
5. ΠΟΛΥΣΤΑΔΙΑΚΗ ΔΕΙΓΜΑΤΟΛΗΨΙΑ (MULTISTAGE SAMPLING)

Η υποδειγματοληψία που αποτελεί τεχνική ρουτίνας στις εργαστηριακές μελέτες δεν είναι τίποτε άλλο παρά μια περίπτωση δισταδιακής δειγματοληψίας.

Η πολυσταδιακή δειγματοληψία είναι μια κοινότατη στρατηγική. Η βασική της φιλοσοφία βασίζεται στην εκμετάλλευση της ιεραρχημένης δομής του πληθυσμού στόχου. Πράγματι συχνά οι πληθυσμοί αποτελούνται από ομάδες ή υποσύνολα που ονομάζονται πρωτογενείς μονάδες. Κάθε μια από αυτές μπορεί να αποτελείται από ενότητες, ομάδες, συναθροίσεις κ.λ.π. που αναφέρονται σε δευτερογενείς μονάδες. Κάθε δευτερογενής μονάδα μπορεί να αποτελείται από άτομα ή στοιχεία που αποτελούν τις τριτογενείς μονάδες. Η ιεραρχική αυτή δομή μπορεί να συνεχιστεί ανάλογα με την περίπτωση σε αρκετά επίπεδα. Σε κάθε ένα από αυτά τα επίπεδα μπορεί να γίνει μια τυχαία δειγματοληψία.

Για παράδειγμα οι φοιτητές της επικράτειας είναι το σύνολο των φοιτητών των διαφόρων Πανεπιστημίων (πρωτογενείς μονάδες). Κάθε Πανεπιστήμιο αποτελείται από τμήματα (δευτερογενείς μονάδες), κάθε τμήμα από φοιτητές (τριτογενείς μονάδες).

Στην περίπτωση μιας αλιευτικής μελέτης σε μια περιφέρεια, τα λιμάνια αποτελούν τις πρωτογενείς μονάδες, τα αλιευτικά σκάφη κάθε λιμανιού τις δευτερογενείς μονάδες και οι κάσες με το αλιεύμα του κάθε σκάφους τις τριτογενείς μονάδες. Για την εκτίμηση λοιπόν κάποιων βιομετρικών χαρακτηριστικών των αλιευμάτων επιλέγουμε τυχαία κάποια από τα λιμάνια της περιφέρειας, σε κάθε ένα από αυτά επιλέγουμε τυχαία κάποια σκάφη και από αυτά επιλέγουμε τυχαία κάποιες κάσες με ψάρια. Όλα τα άτομα κάθε κασέλας θα καταγραφούν και θα μετρηθούν. Έχουμε με αυτό τον τρόπο οργανώσει μια τρισταδιακή δειγματοληψία.

Μια κοινότατη εφαρμογή της δισταδιακής δειγματοληψίας που εφαρμόζεται σε μέθοδο ρουτίνας στις εργαστηριακές αναλύσεις είναι η υποδειγματοληψία. Από ένα σύνολο πειραματόζωων επιλέγουμε κάποια στην τύχη. Σε κάθε ένα από αυτά γίνεται λήψη ποσότητας αίματος π.χ. 5ml και από την ποσότητα αυτή μόνο 0.5 ml χρησιμοποιούνται στις

αναλύσεις. Μπορούμε δηλαδή να θεωρήσουμε ότι τα N πειραματόζωα από τα οποία επιλέγουμε n αποτελούν τις πρωτογενείς μονάδες η δε ποσότητα των 0.5ml που χρησιμοποιείται στις αναλύσεις είναι μια από τις 10 δευτερογενείς μονάδες που μπορεί να δώσει το δείγμα των 5ml.

Όταν οι μονάδες του πρώτου επιπέδου έχουν το ίδιο μέγεθος η ανάλυση και οι υπολογισμοί απλοποιούνται. Δυστυχώς όμως στη φύση η περίπτωση αυτή είναι σπάνια και έτσι θα δούμε στη συνέχεια τους εκτιμητές που έχουν τη γενικότερη εφαρμογή χωρίς να απαιτούν ιδιαίτερες συνθήκες για την εφαρμογή τους. Αλλά και σε αυτή την γενική περίπτωση κάποιες διευκρινίσεις είναι απαραίτητες. Είναι δυνατόν για παράδειγμα ο συνολικός αριθμός των δευτερογενών μονάδων που αποτελούν τον πληθυσμό να είναι γνωστός (π.χ. ο συνολικός αριθμός των σκαφών στην περίπτωση της αλιευτικής μελέτης) ενώ σε άλλες περιπτώσεις είναι άγνωστος και εκτιμάται από τα δεδομένα του δείγματος (παράδειγμα 5.1).

