# MEM-205 Περιγραφική Στατιστική

## Τμήμα Μαθηματικών και Εφ. Μαθηματικών, Πανεπιστήμιο Κρήτης

Κώστας Σμαραγδάκης (kesmarag@gmail.com)

Θεωρία 9ης εβδομάδας

$X^{(1)}, X^{(2)}, \dots, X^{(k)}$

$Y$

$$\hat{y} = \underline{\alpha} + \underline{b}^{(1)} X^{(1)} + \dots + \underline{b}^{(k)} X^{(k)}$$

$\alpha, b^{(1)}, \dots, b^{(k)}$

$$\beta = \begin{bmatrix} \alpha \\ b^{(1)} \\ \vdots \\ b^{(k)} \end{bmatrix}$$

Δεδομένα

$$\{(x_1^{(1)}, \dots, x_1^{(k)}, y_1), \dots, (x_n^{(1)}, \dots, x_n^{(k)}, y_n)\}$$

$n \times (k+1)$

$$\hat{\beta} = (X^T X)^{-1} X^T y \in \mathbb{R}^{k+1}$$

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad (n \times 1)$$

$$X = \begin{bmatrix} 1 & x_1^{(1)} & \cdots & x_1^{(k)} \\ 1 & & & \\ \vdots & & & \\ 1 & x_n^{(1)} & \cdots & x_n^{(k)} \end{bmatrix}$$

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

$X^{(3)}$ $X^{(4)}$

$\hat{y} = \alpha \ldots$

tem $\leftrightarrow$ [1, 0, 0]

math $\leftrightarrow$ [0, 1, 0]

csd $\leftrightarrow$ [0, 0, 1]

tem : $X^{(3)} = 1$ $X^{(4)} = 0$

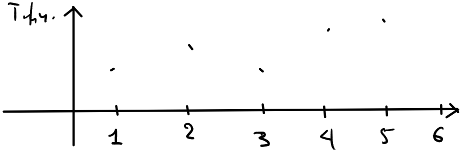
math : $X^{(3)} = 0$ $X^{(4)} = 1$

csd : $X^{(3)} = 0$ $X^{(4)} = 0$

$X^{(5)} = 0$ ή $1$

$$\hat{b} = (X^T X)^{-1} X^T y$$

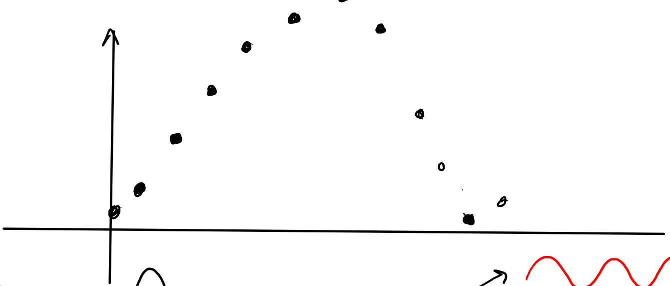
$$\hat{y} = \alpha + b^{(1)} X^{(1)} + b^{(2)} X^{(2)} + b^{(3)} X^{(3)} + b^{(4)} X^{(4)} + b^{(5)} X^{(5)}$$

