τραμεταβλητό μαθηματικό υπόδειγμα.

Με τα διμεταβλητά μαθηματικά υποδείγματα εξετάζουμε αν υπάρχει σχέση εξαρτήσεως της μιάς μεταβλητής Ψ από μιά άλλη μεταβλητή Χ. Η μεταβλητή Χ - η οποία ενδέχεται να επιδρά και να διαμορφώνει τις τιμές της μεταβλητής Ψ - ονομάζεται **ανεξάρτητη μεταβλητή** (Independent Variable), ενώ η μεταβλητή Ψ - η οποία δέχεται τις επιδράσεις της μεταβλητής Χ - ονομάζεται **εξαρτημένη μεταβλητή** (Dependent Variable).

Το κεντρικό πρόβλημα της αναλύσεως διμεταβλητών στατιστικών πληθυσμών είναι να διαπιστώσουμε αν υπάρχει σχέση εξαρτήσεως (συσχέτιση) μεταξύ των μεταβλητών Ψ και Χ (δηλαδή αν η ανεξάρτητη μεταβλητή επηρεάζει τη διαμόρφωση των τιμών της εξαρτημένης μεταβλητής) και να μετρήσουμε το **βαθμό (την ένταση) της συσχετίσεως**. Ο προσδιορισμός και η μέτρηση της συσχετίσεως είναι πολύ χρήσιμος για την επιχειρηματική πρόβλεψη, για τον προγραμματισμό της δράσεως μιάς επιχειρήσεως και γενικά για τη λήψη ορθών επιχειρηματικών αποφάσεων.

Υπάρχουν δύο μέθοδοι σπουδής της εξαρτήσεως (συνάφειας) μεταξύ δύο μεταβλητών Χ και Ψ.

α) Η πρώτη μέθοδος ονομάζεται **ανάλυση παλινδρομήσεως*** (Regression Analysis) και αποβλέπει στον προσδιορισμό μιάς γενικής σχέσεως εξαρτήσεως μεταξύ των συνεξεταζόμενων μεταβλητών Χ και Ψ, δηλαδή στον προσδιορισμό μιάς μαθηματικής εξισώσεως, η οποία ονομάζεται **Εξίσωση Παλινδρομήσεως** (Regression Equation). Με άλλα λόγια, στην Ανάλυση Παλινδρομήσεως εξετάζουμε τη σχέση εξαρτήσεως μεταξύ των μεταβλητών Ψ και Χ έτσι, ώστε να μπορέσουμε να προβλέψουμε μελλοντικά την τιμή της μιάς μεταβλητής με βάση την άλλη μεταβλητή. Η μεταβλητή η οποία λαμβάνεται σαν βάση της προβλέψεως καλείται **ανεξάρτητη μεταβλητή** (= Χ). Σαν ανεξάρτητη μεταβλητή θεωρείται εκείνη της οποίας οι τιμές είναι προκαθορισμένες ή επιλέγονται κατά κάποιο τρόπο. Π.χ. επιλέγουμε άτομα ορισμένου εισοδήματος για να εξετάσουμε την επίδραση που έχει το εισόδημα (= Χ) επί της δαπάνης για ψυχαγωγία (= Ψ). Η μεταβλητή που θέλουμε να προβλέψουμε και η οποία δέχεται τις επιδράσεις της ανεξάρτητης μεταβλητής, καλείται **εξαρτημένη μεταβλητή** (= Ψ) και είναι μία τυχαία μεταβλητή.

