

## ΠΡΟΛΟΓΟΣ

Οι ανάγκες της διαρκώς εξελισσόμενης κοινωνίας μας ολοένα και πλυθένουν . Μια απ'τις μεγαλύτερες και σπουδαιότερες είναι αυτή της ανάγκης για ανταλλαγή πληροφοριών γι'αυτό άλλωστε και αρκετοί μιλάνε για την εποχή της πληροφορίας .

Ένας απ'τους τομείς που ασχολήται με την ανταλλαγή πληροφοριών είναι και αυτός της τηλεφωνίας .

Η παρούσα διπλωματική έχει σαν στόχο μελετώντας την παρούσα τηλεφωνική κατάσταση στην Ελλάδα να βγάλει συμπεράσματα για το μέλλον αυτής . Ποιο συγκεκριμένα έχει σαν σκοπό να κάνει πρόβλεψη στην εγκατεστημένη χωρητικότητα των τηλεφωνικών κέντρων χρησιμοποιώντας μια σειρά από μεθόδους . Η πρόβλεψη της εγκατεστημένης χωρητικότητας θεωρήται απ'τον Ο.Τ.Ε πλέον σημαντική μιας και βάση αυτής παίρνονται οι αποφάσεις για το πόσα τηλεφωνικά κέντρα πρέπει να φτιαχθούν , τι χωρητικότητας πρέπει να είναι και όλα όσα συνεπάγονται αυτών των αποφάσεων . Αποφάσεις βέβαια που μπορεί να κοστίσουν ακριβά όταν παρθούν λάθος ή με αρκετά μεγάλη απόκλιση απ'την πραγματικότητα .

# ΚΕΦΑΛΑΙΟ 1

## **1.1 Γενικά ιστορικά**

Απ' την στιγμή που ο άνθρωπος έκανε την εμφάνισή του πάνω στη γη άρχισε να συνειδητοποιεί την σημασία που έχει η επικοινωνία από απόσταση με τους όμοιους του . Για το λόγο αυτό άρχισε να ψάχνει τρόπους με τους οποίους θα μπορούσε να επικοινωνήσει γρήγορα και εύκολα .

Γνωστό απ' την ιστορία μας είναι ότι οι Σίνες πριν από 30 αιώνες χρησιμοποιούσαν σημεία με φωτιές για να στέλνουν τα μηνύματά τους κάτι το οποίο έκαναν αργότερα και οι Πέρσες και οι Ρωμαίοι . Παράλληλα επιχειρήθηκε και συνεννόηση από απόσταση με χρήση ηχητικών σημάτων εκπεμπόμενα από σειρήνες κώδωνες κ.λ.π.

Στα τέλη του 18ου αιώνα ο Γάλλος Cheppe ανακαλύπτει τον οπτικό τηλεγράφο ο οποίος με κινήσεις πλαισίων και αξόνων μεταβίβαζε γράμματα και λέξεις. Ο τρόπος όμως αυτός παρουσίαζε δυσκολίες κάτω από δυσμενείς καιρικές συνθήκες.

Αρκετά αργότερα ο Samuel Morse δίνει τη δυνατότητα με τον τηλεγράφο να μεταβιβάζονται 120-140 γράμματα το λεπτό. Στη συνέχεια και καθώς οι ανάγκες του κόσμου μεγάλωναν έγινε εφαρμογή του συστήματος κατά το οποίο επιτρεπόταν η πραγματοποίηση δύο ή περισσότερων μεταβιβάσεων στην ίδια γραμμή . Καθώς οι τηλεγραφικές γραμμές μεγάλωναν άρχισε και η εφαρμογή του ενισχυτικού ηλεκτρονόμου ο οποίος πραγματοποιούσε ενίσχυση της αναμεταβίβασης των σημάτων από το μέσο ως και το τέλος της γραμμής και στη συνέχεια η τοποθέτηση σε ανταποκρινόμενο σταθμό συμμετρικών

ηλεκτρονόμων λήψεως , με τη βοήθεια των οποίων ήταν δυνατή η ταυτόχρονη εκπομπή και λήψη και στις δύο κατευθύνσεις της γραμμής που μέχρι τότε ήταν αδύνατη .

Το τηλέφωνο θα ανακαλύψει το 1905 ο Alex Graham Bell όταν ήταν δάσκαλος κωφάλαλων στη Βοστώνη ο οποίος και δίκαια θεωρείται ο πατέρας της τηλεφωνίας . Στην αρχή το τηλέφωνο αποτελείτο από δύο ξεχωριστά εξαρτήματα , τον πομπό δηλαδή το μικρόφωνο και το δέκτη , δηλαδή το μεγάφωνο . Μετά από αρκετές τροποποιήσεις και απλουστεύσεις κατέληξαν στο ακουστικό .

## **1.2 Τηλεπικοινωνίες στην Ελλάδα**

Οι τηλεφωνικές επικοινωνίες παρέχονται στην Ελλάδα υπό καθεστώς μονοπωλίου από τον Ο.Τ.Ε (Οργανισμός Τηλεπικοινωνιών της Ελλάδας) . Ο Ο.Τ.Ε ιδρύθηκε το 1949 ως Δημόσια Επιχείρηση Κοινής Ωφελείας (Δ.Ε.Κ.Ο) .

Ο Ο.Τ.Ε διοικείται από Διοικητικό συμβούλιο , Διοικητή και Υποδιοικητή που ορίζονται από την κυβέρνηση . Οι υπηρεσίες του Οργανισμού διακρίνονται σε κεντρικές και περιφερειακές όπου οι περιφερειακές ελέγχονται και καθοδηγούνται σε μεγάλο βαθμό από τις κεντρικές .

Οι στόχοι του Οργανισμού χωρίζονται σε δύο κατηγορίες από όπου πρέπει να διακρίνεται ο διπλός ρόλος και ο χαρακτήρας του Οργανισμού ως Δ.Ε.Κ.Ο και ως Α.Ε .

Ως Δ.Ε.Κ.Ο ο Ο.Τ.Ε οφείλει :

1. Την παροχή ποιοτικά και ποσοτικά ικανοποιητικών τηλεπικοινωνιών βελτιώνοντας και επεκτείνοντας το δίκτυο.

2. Την παροχή φθηνών τηλεπικοινωνιών , προσιτών σε κάθε κοινωνικό στρώμα .

3. Την παροχή “κοινωνικών” υπηρεσιών , δηλαδή τη δημιουργία και τη διατήρηση δικτύου ακόμα και όταν είναι ζημιογόνο .

Ως Α.Ε όμως πρέπει :

1. Να διασφαλίζει κέρδη για να επεκτείνεται

2. Να έχει τη μέγιστη δυνατή απόδοση .

### **1.3 Υπηρεσίες Ο.Τ.Ε**

Οι υπηρεσίες που παρέχει ο Ο.Τ.Ε είναι :

- Αστική τηλεφωνία με διάφορες ειδικές υπηρεσίες για τους καταναλωτές .
- Υπεραστική και Διεθνή τηλεφωνία .
- Τηλεγραφία εσωτερικού και εξωτερικού .
- Παράκτια ραδιοεπικοινωνία (ραδιοτηλεφωνία και ραδιοτηλεγραφία)
- Τηλετυπία (Telex) και τηλεομοιοτυπία (Fax) .

#### **1.3.1 Νέες υπηρεσίες**

Ο Ο.Τ.Ε ανταποκρινόμενος στα νέα δεδομένα της τηλεπικοινωνιακής αγοράς , που καθημερινά διευρύνεται , στις διεθνείς τεχνολογικές και εμπορικές εξελίξεις και σε

συνδυασμό με την εξασφάλιση ανοικτή πρόσβασης στα δημόσια δίκτυα για τους ιδιώτες και με την εισαγωγή των διευρωπαϊκών δικτύων , τα οποία έχουν τη δυνατότητα να υποστηρίζουν πλήθος εξελιγμένων υπηρεσιών , παρέχει ολοένα και περισσότερα προϊόντα/υπηρεσίες προηγμένης τεχνολογίας στον πελάτη .

