Συμπληρωματικές Ασκήσεις Στατιστικής ΙΙΙ

ΚΕΦΑΛΑΙΑ 1-3

Άσκ. 1. Η κατανομή των βαρών των μαθητών ενός σχολείου είναι κανονική με μ=60kg και σ=5kg. Να υπολογιστεί η πιθανότητα επιλέγοντας τυχαία έναν μαθητή να είναι βαρύτερος από 70kg.

Άσκ. 2. Σε μία παραλία το βάρος X των χαλικιών ακολουθεί κανονική κατανομή $N(30,10^2)$.
(i) Αν πάρουμε τυχαία ένα χαλίκι
  (α) Ποια η πιθανότητα να έχει βάρος $X > 30$gr
  (β) Ποια η πιθανότητα να έχει βάρος $28 < X < 40$gr
  (γ) Πήραμε 10 χαλίκια τυχαία. Ποια η πιθανότητα το συνολικό βάρος τους να είναι πάνω από 300gr
(ii) Αν πάρουμε 4 χαλίκια τυχαία, ποια η πιθανότητα να βρούμε 2 ακριβώς με βάρος $28 < X < 40$gr.

Άσκ. 3. Τα αποτελέσματα σε ένα τεστ δεξιοτήτων ακολουθούν κανονική κατανομή με μ=500 και σ=100. Ποιος είναι ο μεγαλύτερος βαθμός που μπορεί να έχει ένας μαθητής, ώστε να βρίσκεται στο 20% της μικρότερης βαθμολογίας της κατανομής;

Άσκ. 4. Από το αρχείο ενός ιατρείου προκύπτει ότι η πιθανότητα αναμονής ενός ασθενούς για χρόνο περισσότερο από 20 λεπτά είναι 0,0239. Αν ο χρόνος αναμονής ακολουθεί κανονική κατανομή με σ=3,75 λεπτά, να υπολογίσετε: (α) το μέσο χρόνο αναμονής στο ιατρείο, (β) την πιθανότητα, ένας ασθενής να περιμένει στο ιατρείο για χρόνο μεταξύ 10 και 15 λεπτών.

Άσκ. 5. Τα ανδρικά πουκάμισα ταξινομούνται ανάλογα με το μέγεθός τους σε 4 κατηγορίες S, M, L, και XL που αντιστοιχούν σε μήκος περιφέρειας λαιμού ανδρών λιγότερο από 15 ίντσες για το S, μεταξύ 15 ιντσών και 16 ιντσών για το M, μεταξύ 16 ιντσών και 17 ιντσών για το L, και περισσότερο από 17 ίντσες για το XL. Έστω ότι το μήκος της περιφέρειας του λαιμού των ανδρών σε ίντσες είναι τ.μ. X που ακολουθεί την κατανομή $N(15.75, 0.49)$. (α) Να βρεθεί πόσα πουκάμισα ανά κατηγορία πρέπει να κατασκευάσει μια βιοτεχνία αν η παραγωγή της είναι 10000 πουκάμισα. (β) Να καθορισθούν τα όρια του μήκους της περιφέρειας του λαιμού των ανδρών για κάθε μια από τις 4 κατηγορίες S, M, L, και XL, έτσι ώστε σε κάθε μια κατηγορία να αντιστοιχεί το 25% του πληθυσμού των ανδρών.

Άσκ. 6. Η ημερήσια ζήτηση κρέατος σε κιλά στην Κεντρική Αγορά Αθηνών (ΚΑΑ) ακολουθεί την κανονική κατανομή με μέση τιμή 5000Kg και τυπική απόκλιση 300Kg. (α) Αν σε μια ημέρα η ΚΑΑ εφοδιαστεί με 5300Kg τότε ποια είναι η πιθανότητα να καλυφθεί όλη η ζήτηση; (β) Ποιά είναι η ποσότητα q που πρέπει να εφοδιαστεί η ΚΑΑ σε μια ημέρα έτσι ώστε να καλυφθεί η ζήτηση με πιθανότητα 0.9;

Άσκ. 7. Από τα στατιστικά στοιχεία που εξέδωσε το υπουργείο εθνικής αμύνης των ΗΠΑ προκύπτει ότι το 8.08% των αμερικανών στρατιωτών έχει ύψος μικρότερο από 67.2 ίντσες, ενώ το 13.57% των αμερικανών στρατιωτών έχει ύψος μεγαλύτερο από 72.2 ίντσες. Αν υποτεθεί ότι το ύψος των αμερικανών στρατιωτών σε ίντσες είναι τ.μ. X που ακολουθεί την κανονική κατανομή, να βρεθεί η μέση τιμή και η διακύμανση της τ.μ. X.

Άσκ. 8. Η διάρκεια ζωής X (σε στροφές$\times 10^6$) ενός ανταλλακτικού κινητήρα εσωτερικής καύσης ακολουθεί την κατανομή $N(50,5^2)$ αν είναι τύπου Α, και την κατανομή $N(30,5^2)$ αν είναι τύπου Β. Αν το 25% των κινητήρων είναι εφοδιασμένα με ανταλλακτικό τύπου Β, να βρεθεί η πιθανότητα να έχει ένας κινητήρας ανταλλακτικό τύπου Β αν η διάρκεια ζωής του ήταν μικρότερη από 40×10^6 στροφές.

Άσκ. 9. Το βάρος των ατόμων ενός πληθυσμού ακολουθεί κανονική κατανομή $N(81, 10^2)$. Από τον πληθυσμό, επιλέγουμε τυχαία 6 άτομα. Να βρεθεί η πιθανότητα το συνολικό τους βάρος να υπερβαίνει τα 500kgr

Άσκ. 10. Έστω X, Y ανεξάρτητες τ.μ. με $X \sim N(\mu_X,\sigma_X^2)$, $Y \sim N(\mu_Y,\sigma_Y^2)$. Δείξτε ότι

$$Var(XY) = \sigma_X^2\sigma_Y^2 + \mu_Y^2\sigma_X^2 + \mu_X^2\sigma_Y^2 .$$

Άσκ. 11. Έστω X, Y είναι δύο τ.μ. οι οποίες παίρνουν τις τιμές −1 ή 1. Αν $E(X) = E(Y) = 0$, να δείξετε ότι

$$P(X=1,Y=1) = P(X=-1,Y=-1) \quad \text{και} \quad P(X=1,Y=-1) = P(X=-1,Y=1)$$

Αν επιπλέον είναι γνωστό ότι $p = 2P(X=1,Y=1)$ βρείτε την $V(X)$, $V(Y)$ και $\mathrm{Cov}(X,Y)$.

Άσκ. 12. Έστω X και Y είναι ανεξάρτητες τ.μ. οι οποίες ακολουθούν διωνυμική κατανομή με παραμέτρους n, p και m, p αντίστοιχα. Αποδείξτε (με επιχειρήματα, χωρίς να κάνετε υπολογισμούς) ότι η νέα τ.μ. $X+Y$ ακολουθεί και αυτή διωνυμική κατανομή με παραμέτρους $n+m$, p. (Υπόδ. Χρησιμοποιήστε το γεγονός ότι οι X, Y εκφράζουν το αριθμό των επιτυχιών σε n και m δοκιμές αντίστοιχα).

Άσκ. 13. Έστω $X_1,X_2,\ldots$ μία ακολουθία από ανεξάρτητες τ.μ. που ακολουθούν την ίδια συνεχή κατανομή. Θα λέμε ότι ένα «ρεκόρ» συμβαίνει στο χρόνο n αν ισχύει ότι $X_n > \max\{X_1,X_2,\ldots,X_{n-1}\}$. Δηλαδή η τ.μ. X_n αποτελεί «ρεκόρ» αν είναι μεγαλύτερη από όλες τις προηγούμενες τ.μ. $X_1,X_2,\ldots,X_{n-1}$.
(i) Δείξτε ότι P(*παρατηρείται ρεκόρ στο χρόνο* n) $=1/n$.

(ii) Αν R_n είναι μία τ.μ. που εκφράζει το πλήθος των ρεκόρ έως και το χρόνο n, δείξτε ότι

$$E(R_n) = \sum_{i=1}^{n} \frac{1}{i} \quad \text{και} \quad V(R_n) = \sum_{i=1}^{n} \frac{i-1}{i^2}$$

Άσκ. 14. Ρίχνουμε δύο ζάρια 180 φορές και καταγράφουμε το άθροισμα των εδρών του. α) Να βρεθεί προσεγγιστικά η πιθανότητα το άθροισμα 9 να εμφανιστεί (i) τουλάχιστον 30 φορές, (ii) το πολύ 30 φορές, iii) από 20 έως 30 φορές. (β) Να βρεθεί προσεγγιστικά η πιθανότητα, το συνολικό άθροισμα των εδρών και των 180×2 ζαριών να είναι μεγαλύτερο από 1300.

Άσκ. 15. Σε μία μεγάλη πόλη θέλουμε να εκτιμήσουμε το ποσοστό των ψηφοφόρων ενός κόμματος Α και για το σκοπό αυτό ρωτήθηκαν 200 τυχαία επιλεγμένοι ψηφοφόροι. Ποια είναι (προσεγγιστικά) η πιθανότητα η πλειοψηφία των ψηφοφόρων του δείγματος να είναι υπέρ του κόμματος Α, ενώ στην πραγματικότητα μόλις το 45% των ψηφοφόρων της πόλης να είναι υπέρ του κόμματος Α.

Άσκ. 16. Το 10% από τα προϊόντα μιας μηχανής είναι ελαττωματικά. Υπολογίστε προσεγγιστικά την πιθανότητα ανάμεσα σε 400 τυχαία επιλεγμένα κομμάτια του προϊόντος να είναι ελαττωματικά (α) το πολύ 30, (β) 30 έως 50, (γ) 35 έως 45, (δ) 55 ή περισσότερα.

Άσκ. 17. Έστω ότι ο αριθμός των ελαττωμάτων μιας μαγνητικής ταινίας μήκους t μέτρων ακολουθεί την κατανομή Poisson με μέση τιμή $t/200$. Να υπολογιστεί κατά προσέγγιση η πιθανότητα όπως ο συνολικός αριθμός ελαττωμάτων 100 ταινιών μήκους 800 μέτρων η κάθε μία δεν υπερβαίνει τα 430.

Άσκ. 18. Το πλήθος των κλήσεων σε ένα τηλεφωνικό κέντρο κατά τη διάρκεια 1 λεπτού ακολουθεί την κατανομή Poisson με μέση τιμή λ = 1. Αν κάθε τηλεφώνημα αφήνει κέρδος στον ΟΤΕ 0.05 €, να βρεθεί προσεγγιστικά η πιθανότητα, το κέρδος του ΟΤΕ από τη συνεχή λειτουργία 10 ωρών να είναι τουλάχιστον 28 €.

Άσκ. 19. Μία μεγάλη εταιρία διαθέτει 5000 πελάτες. Έχει βρεθεί ότι η πιθανότητα να καλέσει κάποιος πελάτης την εταιρία για υποστήριξη (μία συγκεκριμένη ώρα της ημέρας) είναι 0.1%. Να βρεθεί ο μικρότερος αριθμός υπαλλήλων στο τμήμα υποστήριξης ώστε το λιγότερο το 90% των πελατών να εξυπηρετούνται άμεσα.

Άσκ. 20. Σε ένα πολυκατάστημα πραγματοποιήθηκε μετατροπή των τιμών των προϊόντων από δραχμές σε ευρώ. Μετά την μετατροπή έγινε και στρογγυλοποίηση «προς τα άνω» στο άνω πλησιέστερο δέκατο του ευρώ (δηλ. αν ένα προϊόν κόστιζε π.χ. 550δρχ = 1.61409 €, τώρα θα κοστίζει 1.7 €). Ένας πελάτης αγοράζει (τυχαία) 150 προϊόντα που πριν την μετατροπή κόστιζαν 100000δρχ. (=293.47 €). Θεωρούμε ότι οι «διαφορές» X_1, X_2, ..., X_n στις τιμές (σε ευρώ πριν και μετά την «προς τα άνω» στρογγυλοποίηση) των n = 100 προϊόντων είναι ανεξάρτητες τ.μ. η κάθε μία από τις οποίες ακολουθεί την ομοιόμορφη κατανομή στο διάστημα (α,β) (δηλαδή θεωρούμε ότι μία «διαφορά» X_i είναι το ίδιο πιθανό να πάρει οποιαδήποτε τιμή μεταξύ του α και του β). Αφού προσδιορίσετε με λογικά επιχειρήματα τα α και β, να βρείτε:
(i) το αναμενόμενο ποσό που θα κληθεί τώρα να πληρώσει ο πελάτης
(ii) την κατανομή του ποσού αυτού (προσεγγιστικά)
(iii) την πιθανότητα να πληρώσει τώρα περισσότερο από 300 €

Άσκ. 21. Ο χρόνος εξυπηρέτησης X ενός πελάτη από τον ταμία μιας τράπεζας είναι τ.μ. με μέση τιμή μ = 5 λεπτά και τυπική απόκλιση σ = 2 λεπτά. Αν στο χώρο αναμονής της τράπεζας περιμένουν να εξυπηρετηθούν n = 55 πελάτες, ποια είναι (προσεγγιστικά) η πιθανότητα να χρειαστεί ο ταμίας περισσότερες από 5 ώρες για να τους εξυπηρετήσει όλους (υποθέτουμε ότι οι πελάτες εξυπηρετούνται διαδοχικά ο ένας μετά τον άλλο, και οι χρόνοι εξυπηρέτησής τους είναι ανεξάρτητοι).

Άσκ. 22. Μία τυχαία μεταβλητή X έχει μέση τιμή 3 και διασπορά 2. Χρησιμοποιώντας την ανισότητα του Chebyshev υπολογίστε ένα άνω φράγμα για τις $P(|X-3|\geq 2)$, $P(|X-3|\geq 1)$.

Άσκ. 23. Έστω $X_1,X_2,...,X_{100}$ ανεξάρτητες τ.μ. που ακολουθούν την κατανομή Poisson με μέση τιμή 1. (α) Χρησιμοποιήστε την ανισότητα Markov για να βρείτε ένα φράγμα για την $P(X_1+ ...+X_{100} \geq 120)$. (β) Χρησιμοποιήστε το Κ.Ο.Θ. για να βρείτε μία προσέγγιση για την $P(X_1+...+X_{100} \geq 120)$.

ΚΕΦΑΛΑΙΟ 4

Άσκ. 24. Έστω $X_1,X_2,...,X_n$ ένα τ.δ. από έναν πληθυσμό με μέση τιμή μ και διασπορά σ^2. Να βρείτε το c για το οποίο η στατιστική συνάρτηση

$$S^2 = c\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2\right)$$

είναι αμερόληπτη εκτιμήτρια του σ^2.

Άσκ. 25. Έστω X_1, X_2,..., X_n και Y_1, Y_2,..., Y_k δύο ανεξάρτητα μεταξύ τους τ.δ., το πρώτο από μία κατανομή με μέση τιμή μ και διασπορά σ^2, και το δεύτερο από μία κατανομή με μέση τιμή $c\cdot\mu$ και διασπορά $d\cdot\sigma^2$ ($c \in R$, $d > 0$ γνωστά). Να βρεθούν να a και b έτσι ώστε η εκτιμήτρια της μορφής $a\overline{X} + b\overline{Y}$ να είναι η «καλύτερη δυνατή» αμερόληπτη εκτιμήτρια του μ.

Άσκ. 26. Έστω ένα τ.δ. $X_1,X_2,...,X_n$ από μία κατανομή $F(x;a,b)$. Να βρεθούν οι εκτιμήτριες ροπών των παραμέτρων a,b αν είναι γνωστό ότι $E(X_i)=a/b$, $Var(X_i)=a/b^2$.

Άσκ. 27. Έστω ένα τ.δ. $X_1,X_2,...,X_n$ από μία κατανομή $F(x;a,b)$. Να βρεθούν οι εκτιμήτριες ροπών των παραμέτρων a,b αν είναι γνωστό ότι $E(X_i)=a-b$, $Var(X_i) = 4ab$, $a,b>0$.

Άσκ. 28. Έστω ένα τ.δ. $X_1,X_2,...,X_n$ από την κατανομή G(α,λ) (Γάμμα με παραμέτρους (α, λ)). Αν έχει βρεθεί ότι, $\overline{X} = 4.52, \sum X_i^2 = 255.3$ (n=10), να εκτιμηθούν τα α, λ χρησιμοποιώντας τις εκτιμήτριες ροπών των α, λ (*υπενθυμίζεται ότι η μέση τιμή και η διασπορά της* G(α,λ) *είναι* α/λ *και* α/λ^2 *αντίστοιχα*).

Άσκ. 29. Έστω $X_1,X_2,...,X_n$ τ.δ. από κατανομή με σ.π.π.

$$f(x;\lambda) = \frac{\lambda^3}{2} x^2 e^{-\lambda x}, \quad x > 0 .$$

Βρείτε την ε.μ.π. του λ.

Άσκ. 30. Έστω $X_1,X_2,...,X_n$ τ.δ. από κατανομή με σ.π.π.

$$f(x;\lambda) = 2\lambda x e^{-\lambda x^2}, \quad x > 0 .$$

Βρείτε την ε.μ.π. του λ.

Άσκ. 31. Έστω $X_1,X_2,...,X_n$ τ.δ. από κατανομή με σ.π.π.

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}, \; x \geq 0 .$$

Βρείτε την ε.μ.π. του μ.

