Στοχαστική Ανάλυση Χρονοσειρών

Περιεχόμενο:

α) Παρουσίαση Μαθήματος

β) Γνώριμες έννοιες από τις Πιθανότητες και τη Στατιστική

Επαμεινώνδας Διαμαντόπουλος
epdiaman@ee.duth.gr
Παρασκευή, 8 Νοεμβρίου 2024

Βασικοί Ορισμοί

Με τον όρο **χρονοσειρά** αναφερόμαστε σε μια ακολουθία τυχαίων μεταβλητών $\{X_t : t = 1, 2, ...\}$ η οποία εκφράζει την εξέλιξη της κατάστασης ενός συστήματος στο χρόνο.

Η μεταβλητή X_t δίνει την κατάσταση του συστήματος (αριθμητική τιμή) τη χρονική στιγμή t.

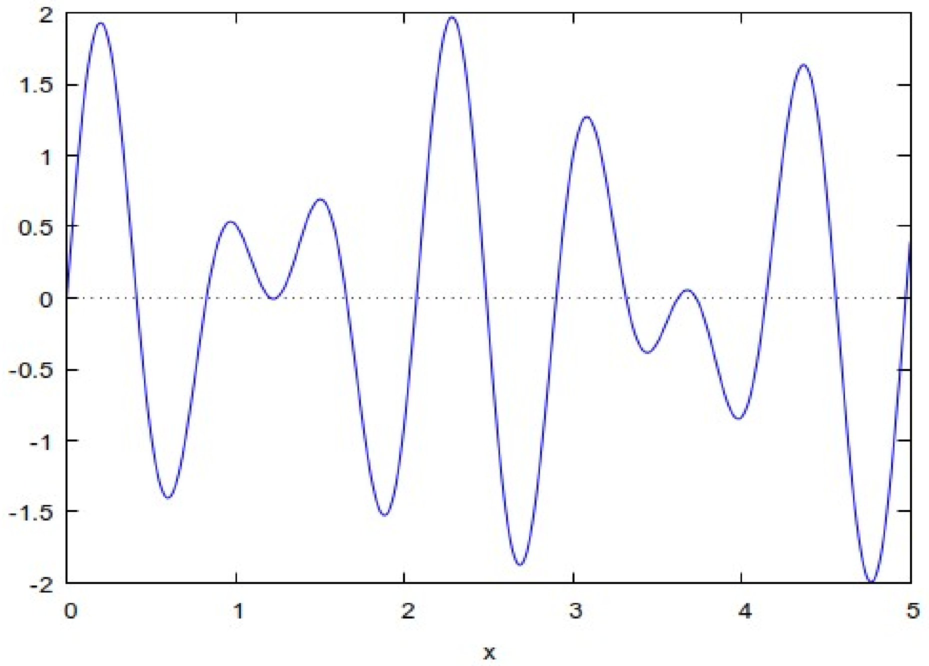
Οι χρονοσειρές διακρίνονται ως προς το είδος του χρόνου t και το είδος των δεδομένων x_t στις εξής κατηγορίες:

- Διακριτά μεγέθη x_t σε διακριτό χρόνο t.
- Συνεχή μεγέθη x_t σε διακριτό χρόνο t.
- Διακριτά μεγέθη x_t σε συνεχή χρόνο t.
- Συνεχή μεγέθη x_t σε συνεχή χρόνο t.

Οι χρονοσειρές διακρίνονται και ως προς τη διάσταση της μεταβλητής X_t σε μονομεταβλητές ή διανυσματικές χρονοσειρές.

Βασικοί Ορισμοί

Μία χρονοσειρά μπορεί να είναι **ντετερμινιστική** αν καθορίζεται από κάποιο γνωστό κανόνα, δηλαδή, αν οι τιμές της προκύπτουν ως αποτέλεσμα μίας γνωστής συνάρτησης όπως για παράδειγμα η $f(t) = \sin(2\pi t)+\sin(2\sqrt{2}\pi t)$



Βασικοί Ορισμοί

Τυχαία ή **μη ντετερμινιστική** χρονοσειρά είναι αυτή που δεν μπορεί να περιγραφεί με κάποια συγκεκριμένη συναρτησιακή έκφραση. Μια χρονοσειρά μπορεί να είναι μη ντετερμινιστική επειδή:

- Δεν είναι γνωστές όλες οι πηγές της μεταβλητότητας των τιμών της, όπως π.χ. το πλήθος των ατόμων που θα επισκεφτούν ένα αρχαιολογικό χώρο μία μέρα του Ιούνη.

- Η φύση της διαδικασίας που περιγράφεται με αυτήν είναι εγγενώς τυχαία, όπως π.χ. το πλήθος των σωματιδίων που καταφτάνουν σε έναν ανιχνευτή στραμμένο προς το διάστημα.