Συγκεκριμένα ο Ο.Τ.Ε προσφέρει πλέον :

- Το δίκτυο Hellaspac που προσφέρει στους χρήστες του μία αποδοτική και οικονομική μέθοδο επικοινωνίας ακόμη και μεταξύ υπολογιστών και τερματικών διαφορετικού τύπου και ταχύτητας .
  - Το δίκτυο Hellascom το οποίο είναι ένα υψηλότερου επιπέδου και αξιοπιστίας δίκτυο που επιτρέπει με μεγάλες ταχύτητες έως 2Mbps , τη μετάδοση δεδομένων , μέσο της παροχής σταθεροζευκτικών ψηφιακών κυκλωμάτων στους ενδιαφερόμενους χρήστες .
- Το δίκτυο Τηλεεικονογραφίας (Videotext) “HELLASTEL”. Ο Ο.Τ.Ε με το δίκτυο αυτό εξασφαλίζει στους χρήστες την πληροφόρηση τους που χωρίς αμφιβολία είναι πιο εύκολη και ευέλικτη από την πληροφόρηση που επιτυγχάνεται με τον παραδοσιακό τρόπο άντλησης των πληροφοριών , μιας και η νέα μορφή πληροφόρησης δίνεται με την μορφή γραπτού κειμένου ή σχεδιαγράματος στην οθόνη του τερματικού του χρήστη , είναι διαθέσιμη όλο το 24ώρο χωρίς κανένα χρονικό περιορισμό , είναι επίκαιρη και συνεχώς ενημερωμένη στις τελευταίες εξελίξεις και αλλαγές του αντικειμένου της .
- Το δίκτυο διαχείρισης μηνυμάτων ηλεκτρονικής ανταλλαγής παραστατικών εγγράφων και καταλόγου

“ERMIS” κατα το οποίο αναμένεται να συμβάλει αποτελεσματικά στην ικανοποίηση των βραχυπρόθεσμων επιτακτικών αναγκών του δημόσιου και ιδιωτικού τομέα σε επικοινωνίες ηλεκτρονικού ταχυδρομείου .

- Το ψηφιακό δίκτυο “ISDN” ενοποιημένων υπηρεσιών που υποστηρίζει τη μετάδοση φωνής , εικόνας , κειμένου και δεδομένων μέσα από ένα ζεύγος δικτύου .
- Την Audiotex , που πρόκειται για μία προστιθέμενης αξίας υπηρεσία που παρέχει τη δυνατότητα στους χρήστες του τηλεφωνικού δικτύου να έχουν άμεση πρόσβαση σε τράπεζες πληροφοριών αντλώντας ζωντανές ή μαγνητοφωνημένες πληροφορίες .
- Την “Τηλεδιάσκεψη” η οποία δίνει τη δυνατότητα στους χρήστες της να επικοινωνούν μεταξύ τους με στόχο την περιφερειακής ανάπτυξη και αναβάθμιση των δυνατοτήτων επαφών σε εθνικό επίπεδο επιχειρήσεων και οργανισμών .

## **ΣΥΜΠΕΡΑΣΜΑΤΑ**

Κατα γενική ομολογία , ο Ο.Τ.Ε αποτελεί στρατηγικής σημασίας επιχείρηση και η συμμετοχή του στις οικονομικές εξελίξεις είναι καθοριστική . Αποτελεί μοχλό οικονομικής ανάπτυξης και κοινωνικής παρέμβασης και η ανάδειξή του σε “ κυρίαρχο παράγοντα ” αποτελεί προϋπόθεση για την ανάπτυξη της χώρας .

## ΚΕΦΑΛΑΙΟ 2

### ΤΕΧΝΟΛΟΓΙΚΗ ΠΡΟΒΛΕΨΗ

#### 2.1 Ανάγκη για πρόβλεψη

Όπως είδαμε στο προηγούμενο κεφάλαιο η ανάπτυξη της τεχνολογίας σ'όλο το φάσμα των επιστημών αλλά και ειδικότερα στις τηλεπικοινωνίες (μιας και αυτές μελετάμε) έχει παίξει σημαντικό ρόλο στην ανάπτυξη του ανθρώπινου είδους . Καταλαβαίνουμε λοιπόν ότι η μελέτη και η προσπάθεια κατανόησης της γενικής μελλοντικής τάσης του κοινού έχει μεγάλη σημασία μιας και θα μπορούσαμε να βαδίσουμε με μεγαλύτερη σιγουριά προς τα εμπρός .

Μια διαδικασία υπολογισμού ενός μελλοντικού γεγονότος με τη βοήθεια παρελθόντος πληροφορίας επεξεργασμένη συστηματικά με ένα προκαθορισμένο τρόπο καλείται **Πρόβλεψη** .

## **2.2 Η διαδικασία προβλέψεων**

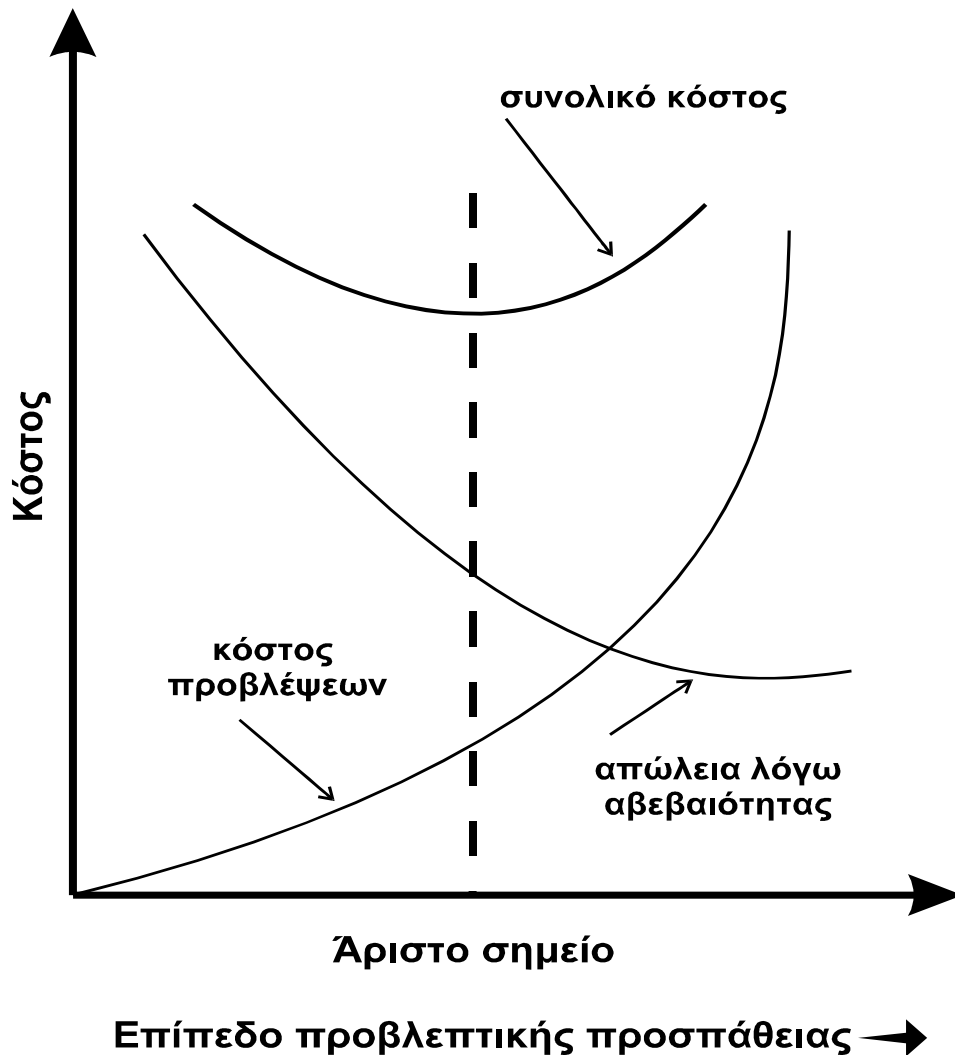
Ο σκοπός της προβλεπτικής διαδικασίας είναι να μειώσει τον κίνδυνο που πάντα υπάρχει στη διαδικασία λήψης απόφασης σε συνθήκες οικονομιών αγοράς .

Οι προβλέψεις συνήθως παρουσιάζουν μικρό ή μεγάλο λάθος , αλλά το μέγεθος του προβλεπικού αυτού λάθους εξαρτάται απ'το σύστημα πρόβλεψης που χρησιμοποιείται . Γενικά η αφιέρωση μεγαλύτερης προσπάθειας στην προβλεπτική διαδικασία οδηγεί στην επίτευξη μεγαλύτερης ακρίβειας , με συνέπεια την μείωση των απωλειών που προκύπτουν εξαιτίας της αβεβαιότητας που υπάρχει στη διαδικασία λήψης αποφάσεων σε συνθήκες οικονομίας αγοράς . Αυτή η έννοια παρουσιάζεται στο παρακάτω διάγραμμα :



# ΔΙΑΓΡΑΜΜΑ 1

## Το κόστος ως συνάρτηση του επιπέδου προβλεπτικής προσπάθειας



Καθώς το κόστος προβλέψεων αυξάνεται, οι απώλειες λόγω αβεβαιότητας μειώνονται. Σε κάποιο επίπεδο προβλεπτικής προσπάθειας το συνολικό κόστος είναι ελάχιστο.

Το εννοιολογικό μοντέλο του διαγράμματος 1 βασίζεται στην υπόθεση της φθίνουσας οριακής χρησιμότητας της πρόβλεψης. Κάθε πρόσθετη δραχμή που ξοδεύεται για την πρόβλεψη καταλήγει σε μικρότερη μείωση των απωλειών αβεβαιότητας από ό,τι η προηγούμενη της δραχμή. Και πέρα από κάποιο

σημείο , πρόσθετη προσπάθεια δεν φέρνει καμιά βελτίωση . Το μέσο λάθος πρόβλεψης είναι αδύνατο να μειωθεί κάτω από ένα συγκεκριμένο επίπεδο , ανεξάρτητα από την πολυπλοκότητα της μεθόδου προβλέψεων που χρησιμοποιείται .

