Κεφάλαιο 3

Στοιχεία Στατιστικής

Η στατιστική ασχολείται με τις εφαρμογές της θεωρίας των πιθανοτήτων τυχαίων μεταβλητών σε πραγματικά προβλήματα και συνίσταται στην εξαγωγή συμπερασμάτων που βασίζονται στις παρατηρήσεις. Εύκολα μπορεί να καταλάβει κάποιος τη σύνδεση πιθανολογικών εννοιών με την πραγματικότητα από την προσέγγιση της πιθανότητας που δώσαμε στην (2.1). Στην πραγματικότητα έχουμε n παρατηρήσεις που αποτελούν το δείγμα και αν η τιμή x_i μιας διακριτής τ.μ. X εμφανίζεται n_i φορές στο δείγμα, μπορούμε να προσεγγίσουμε την πιθανότητα εμφάνισης της x_i, $p = \mathrm{P}(X = x_i)$, ως

$$\hat{p} = n_i/n. \tag{3.1}$$

Η τιμή $\hat{p}$ αποτελεί την εκτίμηση της p με βάση το δείγμα. Σημειώνεται ότι το p μπορούμε να το ονομάσουμε και αναλογία εμφάνισης της τιμής x_i στο σύνολο των δυνατών τιμών της X. Η **εκτίμηση** αυτή είναι **σημειακή** (point estimation) και δίνει την καλύτερη προσέγγιση με μια τιμή που μπορούμε να δώσουμε στην πραγματική αλλά άγνωστη αναλογία p με βάση το δείγμα. Σε πολλές περιπτώσεις θα θέλαμε να εκτιμήσουμε ένα **διάστημα εμπιστοσύνης** (confidence interval) σε κάποιο επίπεδο σημαντικότητας α (ή αντίστοιχα επίπεδο εμπιστοσύνης $1 - \alpha$) που να περιέχει την πραγματική αλλά άγνωστη αναλογία p. Σε άλλες περιπτώσεις μας ενδιαφέρει μόνο να ελέγξουμε αν η αναλογία μπορεί να πάρει ή να υπερβεί κάποια τιμή και για αυτό κάνουμε **έλεγχο υπόθεσης** (hypothesis test).

Σε πολλές μελέτες τα δεδομένα είναι αριθμητικά και το ενδιαφέρον είναι στον προσδιορισμό του κέντρου (που ορίζεται με τη μέση τιμή ή διάμεσο) της κατανομής της παρατηρούμενης τ.μ. ή της διασποράς της. Γενικά για να εκτιμήσουμε με σημειακή εκτίμηση ή διάστημα εμπιστοσύνης κάποια άγνωστη παράμετρο θ (π.χ. μέση τιμή ή διασπορά) ή να ελέγξουμε αν αυτή μπορεί να πάρει κάποια τιμή υπάρχουν τρεις προσεγγίσεις. Η **παραμετρική**

προσέγγιση (parametric approach) υποθέτει ότι τα δεδομένα του προβλήματος προέρχονται από κάποια γνωστή κατανομή. Αυτή η προσέγγιση είναι απλή στην πραγματοποίηση της. Από τη γνωστή κατανομή υπολογίζεται η κατανομή του εκτιμητή και στη συνέχεια το διάστημα εμπιστοσύνης ή σχηματίζεται η πιθανότητα της ορθότητας της μηδενικής υπόθεσης του ελέγχου. Η προσέγγιση αυτή είναι η πιο ακριβής αν η υπόθεση για την κατανομή είναι σωστή. Αντίθετα η **μη-παραμετρική** (nonparametric) προσέγγιση δεν υποθέτει κάποια γνωστή κατανομή για τα δεδομένα. Είναι λιγότερη ακριβής για γνωστές κατανομές αλλά είναι πιο κατάλληλη από την παραμετρική όταν τα δεδομένα δεν προσαρμόζονται καλά σε κάποια γνωστή κατανομή. Η τρίτη προσέγγιση επίσης δε θεωρεί γνωστή κατανομή για τα δεδομένα και χρησιμοποιεί **επαναδειγματοληψία** (resampling) για να δημιουργήσει νέα δείγματα. Από αυτά τα δείγματα υπολογίζεται ένα πλήθος τιμών του εκτιμητή, ένα για κάθε δείγμα, σχηματίζεται η κατανομή του και υπολογίζεται έτσι το διάστημα εμπιστοσύνης ή γίνεται ο έλεγχος υπόθεσης. Αυτή είναι η πιο ακριβής προσέγγιση για πραγματικά προβλήματα, όπου συνήθως τα δεδομένα δεν προσαρμόζονται πιστά σε γνωστές κατανομές, αλλά η ακρίβεια της απαιτεί πολλούς περισσότερους υπολογισμούς. Για αυτό και αυτή η προσέγγιση άρχισε να χρησιμοποιείται ευρέως τα τελευταία χρόνια με την ανάπτυξη ισχυρής υπολογιστικής τεχνολογίας.

Στο κεφάλαιο αυτό θα επικεντρωθούμε στην πρώτη προσέγγιση για τον υπολογισμό διαστημάτων εμπιστοσύνης παραμέτρου και την πραγματοποίηση ελέγχου υπόθεσης παραμέτρου. Ειδικότερα θα μελετήσουμε τις παραμέτρους της μέσης τιμής και της διασποράς. Τέλος θα γενικεύσουμε τη χρήση του ελέγχου υπόθεσης και σε άλλα προβλήματα, όπως στην προσαρμογή της κατανομής των δεδομένων σε κάποια γνωστή κατανομή.

3.1 Σημειακή εκτίμηση

Η σημειακή εκτίμηση μιας παραμέτρου ϑ είναι το στατιστικό (ή η στατιστική) $\hat{\vartheta}$ που υπολογίζουμε από το δείγμα για να προσδιορίσουμε την άγνωστη ϑ, δηλαδή είναι μια τιμή, που υπολογίζεται με βάση τα δεδομένα του δείγματος και αντιπροσωπεύει την πραγματική τιμή της αντίστοιχης παράμετρου του πληθυσμού.

Έστω X μια τ.μ. με αθροιστική συνάρτηση κατανομής $F_X(x;\vartheta)$ που εξαρτάται από την παράμετρο ϑ την οποία θέλουμε να εκτιμήσουμε (το ';' στον παραπάνω συμβολισμό ξεχωρίζει μεταβλητές από παραμέτρους). Έστω ακόμα ότι έχουμε παρατηρήσεις $\{x_1,\ldots,x_n\}$ της X από ένα δείγμα μεγέθους n. Τότε η σημειακή εκτίμηση της ϑ δίνεται από μια *εκτιμήτρια συνάρτηση* των τιμών του δείγματος, $\hat{\vartheta} = g(x_1,\ldots,x_n)$ και το $\hat{\vartheta}$ ονομάζεται **εκτιμητής** (estimator) της ϑ.

Το βασικό (και λεπτό σημείο) στην εκτίμηση παραμέτρων είναι να καταλάβουμε ότι η $\hat{\vartheta}$ είναι τυχαία μεταβλητή. Οι παρατηρήσεις $\{x_1, \dots, x_n\}$ παίρνουν τυχαίες τιμές σε κάθε δείγμα μεγέθους n και πάντα με την ίδια κατανομή $F_X(x; \vartheta)$. Άρα μπορούμε να θεωρήσουμε τις $\{x_1, \dots, x_n\}$ ως τ.μ. και τότε και η $\hat{\vartheta}$ είναι τ.μ. ως συνάρτηση n τ.μ.. Βέβαια για ένα συγκεκριμένο δείγμα οι $\{x_1, \dots, x_n\}$ παίρνουν πραγματικές τιμές που δίνουν την τιμή της $\hat{\vartheta}$ για αυτό το δείγμα.

Ως τ.μ. η $\hat{\vartheta}$ ακολουθεί κάποια κατανομή με μέση τιμή $\mu_{\hat{\vartheta}} \equiv \mathrm{E}[\hat{\vartheta}]$ και διασπορά $\sigma^2_{\hat{\vartheta}} \equiv \mathrm{Var}[\hat{\vartheta}]$.

3.1.1 Μέση τιμή και διασπορά

Δύο σημαντικές παράμετροι μιας τ.μ. X που θέλουμε να εκτιμήσουμε είναι η μέση τιμή της μ και η διασπορά της σ^2. Ο εκτιμητής της μ δίνεται από το γνωστό μέσο όρο των $\{x_1, \dots, x_n\}$ και ονομάζεται *δειγματική μέση τιμή* ή απλά *μέσος όρος* $\bar{x}$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{3.2}$$

Ο εκτιμητής της διασποράς σ^2 είναι η *δειγματική διασπορά* s^2

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right). \tag{3.3}$$

Ένας άλλος εκτιμητής της σ^2 δίνεται ως

$$\tilde{s}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2. \tag{3.4}$$

Οι εκτιμητές s^2 και $\tilde{s}^2$ διαφέρουν μόνο ως προς το συντελεστή του αθροίσματος ($\frac{1}{n-1}$ και $\frac{1}{n}$ αντίστοιχα). Για μεγάλο n οι δύο εκτιμητές συγκλίνουν στην ίδια τιμή.

3.1.2 Βαθμοί ελευθερίας

Η χρήση του $n-1$ στον τύπο της διασποράς στην (3.3) είναι σε συμφωνία με τους *βαθμούς ελευθερίας* (degrees of freedom) του προβλήματος εκτίμησης της διασποράς. Οι βαθμοί ελευθερίας δηλώνουν τις ελεύθερες (τυχαίες) τιμές που υπάρχουν στο πρόβλημα που μελετάμε. Εδώ αρχικά οι βαθμοί ελευθερίας είναι n, όσες και οι παρατηρήσεις στο δείγμα, αλλά επειδή στον ορισμό της δειγματικής διασποράς περιλαμβάνεται η δειγματική μέση τιμή δεσμεύονται οι n ελεύθερες τιμές με την συνθήκη να ικανοποιούν την εξίσωση (3.2), δηλαδή να δίνουν την $\bar{x}$. Έτσι χάνεται ένας βαθμός ελευθερίας και οι βαθμοί έλευθερίας είναι $n-1$.

3.1.3 Κριτήρια καλών εκτιμητών

Παραπάνω ορίσαμε κάπως αυθαίρετα τους εκτιμητές της μέσης τιμής μ και της διασποράς σ^2 χωρίς να γνωρίζουμε αν είναι 'καλοί' εκτιμητές ή όχι. Γενικά για τον ορισμό βέλτιστου εκτιμητή $\hat{\vartheta}$ κάποιας παραμέτρου ϑ θέτουμε κάποια βασικά κριτήρια που αποτελούν και ιδιότητες του εκτιμητή $\hat{\vartheta}$. Επικεντρώνουμε την προσοχή μας σε δύο βασικές ιδιότητες ενός εκτιμητή, όπου η πρώτη έχει να κάνει με τη μέση τιμή του $\mu_{\hat{\vartheta}}$ και η δεύτερη με τη διασπορά του $\sigma^2_{\hat{\vartheta}}$.

