

ACE2013

NEWCASTLE UNIVERSITY

SEMESTER 2 2011/2012

ACE2013

Statistics for Marketing and Management

Time allowed: 2 hours

Candidates should attempt ALL questions. Marks for each question are indicated. However you are advised that marks indicate the relative weight of individual questions, they do not correspond directly to marks on the University scale.

There are FIVE questions on this paper.

Answers to questions should be entered directly on this question paper in the spaces provided. This question paper must be handed in, attached inside an anonymised cover sheet, at the end of the examination.

University-approved calculators are allowed. Statistical tables, and a list of useful formulae, are provided in a separate booklet.

1. (a) *Briefly* explain the purpose for which each of the following inferential statistical data analysis techniques are applied in practice (use examples to explain your answers).

(i) **Independent samples t -test**

Answer:

(ii) **Paired (or related) samples t -test**

Answer:

(iii) **Analysis of Variance (ANOVA)**

Answer:

(iv) **Multiple linear regression**

Answer:

(v) **Logistic regression**

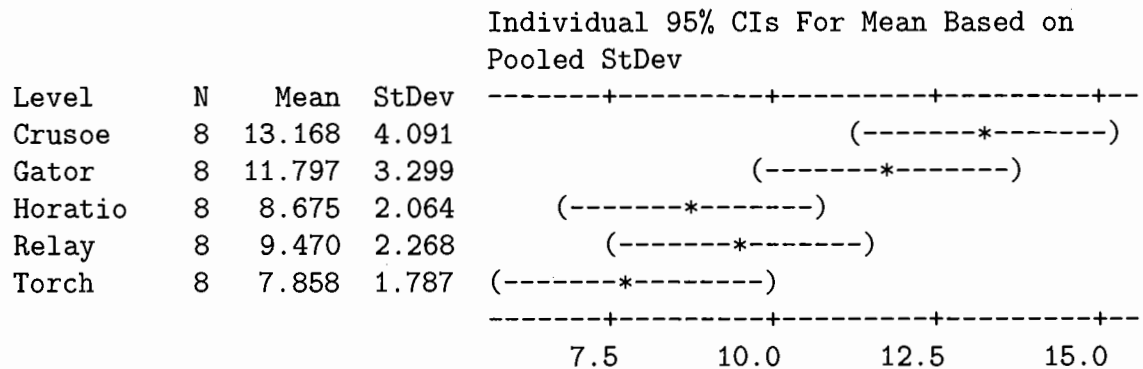
Answer:

- (b) An experiment was conducted to compare the yields of five new varieties of winter wheat named *Crusoe*, *Gator*, *Horatio*, *Relay* and *Torch*. Researchers sowed eight randomly chosen plots with each type of seed and recorded the yield (in tonnes/Ha) of each variety. A one-Way ANOVA was used to analyse the data and the results from Minitab are shown overleaf. Look at this output, and then answer the following questions.

One-way ANOVA: Yield(tonnes/Ha) versus Variety

Source	DF	SS	MS	F	P
Variety	4	157.65	39.41	4.90	0.003
Error	35	281.49	8.04		
Total	39	439.13			

S = 2.836 R-Sq = 35.90% R-Sq(adj) = 28.57%



Pooled StDev = 2.836

- (i) State the null and alternative hypotheses for the study.

Answer:

- (ii) Based on your interpretation of the results, compare the mean yields between the five wheat varieties.

Answer:

[20 marks]

2. *Hotcow Marketing* is considering two alternative advertising plans for a client.

The first plan uses radio and television commercials that will cost £100,000. The agency estimates a 40% chance that the client's revenues will increase by £80,000, a 25% chance that revenues will increase by £110,000, and a 35% chance that revenues will increase by £120,000 as a result of this advertising approach.

The second plan is to use newspaper and magazine advertisements, at a cost of £20,000. The agency estimates the chances are 20% that this print media approach will increase revenues by £14,000, 50% that revenues will increase by £30,000, and 30% that revenues will increase by £40,000.

