



Module 4

Modelling in management: Optimisation and forecasting

In this module we address the advantages and possible dangers of modelling; the use of Markov processes; simulation, linear and non-linear programming; forecasting in decision making; and time series and linear regression. We will also look at applications in Excel.

Upon completion of this module students will be able to:



Outcomes

- **understand** the philosophy, methods and risks of modelling;
- **apply** modelling to stochastic processes;
- **solve** optimisation problems with linear, integer and non-linear programming techniques;
- **use** simulation and forecasting methods, including regression, to inform decisions; and
- **implement** modelling tools in Excel.

Unit 13

Optimisation and forecasting

Introduction to modelling and decision making

Upon completion of this unit students will be able to:



Outcomes

- **describe** the steps in the modelling process;
- **explain** the importance and limitations of models;
- **distinguish** between modelling assumptions and reality; and
- **apply** Markov chains to the modelling of stochastic processes.

Advise students of possible resources (if necessary). Research and reports can be done by groups.

The modelling process and model risk



Case study:
Model risk and crises

Choose one of the historical financial crises mentioned in this unit. Research it and write a short report. If possible, mention the positive and/or negative roles that quantitative models played or could have played in the crisis.

Solutions

It is the Lecturer's decision whether the student's work is evaluated.

Models and assumptions

Activity 4.1



Activity
Consider random walk

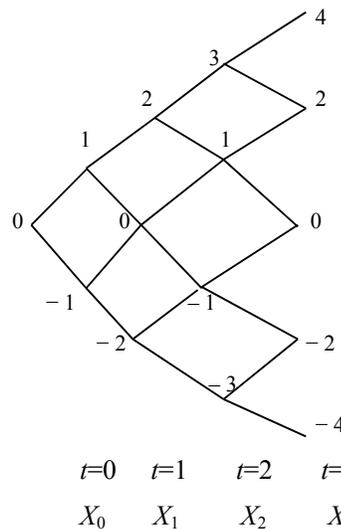
1. Consider the random walk of Example 2: suppose you are standing at point $X_0 = 0$. A coin is flipped every second and if it is heads you take one step forward and if it is tails you take one step backwards. Your position at any time t is a random variable X_t .
 - a) Construct a binomial tree with four time steps for X_t .
 - b) Show that the expected value $E[X_t] = 0$ for each t .
 - c) Prove that $\text{var}[X_t] = t$.



- d) Show that at each node on the tree the value of X_t equals the expected value of the two values X_{t+1} following from that node. Use this fact to reason that if the values of X at the end nodes are given, then you can work backwards to find the value X_0 at time $t = 0$.
2. Explain the link between the efficient market hypothesis and the Markov model.

Solutions

1. a) The process is a simple random walk.



b)
$$E[X_3] = \frac{1}{8}(X_0 + 3) + (3 \times \frac{1}{8})(X_0 + 1) + (3 \times \frac{1}{8})(X_0 - 1) + \frac{1}{8}(X_0 - 3)$$

$$= X_0$$

$$= 0$$

Probabilities are the product of individual probabilities of each individual independent step.

(Note that there are three paths each leading to positions 1 and -1 at $t = 3$.)

This proof can be repeated for each time t .

- c) At $t = 2$:

$$\begin{aligned} \text{var}[X_2] &= E[X_2^2] - (E[X_2])^2 \\ &= (2)^2 (\frac{1}{4}) + (0)^2 (2 \times \frac{1}{4}) + (-2)^2 (\frac{1}{4}) \\ &= 1 + 1 \\ &= 2 \\ &= t \end{aligned}$$

This proof can be repeated for each time t .

- d) Consider final nodes $X_4 = 4$ and $X_4 = 2$, which follow from $X_3 = 3$: the probabilities of reaching positions $X_4 = 4$ or $X_4 = 2$ given that we are at position $X_3 = 3$ are $\frac{1}{2}$ each. The expected value we are computing is a *conditional* expectation.

$$\begin{aligned} E[X_4 = 4 \text{ and } X_4 = 2 \mid X_3 = 3] &= 4(\frac{1}{2}) + 2(\frac{1}{2}) \\ &= 3 \\ &= X_3 \end{aligned}$$

Properties a–d are the properties of random walk or Wiener process X_t . This stochastic process X_t is called a martingale.

Moving backwards from pairs of values in this way brings us to the value at $t = 0$. This allows us to find the price of products such as options from their final values.

2. Briefly, our possible positions at any time t are fully determined by our information of our position at the previous time $t - 1$. Information about our position at $t - 2$, for example, does not add any additional knowledge. This is one way of looking at the efficient market hypothesis.



Unit 14

Simulation



Outcomes

Upon completion of this unit students will be able to:

- **describe** the steps in running a simulation;
- **generate** random numbers; and
- **understand** how to analyse uncertain problems using simulation.

Simulation imitates the behaviour of stochastic situations. Monte Carlo simulation uses random number generation to simulate possible outcomes.

Activities can be constructed along the lines of Activity 4.3. It is not possible to give specific answers to such problems, since random numbers are, well, random.

Unit 15

Linear and non-linear programming

Upon completion of this unit students will be able to:



Outcomes

- **calculate** the optimal value of quantities under constraints using linear programming methods;
- **understand** the basics of integer programming;
- **determine** maxima or minima with non-linear programming; and
- **apply** the methods using Excel.

