Γραμμικά Μοντέλα

Βιολέττα Ε. Πιπερίγκου

Τμήμα Μαθηματικών
Πανεπιστήμιο Πατρών

vpiperig@math.upatras.gr

http://www.math.upatras.gr/~vpiperig

Γραφείο 213, τηλ. 2610 997285

Περιεχόμενα

Ας πάμε στον Οδηγό Σπουδών...

ST434 Γραμμικά Μοντέλα

Εισαγωγή στην απλή γραμμική παλινδρόμηση και σχέσεις ευθείας γραμμής μεταξύ δυο μεταβλητών. Το απλό γραμμικό μοντέλο. Προσαρμογή ευθείας γραμμής, εκτίμηση των παραμέτρων με τη μέθοδο των ελαχίστων τετραγώνων. Υποθέσεις των Gauss - Markov για τα υπόλοιπα και ιδιότητες των εκτιμητών ελαχίστων τετραγώνων. Πίνακας ανάλυσης διασποράς, έλεγχοι υποθέσεων και διαστήματα εμπιστοσύνης. Εξέταση των υπολοίπων. Μελέτη της γραμμικής παλινδρόμησης με πίνακες. Πολλαπλή γραμμική παλινδρόμηση. Το πολλαπλό γραμμικό μοντέλο. Πολυωνυμικά μοντέλα. Η χρήση εικονικών μεταβλητών. Διαδικασία επιλογής της καλύτερης εξίσωσης προσαρμογής.

Μάθημα [ST434]: Γραμμικά Μοντέλα

Εξάμηνο 7 - Χειμερινό

Επιλογές Συγγραμμάτων:

1. Βιβλίο [77115860]: Ανάλυση Παλινδρόμησης Θεωρία και Εφαρμογές, Κούτρας Μ. Ευαγγελάρας Χ. Λεπτομέρειες
2. Βιβλίο [31339]: Εισαγωγή στην οικονομετρία, Χρήστου Γεώργιος Κ. Λεπτομέρειες
3. Βιβλίο [68388695]: Εφαρμοσμένη ανάλυση παλινδρόμησης, Draper N. R.,Smith H. Λεπτομέρειες

Πρόσθετο Διδακτικό Υλικό:

- Βιβλίο [320222]: ΕΙΣΑΓΩΓΗ ΣΤΟΝ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟ ΚΑΙ ΣΤΗ ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΜΕ R, ΙΩΑΝΝΗΣ ΝΤΖΟΥΦΡΑΣ Λεπτομ

ΠΡΟΑΠΑΙΤΟΥΜΕΝΑ ΜΑΘΗΜΑΤΑ:	Συνιστώμενη προαπαιτούμενη γνώση: ΘΕΩΡΙΑ ΠΙΘΑΝΟΤΗΤΩΝ Ι και ΙΙ, ΣΤΑΤΙΣΤΙΚΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ Ι και ΙΙ, ΓΡΑΜΜΙΚΗ ΑΛΓΕΒΡΑ Ι

2. ΜΑΘΗΣΙΑΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Μαθησιακά Αποτελέσματα

Περιγράφονται τα μαθησιακά αποτελέσματα του μαθήματος οι συγκεκριμένες γνώσεις, δεξιότητες και ικανότητες καταλλήλου επιπέδου που θα αποκτήσουν οι φοιτητές μετά την επιτυχή ολοκλήρωση του μαθήματος.

