Appendix

Step-by-Step: Quantitative analysis of cost functions in hotels and restaurants at the property level

Introduction

This appendix explains the steps in using quantitative methods for analysing and evaluating cost function behaviour patterns to assist in practical routine, day-to-day, decision making.

- Cost behaviour is relevant to managers and management accountants as it offers insights into the operating profile of business undertakings and in doing so assists in the decision making process. How costs behave are the building blocks of *cost structure* (also called *operating leverage*) and form the basis for determining the business orientation of an enterprise.
- Kotas (1973) first identified the concept of *business orientation*. He drew attention to business orientation as a function of cost structure, introducing the terms *cost orientation* and *market orientation* where low fixed cost undertakings are product (cost) oriented and high fixed cost undertakings are market (revenue) oriented.
- Managers should recognise the influence cost structure has on the orientation of their businesses and engage in understanding how costs behave. Knowledge of a property's business orientation provides an important operating perspective in straightforward financial terms and guides the approach to strategy formulation (*see Chapter 7: Cost structure, pp. 104-118*).
- The way direct department and undistributed operating costs behave affects routine decision making. Day-to-day decisions relating to pricing strategies, competitive bidding, new business development and transaction negotiation are all influenced by the way costs react to the decisions being considered.
- As economic conditions change and business strategy adapts, financial controllers can help managers understand evolving cost relationships and estimate new fixed and variable costs. Whilst managers are acutely aware of the costs involved in operating their properties, they are often less sure of how the various costs behave in day-to-day decisions.
- There are useful approaches to determining how costs behave (*see Chapter 5: Practical cost behaviour for decisions, pp. 68-80*), but the emphasis here is on quantitative analysis; using pragmatic statistical methods to analyse historical costs – *learning from the past to predict the future.*

- Quantitative analysis should never replace other methods of analysing costs, but act as a cross check against methods based on judgement and experience. All cost analysis methods rely on managers' working knowledge of their own business – *this remains crucial to informed cost decisions.*

Steps for quantitative analysis of cost functions (behaviour)

The steps for analysing, estimating and predicting cost functions using quantitative techniques of past cost relationships are as follows:

- Selecting costs for analysis (dependent variables '*y*')
- Identifying cost drivers (independent variables '*x*')
- Collecting data on cost items and cost drivers
- Plotting the data
- Estimating cost functions
- Determining the quality of data fit
- Testing the significance of cost drivers
- Evaluating cost drivers of estimated cost functions

The steps presented below explain the process and method of quantitative analysis of cost behaviour. The data analysis and graphs are prepared with the assistance of Microsoft Excel.

Step 1: Selecting costs for analysis

Clearly, once we begin applying quantitative techniques to analyse cost behaviour, it is natural to assume all operating costs should be included in the same process. In principle this makes sense, but the early stages of developing and applying cost estimation models in practice is a detailed and time consuming activity.

In attempting the process of analysing operating costs for the first time in the live business situation, it is worth identifying and prioritising the costs to be initially analysed; enabling familiarity with the procedures and confidence with the overall process.

Consider the following in selecting cost items for initial analysis:

- Referring to the profit and loss statement, the three main behavioural cost groups are fixed costs, semi-variable costs and variable costs. Of these, the semi-variable costs are generally regarded as the most ambiguous and challenging cost items to analyse into fixed and variable elements.

- Due to the nature of service businesses, semi-variable cost items are commonly found to represent a high proportion of operating expenses relative to sales revenue; and this is particularly so with costs such as payroll and related expenses in the hotel and restaurant sectors.
- In hospitality properties, semi-variable costs include department direct payroll expenses and other direct expenses; and undistributed operating expenses that include payroll and other expenses associated with each of the main cost (service) centres, including administration and general; property and maintenance; and energy and utilities.

As the above points imply, the nature and prominence of semi-variable costs tend to highlight their presence as relevant cost items for initial analysis into fixed and variable elements. Other expense items are generally perceived as being *wholly* fixed or variable, such as discretionary fixed costs, e.g. marketing expenses, and variable costs, e.g. F&B cost of sales. These are examples of potential expense items which, in due course, may also warrant analysis in order to establish their particular behavioural patterns in specific properties.

Note: It is not inevitable that all operating expenses should immediately be analysed by quantitative techniques. Where well-grounded non-numerical methods are used to determine how costs behave, selected expense items can be analysed numerically for comparison with the technical (knowledge and experience judgements) estimates made by financial controllers and managers.

Step 2: Identifying cost drivers

A cost driver is a variable, such as the level (occupancy %) or volume (rooms let) of business activity, which prompts costs to change over time; in effect, a cause-and-effect relationship present between a change in the level of activity and a change in the level of a cost. For example, if hotel room guest supplies cost change with the number of rooms occupied, the number of rooms occupied is a cost driver.

In statistical cost estimation terms, the cost driver represents the *independent variable* signified as 'x' and the cost item represents the *dependent variable* signified by 'y'. In determining cost behaviour x is used to estimate and predict y.

Causal relationships

We may sometimes assume that intuitive feelings and strong numerical associations between two variables indicate a causal relationship, with variations in one causing variations in the other. However, personal intuition and high correlations (explained later) between cost drivers and cost variations (behaviour) cannot, in themselves, imply causal links between variables. Credible assertions of interdependence between costs and cost driver variables should be based upon convincing evidence of a plausible relationship.

Economic plausibility

A cost driver should be economically – and operationally – plausible in relation to the cost function being estimated. Horngren *et al* (2009) suggests the cause-and-effect relationship might arise as a result of:

- **A physical relationship between the level of activity and costs.** For example, where the number of hotel rooms occupied is used as the activity measure that affects laundry costs, letting more rooms prompts additional laundering, which results in higher total laundry costs.
- **A contractual arrangement.** For example, where hotel room servicing is outsourced to an external provider at a contracted price per room, only the number of rooms serviced affects the total cost incurred.
- **Knowledge of operations.** For example, where hotel capacity utilisation (occupancy) is used as the measure that affects energy consumption costs, the number of guests (sleepers) will normally provide a more accurate measure than the number of rooms occupied.

Note: Unless the relationship between a chosen cost driver and the particular cost itself makes economic and/or operational sense managers will not be confident to incorporate the variables in their routine, day-to-day, business decisions – *this point cannot be over-emphasised*.

Relevant range

The *relevant range* is an important assumption that relates to the normal band of activity level in which a business operates over a given period and/or where individual cost functions are *assumed* to remain in a constant state (in total or per unit of activity).

- For example, fixed costs such as department manager salaries remain constant in total only over a given range of activity (at which the business is expected to operate) and only for a given period of time (usually a budget period or financial year).
- Similarly, the relevant range assumption also applies to variable costs. Outside the relevant range individual variable costs per unit, such as F&B cost of sales may not change in proportion to restaurant sales volume. For example, above a given volume of activity cost of sales may increase at a lower rate due to bulk discounts received on the larger quantities.
- With regard to semi-variable costs, such as energy and utilities, the relevant range assumption may only affect the variable cost element. For example, the standing charge for water service to a property will normally represent the fixed cost element and the water charge per gallon/litre consumed will represent the variable element; outside the water consumption price band (relevant range) the variable cost element may increase to a higher rate per gallon/litre. In the case of department payroll, a sustained change in the level of business may prompt changes in the fixed and/or variable elements of the payroll cost.

Note: The relevant range assumption provides a useful practical context for interpreting the economics concept of marginal analysis in everyday management accounting terms. In this context, *marginal cost* can be explained as 'the average variable cost within the relevant range'.

Step 3: Collecting data on cost items and cost drivers

In order to avoid analytical pitfalls, an important stage in the use of quantitative techniques in analysing cost behaviour is the collection and processing of data:

- **Observation time period.** Unit of time for observations should enable the accounting procedures to match activity and cost; with accounting records for accumulating data kept on an accrual basis. Monthly observation periods are normally regarded as reasonable for hospitality studies and manageable for most management accounting systems, though weekly observation periods would facilitate larger samples.
- **Time span of analysis.** Number of (monthly) observations in a sample should be large enough to be representative of data in terms of encompassing a wide range of business variability. The data should be as recent as possible and reflect – so far as is possible – a constant state of past and future operations (continuity) in terms of products and services offered and working practices.
- **Measures of activity.** Methods of developing cost drivers should be carried out with care in order to identify plausible relationships between costs and activities and avoid spurious results. Various measures are available, such as business mix, events held, volumes of rooms let and covers sold; plus broader capacity measures, including rooms occupied, sleeper-nights, headcount, meals/dishes served and labour hours worked.
- **Extreme observation values.** Unusual, erroneous and exceptional observations may be present in the analysis. These can include numerical entry (recording) errors; non-accrual of purchase and expense payments; others may be exceptional 'one-off' cost items. Each untypical observation should be investigated and adjusted or omitted before estimating a cost relationship.
- **Inflation.** Inflationary price increases should be removed from cost data using an appropriate price index on a month by month basis. This is important if inflation is severe and/or where the analysis involves a large sample (normally regarded as over 30 observations) and data is available only monthly, implying the use of a three-year time span.

Step 4: Plotting the data

Let's take an example of an independent restaurant business. An initial attempt is being made to understand the cost behaviour of direct payroll expense (relating to the preparation and service of F&B in the restaurant).

After an initial discussion, management believes direct payroll is influenced by *number of covers sold* (cost driver) per month. It is economically plausible that number of covers sold would help explain the variations in direct payroll costs in the restaurant.

The data presented in Table 1 relate to monthly *number of covers sold* (cost driver) and monthly *direct payroll cost* of a restaurant property for a one year period. This forms the basis to analyse the restaurant payroll cost function related to covers sold.

