

MEM-205 Περιγραφική Στατιστική
Τμήμα Μαθηματικών και Εφ. Μαθηματικών, Πανεπιστήμιο Κρήτης

Κώστας Σμαραγδάκης (kesmarag@gmail.com)

30-03-2020

Γραμμική Παλινδρόμηση και Ψευδομεταβλητές

Παράδειγμα

- Y - Ο τελικός βαθμός σε ένα συγκεκριμένο μάθημα του 4ου έτους σπουδών
- $X^{(1)}$ - Ο βαθμός στη πρόοδο του μαθήματος
- $X^{(2)}$ - Ο μέσος όρος βαθμολογίας του φοιτητή/τριας $N=60$
- Το τμήμα του φοιτητή/τριας (πχ. tem, math, csd)
- Παρακολούθηση τουλάχιστον των μισών μαθημάτων μετά τη πρόοδο $N=41/0=41$

$$x^{(1)} = 7$$

$$x^{(2)} = 5$$

$$x^{(2)} = 6.5$$

$$x^{(2)} = 7$$

$$x^{(1)} = 1$$

$$x^{(1)} = 0$$

$$N=41 \rightarrow d^{(1)} = 1$$

$$0 \times 1$$

$$y = 8$$

$$y = ?$$

500 φοιτητές

$$+tem \rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$+math \rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$+csd \rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$d^{(1)} \rightarrow \begin{matrix} 0 & 0 \times 1 \\ 1 & N=41 \end{matrix} \quad y = A + B^{(1)}x^{(1)} + B^{(2)}x^{(2)} + C^{(1)}d^{(1)} + C^{(2)}d^{(2)} + C^{(3)}d^{(3)} + \epsilon$$

6 παραβλέψεις

$$p = [a, b^{(1)}, b^{(2)}, c^{(1)}, c^{(2)}, c^{(3)}]^T \quad y = [y_1, \dots, y_n]^T$$

$$X = \begin{bmatrix} 1 & x_1^{(1)} & x_1^{(2)} & d_1^{(1)} & d_1^{(2)} & d_1^{(3)} \\ 1 & x_2^{(1)} & x_2^{(2)} & d_2^{(1)} & d_2^{(2)} & d_2^{(3)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n^{(1)} & x_n^{(2)} & d_n^{(1)} & d_n^{(2)} & d_n^{(3)} \end{bmatrix}$$

$N \times 6$

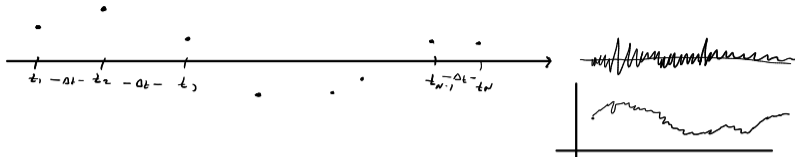
$$P = (X^T X)^{-1} X^T y \quad 6 \times 1$$

$$\hat{y} = a + b^{(1)} \cdot 5 + b^{(2)} \cdot 7 + \cancel{c^{(1)} \cdot 0} + c^{(2)} \cdot 1 + \cancel{c^{(3)} \cdot 0}$$

Χρονολογικές Σειρές (Time Series)

