ΣΧΟΛΗ ΣΔΟ

Τμήμα

Επιχειρηματικού Σχεδιασμού και Πληροφοριακών Συστημάτων

ΠτυχιακήΕργασία

Βασικές Έννοιες της Θεωρίας της Περιγραφικής Στατιστικής και Εφαρμογές με την χρήση του στατιστικού πακέτου SPSS

Ονοματεπώνυμο: **Γεώργιος Καραβάς**

Ανδρέας Καραντώνης

Παναγιώτης Αλεξάκης

Επιβλέπουσα Καθηγήτρια: **Α. Καλαπόδη**

Πάτρα

Ιούνιος 2013

Πίνακας Περιεχομένων

Λίστα Πινάκων

Πίνακας Διαγραμμάτων

Πρόλογος

Σκοπός της παρούσας εργασίας είναι η μελέτη των βασικών εννοιών της θεωρίας της Περιγραφικής Στατιστικής ενώ επιπλέον παρουσιάζονται και κάποιες στατιστικές εφαρμογές με χρήση του στατιστικού πακέτου SPSS.

Η επιστήμη της Στατιστικής σκοπό έχει την ποσοτικοποίηση της αβεβαιότητας ώστε οι επιλογές που θα γίνουν από τους λήπτες αποφάσεων να είναι όσο το δυνατόν απαλλαγμένες από το ενδεχόμενο κινδύνου. Ως εκ τούτου, η στατιστική θεωρία έχει αναπτύξει σύγχρονα εργαλεία προκειμένου να μπορέσει να ερμηνεύσει την πολύπλοκη πραγματικότητα που χαρακτηρίζεται από πολλές και αλληλεξαρτώμενες σχέσεις αιτιότητας.

Στο παρόν πλαίσιο, θα χρησιμοποιήσουμε ένα από αυτά τα λογισμικά και συγκεκριμένα το SPSS18 προκειμένου να περιγράψουμε τις μεταβλητές για τις οποίες έχουμε δεδομένα. Με την βοήθεια του λογισμικού αυτού θα εξάγουμε χρήσιμα συμπεράσματα για την συμπεριφορά των μεταβλητών μας καθώς και για τις σχέσεις που τις χαρακτηρίζουν.

Προκειμένου να παρέχουμε ένα θεωρητικό πλαίσιο ώστε να κατανοήσουμε στην συνέχεια τα αποτελέσματα της στατιστικής ανάλυσης, τα πρώτα τρία κεφάλαια της παρούσας πτυχιακής εργασίας παρέχουν το θεωρητικό πλαίσιο ενώ το τελευταίο περιλαμβάνει τα αποτελέσματα της ανάλυσης. Πιο συγκεκριμένα, η δομή αναπτύσσεται παρακάτω.

Το πρώτο κεφάλαιο αποτελεί μια εισαγωγή στην Περιγραφική Στατιστική. Στην πρώτη υποενότητα αναλύονται οι κατηγορίες της Στατιστικής. Στη συνέχεια παρουσιάζονται εκτενώς τα περιγραφικά Μέτρα Θέσης ή Κεντρικής Τάσης, Διασποράς καθώς και τα Μέτρα Ασυμμετρίας και Κύρτωσης. Το κεφάλαιο ολοκληρώνεται με τα Διαγράμματα τα οποία χρησιμοποιούνται προκείμενου να γίνει η παρουσίαση των στατιστικών δεδομένων (Ιστόγραμμα, Διάγραμμα Πίτας, Διάγραμμα Μίσχου-Φύλλου, κ.α.).

Στο δεύτερο κεφάλαιο γίνεται μια εκτενής αναφορά στις Τυχαίες Μεταβλητές. Μετά την ανάλυση των όρων (τυχαίες μεταβλητές, κατηγορίες μεταβλητών, συνάρτηση πυκνότητας πιθανότητας) σειρά έχει η παρουσίαση των συναρτήσεων κατανομής τόσο των Διακριτών (ομοιόμορφη, διωνυμική, γεωμετρική και κατανομή Poisson) όσο και των Συνεχών (ομοιόμορφη, κανονική, εκθετική και κατανομή Γάμμα).

Το τρίτο κεφάλαιο παρουσιάζει την έννοια της Οικονομετρίας καθώς και τους σκοπούς αυτής. Παρουσιάζεται το Γραμμικό Υπόδειγμα και οι Βασικές τουυποθέσεις καιαναλύεται η Μέθοδος Ελαχίστων Τετραγώνων η οποία αποτελεί το δημοφιλέστερο τρόπο εκτίμησης των παραμέτρων της γραμμικής παλινδρόμησης. Το κεφάλαιο κλείνει με ένα παράδειγμα έτσι ώστε να γίνει αντιληπτό στον αναγνώστη το πώς χρησιμοποιούνται οι οικονομετρικές μέθοδοι ώστε να ποσοτικοποιήσουμε σχέσεις που παρατηρούνται στον πραγματικό κόσμο.

Το τέταρτο κεφάλαιο αποτελεί τον επίλογο της παρούσας εργασίας με στατιστικές εφαρμογές στο SPSS. Συγκεκριμένα, από το Τμήμα Προσωπικού μίας μεγάλης χρηματιστηριακής εταιρείας συλλέξαμε στοιχεία που αφορούν το σύνολο των 190 απασχολουμένων και μέσω του συγκεκριμένου στατιστικού προγράμματος, έγινε ανάλυση των δεδομένων προκειμένου να διαπιστωθεί ποιοι παράγοντες επιδρούν στην διαμόρφωση του μισθού των εργαζομένων.

Τέλος, το πέμπτο κεφάλαιο αφορά στα συμπεράσματα της παρούσας εργασίας.

Κεφάλαιο 1.

Εισαγωγή στην Περιγραφική Στατιστική

1.1 Κατηγορίες Στατιστικής

Προτού προχωρήσουμε σε παρουσίαση και ανάλυση των μέτρων περιγραφικής στατιστικής κρίνεται σκόπιμο να αναφερθούμε στις επιμέρους κατηγορίες της. Η Στατιστική χωρίζεται σε δυο διακριτές μεταξύ τους κατηγορίες οι οποίες είναι η Περιγραφική και η Επαγωγική Στατιστική.

Η Περιγραφική Στατιστική στοχεύει στο να πληροφορήσει τον ερευνητή για τα χαρακτηριστικά του δείγματος που έχει κάθε φορά στην διάθεση του ενώ η Επαγωγική Στατιστική έχει ως αντικείμενο της την εξαγωγή συμπερασμάτων για τον πληθυσμό χρησιμοποιώντας τα χαρακτηριστικά του δείγματος σε συνδυασμό με τον έλεγχο υποθέσεων για διαφορετικές κάθε φορά παραμέτρους.

Στην παρούσα εργασία θα εστιάσουμε την προσοχή μας στην ανάλυση των μέτρων Περιγραφικής Στατιστικής.

Για το κεφάλαιο αυτό χρησιμοποιήσαμε στοιχεία απο τις πηγές :[1], [3], [5], [7].

1.2 Περιγραφικά Μέτρα: *Μέτρα θέσης ή Κεντρικής Τάσης*

Τα περιγραφικά στατιστικά μέτρα χωρίζονται σε τρεις επιμέρους κατηγορίες και κάθε μία περιλαμβάνει διαφορετικά μέτρα. Αυτές οι κατηγορίες είναι οι εξής:

- Τα μέτρα θέσης ή κεντρικής τάσης,
- Τα μέτρα διασποράς και
- Τα μέτρα ασυμμετρίας και κύρτωσης

Έτσι, τα μέτρα κεντρικής τάσης ή θέσης μας πληροφορούν για τον τρόπο με τον οποίο οι παρατηρήσεις του δείγματος μας εκτείνονται γύρω από το κέντρο τους και περιλαμβάνουν τον *μέσο*, την *διάμεσο*, την *επικρατούσα τιμή* και τα *εκατοστημόρια* ενώ τα μέτρα διασποράς μας ενημερώνουν για τον τρόπο με τον οποίο οι παρατηρήσεις μας εκτείνονται γύρω από το κέντρο του δείγματος και αποτελούνται από την *διακύμανση*, την *τυπική απόκλιση* και τον *συντελεστή μεταβλητότητας* και το *ενδο-τεταρτημοριακό εύρος*. Τέλος, τα μέτρα ασυμμετρίας και κύρτωσης περιλαμβάνουν τον συντελεστή ασυμμετρίας και κύρτωσης αντίστοιχα και μας πληροφορούν για την μορφή της κατανομής των παρατηρήσεων (Τσαγκρής, 2008).

Παρακάτω, θα αναλύσουμε τα μέτρα κάθε κατηγορίας ξεχωριστά.

Η κεντρική τάση αναφέρεται στην θέση μιας κατανομής. Θα μετρήσουμε την κεντρική τάση για πληθυσμούς (δηλαδή, μια συλλογή από όλα τα στοιχεία που περιγράφουμε) αλλά και για δείγματα που προέρχονται από πληθυσμούς όπως επίσης και από ομαδοποιημένα ή μη-ομαδοποιημένα δεδομένα.

1.2.1 Ο Αριθμητικός Μέσος ή Μέσος Όρος

Ο αριθμητικός μέσος ή μέσος όρος ενός πληθυσμού συμβολίζεται με μ ενώ ενός δείγματος με το $\overline{X}$.

- Για ομαδοποιημένα δεδομένα, τα μ *και* $\overline{X}$ υπολογίζονται ως εξής:

$$m = \frac{\sum_{i=1}^{N} f_i X_i}{N} \text{ και } \overline{X} = \frac{\sum_{i=1}^{n} f_i X_i}{n} \qquad (1, 2)$$

όπου $\sum f_i X_i$ αναφέρεται στο άθροισμα συχνοτήτων της κάθε κλάσης f επί το ενδιάμεσο σημείο κάθε κλάσης Χ.

- Για μη-ομαδοποιημένα δεδομένα, τα μ *και* $\overline{X}$ υπολογίζονται ως εξής:

$$m = \frac{\sum_{i=1}^{N} X_i}{N} \text{ και } \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad (3, 4)$$

όπου $\sum X_i$ αναφέρεται στο άθροισμα όλων των παρατηρήσεων. Τα N και n αναφέρονται στον αριθμό παρατηρήσεων στον πληθυσμό και στο δείγμα αντίστοιχα.

Ο αριθμητικός μέσος είναι ο πιο δημοφιλής μέσος και η πιο ενδιαφέρουσα στατιστική παράμετρος μιας κατανομής. Το σημείο στο οποίο αντιστοιχεί η τιμή του μέσου όρου πάνω στο στον άξονα των τιμών της μεταβλητής Χ, είναι το σημείο στο οποίο βρίσκεται συγκεντρωμένη όλη η μάζα της μεταβλητής δηλαδή το κέντρο βάρους της κατανομής. Το κύριο μειονέκτημα του μέσου όρου είναι ότι δεδομένου ότι χρησιμοποιεί όλες της παρατηρήσεις, ως εκ τούτου επηρεάζεται από τις πολύ χαμηλές και τις ακραίες τιμές και ενδεχομένως να οδηγήσει σε παραπλανητικά συμπεράσματα σχετικά με την θέση της κατανομής.

Υπάρχουν κάποιες περιπτώσεις όπου όλες οι τιμές της μεταβλητής Χ δεν έχουν την ίδια βαρύτητα. Σε αυτήν την περίπτωση χρησιμοποιείται μια παραλλαγή του αριθμητικού μέσου, ο Σταθμικός Αριθμητικός Μέσος όπου σε κάθε τιμή αντιστοιχεί ένας συντελεστής βαρύτητας ή στάθμισης. Ο Σταθμικός Αριθμητικός Μέσος υπολογίζεται με τον εξής τύπο:

$$\overline{X} = \frac{\sum_{i=1}^{k} w_i X_i}{\sum_{i=1}^{k} w_i} \tag{5}$$

1.2.2 Η Διάμεσος

Η διάμεσος (Μ) είναι εκείνη η τιμή της μεταβλητής Χ, η οποία διαχωρίζει τον πληθυσμό σε δυο μέρη, όπου το 50% των στοιχείων του έχουν τιμή ίση ή μικρότερη από την διάμεσο Μ δηλαδή, βρίσκονται αριστερά αυτής και το υπόλοιπο 50% των στοιχείων του πληθυσμού έχουν τιμή ίση ή μεγαλύτερη από Μ και ως εκ τούτου βρίσκονται δεξιά της Μ.

Θα πρέπει λοιπόν να κατατάξουμε τα δεδομένα σε αύξουσα ή φθίνουσα σειρά για να υπολογίσουμε την τιμή της διαμέσου.

Η διάμεσος για μη-ομαδοποιημένα δεδομένα με περιττό αριθμό στοιχείων είναι η τιμή του μεσαίου στοιχείου, όταν όλα τα στοιχεία έχουν καταταχθεί είτε σε αύξουσα ή φθίνουσα τάξη μεγέθους και δίνεται από τον τύπο:

$$Median = to\left(\frac{N+1}{2}\right) - ios\,tó \text{ στοιχείο στην κατάταξη των δεδομένων} \tag{6}$$

όπου *N* αναφέρεται στον αριθμό των στοιχείων του πληθυσμού (*n*για ένα δείγμα), και στην περίπτωση που ο αριθμός των στοιχείων είναι άρτιος τότε η διάμεσος ορίζεται ως ο μέσος όρος των δυο τιμών της Χ που βρίσκονται στις θέσεις $N/2\; kai\; N/2 + 1$.

Ενώ για ομαδοποιημένα δεδομένα, η διάμεσος δίνεται από τον τύπο:

$$Median = L + \frac{n/2 - F}{f_m} c \tag{7}$$

όπου:

L= το χαμηλότερο όριο της ενδιάμεσης τάξης (δηλαδή, η κλάση οποία περιλαμβάνει το διάμεσο αντικείμενο της κατανομής)

n= ο αριθμός των παρατηρήσεων στο σύνολο δεδομένων που έχουμε στην διάθεσή μας

F= άθροισμα των συχνοτήτων μέχρι και την διάμεση τάξη, χωρίς όμως να την περιλαμβάνει

f_m = συχνότητα της διάμεσης τάξης

c= πλάτος της κλάσης

1.2.3 Η Επικρατούσα Τιμή

Η Επικρατούσα τιμή ή τύπος, **T** είναι η τιμή με την μεγαλύτερη συχνότητα μέσα στο σύνολο δεδομένων και για ομαδοποιημένα δεδομένα δίνεται από τον τύπο:

$$Mode = L + \frac{d_1}{d_1 + d_2} c \qquad (8)$$

όπου:

L= χαμηλότερο όριο της κλάσης με την μεγαλύτερη συχνότητα

d_1 =συχνότητα της κλάσης που περιλαμβάνει την επικρατούσα τιμή μείον την συχνότητα της προηγούμενης κλάσης

d_2 = συχνότητα της κλάσης που περιλαμβάνει την επικρατούσα τιμή κλάσης μείον την συχνότητα της επόμενης κλάσης

c= πλάτος της κλάσης

Η επικρατούσα τιμή αποτελεί μέτρο κεντρικής τάσης της κατανομής διότι δείχνει την θέση όπου υπάρχει η μεγαλύτερη συγκέντρωση των τιμών της μεταβλητής Χ. Η επικρατούσα τιμή μπορεί να μην είναι μοναδική αλλά στην περίπτωση που κάτι τέτοιο συμβαίνει τότε η κατανομή λέγεται μονοκόρυφη.

Από την άλλη μεριά όταν η κατανομή είναι πολυκόρυφη, τότε υπάρχει μια απόλυτη επικρατέστερη τιμή και άλλες επιμέρους ή σχετικές επικρατέστερες τιμές.

1.2.4 Τα Τεταρτημόρια

Τα τεταρτημόρια αποτελούν γενίκευση της έννοιας της διάμεσης τιμής και χρησιμοποιούν την ίδια λογική για τον χωρισμό του πληθυσμού σε υπό-πληθυσμούς όπως η διάμεση τιμή.

Πιο συγκεκριμένα:

- $D_1, D_2, \mathbf{KK}, D_9$ που χωρίζουν τον πληθυσμό σε 10 ίσους πληθυσμούς, (10% ο καθένας), ***δεκατημόρια***

- $C_1, C_2, \ldots, C_{99}$ που χωρίζουν τον πληθυσμό σε 100 ίσους υπό-πληθυσμούς, (1% ο καθένας), ***εκατοστημόρια***
- Q_1, Q_2, Q_3 που χωρίζουν τον πληθυσμό σε 4 ίσους υπό-πληθυσμούς, (25% ο καθένας), ***τεταρτημόρια***

Το Q_1 είναι εκείνη η τιμή της μεταβλητής X, στα αριστερά της οποίας βρίσκονται το 25% των παρατηρήσεων ενώ το Q_3 είναι η τιμή της X, στα αριστερά της οποίας βρίσκονται το 75% των παρατηρήσεων και στα δεξιά της το 25%. Είναι φανερό ότι $Q_2 = M$.

Οι μαθηματικές εκφράσεις των Q_1 και Q_3 δίνονται παρακάτω:

$$Q_1 = e_{i-1} + \frac{e_i - e_{i-1}}{f_i}\left(\frac{N}{4} - \Phi_{i-1}\right) \tag{9}$$

όπου αν ο αριθμός N/4 περιλαμβάνεται μεταξύ των τιμών Φ_{i-1} και Φ_i της αθροιστικής συχνότητας και η Φ_i αντιστοιχεί στο άνω πέρας e_i του διαστήματος $\lfloor e_{i-1,} e_i \rfloor$, εντός του οποίου υπάρχουν f_i παρατηρήσεις (f_i συχνότητα) για το πρώτο τεταρτημόριο ενώ

$$Q_3 = e_{j-1} + \frac{e_j - e_{j-1}}{f_i}\left(3\frac{N}{4} - \Phi_{j-1}\right) \tag{10}$$

όπου αν ο αριθμός 3N/4 περιλαμβάνεται μεταξύ των τιμών Φ_j και Φ_{j-1} της αθροιστικής συχνότητας και η Φ_j αντιστοιχεί στο άνω πέρας e_j του διαστήματος $\lfloor e_{j-1,} e_j \rfloor$, εντός του οποίου υπάρχουν f_j παρατηρήσεις (f_i συχνότητα) για το τρίτο τεταρτημόριο.

1.2.5 Επιπλέον Περιγραφικά Μέτρα

Για λόγους πληρότητας των περιγραφικών μέτρων, θα αναφέρουμε συνοπτικά και τον *Γεωμετρικό (απλό και σταθμισμένο)*, τον *Αρμονικό*, τον *Τετραγωνικό (απλός και σταθμισμένο)* καθώς και τον *Γενικευμένο Μέσο*.

Α) Ο Απλός Γεωμετρικός Μέσος

Ο γεωμετρικός μέσος συμβολίζεται με G και ορίζει το σημείο στο οποίο αντιστοιχεί η τιμή της μεταβλητής στην οποία τείνουν να συγκεντρωθούν όλες οι επιμέρους τιμές της.

Ο απλός γεωμετρικός μέσος υπολογίζεται από τους εξής τύπους:

$$G = \sqrt[n]{X_1 \cdot X_2 \cdot \cdots X_n} \text{ ή } G = \sqrt[n]{\prod_{i=1} X_i} \qquad (11)$$

Συχνά ο υπολογισμός της τετραγωνικής ρίζας δεν είναι απλός, μπορούμε να λογαριθμήσουμε τον γεωμετρικό μέσο και να υπολογίσουμε τον αριθμητικό μέσο των λογαρίθμων των αριθμών αυτών:

$$\begin{aligned} & G = \sqrt[n]{X_1 \cdot X_2 \cdot \cdots X_n} \Rightarrow \\ & \log G = \frac{1}{n}\log\left(X_1 \cdot X_2 \cdot \cdots X_n\right) = \frac{1}{n}(\log(X_1) + \log(X_2) + \cdots + \log(X_n)) \Rightarrow \qquad (12) \\ & \Rightarrow \log G = \frac{1}{n}\sum_{i=1}^{n} \log X_i \end{aligned}$$

B) Ο Σταθμικός Γεωμετρικός Μέσος

Αν τα δεδομένα δίνονται με την μορφή της κατανομής συχνοτήτων, δηλαδή αν $f_1, f_2, \cdots \cdots f_k$ είναι οι συχνότητες (για τις οποίες ισχύει $\sum_{i=1}^{k} f_i = n$), ο σταθμικός γεωμετρικός μέσος υπολογίζεται από τον εξής τύπο:

$$G = \sqrt[n]{X_1^{f1} \cdot X_2^{f2} \cdot \cdots X_k^{fk}} \text{ ή } G = \left(X_1^{f1} \cdot X_2^{f2} \cdots \cdots X_k^{fk}\right)^{\frac{1}{n}} = X_1^{\frac{f1}{n}} \cdot X_2^{\frac{f2}{n}} \cdots \cdots X_k^{\frac{fk}{n}} \qquad (13)$$

και αν λογαριθμήσουμε την σχέση παίρνουμε:

$$\log G = \frac{1}{n}\left(f_1 \log X_1 + f_2 \log X_2 + \cdots \cdots + f_k \log X_k\right) = \frac{\sum_{i=1}^{k} f_i \log\left(X_i\right)}{n} \qquad (14)$$

Πρέπει να αναφέρουμε πως ο γεωμετρικός μέσος δεν μπορεί να εφαρμοστεί όταν η μεταβλητή έχει αρνητικές τιμές, είναι όμως η κατάλληλη έννοια μέσου για δεδομένα που αυξάνονται ή ελαττώνονται με γεωμετρική πρόοδο.

Γ) Ο Αρμονικός Μέσος

Ο Αρμονικός μέσος χρησιμοποιείται κυρίως για ρυθμούς και ορίζεται ως ο αντίστροφος του αριθμητικού μέσου των αντιστρόφωντων n τιμών $X_1, X_2, \ldots, X_n$ της μεταβλητής X και δίνεται από τον τύπο:

$$H = \frac{1}{\frac{1}{n}\left(\frac{1}{X_1} + \frac{1}{X2} + \ldots + \frac{1}{X_n}\right)} = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}} \qquad (15)$$

Ο Αρμονικός μέσος δεν έχει μεγάλες εφαρμογές πέρα από κάποιες κατηγορίες δεδομένων. Είναι όμως ο καταλληλότερος για τον προσδιορισμό της κεντρικής τάσης των τιμών μιας μεταβλητής που αντιστοιχεί σε σταθερή ποσότητα μιας άλλης. Επίσης, ο υπολογισμός του δεν είναι δυνατός αν οι τιμές μιας μεταβλητής περιλαμβάνουν το μηδέν.

Δ) Ο Σταθμισμένος Αρμονικός Μέσος

Αν οι τιμές $X_1, X_2, \ldots, X_k$ της μεταβλητής X εμφανίζονται στον πληθυσμό n με συχνότητες $n_i : i = 1, \ldots, k$ αντίστοιχα, σταθμισμένος αρμονικός μέσος της X δίνεται από τον τύπο:

$$H = \frac{1}{\frac{1}{n}\left(\frac{n_1}{X_1} + \frac{n_2}{X_2} + \ldots + \frac{n_k}{X_k}\right)} = \frac{n}{\sum_{i=1}^{k} \frac{n_i}{X_i}} \qquad (16)$$

όπου $\sum_{i=1}^{k} n_i = n$

Ε) Ο Τετραγωνικός Μέσος

Ο τετραγωνικός μέσος**Q**των τιμών $X_1, X_2, \ldots, X_n$ της μεταβλητής X, χρησιμοποιείται κυρίως για τον υπολογισμό των δεικτών διασποράς και δίνεται από τον τύπο :

$$Q^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \Rightarrow Q = \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2} \qquad (17)$$

ΣΤ) Ο Σταθμισμένος Τετραγωνικός Μέσος

Όταν οι τιμές $X_1, X_2, \ldots, X_k$ της μεταβλητής Χ εμφανίζονται με συχνότητες αντίστοιχα $n_1, n_2, \ldots, n_k$ ο σταθμισμένος τετραγωνικός μέσος δίνεται από την σχέση:

$$Q = \sqrt{\frac{1}{n}\sum_{i=1}^{k} n_i X_i^2} \qquad (18)$$

όπου $\sum_{i=1}^{k} n_i = n$

Ζ) Γενικευμένος Μέσος ή Μέσος R τάξεως

Ο γενικευμένος μέσος είναι μια γενίκευση του αριθμητικού μέσου και δίνεται από την σχέση:

$$M_r = \left[\frac{1}{n}\sum_{i=1}^{n} X_i^r\right]^{\frac{1}{r}} \text{ ή } M_r = \sqrt[r]{\frac{1}{n}\sum_{i=1}^{n} X_i^r} \qquad (19)$$

Η) Ο Σταθμισμένος Γενικευμένος R Μέσος τάξεως

Ο Σταθμισμένος Γενικευμένος R Μέσος τάξεωςδίνεται από την σχέση:

$$M_r = \left[\frac{1}{n}\sum_{i=1}^{n} n_i X_i^r\right]^{\frac{1}{r}} \text{ ή } M_r = \sqrt[r]{\frac{1}{n}\sum_{i=1}^{n} n_i X_i^r} \qquad (20)$$

όπου $\sum_{i=1}^{k} n_i = n$

Διαπιστώνεται ότι για διάφορες τιμές του rπαίρνουμε γνωστούς μέσους :

Για r=1, παίρνουμε τον αριθμητικό μέσο,

Για r=2, παίρνουμε τον τετραγωνικό μέσο,

Για r=-1, παίρνουμε τον αρμονικό μέσο,

Για r=ε, παίρνουμε τον γεωμετρικό μέσο, όπου ε= πολύ μικρός θετικός αριθμός

Συνοπτική Σύγκριση των Περιγραφικών Μέτρων

Σύμφωνα με τα παραπάνω, μπορούμε να εντοπίσουμε τις διαφορές μεταξύ των δυο μέσων:

- Ο γεωμετρικός μέσος επηρεάζεται λιγότερο ή καθόλου από πολύ χαμηλές ή ακραίες τιμές σχετικά με τον αριθμητικό μέσο.
- Ο γεωμετρικός δεν έχει την ευρύτητα των εφαρμογών του αριθμητικού και χρησιμοποιείται κυρίως στην κατάρτιση των αριθμοδεικτών και στον υπολογισμό του μέσου ποσοστού μεταβολής ή λόγων.
- Ο αριθμητικός μέσος είναι πιο ενδιαφέρον μέτρο σε σχέση με τον γεωμετρικό διότι υπολογίζεται και ερμηνεύεται εύκολα, μπορεί να υπολογιστεί ακόμα και στην περίπτωση που μια μεταβλητή παίρνει μηδενική ή ακόμα και αρνητική τιμή και δεν μηδενίζεται όταν μια τιμή είναι μηδέν.
- Ο αριθμητικός μέσος προσφέρεται εύκολα σε παραπέρα υπολογισμούς.
- Όλες οι παρατηρήσεις συμβάλλουν στην διαμόρφωση του αριθμητικού μέσου ενώ στην διάμεσο όχι, όπου το πόσο διαφέρει μια παρατήρηση από την διάμεσο δεν έχει σημασία.
- Η τιμή του μέσου όρου είναι μοναδική ενώ η επικρατούσα τιμή μπορεί να μην είναι και να μην είναι.
- Για ποιοτικά δεδομένα (π.χ. το χρώμα ματιών κάποιου), ο υπολογισμός του μέσου όρου δεν έχει έννοια. Σε αυτή την περίπτωση, καταλληλότερο μέτρο είναι η επικρατούσα τιμή.
- Ισχύει ότι $H < G \leq \overline{X} < Q$.
- Η διάμεση τιμή σε αντίθεση με τον αριθμητικό μέσο, δεν επηρεάζεται από τις πολύ μικρές ή τις πολύ μεγάλες τιμές των παρατηρήσεων διότι για τον υπολογισμό της λαμβάνεται υπόψη κυρίως η διάταξη των δεδομένων και όχι το πραγματικό τους μέγεθος.
- Η διάμεση τιμή δεν προσφέρεται για περαιτέρω αλγεβρικούς χειρισμούς σε σχέση με τα υπόλοιπα περιγραφικά μέτρα.

