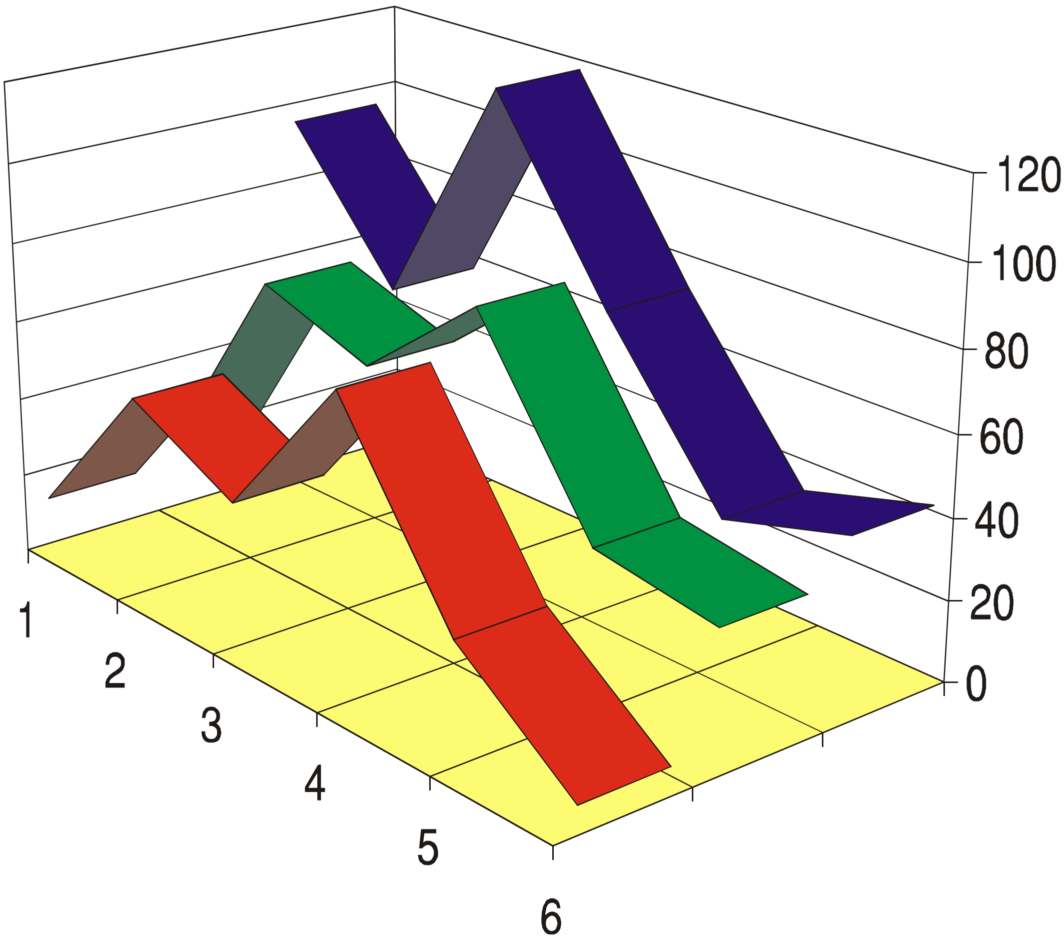
ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ
ΤΜΗΜΑ ΑΓΡΟΤΙΚΗΣ ΑΝΑΠΤΥΞΗΣ
Θεόδωρος Χ. Κουτρουμανίδης
Αναπληρωτής Καθηγητής ΔΠΘ

ΕΦΑΡΜΟΣΜΕΝΗ ΟΙΚΟΝΟΜΙΚΗ ΣΤΑΤΙΣΤΙΚΗ



Ορεστιάδα 2007

ΠΕΡΙΕΧΟΜΕΝΑ

Π ρ ό λ ο γ ο ς

Σκοπός του παρόντος βοηθήματος είναι να δώσει στον φοιτητή του τμήματος της Αγροτικής Ανάπτυξης, με όσο το δυνατόν μεγαλύτερη απλότητα και πιο ουσιαστικά, τις βασικές γνώσεις της Στατιστικής και να παρουσιάσει κύριους τομείς εφαρμογής της Στατιστικής στο χώρο της Οικονομικής διαχείρισης και της διαχείρισης των επιχειρήσεων με έμφαση στον τομέα της Αγροτικής Οικονομίας και Ανάπτυξης.

Η εφαρμοσμένη Στατιστική αποτελεί σημαντικό εργαλείο δουλειάς για ένα μελλοντικό επιστήμονα που θα ασχοληθεί με θέματα Αγροτικής Οικονομίας και Ανάπτυξης.

Για τον λόγο αυτό ο φοιτητής-α πρέπει να εμπεδώσει τις βασικές έννοιες της Στατιστικής και να ασχοληθεί με πρακτικά προβλήματα που αναδεικνύουν τον σπουδαίο ρόλο της Στατιστικής στο πεδίο της εφαρμογής.

Το ξεκίνημα αυτής της προσπάθειας θέλω να πιστεύω ότι θα έχει συνέχεια και ότι λάθη και παραλήψεις στο παρόν εγχείρημα θα διορθωθούν στην πορεία με τελικό στόχο να προκύψει ένα βοήθημα απλό – ουσιαστικό και κατανοητό για όλους όσους ενδιαφέρονται και ασχολούνται άμεσα με τα θέματα της Αγροτικής Οικονομίας και Ανάπτυξης.

Θ. Χ. Κουτρουμανίδης

Ε ι σ α γ ω γ ή

Η Στατιστική πρωτοεμφανίστηκε στην αρχαιότητα ως πρακτική μεθοδολογία καταμέτρησης δεδομένων και στην πορεία εξελίχθηκε σ΄ ένα κλάδο επιστημονικό με την αντίστοιχη θεωρητική υποδομή κυρίως με την εισαγωγή της θεωρίας των πιθανοτήτων.

Ποτέ όμως δεν έπαψε να αποτελεί εργαλείο δουλειάς και έρευνας για όλους τους επιστημονικούς κλάδους θεωρητικούς και θετικούς.

Σήμερα με την ραγδαία εξέλιξη της πληροφορικής η Στατιστική έχει ενσωματωθεί σε πάρα πολλές επιστήμες και αποτελεί αναπόσπαστο τμήμα τους. Παράλληλα επικρατεί η άποψη ότι μειώνεται ο ρόλος του επιστήμονα – ερευνητή σ΄ ένα περιβάλλον που κατακλύζεται από στατιστικά προγράμματα μέσω Η/Υ.

Εδώ θα πρέπει να είμαστε σαφείς και κατηγορηματικοί ο ερευνητής έχει τον πρώτο και τον τελευταίο λόγο στην επεξεργασία των δεδομένων μέσω στατιστικών πακέτων και θα πρέπει να έχει κατανοήσει καλά τις βασικές έννοιες της Στατιστικής (περιγραφικής και επαγωγικής) για να μπορέσει να κάνει σωστά τη δουλειά του. Άρα προέχει η εκπαίδευση του σε θέματα Στατιστικής και δεν μπορεί κανένα στατιστικό πρόγραμμα να υποκαταστήσει τον ερευνητή.

Το παρόν σύγγραμμα κινείται σ΄ αυτήν την αντίληψη, να δώσει δηλαδή απλά και ουσιαστικά τις βασικές γνώσεις της Στατιστικής και να παρουσιάσει εφαρμογές της Στατιστικής στον χώρο της Αγροτικής Οικονομίας και Ανάπτυξης, ώστε να μπορεί ο μελλοντικός επιστήμονας που θα ασχοληθεί με τα θέματα της Αγροτικής Ανάπτυξης να χρησιμοποιεί στατιστικές μεθόδους και τεχνικές για την δουλειά του.

Η δομή του συγγράμματος περιλαμβάνει τα εξής μέρη: Παράγωγες κατανομές (Κεφάλαιο1^{ο}), Κατανομές στατιστικών δείγματος (Κεφάλαιο 2^{ο}), Εκτιμητική (Κεφάλαιο 3), Έλεγχοι υποθέσεων (Κεφάλαιο 4), Έλεγχο προσαρμοστικότητας (Κεφάλαιο 5), Ανάλυση κατηγορικών μεταβλητών (Κεφάλαιο 6), Απλή Γραμμική Παλινδρόμηση (Κεφάλαιο 7), Χρονολογικές σειρές (Κεφάλαιο 8), Αριθμοδείκτες (Κεφάλαιο 9).

Κεφάλαιο 1ο

Παράγωγες Κατανομές

1.1. Η X^2 - Κατανομή

Θεωρούμε τις τυχαίες μεταβλητές $X_1, X_2, \ldots, X_v$ που είναι ανεξάρτητες μεταξύ τους και όπου η κάθε μια μεταβλητή έχει τυπική κανονική κατανομή N(0,1).

Δημιουργούμε την τυχαία μεταβλητή $X^2 = X_1^2 + X_2^2 + \ldots + X_v^2$.

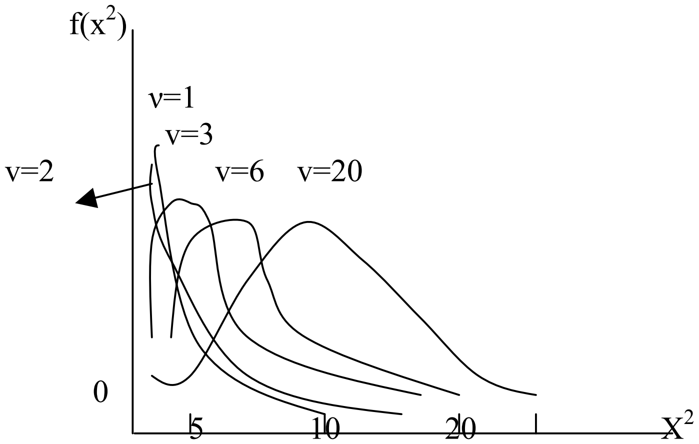
Η τυχαία αυτή μεταβλητή ακολουθεί μια κατανομή που ονομάζεται **Χ - τετράγωνο κατανομή, X^2**.

Η συνάρτηση πυκνότητας πιθανότητας αυτής η $f(x^2)$ έχει **v-βαθμούς ελευθερίας:**

$$f(x^2) = \begin{cases} \dfrac{1}{2^{v/2}\,\Gamma(v/2)}\, e^{-\frac{x}{2}} \cdot x^{\frac{v}{2}-1}, & x > 0 \\ 0, & x < 0 \end{cases}$$

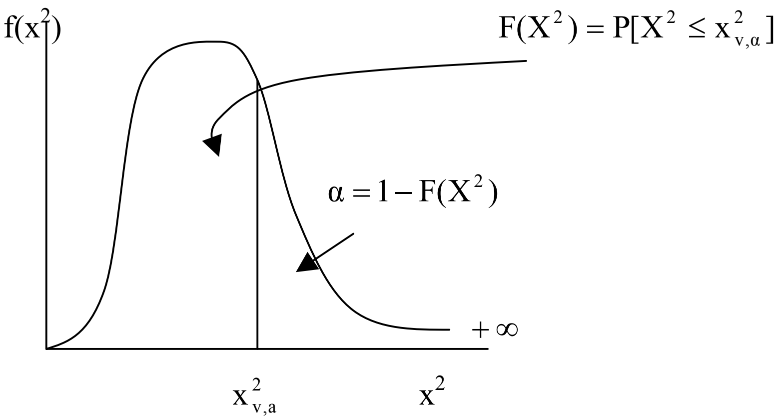
όπου $\Gamma(\kappa) = (\kappa-1)!$

Έχουμε $E[X^2] = v$ και $Var[X^2] = 2v$.



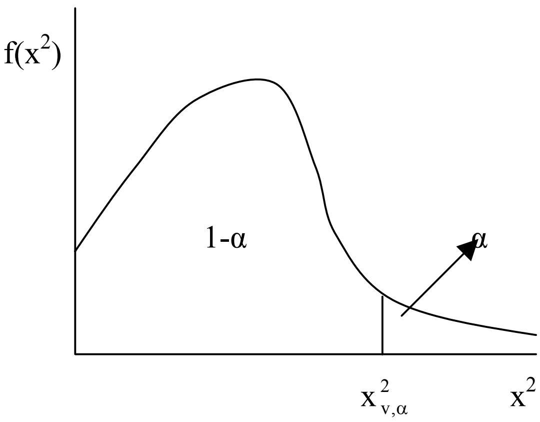
Όταν ν>30 η κατανομή X^2 είναι σχεδόν συμμετρική και λέμε ότι τείνει προς την κανονική.

Η αθροιστική συνάρτηση $F(x^2)$ εκφράζει το εμβαδόν τμήματος.



Ο πίνακας δίνει τις τιμές $x^2_{\nu,\alpha}$ της X για τις οποίες έχουμε την πιθανότητα

$P[X^2 > x^2_{\nu,\alpha}] = \alpha,$ όπου α η δοσμένη πιθανότητα και ν - βαθμοί ελευθερίας γνωστοί.



1.2. Η t - Κατανομή.

Θεωρούμε δύο ανεξάρτητες μεταξύ τους τυχαίες μεταβλητές X και Y από τις οποίες η X έχει κανονική κατανομή και η Y ακολουθεί την κατανομή X^2 με ν-βαθμούς ελευθερίας.

Η τυχαία μεταβλητή $T = \dfrac{X}{\sqrt{Y/\nu}}$ ακολουθεί την κατανομή t με ν-βαθμούς ελευθερίας.

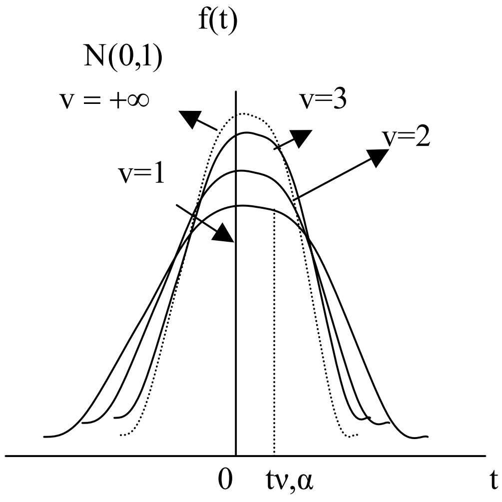
Η Συνάρτηση πυκνότητας πιθανότητας της κατανομής αυτής

$f(t)=\beta\left(1+\frac{t^2}{v}\right)^{-\frac{v+1}{2}}$, $-\infty < t < +\infty$, όπου $\beta=\frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{\pi v}}$ έχει ν-βαθμούς ελευθερίας.

Με E[T]=0 και $\mathrm{Var}[T]=\frac{v}{v-2}$.

$$\lim_{v\to+\infty} Var[T]=1.$$

Όταν ο ν (βαθμοί ελευθερίας) είναι μεγάλος η κατανομή t τείνει στην τυπική κανονική κατανομή N (0,1).

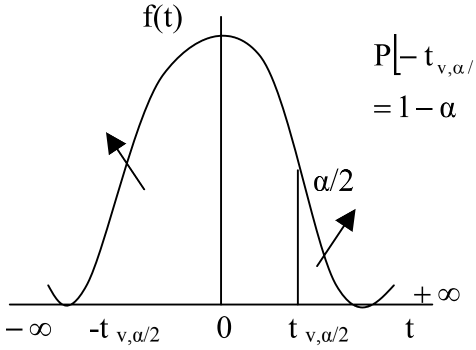


Έχουμε :

Σε πίνακα δίνονται οι τιμές $t_{v,\alpha}$ της T για τις οποίες έχουμε την πιθανότητα

$P[T<t_{v,\alpha}] = \alpha$ και ν – οι βαθμοί ελευθερίας .

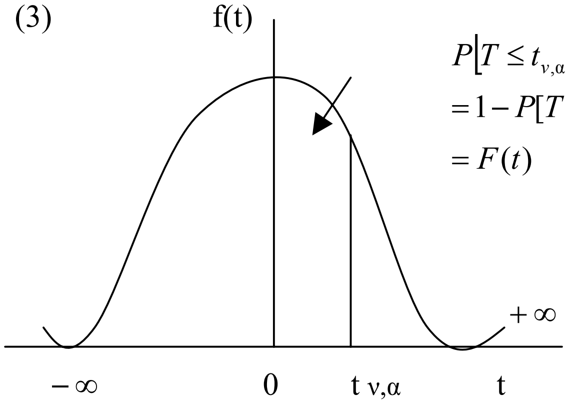
(2)



$$P\left[-t_{\nu,\alpha/2} < T < t_{\nu,\alpha/2}\right] = \\ = 1-\alpha$$

Επίσης

(3)



$$P\left[T \le t_{\nu,\alpha}\right] = \\ = 1 - P[T > t_{\nu,\alpha}] = \\ = F(t)$$

1.3. Η F - Κατανομή

Μια τυχαία μεταβλητή που παράγεται ως λόγος δύο άλλων τυχαίων μεταβλητών ακολουθεί την F – Κατανομή η οποία εκφράζεται με δύο βαθμούς ελευθερίας το ν_1, που είναι ο βαθμός ελευθερίας του αριθμητή και το ν_2, που είναι ο βαθμός ελευθερίας του παρανομαστή. Έτσι συμβολίζεται με F ν_1, ν_2.

Υπάρχουν πίνακες που δίνουν τις κριτήριες τιμές F $\nu_1, \nu_2 ;\alpha$ οι οποίες ορίζονται από τη σχέση:

$$P(X > F\ \nu_1, \nu_2 ;\alpha) = \alpha$$

Για τις διάφορες τιμές των α, ν_1, ν_2.

Κεφάλαιο 2ο

Κατανομές στατιστικών δείγματος

2.1. Κατανομή μέσης τιμής δείγματος

Αν πάρουμε πολλά τυχαία δείγματα μεγέθους n από ένα πληθυσμό τότε προκύπτει η τυχαία μεταβλητή $\bar{X}$, από τις τιμές των μέσων τιμών των δειγμάτων. Όταν το δείγμα προέρχεται από **κανονικό πληθυσμό** N(μ, σ^2) ανεξάρτητα από το μέγεθός του έχουμε ότι η $\bar{X}$ ακολουθεί κανονική κατανομή N(μ, $\frac{\sigma^2}{n}$), το ίδιο ισχύει και στην περίπτωση που το δείγμα δεν προέρχεται από κανονικό πληθυσμό αλλά το μέγεθός του είναι μεγάλο **n $\geq$ 30.**

Τότε η $Z = \dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ ή $Z = \dfrac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}$ ακολουθεί **τυπική κανονική κατανομή**.

Όταν το μέγεθος του δείγματος είναι **μικρό**, η **διακύμανση του πληθυσμού άγνωστη** και ο πληθυσμός, από όπου προέρχεται το δείγμα, **κανονικός** τότε η ποσότητα:

$t = \dfrac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}$ ακολουθεί την **t-κατανομή με n-1 βαθμούς ελευθερίας.**

2.2. Κατανομή διαφοράς μέσων τιμών δύο δειγμάτων

Αν πάρουμε **δύο δείγματα μεγάλα** μεγέθους ν και κ με μέσες τιμές $\bar{x}$ και $\bar{y}$ και διακυμάνσεις s_1^2, s_2^2 και οι πληθυσμιακές διακυμάνσεις είναι σ_1^2, σ_2^2 αντίστοιχα τότε έχουμε ότι η μεταβλητή:

$\dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{\nu}+\frac{\sigma_2^2}{\kappa}}}$ ακολουθεί N(0,1).

Εφόσον οι πληθυσμιακές διακυμάνσεις είναι **άγνωστες** αντικαθίστανται από τις δειγματικές διακυμάνσεις s_1^2, s_2^2 και έχουμε ότι η μεταβλητή:

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{\nu}+\frac{s_2^2}{\kappa}}}$$ ακολουθεί N(0,1).

Αν πάρουμε δύο δείγματα **μικρά ανεξάρτητα** μεγέθους ν και κ με μέσες τιμές $\bar{x}$ και $\bar{y}$ και διακυμάνσεις s_1^2, s_2^2 αντίστοιχα προερχόμενα από **κανονικούς πληθυσμούς** με **άγνωστες αλλά ίσες διακυμάνσεις** $\sigma_1^2 = \sigma_2^2 = \sigma^2$ τότε έχουμε ότι η μεταβλητή:

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{(\nu-1)s_1^2+(\kappa-1)s_2^2}{\nu+\kappa-2}}\sqrt{\frac{1}{\nu}+\frac{1}{\kappa}}}$$ ακολουθεί **t-κατανομή με ν+κ-2 βαθμούς ελευθερίας**

Αν πάρουμε δύο δείγματα **μικρά ανεξάρτητα** μεγέθους ν και κ με μέσες τιμές $\bar{x}$ και $\bar{y}$ και διακυμάνσεις s_1^2, s_2^2 αντίστοιχα προερχόμενα από **κανονικούς πληθυσμούς** με **άγνωστες και διαφορετικές διακυμάνσεις** $\sigma_1^2 \neq \sigma_2^2$ τότε έχουμε ότι η μεταβλητή:

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{\nu}+\frac{s_2^2}{\kappa}}}$$

ακολουθεί **t – κατανομή** με **βαθμούς ελευθερίας λ = 2(ν-1)** όταν ν = κ και $\lambda = \dfrac{\left(\dfrac{s_1^2}{\nu} + \dfrac{s_2^2}{\kappa}\right)^2}{\dfrac{\left(\dfrac{s_1^2}{\nu}\right)^2}{\nu - 1} + \dfrac{\left(\dfrac{s_2^2}{\kappa}\right)^2}{\kappa - 1}}$ (στρογγυλεμένο στον πλησιέστερο ακέραιο) όταν $\nu \neq \kappa$

2.3. Κατανομή διακύμανσης δείγματος

Η ποσότητα:

$\mathrm{X}^2 = \dfrac{(n-1)s^2}{\sigma^2}$ ακολουθεί την $\mathbf{X}^2$ **κατανομή** με **n-1 βαθμούς ελευθερίας.**

Ο πληθυσμός από όπου πάρθηκε το δείγμα μεγέθους n θεωρείται **κανονικός** με σ^2 και s^2 είναι η δειγματική διακύμανση.

2.4. Κατανομή λόγου διακυμάνσεων δύο δειγμάτων

Σύμφωνα με την θεωρία αν πάρουμε δύο δείγματα ανεξάρτητα μεγέθους n και m αντίστοιχα με s_1^2, s_2^2 τις δειγματικές διακυμάνσεις τους από δύο **κανονικούς πληθυσμούς** με σ_1^2, σ_2^2 τις πληθυσμιακές διακυμάνσεις, τότε ισχύει ότι η ποσότητα:

$\dfrac{s_1^2}{s_2^2}\dfrac{\sigma_2^2}{\sigma_1^2}$ ακολουθεί την F κατανομή με n-1 και m-1 βαθμούς ελευθερίας

2.5. Κατανομή ποσοστού (αναλογίας) δείγματος

Αν εκτιμήσουμε το ποσοστό (αναλογία) $p = \dfrac{x}{n}$ των στοιχείων στο **μεγάλο** δείγμα μεγέθους n που έχουν κάποιο συγκεκριμένο χαρακτηριστικό, τότε σύμφωνα με την

θεωρία το δειγματικό ποσοστό κατανέμεται κανονικά:

Δηλαδή η δειγματική μεταβλητή P_{Δ}, που παράγεται από τις τιμές p, ακολουθεί την κανονική κατανομή:

$$\mathrm{N}\left(P_{\pi\lambda\eta\theta}, \frac{p_{\pi\lambda\eta\theta}(1 - P_{\pi\lambda\eta\theta})}{n}\right)$$

Όπου $P_{\pi\lambda\eta\theta}$ το πραγματικό ποσοστό στον πληθυσμό.

Συνεπώς έχουμε ότι η μεταβλητή:

$$\frac{P_{\Delta} - P_{\pi\lambda\eta\theta}}{\sqrt{\dfrac{P_{\pi\lambda\eta\theta}(1-P_{\pi\lambda\eta\theta})}{n}}}$$ ακολουθεί την **N(0,1).**

2.6. Κατανομή διαφοράς ποσοστών (αναλογιών) δύο δειγμάτων

Έστω $p_1 = \dfrac{x}{n}$ και $p_2 = \dfrac{y}{l}$ είναι οι δειγματικές αναλογίες των στοιχείων, που έχουν κάποιο συγκεκριμένο χαρακτηριστικό, δύο δειγμάτων μεγέθους n, l από δύο πληθυσμούς, όπου οι πραγματικές αναλογίες είναι αντίστοιχα $P_{1\ \pi\lambda\eta\theta}$, $P_{2\ \pi\lambda\eta\theta}$.

Τότε έχουμε ότι οι διαφορές των δειγματικών μεταβλητών $P_{1\ \Delta}$ - $P_{2\ \Delta}$ ακολουθούν κανονική κατανομή:

$$N(P_{1\ \pi\lambda\eta\theta} - P_{2\ \pi\lambda\eta\theta},\ \frac{P_{1\pi\lambda\eta\theta}(1-P_{1\pi\lambda\eta\theta})}{n} + \frac{P_{2\pi\lambda\eta\theta}(1-P_{2\pi\lambda\eta\theta})}{l})$$

Κεφάλαιο 3ο

Εκτιμητική

3.1. Σημειακή εκτίμηση μιας άγνωστης παραμέτρου ενός πληθυσμού ως προς μια μεταβλητή.

Όπως έχει αναφερθεί μέχρι τώρα για να μελετήσουμε ένα μεγάλο πληθυσμό ως προς κάποια μεταβλητή X καταφεύγουμε στην δειγματοληψία.

Λαμβάνουμε ένα δείγμα n - στοιχείων του πληθυσμού και μελετάμε αυτό ως προς την μεταβλητή X. Τα συμπεράσματα που προκύπτουν χαρακτηρίζουν, μέσω της επαγωγικής σκέψης, και το πληθυσμό.

Αυτό άλλωστε αποτελεί και τη βασική σκέψη της **Επαγωγικής Στατιστικής** της οποίας ένα από τα σπουδαιότερα κεφάλαια είναι αυτό της **Εκτιμητικής.**

Σε κάθε δείγμα προσδιορίζονται οι στατιστικές παράμετροι αυτού, όπως είναι η μέση τιμή $\overline{x}$, η διακύμανση s^2, η τυπική απόκλιση s κλπ.

Η τιμή μιας από αυτές τις παραμέτρους ονομάζεται **σημειακή εκτιμήτρια** της αντίστοιχης παραμέτρου του πληθυσμού.

Γενικά αν ονομάσουμε $\widehat{\Theta}$ την σημειακή εκτιμήτρια της αντίστοιχης παραμέτρου Θ του πληθυσμού αυτή θα πρέπει να ικανοποιεί κάποιες βασικές ιδιότητες.

Κατ΄ αρχήν επειδή μπορούμε να πάρουμε πολλά δείγματα ίσου μεγέθους από τον πληθυσμό θα έχουμε και τις αντίστοιχες σημειακές εκτιμήτριες των δειγμάτων αυτών για την άγνωστη παράμετρο του πληθυσμού.

Δηλαδή αν πάρουμε ν δείγματα ίσου μεγέθους από τον πληθυσμό θα έχουμε και τις αντίστοιχες $\widehat{\Theta}_\nu$ εκτιμήτριες της Θ.

Γίνεται λοιπόν φανερό ότι η εκτιμήτρια $\widehat{\Theta}$ είναι και αυτή τυχαία μεταβλητή. Η τιμή της τυχαίας μεταβλητής $\widehat{\Theta}$ σε ένα συγκεκριμένο i-δείγμα $\widehat{\Theta}_i$ i =1,2,3,4,…,ν λέγεται σημειακή εκτίμηση της παραμέτρου Θ.

Οι βασικές ιδιότητες που πρέπει να πληρούνται από την σημειακή εκτιμήτρια $\widehat{\Theta}$ είναι:

i. **Η Αμεροληψία:** θα πρέπει δηλαδή η μέση τιμή της τυχαίας μεταβλητής $\widehat{\Theta}$ να είναι η άγνωστη παράμετρος Θ :

$$E(\widehat{\Theta})=\Theta.$$

Αν $E(\widehat{\Theta})\neq\Theta$ τότε η εκτιμήτρια είναι **μεροληπτική** και έχουμε το **σφάλμα μεροληψίας ή σφάλμα εκτίμησης:** $E(\widehat{\Theta})-\Theta$.

ii. Η Αποτελεσματικότητα:

Θα πρέπει η διακύμανση της αμερόληπτης εκτιμήτριας $\widehat{\Theta}$, η $\mathrm{Var}(\widehat{\Theta})$, και είναι μικρότερη ή ίση από την διακύμανση οποιαδήποτε άλλης αμερόληπτης εκτιμήτριας $(\widehat{\Theta})^*$:

$$Var(\widehat{\Theta}) \leq Var(\widehat{\Theta}^*).$$

iii. Η συνέπεια:

Μια σημειακή εκτιμήτρια $\widehat{\Theta}$ είναι συνεπής όταν το σφάλμα μεροληψίας και η διακύμανσή της τείνουν στο μηδέν καθώς το μέγεθος του δείγματος n τείνει στο άπειρο:

$$\lim_{n \to +\infty} E(\widehat{\Theta}) = \Theta \qquad \kappa\alpha\iota \qquad \lim_{n \to +\infty} Var(\widehat{\Theta}) = 0.$$

3.2. Σημειακές εκτιμήτριες της μέσης τιμής και της διακύμανσης ενός πληθυσμού ως προς μια μεταβλητή.

i. Σημειακή εκτίμηση της μέσης τιμής ενός πληθυσμού ως προς μια μεταβλητή.

Έστω $x_1, x_2, \ldots x_n$ οι παρατηρήσεις της μεταβλητής X και μ η άγνωστη μέση τιμή του πληθυσμού.

Έχουμε προσδιορίσει την δειγματική μέση τιμή $\overline{x}$, την σημειακή εκτιμήτρια της μ του πληθυσμού, από τον τύπο:

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{\nu} xi$$

Αν πάρουμε πολλά δείγματα ίσου μεγέθους μπορούμε να ορίσουμε την τυχαία μεταβλητή $\overline{X}$:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{\nu} Xi$$

όπου $X_1, X_2, \ldots X_n$ τυχαίες μεταβλητές με κατανομή την ίδια με αυτή της X.

Η μέση τιμή $E(\overline{X})$ είναι μια αμερόληπτη σημειακή εκτιμήτρια της πραγματικής μέσης τιμής μ του πληθυσμού και ισχύει:

$$E(\overline{X}) = \mu, Var(\overline{X}) = \frac{\sigma^2}{n} \text{ και τυπική απόκλιση } \frac{\sigma}{\sqrt{n}}.$$

Στις συμμετρικές κατανομές ως σημειακή εκτιμήτρια μπορεί να ληφθεί και η διάμεσος Μ, διότι έχουμε:

$$E(M) = \mu.$$

ii. Σημειακή εκτίμηση της διακύμανσης ενός πληθυσμού ως προς μια μεταβλητή.

Ομοίως η άγνωστη διακύμανση ενός πληθυσμού σ^2 εκτιμάται από την διακύμανση ενός δείγματος x_1, x_2,…….x_n τιμών ως προς την μεταβλητή Χ του πληθυσμού.

Η δειγματική διακύμανση συμβολίζεται με s^2 και είναι σύμφωνα με όσα αναφέρθησαν:

$$s^2 = \frac{\sum_{i=1}^{\nu} (xi - \overline{x})^2}{n-1}$$

Η δειγματική αυτή διακύμανση είναι τυχαία μεταβλητή S^2, επειδή μπορούμε να λάβουμε πολλά δείγματα ίσου μεγέθους του πληθυσμού και να ορίσουμε έτσι τις αντίστοιχες διακυμάνσεις.

$$S^2 = \frac{\sum_{i=1}^{\nu} (Xi - \overline{X})^2}{n-1}$$

όπου Xi οι προαναφερόμενες τυχαίες μεταβλητές και $\overline{X}$ η μέση τιμή αυτών.

Αποδεικνύεται ότι $E(S^2) = \sigma^2$, δηλαδή η διακύμανση του δείγματος s^2 είναι μια τιμή της τυχαίας μεταβλητής S^2 που έχει μέση τιμή $E(S^2)$ τη διακύμανση σ^2 του πληθυσμού.

3.3. Διάστημα εμπιστοσύνης της παραμέτρου Θ ενός πληθυσμού.

Επειδή η δειγματική σημειακή εκτιμήτρια $\widehat{\Theta}$ της πραγματικής τιμής Θ του πληθυσμού δεν μας δίνει πληροφορίες περί του βαθμού ακρίβειάς της, δηλαδή πόσο κοντά στην τιμή Θ βρίσκεται, καταφεύγουμε στο να υπολογίσουμε ένα διάστημα που με κάποια **προκαθορισμένη πιθανότητα** θα περιέχει την άγνωστη τιμή του πληθυσμού.

Το διάστημα αυτό το ονομάζουμε **«διάστημα εμπιστοσύνης»** της τιμής Θ.

Το διάστημα αυτό (β, γ) είναι το διάστημα στο οποίο εκτιμούμε ότι θα βρίσκεται η τιμή Θ του πληθυσμού με ορισμένη **πιθανότητα ή επίπεδο εμπιστοσύνης 1-α.**

$$P(\beta \le \Theta \le \gamma) = 1 - \alpha \qquad \mu\varepsilon \qquad 0 \le \alpha \le 1.$$

Τα β, γ ονομάζονται **όρια εμπιστοσύνης** και η πιθανότητα α καλείται **επίπεδο σημαντικότητας.**

3.4. Εκτίμηση διαστήματος εμπιστοσύνης της μέσης τιμής μ πληθυσμού.

i. Κανονικός πληθυσμός (γνωστή η διακύμανση σ^2).

Είναι δεκτό ότι εφόσον ο πληθυσμός που αντιστοιχεί στην τυχαία μεταβλητή X είναι κανονικής κατανομής τότε και η δειγματική εκτιμήτρια $\overline{x}$ είναι κανονικής κατανομής.

Δηλαδή ισχύει ότι η τυχαία μεταβλητή $\bar{X}$, που παίρνει τιμές τις $\overline{x}$, ακολουθεί την κανονική κατανομή $N(\mu, \frac{\sigma}{\sqrt{n}})$.

Επομένως η τυχαία μεταβλητή $\frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ με τιμές $\frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ ακολουθεί μια τυπική κανονική κατανομή.

Η πιθανότητα η τιμή $\frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ να βρίσκεται μέσα σ' ένα δοσμένο διάστημα για παράδειγμα το $\pm$ 1,96 δίδεται από τη σχέση:

$$P\left[-1{,}96 \le \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \le 1{,}96\right] = \phi(1{,}96) - \phi(-1{,}96) = 0{,}95.$$

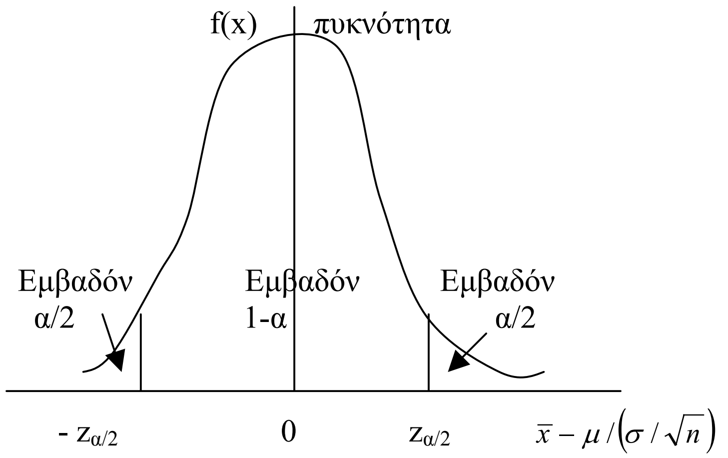
Από την σχέση αυτή αν είναι γνωστή η σ μπορούμε να πούμε ότι:

$\bar{x} - 1{,}96\sigma / \sqrt{n} \le \mu \le \bar{x} + 1.96\sigma / \sqrt{n}$, δηλαδή η μέση τιμή μ του πληθυσμού με πιθανότητα 95% βρίσκεται σε απόσταση $\pm 1{,}96\sigma / \sqrt{n}$ από την δειγματική μέση τιμή $\overline{x}$.

Παρατήρηση:

Μια άλλη ερμηνεία της σχέσεως αυτής είναι ότι «αν χρησιμοποιηθούν διάφορα δείγματα μεγέθους n για τον προσδιορισμό **«διαστημάτων πιθανότητας 95%»** τότε κατά μέσο όρο το 95% από τα διαστήματα αυτά θα περιέχουν την αληθινή τιμή μ..

Σύμφωνα με τα ανωτέρω αν συμβολίσουμε με 1-α το προκαθορισμένο **«επίπεδο εμπιστοσύνης»** και με $\mp z_{\alpha/2}$ τις τιμές της τυπικής κανονικής μεταβλητής με αντίστοιχες τιμές της Αθροιστικής Συνάρτησης κατανομής α/2 και 1-α/2, όπως φαίνεται στο παρακάτω σχήμα:



Τότε έχουμε την πιθανότητα:

$$P\left[-z_{\alpha/2} \leq \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right] = 1-\alpha.$$

$$\text{ή } P\left[\bar{x}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha.$$

Μπορούμε να πούμε ότι υπάρχει εμπιστοσύνη επιπέδου 1-α ότι το διάστημα που εκτιμήθηκε περιέχει την άγνωστη τιμή της μ. Το διάστημα αυτό ονομάζεται **«διάστημα εμπιστοσύνης της μ σε επίπεδο 1-α»** και δίδεται από τη σχέση:

$$\text{(I) } <\mu>_{1-\alpha} = \left[(\bar{x}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}),(\bar{x}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}})\right], \quad \mu\varepsilon \quad \sigma = \gamma\nu\omega\sigma\tau\acute{\eta}.$$

Παρατήρηση:

Η ανωτέρω σχέση ισχύει για κανονικές τυχαίες μεταβλητές που γνωρίζουμε τη σ. Για μη κανονικούς πληθυσμούς ισχύει προσεγγιστικά και ο βαθμός προσέγγισης αυξάνει όσο μεγαλώνει το μέγεθος του δείγματος.

Σχόλιο

Η διαδικασία προσδιορισμού διαστημάτων εμπιστοσύνης της μ, όταν η σ είναι γνωστή ακολουθεί τα εξής βήματα:

1^o : Επιλέγουμε το επίπεδο εμπιστοσύνης 1-α.

2^o : Υπολογίζουμε την τιμή $z_{\alpha/2}$ από τον πίνακα της τυπικής κανονικής κατανομής.

$$z_{\alpha/2} = \phi^{-1}\left(1 - \frac{\alpha}{2}\right).$$

3^o : Χρησιμοποιούμε τη σχέση (I) θέτοντας στη θέση του $\overline{x}$ τη δειγματική μέση τιμή των n - παρατηρήσεων.

Παραδείγματα.