However, prior to performing a quantitative analysis it is beneficial to obtain a visual impression of the data plot on a scatter diagram from Excel, shown in Figure 1, in order to:

- Determine the general relationship which exists between restaurant direct payroll cost and number of covers sold variables
- Identify any abnormal or extreme observations which may require further investigation and subsequent adjustment or omission.
- Provide an indication of the cost function (behaviour) in terms of linearity; and the extent of the relevant range.

Figure 1: Scatter diagram of restaurant monthly direct payroll cost against number of covers sold

The data plot in Figure 1 indicates there are no apparent unusual or extreme observations present and that a positive linear trend exists between the restaurant payroll cost and covers sold; allowing a linear analysis to be carried out. 'In most analysis, a straight line is adequate because it is a reasonable approximation of cost behaviour within the relevant range' Matz and Usry (1980).

Step 5: Estimating cost functions

A cost function is a quantitative expression of how a cost changes (behaves) in relation to changes in levels of activity. Estimating a cost function quantitatively is carried out using regression analysis techniques.

Linear regression analysis

Linear regression is a statistical technique for specifying the relationship between variables:

- The technique uses a formal model (equation) to measure the *average* amount of change in a given dependent variable (cost) that is associated with changes in one or more independent variables (cost drivers), such as rooms occupied, sleeper-nights and covers sold.
- Where only one independent variable is included in an equation the technique is termed *simple linear regression analysis*. Where two or more independent variables are included in the equation the technique is known as *multiple linear regression analysis*.
- Linear regression analysis uses a sample of past costs to estimate how the population of costs behave. The technique used is termed the *method of least squares* which determines the *line of best fit* for a given set of data.
- The least-squares line is so called because the sum of the squares of the vertical distances (known as *residuals*) from the regression line to the actual data points is less than for any other line.

The least-squares line is represented by the following equation:

$$y_C = a + bx$$

where y_c = estimated total cost (dependent variable); a = fixed cost (constant); b = average variable cost per unit of activity (slope coefficient); and x = the cost driver (independent variable e.g. rooms occupied).

The purpose of the analysis is to determine the a and b values of a given regression equation; obtained by the simultaneous solution of two normal equations:

$$y = na + b(x)$$

$$xy = a(x) + b(x^2)$$

where y = observed values (actual data points); n = number of observations in the sample; Σx = sum of the observations of the independent variable(s); Σy = sum of the observations of the dependent variable(s); Σx^2 = sum of the squares of the x observations; Σxy = sum of the product of each pair of observations.

Simple linear regression: restaurant illustration

Drawing on the appropriate columns in Table 1, $\Sigma x = 178$, $\Sigma y = 798$, $\Sigma x^2 = 2{,}902$, $\Sigma xy = 12.492$ and $n = 12$, substitution of the data into the two normal equations provides the following:

(1)	$798 = 12a + b178$
(2)	$12{,}492 = 178a + b2{,}902$
(1) × 14.8333 [178a ÷ 12a]	$11{,}837 = 178a + b2{,}640$
(2) – (1)	$655 = 0 + b262$
∴	$b = 655 \div 262 = 2.50$

Note: When calculated in Excel, $b = 2.5032$ compared to $b = 2.50$, when calculated from Table 1 analysis (differences due to rounding in Table 1 analysis).

(1) $798 = 12a + 2.5032(178)$

$798 = 12a + 445.5696$

∴ $a = 352.4304 \div 12 = 29{,}369$

The solution gives $a = 29{,}369$ and $b = 2.50$, but as an aid to computation the two normal equations can be expressed as follows:

$$a = \frac{(y)(x^2) - (x)(xy)}{n(x^2) - (x^2)}$$

$$b = \frac{n(xy) - (x)(y)}{n(x^2) - (x^2)}$$

Table 1: Restaurant monthly covers sold and direct payroll expense data for regression analysis computations.

	Number of covers sold 000s	Direct payroll expense £000s	000,000s	000,000s	-	000,000s	000,000s	000,000s
Month	x	y	x^2	xy	y_c	$(y - \bar{y})^2$	$(y - y_c)^2$	$(x - \bar{x})^2$
1	22	88	484	1,936	84,439	462.250	12.667	51.366
2	7	56	49	392	46,892	110.25	82.961	61.356
3	20	80	400	1,600	79,433	182.25	0.321	26.698
4	8	40	64	320	49,395	702.25	88.264	46.690
5	19	84	361	1,596	76,930	306.25	49.986	17.364
6	10	52	100	520	54,401	210.25	5.766	23.358
7	12	68	144	816	59,408	2.25	73.829	8.026
8	19	64	361	1,216	76,930	6.25	167.183	17.364
9	13	60	169	780	61,911	42.25	3.651	3.360
10	17	76	289	1,292	71,924	90.25	16.617	4.696
11	15	56	225	840	66,917	110.25	119.185	0.028
12	16	74	256	1,184	69,420	56.25	20.973	1.362
Total	178	798	2,902	12,492	798,000	2,281.00	641.414	261.667
	Σx	Σy	Σx^2	Σxy	Σy_c	$\Sigma(y - \bar{y})^2$	$\Sigma(y - y_c)^2$	$\Sigma(x - \bar{x})^2$

$x = 14{,}833$ $\bar{y} = £66{,}500$

Note: The final values of the results drawn from this table are presented in full in the text. The y_C column is the predicted value of y for each observed value of x in the regression equation $y_C = 29{,}369 + 2.50x$. For example, for Month 2, $y_C = 29{,}369 + 2.50\ (7{,}000) = 46{,}869$ which agrees closely with the value in the table i.e. 46,869 compared to 46,892 (difference due to rounding £2.5032 to £2.50 in the regression equation).

The illustration data now gives:

$$a = \frac{(798)(2{,}902) - (178)(12{,}492)}{(12(2{,}902) - (31{,}684)} = 29,\ 369$$

$$b = \frac{12(12{,}492) - (178)(178)}{12(2{,}902) - (31{,}684)} = 2.50$$

Entering the computed a and b values into the regression equation gives:

$$y_C = £29{,}369 + £2.50x$$

signifying the fixed cost element £29,369 and variable cost element £2.50 per cover sold of restaurant direct payroll cost. Therefore, y_C represents the estimated average direct payroll cost for any number of covers sold within the relevant range, indicated by the regression line in Figure 2. If 15,000 covers are sold the estimated direct payroll cost would be £29,369 + £2.50 (15,000) = £66,869. Again, £66,869 is close to Month 11= £66,917 in Table 1 (difference due to rounding £2.5032 down to £2.50).

Note: The calculations can be performed on most computer software, such as Excel, or alternatively on electronic calculators which contain statistical functions; performed by entering the pairs of observations using linear regression mode and displaying the totals through the appropriate function keys.

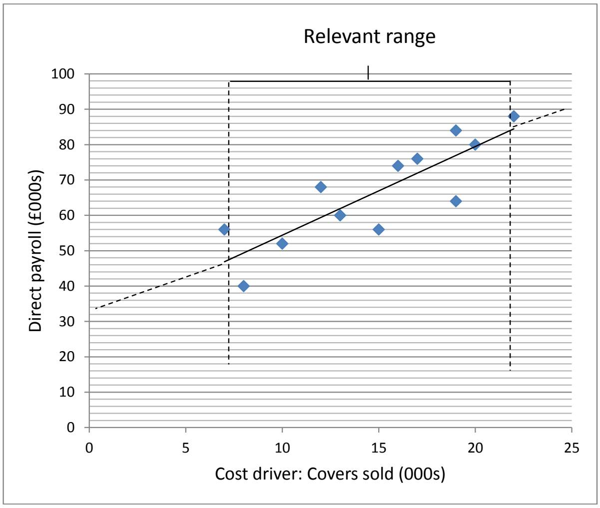

Figure 2: Regression line of best fit for restaurant monthly direct payroll cost and number of covers sold

We have computed a regression analysis which indicates the nature of the relationship between restaurant direct payroll cost (y) and number of covers sold (x) in the data sample. As illustrated above, the use of linear regression allows a line of best fit equation to be computed which may subsequently be used to predict the future level of payroll costs.

Step 6: Determining the quality of data fit

A major concern with using regression is that an equation can be fitted to any sample pairs of observations, regardless of whether there is any measurable association or plausible relationship between the variables. Therefore, having determined the equation the next step is to access the quality of the data fit, known as the *goodness of fit*, which measures how closely the predicted values (y_C) based on the cost driver (x) relate to the actual cost observations (y).

Correlation

Goodness of fit can be measured using correlation analysis techniques. One such measure of correlation is the *coefficient of determination*, denoted by r^2, explained below:

- The coefficient of determination measures the extent to which the dependent variable y (direct payroll cost) is explained by the independent variable x (covers sold), indicated in Figure 3.
- More importantly, r^2 indicates in percentage terms how much of the total variation of the y values can be attributed (explained) to the relationship between the x and y variables and how much can be attributed to chance (unexplained), as depicted in Figure 3.