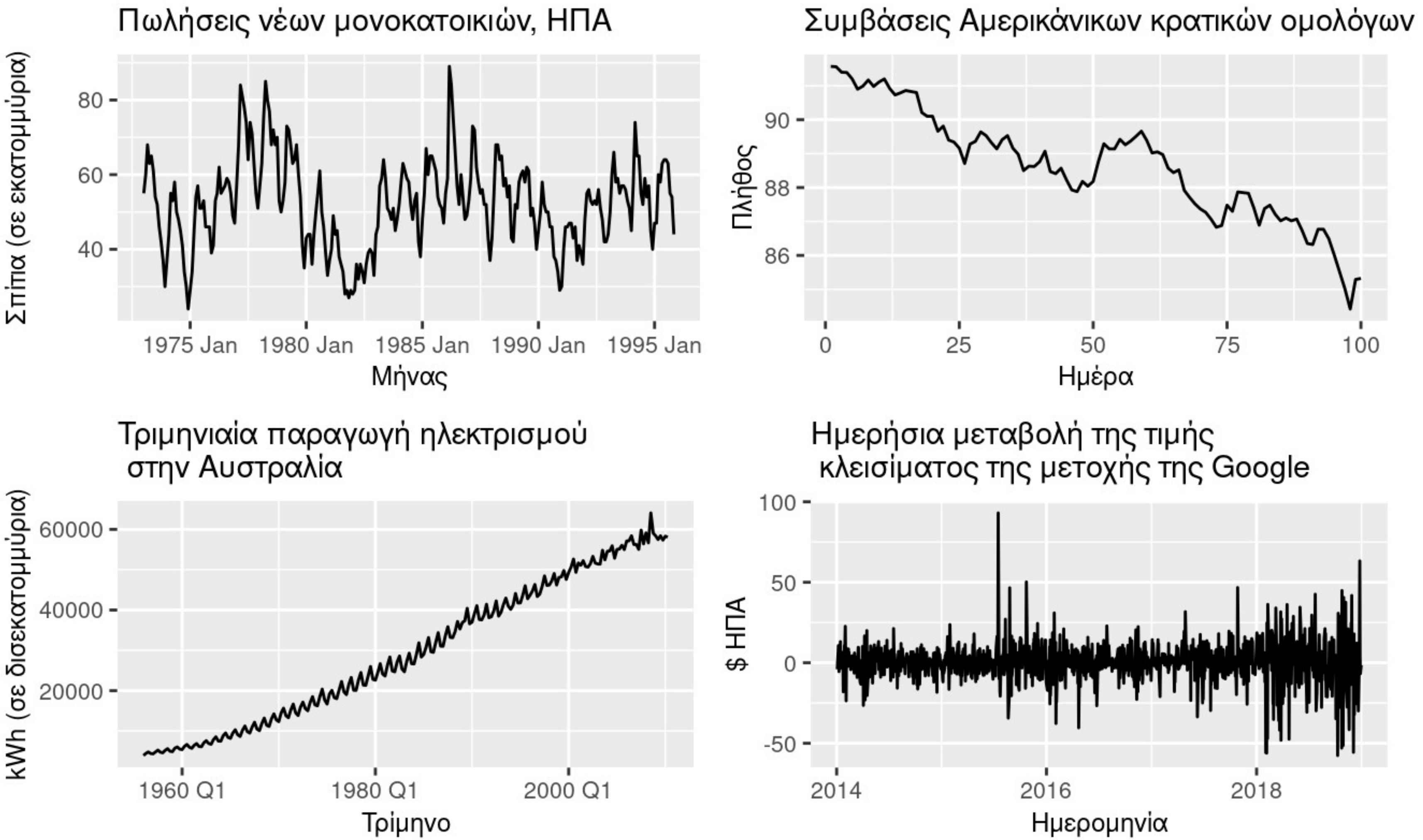
Βασικοί Ορισμοί

Καθώς οι μη ντετερμινιστικές χρονοσειρές εξελίσσονται με τυχαίο τρόπο, η περιγραφή της εξέλιξής τους αρμόζει να γίνει με όρους της θεωρίας πιθανοτήτων.

Στο παραπάνω πλαίσιο, θεωρούμε πως πίσω από κάθε τυχαία χρονοσειρά κρύβεται κάποια στοχαστική διεργασία $\{X_t, t = 1, 2, \ldots\}$.

Το σύνολο των πιθανών εξελίξεων μιας στοχαστικής διεργασίας είναι όλος ο στατιστικός πληθυσμός.

Μία παρατηρούμενη χρονοσειρά είναι μια δειγματική μονάδα.



Πηγή: https://otexts.com/fppgr/tspatterns.html

Σημαντικά ερωτήματα

Μερικά σημαντικά ερωτήματα που προκύπτουν κατά την παρατήρηση μιας χρονοσειράς είναι:

• Υπάρχει **τάση** στη σειρά, δηλαδή, οι μετρήσεις τείνουν κατά μέσο όρο να αυξάνονται (ή να μειώνονται) με την πάροδο του χρόνου;

• Υπάρχει **εποχικότητα**, δηλαδή υπάρχει ένα επαναλαμβανόμενο μοτίβο με υψηλές και χαμηλές τιμές που να σχετίζονται με την χρονική στιγμή;

• Υπάρχουν **ιδιάζουσες** ή **ακραίες** τιμές, δηλαδή τιμές που απέχουν ασυνήθιστα πολύ από τις υπόλοιπες παρατηρήσεις;

• Υπάρχει **μακροχρόνιος κύκλος** ή κάποια άλλη περιοδικότητα που δεν σχετίζεται με εποχιακούς παράγοντες;

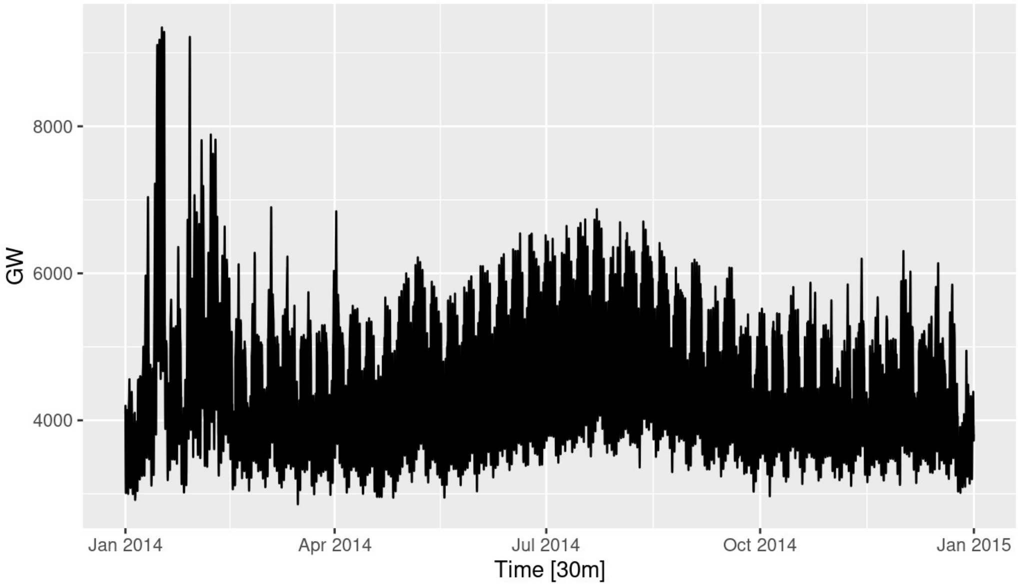
• Υπάρχει **μεταβολή στη διακύμανση** των τιμών με την πάροδο του χρόνου;

Στόχοι της ανάλυσης μίας χρονοσειράς

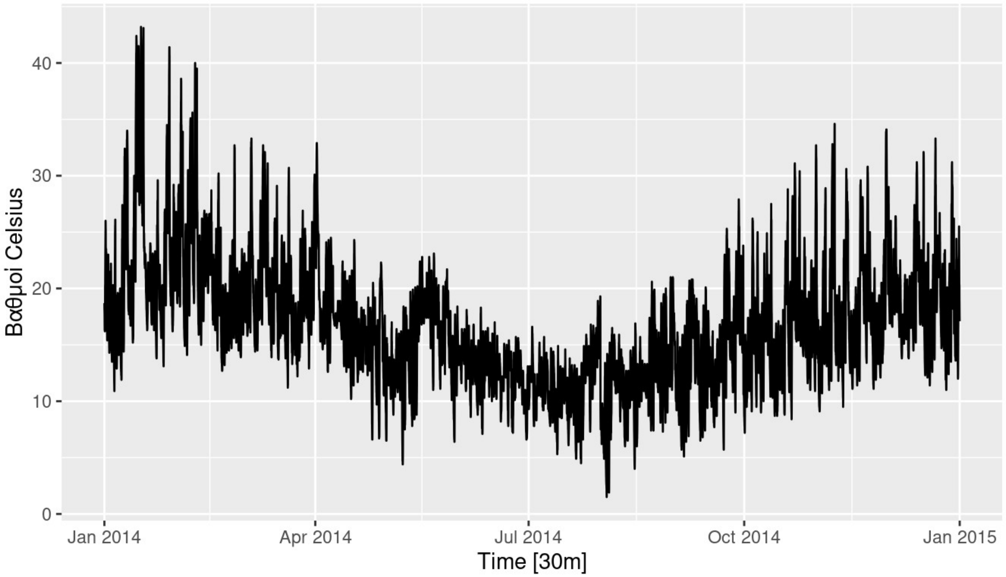
Ο βασικός στόχος της ανάλυσης είναι ο προσδιορισμός ενός μοντέλου που περιγράφει το μοτίβο που ακολουθεί η χρονοσειρά. Ένα τέτοιο μοντέλο, θα μπορούσε να χρησιμοποιηθεί για:

- Να περιγράψει τα σημαντικά χαρακτηριστικά της χρονοσειράς.
- Να εξηγήσει πώς το παρελθόν επηρεάζει το μέλλον.
- Να προβλέψει τις μελλοντικές τιμές της σειράς.

