

UNIT IV Investigating relationship between variables- Correlation and Regression

4.1 Introduction

4.2 Objectives

4.3 Correlation

4.4 Pearson Correlation

4.5 Spearman Rank Correlation

4.6 Partial Correlation

4.7 Regression

4.8 Summary

4.9 Glossary

4.10 Answer to Check Your Progress

4.11 Reference/ Bibliography

4.12 Suggested Readings

4.13 Terminal & Model Questions

4.1 INTRODUCTION

SPSS can easily find the co-movement determination of variables. In this unit you would be learning how to use SPSS for finding relationship between two or more interdependent variables with changes in one variable being associated with the changes in other. Further, you would also be able to estimate or predict the unknown value of one variable from the known value of the other variable. SPSS is skilled with calculating effortlessly the nature and degree of relationship between two or more variables and can be used for finding out with certain measure of reliability the value of one variable i.e. dependent variable, corresponding to some known value of one or more variables i.e. independent variables. This unit describes how to instruct SPSS to use your data for examining the mutual dependence of interrelated variables, executing the command bringing out strength of relationship between the dependent and independent variables and generate conclusions.

4.2 OBJECTIVES

After reading this unit you will be able to:

- ❖ Learn about procedure for calculating correlations using SPSS.
- ❖ Conversant in using SPSS for calculating bivariate, rank and partial correlation and interpreting results.
- ❖ Work with the functions of SPSS for undertaking regression and interpreting results.

4.3 CORRELATION

Correlation assesses the magnitude of association or relationship between the two variables. It expresses strength of relationship that exists between two variables. Hence, it is a measure of the linear relationship between two variables. Thus, coefficient of correlation measures the relative strength of a linear relationship between two numerical variables. It reveals the magnitude and the

direction of the relationships. The values of the coefficient of correlation range from -1 to +1 for a perfect negative correlation to +1 for a perfect positive correlation. This means values which are near to 1 explain a firm association between the variables that are associated. The value which are nearer to 0 signifies no or insignificant association between variables.

However, the relationships describe as tendencies and do not show as causes and effects and hence it does not imply causation. The coefficient does not distinguish between independent and dependent variables and treats them symmetrically.

Generally, scatter diagram is used to examine whether the correlation exists between the variables or not.

Assumptions of correlations

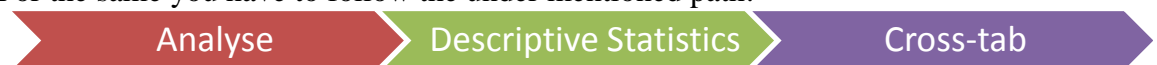
- Data should be measured on ratio or interval scale for estimating Pearson's correlation coefficient. In case of the data measured on ordinal scale, Spearman Rho should be used and if measured on nominal scale then Biserial correlations is used.
- Variables are normally distributed.
- The variance for both the variables should be equal.
- Data should be available on related pairs.

Through SPSS correlation can be work out through two ways; one way is through cross tabulation and the other is through correlation.

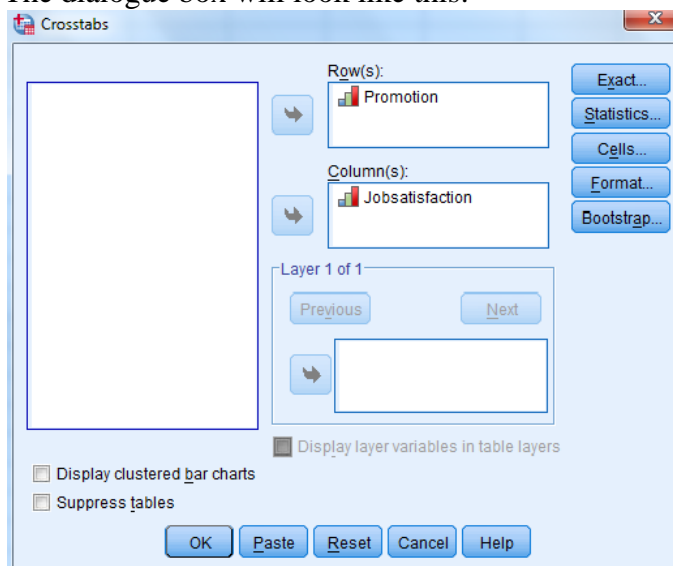
Crosstab

A correlation between two dichotomous or categorical variables is called Phi coefficient and is available in SPSS through Crosstab;

For the same you have to follow the under mentioned path:

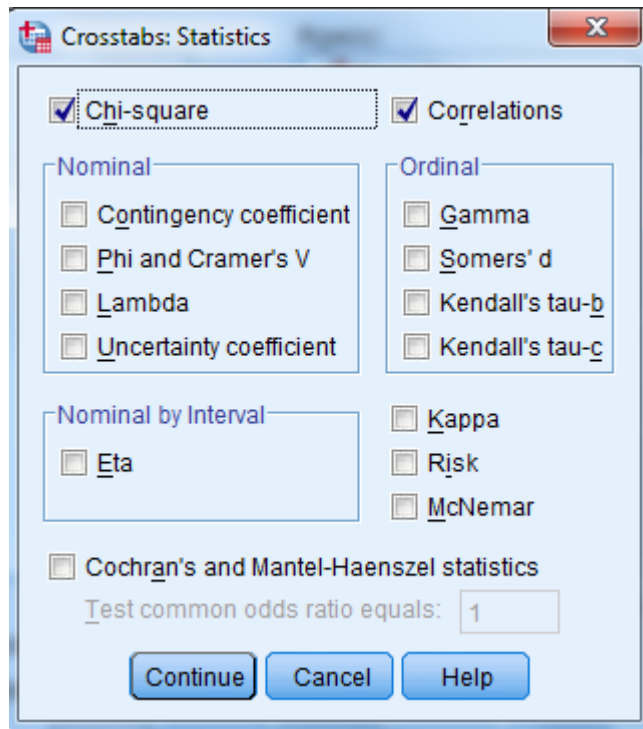


The dialogue box will look like this:



Img:4.1

You have to select the variables in row and column using arrow button. You have to click Statistics and select Chi-square and Correlations to find out the relationship.



Img 4.2

After clicking Continue, the output would include case processing summary and cross tabulation of two variables in terms of Chi-Square tests and Symmetric measure.

→ Crosstabs

[DataSet1]

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Promotion * Jobsatisfaction	14	100.0%	0	0.0%	14	100.0%

Promotion * Jobsatisfaction Crosstabulation

Count		Jobsatisfaction				Total
		Not Satisfied	Moderately Satisfied	Satisfied	Highly satisfied	
Promotion	Never Promoted	1	0	0	0	1
	Rarely Promoted	2	3	0	0	5
	Periodically Promoted	1	1	1	0	3
	Prompted quarterly	0	1	1	3	5
Total		4	5	2	3	14

Img 4.3

The above output depicts Case processing summary with the valid and missing values of the variables and the cross tabulation of the two variables.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	12.133 ^a	9	.206
Likelihood Ratio	14.520	9	.105
Linear-by-Linear Association	7.529	1	.006
N of Valid Cases	14		

a. 16 cells (100.0%) have expected count less than 5. The minimum expected count is .14.

Symmetric Measures

		Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Interval by Interval	Pearson's R	.761	.097	4.064	.002 ^c
Ordinal by Ordinal	Spearman Correlation	.756	.110	3.998	.002 ^c
N of Valid Cases		14			

a. Not assuming the null hypothesis.

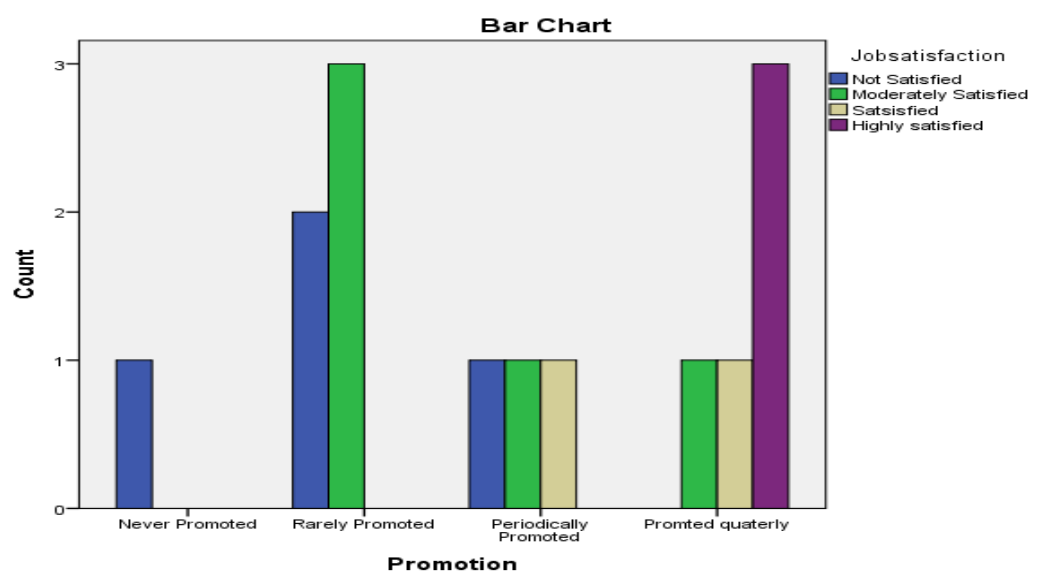
b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

Img:4.4

The table above shows that Chi-square value of 12.133 (df=9 and N=14). Further, symmetric measures narrate strong correlation between the variables and are statistically significant.

