QUESTION 1

a) Calculation of hypothesis testing. Given information:

Male students mean (x1) = 83 Sample size (x1) = 160 Standard deviation (S1) = 5

Female students mean (x2) = 84 Sample size (x2) = 272 Standard deviation (S1) = 7

Level of significance = 1% or 99% confidence

Critical values of z values are -2.58 and 2.58, therefore critical region z ≤ -2.58 or 2.58 ≥

Null hypothesis (Ho): U1 – U2 = 0, that is to say, Sarah's sample claim was drawn from same population.

Alternative Hypothesis (H1): u1 – u2 ≠ 0, which means, Sarah's claim was not drawn from same population.

The standard error of difference between the two means is given as:

$$\partial(x1 - x2) = \sqrt{\frac{S^2 1}{n1} + \frac{S^2 2}{n2}}$$

$$\sqrt{\frac{5^2}{160} + \frac{7^2}{272}}$$

$$\sqrt{0.15625 + 18015}$$

$$\sqrt{0.33} = 0.574$$

$$Z = \frac{x1 - x2}{\partial\ (x1 - x2)} = \frac{83 - 84}{0.574} = -1.74$$

Therefore Ho is accepted since the calculated Z value falls or lies within the critical region i.e. -1.74 is greater than -2.58 hence Sarah's claim is right or true.

b) Additional assumptions required to validate the hypothesis test above.

There are usually two assumptions

i) Sample analyzed using large sample, meaning that sample selected is assumed to be more than 30 items to be counted. Where the sample is selected the hypothesis more

ii) Sample analyzed using small sample, meaning that sample selected had equal likely chances of unselected ones. When they are drawn each element must be represented in an equal way.

QUESTION 2

a) Computing least square regression

X	Y	X^2	Y^2	XY
336	415	112896	172225	139440
14	21	196	441	294
35	38	1225	1444	1330
80	95	6400	9025	7600
113	152	12769	23104	17176
146	174	21316	30276	25404
130	144	16900	20736	18720
288	321	82944	103041	92448
496	541	246016	292681	268336
173	203	29929	41209	35119
26	39	676	1521	1014

$\sum X$=1,837 $\sum Y$= 2143 $\sum X^2$=531,267 $\sum Y^2$= 695,703 $\sum XY$=606,881

Regression equation of X on Y is given as

X = a + bY

The normal equations for regression are

$\sum X = Na + b\sum Y$

$\sum XY = a\sum Y + b\sum Y^2$

Substituting with the above provided values, we get

1,837 = 11a + 2,143b……(i)

606,881 = 2,143a + 695,703b…(ii)

Solving equations (i) and (ii) using elimination method, we multiply 194.8 with (i) equation above to make values of (a) equal, we get by subtracting equations (i) with (ii)

357,848 = 2,143a + 417,456b

606,881 = 2,143a + 695,703b

249,033 = 0a + 278,247b

Value of b = 278,247/249,033 = 1.2

Substituting b values in equation (i) will be

1,837 = 11a + (2,143 x 1.2)

11a = 2572 – 1,837, a = 735/11 = 66.8

Therefore the required regression equation of X on Y is gives as

X = 66.8 + 1.2Y

Regression equation of Y on X is calculated as: Y = a + bX

And normal equations are:

$\sum Y = Na + b\sum X$

$\sum XY = a\sum X + b\sum X^2$

We substitute the figures with the formula

2,143 = 11a + 1,837b…….(i)

606,881 = 1,837a + 531,267b….(ii)

Solving equations (i) and (ii) using elimination method, we multiply 167 with (i) equation above to make values of (a) equal, we get by subtracting equations (i) with (ii).

357,881 = 1,837a + 306,779

606,881 = 1,837a + 531,267b

-249,000 = 0 + 224,488b

Therefore b = 1.1

Regression equation of Y on X is calculated as: Y = a + bX

And normal equations are:

$\sum Y = Na + b\sum X$

$\sum XY = a\sum X + b\sum X^2$

We substitute the figures with the formula

2,143 = 11a + 1,837b…….(i)

606,881 = 1,837a + 531,267b….(ii)

Solving equations (i) and (ii) using elimination method, we multiply 167 with (i) equation above to make values of (a) equal, we get by subtracting equations (i) with (ii).