Άσκ. 32. Έστω ότι το πλήθος των προσπαθειών μέχρι και την επιτυχία ενός στόχου από έναν σκοπευτή ακολουθεί την γεωμετρική κατανομή με σ.π. $f(x;p) = (1-p)^{x-1} p, \quad x = 1,2,...$. Αν σε 5 αγώνες ο σκοπευτής πέτυχε για πρώτη φορά το στόχο στην 3η, 4η, 1η, 3η και 2η προσπάθεια αντίστοιχα, να εκτιμήσετε την πιθανότητα επιτυχίας p του στόχου από τον συγκεκριμένο σκοπευτή. Να χρησιμοποιήσετε (αφού πρώτα βρείτε) την εκτιμήτρια μέγιστης πιθανοφάνειας του p.

ΚΕΦΑΛΑΙΟ 5

Άσκ. 33. Έστω ότι επιθυμούμε να εκτιμήσουμε το μέσο χρόνο που κάνει ένα τρένο του Μετρό για να μεταβεί από το σταθμό Α στο σταθμό Β. Χρονομετρώντας τη διαδρομή αυτή 5 φορές σημειώνουμε τους χρόνους (σε seconds)

137, 151, 157, 152, 160.

α) Να δοθεί ένα δ.ε. συντελεστού 99% για το μέσο χρόνο μετάβασης. β) Να δοθεί ένα δ.ε. συντελεστού 99% για την τυπική απόκλιση του χρόνου μετάβασης. (Υποθέστε ότι οι χρόνοι είναι κανονικοί).

Άσκ. 34. Έστω ότι οι αποδοχές των εργαζομένων σε μία μεγάλη επιχείρηση ακολουθούν την κανονική κατανομή με μέση τιμή μ και διασπορά σ^2. Οι μηνιαίες αποδοχές (σε ευρώ) 7 τυχαία επιλεγμένων εργαζομένων είναι

1098, 1120, 1102, 1052, 1032, 977, 890

(i) Να κατασκευάσετε ένα διάστημα εμπιστοσύνης συντελεστού 95% για τον μέσο μισθό μ όλων των εργαζομένων στην επιχείρηση. (ii) Ποιό θα ήταν το αντίστοιχο δ.ε. αν ήταν γνωστό ότι σ = 70; Πόσο δείγμα πρέπει να πάρουμε σε αυτή την περίπτωση για να κατασκευάσουμε δ.ε. με εύρος 20 ευρώ;

Άσκ. 35. Έστω ότι θέλουμε να εκτιμήσουμε την μέση τιμή πώλησης μ ενός καφέ (συγκεκριμένου τύπου) που σερβίρεται στις καφετέριες του λεκανοπεδίου των Αθηνών. Για το σκοπό αυτό καταγράφηκαν (σε ευρώ) οι αντίστοιχες τιμές X_1, X_2, ..., X_n σε n = 50 τυχαία επιλεγμένες καφετέριες. Βρέθηκε ότι,

$$\sum_{i=1}^{n} X_i = 125, \qquad \sum_{i=1}^{n} X_i^2 = 325.$$

α) Να βρείτε ένα (προσεγγιστικό) δ.ε. συντελεστού 95% για τη μέση τιμή πώλησης μ και β) Να βρείτε ένα (προσεγγιστικό) δ.ε. συντελεστού 95% για την τυπική απόκλιση των X_i.

Άσκ. 36. Σε μία έρευνα αγοράς επιθυμούμε να εκτιμήσουμε το ποσοστό p των καταναλωτών μιας περιοχής που χρησιμοποιούν ένα συγκεκριμένο προϊόν. Για το σκοπό αυτό επιλέχθηκαν τυχαία n = 100 καταναλωτές από τους οποίους 20 απάντησαν ότι χρησιμοποιούν το προϊόν (ενώ οι υπόλοιποι 80 δεν το χρησιμοποιούν). (α) Να δώσετε ένα δ.ε. για το πληθυσμιακό ποσοστό p με συντελεστή εμπιστοσύνης 95%. (β) Πόσο περίπου παραπάνω δείγμα πρέπει να επιλέξουμε για να πάρουμε δ.ε. συντελεστού εμπιστοσύνης 95% και εύρους 5%.

Άσκ. 37. Σε μία έρευνα απασχολήσεως επιθυμούμε να προσδιορίσουμε το ποσοστό των ανέργων p μιας πόλης. Το αντίστοιχο ποσοστό σε ένα τυχαίο δείγμα n = 100 ατόμων βρέθηκε ίσο με 11%. α) Μεταξύ ποιων ορίων βρίσκεται το πραγματικό ποσοστό p με συντελεστή εμπιστοσύνης 95%; β) Αν η πόλη έχει εργατικό δυναμικό m=100000 άτομα, να δώσετε δ.ε. 95% για το πλήθος των ανέργων στην πόλη. γ) Πόσο περίπου παραπάνω δείγμα πρέπει να πάρουμε για να έχουμε δ.ε. 95% εύρους 3%.

Άσκ. 38. Μετά από δειγματοληπτική έρευνα σε 4 καταστήματα μιας περιοχής Α και 3 καταστήματα μιας περιοχής Β βρέθηκε ότι το ίδιο προϊόν κοστίζει (σε ευρώ)

5.1, 2.8, 2.2, 3.4 (περιοχή Α) και 3.5, 2.9, 3.2 (περιοχή Β)

αντίστοιχα. Υποθέτοντας ότι οι τιμές κατανέμονται κανονικά (N(μ_1, σ_1^2) και N(μ_2, σ_2^2) αντίστοιχα), βρείτε ένα διάστημα εμπιστοσύνης συντελεστού 95% για το πηλίκο σ_2^2/σ_1^2. Μπορούμε από αυτό το δ.ε. να πούμε (σε ε.σ. 5%) ότι οι τιμές του προϊόντος παρουσιάζουν διαφορετική μεταβλητότητα στις δύο περιοχές;

ΚΕΦΑΛΑΙΟ 6

Άσκ. 39. Από κανονικό πληθυσμό N(μ, σ^2 = 4) λαμβάνεται τυχαίο δείγμα μεγέθους n = 16, προκειμένου να ελεγχθεί η υπόθεση H_0: μ = 10 έναντι της H_1: μ > 10 σε επίπεδο σημαντικότητας α = 0.05. i) Να προσδιοριστεί η περιοχή απόρριψης. ii) Αν ίσχυε ότι μ = 11 με τι πιθανότητα θα παίρναμε σωστή απόφαση;

Άσκ. 40. Θέλοντας να εκτιμήσουμε τη μέση τιμή μ του λίτρου της βενζίνης στα πρατήρια των Αθηνών επισκεφτήκαμε τυχαία n=10 βενζινάδικα από όπου καταγράψαμε τις τιμές (σε δρχ):

280.3, 282.8, 278.5, 283.1, 290.0, 284.9, 284.4, 279.8, 291.1, 286.7 .

Να ελέγξετε σε επίπεδο σημαντικότητας 0.05 την υπόθεση μ=290 έναντι της μ<290 α) στην περίπτωση που το σ είναι άγνωστο και β) στην περίπτωση που γνωρίζουμε ότι σ=4.

Άσκ. 41. Έστω ότι επιθυμούμε να εκτιμήσουμε το μέσο χρόνο μ που κάνει ένα τρένο του Μετρό για να μεταβεί από το σταθμό Α στο σταθμό Β. Χρονομετρώντας τη διαδρομή αυτή 10 φορές σημειώνουμε τους χρόνους (σε seconds)

357, 337, 351, 357, 350, 352, 360, 353, 377, 372 .

α) Να ελέγξετε σε επίπεδο σημαντικότητας 0.01 την υπόθεση μ=360 έναντι της μ≠360.
β) Να ελέγξετε σε επίπεδο σημαντικότητας 0.01 την υπόθεση σ=30 έναντι της σ<30.

Άσκ. 42. Έστω ότι θέλουμε να εκτιμήσουμε το μέσο χρόνο ζωής των αρρένων κατοίκων μιας συγκεκριμένης περιοχής. Για το σκοπό αυτό ελήφθη τ.δ. μεγέθους n=200 ανδρών. Βρέθηκε ότι,

$$\sum_{i=1}^{200} X_i = 13959.6, \qquad \sum_{i=1}^{200} X_i^2 = 977265.$$

Υποθέτοντας ότι οι χρόνοι ζωής είναι κανονικοί,
α) Να ελέγξετε σε επίπεδο σημαντικότητας 0.05 την υπόθεση μ=71 έναντι της μ<71.
β) Να ελέγξετε σε επίπεδο σημαντικότητας 0.05 την υπόθεση σ=4 έναντι της σ≠4.

Άσκ. 43. Έστω Y_1, Y_2, …,Y_n οι τιμές κλεισίματος μιας μετοχής Α σε ένα χρηματιστήριο n διαδοχικές ημέρες. Υποθέτουμε ότι ισχύει $Y_i = Y_{i-1}e^{X_i}$ για i = 1,2,…,n όπου X_1,X_2,...,X_n είναι ανεξάρτητες τ.μ. που ακολουθούν την N(μ,σ^2) και Y_0 = 100. Αν σε n = 5 διαδοχικές ημέρες έχουν καταγραφεί οι τιμές (Y_1 έως Y_5)

96.64, 87.78, 82.05, 88.15, 90.94

(α) να ελέγξετε αν μ = 0 (σε ε.σ. 5%) έναντι της μ < 0. (β) Να κάνετε τον ίδιο έλεγχο γνωρίζοντας ότι σ = 0.1. Σε αυτή την περίπτωση, να βρεθεί το p-value και η συνάρτηση ισχύος. Αν ίσχυε ότι μ = –0.1 με τι πιθανότητα θα παίρναμε σωστή απόφαση; (*Υπόδειξη: από τα παραπάνω Y_i να προσδιορίσετε πρώτα τα αντίστοιχα X_i*).

Άσκ. 44. Έστω X μία τ.μ. από την N(μ,σ^2 = 9). Επιθυμούμε με βάση αυτή να ελέγξουμε την υπόθεση

H_0: μ = 10 έναντι της H_1: μ = 20.

(i) Βρείτε, χρησιμοποιώντας το λήμμα Neymann - Pearson τον βέλτιστο έλεγχο για την παραπάνω υπόθεση σε επίπεδο σημαντικότητας α. (ii) Ποια είναι η ισχύς του ελέγχου; (iii) Εάν X = 14 δεχόμαστε ή απορρίπτουμε την H_0; (ε.σ. α=0.05).

Άσκ. 45. Έστω ότι ένα μεγάλο κόμμα θέλει να εκτιμήσει το ποσοστό p των ψηφοφόρων μιας μεγάλης πόλης που προτίθενται να το ψηφίσουν στις επερχόμενες βουλευτικές εκλογές. Το αντίστοιχο ποσοστό σε ένα τυχαίο δείγμα n=500 ψηφοφόρων βρέθηκε ίσο με 40%. Να ελέγξετε σε επίπεδο σημαντικότητας 0.05 την υπόθεση $p \le 35\%$ έναντι της $p > 35\%$.

Άσκ. 46. Ένας υποψήφιος δήμαρχος ισχυρίζεται ότι στις εκλογές που έγιναν σήμερα θα εκλεγεί από τον πρώτο γύρο (δηλ. θα λάβει ποσοστό μεγαλύτερο του 50%). Έχουμε αρκετά στοιχεία για να απορρίψουμε τον παραπάνω ισχυρισμό, αν σε ένα τ.δ. n=300 ψηφοφόρων (που ελήφθη με exit poll, δηλαδή οι ψηφοφόροι ρωτήθηκαν αμέσως μετά την έξοδό τους από τυχαία επιλεγμένα εκλογικά τμήματα) μόνο το 44% δήλωσαν ότι ψήφισαν τον συγκεκριμένο υποψήφιο (α=0.01). Αν βρεθεί ότι δεν έχουμε αρκετά στοιχεία, πόσο θα έπρεπε να είναι το ελάχιστο μέγεθος του δείγματος ώστε, με το ίδιο δειγματικό ποσοστό, να απορρίπταμε τον παραπάνω ισχυρισμό;

Άσκ. 47. Έστω μ_1 και μ_2 οι μέσοι χρόνοι εξυπηρέτησης των πελατών από δύο ταμίες μιας τράπεζας. Αν

234, 99, 234, 174, 188, 107, 173, 172 και 105, 194, 77, 33, 159, 150, 167, 127, 169, 166

είναι δειγματοληπτικά κάποιοι χρόνοι (σε sec) εξυπηρέτησης των δύο αυτών υπαλλήλων αντίστοιχα, μπορούμε σε επίπεδο σημαντικότητας α=0.05 α) να πούμε ότι οι δύο υπάλληλοι έχουν διαφορετική απόδοση; β) να πούμε ότι ο πρώτος υπάλληλος έχει μεγαλύτερη απόδοση από το δεύτερο; (είναι γνωστό ότι $\sigma_1=\sigma_2=40$)

Άσκ. 48. Έστω μ_1 η μέση τιμή πώλησης ενός προϊόντος σε μία περιοχή Α και μ_2 η μέση τιμή πώλησης του ίδιου προϊόντος σε μία περιοχή Β. Υποθέτουμε ότι οι τιμές κατανέμονται κανονικά και με ίση διασπορά στις δύο περιοχές. Επιλέγουμε τυχαία 3 τιμές από την περιοχή Α και 3 τιμές από την περιοχή Β. Αν οι τιμές αυτές είναι

112, 106, 121 (περιοχή Α) και 107, 93, 97 (περιοχή Β),

μπορούμε, σε επίπεδο σημαντικότητας 5%, να πούμε ότι η μέση τιμή πώλησης στην περιοχή Α είναι υψηλότερη από την αντίστοιχη στην περιοχή Β; (H_0: $\mu_1 = \mu_2$, H_1:$\mu_1 > \mu_2$)

Άσκ. 49. Έστω μ_1 η μέση τιμή πώλησης ενός προϊόντος σε μία περιοχή Α και μ_2 η μέση τιμή πώλησης του ίδιου προϊόντος σε μία περιοχή Β. Η μέση τιμή και η διασπορά ενός τ.δ. 10 τιμών πώλησης από την περιοχή Α βρέθηκε 100.9 και 8.76667 αντίστοιχα. Επίσης, η μέση τιμή και η διασπορά ενός τ.δ. 20 τιμών πώλησης από την περιοχή Β βρέθηκε 104.45 και 12.9974 αντίστοιχα. Αν υποθέσουμε ότι οι τιμές κατανέμονται κανονικά και με ίση (αλλά άγνωστη) διασπορά και στις δύο περιοχές, να ελέγξετε σε επίπεδο σημαντικότητας 5%
α) αν η μέση τιμή πώλησης στην Α είναι διαφορετική από την αντίστοιχη στην Β.
β) αν η μέση τιμή πώλησης στην Α είναι χαμηλότερη από την αντίστοιχη στην Β.

Άσκ. 50. Στην προηγούμενη άσκηση που αφορούσε τις τιμές πώλησης ενός προϊόντος σε δύο περιοχές Α και Β υποθέσαμε ότι οι διασπορές σ_1^2 και σ_2^2 των τιμών στις περιοχές αυτές είναι ίσες. Να ελέγξετε σε επίπεδο σημαντικότητας 5% αν όντως οι δύο αυτές διασπορές είναι ίσες.

Άσκ. 51. Έστω ότι οι χρόνοι ζωής δύο τύπων εξαρτημάτων Α, Β ακολουθούν κανονική κατανομή $\mathrm{N}(\mu_1, \sigma_1^2)$ και $\mathrm{N}(\mu_2, \sigma_2^2)$ αντίστοιχα. Αν σε ένα τ.δ. από εξαρτήματα τύπου Α βρέθηκε ότι S_1=100 (n_1=50) και σε ένα τ.δ. από εξαρτήματα τύπου Β βρέθηκε ότι S_2=120 (n_2=50), μπορούμε σε ε.σ. 5% να απορρίψουμε ότι οι χρόνοι ζωής των εξαρτημάτων παρουσιάζουν όμοια μεταβλητότητα; (H_0: $\sigma_1=\sigma_2$ έναντι της H_1: $\sigma_1 < \sigma_2$).

Άσκ. 52. Από τα 400 εξαρτήματα που παίρνουμε στην τύχη από μία μηχανή που τα κατασκευάζει, τα 16 είναι ελαττωματικά, ενώ από τα 300 μιας άλλης μηχανής, τα 24 βρέθηκαν ελαττωματικά. Να ελέγξετε σε επίπεδο σημαντικότητας 1% α) αν υπάρχει διαφορά στην παραγωγή ελαττωματικών μεταξύ των δύο μηχανών και β) αν η δεύτερη μηχανή είναι χειρότερη από την πρώτη.