- (a) Set up a decision tree for this problem.

Answer:

- (b) Calculate the *Expected Monetary Value* (EMV) for the Radio/television advertising plan and the EMV for the Newspaper/Magazine plan.

Answer:

- (c) Which advertising plan should *Hotcow* recommend to the client? Why?

Answer:

- (d) Calculate the expected value of perfect information and explain what it means to the client.

Answer:

[20 marks]

3. You are a rural developer specialising in summer cottage properties, and you are considering purchasing a large tract of land adjoining a lake. The current owner has already divided the land into several plots and has prepared these plots by removing some of the trees. You want to estimate the value of each plot of land.

From a nearby area, you gather the following data from 60 recently sold plots of land:

y = Price, in thousands of pounds
 x_1 = Plot size, in thousands of square feet
 x_2 = Number of mature trees
 x_3 = Distance from the lake, in metres

- (a) Minitab was used to fit a multiple linear regression model to these data; the resulting (edited) output for the “full” model is shown below. Look at it, and then answer the following questions.

Regression Analysis: y versus x1, x2, x3

The regression equation is

$y = ?$ _____

Predictor	Coef	SE Coef	T	P
Constant	182.576	9.867	18.50	0.000
x1	0.3091	0.2345 ?_____		0.193
x2	-0.20499	0.09621	-2.13	0.038
x3	?_____	0.08192	-14.83	0.000

S = 16.8855 R-Sq = 81.8% R-Sq(adj) = 80.8%

- (i) Complete the regression output above by filling in the blanks indicated by the question marks (?). If you need to, you may use the space below to show you working.

- (ii) Complete the following statement about the residuals ϵ_i :

$$\epsilon_i \sim N\left(\underline{\hspace{2cm}}, \underline{\hspace{2cm}}\right).$$

- (iii) With reference to the estimated coefficients, explain the *direction* of the relationship between each of the predictors and the response.

Answer:

- (iv) Which variable should be removed before a reduced model is fitted? Explain.

Answer:

- (b) The output from Minitab shown below represents the “final” model. Look at it, and then answer the following questions.

Regression Analysis: y versus x3

The regression equation is

$$y = 187 - 1.25 x3$$

Predictor	Coef	SE Coef	T	P
Constant	186.844	4.537	41.19	0.000
x3	-1.25038	0.08167	-15.31	0.000

S = 17.3204 R-Sq = 80.2% R-Sq(adj) = 79.8%

- (i) Briefly explain what has happened between the fit of the “full” and “final” models.