The cross tab function also have an option of displaying bar chart of the variables Job Satisfaction and Promotion given cross tabulation results. For the same, you have to select the 'Displaying Clustered Bar Charts' in the main dialogue box of cross tabs.



Img 4.5

Correlate

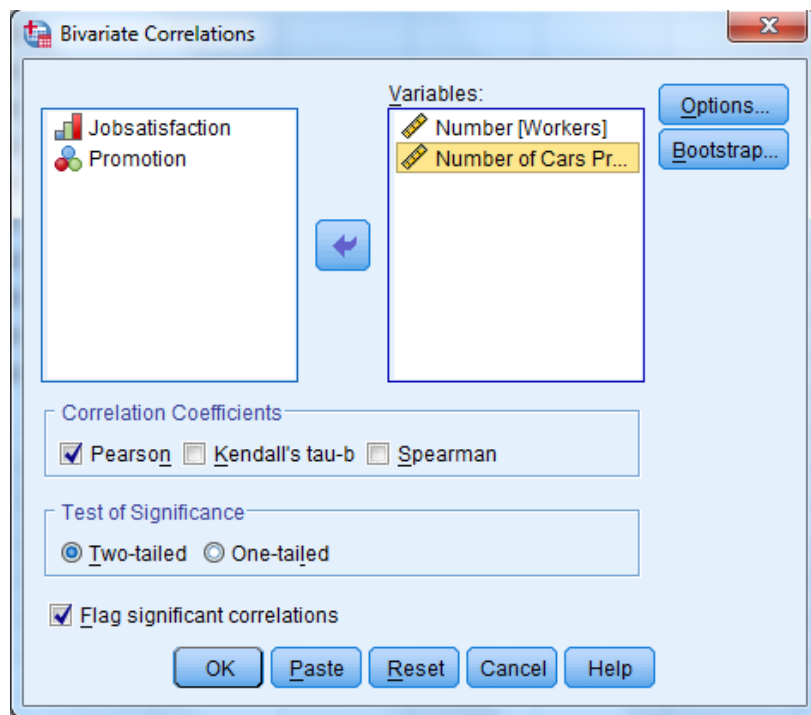
The other way for finding out correlation is through Correlate Function in Analyze Menu. Through Correlate you can calculate three types of correlation which are Bivariate, Partial and Distance.

Bivariate Correlations: As the term in itself reveals that bivariate correlation measures the nature of relationship between two variables.

The following steps for finding out Bivariate correlation may be referred –



After clicking Bivariate, the dialogue box would appear like this:



Img 4.6

Let us take the example that a car manufacturing company wants to know that is there is any relationship between workers employed and car produced. The company took data of eleven years. Then the calculation of correlation for the dataset would be done as under:

Select the variables and move them into the variables box. Further, ensure that the order of selection of variables influences the order in which the variables are listed in the output.

The correlation analysis using Pearson and Spearman Method has been separately dealt in the next part of the unit

(Note: Kendall's tau-b and Spearman are opted for ranked data.)

For further reading refer to the link:
http://en.wikipedia.org/wiki/Correlation_and_dependence

4.4 PEARSON CORRELATION

Pearson is selected for the continuous data since this measures the degree of relationship in real numbers, independent of the units in which the variables have been expressed and also it indicates the direction of the correlation.

You have to then specify the test of significance between Two-tailed or One-tailed to find the direction of your correlations. This will assess whether the hypothesis which is being tested creates a forecast about the direction of effect between the two variables i.e. positive or negative relationship between the variables. By default Two tailed is generally selected by SPSS.

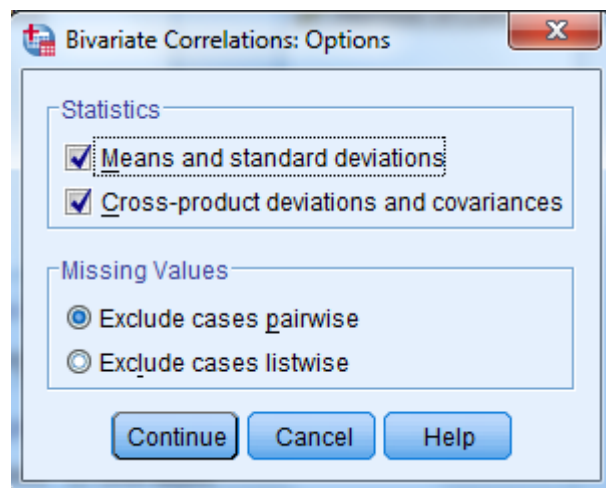
Flag Significant Correlations is selected default and places an asterisk (*) or double asterisk (**) next to correlations that attain a particular level of significance (usually 0.05 and 0.01). Whether or not significant values are flagged, the correlation, the significance accurate to three decimals and the number of subjects involved in each correlation will be included.

