UGRC 120
Numeracy Skills

Session 7

MEASURE OF LINEAR ASSOCIATION & RELATION

Lecturer: Dr. Ezekiel N. N. Nortey/Mr. Enoch Nii Boi Quaye, Statistics
Contact Information: ennortey@ug.edu.gh/enbquaye@ug.edu.gh

UNIVERSITY OF GHANA
College of Education
School of Continuing and Distance Education
2014/2015 – 2016/2017

Session Overview

OVERVIEW

- It is possible to find a relationship or some form of association between variables of a dataset. Being a researcher, it is first expedient to determine if some degree of association or relationship exists between variables and further ascertain the nature of existing relationship.

- Though several forms of relationships can be identified. However, this section of the study guide considers measures of linear relationships. This is commonly applied to bivariate data.

Goals and Objectives

At the end of the session, the student will be able to:

- Understand the essence of correlation between two variables.
- Carry out computations involving linear and rank correlation.
- Compute a correlation coefficient.
- Interpret the value of computed correlation coefficient.

Session Outline

The key topics to be covered in the session are as follows:

- Correlation Analysis
- Establishing Existence of Relationships
- Direction of Linear Relationship
- Measuring the Degree of Linear Relationship
- Rank Correlation Coefficient
- Guide to Interpreting the Correlation Coefficient

Reading List

- Refer to Unit 6 of Recommended Text:

Nortey, E. and Afrim, J. (2013). *Numeracy skills: The basics and beyond.* Accra: Dieco Ventures

Topic One

CORRELATION ANALYSIS

INTRODUCTION

- Correlation analysis deals with the analysis of the relationship between two quantitative variables.

- The two variables are designated as X and Y. The variable which we wish to explain or estimate is referred to as the dependent variable. In this instance, it is denoted by the symbol Y

- The variable or factor from which the estimates are made is called the independent variable and is denoted by the symbol X.

INTRODUCTION - Continued

- The terms dependent and independent do not necessarily imply any cause and effect relationship between the two variables.

- We simply mean that estimates of the values of the dependent variable (*Y*) may be obtained for given values of the independent variable *X*. For that reason the values of *Y* are dependent upon the values of *X*.

INTRODUCTION - Continued

- The *X* variable may or may not be causing changes in the *Y* variable.

- For example if we are estimating sales of a product from figures on advertising expenditures, sales is the dependent variable and advertising expenditures is the independent variable.

- There may or may not be a causal connection between these two factors in the sense that changes in advertising expenditures cause changes in sales.

INTRODUCTION - Continued

- In fact in certain situations, the cause-effect relation may be just the opposite of what appears to be the obvious one.

- For example suppose a company budgets a product's advertising expenditures for the next year as a flat percentage of the sales of that product during the preceding year. Then advertising expenditures are more directly dependent on sales than vice versa.

INTRODUCTION - Continued

- There are numerous illustrations of variables that can reasonably be assumed related to one another, that is, they are correlated.

For example:

- consumption expenditure (Y) and income (X)
- personal net savings (Y) and disposable income (X)
- success in University (Y) and examination grades (X)

- For each pair, the first named is to be estimated and thus is the dependent variable and the second factor is the independent variable

Activity 1.1

- Refer to **“session 7 examples and activities” NOW** and complete Activity 1.1

INTRODUCTION - Continued

1. Establishing Existence i.e. finding out whether there exists any relationship at all between the two variables. This is done with the use of a scatter diagram;

2. There is the establishment of direction i.e. establishing whether the relationship is positive or negative and

3. There is the determination of the degree of relationship i.e. calculating the value of coefficient of correlation which is denoted by r.

INTRODUCTION - Continued

- Refer to **"session 7 examples and activities" NOW** and complete Activity 1.2

Topic Two

ESTABLISHING EXISTENCE OF RELATIONSHIP

INTRODUCTION

- ***A very key technique used in establishing existence of relationship between variables is the use of a scatter diagram or scattergram.***

Scatter Diagram

- A very useful aid in studying the relationship between two variables is to plot the data on a graph. This allows a visual examination of the extent to which the variables are related.