Ημιωριαία ζήτηση ηλεκτρισμού: Victoria Αυστραλίας



Ημιωριαία θερμοκρασία: Μελβούρνη Αυστραλίας



Πηγή: https://otexts.com/fppgr/scatterplots.html

Μοντέλο κινούμενου μέσου όρου (Moving Average)

Γέννεση μοντέλων Moving Average

Παράδειγμα Ι

Υποθέτουμε ότι κάθε χρονική περίοδο σκοπεύουμε να παράγουμε **m** μονάδες από κάποιο αγαθό. Ωστόσο, αναμένονται διάφορα i.i.d. τυχαία σφάλματα $\boldsymbol{\varepsilon_t}$ με $E(\varepsilon_t) = 0$, τα οποία θα αυξομειώνουν την παραγωγή. Άρα, η παραγωγή $\boldsymbol{x_t}$ τη χρονική στιγμή t είναι $\boldsymbol{x_t = m + \varepsilon_t}$. Έστω τώρα θ є (0, 1). Κάθε περίοδο σκοπεύουμε να διαθέτουμε το **1 – θ** μέρος της παραγωγής και να αποθηκεύουμε το μέρος **θ** για να το πουλήσουμε την επόμενη περίοδο. Θέλουμε να προσδιορίσουμε το απόθεμα $\boldsymbol{y_t}$ που υπάρχει στην αποθήκη τη στιγμή t.

Είναι **y_t = (παραγωγή της περιόδου t) + (το θ μέρος παραγωγής περιόδου t – 1).**

Γράφουμε: $y_t = x_t + \theta x_{t-1} = (m + \varepsilon_t) + \theta(m + \varepsilon_{t-1}) = (\theta + 1)m + \theta\varepsilon_{t-1} + \varepsilon_t$

Δηλαδή:

$$\boldsymbol{y_t = \mu + \theta\varepsilon_{t-1} + \varepsilon_t,}$$

όπου $\mu = (\theta + 1)m$.

Γέννεση μοντέλων Moving Average

Παράδειγμα ΙΙ

Έστω ότι ο ίδιος παραγωγός σκοπεύει κάθε περίοδο να αποθηκεύει μέρος θ_1 της παραγωγής για να το διαθέσει την επόμενη περίοδο και μέρος θ_2 για να το διαθέσει τη μεθεπόμενη περίοδο.

Στην περίπτωση αυτή, το απόθεμα y_t, είναι

y_t = (παραγωγή της περιόδου t) + (θ_1 μέρος t – 1 παραγωγής) + (θ_2 μέρος t – 2 παραγωγής).

Γράφουμε:

$$y_t = x_t + \theta_1 y_{t-1} + \theta_2 y_{t-1} = (m + \varepsilon_t) + \theta_1(m + \varepsilon_{t-1}) + \theta_2(m + \varepsilon_{t-2})$$

$$= (\theta_1 + \theta_2 + 1)m + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \varepsilon_t$$

Δηλαδή:

$$\mathbf{y_t = \mu + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \varepsilon_t,}$$

όπου $\mu = (\theta_1 + \theta_2 + 1)m$.

Γέννεση μοντέλων Moving Average

Το υποθετικό σενάριο των παραδειγμάτων, εμφανίζεται συχνά σε πραγματικά συστήματα, στα οποία υπάρχουν άγνωστες πηγές μεταβλητότητας που επηρεάζουν τη χρονοσειρά.

Παράδειγμα ΙΙΙ

Έστω X_t οι πωλήσεις ενός προϊόντος τη χρονική στιγμή t. Η τιμή του X_t, επηρεάζεται από τις εποχιακές προωθητικές ενέργειες που ποικίλουν σε ένταση και επηρεάζουν την τιμή για ένα πλήθος μελλοντικών στιγμών ενώ ενδέχεται να επικαλύπτονται με άλλες παράλληλες προσφορές. Μία τέτοια χρονοσειρά είναι πιθανό να περιγράφεται ικανοποιητικά από ένα μοντέλο MA(q).

Παράδειγμα IV

Έστω η σειρά X_t της τιμής κλεισίματος μιας μετοχής. Η τιμή κλεισίματος κάθε ημέρας προκύπτει από μια τάση (π.χ. γραμμική) συν τις σταθμισμένες επιπτώσεις των ημερήσιων κραδασμών από προηγούμενες ημέρες. Πιθανώς, η επίδραση του κραδασμού την ημέρα t – 1 θα έχει ισχυρότερη επιρροή στην τιμή την στιγμή t από ότι ο κραδασμός την ημέρα t – 2, κ.λπ. Επομένως, η τιμή κλεισίματος της μετοχής την ημέρα t θα αντανακλά το σταθμισμένο άθροισμα των προηγούμενων κραδασμών μέχρι την προηγούμενη ημέρα t – 1, διαδικασία που ταιριάζει σε ένα MA(q) μοντέλο.

Μοντέλο κινούμενου μέσου όρου (Moving Average)

Έστω $\{X_t, t \geq 1\}$ μία χρονοσειρά και $\{w_t\}$, $t \geq 1$, μία οικογένεια ισόνομων και ανεξάρτητων τυχαίων μεταβλητών που ακολουθούν την $N(0, \sigma^2)$ κατανομή (δηλαδή WN).

Το μοντέλο κινητού μέσου 1ης τάξης, που συμβολίζεται με MA(1) είναι:

$$X_t = \mu + w_t + \theta_1 \cdot w_{t-1}.$$

Το μοντέλο κινητού μέσου 2ης τάξης, που συμβολίζεται με MA(2) είναι:

$$X_t = \mu + w_t + \theta_1 \cdot w_{t-1} + \theta_2 \cdot w_{t-2}.$$

Το μοντέλο κινητού μέσου όρου q^{th} τάξης, που συμβολίζεται με MA(q) είναι:

$$X_t = \mu + w_t + \theta_1 \cdot w_{t-1} + \theta_2 \cdot w_{t-2} + \dots + \theta_q \cdot w_{t-q}.$$

Μοντέλο κινούμενου μέσου όρου (Moving Average)

Έστω B ο τελεστής μετατόπισης της ακολουθίας μία θέση προς τα πίσω (ή τελεστής υστέρησης ή τελεστής ολίσθησης, backward shift) που στέλνει κάθε στοιχείο της ακολουθίας στο αμέσως προηγούμενό του:

$$\mathbf{B(w_n) = w_{n-1}}$$

Φανερά $\mathbf{B^K(w_n) = w_{n-\kappa}}$. Με χρήση του τελεστή υστέρησης B, η γενική εξίσωση ενός MA(q) μοντέλου

$$X_t = \mu + w_t + \theta_1 \cdot w_{t-1} + \theta_2 \cdot w_{t-2} + \ldots + \theta_q \cdot w_{t-q}.$$

εκφράζεται ως εξής:

$$X_t = \mu + \Theta(B)w_t,$$

όπου $\mathbf{\Theta(\omega) = 1 + \theta_1\omega + \theta_2\omega^2 + \ldots + \theta_q\omega^q}$, το χαρακτηριστικό πολυώνυμο της MA(q).

