# ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΘΕΩΡΙΑ ΑΠΟΦΑΣΕΩΝ

ΔΙΑΛΕΞΗ: ΠΑΛΙΝΔΡΟΜΗΣΗ

Διδάσκουσα: Ε. Γάκη, Επίκ. Καθηγήτρια

# Περιεχόμενα Ενότητας

- Ανάλυση Παλινδρόμησης
- Εκτίμηση Ευθείας παλινδρόμησης
- Εκτίμηση των συντελεστών της ευθείας παλινδρόμησης
- Στατιστική Συμπερασματολογία στην ανάλυση παλινδρόμησης
  - Διαστήματα Εμπιστοσύνης
  - Έλεγχοι Υποθέσεων
  - Συντελεστής Προσδιορισμού
- Πολλαπλή Παλινδρόμηση
- Μη γραμμικά μοντέλα

# Περιεχόμενα Μαθήματος

- Ανάλυση Παλινδρόμησης
- Εκτίμηση Ευθείας παλινδρόμησης
- Εκτίμηση των συντελεστών της ευθείας παλινδρόμησης
- Στατιστική Συμπερασματολογία στην ανάλυση παλινδρόμησης
  - Διαστήματα Εμπιστοσύνης
  - Έλεγχοι Υποθέσεων

# Στοχαστικό ή Ντετερμινιστικό Μοντέλο?

Έστω

$$Y = α + β X$$

Ντετερμινιστικό μοντέλο

αν X το εισόδημα &
Y οι δαπάνες κατανάλωσης

Αυτό σημαίνει ότι όλοι όσοι έχουν το ίδιο εισόδημα καταναλώνουν το ίδιο ποσό.

**ΙΣΧΥΕΙ??**

**ΌΧΙ.**

Γι αυτό στην πράξη θα πρέπει να χρησιμοποιούνται μοντέλα που να περικλύουν το στοιχείο της **τυχαιότητας**, πράγμα το οποίο επιτυγχάνεται με την προσθήκη μιας τυχαίας μεταβλητής ε.

$$Y = α + βX + ε$$

Το μοντέλο αυτό ονομάζεται **στοχαστικό** ή **μοντέλο πιθανότητας** (stochastic model or probabilistic model).

# Ανάλυση Παλινδρόμησης

- Η ανάλυση παλινδρόμησης ( regression analysis) έχει ως αντικειμενικό σκοπό την πρόβλεψη.

- Στόχος είναι η επιλογή κατάλληλου στοχαστικού μοντέλου το οποίο θα χρησιμοποιηθεί για την πρόβλεψη των τιμών μιας εξαρτημένης μεταβλητής (dependent random variable) Y από τις τιμές μιας τουλάχιστον ανεξάρτητης τυχαίας μεταβλητής (Independent random variable) X.

- Αν η τυχαία μεταβλητή Y είναι γραμμική συνάρτηση των παραμέτρων του μοντέλου τότε μιλάμε για ένα γραμμικό μοντέλο (linear model).

- Αν η Y εξαρτάται από μία μόνο ανεξάρτητη τυχαία μεταβλητή X έχουμε το μοντέλο της απλής γραμμικής παλινδρόμησης (simple linear regression model):

$$Y = \alpha + \beta X + \varepsilon$$

- Αν η Y εξαρτάται από περισσότερες τυχαίες μεταβλητές $X_1, X_2, \ldots, X_K$ έχουμε το μοντέλο της πολλαπλής παλινδρόμησης (multiple regression model).

