Frequently Asked Questions

1. What do you mean by bivariate data?

Answer:
If we have any data set with two variables, then it is called bivariate data. Bivariate data consists of two *quantitative variables* for each individual.

2. Write examples for bivariate data.

Answer:

- Large health studies of populations it is common to obtain variables such as age, sex, height, weight, blood pressure, and total cholesterol on each individual
- Economic studies may be interested in, among other things, personal income and years of education.
- Most of the university admissions committees ask for an applicant's high school grade point average and standardized admission test scores (e.g., SAT).

3. Which are the different types of relationships exist between two variables in a bivariate data?

Answer:
There can be positive relationship, negative or inverse relationships and then there are variables which have no relationship with one another. A positive relationship would exist if as one variable increases, the other variable increases or, if as one variable decreases, the other variable decreases. And a negative or inverse relationship would exist of as one variable increases, the other variable decreases or vice versa.

4. Give examples for positive and negative relationship between two variables in a bivariate data.

Answer:

- An example of a positive relationship would be between the two variables temperature and population at the beach.
- An example for negative correlation - Often as the amount of time you exercise increases, the less time it takes to run a mile.

5. How to identify independent and dependent variables?

Answer:

Whenever a relationship between two sets of variables is being examined, it is useful to know whether one of the variables depends on the other. Often we can make a judgement about this but sometimes we can't.

Consider a case where a study of comparing the heights of the company employees against their salaries. Common sense would suggest that the height of the company's employee would not depend on the person's salary nor would the annual salary of the company's employee depend on the person's height. In this case it is not appropriate to designate one variable as independent and other as dependent. In the case where the age of the company employees are compared with their annual salaries, we might reasonably expect that the annual salary of an employee would depend on the person's age. In this case, the age of the employee is the independent variable where as the salary of the employee is the dependent variable. Hence independent variable controls the dependent variable.

6. Give an example for raw data.

Answer:

Here we write two variables and values taken by these two variables. Consider the following example.

Number of hours of study and marks scored in the test of nine students are recorded. Since here we have only 09 students, we can write in the raw form as follows.

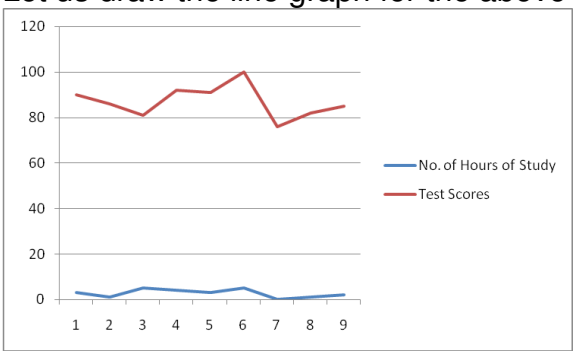
Number of hours of study	Test scores
3	90
1	86
5	84
4	92
3	91
5	100
0	76
1	82
2	85

7. Following table give the number of hours of study and marks score in the test by 9 students. By plotting the bivariate data, verify whether the two variables are associated.

Number of hours of study	Test scores
3	90
1	86
5	84
4	92
3	91
5	100

0	76
1	82
2	85

Answer:
Let us draw the line graph for the above data.

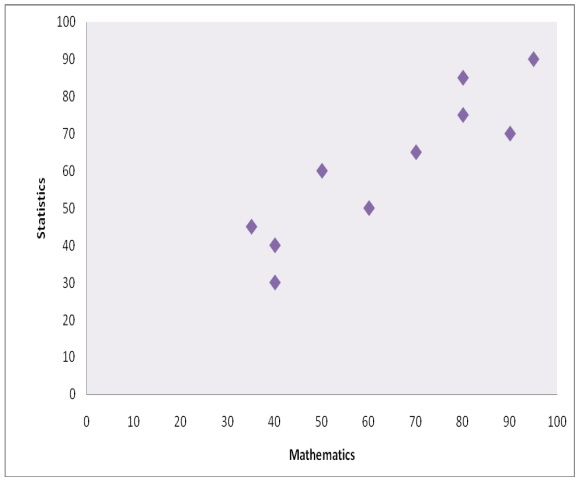


Observe that, there is positive association between number of hours of study and test scores.

8. Following data indicates the marks score by 10 student in Mathematics and Statistics. By plotting the data, verify whether there is any association between two variables.

Marks in Statistics	Marks in Mathematics
45	35
70	90
65	70
30	40
90	95
40	40
50	60
75	80
85	80
60	50

Answer:
Let us plot the given data using scatter diagram.