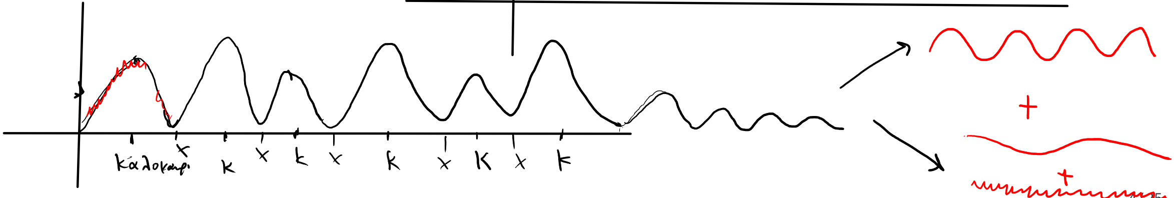


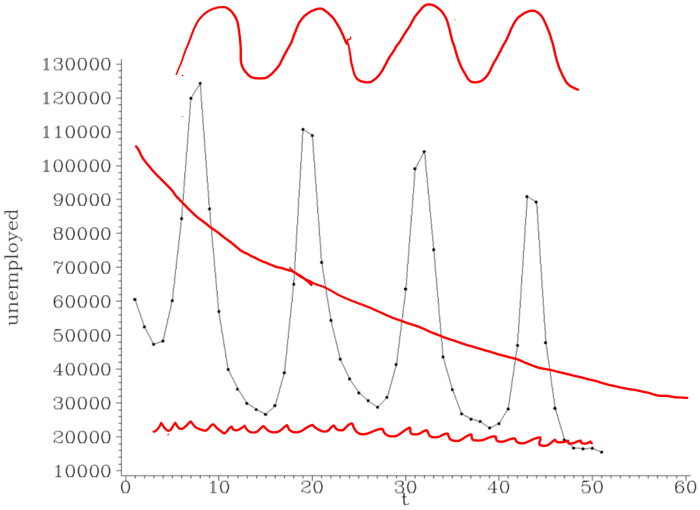


Μια χρονολογική σειρά είναι ένα σύνολο παρατηρήσεων που παρουσιάζονται σε χρονολογική διάταξη.

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







130000
120000
110000
100000
90000
80000
70000
60000
50000
40000
30000
20000
10000
unemployed
0
10
20
30
40
50
60
t

$Y_t = Y(t)$

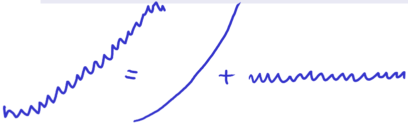
### Το προσθετικό μοντέλο για χρονολογικές σειρές

$$Y_t = \underline{T_t} + \cancel{C_t}^{\,0} + \underline{S_t} + R_t, \quad t = 1, \ldots, N$$

- $T_t$ : 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

$f(t; \beta_1, \beta_2, \dots, \beta_p)$



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

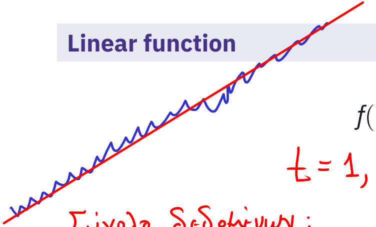
$$\mathbb{E}\{R_t\} = 0, \quad \mathbb{E}\{Y_t\} = T_t \equiv f(t)$$

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

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

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

$Y = A + Bx + \varepsilon$



## Linear function

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

$t = 1, \dots, n$

Σύνολο δεδομένων : $\{(1, y_1), \dots, (n, y_n)\}$

$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$

$\alpha = \bar{Y} - b \frac{1}{n} \sum_{i=1}^{n} i$

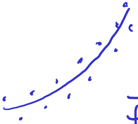
$b = \frac{SS_{xy}}{SS_{tt}}$

$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & n \end{bmatrix}$ $\quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$

$\hat{\beta} = (X^T X)^{-1} X^T y$

## Logistic function

$t = 1, \dots, n$



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

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

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

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

$$\frac{1}{y_t} = \underbrace{\frac{1-\exp\{-\beta_1\}}{\beta_3}}_{\alpha} + \underbrace{\exp\{-\beta_1\}}_{b}\,\frac{1}{f(t-1)}$$

$$y_t = f(t)$$

$$\frac{1}{y_t} = \frac{1}{f(t)}$$

$$\frac{1}{y_t} \;\text{||}$$

$$\frac{1}{f(t)} = \alpha + b\,\frac{1}{f(t-1)}$$

$$y = \alpha + bx$$

$$(x_1, y_1)$$

Νεο συνολο δεδομενων.

$$\left\{\left(\frac{1}{y_1}, \frac{1}{y_2}\right), \left(\frac{1}{y_2}, \frac{1}{y_3}\right), \ldots, \left(\frac{1}{y_{n-1}}, \frac{1}{y_n}\right)\right\} \quad n-1 \longrightarrow \hat{\alpha}, \hat{b}$$

με γραμμική παλινδρομηση.

