

---

# Data Analytics και Ευφυή Συστήματα Πρόβλεψης Δεδομένων σε Χρονοσειρά

Εφαρμογή στον Εναρμονισμένο Δείκτη Τιμών Καταναλωτή

13ο Φοιτητικό Συνέδριο Διοικητικής Επιστήμης και Τεχνολογίας  
12 Μαΐου 2016

# OUTLINE

---

01

## Εναρμονισμένος Δείκτης Τιμών Καταναλωτή & Data Analytics Χρονοσειρών

- *Περί Εναρμονισμένου Δείκτη Τιμών Καταναλωτή*
- *Αυτοσυσχέτιση και Αυτοπαλινδρόμηση*
- *Data Analytics Χρονοσειρών*
- *Contribution*

02

## Αλγόριθμοι Τεχνικών Προβλέψεων

- *NNETAR - Τεχνητά Νευρωνικά Δίκτυα Αυτοπαλινδρόμησης (NNAR)*
- *auto.arima - Μοντέλα ARIMA*

03

## Αποτελέσματα

- *Charts & Analysis*
- *Infographic Αποτελεσμάτων*

# Περί Εναρμονισμένου Δείκτη Τιμών Καταναλωτή

---

- Έχει ως σκοπό να παρέχει συγκρίσιμα στοιχεία για την εκτίμηση του κριτηρίου σύγκλισης της σταθερότητας των τιμών στο πλαίσιο των υποχρεώσεων των χωρών της Ε.Ε.
- Υπολογίζεται βάσει των καταναλωτικών δαπανών, που πραγματοποιούνται στην οικονομική επικράτεια κάθε κράτους.

Συγκεκριμένα συμπεριλαμβάνονται

1. οι δαπάνες των ιδιωτικών νοικοκυριών
2. οι δαπάνες των ατόμων που διαμένουν σε ίδρυμα

*\*Εξαιρούνται οι δαπάνες των μόνιμων κατοίκων, όταν βρίσκονται στο εξωτερικό*

# Αυτοσυσχέτιση και Αυτοπαλινδρόμηση

---

- Αυτοσυσχέτιση ονομάζεται η ιδιότητα των παρατηρήσεων μιας χρονοσειράς να εξαρτώνται από τις προηγούμενές τους.
- Για την ανάλυση αυτοσυσχετιζόμενων χρονοσειρών χρησιμοποιούνται οι τιμές υστέρησης (lag values).

Προκύπτουν με μετατόπιση των τιμών της χρονοσειράς κατά ένα αριθμό θέσεων προς τα εμπρός.

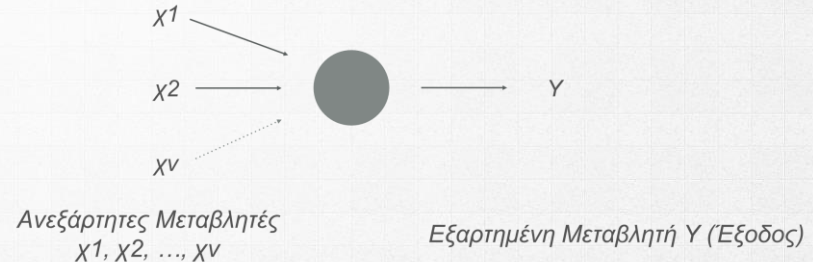


# Αυτοσυσχέτιση και Αυτοπαλινδρόμηση

---

- Στις περιπτώσεις που κάποιος (ή κάποιοι) συντελεστής αυτοσυσχέτισης είναι σημαντικός μπορεί να χρησιμοποιηθεί ένα μοντέλο αυτοπαλινδρόμησης.

- Τα μοντέλα αυτά στηρίζονται σε μια εξίσωση παλινδρόμησης με ανεξάρτητες μεταβλητές, τις τιμές υστέρησης.



# Data Analytics Χρονοσειρών



Τάση



Μακροπρόθεσμο μοτίβο μιας χρονοσειράς. Μπορεί να είναι θετική ή αρνητική.

