Μετρήσεις

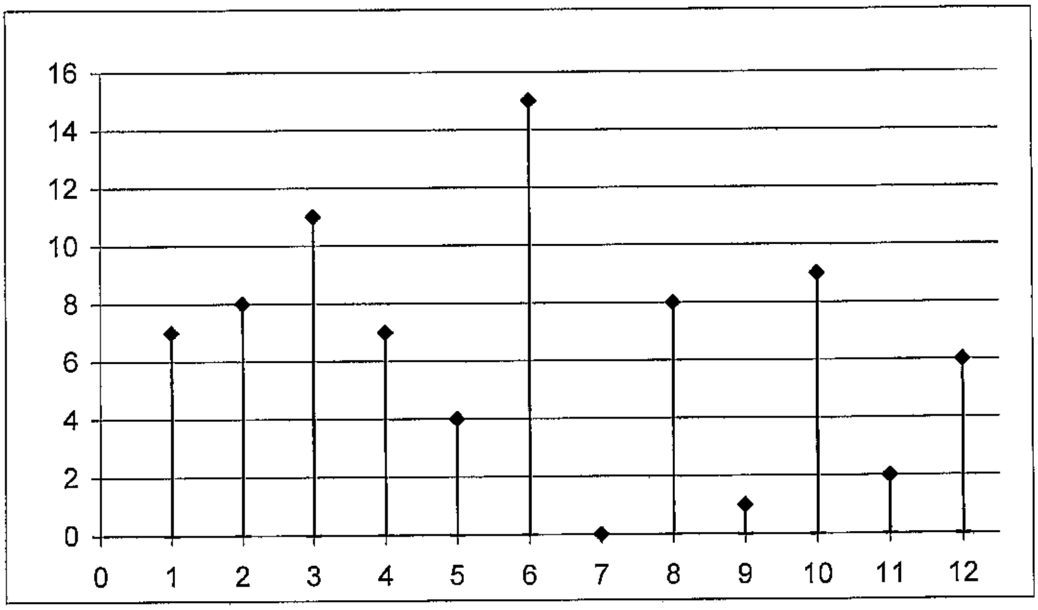
Όπως είναι γνωστό, οι ποσοτικές μεταβλητές διακρίνονται σύμφωνα με το μέγεθος το οποίο μετρούν, στις διακριτές, όπως π.χ. ο αριθμός των μελών μιας οικογένειας και στις συνεχείς, όπως π.χ. η παραγωγή σιτηρών της Βουλγαρίας Οι γραφικές απεικονίσεις αυτών των δύο κατηγοριών ποσοτικών μεταβλητών είναι διαφορετικές και η διαφοροποίηση τους οφείλεται στο ότι μ' αυτές, εκτός της παρουσίασης των κατανομών τους πρέπει να είναι εμφανές και το είδος της μεταβλητής που περιγράφουν.

Οι κατανομές των διακριτών μεταβλητών απεικονίζονται με **ακιδωτά διαγράμματα**. Στα ακιδωτά διαγράμματα, οι τιμές της ανεξάρτητης μεταβλητής, δηλαδή της διακριτής μεταβλητής, της οποίας την κατανομή συχνότητας επιθυμούμε να απεικονίσουμε, αντιστοιχούν σε διακεκριμένα σημεία του άξονα των τετμημένων (οριζόντιος άξονας). Στον άξονα των τεταγμένων (κατακόρυφος άξονας) αντιστοιχούν οι συχνότητες εμφάνισης των τιμών της μεταβλητής. Με τον τρόπο αυτό, σε κάθε τιμή της διακριτής μεταβλητής θα αντιστοιχεί ένα κατακόρυφο ευθύγραμμο τμήμα (ακίδα) του οποίου το μήκος εξαρτάται από την απόλυτη συχνότητα ή την σχετική συχνότητα αυτής της τιμής. Για παράδειγμα ο παρακάτω πίνακας περιέχει την κατανομή συχνοτήτων των εργαζόμενων του δείγματος που ήταν άνεργοι τον τελευταίο χρόνο ανάλογα με τους μήνες ανεργίας. Το αντίστοιχο ακιδωτό διάγραμμα εμφανίζεται στο γράφημα 2.7.

