ΛΥΜΕΝΕΣ ΑΣΚΗΣΕΙΣ ΣΤΑΤΙΣΤΙΚΗΣ - ΠΙΘΑΝΟΤΗΤΩΝ

ΑΣΚΗΣΗ 1

Α. Θα συμβολίζουμε με Α τα αποτελέσματα του δείγματος των τιμών ανοίγματος και με Κ τα αποτελέσματα των τιμών κλεισίματος.

Μέσος Όρος Α $\hat{x}_A$ = Άθροισμα όλων των τιμών / πλήθος παρατηρήσεων = (47.820 + 47.610 + … + 49.570) / 20 = 961.319 / 20 = **48.066**

Τυπική Απόκλιση $s_A = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{x}_A)^2} = \sqrt{\frac{1}{19}\sum_{i=1}^{n}(x_i - 48.066)^2} =$

$\sqrt{\frac{1}{19}[(47.820 - 48.066)^2 + (47.610 - 48.066)^2 + \cdots + (49.570 - 48.066)^2} = \sqrt{0.974} = \mathbf{0.987}$

όπου x_i η τιμή της i παρατήρησης και $\hat{x}_A$ ο μέσος των τιμών ανοίγματος.

Η **Διάμεσος Α** είναι η μεσαία παρατήρηση. Σε αυτήν την περίπτωση έχουμε 20 στοιχεία (άρτιο πλήθος), οπότε η διάμεσος βρίσκεται ανάμεσα στην $10^η$ και $11^η$ παρατήρηση. Αρχικά ταξινομούμε τα δεδομένα μας σε αύξουσα σειρά για διευκόλυνση στους υπολογισμούς μας:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
46.334	46.360	46.450	46.955	47.450	47.610	47.700	47.820	48.119	48.279	48.305	48.345	48.369	48.384	48.785	48.794	48.910	49.205	49.570	49.575

δηλαδή η τιμή (48.279+48.305)/2 = **48.292**

Το **Πρώτο Τεταρτημόριο Α** είναι η τιμή μέχρι την οποία περιλαμβάνεται το 25% των διατεταγμένων τιμών των παρατηρήσεων (από την μικρότερη στην μεγαλύτερη). Η θέση του πρώτου τεταρτημόριου δίνεται από τον τύπο (n+1)·p/100 όπου p το ποσοστό που επιθυμούμε. Επόμενος για n=20 και p=25 το $1^ο$ τεταρτημόριο είναι στη θέση 5,25. Στην θέση αυτή είναι η τιμή 47.450+0,25·(47.610-47.450) = **47.490.**

Το **Τρίτο Τεταρτημόριο Α** είναι η τιμή μέχρι την οποία περιλαμβάνεται το 75% των διατεταγμένων τιμών των παρατηρήσεων (από την μικρότερη στην μεγαλύτερη). Η θέση του πρώτου τεταρτημόριου δίνεται από τον τύπο (n+1)·p/100 όπου p το ποσοστό που επιθυμούμε. Επόμενος για n=20 και p=75 το $3^ο$ τεταρτημόριο είναι στη θέση 15,75. Στην θέση αυτή είναι η τιμή 48.785+0,75·(48.794-48.785) = **48.792.**

Συντελεστής Ασυμμετρίας Α $= \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_A)^3}{{s_A}^3} = \frac{\frac{1}{20}(-6.362)}{0.987^3} = \mathbf{-0.331}$

Μέσος Όρος Κ $\hat{x}_K$ = Άθροισμα όλων των τιμών / πλήθος παρατηρήσεων = (47.764 + 46.849 + … + 49.005) / 20 = 961.692 / 20 = **48.085**

Τυπική Απόκλιση $s_K = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{x}_K)^2} = \sqrt{\frac{1}{19}\sum_{i=1}^{n}(x_i - 48.085)^2} =$

$$\sqrt{\frac{1}{19}[(47.764 \ - 48.085)^2 + (46.849 - 48.085)^2 + \cdots + (49.005 - 48.085)^2} = \sqrt{1.275} = \mathbf{1.129}$$

όπου x_i η τιμή της i παρατήρησης και $\hat{x}_K$ ο μέσος των τιμών ανοίγματος.

