BSc DEGREES AND GRADUATE DIPLOMAS IN ECONOMICS, MANAGEMENT, FINANCE AND THE SOCIAL SCIENCES, THE DIPLOMA IN ECONOMICS AND SOCIAL SCIENCES AND THE CERTIFICATE IN EDUCATION IN SOCIAL SCIENCES

Summer 2021 Online Assessment Instructions

ST3188 Statistical Methods for Market Research

Tuesday, 4 May 2021: 06:00 – 10:00 (BST)

The assessment will be an **open-book take-home online assessment within a 4-hour window**. The requirements for this assessment remain the same as the closed-book exam, with an expected time/effort of **2 hours.**

Candidates should answer the **ONE** question in Section A and **TWO** questions from Section B. Section A carries 40 marks. Questions in Section B carry 30 marks each. Candidates are strongly advised to divide their time accordingly.

You should complete this paper **using pen and paper.** Please use **BLACK INK** only.

Handwritten work then needs to be scanned, converted to PDF and then uploaded to the VLE as **ONE individual file** including the coversheet. Each scanned sheet should have your **candidate number** written clearly in the header. Please **do not write your name anywhere** on your submission.

You have until 10:00 (BST) on Tuesday, 4 May 2021 to upload your file into the VLE submission portal. However, you are advised not to leave your submission to the last minute.

Workings should be submitted for all questions requiring calculations. Any necessary assumptions introduced in answering a question are to be stated.

You may use *any* calculator for any appropriate calculations, but you may not use any computer software to obtain solutions. Credit will only be given if all workings are shown.

If you think there is any information missing or any error in any question, then you should indicate this but proceed to answer the question stating any assumptions you have made.

The assessment has been designed with a duration of 4 hours to provide a more flexible window in which to complete the assessment and to appropriately test the course learning outcomes. As an open-book exam, the expected amount of effort required to complete all questions and upload your answers during this window is no more than 2 hours. Organise your time well.

You are assured that there will be no benefit in you going beyond the expected 2 hours of effort. Your assessment has been carefully designed to help you show what you have learned in the hours allocated.

This is an open book assessment and as such you may have access to additional materials including but not limited to subject guides and any recommended reading. But the work you submit is expected to be 100% your own. Therefore, unless instructed otherwise, you must not collaborate or confer with anyone during the assessment. The University of London will carry out checks to ensure the academic integrity of your work. Many students that break the University of London's assessment regulations did not intend to cheat but did not properly understand the University of London's regulations on referencing and plagiarism. The University of London considers all forms of plagiarism, whether deliberate or otherwise, a very serious matter and can apply severe penalties that might impact on your award.

The University of London 2020-21 Procedure for the consideration of Allegations of Assessment Offences is available online at:

Assessment Offence Procedures - University of London

SECTION A: Compulsory

1. (a) As countries around the world embark on rolling out their respective Covid-19 vaccination programmes, many 'physical' businesses (i.e. those not online) are eager for customers to return. A national restaurant chain which operates in many cities recognises that, despite the vaccines, some people may remain concerned about dining out in one of their restaurants.

    The company has a database of members of its loyalty scheme, many of whom have not visited a restaurant for over a year. While the company wants these customers to return, it also wants to win back 'casual diners', i.e. those who might visit a restaurant only occasionally.

    To better understand potential customer anxieties, the company's management has decided to use a survey of all types of customers and has asked you to devise an appropriate sampling scheme. Explain in detail how each of the following sampling methods could be applied to the overall sampling strategy for this study. Make sure you describe the merits and limitations of each as well as how each would be applied in practice.

    i. Judgemental sampling.

    ii. Quota sampling.

    iii. Simple random sampling.

    iv. Cluster sampling.

    (20 marks)

    (b) Suppose we are interested in estimating the mean of a population using a simple random sample of size n. *In your own words*, answer the following.

    i. State a suitable estimator of the population mean as well as its sampling distribution. Mention any assumptions which you make.

    ii. Explain statistically how to determine the minimum sample size necessary to estimate a population mean to within e units with 90% confidence.

    iii. When several parameters are being estimated, what is the procedure for determining the sample size?

    iv. Explain the difference between the *incidence rate* and the *completion rate* in survey sampling.