Πίνακας 2.11 Κατανομή συχνοτήτων ανάλογα με τους μήνες ανεργίας

ΜΗΝΕΣ ΑΝΕΡΓΙΑΣ	ΣΥΧΝΟΤΗΤΑ
1	7
2	8
3	11
4	7
5	4
6	15
7	0
8	8
9	1
10	9
11	2
12	6
Σύνολο	78

Πηγή: Μελέτη Διερεύνησης Κοινωνικών Αναγκών στους Δήμους του Νομού Θεσσαλονίκης (2003)



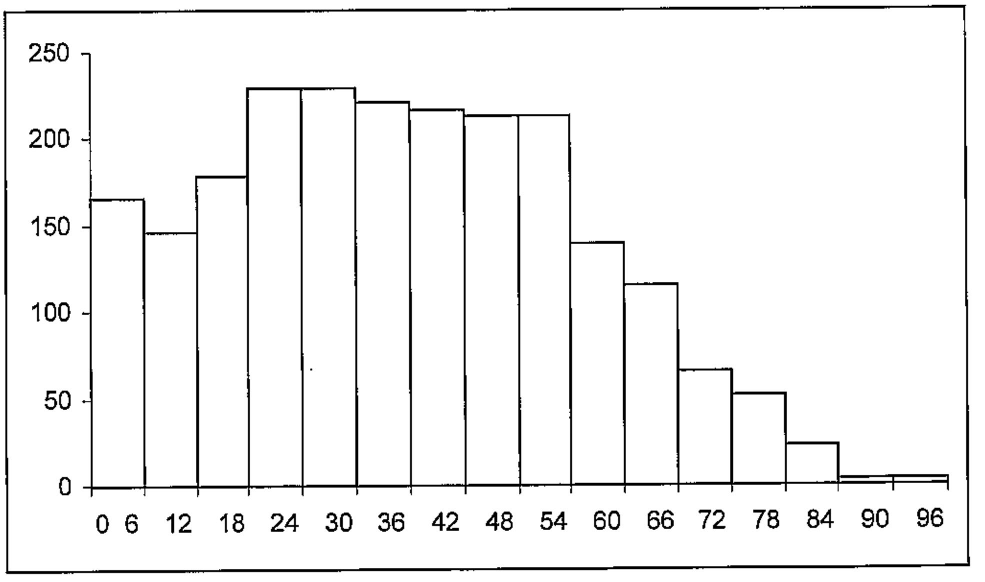
Γράφημα 2.7 Κατανομή συχνοτήτων μηνών ανεργίας

Για τη γραφική απεικόνιση συνεχών ποσοτικών μεταβλητών απαιτείται ο διαχωρισμός των τιμών τους σε κλάσεις. Η απαίτηση του διαχωρισμού σε κλάσεις θεωρείται απαραίτητη ώστε να είναι δυνατή η παρουσίαση της κατανομής τους. Δεν είναι εύκολη η παρουσίαση μιας κατανομής συνεχούς ποσοτικής μεταβλητής όταν έχουμε ένα μεγάλο πλήθος μεμονωμένων τιμών. Η κατανομή μιας συνεχούς ποσοτικής μεταβλητής χωρισμένης σε κλάσεις μπορεί να παρασταθεί γραφικά με τα ιστογράμματα και τα πολύγωνα συχνοτήτων.

Το **ιστόγραμμα** είναι ίσως η πιο σημαντική γραφική παράσταση των δεδομένων καθώς παίζει σημαντικό ρόλο στην στατιστική συμπερασματολογία. Αποτελεί μια ακολουθία κ ορθογώνιων παραλληλογράμμων, όσα και οι τάξεις της κατανομής. Αν οι συχνότητες είναι σχετικές (αθροιστικές), τότε το ιστόγραμμα αναφέρεται ως **ιστόγραμμα σχετικών (αθροιστικών) συχνοτήτων**. Κατά την κατασκευή ενός ιστογράμματος θα πρέπει να προσεχθούν τα παρακάτω:

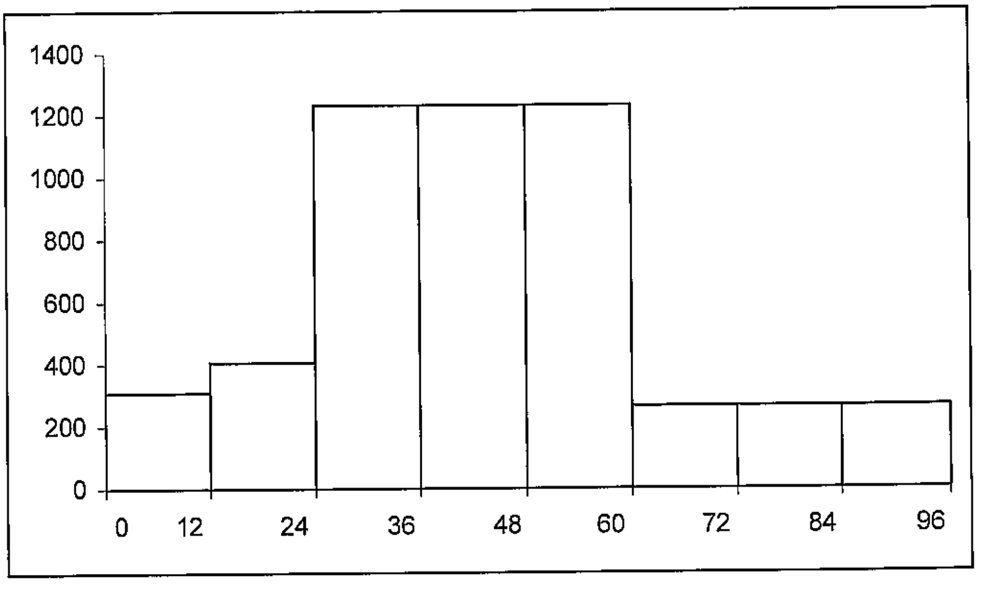
- Η βάση του κάθε ορθογωνίου παραλληλογράμμου αντιστοιχεί στο εύρος της κλάσης στην οποία αντιστοιχεί. Γενικά, στην παρουσίαση της κατανομής συνεχών μεταβλητών, επιδιώκεται όλες οι κλάσεις να έχουν ίσο εύρος, αν αυτό δεν είναι εύκολο ή σκόπιμο, χρειάζεται ιδιαίτερη προσοχή στην κατασκευή του ιστογράμματος.
- Τα εμβαδά των παραλληλογράμμων είναι ανάλογα του περιεχομένου (απόλυτη ή σχετική συχνότητα) των αντίστοιχων κλάσεων.

Αν λοιπόν, όλες οι κλάσεις έχουν ίσο εύρος, η συχνότητα τους μπορεί να απεικονισθεί με το ύψος του παραλληλογράμμου. Το συνολικό εμβαδόν, όλων των παραλληλογράμμων, είναι ανάλογο με το σύνολο των παρατηρήσεων. Συγκεκριμένα αν το εύρος των κλάσεων ισούται με τη μονάδα τότε το συνολικό εμβαδόν του ιστογράμματος απόλυτων (σχετικών) συχνοτήτων ισούται με το σύνολο των παρατηρήσεων n (1). Αν το εύρος των κλάσεων είναι c 1 τότε το συνολικό εμβαδόν του ιστογράμματος απόλυτων (σχετικών) συχνοτήτων ισούται με $c*n$ (c). Στο παρακάτω γράφημα παρουσιάζεται η κατανομή συχνοτήτων ηλικίας του πίνακα 2.8.



Γράφημα 2.8 Κατανομή συχνοτήτων ηλικίας

Πολύ συχνά τα δεδομένα που πρέπει να απεικονισθούν γραφικά δεν είναι ομαδοποιημένα σε κλάσεις ίσων διαστημάτων. Άνισα διαστήματα, συνήθως, εμφανίζονται στις κατανομές εισοδήματος, ηλικιών, μεγέθους επιχειρήσεων, οικονομικής δραστηριότητας κλπ. Το πρόβλημα σε αυτές τις περιπτώσεις συνίσταται στο πώς θα σχεδιαστεί ιστόγραμμα κατανομών με κλάσεις που αντιστοιχούν σε άνισα εύρη. Όταν η κατανομή έχει άνισες τάξεις το ύψος των παραλληλογράμμων θα πρέπει να τροποποιηθεί έτσι ώστε και πάλι τα εμβαδά των επιφανειών που περικλείουν να είναι ανάλογα με τις συχνότητες. Έτσι π.χ. αν μια τάξη είναι διπλάσια από τις υπόλοιπες, τότε το ύψος του παραλληλογράμμου της θα πρέπει να ισούται με το μισό της συχνότητας της. Δηλαδή, αν για την ηλικία είχαμε τον παρακάτω πίνακα συχνοτήτων τότε το ιστόγραμμα στο γράφημα 2.9 θα ήταν λανθασμένο.



