Worksheet 3 – Paper 2

Question 1
Explain the difference between Numerical discrete data and numerical continuous data.

Question 2
Explain the process of stratified random sampling?

Question 3
The weight of 100 first year students' bags was measured at the beginning of a school year. The weight of the bags was normally distributed.
The average weight of the school bags was found to be 12kg with a standard deviation of 3kg.

(a)

(i) Simon is a first year student and is not strong enough to carry a bag heavier than 14.5kgs. Find the probability that Simon will not be able to carry his bag.

(ii) Sorcha's bag is heavier than 72% of all of the other students. How much does Sorcha's bag weigh?

(iii) Simone's bag is lighter than 72% of the other students. How much does Simone's bag weigh?

(iv) Joseph is a first year student. Joseph was absent from school the day the study was carried out. Construct a 95% confidence interval to describe the probable weight of Joseph's bag.

(b)

(i) At the end of the school year the weight of 100 first year students bags was measured again. The average weight of the bags was 11.5kg with a standard deviation of 3kg. Carry out a hypothesis test at the 5% level of significance to investigate if the weight of the bags has changed throughout the school year.

(ii) Find the p-value for this test and explain what it means in the context of the question.

(c)
The Principal wants to carry out a similar study on the weight of second year students' bags. However the weight of second year students' bags is skewed.

(i) If the mean weight of the second year students' bags was 10.5 kg and the mode was 9kg, draw a sketch of the distribution of weight of the second year students' bags.

(ii) Explain how the central limit theorem could be used to generate a normal distribution from this data.

Q4 The length of time it takes for 100 students to travel to school in County Mayo was recorded.

(10 – 20 means more than 10 minutes and $\leq$ 20 minutes)

Time in minutes	≤ 10	10 – 20	20 – 30	≥ 30
Number of students	20	45	30	X

(i) Find the % of students who spend 30 minutes or more travelling to school
(ii) Illustrate this information on a pie chart.
(iii) Kyle wants to illustrate this information on a bar chart explain why he cannot do this.
(iv) In which interval does the median lie?

Q5 A politician claims that support for his party is at an all-time high of 33%. A survey of 10000 constituents found that 3350 of them supported the politicians party. Test the hypothesis at the 5% level of significance that the politicians claim is false.

Q6 (2013 P2 Q2)

The heights of a certain type of plant, when ready to harvest, are known to be normally distributed, with a mean of μ. A company tests the effects of three different growth hormones on this type of plant. The three hormones were used on a different large sample of the crop. After applying each hormone, it was found that the heights of the plants in the samples were still normally distributed at harvest time.

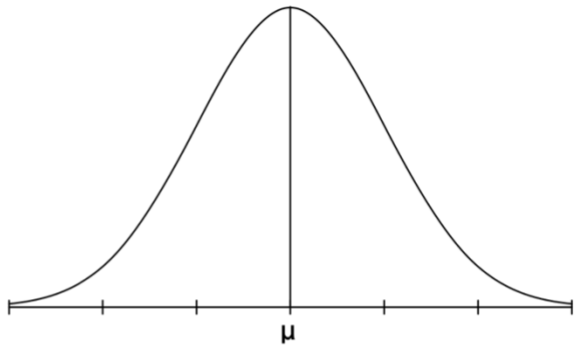
The diagrams A, B and C, on the next page, show the expected distribution of the heights of the plants, at harvest time, without the use of the hormones.

The effect, on plant growth, of each of the hormones is described on the next page. Sketch, on each diagram, a new distribution to show the effect of the hormone.

Hormone A

The effect of hormone A was to increase the height of all of the plants.

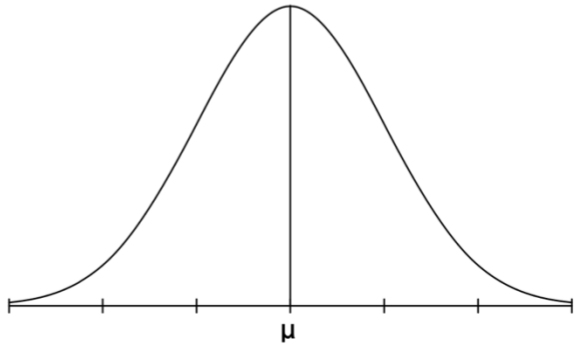
Diagram A



Hormone B

The effect of hormone B was to reduce the number of really small plants and the number of really tall plants. The mean was unchanged.

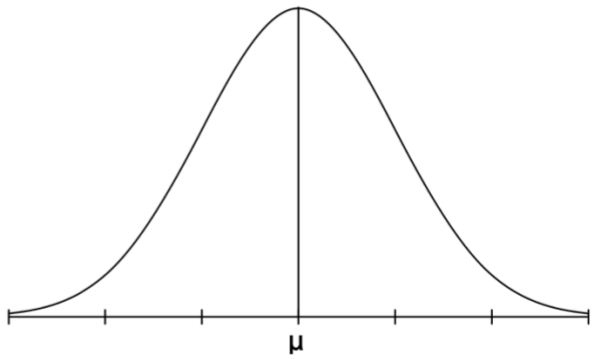
Diagram B



Hormone C

The effect of hormone C was to increase the number of small plants and the number of tall plants. The mean was unchanged.