Introduction to linear programming (LP)

Activity 4.6



Activity

Find the maximum

What will you do?

1. Find the maximum of the function F in Activity 4.5 using the corner-point method.
2. Find the minimum of the function F in Activity 4.5 using the graphic and corner-point methods.
3. Find the maximum of function $G(x, y) = 4x + 2y + 8$ subject to the constraints $0 \leq x \leq 3$, $y \geq 0$ and $y - x \leq 1$.
4. Find the minimum of function $F(x, y) = x - 2y + 6$ subject to the constraints $0 \leq x \leq 3$ and $0 \leq y \leq 3$.

Solutions

1. $F(x, y) = 2x + 4y + 5$. The units of F are in thousands of dollars. Variables x and y represent quantities of two items A and B that are sold, and their units are in thousands. The constraints are $x \geq 0$, $y \geq 0$ and $x + y \leq 1$. (Quantities of items sold must be non-negative and we cannot sell more than 1,000 items altogether.)

Corner-point method: the feasible region has corner points $(0, 0)$, $(1, 0)$ and $(0, 1)$.

$$F(0, 0) = 2(0) + 4(0) + 5 = 5$$

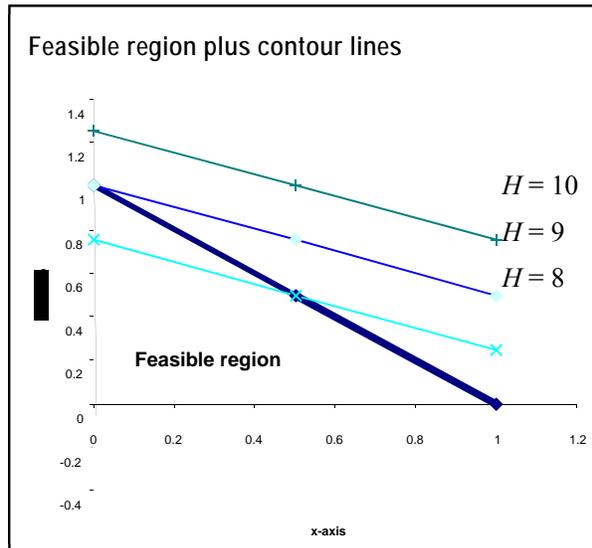
$$F(1, 0) = 2(1) + 4(0) + 5 = 7$$

$$F(0, 1) = 2(0) + 4(1) + 5 = 9$$

The maximum is at $(x, y) = (0, 1)$ and the value is 9.



2.



2. From the graph, it is clear that the contour line $(y = -0.5x + \frac{H-5}{4})$ which touches the feasible region at $(0, 0)$ gives the minimum of F .
 $\text{Min } F = F(0, 0) = 5$

3. $G(x, y) = 4x + 2y + 8$ subject to the constraints $0 \leq x \leq 3$, $y \geq 0$ and $y - x \leq 1$.

Corner points of feasible region: $(0, 0)$, $(3, 0)$, $(3, 4)$ and $(0, 1)$.

$$G(0, 0) = 8$$

$$G(3, 0) = 20$$

$$G(3, 4) = 28$$

$$G(0, 1) = 10$$

The maximum is at $(x, y) = (3, 4)$ and the value is 28.

4. Minimum of function $F(x, y) = x - 2y + 6$ subject to the constraints: $0 \leq x \leq 3$ and $0 \leq y \leq 3$

Corner points of feasible region: $(0, 0)$, $(3, 0)$, $(3, 3)$ and $(0, 3)$.

$$F(0, 0) = 6$$

$$F(3, 0) = 9$$

$$F(3, 3) = 3$$

$$F(0, 3) = 0$$

The minimum is at $(x, y) = (0, 3)$ and the value is 0.

Introduction to non-linear programming

Activity 4.7



Activity
Maximise and
minimise

What will you do?

Classify the problems as LP, ILP or NLP. Draw all feasible regions and a level line for the object function where possible.

1. Maximise $F(x, y) = x + y$ subject to the constraints $0 \leq x \leq 3$, $y \geq 0$ and $y^2 + x^2 \leq 1$.
2. Minimise $G(x, y) = 2x - 3y$ subject to the constraints $x, y \geq 0$ and $y + x \leq 1$.
3. Maximise $F(x, y) = xy + 1$ subject to the constraints $0 \leq x \leq 3$ and $0 \leq y \leq 1$.
4. Maximise $K(x, y) = x + y$ subject to the constraints $0 \leq x \leq 3$, $y \geq 0$, $y + x \leq 3$ and x and y both integers.

Solutions

1. NLP. Feasible region is the quarter circle segment of radius 1 in the first quadrant. Level lines have the equation $y = H - x$ for varying H .
2. LP. Feasible region is the triangle in the first quadrant with corner points $(0, 0)$, $(0, 1)$, $(1, 0)$. Level lines have the equation $y = (2/3)x - (1/3)H$ for varying H .
3. NLP. Feasible region is the rectangle in the first quadrant with corner points $(0, 0)$, $(3, 0)$, $(3, 1)$ and $(0, 1)$. Level lines have the equation $y = (H - 1)/x$ for varying H .
4. ILP. Feasible region consists of the 10 integer pairs (x, y) inside the triangle in the first quadrant with corner points $(0, 0)$, $(0, 3)$ and $(3, 0)$. Level lines have the equation $y = H - x$ for varying H .

