

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
Μονάδα Προβλέψεων & Στρατηγικής

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

Προετοιμασία & Ανάλυση Χρονοσειράς
Διάλεξη 1



Απεικόνιση δεδομένων

Γραφική Αναπαράσταση

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

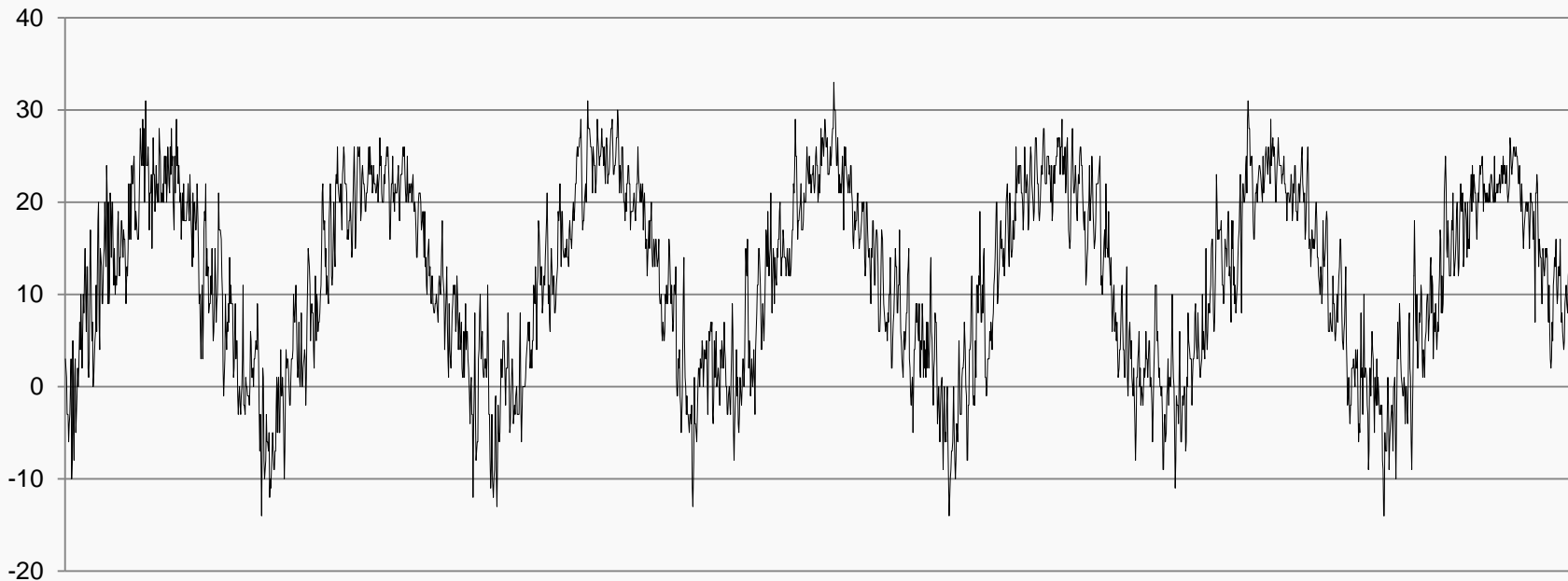
- Η γραφική αναπαράσταση των δεδομένων αποτελεί βασικό εργαλείο της ανάλυσης χρονοσειρών αλλά και της διαδικασίας πρόβλεψης
- Η αναπαράσταση έγκειται ουσιαστικά στη δισδιάστατη γραφική απεικόνιση των πραγματικών τιμών των διαθέσιμων ιστορικών δεδομένων ως προς το χρόνο
- Από την αναπαράσταση των δεδομένων καθίστανται εμφανή τα ποιοτικά χαρακτηριστικά της χρονοσειράς (τάση, εποχιακότητα, κύκλος και τυχαιότητα), τα οποία βοηθούν τον αναλυτή να επιλέξει μεταξύ των εναλλακτικών τεχνικών, τις πλέον κατάλληλες ούτως ώστε να έχει τα βέλτιστα αποτελέσματα και το μικρότερο σφάλμα πρόβλεψης
- Επιπλέον, η γραφική απεικόνιση των δεδομένων ενδέχεται να αποκαλύψει ακραίες ή εσφαλμένες τιμές. Ο αναλυτής μπορεί βάσει των παρατηρήσεών του να προχωρήσει σε κατάλληλες κινήσεις ούτως ώστε να διορθώσει τις εν λόγω τιμές

Απεικόνιση δεδομένων

Γραφική Αναπαράσταση

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

Aaronsburg (Pennsylvania) - Daily temperature (in celsius degrees)

