# Examiners' commentaries 2022

## ST3188 Statistical methods for market research

## Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2021–22. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

## Information about the subject guide and the Essential reading references

Unless otherwise stated, all cross-references will be to the latest version of the course (2021). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refer to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

## General remarks

### Learning outcomes

At the end of the course and having completed the essential reading and activities you should be able to:

- define a market research problem and create an appropriate research design
- perform independent data analysis in a market research setting
- determine which statistical method is appropriate in a given situation and be able to discuss the merits and limitations of a particular method
- use statistical software to analyse datasets and be able to interpret output
- draw appropriate conclusions following empirical analysis and use to form the basis of managerial decision-making
- demonstrate greater commercial awareness.

### Format of the examination

The examination is two hours long and you must answer the question in Section A and two questions out of three in Section B. The examination is worth 70% of the final grade. The other 30% is determined by the coursework component. (The coursework comprised the production of a market research proposal – see the 'Assessment' section in the VLE for details.)

### Overall performance

The performance of candidates in the examination was generally pleasing, with some excellent answers. Use of SPSS is not directly examined, rather some questions in Section B required the interpretation of SPSS output. Some answers lacked sufficient depth of explanation – remember to comment in detail on the output statistically. For example, when reporting on $p$-values explicitly right out the hypotheses being tested, i.e. $H_0$ and $H_1$. An excellent answer would also state the test statistic being used and relate this to the relevant test statistic value.

Although this is an applied statistics course, candidates are reminded that commercial insight is also important. Always think about which business decisions could be taken as a consequence of the market research, justifying the decision(s) based on the results – the course is about market research after all! It is likely any decision will relate to one (or more) of the marketing mix variables – the four 'p's (product, price, placement and promotion).

## Examination revision strategy

Many candidates are disappointed to find that their examination performance is poorer than they expected. This may be due to a number of reasons, but one particular failing is '**question spotting**', that is, confining your examination preparation to a few questions and/or topics which have come up in past papers for the course. This can have serious consequences.

We recognise that candidates might not cover all topics in the syllabus in the same depth, but you need to be aware that examiners are free to set questions on **any aspect** of the syllabus. This means that you need to study enough of the syllabus to enable you to answer the required number of examination questions.

The syllabus can be found in the Course information sheet available on the VLE. You should read the syllabus carefully and ensure that you cover sufficient material in preparation for the examination. Examiners will vary the topics and questions from year to year and may well set questions that have not appeared in past papers. Examination papers may legitimately include questions on any topic in the syllabus. So, although past papers can be helpful during your revision, you cannot assume that topics or specific questions that have come up in past examinations will occur again.

**If you rely on a question-spotting strategy, it is likely you will find yourself in difficulties when you sit the examination. We strongly advise you not to adopt this strategy.**

# Examiners' commentaries 2022

## ST3188 Statistical methods for market research

## Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2021–22. The format and structure of the examination may change in future years, and any such changes will be publicised on the virtual learning environment (VLE).

## Information about the subject guide and the Essential reading references

Unless otherwise stated, all cross-references will be to the latest version of the course (2021). You should always attempt to use the most recent edition of any Essential reading textbook, even if the commentary and/or online reading list and/or subject guide refer to an earlier edition. If different editions of Essential reading are listed, please check the VLE for reading supplements – if none are available, please use the contents list and index of the new edition to find the relevant section.

## Comments on specific questions – Zones A and B

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.**

**Section A: Compulsory**

**Question 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)**

**Reading for this question**

Block 9 on the VLE covers sampling – design and procedures.

**Approaching the question**

Candidates should avoid generic 'textbook' descriptions of the named sampling techniques. Rather, it is necessary to *explain* how each sampling scheme may be used in the specified application (i.e. researching motorist concerns about electric cars in this question).

Clearly distinguish between non-probability and probability methods. For the latter, the sampling frame should be identified, and any possible limitations mentioned – such as any difference(s) between the target population and sampling population, and the implications.

As explicitly mentioned in the question, as well as the mechanics of each method the merits and limitations should be stated. For example, the non-probability methods suffer from selection bias, while cluster sampling requires that the clusters are sufficiently homogeneous – is this likely in practice?

