ΜΕΜ-205 Περιγραφική Στατιστική

Τμήμα Μαθηματικών και Εφ. Μαθηματικών, Πανεπιστήμιο Κρήτης

Κώστας Σμαραγδάκης (kesmarag@gmail.com)

Θεωρία 9ης εβδομάδας

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
- Το τμήμα του φοιτητή/τριας (πχ. tem, math, csd)
- Παρακολούθηση τουλάχιστον των μισών μαθημάτων μετά τη πρόοδο

$$\hat{y} = \alpha + \beta^{(1)}x^{(1)} + \beta^{(2)}x^{(2)} + \beta^{(3)}x^{(3)} + \beta^{(4)}x^{(4)} + \beta^{(5)}x^{(5)}$$

0 - < 50%
1 - ≥ 50%
$X^{(3)}$

$$\begin{array}{c|ccc} & X^{(4)} & X^{(5)} & \\ \text{tem} & 1 & 0 & 0 \\ \text{math} & 0 & 1 & 0 \\ \text{csd} & 0 & 0 & 1 \end{array}$$

Dataset

Y						
8	7	6.8	1	0	1	math
9	6	7	0	0	0	csd
5	3	7.5	0	1	0	tem

$$X = \begin{bmatrix} 1 & 7 & 6.8 & 1 & 0 & 1 \\ 1 & 6 & 7 & 0 & 0 & 0 \\ 1 & 3 & 7.5 & 0 & 1 & 0 \end{bmatrix}$$

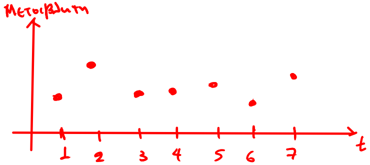
$$p = \begin{bmatrix} \alpha \\ \beta^{(1)} \\ \beta^{(2)} \\ \beta^{(3)} \\ \beta^{(4)} \\ \beta^{(5)} \end{bmatrix} \qquad p = (X^T X)^{-1} X^T y \qquad y = \begin{bmatrix} 8 \\ 9 \\ 5 \end{bmatrix}$$

or

$$(X^T X)\, p = X^T y$$

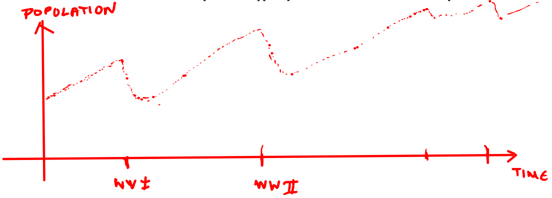
2 7 6.5 0 0 1

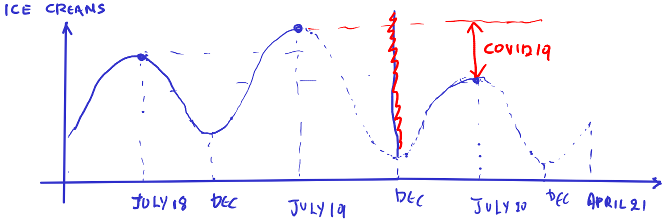
$$\alpha + \beta^{(1)} \cdot 7 + \beta^{(2)} \cdot 6.5 + \beta^{(3)} \cdot 0 + \beta^{(4)} \cdot 0 + \beta^{(5)} \cdot 1$$



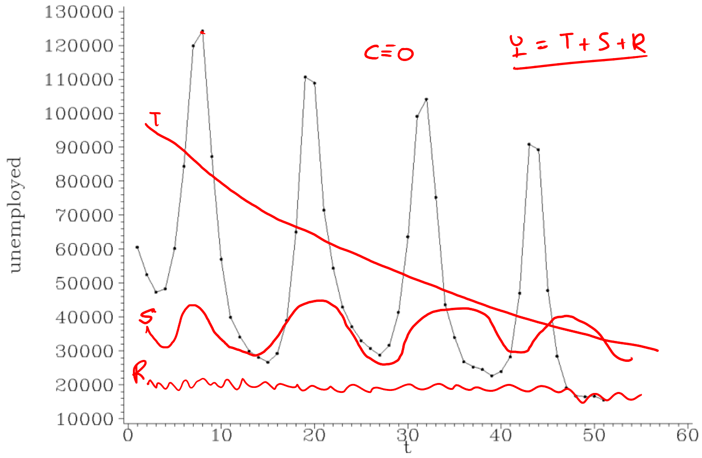
Μια χρονολογική σειρά είναι ένα σύνολο παρατηρήσεων που παρουσιάζονται σε χρονολογική διάταξη.