Άσκ. 53. Βρέθηκε ότι 78 από 200 τυχαία επιλεγμένους ψηφοφόρους μιας μεγάλης πόλης Α προτίθενται να ψηφίσουν ένα συγκεκριμένο κόμμα, ενώ 240 από 500 τυχαία επιλεγμένους ψηφοφόρους μιας άλλης μεγάλης πόλης Β προτίθενται να ψηφίσουν το ίδιο κόμμα. Μπορούμε, να πούμε ότι το ποσοστό στην πόλη Β είναι «σημαντικά» μεγαλύτερο του ποσοστού στην πόλη Α σε επίπεδο σημαντικότητας 0.05;

Άσκ. 54. Ένας δημοσιογράφος σε μία τηλεοπτική συζήτηση ισχυρίζεται ότι συμβαίνει κάτι ύποπτο στην εταιρία μέτρησης τηλεθέασης AGB. Η εταιρία έχει διαθέσει 1150 μηχανάκια σε ισάριθμους τηλεθεατές, 150 από τα οποία έχουν δοθεί σε νέους συνεργάτες (τηλεθεατές που έχουν ξεκινήσει τη συνεργασία τους με την AGB το τελευταίο εξάμηνο). Ο δημοσιογράφος διαπιστώνει ότι, σύμφωνα με την AGB, το κανάλι ALPHA παρουσιάζει τηλεθέαση 30% στο σύνολο (στους 1000+150 τηλεθεατές) ενώ έχει μόνο 22% στους νέους (στους 150) τηλεθεατές. Ο δημοσιογράφος το θεωρεί αυτό πολύ ύποπτο και κάποιος πολιτικός μάλιστα δηλώνει ότι «οι δύο αυτές τιμές δεν συμφωνούν καθόλου» (πρόκειται για πραγματικό περιστατικό). Είναι πράγματι αυτό «πολύ ύποπτο»; Θα μπορούσε δηλαδή η διαφορά αυτή να είναι τυχαία; Συγκεκριμένα, αν p_1, p_2 είναι τα ποσοστά της τηλεθέασης του ALPHA στους πληθυσμούς από τους οποίους προέρχονται οι δύο ομάδες τηλεθεατών («νέοι» και «παλαιοί» συνεργάτες), ελέγξτε σε ε.σ. α = 0.01 την υπόθεση H_0: $p_1 = p_2$ έναντι της H_1: $p_1 > p_2$ (αν απορρίψουμε την H_0 μπορούμε πράγματι να μιλήσουμε για κάτι «ύποπτο», διότι το λογικό είναι να ισχύει $p_1 = p_2$ αφού και οι δύο ομάδες συνεργατών προέρχονται από τον ίδιο πληθυσμό).

Λύσεις Συμπληρωματικών Ασκήσεων Στατιστικής ΙΙΙ

ΚΕΦΑΛΑΙΑ 1-3

Λύση Ασκ. 1. Έστω $X \sim$ N(60, 5^2) το βάρος του τυχαία επιλεγμένου μαθητή (άρα $(X-60)/5 \sim$ N(0,1)). Ζητείται η πιθανότητα

$$P(X > 70) = P(\frac{X-60}{5} > \frac{70-60}{5}) = 1 - \Phi(\frac{70-60}{5}) = 1 - \Phi(2) = 1 - 0.9772 = 0.0228 \,.$$

Λύση Άσκ. 2. Αν X είναι η τ.μ. που εκφράζει το βάρος των ενός χαλικιού σε gr τότε $Z = \frac{X-30}{10} \sim N(0,1)$

i) α) $P(X > 30) = P(\frac{X-30}{10} > \frac{30-30}{10}) = 1 - P(Z \le 0) = 1 - \Phi(0) = 0.5$

β) $P(28 < X < 40) = P(\frac{28-30}{10} < \frac{X-30}{10} < \frac{40-30}{10}) = P(-0.2 < Z < 1)$

$$= \Phi(1) - \Phi(-0.2) = \Phi(1) - 1 + \Phi(0.2) = 0.8413 - 1 + 0.5793 = 0.4206$$

γ) Έστω $X_1, X_2, ..., X_{10}$ το βάρος των 10 χαλικιών. Είναι γνωστό ότι το άθροισμα κανονικών τ.μ. ακολουθεί και πάλι κανονική τ.μ. Επομένως, αν

$$W = \sum_{i=1}^{10} X_i$$

είναι το συνολικό βάρος των 10 χαλικιών τότε η W ακολουθεί κανονική κατανομή με

$$E(W) = E(\sum_{i=1}^{10} X_i) = \sum_{i=1}^{10} E(X_i) = 10E(X_1) = 10 \cdot 30 = 300 \quad \text{και} \quad Var(W) = Var(\sum_{i=1}^{10} X_i) = \sum_{i=1}^{10} Var(X_i) = 10Var(X_1) = 10 \cdot 10^2 = 1000$$

(οι τ.μ. $X_1, X_2, ..., X_{10}$ είναι ανεξάρτητες τ.μ.). Άρα $W \sim N(300,1000)$ και συνεπώς η τ.μ.

$$Y = \frac{W-300}{\sqrt{1000}} \sim N(0,1) \,.$$

Ζητείται η πιθανότητα

$$P(W > 300) = P(\frac{W-300}{\sqrt{1000}} > \frac{300-300}{\sqrt{1000}}) = P(Y > 0) = 1 - \Phi(0) = 0.5$$

ii) Αν συμβολίσουμε με U την τ.μ. που εκφράζει το πλήθος των χαλικιών στο δείγμα με βάρος $28 < X < 40$gr, τότε $Y \sim B(4, p = P(28 < X < 40) = 0.4206)$. Άρα τελικά,

$$P(Y = 2) = \binom{4}{2} p^2 (1-p)^{4-2} = 6 \cdot (0.4206)^2 \cdot (0.5794)^2 = 0.3563 \,.$$

Λύση Άσκ. 3. Έστω $X \sim$ N(500,100^2) ο βαθμός του μαθητή. Θα είναι $Z=(X-500)/100 \sim$ N(0,1). Έστω επίσης c ο μεγαλύτερος βαθμός που μπορεί να έχει, ώστε να βρίσκεται στο 20% της μικρότερης βαθμολογίας της κατανομής. Για το c θα πρέπει να ισχύει ότι,

$$P(X \le c) = 0.2 \Leftrightarrow P(\frac{X-500}{100} \le \frac{c-500}{100}) = 0.2 \;\Rightarrow \Phi(\frac{c-500}{100}) = 0.2 = 1 - 0.8 = 1 - \Phi(0.84) = \Phi(-0.84)$$

$$\Rightarrow \frac{c-500}{100} = -0.84 \Rightarrow c = 500 - 84 = 416 \,.$$

Λύση Άσκ. 4. (α) Έστω X ο χρόνος αναμονής ενός πελάτη στο ιατρείο. Γνωρίζουμε ότι $X \sim$ N($\mu,\sigma^2=3.75^2$) και συνεπώς $(X-\mu)/3.75 \sim$ N(0,1). Επίσης, γνωρίζουμε ότι

$$P(X > 20) = 0.0239 \Rightarrow P(\frac{X-\mu}{3.75} > \frac{20-\mu}{3.75}) = 0.0239 \Rightarrow 1 - \Phi(\frac{20-\mu}{3.75}) = 0.0239 \;\Rightarrow \Phi(\frac{20-\mu}{3.75}) = 0.9761 \approx \Phi(1.98)$$

Συνεπώς, (η Φ είναι συνάρτηση γνήσια αύξουσα και άρα 1-1)

$$\frac{20-\mu}{3.75} \approx 1.98 \Rightarrow \mu \approx 20 - 1.98 \cdot 3.75 \approx 12.575 \,.$$

Δηλαδή, $X \sim$ N(12.575, 3.75^2).

(β) Η πιθανότητα, ένας ασθενής να περιμένει στο ιατρείο για χρόνο μεταξύ 10 και 15 λεπτών, είναι

$$P(10 < X < 15) = P(\frac{10-12.575}{3.75} < \frac{X-12.575}{3.75} < \frac{15-12.575}{3.75}) = \Phi(\frac{15-12.575}{3.75}) - \Phi(\frac{10-12.575}{3.75})$$
$$\approx \Phi(0,6466) - \Phi(-0.6866) = \Phi(0,6466) + \Phi(0.6866) - 1 \approx 0.741 + 0.754 - 1 = 0,495$$

Λύση Άσκ. 5. Γνωρίζουμε ότι $X \sim \mathrm{N}(15.75, 0.7^2)$ και συνεπώς η τ.μ. $Z = (X-15.75)/0.7 \sim \mathrm{N}(0,1)$. Αρχικά θα υπολογίσουμε τι πιθανότητες

$$P(X \in \mathrm{S}) = P(X \le 15) = P(\frac{X-15.75}{0.7} \le \frac{15-15.75}{0.7}) = P(Z \le -1.071) = \Phi(-1.071) = 1 - \Phi(1.071) \approx 1 - 0.8577 = \mathbf{0.1423},$$

$$P(X \in \mathrm{M}) = P(15 < X \le 16) = P(\frac{15-15.75}{0.7} < \frac{X-15.75}{0.7} \le \frac{16-15.75}{0.7}) = P(-1.071 < Z \le 0.3571)$$
$$= \Phi(0.3571) - \Phi(-1.071) = \Phi(0.3571) + \Phi(1.071) - 1 \approx 0.64 + 0.8577 - 1 = \mathbf{0.4977},$$

$$P(X \in \mathrm{L}) = P(16 < X \le 17) = P(\frac{16-15.75}{0.7} < \frac{X-15.75}{0.7} \le \frac{17-15.75}{0.7}) = P(0.3571 < Z \le 1.786)$$
$$= \Phi(1.786) - \Phi(0.3571) \approx 0.963 - 0.64 = \mathbf{0.323},$$

$$P(X \in \mathrm{XL}) = P(X > 17) = P(\frac{X-15.75}{0.7} > \frac{17-15.75}{0.7}) = P(Z > 1.786) = 1 - \Phi(1.786) = 1 - 0.963 \approx \mathbf{0.037}.$$

Θεωρούμε ότι τα 10000 πουκάμισα θα αγοραστούν από ισάριθμους πελάτες (άνδρες). Ας δούμε πόσοι περίπου από αυτούς θα χρειαστούν πουκάμισο κατηγορίας S. Έστω Y το πλήθος των πελατών από τους n που θα χρειαστούν πουκάμισο κατηγορίας S. Η τ.μ. Y μπορεί να θεωρηθεί ότι εκφράζει το πλήθος των επιτυχιών σε n ανεξάρτητες δομικές (n πελάτες) με πιθανότητα επιτυχίας $p = P(X \in \mathrm{S})$ (επιτυχία: «ο πελάτης χρειάζεται πουκάμισο S»). Επομένως, $Y \sim B(n,p)$ και άρα $E(Y) = np$. Συνεπώς, η αναμενόμενη ζήτηση πουκαμίσων κατηγορίας S θα είναι $nP(X \in \mathrm{S}) = 10000 \cdot 0.1423 = 1423$ πουκάμισα. Όμοια, η αναμενόμενη ζήτηση πουκαμίσων κατηγορίας M, L, XL θα είναι αντίστοιχα,

$$nP(X \in \mathrm{M}) = 10000 \cdot 0.4977 = 4977, \quad nP(X \in \mathrm{L}) = 10000 \cdot 0.323 = 3230, \quad nP(X \in \mathrm{XL}) = 10000 \cdot 0.037 = 370 \text{ πουκάμισα.}$$

Άρα, για να καλυφθεί η (αναμενόμενη) ζήτηση θα πρέπει να κατασκευαστούν 1423 S, 4977 M, 3230 L και 370 XL πουκάμισα.

Εναλλακτικά, θα μπορούσαμε να θεωρήσουμε δίτιμες τ.μ. Y_1, Y_2, ..., Y_n, έτσι ώστε Y_i=1 αν ο i-πελάτης χρειάζεται πουκάμισο S και Y_i=0 διαφορετικά. Προφανώς, $Y = Y_1+Y_2+...+Y_n$. Από το νόμο των μεγάλων αριθμών, γνωρίζουμε τότε ότι

$$\frac{Y}{n} = \frac{1}{n}\sum_{i=1}^{n} Y_i \approx E(Y_i) = 1 \cdot P(Y_i = 1) + 0P(Y_i = 0) = P(Y_i = 1) = P(X \in S)$$

και άρα η ζήτηση Y πουκαμίσων κατηγορίας S θα είναι περίπου ίση με $Y \approx nP(X \in S) = 1423$ πουκάμισα. Όμοια εργαζόμαστε και για τις υπόλοιπες κατηγορίες M, L, XL.

(β) Έστω a_1, a_2, a_3 τα όρια του μήκους της περιφέρειας του λαιμού των ανδρών για τις 4 κατηγορίες, δηλαδή S=$(0,a_1]$, M=$(a_1,a_2]$, L=$(a_2,a_3]$, XL=(a_3,∞). Τα a_1, a_2, a_3 θα πρέπει να είναι τέτοια ώστε

$$P(X \in \mathrm{S}) = P(X \le a_1) = 0.25 \Rightarrow P(\frac{X-15.75}{0.7} \le \frac{a_1-15.75}{0.7}) = 0.25 \Rightarrow P(Z \le \frac{a_1-15.75}{0.7}) = 0.25$$

και επειδή $0.25 = 1 - 0.75 = 1 - \Phi(0.675) = \Phi(-0.675)$ θα πρέπει

$$\Phi(\frac{a_1-15.75}{0.7}) = \Phi(-0.675) \Rightarrow \frac{a_1-15.75}{0.7} = -0.675 \Rightarrow a_1 = 15.75 - 0.675 \cdot 0.7 = 15.2775$$

Όμοια, θα πρέπει

$$P(X \in \mathrm{M}) = P(a_1 < X \le a_2) = 0.25 \Rightarrow P(\frac{a_1-15.75}{0.7} < \frac{X-15.75}{0.7} \le \frac{a_2-15.75}{0.7}) = 0.25$$
$$\Rightarrow \Phi(\frac{a_2-15.75}{0.7}) - \Phi(\frac{a_1-15.75}{0.7}) = 0.25 \Rightarrow \Phi(\frac{a_2-15.75}{0.7}) = 0.5 = \Phi(0) \Rightarrow \frac{a_2-15.75}{0.7} = 0 \Rightarrow a_2 = 15.75$$

και

$$P(X \in \mathrm{L}) = P(a_2 < X \le a_3) = 0.25 \Rightarrow P(\frac{a_2-15.75}{0.7} < \frac{X-15.75}{0.7} \le \frac{a_3-15.75}{0.7}) = 0.25$$
$$\Rightarrow \Phi(\frac{a_3-15.75}{0.7}) - \Phi(\frac{a_2-15.75}{0.7}) = 0.25 \Rightarrow \Phi(\frac{a_3-15.75}{0.7}) = 0.75 = \Phi(0.675) \Rightarrow a_3 = 16.2225\,.$$

Λύση Άσκ. 6. Έστω $X \sim N(5000, 300^2)$ ή ζήτηση κρέατος σε κιλά στην ΚΑΑ την ημέρα που εξετάζουμε (άρα $(X–5000)/300 \sim N(0,1)$). (α) Η πιθανότητα να καλυφθεί όλη η ζήτηση θα είναι

$$P(X \le 5300) = P(\frac{X-5000}{300} \le \frac{5300-5000}{300}) = \Phi\left(\frac{5300-5000}{300}\right) = \Phi(1) = 0.8413 .$$

(β) Για την ποσότητα q θα πρέπει να ισχύει ότι

$$P(X \le q) = 0.9 \Leftrightarrow P(\frac{X-5000}{300} \le \frac{q-5000}{300}) = 0.9 \Leftrightarrow \Phi(\frac{q-5000}{300}) = 0.9 \approx \Phi(1.28)$$

Συνεπώς, θα πρέπει (η Φ είναι συνάρτηση γνήσια αύξουσα και άρα 1-1),

$$\frac{q-5000}{300} \approx 1.28 \Leftrightarrow q \approx 5000 + 300 \cdot 1.28 = 5384 \text{ kgr}.$$

Λύση Άσκ. 7. Έστω ότι $X \sim N(\mu,\sigma^2)$. Σύμφωνα με την εκφώνηση θα πρέπει

$$P(X \le 67.2) = 0.0808 \Rightarrow P(\frac{X-\mu}{\sigma} \le \frac{67.2-\mu}{\sigma}) = 0.0808 \Rightarrow \Phi(\frac{67.2-\mu}{\sigma}) = 0.0808 = 1-0.9192$$

$$\Rightarrow \Phi(\frac{67.2-\mu}{\sigma}) = 1-\Phi(1.4) = \Phi(-1.4) \Rightarrow \frac{67.2-\mu}{\sigma} = -1.4$$

και

$$P(X > 72.2) = 0.1357 \Rightarrow P(\frac{X-\mu}{\sigma} > \frac{72.2-\mu}{\sigma}) = 0.1357 \Rightarrow 1-\Phi(\frac{72.2-\mu}{\sigma}) = 0.1357$$

$$\Rightarrow \Phi(\frac{72.2-\mu}{\sigma}) = 0.8643 = \Phi(1.1) \Rightarrow \frac{72.2-\mu}{\sigma} = 1.1 .$$

Λύνοντας τις παραπάνω δύο εξισώσεις ως προς μ και σ βρίσκουμε ότι $\mu=70$, $\sigma=2$.

