Assignment 1 Marking Scheme

LAST NAME: STRUTHERS

FIRST NAME: CYNTHA

USERID: castruth

UWaterloo ID: 20456458

Data must be generated using your ID number not the ID number in the posted example; otherwise 4 marks (10%) are deducted.

Problem 1:

The first five numbers in your Gaussian data set are:

-12.89	**-5.67**	**-2.60**	**-1.54**	**-0.31**

Sample mean = 8.114650

Sample standard deviation = 4.812293

2 marks

The five number summary is:

-12.89	**5.36**	**7.815**	**11.320**	**20.770**

Sample median = 7.815

Range = 20.770 – (-12.89) = 33.66

IQR = 11.32 – 5.36 = 5.96

Sample skewness = -0.2029152

Sample kurtosis = 4.486426

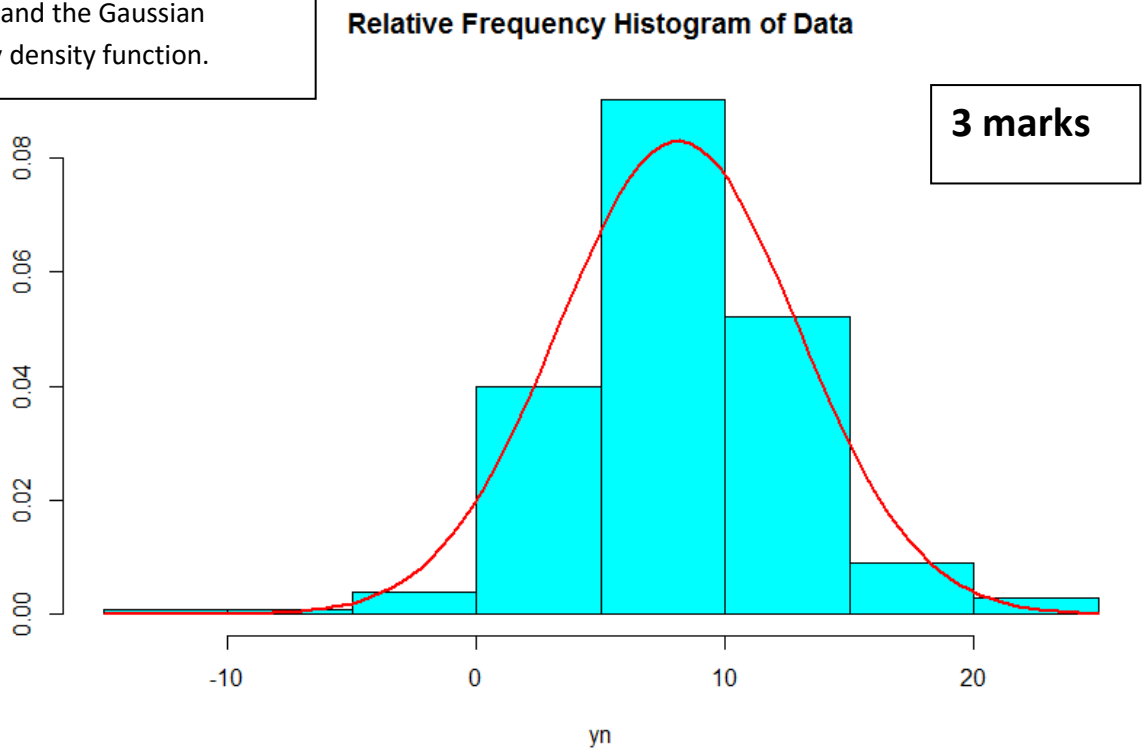
2 marks

Sample mean and median should be close in value.

