Specimen examination paper

ST3188

Statistical methods for market research

Instructions to candidates

This paper contains two sections. Section A is compulsory and contains one question. Section B contains three questions.

Answer THE question in Section A and TWO questions from Section B.

Section A carries 40 marks and Section B carries 60 marks. All questions in Section B will be given equal weight (30 marks).

The coursework mark (worth 30% of the final mark) will be combined with the examination mark (worth 70%) to give the overall percentage for the module.

A calculator may be used when answering questions on this paper and it must comply in all respects with the specification given with your Admission Notice. The make and type of machine must be clearly stated on the front cover of the answer book.

SECTION A: Compulsory

1. (a) Transport for London is the local government body responsible for most aspects of the transport system in London. Its role is to implement the transport strategy and to manage transport services across London.

TfL launched the Night Tube service on the London Underground in August 2016. Currently, the Night Tube operates on Friday and Saturday nights across five lines (the Central, Jubilee, Northern, Piccadilly and Victoria lines). TfL management would consider expansion of the service to more lines and/or other nights of the week, subject to consumer demand.

Some 1.34 billion passenger journeys are made each year on London Underground (albeit with many individuals making multiple journeys), however there is no sampling frame with full coverage of all distinct passengers available. Despite this limitation, TfL management wants to conduct market research to assess consumer appetite for an extension to the Night Tube service, which may have fare implications. At present Night Tube fares are classified as off-peak, although a special night-off-peak fare could be introduced, depending on passengers' price elasticities.

TfL has invited you to devise an appropriate sampling scheme to research attitudes towards developing the Night Tube. Explain in detail how each of the following sampling methods could be applied to the overall sampling strategy. Make sure you describe the merits and limitations of each as well as how each would be applied in practice.

i. Convenience sampling.

ii. Simple random sampling.

iii. Systematic sampling.

iv. Stratified sampling.

(20 marks)

(b) Suppose we are interested in estimating the mean of a population with a finite variance using a simple random sample of size n.

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

ii. Explain how the sampling distribution derived in i. should be interpreted.

iii. Explain how to determine the minimum sample size necessary to estimate a population mean to within e units assuming the population standard deviation is known. If the population standard deviation was unknown, how would you deal with this?

iv. Explain the purpose of the finite population correction factor (including a formula) and when it should be used.

(20 marks)

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

2. (a) A national newspaper was interested in researching how effective paid-for advertising was in its newspapers. The editors suspected the number of customer enquiries resulting from advertisements depended on two key factors – the day of the week (Monday through Friday), and the section of the newspaper where the advertisement appeared (news, business and sports).

   Analyse the selected SPSS output in Figure 1 (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:

   - 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 pricing strategy for paid-for advertisements in this national newspaper.

   (20 marks)

   (b) i. What are projective techniques? Under what circumstances should projective techniques be used?

   ii. Describe the word association technique. Give an example of a situation in which this technique is especially useful.

   iii. Describe the story completion technique. Give an example of the type of respondent and the context in which such a technique would work.

   (10 marks)

Figure 1

Descriptive Statistics

Dependent Variable: Number of customer enquiries

Day of the week	Section of newspaper	Mean	Std. Deviation	N
Monday	News	8.25	2.062	4
	Business	11.50	1.291	4
	Sports	4.50	1.291	4
	Total	8.08	3.315	12
Tuesday	News	10.25	1.258	4
	Business	8.75	1.708	4
	Sports	6.50	1.291	4
	Total	8.50	2.067	12
Wednesday	News	9.25	1.258	4
	Business	8.50	1.291	4
	Sports	6.75	1.708	4
	Total	8.17	1.697	12
Thursday	News	4.25	.957	4
	Business	7.75	1.258	4
	Sports	6.00	.816	4
	Total	6.00	1.758	12
Friday	News	12.50	1.291	4
	Business	9.00	.816	4
	Sports	11.25	.957	4
	Total	10.92	1.782	12
Total	News	8.90	3.059	20
	Business	9.10	1.744	20
	Sports	7.00	2.575	20
	Total	8.33	2.653	60

