భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

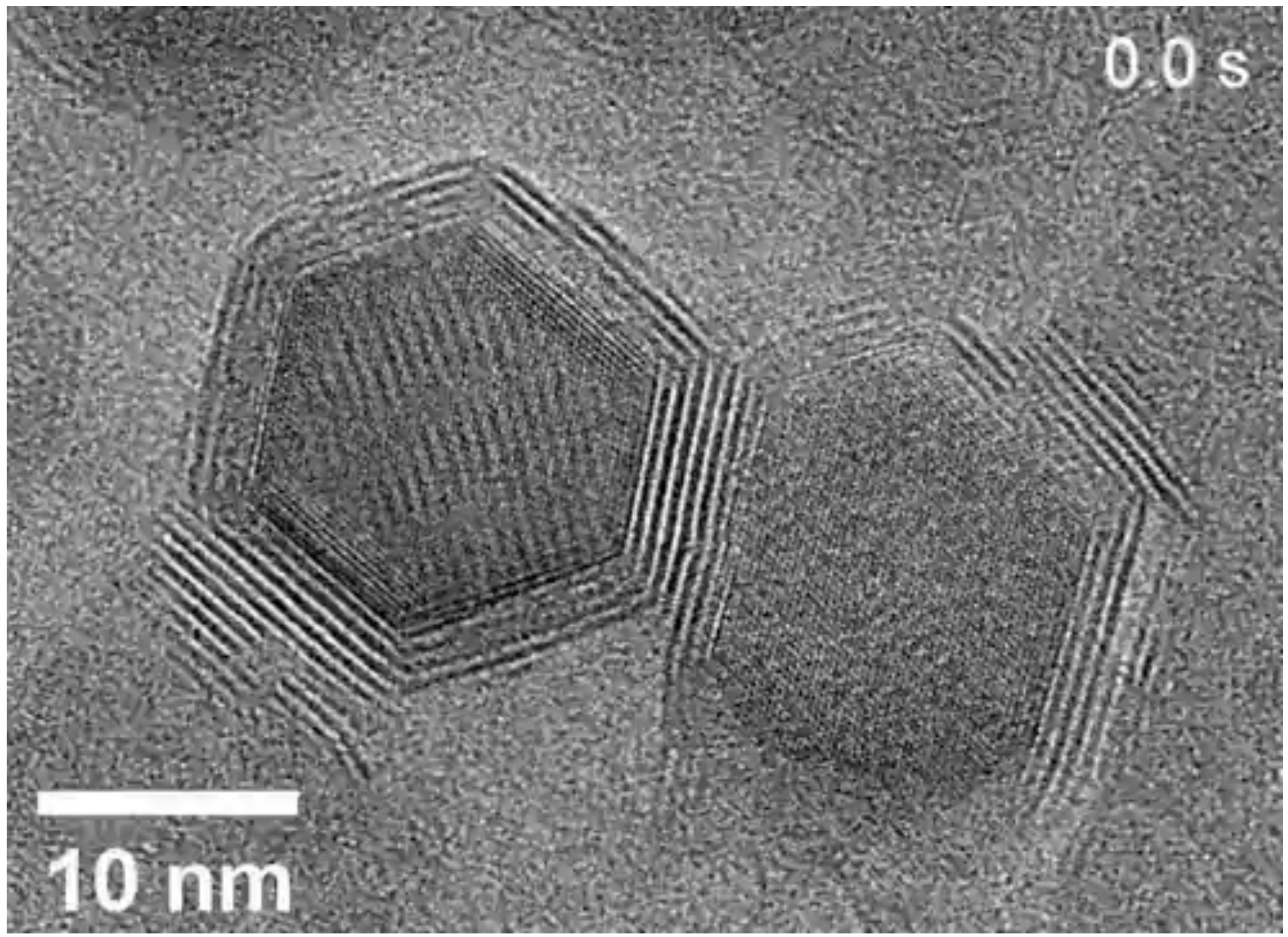
Biostatistics
BT2023

Lecture 11
Correlation and regression

Himanshu Joshi
12 September 2023

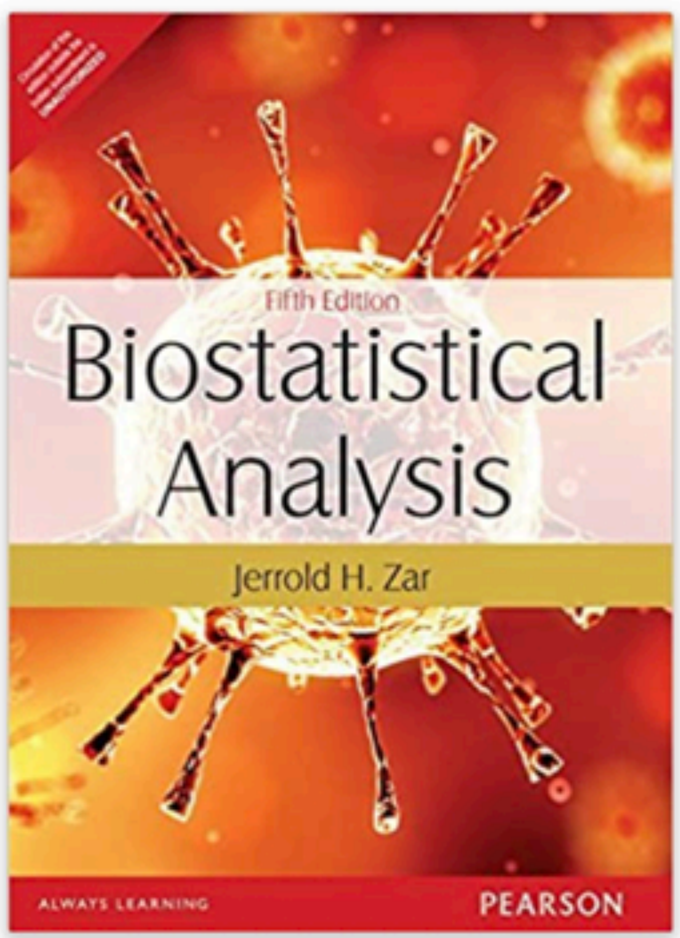
భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Defect-mediated ripening of two Cd-CdCl2 core-shell nanoparticles



Reference Nature Communications
volume 13, Article number: 2211 (2022)

Biostatistical Analysis
Jerrold H Zar
Pearson Publishers



See this image

Biostatistical Analysis, 5e Paperback – 1 January 2014

by Zar (Author)

★★★★☆ 48 ratings

See all formats and editions

Paperback
₹655.00

1 Used from ₹700.00
30 New from ₹655.00

Save Extra with 3 offers

Cashback (3): Get 5% back with Amazon Pay ICICI Bank credit card for Prime members. 3% back for others | See All

No Cost EMI: Avail No Cost EMI on select cards for orders above ₹3000 | Details

See 1 more

Covariance

Covariance in two quantities variable X and Y on a given set is given by

$$Cov.\,(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n}(x_i)\frac{1}{n}\sum_{i=1}^{n}(y_i)$$

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

Correlation coefficient

$$r = \frac{\sum_{i=1}^{n} \frac{(x_i - \bar{x})}{\sigma_x} \frac{(y_i - \bar{y})}{\sigma_y}}{n}$$

$$\implies r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$$where\ \sigma_{xy} = Cov(X, Y)$$

