*ΕΡΓΑΣΙΑ ΣΤΟ ΜΑΘΗΜΑ ΤΗΣ ΠΕΡΙΓΡΑΦΙΚΗΣ ΣΤΑΤΙΣΤΙΚΗΣ*

*ΚΩΝΣΤΑΝΤΙΝΟΣ-ΕΜ. ΖΙΑΚΑΣ*

*ΑΡΙΘΜΟΣ ΜΗΤΡΟΥ : Σ/10058*

*Εξάμηνο 1ο*

Στην εργασία αυτή θα παρουσιάσουμε και θα μελετήσουμε κάποια δεδομένα μέσα από ένα δείγμα 48 ασφαλισμένων σε ένα συγκεκριμένο πρόγραμμα και τον αριθμό των ημερών που νοσηλεύτηκαν στο διάστημα 5 ετών.

Στο παρακάτω διάγραμμα παρουσιάζονται τα δεδομένα αυτά

Διάγραμμα 1.1

Στο δείγμα μας υπάρχουν:

* 7 άτομα που δεν νοσηλεύτηκαν καμία ημέρα
* 5 άτομα που νοσηλεύτηκαν για 1 ημέρα
* 6 άτομα που νοσηλεύτηκαν για 2 ημέρες
* 6 άτομα που νοσηλεύτηκαν για 3 ημέρες
* 6 άτομα που νοσηλεύτηκαν για 4 ημέρες
* 5 άτομα που νοσηλεύτηκαν για 5 ημέρες
* 2 άτομα που νοσηλεύτηκαν για 6 ημέρες
* 1 άτομο που νοσηλεύτηκε για 7 ημέρες
* 1 άτομο που νοσηλεύτηκε για 8 ημέρες
* 1 άτομο που νοσηλεύτηκε για 9 ημέρες
* 1 άτομο που νοσηλεύτηκε για 10 ημέρες
* 2 άτομο που νοσηλεύτηκαν για 12 ημέρες
* 1 άτομο που νοσηλεύτηκε για 16 ημέρες
* 1 άτομο που νοσηλεύτηκε για 17 ημέρες
* 1 άτομο που νοσηλεύτηκε για 19 ημέρες
* 1 άτομο που νοσηλεύτηκε για 20 ημέρες
* 1 άτομο που νοσηλεύτηκε για 23 ημέρες

Το σύνολο των ατόμων είναι 48. Eπίσης,

Το πρώτο πράγμα που παρατηρούμε είναι η μέγιστη και η ελάχιστη τιμή του διαγράμματος.Xmax=23, Xmin=0 .

Στον παρακάτω πίνακα παρουσιάζονται τα δεδομένα

Πίνακας 1.1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *ΑΡΙΘΜΟΣ ΗΜΕΡΩΝ* | *ΑΤΟΜΑ*  *ΑΠΟΛΥΤΕΣ ΣΥΧΝΟΤΗΤΕΣ fi* | *ΑΠΟΛΥΤΕΣ ΑΘΡΟΙΣΤΙΚΕΣ ΣΥΧΝΟΤΗΤΕΣ Fi* | *ΣΥΧΝΟΤΗΤΕΣ* | *ΠΟΣΟΣΤΑ ΣΥΧΝΟΤΗΤΩΝ* | *ΑΘΡΙΣΤΟΙΚΕΣ ΣΥΧΝΟΤΗΤΕΣ* | *ΠΟΣΟΣΤΑ ΑΘΡΟΙΣΤΙΚΩΝ ΣΥΧΝΟΤΗΤΩΝ* |
| 0 | 7 | 7 | 0,146 | 14,6 | 0,146 | 14,6 |
| 1 | 5 | 12 | 0,104 | 10,4 | 0,250 | 25,0 |
| 2 | 6 | 18 | 0,125 | 12,5 | 0,375 | 37,5 |
| 3 | 6 | 24 | 0,125 | 12,5 | 0,500 | 50,0 |
| 4 | 6 | 30 | 0,125 | 12,5 | 0,625 | 62,5 |
| 5 | 5 | 35 | 0,104 | 10,4 | 0,729 | 72,9 |
| 6 | 2 | 37 | 0,042 | 4,2 | 0,771 | 77,1 |
| 7 | 1 | 38 | 0,021 | 2,1 | 0,792 | 79,2 |
| 8 | 1 | 39 | 0,021 | 2,1 | 0,813 | 81,3 |
| 9 | 1 | 40 | 0,021 | 2,1 | 0,833 | 83,3 |
| 10 | 1 | 41 | 0,021 | 2,1 | 0,854 | 85,4 |
| 12 | 2 | 43 | 0,042 | 4,2 | 0,896 | 89,6 |
| 16 | 1 | 44 | 0,021 | 2,1 | 0,917 | 91,7 |
| 17 | 1 | 45 | 0,021 | 2,1 | 0,938 | 93,8 |
| 19 | 1 | 46 | 0,021 | 2,1 | 0,958 | 95,8 |
| 20 | 1 | 47 | 0,021 | 2,1 | 0,979 | 97,9 |
| 23 | 1 | 48 | 0,021 | 2,1 | 1 | 100 |
| *ΣΥΝΟΛΟ* | 48 | - | 1 | 100 | - | - |