Η πολυσταδιακή δειγματοληψία έχει δυο εμφανή πλεονεκτήματα:

- Δεν απαιτεί τον αναλυτικό κατάλογο όλων των στοιχείων του πληθυσμού ώστε να επιλεγούν τυχαία κάποια από αυτά (π.χ. η λίστα με όλες τις κασέλες αλιευμάτων όλων των πλοίων, όλων των λιμανιών της περιφέρειας) αλλά περιοριζόμαστε στη λίστα των στοιχείων της δευτερογενούς μονάδας που επιλέχθηκε (το συγκεκριμένο πλοίο του συγκεκριμένου λιμανιού που είναι πλέον εφικτό)
- Σε σχέση με άλλες στρατηγικές περιορίζει σε μεγάλο βαθμό τις μετακινήσεις με αποτέλεσμα τη μείωση του κόστους της μελέτης. Στην περίπτωση των Πανεπιστημίων που αναφέρθηκε αφού επιλεγεί ένα ίδρυμα (πρωτογενείς μονάδα) οι μετακινήσεις περιορίζονται στο εσωτερικό του ιδρύματος.

5.1. ΕΚΤΙΜΗΣΗ ΠΑΡΑΜΕΤΡΟΥ ΤΟΥ ΠΛΗΘΥΣΜΟΥ ΣΤΗ ΔΙΣΤΑΔΙΑΚΗ ΔΕΙΓΜΑΤΟΛΗΨΙΑ

ΕΚΤΙΜΗΤΕΣ

Ο πληθυσμός είναι χωρισμένος σε I ομάδες (πρωτογενείς μονάδες). Κάθε πρωτογενής μονάδα αποτελείται από στοιχεία (δευτερογενείς μονάδες) ο αριθμός των οποίων μπορεί να κυμαίνεται από τη μία ομάδα στην άλλη. Αυτή είναι και η γενικότερη περίπτωση. Αν ο αριθμός των στοιχείων δηλ. το μέγεθος των πρωτογενών μονάδων είναι σταθερός τότε οι τύποι που ακολουθούν απλοποιούνται σε μεγάλο βαθμό.

Από τις N ομάδες επιλέγονται n . Από τις ομάδες αυτές επιλέγονται στοιχεία. Έτσι από την ομάδα i που αποτελείται από M_i στοιχεία επιλέγονται τυχαία m_i .

Δεδομένα

ομάδα 1: από τα M_1 στοιχεία επελέγησαν m_1 ; $y_{11}, y_{12}, y_{13}, \dots, y_{1j}, \dots, y_{1m1}$

ομάδα 2: από τα M_2 στοιχεία επελέγησαν m_2 ; $y_{21}, y_{22}, y_{23}, \dots, y_{2j}, \dots, y_{2m2}$

ομάδα i : από τα M_i στοιχεία επελέγησαν m_i ; $y_{i1}, y_{i2}, y_{i3}, \dots, y_{ij}, \dots, y_{imi}$

ομάδα n : από τα M_n στοιχεία επελέγησαν m_n ; $y_{n1}, y_{n2}, y_{n3}, \dots, y_{nj}, \dots, y_{nmn}$

Ο δείκτης i αναφέρεται στις πρωτογενείς αυτές μονάδες και ο δείκτης j στα στοιχεία των ομάδων (δευτερογενείς μονάδες).

Παράμετροι του δείγματος

Τά ακόλουθα σύμβολα αναφέρονται στη στρώση h . Ανάλογοι τύποι ισχύουν και για τις υπόλοιπες στρώσεις του πληθυσμού.

N : συνολικός αριθμός πρωτογενών μονάδων του πληθυσμού

n : ο αριθμός των πρωτογενών μονάδων στο δείγμα (πρώτο στάδιο δειγματοληψίας)

M_i : ο συνολικός αριθμός δευτερογενών μονάδων της ομάδας i (μιας από τις n που επιλέχθηκαν).

m_i : ο αριθμός δευτερογενών μονάδων που επιλέχθηκαν από την i ομάδα .

y_{ij} : στοιχείο της ομάδας i .