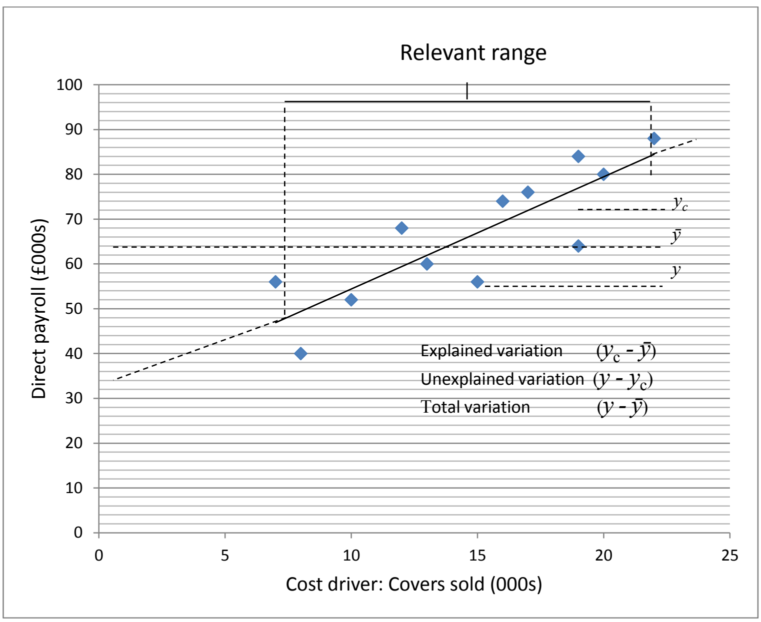

Figure 3: Measure of cost variation about the regression line of best fit

The coefficient of determination (r^2) can be expressed as follows:

$$r^2 = 1 - \frac{(y - y_C)^2}{(y - \bar{y})^2} = 1 - \frac{\textit{Unexplained variation}}{\textit{Total variation}}$$

Using the appropriate data from Table 1, $(y - y_C)^2 = 641.414$ and $(y - \bar{y})^2 = 2{,}281.00$, gives r^2 which indicates the percentage of the variation in y (dependent variable) that can be explained by x (independent variable) which is:

$$r^2 = 1 - \frac{641.414}{2{,}281.00} = 0.7188 \text{ or } 72\%$$

Thus, it can be stated that 72% of restaurant direct payroll cost is explained by number of covers sold and 28% can be attributed to chance variation and the effect of other variables not included in the regression equation. Generally, an r^2 of greater than 0.30 is regarded as acceptable for cost estimation purposes.

Note: As indicated in Figure 3, the total variation in payroll cost from its mean $(y - \bar{y})$ can be analysed into two parts. Firstly, the variation between the regression line and the mean $(y_C - \bar{y})$ which is explained by the given value of x and secondly, the variation between the payroll cost and the regression line $(y - y_C)$ which is not explained by x. Also, notice the regression line always passes through the mean of the data set.

The coefficient of determination (r^2) is not, perhaps, so well known as the coefficient of correlation (r) explained below, but it is far more meaningful measure of cost variation. The value of r^2 cannot, of course, be greater than 1 since it cannot explain a greater proportion of total variation than the whole; nor can it be less than zero since there cannot be less than no variation explained.

The square root of 0.7188 is called the *coefficient of correlation* (r):

$$r^2 = \pm\sqrt{1 - \frac{(y - y_C)^2}{(y - \bar{y})^2}} = \sqrt{1 - \frac{641.414}{2{,}281.00}} = +0.8478 \text{ or } +0.85$$

The range of the coefficient of correlation (r) is from -1 (perfect negative correlation) through 0 (no correlation) to +1 (perfect positive correlation). In the case of a perfect positive fit, the regression line will pass through every observed value of y. In such a case, the sum of the squares of the residuals from the regression line to the data points will be zero and r will equal +1.

Note: the coefficient of correlation (r) may give the impression of a higher degree of association between an independent variable (x) and dependent variable (y) than is applicable. In our example above, 72% of the variance of y is explained by x, $r^2 = 0.7188$, but $r = \pm\sqrt{0.7188} = +0.8478$ or $+0.85$. This occurs because r is a relative measure of the relationship between two variables, whereas r^2 indicates the *proportion* of the total variance that is explained by the independent variable x (cost driver).

Cause-and-effect

As referred to earlier in the cost drivers section, we often assume that high correlations between two variables indicate a causal relationship. Correlation analysis measures the numerical strength of association between pairs of data and does not, in itself, imply any causal link between the data (variables).

Remember, in order to assert an interdependent (mutual) relationship between two variables there should be convincing evidence of economic and/ or operational plausibility.

Standard error of the estimate

Having ascertained the degree of association between the x and y variables the next step is to assess how accurate the regression line is as a basis for prediction. The accuracy, goodness of fit, of the equation line can be measured by the *standard error of the estimate* (S_e), explained below:

- The purpose of the standard error of estimate is to measure how closely the predicted costs (y_c) can be expected to match to the actual costs (y).
- The standard error of the estimate measures the unexplained variation determined by the regression equation line of best fit, as indicated in Figure 3.

The smaller the standard error of the estimate, the better the regression line fits the data. The standard error of the estimate for a population is estimated from a sample of past costs as follows:

$$S_e = \sqrt{\frac{(y - y_c)^2}{\textit{Degrees of freedom}}} = \sqrt{\frac{(y - y_c)^2}{n-2}}$$

where n is the sample size. The denominator, n-2 is called the *degrees of freedom*. Compared with the coefficient of correlation which is a relative measure, the standard error of estimate provides an absolute (numerical) measure, in the case of payroll cost representing £s sterling, as illustrated below

Note: One degree of freedom is lost for each value that has been calculated in the regression equation. In our restaurant illustration, the intercept a together with one slope coefficient b were estimated to establish the regression line, therefore, two degrees of freedom are lost.

Using the data in Table 1 the standard error of the estimate is:

$$S_e = \sqrt{\frac{641.414}{12-2}} = 8.009 \text{ or } £8{,}009$$

If the assumptions which underlie linear regression analysis are satisfied (*linearity, normality, constant variance and independence of residuals* - explained later), then the standard error of the estimate indicates the range of values of the dependent variable (payroll cost) within which there can be some degree of confidence that the true value lies. For example, if 15,000 covers are sold then the predicted payroll cost will be as follows:

$$y_c = £29{,}369 + £2.50\ (15{,}000)$$
$$= £66{,}869$$

With the standard error of the estimate, also sometimes referred to as the range of probable error, of £8,009:

$$Se = £66{,}869 \pm\ £8{,}009\ (1.0)$$

As approximately two thirds of the data points in a normal distribution should fall within one standard error, it is possible to predict that 15,000 covers sold should incur an actual direct payroll cost of between £58,860 and £74,878 with approximately two chances out of three that the confidence interval will contain the true cost; due to 68% (approximately two-thirds) of the data points in a normal distribution fall within a range of ± one standard error.

Statistical theory suggests that for linear regression analysis the data points are *t*-distributed around the regression line and that the distribution becomes normal as the number of observations reaches thirty. In sampling terms, 30 or more observations are categorised as a large sample. The *t*-distribution is presented in the form of a table in Table 2.

Table 2: *t*-distribution table

df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$
1…………………………..	3.078	6.314	12.706	31.821	63.657
2…………………………..	1.886	2.920	4.303	6.965	9.925
3…………………………..	1.638	2.353	3.182	4.541	5.841
4…………………………..	1.533	2.132	2.776	3.747	4.604
5…………………………..	1.476	2.015	2.571	3.365	4.032
6…………………………..	1.440	1.943	2.447	3.143	3.707
7…………………………..	1.415	1.895	2.365	2.998	3.499
8…………………………..	1.397	1.860	2.306	2.896	3.355
9…………………………..	1.383	1.833	2.262	2.821	3.250
10………………………….	1.372	1.812	2.228	2.764	3.196
11………………………….	1.363	1.796	2.201	2.718	3.106
12………………………….	1.356	1.782	2.179	2.681	3.055
13………………………….	1.350	1.771	2.160	2.650	3.012
14………………………….	1.345	1.761	2.145	2.624	2.977
15………………………….	1.341	1.753	2.131	2.602	2.947
16………………………….	1.337	1.746	2.120	2.583	2.921
17………………………….	1.333	1.740	2.110	2.567	2.898
18………………………….	1.330	1.734	2.101	2.552	2.878
19………………………….	1.328	1.729	2.093	2.539	2.861
20………………………….	1.325	1.725	2.086	2.528	2.845
21………………………….	1.323	1.721	2.080	2.518	2.831
22………………………….	1.321	1.717	2.074	2.508	2.819
23………………………….	1.319	1.714	2.069	2.500	2.807
24………………………….	1.318	1.711	2.064	2.492	2.797
25………………………….	1.316	1.708	2.060	2.485	2.787
26………………………….	1.315	1.706	2.056	2.479	2.779
27………………………….	1.314	1.703	2.052	2.473	2.771
28………………………….	1.313	1.701	2.048	2.467	2.763
29………………………….	1.311	1.699	2.045	2.462	2.756
Inf…………………………	1.282	1.645	1.960	2.326	2.576

The *t*-value describes the sampling distribution of a deviation from a population value divided by the standard error. Probabilities indicated in the subordinate of *t* in the headings refer to the sum of the two-tailed areas under the curve that lie outside the points ± *t* degrees of freedom are listed in the first column (*df*). For example, in the distribution of means in the sample size $n = 12$, $df = n\text{-}2 = 12\text{-}2 = 10$, then $_{.05}$ of the area under the curve falls in the two tails of the curve outside the interval $t \pm 2.228$, which is taken from the $t_{.025}$ column of the table.

If, for example, the above restaurant direct payroll cost estimate specified a 95% confidence interval, the range of error would be:

$$Se = £66,869 \pm £8,009\ (2.228)$$
$$= £66,869 \pm £17,844$$

Thus, it is possible to predict that 15,000 covers sold should attract a direct payroll cost of between £49,025 and £84,713, with 95 chances out of 100 (19 out of 20) the confidence interval will contain the actual cost.