Properties of the correlation coefficient

1. It will range from -1 to +1
2. Measures the closeness of the fit

$$dx = x - \bar{x}$$

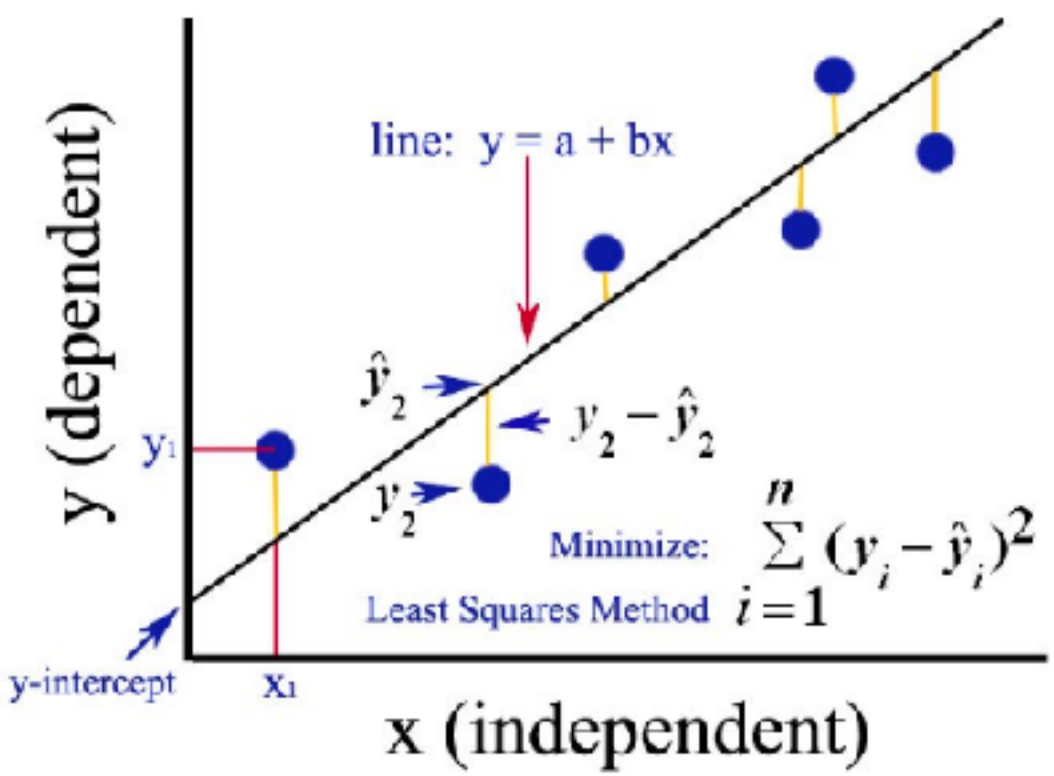
$$dy = y - \bar{y}$$

$$r_{xy} = \frac{\sum_{i=1}^{n} dxdy - \left(\sum dx \sum dy\right)}{\sqrt{\sum dx^2 - \frac{\left(\sum dx\right)^2}{n}} \times \sqrt{\sum dy^2 - \frac{\left(\sum dy\right)^2}{n}}}$$

Linear regression by least square method

In least square method, two conditions needs to be satisfied,

$$\sum_{i=0}^{n}(y_i - \hat{y}) = 0$$



Normal equations

$$\sum y = na + b\sum x$$

$$\sum xy = a\sum x + b\sum x^2$$

$$b_{yx} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

Correlation and Regression

$$b_{yx} = \frac{n\sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

$$= \frac{Cov(X,Y)}{\sigma_x}$$

$$= r\frac{\sigma_y}{\sigma_x}$$

$$r = \frac{\sum_{i=1}^{n} \frac{(x_i - \bar{x})}{\sigma_x}\frac{(y_i - \bar{y})}{\sigma_y}}{n}$$

$$\implies r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$$where\ \sigma_{xy} = Cov(X,Y)$$