Αυτοπαλινδρομικό μοντέλο AR(p)

Γέννεση Αυτοπαλινδρομικών Μοντέλων

Οι τιμές μετοχών συχνά παρουσιάζουν εξάρτηση από προηγούμενες τιμές, λόγω της τάσης των επενδυτών να λαμβάνουν αποφάσεις βασισμένες στην πρόσφατη συμπεριφορά της αγοράς. Αυτό σημαίνει πως X_t είναι η τιμή της μετοχής τη στιγμή t, ένα μοντέλο που αναμένεται να προσαρμόζεται καλά στα δεδομένα είναι το αυτοπαλινδρομικό 1ης τάξης:

$$X_{t+1} = \alpha X_t + \beta.$$

Αντίστοιχες παρατηρήσεις μπορούν να γίνουν:

- Για τη θερμοκρασία μιας ημέρας η οποία εξαρτάται σε κάποιο βαθμό από τις προηγούμενες ημέρες λόγω της αδράνειας του κλίματος.
- Για την κατανάλωση ενέργειας (σε καθημερινή ή ωριαία βάση) η οποία συχνά εξαρτάται από τις τιμές των προηγούμενων χρονικών διαστημάτων, καθώς οι συνήθειες κατανάλωσης είναι συνήθως σχετικά σταθερές και επηρεάζονται από το παρελθόν.
- Για τη σεισμική δραστηριότητα όπου μια δόνηση αυξάνει την πιθανότητα περαιτέρω δονήσεων σε σύντομο χρονικό διάστημα.

Γενική εξίσωση AR(p)

Η γενική εξίσωση ενός AR(p) μοντέλου είναι

$$X_t = \delta + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim IWN(0, \sigma^2).$$

Γενική εξίσωση AR(p)

Έστω B ο τελεστής μετατόπισης της ακολουθίας μία θέση προς τα πίσω (ή τελεστής υστέρησης ή τελεστής ολίσθησης, backward shift) που στέλνει κάθε στοιχείο της ακολουθίας στο αμέσως προηγούμενό του:

$$\mathbf{B(X_n) = X_{n-1}}$$

Φανερά $\mathbf{B^{\kappa}(X_n) = X_{n-\kappa}}$. Με χρήση του τελεστή υστέρησης B, η γενική εξίσωση ενός AR(p) μοντέλου

$$\mathbf{X_t = \delta + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim IWN(0, \sigma^2),}$$

εκφράζεται ως εξής:

$$(1 - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_p B^p)X_t = \delta + \varepsilon_t,$$

ή ακόμα και ως

$$\mathbf{\Phi(B)X_t = \delta + \varepsilon_t,}$$

όπου $\mathbf{\Phi(\omega) = 1 - \alpha_1\omega - \alpha_2\omega^2 - \ldots - \alpha_p\omega^p}$, το **χαρακτηριστικό πολυώνυμο** της AR(p).

Μοντέλα ARMA(p, q)

Μοντέλα ARMA(p, q)

Τα μοντέλα χρονοσειρών που είναι γνωστά ως μοντέλα ARMA μπορεί να περιλαμβάνουν αυτοπαλινδρομικούς όρους και όρους κινητού μέσου όρου.

Τα στοιχεία στο μοντέλο καθορίζονται με τη σειρά (AR, MA).

Ένα μοντέλο με δύο όρους AR μόνο, θα προσδιορίζεται ως ARMA τάξης (2,0).

Ένα μοντέλο MA(2) θα προσδιορίζεται ως ARMA τάξης (0,2).

Μοντέλα ARMA(p, q)

Η γενική εξίσωση ενός ARMA(p, q) μοντέλου είναι ($\varepsilon_t \sim IWN(0, \sigma^2)$):

$$X_t = \delta + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q},$$

Με χρήση των τελεστών Β, Θ, η τελευταία σχέση εκφράζεται ως εξής:

$$\mathbf{\Phi(B)X_t = \delta + \Theta(B)\varepsilon_t}$$

όπου $\mathbf{\Phi(\omega) = 1 - \alpha_1\omega - \alpha_2\omega^2 - \ldots - \alpha_p\omega^p}$, το χαρακτηριστικό πολυώνυμο της AR(p),

και $\mathbf{\Theta(\omega) = 1 + \theta_1\omega + \theta_2\omega^2 + \ldots + \theta_q\omega^q}$, το χαρακτηριστικό πολυώνυμο της MA(q).

Η αναγκαιότητα της στασιμότητας

Τα ARMA(p, q) μοντέλα απαιτούν στασιμότητα.

Αν η σειρά δεν είναι στάσιμη, τότε οι συντελεστές του μοντέλου δεν υπολογίζονται με αξιοπιστία.

Μία πρακτική για την αφαίρεση της τάσης είναι η αντικατάσταση της χρονοσειράς με τις διαφορές διαδοχικές όρων.

Η διαδικασία αυτή ονομάζεται ολοκλήρωση (integration) της χρονοσειράς.

Τα μοντέλα που προκύπτουν συμβολίζονται με **ARIMA(p, d, q).**

Η μοντελοποίηση με ARIMA μίας χρονοσειράς δεν έχει προϋποθέσεις σχετικά με την κατανομή των τιμών και για το λόγο αυτό είναι καθολική. Είναι σαν προσαρμογή πολυωνυμικής καμπύλης - δεν ενδιαφέρει ποια είναι η αληθινή συνάρτηση, αυτό που έχει σημασία είναι η προσέγγισή της με ένα πολυώνυμο κάποιου βαθμού.

Βήματα ανάλυσης χρονοσειράς

1ο Στάδιο: Ταυτοποίηση.

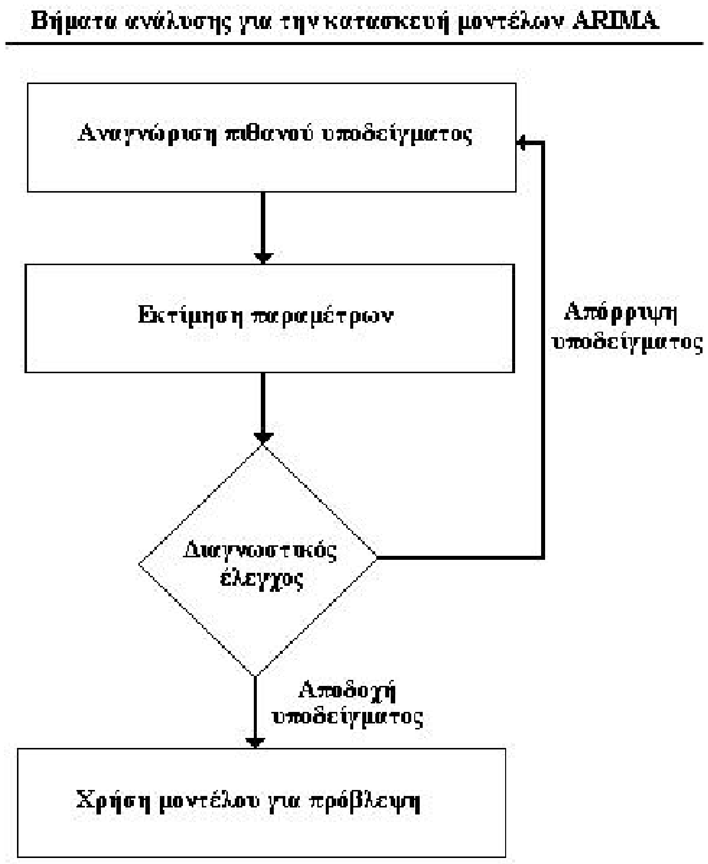
Σε αυτό γίνεται μια προσπάθεια να αναγνωριστούν τα πιθανά υποδείγματα που ταιριάζουν με τα δεδομένα. Η αναγνώριση γίνεται μέσα από τα διαγράμματα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης και με σύγκριση τυπικών υποδειγμάτων.