$$Y = \alpha + \beta_1 X_1 + \ldots \beta_K X_K + \varepsilon$$

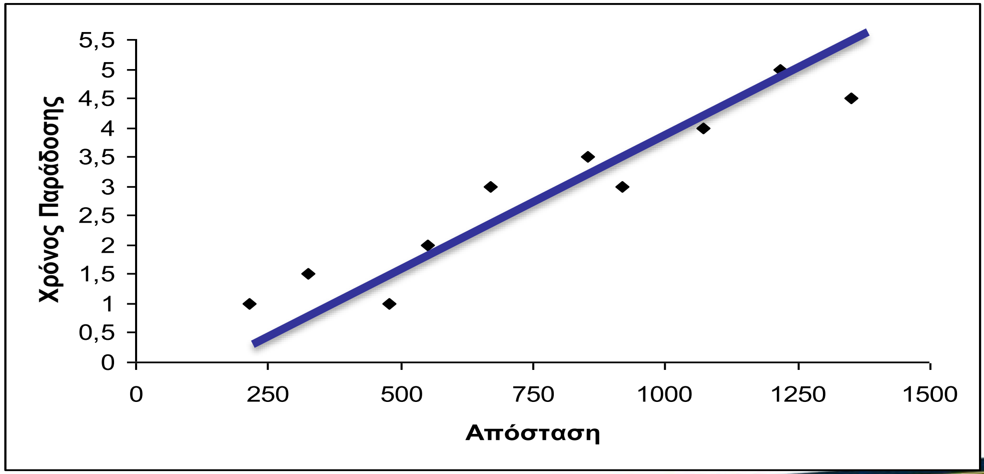
# Διάγραμμα Διασποράς

Σε μια μελέτη σχετική με τη γονιμότητα, επιλέγεται τυχαίο δείγμα 10 νιόπαντρων ζευγαριών και ρωτούνται για τον αριθμό των παιδιών που θα ήθελαν να αποκτήσουν (επιθυμητός αριθμός παιδιών X). Είκοσι χρόνια αργότερα και τα 10 αυτά ζευγάρια ρωτήθηκαν για τον αριθμό των παιδιών που τελικά απέκτησαν (πραγματικός αριθμός παιδιών Y).Τα αποτελέσματα της έρευνας συγκεντρώθηκαν στον παρακάτω πίνακα.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Επιθυμητός αριθμός παιδιών (X) | 0 | 1 | 2 | 1 | 0 | 3 | 4 | 2 | 2 | 1 |
| Πραγματικός Αριθμός παιδιών (Y) | 0 | 2 | 1 | 3 | 1 | 3 | 4 | 2 | 1 | 2 |

Ο ερευνητής ενδιαφέρεται να εξετάσει αν ο αριθμός των παιδιών που κάθε ζευγάρι τελικά απέκτησε (Y) μπορεί να ερμηνευτεί από τον αριθμό των παιδιών που τα ζευγάρια έλεγαν ότι θα επιθυμούσαν να αποκτήσουν κατά το χρόνο του γάμου τους (X).

Το πρώτο βήμα στη μελέτη τέτοιων προβλημάτων είναι η κατασκευή του **διαγράμματος διασποράς** (scatter diagram). Απεικονίζουμε δηλαδή σε σύστημα καρτεσιανών συντεταγμένων το τυχαίο δείγμα των παρατηρήσεων $(x_1,y_1),(x_2, y_2)$ ,… , $(x_n, y_n)$.Το διάγραμμα διασποράς που προκύπτει μας δίνει χρήσιμες πληροφορίες για τη μορφή του στοχαστικού μοντέλου το οποίο πρέπει να ελέγξουμε



# Εκτίμηση της ευθείας της Παλινδρόμησης

$$Y = \alpha + \beta X + \varepsilon$$

τυχαίο λάθος

Από τη σχέση προκύπτει ότι η τιμή $y_i$ της Y που παρατηρούμε για μια συγκεκριμένη τιμή $x_i$ της X είναι μια τυχαία τιμή από τον πληθυσμό των τιμών της τυχαίας μεταβλητής Y και αντιστοιχούν στο συγκεκριμένο $x_i$.

(Y τυχαία μεταβλητή, xi, yi τιμές *CLEAR*???)

Τι συμβαίνει με τα σφάλματα??

Για το τυχαίο σφάλμα $\varepsilon_i$ υποθέτουμε ότι είναι τυχαία μεταβλητή με τη μέση τιμή μηδέν

$$E(\varepsilon_i) = 0$$

Κατά συνέπεια η τυχαία μεταβλητή Y ακολουθεί μία κατανομή για την οποία ισχύει:

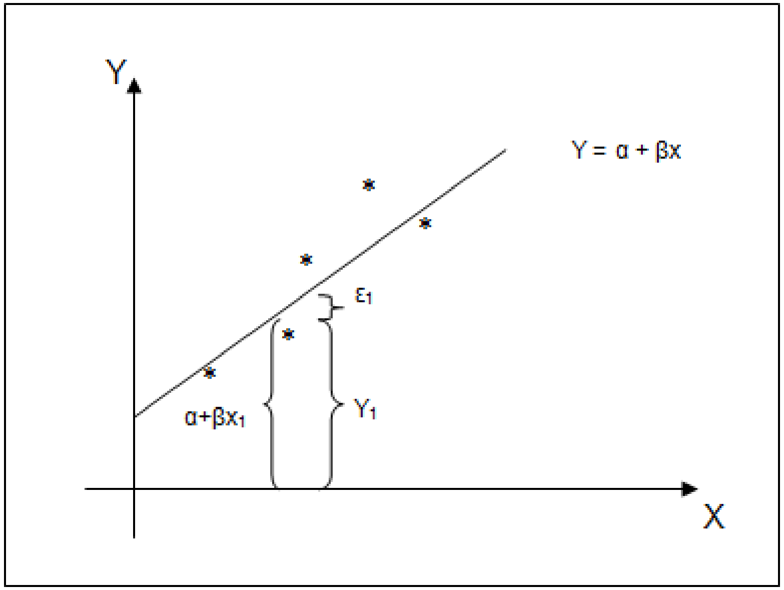
ΠΡΟΣΟΧΗ: Η μέση τιμή της τυχαίας μεταβλητής Y συνδέεται γραμμικά με το $x_i$ για κάθε συγκεκριμένο $x_i$.

