STATISTICS

EX 1

Given the following series of data on Gender and Height for 8 patients.

Id	Height	Gender
1	165	M
2	157	F
3	168	F
4	178	M
5	171	F
6	182	M
7	182	M
8	153	F

Fill the table.

	Mean	Median	Variance	Standard Deviation	Degrees of Freedom
M					
F					
Population					

EX 2

186 patients were given a therapy for a certain disease; 122 had therapy A, the other 64 had therapy B. In group A, there were 37 responders (patients who had benefit from the therapy). In group B, there were 32 responders. Which was the best treatment? How can we measure the advantage with this treatment? Among responders, how many had treatment B?

EX 3

A certain treatment is used in two different centres, A and B; patients in centre A were 25 and were on average 54 years old; patients treated in centre B were 62 and had mean age equal to 58 years. What is the overall mean among all patients who got the treatment?

EX 4

Pregnant women (within month 4) who are being followed-up by a nutritionist had weights (kg) equal to: 64.3; 65.2; 70.0; 54.5; 58.8; 81.5; 61.0; 62.0. What was the mean? and the median? Do data suggest a strong skewness of the distribution of the Weight?

EX 5

The Age quartiles in a sample of participants in a clinical trial were respectively 27, 41, 59.

a) This means that:
- **1 out of 4 was younger than ____ years o**
- **1 out of 4 was older than ____ years o**
- **2 out of 4 were between ____ and ___ years old o**
- **half of them was more than___ years old**

b) Additionally we know that mean and standard deviation were respectively equal to 42 and 12. Can we say whether the distribution was approximately Normal or not?

c) Which index of position is appropriate to give a Better description of the distribution?

EX 6

The guaranteed average life of a certain type of electric bulb is 1000 hours with a standard deviation of 125 hours. It is decided to sample the output so as to ensure that 90% of the bulbs do not fall short of the guaranteed average by more than 2.5%.Use Central limit theorem to find minimum sample size.

EX 7

A salesman in a mall claims that at most 60 percent of the shoppers entering the store leave without making a purchase. A random sample of 50 shoppers showed that 35 of them left without making a purchase. Find out whether the claims of the salesman are consistent with that of the sample results. Use a level of significance of 0.05.

EX 8

The weight distribution of a sample of adults with physical inabilities is approximately Normal, with mean 72 and standard deviation 8.

Find an interval of values around the mean such that:

a) includes 95% of the observed values

b) includes almost all observed values (and thus coincides with the range, min-max)

c) includes 50% of the observed values

EX 9

Examine the below tables. If you are to build a machine-learning model that predicts the sale of the product, which relationships between variables will you choose? Why?

Correlation Table:

	Credit-History	Gender	Credit-Amount	Age	Housing	Job
Credit-History	1	-0.78	0.83	0.1	-0.21	0.3
Gender	-0.78	1	0.97	0.6	0.21	-0.52
Credit-Amount	0.83	0.97	1	0.03	0.31	0.87
Age	0.1	0.6	0.03	1	0.15	0.21
Housing	-0.21	0.21	0.31	0.15	1	0.21
Job	0.3	-0.52	0.87	0.21	0.35	1

Covariance Table:

	Credit-History	Gender	Credit-Amount	Age	Housing	Job
Credit-History	98.343	-532.78	110.13	234.8	-54.12	80.34
Gender	-532.58	145.32	23.7	340.9	123.1	-10.23
Credit-Amount	110.13	23.7	456.34	17.33	12.31	1897.77
Age	234.8	340.9	17.33	34.56	120.5	98.12
Housing	-54.12	123.1	12.31	120.5	23.41	542.98
Job	80.34	-10.23	1897.77	98.12	542.98	10000.23

includes 50% of the observed values