Name:

EMPA PREPARATION

Jack's magic beans

On his way to market to sell the family's only cow, Jack met a stranger who offered him a handful of beans for the cow. These were no ordinary beans, the stranger said but 'magic' beans, magic in the sense of being 'bigger' than ordinary beans. Being interested in beans Jack exchanged them for the family's only cow. On returning home decided to statistically analyse his magic beans by comparing them with ordinary beans (he had no friends!). Jack weighed the mass of samples of each type of bean and then statistically analysed the results.
Answer the questions below.

Mass of beans /g	
Jack's 'magic' beans	**'Ordinary'beans**
10.3	9.8
11.2	8.2
9.9	7.6
10.7	8.3
11.0	7.6
9.9	7.9
11.2	8.3
10.6	8.4
10.4	8.5
10.8	7.6

Processing your data

5 Use a statistical test to analyse your data and test your null hypothesis. You may use a calculator and the Students' Statistical Sheet that has been provided.
You are provided with a sheet of graph paper. You may use this if you wish.

5 (a) State your null hypothesis.

.....

.....

.....

.....*(1 mark)*

5 (b) Give your choice of statistical test.

.....*(1 mark)*

5 (c) Give reasons for your choice of statistical test.

.....

.....

.....

.....*(1 mark)*

5 (d) Calculate the test statistic. *(1 mark)*

5 (e) Interpret the test statistic in relation to your null hypothesis. Use the words probability and chance in your answer.

.....

.....

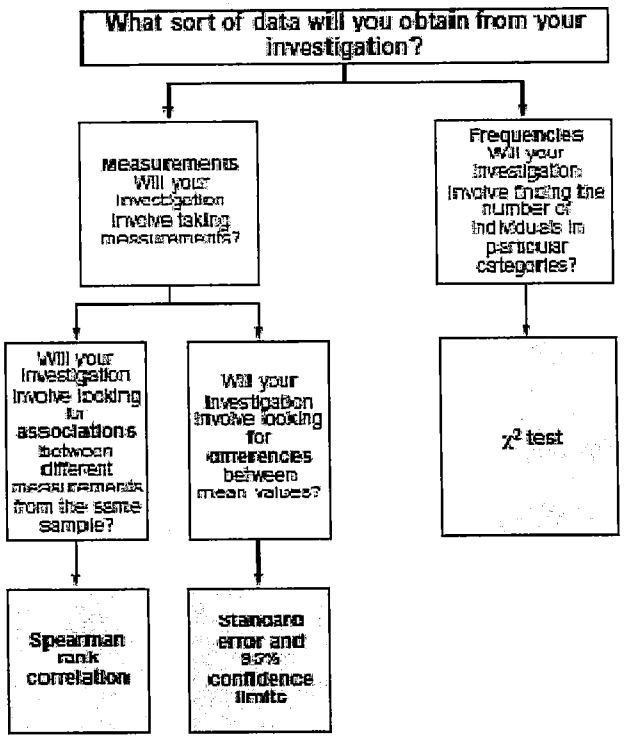
.....

.....

.....

..... *(2 marks)*

AQA Students' Statistics Sheet (version 3)



Standard error and 95% confidence limits

Calculate the standard error of the mean, SE, for each sample from the following formula

$$SE = \frac{SD}{\sqrt{n}}$$

where SD = the standard deviation
and n = sample size

95% confidence limits = 2 x SE above and below the mean

The χ^2 test

The chi square (χ^2) test is based on calculating the value of χ^2 from the equation

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

where O represents the results you observe in the investigation
and E represents the results you expect.