* O αριθμητικός μέσος είναι: *\bar{x} \!\,*= ==5,18
* Ο συνολικός αριθμός των ατόμων είναι **48** (ζυγός) άρα η **Διάμεσος** είναι το ***ημιάθροισμα*** των 2 μεσαίων παρατηρήσεων έχοντας τοποθετήσει τα δεδομένα μας σε ***αύξουσα σειρά***, δηλαδή στο δείγμα μας η 24η και η 25η δια 2 σε διατεταγμένη σειρά.

m= = = = 3,5

όπου n το συνολικό πλήθος του δείγματος μας (n=48).

* Η επικρατούσα τιμή είναι η τιμή με την μεγαλύτερη συχνότητα δηλαδή η τιμή 0 στον πίνακα 1.1 .
* Το ενδοτεταρτομοριακό εύρος (IQR) Αναφέρεται στην διαφορά του 3ου με του 1ου τεταρτημορίου. (Q3-Q1). Χωρίζουμε δηλαδή το δείγμα σε 4 τεταρτημόρια ως προς τα ποσοστά. Το πρώτο τεταρτημόριο είναι το σημείο ***κάτω*** από το οποίο βρίσκεται το ***25%*** των διατεταγμένων τιμών του δείγματος.

Δηλαδή στον πίνακα

Πίνακας 1.2

|  |  |  |  |
| --- | --- | --- | --- |
| *i* | *ΑΡΙΘΜΟΣ ΗΜΕΡΩΝ* | *ΑΤΟΜΑ* | *ΠΟΣΟΣΤΑ ΣΥΧΝΟΤΗΤΩΝ* |
| 1 | 0 | 7 | 14,6 |
| 2 | 1 | 5 | 10,4 |
| ΣΥΝΟΛΟ | | **12** | **25** |

Άρα το σημείο κάτω από το οποίο βρίσκεται το 25% των παρατηρήσεων του δείγματος μας είναι τα 12 άτομα.

Το τρίτο τεταρτημόριο είναι το σημείο ***πάνω*** από το οποίο βρίσκεται το ***25%*** των διατεταγμένων τιμών του δείγματος. Στον παρακάτω πίνακα έχουμε την διατεταγμένη αναπαράσταση του 3ου τεταρτημορίου.

Πίνακας 1.3

|  |  |  |
| --- | --- | --- |
| *i* | *ΑΡΙΘΜΟΣ ΗΜΕΡΩΝ xi* | *ΑΤΟΜΑ* |
| 8 | 7 | 1 |
| 9 | 8 | 1 |
| 10 | 9 | 1 |
| 11 | 10 | 1 |
| 12 | 12 | 2 |
| 13 | 16 | 1 |
| 14 | 17 | 1 |
| 15 | 19 | 1 |
| 16 | 20 | 1 |
| 17 | 23 | 1 |
|  | ***ΣΥΝΟΛΟ*** | 11 |

Αθροίζοντας τα άτομα και τα ποσοστά βρίσκουμε 11 άτομα και το συνολικό ποσοστό στο πίνακα ποσοστιαίων συχνοτήτων 22,9%. Το αποτέλεσμα αυτό όμως δεν εξυπηρετεί το 3ο τεταρτημόριο καθώς χρειαζόμαστε το 25% των παρατηρήσεων. Παρατηρούμε στη συνέχεια ότι στον πίνακα 1.5, ότι 2 άτομα νοσηλεύτηκαν για 6 ημέρες (i=7). Το ποσοστό του κάθε ατόμου με βάση τον πίνακα 2 είναι 2,1%. Το ποσοστό αυτό είναι αυτό που μας ‘’*λείπει*’’ ώστε να έχουμε το πλήρες πλέον 25% του 3ου τεταρτημορίου. Επαληθεύοντας (25%-22,9%=2,1%). Άρα θα χρειαστούμε την μια παρατήρηση από τις δύο στα άτομα που νοσηλεύτηκαν για 6 ημέρες.