Εποχικότητα

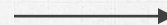


Παρουσιάζεται όταν η χρονοσειρά παρουσιάζει τακτικές διακυμάνσεις κατά τη διάρκεια των ίδιων μηνών κάθε χρόνο.

Τυχαιότητα



Περιλαμβάνει τυχαίες διακυμάνσεις μέσα στη χρονοσειρά που δεν είναι μέρος ούτε του εποχικού μοτίβου, ούτε της τάσης.



# Data Analytics Χρονοσειρών

---

## Κατάλοιπα (Residuals)

Τα κατάλοιπα από ένα μοντέλο παλινδρόμησης υπολογίζονται ως η διαφορά μεταξύ των πραγματικών και των προσαρμοσμένων στο μοντέλο παρατηρήσεων.

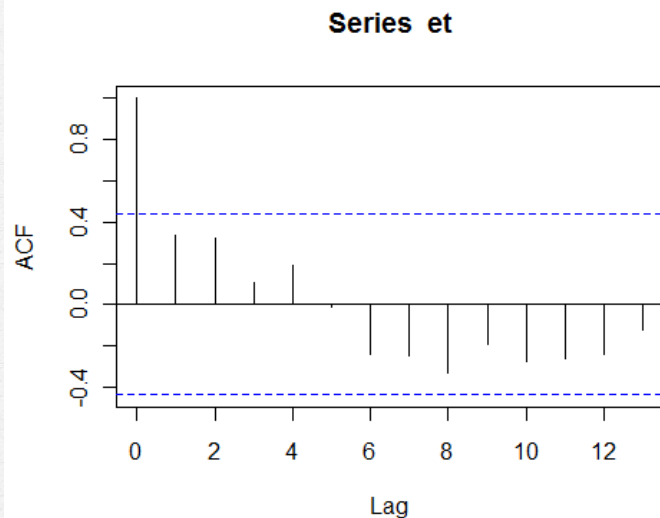
Κάθε κατάλοιπο είναι η μη προβλέψιμη συνιστώσα της αντίστοιχης παρατήρησης.

Είναι της μορφής...

$$e_i = y_i - \hat{y}_i$$

# Data Analytics Χρονοσειρών

## Κατάλοιπα (Residuals)



Τα κατάλοιπα παρουσιάζουν πιο χρήσιμες πληροφορίες μέσα από την Συνάρτηση Αυτοσυσχέτισης (ACF).

Εκεί θα αναδειχθεί εάν υπάρχει αυτοσυσχέτιση καταλοίπων υποδεικνύοντας ότι υπάρχουν δεδομένα τα οποία δεν έχουν ληφθεί υπόψιν για τη δημιουργία του μοντέλου.



# Data Analytics Χρονοσειρών

## Κατάλοιπα (Residuals)

2 στατιστικοί ελέγχοι καταλοίπων

### Box-Ljung Q Statistic

Ο στατιστικός έλεγχος Ljung-Box, (ή τροποποιημένος έλεγχος Box-Pierce) είναι η συνάρτηση των συσσωρευμένων αυτοσυσχετίσεων  $r_j$  σε οποιοδήποτε σημείο τιμής υστέρησης  $m$ . Συναρτηήσει του  $m$ , εκφράζεται ως:

$$Q(m) = n(n+2) \sum_{j=1}^m \frac{r_j^2}{n-j}$$

$n$  = ο αριθμός σημείων δεδομένων που μπορεί να χρησιμοποιηθεί έπειτα από διαδικασίες διαφορίσης

### Durbin-Watson Statistic

Ο στατιστικός έλεγχος Durbin Watson είναι ένας αριθμός που δοκιμάζει εάν υπάρχει αυτοσυσχέτιση στα κατάλοιπα από την ανάλυση παλινδρόμησης. Εκφράζεται ως:

$$d = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2}$$

$e_t$  = το κατάλοιπο της χρονικής στιγμής  $t$   
 $T$  = ο αριθμός των παρατηρήσεων

# Contribution

---

Στόχος μας είναι να

- προσφέρουμε μια αναφορά για το ποια είναι τα σημεία αναφοράς της ανάλυσης δεδομένων χρονοσειρών
- παρουσιάσουμε μια σύγκριση της αποτελεσματικότητας των επιλεγμένων μοντέλων πρόβλεψης χρονοσειρών με τη χρήση δεικτών ασφαμάτων