    (20 marks)

SECTION B: Answer two questions. Each question carries equal weight.

2. (a) An e-commerce retailer suspects that not all consumers are equally confident with the security of shopping online. To investigate this, a survey of $n = 30$ consumers asked for their confidence rating (on a 1 to 100 scale, with a score of 100 representing complete confidence) about purchasing products online and asked for some demographic information (classified by gender and age group). A two-way analysis of variance is conducted.

   Analyse the selected SPSS output in Figure 1 (on the next page) and discuss what conclusions can be drawn from the data. In your analysis, be sure to address at least the following:

   - Describe the strength of the joint effect of the factors.

   - Test the significance of the variables individually and the interaction between them and interpret the results.

   - How you would use the results to inform your decision about convincing consumers to feel safer shopping online.

   (20 marks)

   (b) *In your own words*, answer the following. **Write a maximum of 250 words in total.**

   Define a 'response rate' and 'non-response'. What strategies are available for minimising non-response and adjusting for it? In your response, consider how these are tackled across different data collection methods.

   (10 marks)

Figure 1

Descriptive Statistics

Dependent Variable: Confidence

Gender	Age group	Mean	Std. Deviation	N
Male	18-30	81.80	6.870	5
	31-60	68.80	5.020	5
	61+	59.00	5.831	5
	Total	69.87	11.128	15
Female	18-30	91.60	3.647	5
	31-60	79.40	6.504	5
	61+	53.40	7.092	5
	Total	74.80	17.383	15
Total	18-30	86.70	7.319	10
	31-60	74.10	7.824	10
	61+	56.20	6.795	10
	Total	72.33	14.559	30

Tests of Between-Subjects Effects

Dependent Variable: Confidence

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Model	162260.800[a]	6	27043.467	764.300	.000
Gender	182.533	1	182.533	5.159	.032
Age_group	4698.067	2	2349.033	66.388	.000
Gender * Age_group	416.867	2	208.433	5.891	.008
Error	849.200	24	35.383		
Total	163110.000	30			

Estimated Marginal Means of Confidence
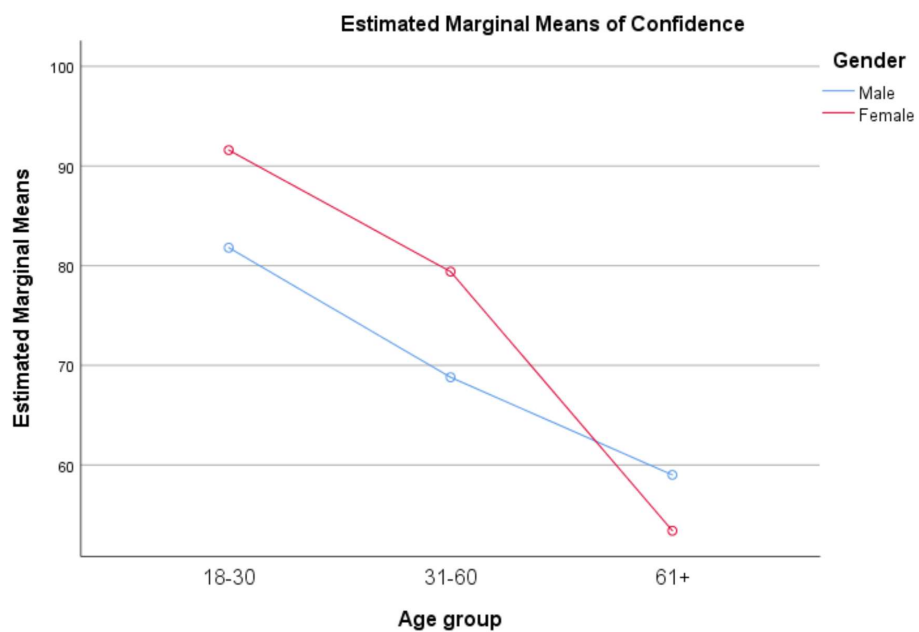
3. (a) A supermarket chain has stores in three separate regions: The North, Central and The South. To maximise sales overall, it is important that each store caters for the tastes and preferences of its consumers by offering the right product mix at each store. The supermarket chain wants to identify which products typically comprise the baskets of customers in the different regions. Ultimately, they wish to identify the optimal mix of products to stock in stores in each region. Five products were selected to be members of a representative basket:

   * Bread and bakery
   * Dairy and milk
   * Fresh fruit
   * Wines and spirits
   * Fresh meat.

