Σχολή Διοίκησης & Οικονομίας

Τμήμα Επιχειρηματικού Σχεδιασμού & Πληροφοριακών Συστημάτων

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

«ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΚΑΙ ΚΑΤΑΝΟΜΕΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ»

Επιβλέπων καθηγητής:

Μπουμπούλη Αθανασία

Σπουδαστές:

Γιακουμάκης Εμμανουήλ

Κωστάκης Στυλιανός

ΠΑΤΡΑ 2010

Σχολή Διοίκησης & Οικονομίας

Τμήμα Επιχειρηματικού Σχεδιασμού & Πληροφοριακών Συστημάτων

ΑΠΡΙΛΙΟΣ 2010

Πρόλογος

Σε ένα πληθυσμό που έχει κ στοιχεία και θέλουμε να εξετάσουμε κάποια χαρακτηριστικά του τότε είτε εξετάζουμε όλα τα στοιχεία του πληθυσμού είτε εξετάζουμε ένα δείγμα από τον πληθυσμό και αυτό λέγεται δειγματοληψία.

Δείγμα είναι ένα μέρος του στατιστικού πληθυσμού που εξετάζουμε με σκοπό τη συλλογή κάποιων παρατηρήσεων. Για να πάρουμε ένα δείγμα μπορούμε:

1. Να παίρνουμε ένα-ένα στοιχείο από τον πληθυσμό και να το εξετάζουμε χωρίς όμως να το ξανατοποθετούμε στον ίδιο τον πληθυσμό.
2. Να παίρνουμε ένα-ένα στοιχείο από τον πληθυσμό να το εξετάζουμε και να το ξανατοποθετούμε στον ίδιο τον πληθυσμό.
3. Να παίρνουμε κ στοιχεία από τον πληθυσμό μας και να τα εξετάζουμε.

Στη στατιστική έχει μεγάλη σημασία η δειγματοληψία και οι πληροφορίες που παίρνουμε από το δείγμα. Το δείγμα μπορεί είτε να είναι μικρό, είτε να αποτελείται από ένα μεγάλο αριθμό στατιστικών στοιχείων. Υπάρχει βέβαια και η ακραία περίπτωση όπου το δείγμα είναι όλος ο πληθυσμός και στην περίπτωση αυτή δείγμα και πληθυσμός συμπίπτουν.

Προκειμένου να γενικεύσουμε τα συμπεράσματα της έρευνάς μας από το δείγμα στον πληθυσμό, (από όπου αυτό προέρχεται), είναι απαραίτητο το δείγμα να είναι αντιπροσωπευτικό. Για να είναι ένα δείγμα αντιπροσωπευτικό σημαίνει ότι δίνεται η ίδια ευκαιρία σε κάθε μονάδα του πληθυσμού να είναι μονάδα του δείγματος. Ο απλούστερος τρόπος για να το επιτύχει κανείς αυτό είναι να σχηματίσει ένα απλό τυχαίο δείγμα. Η επιλογή των μελών του δείγματος αυτού γίνεται κυρίως με τη χρήση των τυχαίων αριθμών που τους παίρνουμε από τους πίνακες των τυχαίων αριθμών.

Άλλος τρόπος σχηματισμού ενός στατιστικού δείγματος είναι η ενστρωμάτωση όπου γίνεται κατανομή του πληθυσμού σε ομάδες ιδίων χαρακτηριστικών των στρωμάτων (strata). Στη δειγματοληψία θα πρέπει να έχουμε υπόψη μας και κάποιες δυσκολίες που προέρχονται: από τη δυσκολία που έχουμε κάποιες φορές στο να βρούμε τα άτομα που έχουμε επιλέξει, από τις ελλιπείς απαντήσεις και από την δημιουργία ενός δείγματος που εμάς μας εξυπηρετεί -" βολεύει" στην έρευνά μας.

ΠΕΡΙΕΧΟΜΕΝΑ

Πρόλογος σελ. 3-4

ΚΕΦΑΛΑΙΟ 1^{O} σελ. 8-10

ΓΕΝΙΚΑ ΣΤΟΙΧΕΙΑ ΣΤΑΤΙΣΤΙΚΗΣ

1.1Γενικά στοιχεία στατιστικής σελ. 9

ΚΕΦΑΛΑΙΟ 2^{O} σελ. 11-22

ΘΕΩΡΙΑ ΠΙΘΑΝΟΤΗΤΩΝ

Εισαγωγή σελ. 12

2.1 Στοιχεία από τη θεωρεία των πιθανοτήτων. Βασικές έννοιες σελ. 14

2.2 Πράξεις με ενδεχόμενα σελ. 15

2.3 Έννοια και Ορισμός της Πιθανότητας σελ. 15

2.4 Αξιωματική Θεμελίωση των Πιθανοτήτων σελ. 16

2.5 Αρχές Απαρίθμησης σελ. 16

2.6 Δεσμευμένη Πιθανότητα σελ. 20

2.7 Θεώρημα του Bayes σελ. 21

ΚΕΦΑΛΑΙΟ 3^{O} σελ. 23-63

ΔΕΙΓΜΑΤΟΛΗΨΙΑ

3.1 Εισαγωγή σελ. 24

3.2 Βασικές Αρχές Δειγματοληψίας σελ. 24

3.3 Απογραφή και Δειγματοληψία σελ. 25

3.4 Θεωρητική Βάση Δειγματοληψίας σελ. 26

3.5 Δείγματα σελ. 29

3.6 Στατιστικά Σφάλματα σελ. 31

3.7 Σχέδια Δειγματοληψίας σελ. 32

3.7.1 Απλή Τυχαία Δειγματοληψία σελ. 33

3.7.2 Στρωματοποιημένη Δειγματοληψία σελ. 34

3.7.3 Δειγματοληψία κατά ομάδες σελ. 35

3.7.4 Συστηματική Δειγματοληψία σελ. 36

3.7.5 Διπλή Πολλαπλή και Προοδευτική Δειγματοληψία σελ.37

3.7.6 Δειγματοληψία λογική σελ. 37

3.7.7 Δειγματοληψία ποσοστών σελ. 38

3.7.8 Δειγματοληψία με άνεση σελ. 39

3.8 Πληθυσμός και Δείγμα Στατιστική Συμπερασματολογία σελ. 39

3.9 Δειγματοληψία με ή χωρίς Επανατοποθέτηση σελ. 45

3.10 Σχεδιασμός δειγματοληψίας σελ. 46

ΚΕΦΑΛΑΙΟ 4^Ο^ σελ.64-112

ΚΑΤΑΝΟΜΕΣ

4.1 Εισαγωγή σελ. 65

4.2 Θεωρητικές Κατανομές σελ. 66

4.3 Διακριτές Κατανομές σελ. 68

4.3.1 Δυωνυμική Κατανομή σελ. 69

4.3.2 Κατανομή Poisson σελ. 82

4.4 Συνεχείς Κατανομές σελ. 84

4.5 Κατανομές Δειγματοληψίας σελ. 96

4.6 Κατανομή Δειγματοληψίας του Μέσου σελ. 110

4.6.1 Παρατηρήσεις στην κατανομή δειγματοληψίας του μέσου

ΚΕΦΑΛΑΙΟ 5^{O} σελ. 113- 121

ΓΕΝΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ

5.1 Γενικά συμπεράσματα στη δειγματοληψία σελ. 114

5.2 Γενικά συμπεράσματα στις κατανομές δειγματοληψίας σελ. 117

Πηγές-Βιβλιογραφία σελ.120

ΚΕΦΑΛΑΙΟ 1^{Ο}

ΓΕΝΙΚΑ ΣΤΟΙΧΕΙΑ ΣΤΑΤΙΣΤΙΚΗΣ

1.1 ΓΕΝΙΚΑ ΣΤΟΙΧΕΙΑ ΣΤΑΤΙΣΤΙΚΗΣ

Η ανάγκη για καταγραφή και επεξεργασία δεδομένων οδήγησε στην δημιουργία αρχικά στατιστικών υπηρεσιών κατάλληλων για την αποθήκευση και αρχειοθέτηση δεδομένων καθώς και την αποδοτικότερη επεξεργασία και συλλογή αποτελεσμάτων. Η διαδικασία της επεξεργασίας δεδομένων και συλλογής αποτελεσμάτων περιλαμβάνει ένα προστάδιο, το στάδιο της ανάλυσης των δεδομένων.

Με τον όρο αυτό εννοούνται όχι μόνο οι τεχνικές και οι μέθοδοι επεξεργασίας πληροφοριών που προέκυψαν από πραγματικά ή εικονικά πειράματα ή παρακολούθηση φαινομένων αλλά και η θεσμοθέτηση κοινά παραδεκτών τεχνικών με σκοπό την εκτίμηση των χαρακτηριστικών του πληθυσμού.

Παράλληλα η χρήση των υπαρχόντων στατιστικών – μαθηματικών εργαλείων ανάλυσης δεδομένων οδήγησε στην αποδοχή κοινών μεθοδολογιών για την αποδοτικότερη επεξεργασία δεδομένων. Ο διαχωρισμός των υπάρχων μεθοδολογιών περιλαμβάνει:

- Περιγραφική στατιστική,
- Στατιστική συμπερασματολογία – Επαγωγική στατιστική,
- Ανάλυση παλινδρόμησης και διακύμανσης,
- Στοχαστική ανάλυση,

- Μπεϋζιανή ανάλυση,

- Πολυμεταβλητή ανάλυση – Ανάλυση κατηγορικών δεδομένων,

- Μη- παραμετρική στατιστική.

Είναι παραδεκτό ότι όλες οι ερευνητικές τεχνικές που παράγουν δεδομένα επιδέχονται στατιστική επεξεργασία. Μερικές από τα κύρια εργαλεία της επεξεργασίας είναι: (1) Περιγραφική Στατιστική και (2) Στατιστική Συμπερασματολογία (Επαγωγική).

Σκοπός της **Περιγραφικής Στατιστικής** είναι γενικά η άθροιση και σύνοψη δεδομένων. Ειδικότερα αποτελεί ένα στατιστικό εργαλείο με σκοπό την συγκέντρωση ταξινόμηση και παρουσίαση πρωτογενών δεδομένων σε κατανοητή μορφή. Γίνεται με την χρήση **πινάκων** (συχνοτήτων, διπλής εισόδου), γραφημάτων (ραβδογράμματα, θηκογράμματα, διασποράς), και στατιστικών μέτρων (μέτρα κεντρικής τάσης, μέτρα κύμανσης, και μεταβλητότητας).

Σκοπός της **Στατιστικής Συμπερασματολογίας** είναι η διεξαγωγή από τα δεδομένα νόμων, κανόνων και συμπερασμάτων των οποίων η ισχύς ξεπερνά το επίπεδο των παρατηρήσεων. Οι προτεινόμενοι κανόνες καθορίζουν ένα **μαθηματικό μοντέλο** με σκοπό την καλύτερη και απλούστερη ερμηνεία των δεδομένων.