1. Η ημερήσια συγκέντρωση διαλυμένου οξυγόνου σ' ένα σταθμό μέτρησης ενός ποταμού έχει μετρηθεί για 30 ημέρες. Από προηγούμενες εμπειρίες είναι γνωστό ότι η διασπορά της ημερήσιας συγκέντρωσης είναι 4,2 $(mg/l)^2$. Για τις 30 μετρήσεις η μέση δειγματική τιμή $\overline{x} = 2{,}52 mg/l.$ Να υπολογισθεί το διάστημα εμπιστοσύνης επιπέδου 99% για τη μέση ημερήσια συγκέντρωση διαλυμένου οξυγόνου. Θεωρούμε ότι ο πληθυσμός των μετρήσεων της ημερήσιας συγκέντρωσης διαλυμένου οξυγόνου κατανέμεται κανονικά.

$$1 - \alpha = 0{,}99 \rightarrow \alpha/2 = 0{,}005 \rightarrow z_{0,005} = \phi^{-1}(0{,}995) = 2{,}58$$

$$\sigma/\sqrt{n}\, z_{\alpha/2} = \frac{\sqrt{4{,}2}}{\sqrt{30}} \cdot 2{,}58 = 0{,}965.$$

$$< \mu >_{0.99} = (\overline{x} - 0{,}965, \overline{x} + 0{,}965) = (1{,}56\ ,\ 3{,}49)\, mg/l.$$

Αυτό είναι το διάστημα εμπιστοσύνης της μ σε επίπεδο 99%.

2. Σε ένα δείγμα 10 κουτιών παστεριωμένου γάλακτος που παράγει μια βιομηχανία γάλακτος το μέσο βάρος των κουτιών είναι 250 γραμμάρια. Από προηγούμενες μετρήσεις είναι γνωστή η διακύμανση που παρατηρείται και ίση, με 60 γραμμάρια. Να προσδιοριστεί το διάστημα εμπιστοσύνης στο οποίο θα βρίσκεται το μέσο βάρος του συνόλου των κουτιών γάλακτος που παράγονται με πιθανότητα 99,60%. Θεωρούμε ότι ο πληθυσμός των βαρών των κουτιών κατανέμεται κανονικά.

Έχουμε $1-\alpha = 0{,}996 \rightarrow \alpha/2 = 0{,}002$.

Άρα $z_{\alpha/2} = z_{0,002} = \phi^{-1}\left(1-\dfrac{\alpha}{2}\right) = \phi^{-1}(0{,}998) = 2{,}88$.

Επίσης $\sigma/\sqrt{n} \cdot z_{\alpha/2} = \dfrac{\sqrt{60}}{\sqrt{10}} \cdot 2{,}88 = 7{,}05$.

Το διάστημα εμπιστοσύνης:

$$<\mu>_{1-\alpha} = \left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \Rightarrow$$

$$<\mu>_{0.996} = (\bar{x} - 7{,}05, \bar{x} + 7{,}05) = (250 - 7{,}05, 250 + 7{,}05) \Rightarrow$$

$$<\mu>_{0.996} = (242{,}95 \quad , \quad 257{,}05) \text{ γραμμάρια.}$$

Το διάστημα εμπιστοσύνης της μ με πιθανότητα ή επίπεδο εμπιστοσύνης 99,6%.
Αν ο πληθυσμός της ημερήσιας συγκέντρωσης είναι κανονικός τότε το παραπάνω διάστημα είναι ακριβές αν όχι τότε το δεχόμαστε προσεγγιστικά.

ii. Δείγμα μεγάλο άγνωστή η διακύμανση σ^2.

Όταν το δείγμα είναι μεγάλο σύμφωνα με το κεντρικό οριακό θεώρημα η τυχαία μεταβλητή $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ με τιμές $\dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ ακολουθεί μια τυπική κανονική κατανομή.

Συνεπώς το **«διάστημα εμπιστοσύνης της μ σε επίπεδο 1-α»** δίδεται επίσης από τη σχέση:

$$(\mathrm{I})\ <\mu>_{1-\alpha}=\left[(\bar{x}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}),(\bar{x}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}})\right],$$

εφόσον η διακύμανση σ^2 είναι άγνωστη αντικαθίσταται από την s^2 οπότε καταλήγουμε στη σχέση:

$$<\mu>_{1-\alpha}=\left[(\bar{x}-z_{\alpha/2}\frac{s}{\sqrt{n}}),(\bar{x}+z_{\alpha/2}\frac{s}{\sqrt{n}})\right]$$

Παράδειγμα:

Η ημερήσια συγκέντρωση διαλυμένου οξυγόνου σ' ένα σταθμό μέτρησης ενός ποταμού έχει μετρηθεί για 29 ημέρες και είχαμε $\bar{x}$=2,52 mg/l και s^2=4,2 (mg/l)^2 (άγνωστη η σ^2). Να υπολογισθεί το διάστημα εμπιστοσύνης επιπέδου 99% για τη μέση ημερήσια συγκέντρωση διαλυμένου οξυγόνου.

Το διάστημα εμπιστοσύνης για τη μέση ημερήσια συγκέντρωση διαλυμένου οξυγόνου είναι: 1-α = 0,99→ α/2 = 0,005.

ν= 29-1 = 28.

Άρα $t_{\alpha/2,}\,\nu = t_{0,005\,,\,28} = 2,763$ (παράρτημα).

Οπότε:

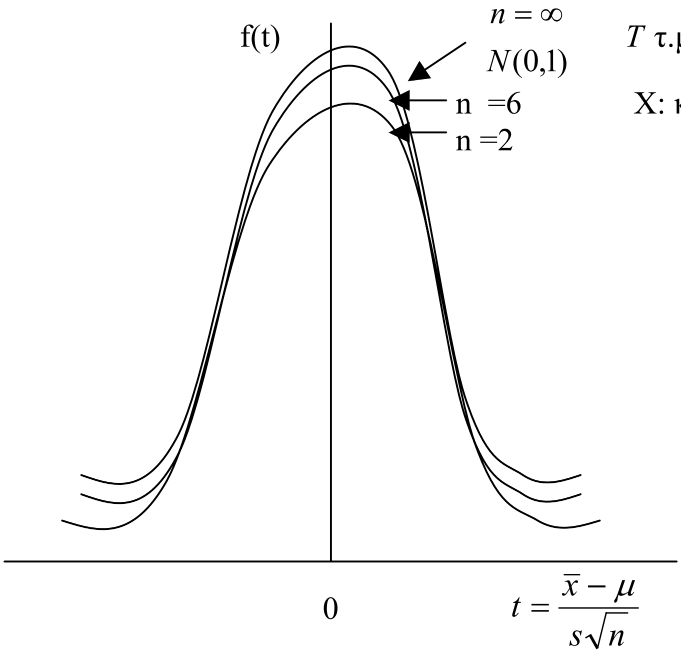
$$<\mu>_{0.99}=\left(2,52-2,763\frac{\sqrt{4,2}}{\sqrt{29}},2,52+2,763\frac{\sqrt{4,2}}{\sqrt{29}}\right)=$$

$$=(\ 1,47,\ \ 3,57)\ mg/l.$$

ii. Δείγμα μικρό άγνωστή η διακύμανση σ^2.

Για n μικρό πολύ π.χ. n < 10 τα διαστήματα εμπιστοσύνης της μ από την (Ι) με s^2 στη θέση του σ^2 είναι αρκετά ανακριβή (όταν δεν είναι γνωστή η διασπορά σ).

Όταν η κατανομή του πληθυσμού της Χ είναι κανονική τότε και αν ακόμα ή σ^2 δεν είναι γνωστή μπορούν να προσδιοριστούν ακριβή διαστήματα εμπιστοσύνης για τη μ. Έχει αποδειχθεί ότι η τυχαία μεταβλητή $T\ \mu\varepsilon\ \tau\iota\mu\acute{\varepsilon}\varsigma=\frac{\bar{x}-\mu}{s\sqrt{n}}$ έχει κατανομή t με n-1 βαθμούς ελευθερίας και με συνάρτηση πυκνότητας f(t).



T τ.μ. $\text{με τιμές} = \dfrac{\overline{x}-\mu}{s\sqrt{n}}$

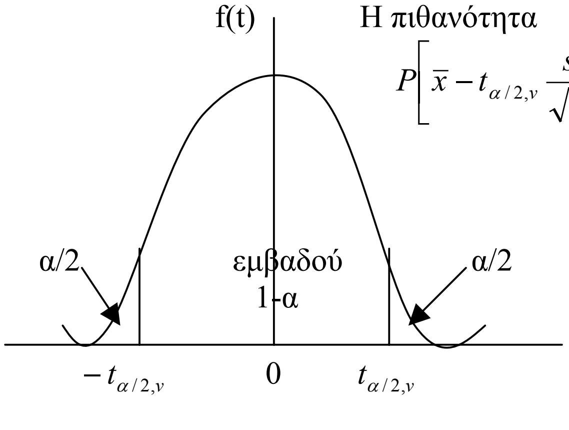
X: κανονική μεταβλητή

Η κατανομή t είναι συμμετρική ως προς το μηδέν και έχει σχήμα παρόμοιο της κανονικής κατανομής, όταν το n μεγαλώνει (ν = n -1 = βαθμοί ελευθερίας) τότε η κατανομή t προσεγγίζει την κανονική κατανομή.

Έχουμε λοιπόν:

$$P\left[-t_{\alpha/2,v} \le \frac{\overline{x}-\mu}{s\sqrt{n}} \le t_{\alpha/2,v}\right]=1-\alpha.$$

Το $t_{\alpha/2,v}$ είναι η τιμή της τυχαίας μεταβλητής Τ που αντιστοιχεί σε τιμή της Αθροιστικής συνάρτησης κατανομής ίση με 1-α/2, για ν = n-1 βαθμούς ελευθερίας. Οι τιμές του $t_{\alpha/2,v}$ δίδονται από πίνακα (παράρτημα).



Η πιθανότητα

$$P\left[\overline{x}-t_{\alpha/2,v}\frac{s}{\sqrt{n}} \le \mu \le \overline{x}+t_{\alpha/2,v}\frac{s}{\sqrt{n}}\right]=1-\alpha.$$

μας οδηγεί στη σχέση

$$< \mu >_{1-\alpha} = \left[(\bar{x} - t_{\alpha/2,v} \frac{s}{\sqrt{n}}), (\bar{x} + t_{\alpha/2,v} \frac{s}{\sqrt{n}}) \right]$$

όπου

$\overline{\mathrm{x}}$: δειγματική μέση τιμή

s: δειγματική τυπική απόκλιση. Αυτό είναι το διάστημα εμπιστοσύνης της μ, με άγνωστη τη σ, σε επίπεδο 1-α.

Παρατήρηση: Όταν ο αριθμός των βαθμών ελευθερίας είναι πολύ μεγάλος τότε $t_{\alpha/2}, v \rightarrow z_{\alpha/2}$ διότι η κατανομή t τείνει τότε στην τυπική κανονική κατανομή.

Παράδειγμα:

Σε ένα δείγμα 20 κουτιών παστεριωμένου γάλακτος που παράγει μια βιομηχανία γάλακτος το μέσο βάρος των κουτιών είναι 250 γραμμάρια και η δειγματική διακύμανση $s^2 = 56{,}25$. Να προσδιοριστεί το διάστημα εμπιστοσύνης στο οποίο θα βρίσκεται το μέσο βάρος του συνόλου των κουτιών γάλακτος που παράγονται με πιθανότητα 99,60%.

Γνωρίζουμε ότι:

$$< \mu >_{1-\alpha} = \left[(\bar{x} - t_{\alpha/2,v} \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2,v} \frac{s}{\sqrt{n}}) \right]$$

$$1 - \alpha = 0{,}998 \rightarrow \frac{\alpha}{2} = 0{,}001$$

$$v = 20 - 1 = 19$$

$$t_{\alpha/2,v} = t_{0{,}001\,,19} = 3{,}883. (\text{παράρτημα}).$$

$$t_{\alpha/2,v} \frac{s}{\sqrt{n}} = 3{,}883 \cdot \frac{7{,}5}{\sqrt{20}} = 6{,}512.$$

Άρα $< \mu >_{0{,}998} = (250 - 6{,}512 \quad , \quad 250 + 6{,}512)\mathrm{kgr} \Rightarrow$

$< \mu >_{0{,}998} = (243{,}488 \quad , \quad 256{,}512)\ \mathrm{kgr}$, το διάστημα εμπιστοσύνης.

Μονόπλευρα όρια εμπιστοσύνης για τη μέση τιμή μ.

i. **Η διακύμανση σ γνωστή:** Χρησιμοποιείται η τ.μ. με τιμές $\dfrac{\overline{x}-\mu}{\sigma\sqrt{n}}$ κανονικής κατανομής.

$$<\mu>_{1-\alpha}=\left(\overline{x}-z_\alpha\frac{\sigma}{\sqrt{n}}\right) \qquad z_\alpha=\phi^{-1}(1-\alpha).$$

Αυτό δηλώνει ότι η μέση τιμή του πληθυσμού θα είναι μεγαλύτερη από το όριο αυτό με πιθανότητα 1-α.

* Αυτό προκύπτει από το ότι πρέπει η:

$$P=\left[\frac{\overline{x}-\mu}{\sigma\sqrt{n}}\leq z_\alpha\right]=1-\alpha., \text{ άρα } \quad P\left[\mu\geq\overline{x}-z_\alpha\frac{\sigma}{\sqrt{n}}\right]=1-\alpha.$$

με $z_\alpha=\phi^{-1}(1-\alpha)$.

Ανώτατο όριο εμπιστοσύνης της μ επιπέδου 1-α.

$$(\mu>_{1-\alpha}=\left(\overline{x}-z_\alpha\frac{\sigma}{\sqrt{v}}\right) \quad , \quad z_\alpha=\phi^{-1}(1-\alpha).$$

i) **Η διακύμανση σ άγνωστη (μικρό δείγμα):** χρησιμοποιείται η t - κατανομή για τον προσδιορισμό ανώτατου και κατώτατου ορίου εμπιστοσύνης.

Κατώτατο όριο εμπιστοσύνης της μ, επιπέδου 1-α.

$$<\mu)_{1-\alpha}=\overline{x}-t_{\alpha,v}\frac{s}{\sqrt{n}} \quad , \quad t_{\alpha,v} \text{ από πίνακα.}$$

Ανώτατο όριο εμπιστοσύνης της μ, επιπέδου 1-α.

$$(\mu>_{1-\alpha}=\overline{x}-t_{\alpha,v}\frac{s}{\sqrt{n}} \quad , \quad t_{\alpha,v} \text{ από πίνακα}$$

Παράδειγμα.

Τα εργαστηριακά αποτελέσματα 100 δοκιμίων χάλυβα A36 που διαλέχτηκαν τυχαία δείχνουν για την τάση ροής μια μέση τιμή $\overline{\mathrm{x}} = 2200\mathrm{kp/cm}^2$ και μια τυπική απόκλιση $220\mathrm{kp/cm}^2$.

Να προσδιορισθεί το κατώτατο όριο εμπιστοσύνης της μέσης τιμής μ της τάσης ροής του χάλυβα αυτού σε επίπεδο 95%.

Λόγω μεγάλου μεγέθους n = 100 σ $\simeq$ s = 220.

$$1-\alpha = 0{,}95 \rightarrow \alpha = 0{,}05 \rightarrow z_{0,05} = \phi^{-1}(1-0{,}05) = \phi^{-1}(0{,}95) = 1{,}65. \qquad \textit{παράρτημα}$$

οπότε κατώτατο όριο:

$$<\mu>_{1-\alpha} = \bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}} = 2200 - 1{,}65\frac{220}{\sqrt{100}} = 2164\,kgr.$$

Γενική παρατήρηση:

Οι αποφάσεις ή εκτιμήσεις στη Στατιστική έχουν στοχαστικό χαρακτήρα.
Δεν αποτελούν αποδείξεις.

Έτσι εκτιμούμε ότι ο μέσος ενός πληθυσμού $<\mu>_{1-\alpha} = (\alpha_1, \beta_1)$. Δηλαδή ότι ο μέσος μ βρίσκεται στο διάστημα (α_1, β_1) δίνοντας συγχρόνως την πιθανότητα σφάλματος αυτής της εκτίμησης.

Έτσι λέμε ότι σε επίπεδο εμπιστοσύνης 95% = 1-α η μέση τιμή (μ) του πληθυσμού ανήκει στο διάστημα: (1,36 , 2,18).

3.5. Εκτίμηση διαστήματος εμπιστοσύνης της διαφοράς των μέσων τιμών δύο πληθυσμών.

Θεωρώ την διαφορά των μέσων τιμών δύο δειγμάτων $\bar{x}_1 - \bar{x}_2$ από δύο πληθυσμούς. Η διαφορά αυτή είναι μια τυχαία μεταβλητή:

1. Αν πάρουμε **δύο δείγματα ανεξάρτητα μεγάλα** μεγέθους ν και κ με μέσες τιμές $\bar{x}$ και $\bar{y}$, διακυμάνσεις s_1^2, s_2^2 και **οι πληθυσμιακές διακυμάνσεις είναι γνωστές** σ_1^2 , σ_2^2 αντίστοιχα, τότε έχουμε ότι η μεταβλητή:

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{\nu}+\frac{\sigma_2^2}{\kappa}}}$$ ακολουθεί N(0,1).

Τότε ένα **100(1-α)% διάστημα εμπιστοσύνης** για την διαφορά $\mathbf{\mu_1 - \mu_2}$ των δύο πληθυσμών όταν οι **διακυμάνσεις τους είναι γνωστές** είναι:

$$\bar{x} - \bar{y} - Z_{\frac{a}{2}}\sqrt{\frac{\sigma_1^2}{\nu}+\frac{\sigma_2^2}{\kappa}} \le \mathbf{\mu_1 - \mu_2} \le \bar{x} - \bar{y} + Z_{\frac{a}{2}}\sqrt{\frac{\sigma_1^2}{\nu}+\frac{\sigma_2^2}{\kappa}}$$

Εάν **οι πληθυσμιακές διακυμάνσεις είναι άγνωστες** αντικαθίστανται από τις δειγματικές διακυμάνσεις s_1^2, s_2^2 και έχουμε ότι η μεταβλητή:

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{\nu}+\frac{s_2^2}{\kappa}}}$$ ακολουθεί N(0,1).

Τότε ένα **100(1-α)% διάστημα εμπιστοσύνης** για την διαφορά $\mathbf{\mu_1 - \mu_2}$ των δύο πληθυσμών όταν **οι διακυμάνσεις τους είναι άγνωστες** είναι:

$$\bar{x} - \bar{y} - Z_{\frac{a}{2}}\sqrt{\frac{s_1^2}{\nu}+\frac{s_2^2}{\kappa}} \le \mathbf{\mu_1 - \mu_2} \le \bar{x} - \bar{y} + Z_{\frac{a}{2}}\sqrt{\frac{s_1^2}{\nu}+\frac{s_2^2}{\kappa}}$$

2. Αν πάρουμε δύο δείγματα **μικρά ανεξάρτητα** μεγέθους ν και κ με μέσες τιμές $\bar{x}$ και $\bar{y}$ και διακυμάνσεις s_1^2, s_2^2 αντίστοιχα προερχόμενα από **κανονικούς πληθυσμούς** με **άγνωστες αλλά ίσες διακυμάνσεις** $\boldsymbol{\sigma_1^2 = \sigma_2^2 = \sigma^2}$ έχουμε ότι η μεταβλητή:

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{(\nu-1)s_1^2+(\kappa-1)s_2^2}{\nu+\kappa-2}}\sqrt{\frac{1}{\nu}+\frac{1}{\kappa}}}$$ ακολουθεί **t-κατανομή με ν+κ-2 βαθμούς ελευθερίας**

Τότε ένα **100(1-α)% διάστημα εμπιστοσύνης** για την διαφορά $\mathbf{\mu_1 - \mu_2}$ των δύο πληθυσμών όταν **οι διακυμάνσεις τους είναι άγνωστες αλλά ίσες** είναι:

$$\bar{x} - \bar{y} - t_{\nu+\kappa-2;\frac{\alpha}{2}}\, s\sqrt{\frac{1}{\nu}+\frac{1}{\kappa}} \leq \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \leq \bar{x} - \bar{y} + t_{\nu+\kappa-2;\frac{\alpha}{2}}\, s\sqrt{\frac{1}{\nu}+\frac{1}{\kappa}}$$

$$\text{όπου } s = \sqrt{\frac{(\nu-1)s_1^2 + (\kappa-1)s_2^2}{\nu+\kappa-2}}$$

3. Αν πάρουμε δύο δείγματα **μικρά ανεξάρτητα** μεγέθους ν και κ με μέσες τιμές $\bar{x}$ και $\bar{y}$ και διακυμάνσεις s_1^2, s_2^2 αντίστοιχα προερχόμενα από **κανονικούς πληθυσμούς** με **άγνωστες και διαφορετικές διακυμάνσεις** $\boldsymbol{\sigma}_1^2 \neq \boldsymbol{\sigma}_2^2$ έχουμε ότι η μεταβλητή:

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{\nu} + \frac{s_2^2}{\kappa}}}$$

ακολουθεί **t – κατανομή με βαθμούς ελευθερίας**

λ = 2(ν-1) όταν **ν = κ**

και $\lambda = \dfrac{(\frac{s_1^2}{\nu} + \frac{s_2^2}{\kappa})^2}{\dfrac{(\frac{s_1^2}{\nu})^2}{\nu-1} + \dfrac{(\frac{s_2^2}{\kappa})^2}{\kappa-1}}$ (στρογγυλεμένο στον πλησιέστερο ακέραιο) όταν $\boldsymbol{\nu \neq \kappa}$

Ένα **100(1-α)% διάστημα εμπιστοσύνης** για την διαφορά $\mu_1 - \mu_2$ των δύο πληθυσμών όταν **οι διακυμάνσεις τους είναι άγνωστες και διαφορετικές** $\boldsymbol{\sigma}_1^2 \neq \boldsymbol{\sigma}_2^2$ είναι:

$$\bar{x} - \bar{y} - t_{\lambda;\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{\nu}+\frac{s_2^2}{\nu}} \leq \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \leq \bar{x} - \bar{y} + t_{\lambda;\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{\nu}+\frac{s_2^2}{\nu}} \quad \text{για } \boldsymbol{\nu = \kappa}$$

ή

$$\bar{x} - \bar{y} - t_{\lambda;\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{\nu} + \frac{s_2^2}{\kappa}} \leq \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \leq \bar{x} - \bar{y} + t_{\lambda;\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{\nu} + \frac{s_2^2}{\kappa}} \quad \text{για } \boldsymbol{\nu} \neq \boldsymbol{\kappa}$$

Παραδείγματα:

1. Δύο εργοστάσια κατασκευάζουν το ίδιο εξάρτημα για μια μηχανή.

Παίρνουμε ένα δείγμα 30 εξαρτημάτων από το πρώτο εργοστάσιο και διαπιστώνουμε ότι το μέσο βάρος τους είναι 700 kgr και έχουν δειγματική διακύμανση 400 kgr. Λαμβάνουμε επίσης ένα δείγμα 40 εξαρτημάτων από το δεύτερο εργοστάσιο και βρίσκουμε ότι έχει μέσο βάρος 720 kgr με δειγματική διακύμανση 450 kgr.

Να βρεθεί το διάστημα εμπιστοσύνης της διαφοράς μ_1 - μ_2 των πραγματικών μέσων των πληθυσμών με πιθανότητα 99%. Ποιου εργοστασίου το εξάρτημα είναι βαρύτερο κατά μέσο όρο;

Έχουμε $\bar{x}_1 = 700 \qquad \bar{x}_2 = 720$

$$s_1^2 = 400, s_2^2 = 450$$
$$n_1 = 30, \qquad n_2 = 40.$$

$$\alpha = 1 - 0{,}99 = 0{,}01 \qquad \rightarrow \frac{\alpha}{2} = 0{,}005 \rightarrow z_{\frac{\alpha}{2}} = \phi^{-1}\left(1 - \frac{\alpha}{2}\right) = \phi^{-1}(0{,}995) = 2{,}58.$$

$$s^2_{\bar{x}_1 - \bar{x}_2} = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} = \frac{400}{30} + \frac{450}{40} = 24{,}5833$$

$$s_{\bar{x}_1 - x_2} = 4{,}9581 \qquad \bar{x}_1 - \bar{x}_2 = -20$$

$$-20 - 2{,}58 \cdot 4{,}9581 < \mu_1 - \mu_2 < -20 + 2{,}58 \cdot 4{,}9581$$

$$-32{,}79 < \mu_1 - \mu_2 < -7{,}21 \text{ ή}$$

$$32{,}79 > \mu_1 - \mu_2 > 7{,}21.$$

Το εξάρτημα του δεύτερου εργοστασίου είναι βαρύτερο κατά μέσο όρο.

2. Αν στο πιο πάνω παράδειγμα γνωρίζουμε ότι ύστερα από μετρήσεις οι διακυμάνσεις των πληθυσμών είναι $\sigma_1^2 = 420 και \sigma_2^2 = 480$ τότε το διάστημα εμπιστοσύνης της διαφοράς μ_1-μ_2 , με πιθανότητα 99% , θα είναι:

$$z_{\frac{\alpha}{2}} = 2{,}58$$

$$\sigma^2_{\bar{x}_1-\bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{420}{30} + \frac{480}{40} = 26 \rightarrow \sigma^2_{\bar{x}_1-\bar{x}_2} = 5{,}099$$

$$-20 - 2{,}58 \cdot 5{,}099 < \mu_1 - \mu_2 < -20 + 2{,}58 \cdot 5{,}099$$

$$-33{,}16 < \mu_1 - \mu_2 < -6{,}84 \Rightarrow 33{,}16 > \mu_2 - \mu_1 > 6{,}84.$$

3. Παίρνουμε ένα δείγμα 10 κονσερβών από ένα εργοστάσιο κονσερβοποιίας και διαπιστώνουμε ότι το μέσο βάρος είναι 700 γρ. με διακύμανση 400 $γρ^2$. Λαμβάνουμε επίσης ένα δεύτερο δείγμα 18 κονσερβών από ένα δεύτερο εργοστάσιο κονσερβοποιίας και διαπιστώνουμε ότι το μέσο βάρος είναι 720 γρ. με διακύμανση 450 γρ.

Να βρεθεί το διάστημα εμπιστοσύνης της διαφοράς μ_1-μ_2 των πραγματικών μέσων βαρών των κονσερβών που παράγονται στα δύο εργοστάσια με πιθανότητα 99%.

Θεωρούμε ότι οι πληθυσμοί των βαρών των κονσερβών των δύο εργοστασίων είναι κανονικοί και ότι οι διακυμάνσεις των βαρών των κονσερβών στα δύο εργοστάσια είναι άγνωστες αλλά ίσες.

Ένα 100(1-α)% διάστημα εμπιστοσύνης για την διαφορά μ_1 - μ_2 των δύο πληθυσμών όταν οι διακυμάνσεις τους είναι άγνωστες αλλά ίσες είναι:

$$\bar{x} - \bar{y} - t_{\nu+\kappa-2;\frac{\alpha}{2}} s\sqrt{\frac{1}{\nu}+\frac{1}{\kappa}} \le \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \le \bar{x} - \bar{y} + t_{\nu+\kappa-2;\frac{\alpha}{2}} s\sqrt{\frac{1}{\nu}+\frac{1}{\kappa}}$$

$$\text{όπου } s = \sqrt{\frac{(\nu-1)s_1^2 + (\kappa-1)s_2^2}{\nu+\kappa-2}}$$

Έχουμε $\bar{x} = 700, \bar{y} = 720$

$s_1^2 = 400, s_2^2 = 450$, $n_1 = 10, n_2 = 18.$, $n_1 + n_2 - 2 = 26$,

$\alpha = 1 - 0{,}99 = 0{,}01, \frac{\alpha}{2} = 0{,}005$ και $t_{\nu+\kappa-2;\frac{\alpha}{2}} = 2{,}779$

$$s = \sqrt{\frac{(10-1)400+(18-1)450}{26}} = 21{,}62$$

$$s\sqrt{\frac{1}{\nu}+\frac{1}{\kappa}} = 21{,}62 * 0{,}39 = 8{,}54$$

$$\bar{x}_1 - \bar{x}_2 = -20$$

$$-20 - 2{,}779 \cdot 8{,}54 < \mu_1 - \mu_2 < -20 + 2{,}779 \cdot 8{,}54$$

$$-43{,}73 < \mu_1 - \mu_2 < -3{,}73$$

4. Παίρνουμε ένα δείγμα 10 κονσερβών από ένα εργοστάσιο κονσερβοποιίας και διαπιστώνουμε ότι το μέσο βάρος είναι 700 γρ. με διακύμανση 400 γρ^2. Λαμβάνουμε επίσης ένα δεύτερο δείγμα 10 κονσερβών από ένα δεύτερο εργοστάσιο κονσερβοποιίας και διαπιστώνουμε ότι το μέσο βάρος είναι 720 γρ. με διακύμανση 450 γρ^2.

Να βρεθεί το διάστημα εμπιστοσύνης της διαφοράς μ_1-μ_2 των πραγματικών μέσων βαρών των κονσερβών που παράγονται στα δύο εργοστάσια με πιθανότητα 99%.

Θεωρούμε ότι οι πληθυσμοί των βαρών των κονσερβών των δύο εργοστασίων είναι κανονικοί και ότι οι διακυμάνσεις των βαρών των κονσερβών στα δύο εργοστάσια είναι άγνωστες και διαφορετικές.

Ένα 100(1-α)% διάστημα εμπιστοσύνης για την διαφορά μ_1- μ_2 των δύο πληθυσμών όταν οι διακυμάνσεις τους είναι άγνωστες και διαφορετικές $\sigma_1^2 \neq \sigma_2^2$ είναι:

$$\bar{x} - \bar{y} - t_{\lambda;\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{\nu}+\frac{s_2^2}{\nu}} \le \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \le \bar{x} - \bar{y} + t_{\lambda;\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{\nu}+\frac{s_2^2}{\nu}} \quad \text{για } \boldsymbol{\nu = \kappa}$$

Έχουμε $\bar{x} = 700, \bar{y} = 720$

$s_1^2 = 400, s_2^2 = 450$, $n_1 = 10, n_2 = 10$,

$\alpha = 1 - 0{,}99 = 0{,}01, \frac{\alpha}{2} = 0{,}005$, λ = 2(10-1) = 18

και $t_{\lambda;\frac{\alpha}{2}} =$ 2,878

$$\sqrt{\frac{400}{10}+\frac{450}{10}}=9{,}22$$

$$-20-2{,}878\cdot 9{,}22<\mu_1-\mu_2<-20+2{,}878\cdot 9{,}22$$

ή

$$-46{,}54<\mu_1-\mu_2<6{,}54$$

5. Παίρνουμε ένα δείγμα 10 κονσερβών από ένα εργοστάσιο κονσερβοποιίας και διαπιστώνουμε ότι το μέσο βάρος είναι 700 γρ. με διακύμανση 400 γρ ^{2}. Λαμβάνουμε επίσης ένα δεύτερο δείγμα 16 κονσερβών από ένα δεύτερο εργοστάσιο κονσερβοποιίας και διαπιστώνουμε ότι το μέσο βάρος είναι 720 γρ. με διακύμανση 450 γρ ^{2}.

Να βρεθεί το διάστημα εμπιστοσύνης της διαφοράς μ_1-μ_2 των πραγματικών μέσων βαρών των κονσερβών που παράγονται στα δύο εργοστάσια με πιθανότητα 99%.

Θεωρούμε ότι οι πληθυσμοί των βαρών των κονσερβών των δύο εργοστασίων είναι κανονικοί και ότι οι διακυμάνσεις των βαρών των κονσερβών στα δύο εργοστάσια είναι άγνωστες και διαφορετικές.

Ένα 100(1-α)% διάστημα εμπιστοσύνης για την διαφορά μ_1- μ_2 των δύο πληθυσμών όταν οι διακυμάνσεις τους είναι άγνωστες και διαφορετικές $\sigma_1^2 \neq \sigma_2^2$ είναι:

$$\bar{x}-\bar{y}-t_{\lambda;\frac{\alpha}{2}}\sqrt{\frac{s_1^2}{\nu}+\frac{s_2^2}{\nu}} \le \boldsymbol{\mu}_1-\boldsymbol{\mu}_2 \le \bar{x}-\bar{y}+t_{\lambda;\frac{\alpha}{2}}\sqrt{\frac{s_1^2}{\nu}+\frac{s_2^2}{\nu}} \quad \text{για } \boldsymbol{\nu=\kappa}$$

Έχουμε $\bar{x}=700, \bar{y}=720$

$$s_1^2=400, s_2^2=450, \quad n_1=10, n_2=16$$

$$\alpha=1-0{,}99=0{,}01, \frac{\alpha}{2}=0{,}005 \ ,$$

$$\lambda=\frac{\left(\frac{400}{10}+\frac{450}{16}\right)^2}{\frac{\left(\frac{400}{10}\right)^2}{10-1}+\frac{\left(\frac{450}{16}\right)^2}{16-1}}=20$$

και $t_{\lambda;\frac{\alpha}{2}} = 2{,}845$

$$\sqrt{\frac{400}{10}+\frac{450}{10}} = 9{,}22$$

$$-20-2{,}845\cdot 9{,}22 < \mu_1-\mu_2 < -20+2{,}845\cdot 9{,}22$$

ή

$$-46{,}23 < \mu_1 - \mu_2 < 6{,}23$$

4. Για **δείγματα μικρά εξαρτημένα (ζευγαρωτές παρατηρήσεις)** που προέρχονται από μετρήσεις της ίδιας ομάδας σε δυο διαφορετικές χρονικές στιγμές (ζευγαρωτές παρατηρήσεις) ορίσουμε x_i και y_i, i=1,2,….,n τις παρατηρήσεις στα δύο δείγματα και δημιουργούμε τις αντίστοιχες διαφορές $z_i = x_i - y_i$, i=1,2,….,n που τις θεωρούμε διαφορετικές. Ο πληθυσμός από όπου πήραμε τα ζεύγη θεωρείται **κανονικός.**

Οι παρατηρήσεις z_i ακολουθούν την t – κατανομή με n -1 βαθμούς ελευθερίας.

$\dfrac{\bar{Z}}{\frac{s_Z}{\sqrt{n}}}$ που ακολουθεί t – κατανομή με n -1 βαθμούς ελευθερίας

Ένα 100(1-α)% διάστημα εμπιστοσύνης για τη διαφορά $\boldsymbol{\mu}_1$-$\boldsymbol{\mu}_2$= $\boldsymbol{\mu}_Z$ των δύο πληθυσμών είναι:

$$\bar{Z} - t_{n-1;\frac{a}{2}}\frac{s_z}{\sqrt{n}} \leq \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{\mu}_Z \leq \bar{Z} - t_{n-1;\frac{a}{2}}\frac{s_z}{\sqrt{n}}$$

Όταν το δείγμα είναι μικρό τότε έχουμε : $t_{n-1;\frac{a}{2}} = Z_{\frac{a}{2}}$

Παράδειγμα:

Έχουμε τις παρακάτω ζευγαρωτές παρατηρήσεις:

X: 4 5 6 4,2 5,2 5,3 6,4 4,8 5,3 5

Y: 4,9 4,8 5,7 5 6 5.2 6,5 5,9 4,8 5,7

Να βρεθεί ένα 90% διάστημα εμπιστοσύνης για την πραγματική διαφορά μ_1 - μ_2 .

Έχουμε Z: -0,9 0,2 0,3 -0,8 -0,8 0,1 -0,1 -1,1 0,5 -0,7

Και $\bar{Z}$ = -0,33 , S_Z = 0,72 , $t_{9;0,05}$ = 1,833

Επομένως ένα 90% διάστημα εμπιστοσύνης για την πραγματική διαφορά μ_1 - μ_2 είναι:

$$\bar{Z} - t_{9;0,05}\ S_Z/\sqrt{n} \le \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{\mu}_Z \le \bar{Z} + t_{9;0,05}\ S_Z/\sqrt{n}$$

ή

$$-0{,}33 - 1{,}833 \times 0{,}72/ 3{,}16 \le \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{\mu}_Z \le -0{,}33 + 1{,}833 \times 0{,}72/ 3{,}16$$

ή

$$-0{,}75 \le \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{\mu}_Z \le 0{,}09$$

3.6. Εκτίμηση διαστήματος εμπιστοσύνης για την διακύμανση του πληθυσμού

Στην θεωρία διατυπώθηκε ότι η ποσότητα:

$$X^2 = \frac{(n-1)s^2}{\sigma^2}$$ ακολουθεί την X^2 κατανομή με n-1 βαθμούς.

Ο πληθυσμός από όπου πάρθηκε το δείγμα θεωρείται κανονικός με σ^2 και s^2 η δειγματική διακύμανση.

Ένα 100(1-α)% διάστημα εμπιστοσύνης για την διακύμανση σ^2 του πληθυσμού είναι:

$$\frac{(n-1)s^2}{X^2_{n-1;\frac{a}{2}}} \le \sigma^2 \le \frac{(n-1)s^2}{X^2_{n-1;1-\frac{a}{2}}}$$

Παράδειγμα:

Ένα δείγμα από 51 αγρότες παρουσιάζει διακύμανση των ετήσιων εισοδημάτων τους $s^2 = 156$ ευρώ^2. Να βρεθεί ένα 95% διάστημα εμπιστοσύνης για την διακύμανση σ^2 των ετήσιων εισοδημάτων του πληθυσμού των αγροτών.

Έχουμε ότι:

$$\frac{(n-1)s^2}{X^2_{n-1;\frac{a}{2}}} \leq \sigma^2 \leq \frac{(n-1)s^2}{X^2_{n-1;1-\frac{a}{2}}}$$

ή

$$\frac{(51-1)156}{71{,}42} \leq \sigma^2 \leq \frac{(51-1)156}{32{,}36}$$

$[X^2_{50;0,025} = 71{,}42\ ,\ X^2_{50;0,975} = 32{,}36]$

ή

$$109{,}21 \leq \sigma^2 \leq 241{,}04$$

3.7. Εκτίμηση διαστήματος εμπιστοσύνης για τον λόγο των διακυμάνσεων δύο πληθυσμών

Σύμφωνα με την θεωρία αν πάρουμε δύο δείγματα ανεξάρτητα μεγέθους n και m αντίστοιχα με s_1^2, s_2^2 τις δειγματικές διακυμάνσεις από δύο **κανονικούς πληθυσμούς** αντίστοιχα με σ_1^2, σ_2^2 τις πληθυσμιακές διακυμάνσεις τότε ισχύει ότι η ποσότητα:

$\dfrac{s_1^2}{s_2^2}\dfrac{\sigma_2^2}{\sigma_1^2}$ ακολουθεί την F κατανομή με n-1 και m-1 βαθμούς ελευθερίας

Οι πληθυσμοί είναι κανονικοί.

Ένα 100(1-α)% διάστημα εμπιστοσύνης για τον λόγο των διακυμάνσεων σ_1^2, σ_2^2 των δύο πληθυσμών είναι:

$$\frac{s_1^2}{s_2^2}\frac{1}{F_{n-1,m-1;\frac{a}{2}}} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{s_1^2}{s_2^2}F_{m-1,n-1;\frac{a}{2}}$$

Παράδειγμα:

Πήραμε δύο δείγματα αγροτών από δύο διοικητικές περιφέρειες της Ελλάδας και εξετάσαμε τα ετήσια εισοδήματά τους. Είχαμε τα ακόλουθα δεδομένα:

n= 31, S_1^2 = 200 ευρώ ^{2} και m =41, S_2^2 = 220 ευρώ ^{2}

Να βρεθεί ένα 98% διάστημα εμπιστοσύνης για τον λόγο των πραγματικών διακυμάνσεων σ_1^2, σ_2^2 των ετήσιων εισοδημάτων των αγροτικών πληθυσμών στις δύο διοικητικές περιφέρειες της Ελλάδας.

Έχουμε ότι:

$$\frac{s_1^2}{s_2^2}\frac{1}{F_{n-1,m-1;\frac{a}{2}}} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{s_1^2}{s_2^2}F_{m-1,n-1;\frac{a}{2}}$$

$F_{30,40;0,01} = 2,2$ $F_{40,30;0,01} = 2,3$

$$\frac{200}{220}\frac{1}{2,2} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{200}{220}2,3$$

ή

$$0,41 \le \frac{\sigma_1^2}{\sigma_2^2} \le 2,09$$

3.8. Εκτίμηση διαστήματος εμπιστοσύνης αναλογίας σε πληθυσμό.