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





Χρονολογικές Σειρές (Time Series)



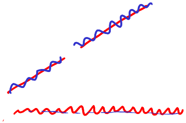
Χρονολογικές Σειρές (Time Series)

Το προσθετικό μοντέλο για χρονολογικές σειρές

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

- 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



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

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

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

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

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

Linear function

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

$t = 1, \ldots, N$ $\qquad Y_1, \ldots, Y_N$

$\{(1, Y_1), \ldots, (N, Y_N)\}$

$a \quad b \leftarrow$ της γραμμικής παλινδρόμησης

$b = \dfrac{SS_{xy}}{SS_{yy}} \qquad a = \bar{Y} - b\bar{X}$

$t_1 = 1,\ t_2 = 2,\ \ldots,\ t_N = N$

$y_1 \quad y_2 \quad \ldots \quad y_N$

Logistic function

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

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

$$= \frac{1 + \exp\{-\beta_1\} - \exp\{-\beta_1\} + \beta_2 \exp(-\beta_1) \exp\{-\beta_1 (t-1)\}}{\beta_3} =$$

$$= \frac{1 - \exp\{-\beta_1\} + \exp\{-\beta_1\} (1 + \beta_2 \exp\{-\beta_1 (t-1)\})}{\beta_3} =$$

$$= \frac{1 - \exp\{-\beta_1\}}{\beta_3} + \exp\{-\beta_1\} \cdot \frac{1 + \beta_2 \exp\{-\beta_1 (t-1)\}}{\beta_3}$$

$$\frac{1}{f(t-1)}$$

$\{(1,Y_1),\dots,(N,Y_N)\}$

Logistic function

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

όπου $\alpha = \dfrac{1-\exp\{-\beta_1\}}{\beta_3}$

$b = \exp\{-\beta_1\}$

$$\Rightarrow \underbrace{\frac{1}{f(t)}}_{y_t} = a + b\cdot\underbrace{\frac{1}{f(t-1)}}_{y_{t-1}} \longleftrightarrow \hat{y} = \alpha + bX$$

New Dataset, N-1 στοιχεία!!!

$\{(Y_1,Y_2),(Y_2,Y_3),\dots,(Y_{N-1},Y_N)\}$

$\hat{y}_t = \alpha + b\,y_{t-1}$, $t=2,\dots,N$

$\hat{Y}_{t+1} = \alpha + bY_t$, $t=1,\dots,N-1$

Μπορούμε να υπολογίσουμε τα α, b (βλέπε γραμμική παλινδρόμηση).

Έχοντας $b=\exp\{-\beta_1\} \Rightarrow \boxed{\hat{\beta}_1 = -\ln b} \Rightarrow \boxed{\hat{\beta}_3 = \dfrac{1-b}{a}}$

Logistic function

$$\{(1, Y_1), (2, Y_2), \ldots, (N, Y_N)\}$$

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

$$\hat{y} = \frac{\hat{\beta}_3}{1 + \beta_2 \exp\{-\hat{\beta}_1 t\}}$$

Παραδοχή

$$t = 1 \Rightarrow \hat{y} = 1 \Rightarrow \frac{\hat{\beta}_3}{1 + \beta_2 \exp\{-\hat{\beta}_1\}} = Y_1$$

και λύνουμε ως προς β_2

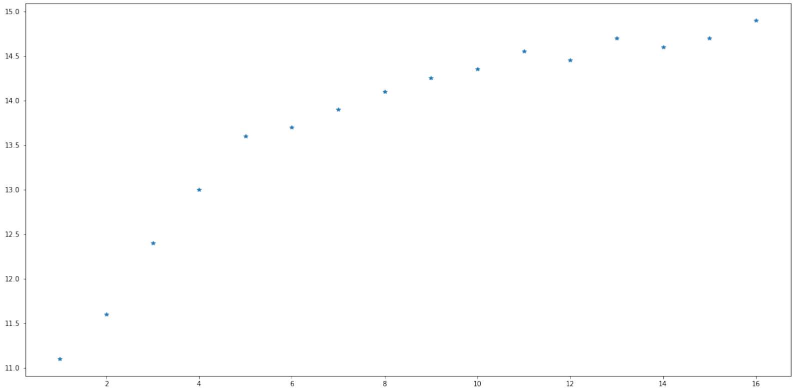
$$\boxed{\hat{\beta}_2 = (Y_1 \hat{\beta}_3 - 1)\exp(\hat{\beta}_1)}$$