Σχολή Διοίκησης & Οικονομίας

Τμήμα Επιχειρηματικού Σχεδιασμού & Πληροφοριακών Συστημάτων

ΚΕΦΑΛΑΙΟ 2^{O}

ΘΕΩΡΙΑ ΠΙΘΑΝΟΤΗΤΩΝ

Εισαγωγή

Τα φυσικά και κοινωνικά φαινόμενα διακρίνονται σε προκαθορισμένα και σε φαινόμενα τύχης. Η επανάληψη (αν μπορεί να γίνει) ενός φαινομένου τύχης συνιστά ένα πείραμα τύχης (random experiment). Κάθε πείραμα τύχης έχει τα εξής χαρακτηριστικά: είναι γνωστές οι συνθήκες κάτω από τις οποίες διεξάγεται. Μπορεί να επαναληφθεί όσες φορές θέλουμε. Το αποτέλεσμα που παίρνουμε κάθε φορά δεν μπορεί να προβλεφθεί γιατί εξαρτάται είτε από παράγοντες που δεν μπορούμε να ελέγξουμε είτε από παράγοντες που δεν γνωρίζουμε. Η εκτέλεση ενός πειράματος τύχης καλείται δοκιμή. Το αποτέλεσμα μιας δοκιμής καλείται απλό ή στοιχειώδες ενδεχόμενο ή δειγματοσημείο (simple event). Το σύνολο όλων των διαφορετικών μεταξύ τους δειγματοσημείων αποτελεί το δειγματικό χώρο (sample space) του πειράματος τύχης Μπορούμε να τα διαχωρίσουμε τα πειράματα τύχης σε απλά και σε σύνθετα πειράματα. Τα πειράματα τύχης μπορούμε ακόμα να τα διαχωρίσουμε σε ανεξάρτητα και σε εξαρτημένα πειράματα. Ανεξάρτητα είναι δύο ή περισσότερα πειράματα τύχης όταν το αποτέλεσμα του ενός δεν επηρεάζει ούτε επηρεάζεται από τα αποτελέσματα των άλλων, ενώ στην αντίθετη περίπτωση θεωρούνται εξαρτημένα. Ένα υποσύνολο του δειγματικού χώρου ενός πειράματος τύχης είναι ένα ενδεχόμενο (ή γεγονός ή συμβάν) (event) αυτού. Ένα ενδεχόμενο συμβολίζεται συνήθως μ’ ένα κεφαλαίο γράμμα και περιγράφεται είτε λεκτικά με μια φράση είτε αναλυτικά με καταγραφή των στοιχειωδών ενδεχομένων από τα οποία αποτελείται.

Αν τώρα θεωρήσουμε ότι τα Α και Β είναι ενδεχόμενα του ίδιου πειράματος τύχης με δειγματικό χώρο Ω, τότε ορίζουμε τις εξής σχέσεις μεταξύ αυτών:

1. $A \cup B$ είναι το γεγονός «ή Α ή Β ή και τα δυο»
2. $A \cap B$ είναι το γεγονός «και Α και Β»
3. Α’ είναι το γεγονός «όχι Α»
4. Α-Β είναι το γεγονός «Α αλλά όχι και Β»

2.1 ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΣΤΟΙΧΕΙΑ ΑΠΟ ΤΗ ΘΕΩΡΙΑ ΤΩΝ ΠΙΘΑΝΟΤΗΤΩΝ

Κάθε ανθρώπινη ενέργεια οσοδήποτε προγραμματισμένη και αν είναι περικλείει πάντοτε ένα στοιχείο αβεβαιότητας έτσι ώστε το αποτέλεσμα της να εξαρτάται άλλοτε περισσότερο και άλλοτε λιγότερο από την τύχη δηλαδή από τους παράγοντες που δεν μπορούν να προβλεφτούν / ελεγχθούν εκ των προτέρων.

Σε περιπτώσεις που δεν μπορούμε να κάνουμε ακριβή πρόβλεψη του αποτελέσματος μιας ενέργειας έχουμε τη δυνατότητα να μετρήσουμε το βαθμό αβεβαιότητας ενός ορισμένου αποτελέσματός της αφού σε μια ενέργεια που έχει αβέβαιη έκβαση τελικά θα εμφανιστεί ένα μόνο από τα δυνατά ενδεχόμενα της.

Τυχαίο φαινόμενο είναι ένα φαινόμενο η κατάληξη του οποίου δεν είναι γνωστή εκ των προτέρων.

Τυχαίο πείραμα είναι μια διαδικασία της οποίας το αποτέλεσμα δεν είναι γνωστό εκ των προτέρων (π.χ ρίξιμο ζαριού, στρίψιμο νομίσματος, γέννηση παιδιού κ.λ.π.).

Απλά (ή Στοιχειώδη) Ενδεχόμενα ονομάζονται τα δυνατά αποτελέσματα ενός πειράματος. (Το i ενδεχόμενο συμβολίζεται, για παράδειγμα, με Ai).

Δειγματικός χώρος είναι το σύνολο όλων των δυνατών αποτελεσμάτων ενός τυχαίου πειράματος. Ο δειγματικός χώρος συμβολίζεται συνήθως με το γράμμα S.

Ενδεχόμενο καλείται κάθε υποσύνολο του δειγματικού χώρου.

Ο Δειγματικός χώρος S ονομάζεται και **βέβαιο ενδεχόμενο.**

2.2 ΠΡΑΞΕΙΣ ΜΕ ΕΝΔΕΧΟΜΕΝΑ

Ένωση δύο ενδεχομένων καλείται το ενδεχόμενο το οποίο συμβαίνει όταν συμβαίνει τουλάχιστον ένα από τα ενδεχόμενα.

Τομή δύο ενδεχομένων A_1, A_2 καλείται το ενδεχόμενο το οποίο συμβαίνει όταν συμβαίνουν και τα δύο ενδεχόμενα, $A_1 \cap A_2$

Αν η τομή των δύο ενδεχομένων είναι το αδύνατο ενδεχόμενο τότε αυτά καλούνται **ασυμβίβαστα ή ξένα ενδεχόμενα.**

Συμπλήρωμα του ενδεχομένου Α καλείται το ενδεχόμενο Α΄ το οποίο περιέχει τα αποτελέσματα του δειγματικού χώρου S τα οποία δεν περιέχονται στο Α.

2.3 ΈΝΝΟΙΑ ΚΑΙ ΟΡΙΣΜΟΣ ΤΗΣ ΠΙΘΑΝΟΤΗΤΑΣ

Στη μελέτη τυχαίων φαινομένων και πειραμάτων το πρόβλημα βρίσκεται στον προσδιορισμό του δειγματικού χώρου, στην επιλογή των ενδεχομένων που μας ενδιαφέρουν και στον υπολογισμό των πιθανοτήτων που έχουν αυτά να συμβούν.

Πριν ορίσουμε την πιθανότητα θα πρέπει να σημειώσουμε ότι οι διάφοροι ερευνητές έχουν προσεγγίσει την έννοια της πιθανότητας με διαφορετικούς τρόπους και κατά συνέπεια υπάρχουν διαφορετικοί ορισμοί.

Ο κλασικός ορισμός της πιθανότητας οφείλεται στον Laplace ο οποίος, με την προϋπόθεση ότι όλα τα απλά ενδεχόμενα είναι ισοπίθανα, όρισε την πιθανότητα ενός ενδεχομένου ως το λόγο του πλήθους των ευνοϊκών για το ενδεχόμενο περιπτώσεων δια του συνολικού αριθμού των περιπτώσεων (ως περίπτωση θεωρείται το απλό ενδεχόμενο):

2.4 ΑΞΙΩΜΑΤΙΚΗ ΘΕΜΕΛΙΩΣΗ ΤΩΝ ΠΙΘΑΝΟΤΗΤΩΝ

Η αξιωματική θεμελίωση των Πιθανοτήτων προτάθηκε το 1933 από το μαθηματικό A.N. Kolmogorov και έγινε κοινά αποδεκτή. Σύμφωνα με τη θεωρία του Kolmogorov η πιθανότητα είναι μια συνάρτηση που ορίζεται ως εξής : Έστω S ένας δειγματικός χώρος και έστω **B** το σύνολο όλων των ενδεχομένων του S. Ορίζουμε ως **συνάρτηση πιθανότητας** μια συνάρτηση P:

P: **B** → R

2.5 ΑΡΧΕΣ ΑΠΑΡΙΘΜΗΣΗΣ

Σύμφωνα με τον κλασικό ορισμό η πιθανότητα εμφάνισης ενός ενδεχομένου A εκφράζεται ως ο λόγος του αριθμού των ευνοϊκών για το ενδεχόμενο αποτελεσμάτων προς τον αριθμό του συνόλου των δυνατών αποτελεσμάτων του:

Σε απλά προβλήματα τόσο ο αριθμός των ευνοϊκών όσο και ο αριθμός των δυνατών αποτελεσμάτων μπορούν εύκολα να μετρηθούν. Σε πιο σύνθετα όμως προβλήματα η απαρίθμηση των ευνοϊκών και των δυνατών αποτελεσμάτων είναι εξαιρετικά δύσκολη αν όχι αδύνατη. Για το λόγο αυτό χρησιμοποιούμε τις Αρχές Απαρίθμησης.

2.5.1 ΠΟΛΛΑΠΛΑΣΙΑΣΤΙΚΗ ΑΡΧΗ

Έστω ότι η σύνθετη ενέργεια Ε συνίστανται στην εκτέλεση των επί μέρους ενεργειών Ε1, Ε2,...,Εk . Έστω επιπλέον ότι καθεμιά από τις ενέργειες αυτές έχει n_i $(1 \leq n \leq k)$ τρόπους για να πραγματοποιηθεί. Τότε η ενέργεια Ε έχει $n_1 n_{2***} n_k$ δυνατά αποτελέσματα.

2.5.2 ΜΕΤΑΘΕΣΕΙΣ

Οι μεταθέσεις n στοιχείων εκφράζουν τους διαφορετικούς τρόπους με τους οποίους μπορεί κανείς να τοποθετήσει τα στοιχεία αυτά. Το πλήθος των διαφορετικών αυτών τρόπων δίνεται από τον τύπο

! nPn=

Υπενθυμίζεται ότι το n! (n παραγοντικό) ορίζεται ως εξής:

n! = 1*2*3*…*n

Επιπλέον σημειώνεται ότι εξ ορισμού 0! =1

Παράδειγμα

Να υπολογιστούν οι διαφορετικοί τρόποι με τους οποίους τρεις φοιτητές μπορούν να καθίσουν σε τρεις καρέκλες.

Λύση

3! = 6

Αν μεταξύ των n στοιχείων που θέλουμε να τοποθετήσουμε υπάρχουν και ορισμένα όμοια μεταξύ τους τότε το σύνολο των n στοιχείων μπορεί να χωρισθεί σε k υποσύνολα έτσι ώστε το πρώτο να περιέχει n1 όμοια στοιχεία, το δεύτερο n2 όμοια, το τρίτο n3 όμοια κ.ο.κ. με n1+n2+n3+…+nk = n. Ο αριθμός των μεταθέσεων θα δίνεται από τον τύπο:

$$_{n}P_{n1,n2,n3,\ldots,nk} = \frac{n!}{n1!*n2!*n3!***nk!}$$

Παράδειγμα

Να υπολογιστούν οι διαφορετικοί τρόποι με τους οποίους μπορούν να τοποθετηθούν 6 όμοιοι κύβοι από τους οποίους 2 είναι κόκκινοι, 2 πράσινοι, 1 κίτρινος και 1 λευκός

Λύση

$$_{6}P_{1,1,2,2} = \frac{6!}{1!*1!*2!*2!} = 180$$

2.5.3 ΔΙΑΤΑΞΕΙΣ

Οι Διατάξεις των n στοιχείων ενός συνόλου ανά x εκφράζουν τους διαφορετικούς τρόπους με τους οποίους μπορεί να επιλέξει κάποιος x από τα n nx≤ στοιχεία ενός συνόλου όταν τους ενδιαφέρει η σειρά επιλογής.

