Student Number_____

Desk Number_____

UNIVERSITY OF WARWICK

Summer

Business Analytics

EXAMPLE PAPER 2
Example Solutions

Instructions

This is a CLOSED book examination.

Time Allowed: 1.5 hours

One silent calculator, which is not capable of data storage or retrieval is permitted. Electronic devices such as, for example, a mobile phone, tablet, smart watch, fitbit or similar device are not permitted. Graph paper will be provided. Graph paper will be provided.

This examination paper consists of 4 questions. **You should answer ALL questions.** The sum of obtainable marks across all questions is 100.

Answers should be entered on the examination paper in the spaces provided. If you run out of space continue on the back of the page but make sure that you number the answers clearly. Where graph paper is used, make sure that you write your student number on the paper and attach it securely to the examination paper. **Do not append any material that you bring with you to the examination as only material written during the examination will be marked.**

Question papers MUST NOT be removed from the Exam Hall.

Add your student number and desk number to the top of this examination paper and make sure that you hand the paper to an invigilator (together with any answer book if required) at the end of the examination.

The next pages (2-4) contains some formulae and statistical tables for your use if you wish. The questions start on page 5.

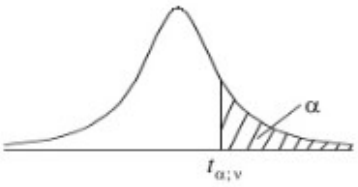
Answer all questions

FOR MARKING PURPOSES ONLY	Q1	Q2	Q3	Q4
Mark				
Marker Initials				

Selected formulae and statistical tables:

Table of the Student's *t*-distribution

The table gives the values of $t_{\alpha;\nu}$ where
$\Pr(T_\nu > t_{\alpha;\nu}) = \alpha$, with ν degrees of freedom



ν \ α	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.078	6.314	12.076	31.821	63.657	318.310	636.620
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

https://cdn.slidesharecdn.com/ss_thumbnails/t-distributiontable-170723122602-thumbnail-4.jpg?cb=1500812899

Formula sheet continued over page....\

Formula sheet continued…\

- **Confidence Interval around a regression coefficient:**

$$Regression\ Coefficient\ Value\ \pm \left[t_{\left(\frac{\alpha}{2}, n-k-1\right)} \times SE(regression\ coefficient) \right]$$

Where k is the number of predictors in the model (not including the intercept).

- **T-Test Statistic for regression coefficient:**

$$t = \frac{coefficient\ estimate - Assumed\ value\ from\ H_0}{SE(regression\ coefficient)}$$

The degrees of freedom for this test is $n - k - 1$, where k is the number of predictors in the model (not including the intercept).

- **F-Test statistic value** = $MS_{regression}/MS_{residual}$

- **Adjusted R-Squared:** $Adj\ R^2 = R^2 - \frac{(1-R^2)k}{(n-k-1)}$, where k is the number of predictors in the model (not including the intercept).

- **The Durbin-Watson (DW) Statistic:**

$$DW = \frac{\sum_{t=2}^{n} \left(e_t - e_{t-1}\right)^2}{\sum_{t=1}^{n} e_t^2}$$

Where $e_{t-1} = Y_{t-1} - \hat{Y}_{t-1}$, $e_t = Y_t - \hat{Y}_t$ and n is the number of data points used to fit the model. *Note: Y_t is the observed Y value at time period t and $\hat{Y}_t$ is the predicted Y value at time period t.*

- **Predictive Interval (general form):**

$$Forecast_{t+p} \pm t_{\alpha/2,df} \times SE\left(Y_{t+p} - Forecast_{t+p}\right)$$

$p = 1,2,3,\ldots$ and indicates how far ahead the forecast is, df represents the degrees of freedom and α is the significance level.

- **The CDF formula for an Exponential distribution** with a mean of λ: $F(x) = 1 - e^{-\lambda x}$

- **The CDF formula for the continuous Uniform distribution**: $U(a, b)$ is $F(x) = \frac{x-a}{b-a}$

Formula sheet continued over page....\

Formula sheet continued…\

- **Project scheduling formula:**

  EST(i) = earliest start time for activity i, EFT(i) = earliest finish time for activity i

  LST(i) = latest start time for activity i, LFT(i) = latest finish time for activity i

  t_i : duration of activity i

  P_i: set of immediate predecessors of activity i, S_i: set of immediate successors of activity i

  $EST(i) = \max\{EFT(j) : j \in P_i\}$

  $EFT(i) = EST(i) + t_i$

  LF$T(i) = \min\{LST(j) : j \in S_i\}$

  $LST(i) = LFT(i) - t_i$

Question 1 over the page....\

Question 1 (33 marks)

The undergraduate program team in the Business School of a British University is interested in what appears to effect student marks in a 2nd year core quantitative module called "Business Modelling". The manager collects some data for a randomly selected sample of students who all took this module and sat the same exam. They are courious as to whether the exam score is related to any of the following demographic or academic variables: Gender, A-level/GCSE results, Course of study, class attendance, completion of online quizzes, first language, result in pre-requisite module, first year average. The data set, referred to as the Grades data set from here on, includes 119 students. They decide to analyse this data using multiple regression. Details of the Grades data set are shown in Table 1.1 and data relating to the first 5 students in the data set is shown in Table 1.2.

Variable	Description	Categories
ID	Unique Identifier	N/A
Exam	Mark (out of 100%) attained for the Business Modelling module.	N/A
Maths	The highest (equivalent) level/grade attained in school age maths exams.	1 = A-level C or above. 2 = A-level D or lower. 3 = AS-level. 4 = GCSE C or above.
Degree	The degree course that the student in registered on.	1 = Management 2 = Finance 3 = Economics
Attendance	Percentage of classes attended by a student for the Business Modelling module. There were 20 hours of classes over the term.	N/A
Quiz	Percentage of online quizzes completed by a student for the Business Modelling module. There were 10 quizzes and each quiz was either completed or not.	N/A
Gender	Whether the student is male or female.	0 = Male 1 = Female
Language	Indicates whether the student's first language is English.	0 = Not English 1 = English
Pre-requisite	The result (out of 100%) obtained in the pre-requisite first year module.	N/A
Year1	The average mark received by a student across all their first year modules (out of 100%).	N/A

Table 1.1: Details of the variables in the Grades data set

Question 1 continued over page....\

Question 1 continued…\

ID	Exam	Maths	Degree	Attendance	Quiz	Gender	Language	Pre-requisite	Year1
1	39	1	1	10	40	0	1	38	25
2	39	4	1	65	20	0	0	64	36
3	36	4	1	20	60	0	1	83	92
4	42	2	1	55	20	1	1	60	63
5	50	4	1	45	100	0	0	73	89
:	:	:	:	:	:	:	:	:	:

Table 1.2: Sample of the Grades data set

The manager first graphs each of the explanatory variables against the response variable Exam, as well as each explanatory variable against all other explanatory variables. The following Figures 1.1 to 1.7 are some of these graphs produced by the manager.