Significant correlations are highlighted underneath the output table with a one asterisk (*) for a significance of $p < 0.05$ and double asterisk (**) for $p < 0.01$.

You can then select option which will result into its sub dialogue box for selecting descriptive statistics as means, standard deviation, cross-product deviations and covariances along with the summary of correlation.

To exclude cases pair wise means that for a particular correlation in the matrix, if a case has one or two missing values that for finding the relationship or comparison its influence will not be included in finding correlation.

To exclude Cases list wise means that if a subject has any missing values, all data from that subject will be eliminated from any analyses.



Img 4.7

The output shall be displayed in the output viewer window:

➔ **Correlations**

[DataSet1]

Descriptive Statistics			
	Mean	Std. Deviation	N
Number of workers employed	409.0000	21.13291	11
Number of Cars Produced	36.7273	8.03854	11

Correlations			
		Number of workers employed	Number of Cars Produced
Number of workers employed	Pearson Correlation	1	.888**
	Sig. (2-tailed)		.000
	Sum of Squares and Cross-products	4466.000	1509.000
	Covariance	446.600	150.900
	N	11	11
Number of Cars Produced	Pearson Correlation	.888**	1
	Sig. (2-tailed)	.000	
	Sum of Squares and Cross-products	1509.000	646.182
	Covariance	150.900	64.618
	N	11	11

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

Img 4.8

The bi-variate correlation is undertaken between the numbers of workers employed and cars produced. It is hereby assessed that a relationship exists between number of workers employed and cars produced. The results depicts that there exists a positive relationship between number of workers employed and cars produced and is statistically significant as the p-value is less than 0.01. (r=.888, p<0.01)

SPSS generates output in a correlation matrix stating all the possible correlations between the pairs of variables and indicates the number of cases to compute them.

The interrelationship between three variables along with the number of cases used to compute them would be depicted in the output. The top most number is the value of correlation coefficient, the number below it is the two tailed p-value for the correlation and the last number represents number of observations.

Descriptive Statistics			
	Mean	Std. Deviation	N
Number of workers employed	409.0000	21.13291	11
Number of Cars Produced	36.7273	8.03854	11
Promotion granted	2.8571	1.02711	14

Correlations				
		Number of workers employed	Number of Cars Produced	Promotion granted
Number of workers employed	Pearson Correlation	1	.888**	-.433
	Sig. (2-tailed)		.000	.183
	Sum of Squares and Cross-products	4466.000	1509.000	-94.000
	Covariance	446.600	150.900	-9.400
	N	11	11	11
Number of Cars	Pearson Correlation	.888**	1	-.231

Produced	Sig. (2-tailed)	.000		.494
	Sum of Squares and Cross-products	1509.000	646.182	-19.091
	Covariance	150.900	64.618	-1.909
	N	11	11	11
	Pearson Correlation	-.433	-.231	1
Promotion granted	Sig. (2-tailed)	.183	.494	
	Sum of Squares and Cross-products	-94.000	-19.091	13.714
	Covariance	-9.400	-1.909	1.055
	N	11	11	14

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

Img 4.9

The result can be summarized as under:

- ✚ There is a strong positive correlation between Number of Workers Employed and Number of cars produced. The correlation coefficient is .888 and is statistically significant as the p-value is less than 0.01. SPSS also illustrates with an asterisk * that it is significant at the 0.01 level for a two-tailed prediction. The actual p value is shown to be 0.000. These figures are duplicated in the matrix.
- ✚ There is negative correlation between number of workers employed and promotion granted.
- ✚ Number of Cars produced and promotion granted are negatively correlated.

Significance is the likelihood that these results could happen by chance. The large p value indicates that observed values do not differ significantly from expected values.

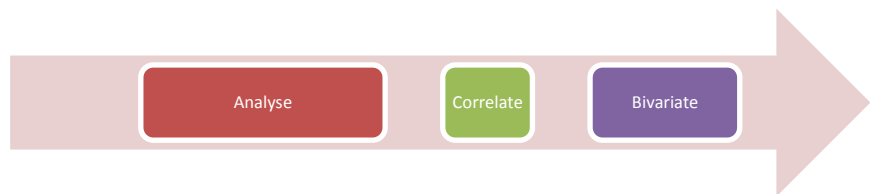
4.5 SPEARMAN RANK CORRELATION

The information collected by the researchers may be in form of quantitative or qualitative form. The Karl Pearson's coefficient of correlation is opted when the data is expressed in numbers. When the nature of data is expressed in qualitative attributes like honesty, efficiency, average, effectiveness, etc. or the data is provided in ordinal scale i.e. ranking; then Spearman's rho correlation or rank correlation is used.

