1+n_2-2}$$
$$= t_{1-\frac{0.02}{2},\ 50+60-2}$$
$$= t_{0.99,108} = 2.36$$

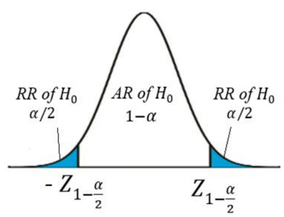
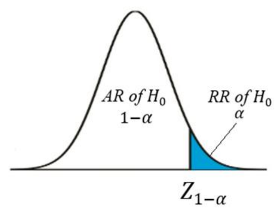
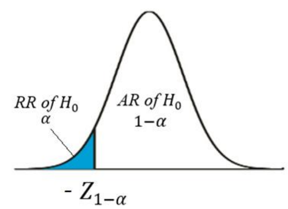


A	2.60 or − 2.60	B	2.06 or − 2.06	C	2.36 or − 2.36	D	2.58

5) Your decision is:

A	Reject H_0	B	Accept H_0	C	Accept H_A	D	No decision

3- Single proportion:

Hypothesis *Null H_0* *Alternative (Research) H_A*	$H_0: p = p_o$ $H_A: p \neq p_o$	$H_0: p \leq p_o$ $H_A: p > p_o$	$H_0: p \geq p_o$ $H_A: p < p_o$
Test Statistics (TS)	$Z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0 q_0}{n}}} \sim N(0,1)$		
Rejection Region (RR) of H_0 *Acceptance Region (AR) of H_0*			
Decision	We reject H_0 at the significance level α if		
	$Z < -Z_{1-(\alpha/2)}$ *or* $Z > Z_{1-(\alpha/2)}$ *Two sides test*	$Z > Z_{1-\alpha}$ *One side test*	$Z < -Z_{1-\alpha}$ *One side test*

Question 8:

Toothpaste (معجون الأسنان) company claims that more than 75% of the dentists recommend their product to the patients. Suppose that 161 out of 200 dental patients reported receiving a recommendation for this toothpaste from their dentist. Do you suspect that the proportion is actually morethan75%. If we use 0.05 level of significance to test $H_0: P \le 0.75$, $H_A: P > 0.75$, then:

(1) The sample proportion $\hat{p}$ is:

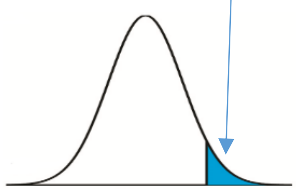
$$n = 200, \ \hat{p} = \frac{161}{200} = 0.8050$$

(2) The value of the test statistic is:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.805 - 0.75}{\sqrt{\frac{(0.75)(0.25)}{200}}} = \boxed{1.7963}$$

(3) The decision is:

$$\alpha = 0.05$$
$$Z_{1-\alpha} = Z_{0.95} = 1.645$$



A	Reject H_0 (agree with the claim)
B	Do not reject (Accept) H_0
C	Accept both H_0 and H_A
D	Reject both H_0 and H_A

$$P - value = P(Z > 1.7963) = 1 - P(Z < 1.7963) = 1 - 0.96407 = 0.03593 < 0.05$$

Question 9:

A researcher was interested in studying the obesity (السمنة) disease in a certain population. A random sample of 400 people was taken from this population. It was found that 152 people in this sample have the obesity disease. If p is the population proportion of people who are obese. Then, to test if p is greater than 0.34 at level 0.05 ($H_0: p \leq 0.34$ vs $H_A: p > 0.34$) then:

(1) The value of the test statistic is:

$$n = 400, \quad \hat{p} = \frac{152}{400} = 0.38$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.38 - 0.34}{\sqrt{\frac{0.34 \times 0.66}{400}}} = \boxed{1.69}$$