Ο εκτιμητής $\hat{\vartheta}$ είναι **αμερόληπτος** (unbiased) αν η μέση τιμή του είναι ίση με την παράμετρο ϑ, δηλαδή αν ισχύει

$$\mathrm{E}[\hat{\vartheta}] = \vartheta.$$

Αλλιώς λέγεται μεροληπτικός με μεροληψία

$$b(\hat{\vartheta}) = \mathrm{E}[\hat{\vartheta}] - \vartheta.$$

Ένα δεύτερο σημαντικό κριτήριο καλού εκτιμητή είναι η **αποτελεσματικότητα** (efficiency) αναφέρεται στη διασπορά του εκτιμητή και δίνεται συγκριτικά. Ένας εκτιμητής $\hat{\vartheta}_1$ της ϑ είναι πιο αποτελεσματικός (effective) από έναν άλλο εκτιμητή $\hat{\vartheta}_2$ αν έχει μικρότερη διασπορά, αν δηλαδή ισχύει $\sigma^2_{\hat{\vartheta}_1} < \sigma^2_{\hat{\vartheta}_2}$.

Οι εκτιμητές $\bar{x}$ για την παραμέτρο μ και s^2 για την παράμετρο σ^2, που ορίσαμε αυθαίρετα, είναι και οι δύο αμερόληπτοι και οι πιο αποτελεσματικοί (δες επίσης 3.2.1). Ο εκτιμητής $\tilde{s}^2$ είναι μεροληπτικός εκτιμητής της σ^2 με μεροληψία $b(\tilde{s}^2) = -\sigma^2/n$. Είναι κατανοητό ότι καθώς το μέγεθος του δείγματος αυξάνει η μεροληψία τείνει προς το μηδέν και για αυτό θεωρούμε πως ο εκτιμητής $\tilde{s}^2$ είναι *ασυμπτωτικά* (asymptotic) αμερόληπτος.

Σε κάποια προβλήματα είναι δύσκολο ή αδύνατο να βρούμε εκτιμητή που είναι και αμερόληπτος και ο πιο αποτελεσματικός. Για αυτό συνθέτουμε το κριτήριο του **μέσου τετραγωνικού σφάλματος** (mean square error) της εκτίμησης ως το άθροισμα της διασποράς του εκτιμητή και του τετραγώνου της μεροληψίας

$$\mathrm{MSE}[\hat{\vartheta}] = b(\hat{\vartheta})^2 + \sigma^2_{\hat{\vartheta}} = (\mathrm{E}[\hat{\vartheta}] - \vartheta)^2 + \mathrm{E}[\hat{\vartheta}^2] - (\mathrm{E}[\hat{\vartheta}])^2 = \mathrm{E}[(\hat{\vartheta} - \vartheta)^2]. \quad (3.5)$$

Στον ορισμό των εκτιμητών $\bar{x}$ και s^2 δεν κάναμε κάποια υπόθεση για την κατανομή της τ.μ. X και άρα μπορούμε να τους χρησιμοποιήσουμε για οποιαδήποτε τ.μ. X που παρατηρούμε.

3.1.4 Μέθοδος της μέγιστης πιθανοφάνειας

Η σημαντικότερη μέθοδος της στατιστικής για την εκτίμηση παραμέτρων είναι η μέθοδος της **μέγιστης πιθανοφάνειας** (maximum likelihood). Η μέθοδος αυτή δίνει την εκτίμηση που έχει τη μέγιστη πιθανοφάνεια, δηλαδή δίνει την τιμή της παραμέτρου η οποία, μεταξύ όλων των δυνατών τιμών της παραμέτρου, είναι η πιο πιθανή με βάση το δείγμα.

Υποθέτουμε ότι η τ.μ. X έχει κάποια γνωστή κατανομή, δηλαδή γνωρίζουμε τη γενική μορφή της ασκ $F_X(x;\vartheta)$ και της σππ $f_X(x;\vartheta)$. Η παράμετρος ϑ της κατανομής είναι άγνωστη και θέλουμε να την εκτιμήσουμε από ένα δείγμα ανεξάρτητων παρατηρήσεων $\{x_1,\ldots,x_n\}$.

Επειδή $\{x_1,\ldots,x_n\}$ είναι ανεξάρτητες τ.μ., η πιθανότητα να τις παρατηρήσουμε σε ένα τυχαίο δείγμα μεγέθους n δίνεται από τη **συνάρτηση πιθανόφανειας** (likelihood function) ως προς ϑ

$$L(x_1,\ldots,x_n;\vartheta) = f(x_1;\vartheta)\cdots f(x_n;\vartheta).$$

Αν λοιπόν $L(x_1,\ldots,x_n;\vartheta_1) > L(x_1,\ldots,x_n;\vartheta_2)$ για δύο τιμές ϑ_1 και ϑ_2 της ϑ, τότε η τιμή ϑ_1 είναι πιο αληθοφανής από τη ϑ_2 γιατί δίνει μεγαλύτερη πιθανότητα να παρατηρήσουμε το συγκεκριμένο δείγμα των $\{x_1,\ldots,x_n\}$. Θέλουμε λοιπόν να βρούμε την 'πιο αληθοφανή' τιμή της ϑ, δηλαδή την τιμή $\hat{\vartheta}$ που μεγιστοποιεί τη $L(x_1,\ldots,x_n;\vartheta)$ ή καλύτερα (για ευκολότερους υπολογισμούς) τη $\log L(x_1,\ldots,x_n;\vartheta)$. Άρα ο **εκτιμητής μέγιστης πιθανοφάνειας** (maximum likelihood estimator) $\hat{\vartheta}$ βρίσκεται από τη σχέση

$$\frac{\partial \log L(x_1,\ldots,x_n;\vartheta)}{\partial\vartheta} = 0. \tag{3.6}$$

Αν θέλουμε να εκτιμήσουμε δύο ή περισσότερες παραμέτρους $\vartheta_1,\ldots,\vartheta_m$, η συνάρτηση πιθανόφανειας είναι $L(x_1,\ldots,x_n;\vartheta_1,\ldots,\vartheta_m)$ και οι εκτιμητές $\hat{\vartheta}_1,\ldots,\hat{\vartheta}_m$ βρίσκονται λύνοντας το σύστημα των m εξισώσεων

$$\frac{\partial \log L(x_1,\ldots,x_n;\vartheta_1,\ldots,\vartheta_m)}{\partial\vartheta_j} = 0 \quad \text{για} \quad j = 1,\ldots,m. \tag{3.7}$$

Παράδειγμα 3.1. Έχουμε ένα τυχαίο δείγμα $\{x_1,\ldots,x_n\}$ από κανονική κατανομή $\mathrm{N}(\mu,\sigma^2)$ και θέλουμε να εκτιμήσουμε τη μέση τιμή μ θεωρώντας τη σ^2 γνωστή. Η συνάρτηση πυκνότητας πιθανότητας της κανονικής κατανομής δίνεται από την (2.23). Η συνάρτηση πιθανόφανειας (για την οποία μόνο η παράμετρος μ είναι άγνωστη) είναι

$$L(x_1,\ldots,x_n;\mu) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right],$$

όπου $\exp(x) \equiv e^x$. Ο λογάριθμος της συνάρτησης πιθανόφανειας είναι

$$\log L(x_1,\ldots,x_n;\mu) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Ο εκτιμητής μέγιστης πιθανοφάνειας $\hat{\mu}$ βρίσκεται μηδενίζοντας την παράγωγο της $\log L$

$$\frac{\partial \log L}{\partial \mu} = 0 \quad \Rightarrow \quad \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0 \qquad (3.8)$$

που δίνει τη λύση

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}x_i = \bar{x},$$

δηλαδή είναι ίδιος με τον εκτιμητή $\bar{x}$ της μέσης τιμής μ που ορίσαμε για οποιαδήποτε κατανομή της τ.μ. X.

Παράδειγμα 3.2. Ας υποθέσουμε στο προηγούμενο παράδειγμα πως και η διασπορά σ^2 είναι άγνωστη. Τότε στην παραπάνω εξίσωση (3.8) προστίθεται και η εξίσωση

$$\frac{\partial \log L}{\partial \sigma^2} = 0 \quad \Rightarrow \quad -\frac{n}{2\sigma^2} + \frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0. \qquad (3.9)$$

Η επίλυση του συσήματος των εξισώσεων (3.8) και (3.9) δίνει την ίδια λύση για τη μ και για τη σ^2 έχουμε

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2.$$

Οι εκτιμητές μέγιστης πιθανοφάνειας λοιπόν για τη μέση τιμή μ και τη διασπορά σ^2 μιας τ.μ. που ακολουθεί κανονική κατανομή είναι απλά η δειγματική μέση τιμή και η δειγματική διασπορά αντίστοιχα, αλλά για τη διασπορά έχουμε τον ασυμπτωτικά αμερόληπτο εκτιμητή $\tilde{s}^2$ (σχέση (3.4)).

Η μέθοδος μέγιστης πιθανοφάνειας είναι η καλύτερη μέθοδος εκτίμησης αν γνωρίζουμε την κατανομή της τ.μ. X και μπορεί να εφαρμοσθεί σε οποιοδήποτε πρόβλημα εκτίμησης παραμέτρων από δείγμα ανεξάρτητων παρατηρήσεων.

3.2 Εκτίμηση διαστήματος εμπιστοσύνης

Η σημειακή εκτίμηση $\hat{\vartheta}$ από κάποιο δείγμα δεν περιέχει καμιά πληροφορία για την ακρίβεια της εκτίμησης της ϑ. Η εκτίμηση $\hat{\vartheta}$ που παίρνουμε από ένα δείγμα είναι μια τιμή που δε γνωρίζουμε πόσο κοντά είναι στην πραγματική τιμή της ϑ και επίσης η τιμή αυτή αλλάζει με το δείγμα. Για παράδειγμα, υπολογίζουμε τη δειγματική μέση τιμή $\bar{x}$ από ένα τυχαίο δείγμα μεγέθους n. Αν πάρουμε ένα άλλο τυχαίο δείγμα ίδιου μεγέθους, η τιμή της $\bar{x}$ θα είναι διαφορετική. Μπορεί να είναι πιο κοντά ή πιο μακριά στην πραγματική τιμή της μ απ' ότι αυτή από το προηγούμενο δείγμα. Γι αυτό στην εκτίμηση της ϑ είναι σημαντικό εκτός από τη σημειακή εκτίμηση $\hat{\vartheta}$ να υπολογίσουμε και διάστημα $[\vartheta_1, \vartheta_2]$ που να μπορούμε να πούμε με κάποια πιθανότητα $1 - \alpha$ ότι θα περιέχει την πραγματική τιμή της παραμέτρου ϑ (δηλαδή η πιθανότητα σφάλματος είναι α).