Η μορφή της εξαρτήσεως των διαφόρων μεταβλητών μπορεί να είναι **γραμμική ή καμπυλόγραμμη**. Λέγοντας ότι η σχέση εξαρτήσεως μεταξύ δύο μεταβλητών είναι γραμμική, εννοούμε ότι ο μέσος όρος της εξαρτήσεως μεταξύ των μεταβλητών μπορεί να αναπαρασταθεί με μιά **ευθεία γραμμή** $Ψ = α + βΧ$. Ενώ όταν τα δεδομένα της παρατηρήσεως μπορούν να αναπαρασταθούν με μιά

* Ο όρος "παλινδρόμηση" οφείλεται στο Sir Francis Galton (1822-1911), ο οποίος παρατήρησε ότι πατέρες χαμηλού αναστήματος αποκτούσαν γιούς, των οποίων το μέσο ανάστημα "παλινδρομεί" (μετακινείται) προς το μέσο ανάστημα του πληθυσμού. Στη Στατιστική σήμερα, ο όρος "παλινδρόμηση" χρησιμοποιείται για να υποδηλώσει την εξάρτηση (συνάφεια) μεταξύ δύο ή περισσοτέρων μεταβλητών.

καμπύλη, τότε λέμε ότι η σχέση εξαρτήσεως μεταξύ των μεταβλητών είναι **καμπυλόγραμμη**, Π.χ. $Ψ = α + βΧ + γΧ^2$. Η πρώτη περίπτωση λέγεται **γραμμική παλινδρόμηση** (Linear Regression) και η δεύτερη **καμπυλόγραμμη παλινδρόμηση** (Curvilinear Regression). Επίσης, η Ανάλυση Παλινδρομήσεως μπορεί να περιλαμβάνει περισσότερες από δύο μεταβλητές. Π.χ. η ζητούμενη ποσότητα (= Ψ) ενός αγαθού εξαρτάται από την τιμή (= Χ) του αγαθού, από το διαθέσιμο εισόδημα (= Υ), από την τιμή των συναγωνιζόμενων αγαθών (= Ζ), κ.λ.π. Στην περίπτωση αυτή λέμε ότι έχουμε μία **πολλαπλή** σχέση εξαρτήσεως, η οποία ονομάζεται **πολλαπλή παλινδρόμηση** (Multiple Regression) και η οποία μπορεί να είναι γραμμική ή καμπυλόγραμμη.

β) Η δεύτερη μέθοδος σπουδής της εξαρτήσεως (συσχετίσεως) της μιάς μεταβλητής Ψ από μιά άλλη μεταβλητή Χ, ονομάζεται **Συσχέτιση** (Correlation). Η Συσχέτιση ασχολείται με τον ποσοτικό προσδιορισμό του **βαθμού** (της εντάσεως) της εξαρτήσεως μεταξύ των μεταβλητών και με τη **φύση** της συσχετίσεως (θετική ή αρνητική). Η μέτρηση του βαθμού της συσχετίσεως μεταξύ δύο ή περισσοτέρων μεταβλητών γίνεται με μιά στατιστική παράμετρο, η οποία ονομάζεται **Συντελεστής Συσχετίσεως** (Correlation Coefficient). Η συσχέτιση διακρίνεται σε θετική και αρνητική. **Θετική** καλείται η συσχέτιση, όταν σε κάθε αύξηση (ή μείωση) της μιάς μεταβλητής αντιστοιχεί αύξηση (ή μείωση) και της άλλης μεταβλητής. Π.χ. αύξηση του εισοδήματος (= Χ) συνεπάγεται αύξηση της δαπάνης για ψυχαγωγία (=Ψ) και αντίστροφα. **Αρνητική** καλείται η συσχέτιση, όταν σε κάθε αύξηση (ή μείωση) της μιάς μεταβλητής αντιστοιχεί μείωση (ή αύξηση) της άλλης μεταβλητής. Π.χ. αύξηση της τιμής ενός αγαθού (= Χ) συνεπάγεται μείωση της ζητούμενης ποσότητας (=Ψ) του αγαθού και αντίστροφα. Αν τώρα τα μεταβλητά Χ και Ψ είναι ανεξάρτητα μεταξύ τους, τότε θα είναι και ασυσχέτιστα, δηλαδή η μεταβολή των τιμών της μιάς μεταβλητής (=Χ) δεν θα προκαλεί επιδράσεις στις τιμές της άλλης μεταβλητής (= Ψ). Άρα δεν υπάρχει πρόβλημα αναλύσεως παλινδρομήσεως και συσχετίσεως.