A	0.023	B	1.96	C	2.50	D	1.69

(2) The P-value is

$$P - value = P(Z > 1.69) = 1 - P(Z < 1.69) = 1 - 0.9545 = 0.0455$$

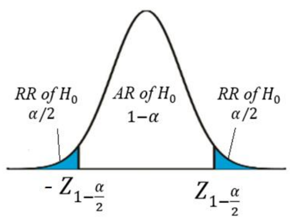
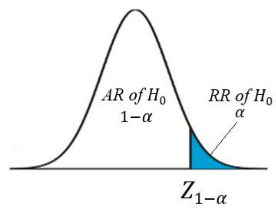
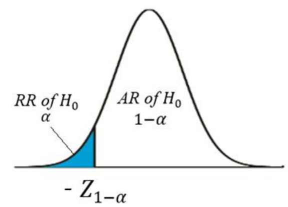
A	0.9545	B	0.0910	C	0.0455	D	1.909

(3) The decision is:

$$P - value = 0.0455 < 0.05$$

A	Reject H_0	B	Accept H_0	C	No decision	D	None of these

4-Two proportions:

<table>
<tr><td>Hypothesis
Null H_0
Alternative (Research) H_A</td><td>$H_0: p_1 - p_2 = d$
$H_A: p_1 - p_2 \neq d$</td><td>$H_0: p_1 - p_2 \leq d$
$H_A: p_1 - p_2 > d$</td><td>$H_0: p_1 - p_2 \geq d$
$H_A: p_1 - p_2 < d$</td></tr>
<tr><td>Test Statistics
(TS)</td><td colspan="3">$Z = \frac{(\hat{p}_1 - \hat{p}_2) - d}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}} = \frac{(\hat{p}_1 - \hat{p}_2) - d}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$</td></tr>
<tr><td>Rejection Region
(RR) of H_0
Acceptance Region
(AR) of H_0</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td rowspan="2">Decision</td><td colspan="3">We reject H_0 at the significance level α if</td></tr>
<tr><td>$Z < -Z_{1-(\alpha/2)}$
or
$Z > Z_{1-(\alpha/2)}$
Two sides test</td><td>$Z > Z_{1-\alpha}$

One side test</td><td>$Z < -Z_{1-\alpha}$

One side test</td></tr>
</table>

Question 10:

In a first sample of 200 men, 130 said they used seat belts and a second sample of 300 women, 150 said they used seat belts. To test the claim that men are more safety-conscious than women ($H_0: p_1 - p_2 \leq 0, H_1: p_1 - p_2 > 0$), at 0.05 level of significant:

(1) The value of the test statistic is:

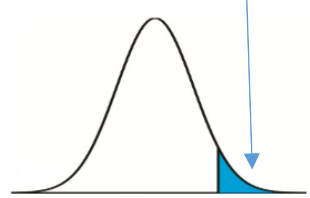
$$n_1 = 200,\ \ \hat{p}_1 = \frac{130}{200} = 0.65 \qquad n_2 = 300,\ \ \hat{p}_2 = \frac{150}{300} = 0.5$$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{130+150}{200+300} = 0.56$$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(0.65-0.5)}{\sqrt{(0.56)(0.44)\left(\frac{1}{200} + \frac{1}{300}\right)}} = \boxed{3.31}$$

(2) The decision is:

$$Z_{1-\alpha} = Z_{1-0.05} = Z_{0.95} = 1.645$$



1.645

A	Reject H_0
B	Do not reject (Accept) H_0
C	Accept both H_0 and H_A
D	Reject both H_0 and H_A

$$P-value = P(Z > 3.31) = 1 - P(Z < 3.31) = 1 - 0.99953 = 0.00047 < 0.05$$

Question 11:

In a study of diabetes, the following results were obtained from samples of males and females between the ages of 20 and 75. Male sample size is 300 of whom 129 are diabetes patients, and female sample size is 200 of whom 50 are diabetes patients. If P_M, P_F are the diabetes proportions in both populations and $\hat{p}_M, \hat{p}_F$ are the sample proportions, then:
A researcher claims that the Proportion of diabetes patients is found to be more in males than in female ($H_0: P_M - P_F \leq 0$ vs $H_A: P_M - P_F > 0$) . Do you agree with his claim, take $\alpha = 0.10$

$$n_m = 300 \quad , \quad x_m = 129 \quad \Longrightarrow \hat{p}_1 = \frac{129}{300} = 0.43$$

$$n_f = 200 \quad , \quad x_f = 50 \quad \Longrightarrow \hat{p}_2 = \frac{50}{200} = 0.25$$