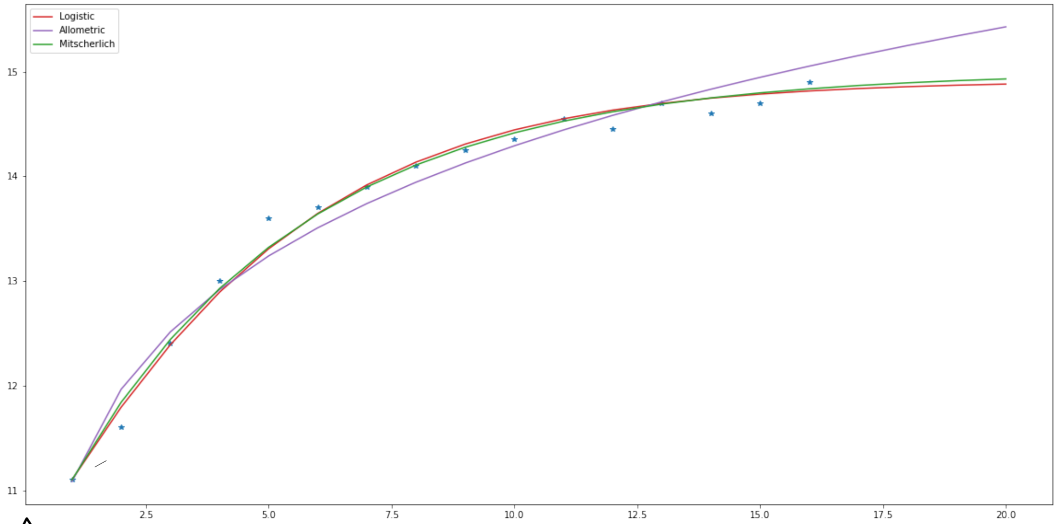
Παράδειγμα

Άσκηση: Γράψτε το Νέο Dataset για την επίλυση του προβλήματος

t 1 2 3 4 5 6 7 16

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





Τέλος 9ης εβδομάδας.

Το προσθετικό μοντέλο για χρονολογικές σειρές

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

- 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

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

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

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

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

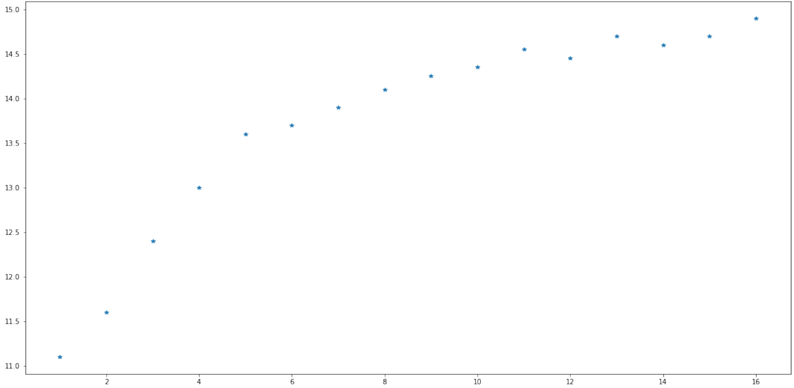
$$\hat{y}_t = f(t; \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p)$$

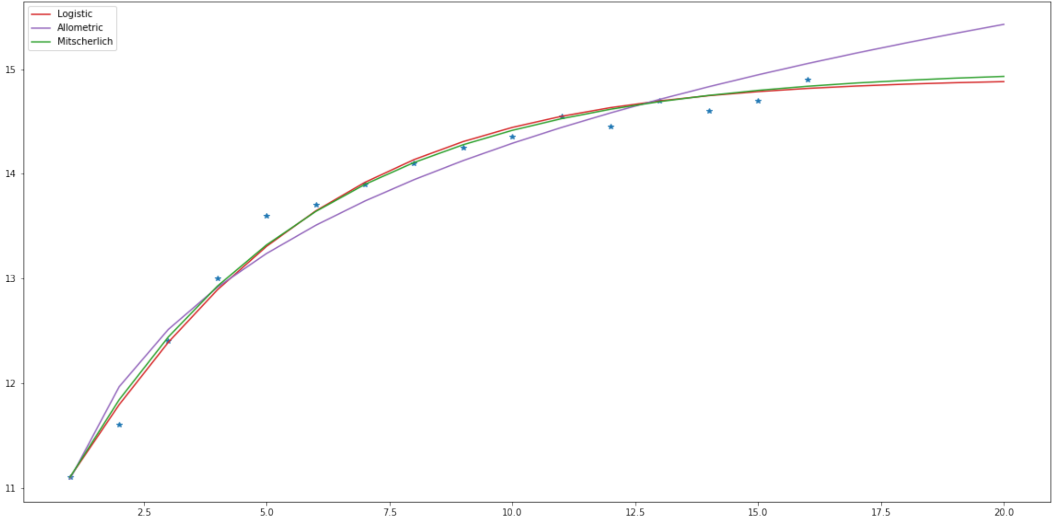
Logistic function

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

Παράδειγμα

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




Logistic
Allometric
Mitscherlich
15
14
13
12
11
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

