Examiners' commentaries

ST3188 Statistical methods for market research

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2018–19. 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 (2018). 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.

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.

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

ST3188 Statistical methods for market research

Important note

This commentary reflects the examination and assessment arrangements for this course in the academic year 2018–19. 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 (2018). 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

Section A

Question 1

(a) Reading for this question

Block 9 of the course.

Approaching the question

Standard responses expected here, noting which are probability and non-probability methods. For (ii., iii. and iv.), a viable sampling frame should be identified (such as Oyster card holders). As explicitly mentioned in the question, as well as the mechanics of each method the merits and limitations (as covered in lectures) should be stated.

(b) Reading for this question

Block 10 of the course.

Approaching the question

i. $\bar{X} \sim N(\mu, \sigma^2/n)$ either exactly (if sampling from $N(\mu, \sigma^2)$), or approximately by the central limit theorem, as $n \to \infty$.

ii. The sampling distribution represents the distribution of point estimates obtained over infinitely many independent random samples of size n. Since $\mathrm{E}(\bar{X}) = \mu$, $\bar{X}$ is an unbiased estimator of μ, meaning that on average the point estimate is correct. The variance of the sampling distribution, $\mathrm{Var}(\bar{X}) = \sigma^2/n$, decreases as n increases, i.e. the precision of the estimate improves with larger samples.

iii. Should state the condition on finding n as:

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

and hence:

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

$z_{\alpha/2}$ depends on the confidence level – scope to cite an example or two. If σ^2 was unknown, its value could be assumed (i.) based on previous research, or (ii.) estimated from a pilot study.

iv. When $n \geq 0.1N$, the standard error will be (non-negligibly) overestimated, hence we require a finite population correction factor defined by:

$$\sqrt{\frac{N-n}{N-1}}$$

in which case:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}.$$

Section B

Answer two questions. Each question carries equal weight.

Question 2

(a) Reading for this question

Block 13 of the course.

Approaching the question

Output interpretation should include at least the following:

- Calculation of multiple η^2, and interpretation.
- Both factors and the interaction are highly significant. Hypotheses, test statistics, test statistic values and p-values should all be cited and correctly interpreted.
- Discussion of plots, including disordinal interaction with crossover.
- Comment on how more effective advertising can command a higher price when the newspaper decides its pricing strategy for paid-for advertisements.

(b) Reading for this question

Block 4 of the course.

Approaching the question

i. Projective techniques are unstructured and indirect data collection techniques. They may be defined as a form of questioning which encourages the respondent to project their underlying motivations, beliefs, attitudes or feelings regarding the issues of concern. Respondents are not directly asked about their own behaviour, but are asked to reflect on the behaviour of others, indirectly projecting their own motivations, beliefs, attitudes or feelings. Projective techniques should be employed when the required information cannot be accurately obtained by direct methods because the information is not part of conscious memory.

Respondents may not be able to put into words, or even be aware of, why they behave in a particular manner or feel about an issue in a certain way. Direct questioning in these circumstances would generate shallow and meaningless responses. By using projective techniques, respondents can express themselves in ways which can allow the researcher to access their underlying feelings.

ii. In the word association technique, respondents are presented with a list of words, one at a time. After each word they are asked to give the first word which comes to mind. The underlying assumption is that by freely associating with certain words, respondents will reveal their inner feelings about the topic of interest. Word association is frequently used in testing brand names, and occasionally for measuring attitudes about particular products, services, brands, packages or advertisements. For example, imagine the airline KLM wishing to understand the inner feelings of customers about their service delivery. They could present a series of words which encapsulate different aspects of their service delivery such as 'check-in', 'seating' or 'cabin crew'. Presented to target respondents, the first word which comes to mind for a respondent could reveal their inner feelings which may not emerge if they were to be questioned directly or had time to think through and rationalise their response.

iii. With the story completion technique, respondents are given part of a story, enough to direct attention to a particular topic but not to hint at the ending. They are required to give the conclusion in their own words, to complete the story as they see it. The type of respondent could be of any age or type, i.e. a business or consumer respondent. One characteristic they would need is a willingness to let their imagination flow. As with the KLM service delivery example, imagine interviewing travellers waiting to board a long-distance KLM flight. For example, a lone business traveller with a little time on their hands, not engaged with work on their mobile phone or laptop, could help. They would be part-way through their service experience so their expectations of the remainder of their flight would be much clearer than at any other time. With the right introduction and motivation, they could be given the start of a story about an airline service experience, and then left alone to complete the story with the interviewer calling back to collect the details and perhaps ask a few other questions.

