Lesson: Quantitative Data Analysis part -I

Research Methodology - COMC/CMOE/ COMT 41543

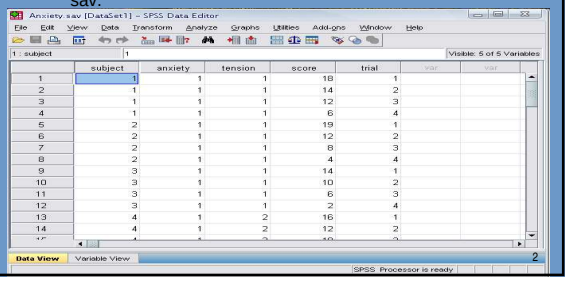
By
Srinath Dissanayake
DCFM

1

The Four Windows: Data Editor

- Data Editor

  Spreadsheet-like system for defining, entering, editing, and displaying data. Extension of the saved file will be "sav."

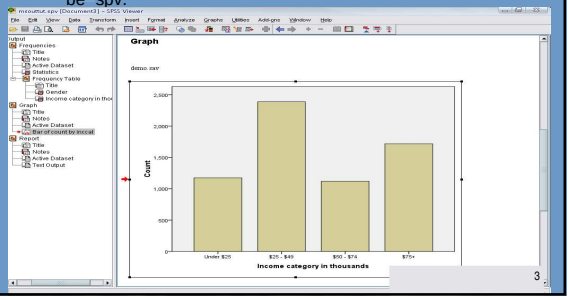


2

The Four Windows: Output Viewer

- Output Viewer

  Displays output and errors. Extension of the saved file will be "spv."

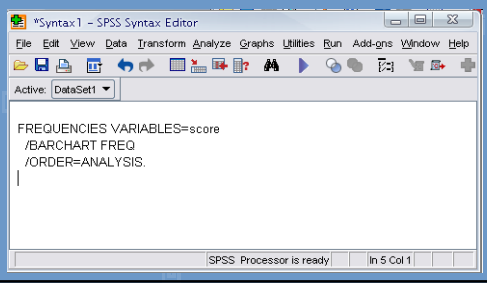


3

The Four Windows: Syntax editor

- Syntax Editor

  Text editor for syntax composition. Extension of the saved file will be "sps."

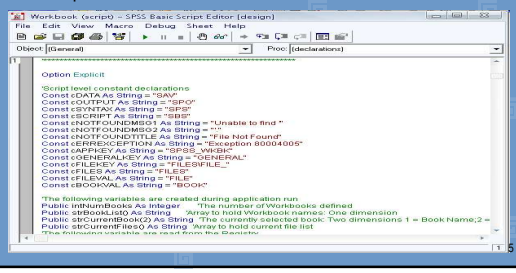


4

The Four Windows: Script Window

- Script Window

  Provides the opportunity to write full-blown programs, in a BASIC-like language. Text editor for syntax composition. Extension of the saved file will be "sbs."

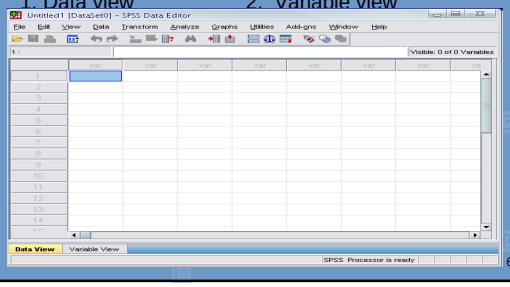


5

Opening SPSS

- The default window will have the data editor
- There are two sheets in the window:

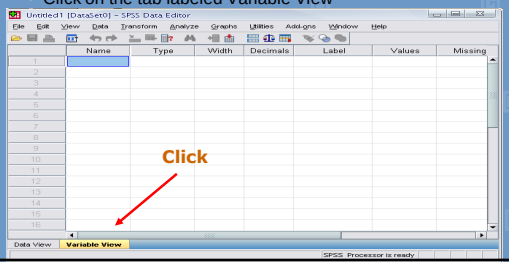
  1. Data view
  2. Variable view



6

Data View window

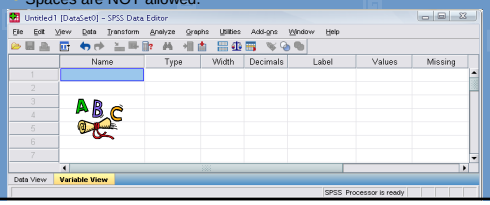
- The Data View window
  This sheet is visible when you first open the Data Editor and this sheet contains the data
- Click on the tab labeled Variable View