(1) The pooled proportion is:

$$\bar{p} = \frac{x_m + x_f}{n_m + n_f} = \frac{129 + 50}{300 + 200} = 0.358$$

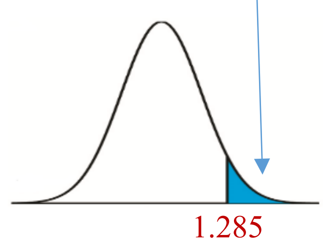
A	0.43	B	0.18	C	0.358	D	0.68

(2) The value of the test statistic is:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(0.43 - 0.25)}{\sqrt{(0.358)(1 - 0.358)\left(\frac{1}{300} + \frac{1}{200}\right)}} = \boxed{4.11}$$

(3) The decision is:

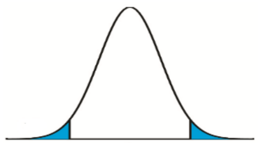
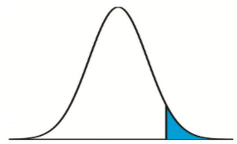
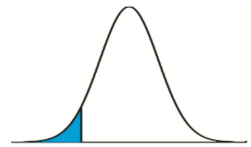
$$Z_{1-\alpha} = Z_{1-0.10} = Z_{0.90} = 1.285$$



A	Agree with the claim (Reject H_o)
B	do not agree with the claim
C	Can't say

$$P - value = P(Z > 4.11) = 1 - P(Z < 4.11) = 1 - 1 = 0 < 0.05$$

- *P – value:*

Hypothesis	$H_0: \mu = \mu_o$ $H_A: \mu \neq \mu_o$	$H_0: \mu \leq \mu_o$ $H_A: \mu > \mu_o$	$H_0: \mu \geq \mu_o$ $H_A: \mu < \mu_o$
RR			
P-value	$2 \times P(Z > \lvert TS \rvert)$	$P(Z > TS)$	$P(Z < TS)$

$2 \times P(Z > TS)$ *If* $TS > 0$	$2 \times P(Z < TS)$ *If* $TS < 0$

	population normal or not normal n large ($n \geq 30$)		population normal n small ($n < 30$)	
	σ known	σ unknown	σ known	σ unknown
Testing	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$	$Z = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$	$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$

- **Two Samples Test for Paired Observation**

Question 1:

The following contains the calcium levels of eleven test subjects at zero hours and three hours after taking a multi-vitamin containing calcium.

Pair	0 hour (X_i)	3 hours (Y_i)	Difference $D_i = X_i - Y_i$
1	17.0	17.0	0.0
2	13.2	12.9	0.3
3	35.3	35.4	-0.1
4	13.6	13.2	0.4
5	32.7	32.5	0.2
6	18.4	18.1	0.3
7	22.5	22.5	0.0
8	26.8	26.7	0.1
9	15.1	15.0	0.1

The sample mean and sample standard deviation of the differences D are 0.144 and 0.167, respectively. To test whether the data provide sufficient evidence to indicate a difference in mean calcium levels ($H_0: \mu_1 = \mu_2$ against$H_1: \mu_1 \neq \mu_2$) with $\alpha = 0.10$ we have: $\boxed{\bar{D} = 0.144\ ,\ \ S_d = 0.167\ \ ,\ \ n = 9}$

[1]. The reliability coefficient (the tabulated value) is:

$$t_{1-\frac{\alpha}{2},n-1} = t_{1-\frac{0.1}{2},9-1} = t_{0.95,8} = \boxed{1.860}$$