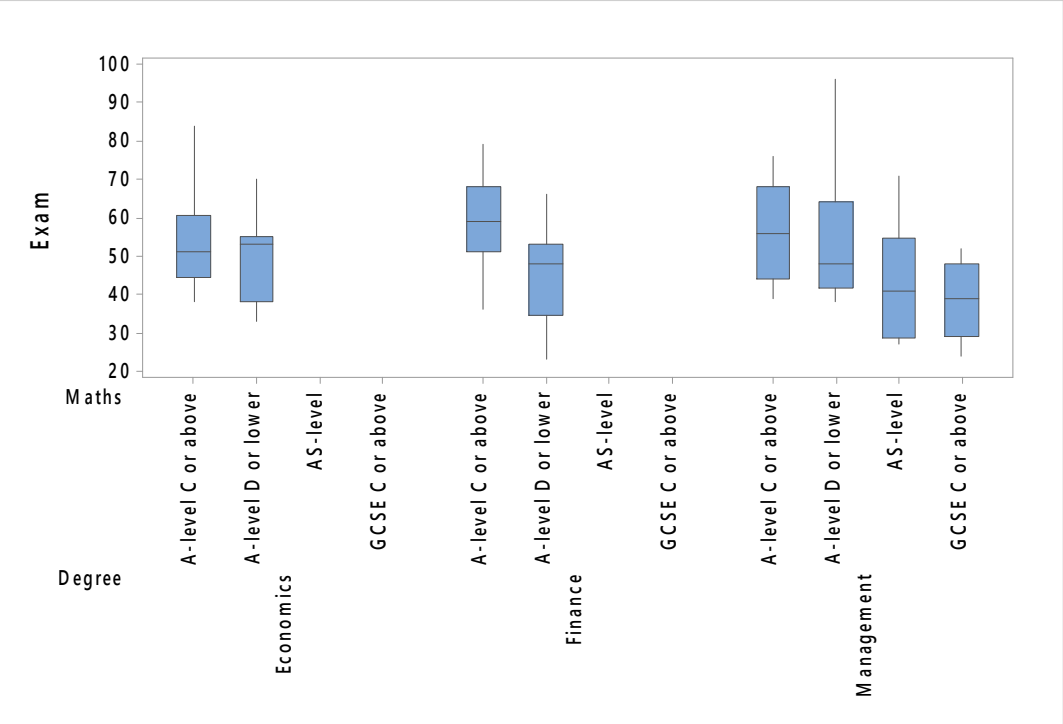


Figure 1.1: box plot of exam mark versus Maths pre-knowledge grouped by Degree course.

Question 1 continued over page....\

Question 1 continued…\

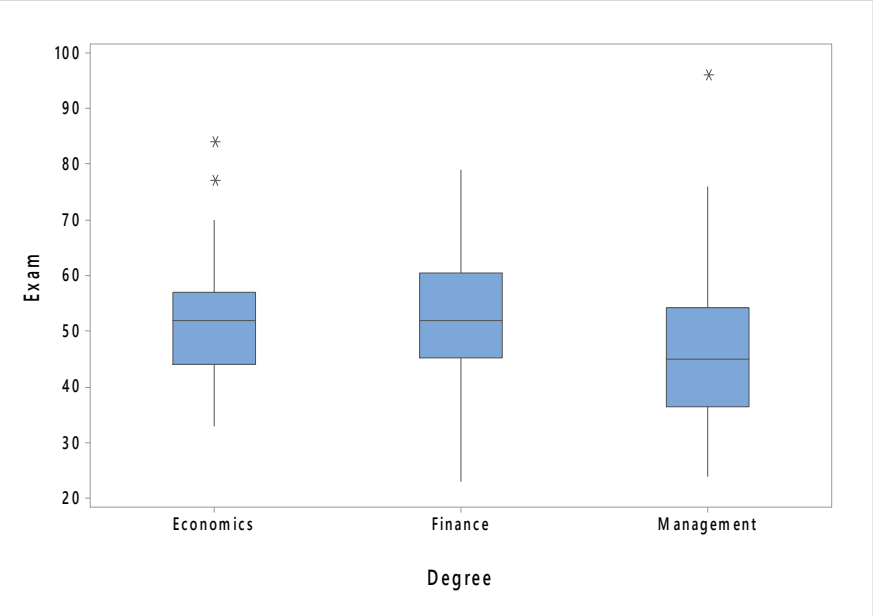


Figure 1.2: box plot of exam mark versus student degree course.

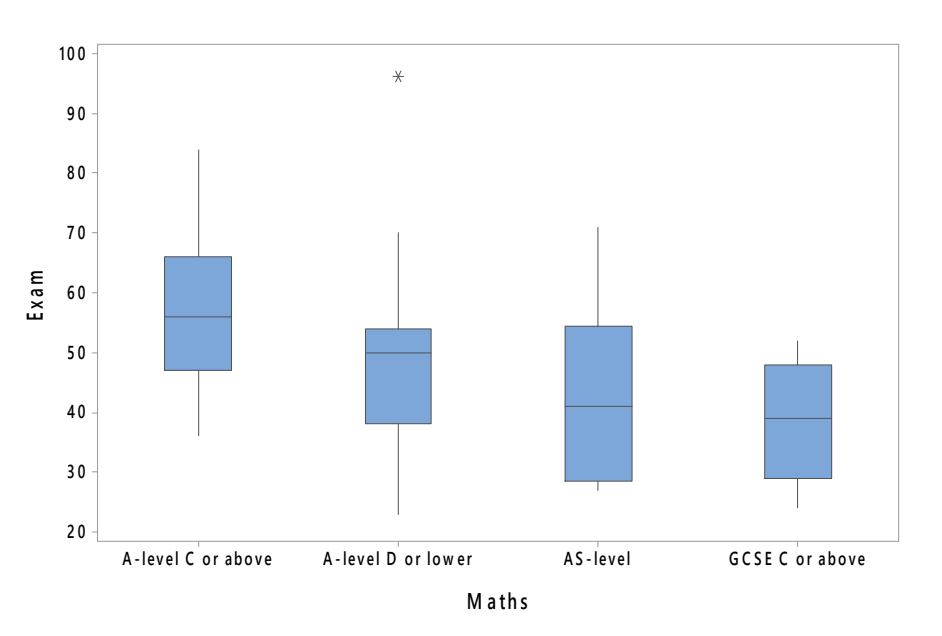


Figure 1.3: box plot of exam mark versus Maths pre-knowledge.

Question 1 continued over page....\

Question 1 continued…\

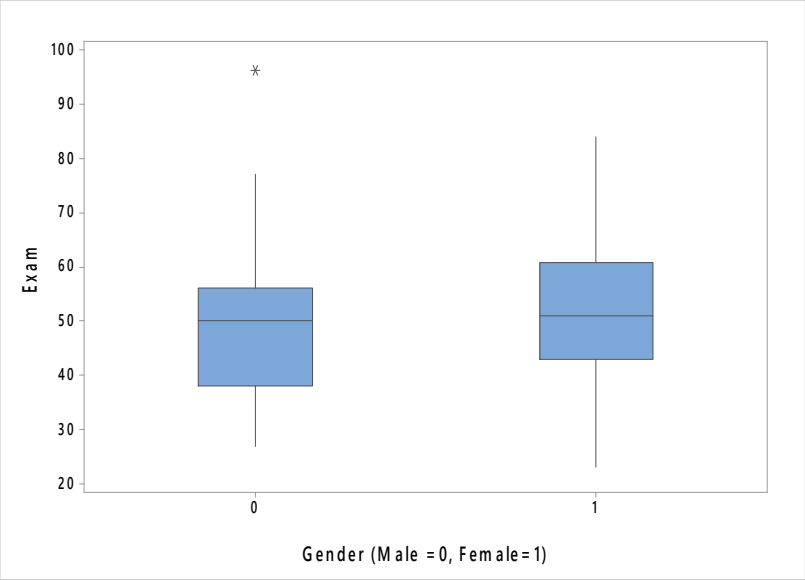


Figure 1.4: box plot of exam mark versus student gender.