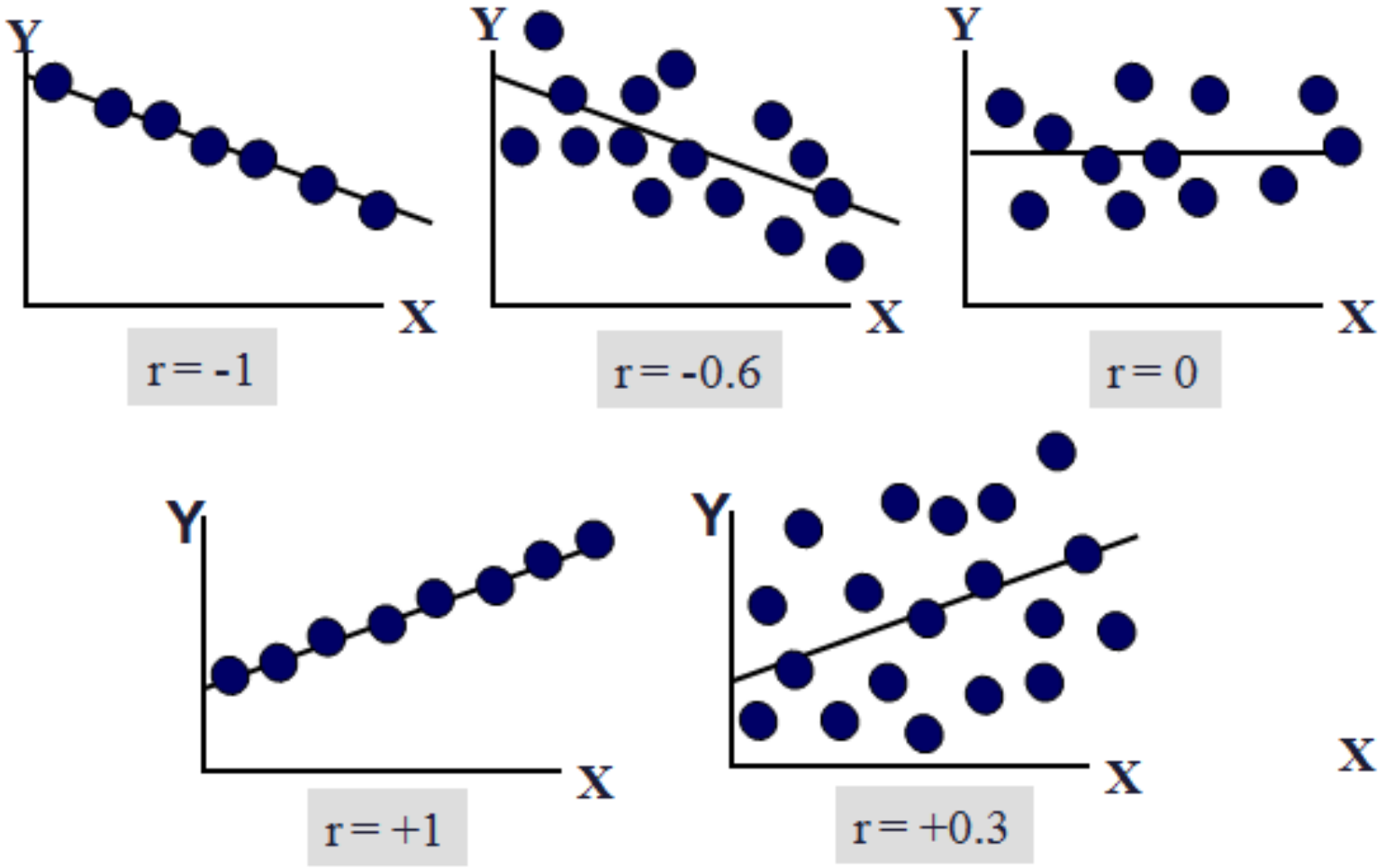
INTRODUCTION - Continued

- The chart used for this purpose is known as a scatter diagram which is a graph on which each plotted point represents an observed pair of values of the dependent and independent variables.

- The establishment of whether there exists any relationship between any variables is done by the use of a scatter diagram.