Λύση Άσκ. 8. Έστω A, B τα ενδεχόμενα να είναι εφοδιασμένος ο κινητήρας με ανταλλακτικό τύπου A ή B αντίστοιχα. Ζητείται η πιθανότητα $P(B \mid X < 40)$. Από το Θεώρημα Bayes προκύπτει ότι

$$P(B \mid X < 40) = \frac{P(X < 40 \mid B)P(B)}{P(X < 40 \mid A)P(A) + P(X < 40 \mid B)P(B)}$$

Αλλά,

$$P(X < 40 \mid B) = P(X < 40 \mid X \sim N(30,5^2)) = P(\frac{X-30}{5} < \frac{40-30}{5} \mid X \sim N(30,5^2)) = \Phi(\frac{40-30}{5}) = \Phi(2) = 0.9772$$

και

$$P(X < 40 \mid A) = P(X < 40 \mid X \sim N(50,5^2)) = P(\frac{X-50}{5} < \frac{40-50}{5} \mid X \sim N(30,5^2))$$

$$= \Phi(\frac{40-50}{5}) = \Phi(-2) = 1-\Phi(2) = 1-0.9772 = 0.0228$$

ενώ από την εκφώνηση γνωρίζουμε ότι $P(B) = 0.25$ και άρα $P(A) = 1-P(B) = 0.75$. Επομένως τελικά θα είναι

$$P(B \mid X < 40) = \frac{0.9772 \cdot 0.25}{0.0228 \cdot 0.75 + 0.9772 \cdot 0.25} = \frac{0.2443}{0.0171 + 0.2443} = \frac{0.2443}{0.0171 + 0.2443} = 0.9346 .$$

Λύση Άσκ. 9. Έστω $X_1,X_2,...,X_6$ οι τ.μ. που εκφράζουν τα βάρη των 6 ατόμων που επιλέχθηκαν τυχαία από τον πληθυσμό. Ζητείται η πιθανότητα το συνολικό τους βάρος να υπερβαίνει τα 500kgr, δηλαδή η

$$P(\sum_{i=1}^{6} X_i > 500) .$$

Γνωρίζουμε ότι το άθροισμα ανεξάρτητων κανονικών τ.μ. ακολουθεί και αυτό κανονική κατανομή. Συνεπώς η τ.μ. $X_1+X_2+...+X_6$ ακολουθεί κανονική κατανομή N(486, 600) διότι,

$$E(\sum_{i=1}^{6} X_i) = \sum_{i=1}^{6} E(X_i) = 6 \cdot 81 = 486 , \qquad V(\sum_{i=1}^{6} X_i) = \sum_{i=1}^{6} V(X_i) = 6 \cdot 100 = 600 .$$

Άρα (από γνωστή πρόταση) η τ.μ.

$$Z = \frac{\sum_{i=1}^{6} X_i - 486}{\sqrt{600}}$$

θα ακολουθεί τυπική κανονική κατανομή N(0,1). Επομένως τελικά, η πιθανότητα που ζητείται θα είναι

$$P(\sum_{i=1}^{6} X_i > 500) = P(\frac{\sum_{i=1}^{6} X_i - 486}{\sqrt{600}} > \frac{500-486}{\sqrt{600}}) \approx P(Z > 0.5715) \; = 1-\Phi(0.5715) \approx 1-0.7157 = 0.2843\,.$$

Λύση Άσκ. 10. Λόγω του ότι X, Y είναι ανεξάρτητες, θα ισχύει ότι $E(XY) = E(X)E(Y)$. Επίσης, οι τ.μ. X^2, Y^2 θα είναι και αυτές ανεξάρτητες και άρα $E(X^2Y^2) = E(X^2)E(Y^2)$. Επομένως,

$$\begin{aligned} Var(XY) &= E(X^2Y^2) - E(XY)^2 = E(X^2)E(Y^2) - E(X)^2E(Y)^2 = \left(V(X)+E(X)^2\right)\left(V(Y)+E(Y)^2\right) - E(X)^2E(Y)^2 \\ &= V(X)V(Y) + V(X)E(Y)^2 + E(X)^2V(Y) + E(X)^2E(Y)^2 - E(X)^2E(Y)^2 = V(X)V(Y) + V(X)E(Y)^2 + E(X)^2V(Y)\,. \end{aligned}$$

το οποίο είναι και το ζητούμενο.

Λύση Άσκ. 11. Η μέση τιμή των τ.μ. X, Y είναι ίση με

$$E(X) = 1\cdot P(X=1) + (-1)\cdot P(X=-1) = P(X=1) - P(X=-1)\,, \quad E(Y) = 1\cdot P(Y=1) + (-1)\cdot P(Y=-1) = P(Y=1) - P(Y=-1)\,,$$

και επειδή $E(X)$=0, $E(Y)$=0, συμπεραίνουμε ότι $P(X=1) = P(X=-1)$ και $P(Y=1) = P(Y=-1)$. Από το γεγονός τώρα ότι

$$\begin{aligned} P(X=1) &= P(X=1, Y=-1) + P(X=1, Y=1)\,, \quad P(X=-1) = P(X=-1, Y=-1) + P(X=-1, Y=1) \\ P(Y=1) &= P(X=-1, Y=1) + P(X=1, Y=1)\,, \quad P(Y=-1) = P(X=1, Y=-1) + P(X=-1, Y=-1) \end{aligned}$$

θα έχουμε

$$\begin{aligned} P(X=1, Y=-1) + P(X=1, Y=1) &= P(X=-1, Y=-1) + P(X=-1, Y=1) \\ P(X=-1, Y=1) + P(X=1, Y=1) &= P(X=1, Y=-1) + P(X=-1, Y=-1) \end{aligned}$$

Αφαιρώντας τις δύο παραπάνω εξισώσεις κατά μέλη προκύπτει ότι $P(X=1, Y=-1) = P(X=-1, Y=1)$, ενώ προσθέτοντας τις δύο παραπάνω εξισώσεις κατά μέλη προκύπτει τελικά ότι $P(X=1, Y=1) = P(X=-1, Y=-1)$.

Έστω τώρα ότι $p = 2P(X=1, Y=1)$. Άρα και $p = 2P(X=-1, Y=-1)$. Επίσης, επειδή

$$P(X=1, Y=-1) + P(X=-1, Y=1) + P(X=1, Y=1) + P(X=-1, Y=-1) = 1$$

θα ισχύει και ότι

$$P(X=1, Y=-1) + P(X=-1, Y=1) + p = 1 \Rightarrow 2P(X=1, Y=-1) = 2P(X=-1, Y=1) = 1-p\,.$$

Επομένως, είναι εύκολο να δούμε ότι $P(X=1) = P(X=-1) = P(Y=1) = P(Y=-1) = 1/2$. Η διασπορά $V(X)$ της τ.μ. X θα είναι

$$V(X) = E(X^2) - E(X)^2 = (1^2P(X=1) + (-1)^2P(X=-1)) - 0 = 1$$

και όμοια, $V(Y) = 1$. Τέλος,

$$\begin{aligned} Cov(X,Y) &= E(XY) - E(X)E(Y) = E(XY) = \sum_{x=-1,1,\; y=-1,1} xyP(X=x, Y=y) \\ &= (-1)(-1)P(X=-1, Y=-1) + (-1)\cdot 1\cdot P(X=-1, Y=1) + 1\cdot(-1)P(X=1, Y=-1) \; + 1\cdot 1\cdot P(X=1, Y=1) \; = 2p-1\,. \end{aligned}$$

Λύση Άσκ. 12. Η τ.μ. X ακολουθεί διωνυμική κατανομή $B(n,p)$ και επομένως μπορεί να θεωρηθεί ότι εκφράζει το πλήθος των επιτυχιών σε n ανεξάρτητες δοκιμές με πιθανότητα επιτυχίας p. Όμοια, η τ.μ. Y μπορεί να θεωρηθεί ότι εκφράζει το πλήθος των επιτυχιών σε m ανεξάρτητες δοκιμές με πιθανότητα επιτυχίας p. Μάλιστα, εφόσον οι τ.μ. X, Y είναι ανεξάρτητες, μπορούμε να θεωρήσουμε ότι οι n δοκιμές για την X και οι m δοκιμές για την Y είναι μεταξύ τους ανεξάρτητες. Άρα, μπορούμε να θεωρήσουμε συνολικά n+m ανεξάρτητες δοκιμές η κάθε μία από τις οποίες είναι «επιτυχία» με πιθανότητα p. Η τ.μ. X+Y τώρα εκφράζει το πλήθος των επιτυχιών στις n πρώτες δοκιμές, συν το πλήθος των επιτυχιών στις m τελευταίες δοκιμές. Με άλλα λόγια η X+Y εκφράζει το συνολικό πλήθος των επιτυχιών στις n+m δοκιμές. Γνωρίζουμε όμως ότι το πλήθος των επιτυχιών σε n+m ανεξάρτητες δοκιμές (με πιθ. επιτ. p) ακολουθεί διωνυμική κατανομή με παραμέτρους n+m, p. Άρα τελικά $X+Y \sim B(n+m,p)$.

Λύση Άσκ. 13. (i) Έστω A_i το ενδεχόμενο να είναι η τ.μ. X_i η μεγαλύτερη από τις $X_1,X_2,\ldots,X_n$. Τα ενδεχόμενα $A_1,A_2,\ldots,A_n$ είναι ισοπίθανα διότι δεν υπάρχει κάποια διάκριση μεταξύ των n τ.μ. $X_1,X_2,\ldots,X_n$ (όλες προέρχονται από την ίδια κατανομή). Επίσης, $P(A_1)+P(A_2)+\ldots+P(A_n)=1$ και άρα $P(A_1) = \ldots = P(A_n) = 1/n$. Επομένως και P(*παρατηρείται ρεκόρ στο χρόνο* n) = $P(A_n) = 1/n$.

(ii) Η τ.μ. R_n μπορεί να γραφεί στη μορφή $R_n = Y_1+Y_2+\ldots+Y_n$ όπου $Y_i = 1$ αν παρατηρείται ρεκόρ στο χρόνο i και $Y_i = 0$ διαφορετικά. Από το (i) γνωρίζουμε ότι $E(Y_i) = P(Y_i=1) = 1/i$, i=1,2,..,n και συνεπώς

$$E(R_n) = E(\sum_{i=1}^{n} Y_i) = \sum_{i=1}^{n} E(Y_i) = \sum_{i=1}^{n} \frac{1}{i}\,.$$

Επίσης, αποδεικνύεται* ότι οι τ.μ. Y_1, Y_2, ..., Y_n είναι ανεξάρτητες οπότε

$$V(R_n) = V(\sum_{i=1}^{n} Y_i) = \sum_{i=1}^{n} V(Y_i) = \sum_{i=1}^{n} \frac{1}{i}\left(1-\frac{1}{i}\right).$$

*(Έστω B_n η τ.μ. που δείχνει τη σχετική θέση της n-παρατήρησης X_n ανάμεσα στις $X_1,X_2,...,X_n$, δηλαδή B_n=1 αν η X_n είναι η μεγαλύτερη από τις $X_1,X_2,...,X_n$, B_n=2 αν η X_n είναι η δεύτερη μεγαλύτερη από τις $X_1,X_2,...,X_n$ κ.ο.κ. ... , $B_n=n$ αν η X_n είναι η μικρότερη από τις $X_1,X_2,...,X_n$. Παρατηρούμε ότι κάθε μία από τις $n!$ δυνατές διατάξεις των $X_1,X_2,...,X_n$ έχει την ίδια πιθανότητα εμφάνισης (με πιθανότητα 1 δεν εμφανίζονται "ισοπαλίες" μεταξύ των $X_1,...,X_n$ διότι είναι συνεχείς τ.μ.). Παρατηρούμε επίσης ότι υπάρχει μία 1-1 σχέση μεταξύ των δυνατών διατάξεων και των ενδεχομένων

$$[B_1 = b_1, B_2 = b_2,..., B_n = b_n],\quad b_1 \in \{1\}, b_2 \in \{1,2\},..., b_n \in \{1,2,...,n\}$$

(π.χ. αν n=3 τότε $X_3<X_1<X_2$ αν και μόνο αν B_1=1, B_2=1, B_3=3). Άρα τελικά

$$P(B_1 = b_1, B_2 = b_2,..., B_n = b_n) = \frac{1}{n!},\quad b_1 \in \{1\}, b_2 \in \{1,2\},..., b_n \in \{1,2,...,n\}\ .$$

Επίσης, παρατηρούμε ότι $P(B_i=1)= P(B_i=2)=...= P(B_i=i) = 1/i$ διότι η X_i έχει την ίδια πιθανότητα να είναι μεγαλύτερη, δεύτερη μεγαλύτερη, ..., μικρότερη από τις $X_1,X_2,...,X_i$ (διαισθητικά αυτό φαίνεται διότι δεν υπάρχει κάποια διάκριση μεταξύ των n τ.μ. $X_1,X_2,...,X_n$). Άρα τελικά

$$P(B_1 = b_1, B_2 = b_2,..., B_n = b_n) = \frac{1}{n!} = P(B_1 = b_1)P(B_2 = b_2)...P(B_n = b_n),$$

για $b_1 \in \{1\}, b_2 \in \{1,2\},..., b_n \in \{1,2,...,n\}$, το οποίο υποδηλώνει ότι οι τ.μ. $B_1,B_2,...,B_n$ είναι ανεξάρτητες. Συνεπώς και οι $Y_1,Y_2,...,Y_n$ θα είναι ανεξάρτητες ως συναρτήσεις των B_i, i=1,2,...,n (Y_i=1 ή 0 ανάλογα με το αν B_i=1 ή B_i>1 αντίστοιχα)).

Λύση Άσκ. 14. Έστω τ.μ. $Z_1,Z_2,...,Z_n$, n=180, έτσι ώστε Z_i=1 ή 0 ανάλογα με το αν στην i-ρίψη εμφανίστηκε άθροισμα 9 ή όχι. Θα είναι

$P(Z_i = 1) = P(\textit{άθροισμα ζαριών } 9)$

$= P(1^o\ \textit{ζάρι } 3, 2^o\ \textit{ζάρι } 6) + P(1^o\ \textit{ζάρι } 4, 2^o\ \textit{ζάρι } 5) + P(1^o\ \textit{ζάρι } 5, 2^o\ \textit{ζάρι } 4) + P(1^o\ \textit{ζάρι } 6, 2^o\ \textit{ζάρι } 3) = 4/36 = 1/9.$

(α)(i) Η τ.μ. $\sum_{i=1}^{n} Z_i$ εκφράζει το πλήθος των εμφανίσεων του 9 στις 180 ρίψεις. Επομένως, ζητείται η πιθανότητα $P(\sum_{i=1}^{n} Z_i \geq 30)$. Η τ.μ. $\sum_{i=1}^{n} Z_i$ μπορεί να θεωρηθεί ως το πλήθος των επιτυχιών σε n = 180 ανεξάρτητες δοκιμές με πιθ. επιτυχίας p = 1/9 και επομένως θα ακολουθεί διωνυμική κατανομή Β(n,p). Επειδή όμως το n είναι αρκετά μεγάλο θα χρησιμοποιήσουμε προσέγγιση από κανονική κατανομή. Με βάση το Κ.Ο.Θ. θα ισχύει ότι προσεγγιστικά (οι τ.μ. $Z_1,...,Z_n$ είναι ανεξάρτητες και ισόνομες),

$$\frac{\sum_{i=1}^{n} Z_i - nE(Z_1)}{\sqrt{nV(Z_1)}} = \frac{\sum_{i=1}^{n} Z_i - np}{\sqrt{np(1-p)}} \sim N(0,1)\ ,$$

και συνεπώς,

$$P(\sum_{i=1}^{n} Z_i \geq 30) = P(\frac{\sum_{i=1}^{n} Z_i - np}{\sqrt{np(1-p)}} \geq \frac{30-np}{\sqrt{np(1-p)}}) \approx 1-\Phi(\frac{30-180/9}{\sqrt{180(1/9)(8/9)}}) = 1-\Phi(2.37) = 0.0089$$

(ii) Θα είναι, $P(\sum_{i=1}^{n} Z_i < 30) = 1 - P(\sum_{i=1}^{n} Z_i \geq 30) \approx 1 - 0.0089 = 0.9911$.

