

3.2 ΒΑΣΙΚΕΣ ΘΕΩΡΗΤΙΚΕΣ ΚΑΤΑΝΟΜΕΣ ΔΙΑΚΡΙΤΩΝ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

3.2.1 Διωνυμική κατανομή

Ένα τυχαίο πείραμα στο οποίο υπάρχουν δύο μόνο δυνατά αποτελέσματα λέγεται **δοκιμή Bernoulli**. Το ένα από τα δύο αποτελέσματα της δοκιμής Bernoulli λέγεται **επιτυχία** και το άλλο **αποτυχία**. Για κάθε δοκιμή **Bernoulli** ορίζουμε μια τυχαία μεταβλητή X ως εξής: Αν το αποτέλεσμα s του πειράματος είναι **επιτυχία** τότε $X(s)=1$, αλλιώς $X(s)=0$. Η κατανομή πιθανοτήτων της X είναι:

$$P(x) = \begin{cases} p, & \text{αν } x=1 \\ 1-p, & \text{αν } x=0 \end{cases} \quad \text{όπου } p \text{ είναι η πιθανότητα επιτυχίας.}$$

Έστω μια **δοκιμή Bernoulli** με πιθανότητα επιτυχίας p , επαναλαμβάνεται n φορές, έτσι ώστε το αποτέλεσμα κάθε δοκιμής να είναι ανεξάρτητο από κάθε άλλη δοκιμή. Ορίζουμε την τυχαία μεταβλητή X που εκφράζει τον αριθμό των επιτυχιών σε n ανεξάρτητες επαναλήψεις του πειράματος (άρα $X = 0, 1, 2, \dots, n$). Η **κατανομή πιθανοτήτων** ή **συνάρτηση πιθανότητας** της X είναι

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{όπου} \quad \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Η διωνυμική κατανομή προσδιορίζεται πλήρως από τις παραμέτρους n και p . Γράφουμε $X \sim B(n, p)$ και διαβάζουμε «η X ακολουθεί τη διωνυμική κατανομή με παραμέτρους n και p ».

- Η μέση τιμή της διωνυμικής κατανομής είναι $E(X) = n p$.
- Η διασπορά της διωνυμικής κατανομής είναι $\text{Var}(X) = n p (1-p)$.

Π.χ. Το 25% των μαθητών ενός σχολείου δήλωσαν ότι δεν θα πάνε στην σχολική εκδρομή. Επιλέγουμε τυχαία 15 μαθητές. Να βρεθεί:

- α) η πιθανότητα οι 3 από τους 15 μαθητές να μην πάνε στην εκδρομή.
- β) η πιθανότητα τουλάχιστον 2 από τους 15 μαθητές να μην πάνε στην εκδρομή.
- γ) η μέση τιμή και η διακύμανση των μαθητών.

Λύση

Έστω η μεταβλητή X = αριθμός μαθητών που δεν θα πάνε εκδρομή, η οποία ακολουθεί τη διωνυμική κατανομή με $n = 15$ και $p = 0,25$.

$$\alpha) P(X=3) = \binom{15}{3} 0,25^3 (1-0,25)^{15-3} = \mathbf{0,2246}$$

$$\begin{aligned} \beta) P(X \geq 2) &= 1 - P(X < 2) = 1 - [P(X=0) + P(X=1)] = \\ &= 1 - \left[\binom{15}{0} 0,25^0 (1-0,25)^{15-0} + \binom{15}{1} 0,25^1 (1-0,25)^{15-1} \right] = \\ &= 1 - 0,2138 = \mathbf{0,7862} \end{aligned}$$

$$\gamma) E(X) = n p = 15 \cdot 0,25 = \mathbf{3,75} \quad \text{και} \quad \text{Var}(x) = n p (1-p) = 15 \cdot 0,25 (1-0,25) = \mathbf{2,8125}$$

3.2.2 Κατανομή Poisson

Η κατανομή Poisson είναι η κατανομή των σπάνιων γεγονότων και μας ενδιαφέρει όταν ψάχνουμε τον αριθμό και την πιθανότητα των «συμβάντων» στην μονάδα μέτρησης (π.χ. τυπογραφικά λάθη ανά σελίδα, τηλεφωνικές κλήσεις ανά ώρα κ.τ.λ.). Η τυχαία μεταβλητή που ακολουθεί κατανομή Poisson και εκφράζει το πλήθος των «συμβάντων» ανά μονάδα μέτρησης έχει συνάρτηση πιθανότητας

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

όπου $x = 0, 1, 2, \dots$ και $\lambda > 0$ (παράμετρος).

- Η μέση τιμή της κατανομής Poisson είναι $E(X) = \lambda$.
- Η διασπορά της κατανομής Poisson είναι $Var(X) = \lambda$.

Π.χ. Για μια μεταφορική εταιρεία γνωρίζουμε ότι περίπου 12 φορτηγά ανά 5 ημέρες χρειάζονται επισκευή. Αν η εταιρεία έχει 2 εφεδρικά φορτηγά, να βρεθεί η πιθανότητα μια οποιαδήποτε ημέρα

- α) να μην χρειαστεί η εταιρεία κανένα εφεδρικό φορτηγό
- β) ο αριθμός των εφεδρικών φορτηγών να μην είναι επαρκής.
- γ) Να βρεθεί η μέση τιμή και η τυπική απόκλιση του αριθμού των φορτηγών που παθαίνουν βλάβη κάθε μέρα.