While there is no single 'right' answer to such a question, the examiners rewarded responses which directly related to the motorists' concerns about electric cars – in particular the different types of motorists (by country, gender and age).

As with any sampling, for the results to be meaningful the objective is to obtain a representative sample, which different sampling techniques achieve to varying degrees. The extent of representativeness for each technique should be addressed.

(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)**

**Reading for this question**

Block 10 on the VLE covers sampling distributions and sample size determination.

**Approaching the question**

i. Here the examiners expected a discussion of the mean as the preferred measure of central tendency in market research for most variables (provided a variable's distribution is not too heavily skewed, in which case the median may be preferred). The mode is only useful for categorical variables. Also, confidence interval and hypothesis testing procedures for $\mu$ are well-known and easily applicable to sample datasets.

ii. Let $\{X_1, X_2, \ldots, X_n\}$ be a simple random sample of size $n$ from a distribution, where the $X_i$s are independent and identically distributed. We have:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

approximately, by the central limit theorem as $n \to \infty$ if the population is non-normal and $\sigma^2 < \infty$, and exactly if $X_i \sim N(\mu, \sigma^2)$.

iii. For a $100(1-\alpha)\%$ confidence interval, we have:

$$z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} \leq e$$

where $e$ is the maximum tolerance for the sampling error, and so:

$$n \geq \frac{(z_{\alpha/2})^2 \sigma^2}{e^2}.$$

All terms should be defined.

iv. This was a deliberately open-ended question, for which the examiners accepted any reasonable justifications based on the market research example chosen by the candidate.

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

**Question 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]**

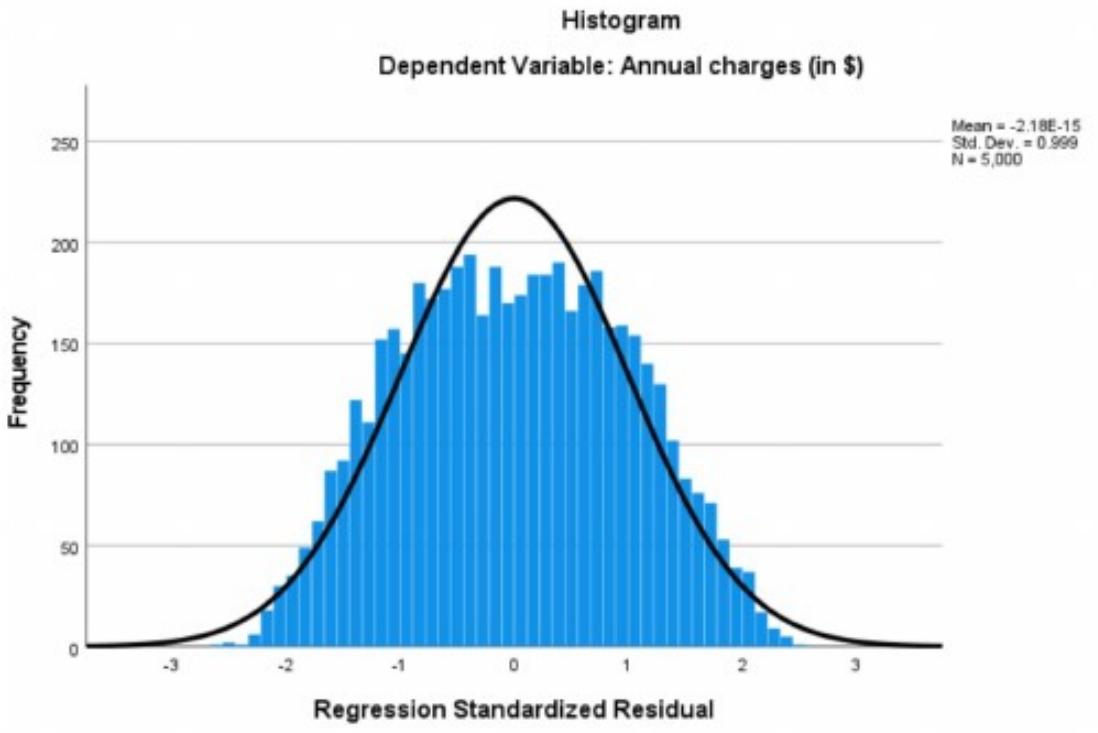
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .591[a] | .349 | .348 | 4903.278836 |

**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 |



**Reading for this question**

For part (a), Block 14 on the VLE covers regression. For part (b), Block 13 on the VLE covers two-way analysis of variance.