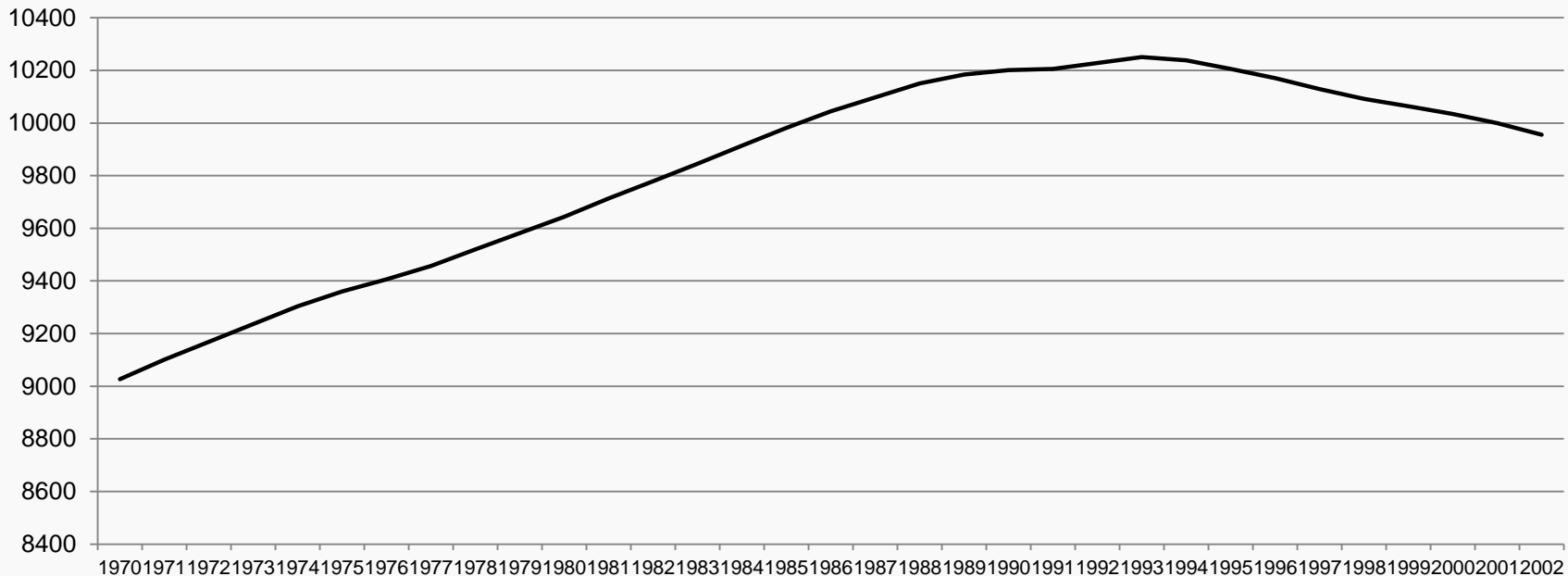


Απεικόνιση δεδομένων

Γραφική Αναπαράσταση

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

Belarus - Midyear population (in thousands)



Απεικόνιση δεδομένων

Γραφική Αναπαράσταση

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

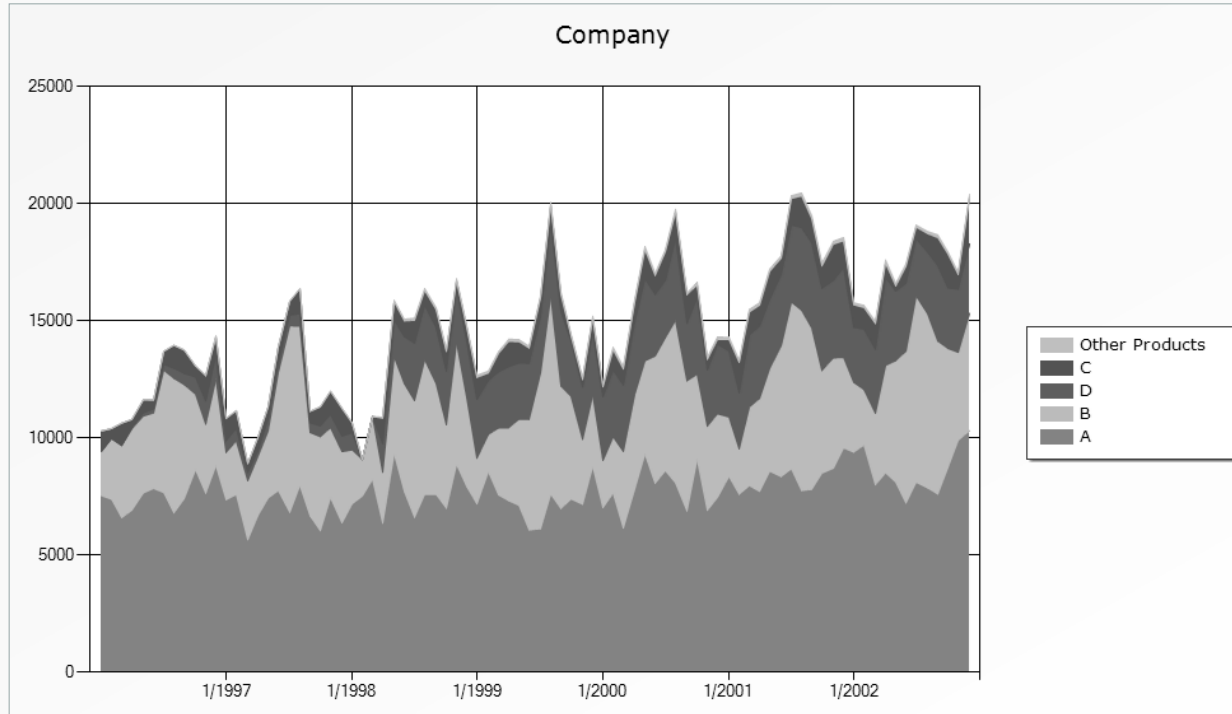
Gold Price (in \$ per ounce)