Έστω η μεταβλητή X = αριθμός των φορτηγών που παθαίνουν βλάβη κάθε μέρα. Η X ακολουθεί κατανομή Poisson με $\lambda = 12/5 = 2,4$ (12 βλάβες ανά 5 ημέρες).

$$\alpha) P(X = 0) = e^{-2,4} \frac{2,4^0}{0!} = 0,0907$$

$$\begin{aligned} \beta) P(X > 2) &= 1 - P(X \leq 2) = 1 - [P(X=0) + P(X=1) + P(X=2)] = \\ &= 1 - \left[e^{-2,4} \frac{2,4^0}{0!} + e^{-2,4} \frac{2,4^1}{1!} + e^{-2,4} \frac{2,4^2}{2!} \right] = \\ &= 1 - (0,0907 + 0,21777 + 0,2613) = 0,4303 \end{aligned}$$

$$\gamma) \text{ Μέση τιμή } E(X) = \lambda = 2,4$$

$$\text{Διασπορά } Var(X) = \lambda = 2,4 \Rightarrow \text{Τυπική απόκλιση } \sigma = \sqrt{Var(X)} = \sqrt{2,4} = 1,55$$

3.3 ΚΑΤΑΝΟΜΕΣ ΠΙΘΑΝΟΤΗΤΩΝ ΣΥΝΕΧΩΝ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

Μια τυχαία μεταβλητή ονομάζεται *συνεχής*, αν το πεδίο τιμών της R_X είναι ένα διάστημα ή μια συλλογή διαστημάτων του R . Αν η X είναι μια συνεχής τυχαία μεταβλητή, η πιθανότητα να πάρει η X μια ορισμένη τιμή είναι γενικά μηδέν. Συνεπώς δεν μπορούμε να ορίσουμε μια συνάρτηση πιθανότητας με τον ίδιο τρόπο, όπως μια διακριτή μεταβλητή. Για να ορίσουμε την κατανομή πιθανότητας για μια συνεχή τυχαία μεταβλητή, παρατηρούμε ότι η πιθανότητα να βρίσκεται η X μεταξύ δύο διαφορετικών τιμών έχει νόημα.

Συνάρτηση πυκνότητας πιθανότητας ή συνάρτηση πυκνότητας της συνεχούς τυχαίας μεταβλητής X ονομάζεται η συνάρτηση $f(x)$ για την οποία ισχύουν οι σχέσεις

(1) $f(x) \geq 0$ για κάθε $x \in R_X$

(2) $\int_{R_X} f(x) dx = 1$

(3) $P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x) dx$

Παρατηρήσεις

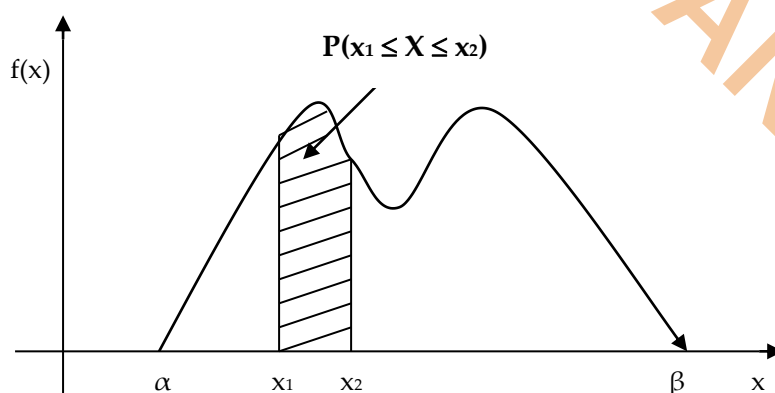
- Το εμβαδόν κάτω από την καμπύλη της συνάρτησης πυκνότητας πιθανότητας $f(x)$ ισούται με 1.

- $P(x_1 \leq X \leq x_2)$ είναι το εμβαδόν κάτω από την καμπύλη της συνάρτησης $f(x)$ μεταξύ των σημείων x_1 και x_2 .

Επίσης ισχύουν τα εξής:

(1) Αν x_1 και x_2 δυο τιμές της X με $x_1 < x_2$ τότε $F(x_1) \leq F(x_2)$.

(2) $P(x_1 \leq X \leq x_2) = P(X \leq x_2) - P(X \leq x_1) = F(x_2) - F(x_1)$



Αθροιστική συνάρτηση πυκνότητας πιθανότητας ή συνάρτηση κατανομής της συνεχούς τυχαίας μεταβλητής X ονομάζεται η συνάρτηση

$F(x)$ για την οποία ισχύει: $F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$

3.4 ΒΑΣΙΚΕΣ ΘΕΩΡΗΤΙΚΕΣ ΚΑΤΑΝΟΜΕΣ ΣΥΝΕΧΩΝ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

ΚΑΝΟΝΙΚΗ ΚΑΤΑΝΟΜΗ

Αν μια συνεχής τυχαία μεταβλητή έχει *συνάρτηση πυκνότητας πιθανότητας* την

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{όπου } E(X) = \mu \text{ και } \text{Var}(X) = \sigma^2.$$

- Η κανονική κατανομή προσδιορίζεται πλήρως από τις παραμέτρους μ και σ^2 . Γράφουμε $X \sim N(\mu, \sigma^2)$ και διαβάζουμε «η X ακολουθεί τη κανονική κατανομή με παραμέτρους μ και σ^2 ».