7

Variable View window

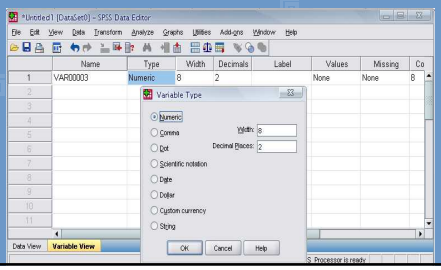
- This sheet contains information about the data set that is stored with the dataset
- Name
  - The first character of the variable name must be alphabetic
  - Variable names must be unique, and have to be less than 64 characters.
  - Spaces are NOT allowed.



8

Variable View window: Type

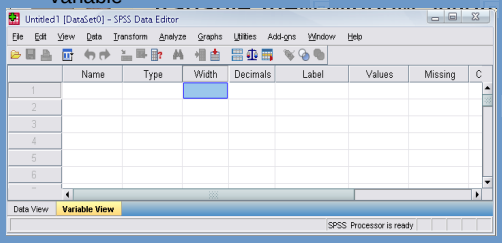
- Type
  - Click on the 'type' box. The two basic types of variables that you will use are numeric and string. This column enables you to specify the type of variable.



9

Variable View window: Width

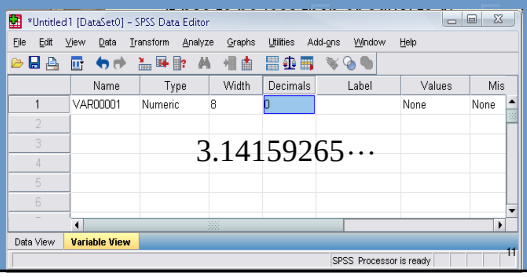
- Width
  - Width allows you to determine the number of characters SPSS will allow to be entered for the variable



10

Variable View window: Decimals

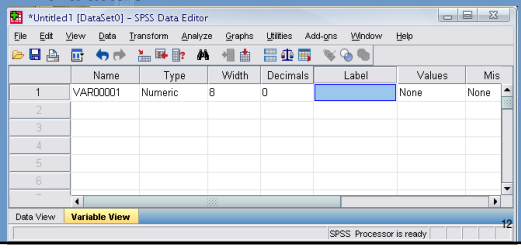
- Decimals
  - Number of decimals
  - It has to be less than or equal to 16



11

Variable View window: Label

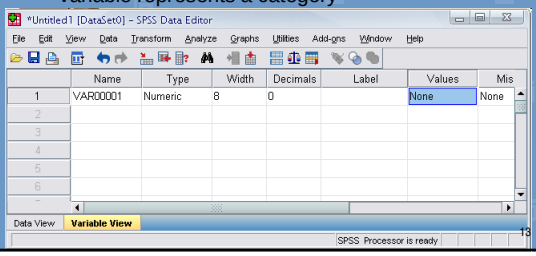
- Label
  - You can specify the details of the variable
  - You can write characters with spaces up to 256 characters



12

Variable View window: Values

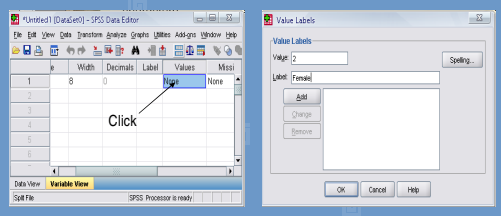
- Values
  - This is used and to suggest which numbers represent which categories when the variable represents a category



13

Defining the value labels

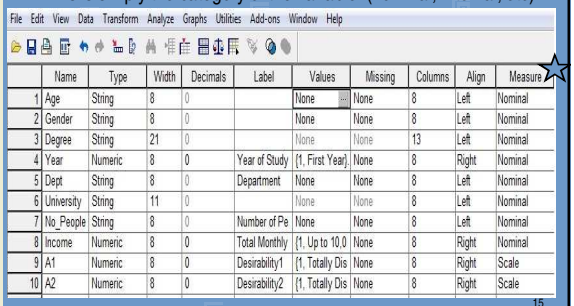
- Click the cell in the values column as shown below
- For the value, and the label, you can put up to 60 characters.
- After defining the values click add and then click OK.