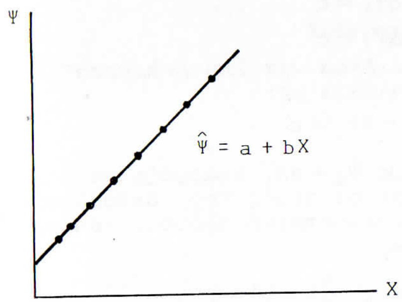
Αν διαθέτουμε ένα δείγμα n ζευγών - τιμών των μεταβλητών Ψ και Χ και θέλουμε να έχουμε μία πρώτη εικόνα για την ύπαρξη ή όχι εξαρτήσεως (συσχετίσεως) μεταξύ των μεταβλητών Ψ και Χ, τότε απεικονίζουμε τις τιμές των Ψ και Χ πάνω σ' ένα σύστημα ορθογώνιων αξόνων. Τα σημεία του επιπέδου, τα οποία έχουν τεταγμένες τις τιμές της μεταβλητής Ψ και τετμημένες τις τιμές της μεταβλητής Χ, δημιουργούν ένα "νέφος" σημείων, το οποίο ονομάζεται **Διάγραμμα Διασποράς** (Scatter Diagram).

Το Διάγραμμα Διασποράς είναι χρησιμότατο, γιατί μας δίνει μία πρώτη εικόνα για το είδος της εξαρτήσεως της μεταβλητής Ψ από τη μεταβλητή Χ (δηλ. αν η εξάρτηση είναι γραμμική ή καμπυλόγραμμη, θετική ή αρνητική), και επιπλέον μας δείχνει το **βαθμό** (ένταση) της συσχετίσεως μεταξύ των μεταβλητών Ψ και Χ.

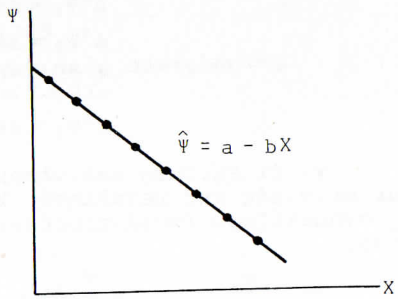
Στο Διάγραμμα 12.1 παραθέτουμε αντιπροσωπευτικά παρα-

Διάγραμμα 12.1

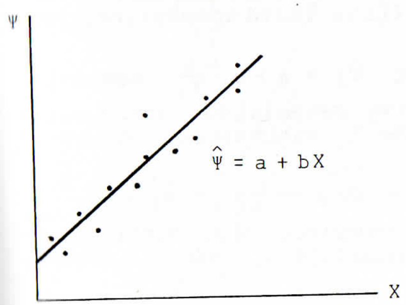
Παραδείγματα γραμμικής εξαρτήσεως και συσχετίσεως



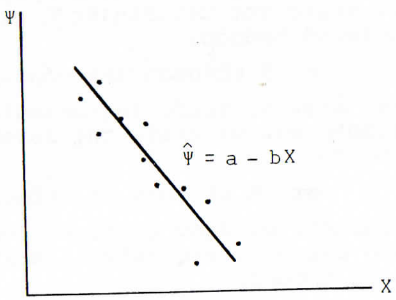
α) Τέλεια θετική εξάρτηση και συσχέτιση



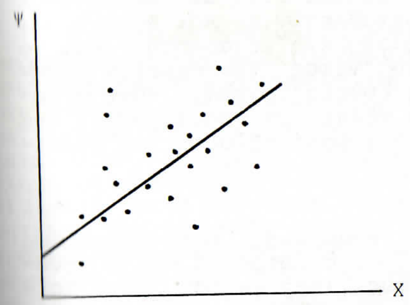
β) Τέλεια αρνητική εξάρτηση και συσχέτιση