Η **Διάμεσος Κ** είναι η μεσαία παρατήρηση. Σε αυτήν την περίπτωση έχουμε 20 στοιχεία (άρτιο πλήθος), οπότε η διάμεσος βρίσκεται ανάμεσα στην $10^{η}$ και $11^{η}$ παρατήρηση. Αρχικά ταξινομούμε τα δεδομένα μας σε αύξουσα σειρά για διευκόλυνση στους υπολογισμούς μας:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
45.715	46.320	46.400	46.849	47.125	47.674	47.764	47.770	48.105	48.235	48.430	48.450	48.505	48.825	48.935	49.005	49.020	49.255	49.430	49.880

δηλαδή η τιμή (48.235+48.430)/2 = **48.333.**

Το **Πρώτο Τεταρτημόριο Κ** είναι η τιμή μέχρι την οποία περιλαμβάνεται το 25% των διατεταγμένων τιμών των παρατηρήσεων (από την μικρότερη στην μεγαλύτερη). Η θέση του πρώτου τεταρτημόριου δίνεται από τον τύπο (n+1)·p/100 όπου p το ποσοστό που επιθυμούμε. Επόμενος για n=20 και p=25 το $1^{ο}$ τεταρτημόριο είναι στη θέση 5,25. Στην θέση αυτή είναι η τιμή 47.125+0,25·(47.674-47.125) = **47.262.**

Το **Τρίτο Τεταρτημόριο Κ** είναι η τιμή μέχρι την οποία περιλαμβάνεται το 75% των διατεταγμένων τιμών των παρατηρήσεων (από την μικρότερη στην μεγαλύτερη). Η θέση του πρώτου τεταρτημόριου δίνεται από τον τύπο (n+1)·p/100 όπου p το ποσοστό που επιθυμούμε. Επόμενος για n=20 και p=75 το $3^{ο}$ τεταρτημόριο είναι στη θέση 15,75. Στην θέση αυτή είναι η τιμή 48.935+0,75·(49.005-48.935) = **48.988.**

Συντελεστής Ασυμμετρίας Κ = $\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_K)^3}{{s_K}^3} = \frac{\frac{1}{20}(-13.871)}{1.129^3} = \mathbf{-0.481}$

Να σχολιάστε την ασυμμετρία της κατανομής του δείγματος με βάση το μέσο, τη διάμεσο και το συντελεστή ασυμμετρία:

<u>Και για τα δύο δείγματα</u>: Επειδή η διάμεσος είναι μεγαλύτερη από το μέσο, υποδεικνύεται μια αρνητική ασυμμετρία. Ο αρνητικός συντελεστής ασυμμετρίας επιβεβαιώνει την ασυμμετρία προς τα αριστερά της κατανομής. Ο μέσος όρος είναι κοντά στη διάμεσο, αλλά η αρνητική ασυμμετρία υποδεικνύει ότι υπάρχουν περισσότερες ακραίες τιμές στο αριστερό άκρο της κατανομής του δείγματος, ενώ η ουρά της κατανομής εκτείνεται προς τα αριστερά του μέσου. Οι περισσότερες τιμές στο δείγμα συγκεντρώνονται γύρω από το χαμηλότερο άκρο της κατανομής τους, με λιγότερες τιμές να παρατηρούνται προς το ανώτερο άκρο.

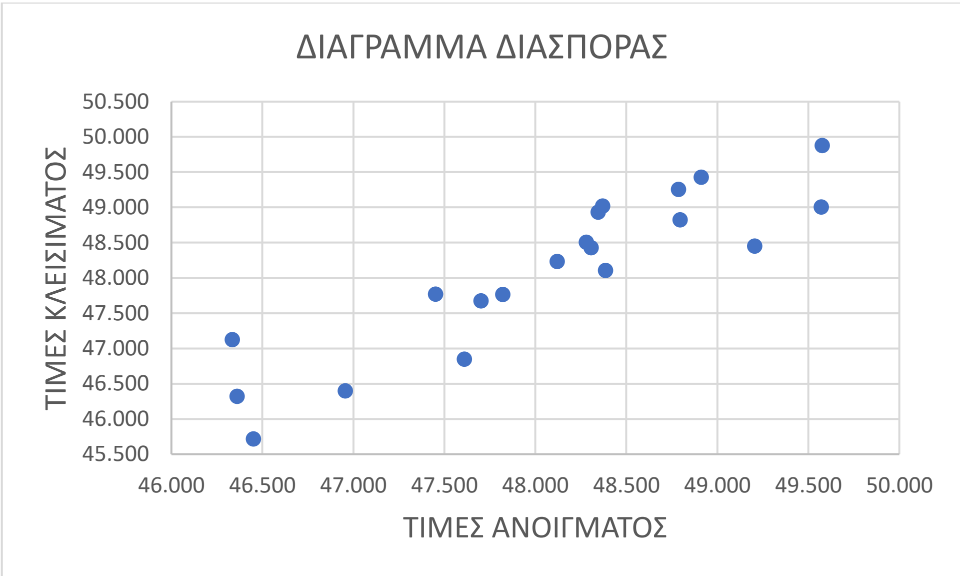
Β.

