

Full Length Research Paper

Exploratory study of selected statistical techniques with special relevance to business

Angus Okechukwu Unegbu^{1*}, Chinyere Unegbu² and Tasie George³

¹Department of Accounting, American University of Nigeria, Yola, Adamawa State, Nigeria.

²Federal University of Technology, Yola, Adamawa State, Nigeria.

³Faculty of Management Science, American University of Nigeria, Yola, Adamawa State, Nigeria.

Accepted 18 August, 2011

This paper examines selected statistical tools of correlation, regression analysis, t-test, Chi-square and analysis of variance (ANOVA) and their areas of relevance to business decisions. The method of research adopted is exploratory. In search for solution to the advocated research questions, near live examples were given and solved using both the stated statistical formula and computer SPSS. The study found out that, each of the selected statistical tools has very unique importance to the business decisions and they are recommended for use as aid to decision to accountants, financial analysts, business managers, investors and business owners.

Key words: Selected statistical techniques, accounting and business decision aids.

INTRODUCTION

Accounting has become a major tool for planning and controlling business activities. Bull (1981) described accounting as an area of study concerned with the quantification of economic events in monetary terms in order to collect, record, evaluate and communicate the results of past events and to aid in business decision making. The collection of accounting data, recording, evaluation, quantification and communication of these economic events many a time requires statistical techniques and the use of computer.

In the auditing and investigation podium, the use of analytical techniques are very essential tool of extracting evidence to aid in the formation of opinion and recreation of real events from available samples. Analytical procedures employ statistical techniques. The aforementioned assertion is given a helping hand by Messier et al. (2008)

when they cited "Accounting and Auditing Pronouncements No 329" which defined analytical procedures as consisting of evaluations of financial information made by a study of plausible relationships among both financial and nonfinancial data. Thus analytical tools broadens accounting quest for the use of statistical techniques and computer.

Statement of the problem

Despite the established relevance of statistical techniques in business and accounting data capture, presentation and communication to different users of accounting information, there are less exploratory studies on these statistical techniques and areas of their importance in the accounting parlance, especially correlation, t-test, regression, Chi-square and analysis of variance.

Research questions

In a bid to carry out a detailed exploratory study of statistical techniques such as correlation, regression, t-test, Chi-square and ANOVA, the study raise the

*Corresponding author. E-mail: unegbu4@yahoo.com. Tel: +2348033397607.

Abbreviations: ANOVA, analysis of variance; VAT, value added tax; SSE, sum of squares for errors; RER, respiratory exchange ratio.

following questions for each of the earlier mentioned statistical techniques;

How is it defined?
 How can it be vividly explained?
 How does it operate?
 What type of data is it meant for?
 How can it be applied?
 What is the decision rule?
 Can it be applied in accounting field of study, citing two examples?
 What are the short comings of the statistical technique in question?
 Does each of the selected statistical technique possess any uniqueness from other statistical techniques and advantages?

Significance of the study

The study will high light the relevance of the selected statistical techniques in business as a field of study. It will vividly explain how to use computer and computer programs to enable each of the named statistical techniques. Business managers will find this study very important as a working paper for the adoption of the selected statistical tools in data capture and communication decisions. The study will also add to existing academic literature and assertions.

Research methods

The research method adopted is exploratory study. Each of the selected statistical techniques will be academically explored extensively using existing academic materials and expert assertions.

Limitations and scope of the study

The study will be limited to exploratory study of existing academic material. The mathematical and statistical proof of each of the selected statistical techniques will be avoided so as to ensure a smooth flow of impacting knowledge and understanding.

LITERATURE REVIEW, PRESENTATION AND ANALYSIS OF DATA

Correlation: Introduction and definitions

Correlation is a statistical technique that can show whether and how strongly pairs of variables are related. Variables refer to items, events or business transactions, example sales and net income. A correlation may seek to verify how strongly a firm's net income is correlated with her sales during a given period. The afore definition is

supported by experiment-resources.com (17/06/2010) which asserts that statistical correlation is a statistical technique which tells us if two variables are related. Lind et al. (2006) opined that, correlation analysis consist of a group of techniques that measure the association between two variables. The basic idea of correlation analysis is to report the association between two variables.

Variables are conveniently classified into two of dependent and independent variables. Bluman (2007) and Lind et al. (2006) both are of the view that dependent variable is the event or outcome that is being predicted or estimated and that, independent variable is the event or issue that provides the basis for estimation (it is the predictor variable). In the accounting arena, a study of the effect of value added tax (VAT) on the prices of building materials is an example of event. The variables from the event are VAT and prices of building materials. Here, the outcome that is being predicted is 'prices of building materials' which is the dependent variable and the independent variable is VAT, which provides the basis of prediction. In other words, the independent variable is the causal event, while the dependent variable is the outcome, target or resultant event. Adefila (2008) simplified the explanations as Adefila stated that dependent variable is the element of relationship under investigation, which is observed. It is the response variable. Adefila went further to define independent variable as the factor that cause or at least influence that problem under investigation. Adefila (2008) introduced a third variable which he called 'intervening variable'. Continuing, he stipulated that intervening variable is a factor that interferes between the dependent and the independent variables even though it cannot be measured directly.

Buglear (2005) had earlier asserted in line with Adelifa (2008) opinion of the description of dependent variable as the 'response' by stating that, dependent variable is represented by letter Y in graphs, which is the variable whose values are believed to depend on the other variable. The other variable is represented by the letter X, is called in the independent variable. The dependent variable is also known as the response because it is believed to respond to changes in the value of the X variable. The X variable is also known as the predictor because it might help to predict the values of Y.

The decision rule criteria for correlation

In the view of Pinkmonkey.com (17/06/2010), correlation outcomes may be classified into two namely;

- a. Positive and negative correlation
- b. Linear and non-linear correlation

A .1: If two variables change in the same direction (that is if one increases the other also increases, or if one

decreases, the other also decreases), then this is called a positive correlation. For example: Advertising and sales.

A.2: If two variables change in the opposite direction (that is if one increases, the other decreases and vice versa), then the correlation is called a negative correlation. For example: T.V. registrations and cinema attendance.

B.1: The nature of the graph gives us the idea of the linear type of correlation between two variables. If the graph is in a straight line, the correlation is called a "linear correlation"

B.2: If the graph is not in a straight line, the correlation is non-linear or curvi-linear.

For example, if variable VAT (x) changes by a constant quantity, say 20 then prices of building materials (y) also changes by a constant quantity, say 4. The ratio between the two always remains the same (1/5 in this case). In case of a curvi-linear correlation this ratio does not remain constant.