Αν εκτιμήσουμε το ποσοστό (αναλογία) p = $\frac{x}{n}$ των στοιχείων στο **μεγάλο** δείγμα μεγέθους n που έχουν κάποιο συγκεκριμένο χαρακτηριστικό, τότε σύμφωνα με την

θεωρία το δειγματικό ποσοστό κατανέμεται κανονικά:

Δηλαδή η δειγματική μεταβλητή P_Δ, που παράγεται από τις τιμές p, ακολουθεί την κανονική κατανομή:

$$\mathrm{N}(\mathrm{P}_{\pi\lambda\eta\theta},\ \frac{p_{\pi\lambda\eta\theta}(1-P_{\pi\lambda\eta\theta})}{n})$$

Όπου $\mathrm{P}_{\pi\lambda\eta\theta}$ το πραγματικό ποσοστό στον πληθυσμό.

Συνεπώς έχουμε ότι η μεταβλητή:

Συνεπώς έχουμε ότι η μεταβλητή:

$$\frac{P_{\Delta}-P_{\pi\lambda\eta\theta}}{\sqrt{\dfrac{P_{\pi\lambda\eta\theta}(1-P_{\pi\lambda\eta\theta})}{n}}}$$ ακολουθεί την **N(0,1).**

Ένα 100(1-α)% διάστημα εμπιστοσύνης για το ποσοστό (αναλογία) του πληθυσμού είναι:

$$\rho_{\Delta}-z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{\rho_{\Delta}(1-\rho_{\Delta})}{n}}\leq P_{\pi\lambda\eta\theta}\leq\rho_{\Delta}+z_{\frac{\alpha}{2}}\sqrt{\frac{\rho_{\Delta}(1-\rho_{\Delta})}{n}}$$

Παράδειγμα:

Ο αριθμός των ελαττωματικών προϊόντων μιας μεταποιητικής βιομηχανίας αγροτικών προϊόντων που εντοπίσθηκαν σ' ένα δείγμα 300 προϊόντων είναι 24 προϊόντα.

Να βρεθούν τα όρια μέσα στα οποία θα βρίσκεται το πραγματικό ποσοστό Π των ελαττωματικών προϊόντων της συνολικής παραγωγής με πιθανότητα 98%.

Έχουμε n = 300 , m = 24 $\rightarrow$ $\widehat{\rho}=\dfrac{24}{300}=8\%=0{,}08.$

Οπότε $\widehat{q}$=0,92.

$$\left.\begin{aligned}\alpha&=1-0{,}98=0{,}02\\ \frac{\alpha}{2}&=0{,}01\end{aligned}\right\}\rightarrow z_{\frac{\alpha}{2}}=\phi^{-1}\left(1-\frac{\alpha}{2}\right)=\phi^{-1}(0{,}99)=2{,}33.$$

$$S\widehat{\rho}=\frac{\sqrt{\widehat{\rho}\cdot\widehat{q}}}{\sqrt{n}}=\frac{\sqrt{0{,}08\cdot0{,}92}}{\sqrt{300}}=0{,}0157$$

$$\rho-z_{\frac{a}{2}}\cdot S\widehat{\rho}<\mathrm{P}<\rho+z_{\frac{a}{2}}-S\widehat{\rho}$$

$$0{,}08 - 0{,}0157 \cdot 2{,}33 < \mathrm{P} < 0{,}08 + 0{,}0157 \cdot 2{,}33$$

$$0{,}0434 < \mathrm{P} < 0{,}1165 \Rightarrow 4{,}34\% < \mathrm{P} < 11{,}6\%.$$

3.9. Εκτίμηση διαστήματος εμπιστοσύνης για τη διαφορά των αναλογιών στοιχείων δύο πληθυσμών.

Έστω $\rho_{1\,\Delta} = \dfrac{x}{n}$ και $\rho_{2\,\Delta} = \dfrac{y}{l}$ είναι οι δειγματικές αναλογίες των στοιχείων, που έχουν κάποιο συγκεκριμένο χαρακτηριστικό, δύο μεγάλων δειγμάτων μεγέθους n, l από δύο πληθυσμούς, όπου οι πραγματικές αναλογίες είναι αντίστοιχα $\mathrm{P}_{1\ \pi\lambda\eta\theta}$, $\mathrm{P}_{2\ \pi\lambda\eta\theta}$.

Τότε έχουμε ότι οι διαφορές των δειγματικών μεταβλητών $\mathrm{P}_{1\,\Delta}$ - $\mathrm{P}_{2\,\Delta}$ ακολουθούν κανονική κατανομή:

$$\mathrm{N}(\mathrm{P}_{1\ \pi\lambda\eta\theta} - \mathrm{P}_{2\ \pi\lambda\eta\theta},\ \frac{P_{1\pi\lambda\eta\theta}(1-P_{1\pi\lambda\eta\theta})}{n} + \frac{P_{2\pi\lambda\eta\theta}(1-P_{2\pi\lambda\eta\theta})}{l})$$

Ένα 100(1-α)% διάστημα εμπιστοσύνης για τη διαφορά των ποσοστών (αναλογιών) των δύο πληθυσμών $\mathrm{P}_{1\ \pi\lambda\eta\theta}$, $\mathrm{P}_{2\ \pi\lambda\eta\theta}$ είναι:

$$\rho_{1\,\Delta} - \rho_{2\,\Delta} - Z_{\frac{a}{2}} \sqrt{\frac{\rho_{1\Delta}(1-\rho_{1\Delta})}{n} + \frac{\rho_{2\Delta}(1-\rho_{2\Delta})}{l}} \le$$

$$\mathbf{P}_{1\ \pi\lambda\eta\theta} - \mathbf{P}_{2\ \pi\lambda\eta\theta} \le \rho_{1\,\Delta} - \rho_{2\,\Delta} + Z_{\frac{a}{2}} \sqrt{\frac{\rho_{1\Delta}(1-\rho_{1\Delta})}{n} + \frac{\rho_{2\Delta}(1-\rho_{2\Delta})}{l}}$$

Παράδειγμα:

Παίρνουμε δύο μεγάλα δείγματα μιας ποικιλίας ενός φυτού από δύο διαφορετικούς πληθυσμούς και εξετάζουμε πόσα από αυτά ασθένησαν μέσα σε ένα συγκεκριμένο χρονικό διάστημα. Τα δεδομένα που προέκυψαν ήταν:

Από τα n=120 του πρώτου δείγματος ασθένησαν τα 12 και από τα m=130 του δεύτερου δείγματος ασθένησαν τα 18. Να βρεθεί ένα 95% διάστημα εμπιστοσύνης για τη διαφορά των πραγματικών ποσοστών (αναλογιών) των ασθενούντων φυτών των δύο πληθυσμών $\mathrm{P}_{1\ \pi\lambda\eta\theta}$, $\mathrm{P}_{2\ \pi\lambda\eta\theta}$.

Έχουμε ότι:

$$\mathrm{p}_1 - \mathrm{p}_2 - Z_{\frac{a}{2}} \sqrt{\frac{\rho_{1\Delta}(1-\rho_{1\Delta})}{n} + \frac{\rho_{2\Delta}(1-\rho_{2\Delta})}{l}} \le$$

$$\mathbf{P}_{1\ \pi\lambda\eta\theta} - \mathbf{P}_{2\ \pi\lambda\eta\theta} \leq p_1 - p_2 + Z_{\frac{a}{2}} \sqrt{\frac{\rho_{1\Delta}(1-\rho_{1\Delta})}{n} + \frac{\rho_{2\Delta}(1-\rho_{2\Delta})}{l}}$$

$\rho_{1\ \Delta} = 12/120 = 0{,}1$, $\rho_{2\ \Delta} = 18/130 = 0{,}14$

Έχουμε: $\rho_{1\ \Delta} - \rho_{2\ \Delta} = -0{,}04$

$Z_{0,025} = 1{,}96$

$$\sqrt{\frac{0{,}1x0{,}9}{120} + \frac{0{,}14x0{,}86}{130}} = 0{,}04$$

Άρα ένα 95% διάστημα εμπιστοσύνης για τη διαφορά των πραγματικών ποσοστών (αναλογιών) των ασθενούντων φυτών των δύο πληθυσμών $P_{1\ \pi\lambda\eta\theta}$, $P_{2\ \pi\lambda\eta\theta}$ είναι:

$$-0{,}04 - 1{,}96 \text{ x } 0{,}04 \leq \mathbf{P}_{1\ \pi\lambda\eta\theta} - \mathbf{P}_{2\ \pi\lambda\eta\theta} \leq -0{,}04 + 1{,}96 \text{ x } 0{,}04$$

ή

$$-0{,}118 \leq \mathbf{P}_{1\ \pi\lambda\eta\theta} - \mathbf{P}_{2\ \pi\lambda\eta\theta} \leq 0{,}038$$

Κεφάλαιο 4ο

Έλεγχοι υποθέσεων

4.1 . Γενικά

Εκτός του προσδιορισμού του διαστήματος εμπιστοσύνης μιας αγνώστου παραμέτρου Θ του πληθυσμού πολλές φορές απαιτείται να κάνουμε υποθέσεις για την τιμή που μπορεί να πάρει η Θ, τις οποίες και ελέγχουμε.

Σημαντικό ρόλο στον έλεγχο της υπόθεσης που κάνουμε για την άγνωστη παράμετρο Θ του πληθυσμού παίζει η **εκτιμήτρια θ** και το **στατιστικό του ελέγχου** από το δείγμα.

Καταρχήν η υπόθεση που διατυπώνουμε για την άγνωστη παράμετρο Θ του πληθυσμού καλείται H_0 και είναι ης μορφής:

$H_0 : \Theta = \theta^*$ όπου θ^* είναι μια συγκεκριμένη τιμή που υποθέτουμε ότι μπορεί να ην πάρει η Θ.

Η υπόθεση αυτή ελέγχεται αν ισχύει η όχι και καλείται **μηδενική υπόθεση**.

Οι εναλλακτικές υποθέσεις είναι τις μορφής $H_1 : \Theta \neq \theta^*$, $H_1 : \Theta > \theta^*$, $H_1 : \Theta < \theta^*$

Έτσι διαμορφώνονται οι ακόλουθες υποθέσεις προς έλεγχο:

$H_0 : \Theta = \theta^*$

$H_1 : \Theta \neq \theta^*$

έχουμε τότε δίπλευρο έλεγχο.

$H_0 : \Theta = \theta^*$

$H_1 : \Theta > \theta^*$

έχουμε τότε μονόπλευρο έλεγχο (δεξιά)

$H_0 : \Theta = \theta^*$

$H_1 : \Theta < \theta^*$

έχουμε τότε μονόπλευρο έλεγχο (αριστερά)

Για τον έλεγχο υπολογίζεται η **απορριπτική περιοχή R** της H_0 ,δηλαδή η περιοχή στα σημεία τη οποίας η H_0 απορρίπτεται. Αυτή προσδιορίζεται από την κατανομή που ακολουθεί το στατιστικό του ελέγχου και το **σφάλμα α** που λαμβάνεται υπόψη.

Συνεπώς τα στοιχεία ενός ελέγχου μηδενικής υπόθεσης είναι τα ακόλουθα:

1. Ορισμός της μηδενικής υπόθεσης
2. Ορισμός της εναλλακτικής υπόθεσης
3. Ορισμός του στατιστικού του ελέγχου από το δείγμα
4. Ορισμός της απορριπτικής περιοχής R της H_0
5. Εξαγωγή συμπερασμάτων.

4.2 . Σφάλματα – στάθμη σημαντικότητας – περιοχή απόρριψης της H_0

Το α είναι η πιθανότητα να απορρίψουμε την H_0 ενώ είναι σωστή:

α = P(απόρριψη της H_0 / H_0 σωστή)

Το α καλείται και **σφάλμα τύπου Ι.**

Το β είναι η πιθανότητα να δεχτούμε την H_0 ενώ είναι λάθος:

β = P(αποδοχή της H_0 / H_0 λάθος)

Το β καλείται και **σφάλμα τύπου ΙΙ.**

Το γ = 1 – β και εκφράζει την πιθανότητα απόρριψης της H_0 όταν η H_0 είναι πράγματι λάθος.

Το γ καλείται και **ισχύς** του στατιστικού του ελέγχου**.**

Η απορριπτική περιοχή της R της H_0 ορίζεται βάσει του σφάλματος α που καλείται **στάθμη σημαντικότητας ή επίπεδο σημαντικότητας (σ.σ).**

Συγκεκριμένα επίπεδο σημαντικότητας α ενός ελέγχου μηδενικής υπόθεσης H_0 ονομάζουμε την πιθανότητα να παρατηρηθεί μια τιμή του στατιστικού του ελέγχου μεγαλύτερη από αυτή που έδωσε το δείγμα.

Δηλαδή η πιθανότητα $P(Y>|y|/\ H_0$ σωστή), όπου Y η τ.μ που αντιστοιχεί στο στατιστικό και y η τιμή του στατιστικού από το συγκεκριμένο δείγμα.

Η πιθανότητα αυτή αναφέρεται σε μονόπλευρους ελέγχους ενώ σε δίπλευρους ελέγχους η πιθανότητα διπλασιάζεται.

Η υπόθεση H_0 απορρίπτεται εάν η παρατηρούμενη πιθανότητα α = P(απόρριψη της H_0 / H_0 σωστή) είναι μικρότερη μιας ορισμένης στάθμης σημαντικότητας που επιλέγεται από αυτόν που βγάζει τα στατιστικά συμπεράσματα.

Πως ορίζεται το στατιστικό και η απορριπτική περιοχή R;

Εάν η εκτιμήτρια θ ακολουθεί κανονική κατανομή ή προσεγγιστικά κανονική κατανομή τότε βάσει της θεωρίας η μεταβλητή:

$\dfrac{|\theta-\theta^*|}{\tau}$ ακολουθεί τυπική κανονική κατανομή.

Όπου τ το τυπικό σφάλμα (τυπική απόκλιση) της κατανομής της εκτιμήτριας θ.

Η υπόθεση H_0 απορρίπτεται όταν $\frac{\left|\theta-\theta^*\right|}{\tau} > Z_{\frac{\alpha}{2}}$ με α το επίπεδο σημαντικότητας του ελέγχου.

Γενικά όταν το δείγμα προέρχεται από **κανονικό πληθυσμό** με την προϋπόθεση ότι ισχύει η H_0 η ποσότητα $\frac{\theta-\theta^*}{\tau}$ ακολουθεί γνωστή κατανομή και η περιοχή απόρριψης της H_0 είναι εκεί όπου:

$$\frac{\theta-\theta^*}{\tau} > \Phi_{\alpha}, \quad \frac{\theta-\theta^*}{\tau} < -\Phi_{\alpha} \quad \text{ή} \quad \left|\frac{\theta-\theta^*}{\tau}\right| > \Phi_{\frac{\alpha}{2}}$$

όταν οι εναλλακτικές υποθέσεις είναι αντίστοιχα:

$H_1: \Theta > \theta^*$, $H_1: \Theta < \theta^*$ ή $H_1: \Theta \neq \theta^*$

Οι Φ_{α} , $\Phi_{\frac{\alpha}{2}}$ είναι τιμές της κατανομής που ακολουθεί η ποσότητα $\frac{\theta-\theta^*}{\tau}$

ώστε $P(\frac{\theta-\theta^*}{\tau} > \Phi_{\alpha}) = \alpha$, $P(\frac{\theta-\theta^*}{\tau} < -\Phi_{\alpha}) = \alpha$ και $P(\left|\frac{\theta-\theta^*}{\tau}\right| > \Phi_{\frac{\alpha}{2}}) = \alpha$

4.3. Έλεγχοι υποθέσεων

4.3.1. Έλεγχος υπόθεσης για τη μέση τιμή του πληθυσμού

Ι. Έλεγχος υπόθεσης για τη μέση τιμή του πληθυσμού (όταν n≥30 και η διακύμανση του πληθυσμού να είναι γνωστή ή άγνωστη).

Οι έλεγχοι υποθέσεων στην προκειμένη περίπτωση παίρνουν τις ακόλουθες μορφές:

$H_0: \mu = \mu^*$
$H_1: \mu \neq \mu^*$
δίπλευρος έλεγχος.

$H_0: \mu = \mu^*$
$H_1: \mu > \mu^*$
μονόπλευρος έλεγχος (δεξιά)

$H_0 : \mu = \mu^*$
$H_1 : \mu < \mu^*$
μονόπλευρος έλεγχος (αριστερά)

Ο σημειακός εκτιμητής του μ είναι το $\bar{x}$. Όταν το δείγμα προέρχεται από κανονικό πληθυσμό N(μ, σ^2) ανεξάρτητα από το μέγεθός του έχουμε ότι η $\bar{X}$ ακολουθεί κανονική κατανομή N(μ, $\frac{\sigma^2}{n}$), το ίδιο ισχύει και στην περίπτωση που το δείγμα δεν προέρχεται από κανονικό πληθυσμό αλλά το μέγεθός του είναι $\mathbf{n \geq 30}$.

Το στατιστικό για τον έλεγχο είναι το:

$$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \quad \text{ή} \quad \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}$$

Στην πράξη παίρνουμε:

$$\frac{\bar{X}-\mu^*}{\frac{\sigma}{\sqrt{n}}} \quad \text{ή} \quad \frac{\bar{X}-\mu^*}{\frac{s}{\sqrt{n}}}$$

Οι απορριπτικές περιοχές για τους προαναφερόμενους ελέγχους είναι αντίστοιχα:

Για τον δίπλευρο έλεγχο R = $\{|z| > z_{\frac{a}{2}}\}$, για τον μονόπλευρο έλεγχο (δεξιά) R = $\{z > z_a\}$ και για τον μονόπλευρο έλεγχο (αριστερά) R = $\{z < - z_a\}$

ΙΙ. Έλεγχος υπόθεσης για τη μέση τιμή του πληθυσμού (όταν n<30 και η διακύμανση του πληθυσμού να είναι άγνωστη).

Όταν το μέγεθος του δείγματος είναι μικρό, η διακύμανση του δείγματος άγνωστη και ο πληθυσμός από όπου προέρχεται το δείγμα κανονικός τότε η ποσότητα:

$\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}$ ακολουθεί την t κατανομή με n-1 βαθμούς ελευθερίας.

Η μεταβλητή αυτή παίρνεται ως το στατιστικό ελέγχου.

Οι έλεγχοι υποθέσεων:

$H_0: \mu = \mu^*$
$H_1: \mu \neq \mu^*$
δίπλευρος έλεγχος.

$H_0: \mu = \mu^*$
$H_1: \mu > \mu^*$
μονόπλευρος έλεγχος (δεξιά)

$H_0: \mu = \mu^*$
$H_1: \mu < \mu^*$
μονόπλευρος έλεγχος (αριστερά)

έχουν ως απορριπτικές περιοχές αντίστοιχα:

ο δίπλευρος έλεγχος R = $\{|t| > t_{n-1\,;\,\frac{a}{2}}\}$, ο μονόπλευρος έλεγχος (δεξιά) R = $\{t > t_{n-1;\,a}\}$και ο μονόπλευρος έλεγχος (αριστερά) R = $\{t < - t_{n-1;\,a}\}$

Παραδείγματα:

1. Παίρνουμε ένα δείγμα ημερήσιων αμοιβών n=40 εργατών με μέση ημερήσια αμοιβή $\bar{x}$ = 28 ευρώ. Θεωρούμε ότι οι ημερήσιες αμοιβές των εργατών κατανέμονται κανονικώς $N(29, 4^2)$. Μπορούμε να πούμε ότι η μέση ημερήσια αμοιβή στον πληθυσμό είναι μικρότερη του 29.

Έχουμε τον έλεγχο της μηδενικής υπόθεσης: (α = 0,05)

$H_0: \mu = 29$
$H_1: \mu < 29$
μονόπλευρος έλεγχος (αριστερά)

Η απορριπτική περιοχή είναι:

$R = \{z < -z_a\}$ με $Z = \dfrac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{28 - 29}{\frac{4}{\sqrt{40}}} =$ -1,58

$z_a = z_{0,05} = 1{,}64$

Συνεπώς δεν απορρίπτουμε την H_0

2. Παίρνουμε ένα δείγμα ημερήσιων αμοιβών n = 50 εργατών με μέση ημερήσια αμοιβή $\bar{x}$ = 38 ευρώ και δειγματική διακύμανση 6^2. Μπορούμε να πούμε ότι η μέση ημερήσια αμοιβή στον πληθυσμό είναι μεγαλύτερη του 39.

Έχουμε τον έλεγχο της μηδενικής υπόθεσης: (α=0,05)
H_0: μ = 39
H_1: μ > 39
μονόπλευρος έλεγχος (δεξιά)

Η απορριπτική περιοχή είναι:

$R = \{z > z_a\}$ με $Z = \dfrac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{38 - 39}{\frac{6}{\sqrt{50}}} =$ -1,18

$z_a = z_{0,05} = 1{,}64$

Συνεπώς δεν απορρίπτουμε την H_0

3. Παίρνουμε ένα δείγμα ημερήσιων αμοιβών n=20 εργατών με μέση ημερήσια αμοιβή $\bar{x}$ = 33 ευρώ και δειγματική διακύμανση 6^2. Μπορούμε να πούμε ότι η μέση ημερήσια αμοιβή στον πληθυσμό διαφέρει του 30.

Έχουμε τον έλεγχο της μηδενικής υπόθεσης: (α=0,05)
H_0: μ = 30
H_1: μ $\neq$ 30
δίπλευρος έλεγχος

Η απορριπτική περιοχή είναι:

$R = \{ |t| > t_{n-1\,;\,\frac{a}{2}} \}$, με $t = \dfrac{\bar{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}} = \dfrac{33 - 30}{\dfrac{6}{\sqrt{20}}} = 2{,}24$

$t_{n-1\,;\,\frac{a}{2}} = t_{19;0,025} = 2{,}093$

Συνεπώς απορρίπτουμε την H_0

4.3.2. Έλεγχος υπόθεσης για τη διαφορά των μέσων τιμών μ_1, μ_2 δύο πληθυσμών

Ι. Έλεγχος υπόθεσης για τη διαφορά των μέσων τιμών δύο πληθυσμών (δείγματα μεγάλα ανεξάρτητα, διακυμάνσεις γνωστές ή άγνωστες)

Από τις κατανομές των στατιστικών ενός δείγματος γνωρίζουμε ότι:

Αν πάρουμε δύο δείγματα μεγάλα μεγέθους ν και κ με μέσες τιμές $\bar{x}$ και $\bar{y}$ και διακυμάνσεις s_1^2, s_2^2 αντίστοιχα τότε έχουμε ότι η μεταβλητή:

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{\nu} + \dfrac{\sigma_2^2}{\kappa}}} \text{ ακολουθεί } N(0,1).$$

Εφόσον οι διακυμάνσεις είναι άγνωστες αντικαθίστανται από τις δειγματικές διακυμάνσεις s_1^2, s_2^2 και έχουμε ότι η μεταβλητή:

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{\nu} + \dfrac{s_2^2}{\kappa}}} \text{ ακολουθεί } N(0,1).$$

Οι έλεγχοι υπόθεσης για τη <u>διαφορά των μέσων τιμών δύο πληθυσμών</u> **(δείγματα μεγάλα ανεξάρτητα, διακυμάνσεις γνωστές ή άγνωστες)** είναι οι ακόλουθοι:

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

δίπλευρος έλεγχος.

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 > \mu_2$

μονόπλευρος έλεγχος (δεξιά)

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 < \mu_2$
μονόπλευρος έλεγχος (αριστερά)

Χρησιμοποιείται ως στατιστικό ελέγχου το:

$$\frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_1^2}{\nu}+\frac{\sigma_2^2}{\kappa}}} \quad \acute{\eta} \quad \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_1^2}{\nu}+\frac{s_2^2}{\kappa}}}$$

και έχουμε ως απορριπτικές περιοχές αντίστοιχα:
Για τον δίπλευρο έλεγχο R = $\{|z| > z_{\frac{a}{2}}\}$, για τον μονόπλευρο έλεγχο (δεξιά) R = $\{z > z_a\}$ και για τον μονόπλευρο έλεγχο (αριστερά) R = $\{z < - z_a\}$

Παραδείγματα:

1. Από ένα πληθυσμό ημερήσιων αμοιβών εργατών παίρνουμε ένα δείγμα n=40 ημερήσιων αμοιβών εργατών με μέση ημερήσια αμοιβή $\bar{x}$ = 33 ευρώ και δειγματική διακύμανση 6^2 και από ένα δεύτερο πληθυσμό ημερήσιων αμοιβών εργατών παίρνουμε ένα δεύτερο δείγμα n=50 ημερήσιων αμοιβών εργατών με μέση ημερήσια αμοιβή $\bar{y}$ = 30 ευρώ και δειγματική διακύμανση 5^2. Μπορούμε να πούμε ότι η μέση ημερήσια αμοιβή μ_1 στον πρώτο πληθυσμό διαφέρει της μέσης ημερήσιας αμοιβής μ_2 στον δεύτερο πληθυσμό σε επίπεδο σημαντικότητας (α=0,05)

Ο έλεγχος μηδενικής υπόθεσης για τη <u>διαφορά των μέσων τιμών δύο πληθυσμών</u> είναι ο ακόλουθος:

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$
δίπλευρος έλεγχος.
Η απορριπτική περιοχή είναι:
R = $\{|z| > z_{\frac{a}{2}}\}$,
Για τον έλεγχο της μηδενικής υπόθεσης χρησιμοποιούμε το στατιστικό

$$Z = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_1^2}{\nu}+\frac{s_2^2}{\kappa}}} = \frac{33-30}{\sqrt{\frac{36}{40}+\frac{25}{50}}} = 2{,}54$$

Έχουμε ότι: $z_{\frac{a}{2}} = 1{,}96$

Επειδή $Z > z_{\frac{a}{2}}$ απορρίπτεται η H_0. Άρα σε επίπεδο σημαντικότητας α=5% δεν δεχόμαστε ότι $\mu_1 = \mu_2$.

2. Από ένα πληθυσμό ημερήσιων αμοιβών εργατών παίρνουμε ένα δείγμα n=38 ημερήσιων αμοιβών εργατών με μέση ημερήσια αμοιβή $\bar{x}$ = 37 ευρώ. Από ένα δεύτερο πληθυσμό ημερήσιων αμοιβών εργατών παίρνουμε ένα δεύτερο δείγμα n=45 ημερήσιων αμοιβών εργατών με μέση ημερήσια αμοιβή $\bar{y}$ = 34 ευρώ. Γνωρίζουμε τις πληθυσμιακές διακυμάνσεις των ημερήσιων αμοιβών ότι είναι 4^2 και 5^2 αντίστοιχα. Μπορούμε να πούμε ότι η μέση ημερήσια αμοιβή μ_1 στον πρώτο πληθυσμό διαφέρει της μέσης ημερήσιας αμοιβής μ_2 στον δεύτερο πληθυσμό σε επίπεδο σημαντικότητας (α=0,05)

Ο έλεγχος μηδενικής υπόθεσης για τη διαφορά των μέσων τιμών δύο πληθυσμών είναι ο ακόλουθος:

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$
δίπλευρος έλεγχος.
Η απορριπτική περιοχή είναι:
$R = \{ |z| > z_{\frac{a}{2}} \}$,
Για τον έλεγχο της μηδενικής υπόθεσης χρησιμοποιούμε το στατιστικό

$$Z = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_1^2}{\nu}+\frac{\sigma_2^2}{\kappa}}} = \frac{37-34}{\sqrt{\frac{16}{38}+\frac{25}{45}}} = 3{,}03$$

Έχουμε ότι: $z_{\frac{a}{2}} = 1{,}96$

Επειδή Z> $z_{\frac{a}{2}}$ απορρίπτεται η H_0. Άρα σε επίπεδο σημαντικότητας α=5% δεν δεχόμαστε ότι $\mu_1 = \mu_2$.

ΙΙ. Έλεγχος υπόθεσης για τη διαφορά των μέσων τιμών δύο πληθυσμών (δείγματα μικρά ανεξάρτητα, διακυμάνσεις άγνωστες και ίσες)

Από τις κατανομές των στατιστικών ενός δείγματος γνωρίζουμε ότι:

Αν πάρουμε δύο δείγματα μικρά ανεξάρτητα μεγέθους ν και κ με μέσες τιμές $\bar{x}$ και $\bar{y}$ και διακυμάνσεις s_1^2, s_2^2 αντίστοιχα προερχόμενα από κανονικούς πληθυσμούς με άγνωστες αλλά ίσες διακυμάνσεις $\sigma_1^2 = \sigma_2^2 = \sigma^2$ τότε έχουμε ότι η μεταβλητή:

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{(\nu-1)s_1^2 + (\kappa-1)s_2^2}{\nu+\kappa-2}}\sqrt{\frac{1}{\nu}+\frac{1}{\kappa}}}$$ ακολουθεί t-κατανομή με ν+κ-2 βαθμούς ελευθερίας

Οι έλεγχοι υπόθεσης για τη <u>διαφορά των μέσων τιμών δύο πληθυσμών</u> (δείγματα μικρά ανεξάρτητα, διακυμάνσεις άγνωστες και ίσες) είναι οι ακόλουθοι:

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$
δίπλευρος έλεγχος.

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 > \mu_2$
μονόπλευρος έλεγχος (δεξιά)

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 < \mu_2$
μονόπλευρος έλεγχος (αριστερά)

Χρησιμοποιείται ως στατιστικό ελέγχου το:

$$\frac{\bar{x}-\bar{y}}{\sqrt{\frac{(\nu-1)s_1^2+(\kappa-1)s_2^2}{\nu+\kappa-2}}\sqrt{\frac{1}{\nu}+\frac{1}{\kappa}}}$$

και έχουμε ως απορριπτικές περιοχές αντίστοιχα:

ο δίπλευρος έλεγχος R = $\{|t| > t_{\nu+\kappa-2\,;\,\frac{a}{2}}\}$, ο μονόπλευρος έλεγχος (δεξιά) R = $\{t > t_{\nu+\kappa-2;\,a}\}$ και ο μονόπλευρος έλεγχος (αριστερά) R = $\{t < -\,t_{\nu+\kappa-2;\,a}\}$

Παράδειγμα:

Από ένα πληθυσμό στρεμματικών αποδόσεων μιας καλλιέργειας σε μια περιοχή παίρνουμε ένα δείγμα ν =16 στρεμματικών αποδόσεων με μέση τιμή $\bar{x}$ = 370κιλά/στρεμ. Από ένα δεύτερο πληθυσμό στρεμματικών αποδόσεων της ίδιας καλλιέργειας σε μια άλλη περιοχή παίρνουμε ένα δεύτερο δείγμα κ = 15 στρεμματικών αποδόσεων με μέση τιμή $\bar{y}$ = 340 κιλά/στρεμ. Οι δειγματικές διακυμάνσεις είναι αντίστοιχα 41^2 και 52^2.

Μπορούμε να πούμε ότι η μέση στρεμματική απόδοση μ_1 στον πρώτο πληθυσμό διαφέρει της μέσης στρεμματική απόδοση μ_2 στον δεύτερο πληθυσμό σε επίπεδο σημαντικότητας (α = 0,05). Θεωρούμε ότι οι πληθυσμιακές διακυμάνσεις είναι άγνωστες και ίσες.

Ο έλεγχος μηδενικής υπόθεσης για τη <u>διαφορά των μέσων τιμών δύο πληθυσμών</u> είναι ο ακόλουθος:

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

δίπλευρος έλεγχος.

Χρησιμοποιείται το στατιστικό:

$$t=\frac{\bar{x}-\bar{y}}{\sqrt{\frac{(\nu-1)s_1^2+(\kappa-1)s_2^2}{\nu+\kappa-2}}\sqrt{\frac{1}{\nu}+\frac{1}{\kappa}}}=\frac{370-340}{\sqrt{\frac{(16-1)1681+(15-1)2704}{29}}\sqrt{\frac{1}{16}+\frac{1}{15}}}=0{,}34$$

η απορριπτική περιοχή είναι:

$R = \{ |t| > t_{\nu+\kappa-2\ ;\ \frac{a}{2}} \}$, $t_{\nu+\kappa-2\ ;\ \frac{a}{2}} = t_{29\ ;\ 0,025} = 2{,}045$

Εφόσον $t < t_{\nu+\kappa-2\ ;\ \frac{a}{2}}$ δεν απορρίπτεται η H_0. Άρα σε επίπεδο σημαντικότητας α = 5% δεχόμαστε ότι $\mu_1 = \mu_2$ και δεν μπορούμε να ισχυριστούμε ότι διαφέρουν μεταξύ τους.

III. Έλεγχος υπόθεσης για τη διαφορά των μέσων τιμών δύο πληθυσμών (δείγματα μικρά ανεξάρτητα, διακυμάνσεις άγνωστες και διαφορετικές)

Από τις κατανομές των στατιστικών ενός δείγματος γνωρίζουμε ότι:

Αν πάρουμε δύο δείγματα μικρά ανεξάρτητα μεγέθους ν και κ με μέσες τιμές $\bar{x}$ και $\bar{y}$ και διακυμάνσεις s_1^2, s_2^2 αντίστοιχα προερχόμενα από κανονικούς πληθυσμούς με άγνωστες και διαφορετικές διακυμάνσεις $\sigma_1^2 \neq \sigma_2^2$ τότε έχουμε ότι η μεταβλητή:

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{\nu} + \frac{s_2^2}{\kappa}}}$$

ακολουθεί t – κατανομή με βαθμούς ελευθερίας λ = 2(ν-1) όταν ν = κ

και $\lambda = \dfrac{(\frac{s_1^2}{\nu} + \frac{s_2^2}{\kappa})^2}{\frac{(\frac{s_1^2}{\nu})^2}{\nu - 1} + \frac{(\frac{s_2^2}{\kappa})^2}{\kappa - 1}}$ (στρογγυλεμένο στον πλησιέστερο ακέραιο) όταν $\nu \neq \kappa$

Οι έλεγχοι υπόθεσης για τη <u>διαφορά των μέσων τιμών δύο πληθυσμών</u> (δείγματα μικρά ανεξάρτητα, διακυμάνσεις άγνωστες και διαφορετικές) είναι οι ακόλουθοι:

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$
δίπλευρος έλεγχος.

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 > \mu_2$
μονόπλευρος έλεγχος (δεξιά)

$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 < \mu_2$
μονόπλευρος έλεγχος (αριστερά)
χρησιμοποιούν ως στατιστικό ελέγχου το:

$$\frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{\nu} + \frac{s_2^2}{\kappa}}}$$

και έχουν ως απορριπτικές περιοχές αντίστοιχα:

ο δίπλευρος έλεγχος R = $\{|t| > t_{\lambda\,;\,\frac{a}{2}}\}$, ο μονόπλευρος έλεγχος (δεξιά) R = $\{t > t_{\lambda;\,a}\}$και

ο μονόπλευρος έλεγχος (αριστερά) R = $\{t < - t_{\lambda;\,a}\}$

Παραδείγματα:

1. Από ένα πληθυσμό στρεμματικών αποδόσεων μιας καλλιέργειας σε μια περιοχή παίρνουμε ένα δείγμα ν=16 στρεμματικών αποδόσεων με μέση τιμή $\bar{x}$ = 370κιλά/στρεμ. Από ένα δεύτερο πληθυσμό στρεμματικών αποδόσεων της ίδιας καλλιέργειας σε μια άλλη περιοχή παίρνουμε ένα δεύτερο δείγμα κ=16 στρεμματικών αποδόσεων με μέση τιμή $\bar{y}$ = 340 κιλά/στρεμ. Οι δειγματικές διακυμάνσεις είναι αντίστοιχα 41^2 και 52^2.

Μπορούμε να πούμε ότι η μέση στρεμματική απόδοση μ_1 στον πρώτο πληθυσμό διαφέρει της μέσης στρεμματική απόδοση μ_2 στον δεύτερο πληθυσμό σε επίπεδο σημαντικότητας (α=0,05). Θεωρούμε ότι οι πληθυσμιακές διακυμάνσεις είναι άγνωστες και διαφορετικές.

Ο έλεγχος μηδενικής υπόθεσης για τη <u>διαφορά των μέσων τιμών δύο πληθυσμών</u> είναι ο ακόλουθος:

$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$
δίπλευρος έλεγχος.
Χρησιμοποιείται το στατιστικό:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{\nu} + \frac{s_2^2}{\kappa}}} = \frac{370 - 340}{\sqrt{\frac{41^2}{16} + \frac{52^2}{16}}} = 1{,}81$$

η απορριπτική περιοχή είναι:

$R = \{ |t| > t_{\lambda\,;\,\frac{a}{2}} \}$, $t_{\lambda\,;\,\frac{a}{2}} = t_{30\,;\,0{,}025} = 2{,}042$
$\lambda = 2(\nu\text{-}1) = 30$

Εφόσον $t < t_{\lambda\,;\,\frac{a}{2}}$ δεν απορρίπτεται η H_0. Άρα σε επίπεδο σημαντικότητας α=5% δεχόμαστε ότι $\mu_1 = \mu_2$ και δεν μπορούμε να ισχυριστούμε ότι διαφέρουν μεταξύ τους.

2. Από ένα πληθυσμό στρεμματικών αποδόσεων μιας καλλιέργειας σε μια περιοχή παίρνουμε ένα δείγμα ν=10 στρεμματικών αποδόσεων με μέση τιμή $\bar{x}$ = 370κιλά/στρεμ. Από ένα δεύτερο πληθυσμό στρεμματικών αποδόσεων της ίδιας καλλιέργειας σε μια άλλη περιοχή παίρνουμε ένα δεύτερο δείγμα κ=16 στρεμματικών αποδόσεων με μέση τιμή $\bar{y}$ = 340 κιλά/στρεμ. Οι δειγματικές διακυμάνσεις είναι αντίστοιχα 41^2 και 52^2.

Μπορούμε να πούμε ότι η μέση στρεμματική απόδοση μ_1 στον πρώτο πληθυσμό διαφέρει της μέσης στρεμματική απόδοση μ_2 στον δεύτερο πληθυσμό σε επίπεδο σημαντικότητας (α=0,05). Θεωρούμε ότι οι πληθυσμιακές διακυμάνσεις είναι άγνωστες και διαφορετικές.

Ο έλεγχος μηδενικής υπόθεσης για τη <u>διαφορά των μέσων τιμών δύο πληθυσμών</u> είναι ο ακόλουθος:

$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$
δίπλευρος έλεγχος.
Χρησιμοποιείται το στατιστικό:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{\nu} + \frac{s_2^2}{\kappa}}} = \frac{370 - 340}{\sqrt{\frac{41^2}{10} + \frac{52^2}{16}}} = 1{,}63$$

η απορριπτική περιοχή είναι:

$$R = \{ |t| > t_{\lambda\ ;\ \frac{a}{2}} \},\ t_{\lambda\ ;\ \frac{a}{2}} = t_{23\ ;\ 0,025} = 2,069$$

$$\lambda = \frac{\left(\frac{s_1^2}{\nu} + \frac{s_2^2}{\kappa}\right)^2}{\frac{\left(\frac{s_1^2}{\nu}\right)^2}{\nu - 1} + \frac{\left(\frac{s_2^2}{\kappa}\right)^2}{\kappa - 1}} = 23$$

Εφόσον $t < t_{\lambda\ ;\ \frac{a}{2}}$ δεν απορρίπτεται η H_0. Άρα σε επίπεδο σημαντικότητας α = 5% δεχόμαστε ότι $\mu_1 = \mu_2$ και δεν μπορούμε να ισχυριστούμε ότι διαφέρουν μεταξύ τους.