Table showing the critical values of χ^2 at P = 0.05 for different degrees of freedom

Degrees of Freedom	Critical value
1	3.84
2	5.99
3	7.82
4	9.49
5	11.07
6	12.59
7	14.07
8	15.51
9	16.92
10	18.31

Spearman rank correlation test

Calculate the value of the Spearman rank correlation, r_s, from the equation

$$r_s = 1 - \left[\frac{6 \times \Sigma D^2}{n^3 - n}\right]$$

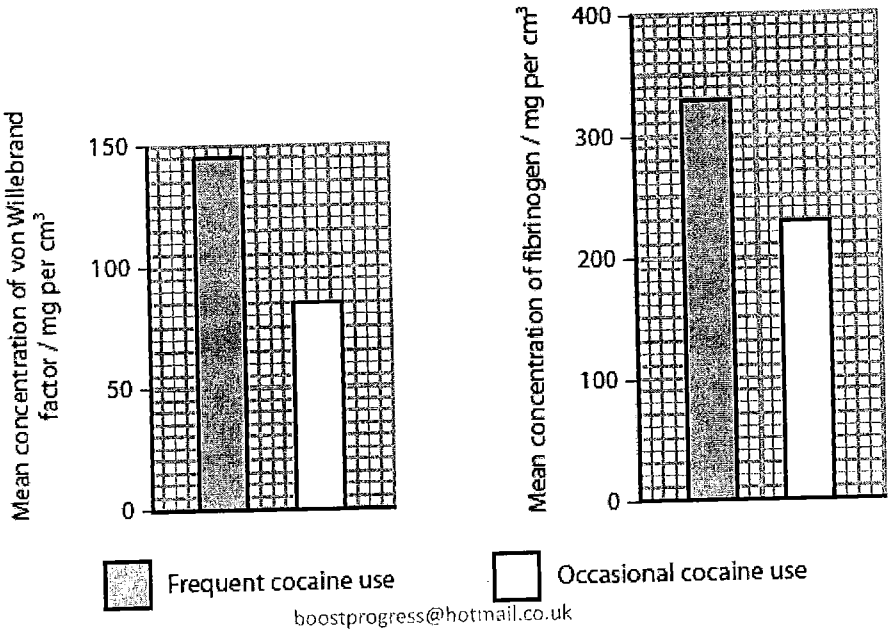
where n is the number of pairs of items in the sample and D is the difference between each ranked pair of measurements.

Table showing the critical values of r_s at P = 0.05 for different numbers of paired values

Number of pairs of measurements	Critical value
5	1.00
6	0.89
7	0.79
8	0.74
9	0.68
10	0.65
12	0.59
14	0.54
16	0.51
18	0.48

Cocaine use increases the risk of a heart attack.
Cocaine also affects the levels of a number of blood components, including von Willebrand factor and fibrinogen. These two components are involved in blood clotting.

(a) The normal range for von Willebrand factor is 50 to 150 mg per cm^3 and for fibrinogen is 150 to 300 mg per cm^3.
The graphs below show the effects of frequent and occasional cocaine use on the mean concentration of von Willebrand factor and fibrinogen in the blood.



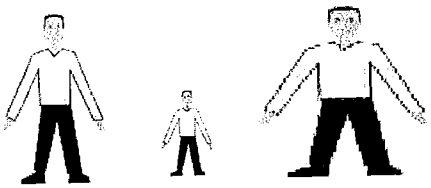
Cocaine

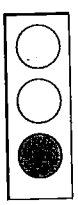
a) Why were both the independent variables: factor and fibrinogen measured in the same units?

b) How would you describe the dependent variable?

