# UNIVERSITY *of* LIMERICK
OLLSCOIL LUIMNIGH

## FACULTY OF SCIENCE AND ENGINEERING

## DEPARTMENT OF MATHEMATICS & STATISTICS

## END OF SEMESTER ASSESSMENT PAPER

MODULE CODE: MA4605

SEMESTER: Autumn 2011

MODULE TITLE: Chemometrics

DURATION OF EXAMINATION: 2.5hrs

LECTURER: Ana-Maria Magdalina

PERCENTAGE OF TOTAL MARKS: 100%

EXTERNAL EXAMINER: Professor Brendan Murphy

### INSTRUCTIONS TO CANDIDATES:

- Answer all five questions. All questions carry equal marks.
- Mark clearly on the answer sheet the question number that you are answering.
- Calculators may be used.
- The exam is worth 100% of your final grade.

1. In a series of experiments on the determination of tin in canned food, samples of food were boiled with hydrochloric acid under reflux for two different times: 30 and 75 minutes. The size of both samples is 10.

| Refluxing time(min) | Tin found (mg/kg) |
|---|---|
| 30 | 55,57,59,56,56,59,56,60,54,55 |
| 75 | 57,55,58,59,59,59,58,53,56,57 |

(a) Are these data suitable for a paired t-test? Justify your answer. [2%]

(b) The normality of the two samples is checked below. Interpret the *R* output. Clearly indicate the null and alternative hypothesis. [4%]

**Anderson-Darling normality test**
data: sample1
A = 0.4596, p-value = 0.2036

**Anderson-Darling normality test**
data: sample2
A = 0.4413, p-value = 0.2281

(c) Test the assumption that the two samples are extracted from populations with equal variances. Clearly indicate the null and alternative hypothesis. [6%]
Interpret the 95% confidence interval for the ratio of variances.

**F test to compare two variances**
data: sample1 and sample2
F = 1.0344, num df = 9, denom df = 9, p-value = 0.9607
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
0.2569263 4.1644238
sample estimates:
ratio of variances 1.034384

(d) Test the claim that mean amount of tin found differs significantly for the two boiling times. The results of the t-test for equality of means are presented below. [8%]
Clearly indicate the null and alternative hypothesis.

**Two Sample t-test** data: sample1 and sample2
t = -0.4504, df = 18, p-value = 0.6578
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-2.266027 1.466027

2. (a) Three investigators performed five determination of C of 2,4-dinitrophenol in water using the same procedure. The results in $C/\mu M$ were: [10%]

| Investigator 1 | Investigator 2 | Investigator 3 |
|---|---|---|
| 701 | 550 | 510 |
| 677 | 540 | 520 |
| 680 | 570 | 540 |
| 660 | 530 | 540 |
| 654 | 520 | 560 |

The analysis of variance is used to determine if there is a significant difference between the mean of the determinations made by the three investigators. The following output is obtained in $R$.

$investigator <- c(701, 677, 680, 660, 654, 550, 540, 570, 530, 520, 510, 520, 540, 540, 560)$
$group <- factor(rep(1:3, each = 6))$

$anova(lm(investigator \sim group))$
Analysis of Variance Table

| | Df | SumSq | MeanSq | Fvalue | Pr(>F) |
|---|---|---|---|---|---|
| group | ? | 62177 | ? | ? | 8.011e-08 |
| Residuals | 12 | ? | ? | | |
| Total | ? | 66546 | | | |

Fill in the missing values in the ANOVA table and explain how these statistics can be used to assess if there is any difference between the mean determinations made by the three investigators. Clearly state the null and alternative hypothesis.

(b) The following analysis was performed in $R$ to analyze the relationship between the concentration of nicotinamide mononucleotide (X), in mM, and the initial rate of nicotinamide-adenine dinucleotide formed (Y), in micromoles. Standard deviations $s_i$ are also given together with $(x_i, y_i)$ [10%]

| Concentration(X) | 0 | 2 | 4 | 6 | 8 | 10 |
|---|---|---|---|---|---|---|
| Standard deviation($s_i$) | 0.001 | 0.004 | 0.010 | 0.013 | 0.017 | 0.022 |
| Absorbance(Y) | 0.009 | 0.158 | 0.301 | 0.472 | 0.577 | 0.739 |

Are these data homoscedastic or heteroscedastic?
What kind of analysis is most appropriate to this kind of data? Explain what are its advantages.