Note: The 2.228 standard errors, representing the 95% confidence interval is obtained by referring to the *t*-table in Table 2. The illustration sample size is $n = 12$, degrees of freedom (df) = $n - 2 = 10$. Therefore, for a two-tailed *t*-test with 10 degrees of freedom, 5% of the area under a normal distribution curve falls in the two-tails of the curve outside the interval $t \pm 2.228$, taken from the $t_{.025}$ column of the table.

The standard error of the estimate at the 68% confidence interval ($1S_e$) and at the 95% confidence interval ($2.228\ S_e$) for a small sample is illustrated in Figure 4.

In principle, the standard error of the estimate is similar to the standard deviation in normal probability analysis, the difference being that whereas the standard deviation measures the dispersion of data points around the mean, the standard error of the estimate measures the variability around the regression line.

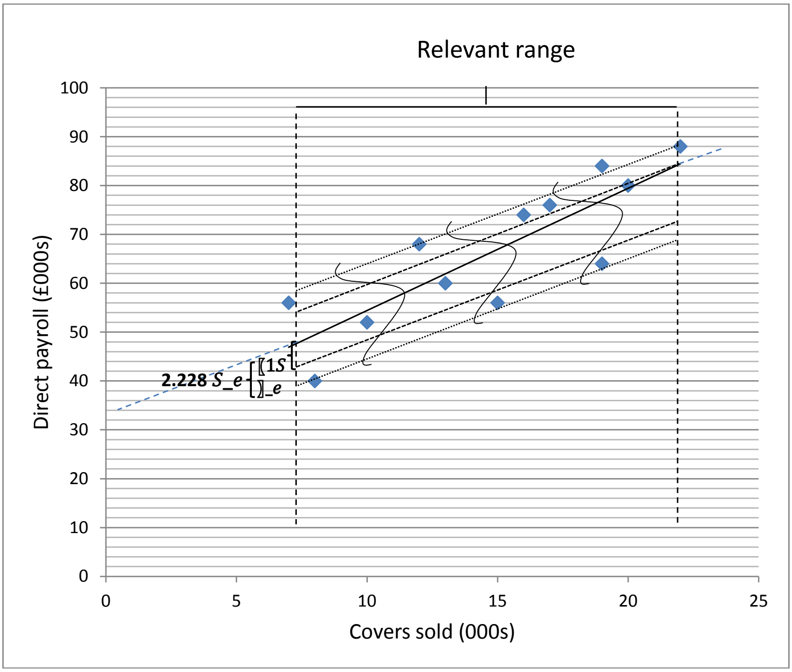

Figure 4: Regression line showing standard errors of estimate

Standard error of a prediction

As mentioned earlier, one of the reasons for engaging in cost analysis is to provide a basis for cost and profit projections. The standard error of estimate measures how closely predicted costs can be expected to match the actual costs in the same sample.

However, where an equation is employed to predict costs for a *future* period – thus using data not incorporated in the initial equation – then it becomes necessary to introduce a correction factor to the standard error of the estimate. This arises because with repeated sampling the estimated value of y will vary:

- The *standard error of a predication* (S_p) consists of the standard error of estimate with a correction factor in respect of *each* prediction.
- Remember, the estimated value of y for any given value of x is y_c. With each new sample the estimates of the intercept and slope coefficient will very to some extent. Hence, each sample will produce a slightly different regression line and, thus, a different y_c value for a given value of x.

The standard error of a prediction (is computed as follows:

$$S_p = S_e\sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{(x - \bar{x})^2}}$$

where x_p indicates the value of the new prediction (observation). Hence, the standard error of the estimate obtained from historical data is multiplied be the correction factor. The standard error of prediction is minimised when $x_p = \bar{x}$. This occurs because the regression line must always go through the mean point of the data.

Knowledge of the height of the line is most certain at the mean. Therefore, as observation values move further from the mean small changes in the slope of the line will cause increasing uncertainty as to the height of the line.

The result of this is the confidence interval band is not parallel to the regression line and so reflects a greater risk, and thus a wider interval, at the extremes of the range. Therefore, the further the covers sold prediction is from the mean of future covers sold, the wider the prediction interval will become, illustrated in Figure 5.

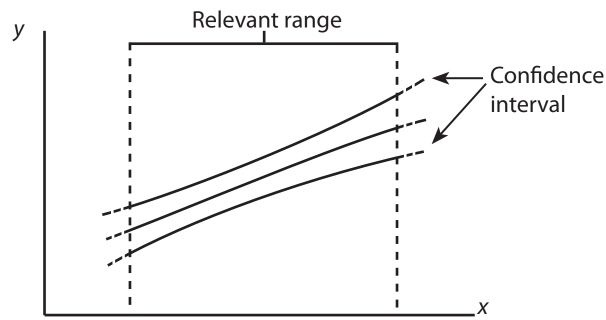

Figure 5: Confidence interval of a prediction

If 15,000 covers are predicted in a future period, and a 95% confidence interval is specified, then the standard error of the prediction will be:

$$S_p = S_e\sqrt{1 + \frac{1}{12} + \frac{(15{,}000 - 14{,}833)^2}{261.667}}$$
$$= 8.009\ \sqrt{1.0834339915}$$
$$= 8.009 \times 1.0408841987$$
$$= 8.336 \text{ or } £8{,}336$$

$$\therefore\ y_c = 366{,}869 \pm (£8{,}336)\ (2.228)$$
$$= £66{,}869\ \ \pm £18{,}573$$

It is, therefore, 95% certain that if 15,000 covers are sold, the actual payroll cost will be between £48,296 and £85,442.

Note: The higher standard error of the prediction of £18,573 compared to the standard error of the estimate of £17,844 is due to including the correction factor for a regression equation computed from a sample taken in one period to predict costs in a future period where the sample mean is not known.

Standard error of the variable cost coefficient (b)

In addition to determining the accuracy of total cost estimates and predictions, it is important to assess the reliability of the *b* coefficient (variable cost). The standard error of the regression coefficient (S_b) is computed as follows:

$$S_b = \frac{S_e}{\sqrt{(x - \bar{x})^2}}$$

Using the illustration data gives:

$$S_b = \frac{8.009}{\sqrt{261.667}} = 0.50 \text{ or } £0.50$$

If a 95% confidence interval is specified the range of probable error for the variable payroll cost *b* will be:

$$S_b = £2.50 \pm (£0.50)\ (2.228)$$
$$= £2.50 \pm £1.11$$

Therefore, there are approximately 95 chances out of 100 (19 out of 20) the true variable cost element of the restaurant direct payroll expense lies within the range £1.39 to £3.61.

Note: The standard error of 2.228 was obtained from the *t*-table critical values in Table 2 by referring to degrees of freedom row 10 under column $t_{.025}$.

Step 7: Testing the significance of cost drivers

We have computed a regression and correlation analysis which indicates the nature of the relationship between restaurant direct payroll costs (y) and number of covers sold (x), the cost driver. In this case, the results are obtained from a small sample of 12 observations. However, the question arises as to whether the results obtained from the data are significant. Do changes in the cost driver result in statically significant changes in the cost function?

Significance test for the variable cost coefficient (b)

Having computed the standard error of the regression coefficient, *b*, it is possible to test if a significant explanatory relationship exists between the *y* and *x* variables. In our restaurant example, the *b* coefficient suggests a £2.50 change in the average variable cost of direct payroll cost for each additional cover sold.

- The *b* coefficient of £2.50 is an estimate of the true variable payroll cost for the population 'B', but a particular sample may indicate a relationship by chance, even though none exists.
- If there is no relationship between *y* and *x* the true slope of the regression line will be zero; in other words *b* will be zero and restaurant payroll cost will be regarded as a fixed cost.

The relationship between the *x* and y variables can be tested by using the *null hypothesis* (H_0) and the *alternative hypothesis* (H_1). If we assume the restaurant sample has been drawn from a population with a zero *b* coefficient, then:

$$H_0 : B = 0 \text{ (}\textit{no relationship}\text{)}$$

$$H_1 : B \neq 0 \text{ (}\textit{direct payroll cost varies with covers sold}\text{)}$$

To test the hypothesis it is necessary to compute how many standard errors the sample *b* is away from the population B, and compare the computed *t*-value with the *t*-table critical value. Under the null hypothesis (H_0), which assumes b = 0 (no relationship) the computed *t*-value is:

$$t\text{-value} = \frac{b-0}{S_b} = \frac{£2.50-0}{£0.50} = 5.0$$

Therefore, *b* is determined to be 5 standard errors from zero. If a 95% confidence interval is specified then reference to the *t*-table for a two-tailed test indicates a critical value of 2.228. Thus, as a deviation of above 2 standard errors is usually regarded as significant, it is unlikely that a deviation as large as 5 standard errors could occur by chance.

At the 95% confidence level there are only 5 chances out of 100 (1 in 20) that a sample indicates a significant relationship where none exists. Thus, in this case, the null hypothesis H_0 can be rejected and the alternative hypothesis H_1 is accepted i.e. that a highly significant relationship does exist between restaurant payroll cost and covers sold (assuming the specification analysis assumptions mentioned later hold).

Significance test for the correlation coefficient (r)

Having computed the coefficient of correlation *r* it is possible to test if there is a significant numerical relationship between *x* and *y* variables. In our restaurant example, $r = 0.8478$ (*or* 0.85) suggesting 71.88% (*or* 72%) of direct payroll cost behaviour is explained by covers sold.

