Subject: Leaving Certificate Maths

Teacher: Mr Murphy

Lesson 13: Statistics II

13.1 Learning Intentions

After this week's lesson you will be able to;

- Describe the measures of central tendency.
- Explain when to use each of the measures of central tendency.
- Describe and calculate the measures of variation of a set.
- Describe data relative to the sample (percentiles).
- Describe the normal distribution and empirical rule
- Comment on the standard deviations of a distribution.
- Calculate margin of error.

13.2 Specification

Students learn about	Students working at OL should be able to	In addition, students working at HL should be able to
1.7 Analysing, interpreting and drawing inferences from data	– recognise how sampling variability influences the use of sample information to make statements about the population – use appropriate tools to describe variability drawing inferences about the population from the sample – interpret the analysis and relate the interpretation to the original question – interpret a histogram in terms of distribution of data – make decisions based on the empirical rule – recognise the concept of a hypothesis test – calculate the margin of error ($\frac{1}{\sqrt{n}}$) for a population proportion* – conduct a hypothesis test on a population proportion using the margin of error	– build on the concept of margin of error and understand that increased confidence level implies wider intervals – construct 95% confidence intervals for the population mean from a large sample and for the population proportion, in both cases using z tables – use sampling distributions as the basis for informal inference – perform univariate large sample tests of the population mean (two-tailed z-test only) – use and interpret p-values
	Numerical – recognise standard deviation and interquartile range as measures of variability – use a calculator to calculate standard deviation – find quartiles and the interquartile range – use the interquartile range appropriately when analysing data – recognise the existence of outliers	**Numerical** – recognise the effect of outliers – use percentiles to assign relative standing

13.3 Chief Examiner's Report

Section	Q	Mean Mark	Mean Mark (%)	Mark Ranking (Paper)	Main Topic
A	1	21·6	84	1	Probability
A	2	14·8	59	6	Inferential statistics

13.4 Central Tendency

These measures are commonly known as AVERAGES

Mean

Add up all numbers
Divide by the amount of numbers

ONLY works with Numerical Data

Uses All the data

Very sensitive to outliers

Median

Middle Value
Data in order

ONLY works with Numerical Data

Can be used With outliers

Mode

Only option for Categorical Data

Can be used with any data type

Easy to find

Not always a mode*

13.5 Measure of Variation

Mean

Add up all numbers
Divide by the amount of numbers

ONLY works with Numerical Data

Uses All the data

Very sensitive to outliers

Median

Middle Value
Data in order

ONLY works with Numerical Data

Can be used With outliers

Mode

Only option for Categorical Data

Can be used with any data type

Easy to find

Not always a mode*

Measures of central values only give us an one description of the data. They don't give any information on the spread of the data.

Range

Range = Max - Min

Interquartile Range

I.Q.R. = $Q_3 - Q_1$

3 Quartiles:
Lower
Median
Upper

Standard Deviation

Average spread from the mean

Caters for all values in set unlike range or I.Q.R.

$$\sigma = \sqrt{\sum \frac{(x - \bar{x})^2}{n}}$$

$\bar{x}$ = mean

Standard Deviation

The below data set is the cost of a packet of multivitamin tablets:

Roots = €28.99
VitaBita= €27.99
Nurture's Hay = €29.99
X= €30.50
The health shore= €31.29

$$\sigma = \sqrt{\sum \frac{(x-\bar{x})^2}{n}}$$

x	$\bar{x}$	Difference ($d = x - \bar{x}$)	Difference Squared ($d^2 = (x - \bar{x})^2$)
28.99	29.75	-0.76	0.58
27.99	29.75	-1.76	3.1
29.99	29.75	0.24	0.06
30.50	29.75	0.75	0.56
31.29	29.75	1.54	2.37
			6.67

13.6 Percentile

Another analysis tool we have is to compare the data points to the other points in the set.

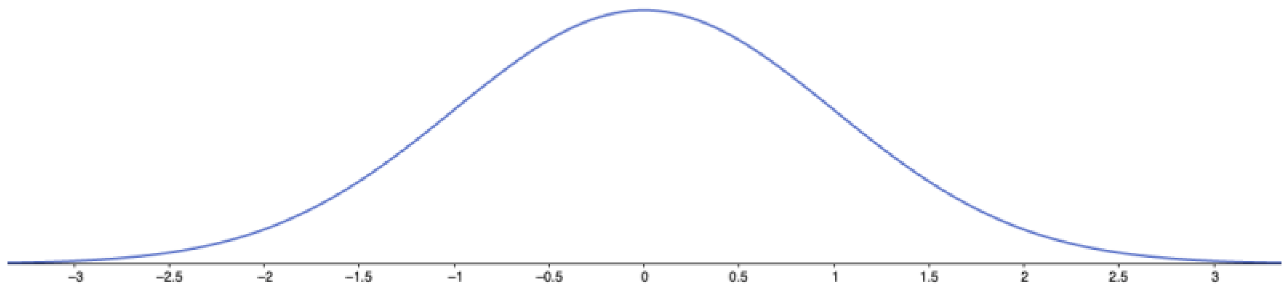
This method divides the data into 100 equal parts
$P_1, P_2, \ldots\ldots P_{99}$
Each percentile represents the percentage data points below that score.