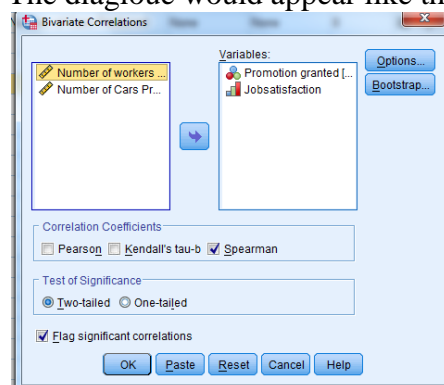
Spearman rank Correlation is used when:

- When the data is is ordinal scale- one or both the variables are not in interval scale.
- When there is presence of extreme outliers
- when data is not normally distributed
- when other assumptions of the Pearson correlation are violated
- when the two variables is non-linear.

Rank correlation can be calculated by opting the same menu of commands as used for Pearson Coefficient of Correlation.



The dialogue would appear like this :



Img 4.10

You have to then select Spearman’s Rank Correlation and opt for two tailed test of signifcance and exclude cases pairwise. Correlation table is displayed under the heading Non parametric Correlations.

Nonparametric Correlations

		Correlation Coefficient	1.000	.756**
	Promotion granted	Sig. (2-tailed)	.	.002
		N	14	14
Spearman's rho		Correlation Coefficient	.756**	1.000
	Job satisfaction	Sig. (2-tailed)	.002	.
		N	14	14

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

```

DATASET ACTIVATE DataSet1.
NONPAR CORR
  /VARIABLES=Promotion Jobsatisfaction
  /PRINT=SPEARMAN TWOTAIL NOSIG
  /MISSING=PAIRWISE.

```

➔ Nonparametric Correlations

[DataSet1]

Correlations			Promotion granted	Jobsatisfaction
Spearman's rho	Promotion granted	Correlation Coefficient	1.000	.756**
		Sig. (2-tailed)	.	.002
		N	14	14
	Jobsatisfaction	Correlation Coefficient	.756**	1.000
		Sig. (2-tailed)	.002	.
		N	14	14

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

Img 4.11

Correlation table is displayed as Non parametric correlations in the output viewer. The non-parametric correlations are interpreted as you interpret output of Pearson coefficient of correlation. Above table narrates that there is a positive correlation between promotion given and job satisfaction level among employee as the correlation coefficient is .756 and it is statistically significant.

$$r = .756, p < 0.01$$

4.6 PARTIAL CORRELATION

Partial correlation measures an association between two variables with effects of one or more variables factored out. The partial correlation coefficients measure the net correlation between dependent variable and one of the two independent variables (holding the other independent variable constant or eliminating its relationship with the dependent one). Partial Correlation analysis seeks to measure separately the relationship between two variables in such connection. Thus, the subscript to the right of the decimal denotes that the variable is held constant or relationship with the dependent variable has been eliminated while the subscript to the left of decimal point indicates the variables (two) between which net correlation is computed.

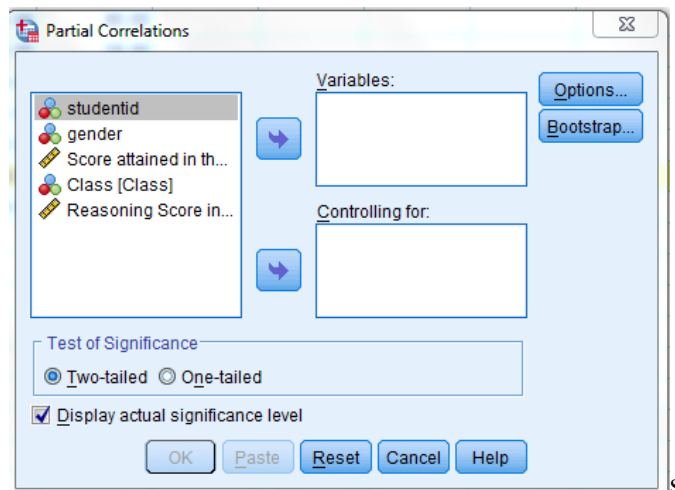
For finding out the partial correlation the steps are provided as under:

Analyse

Correlate

Partial

After this, a Partial Correlations dialog box would appear. Click the variables you want to correlate and move them to the Variables box and then transfer the variable you wish to control to the 'Controlling For' box. Make sure the Two-tailed option is selected in the Test of Significance box and 'Display Actual Significance Level' is selected. Then Click Options, and choose Means and standard deviations in the Statistics box. Then click options and continue.



The table shows the association of Job Satisfaction and Promotion Granted:

Nonparametric Correlations

[DataSet1] C:\Users\magarwal\Documents\Untit

Double-click to activate

Correlations

			Jobsatisfac ion	Promotion granted
→	Spearman's rho	Jobsatisfaction	1.000	.838**
		Correlation Coefficient		
		Sig. (2-tailed)	.	.000
		N	14	14
	Promotion granted	Correlation Coefficient	.838**	1.000
		Sig. (2-tailed)	.000	.
N		14	14	

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

Img 4.12

In this example, if you want to assess the relationship between job satisfaction and promotion granted however controlling the number of workers employed.