1.3 Περιγραφικά Μέτρα: Μέτρα Διασποράς

Ο αριθμητικός μέσος και οι άλλες παράμετροι θέσης όπως είδαμε παραπάνω προσδιορίζουν την θέση της κατανομής πάνω στον οριζόντιο άξονα αλλά δεν μας λένε τίποτα για το πώς είναι σκορπισμένα γύρω από το κέντρο τους. Από την άλλη μεριά όμως, η διασπορά των δεδομένων αναφέρεται στην μεταβλητότητα των δεδομένων, δηλαδή το πώς αυτά εκτείνονται γύρω από τον μέσο όρο τους.

Τα πιο σημαντικά μέτρα μεταβλητότητας είναι: διακύμανση, η τυπική απόκλιση, ο συντελεστής μεταβλητότητας και το ενδοτεταρτημοριακό εύρος.

1.3.1 Η Διακύμανση

Η μέση τετραγωνική απόκλιση από την μέση τιμή $\overline{X}$ ονομάζεται διακύμανση ή διασπορά και συμβολίζεται s^2 και για τον υπολογισμό της χρησιμοποιούνται όλα τα στοιχεία. Η διακύμανση του πληθυσμού συμβολίζεται με s^2 και η δειγματική διακύμανση s^2.

- Για μη-ομαδοποιημένα δεδομένα δίνονται από τους ακόλουθους τύπους:

$$s^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - m)^2 \text{ η διακύμανση του πληθυσμού} \quad (21)$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \text{ η διακύμανση του δείγματος} \quad (22)$$

όπου N είναι το πλήθος των στοιχείων του πληθυσμού, μ ο αριθμητικός μέσος του πληθυσμού και X_iείναι η τιμή της μεταβλητής X ενώ για την δεύτερη περίπτωση, nείναι το μέγεθος του δείγματος και $\overline{X}$ είναι ο μέσος όρος των στοιχείων του δείγματος.

- Για ομαδοποιημένα δεδομένα (με μορφή κατανομής συχνοτήτων) δίνονται από τους ακόλουθους τύπους:

$$s^2 = \frac{1}{N}\sum_{i=1}^{k} f_i (X_i - m)^2 \text{ η διακύμανση του πληθυσμού} \quad (23)$$

όπου $\sum_{i=1}^{k} f_i = N$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k} f_i\left(X_i - \overline{X}\right)^2$$ η διακύμανση του δείγματος (24)

όπου $\sum_{i=1}^{k} f_i = n$

Οι παραπάνω τύποι συχνά εκφράζονται και ως εξής:

$$s^{\,2} = \frac{1}{\mathrm{N}}\sum_{i=1}^{k} f_i X_i^2 - m^2$$ για τον πληθυσμό (23.1)

όπου $\sum_{i=1}^{k} f_i X_i = Nm$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k} f_i X_i^2 - n\overline{X}^2$$ για το δείγμα (24.1)

όπου $\sum_{i=1}^{k} f_i = n$

Αξίζει να σημειώσουμε ότι η διακύμανση δεν μετριέται στις ίδιες μονάδες που μετριέται η X, αλλά στα τετράγωνα των μονάδων αυτών κάτι που δυσκολεύει την ερμηνεία της.

1.3.2 Η Τυπική Απόκλιση

Το πρόβλημα ερμηνείας της διακύμανσης παύει να υπάρχει με την χρήση της τετραγωνικής της ρίζας κάτι που μας δίνει την τυπική απόκλιση η οποία συμβολίζεται με σ και αποτελεί το κυριότερο μέτρο διασποράς και μετριέται στις ίδιες μονάδες με αυτές της μεταβλητής X.

- Για μη-ομαδοποιημένα δεδομένα η τυπική απόκλιση δίνεται από τους ακόλουθους τύπους:

$$s = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X_i - m\right)^2}$$ η τυπική απόκλιση του πληθυσμού (25)

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$ η τυπική απόκλιση του δείγματος (26)

όπου *N* είναι το πλήθος των στοιχείων του πληθυσμού, *μ* ο αριθμητικός μέσος του πληθυσμού και X_iείναι η τιμή της μεταβλητής X ενώ για την δεύτερη περίπτωση, *n*είναι το μέγεθος του δείγματος και $\overline{X}$ είναι ο μέσος όρος των στοιχείων του δείγματος.

- Για ομαδοποιημένα δεδομένα (με μορφή κατανομής συχνοτήτων) δίνεται από τους ακόλουθους τύπους:

$$s = \sqrt{\frac{1}{N}\sum_{i=1}^{k} f_i\left(X_i - m\right)^2} \text{ η τυπική απόκλιση του πληθυσμού} \qquad (27)$$

όπου $\sum_{i=1}^{k} f_i = N$

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{k} f_i\left(X_i - \overline{X}\right)^2} \text{ η τυπική απόκλιση του δείγματος} \qquad (28)$$

όπου $\sum_{i=1}^{k} f_i = n$

Η ερμηνεία της τυπικής απόκλισης είναι δύσκολη ή και αδύνατη με απλά λόγια διότι η έννοια της είναι αφηρημένη. Ωστόσο, είναι χρήσιμη όταν χρειάζεται να συγκρίνουμε την διασπορά δύο μεταβλητών που μετριούνται στις ίδιες μονάδες και του έχουν τον ίδιο μέσο. Είναι επίσης χρήσιμη όταν θέλουμε να εκτιμήσουμε το ποσοστό των παρατηρήσεων που βρίσκονται σε διάφορα διαστήματα τιμών της μεταβλητής και γνωρίζουμε την πιθανότητα οι τιμές που παίρνει να είναι σε απόσταση μιας, δύο ή τριών τυπικών αποκλίσεων.

1.3.3 Ο Συντελεστής Μεταβλητότητας

Στην περίπτωση που οι μεταβλητές είναι ανομοιογενείς, δεν έχουν μετρηθεί στις ίδιες μονάδες μέτρησης και δεν έχουν τον ίδιο μέσο. Για να συγκριθούν οι διασπορές τους, θα πρέπει να ικανοποιούνται οι προαναφερθείσες συνθήκες.

Αν κάποια από αυτές δεν ικανοποιείται, τότε χρησιμοποιούμε μέτρα σχετικής διασποράς, όπως είναι ο συντελεστής μεταβλητότητας, CV (coefficientofvariation).

Ο συντελεστής μεταβλητότητας δίνεται από τους ακόλουθους τύπους:

$$CV = \frac{s}{m} \text{ για τον πληθυσμό} \qquad (29)$$

όπου σ είναι η τυπική απόκλιση του πληθυσμού και μ ο αριθμητικός του μέσος

$$CV = \frac{s}{\overline{X}} \text{ για το δείγμα} \qquad (30)$$

όπου s είναι η τυπική απόκλιση του δείγματος και $\overline{X}$ ο αριθμητικός του μέσος

Πρέπει να σημειώσουμε ότι η ποσότητα CV είναι απαλλαγμένος από μονάδες μέτρησης, δηλαδή είναι καθαρός αριθμός και προφανώς ο υπολογισμός του είναι τετριμμένος στην περίπτωση όπου μή $\overline{X}$ είναι μηδέν.

1.3.4 Εύρος και ημι-ενδοτεταρτημοριακό εύρος

Το Εύρος, d, μιαw μεταβλητήςX είναι η διαφορά της μικρότερης από την μεγαλύτερη τιμή της. Το βασικό μειονέκτημα είναι ότι βασίζεται μόνο στις δυο ακραίες τιμές αγνοώντας όλες τις ενδιάμεσες. Αν έχουμε δυο κατανομές που έχουν το ίδιο εύρος, αλλά στην πρώτη κατανομή οι περισσότερες παρατηρήσεις είναι πιο συγκεντρωμένες γύρω από τον μέσο σε σύγκριση με την δεύτερη, για να μετρήσουμε την διασπορά θα πρέπει να χρησιμοποιήσουμε ένα μέτρο που να δείχνει μικρότερη διασπορά για την πρώτη σχετικά με την δεύτερη.

Ένα άλλο μέτρο διασποράς είναι το ημι-ενδοτεταρτημοριακό εύρος και ορίζεται ως εξής:

$$d_Q = \frac{Q_3 - Q_1}{2} \tag{31}$$

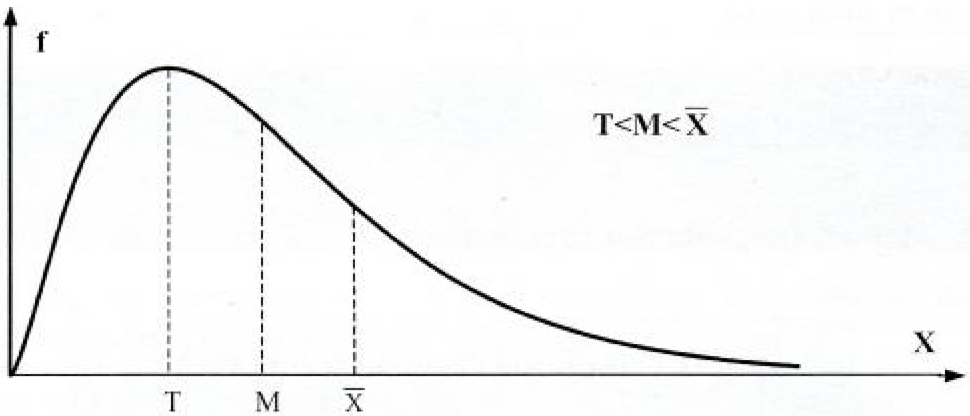
Είναι προφανές ότι πάσχει από το ίδιο πρόβλημα με το εύρος, διότι και αυτό το μέτρο λαμβάνει υπόψη του μόνο ακραίες τιμές.

1.4 Μέτρα Ασυμμετρίας και Κύρτωσης

1.4.1 Μέτρα Ασυμμετρίας

Σε αυτή την ενότητα θα εξετάσουμε την ασυμμετρία μιας κατανομής αλλά θα χρησιμοποιήσουμε το μέτρο ασυμμετρίας που χρησιμοποιείται πιο συχνά, τον συντελεστή ασυμμετρίας $α_3$, ο οποίος ορίζεται ως εξής:

$$a_3 = \frac{m_3}{s^3} \tag{32}$$

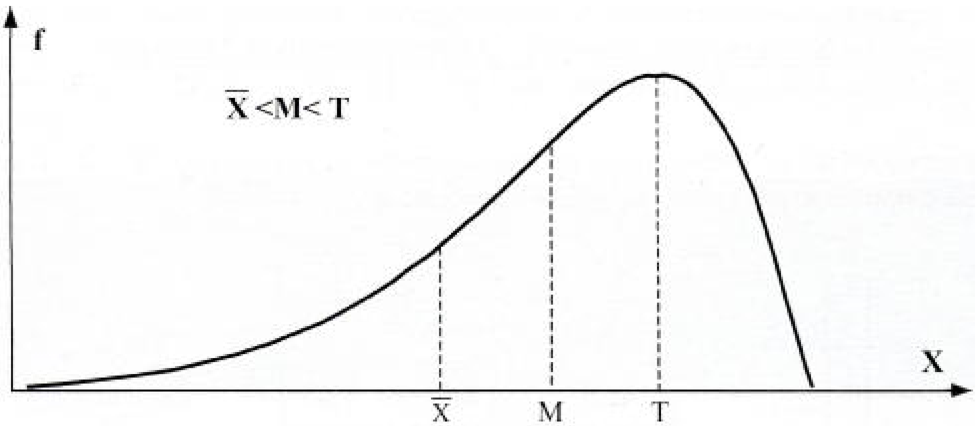


Διάγραμμα 1: Θετικά Ασύμμετρη Κατανομή (Παπαδημητρίου, 2005, σελίδα 339)

Για μια θετικά ασύμμετρη κατανομή έχουμε ότι η μάζα της πιθανότητας είναι μαζεμένη στην αριστερή πλευρά.

Επίσης ισχύει ότι επικρατούσα τιμή < διάμεσο < αριθμητικό μέσο καθώς επίσης και ο συντελεστής ασυμμετρίας $α_3>0$, όπου $α_3$ είναι η τρίτη ροπή της μεταβλητής.

Αντίστοιχα, για μια αρνητικά ασύμμετρη κατανομή έχουμε ότι αριθμητικός μέσος< επικρατούσα τιμή < διάμεσο καθώς και ότι ο συντελεστής ασυμμετρίας είναι $α_3<0$.



Διάγραμμα 2: Αρνητικά Ασύμμετρη Κατανομή (Παπαδημητρίου 2005, σελίδα 339)

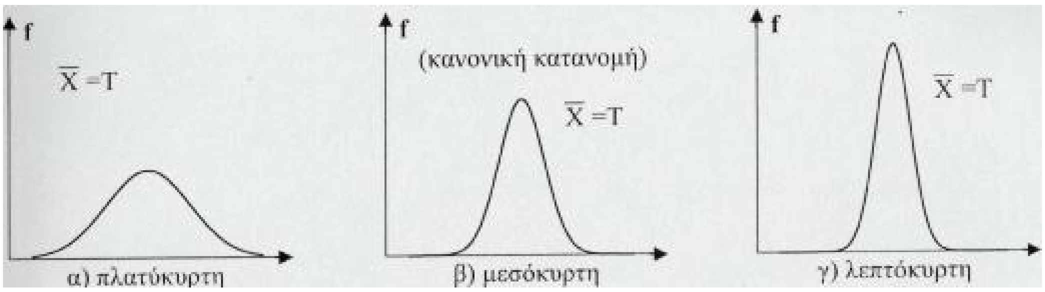
Σε αυτό το σημείο πρέπει να πούμε πως η γνώση της τιμής του $α_3$δεν αρκεί για να συμπεράνουμε αν μια κατανομή είναι συμμετρική ή όχι (προφανώς στην περίπτωση της συμμετρικής κατανομής $α_3$=0 και $\overline{X} = M = T$) δηλαδή η γνώση του $α_3$αποτελεί μια αναγκαία αλλά όχι ικανή συνθήκη για να είναι μια κατανομή συμμετρική. Αυτό είναι το βασικό μειονέκτημα του συντελεστή $α_3$ως μέτρο ασυμμετρίας.

1.4.2 Μέτρα Κύρτωσης

Οι συμμετρικές κατανομές διαφέρουν μεταξύ τους ως προς την κυρτότητα ή την αιχμηρότητα τους στην περιοχή τα επικρατέστερης τιμής τους.

Για τον προσδιορισμό της κυρτότητας μια κατανομής συχνοτήτων, η τελευταία συγκρίνεται με την κατανομή Gaussή κανονική κατανομή που θεωρείται μεσόκυρτη (περιοχή β του Διαγράμματος 3).

Στην περίπτωση που η κατανομή παρουσιάζει αιχμηρότητα μεγαλύτερη από την μεσόκυρτη, δηλαδή η κορυφή τους βρίσκεται υψηλότερα από την αντίστοιχη της μεσόκυρτης ονομάζονται λεπτόκυρτες (περιοχή γ του Διαγράμματος 3) ενώ στην αντίθετη περίπτωση πλατύκυρτες (περιοχή α του Διαγράμματος 3).



Διάγραμμα 3: Κατανομές διαφορετικής κυρτότητας (Παπαδημητρίου 2005, σελίδα 345)

Επίσης για την μέτρηση της κυρτότητας χρησιμοποιούμε τον συντελεστή κύρτωσης $α_4$ και δίνεται από τον εξής τύπο:

$$a_4 = \frac{m_4}{s^4} \tag{33}$$

Για την μεσόκυρτη κατανομή ισχύει $α_4$=3 ($α_4$ αντίστοιχα είναι η τέταρτη ροπή), για την λεπτόκυρτη $α_4$>3 ενώ $α_4$<3 για την πλατύκυρτη. Επίσης παρότι ο συντελεστής $α_4$ χρησιμοποιείται συχνά ως δείκτης κύρτωσης, δεν είναι αρκετός για να συμπεράνουμε σε ποια από τις παραπάνω κατηγορίες ανήκει η κατανομή.

1.5 Διαγραμματική Παρουσίαση των Δεδομένων

1.5.1 Το Ιστόγραμμα

Στην τελευταία αυτή ενότητα του πρώτου κεφαλαίου, κρίνεται σκόπιμο για λόγους ολοκληρωμένης παρουσίασης των εννοιών που μέχρι τώρα συζητήθηκαν να αναφερθούμε και στην διαγραμματική παρουσίαση των στατιστικών δεδομένων.

Για να αποφευχθούν μακροσκελείς στατιστικές παρουσιάσεις αναφερόμενες σε έννοιες όπως συχνότητα, σχετική συχνότητα, αθροιστική συχνότητα κλπ. θεωρήθηκε πιο χρήσιμο να αναφερθούμε σε αυτές τις έννοιες μέσω παραδειγμάτων. Κατ' αυτόν τον τρόπο, θα αναφερθούμε στις έννοιες αυτές και ταυτόχρονα θα παρουσιάσουμε και τα πιο δημοφιλή διαγράμματα στατιστικών αναλύσεων.

Τα παρακάτω διαγράμματα και περιγραφικά μέτρα υπολογίστηκαν χρησιμοποιώντας το αρχείο του SPSS 16, Car_sales.sav

Ο παρακάτω πίνακας παρουσιάζει τα περιγραφικά μέτρα για την μεταβλητή Horsepower, όπως αυτά συζητήθηκαν παραπάνω.

Περιγραφικά Μέτρα της μεταβλητής Horsepower	
Έγκυρες τιμές	156
Τιμές που λείπουν	1
Μέσος	185.95
Διάμεσος	177.5
Επικρατούσα τιμή	150 *
Τυπική Απόκλιση	56.7
Διακύμανση	3214.926
Ασυμμετρία	1.001
Κύρτωση	2,407
Εύρος	395

*Υπάρχουν πολλαπλές Επικρατούσες Τιμές. Εδώ, φαίνεται η μικρότερη τιμή

Πίνακας 1: Περιγραφικά Μέτρα για της μεταβλητής Horsepower

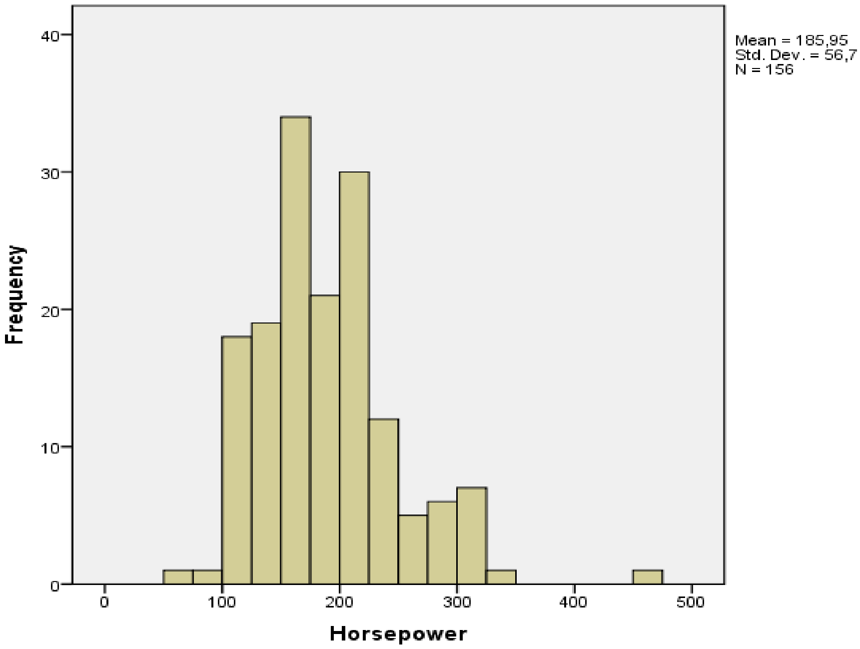
Παρακάτω παρουσιάζεται ο πίνακας συχνοτήτων για την μεταβλητή Horsepower. Η Συχνότητα δείχνει πόσες φορές παρουσιάζεται η τιμή της μεταβλητής στο δείγμα και το Ποσοστό δείχνει σε τι ποσοστό του δείγματος αντιστοιχεί η τιμή αυτή ενώ η Αθροιστική Συχνότητα προκύπτει αν προσθέσουμε το ποσοστό της κλάσης που βρισκόμαστε κάθε φορά και της επόμενης.

Έγκυρες Τιμές	Συχνότητα	Ποσοστό	Έγκυρο Ποσοστό	Αθροιστικό Ποσοστό
55	1	,6	,6	,6
92	1	,6	,6	1,3
100	2	1,3	1,3	2,6
106	1	,6	,6	3,2
107	1	,6	,6	3,8
110	1	,6	,6	4,5
113	1	,6	,6	5,1
115	6	3,8	3,8	9,0
119	1	,6	,6	9,6
120	4	2,5	2,6	12,2
124	1	,6	,6	12,8
125	2	1,3	1,3	14,1
126	1	,6	,6	14,7
127	1	,6	,6	15,4
132	4	2,5	2,6	17,9
133	1	,6	,6	18,6
135	1	,6	,6	19,2
137	2	1,3	1,3	20,5
140	3	1,9	1,9	22,4
142	1	,6	,6	23,1
145	1	,6	,6	23,7
146	1	,6	,6	24,4
148	1	,6	,6	25,0
150	9	5,7	5,8	30,8
153	1	,6	,6	31,4
154	1	,6	,6	32,1
155	2	1,3	1,3	33,3
160	2	1,3	1,3	34,6
161	1	,6	,6	35,3
163	2	1,3	1,3	36,5
165	2	1,3	1,3	37,8
168	4	2,5	2,6	40,4
170	9	5,7	5,8	46,2
173	1	,6	,6	46,8
175	5	3,2	3,2	50,0
180	2	1,3	1,3	51,3
185	5	3,2	3,2	54,5

190	4	2,5	2,6	57,1
193	2	1,3	1,3	58,3
194	1	,6	,6	59,0
195	2	1,3	1,3	60,3
200	8	5,1	5,1	65,4
201	1	,6	,6	66,0
202	1	,6	,6	66,7
205	4	2,5	2,6	69,2
210	7	4,5	4,5	73,7
215	4	2,5	2,6	76,3
217	1	,6	,6	76,9
220	2	1,3	1,3	78,2
221	1	,6	,6	78,8
222	1	,6	,6	79,5
225	3	1,9	1,9	81,4
227	1	,6	,6	82,1
230	4	2,5	2,6	84,6
236	1	,6	,6	85,3
240	3	1,9	1,9	87,2
250	1	,6	,6	87,8
253	3	1,9	1,9	89,7
255	1	,6	,6	90,4
275	5	3,2	3,2	93,6
290	1	,6	,6	94,2
300	4	2,5	2,6	96,8
302	2	1,3	1,3	98,1
310	1	,6	,6	98,7
345	1	,6	,6	99,4
450	1	,6	,6	100,0
Σύνολο	156	99,4	100,00	
Τιμές που λείπουν	1	,6		
Σύνολο	157	100,00		

Πίνακας 2: Πίνακας Συχνοτήτων για την μεταβλητή Horsepower

Χρησιμοποιώντας τον παραπάνω πίνακα συχνοτήτων και κατασκευάζοντας κλάσης μεγέθους μπορούμε να κατασκευάσουμε ένα ιστόγραμμα. Το ιστόγραμμα είναι κατάλληλο για ποιοτικά δεδομένα ενώ το πλάτος κάθε στήλης αντιστοιχεί στο εύρος της αντίστοιχης κλάσης. Στον οριζόντιο άξονα βρίσκονται οι τιμές της μεταβλητής ενώ στον κάθετο φαίνεται η συχνότητα εμφάνισης της κλάσης, π.χ. περίπου 22 αυτοκίνητα του δείγματος έχουν ιπποδύναμη (horsepower) μεταξύ 175 και 200 ίππων.