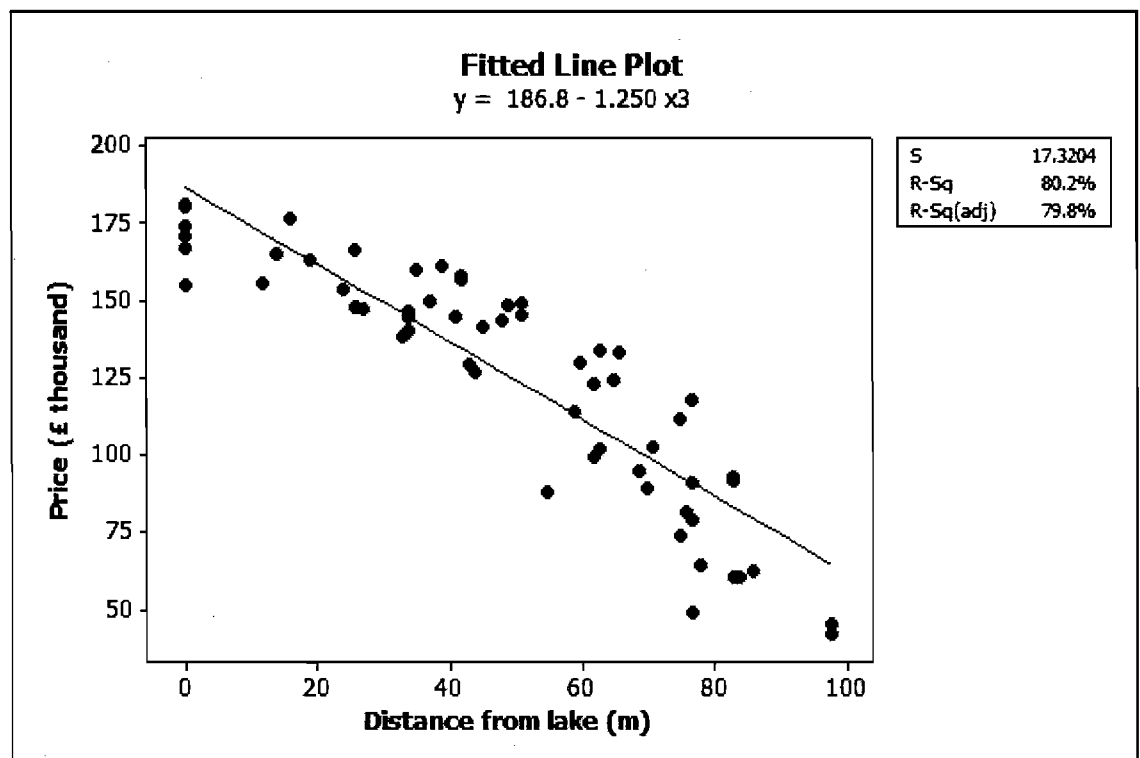
Answer:

- (ii) Comment on the change in the R^2 statistic from the “full” to the “final” model.

Answer:

- (iii) The graph below shows a plot of y against x_3 , with the fitted regression line from the output shown above superimposed. Comment.

Answer:



- (c) A further analysis was performed in Minitab, the output from which is shown below. Look at it, and then answer the following questions.

Regression Analysis: y versus x3, x3squared

The regression equation is

$$y = 171 - 0.128 x3 - 0.0125 x3squared$$

Predictor	Coef	SE Coef	T	P
Constant	171.071	4.897	34.93	0.000
x3	-0.1280	0.0621	-2.06	0.044
x3squared	-0.012469	0.002448	-5.09	0.000

$$S = 14.4835 \quad R\text{-Sq} = 91.4\% \quad R\text{-Sq}(\text{adj}) = 89.8\%$$

- (i) Explain the link between the scatter plot shown previously and the output given above.

Answer:

- (ii) Compare the fit of the model above to the simple linear regression model fitted in part (b).

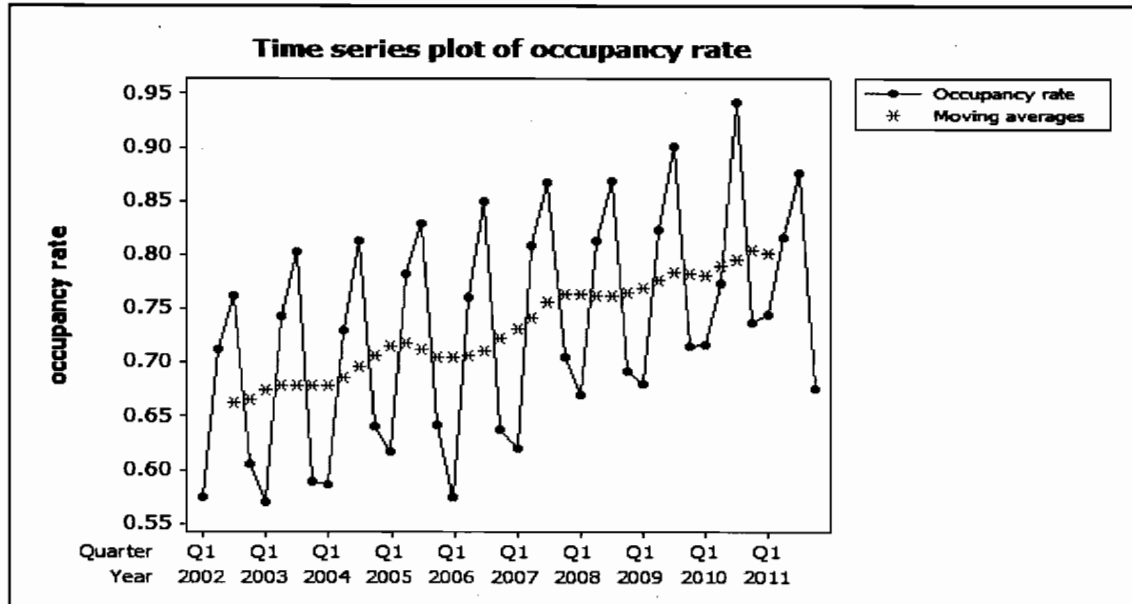
Answer:

- (iii) Using the output given above, estimate the value of a plot of land that is 50 metres from the lake.

Answer:

[20 marks]

4. As the manager of a hotel in the Lake District you have recorded the average occupancy rate of your hotel, per quarter, for the last ten years (2002–2011, inclusive); this information is shown in the time series plot below. The table underneath gives the occupancy rates for the last three years (2009–2011 inclusive).



	Jan–Mar	Apr–Jun	Jul–Sep	Oct–Dec
2009	0.68	0.82	0.90	0.71
2010	0.72	0.77	0.94	0.74
2011	0.74	0.82	0.88	0.68

- (a) Other than an obvious positive trend, comment on any interesting features revealed by the time series plot.

Answer:

- (b) We can use centred moving averages (y_t^*) to isolate the trend. Show that the *last* such moving average is

$$y_{38}^* = 0.7875,$$

and superimpose this on the time series plot above.

Answer:

- (c) Let T represent time, and let $t_i, i = 1, 2, \dots$ correspond to Jan–Mar (2002), Apr–Jun (2002), $\dots$ (respectively). Suppose you calculate

$$S_{TY^*} = \sum_{i=3}^{38} t_i y_i - n \bar{t} \bar{y}^* = -18.651$$

and

$$S_{TT} = \sum_{i=3}^{38} t_i^2 - n \bar{t}^2 = 3885,$$

where $\bar{t} = 20.5$ and $\bar{y}^* = 0.734$ are the mean of the time-points, and moving averages, respectively.

- (i) Without performing any further calculations, how do you know your answer for S_{TY^*} must be wrong?

Answer:

- (ii) Suppose that, on recalculating, you find that $S_{TY^*} = 15.991$. Estimate the slope term β_1 in the simple linear regression model

$$Y^* = \beta_0 + \beta_1 T + \epsilon,$$

where $\epsilon \sim N(0, \sigma^2)$.

Answer:

- (iii) Now also estimate the intercept term β_0 , and hence write down the estimated linear trend equation.

Answer:

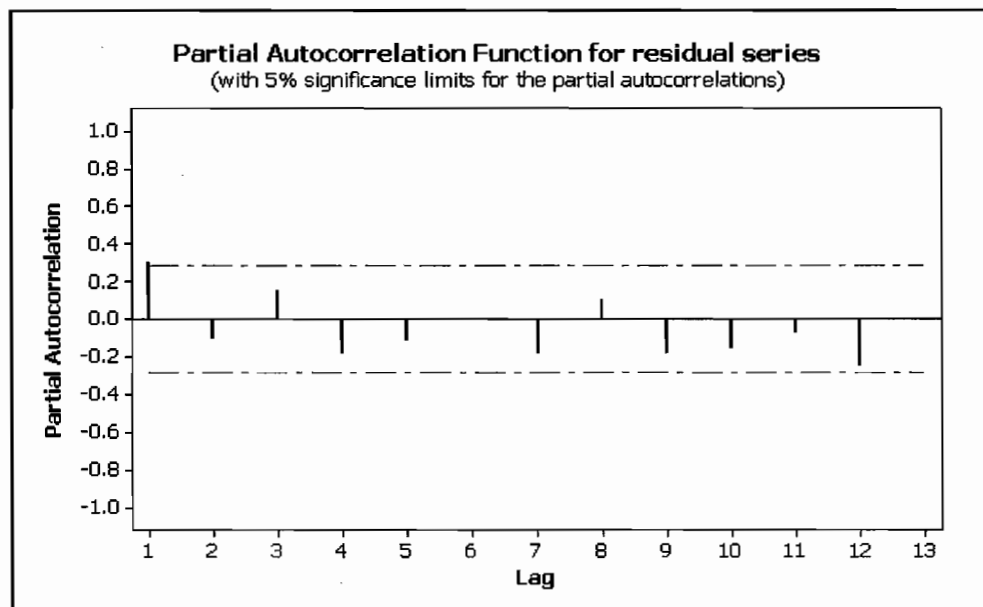
- (iv) Use your equation in part (iii) to estimate the trend for July–September 2012.

Answer:

(d) The following (edited) output has been obtained from Minitab:

Period	1	2	3	4
Seasonal Indices	-0.0860	0.0570	-----	-----

- (i) One of the missing seasonal indices is 0.1225. Enter this in the correct space in the Minitab output above.
- (ii) Now calculate the remaining seasonal effect, and enter this in the remaining space above.
- (e) You now use Minitab to subtract the estimated trend in part (c), and the seasonal effects in part (d), from the original data shown in the time series plot. You then examine the autocorrelation in the resulting residual series by producing a plot of the partial autocorrelation function, as shown in the graph below.



- (i) Use this plot to estimate the lag 1 autocorrelation.

Answer:

$$r_1 \approx \underline{\hspace{2cm}}$$

- (ii) Does this plot suggest an $AR(1)$ or an $AR(2)$ model might be most appropriate for the residual series? Briefly explain.

Answer:

- (f) You now use Minitab to fit the most suitable autoregressive model, as suggested by the plot in part (e), to the residual series. This gives the forecasted values shown below:

Forecasts from period 40

		95% Limits	
Period	Forecast	Lower	Upper
41	-0.0104	-0.0591	0.0383
42	-0.0033	-0.0524	0.0457
43	-0.0025	-0.0515	0.0466
44	-0.0024	-0.0514	0.0467
45	-0.0024	-0.0514	0.0467

Using the output above, your answer to part (c)(iv) and the seasonal effects in part (d), forecast the average occupancy rate of your hotel in July–September 2012, and provide a 95% confidence interval for this forecast. Comment.

Answer:

[20 marks]

5. (a) In the context of *Statistical Process Control*, outline **two** ways a production process could go *out of control*.

Answer:

- (b) It is thought that *Econase XT*, an enzyme added to hen feed, can increase the size of eggs produced by factory-farmed hens. During the feed-production process, a machine adds the enzyme to standard hen feed. Too much of the enzyme can be extremely harmful to hens, and so the amount added must be continuously monitored.

Every hour, a random sample of three batches of hen feed is inspected, and the amount of enzyme (milligrams) in the feed recorded. Results for a ten hour period are shown in the table below.

- (i) Complete the table of results shown below for the last two samples.