Γράφημα 2.9 Λανθασμένο ιστόγραμμα συχνοτήτων ηλικίας

Πίνακας 2.12 Απόλυτες (n) και σχετικές (f) συχνότητες ανά ηλικία

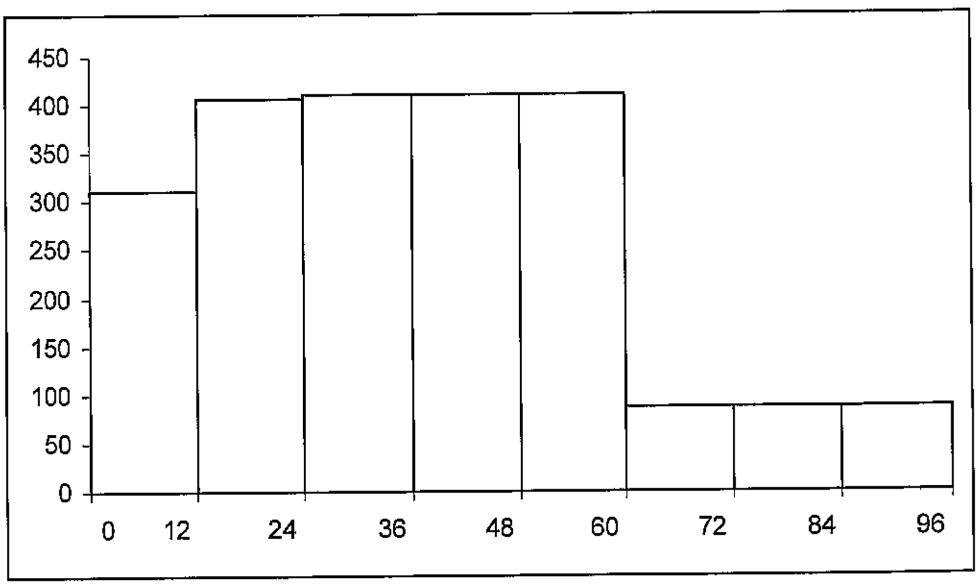
ΗΛΙΚΙΑ	n	f
[0, 12)	312	0,141
[12, 24)	407	0,184
[24 ,60)	1232	0,556
[60, 96)	263	0,119
Σύνολο	2214	1,000

Πηγή: Μελέτη Διερεύνησης Κοινωνικών Αναγκών στους Δήμους του Νομού Θεσσαλονίκης (2003)

Γενικά, για να κατασκευάσουμε το ιστόγραμμα μιας κατανομής με άνισες τάξεις υπολογίζουμε τις διορθωμένες συχνότητες ως εξής:

1. Διαλέγουμε την τάξη της οποίας το εύρος θέλουμε να πάρουμε ως μονάδα.
2. Διαιρούμε το εύρος της κάθε τάξης δια του εύρους-μονάδα.
3. Θέτουμε ως διορθωμένη συχνότητα το πηλίκο της διαίρεσης της αρχικής συχνότητας δια του αντίστοιχου αριθμού που προέκυψε από το βήμα 2.

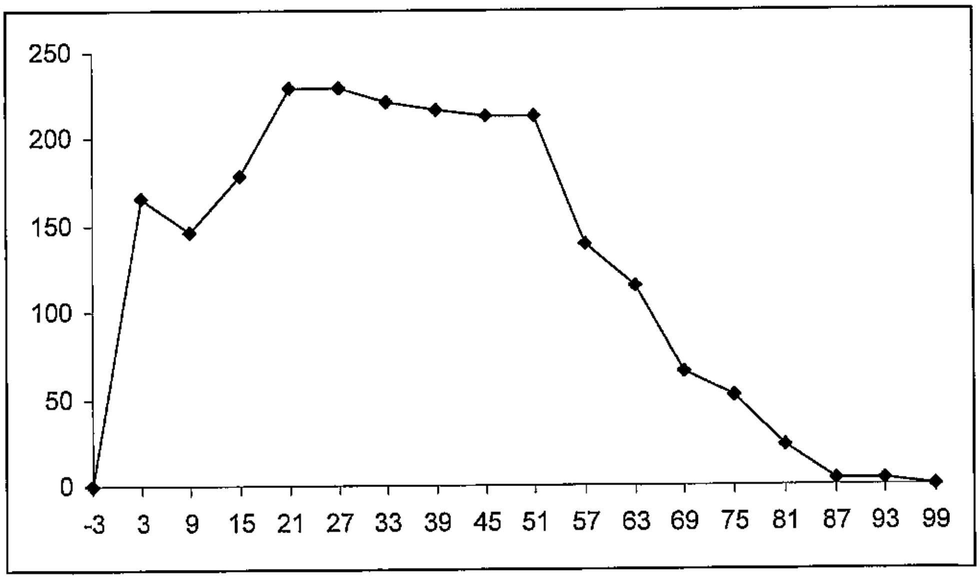
Έτσι το σωστό ιστόγραμμα για τα δεδομένα του πίνακα 2.12 εμφανίζεται στο γράφημα 2.10 στο οποίο ορίστηκε ως μοναδιαίο εύρος το 12 και ως συνέπεια των παραπάνω οι συχνότητες των δύο τελευταίων τάξεων του πίνακα 2.12 είναι 410,67 και 87,67 αντίστοιχα.