Στη συνέχεια θα παρουσιάσουμε την παραμετρική διαδικασία υπολογισμού διαστημάτων εμπιστοσύνης χρησιμοποιώντας ως παράμετρο ϑ τη μέση τιμή μ και τη διασπορά σ^2. Θεωρούμε και πάλι πως οι παρατηρήσεις $x_1, \ldots, x_n$ είναι ανεξάρτητες. Για τον υπολογισμό του διαστήματος εμπιστοσύνης θα πρέπει να γνωρίζουμε την κατανομή του εκτιμητή (π.χ. $\bar{x}$ ή s^2) και με βάση αυτήν την κατανομή ορίζεται το διάστημα εμπιστοσύνης για την παράμετρο.

3.2.1 Διάστημα εμπιστοσύνης της μέσης τιμής μ

Το διάστημα εμπιστοσύνης της μ υπολογίζεται με βάση την κατανομή της τ.μ. $\bar{x}$, που είναι ο καλύτερος εκτιμητής της μ. Πράγματι στην παράγραφο 3.1.3 αναφέρθηκε ότι η $\bar{x}$ είναι αμερόληπτος εκτιμητής της μ, δηλαδή ισχύει

$$\mu_{\bar{x}} \equiv \mathrm{E}[\bar{x}] = \mu \tag{3.10}$$

και άρα η μέση τιμή της $\bar{x}$ είναι το ίδιο το μ. Η διασπορά της $\bar{x}$ είναι

$$\sigma_{\bar{x}}^2 \equiv \mathrm{Var}[\bar{x}] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[x_i] = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}. \tag{3.11}$$

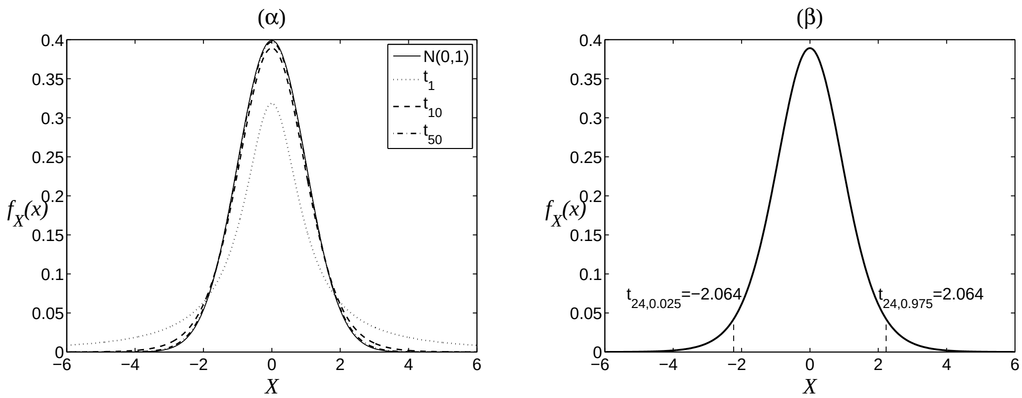
και άρα η τυπική απόκλιση της είναι $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ και ονομάζεται **σταθερό σφάλμα** (standard error) του εκτιμητή $\bar{x}$. Η μορφή της κατανομής της $\bar{x}$ εξαρτάται από το μέγεθος του δείγματος n, από το αν η κατανομή της X είναι κανονική και επίσης από το αν γνωρίζουμε τη διασπορά της.

Αν η κατανομή της τ.μ. X είναι κανονική, $X \sim N(\mu, \sigma^2)$, τότε οι παρατηρήσεις $x_1, \ldots, x_n$ ακολουθούν την ίδια κατανομή και άρα ο μέσος όρος τους $\bar{x}$ ακολουθεί επίσης κανονική κατανομή και ισχύει $\bar{x} \sim N(\mu, \sigma^2/n)$. Από το

ΚΟΘ στην παράγραφο 2.3.4 το ίδιο ισχύει ακόμα και όταν δε γνωρίζουμε την κατανομή της τ.μ. X αλλά το δείγμα είναι μεγάλο ($n > 30$). Στην πράξη όμως δε γνωρίζουμε τη διασπορά σ^2 της τ.μ. X. Αν θεωρήσουμε αντί για σ^2 τη δειγματική διασπορά s^2, η υπόθεση $\bar{x} \sim N(\mu, s^2/n)$ δεν είναι πλέον ακριβής αλλά η $\bar{x}$ ακολουθεί μια άλλη κατανομή, επίσης συμμετρική και με κωνοειδές σχήμα, αλλά με πιο παχιές ουρές από την κανονική. Ειδικότερα η κανονικοποίηση της $\bar{x}$, $(\bar{x} - \mu)/(s/\sqrt{n})$, ακολουθεί την **κατανομή Student** ή t-κατανομή με $n - 1$ βαθμούς ελευθερίας

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}. \tag{3.12}$$

Οι βαθμοί ελευθερίας είναι $n - 1$ γιατί υπάρχουν n ελεύθερες παρατηρήσεις στο δείγμα που δεσμεύονται με μια συνθήκη, να δίνουν την s^2. Στο Σχήμα 3.1α δίνεται το γράφημα της σππ της κατανομής Student για διαφορετικούς βαθμούς ελευθερίας. Για ένα βαθμό ελευθερίας (δείγμα δύο μόνο



Σχήμα 3.1: (α) Η σππ της τυπικής κανονικής κατανομής και της Student για βαθμούς ελευθερίας όπως δίνονται στο ένθετο. (β) Η σππ της Student για 24 βαθμούς ελευθερίας και οι κρίσιμες τιμές για $a = 0.05$.

παρατηρήσεων) η κατανομή Student διαφέρει φανερά από την τυπική κανονική κατανομή και έχει πολύ παχιές ουρές. Αυτή η κατανομή είναι γνωστή με πολλά ονόματα (**Cauchy, Lorentzian, Breit-Wigner**) και χρησιμοποιείται συχνά στη φυσική, όπως στη φυσική υψηλής ενέργειας ως μοντέλο σππ για την ενέργεια που εμφανίζεται συντονισμός. Καθώς οι βαθμοί ελευθερίας πληθαίνουν η κατανομή Student συγκλίνει στην τυπική κανονική κατανομή. Πρακτικά για μεγάλα δείγματα δεν υπάρχει διαφορά μεταξύ της κατανομής Student και της τυπικής κανονικής κατανομής.

Στην Παράγραφο 2.3.3 είχε δειχθεί ότι για την τυπική κανονική κατανομή σε κάθε πιθανότητα, έστω $1 - a$, αντιστοιχεί ένα διάστημα τιμών της z

συμμετρικό ως προς το 0, $[-z_{1-a/2}, z_{1-a/2}]$, έτσι ώστε

$$\mathrm{P}(-z_{1-a/2} < z \leq z_{1-a/2}) = \Phi(z_{1-a/2}) - \Phi(-z_{1-a/2}) = 1 - a,$$

όπου $\Phi(z)$ είναι η ασκ της τυπικής κανονικής κατανομής. Αντίστοιχα για την κατανομή Student με $n-1$ βαθμούς ελευθερίας σε κάθε πιθανότητα $1-a$ αντιστοιχεί ένα διάστημα τιμών της t, $[-t_{n-1,1-a/2}, t_{n-1,1-a/2}]$, έτσι ώστε ισχύει

$$\mathrm{P}(-t_{n-1,1-a/2} < t \leq t_{n-1,1-a/2}) = 1 - a, \tag{3.13}$$

Για παράδειγμα για $a = 0.05$ και 24 βαθμούς ελευθερίας το διάστημα $[-2.064, 2.064]$ περιέχει την τ.μ. t με πιθανότητα 0.95, όπου $t_{24,0.975} = 2.064$, όπως φαίνεται στο Σχήμα 3.1β. Η τιμή που ορίζει την ουρά της κατανομής λέγεται και *κρίσιμη τιμή* (critical value) και παραδοσιακά δίνεται από κατάλληλο στατιστικό πίνακα, αλλά μπορεί να υπολογισθεί εύκολα (π.χ. στο `matlab` η κρίσιμη τιμή για την κατανομή Student δίνεται από τη συνάρτηση `tinv`).

Για να βρούμε το διάστημα που με πιθανότητα $1-a$ περιέχει την παράμετρο μ αντικαθιστούμε το $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ στη σχέση (3.13) και λύνουμε τις ανισότητες ως προς μ

$$\mathrm{P}(\bar{x} - t_{n-1,1-a/2}\frac{s}{\sqrt{n}} < \mu \leq \bar{x} - t_{n-1,1-a/2}\frac{s}{\sqrt{n}}) = 1 - a. \tag{3.14}$$

Η παραπάνω σχέση ορίζει πως το διάστημα

$$\bar{x} \pm t_{n-1,1-a/2}\frac{s}{\sqrt{n}} \quad \text{ή} \quad \left[\bar{x} - t_{n-1,1-a/2}\frac{s}{\sqrt{n}}, \quad \bar{x} + t_{n-1,1-a/2}\frac{s}{\sqrt{n}}\right] \tag{3.15}$$

περιέχει την πραγματική μέση τιμή μ για κάποια δοθείσα πιθανότητα $1-a$ που είναι το προκαθορισμένο **επίπεδο (ή στάθμη) εμπιστοσύνης** (confidence level) και λέγεται **διάστημα εμπιστοσύνης** (confidence interval) της μ σε επίπεδο εμπιστοσύνης $1-a$.

Όταν το δείγμα είναι μεγάλο η κατανομή Student συγκλίνει στην τυπική κανονική κατανομή και άρα στο διάστημα εμπιστοσύνης η κρίσιμη τιμή από την κατανομή Student $t_{n-1,1-a/2}$ μπορεί να αντικατασταθεί από αυτήν της τυπικής κανονικής $z_{1-a/2}$. Αν το δείγμα είναι μικρό και δε μπορούμε να υποθέσουμε πως η κατανομή της τ.μ. X είναι κανονική, π.χ. γιατί φαίνεται από ένα ιστόγραμμα να είναι ισχυρά ασύμμετρη (λοξή), τότε δε μπορούμε να χρησιμοποιήσουμε το παραμετρικό διάστημα εμπιστοσύνης στην (3.15) για τη μέση τιμή μ. Σε αυτήν την περίπτωση θα πρέπει να καταφύγουμε στη μη-παραμετρική προσέγγιση. Η πιο συχνή μη-παραμετρική μέθοδος χρησιμοποιεί τις *τάξεις* (ranks) των δεδομένων, δηλαδή τη σειρά τους όταν αυτά κατατάσσονται σε αύξουσα σειρά.