Αν η τυχαία μεταβλητή X ακολουθεί την κανονική κατανομή $N(\mu, \sigma^2)$, τότε η μεταβλητή $Z = \frac{X - \mu}{\sigma}$ ακολουθεί την τυπική κανονική κατανομή $N(0,1)$.

- Η κατανομή της Z ονομάζεται *τυποποιημένη* ή *τυπική κανονική κατανομή*, με συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad \text{όπου } \mu = 0 \text{ και } \sigma^2 = 1$$

- Η *συνάρτηση αθροιστικής κατανομής* της τυπικής κανονικής κατανομής $N(0,1)$ συμβολίζεται με $\Phi(x) = P(Z \leq x)$ και έχει τις εξής ιδιότητες:

$$\Phi(-x) = 1 - \Phi(x)$$

$$P(\alpha \leq Z \leq \beta) = \Phi(\beta) - \Phi(\alpha)$$

- Αν $F_X(x)$ η *συνάρτηση αθροιστικής κατανομής* της $X \sim (\mu, \sigma^2)$ και $F_Z(z) = \Phi(z)$ η *συνάρτηση αθροιστικής κατανομής* της $Z \sim (0, 1)$ τότε

$$F_X(x) = F_Z\left(\frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

$$P(a \leq X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

- Οι πιθανότητες $\Phi(z) = P(Z \leq z)$ είναι γνωστές και δίνονται σε πίνακα.

3.5 ΚΕΝΤΡΙΚΟ ΟΡΙΑΚΟ ΘΕΩΡΗΜΑ

Το θεώρημα αυτό αποδεικνύει ότι η κατανομή των μέσων τυχαίων δειγμάτων με μέγεθος σχετικά μεγάλο, πρακτικά μεγαλύτερο ή ίσο του τριάντα, δεν εξαρτάται από την κατανομή του αρχικού πληθυσμού αλλά ακολουθεί, κάτω από ορισμένες προϋποθέσεις, οι οποίες συνήθως πληρούνται, την κανονική κατανομή. Όταν το μέγεθος του δείγματος είναι μικρότερο του τριάντα, για να ακολουθεί η κατανομή δειγματοληψίας των μέσων την κανονική κατανομή, θα πρέπει και ο πληθυσμός από τον οποίο προέρχεται το δείγμα να ακολουθεί την κανονική κατανομή.

Αν οι τυχαίες μεταβλητές $X_1, X_2, \dots, X_n$ έχουν την ίδια κατανομή πιθανοτήτων, τότε ονομάζονται **ισόνομες τυχαίες μεταβλητές**.

Κεντρικό οριακό θεώρημα

Αν $X_1, X_2, \dots, X_n$ είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με μέση τιμή μ και πεπερασμένη διακύμανση σ^2 , τότε όσο αυξάνει το n , η κατανομή της τυχαίας μεταβλητής

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

προσεγγίζει την τυπική κανονική κατανομή.

3.6 Ασκήσεις

3.6.1. Έστω κάποιος αγοράζει 10 μετοχές προς 50 ευρώ τη μια. Πληροφορείται ότι σε έναν χρόνο:

- οι μετοχές δεν θα αξίζουν τίποτα, με πιθανότητα 0,1
- οι μετοχές θα αξίζουν όσο ήταν η αρχική τους τιμή με πιθανότητα 0,5
- οι μετοχές θα αξίζουν διπλάσια από την αρχική τους τιμή με πιθανότητα 0,4.

Αν X το κέρδος από την αγορά 10 μετοχών, να βρεθούν

- (α) η συνάρτηση πιθανότητας της X
- (β) η αθροιστική συνάρτηση πιθανότητας της X
- (γ) η μέση τιμή της X
- (δ) η διακύμανση και η τυπική απόκλιση της X

3.6.2. Έστω ότι για κάθε παιδί που γεννιέται η πιθανότητα να είναι κορίτσι είναι 0.5. Να υπολογιστεί η πιθανότητα σε μια οικογένεια με 10 παιδιά:

- (α) να είναι όλα κορίτσια
- (β) να είναι κορίτσια τουλάχιστον τα 5
- (γ) να είναι το πολύ 8 κορίτσια.

3.6.3. Έστω μια παραγγελία ηλεκτρικών συσκευών περιλαμβάνει 10% ελαττωματικούς. Εάν επιλέξουμε τυχαία 4 συσκευές, να βρεθεί η πιθανότητα

- (α) ακριβώς μια συσκευή να είναι ελαττωματική

(β) τουλάχιστον μια συσκευή να είναι ελαττωματική, στο δείγμα των 4 συσκευών
(γ) Η μέση τιμή και η διακύμανση των συσκευών που είναι ελαττωματικές.

