STAT 231

Nov 16, 2016

No Tutorial today.

Next week → 6 pm Tutorial DC 1351.

Roadmap

Simple Linear Regression Model

(i) Definitions

(ii) Different Assumptions of the model.

(iii) Estimating unknown parameters

→ Method of ML.

→ Method of Least Squares

Y = response variate, a r.v.

We try to explain the variability of Y.

We use an explanatory variable (X, which is known) to

(i) Estimate the relationship between X and Y

(ii) Explain the variability of Y using X

X = independent variable.

Examples:

(i) Y = STAT 231 score.

X = STAT 230 score.

Part of the variability of STAT 231 score can be explained by your STAT 230 score.

(ii) Y = lifetime income.

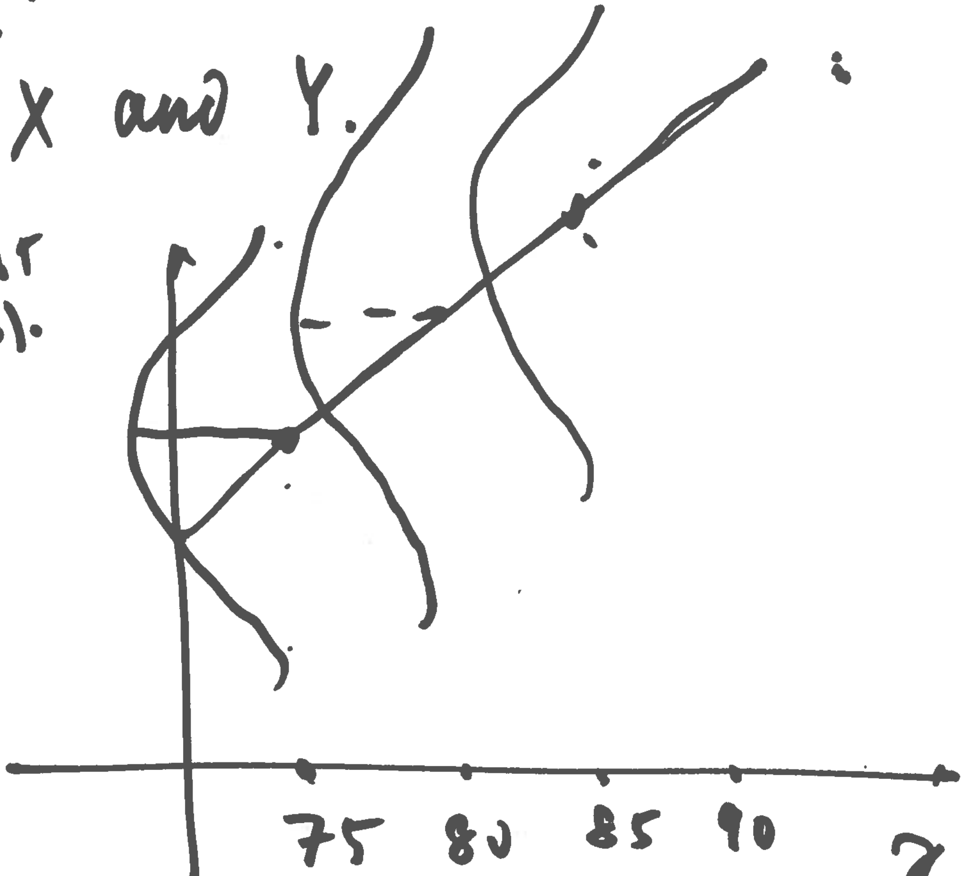
X = man/woman.

(iii) Y = lifetime income

X = # of hours we spend on Facebook.

Note: We make some assumptions of X and Y.

STAT 231.



x = STAT

For each value of $X = x$, Y's are Normally distributed with mean $\mu(x)$ and variance σ^2, and independent.

with

$$\mu(x) = \alpha + \beta x.$$

SIMPLE LINEAR REGRESSION MODEL

We are assuming that σ^2 is independent of the value of X.

HOMOSCEDASTICITY ASSUMPTION

Objective: To estimate α, β, σ from our sample.

