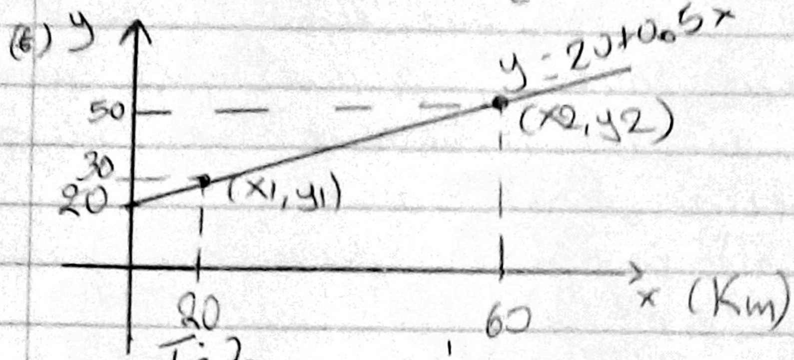
Παλινδρόμηση - (Απλή) Γραμμική Παλινδρόμηση
(Regression Analysis - (simple) Linear Regression)



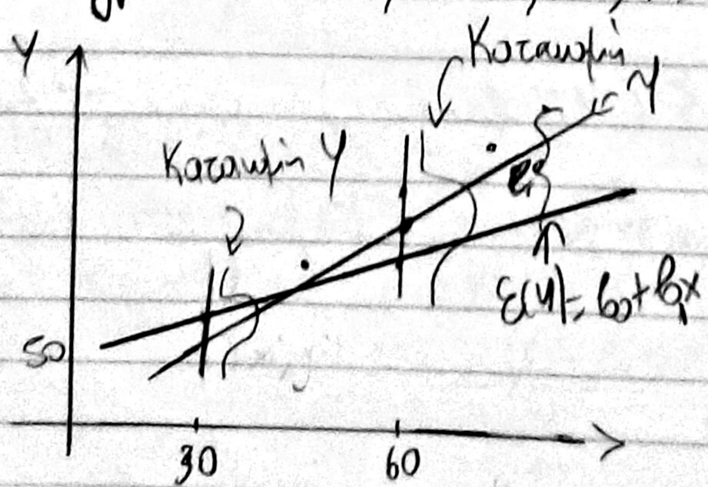
$y = f(x) = 20 + 0{,}5x$

Τέλεια σχέση

Μήνας (i) Παραγωγής	Μέγεθος (x) Παραγγελίας	Ώρες (y) Εργασίας
1	30	73
2	20	50
3	60	128
4	80	170
5	40	87
6	50	108
7	60	135
8	30	69
9	70	148
10	60	132

Παράδειγμα EBO, $n = 10$, (x_i, y_i)

Διάγραμμα Διασποράς



Η στατιστική σχέση δεν είναι μια τέλεια σχέση!

Απλή Γραμμική Παλινδρόμηση

Για (x_i, y_i), $i=1,..,n$, των x: ελεγχόμενη, y: εξαρτημένη
$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, όπου β_0, β_1: συντελεστές παλινδρόμησης (παράμετροι) κ' ε_i: τυχαία σφάλματα, ανά δύο ανεξάρτητες τ.μ. με $E(\varepsilon_i)=0$, $Var(\varepsilon_i)=\sigma^2 \ \forall x$.

Έστι: $E(y_i) = \beta_0 + \beta_1 x_i$ ← άγνωστη συν. παλινδρόμησης

$Var(y_i) = \sigma^2 \ \forall x$

$y_i \sim (E(y_i), \sigma^2)$

Εκτίμηση Παραμέτρων με τη μέθοδο των Ελαχίστων Τετραγώνων

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - E(y_i))^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

$\varepsilon_i = y_i - (\beta_0 + \beta_1 x_i)$
$= y_i - E(y_i)$
άγνωστη απόκλιση

Εκτιμητές ελ. τετρ. $\hat{\beta}_0, \hat{\beta}_1 \rightarrow \min Q(\beta_0, \beta_1)$

$$\frac{dQ}{d\beta_0} = -2 \sum_{i=1} (y_i - \beta_0 - \beta_1 x_i) \ (=0) \rightarrow \sum_{i=1} y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i \quad (1^{\eta})$$

$$\frac{dQ}{d\beta_1} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) x_i \ (=0)$$ — Καν. Εξ.