Πίνακας 1.4

|  |  |  |
| --- | --- | --- |
| *i* | *ΑΡΙΘΜΟΣ ΗΜΕΡΩΝ xi* | *ΑΤΟΜΑ* |
| 1 | 0 | 7 |
| 2 | 1 | 5 |
| 3 | 2 | 6 |
| 4 | 3 | 6 |
| 5 | 4 | 6 |
| 6 | **5** | **5** |
| 7 | **6** | **2** |
| 8 | 7 | 1 |
| 9 | 8 | 1 |
| 10 | 9 | 1 |
| 11 | 10 | 1 |
| 12 | 12 | 2 |
| 13 | 16 | 1 |
| 14 | 17 | 1 |
| 15 | 19 | 1 |
| 16 | 20 | 1 |
| 17 | 23 | 1 |
|  | ***ΣΥΝΟΛΟ*** | 48 |

Άρα τώρα το πλήθος δεν είναι 11 όπως αναφέραμε στην 1η παρουσίαση του 3ου τεταρτημορίου αλλά 12 και το ποσοστό αντίστοιχα 25%.Άρα το ενδοτεταρτομοριακό εύρος είναι IQR=Q1-Q3=12-12=0.

* Η διακύμανση, S2*=* . Στον παρακάτω πίνακα παρουσιάζουμε τις διάφορες τιμές για την διακύμανση.

Πίνακας 1.6

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *i* | *xi* | *fi* | *(xi-x) fi* | *(xi-x)2 fi* |
| 1 | 0 | 7 | -36,31 | 188,37 |
| 2 | 1 | 5 | -20,94 | 87,68 |
| 3 | 2 | 6 | -19,13 | 60,96 |
| 4 | 3 | 6 | -13,13 | 28,71 |
| 5 | 4 | 6 | -7,13 | 8,46 |
| 6 | 5 | 5 | -0,94 | 0,18 |
| 7 | 6 | 2 | 1,63 | 1,32 |
| 8 | 7 | 1 | 1,81 | 3,29 |
| 9 | 8 | 1 | 2,81 | 7,91 |
| 10 | 9 | 1 | 3,81 | 14,54 |
| 11 | 10 | 1 | 4,81 | 23,16 |
| 12 | 12 | 2 | 13,63 | 92,82 |
| 13 | 16 | 1 | 10,81 | 116,91 |
| 14 | 17 | 1 | 11,81 | 139,54 |
| 15 | 19 | 1 | 13,81 | 190,79 |
| 16 | 20 | 1 | 14,81 | 219,41 |
| 17 | 23 | 1 | 17,81 | 317,29 |
|  | ***ΣΥΝΟΛΟ*** | 48 | 0 | **1501,31** |

Άρα η διακύμανση είναι S2== = 31,94

Και η τυπική απόκλιση είναι s===5,65

* Ο συντελεστής ασυμμετρίας β1=

**Πίνακας 1.7**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *i* | *xi* | *fi* | *(xi-x)fi* | *(xi-x)2fi* | *(xi-x)3fi* |
| 1 | 0 | 7 | -36,31 | 188,37 | -977,18 |
| 2 | 1 | 5 | -20,94 | 87,68 | -367,14 |
| 3 | 2 | 6 | -19,13 | 60,96 | -194,31 |
| 4 | 3 | 6 | -13,13 | 28,71 | -62,81 |
| 5 | 4 | 6 | -7,13 | 8,46 | -10,05 |
| 6 | 5 | 5 | -0,94 | 0,18 | -0,03 |
| 7 | 6 | 2 | 1,63 | 1,32 | 1,07 |
| 8 | 7 | 1 | 1,81 | 3,29 | 5,95 |
| 9 | 8 | 1 | 2,81 | 7,91 | 22,25 |
| 10 | 9 | 1 | 3,81 | 14,54 | 55,42 |
| 11 | 10 | 1 | 4,81 | 23,16 | 111,46 |
| 12 | 12 | 2 | 13,63 | 92,82 | 632,34 |
| 13 | 16 | 1 | 10,81 | 116,91 | 1264,09 |
| 14 | 17 | 1 | 11,81 | 139,54 | 1648,26 |
| 15 | 19 | 1 | 13,81 | 190,79 | 2635,22 |
| 16 | 20 | 1 | 14,81 | 219,41 | 3250,01 |
| 17 | 23 | 1 | 17,81 | 317,29 | 5651,64 |
|  | ***ΣΥΝΟΛΟ*** | 48 | **0** | 1501,31 | **+13666,2** |

β1= =1,58 >0 .

