Ασκήσεις Περιγραφικής Στατιστικής

1) α) Σε έναν διαγωνισμό είχαμε τις ακόλουθες επιδόσεις:

8, 8, 9, 10, 10, 11, 11, 12, 12, 12, 13, 13, 13, 14, 18

Ποιά η διάμεσος;

Λύση

Τα δεδομένα είναι ήδη τοποθετημένα σε αύξουσα διάταξη. Άρα, προσδιορίζουμε τη θέση της διαμέσου από τον τύπο

$\frac{n+1}{2} = \frac{15+1}{2} = 8$ όπου $n=15$ αριθμός παρατηρήσεων.

Άρα $M=12$ η τιμή της μεταβλητής που βρίσκεται στην 8η θέση.

β). Δίνονται τα ακόλουθα 3.000 εισοδήματα

Εισοδήματα	Στήλη Δεδομένων: Απόλυτες Συχνότητες	Στήλη Υπολογισμών: Αθροιστικές Συχνότητες
[0, 500)	400	400
[500, 1000)	800	1200
[1000, 2000)	600	1800
[2000, 4000)	1200	3.000

$> \frac{n}{2} = 1500$

Ποιά η διάμεσος;

Λύση:

Τα δεδομένα εδώ είναι ομαδοποιημένα. Ο παραπάνω πίνακας θα πρέπει να συμπληρωθεί με την στήλη των Αθροιστικών Συχνοτήτων. Από τη στήλη των Αθροιστικών Συχνοτήτων προσδιορίζω το $\frac{n}{2} = \frac{3000}{2} = 1.500$ που βρίσκεται ανάμεσα στο 1200 και 1.800

F_{i-1} F_i

$F_{i-1} = 1200$ και $F_i = 1800$

$$M = a_{i-1} + \frac{\delta}{f_i}\left[\frac{N}{2} - F_{i-1}\right] =$$

$$= 1000 + \frac{1000}{600}(1500 - 1200) = 1500$$

Άρα το μισό εισοδήματα είναι κάτω από 1500 και το υπόλοιπο μισό πάνω από 1500

2). Σε ένα εργοστάσιο Ε, που παράγει κονσέρβες μετρήθηκε σε mg η περιεκτικότητα ενός τ.δ. 50 κονσερβών σε κάποια ουσία. Οι μετρήσεις είναι οι ακόλουθες:

26	19	19	28	26	21	21	17	31	27
15	26	23	19	21	28	25	22	17	23
27	16	24	21	25	23	32	26	26	28
22	31	22	31	19	21	24	21	28	29
28	23	19	19	18	33	25	30	17	19

α. Να υπολογισθούν ο αριθμητικός μέσος, η διάμεσος, η επικρατούσα τιμή, το εύρος, το ενδοτεταρτημοριακό εύρος, η τυπική απόκλιση καθώς και ο συντελεστής ασυμμετρίας κατά Pearson. Να δοθούν οι κατάλληλες ερμηνείες.

β. Να κατασκευαστεί το boxplot (θηκόγραμμα) των περιεκτικοτήτων και να ερμηνευτεί.

γ. Να γίνει ομαδοποίηση των δεδομένων σε διαστήματα εύρους 3mg με κεντρική τιμή του πρώτου τη μικρότερη παρατήρηση

δ. Θεωρώντας τα δεδομένα ομαδοποιημένα να υπολογισθούν ο αριθμητικός μέσος, η επικρατούσα τιμή, το πρώτο εργαστήριο και η τυπική απόκλιση των περιεκτικοτήτων.

ε). Από ένα άλλο εργοστάσιο E_2 παίρνουμε ένα δείγματο τ.δ. (τυχαίο δείγμα) από το οποίο υπολογίζεται ότι $\bar{x}_2 = 38{,}4\,mg$ και $s_2 = 5{,}1\,mg$. Να εξετάσετε σε ποιο από τα 2 εργοστάσια παρατηρείται μεγαλύτερη διασπορά στην περιεκτικότητα

στ). Ο υπεύθυνος παραγωγής για να εναρμονιστεί με τις οδηγίες της ΕΕ είναι υποχρεωμένος να μειώσει την περιεκτικότητα εφαρμόζοντας τον παρακάτω μτχ

$$Y = 0.4X - 1$$

Να υπολογιστεί ο αριθμητικός μέσος και η τυπική απόκλιση.