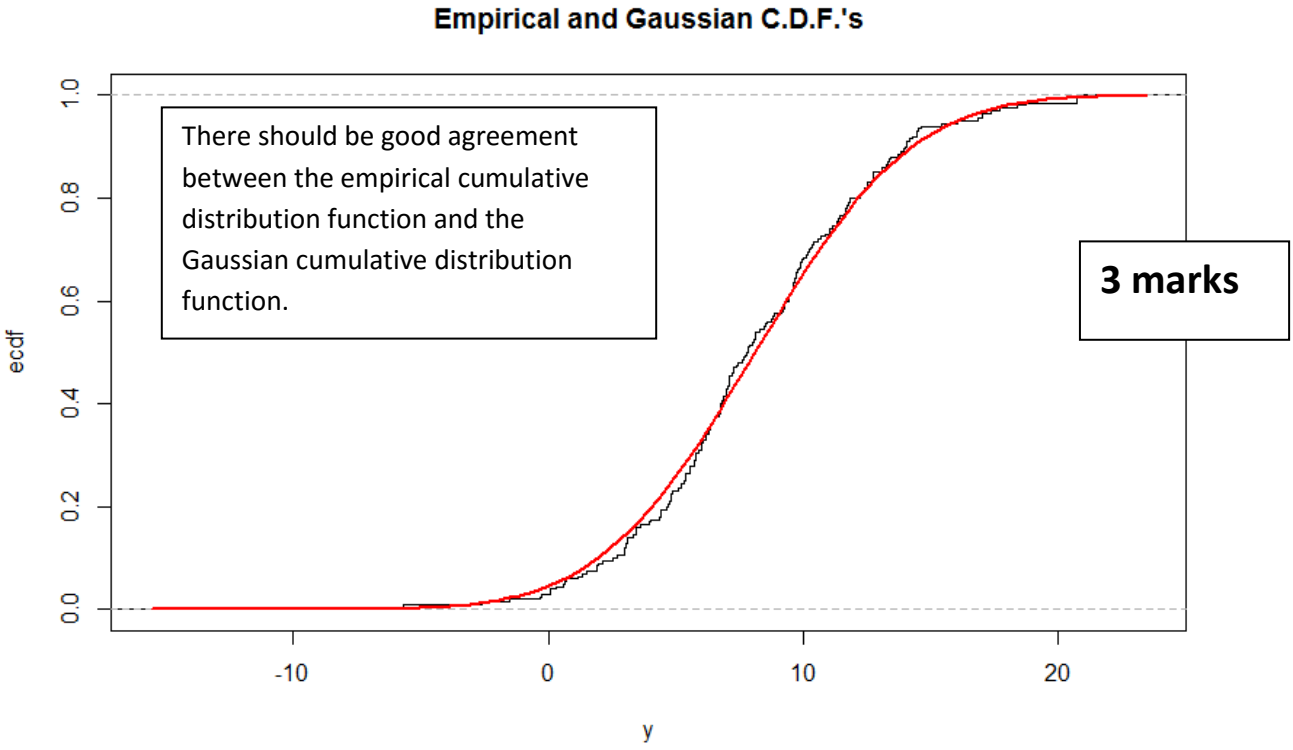
Sample skewness should be “close” to 0.

Sample kurtosis should be “close” to 3.

There should be good agreement between the relative frequency histogram and the Gaussian probability density function.



3 marks



There should be good agreement between the empirical cumulative distribution function and the Gaussian cumulative distribution function.

3 marks

<u>Problem 2:</u> The first five numbers in your Exponential data set are:

0.01	**0.13**	**0.18**	**0.24**	**0.26**

2 marks

Sample mean = 7.916900

Sample standard deviation = 9.249768

Sample mean should be larger than the sample median.

Sample mean and the sample standard deviation should be "close" in value.