3.6.4. Σε μια ποσότητα μήλων, το 15% είναι χαλασμένα. Παίρνουμε τυχαία 5 μήλα. Να βρεθεί η πιθανότητα:

(α) ακριβώς ένα μήλο είναι χαλασμένο

(β) κανένα μήλο δεν είναι χαλασμένο

(γ) τουλάχιστον ένα μήλο είναι χαλασμένο

(γ) Να βρεθεί η μέση τιμή και η διακύμανση του αριθμού των χαλασμένων μήλων.

3.6.5. Έστω η πιθανότητα να κάνει κάποιος ένα λάθος στην δακτυλογράφηση μιας σελίδας είναι 0.01. Ποια η πιθανότητα σε ένα κείμενο 350 σελίδων

(α) να υπάρχει μια σελίδα με τουλάχιστον ένα σφάλμα

(β) το πολύ 4 σελίδες με τουλάχιστον ένα σφάλμα σε κάθε μια σελίδα

3.6.6. Οι πελάτες ενός καταστήματος φτάνουν στο ταμείο με ρυθμό 8 πελατών /ώρα. Αν ο αριθμός των αφίξεων των πελατών ακολουθεί κατανομή Poisson, τότε να βρεθεί για μια δεδομένη ώρα, η πιθανότητα

(α) να φτάνουν ακριβώς 8 πελάτες

(β) να φτάνουν το πολύ 3 πελάτες

(γ) να φτάνουν τουλάχιστον 2 πελάτες

3.6.7. Ο αριθμός X των πλοίων που φτάνουν σε ένα λιμάνι σε μια ημέρα ακολουθεί κατανομή Poisson με $\lambda=3$ πλοία. Αν στο λιμάνι δεν μπορούν να εξυπηρετηθούν πάνω από 5 πλοία την ημέρα, ποια η πιθανότητα σε μια ημέρα να υπάρξουν πλοία που δεν θα εξυπηρετηθούν;

3.6.8. Αν η τυχαία μεταβλητή Z ακολουθεί την τυπική κανονική κατανομή, να υπολογισθούν οι πιθανότητες

α) $P(Z \leq 1,25)$

β) $P(Z \leq -1,25)$

γ) $P(0,3 \leq Z \leq 1,3)$

δ) $P(Z \geq 2,1)$

3.6.9. Αν $X \sim N(200,100)$ να υπολογισθούν οι πιθανότητες

α) $P(190 < X < 210)$

β) $P(180,4 < X \leq 219,6)$

3.6.10. Αν $X \sim N(4,9)$ να υπολογισθεί η πιθανότητα $P(-2 \leq X \leq 5)$.

Κεφάλαιο 4

ΕΚΤΙΜΗΤΙΚΗ

4.1 Εισαγωγή

Στατιστική συμπερασματολογία είναι η μεθοδολογία με την οποία συγκεντρώνονται συμπεράσματα για έναν στατιστικό πληθυσμό βάση αποτελεσμάτων που λαμβάνονται από ένα τυχαίο δείγμα.

Ένας **εκτιμητής σε σημείο** είναι μια συνάρτηση των δεδομένων του τυχαίου δείγματος. Ο **εκτιμητής σε σημείο** είναι μια τιμή ενός στατιστικού που χρησιμοποιείται ως η καλύτερη προσέγγιση στην αντίστοιχη παράμετρο του πληθυσμού την οποία εκτιμά ο εκτιμητής.

π.χ. ο δειγματικός μέσος $\bar{x}$ είναι ο εκτιμητής σε σημείου του πληθυσμιακού μέσου μ ενός πληθυσμού.

Ο **εκτιμητής σε διάστημα ή διάστημα εμπιστοσύνης** είναι ένα διάστημα για το οποίο μπορεί να καθοριστεί με διάφορους βαθμούς εμπιστοσύνης ότι περιέχει την παράμετρο του πληθυσμού που εκτιμάται.

Συντελεστής ή βαθμός εμπιστοσύνης ονομάζεται η πιθανότητα με την οποία η εκτιμώμενη παράμετρος περιέχεται στο διάστημα εμπιστοσύνης. Αν ο συντελεστής εμπιστοσύνης είναι $1-\alpha$ τότε το διάστημα εμπιστοσύνης λέγεται $100(1-\alpha)\%$ διάστημα εμπιστοσύνης.

4.2. ΔΙΑΣΤΗΜΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΓΙΑ ΤΗ ΜΕΣΗ ΤΙΜΗ ΕΝΟΣ ΠΛΗΘΥΣΜΟΥ

Α. Αν είναι γνωστή η διασπορά σ^2 ενός πληθυσμού τότε το $100(1-\alpha)\%$ διάστημα εμπιστοσύνης για τη μέση τιμή μ του είναι

$$\left(\bar{x} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

π.χ. Το καθαρό βάρος X μιας συσκευασίας βούτυρο ακολουθεί την κανονική κατανομή με τυπική απόκλιση $\sigma = 5$ gr. Σε τυχαίο δείγμα 25 πακέτων υπολογίσαμε το μέσο καθαρό βάρος $\bar{x} = 243$ gr. Να εκτιμηθεί το διάστημα εμπιστοσύνης με πιθανότητα 0.99 για το μέσο καθαρό βάρος στο σύνολο των συσκευασιών.

