UNIVERSITY OF LONDON

ST3188

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 2022 Online Assessment Instructions

ST3188 Statistical methods for market research

Friday 6 May 2022: 09:00 - 12:00 (BST)

The assessment will be a **closed-book take-home online assessment within a 3-hour window**. The expected time/effort to answer all questions is **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 exam platform as ONE individual file. Please ensure that your candidate number is written clearly at the top of each page included in the scan. Please do not write your name anywhere on your submission.

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.

You have until 12:00 (BST) on Friday 6 May 2022 to submit your answers. However, you are advised not to leave your submission to the last minute in order to allow sufficient time to submit your work.

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 **3 hours** to provide a more flexible window in which to complete the assessment. As a closed-book exam, the

expected amount of effort required to complete all questions is no more than **2 hours.** Organise your time well. You are assured that in terms of answering all questions, 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.

By accessing this question paper, you agree not to commit any assessment offence. Assessment offences include (but are not limited to) **committing plagiarism** and the **use or access of any paid-for or any other services offering live assistance** during an examination. You must not confer with anyone else during a live examination; and we take conferring to include any exchange of information or discussion about the assessment with others in any way that could potentially give you or another student an advantage in the examination. As such, any exchanging with others of exam questions; or any accessing of websites, blogs, forums **or any other form of oral or written communication with others which involves any discussion of live examination questions or potential answers/solutions to exam questions** will be considered an assessment offence.

The University of London will conduct 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's *Procedure for the Consideration of Allegations of Assessment Offences* is available online at:

Assessment Offence Procedures - University of London

UNIVERSITY OF LONDON

ST3188

ST3188 Statistical methods for market research

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.

Please find questions on the following page.

SECTION A: Compulsory

1. (a) Globally, registrations of new electric cars have increased in recent years, despite the pandemic, as motorists start to embrace all-electric vehicles to reduce CO_2 emissions. However, many motorists have concerns such as price, a lack of availability of public charging points, and driving range limitations on a single battery charge.

   You work in the marketing department of a global car manufacturer that is in the early stages of developing their electric vehicle range of models. The company wants to better understand drivers' opinions about electric cars and their intention to purchase an electric car when they next trade in their vehicle. The company suspects that sentiment may vary from country to country, as well as by gender and age. The company has a database of customers who have previously bought at least one of their non-electric vehicles which includes their contact details and main demographic attributes.

   To better understand where to target their research and development (R&D) budget, 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. Quota sampling.

   ii. Snowball 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. Explain why it would be preferable to estimate the mean of a population in a market research context, rather than estimating the median or mode.

   ii. State a suitable estimator of the population mean as well as its sampling distribution. Mention any assumptions which you make and define all terms used.

   iii. Explain statistically how to determine the minimum sample size necessary to estimate a population mean to within e units, defining all terms used.

   iv. Based on your answer to part iii., briefly explain how you would choose numerical values for each term in the sample size determination problem using any market research example of your choice.

   (20 marks)

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

2. (a) A credit card company wants to understand the variation in credit card charges accrued by its customers. While the company profits from these charges, it is concerned that customers with very high charges may be at greater risk of default – a bad outcome for both the customer and the company. It was decided to perform a multiple linear regression.

   A random sample of $n = 5{,}000$ credit card customers was obtained, with information on the following variables:

   - Annual charges, in $
   - Annual income, in $000s
   - Age
   - Household size
   - Sex (1 = female, 0 = male)
   - Exceeded credit limit in past 12 months? (1 = yes, 0 = no).

   Selected SPSS output is provided in Figure 1 (on the next page). Analyse the regression results. In your analysis, be sure to address at least the following:

   - Write out the full regression model, including any assumptions, and the estimated model.
   - Interpret the estimated coefficient of the 'Exceeded credit limit in past 12 months?' variable.
   - Comment on the relative importance of the predictor variables.
   - Briefly discuss any changes you would recommend making to the model.

   (20 marks)

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

   In a two-way analysis of variance (ANOVA), explain how you would determine the relative importance of the two factors.