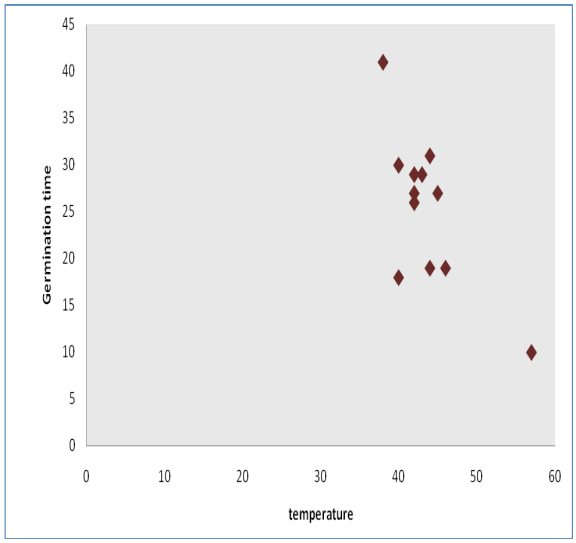


From the scatter diagram, we can see that there is a positive association between the two variables.

9. The table (in the right) shows the soil temperature and the germination time required for the seeds at various places. Verify whether there is any relationship between the two variables by plotting the data

Temperature	Germination time
57	10
42	26
40	30
38	41
42	29
45	27
42	27
44	19
40	18
46	19
44	31
43	29

Answer:
Let us plot the scatter diagram as follows.

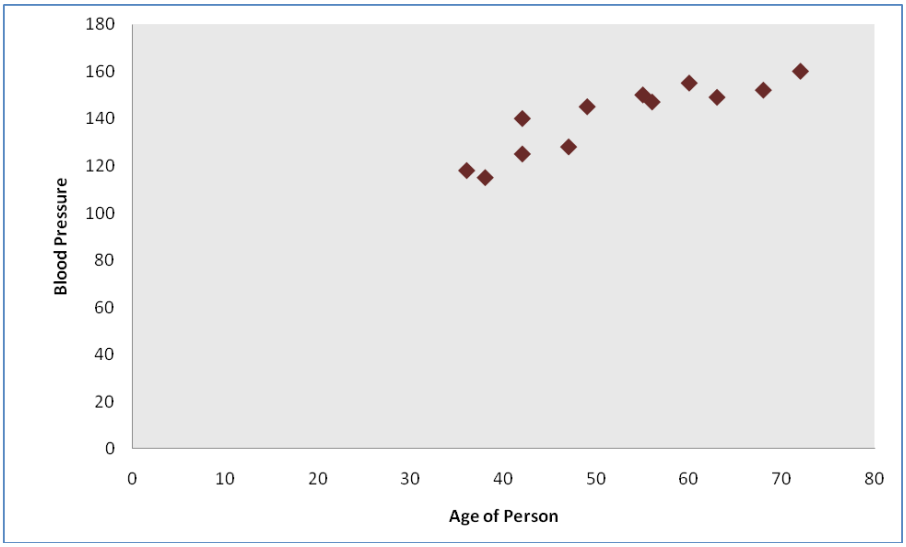


From the above scatter plot we can see that there is slightly negative association between soil temperature and germination time.

10. Following table shows the age and blood pressure of 12 persons. By plotting the data verify whether there is any association between the variables.

Age of a person	Blood Pressure
56	147
42	125
72	160
36	118
63	149
47	128
55	150
49	145
38	115
42	140
68	152
60	155

Answer:



Observe that there is a positive association between blood pressure and age of a person.

11. The following data gives the data regarding the height from which a ball was dropped and its rebound height. Verify whether there is any relationship between these two variables by plotting the data.

Table 8.1 Ball drop height and rebound height

DROP HEIGHT	REBOUND MEASUREMENT 1	REBOUND MEASUREMENT 2	REBOUND MEASUREMENT 3
12	3	6	5
18	7	8	11
24	13	14	16
30	19	18	17
36	20	21	20.5
42	22	21.5	21
48	24	26	25
54	27	28	30
60	25	31	32
66	37	39	38
72	45	44	42

Answer:

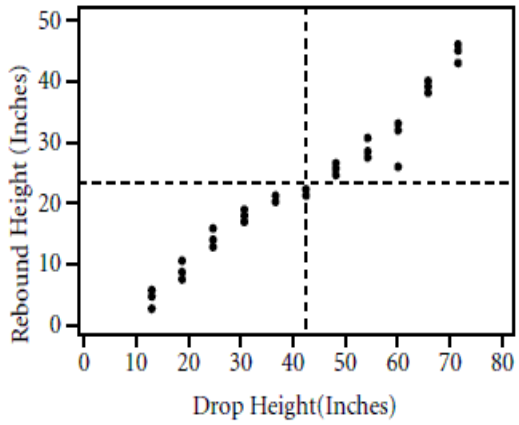


Figure 8.2

The rebound height tends to increase in a linear manner as the drop height increases.

12. What do you mean by contingency table?

Answer:

If there is a huge data, that is number of observations under each variable is very large, then we can write the data in the tabular form either by giving frequencies that each value take when the number repeats or give class interval and see how many will belong to each of the class interval. And this arrangement is known as contingency table.

