FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF MATHEMATICS AND STATISTICS

END OF SEMESTER EXAMINATION PAPER 2015

MODULE CODE: MA4605

SEMESTER: Autumn 2015

MODULE TITLE: Chemometrics

DURATION OF EXAM: 2.5 hours

LECTURER: Mr. Kevin O'Brien

GRADING SCHEME: 100 marks
70% of module grade

EXTERNAL EXAMINER: Dr. C.F. Ryback

INSTRUCTIONS TO CANDIDATES

Scientific calculators approved by the University of Limerick can be used.
Formula sheet and statistical tables provided at the end of the exam paper.
There are 5 questions in this exam. Students must attempt any 4 questions.

Question 1

- Basic Definitions (Type I and Type II error, what are p-values etc)
- Testing that Data is normally distributed (may appear elsewhere)
    - Shapiro Wilk Test
    - Anderson Darling Test
    - qq plots
    - *This is a good question to test your knowledge in Type II error and Power.*
- Transformation of Data (Tukey's Ladder - may appear in Q5)
- Outliers (Grubbs Test, Dixon Q-test)
- Interpreting Output for Inference Procedures (See Clinical Trials Scenarios for R Practical exam)
    - Two sample t-tests (both types) and paired t-tests
    - F-test for Equality of Variance
    - Non-Parametric Procedures (e.g. Wilcoxon test, Kolmogorov Smirnov Test) *(We are not doing much on this - just be familiar with the R procedures).*
- One Way ANOVA Tests for Multiple Means (may appear elsewhere) (HAND)
    - Calculate Between Group Sum of Squares
    - Calculate Within Group Sum of Squares
    - Calculate Total Sum of Squares

Question 2

What is going here?

- Linear Regression
    - Reading R output for Model Summary
    - Hypothesis tests for Regression Coefficients
    - Confidence intervals for Regression Coefficients
    - *(We didnt do a lot on prediction intervals and confidence intervals for fitted values this year, so dont expect a question on that)*
- Regression ANOVA (HAND)
- Simple and Multiple Linear Regression
- Model Fit Metrics such as AIC, and both R Squared
- Residuals - Heteroscedascity, Interpreting Diagnostic Plots 1 and 2, Autocorrelation, Cook's Distances, `ncvTest`, `outliersTest`, Durbin Watson Test
- Robust Regression, Huber Weighting
- Short Questions on Theory Topics Such as Overfitting, Law of Parsimony
- Method Comparison, Deming Regression
- Limits of Detection *(short question - with a bit or arithmetic to find both Y and X values)*
- Standard Addtions Method - *just a bit of theory on how what both variables mean. Dont expect any arithmetic this year.*

Question 3 and 4

What is going here?

- One Way Anova Tests for Multiple Means (HAND) (possible Q1 also)
- Understand the relationship between Variance and SStot
    - Calculate Between Group Sum of Squares
    - Calculate Within Group Sum of Squares
    - Calculate Total Sum of Squares
- Two Way ANOVA Testing (HAND)
    - No Replicates (i.e. give s_R^2 and S_C^2)
    - With Replicates - Partial Completion
    - Be very clear about Degrees of Freedom
- Checking Model Assumptions
    - Bartlett Test
    - Constant variance of Residuals
    - Normality of Residuals
- Factorial Design : Contrasts, Effects Sums of Squares
- Regression Equation based on Factorial Design
- Interaction Plots
- *(I mentioned material related to Treatments and Blocks, but decided to not to use it - instead using a very simple set up for Two Way ANOVA)*

Question 5

What is going here?

- Statistical Process Control
- Control Charts : Rules and Interpretation (very importiant)
    - State Rule
    - Sketch
    - Probability of event
- Process Capability Indices (HAND)
    - C_{pm} not included
    - Sketches are important
- Testing for Univariate Normality (may appear in Q1)

Critical Values for Dixon Q Test

N	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
3	0.941	0.970	0.994
4	0.765	0.829	0.926
5	0.642	0.710	0.821
6	0.560	0.625	0.740
7	0.507	0.568	0.680
8	0.468	0.526	0.634
9	0.437	0.493	0.598
10	0.412	0.466	0.568
11	0.392	0.444	0.542
12	0.376	0.426	0.522
13	0.361	0.410	0.503
14	0.349	0.396	0.488
15	0.338	0.384	0.475
16	0.329	0.374	0.463

Two Way ANOVA

$$MS_A = c \times S_r^2$$
$$MS_B = r \times S_c^2$$

Control Limits for Control Charts

$$\bar{\bar{x}} \pm 3\frac{\bar{s}}{c_4\sqrt{n}}$$

$$\bar{s} \pm 3\frac{c_5\bar{s}}{c_4}$$

$$\left[\bar{R}D_3, \bar{R}D_4\right]$$

Process Capability Indices

$$\hat{C}_p = \frac{\text{USL} - \text{LSL}}{6s}$$

$$\hat{C}_{pk} = \min\left[\frac{\text{USL} - \bar{x}}{3s}, \frac{\bar{x} - \text{LSL}}{3s}\right]$$

$$\hat{C}_{pm} = \frac{\text{USL} - \text{LSL}}{6\sqrt{s^2 + (\bar{x} - T)^2}}$$

2^3 Design: Interaction Effects

$$AB = \frac{1}{4n}\left[abc - bc + ab - b - ac + c - a + (1)\right]$$

$$AC = \frac{1}{4n}\left[(1) - a + b - ab - c + ac - bc + abc\right]$$

$$BC = \frac{1}{4n}\left[(1) + a - b - ab - c - ac + bc + abc\right]$$

$$ABC = \frac{1}{4n}\left[abc - bc - ac + c - ab + b + a - (1)\right]$$

Factorial Design: Sums of Squares

$$\text{Effect} = \frac{\text{Contrast}}{4n}$$

$$\text{Sums of Squares} = \frac{(\text{Contrast})^2}{8n}$$

Factors for Control Charts

Sample Size (n)	c4	c5	d2	d3	D3	D4
2	0.7979	0.6028	1.128	0.853	0	3.267
3	0.8862	0.4633	1.693	0.888	0	2.574
4	0.9213	0.3889	2.059	0.88	0	2.282
5	0.9400	0.3412	2.326	0.864	0	2.114
6	0.9515	0.3076	2.534	0.848	0	2.004
7	0.9594	0.282	2.704	0.833	0.076	1.924
8	0.9650	0.2622	2.847	0.82	0.136	1.864
9	0.9693	0.2459	2.970	0.808	0.184	1.816
10	0.9727	0.2321	3.078	0.797	0.223	1.777
11	0.9754	0.2204	3.173	0.787	0.256	1.744
12	0.9776	0.2105	3.258	0.778	0.283	1.717
13	0.9794	0.2019	3.336	0.770	0.307	1.693
14	0.9810	0.1940	3.407	0.763	0.328	1.672
15	0.9823	0.1873	3.472	0.756	0.347	1.653
16	0.9835	0.1809	3.532	0.750	0.363	1.637
17	0.9845	0.1754	3.588	0.744	0.378	1.622
18	0.9854	0.1703	3.64	0.739	0.391	1.608
19	0.9862	0.1656	3.689	0.734	0.403	1.597
20	0.9869	0.1613	3.735	0.729	0.415	1.585
21	0.9876	0.1570	3.778	0.724	0.425	1.575
22	0.9882	0.1532	3.819	0.720	0.434	1.566
23	0.9887	0.1499	3.858	0.716	0.443	1.557
24	0.9892	0.1466	3.895	0.712	0.451	1.548
25	0.9896	0.1438	3.931	0.708	0.459	1.541