   (10 marks)

Figure 1

Model Summary[b]

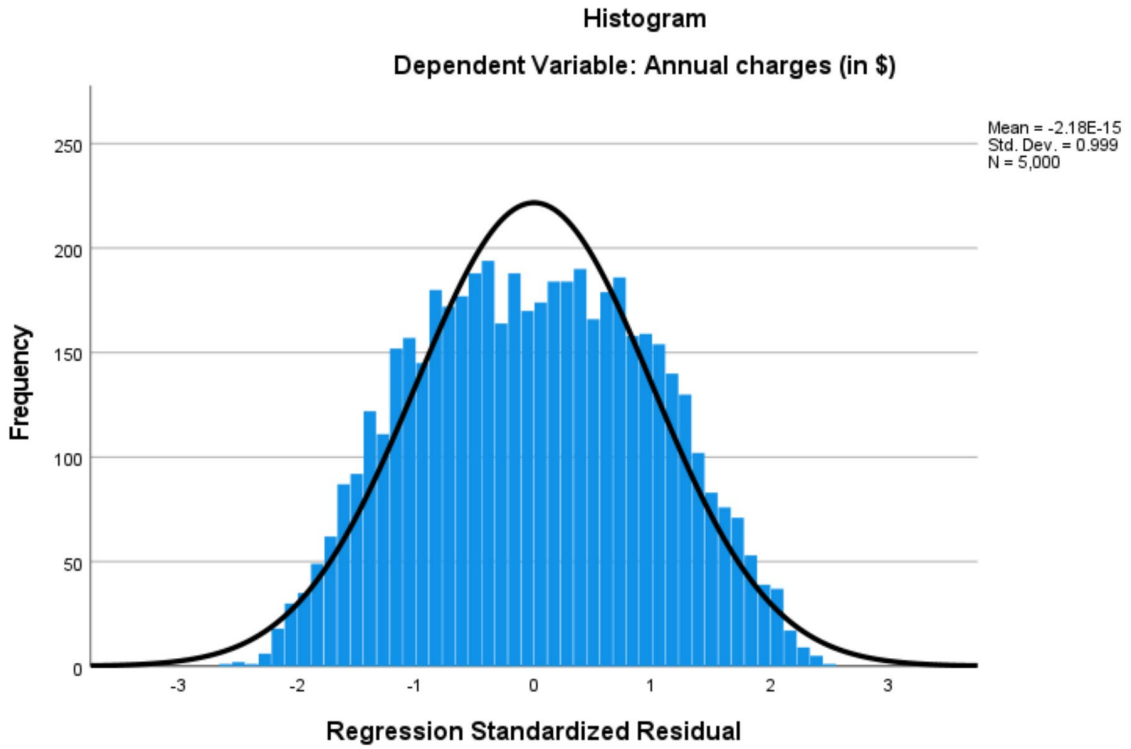
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.591[a]	.349	.348	4903.278836

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6.439E+10	5	1.288E+10	535.669	.000[b]
	Residual	1.201E+11	4994	24042143.34		
	Total	1.845E+11	4999			

a. Dependent Variable: Annual charges (in $)

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	791.790	435.050		1.820	.069
	Annual income (in $000s)	120.882	2.468	.559	48.987	.000
	Age	-9.256	6.470	-.016	-1.431	.153
	Household size	528.069	33.458	.180	15.783	<.001
	Sex	-198.509	139.359	-.016	-1.424	.154
	Exceeded credit limit in past 12 months?	543.669	238.152	.026	2.283	.022



3. (a) A well-known retailer recently conducted a customer satisfaction survey, with $n = 200$ respondents, to determine satisfaction levels across the following attributes:

- price satisfaction
- product quality satisfaction
- product range satisfaction
- packaging satisfaction
- customer service satisfaction.

Each attribute was scored on a 5-point Likert scale, ranging from 1 = strongly negative to 5 = strongly positive.

The retailer wants to identify the underlying dimensions which explain the correlations among responses to the above five attributes, so decided to undertake a factor analysis.