IV. Έλεγχος υπόθεσης για τη διαφορά των μέσων τιμών δύο πληθυσμών (δείγματα μικρά εξαρτημένα)

Στην περίπτωση που έχουμε **δείγματα μικρά εξαρτημένα** που προέρχονται από μετρήσεις της ίδιας ομάδας σε δυο διαφορετικές χρονικές στιγμές (ζευγαρωτές παρατηρήσεις) ορίσουμε x_i και y_i, i=1,2,….,n τις παρατηρήσεις στα δύο δείγματα και δημιουργούμε τις αντίστοιχες διαφορές $z_i = x_i - y_i$, i=1,2,….,n.

Οι παρατηρήσεις z_i ακολουθούν την t – κατανομή με n -1 βαθμούς ελευθερίας.

Ως δύο πληθυσμοί θεωρούνται

Οι έλεγχοι υπόθεσης για τη διαφορά των μέσων τιμών δύο πληθυσμών **(δείγματα μικρά εξαρτημένα)** είναι οι ακόλουθοι:

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$
δίπλευρος έλεγχος.

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 > \mu_2$
μονόπλευρος έλεγχος (δεξιά)

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 < \mu_2$
μονόπλευρος έλεγχος (αριστερά)
χρησιμοποιούν ως στατιστικό ελέγχου το:

$$\frac{\bar{Z}}{\frac{s_Z}{\sqrt{n}}}$$ που ακολουθεί t – κατανομή με n -1 βαθμούς ελευθερίας.

και έχουν ως απορριπτικές περιοχές αντίστοιχα:

ο δίπλευρος έλεγχος R = $\{|t| > t_{n-1\,;\,\frac{a}{2}}\}$, ο μονόπλευρος έλεγχος (δεξιά) R = $\{t > t_{n-1;\,a}\}$ και ο μονόπλευρος έλεγχος (αριστερά) R = $\{t < -\,t_{n-1;\,a}\}$

Παράδειγμα:

Έχουμε τις παρακάτω ζευγαρωτές παρατηρήσεις:

X:	4	5	6	4,2	5,2	5,3	6,4	4,8	5,3	5
Y:	4,9	4,8	5,7	5	6	5.2	6,5	5,9	4,8	5,7

Να ελεγχθεί η μηδενική υπόθεση:

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 < \mu_2$
μονόπλευρος έλεγχος (αριστερά), σε επίπεδο σημαντικότητας α=0,05

Έχουμε Z: -0,9 0,2 0,3 -0,8 -0,8 0,1 -0,1 -1,1 0,5 -0,7

και $\bar{Z}$ = -0,33 , S_Z = 0,72 , $t_{9;0,05}$ = 1,833

$$t = \frac{\bar{Z}}{\frac{s_Z}{\sqrt{n}}} = \frac{-0{,}33}{\frac{0{,}72}{\sqrt{10}}} = -1{,}43$$

Η απορριπτική περιοχή είναι: R = $\{t < -\,t_{n-1;\,a}\}$
Επειδή όμως $t > -\,t_{n-1;\,a}$ δεν απορρίπτεται η $H_0: \mu_1 = \mu_2$

4.4. Έλεγχος υπόθεσης για τη διακύμανση ενός πληθυσμού

Στην θεωρία διατυπώθηκε ότι η ποσότητα:

$$X^2 = \frac{(n-1)s^2}{\sigma^2} \text{ ακολουθεί την } X^2 \text{ κατανομή με n-1 βαθμούς.}$$

Ο πληθυσμός από όπου πάρθηκε το δείγμα θεωρείται κανονικός με σ^2 και ελέγχεται αν η πληθυσμιακή διακύμανση σ^2 παίρνει την τιμή σ_0^2.
Όπου s^2 η δειγματική διακύμανση.
Οι έλεγχοι υποθέσεων στην προκειμένη περίπτωση παίρνουν τις ακόλουθες μορφές:

$H_0: \sigma^2 = \sigma_0^2$
$H_1: \sigma^2 \neq \sigma_0^2$
δίπλευρος έλεγχος.

$H_0: \sigma^2 = \sigma_0^2$
$H_1: \sigma^2 > \sigma_0^2$
μονόπλευρος έλεγχος (δεξιά)

$H_0: \sigma^2 = \sigma_0^2$
$H_1: \sigma^2 < \sigma_0^2$
μονόπλευρος έλεγχος (αριστερά)

Το στατιστικό ελέγχου που χρησιμοποιείται είναι το:

$$X^2 = \frac{(n-1)s^2}{\sigma_0{}^2} \text{ που ακολουθεί } X^2 \text{ κατανομή με n-1 βαθμούς}$$

Οι απορριπτικές περιοχές για τους προαναφερόμενους ελέγχους είναι αντίστοιχα:

Για τον δίπλευρο έλεγχο R = { $X^2 > \chi^2_{n-1;\frac{a}{2}}$ ή $X^2 < \chi^2_{n-1;1-\frac{a}{2}}$ }, για τον μονόπλευρο έλεγχο (δεξιά) R = {$X^2 > \chi^2_{n-1;a}$}και για τον μονόπλευρο έλεγχο (αριστερά) R = {$X^2 < \chi^2_{n-1;1-a}$}

Παράδειγμα:

Ένα δείγμα από 51 αγρότες παρουσιάζει διακύμανση των ετήσιων εισοδημάτων τους $s^2 = 156$ ευρώ^2. Η πληθυσμιακή διακύμανση είναι η σ^2.

Να ελεγχθεί η υπόθεση:
$H_0: \sigma^2 = 140$
$H_1: \sigma^2 \neq 140$
Σε επίπεδο σημαντικότητας α=5%

Έχουμε το στατιστικό:

$$X^2 = \frac{(n-1)s^2}{\sigma_0{}^2} = \frac{(51-1)156}{140} = 55{,}71$$

$$[\chi^2_{n-1;\frac{a}{2}} = X^2_{50;0,025} = 71{,}42\ ,\ \chi^2_{n-1;1-\frac{a}{2}} = X^2_{50;0,975} = 32{,}36]$$

Για τον δίπλευρο έλεγχο η απορριπτική περιοχή είναι:

$R = \{ X^2 > \chi^2_{n-1;\frac{a}{2}}$ ή $X^2 < \chi^2_{n-1;1-\frac{a}{2}} \}$,

Εφόσον $\chi^2_{n-1;1-\frac{a}{2}} < \mathbf{X}^2 < \chi^2_{n-1;\frac{a}{2}}$ η υπόθεση $H_0: \sigma^2 = 140$ δεν απορρίπτεται.

4.5. Έλεγχος υπόθεσης για το λόγο των διακυμάνσεων δύο πληθυσμών

Σύμφωνα με την θεωρία αν πάρουμε δύο δείγματα ανεξάρτητα μεγέθους n και m αντίστοιχα με s_1^2, s_2^2 τις δειγματικές διακυμάνσεις από δύο κανονικούς πληθυσμούς αντίστοιχα με σ_1^2, σ_2^2 τις πληθυσμιακές διακυμάνσεις τότε ισχύει ότι η ποσότητα:

$\frac{s_1^2}{s_2^2}\frac{\sigma_2^2}{\sigma_1^2}$ ακολουθεί την F κατανομή με n-1 και m-1 βαθμούς ελευθερίας

Οι έλεγχοι υποθέσεων στην προκειμένη περίπτωση παίρνουν τις ακόλουθες μορφές:

$H_0: \dfrac{\sigma_1^2}{\sigma_2^2} = 1$

$H_1: \dfrac{\sigma_1^2}{\sigma_2^2} \neq 1$

δίπλευρος έλεγχος.

$H_0: \dfrac{\sigma_1^2}{\sigma_2^2} = 1$

$H_1: \dfrac{\sigma_1^2}{\sigma_2^2} > 1$

μονόπλευρος έλεγχος (δεξιά)

$H_0: \dfrac{\sigma_1^2}{\sigma_2^2} = 1$

$H_1: \dfrac{\sigma_1^2}{\sigma_2^2} < 1$

μονόπλευρος έλεγχος (αριστερά)

Το στατιστικό ελέγχου που χρησιμοποιείται είναι το:

$F = \dfrac{s_1^2}{s_2^2}$ (όταν $s_1^2 > s_2^2$)

που ακολουθεί την F κατανομή με n-1 και m-1 βαθμούς ελευθερίας

ή

$F = \dfrac{s_2^2}{s_1^2}$ (όταν $s_2^2 > s_1^2$)

που ακολουθεί την F κατανομή με m-1 και n-1 βαθμούς ελευθερίας

Οι απορριπτικές περιοχές για τους προαναφερόμενους ελέγχους είναι αντίστοιχα:

Για τον δίπλευρο έλεγχο R = { $F > F_{n-1,m-1;\frac{a}{2}}$ }, για τον μονόπλευρο έλεγχο (δεξιά) R = {$F > F_{n-1,m-1;a}$ }και για τον μονόπλευρο έλεγχο (αριστερά) R = {$F > F_{n-1,m-1;a}$ }

Παραδείγματα:

1. Πήραμε δύο δείγματα αγροτών από δύο διοικητικές περιφέρειες της Ελλάδας και εξετάσαμε τα ετήσια εισοδήματά τους. Είχαμε τα ακόλουθα δεδομένα:

n= 31, S_1^2 = 200 ευρώ ^{2} και m = 41, S_2^2 = 220 ευρώ ^{2}
Να ελεγχθεί η υπόθεση:

$H_0: \dfrac{\sigma_1^2}{\sigma_2^2} = 1$

$H_1: \dfrac{\sigma_1^2}{\sigma_2^2} < 1$ σε επίπεδο σημαντικότητας α=1%

Όπου σ_1^2, σ_2^2 οι πραγματικές διακυμάνσεις των ετήσιων εισοδημάτων των αγροτικών πληθυσμών στις δύο διοικητικές περιφέρειες της Ελλάδας.

Το στατιστικό ελέγχου είναι:

$F = \dfrac{s_2^2}{s_1^2} = 1,1$ $(S_1^2 < S_2^2)$ έχουμε $F_{40,30;0,01} = 2,3$

Έφόσον $F < F_{40,30;0,01} = 2,3$ άρα η $H_0: \dfrac{\sigma_1^2}{\sigma_2^2} = 1$ δεν απορρίπτεται.

2. Πήραμε δύο δείγματα αγροτών από δύο διοικητικές περιφέρειες της Ελλάδας και εξετάσαμε τα ετήσια εισοδήματά τους. Είχαμε τα ακόλουθα δεδομένα:

n= 31, S_1^2 = 220 ευρώ ^{2} και m =41, S_2^2 = 200 ευρώ ^{2}
Να ελεγχθεί η υπόθεση:

$H_0: \dfrac{\sigma_1^2}{\sigma_2^2} = 1$

$H_1: \dfrac{\sigma_1^2}{\sigma_2^2} < 1$ σε επίπεδο σημαντικότητας α=1%

Όπου σ_1^2, σ_2^2 οι πραγματικές διακυμάνσεις των ετήσιων εισοδημάτων των αγροτικών πληθυσμών στις δύο διοικητικές περιφέρειες της Ελλάδας.

Το στατιστικό ελέγχου είναι:

$F = \dfrac{s_1^2}{s_2^2} = 1,1$ $(S_1^2 > S_2^2)$ έχουμε $F_{30,40;0,01} = 2,2$

Έφόσον $F < F_{30,40;0,01} = 2{,}2$ άρα η $H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1$ δεν απορρίπτεται.

3. Πήραμε δύο δείγματα αγροτών από δύο διοικητικές περιφέρειες της Ελλάδας και εξετάσαμε τα ετήσια εισοδήματά τους. Είχαμε τα ακόλουθα δεδομένα:

n= 31, $S_1^2 = 220$ ευρώ^2 και m = 41, $S_2^2 = 200$ ευρώ^2

Να ελεγχθεί η υπόθεση:

$H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1$

$H_1: \frac{\sigma_1^2}{\sigma_2^2} \neq 1$ σε επίπεδο σημαντικότητας α =2%

Όπου σ_1^2, σ_2^2 οι πραγματικές διακυμάνσεις των ετήσιων εισοδημάτων των αγροτικών πληθυσμών στις δύο διοικητικές περιφέρειες της Ελλάδας.

Το στατιστικό ελέγχου είναι:

$F = \frac{s_1^2}{s_2^2} = 1{,}1$ $(S_1^2 > S_2^2)$ έχουμε $F_{30,40;0,01} = 2{,}2$, όπου $\frac{a}{2} = 0{,}01$

Εφόσον $F < F_{30,40;0,01} = 2{,}2$ άρα η $H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1$ δεν απορρίπτεται.

4.6. Έλεγχος υπόθεσης για το ποσοστό των στοιχείων ενός πληθυσμού

Αν εκτιμήσουμε το ποσοστό (αναλογία) $p = \frac{x}{n}$ των στοιχείων στο μεγάλο δείγμα μεγέθους n που έχουν κάποιο συγκεκριμένο χαρακτηριστικό, τότε σύμφωνα με την

θεωρία το δειγματικό ποσοστό κατανέμεται κανονικά:

Δηλαδή η δειγματική μεταβλητή P_Δ, που παράγεται από τις τιμές p, ακολουθεί την κανονική κατανομή:

$$N(P_{\pi\lambda\eta\theta}, \frac{p_{\pi\lambda\eta\theta}(1-P_{\pi\lambda\eta\theta})}{n})$$

Όπου $P_{\pi\lambda\eta\theta}$ το πραγματικό ποσοστό στον πληθυσμό.

Συνεπώς έχουμε ότι η μεταβλητή:

$$\frac{P_{\Delta} - P_{\pi\lambda\eta\theta}}{\sqrt{\frac{P_{\pi\lambda\eta\theta}(1 - P_{\pi\lambda\eta\theta})}{n}}}$$ ακολουθεί την N(0,1).

Οι έλεγχοι υποθέσεων στην προκειμένη περίπτωση παίρνουν τις ακόλουθες μορφές:

$H_0: P_{\pi\lambda\eta\theta} = p^*$
$H_1: P_{\pi\lambda\eta\theta} \neq p^*$
δίπλευρος έλεγχος.

$H_0: P_{\pi\lambda\eta\theta} = p^*$
$H_1: P_{\pi\lambda\eta\theta} > p^*$
μονόπλευρος έλεγχος (δεξιά)

$H_0: P_{\pi\lambda\eta\theta} = p^*$
$H_1: P_{\pi\lambda\eta\theta} < p^*$
μονόπλευρος έλεγχος (αριστερά)

Το στατιστικό ελέγχου που χρησιμοποιείται είναι το:

$$\frac{p_{\Delta} - p^*}{\sqrt{\frac{p^*(1 - p^*)}{n}}}$$ που ακολουθεί την τυπική κανονική κατανομή.

Οι απορριπτικές περιοχές για τους προαναφερόμενους ελέγχους είναι αντίστοιχα:

Για τον δίπλευρο έλεγχο R = $\{|z| > z_{\frac{a}{2}}\}$, για τον μονόπλευρο έλεγχο (δεξιά) R = $\{z > z_a\}$ και για τον μονόπλευρο έλεγχο (αριστερά) R = $\{z < - z_a\}$

Παράδειγμα:

Ο αριθμός των ελαττωματικών προϊόντων μιας μεταποιητικής βιομηχανίας αγροτικών προϊόντων που εντοπίσθηκαν σ' ένα δείγμα 300 προϊόντων είναι 24 προϊόντα.
Να ελεγχθεί σε επίπεδο σημαντικότητας α = 2% η μηδενική υπόθεση ότι το πραγματικό ποσοστό $P_{\pi\lambda\eta\theta}$ των ελαττωματικών προϊόντων της συνολικής παραγωγής είναι μεγαλύτερο του10%.

$H_0: P_{\pi\lambda\eta\theta} = 0{,}10$
$H_1: P_{\pi\lambda\eta\theta} > 0{,}10$
μονόπλευρος έλεγχος (δεξιά)

Το στατιστικό ελέγχου που χρησιμοποιείται είναι το:

$$Z = \frac{p_\Delta - p^*}{\sqrt{\dfrac{p^*(1-p^*)}{n}}}$$ που ακολουθεί την τυπική κανονική κατανομή.

Η απορριπτική περιοχή είναι: $R = \{z > z_a\}$

Έχουμε n = 300 , m = 24 $\rightarrow \quad \widehat{\rho}_\Delta = \dfrac{24}{300} = 8\% = 0{,}08.$

Οπότε $\widehat{q}$=0,92.

$$\left.\alpha = 1 - 0{,}98 = 0{,}02\right\} \rightarrow z_a = \phi^{-1}(1-a) = \phi^{-1}(0{,}98) = 2{,}06.$$

$$Z = \frac{p_\Delta - p^*}{\sqrt{\dfrac{p^*(1-p^*)}{n}}} = \frac{0{,}08 - 0{,}10}{\sqrt{\dfrac{0{,}10 * 0{,}90}{300}}} = -1$$

Εφόσον $z < z_a$ δεν απορρίπτεται η H_0: $P_{\pi\lambda\eta\theta}$ = 0,10 άρα δεν μπορούμε να ισχυριστούμε ότι το πραγματικό ποσοστό $P_{\pi\lambda\eta\theta}$ των ελαττωματικών προϊόντων της συνολικής παραγωγής είναι μεγαλύτερο του10%.

4.7. Έλεγχος υπόθεσης για τη διαφορά των ποσοστών των στοιχείων δύο πληθυσμών

Έστω $p_1 = \dfrac{x}{n}$ και $p_2 = \dfrac{y}{l}$ είναι οι δειγματικές αναλογίες των στοιχείων, που έχουν κάποιο συγκεκριμένο χαρακτηριστικό, δύο δειγμάτων μεγέθους n, l από δύο πληθυσμούς, όπου οι πραγματικές αναλογίες είναι αντίστοιχα $P_{1\,\pi\lambda\eta\theta}$, $P_{2\,\pi\lambda\eta\theta}$.

Τότε έχουμε ότι οι διαφορές των δειγματικών μεταβλητών $P_{1\,\Delta}$ - $P_{2\,\Delta}$ ακολουθούν κανονική κατανομή:

$N(P_{1\,\pi\lambda\eta\theta} - P_{2\,\pi\lambda\eta\theta},\ \frac{P_{1\pi\lambda\eta\theta}(1-P_{1\pi\lambda\eta\theta})}{n} + \frac{P_{2\pi\lambda\eta\theta}(1-P_{2\pi\lambda\eta\theta})}{l})$

Οι έλεγχοι υποθέσεων στην προκειμένη περίπτωση παίρνουν τις ακόλουθες μορφές:

$H_0: P_{1\,\pi\lambda\eta\theta} = P_{2\,\pi\lambda\eta\theta}$
$H_1: P_{1\,\pi\lambda\eta\theta} \neq P_{2\,\pi\lambda\eta\theta}$
δίπλευρος έλεγχος.

$H_0: P_{1\,\pi\lambda\eta\theta} = P_{2\,\pi\lambda\eta\theta}$
$H_1: P_{1\,\pi\lambda\eta\theta} > P_{2\,\pi\lambda\eta\theta}$
μονόπλευρος έλεγχος (δεξιά)

$H_0: P_{1\,\pi\lambda\eta\theta} = P_{2\,\pi\lambda\eta\theta}$
$H_1: P_{1\,\pi\lambda\eta\theta} < P_{2\,\pi\lambda\eta\theta}$
μονόπλευρος έλεγχος (αριστερά)

Το στατιστικό ελέγχου που χρησιμοποιείται είναι:

$$\frac{p_{1\Delta} - p_{2\Delta}}{\sqrt{pq(\frac{1}{n}+\frac{1}{l})}}$$ ακολουθεί την τυπική κανονική κατανομή

όπου $p = \frac{x+y}{n+l}$ εκτιμά την $\mathbf{P} = \mathbf{P}_{1\,\pi\lambda\eta\theta} = \mathbf{P}_{2\,\pi\lambda\eta\theta}$ και $q = 1 - p$

Οι απορριπτικές περιοχές για τους προαναφερόμενους ελέγχους είναι αντίστοιχα:

Για τον δίπλευρο έλεγχο $R = \{|z| > z_{\frac{a}{2}}\}$, για τον μονόπλευρο έλεγχο (δεξιά) $R = \{z > z_a\}$ και για τον μονόπλευρο έλεγχο (αριστερά) $R = \{z < - z_a\}$

Παράδειγμα:

Παίρνουμε δύο μεγάλα δείγματα μιας ποικιλίας ενός φυτού από δύο διαφορετικούς πληθυσμούς και εξετάζουμε πόσα από αυτά ασθένησαν μέσα σε ένα συγκεκριμένο χρονικό διάστημα. Τα δεδομένα που προέκυψαν ήταν:

Από τα n=120 του πρώτου δείγματος ασθένησαν τα 12 και από τα m =130 του δεύτερου δείγματος ασθένησαν τα 18.
Να ελεγχθεί η μηδενική υπόθεση
$H_0: P_{1\ \pi\lambda\eta\theta} = P_{2\ \pi\lambda\eta\theta}$
$H_1: P_{1\ \pi\lambda\eta\theta} \neq P_{2\ \pi\lambda\eta\theta}$
σε επίπεδο σημαντικότητας α = 5%.

Έχουμε ότι:

$\rho_{1\ \Delta} = 12/120 = 0{,}1$, $\rho_{2\ \Delta} = 18/130 = 0{,}14$ και $\rho_{1\ \Delta} - \rho_{2\ \Delta} = -0{,}04$

$z_{\frac{a}{2}} = Z_{0,025} = 1{,}96$

$p = \dfrac{x+y}{n+l} = 0{,}12$, $q = 0{,}88$

$$Z = \frac{p_{1\Delta} - p_{2\Delta}}{\sqrt{pq(\frac{1}{n} + \frac{1}{l})}} = \frac{0{,}1 - 0{,}14}{\sqrt{0{,}12 * 0{,}88(\frac{1}{120} + \frac{1}{130})}} = -0{,}98$$

Η απορριπτική περιοχή είναι: $R = \{ |z| > z_{\frac{a}{2}} \}$

Εφόσον $|z| < z_{\frac{a}{2}}$ η $H_0: P_{1\ \pi\lambda\eta\theta} = P_{2\ \pi\lambda\eta\theta}$ δεν απορρίπτεται.

4.8. Έλεγχος υπόθεσης και διάστημα εμπιστοσύνης

Εάν υπολογίσουμε το 100(1-α) % διάστημα εμπιστοσύνης μιας παραμέτρου Θ (μ, σ, p) ενός πληθυσμού τότε για τον έλεγχο μηδενικής υπόθεσης H_0: Θ = θ (η εκτιμήτριά της από ένα δείγμα) μπορούμε ότι:

Αν το θ ανήκει στο 100(1-α) % διάστημα εμπιστοσύνης της παραμέτρου Θ τότε η :

H_0: Θ = θ γίνεται δεκτή, αλλιώς απορρίπτεται

Η εναλλακτική υπόθεση είναι πάντα η H_1: $\Theta \neq \theta$

Κεφάλαιο 5ο

Ανάλυση κατηγορικών δεδομένων

5.1. Γενικά

Αρκετές φορές δεδομένα κατηγοριοποιούνται ως προς δύο χαρακτηριστικά γνωρίσματα τέτοια δεδομένα συνήθως δίνονται με πίνακες που καλούνται **πίνακες συνάφειας.**
Ένας πίνακας συνάφειας είναι της μορφής:

	$B_1 \quad B_2 \ldots\ldots\ldots B_\lambda$	Αθροίσματα γραμμών
A_1 A_2 . . A_κ	$n_{11} \quad n_{12} \ldots\ldots\ldots n_{1\lambda}$ $n_{21} \quad n_{22} \ldots\ldots\ldots n_{2\lambda}$. $n_{\kappa 1} \quad n_{\kappa 2} \ldots\ldots\ldots n_{\kappa\lambda}$	n_{1*} n_{2*} . $n_{\kappa *}$
Αθροίσματα στηλών	$n_{*1} \quad n_{*2} \ldots\ldots\ldots n_{*\lambda}$	n

Τα A και B είναι τα χαρακτηριστικά, το A χωρίζεται σε κ – κατηγορίες και το B σε λ – κατηγορίες.
Το πλήθος των μετρήσεων που ανήκουν ταυτόχρονα στην A_i και στην B_j είναι το n_{ij}
Το άθροισμα της i - γραμμής είναι το n_{i*}, το άθροισμα της j – στήλης είναι το n_{*j} τέλος το σύνολο των μετρήσεων είναι το n.
Η πιθανότητα του ij κελιού συμβολίζεται με p_{ij}.

5.2. Ο έλεγχος ανεξαρτησίας

Στην περίπτωση αυτή ελέγχουμε την ανεξαρτησία των δύο χαρακτηριστικών.
Σχηματίζουμε την μηδενική υπόθεση:

H_0: τα δύο χαρακτηριστικά είναι ανεξάρτητα
H_1: τα δύο χαρακτηριστικά είναι εξαρτημένα

Χρησιμοποιούμε έλεγχο X^2.
Για τον έλεγχο της μηδενικής υπόθεσης χρησιμοποιούμε το στατιστικό:

$$X^2 = \sum_{i=1}^{\kappa} \sum_{j=1}^{\lambda} \frac{(n_{ij} - \theta_{ij})^2}{\theta_{ij}} = \sum_{i=1}^{\kappa} \sum_{j=1}^{\lambda} \frac{n_{ij}^{\ 2}}{\theta_{ij}} - n$$

Όπου θ_{ij} είναι η θεωρητική συχνότητα του ij – κελιού (που μπορεί να προκύψει από την προσαρμογή μιας θεωρητικής κατανομής στα δεδομένα).

Για την εκτίμηση της θ_{ij} παίρνουμε την ποσότητα:

$$\frac{n_{i*}n_{*j}}{n}$$

Οπότε στον υπολογισμό της ποσότητας X^2 στη θέση της θ_{ij} βάζουμε την εκτίμησή της $\frac{n_{i*}n_{*j}}{n}$.

Η απορριπτική περιοχή της H_0 είναι:

$$R = \{ X^2 > \chi^2_{(\kappa-1)(\lambda-1);\alpha} \}$$

Πρέπει η εκτίμηση της θ_{ij} δηλαδή η ποσότητα $\frac{n_{i*}n_{*j}}{n} \geq 5$ για κάθε i και j.

- Αν υπάρχουν $\theta_{ij} < 5$ συμπτύσσουμε τις κατηγορίες ώστε να πάρουμε $\theta_{ij} \geq 5$.
- Αν παρόλα αυτά εξακολουθεί να υπάρχει $\theta_{ij} < 5$ τότε εφόσον το πλήθος των κατηγοριών είναι μικρό δεχόμαστε μια κατηγορία με $\theta_{ij} < 5$ σύμφωνα με τον περιορισμό του Cochran.

Παράδειγμα:

Ρωτήθηκαν εργαζόμενοι εργατοϋπάλληλοι από μια περιοχή όσον αφορά τις μηνιαίες αμοιβές τους και είχαμε τα ακόλουθα αποτελέσματα:

	Χαμηλή Αμοιβή	Μεσαία Αμοιβή	Υψηλή Αμοιβή
Άνδρες	64	43	13
Γυναίκες	69	47	9

Μπορούμε να ισχυριστούμε ότι τα χαρακτηριστικά φύλο και μηνιαία αμοιβή είναι ανεξάρτητα (α=0,05);

Έχουμε έλεγχο ανεξαρτησίας και ο πίνακας συνάφειας είναι:

	Χαμηλή Αμοιβή	Μεσαία Αμοιβή	Υψηλή Αμοιβή	
Άνδρες	64	43	13	120
Γυναίκες	69	47	9	125
	133	90	22	245

Η υπόθεση είναι:

H_0: τα δύο χαρακτηριστικά φύλο και μηνιαία αμοιβή είναι ανεξάρτητα
H_1: τα δύο χαρακτηριστικά είναι εξαρτημένα

Χρησιμοποιούμε έλεγχο X^2.
Για τον έλεγχο της μηδενικής υπόθεσης χρησιμοποιούμε το στατιστικό:

$$X^2 = \sum_{i=1}^{\kappa} \sum_{j=1}^{\lambda} \frac{(n_{ij} - \theta_{ij})^2}{\theta_{ij}} = \sum_{i=1}^{\kappa} \sum_{j=1}^{\lambda} \frac{n_{ij}^2}{\theta_{ij}} - \mathrm{n}$$

Υπολογίζουμε τις θεωρητικές κατανομές $\theta_{ij} = \frac{n_{i*} n_{*j}}{n}$ και έχουμε:

$\theta_{11} = 65{,}14$ $\theta_{12} = 44{,}08$ $\theta_{13} = 10{,}78$

$\theta_{21} = 67{,}86$ $\theta_{22} = 45{,}92$ $\theta_{23} = 11{,}22$

$$X^2 = \sum_{i=1}^{\kappa} \sum_{j=1}^{\lambda} \frac{n_{ij}^2}{\theta_{ij}} - \mathrm{n} = 1 < \chi^2_{(\kappa-1)(\lambda-1);\alpha} = \chi^2_{2;0{,}05} = 5{,}99$$

Άρα δεν απορρίπτεται η H_0: τα δύο χαρακτηριστικά φύλο και μηνιαία αμοιβή είναι ανεξάρτητα

5.3. Έλεγχος ομοιογένειας

Αρκετές φορές χωρίζουμε ένα πληθυσμό σε υπο-πληθυσμούς και εξετάζουμε συγκεκριμένου μεγέθους δείγματα από τους υπο-πληθυσμούς αυτούς ως προς ένα χαρακτηριστικό. Τα δείγματα ταξινομούνται σε κατηγορίες ως προς το χαρακτηριστικό.

Πίνακας συνάφειας:

πληθυσμός	Κατηγορία 1 (χαρακτηριστικού)	Κατηγορία 2 (χαρακτηριστικού)	Σύνολα γραμμών
Υποπληθυσμός 1 Υποπληθυσμός 2 Υποπληθυσμός 3	n_{11} n_{21} n_{31}	n_{12} n_{22} n_{32}	ΝΓ1 ΝΓ2 ΝΓ3
Συν. Στηλών	ΝΣ1	ΝΣ2	Ν

Η μηδενική υπόθεση H_0, αφορά την ομοιογένεια του πληθυσμού.
Συγκεκριμένα έχουμε:

H_0: Οι παρατηρούμενες κατανομές ως προς τις κατηγορίες των στηλών είναι για όλες τις γραμμές ίδιες (πληθυσμός ομοιογενής).

Ο έλεγχος της H_0 γίνεται με το X^2
Για τον έλεγχο της μηδενικής υπόθεσης χρησιμοποιούμε πάλι το στατιστικό:

$$X^2 = \sum_{i=1}^{\kappa} \sum_{j=1}^{\lambda} \frac{(n_{ij} - \theta_{ij})^2}{\theta_{ij}} = \sum_{i=1}^{\kappa} \sum_{j=1}^{\lambda} \frac{n_{ij}^{\ 2}}{\theta_{ij}} - n$$

Η απορριπτική περιοχή της H_0 επίσης είναι:

$$R = \{ X^2 > \chi^2_{(\kappa-1)(\lambda-1);\alpha} \}$$

Ισχύουν τα ίδια με το έλεγχο ανεξαρτησίας.

Παράδειγμα:

Ρωτήθηκαν αγρότες από μια περιοχή όσον αφορά τις μηνιαίες δαπάνες τους για αλκοολούχα ποτά και είχαμε τα ακόλουθα αποτελέσματα:

	Χαμηλή Δαπάνη	Μεσαία Δαπάνη	Υψηλή Δαπάνη
Νέοι	64	43	13
Μεσήλικες	69	47	9
Ηλικιωμένοι	53	40	20

Μπορούμε να ισχυριστούμε ότι ο ανδρικός αγροτικός πληθυσμός παρουσιάζει ομοιογένεια ως προς τις δαπάνες για αλκοολούχα ποτά (α=0,05);

Έχουμε έλεγχο ομοιογένειας και ο πίνακας συνάφειας είναι:

	Χαμηλή Δαπάνη	Μεσαία Δαπάνη	Υψηλή Δαπάνη	
Νέοι	64	43	13	120
Μεσήλικες	69	47	9	125
Ηλικιωμένοι	53	40	20	113
	186	130	42	358

Η υπόθεση είναι:

H_0: Οι παρατηρούμενες κατανομές ως προς τις κατηγορίες των στηλών είναι για όλες τις γραμμές ίδιες (πληθυσμός ομοιογενής).

Ο έλεγχος της H_0 γίνεται με το X^2

Για τον έλεγχο της μηδενικής υπόθεσης χρησιμοποιούμε το στατιστικό:

$$X^2 = \sum_{i=1}^{\kappa} \sum_{j=1}^{\lambda} \frac{n_{ij}^2}{\theta_{ij}} - n$$

Υπολογίζουμε τις θεωρητικές κατανομές $\theta_{ij} = \frac{n_{i*} n_{*j}}{n}$ και έχουμε:

$\theta_{11} = 62{,}35$ $\theta_{12} = 43{,}58$ $\theta_{13} = 14{,}08$

$\theta_{21} = 64{,}94$ $\theta_{22} = 45{,}39$ $\theta_{23} = 14{,}66$

$\theta_{31} = 58{,}71$ $\theta_{32} = 41{,}03$ $\theta_{33} = 13{,}26$

$$X^2 = \sum_{i=1}^{\kappa} \sum_{j=1}^{\lambda} \frac{n_{ij}^2}{\theta_{ij}} - n = 0 < \chi^2_{(\kappa-1)(\lambda-1);\alpha} = \chi^2_{4;0,05} = 9{,}49$$

Άρα δεν απορρίπτεται H_0: Οι παρατηρούμενες κατανομές ως προς τις κατηγορίες των στηλών είναι για όλες τις γραμμές ίδιες (πληθυσμός ομοιογενής).

Κεφάλαιο 6ο
Έλεγχος προσαρμοστικότητας

6.1. Έλεγχος προσαρμοστικότητας της υποθετικής κατανομής.

Όταν έχουμε υποθέσει μια θεωρητική κατανομή, προσδιορισμένη από το γενικό σχήμα του ιστογράμματος η προσαρμοστικότητα της υποθετικής κατανομής μπορεί να επιβεβαιωθεί ή όχι με στατιστικό τρόπο χρησιμοποιώντας μεθόδους που ονομάζονται **έλεγχοι καλής προσαρμογής ή έλεγχοι προσαρμοστικότητας**. Υπάρχουν δύο τέτοιοι έλεγχοι για την προσαρμογή των κατανομών: Ο **έλεγχος X^2** και ο **έλεγχος κατά Kolmogorov-Smirnov**.

6.2. Έλεγχος X^2.

Ας πάρουμε ένα δείγμα n - παρατηρήσεων μιας τυχαίας μεταβλητής X.

Με τον έλεγχο X^2 συγκρίνονται οι συχνότητες $v_1, v_2, \ldots, v_\kappa$ για κ συγκεκριμένες τιμές (ή κ διαστήματα), όπως αυτές προκύπτουν από τις παρατηρήσεις της τυχαίας μεταβλητής X με τις αντίστοιχες συχνότητες $e_1, e_2, \ldots, e_\kappa$ μιας υποθετικής θεωρητικής κατανομής.

Η βάση για την ποιοτική αξιολόγηση της σύγκρισης αυτής είναι η κατανομή της ποσότητας:

$$\sum_{i=1}^{\kappa} \frac{(v_i - e_i)^2}{e_i}$$

η οποία τείνει να γίνει κατανομή X^2 όταν $n \to +\infty$ με βαθμούς ελευθερίας $v = \kappa - 1$.

Αν όμως οι παράμετροι της θεωρητικής κατανομής είναι άγνωστοι και πρέπει να εκτιμηθούν από τα δεδομένα, τότε η προηγούμενη πρόταση θα ισχύει μόνο αν για κάθε άγνωστη παράμετρο που πρέπει να εκτιμηθεί, οι βαθμοί ελευθερίας μειωθούν κατά μια μονάδα.

Με βάση τα παραπάνω, αν για μια υποθετική κατανομή ισχύει η σχέση:

$$\sum_{i=1}^{\kappa} \frac{(v_i - e_i)^2}{e_i} < C_{1-\alpha,v}$$

όπου $C_{1-\alpha,v}$ είναι η τιμή της κατάλληλης κατανομής X_v^2 που αντιστοιχεί στην αθροιστική πιθανότητα 1-α, τότε:

Η υποθετική θεωρητική κατανομή είναι ένα αποδεκτό μοντέλο σε επίπεδο σημαντικότητας (α). Αν η σχέση δεν ισχύει τότε τα δεδομένα των παρατηρήσεων δεν επαληθεύουν την υποθετική κατανομή.

Στην εφαρμογή του ελέγχου X^2 είναι γενικά απαραίτητο (για να έχουμε ικανοποιητικά αποτελέσματα) να έχουμε $\kappa \geq 5$ και $e_i \geq 5$.

Παραδείγματα.

1. Ο αριθμός των αυτοκινήτων που φτάνουν, ανά λεπτό, σε μια διασταύρωση έχει παρατηρηθεί, με τα παρακάτω αποτελέσματα:

0,3,1,2,0,1,1,1,2,0,1,4,3,1,1,0,0,1,0,2.

Να γίνει ένας X^2 έλεγχος για να εξετασθεί αν ο αριθμός των αφίξεων στη διασταύρωση είναι διαδικασία Poisson σε 1% επίπεδο σημαντικότητας.

Πίνακας.

Αριθ.αυτοκ/λεπτό	**Παρατηρ.συχνοτ.v_i**	**Θεωρητ.συχν. e_i**	**$(v_i - e_i)^2$**	**$(v_i-e_i)^2/e_i$**
0	6	6,02	0,0002	$3{,}322\cdot10^{-5}$
1	8	7,22	0,6084	0,0842
2	3	4,34	1,7956	0,4137
≥ 3	3	2,42	0,58	0,2396
	20	20		0,7376

Έχουμε εκτιμώμενο το μέσο όρο αφίξεων nt, δεν μας δίνεται:

Εκτίμηση $\hat{nt} = (0\cdot6+1\cdot8+2\cdot3+3\cdot2+4\cdot1)\cdot\frac{1}{20} = 1{,}2$ του μέσου όρου αφίξεων nt.

Θεωρώ την διαδικασία Poisson: $P(X=x)\frac{(nt)^x\cdot e^{-nt}}{x!}$ απ΄ όπου προκύπτουν οι πιθανότητες:

$P(X=0)=0{,}301$

$P(X=1)=0{,}361$

$P(X=2)=0{,}217$

$P(X\geq3)=1-P(X<3)=1-\left[P(x=0)+P(x=1)+P(x=2)\right]=0{,}121$

Οι αντίστοιχες θεωρητικές συχνότητες μέσω της διαδικασίας Poisson είναι:

$$e_1 = P(x=0)\cdot 20 = 0{,}301\cdot 20 = 6{,}02$$

$$e_2 = P(x=1)\cdot 20 = 0{,}361\cdot 20 = 7{,}22$$

$$e_3 = P(x=2)\cdot 20 = 0{,}217\cdot 20 = 4{,}34$$

$$e_4 = P(x=3)\cdot 20 = 0{,}121\cdot 20 = 2{,}42$$

Επειδή η παράμετρος μέσος όρος αφίξεων nt εκτιμάται από την $\hat{nt}, n$ ποσότητα $\sum_{i=1}^{\kappa} (v_i - e_i)^2 / e_i$ έχει κατανομή X^2 με v=(κ-1)-1=(4-1)-1=2 βαθμοί ελευθερίας. (εκτιμάται ή Υ άρα αφαιρείται μια μονάδα).