Εφαρμογή γραμμικού φίλτρου στη χρονολογική σειρά

$$\mathbf{a} = [a_{-s}, \ldots, a_s]^T, \quad \mathbf{a}^T\mathbf{a} = 1,\ a_u \geq 0$$

$$Y_t^* = \sum_{u=-s}^{s} a_u Y_{t+u}$$

Απλός Κινητός Μέσος (Simple Moving Average)

- Απλός κινητός μέσος τάξης $2s+1$

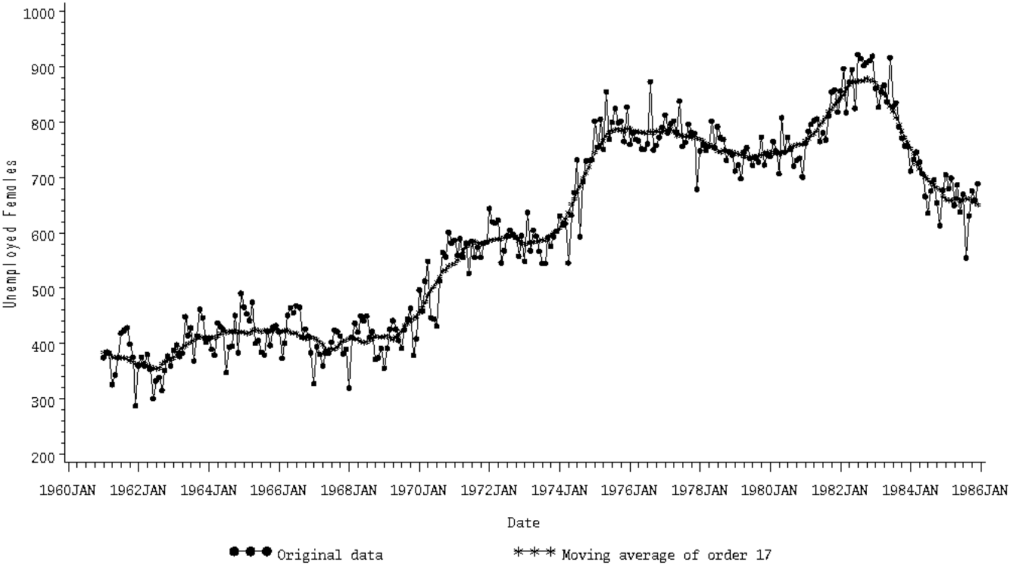
$$a_u = \frac{1}{2s+1}, \quad u = -s, \ldots, s$$

- Απλός κινητός μέσος τάξης $2s$

$$a_u = \frac{1}{2s}, \quad u = -s+1, \ldots, s-1, \quad a_{-s} = a_s = \frac{1}{4s}$$

Παράδειγμα

Ποιά είναι τα διανύσματα συντελεστών για τα γραμμικά φίλτρα που αντιστοιχούν στους κινητούς μέσους με τάξεις 4 και 5;


1000
900
800
700
600
500
400
300
200
Unemployed Females
1960JAN 1962JAN 1964JAN 1966JAN 1968JAN 1970JAN 1972JAN 1974JAN 1976JAN 1978JAN 1980JAN 1982JAN 1984JAN 1986JAN
Date
Original data
Moving average of order 17

Παράδειγμα

Εφαρμόστε το φίλτρο για τον απλό κινητό μέσο 3ης τάξεως στην παρακάτω χρονολογική σειρά

$$\{1,\ 3,\ 5,\ 4,\ 6,\ 5,\ 7\}$$

Απλός Κινητός Μέσος (Simple Moving Average)

Έστω ότι η S_t είναι p-periodic συνάρτηση, δηλαδή

$$S_t = S_{t+p}, \quad t = 1, \ldots, N - p$$

Εάν εφαρμόσουμε το φίλτρο για τον απλό κινητό μέσο p τάξεως, θα έχουμε:

$$S_t^* = S, \quad t = 1 + s, 1 + s + 1, \ldots, N - s$$

Παράδειγμα

$$\{0,\ 2,\ 4,\ 3,\ 1,\ 0,\ 2,\ 4,\ 3,\ 1,\ 0,\ 2,\ 4,\ 3,\ 1\}$$

Παράδειγμα

$$\{0,\ 3,\ 4,\ 1,\ 0,\ 3,\ 4,\ 1,\ 0,\ 3,\ 4,\ 1\}$$