Tests of Between-Subjects Effects

Dependent Variable: Number of customer enquiries

Source	Type I Sum of Squares	df	Mean Square	F	Sig.
Model	4503.000[a]	15	300.200	171.000	.000
Day	4313.500	5	862.700	491.411	.000
Section	53.733	2	26.867	15.304	.000
Day * Section	135.767	8	16.971	9.667	.000
Error	79.000	45	1.756		
Total	4582.000	60			

a. R Squared = .983 (Adjusted R Squared = .977)

Levene's Test of Equality of Error Variances[a]

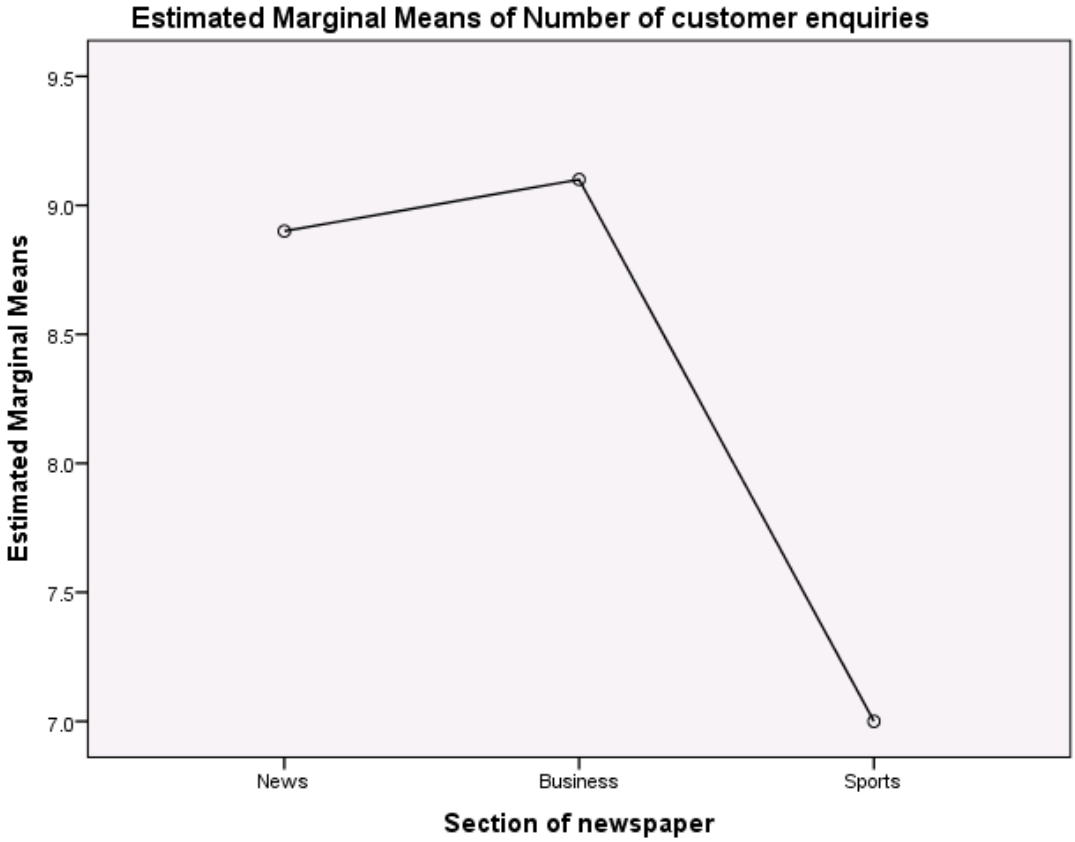
Dependent Variable: Number of customer enqı