Lecturers should review graphs for correctness.



Applications in Excel

The Solver tool for linear programming

Activity 4.8



Activity

Minimise the cost

What will you do?

The cost C associated with manufacturing cables depends on the cost of copper wire and plastic tubing. Let x denote the number of metres of copper wire and y the number of metres of plastic tubing. There are fixed costs of INR 200,000 before manufacturing even starts. The cost of copper wire is INR 200 per metre and the cost of plastic tubing is INR 5.45 per metre. Material is only delivered if you order a combined length of copper and plastic that is more than 500 m. On the other hand, restrictions on storage space mean the combined length of copper and plastic can be 1,000 m at most. As well, operationally you should always have at least 200 m of copper wire and 100 m of plastic tubing at the factory.

Use Solver to find the numbers of metres of copper wire and of plastic tubing that will minimise the cost.

Solution

The model is:

$$\text{Minimise cost } C(x, y) = 200,000 + 200x + 5.45y$$

$$\text{Subject to constraints } 500 \leq x + y \leq 1,000$$

$$x \geq 200$$

$$y \geq 100$$

Create a workbook in Excel. Enter:

Cell A1: "Cost F" Cell B1: =(200000 + 200*B2+5.45*B3)

Cell A2: "x" Cell B2: 0

Cell A3: "y" Cell B3: 0

Cell A4: "x+y" Cell B4: =(B2+B3)

The values of 0 in B2 and B3 are "starter" values for x and y and you could enter any reasonable values. We need these cells because they contain the variables for which we are optimising F . They also appear in the constraints. We need cell B4 because " $x+y$ " appears in a constraint.

Cost F	200000
x	0
y	0
x+y	0

Click on B1. Now click <Tools> and then Solver. (To see that the Solver Add-in has been activated, click <Tools>, <Add-ins> and then ensure that Solver Add-In has been ticked.)

A box with Solver Parameters will appear. \$B1\$ should be in “Set Target Cell”. Remember that the \$ symbol on either side of a cell name means that the cell is an absolute reference and will not change.

In the next line click “Min” because we want to minimise F . In “By changing cells” enter B2,B3.

Click in “Subject to Constraints” box and then click on <Add>. The “Add Constraint” box pops up. In “Cell Reference” enter B2, in the next box choose inequality \geq and in the final box enter 200. Click <OK>. You have entered constraint $x \geq 200$. Click <Add>.

The “Add Constraint” box pops up. In “Cell Reference” enter B3, in the next box choose inequality \geq and in the final box enter 100. Click <OK>. You have entered constraint $y \geq 100$. Click <Add>.

The “Add Constraint” box pops up. In “Cell Reference” enter B4, in the next box choose inequality \leq and in the final box enter 1000. Click <OK>. You have entered constraint $x + y \leq 1000$. Click <Add>.

The “Add Constraint” box pops up. In “Cell Reference” enter B4, in the next box choose inequality \geq and in the final box enter 500. Click <OK>. You have entered constraint $x + y \geq 500$.

Now click <Solve>. The following box appears.

Cost F	241635
x	200
y	300
x+y	500

The cost is minimised by keeping 200 m of copper and 300 m of plastic for manufacturing. The minimum cost is 241,635 rupees.



The Solver tool for non-linear programming

Activity 4.9



Activity Solve an ILP problem

What will you do?

- Use Excel to solve this ILP problem.
Maximise $(2x + 3y - 5z)$ subject to:
 $x + y \leq 12, x + y + z \leq 20, x \leq 10, x, y$ and z positive integers.
- In the portfolio problem above, why can you not use these starter values: a: $x_1 = 1, x_2 = 1$ and $x_3 = 0$?

Change the co-variances to positive values:

$$\sigma_{x_1, x_2} = 0.025, \sigma_{x_1, x_3} = 0.044 \text{ and } \sigma_{x_2, x_3} = 0.0542$$

and repeat the optimisation problem.

Discuss your findings.

Solutions

- Do as for LP, with additional constraints for positivity and integer values.
Click <Add> in the “Add Constraint” box. In “Cell Reference” enter B2, and in the next box choose “int” and click <OK>. You have entered constraint “ x is an integer”. Click <Add> again and in “Cell Reference” enter B3, in the next box choose “int” and click <OK>. You have entered constraint “ y is an integer”. And so on. The result is:

F	36
x	0
y	12
z	0
x+y	12
x+y+z	12

- These seed values do not satisfy constraint $x_1 + x_2 + x_3 = 1$.

We consider a portfolio P of three shares, X1, X2 and X3. It is known that their average annual rates of return over the last five years were:

$$E(R_{X1}) = 7.6\%, E(R_{X2}) = 14.4\% \text{ and } E(R_{X3}) = 16.9\%$$

Their annual volatilities are estimated as:

$$\sigma_{X1} = 25.1\%, \sigma_{X2} = 45.25\% \text{ and } \sigma_{X3} = 49.2\%$$

The new co-variances are:

$$\sigma_{X1, X2} = 0.025, \sigma_{X1, X3} = 0.044 \text{ and } \sigma_{X2, X3} = 0.0542$$

The problem is to determine the proportions x_1 , x_2 and x_3 of the investor's money that should be put in shares X1, X2 and X3, respectively. The investor wants to minimise her risk while requiring a rate of return of at least 12% on portfolio P. No short-selling is allowed.