2ο Στάδιο: Εκτίμηση και δοκιμή.

Στο στάδιο αυτό γίνεται η εκτίμηση των παραμέτρων του υποδείγματος που αναγνωρίστηκε. Στη συνέχεια γίνεται η δοκιμή του υποδείγματος που επιλέχθηκε, ώστε να αναγνωριστούν οι συντελεστές που είναι στατιστικά σημαντικοί και να αφαιρεθούν οι υπόλοιποι. Τέλος γίνεται έλεγχος της ικανότητας πρόβλεψης και γενικά της αποδοχής ή μη του υποδείγματος. Σε περίπτωση μη αποδοχής του υποδείγματος, επαναλαμβάνεται η διαδικασία για να επιλεγεί ένα πιο κατάλληλο υπόδειγμα.

3ο Στάδιο: Χρήση.

Σε περίπτωση που το υπόδειγμα γίνει αποδεκτό ακολουθεί το τρίτο στάδιο, δηλαδή η χρήση του μοντέλου για πρόβλεψη.



Μοντέλα Κατάστασης – Χώρου

Πέρα από τα ARMA: Μοντέλα Κατάστασης – Χώρου

Έστω X_t η τιμή μίας χρονοσειράς τη χρονική στιγμή t, συνάρτηση των $\theta_t = (\theta^1_t, \theta^2_t, ..., \theta^d_t)'$.

$$\mathbf{X_t = h\,\theta_t + \zeta_t = h_1\,\theta^1_t + h_2\,\theta^2_t + ... + h_d\,\theta^d_t + \zeta_t,\ \zeta_t \sim N(0, \sigma_\zeta^2)}$$

Η εξίσωση αυτή είναι γνωστή ως η **εξίσωση μέτρησης** ή **εξίσωση παρατήρησης**.

Στην εξίσωση $\mathbf{X_t = h\,\theta_t + \zeta_t}$, η χρονοσειρά X_t είναι αυτή που παρατηρείται, ενώ οι θ_t γνωρίζουμε ότι υπάρχουν χωρίς όμως να παρατηρούνται. Ωστόσο, μπορούμε να καταγράψουμε τις εξισώσεις που περιγράφουν τον τρόπο που εξελίσσονται οι θ_t στο χρόνο:

$$\mathbf{\theta_t = G_t\,\theta_{t-1} + w_t,\ t > 2.\ w_t \sim N(0, \Sigma_w)}$$

Η τελευταία εξίσωση είναι γνωστή ως η **εξίσωση μετάβασης**. Οι εξισώσεις

$$\mathbf{X_t = h\,\theta_t + \zeta_t}$$

$$\mathbf{\theta_t = G_t\,\theta_{t-1} + w_t,}$$

αποτελούν ένα γενικό μοντέλο κατάστασης - χώρου (state - space model). Στην παραπάνω προσέγγιση το space είναι ο χώρος με διαστάσεις τις d μεταβλητές $(\theta_1^t, \theta_2^t, ..., \theta^d_t)$, ενώ το state είναι το διάνυσμα $h = (h_1, h_2, ...,h_d)$ που ορίζει την τιμή της χρονοσειράς X_t.

Πέρα από τα ARMA: Μοντέλα Κατάστασης – Χώρου

Τα μοντέλα χώρου κατάστασης διαφέρουν από τα μοντέλα ARIMA ως προς τη δομή, την ευελιξία και τη δυνατότητα εφαρμογής τους.

Ειδικότερα, τα μοντέλα χώρου κατάστασης είναι πιο ευέλικτα και φιλοξενούν τόσο μονομεταβλητά όσο και πολυμεταβλητά δεδομένα με υποκείμενες λανθάνουσες διαδικασίες που εξελίσσονται με την πάροδο του χρόνου.

Επιπλέον, σε αντίθεση με τα μοντέλα ARIMA δεν προϋποθέτουν τη στασιμότητα και είναι κατάλληλα για πρόβλεψη όταν τα δεδομένα έχουν χρονικές εξαρτήσεις χωρίς σημαντικές εξωτερικές επιρροές.

Πέρα από τα ARMA: Μοντέλα Κατάστασης – Χώρου

Παράδειγμα

Για να μετρήσουμε την εξέλιξη της στάθμης ενός ποταμού, πηγαίνουμε κάθε χρόνο σε ένα συγκεκριμένο σημείο και κάνουμε τη μέτρηση. Όμως, η μέτρησή επηρεάζεται από τυχαίους παράγοντες που δεν γνωρίζουμε, όπως για παράδειγμα μία πιθανή βροχή που προηγήθηκε ή ένα σφάλμα του εργαλείου μέτρησης. Οπότε, μετρούμε τη στάθμη του νερού με μια πρόσθετη, μη ελεγχόμενη και τυχαία ανακρίβεια. Η κατάσταση αυτή αντιπροσωπεύεται στην εξίσωση:

$$\textbf{Μέτρηση Στάθμης}_t = \textbf{Πραγματική Στάθμη}_t + \varepsilon_t,$$

όπου ε_t το σφάλμα της μέτρησης.

Επιπλέον, είναι λογικό να υποθέσουμε ότι η πραγματική στάθμη του νερού του ποταμού αλλάζει με την πάροδο του χρόνου όπως για παράδειγμα στην περίπτωση όπου χτιστεί ένα φράγμα. Δηλαδή, η τρέχουσα πραγματική στάθμη εξαρτάται από την στάθμη του προηγούμενου έτους συν την τυχαία μεταβολή της τελευταίας περιόδου. Συνεπώς, είναι λογικό να θεωρήσουμε την ακόλουθη εξίσωση:

$$\textbf{Πραγματική Στάθμη}_t = \textbf{Πραγματική Στάθμη}_{t-1} + \eta_t$$

όπου η_t το σφάλμα της τιμής.

Πέρα από τα ARMA: Μοντέλα Κατάστασης – Χώρου

Παράδειγμα 1

Οι δύο εξισώσεις

$$\textbf{Μέτρηση Στάθμης}_t = \textbf{Πραγματική Στάθμη}_t + \varepsilon_t$$

$$\textbf{Πραγματική Στάθμη}_t = \textbf{Πραγματική Στάθμη}_{t-1} + \eta_t$$

αποτελούν μία περίπτωση State Space μοντέλου (η οποία είναι γνωστή ως Local Level μοντέλο).

Πέρα από τα ARMA: Μοντέλα Στοχαστικών Συνιστωσών

Έστω X_t η τιμή μίας χρονοσειράς τη χρονική στιγμή t. Οι μέθοδοι αποσύνθεσης επιδιώκουν την ανάλυση της X_t σε μέρη

$$X_t = \mu_t + \gamma_t + \omega_t + \varepsilon_t,$$

όπου: - μ_t η τάση, - γ_t η περιοδική συνιστώσα,

- ω_t η κυκλική συνιστώσα, - ε_t το τυχαίο σφάλμα.