357,881 = 1,837a + 306,779

<u>606,881 = 1,837a + 531,267b</u>

-249,000 = 0 + 224,488b

Therefore b = 1.1

Substituting b values in equation (i) will be

2,143 = 11a + (1,837 x 1.1)

11a = 2021 – 2,143 = -122

Therefore, a = -122/11 = 11

Hence, the required regression equation of X on Y is gives as

Y = 11 + 1.1X

b) If computer generated balance was 100, then the verified value confirmed by accountant will be

Y =11 + 1.1(100) = 121

c) Calculation of the 95% interval estimate will be

P = 100/121 = 0.8

Q= 1-p (1 – 0.8) = 0.2

N= 121

$\partial p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.8\ x\ 0.2)}{121}} = 0.16$

Population proportion = p± 1.96 (∂p)

0.8 ± 1.96 (0.16)

0.8 ± 0.3

0.5 to 1.1 which is between 50% to 110%

d) Calculating p- values and interpret it

Therefore Ho is accepted since the calculated Z value falls or lies within the critical region i.e. -1.1 is less than +2.58 hence hypothesis in this case is strongly supported right or true because it falls within the accepted region.

QUESTION 3

	X	Y	X^2
1	2005	329670	4020025
2	2136	332980	4562496
3	1897	326810	3530641
4	2245	361240	5040025
5	2053	347380	4214809
6	1920	309990	3686400
7	2989	453420	8934121
8	2952	444960	8714304
9	2353	372900	5536609
10	1860	302840	3459600
11	2028	320600	4112784
12	2040	328800	4161600
13	2029	341550	4116841
14	2233	366630	4986289
15	1932	305920	3732624
16	1872	303720	3504384
17	2653	422510	7038409
18	2005	337920	4020025
19	2509	394020	6295081
20	1993	333410	3972049
	69631	**7037270**	**97639116**

a) Regression equation of X on Y is given as

X = a + bY

The normal equations for regression are

$\sum X = Na + b\sum Y$

$\sum XY = a\sum Y + b\sum Y^2$

Substituting with the above provided values, we get

43,704 = 20 a + 7037270 b……(i)

606,881,675 = 7037270 a + 695,703,854b…(ii)

Solving equations (i) and (ii) using elimination method, we multiply 351863.5 with (i) equation above to make values of (a) equal, we get by subtracting equations (i) with (ii)

43704 = 7037270a + 4,468,456b

<u>606,881,675 = 7037270a + 695,703b</u>

606837971 = 0a + 3772753b

Value of b = 606837971 / 3772753 = 16

Substituting b values in equation (i) will be

43704 = 20a + (4,468,456 x 16)

11a = 7149296 – 43704, a = 1392620 / 20 = 69631

Therefore the required regression equation is gives as

Price = 69631-3630 type + 90.4 Square Feet

b) If the condominium is the residence, the price is likely to decrease on average by $3630. With increase of one unit in square feet the price is likely to increase by $100 which causes an average of $90.4

c) Calculation of the 95% interval estimate will be

P-value = 16/20 = 0.822

Q= 1-p (1 – 0.8) = 0.2

N= 20

$\partial p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.8\ x\ 0.2)}{20}} = 0.16$

Population proportion = p± 1.96 (∂p)

0.8 ± 1.96 (0.16)

0.8 ± 0.3

0.5 to 1.1 which is between 50% to 110%

Therefore, the test is equivalent to Ho: β 1 = 0, t value = -0.23 (reading from T-square tables), P-value = 0.822. Since the p-value of 0.822 is greater than level of significance of 0.05, we reject the null hypothesis and we have to conclude that the relationship that is existing between the selling price and the square footage is not different from condominiums and single family homes.

d) i) The condominiums equation from (a) above will give us, Price = 66,001 + 90.4 Square Feet

ii) The single family equation is also given as Price = 69,631 + 90.4 Square Feet

QUESTION 4

a) The calculation of the multiple regression analysis using dummy variables to determine the revenue.