The output is:

E(RP)	0.12
vol(RP)	0.267238
x1	0.448883
x2	0.290155
x3	0.260962
Sum(xi)	1

Compare this to the output for the original problem with the negative co-variances:

E(RP)	0.12
vol(RP)	0.201919
x1	0.434607
x2	0.343263
x3	0.22213
Sum(xi)	1

The volatility for the portfolio with positive co-variances only is larger than for the original problem with some negative co-variances. This means more risk.



**Case study:
Employee work
schedules**

A factory operates seven days a week, 24 hours a day, with workers doing shifts. On certain days, more workers are needed to finish off and load products for shipping. The manager estimates the minimum numbers of workers needed on different days are:

Mon	Tue	Wed	Thu	Fri	Sat	Sun
18	12	20	12	22	20	8

On Mondays at least 18 workers are needed; on Tuesdays at least 12 workers, and so on. The shifts are 24 hours long but designed to let workers rest in between. Certain shifts are better paid because they extend late at night or are on Saturdays. There are four kinds of shift, described as:

Shift type	Days on duty	Weekly wage
1	Monday, Wednesday, Saturday	USD 560
2	Tuesday, Friday, Sunday	USD 570
3	Monday, Thursday, Saturday	USD 550
4	Wednesday, Friday	USD 300



Problem: The manager wants to minimise wage expenses.

How many workers should the manager assign to each shift so that the minimum number of workers will be available each day as required?

Hint: Assign variables as follows:

x_1 = number of workers assigned to Shift 1, x_2 = number of workers assigned to Shift 2, x_3 = number of workers assigned to Shift 3 and x_4 = number of workers assigned to Shift 4.

Formulate the object function as the total weekly wage expense. Then formulate the seven constraints and use Solver in Excel. Remember, this is an ILP problem. To help you, here is one constraint: since you must have at least 18 workers on Mondays, you must have $x_1 + x_3 \geq 18$.

Solutions

Total weekly wages function:

$$F(x_1, x_2, x_3) = 560 x_1 + 570 x_2 + 550 x_3 + 300 x_4$$

Constraints: $x_1 + x_3 \geq 18$, $x_2 \geq 12$, $x_1 + x_4 \geq 20$, $x_3 \geq 12$, $x_2 + x_4 \geq 22$,
 $x_1 + x_3 \geq 20$, $x_2 \geq 8$

$$x_1, x_2, x_3, x_4 \geq 0$$

x_1, x_2, x_3, x_4 are integers.

The constraints can be reduced: for example, the two constraints $x_1 + x_3 \geq 18$ and $x_1 + x_3 \geq 20$ reduce to one constraint: $x_1 + x_3 \geq 20$.

Use Solver. The output is:

Objective function F	21520
x1	8
x2	12
x3	12
x4	12
x1+x3	20
x1+x4	20
x2+x4	24

This gives the number of workers per shift as 8, 12, 12, 12. The minimum total wage cost per week is USD 21,520. (The question of whether cost is the only determinant for assigning workers to shifts is another matter.)

Unit 16

Forecasting: Time series and linear regression

Upon completion of this unit students will be able to:



Outcomes

- **describe** different methods for forecasting in management;
- **calculate** trends with moving averages;
- **apply** forecasting with time series;
- **apply** basic linear regression in forecasting; and
- **explain** the limitations of forecasting methods.

Moving averages and exponential smoothing

Activity 4.12



Activity
Forecast

What will you do?

1. Consider the table of observed values y_t for six monthly time steps t .

t	1	2	3	4	5	6
y_t	9.4	10	11.02	10.8	11	11.3

- a) Use the simple averages method to find a forecast for F_7 .
 - b) Use the two-month moving averages method to find a forecast for F_7 .
 - c) Use the four-month moving averages method to find a forecast for F_7 .
 - d) Apply exponential smoothing with $\alpha = 0.2$ to find a forecast for F_7 . Let $F_1 = 9$.
2. Consider the table of observed and forecast values.

t	1	2		4
y_t	2.1	2.3	2.6	2.4
F_t	2	2.4	2.5	2.6

Determine the ME, MAD and MSE.



Solutions

$$1. \text{ a) } F_7 = \frac{\sum_{t=1}^6 y_t}{6} = 10.59$$

$$\text{b) } F_7 = \frac{\sum_{t=5}^6 y_t}{2} = 11.15$$

$$\text{c) } F_7 = \frac{\sum_{t=3}^6 y_t}{4} = 11.03$$

$$\begin{aligned} \text{d) } F_7 &= \sum_{k=1}^6 0.2(0.8)^{k-1} y_{7-k} + (0.8)^6 (9) \text{ (with } \alpha = 0.2) \\ &= 0.2(11.3) + 0.2(0.8)(11) + 0.2(0.8)^2(10.8) + 0.2(0.8)^3(11.02) \\ &\quad + 0.2(0.8)^4(10) + 0.2(0.8)^5(9.4) + (0.8)^6(9) \\ &= 7.966 + 2.359 \\ &= 10.325 \end{aligned}$$