Επειδή η διαδικασία πρόβλεψης ποτέ δεν απομακρύνει τελείως τον κίνδυνο , είναι αναγκαίο στη διαδικασία για τη λήψη απόφασης να παίρνουμε ρητά υπόψη την αβεβαιότητα που περικλείει η πρόβλεψη .

Συχνά η απόφαση σχετίζεται εννοιολογικά με την πρόβλεψη σύμφωνα με το παρακάτω :

***Πραγματική απόφαση = απόφαση υποθέτοντας την πρόβλεψη σωστή + προβλεπτικό λάθος .***

Αυτό σημαίνει ότι ένα σύστημα προβλέψεων πρέπει να δίνει την περιγραφή του προβλεπτικού λάθους ταυτόχρονα με την πρόβλεψη . Στην ιδεώδη περίπτωση , η προβλεπτική διαδικασία πρέπει να καταλήγει στην εκτίμηση της κατανομής πιθανότητας της μεταβλητής που προβλέπεται.

Εντούτοις , στην πράξη είναι σχεδόν αδύνατο να προσδιορίσουμε τη συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής που θα προβλεφθεί . Έτσι περιοριζόμαστε στον προσδιορισμό κάποιου διαστήματος εμπιστοσύνης ή κάποιας συγκεκριμένης τιμής - σημειακής πρόβλεψης - που αντιπροσωπεύει κατά “άριστο” τρόπο τη συγκεκριμένη τυχαία μεταβλητή.

Για να επιτευχθεί η “άριστη” τιμή μιας σημειακής πρόβλεψης , απαιτείται η ύπαρξη κριτηρίου με βάση το οποίο να αξιολογούνται οι διάφορες εναλλακτικές τιμές . Ένα τέτοιο

κριτήριο επιτυγχάνεται με την εισαγωγή της έννοιας της συνάρτησης κόστους . Ας υποθέσουμε ότι το κόστος ενός λάθους μεγέθους  $e$  είναι  $C(e)$  , με  $C(0)=0$ . Αν η σημειακή πρόβλεψη μιας τυχαίας μεταβλητής  $X_{n+h}$  είναι  $f_{n,h}$  με λάθος :

$$\underline{e_{n,h} = X_{n+h} - f_{n,h}}$$

τότε επειδή δεν έχει ληφθεί η άριστη απόφαση , πραγματοποιείται ένα κόστος :

$$(C_{e_{n,h}})$$

έτσι το κριτήριο που απαιτείται διαμορφώνεται ως εξής :

Με δοσμένη μια τέτοια συνάρτηση κόστους , επιλέγεται η σημειακή πρόβλεψη , έτσι ώστε το αναμενόμενο κόστος :

$$\underline{E \{ (C_{e_{n,h}}) \}}$$

να ελαχιστοποιείται θεωρώντας ως δέσμευση το σύνολο πληροφόρησης  $I_n$  , που περικλείει εμπειρικές παρατηρήσεις, θεωρίες και υποθέσεις σχετικά με τις ιδιότητες της υπό μελέτης στοχαστικής διαδικασίας στο χρόνο  $n$  (τώρα) . Σύμφωνα με την κλασική στατιστική θεωρία , η σημειακή πρόβλεψη που ελαχιστοποιεί αυτή τη συνάρτηση κόστους και που αντιστοιχεί στο κριτήριο των ελαχίστων τετραγώνων είναι ο δεσμευμένος μέσος :

$$f_{n,h} = E_c \{ X_{n+h} \}$$

Αυτό το συμπέρασμα εξαρτάται από τη μορφή της συνάρτησης κόστους .

### **2.3 Κατηγορίες προβλεπτικών μεθόδων .**

Μπορούμε να ταξινομήσουμε τους διάφορους τύπους προβλεπτικών μεθόδων σε τρεις μεγάλες κατηγορίες :

- 1.Υποκειμενικές προβλέψεις , οι οποίες βασίζονται στην προσωπική κρίση του ερευνητή .
- 2.Μονομεταβλητές προβλέψεις , οι οποίες βασίζονται ολοκληρωτικά στις τρέχουσες και περασμένες τιμές μιας δοσμένης χρονολογικής σειράς .
- 3.Πολυμεταβλητές προβλέψεις οι οποίες διαμορφώνονται παίρνοντας υπόψη και άλλες μεταβλητές .

Η διαπίστωση πως αρκετές διαδικασίες εξέλιξης είχαν ομοιότητες έκανε δυνατή την ομαδοποίηση και περιγραφή τους με την βοήθεια μαθηματικών τύπων . Έτσι λοιπόν θεωρώντας ότι και οι τηλεπικοινωνίες ανήκουν σε κάποια μεγάλη ομάδα θα προσπαθήσουμε να βρούμε ποια είναι τα κατάλληλα εκείνα μαθηματικά μοντέλα με τα οποία θα μπορούσαμε να αντιμετωπίσουμε τα προβλήματα πρόβλεψης που υφίστανται σ' αυτό το κλάδο .

## **ΚΕΦΑΛΑΙΟ 3**

## ΜΟΝΟΜΕΤΑΒΛΗΤΕΣ ΠΡΟΒΛΕΨΕΙΣ

### 3.1 Γενικά

Στην περίπτωση των μονομεταβλητών υποδειγμάτων η πρόβλεψη συνίσταται στην προβολή ενός αποτελεσματικά εκτιμημένου μονομεταβλητού υποδείγματος στο μέλλον . Οι προβλέψεις που προκύπτουν ονομάζονται μονομεταβλητές , επειδή στη διαμόρφωσή τους χρησιμοποιούνται μόνο τρέχουσες και περασμένες τιμές της σειράς που θα προβλεφθεί . Στην περίπτωση αυτή το σύνολο πληροφόρησης που χρησιμοποιείται είναι μάλλον περιορισμένο .

### 3.2 Διωνυμικά μοντέλα πρόβλεψης

Τα διωνυμικά μοντέλα προέρχονται από την παρακάτω διαφορική εξίσωση ή από επεκτάσεις και μετασχηματισμούς αυτής :

$$\dot{f} = bF \left(1 - \frac{f}{F}\right)^{\hat{a}} \left(\frac{f}{F}\right)^a$$

Όπου  $a, \beta$  παράμετροι .

Η παραπάνω διαφορική εξίσωση προέρχεται από την εμπειρική διαπίστωση ότι η διάδοση μιας καινοτομίας είναι ένα φαινόμενο κατά το οποίο αλληλεπιδρούν το ποσοστό  $f/F$

των ατόμων που έχουν αποδεχθεί την καινοτομία και το ποσοστό  $(I-f/F)$  αυτών που δεν την έχουν ακόμα αποδεχθεί . Οι αδιάστατες παράμετροι  $\alpha, \beta$  εκφράζουν το βαθμό επίδρασης κάθε μιας από τις δυο προηγούμενες κύριες ομάδες στην όλη πορεία της διαδικασίας διάδοσης .

Στην συνέχεια θα αναλύσουμε εν συντομία τα βασικά διωνυμικά μοντέλα τα οποία θα χρησιμοποιήσουμε στο πρώτο μέρος της πρόβλεψής μας .

### 3.3 Βασικά διωνυμικά μοντέλα

Για να γίνουν καλύτερα κατανοητά τα μοντέλα αυτά θα ήταν καλύτερο απ'την αρχή να δώσουμε την σημασία των δεικτών .

Ορίζουμε ως :

- **F** το άνω όριο των δεκτών της όποιας καινοτομίας ή το σύνολο των δεκτών του πληθυσμού που είναι δυνατό να την κάνει αποδεκτή .
- **f<sub>0</sub>** τους αρχικούς αποδέκτες της καινοτομίας στο χρόνο  $t=0$  .
- **f** το σύνολο των δεκτών στο χρόνο μεταξύ  $t$  και  $t+dt$  .
- **f'** το ποσοστό των δεκτών στο χρόνο μεταξύ  $t$  και  $t+dt$  , δηλαδή  $f'=df/dt$  . Το  $f'$  ονομάζεται ακόμη και ρυθμός διάδοσης ή ρυθμός αποδοχής στο χρόνο  $t$  .

Τώρα είμαστε σε θέση να δούμε τα διωνυμικά μοντέλα .

1. **Coleman** . Το μοντέλο αυτό στηρίζεται στις ακόλουθες παραδοχές .

- Το σύνολο των πιθανών δεκτών  $F$  είναι σταθερό .
- Το σύνολο  $F$  αποτελείται από ομοιόμορφο πληθυσμό που δέχεται τυχαία το νεωτερισμό .
- Η πηγή που επηρεάζει τον πληθυσμό , είναι σταθερή και ανεξάρτητη από αυτόν .
- Η επίδραση της σταθερής και ανεξάρτητης πηγής στον πληθυσμό  $F$  είναι σταθερό .
- Ο ρυθμός μίμησης είναι μηδέν .