**Approaching the question**

(a) An excellent answer would include the following:

- Full specification of the theoretical and estimated regression models. The theoretical regression model is:

  $$\text{Annual charge} = \beta_0 + \beta_1 \text{Annual income} + \beta_2 \text{Age} + \beta_3 \text{Household size} + \beta_4 \text{Sex} + \beta_5 \text{Credit limit} + \varepsilon$$

  while the estimated regression model is:

  $$\widehat{\text{Annual charge}} = 791.79 + 120.88 \times \text{Annual income} - 9.26 \times \text{Age} + 528.07 \times \text{Household size} - 198.51 \times \text{Sex} + 543.67 \times \text{Credit limit}.$$

- $F$ and $t$ tests were expected by the examiners to analyse the statistical significance of the model (with an $F$ test) and the individual predictor variables (with $t$ tests).
- Since 'Credit limit' is a dummy variable, we interpret $\widehat{\beta}_5 = 543.67$ as being the expected increase in annual charges (in dollars) for a credit card customer who has exceeded their credit limit in the past 12 months relative to a customer who had not, other things equal.
- Discussion of the standardised coefficients was expected by the examiners to determine the relative importance of the predictor variables, with annual income being the most dominant. However, with $R^2 = 0.349$ the multiple linear regression model has quite low explanatory power (34.9% of the variation in annual charges can be explained by the model) suggesting there are other important predictor variables which have not been included in the model.
- Any reasonable model changes suggested by candidates were rewarded by the examiners, such as suggestions for sensible additional predictor variables, as well as proposing removing insignificant variables in the model, i.e. age and sex.
- References to the standardised residual histogram and goodness-of-fit to normality were desirable, relating to the model assumption that $\varepsilon \sim N(0, \sigma^2)$.

(b) The most commonly used measure in ANOVA is omega squared, $\omega^2$. This measure indicates what proportion of the variation in the dependent variable is related to a particular independent variable or factor.

The relative contribution of a factor $X$ is calculated as follows:

$$\omega_X^2 = \frac{SS_X - (df_X \times MS_{Error})}{SS_{Total} + MS_{Error}}.$$

Normally, $\omega^2$ is interpreted only for statistically significant effects. As a guide to interpreting $\omega^2$, a large experimental effect produces an index of 0.15 or greater, a medium effect produces an index of around 0.06, and a small effect produces an index of 0.01.