The relationship between the x and y variables can also be tested by using the 'null hypothesis' (H_o) and the alternative hypothesis (H_1) expressed as follows:

$$H_0 : r = 0 \ (\textit{no relationship})$$

$$H_1 : r \neq 0 \ (\textit{direct payroll cost varies with covers sold})$$

To test the hypothesis it is necessary to compute how many standard errors r is away from B, and compare the computed t-value with the t-table critical value. Under the null hypothesis (H_0), which assumes $r = 0$ (no relationship) the computed t-value is:

$$S_r = \sqrt{\frac{1 - r^2}{n - 2}} = \sqrt{\frac{1 - 0.7188}{12 - 2}} = \sqrt{0.02812} = 0.1677$$

$$t\text{-}value = \frac{r}{S_r} = \frac{0.8478 - 0}{0.1677} = 5.0$$

Therefore, r is determined to be 5 standard errors from zero. If a 95% confidence interval is specified then reference to the t-table for a two-tailed test indicates a critical value of 2.228. Thus, as a deviation of above 2 standard errors is usually regarded as significant, it is unlikely that a deviation as large as 5 standard errors could occur by chance.

Again, at the 95% confidence level there are only 5 chances out of 100 (1 in 20) that a significant relationship is indicated where none exists. Thus, the null hypothesis H_0 can be be rejected and the alternative hypothesis H_1 is accepted (assuming the assumptions mentioned later hold).

Note: The sign attached to r is the sign of b in the regression equation Therefore, the result and interpretation of the significance test for r is similar to that of the b coefficient, highly significant. However, testing r is included here as correlation analysis is a useful technique to apply in the search for suitable cost drivers of cost functions.

Step 8: Evaluating cost drivers of estimated cost functions

Once the sample data is processed the final step is to evaluate the cost driver of the estimated cost function as a predictor of future costs. So, the question arises as to how managers and management accountants assess cost drivers of cost functions?

Cost driver: restaurant illustration

There are four key selection criteria that can be used to evaluate cost drivers of cost functions derived from regression analysis and these are explained below:

- **Criterion 1: Plausibility.** The positive physical relationship between monthly *number of covers sold* and *restaurant direct payroll cost* is economically and operationally plausible; producing and selling more meals requires more restaurant labour, which results in higher total direct payroll costs.
- **Criterion 2: Goodness of fit.** A cost driver should explain a considerable amount of the variation in a cost function. The *coefficient of determination* (r^2) is a particularly useful indicator of the goodness of fit as it measures the percentage variation in the cost function explained by the cost driver; in this case 72% of *direct payroll cost* is explained by *number of covers sold*. Since an r^2 of greater than 0.30 is usually regarded as acceptable an $r^2 = 0.72$ indicates a strong relationship between the variables.
- **Criterion 3: Significance of the independent variable.** The *t*-value of a variable cost coefficient (b) measures the significance of the relationship between the changes in the cost function and the changes in the cost driver; in this case $b = 5.0$ standard errors from zero at the 95% confidence interval. Generally, if there are 30 or more observations in a sample and the *t*-value is greater than 2.0 at the 95% confidence interval, then a variable cost element can assume to be present in the cost population (B) as a whole. It is unlikely for a deviation as large as 5.0 standard errors could occur by chance.
- **Criterion 4: Specification analysis.** There are underlying assumptions of regression analysis – referred to in earlier sections – required to be satisfied in order to make valid estimates and predictions from sample data about population relationships, and these are as follows:
  - **Linearity**. This can be identified visually from a scatter diagram and should reflect a general linear (straight line) trend in the data. Figure 1 reveals that linearity is a likely reasonable assumption in the case of the restaurant *direct payroll cost* and *number of covers sold* data.
  - **Constant variance.** The standard errors and variance of the residuals should be constant for all values of x, which means there is a uniform dispersion of points about the regression line, as shown in Figure 6 (Diagram 1). Where this is not so, as shown in Figure 6 (Diagram 2) the reliability of the slope coefficient (b) is reduced; not affecting the accuracy of the equation estimates, but affecting the reliability of standard error estimates. Figure 2 reveals the restaurant data has a constant variance.
  - **Normality**. Distribution of data points about the regression line should approximately follow a normal curve, i.e. the residuals are normally distributed, as appears to be the case with the restaurant data in Figure 4. However, this is difficult to determine with small samples of data.
  - **Independence.** Residual values should be independent of one another. This means the deviation of one data point about the regression line should be unrelated to the deviation of any other data point. Where this is not so then serial correlation is said to

be present. One measure used to determine the presence serial correlation in sample data is the Durban-Watson statistic which is incorporated in many computer programmes. However, this condition can be checked on a scientific calculator by computing the coefficient of correlation (r) or coefficient of determination (r^2) of the cost residuals which, in the case of the restaurant data is $r = 0.39$ ($r^2 = 0.15$).

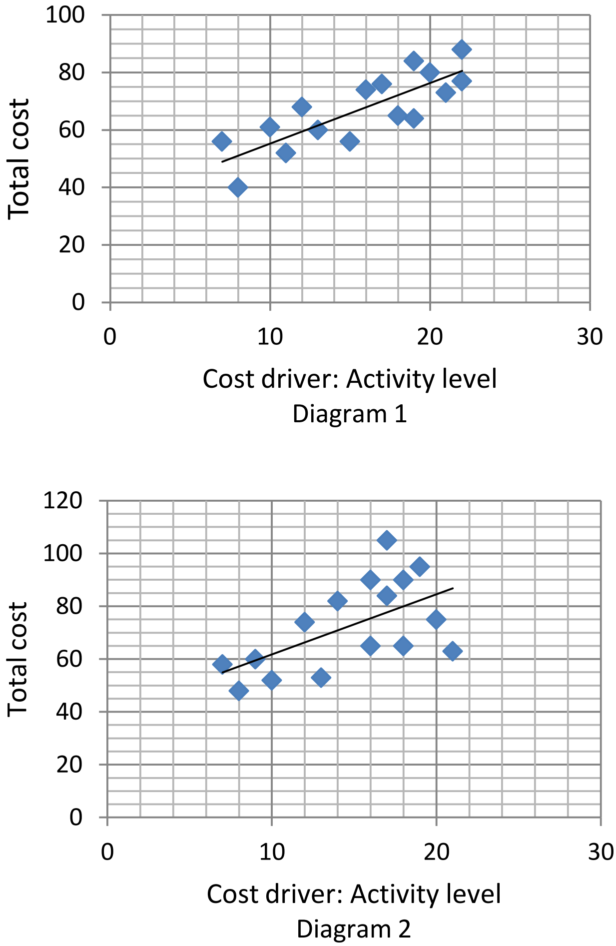

Figure 6: Examples of constant and non-constant variance of residuals

In cases where linearity, constant variance, normality and independence assumptions are satisfied the regression coefficients and standard errors determined from a sample can be regarded as efficient, linear and unbiased estimates of the true population values.

Conclusion

A summary of the above restaurant evaluation of the cost driver *number of covers sold* is presented in Table 3. The findings indicate the restaurant data sample generally satisfies the selection criteria. In hotel restaurants, and indeed stand-alone restaurants, direct payroll cost is commonly considered to change with number of covers sold and, therefore, represents a plausible operational relationship. This is supported by the data with a strong coefficient of determination and significant *t*-value.

Table 3: Restaurant cost driver: number of covers sold evaluation summary

Criteria	
1. ***Plausibility***	Positive relationship between direct payroll and covers sold makes economic and operational sense.
2. ***Goodness of fit***	$r^2 = 0.72$ indicates a strong degree of association between the variables.
3. ***Significance of the independent variable***	*t*-value = 5.0 indicates a significant relationship at the 95% confidence interval; and the 99% confidence interval.
4. ***Specification analysis***	
Linearity	There appears to be a clear linear trend present in the data plot.
Constant variance	This appears to be satisfied, but is based on only 12 observations.
Normality	Difficult to draw conclusions from 12 observations.
Independence	$r = 0.39$ indicates the assumption of independence is reasonable in the regression equation.

Note: In the case of the restaurant illustration the convincing results need to be weighed against the small size of the data sample (less than 30 pairs of observations). A balance must be struck between collecting representative data samples from periods where the operating situation may have been different to current working practices and economic conditions.

Fixed cost element

Explanation has centred on testing and evaluating the variable cost coefficient, but it is important from a practical standpoint to understand the role of the fixed cost element in cost analysis. A business undertaking tends to operate within a particular band of activity (relevant range) and, therefore, it is inappropriate and often dangerous to make estimates beyond the range of the observed data.

In his seminal article, Benston (1966) pointed out that it is tempting to interpret the constant term, *a*, as a fixed cost by extending the regression line back to zero activity. This presupposes a linear relationship which, as indicated by Figure 7, may not be a valid assumption.

In the case of our restaurant illustration, the regression line was fitted from the equation, where the data points are the observed values of the cost and activity. The line provides an estimate of the fixed cost if the range of observation included the point where activity is zero. However, if additional cost and activity observations were available they might show that the broken curve fitted with the fixed cost *a*, itself, being zero at zero activity.