Από τον πίνακα έχουμε (α=0,01): $C_{0,99\,,\,2} = 9{,}21 > 0{,}7376 = \sum_{i=1}^{\kappa} \dfrac{(v_i - e_i)^2}{e_i}$

Συνεπώς $C_{1-\alpha,v} > \sum_{i=1}^{\kappa} (v_i - e_i)^2 / e_i$ οπότε η Poisson είναι αξιόπιστο μοντέλο για την εξέταση των αφίξεων σε επίπεδο σημαντικότητας 1%.

2. Στον πίνακα που ακολουθεί αναγράφονται δεδομένα παρατηρήσεων για το ρυθμό οξυγόνωσης (K) του ποταμού Ohio River, σε μια συγκεκριμένη τοποθεσία και σε 20°C. Για να περιγραφεί ο ρυθμός οξυγόνωσης προτείνεται ένα μοντέλο κανονικής κατανομής μέσης τιμής 0,173 και τυπικής απόκλισης 0,066 (και οι δύο τιμές έχουν εκτιμηθεί με βάση τα δεδομένα παρατηρήσεων).

Να εφαρμοσθεί ένας έλεγχος X^2 σε επίπεδο σημαντικότητας 1% για την προσαρμοστικότητα της προτεινόμενης κατανομής στα δεδομένα του πίνακα.

Ο (ανά μέρα)	**Παρατηρ. Συχνοτ.**
0,000-0,049	1
0,050-0,099	11
0,100-0,149	20
0,150-0,199	23
0,200-0,249	15
0,250-0,299	11
0,300-0,349	2

$$P(\alpha \le X \le \beta) = \phi\left(\frac{\beta-\mu}{\sigma}\right) - \phi\left(\frac{\alpha-\mu}{\sigma}\right).$$

Για να έχουμε $e_i \ge 5$ μειώνουμε των αριθμό των διαστημάτων.

$P(X \le 0{,}099) = \phi(-1{,}121) = 1 - \phi(1{,}121) = 0{,}1314$ έχουμε:

$P(0{,}100 \le X \le 0{,}149) = 0{,}2259$ $\qquad e_1 = 0{,}1314 \cdot 83 = 10{,}91$

$P(0{,}150 \le X \le 0{,}199) = 0{,}2902 \longrightarrow$ $\qquad e_2 = 0{,}2259 \cdot 83 = 18{,}75$

$P(0{,}200 \le X \le 0{,}249) = 0{,}2195$ $\qquad e_3 = 0{,}2902 \cdot 83 = 24{,}09$

$P(X \ge 0{,}250) = 1 - P(X < 0{,}250) = 0{,}1329$ $\qquad e_4 = 0{,}2195 \cdot 83 = 18{,}22$

$e_5 = 0{,}1329 \cdot 83 = 11{,}03$

Πίνακας

Ο (ανά μέρα)	**Παρατ.συχνοτ.v_i**	**Θεωρητ.συχν. e_i**	**$(v_i - e_i)^2$**	**$(v_i - e_i)^2/e_i$**
$\le 0{,}099$	12	10,91	1,1881	0,1089
0,100-0,149	20	18,75	1,5625	0,0833
0,150-0,199	23	24,09	1,1881	0,0493
0,200-0,249	15	18,22	10,3684	0,5690
$\ge 0{,}250$	13	11,03	3,8809	0,3518
	83	83		1,1623

κ = 5, βαθμοί ελευθερίας : ν=(κ-1)-2 = 2. (έχουν εκτιμηθεί 2 παράμετροι: μέση τιμή – τυπική απόκλιση, άρα αφαιρούμε το 2).

Από πίνακα για α = 0,01 έχουμε: $C_{0,99\,,\,2} = 9{,}21 > 1{,}1623$.

Επομένως επιβεβαιώνεται η πιστότητα της κανονικής κατανομής.

6.3. Έλεγχος Kolmogorov-Smirnov (K-S).

Η βασική διαδικασία που περιέχεται στον έλεγχο αυτό είναι η σύγκριση της πειραματικής αθροιστικής συχνότητας με μια υποθετική θεωρητική συνάρτηση κατανομής.

Αν η διαφορά είναι μεγάλη, σε σχέση μ' αυτή που κανονικά αναμέναμε από ένα δοσμένο μέγεθος δείγματος, τότε η θεωρητική κατανομή απορρίπτεται.

Για ένα δείγμα μεγέθους n, ανακατατάσσουμε τα δεδομένα σε σειρά αυξανόμενου μεγέθους. Από αυτά τα ταξινομημένα δεδομένα κατασκευάζουμε στη συνέχεια μια κλιμακούμενη ΑΣΚ ως εξής.

$$Sn(x) = \begin{cases} 0, & \gamma\iota\alpha \quad x < x_1 \\ \kappa / n & \gamma\iota\alpha \quad x_\kappa \le x < x_{\kappa+1} \\ 1, & \gamma\iota\alpha \quad x \ge x_n \end{cases}$$

όπου $x_1, x_2, \ldots, x_n$ είναι οι δειγματικές τιμές και n το μέγεθος του δείγματος.

Στον έλεγχο K-S, η μέγιστη διαφορά μεταξύ των $S_n(x)$ και $F(x)$, μέσα στο πεδίο ορισμού της X, δίδει το μέτρο της ασυμφωνίας μεταξύ του θεωρητικού μοντέλου και των παρατηρηθέντων δεδομένων.

Συμβολίζουμε αυτή τη μέγιστη διαφορά με:

$$D_n = \max_x \left| F(x) - S_n(x) \right|.$$

θεωρητικά, η D_n είναι μια τυχαία μεταβλητή που η κατανομή της εξαρτάται από το n. Για ένα ορισμένο επίπεδο σημαντικότητας α, ο έλεγχος K-S συγκρίνει την D_n με την D_n^α που ορίζεται από τη σχέση.

$$P(D_n \le D_n^\alpha) = 1 - \alpha.$$

Αν η παρατηρηθείσα τιμή D_n είναι μικρότερη από την κρίσιμη τιμή D_n^α, τότε η προτεινόμενη κατανομή είναι αποδεκτή στο συγκεκριμένο επίπεδο σημαντικότητας α αλλιώς η υποτιθέμενη κατανομή απορρίπτεται.

Το πλεονέκτημα του ελέγχου K-S σε σύγκριση με τον έλεγχο X^2 είναι ότι δεν χρειάζεται να διαιρέσουμε τα δεδομένα σε διαστήματα επομένως τα προβλήματα που παρουσιάζονται με την προσέγγιση X^2 όταν τα e_i είναι μικρά ή το κ μικρό, δεν υπάρχουν στον έλεγχο K-S.

Παραδείγματα

1. Στον παρακάτω πίνακα δίνονται οι τιμές της αθροιστικής συχνότητας F(x) της κανονικής κατανομής που χρησιμοποιήθηκε σ' ένα δείγμα. Να εξετασθεί η πιστότητα ή μη της κανονικής κατανομής με τον έλεγχο K-S.

θεωρώ την υποθετική συνάρτηση κατανομής.

$$S_n(x) = \begin{cases} 0, & \gamma\iota\alpha \quad x < x_1 \\ \kappa / n = \kappa /_{15}, & \gamma\iota\alpha \quad x_\kappa \le x < x_{\kappa+1} \\ 1, & \gamma\iota\alpha \quad x \ge x_{15} \end{cases}$$

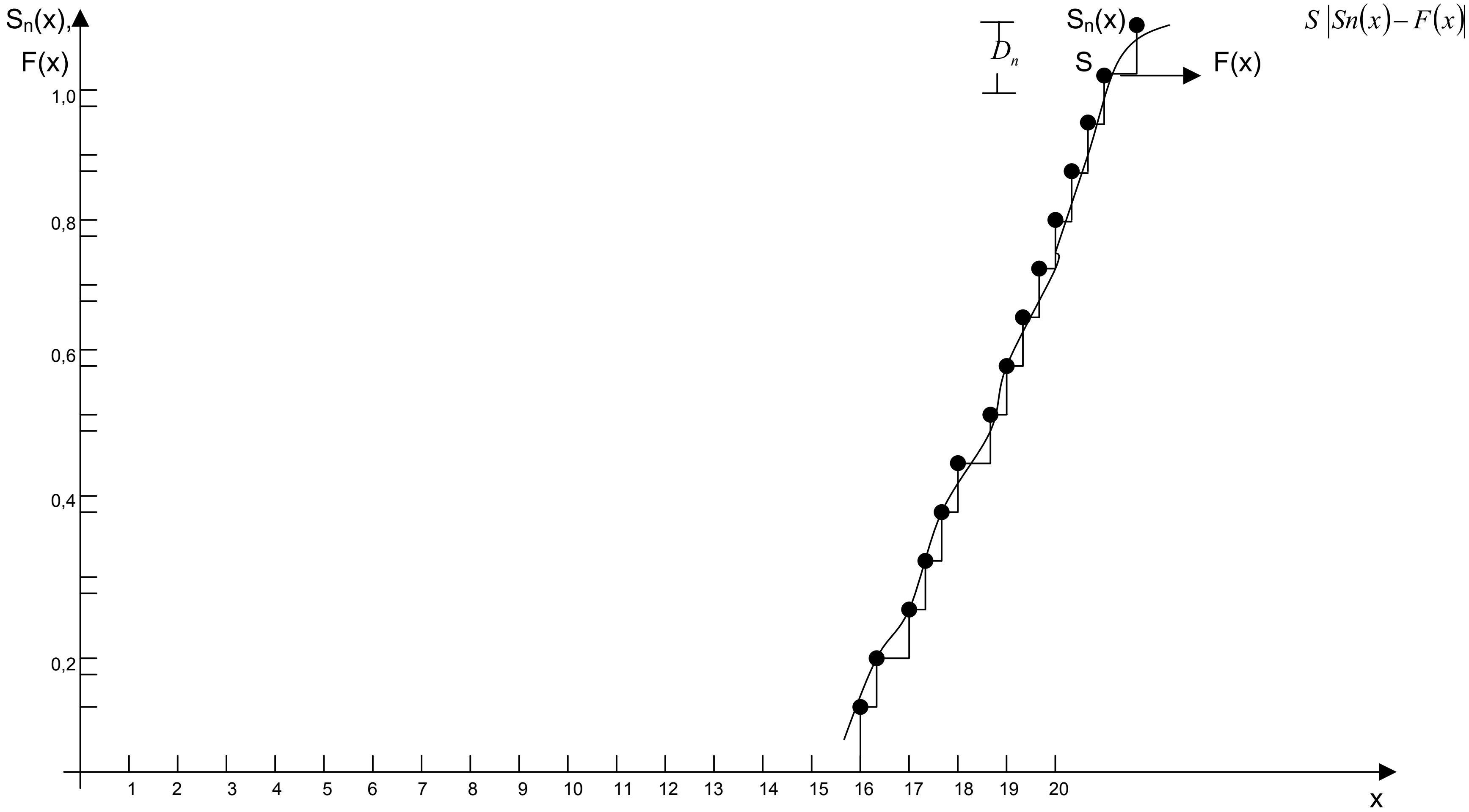
<u>Πίνακας</u>

κ	**F(x)**	**S_n(x)**	$\lvert F(x) - Sn(x) \rvert$
1	0,063	0,066	0,003
2	0,125	0,133	0,008
3	0,188	0,2	0,012
4	0,25	0,266	0,016
5	0,313	0,333	0,02
6	0,375	0,4	0,025
7	0,438	0,466	0,028
8	0,5	0,533	0,033
9	0,563	0,6	0,037
10	0,625	0,666	0,041
11	0,688	0,733	0,045
12	0,75	0,8	0,05
13	0,813	0,866	0,053
14	0,875	0,933	0,058
15	0,938	1	0,062

$$D_n = \max_{v} \lvert F(x) - S_n(X) \rvert = 0{,}062.$$

$$\left.\begin{array}{l} n=15 \\ \\ \alpha=0{,}01 \end{array}\right\} \quad \text{Από πίνακα} \qquad D_n^{\alpha} = 0{,}40 > D_n$$

Επομένως η προτεινόμενη κανονική κατανομή είναι αποδεκτή.



ΣΥΓΚΡΙΣΗ ΤΩΝ $S_n(x)$, $F(x)$.

2. Στον πίνακα που ακολουθεί αναγράφεται η αθροιστική συχνότητα F(x) όπως προέκυψε με την εφαρμογή μιας θεωρητικής κατανομής Poisson σε ένα δείγμα για να ελεγχθεί αν η κατανομή Poisson προσαρμόζεται καλά. Να χρησιμοποιηθεί η μέθοδος K-S με επίπεδο σημαντικότητας 1%.

$$S_n(X) = \begin{cases} 0, & \gamma\iota\alpha \quad x < x_1 \\ \kappa / n & \gamma\iota\alpha \quad x_\kappa \le x < x_{\kappa+1} \\ 1, & \gamma\iota\alpha \quad x \ge x_5 \end{cases}$$

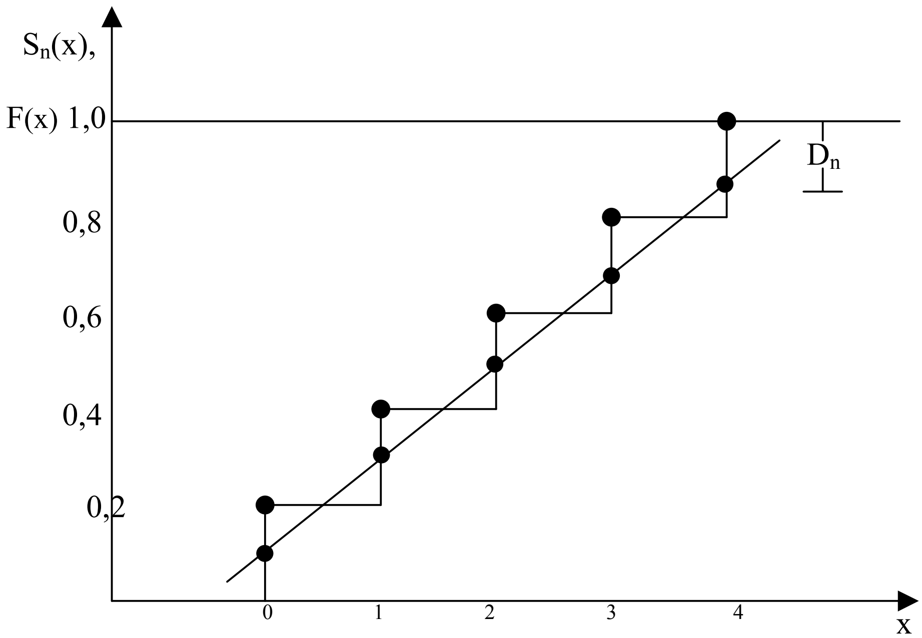
κ	**F(x)**	**S_n(x)**	$\lvert F(x) - S_n(x) \rvert$
1	0,166	0,2	0,034
2	0,333	0,4	0,067
3	0,5	0,6	0,1
4	0,66	0,8	0,134
5	0,833	1,0	0,167

$$D_n = \max_v \lvert F(x) - S_n(X) \rvert = 0{,}167.$$

$$\left. \begin{array}{l} n=5 \\ \\ \alpha=0{,}01 \end{array} \right\} \quad \text{Από πίνακα} \longrightarrow D_n^{\alpha} = 0{,}67 > D_n$$

Επομένως η προτεινόμενη κατανομή Poisson είναι αποδεκτή.

Η σύγκριση των F(x), S_n(x) παρουσιάζεται στο παρακάτω διάγραμμα.



Σύγκριση των $s_n(x)$, F(x).

Κεφάλαιο 7ο

Απλή παλινδρόμηση –Το γραμμικό μοντέλο

7.1. Η Έννοια της συσχέτισης.

Θέλουμε να βρούμε αν υπάρχει σχέση μεταξύ των τιμών των δύο μεταβλητών X, Y σ' ένα πλήθος ν ατόμων και να την προσδιορίσουμε.

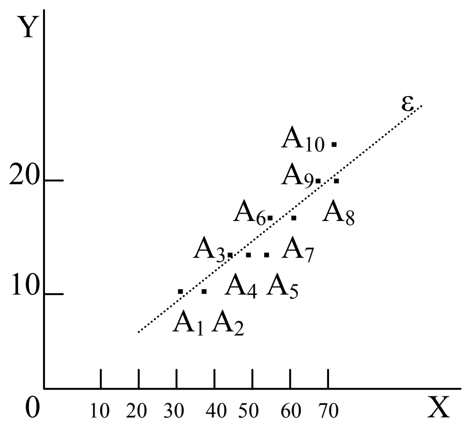
Το κάθε άτομο θα έχει ένα ζεύγος τιμών (X,Y) των τυχαίων μεταβλητών X, Y. Έτσι θα δημιουργηθούν ν – διατεταγμένα ζεύγη:

$(X_1, Y_1), (X_2, Y_2), ..., (X_\nu, Y_\nu)$. Έστω ότι X είναι το βάρος και Y η ηλικία 10 ατόμων, στον παρακάτω πίνακα.

Πίνακας

Άτομα	X	Y	Άτομα	X	Y
1	40	12	6	55	1
2	45	13	7	60	17
3	45	15	8	65	17
4	50	15	9	65	18
5	55	16	10	70	19

Στο ορθογώνιο επίπεδο δημιουργούμε τα 10 σημεία. Το σύνολο των σημείων ονομάζεται **νέφος σημείων ή διάγραμμα διασποράς**



Τα 10 σημεία είναι διασπαρμένα γύρω από την ευθεία ε με προσανατολισμό κάτω αριστερά-πάνω δεξιά.

Η ευθεία γραμμή **(ε)** εκφράζει την σχέση των X, Y των 10 ατόμων.

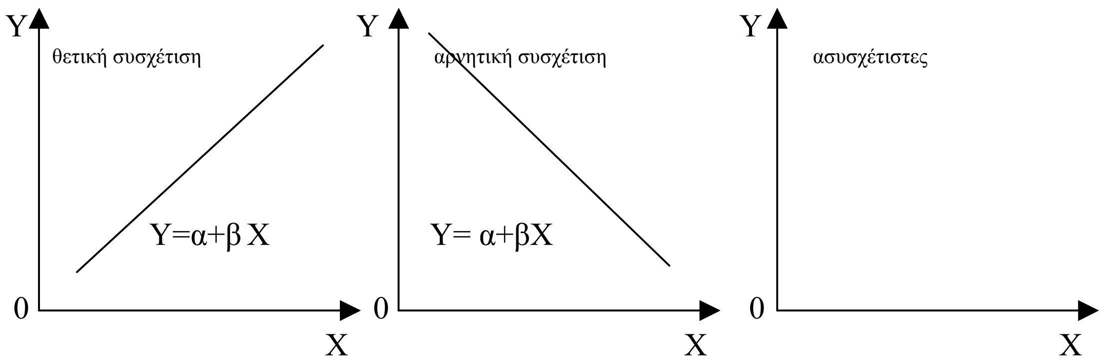
Η αναλυτική της έκφραση είναι: $Y = a + \beta X$

Οι μεταβλητές X, Y λέμε ότι συσχετίζονται.

Θετική συσχέτιση: όταν οι χαμηλές τιμές της μιας μεταβλητής αντιστοιχούν στις χαμηλές τιμές της άλλης μεταβλητής και οι υψηλές τιμές της μιας στις υψηλές τιμές της άλλης.

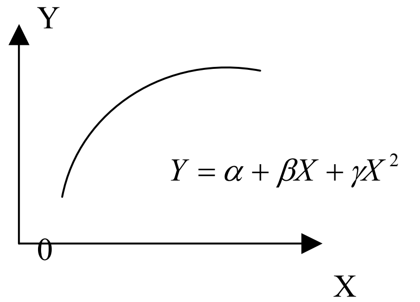
Αρνητική συσχέτιση: όταν οι χαμηλές τιμές της μιας μεταβλητής αντιστοιχούν στις υψηλές τιμές της άλλης και αντίστροφα.

Ασυσχέτιστες μεταβλητές: όταν δεν υπάρχει προσαρμογή μιας καμπύλης στο διάγραμμα διασποράς των δύο μεταβλητών.



Συχνά από το διάγραμμα το σύνολο των σημείων υποδεικνύουν το είδος της καμπύλης που προσαρμόζεται σ' αυτό.

Έτσι μπορεί να χρησιμοποιηθεί μια ευθεία $Y = \alpha + \beta X$, όπως στα παραπάνω σχήματα ή μια παραβολή (δευτεροβάθμια καμπύλη) $Y = \alpha + \beta X + \gamma X^2$ στο παρακάτω σχήμα.

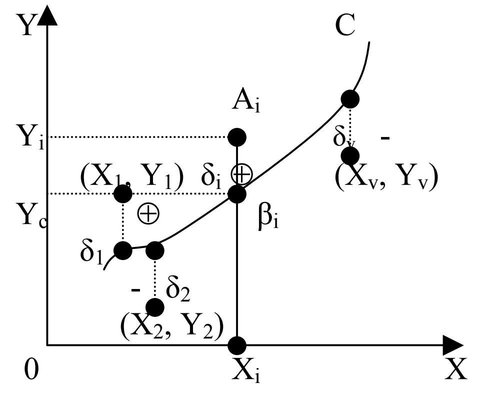


Μερικές φορές είναι καλύτερο να χρησιμοποιούμε διαφορετικούς άξονες ή αλλιώς να μετασχηματίζουμε τις μεταβλητές. Π.χ. αν ο lnY εκφράζεται γραμμικά με τον X, δηλαδή lnY = α + βX τότε μπορούμε να πάρουμε αυτή την ευθεία για προσεγγιστική καμπύλη.

7.2. Η Μέθοδος των ελαχίστων τετραγώνων.

Σ' ένα νέφος σημείων μπορούν να προσαρμοσθούν περισσότερη από μιας καμπύλες μιας ορισμένης μορφής. Για να αποφευχθούν υποκειμενικές τοποθετήσεις για το πόσο καλύτερα προσαρμόζεται ένας συγκεκριμένος τύπος καμπύλης στο νέφος των σημείων καταφεύγουμε στη θεωρία των ελαχίστων τετραγώνων.

Έστω ένα νέφος σημείων (X_1, Y_1), (X_2, Y_2),…., (X_v, Y_v) και C η καμπύλη που αντιστοιχεί:



Σε κάθε τιμή X_i αντιστοιχούν δύο τιμές η παρατήρηση Y_i και η τιμή επί της καμπύλης Y_c , η διαφορά αυτών των δύο τιμών είναι το e_i : $e_i = Y_i - Y_c$ και μπορεί να είναι > 0 ή < 0 ή = 0. Το e_i καλείται **απόκλιση ή σφάλμα ή υπόλοιπο.**

Ένα μέτρο του πόσο καλή είναι η προσαρμογή της C στα σημεία είναι η ποσότητα: $e_1^2 + e_2^2 + ... + e_\nu^2$ να είναι ελάχιστη.

Απ' όλες τις προσεγγιστικές καμπύλες για ένα δεδομένο πλήθος σημείων η καμπύλη με την ιδιότητα: $e_1^2 + e_2^2 + ... + e_\nu^2$ = ελάχιστο είναι η καμπύλη με την καλύτερη προσαρμογή.

Η καμπύλη αυτή καλείται καμπύλη ελάχιστων τετραγώνων (είναι ευθεία, παραβολή κλπ.).

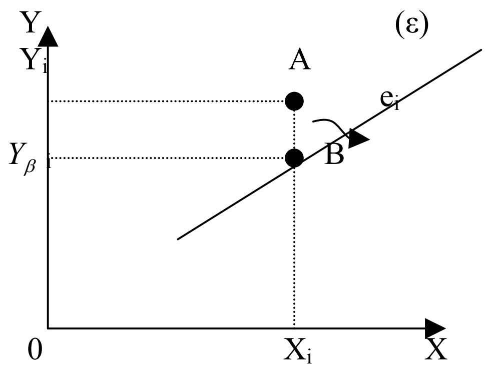
7.3. Η ευθεία των ελαχίστων τετραγώνων (γραμμική παλινδρόμηση).

Έστω ότι τα σημεία είναι τέτοια που μας επιτρέπουν να αναζητήσουμε την εξίσωση μιας ευθείας η οποία περνά όσο γίνεται πιο κοντά απ' αυτά.

Ζητάμε να προσδιορίσουμε τους εκτιμητές $\widehat{a}, \widehat{\beta}$ των συντελεστών α, β της εξίσωσης Y = α + βX, η οποία θεωρούμε ότι υπάρχει στον πληθυσμό από πάρθηκε το δείγμα των σημείων.

Η εξίσωση που παίρνουμε είναι: $\hat{Y} = \widehat{a} + \widehat{\beta}$ X .

Έστω τώρα δύο σημεία A(X_i,Y_i) και B (X_i, Y_β).



Ο υπολογισμός των εκτιμητριών $\widehat{a}, \widehat{\beta}$ των α, β γίνεται ως εξής:

Η απόκλιση του σημείου B επί της ευθείας (ε) από το σημείο A της παρατήρησης είναι:

$e_i = Y_i - Y_\beta = Y_i - a - \beta X_i,$ $\quad εφόσον \quad$ $B_i(X_i, Y_\beta)$ ανήκει στην ευθεία που θέλουμε να προσδιορίσουμε.

Το άθροισμα των τετραγώνων των αποκλίσεων των σημείων Α από τα αντίστοιχα σημεία Β από την ευθεία (ε) θα ισούται με:

$$\Delta = e_1^2 + e_2^2 + \ldots.. + e_v^2 = \sum_{i=1}^{v} e_i^2$$

$$\Delta = (Y_1 - a - \beta X_1)^2 + (Y_2 - a - \beta X_2)^2 + \ldots. + (Y_v - \alpha - \beta X_v)^2 = \sum_{i=1} (Yi - a - \beta X_i)^2$$

Σύμφωνα με την θεωρία των ακρότατων η Δ θα παίρνει το ελάχιστο για τις τιμές $\widehat{a}, \widehat{\beta}$ όπου οι μερικές παράγωγοί της ,ως προς τα α, β, είναι ίσες με μηδέν:

$$\frac{\partial \Delta}{\partial \alpha} = 0 \qquad \kappa\alpha\iota \qquad \frac{\partial \Delta}{\partial \beta} = 0 \Rightarrow \widehat{a}\,,\, \widehat{\beta}$$

Οπότε: $$\left.\begin{array}{ll} \dfrac{\partial \Delta}{\partial \alpha} = 0 & \Rightarrow \displaystyle\sum_{i=1}^{v} Y = va + \beta \sum_{i=1}^{v} Xi \\ \kappa\alpha\iota\, \dfrac{\partial \Delta}{\partial \beta} = 0 & \Rightarrow \displaystyle\sum_{i=1}^{v} XiYi = a \sum_{i=1}^{v} Xi + \beta \sum_{i=1}^{v} Xi^2 \end{array}\right\} \Rightarrow A\pi\acute{o}\ \tau\iota\varsigma$$

σχέσεις αυτές προκύπτουν οι εκτιμήτριες $\widehat{\alpha}$ $\widehat{\beta}$

$$\widehat{\beta} = \frac{v \sum_{i=1}^{v} XiYi - \sum_{i=1}^{v} Xi \sum_{i=1}^{v} Yi}{v \sum_{i=1}^{v} Xi^2 - (\sum_{i=1}^{v} Xi)^2} \kappa\alpha\iota \qquad \widehat{a} = \bar{Y} - \widehat{\beta}\, \bar{X}$$

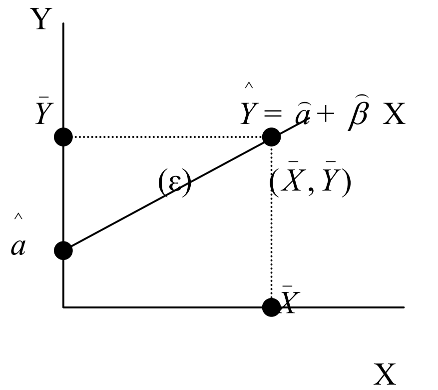
Το $\widehat{\beta}$ μπορεί να γραφεί και ως:

$$\widehat{\beta} = \frac{\sum_{i=1}^{v} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{v} (Xi - \bar{X})^2}, \acute{o}\pi o\upsilon \qquad \bar{X} = \sum_{i=1}^{v} Xi / v, \qquad \bar{Y} = \sum_{i=1}^{v} Yi / v$$

Η ευθεία $\hat{Y} = \hat{a} + \hat{\beta}$ X καλείται **ευθεία παλινδρόμηση** του δείγματος, όπου $\hat{Y}$ οι εκτιμώμενες τιμές της εξαρτημένης μεταβλητής Y στο δείγμα.
Τα σφάλματα που εκτιμώνται από την ευθεία παλινδρόμηση του δείγματος ορίζονται ως $\hat{e}_i = Y_i - \hat{Y}_i$ (παρατήρηση μείον την εκτίμηση)

Η ευθεία ε της γραμμικής παλινδρόμησης περνά από το σημείο $(\bar{X}, \bar{Y})$ του δείγματος.

Η εκτιμήτρια $\hat{a}$ εκφράζει την θέση όπου η ευθεία ε της γραμμικής παλινδρόμησης τέμνει τον Y.



Η εκτιμήτρια $\hat{\beta}$ είναι ο συντελεστής διεύθυνσης της ευθείας ε. Αν $\hat{\beta} > 0$ έχουμε θετική εξάρτηση των X,Y και αν $\hat{\beta} < 0$ έχουμε αρνητική εξάρτηση των X,Y. Τέλος η εκτιμήτρια $\hat{\beta}$ εκφράζει την μεταβολή της y, όταν η X μεταβληθεί κατά μονάδα.

Αν X_1, X_2 τιμές της X με $X_2 = X_1 + 1$ τότε $\hat{Y}_2 - \hat{Y}_1 = \hat{a} + \hat{\beta}X_2 - (\hat{\alpha} + \hat{\beta}X_1) =$
$= \hat{\alpha} + \hat{\beta}X_2 - \hat{\alpha} - \hat{\beta}X_1 = \hat{\beta}(X_1 + 1) - \hat{\beta}X_1 = \hat{\beta}.$

7.4. Μέσο τετραγωνικό σφάλμα - Συντελεστής προσδιορισμού- Συντελεστής συσχέτισης.

Ένα κριτήριο για να ελέγχουμε πόσο καλά ή όχι προσαρμόζεται η ευθεία παλινδρόμησης $\hat{Y} = \hat{a} + \hat{\beta}$X, στο διάγραμμα διασποράς του νέφους των σημείων (x_i,y_i) είναι το εκτιμώμενο **μέσο τετραγωνικό σφάλμα** $\hat{\sigma}^2$.

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{\nu} \hat{e}_i^2}{\nu - 2} = \frac{\sum_{i=1}^{\nu} (Y_i - \bar{Y})^2 - \hat{\beta} \sum_{i=1}^{\nu} (X_i - \bar{X})(Y_i - \bar{Y})}{\nu - 2}$$

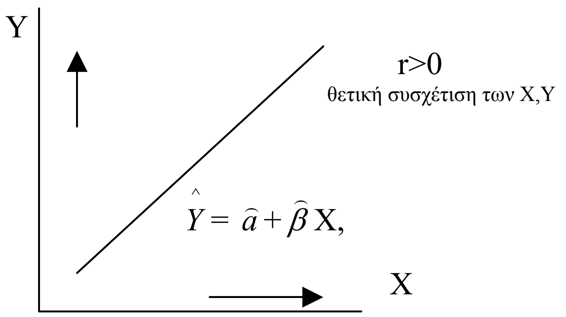
Αν θέσουμε $x_i = X_i - \bar{X}$ και $y_i = Y_i - \bar{Y}$ τότε έχουμε τον τύπο:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{\nu} \hat{e}_i^2}{\nu - 2} = \frac{\sum_{i=1}^{\nu} y_i^{\,2} - \hat{\beta} \sum_{i=1}^{\nu} x_i\, y_i}{\nu - 2}$$

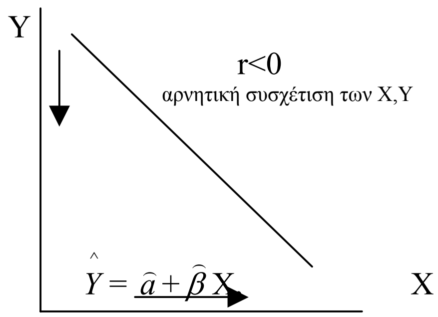
Οι τιμές του $\hat{\sigma}^2$ είναι ένας οποιοσδήποτε θετικός αριθμός και **όσο πιο μικρή είναι η τιμή του τόσο πιο καλύτερη προσαρμογή έχουμε της ευθείας παλινδρόμησης στο διάγραμμα διασποράς του νέφους των σημείων.**

Η έκφραση μικρός αριθμός είναι οπωσδήποτε κάτι το ερμηνεύσιμο υποκειμενικά για τον λόγο αυτό λαμβάνεται ως απόλυτο πλέον κριτήριο βαθμού προσαρμογής της ευθείας παλινδρόμησης ο **συντελεστής προσδιορισμού ή καλής προσαρμογής r^2**.

$$r^2 = \hat{\beta} \frac{\sum_{i=1}^{\nu} x_i y_i}{\sum_{i=1}^{\nu} y_i^2} = \frac{(\sum_{i=1}^{\nu} x_i y_i)^2}{(\sum_{i=1}^{\nu} x_i^2)(\sum_{i=1}^{\nu} y_i^2)} \quad \text{με: } 0 \le r^2 \le 1 .$$



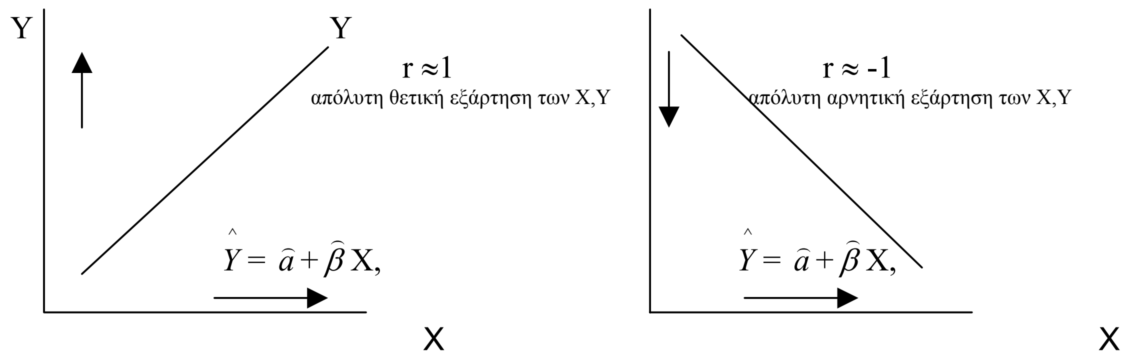
Αυξάνει το Χ αυξάνει και το Υ



Αυξάνει το Χ ελαττώνεται το Υ.

- Όταν το $|r|$ είναι πολύ κοντά στο 1, τότε η διασπορά των τιμών του δείγματος γύρω από την ευθεία παλινδρόμησης είναι μικρή.

- Όταν το $|r|$ είναι πολύ μικρό τότε η διασπορά των τιμών του δείγματος γύρω από την ευθεία παλινδρόμησης είναι μεγάλη.



Επίσης χρησιμοποιείται και ο **συντελεστής συσχέτισης** που είναι ο $r = \sqrt{r^2}$.

Όσο πιο κοντά στην μονάδα βρίσκεται ο συντελεστής προσδιορισμού τόσο πιο καλή προσαρμογή έχουμε την ευθεία παλινδρόμησης στα δεδομένα.

Παρατήρηση: Όταν $r^2 = 1$ τότε η ευθεία παλινδρόμησης $\hat{y} = \hat{\alpha} + \hat{\beta}\chi$ περνά απ' όλα τα σημεία (xi,yi) του νέφους του διαγράμματος διασποράς.

Παρατήρηση: Ο συντελεστής r^2 φανερώνει το ποσοστό των τιμών της μεταβλητής X που ερμηνεύονται από τις μεταβολές της μεταβλητής Y.

Π.χ. $r^2 = 0{,}78$ σημαίνει ότι το 78% της μεταβλητότητας της μεταβλητής Y οφείλεται στη σχέση που έχει αναπτυχθεί μεταξύ των X, Y.

Παραδείγματα

1. Δίνεται ο πίνακας των τιμών των δύο μεταβλητών X, Y ως προς ένα δείγμα.

Να προσαρμοσθεί η ευθεία ελαχίστων τετραγώνων στα δεδομένα αυτά.

Πίνακας

	X_i	**Y_i**	**$X_i\,Y_i$**	**X_i^2**	**ΣYi^2**
1	1	1	1	1	1
2	3	2	6	9	4
3	4	4	16	16	16
4	6	4	24	36	16
5	8	5	40	64	25
6	9	7	63	81	49
7	11	8	88	121	64
8	14	9	126	196	81

$$\sum Xi = 56 \qquad \sum Yi = 40 \qquad \sum XiYi = 364 \qquad \sum Xi^2 = 524 \qquad \sum Yi^2 = 256$$

$$\widehat{\beta} = \frac{v\sum_{i=1}^{v} XiYi - \sum_{i=1}^{v} Xi \sum_{i=1}^{v} Yi}{v\sum_{i=1}^{v} Xi^2 - \left(\sum_{i=1}^{v} Xi\right)^2} = \frac{8\cdot 364 - 56\cdot 40}{8\cdot 524 - 56^2} = 0{,}636$$

$$\bar{X} = \frac{56}{8} = 7, \qquad \bar{Y} = \frac{40}{8} = 5$$

$$\widehat{\alpha} = \bar{Y} - \widehat{\beta}\,\bar{X} = 5 - 0{,}636\cdot 7 = 0{,}548$$

άρα $\hat{Y} = 0{,}548 + 0{,}636X$ η ευθεία γραμμικής παλινδρόμησης της μεταβλητής Y πάνω στην X.

Αν ζητήσουμε την γραμμική παλινδρόμηση της μεταβλητής X πάνω στην Y τότε θα αλλάξουμε τους ρόλους των μεταβλητών, το Y θα γίνει X και το X θα γίνει Y. Αυτό δείχνεται στο πίνακα.

Πίνακας

	X_i	Y_i	$X_i\,Y_i$	X_i^2	Yi^2
1	1	1	1	1	1
2	2	3	6	4	9
3	4	4	16	16	16
4	4	6	24	16	36
5	5	8	40	25	64
6	7	9	63	49	81
7	8	11	88	64	121
8	9	14	126	81	196

$\sum Xi = 40 \quad \sum Yi = 56 \quad \sum XiYi = 364 \quad \sum Xi^2 = 256 \quad \sum Yi^2 = 524$

$\bar{X} = 5$

$\bar{Y} = 7$

$$\widehat{\beta} = \frac{8\cdot 364 - 40\cdot 56}{8\cdot 256 - 40^2} = \frac{2912 - 2240}{2048 - 1600} = 1{,}5$$

$\widehat{\alpha} = 7 - 1{,}5\cdot 5 = -0{,}5$

άρα $\hat{Y} = -0{,}5 + 1{,}5X$

Παρατήρηση:

Βλέπουμε ότι οι δύο ευθείες παλινδρόμησης της Y πάνω στην X και της X πάνω στην Y είναι διαφορετικές: $\hat{Y} = 0{,}548 + 0{,}636X$ και $\hat{Y} = 0{,}5 + 1{,}5X$.