Γράφημα 2.10 Σωστό ιστόγραμμα συχνοτήτων ηλικίας

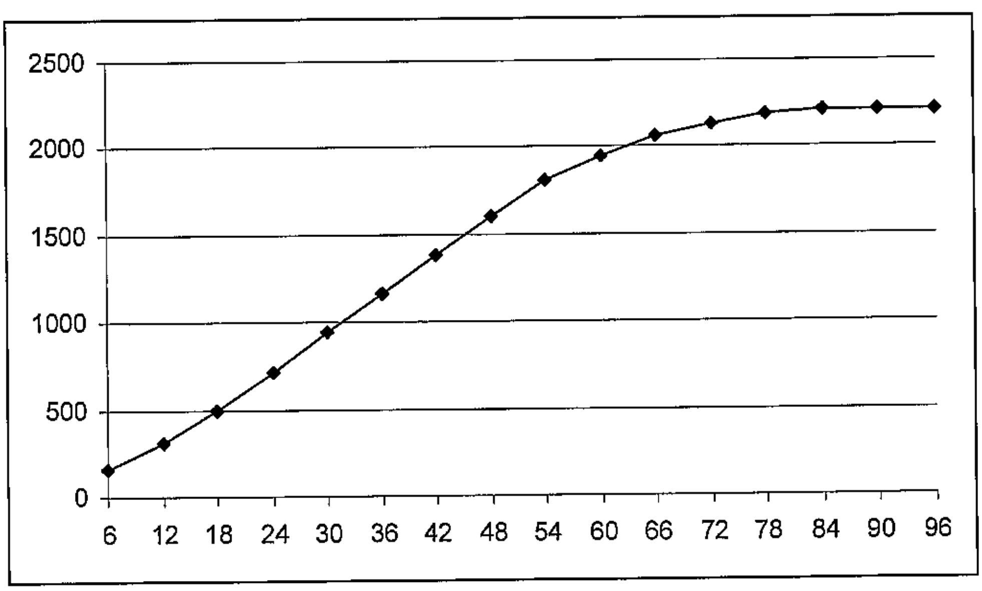
Αν σε ένα σύστημα ορθογωνίων αξόνων, όπου έχουν ορισθεί οι κλίμακες των μεγεθών που μετρούν, προσδιορισθούν τα σημεία με τετμημένες τις κεντρικές τιμές των κλάσεων (μέσα των διαστημάτων που αντιστοιχούν σε κάθε κλάση) και τεταγμένες τις αντίστοιχες συχνότητες, και συνδεθούν διαδοχικά αυτά τα σημεία με ευθύγραμμα τμήματα, δημιουργείται ένα διάγραμμα, τεθλασμένη γραμμή, που είναι γνωστή ως «πολυγωνική γραμμή» συχνοτήτων ή ως **πολύγωνο συχνοτήτων**. Συμβατικά, η πολυγωνική γραμμή επεκτείνεται από τα δύο άκρα στον οριζόντιο άξονα και σε απόσταση μισού ταξικού εύρους. Έτσι, η επιφάνεια που περικλείει, ισούται με την επιφάνεια του ιστογράμματος.

Σε πολυπληθείς πληθυσμούς με μεγάλο αριθμό κλάσεων πολύ μικρού εύρους, ενώνοντας τα μέσα των άνω πλευρών των διαδοχικών κατακόρυφων στηλών, είναι εύκολο αντί για τεθλασμένη να έχουμε καμπύλη γραμμή, που λέγεται **καμπύλη συχνοτήτων**. Στο παρακάτω γράφημα παρουσιάζεται το πολύγωνο συχνοτήτων ηλικίας του πίνακα 2.8.