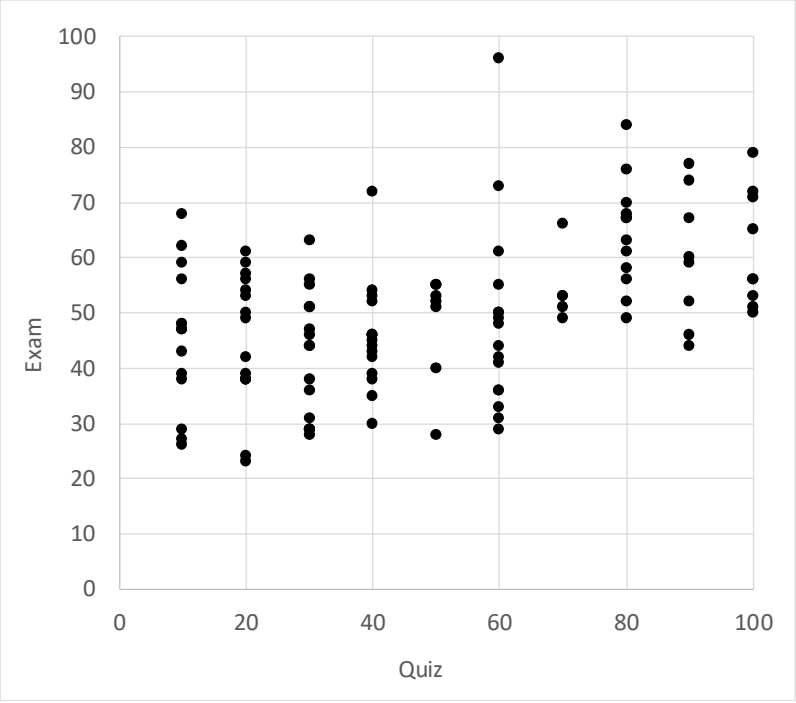


Figure 1.5: scatter plot of Exam versus completion of online quizzes (%).

Question 1 continued over page....\

Question 1 continued…\

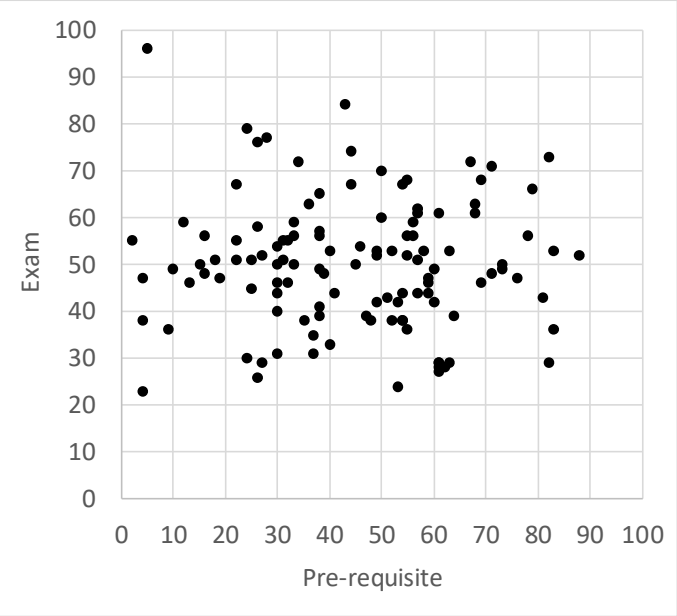


Figure 1.6: scatter plot of Exam versus mark on pre-requisite module (%).

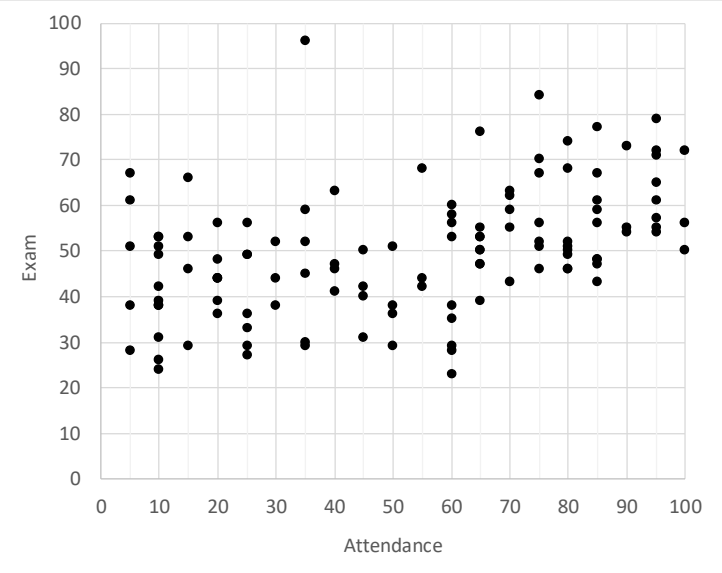


Figure 1.7: scatter plot of Exam versus attendance at module classes (%).

a) Carefully examine each of these graphs (Figures 1.1 to 1.7), fully explain and interpret the information each one gives the manager and hence give recommendations regarding the variables he could use in constructing his multiple regression model of this Grades data.
(Write your answers in the box over the page) **(10 marks)**

Question 1a) ANSWER BOX over page....\

Question 1a) ANSWER BOX …\

Sensible and correct full statements that commented on the important characteristics of the given graphs that are connected with constructing the regression model could gain marks:

- Fig 1.1: shows that there is a correlation/correspondence between Maths and Degree, as no students taking Finance or Economics have less than A-level (and/or Only some Management students have GCSE only.)
- Use only one of these variables (Maths, Degree) to avoid multicolinearity issues.
- Choose Maths as more information (and also) Fig 1.2 doesn't show any obvious correlation OR shows only weak correlation between Degree and Exam.

- Fig 1.1 & 1.3 shows correlation between Maths and Exam, so recommend including Maths variable in model construction.

- Figure 1.4 shows maybe very weak correlation between gender and Exam so could try this in model; or calculate correlation or try fitting in simple regression model with response to see further if should try in model.

- Fig 1.5 shows some correlation between Completion and Exam, so recommend including Completion variable in model construction.

- Fig 1.6 shows no obvious correlation between pre-req and exam, so could drop this from model construction or calculate correlation or try fitting in simple regression model with response to see further if should try in model.

- Fig 1.7 shows some correlation between Attendance and Exam, so recommend including Attendance variable in model construction.

- Figs 1.2 to 1.7 all show presence of possible outliers [This statement is only counted once for marking, since most graphs show these same outliers]
- The outliers should be investigated for validity and if valid, recommend fitting model with and without outliers to see what difference is made.

- Variance appears reasonably constant in all graphs [is also considered a valid comment – but like outliers, only counted once in marks].