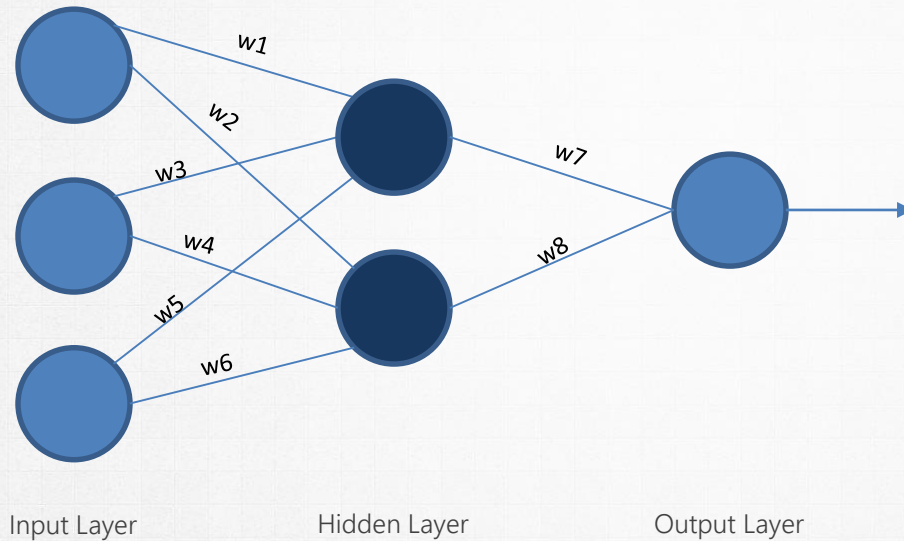
## auto.arima - Μοντέλα ARIMA

---

- Επιστρέφει το καλύτερο δυνατό μοντέλο σύμφωνα με την τιμή ενός από τα κριτήρια AIC, ή BIC.
- Η λειτουργία του εμπεριέχει μια αναζήτηση σε όλα τα δυνατά μοντέλα μέσα στα όρια των μέγιστων τιμών των κριτηρίων.
  - ❑ Ένα από τα μεγαλύτερα πλεονεκτήματα του αλγορίθμου είναι ο χρόνος εξαγωγής ενός ολοκληρωμένου μοντέλου.