Λύση

α) $\bar{x}_1 = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1181}{50} = 23{,}62\ mg$

Διάμεσος M

n άρτιος. Προσδιορίζω τη θέση αφού πρώτα τακτοποιήσω τις παρατηρήσεις στην φυσική τους διάταξη. $\frac{n+1}{2} = \frac{50+1}{2} = 25{,}5$ άρα η διάμεσος βρίσκεται μεταξύ της 25ης και 26ης παρατήρησης οπότε $M = \frac{x_{25} + x_{26}}{2} = \frac{23+23}{2} = 23\ mg$

άρα οι μισές συσκευασίες έχουν περιεκτικότητα κάτω από 23 mg και οι υπόλοιπες μισές πάνω από 23 mg.

Επικρατούσα Τιμή ή Κορυφή M_o

είναι η τιμή που εμφανίζεται τις περισσότερες φορές. $M_o = 19$

Εύρος R

$R = Max - Min = 33 - 15 = 18$

Ενδοτεταρτημοριακό Εύρος $Q_3 - Q_1$.

άρα πρέπει να υπολογίσουμε το Q_3 και το Q_1.

Αρχικά, προσδιορίζουμε τη θέση των Q_3, και Q_1.

Είναι $\frac{3(n+1)}{4} = \frac{3(51)}{4} = 3 \cdot 12,75 = 38,25$ και

$\frac{n+1}{4} = 12,75$ αντίστοιχα.

Άρα, έχουμε για το $Q_3 = \frac{27+28}{2} = 27,5$

$Q_1 = \frac{19+19}{2} = 19$

Άρα $Q_3 - Q_1 = 8,5$

Τυπική Απόκλιση

$$s^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2 = \frac{1}{n-1}\left[\sum x_i^2 - n\bar{x}^2\right] = \frac{1}{49}\left[\sum x_i^2 - 50 \cdot 23,6^2\right]$$

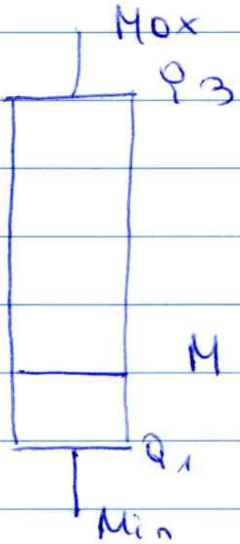
$= 21,22$

$s = \sqrt{21,22} = 4,6$

Συντελεστής Ασυμμετρίας κατά Pearson

$$\gamma_1 = \frac{\bar{x} - M_o}{s} = 1.004 > 0 \Rightarrow \text{Θ.Α. (Θετική Ασυμμετρία)}$$

β). Το θηκόγραμμα αποτελείται από τα ακόλουθα μέτρα $Min = 15$, $Max = 33$, $M = 23$, $Q_1 = 19$, $Q_3 = 27{,}5$



Από το θηκόγραμμα καταλαβαίνουμε ότι έχουμε Θ.Α και μικρή μεταβλητότητα

γ). Η μικρότερη παρατήρηση είναι το 15 και η τιμή αυτή θα πρέπει να είναι η κεντρική τιμή του $1^{ου}$ διαστήματος ~~άρα~~ αφού 5 άρα το $1^{ο}$ διάστημα είναι [12,5   17,5)
(17,5   22,5] και συνεχίζουμε μέχρι να καλύψουμε όλες τις παρατηρήσεις, οπότε συνολικά φτιάχνουμε 5 διαστήματα. Οπότε έχουμε:

Τάξεις	x_i	f_i	Σχ. Συχν	F_i	ΣχΑθρ.	$x_i f_i$	$x_i^2 f_i$
[12,5-17,5)	15	5	0,1	5	0,1	15·5	$15^2 \cdot 5$
[17,5-22,5)	20	17	0,34	22	0,44	20·17	$20^2 \cdot 17$
[22,5-27,5)	25	16	0,32	38	0,76	25·16	$25^2 \cdot 16$
[27,5-32,5)	30	11	0,22	49	0,98	30·11	$30^2 \cdot 11$
[32,5-37,5)	35	1	0,02	50	1	35·1	$35^2 \cdot 1$
						1180	

$$\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{1180}{50} = 23{,}6$$ Αριθμητικός Μέσος

Κορυφή

Η μεγαλύτερη συχνότητα είναι το 17 που βρίσκεται στη 2ο τάξη → στο [17,5 22,5)

$$M_0 = a_{i-1} + \delta \frac{\Delta_1}{\Delta_1+\Delta_2} = 17,5 + 5\,\frac{17-5}{(17-5)+(17-6)} = 22,12$$