Απεικόνιση δεδομένων

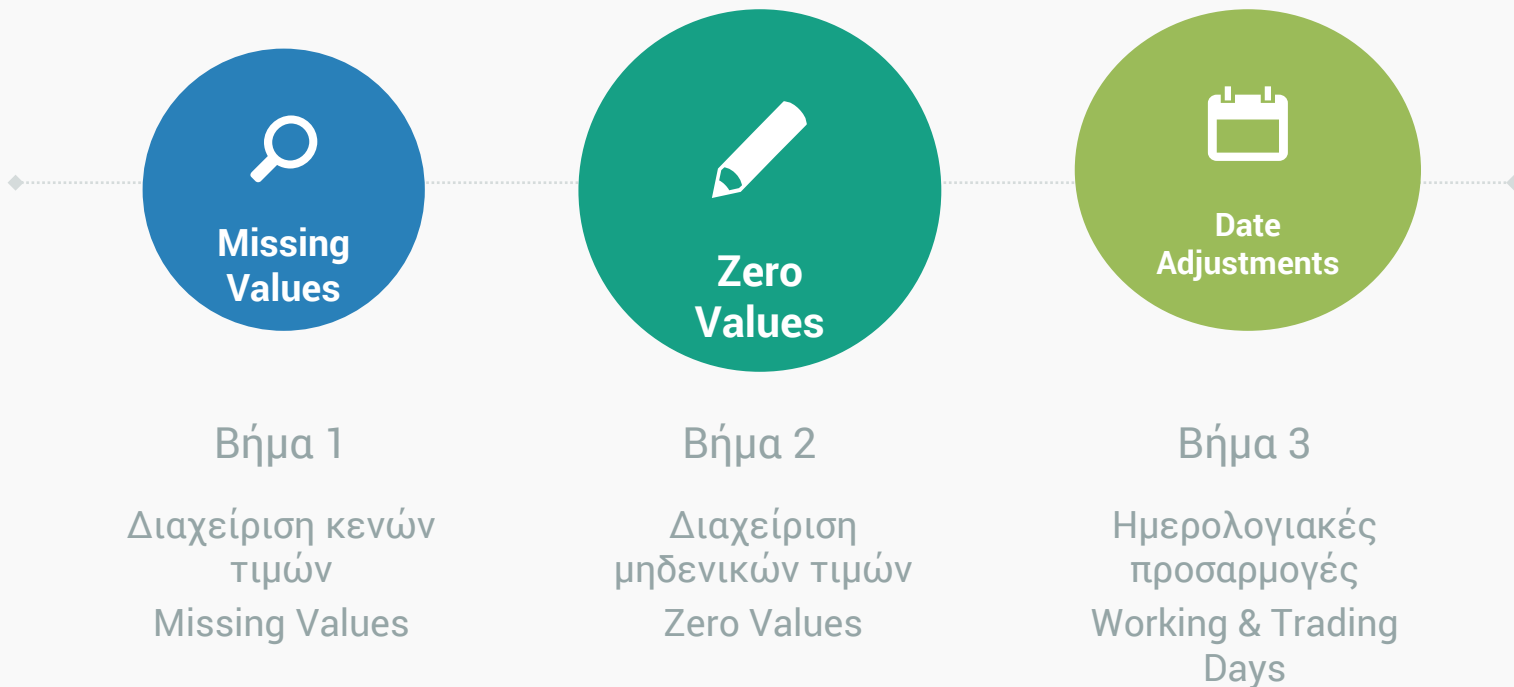
Γραφική Αναπαράσταση

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές



Προσαρμογή Δεδομένων

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές



Διαχείριση κενών τιμών

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές



Γίνεται προσπάθεια αντικατάστασης της κενής τιμής βάσει άλλων που είναι διαθέσιμες ή σχετικής πληροφόρησης

Η κενή τιμή συμπληρώνεται συνήθως με το **ημιάθροισμα** (μέσος όρος) της προηγούμενης και της επόμενης παρατήρησης

Βασική προϋπόθεση για να είναι αυτή η τεχνική αποτελεσματική, είναι η χρονοσειρά να χαρακτηρίζεται από **στασιμότητα**