The five number summary is:

0.010	**2.070**	**5.095**	**11.120**	**90.520**

2 marks

Sample median = 5.095

Range = 90.520 – 0.010 = 90.51

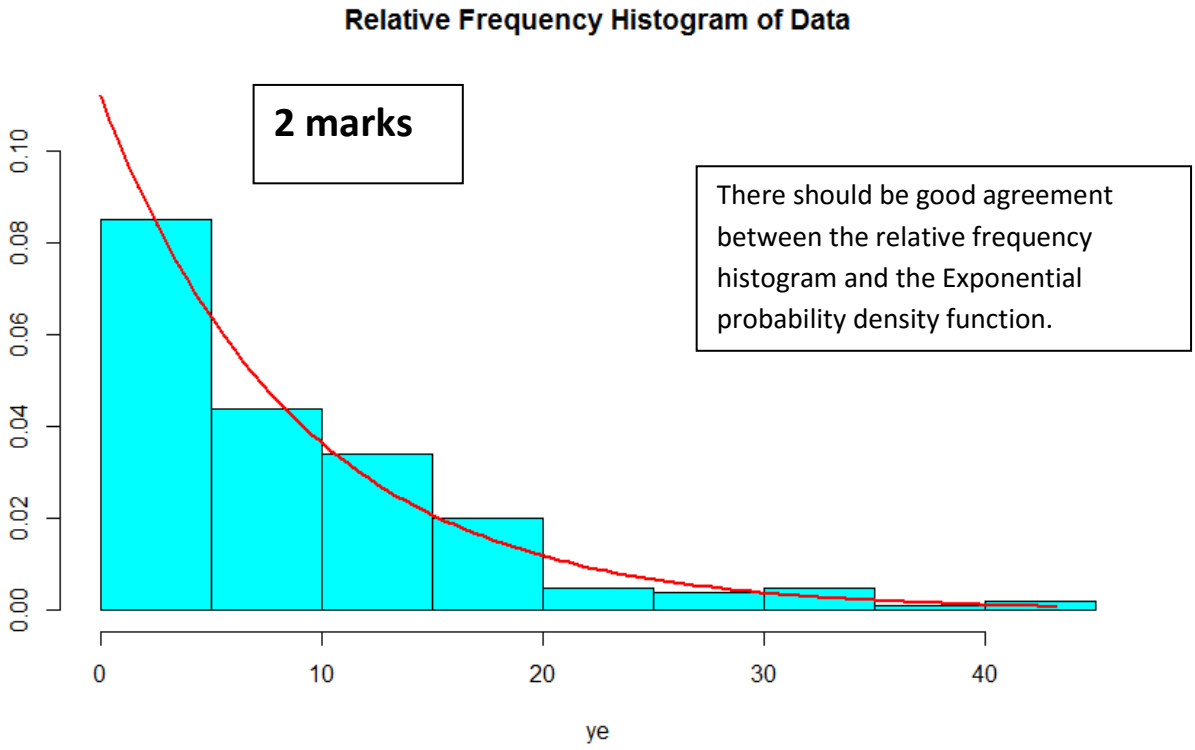
IQR = 11.120 – 2.070 = 9.05

