ΣΤΑΤΙΣΤΙΚΗ

Διάλεξη 14η

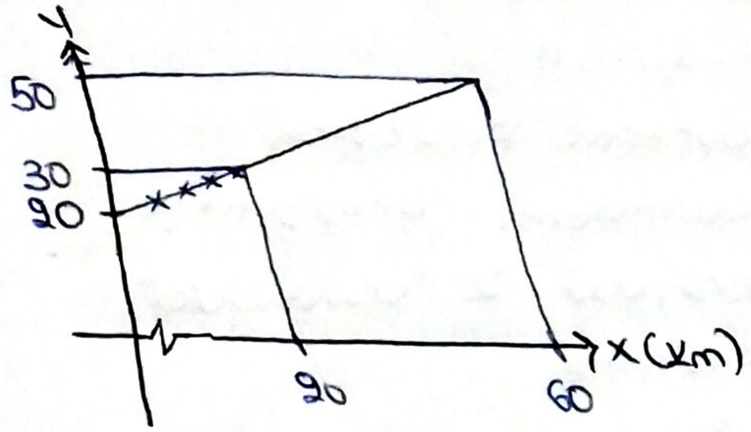
25/04/2018

Ανάλυση παλινδρόμησης (Regression Analysis) / Γραμμική παλινδρόμηση (Linear Regression)

■ Διερεύνηση ύπαρξης σχέσεων μεταξύ μεταβλητών (ποσοτικών) και συναφή προβλήματα.

$y=f(x)$, x = ανεξάρτητη μεταβλητή

y = εξαρτημένη μεταβλητή

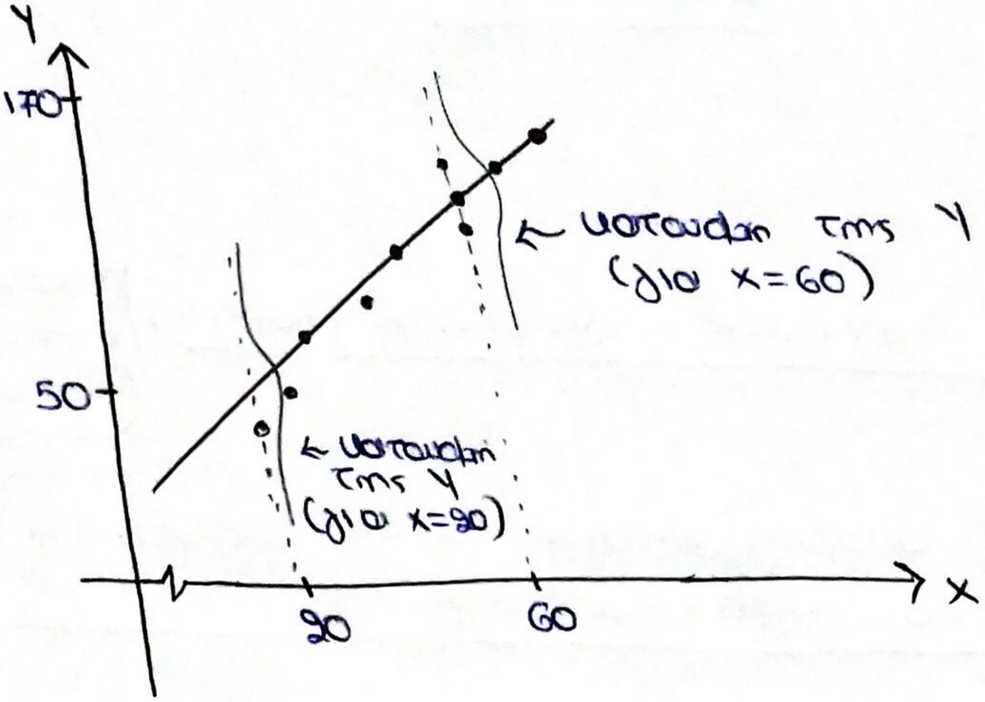


Ενοικίαση αυτοκινήτων (€)

