Άσκηση 1

Ο αριθμός των ωρών ανά εβδομάδα που περνούν στο «campus» οι φοιτητές ενός μεγάλου πανεπιστημίου ακολουθεί μια κανονική κατανομή με τυπική απόκλιση 5 ώρες. Παίρνουμε ένα τυχαίο δείγμα n φοιτητών. Πόσο μεγάλο πρέπει να είναι το δείγμα ώστε η πιθανότητα ο μέσος του δείγματος να διαφέρει λιγότερο από μια ώρα από το μέσο του πληθυσμού να είναι τουλάχιστον 0,90;

Λύση

Έστω X= αριθμός των ωρών ανά εβδομάδα που βρίσκονται στο «campus» οι φοιτητές, τότε $X \sim N(\mu, 5^2)$. Θέλουμε να βρούμε n ώστε

$P(|\overline{X} - \mu| < 1) \geq 0{,}90$. Από τα στοιχεία πιο πάνω ξέρουμε ότι $\overline{X} \sim N\left(\mu, \frac{5^2}{n}\right)$, τότε,

$$P(|\overline{X}-\mu|<1)=P\left(\frac{|\overline{X}-\mu|}{5/\sqrt{n}}<\frac{1}{5/\sqrt{n}}\right)=P\left(|Z|<\frac{1}{5/\sqrt{n}}\right)$$ (|...| παριστά την απόλυτη τιμή).

$$=P\left(|Z|<\frac{\sqrt{n}}{5}\right)$$

Από το πίνακα της τυποποιημένης κανονικής κατανομής πρέπει να βρούμε εκείνο το αριθμό που αφήνει 5% του εμβαδού δεξιά, δηλαδή 1,645. Τότε έχουμε,

$\frac{\sqrt{n}}{5}=1{,}645$, λύνουμε και βρίσκουμε n=68.

Άσκηση 2

Έστω ότι από κάποιο πληθυσμό με μέσο μ και διακύμανση σ^2, έχουμε πάρει ένα τυχαίο δείγμα τριών παρατηρήσεων X_1,X_2,X_3. Ποια από τις παρακάτω εκτιμήτριες συναρτήσεις είναι αμερόληπτη εκτίμηση του μέσου μ του πληθυσμού και ποια είναι πιο αποτελεσματική;

$\alpha)\ T_1=\frac{1}{10}(X_1+2X_2+7X_3)$

$\beta)\ T_2=\frac{1}{9}(X_1+2X_2+2X_3)$

$\gamma)\ T_3=\frac{1}{7}(6X_1+2X_2-5X_3)$

$\delta)\ T_4=\frac{1}{3}(X_1+X_2+X_3)$

Λύση

Αμεροληψία:

$$E(T_1)=E\left(\frac{1}{10}(X_1+2X_2+7X_3)\right)=\frac{1}{10}\big(E(X_1)+2E(X_2)+7E(X_3)\big)$$

$$=\frac{1}{10}(\mu+2\mu+7\mu)=\mu$$

T_1 είναι αμερόληπτη.

$$E(T_2)=E\left(\frac{1}{9}(X_1+2X_2+2X_3)\right)=\frac{1}{9}\big(E(X_1)+2E(X_2)+2E(X_3)\big)$$

$$=\frac{1}{9}(\mu+2\mu+2\mu)=\frac{5}{9}\mu$$

T_2 είναι μεροληπτική.

$$E(T_3)=E\left(\frac{1}{7}(6X_1+2X_2-5X_3)\right)=\frac{1}{7}\big(6E(X_1)+2E(X_2)-5E(X_3)\big)$$

$$=\frac{1}{7}(6\mu+2\mu-5\mu)=\frac{3}{7}\mu$$

T_3 είναι μεροληπτική.

$$E(T_4) = E\left(\frac{1}{3}(X_1 + X_2 + X_3)\right) = \frac{1}{3}\big(E(X_1) + E(X_2) + E(X_3)\big)$$

$$= \frac{1}{3}(\mu + \mu + \mu) = \mu$$

T_4 είναι αμερόληπτη.