$$1 - \alpha = 0.99 \Leftrightarrow \alpha = 0.01 \Leftrightarrow 1 - \alpha/2 = 0.995 \rightarrow \Phi(z_{1-\alpha/2}) = 1 - \alpha/2$$

$$\Phi(z_{1-\alpha/2}) = 0.995 \rightarrow z_{1-\alpha/2} = 2.58$$

Άρα το $100(1-\alpha)\%$ διάστημα εμπιστοσύνης για τη μέση τιμή μ είναι:

$$\left(\bar{x} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right) \Leftrightarrow \left(243 - 2.58 \frac{5}{\sqrt{25}}, 243 + 2.58 \frac{5}{\sqrt{25}} \right) \Leftrightarrow (240.42, 245.58)$$

B. Αν δεν είναι γνωστή η διασπορά σ^2 ενός πληθυσμού και το δείγμα είναι μεγάλο τότε το $100(1-\alpha)\%$ διάστημα εμπιστοσύνης για τη μέση τιμή μ του είναι

$$\left(\bar{x} - z_{1-\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2} \frac{s}{\sqrt{n}} \right)$$

όπου s είναι η δειγματική τυπική απόκλιση.

π.χ. Σε φορολογικό έλεγχο που έγινε σε 255 καταστήματα υπολογίστηκε το μέσο ποσό φοροκλοπής σε 540 ευρώ, με τυπική απόκλιση 110 ευρώ. Να εκτιμηθεί διάστημα εμπιστοσύνης με πιθανότητα 99% για το μέσο ποσό φοροκλοπής στο σύνολο των καταστημάτων.

$$1 - \alpha = 0.99 \Leftrightarrow \alpha = 0.01 \Leftrightarrow 1 - \alpha/2 = 0.995 \rightarrow \Phi(z_{1-\alpha/2}) = 1 - \alpha/2$$

$$\Phi(z_{1-\alpha/2}) = 0.995 \rightarrow z_{1-\alpha/2} = 2.58$$

Άρα το $100(1-\alpha)\%$ διάστημα εμπιστοσύνης για τη μέση τιμή μ είναι:

$$\begin{aligned} \bar{x} - z_{1-\alpha/2} \frac{s}{\sqrt{n}} < \mu < \bar{x} + z_{1-\alpha/2} \frac{s}{\sqrt{n}} &\Leftrightarrow \\ 540 - 2.58 \frac{110}{\sqrt{255}} < \mu < 540 + 2.58 \frac{110}{\sqrt{255}} &\Leftrightarrow \\ 522.228 < \mu < 557.772 \end{aligned}$$

4.3. Ασκήσεις

4.3.1. Σε τυχαίο δείγμα 9 συσκευασιών γάλατος υπολογίσαμε το μέσο βάρος $\bar{x}=96$ gr. Αν είναι γνωστό ότι το βάρος X μιας συσκευασίας γάλατος ακολουθεί κανονική κατανομή με τυπική απόκλιση $\sigma = 4.5$ gr, να εκτιμηθεί το διάστημα εμπιστοσύνης με πιθανότητα 95% για το μέσο βάρος σε όλες τις συσκευασίες.

4.3.2. Σε τυχαίο δείγμα 35 μπαταριών υπολογίσαμε τη μέση διάρκεια ζωής 217 ώρες. Αν η τυπική απόκλιση του δείγματος είναι 47 ώρες, να εκτιμηθεί το διάστημα εμπιστοσύνης με πιθανότητα 95% για τη μέση διάρκεια ζωής στο σύνολο των μπαταριών.

Κεφάλαιο 5

ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ

5.1. Εισαγωγή

Υπόθεση ονομάζεται η απόφαση που παίρνουμε για τα θέματα σχετικά με τους πληθυσμούς, βασιζόμενοι στις πληροφορίες που παίρνουμε από τα δείγματα των πληθυσμών. **Έλεγχος υποθέσεων ή έλεγχος σημαντικότητας ή κανόνες αποφάσεων** ή στατιστικός έλεγχος ονομάζεται η διαδικασία που χρησιμοποιείται ώστε να αποφασίσουμε αν θα δεχτούμε ή θα απορρίψουμε τις υποθέσεις που έχουμε κάνει. **Μηδενική υπόθεση H_0** ορίζεται ως η υπόθεση που κάνουμε αρχικά με σκοπό να την απορρίψουμε. **Εναλλακτική υπόθεση H_1** ορίζεται ως η ασυμβίβαστη υπόθεση σε σχέση με την μηδενική υπόθεση. Η απόφαση αν θα γίνει δεκτή ή αν θα απορριφθεί η μηδενική υπόθεση H_0 στηρίζεται σε ένα στατιστικό που ονομάζεται **στατιστικό του τεστ**, το οποίο υπολογίζεται από τα δεδομένα του δείγματος. **Απορριπτική περιοχή R της H_0** ονομάζεται η περιοχή στα σημεία της οποίας η H_0 απορρίπτεται.