$$2. \text{ ME} = \frac{\sum_{t=1}^4 e_t}{4} \text{ where } e_t = y_t - F_t$$

$$\text{MAD} = \frac{\sum_{t=1}^4 |e_t|}{4}$$

$$\text{MSE} = \frac{\sum_{t=1}^4 (e_t)^2}{4}$$

Determine errors: $e_1 = 0.1, e_2 = -0.1, e_3 = 0.1, e_4 = -0.2$

$$|e_1| = 0.1, |e_2| = 0.1, |e_3| = 0.1, |e_4| = 0.2$$

$$(e_1)^2 = 0.01, (e_2)^2 = 0.01, (e_3)^2 = 0.01, (e_4)^2 = 0.04$$

$$\text{ME} = -0.025, \text{MAD} = 0.125, \text{MSE} = 0.0175$$

Time series models

The method of simple linear regression

Activity 4.14



Activity
Consider sales data

What will you do?

Consider data (t, y_t) , $t = 1, 2, 3$ for sales in your shop. Sales are in thousands of items.

Year (t)	Sales (y_t)
1	10
2	8
3	5

1. Use linear regression to find the equation of the line of best fit for the data in the table.
2. Draw the graph of the regression model for sales.
3. Determine the value of one of the measures of goodness of fit. Discuss its meaning.
4. Forecast the values of y at $t = 4$, and $t = 4.5$
5. After how many months will you have no sales?

Solutions

$$\begin{aligned}
 1. \text{ Slope } m &= \frac{3 \sum t y_t - \sum t \sum y_t}{3 \sum t^2 - (\sum t)^2} \\
 &= \frac{3(10 + 16 + 15) - 6(23)}{3(14) - 6^2} \\
 &= -2.5
 \end{aligned}$$

$$\begin{aligned}
 \text{y-intercept } c &= \bar{y} - m\bar{t} \\
 &= (23/3) - (-2.5)2 \\
 &= 12.67
 \end{aligned}$$

Straight line of best fit: $y = -2.5t + 12.67$

2. The graph is the straight line with y -intercept of 12.67 and negative slope -2.5 , and can be drawn by hand or in Excel. (See also the additional section below Activity 4.14.)

Lecturers should review graphs for correctness.



3. The co-efficient of determination, r^2 , is such a measure.

$$r^2 = \frac{[3\sum ty_t - \sum t \sum y_t]^2}{[3\sum t^2 - (\sum t)^2] \times [3\sum y_t^2 - (\sum y_t)^2]}$$

$$= \frac{225}{228}$$

$$= 0.987$$

This is close to 1: there is a reasonably strong linear relation between x and y values — the line is a good fit.

4. $y(t) = -2.5t + 12.67$ implies $y(4) = 2.67$ and $y(4.5) = 1.42$
5. No sales means $y = 0$. Solve for t : $t = 5.068$ years = 61 months (rounded off)

Linear regression with Excel

Activity 4.15



Activity

Do a regression analysis

What will you do?

- Repeat Activity 4.14 using Excel. Discuss as many regression statistics as you can.
- Do a regression analysis for the following data from five companies and interpret the output.

Bonus of CEO (millions of CAD)	Profit of company (millions of CAD)
1.40	110.890
2.00	148.567
4.58	200.55
6.45	185.86
10.1	260.380

Solutions

1. We show part of the output table:

SUMMARY OUTPUT

<i>Regression statistics</i>	
Multiple R	0.993399
R square	0.986842
Adjusted R square	0.973684
Standard error	0.408248
Observations	3

ANOVA			
	<i>df</i>	<i>SS</i>	<i>MS</i>
Regression	1	12.5	12.5
Residual	1	0.1666667	0.166667
Total	2	12.6666667	

	<i>Co-efficients</i>	<i>Standard error</i>	<i>t Stat</i>
Intercept	12.66667	0.62360956	20.31185
X Variable 1	-2.5	0.28867513	-8.66025

R and R square are close to 1, indicating good fit of linear relation.

Follow the Course Manual for further discussion.

2. Use the Bonus column as the Y range (dependent variable).

<i>Regression statistics</i>	
Multiple R	0.944596
R square	0.892261
Adjusted R square	0.856349
Standard error	1.342756
Observations	5

ANOVA			
	<i>df</i>	<i>SS</i>	<i>MS</i>
Regression	1	44.795736	44.79574
Residual	3	5.4089836	1.802995
Total	4	50.20472	

	<i>Co-efficients</i>	<i>Standard error</i>	<i>t Stat</i>
Intercept	-5.86951	2.2436591	-2.61604
X Variable 1	0.059451	0.0119272	4.984495



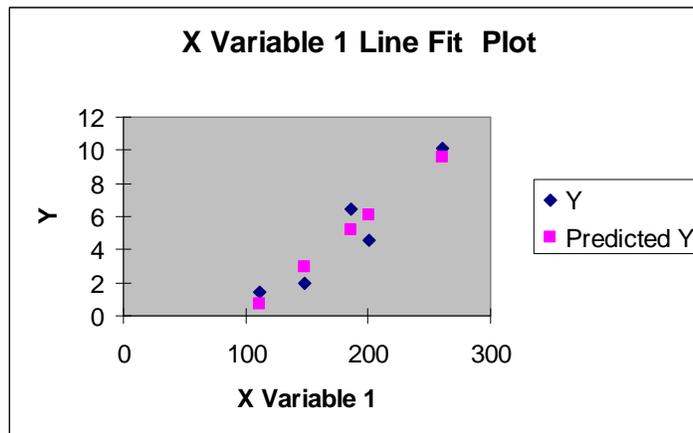
Note: Use the Bonus column as the Y range (dependent variable).