2. Στο παράδειγμα 1 να προσδιορισθεί το εκτιμώμενο **μέσο τετραγωνικό σφάλμα** $\hat{\sigma}^2$. Επίσης να προσδιορισθεί ο συντελεστής καλής προσαρμογής της ευθείας παλινδρόμησης ή ο συντελεστής προσδιορισμού r^2, ο συντελεστής συσχέτισης r και να γίνει το διάγραμμα διασποράς με την ευθεία παλινδρόμησης.

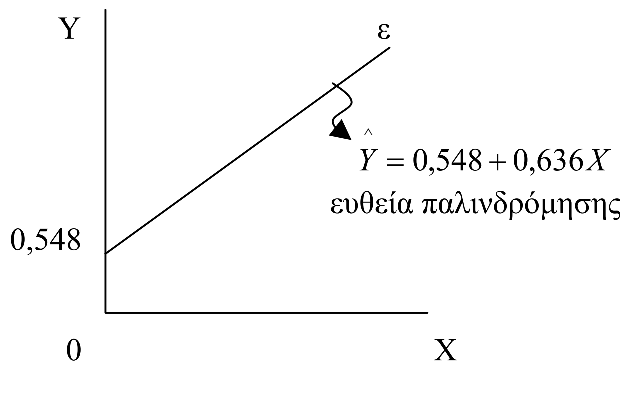
$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{\nu} \hat{e}_i^2}{\nu-2} = \frac{\sum_{i=1}^{\nu} y_i^{\ 2} - \hat{\beta}\sum_{i=1}^{\nu} x_i\, y_i}{\nu-2} = \frac{132-0,636*84}{6} = 13,10$$

Όπου $\sum y_i^2 = 132,\ \sum x_i y_i = 84,\ \sum x_i^2 = 56$

$$r^2 = \hat{\beta}\frac{\sum_{i=1}^{\nu} x_i y_i}{\sum_{i=1}^{\nu} y_i^2} = \frac{(\sum_{i=1}^{\nu} x_i y_i)^2}{(\sum_{i=1}^{\nu} x_i^2)(\sum_{i=1}^{\nu} y_i^2)} = \frac{84^2}{56*132} = 0,95$$

Αυτό δείχνει ότι η ευθεία παλινδρόμησης, $\hat{Y} = 0,548 + 0,636X$ έχει μια πολύ καλή προσαρμογή στα δεδομένα του προβλήματος.

Ο συντελεστής συσχέτισης είναι r=0,97.



2. Δίνεται η τιμή ενός προϊόντος σε χιλιάδες δραχμές (X) και η ζητούμενη ποσότητα του προϊόντος αυτού σε τόνους (Y).

X:	500	400	300	600	450	350	410	550
Y:	8	10	20	7	12	25	15	6

α. Να προσδιορισθεί η εξίσωση γραμμικής παλινδρόμησης της Y πάνω στη X.

β. Να κατασκευασθεί η ευθεία γραμμικής παλινδρόμησης και το διάγραμμα διασποράς.

γ. Να βρεθούν οι συντελεστές συσχέτισης και προσδιορισμού.

δ. Να εξαχθεί το συμπέρασμα αν η ευθεία παλινδρόμησης προσαρμόζεται καλά στα δεδομένα.

X_i	Y_i	$X_i Y_i$	X_i^2	Yi^2
500	8	4000	250.000	64
400	10	4000	160.000	100
300	20	6000	90.000	400
600	7	4200	360.000	49
450	12	5400	202.500	144
350	25	8750	122.500	625
410	15	6150	168.100	225
550	6	3300	302.500	36

$\sum Xi = 3560$, $\sum Yi = 103$, $\sum XiYi = 41.800$, $\sum Xi^2 = 1.655.600$, $\sum Yi^2 = 1643$

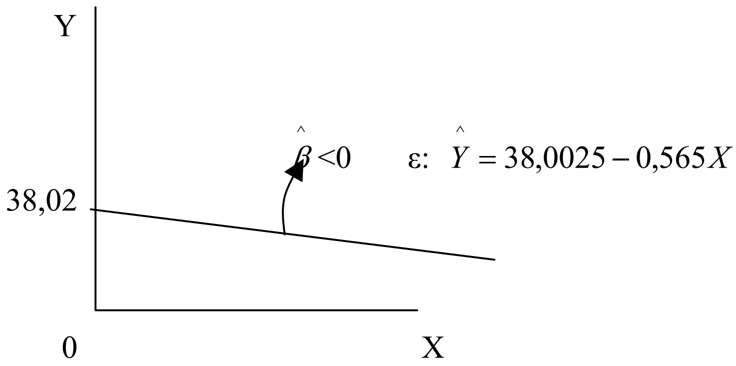
$$\widehat{\beta} = \frac{8 \cdot 41.800 - 3560 \cdot 103}{8 \cdot 1.655.600 - (3560)^2} = \frac{334.400 - 366.680}{13.244.800 - 12.673.600} = \frac{-32.280}{571.200} = -0,0565$$

$$\bar{X} = \frac{3560}{8} = 445 \qquad \bar{Y} = \frac{103}{8} = 12,88$$

$$\widehat{\alpha} = 12,88 + 0,0565 \cdot 445 = 38,0225$$

Συνεπώς έχουμε: $\hat{Y} = 38,0225 - 0,0565X$

β.



γ. Για τον συντελεστή προσδιορισμού έχουμε:

$$r^2 = \hat{\beta}\frac{\sum_{i=1}^{\nu} x_i y_i}{\sum_{i=1}^{\nu} y_i^2} \;=\; (-0,0565)\frac{(-2069,62)}{316,84} = 0,37$$

Ο συντελεστής συσχέτισης r είναι - 0,61

Ο συντελεστής συσχέτισης r μπορεί να υπολογισθεί και μέσω του τύπου:

$$r = \frac{\nu\sum xiyi - \sum xi \sum yi}{\sqrt{\nu\sum xi^2 - \left(\sum xi\right)^2}\,\sqrt{\nu\sum yi^2 - \left(\sum yi\right)^2}}$$

ή του τύπου: $r = \dfrac{\sum (xi-\overline{x})(yi-\overline{y})}{\sqrt{\sum (xi-\overline{x})^2 \cdot \sum (yi-\overline{y})^2}}$

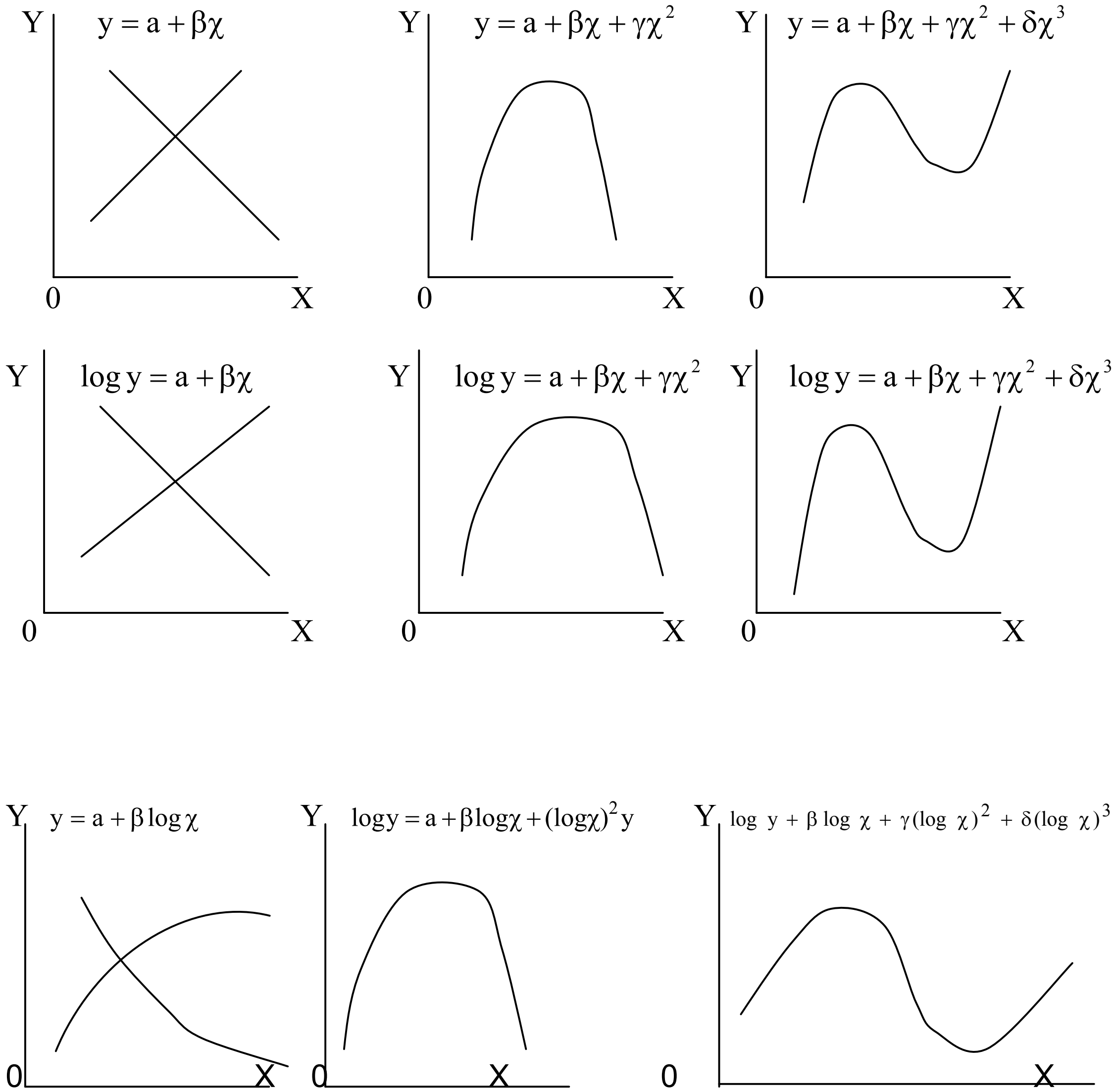
δ. Επειδή ο συντελεστής προσδιορισμού $r^2 = 0{,}37$ τούτο σημαίνει ότι μόνο ποσοστό 37% των μεταβολών της ζητούμενης ποσότητας Y ερμηνεύεται από τις μεταβολές της τιμής X του προϊόντος μέσω της παλινδρόμησης.

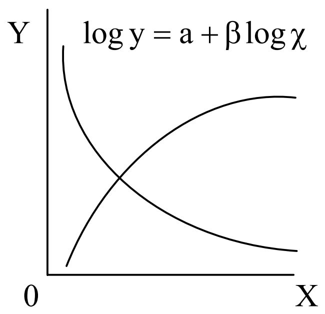
Η ευθεία παλινδρόμησης δεν έχει καλή προσαρμογή στα δεδομένα.

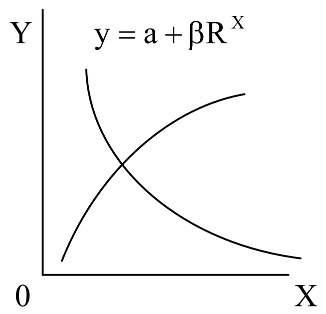
Επίσης επειδή $r = -0{,}61 < 0$ έχουμε αρνητική εξάρτηση των X,Y δηλαδή σε αύξηση του X αντιστοιχεί μείωση του Y και αντίστροφα.

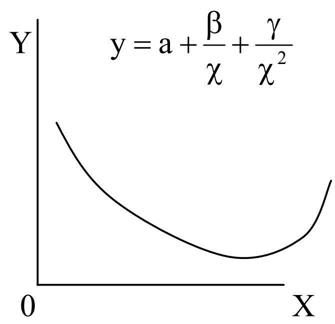
Τέλος επειδή $|r|$ αρκετά μεγάλο $|r| = 0{,}61$ αυτό σημαίνει ότι η διασπορά των τιμών του δείγματος είναι μικρή γύρω από την ευθεία παλινδρόμησης.

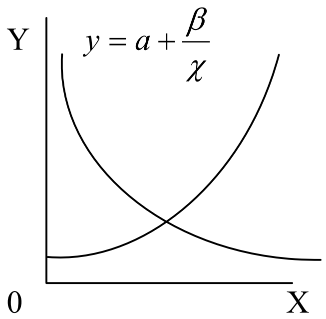
7.5. Μοντέλα απλής παλινδρόμησης (γραμμικά – καμπυλόγραμμα)











7.6. Μοντέλα απλής παλινδρόμησης που ανάγονται σε γραμμικά.

1. Το λογαριθμικό μοντέλο ή μοντέλο Cobb-Douglas:

$y = ax^{\beta}, \quad x, y > 0$

Λογαριθμίζω (με δεκαδικό λογάριθμο).

$\log y = \log a + \beta \log x$

θέτω $\log y = y'$, $\quad \log x = x'$, $\quad \log a = a'$

γραμμικό μοντέλο $\leftarrow$ $y' = a' + \beta x'$

Τα $\widehat{a}', \widehat{\beta}$ είναι εκτιμητές των α΄, β αντίστοιχα, μέσω της μεθόδου των ελαχ ίστων τετραγώνων.

Από το $\widehat{\alpha}'$ προκύπτει και το α΄ άρα και το $a : \widehat{\alpha} = 10^{\widehat{\alpha}'}$

2. Το αντίστροφο μοντέλο:

$$y = a + \frac{\beta}{\chi}, \qquad x \neq 0$$

θέτω $\frac{1}{\chi} = x' \rightarrow y = \alpha + \beta x' \rightarrow$ γραμμικό μοντέλο.

3. Το εκθετικό μοντέλο:

$y = a \cdot e^{\beta\chi} \rightarrow$ λογαριθμίζω (με νεπέρειο λογάριθμο).

$\ln y = \ln a + \beta\chi$ θέτω $\ln y = y', \ln a = a'$

και έχω $y' = a' + \beta\chi \rightarrow$ γραμμικό μοντέλο

4. Το ημιλογαριθμικό ως προς Χ μοντέλο:

$x = a\beta^{y}, \quad x > 0 \quad \beta > 0$

Λογαριθμίζω (με δεκαδικό λογάριθμο)

$\log x = \log a + y \log\beta$

$y = (\log x - \log a)/\log\beta$

$$\Rightarrow y = -\frac{\log a}{\log\beta} + \frac{1}{\log\beta} \cdot \log x$$

$$\Rightarrow y = \alpha' + \beta' \log x (\mathrm{I}), \quad a' = -\frac{\log a}{\log\beta}, \beta' = \frac{1}{\log\beta}$$

Οι εκτιμητές των α, β οι $\widehat{\alpha}, \widehat{\beta}$ προκύπτουν
Από τις σχέσεις $\widehat{\beta} = 10^{1/\widehat{\beta}'}, \quad \widehat{\alpha} = 10^{-\alpha'/\widehat{\beta}'}$

Όπου $\widehat{\alpha}, \widehat{\beta}$ οι εκτιμητές των α΄, β΄ από τη σχέση (Ι)

5. Το μοντέλο των Mitsherlish Spillman:

$y = a + \beta R^{X}, \quad 0 < R < 1$ γνωστή σταθερά

θέτω $x' = R^{X} \rightarrow y = a + \beta x' \rightarrow$ γραμμικό μοντέλο.

6. Το πολυωνυμικό μοντέλο:

Μορφές αυτού:

$$y = a + \beta x + \gamma x^2$$

$$\log y = a + \beta x + \gamma x^2, \qquad y > 0$$

$$\log y = a + \beta(\log x) + \gamma(\log x)^2, \qquad x, y > 0$$

$$y = \alpha + \beta x + \gamma x^2 + \delta x^3$$

$$\log y = \alpha + \beta x + \gamma x^2 + \delta x^3 \quad , \quad y > 0$$

$$\log y = \alpha + \beta(\log x) + \gamma(\log x)^2 + \delta(\log x)^3 \quad , \quad x, y > 0$$

$$y = a + \frac{\beta}{x} + \frac{\gamma}{x^2} \qquad x \neq 0$$

όλες οι μορφές αυτές μετατρέπονται σε γραμμικό πολυμεταβλητό μοντέλο:

$$y = a + \beta x_1 + \gamma x_2 + ...$$

Κεφάλαιο 8ο

Χρονολογικές Σειρές

8.1. Εισαγωγή

Είναι αναγκαίο οι διευθύνοντες μιας επιχείρησης να προγραμματίζουν τις επιχειρηματικές τους δραστηριότητες στηριζόμενοι στην εμπειρία του παρελθόντος και στην πρόβλεψη του μέλλοντος.

Αυτό θα τους οδηγήσει εκτός απροόπτων καταστάσεων σε μια σωστή απόφαση για το μέλλον της επιχείρησής τους, διότι θα στηριχθούν στην γνώση του παρελθόντος και θα κινηθούν με την μελέτη των μακροχρόνιων τάσεων που παρουσιάζει η αγορά την δεδομένη περίοδο.

Παραδείγματος χάριν αν η πρόβλεψη για το μέλλον μιλά για μείωση του πληθυσμού κατά 10% στην επόμενη 5ετία ασφαλώς και θα πρέπει να ληφθεί υπόψιν για την μελλοντική ζήτηση των αγαθών.

Η σπουδαιότητα της πρόβλεψης στην λήψη ορθών αποφάσεων στον προγραμματισμό μιας επιχείρησης και όχι μόνο, αλλά και αυτών ακόμη των κυβερνήσεων, αποτέλεσε σοβαρό στοιχείο στο να αναπτυχθεί μια Στατιστική μεθοδολογία που μελετά το παρελθόν και κάνει προβλέψεις για το μέλλον.

Η Στατιστική μεθοδολογία αυτή ονομάζεται **«Ανάλυση Χρονολογικών Σειρών».**

Ορισμός: **Χρονολογική σειρά ή χρονοσειρά** είναι η σειρά των τιμών τις οποίες παίρνει μια μεταβλητή σε διαδοχικά χρονικά διαστήματα.

Παραδείγματα χρονολογικών σειρών

Η ετήσια παραγωγή σταριού, καπνού, βαμβακιού, ελαιόλαδου κλπ. τα τελευταία 15 χρόνια στην Β. Ελλάδα ή στην χώρα ολόκληρη.

Οι ετήσιες παραγωγές (ποσότητες) μιας επιχείρησης τα τελευταία 15 χρόνια.

Οι τραπεζικές καταθέσεις την τελευταία 20ετία στην Ελλάδα.

Οι δείκτες της βιομηχανικής παραγωγής την τελευταία 20ετία

Οι δημογραφικές μεταβολές στην Ελλάδα ανά έτος για την τελευταία 20ετία κλπ.

Μια χρονολογική σειρά συμβολίζεται με ένα γράμμα π.χ. y με δείκτη το t (χρόνος): $\{ y_t \}$

Αν δηλαδή έχουμε τον δείκτη τιμών καταναλωτή τα έτη 1984-1994, η χρονολογική σειρά συμβολίζεται με

$\{ y_t \}$ όπου $t = 0, 1, 2, \ldots\ldots, 11$ με $t = 0$ την 30/6/84 ή 1/7/84

Η χρονολογική σειρά y_t είναι μια μεταβλητή στο χρόνο που μπορεί να είναι συνεχή (θερμοκρασία περιβάλλοντος) ή ασυνεχής (συλλογή γεωργικών προϊόντων που γίνεται σε συγκεκριμένους μήνες του έτους).

Κάποιες χρονολογικές σειρές δίνονται αθροιστικά σε ορισμένες χρονικές περιόδους όπως η κατανάλωση του ηλεκτρικού ρεύματος για βιομηχανική ή οικιακή χρήση, ο όγκος των εισαγωγών ή των εξαγωγών ανά έτος t (που είναι το άθροισμα των όγκων των επιμέρους 12 μηνών).

Άλλες χρονολογικές σειρές ορίζονται ως μέσοι όροι (ετήσιοι) όπως οι οικονομικοί δείκτες που προσδιορίζονται ως οι μέσοι όροι των 12 μηνιαίων δεικτών.

Επίσης παίζει ρόλο πότε θα παίρνετε ο μέσος όρος των μηνών ή του έτους, και συνήθως λαμβάνεται το μέσο του μήνα ή του έτους (μεταξύ 30/6 και 1/7).

Όταν αναφερόμαστε στον κάθε μήνα σε μια χρονολογική σειρά πρέπει να παίρνουμε το μέσο όρο για τον κάθε μήνα (εφόσον έχουμε διαφορετικό αριθμό σε κάθε μήνα).

Όταν έχουμε χρονολογική σειρά που αναφέρεται σε χρηματικά ποσά (π.χ. εισόδημα αγροτικό) πρέπει να μετατρέπονται αυτά τα ποσά των διαφόρων ετών σε συγκρίσιμα ποσά δηλαδή να γίνεται ο αποπληθωρισμός αυτών με ένα έτος βάσης που θα λαμβάνεται εξ' αρχής.

Όταν οι χρονολογικές σειρές αναφέρονται σε εισόδημα ή κατανάλωση πρέπει να λαμβάνεται υπόψη και η αυξομείωση του πληθυσμού της χώρας ή της περιοχής όπου αναφερόμαστε. Γι΄ αυτό θα πρέπει η χρονολογική σειρά να εκφράζεται σε μεγέθη κατά κεφαλή.

8.2. Συνιστώσες των χρονολογικών σειρών.

Η κάθε χρονολογική σειρά παρουσιάζει μεταβολές – διακυμάνσεις στο χρόνο αυτές μπορεί να είναι:

i) **Μακροχρόνιες τάσεις**

ii) **Κυκλικές κυμάνσεις**

iii) **Εποχικές κυμάνσεις**

και iv) **Τυχαίες ή άρρυθμες κινήσεις**

Οι κυκλικές και οι εποχικές κυμάνσεις καλούνται και **περιοδικές κυμάνσεις**.

i) **Μακροχρόνια τάση ή απλά τάση**

Είναι η ομαλή κίνηση στο χρόνο που εμφανίζει μια χρονολογική σειρά η οποία μπορεί να είναι ανοδική-καθοδική ή σύνθετη. Είναι σαν μια δύναμη που οδηγεί την χρονολογική σειρά (ανοδικά ή καθοδικά) και εξαρτάται από ποικίλους παράγοντες που μπορεί να είναι η πληθυσμιακή αύξηση σε μια περιοχή – χώρα, η τεχνολογική εξέλιξη, η αύξηση του βιοτικού επιπέδου των κατοίκων μιας περιοχής κλπ.

Παραδείγματος χάριν η χρονολογική σειρά που αναφέρεται στους όγκους των καταναλωτικών αγαθών π.χ. (τροφίμων) σε μια δυτική κοινωνία οφείλεται αφενός μεν στην αύξηση του βιοτικού επιπέδου των κατοίκων (αύξηση εισοδήματος) αλλά και στην αύξηση του καταναλωτισμού με την αλλαγή καταναλωτικών συνηθειών.

Η τάση των χρονολογικών σειρών που αναφέρονται στην παραγωγή αγροτικών προϊόντων οφείλεται στην τεχνολογική πρόοδο που βελτίωσε τις μεθόδους καλλιέργειας και συλλογής των αγροτικών προϊόντων, στην καλύτερη οργάνωση των επιχειρήσεων που ασχολούνται με την αγροτική παραγωγή και μεταποίηση των αγροτικών προϊόντων.

Μια μακροχρόνια τάση ασφαλώς μπορεί να παρουσιάζει και μείωση και αυτό μπορεί να οφείλεται είτε σε πληθυσμιακές μεταβολές (μειώσεις πληθυσμού) είτε σε ανταγωνισμό (κάποιο νέο προϊόν ανταγωνιστικό μπήκε στην αγορά και εκτόπισε το προϋπάρχον) και επομένως στην μείωση της ζήτησης του εν λόγω προϊόντος.

ii. Κυκλική κύμανση

Οι κινήσεις αυτές έχουν διάρκεια μεγαλύτερη του έτους και εμφανίζουν ανοδική κίνηση και κατόπιν καθοδική.

Οι κυκλικές κυμάνσεις μπορεί να έχουν περίοδο από δύο, τρία, και περισσότερα χρόνια.

Σ’ αυτές εντάσσονται και οι γνωστές στην Οικονομία. **Οικονομικές διακυμάνσεις (οικονομικός ή επιχειρηματικός κύκλος)**. Όλα τα μακροοικονομικά μεγέθη (απασχόληση, επενδύσεις - εθνικό εισόδημα κλπ.) έχουν κυκλική κύμανση.

Επίσης κυκλική κύμανση παρουσιάζουν οι χρονολογικές σειρές που αναφέρονται στην ετήσια παραγωγή ορισμένων γεωργικών προϊόντων (σιτάρι-αραβόσιτος – ελαιόλαδο – καπνός - βαμβάκι κλπ.) και έχουν περίοδο κύμανσης 2,3,4, ή 5 χρόνων.

Οι κυριότεροι παράγοντες που επηρεάζουν τις χρονολογικές σειρές και προκαλούν κυκλικές κυμάνσεις πρέπει να θεωρηθούν η ανισορροπία μεταξύ προσφοράς και ζήτησης οι μετεωρολογικές συνθήκες, οι τεχνολογικές εξελίξεις, τα δυσάρεστα πολιτικά γεγονότα κλπ.

iii) Εποχιακή κύμανση

Διαρκεί ένα έτος συνήθως και εκεί εμφανίζονται οι ανοδικές και οι καθοδικές κινήσεις της χρονολογικής σειράς.

Καθοριστικό ρόλο για αυτές τις κινήσεις παίζουν η εποχιακή μεταβολή, οι κλιματολογικές συνθήκες, οι εορτές κλπ.

iv) Τυχαίες ή άρρυθμες κινήσεις

Οι κινήσεις αυτές είναι μικρές ή μεγάλες, θετικές ή αρνητικές και οφείλονται σε σοβαρά και απρόσμενα γεγονότα.

Θεομηνίες, επαναστάσεις, απεργίες, έκτακτα οικονομικά μέτρα κλπ.

Αν ορίσουμε T την μακρόχρονη τάση μιας χρονολογικής σειράς, S των εποχικών κυμάνσεων, C των κυκλικών κυμάνσεων και I των τυχαίων κινήσεων.

Τότε μπορούμε θα θεωρήσουμε ότι οι τιμές της χρονολογικής σειράς y_t δίνονται i) πρόσθετα από τις τιμές των συνιστωσών ή ii) πολλαπλασιαστικά

Δηλαδή: $y_t = T + S + C + I$

ή $y_t = T \cdot S \cdot C \cdot I$

8.3. Στατιστικός προσδιορισμός της τάσης

Ο προσδιορισμός της τάσης μιας χρονολογικής σειράς που συνήθως θα είναι η μια ευθεία ή μια καμπύλη γραμμή μας επιτρέπει να αναλύσουμε αυτή στις συνιστώσες της και να κάνουμε πρόβλεψη των όρων της σειράς.

Για τον στατιστικό προσδιορισμό της τάσης υπάρχουν οι εξής μέθοδοι:

i) Χάραξη της τάσης με το χέρι
ii) Μέθοδος των μέσων σημείων
iii) Μέθοδος των κινητών μέσων όρων
iv) Μέθοδος των ελαχίστων τετραγώνων

i) Χάραξη με το χέρι - γραφική μέθοδος:

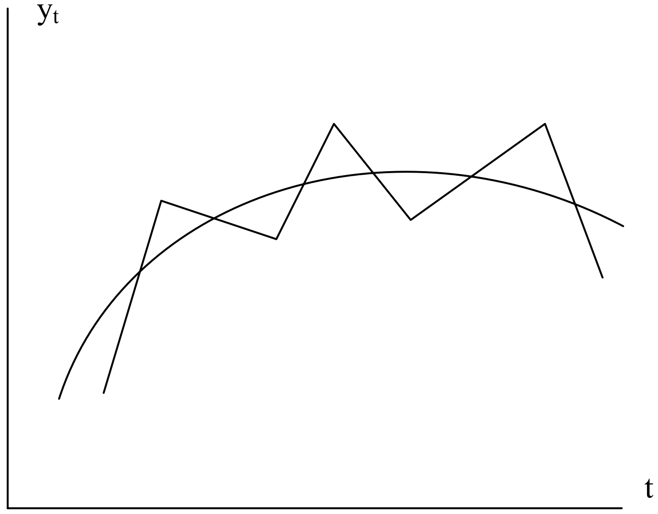
Ο ερευνητής με μεγάλη προσοχή χαράζει με το χέρι μια γραμμή ελεύθερα πάνω στο χρονόγραμμα της χρονολογικής σειράς η οποία να περνά ανάμεσα από την πολυγωνική γραμμή της χρονολογικής σειράς και να ακολουθείται η μακροχρόνια κίνηση της χρονολογικής σειράς.

- ***Τα πλεονεκτήματα της μεθόδου:***

Είναι απλή, δεν χρησιμοποιούμε μαθηματική εξίσωση αν γίνει προσεκτικά η γραμμή δίνει πολύ καλά αποτελέσματα και προσεγγίζει και την μαθηματική εξίσωση της τάσης της χρονολογικής σειράς.

- ***Μειονεκτήματα της μεθόδου:***

Είναι υποκειμενική, απαιτείται μεγάλη εξάσκηση.
Χρησιμοποιείται πολύ περιορισμένα.



Γραφική μέθοδος

ii) Μέθοδος των μέσων σημείων.

α) Αν από τη γραφική απεικόνιση της χρονολογικής σειράς διαπιστωθεί ότι έχουμε μια γραμμική τάση τότε οι τιμές της χρονολογικής σειράς χωρίζονται σε δύο ομάδες. Αν ο αριθμός των όρων της χρονολογικής σειράς είναι άρτιος τότε ο χωρισμός είναι εύκολος, αν είναι ο αριθμός των όρων περιττός παραλείπουμε το μεσαίο όρο και σχηματίζουμε δύο ομάδες με τους λοιπούς όρους.

Υπολογίζονται οι μέσοι όροι της μεταβλητής y και της t και προσδιορίζεται η ευθεία που περνά από τα σημεία $(t_1\ y_1)$, $(t_2\ y_2)$ όπου

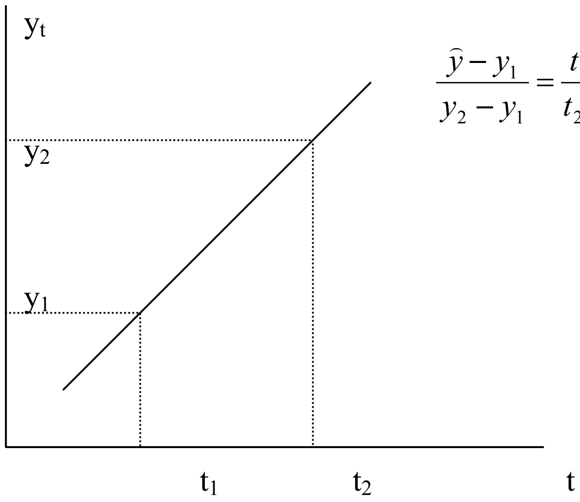
y_1 ο μέσος όρος των δεδομένων της 1ης ομάδας

y_2 ο μέσος όρος των δεδομένων της 2ης ομάδας

t_1 ο μέσος όρος της χρονικής περιόδου της 1ης ομάδας

t_2 ο μέσος όρος της χρονικής περιόδου της 2ης ομάδας

Έχουμε: $\widehat{y} = \dfrac{y_2 - y_1}{t_2 - t_1} \cdot (t - t_1) \Rightarrow \dfrac{\widehat{y} - y_1}{y_2 - y_1} = \dfrac{t - t_1}{t_2 - t_1}$



$$\frac{\widehat{y}-y_1}{y_2-y_1}=\frac{t-t_1}{t_2-t_1}$$

Γραμμική τάση, χρονολογικής σειράς

β) Αν από την γραφική απεικόνιση προκύψει ότι η χρονολογική σειρά έχει καμπυλόγραμμη τάση παραβολή 2^{ου} βαθμού της μορφής $\widehat{y}=\widehat{\alpha}+\widehat{\beta}t+\widehat{\gamma}t^2$, τότε οι τιμές της χρονολογικής σειράς υποδιαιρούνται σε τρεις κατά το δυνατόν ισοπληθείς ομάδες και υπολογίζονται οι αντίστοιχοι τρεις μέσοι όροι αυτών $(t_1,y_1),(t_2,y_2),(t_3,y_3)$.

Οι εκτιμήτριες $\widehat{a},\widehat{\beta},\widehat{\gamma}$ στην σχέση $\widehat{y}=\widehat{\alpha}+\widehat{\beta}t+\widehat{\gamma}t^2$ προσδιορίζονται από την λύση του συστήματος.

$$y_1=\alpha+\beta t_1+\gamma t_1^2$$

$$y_2=\alpha+\beta t_2+\gamma t_2^2$$

$$y_3=\alpha+\beta t_3+\gamma t_3^2$$

Η μέθοδος των μέσων σημείων εφαρμόζεται συνήθως σε γραμμικές τάσεις ή το πολύ σε τάσεις 2^{ου} βαθμού.

Παράδειγμα:

Δίνεται η κατανάλωση ενός αγροτικού προϊόντος τα έτη 1986-1999 σε χιλιάδες τεμάχια ότι είναι η ακόλουθη:

Έτη	Κατανάλωση (y)		Μεταβλητή χρόνου (x)
1986	75		0
1987	90		1
1988	90		2
1989	85	$\frac{700}{7} = 100 = y_1$	3 $= x_1$ *
1990	105		4
1991	125		5
1992	130		6
1993	135		7
1994	140		8
1995	150		9
1996	170	$\frac{1155}{7} = 165 = y_2$	10 $= x_2$ *
1997	195		11
1998	185		12
1999	180		13

Έχουμε τα έτη 1989 με τιμή τάσης 100
1996 με τιμή τάσης 165

Η ευθεία γραμμή περνά από τα σημεία (3,100) και (10,165) ή (1989,100) και (1996,165).

Η κλίση της ευθείας είναι:

$$\frac{y_2 - y_1}{x_2 - x_1} = \frac{165 - 100}{10 - 3} = \frac{65}{7} = 9{,}2857$$

έχουμε επομένως την ευθεία τάσης:

$$\hat{y} = 9{,}2857x + 72{,}143 = 72{,}143 + 9{,}2857x$$

iii) Μέθοδος των κινητών μέσων όρων

Η μέθοδος χρησιμοποιείται για να εξαλειφθούν οι κυκλικές και εποχιακές κυμάνσεις, όπως και οι τυχαίες επιδράσεις.

Βασική προϋπόθεση που τίθεται είναι ότι η περίοδος του κύκλου (ή της περιοδικής κίνησης) είναι σταθερή.

Η διαδικασία αναφέρεται μέσω του παρακάτω παραδείγματος:

Α. περίπτωση: μήκος κύκλου= περιττό πλήθος ετών.

Υπολογισμός κινητών μέσων όρων 5 ετών της χρονολογικής σειράς y_t = "παραγωγή βάμβακος στην Ελλάδα (1970-1981)" για να εκτιμηθεί η γραμμή τάσης.

Έτη	y_t		Κινητά αθροίσματα		Κινητοί μέσοι όροι
1970	308		-----		-----
1971	330		-----		-----
1972	395	+ →	1764*	: 5 →	352,8*
1973	361	+ →	1822	: 5 →	364,4
1974	370		1821	: 5 →	364,2
1975	366		1808		361,6
1976	329		1855		371
1977	382		1810		362
1978	408		1801		360,2
1979	325		1857		371,4
1980	357		1786		357,2
1981	385		-----		-----
1982	311		-----		-----

* 1764=308+330+395+361+370
* 352,8=1764:5

Αν ενώσουμε τους κινητούς μέσους όρους δηλαδή τα σημεία (72, 352,8),(73,364,4), κ.ο.κ.

Παίρνουμε την ευθεία τάσης των κινητών μέσων 5 - ετιών που εκφράζει την ευθεία τάσης της χρονολογικής σειράς.

Β. Περίπτωση: μήκος κύκλου= άρτιο πλήθος ετών.

Υπολογισμός κινητών μέσων όρων τεσσάρων ετών της χρονολογικής σειράς y_t = παραγωγή ελαιόλαδου (1960-1981) σε χιλ. τόνους ΕΣΥΕ.

Έτος	y_t	Κινητά αθροίσματα	Κινητοί μέσοι	Άθροισμα. Κινητ. μέσων	Κινητοί μέσοι
1970	198	-----	-----	-----	-----
1971	183	-----	-----	-----	-----
		858	214,5		
1972	262			449,75	224,88
		901	235,25		
1973	215			481	240,5
		983	245,75		
1974	241			488,75	244,38
		972	243		
1975	265			497,75	248,88
		1019	254,75		
1976	251			516,5	258,25
		1047	261,75		
1977	262			526,25	263,13
		1058	264,8		
1978	269			571,75	285,88
		1229	307,25		
1979	276			641,5	320,75
		1337	334,25		
1980	422			691,25	345,6
		1428	357		
1981	370	-----	-----	-----	-----
1982	360	-----	-----	-----	-----

Το μειονέκτημα της εν λόγω μεθόδου είναι ότι δεν είναι δυνατόν να υπολογιστούν στην αρχή και στο τέλος τα κινητά μέσα, χάνονται δηλαδή παρατηρήσεις και εφόσον ο αριθμός των παρατηρήσεων είναι σχετικά μικρός η απώλεια αυτή είναι σημαντική. Δεν απαλείφονται οι τυχαίες επιδράσεις (παραμένουν οι πολύ μικρές ή πολύ μεγάλες τιμές) και τελικά και επηρεάζουν τους μέσους όρους.

Η μέθοδος αυτή δεν προσφέρεται για προβλέψεις στο μέλλον.

iv.Μέθοδος των ελαχίστων τετραγώνων:

α: Προσδιορισμός γραμμικής τάσης μιας χρονολογικής σειράς.

Αν οι τιμές μιας χρονολογικής σειράς y_t σχηματίζουν αριθμητική πρόοδο τότε για να εκφρασθεί η τάση χρησιμοποιούμε την γραμμική μορφή: $\widehat{y}_t = \widehat{a} + \widehat{\beta} t$

Για να προσδιορισθούν τα $\widehat{a}, \widehat{\beta}$ που είναι εκτιμήτριες των α, β λύνουμε το σύστημα:

$$\left.\begin{aligned} \sum y_i &= na + \beta \sum_i t_i \\ \sum t_i y_i &= a \sum t_i + \beta t_i^2 \end{aligned}\right\} \rightarrow \widehat{a}, \widehat{\beta}$$

Αν διατάξουμε τα δεδομένα μας ώστε $\Sigma t_i = 0$ (λαμβάνεται ως $t = 0$ το κεντρικό έτος της χρονολογικής σειράς)

Τότε

$$\widehat{a} = \sum y_i / n \qquad \widehat{\beta} = \sum t_i y_i / \sum t_i^2$$

Το μέσο τετραγωνικό σφάλμα εκτίμησης των τιμών της χρονολογικής σειράς γύρω από τη γραμμή της τάσεως όσο τείνει προς το μηδέν τόσο καλύτερα η εξίσωση της τάσης προσαρμόζεται στα εμπειρικά δεδομένα της χρονολογικής σειράς.

Το μέσο τετραγωνικό σφάλμα δίνεται από τον τύπο $\hat{\sigma}^{2} = \dfrac{\sum_{i=1}^{\nu} \hat{e}_i^2}{\nu - 2}$

Όπου ν το πλήθος της χρονολογικής σειράς

Παράδειγμα

Ο αριθμός των όρων της χρονολογικής σειράς είναι περιττός.

