STATISTICS 11 AUGUST 2014

Lesson Description

In this lesson we:

- Revise the concepts learnt in Grade 11 Statistics
- Discuss the concepts of scatterplots, line of best fit, interpolation, extrapolation and correlation coefficients.
- Do examples to practise these concepts.

Summary

Grade 11 Revision

Measures of central tendency

- **Mode:** The most common value in a data set.
- **Median (Q_2):** The median is the middle value in an ordered data set.
- **Mean ($\bar{x}$):** The average of a data set.

Five Number Summary

- **Minimum Value in the set:** The lowest value in a data set.
- **Lower Quartile (Q_1):** The median of the lower half of an ordered data set.
- **Median (Q_2):** The median is the middle value of an ordered data set.
- **Upper Quartile (Q_3):** The median of the upper half of an ordered data set.
- **Maximum Value in a set:** The highest value in a data set.

The Five Number Summary is represented visually on a box and whisker diagram.

Measures of Dispersion

- **Range:** The difference between the largest value and smallest value in the data set.

Range = largest value – smallest vaue.

- **Inter-quartlie Range (IQR):** The difference between the upper and lower quartiles.

IQR = $Q_3 - Q_1$

Calculation of the Mean

- It is commonly known as the average.
- Mean =

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{sum\ of\ the\ values\ in\ the\ data\ set}{number\ of\ elements\ in\ the\ data\ set}$$

Ogives

- An ogive is a graph which represents the **Cumulative Frequency** of a set of data.
- The graph is usually an s-shape and is a smooth curve (not joined by straight lines)

Variance and Standard Deviation

- Standard Deviation is representation of how much variation or dispersion there is around the mean.
- A low standard deviation shows that the data points lie close to the mean and the data set is generally consistent.
- $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$

Scatter Plots

- Regression Lines
- Least Squares Regression Line
- The Correlation Coefficient

Test Yourself

Lerato and Jane are Grade 12 learners who both take IT as a subject. In order to earn extra money they decide to write an Andriod App. The table below shows the number of sales, in thousands, which similar applications have made at various selling prices.

Price (R)	1	4	5	7	10	11	15	18
Number of sales (in thousands)	20	16	12	11	8	4	2	1

Question 1

Determine the equation of the line of best fit for the data.

a) $y = 19{,}39x - 1{,}14$ b) $y = 1{,}14x - 19{,}39$

c) $y = -1{,}14x + 19{,}39$ d) $y = -11{,}4x - 19$

Question 2

Determine the mean point

a) $(8{,}875; 9{,}25)$ b) $(9; 9)$ c) $(9{,}25; 8{,}875)$ d) (8;9)

Question 3

If they set the selling price to be R8,24, estimate the number of sales they will make. Give your answer correct to the nearest thousand.

a) 1000 b) 10 000 c) 100 000 d) 1000 000

Question 4

At what approximate price will no sales be made?

a) R18 b) R17 c) R50 d) R25

The table below shows the amount of money spent on advertising and the income of the company in thousands of Rand, over a 6 month period.

Month	1	2	3	4	5	6
Advertising	3	4.5	1	5	7	2.4
Omcome	32	56	18	48	60	25

Question 5

Determine the mean point of the data.

a) (39,83;3,82) b) (4;40) c) (40;4)

d) $(3{,}82;\ 39{,}83)$

Question 6

Determine the equation of the least squares regression line.

a) $y = 7{,}69x + 10{,}49$ b) $y = -7{,}69x - 10{,}49$

c) $y = 8x + 11$ d) $y = 6{,}345x + 23{,}45$

Question 7

Determine the correlation coeffiencient of the data.

a) $0{,}45$ b) $0{,}59$ c) $0{,}95$ d) $0{,}99$

Question 8

The owner of the company states that there is a strong relationship between the advertising expenditure and company's income. Explain this statement with reference to the correlation coefficient.

a) the correlation coefficient is a negative value very close to -1

b) the correlation coefficient is a negative value very close to -0.5

c) the correlation coefficient is a positive value very close to 1

d) the correlation coefficient is a positive value very close to 0.5

Question 9

Predict the company's income in a month where R3 500 is spent on advertising.

a) $R7\ 405$ b) $R30\ 134$ c) $R3\ 405$

d) $R37\ 405$

Question 10

Correlation coeffiecient of 0,75 represents a

a) Moderate positive association

b) weak positive association

c) perfect positive association

d) strong negative association

Improve your Skills

Question 1

The table below represents the distance in meters required by a car to apply brakes and reach a standstill when it is travelling at a given speed.

Speed Km/h	Braking distance
20	6
40	16
60	30
80	48
100	70
120	80
140	110

i) Draw a scatter plot to represent this data.

ii) Explain whether a linear, quadratic or exponential curve would be a line or curve of best fit.

iii) Determine the equation of the regression line.

iv) Draw the regression line on the scatterplot diagram.

v) Use your line to estimate the breaking distance at a speed of a) 150km/h b) 130 km/h

Question 2

A leading nursery in Johannesburg recorded the effect of temperature on the growth of a new plant that has recently been imported in the country, The goal of the study is to determine what temperature is ideal for maximum flowering for a particular plant.

Temperature	Number of flowers
$25^{\circ}C$	2
$26^{\circ}C$	3
$27^{\circ}C$	4
$28^{\circ}C$	6
$29^{\circ}C$	7
$30^{\circ}C$	7
$31^{\circ}C$	8
$32^{\circ}C$	9
$33^{\circ}C$	10
$34^{\circ}C$	14

i) Draw a scatterplot to represent this bivariate data.

ii) Describe the trends shown by this scatterplot.

iii) Determine the equation of the regression line (called the line of best fit).

iv) Predict the number of flowers if the temperature is $38^{\circ}C$.

Question 3

The table below records the results of a study investigating how the number of times exercised per week reduces the number of stress related headaches.

Calculate the correlation coefficient and comment on the strength of the linear association.

Number of days exercised	Number of headaches
1	20
2	15
3	12
4	8
5	3
6	1

Question 4

During the Gauteng winter month of July, a number of patients visited a local hospital suffering from influenza. The table below shows the number of patients treated.

Dates in the month of July	4	6	10	12	16	20	22	26
Number of patients treated	260	280	380	420	600	680	800	820

i) Draw a scatterplot of the above data

ii) Determine the equation of the least squares regression line for the data.

iii) Draw the line of best fit on your scatterplot diagram.

iv) Estimate how many patients were treated as at 28 June.

v) Estimate how many patients were treated as at 23 July.

vi) Estimate how many patients were treated as at 31 July.

vii) Calculate the correlation coefficient for the data. Interpret this result.