Τεχνικές Προβλέψεων

## NNETAR - Τεχνητά Νευρωνικά Δίκτυα Αυτοπαλινδρόμησης (NNAR)



Το μοντέλο NNAR ακολουθεί τα πρότυπα ενός δικτύου Feed Forward



Αποτελέσματα

# Δείκτες Σφαλμάτων

01

RMSE: Root Mean Squared Error  
Μέσο Τετραγωνικό Σφάλμα

$$RMSE = \sqrt{\text{mean}(e_i^2)}$$

$e_i = \text{Πραγματική Τιμή} - \text{Προβλεπόμενη Τιμή } (y_i - \hat{y}_i)$

02

MAPE: Mean Absolute Percentage Error  
Μέσο Απόλυτο Ποσοστιαίο Σφάλμα

$$MAPE = \text{mean}(|p_i|)$$

Ποσοστιαίο σφάλμα πρόβλεψης:  $p_i = 100e_i/y_i$

03

MASE: Mean Absolute Scaled Error  
Μέσο Απόλυτο Κανονικοποιημένο Σφάλμα

$$MASE = \text{mean}(|q_j|)$$

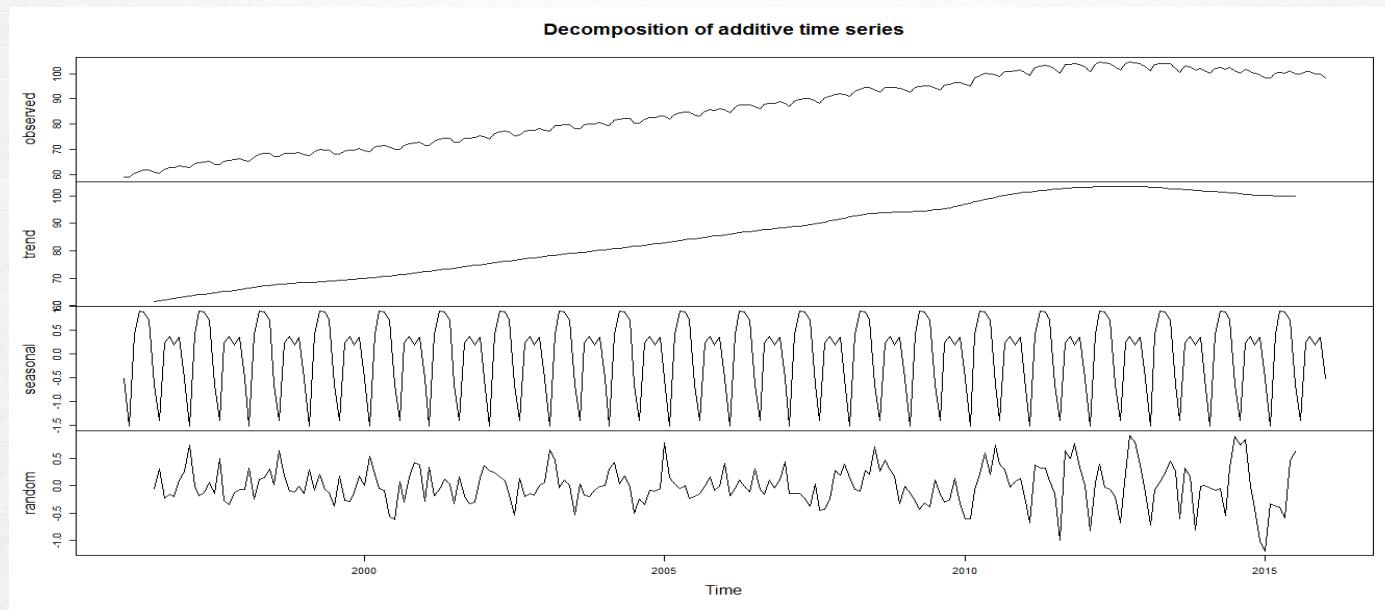
Κανονικοποιημένο σφάλμα:

$$q_j = \frac{e_j}{\frac{1}{T-m} \sum_{t=m+1}^T |y_t - y_{t-m}|}$$

Αποτελέσματα

# CHARTS & ANALYSIS

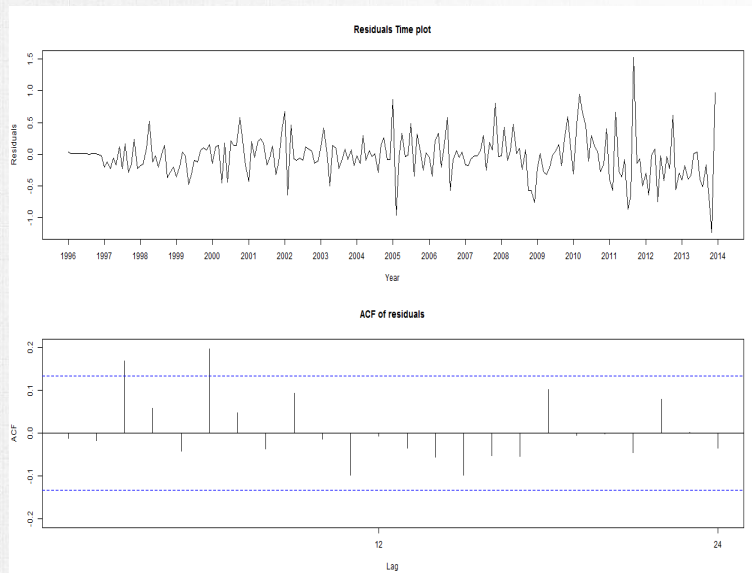
## Ανάλυση Χρονοσειράς



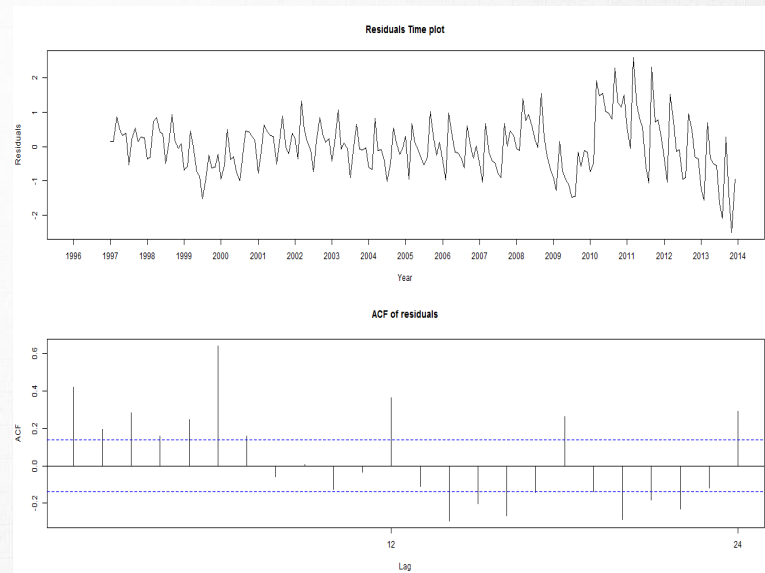
Αποτελέσματα

# CHARTS & ANALYSIS

Σύγκριση καταλοίπων και διαγραμμάτων ACF



ARIMA

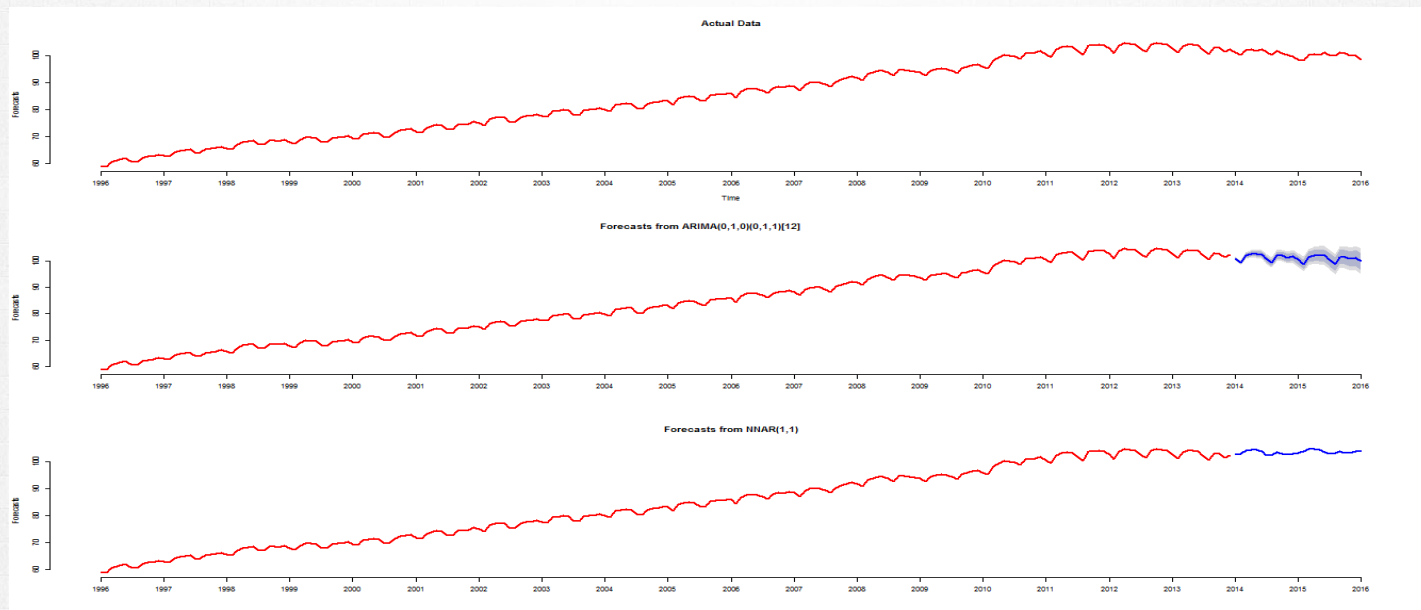


NNAR

Αποτελέσματα

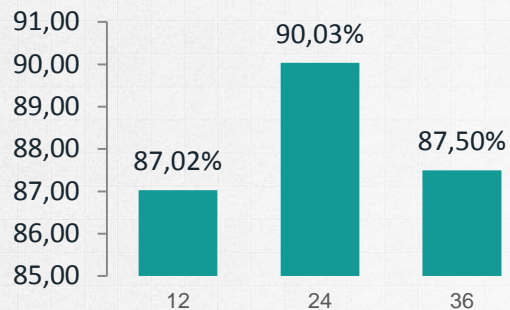
# CHARTS & ANALYSIS

Σύγκριση Πρόβλεψης ARIMA, NNAR με τις πραγματικές παρατηρήσεις





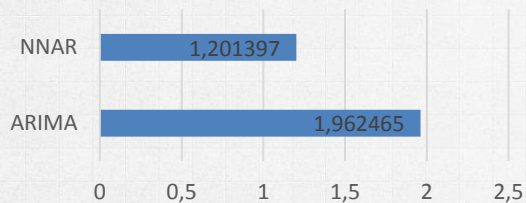
Box-Ljung Test  