   Analyse the selected SPSS output in Figure 2 (spread over the next two pages) and discuss what conclusions can be drawn from the data. In your analysis, be sure to address at least the following:

   • State the theoretical and estimated discriminant analysis models.
   • Comment on the relative importance of the predictor variables.
   • Comment on any other predictor variables not included in the model which you think might be relevant.
   • Determine the predictive accuracy of the model.

   (20 marks)

   (b) *In your own words*, answer the following. **Write a maximum of 250 words in total.**

      i. Describe the semantic differential scale and the Likert scale. For what purposes are these scales used? Provide an example of each scale.

      ii. How does the nature and degree of verbal description affect the response to itemised rating scales?

      iii. Construct a simple example of a question with an itemised rating scale.

   (10 marks)

Figure 2

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	2.793[a]	96.9	96.9	.858
2	.090[a]	3.1	100.0	.288

a. First 2 canonical discriminant functions were used in the analysis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.242	78.095	10	.000
2	.917	4.765	4	.312

Structure Matrix

	Function 1	Function 2
Bread and bakery	.677[*]	-.166
Wines and spirits	.658[*]	-.139
Dairy and milk	.124[*]	.005
Fresh meat	.427	.830[*]
Fresh fruit	.410	-.704[*]

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions
Variables ordered by absolute size of correlation within function.

*. Largest absolute correlation between each variable and any discriminant function

Standardized Canonical Discriminant Function Coefficients

	Function 1	Function 2
Bread and bakery	.498	-.016
Dairy and milk	.030	.050
Fresh fruit	.321	-.535
Wines and spirits	.506	-.130
Fresh meat	.456	.726

Canonical Discriminant Function Coefficients

	Function 1	Function 2
Bread and bakery	.222	-.007
Dairy and milk	.027	.046
Fresh fruit	.091	-.152
Wines and spirits	.181	-.047
Fresh meat	.182	.290
(Constant)	-10.819	-.290

Unstandardized coefficients

Functions at Group Centroids

Region	Function 1	Function 2
The North	-2.202	-.122
Central	.513	.404
The South	1.689	-.282

Unstandardized canonical discriminant functions evaluated at group means

Figure 2 (continued)

Canonical Discriminant Functions
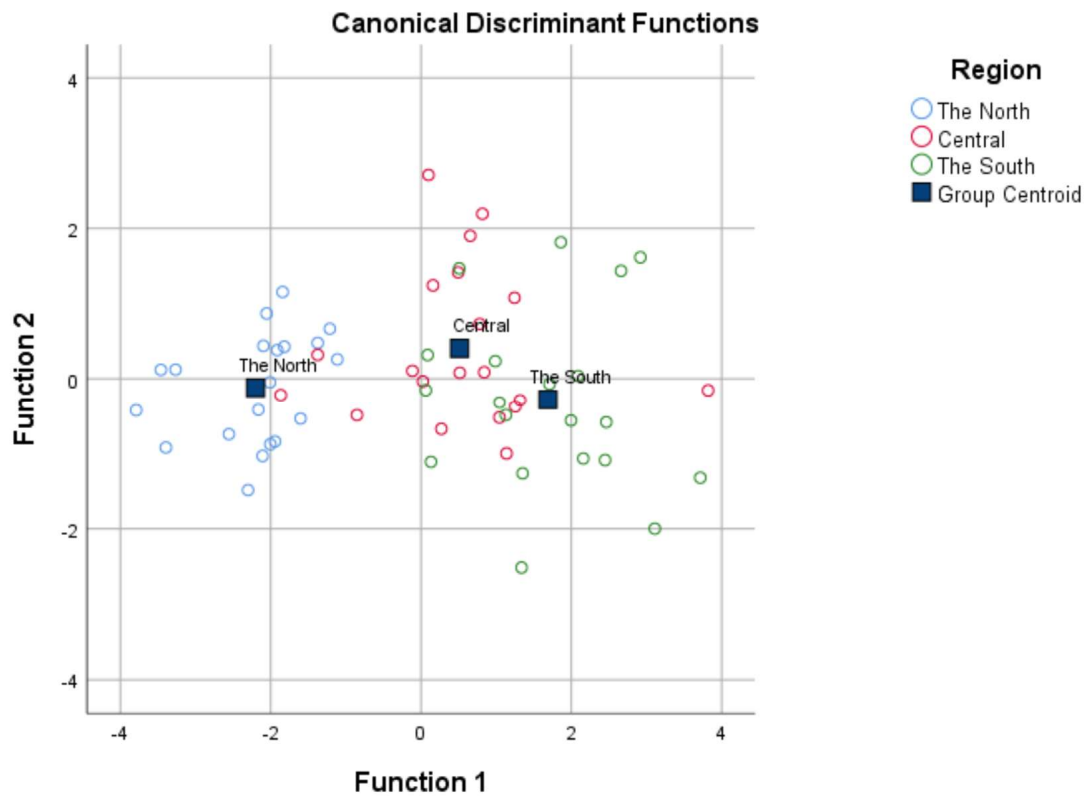
Classification Results[a,c]

