Sri Lanka Institute of Information Technology

Year 02 – Semester II – 2020

Probability and Statistics – IT2110

Tutorial 09

Scatter plot & Types of Relationships

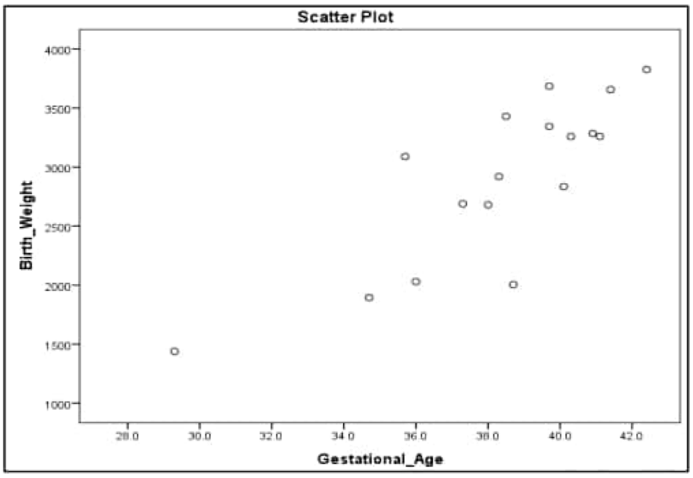
1) A small study is conducted involving 17 infants to investigate the association between gestational age at birth (in weeks) and birth weight (in grams). Here, researcher is interested in identifying whether birth weight of an infant has an effect from gestational age. [This data set extracted from "***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 : Birth Weight
Independent Variable : Gestational Age

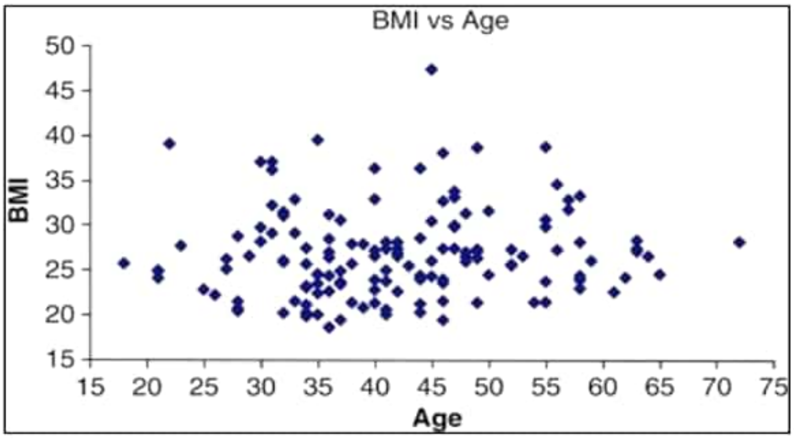
ii. You have given the scatter plot of gestational age and birth weight. Identify whether there is any relationship between gestational age and birth weight.



There is a positive linear relationship between Birth weight and Gestational age.

2) Identify the type of relationship.

i.



There is no linear relationship between BMI and Age

ii.

There is a negative linear relationship between Driver Age and Sign legibility Distance.

(03)

i) There is a positive linear relationship between Annual sales and ~~amount of squre feet~~ Store size.

ii)

Analysis of variance table

Response : Annual sales.

	df	Sum sq	Mean sq	F valu	P(>F)
Squre feet	A	105.7476	E	G	0.000
Residuals	B	11.2067	F		
Total	C	D			

C

total $df = n-1$

$C = 14-1$

$C = 13$

D.

$D = SSR + SSE$

$D = 105.7476 + 11.2067$

$D = 116.9543$

A

$A =$ no. of parameters -1

$A = 2-1$

$A = 1$

E

$E = \frac{SSR}{df}$

$E = \frac{105.7476}{1}$

$E = 105.7476$

B.

$A + B = C$

$1 + B = 13$

$B = 12$

F

$$F = \frac{11.2067}{B}$$

$$F = \frac{11.2067}{12}$$

$$F = 0.9339$$

G

$$G = \frac{MSR}{MSE}$$

$$G = \frac{105.7476}{0.9339}$$

$$G = 113.2323$$

iii) Estimation regression equation is,

$\widehat{\text{Annual Sales}} = 0.9645 + 1.6699 \times$ (~~Squre feet~~ store size).

* if they increse the size by 1000 squre unit (by 1 unit in the graph), annual sales will be increase by 1.6699 million dollars.

iv)

$$r_{xy} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{(n-1) S_x S_y}$$

[This equation can also use to calculate correlation]

$$= \frac{302.30 - 14 \times \frac{40.9}{14} \times \frac{81.8}{14}}{13 \times 1.71 \times 2.98}$$

$$= \frac{302.30 - 238.97}{66.25}$$

$$= 0.9559$$

vi) $\widehat{\text{Annual Sales}} = 0.9645 + 1.6699 \times (\text{Store size})$ ~~Square feet~~

When Store size ~~Square feet~~ = 10 000 Sf.

$\widehat{\text{Annual sales}} = 0.9645 + 1.6699 \times 10$

scale of the "x" axis is in '000'.

$= \underline{\underline{17.6635}}$ \$ million.

v) * $R^2 = r^2 = (0.9559)^2 = 0.9138.$

$\therefore$ As a percentage $R^2 = \underline{\underline{91.38\%}}$ (0.9138×100)

* Interpretation :- 91.38% of variation in annual sales has enplained by Store size in the model.

Scanned with CamScanner