Παράδειγμα 3.3. Θέλουμε να μελετήσουμε κατά πόσο ασφάλειες ενός τύπου καίγονται σε ένταση ρεύματος 40 αμπέρ όπως είναι η ένδειξη τους. Στον Πίνακα 3.1 δίνονται οι μετρήσεις έντασης του ηλεκτρικού ρεύματος στις οποίες κάηκαν 25 ασφάλειες που δοκιμάσαμε. Η τ.μ. X που μας ενδιαφέρει

40.9	40.3	39.8	40.1	39.0	41.4	39.8	41.5	40.0	40.6
38.3	39.0	40.9	39.1	40.3	39.3	39.6	38.4	38.4	40.7
39.7	38.9	38.9	40.6	39.6					

Πίνακας 3.1: Δεδομένα ορίου έντασης ηλεκτρικού ρεύματος που κάηκαν 25 ασφάλειες των 40 αμπέρ.

είναι το όριο έντασης ηλεκτρικού ρεύματος που καίγονται ασφάλειας των 40 αμπέρ. Η δειγματική μέση τιμή της είναι

$$\bar{x} = \frac{1}{25}\sum_{i=1}^{25} x_i = \frac{1}{25}995.1 = 39.80$$

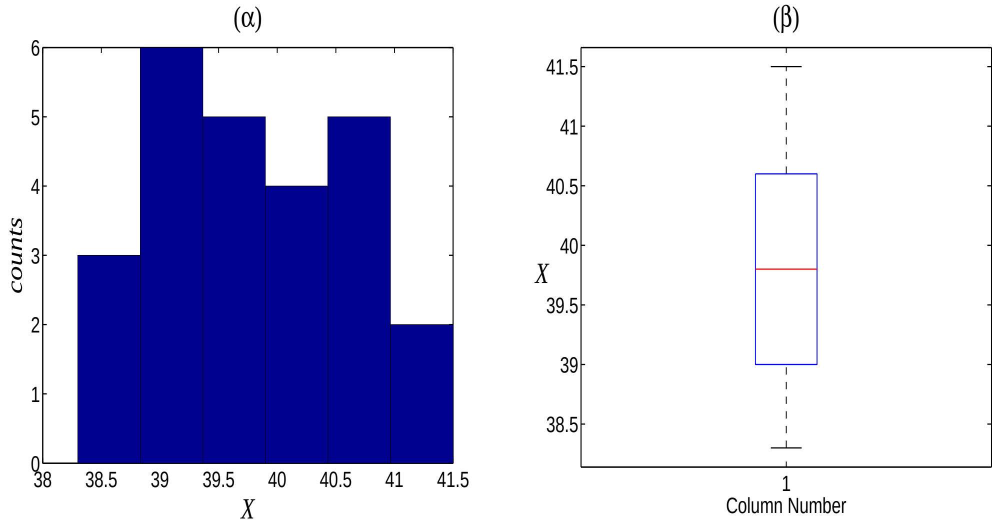
και η δειγματική διασπορά της είναι

$$s^2 = \frac{1}{24}\left(\sum_{i=1}^{25} x_i^2 - 25\bar{x}^2\right) = \frac{1}{24}(39629 - 25\cdot 39.80^2) = 0.854.$$

Με βάση αυτό το δείγμα η εκτίμηση της μέση τιμής μ είναι $\bar{x} = 39.80$ αμπέρ και της διασποράς σ^2 είναι $s^2 = 0.854\,(\text{αμπέρ})^2$.

Έστω τώρα ότι θέλουμε να εκτιμήσουμε διάστημα εμπιστοσύνης σε επίπεδο 95% για το μέσο όριο έντασης ηλεκτρικού ρεύματος για ασφάλειες των 40 αμπέρ. Το δείγμα είναι μικρό ($n = 25 < 30$). Εξετάζουμε τη δειγματική κατανομή του ορίου έντασης ηλεκτρικού ρεύματος από τα δεδομένα μας. Για αυτό σχεδιάζουμε το ιστόγραμμα και το θηκόγραμμα των δεδομένων του Πίνακα 3.1, τα οποία παρουσιάζονται στο Σχήμα 3.2.

Το **θηκόγραμμα** (boxplot) είναι η απεικόνιση των τεταρτομορίων των δεδομένων, και δίνεται από τους παρακάτω 5 αριθμούς: τη μικρότερη παρατήρηση, την παρατήρηση με τάξη στο 25% των παρατηρήσεων (πρώτο τεταρτομόριο), τη διάμεσο, την παρατήρηση με τάξη στο 75% των παρατηρήσεων (τρίτο τεταρτομόριο) και τη μεγαλύτερη τιμή. Στο θηκόγραμμα, η θήκη (κουτί) διαγράφεται μεταξύ του πρώτου και τρίτου τεταρτομορίου και οι μύστακες ενώνουν τα τεταρτομόρια με τα άκρα (ή διακόπτεται η γραμμή και οι ακραίες τιμές δηλώνονται με κάποιο σύμβολο όταν είναι απόμακρες). Τέλος σχηματίζεται μια γράμμη στη θήκη στη θέση της διαμέσου. Από το θηκόγραμμα μπορούμε να κρίνουμε τη συμμετρία της κατανομής που αναφέρονται τα δεδομένα (αν η γραμμή της διαμέσου βρίσκεται προς το κέντρο της θήκης και



Σχήμα 3.2: Ιστόγραμμα στο (α) και θηκόγραμμα στο (β) των δεδομένων του ορίου έντασης ηλεκτρικού ρεύματος για ασφάλειες των 40 αμπέρ του Πίνακα 3.1.

οι μύστακες έχουν περίπου στο ίδιο μήκος) και κατ' επέκταση αν η κατανομή είναι κανονική.

Από το ιστόγραμμα και το θηκόγραμμα βλέπουμε ότι η κατανομή του ορίου έντασης ηλεκτρικού ρεύματος φαίνεται να είναι κανονική (είναι συμμετρική και δεν έχει μακριές ουρές). Άρα μπορούμε να υποθέσουμε ότι το όριο έντασης ηλεκτρικού ρεύματος X ακολουθεί κανονική κατανομή και μπορούμε να χρησιμοποιήσουμε το διάστημα εμπιστοσύνης από την κατανομή Student που δίνεται στην (3.15). Βρίσκουμε την κρίσιμη τιμή $t_{n-1,1-\alpha/2}$ για $1-\alpha/2 = 0.975$ και $n-1=24$, $t_{24,0.975} = 2.064$ (από το στατιστικό πίνακα για την κατανομή Student ή υπολογίζοντας απευθείας την αντίστροφη ασκ σε κάποιο πρόγραμμα όπως το `matlab`, δες επίσης Σχήμα 3.1β). Το διάστημα εμπιστοσύνης για τη μ είναι

$$39.80 \pm 2.064\frac{\sqrt{0.854}}{5} \quad \rightarrow \quad [39.42, 40.18].$$

Με βάση το παραπάνω 95% διάστημα εμπιστοσύνης μπορούμε να πούμε ότι η σημειακή εκτίμηση $\bar{x} = 39.80$ είναι αρκετά ακριβής αφού το αντίστοιχο 95% διάστημα εμπιστοσύνης είναι αρκετά μικρό. Επίσης παρατηρούμε ότι το 95% διάστημα εμπιστοσύνης περιέχει το 40, δηλαδή με 95% εμπιστοσύνη μπορούμε να συμπεράνουμε ότι κατά 'μέσο όρο' οι ασφάλειες διασφαλίζουν την ένδειξη τους και καίγονται πράγματι σε ένταση ηλεκτρικού ρεύματος 40 αμπέρ.

3.2.2 Διάστημα εμπιστοσύνης της διασποράς σ^2

Για να ορίσουμε διάστημα εμπιστοσύνης για τη μέση τιμή μ κανονικοποιήσαμε τον εκτιμητή $\bar{x}$ ως $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ ώστε να ακολουθεί κατανομή Student. Το ίδιο θα κάνουμε και εδώ για τον εκτιμητή s^2 της διασποράς σ^2. Η κανονικοποίηση είναι $(n-1)s^2/\sigma^2$ που ακολουθεί την κατανομή $\mathcal{X}^2$ με $n-1$ βαθμούς ελευθερίας. Για να δείξουμε αυτό το αποτέλεσμα ας δούμε πρώτα πως προκύπτει η κατανομή αυτή.

Η κατανομή $\mathcal{X}^2$ χρησιμοποιείται σε πολλά προβλήματα στατιστικής συμπερασματολογίας. Έστω ότι η τ.μ. χ^2 είναι το άθροισμα τετραγώνων των διαφορών μεταξύ παρατηρούμενων και προσδοκώμενων ή μέσων τιμών διαιρούμενο με τη διασπορά τους. Είναι φανερό από τον ορισμό της ότι η χ^2 εξαρτάται από το πλήθος των παρατηρήσεων n. Ειδικότερα όταν οι παρατηρήσεις προέρχονται από την ίδια κατανομή η χ^2 ορίζεται ως

$$\chi^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\sigma^2} \tag{3.16}$$

και ακολουθεί κατανομή $\mathcal{X}^2$ με $n-1$ βαθμούς ελευθερίας (ένας βαθμός ελευθερίας χάνεται λόγω της συνθήκης της δειγματικής μέσης τιμής $\bar{x}$). Συνδυάζοντας της σχέση (3.16) με τη σχέση (3.3) που ορίζει τη δειγματική διασπορά s^2 έχουμε το ζητούμενο αποτέλεσμα

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \sim \mathcal{X}^2_{n-1}. \tag{3.17}$$

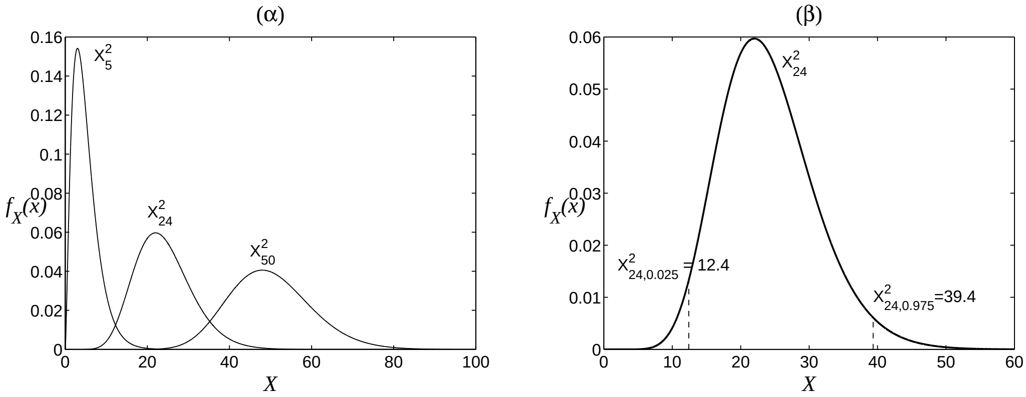
Στο Σχήμα 3.3α δίνεται το γράφημα της σππ της κατανομής $\mathcal{X}^2$ για διαφορετικούς βαθμούς ελευθερίας. Για λίγους βαθμούς ελευθερίας η κατανομή $\mathcal{X}^2$ είναι φανερά ασύμμετρη αλλά καθώς οι βαθμοί ελευθερίας πληθαίνουν γίνεται πιο συμμετρική και κωνοειδής. Για πολλούς βαθμούς ελευθερίας προσεγγίζει την κανονική κατανομή. Γενικά επειδή η κατανομή $\mathcal{X}^2$ δεν είναι συμμετρική γύρω από το 0, οι ουρές της ορίζονται με δύο κρίσιμες τιμές, την αριστερή κρίσιμη τιμή $\chi^2_{n-1,a/2}$ και τη δεξιά κρίσιμη τιμή $\chi^2_{n-1,1-a/2}$ για κάποιο επίπεδο σημαντικότητας a. Στο Σχήμα 3.3β φαίνονται οι ουρές για $n = 25$ και $a = 0.05$.