Διάγραμμα 4: Ιστόγραμμα Συχνοτήτων για την μεταβλητή Horsepower

1.5.2 Το Διάγραμμα Πίτας

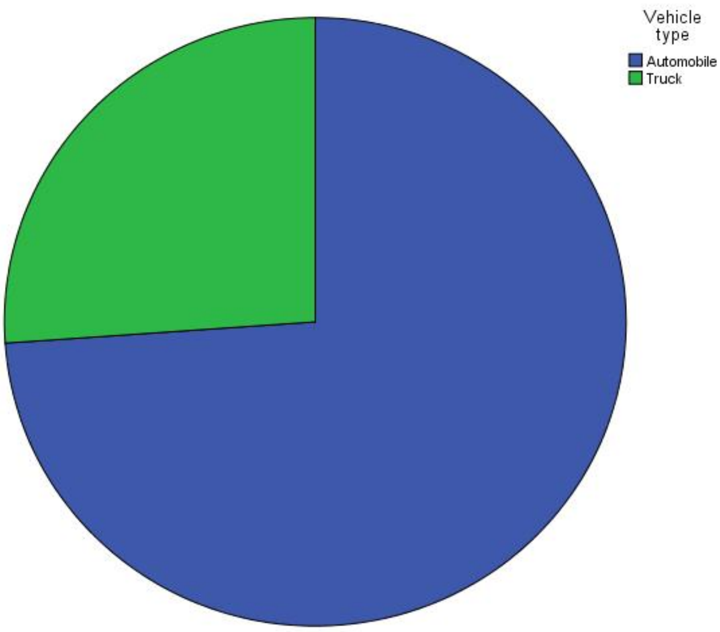
Ένα άλλο, εξίσου δημοφιλές διάγραμμα παρουσίασης ποιοτικών δεδομένων αυτή την φορά είναι το διάγραμμα πίτας ή απλά πίτα όπως έχει επικρατήσει να λέγεται.

Η πίτα χωρίζεται σε μέρη ανάλογα με το ποσοστό που καταλαμβάνει το χαρακτηριστικό που θέλουμε να παραστήσουμε στο δείγμα μας.

Το παρακάτω διάγραμμα πίτας αναφέρεται στον τύπο του αυτοκινήτου που υπάρχει στο δείγμα μας. Έτσι, σύμφωνα με τους πίνακες, έχουμε δύο τύπους αυτοκινήτων, τα συμβατικά αυτοκίνητα (automobile) και τα φορτηγά (trucks).

	Συχνότητα	**Ποσοστό**	**Έγκυρο Ποσοστό**	**Αθροιστικό Ποσοστό**
Συμβατικά αυτοκίνητο (automobile)	116	73,9	73,9	73,9
Φορτηγά (trucks)	41	26,1	26,1	100,00
	157	100,00	100,00	

Διάγραμμα 5: Πίνακας Συχνοτήτων για την μεταβλητή Τύπος Αυτοκινήτου (VehicleType)

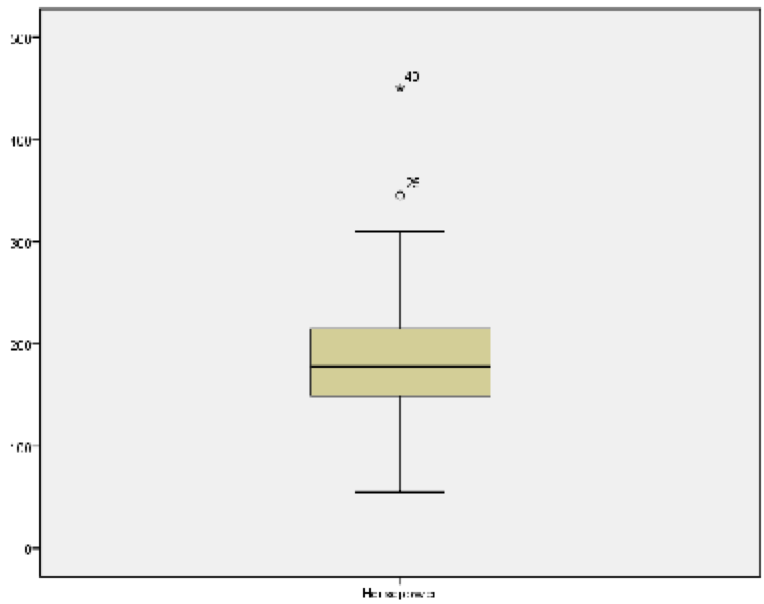


Διάγραμμα 6: Διάγραμμα πίτας για την μεταβλητή Τύπος Αυτοκινήτου (VehicleType)

1.5.3 Το Διάγραμμα Μίσχου-Φύλλου

Ένα τελευταίο διάγραμμα που δεν τυγχάνει την καθολική αναγνώριση των παραπάνω είναι το Διάγραμμα Μίσχου-Φύλλου ή BoxandWhiskeyPlotτο οποίοχρησιμοποιείται για αριθμητικά δεδομένα και την σύγκριση δειγμάτων. Το διάγραμμα που παρουσιάζεται παρακάτω αφορά ένα μόνο δείγμα .

Στο διάγραμμα Μίσχου-Φύλλου οι δυο ακραίες οριζόντιες γραμμές αντιστοιχούν στην μέγιστη και ελάχιστη παρατήρηση αντίστοιχα, ενώ η πάνω πλευρά του ορθογωνίου είναι το Q_3 (το 25% του δείγματος βρίσκεται πάνω από αυτή την τιμή) και η κάτω το Q_1 (το 25% των παρατηρήσεων του δείγματος βρίσκεται κάτω από αυτή την τιμή) αντίστοιχα. Η ευθεία γραμμή που βρίσκεται μέσα στο ορθογώνιο, δεν αντιστοιχεί στον αριθμητικό μέσο της μεταβλητής αλλά στην διάμεσοΟι τιμές που φαίνονται πάνω από την μέγιστη τιμή ή κάτω από την ελάχιστη αποτελούν τις ακραίες τιμές.



Διάγραμμα 7: Διάγραμμα Μίσχου-Φύλλου για την μεταβλητή Horsepower

1.5.4 Επιπλέον Διαγράμματα

Ενδεικτικά αναφέρουμε και άλλες κατηγορίες διαγραμμάτων όπως το διάγραμμα σημείων (dotplot) που είναι κατάλληλο για λίγες παρατηρήσεις και κάθε τιμή αντιστοιχεί σε μια τελεία και το ραβδόγραμμα που αποτελείται και αυτό από κάθετες ράβδους όπως το ιστόγραμμα με την διαφορά ότι οι ράβδοι έχουν απόσταση μεταξύ τους και επιπλέον αναφέρεται σε ποιοτικά δεδομένα.

Κεφάλαιο 2

Τυχαίες Μεταβλητές και οι Κατανομές τους

Σε αυτό το κεφάλαιο θα παρουσιάσουμε τις κατηγορίες τυχαίων μεταβλητών δηλαδή τις συνεχείς και τις διακριτές τυχαίες μεταβλητές καθώς και τις κατανομές τους. Επίσης, θα παρουσιάσουμε τα διαγράμματα πυκνότητας πιθανότητας για κάθε μια από αυτές και θα παρουσιάσουμε και τις πρώτες δυο ροπές τους, δηλαδή τον μέσο και την διακύμανση.

Για το κεφάλαιο αυτό χρησιμοποιήσαμε στοιχεία απο τις πηγές :[2], και τις διαδικτυακές πηγές : [9], [10], [11].

2.1 Τυχαίες Μεταβλητές

Ως τυχαία μεταβλητή (η οποία συνήθως συμβολίζεται με κεφαλαία γράμματα) καλείται η συνάρτηση εκείνη η οποία απεικονίζει τα δυνατά αποτελέσματα (ενδεχόμενα) ενός πειράματος τύχης , τα οποία αποτελούν τα σημεία του δειγματικού χώρου Sτου πειράματος, στο σύνολο των πραγματικών αριθμών R.

Το πείραμα τύχης μπορεί να ενδιαφέρει την μελέτη ενός συγκεκριμένου χαρακτηριστικού όπως το εισόδημα ή η κατανάλωση των ατόμων διαφορετικής ηλικίας. Τις τιμές που παίρνουν οι τυχαίες μεταβλητές τις συμβολίζουμε με μικρά γράμματα. Για παράδειγμα αν Χ συμβολίζει μια τυχαία μεταβλητή, τότε οι τιμές της θα συμβολίζονται ως x.

2.2 Κατηγορίες τυχαίων μεταβλητών

Οι τυχαίες μεταβλητές χωρίζονται σε δύο κατηγορίες: τις *διακριτές* και τις *συνεχείς*. Οι πρώτες περιέχουν ένα πεπερασμένο (αριθμήσιμο) πλήθος τιμών, ενώ οι δεύτερες περιλαμβάνουν το σύνολο των πραγματικών αριθμών ή διαστήματα αυτών.

2.3 Πιθανότητα και Συνάρτηση Πυκνότητας Πιθανότητας

Εφόσον οι τιμές των τυχαίων μεταβλητών αποτελούν το αποτέλεσμα ενός πειράματος τύχης, αυτές θα εμφανίζονται με κάποια συχνότητα η οποία καλείται πιθανότητα. Η συνάρτηση μέσω της οποίας οι τιμές μιας τυχαίας μεταβλητής απεικονίζονται στις πιθανότητές τους συνήθως συμβολίζεται ως *φ(x)* και καλείται *συνάρτηση πυκνότητας πιθανότητας* (probabilitydensityfunction) της μεταβλητής ή *κατανομή τηςX*.

Για διακριτές τυχαίες μεταβλητές, των οποίων οι τιμές είναι αριθμήσιμες, η συνάρτηση κατανομής ορίζεται ως $f(c) = P(X = x)$. Αυτή δίνει την πιθανότητα η μεταβλητή X να πάρει μια συγκεκριμένη τιμή, π.χ. την x=0. Αν η τυχαία μεταβλητή είναι συνεχής, τότε η συνάρτηση κατανομής *φ(x)* ορίζεται ως η πιθανότητα η τυχαία μεταβλητή X να πάρει μια τιμή x σε κάποιο διάστημα του συνόλου των πραγματικών αριθμών, έστω το (α,b). Δηλαδή, αυτή ορίζεται ως $f(c) = P(a < X \le b)$.

2.4 Συναρτήσεις κατανομής

2.4.1 Διακριτές κατανομές

- **Ομοιόμορφη διακριτή κατανομή**

Η Ομοιόμορφη διακριτή κατανομή εμφανίζεται στις περιπτώσεις όπου η υπό εξέταση τυχαία μεταβλητή X παίρνει πεπερασμένο πλήθος τιμών (π.χ. $X \in \{1,2,...,n\}$) και όλες οι πιθανότητες $P(X= i)$ είναι ισοπίθανες.

Ο ορισμός της ομοιόμορφης διακριτής κατανομής παρουσιάζεται παρακάτω:

Η κατανομή με συνάρτηση πιθανότητας

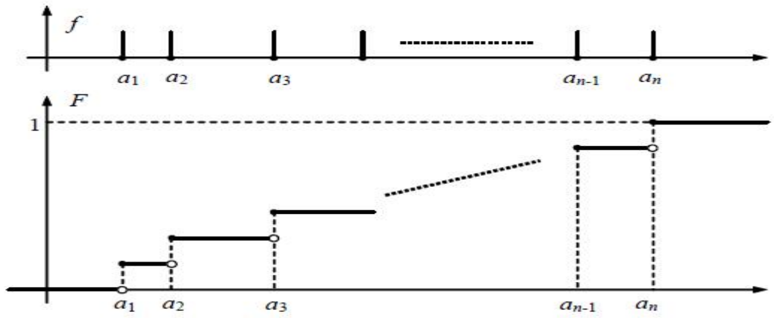
$$f(a_i) = \frac{1}{n} \quad \mathrm{i} = 1,2,\ldots \mathrm{n} \tag{1}$$

καλείται διακριτή ομοιόμορφη κατανομή στο σύνολο $\{a_1, a_2, \ldots, a_n\}$

Η συνάρτηση κατανομής της διακριτής ομοιόμορφης θα είναι

$$F(a_k) = \sum_{i=1}^{k} f(a_i) = \sum_{i=1}^{k} \frac{1}{n} = \frac{k}{n} \tag{2}$$

k=1,2,…n



Διάγραμμα 8: Συνάρτηση πιθανότητας και συνάρτηση κατανομής της διακριτής ομοιόμορφης κατανομής

Αν X ακολουθεί την διακριτή ομοιόμορφη κατανομή με πιθανότητα P(X=x)=1/n τότε

$$m = E(X = \frac{n+1}{2}) \text{ και } s^2 = V(X) = \frac{n^2-1}{12} \qquad (3,4)$$

Ø Διωνυμική κατανομή

Η Διωνυμική κατανομή εφαρμόζεται σε πειράματα όπου πραγματοποιούνται *n* επαναλήψεις της ίδιας διαδικασίας (π.χ. *n*ρίψεις ενός νομίσματος) και υπάρχουν μόνο τα ενδεχόμενα επιτυχία και αποτυχία .Η πιθανότητα επιτυχίας παραμένει σταθερή και οι επαναλήψεις του πειράματος είναι ανεξάρτητες. Η διωνυμική κατανομή μας δίνει την πιθανότητα να έχουμε ακριβώς *x* επιτυχίες στις ανεξάρτητες επαναλήψεις του πειράματος αν p είναι η πιθανότητα επιτυχίας και 1-p=q η πιθανότητα αποτυχίας. Το αποτέλεσμα κάθε ενός από αυτά τα πειράματα μπορεί να είναι είτε 1:επιτυχία είτε 0:αποτυχία.

Επομένως ο ορισμός της Διωνυμικής κατανομής μπορεί να γραφτεί ως εξής:

Η κατανομή με συνάρτηση πιθανότητας

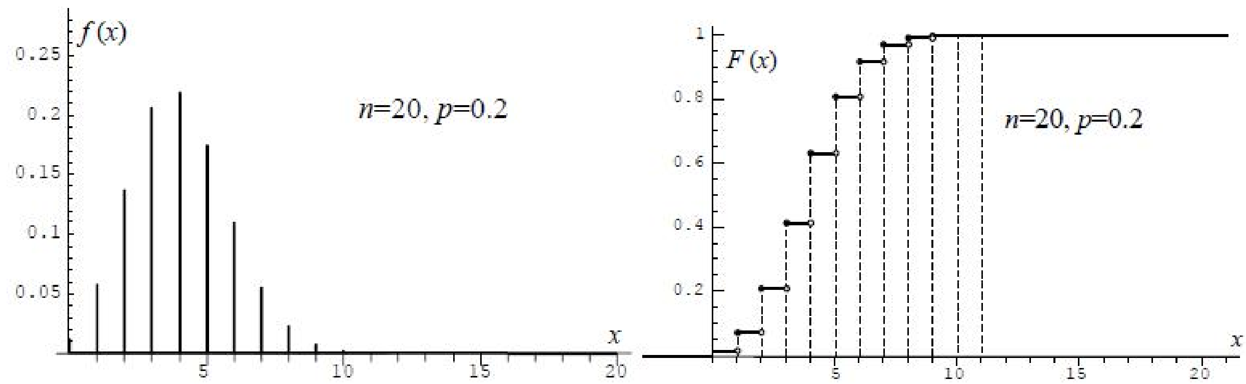
$$f(x) = P(X \le x) = \sum_{i=0}^{x} f(i) = \sum_{i=0}^{x} \binom{n}{i} p^x (1-p)^{n-x}, x = 0,1,\mathbf{KK}n \qquad (5)$$

καλείται διωνυμική κατανομή με παραμέτρους n>0, p∈ (0,1)

Η συνάρτηση κατανομής της διωνυμικής κατανομής θα είναι

$$F(x) = P(X \leq x) = \sum_{i=0}^{x} f(i) = \sum_{i=0}^{x} \binom{n}{i} p^{i} (1-p)^{n-1}, x = 0,1,\ldots n$$

(6)



Διάγραμμα 9: Συνάρτηση πιθανότητας και συνάρτηση κατανομής της διωνυμικής κατανομής για n=20 και p=0.2

Τέλος η μέση τιμή και η διακύμανση της παρούσας κατανομής με παραμέτρους n και p δίνεται από τους τύπους

$\mu = E(X) = np$ και $\sigma^2 = V(X) = npq$ (7,8)

Ø Γεωμετρική κατανομή

Αν X η τυχαία μεταβλητή που εκφράζει το πλήθος των δοκιμών μέχρι και την εμφάνιση της πρώτης επιτυχίας, τότε η X ακολουθεί την γεωμετρική κατανομή με παράμετρο p.

Ο ορισμός της γεωμετρικής κατανομής είναι ο ακόλουθος:

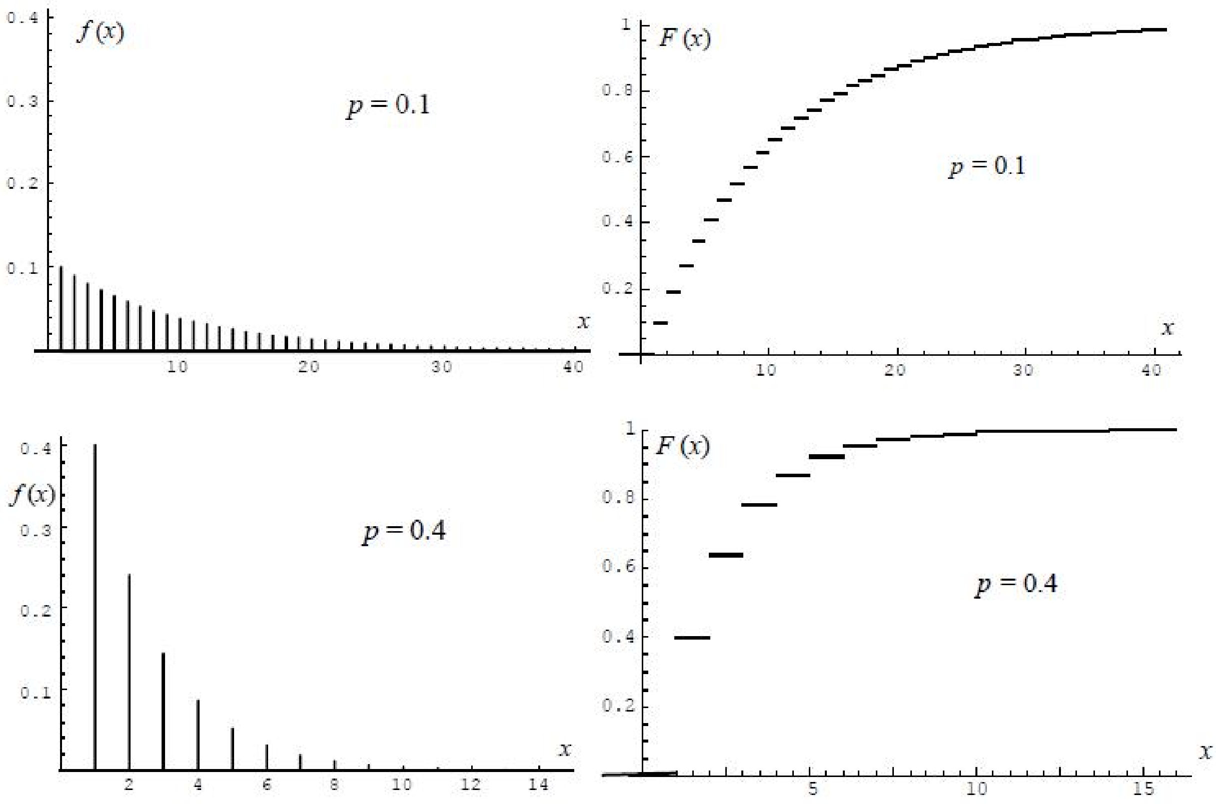
Η κατανομή με συνάρτηση πιθανότητας

$$f(x) = (1-p)^{x-1} p, x = 1,2,\ldots \quad (9)$$

Η συνάρτηση κατανομής της γεωμετρικής κατανομής θα είναι

$$F(x) = P(X \leq x) = \sum_{i=1}^{x} f(i) = 1 - (1-p)^{x}, x = 1,2,\ldots \quad (10)$$

Η συνάρτηση πιθανότητας και η συνάρτηση κατανομής της γεωμετρικής κατανομής δίνεται στα επόμενα σχήματα για τις ακόλουθες τιμές του *p*.



Διάγραμμα 10: Συνάρτηση πιθανότητας και συνάρτηση κατανομής της γεωμετρικής κατανομής για p=0.1 και p=0.4

Η μέση τιμή και η διακύμανση της γεωμετρικής κατανομής δίνονται από τους παρακάτω τύπους:

$$\mu = E(X) = \frac{1}{p} \text{ και } s^2 = V(X) = \frac{1-p}{p^2} \qquad (11,12)$$

Ø Κατανομή Poisson

Γενικός ορισμός:

Έστω μια ακολουθία τυχαίων ενδεχομένων για τα οποία:

(α) τα ενδεχόμενα είναι ανεξάρτητα

(β) τα ενδεχόμενα πραγματοποιούνται με ένα σταθερό μέσο ρυθμό λ ανά **μονάδα χρόνου** (γ) τα ενδεχόμενα δεν πραγματοποιούνται ταυτόχρονα. Πραγματοποιούνται σε διαφορετικά χρονικά διαστήματα και έστω **X = ο αριθμός των ενδεχομένων που πραγματοποιούνται στη μονάδα του χρόνου**

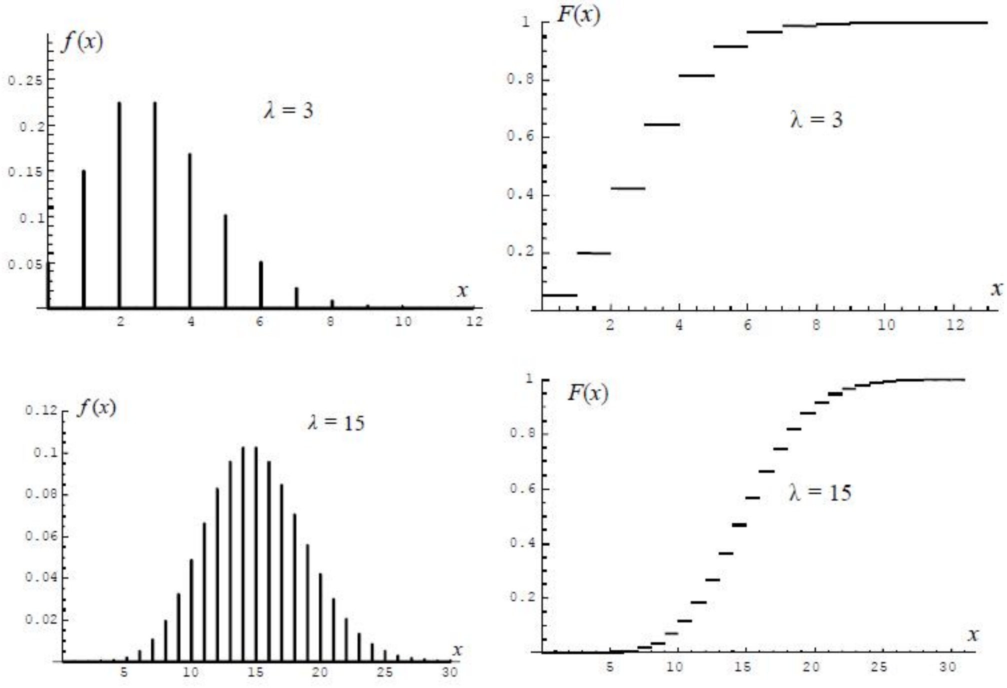
Τότε η X ακολουθεί την κατανομή Poisson με παράμετρο λ.

Επομένως η κατανομή με συνάρτηση πιθανότητας

$$f(x) = e^{-\lambda}\frac{\lambda^x}{x!}, x = 0,1,2\ldots \qquad (13)$$

καλείται κατανομή Poisson με παράμετρο λ και συμβολίζεται με $P_0(\lambda)$ και η συνάρτηση κατανομής της κατανομής Poisson θα είναι

$$F(x) = P(X \le x) = \sum_{i=0}^{x} f(i) = e^{-\lambda}\sum_{i=0}^{x}\frac{\lambda^i}{i!}, x = 0,1,2\ldots \qquad (14)$$



Διάγραμμα 11: Συνάρτηση πιθανότητας και συνάρτηση κατανομής της κατανομής poisson για λ= 3 και λ=15

Τέλος η μέση τιμή και η διάμεσος της κατανομής Poisson δίνεται από τους παρακάτω τύπους:

$\mu = E(X)=\lambda$ και $\sigma^2 = V(X) = \lambda$ (15,16)

2.4.2 Συνεχείς κατανομές

- **Ομοιόμορφη Συνεχής Κατανομή**

Η τυχαία μεταβλητή Χ λέμε ότι ακολουθεί την ομοιόμορφη κατανομή στο διάστημα [α, b] όταν το Χ είναι το ίδιο πιθανό να πάρει τιμές σε οποιαδήποτε υποδιάστημα μεταξύ του α και του b.

