DEPARTMENT OF CHEMISTRY

UNIVERSITY OF SWAZILAND

C610 RESEARCH METHODS

MAY 2015 FINAL EXAMINATION

Time Allowed: Three (3) Hours

---

*Instructions*

1. This examination has six (6) questions and one data sheet. The total number of pages is seven (7) including this page.
2. Answer any two (2) questions from Section A, and any two (2) from Section B; diagrams should be clear, large and properly labeled. Marks will be deducted for improper units and lack of procedural steps in calculations.
3. Each question is worth 25 marks.

*Special Requirements*

1. Data sheet.
2. Graph Paper.
3. Statistical Tables.

YOU ARE NOT SUPPOSED TO OPEN THIS PAPER UNTIL PERMISSION TO DO SO HAS BEEN GIVEN BY THE CHIEF INVIGILATOR.

## SECTION A

Answer any two (2) questions from this section.

**Question 1 [25]**

a) Certified reference materials are useful in the evaluation of reliability and validity of analytical data, especially when the analyte is in a complex matrix. In the determination of zinc in soils from a 100-ha pineapple field,

(i) What kind of certified reference materials would be suitable for this analysis? [1]

(ii) How would bulk sampling be carried out to source in material? [3]

(iii) Outline the processes that such a material would undergo during certification. [4]

(iv) Explain how this material would be used to evaluate validity and reliability of zinc measurements in the soil. [3]

(v) What is the major drawback in using certified reference materials? [1]

b) Interlaboratory comparisons are useful in the evaluation of reliability and validity of analytical data. In the measurement of zinc in the soils in (a) above, "LAB A" ran ten replicate measurements on a sample from the field, and requested "LAB B" to do the same with the remainder of the sample. The following results were obtained:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **LABA (ppm)** | **125** | **123** | **121** | **214** | **125** | **122** | **120** | **122** | **121** | **120** |
| **LAB B (ppm)** | **123** | **129** | **122** | **118** | **115** | **121** | **125** | **129** | **132** | **121** |

(i) Comment on the validity of the results at the 95% confidence level [3]

(ii) Comment on the relative precisions of the two laboratories at the 95% confidence level [3]

c) Quality control charts are useful in ensuring that repetitive day to day measurements are under statistical control. An in-house reference material was used to generate the following data over a period of 10 days of measurement of zinc in the soils in (b) above:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Day #** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| **Zn, ppm** | **121** | **119** | **140** | **118** | **120** | **113** | **112** | **121** | **118** | **124** |

(i) What is meant by an "in-house reference material"? [1]

(ii) Draw the quality control chart for the zinc measurements, assuming that the in-house reference material is 120±3 ppm Zn. [3]

(iii) Which days were the measurements not under statistical control and what actions would you take on the results obtained on these days? [3]

**QUESTION 2 [25]**

a) In data acquisition, noise is an important concept in instrumental analysis as it is a predominant factor in determining errors in output quality, precision and detection limits.

(i) What is meant by "signal" in analytical data acquisition? [2]

(ii) What is meant by "noise" in analytical data acquisition? [2]

(iii) What is the significance of the concept "signal-to-noise ratio" in analytical data acquisition? [2]

(iv) Give the operational definition of "detection limit" in instrumental analysis [2]

b) In regard to Johnson Noise,

(ii) Explain its origins in analytical instrumentation [2]

(iii) Write down the equation relating the magnitude of this noise to the bandwidth, and explain all terms appearing in it [3]

c) In regard to Shot Noise,

(i) Explain its origins in analytical instrumentation [2]

(ii) Write down the equation relating the magnitude of this noise to the bandwidth, and explain all terms appearing in it [3]

d) When atomic absorption analog signals are sampled, the rate of sampling is important.

(ii) Use diagrams to explain how undersampling leads to errors due to aliasing [4]

(iii) Use the Nyguist Theorem to explain how aliasing errors can be eliminated during data acquisition [3]

**QUESTION 3 [25]**

a) The following absorbance data was obtained in triplicate during measurement of lead in a river sediment sample using atomic absorption, AA, following a classical dissolution of a 500-mg portion:

| **Lead, Pb** | **Absorbance Readings** | | |
|---|---|---|---|
| **Standard, 1.16 ppm** | **0.120** | **0.125** | **1.130** |
| **Standard, 2.32 ppm** | **0.248** | **0.255** | **0.252** |
| **Standard, 3.48 ppm** | **0.382** | **0.385** | **0.384** |
| **Standard, 4.65 ppm** | **0.504** | **0.506** | **0.502** |
| **Sample** | **0.337** | **0.335** | **0.340** |