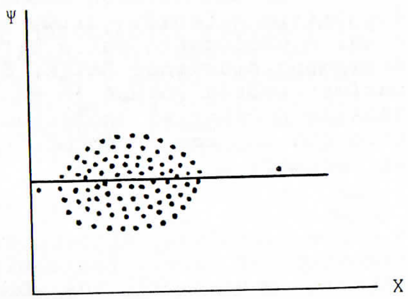
γ) Έντονη θετική εξάρτηση και συσχέτιση



δ) Έντονη αρνητική εξάρτηση και συσχέτιση



ε) Ασθενής εξάρτηση και συσχέτιση



στ) Ασυσχέτιστα μεταβλητά

δείγματα γραμμικής εξαρτήσεως και συσχετίσεως μεταξύ των μεταβλητών Ψ και X, με τις αντίστοιχες εξισώσεις γραμμικής παλινδρομήσεως.

Από το Διάγραμμα 12.1 συνάγονται τα ακόλουθα:

α) Όταν τα σημεία σχηματίζουν ευθεία γραμμή, η οποία έχει **θετικό γωνιακό συντελεστή*** (περίπτωση 12.1α), τότε έχουμε θετική εξάρτηση και τέλεια θετική συσχέτιση, δηλαδή και οι δύο μεταβλητές κινούνται προς την ίδια κατεύθυνση, ενώ όταν τα σημεία σχηματίζουν ευθεία γραμμή, η οποία έχει **αρνητικό γωνιακό συντελεστή** (περίπτωση 12.1β), τότε υπάρχει αρνητική εξάρτηση και τέλεια αρνητική συσχέτιση, δηλαδή οι μεταβλητές κινούνται προς αντίθετη κατεύθυνση.

β) Όταν τα σημεία δεν σχηματίζουν ευθεία γραμμή αλλά και δεν απομακρύνονται πολύ απ'αυτή (περίπτωση 12.1γ και 12.1δ), τότε υπάρχει θετική ή αρνητική εξάρτηση ανάλογα αν η ευθεία που περνάει μέσα από το "νέφος" των σημείων έχει θετικό ή αρνητικό γωνιακό συντελεστή και η συσχέτιση είναι έντονη. Όταν όμως τα σημεία απομακρύνονται από την ευθεία (περίπτωση 12.1ε), τότε εξακολουθεί να υπάρχει εξάρτηση, αλλά η συσχέτιση γίνεται πιό ασθενής.

γ) Τέλος, όταν τα σημεία σχηματίζουν κύκλο (περίπτωση 12.1στ) ή τετράγωνο, τότε δεν υπάρχει καμιά εξάρτηση και άρα καμιά συσχέτιση.

Στο Διάγραμμα 12.2 παραθέτουμε ορισμένες μορφές καμπυλόγραμμης εξαρτήσεως μεταξύ των μεταβλητών Ψ και X, με τις αντίστοιχες εξισώσεις καμπυλόγραμμης παλινδρόμησης.

Εκτός από τα διαγράμματα διασποράς, υπάρχουν και ορισμένα άλλα κριτήρια για την εκλογή της καταλληλότερης μορφής μιάς εξισώσεως παλινδρομήσεως, την οποία θα προσαρμόζουμε κάθε φορά στα δεδομένα της παρατηρήσεως. Ειδικότερα πρέπει να έχουμε υπόψη μας ότι ισχύουν τα εξής κριτήρια:

α) Η εξίσωση παλινδρομήσεως $\hat{\Psi}_i = a + bX_i$ εφαρμόζεται όταν και οι τιμές της μεταβλητής X_i και οι τιμές της μεταβλητής Ψ_i σχηματίζουν αριθμητική πρόοδο ή όταν οι πρώτες διαφορές των μεταβλητών X_i και Ψ_i είναι σταθερές ή περίπου σταθερές. Δηλαδή όταν:

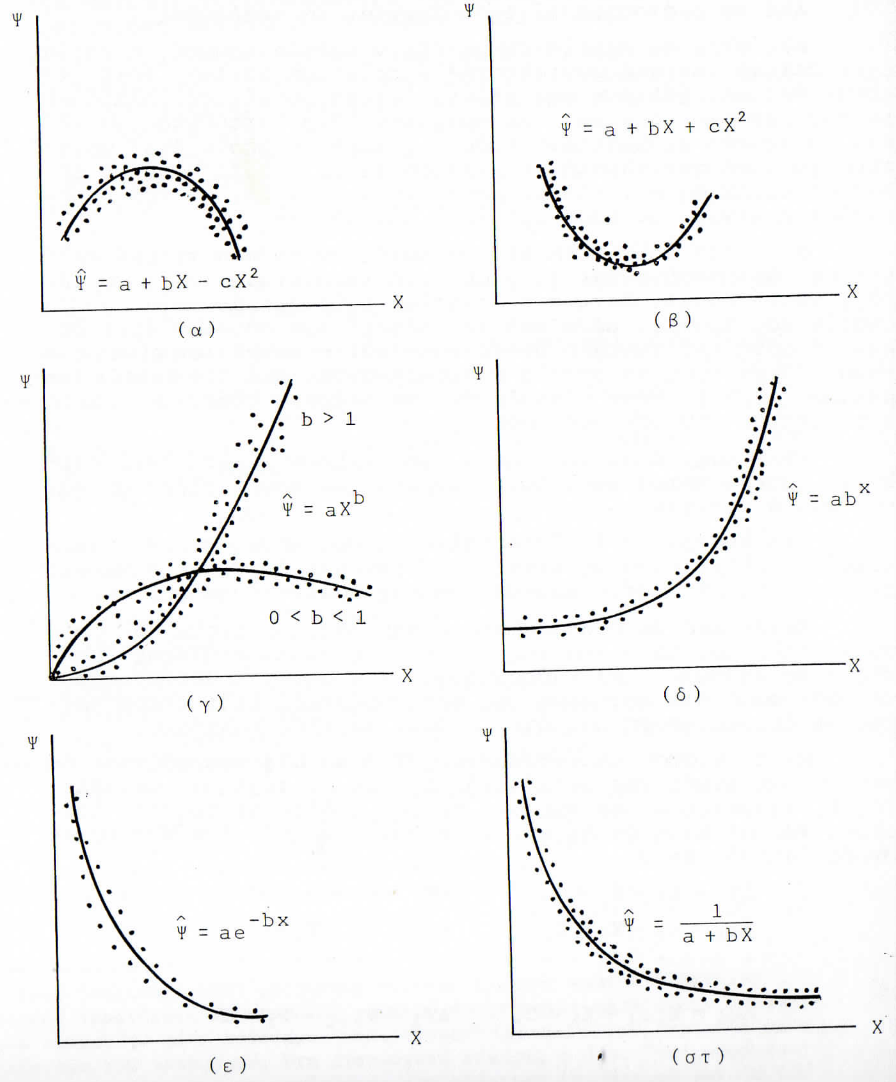
$$\Delta X_1 = X_2 - X_1 = C, \qquad \Delta\Psi_1 = \Psi_2 - \Psi_1 = C$$
$$\Delta X_2 = X_3 - X_2 = C, \qquad \Delta\Psi_2 = \Psi_3 - \Psi_2 = C$$
$$\cdots\cdots\cdots \qquad \cdots\cdots\cdots$$
$$\cdots\cdots\cdots \qquad \cdots\cdots\cdots$$
$$\Delta X_i = X_{i+1} - X_i = C, \qquad \Delta\Psi_i = \Psi_{i+1} - \Psi_i = C$$

---

* Γωνιακός συντελεστής ή συντελεστής κατευθύνσεως ή κλίση (slope) της ευθείας $\Psi = \alpha + \beta X$ είναι η εφαπτομένη της γωνίας, η οποία σχηματίζεται από την τομή της ευθείας $\Psi = \alpha + \beta X$ με το θετικό ημιάξονα OX (Βλ. περισσότερα στην επόμενη παράγραφο 12.2).