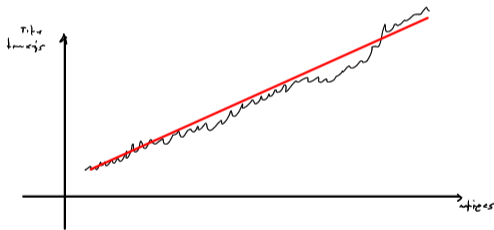
$$y = [y_1, y_2, y_3, \dots, y_n]$$

Δt Δt
 $t_1 \rightarrow t_2 \rightarrow t_3$

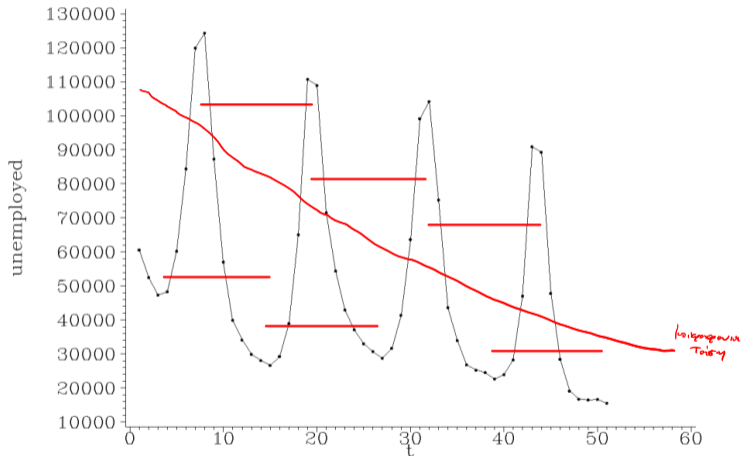


Μια χρονολογική σειρά είναι ένα σύνολο παρατηρήσεων που παρουσιάζονται σε χρονολογική διάταξη.

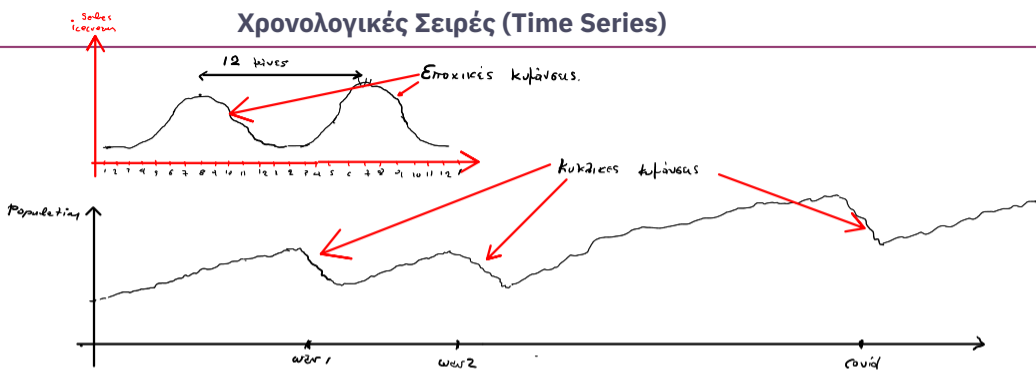
- ▶ Μακροχρόνια τάση
- ▶ Εποχικές κυμάνσεις
- ▶ Κυκλικές κυμάνσεις
- ▶ Τυχαίες κυμάνσεις



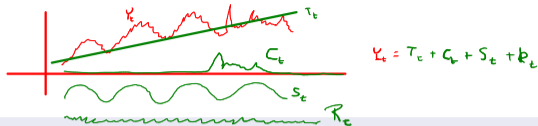
Χρονολογικές Σειρές (Time Series)



Χρονολογικές Σειρές (Time Series)



Χρονολογικές Σειρές (Time Series)



Το προσθετικό μοντέλο για χρονολογικές σειρές

$$Y_t = T_t + C_t + S_t + R_t, \quad t = 1, \dots, N$$

- ▶ T_t : Η Μακροχρόνια τάση για την t -χρονική περίοδο.
- ▶ S_t : Ο δείκτης εποχικότητας για την t -χρονική περίοδο.
- ▶ C_t : Η κυκλική κύμανση για την t -χρονική περίοδο.
- ▶ R_t : Η τυχαία κύμανση για την t -χρονική περίοδο.

Απλουστευμένο μοντέλο

$$Y_t = T_t + R_t, \quad t = 1, \dots, N$$

$$\mathbb{E}\{R_t\} = 0, \quad \mathbb{E}\{Y_t\} = T_t \equiv f(t)$$

- ▶ $f(t) = f(t; \beta_1, \beta_2, \dots, \beta_p)$
- ▶ Έυρεση εκτιμήσεων $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$ των παραμέτρων της f .

$$y_t = f(t; \beta_1, \beta_2, \dots, \beta_p) + r(t)$$

$$\hat{y}_t = f(t; \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)$$

Χρονολογικές Σειρές (Time Series)

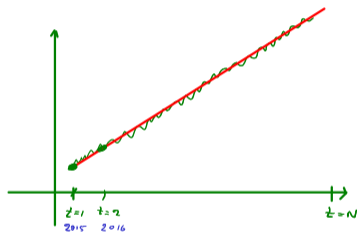
$$\{y_1, y_2, \dots, y_N\} \equiv \{(t, y_t) \mid t=1, 2, \dots, N\}$$