Υπολογισμός Q_1

Προσδιορισμός της θέσης $\frac{n}{4} = 12,5$

Βρίσκεται από τη στήλη F_i. Το $\frac{n}{4}$ είναι μεταξύ της $F_{i-1}=5$ και $F_i = 22$

Η $F_i = 22$ → στο 2ο διάστημα οπότε έχουμε.

$$Q_1 = a_{i-1} + \frac{\delta}{f_i}\left[\frac{n}{4} - F_{i-1}\right] = 17,5 + \frac{5}{17}(12,5-5) = 19,71$$

~~Υπολογισμός Q_3~~

~~Προσδιορισμός της θέσης $\frac{3n}{4} =$~~

Ερμηνεία

Το 25% των ανθρώπων έχουν περιεκτικότητα κάτω από 19,71mg, και το 75% έχουν περιεκτικότητα πάνω από 19,71mg.

Τυπική Απόκλιση

$$s = \sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2 f_i} = \sqrt{\frac{1}{n-1}\left[\sum x_i^2 f_i - n\bar{x}^2\right]}$$

Συμπληρώνουμε τον πίνακα συχνοτήτων με τη στήλη $x_i^2 f_i$ και κάνοντας τους υπολογισμούς βρίσκουμε

$s = 4,95$ mg

ε). Για τη σύγκριση των διασπορών υπολογίζουμε τον CV για το κάθε εργοστάσιο.

$$CV_1 = \frac{s_1}{\bar{x}_1} = \frac{4,95}{23,6} \cdot 100 = 20,97\%$$

$$CV_2 = \frac{s_2}{\bar{x}_2} \cdot 100 = \frac{5,1}{38,4} \cdot 100 = 13,28\%$$

$CV_2 < CV_1$ άρα το δεύτερο εργοστάσιο

έχει μικρότερη διασπορά

στ). $Y = 0,4X - 1$

$$\bar{x}_y = 0,4\bar{x}_x - 1 = 8,448$$

$$s_y = 0,4 \cdot s_x = 1,84$$

Σύμφωνα με τις ιδιότητες για μέσου και της διασποράς (γραμμικής συνάρτησης)

2) Αν σε όλες τις τιμές μιας μεταβλητής προσθέσουμε τον αριθμό −2, τότε η μέση τιμή

- ☐ αυξάνεται κατά 2
- ☑ μειώνεται κατά 2
- ☐ παραμένει ίδιο

3) Αν σε όλες τις τιμές μιας μεταβλητής πολ/με επί 2, τότε η μέση τιμή

- ☐ αυξάνεται κατά 2
- ☑ διπλασιάζεται
- ☐ παραμένει ίδιο

4) Η μέση τιμή των αριθμών 3, 6, 7, α, 14 είναι 8. Να υπολογιστεί η διασπορά τους.

Λύση

$$\bar{x} = \frac{\Sigma x_i}{n} \Rightarrow 8 = \frac{3+6+7+a+14}{5} \Rightarrow 30+a=40 \Rightarrow$$

$$\Rightarrow a = 10$$

$$s^2 = \frac{1}{n-1}\Sigma(x_i-\bar{x})^2 = \frac{1}{n-1}\left[\Sigma x_i^2 - n\bar{x}^2\right] =$$

$$= \frac{1}{4}\left[(9+36+49+100+196) - 5 \cdot 64\right] =$$

$$= \frac{1}{4}\left[390-320\right] = \frac{70}{4} = 17,5$$

5). Αν $\bar{x} = 2$ και $CV = 50\%$ τότε να υπολογισθεί η τυπική απόκλιση

Λύση

$$CV = \frac{s}{\bar{x}} \Rightarrow s = CV \cdot \bar{x} \Rightarrow s = 0,5 \cdot 2 = 1$$

6). Οι μηνιαίες αποδοχές 100 υπαλλήλων μιας επιχείρησης κυμαίνονται μεταξύ 2.000 και 3000 ευρώ. Ακόμη γνωρίζουμε ότι:
10 υπάλληλοι αμείβονται το πολύ με 2200
30 υπάλληλοι αμείβονται το πολύ με 2.400
40 υπάλληλοι έχουν μισθούς μεγαλύτερους από 2600 και 10 υπάλληλοι έχουν μισθούς μεγαλύτερους από 2800 €.
Να υπολογισθούν: η μέση τιμή, η διάμεσος, το Q_1 και το Q_3

