Queen Mary
University of London

Main Examination period 2017

MTH6134 / MTH6134P: Statistical Modelling II

Duration: 2 hours

Student number									

Desk number			

Name and type of calculator used _____

Apart from this page, you are not permitted to read the contents of this question paper until instructed to do so by an invigilator.

Write your solutions in the spaces provided in this exam paper. If you need more paper, ask an invigilator for an additional booklet and attach it to this paper at the end of the exam.

You should attempt ALL questions. Marks available are shown next to the questions.

Only non-programmable calculators that have been approved from the college list of non-programmable calculators are permitted in this examination. Please state on your answer book the name and type of machine used.
The New Cambridge Statistical Tables are provided.

Complete all rough work in the answer book and cross through any work that is not to be assessed.

Possession of unauthorised material at any time when under examination conditions is an assessment offence and can lead to expulsion from QMUL. Check now to ensure you do not have any notes, mobile phones, smartwatches or unauthorised electronic devices on your person. If you do, raise your hand and give them to an invigilator immediately.

It is also an offence to have any writing of any kind on your person, including on your body. If you are found to have hidden unauthorised material elsewhere, including toilets and cloakrooms, it shall be treated as being found in your possession. Unauthorised material found on your mobile phone or other electronic device will be considered the same as being in possession of paper notes. A mobile phone that causes a disruption in the exam is also an assessment offence.

Exam papers must not be removed from the examination room.

Examiners: H. Maruri-Aguilar, J. Griffin

This page is for marking purposes only.

Do not write on it.

Question	Mark	Comments
1		
2		
3		
4		
Total		

Question 1. [25 marks] The effect of three chemicals to mitigate dry rot in potatoes was studied. The study involved samples from two farms and as potatoes from the same farm are considered similar, farms form a system of blocks. The block model for the amount of dry rot y_{ij} is $y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$ where $\alpha_1, \alpha_2, \alpha_3$ are the effects of the chemicals; β_1, β_2 are the effects of farm; μ is the mean and the errors $\varepsilon_{11}, \cdots, \varepsilon_{32}$ are assumed to be independent $N(0, \sigma^2)$ random variables. Data are given below.

Chemical	Farm	y
1	1	52
1	2	60
2	1	64
2	2	66
3	1	70
3	2	74

(a) Calculate the least squares estimates of the model parameters. You are not required to derive formulæ but to only use results seen in lectures. [7]

(b) The following two GenStat outputs are given.

Output 1

```
Estimates of parameters

Parameter  estimate  s.e.      t(2)  t pr.
Constant     64.333   0.882   72.95   <.001
W1            -8.33   1.25    -6.68   0.022
W2             0.67   1.25     0.53   0.646
U1           -2.333   0.882   -2.65   0.118
```

Output 2

```
Estimates of parameters

Parameter  estimate  s.e.    t(2)  t pr.
Constant     65.00   1.53   42.55  <.001
S1           -9.00   2.16   -4.17  0.053
S2            7.00   2.16    3.24  0.083
T1          -2.333   0.882  -2.65  0.118
```

One of the two given outputs is for the matrix version of the block model

$$Y = \tilde{X}\tilde{\beta} + \varepsilon, \tag{1}$$

as was described in the preamble of this question while the other output is a baseline parameterization

$$Y = X_b\beta_b + \varepsilon. \tag{2}$$

However it is **not** known which output corresponds to which of the two models above. Identify which output corresponds to which of models (1) and (2). For the baseline parameterization, also state which chemical is the baseline. Justify your answer. **[6]**

(c) For the matrix version of the block model, write down the matrix $\tilde{X}$ and parameter vector $\tilde{\beta}$ used. [**6**]

(d) For the baseline model, write down the matrix X_b and parameter vector β_b used. [**6**]

Question 2. [28 marks] A factorial experiment was designed to determine the effect of machining factors on the resistance of manufactured units, measured by the response variable ceramic strength (`strength`). Three factors were involved, each with two levels coded according to the table below.