| | | | Predicted Group Membership | | | |
		Region	The North	Central	The South	Total
Original	Count	The North	20	0	0	20
		Central	3	12	5	20
		The South	0	6	14	20
	%	The North	100.0	.0	.0	100.0
		Central	15.0	60.0	25.0	100.0
		The South	.0	30.0	70.0	100.0
Cross-validated[b]	Count	The North	20	0	0	20
		Central	3	12	5	20
		The South	0	7	13	20
	%	The North	100.0	.0	.0	100.0
		Central	15.0	60.0	25.0	100.0
		The South	.0	35.0	65.0	100.0

4. (a) A publisher of textbooks is interested in whether countries can be clustered based on the:

- percentage of eligible people pursuing lifelong learning
- percentage of GDP spent on education.

Data were obtained on a sample of 20 countries.

Figure 3 (spread over the next two pages) presents selected SPSS output from a cluster analysis using Ward's procedure and squared Euclidean distance. Interpret the output. In your analysis, be sure to address at least the following:

- Explain how Ward's procedure employed in the cluster analysis works.
- Explain how the squared Euclidean distance between objects i and j is calculated when a cluster analysis is based on p objects.
- Explain, with reasons, the appropriate number of clusters according to the SPSS output.

(20 marks)

(b) *In your own words*, answer the following. **Write a maximum of 250 words in total.**