(other valid, logical complete statements could also gain marks)

Question 1a) ANSWER BOX continued over page....\

Question 1a) ANSWER BOX continued…\

Question 1 continued over page....\

Question 1 continued…\

The manager decides to work with all the data collected (119 student cases) and selects Exam as the response variable and Maths, Attendance, Quiz, Gender, Language, Pre-requisite and Year1 as explanatory variables. He includes the categorical variables into the model as dummy variables using the coding as stated in Tables 1.3 to 1.5.

MATHS:	$d^M{}_1$	$d^M{}_2$	$d^M{}_3$
A-LEVEL C OR ABOVE.	0	0	0
A-LEVEL D OR LOWER.	1	0	0
AS-LEVEL.	0	1	0
GCSE C OR ABOVE	0	0	1

Table 1.3: Coding of variable Maths in the multiple regression model.

GENDER:	$d^G{}_1$
MALE	0
FEMALE	1

Table 1.4: Coding of variable Gender in the multiple regression model.

LANGUAGE:	$d^L{}_1$
NOT ENGLISH	0
ENGLISH	1

Table 1.5: Coding of variable Language in the multiple regression model.

Due to the small size of the data set the manager decides to use a manual backward stepwise regression approach. He fits the regression model using MSExcel. Figure 1.8 below shows selected MSExcel regression output for the first model fitted in the backward stepwise procedure.

Question 1 continued over page....\

Question 1 continued…\

SUMMARY OUTPUT

Regression Statistics	
R Square	?
Adjusted R Square	0.629545239
Standard Error	8.32828353
Observations	119

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	9	14532.88625	1614.76514	23.2808247	1.172E-21
Residual	109	7560.273415	69.3603066		
Total	118	22093.15966			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	22.85946758	3.463236695	6.60060793	1.5348E-09	15.99544504	29.7234901
Attendance	0.218545593	0.026748433	8.17040728	6.0044E-13	0.165531068	0.27156012
Quiz	0.210239235	0.028711497	7.32247559	4.4309E-11	0.153333982	0.26714449
Gender	0.090992319	1.667351813	0.05457296	0.95657858	-3.21364477	3.39562941
Language	-2.1828386	1.625500964	-1.34287131	0.18210333	-5.40452867	1.03885146
Maths						
A-level D or lower	-7.39444934	1.706296633	-4.33362476	3.2789E-05	-10.7762738	-4.01262489
AS-level	-18.0328069	3.872847183	-4.65621443	9.1356E-06	-25.7086642	-10.3569497
GCSE C or above	-22.4884642	3.162868756	-7.11014776	1.2735E-10	-28.7571672	-16.2197611
Pre-requisite	0.182530813	0.045262353	4.03272922	0.00010252	0.092822301	0.27223932
Year1	0.098510583	0.037477019	2.62855973	0.00981221	0.02423235	0.17278882

Figure 1.8: Regression output from MSExcel for the fitting of the Grades data regression model 1.

b) By examining the results in Figure 1.8 calculate the value of R Square and interpret the results of the (2 tailed) t-test for the c*oefficients of the variables: Quiz and Gender.*
(Write your answers in the box over the page)

(5 marks)

Question 1b) ANSWER BOX over page....\

Question 1b) ANSWER BOX …\

SS
14532.88625
7560.273415
22093.15966

R-square = SS Regression/SS Total = 14532.88625 / 22093.15966 = 0.657800

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	22.85946758	3.463236695	6.60060793	1.5348E-09	15.99544504	29.7234901
Attendance	0.218545593	0.026748433	8.17040728	6.0044E-13	0.165531068	0.27156012
Quiz	0.210239235	0.028711497	7.32247559	4.4309E-11	0.153333982	0.26714449
Gender	0.090992319	1.667351813	0.05457296	0.95657858	-3.21364477	3.39562941
Language	-2.1828386	1.625500964	-1.34287131	0.18210333	-5.40452867	1.03885146

Quiz

- P-value (of 4.4 x 10^{-11}) <<<< 0.05
- Therefore this coefficient is significant at the 5% (or 1%...) level **OR** it is significantly different to zero
- And the quiz variable is therefore explaining some of the variance in the response.

Gender

- P-value (of 0.957) >> 0.05
- Therefore this coefficient is NOT significant (insignificant) at the 5% (or 10%...) level **OR** it is NOT significantly different to zero
- And the gender variable is therefore NOT significantly explaining the variance in the response.

Question 1b) ANSWER BOX continued over page....\

Question 1b) ANSWER BOX continued…\

Question 1 continued over page....\

Question 1 continued…\

After completing the manual backward stepwise regression, the manager decides upon his final model. The MSExcel regression output for this final model is shown in Figure 1.9 and graphs of the residuals for this model are shown in Figure 1.10.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.807444643
R Square	0.651966852
Adjusted R Square	0.630018816
Standard Error	8.322958534
Observations	119

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	7	14404.00776	2057.71539	29.7050197	9.51868E-23
Residual	111	7689.151903	69.2716388		
Total	118	22093.15966			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	22.1925786	3.390593017	6.54533838	1.9024E-09	15.47389219	28.91126501
Attendance	0.213414673	0.026379889	8.0900521	8.2306E-13	0.161141163	0.265688184
Quiz	0.210702285	0.027790794	7.58172957	1.1125E-11	0.155632971	0.265771598
Maths						
A-level D or lower	-7.586654382	1.696713807	-4.4713813	1.8852E-05	-10.9488061	-4.22450271
AS-level	-17.2184202	3.767048385	-4.57079879	1.2683E-05	-24.683078	-9.75376234
GCSE C or above	-21.77280814	2.987987735	-7.28677962	4.931E-11	-27.6937053	-15.851911
Pre-requisite	0.17944275	0.043817831	4.09519924	8.0391E-05	0.092614792	0.266270708
Year1	0.098421579	0.037013073	2.65910317	0.00899358	0.025077704	0.171765453

Figure 1.9: Excel regression output for the final model fitted to the Grades data.

c) Using the output in Figure 1.9, write down the fitted model equation. Fully interpret, in context, the regression coefficients for *Quiz and Maths – 'AS-level'*.
(Write your answers in the boxes over the page). **(8 marks)**

Question 1c) ANSWER BOX over page....\

Question 1c) ANSWER BOX …\

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	22.1925786	3.390593017	6.54533838	1.9024E-09	15.47389219	28.91126501
Attendance	0.213414673	0.026379889	8.0900521	8.2306E-13	0.161141163	0.265688184
Quiz	0.210702285	0.027790794	7.58172957	1.1125E-11	0.155632971	0.265771598
Maths						
A-level D or lower	-7.586654382	1.696713807	-4.4713813	1.8852E-05	-10.9488061	-4.22450271
AS-level	-17.2184202	3.767048385	-4.57079879	1.2683E-05	-24.683078	-9.75376234
GCSE C or above	-21.77280814	2.987987735	-7.28677962	4.931E-11	-27.6937053	-15.851911
Pre-requisite	0.17944275	0.043817831	4.09519924	8.0391E-05	0.092614792	0.266270708
Year1	0.098421579	0.037013073	2.65910317	0.00899358	0.025077704	0.171765453

Predicted Exam mark = 22.193 + 0.213 Attendance + 0.211 Quiz – 7.587 A-level D or lower (or $d^M{}_1$) – 17.218 AS-level ($d^M{}_2$) – 21.773 GCSE ($d^M{}_3$) + 0.179 Pre-req + 0.098 Yr1

Quiz:

- The exam mark increases by 0.211 *(or by 2.11)*
- On average
- For a one *(or 10% if quoted 2.11 above)* percent increase in quiz completion or for one extra completed quiz
- When all other variables remain fixed.