Η εξίσωση του μοντέλου είναι :

$$\dot{f} = b(F - f)$$

απ' όπου με ολοκλήρωση παίρνουμε :

$$\underline{f = F - (F - f)e^{-bt}}$$

Το μοντέλο Colleman , έχει εφαρμογές όπου οι δέκτες του νεωτερισμού δεν μπορούν να επηρεάσουν τη διαδικασία διάδοσης , όπως στις περιπτώσεις διαφημίσεων προϊόντων από μέσα μαζικής ενημέρωσης ή από πλασιέ κτλ .

**2. Dodd** . Το μοντέλο αυτό εφαρμόζει τη λογιστική καμπύλη  $S$  .

Στηρίζεται στις δύο πρώτες παραδοχές του μοντέλου Colleman και προσθέτει άλλες δύο :

- Όλος ο πληθυσμός των μη δεκτών ( $F-f$ ) , δέχεται το νεωτερισμό μιμούμενος τους ήδη δέκτες  $f$  , με τους οποίους έρχεται σε επαφή .
- Οι δέκτες επικοινωνούν με τους μη δέκτες με ομοιόμορφο τρόπο , ώστε να μην δημιουργούνται εστίες επηρεασμού .

Το μοντέλο Dodd θεωρεί σαν πηγή επηρεασμού το πλήθος  $f$  που έχει ήδη δεχθεί την καινοτομία . Είναι φανερό ότι η πηγή γίνεται ένα αυτορυθμιζόμενο σύστημα που γίνεται ισχυρότερο καθώς αυξάνει το  $f$  . Για σταθερό ολικό πληθυσμό  $F$  , η ικανότητα της πηγής να επηρεάζει θα είναι ανάλογη του λόγου  $f/F$  . Έτσι ορίζουμε τον συντελεστή διάδοσης ή μίμησης  $b$  :

$$\underline{b(t) = b \frac{f}{F}}$$

Η διαφορική εξίσωση του μοντελου Dodd είναι :

$$\underline{\dot{f} = bf \frac{F-f}{F}}$$



από όπου με ολοκλήρωση παίρνουμε :

$$\ln \frac{f}{F - f} = c + bt$$

$$\underline{c = - \ln\left(\frac{F - f_0}{f_0}\right)}$$

**3.Bass.** Το μοντέλο αυτό συνδιάζει τα δύο προηγούμενα και στηρίζεται στις ακόλουθες παραδοχές :

- Οι ηγέτες γνώμης , δηλαδή τα άτομα εκείνα που η γνώμη τους η συμπεριφορά τους και οι προτιμήσεις τους έχουν ιδιαίτερη επίδραση στον υπόλοιπο πληθυσμό , είναι ένα σταθερό ποσοστό του πληθυσμού και δέχονται την καινοτομία με την ίδια διαδικασία που ακολουθείται και για τον υπόλοιπο πληθυσμό . Έτσι οι ηγέτες γνώμης , περιέχονται πάντα κατα το ίδιο ποσοστό , τόσο στον αρχικό πληθυσμό όσο και στους αποδέκτες  $f$  στη συνέχεια της διαδικασίας διάδοσης της καινοτομίας .
- Οι ηγέτες γνώμης , όταν δρουν ως πηγή επηρεασμού , δε μεταβάλλουν το συντελεστή διάδοσης , ενώ συγχρόνως δεν απομακρύνονται από το σύστημα .

Οι δύο αυτές παραδοχές , εξασφαλίζουν τη σταθερότητα του συντελεστή μίμησης και του ολικού πληθυσμού  $F$  .

Η διαφορική εξίσωση του μοντέλου είναι :

$$\dot{f} = a(F-f) + b \frac{F-f}{F} f$$

και με ολοκλήρωση :

$$f = F \frac{1 - a\tilde{a}e^{-(a+b)t}}{1 - b\tilde{a}e^{-(a+b)t}}$$

όπου :

$$\tilde{a} = \frac{F - f_0}{aF + bf_0}$$

—

**4. Floyd** . Το μοντέλο αυτό προϋποθέτει την ύπαρξη κάποιου σχετικού πλεονεκτήματος που έχουν οι αρχικοί αποδέκτες μιας καινοτομίας έναντι των επομένων . Στο μοντέλο αυτό ο συντελεστής διάδοσης μειώνεται χρονικά σύμφωνα με την σχέση :

$$\underline{b(t) = b \left(1 - \frac{f}{F}\right)}$$

Η διαφορική εξίσωση του μοντέλου είναι :

$$\underline{\dot{f} = bf \left(1 - \frac{f}{F}\right)^2}$$

απ' όπου με ολοκλήρωση παίρνουμε :

$$\underline{\ln \frac{f}{F-f} + \ln \frac{F}{F-f} = c + bt}$$

όπου το c υπολογίζεται αν θέσουμε t=0 .

**5 . Sharif-Kabir .** Το μοντέλο αυτό προσπάθησε να βελτιώσει τα μοντέλα Dodd , Floyd με την προσθήκη ενός “συντελεστή καθυστέρησης” σ στην προηγούμενη εξίσωση ως εξής :

$$\underline{\ln \frac{f}{F-f} + \sigma \ln \frac{F}{F-f} = c + bt}$$

για  $0 \leq \delta \leq 1$

ενώ η διαφορική εξίσωση του μοντέλου είναι :

$$\dot{f} = \frac{b}{F} f \frac{(F-f)^2}{F-(1-\delta)f}$$

Το οποίο προσεγγίζεται απ' την εξίσωση :

$$\dot{f} = b \left(1 - \frac{f}{F}\right)^2 f$$

**6. Von-Bertalanfly** . Πρόκειται για ένα μοντέλο που αρχικά χρησιμοποιήθηκε στη βιολογία για να εκφράσει την αύξηση του βάρους των ζώων . Η εξίσωση του μοντέλου είναι :

$$f = F \left(1 - a e^{-bt}\right)^{\frac{1}{m}}$$

όπου  $m$  μια θετική παράμετρος . Η σταθερά  $a$  ορίζεται αν θέσουμε στην πιο πάνω εξίσωση  $t=0$  . Η διαφορική εξίσωση προκύπτει απ' την προηγούμενη με παραγωγή και γίνεται:

$$\dot{f} = b \left( \left( \frac{f}{F} \right)^{-m} - 1 \right) F$$

**7. Grm I .** Το μοντέλο αυτό χρησιμοποιεί μια ρητή διαφορική εξίσωση , επέκταση αυτής του Sharif-Kabir . Ο μεταβαλλόμενος συντελεστής διάδοσης του μοντέλου είναι :

$$b(f) = bf \frac{(F-f)^{n-1}}{F^{n-1}(F-(1-\sigma)f)}$$

και η διαφορική εξίσωση :

$$\dot{f} = b(f) (F - f)$$

όπου  $n=1,2,3,4,\dots$  και  $\sigma$  ανήκει στο  $[0,1]$  .

**8. S-3 .** Το μοντέλο αυτό έχει την ακόλουθη διαφορική εξίσωση :

$$\dot{f} = \mu \frac{f^2}{f^2 + kf + k^2}$$

με  $\mu = bk$

απ' όπου ολοκληρώνοντας παίρνουμε :

$$\underline{k \ln f + f - \frac{k^2}{f} = k \ln f_0 + f_0 - \frac{k_2}{f_0} + \dot{\lambda} t}$$

### 3.4 Αξιολόγηση μοντέλων .

Για να γίνει αξιολόγηση ανάμεσα στα μοντέλα μπορούμε να χρησιμοποιήσουμε ήτε το συντελεστή προσδιορισμού ( $R^2$ ) ήτε το μέσο τετραγωνικό σφάλμα (M.S.E) .

- Ο συντελεστής προσδιορισμού εκφράζει το ποσό που προσεγγίζει τα πραγματικά δεδομένα η καμπύλη που προκύπτει απ' το μοντέλο . Δίνεται απ' τον τύπο :

$$R^2 = 1 - \frac{\sum_{t=0}^n (y - \hat{f})^2}{\sum_{t=0}^n (y - \bar{y})^2}$$

όπου  $y$  είναι τα πραγματικά δεδομένα ,  $\hat{f}$  οι υπολογιζόμενες τιμες απ' την ανάλυση και  $\bar{y}$  η μέση τιμή των  $y$  αντίστοιχα .

## ΚΕΦΑΛΑΙΟ 4

### Εφαρμογή στο πακέτο Forbim

#### 4.1 Γενικά

Ένα απ'τα πακέτα που έχουν φτιαχθεί έτσι ώστε να εφαρμόζονται στην πρόβλεψη χρονολογικών σειρών χρησιμοποιώντας διωνυμικά μοντέλα είναι και το Forbim (Forecasting binomial models) .

Όπως έχουμε εξηγήσει και στον πρόλογο μέσα απ'την παρούσα διπλωματική θα γίνει μια προσπάθεια πρόβλεψης των αναγκών σε τηλεφωνικά κέντρα και πιο συγκεκριμένα της εγκατεστημένης χωρητικότητας των τηλεφωνικών κεντρων .