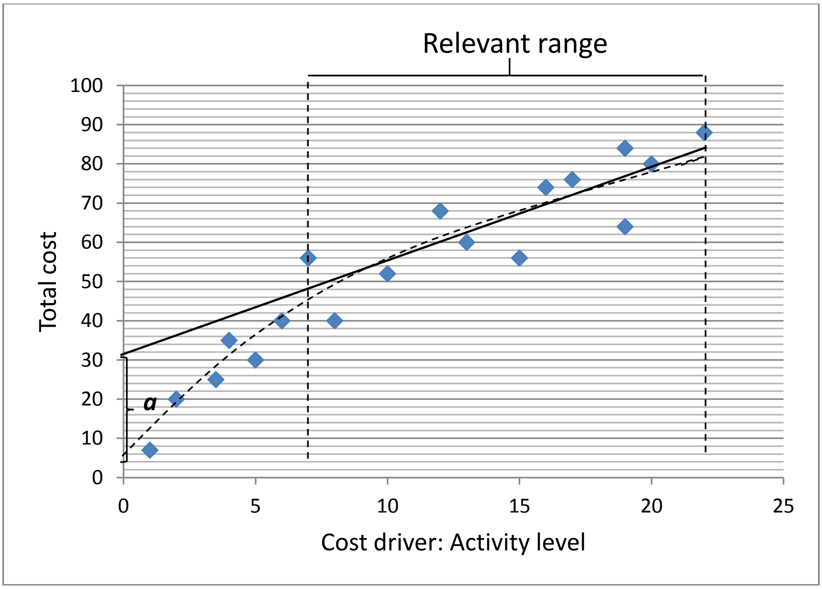

Figure 7: Restaurant fixed cost and the relevant range, adapted from Benston (1966)

It, therefore, becomes apparent that *a* is not the cost that would necessarily be found if the level of activity was at zero, but simply the value that is obtained as a result of the regression line computed from the available data. This helps to explain why the *t*-value of the fixed cost element has not been computed for our restaurant illustration. Since the key objective is to estimate and predict how costs behave as activity levels change over the relevant range, which is usually not at zero activity, the *t*-value of *a* is not normally relevant.

Simple linear regression: hotel illustration

So far, we have explained the step-by-step process and method of quantitative analysis using regression techniques with an independent restaurant business direct payroll expense.

Now, with knowledge of the process, we can consider how to analyse cost functions (behaviour) routinely on a day-to-day basis. Let's take an example from hotel indirect expenses (overhead) by illustrating an undistributed operating expense, such as utility costs.

An attempt is being made to understand the cost behaviour of hotel utility costs, arising from energy and water consumption. After an initial assessment management believes that utility costs are primarily influenced by the level of room occupancy in terms of *number of rooms occupied* (cost driver). It is economically plausible that number of rooms occupied would assist in explaining the variations in utility costs in the hotel. Data from a hotel's utility costs and rooms occupied for a one year period is presented in Table 4.

Table 4: Hotel monthly rooms occupied and utility costs

Month	No. of rooms occupied	Utility cost £
1	1,220	20,350
2	1,665	23,243
3	2,435	24,550
4	4,746	27,114
5	4,497	23,732
6	5,629	25,879
7	5,110	25,200
8	6,024	27,860
9	3,346	26,105
10	4,100	27,283
11	5,768	30,047
12	2,910	22,945
Total	47,450	304,308

The summary simple linear regression analysis of monthly utility costs against number of rooms occupied can be plotted on a scatter diagram from Excel; including the regression equation and the coefficient of determination (r^2) displayed on the chart, presented in Figure 8.

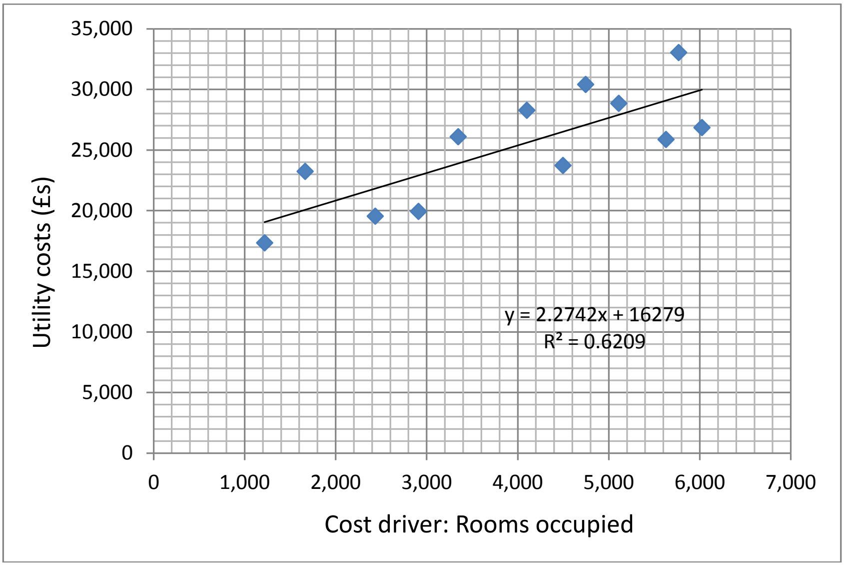

Figure 8: Simple linear regression of monthly utility costs v. number of rooms occupied

The coefficient of determination (r^2) in Figure 8, denoted as *R Square* in Excel format, indicates 62% of utility cost variation (behaviour) is explained by number of rooms occupied.

In addition to the results in Figure 8, a summary of the key results of the analysis, including the quality of data fit and the significance of the cost driver, from Excel data analysis, is presented in Table 5.

Table 5: Summary simple regression results for monthly utility costs against one independent variable (cost driver): number of rooms occupied.

Regression Statistics:		Coefficients	Standard Error	t-Statistic
Multiple R	0.79			
R Square	0.62			
Standard Error	3,031			
Intercept		16,279	2,388	6.82
Rooms occupied		2.27	0.56	4.05

Referring to Table 5, an interpretation of Excel results for the simple linear regression analysis of rooms occupied and utility costs indicates the following:

- The r^2 of 0.62 for the simple linear regression.
- The *t*-values (statistic) of the independent variable coefficient (cost driver) number of rooms occupied (£2.27) is:

$$t \text{ value} = \frac{b - 0}{S_b} = \frac{£2.27 - 0}{£0.56} = 4.05$$

- which is significantly different from zero at the 95% confidence interval with (10 degrees of freedom) as it is above the critical 2.228 value, indicated in Table 2.
- As explained earlier, under the *fixed cost element* section, the intercept *t*-value (in this case 6.82) is not relevant to the evaluation of results as this simply represents the value obtained from computing the regression equation from the available data (relevant range) and is not necessarily the fixed cost when the level of activity is zero.

Note: In Table 5, *Multiple R*, as denoted in Excel, facilitates the use of multiple independent variables (cost drivers) are used in an analysis (explained in the following section). As only one cost driver (number of rooms occupied) is included to analyse utility costs, the $r = 0.79$ here simply reflects the results using one cost driver, giving $r^2 = 0.62$.

Note: Utility consumption costs are likely to be more closely related to the number of guests staying in a hotel rather than the number of rooms occupied. Therefore, whilst the number of rooms occupied cost driver produced a strong r^2, it is possible that a more refined cost driver, such as number of guests/sleeper-nights, could provide a higher r^2.

Conclusion

A summary of the hotel utility costs the cost driver *number of rooms occupied* is presented in Table 6. The findings indicate the room occupancy and utility costs data sample generally satisfies the selection criteria. In the nature of hotel operations, utility costs are influenced by room occupancy, therefore, representing a plausible operational relationship. This is supported by the data with a strong coefficient of determination and significant *t*-value.

Table 6: Hotel cost driver: number of rooms occupied evaluation summary

Criteria:	
1. Plausibility	Positive relationship between rooms occupied against utility costs makes economic and operational sense.
2. Goodness of fit	$r^2 = 0.62$ indicates a strong degree of association between the variables.
3. Significance of the independent variable	*t*-values = 4.05 indicates a significant relationship at the 95% confidence interval.
4. Specification analysis	
Linearity	There appears to be a linear trend present in the data plot.
Constant variance	This appears to be satisfied, but based on only 12 observations
Normality	Difficult to draw conclusions from 12 observations.
Independence	$r = 0.45$ indicates the assumption of independence holds in the regression equation.

As the hotel example indicates, once the process and method of regression analysis is familiar, the estimating equations, quality of data fit and significance tests can be carried out rapidly with the assistance of software programmes, such as Excel. However, the relevance and quality of the cost function (behaviour) information produced rests on two important elements:

- Managers' knowledge of the operating characteristics their own properties when selecting costs and cost drivers for analysis.
- The reliability of the routine accounting and recording systems generating the costs and cost driver data sets to be incorporated in the analysis.

Note: In the hotel illustration, the convincing results need to be weighed against the small size of the data sample (less than 30 pairs of observations). A balance must be struck between collecting representative data samples from periods where the operating situation in the past may have been different to current working practices and economic conditions.

Multiple linear regression analysis

Our restaurant payroll and hotel utility cost examples illustrate the approach to determining cost behaviour using simple linear regression analysis. In these instances, satisfactory estimation of the payroll and utility cost functions were achieved using single independent variables (cost drivers), *number of covers sold* and *number of rooms occupied*. However, in some cases, developing a cost estimation equation with more than one independent variable (using multiple cost drivers) can enhance plausibility and improve data fit.

Multiple linear regression, which is an extension of simple linear regression, provides the means to measure the joint effect of two or more independent variables upon a dependent variable. The equation to express the relationships between multiple independent variables is as follows:

$$y_c = a + b_1 x_1 + b_2 x_2 + \dots\ b_n x_n$$

where, y_c = estimated total cost (dependent variable); a = fixed cost (constant); $b_1 b_2$ = average variable cost per unit of activity (slope coefficients); and $x_1 x_2$= the cost drivers (independent variables e.g. rooms occupied, covers sold):

- In a multiple regression analysis with two independent variables, the regression line takes the form of a plane which is fitted using a modified version of the equation of a straight line.
- The terms b_1 and b_2 are net regression coefficients and each one measures the net change in the particular independent variable.
- Since it is the simultaneous influence of all variables on y which is being measured, the net effect of x_1, or any other x, must be determined apart from any influence of other variables
- For example, b_1 measures the change y_c in per unit of change in x_1 whilst holding other independent variables constant; similar in principle to the calculation of flexible budget variances where, for instance, prices and costs are held constant in order to measure the effect on profit of a change in the volume of products or services sold.