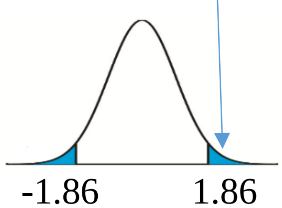
[2]. The value of the test statistic is:

$$\boxed{\begin{matrix} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{matrix}} \Longrightarrow \boxed{\begin{matrix} H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_1 - \mu_2 \neq 0 \end{matrix}} \Longrightarrow \boxed{\begin{matrix} H_0: \mu_D = 0 \\ H_1: \mu_D \neq 0 \end{matrix}}$$

$$T = \frac{\bar{D} - \mu_D}{S_d/\sqrt{n}} = \frac{0.144 - 0}{0.167/\sqrt{9}} = \boxed{2.5868}$$

[3]. The decision is:

We reject H_0



Question 2:

Scientists and engineers frequently wish to compare two different techniques for measuring or determining the value of a variable. Reports the accompanying data on amount of milk ingested by each of 14 randomly selected infants.

Pair	DD method (X_i)	TW method (Y_i)	Difference $D_i = X_i - Y_i$
1	1509	1498	11
2	1418	1254	164
3	1561	1336	225
4	1556	1565	-9
5	2169	2000	169
6	1760	1318	442
7	1098	1410	-312
8	1198	1129	69
9	1479	1342	137
10	1281	1124	157
11	1414	1468	-54
12	1954	1604	350
13	2174	1722	452
14	2058	1518	540

1. The sample mean of the differences $\overline{D}$ is:

$$\overline{D} = \frac{11+164+225-9+169+442-312+\cdots+540}{14} = 167.21$$

A	167.21	B	0.71	C	0.61	D	0.31

2. The sample standard deviation of the differences S_D is:

$$S_D = \sqrt{\frac{\sum(D_i - \overline{D})^2}{n-1}} = 228.21$$

A	3.15	B	−0.71	C	71.53	D	228.21

3. The reliability coefficient to construct 90% confidence interval for the true average difference between intake values measured by the two methods μ_D is:

$$\text{The reliability coefficient} = t_{1-\frac{\alpha}{2},n-1} = t_{0.95,13} = 1.771$$

A	1.96	B	1.771	C	2.58	D	1.372

4. The 90% lower limit for μ_D is:

$$= \bar{D} - \left(t_{1-\frac{\alpha}{2},n-1} \times \frac{S_D}{\sqrt{n}}\right)$$

$$= 167.21 - \left(1.771 \times \frac{228.12}{\sqrt{14}}\right) = 59.19$$

A	24.92	B	22.55	C	59.19	D	44.96

5. The 90% upper limit for μ_D is:

$$= \bar{D} + \left(t_{1-\frac{\alpha}{2},n-1} \times \frac{S_D}{\sqrt{n}}\right)$$

$$= 167.21 + \left(1.771 \times \frac{228.12}{\sqrt{14}}\right) = 275.23$$

A	224.92	B	322.55	C	275.23	D	24.96

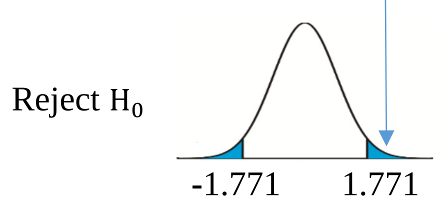
To test $H_0: \mu_D = 0$ versus $H_A: \mu_D \neq 0$, $\alpha = 0.10$ as a level of significance we have:

6. The value of the test statistic is:

$$T = \frac{\bar{D}-\mu_D}{S_d/\sqrt{n}} = \frac{167.21-0}{228.12/\sqrt{14}} = \boxed{2.74}$$

A	2.74	B	−0.7135	C	−7.153	D	−0.3157

7. The decision is:



A	Reject H_0	B	Accept H_0	C	No decision	D	None of these

Question 3:

In a study of a surgical procedure used to decrease the amount of food that person can eat. A sample of 10 persons measures their weights before and after one year of the surgery, we obtain the following data:

Before surgery (X)	148	154	107	119	102	137	122	140	140	117
After surgery (Y)	78	133	80	70	70	63	81	60	85	120
$D_i = X_i - Y_i$	70	21	27	49	32	74	41	80	55	-3

We assume that the data comes from normal distribution.
For 90% confidence interval for μ_D, where μ_D is the difference in the average weight before and after surgery.