Correlation is measured according to the level (strength) of relationship. In the assertion of surveysystem.com (16/06/2010) that through the coefficient of correlation, the study can measure the degree or extent of the correlation between two variables. On the basis of the coefficient of correlation the study can also determine whether the correlation is positive or negative and also its degree or extent. Pinkmonkey.com (17/06/2010) in addition to agreeing with surveysystem.com, went further to state that correlation measurement is best captured in the under listed three levels;

1. Perfect correlation: If two variables changes in the same direction and in the same proportion, the correlation between the two is perfect positive. According to Karl Pearson the coefficient of correlation in this case is +1. On the other hand if the variables change in the opposite direction and in the same proportion, the correlation is perfect negative. Its coefficient of correlation is -1.

2. Absence of correlation: If two series of two variables exhibit no relations between them or change in variable does not lead to a change in the other variable, then the study can firmly say that there is no correlation or absurd correlation between the two variables. In such a case the coefficient of correlation is 0.

3. Limited degrees of correlation: If two variables are not perfectly correlated or is there a perfect absence of correlation, then we term the correlation as limited correlation. It may be positive, negative or zero but lies with the limits ± 1 .

Methods of determination of correlation coefficients

Buglear (2005) vividly explained the meaning of

'correlation coefficient' as a precise way of assessing how closely the variables are related. Buglear (2005) while acknowledging various measurements and theories of correlation coefficients asserts that, there are two most commonly used correlation coefficients namely Pearson's product moment correlation coefficient and Spearman's rank correlation coefficient.

Pearson product moment correlation coefficient

Lind et al. (2006) traced the origin of Pearson product moment correlation coefficient to Karl Pearson who formulated the model in 1900. They asserted that Pearson product moment correlation coefficient measures the strength of the relationship between two sets of interval-scaled or ratio-scaled variables. Adefila (2008) described interval scales relationship of events that possess both order and distance but without specific origin unless such is specifically assigned unto them. A good example is thermometer which has temperature ranging from zero upwards and another example is ages of students in this presentation. In conclusion interval scale describes a business transaction containing lowest and uppermost limits of acceptable actions to maintain the validity of the contract. Example is a business where a company hereby called 'Apha' issues an invoice for N300, 000 to Ihenacho and Sons dated March 22, with the term 2/10, n/30. Interpreting this transaction will mean that a 2% discount was offered to Ihenacho and Sons if they can pay from the date of issue to the next ten days, thereafter the discount offer will elapse. The offer notwithstanding, the credit period will only last for 30 days.

Adefila (2008) is of the view that ratio scales measures events expressed in real numbers and their comparative outcomes are usually expressed as percentages. Examples include time, height and distance. The letter *r* is used to represent the Pearson correlation coefficient for sample data, while the letter *p* is used to represent it for population data.

How Pearson correlation coefficient works

Pearson's correlation coefficient is a ratio that compares the co-ordinate scatter to the total scatter and using the covariance of the values (for example *X* and *Y*) to measure the degree of co-ordinate scatter and dividing the covariance values by the product of the two standard deviations. Thus the formula is

$$r = \frac{\text{Cov } xy}{(sx \times sy)}$$

The resultant answer according to Buglear (2005) can be either positive or negative but it cannot be larger than 1 or lower than -1.

Table 1. Computation of covariance.

Price (X)	$(X - \bar{X})$	$(X - \bar{X})^2$	Y	$(Y - \bar{Y})$	$(Y - \bar{Y})^2$	$(X - \bar{X})(Y - \bar{Y})$
18	-7	49	8	4	16	-28
20	-5	25	6	2	4	-10
25	0	0	5	1	1	0
27	2	4	2	-2	4	-4
28	3	9	2	-2	4	-6
32	7	21	1	-3	9	-21
$\Sigma 136$			$\Sigma 38$			$\Sigma -69$

Example of Pearson correlation coefficient using manual solution

Illustration 1: Unimaid ventures conducted a sales promotion for six weeks. The under listed represents the price of Unimaid waters for each of the six weeks and quantity (trailer load) of the water sold (Appendix 1).

Required: To calculate the covariance and the Pearson correlation coefficient (r).

Solution 1: It is right to assume that lorry loads (quantity) of water sold depends on weekly price fixed, hence price is the independent variable (X) while quantity sold (Y) is the dependent variable. Computation of the mean of X gave 25 and that of Y gave 4. Further computations are as shown in Table 1. Given that covariance is computed thus:

$$\text{cov}(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N - 1}$$

Therefore $-69/5 = -13.8$

$$S_x = \sqrt{\sum \frac{(x - \bar{x})^2}{N - 1}} = \sqrt{136/5} = 5.215$$

$$S_y = \sqrt{\sum \frac{(y - \bar{y})^2}{N - 1}} = \sqrt{38/5} = 2.757$$

Correlation coefficient: $r = (-13.8) / (5.215 * 2.757) = -0.960$

Interpretation of result: Increases in selling price of lorry load of water correlates negatively with the quantity of water purchased each week. A direct approach to the computation of correlation coefficient is the use of a direct formula thus (Table 2).

$$r = \frac{n \sum xy - (\sum x \sum y)}{\sqrt{(n \sum x^2 - (\sum x)^2)(n \sum y^2 - (\sum y)^2)}}$$

Therefore

$$r = \frac{6 * 531 - 150 * 24}{\sqrt{(6 * 3886 - 150^2)(6 * 134 - 24^2)}}$$

$r = -0.960$

Using SPSS to compute correlation coefficient

Step 1: Create a data variable;

1st Column is tagged 'week'

2nd Column is tagged 'price'

3rd Column is tagged 'quantity sold'

Step 2: Enter the various values in the data editor.

Step 3: According to Field (2005) Pearson's correlation is bivariate correlation. A bivariate correlation is a correlation between two variables.

Field (2005) stipulated that the command for SPSS command for Pearson's correlation is 'analyze — correlate—bivariate'

Following step 1 and 2 the data entry appear as shown in Appendix 2. The result of step 3 is as shown in Appendix 3.

Spearman's correlation coefficient

This correlation is applied where the data are ranked or relate to ordinal scaled data Buglear (2005) and Field (2005). Though, it is possible to use Spearman coefficient with interval and ratio data, provided the data are ranked.

Formula for computation of Spearman correlation coefficient

Bluman (2007) citing Spearman's work gave the formula

Table 2. Computation of correlation coefficient.