Figure 2 (spread over the next two pages) presents selected SPSS output from a factor analysis with principal components extraction, using the varimax rotation procedure. Interpret the output. In your analysis, be sure to address at least the following:

- Explain how you determine the number of factors and interpret the extracted factor(s).
- Comment on the suitability of the 5-point Likert scale for performing a factor analysis.
- Explain whether you recommend the use of any surrogate variables instead of an extracted factor.
- Briefly discuss for what purpose(s) any extracted factors and/or surrogate variables could be used.

(20 marks)

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

Explain the purpose of *online focus groups* in market research, and describe their main characteristics.

(10 marks)

Figure 2

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.659
Bartlett's Test of Sphericity	Approx. Chi-Square	179.863
	df	10
	Sig.	<.001

Communalities

	Initial	Extraction
Price satisfaction	1.000	.861
Product quality satisfaction	1.000	.653
Product range satisfaction	1.000	.723
Packaging satisfaction	1.000	.249
Customer service satisfaction	1.000	.686

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	2.133	42.665	42.665
2	1.038	20.763	63.427
3	.936	18.717	82.144
4	.524	10.487	92.631
5	.368	7.369	100.000

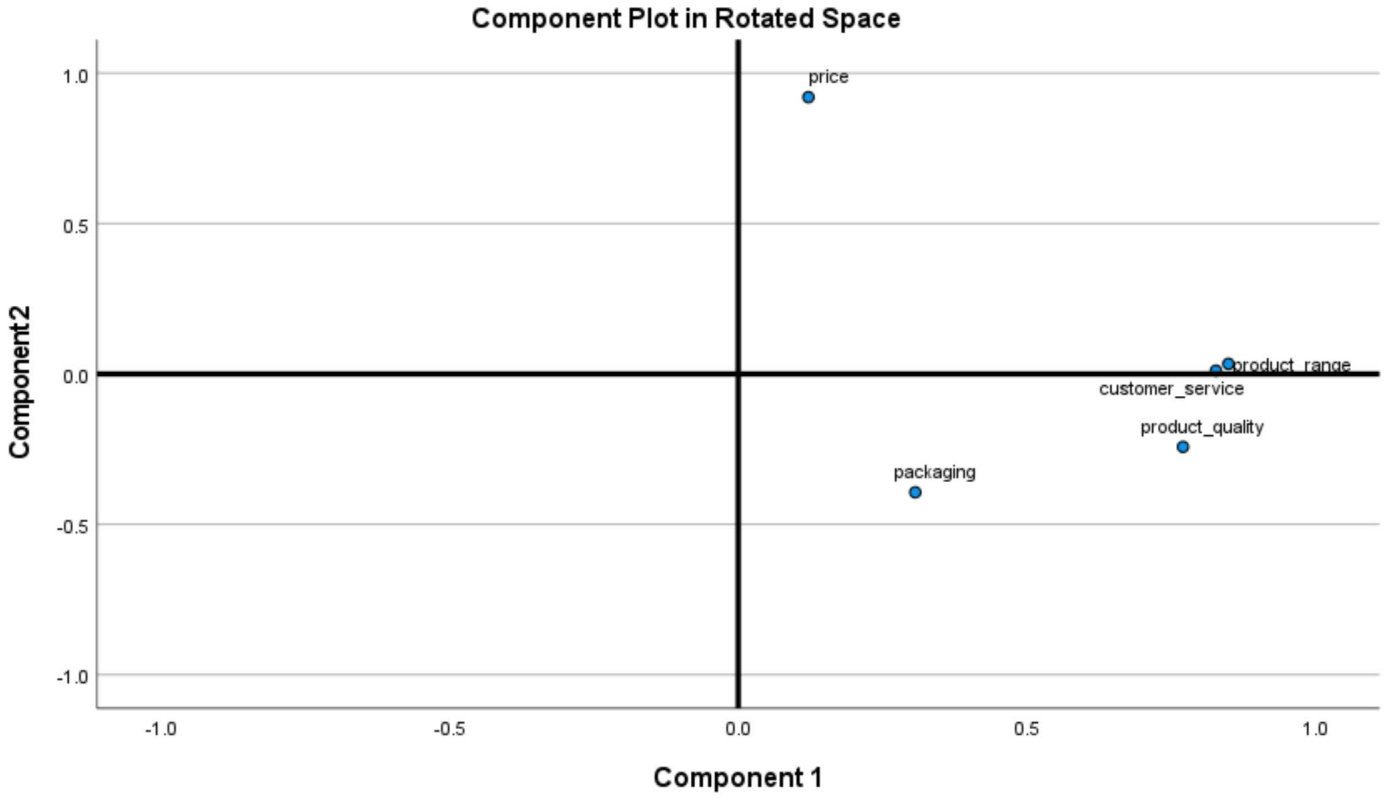


Figure 2 (continued)

Rotated Component Matrix[a]

	Component	
	1	2
Price satisfaction	.122	.920
Product quality satisfaction	.771	-.242
Product range satisfaction	.850	.034
Packaging satisfaction	.307	-.393
Customer service satisfaction	.828	.011

Component Score Coefficient Matrix

	Component	
	1	2
Price satisfaction	.124	.885
Product quality satisfaction	.352	-.176
Product range satisfaction	.410	.093
Packaging satisfaction	.119	-.353
Customer service satisfaction	.398	.069

Reproduced Correlations

		Price satisfaction	Product quality satisfaction	Product range satisfaction	Packaging satisfaction	Customer service satisfaction
Reproduced Correlation	Price satisfaction	.861[a]	-.129	.134	-.325	.111
	Product quality satisfaction	-.129	.653[a]	.647	.332	.635
	Product range satisfaction	.134	.647	.723[a]	.247	.704
	Packaging satisfaction	-.325	.332	.247	.249[a]	.250
	Customer service satisfaction	.111	.635	.704	.250	.686[a]