(iii) Όμοια με το (i) θα είναι

$$P(20 \leq \sum_{i=1}^{n} Z_i < 30) = P(\frac{20-np}{\sqrt{np(1-p)}} \leq \frac{\sum_{i=1}^{n} Z_i - np}{\sqrt{np(1-p)}} < \frac{30-np}{\sqrt{np(1-p)}})$$

$$\approx \Phi(\frac{30-180/9}{\sqrt{180(1/9)(1-1/9)}}) - \Phi(\frac{20-180/9}{\sqrt{180(1/9)(1-1/9)}}) \approx \Phi(2.37) - \Phi(0) = 0.9911 - 0.5 = 0.4911$$

(β) Έστω X_i,Y_i το αποτέλεσμα της i-ρίψης των δύο ζαριών, i = 1,2,...,n=180, και $W_i = X_i+Y_i$. Ζητείται η πιθανότητα $P(\sum_{i=1}^{n} W_i > 1300)$. Από το Κ.Ο.Θ θα ισχύει ότι προσεγγιστικά, (οι τ.μ. $W_1,...,W_n$ είναι ανεξάρτητες και ισόνομες),

$$\frac{\sum_{i=1}^{n} W_i - nE(W_i)}{\sqrt{nV(X)}} \sim N(0,1)\ .$$

Αλλά,

$$E(X_i) = E(Y_i) = 1P(X_i = 1) + ... + 6P(X_i = 6) = (1+2+...+6)/6 = 3.5\ ,$$

$$E(X_i^2) = E(Y_i^2) = 1^2 P(X_i = 1) + ... + 6^2 P(X_i = 6) = (1^2 + 2^2 + ... + 6^2)/6 = 91/6$$

και $V(X_i) = V(Y_i) = 91/6 - 3.5^2 \approx 2{,}916$. Άρα, $E(W_i) = E(X_i) + E(Y_i) = 7$, $V(W_i) = V(X_i) + V(Y_i) \approx 5.832$ και συνεπώς,

$$\frac{\sum_{i=1}^{n} W_i - 180\cdot 7}{\sqrt{180\cdot 5.832}} = \frac{\sum_{i=1}^{n} W_i - 1260}{32.4} \sim N(0,1)\,.$$

Τελικά, λαμβάνουμε ότι

$$P(\sum_{i=1}^{n} W_i > 1300) = P(\frac{\sum_{i=1}^{n} W_i - 1260}{32.4} > \frac{1300-1260}{32.4}) \approx 1-\Phi(\frac{1300-1260}{32.4}) \approx 1-\Phi(1,23) \approx 0.1093\,.$$

Λύση Άσκ. 15. Έστω $X_1,X_2,...,X_n$, οι απαντήσεις των n=200 τυχαία επιλεγμένων ψηφοφόρων (δηλαδή, X_i = 1 αν ο i-ψηφοφόρος του δείγματος είναι υπέρ του κόμματος Α και X_i=0 διαφορετικά). Εφόσον το 45% των ψηφοφόρων της πόλης να είναι υπέρ του κόμματος Α, θα ισχύει ότι $P(X_i=1) = 0.45 = p$, $P(X_i=0) = 0.55 = 1-p$. Η πιθανότητα η πλειοψηφία των ψηφοφόρων του δείγματος να είναι υπέρ του κόμματος Α εκφράζεται ως εξής: $P(\sum_{i=1}^{n} X_i > n/2)$. Για τον προσεγγιστικό υπολογισμό αυτής της πιθανότητας θα χρησιμοποιήσουμε το Κ.Ο.Θ. Συγκεκριμένα θα ισχύει ότι προσεγγιστικά (οι $X_1,X_2,...,X_n$ είναι ανεξάρτητες και ισόνομες τ.μ.)

$$\frac{\sum_{i=1}^{n} X_i - nE(X_i)}{\sqrt{nV(X_i)}} = \frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} \sim N(0,1)$$

και άρα τελικά,

$$P(\sum_{i=1}^{n} X_i > \frac{n}{2}) = P(\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} > \frac{n/2-np}{\sqrt{np(1-p)}}) \approx 1-\Phi(\frac{n/2-np}{\sqrt{np(1-p)}}) = 1-\Phi(\sqrt{200}\,\frac{0.5-0.45}{\sqrt{0.45\cdot 0.55}}) \approx 1-\Phi(1.42) \approx 0.0788\,.$$

Λύση Άσκ. 16. Έστω οι τ.μ. $X_1,X_2,...,X_n$, n=400 έτσι ώστε X_i = 1 αν το i κομμάτι του δείγματος είναι ελαττωματικό και X_i = 0 διαφορετικά, i=1,2,...,n. Το 10% από τα προϊόντα μιας μηχανής είναι ελαττωματικά και συνεπώς $P(X_i=1) = 0.1 = p$.

(α) Ζητείται η πιθανότητα $P(\sum_{i=1}^{n} X_i \le 30)$. Από το Κ.Ο.Θ θα ισχύει ότι προσεγγιστικά,

$$\frac{\sum_{i=1}^{n} X_i - nE(X_i)}{\sqrt{nV(X_i)}} = \frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} \sim N(0,1)$$

και άρα

$$P(\sum_{i=1}^{n} X_i \le 30) = P(\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} \le \frac{30-np}{\sqrt{np(1-p)}}) \approx \Phi(\frac{30-np}{\sqrt{np(1-p)}}) = \Phi(\frac{30-40}{6}) \approx \Phi(-1.66) =$$

$$= 1-\Phi(1.66) = 1-0.9515 = 0.0485\,.$$

Όμοια λύνονται και τα ερωτήματα (β), (γ) και (δ).

Λύση Άσκ. 17. Έστω $X_1,X_2,...,X_n$, n=100 οι αριθμοί των ελαττωμάτων στην 1η, 2η,..., n-οστή ταινία, αντίστοιχα. Οι τ.μ. $X_1,X_2,...,X_n$ είναι ανεξάρτητες (προέρχονται από ανεξάρτητα πειράματα) και ακολουθούν κατανομή Poisson με μέση τιμή 800/200 = 4. Ζητείται η πιθανότητα

$$P(\sum_{i=1}^{n} X_i \le 430)\,.$$

Από το Κ.Ο.Θ θα ισχύει προσεγγιστικά ότι, $(\sum_{i=1}^{n} X_i - n\cdot E(X))/\sqrt{n\cdot V(X)} \sim N(0,1)$ και άρα

$$P(\sum_{i=1}^{n} X_i \le 430) = P(\frac{\sum_{i=1}^{n} X_i - 100\cdot 4}{\sqrt{100\cdot 4}} \le \frac{430-100\cdot 4}{\sqrt{100\cdot 4}}) = \Phi(\frac{430-100\cdot 4}{\sqrt{100\cdot 4}}) = \Phi(1.5) = 0.9332\,.$$

Λύση Άσκ. 18. Χωρίζουμε το χρόνο των 10 ωρών σε 600 λεπτά και έστω X_i η τ.μ. που εκφράζει το πλήθος των τηλεφωνημάτων κατά τη διάρκεια του i-λεπτού (i = 1,2,...,n=600). Οι τ.μ. $X_1,X_2,...,X_n$ είναι ανεξάρτητες και ακολουθούν κατανομή Poisson με λ=1 (και άρα $E(X_i) = V(X_i) = \lambda = 1$). Από το Κ.Ο.Θ. θα ισχύει ότι προσεγγιστικά,

$$\frac{\sum_{i=1}^{n} X_i - nE(X_i)}{\sqrt{nV(X_i)}} = \frac{\sum_{i=1}^{n} X_i - n\lambda}{\sqrt{n\lambda}} \sim N(0,1)$$

και για τη ζητούμενη πιθανότητα θα ισχύει ότι

$$P(\sum_{i=1}^{n} 0.05X_i \ge 28) = P(\sum_{i=1}^{n} X_i \ge 560) = P(\frac{\sum_{i=1}^{n} X_i - n\lambda}{\sqrt{n\lambda}} \ge \frac{560-n\lambda}{\sqrt{n\lambda}}) \approx 1-\Phi(\frac{560-n\lambda}{\sqrt{n\lambda}})$$

$$= 1 - \Phi(\frac{560 - 600 \cdot 1}{\sqrt{600 \cdot 1}}) \approx 1 - \Phi(-0.61) = \Phi(0.61) \approx \ 0.7291$$

και συνεπώς το κέρδος θα είναι μεγαλύτερο των 28 € με πιθανότητα 72.91%.

Λύση Άσκ. 19. Αν θέσουμε X_i=1 αν ο i-πελάτης καλεί την εταιρία (μία συγκεκριμένη στιγμή) και 0 διαφορετικά, i=1,2,…,n=5000, τότε το πλήθος των πελατών που καλούν την εταιρία θα είναι ίσο με $\sum_{i=1}^{n} X_i$. Αν c είναι ο αριθμός των υπαλλήλων στο τμήμα υποστήριξης, θα πρέπει (χρησιμοποιώντας και το Κ.Ο.Θ.) να ισχύει ότι

$$P(\sum_{i=1}^{n} X_i \le c) \ge 1 - a \Leftrightarrow P(\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} \le \frac{c - np}{\sqrt{np(1-p)}}) \ge 1 - a \quad \Leftrightarrow P(Z \le \frac{c - np}{\sqrt{np(1-p)}}) \ge 1 - a$$

$$\Leftrightarrow \frac{c - np}{\sqrt{np(1-p)}} \ge \Phi^{-1}(1 - a) = Z_a \quad \Leftrightarrow c \ge np + \sqrt{np(1-p)} Z_a$$

(1−α=90%) και από το (α) θα είναι (p=0.1%=0.001)

$$c \ge np + \sqrt{np(1-p)} \mathrm{Z}_a = 5000 \cdot 0.001 + \sqrt{5000 \cdot 0.001 \cdot 0.999} \mathrm{Z}_{0.1} = \ 5 + 2.235 \cdot 1.28 \approx 7.86 \ .$$

Άρα τελικά c = 8.

Λύση Άσκ. 20. Προφανώς θα είναι α = 0 και β = 0.1€. Συνεπώς ο πελάτης θα κληθεί να πληρώσει

$$Y = 293.47 + \sum_{i=1}^{n} X_i$$

η μέση τιμή του οποίου είναι

$$E(Y) = 293.47 + \sum_{i=1}^{n} E(X_i) = 293.47 + n\frac{a + \beta}{2} = 293.47 + 150\frac{0.1}{2} = 300.97$$

Από το Κ.Ο.Θ., θα ισχύει (προσεγγιστικά) ότι

$$Z = \frac{\sum_{i=1}^{n} X_i - nE(X)}{\sqrt{nV(X)}} = \frac{\sum_{i=1}^{n} X_i - n\beta / 2}{\sqrt{n\beta^2 / 12}} \sim N(0,1)$$

και αν γράψουμε την τ.μ. Y ως εξής

$$Y = 293.47 + \sum_{i=1}^{n} X_i = 293.47 + n\frac{\beta}{2} + \beta\sqrt{\frac{n}{12}} \cdot Z$$

συμπεραίνουμε ότι αυτή θα ακολουθεί (προσεγγιστικά) κανονική κατανομή (ως γραμμική συνάρτηση μιας κανονικής τ.μ.) με μέση τιμή και διασπορά

$$E(Y) = 293.47 + n\beta / 2 + \beta\sqrt{n / 12} \cdot E(Z) = 293.47 + n\beta / 2 = 300.97$$

$$V(Y) = V(293.47 + n\beta / 2 + \beta\sqrt{n / 12} \cdot Z) = V(\beta\sqrt{n / 12} \cdot Z) = \frac{n\beta^2}{12} V(Z) = \frac{n\beta^2}{12} = 0.125$$

Τέλος, η πιθανότητα να πληρώσει τώρα περισσότερα από c = 300€ είναι

$$P(Y > 300) \approx P(\frac{Y - 300.97}{\sqrt{0.125}} \ge \frac{300 - 300.97}{\sqrt{0.125}}) \approx 1 - \Phi(\frac{300 - 300.97}{\sqrt{0.125}}) \ \approx 1 - \Phi(-2.74) = \Phi(2.74) \approx 0.9969.$$

Λύση Άσκ. 21. Έστω X_1,X_2,...,X_n, n = 55 οι τ.μ. που εκφράζουν τους χρόνους εξυπηρέτησης των 55 πελατών. Οι τ.μ. X_1,...,X_n είναι ανεξάρτητες και ισόνομες με $E(X_i) = 5$, $Var(X_i) = 2^2$. Η τ.μ. $\sum_{i=1}^{n} X_i$ εκφράζει το συνολικό χρόνο εξυπηρέτησης. Σύμφωνα με το Κ.Ο.Θ. θα ισχύει, με μεγάλη προσέγγιση, οτι

$$Z = \frac{\sum_{i=1}^{55} X_i - 55 \cdot 5}{2\sqrt{55}} \sim N(0,1) \ .$$

Ζητείται η πιθανότητα

$$P(\sum_{i=1}^{55} X_i > 300) = P(\frac{\sum_{i=1}^{55} X_i - 55 \cdot 5}{2 \cdot \sqrt{55}} > \frac{300 - 55 \cdot 5}{2 \cdot \sqrt{55}}) = P(Z > 1.6855) = 1 - P(Z \le 1.6855) \ \approx 1 - \Phi(1.69) \approx 1 - 0.9545 = 0.0455 \ .$$

Λύση Άσκ. 22. Θα ισχύει άμεσα ότι

$$P(|X-3|\geq 2)\leq \frac{V(X)}{2^2}=\frac{2}{2^2}=0.5 \quad \text{και} \quad P(|X-3|\geq 1)\leq \frac{V(X)}{2^2}=\frac{2}{1^2}=2\,.$$

Παρατηρούμε ότι στη δεύτερη περίπτωση δεν εξάγουμε κάποια χρήσιμη πληροφορία αφού έτσι και αλλιώς $P(|X-3|\geq 1) \leq 1$.

Λύση Άσκ. 23. (α) Η τ.μ. $X = X_1+...+X_{100}$ είναι θετική και άρα από την ανισότητα Markov θα ισχύει ότι

$$P(X\geq 120)\leq \frac{E(X)}{120}=\frac{E(X_1)+...+E(X_{100})}{120}=\frac{100}{120}\approx 0.833$$

Επίσης, χρησιμοποιώντας την ανισότητα Chebyshev προκύπτει ακόμη καλύτερα ότι

$$P(X\geq 120)\leq P(|X-100|\geq 20)\leq \frac{V(X)}{20^2}=\frac{V(X_1)+...+V(X_{100})}{20^2}=\frac{100}{400}=0.25$$

(β) Από το Κ.Ο.Θ θα έχουμε ότι, προσεγγιστικά

$$\frac{\sum_{i=1}^{100}X_i-nE(X_1)}{\sqrt{nV(X_1)}}\sim N(0,1)$$

και άρα τελικά

$$P(X\geq 120)=P(\frac{X-100}{\sqrt{100}}\geq \frac{120-100}{\sqrt{100}})\approx 1-\Phi(\frac{120-100}{\sqrt{100}})=1-\Phi(2)=1-0.9772=0.0228\,.$$

ΚΕΦΑΛΑΙΟ 4

Λύση Άσκ. 24. θα είναι

$$E(S^2)=c(\frac{1}{n}\sum_{i=1}^{n}E(X_i^2)-E(\overline{X}^2))=c(\frac{1}{n}\sum_{i=1}^{n}(V(X_i)+E(X_i)^2)-(V(\overline{X})+E(\overline{X})^2))$$

$$=c(\frac{1}{n}\sum_{i=1}^{n}(\sigma^2+\mu^2)-(\frac{\sigma^2}{n}+\mu^2))=c(\sigma^2+\mu^2-\frac{\sigma^2}{n}-\mu^2)=c(\frac{n-1}{n})\sigma^2$$

και άρα $c = n/n-1$.

Λύση Άσκ. 25. Θα αναζητήσουμε μία αμερόληπτη εκτιμήτρια της μορφής $a\overline{X}+b\overline{Y}$ που να έχει τη μικρότερη δυνατή διασπορά. Θα πρέπει

$$E(a\overline{X}+b\overline{Y})=\mu \Leftrightarrow aE(\overline{X})+bE(\overline{Y})=\mu \Leftrightarrow a\mu+bc\mu=\mu \Leftrightarrow b=(1-a)/c\,.$$

Επίσης,

$$Var(a\overline{X}+b\overline{Y})=a^2Var(\overline{X})+b^2Var(\overline{Y})=a^2\frac{\sigma^2}{n}+b^2\frac{d\,\sigma^2}{k}=\frac{ka^2c^2+n(1-a)^2d}{nkc^2}\sigma^2\,.$$

Αρκεί να βρούμε το a που ελαχιστοποιεί την παράσταση $f(a)=ka^2c^2+n(1-a)^2d$. Ισχύει ότι

$$f'(a)=2akc^2-2n(1-a)d=2akc^2-2nd+2nad=0 \Leftrightarrow a=\frac{nd}{kc^2+nd} \qquad (\,f''(a)=2kc^2+2nd>0\,)$$

και επομένως η εκτιμήτρια

$$a\overline{X}+b\overline{Y}=\frac{nd}{kc^2+nd}\overline{X}+\frac{1}{c}\left(1-\frac{nd}{kc^2+nd}\right)\overline{Y}=\frac{nd\overline{X}+kc\overline{Y}}{kc^2+nd}$$

είναι αμερόληπτη και έχει τη μικρότερη διασπορά ανάμεσα στις εκτιμήτριες της μορφής $a\overline{X}+b\overline{Y}$.