3. The mercury level of several tests of sea-water from costal areas was determined by atomic-absorption spectrometry. The results obtained are as follows

| Concentration in $\mu g l^{-1}$ | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Absorbance | 0.321 | 0.834 | 1.254 | 1.773 | 2.237 | 2.741 | 3.196 | 3.678 | 4.217 | 4.774 | 5.261 |

The analysis of the relationship between concentration and absorbance is obtained in $R$ and presented below.

$x < -seq(0, 100, by = 10)$

$y < -c(0.321, 0.834, 1.254, 1.773, 2.237, 2.741, 3.196, 3.678, 4.217, 4.774, 5.261)$

$model < -lm(y \sim x)$

$summary(model)$

$Call:$

$lm(formula = y \sim x)$

$Coefficients:$

| | Estimate | Std. Error | t value | $Pr(>\|t\|)$ |
|---|---|---|---|---|
| (Intercept) | 0.2933636 | 0.0234754 | 12.50 | 5.45e-07 |
| x | 0.0491982 | 0.0003968 | 123.98 | 7.34e-16 |

Residual standard error: 0.04162 on 9 degrees of freedom

Multiple R-squared: 0.9994, Adjusted R-squared: 0.9993

F-statistic: 1.537e+04 on 1 and 9 DF, p-value: 7.337e-16

(a) Sketch the data and describe the relationship between the two variables. [2%]

(b) Determine the slope and the intercept of the calibration plot. [2%]

(c) Knowing that qt(0.975,df=9) is 2.262, calculate the 95% confidence interval for the slope and the intercept. [4%]

(d) Using the fitted line calculate the predicted value of the Absorbance when the Concentration is 75. [2%]

(e) Using the fitted line calculate the predicted value of the Hg concentration when the Absorbance is 3. Estimate the 95% confidence interval for this concentration. The following results obtained in R are necessary for calculating the standard error of the concentration. [6%]

$mean(y)$
[1]2.753273
$mean(x)$
[1]50
$sum((x - mean(x))^2)$
[1]11000

The formula for the standard error of the concentration is:

$$s_{x_0} = \frac{s_{y/x}}{b}\sqrt{1 + \frac{1}{n} + \frac{(y_0 - \overline{y})^2}{b^2 \sum_i (x_i - \overline{x})^2}}, \text{ where } s_{y/x} = 0.04162.$$

(f) Estimate the limit of detection of the mercury analysis $x_0$, knowing that the corresponding $y_0$ value is obtained with the formula $y_0 = a + 3s_{y/x}$. $\boxed{4\%}$

4. **Part I**

Four standard solutions of chloride were prepared. Three titration methods, each with a different technique of end-point determination, were used to analyze each standard solution. The order of the experiments was randomized. The results of chloride found are shown below.

| Solution | Method 1 | Method 2 | Method 3 |
|---|---|---|---|
| 1 | 10.03 | 10.13 | 10.12 |
| 2 | 10.13 | 10.15 | 10.12 |
| 3 | 10.09 | 10.02 | 9.97 |
| 4 | 10.05 | 9.94 | 10.10 |

The following output table is the result of performing a two-way ANOVA without interactions in *R*.

**Analysis of Variance Table**

Response: chloride

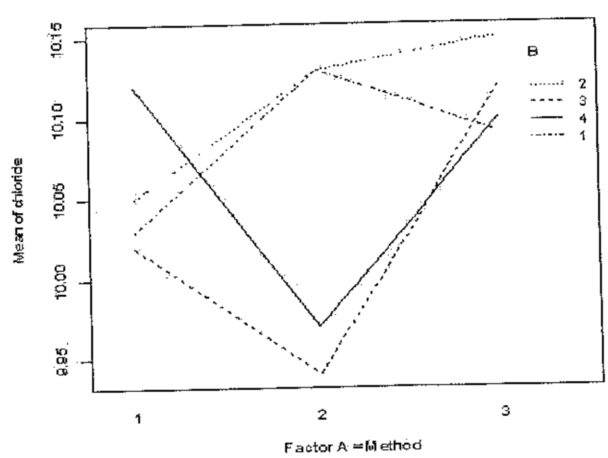
| | Df | Sum Sq | MeanSq | F value | Pr(>F) |
|---|---|---|---|---|---|
| A | 2 | 0.012017 | ? | 1.2791 | 0.3446 |
| B | ? | 0.011092 | ? | 0.7871 | 0.5435 |
| Residuals | 6 | 0.028183 | ? | | |