Στο μάθημα των Γραμμικών Μοντέλων εισάγονται και μελετώνται οι τεχνικές της ανάλυσης παλινδρόμησης που αποτελούν ένα από τα πιο χρήσιμα και διαδεδομένα εργαλεία της εφαρμοσμένης στατιστικής. Η ανάλυση παλινδρόμησης είναι μια στατιστική μεθοδολογία που χρησιμοποιείται σχεδόν σε όλες τις περιπτώσεις όπου υπάρχει η ανάγκη της ταυτόχρονης μελέτης δύο ή περισσοτέρων μεταβλητών με στόχο την μελέτη και πρόβλεψη μιας εξ'αυτών συναρτήσει των τιμών μιας άλλης ή κάποιων άλλων μεταβλητών που σχετίζονται με αυτήν.
Μετά την επιτυχή ολοκλήρωση του μαθημάτος ο φοιτητής θα μπορεί να εφαρμόζει και να μελετά στοχαστικά μοντέλα γραμμικής παλινδρόμησης.

- **Πιθανότητες:** πρακτική χρήση της κανονικής κατανομής
- **Στατιστική:** εμπέδωση μεθόδων εκτίμησης παραμέτρων, κατανόηση των διαστημάτων εμπιστοσύνης και των ελέγχων υποθέσεων
- **Γραμμική Άλγεβρα:** χρήση πινάκων σε προβλήματα βελτιστοποίησης

Τι είναι η ανάλυση παλινδρόμησης (regression);

είναι μια στατιστική τεχνική για να διερευνηθεί και να μοντελοποιηθεί η σχέση ανάμεσα σε μεταβλητές. Συγκεκριμένα μελετάται η εξάρτηση της (τυχαίας) μεταβλητής y από ανεξάρτητες μεταβλητές $\mathbf{x}$, οι οποίες δεν είναι τυχαίες, άλλα οι τιμές τους επιλέγονται από εκείνον που διεξάγει το πείραμα.

Η ανάλυση παλινδρόμησης μπορεί να θεωρηθεί ως η στατιστική θεωρία της προβλέψεως της εξαρτημένης μεταβλητής y από τις ανεξάρτητες μεταβλητές $\mathbf{x}$, οι οποίες γιαυτό ονομάζονται και προβλέπουσες

- **Οικονομία:** να προβλεφθεί το εθνικό εισόδημα από τα σχεδιαζόμενα κυβερνητικά μέτρα και επενδύσεις
- **Βιομηχανία:** προκαταρκτικά πειράματα για την μελέτη των επιδράσεων διαφόρων συνθηκών παραγωγής (θερμοκρασία, υγρασία, πίεση κλπ) στην ποιότητα του προϊόντος (πχ σκληρότητα χάλυβα)
- **Μετεωρολογία:** πρόβλεψη του καιρού βάσει διαφόρων ατμοσφαιρικών μετρήσεων (πίεση, υγρασία, ένταση ανέμου)

Βήματα Εργασίας

- **Συλλογή των δεδομένων:** παλαιά δεδομένα, τρέχουσες παρατηρήσεις, πειραματικός σχεδιασμός
- **Περιγραφή των δεδομένων:** scatter diagram ή dot plot (γράφημα διασποράς ή σημειόγραμμα), ένα γραμμικό **μοντέλο**
- **Εκτίμηση παραμέτρων**
- **Πρόβλεψη και εκτίμηση:** ο ρόλος των υπολογιστικών πακέτων στην αξιολόγηση της καταλληλότητας του μοντέλου
- **Χρήση:** πχ στον έλεγχο διεργασιών

Στην αναζήτηση της σχέσης των μεταβλητών

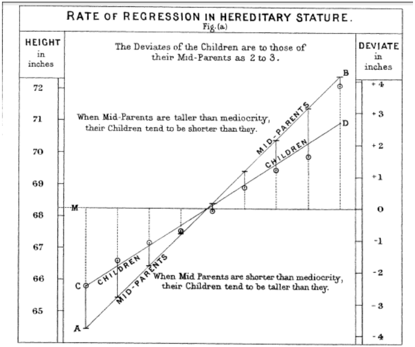
Ο όρος regression εμφανίζεται σε εργασία του ανθρωπολόγου Sir Francis Galton (1886) Regression Towards Mediocrity in Hereditary Stature. *The Journal of the Anthropological Institute of Great Britain and Ireland* Vol. 15, pp. 246-263

ANTHROPOLOGICAL MISCELLANEA.