Price (X)	X^2	Y	Y^2	XY
18	324	8	64	144
20	400	6	36	120
25	625	5	25	125
27	729	2	4	54
28	784	2	4	56
32	1024	1	1	32
$\sum X = 150$	$\sum X^2 = 3886$	$\sum Y = 24$	$\sum Y^2 = 134$	$\sum XY = 531$

Table 3. Computation of spearman correlation coefficient.

Week	Price (X)	Quantity sold (Y)	X ranks	Y ranks	d (x _r - y _r)	d ²
1	18.0	8.0	1	6	-5	25
2	20.0	6.0	2	5	-3	9
3	25.0	5.0	3	4	-1	1
4	27.0	2.0	4	2.5	1.5	2.25
5	28.0	2.0	5	2.5	2.5	6.25
6	32.0	1.0	6	1	5	25

$$\sum d^2 = 68.5$$

to be;

$$r = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

Illustration 2: Using the facts of illustration 1, the ranks will appear thus:

$$r = 1 - \frac{6 * 68.5}{6(6^2 - 1)} = -0.957$$

The result shows that price and quantity sold are negatively correlated (Table 3). The result is reported similar answers with that of Pearson.

Using computer SPSS to calculate Spearman's correlation

The procedure for keying the variables are the same with Pearson correlation earlier discussed, except that the study have to tick 'Spearman' in bivariate correction rather than ticking 'Pearson' the result is as shown in Appendix 4.

The answer also shows that price of lorry load of water and quantity sold is significantly negatively correlated.

Other ways of classifying correlation analysis

Other methods of classifying correlation apart from Pearson and Spearman include:

Kendall's Tau correlation: It is represented by $\hat{\tau}$. It is non-parametric correlation. According to Field (2005), Kendall's correlation are more relevant to Spearman's coefficient when there is a small data set but with many rank ties. Kendall correlation leads to more accurate generalizations than from Spearman's, Field (2005) asserted, citing (Howell, 1997).

According to Wikipedia.org (21/06/2010), the Kendall τ coefficient is defined as:

$$\tau = \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{\frac{1}{2}n(n-1)}$$

To clarify concordant and discordant pairs according to Wikipedia.org (21/06/2010), it is better explain thus; Let $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ be a set of joint observations from two random variables X and Y respectively, such that all the values of (x_i) and (y_i) are unique.

Any pair of observations (x_i, y_i) and (x_j, y_j) are said to be concordant if the ranks for both elements agree: that is, if both $x_i > x_j$ and $y_i > y_j$ or if both $x_i < x_j$ and $y_i < y_j$; otherwise they are said to be discordant.

Kendal correlation computation

Illustration 3: Using the previous illustration, the number of concordant pairs = 0 and number of discordant pairs = 6

Substituting using the formula will be;

$$r = \frac{0 - 6}{\frac{1}{2}(6)} = -0.400$$

Kendall rank correlation computing with SPSS

Statisticsolutions.com (21/06/2010) asserts that most of the procedures for Kendall rank correlation will be the same as the Pearson correlation in SPSS.

It is executed by selecting 'correlate' from the analysis menu and the next to select 'bivariate' from the correlate menu.

From the window, 'Kendall test' will be highlighted. The result is as shown in Appendix 5.

Partial and multiple correlations

Shenoy and Pant (2007) opined that, where two variables are involved, it is called simple correlation but where three or more variables are studied simultaneously, it is called multiple correlations.

Field (2005) wrote that a correlation between two or more variables in which the effects of other variables are held constant is known as a partial correlation.

Shortcomings of correlation

- Positive or negative correlation result does not necessarily mean that there is a causal relationship between the variables under consideration (McClave et al., 2005).
- Data result interpretations from correlation analysis are usually subjective (Levine et al., 2005).

REGRESSION ANALYSIS

Definition and explanations

Shenoy and Pant (2007) defines 'regression analyses' as a statistical method that deals with the analysis of mathematical models expressing the relationship between several variables. In other words, regression analysis is the study of the pattern of relationship or the actual form of relationship so that the knowledge of relationship between the variables can be used for decision making. Field (2005) differentiated between

correlation and regression as he stated that correlation concerns itself with pattern of relationship but each of the variables can only be used to predict the outcome of the other on regression analysis, thus regression is used primarily for the purpose of prediction. Levine et al. (2005) stated that in addition to prediction, regression is also used to study the relationship between dependent variable based on the values of independent variable. In a nutshell, regression analysis is basically useful for the following primary purposes namely;

- It can be used to predict the value of one variable when the value(s) of the other related variable(s) is/are known
- It is also used to examine whether there is a relationship between two or more variables
- It can further be used to describe the strength of variable relationship.

Types of regression analysis

McClave et al. (2005) classified regression analysis into two of simple and multiple regressions. Shenoy and Pant (2007) describes simple regression as a form of mathematical relationship used between two variables. Buglear (2005) asserted that a regression is called simple when it analyses two variables only. It is called linear when it is about finding a straight line. It is called multiple regressions when the analysis involves more than two variables.

How to apply the regression analysis

In the words of Buglear (2005), Shenoy and Pant (2007) and Field (2005) regression analysis is finding 'the line of best fit' of relationship between variables. This line of best fit can be ascertained thus:

$$Y = a + bX + e$$

Where: Y is the 'dependent variable' that is the outcome, X represents the 'independent variable' that is the predictor and 'b' represents slope of the line. The formula for finding b is given as;

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

A more advanced formula for b the slope is;

$$b = r \frac{\sigma_y}{\sigma_x}$$

Where: 'r' is the correlation coefficient and 'a' is the intercept. The formula to ascertain the intercept is;

Table 4. Variables for computation of the intercept 'a' and slope 'b'.

Week	Price (X)	Quantity sold (Y)	X ²	XY
1	18.0	8.0	324	144
2	20.0	6.0	400	120
3	25.0	5.0	625	125
4	27.0	2.0	729	54
5	28.0	2.0	784	56
6	32.0	1.0	1024	32
$\sum x = 150$		$\sum y = 24$	$\sum x^2 = 3886$	$\sum xy = 531$

$$a = \frac{\sum y}{n} - \frac{b \sum x}{n}$$

A direct formula for obtaining 'a' is

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

Where: 'e' is the error term.

On the process of regression

Illustration 4: Using the facts of illustration 1 and illustration 3, the results is as shown in Table 4.

