

**1<sup>ο</sup> Θέμα [40]** Τα τελευταία χρόνια παρατηρείται συνεχώς αυξανόμενο ενδιαφέρον για τη μελέτη της συγκέντρωσης τοξικών στοιχείων στον οργανισμό των θαλάσσιων θηλαστικών. Στο πλαίσιο μιας σχετικής μελέτης για τη συγκέντρωση, έστω  $X$ , υδραργύρου στο συκώτι δελφινιών στη Μεσόγειο, έγιναν σχετικές μετρήσεις σε ένα τυχαίο δείγμα 28 δελφινιών με τα εξής αποτελέσματα (σε  $\mu\text{gr}/\text{gr}$ ):

1.70	101	168	481	252	278	397
1.72	85.40	218	485	329	286	209
8.80	118	180	221	316	315	314
5.90	183	264	406	445	241	318

Η μέση τιμή, η τυπική απόκλιση, το πρώτο τεταρτημόριο, το δεύτερο τεταρτημόριο και το ενδοτεταρτημοριακό εύρος του δείγματος είναι, αντίστοιχα, 236.73, 141.55, 130.5, 246.5, 187 (σε  $\mu\text{gr}/\text{gr}$ ). **α)** Να κατασκευάσετε το θηκόγραμμα του δείγματος. Υπάρχουν ακραίες τιμές στο δείγμα; **β)** Να υπολογίσετε τη z-τιμή της τιμής  $x = 118 \mu\text{gr}/\text{gr}$  και να εξηγήσετε τι εκφράζει. **γ)** Αν  $P_{90} = 452.5 \mu\text{gr}/\text{gr}$ , τι μπορούμε να πούμε για το δελφίνι στο οποίο βρέθηκε συγκέντρωση υδραργύρου  $x = 481 \mu\text{gr}/\text{gr}$ ; **δ)** Ποιο ποσοστό των παρατηρήσεων του δείγματος βρίσκεται μεταξύ 130.5 και  $317.5 \mu\text{gr}/\text{gr}$ ; **ε)** Βρείτε ένα 95% διάστημα εμπιστοσύνης για τη μέση συγκέντρωση υδραργύρου στο συκώτι δελφινιών στη Μεσόγειο. Πώς αντιλαμβάνεσθε (πώς ερμηνεύετε) αυτό το διάστημα; **στ)** Ποιο είναι το περιθώριο σφάλματος της εκτίμησης που δώσατε στο (ε) για τη μέση συγκέντρωση υδραργύρου στο συκώτι δελφινιών στη Μεσόγειο; Αν από το ίδιο δείγμα κατασκευάσουμε για τη μέση συγκέντρωση υδραργύρου στο συκώτι δελφινιών στη Μεσόγειο ένα άλλο διάστημα εμπιστοσύνης με μεγαλύτερο συντελεστή εμπιστοσύνης, το περιθώριο σφάλματος θα αυξηθεί, θα μειωθεί ή θα παραμείνει ίδιο; **ζ)** Για να απαντήσετε στο ερώτημα (ε), χρειάστηκε να κάνετε κάποια υπόθεση; Αν ναι, να αναφέρετε έναν τρόπο ελέγχου αυτής της υπόθεσης. **η)** Γνωρίζουμε ότι η δειγματική μέση τιμή  $\bar{X}$ , είναι μια συνεπής εκτιμήτρια της μέσης τιμής  $\mu$  ενός πληθυσμού. Πώς αντιλαμβάνεσθε αυτή την ιδιότητα της  $\bar{X}$ ;

