University *of* Limerick
Ollscoil Luimnigh

FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF MATHEMATICS & STATISTICS

END OF SEMESTER ASSESSMENT PAPER

MODULE CODE: MA4007 | **SEMESTER: Autumn 2008/2009**

MODULE TITLE: Experimental Design | **DURATION OF EXAM: 2.5 hours**

LECTURER: Dr. N. Coffey | **GRADING SCHEME: Examination: 90%**

EXTERNAL EXAMINER: Prof. A. Bowman

INSTRUCTIONS TO CANDIDATES

Answer Question 1 (30%) and THREE other questions (20% each).

Statistical tables available from the invigilators. A set of formulae is attached to this paper. Calculators may be used.

Q1. (a). The following data give the recovery of bromide from spike samples of vegetable matter measured using a gas liquid chromatographic method. The same amount of bromide was added to each specimen. Also given is a sample of Minitab output calculated from these data.

Tomato:	777	790	759	790	770	758	764mg g^{-1}
Cucumber:	782	773	778	765	789	797	782mg g^{-1}

```
          N   Mean  StDev  SE Mean
Tomato    7  772.6   13.6      5.1
Cucumber  7  780.9   10.4      3.9
```

Do the recoveries from the two vegetables have variances that differ significantly? State clearly the hypothesis to be tested and your conclusion. Use $\alpha = 0.05$.

[5 marks]

(b). Distinguish between matched pairs and independent groups? Which category do the data in part (a) fall into? Do the mean recovery rates for the data in part (a) differ significantly? State clearly the hypothesis to be tested and your conclusion. Use $\alpha = 0.05$.

[5 marks]

(c). Complete the following ANOVA table:

Source	DF	SS	MS	F
A	1		3088	
B		3400		
AB	3	49000		
Error				
Total	23	63000		

How many replicates were used?

[5 marks]

(d). From the ANOVA table in part (c), state the hypotheses to be tested, run the tests and state your conclusions.

[5 marks]

(e). Set up a scheme for confounding a 2^4 factorial design into two blocks of eight each

(i) confounding ABCD with blocks

(ii) confounding ABD with blocks

[5 marks]

(f). Explain the term 'multicollinearity'. Test the following correlation table for evidence of multicollinearity assuming that $n = 12$ and $\alpha = 0.05$. State clearly the hypothesis being tested.

	$\mathbf{Y}$	$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$
$\mathbf{Y}$	1.0	0.398	-0.224	0.606
$\mathbf{X}_1$		1.0	-0.428	0.123
$\mathbf{X}_2$			1.0	-0.079
$\mathbf{X}_3$				1.0

[5 marks]

Q2. In the production of synthetic fibre, extrusion temperature and the machine used during manufacture are important in determining strength. Five temperatures are used on each of four machines, and each combination is sampled twice after running. The results are as follows:

	Temperature					
Machine	250	255	260	265	270	T_j
I	9,7	13,17	15,7	13,8	11,11	111
II	9,7	7,9	12,14	9,15	16,13	111
III	4,0	6,9	8,8	11,16	20,19	101
IV	0,7	1,3	6,11	12,10	19,21	90
T_i	43	65	81	94	130	T = 413

$$S = \sum\sum\sum y_{ijk}^2 = 5329$$

(a). Is it possible to estimate an interaction effect between machine and temperature? Justify your answer.

[2 marks]

(b). Write down the appropriate mathematical model based on your conclusion in part (a).

[2 marks]

(c). Determine the total sum of squares, the sum of squares for machine, the sum of squares for temperature and the sum of squares for the interaction (if appropriate).

[8 marks]

(d). Set up the ANOVA table and test all appropriate hypotheses. Clearly state the hypotheses under consideration and your conclusions in each case.

[6 marks]

(e). Give three possible reasons for an inflated residual sum of squares.

[2 marks]

Q3. An engineer wishes to maximise the yield from a particular process. There are three factors that could affect the yield; temperature, concentration and type of catalyst used. Each factor has two settings:

Factor	-1	+1
A (Temperature)	160 deg	180 deg
B (Concentration)	20	40
C (Catalyst)	A	B

An experiment was carried out to determine what factors have an effect on the yield, with the following results.