Use Profits as the X range (independent variable).

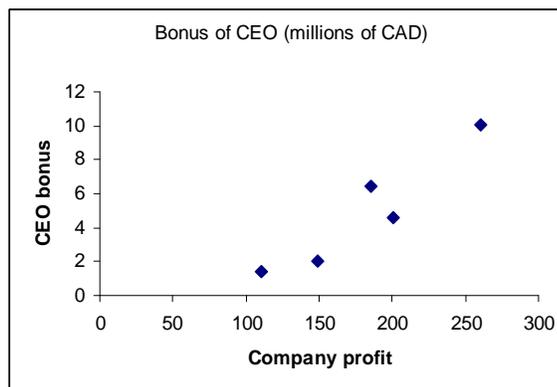
We note that the slope is 0.06 and the intercept at -5.9 . The difference in scale between x and y causes the small slope. The value of R^2 is 0.89, which still indicates a reasonably good fit.

Graphs for linear regression (not in Course Manual)

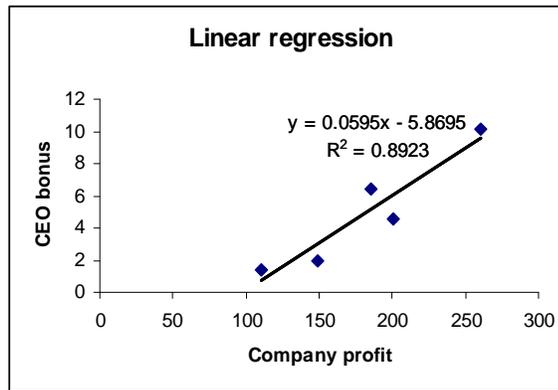
Instead of using Chart Wizard to draw the straight line from the equation, you can click the box “Line Fit Plots” in the Regression dialogue box. This produces the following, whose scale makes for a better graph.



Alternatively, use Chart Wizard (XY Scatter) to plot the separate data points.



Then right-click in the chart area. A new drop-down menu <Chart> will appear in the top toolbar next to <Tools>. Click on <Chart> and then click <Add Trendline>. A box will open. Choose “Linear” under the Type tab and then click the Options tab. Tick the two boxes “Display equation on Chart” and “Display R squared value on chart”. Click <OK>. The graph will look like this:





Assessment



Assessment — Module 4

- Draw a binomial tree to represent the Markov price process $\{S_t : t = 0, 1, 2, 3\}$ for a share that starts with initial price $S_0 = 10$ and then either increases by 10% per time-step t or decreases by 10% per time-step t . The probability of an increase is 0.5 (50%).
 - Is the tree recombining?
 - Calculate $E[S_1]$ as well as $E[S_2]$ and compare the values with S_0 .
 - Repeat the calculations in b) above for the case where the probability of an increase is 0.7 (70%). What conclusions can you draw?
- A fast-food company wants to study the arrival of customers at the counter. It has estimated the probabilities of certain numbers of customers arriving at the counter during any two-minute period. The data is:

Number of customers	0	1	2	3
Probability	0.1	0.45	0.25	0.2

The company wants to simulate the number of customers arriving over 10 two-minute periods by generating strings of random numbers. The aim is to try to find out how many customers may arrive in a 20-minute period. This is important because it cannot cope if more than 18 people arrive during a typical 20-minute period.

- Generate a column of 10 random numbers using function `RANDBETWEEN(0,99)` in Excel. Discuss your method.
 - Determine how many people arrived in a 20-minute period in this simulation.
 - Repeat the process 20 times (a sample of 20 runs) and determine the average number of people that will arrive in a 20-minute period. Discuss your results and compare with those of fellow students.
- A car manufacturer makes USD 1,000 profit on each model A car and USD 1,350 profit on each model B car. The assembly team can assemble a model A car in 30 minutes and a model B car in 45 minutes. A working day consists of eight hours.
 - Write down an expression for the per-day profit function P . Use a spreadsheet package to represent this by a plane in 3-D space.
 - Write down the daily time constraint. Represent the feasible region in a spreadsheet using Chart Wizard.
 - Give a graphical representation of the feasible region showing the profit function contour line that gives maximum daily profit.

- d) Find the optimal solution graphically.
- e) Verify the solution with the corner-point method.
- f) Use the Solver tool to find the optimal solution.
4. A profit function has the form $P(Q) = Q^3 - 2R^2 + 4QR - 3R$. The following constraints apply: $Q + R \geq 100$, $Q \leq 150$ and $R \leq 300$
- a) Find the values of Q and R that maximise the profit. Calculate the profit in this case.
- b) If there were no constraints, what would the maximum profit be? Use both analytical (calculus) methods and the Solver tool.
5. Your country's Internal Revenue Service (IRS) asks you to help develop a model to detect whether companies inflate the figures of their contributions to charity in order to gain greater deductions on their tax returns. You have the following data for eight companies whose tax returns have been audited and accepted as truthful.
- a) Draw a scatter graph for the data and decide whether a linear relationship is likely.
- b) Do a regression analysis and derive a linear equation for the relation between gross income (G) and charitable contribution (C).
- c) Interpret the value of R^2 and evaluate your model.
- d) Write a report to explain to the IRS how it could use your model to detect instances where companies claim unusually high contributions to charity.