Παράδειγμα 5.2

Αριθμός Παραγ. (i)	Μέγεθ. παραγ. (x)	Ώρες εργασίας (y)
1	30	73
2	20	50
3	60	128
4	80	170
5	40	87
6	50	106
7	60	135
8	30	69
9	70	148
10	60	132

Διάγραμμα Διασποράς (κ' στατιστικής σχέσης)



Απλή γραμμική παλινδρόμηση (υποθέσεις):

$Y_i = B_0 + B_1 x_i + \varepsilon_i \quad , i = 1, \dots, n$

x_i: τιμές της ανεξάρτητης μεταβλητής

Y_i: τιμές της εξαρτημένης μεταβλητής

ε_i: τυχαία σφάλματα, ανά 2 "ασυσχέτιστα" τυχαίες μεταβλητές.

$E(\varepsilon_i)=0 \, , \; Var(\varepsilon_i)=\sigma^2 \; (\forall x)$

Τα B_0 και B_1 είναι συντελεστές παλινδρόμησης (παράμετροι)

$\underbrace{E(Y_i) = B_0 + B_1 x_i}$ και $Var(Y_i) = \sigma^2 \, , \forall x$

η άγνωστη σχέση είναι αυτή, διότι τα B_0 και B_1 είναι άγνωστα

Με βάση το τύπο ορίζουμε τα παρακάτω έχουμε: $\varepsilon_i = Y_i - E(Y_i) = Y_i - B_0 - B_1 x_i$ (οι οποίες είναι άγνωστες ποσότητες).

$\Downarrow$

$E(Y_i) = B_0 + B_1(x_i)$ (= Άγνωστη συνάρτηση παλινδρόμησης)

Εκτίμηση παραμέτρων (β_0, β_1) με την μέθοδο των ελαχίστων τετραγώνων.

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

- $$\frac{\partial Q}{\partial \beta_1} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) \;(=0) \Rightarrow \underbrace{\sum_{i=1}^{n} y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i}_{1}$$

- $$\frac{\partial Q}{\partial \beta_0} = -2 \sum_{i=1}^{n} x_i (y_i - \beta_0 - \beta_1 x_i) (=0) \Rightarrow \underbrace{\sum_{i=1}^{n} x_i y_i = \hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2}_{2}$$

Οι εξισώσεις $\boxed{(1 =) \sum_{i=1}^{n} y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i}$

και $\boxed{(2 =) \sum_{i=1}^{n} x_i y_i = \hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2}$ ονομάζονται

κανονικές εξισώσεις.

~o~o~o~o~o~o

$$1 \rightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$2 \rightarrow \sum_{i=1}^{n} x_i y_i = [\bar{y} - \hat{\beta}_1 \bar{x}] \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2$$

$$= \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n} + \hat{\beta}_1 \left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} \right] \rightarrow$$

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

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

και

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

$$A = \begin{pmatrix} \frac{\partial^2 Q}{\partial \beta_0^2} = 2n & \frac{\partial^2 Q}{\partial \beta_0 \partial \beta_1} = 2\sum x_i \\ \frac{\partial^2 Q}{\partial \beta_0 \partial \beta_1} = 2\sum x_i & \frac{\partial^2 Q}{\partial \beta_1^2} = 2\sum x_i^2 \end{pmatrix} \rightsquigarrow$$

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

Έχω $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ και για $x = x_0 \to \hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$

Με βάση τα παραπάνω. και για $\left.\begin{array}{l} \hat{\beta}_0 = 10 \\ \hat{\beta}_1 = 2.0 \end{array}\right\}$

Κάνοντας πράξεις έχω: $\hat{y} = 10 + 2.0 * x$.

αν $x_0 = 5.5 \rightsquigarrow \hat{y}_0 = 10 + 2.0 * 5.5 = 120$ ώρες εργασία.

$e_i = y_i - \hat{y}_i$ = κατάλοιπα (residuals)