MTH6134 / MTH6134P: Statistical Modelling II

Duration: 2 hours

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

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: D. S. Coad, H. Maruri-Aguilar

Turn Over

Question 1. **[27 marks]** A fast food franchise is test marketing three new menu items. To find out if they have the same popularity, 18 franchisee restaurants are randomly chosen for participation in the study. Six of the restaurants are randomly chosen to test market each of the new menu items. Below are the sales figures, in number of orders per day, of the three new menu items in these restaurants after a week of test marketing.

Item	Sales						Total
1	22	42	44	52	45	37	242
2	52	33	8	47	43	32	215
3	16	24	19	18	34	39	150

The sum of squares of the observations is $\sum_{i=1}^{3}\sum_{j=1}^{6} y_{ij}^2 = 23,415$.

(a) Write down a suitable model for these data and any necessary assumptions, explaining your notation. **[5]**

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

(c) Using an appropriate constraint, write down in matrix form the multiple regression model that is equivalent to the analysis of variance model for the data. **[6]**

(d) By using the model in part (c), find the least squares estimates of the parameters. **[6]**

Question 2. [22 marks] An engineer conducted a randomised block experiment to compare three metals, nickel, iron and copper, for bonding agents for an alloy material. Components of the alloy were bonded using the metals and the pressures required, in pascals, to break the bonds were measured. Seven ingots of the alloy material were available for the experiment and the engineer recognised the possibility of variability between the ingots. Thus, the engineer constructed bonds of components from each ingot using each metal. The data are given below.

	Metal		
Ingot	Nickel	Iron	Copper
1	67.0	71.9	72.2
2	67.5	68.8	66.4
3	76.0	82.6	74.5
4	72.7	78.1	67.3
5	73.1	74.2	73.2
6	65.8	70.8	68.7
7	75.6	84.9	69.0

An analysis of variance was performed using `GenStat` and the output is as follows:

```
Analysis of variance

Variate: pressure

Source of variation d.f.    s.s.     m.s.   v.r.   F pr.
metal                 2    131.90   65.95  6.36   0.013
ingot                 6    268.29   44.71  4.31   0.015
Residual             12    124.46   10.37
Total                20    524.65
```

(a) Briefly explain how you would enter these data into `GenStat`. What expression should be entered in Treatment Structure in the analysis of variance Dialogue Box? **[4]**

(b) Draw conclusions from the above output. **[4]**

(c) Use the least significant difference method to compare pairs of metal means. **[9]**

(d) Analyse the data ignoring the block structure to test factor metal. **[5]**

Turn Over

Question 3. [18 marks] A factorial experiment was designed for making microwave popcorn. The effect of time of cooking, in minutes, and microwave power setting, as a percentage of the total available, each at two levels, was studied, and the response variable was the amount of inedible "bullets" or unpopped kernels. Below are the results of two replicates of the experiment.

Time	Power	Bullets	
4	75	3.5	3.1
4	100	0.7	0.7
6	75	1.6	1.2
6	100	0.3	0.5

The model to be used for these data is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

for $i = 1, 2$, $j = 1, 2$ and $k = 1, 2$, where α_i is the effect of the ith time, β_j is the effect of the jth power, and γ_{ij} is the effect of interaction between the ith time and the jth power. The constraints on the parameters are $\sum_{i=1}^{2} \alpha_i = 0$, $\sum_{j=1}^{2} \beta_j = 0$, $\sum_{i=1}^{2} \gamma_{ij} = 0$ and $\sum_{j=1}^{2} \gamma_{ij} = 0$, and it is assumed that $\varepsilon_{ijk} \sim N(0, \sigma^2)$, all independent.

(a) Derive expressions for the least squares estimates of μ, α_i, β_j and γ_{ij}.   [**10**]

(b) Derive expressions for the fitted values and residuals.   [**3**]

(c) Write down contrasts for testing for the main effects of time and power. Construct another contrast which is orthogonal to these. What is this contrast testing?   [**5**]

Question 4. [21 marks] A company is interested in testing the uniformity of its film-coated pain tablets. Three batches were randomly sampled from each of its two blending sites and five tablets were assayed from each batch. Below are the responses, in micrometres.

Site		1			2	
Batch	1	2	3	1	2	3
	5.03	4.64	5.10	5.05	5.46	4.90
	5.10	4.73	5.15	4.96	5.15	4.95
	5.25	4.82	5.20	5.12	5.18	4.86
	4.98	4.95	5.08	5.12	5.18	4.86
	5.05	5.06	5.14	5.05	5.11	5.07

The sum of squares of the observations is $\sum_{i=1}^{2} \sum_{j=1}^{3} \sum_{k=1}^{5} y_{ijk}^2 = 763.8188$.

(a) Write down a suitable model for these data and any necessary assumptions, explaining your notation.   [**5**]

(b) Compute the analysis of variance table, and test factors site and batch.   [**13**]

(c) Estimate the variance components.   [**3**]

Question 5. [12 marks] Consider a completely randomised design with four treatments. Let the data vector be $\mathbf{y} = (y_{11}, y_{12}, y_{21}, y_{22}, y_{31}, y_{32}, y_{41}, y_{42})^\top$.

(a) Define the treatment subspace V_T. [4]

(b) Compute the projections $P_{V_T}\mathbf{y}$ and $P_{V_T^\perp}\mathbf{y}$. [4]

(c) In the usual analysis of variance terminology, what are the names of the projections that you computed in part (b)? [2]

(d) Define the sum of squares and the degrees of freedom for $V_T^\perp$. [2]

End of Paper.