Λύση

Από τα δεδομένα της άσκησης σχηματίζουμε τον παρακάτω πίνακα

Στήλες Δεδομένων | Στήλες Υπολογισμών

Μηνιαίες Αποδοχές	Αριθμός Υπαλλήλων	x_i	$x_i f_i$	F_i	
[2000 2200)	10	2100	21000	10	
					$\frac{n}{4} = 25$
[2200 2400)	20	2300	46000	30	
					$\frac{n}{2} = 50$
[2.400 2600)	30	2500	75000	60	
					$\frac{3n}{4} = 75$
[2600 2800)	30	2700	81000	90	
[2800 3000)	10	2900	29000	100	
Σύνολο	100		252000		

$$\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{252000}{100} = 2520 \text{ ευρώ}$$

Υπολογισμός Μ

Από τη στήλη Fi υπολογίζω τη θέση της διαμέσου. Το $\frac{n}{2}=50$ βρίσκεται μεταξύ 30-60

άρα έχουμε

$$M = a_{i-1} + \frac{\delta}{f_i}\left[\frac{n}{2} - F_{i-1}\right] = 2400 + \frac{200}{30}(50-30) =$$

$$= 2.533,33$$

Οι μισοί υπάλληλοι έχουν αποδοχές κάτω από 2533,33 ενώ οι υπόλοιποι μισοί πάνω από 2533,33 €

Υπολογισμός Q_1

Η θέση του Q_1 βρίσκεται μεταξύ 10 και 30 της Fi, οπότε έχουμε:

$$Q_1 = a_{i-1} + \frac{\delta}{f_i}\left[\frac{n}{4} - F_{i-1}\right] = 2200 + \frac{200}{20}\left[25-10\right]$$

$$= 2350 \text{ €}$$

Το 25% των υπαλλήλων έχουν μηνιαίες αποδοχές κάτω από 2350 € και το υπόλοιπο 75% πάνω από 2350 €

Υπολογισμός Q_3

Η θέση του Q_3 βρίσκεται μεταξύ 60 και 90 της Fi, οπότε έχουμε

$$Q_3 = a_{i-1} + \frac{\delta}{f_i}\left[\frac{3n}{4} - F_{i-1}\right] = 2600 + \frac{200}{30}[75-60] = 2700$$

Το 75% των υπαλλήλων έχουν μηνιαίες αποδοχές κάτω από 2700 € και το υπόλοιπο 25% πάνω από 2700 €

7). Η επίδοση 50 φοιτητών στο μάθημα της Στατιστικής ήταν: 4, 2, 9, 5, 5, 6, 7, 7, 8, 8, 2, 3, 4, 6, 6, 4, 5, 3, 2, 1, 6, 8, 7, 9, 0, 9, 4, 1, 0, 3, 1, 8, 6, 7, 7, 6, 8, 3, 5, 2, 9, 9, 3, 7, 4, 1, 5, 8, 7, 6

i). Να παραστήσετε τα δεδομένα σ' έναν πίνακα συχνοτήτων

ii). Αν η σχολή πρόκειται να δώσει υποτροφία σε όσους πήραν βαθμό μεγαλύτερο ή ίσο του 8, πόσοι φοιτητές θα πάρουν υποτροφία;

iii). Αν η σχολή δώσει υποτροφία στο 36% των καλύτερων φοιτητών, τι επίδοση πρέπει να έχει κάποιος για να πάρει υποτροφία;

Λύση:

Για να βρούμε τις διαφορετικές τιμές της επίδοσης των φοιτητών για να κατασκευάσουμε ο πίνακας συχνοτήτων θα πρέπει τα δεδομένα να τα τοποθετήσουμε στην φυσική τους διάταξη

Ο πίνακας συχνοτήτων είναι ο ακόλουθος:

x_i	f_i	$f_i\%$	F_i	$F_i\%$
0	2	4	2	4
1	4	8	6	12
2	4	8	10	20
3	5	10	15	30
4	5	10	20	40
5	5	10	25	50
6	7	14	32	64
7	7	14	39	78
8	6	12	45	90
9	5	10	50	100
	50	100		

ii) Θέλουμε επίδοση μεγαλύτερη ή ίση του 8. $X \geq 8$. Άρα, η σχολή θα πάρει $6+5=11$ άτομα, δηλ. το 22% των υποψηφίων

iii). Αφού η σχολή θα πάρει το 36% των υποψηφίων, αυτό σημαίνει ότι το 64% των υποψηφίων δεν θα επιτύχουν, δηλ. όσοι έχουν επίδοση μικρότερη ή ίση του 6. Άρα θα επιτύχουν όσοι έχουν επίδοση μεγαλύτερη ή ίση του 7.