Therefore,

$$b = \frac{6 * 531 - (150 * 24)}{6 * 3886 - 150^2} = -0.507, \text{ and}$$

$$a = \frac{24}{6} - \frac{-0.507 * 150}{6} = 16.675$$

Therefore $Y = 16.675 - 0.507X$

Interpretation and decision rule of regression analysis

Using the aforementioned result, the interpretation and prediction will be that;

- For every one naira increase in the price of lorry load of water, the quantity sold will reduce by 0.507 loads
- Management can from above also predict that for every one naira decrease in the price of lorry load of water, the quantity sold will increase by 0.507 of lorry loads
- The 'a' intercept shows that at zero price, the quantity to be sold will almost be 17 lorry loads of water.

Using the computer SPSS to compute simple regression

The step to follow is: analyze to regression to linear, noting the dependent and independent variables. The result is shown in Appendix 6. The interpretation of the result is as earlier stated. The Computer result further tells us that the result reflects a significant effect since .002 is less than 0.05.

Coefficient of determination

In the words of Shenoy and Pant (2007), the coefficient of determination represented by R^2 is widely used as a measure of the goodness of fit of a regression line. It measures the extent of or strength of the association that exists between the variables. In other to avoid the confusion that results among the use of the terms; standard error of estimate, correlation coefficient and coefficient of determination, it will be relevant re-explaining each vividly. Standard error of estimate measures how close the actual values are to the regression line (Lind et al., 2006). When the standard error (estimated error) is small, it shows that the two variables are closely related (Keller, 2009). The formula for the computation of standard error of estimate is

$$S_{y.x} = \sqrt{\frac{\sum (Y - \bar{Y})^2}{n - 2}}$$

The correlation coefficient measures the strength of the linear association between two variables (Lind et al., 2006). It should be noted that, standard error of estimate and the coefficient of correlation relate the same information but uses different scale to report the strength of the association. See the formula for Pearson's correlation coefficient and Spearman's.

Table 5. Showing salesgirls' performances and effort index.

Salesgirl code	Sales performance (n,000)	Aptitude test score	Effort index
1	55	50	0.8
2	60	45	0.9
3	70	80	0.3
4	75	65	1.0
5	90	70	1.2
6	95	75	1.3
7	80	80	1.1
8	90	85	0.5
9	65	70	1.0
10	70	60	0.4
11	80	55	1.6
12	60	75	0.3
13	120	90	2.0
14	105	75	1.8
15	110	80	1.6

Sources: Shenoy and Pant (2007).

The coefficient of determination measures the percent of the variation of dependent variable (Y) that is explained by the variation of the independent variable (X). The square of correlation efficient is equal to coefficient of determination. The formula is;

$$r^2 = \frac{\text{explained variation}}{\text{total variation}}$$

Multiple regression

According to Bluman (2007), multiple regression is used to find the line of best fit where there are several independent variables but one dependent variable. Except the number of independent variables regression, all the concepts of simple regression apply to multiple regression analysis. A multiple regression equation with two independent variables (X_1 and X_2) with one dependent variable (Y) will be represented thus:

$$Y = a + b_1X_1 + b_2X_2$$

Multiple regression formula and illustration

Maddala (2001) advocated that the least squares method is best suited for manual method of computing multiple regressions. The steps to follow are as listed;

Step 1: State the equation, representing the number of variables under consideration. Example with two independent variables will be;

$$Y = a + b_1X_1 + b_2X_2$$

Step 2: State the normal equations from the least square equations;

$$\sum Y = an + b_1\sum X_1 + b_2\sum X_2$$

$$\sum X_1Y = a\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1X_2$$

$$\sum X_2Y = a\sum X_2 + b_1\sum X_1X_2 + b_2\sum X_2^2$$

Illustration 5: Given that sales variables of Unimaid ventures is as shown in Table 5.

Manual solution 5:

Step 1: Compute for each of the known variables in the three normal equations (Table 6). On substitution, the normal equation will appear thus;

$$1225 = 15a + 1055b_1 + 15.8b_2$$

$$88525 = 1055a + 76575b_1 + 1129b_2$$

$$1402.5 = 15.8a + 1129b_1 + 20.74b_2$$

Applying the principle of elimination and substitution for any of the two equations, the final answers will appear thus;

$$a = 1.7$$

$$b_1 = 0.78$$

$$b_2 = 23.92$$

Table 6. Computation of the variables for multiple regressions.

Y	X ₁	X ₂	X ₁ ²	X ₂ ²	X ₁ X ₂	X ₁ Y	X ₂ Y
55	50	0.8	2500	0.64	40.0	2750	44.0
60	45	0.9	2025	0.81	40.5	2700	54.0
70	80	0.3	6400	0.09	24.0	5600	21.0
75	65	1.0	4225	1.00	65.0	4875	75.0
90	70	1.2	4900	1.44	84.0	6300	108.0
95	75	1.3	5625	1.69	97.5	7125	123.5
80	80	1.1	6400	1.21	88.0	6400	88.0
90	85	0.5	7225	0.25	42.5	7650	45.0
65	70	1.0	4900	1.00	70.0	4550	65.0
70	60	0.4	3600	0.16	24.0	4200	28.0
80	55	1.6	3025	2.56	88.0	4400	128.0
60	75	0.3	5625	0.09	22.5	4500	18.0
120	90	2.0	8100	4.00	180.0	10800	240.0
105	75	1.8	5625	3.24	135.0	7875	189.0
110	80	1.6	6400	2.56	128.0	8800	176.0
Σ1225	1055	15.8	76575	20.74	1129.0	88525	1402.5

From the aforementioned answers, the estimated linear multiple regression function will appear thus;

$$Y = 1.7 + 0.78X_1 + 23.92X_2$$

Decision rule using multiple regression analysis

There are two major decision areas in the use of multiple regression analysis namely to predict further or expected outcomes and to evaluate the strength of the relationship.

i. Using the aforementioned illustration, to predict a salesgirl's performance given that her aptitude test score is 50 and effort index is 0.6. Computing the expected sales level will easily be solved thus:

$$Y = 1.7 + 0.78(50) + 23.92(0.6) = 55.052 \text{ i.e. N55, 052 sales.}$$

ii. Evaluation of the strength of the relationship means that the coefficient of determination has to be computed. The formula of coefficient of determination according to Keller (2008) is:

$$R^2_{y.12} = \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - Y)^2} \text{ or } R^2 = 1 - \frac{SSE}{\sum(y_i - \bar{y})^2}$$

$$\text{And for small samples } R^2 = 1 - \frac{SSE/(n-k-1)}{\sum(Y_i - Y)^2/(n-1)} \text{ or}$$

$$1 - \frac{MSE}{S_y^2} \text{ where } k \text{ is the number of independent}$$

variables, and sum of squares for errors (SSE) according to Shenoy and Pant (2007) is ascertained by the formula

$$SSE = \sqrt{\frac{\sum(Y^2 - a\sum Y - b_1\sum X_1Y - b_2\sum X_2Y)}{n-k-1}}$$

Substituting will appear thus;

$$\sqrt{\frac{105425 - (1.7)(1225) - (0.78)(88525) - (23.92)(1402.5)}{15-2-1}} = 7.88$$

$$\text{Therefore } R^2 = 1 - \frac{(7.88)^2}{358.889} = 0.827$$

The interpretation of the result shows that 82.7% variation in Y (Sales girl's Performance) has been accounted for by the relationship between aptitude test score (X₁) and the sales effort (X₂)

Using computer SPSS to compute multiple regression analysis

Step 1: Code the variable in the variable view, each variable with its own column.

