Θεματική Ενότητα: TSM33 Ποσοτικές Μέθοδοι
Ακαδ. Έτος: 2023-2024

Γραπτή Εργασία 1 (ΓΕ1)-Ενδεικτικές Απαντήσεις

Ερώτηση 1 (4 μονάδες)

Εκφώνηση

Στον πίνακα που ακολουθεί παρουσιάζεται ο **αριθμός των followers** (ακολούθους) που αποκτούν ανά ημέρα οι **50** δημοφιλέστεροι τραγουδιστές της ελληνικής μουσικής σκηνής.

Πίνακας 1. Αριθμός followers (ακολούθων) 50 δημοφιλέστερων τραγουδιστών ελληνικής μουσικής σκηνής

21	43	50	25	55	30	28	40	31	51
18	47	52	34	47	32	27	41	35	54
30	48	36	43	38	33	27	39	41	43
22	22	46	52	29	32	34	34	42	36
35	28	57	56	20	38	27	27	40	35

i) Να κατασκευάσετε τον **πίνακα συχνοτήτων, σχετικών συχνοτήτων και αθροιστικών συχνοτήτων**

ii) Να πραγματοποιηθεί η **γραφική απεικόνιση** της **συχνότητα**ς και της **σχετικής συχνότητας** των δεδομένων

iii) Να υπολογιστούν και να σχολιαστούν **τα μέτρα κεντρικής τάσης:** αριθμητικός μέσος, διάμεσος, επικρατούσα τιμή, πρώτο και τρίτο τεταρτημόριο

iv) Να υπολογιστούν και να σχολιαστούν **τα μέτρα διασποράς**: εύρος, διακύμανση, τυπική απόκλιση και ο συντελεστής μεταβλητότητας

v) Να υπολογιστούν και να σχολιαστούν: ο **συντελεστής ασυμμετρίας** και ο **συντελεστής κύρτωσης** του Pearson

Απάντηση

i) Δεδομένα:

Πλήθος παρατηρήσεων, n=**50**

Εύρος τιμών **R=max-min**=57-18= 39

Αριθμός κλάσεων **K**=6, (με βάση τον εμπειρικό κανόνα, βλέπε διαφάνεια 11, 1ης εκπαιδευτικής συνάντησης)

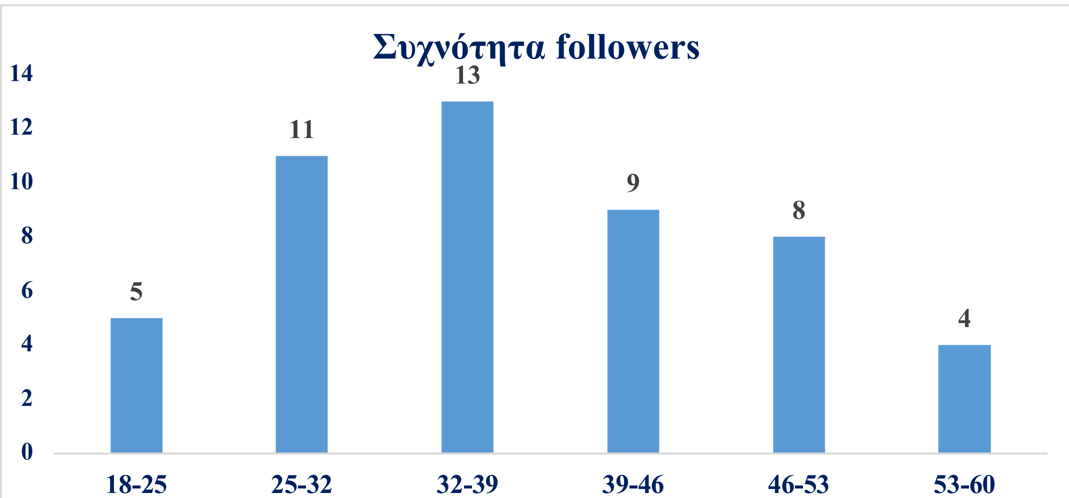
Εύρος κλάσεων **C=R/K**=39/6=6,5 ≈7

Συνεπώς με βάση τα παραπάνω οργανώνω την παρουσίαση των δεδομένων μου ως εξής:

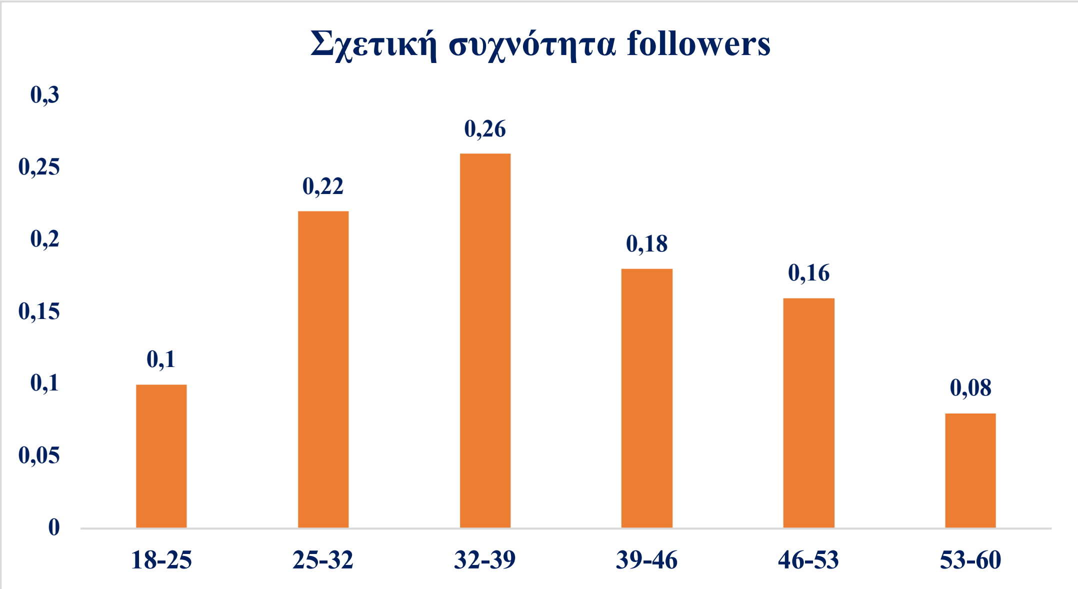
Πίνακας 2. Πίνακας συχνοτήτων, σχετικών και αθροιστικών συχνοτήτων

Κλάσεις/Τάξεις	Συχνότητα (n_i)	Σχετική συχνότητα $f_i=(n_i/n)$	Αθροιστική συχνότητα (Ni)
18-25	5	0,10	5
25-32	11	0,22	16
32-39	13	0,26	29
39-46	9	0,18	38
46-53	8	0,16	46
53-60	4	0,08	50
Σύνολο	**50**	**1,00**	

ii



Σχήμα 1. Συχνότητα followers (αριθμός)



Σχήμα 2. Σχετική συχνότητα followers (αριθμός)

Αριθμητικός μέσος: $\bar{x} = \frac{\sum x_i}{n} = \frac{18+20+21+22.......+55+51+54}{50} = \frac{1851}{50} = 37{,}02$

Διάμεσος: Αρχικά τοποθετώ τα δεδομένα σε αύξουσα σειρά

18	20	21	22	22	25	27	27	27	27
28	28	29	30	30	31	32	32	33	34
34	34	35	35	35	36	36	38	38	39
40	40	41	41	42	43	43	43	46	47
47	48	50	51	52	52	54	55	56	57

Δεδομένου ότι ο αριθμός των παρατηρήσεων μου είναι άρτιος (n=50) αυτό σημαίνει ότι η **διάμεσος** είναι η μέση τιμή των δύο (2) τιμών που βρίσκονται στην $25^{η}$ και $26^{η}$ θέση. Συνεπώς $M = \frac{X_{25}+X_{26}}{2}$ =35+36/2 = 35,5

Η **επικρατούσα τιμή, T_0** είναι η τιμή παρουσιάζει την μεγαλύτερη συχνότητα, στην περίπτωση μας είναι το 27 (4 φορές) και μάλιστα είναι μονοκόρυφη, (έχουμε μόνο μία).

Πρώτο τεταρτημόριο

(n+1)/4=12,75 άρα το $1^{ο}$ τεταρτημόριο είναι η τιμή που βρίσκεται στην θέση AQ = 12 και ΔQ = 0,75.

