CORK INSTITUTE OF TECHNOLOGY

INSTITIÚID TEICNEOLAÍOCHTA CHORCAÍ

Autumn Examinations 2017/18

Module Title: Statistical Quality Control for Chemists.

Module Code: **STAT8003**

School: Science and Informatics

Programme Title:
Bachelor of Science (Honours) in Analytical Chemistry with Quality Assurance - Award

Programme Code:
SACQA_8_Y4

External Examiner(s): **Prof. M. Wallace**
Internal Examiner(s): **Dr. C. Palmer**

Instructions: Answer **any three questions**. All questions carry equal marks

Duration: **2 HOURS**

Sitting: Autumn 2018

Requirements for this examination: Statistical Tables by Murdoch and Barnes.

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

1.

(a) A company is considering the following acceptance plan:

Take a random sample of 50 items from the batch. If the sample contains 2 or fewer defectives, accept the batch; otherwise reject it.

Calculate the probability of accepting the batch if it is:

(i) 0% defective

(ii) 1% defective

(iii) 5% defective

(iv) 10% defective (4 marks)

(b) Use your answers from question 1 to create a plot showing how the probability of accepting a batch varies with the percentage of defective items in the batch.

(2 marks)

(c) The amount of copper in 1 gram of a particular type of brass alloy is known to follow a Normal distribution with a mean of 0.6 grams and a standard deviation of 0.019 grams.

(i) What is the probability that a randomly chosen sample of 1g of the alloy will contain more than 0.65g of copper?

(ii) 20% of samples will contain more than k grams of copper. What is the value of k?

(iii) Suppose 5 samples of 1g of the alloy are chosen at random and the mean amount of copper per 1g is calculated. Let $\bar{X}$ represent the sample mean amount of copper per 1g. What is the probability that $\bar{X}$ will lie between 0.59 and 0.62g?

(7 marks)

(d) To monitor the quality of components produced on a production line 40 samples, each containing 4 components, are taken at random and weighed to the nearest 0.01g. The mean and range are determined for each sample and the following summary data are obtained: $\bar{\bar{x}} = 3.13$ g and $\bar{R} = 0.12$ g.

(i) What are the 3-sigma control limits for $\bar{x}$ and R?

(ii) Assume that the process output is normally distributed. Suppose that the lower specification limit for the component is 3.05 g and the upper specification limit for the weight of the component is 3.18 g. Calculate the percentage of components produced that fail to meet the specification limits.

(iii) Calculate the process capability ratio of the production line.

Note, the estimate for σ is $\frac{\bar{R}}{d_2}$. (12 marks)

2.

(a) In the case of each of the following distributions, determine the lower- and upper-tail critical values which between them contain the central 99% of values in the distribution:

(i) a normal distribution, with mean 120 and standard deviation 15;

(ii) a t-distribution, with 8 degrees of freedom;

(iii) a χ^2 distribution, with 15 degrees of freedom (6 marks)

(b) A company produces a therapeutic drug. A random sample of 8 measurements of the concentration (g/L) of the active ingredient found in a dose of the drug yielded the following data.

14.21 14.25 14.18 14.31 14.26 14.22 14.19 14.25

(i) Calculate a 99% confidence interval for the *mean* concentration of the active ingredient and interpret the confidence interval.

(ii) Produce a 95% confidence interval for the *variance* of the concentration of the active ingredient.

(5 marks)

(c) A company that manufactures a specific type of laboratory reagent in 10mL packs claims that the standard deviation of the packs is at most 0.05 mL. One of its clients decides to test this claim by taking a sample of 20 packs and checking the quantity in each pack.

(i) Suppose that the standard deviation of the sample was estimated to be 0.09 mL. Does this undermine the company's claim?

(ii) What assumption do you need to make in order to perform this test?

(6 marks)

(d) The concentration of arsenic present in two different brands of rice milk is to be compared. Seven random samples from brand A and seven random samples from brand B were analysed and the concentration of arsenic (μg/L) is shown in the table below. Test whether the mean concentrations for the brands differ significantly. You may assume that the pooled variance approach is appropriate here.