Τα στοιχεία ενός στατιστικού ελέγχου είναι τα εξής:

- Ορίζεται η μηδενική υπόθεση H_0
- Ορίζεται η εναλλακτική υπόθεση H_1
- Ορίζεται το στατιστικό του ελέγχου από τα δεδομένα του δείγματος
- Ορίζεται η απορριπτική περιοχή R της υπόθεσης H_0
- Εξάγονται τα συμπεράσματα

Μονόπλευρος έλεγχος ορίζεται ως ο έλεγχος που κάνουμε όταν θέλουμε να δούμε αν η υπόθεση ότι η διαδικασία που επιλέξαμε είναι καλύτερη από μία άλλη είναι ευσταθής. Δηλαδή αν $H_0 : \theta = \theta_0$, τότε $H_1 : \theta > \theta_0$ ή $H_1 : \theta < \theta_0$. **Δίπλευρος έλεγχος** ορίζεται ως ο έλεγχος που γίνεται στις τιμές των άκρων της κατανομής που μελετάμε. Δηλαδή αν $H_0 : \theta = \theta_0$, τότε $H_1 : \theta \neq \theta_0$.

Σφάλμα τύπου I ονομάζεται η απόρριψη της υπόθεσης H_0 ενώ είναι σωστή. Η πιθανότητα αυτού του σφάλματος συμβολίζεται με α και ονομάζεται **επίπεδο σημαντικότητας ή σημαντικότητα ενός ελέγχου**. **Σφάλμα τύπου II** ονομάζεται η αποδοχή της υπόθεσης H_0 ενώ είναι λάθος. Η πιθανότητα αυτού του σφάλματος συμβολίζεται με β .

Ο πληθυσμός βάση κάποιου κριτηρίου χωρίζεται σε δυο περιοχές: στην περιοχή αποδοχής της H_0 και στην περιοχή απόρριψης της που είναι η μικρότερη. Κατόπιν γίνεται ο έλεγχος βάση του κριτηρίου που ορίστηκε και αποφασίζεται που ανήκει το δείγμα. Εάν ανήκει στην περιοχή αποδοχής, τότε η H_0 γίνεται δεκτή. Θεωρούμε ότι ο εκτιμητής της παραμέτρου που εξετάζουμε στον έλεγχο ακολουθεί κανονική κατανομή.

5.2. ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΗΣ ΓΙΑ ΤΗ ΜΕΣΗ ΤΙΜΗ ΠΛΗΘΥΣΜΟΥ

5.2.1. Έλεγχος υπόθεσης για τη μέση τιμή μ ενός πληθυσμού αν σ^2 γνωστό

Αν $\bar{x}$ είναι η δειγματική μέση τιμή και σ η τυπική απόκλιση και μ_0 η επιθυμητή μέση τιμή, τότε υπολογίζουμε το στατιστικό:

$$z_{\pi} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Μονόπλευρο τεστ

$H_0: \mu = \mu_0, H_1: \mu > \mu_0 \Rightarrow$ Απορριπτική περιοχή $R = \{z_{\pi} > z_{1-\alpha}\}$ ή

$H_0: \mu = \mu_0, H_1: \mu < \mu_0 \Rightarrow$ Απορριπτική περιοχή $R = \{z_{\pi} < -z_{1-\alpha}\}$

Δίπλευρο τεστ

$H_0: \mu = \mu_0, H_1: \mu \neq \mu_0 \Rightarrow$ Απορριπτική περιοχή $R = \{|z_{\pi}| > z_{1-\alpha/2}\}$

5.2.2. Έλεγχος υπόθεσης για τη μέση τιμή μ ενός πληθυσμού αν σ^2 άγνωστο και $n \geq 30$

Ομοίως με πριν αλλά το σ αντικαθίσταται με το s (δειγματική τυπική απόκλιση).

5.2.3. Έλεγχος υπόθεσης για τη μέση τιμή μ ενός πληθυσμού αν σ^2 άγνωστο και $n < 30$

Στην περίπτωση αυτή η μεταβλητή $t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$ ακολουθεί την κατανομή Student

με $v = n-1$. Αν μ_0 είναι η επιθυμητή μέση τιμή, τότε τα κριτήρια απόφασης είναι:

Μονόπλευρο τεστ

$H_0: \mu = \mu_0, H_1: \mu > \mu_0 \Rightarrow$ Απορριπτική περιοχή $R = \{t_{\pi} > t_{v,\alpha}\}$ ή

$H_0: \mu = \mu_0, H_1: \mu < \mu_0 \Rightarrow$ Απορριπτική περιοχή $R = \{t_{\pi} < -t_{v,\alpha}\}$

Δίπλευρο τεστ

$H_0: \mu = \mu_0, H_1: \mu \neq \mu_0 \Rightarrow$ Απορριπτική περιοχή $R = \{|t_{\pi}| > t_{v,\alpha/2}\}$

5.3. Ασκήσεις

5.3.1. Το βάρος X των βρεφών που γεννήθηκαν το 2000 ακολουθεί την κανονική κατανομή $N(3300,900)$. Σε τυχαίο δείγμα $n=20$ βρεφών το 2009 βρέθηκε ότι $\bar{x}=3380$. Να ελεγχθεί η υπόθεση ότι το μέσο βάρος των βρεφών δεν αυξήθηκε κατά την περίοδο 2000-2009, σε στάθμη σημαντικότητας $\alpha=0.05$.