Step 2: The records are entered on each row according to events.

Step 3: In the regression menu, select: analyze then regression and linear, choosing the dependent variable and independent variables.

Using the aforementioned illustration, the result obtained are shown in Appendix 7, 8, 9 10 and 11.

Interpretation of the SPSS computed results

The descriptive statistics result showed the means, while the correlation shows three things. First, it shows the value of the Pearson correlation coefficient between aptitude test and effort index variables bears to sales performance. It is very clear that effort index had a large positive correlation with sales performance. Secondly, the one-tailed significance of each correlation is displayed in the correlation's result, which shows that effort index correlation is very significant since $p = 0.001$. Thirdly, it shows the number of cases contributing to each correlation that is $N = 15$.

The model summary results tell us whether model is successful in predicting sales performance. The result shows that there are two models. Model 1 is only when aptitude test is used as a predictor. Model 2 refers to when the two predictors were used. The adjusted value of R shows that aptitude test alone accounts for 66.2% of the variation in sales performance while the combined predictors accounts for 92.6% of the variations in sales performance.

The difference between R^2 and adjusted R^2 (0.857 – 0.833) is very small. It is approximately 2.4%, which means that if the model were derived from the Population rather than the sample, it would account for 2.4% less variance in the outcome.

The coefficients result is the most important under our consideration. It shows the variables for predicting future sales performance. The equation is construed thus;

$$\text{Sales performance} = -1.05 + 0.82(\text{aptitude test}) + 23.83(\text{effort Index})$$

Using the assumption of manual computation of 50 and 0.6 for aptitude test and effort index respectively, a sales performance of (N54, 250) is predicted thus; $-1.05 + 0.82 * 50 + 23.83 * 0.6$. The result approximates the manual computed result of N55, 052 sales performances. Looking at second coefficients result, the average VIP is not substantially greater than 1. It is 1.033; therefore, this confirms that collinearity is not a problem for this model.

t-test

Definition and explanation

The t-test is used to compare the values of the means

from two samples and test whether it is likely that the samples are from same populations having different mean values (University of the West of England, 2010).

When two samples are taken from the same population it is very unlikely that the means of the two samples will be identical. When two samples are taken from two populations with very different means values, it is likely that the means of the two samples will differ. The problem is how to differentiate between these two situations using only the data from the two samples. Research (02/07/2010) defined t-test as; 'The t-test assesses whether the means of two groups are statistically different from each other. This analysis is appropriate whenever you want to compare the means of two groups'. This statistical technique is very useful in cost accounting in variance analysis. It is also useful in finding out whether sales promotion exercise did improve sales at a given level of significance. According to McClave et al. ((2005), t-test was discovered by William Gosset in 1908. Another name for t-test is student test. By the works of Bluman (2007), where there exists population mean and sample mean can be inferred, the formula for t-test is given thus:

$$t = \frac{\bar{X} - \mu}{S / \sqrt{n}}, \text{ where } \bar{X} \text{ is the sample mean}$$

Where:

μ = The population mean, S = Standard error of differences between the means and N = Number of samples under consideration.

This formula is best suited for situations where samples are drawn from the sample population. In a situation where the samples come from different populations, Adefila (2008) stipulated the formula to be;

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n(n-1)}}$$

Where:

d = differences between sample 1 and sample 2.

\bar{X} and \bar{Y} are the mean of sample X and that of Y respectively.

Field (2005) agreed with both Bluman (2007) and Adefila (2008) formula but went further to explain that the first formula will be used when 'dependent means t-test' is carried out with a large population and the latter in a paired, repeated or controlled test with a small population.

Table 7. Showing the participants' performances under each condition.

Man code	RER(%)	
	Placebo	Caffeine
1	105	96
2	119	99
3	100	94
4	97	89
5	96	96
6	101	93
7	94	88
8	95	105
9	98	88
Mean	100.56	94.23
SD	7.70	5.61

Dependent means t-test is when there are two experimental conditions and the same participants took part in both conditions of the experiment. Independent t-test is used when there are two experimental conditions and different participants were assigned to each condition. Field (2005) also stipulated formula for small population in computing independent samples of equal sizes thus;

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{N_1} + \frac{S_Y^2}{N_2}}}$$

Field (2005) went further to assert that, where there exist two groups that contain different numbers of participants (that is sizes), that the above formula is not appropriate. The right formula to use is

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$$

$$\text{Given that } S_p^2 = \frac{(n_1 - 1)s_x^2 + (n_2 - 1)s_y^2}{n_1 + n_2 - 2}$$

Worked example:

Illustration 6: A study of the effect of caffeine on muscle metabolism used male volunteers who each underwent arm exercise tests. Nine of the men were randomly selected to take a capsule containing pure caffeine one hour before the test. The subject's respiratory exchange ratio (RER) was measured before and after the caffeine intakes. The question of interest to the experimenter was whether, on average, caffeine changes RER significantly at 95% level of confidence. The results are as shown

Table 7.

Manual solution 6

1. Given that this is a controlled experiment, paired test or dependent test, with small samples therefore the formula that will be appropriate is

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n(n-1)}}}$$

This is tabulated in Table 8. Substituting for the formula will be;

$$t = \frac{100.56 - 94.23}{\sqrt{\frac{881 - \frac{57^2}{9}}{9(9-1)}}} = \frac{6.33}{\sqrt{\frac{520}{72}}}$$

Therefore $t = 2.36$

The decision rule in t-test

The computed value of t-test is compared with the critical table value. Where the computed value is greater than or lower than the critical table, the null hypothesis is rejected, but where they are equal the null hypothesis is accepted. According to Shenoy and Pant (2007), Null hypothesis for t-test is to ascertain the existence or otherwise of equality of means between the samples mean and the population mean. Thus t-test null hypothesis is stated this way;

$H_0: \mu_1 = \mu_2$ (or equivalently the difference between the two

Table 8. Computation of variables for t-test.

