Queen Mary
University of London

MTH6134 / MTH6134P: Statistical Modelling II

Duration: 2 hours

Date and time: 1st June 2016, 10:00–12:00

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 awarded are shown next to the questions.

Statistical functions provided by the calculator may be used provided that you state clearly where you have used them.
The New Cambridge Statistical Tables are provided.

Complete all rough workings 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.

Examiner(s): H. Maruri-Aguilar and J. Míguez Arenas

Question 1 (25 marks). An experiment in teaching performance intends to compare scores for two modules. The study was carried out in three schools selected at random from a large candidate set. The modules are considered to be equal among schools so a standard block design was used.

The score is modelled as $y_{ij} = \mu + \alpha_i + b_j + \varepsilon_{ij}$, where $b_j \sim N(0, \sigma_B^2)$ and $\varepsilon_{ij} \sim N(0, \sigma^2)$ are independent random terms. The indices range as $i = 1, 2$ for modules and $j = 1, 2, 3$ for schools. The covariance structure of the model is also given

$$\text{Cov}(y) = \begin{pmatrix} \sigma^2+\sigma_B^2 & 0 & 0 & \sigma_B^2 & 0 & 0 \\ 0 & \sigma^2+\sigma_B^2 & 0 & 0 & \sigma_B^2 & 0 \\ 0 & 0 & \sigma^2+\sigma_B^2 & 0 & 0 & \sigma_B^2 \\ \sigma_B^2 & 0 & 0 & \sigma^2+\sigma_B^2 & 0 & 0 \\ 0 & \sigma_B^2 & 0 & 0 & \sigma^2+\sigma_B^2 & 0 \\ 0 & 0 & \sigma_B^2 & 0 & 0 & \sigma^2+\sigma_B^2 \end{pmatrix},$$

where the vector of reponses is $y = \begin{pmatrix} y_{11} & y_{12} & y_{13} & y_{21} & y_{22} & y_{23} \end{pmatrix}^T$. Means per treatment (module) are defined as usual as $y_{1\cdot} = T_1/3$ and $y_{2\cdot} = T_2/3$ with $T_i = \sum_{j=1}^{3} y_{ij}$.

(a) For these data and this model, there is only one contrast of interest $L = \alpha_1 - \alpha_2$. Show that $\hat{L} = y_{1\cdot} - y_{2\cdot}$ is an unbiased estimator of L. **[2]**

(b) Compute $\text{V}(y_{1\cdot})$. **[4]**

(c) Using your previous result for $\text{V}(y_{1\cdot})$ and the facts that $\text{V}(y_{1\cdot}) = \text{V}(y_{2\cdot})$ and $\text{Cov}(y_{1\cdot}, y_{2\cdot}) = \sigma_B^2/3$, find $\text{V}(\hat{L})$. **[4]**

(d) Using the results for constrasts seen in lectures, the sum of squares for treatments is written in terms of $\hat{L}$ as $S_T = \frac{3}{2}\hat{L}^2$. Find $E(S_T)$. **[6]**

(e) Copy the anova table below and complete the missing information. **[9]**

Source	SS	d.o.f.	EMS	Null hypothesis
Module (tmnt)	S_T			H_0 :
School (block)	S_B		$\sigma^2 + 2\sigma_B^2$	H_0 :
Residual	S_E			
Total	S_G			

Question 2 (26 marks). A researcher is interested in determining the influence of genre and artist in the lengths of music samples. He selected 3 music genres and for each genre, 2 artists. From each artist, he used 2 music samples to give a total of 12 samples for his study.

The genres under study were selected at random from a group of available genres and within each genre, the artists were also selected at random from a collection of artists available. In this setting, an artist in a given music genre is unrelated to another artist in a different genre.

The data are given below, with lengths in minutes and fractions of minutes suitably converted to decimal format.

Genre	Artist	Sample	Length
Blues	1	1	9.31
Blues	1	2	2.6
Blues	2	1	8.44
Blues	2	2	5.1
Jazz	3	1	2.1
Jazz	3	2	3.62
Jazz	4	1	3.61
Jazz	4	2	1.87
Disco	5	1	2.3
Disco	5	2	1.89
Disco	6	1	1.79
Disco	6	2	1.78

(a) Write a suitable model for the data and describe its component elements. **[6]**

(b) Compute the relevant sums of squares and complete the analysis of variance table. **[10]**

(c) Perform the relevant hypothesis tests for factors Genre and Artist. **[6]**

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

Treatment Structure:	
Block Structure:	

Question 3 (26 marks). An industrial bread maker wants to compare the relative humidity in bread loaves produced under five different processes. He sampled two loaves per process. The data and means are given below.