$$(=0) \rightarrow \sum_{i=1}^{n} x_i y_i = \hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 \quad (2^{\eta})$$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

$2^{η} \to \sum_{1}^{n} x_i y_i = (\bar{y} - \hat{\beta}_1 \bar{x}) \sum_{1}^{n} x_i + \hat{\beta}_1 \sum_{1}^{n} x_i^2$

$\to \sum_{1}^{n} x_i y_i = \frac{\sum x_i \sum y_i}{n} - \hat{\beta}_1 \left(\frac{\sum x_i}{n}\right)^2 + \hat{\beta}_1 \sum_{1}^{n} x_i^2$

$\to \hat{\beta}_1 = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum_{1}^{n} x_i^2 - \frac{(\sum_{1}^{n} x_i)^2}{n}}$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum x_i \sum^{n} y_i}{n}}{\sum_{1}^{n} x_i^2 - (\sum x_i)^2 / n} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$A = \begin{pmatrix} \frac{d^2 \alpha}{d\beta_0^2} = 2n & \\ \frac{d^2 \alpha}{d\beta_0 d\beta_1} = 2\sum x_i \xrightarrow{} 2\sum x_i & \frac{d\alpha^2}{d\beta_1^2} = 2\sum x_i^2 \end{pmatrix} \to$$

$\to |A| = 4n \sum x_i^2 - 4(\sum x_i)^2 = 4n \sum (x_i - \bar{x})^2 > 0$

άρα έχουμε ελάχιστο

$\boxed{\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}}$

$\Sigma\epsilon i = 0$, $\bar{y} = \bar{\hat{y}}_i$

$\Sigma \epsilon_i x_i = 0$

$$\boxed{\hat{\beta}_1 = \frac{\sum^{n} x_i y_i - \frac{\sum^{n} x_i \sum y_i}{n}}{\sum_{1}^{n} x_i^2 - (\sum x_i)^2 / n} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}}$$

Για το πα του ΕΒ

$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$, Εφο: $\hat{\beta}_1 = 2.0$, $\hat{\beta}_0 = 10.0 \to \hat{y}_0 = 10 + 2.0 x_0$ π.χ. $x_0 = 55 \to \hat{y}_0 = 120$

για τα x_i των δεδομένων: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

και e_i: υπόλοιπα, $e_i = y_i - \hat{y}_i$, $i = 1, \ldots, n$

$= y_i - (\hat{\beta}_0 - \hat{\beta}_1 x_i)$ γνωστά

Ιδιότητες της εκτιμώμενης ευθείας παλινδρόμησης

i) Το κέντρο βάρους των παρατηρήσεων $(\bar{X},\bar{Y})$ είναι σημείο της ευθείας γιατί για $x_0=\bar{x}$, $\hat{Y}_0=\hat\beta_0+\hat\beta_1\bar{x}=\bar{Y}-\hat\beta_1\bar{x}+\hat\beta_1\bar{x}=\bar{Y}$

ii) $\sum e_i=0$ γιατί από 1η Κ.Ε.:

$$\sum_1^n e_i=\sum_1^n(Y_i-\hat Y_i)=\sum_1^n Y_i-\sum_1^n(\hat\beta_0+\hat\beta_1x_i)$$

$$=\sum Y_i-n\hat\beta_0-\hat\beta_1\sum X_i$$

$$=0$$

iii) $\sum_1^n e_ix_i=0$ γιατί από 2η Κ.Ε. είναι:

$$\sum_1^n e_ix_i=\sum(Y_i-\hat Y_i)x_i=\sum_1^n x_iY_i-\sum_1^n(\hat\beta_0+\hat\beta_1x_i)x_i$$

$$=\sum_1^n X_iY_i-\hat\beta_0\sum_1^n x_i-\hat\beta_1\sum_1^n x_i^2$$

σ^2, $\hat\sigma^2=S^2=\frac{1}{n-2}\sum_1^n e_i^2=\frac{1}{n-2}\sum_1^n(Y_i-\hat Y_i)^2$

Αμερόληπτος
εκτιμητής του σ^2