$$E[Y_i / X_i] = E[\alpha + \beta x_i + \varepsilon_i] = \alpha + \beta E(x_i) + E(\varepsilon_i) = \alpha + \beta E(x_i)$$

# Εκτίμηση της ευθείας της Παλινδρόμησης

$$E[Y_i/X_i] = \alpha + \beta x_i$$

Η παραπάνω σχέση ονομάζεται ευθεία παλινδρόμησης (regression line) και οι συντελεστές α και β ονομάζονται συντελεστές γραμμικής παλινδρόμησης (linear regression coefficients).




Από το διάγραμμα φαίνεται ότι

- ✓το α είναι το σημείο στο οποίο η ευθεία παλινδρόμησης τέμνει τον άξονα των y, είναι δηλαδή η τιμή της Y που αντιστοιχεί στο x = 0
- ✓το β είναι η κλίση της ευθείας παλινδρόμησης και εκφράζει την αύξηση (μείωση) της Y που αντιστοιχεί σε αύξηση της X κατά μία μονάδα.

*Σημείωση: οι συντελεστές του μοντέλου είναι οι συντελεστές παλινδρόμησης στον πληθυσμό*
*Συνήθως έχουμε δείγμα.* ***Τι κάνουμε???***

# Εκτίμηση των συντελεστών της Παλινδρόμησης

Θυμάστε τις εκτιμήτριες?? (Lecture 6: Estimation)

1.Θα προσδιορίσουμε τις εκτιμήτριες $\hat{\alpha}, \hat{\beta}$ για τους συντελεστές α και β

2.Σε κάθε συγκεκριμένο δείγμα θα χρησιμοποιήσουμε τις [παραπάνω εκτιμήτριες και θα υπολογίσουμε τις σημειακές εκτιμήσεις τις οποίες συμβολίζουμε με a και b.

ΕΤΣΙ αν συμβολίσουμε με

- $\hat{Y}$ την εκτιμήτρια της $E(Y_i/x_i)$ και
- $\hat{y}_i$ μια εκτίμηση της

θα οδηγηθούμε σε μία σημειακή εκτίμηση της ευθείας παλινδρόμησης η οποία είναι

$$\hat{Y} = \hat{\alpha} + \hat{\beta}x$$

και

$$\hat{y} = a + bx$$

$$E[Y_i/X_i] = \alpha + \beta x_i$$

Για να υπολογίσουμε τα α και β θα χρησιμοποιήσουμε τη Μέθοδο Ελαχίστων Τετραγώνων (Lecture 6)

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
διοικώ σημαίνει αποφασίζω

# Εκτίμηση των συντελεστών της Παλινδρόμησης

Σύμφωνα με τη μέθοδο των ελαχίστων τετραγώνων διαλέγουμε τα α και β έτσι ώστε να ελαχιστοποιείται η παράσταση

$$Q = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left( Y_\iota - (\alpha + \beta X_i) \right)^2$$

Τα αποτελέσματα που προκύπτουν μας δίνουν

$$\hat{\alpha} = \frac{\left(\sum x_i^2\right)\left(\sum Y_i\right) - \left(\sum x_i\right)\left(\sum x_i Y_i\right)}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2} \quad \text{και} \quad \hat{\beta} = \frac{n\left(\sum x_i Y_i\right) - \left(\sum x_i\right)\left(\sum Y_i\right)}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}$$

Χρησιμοποιώντας τους τελευταίους τύπους παίρνω τις εκτιμήσεις για a και b

$$a = \frac{\left(\sum x_i^2\right)\left(\sum y_i\right) - \left(\sum x_i\right)\left(\sum x_i y_i\right)}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2} \qquad b = \frac{n\left(\sum x_i y_i\right) - \left(\sum x_i\right)\left(\sum y_i\right)}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}$$

# Πως προσδιορίζω τους συντελεστές στην πράξη?

Από τις τελευταίες 2 σχέσεις με μαθηματικούς μετασχηματισμούς προκύπτει

$$a = \bar{y} - b\bar{x} \quad \text{όπου} \quad b = \frac{\sum_i (X_i - \bar{x})(Y_i - \bar{y})}{\sum_i (X_i - \bar{x})^2}$$