Πέρα απ'αυτήν όμως την χρονολογική σειρά θα χρησιμοποιήσουμε στην πρόβλεψή μας και τις παρακάτω για την σωστότερη αντιμετώπιση του προβλήματος .

Έτσι θα χρησιμοποιηθούν :

- Ο συνολικός αριθμός τηλεφώνων στην Ελλάδα .
- Ο αριθμός των ανικανοποίητων αιτήσεων .
- Και το κατά κεφαλήν ακαθάριστο εθνικό εισόδημα (σε σταθερες τιμές 1970) .

Στη συνέχεια γίνεται εφαρμογή του πακέτου για κάθε μια απ' τις χρονολογικές σειρές ξεχωριστά .

Η διαδικασία που ακολουθείται είναι η εξής :

Για κάθε μια απ' τις σειρές δίνονται τα στοιχεία και κατόπιν γίνεται το γράφημα τόσο της εξέλιξης αυτής της σειράς στο χρόνο όσο και το γράφημα της παραγώγου στο χρόνο απ' το οποίο μπορεί να δει κανείς κατά πόσο ήταν ομαλή η εξέλιξη αυτή .

Αφού γίνει αυτό δίνουμε σε κάθε μοντέλο ξεχωριστά στοιχεία λιγότερα απ' αυτά που έχουμε διαθέσιμα συνήθως 3 με 5 και παρατηρούμε κατά πόσο οι προβλέψεις πλησιάζουν τις πραγματικές τιμές . Στη συνέχεια χρησιμοποιώντας το μέσο τετραγωνικό σφάλμα (MSE) το συντελεστή προσδιορισμού ( $R^2$ ) και το ποσοστό επιτυχίας ( προβλεπόμενη τιμή / πραγματική τιμή ) προσπαθούμε να βρούμε ποιο είναι εκείνο το μοντέλο που κάνει τη καλύτερη προσαρμογή στα δεδομένα μας .

Εφαρμόζοντας τα παραπάνω σε κάθε μια απ' τις χρονοσειρές που εξετάζουμε παίρνουμε τους πίνακες και τα γραφήματα που παρουσιάζουμε στη συνέχεια βάση των οποίων γίνεται η πρόβλεψη :



## 4.2 Συμπεράσματα .

Παρατηρώντας κανείς τους προηγούμενους πίνακες μπορεί να πει τα εξής :

1. Για τον συνολικό αριθμό τηλεφώνων στην Ελλάδα η εξέλιξη είναι σχετικά ομαλή γι'αυτό και τα μοντέλα μας δουλεύουν αρκετά καλά εκτός βέβαια απ'τον Von Bert . Έτσι για την πρόβλεψη μας θα χρησιμοποιήσουμε το μοντέλο Coleman γιατί αν και έχει τον ίδιο αριθμό αστεριών με το GRM 1 δηλαδή τον ίδιο αριθμό καλών προβλέψεων έχει καλύτερο συντελεστή συσχέτισης και καλύτερο μέσο τετραγωνικό σφάλμα . Στη συνέχεια λοιπόν δίνουμε και τις 24 παρατηρήσεις στο μοντέλο Coleman και παίρνουμε τις προβλέψεις για τα επόμενα δέκα χρόνια .
2. Όσον αφορά την εγκατεστημένη χωρητικότητα των κέντρων παρατηρούμε ότι και αυτή έχει μια σχετικά καλή εξέλιξη στο χρόνο . Και σε αυτήν την χρονολογική σειρά τα μοντέλα του Forbim δουλεύουν αρκετά καλά (εκτός του Von Bert) . Παρατηρούμε δε ότι αν και το GRM 1 έχει το

μικρότερο μέσο τετραγωνικό σφάλμα καθώς και το μικρότερο συντελεστή συσχέτισης έχει συνολικά τα περισσότερα αστέρια . Έτσι λοιπόν χρησιμοποιούμε το GRM 1 για την πρόβλεψη των επόμενων δέκα χρόνων .

3. Για την χρονοσειρά τώρα των αιτήσεων υπό εκκρεμότητα παρατηρούμε απ' το γράφημα ότι ενώ μέχρι το 1990 είχαμε κάποια αύξηση μετά οι τιμές μιώνονται σημαντικά σχεδόν κατακόρυφα . Το φαινόμενο αυτό εξηγείται απ' το γεγονός ότι στο χρονικό εκείνο σημείο άρχισε η εισαγωγή ψηφιακών τηλεφωνικών κέντρων , των οποίων η δυνατότητα σε παροχές τηλεφωνικών γραμμών είναι τεράστια . Για το λόγο αυτό η εφαρμογή των διωνυμικών μοντέλων σ' αυτή τη χρονοσειρά παρουσιάζει πρόβλημα καθώς κανένα απ' τα μοντέλα όπως φαίνεται και στα γραφήματα αλλά και στους πίνακες των αποτελεσμάτων δεν μπορούν να προβλέψουν την μείωση των αιτήσεων . Έτσι λοιπόν είναι καλύτερα να μην χρησιμοποιήσουμε το Forbim για την πρόβλεψη αυτής της χρονοσειράς

4. Εξετάζοντας και την τελευταία χρονοσειρά που μας ενδιαφέρει δηλαδή το κατά κεφαλήν ακαθάριστο εθνικό εισόδημα παρατηρούμε ότι τα μοντέλα μας μπορούν να κάνουν μια σχετικά καλή προσαρμογή και κατ' επέκταση καλή πρόβλεψη .

Γενικά μπορούμε να πούμε ότι το Forbim εκτός από εξαιρέσεις μπορεί να κάνει μια καλή μονοδιάστατη πρόβλεψη .

## **ΚΕΦΑΛΑΙΟ 5**

### **Μονομεταβλητές προβλέψεις στο EXCEL**

#### **5.1 Εφαρμογή**

Το EXCEL μας δίνει τη δυνατότητα να χρησιμοποιήσουμε μια σειρά από μαθηματικά μοντέλα στην μονοδιάστατη πρόβλεψη χρονοσειρών .

Στη συνέχεια γίνεται εφαρμογή των μοντέλων αυτών στην χρονοσειρά της εγκατεστημένης χωρητικότητας των κέντρων (I.C.T.C) απ' όπου παίρνουμε τα παρακάτω διαγράμματα :

## Συμπεράσματα

Όσον αφορά το λογαριθμητικό μοντέλο βλέπουμε ότι η προσαρμογή της καμπύλης δεν είναι ιδιαίτερα αξιόλογη εξού και ο μικρός συντελεστής συσχέτισης μόλις 0.775 .

Στη συνέχεια το εκθετικό μοντέλο μπορούμε να πούμε ότι κάνει μια αρκετά καλή προσαρμογή στα δεδομένα μας φθάνοντας το συντελεστή συσχέτισης στο 0.9934 .

Ενδιαφέρον παρουσιάζει το πολυωνυμικό μαθηματικό μοντέλο στο οποίο μάλιστα έχοντας την δυνατότητα να αλλάξουμε το βαθμό μπορούμε να καταφέρουμε πολύ καλή εφαρμογή στα δεδομένα μας . Έτσι λοιπόν ξεκινώντας μ'ένα πολυώνυμο 3ου βαθμού έχουμε συντελεστή συσχέτισης 0.9962 με 5ου 0.9984 και τέλος 6ου καταφέρνουμε να φθάσουμε το 0.9992 . Παρατηρούμε όμως ότι το πολυώνυμο

του 6ου βαθμού παρόλο που κάνει πάρα πολύ καλή προσαρμογή στα δεδομένα μας δεν μπορεί να μας δώσει μια μακροπρόθεσμη πρόβλεψη γιατί μετά από 6 περιόδους πρόβλεψης μας δίνει αρνητικές τιμές . Θα μπορούσαμε όμως να χρησιμοποιήσουμε το μοντέλο αυτό για μικρό χρονικό διάστημα .

## **ΚΕΦΑΛΑΙΟ 6**

### **Προβλέψεις του Ο.Τ.Ε**

#### **6.1 Γενικά**

Ο Ο.Τ.Ε στην προσπάθειά του για ανάπτυξη έχει εφεύρει αρκετές μεθόδους με τις οποίες προσπαθεί να προβλέψει τις μελλοντικές ανάγκες του οργανισμού .

Μια απ' τις σημαντικότερες ανάγκες είναι και η ζήτηση τηλεφωνικών συνδέσεων στην χώρα .

Για το λόγο αυτό η υποδιεύθυνση βασικού προγραμματισμού έχει αναπτύξει ένα προβλεπτικό υπολογιστικό πακέτο το EST-DEM .