Ο συντελεστής ασυμμετρίας β1 είναι θετικός οπότε βγάζουμε το συμπέρασμα ότι η κατανομή έχει συγκεντρωμένες τις περισσότερες τιμές “*μαζεμένες*” κοντά στον άξονα των Y. Πράγματι στο παρακάτω διάγραμμα παρουσιάζονται οι διακυμάνσεις των τιμών του δείγματος.

**Διάγραμμα 1.2**

***Ομαδοποίηση Δεδομένων***

Το εύρος είναι R=Xmax-Xmin=23-0=23

Πλάτος διαστήματος τάξεων: δ = = = =3,49 ο μικρότερος ακέραιος από αυτόν τον αριθμό είναι ο αριθμός 3. Άρα χωρίζουμε τις κλάσεις σε 3 ίσα πλάτη δημιουργώντας 8 κλάσεις όπως στον παρακάτω πίνακα

***Πίνακας 2.1***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ΑΡΙΘΜΟΣ ΗΜΕΡΩΝ | ΑΤΟΜΑ  *ΑΠΟΛΥΤΕΣ ΣΥΧΝΟΤΗΤΕΣ*  *fi* | *ΑΠΟΛΥΤΕΣ ΑΘΡΟΙΣΤΙΚΕΣ ΣΥΧΝΟΤΗΤΕΣ*  *Fi* | *ΣΥΧΝΟΤΗΤΕΣ* | *ΠΟΣΟΣΤΑ ΣΥΧΝΟΤΗΤΩΝ* | *ΑΘΡOΙΣΤΙΚΕΣ ΣΥΧΝΟΤΗΤΕΣ* | *ΠΟΣΟΣΤΑ ΑΘΡΟΙΣΤΙΚΩΝ ΣΥΧΝΟΤΗΤΩΝ* |
| 0 ως 2 | 18 | 18 | 0,38 | 38 | 0,38 | 38 |
| 3 ως 5 | 17 | 35 | 0,35 | 35 | 0,73 | 73 |
| 6 ως 8 | 4 | 39 | 0,08 | 8 | 0,81 | 81 |
| 9 ως 11 | 2 | 41 | 0,04 | 4 | 0,85 | 85 |
| 12 ως 14 | 2 | 43 | 0,04 | 4 | 0,90 | 90 |
| 15 ως 17 | 2 | 45 | 0,04 | 4 | 0,94 | 94 |
| 18 ως 20 | 2 | 47 | 0,04 | 4 | 0,98 | 98 |
| 21 ως 23 | 1 | 48 | 0,02 | 2 | 1 | 100 |
| ΣΥΝΟΛΟ | **48** | - | 1 | 100 | - | - |

Σε αυτό το διάγραμμα έχουμε ομαδοποιήσει τα δεδομένα σε 8 ίσες κλάσεις.

***Διάγραμμα 2.1***

Θα χρειαστούμε να κατασκευάσουμε έναν νέο πίνακα προσθέτοντας διάφορα δεδομένα που θα μας βοηθήσουν να βρούμε μερικά περιγραφικά στατιστικά μέτρα.

1. *Αριθμητικός Μέσο*ς

***Πίνακας 2.2***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| i | ΑΡΙΘΜΟΣ ΗΜΕΡΩΝ | ΑΠΟΛΥΤΕΣ ΣΥΧΝΟΤΗΤΕΣ | *ΚΕΝΤΡΙΚΗ ΤΙΜΗ Wi* | *fiWi* |
| 1 | 0 ως 2 | 18 | 1 | 18 |
| 2 | 3 ως 5 | 17 | 4 | 68 |
| 3 | 6 ως 8 | 4 | 7 | 28 |
| 4 | 9 ως 11 | 2 | 10 | 20 |
| 5 | 12 ως 14 | 2 | 13 | 26 |
| 6 | 15 ως 17 | 2 | 16 | 32 |
| 7 | 18 ως 20 | 2 | 19 | 38 |
| 8 | 21 ως 23 | 1 | 22 | 22 |
|  | **ΣΥΝΟΛΟ** | **48** | - | **252** |

* ===5,25

1. *Διάμεσος*

*m = Li + .*

Για i=2 και n=48 έχουμε, m = L2 + ) = 3+ )=3,7

1. *Επικρατούσα Τιμή*

*τ = Li+δi*    
για i=1 (διότι στο 1ο διάστημα [0-2] βρίσκεται η επικρατούσα τιμή) έχουμε

τ = L1+δ1 = 0+2 = 1.