και

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \quad \text{και} \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Στην πράξη χρησιμοποιώ τους εξής συμβολισμούς

$$S_{xx} = \sum_i \left(x_i - \bar{X}\right)^2 = \sum_i x_i^2 - \frac{\left(\sum x_i\right)}{n}$$

$$S_{yy} = \sum_i \left(y_i - \bar{Y}\right)^2 = \sum_i y_i^2 - \frac{\left(\sum y_i\right)^2}{n}$$

$$S_{xy} = \sum_i \left(x_i - \bar{X}\right)\left(y_i - \bar{Y}\right) = \sum_i \left(x_i - \bar{X}\right) y_i$$

ΟΠΟΤΕ έχουμε

$$\boxed{a = \bar{Y} - \frac{S_{xy}}{S_{xx}}\bar{X}}$$

$$\boxed{b = \frac{S_{xy}}{S_{xx}}}$$

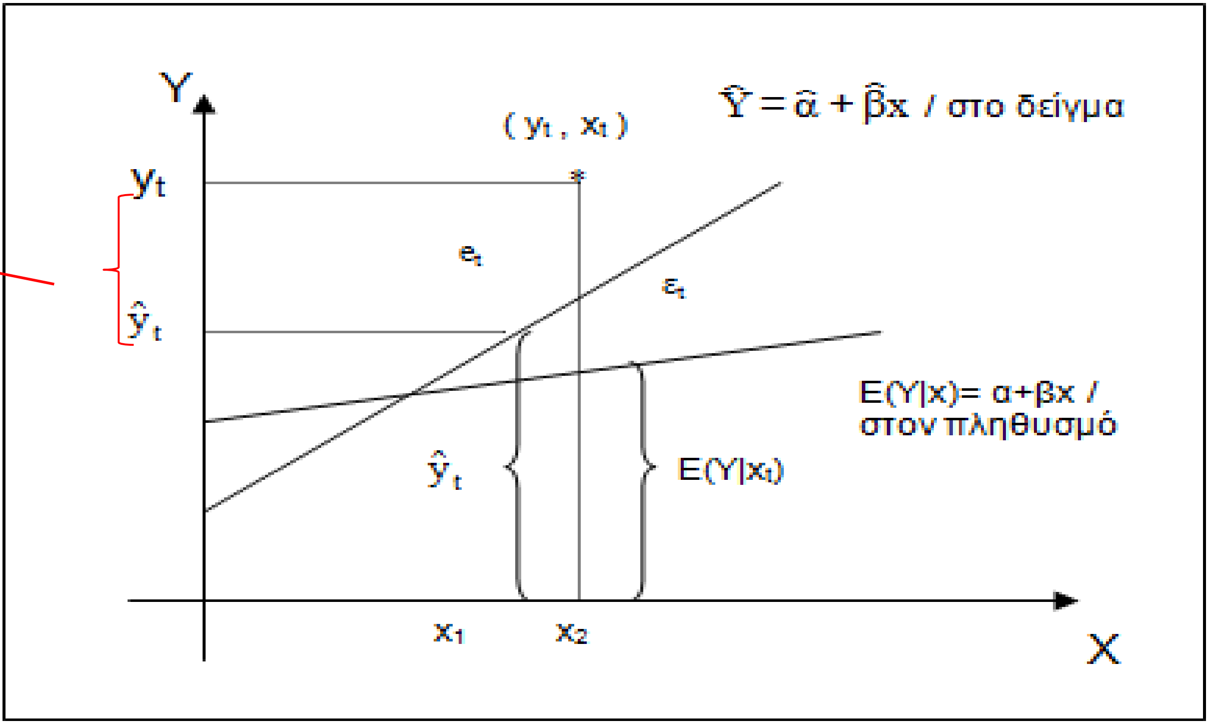
# Γραφικά για να δούμε τι συμβαίνει…..