Residual[b]	Price satisfaction		.037	-.102	.291	-.058
	Product quality satisfaction	.037		-.092	-.139	-.183
	Product range satisfaction	-.102	-.092		-.163	-.124
	Packaging satisfaction	.291	-.139	-.163		-.024
	Customer service satisfaction	-.058	-.183	-.124	-.024	

4. (a) An entrepreneur is considering opening a new 'fast fashion' company, which would regularly use promotions to attract budget-conscious consumers. To assist with decision-making, the entrepreneur decided to carry out some exploratory research to identify any distinct market segments related to brand loyalty and appetite for promotions.

A short questionnaire was completed by a sample of 25 people. The following seven statements required a response on a 7-point Likert scale, ranging from 1 = strongly disagree to 7 = strongly agree.

- I am more likely to buy brands with a promotion.
- Gamified promotions are fun.
- I spend more money with brands when they offer a promotion.
- I feel more emotionally connected to a brand when it offers a promotion.
- I find brands more appealing when they offer a promotion.
- I am more likely to notice a brand when it offers a promotion.
- I am more loyal to a brand once it's offered a promotion (i.e. I don't switch to other brands afterwards).

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, with reasons, the appropriate number of clusters according to the SPSS output.
- Profile the clusters. Hint: You may wish to conduct the profiling based on the age and income variables.
- Briefly discuss for what purpose(s) any identified clusters could be used.

(20 marks)

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

Explain the purpose of a *paired comparison scale* and a *constant sum scale* in market research, and provide an example of each.

(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	4	14	1.000	0	0	9
2	10	12	3.500	0	0	9
3	8	25	6.500	0	0	10
4	11	16	9.500	0	0	7
5	1	18	13.500	0	0	14
6	2	5	17.500	0	0	14
7	11	21	21.833	4	0	21
8	7	23	27.333	0	0	11
9	4	10	33.083	1	2	12
10	8	17	39.417	3	0	13
11	6	7	45.917	0	8	19
12	4	19	53.067	9	0	17
13	8	9	61.983	10	0	22
14	1	2	70.983	5	6	19
15	3	22	80.983	0	0	20
16	20	24	91.483	0	0	18
17	4	15	104.083	12	0	20
18	13	20	118.917	0	16	21
19	1	6	138.488	14	11	23
20	3	4	161.988	15	17	22
21	11	13	187.821	7	18	23
22	3	8	217.821	20	13	24
23	1	11	248.904	19	21	24
24	1	3	494.960	23	22	0