5.3.2. Ο αριθμός X των αμειβόμενων υπερωριών μιας επιχείρησης ακολουθεί κανονική κατανομή $N(43,100)$. Η διοίκηση της επιχείρησης έδωσε εντολή να μειωθούν οι υπερωρίες. Σε τυχαίο δείγμα $n=20$ εργαζομένων βρέθηκε ότι ο μέσος όρος των υπερωριών είναι 32 ώρες. Να ελεγχθεί η υπόθεση ότι δεν μειώθηκε ο μέσος όρος των υπερωριών, σε στάθμη σημαντικότητας $\alpha=0.01$.

5.3.3. Επιχείρηση σοκολατοποιίας διαφημίζει ότι τα προϊόντα της έχουν μέση επικάλυψη σοκολάτας ίση με 10 γραμμάρια. Η επικάλυψη σοκολάτας X είναι γνωστό ότι ακολουθεί κανονική κατανομή με τυπική απόκλιση $\sigma=1$ γραμμάρια. Αν ο ιδιοκτήτης δέχεται να καταστραφεί η ημερήσια παραγωγή 1 στις 100 περιπτώσεις, και ο ποιοτικός έλεγχος μια ημέρα σε $n = 25$ προϊόντα έδειξε ότι $\bar{x}=9.3$ γραμμάρια, τότε θα αχρηστευτεί η ημερήσια παραγωγή;

5.3.4. Έστω το μέσο βάρος κάποιων προϊόντων δεν διαφέρει από 700 γραμμάρια σε επίπεδο σημαντικότητας $\alpha = 0.05$. Αν σε τυχαίο δείγμα $n=18$ προϊόντων υπολογίσαμε μέσο βάρος $\bar{x} = 692$ γραμμάρια και τυπική απόκλιση $s = 20$ γραμμάρια, και η κατανομή του βάρους μπορεί να υποτεθεί κανονική, τότε να ελεγχθεί η παραπάνω υπόθεση.

Κεφάλαιο 6

ΕΠΑΝΑΛΗΠΤΙΚΕΣ ΑΣΚΗΣΕΙΣ

6.1. Δίνονται τα παρακάτω δεδομένα που παριστάνουν το πλήθος των χρωστούμενων μαθημάτων από μια ομάδα 25 φοιτητών του τμήματος :

1, 0, 0, 4, 5, 10, 7, 6, 2, 2, 3, 3, 4, 1, 0, 0, 1, 2, 2, 3, 3, 1, 1, 0, 0.

- α) Ποιο είναι το ποσοστό των φοιτητών που χρωστούν δύο μαθήματα;
- β) Βρείτε τη διάμεσο, το πρώτο και τρίτο τεταρτημόριο των παραπάνω παρατηρήσεων.
- γ) Σχεδιάστε το θηκόγραμμα των παραπάνω παρατηρήσεων.
- δ) Υπολογίστε το μέσο πλήθος των χρωστούμενων μαθημάτων και την τυπική απόκλιση αυτού.

6.2. Μια έρευνα έδειξε ότι το 35% των νέων αγοραστών αυτοκινήτων είναι γυναίκες. Επιλέγουμε τυχαία 5 νέους αγοραστές αυτοκινήτων. Να βρεθεί:

- α) η πιθανότητα και οι 5 αγοραστές να είναι γυναίκες.
- β) η πιθανότητα ακριβώς 3 από τους 5 αγοραστές να είναι γυναίκες
- γ) το πολύ 2 αγοραστές να είναι γυναίκες
- δ) η μέση τιμή και η διακύμανση του αριθμού των γυναικών που θα μπορούσαν να εκλεγούν.

6.3. Ένα εργοστάσιο παράγει μπαταρίες συγκεκριμένου τύπου για χρήση σε συσκευές κινητής τηλεφωνίας τις οποίες συσκευάζει σε πακέτα των 120. Σύμφωνα με στοιχεία του εργοστασίου ο χρόνος ζωής μιας μπαταρίας του τύπου αυτού μπορεί να θεωρηθεί ότι ακολουθεί την κανονική κατανομή με μέση τιμή $\mu=40$ ώρες και τυπική απόκλιση $\sigma=5$ ώρες. Ακόμη σύμφωνα με τις προδιαγραφές που έχει θέσει η Ευρωπαϊκή Ένωση αν μία μπαταρία του συγκεκριμένου τύπου έχει χρόνο ζωής μικρότερο από 30 ώρες θεωρείται ελαττωματική. Με βάση τα στοιχεία αυτά:

- α) Αν επιλεγεί τυχαία ένα πακέτο μπαταριών να υπολογισθεί ο αριθμός των μπαταριών των οποίων ο χρόνος ζωής θα ξεπερνά τις 45 ώρες καθώς και ο αριθμός των μπαταριών των οποίων ο χρόνος ζωής θα είναι μεταξύ 30 και 50 ώρες.
- β) Αν επιλεγεί τυχαία ένα πακέτο μπαταριών να υπολογισθεί ο αριθμός των ελαττωματικών μπαταριών που περιέχει.
- γ) Αν επιλεγούν τυχαία 4 μπαταρίες από ένα τυχαία επιλεγμένο πακέτο να υπολογισθεί η πιθανότητα δύο τουλάχιστον να είναι ελαττωματικές;

6.4. Ο χρόνος εκτύπωσης των αρχείων που αποστέλλονται από τους υπαλλήλους μιας εταιρείας στον κεντρικό εκτυπωτή ακολουθεί κανονική κατανομή με μέση τιμή 2 min και διακύμανση $6,25 \text{ min}^2$.