Η πιθανότητα $1-a$ η τ.μ. χ^2 να βρίσκεται μεταξύ των ορίων που δίνονται από την αριστερή και δεξιά κρίσιμη τιμή είναι

$$\mathrm{P}(\chi^2_{n-1,a/2} < \chi^2 < \chi^2_{n-1,1-a/2}) = 1-a. \tag{3.18}$$

Αντικαθιστώντας $\chi^2 = (n-1)s^2/\sigma^2$ και λύνοντας τις δύο ανισότητες ως προς σ^2 στη σχέση (3.18) βρίσκουμε το $(1-a)\%$ διάστημα εμπιστοσύνης για τη σ^2

$$\left[\frac{(n-1)s^2}{\chi^2_{n-1,1-a/2}}, \frac{(n-1)s^2}{\chi^2_{n-1,a/2}}\right]. \tag{3.19}$$



Σχήμα 3.3: (α) Η σππ της κατανομής $\mathcal{X}^2$ για βαθμούς ελευθερίας όπως δίνονται στο σχήμα. (β) Η σππ της κατανομής $\mathcal{X}^2$ για 24 βαθμούς ελευθερίας και οι κρίσιμες τιμές για $\alpha = 0.05$.

Το 95% δ.ε. για την τυπική απόκλιση σ έχει ως άκρα τις τετραγωνικές ρίζες των αντίστοιχων άκρων του 95% δ.ε. για τη διασπορά σ^2.

Παράδειγμα 3.4. Από τα δεδομένα για το όριο έντασης ηλεκτρικού ρεύματος ασφάλειας που χρησιμοποιήθηκαν στο Παράδειγμα 3.3 θέλουμε να εκτιμήσουμε τη διασπορά σ^2 του ορίου έντασης. Η σημειακή εκτίμηση βρέθηκε να είναι $s^2 = 0.854\,(\text{αμπέρ})^2$. Για $n-1 = 24$ και $\alpha = 0.05$ από τον στατιστικό πίνακα για τη $\mathcal{X}^2$ (ή με απευθείας υπολογισμό στο `matlab` με τη συνάρτηση `chi2inv`) βρίσκουμε $\chi^2_{24,\,0.025} = 12.4$ και $\chi^2_{24,\,0.975} = 39.4$ (δες επίσης Σχήμα 3.3β). Το 95% δ.ε. για τη διασπορά σ^2 είναι

$$\left[\frac{24\cdot 0.854}{39.4},\ \frac{24\cdot 0.854}{12.4}\right] = [0.52,\ 1.65].$$

Το 95% δ.ε. για την τυπική απόκλιση σ του ορίου ηλεκτρικού ρεύματος είναι

$$[\sqrt{0.52},\ \sqrt{1.65}] = [0.72,\ 1.28],$$

που δείχνει ότι οι ασφάλειες δεν καίγονται με μεγάλη ακρίβεια γύρω από το όριο των 40 αμπέρ αλλά με μια τυπική απόκλιση που περιμένουμε να κυμαίνεται μεταξύ 0.7 και 1.3 αμπέρ. Αυτό το συμπέρασμα φαίνεται να είναι σε αντίθεση με το συμπέρασμα για την καλή ακρίβεια εκτίμησης του μέσου ορίου ηλεκτρικού ρεύματος μ που βρήκαμε στο Παράδειγμα 3.3. Εδώ θα πρέπει να τονισθεί ότι η κατανομή του εκτιμητή $\bar{x}$ της μέσης τιμής είναι πολύ πιο 'στενή' από την κατανομή της ίδιας της τ.μ. X και μάλιστα η διασπορά της όπως εκτιμάται από το δείγμα είναι s^2/n δηλαδή η τυπική της απόκλιση (το τυπικό σφάλμα) είναι $\sqrt{n}$ φορές μικρότερη από την τυπική απόκλιση της X. Μια τυπική μέτρηση του ορίου έντασης ηλεκτρικού ρεύματος που καίγεται η ασφάλεια των 40 αμπέρ με βάση το δείγμα των 25 παρατηρήσεων, θα περιμέναμε να κυμαίνεται στο διάστημα $\bar{x} \pm s = 39.80 \pm \sqrt{0.854} = 39.80 \pm 0.925$. Επίσης το 95 % των μετρήσεων του ορίου ηλεκτρικού ρεύματος θα περιμέναμε να βρίσκεται στο διάστημα

$$\bar{x} \pm t_{n-1,1-a/2} s. \tag{3.20}$$

δηλαδή $39.80 \pm 2.064 \cdot 0.925 = 39.80 \pm 1.763$, που δηλώνει ότι θα εμφανίζονται ασφάλειες με σοβαρές αποκλίσεις από την ονομαστική ένδειξη της ασφάλειας.

3.3 Έλεγχος υπόθεσης

Σε πολλά προβλήματα δεν ενδιαφερόμαστε να εκτιμήσουμε με κάποια ακρίβεια την τιμή της παραμέτρου αλλά να διαπιστώσουμε αν η παραμέτρος είναι μικρότερη ή μεγαλύτερη από μια δεδομένη τιμή που έχει φυσική σημασία για το πρόβλημα μας. Στο προηγούμενο παράδειγμα μπορεί να μας ενδιαφέρει να ελέγξουμε αν το μέσο όριο έντασης ηλεκτρικού ρεύματος που καίγονται οι ασφάλειες των 40 αμπέρ διαφέρει σημαντικά από 40 αμπέρ. Σε όργανα μέτρησης μας ενδιαφέρει να ελέγξουμε αν η μέτρηση είναι σωστή και δεν υπάρχει συστηματικό λάθος στο όργανο μέτρησης. Βέβαια την απάντηση σε τέτοια ερωτήματα μπορεί να τη δώσει η εκτίμηση κατάλληλου διαστήματος εμπιστοσύνης ελέγχοντας αν η δεδομένη τιμή ανήκει σ' αυτό το διάστημα ή όχι, αλλά εδώ θα δούμε μια διαφορετική προσέγγιση, θέτοντας κατάλληλη στατιστική υπόθεση και ελέγχοντας αν είναι αποδεκτή ή όχι.

Ο **έλεγχος υπόθεσης** (hypothesis testing) επεξεργάζεται στατιστικά εργαλεία (τον εκτιμητή και την κατανομή του) σε μια διαδικασία λήψης απόφασης. Για τη διαδικασία ελέγχου μιας στατιστικής υπόθεσης πρώτα ορίζουμε τη στατιστική υπόθεση, μετά υπολογίζουμε το στατιστικό ελέγχου και την περιοχή απόρριψης και τέλος αποφασίζουμε για την υπόθεση με βάση την ένδειξη που έχουμε από το δείγμα.

Η *στατιστική υπόθεση* (statistical hypothesis) μπορεί να είναι μια οποιαδήποτε 'στατιστική' δήλωση ή πρόταση που θέτουμε υπό έλεγχο με βάση τις παρατηρήσεις. Στην αρχή θα μελετήσουμε υποθέσεις για την τιμή μιας παραμέτρου και στη συνέχεια για την κατανομή μιας τ.μ.. Η **μηδενική υπόθεση** (null hypothesis) την οποία θέτουμε υπό έλεγχο συμβολίζεται H_0 ενώ η **εναλλακτική υπόθεση** (alternative hypothesis) την οποία δεχόμαστε αν απορρίψουμε τη H_0 συμβολίζεται H_1. Οι δυνατές αποφάσεις του ελέγχου είναι:

1. *Σωστή απόφαση*: Αποδεχόμαστε την H_0 όταν η H_0 είναι σωστή. Η πιθανότητα αυτής της απόφασης είναι

$$\mathrm{P}(\text{αποδοχή της } H_0 \mid H_0 \text{ σωστή}) = 1 - a.$$

2. **Σφάλμα τύπου II** (type II error): Αποδεχόμαστε την H_0 όταν η H_0 είναι λανθασμένη. Η πιθανότητα αυτού του σφάλματος είναι

$$\mathrm{P}(\text{αποδοχή της } H_0 \mid H_0 \text{ λανθασμένη}) = \beta.$$

3. **Σφάλμα τύπου I** (type I error): Απορρίπτουμε την H_0 όταν η H_0 είναι σωστή. Η πιθανότητα αυτού του σφάλματος είναι το επίπεδο σημαντικότητας

$$\mathrm{P}(\text{απόρριψη της } H_0 \mid H_0 \text{ σωστή}) = a.$$

4. *Σωστή απόφαση*: Απορρίπτουμε την H_0 και η H_0 είναι λανθασμένη. Η πιθανότητα αυτής της απόφασης είναι

$$\mathrm{P}(\text{απόρριψη της } H_0 \mid H_0 \text{ λανθασμένη}) = 1 - \beta$$

   και δηλώνει την **ισχύ του ελέγχου** (power of the test).

Οι 4 δυνατές περιπτώσεις στην αποφάση του ελέγχου δίνονται στον Πίνακα 3.2. Για να είναι ένας έλεγχος ακριβής θα πρέπει το πραγματικό σφάλμα

	Αποδοχή της H_0	Απόρριψη της H_0
H_0 σωστή	ορθή απόφαση $(1 - a)$	σφάλμα τύπου I (a)
H_0 λανθασμένη	σφάλμα τύπου II (β)	ορθή απόφαση $(1 - \beta)$

Πίνακας 3.2: Οι 4 περιπτώσεις στην απόφαση ελέγχου με την αντίστοιχη πιθανότητα σε παρένθεση.