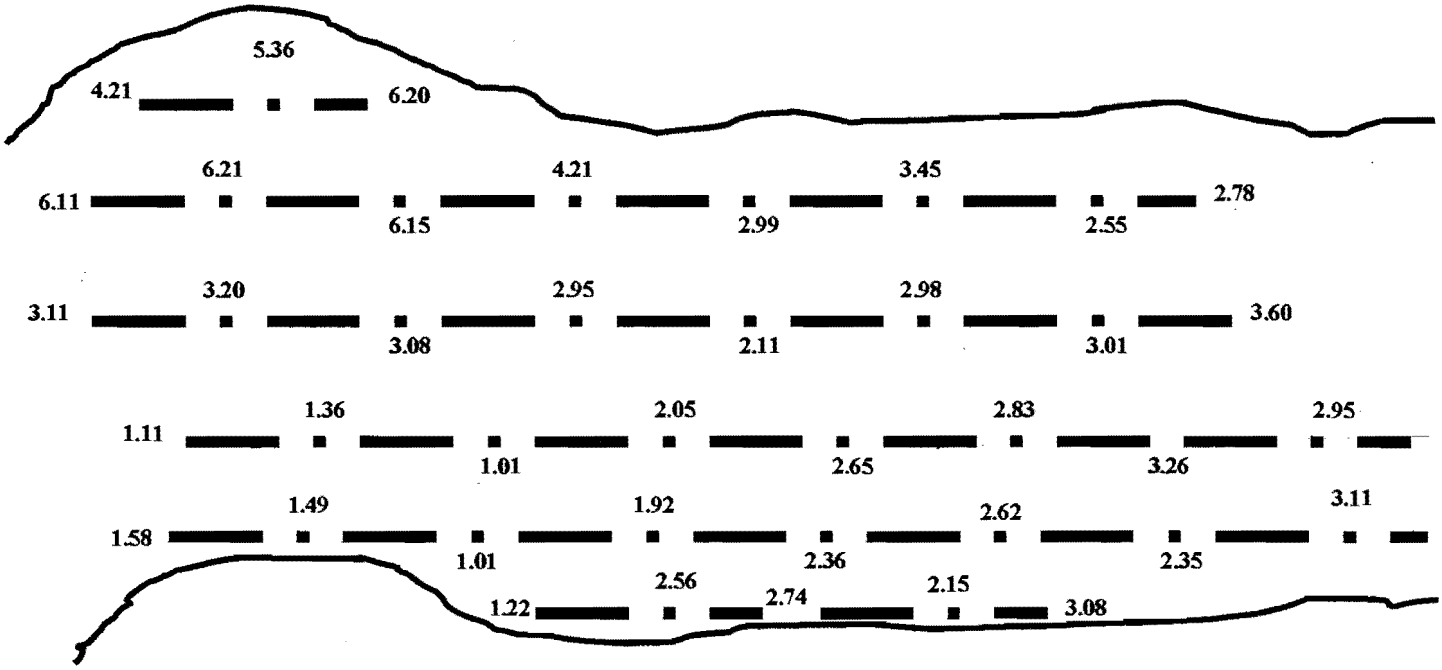
(iv) Perform a linear regression on this data set using the Least Squares Method [5]

(v) Calculate the absolute error associated with the calibration curve, $S_v$ [3]

(vi) Calculate the absolute error associated with the measurement operation, $S_m$ [2]

(vii) Calculate the subsampling uncertainty, $S_{ss}$, if five 500-mg portions of this sample gave the following results of Pb 3.08 ppm; 3.07 ppm; 3.11 ppm; 5.01 ppm; and 309 ppm [2]

b) Forty (40) samples of sediments were taken from a river to map the spatial variability of Pb in the sediments. 500-mg portions of each dried sediment sample were digested and Pb measured by AA on the same day and same instrument as in part 3 (a) above. The spatial distribution of Pb, in ppm, was found to be as follows:




(i) Use the Kolmogorov-Smirnoff test to show that the distribution of Pb in the sediments is not Gaussian. [5]

(ii) What is meant by a "hot spot" or "cold spot" in analytical sampling, and on the distribution of Pb on the sediments as mapped, indicate these spots [3]

(iii) Use the Student's t-test equation to determine the minimum number of samples to be taken from the population, if the average value of Pb is to be within the error due to sampling, at the 95% confidence level [5]

## SECTION B

**Answer any Two(2) questions from this section**

**Question 4 [25 Marks]**

a) Write short notes on "Principal Component Analysis, PCA". In your brief description include uses, applications, weaknesses and any relevant detail of the technique as applied in Chemometrics. [6]

b) Data is sometimes scaled in PCA before application of the techniques. Give two reasons with examples to clarify your answer. [2]