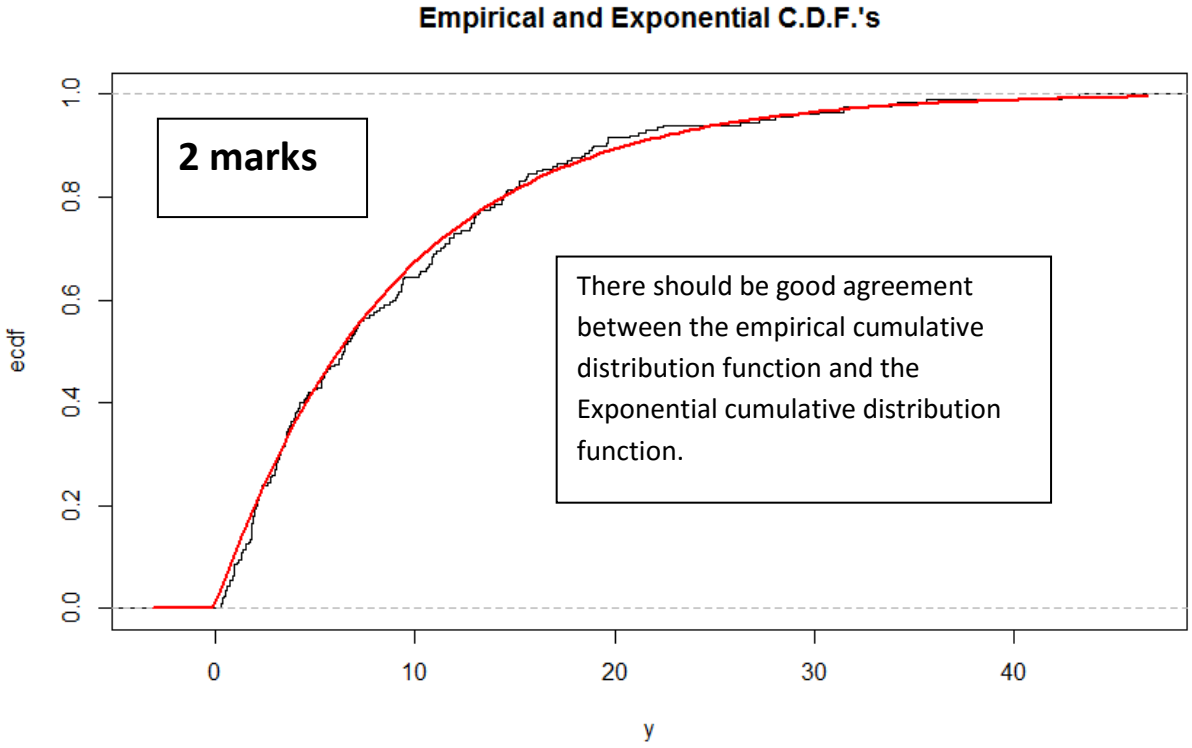
Sample skewness = 4.198336

Sample kurtosis = 33.82573

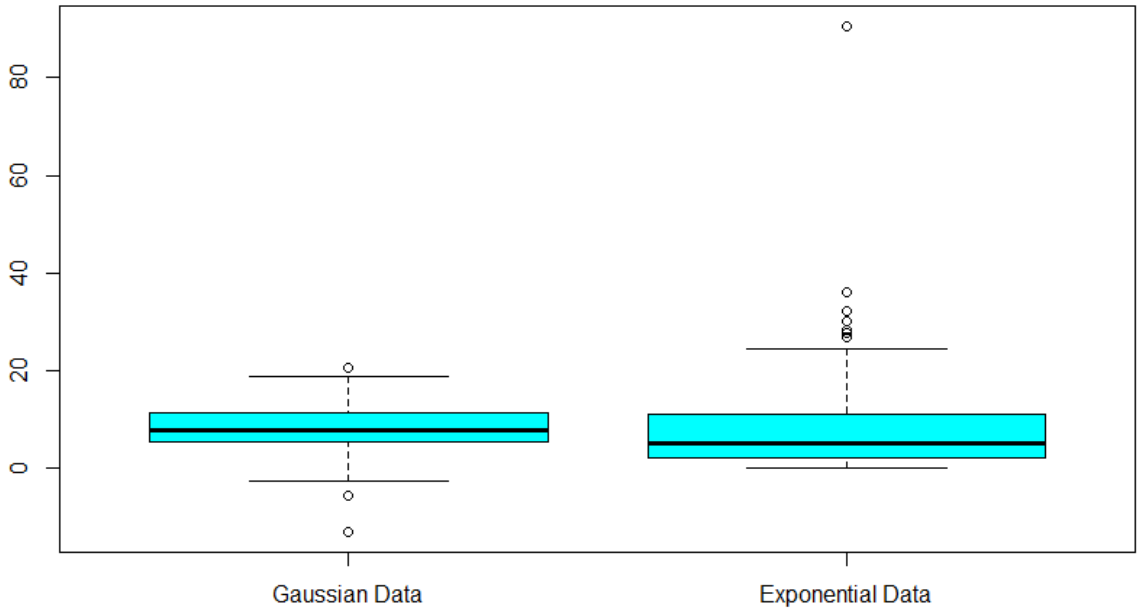
Sample skewness should be positive and not "close" to zero.