Παράδειγμα

Να υπολογιστούν οι διαφορετικοί διψήφιοι αριθμοί που μπορούν να σχηματιστούν από τα ψηφία 1,2,3,4,5 (επανάληψη του ίδιου ψηφίου στον αριθμό δεν επιτρέπεται)

Λύση

$$_5P_2 = \frac{5!}{3!} = 4*5 = 20$$

2.5.4 ΕΠΑΝΑΛΗΠΤΙΚΕΣ ΔΙΑΤΑΞΕΙΣ

Οι επαναληπτικές διατάξεις των n στοιχείων ανά k εκφράζουν τους διαφορετικούς τρόπους με τους οποίους μπορούν να τοποθετηθούν σε σειρά k στοιχεία που λαμβάνουμε από τα n αν κάθε στοιχείο μπορεί να επαναληφθεί μέχρι k φορές (εδώ το k μπορεί να είναι μικρότερο, ίσο ή μεγαλύτερο από το n)

Οι διαφορετικοί τρόποι δύνονται από τον τύπο :

k^n

Παράδειγμα

Να υπολογιστεί το πλήθος των διαφορετικών τριψήφιων αριθμών που μπορούν να σχηματισθούν από τα ψηφία 1,2,3,4 όταν κάθε αριθμός μπορεί να έχει μερικά ή όλα τα ψηφία του ίδια.

Λύση

$4^3 = 64$

2.5.5 ΣΥΝΔΥΑΣΜΟΙ

Οι **συνδυασμοί** των n στοιχείων ενός συνόλου ανά x εκφράζουν τους διαφορετικούς τρόπους με τους οποίους μπορεί να επιλέξει κάποιος x από τα n στοιχεία ενός συνόλου όταν η σειρά επιλογής δεν τον ενδιαφέρει.

Κατά συνέπεια στην περίπτωση των Συνδυασμών δύο επιλογές είναι διαφορετικές μόνο όταν περιέχουν διαφορετικά στοιχεία.

Παράδειγμα

Να υπολογιστούν οι τρόποι με τους οποίους μπορεί να συγκροτηθεί μια τριμελής επιτροπή από μία ομάδα επτά ατόμων

Λύση

$$\binom{7}{3} = {}_7C_3 = \frac{7!}{3!*4!} = 35$$

2.6 ΔΕΣΜΕΥΜΕΝΗ ΠΙΘΑΝΟΤΗΤΑ

Σε πολλές περιπτώσεις όταν ζητάμε να υπολογίσουμε την πιθανότητα ενός ενδεχομένου έχουμε ήδη κάποιες πληροφορίες γι' αυτό. Οι πληροφορίες αυτές στην ουσία περιορίζουν τον αρχικό δειγματικό χώρο. Για παράδειγμα, η πιθανότητα ότι ένα χαρτί τυχαία επιλεγμένο από μια τράπουλα θα είναι σπαθί γίνεται μεγαλύτερη αν γνωρίζουμε ότι ήδη από την τράπουλα αυτή έχουν αφαιρεθεί τα μπαστούνια.

Η πιθανότητα να συμβεί ένα ενδεχόμενο δοθέντος ότι έχει συμβεί ένα ενδεχόμενο ονομάζεται **δεσμευμένη πιθανότητα του δοθέντος του.**

2.7 ΘΕΩΡΗΜΑ ΤΟΥ BAYES

Στην ενότητα αυτή θα παρουσιάσουμε το θεώρημα του Bayes, που είναι η βάση μιας ολόκληρης στατιστικής φιλοσοφίας γνωστής ως Μπεϋζιανής Στατιστικής (Bayesian Statistics). Για να το κάνουμε όμως αυτό πρέπει πρώτα να ορίσουμε τη διαμέριση ενός συνόλου.

Παράδειγμα

Σε ένα εργοστάσιο οι μηχανές Α, Β, και Γ παράγουν το 25%, το 35%,και το 40% της συνολικής παραγωγής. Μετά από έλεγχο βρέθηκε ότι το 5% των παραγομένων τεμαχίων από το μηχάνημα Α είναι ελαττωματικό, επίσης το 4% και το 2% των παραγομένων τεμαχίων από τις μηχανές Β και Γ είναι ελαττωματικό. Η συνολική παραγωγή φτάνει τα 5000 τεμάχια. Ένα τεμάχιο εξάγεται τυχαία και βρίσκεται ελαττωματικό.
Ποια η πιθανότητα κατασκευής του από τη μηχανή Β;

Λύση

Έστω:
Ε το ενδεχόμενο ότι το τεμάχιο που έχει εξαχθεί είναι ελαττωματικό.
Α το ενδεχόμενο ότι το τεμάχιο αυτό έχει εξαχθεί από τη μηχανή Α.
Β το ενδεχόμενο ότι το τεμάχιο αυτό έχει εξαχθεί από τη μηχανή Β.
Γ το ενδεχόμενο ότι το τεμάχιο αυτό έχει εξαχθεί από τη μηχανή Γ.
Θα έχουμε:

P(A)= 25/100, P(B)= 35/100, P(Γ)= 40/100

Επίσης έχουμε:

P(E/A)= 0.05, P(E/B)= 0.04, P(E/Γ)= 0,02

Οπότε:

$$P(B/E)= \frac{P(B)\ P(E/B)}{P(A)\ P(E/A) + P(B)\ P(E/B) + P(\Gamma)\ P(E/\Gamma)}$$

άρα:

$$P(B/E)= \frac{0,35 * 0,04}{0,25 * 0,05 + 0,35 * 0,04 + 0,40 * 0,02}$$

οπότε καταλήγουμε ότι:

P(B/E)= 0,4058

ΚΕΦΑΛΑΙΟ 3^{Ο}

ΔΕΙΓΜΑΤΟΛΗΨΙΑ

3.1 ΕΙΣΑΓΩΓΗ

Στο παρακάτω κεφάλαιο θα ασχοληθούμε με τις βασικές αρχές της δειγματοληψίας και τα εναλλακτικά σχέδια δειγματοληψίας. Στόχος της δειγματοληψίας και ζητούμενο της ενασχόλησής μας είναι η δυνατότητα της εξασφάλισης ικανοποιητικών αποτελεσμάτων και η εξαγωγή συμπερασμάτων για έναν ολόκληρο πληθυσμό παίρνοντας ένα πολύ μικρό δείγμα . Επίσης θέτουμε παρακάτω την έννοια του μέτρου, της ακρίβειας σε σχέση με το βαθμό εμπιστοσύνης που μπορούμε να αποδώσουμε στα αποτελέσματα μιας δειγματοληψίας.

3.2 ΒΑΣΙΚΕΣ ΑΡΧΕΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

Η δειγματοληψία είναι η μέθοδος μελέτης ενός πληθυσμού με βάση όπως προαναφέραμε ένα δείγμα από τον πληθυσμό αυτό. Η μελέτη αυτού του δείγματος συνοψίζεται στα παρακάτω ερωτήματα :

- Ποια τα αίτια για δειγματοληψία
- Ποια η θεωρητική βάση της δειγματοληψίας
- Πως πρέπει να σχεδιάζεται το δείγμα
- Ποιες οι μέθοδοι συλλογής των δεδομένων
- Πως μπορούν οι ζητούμενες πληροφορίες να προκύπτουν από τα δεδομένα του δείγματος

- Πως μπορούν οι δειγματικές πληροφορίες να χρησιμοποιηθούν για να εξαχθούν συμπεράσματα για το σύνολο του πληθυσμού
- Πως μπορεί να μετρηθεί και να εκτιμηθεί η αξιοπιστία των επαγωγικών γενικεύσεων για τον πληθυσμό από το δείγμα

3.3 ΑΠΟΓΡΑΦΗ ΚΑΙ ΔΕΙΓΜΑΤΟΛΗΨΙΑ

Τα στατιστικά σφάλματα χωρίζονται σε δύο είδη, τα μη δειγματοληπτικά και τα δειγματοληπτικά.

Το μη δειγματοληπτικό σφάλμα είναι συνήθως μεγάλο αλλά αυτό μπορεί να αποφευχθεί σημαντικά με κατάλληλα σχεδιασμένη δειγματοληψία. Επιπλέον το μη δειγματοληπτικό σφάλμα δεν μπορεί να εκτιμηθεί ενώ η μέτρηση του δειγματοληπτικού σφάλματος είναι δυνατή.

Γι αυτούς τους λόγους το συνολικό σφάλμα όχι μόνο αναμένεται να είναι μικρότερο σε μια δειγματοληπτική έρευνα αλλά παράλληλα τα δειγματικά αποτελέσματα μπορούν να χρησιμοποιηθούν με μεγαλύτερο βαθμό εμπιστοσύνης εφόσον γνωρίζουμε το πιθανό μέγεθος του σφάλματος.

Συνάμα, οι πληροφορίες που δίνουν οι δειγματοληπτικές απογραφές μπορούν να χρησιμοποιηθούν για τη βελτίωση των εργασιών της γενικής απογραφής.

Μια συγκριτική αξιολόγηση των δύο μεθόδων συλλογής πληροφοριών, απογραφής και δειγματοληψίας είναι ικανή να προσδώσει σοβαρά μειονεκτήματα στη μέθοδο της απογραφής και πολλά πλεονεκτήματα στη μέθοδο της δειγματοληψίας. Μερικά πλεονεκτήματα τα οποία καταγράφονται είναι η χρήση λιγότερων απογραφέων και η αρτιότερη εκπαίδευσή τους καθώς ο χρόνος και το κόστος της έρευνας .

Τα μη δειγματοληπτικά σφάλματα μπορεί να είναι εξίσου σπουδαία όσο και τα σφάλματα δειγματοληψίας και εφόσον εξετάζουμε την πιστότητα των δειγματοληπτικών εξαγόμενων, η διαφορά μεταξύ ενός δειγματικού αποτελέσματος και του αποτελέσματος από μια πλήρη μέτρηση κάτω από τις ίδιες προϋποθέσεις μετριέται ανάλογα με το αν αναφερόμαστε στην ακρίβεια ή την αξιοπιστία του δειγματικού αποτελέσματος.

3.4 ΘΕΩΡΗΤΙΚΗ ΒΑΣΗ ΤΗΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

Η δειγματοληψία είναι ένα ισχυρό εργαλείο με το οποίο μπορούμε να προβλέψουμε και να γενικεύσουμε επαγωγικά την συμπεριφορά μαζικών φαινομένων. Η θεωρία της δειγματοληψίας που καθιστά δυνατή αυτή την επαγωγική διαδικασία έχει να κάνει με την επιρροή που ασκούν οι πολλαπλές δυνάμεις ενός πληθυσμού στις στοιχειώδεις μονάδες του. Αυτή η τεράστια σύνθεση δυνάμεων επιδρά πάνω στις μονάδες με σημαντικό βαθμό ανεξαρτησίας. Αυτό επεξηγεί τις μεταβολές από μονάδα σε μονάδα του πληθυσμού. Τέτοια φαινόμενα μπορεί να συναντήσουμε στο οικογενειακό μας περιβάλλον, στο φυσικό περιβάλλον, στο οικονομικό και κοινωνικό σύνολο.

Αν και η ποικιλία είναι μια καθολική ιδιότητα των μαζικών δεδομένων δεν υπάρχει στατιστικός πληθυσμός του οποίου οι μονάδες θα ποικίλουν μεταξύ τους ως προς το ερευνούμενο χαρακτηριστικό χωρίς όριο. Οι διαπιστώσεις ότι ο οποιοσδήποτε πληθυσμός έχει χαρακτηριστικές ιδιότητες και ότι οι αποκλίσεις μεταξύ των μονάδων του είναι περιορισμένες κάνουν το έργο του ενδιαφερόμενου δυνατό να επιλέγει ένα μικρό δείγμα τυχαίο και αμερόληπτο το οποίο αποτελεί το πλαίσιο των χαρακτηριστικών ιδιοτήτων του πληθυσμού.