14

Define measures

- This is simply the category of the variable. (nominal, Interval, etc)

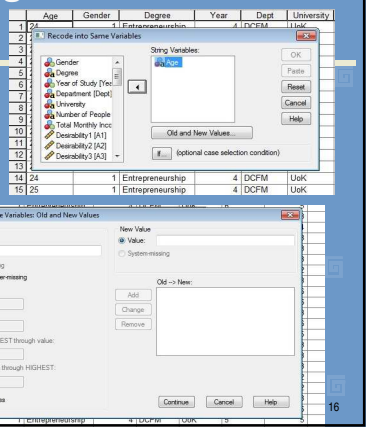


	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	Age	String	8	0		None	None	8	Left	Nominal
2	Gender	String	8	0		None	None	8	Left	Nominal
3	Degree	String	21	0		None	None	13	Left	Nominal
4	Year	Numeric	8	0	Year of Study	{1, First Year}	None	8	Right	Nominal
5	Dept	String	8	0	Department	None	None	8	Left	Nominal
6	University	String	11	0		None	None	8	Left	Nominal
7	No_People	String	8	0	Number of Pe	None	None	8	Left	Nominal
8	Income	Numeric	8	0	Total Monthly	{1, Up to 10,0	None	8	Right	Nominal
9	A1	Numeric	8	0	Desirability1	{1, Totally Dis	None	8	Right	Scale
10	A2	Numeric	8	0	Desirability2	{1, Totally Dis	None	8	Right	Scale

15

Filling missing values

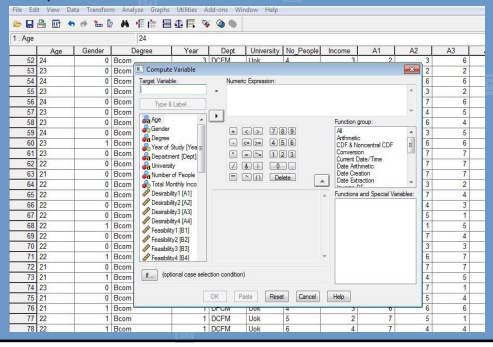
Filling missing value is important in research. This is because, the respondent may not fill all the items in the questionnaire.



16

Data transformation

- In our example (annex 01) number of questions are raised for each variable. Thus, we need to compute one variable from each set of questions.



17

Basic data analysis

- Once the variables are *defined*, *proper measured are set*, *treat missing values*, and *data transformations are computed*, its ready to analyze the data.

1. Frequencies
   1. Bar charts and pie charts
2. Measures of central tendency and dispersion
   1. Measures Central tendency: mean, median, and mode.
   2. Measures of dispersion: range, variance, standard deviation

18

Frequencies

From the menus, chose:

- Analyze
  - *Descriptive statistics*
    - *Frequencies*
    - *(select the relevant variables)*
  - Choose needed:
    - *Statistics ……*
    - *Charts …….*
    - *Format (for the order in which the results are to be displayed)*

Frequency Table

Age

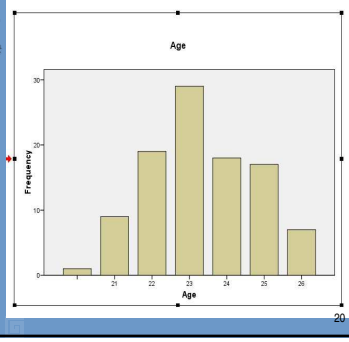
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	21	9	9.0	9.0	10.0
	22	19	19.0	19.0	29.0
	23	29	29.0	29.0	58.0
	24	18	18.0	18.0	76.0
	25	17	17.0	17.0	93.0
	26	7	7.0	7.0	100.0
	Total	100	100.0	100.0	

19

Frequencies *cont'd*

Frequencies can also be visually presented as bar charts, histograms, or pie charts.

Bar Chart



20

Measures of central tendency and dispersion.