Λύση Άσκ. 26. Έχουμε δύο άγνωστες παραμέτρους $\theta_1=a$, $\theta_2=b$ και επομένως θα πάρουμε δύο εξισώσεις με δύο αγνώστους ($k=2$):

$$\begin{cases} m_1=\mu_1'(\theta_1,\theta_2) \\ m_2=\mu_2'(\theta_1,\theta_2) \end{cases} \Leftrightarrow \begin{cases} m_1=E(X_i) \\ m_2=E(X_i^2)=Var(X_i)+E(X_i)^2 \end{cases} \Leftrightarrow \begin{cases} \frac{1}{n}\sum_{i=1}^{n}X_i=\frac{a}{b} \\ \frac{1}{n}\sum_{i=1}^{n}X_i^2=\frac{a}{b^2}+\frac{a^2}{b^2} \end{cases}$$

$$\Leftrightarrow\begin{cases}\overline{X}=\dfrac{a}{b}\\ \dfrac{1}{n}\left(\sum_{i=1}^{n}X_i^2-n\overline{X}^2\right)=\dfrac{a}{b^2}\end{cases}\Leftrightarrow\begin{cases}a=b\overline{X}\\ \dfrac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2=\dfrac{a}{b^2}\end{cases}\Leftrightarrow\begin{cases}\widetilde{a}=\overline{X}^2\left(\dfrac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right)^{-1}\\ \widetilde{b}=\overline{X}\left(\dfrac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right)^{-1}\end{cases}.$$

Λύση Άσκ. 27. Έχουμε δύο άγνωστες παραμέτρους $\theta_1=\alpha$, $\theta_2=b$ και επομένως θα πάρουμε δύο εξισώσεις με δύο αγνώστους (k=2):

$$\begin{cases}m_1=\mu_1'(\theta_1,\theta_2)\\ m_2=\mu_2'(\theta_1,\theta_2)\end{cases}\Leftrightarrow\begin{cases}m_1=E(X_i)\\ m_2=E(X_i^2)=Var(X_i)+E(X_i)^2\end{cases}\Leftrightarrow\begin{cases}\dfrac{1}{n}\sum_{i=1}^{n}X_i=a-b\\ \dfrac{1}{n}\sum_{i=1}^{n}X_i^2=4ab+(a-b)^2\end{cases}$$

$$\Leftrightarrow\begin{cases}\dfrac{1}{n}\sum_{i=1}^{n}X_i=a-b\\ \sqrt{\dfrac{1}{n}\sum_{i=1}^{n}X_i^2}=a+b\end{cases}\Leftrightarrow\begin{cases}b=\dfrac{1}{2}\left(\sqrt{\dfrac{1}{n}\sum_{i=1}^{n}X_i^2}-\dfrac{1}{n}\sum_{i=1}^{n}X_i\right)\\ a=\dfrac{1}{2}\left(\sqrt{\dfrac{1}{n}\sum_{i=1}^{n}X_i^2}+\dfrac{1}{n}\sum_{i=1}^{n}X_i\right)\end{cases}.$$

Λύση Άσκ. 28. Έχουμε δύο άγνωστες παραμέτρους $\theta_1=\alpha$, $\theta_2=\lambda$ και επομένως θα πάρουμε δύο εξισώσεις με δύο αγνώστους (k=2):

$$\begin{cases}m_1=E(X_i)\\ m_2=E(X_i^2)=Var(X_i)+E(X_i)^2\end{cases}\Leftrightarrow\begin{cases}\dfrac{1}{n}\sum_{i=1}^{n}X_i=\dfrac{a}{\lambda}\\ \dfrac{1}{n}\sum_{i=1}^{n}X_i^2=\dfrac{a}{\lambda^2}+\dfrac{a^2}{\lambda^2}\end{cases}\Leftrightarrow\begin{cases}a=\lambda\overline{X}\\ \dfrac{1}{n}\sum_{i=1}^{n}X_i^2=\dfrac{\lambda\overline{X}}{\lambda^2}+\dfrac{\lambda^2\overline{X}^2}{\lambda^2}\end{cases}\Leftrightarrow\begin{cases}\widetilde{a}=\dfrac{\overline{X}^2}{\dfrac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2}=4\\ \widetilde{\lambda}=\dfrac{\overline{X}}{\dfrac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2}=0.886\end{cases}$$

Λύση Άσκ. 29. Ο λογάριθμος της συνάρτησης πιθανοφάνειας του δείγματος θα δίνεται από τον τύπο

$$\ln L(\lambda)=\ln\prod_{i=1}^{n}f(x_i;\lambda)=\sum_{i=1}^{n}\ln f(x_i;\lambda)=\sum_{i=1}^{n}\ln\left(\frac{\lambda^3}{2}x_i^2e^{-\lambda x}\right)=\sum_{i=1}^{n}\left(3\ln\lambda-\ln 2+\ln x_i^2-\lambda x_i\right)=3n\ln\lambda-n\ln 2+\sum_{i=1}^{n}\ln x_i^2-\lambda\sum_{i=1}^{n}x_i\ .$$

Παρατηρούμε ότι η παραπάνω συνάρτηση του λ είναι παραγωγίσιμη σε όλο τον παραμετρικό χώρο $(0,\infty)$. Επομένως, τα πιθανά λ για τα οποία μεγιστοποιείται η $\ln L(\lambda)$ θα είναι ρίζες της εξίσωσης $(\ln L(\lambda))'=0$. Ειδικότερα θα έχουμε ότι

$$\frac{\partial}{\partial\lambda}(\ln L(\lambda))=\frac{\partial}{\partial\lambda}\left(3n\ln\lambda-n\ln 2+\sum_{i=1}^{n}\ln x_i^2-\lambda\sum_{i=1}^{n}x_i\right)=\frac{3n}{\lambda}-\sum_{i=1}^{n}x_i=0$$

και επομένως $\widehat{\lambda}=\dfrac{3}{\overline{X}}$. Απομένει φυσικά να επαληθεύσουμε ότι η παραπάνω τιμή αποτελεί μέγιστο. Πράγματι, ισχύει ότι

$$\frac{\partial^2}{\partial\lambda^2}(\ln L(\widehat{\lambda}))=-\frac{3n}{\widehat{\lambda}^2}=<0\ .$$

Λύση Άσκ. 30. Ο λογάριθμος της συνάρτησης πιθανοφάνειας του δείγματος θα δίνεται από τον τύπο

$$\ln L(\lambda)=\ln\prod_{i=1}^{n}f(x_i;\lambda)=\sum_{i=1}^{n}\ln f(x_i;\lambda)=\sum_{i=1}^{n}\ln\left(2\lambda x_ie^{-\lambda x_i^2}\right)=\sum_{i=1}^{n}\left(\ln 2+\ln\lambda+\ln x_i-\lambda x_i^2\right)=n\ln 2+n\ln\lambda+\sum_{i=1}^{n}\ln x_i-\lambda\sum_{i=1}^{n}x_i^2\ .$$

Παρατηρούμε ότι η παραπάνω συνάρτηση του λ είναι παραγωγίσιμη σε όλο τον παραμετρικό χώρο $(0,\infty)$. Επομένως, τα πιθανά λ για τα οποία μεγιστοποιείται η $\ln L(\lambda)$ θα είναι ρίζες της εξίσωσης $(\ln L(\lambda))'=0$. Ειδικότερα θα έχουμε ότι

$$\frac{\partial}{\partial\lambda}\ln L(\lambda)=\frac{\partial}{\partial\lambda}\left(n\ln 2+n\ln\lambda+\sum_{i=1}^{n}\ln x_i-\lambda\sum_{i=1}^{n}x_i^2\right)=\frac{n}{\lambda}-\sum_{i=1}^{n}x_i^2=0$$

και επομένως $\widehat{\lambda}=\dfrac{n}{\sum_{i=1}^{n}X_i^2}$. Απομένει φυσικά να επαληθεύσουμε ότι η παραπάνω τιμή αποτελεί μέγιστο. Πράγματι, ισχύει ότι

$$\frac{\partial^2}{\partial\lambda^2}(\ln L(\widehat{\lambda})) = -\frac{n}{\widehat{\lambda}^2} < 0 .$$

Λύση Άσκ. 31. Ο λογάριθμος της συνάρτησης πιθανοφάνειας του δείγματος δίνεται από τον τύπο

$$\ln L(\mu) = \ln\prod_{i=1}^{n} f(x_i;\mu) = \sum_{i=1}^{n}\ln f(x_i;\mu) = \sum_{i=1}^{n}\ln\left(\frac{1}{x_i\sqrt{2\pi\sigma^2}}e^{-\frac{(\ln x_i-\mu)^2}{2\sigma^2}}\right)$$

$$= \sum_{i=1}^{n}\left(-\ln x_i - \ln\sqrt{2\pi\sigma^2} - \frac{(\ln x_i-\mu)^2}{2\sigma^2}\right) = -\sum_{i=1}^{n}\ln x_i - n\ln\sqrt{2\pi\sigma^2} - \sum_{i=1}^{n}\frac{(\ln x_i-\mu)^2}{2\sigma^2}$$

Παρατηρούμε ότι η παραπάνω συνάρτηση του λ είναι παραγωγίσιμη σε όλο τον παραμετρικό χώρο (0,∞). Επομένως, τα πιθανά λ για τα οποία μεγιστοποιείται η ln$L$(λ) θα είναι ρίζες της εξίσωσης (ln$L$(μ))΄=0. Ειδικότερα θα έχουμε ότι

$$\frac{\partial}{\partial\mu}(\ln L(\mu)) = -\frac{\partial}{\partial\mu}\sum_{i=1}^{n}\frac{(\ln x_i-\mu)^2}{2\sigma^2} = \sum_{i=1}^{n}\frac{\ln x_i-\mu}{\sigma^2} = 0$$

και επομένως

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\ln X_i$$

Απομένει φυσικά να επαληθεύσουμε ότι η παραπάνω τιμή αποτελεί μέγιστο.

Λύση Άσκ. 32. Θα είναι

$$\ln L(p) = \ln\prod_{i=1}^{n} f(x_i;p) = \sum_{i=1}^{n}\ln f(x_i;p) = \sum_{i=1}^{n}\ln((1-p)^{x_i-1}p) = \ln(1-p)\sum_{i=1}^{n}(x_i-1) + n\ln p$$

και συνεπώς,

$$\frac{d}{dp}\ln L(p) = -\frac{1}{1-p}\sum_{i=1}^{n}(x_i-1) + \frac{n}{p} = 0 \Leftrightarrow p = \bar{x}^{-1} = \frac{5}{3+4+1+3+2} = \frac{5}{13}$$

ΚΕΦΑΛΑΙΟ 5

Λύση Άσκ. 33. Από τις παραπάνω παρατηρήσεις βρίσκουμε ότι $\bar{X} = 151.4$ και $S^2 = 78.3$ α) Το δείγμα είναι σχετικά μικρό και για αυτό θα χρησιμοποιήσουμε την κατανομή t. Ένα δ.ε. συντελεστού 99% μια το μ θα είναι

$$[\bar{X} - \frac{S}{\sqrt{n}}t_{n-1}(a/2), \bar{X} + \frac{S}{\sqrt{n}}t_{n-1}(a/2)] = [151.4 - \frac{\sqrt{78.3}}{\sqrt{5}}4.6,\ 151.4 + \frac{\sqrt{78.3}}{\sqrt{5}}4.6] = [133.197, 169.603] .$$

όπου 1−α = 99% και $t_4(0.005) = 4.60$.

β) Ένα δ.ε. για το σ^2 όταν μ άγνωστο είναι

$$[\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\chi^2_{n-1}(a/2)}, \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\chi^2_{n-1}(1-a/2)}] = [\frac{(n-1)S^2}{\chi^2_4(0.005)}, \frac{(n-1)S^2}{\chi^2_4(0.995)}] = [\frac{4\cdot 78.3}{14.86}, \frac{4\cdot 78.3}{0.21}] = [21.07, 1491.4] .$$

Επομένως ένα δ.ε. για το σ συντελεστού 99% θα είναι το $[\sqrt{21.07}, \sqrt{1491.4}] = [4.59, 38.61]$.

Λύση Άσκ. 34. Από τις παραπάνω 7 παρατηρήσεις βρίσκουμε ότι $\bar{X} = 1038.71$ και $S^2 = 6712.24$.

α) Το δείγμα είναι σχετικά μικρό και για αυτό θα χρησιμοποιήσουμε την κατανομή t. Ένα δ.ε. συντελεστού 95% μια το μ όταν το σ είναι άγνωστο θα είναι

$$[\bar{X} - \frac{S}{\sqrt{n}}t_{n-1}(a/2), \bar{X} + \frac{S}{\sqrt{n}}t_{n-1}(a/2)] = [1038.71 \pm \frac{\sqrt{6712.24}}{\sqrt{7}}2.447,] = [963.39, 1114.94]$$

όπου 1−α=95% και $t_6(0.025) \approx 2.447$. Στην περίπτωση που είναι γνωστό ότι σ=4, θα έχουμε το δ.ε. 95% για το μ:

$$[\bar{X} - \frac{\sigma}{\sqrt{n}}Z_{a/2}, \bar{X} + \frac{\sigma}{\sqrt{n}}Z_{a/2}] = [1038.71 - \frac{70}{\sqrt{7}}1.96, 1038.71 + \frac{70}{\sqrt{7}}1.96] = [987.3, 1091.03] .$$

Τέλος, θα πρέπει:

$$2\frac{\sigma}{\sqrt{n}}Z_{a/2}=20 \Leftrightarrow n=4\sigma^2 Z_{a/2}^2/20^2=4\cdot 70^2\cdot 1.96^2/20^2=27.44$$

Λύση Άσκ. 35. Πριν προχωρήσουμε θα πρέπει να υπολογίσουμε τις εκτιμήτριες $\overline{X}$ και S^2. Για το δειγματικό μέσο θα ισχύει ότι

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i=\frac{1}{50}\sum_{i=1}^{50}X_i=\frac{125}{50}=2.5$$

ενώ για τη δειγματική διασπορά θα έχουμε

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

$$=\frac{1}{n-1}(\sum_{i=1}^{n}X_i^2+n\overline{X}^2-2n\overline{X}^2)=\frac{1}{n-1}(\sum_{i=1}^{n}X_i^2-n\overline{X}^2)=\frac{1}{49}(325-50\cdot 2.5^2)=0.2551\,.$$

α) Σύμφωνα με τα παραπάνω, ένα δ.ε. συντελεστού 95% για το μ θα είναι

$$[\overline{X}-\frac{S}{\sqrt{n}}t_{n-1}(a/2),\overline{X}+\frac{S}{\sqrt{n}}t_{n-1}(a/2)]=[2.5-\frac{\sqrt{0.2551}}{\sqrt{50}}t_{49}(0.025),\,2.5+\frac{\sqrt{0.2551}}{\sqrt{50}}t_{49}(0.025)]\,.$$

Επειδή το μέγεθος του δείγματος είναι αρκετά μεγάλο (>30), προκύπτει ότι $t_{49}(0.025)=Z_{0.025}=1.96$ και άρα τελικά το παραπάνω δ.ε. συντελεστού 95% για το μέσο χρόνο ζωής θα είναι:

$$=[2.5-\frac{\sqrt{0.2551}}{\sqrt{50}}1.96,\,2.5+\frac{\sqrt{0.2551}}{\sqrt{50}}1.96]=[2.36,\,2.64]\,.$$

β) Ένα δ.ε. για το σ^2 όταν μ άγνωστο είναι

$$[\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\chi_{n-1}^2(a/2)},\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\chi_{n-1}^2(1-a/2)}]=[\frac{(n-1)S^2}{\chi_{49}^2(0.025)},\frac{(n-1)S^2}{\chi_{49}^2(0.975)}]\,.$$

και αντικαθιστώντας θα έχουμε τελικά το δ.ε. $[\frac{49\cdot 0.2551}{71.4},\frac{49\cdot 0.2551}{32.3}]=[0.1750,0.3870]$. Τέλος, ένα δ.ε. για το σ συντελεστού 95% θα είναι το $[\sqrt{0.1750},\sqrt{0.387}]=[0.4184,0.622]$.

Λύση Άσκ. 36. α) Έστω $X_1,X_2,\ldots,X_n$, οι απαντήσεις των n=100 ατόμων του τ.δ. ώστε X_i=1 αν ο i-καταναλωτής χρησιμοποιεί το συγκεκριμένο προιόν και X_i=0 διαφορετικά. Προφανώς, το τ.δ. X_1, X_2, ..., X_n προέρχεται από Β(1,p) κατανομή ($P(X_i=1)=p$, $P(X_i=0)=1-p$). Ζητάμε δ.ε. συντελεστού 95% για το p. Θα είναι:

$$[\overline{X}-\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}Z_{a/2},\overline{X}+\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}Z_{a/2}]=[0.2-\sqrt{\frac{0.2\cdot 0.8}{100}}1.96,0.2+\sqrt{\frac{0.2\cdot 0.8}{100}}1.96]=[0.1216,\,0.2784]\,.$$

Άρα το πραγματικό ποσοστό p βρίσκεται μεταξύ του 12.6% και του 27.8% με συντελεστή εμπιστοσύνης 95%.