**2<sup>ο</sup> Θέμα [40]** Έστω  $X$  τυχαία μεταβλητή που εκφράζει την ποσότητα λυκοπενίου που περιέχεται στις συσκευασίες των 200gr ελαφρά συμπυκνωμένου χυμού τομάτας της εταιρείας τροφίμων Α και  $Y$  τυχαία μεταβλητή που εκφράζει την ποσότητα λυκοπενίου που περιέχεται στις συσκευασίες (των 200gr επίσης) ελαφρά συμπυκνωμένου χυμού τομάτας της εταιρείας τροφίμων Β. Στο πλαίσιο της πτυχιακής σας εργασίας πήρατε, με βάση ένα σχέδιο τυχαίας δειγματοληψίας, 25 συσκευασίες ελαφρά συμπυκνωμένου χυμού τομάτας (των 200gr) της εταιρείας τροφίμων Α και 25 συσκευασίες της εταιρείας τροφίμων Β και μετρήσατε την ποσότητα λυκοπενίου που περιέχεται σε κάθε συσκευασία. Η μέση ποσότητα λυκοπενίου στο δείγμα από την εταιρεία Α βρέθηκε ίση με 24mg με τυπική απόκλιση 2mg και στο δείγμα από την εταιρεία Β ίση με 22.6mg με τυπική απόκλιση 1.9mg. **α)** Με βάση τα ευρήματα στα δύο δείγματα να κατασκευάσετε ένα 95% διάστημα εμπιστοσύνης για τη διαφορά της μέσης ποσότητας λυκοπενίου που περιέχεται στις συσκευασίες των 200gr της εταιρείας Β από τη μέση ποσότητα λυκοπενίου που περιέχεται στις συσκευασίες των 200gr της εταιρείας Α. **β)** Με βάση το 95% διάστημα εμπιστοσύνης που υπολογίσατε στο ερώτημα (α), μπορείτε να συμπεράνετε αν οι δύο πληθυσμιακοί μέσοι διαφέρουν ή όχι; **γ)** Σε επίπεδο σημαντικότητας 5%, η

διαφορά που παρατηρείται στους δύο δειγματικούς μέσους είναι άραγε στατιστικά σημαντική; Σε επίπεδο σημαντικότητας 1%; **δ)** Τι πρέπει να υπολογίσουμε για να διαπιστώσουμε πόσο στατιστικά σημαντική είναι αυτή η διαφορά; **ε)** Η *P-τιμή* που δίνουν τα ευρήματα στα δύο δείγματα, θα μπορούσε να είναι ίση με 0.0004; Εξηγήστε. **στ)** Ποιες υποθέσεις χρειάστηκε να κάνετε για να απαντήσετε στα ερωτήματα (α) και (γ);

**3° Θέμα [10]** Μια γαλλική εισαγωγική εταιρεία νωπών λαχανικών, εισάγει στη Γαλλία νωπά λαχανικά από έναν ελληνικό αγροτικό συνεταιρισμό. Από κάθε παρτίδα νωπών λαχανικών που φθάνει στις εγκαταστάσεις της από το συνεταιρισμό, παίρνει ένα τυχαίο δείγμα και κάνει ποιοτικό έλεγχο. Η εταιρεία επιστρέφει την παρτίδα αν με βάση το δείγμα συμπεράνει ότι το ποσοστό των λαχανικών που είναι εκτός προδιαγραφών υπερβαίνει το 7%, αλλιώς αποδέχεται την παρτίδα. **α)** Ποια μηδενική υπόθεση και έναντι ποιας εναλλακτικής πρέπει να ελέγχει η εισαγωγική εταιρεία; **β)** Διατυπώστε με όρους του προβλήματος το σφάλμα τύπου I και το σφάλμα τύπου II του ελέγχου που κάνει η εταιρεία. **γ)** Από τη σκοπιά της εισαγωγικής εταιρείας, ποιο σφάλμα είναι πιο σοβαρό; Αντίστοιχα, από τη σκοπιά του συνεταιρισμού;

**4° Θέμα [20]** Ο πίνακας ANOVA που ακολουθεί προέκυψε από ένα πείραμα που έγινε με βάση το σχέδιο των *τυχαιοποιημένων πλήρων ομάδων*.

Πηγή μεταβλητότητας	B.E.	Άθροισμα τετραγώνων SS	Μέσο άθροισμα τετραγώνων MS	Κριτήριο F
Επεμβάσεις	2	38.00		
Blocks		61.67		
Σφάλμα	6			
Ολική	11	111.00		

**α)** Πόσα blocks δημιουργήθηκαν; **β)** Πόσες παρατηρήσεις ελήφθησαν από κάθε block; **γ)** Να συμπληρώσετε τον πίνακα. **δ)** Σε επίπεδο σημαντικότητας 5%, υποστηρίζουν αυτά τα δεδομένα ότι υπάρχουν στατιστικά σημαντικές διαφορές μεταξύ των μέσων των επεμβάσεων; **ε)** Κρίνετε ότι επιβεβαιώνεται η ορθότητα της επιλογής του συγκεκριμένου πειραματικού σχεδίου;