Άρα έχουμε: Q1 = $X_{12} + \Delta Q\ (X_{13} - X_{12})$ → Q1 =28+0,75* (29-28)= 28,75

Τρίτο τεταρτημόριο
(n+1)*3/4=38,25, άρα το 3^{0} τεταρτημόριο είναι η τιμή που βρίσκεται στην θέση AQ = 38 και ΔQ = 0,25.

Άρα έχουμε: Q3 = $X_{38} + \Delta Q\ (X_{39} - X_{38})$ → Q1 = 43+0,25 (46-43)=43+0,75=43,75

iii) Εύρος τιμών **R=max-min**=57-18= 39

Ομαδοποιημένα δεδομένα

Κλάσεις/Τάξεις	**Συχνότητα (f_i)**	**Κέντρο τάξης c_i**	**$f_i c_i$**	**c_i^2**	**$f_i c_i^2$**
18-25	5	21,5	107,5	462,25	2.311,25
25-32	11	28,5	313,5	812,25	8.934,75
32-39	13	35,5	461,5	1.260,25	16.383,25
39-46	9	42,5	382,5	1.806,25	16.256,25
46-53	8	49,5	396	2.450,25	19.602
53-60	4	56,5	226	3.192,25	12.769
Σύνολο	**50**		**1.887**	**9.983,5**	**76.256,5**

$\overline{X} = \frac{\sum_{i=1}^{m} f_i c_i}{n}$ = 1887/50= 37,74

$S^2 = \frac{\sum_{i=1}^{m} f_i c_i^2}{n} - \overline{X^2}$ = (76.256,50/50) – (37,74)^2 = (1.525,33) – (1.424,30)= 101,03

Τυπική απόκλιση (τετραγωνική ρίζα της διακύμανσης), επομένως:
$S = \sqrt{101{,}03} \rightarrow S = 10{,}051.$

Συντελεστής μεταβλητότητας $CV = \frac{S}{\bar{X}} \rightarrow CV = \frac{10{,}051}{37{,}74} = 0{,}2663$

Συντελεστής ασυμμετρίας Pearson

$Sp = \frac{\bar{X} - T_o}{S} \rightarrow Sp = \frac{37{,}02 - 27}{10{,}051}$ $Sp = 0{,}996$, και αφού $Sp > 0$, τότε έχουμε θετική ασυμμετρία.

Συντελεστής κύρτωσης:

$$\beta_4 = \frac{\frac{1}{n}\sum(x_i - \bar{x})^4}{S^4} = 0{,}7695$$

Δεδομένου ότι $\beta_4<3$, συμπεραίνουμε ότι έχουμε πλατύκυρτη κατανομή (χαμηλή κορυφή).

Βλέπε σημειώσεις 1ης εκπαιδευτικής συνάντησης

Ερώτηση 2 (2 μονάδες)

Α. Έστω ότι ρίχνουμε ένα ζάρι **5** φορές και ενδιαφερόμαστε αν το αποτέλεσμα κάθε ρίψης ήταν «**1**» ή «**όχι 1**». Να απαντηθούν τα ακόλουθα ερωτήματα.

i. Ποια η πιθανότητα να μην έρθει **ούτε μια** φορά στις προσπάθειες το «**1**» ;
ii. Ποια η πιθανότητα να έρθει ακριβώς **τρεις** φορές στις **5** προσπάθειες το «**1**» ;
iii. Ποια η πιθανότητα να έρθει **τουλάχιστον δύο** φορές στις **5** προσπάθειες το «**1**» ;
iv. Ποιος είναι ο αναμενόμενος αριθμός επιτυχιών και η διακύμανση του.

Απάντηση

Έχουμε περίπτωση διωνυμική κατανομής

$$X \sim B\ (n, p)$$

Όπου η συνάρτηση πιθανότητας ορίζεται από τον ακόλουθο τύπο

$$P(X=k)=\binom{n}{k}p^k q^{n-k}=\frac{n!}{k!(n-k)!}p^k q^{n-k},\quad 0\le k\le n,\quad 0<p<1$$

Με βάση την εκφώνηση ορίζουμε ως **επιτυχία p** = η ρίψη του ζαριού να έρθει 1 και ως **αποτυχία q**= την ρίψη του ζαριού να μην έρθει 1. Επίσης ορίζουμε X το πλήθος των φορών που θα έρθει το 1, που αποτελεί παράλληλα το πλήθος των επιτυχιών. Επιπλέον έχουμε πλήθος επαναλήψεων 5, (n=5).