1. The sample mean of the differences $\overline{D}$ is:

$$\overline{D} = \frac{70+21+27+49+32+74+41+80+55-3}{10} = 44.6$$

2. The sample standard deviation of the differences S_D is:

$$S_D = \sqrt{\frac{\sum(D_i - \overline{D})^2}{n-1}} = 26.2$$

3. The 90% upper limit of the confidence interval for μ_D is:

$$t_{1-\frac{\alpha}{2},n-1} = t_{0.95,9} = 1.833$$

$$= \overline{D} + \left(t_{1-\frac{\alpha}{2},n-1} \times \frac{S_D}{\sqrt{n}}\right)$$

$$= 44.6 + \left(1.833 \times \frac{26.2}{\sqrt{10}}\right) = 59.79$$

4. To test $H_0: \mu_D \geq 43$ versus $H_A: \mu_D < 43$, with α=0.10 as a level of significance, the value of the test statistic is:

$$T = \frac{\overline{D} - \mu_D}{S_d/\sqrt{n}} = \frac{44.6-43}{26.2/\sqrt{10}} = \boxed{0.19}$$

5. The decision is:

$$-t_{1-\alpha,n-1} = -t_{0.90,9} = -1.383 \Longrightarrow \boxed{0.19 \notin RR: (-\infty, -1.383)}$$

A	Reject H_0	B	Do not reject H_0	C	No decision	D	None of these

<u>Questions 4:</u>

Trace metals in drinking water affect the flavor and an unusually high concentration can pose a health hazard. Ten pairs of data were taken measuring zinc concentration in bottom water and surface water.
The Data is given below:

	zinc concentration in Bottom water	zinc concentration in Surface water	Difference
1	0.43	0.415	0.015
2	0.266	0.238	0.028
3	0.567	0.39	0.177
4	0.531	0.41	0.121
5	0.707	0.605	0.102
6	0.716	0.609	0.107
7	0.651	0.632	0.019
8	0.589	0.523	0.066
9	0.469	0.411	0.058
10	0.723	0.612	0.111

Note that the mean and the standard deviation of the difference are given respectively by $\overline{D} = 0.0804$ and $S_D = 0.0523$ We want to determine the 95 % confidence interval for $\mu_1 - \mu_2$, where μ_1 and μ_2 represent the true mean zinc concentration in Bottom water and surface water respectively. Assume the distribution of the differences to be approximately normal.

1. The 95% lower limit for $\mu_1 - \mu_2$ equals to:

A	0.02628	B	0.13452	C	0.04299	D	0.11781

2. The 95% upper limit for $\mu_1 - \mu_2$ equals to:

A	0.02628	B	0.13452	C	0.04299	D	0.11781

	Estimation	Testing
Single mean	$\bar{X} \pm Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ σ known	$Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ σ known
	$\bar{X} \pm t_{1-\frac{\alpha}{2},(n-1)} \frac{S}{\sqrt{n}}$ σ unknown	$T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$ σ unknown
Two means	$(\bar{X}_1 - \bar{X}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ σ_1 and σ_2 known	$Z = \frac{(\bar{X}_1-\bar{X}_2)-d}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$ σ_1 and σ_2 known
	$(\bar{X}_1 - \bar{X}_2) \pm t_{1-\frac{\alpha}{2},\ (n_1+n_2-2)}\ S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ σ_1 and σ_2 unknown	$T = \frac{(\bar{X}_1-\bar{X}_2)-d}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ σ_1 and σ_2 unknown
Single proportion	$\hat{p} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}$	$Z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0 q_0}{n}}}$
Two proportions	$(\hat{p}_1 - \hat{p}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$	$Z = \frac{(\hat{p}_1-\hat{p}_2)-d}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

$$S_p^2 = \frac{S_1^2(n_1-1)+S_2^2(n_2-1)}{n_1+n_2-2}$$

	H_0 is true	H_0 is false
Accepting H_0	Correct decision ✓	Type II error (β)
Rejecting H_0	Type I error (α)	Correct decision ✓