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

The regressor and response

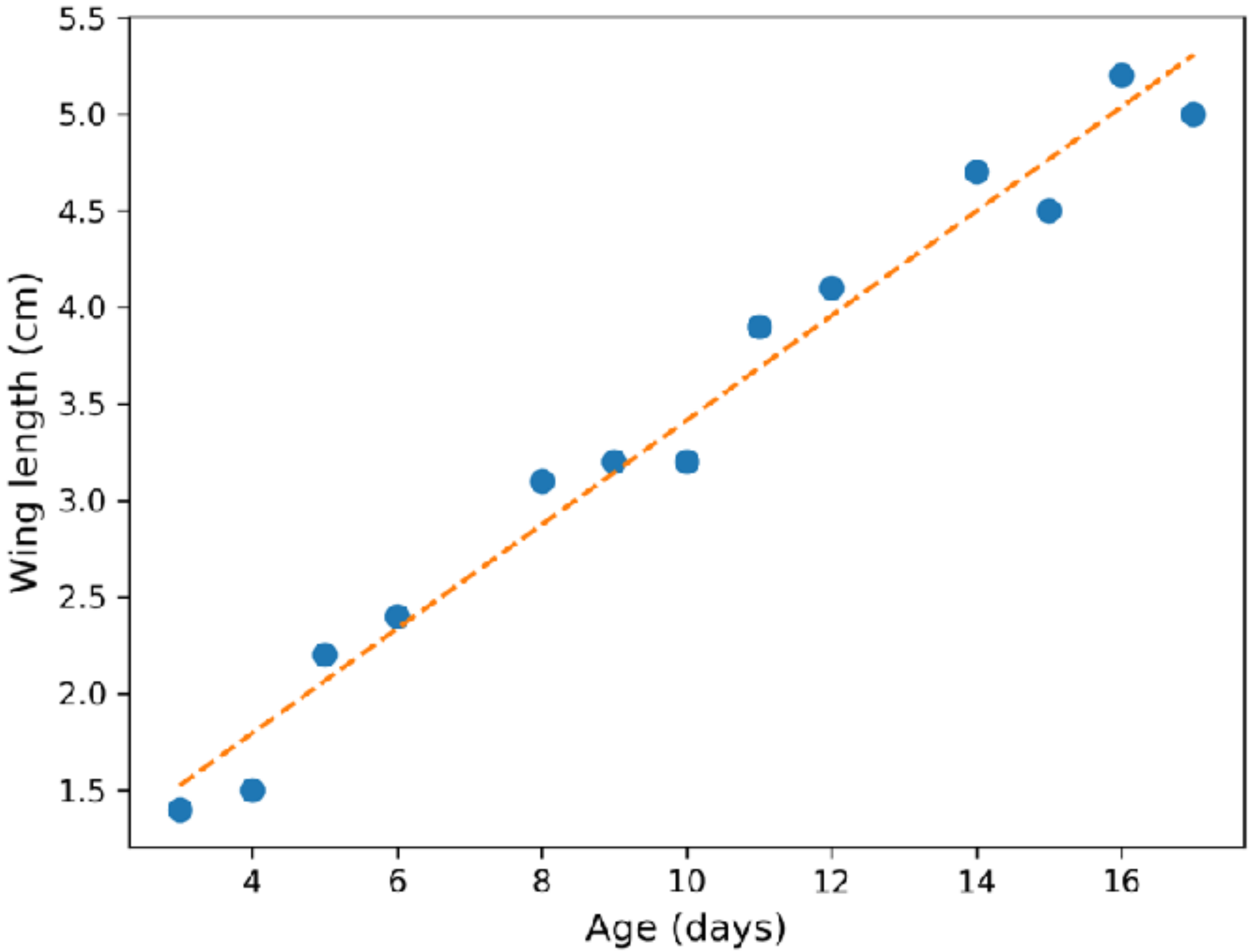
When there is a possibility of functional dependence of one variable to another, such relationship is called regression. The variables are called independent and dependent variables but it doesn’t a cause and effect relationships between two variable.