Sample kurtosis should be large.


Relative Frequency Histogram of Data
2 marks
There should be good agreement between the relative frequency histogram and the Exponential probability density function.
0.00
0.02
0.04
0.06
0.08
0.10
0
10
20
30
40
ye


Empirical and Exponential C.D.F.'s
2 marks
There should be good agreement between the empirical cumulative distribution function and the Exponential cumulative distribution function.
ecdf
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
y

2 marks



The median line in the Gaussian boxplot should cut the box and the whiskers approximately in half. There should be few outliers.

The boxplot for the Exponential data should have many outliers above the top whisker.

<u>Problem 3:</u> The first five numbers in your Gamma data set are:

1.32	**2.27**	**3.62**	**4.18**	**4.66**

Sample mean = 22.89415

1 mark

Sample standard deviation = 13.02463

The five number summary is:

1.320	**13.515**	**20.775**	**30.555**	**74.390**

Sample median = 20.775

1 mark

Range = 74.390 – 1.320 = 73.07

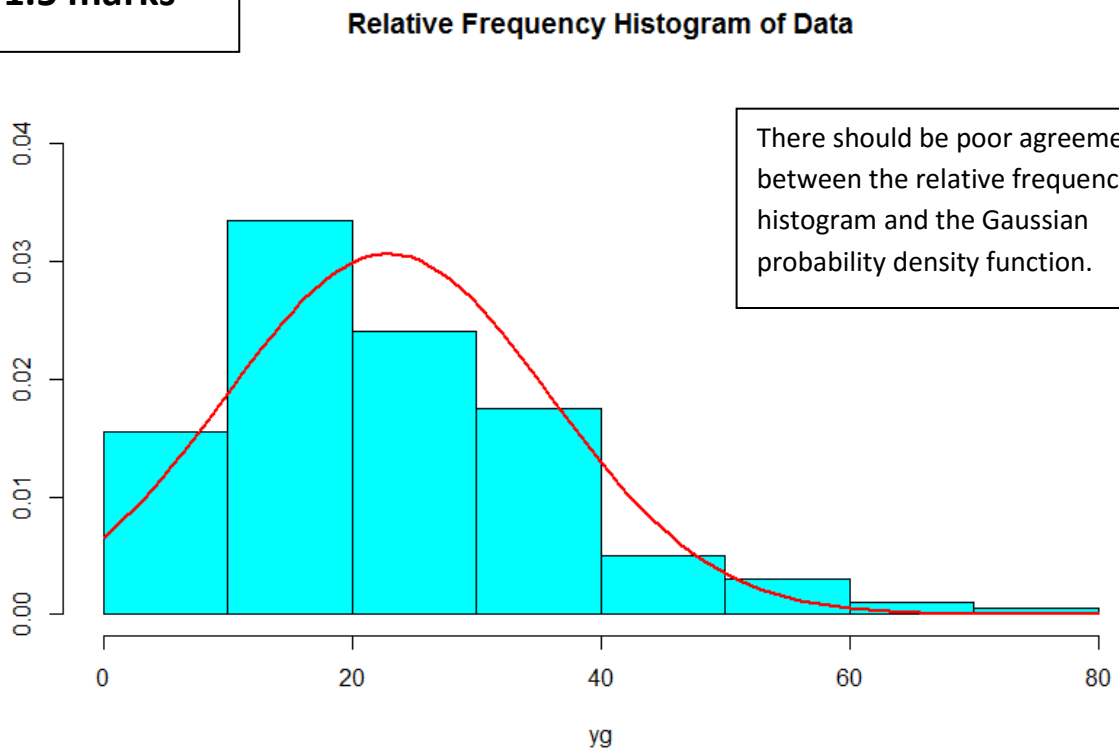
IQR = 30.555 – 13.515 = 17.04

Sample skewness = 0.9927479

Sample kurtosis = 4.116735

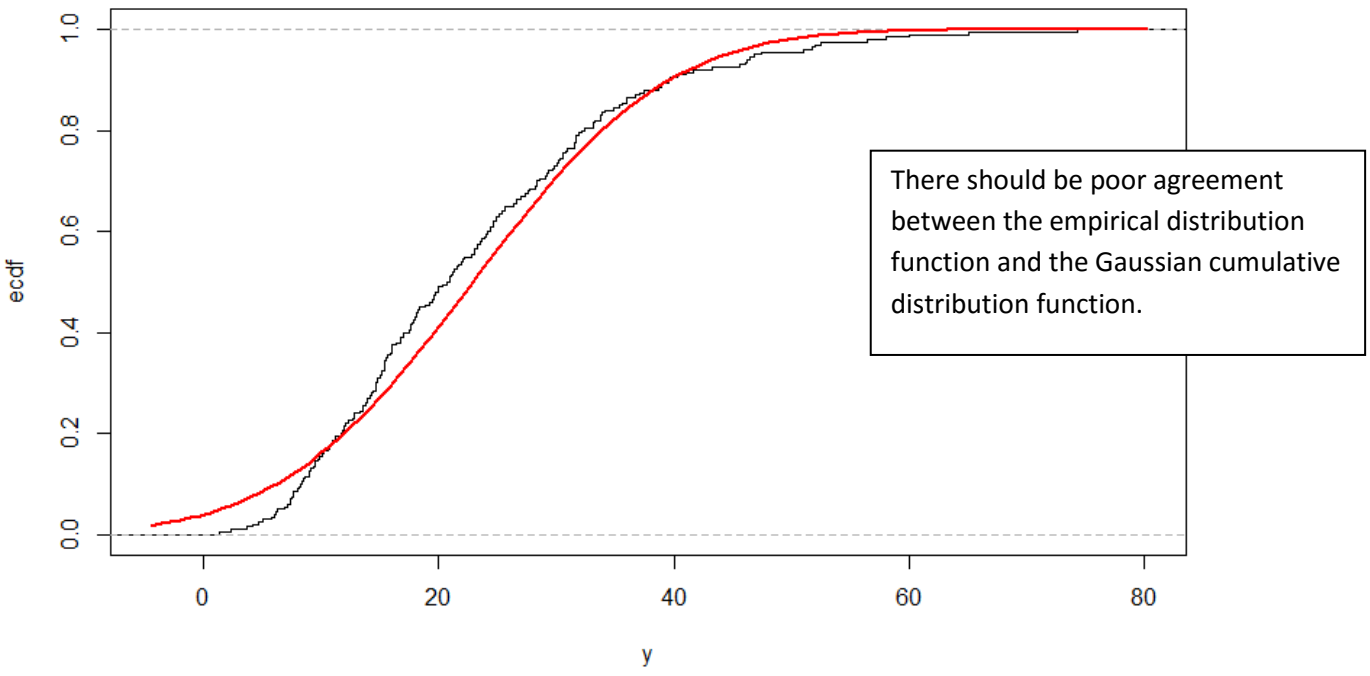
Sample skewness should be positive and the sample kurtosis should be larger than three.

1.5 marks



There should be poor agreement between the relative frequency histogram and the Gaussian probability density function.

1.5 marks



There should be poor agreement between the empirical distribution function and the Gaussian cumulative distribution function.

10 marks

This question is worth 10 marks. A complete answer must include at least five complete statements which support your conclusion.