Year	Quarter	Revenue	t	Q1	Q2	Q3
2015	1	492	1	1	0	0
	2	457	2	0	1	0
	3	485	3	0	0	1
	4	578	4	0	0	0
2016	1	554	5	1	0	0
	2	559	6	0	1	0
	3	574	7	0	0	1
	4	630	8	0	0	0
2017	1	587	9	1	0	0
	2	651	10	0	1	0
	3	645	11	0	0	1
	4	728	12	0	0	0
2018	1	701	13	1	0	0
	2	729	14	0	1	0

3	776	15	0	0	0
4	840	16	0	0	0

Regression analysis

Overall fit

Multiple R	0.99089
R Square	0.97526
Adjusted R Square	0.98185
Standard error	0.59375
Observations	16

	Coeff	Std error	t stat
Intercept	14.327	0.4488	30.73
T	0.6524	0.0139	16.59
Q1	-3.9381	0.4367	-8.23
Q2	-5.7375	0.4321	-13.488
Q3	-2.4938	0.4211	-5.922

Explanation

The data shown is for the two independent variables, that is, square footage and corresponding prices for each footage cost. So we are looking at the relationship in the form of

Sales = a + b1x the square feet + b2 x price. Therefore the above calculation shows the value for the interpretation and variables within the line of best fit identified as

Sales = 585.96 + (9.90 x 0.997)

The coefficient of correlation r = 0.997. That means, a very strong linear relationship exist. Then this is confirmed by determination coefficient, r2, which shows that 99.7%.

The adjusted r2 which removes the bias but is only slightly lesser that that calculated value.

b) Calculation of the 95% and 99% interval estimate will be

Lower at 95% significance level

P = 14/20 = 0.7

Q= 1-p (1 – 0.7) = 0.3

N= 16

$\partial p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.7\ x\ 0.3)}{16}} = 0.01$

Population proportion = p± 1.96 (∂p)

0.7 ± 1.96 (0.01)

0.7 ± 0.01

0.69 to 0.71

Upper at 99% confidence

0.7 ± 2.58 (0.01)

0.7 ± 0.02

0.68 to 0.72

c) Let us assume that sales will constantly increase in year 2019 and 2020 consecutively

Ending revenue as of Q4 2018 was 840 with 16 quarters (4 x 4 years).

By end of Q4 of 2019 sales will be: = 840 + (9.90 x 0.997 x 20 quarters)

840 + 197 = 1037.

By end of Q4 of 2019, sales forecast will be = 1037 + (9.90 x 0.997 x 24 quarters)

1037 + 237 = 1274

QUESTION 5

a) Calculating the best combination of asset stock, bonds and commodities in the financial market.

Asset case	Stocks	Bonds	Commodities
Expected Return	100%	0%	0%
	90%	10%	0%
	80%	20%	10%

	70%	30%	20%
	60%	40%	30%
	50%	50%	40%
	40%	60%	50%
	30%	70%	60%
	20%	80%	70%
	10%	90%	80%
	0%	100%	90%
	0%	0%	100%
Standard deviation	100%	0%	0%
	90%	10%	0%
	80%	20%	10%
	70%	30%	20%
	60%	40%	30%
	50%	50%	40%
	40%	60%	50%
	30%	70%	60%
	20%	80%	70%
	10%	90%	80%
	0%	100%	90%
	0%	0%	100%

b)

Combination	Stocks		Bonds		Commodities	
	ER	SD	ER	SD	ER	SD
	100%	100%	0%	0%	0%	0%
	90%	90%	10%	10%	0%	0%

80%	80%	20%	20%	10%	10%
70%	70%	30%	30%	20%	20%
60%	60%	40%	40%	30%	30%
50%	50%	50%	50%	40%	40%
40%	40%	60%	60%	50%	50%
30%	30%	70%	70%	60%	60%
20%	30%	80%	80%	70%	70%
10%	20%	90%	90%	80%	80%
0%	0%	100%	100%	90%	90%
0%	0%	0%	0%	100%	100%

QUESTION 6

Find the multiple linear regression equation of X1 on X2 and X3 from the data relating to three variables given below

X1:	4	6	7	9	13	15
X2:	15	12	8	6	4	8
X3:	30	24	20	14	10	4

Calculate the regression equation above and estimate the probable revenue for the firm.