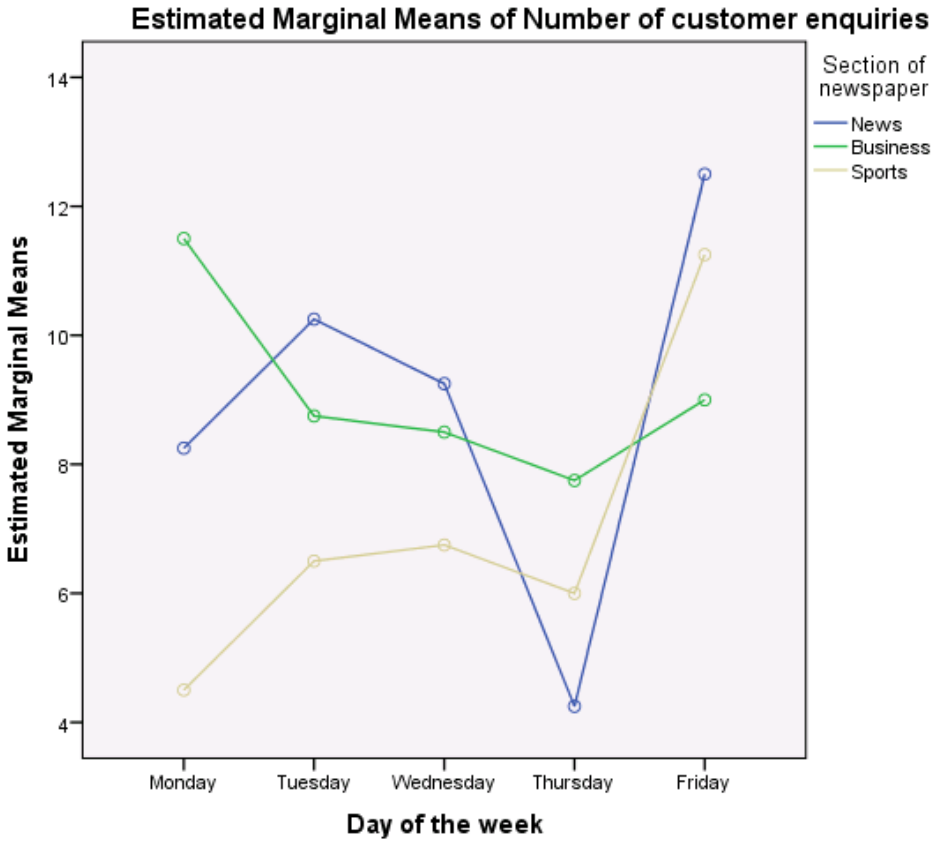
F	df1	df2	Sig.
.482	14	45	.931

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Day + Section + Day * Section

Figure 1 (continued)





3. (a) Many factors determine the attendance at team sporting events, such as when the game is played, the weather, the opponent, and whether the team is having a good season. In an effort to increase ticket sales at matches, some football clubs decided to run promotions such as free concerts after a game, or giveaways of team merchandise etc. A multiple regression was run on the following variables for a football club:

- Paid attendance for each home game, `Attendance`.
- Highest temperature on the fixture day, `Temp`.
- Team's winning percentage at the time of the game, `Win`.
- Opponent team's winning percentage at the time of the game, `OpWin`.
- A dummy variable of 1 if the game is played on Friday, Saturday or Sunday, and 0 otherwise, `Weekend`.
- A dummy variable of 1 if a promotion was held, and 0 otherwise, `Promotion`.

Selected SPSS output is provided in Figure 2. Analyse the regression results, making sure you first write out the full regression model, including any assumptions, and the estimated model. In light of the regression results, propose any changes you would make to the model, including any other explanatory variables which you would consider using, or any you would remove from the existing model. Keep in mind the objective of increasing ticket sales at matches.