Άλλη μια σπουδαία ιδιότητα των μαζικών δεδομένων είναι η κανονικότητα και η ομοιομορφία που παρουσιάζουν. Έτσι λοιπόν η μεταβλητότητα ενός πληθυσμού είναι εξισορροπημένη γύρω από κάποια κεντρική τιμή η οποία μπορεί να αναφερθεί ως τιμή συγκέντρωσης του πληθυσμού. Συνεπώς, οι επιμέρους μονάδες του πληθυσμού τείνουν να διαφέρουν η μία της άλλης αλλά και ταυτόχρονα να προσαρμόζονται προς ορισμένα μαζικά πρότυπα.

Εξαιτίας της στατιστικής κανονικότητας ένα μεγάλο τυχαίο δείγμα εφόσον είναι επιλέξιμο δεν θα διαφέρει σημαντικά από τον πληθυσμό ενώ λόγω της απόκλισης εάν λάβουμε έναν αριθμό τυχαίων δειγμάτων δεν θα έχουμε αυστηρή κανονικότητα. Εάν άλλο ένα μεγαλύτερο τυχαίο δείγμα λαμβάνεται από τον ίδιο πληθυσμό, τα χαρακτηριστικά μέτρα κάθε μεγαλύτερου δείγματος θα τείνουν να διαφέρουν όλο και λιγότερο από τα αντίστοιχα χαρακτηριστικά μέτρα του πληθυσμού. Δηλαδή, όσο μεγαλύτερο είναι το δείγμα, τόσο περισσότερο είναι αντιπροσωπευτικό του πληθυσμού. Παράλληλα οι μέσοι μεγάλων τυχαίων δειγμάτων, να τείνουν να συγκεντρώνονται γύρω από το μέσο του πληθυσμού ενώ θα έχουν μια μικρή διακύμανση στο εύρος της μεταβολής τους.

Η σταθερότητα των μεγάλων αριθμών, συνεπώς καθιστά δυνατόν τον καθορισμό του μεγέθους του δείγματος έτσι ώστε τα χαρακτηριστικά του να μπορούν να περιγράφουν το όλο του πληθυσμού. Στην πράξη κατά την επαγωγική διαδικασία συνήθως λαμβάνουμε μόνο ένα δείγμα είτε αυτό είναι μεγάλο είτε αυτό είναι μικρό είμαστε σχεδόν βέβαιοι ότι τα χαρακτηριστικά του δεν είναι ακριβώς αυτά του πληθυσμού, η τυχαία διαδικασία όμως της δειγματοληψίας είναι αυτή που θα μετρήσει αντικειμενικά τα σφάλματα .

Από τα παραπάνω, προκύπτουν τα εξής συμπεράσματα:

- Ένα τυχαίο δείγμα μπορεί να αναμένεται να αντιπροσωπεύει τον πληθυσμό αρκετά καλά διότι κάθε πληθυσμός παρουσιάζει σχετικά σταθερά χαρακτηριστικά γνωρίσματα.
- Τυχαία δείγματα που επιλέγονται από τον ίδιο πληθυσμό, μπορεί να έχουν πολλά κοινά γνωρίσματα αλλά δεν αναμένεται να έχουν ποτέ σταθερή ομοιομορφία.
- Ένα μεγάλο δείγμα θα διαφέρει λιγότερο από ότι ένα μικρό δείγμα από το γεννήτορα πληθυσμό και όταν αυξάνει το μέγεθος των δειγμάτων η μεταβλητότητα μεταξύ αυτών θα μειώνεται ένεκα της σταθερότητας που παρουσιάζει στους μεγάλους αριθμούς.
- Η δειγματοληπτική μεταβλητικότητα μπορεί να μετρηθεί μόνο αν η δειγματοληψία είναι τυχαία, διότι εκεί ισχύουν οι νόμοι των πιθανοτήτων.

3.5 ΔΕΙΓΜΑΤΑ

Ένα δείγμα μπορούμε να το ορίσουμε με περισσότερη ακρίβεια στο πλαίσιο της δειγματοληψίάς και όχι μόνο σαν ένα μέρος ενός πληθυσμού. Οι μονάδες που περιέχονται σε ένα πληθυσμό αποκαλούνται μονάδες δειγματοληψίας και μπορούν να ταξινομηθούν σε στοιχειώδεις μονάδες δειγματοληψίας και πρωταρχικές μονάδες δειγματοληψίας.

Οι στοιχειώδεις μονάδες δειγματοληψίας είναι όλες οι μονάδες που περιέχονται στον πληθυσμό και οι χαρακτηριστικές τους ιδιότητες μετρώνται ή απαριθμούνται.

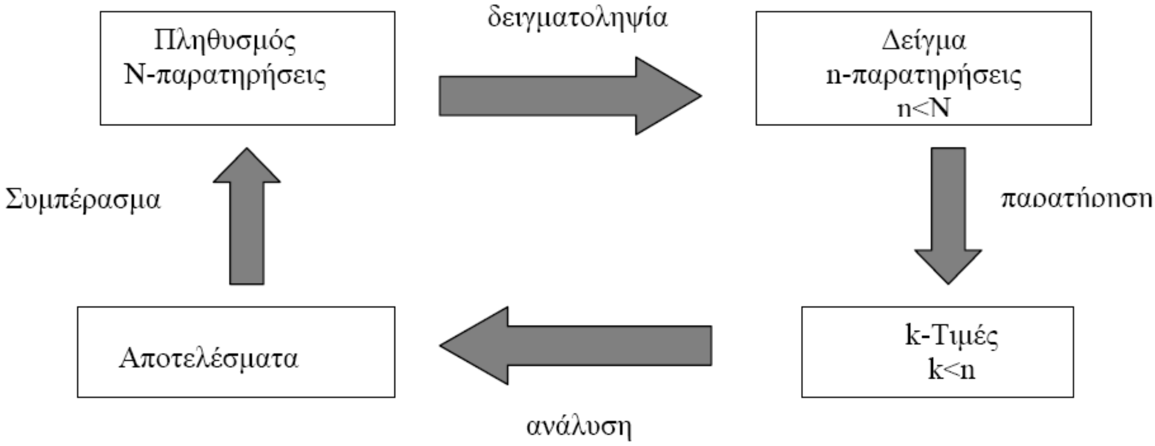
Πρωταρχικές μονάδες δειγματοληψίας μπορεί να είναι οι στοιχειώδεις μονάδες δειγματοληψίας καθώς και ομάδες από στοιχειώδεις μονάδες δειγματοληψίας.

- Το δείγμα είναι μια συλλογή πρωταρχικών μονάδων δειγματοληψίας που επιλέγονται έτσι ώστε να αποτελούν ένα αντιπροσωπευτικό δείγμα από το οποίο μπορούν να εξαχθούν συμπεράσματα για τον πληθυσμό. Ο όρος δείγμα μπορεί να περιέχει τρεις έννοιες:
- Ένα δείγμα μπορεί να περιέχει Πρωταρχικές μονάδες δειγματοληψίας όπως για παράδειγμα οι επιχειρήσεις του βιομηχανικού κλάδου που χρησιμοποιούνται για τη μελέτη των εργαζόμενων στο βιομηχανικό κλάδο.

- Εάν ένα δείγμα δεν είναι απλώς μια φέτα του πληθυσμού αλλά ένα υποσύνολο αναμένεται να είναι αντιπροσωπευτικό του πληθυσμού
- Ο τελικός σκοπός της δειγματοληψίας δεν είναι απλώς η εξασφάλιση δειγματικών συναρτήσεων αλλά η εξαγωγή συμπερασμάτων που αφορούν τον πληθυσμό

Στη δειγματοληψία είναι σημαντικό να διακρίνουμε τα τυχαία δείγματα από τα λογικά δείγματα. Ένα τυχαίο δείγμα επιλέγεται δια της μεθόδου της πιθανότητας σύμφωνα με την οποία δεν μπορούν να αποφασιστούν ποιες μονάδες του πληθυσμού θα συμπεριληφθούν στο δείγμα. Η επιλογή επιτυγχάνεται με τη λειτουργία της τύχης και μόνο. Αντίθετα ένα λογικό δείγμα λαμβάνεται σύμφωνα με την προσωπική κρίση του ερευνητή.

Οι μονάδες οι οποίες θα συμπεριληφθούν στο δείγμα κρίσεως είναι αποτέλεσμα εμπειρικής κρίσης του ερευνητή ως προς την αντιπροσωπευτικότητά τους. Άρα εδώ δεν μιλάμε για πιθανότητα αλλά για προσωπική κρίση του ενδιαφερόμενου.



3.6 ΣΤΑΤΙΣΤΙΚΑ ΣΦΑΛΜΑΤΑ

Με τον όρο σφάλμα εννοούμε τη διαφορά μεταξύ της τιμής της δειγματικής συνάρτησης και της τιμής της αντίστοιχης παραμέτρου του πληθυσμού και όπως προαναφέραμε τα σφάλματα χωρίζονται σε μη δειγματοληπτικά και δειγματοληπτικά. Τα μη δειγματοληπτικά σφάλματα περιέχουν μεροληψίες και λάθη . Παράγοντες που προκαλούν μεροληψίες είναι οι εξής: ο ελλιπής προσδιορισμός του πληθυσμού, η ατέλεια του πλαισίου έρευνας, η αοριστία του ερωτηματολογίου, ο μη σαφής καθορισμός των επιθυμητών πληροφοριών κλπ.

Σφάλμα δειγματοληψίας προκύπτει από την τυχαία επιλογή των μονάδων δειγματοληψίας. Αυτός ο τύπος σφάλματος συμβαίνει διότι μόνο ένα μέρος του πληθυσμού περιέχεται στο δείγμα. Με τον όρο αυτό εννοούμε τη διαφορά μεταξύ του δειγματικού αποτελέσματος και του αποτελέσματος της απογραφής. Το συνολικό σφάλμα μιας στατιστικής έρευνας προκύπτει από το άθροισμα των μη δειγματοληπτικών και δειγματοληπτικών σφαλμάτων.

Πίνακας Τυχαία δείγματα και στατιστικές

Α/Α	Τιμές δείγματος	Αριθμητικός μέσος δείγματος	Διακύμανση δείγματος
1ο	$\{x_{11}, x_{12}, \cdots, x_{1n}\}$	$\overline{x}_1 = \frac{\sum x_{1i}}{n}$	$s_1^2 = \frac{\sum (x_{1i} - \overline{x}_1)^2}{n-1}$
2ο	$\{x_{21}, x_{22}, \cdots, x_{2n}\}$	$\overline{x}_2$	s_2^2
…	…….	…..	…..
kο	$\{x_{k1}, x_{k2}, \cdots, x_{kn}\}$	$\overline{x}_k$	s_k^2
…	…….	…..	…..
$m^o = \binom{N}{n}^o$	$\{x_{m1}, x_{m2}, \cdots, x_{mn}\}$	$\overline{x}_m$	s_m^2
	$\{X_1, X_2, \ldots, X_n\}$	$\overline{X} = \frac{\sum X_i}{n}$	$S^2 = \frac{\sum (X_i - \overline{X})^2}{n-1}$

3.7 ΣΧΕΔΙΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

Τα σχέδια δειγματοληψίας γενικά κατανέμονται σε δύο κατηγορίες, τυχαίας δειγματοληψίας και μη τυχαίας δειγματοληψίας.

Η τυχαία δειγματοληψία που αναφέρεται και ως δειγματοληψία πιθανότητας περιλαμβάνει την απλή τυχαία δειγματοληψία, την περιορισμένη τυχαία δειγματοληψία και την διπλή πολλαπλή και προοδευτική δειγματοληψία.

Η μη τυχαία δειγματοληψία περιλαμβάνει την λογική δειγματοληψία, την quota δειγματοληψία και την δειγματοληψία με άνεση.