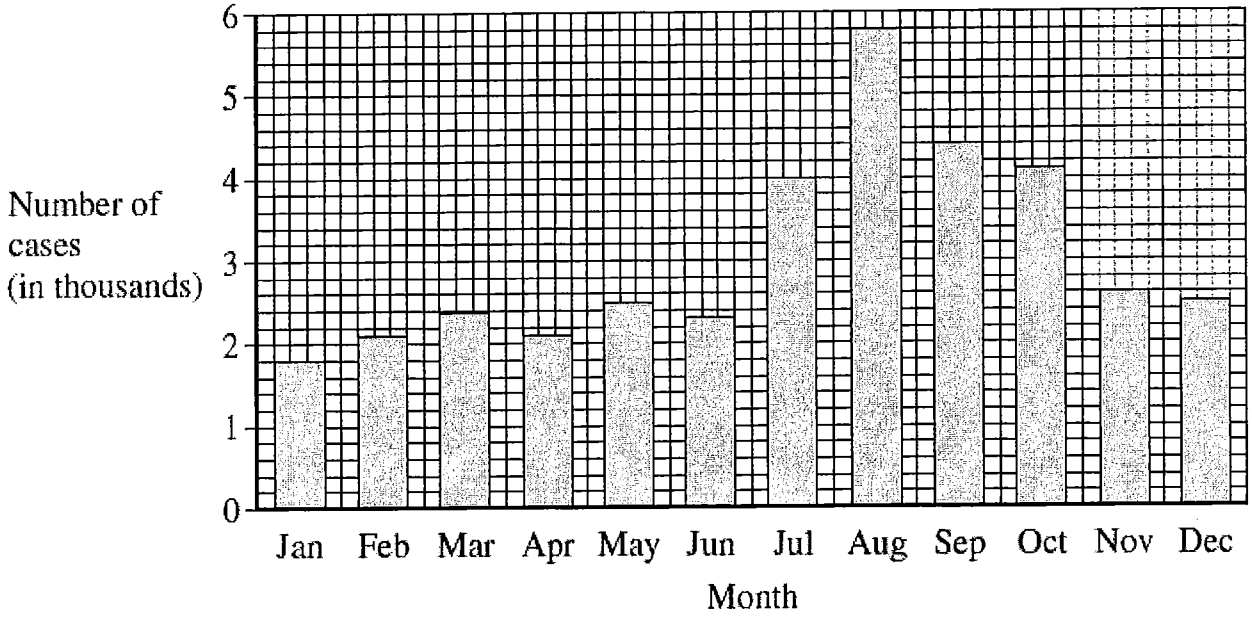
c) Why might the independent variable invalidate the conclusions of this study?

d) What is one major criticism of the graph presented and how would you address this criticism?





Giardiasis is one of the main causes of diarrhoea in the USA. It is usually transmitted by drinking contaminated water. The bar chart shows the number of cases of giardiasis in one state of the USA during one year.



Giardiasis

a) Calculate the percentage change between January to April, May to August and September to December.

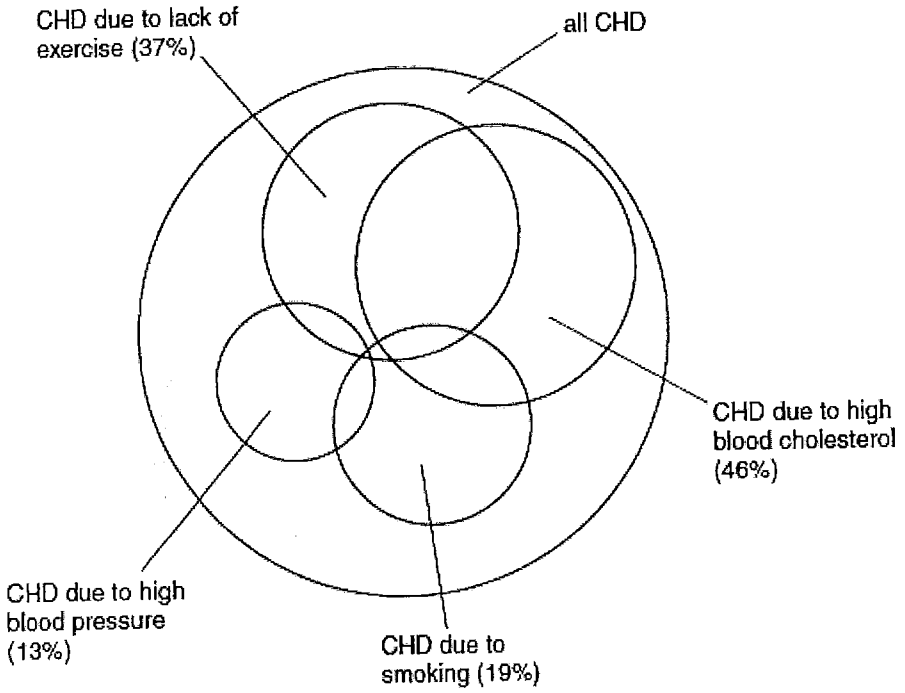
b) Do these different percentage change values accurately reflect the data?

c) Why is percentage change a useful measure when used in this way?