Diagram C

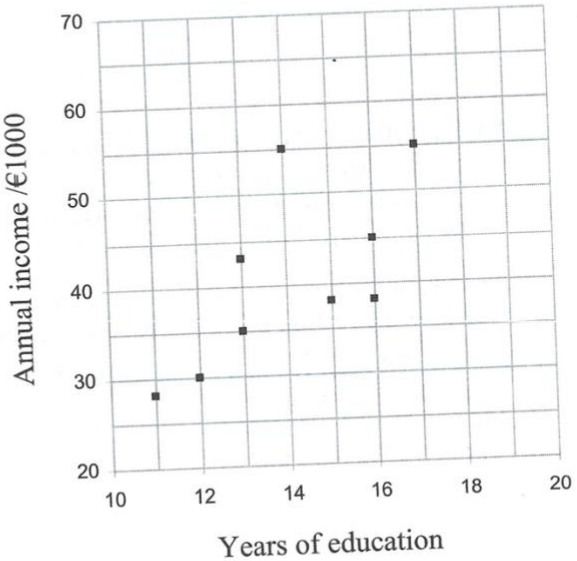


Q7 (2012 SEC Sample P2 Q7)

stion 7

An economics student wants to find out whether the length of time people spend in education affects the income they earn. The student carries out a small study. Twelve adults are asked to state their annual income and the number of years they spent in full-time education. The data are given in the table below, and a partially completed scatter plot is given.

Years of education	Income /€1,000
11	28
12	30
13	35
13	43
14	55
15	38
16	45
16	38
17	55
17	60
17	30
19	58



(i) The last three rows of data have not been included on the scatter plot. Insert them now.

(ii) Calculate the correlation coefficient.

(iii) What can you conclude from the scatter plot and the correlation coefficient?

(iv) Add the line of best fit to the scatter plot above.

(v) Use the line of best fit to estimate the annual income of somebody who has spent 14 years in education.

(vi) By taking suitable readings from your diagram, or otherwise, calculate the slope of the line of best fit.

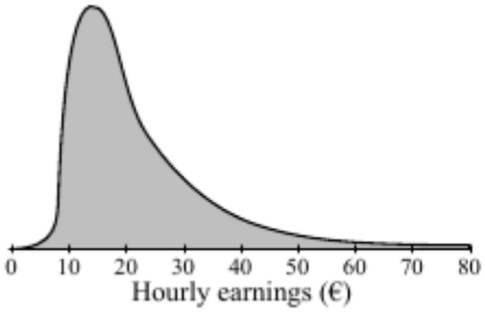
(vii) Explain how to interpret the slope in this context.

(viii) The student collected the data using a telephone survey. Numbers were randomly chosen from the Dublin area telephone directory. The calls were made in the evenings, between 7 and 9pm. If there was no answer, or if the person who answered did not agree to participate, then another number was chosen at random.

List **three** possible problems regarding the sample and how it was collected that might make the result of the investigation unreliable. In each case, state clearly why the issue you mention could cause a problem.

(b) The distribution of the hourly earnings of all employees in Ireland in October 2009 is shown in the diagram. It can be seen that the distribution is positively skewed.

The mean is €22.05.
The median is €17.82.
The standard deviation is €10.64
The lower quartile is €12.80
The upper quartile is €26.05



(Source: adapted from: CSO. *National Employment Survey 2008 and 2009*)

(i) If six employees are selected at random from this population, what is the probability that exactly four of them had hourly earnings of more than €12.80?

Q8

The average lifespan of 100 mayfly in Europe was found to be normally distributed with an average of 24 hours and a standard deviation of 38 minutes.

(i) Construct a 95% confidence interval for the lifespan of a mayfly.

(ii) Find the probability that a mayfly will live for 25 or more hours?

(iii) Mary the Mayfly lived a shorter life than 96% of all mayfly. Find, to the nearest minute, how long Mary the mayfly lived?

(iv) A sample of 5 mayfly was observed in the wild, find the probability that their average lifespan will be $\geq$1470 minutes.

(v) 20 mayfly in North America were found to have an average lifespan of 1420 minutes with a standard deviation of 38 minutes.
Test the hypothesis at the 5% level of significance that the average lifespan of mayfly in North America is different from that of a mayfly in Europe.

Q9 Find:

(i) P(-1.96 $\leq Z \leq$ 1.96)
(ii) P(Z $\leq$ 1.04)
(iii) P(Z $\geq$ 1.04)
(iv) P(0 $\leq Z \leq$ 0.1)
(v) P(0.1$\leq Z \leq$ 0.2)
(vi) P(-1.2 $\leq Z \leq$ - 0.5)

Q10

The average result in a Maths test was normally distributed with an average result of 69% with a standard deviation of 8%.
500 students took this exam.

(a) Use the empirical rule to estimate the number of students who scored :
(i) Between 53% and 77%
(ii) Over 77%

(b) Ues Z scores to find the number of students who scored:
(i) Between 53% and 77%
Over 77%