Let's consider the example of a hotel business which operates food & beverage services, including a banqueting department offering a range of banquet events; where each event forms a discrete, one off, occurrence.

A first attempt is being made to understand the cost behaviour of direct payroll expense in the banqueting department. After an initial assessment management considers that, in addition to number of labour hours worked (cost driver), department payroll is also affected by number of banquet events (cost driver) per month. It is economically plausible that number of banquet labour hours worked and number of banquet events held would help explain the variations in banqueting payroll costs in the hotel. Data for the hotel banqueting department example for a one year period is presented in Table 7.

Table 7: Banqueting monthly number of events, number of labour hours worked and direct payroll expense

Month	Events	Labour hours worked	Payroll costs £
1	11	2,040	22,560
2	15	1,380	21,300
3	26	1,960	30,620
4	32	2,700	36,330
5	25	2,880	43,570
6	21	2,160	27,510
7	23	2,040	35,700
8	20	1,850	23,100
9	24	2,800	30,960
10	28	2,460	38,340
11	34	2,240	35,400
12	29	1,440	28,890
Total	288	25,950	374,280

However, before carrying out a multiple linear regression on the hotel banqueting department payroll data in Table 5, let's first prepare a simple linear regression analysis so we can subsequently compare and assess the multiple regression results.

Simple linear regression: banqueting illustration

The summary simple linear regression analysis of monthly payroll cost against number of labour hours worked and number of banquet events are presented in Figure 9 and Figure 10 respectively.

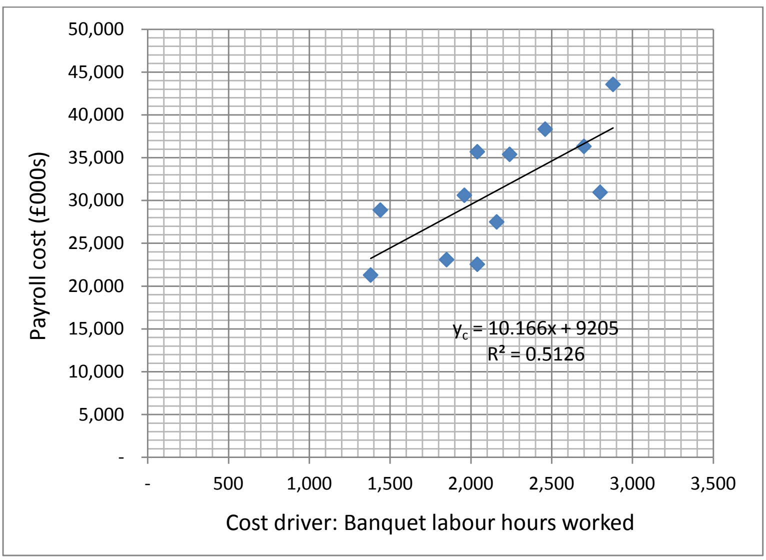

Figure 9: Simple linear regression results of monthly banqueting payroll costs against number of banquet labour hours worked

The coefficient of determination (r^2) in Figure 9, denoted as *R Square* in Excel, indicates 51% of banqueting payroll cost variation (behaviour) is explained by the number of banquet labour hours worked.

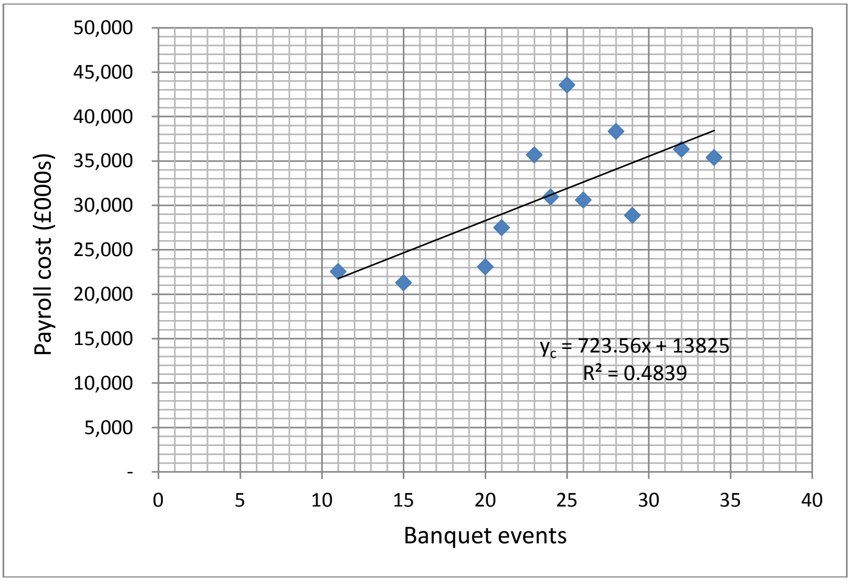

Figure 10: Simple linear regression results of monthly banqueting payroll costs against number of banquet events

The coefficient of determination in Figure 10, indicates 48% of banqueting payroll cost variation (behaviour) is explained by the number of banquet events.

Note: Remember, simple linear regression measures the *average* amount of change in a given dependent variable (cost) that is associated with changes in one independent variable (cost driver), such as in this case banquet payroll against labour hours worked and banquet payroll against banquet events .

Multiple linear regression: banqueting illustration

Estimates of the relationship between number of banquet events and number of banquet labour hours worked per month against banqueting department payroll expense can be carried out in Excel using multiple regression analysis to give the following equation:

$$y_c = £2{,}305 + £7.57\ x_1 + £521\ x_2$$

where x_1 is number of banquet labour hours worked and x_2 is number of banquet events; signifying the fixed cost element £2,305 plus variable cost elements £7.57 per labour hour worked and £521 per banquet held of banqueting total payroll cost. The key results of the analysis, including the quality of data fit and the significance of the cost drivers are presented in Table 8.

Table 8: Summary multiple regression results for monthly banqueting department payroll expense against two independent variables (cost drivers): number of banquet labour hours worked; and number of banquet events.

Regression Statistics:		Coefficients	Standard Error	*t*-Statistic
Multiple R	0.85			
R Square	0.73			
Adjusted R Square	0.67			
Standard Error	3,969			
Intercept		2,305	6,009	0.38
Banquet labour hours worked		7.57	2.64	2.87
Banquet events		521.35	193.84	2.70

Referring to Table 8, an interpretation of Excel results for the multiple linear regression analysis of banquet labour hours worked, banquet events and banqueting department payroll cost indicates the following:

- The r^2 of 0.51 for the simple linear regression using number of labour hours worked (Figure 9) increases to 0.73 with the multiple linear regression.
- Adding additional independent variables always improves r^2, but at the same time reduces the degrees of freedom in the data. In this case, by adding number of banquet events to the equation, the increase in adjusted r^2 appreciably outweighs the degree of freedom lost.
- The standard error of estimate (S_e) of the multiple regression equation that includes number of labour hours worked and number of banquet events as an independent variable (cost driver) is:

  $$S_e = \sqrt{\frac{(y - y_c)^2}{n-3}} = \sqrt{\frac{141{,}768{,}160.2}{9}} = £3{,}969$$

  which is correspondingly lower than the standard error of the simple regression equation, with only labour hours worked as the independent variable (cost driver) is £5,063 (computation not shown).
- The t-values (statistic) of the independent variable coefficients (cost drivers) of both number labour hours worked (£7.57) and number of banquet events (£521.35) which are respectively:

  $$t \text{ value} = \frac{b-0}{S_b} = \frac{£7.57 - 0}{£2.64} = 2.87$$

  $$t \text{ value} = \frac{b-0}{S_b} = \frac{£521.35 - 0}{£193.84} = 2.70$$

  are significantly different from zero at the 95% confidence interval with (9 degrees of freedom) as they are both above the critical 2.26 value, indicated in Table 2.

- Again, as explained earlier, under the *fixed cost element* section, the intercept *t*-value (in this case (0.38) is not relevant to the evaluation of results as this simply represents the value obtained from computing the regression equation from the available data (relevant range) and is not necessarily the fixed cost when the level of activity is zero.

Multicollinearity

An important constraint of applying multiple regression equations is the presence of *multicollinearity*. Where multicollinearity is present in an equation, the independent variables are highly correlated with one another and their individual relationships with the dependent variable cannot be accurately determined; from an operating perspective, making it impossible to separate individual coefficients (marginal costs) of cost drivers.

Managers will normally prefer to determine the marginal costs of each type or group of products or services, but this depends on the relationship between the independent variables. For example, take the case of a full-service hotel comprising accommodation, restaurant and bar departments where the analysis of undistributed operating expenses is being undertaken. If demand for the departments is highly correlated the volume of rooms, covers and drinks will vary together and make it impossible to disaggregate the marginal costs of letting rooms from the costs of producing meals or serving drinks.

However, where the correlation between the independent variables continues, the regression equation can provide accurate predictions of *total* undistributed operating costs of the full-service hotel for e.g. flexible budget preparation.

Referring to our hotel banqueting department example, the presence of multicollinearity can be tested by determining the coefficient of correlation (r) between the two independent variables, number of labour hours worked and number of banquet events, and presented in Figure 11.

A coefficient of correlation greater than 0.70 between independent variables in a multiple regression equation is usually considered indicative of multicollinearity. Therefore, the significantly lower $r = 0.3645$ obtained in Figure 10 suggests the analysis is unlikely to contain problems of multicollinearity, reflected in the coefficient of determination $r^2 = 0.1329$, indicating that only 13% of number of banquet labour hours worked is explained by number of banquet events.