Θεωρώ $t = 0$ το 1970 (1/6/1970). Εξετάζω την χρονολογική σειρά $y_t =$ παραγωγή καπνού (σε χιλ. τόνους) στην Ελλάδα (1970-1982) και ζητώ την ευθεία τάσης.

Έτος	y	t	y·t	t^2	t	yt	t^2	$\widehat{y}$ για t=0: 30/6/76
1970	95	0	0	0	-6	-570	36	83,89
1971	89	1	89	1	-5	-445	25	88,66
1972	85	2	170	4	-4	-340	16	93,43
1973	91	3	273	9	-3	-273	9	98,2
1974	83	4	332	16	-2	-166	4	102,27
1975	119	5	595	25	-1	-119	1	107,74
1976	142	6	852	36	0	0	0	112,51
1977	120	7	840	49	1	120	1	117,28
1978	129	8	1032	64	2	258	4	122,05
1979	128	9	1152	81	3	384	9	126,82
1980	118	10	1180	100	4	472	16	131,59
1981	131	11	1441	121	5	655	25	136,36
1982	133	12	1596	144	6	798	36	141,13

Παίρνω τους τύπους: $\left.\begin{array}{l} \Sigma y = n\alpha + \beta\Sigma t \\ \Sigma yt = \alpha\Sigma t + \beta\cdot\Sigma t^2 \end{array}\right\}$

$$\left.\begin{array}{l} 1463 = 13\alpha + \beta 78 \\ 9.552 = \alpha 78 + \beta 650 \end{array}\right\} \Rightarrow \begin{array}{l} \hat{\alpha} = 87{,}04 \qquad \text{και} \qquad \hat{\beta} = 4{,}25 \\ \hat{\beta} > 0 \ \text{ Η σειρά έχει ανοδική τάση.} \end{array}$$

Και $\widehat{y}$= 87,04 + 4,25 t με t = 0,1,2,3,…. και t = 0 το 1970 (30/6/70).

Υπολογίζω τις τιμές της τάσης $\widehat{y}$ για t = 0,1,2,…..

Συγκρίνω τις τιμές της τάσης $\widehat{y}$ με τις τιμές της χρονολογικής σειράς και διαπιστώνω μια καλή προσαρμογή της ευθείας τάσης με τη δοσμένη χρονολογική σειρά.

Το μέσο τετραγωνικό σφάλμα εκτίμησης είναι:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{\nu} \hat{e}_i^2}{\nu - 2} = 177{,}34$$

* Αν πάρουμε ως αρχή t = 0 το κεντρικό έτος 1976 (30/6/76).

Τότε $\hat{\alpha} = \dfrac{\Sigma y}{n} = \dfrac{1463}{13} = 112{,}5$

$\widehat{\beta} = \dfrac{\Sigma yt}{\Sigma t^2} = \dfrac{774}{182} = 4{,}25$ και $\widehat{y} = 112{,}5 + 4{,}25t$

με t = -6, -5, -4, -3, -2, -1, 0, 1, 2, 3,....   από 30/6/7

Αν μετακινηθούμε στο έτος 1970 βάζοντας t = - 6 θα έχουμε $\widehat{y}$ =112,5+4,25

$\widehat{y}$ =112,5+4,25(-6) = 87$\simeq$ 87,04 (t=0).

Συνεπώς η ευθεία τάσης είναι:

$\widehat{y}$ = 87,04+4,25t με πολύ καλή προσαρμογή στη χρονολογική σειρά.

Παράδειγμα

Ο αριθμός των όρων της χρονολογικής σειράς είναι άρτιος.

Να γίνει πρόβλεψη για το έτος 2005.

Έτη	**(t=0) το 1986**	**t***	**y_t**	**$t\ y_t$**	**t^2**
1986	0	-6,5	75	-487,5	42,25
1987	1	-5,5	90	-495,0	30,25
1988	2	-4,5	90	-405,0	20,25
1989	3	-3,5	85	-297,5	12,25
1990	4	-2,5	105	-262,5	6,25
1991	5	-1,5	125	-187,5	2,25
1992	6	-0,5	130	-65	0,25
1993	7	+0,5	135	67,5	0,25
1994	8	1,5	140	210	2,25
1995	9	2,5	150	375	6,25
1996	10	3,5	170	595	12,25
1997	11	4,5	195	877,5	20,25
1998	12	5,5	185	1017,5	30,25
1999	13	6,5	180	1170,5	42,25
	91	0	1885	2112,5	227,50

$\hat{\alpha} = 132,5 \qquad \hat{\beta} = 9,2857$

$\hat{y} = 132,5+9,2857\ t^*$, με $t^* = -6,5, -5,5, -4,5,\ldots, -0,5, +0,5, +1,5\ldots$

Για να βρούμε την ευθεία της τάσης με αρχή το έτος 1986 θέτουμε όπου

$t^* = t - 6,5 \rightarrow$ άρα $t = t^* + 6,5$

και $\hat{y} = 132,5+9,2857\ (t - 6,5)$ με $t = 0, 1,2,\ldots$ από το 1986

οπότε $\hat{y} = 72,143+ 9,2857 \cdot t$ με $t = 0,1,2,\ldots$

Η τάση (τιμή) για το 2005 είναι $\hat{y}_{2005} = 72,143 + 9,2857 \cdot \underset{=}{19} = 248,5$

β. Προσδιορισμός καμπυλόγραμμης τάσης μιας χρονολογικής σειράς.

Οι καμπύλες μιας τέτοιας τάσης μπορεί να είναι παραβολή (καμπύλη β' βαθμού), καμπύλη τρίτου βαθμού και εκθετική καμπύλη.

Ι. ***Δευτέρου βαθμού καμπύλη.***

$\hat{y} = \hat{\alpha} + \hat{\beta}t + \hat{\gamma}t^2$, όπου τα $\hat{a}, \hat{\beta}, \hat{\gamma}$ προσδιορίζονται από το σύστημα των εξισώσεων:

$$\left.\begin{aligned} \Sigma y &= n\alpha + \beta\Sigma t + \gamma\Sigma t^2 \\ \Sigma yt &= \alpha \cdot \Sigma t + \beta\Sigma t^2 + \gamma\Sigma t^3 \\ \Sigma yt^2 &= \alpha\Sigma t^2 + \beta\Sigma t^3 + \gamma\Sigma t^4 \end{aligned}\right\} \text{(I)}$$

Αν ως αρχή t = 0 ληφθεί το κεντρικό έτος της σειράς τότε $\Sigma t = 0$, $\Sigma t^3 = 0$ οπότε:

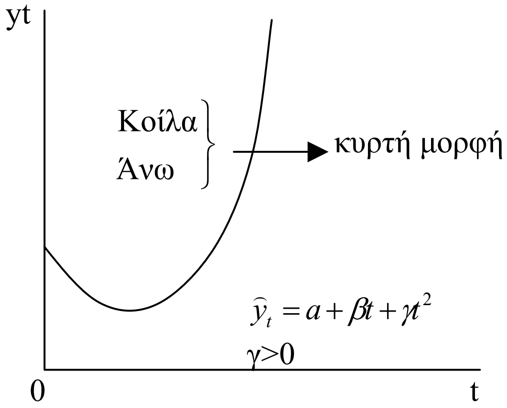
$$\left.\begin{aligned} \Sigma y &= n\alpha + \gamma\Sigma t^2 \\ \Sigma yt &= \quad \beta\Sigma t^2 \\ \Sigma yt^2 &= \alpha\Sigma t^2 + \quad + \gamma\Sigma t^4 \end{aligned}\right\} \text{(II)}$$

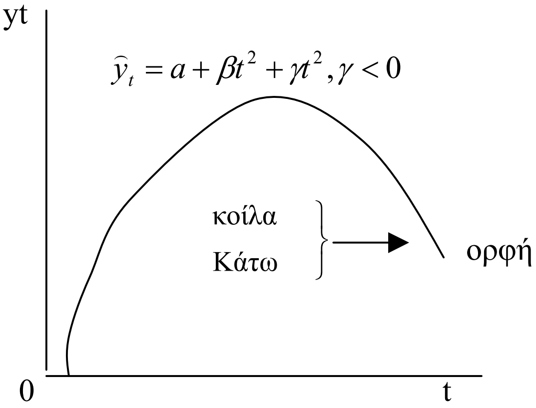
Το μέσο τετραγωνικό σφάλμα εκτίμησης είναι:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{\nu} \hat{e}_i^2}{\nu - 3}$$

Προσοχή.

Για να χρησιμοποιηθεί παραβολή 2ου βαθμού, πρέπει ή οι πρώτες διαφορές της χρονολογικής σειράς $\widehat{y}_{t+1} - y_t$ να σχηματίζουν αριθμητική πρόοδο ή το διάγραμμα της χρονολογικής σειράς να στρέφει τα κοίλα άνω ή κάτω.





Καμπυλόγραμμη τάση $2^{ου}$ βαθμού χρονολογικής σειράς.

Παράδειγμα:

Δίνεται η παραγωγή συνθετικού καουτσούκ στις ΗΠΑ στα έτη 1944-58 σε μέση μηνιαία παραγωγή – σε χιλ. τόνους (πηγή Β.Κ. Μπένος)

Έτη	**t=0:1951**	**Παραγωγή** y_t	t^2	**yt**	yt^2	$\widehat{y}_t$
1944	-7	63,6	49	-445,2	3116,4	63,1
1945	-6	68,4	36	-410,4	2462,4	58,2
1946	-5	61,7	25	-308,5	1542,5	54,4
1947	-4	42,4	16	-169,6	678,4	51,8
1948	-3	40,7	9	-122,1	366,3	50,3
1949	-2	32,8	4	-65,6	131,2	50
1950	-1	39,7	1	-39,7	39,7	51
1951	0	70,4	0	0	0	53,1
1952	1	66,5	1	66,5	66,5	56,3
1953	2	70,7	4	141,4	282,8	60,8
1954	3	51,9	9	155,7	467,1	66,4
1955	4	80,9	16	323,6	1294,4	73,2
1956	5	90	25	450	2250	81,2
1957	6	93,2	36	559,2	3355,2	90,3
1958	7	87,9	49	615,3	4307,1	100,6
		960,8	280	750,6	20360	

Από τις σχέσεις (II)
Έχουμε:

$$\left.\begin{array}{l} 980{,}8 = 15\alpha + 280\gamma \\ 750{,}6 = 280\beta \\ 20.360 = 280\alpha + 9.352\gamma \end{array}\right\} \longrightarrow \quad \widehat{\alpha} = 53{,}08, \widehat{\beta} = 2{,}6807, \widehat{\gamma} = 0{,}58785$$

Άρα η εξίσωση της τάσης θα είναι:

$$\widehat{y}_t = 53{,}08 + 2{,}6807t + 0{,}58785t^2 \quad \text{με} \quad t = 0 \rightarrow 30/6/51$$

$$t = -7,-6,-5,-4,-3,-2,-1,0,1,2,3,....$$

Αν ζητήσουμε πρόβλεψη για το έτος 1968 θα θέσω $\mathrm{t} = 17$ στην εξίσωση:

$$\widehat{y}_t = 53{,}08 + 2{,}6807 \cdot 17 + 0{,}58785 \cdot 17^2 = 268{,}54$$

Επίσης μπορεί να υπολογισθεί και το μέσο τετραγωνικό σφάλμα εκτίμησης που είναι:

$$\hat{\sigma}^{2} = \frac{\sum_{i=1}^{\nu} \hat{e}_i^2}{\nu - 3}$$

γ. Προσδιορισμός εκθετικής τάσης

Η εκθετική μορφή είναι καταρχήν η απλή $y_t = \alpha \cdot \beta^t$.

Χρησιμοποιείται όταν οι τιμές της χρονολογικής σειράς σχηματίζουν γεωμετρική πρόοδο.

Αν λογαριθμίσουμε (με δεκαδικό λογάριθμο) θα πάρουμε:

$$\lambda o\gamma y_t = \lambda o\gamma \alpha + t\lambda o\gamma \beta$$

Η εξίσωση αυτή παριστά ευθεία αν θέσουμε όπου $\lambda o\gamma y_t = y_t'$, $\lambda o\gamma \alpha = \alpha'$ και $\lambda o\gamma \beta = \beta'$ δηλαδή : $y_t' = \alpha' + t \cdot \beta'$

Οι εκτιμήσεις $\widehat{\alpha}, \widehat{\beta}$ των α, β προσδιορίζονται από το σύστημα των εξισώσεων:

$$\left.\begin{aligned} \Sigma\lambda o\gamma y &= n \cdot \lambda o\gamma \alpha + \lambda o\gamma \beta \cdot \Sigma t \\ \Sigma t \cdot \lambda o\gamma y &= \lambda o\gamma \alpha \cdot \Sigma t + \lambda o\gamma \beta \cdot \Sigma t^2 \end{aligned}\right\} \quad (\mathrm{I})$$

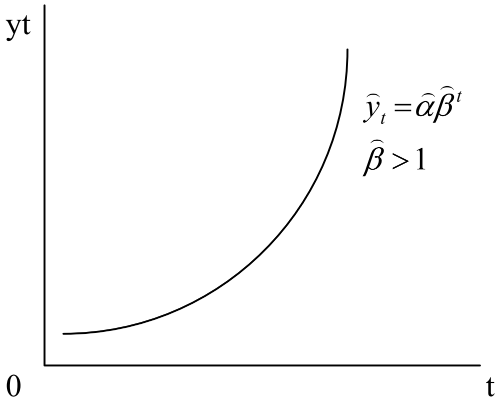
Αν επίσης ως αρχή χρόνου t = 0 ληφθεί το κεντρικό έτος της χρονολογικής σειράς άρα θα έχουμε Σt = 0 το σύστημα (I) παίρνει την απλή μορφή:

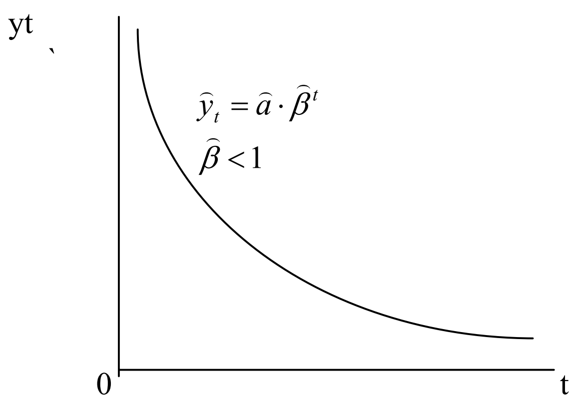
$$\left.\begin{aligned} \Sigma\lambda o\gamma y = n \cdot \lambda o\gamma \alpha &\Rightarrow \lambda o\gamma \alpha = \frac{\Sigma\lambda o\gamma y}{n} \\ \Sigma t \cdot \lambda o\gamma y = \lambda o\gamma \beta \cdot \Sigma t^2 &\Rightarrow \lambda o\gamma \beta = \frac{\Sigma t \cdot \lambda o\gamma y}{\Sigma t^2} \end{aligned}\right\} \quad \widehat{\alpha}, \widehat{\beta} \qquad (\mathrm{II})$$

Ομοίως μπορούμε να εκτιμήσουμε και το μέσο τετραγωνικό σφάλμα από τον τύπο:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{\nu} \hat{e}_i^2}{\nu - 2}$$

Επειδή οι όροι μιας χρονολογικής σειράς εμφανίζουν σταθερό ποσοστό μεταβολής, όταν μεταβάλλονται κατά γεωμετρική πρόοδο, η απλή εκθετική τάση χρησιμοποιείται για την εύρεση **μέσου ετήσιου ποσοστού μεταβολής (αύξησης ή μείωσης)** μιας χρονολογικής σειράς.





Παράδειγμα:

Υποθετικά δεδομένα (Πηγή: κ. Μπένος)

Έτη	**t**	**y_t**	**log y**	**t log y**	**t^2**
1955	-1	2	0,3010	-0,3010	1
1956	0	8	0,9031	0	0
1957	1	40	1,6021	1,6021	1
		50	2,8062	1,3011	2

$$\log \alpha = \frac{\Sigma \log y}{n} = \frac{2,8062}{3} = 09354 \cdot \Rightarrow \hat{a} = 8,62$$

$$\log \beta = \frac{\Sigma t \cdot \log y}{\Sigma t^2} = \frac{1,3011}{2} = 0,6505 \cdot \Rightarrow \hat{\beta} = 4,47$$

Άρα $\log y_t = 09354 + 06505 \cdot t$

$\Rightarrow \widehat{y}_t = 8{,}62 \cdot (4{,}47)^t \cdot$ με t=0 την 1/7/56
t = -1, 0, 1.

Υπολογίζουμε και το μέσο τετραγωνικό σφάλμα από τον τύπο:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{\nu} \hat{e}_i^2}{\nu - 2}$$

Επειδή το $\widehat{\beta} = 4{,}47 > 1$ η χρονολογική σειρά έχει ανοδική τάση.

Παρατήρηση:

Εκτός των αναφερομένων μοντέλων παλινδρόμησης μπορούν να αναφερθούν και τα ακόλουθα μοντέλα:

1. **Το εκθετικό μοντέλο**: (για καμπυλόγραμμη τάση).
   $\widehat{y}_t = \widehat{a} \cdot e^{\widehat{\beta} t}$, όπου e η βάση του νεπέριου λογάριθμου.

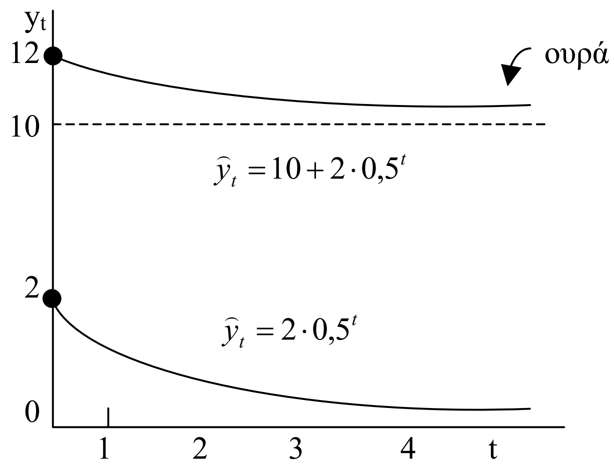
Σ’ αυτή την περίπτωση έχουμε: από την σχέση προς εκτίμηση $y_t = a \cdot e^{\beta t}$
$\ln y_t = \ln a + \beta t,$ οπότε αν θέσουμε:

$\ln y_t = y_t',\quad \ln a = a'$ έχουμε $y_t' = a' + \beta t$ και προσδιορίζονται τα $\widehat{\alpha}, \widehat{\beta}$ από τις αντίστοιχες εξισώσεις.

2. **Το απλό τροποποιημένο εκθετικό μοντέλο:**

$\widehat{y}_t = k + \widehat{a}\widehat{\beta}^t$ το οποίο διαφοροποιείται από το απλό εκθετικό μοντέλο κατά την σταθερά k (χρησιμοποιείται για χρονολογικές σειρές που καταλήγουν σε ουρά).

Αν π. χ $\widehat{y}_t = 2 \cdot 0{,}5^t$ με t = 0, 1, 2,… και το τροποποιημένο εκθετικό μοντέλο ήταν $\widehat{y}_t = 10 + 2 \cdot 0{,}5^t$ θα είχαμε τις μορφές

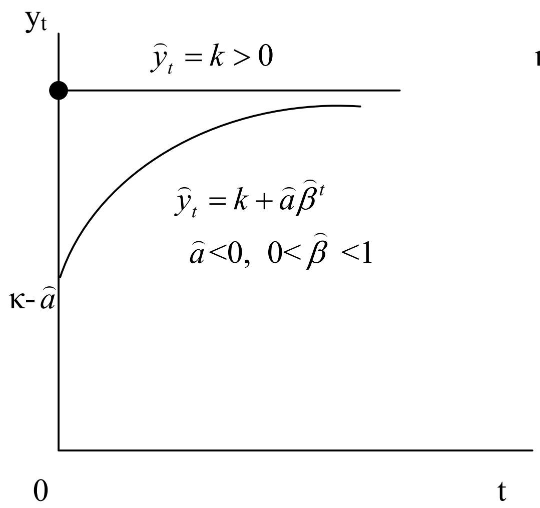


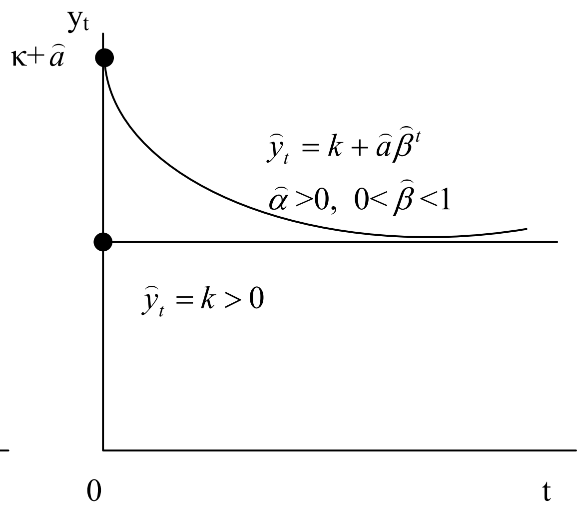
Στην περίπτωση αυτή έχουμε με k>0, $\widehat{a}$<0, 0<$\widehat{\beta}$ <1 το κ να είναι το κατώτερο επίπεδο όπου κατέρχεται η καμπύλη:

$$\widehat{y}_t = k + \widehat{a}\widehat{\beta}^t$$

Επίσης υπάρχει η περίπτωση με κ>0, $\widehat{a}$<0, 0<$\widehat{\beta}$ <1 το κ να είναι το ανώτερο επίπεδο όπου ανέρχεται η καμπύλη:

$$\widehat{y}_t = k + \widehat{a}\widehat{\beta}^t$$





Η $\widehat{y}_t = k > 0$ καλείται ασύμπτωτος της τάσης.

3. **Το πολυωνυμικό μοντέλο** (για ευθύγραμμη ή καμπυλόγραμμη τάση).

$$\widehat{y}_t = \widehat{\alpha} + \widehat{\beta}_1 t + \widehat{\beta}_2 t^2 + \widehat{\beta}_3 t^3 + ... \qquad ή$$

$$\widehat{y}_t = \sum_{i=1} \widehat{\beta}_i t^i = \widehat{\beta}_0 + \widehat{\beta}_1 t + \widehat{\beta}_2 t^2 + \widehat{\beta}_3 t^3 + ...$$

Έχουμε εξετάσει μέχρι $2^{ου}$ βαθμού.

4.**Το λογιστικό μοντέλο**: (για σειρές χρονολογικές που εμφανίζουν σχήμα S).

$$\frac{1}{\widehat{y}_t} = \widehat{\gamma} + \widehat{\alpha}\widehat{\beta}^t$$

5. **Το μοντέλο Compertz**: (χρονολογικές σειρές που εμφανίζουν σχήμα S).

$$\log y_t = \log \gamma + (\log a)\beta^t$$

6. **Η ανάλυση σε σειρές Fourier**: (για σειρές που παρουσιάζουν κυκλικές ή εποχιακές διακυμάνσεις) π.χ. εβδομαδιαία, κατανάλωση κρέατος.

$$\hat{y}_t = \sum_{k=1}^{n} k(a_k \sigma\upsilon\nu\omega t + \beta_\kappa \eta\mu\omega t)$$

$\omega = \frac{2\pi}{\nu}$, ν η βασική συχνότητα του εποχιακού κύκλου, n ο αριθμός των παρατηρήσεων σ΄ ένα εποχιακό κύκλο, α_κ, β_κ παράμετροι που υπολογίζονται με τη μέθοδο των ελαχίστων τετραγώνων, κ ο αριθμός των αρμονικών κυμάνσεων.

8.4. Στατιστικός προσδιορισμός της εποχικότητας.

Αναφέρθηκε ότι η εποχική κίνηση ή εποχικότητα είναι σύντομη και εξαντλείται στο χρόνο, έτσι εμφανίζονται περίοδοι μηνιαίοι, τριμηνιαίοι, εβδομαδιαίοι κλπ.

Μέσα στο έτος η χρονολογική σειρά εξαντλεί όλες τις ανοδικές και τις ισοδύναμες καθοδικές τάσεις.

Ο στατιστικός προσδιορισμός της εποχικής συνιστώσας (S) είναι ο υπολογισμός αριθμοδεικτών των **δεικτών εποχικότητας** που συμβολίζονται με S_j j=1,2,….,12 αν έχουμε μηνιαία στοιχεία ή j =1,2,3,4 αν έχουμε τριμηνιαία στοιχεία.

Οι μέθοδοι υπολογισμού των δεικτών εποχικότητας είναι:

Ι) Η μέθοδος των ποσοστών ως προς το μηνιαίο μέσο.

Η μέθοδος αυτή εφαρμόζεται σε χρονολογικές σειρές που δεν έχουν τάση και κυκλικές κυμάνσεις.
Η μέθοδος παρουσιάζεται στο παρακάτω παράδειγμα:

Η Μηνιαία κατανάλωση ηλεκτρικής ενέργειας ενός εργοστασίου την περίοδο (1978-1982) σε χιλ. ΩΧΒ. είναι:

Έτος	Ιαν.	Φεβρ.	Μαρτ.	Απρ.	Μάιο	Ιουν.	Ιουλ.	Αυγ.	Σεπτ.	Οκτ.	Νοεμ	Δεκ.
1978	503	506	428	391	401	351	316	302	310	331	402	468
1979	543	524	491	452	441	392	180	165	511	492	403	502
1980	587	618	576	561	482	410	370	350	348	376	431	503
1981	634	576	653	524	478	413	381	368	382	381	452	523
1982	686	676	638	607	537	446	461	391	401	410	472	581

Έτος	Ετήσια Αθρ.	Μηνιαίοι Μέσοι
1978	4709	392,4
1979	5096	424,7
1980	5612	467,7
1981	5765	480,4
1982	6306	525,5

Διαιρώ τις καταναλώσεις κατά μήνα με τις αντίστοιχες μέσες μηνιαίες καταναλώσεις το χρόνο και πολ/ζω επί 100. Προκύπτει έτσι ο κάτωθι πίνακας

Πίνακας:

Έτος	**Ιαν.**	**Φεβρ**	**Μαρτ**	**Απρ.**	**Μαίο**	**Ιουν.**	**Ιουλ.**	**Αυγ.**	**Σεπτ**	**Οκτ.**	**Νοεμ**	**Δεκ.**
1978	128,2	129	109,1	99,6	102,2	89,4	80,5	77	79	84,4	102,4	119,3
1979	127,9	123,4	115,6	106,4	103,8	92,3	42,4	38,9	120,3	115,8	94,9	118,2
1980	125,5	132,1	123,2	120	103,7	87,7	79,1	74,8	74,4	80,4	92,2	107,5
1981	132	120	136	109,1	99,5	86,0	79,3	76,6	79,5	79,3	94,1	120,9
1982	130,5	128,6	121,4	115,5	102,2	84,9	87,7	74,4	76,3	78	89,8	110,6
Αθρ.	644,1	633,1	605,3	550,6	511,4	440,3	369	341,7	429,5	437,9	473,4	576,5
Δεί/της Εποχ. Sj	128,8	126,6	121,1	110,1	102,3	88,1	73,8	68,3	85,9	87,6	94,7	115,3

j= 1,2,….,12 π.χ. $S_7 = 73{,}8$ Ιούλιος.

Το άθροισμα $\sum_{j=1}^{12} S_j$ πρέπει να είναι 1200%. Αν $\sum Sj \neq 1200$ τότε διορθώνουμε τους δείκτες εποχικότητας Sj πολλαπλασιάζοντας καθένα με το **συντελεστή διόρθωσης.**

$$\delta = \frac{1200}{\Sigma Sj}$$

Έχουμε στο παράδειγμα:
$\Sigma S_j = 1202{,}6\% \simeq 1200\%$ δεν χρειάζονται οι δείκτες εποχικότητας διόρθωση.

Ερμηνεία: Ο δείκτης εποχικότητας π.χ. του Σεπτεμβρίου 85,9% σημαίνει ότι η κατανάλωση της ενέργειας τον μήνα αυτό παρουσιάζει μείωση κατά 14,1% σε σχέση με τη μέση μηνιαία κατανάλωση.

Παρατήρηση:

Αν διαθέτουμε τριμηνιαία στοιχεία μιας χρονολογικής σειράς για κάποια έτη και θέλουμε να προσδιορίσουμε τους δείκτες εποχικότητας εργαζόμαστε όπως στο προηγούμενο παράδειγμα με τη διαφορά τα ετήσια αθροίσματα τα διαιρούμε με το 4 (4 τρίμηνα) και υπολογίζουμε τους τριμηνιαίους μέσους κάθε έτους. Κατόπιν διαιρούμε τα τριμηνιαία στοιχεία με τους αντίστοιχους τριμηνιαίους μέσους κάθε έτους και πολλαπλασιάζουμε με το 100. Ο μέσος όρος των ποσοστών κάθε τριμήνου για μια σειρά ετών αποτελεί το δείκτη εποχικότητας του τριμήνου.

Πρέπει να ισχύει:

$$\sum_{j=1}^{4} S_j = 400(4x100)$$

Αν $\sum_{j=1}^{4} S_j \neq 400$ πολλαπλασιάζουμε κάθε S_j επί τον συντελεστής διόρθωσης

$$\delta = \frac{400}{\Sigma Sj}$$

II) Η **μέθοδος των ποσοστών ως προς τη μηνιαία τάση.**

Θεωρούμε ότι y_t = T.S.C.I. (το πολλαπλασιαστικό υπόδειγμα).
Τα δεδομένα κάθε μήνα εκφράζονται ως ποσοστό % της αντίστοιχης μηνιαίας τιμής της τάσης.

Ο αριθμητικός μέσος των ποσοστών αυτών για τους αντίστοιχους μήνες δίνει τους δείκτες εποχικότητας (S_j) που πρέπει να ικανοποιούν τη σχέση $\sum Sj \neq 1200\%$ αλλιώς διορθώνονται με το συντελεστή διόρθωσης $\delta = \frac{1200}{\Sigma Sj}$.

Το μειονέκτημα της μεθόδου των ποσοστών ως προς τη μηνιαία τάση βρίσκεται στο σημείο όπου ο λόγος που ζητείται:

$$\frac{y_t}{\underset{(\text{τάση})}{T}} = S.C.I$$

μας δείχνει ότι οι δείκτες εποχικότητας περιέχουν εποχικές (S) και κυκλικές (C) κυμάνσεις όπως και τυχαίες επιδράσεις (I), όταν η χρονολογική σειρά είναι και αρκετά μεγάλη.

Παράδειγμα

Στο παράδειγμα της προηγούμενης μεθόδου καθορίζουμε την μηνιαία τάση:

Έχουμε:

Έτος	**y** **Μηνιαίοι μέσοι**	**t**	**yt**	**t^2**
1978	392,4	-2	-784,8	4
1979	424,7	-1	-424,7	1
1980	467,7	0	0	0
1981	480,4	1	480,4	1
1982	525,5	2	1051	4
	Σy=2290,7	---	Σyt=321,9	Σt^2=10

$$\widehat{\alpha} = \frac{\Sigma y}{n} = \frac{2290{,}7}{5} = 458{,}14$$

$$\widehat{\beta} = \frac{\Sigma yt}{\Sigma t^2} = \frac{321{,}9}{10} = 32{,}19$$

Η εξίσωση τάσης είναι $\widehat{y}_t = 458{,}14 + 32{,}19t$ με αρχή $\mathrm{t} = 0 \rightarrow 30/6/80$
Μετατρέπουμε την ετήσια περίοδο της τάσης σε μηνιαία.

Τα y είναι μέσα ετήσια και γι΄ αυτό διαιρούμε ***μόνο*** το $\widehat{\beta}$ με το 12. Αν τα y ήταν ετήσια αθροίσματα θα διαιρούσαμε το $\widehat{\alpha}$ με το 12 και το $\widehat{\beta}$ με το 12·12=144. Έτσι παίρνουμε:

$\widehat{y}_t = 458{,}14 + 2{,}68t$ με $\mathrm{t} = 0 \rightarrow 30/6/80$ και ο χρόνος t μετριέται σε μήνες.

Μετακινούμε την αρχή στις 15/1/78 θέτοντας όπου:

$$\mathrm{t} = \frac{12 \cdot \mathrm{n} - 1}{2} = -\frac{12\mathrm{x}5 - 1}{2} = -29{,}5\text{μήνες}$$

(- 29,5 μήνες από 30/6/80 έως 15/1/78)

τότε η μηνιαία τάση τον Ιανουάριο του 1978 θα ήταν:

$\widehat{y}_t = 458{,}14 + 2{,}68 \cdot (-29{,}5) = 379{,}1$ μηνιαία τάση τον Ιανουάριο του 1978.

Άρα η εξίσωση τάσης ανά μήνα είναι:

$\widehat{y}_t = 379{,}1 + 2{,}68t$ από 15/1/78 ο χρόνος σε μήνες t σε μήνες

Αν τώρα τα αρχικά δεδομένα κάθε μήνα διαιρεθούν με τις αντίστοιχες μηνιαίες τιμές τάσης, εκφρασθούν επί τοις % και βρεθεί ο μέσος όρος κάθε μήνα ή η διάμεσος της πενταετίας (για κάθε μήνα) έχουμε τους δείκτες εποχικότητας κατά μήνα.

Π.χ. Για τον Ιανουάριο του 1978: $\frac{503}{379{,}1} = 1{,}326\mathrm{x}100 = 132{,}6\%$ (η μηνιαία τάσης).

Αν υπολογίσω για τα άλλα έτη τον Ιανουάριο προχωρώ όπως αναφέρεται πιο πάνω.

III) **Η μέθοδος των ποσοστών ως προς τους μηνιαίους κινητούς μέσους.**

Θεωρούμε το πολλαπλασιαστικό υπόδειγμα: $\hat{y}_t = T \cdot S \cdot C \cdot I$

Η μέθοδος αποβλέπει στην απαλοιφή των εποχικών κυμάνσεων και των τυχαίων κινήσεων.

Οι κινητοί μέσοι εκτιμούν ικανοποιητικά την τάση και τις κυκλικές κυμάνσεις δηλαδή ισοδυναμούν με TC.

Αν διαιρέσουμε τα αρχικά δεδομένα με τους αντίστοιχους κινητούς μέσους (TC) παίρνουμε δεδομένα που έχουν μόνο (S.I) εποχιακές κυμάνσεις και τυχαίες κινήσεις $\widehat{y}_t / TC = S.I.$

Τέλος με τον υπολογισμό του μέσου όρου των ποσοστών απαλείφονται και οι τυχαίες κινήσεις (I) και μένουν οι εποχικές.

Παράδειγμα:

	1 Έτος	**2** y	**3** Κιν.Αθρ. 12 - μηνών	**4** Κιν.Αθρ. 2 - διαδοχ.μην.	**5 = 4:24** Κιν.μέσοι 12 μηνών	**6 = (2:5)·100** ποσοστό %
1979	Ιαν.	500	-	-	-	-
	Φεβ.	600	-	-	-	-
	Μάρ.	700	-	-	-	-
	Απρ.	500	-	-	-	-
	Μαϊος	600	-	-	-	-
	Ιούν.	700	-	-	-	-
	Ιούλ.	300	6300	12300	512,5	58,54
	Αυγ.	800	6000	11900	495,8	16,14
	Σεπτ.	600	5900			
	Οκτ.	500				
	Νοεμ.	200				
	Δεκ.	300				
1980	Ιαν.	200				
	Φεβ.	500				

Αν συγκεντρώσουμε τα ποσοστά της 6ης στήλης σε ένα πίνακα:

Έτη	**Ιανουάριο……**	**Ιούλιος**	**Αύγουστος……..**
1978			
1979		58,54	16,14
1980			
1981			
1982			
Άθροισμα		α	
Μέσος		α/5	
Δείκτες Εποχικότητας		α/5 ή διορθωμένος*	
Ή διάμεσοι		β	
Δείκτες εποχικότητας.		β ή διορθωμένος*	

* Η διόρθωση γίνεται με τον δείκτη $\delta = \dfrac{1200}{\Sigma Sj}$ και εφόσον $\sum_{j=1}^{12} S_j \neq 1200$

Απαλοιφή της εποχικότητας.

Οι δείκτες εποχικότητας έχουν μεγάλη σημασία διότι μπορούμε να απαλλάξουμε τους όρους μιας χρονολογικής σειράς από την επίδραση της εποχικής κύμανσης εφόσον διαιρεθούν τα δεδομένα μιας χρονολογικής σειράς με τους αντίστοιχους δείκτες εποχικότητας (μηνιαία-τριμηνιαία δεδομένα) κλπ.

Η διαδικασία αυτή καλείται απαλοιφή της εποχικότητας και έτσι προκύπτουν συγκρίσιμα μεταξύ τους στοιχεία.

Αυτό επιτρέπει και τους ιθύνοντες των επιχειρήσεων να μπορούν να μετρούν τις εποχικές μεταβολές και προγραμματίζουν για το μέλλον.

8.5. Στατιστικός προσδιορισμός των κυκλικών κυμάνσεων.

Η μεθοδολογία που ακολουθείται είναι η «**μέθοδος των καταλοίπων**».

Ι)Πρώτα κάνουμε απαλοιφή της εποχικότητας.
Στο πολλαπλασιασμό υπόδειγμα έχουμε:

$$\frac{TSCI}{S} = TCI$$

ΙΙ)Διαιρούμε τα στοιχεία με την τάση (Τ).

$$\frac{TCI}{T} = CI$$

ΙΙΙ)Αν εφαρμόσουμε ένα κατάλληλο κινητό μέσο (3 ή 5 μηνών) μπορούμε να απαλείψουμε και τις άρρυθμες κινήσεις (Ι) και έτσι να προκύψουν οι κυκλικές κυμάνσεις (C).

Παράδειγμα:
Έστω η μηνιαία κατανάλωση ηλεκτρικής ενέργειας στην Ελλάδα για οικιακή χρήση (1978-82). (Στοιχεία ΔΕΗ) σε εκατ. ΩΧΒ.

Πίνακας(α):

Έτος	Ιαν.	Φεβρ	Μαρτ.	Απρ.	Μαίο	Ιουν.	Ιουλ.	Αυγ.	Σεπτ.	Οκτ.	Νοεμ	Δεκ.
1978	501	516	438	407	403	358	323	301	315	339	401	477
1979	542	534	489	482	448	389	167	160	512	499	406	501
1980	584	621	585	570	491	413	381	344	350	374	441	500
1981	645	584	662	555	481	416	391	373	386	392	464	554
1982	698	686	647	608	543	454	424	388	406	413	479	576

Τα δεδομένα αυτά τα απαλείφουμε από την εποχικότητα και την τάση. Η απαλοιφή της εποχικότητας γίνεται μέσω των συντελεστών εποχικότητας, εφόσον διαιρεθούν οι τιμές του πίνακα (α) με τους αντίστοιχους συντελεστές εποχικότητας.

Οι συντελεστές εποχικότητας υπολογίζονται με την μέθοδο των μηνιαίων κινητών μέσων, όπως αναπτύχθηκε.

Συγκεκριμένα από τον πίνακα (α) δημιουργούμε τον πίνακα (β) ως εξής:

Πίνακας (β):

	1 Έτος	2 y: Καταν. ενέργειας	3 Κιν. Αθρ. 12-μηνών	4 Κιν. Αθρ. 2 - διαδοχ.μην.	5=4:24 Κιν. μέσοι 12 μηνών	6=(2:5)·100 ποσοστό %
1978	Ιαν.	501	-	-	-	-
	Φεβ.	576	-	-	-	-
	Μαρ.	438	-	-	-	-
	Απρ.	407	-	-	-	-
	Μαιος	403	-	-	-	-
	Ιουν.	358	-	-	-	-
	Ιουλ.*	323	4779	9,599	399,9	80,8
	Αυγ.	301	4820			
				κ.ο.κ.		