Είναι φανερό ότι οι υπολογιζόμενες $\widehat{y}_i$ τιμές δεν είναι όλες ίσες με τις πραγματικές τιμές $y_i$

Η διαφορά μεταξύ των πραγματικών τιμών $y_i$ και των υπολογιζόμενων $\widehat{y}_i$

$$e_i = y_i - \widehat{y}_i$$

ονομάζεται κατάλοιπο (residual) και μπορεί να θεωρηθεί ως εκτίμηση της άγνωστης τιμής του διαταρακτικού όρου $\varepsilon_i$



# Στατιστική Συμπερασματολογία στην Ανάλυση Παλινδρόμησης

# Ας ασχοληθούμε με τον διαταρακτικό όρο

Το μοντέλο της ευθείας παλινδρόμησης είναι $Y_i = \alpha + \beta X_i + \varepsilon_i$

- ✓ όλα τα $\varepsilon_i$ ακολουθούν κανονική κατανομή με μέση τιμή μηδέν και την ίδια διακύμανση έστω $\sigma^2$. Δηλαδή $\varepsilon_i \sim N(0,\sigma^2)$ για όλα τα i=1,2,…
- ✓ Τα $\varepsilon_i$ είναι ανεξάρτητα μεταξύ τους.
- ✓ Τα $x_i$ είναι προκαθορισμένα

ΤΙ ΣΗΜΑΙΝΟΥΝ ΑΥΤΆ???

✓ότι η διακύμανση της τυχαίας μεταβλητής $\varepsilon_i$ είναι σταθερή για όλες τις τιμές του x. Στην περίπτωση που η διακύμανση παραμένει σταθερή, ο διαταρακτικός όρος είναι ομοσκεδαστικός (homoskedastic)

*Αν η διακύμανση δεν είναι σταθερή, ο διαταρακτικός όρος είναι **ετεροσκεδαστικός** (heteroskedastic)*

✓ότι οι διαταρακτικοί όροι είναι ασυσχέτιστοι μεταξύ τους.

$$Cov(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i, \varepsilon_j) - E(\varepsilon_i)E(\varepsilon_j) = 0 \quad \gamma\iota\alpha \quad i \neq j$$

**Οι βασικές ιδιότητες του μοντέλου της απλής γραμμικής παλινδρόμησης είναι:**

**κανονικότητα, ομοσκεδαστικότητα και ανεξαρτησία λαθών.**

# Ομοσκεδασστικότητα, Κανονικότητα

Οπότε

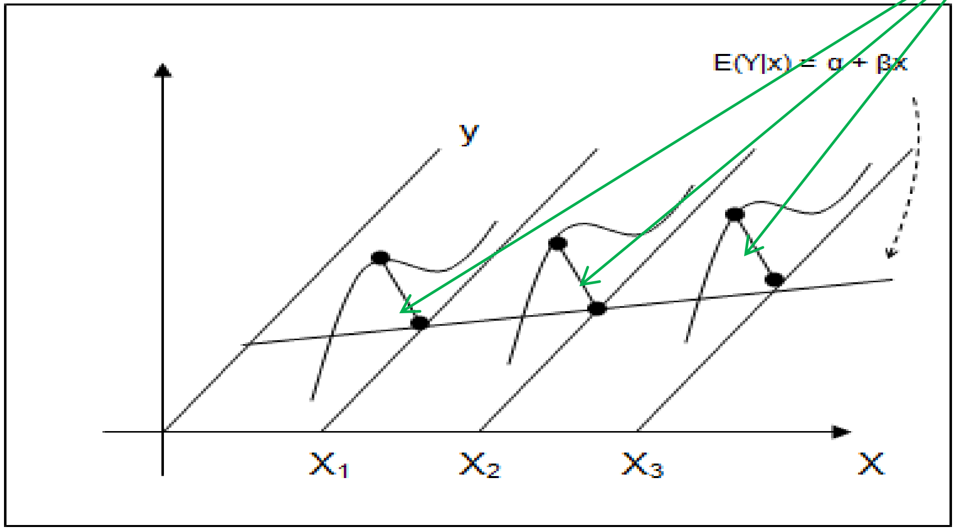
i. Τα $Y_i$ είναι $N(\alpha + \beta x_i, \sigma^2)$, $i=1,2,\ldots$

ii. Τα $Y_i$ είναι ανεξάρτητα μεταξύ τους

και επίσης

iii. Τα xi είναι προκαθορισμένα.

Λόγω της υπόθεσης **της κανονικότητας**, οι τιμές της Y ακολουθούν την κανονική κατανομή για κάθε τιμή του x




Λόγω της υπόθεσης της **ομοσκεδαστικότητας** η μεταβλητότητα γύρω από την ευθεία παλινδρόμησης είναι σταθερή για όλες τις τιμές του x.

Ο τρίτος άξονα αναφέρεται στην συνάρτηση πυκνότητας του διαταρακτικού όρου

# Εκτίμηση της Διακύμανσης των Τυχαίων Σφαλμάτων

- Με τη μέθοδο των ελαχίστων τετραγώνων οδηγούμεθα σε μία ευθεία η οποία προσαρμόζεται στα δεδομένα έτσι ώστε το άθροισμα των τετραγώνων των καταλοίπων να ελαχιστοποιείται και η οποία χρησιμεύει για να εκτιμήσουμε μια τιμή της Y για συγκεκριμένο x.