Maths: 'AS-level'.

- The exam mark decreases by 17.218
- On average
- If the student has AS-level maths as their highest (equivalent) level/grade attained in school age maths exams.
- Rather than A-level maths grade C or higher
- When all other variables remain fixed.

[other valid ways of stating the above, or different order of statements was fine]

Question 1c) ANSWER BOX continued over page....\

Question 1c) ANSWER BOX continued…\

Question 1 continued over page....\

Question 1 continued…\

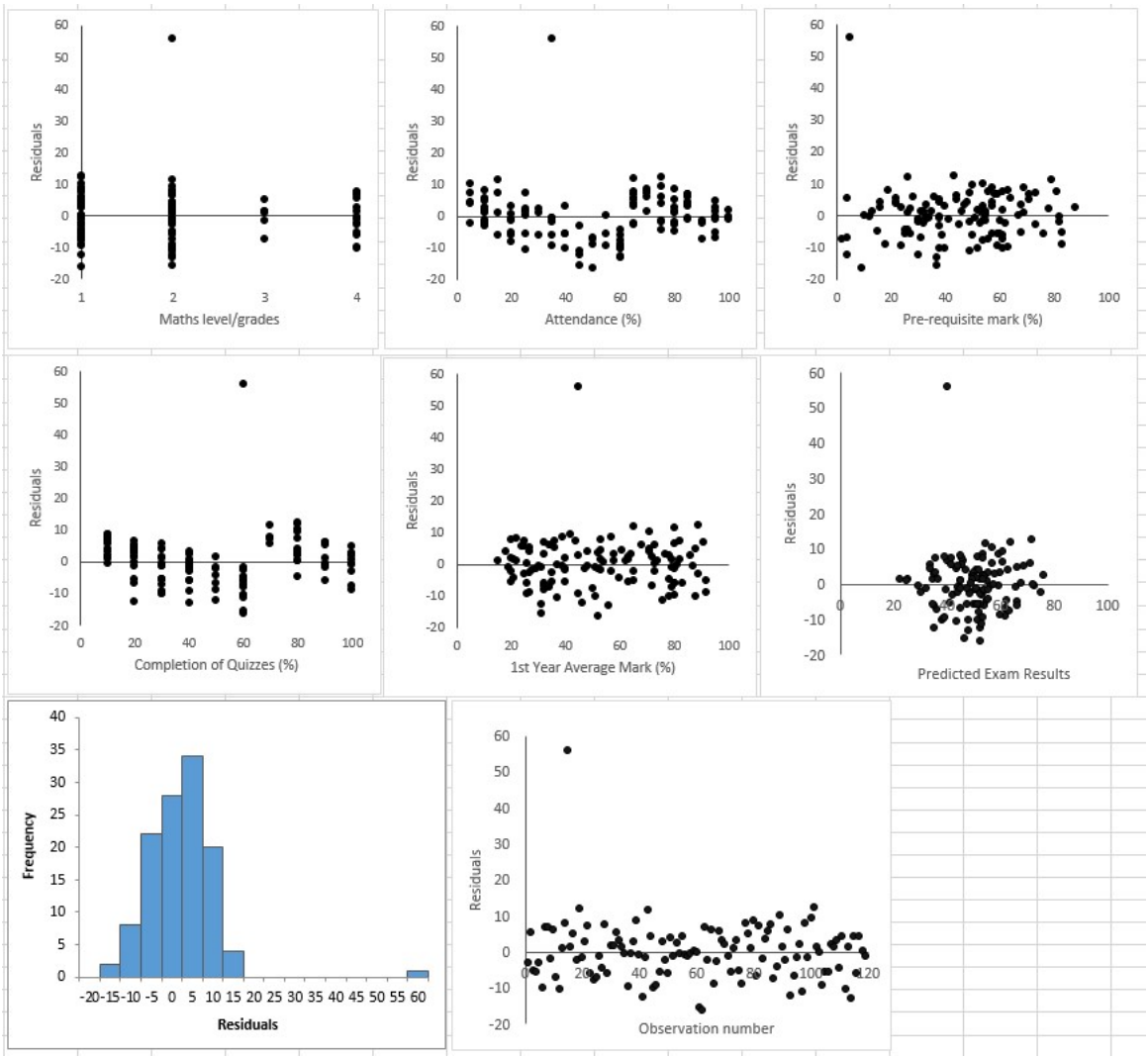


Figure 1.10: Plots of the residuals from the final regression model fitted to the house data.

d) By examining the information in Figures 1.9 and 1.10, comment on the adequacy of the fitted regression model and hence describe any further actions you might want to take as a result. **(Write your answers in the box over the page).** **(10 marks)**

Question 1d) ANSWER BOX over page....\

Question 1d) ANSWER BOX …\

Must have at least one statement from each category (A and B) to score full marks:

A) Comments of adequacy:

- All residual graphs show an obvious outlier
- Graphs of residuals versus Quiz and Attendance show possible strange S (curved, wave…) pattern.
- Graphs of residuals versus pre-req, year1, maths, observation number and pred exams appear reasonable randomly scattered
- Histogram (even if ignore outlier) is a little skewed.
- The residual scatter plots appear to be centred around zero.
- Variance in all scatter plots appears reasonably constant, if ignore outlier.
- Maths level/grades plot may possibly have some decreasing variance
- Adjusted R-squared (R-square) indicates model is explaining around 63% (65%) of the variation in exam marks.
- The Adjusted R-squared value is reasonably high, but around 37% of the variation in exam mark is not explained by the model.
- All explanatory variables are significant in the model ($p<0.05$)
- F-test is significant (<0.05) so model is significantly explaining variability in the response better than an intercept only model.

B) Suggestions of actions to take:

Re. outlier:

- Should check this observation to see if valid value
- If valid value, try fitting model without this outlier to see what difference it makes
- If not valid value, should delete case from sample

Re. pattern in residuals (attendance and quiz):

- Consider transformation of variables

Re. Adj R-square value:

- Could try to find other variables to include in model to explain more of the other 37% of variation in exam mark.
- Increase sample size if possible: since working with 11 potential explanatory variables, 119 observations is quite a small data sample size.

(other valid, relevant and complete statements could gain marks)

Question 2 over the page....\

Question 2 (16 marks)

Cedric Achebe works as a data analyst in the United States National Travel and Tourism Office (NTTO). The goal of NTTO is to enhance the international competitiveness of the U.S. travel and tourism industry and increase its exports, thereby creating U.S. employment and economic growth. Cedric's job entails working with data to help the NTTO to make informed decisions regarding the future economic conditions of the country.

Cedric is asked to analyse data (from this point on referred to as the travel data) of the number of visits to the US from Mexico. These observations are given for each quarter of the year from 1st quarter (Q1) of 2009 to the 4th quarter (Q4) of 2019 (44 data points in total). Table 2.1 below shows the months included within each quarter. Cedric generates a time series plot of the travel data using MSExcel, as shown in Figure 2.1 below.

Quarter 1	Quarter 2	Quarter 3	Quarter 4
January	April	July	October
February	May	August	November
March	June	September	December

Table 2.1: Months of the year included within each Quarter of the year.