Επομένως ο ορισμός της ομοιόμορφης συνεχής κατανομής είναι ο εξής:

Η συνεχής κατανομή με συνάρτηση πυκνότητας πιθανότητας

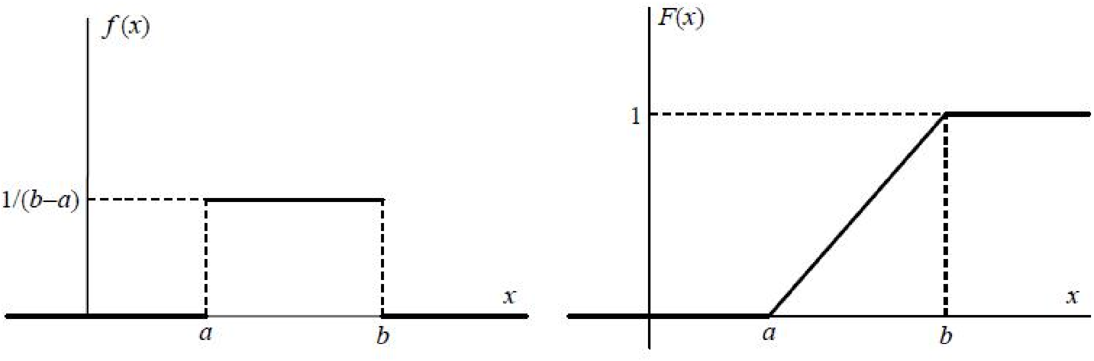
$$f(x) = \frac{1}{b-a}, x \in [a,b] \tag{17}$$

(f(x)=0, x∉ [α,b]) καλείται συνεχής ομοιόμορφη κατανομή στο διάστημα [α, b].

Η συνάρτηση κατανομής της συνεχούς ομοιόμορφης κατανομής είναι

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{a}^{x} \frac{1}{b-a}dt = \frac{x-a}{b-a}, x \in [a,b] \tag{18}$$

ενώ$F(x) = 0$ an$x\backslash\ a$, $F(x) = 1$ an$x|\ b$.



Διάγραμμα 12: Συνάρτηση πιθανότητας και συνάρτηση κατανομής της ομοιόμορφης συνεχής κατανομής

Τέλος η μέση τιμή και διασπορά της συνεχούς ομοιόμορφης κατανομής είναι:

$$m = E(X) = \frac{a+b}{2} \text{ και } s^2 = V(X) = \frac{(b-a)^2}{12} \tag{19}$$

Ø Κανονική Κατανομή

Η κανονική κατανομή είναι η σημαντικότερη κατανομή με τις περισσότερες εφαρμογές. Σύμφωνα με το *κεντρικό οριακό θεώρημα* κάθε φυσική ποσότητα της οποίας η τιμή μπορεί να θεωρηθεί ότι διαμορφώνεται από έναμεγάλο αριθμό (ανεξάρτητων) παραγόντων ακολουθεί προσεγγιστικά κανονική κατανομή.

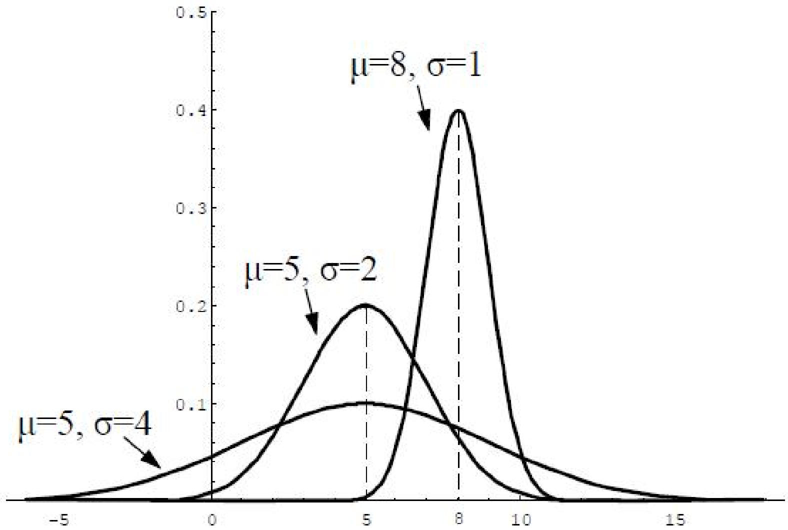
Ορισμός: Η συνεχής κατανομή με συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \frac{1}{s\sqrt{2p}} e^{\frac{(x-m)^2}{2s^2}} \qquad (20)$$

x∈R (μ ∈R, σ>0)

καλείται κανονική κατανομή με παραμέτρους μ, σ^2 και συμβολίζεται με $N(\mu,\sigma^2)$.

Το γράφημα της συνάρτησης πυκνότητας πιθανότητας της $N(\mu,\sigma^2)$ δίνεται στο επόμενο σχήμα για α) μ = 8, σ = 1, β) μ = 5, σ = 2, γ) μ = 5, σ = 4.



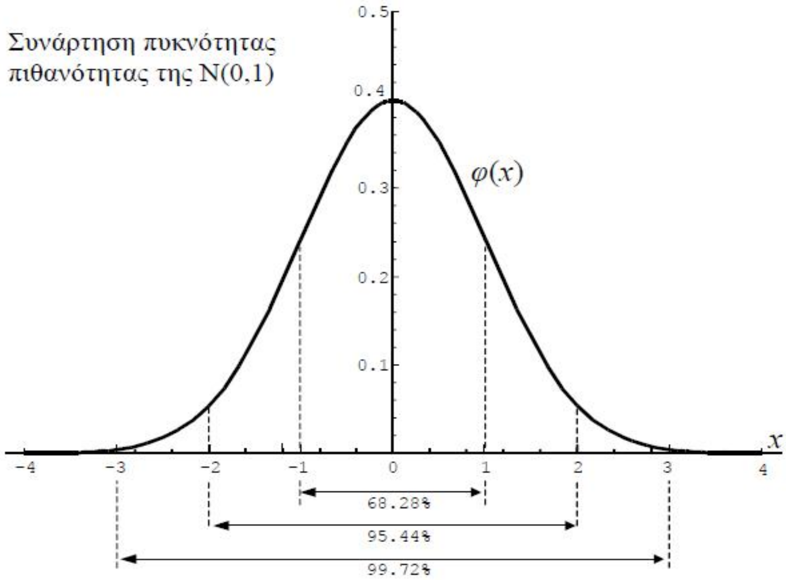
Διάγραμμα 13: Συνάρτησης πυκνότητας πιθανότητας της N(μ,σ2) για α) μ = 8, σ = 1, β) μ = 5, σ = 2, γ) μ = 5, σ = 4

Όσον αφορά την μέση τιμή και την διάμεσο αποδεικνύεται ότι:

$$E(X) = m \ \ \textit{και} \ \ V(X) = s^2 \qquad (21, 22)$$

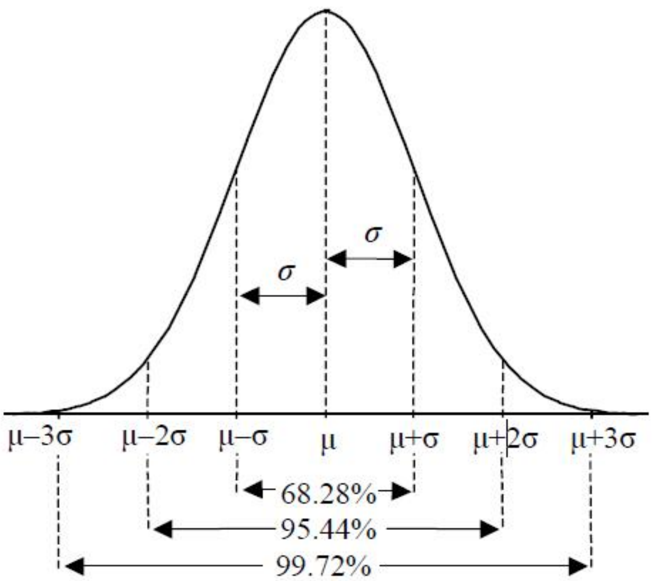
Η κανονική κατανομή N(0,1) (δηλ. μ=0, σ=1) θα καλείται ***τυπική κανονική κατανομή*** με συνάρτηση πυκνότητας πιθανότητας

Η συνάρτηση πυκνότητας πιθανότητας της τυπικής κανονικής κατανομής έχει τη μορφή:



Διάγραμμα 14: Συνάρτηση πυκνότητας πιθανότητας της τυπικής κανονικής κατανομής

Επίσης αν μία τυχαία μεταβλητή X ακολουθεί N(μ,σ2) τότε παίρνει τιμές μεταξύ του μ−3σ και του μ+3σ με πιθανότητα σχεδόν 1, τιμές μεταξύ του μ−2σ και του μ+2σ με πιθανότητα περίπου 95% και τιμές μεταξύ του μ−σ και του μ+σ με πιθανότητα περίπου 68%.



Διάγραμμα 15: Τυποποιημένη Κανονική Κατανομή και Ποσοστό Παρατηρήσεων

Ø Εκθετική κατανομή

Η εκθετική κατανομή εμφανίζεται συνήθως σε περιπτώσεις όπου μελετάμε *το χρόνο αναμονής μέχρι την πραγματοποίηση ενός γεγονότος.*

Ο ορισμός της εκθετικής κατανομής δίνεται παρακάτω:

Η συνεχής κατανομή με συνάρτηση πυκνότητας πιθανότητας

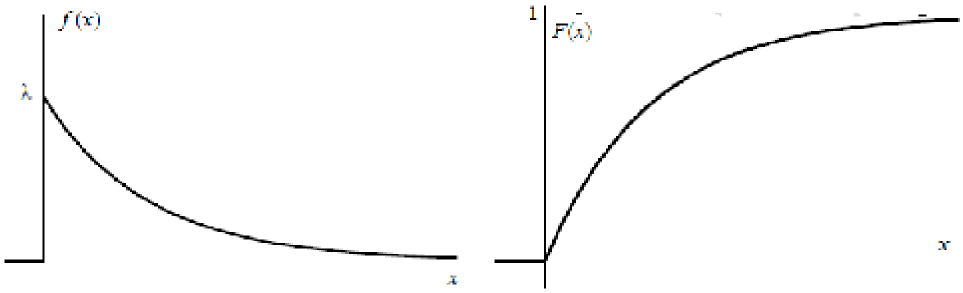
$$f(x) = l e^{-lx}, x \geq 0 (l \rangle 0) \qquad (23)$$

(f(x)=0,x<0) καλείται εκθετική κατανομή με παράμετρο λ.

Επιπλέον η συνάρτηση κατανομής της εκθετικής κατανομής θα είναι

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{0}^{x} l e^{-lt}dt = 1 - e^{-lx}, x \geq 0 \qquad (24)$$

ενώ F(x)=0 αν x<0



Διάγραμμα 16: Συνάρτηση πιθανότητας και συνάρτηση κατανομής της εκθετικής κατανομής

Τέλος η μέση τιμή και η διακύμανση της εκθετικής κατανομής δίνονται από τους παρακάτω τύπους:

$$E(X) = m = \frac{1}{l} \;\; \text{και} \;\; V(X) = s^{2} = \frac{1}{l^{2}} \qquad (25, 26)$$

Ø Κατανομή Γάμμα

Τέλος μια άλλη σημαντική κατανομή είναι η *κατανομή Γάμμα* η οποία μπορεί να θεωρηθεί ως γενίκευση της εκθετικής κατανομής.

Ορισμός: Η κατανομή με συνάρτηση πυκνότητας πιθανότητας

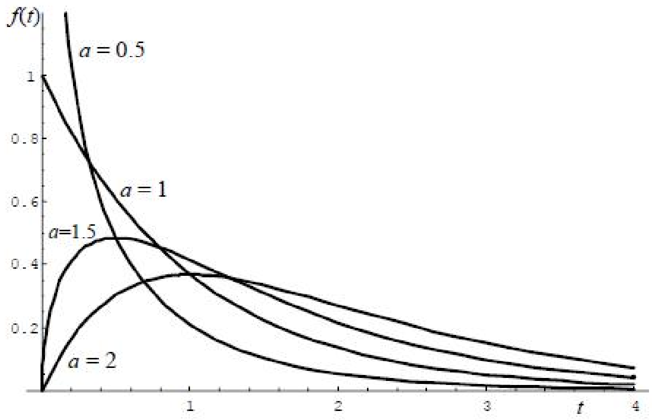
$$f(t) = \frac{l^a}{\Gamma(a)} t^{a-1} e^{-lt}, t \geq 0 \qquad (27)$$

όπου λ > 0, *και a*≥ 0 *καλείται κατανομή Γάμμα με παραμέτρους* (*λ*, *a*).

Ακόμα η συνάρτηση της Γάμμα κατανομής είναι η εξής:

$$\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx \qquad (28)$$

($\Gamma(a+1) = a\Gamma(a)$, $\Gamma(n) = (n-1)!$, $n \in N$, $\Gamma(1/2) = \pi$).



Διάγραμμα 17: Συνάρτηση πιθανότητας της Γάμμα κατανομής

Η μέση τιμή και η διακύμανση της Γάμμα κατανομής δίνονται από τους παρακάτω τύπους:

$$E(X) = m = \frac{a}{l} \text{ και } V(X) = s^2 = \frac{a}{l^2} \qquad (29)$$

Κεφάλαιο 3

Ανώτερα Θέματα Στατιστικής

3.1 Τι είναι η Οικονομετρία

Η Οικονομική επιστήμη, όπως η Βιολογία και η Φυσική, είναι εμπειρική επιστήμη, αντίθετα με τα Μαθηματικά ή τη Λογική, που είναι apriori επιστήμες. Επομένως οι οικονομικές θεωρίες ή προτάσεις μπορούν και πρέπει να ελέγχονται και να αξιολογούνται με βάση τα πραγματικά δεδομένα της οικονομικής ζωής.

Η Οικονομετρία αναφέρεται κυρίως στην ποσοτική πλευρά της Οικονομικής Επιστήμης και προσπαθεί να δώσει εμπειρικό περιεχόμενο στις «αφηρημένες» σχέσεις της οικονομικής θεωρίας.

Στην Οικονομετρία, η μαθηματικο-οικονομική και η στατιστική ανάλυση και έρευνα χρησιμοποιούνται συνδυασμένα, με κύριο αντικειμενικό σκοπό την εμπειρική εκτίμηση των σχέσεων αυτών αλλά και την επαλήθευση (έλεγχο) της οικονομικής θεωρίας.

Με άλλα λόγια, η οικονομετρική ανάλυση χρησιμοποιεί τις συναρτησιακές σχέσεις της οικονομικής θεωρίας και αφού τις μετατρέψει σε μαθηματικές, δηλαδή αφού κατασκευάσει ένα *υπόδειγμα* (model), προσπαθεί να τις εκτιμήσει εμπειρικά. Γι αυτή την εκτίμηση χρησιμοποιούνται στατιστικές μέθοδοι προσαρμοσμένες στα ιδιαίτερα χαρακτηριστικά των οικονομικών φαινομένων.

Για το κεφάλαιο αυτό χρησιμοποιήσαμε στοιχεία απο τις πηγές :[4], [6], [8], [12].

3.2 Σκοποί της Οικονομετρίας

Η οικονομετρική ανάλυση ενός οικονομικού φαινομένου και η εμπειρική εκτίμηση των παραμέτρων μια οικονομικής σχέσεως επιχειρείται για έναν ή περισσότερους από τους παρακάτω τρεις σκοπούς:

- ***Την εμπειρική επαλήθευση ή τον έλεγχο της θεωρίας.*** Ο έλεγχος μια θεωρίας ή μιας υποθέσεως είναι έλεγχος των προβλέψεων της θεωρίας ή της υποθέσεως. Ο όρος *πρόβλεψη* δεν αναφέρεται μόνο στο μέλλον, αλλά μπορεί να αναφέρεται και στο παρελθόν. Η ορθότητα και η ακρίβεια των προβλέψεων της θεωρίας, η ικανότητα δηλαδή της θεωρίας, η ικανότητα δηλαδή της θεωρίας να εξηγεί το φαινόμενο, ελέγχεται μόνο με αναφορά στα πραγματικά δεδομένα. Η θεωρία γίνεται δεκτή, ή λέμε ότι η θεωρία επαληθεύεται, αν επανειλημμένα αποτυγχάνουμε να αποδείξουμε ότι δεν είναι ορθή (Friedman, 1953). Το πρώτο βήμα για τον οικονομετρικό έλεγχο μιας θεωρίας είναι να εκφράσουμε τη θεωρία μαθηματικά, δηλαδή να διατυπώσουμε το *υπόδειγμα* ή τη *διατηρούμενη υπόθεση*. Αν από την αντιπαράθεση του υποδείγματος με τα πραγματικά δεδομένα προκύψει ότι το υπόδειγμα εξηγεί την πραγματική συμπεριφορά των οικονομικών μονάδων, καταναλωτών ή παραγωγών, αν με άλλα λόγια η θεωρία συμβιβάζεται με τα πραγματικά δεδομένα, η θεωρία επαληθεύεται και γίνεται δεκτή. Αν η θεωρία δε συμβιβάζεται με τα πραγματικά δεδομένα, τότε η θεωρία ή δε γίνεται δεκτή ή τροποποιείται με βάση τα καινούργια δεδομένα.

- ***Την άσκηση οικονομικής πολιτικής.*** Οι αριθμητικές τιμές των παραμέτρων των διαφόρων οικονομικών σχέσεων είναι απαραίτητες για την άσκηση οικονομικής πολιτικής. Αν και κατά πόσο θα ελαττωθεί η κατανάλωση, π.χ. ενός αγαθού ή πόσα θα είναι τα φορολογικά έσοδα από την επιβολή ενός άμεσου φόρου, θα εξαρτηθεί κατά κύριο λόγο από την ελαστικότητα της ζητήσεώς του. Επομένως, για να δοθούν απαντήσεις στα προηγούμενα ερωτήματα είναι απαραίτητη η γνώση της αριθμητικής τιμής της ελαστικότητας. Η χρησιμότητα της Οικονομετρίας για την άσκηση οικονομικής πολιτικής είναι συνεπώς και φανερή και σπουδαία.

- ***Την πρόβλεψη των μελλοντικών τιμών των οικονομικών μεταβλητών.*** Η σημασία που έχουν οι προβλέψεις για την μελλοντική πορεία των διαφόρων οικονομικών μεγεθών είναι φανερή. Αποτελούν τη βάση για ένα ορθολογικότερο προγραμματισμό και μια ορθολογικότερη λήψη αποφάσεων, είτε σε επίπεδο μικροοικονομικό (επιχείρηση) είτε σε επίπεδο μακροοικονομικό (κράτος). Με άλλα λόγια, οι προβλέψεις είναι απαραίτητες για τον έλεγχο και τη λήψη των αναγκαίων μέτρων από τους φορείς λήψεως αποφάσεων που θα επηρεάσουν τις τιμές των διαφόρων οικονομικών μεταβλητών. Για τους παραπάνω σκοπούς, με τη βοήθεια των οικονομετρικών υποδειγμάτων, μπορούμε να έχουμε τις απαραίτητες εκτιμήσεις των μελλοντικών τιμών των διαφόρων οικονομικών μεταβλητών, όπως το εισόδημα, η απασχόληση, η κατανάλωση, οι εισαγωγές κ.λπ.

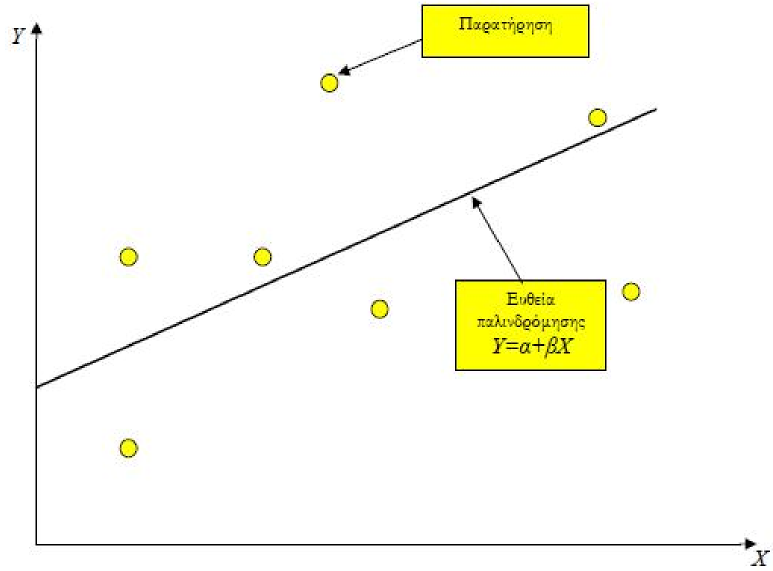
3.3 Το Γραμμικό Υπόδειγμα

Σύμφωνα με τον Keynes, η κατανάλωση είναι μια απλή γραμμική συνάρτηση του εισοδήματος, δηλαδή:

$$Y = a + bX \tag{30}$$

όπου *Y*είναι η κατανάλωση, *X*είναι το εισόδημα, α είναι η λεγόμενη αυτόνομη κατανάλωση και β είναι η οριακή ροπή για κατανάλωση.

Τα οικονομικά δεδομένα δεν σχηματίζουν ποτέ ακριβείς σχέσεις. Έτσι μπορεί να έχουμε μια κατάσταση όπως στο επόμενο διάγραμμα στο οποίο οι παρατηρήσεις δεν βρίσκονται ακριβώς πάνω σε μια ευθεία αλλά είναι επίσηςσαφές από την άλλη μεριά ότι έχουν μια προσεγγιστική γραμμική σχέση.



Διάγραμμα 18: Σχέση Κατανάλωσης-Εισοδήματος (Τσιώνας, 2010, σελ.2)

Όπως φαίνεται στο παραπάνω διάγραμμα, η κατανάλωση και το εισόδημα δεν βρίσκονται πάνω σε μια ευθεία, ούτε σε καμία άλλη προφανή συνάρτηση.

Αυτό συμβαίνει πάντοτε με τα οικονομικά στοιχεία, για δυο βασικούς λόγους:

- *Πρώτον*, οι θεωρίες δεν είναι τέλειες και είναι δυνατόν κάποιες σημαντικές μεταβλητές να έχουν παραλειφθεί από το υπόδειγμα.
- *Δεύτερον*, υπάρχουν σφάλματα μέτρησης στις μεταβλητές. Κατά συνέπεια θα πρέπει να τροποποιήσουμε τη θεωρία στη μορφή:

$$Y = a + bX + u \tag{31}$$

όπου **u**είναι μια μεταβλητή που περιλαμβάνει όλους τους παράγοντες που έχουμε παραλείψει και τα σφάλματα μέτρησης στην κατανάλωση. Η μεταβλητή αυτή λέγεται σφάλμα. Αν συμβολίσουμε με $\mathbf{Y_t}$την κατανάλωση του έτους *t*, με $\mathbf{X}_t$ το εισόδημα του έτους *i*και $\mathbf{u_t}$το σφάλμα του έτους *t*, θα πρέπει να έχουμε:

$$Y_t = a + bX_t + u_t, t = 1,2,\mathbf{K},n \qquad (32)$$

Τα (Y_t, X_t) λέγονται παρατηρήσεις ή στοιχεία. Τα α και β ονομάζονται παράμετροι του υποδείγματος. Είναι άγνωστες σταθερές που έχουν σημασία για τη θεωρία. Για παράδειγμα το $\frac{1}{1-b}$ είναι η τιμή του πολλαπλασιαστή σε μια κλειστή οικονομία με εξωγενή επένδυση και εξωγενή δημόσια δαπάνη.

Αυτή η τιμή, είναι πολύ σημαντική για τη θεωρία γιατί δίνει την αύξηση του εισοδήματος που θα προέλθει από μια αύξηση των δημοσίων δαπανών ή επενδύσεων.

Ωστόσο, η τιμή του πολλαπλασιαστή είναι άγνωστη, γιατί η τιμή του β είναι άγνωστη. Από την παραπάνω εξίσωση είναι φανερό ότι ο αριθμός των παρατηρήσεων είναι *n*, δηλαδή έχουμε στοιχεία για *n*χρόνια.

Τέλος, η μεταβλητή Y_tλέγεται **εξαρτημένη μεταβλητή**και η X_tλέγεται **ερμηνευτική μεταβλητή**.

3.3.1 Βασικές Υποθέσεις του Γραμμικού Υποδείγματος

Μια ολοκληρωμένη εξειδίκευση του Υποδείγματος της σχέσεως ανάμεσα στην Y και στη X περιγράφεται από τις ακόλουθες υποθέσεις:

$$Y_t = a + bX_t + u_t, \qquad (33)$$

Η υπόθεση αυτή αναφέρεται στην γραμμική σχέση που συνδέει τις μεταβλητές Y και X. Δηλαδή, πως κάθε τιμή Y_t είναι μια συνάρτηση της τιμής συν το διαταρακτικό όρο u_t. Η μεταβλητή Y είναι ***εξαρτημένη***μεταβλητή, ενώ η μεταβλητή X είναι η ***ανεξάρτητη***μεταβλητή.

$$u_t \sim \left(0, s^2\right), \qquad (34)$$

όπου u_t είναι τυχαία μεταβλητή,

$$\mathrm{E}(u_t) = 0, \qquad (35)$$

$$E\left(u_t^2\right) = s^2, \qquad (36)$$

Η παραπάνω υπόθεση αναφέρεται στη συμπεριφορά του διαταρακτικού όρου u_t.

Συγκεκριμένα οιυποθέσεις(35) και (36)σημαίνουν τα εξής:

i. Η μεταβλητή u_t είναι τυχαία μεταβλητή που παίρνει θετικές και αρνητικές τιμές, αλλά κατά μέσω όρο η τιμή της είναι μηδέν.

ii. Η διακύμανση της τυχαίας μεταβλητής u_t, είναι σταθερή για όλες τις τιμές Χ. Δηλαδή η διασπορά των τιμών της u_tαπό τον μέσο δεν αλλάζει όταν μεταβάλλεται η τιμή της X_t, αλλά παραμένει η ίδια. Όταν η διακύμανση παραμένει σταθερή, ο διαταρακτικός όρος είναι ***ομοσκεδαστικός*** ή χαρακτηρίζεται από ***ομοσκεδαστικότητα (homoscedasticity)***. Όταν η διακύμανση δεν είναι σταθερή, ο διαταρακτικός όρος είναι ***ετεροσκεδαστικός*** ή χαρακτηρίζεται από ***ετεροσκεδαστικότητα (heteroscedasticity)***. Η υπόθεση (36) επομένως, σημαίνει πως ο διαταρακτικός όρος είναι ομοσκεδαστικός.

$$E(u_t u_s) = 0, t \neq s \quad (37)$$

Η υπόθεση (37) σημαίνει ότι οι διαταρακτικοί όροι δε σχετίζονται με συσχετίζονται μεταξύ τους και ότι επομένως η συνδιακύμανση του διαταρακτικού όρου της παρατηρήσεως t, με το διαταρακτικό όρο οποιασδήποτε άλλης παρατηρήσεως s είναι μηδέν.