For example:

statements about the mean, median, skewness and kurtosis for these data and how they compare with the values which are expected for Gaussian data

statements about the relative frequency histogram and the empirical cumulative distribution function and what you expect to see for Gaussian data.

Problem 4:

Alpha = 8.11465 **Beta = 7.9169**

The first five pairs of numbers in your bivariate data set are:

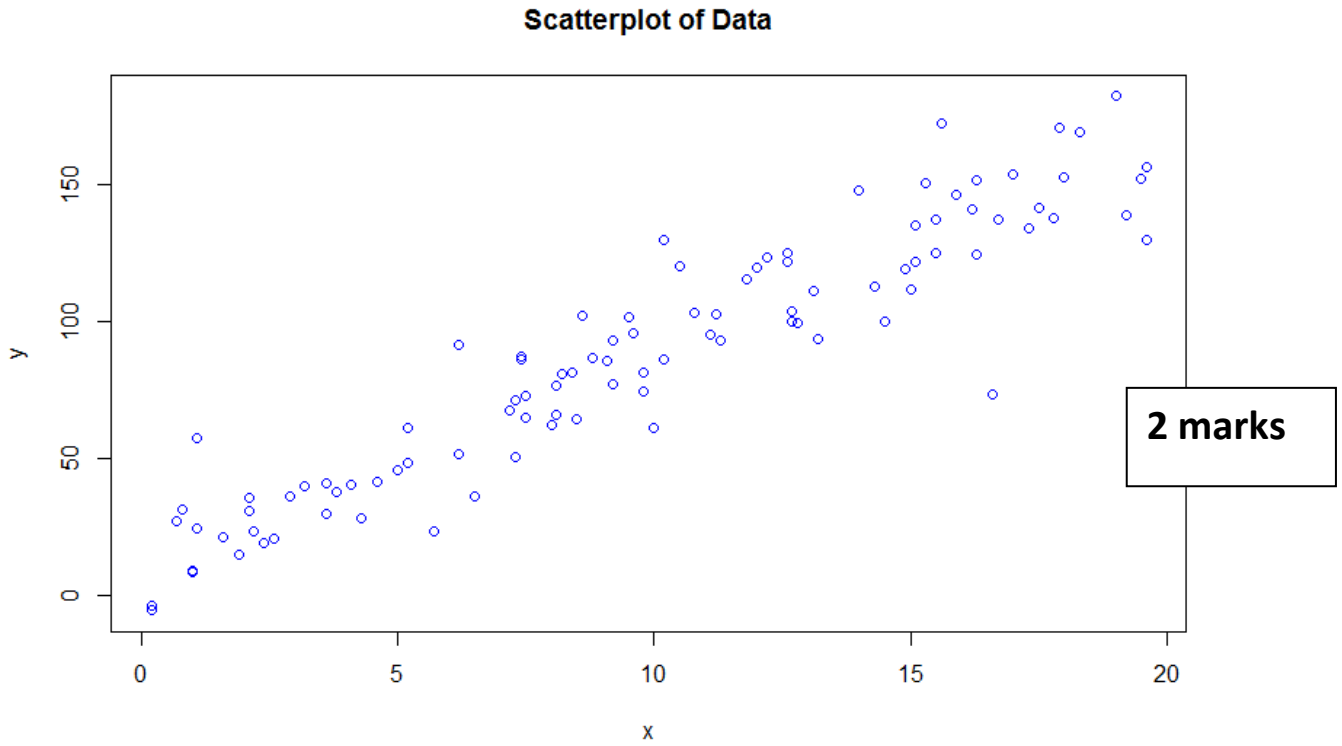
x	y
1.1	24.1
1.9	14.9
8.5	64.1
15.5	136.9
19.6	156.0

2 marks

Sample Correlation = 0.9365159

1 mark

Sample correlation should be "close" to 1 and data should lie along a straight line.



2 marks