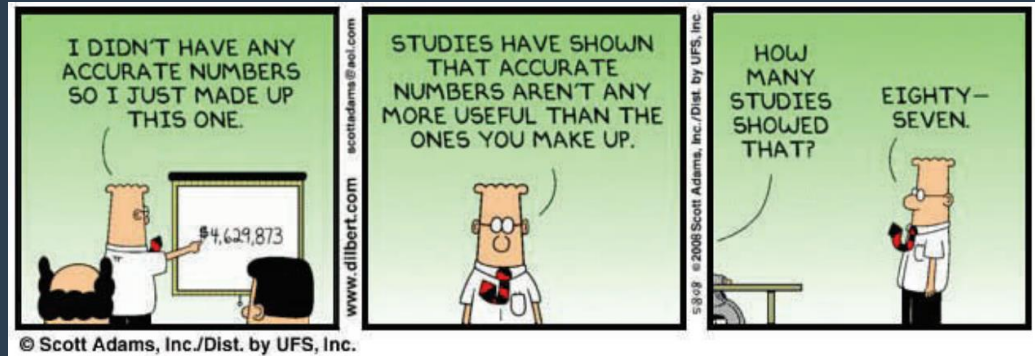
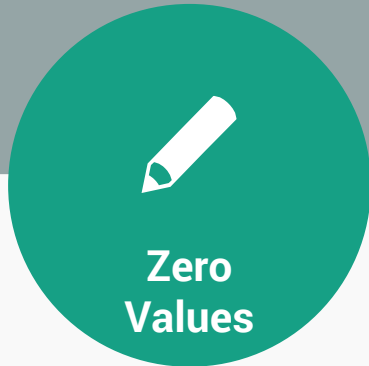
Αν η χρονοσειρά παρουσιάζει **εποχιακή συμπεριφορά**, τότε η κενή τιμή συμπληρώνεται από το μέσο όρο των τιμών των **αντίστοιχων περιόδων**

Για παράδειγμα, αν τα δεδομένα αποτελούνται από μηνιαίες παρατηρήσεις και παρατηρηθεί κενή τιμή στο Μάρτιο κάποιου έτους, τότε η κενή αυτή τιμή ορίζεται ως ο μέσος όρος των λοιπών Μαρτίων

Διαχείριση μηδενικών τιμών

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

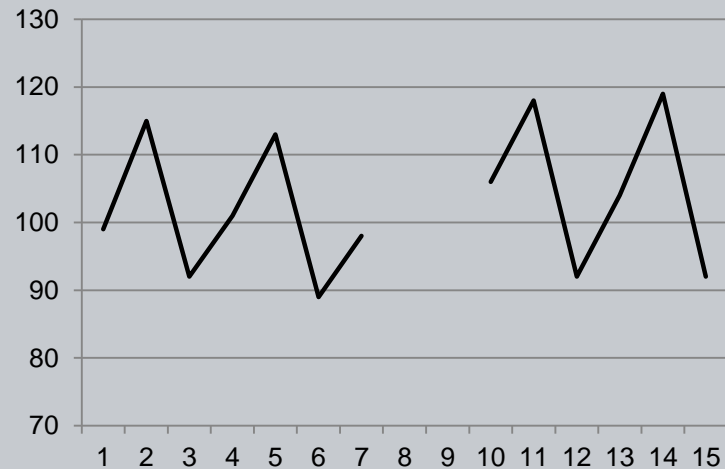
- Καμία αλλαγή στις μηδενικές τιμές όταν είναι λίγες
- Διαχείριση μηδενικών τιμών όπως τις κενές τιμές



Διαχείριση κενών τιμών

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

t	Τετράμηνο	Y_t
1	1ο	99
2	2ο	115
3	3ο	92
4	1ο	101
5	2ο	113
6	3ο	89
7	1ο	98
8	2ο	
9	3ο	
10	1ο	106
11	2ο	118
12	3ο	92
13	1ο	104
14	2ο	119
15	3ο	92