i) Δεδομένα k=0, p= (0,167), άρα q=1-p= (0,833) και n=5

Αξιοποιώντας την συνάρτηση πιθανότητας

$$P(X=k)=\binom{n}{k}p^k q^{n-k}=\frac{n!}{k!(n-k)!}p^k q^{n-k},\quad 0\le k\le n,\quad 0<p<1$$

έχω:

$P\ (X=0)=\frac{n!}{k!(n-k)!}\,p^k q^{n-k} \rightarrow P\ (X=0)=\frac{5!}{0!(5-0)!}\ 0{,}167^0 0{,}833^{5-0}$ =0,40107 ή 40,107%

ii) Δεδομένα k=3, p= (0,167), άρα q=1-p= (0,833) και n=5

$P\ (X = K) = \frac{n!}{k!(n-k)!}\, p^k q^{n-k} \rightarrow P\ (X = 3) = \frac{5!}{3!(5-3)!}\ 0{,}167^3 0{,}833^{5-3}$= 0,031 ή 3,1%

iii) Δεδομένα, k≥2, p= (0,167), άρα q=1-p= (0,833) και n=5

Λειτουργούμε λίγο διαφορετικά. Από το 1 που εκφράζει όλες τις πιθανότητες θα αφαιρέσουμε την πιθανότητες το X να λάβει τις τιμές 0 και 1 που είναι σίγουρα κάτω από το 2 που αναζητούμε. Από το προηγούμενο ερώτημα βρήκαμε ότι το $P\ (X = 0) = 0{,}40107$

Για $P\ (X = 1) = \frac{5!}{1!(4)!}\ 0{,}167^1 0{,}833^4$=0,4019

Άρα $P\ (X \geq 2) = 1 - P\ (X = 0) - P\ (X = 1)$= 1-0,40107-0,40109=0,19784 ή 19,78%

iv) Ο αναμενόμενος αριθμός επιτυχιών (ή μέση τιμή) είναι:

$E(X) = np \rightarrow E(X) = 5 * \frac{1}{6} = 0{,}833$

και η διακύμανση:

$V\ (X) = npq = 5 * \frac{1}{6} * \frac{5}{6}$ =*0,6944*

Βλέπε σημειώσεις 2ης εκπαιδευτικής συνάντησης, διαφάνεια: 17

Ερώτηση 2Β

Αν ο αριθμός των κλήσεων προς το κινητό μας ημερησίως ακολουθεί την κατανομή Poisson και ο μέσος αριθμός κλήσεων είναι **λ= 12**. Να υπολογιστεί η πιθανότητα σε μία συγκεκριμένη μέρα: α) να έχουμε **2** κλήσεις και β) να έχουμε **15** κλήσεις.

Απάντηση 2Β

Φαινόμενο ουράς, τυχαία μεταβλητή που ακολουθεί την κατανομή Poisson με παράμετρο λ,

$$X \sim P(\lambda)$$

Αξιοποιώ την συνάρτηση πιθανότητας της κατανομής Poisson $P(X=k) = \frac{e^{-\lambda}\lambda^{\kappa}}{k!}$

λ= 12, οπότε έχω:

α) Να έχουμε 2 κλήσεις

$$p(X=x) = p_x = \frac{e^{-\lambda}\lambda^x}{x!} \Leftrightarrow p(X=1) = \frac{e^{-12}12^2}{2!}$$
$$= 72e^{-12} = 0{,}00044$$

β) Να έχουμε 15 κλήσεις

$$p(X=x) = p_x = \frac{e^{-\lambda}\lambda^x}{x!} \Leftrightarrow p(X=15) = \frac{e^{-12}12^{15}}{15!}$$
$$= 0{,}0724$$

Βλέπε σημειώσεις 2ης εκπαιδευτικής συνάντησης, διαφάνειες: 28-36

Ερώτηση 3 (4 μονάδες)

Α. Έστω πληθυσμός ατόμων με **Δείκτη Μάζα Σώματος** (Δ.Μ.Σ.) που ακολουθεί κανονική κατανομή μέσης τιμής **25** (μ=25) και τυπική απόκλιση **6** (σ=6). Ζητείται το ποσοστό ατόμων με Δ.Μ.Σ. : i) μικρότερο του **28**, ii) μεγαλύτερο του **20** και iii) από **23 έως 30**.