**5° Θέμα [20]** Σύμφωνα με ένα μοντέλο κληρονομικότητας, οι τρεις τύποι απογόνων, A, B και Γ, που προκύπτουν από διασταύρωση ορισμένου είδους πειραματόζωων, πρέπει να βρίσκονται σε αναλογία 9:3:4, αντίστοιχα. Σε ένα σχετικό πείραμα, από 64 απογόνους που προέκυψαν, 34 βρέθηκαν να είναι τύπου A, 10 τύπου B, τα 20 τύπου Γ. Σε επίπεδο σημαντικότητας 1%, αυτά τα πειραματικά δεδομένα δίνουν άραγε σημαντικές αποδείξεις εναντίον του μοντέλου κληρονομικότητας;

**Θα σας χρειασθούν κάποιες από τις παρακάτω τιμές,  $F_{n,m;\alpha}$ , της κατανομής  $F_{n,m}$  :**

$$F_{4;12;0.05} = 3.26, \quad F_{2;6;0.05} = 5.14, \quad F_{6;2;0.05} = 19.33, \quad F_{6;3;0.05} = 8.94, \quad F_{3;6;0.05} = 4.76, \\ F_{11;6;0.05} = 4.03, \quad F_{11;2;0.05} = 19.40, \quad F_{6;11;0.05} = 3.9, \quad F_{2;11;0.05} = 3.98$$

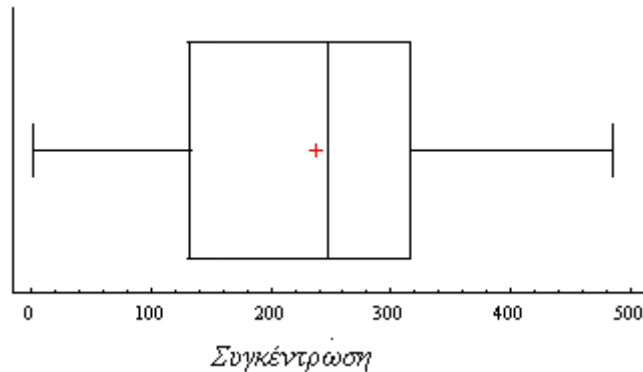
Από τα θέματα 4 και 5 πρέπει να επιλέξετε ένα. Δηλαδή, πρέπει να απαντήσετε είτε στα θέματα 1, 2, 3 και 4 είτε στα θέματα 1, 2, 3 και 5. Για το άριστα (10) απαιτούνται 100 μόρια και για τη βάση (5), 50 μόρια.

### Ενδεικτικές απαντήσεις

**1<sup>ο</sup> Θέμα: α)**  $Q_3 - Q_1 = 187 \Leftrightarrow Q_3 = 187 + Q_1 = 187 + 130.5 = 317.5$ .

Το ανώτερο εσωτερικό φράγμα είναι  $Q_3 + 1.5 \cdot (Q_3 - Q_1) = 317.5 + 1.5 \cdot 187 = 598$ , επομένως η άνω οριακή τιμή είναι η τιμή 485 (η μεγαλύτερη παρατήρηση που είναι ίση ή μικρότερη από 598).

Το κατώτερο εσωτερικό φράγμα είναι  $Q_1 - 1.5 \cdot (Q_3 - Q_1) = 130.5 - 1.5 \cdot 187 = -150$ , επομένως η κάτω οριακή τιμή είναι η τιμή 1.70 (η μικρότερη παρατήρηση που είναι ίση ή μεγαλύτερη από -150).



Ακραίες παρατηρήσεις δεν υπάρχουν.