τύπου I να είναι στο επίπεδο σημαντικότητας a στο οποίο γίνεται ο έλεγχος. Στην πράξη αυτό δεν είναι βέβαια εφικτό αλλά μπορούμε να το διαπιστώσουμε αν έχουμε τη δυνατότητα να κάνουμε προσομοιώσεις. Για να υπολογίσουμε το πραγματικό σφάλμα τύπου I θα πρέπει να γνωρίζουμε ότι η H_0 είναι σωστή και να επαναλάβουμε τον έλεγχο σε M διαφορετικά δείγματα ίδιου τύπου και στο ίδιο επίπεδο σημαντικότητας a. Αν η H_0 απορρίπτεται m φορές σε επίπεδο σημαντικότητας a θα πρέπει για να είναι ο έλεγχος ακριβής (να έχει σωστή σημαντικότητα) η αναλογία m/M να είναι κοντά στο a.

Επίσης μας ενδιαφέρει ο έλεγχος να έχει μεγάλη ισχύ. Την ισχύ του ελέγχου μπορούμε υπολογιστικά να τη μετρήσουμε και πάλι με προσομοιώσεις όπου τώρα θα πρέπει να γνωρίζουμε ότι η H_0 δεν είναι σωστή και επιπλέον ότι ο έλεγχος έχει σωστή σημαντικότητα (σύμφωνα με τα παραπάνω).

Οι παραπάνω προσομοιώσεις συνήθως ακολουθούνται για να αξιολογήσουμε το στατιστικό που χρησιμοποιείται στον έλεγχο. Στη συνέχεια θα παρουσιάσουμε τη διαδικασία του παραμετρικού ελέγχου για τη μέση τιμή.

3.3.1 Έλεγχος μέσης τιμής

Θέλουμε να ελέγξουμε με βάση ένα δείγμα παρατηρήσεων $\{x_1, \ldots, x_n\}$ μιας τ.μ. X αν η μέση τιμή μ της X μπορεί να πάρει κάποια τιμή μ_0, δηλαδή η μηδενική υπόθεση είναι $H_0 : \mu = \mu_0$. Φυσικά όταν υποθέτουμε $\mu = \mu_0$ δεν εννοούμε αυστηρά την ισότητα και θα θέλαμε ο έλεγχος να αποφασίζει ότι η H_0 είναι ορθή όταν η μ βρίσκεται 'κοντά' στην τιμή μ_0 και λανθασμένη αλλιώς. Έτσι η τυχόν απόρριψη της H_0 δεν ερμηνεύεται ως αποδοχή της πρότασης $\mu = \mu_0$ αλλά μη-απόρριψη της, πάντα με βάση το δείγμα.

Ο κατάλληλος εκτιμητής της μ είναι η δειγματική μέση τιμή $\bar{x}$ που υπολογίζεται από το δείγμα. Σύμφωνα και με τα παραπάνω, τιμές της $\bar{x}$ 'κοντά' στη μ_0 υποστηρίζουν την ορθότητα της H_0 και σχηματίζουν την περιοχής αποδοχής της H_0, ενώ τιμές της $\bar{x}$ 'μακριά' από τη μ_0 δεν την υποστηρίζουν και σχηματίζουν την *περιοχή απόρριψης* (rejection region) που συμβολίζουμε R.

Η απόφαση για την αποδοχή ή απόρριψη της H_0 γίνεται με βάση τις πιθανότητες και όπως για τα διαστήματα εμπιστοσύνης έτσι και εδώ ορίζουμε επίπεδο σημαντικότητας α (η επίπεδο εμπιστοσύνης $1 - \alpha$) για την απόφαση ελέγχου. Το α καθορίζει το 'κοντά' και 'μακριά' που αναφέραμε παραπάνω. Στην παραμετρική προσέγγιση που ακολουθούμε εδώ το α ορίζει το εύρος των ουρών της κατανομής του εκτιμητή $\bar{x}$.

Είχαμε δείξει πως η κανονικοποίηση του εκτιμητή $\bar{x}$, $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$, ακολουθεί κατανομή Student θεωρώντας ότι η τ.μ. X ακολουθεί κανονική κατανομή ή ότι το δείγμα είναι μεγάλο. Θεωρώντας ότι ισχύει η H_0, έχουμε $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} \sim t_{n-1}$, όπου t είναι το **στατιστικό ελέγχου** (test statistic). Οι κρίσιμες τιμές του t για δεδομένο α δίνουν τα όρια του R. Τιμές του t που είναι στις ουρές της κατανομής t_{n-1} ανήκουν στο R, δηλαδή δεν είναι πιθανές όταν ισχύει η H_0 και άρα η εμφάνιση τους συνιστά απόρριψη της H_0.

Υπολογίζουμε την τιμή του στατιστικού t από το δείγμα, έστω $\tilde{t}$. Αν το $\tilde{t}$ ανήκει στην περιοχή απόρριψης R που ορίσαμε για κάποιο επίπεδο σημαντικότητας α απορρίπτουμε την H_0. Σημειώνεται ότι αυτή η απόρριψη ισχύει για το α που επιλέχτηκε και μπορεί η απόφαση του ελέγχου να αλλάξει για μικρότερο α. Αντίστροφα αν δεν απορρίπτουμε την H_0 για κάποιο α, μπορεί να την απορρίψουμε για μεγαλύτερο α. Η μικρότερη τιμή του α που δίνει απόρριψη της H_0 λέγεται ***p*-τιμή** (*p*-value) και είναι η πιθανότητα να παρατηρήσουμε για το t μια τιμή τόσο ακραία όσο το $\tilde{t}$ όταν ισχύει η H_0. Άρα

p-τιμή είναι η πιθανότητα το t να είναι στο R που ορίζεται με κρίσιμη τιμή το $\tilde{t}$, δηλαδή

$$p = 2\,\mathrm{P}(t > |\tilde{t}|) = 2\,(1 - \mathrm{P}(t < |\tilde{t}|)). \tag{3.21}$$

Όσο πιο κοντά στο 0 είναι η p-τιμή τόσο πιο σίγουρη είναι η απόρριψη της H_0. Τιμές $p > 0.05$ δηλώνουν πως η H_0 δε μπορεί να απορριφθεί. Η p-τιμή υπολογίζεται εύκολα από την ασκ της κατανομής t_{n-1} (στο `matlab` δίνεται από τη συνάρτηση `tcdf`), αλλά παλιότερα που τέτοιοι υπολογισμοί δεν ήταν εύκολα εφικτοί δε χρησιμοποιούταν η p-τιμή και ο έλεγχος γινόταν σε κάποιο επίπεδο σημαντικότητας α (με απάντηση τύπου ναι / όχι).

Σε κάποιες περιπτώσεις ο έλεγχος μπορεί να είναι **μονόπλευρος** (one-sided), δηλαδή η απορριπτική περιοχή που ενισχύει την H_1 σχηματίζεται μόνο από τη μια ουρά της κατανομής γιατί θεωρούμε πως είναι αδύνατον για το πρόβλημα μας η μ να παίρνει τιμές στην άλλη ουρά της κατανομής του $\bar{x}$. Η επιλογή μονόπλευρου ή δίπλευρου ελέγχου εξαρτάται από την έρευνα που θέλουμε να κάνουμε και από το κατά πόσο μπορούμε να προβλέψουμε το αποτέλεσμα της έρευνας. Για παράδειγμα, έστω ότι θέλουμε να ελέγξουμε αν η μέση απόδοση μ ενός μηχανήματος που σχεδιάσαμε μπορεί να φθάσει την απόδοση αναφοράς μ_0 του 'άριστου' μηχανήματος (μηχάνημα αναφοράς). Σε αυτήν την περίπτωση ο έλεγχος πρέπει να είναι μονόπλευρος ($\mathrm{H}_0 : \mu = \mu_0$ ή ισοδύναμα $\mathrm{H}_0 : \mu \geq \mu_0$ και $\mathrm{H}_1 : \mu < \mu_0$) γιατί *γνωρίζουμε* πως δε μπορεί $\mu > \mu_0$ αφού η απόδοση του νέου μηχανήματος δε μπορεί να ξεπεράσει αυτή του μηχανήματος αναφοράς.

Παράδειγμα 3.5. Με αναφορά στο Παράδειγμα 3.3 έστω ότι θέλουμε να ελέγξουμε αν το μέσο όριο του ηλεκτρικού ρεύματος που καίγονται οι ασφάλειες των 40 αμπέρ είναι πράγματι 40. Οι υποθέσεις του ελέγχου είναι $\mathrm{H}_0 : \mu = 40$ και $\mathrm{H}_1 : \mu \neq 40$.

Είχαμε δεχθεί πως με βάση το δείγμα των 25 ασφαλειών η κατανομή του οριού του ηλεκτρικού ρεύματος που καίγονται οι ασφάλειες των 40 αμπέρ μπορεί να είναι κανονική. Χρησιμοποιούμε λοιπόν ως στατιστικό ελέγχου το $t = \frac{\bar{x}-40}{s/\sqrt{n}}$ που ακολουθεί την t_{n-1} σύμφωνα με την H_0. Για επίπεδο σημαντικότητας $\alpha = 0.05$ η απορριπτική περιοχή ορίζεται από την κρίσιμη τιμή $t_{n-1,1-\alpha/2}$ για $1-\alpha/2 = 0.975$ και $n-1 = 24$, $t_{24,0.975} = 2.064$. Η απορριπτική περιοχή είναι

$$R = \{t \mid t < -2.064 \vee t > 2.064\} = \{t \mid |t| > 2.064\}.$$

Η τιμή του στατιστικού από το δείγμα είναι ($\bar{x} = 39.8$, $s = 0.925$)

$$\tilde{t} = \frac{39.8 - 40}{0.925/5} = -1.081$$

που προφανώς δεν ανήκει στην απορριπτική περιοχή και άρα δε μπορούμε να απορρίψουμε την H_0, ότι οι ασφάλειες καίγονται στο όριο των 40 αμπέρ. Από την ασκ της t_{24}-κατανομής για $\tilde{t} = -1.081$ βρίσκουμε την p-τιμή

$$p = 2(1 - \mathrm{P}(t \leq |\tilde{t}|)) = 2(1 - \mathrm{P}(t \leq 1.081)) = 2(1 - 0.855) = 0.29$$

που δηλώνει ποσοστό εμπιστοσύνης της απόρριψης της H_0 σε επίπεδο περίπου 70%. Πρακτικά αυτό σημαίνει πως δε μπορούμε να απορρίψουμε την H_0.

Αν για κάποιο λόγο αποκλείουμε ότι οι ασφάλειες μπορούν να καίγονται σε υψηλότερο όριο από 40 αμπέρ, τότε ο έλεγχος γίνεται μονόπλευρος, $H_0 : \mu \geq 40$ και $H_1 : \mu < 40$. Η απορριπτική περιοχή για $a = 0.05$ τότε είναι

$$R = \{t \mid t < t_{n-1,a/2}\} = \{t \mid t < t_{24,0.05}\} = \{t \mid t < -1.71\}.$$

Και πάλι όμως η H_0 δεν απορρίπτεται αφού $\tilde{t} \notin R$. Η p-τιμή για το μονόπλευρο έλεγχο είναι

$$p = \mathrm{P}(t \leq \tilde{t}) = 0.145$$

που και πάλι είναι αρκετά υψηλό και δηλώνει πως η H_0 δε μπορεί να απορριφθεί.