## Question 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**

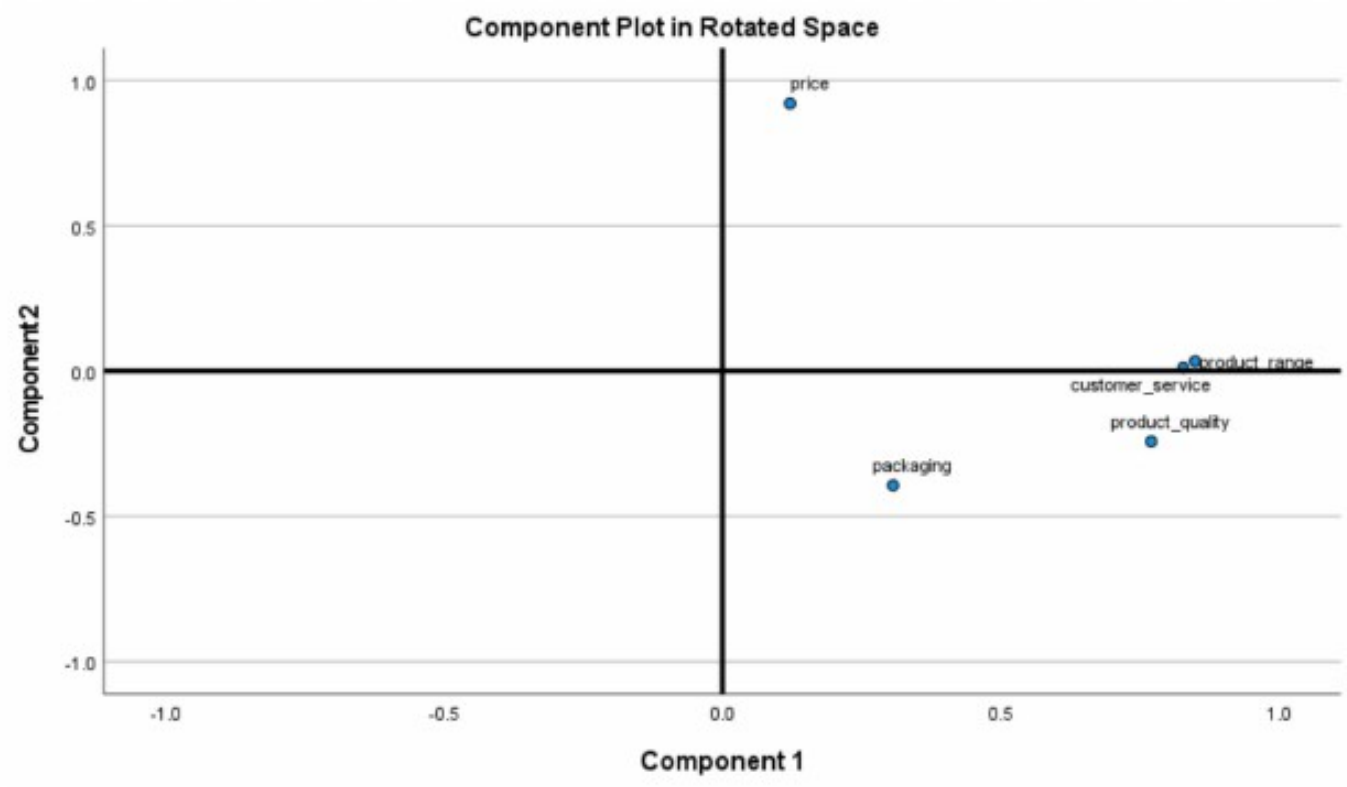
**KMO and Bartlett's Test**

| | | |
|---|---|---|
| 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 | Component 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 | Component 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 | |

**Reading for this question**

For part (a), Block 17 on the VLE covers factor analysis. For part (b), Block 4 presents qualitative research methods, including focus groups.

**Approaching the question**

(a) An excellent answer would consider the statistical and commercial interpretation of the SPSS output. Key 'ingredients' would include the following.

- Examination of eigenvalues and the cumulative percentage variance explained to determine the number of factors are important here. Two factors have eigenvalues greater than 1, retaining 63.4% of the original variation in the data.
- Interpreting the two factors using the rotated component matrix and associated component plot is important. It can be seen that factor 1 is highly correlated with the satisfaction with product range (correlation of 0.850), customer service (correlation of 0.828) and product quality (correlation of 0.771), while factor 2 is highly correlated with price satisfaction (correlation of 0.920). Suggestions of appropriate names for factor 1 (related to non-price attributes) were rewarded by the examiners, while recommending that price satisfaction could be used as a surrogate variable for factor 2 impressed the examiners.
- Model fit is assessed via an examination of the residuals, i.e. the differences between the observed correlations obtained from the input correlation matrix and the reproduced correlations estimated from the factor matrix (i.e. the percentage of residuals $\geq 0.05$ in absolute value). Here 80% of the residuals exceed 0.05 in absolute value, possibly calling into question the model fit.

(b) There are several advantages of running online focus groups. More potential participants can be recruited through the growing use of the internet, and the growing ease of conducting discussions online are key benefits. Participants can be made to feel that they have the ability to contribute, allowing their confidence to be quickly built up. Conflicts in face-to-face focus groups which may stem from participants taking a dislike to other participants due to their physical appearance can be avoided.

A great breadth of information may be collected, through the types and the geographic spread of participants. The practical difficulties of getting individuals together at the same time in the same location can be overcome. The nature of a discussion location which is 'comfortable' to the participant is largely overcome by each participant setting the conditions in which they feel comfortable.