**β)**  $z = \frac{118 - 236.73}{141.55} \approx -0.84$ . Δηλαδή, η τιμή  $118 \mu\text{gr}/\text{gr}$  βρίσκεται 0.84 τυπικές

αποκλίσεις κάτω από τη μέση τιμή του δείγματος. **γ)** Ότι ανήκει στο 10% των δελφινιών του δείγματος με τη μεγαλύτερη συγκέντρωση υδραργύρου (πάνω από  $452.5 \mu\text{gr}/\text{gr}$ . **δ)** Το 50%. **ε)** Με την υπόθεση ότι το δείγμα προέρχεται από κανονικό πληθυσμό, ένα 95% διάστημα εμπιστοσύνης για τη μέση συγκέντρωση υδραργύρου

στο συκώτι δελφινιών στη Μεσόγειο είναι  $\bar{x} \pm t_{n-1;0.025} \frac{s}{\sqrt{n}}$  ή  $236.73 \pm t_{27;0.025} \frac{141.55}{\sqrt{28}}$

ή  $236.73 \pm 2.052 \cdot 26.75$  ή  $236.73 \pm 54.89$  ή  $[181.84, 291.62]$ . Αυτό σημαίνει ότι: Το διάστημα  $[181.84, 291.62]$ , έχει 95% πιθανότητα να περιέχει την άγνωστη (πληθυσμιακή) μέση συγκέντρωση υδραργύρου στο συκώτι δελφινιών στη Μεσόγειο.

**στ)** Με 95% πιθανότητα, το περιθώριο σφάλματος της εκτίμησης είναι  $54.89 \mu\text{gr}/\text{gr}$ . Το περιθώριο σφάλματος θα αυξηθεί. **ζ)** Ότι το δείγμα προέρχεται από κανονικό πληθυσμό. Ένας τρόπος για να ελέγξουμε αν ευσταθεί αυτή η υπόθεση είναι να κάνουμε  $\chi^2$  έλεγχο καλής προσαρμογής των δεδομένων του δείγματος στην κανονική κατανομή με μέση τιμή και διασπορά που θα εκτιμήσουμε από το δείγμα. **η)** Όσο αυξάνεται το μέγεθος του δείγματος, τόσο πιο ακριβής η εκτίμηση της πληθυσμιακής μέσης τιμής  $\mu$  από την  $\bar{X}$ , ή αλλιώς, όσο αυξάνεται το μέγεθος του δείγματος τόσο αυξάνεται η πιθανότητα η εκτίμηση να είναι «κοντά» στην τιμή της πληθυσμιακής μέσης τιμής (και το τυπικό σφάλμα τείνει στο μηδέν).

**2<sup>ο</sup> Θέμα:** Έστω  $\mu_1$  η (άγνωστη) μέση τιμή της  $X$  και  $\mu_2$  η (άγνωστη) μέση τιμή της  $Y$ . Δίνονται:  $\bar{x} = 24 \text{mg}$  με  $s_1 = 2 \text{mg}$ ,  $\bar{y} = 22.6 \text{mg}$  με  $s_2 = 1.9 \text{mg}$  και  $n_1 = n_2 = 25$ .

**α)** Με την υπόθεση ότι τα δείγματα προέρχονται από κανονικούς πληθυσμούς με ίσες διασπορές, ένα 95% διάστημα εμπιστοσύνης για τη διαφορά  $\mu_1 - \mu_2$  είναι:

$$(\bar{x} - \bar{y}) \pm t_{n_1+n_2-2;0.025} \cdot s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad \text{ή} \quad (24 - 22.6) \pm t_{48;0.025} \cdot s \cdot \sqrt{\frac{1}{25} + \frac{1}{25}}$$

$$\text{όπου } s = \sqrt{\frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{24 \cdot 2^2 + 24 \cdot 1.9^2}{48}} = 1.95.$$

Άρα, το ζητούμενο διάστημα είναι,  $1.4 \pm 1.96 \cdot 1.95 \cdot \sqrt{\frac{1}{25} + \frac{1}{25}}$  ή  $1.4 \pm 1.08$  ή  $[0.32, 2.48]$ .

**β)** Επειδή το 95% διάστημα εμπιστοσύνης  $[0.32, 2.48]$ , δεν περιέχει την τιμή 0, μπορούμε να συμπεράνουμε ότι οι δύο πληθυσμιακοί μέσοι διαφέρουν. Η πιθανότητα το συμπέρασμα αυτό να είναι λάθος είναι το πολύ 5%.