Man code	X ₁ (Placebo)	X ₂ (Caffeine)	d	d ²
1	105	96	9	81
2	119	99	20	400
3	100	94	6	36
4	97	89	8	64
5	96	96	0	0
6	101	93	8	64
7	94	88	6	36
8	95	105	-10	100
9	98	88	10	100
Total	905	848	57	881
Means	100.56	94.23		

population means is zero)

Where; μ_1 and μ_2 are the means of first and the second populations respectively. Alternative hypothesis (H_1) could be stated thus;

$H_1: \mu_1 \neq \mu_2$ (two tailed test) or $H_1: \mu_1 > \mu_2$ or $H_1: \mu_1 < \mu_2$ for one tailed test.

To test the null hypothesis that men that take caffeine do not perform significantly better than those without it (Placebo) at $\alpha = 0.05$, in the above illustration, the study compare the computed outcome with the critical table value. The critical table value at 8 degrees of freedom is 2.31, while the computed value is 2.31. The difference of 0.05 is not significant. The study therefore, accepts the null hypothesis which states that caffeine has no significant effect on men's performance.

Where the two samples are drawn from different populations

In this type of situation, different samples (in the example men) are used for each occasion. Using the above illustration, it will be assumed that the two populations being compared are "men who did not take caffeine" and "men who took the caffeine". The null hypothesis is that, men with caffeine intakes do not significantly perform better those with no caffeine at 95% level of confidence.

In this situation, the equation more suitable for the computation of t-test is

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{N_1} + \frac{S_Y^2}{N_2}}}$$

Where, \bar{X} = is the mean of the control variable (men

placebo) that is men without caffeine intakes; \bar{Y} = the mean of the sample or experiment variable (men with caffeine intakes); S = Standard error of differences between the means; and N= Number of samples under consideration; Substituting for the formula will be

$$t = \frac{100.56 - 94.23}{\sqrt{\frac{7.7}{9} + \frac{5.61}{9}}} = 5.1$$

In this situation since the computed value of 5.1 is significantly different from the critical table value of 2.31, the study reject the null hypothesis which states that the performance of men with caffeine intakes are not significantly different from those without caffeine intakes.

Using computer SPSS to compute independent t-test

The steps involved according to University of the West of England (02/072010) are to enter the data from both samples into one column and the group identity in a second column, then select analyze > compare means > independent samples t-test. The data on entry is shown in Appendix 12. The result is given in Appendix 12 and 13.

INTERPRETATION OF COMPUTER SPSS RESULTS

The group statistics shows the same result with manually computed group means (100.56 and 94.22) and standard deviations (7.69 and 7.61) of X_1 , X_2 respectively. In addition, it shows the standard errors, which may be computed manually thus; Std/\sqrt{N} .

The output of independent sample test shows that Levene's test is >0.05, therefore, the study reject the null hypothesis which states that the performance of men with caffeine intakes are not significantly different from those

without caffeine intakes.

The t-test showed two tailed test but one tailed result will be $p = 0.032$ which is < 0.05 . This tells us that there are significant differences between performances of men with caffeine intakes and those without caffeine intakes. In conclusion, the SPSS result tells us that the value of $t = 1.995$ at 16 degrees of freedom and it is significant at $p < 0.05$.

Using computer SPSS to compute dependent t-test

In this situation, each variable and their values are keyed in under a different column. Using analyze > compare means > paired samples t-test. The output is shown in Appendix 14, 15 and 16.

Interpretation of the dependent t-test output

Paired samples statistics result shows the mean and the standard deviations of the two groups. The result is not different from that of the independent t-test. Paired samples correlations result shows the Pearson correlation of the two groups, are positively correlated. On appraisal of Paired samples test result, the study can conclude that exposure to caffeine intakes do not significantly affect men's performance at 8degrees of freedom, $t = 2.36$, $p < 0.05$.

Chi-square

In social sciences, accounting inclusive, there are issues which are either counted as they occur or ranked in degree of occurrences. The population of this kind of data is not known but estimates of their occurrences may be inferred. This type of data is called non-parametric data.

Chi-square test (χ^2) is one of the statistical tools for testing non-parametric data. Adefila (2008) asserted that χ^2 applies only to discrete data, counted rather than measured values. According to Field (2005) χ^2 was invented by Pearson. Its main focus is to compare the frequencies observed in certain events to the frequencies expected to get in those events by chance. In the words of Shenoy and Pant (2007), χ^2 is a test of goodness-of-fit type used to determine whether a significant difference exists between the observed number and the expected ones, and how close are they if they are not equal. Field (2005) asserted that Chi-square statistical tool should not be used to test repeated measures or outcomes and should be conveniently applied where expected outcomes are greater than five. Adefila (2008), Bluman (2007) and Shenoy and Pant (2007) all wrote that the formula for χ^2 is

$$X^2 = \sum \frac{(O - E)^2}{E}$$

Where:

O = Observed frequencies and E = expected frequencies.

Decision rule in Chi-square

Adefila (2008), Bluman (2007) and Shenoy and Pant (2007) all wrote that the Null hypothesis and alternative hypothesis for χ^2 is stated thus:

H_0 = There is no significant difference between observed data and expected data in the transaction or event.

H_1 = There is a significant difference between observed data and expected data in the transaction or event.

Decision rule is to compare the computed value with the critical value of χ^2 at α and K-1. Where the calculated or computed value is greater than the critical value, the null hypothesis will be rejected, while the alternative hypothesis will be accepted.

Bluman (2007) stipulated the following steps for addressing chi-square problems;

Step 1: State the hypothesis and identify the claim.

Step 2: Find the critical value at degrees of freedom (K - 1) and given alpha.

Step 3: Compute the test value using the formula

$$X^2 = \sum \frac{(O - E)^2}{E}$$

Step 4: Make the decision and summarize the results.

Illustration of Chi-square using manual computation

Illustration 7: A Company called Ihena and Sons produces Luco-Sport, Energy-Drink, Orange-Min, Lime-Min and Grape-Juice. The company wants to test a consumption pattern of their drinks on Sundays in Unimaid. The company had already established a demand pattern for the drinks on Fridays in Unimaid as follows; Luco-Sport = 32 cartons, energy-drink = 28 cartons, orange-Min = 16crates, lime-Min = 14 crates and grape-Juice = 10 cartons. The company estimates sales on Sundays to be as follows; Luco-Sport = 20 cartons, energy-drink = 20 cartons, orange-min = 20crates, lime-min = 20 crates and grape-juice = 20 cartons.