Διαχείριση κενών τιμών

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

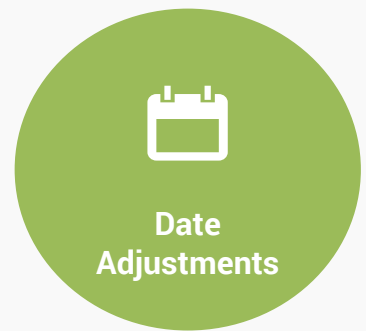
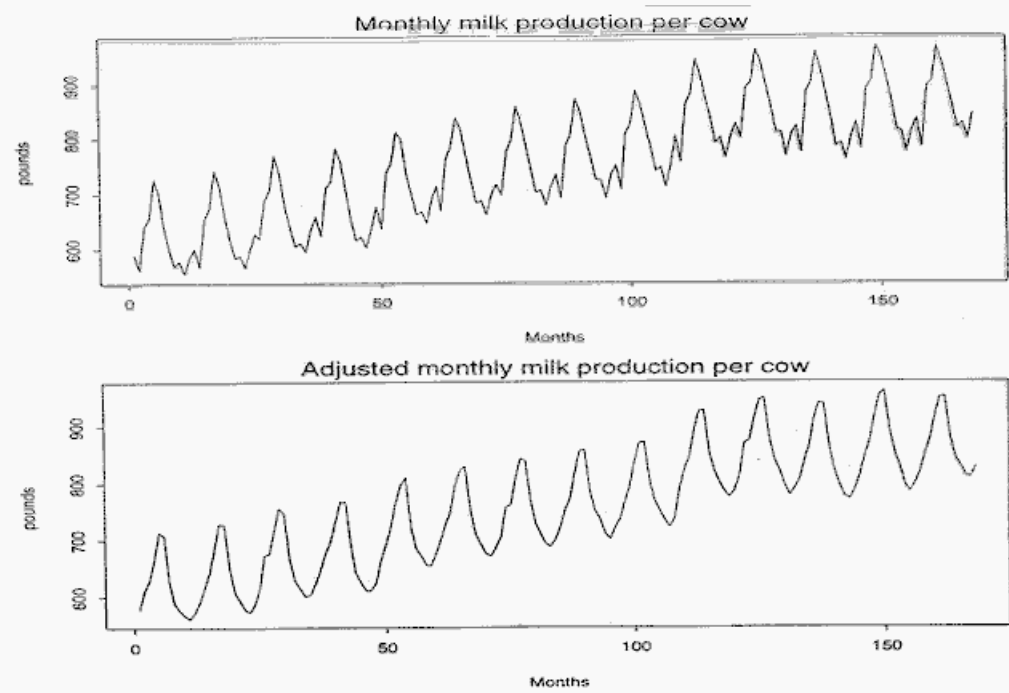
t	Τετράμηνο	Y _t
1	1ο	99
2	2ο	115
3	3ο	92
4	1ο	101
5	2ο	113
6	3ο	89
7	1ο	98
8	2ο	116,25
9	3ο	91,25
10	1ο	106
11	2ο	118
12	3ο	92
13	1ο	104
14	2ο	119
15	3ο	92

$$Y_8 = \frac{Y_2 + Y_5 + Y_{11} + Y_{14}}{4} = \frac{115 + 113 + 118 + 119}{4} = 116,25$$

$$Y_9 = \frac{Y_3 + Y_6 + Y_{12} + Y_{15}}{4} = \frac{92 + 89 + 92 + 92}{4} = 91,25$$

Ημερολογιακές Προσαρμογές

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές



Ημερολογιακές Προσαρμογές

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

1. Καθορισμός των εργάσιμων ημερών ή ημερών συναλλαγών (*trading days*) στη διάρκεια μιας εβδομάδας
2. Καθορισμός της χώρας που εδρεύει η επιχείρηση και εύρεση των επίσημων αργιών (*bank holidays*) αυτής
3. Υπολογισμός, βάσει των παραπάνω, του πλήθους των εργάσιμων ημερών για κάθε περίοδο που συμπεριλαμβάνεται στο χρονικό διάστημα των διαθέσιμων δεδομένων
4. Υπολογισμός του μέσου όρου των εργάσιμων ημερών για όλες τις περιόδους που εξετάζονται.
5. Εξομάλυνση της τιμής κάθε διαθέσιμης περιόδου, σύμφωνα με τον παρακάτω τύπο:

$$Y'_t = Y_t \cdot \frac{\text{number of average trading days in all periods}}{\text{number of trading days in current period (t)}}$$

Ημερολογιακές Προσαρμογές

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

t	Y _t	WD	Y' _t
1	68	118	69,96
2	125	123	123,37
3	73	122	72,64
4	121	121	121,40
5	80	125	77,70
6	115	117	119,32
7	76	120	76,89
8	123	121	123,41
9	79	123	77,97
10	132	124	129,23
Μέσος Όρος		121,4	

$$\begin{aligned} \text{Average} &= \frac{WD_1 + WD_2 + \dots + WD_{10}}{10} \\ &= \frac{118 + 123 \dots + 124}{10} \end{aligned}$$

$$Y'_1 = Y_1 \cdot \frac{\text{Average}}{WD_1} = 68 \cdot \frac{121,4}{118} = 69,96$$

Βασική στατιστική ανάλυση

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

t	Δεδομένα
1	Y_1
2	Y_2
3	Y_3
...	...
n-2	Y_{n-2}
n-1	Y_{n-1}
n	Y_n

- Μέση τιμή (*Average*)

$$\bar{Y} = \frac{1}{n} \cdot \sum_{i=1}^n Y_i$$

- Μέγιστη και ελάχιστη τιμή (*Maximum* και *Minimum*)

- Τυπική απόκλιση (*Standard Deviation*)

$$\sigma_{\text{δείγματος}} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n - 1}}$$

$$\sigma_{\text{πληθυσμού}} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n}}$$