Πιο αποτελεσματική θα είναι εκείνη που είναι αμερόληπτη και έχει μικρότερη διακύμανση.

$$\text{var}(T_1) = \text{var}\left(\frac{1}{10}(X_1 + 2X_2 + 7X_3)\right) = \frac{1}{100}\text{var}\big(X_1 + 2X_2 + 7X_3\big)$$

$$= \frac{1}{100}\left(\sigma^2 + 4\sigma^2 + 49\sigma^2\right) = \frac{54}{100}\sigma^2$$

$$\text{var}(T_4) = \text{var}\left(\frac{1}{3}(X_1 + X_2 + X_3)\right) = \frac{1}{9}\text{var}\big(X_1 + X_2 + X_3\big)$$

$$= \frac{1}{9}\left(\sigma^2 + \sigma^2 + \sigma^2\right) = \frac{1}{3}\sigma^2$$

T_4 είναι πιο αποτελεσματική.

Άσκηση 3

Η κατανάλωση βενζίνης για ένα τυχαίο δείγμα 6 αυτοκίνητων δίνεται πιο κάτω:

18,6 18,4 19,2 20,8 19,4 20,5.

Να κατασκευάσετε ένα 90% διάστημα εμπιστοσύνης για την μέση κατανάλωση στο πληθυσμό κάτω από την υπόθεση ότι η κατανομή του πληθυσμού είναι κανονική.

Λύση

Πρέπει να υπολογίσουμε το μέσο και την διακύμανση του δείγματος.

$$\overline{X} = \frac{1}{6}\sum X_i = 19{,}48$$

$$S_X^2 = \frac{1}{n-1}\left(\sum X_i - n\overline{X}^2\right) = 1{,}12$$

$$S_X = 1{,}06$$

Επειδή έχουμε κανονικό πληθυσμό με άγνωστη διακύμανση το Δ.Ε. δίνεται από,

$$\overline{X} - \frac{t_{n-1,\alpha/2}S_X}{\sqrt{n}} < \mu < \overline{X} + \frac{t_{n-1,\alpha/2}S_X}{\sqrt{n}} \quad (\text{με } \overline{X} = 19{,}48 \;\; S_X = 1{,}06 \;\; n = 6 \;\; \alpha = 0{,}10 \;\; t_{n-1,\alpha/2} = 2{,}015)$$

$$18{,}61 < \mu < 20{,}35$$

Άσκηση 4

Παίρνουμε ένα τυχαίο δείγμα 745 ατόμων που είχαν αποκτήσει βίντεο κάμερα για τουλάχιστον 12 μήνες και λιγότερο από 24 μήνες. Από τα μέλη του δείγματος, 321 δήλωσαν ότι μετά από την αγορά της βίντεο κάμερας πηγαίνουν λιγότερο συχνά στο Κινηματογράφο. Να κατασκευάσετε ένα 95% διάστημα εμπιστοσύνης για την αναλογία στο πληθυσμό.

Λύση

Έχουμε μεγάλο δείγμα οπότε χρησιμοποιούμε την κανονική κατανομή.

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$0{,}395 < p < 0{,}467$

όπου

$\hat{p} = \frac{321}{745} \approx 0{,}43$

$n = 745$

$\alpha = 0{,}05$

$z_{\alpha/2} = 1{,}96$

Άσκηση 5

Ακολουθεί ένα σύνολο εσόδων από πωλήσεις (σε χιλιάδες φύλλα) μιας εφημερίδας σε δύο γειτονικές πόλεις για μια περίοδο λίγων ημερών:

Πόλη Α: 25,13, 14, 19, 23, 30, 35, 29, 28, 17, 17, 16, 13, 18, 20

Πόλη Β: 10,12, 15, 13, 7, 6, 11, 5, 9, 14, 15, 18, 17, 16, 12, 12, 10, 11, 13, 14

Να κατασκευάσετε ένα 99% διάστημα εμπιστοσύνης της διαφοράς των μέσων εάν οι αρχικοί πληθυσμοί είναι κανονικοί και $\sigma_A^2 = 40, \sigma_B^2 = 14$..