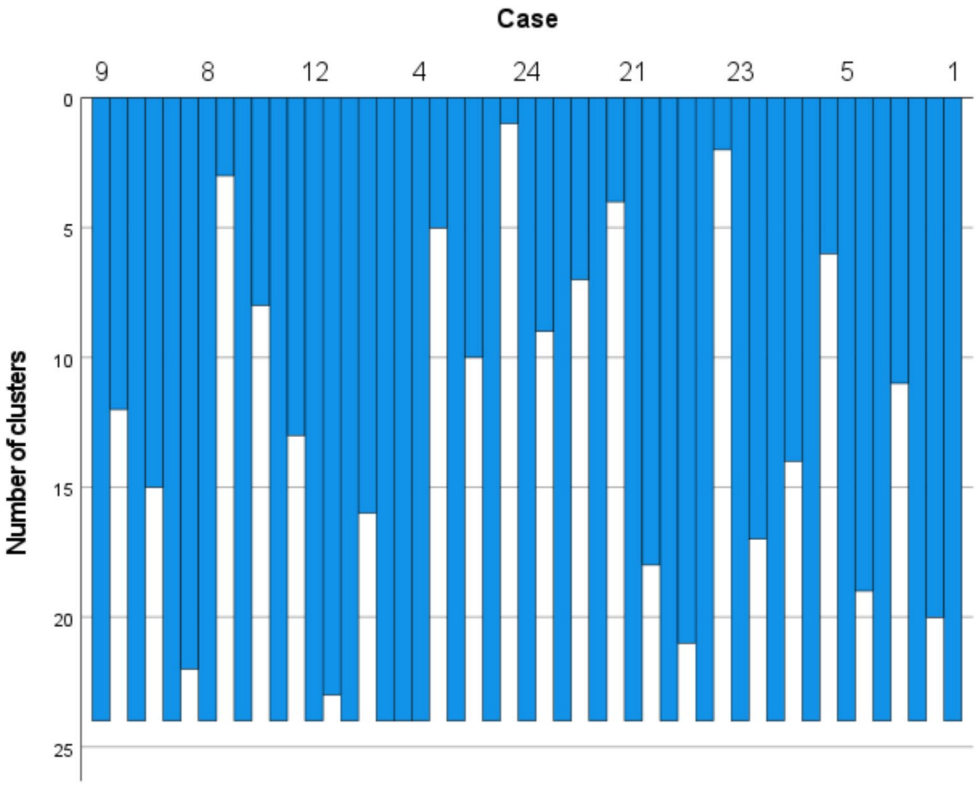
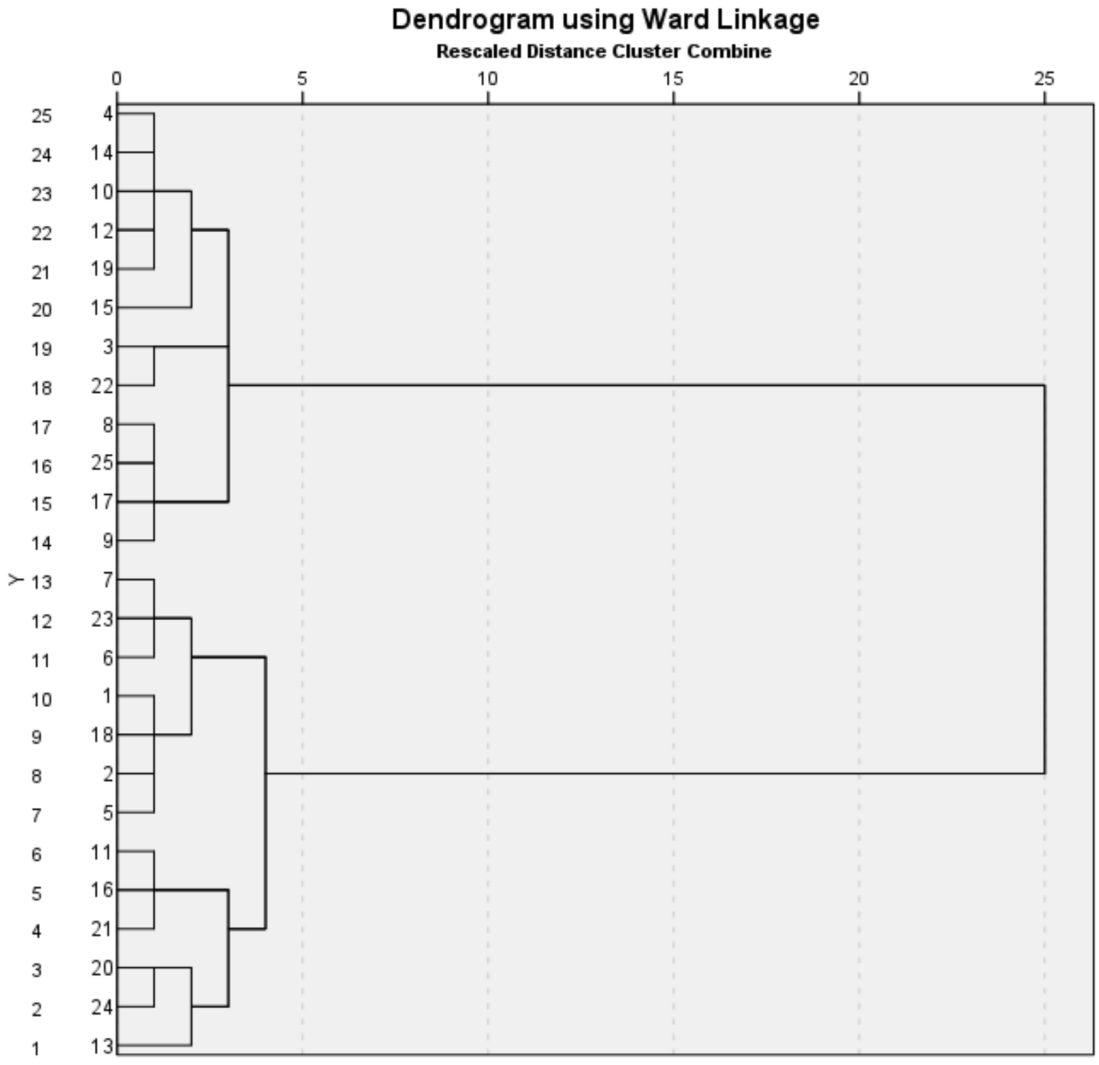


Figure 3 (continued)



Ward Method		I am more likely to buy brands with a promotion	Gamified promotions are fun	I spend more money with brands when they offer a promotion	I feel more emotionally connected to a brand when it offers a promotion	I find brands more appealing when they offer a promotion	I am more likely to notice a brand when it offers a promotion	I am more loyal to a brand once it's offered a promotion (i.e. I don't switch to other brands afterwards)	Age	Income in $000s
1	Mean	3.15	3.46	3.54	3.38	3.15	3.54	3.00	54.08	62.69
	N	13	13	13	13	13	13	13	13	13
	Std. Deviation	1.144	1.391	1.330	1.325	1.405	1.050	1.291	10.805	14.250
2	Mean	5.75	5.67	5.67	5.83	5.08	6.17	5.58	27.33	29.50
	N	12	12	12	12	12	12	12	12	12
	Std. Deviation	.965	1.155	1.435	1.115	1.505	1.030	1.084	4.716	5.486
Total	Mean	4.40	4.52	4.56	4.56	4.08	4.80	4.24	41.24	46.76
	N	25	25	25	25	25	25	25	25	25
	Std. Deviation	1.683	1.686	1.734	1.734	1.730	1.683	1.763	15.954	20.044

[END OF PAPER]