Factor	Level	Coded level
Table Speed	0.025 m/s (slow)	-1
(`speed`)	0.125 m/s (fast)	1
Down Feed Rate	0.05 mm (slow)	-1
(`rate`)	0.125 mm (fast)	1
Wheel Grit	140/170	-1
(`grit`)	80/100	1

The analyst performed an analysis of variance using `GenStat`, whose output is the following:

```
Analysis of variance

Variate: strength

Source of variation d.f.     s.s.    m.s.    v.r.  F pr.
speed                 1   552.25  552.25  25.39  0.001
rate                  1    56.25   56.25   2.59  0.146
grit                  1   484.00  484.00  22.25  0.002
speed.rate            1     2.25    2.25   0.10  0.756
speed.grit            1     0.00    0.00   0.00  1.000
rate.grit             1   441.00  441.00  20.28  0.002
speed.rate.grit       1     4.00    4.00   0.18  0.679
Residual              8   174.00   21.75
Total                15  1713.75
```

(a) Give the correct instruction of what to put in the `GenStat` boxes below in order to analyze these data. [**4**]

Treatment Structure: ☐
Block Structure: ☐

Use the box below for your answer.

(b) Explain the principle of hierarchy in the context of testing effects in a factorial experiment. [**4**]

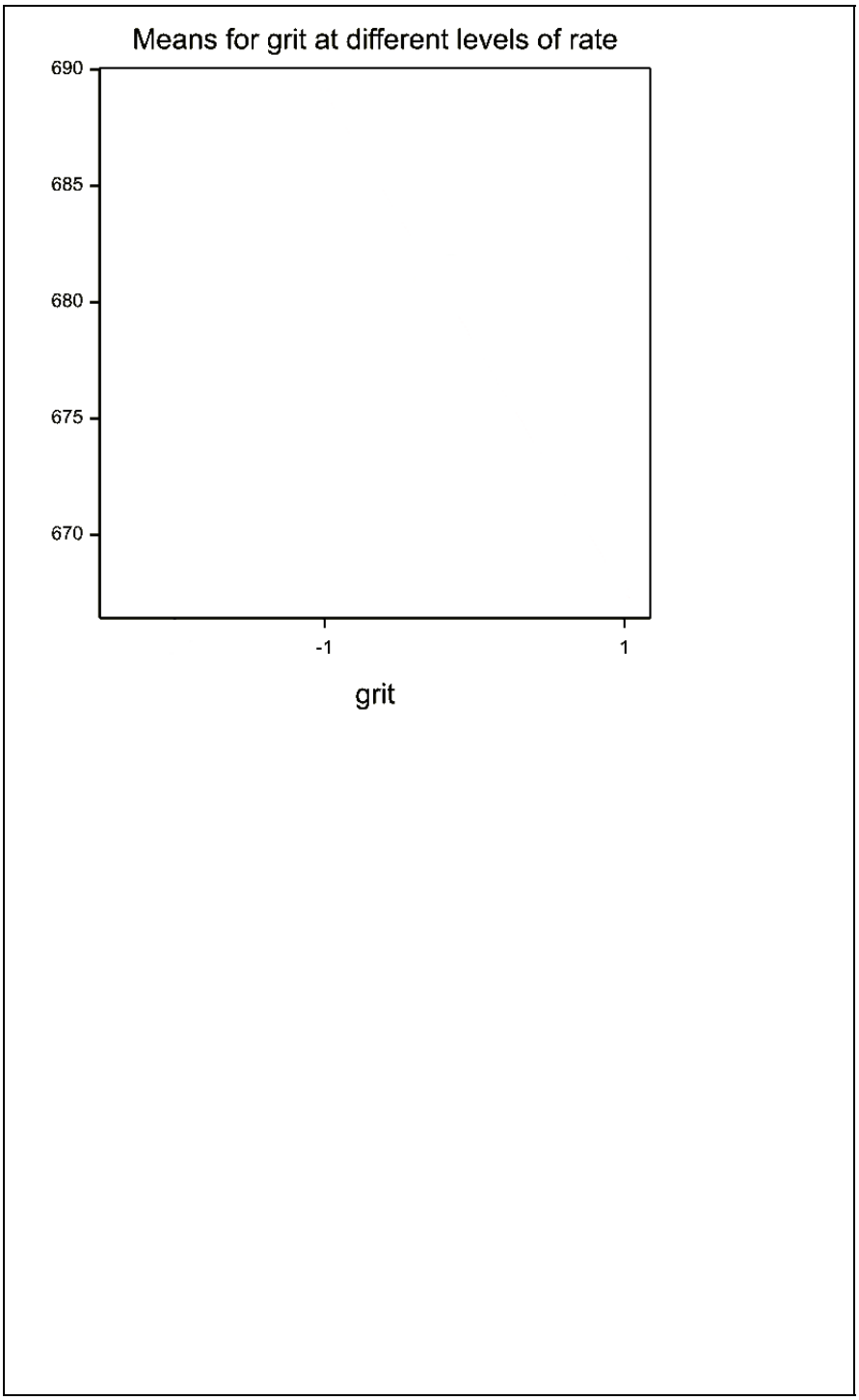
(c) Using the principle of hierarchy and the anova table given, perform hypotheses tests about the effects of machining factors and their double and triple interactions. In your response explain which tests you perform, in which sequence and why. [**8**]