Calculate a value for P_k

$$c = \frac{n \times k}{100}$$

1) If c is a whole number, choose that data point and the next one, divide their sum by 2.
2) If c is not whole, then round value to nearest whole number.
3) The value obtained from either step **1 or 2** is the Percentile

13.7 Normal Distribution & Empirical Rule



In order to work with a normal distribution our data set **must** be normal. Once that is the case, we can take a data point we are interested in and turn it into something called a z-score. This allows us to analyse the data in terms of probability.

To get a z-score we can use the following:

$$z = \frac{\bar{x} - \mu}{\sigma}$$

Once we have a z-score we can then find the probability of a particular event. For example. Take a sample of people's heights. These are normally distributed. If I want to know what is the probability that a person chosen at random will have a height between two values, less than a particular value etc. I can do this using z-scores. To do this we use our Tables book (pages 36 and 37).

A random variable X follows a normal distribution with a mean of 60 and standard deviation of 5.

i) Find $P(X \leq 68)$

ii) Find $P(52 \leq X \leq 68)$

13.8 Confidence Intervals & Margin of Error (Proportion)

In 13.7 we were dealing with proportions or fractions of a sample. For example, 40% of this sample have brown hair. Sometimes, we can be asked to deal with the means of samples. In questions of this type we will be given information of the mean and not the proportion. This changes our interval make up ever so slightly.

The main change is in the error and the fact we are now dealing with a mean. So we do not know the population mean (μ) we only know the sample mean $(\bar{x})$ so if we are trying to make an interval around the population mean it would look something like this:

$\bar{x} - 2\sigma \leq \mu \leq \bar{x} + 2\sigma$

We also have the change to our error, which in this case is:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

So now factoring in the more accurate 1.96 we have an interval for the population mean that looks like this:

$$\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$$

Question:

In a random sample of 500 eye-phones, a group of scientists were testing how many times could they drop the phone from a set height until the screen cracked. The mean number of times was 258 and there was a standard deviation of 13.

i) Construct a 95% confidence interval for the mean number of drops for any eye-phone before the screen cracked

13.9 Recap of the Learning Intentions

After this week's lesson you will be able to;

- Describe the measures of central tendency.
- Explain when to use each of the measures of central tendency.
- Describe and calculate the measures of variation of a set.
- Describe data relative to the sample (percentiles).
- Describe the normal distribution and empirical rule
- Comment on the standard deviations of a distribution.
- Calculate margin of error.

13.10 Homework Task

Question: **60 marks**

Acme Confectionery makes cakes and chocolate bars.

a. (i) Acme Confectionery has launched a new bar called Chocolate Crunch. The weight of these new bars are normally distributed with a mean of 4.65 g and a standard deviation of 0.12 g. A sample of 10 bars is selected at random and the mean weight of the sample is found.

Find the probability that the mean weight of the sample is between 4.6 g and 4.7 g.

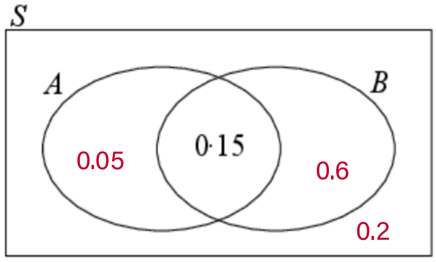
(ii) A company surveyed 400 people, chosen from the population of people who had bought at least one Chocolate Crunch bar.

Of those surveyed, 324 of them said they liked the new bar.
Create teh 95% confidence interval for the population proportion who liked the new bar.
Give your answer correct to 2 decimal places.

13.11 Solutions to 12.15

Two events A and B are such that P(A)=0.2, P(A∩B)=0.15 and P(Á∩B)=0.6.

a. Complete thsi Venn diagram



b. Find the probability that neither A nor B happenps. 0.2

c. Find the conditional probability P(A|B).

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A|B) = \frac{0.15)}{0.75} = 0.2$$

d. State whether A and B are independent events and justify your answer.

$$Yes\ they\ are\ independent\ as, P(A|B) = P(A) = 0.2$$