$f_1 = \frac{n}{N}$ δειγματοληπτικό κλάσμα του πρώτου σταδίου δειγματοληψίας

$f_{2i} = \frac{m_i}{M_i}$ δειγματοληπτικό κλάσμα του δεύτερου σταδίου δειγματοληψίας της ομάδας i .

$\bar{M} = \frac{\sum_{i=1}^N M_i}{N}$ μέσος αριθμός δευτερογενών μονάδων ανά πρωτογενή μονάδα.

$\bar{y}_i = \frac{\sum_{j=1}^{m_i} y_{ij}}{m_i}$ μέση τιμή στοιχείου της πρωτογενούς μονάδας i

Εκτίμηση της μέσης τιμής της δευτερογενούς μονάδας του πληθυσμού

1η περίπτωση: το \bar{M} είναι γνωστό

$$\bar{y} = \frac{1}{n\bar{M}} \sum_{i=1}^n M_i \bar{y}_i \quad s_i^2 = \frac{\sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2}{m_i - 1}$$

$$v(\bar{y}) = (1 - f_1) \frac{1}{n(n-1)} \sum_{i=1}^n \left(\frac{M_i}{\bar{M}} \bar{y}_i - \bar{y} \right)^2 + \frac{1}{nN} \sum_{i=1}^n \frac{M_i^2}{\bar{M}^2} (1 - f_{2i}) \frac{s_i^2}{m_i}$$

$$\text{εάν το } N \text{ είναι μεγάλο τότε } v(\bar{y}) = \frac{\sum_{i=1}^n \left(\frac{M_i}{\bar{M}} \bar{y}_i - \bar{y} \right)^2}{n(n-1)}$$

⚠ Σ' αυτή την περίπτωση το \bar{M} πρέπει να είναι γνωστό καθώς και το N εάν είναι μικρό.

2η περίπτωση: το \bar{M} είναι άγνωστό

σ' αυτή την περίπτωση μια εκτίμησή του είναι $\hat{M} = \frac{\sum_{i=1}^n M_i}{n}$

$$\bar{y} = \frac{\sum_{i=1}^n M_i \bar{y}_i}{\sum_{i=1}^n M_i} \quad s_i^2 = \frac{\sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2}{m_i - 1}$$

$$v(\bar{y}) = (1 - f_1) \frac{1}{n(n-1)} \sum_{i=1}^n \frac{M_i^2}{\hat{M}^2} (\bar{y}_i - \bar{y})^2 + \frac{1}{nN} \sum_{i=1}^n \frac{M_i^2}{\hat{M}^2} (1 - f_{2i}) \frac{s_i^2}{m_i}$$

$$\text{στην περίπτωση που } n/N \text{ είναι μικρό τότε } v(\bar{y}) = \frac{\sum_{i=1}^n \frac{M_i^2}{\hat{M}^2} (\bar{y}_i - \bar{y})^2}{n(n-1)}$$

$$P\{\bar{y} - t_{\alpha/2} \sqrt{v(\bar{y})} < \bar{Y} < \bar{y} + t_{\alpha/2} \sqrt{v(\bar{y})}\} = 1 - \alpha$$

⊗ Όπως αναφέρθηκε και στο κεφάλαιο 3 ο εκτιμητής αυτός είναι μεροληπτικός το μέγεθος όμως της μεροληψίας μειώνεται γρήγορα με την αύξηση του n . Εάν το N είναι μικρό πρέπει και αυτό να είναι γνωστό.