Λύση

Έχουμε κανονικοί πληθυσμοί, μικρά δείγματα, γνωστές διακυμάνσεις.

$\overline{X}_A = 21{,}13$

$n_A = 15$

$\overline{X}_B = 12$

$n_B = 20$

$\overline{X}_A - \overline{X}_B = 9{,}13$

$\alpha = 0{,}01$

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

Τότε,

$$\overline{X}_A - \overline{X}_B - z_{\alpha/2}\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} < \mu_A - \mu_B < \overline{X}_A - \overline{X}_B + z_{\alpha/2}\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$$

$4{,}40 < \mu_A - \mu_B < 13{,}86$

Άσκηση 6

Ένα ζυθοποιείο ξέρει ότι η ποσότητα μπύρας σε ένα κουτί ακολουθεί μια κανονική κατανομή με τυπική απόκλιση 0,2 ουγκιές.
Παίρνοντας ένα τυχαίο δείγμα 25 κουτιών κατασκευάσθηκε ένα διάστημα εμπιστοσύνης για το μέσο του πληθυσμού ως εξής: $11{,}98 < \mu < 12{,}12$ Ποιο είναι το επίπεδο εμπιστοσύνης του παραπάνω διαστήματος;

Λύση

Έχουμε κανονικό πληθυσμό και γνωστή διακύμανση, το (1-α)% ΔΕ για το μέσο δίνεται από,

$\overline{X} - z_{\alpha/2}\dfrac{0{,}2}{\sqrt{25}} < \mu < \overline{X} + z_{\alpha/2}\dfrac{0{,}2}{\sqrt{25}}$. Το εύρος του διαστήματος είναι $2\left(z_{\alpha/2}\dfrac{0{,}2}{\sqrt{25}}\right)$,

Λύνοντας, $12{,}12 - 11{,}98 = 2\left(z_{\alpha/2}\dfrac{0{,}2}{\sqrt{25}}\right)$, έχουμε $z_{\alpha/2} = \dfrac{0{,}14}{2}\dfrac{5}{0{,}2} = 1{,}75$. Χρησιμοποιούμε το πίνακα της $N(0,1)$ και το επίπεδο εμπιστοσύνης είναι 91,98%($1-2(1-\Phi(1{,}75))$)

Άσκηση 7

Έστω ότι από έναν κανονικό πληθυσμό έχει ληφθεί δείγμα 4 μονάδων και οι συγκεκριμένες παρατηρήσεις είναι: 50, 60, 48, 74. Αν ο μέσος του πληθυσμού είναι μ=50, ζητείται να εκτιμηθεί το 95% διάστημα εμπιστοσύνης της διακύμανσης.

Λύση

Το διάστημα της διακύμανσης δίνεται από,

$$\frac{\sum (x_i - \mu)^2}{\chi^2_{n,\alpha/2}} < \sigma^2 < \frac{\sum (x_i - \mu)^2}{\chi^2_{n,1-\alpha/2}}$$

$39{,}5 < \sigma^2 < 916{,}7$

όπου $\alpha = 0{,}05$ $\chi^2_{4,\alpha/2} = 11{,}14$ $\chi^2_{4,1-\alpha/2} = 0{,}48$. Σημειώστε ότι οι βαθμοί ελευθερίας είναι 4 και όχι 3 επειδή χρησιμοποιούμε μ και όχι $\overline{X}$ στον τύπο της διακύμανσης του δείγματος.

Άσκηση 8

Προκειμένου να βρεθεί η διαφορά του βάρους μεταξύ των σπουδαστών της Ανώτατης Εμπορικής και του Πανεπιστήμιο Πειραιώς, έχει ληφθεί από την πρώτη δείγμα $n_1 = 1200$ σπουδαστών και από τη δεύτερη $n_2=250$ σπουδαστών.
Από τα δείγματα αυτά προέκυψε:
$\overline{X}_1 = 72\ kgr$ για τους σπουδαστές της Εμπορικής
$\overline{X}_2 = 70\ kgr$ για τους σπουδαστές του Πανεπ. Πειραιώς.