- Measures of central tendency
- The mean: or the *average* is a measure of central tendency that offers a general picture of the data set.
- The median: the median is the central item in a group of observations when they are arrayed in either ascending or descending order.
- The mode: since median & mode do not necessarily imply a meaningful presentation, this can be signified by the *most frequently occurring phenomenon.*

Statistics

Total Monthly Income

N	Valid	100
	Missing	0
Mean		3.38
Median		3.00
Mode		3

21

Measures of central tendency and dispersion.

- Measures of dispersion
- This is important because; two sets of data can have the same mean but the dispersions could be different.
- Range: range is the *extreme values* in a set of observations.
- Variance: variance gives an indication of how dispersed the data in data set are.
- Standard deviation: this is the square root of the variance.

Statistics

Year of Study

N	Valid	100
	Missing	0
Std. Deviation		1.165
Variance		1.357
Range		3

22

References

- Kumar, R., 2011. *Research Methodology - A step by step guide for beginners.* 3rd ed. New Delhi: Sage Publications India Pvt Ltd.
- Sekaran, U. & Bougie, R., 2010. *Research Methods for Business: A skill building approach.* 5th ed. Delhi: Wiley India Publishers.

23

Lesson: Quantitative Data Analysis part -II

Research Methodology - COMC/CMOE/ COMT 41543

By
Srinath Dissanayake
DCFM

SPSS

24

Testing goodness of data

- **Reliability:** Simply states, reliability measures whether the ***instrument measures the concept*** and it helps assess the 'goodness' of a measure. Cronbach's alpha is the most common measure of internal consistency ("reliability").
- It is most commonly used when you have ***multiple Likert questions*** in a survey/questionnaire that form a scale, and you wish to determine if the scale is reliable.

25

Testing goodness of data

- Cronbach's alpha

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.805	.796	9

26

Hypothesis Testing

27

Testing a hypothesis about a single mean

- The one sample t-test is used to test the ***mean*** of the ***population*** from which a sample is drawn is equal to a ***comparison standard. (one sample t test is used)***
- Example: The Government claims cars traveling past your house average 55 mph, but you think they are actually traveling much faster. You steal a police radar gun and record the speed of the next nine cars that pass your house.
- **50,60,65,55,65,60,55,75,65. *(**Different Data!!)***

28

Testing a hypothesis about a single mean

- ***In this example the null hypothesis is;***
  - H_0 : the speed of the cars travelled by my house is ***equal*** to the claim of the government.
- ***The alternate hypothesis is;***
  - H_1 : the speed of the cars travelled by my house is ***differs*** to the claim of the government.

- To test your assumption, you entered data to SPSS in the **Analyze Menu**, select **Compare Means**, then choose **One sample t-test.**
- Select the variable "speed."
- Set Test Value equal to (in this case 55). You're testing to see if the data you have could really come from a population with a mean of 55.

29

Testing a hypothesis about a single mean

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
speed	9	61.3333	7.56637	2.52212

One-Sample Test

	Test Value = 55					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference: Lower	Upper
speed	2.511	8	.036	6.33333	.5173	12.1494

30

Testing a hypothesis about two related means

- We can also perform a (paired sample) t-test to examine the differences in the **same group *before and after treatment*.**
- Example: Would a group of employees perform better *after* undergoing training than they did *before*?
- *In this case, there would be two observations for each employee, one before training and after training.*

31

Testing a hypothesis about two related means

- ***In this example the null hypothesis is;***
  - H_0 There is no any significant differences in math scores in 1st and 2nd weeks
- ***The alternate hypothesis is;***
  - H_1 There is a significance difference in math scores in 1st and 2nd weeks
- To test your assumption, you entered data to SPSS in the **Analyze menu**, select **Nonparametric tests** and chose **2 Related samples**

32

Testing a hypothesis about two related means

Ranks

		N	Mean Rank	Sum of Ranks
score2 - score1	Negative Ranks	2[a]	3.00	6.00
	Positive Ranks	8[b]	6.13	49.00
	Ties	0[c]		
	Total	10		

a. score2 < score1

b. score2 > score1

c. score2 = score1

Test Statistics[b]

	score2 - score1
Z	-2.222[a]
Asymp. Sig. (2-tailed)	.026

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

33

Testing hypothesis about two unrelated means

- There are many instances when we are interested to know whether **two groups** are different from each other.
- Examples such as: Do MBAs perform better in organizational settings than business students with only a bachelor's degree?
- To answer such cases, **Kruskal-Wallis test** can be used.