3.3.2 Έλεγχος διασποράς

Η στατιστική υπόθεση για τη διασπορά σ^2 είναι όπως και για τη μέση τιμή, δηλαδή $H_0 : \sigma^2 = \sigma_0^2$ με κατάλληλη εναλλακτική υπόθεση H_1 ανάλογα αν ο έλεγχος είναι δίπλευρος ή μονόπλευρος. Το στατιστικό ελέγχου είναι το χ^2 που χρησιμοποιήθηκε στο διάστημα εμπιστοσύνης της σ^2 (δες Παράγραφο 3.2.2)

$$\chi^2 \equiv \frac{(n-1)s^2}{\sigma_0^2} \sim \mathcal{X}_{n-1}^2, \tag{3.22}$$

όπου s^2 είναι ο εκτιμητής της διασποράς. Από την κατανομή $\mathcal{X}^2$ με $n-1$ βαθμούς ελευθερίας και για επίπεδο σημαντικότητας a βρίσκουμε τις κρίσιμες τιμές και η περιοχή απόρριψης R δίνεται για τον κάθε τύπο ελέγχου ως

1. $H_1 : \sigma^2 \neq \sigma_0^2, \quad R = \{\chi^2 | \, \chi^2 < \chi_{n-1,a/2}^2 \; \vee \; \chi^2 > \chi_{n-1,1-a/2}^2\}.$
2. $H_1 : \sigma^2 < \sigma_0^2, \quad R = \{\chi^2 | \, \chi^2 < \chi_{n-1,a}^2\}.$
3. $H_1 : \sigma^2 > \sigma_0^2, \quad R = \{\chi^2 | \, \chi^2 > \chi_{n-1,1-a}^2\}.$

Το στατιστικό ελέγχου από το δείγμα $\tilde{\chi}^2$ υπολογίζεται θέτοντας στη σχέση (3.22) την εκτίμηση s^2 από το δείγμα. Αν $\tilde{\chi}^2 \in R$ η H_0 απορρίπτεται. Η p-τιμή για τα τρία είδη ελέγχου είναι

1. $\mathrm{H}_1 : \sigma^2 \neq \sigma_0^2$, $\quad p = \mathrm{P}(\chi^2 < \tilde{\chi}^2 \ \vee \ \chi^2 > \tilde{\chi}^2\}$.
2. $\mathrm{H}_1 : \sigma^2 < \sigma_0^2$, $\quad p = \mathrm{P}(\chi^2 < \tilde{\chi}^2\}$.
3. $\mathrm{H}_1 : \sigma^2 > \sigma_0^2$, $\quad p = \mathrm{P}(\chi^2 > \tilde{\chi}^2\}$.

Παράδειγμα 3.6. Θέλουμε να ελέγξουμε σε επίπεδο εμπιστοσύνης 99% αν η τυπική απόκλιση σ του όριου έντασης ηλεκτρικού ρεύματος των ασφαλειών 40 αμπέρ μπορεί να είναι 0.7 αμπέρ. Χρησιμοποιούμε τα 25 δεδομένα για το όριο έντασης ηλεκτρικού ρεύματος που χρησιμοποιήθηκαν στο Παράδειγμα 3.3.

Εφαρμόζουμε δίπλευρο έλεγχο για τη διασπορά: $\mathrm{H}_0 : \sigma^2 = 0.49$, $\mathrm{H}_1 : \sigma^2 \neq 0.49$. Το μέγεθος του δείγματος είναι $n = 25$ και η δειγματική διασπορά είναι $s^2 = 0.854\,(\text{αμπέρ})^2$. Η κρίσιμη τιμή του στατιστικού ελέγχου χ^2 για $a = 0.01$ είναι $\chi^2_{24,0.005} = 9.886$ και $\chi^2_{24,0.995} = 45.558$. Η περιοχή απόρριψης είναι

$$R = \{\chi^2 \mid \chi^2 < 9.886 \vee \chi^2 > 45.558\}.$$

Η στατιστική ελέγχου που παίρνουμε από το δείγμα είναι

$$\tilde{\chi}^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{24 \cdot 0.854}{0.49} = 41.829.$$

Η $\tilde{\chi}^2$ δεν ανήκει στην περιοχή απόρριψης και άρα δε μπορούμε να απορρίψουμε την H_0 στο επίπεδο σημαντικότητας $a = 0.01$. Φαίνεται όμως να μπορούμε να απορρίψουμε την H_0 για λίγο υψηλότερο a. Η τιμή αυτή είναι

$$\begin{aligned} p &= \mathrm{P}(\chi^2 < \tilde{\chi}^2 \ \vee \ \chi^2 > \tilde{\chi}^2\} = \mathrm{P}(\chi^2 > 41.829\} \\ &= 1 - \mathrm{P}(\chi^2 < 41.829\} = 1 - 0.986 = 0.014. \end{aligned}$$

και άρα μπορούμε να απορρίψουμε την H_0 σε επίπεδο σημαντικότητας ως και 0.014 (αλλά όχι 0.01). Στην παραπάνω σχέση της πιθανότητας επιλέγουμε μια από τις δύο ανισότητες (για την αριστερή ή δεξιά κρίσιμη τιμή που δίνεται από το $\tilde{\chi}^2$) για να υπολογίσουμε την p-τιμή.

Γενικά όταν θέλουμε να κάνουμε στατιστικό έλεγχο για την τυπική απόκλιση σ υψώνουμε στο τετράγωνο την τιμή της τυπικής απόκλισης που θέλουμε να ελέγξουμε και κάνουμε έλεγχο διασποράς γι αυτήν την τιμή.

3.3.3 Έλεγχος καταλληλότητας $\mathcal{X}^2$

Μια άλλη περίπτωση που χρησιμοποιείται η κατανομή $\mathcal{X}^2$ είναι ο **έλεγχος καλής προσαρμογής** (goodness-of-fit test) γνωστής κατανομής στα δεδομένα, δηλαδή για να ελέγξουμε αν το δείγμα προέρχεται από πληθυσμό με κάποια γνωστή κατανομή.

Για τον έλεγχο $\mathcal{X}^2$ η τ.μ. X πρέπει να είναι διακριτή, ή αν είναι συνεχής να μετατραπεί σε διακριτή (δες Παράγραφο 2.1.1). Πρακτικά αυτό σημαίνει πως αν τα δεδομένα είναι αριθμητικά θα πρέπει να ορισθεί πρώτα διαμέριση των τιμών και να ομαδοποιηθούν. Έστω οι K παρατηρούμενες διακριτές τιμές O_j, $j = 1, \ldots, K$, και οι αντίστοιχες αναμενόμενες τιμές E_j από τη γνωστή κατανομή. Στην περίπτωση της διακριτικοποίησης, O_j είναι η συχνότητα εμφάνισης παρατηρήσεων σε κάθε ομάδα (διάστημα) j από τις K ομάδες της διαμέρισης και E_j η αντίστοιχη αναμενόμενη συχνότητα. Το στατιστικό ελέγχου είναι

$$\chi^2 = \sum_{j=1}^{K} \frac{(O_j - E_j)^2}{E_j}. \tag{3.23}$$

Όταν η παρατηρούμενη τ.μ. X είναι διακριτή τότε οι αναμένομενες τιμές E_j δίνονται από το γινόμενο του πλήθους των δεδομένων n με την αντίστοιχη πιθανότητα της διακριτής κατανομής $f_X(x_j) = \mathrm{P}(X = x_j)$. Όταν η X είναι συνεχής οι αναμένομενες τιμές E_j υπολογίζονται ως

$$E_j = n(F_X(x_j^u) - F_X(x_j^l)), \tag{3.24}$$

όπου $F_X(x)$ είναι η ασκ της X για την τιμή x και x_j^l και x_j^u είναι το κάτω και πάνω άκρο του διαστήματος j, αντίστοιχα.

Οι βαθμοί ελευθερίας της κατανομής $\mathcal{X}^2$ είναι $K - c$, όπου για διακριτή κατανομή $c = 1$ επειδή οι αναμενόμενες συχνότητες E_j δεν είναι όλες ανεξάρτητες αφού το άθροισμα τους πρέπει να είναι n. Για συνεχή κατανομή, οι παράμετροι της εκτιμούνται από τα δεδομένα και άρα χάνονται τόσοι βαθμοί ελευθερίας επιπλέον όσες και οι παράμετροι της κατανομής. Για παράδειγμα για να ελέγξουμε αν τα δεδομένα προσαρμόζονται σε κανονική κατανομή, υπολογίζονται πρώτα οι εκτιμήσεις της μέσης τιμής $\bar{x}$ και διασποράς s^2 από τα δεδομένα για να ορίσουμε την ασκ της κανονικής κατανομής και να υπολογίσουμε στη συνέχεια τις αναμενόμενες τιμές E_j από την σχέση (3.24). Άρα σε αυτήν την περίπτωση οι βαθμοί ελευθερίας της $\mathcal{X}^2$ κατανομής είναι $K - 3$.

Έχοντας ότι κάτω από την H_0 είναι $\chi^2 \sim \mathcal{X}^2_{K-c}$, εξετάζουμε αν το στατιστικό $\tilde{\chi}^2$ από το δείγμα που δίνεται από την σχέση (3.23) ανήκει στην απορριπτική περιοχή $R = \{\chi^2 | \, \chi^2 > \chi^2_{K-c,1-\alpha/2}\}$. Εναλλακτικά μπορούμε να χρησιμοποιήσουμε την p-τιμή που ορίζεται ως $p = \mathrm{P}(\chi^2 > \tilde{\chi}^2\}$.

Ο έλεγχος $\mathcal{X}^2$ μπορεί να εφαρμοσθεί για οποιαδήποτε μονομεταβλητή κατανομή, διακριτή ή συνεχή, για την οποία η ασκ μπορεί να υπολογισθεί. Έχει όμως και κάποια μειονεκτήματα: εξαρτάται από τη διαμέριση για τη διακριτικοποίηση των δεδομένων (όταν είναι αριθμητικά) και απαιτεί αρκετά μεγάλο δείγμα. Υπάρχουν και άλλοι έλεγχοι καλής προσαρμογής κατανομής, όπως ο έλεγχος Kolmogorov-Smirnov που εφαρμόζεται όμως μόνο για συνεχή κατανομή.