- This is illustrated by diagrams in the next slide. **Corresponding values of the correlation coefficients are displayed.**

INTRODUCTION - Continued



INTRODUCTION - Continued

- In the prior slide, each diagram represents an observed pair of values of the two variables X and Y.

- The scatter diagram is usually compressed into an index called the coefficient of correlation (r).

Topic Three

DIRECTION OF RELATIONSHIP

Direction of Relationship

- The coefficient of correlation (r) provides the measure of the strength and direction of the relationship, which can be used for direct comparisons.

- The coefficient of correlation can be either positive or negative.

- The value of r ranges from -1 to $+1$. If r is -1, it means there is a *perfect negative relationship* between the two variables.

Direction of Relationship

- This means that as the X variable increases, the Y variable decreases and shows an inverse relationship.

- If r is $+1$ it means there is a perfect positive relationship between the two variables.

- This means that as the X variable increases the Yvariable also increase.

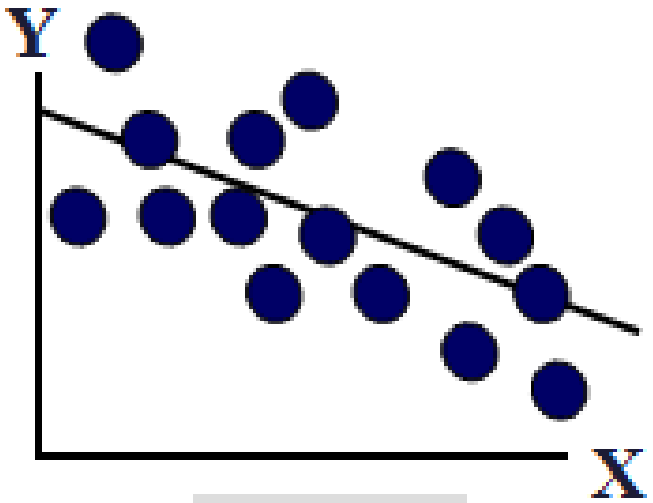
Direction of Relationship

- In the social sciences it is difficult to find two variables that are perfectly correlated in a positive or negative direction.

- This is due to the fact that we always have to take account of the effect of other intervening variables.

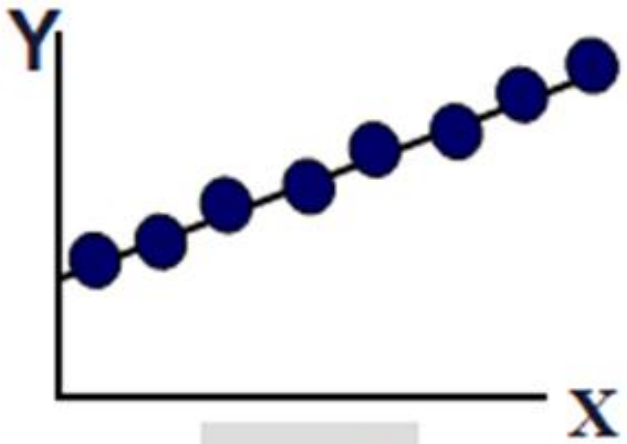
- The influences of these other variables usually reduce the ability of social scientists to predict to precision.

Direction of Relationship



r = -0.6

Negative Relationship



r = +1

Positive Relationship

Direction of Relationship

- A positive r indicates that as the independent variable (X) increases e.g. as the number of years spent on education increases, it influences an increase in the dependent Y variable or the income people earn.

- A negative r indicates an inverse relationship between the X and Y variables.

- This means that an increase in the independent variable (X), influences a decrease in the dependent (Y) variable. Where the $r = 0$ it indicates no relationship between X and Y.

Properties of The Correlation Coefficient

1. Unit free
2. Ranges between –1 and 1
3. The closer to –1, the stronger the negative linear relationship
4. The closer to 1, the stronger the positive linear relationship
5. The closer to 0, the weaker any linear relationship

Correlation Coefficient

Refer to **“session 7 examples and activities” NOW** and complete Activity 1.3

Topic Four

DEGREE OF RELATIONSHIP

DEGREE OF RELATIONSHIP

- This involves the computation of the coefficient of correlation (r) between the X and Y variables.

- Linear r is usually called *Pearsonian correlation* or the *Pearson's Product-Moment correlation*.

