



**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: 15:00 – 19:00 (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 19:00 (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 global cinema chain which operates in many countries worldwide recognises that, despite the vaccines, some people may remain concerned about watching a new film release in one of their cinemas.

The company has a database of members of its loyalty scheme, many of whom have not visited a cinema for over a year. While the company wants these customers to return, it also wants to win back 'casual customers', i.e. those who might visit a cinema 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. Convenience sampling.
- ii. Snowball sampling.
- iii. Systematic sampling.
- iv. Stratified sampling.

**(20 marks)**

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

- i. Explain what a sampling distribution is.
- ii. State a suitable estimator of the population proportion as well as its sampling distribution. Mention any assumptions which you make.
- iii. Explain statistically how to determine the minimum sample size necessary to estimate a population proportion to within  $e$  units.
- iv. When constructing a confidence interval for a proportion, which level of confidence would you propose and why?

**(20 marks)**

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

2. (a) A car manufacturer is aware that some drivers are concerned about the amount of noise pollution created by cars. As such, the manufacturer has decided to investigate what factors may influence the amount of noise. A sample of 36 cars was obtained and for each car its size was recorded (as small, medium or large) and its engine type (normal or new).

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 manufacturing cars which are marketed as being quieter than other cars.

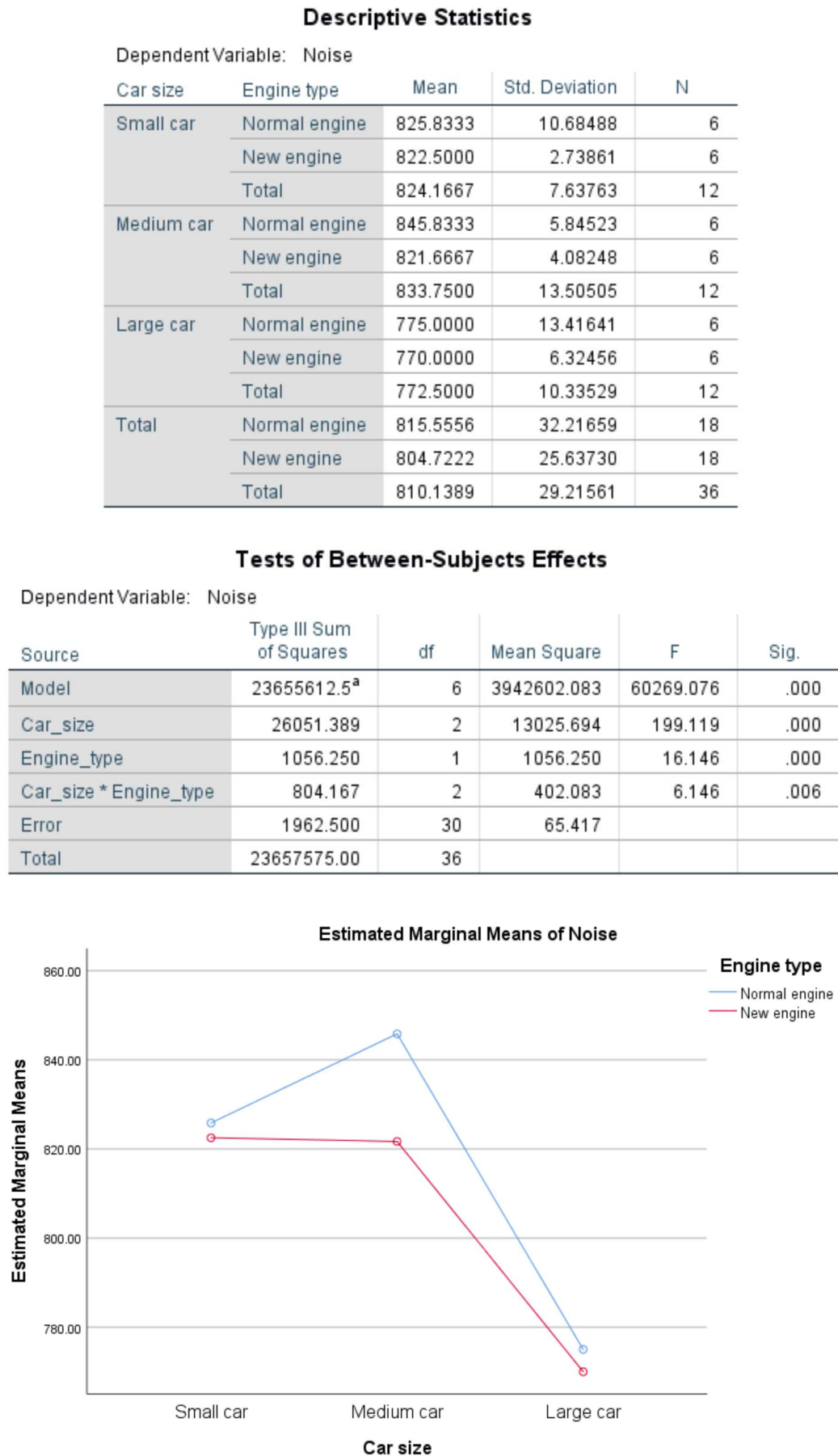
**(20 marks)**

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

Would you argue that a Likert scale variable should be considered to have an ordinal or interval level of measurement? Justify your choice.

**(10 marks)**

Figure 1



3. (a) A large recruitment agency has decided to investigate whether different job classifications appealed to different personality types. Three job classifications were considered and indexed as follows: (1) customer-facing, (2) technical roles, and (3) roles requiring people to work on their own. Psychometric tests were conducted on 165 employees measuring their interest in outdoor activities, interest in social events and their degree of conservativeness.