Απάντηση:

Ορίζουμε ως μεταβλητή X τον Δείκτη Μάζα Σώματος (Δ.Μ.Σ.) η οποία ακολουθεί την κανονική κατανομή με $X \sim N(25, 6^2)$

i) μικρότερο του **28**

$$P(X < 28) = \frac{\text{x}-\text{M}}{\sigma} < \frac{28-25}{6} = \text{P}(\text{Z} < 0{,}5) = \Phi(0{,}5) = 0{,}6915 \text{ ή } 69{,}15\%^*$$

* Από τον πίνακα τυποποιημένης κανονικής κατανομής, βρίσκουμε πως $\Phi(0{,}5)$= 0,6915.

Συνεπώς, το 69,15% του πληθυσμού έχει Δείκτη Μάζα Σώματος μικρότερο του 28.

ii) μεγαλύτερο του **20**

(i) $$P(X > 20) = \frac{\text{x}-\mu}{\sigma} > \frac{20-25}{6} = \text{P}\left(\text{Z} > \frac{-5}{6}\right) = P(Z > -0{,}833) = 1 - P(Z < -0{,}833) = 1 - \Phi(-0{,}833) = 1 - [1 - \Phi(-0{,}833)] = 1 - 1 + \Phi(0{,}833) = \Phi(0{,}833) = 0{,}7967 \text{ ή } 79{,}67\%$$

Από τον πίνακα τυποποιημένης κανονικής κατανομής, βρίσκουμε πως $\Phi(0{,}833)$= 0,7967

Σημείωση= Θυμίζω $\Phi(-z) = 1 - \Phi(z)$

Συνεπώς, το 79,67% του πληθυσμού έχει Δείκτη Μάζα Σώματος μεγαλύτερο του 20.

iii) από **23 έως 30**

$$P(23 \leq X \leq 30) = \frac{23-\mu}{\sigma} \leq \frac{\text{X}-\mu}{\sigma} \leq \frac{30-25}{6} = P(-0{,}333 \leq Z \leq 0{,}8333)$$
$$= \Phi(0.8333) - \Phi(-0{,}0333) = \Phi(0{,}8333) - [1 - \Phi(0{,}333)]$$
$$= 0{,}7967 - 1 + 0{,}6293 = 0{,}426 \text{ ή } 42{,}6\%$$

Συνεπώς, το 42,6% του πληθυσμού έχει Δείκτη Μάζα Σώματος από 23 έως 30.

Βλέπε σημειώσεις 2 ης εκπαιδευτικής συνάντησης, διαφάνειες: 50-66

Β. Σε μια έρευνα που έγινε για το μέσο εβδομαδιαίο εισόδημα των σερβιτόρων της πόλης του Βόλου έδειξε ότι σε δείγμα **75** σερβιτόρων ο μέσος μισθός είναι **227** ευρώ με τυπική απόκλιση **15**. Να κατασκευάσετε το i) **90%** και ii) **80%** διάστημα εμπιστοσύνης για το μέσο εβδομαδιαίο εισόδημα.

Απάντηση:

Στην συγκεκριμένη περίπτωση η κατανομή $\bar{X}$ ακολουθεί την **κανονική κατανομή**. Από τα δεδομένα μας έχουμε $\bar{X} = 227$ και S=15.

α) Επίπεδο εμπιστοσύνης 1-α =1-09,=0.1 άρα α/2= 0,05 και συνεπώς το $Z_{\alpha/2}$= 1,645.

Προσδιορίζω το τυπικό σφάλμα αξιοποιώντας τον παρακάτω τύπο:

$$SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Άρα έχω $SE(\bar{X}) = \frac{15}{\sqrt{75}}$ =1.732

Το διάστημα εμπιστοσύνης προκύπτει από τον ακόλουθο τύπο:

$\bar{X}$± (θεωρητική τιμή κατανομής)*(τυπ. σφάλμα)

Άρα στο παράδειγμα μας : 227 ± (1.645)*(1.732)= 2.85

και συνεπώς έχω άνω όριο: $\bar{X} + 2.85$= 227+2.85=229.85 και

κάτω όριο $\bar{X} - 2.85$= 227-2.85=224.15

άρα 224.15 ≤ $\bar{X}$≤229.85

β) Επίπεδο εμπιστοσύνης 1-α =1-08,=0.2 άρα α/2= 0,10 και συνεπώς το $Z_{\alpha/2}$= 1,282.