Linear function

$$f(t) = f(t; \beta_1, \beta_2) = \beta_1 + \beta_2 t, \quad \beta_1, \beta_2 \in \mathbb{R}$$

$$a \quad b$$

$$\hat{\beta}_1 = a \quad \hat{\beta}_2 = b$$

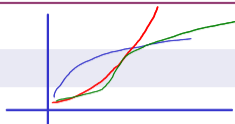


Χρονολογικές Σειρές (Time Series)

Allometric function

$$\{y_1, y_2, \dots, y_N\}$$

$$\hat{\beta}_1 = \hat{\beta}_0 = \hat{\beta}$$

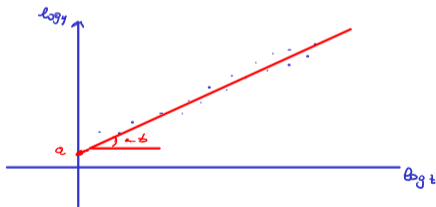


$$f(t) = f(t; \beta_1, \beta_2) = \beta_2 t^{\beta_1}, \quad \beta_1 \in \mathbb{R}, \beta_2 > 0$$

$$\log f(t) = \underbrace{\log \beta_2}_a + \underbrace{\beta_1}_{b} \log t$$

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$$

$$\hat{\beta}_1 = b \quad \hat{\beta}_2 = e^{x \sum a_i} = e^a$$



Mitscherlich function

$$f(t) = f(t; \beta_1, \beta_2, \beta_3) = \beta_1 + \beta_2 \exp(\beta_3 t), \quad \beta_1, \beta_2 \in \mathbb{R}, \beta_3 < 0$$

$$\lim_{t \rightarrow \infty} f(t) = \beta_1 \quad \hat{y}_1 = f(t) = \beta_1 + \beta_2 = y_1 \Leftrightarrow \boxed{\beta_2 = y_1 - \beta_1}$$

$$\begin{aligned} f(t) &= \beta_1 + \beta_2 \exp(\beta_3 t) \exp(\beta_3(t-1)) = \\ &= \beta_1 + \beta_1 \exp(\beta_3) \frac{-\beta_1 \exp(\beta_3) + \beta_2 \exp(\beta_3) \exp(\beta_3(t-1))}{\beta_1 (1 - \exp(\beta_3))} + \exp(\beta_3) \frac{\beta_1}{\beta_1 (1 - \exp(\beta_3))} f(t-1) \end{aligned}$$

$$\{y_1, \dots, y_n\} \rightarrow \sum_{(x,y)}^{N-1 \text{ στοιχεία}} (y_1, y_2), (y_2, y_3), \dots, (y_{n-1}, y_n)$$

$$\boxed{\hat{\beta}_3 = \log b}$$

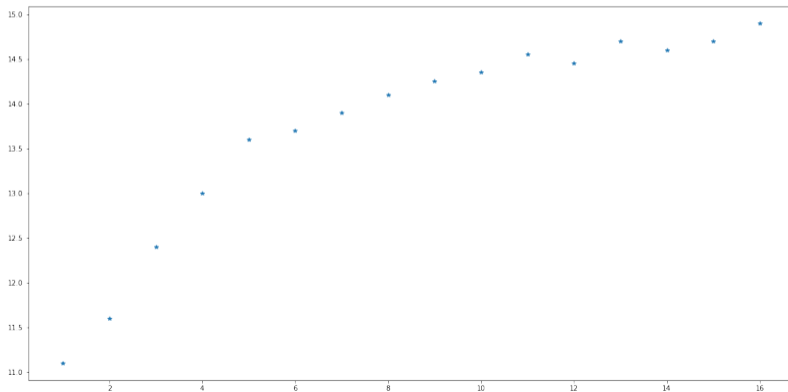
$$\boxed{\hat{\beta}_1 = \frac{a}{1 - \exp(\hat{\beta}_3)}}$$

Logistic function

$$f(t) = f(t; \beta_1, \beta_2, \beta_3) = \frac{\beta_3}{1 + \beta_2 \exp(-\beta_1 t)}, \quad \beta_1, \beta_2 > 0, \beta_3 \in \mathbb{R} - \{0\}$$

Παράδειγμα

{11.1, 11.6, 12.4, 13.0, 13.6, 13.7, 13.9, 14.1, 14.25, 14.35, 14.55, 14.45, 14.7, 14.6, 14.7, 14.9}



Χρονολογικές Σειρές (Time Series)