(Διάγραμμα Ποσοστιαίας Διαφοράς ARIMA  
από NNAR προς Χρονική Υστέρηση)



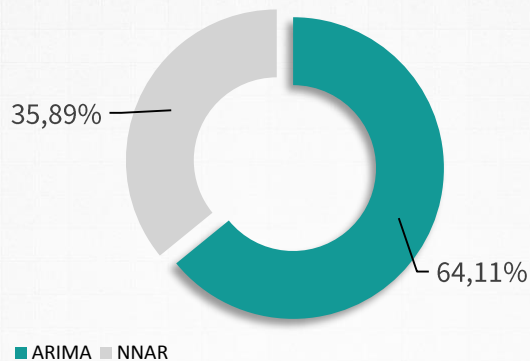
88%

Των καταλοίπων, του μοντέλου NNAR, παρουσιάζουν συσχέτιση με  $p\text{-value} < 0,00022$  ακόμα και μετά από διόρθωση σφαλμάτων

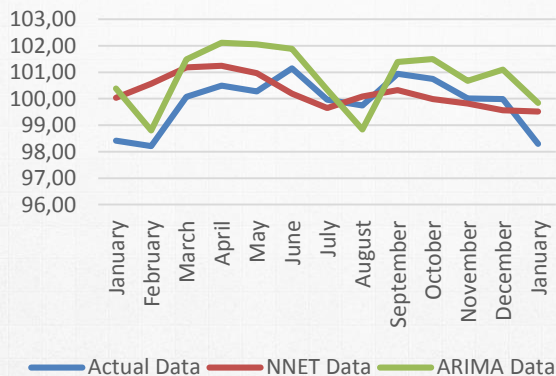
Durbin-Watson Test



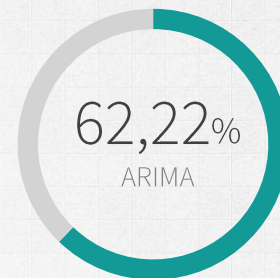
Αποδοτικότητα Πρόβλεψης  
κατά Μέσο Όρο



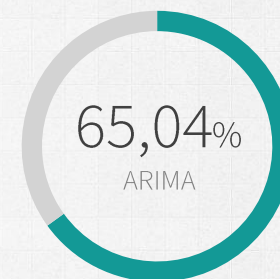
Σύγκριση έτους 2015 &  
Ιανουαρίου 2016



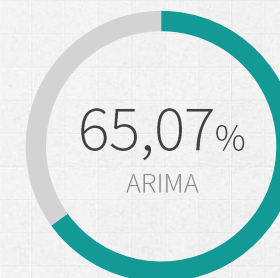
RMSE



MAPE



MASE



Ευχαριστούμε για το χρόνο σας