The sales manager wants to find whether there are significant brand preferences in the sales during the two worship days and brand preferences at 95% of assurance.

Solution to illustration 7:

Step 1: State the hypothesis and identify the claim

Table 9. Chi-square computation table

O (established)	E (estimates)	O - E	(O - E) ²	$\frac{(O - E)^2}{E}$
32	20	12	144	7.2
28	20	8	64	3.2
16	20	-4	16	0.8
14	20	-6	36	1.8
10	20	-10	100	5
χ^2				18

H₀: There are no significant brand preferences during the two days.

H₁: Consumers show brand preferences.

Step 2: Find the critical value at degrees of freedom (K - 1) and given alpha.

The degrees of freedom (5-1) = 4 and $\alpha = 0.05$. CV= 9.49

Step 3: Compute the test value using the formula

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Step 4: Make the decision and summarize the results.

Since the computed χ^2 that is 18 > 9.49, the study rejects the null hypothesis which states that there are no significant brand preferences during the two days (Table 9).

Using computer SPSS to calculate chi-square

McClave et al. (2005) are of the view that to use SPSS to compute for the Chi-square, the variables will be coded and same group entered into one column. The next step is to choose; analyze \Rightarrow descriptive statistics \Rightarrow crosstabs (For a two way tests). Following the above, illustration 7 data will code 0 for Fridays, 1 for Sundays and 2 - 6 for the drinks. The data is shown in Appendix 17. The SPSS output for the Chi-square test on sales of Fridays and Sunday's expectations is given in Appendix 19.

Interpretation of SPSS result

The Chi-square test examines whether there is an association between Friday's sales and expected Sunday sales. Since Pearson's Chi-square test shows that p .075

>0.05, the study reject the null hypothesis which states that there are no differences between Friday sales and expected Sunday sales. Looking at output, the F-ratio of 9.49 is equal to the critical value but chi-square value differed a lot. This is because of the constant expected sales. However, if the hypothesis is to demand preferences between the drinks on Fridays and Sunday's expectations, the SPSS output is shown in Appendix 20.

The result once again leads to the rejection of the null hypothesis with a chi-square value of 17.5 against the computed 18. The later result equates the stated null hypothesis of illustration 7.

Analysis of variance (ANOVA)

ANOVA is a statistical method of comparing the means of several populations (Adefila, 2008). Levine et al. (2005) explained that ANOVA is used to compare the means of the groups and it is an extension of the t-test which compares the means of two populations only.

ANOVA, also called F-test in honor of Ronal A. Fisher who developed the statistical method is employed to make two estimates. The first estimate according to Lind et al. (2006) involves finding the variance of the means. This is also called the 'between-group' variance. The second estimate, that is the 'within-group' variance is made by computing the variance using all the data and is not affected by differences in the means. According to Bluman (2007), the assumptions of F-test are;

1. The populations from which they samples were obtained must be normally or approximately normally distributed
2. The Samples must be independent of one another
3. The variances of the populations must be equal.

Lind et al.(2006) had earlier made the aforeassumption stipulations of Bluman (2007) and went further to assert that the hypothesis for ANOVA are stated thus;

$$H_0: \mu_1 = \mu_2 = \dots \mu_n$$

H₁: At least one mean is different from the others.

Decision rule in analysis of variance (ANOVA)

Both McClave et al. (2005) and Bluman (2007) stated that the decision rules in ANOVA are;

1. If there is no difference in the means, the between-group variance estimate will be approximately equal to the within-group variance estimate and the F-test will be approximately equal to 1. The null hypothesis will be accepted.

2. When the means differ significantly, the between-group variance will be much larger than the within-group variance and the F test value will be significantly greater than

1. The null hypothesis will be rejected. The bases of acceptance or rejection of the null hypothesis will be comparing the computed value of F with the critical value of F. Keller (2009) wrote that, to access the critical values of F, the degrees of freedom are ascertained thus;

- a. Between-group $k - 1$ (where k = No of groups)
- b. Within-group $N - k$ (where N = Sum of sample sizes for groups)

How analysis of variance (ANOVA) operates and applied

The steps to follow in the application and computation of ANOVA are vividly laid out by Bluman (2007) as follows;

Step 1: State the hypothesis and identify the claim

$$H_0: \mu_1 = \mu_2 = \dots \mu_n$$

H_1 : At least one mean is different from the others.

Step 2: Find the critical values of the numerator and denominator as follows $k - 1$ (where k = No of groups) and $N - k$ (where N = Sum of sample sizes for groups) respectively.

Step3: Compute the test value as follows;

- a. Find the mean and variance of each Group
- b. Find the Grand mean of the samples in all the Groups.
- c. Find the between-group variance.
- d. Find the within-group variance
- e. Find the F test value.

Step 4: Make the decision.

Step 5: Summarize the results.

Though the listed steps are very simple but the problem lies with the computation of each subdivisions of step 3. The detailed formula for computing each is as follows;

a. i. Mean of each Group $(\bar{X}) = \frac{\sum X}{N}$, ii. Variance = S^2

b. Grand/Total mean = $(\bar{X}_1 + \bar{X}_2 + \bar{X}_n)/k$, where k is the number of groups

c. The between-group variance (Mean of Squares = SS_M) = S_B^2

$$SS_M = \frac{\sum n_i (\bar{X}_i - \bar{G})^2}{K - 1}$$

where n_1 denotes each group size, \bar{G} = Grand mean
 K = Group size, d. The within-group Variance (S_W^2)

$$S_W^2 = \frac{\sum (n_i - 1) S_i^2}{\sum (n_i - 1)}$$

$$e. F = \frac{S_B^2}{S_W^2}$$

Worked illustration of analysis of variance (ANOVA)

Illustration 8: Given that a company sent out three groups of their sales persons to Bama, Yola and Gobi for sales promotion. The following sales were made on the five days promotion duration (Appendix 21). You are required to determine the claim of the sales manager that there are no difference among means of sales recorded at towns that witnessed the sales promotions at $\alpha = 0.05$. Manual solution to illustration 8:

Step 1: State the hypothesis

$$H_0: \bar{b} = \bar{y} = \bar{g}$$

H_a : At least one mean is different from the others.

Step 2: Find the critical value $k = 3 - 1 = 2$ and $N = 15 - 3 = 12$ at $\alpha = 0.05$. The critical table value of $F = 3.89$.