β) Το εύρος του διαστήματος στο (α) είναι ίσο με 27.8%−12.6% = 15.2%. Έστω ότι για να γίνει ίσο με 5% πρέπει να πάρουμε δείγμα μεγέθους n_1. Έστω επίσης $\overline{X}'$ το αντίστοιχο δειγματικό ποσοστό από το δείγμα αυτό. Θα πρέπει να ισχύει ότι

$$2\sqrt{\frac{\overline{X}'(1-\overline{X}')}{n_1}}Z_{a/2}\approx 0.02 \qquad \text{ή ισοδύναμα} \qquad n_1\approx 4\frac{\overline{X}'(1-\overline{X}')}{0.05^2}Z_{a/2}^2\,,$$

και αν υποθέσουμε ότι θα βρούμε και πάλι $\overline{X}'$ περίπου ίσο με 20% θα έχουμε τελικά ότι

$$n_1\approx 4\frac{0.2(1-0.2)}{0.05^2}1.96^2=983.45\,.$$

Λύση Άσκ. 37. Έστω $X_1,X_2,\ldots,X_n$, οι απαντήσεις των n=100 ατόμων του τ.δ. ώστε X_i=1 αν το i-άτομο προτίθεται να είναι άνεργος και X_i=1 διαφορετικά. Ζητάμε δ.ε. συντελεστού 95% για το p. Το δ.ε. θα είναι:

$$[\overline{X}-\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}Z_{a/2},\overline{X}+\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}Z_{a/2}]=[0.11-\sqrt{\frac{0.11\cdot 0.89}{100}}1.96,\,0.11+\sqrt{\frac{0.11\cdot 0.89}{100}}1.96]=[0.04867,\,0.1713]\,.$$

ii) Το πλήθος των ανέργων στην πόλη θα είναι $m\cdot p$ και επομένως ζητάμε δ.ε. 95% για το $m\cdot p$. Γνωρίζουμε ότι

$$P(L \le p \le U) = 1-a, \quad \text{όπου} \quad L = \overline{X} - \sqrt{\frac{\overline{X}(1-\overline{X})}{n}} Z_{a/2}, \quad U = \overline{X} + \sqrt{\frac{\overline{X}(1-\overline{X})}{n}} Z_{a/2}$$

και επομένως $P(mL \le mp \le mU) = 1-a$. Άρα ένα δ.ε. 95% για τον αριθμό των ψήφων του κόμματος θα είναι το

$$m\cdot[L, U]=100000\cdot[0.04867,\ 0.1713] = [48670,\ 171300] .$$

iii) Έστω ότι για να έχουμε εύρος 3% θα πρέπει να πάρουμε δείγμα μεγέθους n_1. Έστω επίσης $\overline{X}'$ το αντίστοιχο δειγματικό ποσοστό από το δείγμα αυτό. Θα πρέπει να ισχύει ότι

$$2\sqrt{\frac{\overline{X}'(1-\overline{X}')}{n_1}} Z_{a/2} \approx 0.03 \qquad \text{ή ισοδύναμα} \qquad n_1 \approx 4\frac{\overline{X}'(1-\overline{X}')}{0.03^2} Z_{a/2}^2 ,$$

και αν υποθέσουμε ότι θα βρούμε και πάλι $\overline{X}'$ περίπου ίσο με 10% θα έχουμε τελικά ότι

$$n_1 \approx 4\frac{0.11(1-0.11)}{0.03^2} 1.96^2 = 852.8 .$$

Λύση Άσκ. 38. Γνωρίζουμε ότι το δ.ε. συντελεστού 1−α για το πηλίκο σ_2^2 / σ_1^2 είναι

$$[\frac{S_2^2}{S_1^2} F_{n_1-1,n_2-1}(1-\tfrac{a}{2}), \frac{S_2^2}{S_1^2} F_{n_1-1,n_2-1}(\tfrac{a}{2})] .$$

Από πίνακες των άνω α-σημείων της κατανομής F βρίσκουμε ότι:

$$F_{n_1-1,n_2-1}(\tfrac{a}{2}) = F_{3,2}(0.025) = 39.16 , \quad F_{n_1-1,n_2-1}(1-\tfrac{a}{2}) = F_{3,2}(0.975) = 0.062$$

και συνεπώς το δ.ε. θα είναι ($S_1^2 = 1.5625, S_2^2 = 0.09$)

$$[\frac{1.5625}{0.09}0.062,\ \frac{1.5625}{0.09}39.16]= [1.07, 679.8] .$$

Το διάστημα αυτό δεν περιέχει το 1 και άρα μπορούμε να απορρίψουμε ότι $\sigma_2^2 / \sigma_1^2 = 1$ με συντελεστή εμπιστοσύνης 95%.

ΚΕΦΑΛΑΙΟ 6

Λύση Άσκ. 39. Η περιοχή απόρριψης θα είναι

$$K: \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} > Z_a$$

δηλαδή

$$K: \overline{X} > \mu_0 + \frac{\sigma}{\sqrt{n}} Z_a = 10 + \frac{2}{4} 1.645 = 10.8225$$

Η ισχύς θα είναι θα είναι

$\pi(11) = P(\text{απόρριψη της } H_0 \,/\, \text{ισχύει η } H_1) = P(\overline{X} > 10.8225 / \mu = 11)$

$$= P(\frac{\overline{X}-11}{\sigma/\sqrt{n}} > \frac{10.8225-11}{\sigma/\sqrt{n}} / \mu = 11) = 1-\Phi(\frac{10.8225-11}{2/4}) = 1-\Phi(-0.355) = \Phi(0.355) \approx 0.6387$$

Λύση Άσκ. 40. Υπενθυμίζεται ότι από την άσκηση 5.1. είχαμε βρει ότι $\overline{X} = 284.16$ και $S^2 = 17.5827$

α) Πρόκειται για τον έλεγχο H_0:$\mu=\mu_0$, H_1:$\mu<\mu_0$ όπου μ_0=290 όταν το σ είναι άγνωστο. Από την παράγραφο 6.β. έχουμε την περιοχή απόρριψης της H_0:

$$K: \frac{\overline{X}-\mu_0}{S/\sqrt{n}} < -t_{n-1}(a) \quad \text{ή ισοδύναμα} \quad \overline{X} < \mu_0 - \frac{S}{\sqrt{n}} t_{n-1}(a)$$

Βρίσκουμε ότι

$$\frac{\overline{X}-\mu_0}{S/\sqrt{n}} = \frac{284.16-290}{\sqrt{17.5827/10}} = -4.404$$

και επειδή $-t_{n-1}(a)=-t_9(0.05)=-1.833$ τελικά θα έχουμε ότι –4.404 < –1.833, δηλαδή απορρίπτουμε την H_0:μ=290 έναντι της H_1:μ<290 σε ε.σ. 5%. Το p-value του δείγματος που έδωσε $\overline{X}$ = 284.16 θα είναι

$$\text{p-value} = P(\frac{\overline{X}-\mu_0}{S/\sqrt{n}} < \frac{284.16-\mu_0}{S/\sqrt{n}} / \mathrm{H}_0) = F_{t_9}(\frac{284.16-290}{\sqrt{17.58/10}}) = F_{t_9}(-4.404) \approx 0.000855$$

Συνεπώς, η πιθανότητα να πάρουμε ένα τόσο «ακραίο» δείγμα ενώ ισχύει η H_0 είναι μόλις 0.0855%. (φαίνεται ότι το δείγμα που πήραμε είναι αρκετά ακραίο αν ισχύει η H_0). Παρατηρούμε ότι είναι κατά πολύ μικρότερη του 0.05 και συνεπώς απορρίπτουμε την H_0. (Αντί να ελέγξουμε αν $\frac{\overline{X}-\mu_0}{S/\sqrt{n}} < -t_{n-1}(a)$ θα μπορούσαμε να ισοδύναμα ελέγξουμε αν p-value < a).

β. Πρόκειται για τον έλεγχο H_0:μ=μ_0, H_1:μ<μ_0 όπου μ_0=290 όταν το σ είναι γνωστό και ίσο με 4. Από την παράγραφο 6.α. έχουμε την περιοχή απόρριψης της H_0:

$$K\text{:}\ \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} < -Z_a \ \text{ ή ισοδύναμα }\ \overline{X} < \mu_0 - \frac{\sigma}{\sqrt{n}} Z_\alpha$$

Βρίσκουμε ότι

$$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} = \frac{284.16-290}{4/\sqrt{10}} = -4.61$$

και επειδή $-Z_a=-Z_{0.05}=-1.645$ τελικά θα έχουμε ότι –4.61 < –1.645, δηλαδή απορρίπτουμε και πάλι την H_0:μ=290 έναντι της H_1:μ<290 σε ε.σ. 5%. Η πιθανότητα σφάλματος τύπου *I* θα είναι ίση με α=0.05 ενώ η P(II) θα είναι

$$P(\mathbf{X}\in A/H_1) = P(\overline{X} \ge \mu_0 - \frac{\sigma}{\sqrt{n}} Z_a / \mu=\mu_1>\mu_0) = P(\frac{\overline{X}-\mu_1}{\sigma/\sqrt{n}} \ge \frac{\mu_0-\mu_1}{\sigma/\sqrt{n}} - Z_a / \mu=\mu_1>\mu_0)$$
$$= 1-\Phi(\frac{\mu_0-\mu_1}{\sigma/\sqrt{n}} - Z_a) = 1-\Phi(\frac{290-\mu_1}{4/\sqrt{10}} - 1.645)$$

Η ισχύς του ελέγχου θα είναι ίση με π(μ_1) = 1−P(II). Επίσης το p-value του δείγματος που έδωσε $\overline{X}$ =284.16 θα είναι

$$\text{p-value}= P(\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} < \frac{284.16-\mu_0}{\sigma/\sqrt{n}} / \mathrm{H}_0) = \Phi(\frac{284.16-290}{4/\sqrt{10}}) = \Phi(-4.61) \approx 0.000002$$

Συνεπώς, η πιθανότητα να πάρουμε ένα τόσο «ακραίο» δείγμα ενώ ισχύει η H_0 είναι μόλις 0.0002%. (φαίνεται ότι το δείγμα που πήραμε είναι πολύ ακραίο αν ισχύει η H_0). Παρατηρούμε ότι είναι κατά πολύ μικρότερο του 0.05 και συνεπώς απορρίπτουμε την H_0 (αντί να ελέγξουμε αν $\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} < -Z_a$ θα μπορούσαμε ισοδύναμα να ελέγξουμε αν p-value < a)

Λύση Άσκ. 41. Υπενθυμίζεται ότι από την άσκηση 5.2. είχαμε βρει ότι $\overline{X}=$ 356.6 και $S^2=128.711$.
α) Πρόκειται για τον έλεγχο H_0:μ=μ_0, H_1:μ≠μ_0 όπου μ_0=360 όταν το σ είναι άγνωστο. Από την παράγραφο 6.β. έχουμε την περιοχή απόρριψης της H_0:

$$K\text{:}\ \frac{|\overline{X}-\mu_0|}{S/\sqrt{n}} > t_{n-1}(a/2)$$

Βρίσκουμε ότι

$$\frac{|\overline{X}-\mu_0|}{S/\sqrt{n}} = \frac{|356.6-360|}{\sqrt{128.711/10}} = 0.94$$

και επειδή $t_{n-1}(a/2)=t_9(0.005)=3.250$ τελικά θα έχουμε ότι 0.94 < 3.250, δηλαδή δεν μπορούμε να απορρίψουμε την H_0:μ=360 έναντι της H_1:μ≠360 σε ε.σ. 1%. Το p-value του δείγματος που έδωσε $\overline{X}$ =356.6 θα είναι

$$\text{p-value} = P(\frac{|\overline{X}-\mu_0|}{S/\sqrt{n}} > \frac{|356.6-\mu_0|}{\sqrt{128.711/n}} / H_0) = P(\frac{|\overline{X}-\mu_0|}{S/\sqrt{n}} > 0.94/H_0)$$
$$= 1-P(-0.94 \le \frac{\overline{X}-\mu_0}{S/\sqrt{n}} \le 0.94/\mu=\mu_0) = 1-(F_{t_9}(0.94)-F_{t_9}(-0.94)) = 2(1-F_{t_9}(0.94)) = 2(1-0.814) = 0.372$$

Συνεπώς, η πιθανότητα να πάρουμε ένα τόσο «ακραίο» δείγμα ενώ ισχύει η H_0 είναι 37.2% (φαίνεται ότι το δείγμα που πήραμε δεν είναι καθόλου ακραίο αν ισχύει η H_0). Παρατηρούμε ότι είναι αρκετά μεγαλύτερη του 0.01 και συνεπώς δεχόμαστε (ή καλύτερα, λέμε ότι δεν έχουμε αρκετά στοιχεία ώστε να απορρίψουμε) την H_0.

β) Πρόκειται για τον έλεγχο H_0:σ=$σ_0$, H_1:σ<$σ_0$ όπου $σ_0$=30 όταν το μ είναι άγνωστο. Από την παράγραφο 6.δ. έχουμε την περιοχή απόρριψης της H_0:

$$K:\ \frac{(n-1)S^2}{\sigma_0^2} < \chi^2_{n-1}(a)\ .$$

Βρίσκουμε ότι

$$\frac{(n-1)S^2}{\sigma_0^2} = \frac{9\cdot 128.711}{30^2} = 1.287$$

και επειδή $\chi^2_{n-1}(a) = \chi^2_9(0.01)$ =21.67 > 1.287 τελικά απορρίπτουμε την H_0:σ=30 έναντι της H_1:σ<30 σε ε.σ. 1%.

Λύση Άσκ. 42. Από την άσκηση 4.4. έχουμε ότι $\overline{X} = 69.798$, και

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right) = \frac{1}{199}(977265 - 200\cdot 69.798^2) = 14.63\ .$$

α) Ζητείται ο έλεγχος H_0: μ=$μ_0$, H_1: μ<$μ_0$, $μ_0$=71 με $σ^2$ άγνωστο. Η κρίσιμη περιοχή αυτού του ελέγχου γνωρίζουμε ότι είναι:

$$K:\ \frac{\overline{X}-\mu_0}{S/\sqrt{n}} < -t_{n-1}(a)\ .$$

Αντικαθιστώντας θα έχουμε ότι:

$$\frac{\overline{X}-\mu_0}{S/\sqrt{n}} = \frac{69.798-71}{\sqrt{14.63/200}} = -4.44$$

ενώ $-t_{n-1}(a) = -t_{199}(0.05) = -Z_{0.05} = -1.645 < -4.44$ και άρα απορρίπτουμε την H_0 σε ε.σ. 0.05.

β) Ζητείται ο έλεγχος H_0: σ=$σ_0$, H_1: σ≠$σ_0$, $σ_0$=4 με μ άγνωστο. Η κρίσιμη περιοχή αυτού του ελέγχου γνωρίζουμε ότι είναι:

$$K:\ \frac{(n-1)S^2}{\sigma_0^2} < \chi^2_{n-1}(1-\frac{a}{2}) \ \text{ ή }\ \frac{(n-1)S^2}{\sigma_0^2} > \chi^2_{n-1}(\frac{a}{2})$$

Αντικαθιστώντας θα έχουμε ότι:

$$\frac{(n-1)S^2}{\sigma_0^2} = \frac{199\cdot 14.63}{16} = 181.96$$

ενώ

$$\chi^2_{n-1}(1-\frac{a}{2}) = \chi^2_{199}(1-0.025) = 199+\sqrt{2\cdot 199}Z_{1-0.025} = 199-\sqrt{2\cdot 199}Z_{0.025} = 199-\sqrt{2\cdot 199}\cdot 1.96 = 159.89$$

και

$$\chi^2_{n-1}(\frac{a}{2}) = \chi^2_{199}(0.025) = 199+\sqrt{2\cdot 199}Z_{0.025} = 199+\sqrt{2\cdot 199}Z_{0.025} = 238.1$$

και επειδή 159.89 < 181.96 < 238.1 δεχόμαστε (ή καλύτερα, λέμε ότι δεν έχουμε αρκετά στοιχεία ώστε να απορρίψουμε) την H_0.

Λύση Άσκ. 43. Επειδή $Y_i = Y_{i-1}e^{X_i}$, θα είναι $X_i = \ln Y_i - \ln Y_{i-1}$ και συνεπώς, τα $X_1, X_2, ..., X_5$ θα είναι

-0.0341775, -0.096159, -0.0675049, 0.0717111, 0.03116

Βρίσκουμε ότι $\overline{X}$ = -0.018994, S^2 = 0.00482028. Ζητείται ο έλεγχος H_0: μ=$μ_0$, H_1: μ<$μ_0$, $μ_0$=0. Η κρίσιμη περιοχή θα είναι

$$T < -t_{n-1}(a) = -t_4(0.05) = -2.13$$

όπου

$$T = \frac{\overline{X}-\mu_0}{S/\sqrt{n}} = -8.81$$

και άρα απορρίπτουμε την H_0.

(ii) Στη συγκεκριμένη περίπτωση έχουμε ένα τ.δ. $X_1, X_2, ..., X_n$, n=5 από την N(μ, $σ^2$=0.01) και ζητείται ο έλεγχος της υπόθεσης

H_0: μ=$μ_0$ (=0) έναντι της H_0: μ<$μ_0$ (=0)

σε επίπεδο σημαντικότητας α. Η κρίσιμη περιοχή είναι η

$$K\colon\ T < -Z_a = -Z_{0.05} = -1.645$$

όπου τώρα, $T = \dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ = -4.24 και άρα και πάλι απορρίπτουμε την H_0. Η πιθανότητα να έχουμε απορρίψει λανθασμένα είναι α=5%. Το *p*-value του ελέγχου θα είναι

$$P(\frac{\overline{X}-\mu_0}{\sqrt{\sigma^2/n}} < -4.24/H_0) = \Phi(-4.24) = 0$$

Επίσης, η ισχύς του ελέγχου θα είναι

$$\pi(\mu) = 1 - P(II) = 1 - P(\overline{X} \geq \mu_0 - \sqrt{\frac{\sigma^2}{n}}Z_a\ /H_1) = 1 - P(\frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}} \geq \frac{\mu_0-\mu}{\sqrt{\sigma^2/n}} - Z_a\ /H_1)$$

$$= \Phi(\frac{\mu_0-\mu}{\sqrt{\sigma^2/n}} - Z_a) = \Phi(\frac{-\mu}{0.1/\sqrt{5}} - 1.645)\,.$$

Τέλος, αν ίσχυε η H_1 με μ = -0.1 τότε

$$\pi(0.1) = \Phi(\frac{0.1}{0.1/\sqrt{5}} - 1.645) = \Phi(0.591) = 0.7227\,,$$

η οποία είναι στην ουσία η πιθανότητα ορθής απόρριψης της H_0 (πιθανότητα σωστής απόφασης) αν ίσχυε ότι μ = -0.1.