Εκτίμηση της μέσης τιμής της πρωτογενούς μονάδας του πληθυσμού

$$\bar{y} = \frac{\sum_{i=1}^n M_i \bar{y}_i}{n} \quad s_i^2 = \frac{\sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2}{m_i - 1}$$

$$v(\bar{y}) = (1 - f_1) \frac{1}{n(n-1)} \sum_{i=1}^n (M_i \bar{y}_i - \bar{y})^2 + \frac{1}{nN} \sum_{i=1}^n (1 - f_{2i}) M_i^2 \frac{s_i^2}{m_i}$$

εαν το N είναι μεγάλο τότε $v(\bar{y}) = \frac{(1 - f_1)}{n(n-1)} \sum_{i=1}^n (M_i \bar{y}_i - \bar{y})^2$

$$P\{\bar{y} - t_{\alpha/2} \sqrt{v(\bar{y})} < \bar{Y} < \bar{y} + t_{\alpha/2} \sqrt{v(\bar{y})}\} = 1 - \alpha$$

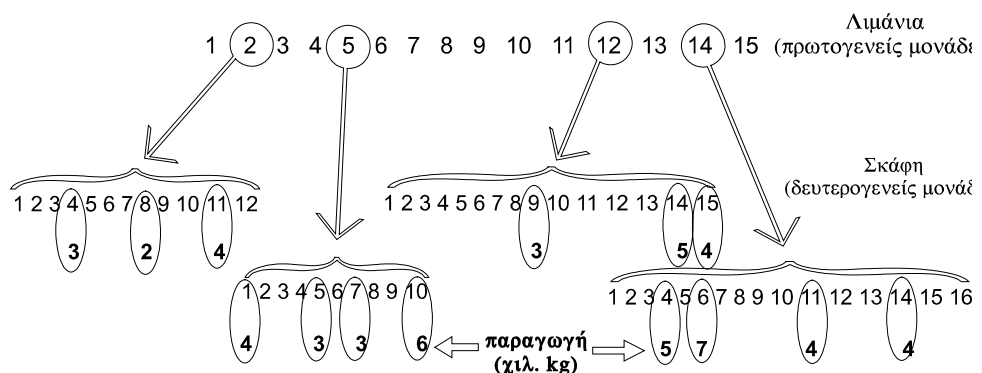
Εκτίμηση του συνόλου (\hat{Y}) του πληθυσμού

$$\hat{Y} = N \cdot \bar{y} \quad v(\hat{Y}) = N^2 \cdot v(\bar{y})$$

$$P\{\hat{Y} - t_{\alpha/2} \sqrt{v(\hat{Y})} < Y < \hat{Y} + t_{\alpha/2} \sqrt{v(\hat{Y})}\} = 1 - \alpha$$

☞ ΠΑΡΑΔΕΙΓΜΑ 5.1

Σε ένα νησί στο οποίο δεν υπάρχει οργανωμένη κεντρική αγορά για τη διακίνηση της παραγωγής της τοπικής αλιείας επιθυμούμε να εκτιμήσουμε το μέγεθος της για συγκεκριμένο είδος ψαριού. Πιο συγκεκριμένα, εν όψη της λήψεως μέτρων η παραγωγή του μηνός Μαρτίου κρίνεται καθοριστική. Ο αλιευτικός στόλος του νησιού, ο οποίος είναι σχετικά ομοιογενής, αποτελείται από 225 σκάφη τα οποία μοιράζονται σε 15 λιμάνια και λιμενίσκους του νησιού. Δεδομένης της δύσκολης πρόσβασης σε όλες τις τοποθεσίες καθώς και των περιορισμένων κονδυλίων για τη μελέτη, ζητείται η οργάνωση και πραγματοποίηση δειγματοληψίας με σκοπό την εκτίμηση της συνολικής παραγωγής του συγκεκριμένου είδους το μήνα Μάρτιο σε όλο το νησί, της μέσης παραγωγής ανά λιμάνι καθώς και της μέσης ανά σκάφος παραγωγής.



Η ιεραρχική δομή με τα δυο επίπεδα οργάνωσης του πληθυσμού των παραγωγών (λιμάνια και σκάφη) που σχηματικά παρουσιάζεται στο παραπάνω διάγραμμα οδηγεί λογικά στην οργάνωση μιας δισταδιακής δειγματοληψίας. Από τα 15 λιμάνια του νησιού επιλέγονται τυχαία 4. Από τα σκάφη αυτών των τεσσάρων λιμανιών επιλέγονται πάλι τυχαία κάποια των οποίων η μηνιαία παραγωγή καταγράφεται. Τα πλεονεκτήματα αυτής της δειγματοληψίας είναι: περιορισμός των μετακινήσεων σε 4 μόνο λιμάνια και μείωση του γενικότερου κόστους λόγω του περιορισμένου αριθμού σκαφών προς παρακολούθηση ανά λιμάνι. Επιπλέον αποφεύγουμε και το ακόλουθο πρόβλημα. Σε περιοχές που οι αποστάσεις μεταξύ αγκυροβολίων είναι μικρές και οι παρεχόμενες υπηρεσίες ελλιμενισμού περιορισμένες οι μετακινήσεις των σκαφών από το ένα λιμάνι στο άλλο είναι συχνές. Έτσι γνωρίζουμε τον συνολικό αριθμό των σκαφών του νησιού αλλά είναι δύσκολη η κατανομή τους σε λιμάνια. Με τη δισταδιακή δειγματοληψία χρειάζεται να σχηματίσουμε τον κατάλογο των σκαφών μόνο για τα 4 λιμάνια που έχουν επιλεγεί..