**Question 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.**

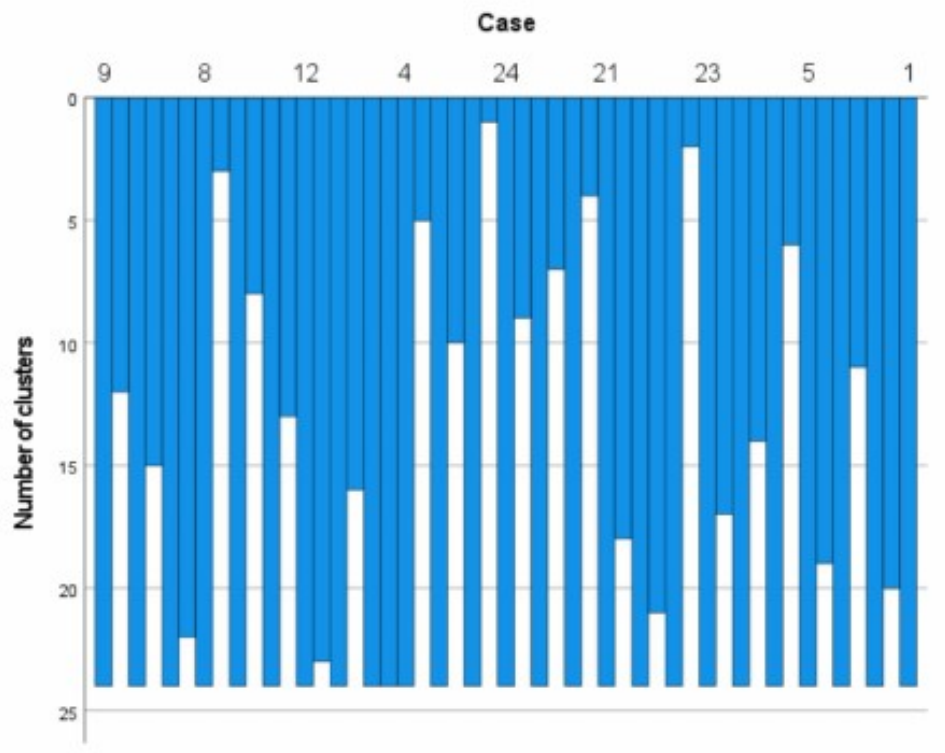
**Explain the purpose of a *paired comparison scale* and a *constant sum scale* in market research, and provide an example of each.**