3.7.1 ΑΠΛΗ ΤΥΧΑΙΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑ

Απλή τυχαία δειγματοληψία είναι η διαδικασία επιλογής ενός δείγματος κατά την οποία κάθε μονάδα του πληθυσμού έχει ίση και ανεξάρτητη πιθανότητα να συμπεριληφθεί στο δείγμα. Η τυχαία επιλογή που εξασφαλίζει τυχαιότητα στη δειγματοληπτική διαδικασία πρέπει να προέρχεται από μια διαδικασία προσεκτικά ελεγχόμενη. Η απλή τυχαία δειγματοληψία κρίνεται αποτελεσματική όταν ο πληθυσμός δεν είναι μεγάλος. Τα πλεονεκτήματα της είναι η δυνατότητα υπολογισμού του δειγματοληπτικού σφάλματος, η αντικειμενικότητα της δειγματοληψίας και η αποφυγή της μεροληψίας, η ευκολία της εφαρμογής της και η κατανόηση της δεδομένου ότι έχει απλό εννοιολογικό σχεδιασμό

Ένα μειονέκτημα της εφαρμογής αυτής στους μεγάλους πληθυσμούς είναι ότι οι μονάδες του πληθυσμού πρέπει να έχουν απαριθμηθεί. Παράλληλα η δειγματοληψία καθίσταται δαπανηρή επειδή η μέθοδος απαιτεί πλήρως ενημερωμένα πλαίσια. Συνάμα σε περιπτώσεις πεπερασμένου συνόλου η μέθοδος κρίνεται χρονοβόρα και έχει μεγάλο κόστος στην συλλογή δεδομένων.

3.7.2 ΣΤΡΩΜΑΤΟΠΟΙΗΜΕΝΗ ΔΕΙΓΜΑΤΟΛΗΨΙΑ

Η στρωματοποιημένη δειγματοληψία είναι μια διαδικασία κατά την οποία με την χρήση διαθέσιμων πληροφοριών που αφορούν στα δεδομένα, προσπαθούμε να σχεδιάσουμε ένα πιο αποτελεσματικό δείγμα από εκείνο που θα παίρναμε από την απλή τυχαία διαδικασία. Η διαδικασία της στρωματοποιήσεως απαιτεί όπως ο πληθυσμός διαιρεθεί σε ομοιογενείς ομάδες ή τάξεις που ονομάζονται στρώματα. Στη συνέχεια ένα δείγμα μπορεί να ληφθεί από κάθε στρώμα δια των απλών και τυχαίων μεθόδων και το δείγμα που προκύπτει τελικά είναι το στρωματοποιημένο δείγμα

Η παραπάνω διαδικασία είναι ικανοποιητική αν δεν υπάρχει μεγάλη διαφορά στη διασπορά από στρώμα σε στρώμα.

Ο σημαντικότερος στόχος της Στατιστικής, και ιδιαίτερα της Στατιστικής συμπερασματολογίας, είναι η εξαγωγή συμπερασμάτων για το σύνολο ενός πληθυσμού, αντλώντας πληροφορίες από ένα μικρό υποσύνολο αυτού.

Τα πλεονεκτήματα της δειγματοληψίας αυτής είναι η ομοιογένεια των δεδομένων ανάλογα το στρώμα με συνέπεια τη μικρότερη διασπορά και μεγαλύτερη ακρίβεια.

Το μειωμένο κόστος είναι ένα χαρακτηριστικό της δειγματοληψίας αυτής, ενώ με τη δειγματοληψία αυτή εξασφαλίζεται η αντιπροσώπευση συγκεκριμένων μονάδων πληθυσμού που διαφορετικά θα μπορούσαν να αποκλειστούν λόγω το ότι αντιπροσωπεύουν μια μικρή αναλογία του συνολικού πληθυσμού. Επιπρόσθετα δίνεται η δυνατότητα αντιμετώπισης των διαφορών που υπάρχουν μεταξύ των πληθυσμιακών ομάδων.

Στα μειονεκτήματα συμπεριλαμβάνεται η σύνθετη επιλογή δείγματος, η ύπαρξη προκαταρκτικής πληροφόρησης, και η κατάταξη και ο ορισμός του συνόλου του πληθυσμού με χαρακτηριστικά της στρωματοποίησης.

3.7.3 ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΚΑΤΑ ΟΜΑΔΕΣ

Η παραπάνω δειγματοληψία αναφέρεται στην διαδικασία διαίρεσης του πληθυσμού σε ομάδες και τον σχηματισμό ενός δείγματος από ομάδες που να αντιπροσωπεύουν τον πληθυσμό. Όταν αυτές επιλεγούν μπορούμε είτε να συμπεριλάβουμε στο δείγμα όλες τις στοιχειώδεις μονάδες των ομάδων είτε να πάρουμε ένα δείγμα με λιγότερες πρωταρχικές μονάδες.

Για παράδειγμα να ξεκινήσουμε περνώντας τα τμήματα ΑΤΕΙ στην Ελλάδα, μετά τα τμήματα των σχολών της Διοίκησης, Οικονομίας και τέλος τους σπουδαστές. Στόχος πάντως είναι η διερεύνηση των στοιχειωδών μονάδων.

Τα πλεονεκτήματα της είναι η συντομία του χρόνου διεξαγωγής, η αποφυγή της πολυπλοκότητας στη συλλογή του δείγματος. Παράλληλα κρίνεται ως η μόνη εφικτή μέθοδος όταν δεν υπάρχει δειγματοληπτικό πλαίσιο. Σαν μειονέκτημα μπορεί να αναφερθεί η αβεβαιότητα, που υπάρχει για το αν τα δείγματα της ομάδας είναι αντιπροσωπευτικά του συνολικού πληθυσμού.

3.7.4 ΣΥΣΤΗΜΑΤΙΚΗ ΔΕΙΓΜΑΤΟΛΗΨΙΑ

Η παραπάνω έχει να κάνει με την επιλογή ενός τυχαίου δείγματος. Είναι η διαδικασία επιλογής της κ μονάδας από τον πληθυσμό με μια τυχαία εκκίνηση. Όπου κ είναι ο λόγος δειγματοληψίας. Βασικό πλεονέκτημα είναι η απλότητα στη σχεδίαση και η αποτελεσματικότητα στην συγκεκριμένη δειγματοληψία.

Τα πλεονεκτήματα της είναι η αποφυγή μεγάλου αριθμού σφαλμάτων συλλογής, ο περιορισμός της μεροληψίας, η συλλογή πληροφοριών για δεδομένο κατά μονάδα κόστος και ομαδοποιείται ο πληθυσμός αφού βάσει του καταλόγου επιλέγεται ένα στοιχείο από κάθε ομάδα για το δείγμα.

Τα μειονεκτήματα είναι η υποχρεωτική κατασκευή ή ύπαρξη καταλόγου, με το σύνολο Ν του πληθυσμού και ο κίνδυνος της περιοδικότητας στις τιμές των μονάδων του πληθυσμού όσον αφορά τη σειρά εμφάνισης τους στη λίστα.

3.7.5 ΔΙΠΛΗ, ΠΟΛΛΑΠΛΗ ΚΑΙ ΠΡΟΟΔΕΥΤΙΚΗ ΔΕΙΓΜΑΤΟΛΗΨΙΑ

Όλα τα σχέδια δειγματοληψίας εκτός από την πολυσταδιακή συσσωρευτική δειγματοληψία μπορούν να αναφερθούν με τον όρο απλή δειγματοληψία επειδή χρησιμοποιούν ένα απλό δείγμα.

Αλλά η καθημερινή πρακτική και ιδιαίτερα η παραγωγική διαδικασία έχει οδηγήσει στην ευρεία εφαρμογή των διπλών και πολλαπλών και προοδευτικών σχεδίων δειγματοληψίας.

Η δυνατότητα αυτή έχει οδηγήσει στην διπλή δειγματοληψία μια διαδικασία κατά την οποία ένα μικρό δείγμα ελέγχεται καταρχήν και η απόφαση για αποδοχή η απόρριψη μπορεί να επιλεγεί από ένα δεύτερο δείγμα. Έτσι αν παίρνουμε μία μονάδα κάθε φορά μπορούμε να έχουμε προοδευτική δειγματοληψία.

3.7.6 ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΛΟΓΙΚΗ

Είναι η μέθοδος συλλογής των μονάδων ενός δείγματος με βάση την κρίση του δειγματολήπτη εφόσον αυτός θεωρεί ότι θα έχει τα ανάλογα αποτελέσματα. Έτσι οι μονάδες του δείγματος που θεωρούνται αντιπροσωπευτικές επιλέγονται σε ένα λογικό δείγμα.

Χρησιμοποιείται για να κρατήσουμε το μέγεθος του δείγματος μικρό καθώς και στην καθημερινότητα για τη λήψη αποφάσεων κυβερνητικής πολιτικής εφόσον οι εκτιμήσεις μπορούν να είναι διαθέσιμες σε σύντομο χρονικό διάστημα.

3.7.7 ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΠΟΣΟΣΤΩΝ

Η παραπάνω δειγματοληψία έχει να κάνει πληθυσμό με συγκεκριμένα χαρακτηριστικά τότε το δείγμα μπορεί να περιέχει συγκεκριμένη αναλογία των χαρακτηριστικών του πληθυσμού. Εξαρτάται (το δείγμα) από την κρίση του ερευνητή.

Πλεονεκτήματα της δειγματοληψίας αυτής είναι το χαμηλό κόστος διεξαγωγής και ο περιορισμένος χρόνος, η δυνατότητα διεξαγωγής σε περιπτώσεις που η τυχαία λήψη των μονάδων του δείγματος δεν μπορεί να γίνει επειδή το αναγκαίο δειγματοληπτικό πλαίσιο είναι ανύπαρκτο. Τα διοικητικά προβλήματα είναι σαφώς λιγότερα λόγω της έλλειψης των προβλημάτων που προκύπτουν από την άρνηση συνέντευξης ενώ δίνεται η δυνατότητα αξιολόγησης των πληροφοριών που περιέχονται στο δείγμα με στατιστικές μεθόδους καθώς και τον υπολογισμό του σφάλματος το οποίο είναι συνυφασμένο με την εκτιμήτρια.

Η έλλειψη τυχαιότητας κατά την επιλογή των μονάδων του δείγματος δεν επιτρέπει να υπολογισθούν τα τυπικά σφάλματα των εκτιμήσεων της δειγματοληψίας ποσοστών, ενώ η μεροληψία και η έλλειψη ελέγχου του συνεντευκτή κατά το χρόνο του έργου του μπορεί να λογισθεί σαν μειονέκτημα.

3.7.8 ΔΕΙΓΜΑΤΟΛΗΨΑ ΜΕ ΑΝΕΣΗ

Συνίσταται στο ότι το δείγμα είναι ένα κλάσμα του πληθυσμού και λαμβάνεται για έρευνα εξαιτίας της άνετης διαθέσεως του, άρα επιλέγεται με ευκολία. Τα αποτελέσματα είναι συνήθως μεροληπτικά και μη ικανοποιητικά. Η παραπάνω διαδικασία μπορεί να χρησιμοποιηθεί σε έρευνες πιλότος.

