

**Q1. [6 marks]**

A) For each of the following variables, specify whether the variable is **discrete or continuous** and give the appropriate **measurement scale**:

a) Patients status after surgery. [1 mark]

b) Height of university female students. [1 mark]

B) For the following data:

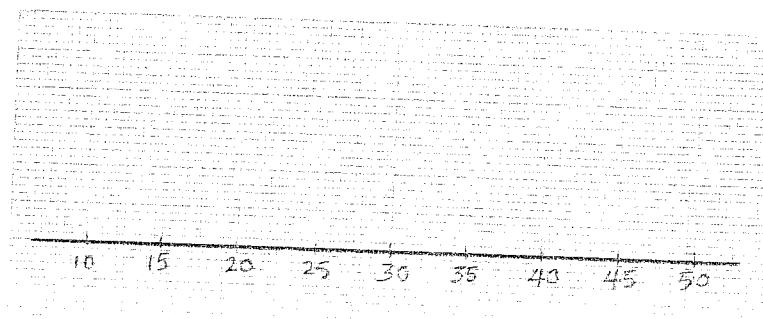
11	13	18	20	20	26	26	30	32	34	38	40
----	----	----	----	----	----	----	----	----	----	----	----

a) Find the cumulative frequency of the observation 30. [0.5 mark]

b) Find the relative frequency of the observation 26. [1 mark]

b) Find the quartiles :  $Q_1$ ,  $Q_2$ ,  $Q_3$ . [1.5 marks]

c) Draw the Boxplot. [1 mark]



**Q2. [6 marks]**

Swine flu has contaminated 200 people. Each of the patients may be hospitalized (HOS) or may not be hospitalized (NHOS). A patient may survive (SUR) or may not survive (NSUR). Assume following frequencies:

	SUR ( $C$ )	NSUR ( $D$ )
HOS ( $A$ )	40	20
NHOS ( $B$ )	90	50

A) Find the probability that:

a) a hospitalized patient will survive.

[1.5 marks]

b) a non-survived patient was hospitalized.

[1.5 marks]

c) a patient is not hospitalized or survived.

[1.5 marks]

B) Are the events  $A$  and  $C$  independent? why?

[1.5 marks]

**Q3. [6 marks]**

A) Suppose that it is well known that, 40% of a certain human **female** population are immune to a certain contagious disease, when it spreads through the population. Find the probability that in a group of 12 females

a) Exactly 8 will be immune. [1 mark]

b) Between 5 and 8 (inclusive) will be immune. [1 mark]

B) The serum iron value for a healthy young man in certain country has a normal distribution with mean 120 and standard deviation 15 micrograms per 100 ml, respectively.

a) Find the probability that a randomly selected healthy young man from this country will have serum iron values more than 105. [1 mark]

b) What is the top 10% serum iron value in young men? [1 mark]

c) Find the probability that 36 randomly selected healthy young men from this country will have **mean** serum iron value at most 115. [2 marks]

**Q4. [6 marks]**

In a study of blood pressure the following results were obtained from samples of young men (30-40 years) in two countries.

	Sample size	Number with high blood pressure
Country A	120	8
Country B	250	10

- a) Construct a 99% confidence interval for the proportion of all young men with high blood pressure in country A. **[3 marks]**

- b) Can we conclude, at 5% significance level, that the proportions of all young men with high blood pressure in the two countries are different? **[3 marks]**

**Q5. [8 marks]**

**A)** The following table gives the urine levels in two samples of preterm infants with and without late metabolic acidosis.

	Sample size	Sample mean	Sample standard deviation
With condition	50	8.5	4.5
Without condition	100	6.8	3.6

Can we conclude, at 5% significance level, that the mean urine levels of the two groups are different?  
Base your conclusion on the  $p$ -value of the test?

**[4 marks]**

**B)** The following data are the Geriatric Depression Scale (GDS) scores for older women with depression pre and post a certain therapy. Higher scores indicate more severe depression symptoms.

Pre-GDS	12	10	16	2	12	18
Post-GDS	11	10	11	3	10	19

Can we conclude, at 5% significance level, that on average the therapy is effective in reducing the GDS scores?

**[4 marks]**

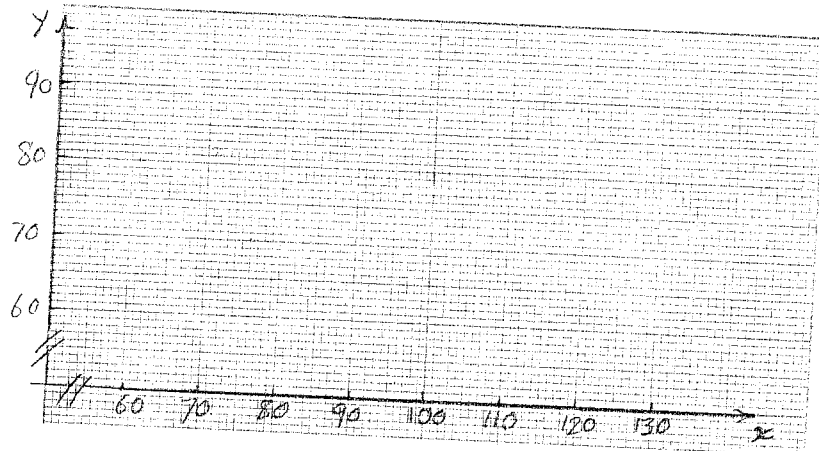
**Q6. [8 marks]**

The Following data show the weight (in kg) before a diet program ( $x$ ) and the weight after of the diet program ( $y$ ) for 8 subjects:

$x$	117	111	98	104	105	100	81	89
$y$	83	85	75	82	82	77	62	69

$$(\sum x = 805, \sum x^2 = 81937, \sum y = 615, \sum y^2 = 47721, \sum xy = 62497)$$

a) Draw the scatter plot.



[1 mark]

b) Write down the equation of the least-squares regression line and plot it on the scatter plot. [2 marks]

c) Find the predicted values of  $y$  and their corresponding errors for  $x = 81$  and  $x = 117$ . [2 marks]

d) Find the correlation coefficient, coefficient of determination, and explain what do they mean.

[3 marks]