1. *Τεταρτημόρια*

Εντοπίζουμε και πάλι το διάστημα στο οποίο ανήκει κάθε τεταρτημόριο. Στο 1ο τεταρτημόριο η θέση του είναι .

Q1= Li + ) για i=2 έχουμε

Q1= L2 + )= 3 + )=2,29.

Ο τύπος για το Q3 θα είναι

Q3= Li + ) για i=3 έχουμε

Q3= L3 + )= 6 + )=6,5.

Άρα το ενδοτερταμοριακό εύρος θα είναι, IQR= Q3- Q1=6,5-2,29=4,21.

1. *Διακύμανση / Τυπική Απόκλιση*

S2=

***Πίνακας 2.3***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| i | ΑΡΙΘΜΟΣ ΗΜΕΡΩΝ | ΑΠΟΛΥΤΕΣ ΣΥΧΝΟΤΗΤΕΣ fi | *ΚΕΝΤΡΙΚΗ ΤΙΜΗ Wi* | *(wi-\bar{x} \!\, )2* | *fi(wi-\bar{x} \!\, )2* |
| 1 | 0 ως 2 | 18 | 1 | 27,56 | 496,13 |
| 2 | 3 ως 5 | 17 | 4 | 441 | 7497 |
| 3 | 6 ως 8 | 4 | 7 | 1350,56 | 5402,25 |
| 4 | 9 ως 11 | 2 | 10 | 2756,25 | 5513 |
| 5 | 12 ως 14 | 2 | 13 | 4658,06 | 9316,13 |
| 6 | 15 ως 17 | 2 | 16 | 7056 | 14112 |
| 7 | 18 ως 20 | 2 | 19 | 9950,06 | 19900,13 |
| 8 | 21 ως 23 | 1 | 22 | 13340,25 | 13340,25 |
|  | **ΣΥΝΟΛΟ** | **48** | - | 39579,8 | **75576** |

S2= = = 1608 και η τυπική απόκλιση είναι =40.

1. *Συντελεστής Ασυμμετρίας* ***β1***

β1=\bar{x} \!\,3

***Πίνακας 2.4***

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| i | ΑΡΙΘΜΟΣ ΗΜΕΡΩΝ | ΑΠΟΛΥΤΕΣ ΣΥΧΝΟΤΗΤΕΣ fi | *ΚΕΝΤΡΙΚΗ ΤΙΜΗ Wi* | *(wi-\bar{x} \!\, )2* | *fi(wi-\bar{x} \!\, )2* | *(wi-\bar{x} \!\, )3* | *fi(wi-\bar{x} \!\, )3* |
| 1 | 0 ως 2 | 18 | 1 | 27,56 | 496,13 | -76,77 | -1381,78 |
| 2 | 3 ως 5 | 17 | 4 | 441 | 7497 | -1,95 | -33,2 |
| 3 | 6 ως 8 | 4 | 7 | 1350,56 | 5402,25 | 5,36 | 21,44 |
| 4 | 9 ως 11 | 2 | 10 | 2756,25 | 5513 | 107,17 | 214,34 |
| 5 | 12 ως 14 | 2 | 13 | 4658,06 | 9316,13 | 465,48 | 930,97 |
| 6 | 15 ως 17 | 2 | 16 | 7056 | 14112 | 1242,3 | 2484,59 |
| 7 | 18 ως 20 | 2 | 19 | 9950,06 | 19900,13 | 2599,61 | 5199,22 |
| 8 | 21 ως 23 | 1 | 22 | 13340,25 | 13340,25 | 4699,42 | 4699,42 |
|  | **ΣΥΝΟΛΟ** | **48** | - | 39579,8 | 75576 | 9040,63 | **+12135** |

β1== **+**3,95 >0.

Θετική ασυμμετρία όπως παρουσιάζεται στο παρακάτω διάγραμμα.

***Διάγραμμα 2.2***

***Σύγκριση Αποτελεσμάτων***

***Πίνακας 2.5***

|  |  |  |
| --- | --- | --- |
|  | *Μη Ομαδοποιημένα* | *Ομαδοποιημένα* |
| Xμέσο | 5,18 | 5,25 |
| m | 3,5 | 3,7 |
| τ | 0 | 1 |
| Q1 | 11 | 2,2 |
| Q3 | 11 | 6,5 |
| IQR | 0 | 4,21 |
| S | 5,65 | 40 |
| β1 | 1,58 | 3,95 |