REGRESSION *towards* MEDIOCRITY *in* HEREDITARY STATURE.
By FRANCIS GALTON, F.R.S., &c.

[WITH PLATES IX AND X.]

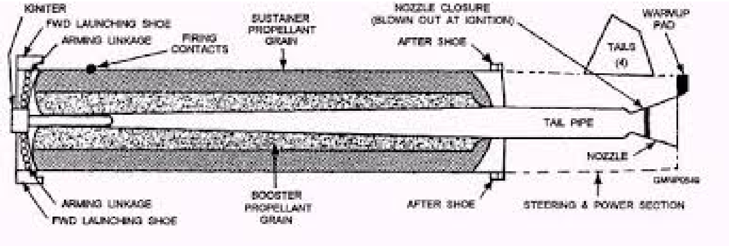
THIS memoir contains the data upon which the remarks on the Law of Regression were founded, that I made in my Presidential Address to Section H, at Aberdeen. That address, which will appear in due course in the Journal of the British Association, has already been published in "Nature," September 24th. I reproduce here the portion of it which bears upon regression, together with some amplification where brevity had rendered it obscure, and I have added copies of the diagrams suspended at the meeting, without which the letterpress is necessarily difficult to follow. My object is to place beyond doubt the existence of a simple and far-reaching law that governs the hereditary transmission of, I believe, every one of those simple qualities which all possess, though in unequal degrees. I once before ventured to draw attention to this law on far more slender evidence than I now possess.



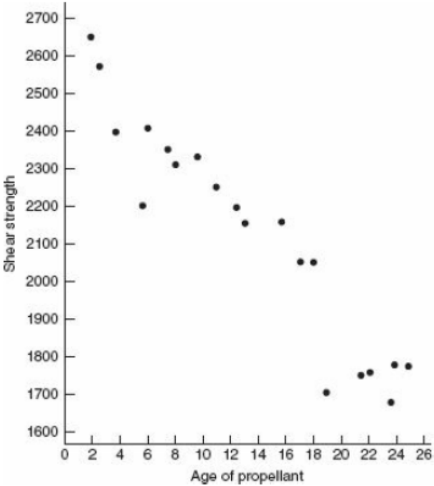
Οπτική αναζήτηση της σχέσης των μεταβλητών

Το παράδειγμα του rocket motor

Προωθητικός μηχανισμός κατασκευάζεται με προσκόλληση του καυσίμου που αναφλέγεται(igniter propellant) σε ένα μέσο συγκρατήσεως εντός μεταλλικού κελύφους. Οι ειδικοί θεωρούν ότι η διατμητική τάση (Shear Strength) σχετίζεται με την ηλικία του μέσου συγκρατήσεως (sustainer propellant)