Consider the following examples;

1. Blood pressure and the age of the patient
2. Length of arm vs length of legs

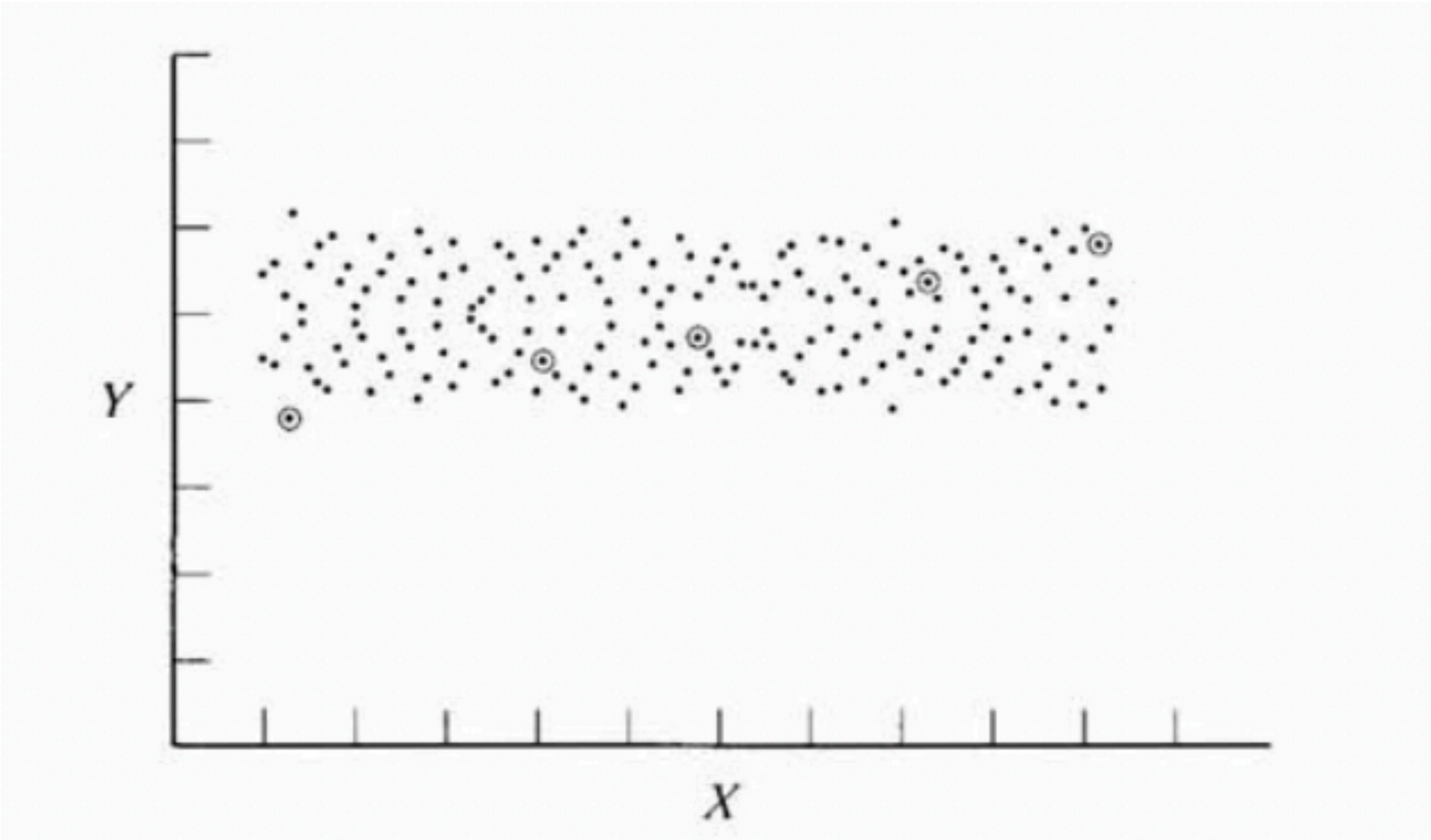
భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Caution with regression



భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Caution with regression



భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Next Class

2:30 PM Friday, 15 September 2022