Under the heading Partial Corr, a correlations table is displayed. You will find that the correlation coefficient of Job satisfaction and promotion granted has changed because we have controlled numbers of workers employed while computing correlations.

Partial Corr				
[DataSet1] C:\Users\magarwal\Documents\Untitled2.sav				
Descriptive Statistics				
	Mean	Std. Deviation	N	
Jobsatisfaction	2.4286	1.22250	14	
Promotion granted	2.8571	1.02711	14	
Number of workers employed	399.9286	34.37008	14	
Correlations				
Control Variables			Jobsatisfaction n	Promotion granted
Number of workers employed	Jobsatisfaction	Correlation	1.000	.860
		Significance (2-tailed)	.	.000
		df	0	11
	Promotion granted	Correlation	.860	1.000
		Significance (2-tailed)	.000	.
		df	11	0

Img 4.13

Though the change in correlation coefficient is small however, by controlling a factor the nature of relationship can show a drastic change. (The variable controlled is termed as lurking variable)

Distance: This helps in calculating either similarities or dissimilarities (distances), either between pairs of variables or among pairs of cases. These similarity or distance measures can then be used with other procedures, such as factor analysis, cluster analysis, or multidimensional scaling, to help analyze complex data sets.



Check Your Progress-A

Q1. When the Kendall's tau-b and Spearman test is is used in SPSS for finding out correlations?

Q2. What is the procedure for computing partial correlations?

Q3. Distinguish between cross tab and correlate function.

Q4. Fill in the Blanks with appropriate word or words.

- a) When the assumptions underlying the correlation cannot be met then can be used.
- b) A Coefficient explains the relationship between two continuous variables and is available in SPSS through Bivariate.
- c)is selected default and places an asterisk (*) or double asterisk (**) next to correlations that attain a particular level of significance.
- d) The correlations can be graphically shown by using

Q5. Which of the following statements are true or false?

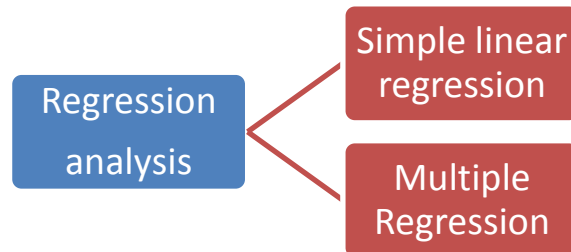
- a) Negative correlations are never depicted in SPSS.
- b) Each correlation appears twice in the square matrix.

4.7 REGRESSION

Correlation analysis assesses the relationship between the variables however; it does not able to explain the extent variables are dependent on each other. Hence, regression analysis is used to make quantitative estimates of various theoretical relationships.

Regression analysis is statistical measure for studying the dependence between the two variables. Regression analysis is used to predict the variability of dependent variables on the basis of the values of independent variable. Hence, it is used to estimate or predict of the unknown value of one variable from the known variable of the other value. In regression, we try to estimate or predict the average value of one variable that rest upon the fixed value of other variable.

Regression analysis is of two types; one is simple regression analysis and the other is multiple regression analysis. Regression analysis that involves two variables is termed as bi-variate linear regression however, analysis that involves more than two variables are termed as multiple regression analysis.



When the relationship between two or more variables is described best by the way of a straight line then it is called as linear relationship. However, if the underlying relationship between two or more variables is best described by a curve or a surface it is called as non-linear relationship or curvilinear relationship.

Simple Linear Regression analysis-

The bi-variate linear regression analysis involves analyzing the straight line relationship

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon$$

where Y represents the dependent variable for observation i

X represents independent variable for observation i

β_0 and β_1 are two constants which are known as regression coefficients.

ε is the error term

β_1 is the slope coefficient i.e. it is the change in value of Y with the corresponding change in the one unit of X.

These independent variables are also called as "Explanatory" or "Exogenous" variables. They are also referred to as "Regressors."

Regression analysis using SPSS:

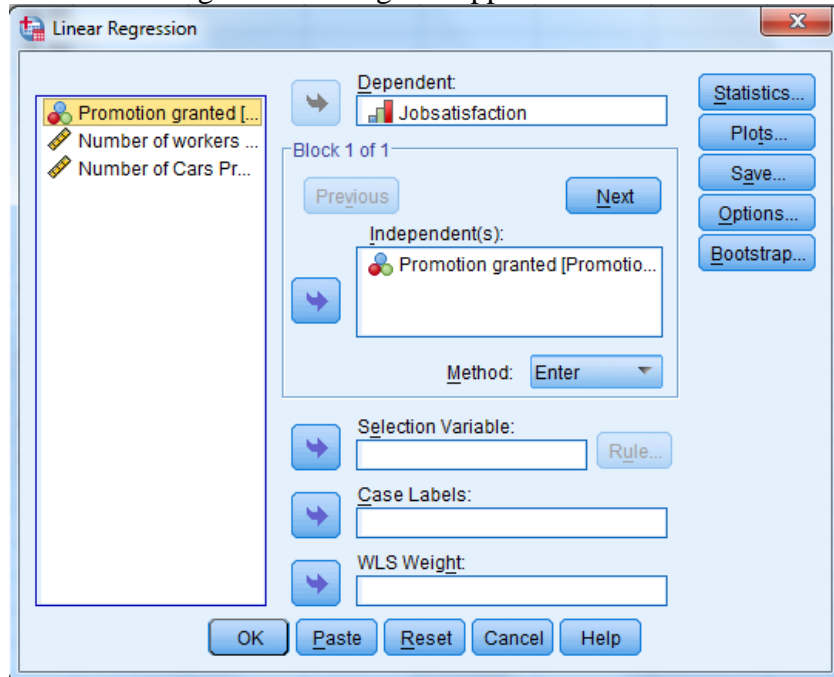
Let us take the previous example of Job Satisfaction of employees and promotion given, is you may wish to further investigate this relationship by examining

whether promotion reliably *predicts* the Job satisfaction. To do this you have to Click **Analyze** and choose Regression, then click **Linear**.

Path:



The Linear Regression dialog box appears as shown below:



Img 4.14

Then select Job Satisfaction in Dependent variable and Promotion given as Independent variable by using arrow button. There is also a method option which states that the method for selecting or excluding variables in that model. These are represented by Stepwise, Remove, Backward, Forward and Enter this gives you an option of removing or retaining the predictor variables in regression equation. By default Enter method is selected.

Through method option you can define how independent variables are entered into the analysis. Using different methods, you can construct a variety of regression models from the same set of variables.

- In Enter Method all the variables are considered in a block in a single attempt.
- In Stepwise Method the independent variable not in the equation which has the smallest probability of F is considered, if that probability is sufficiently small. Variables already in the regression equation are removed if their probability of F becomes sufficiently large. The method ends when no more variables are eligible for inclusion or removal.
- In Remove Method as itself it conveys that while selecting variables all variables in a single attempt are removed.

- In Backward Elimination an independent variables are entered into the equation and then sequentially removed. The variable with the smallest partial correlation with the dependent variable are sequentially removed. The procedure ends when there are no variables in the equation that satisfy the removal criteria.
- In Forward selection the first variable considered for entry into the equation is the one with the largest positive or negative correlation with the dependent variable. This variable is entered into the equation only if it satisfies the criterion for entry. If the first variable is entered, the independent variable (not in the equation) that has the largest partial correlation is considered next. The procedure stops when there are no variables that meet the entry criterion.

Bootstrapping is a method for deriving robust estimates of standard errors and confidence intervals for estimates such as the mean, median, proportion, odds ratio, correlation coefficient or regression coefficient. Plots can aid in the validation of the assumptions of normality, linearity, and equality of variances.

After clicking O.K., the following output is generated:

→ **Regression**

[DataSet1] C:\Users\magarwal\Documents\Untitled2.sav

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Promotion granted ^b	.	Enter

a. Dependent Variable: Jobsatisfaction

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.849 ^a	.721	.697	.67250

a. Predictors: (Constant), Promotion granted

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	14.001	1	14.001	30.959	.000 ^b
	Residual	5.427	12	.452		
	Total	19.429	13			

a. Dependent Variable: Jobsatisfaction

b. Predictors: (Constant), Promotion granted

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		

1	(Constant)	-.458	.549		-.835	.420
	Promotion granted	1.010	.182	.849	5.564	.000

a. Dependent Variable: Job satisfaction

Img 4.15

Variables Entered and Variable Removed

Independent Variables that are considered for regression are considered in Variables Entered whereas list of variables which are removed for computing regression are highlighted in Variables removed.

Model Summary Table-

Method explains the method used for computing regression and model is used to assign numerous models in one regression command. R is the link between the examined and supposed value of a dependent variable. The R Square value in the Model Summary table shows the amount of variance in the dependent variable that can be explained by the independent variable. R-square is called as the coefficient of determination in two variable cases and it is known as multiple coefficient of determination in case when more than two variables are included in model. The Adjusted R Square adjusts for a bias in R Square that when additional independent variables are included then the change in dependent variable is depicted by adjusted R-square. The standard error of the estimate is the standard deviation of the error term and is the square root of the Mean Square Residual.

ANOVA Table-

Model includes source of variance-Regression, Residual and Total. The Total variance is portioned into variance which can be explained by the regression model. The Residual model is termed as error. The sum of square for the regression and residual add up to the total.