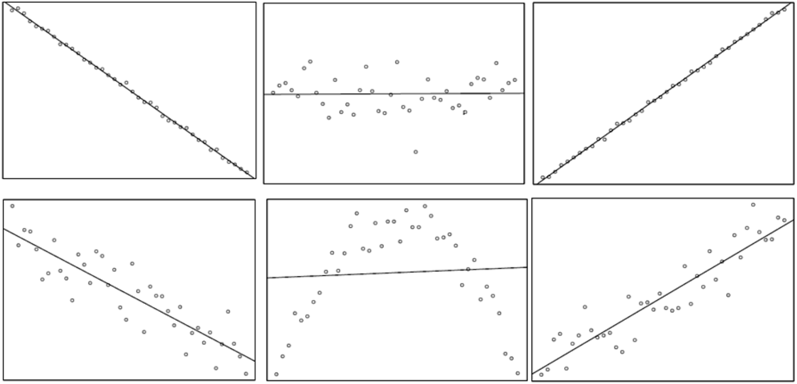


Observation, i	Shear Strength, y_i(psi)	Age of Propellant, x_i(weeks)
1	2158.70	15.50
2	1678.15	23.75
3	2316.00	8.00
4	2061.30	17.00
5	2207.50	5.50
6	1708.30	19.00
7	1784.70	24.00
8	2575.00	2.50
9	2357.90	7.50
10	2256.70	11.00
11	2165.20	13.00
12	2399.55	3.75
13	1779.80	25.00
14	2336.75	9.75
15	1765.30	22.00
16	2053.50	18.00
17	2414.40	6.00
18	2200.50	12.50
19	2654.20	2.00
20	1753.70	21.50



“Ποσοτικοποίηση” της εξάρτησης

- **Ανάλυση συσχέτισης (correlation analysis):** μέτρηση του βαθμού και της κατεύθυνσης της γραμμικής σχέσης δύο τυχαίων μεταβλητών (συντελεστής συσχέτισης του Pearson $\rho = \frac{cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}}$, $-1 \leqslant \rho \leqslant 1$)
- **Ανάλυση παλινδρόμησης:** αφορά στην πρόβλεψη του αποτελέσματος μιας τυχαίας μεταβλητής βασιζόμενοι στην τιμή μιας άλλης (ή άλλων)



Το απλό γραμμικό μοντέλο

εξαρτημένη μεταβλητή

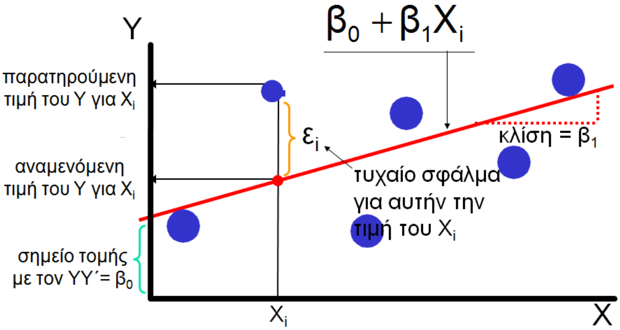
που τέμνει τον άξονα YY΄

κλίση της ευθείας

ανεξάρτητη μεταβλητή

τυχαίο σφάλμα

$$Y_i = \underbrace{\beta_0 + \beta_1 X_i}_{\text{γραμμικό «κομμάτι»}} + \underbrace{\varepsilon_i}_{\text{«κομμάτι» του τυχαίου σφάλματος}}$$



Το απλό γραμμικό μοντέλο

- **απλό:** μία ανεξάρτητη μεταβλητή

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon \qquad \text{(γραμμικό πολλαπλό)}$$

- **γραμμικό:** ως προς τις παραμέτρους

$$Y = \beta_0 + e^{\beta_1 X} \qquad \text{(μη γραμμικό)}$$

- **όχι κατά ανάγκη γραμμικό ως προς την μεταβλητή**

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 \qquad \text{(παραβολή, γραμμικό στις παραμέτρους)}$$

- **Οι ελάχιστες ΥΠΟΘΕΣΕΙΣ**

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad i = 1, \ldots, n$$

ανεξάρτητα σφάλματα $\epsilon_i \quad \Rightarrow \quad cov(\epsilon_i, \epsilon_j) = 0, i \neq j$

$E(\epsilon_i) = 0$, και $var(\epsilon_i) = \sigma^2 \quad i = 1, \ldots, n$

Το απλό γραμμικό μοντέλο

Για τις Y_i ισχύει:

- οι Y_i είναι ανεξάρτητες, επειδή τα σφάλματα ϵ_i είναι ανεξάρτητα (βλέπε συμπληρωματικό υλικό)
- $E(Y_i) = \beta_0 + \beta_1 X_i + E(\epsilon_i) = \beta_0 + \beta_1 X_i, \quad i = 1, \ldots, n$
  δεν έχουν την ίδια μέση τιμή
- $var(Y_i) = var(\epsilon_i) = \sigma^2, \quad i = 1, \ldots, n$
  έχουν την ίδια διασπορά