Question 3

(a) Reading for this question

Block 14 of the course.

Approaching the question

Good answers would include the following:

- Regression model in full with assumptions on the random error term, plus the estimated model.
- Discussion of output results such as R^2, the F statistic, regression coefficients (including standadised coefficients) and their significance, the construction of confidence intervals, and the residual histogram.
- Promotion is by far the most significant and important explanatory variable of match attendance, suggesting scope for an expansion of promotional activities noting the objective of increasing ticket sales at matches.

(b) Reading for this question

Block 10 of the course.

Approaching the question

'Response rate' = # of completions/# in survey. Non-response is refusal to respond. There are two types: (i.) unit non-response (refusal to participate in survey) and (ii.) item non-response (refusal to answer a specific question in the survey). Improve response rates by prior notification, incentives, better questionnaire design, callbacks etc. Strategies for adjusting for non-response include the following.

- Sub-sampling of non-respondents. A concerted effort is made to contact a sub-sample of the respondents, usually by means of telephone or personal interviews.
- Replacement. The non-respondents in the current survey are replaced with non-respondents from an earlier, similar survey.
- Substitution. The non-respondents are substituted with other elements from the sampling frame that are expected to respond.
- Subjective estimates. This involves making a subjective evaluation of the likely effects of non-response based on experience and available information.
- Trend analysis. The researcher tries to discern a trend between early and late respondents. This trend is projected to non-respondents to estimate their characteristic of interest.
- Simple weighting. Depending on the response rates, differential weights are assigned to the data to account for non-response.
- Imputation. Imputing the characteristic of interest to the non-respondents based on the similarity of the variables available for both non-respondents and respondents.

Reference should be made of how these are tackled across different data collection methods.

Question 4

(a) Reading for this question

Block 17 of the course.

Approaching the question

Output interpretation should include at least the following:

- Examination of eigenvalues and cumulative percentage variance explained to determine the number of factors. Interpretation of factors using rotated component matrix and associated component plot. Comment on how good or poor the results are using the reproduced correlations. Discussion of how factor scores can be calculated.
- Model fit is assessed via an examination of residuals, the differences between the observed correlations obtained from the input correlation matrix and the reproduced correlations estimated from the factor matrix. If many large residuals exist, then one can infer that the factor model does not provide a good fit to the data. This analysis is based on the implicit assumption that the observed correlation between the variables is due to the common factors, therefore the correlations between the variables can be reproduced from the estimated correlations between the variables and the factors.
- Discussion of how factor scores can be used in multiple regression and discriminant analysis.

(b) Reading for this question

Block 19 of the course.

Approaching the question

The total variation is defined as the sum of the squared deviations of each measurement from the overall mean and is given by:

$$SS_Y = SS_X + SS_{Error}$$

where:

$$SS_Y = \sum_{j=1}^{c} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y})^2 \qquad SS_X = \sum_{j=1}^{c} n_j (\bar{Y}_j - \bar{Y})^2$$

and:

$$SS_{Error} = \sum_{j=1}^{c} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2$$

where:

- Y_{ij} = the ith observation in the jth category
- $\bar{Y}$ = the mean over the whole sample, i.e. the overall mean
- $\bar{Y}_j$ = the mean for category j
- n_j = the sample size for category j.

$SS_{Between}$ represents the variation among the means of Y in the categories of X and involves the squares of the deviations of various category means from the overall mean.

SS_{Within} represents the variation in Y due to the variation within each category of X and involves the squares of the deviations of each measurement of Y from the corresponding category mean. $SS_{Between}$ is often referred to as SS_X, and SS_{Within} is often referred to as SS_{Error}.