## 6.2 Διαδικασία πρόβλεψης

Η διαδικασία που ακολουθείται στην πρόβλεψη απ' τον Ο.Τ.Ε είναι η ακόλουθη :

- 1.Επιλογή κέντρου . Αποφασίζεται δηλαδή σε ποιο απ' τα τηλεφωνικά κέντρα ή καλύτερα ακόμα σε ποια απ' τις περιοχές που ορίζονται απ' τα τηλεφωνικά κέντρα θέλουμε να γίνει η πρόβλεψη .
- 2.Καθορισμός ορίων απ' το βασικό σχέδιο τηλεφωνίας . Απ' τους χάρτες δηλαδή που έχει σχηματίσει ο Ο.Τ.Ε .
- 3.Χωρισμός ζωνών . Για την καλύτερη μελέτη της περιοχής χωρίζεται σε περιοχές όπως παραλία , κεντρο πόλεως , αγροτική περιοχή κτλ .
- 4.Υπολόγισμός τωρινής και μελλοντικής κατάστασης . Αυτό γίνεται αφού επιλέξουν δείγμα 30-40% από σπίτια επαγγελματικές στέγες εξοχικές κατοικίες κτλ και “περπατήσουν” οι ίδιοι την περιοχή .
- 5.Συλλογή στοιχείων από άλλες υπηρεσίες . Παίρνουν στοιχεία απ' τα ρολόγια της ύδρευσης δηλαδή πόσα υπάρχουν πόσα καινούργια θα μπουύν κτλ . Στοιχεία δίνει επίσης και η Δ.Ε.Η απ' τα δικά της ρολόγια .

Για τα χωριά οι υπεύθυνοι του Ο.Τ.Ε πηγαίνουν οι ίδιοι στις κοινότητες και βλέπουν αν έχει σχολείο , γέρους μαγαζιά κτλ .

### 6.3 Πακέτο EST-DEM .

Το λογισμικό πακέτο EST-DEM αναπτύχθηκε από τον τομέα μηχανογράφησης υποστήριξης σχεδιασμού τηλεπικοινωνιακών δικτύων της διεύθυνσης προγραμματισμού με τη συνεργασία των στελεχών της υποδιεύθυνσης βασικού προγραμματισμού και ειδικότερα του τομέα προβλέψεων συνδρομητών , αντικείμενο του οποίου είναι η πρόβλεψη ζήτησης κύριων τηλεφωνικών συνδέσεων . Το χρησιμοποιούμενο μαθηματικό μοντέλο είναι το περιγραφόμενο απ'την εκθετική υπολογιστική συνάρτηση . Η όλη μεθοδολογία βασίζεται στα ιστορικά στοιχεία του πληθυσμού στα ιστορικά στοιχεία ζήτησης κύριων τηλεφωνικών συνδέσεων προηγούμενων ετών καθώς και στην πυκνότητα κόρου .

Για την ακρίβεια χρησιμοποιείται η συνάρτηση :

$$\frac{D(t)}{D_{\max}} = \frac{1}{(1+e^{-c(t-t_0)})^{1/m}}$$

όπου  $D(t)$  : η τηλεφωνική πυκνότητα τη χρονική στιγμή  $t$ .

$D_{\max}$  : είναι η μέγιστη τηλεφωνική πυκνότητα στο “έτος κόρου” δηλαδή  $D(t)/D_{\max}=f/F$  .

$t_0$  : η αρχή των χρόνων

$t$  : ο χρόνος .

$c$  ,  $m$  : παράμετροι οι οποίοι υπολογίζονται από την ανάλυση παλινδρόμησης . Για την εφαρμογή της συνάρτησης απαιτείται

αρχικά ο υπολογισμός ενός άνω ορίου , της αποκαλούμενης “ πυκνότητας κόρου “ ( $D_{max}$ ) , για το έτος στο οποίο καταλήγει η πρόβλεψη . Πρέπει να σημειωθεί για τα στοιχεία που χρησιμοποιούνται κατά την ανάλυση παλινδρόμησης ότι πρόκειται βασικά για το λόγο :

Τηλεφωνα +εκκρεμείς αιτήσεις

---

νοικοκυριά + εξοχικά + επαγγελματικές στέγες

Στο πακέτο αυτό εξάλλου δίνεται η δυνατότητα για πρόβλεψη πληθυσμού μετά το έτος αρχικής μελέτης σύμφωνα με το μαθηματικό τύπο :

$$P_n = P_0 \cdot (1+r)^n + c \cdot [(1+r)^n - 1] / r$$

Συνολικά τα στοιχεία εισόδου που πρέπει να δώσουμε στο πρόγραμμα είναι τα παρακάτω :

- Ο αρχικός πληθυσμός  $P_0$  για το έτος μελέτης (πχ. 2750) , η μέση ετήσια φυσιολογική αύξηση του πληθυσμού  $r$  (πχ 0.01) , ο μέσος αριθμός μετακινουμένων κατοίκων  $c$  (πχ 100) , το αρχικό έτος υπολογισμών ( πχ 1993) και το πλήθος έτων  $n$  για τους υπολογισμούς (πχ 70) . Το πρόγραμμα δίνει τη δυνατότητα να χρησιμοποιήσουμε διαφορετική τιμή της σταθεράς  $c$  , για τη χρονική περίοδο που έχουμε επιλέξει . Σε μια τέτοια περίπτωση ο υπολογισμός ,της σταθεράς αυτής γίνεται απ’τον τύπο :



$$c = r*(Pn - Po(1+r)^n)/[(1+r)^n - 1]$$

Τα αποτελέσματα εξάλλου της πρόβλεψης με αυτό το πακέτο κατά περιφέρεια αλλά και συνολικά για την χώρα μας φαίνονται παρακάτω :

## **ΚΕΦΑΛΑΙΟ 7**

### **Πρόβλεψη συσχετίζοντας χρονοσειρές**

## 7.1 Εισαγωγικά

Δύο ή και περισσότερες ιδιότητες του ίδιου συστήματος ή ατόμου είναι δυνατόν να σχετίζονται ποσοτικά : πχ η ηλικία και το βάρος ενός παιδιού έχουν κάποια θετική συσχέτιση (όσο πιο μεγάλο είναι το παιδί τόσο πιο βαρύ είναι ) και λέμε ότι το βάρος είναι αύξουσα συνάρτηση της ηλικίας του παιδιού . Η απλούστερη μορφή συσχέτισης είναι η γραμμική ( δηλαδή ένα  $X$  και ένα  $Y$  ) ενώ καθώς αυξάνονται οι ιδιότητες που συσχετίζονται αυξάνεται και ο βαθμός της εξίσωσης της συσχέτισης .

Εξετάζοντας την πιο απλή μορφή της συσχέτισης δηλαδή την μορφή  $Y=\beta X+\alpha$  πρέπει να πούμε ότι λέγεται ευθεία γραμμικής παλινδρόμησης . Εάν η παρατήρηση  $Y$  δεν υπόκειται σε σφάλματα δηλαδή αν για κάποια τιμή της ανεξάρτητης μεταβλητής  $X$  , μπορούμε να προβλέψουμε ακριβώς το  $Y$  , τότε το μοντέλο καλείται προσδιοριστικό . Στην πραγματικότητα όμως δεν μπορούμε να κάνουμε ποτέ ακριβής πρόβλεψη .

## 7.2 Μέθοδος ελαχίστων τετραγώνων .

Ένα μοντέλο που δίνει τη δυνατότητα στο  $Y$ , να μη βρίσκεται ακριβώς στην ευθεία  $Y=\beta X+\alpha$  είναι το :

$$Y=\alpha+\beta X+e$$

όπου  $e$  είναι ένα τυχαίο σφάλμα και παριστάνει τη διαφορά της παρατηρούμενης τιμής  $Y$  για δοσμένο  $X$ , απ'τη θεωρητική τιμή  $\alpha+\beta X$ . Αυτού του είδους τα μοντέλα καλούνται στοχαστικά.

Όσον αφορά τα σφάλματα  $e$  υποθέτουμε ότι είναι τυχαία με μέση τιμή  $E(e)=0$ . Έτσι, αφού τα  $\alpha$ ,  $\beta$  είναι άγνωστες σταθερές ισχύει :

$$E(Y)=\alpha+\beta X$$

Το πρόβλημα έγκυται στο να προσδιορίσουμε τους συντελεστές  $\alpha$ ,  $\beta$ .

Υπάρχουν πολλοί τρόποι να προσδιορίσουμε έναν εκτιμητή της ευθείας  $E(Y)=\alpha+\beta X$  που παριστάνεται :

$$\hat{O} = \hat{\alpha} + \hat{\beta} X$$

Ο προσδιορισμός των εκτιμητών γίνεται με τη μέθοδο των ελαχίστων τετραγώνων διότι μας δίνει αμερόληπτους εκτιμητές ελάχιστης διασποράς και με την υπόθεση ότι τα σφάλματα ακολουθούν κανονική κατανομή, οι εκτιμητές αυτοί είναι επίσης εκτιμητές μέγιστης πιθανοφάνειας.