- Pearson's correlation measures linearity between X and Y. The term "product-moment" is a mathematical term, which refers to the mean of a product.

DEGREE OF RELATIONSHIP

- Pearson's r equals the sum of the product of X and Y deviations from their respective means divided by the value that sum would have if the observations fell perfectly on a straight line.

- By definition, Pearson's r is as follows:

$$r = \frac{\sum \hat{X}\hat{Y}}{\sqrt{\sum \hat{X}^2 \sum \hat{Y}^2}}$$

where $\hat{X} = X - \bar{X}$ and $\hat{Y} = Y - \bar{Y}$

DEGREE OF RELATIONSHIP

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

- The above is the mathematical equivalent of the definitional formula but it avoids the need to compute deviations from the mean for each pair of observations

DEGREE OF RELATIONSHIP

Refer to **“session 7 examples and activities” NOW** and practice with Example 1.1

The Correlation Coefficient Using Microsoft Excel

Refer to **"session 6 examples and activities" NOW** and practice along the excel workbook demonstration.

Topic Five

GUIDE TO INTERPRETING THE CORRELATION COEFFICIENT

Interpretation of Correlation Coefficient (r)

- Generally the following is a rough but useful guide to the degree of relationship indicated by the size of the correlation coefficient.

  - **Less than 0.20:** *Indicates a slight correlation i.e. the relationship is so small as to be negligible.*
  - **0.21 – 0.40**: *Indicates a low correlation i.e. a definite relationship exists between the two variables, but the relationship is a weak one.*

Interpretation of Correlation Coefficient (r)

- **0.41 – 0.70:** *Indicates moderate correlation i.e. a substantial relationship exist between the two variables.*

- **0.71 – 0.99:** *Indicates a very high correlation i.e. a very strong relationship exist between the two variables.*

- **1.00 :** *Perfect correlation i.e. an exact linear relationship between the two variables.*

Interpretation of Correlation Coefficient (r)

- The answer $r = 0.977$ is indication of a very high linear correlation between X and Y. Since the answer is positive it means that there is a very strong relationship between X and Y such that as X (years in school) variable increases it influences an increase in the Y (income) variable.

- It means that as the years spent in school increases, the person acquires higher qualifications and this tends to make the person earn a higher income.

Coefficient of Determination

- The coefficient of determination r^2 measures the extent to which the independent variable (X) explains the dependent variable (Y) i.e. it measures how much of the variation in the Y variable is explained by the X variable.

- This also indicates the strength of the X variable. It is usually computed as a percentage

$$r^2 = (0.977)^2 \times 100$$
$$= 0.954529 \times 100 = 95.45 = 95.5\%$$

Coefficient of Determination

- What this means is that the X variable (years in school) explains as much as 95.5% of the variation in Y (income) i.e. the number of years spent in school accounts for 95.5% of the income people earn.

- This leaves only 4.5% of Y (income) accounted for by factors other than years in school. Years in school (X) is therefore a very strong variable in the determination of income (Y).

Topic Six

RANK CORRELATION ANALYSIS

RANK CORRELATION ANALYSIS - Computation

- This is a special type of correlation coefficient whose computation is based on the rankings of two variables.

- Rank correlation was developed by Spearman. In this computation, the ranks of the values of the two variables are substituted for the actual values and used for the computation.

- Sometimes in the social sciences it is possible to rank the characteristics of an item but not to give it a more specific value.

RANK CORRELATION ANALYSIS - Computation

- Rank correlation is computed from the following expression

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

where $d = difference\ between\ pairs\ of\ ranks$

RANK CORRELATION ANALYSIS - Computation

Refer to **"session 7 examples and activities" NOW** and practice with Example 1.2

Interpretation

- The value of rank correlation (r_{rank}) range between -1 and $+1$. When the value of $r_{rank} > 0$ then there is an agreement between the rankings in the same positive $(+ve)$ direction i.e. high ranks in one variable tend to be associated with high ranks in the second variable.

- If the value of $r_{rank} < 0$ it means the rankings in the two variables are in disagreement.

Interpretation - Continued

Refer to **"session 7 examples and activities" NOW** and complete Activity 1.

- **Complete Assignment 6 and submit**

.