



Sri Lanka Institute of Information Technology

Year 02 – Semester II – 2023

Probability and Statistics – IT2110

Tutorial 09

Scatter plot & Types of Relationships

- 1) A small study is conducted involving 17 infants to investigate the association between gestational age at birth (in weeks) and birth weight (in grams). Here, researcher is interested in identifying whether birth weight of an infant has an effect from gestational age. [This data set extracted from “*Boston University School of Public Health*” web site.]

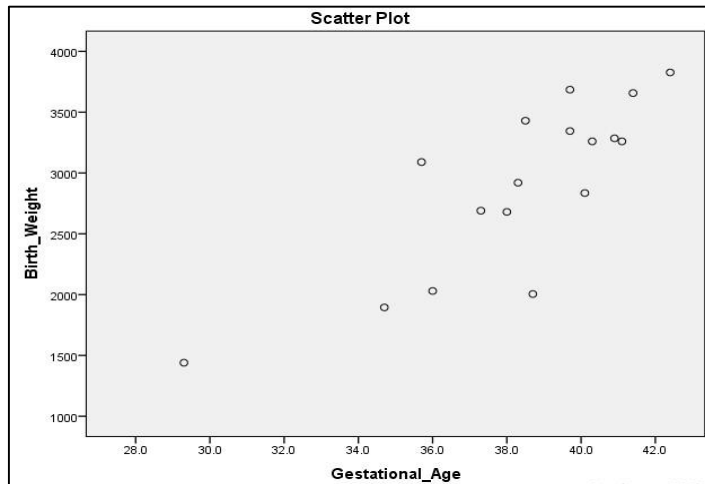
Infant ID	Gestational Age (In weeks)	Birth Weight (In grams)
1	34.7	1895
2	36.0	2030
3	29.3	1440
4	40.1	2835
5	35.7	3090
6	42.4	3827
7	40.3	3260
8	37.3	2690
9	40.9	3285
10	38.3	2920
11	38.5	3430
12	41.4	3657
13	39.7	3685
14	39.7	3345
15	41.1	3260
16	38.0	2680
17	38.7	2005

- i. What is the dependent variable (response variable) & independent variable (predictor variable)?

Dependent Variable :

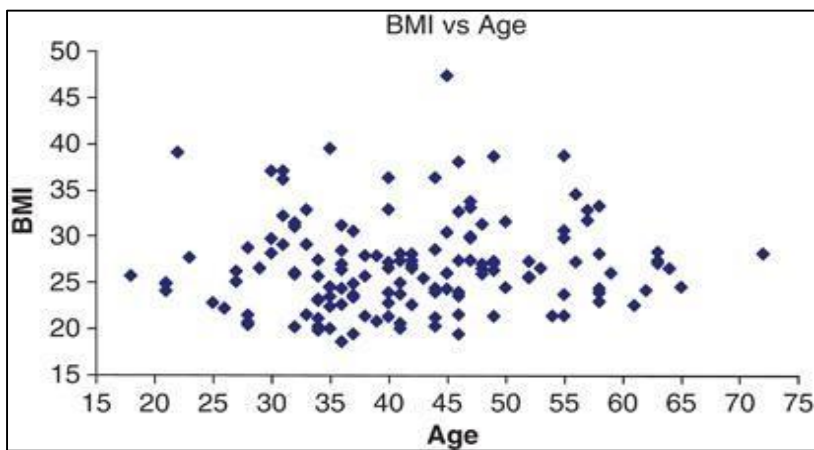
Independent Variable :

- ii. You have given the scatter plot of gestational age and birth weight. Identify whether there is any relationship between gestational age and birth weight.



- 2) Identify the type of relationship.

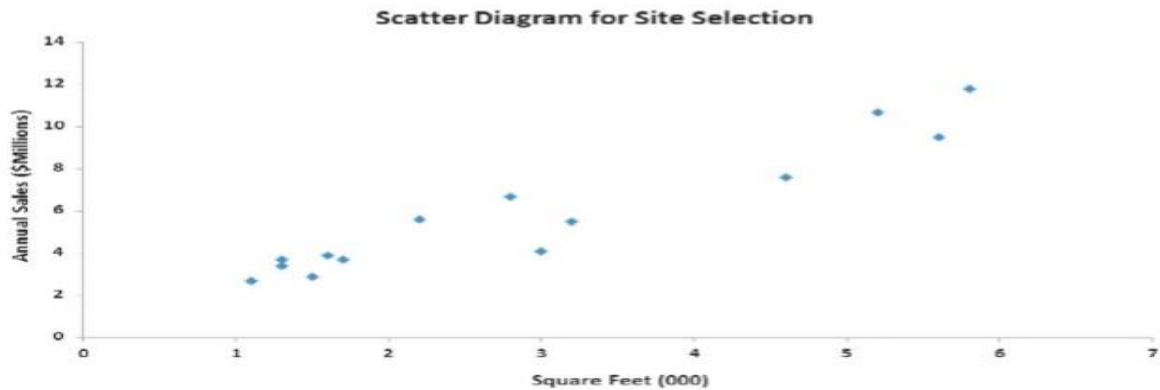
i.



ii.



- 3) The business objective of the director of planning of Sunflowers Apparel Ltd., is to forecast annual sales for all new stores, based on store size. To examine the relationship between the store size in square feet and its annual sales, data were collected from a sample of 14 stores. Figure below displays the scatter plot for the data.



R outputs of the regression model are shown below

Regression Model

Coefficients

Intercept	Square feet
0.9645	1.6699

Analysis of Variance Table

Response: Annual Sales

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Square feet	A	105.7476	E	G	0.000 ***
Residuals	B	11.2067	F		
Total	C	D			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

- What can be concluded using the scatterplot?
- Find values marked A, B, C, D, E, F and G in the ANOVA table (Show workings).
- State the estimated regression equation in the form of $\hat{Y} = \hat{\alpha} + \hat{\beta}X$ and state how much more sales are expected if they increase the store size by 1,000 square feet.
- Following information is given;

$$\sum X = 40.9$$

$$\sum Y = 81.8$$

$$\sum XY = 302.30$$

Standard deviations are $S_x = 1.71$ and $S_y = 2.98$

Where X is the square feet and Y is the annual sales

Calculate the Pearson's linear coefficient of correlation between the two variables.

- Calculate coefficient of determination and interpret it.
- Use the regression equation to predict annual sales, if store size is 10,000 square feet.