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 the suitability of including the ‘highest level of education’ variable.
- 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.**

Explain how you would adjust for non-response in market research surveys.

**(10 marks)**

Figure 2

**Eigenvalues**

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	1.289 <sup>a</sup>	87.8	87.8	.750
2	.179 <sup>a</sup>	12.2	100.0	.390

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	.371	159.825	6	.000
2	.848	26.498	2	.000

**Structure Matrix**

	Function	
	1	2
SOCIAL	-.762 <sup>*</sup>	.231
CONSERVATIVE	.472 <sup>*</sup>	-.289
OUTDOOR	.278	.954 <sup>*</sup>

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	2
OUTDOOR	.370	.927
SOCIAL	-.833	.181
CONSERVATIVE	.556	-.255

**Canonical Discriminant Function Coefficients**

	Function	
	1	2
OUTDOOR	.086	.214
SOCIAL	-.195	.042
CONSERVATIVE	.166	-.076
(Constant)	1.165	-3.214

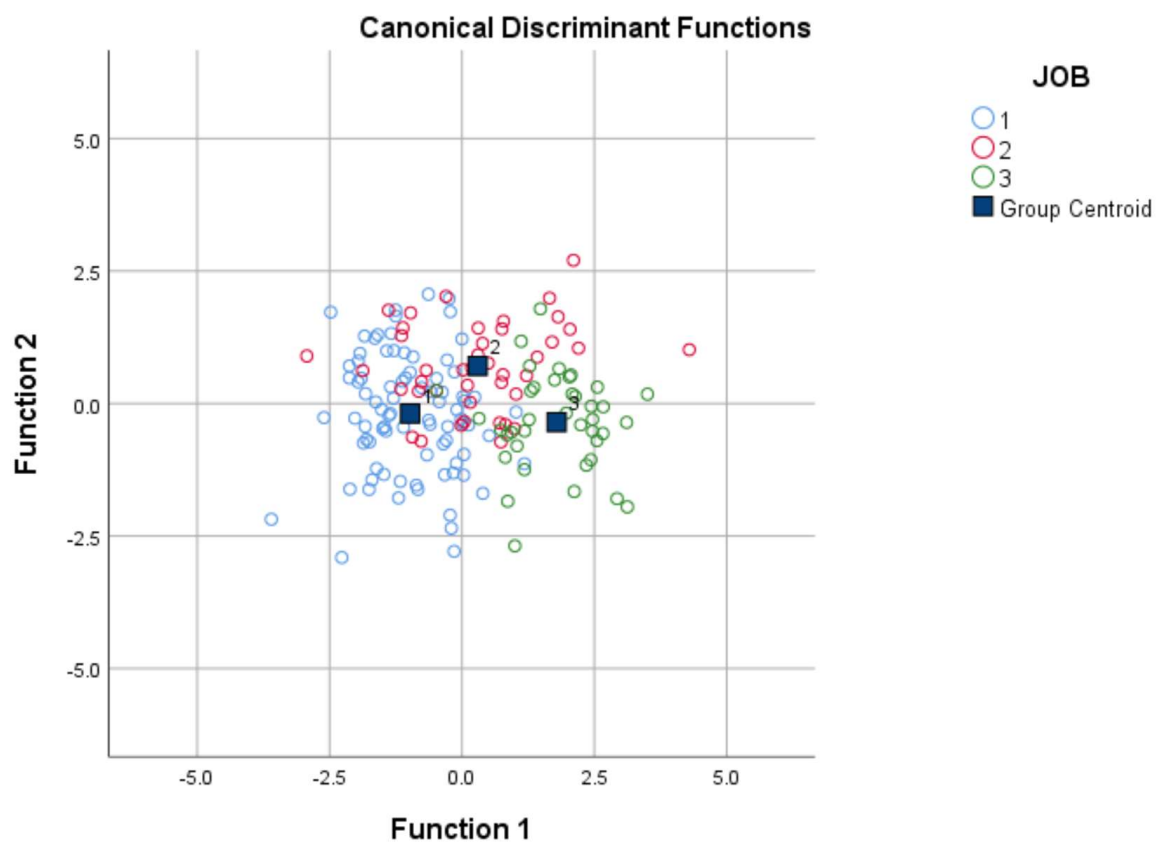
Unstandardized coefficients

**Functions at Group Centroids**

JOB	Function	
	1	2
1	-.979	-.191
2	.297	.709
3	1.788	-.352

Unstandardized canonical discriminant functions evaluated at group means

Figure 2 (continued)



**Classification Results<sup>a,c</sup>**

		Predicted Group Membership				Total
		JOB	1	2	3	
Original	Count	1	68	12	4	84
		2	12	23	7	42
		3	1	4	34	39
	%	1	81.0	14.3	4.8	100.0
		2	28.6	54.8	16.7	100.0
		3	2.6	10.3	87.2	100.0
Cross-validated <sup>b</sup>	Count	1	67	13	4	84
		2	14	19	9	42
		3	1	4	34	39
	%	1	79.8	15.5	4.8	100.0
		2	33.3	45.2	21.4	100.0
		3	2.6	10.3	87.2	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.**

In a multiple linear regression model, how would you determine the relative importance of the independent variables? Make sure that you discuss any possible limitations with your suggestion(s).

**(10 marks)**

Figure 3

**Agglomeration Schedule**

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
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)