Explain the three different types of research design and provide an example of when each would be appropriate.
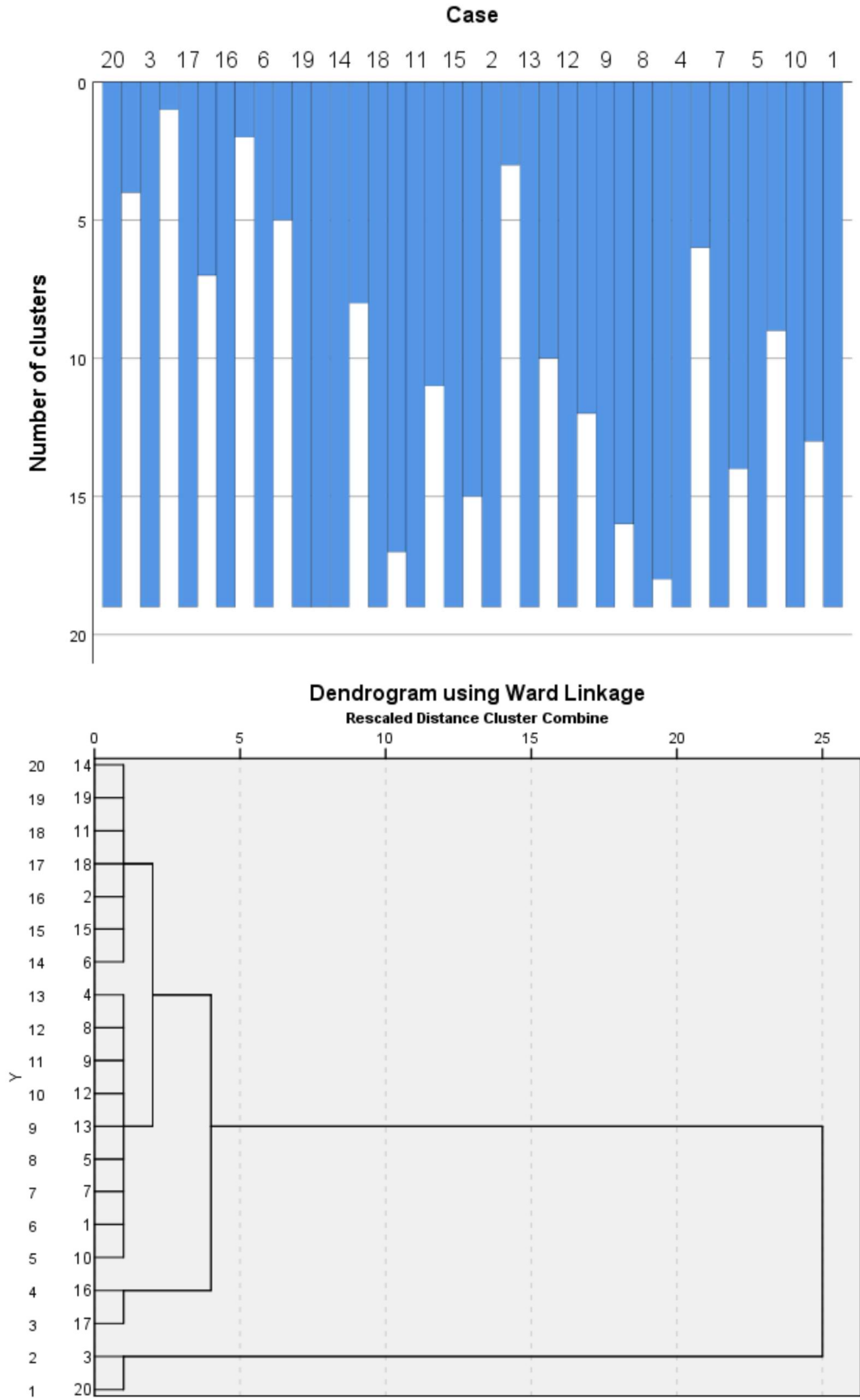
(10 marks)

Figure 3

Agglomeration Schedule

	Cluster Combined			Stage Cluster First Appears		
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
1	14	19	.007	0	0	12
2	4	8	.018	0	0	4
3	11	18	.093	0	0	9
4	4	9	.204	2	0	8
5	2	15	.372	0	0	9
6	5	7	.711	0	0	11
7	1	10	1.150	0	0	11
8	4	12	1.901	4	0	10
9	2	11	2.755	5	3	12
10	4	13	3.652	8	0	14
11	1	5	5.272	7	6	14
12	2	14	8.001	9	1	15
13	16	17	11.162	0	0	18
14	1	4	16.843	11	10	17
15	2	6	24.317	12	0	17
16	3	20	32.753	0	0	19
17	1	2	67.500	14	15	18
18	1	16	192.094	17	13	19
19	1	3	990.630	18	16	0

Ward Method		Percentage of eligible people doing lifelong learning	Percentage of GDP spent on education
1	Mean	5.9813	5.0956
	N	16	16
	Std. Deviation	1.79823	.70228
2	Mean	27.9000	6.8800
	N	2	2
	Std. Deviation	1.83848	2.24860
3	Mean	14.3500	5.3150
	N	2	2
	Std. Deviation	1.76777	.19092
Total	Mean	9.0100	5.2960
	N	20	20
	Std. Deviation	7.15423	.97743

Ward Method		Percentage of eligible people doing lifelong learning	Percentage of GDP spent on education
1	Mean	6.9111	5.1200
	N	18	18
	Std. Deviation	3.21886	.66510
2	Mean	27.9000	6.8800
	N	2	2
	Std. Deviation	1.83848	2.24860
Total	Mean	9.0100	5.2960
	N	20	20
	Std. Deviation	7.15423	.97743

Figure 3 (continued)

Case

Dendrogram using Ward Linkage

END OF PAPER