Ένα από τα προβλήματα τα οποία έχει να αντιμετωπίσει ο στατιστικός κατά τη διενέργεια μιας δειγματοληψίας είναι ο καθορισμός του άριστου μεγέθους δείγματος. Οι παράγοντες που θα ληφθούν υπόψιν για τον καθορισμό του άριστου μεγέθους του δείγματος, είναι το μέγεθος του δειγματοληπτικού σφάλματος. Όσο μεγαλύτερο είναι το δείγμα τόσο μικρότερο είναι το δειγματοληπτικό σφάλμα. Συνάμα όσο μικρότερο είναι το δείγμα τόσο λιγότερο δαπανηρή είναι η έρευνα και τόσο γρηγορότερα τελειώνει. Αν ο πληθυσμός είναι ομοιογενής απατούνται λίγες μονάδες δείγματος. Σε αντίθετη περίπτωση ανομοιογένειας απατούνται πολλές μονάδες δείγματος. Για να είναι άριστο ένα δείγμα δεν αρκεί να αποτελείται από μεγάλο αριθμό μονάδων αλλά θα πρέπει να εφαρμοσθεί και η κατάλληλη μέθοδος δειγματοληπτικής έρευνας.

3.8 ΠΛΗΘΥΣΜΟΣ ΚΑΙ ΔΕΙΓΜΑ. ΣΤΑΤΙΣΤΙΚΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ

Στα πλαίσια της στατιστικής συμπερασματολογίας, η έννοια “πληθυσμός” είναι συνυφασμένη με το σύνολο όλων των υπό εξέταση μονάδων (ατόμων), το δε υπό εξέταση “χαρακτηριστικό” αναφέρεται σε κάποια ποσοτική (και σπανιότερα ποιοτική) μέτρηση που αφορά όλα τα άτομα του πληθυσμού. Για παράδειγμα, σε μία στατιστική μελέτη σχετικά με το μηνιαίο εισόδημα των εργαζομένων στην Ευρωπαϊκή Ένωση, ο όρος “πληθυσμός” αναφέρεται στο σύνολο των εργαζομένων των χωρών της Ευρωπαϊκής Ένωσης, ενώ το υπό μελέτη “χαρακτηριστικό” ενός συγκεκριμένου ατόμου είναι το μηνιαίο εισόδημα αυτού. Παρόμοια, αν ένα εργοστάσιο κατασκευής λαμπτήρων πραγματοποιήσει στατιστική μελέτη σχετικά με τον χρόνο ζωής των λαμπτήρων, τότε “πληθυσμός” είναι το σύνολο των λαμπτήρων που παρασκευάζει το εργοστάσιο, ενώ “χαρακτηριστικό” ενός “ατόμου” (εδώ άτομο = λαμπτήρας) του πληθυσμού είναι ο χρόνος λειτουργίας (ζωής) του συγκεκριμένου λαμπτήρα.

Για να τεθούν σε ενιαία βάση όλες οι παραπάνω περιπτώσεις, είναι απαραίτητο να χρησιμοποιηθεί μία μορφή αντιστοιχίας μεταξύ των όρων πληθυσμός και συνάρτηση κατανομής, καθώς επίσης και χαρακτηριστικό και τυχαία μεταβλητή. Συγκεκριμένα, θεωρούμε ότι η πιθανοθεωρητική συμπεριφορά του πληθυσμού περιγράφεται από κάποια συνάρτηση κατανομής F, και ότι η αντίστοιχη ποσοτική μέτρηση του χαρακτηριστικού περιγράφεται από την αντίστοιχη τυχαία μεταβλητή X, με $X \sim F$. Επομένως, πλήρης γνώση της συνάρτησης κατανομής F θα σήμαινε και πλήρη γνώση της συμπεριφοράς του πληθυσμού. Για παράδειγμα, αν η τ.μ. X παριστάνει το χρόνο ζωής ενός λαμπτήρα και F είναι η αντίστοιχη συνάρτηση κατανομής, τότε οι τιμές

$$F(x) = P(X \leq x)$$

παριστάνουν το ποσοστό των λαμπτήρων με χρόνο ζωής το πολύ x. Συνεπώς, αν ο κατασκευαστής (εργοστάσιο) γνώριζε την F(x) , θα μπορούσε να περιγράψει πλήρως την πιθανοθεωρητική συμπεριφορά του χρόνου ζωής X ενός λαμπτήρα, και άρα την σύνθεση του πληθυσμού. Αυτό όμως δεν συμβαίνει στην πράξη διότι η F είναι άγνωστη, και το μόνο που μπορούμε να κάνουμε είναι να παρατηρήσουμε ορισμένες τιμές (πραγματοποιήσεις, μετρήσεις) της τυχαίας μεταβλητής X, δηλαδή, στο παράδειγμά μας, να θέσουμε σε λειτουργία ν λαμπτήρες και να παρατηρήσουμε το χρόνο ζωής τους, έστω X1, X2 ,..., Xν .

Φυσικά, όλες οι X1,..., Xν προέρχονται από την ίδια συνάρτηση κατανομής F (αφού θεωρήσαμε ότι η F παριστάνει την συνάρτηση κατανομής του “πληθυσμού”), και είναι στοχαστικά ανεξάρτητες τ.μ., επειδή θεωρήσαμε ότι παριστάνουν το χρόνο ζωής (λειτουργίας) ν διαφορετικών λαμπτήρων. Η προηγούμενη ανάλυση οδηγεί φυσιολογικά στον εξής ορισμό.

Ορισμός (Τυχαίο δείγμα). Αν ένας πληθυσμός έχει αντίστοιχη συνάρτηση κατανομής F, τότε τυχαίο δείγμα καλείται ένα σύνολο ανεξαρτήτων και ισόνομων τυχαίων μεταβλητών X1, X2 ,..., Xν με κοινή συνάρτηση κατανομής F. Ο αριθμός ν{1,2,...} καλείται μέγεθος δείγματος. (Συμβολίζουμε X1, X2 ,..., Xν ~ F) .

Παρατήρηση Στην πράξη, μετά τη λήψη του δείγματος (δειγματοληψία), οι τυχαίες μεταβλητές Χ1,..., Χν λαμβάνουν κάποιες πραγματικές τιμές, έστω Χ1 = x1,..., Χν = xν . Οι τιμές x1,..., xν καλούνται επίσης τυχαίο δείγμα, διότι παριστάνουν τις παρατηρηθείσες τιμές των Χ1,..., Χν μετά τη δειγματοληψία, και άρα είναι διαθέσιμες στον ερευνητή που θα πραγματοποιήσει την στατιστική έρευνα. Εντούτοις, στη θεωρία δεν γίνεται διάκριση μεταξύ Χ1,..., Χν και x1,..., xν , επειδή η στατιστική ανάλυση (πρέπει να) προηγείται της δειγματοληψίας. Απλώς, τα στατιστικά συμπεράσματα εφαρμόζονται στη συνέχεια στις παρατηρηθείσες τιμές (μετρήσεις) x1,..., xν .

Παρατήρηση Η συνάρτηση κατανομής F που αντιστοιχεί στον πληθυσμό θεωρείται **άγνωστη**, και η πληροφορία για την συμπεριφορά της πρέπει να αντλείται μόνο από το τυχαίο δείγμα. Αυτό έρχεται σε αντίθεση με την πρακτική που εφαρμόζεται σε προβλήματα πιθανοτήτων, όπου η συνάρτηση κατανομής F θεωρείται γνωστή, και ουσιαστικά διαχωρίζει την επιστήμη των Πιθανοτήτων από αυτήν της Στατιστικής. Θα λέγαμε ότι οι Πιθανότητες είναι “επαγωγική” επιστήμη (με βάση αξιώματα και θεωρήματα συνάγεται το “όλον” από το “μέρος”), ενώ η Στατιστική είναι “επαγωγική” (από παρατήρηση του “μέρους” συμπεραίνονται ιδιότητες που αφορούν το “όλον”).

Παρατήρηση Η Στατιστική διαχωρίζεται σε δύο κύριους κλάδους, τη μη παραμετρική και την παραμετρική. Στη μη παραμετρική στατιστική, η συνάρτηση κατανομής F υποτίθεται εντελώς άγνωστη, ή τουλάχιστον ανήκει σε μία πολύ μεγάλη κλάση κατανομών (π.χ. η F υποτίθεται συνεχής συνάρτηση κατανομής), ενώ στην παραμετρική στατιστική η άγνωστη συνάρτηση κατανομής F περιορίζεται σε μία παραμετρική οικογένεια κατανομών, έτσι ώστε μόνο κάποια παράμετρος αυτής να θεωρείται άγνωστη (στα επόμενα δεν θα ασχοληθούμε με προβλήματα της μη παραμετρικής Στατιστικής). Για παράδειγμα, αν η F είναι η συνάρτηση κατανομής του χρόνου ζωής των λαμπτήρων, τότε μπορεί να υποτεθεί εκ των προτέρων ότι η F είναι εκθετική με παράμετρο θ > 0 (θ άγνωστη). Σε αυτήν την περίπτωση

$$F(x) = F(x;\theta) = 1 - e^{-x\theta}, \quad x > 0,$$

και η προσπάθειά μας εστιάζεται στην εκτίμηση της άγνωστης παραμέτρου θ, με βάση την πληροφορία που αντλείται από ένα τυχαίο δείγμα Χ1,..., Χν . Παρόμοια, αν η τ.μ. Χ παριστάνει το εισόδημα ενός εργαζόμενου στην Ευρωπαϊκή Ένωση, τότε μπορεί να υποτεθεί ότι Χ ~ Ν (μ,σ^2), όπου τουλάχιστον μία από τις παραμέτρους μ, σ^2 θεωρείται άγνωστη (φυσικά, το πιο ρεαλιστικό θα ήταν να θεωρήσουμε και τις δύο παραμέτρους μ και σ^2 άγνωστες).

Στην πράξη θέλουμε συχνά να βγάλουμε συμπεράσματα για μια μεγάλη ομάδα ατόμων ή αντικειμένων. Αντί να μελετήσουμε ολόκληρη την ομάδα, τον πληθυσμό όπως λέμε, που είναι δύσκολο ή αδύνατο, μπορούμε να εξετάσουμε ένα μικρό μέρος αυτού του πληθυσμού, δηλ.

ένα δείγμα. Αυτό γίνεται με σκοπό να συμπεράνουμε κάτι για τον πληθυσμό από τα αποτελέσματα εξετάσεως του δείγματος. Οι αρχές και οι μέθοδοι που χρησιμοποιούνται για το σκοπό αυτό αποτελούν τη Στατιστική Συμπερασματολογία.

Παράδειγμα Για ένα πλήθος 12.000 ενήλικων φοιτητών (αυτός είναι ο πληθυσμός) ζητάμε να μελετήσουμε το βάρος η το ύψος τους εξετάζοντας μόνον 100 φοιτητές (το δείγμα).

Παράδειγμα Ζητάμε να βγάλουμε συμπεράσματα για το ποσοστό των ελαττωματικών βιδών από την εξαήμερη παραγωγή ενός εργοστασίου εξετάζοντας 20 τυχαίες βίδες από την παραγωγή κάθε ημέρας. Στην περίπτωση αυτή το σύνολο των βιδών της εξαήμερης παραγωγής αποτελεί τον πληθυσμό, ενώ οι 120 βίδες το δείγμα.

Παράδειγμα . Ζητάμε να εξετάσουμε εάν ένα νόμισμα είναι κανονικό η όχι ρίχνοντας το πολλές φορές. Ο πληθυσμός περιλαμβάνει όλες τις δυνατές ρίψεις. Ένα δείγμα μπορεί να ληφθεί, εάν ξεχωρίσουμε μερικές ρίψεις, π.χ. τις πρώτες 60, και σημειώσουμε πόσες φορές ήρθε «κεφάλι» και πόσες «γράμματα».

Παράδειγμα Ζητάμε να βγάλουμε συμπεράσματα για τα χρώματα 200 σφαιρών (πληθυσμός) βγάζοντας 20 σφαίρες (δείγμα) από το κουτί που περιέχει τις σφαίρες με επανατοποθέτηση η χωρίς επανατοποθέτηση.

