Silence Please
Do Not turn over this page until advised to by the Invigilator

CIT Semester 1 Examinations 2018/19

Note to Candidates:	Check the Programme Title and the Module Description to ensure that you have received the correct examination. If in doubt please contact an Invigilator.
Module Title:	**Stats & Experimental Design**
Module Code:	**STAT8004**
Programme Title(s):	BEng Hons Chem&Biopharm Eng Y3 BSc Hons Env Sci & Sus Tech Y4
Block Code(s):	**ECPEN_8_Y3** **SESST_8_Y4**
External Examiner(s):	**Dr. Katarina Domijan**
Internal Examiner(s):	Ms. Sarah Murphy
Instructions:	Answer any 3 questions. All questions carry equal marks.
Duration:	2 Hours
Required Items:	Calculator, Murdoch & Barnes Tables, Log/Formulae Tables

Question 1

(a) Of all patients suffering from a particular illness, 35% experience improvement from a particular medication. A random sample of 20 people suffering from the illness were selected and were administered the medication. Let X represent the number of patients who experience improvement.

i. Is X discrete or continuous? Explain
ii. What is the probability that exactly one patient will experience improvement?
iii. What is the probability that more than one patient will experience improvement?

(5 marks)

(b) Serious paint blemishes on a flat automotive panel occur at an average rate of 0.3 per panel with blemishes occurring according to the pattern of a Poisson distribution.

i. If a panel is chosen at random, what is the probability of there being two serious blemishes?
ii. How many panels need to be examined if the chance of observing at least one serious blemish is 80%?

(6 marks)

(c) A device that monitors the level of pollutants has sensors that detect the amount of CO in the air. Placed in a particular location, it is known that the amount of CO is normally distributed with a mean of 5.24 ppm and a standard deviation of 1.05 ppm.

i. What is the probability that the CO level is between 5.4 ppm and 8.3 ppm?
ii. Find the value of k such that the level of CO at this location lies above k 4% of the time.

(6 marks)

(d) Suppose that X is a continuous random variable whose probability density function is given by the following

$$f(x) = \begin{cases} kx(1-x)^2 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

i. Find the value of k.
ii. Find the expected value of X.
iii. What is the probability that X is greater than 0.6.

(8 marks)

Question 2

(a) The contents (in litres) of a random sample of 12 containers of a certain type of chemical were as follows:

9.7	9.6	9.6	10.7	10.2	9.9
10.3	10.5	9.8	9.8	10.1	10.3

i. Calculate a 99% confidence interval for the true mean content off all such containers.
ii. Calculate a 95% confidence interval for the variance of the content off all such containers.

(6 marks)

(b) One hundred households in the Bishopstown area were selected at random and the number of children in each was recorded. The following table shows the results of this survey:

Number of Children	0	1	2	3	4
Frequency	16	23	31	19	11

Are these data consistent with a Poisson distribution? Carry out a suitable test using a 5% significance level.

(10 marks)

(c) The level of chlorobenzene contained in the waste water for a particular chemical plant is under investigation. Eight water samples were taken at random and the level of chlorobenzene (mg/L) in each was measured. These data are shown in the table below:

Chlorobenzene levels (mg/L)	0.178	0.183	0.203	0.204	0.166	0.154	0.163	0.158

The maximum waste water contaminant level for chlorobenzene is 0.2mg/L. The manager of the chemical plant under investigation has claimed that the average chlorobenzene level found in their waste water is less than 0.2mg/L.

i. Carry out a suitable hypothesis test to test the manager’s claim. Clearly state your hypotheses and conclusions. Use a significance level of 1%.
ii. Find an approximate p −value for this test.

(9 marks)

Question 3

(a) It is believed that the environmental temperature at which batteries are activated affects their lifespan. Various different temperatures were examined. The lifespan in minutes of randomly tested batteries are shown in the table below.

Activation Temperature (℃)	**Lifespan in minutes**		
0	55	57	54
10	68	72	70
20	73	75	78
30	59	61	60

i. How many factor levels are involved in this experiment?
ii. How many replicates are involved in this experiment?
iii. Using a 1% level of significance, perform an analysis of variance on these data to investigate whether activation temperature has a significant effect on the lifespan of batteries. Clearly state your hypotheses and conclusions.

(12 marks)

(b) A study is conducted on the effect of temperature (A), time in process (B), and rate of temperature rise (C) on the amount of dye (in ml) left in the residue bath in a dyeing process. The experiment was run at two levels of temperature (120℃ and 140℃), two levels of time in the process (40 min, 55 min) and two rates of temperature rise (R_1 and R_2). Two readings were taken at each combination of factor levels. The resulting set of data is as follows:

	Temperature (A)			
	120 ℃		**140** ℃	
	Time in Process (B)			
Rate (C)	**40 min**	**55 min**	**40 min**	**55 min**
R_1	18.6	17.4	25	19.5
	19.9	16.8	22.8	18.3
R_2	16.1	14.6	18	26.2
	14.5	16.3	27.7	28.3

i. Calculate an estimate of the temperature effect, the rate effect and the temperature-rate interaction effect.
ii. Estimate the error variance.
iii. Test the significance of the effects estimated in part i. Use a significance level of 5%.
iv. Draw the temperature-rate interaction plot and comment.