Process	1		2		3		4		5	
Humidity	14.2	13.9	11	10.7	11	9.9	11.1	10.6	13.5	14.1
Mean	14.05		10.85		10.45		10.85		13.80	

In addition, he performed `GenStat` fixed effects analysis which produced the following output.

```
Analysis of variance

Variate: Humidity

Source of variation  d.f.     s.s.     m.s.    v.r.   F pr.
Process                 4  24.9800   6.2450   31.23   <.001
Residual                5   1.0000   0.2000
Total                   9  25.9800
```

(a) Interpret the `GenStat` output. **[4]**

(b) The bread maker is interested in performing multiple comparisons between processes. Describe the Tukey-Kramer method and as part of your solution, explain why this method is preferable to the LSD method. Note that you do not need to describe the LSD method. **[4]**

(c) For the above data and $\alpha = 0.05$, perform multiple comparisons using the Tukey-Kramer method and draw conclusions. A table of the studentized range distribution is provided in the next page. **[10]**

(d) Write the hypothesis to be tested with Bartlett's test and briefly explain it. **[4]**

(e) An intern in the factory produced partial computations to compare variances using Bartlett's test for this data. He computed $K = 2.451$ and $L = 0.4$. Write the test statistic for Bartlett's test as a function of K and L and compute its numerical value, then perform the test and conclude. **[4]**

The next page contains an appendix for part (c) of this question

Appendix for Question 3: Percentiles of the studentized range distribution for $1-\alpha=0.95$.

Every entry in the table is $q(1-\alpha;\gamma,\nu)$ where
$P\{q(\gamma,\nu)\le q(1-\alpha;\gamma,\nu)\}=1-\alpha$ with $\alpha=0.05$.

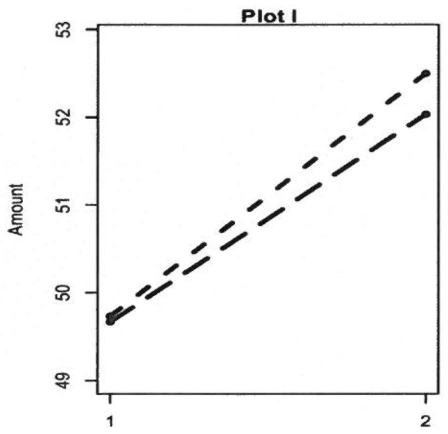
ν \ γ	2	3	4	5	6	7	8	9	10	11
2	6.08	8.33	9.80	10.88	11.73	12.43	13.03	13.54	13.99	14.40
3	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46	9.72
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83	8.03
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.65
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05
9	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.87
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.61

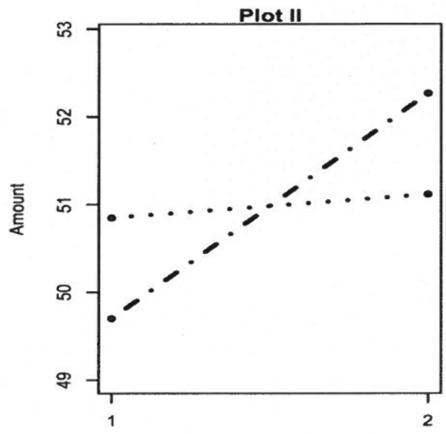
Question 4 (23 marks). The mechanism of certain enzymatic reaction depends upon amounts of two reactants termed A and B. An experiment was performed, using two levels of each A and B which were coded as 1,2. For each combination, three replications were created and the amount of material y produced (in grams) per 100 grams of reactants is shown below.

A	1	1	1	1	1	1	2	2	2	2	2	2
B	1	2	1	2	1	2	1	2	1	2	1	2
y	49.9	51.9	49.8	52.2	49.3	52.0	50.9	52.3	49.3	52.9	49.0	52.3

(a) Describe an appropriate model for this data. [5]

(b) For this data, plots I and II below were produced. One of the plots shows means per treatment combination (interaction plot) and the other shows means per factor level (main effects plot). However it is **not** known which is the main effects plot and which is the interaction plot and your task is to disentangle this. Note that different types of lines in the plots do not have special meaning.





For each of plots I and II,

(i) identify the type of plot (main effects, interaction), [4]

(ii) explain what each of the lines in the plot represent, label the horizontal axis, and [4]

(iii) interpret the plot. [4]

(c) Compute least squares parameter estimates for the model you described in (a). Here you are only required to use the formulæ and not to derive them. [4]

(d) Give the formula for the standard error of the difference of means per treatment combination. [2]

End of Paper.