Treatment	A	B	C	y_1	y_2
(1)	-1	-1	-1	59	61
a	+1	-1	-1	74	70
b	-1	+1	-1	50	58
ab	+1	+1	-1	69	67
c	-1	-1	+1	50	54
ac	+1	-1	+1	81	85
bc	-1	+1	+1	46	44
abc	+1	+1	+1	79	81

$$S = \sum\sum y_{ij}^2 = 68748, \quad T = \sum y_{ij} = 1028$$

(a). Set up the matrix of contrasts for the main effects and interactions. Calculate the contrast, the estimate and the sum of squares for each effect.

[5 marks]

(b). Set up the ANOVA table to show the breakdown of the total sum of squares into all sources of possible variation. Calculate the error sums of squares and test for significant effects. Use $\alpha = 0.05$

[5 marks]

(c). Based on the results of part (b), what settings would you use to maximise the yield? What is the expected value of yield at the optimum settings?

[3 marks]

(d). Write down the appropriate regression equation based on your results from part (b).

[2 marks]

(e). The engineer believes that 3 more factors (D,E,F) may also be of interest in the analysis, but is concerned about the increase in cost that this may incur. Explain the concept of a one half fractional factorial design 2^{6-1}. Using ABCD as the generator of the fraction determine the alias structure and corresponding design resolution.

[5 marks]

Q4. When designing an experiment involving the main effects A, B,C, D, and E at two levels each, the experimenters agreed that the only interactions likely to be active were AB and AC. Columns 1, 2, 4, 6, and 7 of L_8 (2^7) were used as a design matrix, which yielded the coded data in the following array.

Run	**A**	**B**	**AB**	**C**	**AC**	**D**	**E**	y_1	y_2
(1)	1	1	1	1	1	1	1	30	26
cde	1	1	1	2	2	2	2	32	27
bde	1	2	2	1	1	2	2	52	44
bc	1	2	2	2	2	1	1	27	29
ae	2	1	2	1	2	1	2	37	35
acd	2	1	2	2	1	2	1	1	6
abd	2	2	1	1	2	2	1	2	3
abce	2	2	1	2	1	1	2	21	12

$$S = \sum\sum y_{ij}^2 = 12768, \quad T = 384.$$

(a). Set up the response table for the seven effects and comment.

[4 marks]

(b). Draw the line plot for each effect and comment.

[4 marks]

(c). Construct an ANOVA table to test the findings for part (a) and (b) above. Use $\alpha = 0.05$.

[6 marks]

(d). Recommend optimal factor settings, assuming smaller is best, and generate the predicted value at these settings.

[3 marks]

(e). Explain the role of confirmatory runs. How would you interpret a confirmatory run value of 6.5?

[2 marks]

(f). What is the central idea behind the Taguchi philosophy?

[1 mark]

Q5. Data on sales (in thousands of units) last year in 14 sales districts are given below for a maker of asphalt roofing shingles. Initially the management believe that the number of competing brands in a particular district is an important independent variable. As a result the number of competing brands for each of the districts are also shown.

District	**x_1 = Comp Brands**	**y = Sales**
1	10	79.3
2	8	200.1
3	12	163.2
4	7	200.1
5	8	146.0
6	12	177.7
7	12	30.9
8	5	291.9
9	8	160.0
10	5	339.4
11	11	159.6
12	12	86.3
13	6	237.5
14	10	107.2

$$\sum x_{i1} = 126 \qquad \sum x_{i1}^2 = 1224 \qquad \sum (x_{i1} - \bar{x}_1)^2 = 90$$
$$\sum y_i = 2379.2 \qquad \sum y_i^2 = 493667 \qquad \sum (y_i - \bar{y})^2 = 89338.7$$
$$\sum (x_{i1} - \bar{x}_1)(y_i - \bar{y}) = -2264$$

(a). Calculate the coefficient of linear correlation and test it's significance using $\alpha = 0.05$.