(13 marks)

Question 4

(a) An experiment is created to determine the effect of two water treatment methods (A and B) on magnesium uptake. Magnesium levels in grams per cubic centimetre (g/cm^3) are measured and two different time levels (1 hour and 2 hours) are incorporated into the experiment. The data are as follows:

	Treatment	
Time (hr)	***A***	***B***
1	2.16, 2.15, 2.19	2.06, 2.04, 2.01
2	2.10, 2.09, 2.12	1.86, 1.88, 1.91

i. Given that the error sum of squares is 0.003867 and the interaction sum of squares is 0.006075, compile the ANOVA table for this experiment. Clearly state your conclusions using a 5% significance level.

ii. Produce an interaction plot based on this set of data and comment.

(11 marks)

(b) The simple linear regression model takes the following form

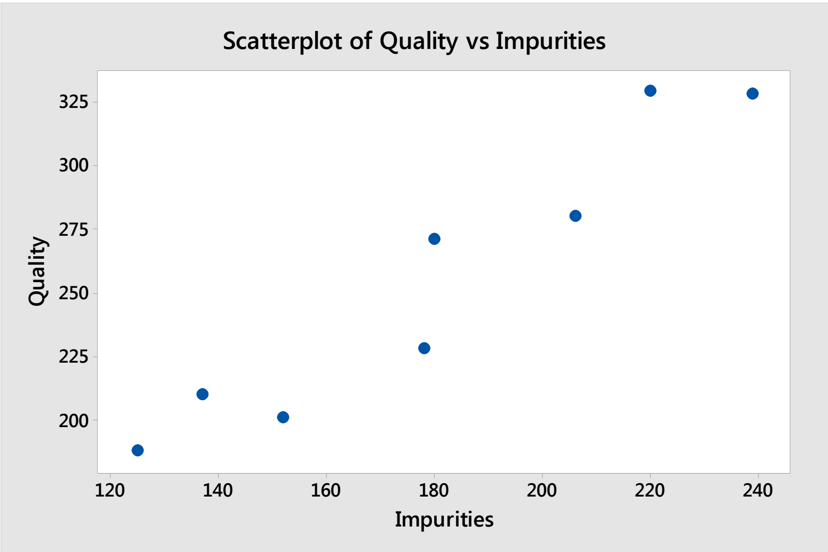
$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

i. Provide a clear interpretation of the parameters β_0 and β_1 in this model.

ii. Outline the meaning of the error term in the model paying particular attention to the distribution of this term.

(5 marks)

(c) A study was conducted to examine the relationship between the quality measure (y) of high grade stainless steel for use in a biopharmaceutical facility and the measure of impurities (x) in the iron used to manufacture this steel. A simple linear regression analysis was conducted in Minitab and some selected output is shown below and on the next page:



Analysis of Variance

Source	DF	SS	MS	F-Value
Regression	1	19973	*	*
Error	*	*	*	
Total	7	21842		

Using the Minitab output, answer the following questions:

i. Comment on the relationship between the two variables.
ii. Fill in the missing entries marked * in the Analysis of Variance table, and say what conclusions may be drawn from the table. Cleary state the relevant hypotheses and conclusions.
iii. Find the value of the correlation coefficient and test its significance using $\alpha = 0.01$.
iv. The equation of the least squares line generated for these data is $\hat{y} = 16.8 + 1.323x$.
    a. The measure of impurity in the iron used to manufacture a particular steel specimen was 170. Predict the quality measure for this measure of impurity.
    b. If the actual quality measure is 257, calculate the residual associated with this observation.

(9 marks)

Statistical Formulae

$$Z = \frac{X - \mu}{\sigma}$$

$$\frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$\bar{x} \pm t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}$$

$$Z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$(\bar{x}_1 - \bar{x}_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$\frac{s_1^2}{s_2^2} \sim F_{n_1 - 1, n_2 - 1}$$

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\left(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}, n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}\right)$$

ANOVA

1. **One-way model**: $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, $i = 1, 2, ..., a$, $j = 1, 2, ..., n.$

Total SS: $\sum \sum y^2{}_{ij} - \frac{y_{..}^2}{an}$

Factor SS: $\sum_{i=1}^{a} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{an}$

2. **AXB factorial model**:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta_{ij}) + \varepsilon_{ijk},\ i = 1,2,...,a;\ j = 1,2,...,b;\ k = 1,2,...,n.$$

Total SS: $\sum\sum\sum\ y^2{}_{ijk} - \frac{y_{...}^2}{abn}$

SSA: $\sum_{i=1}^{a} \frac{y_{i..}^2}{bn} - \frac{y_{...}^2}{abn}$ SSB: $\sum_{j=1}^{b} \frac{y_{.j.}^2}{an} - \frac{y_{...}^2}{abn}$

3. 2^k design, n replicates.

Effect estimate given by $\frac{(Contrast)}{n.2^{k-1}}$

Effect SS given by $\frac{(Contrast)^2}{n.2^k}$

V(effect estimate) $= \frac{\sigma_e{}^2}{n.2^{k-2}}$, where $\sigma_e{}^2$ is the error variance .