Διάγραμμα 12.2

Διάφορες μορφές καμπυλόγραμμης παλινδρόμησης



β) Η εξίσωση παλινδρομήσεως $\hat{\Psi} = a + bX + cX^2$ εφαρμόζεται όταν οι τιμές της μεταβλητής X_i σχηματίζουν αριθμητική πρόοδο και οι δεύτερες διαφορές της μεταβλητής Ψ_i είναι σταθερές ή περίπου σταθερές. Δηλαδή:

$$\Delta^2\Psi_1 = \Delta\Psi_2 - \Delta\Psi_1 = c$$
$$\Delta^2\Psi_2 = \Delta\Psi_3 - \Delta\Psi_2 = c$$
$$\cdots\cdots$$
$$\Delta^2\Psi_i = \Delta\Psi_{i+1} - \Delta\Psi_i = c$$

γ) Οι εξίσωση παλινδρομήσεως $\hat{\Psi}_i = aX_i^b$ εφαρμόζεται όταν και οι τιμές της μεταβλητής X_i και οι τιμές της μεταβλητής Ψ_i σχηματίζουν (κατά προσέγγιση) γεωμετρική πρόοδο. Δηλαδή όταν:

$$\frac{\Psi_2}{\Psi_1} = \frac{\Psi_3}{\Psi_2} = \ldots = \frac{\Psi_{i+1}}{\Psi_i} = \omega \quad \text{και} \quad \frac{X_2}{X_1} = \frac{X_3}{X_2} = \ldots = \frac{X_{i+1}}{X_i} = \omega'$$

δ) Η εξίσωση παλινδρομήσεως $\hat{\Psi}_i = ab^{X_i}$ εφαρμόζεται όταν οι τιμές της μεταβλητής X_i σχηματίζουν αριθμητική πρόοδο και οι τιμές της μεταβλητής Ψ_i σχηματίζουν (κατά προσέγγιση) γεωμετρική πρόοδο.

ε) Η εξίσωση παλινδρομήσεως $\hat{\Psi}_i = a + b \cdot \frac{1}{X_i}$ εφαρμόζεται όταν οι τιμές της μεταβλητής Ψ_i σχηματίζουν αριθμητική πρόοδο και οι τιμές της μεταβλητής X_i σχηματίζουν αρμονική πρόοδο.

στ) Η εξίσωση παλινδρομήσεως $\hat{\Psi}_i = \frac{1}{a + bX_i}$ ή $\hat{\Psi}_i = \frac{X_i}{a + bX_i}$ εφαρμόζεται όταν οι τιμές της X_i σχηματίζουν αριθμητική πρόοδο και οι τιμές (λόγοι) X_i/Ψ_i σχηματίζουν, επίσης, αριθμητική πρόοδο.

12.2. ΔΙΜΕΤΑΒΛΗΤΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

Στην προηγούμενη παράγραφο είπαμε ότι, αν διαθέτουμε ένα δείγμα n ζευγών - τιμών των μεταβλητών X_i και Ψ_i - η X_i είναι η ανεξάρτητη και η Ψ_i η εξαρτημένη μεταβλητή - και το διάγραμμα διασποράς δείξει ότι το "νέφος" των σημείων σχηματίσει ευθεία γραμμή (ή περίπου ευθεία γραμμή), τότε το κατάλληλο μαθηματικό υπόδειγμα που πρέπει να προσαρμοστεί στα δεδομένα της παρατηρήσεως είναι η γνωστή εξίσωση της ευθείας γραμμής:

$$\hat{\Psi}_i = a + bX_i \qquad (12.1)$$

η οποία ονομάζεται **εξίσωση απλής γραμμικής παλινδρομήσεως** (Equation of Linear Regression). Οι σταθερές a και b ονομάζονται **συντελεστές παλινδρομήσεως** (Regression Coefficients) και έχουν τις εξής έννοιες:

Η σταθερά a εκφράζει την τιμή της Ψ όταν X = 0. Όπως φαίνεται και στο Διάγραμμα 12.3, η σταθερά a δείχνει το σημείο τομής της ευθείας $\hat{\Psi} = a + bX$ με τον κατακόρυφο άξονα (=Ψ)