$\hat{b}$ γνωστο. $\Rightarrow \exp\{-\hat{\beta}_1\} = \hat{b} \Rightarrow -\hat{\beta}_1 = \ln\hat{b} \Rightarrow \hat{\beta}_1 = -\ln\hat{b}$

$$\frac{1-\exp\{-\hat{\beta}_1\}}{\hat{\beta}_3}=\hat{\alpha} \Rightarrow \frac{1-\hat{b}}{\hat{\beta}_3}=\hat{\alpha}$$

$$\rightarrow \hat{\beta}_3=\frac{1-\hat{b}}{\hat{\alpha}}$$

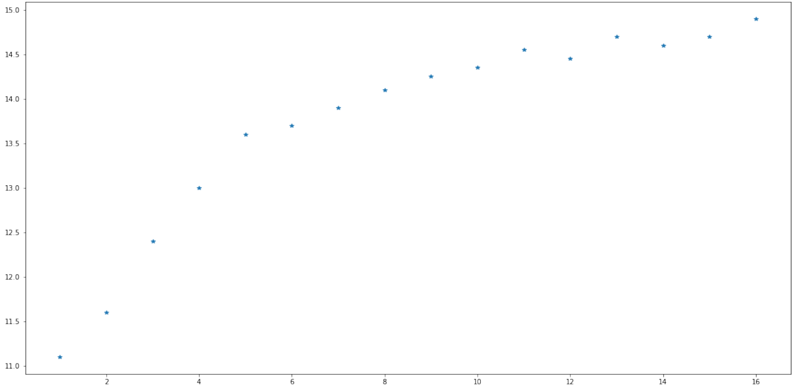
Επιλέγω: μηδενικό σφάλμα στο $t=1$


$$\hat{y}_1 = \boxed{y_1 = \frac{\hat{\beta}_3}{1+\hat{\beta}_2\exp(-\hat{\beta}_1 t)}}$$

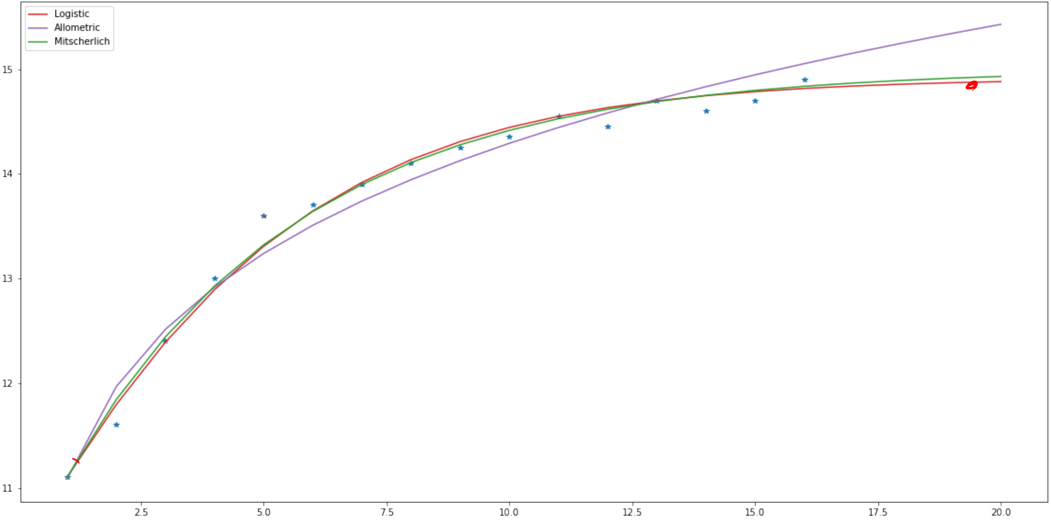
Λύνω ως προς $\hat{\beta}_2$

## 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)



**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 $f$.

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

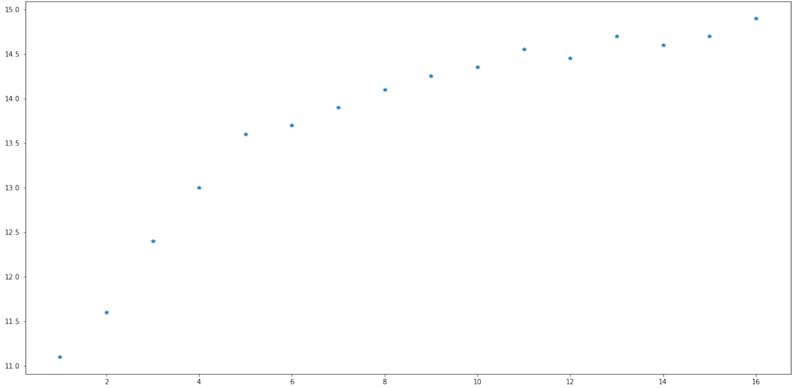
$$\hat{y}_t = f(t; \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p)$$