Η αποσύνθεση μας ενημερώνει για τη δομή μίας χρονοσειράς. Το επόμενο βήμα είναι η πρόβλεψη των τιμών της σειράς με την προσαρμογή ενός κατάλληλου μοντέλου. Το μοντέλο

$$\mathbf{X_t = \mu_t + \gamma_t + \omega_t + \Sigma\varphi_i X_{t-i} + \Sigma\beta_j Y_{jt} + \varepsilon_t.}$$

αντιπροσωπεύει τη γενική περίπτωση ενός μοντέλου πρόβλεψης με στοχαστικές συνιστώσες (Unobserved Components Model ή Structural Time Series Model).

State – Space Time Series Models

$$X_t = h'\theta_t + \zeta_t$$
$$\theta_t = G_i\,\theta_{t-i} + w_t$$

Local Level

$$X_t = \mu_t + \varepsilon_t$$
$$\mu_t = \mu_{t-1} + \eta_t$$

Local Level with (stochastic) trend

$$X_t = \mu_t + \varepsilon_t$$
$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t$$
$$\beta_t = \beta_{t-1} + \zeta_{t.}$$

Unobserved Components Models - UCM

$$X_t = \mu_t + \gamma_t + w_t + \Sigma_i\,\varphi_i\,X_{t-i} + \Sigma_j\,\beta_j\,Y_{jt} + \varepsilon_t$$

Random Walk

$$X_t = X_{t-1} + \varepsilon_t$$

Random Walk with drift

$$X_t = X_{t-1} + \mu + \varepsilon_t$$

Locally Linear Trend

$$X_t = X_{t-1} + \mu_{t-1} + \varepsilon_t$$
$$\mu_t = \mu_{t-1} + \zeta_t$$

ARIMA(p, d, q)

$$\varphi(B)\Delta^d\,X_t = \theta(B)\varepsilon_t$$

SARIMA(p, d, q)(P, D, Q)$_m$

$$\varphi(B)\Phi(B^m)\Delta^d\,\Delta^D_m X_t = \theta(B)\Theta(B^m)\varepsilon_t$$

Διαδικασία Μαθήματος

Το μάθημα θα διδαχθεί:

Οι πρώτες 7 διαλέξεις από το μέλος Ε.ΔΙ.Π. Επαμεινώνδα Διαμαντόπουλο.

Οι επόμενες 6 διαλέξεις από τον Καθηγητή Αλέξανδρο Ρήγα.

Διαδικασία Μαθήματος

Ο (ενδεικτικός) προγραμματισμός για τις 7 πρώτες διαλέξεις είναι ο εξής:

Διάλεξη 1 (08/11/24): Περίγραμμα του μαθήματος. Εισαγωγικές Έννοιες.

Διάλεξη 2 (15/11/24): Διαγνωστικά Εργαλεία (ACF, PACF) – Η έννοια της Στασιμότητας – Έλεγχος Στασιμότητας – Δοκιμασίες Dickey-Fuller και Kwiatkowski–Phillips–Schmidt–Shin (KPSS) – Μετατροπή σειράς σε Στάσιμη.

Διάλεξη 3 (22/11/24): Λευκός Θόρυβος – Τυχαίος Περίπατος – Υπόδειγμα ARIMA.

Διάλεξη 4 (25/11/24): Η περίπτωση της εποχικότητας. Υπόδειγμα Seasonal ARIMA.

Διάλεξη 5 (13/12/24): Τοπικό Μοντέλο (Local Level Model)

Διάλεξη 6 (20/12/24): Αποσύνθεση χρονοσειράς - Μοντέλα Στοχαστικών Συνιστωσών.

Διάλεξη 7 (10/01/25): Παρουσίαση Εργασιών.

Διαδικασία Μαθήματος

Το μάθημα θα υλοποιηθεί με χρήση υπολογιστή.

Για την ανάλυση δεδομένων χρονοσειρών θα χρησιμοποιήσουμε τη γλώσσα R και το R Studio.

Για κάθε μία διαδικασία θα υποδεικνύεται ο κώδικας που την υλοποιεί και θα γίνεται ανάγνωση στο output που προκύπτει.

Θα δοθούν 2 εργασίες οι οποίες θα παρουσιαστούν από τους φοιτητές στην τελευταία συνάντηση (10/01/25).

Η συνέχεια εδώ:

https://utopia.duth.gr/epdiaman/files/dpth/IntroR.html

Έννοιες Στατιστικής και Πιθανοτήτων που Χρησιμοποιούμε στην Ανάλυση Χρονοσειρών

ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ
DEMOCRITUS UNIVERSITY OF THRACE

Διάστημα Εμπιστοσύνης

Διάστημα Εμπιστοσύνης

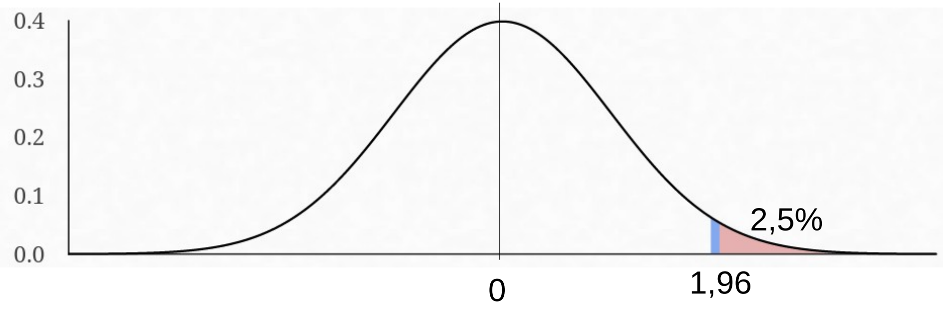
Έστω $Z \sim N(0, 1)$, $0 < \alpha < 1$, και z_α ο αριθμός για τον οποίο

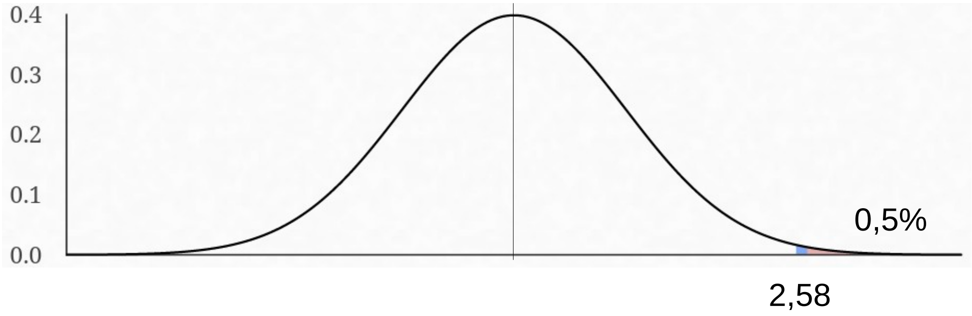
$$\mathbf{P(Z > z_\alpha) = \alpha \text{ ή } \Phi(z_\alpha) = 1 - \alpha.}$$

Το z_α ονομάζεται α – ποσοστιαίο σημείο της κανονικής κατανομής. Ενδεικτικά:

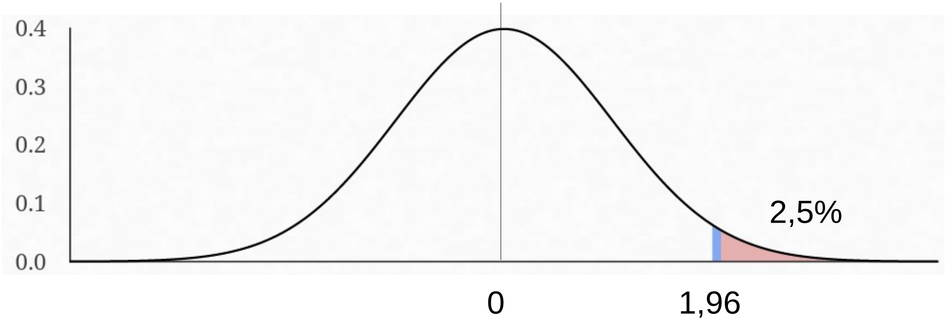
Αν $\alpha = 0{,}025$, τότε $z_{0,025}$ είναι ο αριθμός για τον οποίον $\Phi(z_{0,025}) = 0{,}975$ ή $\mathbf{z_{0,025} = 1{,}96}$.