Παράδειγμα 3.7. Έστω ένα παιχνίδι ζαριών, όπου ο παίχτης πετάει το ζάρι τρεις φορές και κερδίζει ανάλογα με το πλήθος των εξαριών που φέρνει. Ας υποθέσουμε ότι ένας παίχτης παίζει 100 φορές και οι παρατηρούμενες εμφανίσεις εξαριών δίνονται στον Πίνακα 3.3. Αν το παιχνίδι είναι δίκαιο θα

πλήθος εξαριών	παρατηρούμενο πλήθος	αναμενόμενο πλήθος
0	47	57.9
1	36	34.7
2	14	6.9
3	4	0.5

Πίνακας 3.3: Οι εμφανίσεις εξαριών σε 3 ρίψεις ζαριών για 100 επαναλήψεις. Στη δεύτερη στήλη είναι οι παρατηρούμενες συχνότητες και στην τρίτη οι αναμενόμενες από τη διωνυμική κατανομή $B(3, 1/6)$.

πρέπει να εμφανίζεται εξάρι σε κάθε ζαριά με πιθανότητα 1/6 (πιθανότητα 'επιτυχίας'). Στις 3 ζαριές η πιθανότητα να έρθουν 0,1,2, ή 3 εξάρια δίνεται από τη διωνυμική κατανομή $B(m, p)$ για πλήθος επαναλήψεων $m = 3$, πιθανότητα επιτυχίας $p = 1/6$ και δυνατές τιμές της τ.μ. X εμφάνισης εξαριών 0,1,2,και 3. Θέλουμε να ελέγξουμε αν πράγματι μπορούμε να δεχτούμε ότι $X \sim B(3, 1/6)$ με βάση αυτό το δείγμα.

Οι πιθανότητες εμφάνισης 0,1,2, και 3 εξαριών από την $B(3, 1/6)$ είναι (δες Παράγραφο 2.3.1): $\mathrm{P}(X = 0) = 0.579$, $\mathrm{P}(X = 1) = 0.347$, $\mathrm{P}(X = 2) = 0.069$, $\mathrm{P}(X = 3) = 0.005$. Οι αντίστοιχες αναμενόμενες συχνότητες είναι $(n \cdot \mathrm{P}(X = x))$ και δίνονται στην τρίτη στήλη του Πίνακα 3.3. Συγκρίνοντας τις παρατηρούμενες και τις αναμενόμενες τιμές φαίνεται να υπάρχουν σοβαρές διαφορές για τις πετυχημένες ζαριές με 2 και 3 εξάρια.

Εφαρμόζοντας τη σχέση (3.23) βρίσκουμε $\chi^2 = 36.28$. Η τιμή αυτή είναι πολύ μεγαλύτερη από την κρίσιμη τιμή της $\mathcal{X}_3^2$ (βαθμοί ελευθερίας $K - c = 4 - 1$) για $\alpha = 0.05$ που είναι $\chi^2_{3,0.95} = 7.815$. Η p-τιμή του ελέγχου είναι $p = 6.5 \cdot 10^{-8}$, δηλαδή είναι πάρα πολύ απίθανο το δείγμα αυτό να προέρχεται από τη διωνυμική κατανομή $B(3, 1/6)$. Μάλλον λοιπόν ο παίχτης δεν ρίχνει σωστά τα ζάρια αλλά κάτι κάνει και φέρνει πιο συχνά εξάρες!

Ασκήσεις Κεφαλαίου 3

1. Η κατανομή **Poisson** χρησιμοποιείται αντί της διωνυμικής όταν ο αριθμός m των επαναλήψεων των δοκιμών είναι μεγάλος και η πιθανότητα 'επιτυχίας' p σε κάθε δοκιμή είναι μικρή. Το γινόμενο $\lambda = mp$ είναι η παράμετρος της κατανομής Poisson για την τ.μ. της εμφάνισης x 'επιτυχιών' και δίνεται ως

$$f_X(x;\lambda) = e^{-\lambda}\frac{\lambda^x}{x!} \tag{3.25}$$

όπου $x! = 1 \cdot 2 \cdots x$ είναι το παραγοντικό του x.

(α') Έστω τυχαίο δείγμα ανεξάρτητων παρατηρήσεων $\{x_1,\dots,x_n\}$ από κατανομή Poisson με άγνωστη παράμετρο λ. Δείξετε ότι ο εκτιμητής μέγιστης πιθανοφάνειας του λ είναι η δειγματική μέση τιμή.

(β') Φτιάξτε μια συνάρτηση στο `matlab` που θα δημιουργεί M δείγματα μεγέθους n από κατανομή Poisson με δεδομένη παράμετρο λ και θα υπολογίζει τη δειγματική μέση τιμή για κάθε ένα από τα M δείγματα. Στη συνέχεια θα κάνει κατάλληλο ιστόγραμμα των M δειγματικών μέσων τιμών και θα δίνει ως έξοδο το μέσο όρο από τις M δειγματικές μέσες τιμές. Καλέστε τη συνάρτηση για διαφορετικούς συνδυασμούς των M, n και λ. Είναι πάντα (και σύμφωνα με το αποτέλεσμα στο υποερώτημα 1α') το κέντρο της κατανομής της δειγματικής μέσης τιμής (που περιγράφεται από το ιστόγραμμα) στην τιμή του λ;

Βοήθεια (`matlab`*):* Για τη δημιουργία των τυχαίων αριθμών χρησιμοποίησε τη συνάρτηση `poissrnd`.

2. Ένα ασταθές στοιχείο (unstable particle) μετά από κάποιο χρόνο αποσυντίθεται (decays) σε άλλα σωματίδια. Ο χρόνος ζωής ενός ασταθούς σωματιδίου (μέχρι την αποσύνθεση του) είναι τ.μ. που πιστεύεται ότι ακολουθεί **εκθετική** (exponential) κατανομή με σπп

$$f_X(x;\tau) = \frac{1}{\tau}e^{-x/\tau}, \tag{3.26}$$

όπου τ η παράμετρος της κατανομής. Επαναλάβετε τα ερωτήματα 1α' και 1β' για την εκθετική κατανομή.

Βοήθεια (`matlab`*):* Για τη δημιουργία των τυχαίων αριθμών χρησιμοποίησε τη συνάρτηση `exprnd`.

3. Σε συνέχεια της προηγούμενης άσκησης, προσομοιώστε το χρόνο ζωής n σωματιδίων (νετρονίων) δημιουργώντας n τιμές από εκθετική κατανομή με μέσο χρόνο ζωής $\tau = 15$min. Με βάση αυτό το δείγμα υπολογίστε το 95% παραμετρικό διάστημα εμπιστοσύνης για το μέσο χρόνο ζωής και εξετάστε αν περιέχεται σε αυτό η τιμή $\tau = 15$.

   (α') Υπολογίστε $M = 1000$ δείγματα μεγέθους $n = 5$. Σε τι ποσοστό βρίσκεται ο πραγματικός μέσος χρόνος ζωής μέσα στο 95% διάστημα εμπιστοσύνης;

   (β') Κάνετε το ίδιο για $M = 1000$ αλλά $n = 100$. Διαφέρει το ποσοστό αυτό από το παραπάνω;

   Βοήθεια (`matlab`*):* Για τον υπολογισμό διαστήματος εμπιστοσύνης και ελέγχου για τη μέση τιμή με χρήση της κατανομής Student κάλεσε τη συνάρτηση `ttest`.

4. Η τάση διακοπής εναλλασσόμενου ρεύματος ενός μονωτικού υγρού δηλώνει τη διηλεκτρική ανθεκτικότητα του. Πήραμε τις παρακάτω παρατηρήσεις της τάσης διακοπής (kV) σε κάποιο κύκλωμα κάτω από ορισμένες συνθήκες.

41	46	47	47	48	50	50	50	50	50	50	50
48	50	50	50	50	50	50	50	52	52	53	55
50	50	50	50	52	52	53	53	53	53	53	57
52	52	53	53	53	53	53	53	54	54	55	68

   (α') Βρείτε 95% διάστημα εμπιστοσύνης για τη διασπορά της τάσης διακοπής του κυκλώματος.

   (β') Από παλιότερες μετρήσεις είχαμε βρει πως η τυπική απόκλιση της τάσης διακοπής παρόμοιου κυκλώματος ήταν περίπου 5 kV. Με βάση το δείγμα κάνετε έλεγχο για την υπόθεση πως αυτή είναι η τυπική απόκλιση της τάσης διακοπής.

   (γ') Βρείτε 95% διάστημα εμπιστοσύνης για τη μέση τάση διακοπής του κυκλώματος.

   (δ') Μπορούμε να αποκλείσουμε ότι η μέση τάση διακοπής είναι 52 kV;

   (ε') Κάνετε έλεγχο $\mathcal{X}^2$ καλής προσαρμογής σε κανονική κατανομή και βρείτε την p-τιμή του ελέγχου.

   Βοήθεια (`matlab`*):* Για τον υπολογισμό διαστήματος εμπιστοσύνης και ελέγχου για τη διασπορά με χρήση της κατανομής $\mathcal{X}^2$ κάλεσε τη συνάρτηση `vartest`. Για τον έλεγχο $\mathcal{X}^2$ καλής προσαρμογής κάλεσε τη συνάρτηση `chi2gof`.

5. Ο θερμοπίδακας Old Faithful στην Αμερική είναι από τους πιο γνωστούς θερμοπίδακες για το μέγεθος αλλά και την κανονικότητα των εξάρσεων του (eruptions)
(δες `http://en.wikipedia.org/wiki/Old_Faithful`).
Στο αρχείο δεδομένων `eruption.dat` στην ιστοσελίδα του μαθήματος
`http://users.auth.gr/dkugiu/Teach/DataAnalysis/index.html`
δίνονται στην πρώτη και δεύτερη στήλη 298 μετρήσεις (σε λεπτά) του διαστήματος αναμονής (waiting time) και της διάρκειας του ξεσπάσματος (duration) για το 1989 και στην τρίτη στήλη 298 μετρήσεις του διαστήματος αναμονής έξαρσης για το 2006. Για κάθε ένα από τα τρία μετρούμενα μεγέθη κάνετε τα παρακάτω.

   (α΄) Βρείτε 95% διάστημα εμπιστοσύνης για την τυπική απόκλιση του μεγέθους και ελέγξτε αν είναι $10'$ για την αναμονή και $1'$ για τη διάρκεια.

   (β΄) Βρείτε 95% διάστημα εμπιστοσύνης για τη μέση τιμή του μεγέθους και ελέγξτε αν είναι $75'$ για την αναμονή και $2.5'$ για τη διάρκεια.

   (γ΄) Κάνετε έλεγχο $\mathcal{X}^2$ καλής προσαρμογής σε κανονική κατανομή και βρείτε την p-τιμή του ελέγχου.

   Με βάση τις 298 μετρήσεις για το χρόνο αναμονής και διάρκειας έξαρσης το 1989, εξετάστε αν μπορείτε να δεχθείτε τον παρακάτω ισχυρισμό (αντιγραφή από τη διεύθυνση της Wikipedia): ”With an error of 10 minutes, Old Faithful will erupt 65 minutes after an eruption lasting less than 2.5 minutes or 91 minutes after an eruption lasting more than 2.5 minutes.”