Αν Θεωρηθεί γνωστό ότι η κατανομή των βαρών στις δυο Σχολές είναι κανονικές και ότι $\sigma_1^2 = 100, \sigma_2^2 = 81$, ζητείται το διάστημα εμπιστοσύνης της διαφοράς των μέσων $\mu_1 - \mu_2$ σε επίπεδο σημαντικότητας 5%.

Λύση
Το ΔΕ δίνεται από

$$\overline{X}_1 - \overline{X}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < \overline{X}_1 - \overline{X}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$1{,}20 < \mu_1 - \mu_2 < 2{,}80$

όπου $z_{\alpha/2} = 1{,}96$.

Άσκηση 9

Επιθυμούμε να εκτιμήσουμε το μέσο βάρος των παραγομένων πεπονιών σε μια πειραματική καλλιέργεια. Ποιο πρέπει να είναι το απαιτούμενο μέγεθος του τυχαίο δείγματος, ώστε με πιθανότητα 99% το δειγματοληπτικό σφάλμα (το ανώτατο όριο σφάλματος της εκτιμήσεως) να είναι μικρότερο του 1 κιλό όταν η διακύμανση είναι $\sigma^2=4$.

Λύση
Έχουμε $\alpha = 0{,}01$ $z_{\alpha/2} = 2{,}58$. Το σφάλμα ε=1, τότε λύνοντας $\varepsilon = \frac{z_{\alpha/2}\sigma}{\sqrt{n}}$ έχουμε n=26,6

δηλαδή n=27.

Άσκηση 11

Μέχρι το έτος 1987 στις εξετάσεις στο μάθημα της Στατιστικής στη σχολή Ικάρων έχει διαπιστωθεί από παρατηρήσεις πολλών ετών, ότι η μέση βαθμολογία των σπουδαστών ήταν 87 μονάδες με τυπική απόκλιση σ=9 μονάδες. Στις εξετάσεις της περιόδου Ιανουαρίου του τρέχοντας έτους από τυχαίο δείγμα 36 σπουδαστών διαπιστώνεται μέση βαθμολογία 90 μονάδων. Με την προϋπόθεση ότι η τυπική απόκλιση σ=9 ισχύει, μπορούμε να ισχυριστούμε ότι η μέση βαθμολογία δεν έχει αυξηθεί; Επίπεδο σημαντικότητας 5%.

Λύση

Η υπόθεση θα είναι $\begin{matrix} H_0 : \mu = 87 \\ H_1 : \mu > 87 \end{matrix}$

Επειδή σ είναι γνωστή και έχουμε μεγάλο δείγμα θα χρησιμοποιήσουμε την κανονική κατανομή.

Στατιστικό στοιχείο: $\dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} = \dfrac{90-87}{9/6} = 2$

Κριτική τιμή (έλεγχο μιας ουράς): $z_\alpha = z_{5\%} = 1{,}645$

Επειδή 2>1,645 απορρίπτεται ότι η βαθμολογία έχει παραμένει σταθερή (απορ. H_0)

<u>Άσκηση 10</u>

Από έναν κανονικό πληθυσμό πήραμε ένα δείγμα με τιμές 30, 40, 28, 54. Ζητείται να ελεγχθεί η υπόθεση:
H_0: $\sigma^2 = 30$
H_1: $\sigma^2 > 30$
σε επίπεδο σημαντικότητας 5%.

Λύση

Στατιστικό στοιχείο: $\dfrac{(n-1)S^2}{\sigma_0^2} = \dfrac{3(141{,}3)}{30} = 14{,}13$ όπου $S^2 = \dfrac{\sum (X_i - \overline{X})^2}{n-1} = 141{,}3$

Κριτική τιμή: οι βαθμοί ελευθερίας είναι v=n-1=4-1=3, α=5%, οπότε $\chi^2_{\nu,\alpha} = \chi^2_{3,5\%} = 7{,}81$

Επειδή 14,13>7,81 απορρίπτεται η H_o.