Solution:

Regression equation of X1 on X2 and X3 is given as

X1 = a + bX2 + bX3

The constants value are a, b1 and b2 are obtained by solving the following three normal equations as given below

$\sum X1 = Na + b\sum X2 + b\sum X3$

$\sum X1X2 = a\sum X3 + b\sum X^2 2 + b\sum X2X3$

$\sum X1X3 = a\sum X3 + b\sum X2X3 + b\sum X^2 3$

Calculating the required values

X1	X2	X3	X1X2	X1X3	X2X3	$X^2 1$	$X^2 2$	$X^2 3$
4	15	30	60	120	450	225	900	16
6	12	24	72	144	288	144	576	36
7	8	20	56	140	160	64	400	49
9	6	10	54	126	84	36	196	81
13	4	10	52	130	40	16	100	169
15	3	4	45	60	12	9	12	225
54	48	102	339	720	1,034	494	2,188	576

Substituting the values in the normal equations, we get

6a + 48b + 102b = 54 ……..(i)

48a + 494b + 1034b = 339 …….(ii)

102a + 1034b + 2188b = 720 ……..(iii)

Multiplying equation (i) by 8, we get

48a + 384b + 816b = 432 ……(iv)

Subtracting equation (ii) from equation (iv), we get

110b + 218b = -93 …….(v)

Multiplying equation (i) by 17, we get

102a + 816b + 1734b = 918 ……..(vi)

Subtracting equation (iii) from equation (iv), we get

218b + 454b = -198 ……...(vii)

Multiplying equation (v) by 109, we obtain

11990b + 23762b = -10137 ……..(viii)

Multiply equation (vii) by 55, we get

11990b + 24070b = -10890 ……..(ix)

Subtracting equation (viii) from equation (ix), we get

1208b = -753

B = -753 / 1208 = 0.623

Substituting this value of b in equation (v), we get

110b + 218 (-o.623) = -93

110b + 135.814 – 93

B = 42.814 / 110 = +0.389

Substituting the value of b3 and b2

110b + 218(-0.623) in equation (i), we get

6a + 48 (0.389) + 102 (-0.623) = 54

6a = 54 + 63.546 – 18.672 = 98.874

A = 16.479

The regression equation is given as

X1 = 16.479 + 0.389X2 – 0.623X3

Definition

Regression is a statistical method that is very important for the analysis in any business sector in today's modern world. This enables the identification and characteristics of the relationships among two or more variables or multiple variables (Duncan, and Klebanov, 1994). This method also helps to identify the relevant business risk factors the businesses face and enable them to calculate the risk scores for an individual business or sector or department of the business. It is also a graphical method that describes the average association among the two or more variable. Some of the methods applied in regression are many and do vary from business to business or sector.

However, when there is either a perfect positive or negative relationship among the variables, two regression lines will coincide, that is to say, we will have one line. The farther two regression lines are from each other, the lesser is the degree of correlation and the nearer the two lines the higher the degree of correlation. The variables are independent, where r is assumed to be zero and regression lines are a right angle, that is, they are parallel to X-axis. Some of the regression models used explains the important considerations that are put before the analysis is done or performed and how the outcome of the analysis done will look like and will be interpreted to fit the business need of the company (Duncan, and Klebanov, 1994). The beneficiaries of the analysis outcomes should be able to charge or gauge as to whether the methods which has been used are good and correct, and interpreted outcome or result are suitable to the needs of the company.

It must be noted that regression lines cuts each other at the point of average, if from the point where both lines cut each other, we get the mean value of and from the point a horizontal lines are drawn. It may not pass through all the points drawn, but it will lie on some place near or

adjacent to drawn points or in the midst of them. Then from there the collection of points are seen and it will slope in the direction as suggested by the said points

Purpose of Regression

The purpose of this statistical estimation of data is over and over again pronounced as the association among two variables or among many or several ones (Christian, and Bryant, 1998). If for example, the company wants to know whether the sales in the coming month or year will increase by a certain percentage, but also whether there will be a likelihood of high volume of sales in the near future, it will be influenced by customer satisfaction and efficient services offered to customers (Christian, and Bryant, 1998). In most cases, regression permits multiple independent variable, with the adjustment of the coefficient to confound the effects among them. It also permits evaluation and determine the optional values.