Difference
Original **x 100 =**

Coronary heart disease (CHD) can be described as a multifactorial disease. This means that a number of different risk factors contribute to the development of the disease.

This shows the percentage of cases of CHD in a population to which each risk factor contributed.



CHD

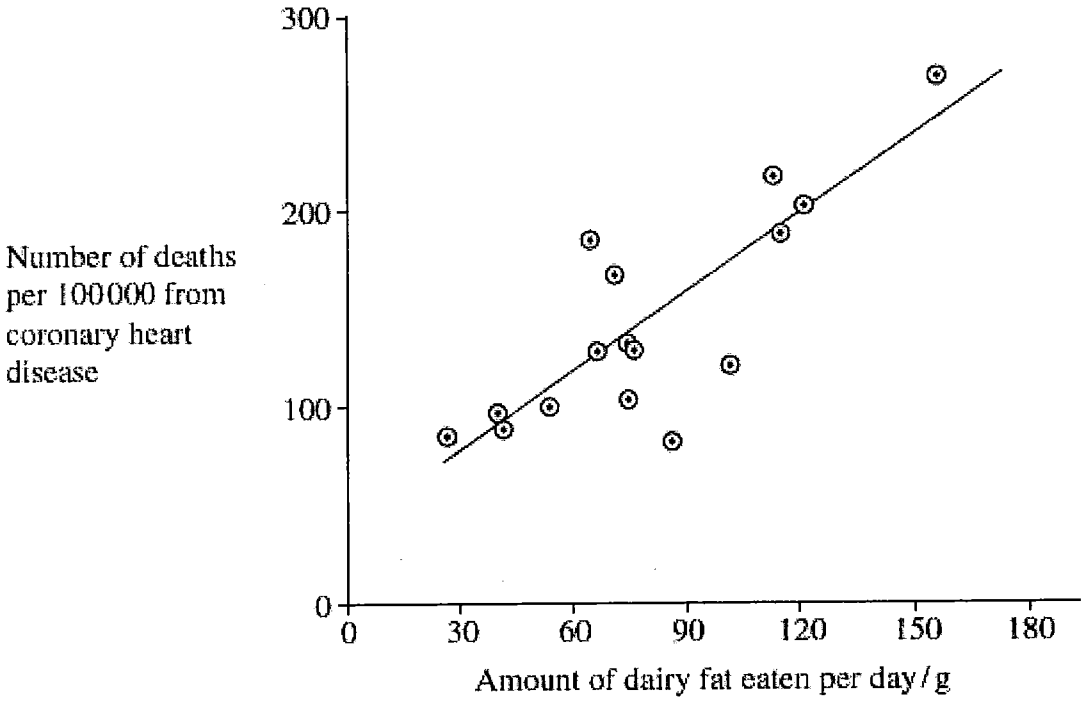
a) What is the advantage of representing data as a Venn diagram?

b) What further information would you want to know about the population sampled in this study?

c) There is a fundamental flaw in the data presented. What is it?

d) How might cigarette produces use this data for their marketing campaign?

The graph shows the relationship between the amount of dairy fat eaten and the deaths from coronary heart disease (CHD) in different countries.



Dairy Fat

a) Why is a line of best fit justified for representing the correlation of this data?