Αν $\alpha = 0{,}005$, τότε $z_{0,005}$ είναι ο αριθμός για τον οποίον $\Phi(z_{0,005}) = 0{,}995$ ή $\mathbf{z_{0,005} = 2{,}58}$.

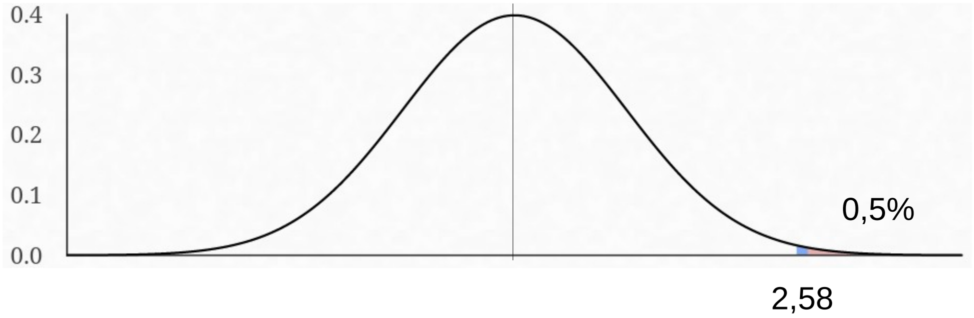




Διάστημα Εμπιστοσύνης



$$Z \sim N(0\,,\,1) \Leftrightarrow P(-1.96 < Z < 1.96) = 0.95\,.$$



$$Z \sim N(0\,,\,1) \Leftrightarrow P(-2.58 < Z < 2.58) = 0.99\,.$$

Στατιστικοί Έλεγχοι Υποθέσεων

ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ | DEMOCRITUS UNIVERSITY OF THRACE

Στατιστικοί Έλεγχοι Υποθέσεων

Στην Επαγωγική Στατιστική προσπαθούμε να ελέγξουμε μία υπόθεση που αφορά τον πληθυσμό από ένα δείγμα. Αυτό, το πετυχαίνουμε σε 4 βήματα:

(α) Προσδιορίζουμε μία στατιστική υπόθεση H_0 που εκφράζει ένα γεγονός για κάποιο στατιστικό μέγεθος (π.χ. X, Y ανεξάρτητες, $\mu = 10$, $\sigma_1 = \sigma_2$, $X_t \sim \text{iid } N(0, 1)$, κλπ)

(β) Υπολογίζουμε ένα στατιστικό $\boldsymbol{\Sigma_\Delta}$ στο οποίο αντανακλάται η διαφοροποίηση του δείγματος από την υπόθεση H_0 *και για το οποίο γνωρίζουμε την θεωρητική κατανομή των τιμών του, δεδομένης της H_0.* (π.χ. χ^2, $t = (x - 10)/SE$, $F = \sigma_1^2 / \sigma_2^2$, Q_m, κλπ)

(γ) Υπολογίζουμε την πιθανότητα **p** το στατιστικό να πάρει περισσότερο ακραία τιμή από την τιμή που υπολογίστηκε. Συνήθως το όριο απόρριψης είναι το 0,05.

(δ) Αν $p < 0{,}05$ τότε λέμε ότι η H_0 απορρίπτεται και αποδεχόμαστε τη συμπληρωματική της πρόταση (H_1).

(ε) Ακολουθεί η καταγραφή του αποτελέσματος.

Στατιστικοί Έλεγχοι Υποθέσεων

Ενδεικτικά παραδείγματα

Δοκιμασία ομοιογένειας χ^2

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(k-1).$$

Δοκιμασία one sample t – test

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t(n-1)$$

Δοκιμασία ανεξαρτησίας χ^2

$$\chi^2 = \sum_{i=1}^{k} \sum_{i=1}^{m} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2(k-1)(m-1).$$

Δοκιμασία independent samples t – test

$$\frac{\bar{X}_1 - \bar{X}_2}{SE_p} \sim t(n_1 + n_2 - 2)$$

Δοκιμασία Box-Pierce[1] $Q(m) = n(n+2)\sum_{j=1}^{m} \frac{r_j^2}{n-j} \sim \chi^2(m)$ ή $\chi^2(m-p-q-1)$

(1) Σχετική αναφορά: https://online.stat.psu.edu/stat510/lesson/3/3.2

Στατιστικοί Έλεγχοι Υποθέσεων

Καταγραφή μίας στατιστικής δοκιμασίας

Αναφέρουμε την τιμή του στατιστικού, τους βαθμούς ελευθερίας και την τιμή της στατιστικής σημαντικότητας p. Ακολουθεί η ερμηνεία του αποτελέσματος.

Παραδείγματα

"Η δοκιμασία χ^2 κατέδειξε πως το ποσοστό των μαθητών που κάνουν δίαιτα διαφέρει σημαντικά ανάμεσα στα δύο φύλα ($\chi^2(1) = 8,4$, $p = 0,003$)."

"Ο στατιστικός έλεγχος t - test για δύο ανεξάρτητα δείγματα κατέδειξε πως οι μέσες αγορές της ομάδας A (M = 1.704,1, SD = 341,3), δεν είχαν στατιστικά διαφορετική μέση τιμή από τις αγορές της ομάδας B (M = 1.578,4, SD = 441,8), $t(19) = 0,734$, $p = 0,472$, (δίπλευρος έλεγχος)."

" Για την ανίχνευση της διαφοροποίησης των μέσων τιμών των πωλήσεων μεταξύ των τριών ανεξάρτητων ομάδων χρησιμοποιήθηκε η ανάλυση διακύμανσης με έναν παράγοντα. Βρέθηκε πως υπάρχει σημαντική διαφοροποίηση των μέσων τιμών των πωλήσεων μεταξύ των τριών διαφορετικών ομάδων, $F(2, 28) = 7,512$, $p = 0,002$."