Gross annual income (millions)	Contribution to charity
66.7	100,000
72.84	115,000
95.1	135,800
110.25	140,000
129.56	165,500
166.4	175,000
182.3	200,000
200.75	225,000



6. Consider the table of observed values y_t for six annual time steps t .

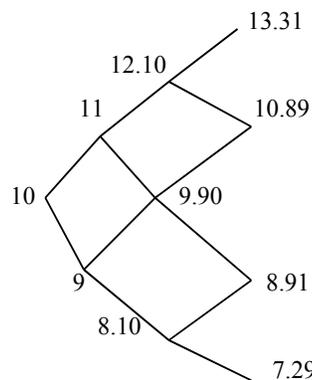
t	1	2	3	4	5	6
y_t	352.5	340	341.5	329.8	320	311.3

- a) Use the simple averages method to find a forecast for F_7 .
 - b) Use the 2-year moving averages method to find a forecast for F_7 .
 - c) Use the 4-year moving averages method to find a forecast for F_7 .
 - d) Apply exponential smoothing with $\alpha = 0.1$ to find a forecast for F_7 . Let $F_1 = 350$.
7. Consider the table of observed values y_t and forecast values $F1_t$ and $F2_t$. The forecasts F1 were obtained by Method 1 and the forecasts F2 were obtained by Method 2. Calculate some measures of error and interpret them. Which method is best?

t	1	2	3
y_t	30	34	38
$F1_t$	28	33	35
$F2_t$	28	35	40

Solutions

1.



$t=0$ $t=1$ $t=2$ $t=3$
 S_0 S_1 S_2 S_3

- a) The tree is recombining.
- b) $E[S_1] = \frac{1}{2}(11 + 9)$
 $= 10$
 $= S_0$

$$\begin{aligned}
 E[S_2] &= \frac{1}{4}(12.10) + \frac{1}{2}(9.90) + \frac{1}{4}(8.10) \\
 &= 10 \\
 &= S_0
 \end{aligned}$$

We have a Markov process — in fact, a martingale process.

c)
$$\begin{aligned}
 E[S_1] &= 0.7(11) + 0.3(9) \\
 &= 10.40 \\
 &> S_0
 \end{aligned}$$

$$\begin{aligned}
 E[S_2] &= \frac{1}{4}(12.10) + \frac{1}{2}(9.90) + \frac{1}{4}(8.10) \\
 &= 10.82 \\
 &> S_0
 \end{aligned}$$

We have a Markov process, but not a martingale. The share price seems to be “drifting” upwards.

2. a) `RANDBETWEEN(0,99)` generates random numbers from 0 to 99. Each of the 100 possible numbers has a 1% probability of being generated.

The idea is that a number between 0 and 9 represents 0 people arriving in a 2-minute interval. Reason: there are 10 numbers from 0 to 9, which represent a 10% probability out of the 100 possible numbers from 0 to 99 that can be generated. Similarly, there are 45 numbers from 10 to 54, which represents a 45% probability out of the 100 possible numbers between 0 and 99. So numbers from 10 to 54 represent one person arriving in a 2-minute interval. Numbers from 55 to 79 represent two people arriving in a two-minute interval and numbers 80 to 99 represent three people arriving in a two-minute interval.

A column (or row) of 10 numbers gives 10 two-minute intervals – that is, 20 minutes. Translate the random numbers to numbers of customers and find the sum for this column to simulate how many customers may arrive in a 20-minute period.

- b) Example of random numbers:

35 60 90 91 68 24 0 84 88 65

Translated numbers with sum:

										SUM
1	2	3	3	2	1	0	3	3	2	20

- c) You can generate a block of size 20×10 . Remember to translate each column (or row) of 10 to represent customers. Sum each column and find the average of the 20 numbers.

3. a) Let x denote number of model A cars assembled per day. Let y denote number of model B cars assembled per day. Profit $P(x, y) = 1,000x + 1,350y$

- b) Constraints: $\frac{1}{2}x + \frac{3}{4}y \leq 8$, $x, y \geq 0$. We may treat this as a LP problem and allow fractions as answers, since we do not need to assemble a complete car in a day. Otherwise, if you want whole cars completed in a day, you can see this as an ILP problem.



The feasible region is a triangle with corner points (0, 0), (0, 10.667) and (16, 0).

c) and d)

Draw the triangle as described in b). The slope of the hypotenuse is -0.67 ; the slope of the profit contour lines is -0.74 , which shows that contour lines have a steeper slope than the hypotenuse. This means the maximum profit contour line is the one touching point (16, 0).

Lecturers should review graphs for correctness.

e) Corner-point method: $P(0, 0) = 0$

$$P(0, 10.667) = 1,350(10.667) = 14,400.45$$

$$P(16, 0) = 1,000(16) = 16,000$$

The optimal solution for maximum profit is to assemble 16 model A cars per day and no model B cars.