N=15		n=	
Μέγεθος πρωτογενών μονάδων (M_i) (αριθμός σκαφών ανά λιμάνι)			
12	10	15	16
$\sum_{i=1}^n M_i = 53$			
Δεδομένα (y_{ij}) (παραγωγή του σκάφους j του λιμανιού i)			
3	4	3	5
2	3	5	7
4	3	4	4
	6		4
m_i			
$f_1 =$			
Δειγματοληπτικό κλάσμα 2γενών μονάδων (f_{2i})			
Μέσες τιμές 2γενων μονάδων ανά 1γενη μονάδα (\bar{y}_i) (μέση παραγωγή ανά σκάφος του λιμανιού i)			
Διασπορές των 2γενων μονάδων ανά 1γενη μονάδα (s^2_{i})			
$M_i \bar{y}_i$			

				$\sum_{i=1}^n M_i \bar{y}_i =$
Μέσος αριθμός στοιχείων (2γενων μον.) ανά 1γενή μονάδα $\bar{M} =$				

Εκτίμηση της μέσης τιμής της 2γενούς μονάδας $\bar{y} =$ (μέση παραγωγή ανά σκάφος)
Διασπορά της μέσης τιμής της 2γενούς μονάδας $s_{\bar{y}}^2 =$
Διάστημα εμπιστοσύνης (95%) ($t=$) $< \bar{y} <$
Εκτίμηση της μέσης τιμής 1γενούς μονάδας $\bar{y} =$ (μέση παραγωγή ανά λιμάνι)
Διασπορά της μέσης τιμής της 1γενούς μονάδας $s_{\bar{y}}^2 =$
Διάστημα εμπιστοσύνης (95%) $< \bar{y} <$
Εκτίμηση του συνόλου του πληθυσμού $\hat{Y} =$ (συνολική παραγωγή του μηνός Μαρτίου στο νησί)
Διασπορά του συνόλου $s_{\hat{Y}}^2 =$
Διάστημα εμπιστοσύνης (95%) $< Y <$

☞ ΠΑΡΑΔΕΙΓΜΑ 5.1

☞ ΠΑΡΑΔΕΙΓΜΑ 5.2

Σε μια περιοχή με 40 μικρές λίμνες έχει προβλεφθεί μελέτη είδους πτηνού και πιο συγκεκριμένα εκτίμηση του συνολικού αριθμού νεοσσών του είδους. Το είδος αυτό χτίζει τη φωλιά του σε απόκρημνες όχθες έτσι ώστε η καταγραφή και επίσκεψη των φωλεών να γίνεται μόνο από το νερό. Θεωρώντας ότι το κόστος μεταφοράς του πλωτού μέσου από τη μια λίμνη στην άλλη είναι πολύ μεγάλο και παράλληλα ότι η επίσκεψη των φωλεών είναι σχετικά δύσκολη ενώ η απλή καταγραφή τους εύκολη το ακόλουθο σχέδιο δειγματοληψίας αποφασίστηκε. Από τις 40 λίμνες επιλέχθηκαν τυχαία 7. Το σύνολο των φωλεών σε κάθε μια από τις 7 λίμνες κατεγράφη και ακολούθως μερικές φωλιές επιλέχθηκαν τυχαία σε κάθε λίμνη και μετρήθηκε ο αριθμός των νεοσσών σε κάθε μια από αυτές. Τα αποτελέσματα δίνονται στον ακόλουθο πίνακα. Με βάση αυτά τα στοιχεία γίνεται η εκτίμηση της αφθονίας των νεοσσών του πληθυσμού αλλά και η μέση τιμή νεοσσών ανά φωλιά και ανά λίμνη.