Συμπεράσματα για τις Y_i:

- οι Y_i δεν είναι ισόνομες
- η κατανομή τους δεν είναι γνωστή, εάν δεν είναι γνωστή η κατανομή των ϵ_i

Πώς θα γίνει η εκτίμηση των παραμέτρων β_0, β_1;

Με τις προηγούμενες υποθέσεις, οι δύο γνωστές μέθοδοι δεν μπορούν να χρησιμοποιηθούν

- **ΕΜΠ, Εκτιμητές μεγίστης πιθανοφάνειας:** επειδή η κατανομή των Y_i δεν είναι γνωστή δεν γνωρίζουμε τη συνάρτηση πιθανοφάνειας
- **Εκτιμητές με την μέθοδο των ροπών:** επειδή οι Y_i δεν είναι ισόνομες δεν μπορούμε να γράψουμε τις εξισώσεις των ροπών

Μέθοδος ελαχίστων τετραγώνων

Ελαχιστοποίηση των τετραγώνων των σφαλμάτων:

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} \epsilon_i^2 = \min_{\beta_0,\beta_1} \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2$$

οφείλεται στους Gauss (1809) και Legendre (1805) και έχει αρχικά χρησιμοποιηθεί στην γεωδαισία (με χρήση στη ναυσιπλοοΐα) και στην ουράνια μηχανική (για τον προσδιορισμό της τροχιάς πλανητών)

Μέθοδος ελαχίστων τετραγώνων

$$S = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} [y_i - \beta_0 - \beta_1 x_i]^2 \Rightarrow$$

$$\frac{\partial S}{\partial \beta_0} = \sum_{i=1}^{n} (-2)[y_i - \beta_0 - \beta_1 x_i] \Rightarrow 0 = \sum_{i=1}^{n} y_i - n\beta_0 - \beta_1 \sum_{i=1}^{n} x_i \Rightarrow$$

και $$\frac{\partial S}{\partial \beta_1} = \sum_{i=1}^{n} (-2x_i)[y_i - \beta_0 - \beta_1 x_i] \Rightarrow 0 = \sum_{i=1}^{n} x_i y_i - \beta_0 \sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} x_i^2 \Rightarrow$$

$$\sum_{i=1}^{n} y_i = \beta_0 n + \beta_1 \sum_{i=1}^{n} x_i \tag{1}$$

$$\sum_{i=1}^{n} x_i y_i = \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 \tag{2}$$

οι (1) και (2) ονομάζονται κανονικές εξισώσεις

Μέθοδος ελαχίστων τετραγώνων

$$\frac{\sum_{i=1}^n y_i}{n} = \beta_0 + \beta_1 \frac{\sum_{i=1}^n x_i}{n}$$
$$\sum_{i=1}^n x_i y_i = \beta_0 \sum_{i=1}^n x_i + \beta_1 \sum_{i=1}^n x_i^2$$

$$\beta_1 = \frac{D_{\beta_1}}{D} = \frac{\begin{vmatrix} 1 & \frac{\sum_{i=1}^n y_i}{n} \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i y_i \end{vmatrix}}{\begin{vmatrix} 1 & \frac{\sum_{i=1}^n x_i}{n} \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{vmatrix}}, \qquad \beta_0 = \frac{\sum_{i=1}^n y_i}{n} - \beta_1 \frac{\sum_{i=1}^n x_i}{n}$$

Παρατηρούμε επίσης ότι

$$\begin{vmatrix} \frac{\partial^2 S}{\partial \beta_0^2} & \frac{\partial^2 S}{\partial \beta_0 \partial \beta_1} \\ \frac{\partial^2 S}{\partial \beta_1 \partial \beta_0} & \frac{\partial^2 S}{\partial \beta_1^2} \end{vmatrix} = \begin{vmatrix} \sum_{i=1}^n 2 & \sum_{i=1}^n 2x_i \\ \sum_{i=1}^n 2x_i & \sum_{i=1}^n 2x_i^2 \end{vmatrix} = 4n\left(\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}\right) \geqslant 0$$