Συντελεστής Μεταβλητότητας Α = $\frac{s_A}{\hat{x}_A}100 = \frac{0.987}{48.066}100 = \mathbf{2.053}\%$

Συντελεστής Μεταβλητότητας Κ = $\frac{s_K}{\hat{x}_K}100 = \frac{1.129}{48.085}100 = \mathbf{2.349}\%$

Ο Συντελεστής Μεταβλητότητας Κ είναι μεγαλύτερος από τον Συντελεστή Μεταβλητότητας Α κατά 0.296%. Αυτό σημαίνει ότι η διασπορά των τιμών για τις τιμές κλεισίματος είναι μεγαλύτερη σε σχέση με τις τιμές ανοίγματος. Η μεγαλύτερη διασπορά στις τιμές κλεισίματος υποδηλώνουν ότι οι τιμές διαφέρουν περισσότερο μεταξύ τους συγκριτικά με τις τιμές ανοίγματος. Επομένως υπάρχει λιγότερη ομοιομορφία στις τιμές κλεισίματος σε σχέση με τις τιμές ανοίγματος που συγκριτικά είναι περισσότερο ομοιόμορφες.

Γ.



Δ.

Από το παραπάνω διάγραμμα διασποράς παρατηρούμε ότι, όσο αυξάνεται η τιμή ανοίγματος, τόσο πιο πιθανό είναι να αυξηθεί και η τιμή κλεισίματος, και αντίστροφα. Αυτό υποδεικνύει μια γραμμική σχέση μεταξύ αυτών των δύο μεταβλητών. Η σχέση αυτή παραιτείται από κάτω αριστερά προς τα πάνω δεξιά, όπου τα σημεία τείνουν να διαμορφώνουν αρκετά καλά μια νοητή ευθεία. Επομένως η συσχέτιση αυτή φαίνεται να είναι πολύ ισχυρή πράγμα το οποίο μπορεί να επιβεβαιωθεί και από τον συντελεστή συσχέτισης Pearson που στην προκειμένη περίπτωση έχει τιμή 0,9.

Η τιμή κλεισίματος μιας ημέρας στο χρηματιστήριο δεν είναι πάντα ίδια με την τιμή ανοίγματος της επόμενης ημέρας για διάφορους λόγους:

Αρχικά, η τιμή μπορεί να επηρεαστεί από τις τρέχουσες εξελίξεις σε διεθνές επίπεδο κατά τη διάρκεια της νύχτας όταν το χρηματιστήριο είναι κλειστό. Οι τιμές επίσης μπορεί να καθορίζονται από τη συνεχή αλληλεπίδραση μεταξύ της προσφοράς και της ζήτησης με αποτέλεσμα η τιμή κλεισίματος μιας ημέρας να επηρεαστεί από πολλούς παράγοντες κατά τη διάρκεια της συνεδρίας, όπως νέες ανακοινώσεις, ανακατατάξεις στις εταιρείες, οικονομικές αλλαγές, κλπ. Τέλος υπάρχει η πιθανότητα οι επενδυτές να αντιδρούν διαφορετικά στις πληροφορίες και τα γεγονότα που προέκυψαν κατά τη διάρκεια μιας συνεδρίας το οποίο μπορεί να οδηγήσει σε μια διαφορετική αξιολόγηση της εταιρείας την επόμενη ημέρα.

ΑΣΚΗΣΗ 2

Α.

(α) Ζητείται η πιθανότητα ένα παιδί να ακούει ροκ μουσική ή να είναι αγόρι.