Brand A	8.67	8.37	8.43	8.21	7.98	8.35	8.22
Brand B	8.36	8.54	8.64	8.41	8.29	8.42	8.25

(8 marks)

3.

(a) A therapeutic drug is manufactured at three different sites. The consistency of the drug is under investigation and five samples of the drug are chosen at random from each site. The concentration of active ingredient in each sample is measured and the results are shown in the table below.

	Concentration of active ingredient (mg/L)			
Site 1	17.9	17.5	17.3	17.8
Site 2	17.2	17.4	17.0	17.2
Site 3	17.4	17.5	17.3	17.6

Carry out a one-way analysis of variance on this set of data to determine whether there is a difference in the mean level of active ingredient between the three manufacturing sites. (12 marks)

(b) A factorial experiment is carried out to determine how the temperature (°C), and the concentration of a reactant (mol/L) affect the reaction time of a chemical process. Two replicates of a 2^2 design produced the following results.

Temperature (°C)	Concentration (mol/L)	Reaction Time (seconds)
40	12	3.65, 3.71
40	14	4.11, 4.08
45	12	3.82, 3.71
45	14	3.76, 3.78

(i) Estimate the main effects and the interaction effect.

(ii) Test the significance of the effect estimates calculated in (ii).

(iii) Draw the interaction plot, and comment. (13 marks)

4.

A study was made to determine the effect of stirring rate on the amount of impurity in a solvent produced by a chemical process. The study yielded the following data.

Stirring rate, rpm (x)	20	22	24	26	28	30	32	34	36
Impurity (%) (y)	9.3	10.6	12.8	11.2	14.4	14.2	15.7	15.9	16.3

$\sum x = 252 \quad \sum y = 120.4 \quad \sum xy = 3476.6 \quad \sum x^2 = 7296 \quad \sum y^2 = 1662.12$

(i) Construct a scatter plot for these data, and comment on the level of correlation. (5 marks)

(ii) Find the equation of the regression line of percentage of impurity on stirring rate. (6 marks)

(iii) Produce the analysis of variance table for the regression, and say what conclusions you draw from it. (7 marks)

(iv) Using relevant information from the ANOVA table, find the value of the correlation coefficient. (2 marks)

(v) Find 95% confidence limits for the population slope. (5 marks)

Statistical Formulae

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

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

$$t_{n-1} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}};$$

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

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

$$t_{n_1+n_2-2} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

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

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

$$\frac{s_1^2}{s_2^2} \sim F^{n_1-1}_{n_2-1}$$

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

$$SE(p) = \sqrt{\frac{p_0(1-p_0)}{n}}$$

ANOVA

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

Total SS: $\sum\sum y^2_{ij} - \frac{y_{..}^2}{an}$ or $\sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{\bar{y}})^2$

Factor SS: $\sum_{i=1}^{a} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{an}$ or $n \sum_{i=1}^{a} (\bar{y}_i - \bar{\bar{y}})^2$

2^k design, *n* replicates

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

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

Variance of an effect estimate is $\frac{\sigma_e^2}{n.2^{k-2}}$

Regression

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} \qquad S_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} \qquad S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$s_e = \hat{\sigma} = \sqrt{\frac{S_{yy} - \widehat{\beta_1}^2 S_{xx}}{n-2}} \qquad s_e = \hat{\sigma} = \sqrt{\frac{SSE}{n-2}}$$

$$S.E.(\hat{\beta}_1) = \frac{s_e}{\sqrt{S_{xx}}} \qquad S.E.(\hat{\beta}_0) = s_e\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}$$

$$SSR = \frac{(S_{xy})^2}{S_{xx}}$$

$$S.E.(\hat{\beta}_0 + \widehat{\beta_1} x_0) = s_e\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

$$S.E.(individual\ \hat{\beta}_0 + \widehat{\beta_1} x_0) = s_e\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$