N=40				n=		
Μέγεθος πρωτογενών μονάδων (M_i)						
50	70	60	30	40	40	60
Δεδομένα (y_{ij})						
2	5	5	3	5	4	5
4	2	4	4	5	3	2
1	3	4	3	4	3	3
3	4	4	3	3		4
3	6	5	1			3
3	3	2				4
	2	3				6
		3				
m_i						
$f_1 =$						
Δειγματοληπτικό κλάσμα 2γενών μονάδων (f_{2i})						
Μέσες τιμές 1γενών μονάδων (\bar{y}_i)						
Διασπορές 1γενών μονάδων ($s^2_{y_i}$)						
$M_i \bar{y}_i$						
Εκτίμηση του μέσου αριθμού στοιχείων / 1γενή μονάδα $\hat{M} =$						

$$\sum_{i=1}^n M_i = 350$$

$$\sum_{i=1}^n M_i \bar{y}_i =$$

Εκτίμηση της μέσης τιμής της 2γενούς μονάδας $\bar{\bar{y}} =$ (αριθμός νεοσσών ανά φωλιά)
Διασπορά της μέσης τιμής της 2γενούς μονάδας $s_{\bar{\bar{y}}}^2 =$
Διάστημα εμπιστοσύνης (95%) ($t=$) $< \bar{\bar{y}} <$
Εκτίμηση της μέσης τιμής 1γενούς μονάδας $\bar{y} =$ (αριθμός νεοσσών ανά λίμνη)

Διασπορά της μέσης τιμής της 1γενούς μονάδας $s_y^2 =$	
Διάστημα εμπιστοσύνης (95%)	$<\bar{y}<$
Εκτίμηση του συνόλου του πληθυσμού (αφθονία των νεοσσών στην περιοχή) $\hat{Y} =$	
Διασπορά του συνόλου $S_{\hat{Y}}^2 =$	
Διάστημα εμπιστοσύνης (95%)	$<Y<$

☞ ΠΑΡΑΔΕΙΓΜΑ 5.2

5.2. ΠΛΕΟΝΕΚΤΗΜΑΤΑ ΚΑΙ ΜΕΙΟΝΕΚΤΗΜΑΤΑ ΤΗΣ ΠΟΛΥΣΤΑΔΙΑΚΗΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

↑ Πλεονεκτήματα ↑

- Δεν απαιτεί τον αναλυτικό κατάλογο όλων των στοιχείων του πληθυσμού ώστε να επιλεγούν τυχαία κάποια από αυτά αλλά περιοριζόμαστε στη λίστα των στοιχείων της πρωτογενούς μονάδας που επιλέχθηκε
- Σε σχέση με άλλες στρατηγικές περιορίζει σε μεγάλο βαθμό τις μετακινήσεις με αποτέλεσμα τη μείωση του κόστους της μελέτης
- Στην περίπτωση που η μελέτη προκαλεί φθορές στο φυσικό πληθυσμό ή το περιβάλλον του η στρατηγική αυτή επιτρέπει τον περιορισμό τους σε ορισμένες περιοχές
- Η τεχνική αυτή προσαρμόζεται εύκολα στην δομή των διαφόρων συστημάτων. Το μέγεθος των πρωτογενών μονάδων καθώς και το μέγεθος του δείγματος των δευτερογενών μπορούν εύκολα να βελτιστοποιηθούν.

↓ Μειονεκτήματα ↓

- Η αποδοτικότητα αυτής της στρατηγικής είναι περιορισμένη στην περίπτωση που τα στοιχεία κάθε πρωτογενούς μονάδας μοιάζουν μεταξύ τους και διαφέρουν από αυτά των άλλων πρωτογενών μονάδων
- Όπως φαίνεται και από τους εκτιμητές το σχέδιο αυτό είναι ή μπορεί να γίνει γρήγορα σχετικά πολύπλοκο στην οργάνωση και την ανάλυση των δεδομένων
- Λόγω της εκ των προτέρων διαίρεσης του πληθυσμού σε στρώσεις (κατηγορίες, ομάδες) κάποιες από τις στατιστικές αναλύσεις δεν εφαρμόζονται άμεσα.