- Κατά συνέπεια χρειαζόμαστε ένα *μέτρο* για τη μεταβλητότητα που υπάρχει μεταξύ της παρατηρούμενης τιμής της Y και της αντίστοιχης υπολογιζόμενης τιμής μέσω του μοντέλου που έχουμε εκτιμήσει.

- Άρα χρειαζόμαστε την ποσότητα $\sigma^2$ τη διακύμανση δηλαδή των τυχαίων σφαλμάτων.

- Η τιμή όμως της διακύμανσης $\sigma^2$ σπανίως είναι γνωστή. Πρέπει λοιπόν να την εκτιμήσουμε, είναι δε λογικό η εκτίμηση αυτή, που θα συμβολίσουμε με $\mathbf{s^2_{Y/x}}$ να βασίζεται στα κατάλοιπα $e_i = y_i - \hat{y}_i$

$$\hat{\sigma}^2 \equiv S^2_{Y/x} = \frac{\sum\left(Y_i - \hat{Y}_i\right)^2}{n-2}$$

$$\boxed{S^2_{Y/x} = \frac{1}{n-2}\left(S_{YY} - \frac{S^2_{xY}}{S_{xx}}\right)}$$

Η τετραγωνική ρίζα της εκτίμησης $s^2_{Y/x}$ ονομάζεται **τυπικό σφάλμα της εκτίμησης της Y**

ΜΒΑχίος
ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
διοικώ σημαίνει αποφασίζω

# Διαστήματα Εμπιστοσύνης
## για τις παραμέτρους του μοντέλου της απλής γραμμικής παλινδρόμησης

# Διάστημα Εμπιστοσύνης για το β

Διάστημα εμπιστοσύνης για το β

Ισχύει $\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{S_{xx}}\right)$

Αποδεικνύεται $\dfrac{\hat{\beta}-\beta}{S_{Y/x}\big/\sqrt{S_{xx}}} \sim t_{n-2}$

Ένα 100(1-α)% διάστημα εμπιστοσύνης για την κλίση β της ευθείας παλινδρόμησης είναι το

$$\left(\hat{\beta} - t_{n-2,1-a/2} S_{\hat{\beta}},\ \ \hat{\beta} + t_{n-2,1-a/2} S_{\hat{\beta}}\right)$$

Όπου $S_{\hat{\beta}} = S_{Y/x}\big/\sqrt{S_{xx}}$ είναι το τυπικό σφάλμα εκτίμησης της .

# Διάστημα Εμπιστοσύνης για το α

Ισχύει $\hat{\alpha} \sim N\left(\alpha, \sigma^2\left(\frac{1}{n}+\frac{x^2}{S_{xx}}\right)\right)$

Αποδεικνύεται $\dfrac{\hat{\alpha}-\alpha}{S_{Y/x}\sqrt{\dfrac{1}{n}+\dfrac{\overline{x}^2}{S_{xx}}}} \sim t_{n-2}$

Ένα 100(1-α)% διάστημα εμπιστοσύνης για το α είναι το

$$\left(\widehat{\alpha}-t_{n-2,1-a/2}S_{\hat{\alpha}},\ \widehat{\alpha}+t_{n-2,1-a/2}S_{\hat{\alpha}}\right)$$

Όπου $S_{\hat{\alpha}} = S_{Y/x}\sqrt{\frac{1}{n}+\frac{\overline{x}^2}{S_{xx}}}$ είναι το τυπικό σφάλμα εκτίμησης της $\widehat{\alpha}$

# Διάστημα Εμπιστοσύνης για την $E(Y/X_t)$

Ισχύει $\hat{Y}_t \sim N\left(E\left(Y/X_t\right), \quad \sigma^2\left(\frac{1}{n}+\frac{(X_t-\bar{X})^2}{S_{XX}}\right)\right)$

Αποδεικνύεται $$\frac{\hat{Y}-E\left(Y/X_t\right)}{s^2{}_{Y/X}\sqrt{\left\{\frac{1}{n}+\frac{(X_t-\bar{X})^2}{S_{XX}}\right\}}} \sim t_{n-2}$$

Ένα 100(1-α)% διάστημα εμπιστοσύνης για την $E(Y/X_t)$

$$\left(\hat{Y}-t_{n-2,1-a/2}S_{\hat{Y}_t/X_t}, \quad \hat{Y}+t_{n-2,1-a/2}S_{\hat{Y}_t/X_t}\right)$$

Όπου $S_{\hat{Y}_t/X_t}=S_{Y/X_t}\sqrt{\left\{\frac{1}{n}+\frac{(X_t-\bar{X})^2}{S_{XX}}\right\}}$