Προσδιορίζω το τυπικό σφάλμα αξιοποιώντας τον παρακάτω τύπο:

$$SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Άρα έχω $SE(\bar{X}) = \frac{15}{\sqrt{75}}$ =1.732

Το διάστημα εμπιστοσύνης προκύπτει από τον ακόλουθο τύπο:

$\bar{X}\pm$ (θεωρητική τιμή κατανομής)*(τυπ. σφάλμα)

Άρα στο παράδειγμα μας : 227 ± (1.282)*(1.732)= 227±2.22

και συνεπώς έχω άνω όριο: $\bar{X} + 2.85$= 227+2.22=229.22 και

κάτω όριο $\bar{X} - 2.85$= 227-2.22=224.78

άρα 224.78 ≤ $\bar{X}$≤229.22

Βλέπε σημειώσεις 3 ης εκπαιδευτικής συνάντησης, διαφάνειες: 21-34

Γ. Η γνωστή ιστοσελίδα TripAdvisor πραγματοποίησε μια έρευνα για την ποιότητα του φαγητού των εστιατορίων της περιοχής της Κυψέλης. Η άριστη ποιότητα βαθμολογείται με **10** ενώ ποιοτικά θεωρούνται τα εστιατόρια με βαθμολογία ***πάνω από* 7**. Ένα δείγμα **12** φοιτητών επιλέχθηκε να ρωτηθεί για το εστιατόριο **«Παράδεισος»** και έδωσαν τις εξής απαντήσεις: **7,8,10,8,6,9,6,7,7,8,9** και **8.** Ο δειγματικός μέσος είναι **7,75** και η τυπική απόκλιση **1,215**. Εάν υποθέσουμε ότι η κατανομή του πληθυσμού ακολουθεί προσεγγιστικά την κανονική κατανομή, μπορούμε να θεωρήσουμε ότι το εστιατόριο **«Παράδεισος»** παρέχει ποιοτικό φαγητό; **(α=0,05)**

Απάντηση

Δεδομένα: α) Μέση τιμή $\bar{x}$**= 7,75**, β) τυπική απόκλιση **σ=1,215** γ) μέγεθος δείγματος **n:12** το οποίο είναι <u>**κάτω από 30**</u> με άγνωστη διακύμανση που σημαίνει ότι αξιοποιήσω την στατιστική συνάρτηση ελέγχου **t-student** και δ) έχω επίπεδο σημαντικότητας **α=0,05**

Κριτήριο αξιολόγησης μου στο συγκεκριμένο έλεγχο είναι η μέση βαθμολογία του ποιοτικού φαγητού που ισοδυναμεί με μία βαθμολογία πάνω από το 7. Άρα με βάση το συγκεκριμένο κριτήριο ορίζω τον παρακάτω μονόπλευρο (δεξιόπλευρο) έλεγχο:

H_0: $\mu \leq 7$ (δεδομένου ότι η μέση βαθμολογία είναι κάτω από 7 δεν παρέχει το εστιατόριο ποιοτικό φαγητό)

H_1: $\mu > 7$ (δεδομένου ότι η μέση βαθμολογία είναι πάνω από 7 παρέχει το εστιατόριο ποιοτικό φαγητό)

Αρχικά υπολογίζω τις τιμές για το $S_{\bar{X}}$ και t αξιοποιώντας του παρακάτω τύπους:

$$S_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{1,215}{\sqrt{12}} = 0,35$$

$t = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{7,75-7}{0,35} = 2,14$ και συγκρίνω με την κρίση τιμή

$$t_{n-1,1-\frac{a}{2}} = t_{12-1,\ 1-\frac{0,05}{2}} = t_{11,\ 0,975} \cong 1,796$$

Και δεδομένου ότι

$t = 2,14 > \ t_{11,\ 0,975 \cong 1,796}$ * Δες πίνακα θεωρητικών τιμών της κατανομής t-student (n=11, a=0.05)

απορρίπτουμαι την μηδενική υπόθεση και οδηγούμαστε στο συμπέρασμα ότι το εστιατόριο «Παράδεισος» προσφέρει ποιοτικό φαγητό.

Βιβλιογραφικές αναφορές

Παπαδόγγονας, Θ. (2022). *«Ποσοτικές μέθοδοι ανάλυσης επιχειρηματικών αποφάσεων*", 2η Έκδοση, Αθήνα: Εκδόσεις Τσότρας.