Η μεταβλητή Χ δεν είναι στοχαστική. Οι τιμές της παραμένουν σταθερές και δεν είναι όλες ίσες μεταξύ τους. (38)

Η παραπάνω υπόθεση αναφέρεται στην ερμηνευτική μεταβλητή Χ. Υποθέτουμε ότι η μεταβλητή Χ δεν είναι στοχαστική και πως οι τιμές της παραμένουν σταθερές σε μια υποθετική διαδικασία επαναλαμβανόμενης δειγματοληψίας.

Με άλλα λόγια, αν υποθέσουμε ότι παίρνουμε ένα μεγάλο αριθμό δειγμάτων για τις Υ και Χ μεγέθους Τ, οι τιμές της Χ δεν μεταβάλλονται από δείγμα σε δείγμα, αλλά παραμένουν σταθερές.

Οι τιμές φυσικά που παίρνει ο διαταρακτικός όρος μεταβάλλονται, όπως επίσης μεταβάλλονται και η τιμές της εξαρτημένες μεταβλητής. Επιπλέον υποθέτουμε ότι οι τιμές της Χ δεν είναι όλες ίσες μεταξύ τους, πράγμα που σημαίνει ότι η διακύμανση της Χ στο δείγμα είναι διαφορετική από το μηδέν.

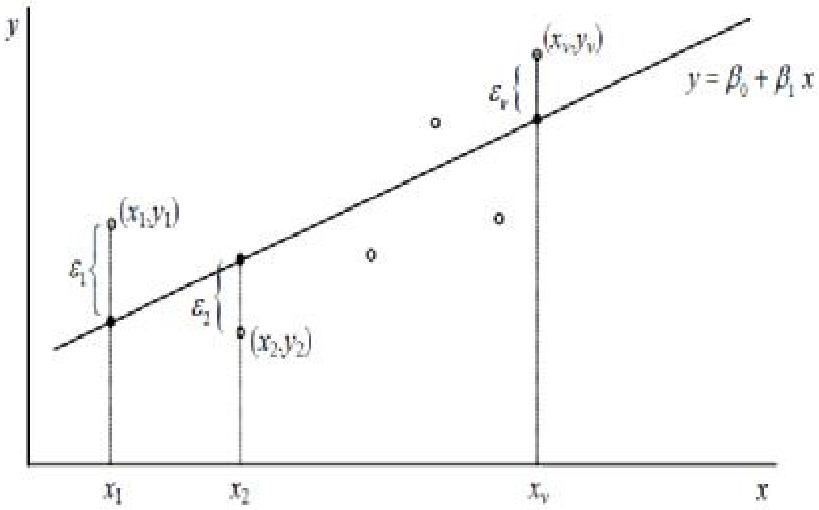
Οι υποθέσεις/σχέσεις (34) έως (38) αποτελούν το ***υπόδειγμα της κλασικής γραμμικής παλινδρόμησης (classicallinearregressionmodel)***, στο οποίο μπορούμε να εφαρμόσουμε τις κλασικές μεθόδους για να εκτιμήσουμε τις άγνωστες παραμέτρους του πληθυσμού α και β.

3.4 Η Μέθοδος των Ελαχίστων Τετραγώνων

Η εκτίμηση των παραμέτρων του υποδείγματος της γραμμικής παλινδρόμησης μπορεί να γίνει με διαφόρους τρόπους. Συνήθως όμως χρησιμοποιείται η ***μέθοδος των ελάχιστων τετραγώνων*** (leastsquaresmethod), καθώς είναι σχετικά απλή και οι εκτιμητές που προκύπτουν από αυτή έχουν πολλές από τις επιθυμητές ιδιότητες.

Ο αριθμός των εκτιμητών για μια άγνωστη παράμετρο του πληθυσμού, στην προκειμένη περίπτωση οι συντελεστές β_0 και β_1, που μπορούμε να έχουμε από ένα δείγμα, είναι στην ουσία άπειρος. Αυτό σημαίνει ότι μπορούμε να κατασκευάσουμε άπειρες γραμμές παλινδρομήσεως όταν έχουμε ένα δείγμα από Τ ζεύγη παρατηρήσεων για τις μεταβλητές Υ και Χ.

Με την μέθοδο των ελαχίστων τετραγώνων, όπως το όνομα φανερώνει, επιλέγουμε εκείνη τη γραμμή για την οποία το άθροισμα των τετραγώνων των αποκλίσεων (καταλοίπων) των παρατηρήσεων της Υ από τη γραμμή παλινδρομήσεως του δείγματος είναι ελάχιστο.



Διάγραμμα 19: Η Γραμμή Παλινδρόμησης και τα κατάλοιπα

Το παραπάνω σχήμα παριστάνει το διάγραμμα διασποράς για ένα δείγμα από Τ παρατηρήσεις για τις μεταβλητές Υ και Χ, έστω $\hat{Y}_v = \hat{b}_0 + \hat{b}_1 X_v$ η γραμμή παλινδρομήσεως που προκύπτει από τη μέθοδο των ελαχίστων τετραγώνων.

Για την τιμή X_1, η απόκλιση από τη γραμμή παλινδρομήσεως είναι $\hat{e}_1 = \boldsymbol{Y}_n - \hat{\boldsymbol{Y}}_1$ και γενικά $\hat{e}_1 = Y_n - \hat{Y}_n$ είναι η απόκλιση της Y_n. Το άθροισμα τώρα των τετραγώνων των αποκλίσεων, δηλαδή

$\hat{e}_1^2 + \hat{e}_2^2 + \mathbf{L} + \hat{e}_n^2 \sum_{n=1}^{n} \hat{e}_n^2$, που προκύπτει από τη μέθοδο των ελαχίστων τετραγώνων είναι ελάχιστο.

Δεν υπάρχει, με άλλα λόγια, άλλη γραμμή παλινδρομήσεως που το άθροισμα των τετραγώνων των αποκλίσεών της να είναι μικρότερο από αυτό που προκύπτει από τη μέθοδο των ελαχίστων τετραγώνων.

3.4.1 Ένα παράδειγμα: Η Εξίσωση Μισθών

Σε αυτή την ενότητα θα δούμε πως χρησιμοποιούνται οι οικονομετρικές μέθοδοι ώστε να ποσοτικοποιήσουμε σχέσεις που παρατηρούνται στον πραγματικό κόσμο και οι οικονομολόγοι ενδιαφέρονται να μελετήσουν και να ερμηνεύσουν.

Παρακάτω θα παρουσιάσουμε ένα παράδειγμα με πραγματικά δεδομένα από 526 εργαζόμενους και αυτό που θέλουμε να εξετάσουμε είναι αν τα έτη εκπαίδευσης ενός ατόμου (*educ*), η εργασιακή του εμπειρία ή αλλιώς προϋπηρεσία (*exper*) καθώς και τα χρόνια συνεργασίας με τον σημερινό εργοδότη του ατόμου (*tenure*) επηρεάζουν και πόσο το ύψος του μισθού (*wage*)ενός εργαζόμενου.

Είναι προφανέςαπό τα παραπάνω πως για να μπορέσουμε να διερευνήσουμε το ερώτημα που θέσαμε θα χρησιμοποιήσουμε ένα μοντέλο πολλαπλής παλινδρόμησης προκειμένου να εξετάσουμε την επίδραση των ανεξαρτήτων μεταβλητών (*educ, exper, tenure*) στην εξαρτημένη μεταβλητή (log(wage)). Η χρήση λογαρίθμων κατά την ανάλυση παλινδρόμησης αποτελεί σύνηθες μετασχηματισμό της εξαρτημένης μεταβλητής αλλά και των ερμηνευτικών.

Θα εκτιμήσουμε το εξής οικονομετρικό μοντέλο:

$$\log(wage_i) = a_1 + b_1 educ_i + b_2 \exp er_i + b_3 tenure_i + u_i \,,$$

Όπου $b_{1,} b_2, b_3$ είναι οι συντελεστές των ανεξάρτητων μεταβλητών και δείχνουν το πόσο επηρεάζουν οι ανεξάρτητες την εξαρτημένη μεταβλητή καθώς και την κατεύθυνση της επίδρασης. Επίσης, τα σύγχρονα εξειδικευμένα οικονομετρικά προγράμματα υπολογίσουν και τα τυπικά σφάλματα και μας δίνετε έτσι η δυνατότητα να πραγματοποιήσουμε ελέγχους στατιστικής σημαντικότητας για τους συντελεστές για να διαπιστώσουμε αν η επίδραση που ασκεί η ερμηνευτική μεταβλητή στην ερμηνευόμενη είναι ή όχι στατιστικά σημαντική. Θα πρέπει να αναφέρουμε πως ο όρος u_i αποτελεί τον διαταρακτικό όρο της γραμμής παλινδρόμησης.

Επιλέγοντας ένα επίπεδο σημαντικότητας και μια κριτική τιμή μπορούμε να προχωρήσουμε σε έλεγχο υποθέσεων για τις παραμέτρους του πληθυσμού.

Έπειτα από εκτίμηση του παραπάνω μοντέλου με την μέθοδο των ελαχίστων τετραγώνων, προέκυψε το εξής εκτιμημένο υπόδειγμα παλινδρόμησης:

$$\log(\hat{w}age) = \underset{(.104)}{.284} + \underset{(.007)}{.092} educ + \underset{.0017}{.0041} \exp er + \underset{(.003)}{.022} tenure,$$

$$n = 526, R^2 = .316$$

Στις παρενθέσεις αναφέρονται τα τυπικά σφάλματα των εκτιμήσεων.

Το επόμενο βήμα μετά την εξειδίκευση του υποδείγματος είναι να προχωρήσουμε σε ερμηνεία των αποτελεσμάτων.

Έτσι λοιπόν έχουμε πως, σύμφωνα με την παραπάνω εξειδίκευση και παραμενόντων όλων των άλλων παραγόντων σταθερών, ένα επιπλέον έτος εκπαίδευσης αυξάνει τον μισθό ενός εργαζόμενο κατά μέσο όρο κατά 9.2%. Η συγκεκριμένη εξειδίκευση, όπου χρησιμοποιήσαμε τον λογάριθμο του μισθού αντί τον μισθό, αλλάζει την ερμηνεία του συντελεστή της μεταβλητής σε σχέση με την περίπτωση που παίρναμε τον μισθό.

Συνεχίζοντας την ερμηνεία του υποδείγματος, ένα επιπλέον έτος εργασιακής εμπειρίας, παραμενόντων όλων των άλλων παραγόντων σταθερών, αυξάνει τον μισθό του ατόμου κατά μέσο όρο4.1% και ένα παραπάνω έτος συνεργασίας με τον σημερινό εργοδότη, παραμενόντων όλων των άλλων παραγόντων σταθερών, αυξάνει τον μισθό του εργαζόμενου κατά μέσο όρο κατά 2.2%.

Το μέγεθος του δείγματος που είναι μεγαλύτερο από 30 ($n = 526$) καθώς και τα τυπικά σφάλματα μας επιτρέπουν να προβούμε σε έλεγχο στατιστικών υποθέσεων χρησιμοποιώντας την tστατιστική ελέγχου.

Παραδείγματος χάρη, μπορούμε να εξετάσουμε αν τα έτη εκπαίδευσης, η προϋπηρεσία ή τα έτη συνεργασίας με τον ίδιο εργοδότη είναι μηδενικά στον πληθυσμό δηλαδή αν έχουν μηδενική επίδραση στον καθορισμό του μισθού ή όχι.

Μπορούμε να γράψουμε το παραπάνω με την μορφή μηδενικής και εναλλακτικής υπόθεσης ως εξής: $H_0 : b_2 = 0$ έναντι της εναλλακτικής $H_1 : b_2 \neq 0$. Το ίδιο ισχύει και για τα b_1, b_3.

Χρησιμοποιώντας την στατιστική t, επίπεδο στατιστικής σημαντικότητας 5% και την αντίστοιχη κριτική τιμή που ισούται με 1.645 μπορούμε να πραγματοποιήσουμε τον τέλεγχο ως εξής:

$$t - sta- = \frac{\hat{b}_2 - b_2}{s\hat{e}\left(\hat{b}_2\right)} = \frac{.0041 - 0}{.0017} = \frac{.0041}{.0017} \approx 2.41 > 1.645$$

Συγκρίνοντας την τιμή του τελέγχου με την κριτική τιμή καταλήγουμε πως η εργασιακή εμπειρία είναι στατιστικά σημαντική σε επίπεδο σημαντικότητας 5% και άρα επηρεάζει το ύψος του μισθού στην συγκεκριμένη εξειδίκευση. Ακολουθώντας το ίδιο σκεπτικό μπορούμε να προβούμε σε έλεγχο στατιστικής σημαντικότητας και για τις άλλες παραμέτρους.

Ένα άλλο σημείο που αξίζει προσοχής είναι το R^2 ή αλλιώς ο συντελεστής προσδιορισμού. Αποτελεί το ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής που ερμηνεύεται από την εκάστοτε παλινδρόμηση και παίρνει τιμές μεταξύ του μηδενός και της μονάδας, $0 \le R^2 \le 1$. Στο συγκεκριμένο παράδειγμα, $R^2 = .316$, που σημαίνει πως το 31.6% της μεταβλητότητας του μισθού ενός εργαζομένου εξηγείται από την παρούσα εξειδίκευση.

Γενικά, όσο μετακινούμαστε από το μηδέν στην μονάδα, μεγαλύτερο ποσοστό της μεταβλητότητας της εξαρτημένης μεταβλητής εξηγείται από την παλινδρόμηση. Εδώ, που έχουμε ένα σχετικά χαμηλό R^2 δεν θα πρέπει να μας προβληματίζει καθώς υπάρχουν επιπλέον μεταβλητές που μπορούν να χρησιμοποιηθούν για την ερμηνεία του ωρομισθίου του εργαζόμενου.

Κεφάλαιο 4

Εφαρμογή στο SPSS

v *Για τα την επίλυση της άσκησης χρησιμοποιήθηκε το SPSSv18*

Από το Τμήμα Προσωπικού μίας μεγάλης χρηματιστηριακής εταιρείας συλλέξαμε στοιχεία που αφορούν το σύνολο των 190 απασχολουμένων. Συγκεκριμένα η βάση δεδομένων περιλαμβάνει μεταξύ άλλων τις εξής πληροφορίες:

- Φύλο(sex): 1=άνδρας, 2=γυναίκα
- Ηλικία(age): σε έτη
- Κατηγορία εργασίας (categ_erg):
    1. Υπάλληλος Γραφείου
    2. Εκπαιδευόμενος Υπάλληλος
    3. Ασφάλεια
    4. Πτυχιούχος
    5. Πτυχιούχος με mba
    6. Χρηματικός Αναλυτής
    7. Προϊστάμενος Τμήματος
- Χρόνια Εκπαίδευσης(educ_years): Συνολικός χρόνος εκπαίδευσης (σε έτη)
- Προϋπηρεσία(proipiresia): Χρόνος προϋπηρεσίας πριν την πρόσληψη
- Επίδοση(epidosi): Μέση βαθμολογία επίδοσης του εργαζομένου κατά την διάρκεια της θητείας του (με άριστα το 100)
- Αρχικός μισθός(archikos_misthos):ο μισθός πρόσληψης κάθε εργαζομένου στην εταιρεία (ετήσιος)
- Σημερινός μισθός(simerinos_misthos):(ετήσιος)

Ακολουθεί ανάλυση των δεδομένων προκειμένου να διαπιστωθεί ποιοι παράγοντες επιδρούν στην διαμόρφωση του μισθού των εργαζομένων.

ΜΕΤΑΒΛΗΤΗ ΦΥΛΟ (SEX)

Στον πίνακα SEX υπάρχουν οι τιμές 1,2. Η τιμή 1 αντιστοιχεί σε άνδρα και η τιμή 2 σε γυναίκα.

- Μέσω του VARIABLE VIEW, στην στήλη VALUES καταχωρούμε τις παραπάνω αντιστοιχίες.
- Δημιουργούμε πίνακα κατανομής συχνοτήτων: Analyze -->Descriptive -->Statistics -->Frequencies.

Φύλο

	Frequency	Percent	Valid Percent	Cumulative Percent
Valıd andras	103	54,2	54,2	54,2
gunaika	87	45,8	45,8	100,0
Total	190	100,0	100,0	

Ο πίνακας FULO μας δείχνει ότι απο τις 190 εγγραφές της μεταβλητής SEX, οι 103 είναι άνδρες και 87 γυναίκες (στήλη FREQUENCY). Επίσης, φαίνεται, ότι οι 103 άνδρες αντιστοιχούν στο 54,2% των συνολικών εγγράφων. οι γυναίκες αντιστοιχούν στο 45,8% .

ΜΕΤΑΒΛΗΤΗ ΗΛΙΚΙΑ (AGE): ΣΕ ΕΤΗ

Επειδή η μεταβλητή AGE είναι συνεχής, την ταξινομούμε σε τάξεις.

-Προσδιορίζουμε τον αριθμό των τάξεων με τον κανόνα του sturges.

$$k = 1 + 3{,}322 logn$$

(n=190)

Μετά από τον υπολογισμό του k, το αποτέλεσμα ισούται με $k=8.57 \approx 9$ τάξεις.

Επιλεγούμε το πλάτος των τάξεων από τον τύπο $\delta = r/k$.

Statistics

hlikia

N	Valid	190
	Missing	0
Range		41,50
Minimum		23,00
Maximum		64,50

R= εύρος μεταβολης=41,5

Οπότε $\delta = 41{,}5/9 = 4.6 \approx 5$

-Ταξινομούμε τις τάξεις συμφώνα με τα παραπάνω:

1^{η}) 23 --> 28

2^{η}) 28 --> 33

3^{η}) 33 --> 38

4^{η}) 38 --> 43

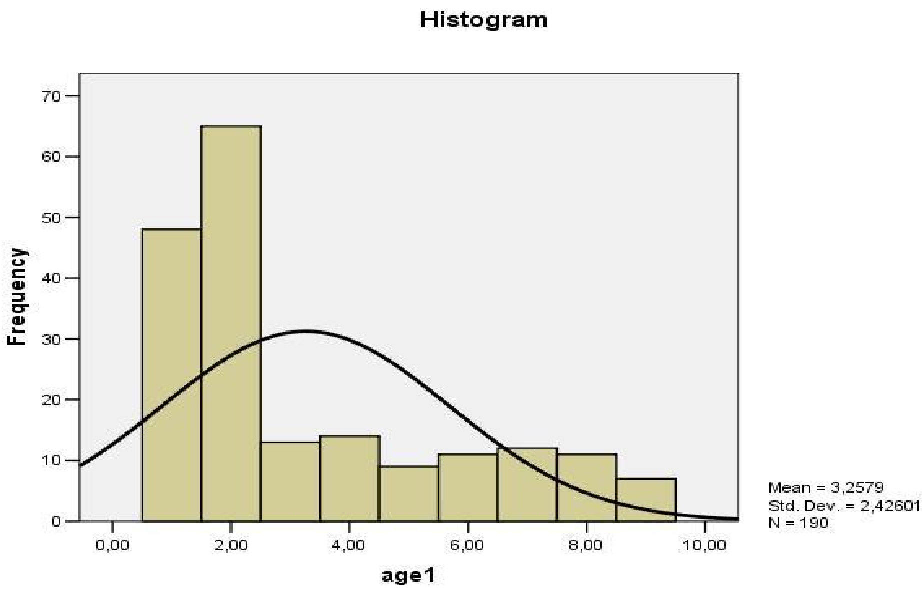
5^{η}) 43 --> 48

6^{η}) 48 --> 53

7^{η}) 53 --> 58

8^{η}) 58 --> 63

9^{η}) 63 --> 68



<u>ΣΧΟΛΙΑΣΜΟΣ ΠΙΝΑΚΑ AGE1</u>

age1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	48	25,3	25,3	25,3
	2,00	65	34,2	34,2	59,5
	3,00	13	6,8	6,8	66,3
	4,00	14	7,4	7,4	73,7
	5,00	9	4,7	4,7	78,4
	6,00	11	5,8	5,8	84,2
	7,00	12	6,3	6,3	90,5
	8,00	11	5,8	5,8	96,3
	9,00	7	3,7	3,7	100,0
	Total	190	100,0	100,0	

-Στον πίνακα βλέπουμε τις συχνότητες (στήλη FREQUENCY) και τα ποσοστά (στήλη PERCENT) κάθε τάξης.

Τα στοιχειά του πίνακα μπορούμε να τα διακρίνουμε και στο παρακάτω ιστόγραμμα:

-Για την κατασκευή αθροιστικού διαγράμματος βρίσκουμε τις κεντρικές τιμές κάθε τάξης.

ΚΕΝΤΡΙΚΗ ΤΙΜΗ

1$^{\eta}$) 23 --> 28 = 25,5

2$^{\eta}$) 28 --> 33 = 30,5

3$^{\eta}$) 33 --> 38 = 35,5

4$^{\eta}$) 38 --> 43 = 40,5

5$^{\eta}$) 43 --> 48 = 45,5

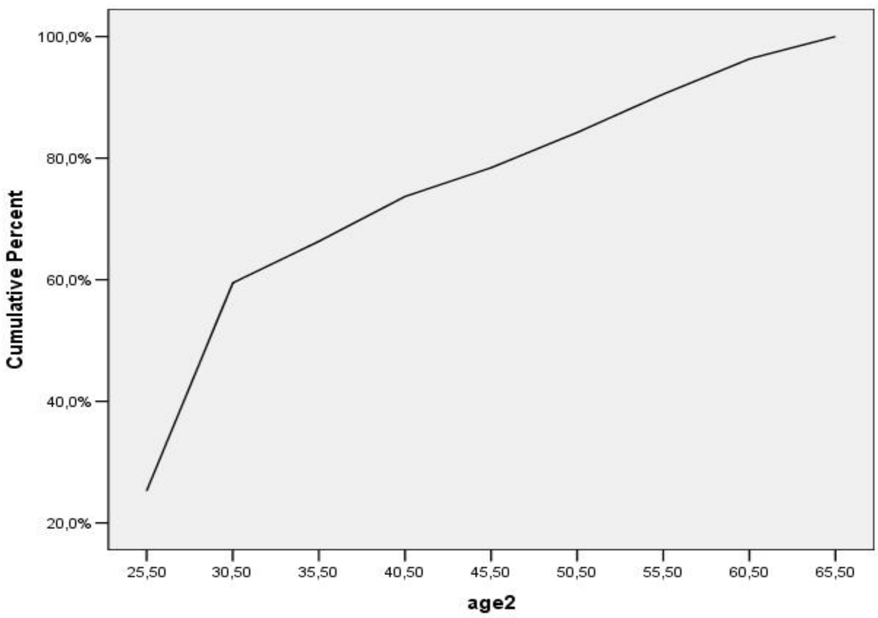
6η) 48 --> 53 = 50,5

7η) 53 --> 58 = 55,5

8η) 58 --> 63 = 60,5

9η) 63 --> 68 = 65,5

-Στηνσυνεχεία, ακλουθούμετηνέξηςδιαδικασία: graphs ->line ->simple ->summariesofgroupofcases ->define ->categoryaxis: age2 ->likerepresents:cum. % ofcases ->ok .



-Επίσης, για την μεταβλητή AGE υπολογίζουμε τα παρακάτω:

Statistics

hlikia

N	Valid	190
	Missing	0
Mean		36,6216
Median		30,9200
Mode		25,58(a)
Std. Deviation		12,13718
Variance		147,311
Range		41,50

Mean: Αριθμητικός μέσος (μ)

Median: Διάμεσος (Μ)

Mode: Επικρατούσα τιμή (M_0)

Std. Deviation: Τυπική απόκλιση (σ)

Variance: Διακύμανση (σ^2)

Range: Εύρος μεταβολής (R)

<u>ΣΧΟΛΙΑΣΜΟΣ ΠΙΝΑΚΑ</u>

- Ο αριθμητικός μέσος μας δείχνει τον μέσο όρο της ηλικίας των 190 καταχωρήσεων.
- Η διάμεσος μας δείχνει ότι το 50% των παρατηρήσεων είναι πάνω από το (μ) και το υπόλοιπο 50% κάτω από αυτή την τιμή.
- Η επικρατούσα τιμή μας δείχνει ότι η πιο συνήθης ηλικία είναι η (μ_0).
- Το εύρος μεταβολής μας δείχνει την διαφορά της μεγαλύτερης ηλικίας από την μικρότερη.
- Μετά, υπολογίζουμε τον συντελεστή μεταβλητότητας (cv) από τον τύπο: $\sigma/\mu = 12{,}137/36{,}621 \approx 0{,}331$

ΚΑΤΗΓΟΡΙΑ ΕΡΓΑΣΙΑΣ

Όπως και η μεταβλητή SEX έτσι και η κατηγορία εργασίας είναι ποιοτική μεταβλητή.

