STAT 2361/STAT 2311

LECTURE NOTES

PREPEARED BY MOHAMMAD MADIAH

CHAPTER 2 PART 2

Descriptive Statistics

Tabular and Graphical Methods

2.4 Crosstabulations and Scatter Diagrams

A cross-tabulation (or crosstab) is a **tabular summary** for two variables (either qualitative or quantitative). The classes for one variable are represented by the rows, the classes of the other variable are represented by the columns. Cross-tabulation tables provide a wealth of information about the **relationship** between the variables.

Example 1

a researcher is interewsted in whether there is a relationship between gender and salary at her company. She randomly sampled some men and women on the company and asked them about their salaries in ILS. Her results appear in the table below.

Salary (ILS)	Gender		
	Men	Women	Total
1450 – 5000	120	80	200
5001 – 7500	140	100	240
7501 – 10000	30	15	45
More than 10000	10	5	15
Total	300	200	**500**

- Variables:
  - Gender, qualitative, ordinal
  - Salary, quantitative, ratio
- Number of elements (sample size): 500

1. What is the sample size?
   Total number of elements = 500 employees.
2. What is the number of employees with salaries more than 7500 ILS?
   Number = 45 + 15 = 60
3. What is the percentage of women employees?
   Percentage of women = (200/500) (100%) = 40%
4. What is the relative frequency of employees with salaries more than 10000 ILS?
   Relative frequency = 15/500 = 0.03.
5. Among the men employees, what is the percentage of those with salaries between 7501 – 10000?

==Percentage = 30 / 300 = 0.1*100%= 10%.==

6. Among the employees with salaries less than or equal to 7500 ILS, what is the percentage of women employees?
==Percentage = (100 + 80)/ (240 +200) = 180 / 440 = 40.91%==

Scatter Diagrams and trend lines.

Crosstabulation is a method that can be used to summarize the data for two variables and help reveal the relationship between the variables. In most cases, a graphical display is more useful for recognizing **patterns** and **trends** in the data.

A scatter diagram (plot) is a graphical display of the relationship between two quantitative (numerical variables x (**independent**), y (**dependent**) variables, and a trend line is a line that provides an approximation of the relationship.

Let x and y be two numerical variables. Select a sample of size n, then we have the following observations for the two variables:
$(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$

The first step to get some idea about possible relationship between two variables is constructing a scatter diagram. It is **an indicator** "about" the relation between the two variables.

To construct a scatter diagram we graph pairs of numerical data, with one variable on each axis. (Independent == Horizontal. Dependent == Vertical).
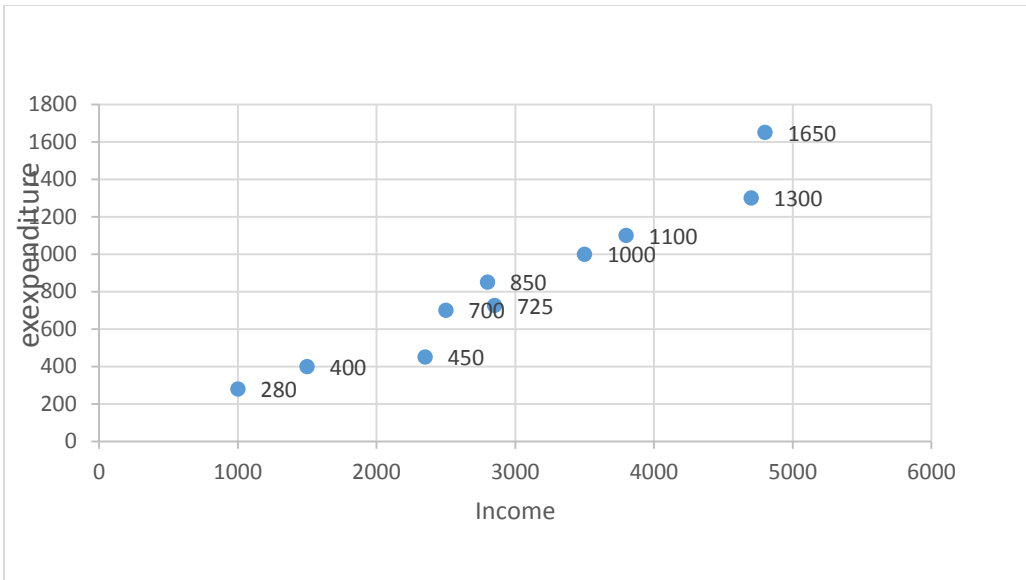
FIGURE 2.8 TYPES OF RELATIONSHIPS DEPICTED BY SCATTER DIAGRAMS
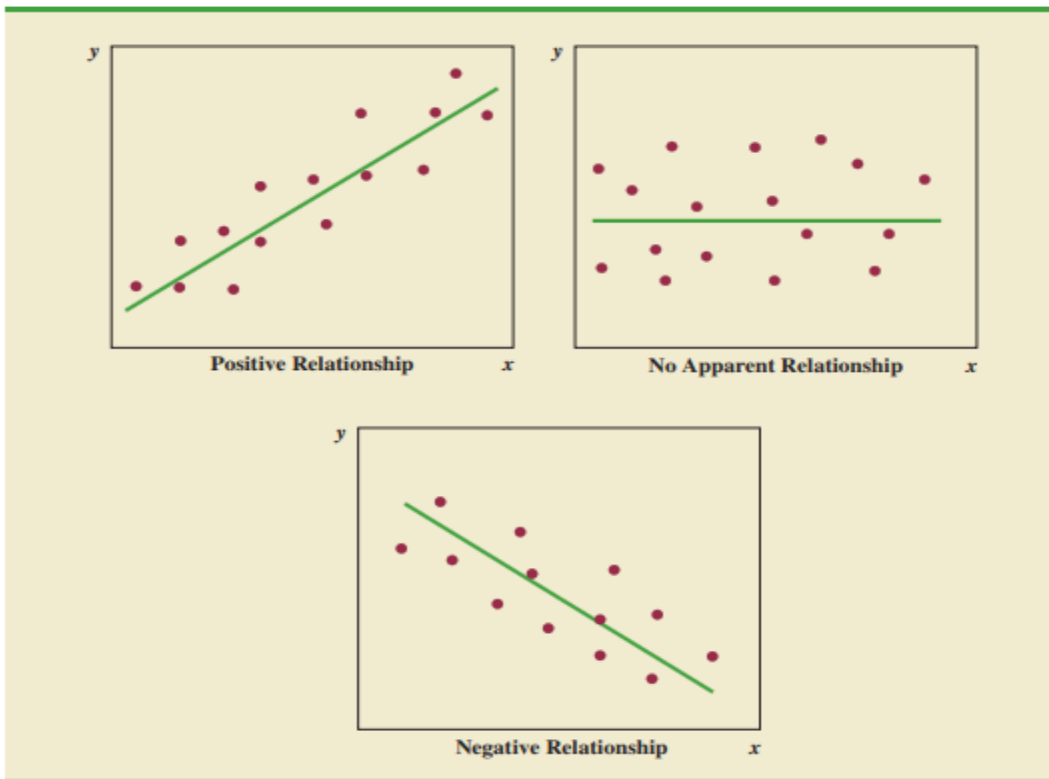
Example 2

The income and food expenditures of ten households is given in the following table. Construct a scatter diagram for this distribution.

Household	Income X	Food expenditure Y
1.	3500	1000
2.	4700	1300
3.	2500	700
4.	3800	1100
5.	1500	400
6.	2800	850
7.	2350	450
8.	4800	1650
9.	1000	280
10.	2850	725

There is a **positive linear relationship** between the income and the daily expenditure