Reviewing the scatter diagram of Figure 11 with Figures 9 and 10 visually contrasts the low degree of correlation between banquet labour hours worked and banquet events, compared to the high correlation between banqueting payroll cost and labour hours worked, and banqueting payroll cost and banquet events; emphasizing the absence of multicolineraity in labour hours worked and banquet events.

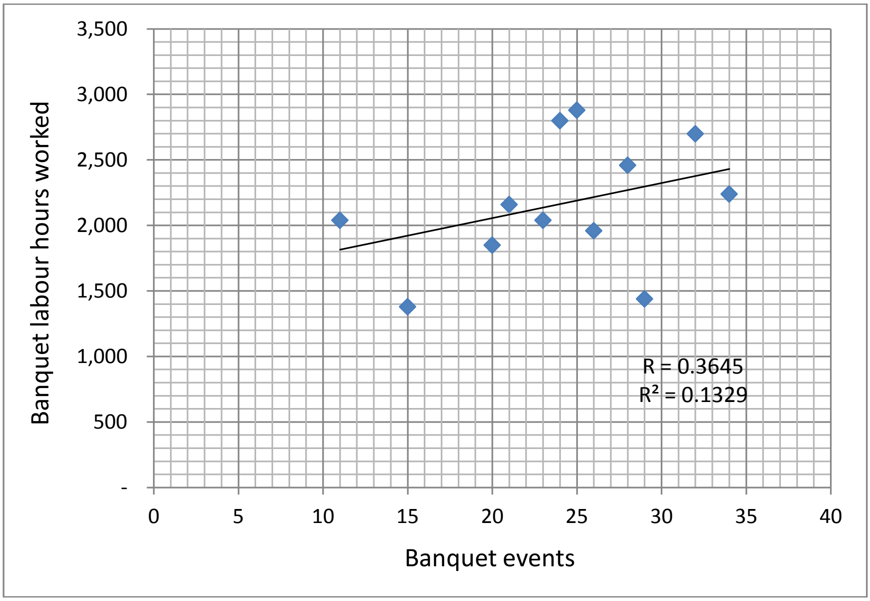

Figure 11: Correlation results of monthly banquet labour hours worked against number of banquet events

Conclusion

A summary of the above hotel banqueting department evaluation of the cost drivers *number of banquet labour hours worked* and *number of banquet events* is presented in Table 9. The findings indicate the banqueting data sample generally satisfies the selection criteria. In the nature of banqueting operations, direct payroll cost is influenced by the type of occasion (e.g. cocktail party, wedding or dinner) level of service (e.g. formal sit-down meal or buffet style), size (number of covers) and quality (e.g. style of menu supplementary requirements) affecting the labour hours worked for the F&B production and service and setup and changeover of events, therefore, representing a plausible operational relationship. This is supported by the data with a strong coefficient of determination and significant *t*-value.

Note: In the banqueting department illustration, the convincing results need to be weighed against the small size of the data sample (less than 30 pairs of observations). A balance must be struck between collecting representative data samples from periods where the operating situation in the past may have been different to current working practices and economic conditions.

Table 9: Banqueting department cost drivers: number of banquet labours hours worked and number of banquet events evaluation summary

Criteria:	
1. Plausibility	Positive relationship between banquet labour hours worked and banquet events against banqueting payroll cost makes economic and operational sense.
2. Goodness of fit	$r^2 = 0.73$ indicates a strong degree of association between the variables.
3. Significance of the independent variable	*t*-values = 2.87 labour hours worked and 2.70 events indicates a significant relationship at the 95% confidence interval.
4. Specification analysis	
Linearity	There appears to be a linear trend present in the data plot.
Constant variance	This appears to be satisfied, but based on only 12 observations
Normality	Difficult to draw conclusions from 12 observations.
Independence	$r = 0.02$ indicates the assumption of independence holds in the regression equation.
Multicollinearity	$r = 0.36$ suggests the analysis is unlikely to encounter problems associated with multicollinearity.

Quantitative analysis of cost functions in decision making

The use of regression for cost estimation and prediction can assist in a wide range of decision-making settings. However, this will only take effect if managers and financial controllers are prepared to engage with analytical techniques of cost estimation to complement the traditional methods of analysing costs based on technical estimates solely using knowledge, judgement and experience. Horngren *et al* (2009) reminds us:

> 'Understanding how costs behave is a valuable technical skill. Managers look to management accountants to help them identify cost drivers, estimate cost relationships, and determine the fixed and variable components of costs'.

Planning Decisions

Regression techniques can be helpful in profit planning, budgeting and pricing decisions as managers are constantly making day-to-day operating decisions which affect revenue, cost and profit. The individual decisions themselves may only involve relatively modest-sized transactions, but when accumulated over a trading period they can have a dramatic influence on operating profit.

- Developing regression equations to determine operating cost behaviour facilitates the assessment of business alternatives and opportunities using break-even analysis, flexible budgets and profit sensitivity techniques.

- Regression equations can be applied to the construction of flexible budgets which can be used to assess the likely impact on profits of `what if' scenarios relating to new business transaction negotiations, pricing structures, sales volumes and business mix decisions.
- Knowledge of the variable cost coefficient can also assist in *ad hoc* pricing decisions where the range of price discretion is required to be determined in order to submit competitive bids.
- Providing estimates of total costs and average variable costs per unit, regression enables the computation of the range of probable error through use of the standard error of the estimate and standard error of the variable cost coefficient.
- The use of crude unqualified estimates as a basis for cost prediction is no longer acceptable in today's operating environment. As far back as the 1960s the use of crude unqualified estimates have been severely criticise. Among others Amey (1961) argued that 'much of their [accountants'] apparent precision is found to be spurious; no self-respecting statistician would present an estimate without indicating the error to which it was thought to be subject'; with apparently little evidence of change occurring in the intervening 50 years, particularly in service sectors, such as hospitality.

Control Decisions

Regression analysis can also contribute to a number of areas of routine cost control, in the following ways:

- The regression equations applied to construct flexible budgets can be used to compare actual costs incurred at a specific level of activity with predicted costs at the same level. Here cost predictions for an original (static) budget are flexed (adjusted) to the relevant level of activity achieved during a period. This can move towards determining variances for subsequent evaluation, such as assessing the efficiency of delivering products and services and measuring the impact of volume changes and business mix on profit margins.
- Cost standards developed from statistical analysis of historical data do not necessarily reflect efficient or optimal performance, but may actually indicate the level of cost behaviour that occurred in the past. The standards can be used to suggest whether current operations have improved or deteriorated from the past, but cannot in themselves suggest whether past activities represented an acceptable level of efficiency. In order to ascertain this detailed examination of the department in question would need to be carried out. As Kaplan (1982) pointed out:

> `High standard errors (or low r^2) are a result of large fluctuations in the cost centre. Thus, if the analyst, when modelling the cost behaviour in a cost centre, observes a poor fit to the historical

data, he or she may conclude that the cost centre is not operating in a state of statistical control – that too many large fluctuations have occurred that cannot be explained by variations in the cost centre's activity levels'. In the event of such findings, operating procedures and working arrangements could be reviewed in an attempt to secure a reduction in the erratic cost behaviour patterns in the department'.

It becomes apparent, therefore, that cost investigations of a particular department can be prompted by either current or past results, i.e. changes in the mean of the actual data (cost observations which are more than two standard errors from the predicted figure), or by a configuration of historical cost data that is regarded as being too widely dispersed.

In conclusion, quantitative cost analysis offers an objective approach to complement the more subjective methods of analysing costs. With the widespread availability of computers and software packages statistical cost analysis is accessible and can provide a powerful contribution to routine, day-to-day, decision making.

Notes for further reference

Amey, L.R. (1961) Professor of Accounting, University of Bristol, *Inaugural Lecture*.

Benston, G.J. (1966) Multiple Regression Analysis of Cost Behaviour, *The Accounting Review*, 41(4), pp 657-672.

An award winning seminal paper that provides an outstanding exploration into the complexities and practicalities involved in statistical cost analysis using multiple variables.

Drury, C. (2007) *Management and Cost Accounting*, 7th ed., Thomson Learning: London.

Provides a detailed and well explained approach to statistical cost estimation.

Harris, P.J. (2011) *Profit Planning for Hospitality and Tourism,* 3rd ed., Goodfellow Publishers: Oxford.

Horngren, C.T., Datar S.M., Foster, G., Rajan, M.V., & Ittner, C. (2009) *Cost Accounting: a managerial emphasis*, 13th ed., Pearson Education: Upper Saddle River, NJ.

Includes an excellent generic overview to understanding cost behaviour and quantitative analysis of cost functions.

Kaplan, R.S. (1982) *Advanced Management Accounting*, 2rd ed., Prentice-Hall: Englewood Cliffs, NJ.

This edition contains extensive explanation and advice on the methods and application of statistical cost analysis.

Kaplan, R.S. & Atkinson, A.A (1998) *Advanced Management Accounting*, 3rd ed., Prentice-Hall International: London.

This edition contains clear guidance on the application of cost estimation techniques for cost planning and control in routine business operations.

Kotas, R. (1973) Market Orientation, *Hotel, Catering and International Management Association Journal*, July, pp 5-7.

A simple, practical, but landmark article in drawing attention to the nature and concept of business orientation.

Matz, A & Usry, M.F. (1980) *Cost Accounting, Planning and Control*, 7th ed., South West Publishing: Cincinnati.

Contains a pragmatic guide to the methods cost analysis and estimation.