Μέσω της ίδιας διαδικασίας καταχωρούμε τις παρακάτω κατηγορίες:

1=Υπαλληλος Γραφείου

2=Εκπαιδευομενος Υπάλληλος

3=Ασφαλεια

4=Πτυχιουχος

5=Πτυχιουχος Με mba

6=Χρηματ. Αναλυτής

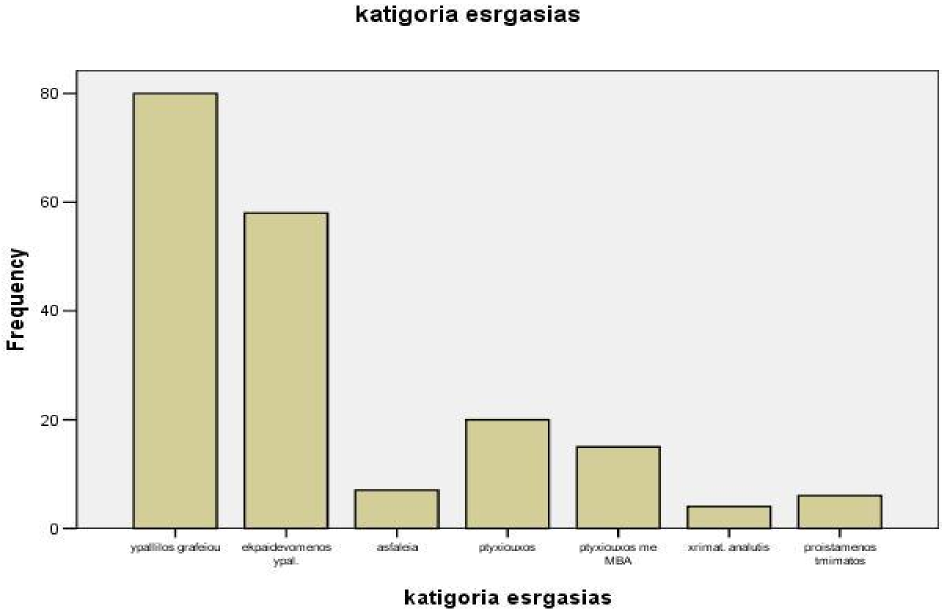
7=Προισταμενος Τμήματος

Μετά δημιουργούμε πίνακα κατανομής συχνοτήτων

katigoria esrgasias

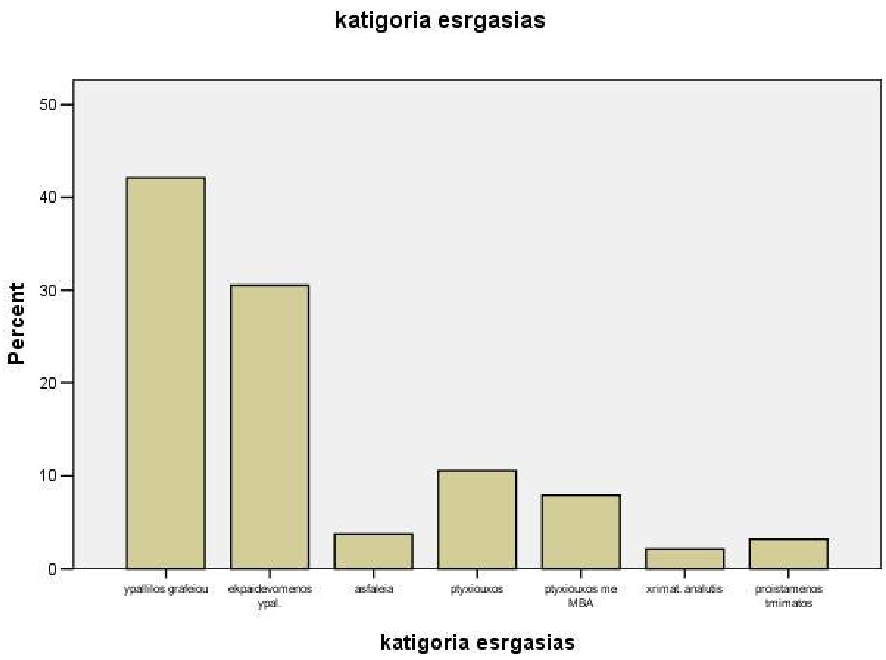
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ypallilos grafeiou	80	42,1	42,1	42,1
	ekpaidevomenos ypal.	58	30,5	30,5	72,6
	asfaleia	7	3,7	3,7	76,3
	ptyxiouxos	20	10,5	10,5	86,8
	ptyxiouxos me MBA	15	7,9	7,9	94,7
	xrimat. analutis	4	2,1	2,1	96,8
	proistamenos tmimatos	6	3,2	3,2	100,0
	Total	190	100,0	100,0	

Ο παραπάνω πίνακας μας δείχνει τις συχνότητες (frequency) και τα ποσοστά (percent) κάθε κατηγορίας. Αυτά φαίνονται και στα παρακάτω σχεδιαγράμματα.



<u>ΧΡΟΝΙΑ ΕΚΠΑΙΔΕΥΣΗΣ (EDUCATION YEARS)</u>

Επειδή η μεταβλητή είναι συνεχής, την ταξινομούμε σε τάξεις.



-Προσδιορίζουμε τον αριθμό των τάξεων με τον κανόνα του Sturges.

$K = 1 + 3.322 logn$

(n=190)

Μετά από τον υπολογισμό του k, το αποτέλεσμα ισούται με $k = 8.57 \approx 9$ τάξεις.

-Επιλέγουμε το πλάτος των τάξεων από τον τύπο $\delta = r/k$.

Statistics

xronia ekpaid.

N	Valid	190
	Missing	0
Range		13
Minimum		8
Maximum		21

R=εύρος μεταβολης=13

Οπότε $\delta = 13/9 = 1{,}44 \approx 1{,}5$

-Ταξινομούμε τις τάξεις σύμφωνα με τα παραπάνω:

1^{η}) 8 --> 9,5

2^{η}) 9,5 --> 11

3^{η}) 11 --> 12,5

4^{η}) 12,5 --> 14

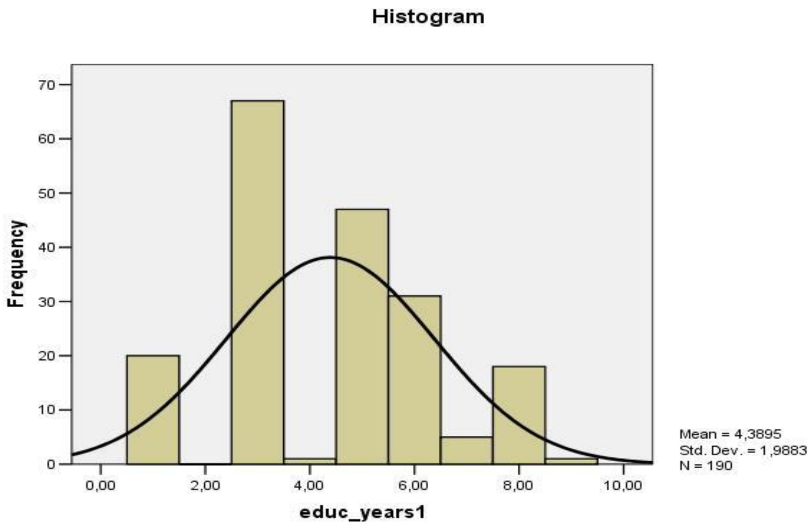
5^{η}) 14 --> 15,5

6^{η}) 15,5 --> 17

7^{η}) 17 --> 18,5

8^{η}) 18,5 --> 20

9^{η}) 20 --> 21,5



ΣΧΟΛΙΑΣΜΟΣΠΙΝΑΚΑEDUC_YEARS

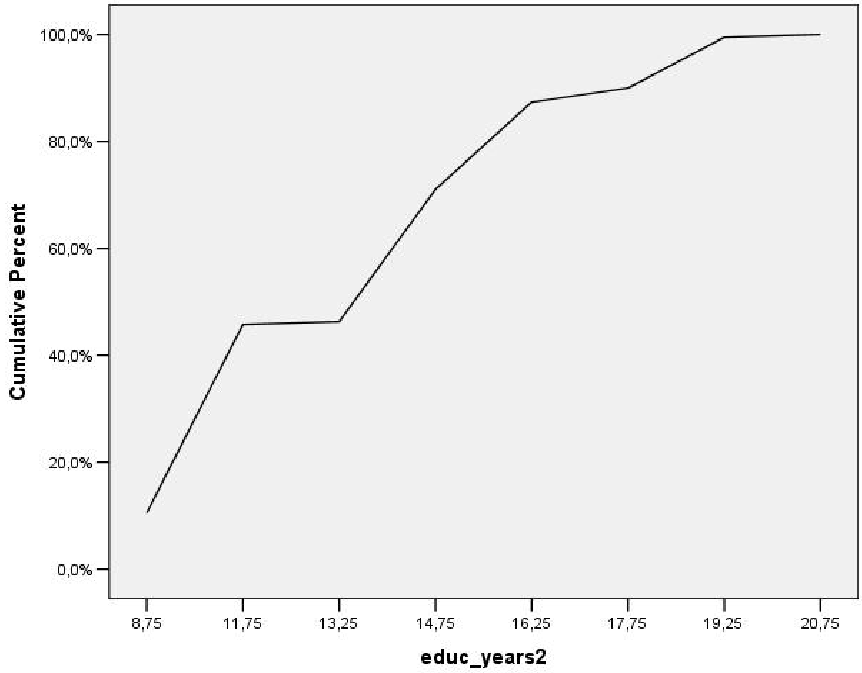
educ_years1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	20	10,5	10,5	10,5
	3,00	67	35,3	35,3	45,8
	4,00	1	,5	,5	46,3
	5,00	47	24,7	24,7	71,1
	6,00	31	16,3	16,3	87,4
	7,00	5	2,6	2,6	90,0
	8,00	18	9,5	9,5	99,5
	9,00	1	,5	,5	100,0
	Total	190	100,0	100,0	

Στον πίνακα βλέπουμε τις συχνότητες (στήλη frequency) και τα ποσοστά (στήλη percent) κάθε τάξης.

Τα στοιχεία του πίνακα μπορούμε να τα διακρίνουμε και στο παρακάτω ιστόγραμμα.

Για την κατασκευή αθροιστικού διαγράμματος βρίσκουμε τις κεντρικές τιμές κάθε τάξης.



ΚΕΝΤΡΙΚΗ ΤΙΜΗ

1η) 8 --> 9,5=8,75
2η) 9,5 --> 11= 10,25
3η) 11 --> 12,5=11,75
4η) 12,5-->14=13,25
5η) 14 -->15,5 =14,75
6η) 15,5--> 17=16,25
7η) 17 -->18,5=17,75
8η) 18,5 --> 20 =19,25
9η) 20 --> 21,5 =20,75

Στηνσυνέχεια, ακλουθούμετηνεξήςδιαδικασία: graphs ->line ->simple ->summariesofgroupofcases ->define ->category axis: educ_years2 ->like represents :cum. % ofcases ->ok

Επίσης, για την μεταβλητή αυτή υπολογίζουμε τα παρακάτω

Statistics

xronia ekpaid.

N	Valid	190
	Missing	0
Mean		13,88
Median		15,00
Mode		12
Range		13

ΣΧΟΛΙΑΣΜΟΣ ΠΙΝΑΚΑ

- Ο αριθμητικός μέσος μας δείχνει τον μέσο όρο των χρόνων εκπαίδευσης των 190 καταχωρήσεων.
- Η διάμεσος μας δείχνει ότι το 50% των παρατηρήσεων είναι πάνω από το (μ) και το υπόλοιπο 50% κάτω από αυτή την τιμή.
- Η επικρατούσα τιμή μας δείχνει ότι ο πιο συνήθης Χρόνος εκπαίδευσης είναι ο (μ_0).
- Το εύρος μεταβολής μας δείχνει την διάφορα του μεγαλύτερου χρόνου εκπαίδευσης από τον μικρότερο.

ΠΡΟΥΠΗΡΕΣΙΑ

Επειδή η μεταβλητή είναι συνεχής, την ταξινομούμε σε τάξεις. Προσδιορίζουμε τον αριθμό των τάξεων με τον κανόνα του Sturges.

$K = 1 + 3.322 log n$ (n=190)

Μετά από τον υπολογισμό του k, το αποτέλεσμα ισούται με $k = 8.57 \approx 9$ τάξεις. Επιλέγουμε το πλάτος των τάξεων από τον τύπο $\delta = r/k$.

Statistics

xronos proip.

N	Valid	190
	Missing	0
Range		36,50
Minimum		,00
Maximum		36,50

R= εύρος μεταβολης=36,50

Οπότε $\delta = 36{,}50/9 = 4{,}05 \approx 4{,}5$

Ταξινομούμε τις τάξεις σύμφωνα με τα παραπάνω:

1^{η}) 0 --> 4,5

2^{η}) 4,5 --> 9

3^{η}) 9--> 13,5

4^{η}) 13,5 --> 18

5^{η}) 18 --> 22,5

6^{η}) 22,5--> 27

7^{η}) 27--> 31,5

8^{η}) 31,5--> 36

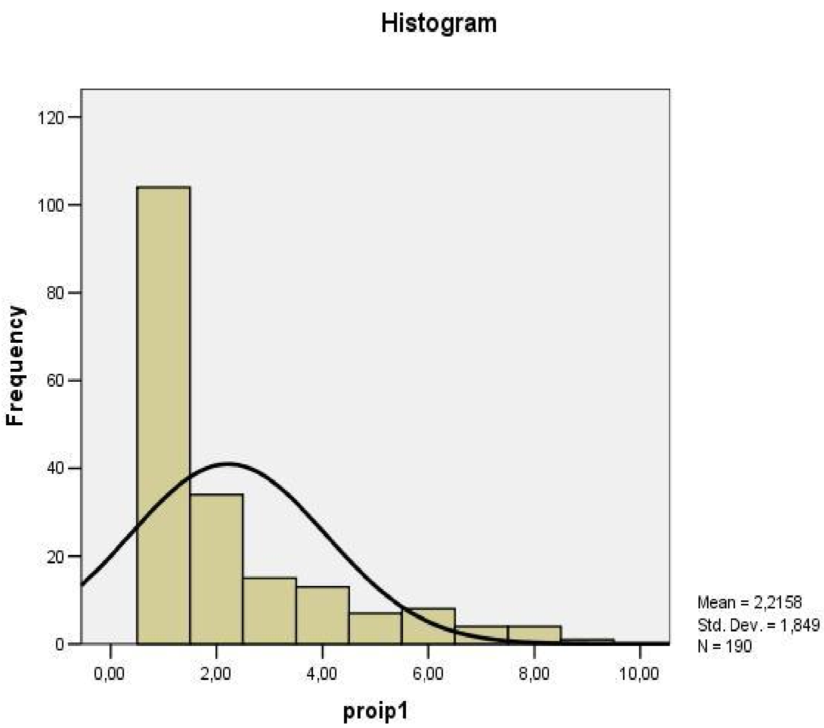
9^{η}) 36 --> 40,5

ΣΧΟΛΙΑΣΜΟΣ ΠΙΝΑΚΑ proipiresia

proip1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	104	54,7	54,7	54,7
	2,00	34	17,9	17,9	72,6
	3,00	15	7,9	7,9	80,5
	4,00	13	6,8	6,8	87,4
	5,00	7	3,7	3,7	91,1
	6,00	8	4,2	4,2	95,3
	7,00	4	2,1	2,1	97,4
	8,00	4	2,1	2,1	99,5
	9,00	1	,5	,5	100,0
	Total	190	100,0	100,0	

Στον πίνακα βλέπουμε τις συχνότητες (στήλη frequency) και τα ποσοστά (στήλη percent) κάθε τάξης.Τα στοιχειά του πίνακα μπορούμε να τα διακρίνουμε και στο παρακάτω ιστόγραμμα:



Για την κατασκευή αθροιστικού διαγράμματος βρίσκουμε τις κεντρικές τιμές κάθε τάξης.

ΚΕΝΤΡΙΚΗ ΤΙΜΗ

1η) 0 --> 4,5=2,25

2η) 4,5 --> 9=6,75

3η) 9 --> 13,5=11,25

4η) 13,5--> 18=15,75

5η) 18 --> 22,5= 20,25

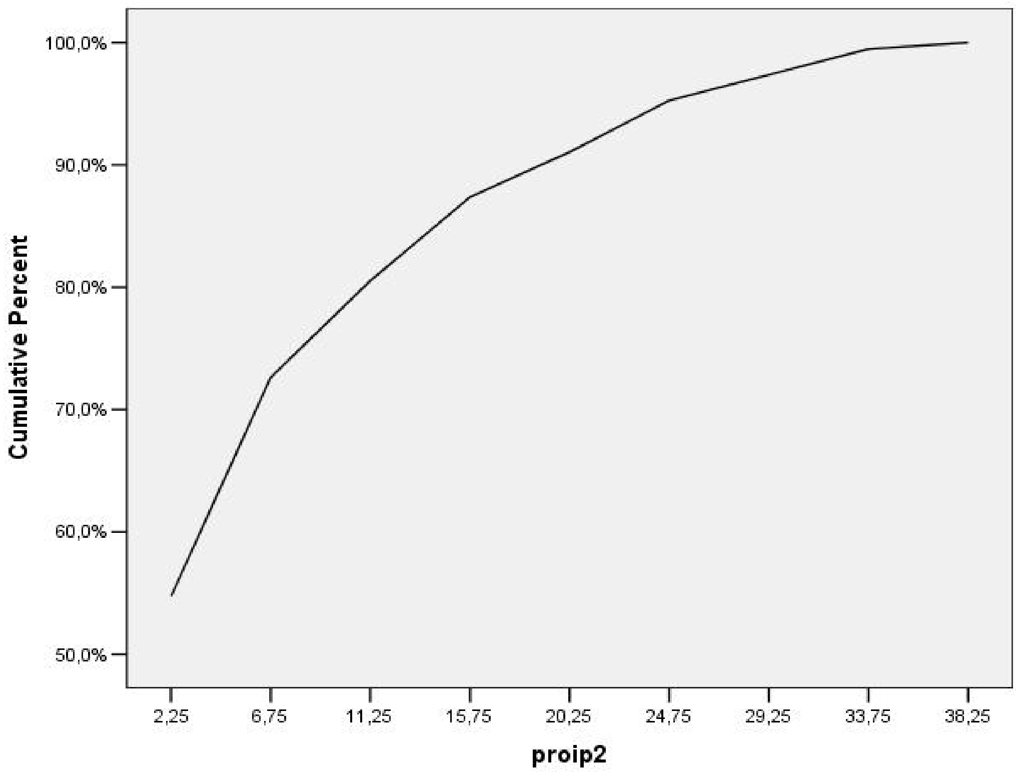
6η) 22,5--> 27=24,75

7η) 27 --> 31,5 =29,25

8η) 31,5--> 36 = 33,75

9η) 36 --> 40,5=38,25

Στηνσυνέχεια, ακλουθούμετηνεξήςδιαδικασία: graphs ->line ->simple ->summariesof group of cases -> define -> category axis: proip2 ->like represents:cum. % of cases ->ok



Επίσης, για την μεταβλητή αυτή υπολογίζουμε τα παρακάτω

Statistics

xronos proip.

N	Valid	190
	Missing	0
Mean		7,4497
Median		4,0850
Mode		,00
Range		36,50

ΣΧΟΛΙΑΣΜΟΣ ΠΙΝΑΚΑ

- Ο αριθμητικός μέσος μας δείχνει τον μέσο όρο της Προϋπηρεσίας των 190 καταχωρήσεων.
- Η διάμεσος μας δείχνει ότι το 50% των παρατηρήσεων είναι πάνω από το (μ) και το υπόλοιπο 50% κάτω από αυτή την τιμή.
- Η επικρατούσα τιμή μας δείχνει ότι η πιο συνήθης προϋπηρεσία είναι η (μ_0).
- Το εύρος μεταβολής μας δείχνει την διάφορα της μεγαλύτερης Προϋπηρεσίας από την μικρότερη.

ΕΠΙΔΟΣΗ

Επειδή η μεταβλητή είναι συνεχής, την ταξινομούμε σε τάξεις: Προσδιορίζουμε τον αριθμό των τάξεων με τον κανόνα του Sturges.

$$K = 1 + 3.322 log n$$

(n=190)

Μετά από τον υπολογισμό του k, το αποτέλεσμα ισούται με $k = 8.57 \approx 9$ τάξεις.

Επιλέγουμε το πλάτος των τάξεων από τον τύπο $\delta = r/k$.

Statistics

mesi bathmologia epidosis

N	Valid	190
	Missing	0
Range		34
Minimum		64
Maximum		98

R= εύρος μεταβολης=34

Οπότε $\delta = 34/9 = 3{,}7 \approx 4$

Ταξινομούμε τις τάξεις σύμφωνα με τα παραπάνω:

$1^{η}$) 64 --> 68

$2^{η}$) 68 --> 72

$3^{η}$) 72 --> 76

$4^{η}$) 76 --> 80

$5^{η}$) 80 --> 84

$6^{η}$) 84 --> 88

$7^{η}$) 88 --> 92

$8^{η}$) 92 --> 96

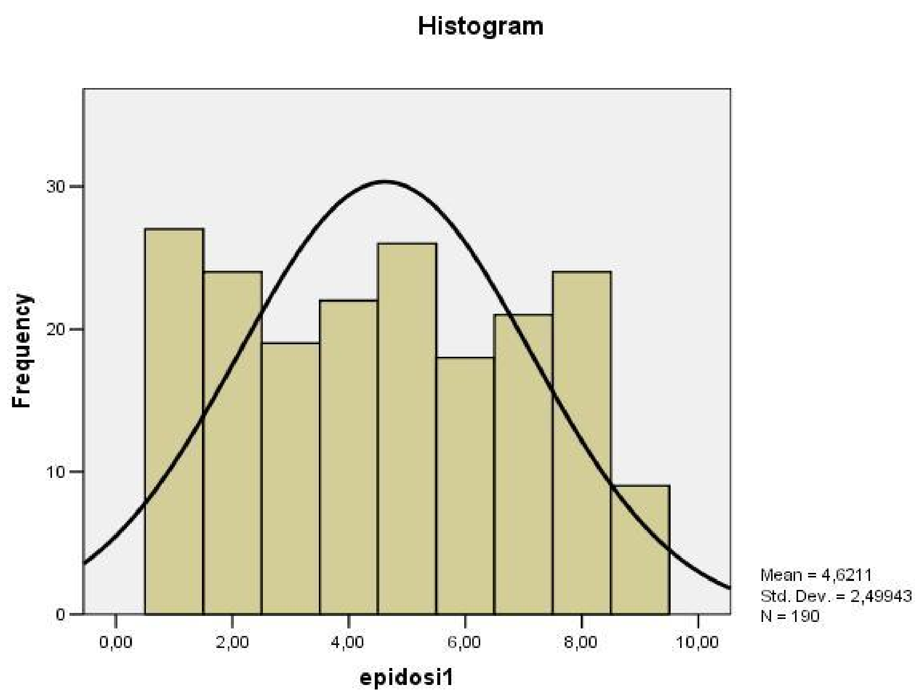
$9^{η}$) 96 --> 100

ΣΧΟΛΙΑΣΜΟΣ ΠΙΝΑΚΑ ΕΠΙΔΟΣΗ

epidosi1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	27	14,2	14,2	14,2
	2,00	24	12,6	12,6	26,8
	3,00	19	10,0	10,0	36,8
	4,00	22	11,6	11,6	48,4
	5,00	26	13,7	13,7	62,1
	6,00	18	9,5	9,5	71,6
	7,00	21	11,1	11,1	82,6
	8,00	24	12,6	12,6	95,3
	9,00	9	4,7	4,7	100,0
	Total	190	100,0	100,0	

Στον πίνακα βλέπουμε τις συχνότητες (στήλη frequency) και τα ποσοστά (στήλη percent) κάθε τάξης.Τα στοιχειά του πίνακα μπορούμε να τα διακρίνουμε και στο παρακάτω ιστόγραμμα:

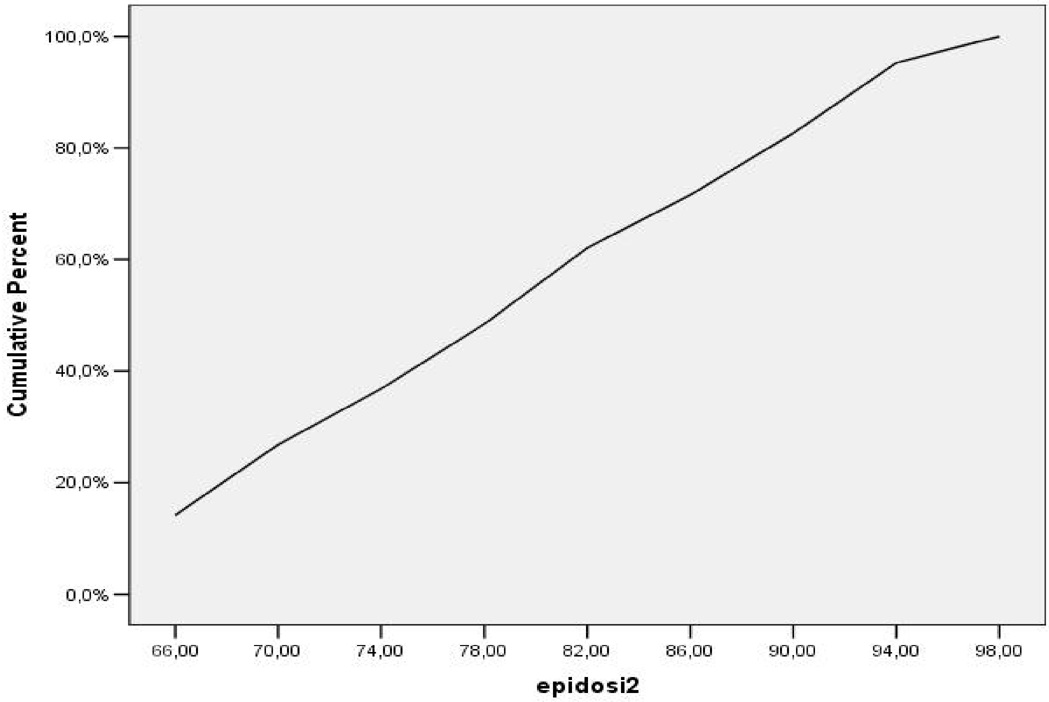


Για την κατασκευή αθροιστικού διαγράμματος βρίσκουμε τις κεντρικές τιμές κάθε τάξης:

ΚΕΝΤΡΙΚΗ ΤΙΜΗ

1η) 64	-->	68	=	66
2η) 68	-->	72	=	70
3η) 72	-->	76	=	74
4η) 76	-->	80	=	78
5η) 80	-->	84	=	82
6η) 84	-->	88	=	86
7η) 88	-->	92	=	90
8η) 92	-->	96	=	94
9η) 96	-->	100	=	98

Στηνσυνέχεια, ακλουθούμετηνεξήςδιαδικασία: graphs -> line -> simple -> summaries of group of cases -> define -> category axis:epidosi2 ->like represents:cum. % of cases ->ok



Επίσης, για την μεταβλητή αυτή υπολογίζουμε τα παρακάτω:

Statistics

mesi bathmologia epidosis

N	Valid	190
	Missing	0
Mean		80,73
Median		81,00
Mode		69
Range		34

ΣΧΟΛΙΑΣΜΟΣ ΠΙΝΑΚΑ

- Ο αριθμητικός μέσος μας δείχνει τον μέσο όρο της επίδοσης των 190 καταχωρήσεων.
- Η διάμεσος μας δείχνει ότι το 50% των παρατηρήσεων είναι πάνω από το (μ) και το υπόλοιπο 50% κάτω από αυτή την τιμή
- Η επικρατούσα τιμή μας δείχνει ότι η πιο συνήθης επίδοση είναι η (μ_0).
- Το εύρος μεταβολής μας δείχνει την διάφορα της μεγαλύτερης επίδοσης από την μικρότερη.