Βασική στατιστική ανάλυση

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

- Διακύμανση (*Variance*)
- Συνδιακύμανση (*Covariance*)
- Συντελεστής γραμμικής συσχέτισης (*Linear Correlation Coefficient*)
 - Αν $r = \pm 1$, υπάρχει τέλεια γραμμική συσχέτιση.
 - Αν $-0,3 < r < 0,3$, δεν υπάρχει γραμμική συσχέτιση. Αυτό, όμως, δεν σημαίνει ότι δεν υπάρχει άλλου είδους συσχέτιση μεταξύ των δύο μεταβλητών.
 - Αν $-0,5 < r \leq -0,3$ ή $0,3 \leq r < 0,5$, υπάρχει ασθενής γραμμική συσχέτιση.
 - Αν $-0,7 < r \leq -0,5$ ή $0,5 \leq r < 0,7$, υπάρχει μέση γραμμική συσχέτιση.
 - Αν $-0,8 < r \leq -0,7$ ή $0,7 \leq r < 0,8$, υπάρχει ισχυρή γραμμική συσχέτιση.
 - Αν $-1 < r \leq -0,8$ ή $0,8 \leq r < 1$, υπάρχει πολύ ισχυρή γραμμική συσχέτιση.

$$\text{Cov}(X, Y) = \frac{1}{n} \cdot \sum_{i=1}^n [(X_i - \bar{X}) \cdot (Y_i - \bar{Y})]$$

$$r_{XY} = \frac{\sum_{i=1}^n [(X_i - \bar{X}) \cdot (Y_i - \bar{Y})]}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \cdot \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

Βασική στατιστική ανάλυση

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

- Συντελεστής αυτοσυσχέτισης (*Autocorrelation Coefficient*).

$$ACF_k = \frac{\sum_{i=1+k}^n [(Y_i - \bar{Y}) \cdot (Y_{i-k} - \bar{Y})]}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

- Συντελεστής μεταβλητότητας (*Coefficient of Variation*).

$$c_v = \frac{\sigma}{\bar{Y}} \cdot 100 \quad (\%)$$

- Μέση τιμή διαστήματος μεταξύ ζητήσεων (*Intermittent Demand Interval*).

Βασική στατιστική ανάλυση

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

Ο δείκτης του **ρυθμού ανάπτυξης** (*growth rate*) αποτελεί ένα μέτρο της αυξητικής ή φθίνουσας πορείας μιας σειράς δεδομένων για ένα συγκεκριμένο χρονικό διάστημα. Εκφράζεται σε ποσοστιαία μορφή και συνήθως αναφέρεται στη σύγκριση του ύψους των δεδομένων του τελευταίου έτους σε σχέση με τα υπόλοιπα διαθέσιμα δεδομένα. Η μαθηματική έκφραση του ρυθμού ανάπτυξης έχει ως εξής:

$$\text{Growth Rate} = \frac{\frac{1}{ppy} \cdot \sum_{i=n-ppy+1}^n Y_i - \frac{1}{n-ppy} \cdot \sum_{i=1}^{n-ppy} Y_i}{\frac{1}{n-ppy} \cdot \sum_{i=1}^{n-ppy} Y_i} \cdot 100 (\%)$$

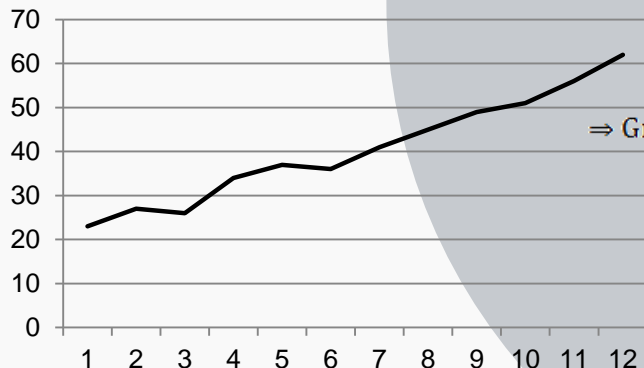
Όπου Y το διάνυσμα των n παρατηρήσεων και ppy το πλήθος των περιόδων στο μήκος ενός έτους (για παράδειγμα, $ppy=12$ αν τα δεδομένα αφορούν μηνιαίες παρατηρήσεις).

Βασική στατιστική ανάλυση

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

Παράδειγμα

t	Y _t
1	23
2	27
3	26
4	34
5	37
6	36
7	41
8	45
9	49
10	51
11	56
12	62



$$\text{Growth Rate} = \frac{\frac{1}{4} \cdot \sum_{i=9}^{12} Y_i - \frac{1}{8} \cdot \sum_{i=1}^8 Y_i}{\frac{1}{8} \cdot \sum_{i=1}^8 Y_i} \cdot 100 (\%)$$

$$\text{Growth Rate} = \frac{\frac{49 + 51 + 56 + 62}{4} - \frac{23 + 27 + \dots + 45}{8}}{\frac{23 + 27 + 26 + 34 + 37 + 36 + 41 + 45}{8}} \cdot 100 (\%)$$

$$\Rightarrow \text{Growth Rate} = \frac{\frac{218}{4} - \frac{269}{8}}{\frac{269}{8}} \cdot 100 (\%) = \frac{54,5 - 33,625}{33,625} \cdot 100 (\%)$$

$$\Rightarrow \text{Growth Rate} = 62,08 \%$$

Στατιστική ανάλυση αξιολόγησης προβλέψεων

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

t	Δεδομένα	Πρόβλεψη
1	Y_1	F_1
2	Y_2	F_2
3	Y_3	F_3
...
n-2	Y_{n-2}	F_{n-2}
n-1	Y_{n-1}	F_{n-1}
n	Y_n	F_n
n+1		F_{n+1}
...		...
n+h		F_{n+h}