$R \text{ Square} = \text{Regression Sum of square} / \text{Total sum of square}$

DF represents degrees of freedom that are linked with the origin of variances. The total variance has N-1 degrees of freedom.

$\text{Mean Square} = \text{Sum of square} / \text{Degree of Freedom}$

The F-value is the mean square regression divided by the mean square residual. The p-value is compared to 0.05.

If $p < 0.05$ is generally interpreted as indicating a statistically significant correlation. If $P > 0.05$ then the strength of association between the two variables is usually not treated as statistically significant. It can be inferred that the independent variables does not show statistically significant relationship with the dependent variable.

This table defines the equation for you in its first column. Assuming that the assumption of the linear relationship is met, income can be predicted with the following:

$$\text{Job Satisfaction} = 1.010 (\text{Promotion Granted}) - .458$$

Multiple regressions

In multiple regression analysis a variable can be predicted on the basis of more than one predictive factor. The multiple regression model is more close to reality and hence generally used in practice. The techniques for estimating the constants of the multiple regression equation are similar as used for simple regression.

The equation can be depicted as under-

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

The dependent variable is required to be measured on a continuous ratio scale. The independent variable can be continuous, categorical or combination of the two.

Regression

[DataSet2] C:\Users\magarwal\Downloads\acadindx.sav

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	female, reading, writing ^b	.	Enter

a. Dependent Variable: academic index

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.789 ^a	.622	.616	10.42169

a. Predictors: (Constant), female, reading, writing

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	34994.282	3	11664.761	107.399	.000 ^b
	Residual	21287.873	196	108.612		
	Total	56282.155	199			

a. Dependent Variable: academic index

b. Predictors: (Constant), female, reading, writing

Img 4.16

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	96.118	4.490		21.409	.000
1 Counseling attended in Hours	.718	.093	.438	7.713	.000
Contributed in Assignments	.791	.104	.446	7.594	.000
Gender	-5.832	1.588	-.173	-3.672	.000

a. Dependent Variable: Examination score

You can interpret the results for multiple variables as you did for linear relationship.

Click on the following links to know more about regression:

- <http://en.wikipedia.org/wiki/Regression>
- <http://www.investopedia.com/terms/r/regression.asp>

4.8 SUMMARY

In this unit we learned how to find the relationship between two variables in SPSS. Thus, this unit provides a correlation analysis technique in SPSS that looks at the indirect relationships in survey data and used to establish the variables which are most closely associated. In the unit, we also looked how to compute regression analysis so that one can predict the value of one variable on the basis of the results of one or more variables.



4.9 GLOSSARY

- **Coefficient of determination:** : It is a measure of the degree of linear association or correlation between the two variables in which one happens to be independent and the other being dependent variable.

- **Partial Correlation:** Partial correlation measures an association between two variables with effects of one or more others variables factored out.
- **Standard Error of the Estimate:** It is used to measure the reliability of the estimating equation.



4.10 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress –A

4.

- a) Non parametric Spearman's rank order correlation
- b) Pearson product moment correlation
- c) Flag Significant Correlations
- d) Scatter diagrams

5.

- a) False
- b) True

4.11 REFERENCES

1. Gupta S.L. and Gupta Hitesh, SPSS 17.0 for Researchers, International Book House Pvt. Ltd., New Delhi
2. Pandya Kiran, Bulsari Smruti, Sinha Sanjay(2012) , SPSS in Simple Steps, Dreamtech press, New Delhi
3. Hooda R. P(2000), Statistics for Business and Economics, Macmillan India Ltd.
4. George Darren and Mallery Paul (2011), SPSS for windows Step by Step, Dorling Kindersley Publishing
5. Research and Communications Methodology, Self Learning Material for MBA Course CP1007, Unit 9, Vikas Publications, Noida



4.12 SUGGESTED READINGS

1. Gupta S.L. and Gupta Hitesh, SPSS 17.0 for Researchers, International Book House Pvt. Ltd., New Delhi
2. Pandya Kiran, Bulsari Smruti, Sinha Sanjay(2012) , SPSS in Simple Steps, Dreamtech press, New Delhi
3. Hooda R. P(2000), Statistics for Business and Economics, Macmillan India Ltd.
4. George Darren and Mallery Paul (2011), SPSS for windows Step by Step, Dorling Kindersley Publishing
5. IBM reference guide posted on the SPSS website.



4.13 *TERMINAL QUESTIONS*

- Q1. Explain the procedure for computing regression analysis using SPSS?
- Q2. Explain the various types of correlations that can be computed using SPSS.
- Q3. Using the funds.sav find out the following
- At the 0.05 level of significance, is there a significant correlation between?
- a) Height and nature of mutual funds selected
 - b) Weight and nature of mutual funds selected
- Q4. Develop hypothetical score for examining the relationship between online classes rendered, practical exercises given to the students of Accounts, and the performance in the annual exam (dependent variable). Perform regression on these data.