Type I error = Rejecting H_0 when H_0 is true P(Type I error) = P(Rejecting $H_0 \| H_0$ is true) = α	Type II error = Accepting H_0 when H_0 is false P(Type II error) = P(Accepting $H_0 \| H_0$ is false) = β

- **Question from previous midterms and finals:**

Q1. In the procedure of testing the statistical hypotheses H_0 against H_A using a significance level α

1. The type I error occur if we:

A	Rejecting H_0 when H_0 is true	B	Rejecting H_0 when H_0 is false	C	Accepting H_0 when H_0 is true	D	Accepting H_0 when H_0 is false

2. The probability of type I error is:

A	β	B	α	C	$1-\beta$	D	$1-\alpha$

3. The level of significance is:

A	The probability of rejecting H_A	B	The probability of rejecting H_0
C	The probability of making a Type I error	D	The probability of making a Type II error

4. When we use P-value method, we reject H_0 if

A	P- value > α	B	P- value < α	C	P- value < β	D	P- value > β

5. If the P-value = 0.0625 and α = 0.05, the decision is:

A	Reject H_0	B	Accept H_0	C	Reject H_A	D	Accept H_A

6. To determine the rejection region for H_0, it depends on:

A	α and H_A	B	H_0	C	α and H_0	D	β

7. Which one is an example of two-tailed test:

A	$H_A: \mu = 0$	B	$H_A: \mu \neq 0$	C	$H_A: \mu < 0$	D	$H_A: \mu > 0$

8. If the P-value = 0.0625 and α = 0.05, the decision is:

A	Reject H_0	B	Accept H_0	C	Reject H_A	D	Accept H_A

9. If the distribution of the random sample is normal and standard deviation of the population is known, which type of the interval should be considered?

A	z - interval	B	x - interval	C	t - interval	D	c - interval

10. An appropriate 95% CI for μ has been calculated as (-1.5 , 3.5) based on $n_1 = 15$, $n_2 = 17$ observations from two independent population with normal distribution. The hypotheses of interest $H_0: \mu_1 = \mu_2$ vs $H_A: \mu_1 \neq \mu_2$. Based on this CI and at $\alpha = 0.05$,

A	We should reject H_0	B	We should not reject H_0

Q2. To compare the mean times spent waiting for a heart transplant for two age groups, you randomly select several people in each age group who have had a heart transplant. The result is shown below. Assume both population is are normally distributed with equal variance.

Sample statistics for heart transplant		
Age group	18-34	35-49
Mean	171 days	169 days
Standard deviation	8.5 days	11.5 days
Sample size	20	17

Do this data provide sufficient evident to indicate a difference among the population means at $\alpha = 0.05$

1. The alternative hypothesis is:

A	$H_A: \mu_1 \neq \mu_2$	B	$H_A: \mu_1 \leq \mu_2$	C	$H_A: \mu_1 > \mu_2$	D	$H_A: \mu_1 = \mu_2$

2. The pooled estimator of the common variance S_p^2 is:

A	9935.82	B	105.5214	C	10.4429	D	99.6786

3. The appropriate test statistics is:

A	$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$	B	$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	C	$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	D	$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$

4. The 95% confidence interval for the different in mean times spent waiting for heart transplant for the two age groups:

A	$(-3.548, 7.565)$	B	$(-0.1306, 4.1306)$	C	$(-4.6862, 8.6862)$	D	$(-4.8519, 8.8519)$

5. Base on the 95% C.I. in the above question, it can be concluded that:

A	$\bar{X}_1 = \bar{X}_2$	B	$\mu_1 \neq \mu_2$	C	$\mu_1 = \mu_2$	D	None of these