Συγκεντρώνουμε κατόπιν όλα τα ποσοστά της στήλης 6 και δημιουργούμε ένα νέο πίνακα (γ):

Πίνακας(γ):

Έτος	**Ιαν.**	**Φεβ.**	**Μαρ.**	**Απρ.**	**Μαίος**	**Ιουν.**	**Ιουλ.**	**Αυγ.**	**Σεπ.**	**Οκτ.**	**Νοεμ**	**Δεκ.**
1978	-	-	-	-	-	-	80,8	74,8	77,7	82,6	96,5	113,9
1979	131,1						38,9					109,4
1980	124,8						*					104,7
1981	134,9						80,4					107,8
1982	133,1					-	79,2	-	-	-	-	-
Αθροισμ.	523,9						-					
Μέσοι	131,0											
Δείκτης**												
Εποχικ.	127,9											
Διάμεσος	129,8											
Δείκτης**												
Εποχικ.	128,6											

*: Δεν λαμβάνεται υπόψιν για την εξαγωγή του μέσου όρου, επειδή είναι μικρή τιμή.

,*: Προκύπτουν με πολλαπλασιασμό αντίστοιχα των μέσων και των διαμέσων επί του συντελεστή διόρθωσης:

δ** = 1200/1229,3 = 0,9762

δ*** = 1200/1210,8 = 0,991

Άρα απαλείφω την εποχικότητα (με δείκτη εποχικότητας π.χ. τον 127,9 (από τον μέσο όρο)) διαιρώντας την αντίστοιχη τιμή του Ιανουαρίου με τον δείκτη εποχικότητας του Ιανουαρίου και προκύπτει ο πίνακας (δ):

Πίνακας (δ):

Έτος Ιανουάριος   Φεβρουάριος   Μάρτιος ………

1978 $\dfrac{501}{127{,}9}\text{x}100 = 392$

1979   ……….

Κατόπιν τα στοιχεία του πίνακα (δ) τα απαλείφουμε από την μακροχρόνια τάση διαιρώντας τα με τις αντίστοιχες μηνιαίες τιμές της τάσης.

Οι μηνιαίες τιμές της τάσης υπολογίζονται όπως προαναφέρθηκε στη μέθοδο των ποσοστών ως προς την μηνιαία τάση.

Συγκεκριμένα:

Δημιουργούμε τον πίνακα (ε) :

Πίνακας (ε):

Έτος	**Μηνιαίοι μέσοι: y**	**t**	**yt**	**t^2**
1978	398,3*	-2	-796,6	4
1979	427,4	-1		
1980	471,2	0		
1981	492,0	1		
1982	526,0	2		
	Σy= 2.314,9		Σy= 2.314,9	Σy= 2.314,9

$$\left.\begin{aligned}\sum y &= n\cdot \alpha\\ \sum yt &= \beta\sum t^2\end{aligned}\right\}\rightarrow \begin{aligned}\widehat{\alpha} &= 462{,}98\\ \widehat{\beta} &= 32\end{aligned}$$

Η εξίσωση τάσης:

$\widehat{y}_t = 462{,}98 + 32t$ t = 0 (έτη) αρχή: 30/6/1980

μετατρέπουμε την ετήσια περίοδο τάσης σε μηνιαία και έχουμε:

$\widehat{y}_t = 462{,}98 + 2{,}67t$ με αρχή την 30/6/1980

Τέλος με την μετακίνηση στις 15/1/78 θέτοντας όπου t = -29,5 έχουμε την τελική εξίσωση μηνιαίας τάσης:

$\widehat{y}_t = 384{,}22 + 2{,}67t$ t=0 όταν έχουμε 15/1/78

384,22=462,98+2,67 (-29,5).

Αν θέσουμε στην σχέση: $\widehat{y}_t = 384{,}22 + 2{,}67t$, t=0,1,2,….,59 βρίσκουμε τις αντίστοιχες μηνιαίες τάσεις και δημιουργούμε τον πίνακα (στ) (Μηνιαίες τάσεις) (Τ).

Πίνακα: (στ)

Έτος	**Ιανουάριος**	**Φεβρουάριος**	**Μάρτιος…..**
1978	384,2		
1979			
1980			
1981			
1982			

Από τον πίνακα (δ) διαιρώντας τα στοιχεία του με τα αντίστοιχα στοιχεία του πίνακα (στ) , που εκφράζουν τις μηνιαίες τάσεις, παίρνω τα στοιχεία απαλλαγμένα και από την μακροχρόνια τάση και δημιουργείται ο πίνακας (ζ):

Πίνακας (ζ)

Έτος Ιανουάριος Φεβρουάριος Μάρτιος

$$1978 \quad \frac{392}{384,2} x100 = 102\%$$

Ο πίνακας (ζ) περιέχει δεδομένα που εμφανίζουν κυκλική κύμανση και άρρυθμες κινήσεις**.**

Αν στα δεδομένα αυτά εφαρμόσουμε ένα κινητό μέσο πέντε μηνών μπορούμε να απαλείψουμε τις άρρυθμες κινήσεις (Τ) και να έχουμε δεδομένα με μόνο κυκλικές κυμάνσεις που καλούνται **δείκτες κυκλικότητας** και χρησιμοποιούνται ως δείκτες εποχικότητας.

Κεφάλαιο 9^{ο}

Αριθμοδείκτες

9.1. Αριθμοδείκτες

Ο Αριθμοδείκτης είναι ένας αριθμός που εκφράζει το επίπεδο των τιμών, των ημερομισθίων, των εισαγωγών-εξαγωγών, των ετησίων εισοδημάτων, των μέσων στρεμματικών αποδόσεων και λοιπών άλλων μεταβλητών σε σχέση με το επίπεδο των μεταβλητών αυτών σε μια άλλη περίοδο που καλείται **περίοδος βάσης.** Είναι στατιστικό μέτρο που φανερώνει την διαχρονική εξέλιξη στις μεταβολές των τιμών μιας μεταβλητής.
Με το μέτρο αυτό η σχετική μεταβολή μιας μεταβλητής εκφράζεται με ένα αριθμό ως εκατοστιαία αναλογία ενός άλλου.

Ένα παράδειγμα αριθμοδείκτη: Θεωρούμε ότι το μέσο ετήσιο εισόδημα του αγρότη το 1987 ήταν 2.000.000 και το αντίστοιχο του 1997 ήταν 2.500.000 έτσι μπορούμε να πούμε ότι έχουμε το 1997 αύξηση του μέσου ετησίου αγροτικού εισοδήματος κατά 25% ή ότι το μέσο ετήσιο αγροτικό εισόδημα το 11997 είναι το 125% εκείνου του 1987. Θεωρούμε ως έτος βάσης το 1987 και παριστάνουμε το μέσο ετήσιο αγροτικό εισόδημα στις δυο χρονικές περιόδους 1987 και 1997 με τους αριθμοδείκτες 100 και 125 που δείχνουν τη μεταβολή στο αγροτικό εισόδημα.

Παρατίθεται ένας πίνακας όπου προσδιορίζονται οι αριθμοδείκτες στο προαναφερόμενο παράδειγμα.

Έτος	Μέσο ετήσιο αγροτικό εισόδημα	αριθμοδείκτης
1977	500.000 (δρχ.)	25
1987	2.000.000 »	100
1997	2.500.000 »	125

Οι αριθμοδείκτες διακρίνονται σε **απλούς δείκτες** οι οποίοι χρησιμοποιούνται για μια μόνο μεταβλητή και οι **σύνθετοι** για μια ομάδα ομοειδών μεταβλητών οι οποίοι διακρίνονται σε **σταθμισμένους** και **αστάθμιστους** αριθμοδείκτες.

9.2. Απλοί Αριθμοδείκτες

Οι απλοί αριθμοδείκτες διακρίνονται σε **αριθμοδείκτες τιμών, αριθμοδείκτες ποσοτήτων, αριθμοδείκτες αξιών.**

1. Οι **αριθμοδείκτες τιμών** ή **τιμάριθμοι** αναφέρονται σε τιμές μετρούν τις ποσοστιαίες μεταβολές (διαχρονικές ή διατοπικές) των τιμών των αγαθών που παράγονται, καταναλώνονται, εισάγονται, εξάγονται κλπ. μεταξύ δύο

χρονικών περιόδων ή μεταξύ δύο γεωγραφικών περιοχών (γεωγραφικοί αριθμοδείκτες) και δίνονται από τον τύπο:

$$I_p = \frac{P_n}{P_o} \cdot 100$$

όπου P_n είναι η τιμή ενός αγαθού το έτος –n που μας ενδιαφέρει και P_O είναι η τιμή του αγαθού το έτος βάσης.

Τέτοιοι δείκτες είναι **ο δείκτης τιμών καταναλωτή, ο δείκτης τιμών χονδρικής πώλησης, ο δείκτης τιμών λιανικής πώλησης, ο δείκτης τιμών των εισαγομένων και των εξαγομένων εμπορευμάτων, ο δείκτης τιμών του χρηματιστηρίου αξιών, ο δείκτης των αποδοχών των μισθωτών στη Βιομηχανία-Βιοτεχνία, ο δείκτης τιμών των οικοδομικών υλικών κλπ.**

2. **Οι αριθμοδείκτες ποσοτήτων**

Μετρούν τις ποσοστιαίες μεταβολές (διαχρονικές ή διατοπικές) των ποσοτήτων των αγαθών που παράγονται, καταναλώνονται, εισάγονται, εξάγονται κλπ. μεταξύ δύο χρονικών περιόδων ή μεταξύ δύο γεωγραφικών περιοχών (γεωγραφικοί αριθμοδείκτες) και δίνονται από τον τύπο:

$$I_q = \frac{Q_n}{Q_o} \cdot 100$$

όπου Q_n αναφέρεται στη ποσότητα ενός αγαθού το έτος – n που μας ενδιαφέρει και Q_o η ποσότητα του αγαθού το έτος βάσης.

Τέτοιοι δείκτες είναι **ο δείκτης όγκου παραγωγής, ο δείκτης όγκου εισαγωγών, ο δείκτης όγκου εξαγωγών κλπ.**

3. **Οι αριθμοδείκτες αξιών**

Μετρούν τις ποσοστιαίες μεταβολές της αξίας διαφόρων αγαθών μεταξύ δύο χρονικών περιόδων μεταξύ δύο γεωγραφικών περιοχών (γεωγραφικοί αριθμοδείκτες) και δίνονται από τον τύπο:

$$I_v = \frac{P_n Q_n}{P_o Q_o} \cdot 100$$

όπου P_n, Q_n, P_o, Q_o τα μεγέθη όπως ορίσθηκαν παραπάνω.

Τέτοιος δείκτης είναι **ο δείκτης αξίας εισαγωγών, ο δείκτης αξίας εξαγωγών, ο δείκτης λιανικών πωλήσεων (τζίρου).**

Παράδειγμα

Θεωρούμε τον **λιανικό δείκτη τιμών** ο οποίος μετρά το κόστος ζωής. Αυτό είναι το σύνολο των εξόδων για μέση οικογένεια και αναφέρεται σε σχέση με τα έξοδα ενός χρόνου που λαμβάνεται ως έτος βάσης. Για παράδειγμα ο δείκτης το 1991 είναι 134,1 με έτος βάσης το 1987, που σημαίνει ότι ο δείκτης αυξήθηκε από το 1987 κατά 34% περίπου.

Στο παρακάτω πίνακα αν πάρουμε τον χρόνο 1 ως χρόνο βάσης τότε ο απλός αριθμοδείκτης τιμών ή τιμάριθμος είναι:

$I_p = \frac{P_n}{P_o} \cdot 100$, όπου P_n είναι ο λιανικός δείκτης τιμών του έτους 1991 που μας ενδιαφέρει και P_o είναι ο λιανικός δείκτης τιμών του έτους βάσης 1987.

Πίνακας : Τιμή και κατανάλωση γάλατος, αυγών και ψωμιού

		Τιμή		**Καταναλωθείσα ποσότητα (π.χ. για ένα μήνα)**	
εμπόρευμα	**μονάς**	**Χρόνος 1** P_o	**Χρόνος 4** P_n	**Χρόνος 1** Q_o	**Χρόνος 4** Q_n
Αυγά	Δωδεκάδα	40	70	1,4	1,3
Ψωμί	0,5 Kg	35	50	3,6	3,5
Γάλα	μισό λίτρο	12	17	20	23

Έτσι έχουμε

$I_p = \frac{70}{40} \cdot 100 = 175$ για τα αυγά ο δείκτης τιμών είναι 175

$I_p = \frac{50}{35} \cdot 100 = 142,9$ για το ψωμί δείκτης τιμών είναι 142,9

$I_p = \frac{17}{12} \cdot 100 = 141,7$ για το γάλα ο δείκτης τιμών είναι 141,7

αυτές οι τιμές καλούνται **σχετικές τιμές** και εκφράζουν την εξέλιξη της τιμής του λιανικού δείκτη.

Ένας αντίστοιχος απλός αριθμοδείκτης ποσοτήτων είναι ο $I_q = \frac{Q_n}{Q_o} \cdot 100$,

όπου Q_n, Q_o αναφέρεται στη ποσότητα του αγαθού που καταναλίσκεται αντίστοιχα στα έτη 1991 και 1987.

Σύμφωνα με τον πίνακα θα έχουμε

για τα αυγά $I_q = \frac{1,3}{1,4} \cdot 100 = 92,9$

για το ψωμί $I_q = \frac{3,5}{3,6} \cdot 100 = 97,2$

για το γάλα $I_q = \frac{23}{20} \cdot 100 = 115,0$

αυτές οι ποσότητες καλούνται **σχετικές ποσότητες.**

Ο δείκτης για την αξία των εμπορευμάτων ορίζεται ως η πολλαπλασιαστική τιμή της ποσότητας σύμφωνα με τον τύπο:

$$I_v = \frac{P_n Q_n}{P_O Q_o} \cdot 100$$

Αν χρησιμοποιήσουμε τα δεδομένα του πίνακα θα βρούμε για την αξία των εμπορευμάτων τα εξής:

$$I_v = \frac{70x1,3}{40x1,4} \cdot 100 = 162,5 \quad \text{για τα αυγά}$$

$$I_v = \frac{50x3,5}{35x3,6} \cdot 100 = 138,9 \quad \text{για το ψωμί}$$

$$I_v = \frac{17x23}{12x20} \cdot 100 = 162,9 \quad \text{για το γάλα}$$

οι απλοί αριθμοδείκτες τιμών φανερώνουν ότι οι τιμές των αυγών, του ψωμιού και του γάλατος έχουν αυξηθεί. Επίσης οι απλοί αριθμοδείκτες ποσοτήτων φανερώνουν ότι η κατανάλωσης των αυγών και του ψωμιού ελαττώθηκε ενώ η κατανάλωση του γάλατος αυξήθηκε.

Τέλος όσον αφορά τον δείκτη για την αξία των εμπορευμάτων έχουμε ότι οι αξίες και των τριών εμπορευμάτων έχει αυξηθεί σε σχέση με την αξία που είχαν στο χρόνο βάσης.

Εάν θέλουμε να αναφερθούμε σε ένα χρόνο πριν το χρόνο βάσης τότε θεωρούμε πάλι την τιμή του λιανικού δείκτη στο χρόνο βάσης το 100 και ανάγουμε την τιμή του υπό εξέταση χρόνου στον χρόνο βάσης, όπως και πριν. Παράδειγμα αν έχουμε δείκτη λιανικής τιμής 36 στο υπό εξέταση χρόνο τότε ο αριθμοδείκτης στο εν λόγω χρόνο θα είναι $\frac{36}{40} \cdot 100 = 90$.

9.3. Σύνθετοι αριθμοδείκτες

9.3.1. Μη σταθμισμένος σύνθετος αριθμοδείκτης

Αναφέρεται στις τιμές ή στις ποσότητες και είναι της μορφής

$$I_P = \frac{\Sigma P_n}{\Sigma P_o} \cdot 100$$

$$\text{ή} \qquad I_q = \frac{\Sigma Q_n}{\Sigma Q_o} \cdot 100$$

με $\mathbf{\Sigma P_n}$, $\mathbf{\Sigma Q_n}$ το άθροισμα των τιμών και των ποσοτήτων στο χρόνο –n που εξετάζουμε και $\mathbf{\Sigma P_0}$, $\mathbf{\Sigma Q_0}$ τα αντίστοιχα μεγέθη στο χρόνο βάσης.

Παράδειγμα

Το έτος 1986 η αμοιβή του εργάτη στην συγκομιδή της ελιάς σε ένα τόπο ήταν ανά ώρα 1000 δραχμές και το μεταφορικό κόστος του ενός σακιού ελιών προς τον ελαιοτριβείο ήταν 100 δραχμές. Το έτος 1996 οι αντίστοιχες τιμές ήταν 2000 και 300 δραχμές. Ο αστάθμητος αριθμοδείκτης για το έτος 1996 με έτος βάσης το 1986 είναι:

$$I_P = \frac{\Sigma P_n}{\Sigma P_o} \cdot 100 = \frac{2.000 + 300}{1.000 + 100} \cdot 100 = 209{,}1 \; (\text{με}\, 1986 = 100).$$

Αν θεωρήσουμε ως αμοιβή τα χρήματα που παίρνει ο εργάτης το 8ωρο τότε θα έχουμε για το αστάθμητο αριθμοδείκτη

$$I_P = \frac{\Sigma P_n}{\Sigma P_o} \cdot 100 = \frac{16.000 + 300}{8.000 + 100} \cdot 100 = 201{,}2 \; (\text{με}\, 1986 = 100).$$

Παρατηρούμε ότι μεταβάλλεται ο αστάθμητος αριθμοδείκτης όταν αλλάξουμε την μονάδα που εκφράζεται η αμοιβή της εργασίας και για τον λόγο αυτό ο αστάθμητος αριθμοδείκτης δεν χρησιμοποιείται γενικά εφόσον παρουσιάζει σοβαρό μειονέκτημα.

Ένας τρόπος για να περιορισθεί αυτή η αδυναμία του αστάθμητου αριθμοδείκτη είναι να ληφθούν **οι σχετικές τιμές** και όχι οι απόλυτες τιμές, τότε συμπίπτει ο **αριθμοδείκτης με το μέσο αριθμητικό των σχετικών τιμών,** δηλαδή

$$I_P = \frac{\Sigma \frac{P_n}{P_o}}{k} \cdot 100$$, όπου κ ο αριθμός των αγαθών.

Για το παράδειγμα στο οποίο αναφερόμαστε θα έχουμε τους σχετικούς λόγους $\frac{2.000}{1.000}, \frac{100}{300}$ και κ = 2 οπότε

$$I_P = \frac{\frac{2.000}{1.000} + \frac{300}{100}}{2} \cdot 100 = 250 \ (\mu\varepsilon\ 1986 = 100)$$

9.3.2. Σταθμισμένος σύνθετος αριθμοδείκτης

Επειδή η σημασία των αγαθών σε ποσότητα και σε τιμές πρέπει να λαμβάνεται υπόψη στον προσδιορισμό των αριθμοδεικτών τιμών ή τιμαρίθμων αλλά και των αριθμοδεικτών ποσοτήτων χρησιμοποιούνται οι σταθμισμένοι αριθμοδείκτες με κατάλληλους συντελεστές που καλούνται **συντελεστές στάθμισης.** Στους αριθμοδείκτες τιμών (τιμάριθμους) χρησιμοποιούμε για συντελεστές στάθμισης τις ποσότητες των αγαθών που καταναλώθηκαν σε ένα ορισμένο χρονικό διάστημα, ενώ στους αριθμοδείκτες ποσοτήτων τις τιμές των αγαθών που καταναλώθηκαν.

- **Σταθμισμένος σύνθετος αριθμοδείκτης τιμών ή τιμάριθμος**

Ειδικότερα όταν αναφερόμαστε στους τιμαρίθμους και πρέπει να λάβουμε υπόψη μας τις ποσότητες που καταναλώθηκαν μπαίνει το ερώτημα ποιες ποσότητες αυτές που καταναλώθηκαν την περίοδο βάσης ή αυτές που καταναλώθηκαν κατά την εξεταζόμενη περίοδο;

Αν χρησιμοποιηθούν για **συντελεστές στάθμισης οι ποσότητες των αγαθών που καταναλώθηκαν κατά την περίοδο βάσης** θα έχουμε για τον σταθμισμένο τιμάριθμο τον τύπο:

$$I_P = \frac{\Sigma P_n Q_O}{\Sigma P_o Q_O} \cdot 100$$ **δείκτης Laspeyres**

P_n είναι η τιμή ενός αγαθού το έτος-n που μας ενδιαφέρει P_o είναι η τιμή του αγαθού το έτος βάσης και Q_o η ποσότητα του αγαθού το έτος βάσης.

Εάν χρησιμοποιηθούν για **συντελεστές στάθμισης οι ποσότητες που καταναλώθηκαν την εξεταζόμενη περίοδο** θα έχουμε για τον σταθμισμένο τιμάριθμο τον τύπο:

$$I_P = \frac{\Sigma P_n Q_n}{\Sigma P_o Q_n} \cdot 100$$ **δείκτης Paasche**

P_n είναι η τιμή ενός αγαθού το έτος –n που μας ενδιαφέρει, P_o είναι η τιμή του αγαθού το έτος βάσης και Q_n η ποσότητα ενός αγαθού το έτος –n που μας ενδιαφέρει

Εάν οι δύο αριθμοδείκτες τιμών διαφέρουν σημαντικά λαμβάνουμε τον μέσο όρο αυτών:

$$I_P = \frac{\frac{\Sigma P_n Q_o}{\Sigma P_O Q_O} + \frac{\Sigma P_n Q_n}{\Sigma P_O Q_n}}{2} \cdot 100$$

Μια καλύτερη προσέγγιση των δύο δεικτών θεωρείται ο γεωμετρικός μέσος όρος ο καλούμενος **ιδανικός δείκτης Fisher** που είναι:

$$I_P = \sqrt{\frac{\Sigma P_n Q_O}{\Sigma P_O Q_O} \frac{\Sigma P_n Q_n}{\Sigma P_o Q_n}} \cdot 100$$

Παρατήρηση

Αρκετές φορές στους υπολογισμούς σταθμισμένων τιμαρίθμων λαμβάνεται η ποσότητα του **έτους απογραφής.**

Παράδειγμα

Στον πίνακα που ακολουθεί δίνονται τα δεδομένα τιμών (σε δραχμές) και ποσοτήτων (σε κιλά) τεσσάρων αγαθών.

Πίνακας

	Τιμές		**Ποσότητες**	
Αγαθά	**Περίοδος 0**	**Περίοδος 1**	**Περίοδος 0**	**Περίοδος 1**
Α	5	10	50	30
Β	10	15	6	4
Γ	12	20	10	7
Δ	20	30	4	3

Να βρεθούν:

1. Οι αριθμοδείκτες τιμών (τιμάριθμοι)
    α. Ο αστάθμητος αριθμοδείκτης τιμών της περιόδου 1 με βάση την περίοδο 0.
    β. Ο αστάθμητος αριθμοδείκτης των σχετικών τιμών της περιόδου 1 με βάση την περίοδο 0.
    γ. Του τύπου Laspeyres της περιόδου 1 με βάση την περίοδο 0.
    δ. Του τύπου Paasche της περιόδου 1 με βάση την περίοδο 0 και να εξηγηθούν τα αποτελέσματα.
2. Ο ιδανικός τιμάριθμος Fisher.

1. Δημιουργούμε τον παρακάτω πίνακα

Πίνακας

Αγαθά	P_0	P_1	Q_O	Q_1	P_OQ_O	P_1Q_O	P_OQ_1	P_1Q_1
Α	5	10	50	30	250	500	150	300
Β	10	15	6	4	60	90	40	60
Γ	12	20	10	7	120	200	84	140
Δ	20	30	4	3	80	120	60	90
Σύνολο	47	75	-	-	510	910	334	590
	ΣP_0	ΣP_1			ΣP_OQ_O	ΣP_1Q_O	ΣP_OQ_1	ΣP_1Q_1

α. $I_P = \frac{\Sigma P_1}{\Sigma P_o} \cdot 100 = \frac{75}{47} \cdot 100 = 159{,}57 \text{ (περίοδος } 0 = 100)$

β. $I_P = \frac{\Sigma \frac{P_1}{P_0}}{\kappa} \cdot 100 = \frac{\frac{10}{5} + \frac{15}{10} + \frac{20}{12} + \frac{30}{20}}{4} \cdot 100 = 166{,}5 \text{ (περίοδος } 0 = 100)$

γ. Τιμάριθμος Laspeyres

$$I_P = \frac{\Sigma P_1 Q_O}{\Sigma P_o Q_O} \cdot 100 = \frac{910}{510} \cdot 100 = 178{,}43 \text{ (περίοδος } 0 = 100)$$

δ. Τιμάριθμος Paasche

$$I_P = \frac{\Sigma P_1 Q_1}{\Sigma P_o Q_1} \cdot 100 = \frac{590}{334} \cdot 100 = 176{,}65 \text{ (περίοδος } 0 = 100)$$

Παρατηρούμε ότι το μέσο επίπεδο τιμών των τεσσάρων αγαθών σημειώνει αύξηση κατά 78,43% κατά Laspeyres την περίοδο 1 σε σχέση με την περίοδο 0 που είναι περίοδος βάσης ενώ κατά Paasche έχουμε αντίστοιχη αύξηση κατά 76,65%.

Παρατήρηση

Συνήθως ο τύπος Laspeyres του υπερεκτιμά την πραγματική μεταβολή των τιμών ενώ ο τύπος του Paasche την υποεκτιμά.

3. Ο ιδανικός δείκτης Fisher είναι:

$$I_P = \sqrt{\frac{\Sigma P_n Q_O}{\Sigma P_O Q_O}\frac{\Sigma P_n Q_n}{\Sigma P_o Q_n}} \cdot 100 = \sqrt{178{,}43x176{,}65} = 177{,}54 \ (\text{περίοδος } 0 = 100)$$

- **Σταθμισμένος σύνθετος αριθμοδείκτης ποσοτήτων (όγκου)**

Όταν αναφερόμαστε στον αριθμοδείκτη ποσοτήτων πρέπει να λάβουμε υπόψη μας τις τιμές των αγαθών που καταναλώθηκαν ως συντελεστές στάθμισης και μπαίνει πάλι το ερώτημα ποιες τιμές θα πάρουμε αυτές των αγαθών που καταναλώθηκαν την περίοδο βάσης ή αυτές που καταναλώθηκαν κατά την εξεταζόμενη περίοδο;

Αν χρησιμοποιηθούν για **συντελεστές στάθμισης οι τιμές των αγαθών που καταναλώθηκαν κατά την περίοδο βάσης** θα έχουμε για τον σταθμισμένο τιμάριθμο τον τύπο:

$$I_q = \frac{\Sigma Q_n P_n}{\Sigma Q_o P_O} \cdot 100 \quad \textbf{δείκτης Laspeyres}$$

όπου Q_n αναφέρεται στη ποσότητα ενός αγαθού το έτος – n που μας ενδιαφέρει και Q_0 η ποσότητα του αγαθού το έτος βάσης, P_o είναι η τιμή του αγαθού το έτος βάσης.

Εάν χρησιμοποιηθούν για **συντελεστές στάθμισης οι τιμές των αγαθών που καταναλώθηκαν την εξεταζόμενη περίοδο** θα έχουμε τον σταθμισμένο τιμάριθμο με τύπο:

$$I_q = \frac{\Sigma Q_n P_n}{\Sigma Q_o P_n} \cdot 100 \quad \textbf{δείκτης Paasche}$$

όπου Q_n αναφέρεται στη ποσότητα ενός αγαθού το έτος - n που μας ενδιαφέρει και Q_0 η ποσότητα του αγαθού το έτος βάσης, P_n είναι η τιμή ενός αγαθού το έτος – n που μας ενδιαφέρει.

Μια καλύτερη προσέγγιση των δύο δεικτών θεωρείται ο γεωμετρικός μέσος όρος ο καλούμενος **ιδανικός δείκτης Fisher** που είναι:

$$I_P = \sqrt{\frac{\Sigma Q_n P_O}{\Sigma Q_O P_O}\frac{\Sigma Q_n P_n}{\Sigma Q_o P_n}} \cdot 100$$

Παράδειγμα

Στον παρακάτω πίνακα δίνονται τα δεδομένα τιμών και ποσοτήτων τεσσάρων γεωργικών προϊόντων κατά τις περιόδους 0 και 1. Να υπολογισθούν οι **αριθμοδείκτες όγκου** της περιόδου 1 με βάση την περίοδο 0:

α. Του τύπου Laspeyres

β. Του τύπου Paasche

γ. Του τύπου Fisher
να εξηγηθούν τα αποτελέσματα.

Στον πίνακα των δεδομένων γίνονται και οι υπολογισμοί.

Γεωργικά προϊόντα	P_0	Q_0	P_1	Q_1	Q_0P_O	Q_1P_O	Q_OP_1	Q_1P_1
Α	3	13	5	16	39	48	65	80
Β	7	26	4	30	182	210	104	120
Γ	6	32	2	18	192	108	64	36
Δ	2	38	5	20	76	40	190	100
Σύνολο	-	-	-	-	489	406	423	336

$\Sigma Q_0P_O = 489$, $\Sigma Q_1P_O = 406$, $\Sigma Q_0P_1 = 423$, $\Sigma Q_1P_1 = 336$

$$\text{α. } I_q = \frac{\Sigma Q_1P_O}{\Sigma Q_oP_O}\cdot 100 = \frac{406}{489}\cdot 100 = 83{,}03$$

$$\text{β. } I_q = \frac{\Sigma Q_1P_1}{\Sigma Q_oP_1}\cdot 100 = \frac{336}{423}\cdot 100 = 79{,}43$$

$$\text{γ. } I_q = \sqrt{\frac{\Sigma Q_1P_O}{\Sigma Q_OP_O}\frac{\Sigma Q_1P_1}{\Sigma Q_oP_1}}\cdot 100 = \sqrt{83{,}03\text{x}79{,}43} = 81{,}21$$

Παρατηρούμε ότι ο όγκος παραγωγής των γεωργικών προϊόντων που εξετάζουμε κατά την περίοδο 1 σε σχέση με την περίοδο 0 σημείωσε **μείωση** κατά 100-83,03=16,97% σύμφωνα με τον τύπο Laspeyres, κατά 20,57% σύμφωνα με τον τύπο του Paasche και κατά 18,79% σύμφωνα με τον τύπο του Fisher.

- Σταθμισμένος σύνθετος αριθμοδείκτης αξίας

$$I_v = \frac{\Sigma P_n Q_n}{\Sigma P_o Q_O} \cdot 100$$

Παρατήρηση
Έχουμε:

$$I_P = \frac{\Sigma P_n Q_O}{\Sigma P_o Q_O} \cdot 100$$ **δείκτης τιμών Laspeyres**

$$I_q = \frac{\Sigma Q_n P_O}{\Sigma Q_o P_O} \cdot 100$$ **δείκτης όγκου Laspeyres**

$$I_P = \frac{\Sigma P_n Q_n}{\Sigma P_o Q_n} \cdot 100$$ **δείκτης τιμών Paasche**

$$I_q = \frac{\Sigma Q_n P_n}{\Sigma Q_o P_n} \cdot 100$$ **δείκτης όγκου Paasche**

$$I_v = \frac{\Sigma P_n Q_n}{\Sigma P_o Q_o} \cdot 100$$ **αριθμοδείκτης αξίας**

Ισχύει:
(δείκτης τιμών Laspeyres)x(δείκτης όγκου Paasche)= αριθμοδείκτης αξίας

και

(δείκτης τιμών Paasche)x(δείκτης όγκου Laspeyres)= αριθμοδείκτης αξίας

9.4. Δείκτες πληθωρισμού – αποπληθωρισμού και αποπληθωρισμός των χρηματικών αξίων.

Ο τιμάριθμος είναι όπως αναφέρθηκε ένας σταθμισμένος αριθμοδείκτης τιμών, και εκφράζει τη μεταβολή του κόστους ζωής. Για να υπολογισθεί λαμβάνονται υπόψη οι τιμές βασικών ειδών διατροφής, ένδυσης, υπόδησης, βασικών υπηρεσιών όπως συγκοινωνίες κλπ. και από την ΕΣΥΕ εκτιμάται το ποσοστό βάρους των εν λόγω ειδών στον τιμάριθμο μετά από δειγματοληπτικές έρευνες που πραγματοποιεί η ΕΣΥΕ κατά καιρούς στα νοικοκυριά της χώρας.

Ο τιμάριθμος μας βοηθά να συσχετίσουμε τα χρηματικά ποσά που δαπανώνται σε διαφορετικές περιόδους για το ίδιο είδος ή που δίνονται ως αμοιβές στο ίδιο άτομο στις διαφορετικές περιόδους.

Αυτό γίνεται με την χρησιμοποίηση των **δεικτών πληθωρισμού** και **αποπληθωρισμού.**

Ο **δείκτης πληθωρισμού** μιας περιόδου (π.χ. ενός έτους) προκύπτει από τον τιμάριθμο της περιόδου αν διαιρεθεί αυτός με το 100:

$$\text{Ο δείκτης πληθωρισμού} = \frac{\text{τιμάριθμος}}{100}$$

Ο δείκτης αποπληθωρισμού μιας περιόδου (ενός έτους) είναι το πηλίκιο του δείκτη πληθωρισμού της περιόδου βάσης προς το δείκτη πληθωρισμού της περιόδου που εξετάζουμε:

$$\text{Ο δείκτης αποπληθωρισμού} = \frac{\text{δείκτης πληθωρισμού περιόδου βάσης}}{\text{δείκτης πληθωρισμού της περιόδου}}$$

Παράδειγμα

Έτος	**τιμάριθμος**	**δείκτης πληθωρισμού**	**Δείκτης αποπληθωρισμού**
1990	100	1	1
1991	118	1,18	1/1,18=0,85
1992	130	1,30	1/1,30=0,77
1993	146	1,46	1/1,46=0,68
1994	175	1,75	1/1,75=0,57

Αν θελήσουμε να αποπληθωρίσουμε την τιμή ενός προϊόντος, το εισόδημα ενός εργαζόμενου κλπ. πρέπει να πολλαπλασιάσουμε τις χρηματικές αυτές αξίες με τον αντίστοιχο δείκτη αποπληθωρισμού.

Έστω ότι στο παραπάνω παράδειγμα η τιμή του βαμβακιού το έτος 1990 ήταν 130 δραχμές το κιλό και το 1994 ήταν 150 δραχμές το κιλό για να αποπληθωρίσουμε την τιμή του 1994 με έτος βάσης το 1990 πολλαπλασιάζουμε την τιμή 150 με τον αντίστοιχο δείκτη αποπληθωρισμού του 1994 που είναι το 0,57 και έχουμε

150 x 0,57 = 85,5 δραχμές το κιλό

Η τιμή του βαμβακιού το 1994, σε τιμές του 1990, είναι 85,5 δραχμές το κιλό, δηλαδή πιο κάτω από ότι ήταν το 1990 κατά 44,5 δραχμές. Η τιμή αυτή αντιστοιχεί στο 65,76% της τιμής του 1990, συνεπώς έχουμε μια μείωση κατά 34,24%.

Παράδειγμα

Δίνονται στο κάτωθι πίνακα οι τιμές ενός αγροτικού προϊόντος στα έτη από το 1984 έως το 1993. Με δεδομένο ότι ο τιμάριθμος αυξάνει κατά 10% κατ΄ έτος να βρεθούν οι δείκτες πληθωρισμού και αποπληθωρισμού των αντιστοίχων ετών όπως και οι τιμές του αγροτικού προϊόντος στα έτη αυτά σε τιμές του έτους 1984 που θεωρείται ως έτος βάσης.

Οι υπολογισμοί γίνονται στον πίνακα

Πίνακας

Έτη	**Τιμή προϊόντος**	**Δείκτης τιμών**	**τιμάριθμος**	**Δείκτης πληθωρισμού**	**Δείκτης αποπληθωρισμού**	**Σε τιμές 1984**
1984	18,16	100	100	1	1	18,16
1985	19,20	105,7	110	1,1	0,91	17,47
1986	23,10	127,2	121	1,21	0,83	19,17
1987	25,36	139,6	133,1	1,33	0,72	18,26
1988	28,52	157,0	146,4	1,46	0,68	19,39
1989	31,42	173,0	161,1	1,61	0,62	19,48
1990	33,05	182,0	177,1	1,77	0,56	18,51
1991	370,8	204,2	194,9	1,95	0,51	18,91
1992	39,41	217	214,4	2,14	0,47	18,52
1993	41,56	229	235,8	2,36	0,42	17,46

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Αποστολόπουλος Θ. : «Στατιστική Επιχειρήσεων», Αθήνα 1991.
2. Αποστολόπουλος Θ. : «Περιγραφική Στατιστική», Αθήνα 1996

3. Δρακάτος Κ. : «Στατιστική», εκδόσεις Παπαζήση, Αθήνα 1984.
4. Δρακάτος Κ. : «Περιγραφική Οικονομική Στατιστική», εκδόσεις Παπαζήση, Αθήνα 1993.
5. Δρακάτος Κ. «Ασκήσεις Στατιστικής», εκδόσεις Παπαζήση, Αθήνα 1984.
6. Ζαχαροπούλου Χ. : «Στατιστική, μέθοδοι-εφαρμογές», Θεσ/νίκη 1996.
7. Λαμπράκη Δ. : «Στατιστική», Αθήνα 1980.
8. Μπένος Β.Κ. : «Στατιστική – Περιγραφική Στατιστική», εκδόσεις Σταμούλη, Πειραιάς 1973.
9. Μπένος Β.Κ. : «Οικονομική Στατιστική», εκδόσεις Σταμούλη, Πειραιάς 1973.
10. Μπένος Β.Κ. : «Στατιστική Επιχειρήσεων», εκδόσεις Σταμούλη, Πειραιάς 1996.
11. Μπόρα Ε. – Κολύβα Φ. : «Στατιστική – Θεωρία – Εφαρμογές», εκδόσεις Ζήτη, Θεσσαλονίκη 1999.
12. Μπόρα Ε. – Μωυσιάδης Χ. : «Εφαρμοσμένη Στατιστική», εκδόσεις Ζήτη, Θεσσαλονίκη 1997.
13. Μάνος Β. : «Γεωργική Οικονομική Στατιστική», εκδόσεις Ζήτη, Θεσ/νίκη 1997.
14. Παναγιωτακόπουλος Δ. : «Εφαρμογές πιθανοτήτων & Στατιστική», Ξάνθη 1997.
15. Παπαδημητρίου Γ. : «Στατιστική» περιγραφική Στατιστική Τεύχος 1, εκδόσεις Παρατηρητής, Θεσ/νίκη 1986.
16. Περσίδης Σ. : «Πιθανότητες και Στατιστική» Schaurns Outline Series. Mc Graun-Hill, N. York. ΕΣΠΙ, Αθήνα 1977.
17. Τζωρτζόπουλος Θ. : «Αριθμοδείκτες» , Αθήνα 1978.
18. Τζωρτζόπουλος Θ. : «Ανάλυση χρονολογικών σειρών» , Αθήνα 1979.
19. GRAIS B. : "Methodes Statistiques" Paris 1978.
20. WAYNE R. OTT: "Environmental Statistics and Data Analysis".
21. P. WHITEHEAD-G. WHITEHEAD: "Statistics for Business" 1992 PITMAN PUBLISHING LONDON.