SLRM

$$Y_i \sim G\left(\alpha + \beta x_i, \sigma^2\right)$$

$i = 1, \ldots n.$

where α, β, σ are unknown parameters

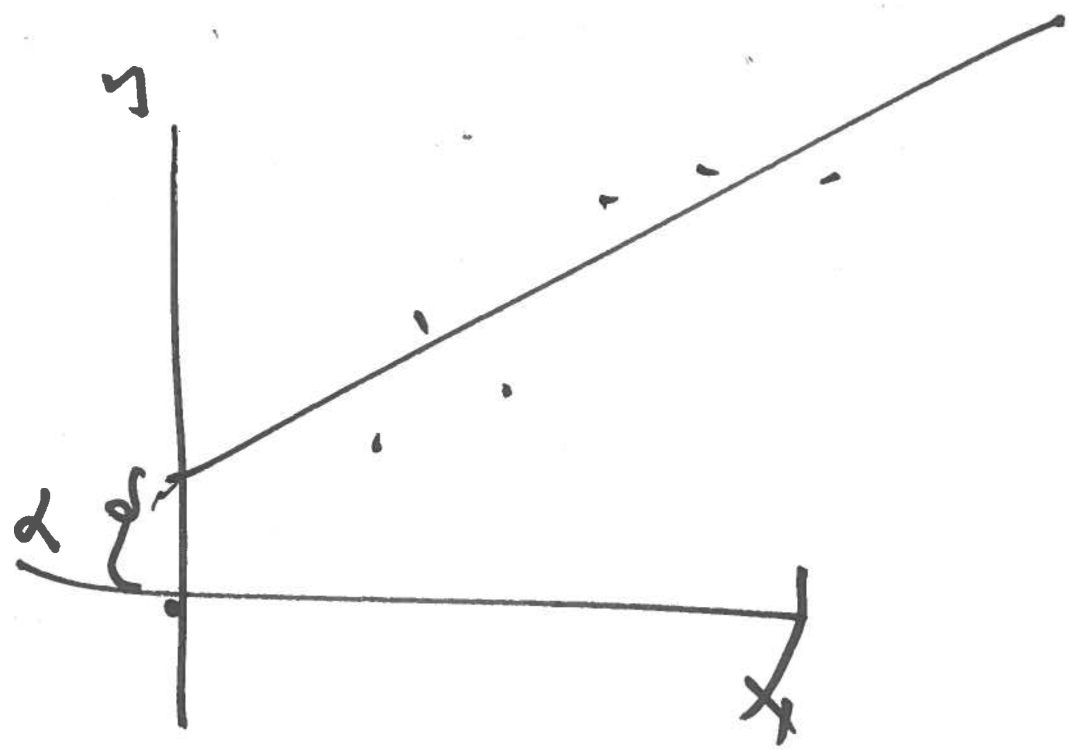
α = Population Average of the Y values when $X = 0$

β = Slope of the line

$\Rightarrow$ If the STAT 230 score $\uparrow$ by 1 unit

the population average STAT 231

score $\uparrow$ β units



So

$$Y_i \sim G(\alpha + \beta x_i, \sigma)$$

$i = 1, \ldots n$

<u>indep.</u>

Sample: $(x_1, y_1), \ldots (x_n, y_n)$

$$f(y_i) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}\left(y_i - (\alpha + \beta x_i)\right)^2}$$

$$L(\alpha, \beta, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}\left(y_i - (\alpha + \beta x_i)\right)^2}$$

$$L(\alpha, \beta, \sigma) = \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-\frac{1}{2\sigma^2} \sum (y_i - \alpha - \beta x_i)^2}$$

$$l(\alpha, \beta, \sigma) = -\frac{n}{2} \ln 2\pi - n \ln \sigma - \frac{1}{2\sigma^2} \sum (y_i - \alpha - \beta x_i)^2$$

do. $$\left.\begin{aligned} \frac{\partial l}{\partial \alpha} &= 0 \\ \frac{\partial l}{\partial \beta} &= 0 \\ \frac{\partial l}{\partial \sigma} &= 0 \end{aligned}\right\}$$

$$\frac{\partial l}{\partial \alpha} = 0 \Rightarrow$$

$$-\frac{2}{2\sigma^2}\sum(y_i - \alpha - \beta x_i)(-1) = 0$$

$$\sum(y_i - \alpha - \beta x_i) = 0$$

$$\sum y_i - \sum \alpha - \beta \sum x_i = 0$$

$$n\bar{y} - n\alpha - \beta n\bar{x} = 0$$

$$\boxed{\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}}$$

$$\frac{\partial l}{\partial \beta} = 0 \Rightarrow$$

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}.$$

$$\frac{\partial l}{\partial \sigma} = 0 \Downarrow$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum (y_i - \hat{\alpha} - \hat{\beta} x_i)^2.$$

$$= \frac{1}{n} \left[S_{yy} - \hat{\beta} S_{xy} \right].$$

$$S_{yy} = \sum (y_i - \bar{y})^2.$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x}$$

$$\hat{\beta} = S_{xy}/S_{xx}.$$

$$\hat{\sigma}^2 = \frac{1}{n}\left[S_{yy} - \hat{\beta}\, S_{xy}\right]$$

The Maximum Likelihood Estimator of the Regression model.