Συντελεστής Συσχέτισης Pearson

Συντελεστής Συσχέτισης Pearson

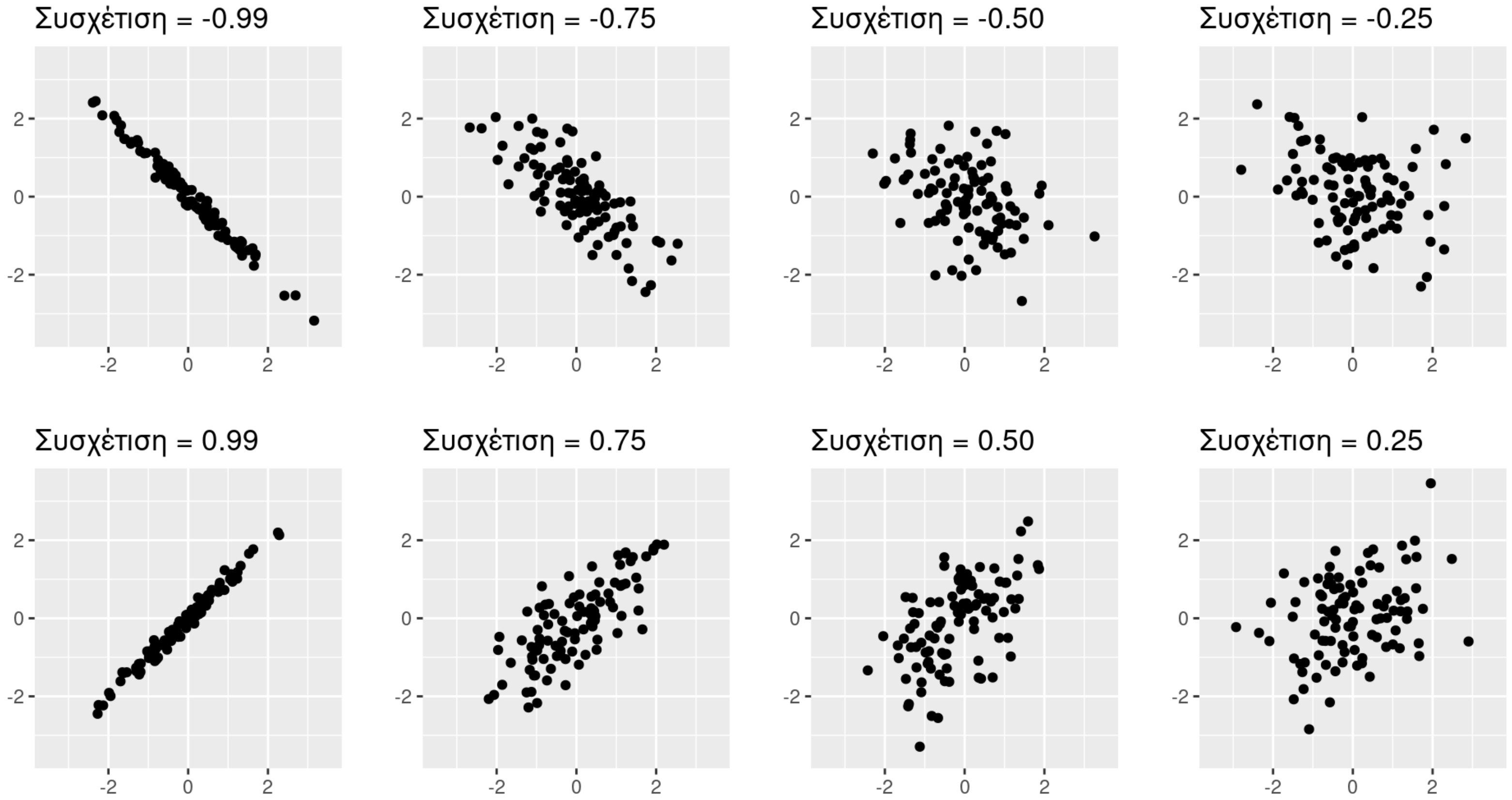
Η συνδιακύμανση Cov(X, Y) είναι μία στατιστική ποσότητα που μας επιτρέπει να ποσοτικοποιήσουμε τον τρόπο με τον οποίο συμμεταβάλλονται δύο τυχαίες μεταβλητές X, Y. Η κανονικοποιημένη εκδοχή της συνδιακύμανσης είναι ο συντελεστής συσχέτισης Pearson $\rho_{X,Y}$, ο οποίος παίρνει τιμές από -1 έως +1.

$$\rho_{X,Y} = \mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\mathrm{E}[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

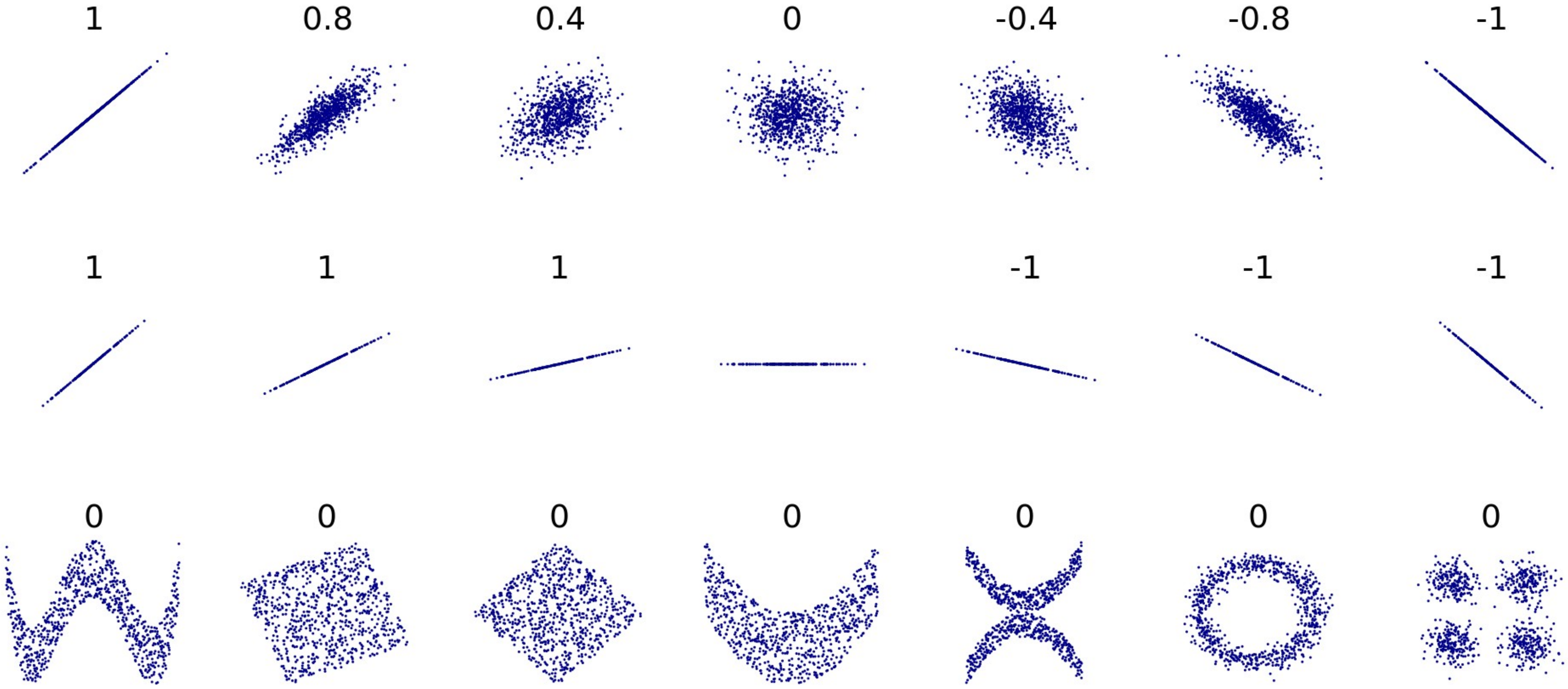
Δειγματικός Συντελεστής Συσχέτισης

$$r_{xy} \stackrel{\text{def}}{=} \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}},$$

Συντελεστής Συσχέτισης Pearson



Συντελεστής Συσχέτισης Pearson



ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ
DEMOCRITUS UNIVERSITY OF THRACE

Ευχαριστώ για την προσοχή σας!