**γ)** Θα κάνουμε τον έλεγχο της μηδενικής υπόθεσης  $H_0 : \mu_1 = \mu_2$  έναντι της εναλλακτικής  $H_1 : \mu_1 \neq \mu_2$  σε επίπεδο σημαντικότητας  $\alpha = 0.05$ . Με την υπόθεση ότι τα δείγματα προέρχονται από κανονικούς πληθυσμούς με ίσες διασπορές, η

απορριπτική περιοχή του ελέγχου είναι:  $t = \frac{|\bar{x} - \bar{y}|}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > t_{n_1+n_2-2; 0.025}$  ή

$$t = \frac{|\bar{x} - \bar{y}|}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > t_{48; 0.025} \quad \text{ή} \quad t = \frac{|\bar{x} - \bar{y}|}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > 1.96$$

και επειδή,  $t = \frac{|1.4|}{1.95 \cdot \sqrt{\frac{1}{25} + \frac{1}{25}}} = 2.538 > 1.96$ , η μηδενική υπόθεση, σε επίπεδο

σημαντικότητας 5% απορρίπτεται και επομένως σε επίπεδο σημαντικότητας 5%, η διαφορά που παρατηρείται στους δύο δειγματικούς μέσους είναι στατιστικά σημαντική. Σε επίπεδο σημαντικότητας 1%, η απορριπτική περιοχή του ελέγχου

είναι:  $t = \frac{|\bar{x} - \bar{y}|}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > t_{48; 0.005}$  ή  $t = \frac{|\bar{x} - \bar{y}|}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > 2.576$  και επειδή, η τιμή

$t = 2.538$  δεν βρίσκεται στην απορριπτική περιοχή, η μηδενική υπόθεση, σε επίπεδο σημαντικότητας 1%, δεν απορρίπτεται, δηλαδή, σε επίπεδο σημαντικότητας 1%, η διαφορά που παρατηρείται στους δύο δειγματικούς μέσους δεν είναι στατιστικά σημαντική.

**δ)** Την  $P$ -τιμή που δίνουν για αυτόν τον έλεγχο τα δύο δείγματα.

**ε)** Όχι, γιατί όπως διαπιστώσαμε στο ερώτημα (γ), η μηδενική υπόθεση σε επίπεδο σημαντικότητας 1% δεν απορρίπτεται ενώ με βάση αυτή την  $P$ -τιμή θα έπρεπε να απορρίπτεται. **στ)** Ότι τα δείγματα προέρχονται από κανονικούς πληθυσμούς με ίσες διασπορές.

**3<sup>ο</sup> Θέμα:** **α)** Έστω  $p$  το (άγνωστο) ποσοστό των εκτός προδιαγραφών λαχανικών της παρτίδας που ελέγχεται. Η εισαγωγική εταιρεία προφανώς πρέπει να ελέγξει (σε κάποιο επίπεδο σημαντικότητας) τη μηδενική υπόθεση,  $H_0 : p = 0.07$  έναντι της εναλλακτικής  $H_1 : p > 0.07$ . **β)** Σφάλμα τύπου I, θα συμβεί όταν η εισαγωγική εταιρεία επιστρέψει την παρτίδα ενώ στην πραγματικότητα το ποσοστό των εκτός προδιαγραφών λαχανικών της παρτίδας δεν ξεπερνάει το 7%. Σφάλμα τύπου II, θα συμβεί όταν η εισαγωγική εταιρεία κάνει δεκτή την παρτίδα ενώ στην πραγματικότητα το ποσοστό των εκτός προδιαγραφών λαχανικών της παρτίδας

ξεπερνάει το 7%. γ) Από τη σκοπιά της εισαγωγικής εταιρείας, προφανώς το σφάλμα τύπου II ενώ από τη σκοπιά του συνεταιρισμού, προφανώς το σφάλμα τύπου I.

4<sup>ο</sup> Θέμα: α) 4 β) 3

γ)

Πηγή μεταβλητότητας	B.E.	Άθροισμα τετραγώνων SS	Μέσο άθροισμα τετραγώνων MS	Κριτήριο F
Επεμβάσεις	2	38.00	19	10.05
Blocks	3	61.67	20.56	10.88
Σφάλμα	6	11.33	1.89	
Ολική	11	111.00		

δ) Θα κάνουμε, σε επίπεδο σημαντικότητας  $\alpha = 0.05$ , τον έλεγχο της μηδενικής υπόθεσης,  $H_0$ : Δεν υπάρχουν διαφορές μεταξύ των μέσων των επεμβάσεων έναντι της εναλλακτικής,  $H_1$ : Υπάρχουν διαφορές μεταξύ των μέσων των επεμβάσεων.