Μερικά σημεία πρέπει να τονιστούν ιδιαίτερα. Πρώτο, η λέξη πληθυσμός δε σημαίνει γενικά ό,τι στην καθημερινή γλώσσα, όπως π.χ. στη φράση «πληθυσμός της Αθήνας». Δεύτερο, η λέξη πληθυσμός δηλώνει συχνά ένα πλήθος παρατηρήσεων η μετρήσεων αντί για πλήθος ατόμων η αντικειμένων. Έτσι μπορούμε να θεωρήσουμε ότι στο Παράδειγμα 1 ο πληθυσμός περιλαμβάνει 12.000 βάρη ή ύψη και ότι στο Παράδειγμα 4 ο πληθυσμός περιλαμβάνει τα χρώματα των 200 σφαιρών. Τρίτο, ο πληθυσμός μπορεί να περιλαμβάνει πεπερασμένο (συνήθως) η άπειρο πλήθος στοιχείων. Το πλήθος αυτό παριστάνεται με N και καλείται μέγεθος του πληθυσμού. Το πλήθος των στοιχείων του δείγματος καλείται μέγεθος του δείγματος και συμβολίζεται με n. Στο Παράδειγμα 1 είναι N = 12.000, n = 100, ενώ στο Παράδειγμα3 το N είναι άπειρο και n = 60.

3.9 ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΜΕ ΚΑΙ ΧΩΡΙΣ ΕΠΑΝΑΤΟΠΟΘΕΤΗΣΗ

Εάν βγάλουμε ένα αντικείμενο από ένα κουτί, μπορούμε να το ξαναβάλουμε η να μην το ξαναβάλουμε πριν βγάλουμε ένα άλλο. Στην πρώτη περίπτωση το ίδιο αντικείμενο μπορεί να βγει πολλές φορές, ενώ στη δεύτερη μόνο μια φορά. Έτσι έχουμε δειγματοληψία με επανατοποθέτηση, όπου ένα στοιχείο του πληθυσμού μπορεί να εκλεγεί πολλές φορές, και δειγματοληψία χωρίς επανατοποθέτηση, όπου ένα στοιχείο του πληθυσμού μπορεί να εκλεγεί το πολύ μια φορά.

Ένας πεπερασμένος πληθυσμός μπορεί να θεωρηθεί άπειρος σε δειγματοληψία με επανατοποθέτηση, επειδή μπορούμε να πάρουμε δείγματα οποιουδήποτε μεγέθους. Σε πολλές περιπτώσεις στην πράξη, ακόμα και δειγματοληψία χωρίς επανατοποθέτηση από έναν πολύ μεγάλο πληθυσμό μπορεί να θεωρηθεί σαν δειγματοληψία από άπειρο πληθυσμό.

3.10 ΣΧΕΔΙΑΣΜΟΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

Βασικές έννοιες δειγματοληψίας

Δειγματοληπτική μονάδα

Κατά το σχεδιασμό της δειγματοληψίας είναι απαραίτητο, πριν προβούμε στην επιλογή του δείγματος, να ορίσουμε το σύνολο των μονάδων που αποτελούν τον ερευνώμενο πληθυσμό, οι οποίες ονομάζονται δειγματοληπτικές μονάδες (sampling units). Π.χ. αν θέλουμε να μελετήσουμε τις αντιδράσεις των υπαλλήλων μιας επιχείρησης, σε μια απόφαση της διοίκησης για αναπροσαρμογή των μισθών, δειγματοληπτικές μονάδες θα είναι όλοι οι υπάλληλοι της επιχείρησης. Αν θέλουμε να μελετήσουμε το ύψος του εισοδήματος των νοικοκυριών μιας πόλης, δειγματολογικές μονάδες θα είναι όλα τα νοικοκυριά της πόλης κ.ο.κ.

Η δειγματοληπτική μονάδα πρέπει να ορίζεται με σαφήνεια, ώστε να μπορούμε να γενικεύσουμε τα συμπεράσματα που προκύπτουν από το δείγμα στο σύνολο του ερευνώμενου πληθυσμού. Π.χ. όταν οι δειγματοληπτικές μονάδες αναφέρονται σε οικογένειες, νοικοκυριά, καταστήματα, επιχειρήσεις κ.λπ. πρέπει να δίνεται στους ερευνητές, ο σχετικός ορισμός των μονάδων αυτών, γιατί κάθε μια από αυτές τις μονάδες έχει και διαφορετική έννοια. Για παράδειγμα το «νοικοκυριό» μπορεί να αποτελείται από δύο ή περισσότερα άτομα, συγγενικά ή μη που διαμένουν στην ίδια κατοικία, προμηθεύονται συνήθως από κοινού τα απαραίτητα για τη συντήρησή τους και τρώνε κατά κανόνα μαζί (πολυπρόσωπο νοικοκυριό). Επίσης νοικοκυριό αποτελεί κάθε άτομο το οποίο διαμένει μόνο του σε μια κατοικία ή διαμένει μαζί με άλλα άτομα στην κατοικία, αλλά δεν προμηθεύεται από κοινού με αυτά τα απαραίτητα για τη συντήρησή του, ούτε τρώει μαζί τους (μονοπρόσωπο νοικοκυριό). Αντίθετα ο όρος «οικογένεια» προϋποθέτει την ύπαρξη συγγένειας μεταξύ των ατόμων.

Είναι φανερό ότι αν σε μια έρευνα μερικοί ερευνητές συγκεντρώνουν πληροφορίες σε επίπεδο νοικοκυριού και μερικοί σε επίπεδο οικογένειας και ειδικότερα όταν κάποια από τα μέλη της οικογένειας διαμένουν σε άλλη πόλη (π.χ. φοιτητές, στρατιώτες κ.λπ.) τότε τα στοιχεία που θα συγκεντρωθούν δεν μπορεί να γενικευθούν επαγωγικά στο σύνολο του ερευνώμενου πληθυσμού.

Δειγματοληπτικό πλαίσιο

Το σύνολο των δειγματοληπτικών μονάδων του ερευνώμενου πληθυσμού, το οποίο είναι καταχωρημένο σε έναν κατάλογο ή παρουσιάζεται υπό μορφή χαρτογραφικών διαγραμμάτων (οικοδομικά τετράγωνα μιας πόλης κ.λπ.), αποτελεί το δειγματοληπτικό πλαίσιο (sampling frame).

Το δειγματοληπτικό πλαίσιο αποτελεί βασική προϋπόθεση για την επιτυχία μιας δειγματοληπτικής έρευνας, δεδομένου ότι πρέπει να περιέχει όλο το δειγματοληπτούμενο πληθυσμό, ώστε να ικανοποιεί την υπόθεση, «ότι κάθε δειγματοληπτική μονάδα έχει την ίδια ευκαιρία επιλογής κατά τη δειγματοληψία», δηλαδή δεν υπάρχουν παραλείψεις δειγματοληπτικών μονάδων και διπλές καταχωρήσεις στο δειγματοληπτικό πλαίσιο. Σε περίπτωση που υπάρχουν παραλείψεις ή διπλές καταχωρήσεις στο πλαίσιο, το δείγμα που θα επιλεγεί, δεν θα είναι αντιπροσωπευτικό και κατά συνέπεια τα στατικά συμπεράσματα που θα γενικευθούν επαγωγικά στο σύνολο του ερευνώμενου πληθυσμού δεν θα είναι αξιόπιστα.

Οι κατάλογοι που χρησιμοποιούνται ως δειγματοληπτικά πλαίσια και περιέχουν τους παρακάτω τύπους λαθών, οδηγούν σε αναξιόπιστα αποτελέσματα:

α) Λανθασμένη πληροφόρηση

Όταν ο κατάλογος – πλαίσιο (δειγματοληπτικό πλαίσιο) δεν είναι πλήρως ενημερωμένος. Π.χ. υπάρχουν στον κατάλογο λανθασμένα ονοματεπώνυμα λόγω μετακίνησης ή θανάτου ορισμένων ατόμων, λανθασμένες διευθύνσεις κατοικιών κ.ο.κ.

β) Ελλιπής πληροφόρηση

Όταν ο κατάλογος – πλαίσιο δεν περιέχει όλες τις δειγματοληπτικές μονάδες. Π.χ. οι εκλογικοί κατάλογοι δεν περιλαμβάνουν άτομα όλων των ηλικιών παρά μόνο αυτούς που ψηφίζουν, επομένως δεν πρέπει να χρησιμοποιούνται ως πλαίσια για έρευνες που αφορούν πληθυσμούς ατόμων όλων των ηλικιών. Επίσης, οι τηλεφωνικοί κατάλογοι αποτελούν πλαίσια με ελλιπή πληροφόρηση, διότι δεν περιλαμβάνονται σ’ αυτούς όλα τα άτομα ή νοικοκυριά και ειδικότερα όταν έχουν πολύ χαμηλό εισόδημα.

γ) Περισσότερες πληροφορίες από ό,τι χρειάζεται

Όταν ο κατάλογος – πλαίσιο περιέχει ορισμένες δειγματοληπτικές μονάδες περισσότερο από μια φορά (π.χ. ονόματα ατόμων να παρουσιάζονται δύο φορές σε κάποιο κατάλογο) ή περιέχει δειγματοληπτικές μονάδες που δεν ανήκουν στο δειγματοληπτούμενο πληθυσμό.

Για τον περιορισμό των παραπάνω λαθών είναι απαραίτητο να ενημερώνονται οι κατάλογοι – πλαίσια, πριν χρησιμοποιηθούν για την επιλογή των τελικών μονάδων (άτομα, νοικοκυριά, καταστήματα κ.λπ.) του δείγματος, ώστε να προκύπτουν αξιόπιστα αποτελέσματα.

Οι κατάλογοι – πλαίσια που χρησιμοποιούνται από τη Γενική Γραμματεία ΕΣΥΕ σε διάφορες δειγματοληπτικές έρευνες καταρτίζονται αρχικά από στοιχεία γενικών απογραφών (πληθυσμού – κατοικιών, γεωργίας – κτηνοτροφίας, κ.λπ.). Π.χ. για έρευνες νοικοκυριών όπως η έρευνα οικογενειακών προϋπολογισμών, καταναλωτή, εργατικού δυναμικού κ.λπ. καταρτίζονται κατάλογοι – πλαίσια για κάθε μονάδα επιφανείας που έχει επιλεγεί στο δείγμα. Στον κατάλογο – πλαίσιο καταχωρούνται όλες οι κατοικίες του οικοδομικού τετραγώνου που αποτελεί η μονάδα επιφανείας με συνεχή αύξουσα αρίθμηση. Αν η μονάδα επιφανείας αποτελείται από δύο ή περισσότερα οικοδομικά τετράγωνα καταχωρούνται πρώτα οι κατοικίες του 1^{ου} οικοδομικού τετραγώνου, μετά του 2^{ου} κ.ο.κ. έως ότου καταχωρηθούν με συνεχή αύξουσα αρίθμηση όλες οι κατοικίες των οικοδομικών τετραγώνων που αποτελούν τη μονάδα επιφανείας. Οι κατάλογοι αυτοί ενημερώνονται πριν χρησιμοποιηθούν για την επιλογή των κατοικιών δείγματος (τελικές μονάδες), σε επόμενες έρευνες.

Σκοπός της ενημέρωσης είναι να διαπιστωθούν οι διαφορές που υπάρχουν μεταξύ της υφιστάμενης κατάστασης από απόψεως κατοικιών σε κάθε μονάδα επιφανείας και αυτής που παρουσιάζεται στον κατάλογο – πλαίσιο, ώστε να εμφανίζεται με ακρίβεια η υφιστάμενη κατάσταση. Συγκεκριμένα πρέπει να εξακριβωθούν οι κατοικίες που δεν υπάρχουν στη μονάδα επιφανείας ή δεν είναι καταχωρημένες στον κατάλογο π.χ. κατοικίες που κατά την αρχική κατάρτιση του καταλόγου είχαν παραλειφθεί ή από επαγγελματικές στέγες που χρησιμοποιούνταν, άλλαξαν χρήση και χρησιμοποιούνται τώρα σαν κατοικίες ή κτίστηκαν αργότερα (μετά την κατάρτιση του καταλόγου).