Σφάλμα:

$$e_i = Y_i - F_i$$

Στατιστική ανάλυση αξιολόγησης προβλέψεων

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

- Μέσο σφάλμα (*Mean Error*)

$$ME = \frac{1}{n} \sum_{i=1}^n (Y_i - F_i)$$

- Μέσο απόλυτο σφάλμα (*Mean Absolute Error*)

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - F_i|$$

- Μέσο τετραγωνικό σφάλμα (*Mean Squared Error*)

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - F_i)^2$$

- Ρίζα μέσου τετραγωνικού σφάλματος (*Root Mean Squared Error*)

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - F_i)^2}$$

Στατιστική ανάλυση αξιολόγησης προβλέψεων

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

- Μέσο απόλυτο ποσοστιαίο σφάλμα (*Mean Absolute Percentage Error*)

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - F_i}{Y_i} \right| \cdot 100 (\%)$$

- Συμμετρικό μέσο απόλυτο ποσοστιαίο σφάλμα (*Symmetric Mean Absolute Percentage Error*).

$$\text{sMAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - F_i}{\left(\frac{Y_i + F_i}{2}\right)} \right| \cdot 100 = \frac{1}{n} \sum_{i=1}^n \left| \frac{2 \cdot (Y_i - F_i)}{Y_i + F_i} \right| \cdot 100 (\%)$$

Αισιόδοξη πρόβλεψη: $Y_t=100$ και $F_t=110 \rightarrow \text{sMAPE}=4,76\%$

Απαισιόδοξη πρόβλεψη: $Y_t=100$ και $F_t=90 \rightarrow \text{sMAPE}=5,26\%$

Στατιστική ανάλυση αξιολόγησης προβλέψεων

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

Παράδειγμα

t	Y _t	F _t	e _t
1	33	31	2
2	49	42	7
3	52	50	2
4	57	61	-4
5	78	73	5
6	83	85	-2
7	90	94	-4
8	112	103	9
9	118	115	3
10	116	124	-8
11		132	
12		141	

$$e_1 = Y_1 - F_1^1 = 33 - 31 = 2$$

$$ME = \frac{1}{n} \sum_{i=1}^n e_i^1 = \frac{2 + 7 + \dots - 8}{10} = \frac{10}{10} = 1$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |e_i^1| = \frac{|2| + |7| + \dots + |-8|}{10} = \frac{46}{10} = 4,6$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (e_i^1)^2 = \frac{2^2 + 7^2 + \dots + (-8)^2}{10} = \frac{272}{10} = 27,2$$

$$\Rightarrow RMSE = \sqrt{MSE} = 5,22$$

Στατιστική ανάλυση αξιολόγησης προβλέψεων

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

Παράδειγμα

t	Y _t	F _t	e _t
1	33	31	2
2	49	42	7
3	52	50	2
4	57	61	-4
5	78	73	5
6	83	85	-2
7	90	94	-4
8	112	103	9
9	118	115	3
10	116	124	-8
11		132	
12		141	

$$\begin{aligned} \text{MAPE} &= \frac{1}{n} \sum_{i=1}^n \frac{|e_i^1|}{Y_i} \cdot 100 (\%) \\ &= \frac{\frac{|2|}{33} + \frac{|7|}{49} + \dots + \frac{|-8|}{116}}{10} \cdot 100 \Rightarrow \end{aligned}$$

$$\text{MAPE} = \frac{0,619}{10} \cdot 100 = 6,19\%$$

$$\begin{aligned} \text{sMAPE} &= \frac{1}{n} \sum_{i=1}^n \frac{2 \cdot |e_i^2|}{Y_i + F_i} \cdot 100 (\%) \\ &= 2 \cdot \frac{\frac{|2|}{33 + 31} + \frac{|7|}{49 + 42} + \dots + \frac{|-8|}{116 + 124}}{10} \cdot 100 \end{aligned}$$

$$\Rightarrow \text{sMAPE} = 2 \cdot \frac{0,317}{10} \cdot 100 = 6,33\%$$

Στατιστική ανάλυση αξιολόγησης προβλέψεων

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

Παράδειγμα

t	Y _t	F _t	e _t
1	33	31	2
2	49	42	7
3	52	50	2
4	57	61	-4
5	78	73	5
6	83	85	-2
7	90	94	-4
8	112	103	9
9	118	115	3
10	116	124	-8
11		132	
12		141	