Scaling Equations:

$$Z = \frac{x_i - \bar{x}}{Sx} \text{ or } Z = \frac{x_i}{Sx}$$

Where 'z' is the standardized variable, $x_i$ is the variable, $\bar{x}$ is the mean and $S_x$ is the standard deviation.

c) Using the data below calculate and discuss the following: [8]

i) Eigen values

ii) Eigen vectors

iii) Loadings factors

iv) Score factors

| Sample sites | Station 1 | Station 2 | Station 3 | Station 4 | Station 5 |
|---|---|---|---|---|---|
| Variables | | | | | |
| $NO_2$ | 26 | 20 | 5 | 4 | 7 |
| NOx | 1 | 2 | 7 | 11 | 4 |

**Show your working. You may use STASTICA to confirm your calculations above.**

**Useful relation**

$$\boxed{|R - \lambda I| = 0} \qquad x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

d) What is the optimum number of principal components, PC's and what is the percentage explained variance as defined by the optimum number of Principal Components? [3]

e) Briefly discuss your findings in your principal component analysis above. In your discussion include comments on sample groups, variable groups, correlations and any observations of vital importance in your findings.[6]

**Save all your working from the computer in a USB**

**Question 5 [25 Marks]**

a) Define the term "Cluster Analysis, CA". In your brief description include uses, applications, strengths/weaknesses and any relevant details of the technique as applied in chemometrics. [10]

b) Using the data below calculate distance matrix d(i,k). [4]

$$d(i,k)=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+(x_3-y_2)^2.....(x_n-y_2)^2}$$

| Sample sites | Station 1 | Station 2 | Station 3 | Station 4 | Station 5 |
|---|---|---|---|---|---|
| Variables | | | | | |
| $NO_2$ | 26 | 20 | 5 | 4 | 7 |
| NOx | 1 | 2 | 7 | 11 | 4 |

c) Using the average linkage method by Lance and Williams determine the clusters of the data above (in b) and draw the appropriate dendrogram. [5]

Lance and Williams equation states that:

$$d(il'2,k)=\alpha 1 di1,k+\alpha 2 di2,k+\beta di1,i2+\gamma|di1,k-di2,k|$$

where:

$\alpha 1$ = 0.5 is the weight between the distance of first joint object to any other object or cluster

$\alpha 2$ =0.5 is the weight between the distance of second joint object to any other object/ cluster

$\beta$ =0 is the weight of the distance of both neigbhouring objects

$\gamma$ = 0 is the weight of the difference between the distance of neigbhouring objects or clusters.

d) Briefly discuss your findings in your cluster analysis above. In your discussion include comments on clusters, correlations and any observations of vital importance in your findings. [6]

**Show all your working. You may use excel/STASTICA to confirm your calculations above. Save all your working from the computer in the USB provided.**

**Question 6 [25 Marks]**

a) Write short notes, comparing and contraction on Factor Analysis with principal component analysis. [10]

b) The data below show analysis of air within the Manzini region air space with 12 sampling points adjacent to major industrial activities. You are to use Multivariate Analysis techniques to study trends and correlations in this data. [15]

| Place | $NO_2$ | $NO_3$ | $CH_4$ | $CH_3CH_2CH_3$ | NO | $NO_X$ |
|---|---|---|---|---|---|---|
| NOKWN | 48 | 26 | 2 | 11 | 11 | 1 |
| NAMBD | 44 | 20 | 1 | 9 | 9 | 2 |
| SWAZWRE | 26 | 5 | 14 | 2 | 3 | 7 |
| SEWRGE | 24 | 4 | 10 | 6 | 2 | 11 |
| LOZTH | 31 | 7 | 7 | 5 | 5 | 4 |
| ZOMBDZ | 38 | 7 | 10 | 6 | 5 | 6 |
| BREWRY | 30 | 18 | 3 | 4 | 9 | 7 |
| FIREDEP | 40 | 24 | 7 | 2 | 10 | 2 |
| ENGEN | 41 | 21 | 2 | 4 | 9 | 3 |
| LUDZLDZ | 32 | 9 | 7 | 6 | 6 | 4 |
| SWPMLL | 28 | 6 | 11 | 16 | 5 | 6 |
| NGBZWN | 35 | 14 | 8 | 5 | 7 | 9 |

Using **ANY ONE** of the techniques PCA, FA or CA briefly discuss your findings. In your discussion include comments on sample groups, variable groups, correlations, dominant pollutants, major polluter and any observations of vital importance in your findings. Cut and paste all analysis from Statistica to Word and include Identity number within the first page of your Word document.

**Show all your working. You may use excel/STASTICA to confirm your calculations above. Save all your working from the computer in the USB provided.**