Γράφημα 2.11 Πολύγωνο συχνοτήτων ηλικίας

Το **πολύγωνο αθροιστικών συχνοτήτων** συνεχούς μεταβλητής προκύπτει από ένα ιστόγραμμα αθροιστικών συχνοτήτων αν συνδέσουμε τα σημεία που αντιστοιχούν στις αθροιστικές συχνότητες των άνω άκρων (και όχι των μέσων) των ταξικών διαστημάτων. Αυτό γιατί όταν στο διάστημα (α, β] αντιστοιχεί η αθροιστική συχνότητα Φ, σημαίνει ότι Φ άτομα έχουν για τη συνεχή μεταβλητή *X* τιμή μικρότερη ή ίση με β και το σημείο Μ(β, Φ) βρίσκεται επί της πολυγώνου της αθροιστικής συχνότητας.



Γράφημα 2.12 Πολύγωνο αθροιστικών συχνοτήτων ηλικίας

Με τον ίδιο τρόπο, είναι δυνατό να ορισθεί η **καμπύλη της αθροιστικής συχνότητας** ως εξής: Καμπύλη αθροιστικής συχνότητας είναι το σύνολο των σημείων καρτεσιανού επιπέδου με τεταγμένες τις παρατηρήσεις (ή το ποσοστό των παρατηρήσεων), που έχουν για την συνεχή μεταβλητή X τιμή μικρότερη ή ίση με την τετμημένη.

Η καμπύλη αθροιστικής συχνότητας συνεχούς μεταβλητής, μετά το πέρας του τελευταίου ταξικού διαστήματος, εκτείνεται παράλληλα προς τον άξονα των τετμημένων, ενώ για τιμές μικρότερες του κάτω ορίου του πρώτου ταξικού διαστήματος συμπίπτει με τον άξονα αυτόν (X).

Είναι σκόπιμο να αναφερθεί ότι οι καμπύλες των αθροιστικών συχνοτήτων είναι δυνατό να κατασκευασθούν, χωρίς να μετατρέπονται σε ίσα μεταξύ τους τα ταξικά διαστήματα των κλάσεων.

Οι καμπύλες αθροιστικών συχνοτήτων που είδαμε μέχρι αυτό το σημείο, απεικονίζουν δεξιόστροφες αθροιστικές σειρές κατανομών και παρέχουν τη δυνατότητα προσδιορισμού του αριθμού των παρατηρήσεων (σε απόλυτη και σχετική συχνότητα) που έχουν τιμή το πολύ ίση με μία οποιαδήποτε συγκεκριμένη τιμή της μεταβλητής. Πολλές φορές, είναι χρήσιμο να προσδιορισθεί ο αριθμός των παρατηρήσεων ή των περιστατικών, που έχουν τιμή τουλάχιστον ίση με μία οποιαδήποτε συγκεκριμένη τιμή της μεταβλητής. Αυτό επιτυγχάνεται υπολογίζοντας αρχικά την αριστερόστροφη αθροιστική σειρά και κατασκευάζοντας στη συνέχεια την καμπύλη της.

Όταν θέλουμε να δώσουμε γραφικά μια συγκριτική εικόνα δύο ή περισσότερων κατανομών συχνοτήτων θα χρησιμοποιήσουμε τις πολυγωνικές γραμμές τους. Δύο ή περισσότερες πολυγωνικές γραμμές μπορούν εύκολα να συγκριθούν αν οι αντίστοιχες κατανομές έχουν τις ίδιες τάξεις και το ίδιο πλήθος παρατηρήσεων. Διαφορετικά, διακρίνουμε τις εξής περιπτώσεις:

- οι κατανομές δεν έχουν το ίδιο πλήθος παρατηρήσεων, οπότε οι συγκρίσεις θα γίνουν από τα πολύγωνα σχετικών συχνοτήτων,
- οι κατανομές δεν έχουν τις ίδιες τάξεις, οπότε οι συγκρίσεις θα γίνουν από τα πολύγωνα των αθροιστικών συχνοτήτων,
- οι κατανομές δεν έχουν ούτε το ίδιο πλήθος παρατηρήσεων, ούτε τις ίδιες τάξεις. Η σύγκριση δύο τέτοιων κατανομών πρέπει να γίνει από τα πολύγωνα των σχετικών αθροιστικών συχνοτήτων.