# Διάστημα Πρόβλεψης

$$\hat{Y}_k \pm t_{n-2,1-a/2} S_{\hat{Y}_k / X_k}$$

$$S_{\hat{Y}_k / x_k} = S_{Y/X} \sqrt{\left\{1 + \frac{1}{n} + \frac{\left(X_k - \bar{X}\right)^2}{S_{XX}}\right\}}.$$

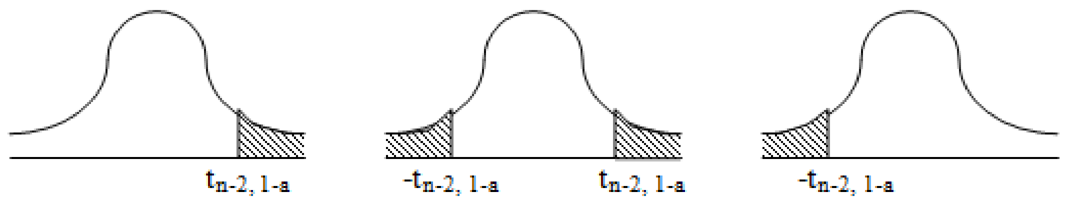
# Έλεγχοι Στατιστικών Υποθέσεων

# Έλεγχος Υπόθεσης για το β

$$H_o : \beta = \beta_o$$

Ως στατιστική συνάρτηση ελέγχου έχουμε $T_o = \dfrac{\hat{\beta} - \beta_o}{S_{Y/X} \big/ \sqrt{S_{XX}}} \sim t_{n-2}$

| $H_o : \beta = \beta_o$<br>$H_1 : \beta > \beta_o$ | $H_o : \beta = \beta_o$<br>$H_1 : \beta \neq \beta_o$ | $H_o : \beta = \beta_o$<br>$H_1 : \beta < \beta_o$ |
| --- | --- | --- |




Δηλαδή,

Απόρριψε την Ho αν

$to > t_{n-2,1-a}$

Απόρριψε την Ho αν

$to > t_{n-2,1-a/2}$

ή

$to < -t_{n-2,1-a/2}$

Απόρριψε την $H_o$ αν

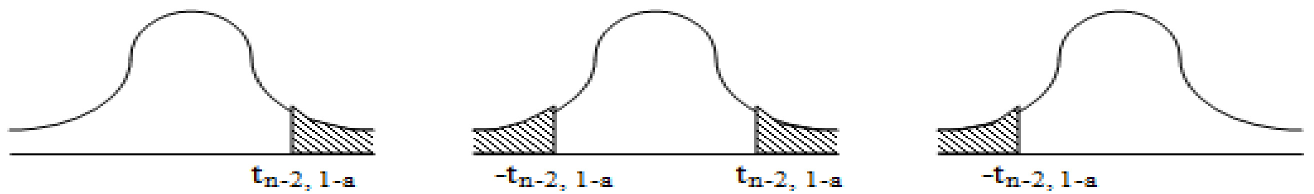
$t_o < -t_{n-2,1-\alpha}$

# Έλεγχος Υπόθεσης για το α

$H_o : \alpha = \alpha_o$

Ως στατιστική συνάρτηση ελέγχου έχουμε $T_o = \dfrac{\widehat{\alpha} - \alpha_o}{S_{Y/X}\sqrt{\dfrac{1}{n} + \dfrac{\overline{X}^2}{S_{XX}}}} \sim t_{n-2}$

| $H_o : \alpha = \alpha_o$<br>$H_1 : \alpha > \alpha_o$ | $H_o : \alpha = \alpha_o$<br>$H_1 : \alpha \neq \alpha_o$ | $H_o : \alpha = \alpha_o$<br>$H_1 : \alpha < \alpha_o$ |
|---|---|---|




Δηλαδή,

| Απόρριψε την $H_o$ αν | Απόρριψε την $H_o$ αν | Απόρριψε την $H_o$ αν |
|---|---|---|
| $t_o > t_{n-2,1-a}$ | $t_o > t_{n-2,1-a/2}$ ή $t_o < -t_{n-2,1-a/2}$ | $t_o < -t_{n-2,1-a}$ |

# Παράδειγμα

Ο Πίνακας περιέχει την βαθμολογία πέντε φοιτητών στα μαθήματα των Μαθηματικών (Χ) και της Στατιστικής (Υ).
α. Να εκτιμηθεί η ευθεία γραμμικής παλινδρόμησης της Υ πάνω στη Χ και να ερμηνευτούν οι συντελεστές της.
β. Να εκτιμηθεί το τυπικό σφάλμα της Υ και να ερμηνευτεί.
γ. Να εκτιμηθεί ο βαθμός ενός φοιτητή στο μάθημα Στατιστικής όταν ο βαθμός του στα Μαθηματικά είναι 4.
δ. Να υπολογιστεί το 95% διάστημα πρόβλεψης του βαθμού της Στατιστικής όταν ο βαθμός των Μαθηματικών είναι 7.