Οι λύση των κανονικών εξισώσεων είναι οι εκτιμητές, $\widehat{\beta_0}$ και $\widehat{\beta_1}$, των παραμέτρων β_0 και β_1

$$\widehat{\beta_1} = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n}} \equiv \frac{S_{XY}}{S_{XX}}$$

$$\widehat{\beta_0} = \frac{\sum_{i=1}^{n} Y_i}{n} - \widehat{\beta_1} \frac{\sum_{i=1}^{n} X_i}{n} \equiv \overline{Y} - \widehat{\beta_1}\overline{X}$$

όπου

$$\overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} \quad \text{και} \quad \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$S_{XX} = \sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n} = \sum_{i=1}^{n} X_i^2 - n\overline{X}^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2 = \sum_{i=1}^{n} (X_i - \overline{X})X_i$$

$$S_{XY} = \sum_{i=1}^{n} X_i Y_i - \frac{(\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)}{n} = \sum_{i=1}^{n} X_i Y_i - n\overline{X}\overline{Y}$$

$$= \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y}) = \sum_{i=1}^{n} (X_i - \overline{X})Y_i - \overline{Y}\sum_{i=1}^{n} (X_i - \overline{X}) = \sum_{i=1}^{n} (X_i - \overline{X})Y_i$$

Παράδειγμα

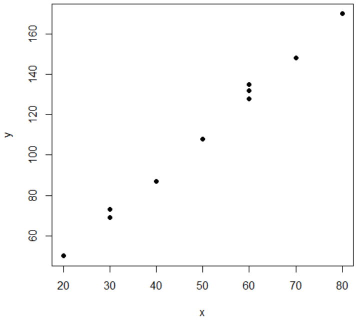
Ξυλουργικό εργοστάσιο κατασκευάζει θρανία μια φορά το μήνα, ανάλογα με τις παραγγελίες. Τους τελευταίους 10 μήνες καταγράφηκαν ο αριθμός των θρανίων (x) και οι αντίστοιχες ανθρωποώρες (y) που χρειάστηκαν, κάτω από όμοιες συνθήκες εργασίας. Για δύο νέες παραγγελίες 50 και 65 θρανίων, πόσες κατά μέσο όρο ανθρωποώρες θα χρειαστούν;

```
       x   y
 [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
```

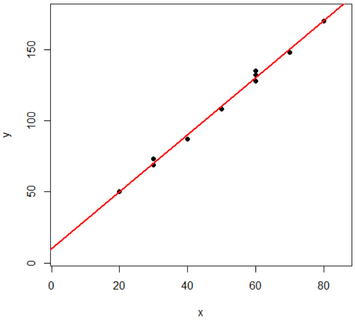
Παράδειγμα

Ξυλουργικό εργοστάσιο κατασκευάζει θρανία μια φορά το μήνα, ανάλογα με τις παραγγελίες. Τους τελευταίους 10 μήνες καταγράφηκαν ο αριθμός των θρανίων (x) και οι αντίστοιχες ανθρωποώρες (y) που χρειάστηκαν, κάτω από όμοιες συνθήκες εργασίας. Για δύο νέες παραγγελίες 50 και 65 θρανίων, πόσες κατά μέσο όρο ανθρωποώρες θα χρειαστούν;

	x	y
[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



```
> cbind(x,y,x*y,x*x)
      x   y
 [1,] 30  73  2190  900
 [2,] 20  50  1000  400
 [3,] 60 128  7680 3600
 [4,] 80 170 13600 6400
 [5,] 40  87  3480 1600
 [6,] 50 108  5400 2500
 [7,] 60 135  8100 3600
 [8,] 30  69  2070  900
 [9,] 70 148 10360 4900
[10,] 60 132  7920 3600
> cbind(sum(x),sum(y),sum(x*y),sum(
     [,1] [,2]  [,3]  [,4]
[1,]  500 1100 61800 28400
```