b) Why is the number of deaths expressed at *per 100,000…?*

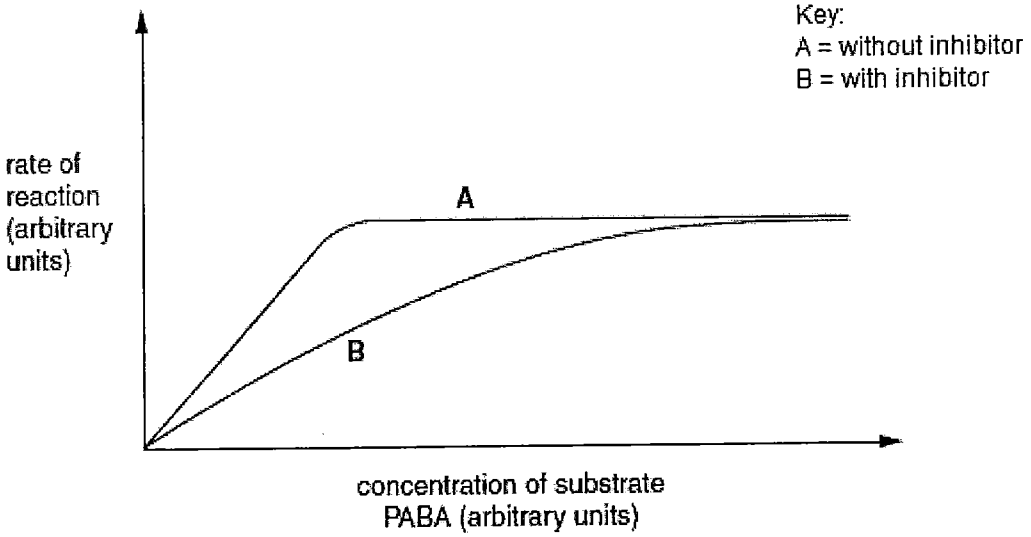
c) Is there any evidence of reliability in this study?

d) What further questions might you ask to fully assess reliability?

boostprogress@hotmail.co.uk

The graph shows the effect of increasing the concentration of the substrate (PABA) on the rate of reaction.

- Curve A shows the rate of reaction without the presence of the competitive inhibitor sulfonamide.
- Curve B shows the rate of reaction in the presence of the competitive inhibitor sulfonamide.



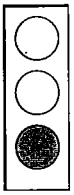
PABA

a) What is meant by arbitrary units?

b) Why is it sometimes useful to use arbitrary units?

c) Compare the patterns shown by graph?

(Don't explain them!!!!)

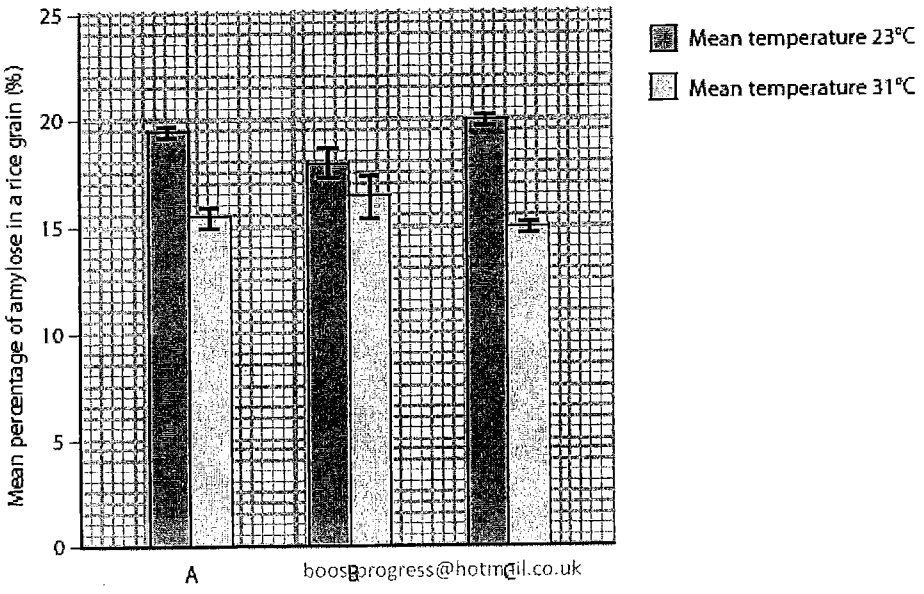


Starch is made up of amylose and amylopectin. An investigation was undertaken to study the effect of temperature on amylose production in rice grains. Three different varieties of rice plant, A, B and C, were grown at a mean temperature of 23°C until they had produced mature rice grains. All other variables were kept constant.

Fifty rice grains were then collected from each variety and the mean percentage of amylose in a rice grain was determined.

This investigation was repeated at a mean temperature of 31°C.

The results are shown in the graph below.



Amylose

a) How were the standard error bars calculated?

b) Which of the three species had the most variable amylose content? How do you know this?

c) Describe the main pattern in the results for each species.

d) Why was percentage of amylose used to measure the amylose content?