Έτσι λοιπόν αν κάνουμε  $n$  παρατηρήσεις έχουμε :

$$Y_i = \alpha + \beta X_i + e_i \quad i=1,2,3,4, \dots, n$$

Θέλουμε να προσδιορίσουμε τα  $\alpha$ ,  $\beta$ , έτσι ώστε το άθροισμα των τετραγώνων των σφαλμάτων :

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - a - \hat{a}x_i)^2$$

να είναι ελάχιστο . Παραγωγίζοντας αυτό το άθροισμα ως προς  $a$ ,  $\beta$  καταλήγουμε στο παρακάτω σύστημα εξισώσεων με δύο αγνώστους που καλούνται κανονικές εξισώσεις .

$$\sum_{i=1}^n y_i = na + \hat{a} \sum_{i=1}^n x_i$$

$$\sum_{i=1}^n x_i y_i = a \sum_{i=1}^n x_i + \hat{a} \sum_{i=1}^n x_i^2$$

Η λύση αυτού του συστήματος ως προς  $a$  και  $\beta$  δίνει :

$$\hat{a} = \frac{S_{xy}}{S_x^2}$$

$$\hat{a} = \bar{y} - \hat{a} \bar{x}$$

όπου :

$$S_{xy} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{n-1}$$

$$S_x^2 = \frac{\sum_{i=1}^n x_i^2 - n \bar{x}^2}{n-1}$$

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

$$S_y^2 = \frac{\sum_{i=1}^n y_i^2 - n \bar{y}^2}{n-1}$$

Η εκτίμηση :  $\hat{y} = \hat{a} + \hat{a}$  της ευθείας παλινδρόμησης , καλείται ευθεία ελαχίστων τετραγώνων απ'το τρόπο υπολογισμού των συντελεστών της .

### 7.3 Γενικά-συντελεστής συσχέτισης .

Ήδη μελετήθηκε η εξάρτηση μιας τ.μ  $y$  από μια άλλη τυχαία μεταβλητή  $x$  που την θεωρήσαμε ανεξάρτητη ή ελεγχόμενη . Είναι όμως δυνατόν να είναι τυχαίες μεταβλητές τόσο το  $y$  όσο και το  $x$  .

Σ'αυτήν την περίπτωση δεν είναι προφανές ποια απ'τις μεταβλητές είναι η ανεξάρτητη και ποια η εξαρτημένη π.χ αν θέλουμε να εξετάσουμε τη σχέση που συνδέει τις ηλικίες των συζύγων σ'ένα ζευγάρι , μπορούμε να θεωρήσουμε σαν ανεξάρτητη μεταβλητή την ηλικία της γυναίκας ή του άνδρα , χωρίς κανένα περιορισμό αφού και οι δύο ηλικίες είναι τ.μ . Κατ'αυτόν τον τρόπο θεωρούμε δύο ευθείες γραμμικής παλινδρόμησης , μία του  $y$  στο  $x$  και μία του  $x$  στο  $y$  ως εξής :

$$\hat{y} = \hat{a} + \hat{a} x$$

$$\hat{x} = \hat{a} + \hat{b} y$$

ì ð

$$\hat{a} = \frac{S_{xy}}{S_x^2}$$

$$\hat{a} = \bar{y} - \hat{a} \bar{x}$$

$$\hat{b} = \frac{S_{xy}}{S_y^2}$$

$$\hat{a} = \bar{x} - \hat{b} \bar{y}$$

Αποδεικνύεται πολύ εύκολα ότι :

$$(\hat{y} - \bar{y}) = \hat{a}(x - \bar{x})$$

&

$$(\hat{x} - \bar{x}) = \hat{b}(y - \bar{y})$$

Πρέπει να σημειωθεί ότι η δύο αυτές ευθείες τέμνονται στο κέντρο βάρους του σμήνους των σημείων .

Στην περίπτωση που οι μεταβλητές  $x$  και  $y$  είναι και οι δύο τυχαίες μεταβλητές , τότε σαν μέτρο συσχέτισης των ορίζεται η ποσότητα :

$$\tilde{r} = \frac{\text{cov}(x,y)}{\sqrt{\text{var } x \text{ var } y}}$$

ή

$$\text{cov}(x,y) = E(x - \bar{x})(y - \bar{y})$$

που καλείται συντελεστής θεωρητικής συσχέτισης .

#### 7.4 Μοντέλα πολλών ανεξάρτητων μεταβλητών .

Ήδη μελετήσαμε το γραμμικό μοντέλο μιας ανεξάρτητης μεταβλητής . Είναι όμως πολύ πιθανό , η μεταβλητή  $y$  , να εξαρτάται από περισσότερες ανεξάρτητες μεταβλητές π.χ η απόδοση  $y$  ενός δένδρου μπορεί να εξαρτάται όχι μόνο απ' την ποσότητα του λιπάσματος  $x_1$  αλλά και από την ποσότητα του νερού  $x_2$  τις ώρες ηλιοφάνειας  $x_3$  κ.ο.κ.

Για το λόγο αυτό κατασκευάζουμε ένα μοντέλο το γενικό γραμμικό μοντέλο το οποίο συνδέει τα αποτελέσματα του πειράματος  $y$  , με ένα σύνολο  $k$  ανεξαρτήτων μεταβλητών :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + e$$



όπου  $\chi_1, \chi_2, \dots, \chi_k$  ανεξάρτητες ή ελεγχόμενες μεταβλητές που η μέτρησή τους δεν υπόκειται σε σφάλματα ,

$\beta_0, \beta_1, \dots, \beta_k$  άγνωστοι παράμετροι ,  
e τυχαία σφάλματα με  $E(e)=0$  .

### 7.5 Συντελεστής πολλαπλής συσχέτισης .

Αν θεωρήσουμε το γενικό γραμμικό μοντέλο της προηγούμενης παραγράφου τότε σαν **συντελεστή προσδιορισμού** ορίζουμε την ποσότητα :

$$R^2 = \frac{SS_y - SSE}{SS_y},$$

$$0 < R^2 < 1$$

ή

$$SS_y = \sum_{i=1}^n (y_i - \bar{y})^2$$

$$SSE = \sum_{i=1}^n (y_i - \hat{y})^2$$

και παριστάνει το ποσοστό της ολικής μεταβλητότητας του y που υπολογίζεται από τις ανεξάρτητες μεταβλητές  $\chi_1, \chi_2, \chi_3, \dots, \chi_k$  ..

Η τετραγωνική ρίζα  $R$  του συντελεστή προσδιορισμού καλείται **συντελεστής πολλαπλής συσχέτισης** και είναι ένα μέτρο συσχέτισης του  $y$  και των  $\chi_1, \chi_2, \dots, \chi_k$ .

## 7.6 Εφαρμογή

Θα προσπαθήσουμε στη συνέχεια να εφαρμόσουμε την μέθοδο πολλαπλής συσχέτισης στις χρονοσειρές που εξετάζουμε. Ποιο συγκεκριμένα θα εξετάσουμε τις χρονοσειρές: κατα κεφαλήν ακαθάριστο εθνικό εισόδημα (**National Private Consumption Expenditure**), συνολικό αριθμό τηλεφώνων στην Ελλάδα (**Total Number of Telephones in Greece**), αριθμό αιτήσεων υπό εκκρεμότητα για νέες τηλεφωνικές συνδέσεις (**Number of Applications for New Telephones Connections**) και κατά πόσο μπορούν οι παραπάνω χρονοσειρές να θεωρηθούν ανεξάρτητες και να μας βοηθήσουν στην πρόβλεψη της εγκατεστημένης χωρητικότητας των τηλεφωνικών κέντρων (**Installed Capacite of Telephones Centers**) θεωρώντας την τελευταία εξαρτημένη.

Στον επόμενο πίνακα δίνουμε τα δεδομένα των χρονοσειρών για τις χρονιές 1973 έως 1993. Θα πρέπει εδώ να παρατηρήσουμε ότι ο συνολικός αριθμός τηλεφώνων ξεπερνάει την εγκατεστημένη χωρητικότητα γιατί υπάρχουν και αρκετά χειροκίνητα τηλέφωνα κάτι βέβαια που δεν μας ενοχλεί στην μελέτη μας.

Τα δεδομένα αυτά τα τοποθετούμε στο EXCEL στο οποίο μπορούμε αυτόματα να πάρουμε τα αποτελέσματα που μας ενδιαφέρουν.

Κατ'αρχήν θα χρησιμοποιήσουμε την συνάρτηση `Linest` η οποία υπολογίζει την ευθεία γραμμή που προσαρμόζεται καλύτερα στα δεδομένα μας και επιστρέφει μια μήτρα που περιγράφει αυτή την ευθεία . Η εξίσωση της ευθείας αυτής είναι :

$$y = m_1x_1 + m_2x_2 + \dots + b$$

Όπου η εξαρτημένη τιμή  $y$  είναι μια συνάρτηση των ανεξάρτητων τιμών  $x$  και  $b$  είναι μια σταθερή τιμή .

Η μήτρα αυτή έχει την παρακάτω μορφή :

$m_n$	$m_{n-1}$	.....	$m_2$	$m_1$	$b$
$se_n$	$se_{n-1}$	.....	$se_2$	$se_1$	$se_b$
$r^2$	$se_y$				
$F$	$df$				
<code>ssreg ssresid</code>					

όπου

$se_n$  οι τιμές των τυπικών σφαλμάτων που αντιστοιχούν στους συντελεστές  $m_1, m_2, \dots, m_n$  .