Simple $\Rightarrow$ One Explanatory variable

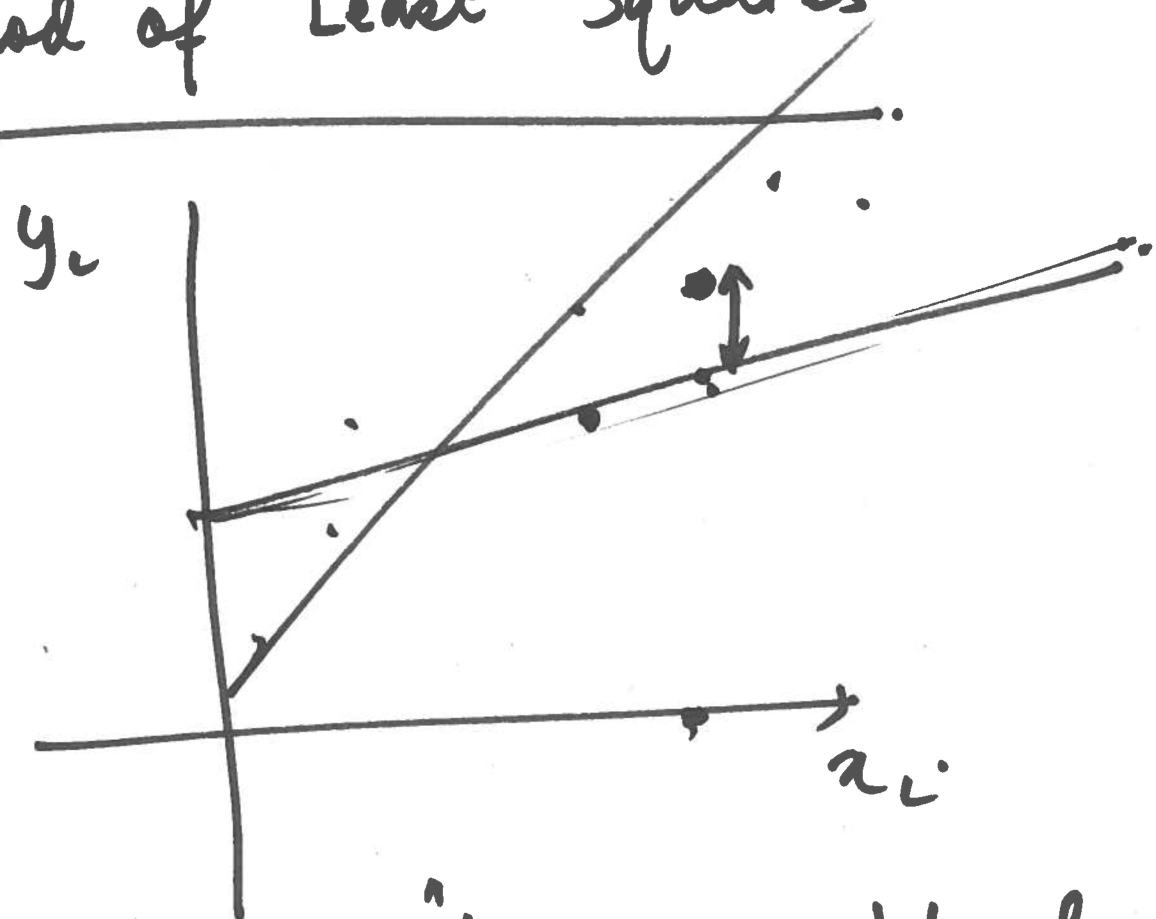
Linear $\Rightarrow$ $\mu(x) = \alpha + \beta x.$

NON-LINEAR REGRESSION MODELS

if $\mu(x)$ is non-linear.

Method 2   Estimating α and β from your sample.

Method of Least Squares



Try to fit the "best possible line" to the data.

The line is chosen in such a way that SS Errors are minimized

$$\min_{\hat{\alpha},\hat{\beta}} \sum \left(y_i - (\hat{\alpha} + \hat{\beta} x_i) \right)^2$$

$$\boxed{\begin{aligned} \hat{\alpha} &= \bar{y} - \hat{\beta}\,\bar{x} \\ \hat{\beta} &= S_{xy} / S_{xx} \end{aligned}}$$

The MLEs are called least square estimates because they are the same.