13. Give an example for contingency table.

Answer:

Suppose that we have two variables, sex (male or female) and handedness (right or left handed). Further suppose that 100 individuals are randomly sampled from a very large population as part of a study of sex differences in handedness. A contingency table can be created to display the numbers of individuals who are male and right-handed, male and left-handed, female and right-handed, and female and left-handed. Such a contingency table is shown below.

	Male	Female	**Total**
Right handed	65	60	125
Left handed	32	43	75
Total	97	103	200

The numbers of the males, females, and right- and left-handed individuals are called marginal totals**. The** grand total**, i.e., the total number of individuals represented in the contingency table, is the number in the bottom right corner.**

14. Here we consider the data regarding the arm span and height of a person given as follows. Construct contingency table using scatter plot.

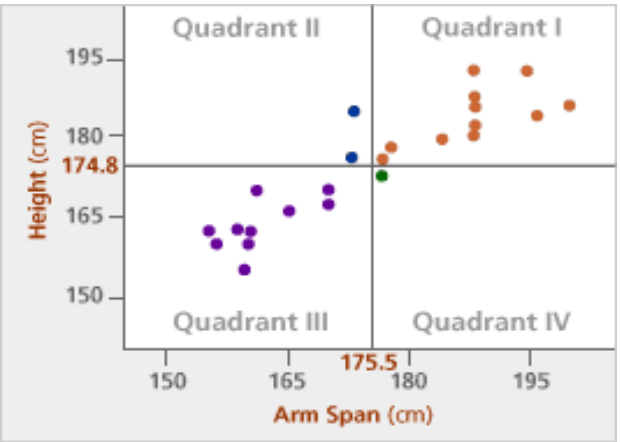
Arm Span	Height	Arm Span	Height
156	162	177	173
157	160	177	176
159	162	178	178
160	155	184	180
161	160	188	188
161	162	188	187
162	170	188	182
165	166	188	181
170	170	188	192
170	167	194	193
173	185	196	184
173	176	200	186

This data we can represent in a contingency table as follows

Since data is sorted by arm span, while constructing the scatter plot, we take arm span along X axis and Height of a person along the Y axis.

Let's take a look at the scatter plot with the quadrants indicated

We find the average height and average arm span using the given data which is equal to 174.8 and 175.5 respectively. Hence we draw the lines parallel to X axis and Y axis at the average points and divide the whole graph into four quadrants as shown below.

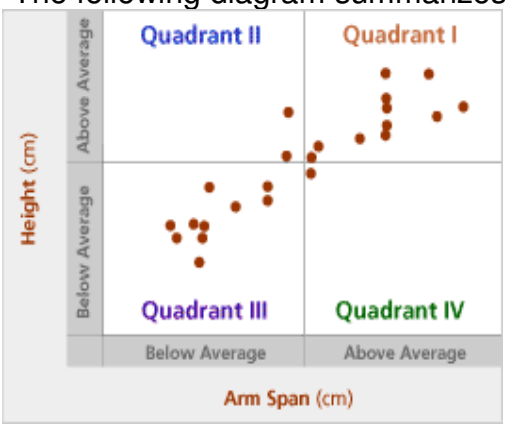


Observe that,

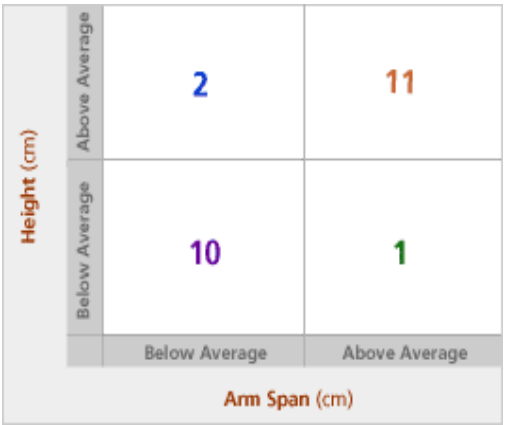
- Quadrant I has points that correspond to people with above-average arm spans and heights.
- Quadrant II has points that correspond to people with below-average arm spans and above-average heights.

- Quadrant III has points that correspond to people with below-average arm spans and heights.
- Quadrant IV has points that correspond to people with above-average arm spans and below-average heights.

The following diagram summarizes this information:



If you count the number of points in each quadrant on the scatter plot, you get the following summary, which is called a contingency table



15. distinguish between Univariate and Bivariate data

Answer:

Univariate Data	Bivariate Data
• involving a **single variable**	• involving **two variables**
• does not deal with causes or relationships	• deals with causes or relationships
• the major purpose of Univariate analysis is to describe	• the major purpose of bivariate analysis is to explain

• central tendency - mean, mode, median • dispersion - range, variance, max, min, quartiles, standard deviation. • frequency distributions • bar graph, histogram, pie chart, line graph, box-and-whisker plot	• analysis of two variables simultaneously • correlations • comparisons, relationships, causes, explanations • tables where one variable is contingent on the values of the other variable. • independent and dependent variables