<u>Άσκηση 11</u>

Από δύο κανονικούς πληθυσμούς πήραμε δυο δείγματα Α και Β, τα οποία μας έδωσαν τις παρακάτω τιμές:
X_i: 0,114 0,127 0,143 0,132
Y_i: 0,131 0,107 0,104 0,111 0,108 0,110

Αν υποθέσουμε ότι $\sigma_1^2 = \sigma_2^2$ να ελεγχθεί η υπόθεση , δηλαδή και τα δυο δείγματα προέρχονται από δύο πληθυσμούς με τον ίδιο μέσο ως προς την εναλλακτική υπόθεση $\mu_1 \neq \mu_2$.

Λύση

$\bar{X} = 0{,}129$

$\bar{Y} = 0{,}1118$

$n_1 = 4$

$n_2 = 6$

$$S_1^2 = \frac{\sum (X_i - \bar{X})^2}{n_1 - 1} = 0{,}000220$$

$$S_2^2 = \frac{\sum (Y_i - \bar{Y})^2}{n_2 - 1} = 0{,}000094$$

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 0{,}0001412$$

$$t = \frac{0{,}129 - 0{,}1118}{\sqrt{0{,}0001412\left(\frac{1}{4} + \frac{1}{6}\right)}} = 2{,}24$$

Οι βαθμοί ελευθερίας είναι $\nu = n_1 + n_2 - 2 = 8$, από τους πίνακες $t_{8,5\%} = 2{,}306$. Επειδή $2{,}24 < 2{,}306$ δεν απορρίπτουμε την H_o σε επίπεδο 5%.

<u>Άσκηση 12</u>

Έστω απο έναν κανονικό πληθυσμό του οποίου δεν γνωρίζουμε τη διακύμανση παίρνουμε δείγμα μεγέθους $n = 14$. Απο το δείγμα αυτό βρέθηκε να έχει δειγματικό μέσο ίσο με 52,52 και τυπική απόκληση ίση με 3,37. Ζητείται το 95% διάστημα εμπιστοσύνης για τον μέσο του πληθυσμού.

Λύση

Επειδή έχουμε αγνωστη πληθυσμιακή διακύμανση και $n = 14 < 30$, χπησιμοποιούμε την κατανομή $t - Student$, με $n - 1 = 13$ βαθμούς ελευθερίας, για να κατασκευάσουμε το 95% διάστημα εμπιστοσύνης για τον μέσο του πληθυσμού. Έτσι έχουμε

$$\overline{X} - t_{n-1,\alpha/2} \frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{n-1,\alpha/2} \frac{S}{\sqrt{n}}$$

$$50{,}58 < \mu < 54{,}47$$

όπου $t_{n-1,\alpha/2} = 2{,}16$.

Άσκηση 13

Προκειμένου να εκτιμηθεί το ποσοστό p των καπνιστών ενός πληθυσμού, ελήφθησαν απο αυτόν 2 τυχαία δείγματα μεγέθους 20 και 80 αντιστοίχα. Εαν p_1 και p_2 τα αντίστοιχα δειγματικά ποσοστά τους, ποιά από τις εκτιμήτριες συναρτήσεις

$$T_1 = \frac{1}{5}(p_1 + 4p_2), \quad T_2 = \frac{1}{2}(p_1 + p_2)$$

είναι πιο αποτελεσματικές.

Λύση

$$E(T_1) = \frac{1}{5}(E(p_1) + 4E(p_2)) = \frac{5p}{5} = p, \quad E(T_2) = \frac{1}{2}(E(p_1) + E(p_2)) = \frac{2p}{2} = p$$

άρα και οι δύο εκτιμήτριες είναι αμερόληπτες.

$$V(T_1) = \frac{1}{25}(V(p_1) + 16V(p_2)) = \frac{1}{25}(\frac{p(1-p)}{20} + 16\frac{p(1-p)}{80}) = 0{,}01p(1-p)$$

$$V(T_2) = \frac{1}{4}(V(p_1) + V(p_2)) = \frac{1}{4}(\frac{p(1-p)}{20} + \frac{p(1-p)}{80}) = 0{,}0156p(1-p)$$

Άρα η T_1 είναι η ποιό αποτελεσματική εκτιμήτρια.