If the equation "a" and "b" are two unknown constant which determine the position of the line completely. These constants are usually parameters of the line. These parameters determine the level of the best fit, that is, distance of the line directly above or below the origin. Parameter "b" usually determines the slope of the line which shows the change in X over the unit change in Y

The variable can be explained in terms of sales so called dependent variables or sometimes called response variable. The one that gives an explanation of sales are independent variable or forecaster variables. The one that measures the association provides an initial impression to the extent at which the statistical dependence that exists between the variables (Christian, and Bryant, 1998). If the depended and independent variables are of a continuous nature such as sales, there must be a correlation coefficient must be calculated to help in measuring the strength of the relationship that exist among themselves (Elder H, 1974). It must be noted that regression lines cuts each other at the point of average, if from the point where both lines cut each other, we get the mean value of and from the point a horizontal lines are drawn.

Scatter diagram is also used in regression whereby its main purpose is to create a straight-line relationship and a connection between which it supports the suggestion. For example, where r is above 0.4 0r 0.5, Then we say that there is a linear correlation between the two variable and their other errors of measurement and even observations which functions to get us a scatter point long the line as an alternative of exactly on it (Elder H, 1974).

To determine such relationship that exist between x and y, we have to understand what straight line need to be drawn through the meeting points of the scatter diagram (Elder H, 1974). It may not pass through all the points drawn, but it will lie on some place near or adjacent to drawn points or in the midst of them. Then from there the collection of points are seen and it will slope in the direction as suggested by the said points. Then they form a regression line which can forecast the future of the sales of the company.

As we look at company's output on a monthly basis where y is taken to be total monthly costs of the manufacturing firm, the scatter diagram bases its records on last year's records. Line drawn

through this points is clearly the one we think that is of best circumstances and many researchers do refer to regression lines as lines of best fit

Shortcomings of regression

Interpretation and performance of linear regression are topic to a variety of drawbacks which can be debated in details late on. It has both opportunities and limitation of use to the most business but it help the managers of the large companies to make economic decisions for their businesses (Elder H, 1974).

In most cases, regression permits multiple independent variable, with the adjustment of the coefficient to confound the effects among them. It also permits evaluation and determine the optional values. Some of the regression models used explains that have no important considerations which are put before the analysis is done or performed and how the outcome of the analysis done will look like and will be interpreted to fit the business need of the company (Duncan, and Klebanov, 1994).

In addition, if there are many inappropriate variables are included in the model, over adjustment is likely to be the outcome, that is to say, irrelevant variables are found to an effect and have a purely change. Some of the variables are usually have random effects (Elder H, 1974). The inappropriate variable allow better fit with the data that is set under the study but since they have random effect, their outcome will get generally applicable outside the data that is set.

The causation connection

The causation's theory has been discussed by many research and for many years in the past. The concerns that have been discussed are many, however, is aimed at the academic need approach for the anticipated research method. It defines the cause of the event as anything which is perceived to be at some varying degree which is in control for the event which will not be said to exist but in the real sense it merely assumed to exist. Causation involved from ordinary knowledge looking to benefit and create generative methods and variables.

References

- Christian, K., and Bryant, F. B. (1998). Correlational analysis and causal inferences. American Anthropologist 62; 624-631
- Clarke, A. and Kurtz-Costes, B. (1997). Correlation and causality; the multivariate case. Social Forces 39: 246-251.
- Coleman, J. S. (1988). Social Statistics. New York, New York, McGraw-Hill Book Co., Inc. 1960.
- Darling-Fisher, C. S., and Tiedje, B. (1990). Evaluating the relative importance of variables. American Sociological Review 26: 866-874.
- Duncan, G. J., and Brooks-Gunn, J. (Eds.). (1997). Causal inferences in natural experiments; some complications in matching designs. Sociometry 30; 300-315.
- Duncan, G. J., Brooks-Gunn, J., and Klebanov, P. K. (1994). Economic deprivation and early childhood development. Child Development, 65, 296-318.
- Elder, G. H. (1974). Children of the Great Depression: Social change in life experience. Chicago: University of Chicago Press.
- Emery, R. E., and Tuer, M. (1993). Parenting and the marital relationship. In T. Luster and L. Okagaki (Eds.), Parenting: An ecological perspective (pp. 121-148). Hillsdale, NJ: Erlbaum.