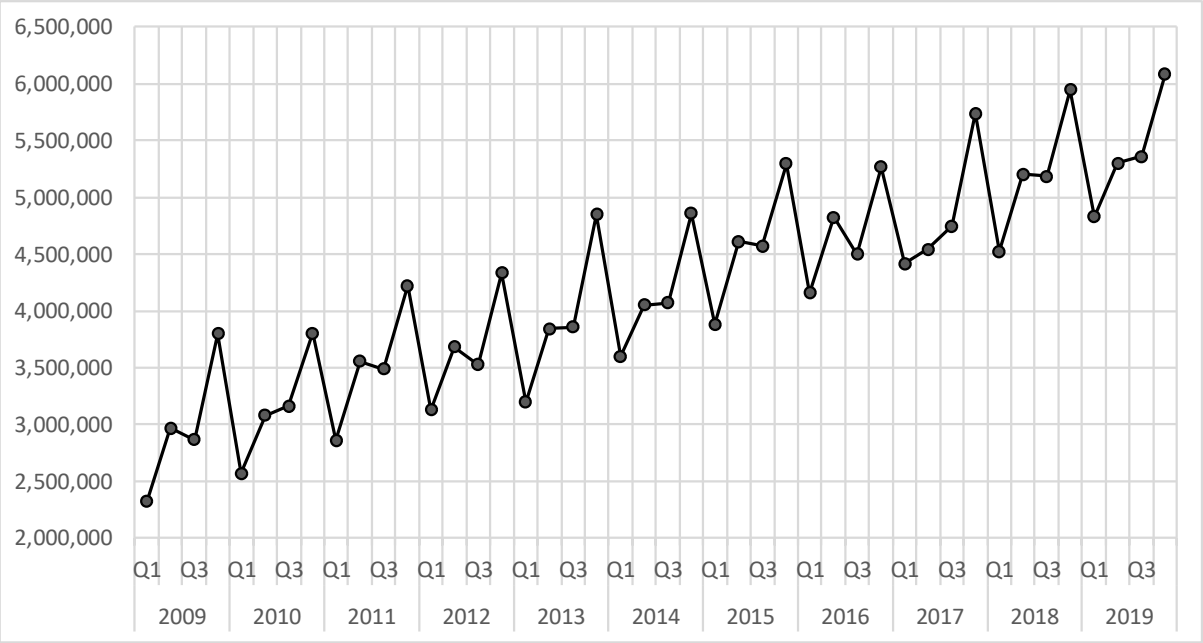


Figure 2.1: Time series plot of the travel data – 2009 Q1 to 2019 Q4 (44 time points)

Question 2 continued over page....\

Question 2 continued …\

a) By looking at the time series plot in Figure 2.1, describe and comment on the main features of this time series. **(Put your answer in the box below)**.

(8 marks)

- Seasonality of period 4
- Upward linear trend
- No obvious outliers by eye, though Q2 of 2017 seems a little unusual
- No obvious cycles
- Constant variance
- Highest visits in 4th Quarter
- Lowest visits in 1st Quarter
- Non-stationary data

(any other applicable, sensible, complete statement can gain marks)

Question 2 continued over page....\

Question 2 continued …\

b) Cedric decides to fit a multiple regression model to this data to forecast future visitor numbers. Describe and explain in detail the structure of a multiple regression model that ought to fit sufficiently well to the important characteristics in the travel data.
(Put your answer in the box below). **(8 marks)**

$$Predicted\ visitor\ numbers = \beta_0 + \beta_1 t + \beta_2 \boldsymbol{Q_1} + \beta_3 \boldsymbol{Q_2} + \beta_4 \boldsymbol{Q_3} + \varepsilon$$

The response variable is the number of visitors. β_0 is the intercept. ε is the errors.
$\beta_i, for\ i = 1,2,3,4$, are the coefficients for each explanatory variable.

The explanatory variables are:

- Time, t, included in the model as a straight forward unit count (t = 1, 2, 3, …, 27) to model the linear upward trend.

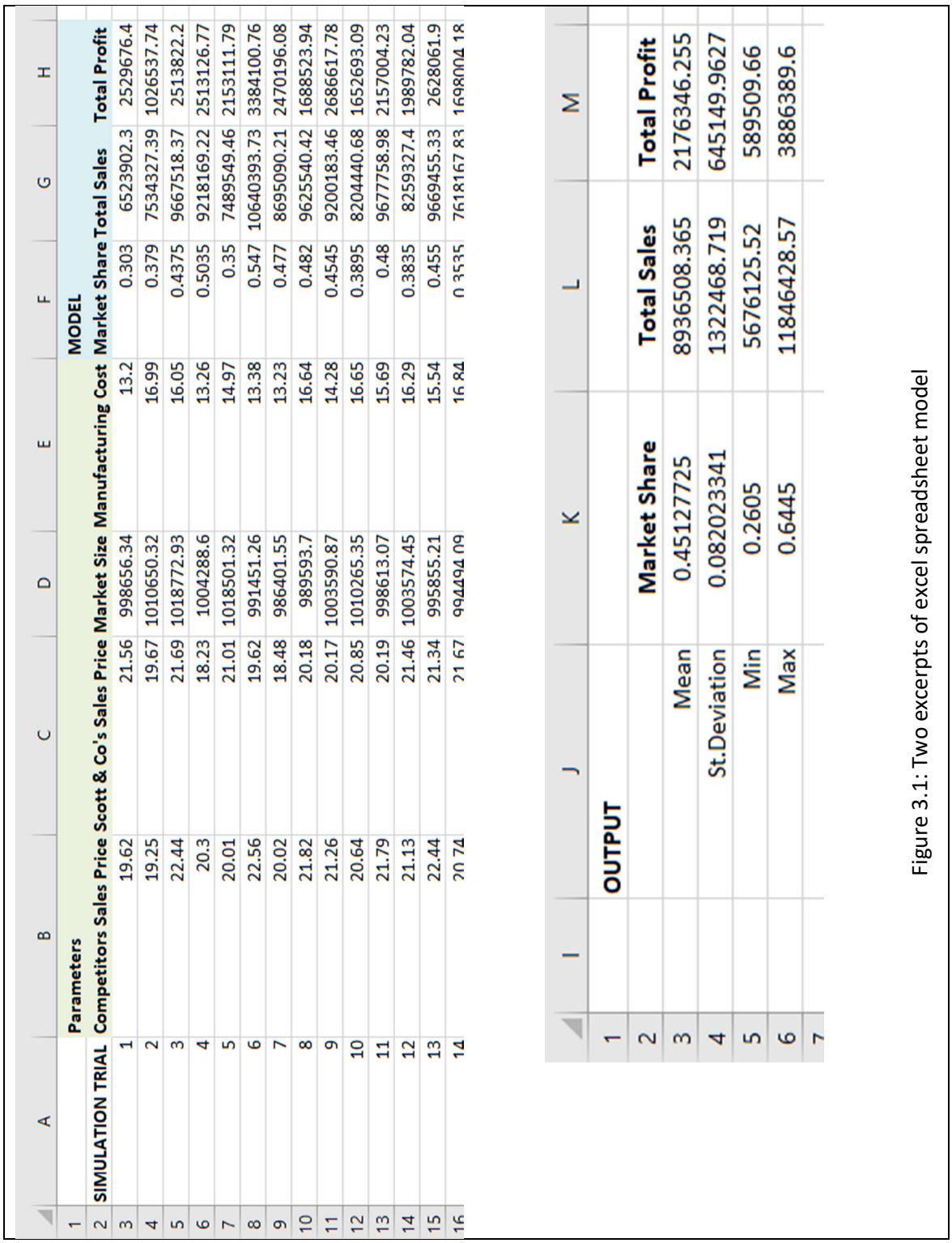
- Seasonality of period 4 included using three dummy variables, coded as shown in table below:

QUARTERS:	$\boldsymbol{Q_1}$	$\boldsymbol{Q_2}$	$\boldsymbol{Q_3}$
1ST QUARTER	0	0	0
2ND QUARTER	1	0	0
3RD QUARTER	0	1	0
4TH QUARTER	0	0	1

Question 3 over the page....\

Question 3 (17 marks)

Figures 3.1 below shows two excerpts from a spreadsheet model used to model the risk associated with the launch of a new product.