ΑΡΧΙΚΟΣ ΜΙΣΘΟΣ

Επειδή η μεταβλητή είναι συνεχής, την ταξινομούμε σε τάξεις.Προσδιορίζουμε τον αριθμό των τάξεων με τον κανόνα του Sturges.

$K = 1 + 3.322 log n$, (n=190)

Μετά από τον υπολογισμό του k, το αποτέλεσμα ισούται με $k = 8.57 \approx 9$ τάξεις.

Επιλέγουμε το πλάτος των τάξεων από τον τύπο $\delta = r/k$.

Statistics

archikos misthos

N	Valid	190
	Missing	0
Range		16855
Minimum		2340
Maximum		19195

R= εύρος μεταβολης=16855

Οπότε $\delta = 16855/9 = 1872{,}7 \approx 1873$

-Ταξινομούμε τις τάξεις σύμφωνα με τα παραπάνω:

$1^{η}$) 2340 --> 4213

$2^{η}$) 4213 --> 6086

$3^{η}$) 6086 --> 7959

$4^{η}$) 7959 --> 9832

$5^{η}$) 9832 --> 11705

$6^{η}$) 11705 --> 13578

$7^{η}$) 13578 --> 15451

$8^{η}$) 15451 --> 17324

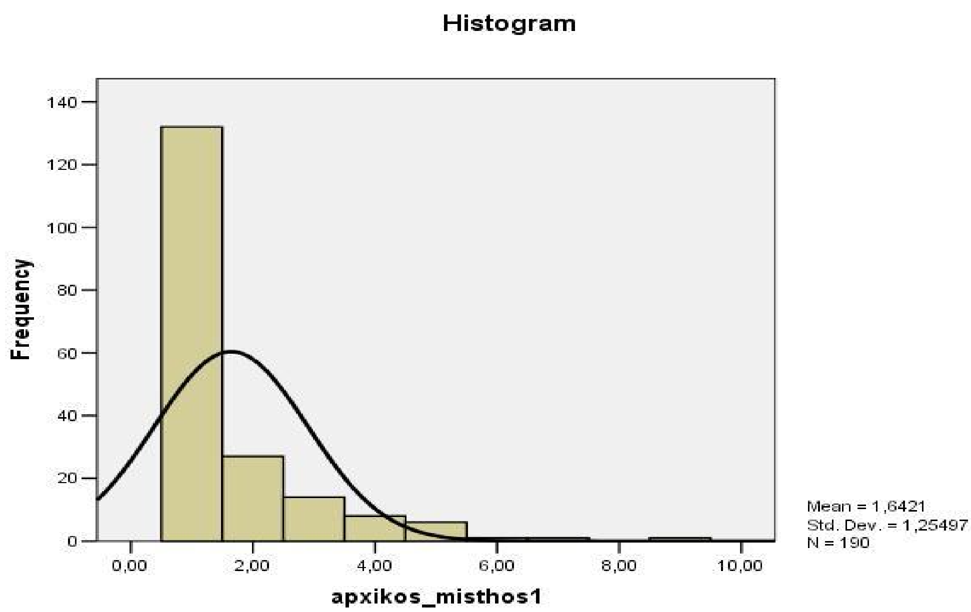
$9^{η}$) 17324 --> 19197

<u>ΣΧΟΛΙΑΣΜΟΣ ΠΙΝΑΚΑ ΑΡΧΙΚΟΥ ΜΙΣΘΟΥ</u>

arxikos_misthos1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	132	69,5	69,5	69,5
	2,00	27	14,2	14,2	83,7
	3,00	14	7,4	7,4	91,1
	4,00	8	4,2	4,2	95,3
	5,00	6	3,2	3,2	98,4
	6,00	1	,5	,5	98,9
	7,00	1	,5	,5	99,5
	9,00	1	,5	,5	100,0
	Total	190	100,0	100,0	

Στον πίνακα βλέπουμε τις συχνότητες (στήλη frequency) και τα ποσοστά (στήλη percent) κάθε τάξης.Τα στοιχειά του πίνακα μπορούμε να τα διακρίνουμε και στο παρακάτω ιστόγραμμα:



Για την κατασκευή αθροιστικού διαγράμματος βρίσκουμε τις κεντρικές τιμές κάθε τάξης:

ΚΕΝΤΡΙΚΗ ΤΙΜΗ

1η) 2340-->4213= 3276,5

2η) 4213 -->6086= 5149,5

3η) 6086 -->7959= 7022,5

4η) 7959--> 9832=8895,5

5η) 9832 -->11705=10768,5

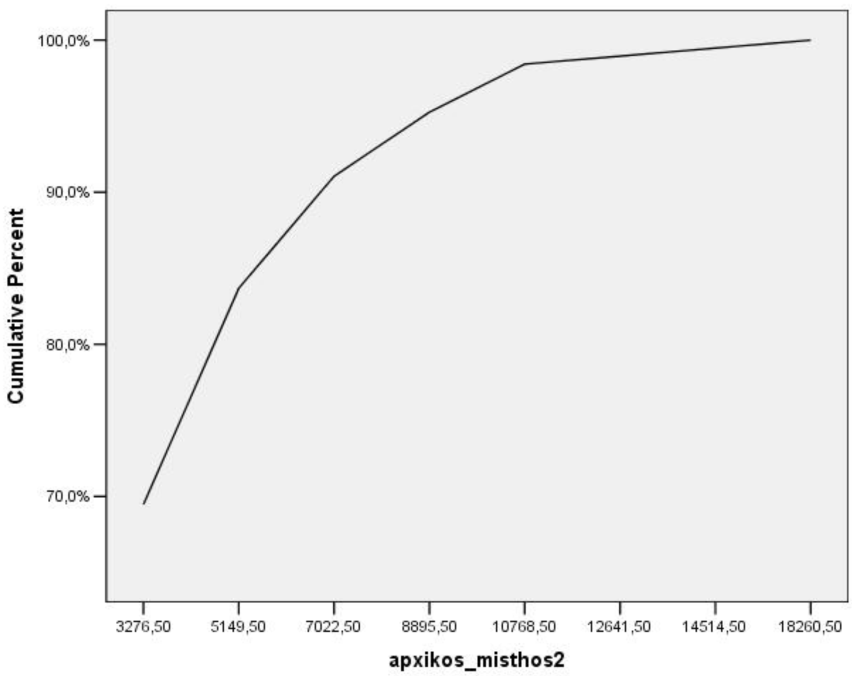
6η) 11705-->13578=12641,5

7η) 13578-->15451=14514,5

8η) 15451-->17324=16387,5

9η) 17324-->19197=18260,5

Στηνσυνέχεια, ακλουθούμετηνεξήςδιαδικασία: graphs -> line -> simple -> summaries of group of cases -> define -> category axis:arxikos_misthos2 ->like represents:cum. % ofcases ->ok



Επίσης, για την μεταβλητή αυτή υπολογίζουμε τα παρακάτω:

Statistics

archikos misthos

N	Valid	190
	Missing	0
Mean		4445,28
Median		3600,00
Mode		3780
Range		16855

<u>ΣΧΟΛΙΑΣΜΟΣ ΠΙΝΑΚΑ</u>

- Ο αριθμητικός μέσος μας δείχνει τον μέσο όρο του αρχικού μισθού των 190 καταχωρήσεων.
- Η διάμεσος μας δείχνει ότι το 50% των παρατηρήσεων είναι πάνω από το (μ) και το υπόλοιπο 50% κάτω από αυτή την τιμή.
- Η επικρατούσα τιμή μας δείχνει ότι ο πιο συνήθης αρχικός Μισθός είναι ο (μ_0).
- Το εύρος μεταβολής μας δείχνει την διάφορα του μεγαλυτέρου αρχικού μισθού από τον μικρότερο.

<u>ΣΗΜΕΡΙΝΟΣ ΜΙΣΘΟΣ</u>

Επειδή η μεταβλητή είναι συνεχής, την ταξινομούμε σε τάξεις.Προσδιορίζουμε τον αριθμό των τάξεων με τον κανόνα του Sturges.

$K = 1 + 3.322logn$, (n=190)

Μετά από τον υπολογισμό του k, το αποτέλεσμα ισούται με $k = 8.57 \approx 9$ τάξεις.

Επιλέγουμε το πλάτος των τάξεων από τον τύπο $\delta = r/k$.

Statistics

simerinos misthos

N	Valid	190
	Missing	0
Range		28620
Minimum		3780
Maximum		32400

R=εύρος μεταβολης=28620

Οπότε $\delta = 28620/9 = 3180$

-Ταξινομούμε τις τάξεις σύμφωνα με τα παραπάνω:

1^{η}) 3780 --> 6960

2^{η}) 6960 --> 10140

3^{η}) 10140 --> 13320

4^{η}) 13320 --> 16500

5^{η}) 16500 --> 19680

6^{η}) 19680 --> 22860

7^{η}) 22860 --> 26040

8^{η}) 26040 --> 29220

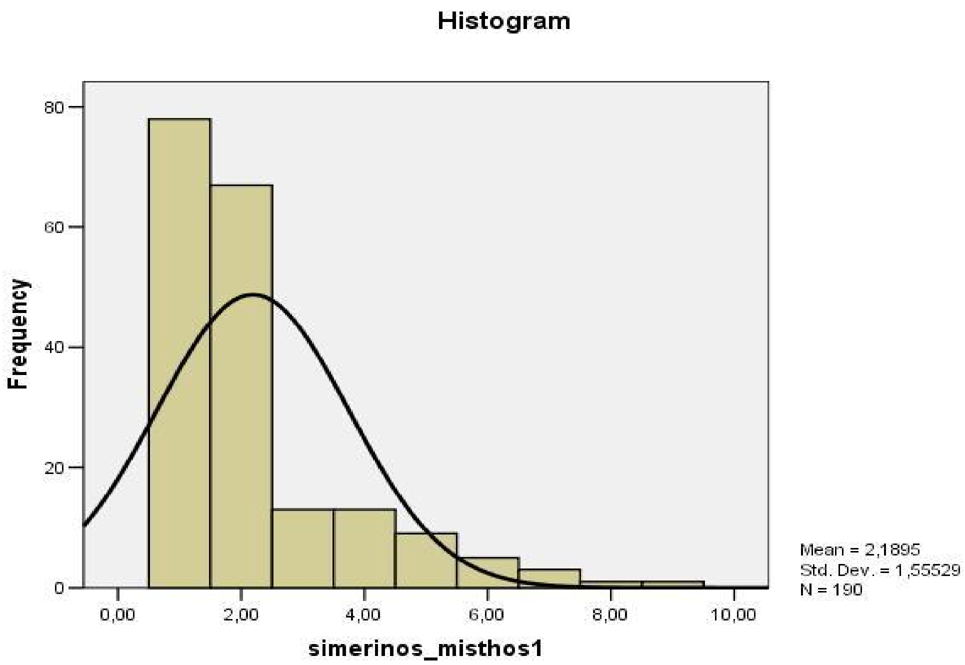
9^{η}) 29220 --> 32400

ΣΧΟΛΙΑΣΜΟΣ ΠΙΝΑΚΑ ΣΗΜΕΡΙΝΟΥ ΜΙΣΘΟΥ

simerinos_misthos1

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1,00	78	41,1	41,1	41,1
2,00	67	35,3	35,3	76,3
3,00	13	6,8	6,8	83,2
4,00	13	6,8	6,8	90,0
5,00	9	4,7	4,7	94,7
6,00	5	2,6	2,6	97,4
7,00	3	1,6	1,6	98,9
8,00	1	,5	,5	99,5
9,00	1	,5	,5	100,0
Total	190	100,0	100,0	

Στον πίνακα βλέπουμε τις συχνότητες (στήλη frequency) και τα ποσοστά (στήλη percent) κάθε τάξης.Τα στοιχειά του πίνακα μπορούμε να τα διακρίνουμε και στο παρακάτω ιστόγραμμα:



Για την κατασκευή αθροιστικού διαγράμματος βρίσκουμε τις κεντρικές τιμές κάθε τάξης:

ΚΕΝΤΡΙΚΗ ΤΙΜΗ

1$^{\eta}$) 3780 --> 6960=5370

2$^{\eta}$) 6960 --> 10140=8550

3$^{\eta}$) 10140 --> 13320=11730

4$^{\eta}$) 13320 --> 1650014910

5$^{\eta}$) 16500 --> 19680=18090

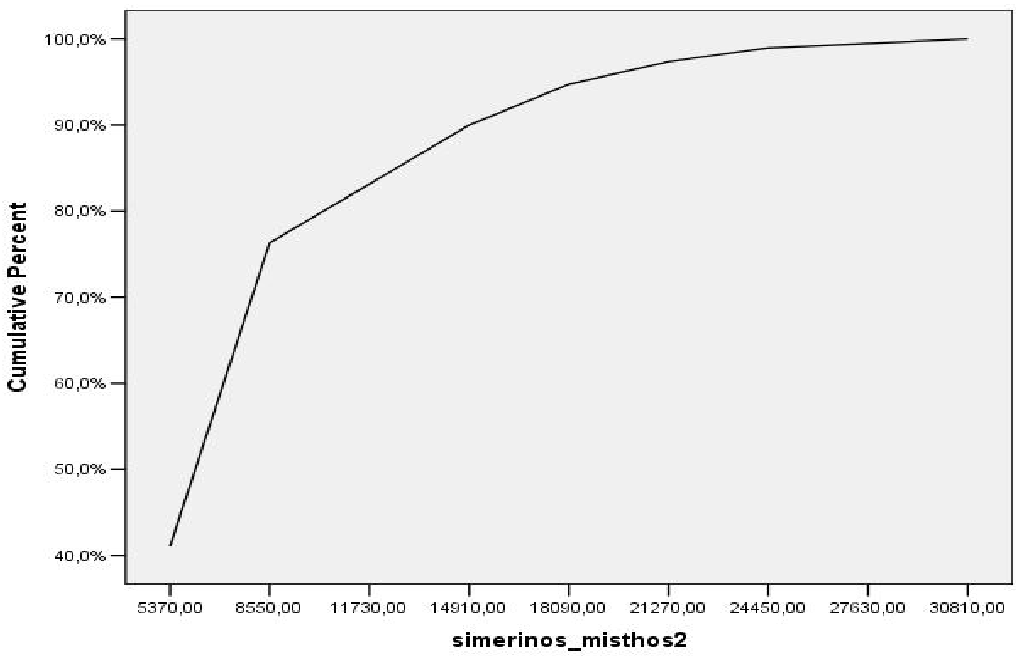
6$^{\eta}$) 19680 --> 22860=21270

7$^{\eta}$) 22860 --> 26040=24450

8$^{\eta}$) 26040 --> 29220=27630

9$^{\eta}$) 29220 --> 32400 = 30810

Στηνσυνέχεια, ακλουθούμετηνεξήςδιαδικασία: graphs -> line -> simple -> summaries of group of cases -> define -> category axis:simerinos_misthos2 ->like represents:cum. % of cases ->ok



Επίσης, για την μεταβλητή αυτή υπολογίζουμε τα παρακάτω

Statistics

simerinos misthos

N	Valid	190
	Missing	0
Mean		9062,90
Median		7380,00
Mode		5760
Range		28620

ΣΧΟΛΙΑΣΜΟΣ ΠΙΝΑΚΑ

- Ο αριθμητικός μέσος μας δείχνει τον μέσο όρο του σημερινού μισθού των 190 καταχωρήσεων.
- Η διάμεσος μας δείχνει ότι το 50% των παρατηρήσεων είναι πάνω από το (μ) και το υπόλοιπο 50% κάτω από αυτή την τιμή.
- Η επικρατούσα τιμή μας δείχνει ότι ο πιο συνήθης Σημερινός Μισθός είναι ο (μ0).
- Το εύρος μεταβολής μας δείχνει την διάφορα του μεγαλυτέρου σημερινού μισθού από τον μικρότερο.

ΠΑΛΙΝΔΡΟΜΗΣΗ ΚΑΙ ΔΙΑΓΡΑΜΜΑ ΔΙΑΣΠΟΡΑΣ

1)Διαγραμμα Διασποράς

2)Υπολογισμος Εξίσωσης Παλινδρόμησης

3)Υπολογισμος Δείκτη Προσδιορισμού

Μεταβλητές:

Αρχικός Μισθός (εξαρτημένη μεταβλητή)

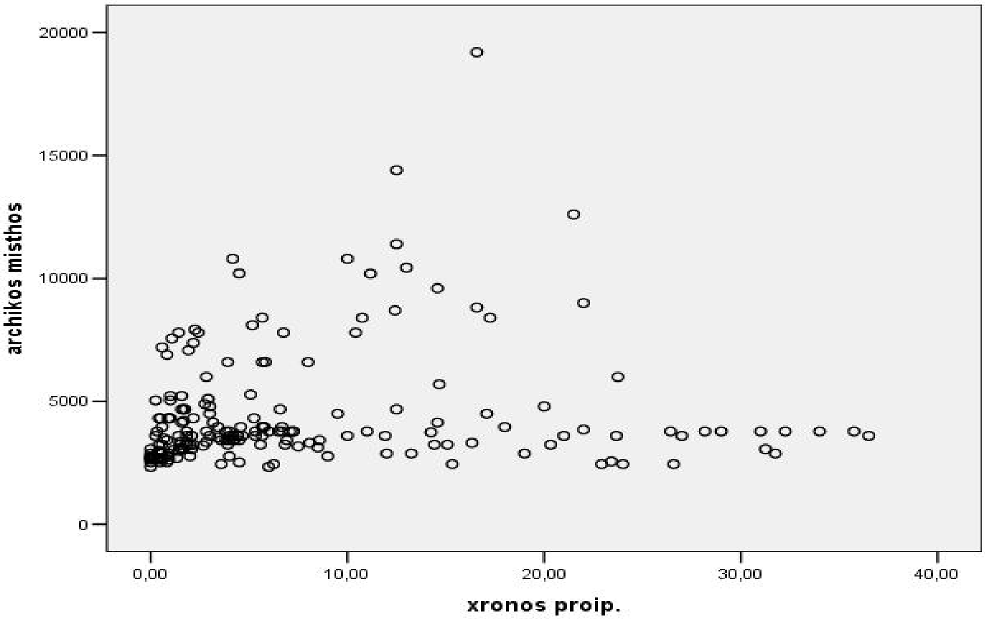
Χρόνος Προϋπηρεσίας (ανεξάρτητη μεταβλητή)

1)Υπολογίζουμε το διάγραμμα διασποράς με την εξής διαδικασία:

Graphs-->scater-->simple-->define-->y_{axis}=αρχικός Μισθός—>x_{axis}=Χρόνος Προϋπηρεσίας-->οκ

Απάντηση:

Το διάγραμμα φανερώνει μια ισχυρή γραμμική συσχέτιση.



2)Υπολογίζουμε την εξίσωση παλινδρόμησης με την εξής διαδικασία: Analyze-->regression-->linear-->dependent= αρχικόςΜισθός -->Independent= ΧρόνοςΠροϋπηρεσίας-->ok

ModelSummary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,130 (a)	,017	,012	2390,750

a Predictors: (Constant), xronos proip.

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	18513403,441	1	18513403,441	3,239	,074 (a)
	Residual	1074549284,774	188	5715687,685		
	Total	1093062688,216	189			

a Predictors: (Constant), xronos proip.

b Dependent Variable: archikos misthos

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4174,603	229,569		18,185	,000
	xronos proip.	36,334	20,188	,130	1,800	,074

a Dependent Variable: archikos misthos

Εξίσωση Παλινδρόμησης: $y = ax + b$

Από τον πίνακα: Coefficients βρίσκουμε ότι το

b=4174,603

a=36,334

Επομένως: $y = 36{,}334x + 4174{,}603$

όπου a και b είναι παράμετροι τις οποίες θέλουμε να υπολογίσουμε ή, όπως λέμε, να “εκτιμήσουμε”, έτσι ώστε η ευθεία που θα προκύψει να μας δίνει όσο το δυνατόν την καλύτερη περιγραφή της σχέσης (εξάρτησης) που υπάρχει μεταξύ των μεταβλητών X (Χρόνος Προϋπηρεσίας) και Y(αρχικός Μισθός).

Η παράμετρος b μας δίνει το σημείο (0,b), όπου η ευθεία αυτή τέμνει τον άξονα y'y, ενώ η παράμετρος a παριστάνει το συντελεστή διεύθυνσης της ευθείας.

3)Απο τον πίνακα Model Summary υπολογίζουμε τον δείκτη προσδιορισμού:

R^2=R Square=0,017

Δηλαδή κατά 1,7% η μεταβλητότητα του Αρχικού Μισθού ερμηνεύεται από τον Χρόνο Προϋπηρεσίας.

Μεταβλητές:

Σημερινός Μισθός (εξαρτημένη μεταβλητή)

Μέση Βαθμολογία Επίδοσης (ανεξάρτητη μεταβλητή)

1)Υπολογίζουμε το διάγραμμα διασποράς με την εξής διαδικασία:

Graphs-->scater-->simple-->define-->y_{axis}= ΣημερινόςΜισθός-->x_{axis}=Επίδοση-->οκ

ΑΠΑΝΤΗΣΗ:

Το διάγραμμα φανερώνει μια ισχυρή γραμμική συσχέτιση.

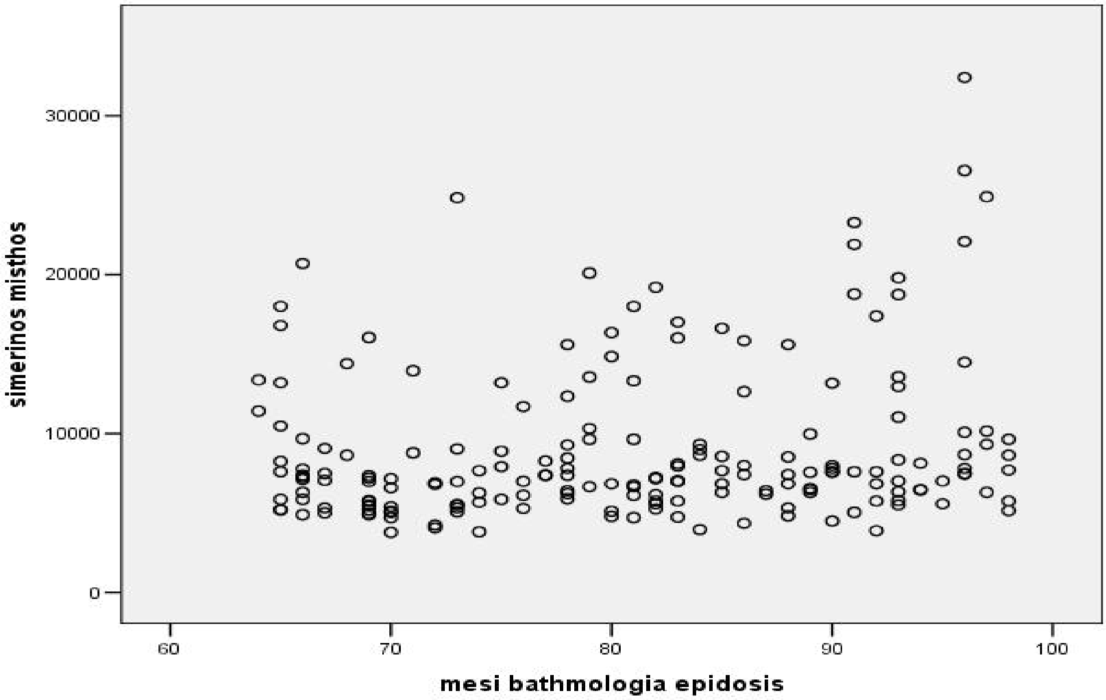
2)Υπολογίζουμε την εξίσωση παλινδρόμησης με την εξής διαδικασία:

Analyze-->Regression-->Linear-->Dependent= ΣημερινόςΜισθός -->Independent= Επίδοση-->ok

ModelSummary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,183(a)	,033	,028	4887,181

a Predictors: (Constant), mesi bathmologia epidosis



ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	155498795,627	1	155498795,627	6,510	,012(a)
	Residual	4490292667,474	188	23884535,465		
	Total	4645791463,100	189			

a Predictors: (Constant), mesi bathmologia epidosis

b Dependent Variable: simerinos misthos

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1872,131	2840,405		,659	,511
	mesi bathmologia epidosis	89,076	34,910	,183	2,552	,012

a Dependent Variable: simerinos misthos

ΕξίσωσηΠαλινδρόμησης:

Από τον πίνακα: Coefficients βρίσκουμε ότι το

b=1872,131

a=89,076

Επομένως: $y = 89{,}076x + 1872{,}131$

όπου a και b είναι παράμετροι τις οποίες θέλουμε να υπολογίσουμε ή, όπως λέμε, να "εκτιμήσουμε", έτσι ώστε η ευθεία που θα προκύψει να μας δίνει όσο το δυνατόν την καλύτερη περιγραφή της σχέσης (εξάρτησης) που υπάρχει μεταξύ των μεταβλητών X (Μέση Βαθμολογία Επίδοσης) και Y(Σημερινός Μισθός).