Discussion: This is probably too simplistic and additional constraints may be necessary. For example, there may be demand for at least three model B cars per day. The added constraint $y \geq 3$ leads to a triangular feasible region with the corner points (0, 3), (0, 10.667) and (11.5, 3). The values of P are:

$$P(0, 3) = 4,050, P(0, 10.667) = 1,350(10.667) = 14,400.45 \text{ and } P(11.5, 3) = 15,550$$

The new optimal solution is 11.5 of model A and three of model B.

f) Solver used for the adapted problem with constraint $y \geq 3$:

P	15550
x	11.5
y	3
$\frac{1}{2}x + \frac{3}{4}y$	8

Solver used with integer constraints on x, y :

P	15400
x	10
y	4
$\frac{1}{2}x + \frac{3}{4}y$	8

1. a) Solver:

Profit	3419551
Q	150
R	149.25
q+r	299.25

b) No constraints: Solver cannot find a solution with starter values (0, 0). Remember the note in the Course Manual: *When you work with non-linear problems, it is best to not use starting values of zero (0) for all the relevant variables. Solver does not always like starting from 0.*

Try $(-1, -1)$ as starting values.

Profit	7.189605
q	-1.86852
r	-2.61852
q+r	-4.48703

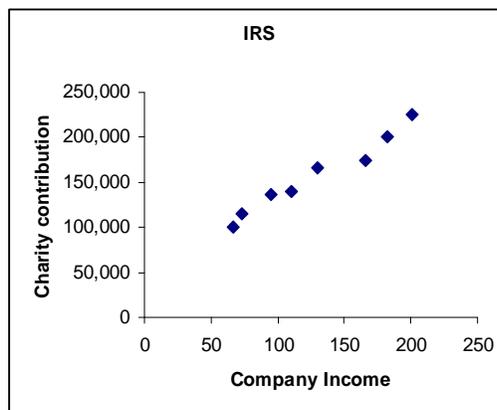
Analytically: Find the derivatives of P for Q and to R . Set both equal to zero and solve. You should find:
 $3Q^2 + 4R = 0$ and $-4R + 4Q - 3 = 0$

The first equation gives $4R = -3Q^2$. Substitute this into the second equation and solve the quadratic for Q . We obtain $Q = -1.867$ and

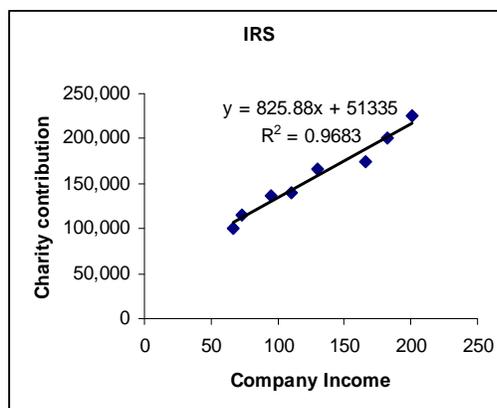
$Q = 0.54$. The second derivatives of P with respect to Q and R are $6Q$ and -4 , respectively. Both should be negative for a maximum P . This is only possible for $Q = -1.867$.

$Q = 0.54$ leads to second derivatives of mixed sign and this means we have a saddle point at $Q = 0.54$.

5. a)



b) Click on the graph and use the Chart drop-down menu as explained previously. Otherwise, use <Tools> and <Data Analysis>.





- c) The value of R^2 indicates a high linear relationship. For every 10 million increase in income, companies increase their contribution to charity by 8,258.80.
- d) If a company claims it contributed 300,000 to charity, its gross income should be about 301 million. The IRS can now check what the company submitted as gross income.
6. a) $F_7 = 332.52$
- b) $F_7 = \frac{\sum_{t=5}^6 y_t}{2} = 315.65$
- c) $F_7 = \frac{\sum_{t=3}^6 y_t}{4} = 325.65$
- d) $F_7 = \sum_{k=1}^6 0.1(0.9)^{k-1} y_{7-k} + (0.9)^6 (350)$ (with $\alpha = 0.1$)
 $= 343.24$

The trend over the past four years is a decreasing one. One should therefore give the first values of y less weight. The four- or two-year moving averages should give better forecasts.

7. The data set is actually too small for a proper error analysis. However, calculations can get messy and we therefore use small data sets.

$$ME = \frac{\sum_{t=1}^3 e_t}{3} \text{ where } e_t = y_t - F_t$$

$$MAD = \frac{\sum_{t=1}^3 |e_t|}{3}$$

$$MSE = \frac{\sum_{t=1}^3 (e_t)^2}{3}$$

Forecasting method 1:

Determine errors: $e_1 = 2, e_2 = 1, e_3 = 3$

$ME = 2, MAD = 2, MSE = 4.67$

Forecasting method 2:

Determine errors: $e_1 = 2, e_2 = -1, e_3 = -2$

$ME = -0.33, MAD = 1.67, MSE = 3$

Interpretation: For Method 1, $ME > 0$ shows that the forecasts are, on average, too low. For Method 2, $ME < 0$ shows that the forecasts are, on average, too high.

MAD gives the average distance of each forecast from the actual observation at that point.

MSE is useful in that it emphasises large errors and minimises small ones. The large error in Method 2 for forecast 3 has been amplified to give it a bigger MSE than Method 1. Method 2 is better overall.