	A	B	C	D	E	F	G	H
1		Parameters			MODEL			
2	SIMULATION TRIAL	Competitors Sales Price	Scott & Co's Sales Price	Market Size	Manufacturing Cost	Market Share	Total Sales	Total Profit
3	1	19.62	21.56	998656.34	13.2	0.303	6523902.3	2529676.4
4	2	19.25	19.67	1010650.32	16.99	0.379	7534327.39	1026537.74
5	3	22.44	21.69	1018772.93	16.05	0.4375	9667518.37	2513822.2
6	4	20.3	18.23	1004288.6	13.26	0.5035	9218169.22	2513126.77
7	5	20.01	21.01	1018501.32	14.97	0.35	7489549.46	2153111.79
8	6	22.56	19.62	991451.26	13.38	0.547	10640393.73	3384100.76
9	7	20.02	18.48	986401.55	13.23	0.477	8695090.21	2470196.08
10	8	21.82	20.18	989593.7	16.64	0.482	9625540.42	1688523.94
11	9	21.26	20.17	1003590.87	14.28	0.4545	9200183.46	2686617.78
12	10	20.64	20.85	1010265.35	16.65	0.3895	8204440.68	1652693.09
13	11	21.79	20.19	998613.07	15.69	0.48	9677758.98	2157004.23
14	12	21.13	21.46	1003574.45	16.29	0.3835	8259327.4	1989782.04
15	13	22.44	21.34	995855.21	15.54	0.455	9669455.33	2628061.9
16	14	20.74	21.67	994494.09	16.84	0.3535	7618167.83	1698004.18

	I	J	K	L	M
1		OUTPUT			
2			Market Share	Total Sales	Total Profit
3		Mean	0.45127725	8936508.365	2176346.255
4		St.Deviation	0.082023341	1322468.719	645149.9627
5		Min	0.2605	5676125.52	589509.66
6		Max	0.6445	11846428.57	3886389.6
7					

Figure 3.1: Two excerpts of excel spreadsheet model

Question 3 continued over page....\

Question 3 continued …\

In this model there are 2000 simulation trials (i.e. the last simulation trial is in row 2002). In this model there are 4 *inputs* that can be used to estimate the company's sales profit for this new product:

- *Competitor's sales price* for a competing product (£),
- *Scott & Co's sales price (£)*,
- *market size (an absolute count)*,
- *manufacturing cost* for the new product (£).

All four inputs have uncertainty / variability in them. The following table shows the distributions that the company have decided best reflect the variability of the four inputs.

NAME	DISTRIBUTION
Competitor's Sales Price	Uniform(£19, £23)
Scott & Co's Sales Price	Uniform(£18, £22)
Market Size	Normal(1,000,000, 10,000)
Manufacturing Cost	Uniform(£13, £17)

The model *outputs* (results) of interest to the company are:

- *Market Share* (which depends on Scott & Co's sales price and the competitor's sales price): expressed in proportion rather than percentage.
- *Total Sales* (which depends on the market size, market share and Scott & Co's sales price): expressed in £s.
- *Total Profit* (which depends on the market size, percent market share, Scott & Co's sales price and the manufacturing costs): expressed in £s.

The formulae for calculating these 3 outputs are as follows:

***Market Share*:** 0.4 – (0.05*Scott & Co's Sales Price) + (0.05*Competitor's Sales Price)
***Total Sales*:** Market Size * Market Share*Scott & Co's Sales Price
***Total Profit*:** Market Size* Market Share*(Scott & Co's Sales Price – Manufacturing Cost)

Question 3 continued over page....\

Question 3 continued …\

a) Write the Excel code required to be in cells B7, G3, H9, K3 and L5 for this model to work correctly.
(Put your answers in the box below). **(10 marks)**

B7: =ROUND(19+((23-19)*RAND()),2)

G3: =ROUND(D3*F3*C3,2)

H9: =ROUND((D9*F9)*(C9-E9),2)

K3: =AVERAGE(F3:F2002)

L5: =MIN(G3:G2002)

Question 3 continued over page....\

Question 3 continued …\

b) Explain in detail what random numbers are.
(Put your answer in the box below). **(7 marks)**

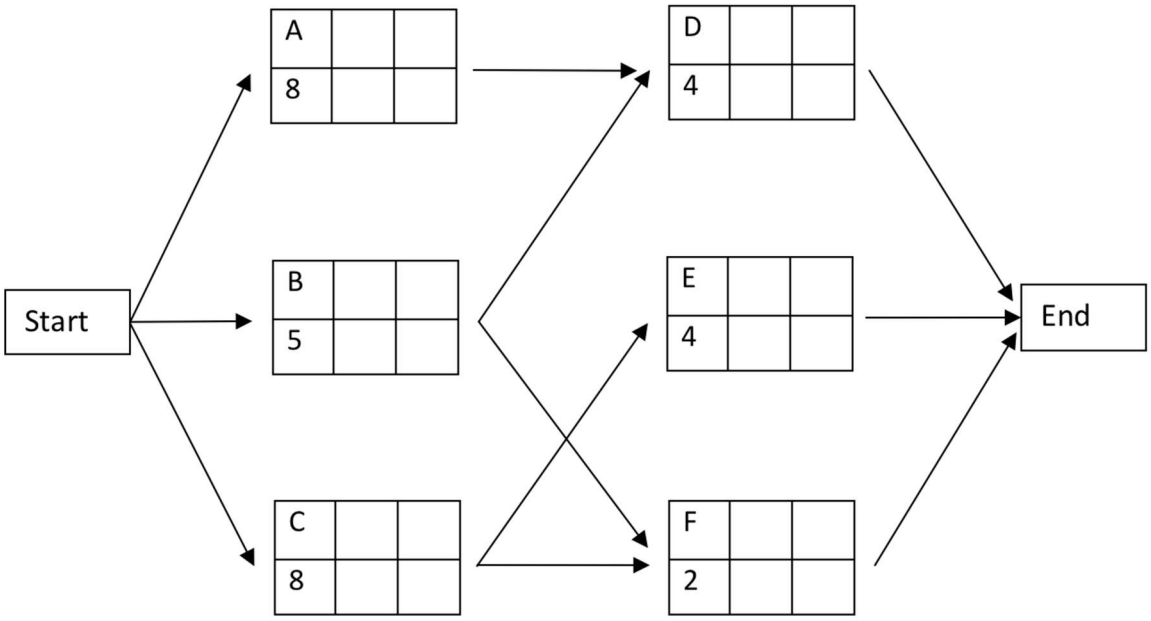
Random numbers are a sequence of numbers that appear in a random order.
They can be whole numbers 0-9, 0-99 or real numbers from 0-1 (not including 1).
They must have two properties:

- The uniform property: each random number has an equal chance of appearing at any point in the random number sequence.
- The independence property: Once a number has appeared in a sequence, this does not affect its chance of appearing again, or the chance of any other number appearing.