Answer:

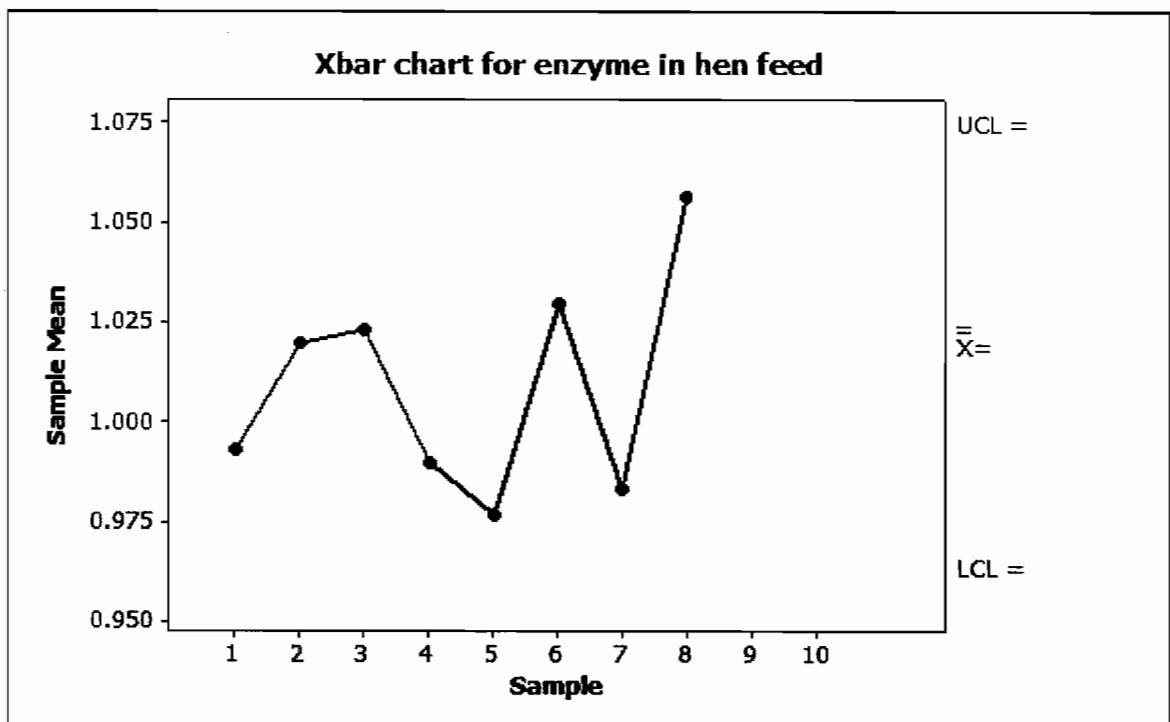
Sample (j)	1st	2nd	3rd	$\bar{x}_j$	s_j
1	1.04	0.96	0.98	0.993	0.042
2	0.99	1.03	1.04	1.020	0.026
3	1.02	0.97	1.08	1.023	0.055
4	0.97	0.97	1.03	0.990	0.035
5	0.98	0.97	0.98	0.977	0.006
6	1.02	1.03	1.04	1.030	0.010
7	0.94	1.02	0.99	0.983	0.040
8	1.12	1.00	1.05	1.057	0.060
9	1.04	1.09	1.02		
10	1.09	1.08	1.02		

(ii) Find $\bar{\bar{x}}$.

Answer:

(iii) It can be shown that the pooled standard deviation is $S = 0.0385$. Use this, and your answer to part (ii), to find the 2.5-sigma control limits for an $\bar{x}$ -chart – and hence complete the graph produced by Minitab below.

Answer:



- (c) Your $\bar{x}$ -chart produced in part (b)(iii) should show that the process is under control.
(i) Show that the probability of a *false alarm* is 0.0124.

Answer:

- (ii) Find the *Average Run Length*, and hence find the expected number of batches of feed produced before a false alarm is called, if the company produces 50 batches per hour.

Answer:

- (d) The next two samples give

$$\bar{x}_{11} = 1.075\text{mg} \quad \text{and} \quad \bar{x}_{12} = 1.08\text{mg}.$$

Add these two points to the $\bar{x}$ -chart on the previous page, and comment.

Answer:

[20 marks]

THE END