Η απορριπτική περιοχή του ελέγχου είναι,  $F_{Tr} = \frac{MSTr}{MSE} > F_{2,6;0.05}$  ή  $F_{Tr} > 5.14$  και

επειδή  $F_{Tr} = 10.05 > 5.14$ , η μηδενική υπόθεση, σε επίπεδο σημαντικότητας 5%, απορρίπτεται και επομένως, σε επίπεδο σημαντικότητας 5%, τα δεδομένα υποστηρίζουν ότι υπάρχουν στατιστικά σημαντικές διαφορές μεταξύ των μέσων των επεμβάσεων.

γ) Για να απαντήσουμε, πρέπει να κάνουμε, σε επίπεδο σημαντικότητας 5%, τον έλεγχο της μηδενικής υπόθεσης,  $H_0$ : Δεν υπάρχουν διαφορές μεταξύ των μέσων των blocks, έναντι της εναλλακτικής,  $H_1$ : Υπάρχουν διαφορές μεταξύ των μέσων των blocks.

Η απορριπτική περιοχή του ελέγχου είναι,  $F_{Bl} = \frac{MSBl}{MSE} > F_{3,6;0.05}$  ή  $F_{Bl} > 4.76$  και

επειδή  $F_{Bl} = 10.88 > 4.76$ , η μηδενική υπόθεση, σε επίπεδο σημαντικότητας 5%, απορρίπτεται και επομένως, σε επίπεδο σημαντικότητας 5%, τα δεδομένα υποστηρίζουν ότι υπάρχουν στατιστικά σημαντικές διαφορές μεταξύ των μέσων των blocks, κάτι που, λογικά, ανέμενε ο ερευνητής και γι' αυτό επέλεξε αυτό το πειραματικό σχέδιο.

5<sup>ο</sup> Θέμα: Έστω  $p_1$  η πιθανότητα ένας απόγονος να είναι τύπου Α,  $p_2$  η πιθανότητα ένας απόγονος να είναι τύπου Β και  $p_3$  η πιθανότητα ένας απόγονος να είναι τύπου Γ.

Σε επίπεδο σημαντικότητας  $\alpha = 0.01$ , θα ελέγξουμε τη μηδενική υπόθεση,

$H_0$ :  $p_1 = \frac{9}{16}$  και  $p_2 = \frac{3}{16}$  και  $p_3 = \frac{4}{16}$  έναντι της εναλλακτικής,  $H_1$ :  $p_1 \neq \frac{9}{16}$  ή

$p_2 \neq \frac{3}{16}$  ή  $p_3 \neq \frac{4}{16}$ . Θα κάνουμε  $\chi^2$  έλεγχο καλής προσαρμογής. Με την υπόθεση

ότι η  $H_0$  είναι αληθής, θα υπολογίσουμε για κάθε τύπο απογόνων την αναμενόμενη (θεωρητική) συχνότητα,  $E_i = 64 \cdot p_i$ ,  $i = 1,2,3$ .

	A	B	Γ	Σύνολο
$O_i$	34	10	20	64
$p_i$	9/16	3/16	4/16	1.00
$E_i = 64 \cdot p_i$	36	12	16	64

Η απορριπτική περιοχή του ελέγχου είναι:  $X^2 = \sum_{i=1}^3 \frac{(O_i - E_i)^2}{E_i} > \chi_{2;0.01}^2$  ή

$$X^2 = \sum_{i=1}^3 \frac{(O_i - E_i)^2}{E_i} > 9.21 \text{ και επειδή,}$$

$$X^2 = \frac{(34-36)^2}{36} + \frac{(10-12)^2}{12} + \frac{(20-16)^2}{16} = 1.44,$$

η μηδενική υπόθεση, σε επίπεδο σημαντικότητας 1% δεν απορρίπτεται και επομένως αυτά τα πειραματικά δεδομένα, σε επίπεδο σημαντικότητας 1%, δε δίνουν στατιστικά σημαντικές αποδείξεις εναντίον του μοντέλου κληρονομικότητας.