Question 4 over the page....\

Question 4 (33 marks)

Consider the following project network with activities A to F, and activity durations (in weeks) as indicated in the figure below. The precedence relations between activities are illustrated with arrows.



a) Determine the project completion time and identify critical path(s). For each activity, find the earliest and latest start-times and finishing-times, as well as the slack. **You can simply enter the numbers in the graph above**.

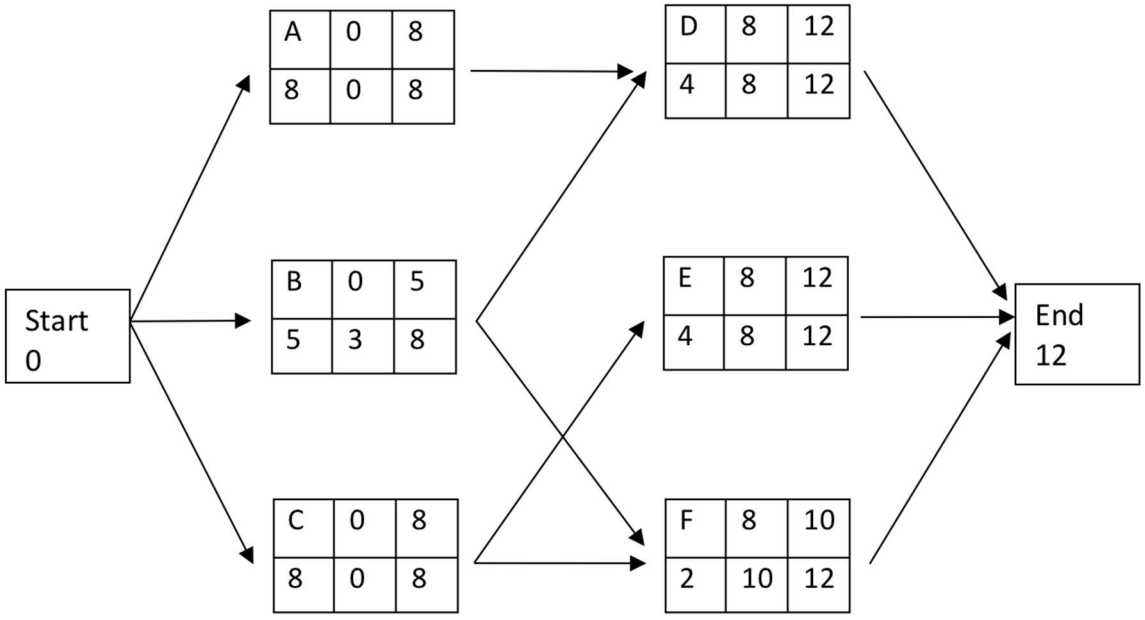
(Please write the additional information in the box below as needed). (13 marks)

Project completion time = 12 weeks

Slack time: S(A) = 0, S(B) = 3, S(C) = 0, S(D)= 0, S(E) =0, S(F) =2

Activity B can be delayed by at most three weeks without delaying the project. Similarly, activity F can be delayed by at most 2 weeks without delaying the project. Activities A, C, D and E are critical. Therefore, delaying these activities will increase the project completion time.

Critical paths: A – D and C – E



Question 4 continued over page....\

Question 4 continued…\

b) What happens if we increase the duration of activity B by 3 weeks and decrease the durations of activities D and F by 2 weeks and 1 weeks, respectively? Does the project completion time change, why? Explain your reasoning.

(Write your answer in the box below). (4 marks)

Activity B has a slack of 3 weeks. Therefore, increasing duration by 3 weeks does not delay the project.

Activity F has a slack of 2 weeks. Decreasing duration of activity F will only increase its slack.

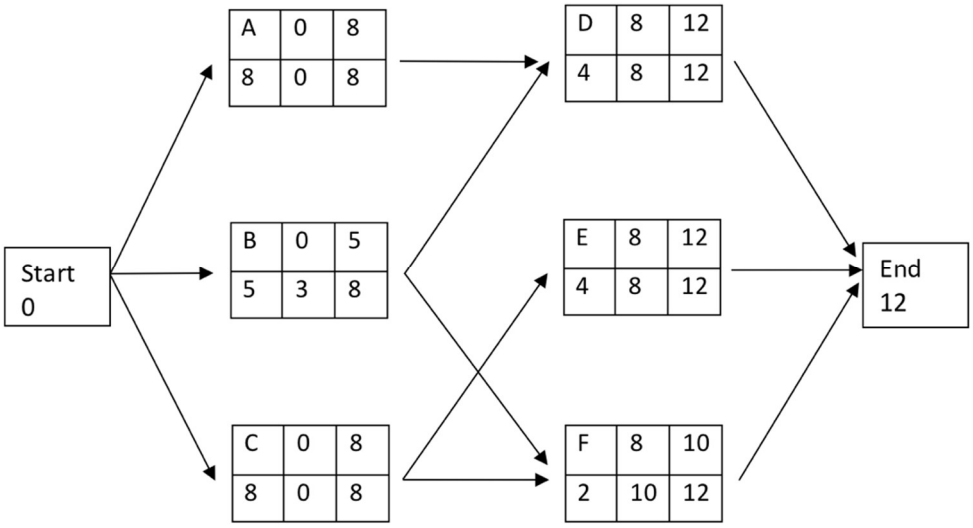
Activity D does not have any slack. Therefore, changing its duration may affect the project completion time. However, we have two parallel critical paths in this problem. Even we complete activity D, we still have to wait for activity E. Therefore, project completion time does not change.

Question 4 continued over page....\

Question 4 continued…\

c) The project manager wants to shorten the overall duration of the project found in part (a) by 1 week. Assume that each activity can be crashed by at most 1 week and the associated costs are listed in the table below. Which activity or activities would you choose to crash by 1 week to obtain the minimum total cost? Why? Explain your answer.
(**Write your answer in the box below and over the page as needed**.) **(10 marks)**

Activity	**A**	**B**	**C**	**D**	**E**	**F**
Cost per week	15	10	8	12	7	5



Since activities B and F have positive slack, shortening their duration by one week does not change the project completion time.

On the other hand, we have two critical paths A-D and C-E. Crashing just one activity does not change the project completion time. Our options are:

Crash A and C with cost 23

Crash A and E with cost 22

Crash C and D with cost 20

Crash D and E with cost 19 (cheapest) Therefore, we crash D and E.

Question 4 continued over page....\

Question 4 continued…\

d) Consider the precedence relation between activities A, B and D. Suppose that x_i denote the starting time of activity i, for $i = A, B, D$. Formulate the linear programming constraints that represent the precedence relations between activities A, B and D.

(Write your answer in the box below). (6 marks)

x_A: starting time of activity A

x_B: starting time of activity B

x_D: starting time of activity D

$x_A + 8 \leq x_D$

$x_B + 5 \leq x_D$

$x_A \geq 0$

$x_B \geq 0$

End of Paper