$$\begin{aligned} \text{MASe} &= \frac{\frac{1}{n} \sum_{i=1}^n |e_i^1|}{\frac{1}{n-1} \sum_{i=2}^n |Y_i - Y_{i-1}|} \\ &= \frac{|2| + |7| + \dots + |-8|}{\frac{10}{|49 - 33| + |52 - 49| + \dots + |116 - 118|}} \end{aligned}$$

$$\Rightarrow \text{MASe} = \frac{\frac{46}{10}}{\frac{87}{9}} = 0,476$$

Στατιστική ανάλυση αξιολόγησης προβλέψεων

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

Παράδειγμα

t	Y _t	F _t ¹	F _t ²	e _t ¹	e _t ²
1	33	31	32	2	1
2	49	42	43	7	6
3	52	50	51	2	1
4	57	61	58	-4	-1
5	78	73	75	5	3
6	83	85	87	-2	-4
7	90	94	95	-4	-5
8	112	103	105	9	7
9	118	115	114	3	4
10	116	124	126	-8	-10
11		132	134		
12		141	143		

$$MAE_1 = \frac{1}{n} \sum_{i=1}^n |e_i^1| = \frac{|2| + |7| + \dots + |-8|}{10} = \frac{46}{10} = 4,6$$

$$MAE_2 = \frac{1}{n} \sum_{i=1}^n |e_i^2| = \frac{|1| + |6| + \dots + |-10|}{10} = \frac{42}{10} = 4,2$$

$$RelMAE_2 = \frac{MAE_2}{MAE_{\text{benchmark}}} = \frac{4,2}{4,6} = 0,913$$

Using R

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

01 Download & Install R
Windows/ MAC (32/64 bit)

02 Download & Install R
studio

03 Install required packages
`install.packages("fpp",dependencies=TRUE)`

04 Install required packages

- forecast
- tseries
- ggplot2
- fpp

Using R

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

```
monthdays<- rep(c(31,28,31,30,31,30,31,31,30,31,30,31),14)
monthdays[26+(4*12)*(0:2)]<-29
par(mfrow=c(2,1))
plot(milk, main="Monthly milk production per cow", ylab="Pounds",xlab="Years")
plot(milk/monthdays, main="Average milk production per cow per day",ylab="Pounds",
xlab="Years")
```

```
monthdays<- rep(c(31, 28, 31, 30, 31, 30, 31, 31, 30, 31, 30, 31), 14)
monthdays[26+(4*12)*(0:2)]<-29
par(mfrow=c(2,1))
plot(milk, main="Monthly milk production per cow", ylab="Pounds",xlab="Years")
plot(milk/monthdays, main="Average milk production per cow per day",ylab="Pounds", xlab="Years")
summary(milk)
```

Using R

Επιχειρηματικές Προβλέψεις: Μέθοδοι & Τεχνικές

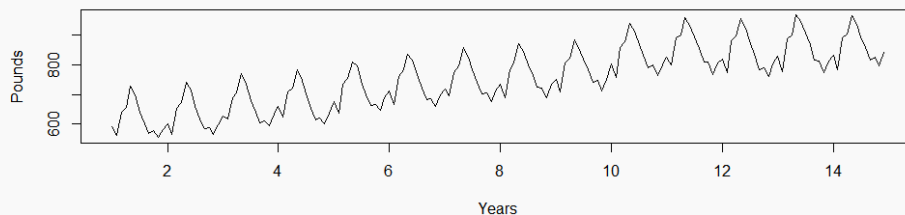
```
monthdays<- rep(c(31,28,31,30,31,30,31,31,30,31,30,31),14)
monthdays[26+(4*12)*(0:2)]<-29
par(mfrow=c(2,1))
plot(milk, main="Monthly milk production per cow", ylab="Pounds",xlab="Years")
plot(milk/monthdays, main="Average milk production per cow per day",ylab="Pounds", xlab="Years")
```

```
summary(milk)
```

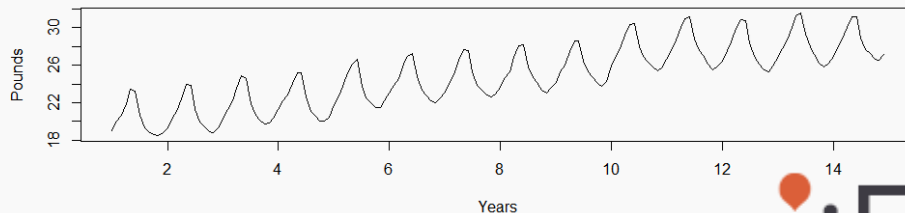
```
summary(milk)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
553.0	677.8	761.0	754.7	824.5	969.0

Monthly milk production per cow



Average milk production per cow per day





Feel free to say hi!

We are friendly and social

Ηρώων Πολυτεχνείου 9, Ζωγράφος
Αττική, 15780, Ελλάδα
Τηλέφωνο: 2107723637 Fax: 2107723740

Κτίριο της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
2ος όροφος - 2.2.1 Εργαστήριο



@FSU NTUA



Μονάδα
Προβλέψεων και
Στρατηγικής ΕΜΠ



spiliotis@fsu.gr