34

Testing hypothesis about two unrelated means

- Working example: a personal manager of a large insurance company wished to evaluate the effectiveness of two **different** sales training programs that had been designed for new employees.
- SPSS output would be as follows

35

Testing hypothesis about two unrelated means

Kruskal-Wallis Test

Ranks

	training	N	Mean Rank
sale	1.00	25	25.84
	2.00	25	25.16
	Total	50	

Test Statistics[a,b]

	sale
Chi-Square	.027
df	1
Asymp. Sig.	.869

a. Kruskal Wallis Test

b. Grouping Variable: training

36

Testing hypothesis about two several means

- An **analysis of variance (ANOVA)** helps to examine the significant mean difference among **more than two groups** on an interval or ratio scaled dependent variables.
- For example: is there a significant difference in the amount of sales by the following **four** groups of salespersons.
  - Those who are sent to training schools
  - Those who are given on-the-job training during field trips
  - Those who have been tutored by the sales manager;
  - And those who have none of the above?

37

Testing hypothesis about two several means

- **Working example:** A manager wants to raise the productivity at his company by increasing the speed at which his employees can use a particular spreadsheet program. As he does not have the skills in-house, he employs an external agency which provides training in this spreadsheet program. **They offer 3 courses**: a beginner, intermediate and advanced course. He is unsure which course is needed for the type of work they do at his company, so he sends 10 employees on the beginner course, 10 on the intermediate and 10 on the advanced course. When they all return from the training, he gives them a problem to solve using the spreadsheet program, and times how long it takes them to complete the problem. He then compares the **three courses** (beginner, intermediate, advanced) to see if there are any differences in the average time it took to complete the problem.

38

Testing hypothesis about two several means

ANOVA

Time

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	91.467	2	45.733	4.467	.021
Within Groups	276.400	27	10.237		
Total	367.867	29			

39

Lesson: Quantitative Data Analysis part -III

Research Methodology - COMC/CMOE/ COMT 41543

By
Srinath Dissanayake
DCFM

SPSS

40

Regression analysis

- **When to use it!; to predict or to determine the "factor" (s) to a specific outcome. And in instances where to explore all types of dependence relationships.**
- **Forecasting models: e.g.; factors to predict national economy based on certain inputs (income levels, business environment etc)**
- **Determine: determinants of profitability in listed companies in Sri Lanka.**
- **Independent and dependent variable???**

41

Regression analysis cont'd

- **Simple regression analysis:** is used in a situation where **one** independent variable is hypothesized to affect one dependent variable.
- *Example: assume that the propensity to buy a product depends only on the perceived quality of that product.*
- **Multiple regression analysis**: is used in a situation where **more than one** independent variable is hypothesized to affect one dependent variable.
- *Example: assume that the weekly family consumption depends on family income, saving patterns, etc.*

42

Regression analysis cont'd

- **Linear relationship!!!**
- Example: let's assume that consumption of a family solely depends on income. Thus we collect data, and results are as follows.

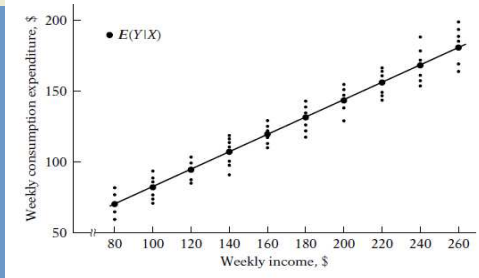
WEEKLY FAMILY INCOME X, $

$Y_\downarrow$ / $X\rightarrow$	80	100	120	140	160	180	200	220	240	260
Weekly family consumption expenditure Y, $	55	65	79	80	102	110	120	135	137	150
	60	70	84	93	107	115	136	137	145	152
	65	74	90	95	110	120	140	140	155	175
	70	80	94	103	116	130	144	152	165	178
	75	85	98	108	118	135	145	157	175	180
	–	88	–	113	125	140	–	160	189	185
	–	–	–	115	–	–	–	162	–	191
Total	325	462	445	707	678	750	685	1043	966	1211
Conditional means of Y, $E(Y \mid X)$	65	77	89	101	113	125	137	149	161	173

43

Regression analysis cont'd

- Once we collect the data we can plot the data in a scatter diagram



- We can see that there is a **liner relationship** between consumption and income levels.