Λύση Άσκ. 44. Διαπιστώνουμε ότι πρόκειται για τον έλεγχο απλής υποθέσεως έναντι απλής οπότε μπορούμε να εφαρμόσουμε το Λήμμα N - P. Θα έχουμε ότι

$$K:\lambda(\mathbf{x}) = \frac{\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(x-\mu_0)^2}{2\sigma^2})}{\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(x-\mu_1)^2}{2\sigma^2})} = \frac{\exp(-\frac{(x-\mu_0)^2}{2\sigma^2})}{\exp(-\frac{(x-\mu_1)^2}{2\sigma^2})} = e^{-\frac{(x-\mu_0)^2}{2\sigma^2}+\frac{(x-\mu_1)^2}{2\sigma^2}} = e^{-\frac{1}{2\sigma^2}\left((x-\mu_0)^2-(x-\mu_1)^2\right)}.$$

Παρατηρούμε στη συνέχεια ότι

$$\lambda(\mathbf{x})<c \Leftrightarrow (x-\mu_0)^2-(x-\mu_1)^2 > -2\sigma^2\ln c \equiv c' \Leftrightarrow \left(x^2+\mu_0^2-2x\mu_0\right)-\left(x^2+\mu_1^2-2x\mu_1\right) > c' \Leftrightarrow 2x\mu_1 - 2x\mu_0 > c' - \mu_0^2 + \mu_1^2 \equiv c''$$

και συνεπώς, επειδή $\mu_1-\mu_0 > 0$, συμπεραίνουμε ότι η ανισότητα λ(**X**) < c είναι ισοδύναμη με την

$$X > \frac{c''}{2(\mu_1-\mu_0)} \equiv c_0\,.$$

Άρα η κρίσιμη περιοχή του ελέγχου (περιοχή απόρριψης) θα έχει τη μορφή K: $X > c_0$ για κάποιο c_0 το οποίο θα καθορίζεται από την $P(I) = P(\mathbf{X}\in K / H_0)$ = α. Επομένως, το c_0 θα πρέπει να είναι τέτοιο ώστε

$$P(I) = P(X > c_0/H_0) = P(\frac{X-\mu_0}{\sqrt{\sigma^2}} > \frac{c_0-\mu_0}{\sqrt{\sigma^2}}/H_0) = a$$

και επειδή, κάτω από την H_0, η τ.μ. $Z = \dfrac{X-\mu_0}{\sqrt{\sigma^2}} \sim N(0,1)$ συμπεραίνουμε τελικά ότι

$$\frac{c_0-\mu_0}{\sqrt{\sigma^2}} = Z_a \Rightarrow c_0 = \mu_0 + \sigma Z_a\,.$$

Άρα, σύμφωνα με το Λήμμα N-P, ο βέλτιστος έλεγχος σε επίπεδο σημαντικότητας α θα έχει περιοχή απόρριψης:

$$X > \mu_0 + \sigma Z_a = 10 + 3\cdot 1.645 = 14.935 \quad \text{ή ισοδύναμα} \quad K\colon\ \frac{X-\mu_0}{\sigma} > Z_a\,.$$

Επειδή λοιπόν X = 14, δεχόμαστε την H_0. Η πιθανότητα σφάλματος τύπου II θα είναι ίση με

$$P(X \leq \mu_0 + \sigma Z_a / H_1) = P(X \leq \mu_0 + \sigma Z_a / \mu = \mu_1) = P(\frac{X-\mu_1}{\sigma} \leq \frac{\mu_0-\mu_1}{\sigma} + Z_a / \mu = \mu_1)$$

και επειδή κάτω από την H_1, ισχύει ότι $Z' = \dfrac{X-\mu_1}{\sigma} \sim N(0,1)$, προκύπτει ότι

$$P(II) = \Phi(\frac{\mu_0 - \mu_1}{\sigma} + Z_a)\ .$$

Τέλος, η ισχύς του ελέγχου θα είναι

$$\pi(\mu_1) = \ 1 - P(II) = 1 - \Phi(\frac{\mu_0 - \mu_1}{\sigma} + Z_a) = \Phi(\frac{\mu_1 - \mu_0}{\sigma} - Z_a) = \Phi(\frac{20-10}{3} - 1.645) = \Phi(1.68) = 0.9535.$$

Λύση Άσκ. 45. Ζητείται ο έλεγχος της υπόθεσης H_0: $p \leq p_0$ έναντι της H_1: $p > p_0$, $p_0 = 35\%$. Η κρίσιμη περιοχή αυτού του ε-λέγχου γνωρίζουμε ότι είναι:

$$K:\ \ \frac{\overline{X} - p_0}{\sqrt{p_0(1-p_0)/n}} > Z_a$$

Αντικαθιστώντας θα έχουμε ότι:

$$\frac{\overline{X} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.4 - 0.35}{\sqrt{0.35(1-0.35)/500}} = 2.344$$

ενώ Z_a=$Z_{0.05}$=1.645 < 2.344 και συνεπώς απορρίπτουμε την H_0 σε ε.σ. α=0.05.

Λύση Άσκ. 46. Στη συγκεκριμένη περίπτωση ζητείται ο έλεγχος της υπόθεσης

$$H_0\text{: } p > p_0 \text{ έναντι της } H_1\text{: } p < p_0$$

όπου p_0=50%. Γνωρίζουμε ότι η κρίσιμη περιοχή θα είναι

$$K:\ \ \frac{\overline{X} - p_0}{\sqrt{p_0(1-p_0)/n}} < -Z_a\ .$$

Αντικαθιστώντας βρίσκουμε ότι

$$\frac{\overline{X} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.44 - 0.5}{\sqrt{0.5(1-0.5)/300}} = -2.07 > -Z_{0.01} = -2.33\ .$$

Επομένως δεν έχουμε αρκετά στοιχεία για να απορρίψουμε τον παραπάνω ισχυρισμό.

Παρατηρούμε ότι για να απορρίπταμε την H_0: p=0.5 λαμβάνοντας ένα μεγαλύτερο δείγμα n' θα έπρεπε να ισχύει ότι

$$\frac{\overline{X}' - p_0}{\sqrt{p_0(1-p_0)/n'}} < -Z_a \ \text{ δηλαδή } \ n' > \frac{p_0(1-p_0)}{(\overline{X}' - p_0)^2} Z_a^2$$

όπου $\overline{X}'$ είναι το αντίστοιχο δειγματικό ποσοστό που θα βρίσκαμε. Βρίσκοντας, σύμφωνα με την εκφώνηση, και πάλι $\overline{X}'$ = 0.44 θα έπρεπε

$$n' > \frac{p_0(1-p_0)}{(\overline{X}' - p_0)^2} Z_a^2 = \frac{0.5(1-0.5)}{(0.44-0.5)^2} 2.33^2 \approx 377$$

Λύση Άσκ. 47. Υπολογίζουμε από τα δύο δείγματα ότι $\overline{X} = 172.625$ και $\overline{Y} = 134.7$.

α) Ζητείται ο έλεγχος της μορφής H_0: μ_1=μ_2, H_1:$\mu_1 \neq \mu_2$ όταν σ_1,σ_2 είναι γνωστά σε ε.σ. α=0.05. Η κρίσιμη περιοχή του ελέγχου σύμφωνα με τα παραπάνω θα είναι

$$K:\ \ |T| > Z_{\alpha/2} \quad \text{όπου} \quad T = \frac{\left(\overline{X} - \overline{Y}\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\ .$$

Αντικαθιστώντας θα έχουμε ότι

$$|T| = \frac{\left|\overline{X} - \overline{Y}\right|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{|172.625 - 134.7|}{40\sqrt{\frac{1}{8} + \frac{1}{10}}} = \frac{37.925}{40\sqrt{0.225}} \approx 2$$

Ενώ $Z_{a/2}$ = $Z_{0.025}$ = 1.96 < 2 και άρα απορρίπτουμε την H_0.

β) Ζητείται ο έλεγχος της μορφής H_0: μ_1=μ_2, H_1:μ_1>μ_2 όταν σ_1,σ_2 είναι γνωστά σε ε.σ. α=0.05. Η κρίσιμη περιοχή του ελέγχου σύμφωνα με τα παραπάνω θα είναι

$$T > Z_a \quad \text{όπου} \quad T = \frac{\left(\overline{X} - \overline{Y}\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

Αντικαθιστώντας θα έχουμε ότι

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \approx 2$$

Ενώ $Z_a = Z_{0.05} = 1.645 < 2$ και άρα απορρίπτουμε και πάλι την H_0.

Λύση Άσκ. 48. α) Ζητείται ο έλεγχος της μορφής H_0: $\mu_1=\mu_2$, H_1:$\mu_1 > \mu_2$ όταν σ_1,σ_2 είναι άγνωστά αλλά ίσα, σε ε.σ. α=0.05. Υπολογίζουμε ότι $\overline{X} = 113, \overline{Y} = 99, S_1^2 = 57, S_2^2 = 52$ και η κρίσιμη περιοχή του ελέγχου θα είναι

$$\text{K: } T > t_{n_1+n_2-2}(a) \quad \text{όπου} \quad T = \frac{\sqrt{\frac{n_1 n_2}{n_1+n_2}}\left(\overline{X} - \overline{Y}\right)}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}}.$$

Αντικαθιστώντας θα έχουμε ότι $T = 2.3226$, ενώ $t_{n_1+n_2-2}(a/2) = t_4(0.05) = 2.132 < 2.3226$ και άρα απορρίπτουμε την H_0.

Λύση Άσκ. 49. α) Ζητείται ο έλεγχος της μορφής H_0: $\mu_1=\mu_2$, H_1:$\mu_1 \neq \mu_2$ όταν σ_1,σ_2 είναι άγνωστά αλλά ίσα, σε ε.σ. α=0.05. Η κρίσιμη περιοχή του ελέγχου σύμφωνα με τα παραπάνω θα είναι

$$\text{K: } |T| > t_{n_1+n_2-2}(a/2) \quad \text{όπου} \quad T = \frac{\sqrt{\frac{n_1 n_2}{n_1+n_2}}\left(\overline{X} - \overline{Y}\right)}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}}.$$

Αντικαθιστώντας θα έχουμε ότι

$$|T| = \frac{\sqrt{\frac{10 \cdot 20}{10+20}}\left|100.9 - 104.45\right|}{\sqrt{\frac{9 \cdot 8.766 + 19 \cdot 12.99}{10+20-2}}} = \left|-2.68\right|$$

Ενώ $t_{n_1+n_2-2}(a/2) = t_{28}(0.025) = 2.048 < 2.68$ και άρα απορρίπτουμε την H_0.

β) Ζητείται ο έλεγχος της μορφής H_0: $\mu_1=\mu_2$, H_1:$\mu_1<\mu_2$ όταν σ_1,σ_2 είναι άγνωστά αλλά ίσα σε ε.σ. α = 0.05. Η κρίσιμη περιοχή του ελέγχου σύμφωνα με τα παραπάνω θα είναι

$$K\text{: } T < -\, t_{n_1+n_2-2}(a) \quad \text{όπου} \quad T = -2.68 .$$

Επειδή $-t_{n_1+n_2-2}(a) = -t_{28}(0.05) = -1.701 > -2.68$ απορρίπτουμε την H_0.

Λύση Άσκ. 50. α) Ζητείται ο έλεγχος της μορφής H_0: $\sigma_1=\sigma_2$, H_1:$\sigma_1 \neq \sigma_2$ όταν τα μ_1, μ_2 είναι άγνωστα σε ε.σ. α=0.05. Η κρίσιμη περιοχή του ελέγχου σύμφωνα με τα παραπάνω θα είναι

$$K\text{: } T < F_{n_1-1,n_2-1}(1-a/2) \text{ ή } T > F_{n_1-1,n_2-1}(a/2) \quad \text{όπου} \quad T = \frac{n_2-1}{n_1-1} \frac{\sum_{i=1}^{n_1}(X_i - \overline{X})^2}{\sum_{j=1}^{n_2}(Y_j - \overline{Y})^2} = \frac{S_1^2}{S_2^2}.$$

Αντικαθιστώντας θα έχουμε ότι

$$T = \frac{8.766}{12.99} = 0.6748$$

Ενώ

$$F_{n_1-1,n_2-1}(1-\tfrac{a}{2}) = F_{9,19}(0.975) = 0.273, \quad F_{n_1-1,n_2-1}(\tfrac{a}{2}) = F_{9,19}(0.025) = 2.84,$$

και επειδή 0.273<0.6748<2.84 δεχόμαστε την H_0 (δηλ. οι πληθυσμοί είναι ομοσκεδαστικοί) σε ε.σ. 0.05.

Λύση Άσκ. 51. Είναι

$$K:\ T < F_{n_1-1,n_2-1}(1-a) = F_{49,49}(0.95) = 0.6221$$

όπου $T = \dfrac{S_1^2}{S_2^2} = \dfrac{100^2}{120^2} = 0.6944$ και επομένως δεν μπορούμε να απορρίψουμε την H_0.

Λύση Άσκ. 52. α) Ζητείται ο έλεγχος της μορφής H_0: $p_1=p_2$, H_1:$p_1 \neq p_2$ σε ε.σ. α=0.05. Η κρίσιμη περιοχή του ελέγχου σύμφωνα με τα παραπάνω θα είναι (n_1,n_2>30)

$$K:\ |T| > Z_{a/2} \text{ όπου } T = \frac{\overline{X}-\overline{Y}}{\sqrt{P(1-P)(1/n_1+1/n_2)}},\quad P = \frac{n_1\overline{X}+n_2\overline{Y}}{n_1+n_2}$$

Αντικαθιστώντας θα έχουμε ότι

$$P = \frac{400\frac{16}{400}+300\frac{24}{300}}{300+400} = \frac{16+24}{700} = 0.0571 \quad \text{και} \quad T = \frac{\frac{16}{400}-\frac{24}{300}}{\sqrt{0.0571(1-0.0571)\left(\frac{1}{300}+\frac{1}{400}\right)}} = -2.257$$

Ενώ $Z_{α/2} = Z_{0.005} = 2.58$ και επειδή 2.257 < 2.58 δεχόμαστε (ή καλύτερα, λέμε ότι δεν έχουμε αρκετά στοιχεία ώστε να απορρίψουμε) την H_0.

β) Ζητείται ο έλεγχος της μορφής H_0: $p_1=p_2$, H_1:$p_1<p_2$ σε ε.σ. α=0.05. Η κρίσιμη περιοχή του ελέγχου σύμφωνα με τα παραπάνω θα είναι (n_1,n_2>30)

$$K:\ T < -Z_a \quad \text{όπου} \quad T = -2.257$$

ενώ $Z_α = Z_{0.01} = 2.33$ και επειδή $T = -2.257 > -2.33$ δεχόμαστε (ή καλύτερα, λέμε ότι δεν έχουμε αρκετά στοιχεία ώστε να απορρίψουμε) την H_0.

Λύση Άσκ. 53. α) Ζητείται ο έλεγχος της μορφής H_0: $p_1=p_2$, H_1:$p_1<p_2$ σε ε.σ. α=0.05. Η κρίσιμη περιοχή του ελέγχου σύμφωνα με τα παραπάνω θα είναι (n_1,n_2>30)

$$K:\ T < -Z_a \text{ όπου } T = \frac{\overline{X}-\overline{Y}}{\sqrt{P(1-P)(1/n_1+1/n_2)}},\quad P = \frac{n_1\overline{X}+n_2\overline{Y}}{n_1+n_2}$$

Αντικαθιστώντας θα έχουμε ότι

$$P = \frac{78+240}{200+500} = 0.454 \quad \text{και} \quad T = \frac{\frac{78}{200}-\frac{240}{500}}{\sqrt{0.454(1-0.454)\left(\frac{1}{200}+\frac{1}{500}\right)}} = \frac{-0.09}{0.041} = -2.16$$

Ενώ $Z_α = Z_{0.05} = 1.645$ και επειδή $T = -2.16 < -1.645$ απορρίπτουμε την H_0.

Λύση Άσκ. 54. Πρόκειται για δύο ανεξάρτητα δείγματα $X_1,X_2,...,X_n \sim B(1,p_1)$ και $Y_1,Y_2,...,Y_m \sim B(1,p_2)$ με n=1000, m=150. Θέλουμε να πραγματοποιήσουμε το έλεγχο H_0: $p_1 = p_2$, H_1: $p_1 > p_2$ έχοντας δειγματικά ποσοστά $\overline{X} = 0.312$, $\overline{Y} = 0.22$ διότι

$$\begin{cases} \overline{Y} = 0.22 \\ \dfrac{n\overline{X}+m\overline{Y}}{n+m} = 0.30 \end{cases} \Rightarrow \begin{cases} \overline{Y} = 0.22 \\ \overline{X} = \dfrac{0.30(n+m)-m\overline{Y}}{n} = \dfrac{1150\cdot 0.30 - 150\cdot 0.22}{1000} = 0.312 \end{cases}$$

Η κρίσιμη περιοχή του ελέγχου θα είναι (n, m>30)

$$K:\ T > Z_a \text{ όπου } T = \frac{\overline{X}-\overline{Y}}{\sqrt{P(1-P)(1/n+1/m)}},\quad P = \frac{n\overline{X}+m\overline{Y}}{n+m}$$

Αντικαθιστώντας θα έχουμε ότι

$$P = \frac{1000\cdot 0.312+150\cdot 0.22}{1150} = 0.3 \quad \text{και} \quad T = \frac{0.312-0.22}{\sqrt{0.3\cdot 0.7\left(\frac{1}{1000}+\frac{1}{150}\right)}} = 2.292$$

ενώ $Z_α = Z_{0.01} = 2.33$ και επειδή $T = 2.292 < 2.33$ δεχόμαστε (ή καλύτερα, λέμε ότι δεν έχουμε αρκετά στοιχεία ώστε να απορρίψουμε) την H_0.