(a) Identify the two factors A and B, and their corresponding levels. Are they controllable or random? Construct the hypothesis statements. [2%]

(b) Fill in the missing values for the mean sum of squares. [2%]

(c) Test whether there are significant differences between the concentration of chloride in the different solutions, and whether there are significant differences between the results obtained by the different methods. [2%]

(d) Interpret the following interactions plot. [2%]



(e) Is it possible to include the interaction term in the model, given the data we have? Why? How can this be changed? [2%]

**Part II**

Three different chemical compounds are measured in three separate laboratories. Each laboratory performed a duplicate analysis on each compound and the results are presented below.

| | Compound | | |
|---|---|---|---|
| Laboratory | 1 | 2 | 3 |
| 1 | 5.1, 5.1 | 5.3, 5.4 | 5.3, 5.1 |
| 2 | 5.8, 5.4 | 5.4, 5.9 | 5.2, 5.5 |
| 3 | 6.5, 6.1 | 6.6, 6.7 | 6.5, 6.4 |

The following analysis was obtained in *R* by coding the compound as factor A and the different laboratories as factor B.

```
model <- lm(y~A+B+A:B)
anova(model)
```

**Analysis of Variance Table**

Response: y

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| A | 2 | 0.1878 | 0.09389 | 2.3151 | 0.1545 |
| B | 2 | 5.0678 | 2.53389 | 62.4795 | 5.281e-06 |
| A:B | 4 | 0.1022 | 0.02556 | 0.6301 | 0.6533 |
| Residuals | 9 | 0.3650 | 0.04056 | | |

(a) Explain what kind of ANOVA design is used for the analysis. [2%]

(b) Write conclusions to the present analysis. [8%]

5. (a) Explain what is the purpose of two-level factorial designs. [5%]

In an experiment for the determination of Titanium(Ti) in glass ceramics, we look at the influence of three elements Mg, Na and Si on the Ti signal. Sensible low(-) and high(+) levels in $\mu g m L^{-1}$ are proposed for each of these elements. Duplicate Ti measurements are obtained for each combination of factors.

| Run | Combination | Mg (a) | Na(b) | Si(c) | Response |
|---|---|---|---|---|---|
| 1 | (1) | (-) | (-) | (-) | 7,9 |
| 2 | a | (+) | (-) | (-) | 8,10 |
| 3 | b | (-) | (+) | (-) | 32,36 |
| 4 | ab | (+) | (+) | (-) | 50,54 |
| 5 | c | (-) | (-) | (+) | 14,18 |
| 6 | ac | (+) | (-) | (+) | 20,24 |
| 7 | bc | (-) | (+) | (+) | 43,47 |
| 8 | abc | (+) | (+) | (+) | 53,59 |

A full $2^3$ design is used to test for significance the main effects, the first and second order interactions. The ANOVA table containing these results is included below.

(b) Write the $R$ code that you would use to obtain this table. [5%]

(c) Interpret the results from the ANOVA table. [5%]

(d) What are your conclusions for the next step of the experiment? [5%]

**Analysis of Variance Table**

Response: y

| | Df | Sum Sq | Mean Sq | F value | $Pr(>F)$ |
|---|---|---|---|---|---|
| Mg | 1 | 4356 | 4356.0 | 562.0645 | 1.067e-08 |
| Na | 1 | 324 | 324.0 | 41.8065 | 0.000195 |
| Si | 1 | 324 | 324.0 | 41.8065 | 0.000195 |
| Mg:Na | 1 | 9 | 9.0 | 1.1613 | 0.312617 |
| Mg:Si | 1 | 121 | 121.0 | 15.6129 | 0.004228 |
| Na:Si | 1 | 1 | 1.0 | 0.1290 | 0.728734 |
| Mg:Na:Si | 1 | 36 | 36.0 | 4.6452 | 0.063254 |
| Residuals | 8 | 62 | 7.7 | | |