$$\widehat{\beta_1} = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n}} = \frac{61800 - \frac{500 \cdot 1100}{10}}{28400 - \frac{(500)^2}{10}} = 2 \quad (\text{ανθρωποώρες})$$

$$\widehat{\beta_0} = \overline{Y} - \widehat{\beta_1}\overline{X} = 110 - 2 \cdot 50 = 10 \quad (\text{ανθρωποώρες})$$

Χάραξη ευθείας

Στο απλό γραμμικό μοντέλο η προσαρμοσμένη ευθεία $\widehat{\beta_0} + \widehat{\beta_1} X$ διέρχεται από τα σημεία

- $\left(0, \widehat{\beta_0}\right)$, εδώ είναι το σημείο $(0, 10)$
- $(\overline{X}, \overline{Y})$ (προκύπτει από τη 2η κανονική εξίσωση), εδώ είναι το $(50, 110)$

Ερμηνεία των τιμών των παραμέτρων

$\widehat{\beta_0}$ **=10:** πριν ξεκινήσει οποιαδήποτε παραγωγή $(x = 0)$ χρειάζονται 10 ανθρωποώρες προετοιμασίας

$\widehat{\beta_1}$ **=2:** η κατασκευή ενός θρανίου απαιτεί (κατά μέσο όρο) 2 ανθρωποώρες

Προβλεπόμενες τιμές

- Για τα αρχικά X_i η $\widehat{Y_i} = \widehat{\beta_0} + \widehat{\beta_1} X_i, i = 1, \ldots, n$ είναι η προβλεπόμενη τιμή για την μεταβλητή Y. Εδώ $X_6 = 50$ οπότε $\widehat{Y_6} = 10 + 2 \cdot 50 = 110$.
- Για μια νέα τιμή X_0 η $\widehat{Y_0} = \widehat{\beta_0} + \widehat{\beta_1} X_0$ είναι η εκτιμώμενη μέση τιμή του Y_0. Εδώ για $X_0 = 65$ έχουμε $\widehat{Y_0} = 10 + 2 \cdot 65 = 140$.

Τα αποτελέσματα χρησιμοποιώντας την R

```
> summary(lm(y~x))

Call:
lm(formula = y ~ x)

Residuals:
   Min     1Q Median     3Q    Max 
  -3.0   -2.0   -0.5    1.5    5.0 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 10.00000    2.50294   3.995  0.00398 ** 
x            2.00000    0.04697  42.583 1.02e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.739 on 8 degrees of freedom
Multiple R-squared:  0.9956,    Adjusted R-squared:  0.9951 
F-statistic:  1813 on 1 and 8 DF,  p-value: 1.02e-10

> summary(aov(y~x))
            Df Sum Sq Mean Sq F value   Pr(>F)    
x            1  13600   13600    1813 1.02e-10 ***
Residuals    8     60       8                     
---
```

Άσκηση

Έστω το γραμμικό μοντέλο

$$Y_i = \beta_1 X_i + \epsilon_i, \quad i = 1, \ldots, n$$

όπου τα ϵ_i είναι ανεξάρτητα σφάλματα με $E(\epsilon_i) = 0$, και $var(\epsilon_i) = \sigma^2$ για $i = 1, \ldots, n$.

- Να βρεθεί ο εκτιμητής, $\widehat{\beta_1}$, της παραμέτρου β_1 με τη μέθοδο των ελαχίστων τετραγώνων.
- Να δοθούν δύο σημεία από τα οποία διέρχεται η προσαρμοσμένη ευθεία $\widehat{\beta_1} X$.
- Να προσδιοριστεί η ευθεία αυτή για το παράδειγμα της διαφάνειας 18.