(20 marks)

(b) 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 2

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.621[a]	.386	.344	7022.575

a. Predictors: (Constant), Promotion, OpWin, Temp, Win, Weekend

b. Dependent Variable: Attendance

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2262333118	5	452466623.5	9.175	.000[b]
	Residual	3600108419	73	49316553.68		
	Total	5862441536	78			

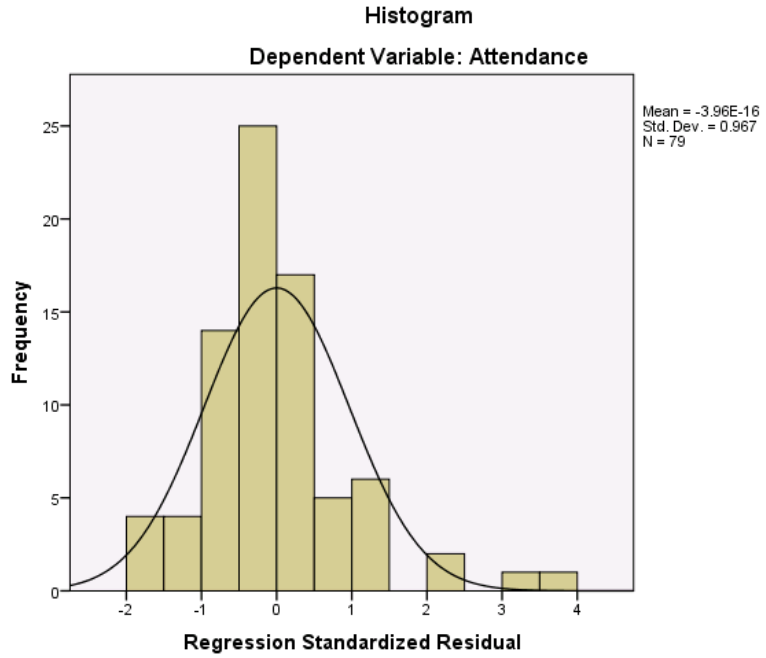
a. Dependent Variable: Attendance

b. Predictors: (Constant), Promotion, OpWin, Temp, Win, Weekend

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-3666.537	11082.837		-.331	.742
	Temp	170.894	67.777	.236	2.521	.014
	Win	-11.246	19.137	-.055	-.588	.559
	OpWin	22.747	12.395	.176	1.835	.071
	Weekend	1823.671	1747.134	.106	1.044	.300
	Promotion	9074.376	1872.070	.479	4.847	.000

a. Dependent Variable: Attendance



4. (a) The dietary characteristics of 200 respondents were recorded in a survey. Respondents were asked to rate the extent to which their diet agreed with the following descriptors:

- fast food
- filling
- hearty
- low meat
- vegetarian.

Figure 3 (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 factors.
- Explain qualitatively and quantitatively how the fit of the factor analysis model should be examined.
- Briefly discuss for what modelling purpose(s) extracted factors could be used.

(20 marks)

(b) Consider a one-way analysis of variance where the independent categorical variable, X, has c levels, and n_j is the sample size for category j for $j = 1, \ldots, c$. Explain how the total variation in the dependent variable, Y, is decomposed. You should support your answer with formulae and define any terms used.

(10 marks)

Figure 3

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.667
Bartlett's Test of Sphericity	Approx. Chi-Square	737.767
	df	10
	Sig.	.000

Communalities

	Initial	Extraction
vegetarian	1.000	.924
low_meat	1.000	.913
fast_food	1.000	.923
filling	1.000	.848
hearty	1.000	.850

Extraction Method: Principal Component Analysis.

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	2.828	56.553	56.553
2	1.631	32.614	89.167
3	.256	5.129	94.296
4	.166	3.316	97.611
5	.119	2.389	100.000

Extraction Method: Principal Component Analysis.

Rotated Component Matrix[a]

	Component	
	1	2
vegetarian	-.041	.960
low_meat	.241	.924
fast_food	.961	.018
filling	.913	.121
hearty	.913	.125

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

Component Score Coefficient Matrix

	Component	
	1	2
vegetarian	-.105	.556
low_meat	.009	.509
fast_food	.375	-.078
filling	.347	-.015
hearty	.347	-.012

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Component Scores.