α) Να υπολογισθεί η πιθανότητα το επόμενο αρχείο που θα σταλεί για εκτύπωση να χρειαστεί χρόνο:

i) Μεγαλύτερο από 2,3 λεπτά.

ii) Μεταξύ 1,35 και 3 λεπτών.

β) Να υπολογισθεί η πιθανότητα μεταξύ 3 αρχείων που στέλνονται ανεξάρτητα για εκτύπωση το ένα ακριβώς να εκτυπωθεί σε χρόνο μικρότερο του 1,5 min.

6.5. Υποθέτουμε ότι ένα περιφερειακό κέντρο υπολογιστών θέλει να υπολογίσει την απόδοση της μνήμης του συστήματος με βάση τον μέσο χρόνο μεταξύ αποτυχιών εκτέλεσης μιας λειτουργίας του δίσκου της μνήμης. Για να υπολογίσει αυτήν την τιμή, το κέντρο ανέγραψε το χρόνο αποτυχιών για ένα τυχαίο δείγμα 45 αποτυχιών. Αν η δειγματική μέση τιμή είναι 1762 ώρες και η τυπική απόκλιση 215 ώρες:

α) να εκτιμήσετε την πραγματική μέση τιμή του χρόνου μεταξύ αποτυχιών με ένα 90% διάστημα εμπιστοσύνης.

β) Εάν ο δίσκος μνήμης του συστήματος τρέχει σωστά, η πραγματική μέση τιμή ξεπερνά τις 1700 ώρες. Βάση του διαστήματος εμπιστοσύνης που εκτιμήσατε στο ερώτημα α, το μπορείτε να συμπεράνετε για το δίσκο μνήμης.

6.6. Το IQ ενός ενήλικα μετρημένο σύμφωνα με ορισμένο τεστ ακολουθεί την κανονική κατανομή με μέση τιμή 100 και διασπορά 48,36. Ποια είναι η πιθανότητα ένας ενήλικας να έχει IQ πάνω από 107?

6.7. Σοκολατοποιία διαφημίζει ότι ορισμένο προϊόν της έχει μέση επικάλυψη σοκολάτας ίση με 10gr. Σε τυχαίο δείγμα 16 προϊόντων μετρήσαμε την επικάλυψη X και πήραμε τα ακόλουθα αποτελέσματα:

8.4, 8.5, 8.5, 8.7, 8.7, 8.8, 8.9, 9.0, 9.0, 9.2, 9.4, 9.5, 9.8, 9.9, 10.0, 10.5

Να βρεθεί το διάστημα εμπιστοσύνης με πιθανότητα 99% για τη μέση επικάλυψη σοκολάτας, αν η X ακολουθεί κανονική κατανομή.

6.8. Ο σύλλογος καταναλωτών για να εξακριβώσει αν τα κουτιά των 100 γραμμαρίων καφέ περιέχουν πραγματικά 100 γραμμάρια, πήρε 37 κουτιά καφέ τυχαία και βρήκε μέσο περιεχόμενο σε καφέ 96 γραμμάρια με τυπική απόκλιση 1,8 γραμμάρια. Να ελεγχθεί η υπόθεση ότι τα κουτιά των 100 γραμμαρίων είναι ελλιποβαρή, σε στάθμη σημαντικότητας $\alpha=0,05$.

6.9. Ο αριθμός X των υπερωριών των εργαζομένων μιας τράπεζας ακολουθεί κανονική κατανομή με μέση τιμή 18 και διασπορά 4. Σε ένα τυχαίο δείγμα από 35 εργαζόμενους βρέθηκε ότι ο μέσος όρος των υπερωριών που έκαναν είναι 20 ώρες. Να ελεγχθεί η υπόθεση ότι δεν αυξήθηκε ο μέσος όρος ωρών των υπερωριών, σε στάθμη σημαντικότητας $\alpha=0.01$.

6.10. Ρωτήθηκαν 10 φοιτητές μιας σχολής που είναι στο πτυχίο για τον αριθμό X των μαθημάτων που χρωστάνε και έστω ότι οι απαντήσεις τους είναι οι εξής:

6, 2, 1, 9, 7, 3, 2, 1, 5, 1, 4, 6.

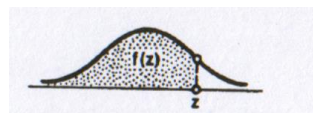
α) Να υπολογίσετε την δειγματική μέση τιμή και τυπική απόκλιση.

β) Αν γνωρίζετε ότι η μεταβλητή X ακολουθεί κανονική κατανομή, να υπολογίσετε το 90% διάστημα εμπιστοσύνης για τη μέση τιμή του δείγματος.

ANGELIKI PAPANA

ΤΥΠΟΛΟΓΙΑ

Πίνακας 1. Αθροιστικές πιθανότητες της τυπικής κανονικής κατανομής



$z/$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997
4.0	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998