44

Regression analysis cont'd

- This relationship can also express as follows

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

45

Regression analysis cont'd

- Knowing the basics of regression analysis, let's move to a complete example,
- *Following description provides facts pertaining to a study which was done in order to ascertain the impact of capital structure on firm's value.*
- *To measure the impact of capital structure on firm's value, six measures of variables are used as independent variables in this research. They are total debt ratio (DBT), liquidity (LIQU), firm size (FSIZE), capital intensity (CI), inventory (INVT), and efficiency (EFFI). The dependent variable used in this study is only the firm's value (FV). Finally the researcher was formulated the following regression model to reveal the impact of capital structure on firm's value.*
- *(2012 June examination, University of Kelaniya)*

46

Working example

- Regression equation

$$FV = \alpha + \beta_1 DBT + \beta_2 EFF I + \beta_3 LIQU + \beta_4 FSIZE + \beta_5 CI + \beta_6 INVT + \in_i$$

- The researcher's SPSS output results were as follows for the above study.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.963[a]	.927	.890	19.3233666

a. Predictors: (Constant), INVT, LIQU, CI, EFFI, DBT, FSIZE

47

Working example cont'd

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	56609.766	6	9434.961	25.268	.000[a]
	Residual	4480.710	12	373.392		
	Total	61090.476	18			

a. Predictors: (Constant), INVT, LIQU, CI, EFFI, DBT, FSIZE

b. Dependent Variable: FV

48

Working example cont'd

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-131.727	116.458		-1.131	.280
	DBT	-5.659	31.802	-.020	-.178	.862
	EFFI	41.963	5.557	.754	7.552	.000
	LIQU	-.864	1.275	-.070	-.678	.511
	FSIZE	30.352	14.752	.232	2.058	.062
	CI	-24.877	7.310	-.360	-3.403	.005
	INVT	-256.523	86.227	-.264	-2.975	.012

a. Dependent Variable: FV

Working example cont'd

- What can you state about the strength of the variables of the study? Interpret the figure. (03 marks)
- What can you state about the explanatory power of the variable of the study? Interpret the figure. (03 marks)
- What can you state about the overall significance of the model? Is the model is significant under 5% level of significance? (03 marks)

50

Working example cont'd

- What can you state about the relative importance of the variables of the study? What are the most important and the least important variables of the study? (03 marks)
- Interpret the B values of the study? Illustrate how Firm Value will behave according to the figures of corresponding B values. (04 marks)
- What are the statistical significance predictor variables of the firm value of the study? (Under 5% level of significance) (04 marks)

(Total 20 marks)

51

Pearson Product-Moment Correlation

- The Pearson product-moment correlation coefficient (or **Pearson correlation coefficient**, for short) is a measure of the **strength of a linear association** between two variables and is denoted by *r*.

Lets take a simple example.

- A researcher wants to know whether a person's **height** is related to how well they **perform** in a long jump. The researcher recruited untrained individuals from the general population, measured their height and had them perform a long jump. The researcher then investigated whether there is an association between height and long jump performance.
- SPSS output is a follows

52

Pearson Product-Moment Correlation

Correlations

		Height	Jump_Dist
Height	Pearson Correlation	1	.777**
	Sig. (2-tailed)		.000
	N	27	27
Jump_Dist	Pearson Correlation	.777**	1
	Sig. (2-tailed)	.000	
	N	27	27

**. Correlation is significant at the 0.01 level (2-tailed).

53

References

- Gujarati, D.N., 2008. *Basic Econometrics*. 4th ed. McGraw Hill.
- Kumar, R., 2011. *Research Methodology - A step by step guide for beginners*. 3rd ed. New Delhi: Sage Publications India Pvt Ltd.
- Sekaran, U. & Bougie, R., 2010. *Research Methods for Business: A skill building approach*. 5th ed. Delhi: Wiley India Publishers.

- "Research has shown that the best way to be happy is to make each day happy."

54