Figure 3 (continued)

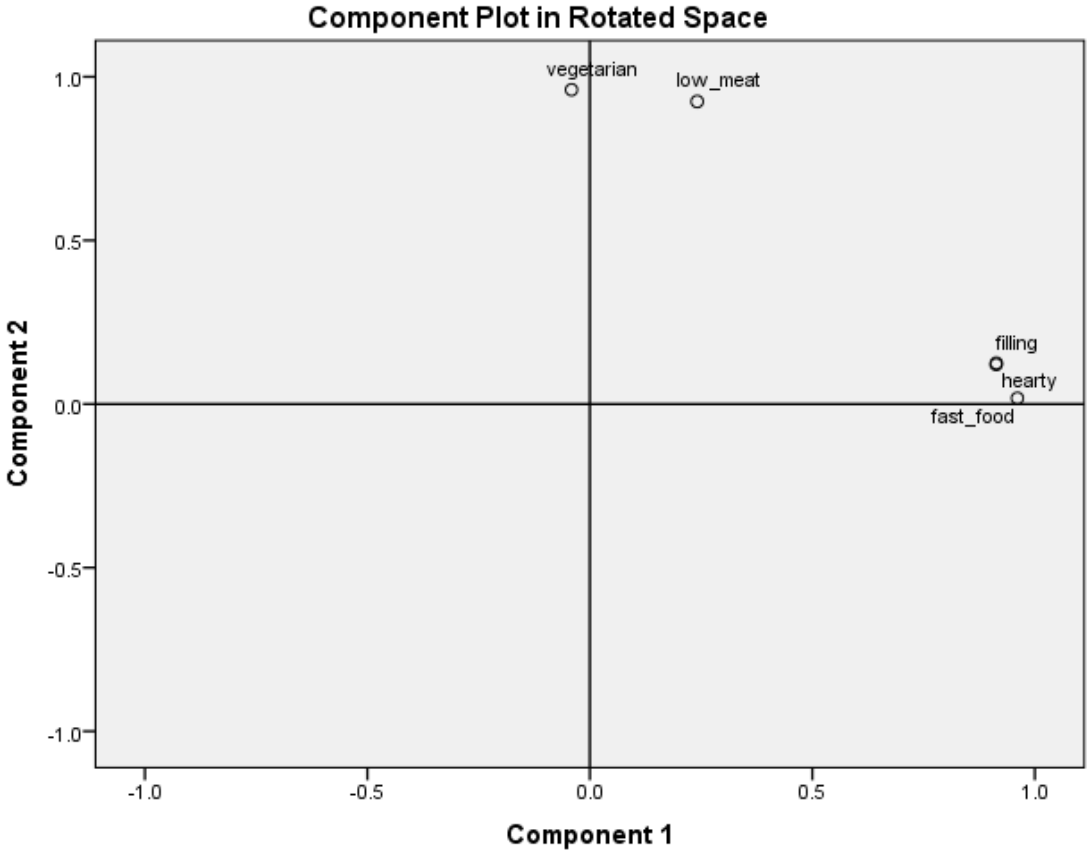
Reproduced Correlations

		vegetarian	low_meat	fast_food	filling	hearty
Reproduced Correlation	vegetarian	.924[a]	.878	-.023	.078	.082
	low_meat	.878	.913[a]	.248	.332	.336
	fast_food	-.023	.248	.923[a]	.879	.880
	filling	.078	.332	.879	.848[a]	.849
	hearty	.082	.336	.880	.849	.850[a]
Residual[b]	vegetarian		-.081	.011	.005	.008
	low_meat	-.081		-.002	-.010	-.014
	fast_food	.011	-.002		-.041	-.039
	filling	.005	-.010	-.041		-.105
	hearty	.008	-.014	-.039	-.105	

Extraction Method: Principal Component Analysis.

a. Reproduced communalities

b. Residuals are computed between observed and reproduced correlations. There are 2 (20.0%) nonredundant residuals with absolute values greater than 0.05.



[END OF PAPER]