(d) The following table has the results of the experiment:

speed	rate	grit	strength	
-1	-1	-1	683	685
1	-1	-1	692	696
-1	1	-1	671	680
1	1	-1	688	690
-1	-1	1	657	666
1	-1	1	671	676
-1	1	1	674	678
1	1	1	682	693

Compute the table of means per treatment combination of factors `rate` and `grit`, i.e. the table of means to study the interaction between factors `rate` and `grit`. [**5**]

(e) Using the previous table of means, complete the interaction plot for factors `rate` and `grit`. Use the figure provided in the box and write your interpretation of the plot. [7]



Question 3. [25 marks] The following data set gives the number of warp breaks per loom, where a loom corresponds to a fixed length of yarn. Three levels of tension (L,M,H) were considered, and there is interest in comparing the mean number of breaks caused by different levels of tension. Summary statistics are $\sum_{i=1}^{3}\sum_{j=1}^{18} y_{ij} = 1520$ and $\sum_{i=1}^{3}\sum_{j=1}^{18} y_{ij}^2 = 52018$, and the data are shown below.

Tension	Number of breaks	Total
L	26 30 54 25 70 52 51 26 67 27 14 29 19 29 31 41 20 44	655
M	18 21 29 17 12 18 35 30 36 42 26 19 16 39 28 21 39 29	475
H	36 21 24 18 10 43 28 15 26 20 21 24 17 13 15 15 16 28	390

(a) Compute the analysis of variance table and test factor tension. [**10**]

(b) Describe the Tukey-Kramer method for multiple comparisons. [**4**]

(c) Give at least one advantage of Tukey-Kramer against the LSD method. Here you need not define LSD. [**2**]

(d) For this data set, perform the multiple comparisons using the Tukey-Kramer method. See the Appendix for the table of the studentized range distribution. [**9**]

Question 4. [22 marks] A completely randomized design has three treatments with replication pattern (3,2,3).

(a) Define the vector spaces V_0 and V_T. [**8**]

(b) Compute projections $P_{V_0}Y$, $P_{V_T}Y$ and $P_{W_T}Y$ for the data vector

$$Y = \begin{pmatrix} 26 \\ 25 \\ 21 \\ 10 \\ 6 \\ 30 \\ 8 \\ 10 \end{pmatrix}.$$

[**9**]

(c) Show numerically that $P_{V_0}Y$ and $P_{W_T}Y$ are orthogonal. [**5**]

End of Paper—An appendix of 1 page follows.

Appendix 1

The table below shows $q(0.05; t, n-t)$, i.e. the upper $\alpha = 0.05$ quantile of the studentized range distribution.

$n-t$	t			
	2	3	4	5
40	2.858232	3.442082	3.790685	4.039123
41	2.856062	3.438874	3.786726	4.034566
42	2.853999	3.435823	3.782961	4.030232
43	2.852033	3.432918	3.779376	4.026105
44	2.850160	3.430150	3.775958	4.022172
45	2.848372	3.427507	3.772697	4.018417
46	2.846664	3.424983	3.769581	4.014830
47	2.845031	3.422569	3.766601	4.011400
48	2.843467	3.420258	3.763749	4.008116
49	2.841969	3.418044	3.761016	4.004969
50	2.840532	3.415921	3.758395	4.001952
51	2.839152	3.413883	3.755879	3.999055
52	2.837827	3.411925	3.753463	3.996273
53	2.836553	3.410043	3.751139	3.993598
54	2.835327	3.408232	3.748904	3.991024
55	2.834147	3.406489	3.746752	3.988545
56	2.833010	3.404809	3.744678	3.986158
57	2.831914	3.403189	3.742679	3.983855
58	2.830856	3.401627	3.740750	3.981634
59	2.829835	3.400118	3.738888	3.979490
60	2.828848	3.398661	3.737089	3.977418

End of Appendix.