CORK INSTITUTE OF TECHNOLOGY
INSTITIÚID TEICNEOLAÍOCHTA CHORCAÍ

Autumn 2018

Module Title: Statistical Calculations

Module Code: STAT6006

School: Science & Informatics

Programme Title(s):
BSc (Hons) in Analytical Chemistry and Quality Assurance (DCH2)
BSc in Chemistry (S2A)
BSc (Hons) in Environmental Science and Sustainable Technology (BEST2)

Programmes Code(s):
CR_SCHQA_8
CR_SCHEM_7
CR_SESST_8

External Examiner(s): Prof. Michael Wallace

Internal Examiner(s): Mr. Justin McGuinness

Instructions:
Answer **ALL** questions
All questions carry equal marks (20 marks each)
All work must be written in the answer book

Duration: 2 Hours

Sitting: Autumn 2018

Note to Candidates: Please check the Programme Title and the Module Title to ensure that you have received the correct examination. If in doubt please contact an Invigilator.

Question 1

(a) Give one example each of discrete and continuous data.

[2 mark]

(b) The following table shows the amount of money spent on Christmas Presents in 2015 by families in a rural village:

Amount of Money Spent (€)	Frequency
0 but less than 100	22
100 but less than 300	18
300 but less than 400	35
400 but less than 600	24
600 but less than 700	16
700 but less than 1000	15

(i) Using the frequency table, estimate the mean and standard deviation for the data set. Round your answers to the nearest cent.

[6 marks]

(ii) Using the frequency table, estimate the third quartile. Round your answer to the nearest cent.

[4 marks]

(iii) Using the graph paper provided, construct an Ogive.

[4 marks]

(iv) Use the Ogive to estimate the Interquartile Range; verifying your calculation for the third quartile in the process.

[2 marks]

(v) Use the Ogive to calculate the percent of families that spent more than €500 on Christmas Presents.

[2 marks]

Question 2

Analysis of rainfall shows that all rain is slightly acidic. The table below shows the pH of rainfall for seven samples along with the corresponding sulphur dioxide emissions.

SO_2 Emissions (millions of tons), x	0.6	0.8	0.9	1.1	1.4	1.7	2.3
pH, y	6.2	5.7	5	4.9	4.3	4.2	4

(a) Using graph paper, plot a scatter diagram for the data.

[3 marks]

(b) Calculate the coefficient of correlation correct to 2 decimal places and comment on your answer.

[9 marks]

(c) Use the least squares method to find the regression equation, rounding the values of b and a to 2 decimal places. Plot the regression line on the scatter diagram from part (a).

[6 marks]

(d) Using the regression equation that you obtained in part (c), estimate

(i) The pH of rainfall with a sulphur dioxide level of 1.0 millions of tons.

(ii) The pH of rainfall with a sulphur dioxide level of 2.5 millions of tons.

(iii) Which estimate is more accurate and why.

[2 marks]

Question 3

A local solid fuel salesman has noticed a decline in sales over the past 4 years. He is going to forecast sales for the upcoming year based on the sales from the previous 4 years. The sales figures are as follows:

Sales (€000s)

	Quarter 1	**Quarter 2**	**Quarter 3**	**Quarter 4**
2009	24	16	5	30
2010	22	15	8	29
2011	21	12	4	26
2012	15	9	3	25

(a) Using the graph paper provided, plot the data on a time series graph to illustrate the fluctuations in the data.

[4 marks]

(b) Calculate a four quarter centred moving average for the data and plot them on your graph from part (a).

[7 marks]

(c) Using a multiplicative model, determine seasonal index values for the data.

[7 marks]

(d) The trend in the data can be depicted by the linear equation:

$$y = 21.26 - 0.556x$$

where x units corresponds to 1 quarter and $x = 1$ gives Quarter 1 in 2009. Use the linear trend equation and the seasonal index to forecast figures for the first and third quarters of 2013.

[2 marks]

Statistical Formulae

Descriptive Statistics

$$\bar{x} = \frac{\sum fx}{\sum f}$$

$$s = \sqrt{\frac{\sum f(x-\bar{x})^2}{\sum f}} \quad \text{or} \quad s = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$$

$$\text{Median} = L_M + C_M\left(\frac{\frac{1}{2}N - F_{M-1}}{f_M}\right)$$

$$\text{Mode} = L_M + C_M\left(\frac{f_M - f_{M-1}}{2f_M - (f_{M-1} + f_{M+1})}\right)$$

$$\text{Coefficient of Variation} = \frac{s}{\bar{x}} \times 100\%$$

$$\text{Coefficient of Skewness} = \frac{3(Mean - Median)}{Standard\ Deviation}$$

Regression and Correlation

$$y = a + bx$$

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

$$a = \frac{\sum y - b\sum x}{n} \qquad or \qquad a = \bar{y} - b\bar{x}$$

$$\text{Coefficient of Correlation} = r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2} \cdot \sqrt{n\sum y^2 - (\sum y)^2}}$$