Step 3: Compute the test values:

$$\bar{b} = (\text{Bama}) (\sum 10 + 12 + 9 + 15 + 13) / 5 = 11.8, S_b^2 = 5.7$$

$$\bar{y} = (\text{Yola}) \frac{\sum 6 + 8 + 3 + 0 + 2}{5} = 3.8, S_y^2 = 10.2$$

$$\bar{g} = (\text{Gobe}) \frac{\sum 5 + 9 + 12 + 8 + 4}{5} = 7.6, S_g^2 = 10.3$$

$$Gm = (\text{Grand mean}) = (11.8 + 3.8 + 7.6)/3 = 7.73$$

$$S_b^2 = (\text{Between-group variance})$$

$$\frac{5(11.8 - 7.73)^2 + 5(3.8 - 7.73)^2 + 5(7.6 - 7.73)^2}{3 - 1} = 80.07$$

$$S_w^2 = (\text{Within-group variance})$$

$$\frac{(5-1)(5.7) + (5-1)(10.2) + (5-1)(10.3)}{(5-1) + (5-1) + (5-1)} = 8.73$$

$$\text{Step 4: Find the F test value} = \left(\frac{S_b^2}{S_w^2} \right) = \frac{80.07}{8.73} = 9.17$$

Decision: Reject the null hypothesis since computed F test 9.17 > 3.89 table value. This gives us enough evidence to reject the claim and conclude that at least one mean is different from others.

Using computer SPSS to compute f test value

Step 1: At the variable view create the three columns of; days, locations and sales but tagging days as 'string'.

Step 2: Enter the sales at the date view. The data is shown in (Appendix 22).

Step 3: To conduct one-way ANOVA, at the dialogue box select: Analyze \Rightarrow Compare Means \Rightarrow One-way ANOVA. The results obtained are shown in Appendix 23, 24, 25 and 26.

Interpretation of SPSS computer analysis of variance (ANOVA) results

From descriptive statistics output, the group means and standard deviations and that of the grand mean correspond to the manually computed means and standard deviations. However, in addition, the standard errors were computed. The confidence interval tells us that the true value of the mean for Bama sales is between 8.84 and 14.76 and a standard error of 1.07. Similar results were also given for Yola and Gobi.

The ANOVA output shows between and within group means to be same as manually computed. The test of whether the group means are the same is represented by the F-ratio for the combined between-group effect. The value of this ratio is 9.168 which are the same as manually calculated. Finally, the result tells us whether this value is likely to have happened by chance. The significance level of 0.004 shows that by 0.4% probability, that the result would have occurred by chance events. Thus the observed significance value is less than 0.05

and the study shall conclude that there is a significant effect on sales by the sales promotion. However, was there significant effect of the sales promotion on all the locations? To answer this question the study shall look at the quadratic term contrast result.

The result shows that the F-ratio is 1.3 and very significant since 0.003 is less than 0.05. Finally, the between group contrasts are shown in multiple comparisons' output. Is the above result contrasting the outcome of manually computed result? The answer is no. In the manual computation, the study tested the null hypothesis that the group means of the effect of the sales' promotions are equal. The computer result that tests the null hypothesis is Levene's test shown in test of homogeneity of variances. Levene's test is designed to test the null hypothesis that the variances of the groups are the same. The result shows F-ratio of 0.384 and significant level of 0.689. It means that the null hypothesis had been violated and should be rejected as p 0.689 is > 0.05. The final conclusion is same with the manual computation.

CONCLUSION AND RECOMMENDATIONS

From the previous analyses, the study conclude that each of the selected statistical tools of correlation, regression analysis, t-test, Chi-square and ANOVA has unique areas of relevance to accounting and business in general. The study hereby recommends each as aid to decision to the accountants, financial analysts, investors and business owners.

REFERENCES

- Adefila JJ (2008). Research Methodology in Behavioural Sciences Apani Publications, Kaduna Nigeria
- Bluman AG (2007). Elementary Statistics (A Step by Step Approach) McGraw-Hill Companies, USA.
- Buglear J (2005). Quantitative Methods for Business (The A-Z of QM) Elsevier Butterworth-Heinemann Burlington, New York.
- Bull RJ (1981): Accounting in Business, 5th ed. England: The Garden City Press Ltd; Letch worth Hert p.4.
- Field A (2005). Discovering Statistics Using SPSS Sage Publications, London.
- Gibson HC (2009). Financial Reporting and Analysis using Financial Accounting Information 11th edition. Thomson, South-Western.
- Keller G (2008). Managerial Statistics 8ed South Western CENGAGE Learning, UK.
- Kieso DE, Weygant JJ, Warfield TD (2007). Intermediate Accounting 12th edition John Wiley and Sons, New York. p 127, 171.
- Levine DM, Stephen D, Krehbiel TC, Berenson ML (2005). Statistics for Managers Using Microsoft Excel 4ed Pearson Education, Inc, New Jersey.
- Lind DA, Marchal WG, Wathen SA (2006). Basic Statistics for Business, Economics 5ed. McGraw-Hill/Irwin New York.
- Maddala GS (2001): Introduction to Econometrics 3rd ed. John Wiley and Sons Ltd, New York.
- Messier WF, Gover S, Prawitt D (2008). Auditing and Assurance Services 6ed McGraw-Hill/Irwin New York
- McClave JT, Benson GP, Sincich T (2005). Statistics for Business and

Economics 9th ed Pearson Education, Inc, New Jersey.
Shenoy GV, Pant M (2007). Statistical Methods in Business and Social Sciences
Macmillan India Ltd, Mumbai India Websites: Experiment-resources.com (17/06/2010). Statistical Correlation
<http://www.experiment-resources.com/statistical-correlation.html#ixzz0r6Gjtwxr>
Pinkmonkey.com (17/06/2010). Types of Correlation
<http://www.pinkmonkey.com/studyguides/subjects/stats/chap6/s0606301.asp>
Research Methods (02/07/2010). The t-test
http://www.socialresearchmethods.net/kb/stat_t.php.
Statisticsolutions.com (21/06/2010). Pearson Correlation
<http://www.statisticssolutions.com/methods-chapter/statistical-tests/correlation-pearson-kendall-spearman/>

Surveysystem.com (16/06/2010). Correlation
<http://www.surveysystem.com/correlation.htm>.
University of the West of England (02/072010). Independent Samples t-test
<http://hsc.uwe.ac.uk/dataanalysis/quantInfDifIndy.asp>
Wikipedia.org (21/06/2010).
http://en.wikipedia.org/wiki/Kendall_tau_rank_correlation_coefficient.