Έστω Ρ το ενδεχόμενο να ακούει Ροκ και Α το ενδεχόμενο να είναι αγόρι. Ψάχνουμε την πιθανότητα:

$$\Pr(\mathrm{P}\cup \mathrm{A})=\Pr(\mathrm{P})+\Pr(\mathrm{A})-\Pr(\mathrm{P}\cap \mathrm{A}) = \frac{31+35}{200}+\frac{26+26+31}{200}-\frac{31}{200}=\frac{35+26+26+31}{200}=\mathbf{0.59}\ (\mathbf{59}\%)$$

(β) Αν ένα παιδί προτιμά το έντεχνο είδος μουσικής, ποια είναι η πιθανότητα να είναι κορίτσι;

Έστω Ε το ενδεχόμενο να προτιμά έντεχνο και Κ το ενδεχόμενο να είναι κορίτσι. Ψάχνουμε την πιθανότητα δεσμευμένη πιθανότητα:

$$\Pr(\mathrm{K}|\mathrm{E})=\frac{P(K\cap E)}{P(E)}=\frac{\frac{44}{200}}{\frac{26+44}{200}}=\frac{44}{26+44}=\frac{44}{70}=\mathbf{0.62857}\ (\mathbf{62.857}\%)$$

(γ) Ο διευθυντής του σχολείου ισχυρίζεται ότι το ενδεχόμενο ένα παιδί να ακούει λαϊκή μουσική είναι ανεξάρτητο του ενδεχομένου να είναι κορίτσι. Ισχύει ο ισχυρισμός του διευθυντή;

Έστω Λ το ενδεχόμενο να ακούει λαϊκή μουσική και Κ το ενδεχόμενο να είναι κορίτσι. Αν όντως τα ενδεχόμενα είναι ανεξάρτητα, τότε θα ισχύει:

$$\Pr(\Lambda \cap \mathrm{K}) = \Pr(\Lambda) \cdot \Pr(\mathrm{K})$$

Έχουμε:

$$\Pr(\Lambda \cap \mathrm{K}) = \frac{38}{200} = 0{,}19 \quad , \quad \Pr(\Lambda) = \frac{26+38}{200} = 0{,}32 \quad , \quad \Pr(\mathrm{K}) = \frac{38+44+35}{200} = 0{,}585$$

Όμως $\Pr(\Lambda) \cdot \Pr(\mathrm{K}) = 0{,}32 \cdot 0{,}585 = 0{,}1872 \neq 0{,}19 = \Pr(\Lambda \cap \mathrm{K})$ άρα τα ενδεχόμενα δεν είναι ανεξάρτητα.

Β.

(α)

Έστω Α το ενδεχόμενο να έχει κατασκευαστεί ένα ελατήριο από την Α μηχανή,

Β το ενδεχόμενο να έχει κατασκευαστεί ένα ελατήριο από την Β μηχανή και

Γ το ενδεχόμενο να έχει κατασκευαστεί ένα ελατήριο από την Γ μηχανή.

Τα ενδεχόμενα Α,Β,Γ είναι ξένα μεταξύ τους και αποτελούν μια διαμέριση διότι $\Omega = A \cup B \cup \Gamma$.

Αν Λ είναι το ενδεχόμενο ένα ελατήριο να είναι λειτουργικό τότε:

- $Pr(\Lambda|A)$ ορίζεται η πιθανότητα ένα ελατήριο να είναι λειτουργικό αν παράχθηκε από τη μηχανή Α = 1-0.01 = 0.99 (99%).
- $Pr(\Lambda|B)$ ορίζεται η πιθανότητα ένα ελατήριο να είναι λειτουργικό αν παράχθηκε από τη μηχανή Β 1-0.04 = 0.96 (96%).
- $Pr(\Lambda|\Gamma)$ ορίζεται η πιθανότητα ένα ελατήριο να είναι λειτουργικό αν παράχθηκε από τη μηχανή Γ, 1-0.06 = 0.94 (94%).

Για να βρούμε την πιθανότητα P(Λ) χρησιμοποιήσουμε το Θεώρημα Ολικής Πιθανότητας:

$$Pr(\Lambda) = Pr(\Lambda|A)\cdot Pr(A) + Pr(\Lambda|B)\cdot Pr(B) + Pr(\Lambda|\Gamma)\cdot P(\Gamma)$$

Επομένως έχουμε: $Pr(\Lambda) = 0.99 \cdot 0.25 + 0.96 \cdot 0.35 + 0.94 \cdot 0.40 =$ **0.9595 (95.95%)**

(β)

Ορίζουμε τις πιθανότητες:

- $Pr(A)$ είναι η πιθανότητα ένα παιδί να είναι μαθητής, η οποία είναι 0.45 (45%),
- $Pr(K)$ είναι η πιθανότητα ένα παιδί να είναι μαθήτρια, η οποία είναι 0.55 (55%) και
- Pr(Δ) είναι η πιθανότητα ένα παιδί να έχει ΔΜΣ >18.