**(10 marks)**

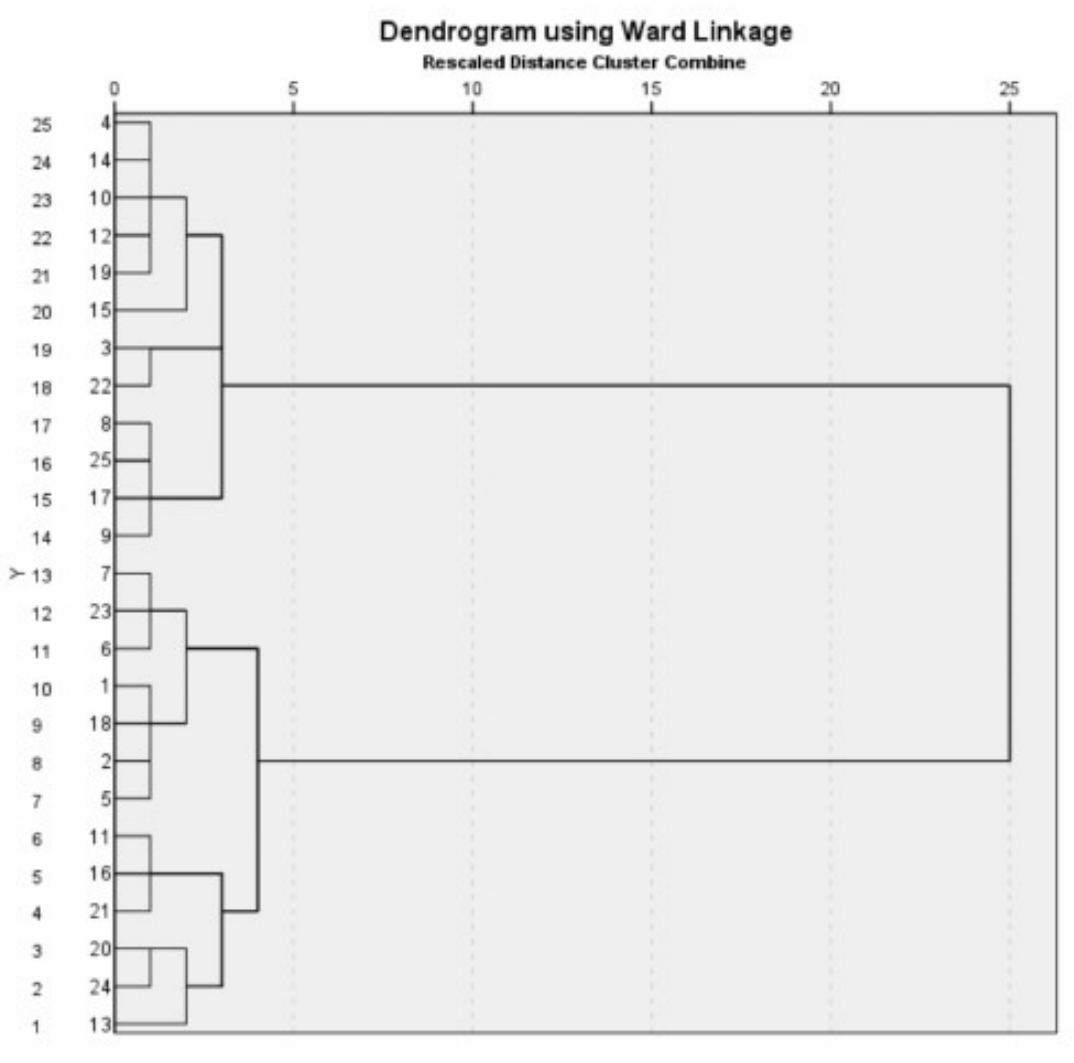
**Figure 3**

**Agglomeration Schedule**

| Stage | Cluster Combined Cluster 1 | Cluster Combined Cluster 2 | Coefficients | Stage Cluster First Appears Cluster 1 | Stage Cluster First Appears Cluster 2 | Next Stage |
|---|---|---|---|---|---|---|
| 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 |

**Reading for this question**

For part (a), Block 18 on the VLE covers cluster analysis. For part (b), Block 7 covers comparative scaling.

**Approaching the question**

(a) An excellent answer would include the following:

- Explanation of Ward's procedure, which could be supported by a diagram, with the aim of minimising the within-cluster variance.
- Explanation of squared Euclidean distance between two objects as:

$$\delta_{ij} = \sum_{k=1}^{p} (x_{ik} - x_{jk})^2.$$

- Here there is a clear two-cluster solution based on the dendrogram, and this is also supported by the agglomeration schedule because of the significant increase in the 'Coefficients' from stage 23 to stage 24 (i.e. when two clusters are merged into one).
- Examining the descriptive statistics of the two clusters (mean, standard deviation, and number of cluster members), Cluster 1 is older (mean age of 54.08) and higher income (mean income of \$62,690), relative to cluster 2 (27.33 and \$29,500, respectively). Note that the age and income socio-economic demographic variables were not used to perform the clustering technique, rather they are used to profile the two identified clusters.
- With reference to the 1–7 Likert scales, cluster 1 scores on average 3.00–3.50 across the attributes, while cluster 2 scores on average 5.00–6.00. Hence the younger cluster seems the more receptive market segment for a new fast fashion company.

(b) Paired comparison scaling is the most widely-used comparative scaling technique. A participant is presented with two objects and asked to select one according to some criterion. The data obtained are ordinal in nature. With $n$ brands, $n(n-1)/2$ paired comparisons are required. Under the assumption of transitivity of preference, it is possible to convert paired comparison data to a rank order.

For constant sum scaling, participants allocate a constant sum of units, such as 100 points, to attributes of a product to reflect their importance. If an attribute is unimportant, the participant assigns it zero points. If an attribute is twice as important as some other attribute, it receives twice as many points. The sum of all the points is 100, hence the name of the scale!