Η παράμετρος b μας δίνει το σημείο (0,b), όπου η ευθεία αυτή τέμνει τον άξονα y'y, ενώ η παράμετρος a παριστάνει το συντελεστή διεύθυνσης της ευθείας.

3)Απο τον πίνακα Model Summary υπολογίζουμε τον δείκτη προσδιορισμού:

R^2=R Square=0,033

Δηλαδή κατά 3,3% η μεταβλητότητα του Σημερινού Μισθού ερμηνεύεται από την Επίδοση.

Μεταβλητές:

Αρχικός Μισθός (εξαρτημένη μεταβλητή)

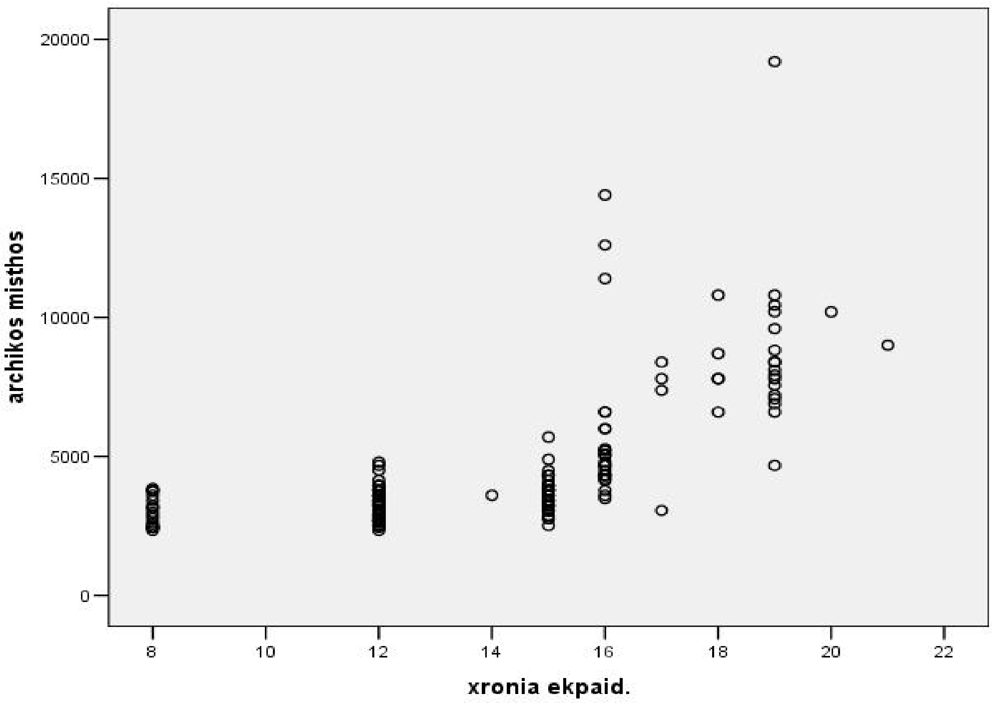
Χρόνια Εκπαίδευσης (ανεξάρτητη μεταβλητή)

1)Υπολογίζουμε το διάγραμμα διασποράς με την εξής διαδικασία:

Graphs-->scater-->simple-->define-->y_{axis}= Αρχικός Μισθός-->x_{axis}= Χρόνια Εκπαίδευσης-->οκ

ΑΠΑΝΤΗΣΗ:

Το διάγραμμα φανερώνει μια ισχυρή γραμμική συσχέτιση.



2)Υπολογίζουμε την εξίσωση παλινδρόμησης με την εξής διαδικασία:

Analyze-->Regression-->Linear-->Dependent= ΑρχικόςΜισθός -->Independent= ΧρόνιαΕκπαίδευσης-->ok

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,651(a)	,423	,420	1831,099

a Predictors: (Constant), xronia ekpaid.

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	462713055,220	1	462713055,220	138,003	,000(a)
	Residual	630349632,996	188	3352923,580		
	Total	1093062688,216	189			

a Predictors: (Constant), xronia ekpaid.

b Dependent Variable: archikos misthos

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2658,518	619,128		-4,294	,000
	xronia ekpaid.	511,840	43,570	,651	11,747	,000

a Dependent Variable: archikos misthos

Εξίσωση Παλινδρόμησης: $y = ax + b$

Από τον πίνακα: Coefficients βρίσκουμε ότι το

b=-2658,518

a=511,840

Επομένως: $y = 511{,}840x - 2658{,}518$

όπου a και b είναι παράμετροι τις οποίες θέλουμε να υπολογίσουμε ή, όπως λέμε, να "εκτιμήσουμε", έτσι ώστε η ευθεία που θα προκύψει να μας δίνει όσο το δυνατόν την καλύτερη περιγραφή της σχέσης (εξάρτησης) που υπάρχει μεταξύ των μεταβλητών X (ΧρόνιαΕκπαίδευσης) και Y(ΑρχικόςΜισθός).

Η παράμετρος b μας δίνει το σημείο (0,b), όπου η ευθεία αυτή τέμνει τον άξονα y'y, ενώ η παράμετρος a παριστάνει το συντελεστή διεύθυνσης της ευθείας..

3)απο τον πίνακα Model Summary υπολογίζουμε τον δείκτη προσδιορισμού:

R^2=R Square=0,423

Δηλαδή κατά 42,3% η μεταβλητότητα του αρχικού μισθού ερμηνεύεται από τα χρόνια εκπαίδευσης.

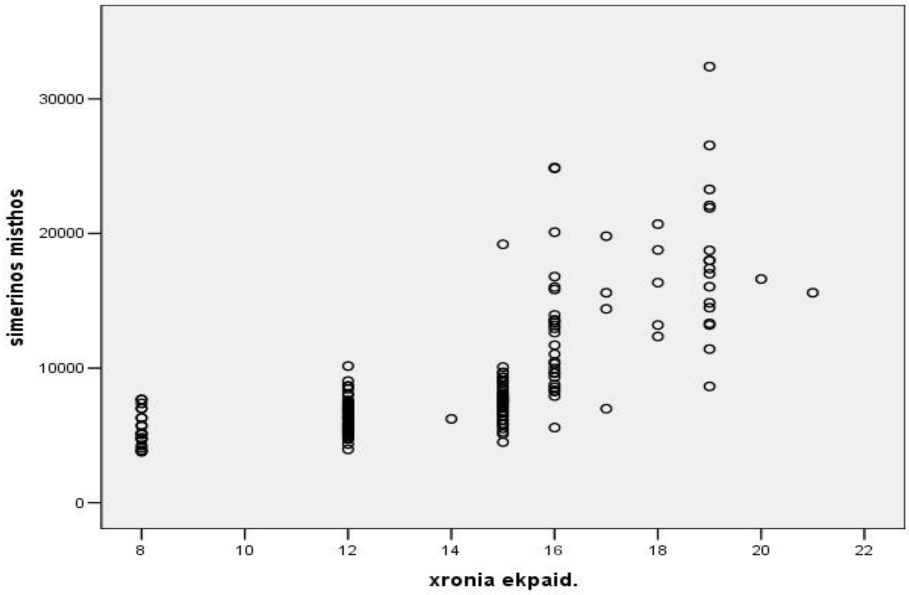
ΜΕΤΑΒΛΗΤΕΣ:

Σημερινός Μισθός (εξαρτημένη μεταβλητή)

Χρόνια Εκπαίδευσης (ανεξάρτητη μεταβλητή)

1)Υπολογίζουμε το διάγραμμα διασποράς με την εξής διαδικασία:

Graphs-->Scater-->Simple-->Define-->y_{axis}= Σημερινός Μισθός-->x_{axis}=χρονια εκπαίδευσης--> οκ



<u>ΑΠΑΝΤΗΣΗ:</u>

Το διάγραμμα φανερώνει μια ισχυρή γραμμική συσχέτιση.

2)Υπολογίζουμε την εξίσωση παλινδρόμησης με την εξής διαδικασία: Analyze-->Regression-->Linear-->Dependent= ΣημερινόςΜισθός -->Independent= χρόνιαεκπαίδευσης-->ok

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,693(a)	,480	,477	3585,552

a Predictors: (Constant), xronia ekpaid.

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2228828504,555	1	2228828504,555	173,366	,000(a)
	Residual	2416962958,545	188	12856185,950		
	Total	4645791463,100	189			

a Predictors: (Constant), xronia ekpaid.

b Dependent Variable: simerinos misthos

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-6528,064	1212,342		-5,385	,000
	xronia ekpaid.	1123,354	85,317	,693	13,167	,000

a Dependent Variable: simerinos misthos

Εξίσωση Παλινδρόμησης: $y = ax + b$

Από τον πίνακα: Coefficients βρίσκουμε ότι το

b=-6528,064

a=1123,354

Επομένως: $y = 1123{,}354x - 6528{,}064$

όπου a και b είναι παράμετροι τις οποίες θέλουμε να υπολογίσουμε ή, όπως λέμε, να “εκτιμήσουμε”, έτσι ώστε η ευθεία που θα προκύψει να μας δίνει όσο το δυνατόν την καλύτερη περιγραφή της σχέσης (εξάρτησης) που υπάρχει μεταξύ των μεταβλητών X (ΧρόνιαΕκπαίδευσης) και Y(ΣημερινόςΜισθός).

Η παράμετρος b μας δίνει το σημείο (0,b), όπου η ευθεία αυτή τέμνει τον άξονα y'y, ενώ η παράμετρος a παριστάνει το συντελεστή διεύθυνσης της ευθείας.

3)Απο τον πίνακα Model Summary υπολογίζουμε τον δείκτη προσδιορισμού:

R^2=R Square=0,480

Δηλαδή κατά 48% η μεταβλητότητα του σημερινού μισθού ερμηνεύεται από τα χρόνια εκπαίδευσης.

Μεταβλητές:

Σημερινός Μισθός (εξαρτημένη μεταβλητή)

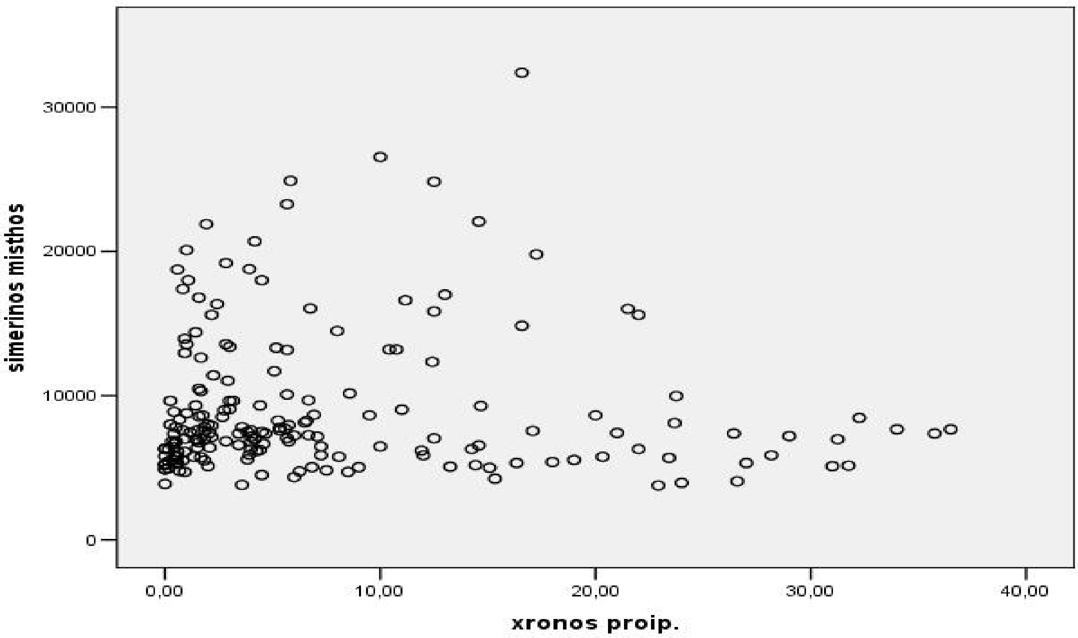
Χρόνος Προϋπηρεσίας (ανεξάρτητη μεταβλητή)

1)Υπολογίζουμε το διάγραμμα διασποράς με την εξής διαδικασία:

Graphs-->Scater-->Simple-->Define-->y_{axis}= Σημερινός Μισθός-->x_{axis}=Χρόνος Προϋπηρεσίας--> οκ

ΑΠΑΝΤΗΣΗ:

Το διάγραμμα φανερώνει μια ισχυρή γραμμική συσχέτιση.



2)Υπολογίζουμε την εξίσωση παλινδρόμησης με την εξής διαδικασία:

Analyze-->Regression-->Linear-->Dependent= ΣημερινόςΜισθός -->Independent= ΧρόνοςΠροϋπηρεσίας-->ok

ModelSummary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,018(a)	,000	-,005	4970,295

a Predictors: (Constant), xronos proip.

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1470883,924	1	1470883,924	,060	,807(a)
	Residual	4644320579,177	188	24703832,868		
	Total	4645791463,100	189			

a Predictors: (Constant), xronos proip.

b Dependent Variable: simerinos misthos

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	9139,195	477,267		19,149	,000
	xronos proip.	-10,241	41,971	-,018	-,244	,807

a Dependent Variable: simerinos misthos

Εξίσωση Παλινδρόμησης: $y = ax + b$

όπου a και b είναι παράμετροι τις οποίες θέλουμε να υπολογίσουμε ή, όπως λέμε, να "εκτιμήσουμε", έτσι ώστε η ευθεία που θα προκύψει να μας δίνει όσο το δυνατόν την καλύτερη περιγραφή της σχέσης (εξάρτησης) που υπάρχει μεταξύ των μεταβλητών X (ΧρόνοςΠροϋπηρεσίας) και Y(ΣημερινόςΜισθός).

Η παράμετρος b μας δίνει το σημείο (0,b), όπου η ευθεία αυτή τέμνει τον άξονα y'y, ενώ η παράμετρος a παριστάνει το συντελεστή διεύθυνσης της ευθείας.

Από τον πίνακα: Coefficients βρίσκουμε ότι το

b=9139,195

a=-10,241

Επομένως: $y = -10{,}241x + 9139{,}195$

3)Απο τον πίνακα Model Summary υπολογίζουμε τον δείκτη προσδιορισμού:

R^2=R Square=0,00

Επειδή $R^2 = 0{,}00$ βγάζουμε το συμπέρασμα ότι η μεταβλητότητα του σημερινού μισθού δεν ερμηνεύεται από τον χρόνο Προϋπηρεσίας.

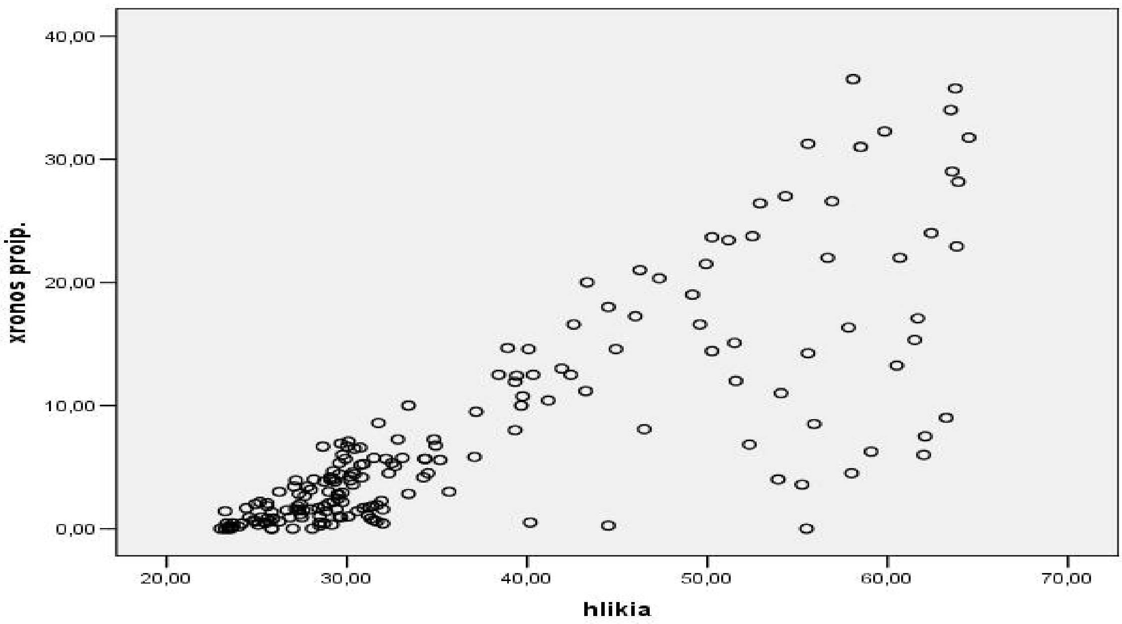
<u>ΜΕΤΑΒΛΗΤΕΣ:</u>

ΧρόνοςΠροϋπηρεσίας(εξαρτημένη μεταβλητή)

Ηλικία(ανεξάρτητη μεταβλητή)

1)Υπολογίζουμε το διάγραμμα διασποράς με την εξής διαδικασία:

Graphs-->Scater-->Simple-->Define-->y_{axis}= ΧρόνοςΠροϋπηρεσίας-->x_{axis}=ηλικία-->οκ



ΑΠΑΝΤΗΣΗ:

Το διάγραμμα φανερώνει μια ισχυρή γραμμική συσχέτιση.

2)Υπολογίζουμε την εξίσωση παλινδρόμησης με την εξής διαδικασία: Analyze-->Regression-->Linear-->Dependent= ΧρόνοςΠροϋπηρεσίας -->Independent=ηλικία--> ok

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,808(a)	,653	,651	5,08663

a Predictors: (Constant), hlikia

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	9159,582	1	9159,582	354,009	,000(a)
	Residual	4864,281	188	25,874		
	Total	14023,864	189			

a Predictors: (Constant), hlikia

b Dependent Variable: xronos proip.

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-13,555	1,176		-11,529	,000
	hlikia	,574	,030	,808	18,815	,000

a Dependent Variable: xronos proip.

Εξίσωση Παλινδρόμησης: $y = ax + b$

Από τον πίνακα: Coefficients βρίσκουμε ότι το

b=-13,555

a=0,574

Επομένως: $y = 0{,}574x - 13{,}555$

όπου a και b είναι παράμετροι τις οποίες θέλουμε να υπολογίσουμε ή, όπως λέμε, να "εκτιμήσουμε", έτσι ώστε η ευθεία που θα προκύψει να μας δίνει όσο το δυνατόν την καλύτερη περιγραφή της σχέσης (εξάρτησης) που υπάρχει μεταξύ των μεταβλητών X (ηλικία) και Y(ΧρόνοςΠροϋπηρεσίας).

Η παράμετρος b μας δίνει το σημείο (0,b), όπου η ευθεία αυτή τέμνει τον άξονα y'y, ενώ η παράμετρος a παριστάνει το συντελεστή διεύθυνσης της ευθείας.

3)Απο τον πίνακα Model Summary υπολογίζουμε τον δείκτη προσδιορισμού:

R^2=R Square=0,653

Δηλαδή κατά 65,3% η μεταβλητότητα του χρόνου Προϋπηρεσίας ερμηνεύεται από την ηλικία.

Συμπεράσματα απο την χρήση του SPSS

Οι μεταβλητές που επιρρεάζουν τον αρχικό μισθό είναι ο χρόνος προϋπηρεσίας και τα χρόνια εκπαίδευσης. Σύμφωνα με τον δείκτη προσδιορισμού R^2 ο αρχικός μισθός επιρρεάζεται περισσότερο απο τα χρόνια εκπαίδευσης.

Ο σημερινός μισθός επιρρεάζεται απο την επίδοση, τα χρόνια εκπαίδευσης και απ'ότι φαίνεται, δεν εξαρτάται καθόλου απο τα χρόνια προϋπηρεσίας. . Σύμφωνα με τον δείκτη προσδιορισμού R^2 ο σημερινός μισθός εξαρτάται περισσότερο απο τα χρόνια εκπαίδευσης, όπως και ο αρχικός μισθός.

Ο χρόνος προϋπηρεσίας επιρρεάζεται κατα πολύ απο την ηλικία

Κεφάλαιο 5: Συμπεράσματα

Ο σκοπός της παρούσας εργασίας ήταν η μελέτη των βασικών εννοιών της θεωρίας της Περιγραφικής Στατιστικής ενώ επιπλέον παρουσιάστηκαν και κάποιες στατιστικές εφαρμογές με χρήση του στατιστικού πακέτου SPSS.

Όπως είδαμε, τα περιγραφικά στατιστικά μέτρα χωρίζονται σε τρεις επιμέρους κατηγορίες ενώ κάθε μία περιλαμβάνει διαφορετικά μέτρα. Αυτές οι κατηγορίες είναι τα μέτρα θέσης ή κεντρικής τάσης, τα μέτρα διασποράς και τα μέτρα ασυμμετρίας και κύρτωσης.

Τα μέτρα κεντρικής τάσης ή θέσης όπου σκοπό έχουν να περιγράψουν το πώς οι παρατηρήσεις του δείγματος μας εκτείνονται γύρω από το κέντρο τους και μεταξύ άλλων περιλαμβάνουν τον *μέσο*, την *διάμεσο*, την *επικρατούσα τιμή* και τα *εκατοστημόρια*.

Από την άλλη μεριά, τα μέτρα διασποράς μας ενημερώνουν για τον τρόπο με τον οποίο οι παρατηρήσεις μας εκτείνονται γύρω από το κέντρο του δείγματος και αποτελούνται από την *διακύμανση*, την *τυπική απόκλιση* και τον *συντελεστή μεταβλητότητας* και το *ενδο-τεταρτημοριακό εύρος*.

Τα μέτρα ασυμμετρίας και κύρτωσης μας πληροφορούν για την μορφή της κατανομής των παρατηρήσεων και περιλαμβάνουν τον συντελεστή ασυμμετρίας και κύρτωσης αντίστοιχα.

Επίσης, μελετήθηκαν οι τυχαίες μεταβλητές οι οποίες χωρίζονται στις *διακριτές* και τις *συνεχείς*. Οι πρώτες περιέχουν ένα πεπερασμένο (αριθμήσιμο) πλήθος τιμών, ενώ οι δεύτερες περιλαμβάνουν το σύνολο των πραγματικών αριθμών ή διαστήματα αυτών. Επίσης παρουσιάστηκαν οι κατανομές τους και οι πρώτες και δεύτερες ροπές τους.

Επίσης, έγινε εμφανής η σημασία της Οικονομετρίας ως βασικό εργαλείο για την μελέτη και ποσοτικοποίηση των οικονομικών θεωριών και των οικονομικών σχέσεων που παρατηρούνται στον πραγματικό κόσμο τόσο για την μελέτη οικονομικών φαινομένων όσο και για την άσκηση οικονομικής πολιτικής.

Τέλος, το βασικό συμπέρασμα της παρούσας εργασίας είναι πως η Στατιστική αποτελεί μια πολύ σημαντική Επιστήμη για την μελέτη των οικονομικών και όχι μόνο φαινομένων και μας βοηθάει να βγάλουμε ορθά συμπεράσματα για τις σχέσεις των υπό εξέταση μεταβλητών αλλά και να αναλύσουμε την συμπεριφορά πολύπλοκων σχέσεων που παρατηρούνται στην πραγματική οικονομία και επηρεάζουν την καθημερινή ζωή μας.

Βιβλιογραφία

Ελληνόγλωσση.

1) Παπαδημητρίου, Γ. (2005), 'Περιγραφική Στατιστική', Τυπωθήτω, Αθήνα

2) Τζαβαλής, Η. (2008), 'Οικονομετρία', Εκδόσεις Οικονομικού Πανεπιστημίου Αθηνών

3) Τσαγκρής, Μ. (2008), 'Στατιστική με την χρήση του πακέτου SPSS 15'

4) Τσιώνας, Ε. (2010), 'Εφαρμοσμένη Οικονομετρία', Οικονομικό Πανεπιστήμιο Αθηνών

5) Χατζηνικολάου, Δ. (2002), Στατιστική για Οικονομολόγους', Δημήτριος Χατζηνικολάου, Β' Έκδοση, Ιωάννινα

6) Χρήστου, Γ. (2007), 'Εισαγωγή στην Οικονομετρία',Τόμος Α', Έκδοση Γ', Εκδόσεις Gutenberg

Ξενόγλωσση.

7) Salvatore, D. & Reagle, D. (2002), 'Theory and Problems of Statistics and Econometrics', 2nd edt., Shaum's Outline Series, McGraw-Hill, United States of America

8) Wooldridge, J. (2006), 'Εισαγωγή στην Οικονομετρία, Μια νέα προσέγγιση', Εκδόσεις Παπαζήση, Τόμος Α΄.

Διαδικτυακές Πηγές

9) http://www.unipi.gr/faculty/mbouts/prob_intro/prob_intro4.pdf

10) http://users.uoi.gr/alapatin/files/Lecture_06_presentation.pdf

11) http://www.unipi.gr/faculty/mkoutras/pithI/Prob_I_ch5a.pdf

12) http://www.unipi.gr/faculty/mkoutras/regres/regres1_1.pdf