*Σημείωση:* Δίνεται ότι το επίπεδο σημαντικότητας είναι α=0.05.

| Μαθηματικά (Χ) | Στατιστική (Υ) |
|---|---|
| 2 | 1 |
| 3 | 2 |
| 5 | 5 |
| 6 | 6 |
| 8 | 7 |

# Παράδειγμα

Με βάση τα δεδομένα της άσκησης δημιουργείται ο παρακάτω Πίνακας.

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | $X_i$ | $Y_i$ | $X_i^2$ | $Y_i^2$ | $\left(X_i - \overline{X}\right)$ | $\left(X_i - \overline{X}\right)Y_i$ |
| | 2 | 1 | 4 | 1 | -2.8 | -2.8 |
| | 3 | 2 | 9 | 4 | -1.8 | -3.6 |
| | 5 | 5 | 25 | 25 | 0.2 | 1.0 |
| | 6 | 6 | 36 | 36 | 1.2 | 7.2 |
| | 8 | 7 | 64 | 49 | 3.2 | 22.4 |
| **ΣΥΝΟΛΑ** | **24** | **21** | **138** | **115** | **24,2** | **24,2** |

# Παράδειγμα

**Υπολογισμός βοηθητικών στοιχείων**

$$\overline{X} = \frac{24}{5} = 4.8, \qquad \overline{Y} = \frac{21}{5} = 4.2$$

$$S_{XX} = \sum X_i - \frac{\left(\sum X_i\right)^2}{n} = 138 - \frac{24^2}{5} = 22.8$$

$$S_{YY} = \sum Y_i - \frac{\left(\sum Y_i\right)^2}{n} = 115 - \frac{21^2}{5} = 26.8$$

$$S_{YX} = \sum (X_i - \overline{X})Y_i = 24.2$$

# Παράδειγμα

**α) Ευθεία Παλινδρόμησης**

$$\alpha = \overline{Y} - \frac{S_{XY}}{S_{XX}}\overline{X} = 4.2 - \frac{24.2}{22.8}4.8 \quad \Rightarrow a = -0.89$$

$$b = \frac{S_{XY}}{S_{XX}} = \frac{24.2}{22.8} \quad \Rightarrow b = 1.06$$

Άρα η ευθεία παλινδρόμησης είναι: $\widehat{Y} = -0.89 + 1.06X$

**β) Τυπικό Σφάλμα της Υ**

$$S^2_{y/x} = \frac{1}{n-2}\left(S_{yy} - \frac{S^2_{xy}}{S_{xx}}\right) = \frac{1}{5-2}\left(26.8 - \frac{(24.2)^2}{22.8}\right) \Rightarrow S^2_{y/x} \cong 0.4$$

Άρα $S_{y/x} = \sqrt{S^2_{y/x}} = 0.6$

# Παράδειγμα

**Γ. Εκτίμηση της τιμής του y**

Η ευθεία παλινδρόμησης είναι:

$$\hat{Y}_i = -0.89 + 1.06 * x$$

Για x = 4 έχουμε $\hat{Y}_i = -0.89 + 1.06 * 4 \Rightarrow \hat{Y}_i = 3.35$

# Παράδειγμα

Δ. **Διάστημα πρόβλεψης της τιμής του y**

Ο γενικός τύπος του διαστήματος πρόβλεψης είναι:

$$\hat{Y}_{\kappa} \pm t_{n-2,1-a/2}\, S_{\hat{Y}_k / X_k} \qquad (1)$$

Στη συγκεκριμένη περίπτωση

$$x_k = 7 \rightarrow \hat{Y}_k = -0.89 + 1.06 * 7 = 6.53$$
$$n = 5,\ a = 0.05$$

$t_{n-2,\ 1-a/2} = t_{3,\ 0.975} = 3.1824$

$$S_{\hat{Y}_k / X_k} = S_{Y/X}\sqrt{\frac{n+1}{n} + \frac{\left(X_k - \overline{X}\right)^2}{S_{xx}}} = 0.06\sqrt{\frac{6}{5} + \frac{4.84}{22.8}} = 0.72$$

Κατά συνέπεια το διάστημα πρόβλεψης είναι: 6.5 ± 3.1824*0.72 δηλαδή 6.53 ±2.29 ή (4.24, 8.82)