$se_b$  Το τυπικό σφάλμα για τη σταθερά  $b$  .

$r^2$  Ο συντελεστής πολλαπλής συσχέτισης .

$se_y$  Το τυπικό σφάλμα για την εκτίμηση του  $y$  .

F Η στατιστική μεταβλητή F την οποία χρησιμοποιούμε για να προσδιορίσουμε αν η παρατηρούμενη σχέση μεταξύ των εξαρτημένων και των ανεξάρτητων μεταβλητών εμφανίζεται κατά τύχη .

df Οι βαθμοί ελευθερίας .

SS<sub>reg</sub> Το άθροισμα των τετραγώνων της παλινδρόμησης .

SS<sub>resid</sub> Το άθροισμα των τετραγώνων των υπολοίπων .

Παρακάτω φαίνονται τα αποτελέσματα που μας επιστρέφει η Linest .

Σύμφωνα με τα παραπάνω αποτελέσματα μπορούμε να γράψουμε την εξίσωση της πολλαπλής παλινδρόμησης :

$$y = -21,758 * x_1 + 1,0029 * x_2 - 0,14057 * x_3 + 512701,6$$

όπου

**x1 N.P.C.E**

**x2 T.N.T.GR**

**x3 N.A.N.T.C**

Παρατηρούμε επίσης ότι ο συντελεστής πολλαπλής συσχέτισης είναι αρκετά καλός αφού φθάνει το **0,993098** .

Θα μπορούσε βέβαια κάποιος να ισχυριστεί ότι η στενή αυτή σχέση μεταξύ των ανεξάρτητων μεταβλητών και της εγκατεστημένης χωρητικότητας οφείλεται στην τύχη , για το λόγο αυτό χρησιμοποιούμε την στατιστική μεταβλητή F .

Για να συμπεράνουμε ότι υπάρχει σχέση μεταξύ των μεταβλητών , πρέπει η παρατηρούμενη μεταβλητή F να είναι μεγαλύτερη απ'την κρίσιμη τιμή της F . Την κρίσιμη τιμή της F μπορούμε να την βρούμε στους πίνακες του παραρτήματος.

Για να διαβάσουμε τον πίνακα -έστω πως πρόκειται για

μονόπλευρο έλεγχο - χρησιμοποιούμε για το Alpha μια τιμή 0.05 και για τους βαθμούς ελευθερίας (που συμβολίζονται με  $\nu_1$  και  $\nu_2$ ) χρησιμοποιούμε  $\nu_1=k=3$  και  $\nu_2=n-(k+1)=21-(3+1)=17$  όπου  $k$  είναι το πλήθος των μεταβλητών στην παλινδρόμηση και  $n$  το πλήθος των σημείων δεδομένων. Η κρίσιμη τιμή της  $F$  είναι **3,20**.

Η παρατηρούμενη τιμή της  $F$  είναι **815,2938** που είναι κατά πολύ μεγαλύτερη απ'την κρίσιμη τιμή της.

Ένας άλλος έλεγχος υποθέσεων προσδιορίζει αν ο κάθε συντελεστής κλίσης είναι κατάλληλος για την εκτίμηση της εγκατεστημένης χωρητικότητας των τηλεφωνικών κέντρων.

Έτσι για παράδειγμα για να ελέγξουμε το συντελεστή των υπό εκκρεμότητα αιτήσεων ως προς τη στατιστική του σημαντικότητα διαιρούμε το -0,14057 (συντελεστής κλίσης) με το 0,109537 απ'την οποία διαίρεση μπορούμε να βρούμε την παρατηρούμενη στατιστική μεταβλητή  $t$ .

$$t = -0,14057 / 0,109537 = -1,2833$$

Αν συμβουλευτούμε τον πίνακα στατιστικής του παραρτήματος για την μεταβλητή  $t$  για μονόπλευρο έλεγχο με 17 βαθμούς ελευθερίας και  $\text{Alpha}=0.05$  βρίσκουμε ότι είναι 1.740 πολύ μεγαλύτερο απ'την απόλυτη τιμή του παρατηρούμενου  $t$  που είναι 1,2833. Αυτό σημαίνει ότι αυτή η μεταβλητή δεν είναι σημαντική για την εκτίμηση της εγκατεστημένης χωρητικότητας άρα θα μπορούσαμε να μην την λάβουμε υπόψη κατά την πρόβλεψη.

Ακολούθως για το συνολικό αριθμό τηλεφώνων και για το κατά κεφαλήν ακαθάριστο εθνικό εισόδημα έχουμε ότι η παρατηρούμενες τιμές της  $t$  είναι 0,640 και 9.5164 απ'τις οποίες μόνο αυτή του κατά κεφαλήν ακαθάριστου εθνικού εισοδήματος είναι κατά απόλυτη τιμή μεγαλύτερη της τιμής

1.740 το οποίο σημαίνει ότι θα μπορούσαμε να χρησιμοποιήσουμε μόνο αυτή τη μεταβλητή για την πρόβλεψή μας .

Στη συνέχεια δίνονται τιμές και στις τρεις μεταβλητές μας για τα έτη 1994 έως 2003 και παίρνουμε την πρόβλεψη της εγκατεστημένης χωρητικότητας για τα αντίστοιχα έτη χρησιμοποιώντας και τις τρεις μεταβλητές στην αρχή και μετά μόνο το κατά κεφαλήν ακαθάριστο εθνικό εισόδημα (linest 1 ,linest 2) .

Στη συνέχεια θα προσπαθήσουμε να κάνουμε προσαρμογή στα δεδομένα μας με μια εκθετική καμπύλη . Η συνάρτηση που μας παρέχει τη δυνατότητα αυτή ονομάζεται Logest .

Παρακάτω βλέπουμε τα αποτελέσματα της προσαρμογής αυτής σε μορφή πίνακα όπως ακριβώς και της Linest .

Όπως και πριν έτσι και τώρα παρατηρούμε το συντελεστή πολλαπλής συσχέτισης που είναι σχετικά καλός αφού φθάνει το 0,986735 .

Όσον αφορά τώρα τη σχέση των μεταβλητών μας βρίσκουμε ότι η παρατηρούμενη τιμή της F είναι 421,5262 ενώ η κρίσιμη τιμή της όπως προκύπτει απ'τους στατιστικούς πίνακες είναι 3,24 το οποίο σημαίνει ότι η εξίσωση της παλινδρόμησης είναι κατάλληλη για την εκτίμηση της εγκατεστημένης χωρητικότητας .

Η εξίσωση λοιπόν βάση της οποίας μπορεί να γίνει η πρόβλεψη και όπως προκύπτει απ'τον πίνακα είναι :

$$\ln y = x_1 \cdot \ln m_1 + x_2 \cdot \ln m_2 + \dots + x_n \ln m_n + \ln b$$

ή

$$\ln y = x_1 \cdot \ln 1,00001 + \ln 825225,8$$

Παρατηρούμε δε ότι μόνο η πρώτη μεταβλητή μας υπησέρχεται στην εξίσωση κάτι που σημαίνει ότι δεν χρειάζεται να λάβουμε υπόψη τις υπόλοιπες στην πρόβλεψή μας κατ'επέκταση δεν χρειάζεται να ελένξουμε τη στατιστική σημαντικότητα ξεχωριστά κάθε μιας απ'τις μεταβλητές αφού μόνο η μια χρειάζεται .

Τα αποτελέσματα της πρόβλεψης βάση της παραπάνω εξίσωσης φαίνονται στην συνέχεια .

### Επίλογος

Παρατηρώντας κανείς τις παραπάνω μεθόδους πρόβλεψης σίγουρα καταλύγει στο συμπέρασμα ότι όλες περικλείουν κινδύνους , για το λόγο αυτό πρέπει να είμαστε ιδιαίτερα προσεκτικοί . Ειδικά μάλιστα όταν απ'τις προβλέψεις μας υπάρχουν κοινωνικά ωφέλη τότε πρέπει να είμαστε ακόμα πιο προσεκτικοί . Θα πρέπει να γίνεται σωστή τοποθέτηση του προβλήματος και των στόχων . Στη συνέχεια πρέπει να γίνεται αντικειμενική συλλογή και μελέτη των στοιχείων που θα μας χρειαστούν στη πρόβλεψη . Και τέλος στη διαρκή εξέταση των όποιων αποτελεσμάτων για την μείωση των λαθών .

## **Βιβλιογραφία**

- 1 . Τεχνολογική πρόβλεψη -Μαθηματικά Μοντέλα  
Εφαρμογές  
Χρήστος Σκιαδάς - Χανιά 1989
2. Practical Business Forecasting  
John A Saunders - John A Sharp - Stephen F Witt
3. Εισαγωγή στη στατιστική



Σ. Κούνια - Φ. Κολυβά Μαχαίρα - Κ. Μπαγιάτη  
Ε. Μπόρα Σέντα

4. Στατιστική Επετηρίδα των Τηλεπικοινωνιών στην Ελλάδα (1993) .