[3 marks]

(b). Fit a linear regression model and test the significance of the model using ANOVA. State clearly the hypothesis being tested and your conclusions. Use $\alpha = 0.05$.

[4 marks]

(c). Calculate R^2 and comment.

[2 marks]

(d). Management then decide that promotional expenditures (x_2) and the number of active accounts (x_3) may also be relevant variables. The table below displays the correlation coefficients of y with all three variables x_1 (Number of competing brands), x_2 (Promotional expenditures) and x_3 (Number of active accounts).

	y	$\mathbf{x_1}$	$\mathbf{x_2}$	$\mathbf{x_3}$
y	1.0	-0.798	0.101	0.716

In a stepwise regression, which variable would be next to enter the model? Justify your answer.

[1 mark]

(e). The following model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i$$

was fitted. The partial output is as follows:

Predictor	**Coef**	**SE Coef**
Constant	178.562	8.262
Comp Brands (x_1)	-22.1272	0.5280
Promo Exp (x_2)	1.9480	0.6600
Active A/Cs (x_3)	3.57424	0.09940

Source	DF	SS	MS	F
Regression		89094		
Residual		245		
Total		89339		

Complete the ANOVA table. Clearly state the hypothesis being tested by the ANOVA table and your conclusion.

[4 marks]

(f). Test using $\alpha = 0.05$:

(i) $H_0: \ \beta_2 = 0$

(ii) $H_0: \ \beta_3 = 0$

[3 marks]

(g). What is the updated value of R^2? Comment.

[2 marks]

(h). Write down the multiple regression model.

[1 mark]

Formulae Sheet - 1 of 2

Parameter (Population Value)	**Statistic (Sample Value)**	**Standard Error**
μ	$\bar{x}$	$SE(\bar{x}) = \frac{s}{\sqrt{n}}$
μ_d	$\bar{x}_d$	$SE(\bar{x}_d) = \frac{s_d}{\sqrt{n}}$ **where** $s_d = \sqrt{\frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{n-1}}$
$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Small sample - n < 30:

Parameter (Population Value)	**Statistic (Sample Value)**	**Standard Error**
$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ **where** $s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$

Formulae sheet - 2 of 2

- $$\textbf{Taguchi: } SS = \frac{T_{High}^2 + T_{Low}^2}{N/2} - \frac{T^2}{N}$$

- $$\begin{aligned} SS_{Total} &= S - T^2/n \\ SS_{Factor1} &= \sum_i T_i^2/n_i - T^2/n \\ SS_{Factor2} &= \sum_j T_j^2/n_j - T^2/n \\ SS_{Interaction} &= \sum\sum T_{ij}^2/n_{ij} - \sum_i T_i^2/n_i - \sum_j T_j^2/n_j + T^2/n \end{aligned}$$

- $$LSD = t_{(\alpha/2,\textbf{df error})}\sqrt{MS_{Error}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

- $$r = \frac{\sum(x_{i1} - \bar{x}_1)(y_i - \bar{y})}{\sqrt{\sum(x_{i1} - \bar{x}_1)^2 \sum(y_i - \bar{y})^2}}$$

- $$\begin{aligned} SS_{xy} &= \sum(x_{i1} - \bar{x}_1)(y_i - \bar{y}) \\ SS_{xx} &= \sum(x_{i1} - \bar{x}_1)^2 \\ SS_{yy} &= \sum(y_i - \bar{y})^2 \end{aligned}$$

- $$\begin{aligned} \hat{\beta}_1 &= \frac{\sum(x_{i1} - \bar{x}_1)(y_i - \bar{y})}{\sum(x_{i1} - \bar{x}_1)^2} = \frac{SS_{xy}}{SS_{xx}} \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1\bar{x}_1 \end{aligned}$$

- $$SS_{Reg} = \sum(\hat{y}_i - \bar{y})^2 = (SS_{xy})^2/SS_{xx}$$

- $$Estimate = \frac{1}{2^{k-1}r}[Contrast] \qquad \frac{1}{2^k r}[Contrast]^2$$