Η πιθανότητα που ψάχνουμε είναι $Pr(K|\Delta)$, δηλαδή η πιθανότητα ένα παιδί να είναι μαθήτρια δεδομένου ότι έχει ΔΜΣ > 18, μπορεί να υπολογιστεί από το θεώρημα του Bayes:

$$Pr(K|\Delta) = \frac{Pr(\Delta|K)\cdot Pr(K)}{P(\Delta)}$$

Για την πιθανότητα $P(\Delta)$ χρησιμοποιούμε όπως και πριν το Θεώρημα Ολικής Πιθανότητας:

$$Pr(\Delta) = Pr(\Delta|A)\cdot Pr(A) + Pr(\Delta|K)\cdot Pr(K)$$

Έχουμε:

$$Pr(\Delta) = 0.05\cdot 0.45 + 0.02\cdot 0.55 = 0.0335\ (3.35\%)$$

Επομένως:

$$Pr(K|\Delta) = \frac{0.02\cdot 0.55}{0.0335} = \mathbf{0.32836\ (32.836}\%)$$

ΑΣΚΗΣΗ 3

Α.

(α)

Για το ταμείο T_1 υπάρχουν διαθέσιμες 9 επιλογές, για το ταμείο T_2 υπάρχουν 8 διαθέσιμες επιλογές (διότι ο ένας υπάλληλος ήδη έχει δεσμεύσει το $1^{ο}$ ταμείο), για το ταμείο T_3 υπάρχουν 7 διαθέσιμες επιλογές και για το ταμείο T_4 υπάρχουν 6 διαθέσιμες επιλογές. Συνεπώς, οι συνολικοί διαφορετικοί τρόποι τοποθέτησής τους είναι:

$$9 \cdot 8 \cdot 7 \cdot 6 = \mathbf{3024} \text{ τρόποι}$$

(β)

Ο αριθμός των διαφορετικών τρόπων για την δέσμευση των ταμείων με τον ίδιο τρόπο όπως και πριν θα ήταν:

$$9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 9! = \mathbf{362880} \text{ τρόποι}$$

Β.

(α)

Συνολικά έχουμε 16 υπαλλήλους και θέλουμε να επιλέξουμε 6 από αυτούς επομένως οι δυνατοί τρόποι επιλογής τους είναι:

$$\frac{16!}{6! \cdot 10!} = \mathbf{8008} \text{ τρόποι}$$

(β)

Πρέπει να διαλέξουμε 3 γυναίκες από τις 6 δηλαδή με $\frac{6!}{3! \cdot 3!}$ τρόπους και 3 άνδρες από τους 10 με $\frac{10!}{3! \cdot 7!}$ τρόπους, οπότε όλες μαζί οι επιλογές θα είναι:

$$\frac{6!}{3! \cdot 3!} \cdot \frac{10!}{3! \cdot 7!} = \frac{4 \cdot 5 \cdot 6}{3!} \cdot \frac{8 \cdot 9 \cdot 10}{3!} = \frac{4 \cdot 5}{1} \cdot \frac{4 \cdot 3 \cdot 10}{1} = \mathbf{2400} \text{ τρόποι}$$

(γ)

Πρέπει να επιλεγούν 2 γυναίκες και 4 άνδρες ή 1 γυναίκα και 5 άνδρες ή 6 άνδρες. Ο αριθμός των τρόπων που μπορούμε να κάνουμε αυτή την επιλογή είναι:

$$\frac{6!}{2! \cdot 4!} \cdot \frac{10!}{4! \cdot 6!} + \frac{6!}{1! \cdot 5!} \cdot \frac{10!}{5! \cdot 5!} + \frac{10!}{6! \cdot 4!} = \frac{5 \cdot 6}{2} \cdot \frac{7 \cdot 8 \cdot 9 \cdot 10}{2 \cdot 3 \cdot 4} + 6 \cdot \frac{6 \cdot 7 \cdot 8 \cdot 9 \cdot 10}{2 \cdot 3 \cdot 4 \cdot 5} + \frac{7 \cdot 8 \cdot 9 \cdot 10}{2 \cdot 3 \cdot 4}$$

$$15 \cdot \frac{7 \cdot 3 \cdot 10}{1} + 6 \cdot \frac{7 \cdot 2 \cdot 9 \cdot 2}{1} + \frac{7 \cdot 3 \cdot 10}{1} = \mathbf{4872} \text{ τρόποι}$$