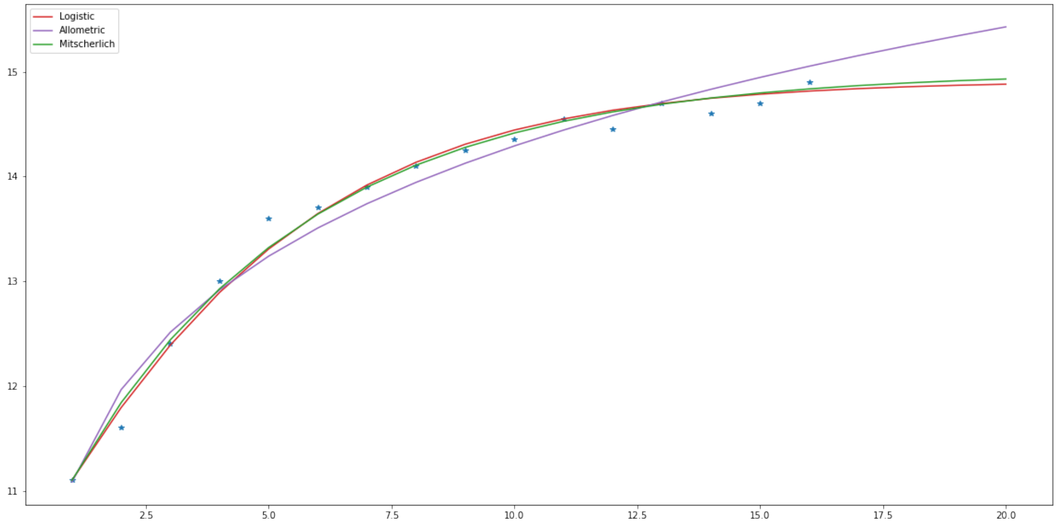
## Logistic function

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

## 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
11
12
13
14
15
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 $2s+1$

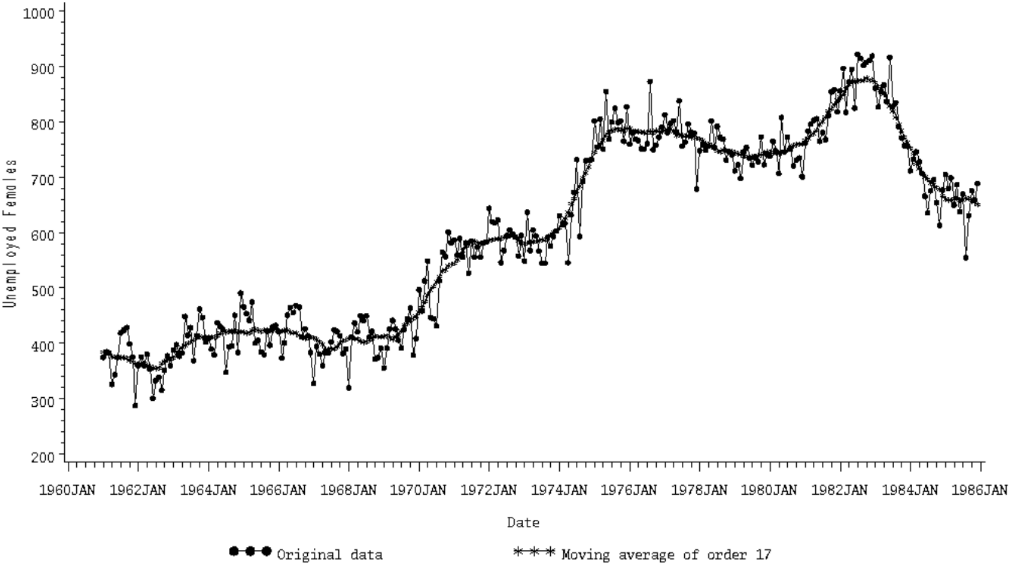
$$a_u = \frac{1}{2s+1}, \quad u = -s, \ldots, s$$

- 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

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