Οι κατάλογοι – πλαίσια που χρησιμοποιούνται για δειγματοληπτικές έρευνες πρωτογενούς τομέα καταρτίζονται με στοιχεία της απογραφής γεωργίας – κτηνοτροφίας και ενημερώνονται σε κάθε έρευνα που διεξάγεται. Η ενημέρωση γίνεται από τον ερευνητή πριν την επιλογή των τελικών μονάδων (γεωργικές ή κτηνοτροφικές εκμεταλλεύσεις) του δείγματος, διότι είναι φυσικό να έχουν επέλθει ορισμένες μεταβολές, όπως π.χ.:

α) Μία ή περισσότερες εκμεταλλεύσεις να μην υπάρχουν κατά το χρόνο διεξαγωγής της έρευνας (πώληση ζώων, γεωργικών εκτάσεων κ.λπ.)

β) Να έχουν δημιουργηθεί νέες εκμεταλλεύσεις

γ) Να υπάρχουν κάτοχοι εκμεταλλεύσεων, που να μην είναι καταχωρημένοι στον κατάλογο (διαφυγές κ.λπ.)

δ) Να υπάρχουν μεταβολές στις γεωργικές εκτάσεις, στον αριθμό ζώων κ.λπ. (π.χ. μια εκμετάλλευση που εμφανίζεται στον κατάλογο με 10 ζώα, να έχει κατά το χρόνο της έρευνας 60 ζώα ή μια άλλη από 300 ζώα να έχει 30 κ.ο.κ.)

ε) Να έχει αλλάξει ο κάτοχος της εκμετάλλευσης

Τυχαία και μη τυχαία δειγματοληψία

Το πόσο τυχαίο είναι το δείγμα του ερευνώμενου πληθυσμού, χαρακτηρίζει τη δειγματοληψία, σε τυχαία και μη τυχαία.

Στην τυχαία δειγματοληψία (random sampling) η επιλογή των δειγματοληπτικών μονάδων γίνεται κατά τρόπο τυχαίο, δηλαδή όλες οι μονάδες του ερευνώμενου πληθυσμού έχουν την ίδια ευκαιρία να συμπεριληφθούν στο δείγμα.

Στη μη τυχαία δειγματοληψία (non random sampling) η επιλογή των δειγματοληπτικών μονάδων γίνεται κατά τρόπο μη τυχαίο, δηλαδή δεν έχουν όλες οι μονάδες του ερευνώμενου πληθυσμού την ίδια ευκαιρία να συμπεριληφθούν στο δείγμα. Με άλλα λόγια η επιλογή του δείγματος γίνεται με υποκειμενικά κριτήρια.

Τα αποτελέσματα που προκύπτουν από τη μη τυχαία δειγματοληψία δεν είναι αντιπροσωπευτικά και επομένως τα στατιστικά συμπεράσματα δε μπορεί να γενικευθούν επαγωγικά στο σύνολο του ερευνώμενου πληθυσμού με αξιοπιστία.

Δειγματοληπτικό σφάλμα

Δειγματοληπτικό σφάλμα (sampling error) είναι η διαφορά μεταξύ της εκτίμησης μιας παραμέτρου που προκύπτει από την τυχαία επιλογή των δειγματοληπτικών μονάδων και της αντίστοιχης υπολογισθείσας πληθυσμιακής παραμέτρου που προκύπτει από την καθολική έρευνα (απογραφή) του πληθυσμού. Φυσικά τα δύο μέτρα (εκτίμηση παραμέτρου και υπολογισθείσα πληθυσμιακή παράμετρος) πρέπει να προκύπτουν από την ίδια διαδικασία μέτρησης.

Όσο αυξάνει το μέγεθος του δείγματος σε μια έρευνα τόσο ελαττώνεται το δειγματοληπτικό σφάλμα. Σε μια καθολική έρευνα το δειγματοληπτικό σφάλμα είναι μηδέν (διότι έχουμε ένα μοναδικό δείγμα).

Μη δειγματοληπτικά σφάλματα

Τα μη δειγματοληπτικά σφάλματα (non sampling errors) δεν οφείλονται στη δειγματοληψία, αλλά είναι σφάλματα τα οποία μπορεί να γίνουν σε κάθε είδους έρευνα, καθολική ή δειγματοληπτική.

Οι σπουδαιότεροι παράγοντες οι οποίοι προκαλούν μη δειγματοληπτικά σφάλματα είναι οι εξής:

α. Ακαταλληλότητα του ερωτηματολογίου

Εάν το ερωτηματολόγιο είναι ακατάλληλο, δηλαδή περιέχει ερωτήματα πολύπλοκα ή ερωτήματα που θίγουν το ερευνώμενο πρόσωπο, οι απαντήσεις μπορεί να είναι εσφαλμένες ή μεροληπτικές.

Επειδή το ερωτηματολόγιο αποτελεί τον πιο βασικό παράγοντα για την επιτυχία μιας έρευνας, θα περιγράψουμε σε επόμενη παράγραφο, τις βασικές αρχές κατάρτισης του ερωτηματολογίου.

β. Σφάλματα ανταπόκρισης

Ο ανταποκρινόμενος μπορεί να δώσει εσφαλμένη απάντηση, είτε επειδή δεν κατάλαβε την ερώτηση του ερευνητή, είτε επειδή δεν θυμάται κάποιο γεγονός, είτε διότι δεν επιθυμεί να δώσει τη σωστή απάντηση. Π.χ. μπορεί να δηλώσει μικρότερη ηλικία από την πραγματική, να δηλώσει υψηλότερο εισόδημα από το πραγματικό για να φανεί στον ερευνητή αξιοσέβαστος κ.ο.κ.

Επίσης οι απαντήσεις του ερευνώμενου προσώπου επηρεάζονται από τα προσωπικά χαρακτηριστικά του ερευνητή, π.χ. φύλο, ηλικία, μόρφωση κ.λπ.

γ. Σφάλματα ερευνητή

Ο ερευνητής μπορεί να καταχωρήσει στο ερωτηματολόγιο εσφαλμένα μια απάντηση που δόθηκε σωστά από τον ανταποκρινόμενο. Π.χ. ο ανταποκρινόμενος λέει ότι είναι 45 ετών και ο ερευνητής γράφει 48.

δ. Σφάλματα επεξεργασίας

Κατά το στάδιο επεξεργασίας (κωδικογράφηση ερωτημάτων και εισαγωγή στοιχείων στα μαγνητικά μέσα) γίνονται λάθη από τους υπαλλήλους που ασχολούνται με την εργασία αυτή.

Τα μη δειγματοληπτικά σφάλματα δεν είναι εύκολο να μετρηθούν και να εντοπισθεί η πιθανή αιτία των σφαλμάτων αυτών. Ένας καλός όμως σχεδιασμός μιας δειγματοληπτικής έρευνας, δηλαδή κατάρτιση ενός κατάλληλου ερωτηματολογίου, επιλογή ειδικευμένων ερευνητών και σωστή εποπτεία στην εργασία αυτών, μπορεί να περιορίσει τα μη δειγματοληπτικά σφάλματα στο ελάχιστο.

Τα δειγματοληπτικά και μη δειγματοληπτικά σφάλματα αποτελούν το συνολικό σφάλμα μιας έρευνας. Επομένως για να αυξήσουμε το βαθμό αξιοπιστίας των αποτελεσμάτων μιας έρευνας, πρέπει κατά το σχεδιασμό να λάβουμε υπόψη μας τους ανωτέρω παράγοντες σφαλμάτων και με το μικρότερο δυνατό κόστος να καταβάλουμε προσπάθεια περιορισμού των σφαλμάτων αυτών.

Μεροληπτική δειγματοληψία

Εάν κατά την επιλογή του δείγματος, από ένα πληθυσμό, σε μερικές μονάδες δίνεται μεγαλύτερη ευκαιρία να συμπεριληφθούν στο δείγμα, τότε έχουμε μεροληπτική δειγματοληψία (biased sampling). Η μεροληπτική επιλογή ενός δείγματος μπορεί να γίνει είτε για λόγους ευκολίας, είτε από άγνοια ενός ερευνητή, ο οποίος δεν έχει την απαραίτητη εμπειρία να σχεδιάσει την επιλογή δείγματος κατά τρόπο ώστε να αποκλεισθεί η μεροληψία. Π.χ. αν γίνει μια δειγματοληπτική έρευνα κοινής γνώμης και ο ερευνητής συγκεντρώσει πληροφορίες από έναν αριθμό ατόμων που διέρχονται τυχαία από την πλατεία Ομόνοιας κατά τις πρωινές ώρες. Η δειγματοληπτική αυτή έρευνα θα είναι μεροληπτική διότι από την πλατεία Ομόνοιας θα περάσουν μόνο άτομα χαμηλής εισοδηματικής τάξης, δηλαδή στην περίπτωση αυτή δεν δίνεται καμία ευκαιρία σε άτομα υψηλής εισοδηματικής τάξης να επιλεγούν στο δείγμα.

Επισημαίνεται ότι ενώ το δειγματοληπτικό σφάλμα ελαττώνεται όταν αυξάνεται το μέγεθος του δείγματος, αντίθετα η μεροληψία δεν ελαττώνεται. Η ύπαρξη μεροληψίας επιδρά σημαντικά στην επιτυχία μιας δειγματοληπτικής έρευνας και είναι δυνατό να οδηγήσει σε εσφαλμένα συμπεράσματα.

Όταν η μεροληψία είναι ανύπαρκτη οι εκτιμήσεις είναι αμερόληπτες, οπότε η αναμενόμενη τιμή είναι ίδια σαν αυτή που θα βρίσκαμε από μια καθολική έρευνα.

Βασικές αρχές κατάρτισης ερωτηματολογίου

Κατά το σχεδιασμό μιας έρευνας (καθολικής ή δειγματοληπτικής) η κατάρτιση του ερωτηματολογίου αποτελεί το πρώτο βήμα στη διαδικασία συλλογής στατιστικών στοιχείων.

Το ερωτηματολόγιο αποτελεί μέσο συλλογής των στατιστικών στοιχείων κατά τη διεξαγωγή μιας έρευνας. Η συμπλήρωση του ερωτηματολογίου γίνεται είτε με προσωπική συνέντευξη (personal interview) είτε με ταχυδρομική αποστολή (mail questionnaire).

Το ερωτηματολόγιο συνιστά μια από τις βασικές προϋποθέσεις για την επιτυχία της έρευνας (καθολικής ή δειγματοληπτικής). Εάν το ερωτηματολόγιο είναι ακατάλληλο, δηλαδή περιέχει ερωτήματα πολύπλοκα ή ερωτήματα που θίγουν το ερευνώμενο πρόσωπο, οι απαντήσεις μπορεί να είναι εσφαλμένες ή μεροληπτικές, γεγονός που οδηγεί στην αύξηση των μη δειγματοληπτικών σφαλμάτων.

Η κατάρτιση του ερωτηματολογίου πρέπει να αρχίζει από την αρχή του σχεδιασμού της έρευνας και να ολοκληρώνεται μετά από τη διεξαγωγή μιας δοκιμαστικής έρευνας.

Το πρώτο βήμα κατά το σχεδιασμό ενός ερωτηματολογίου είναι να αποφασισθεί το είδος και ο αριθμός των ερωτημάτων τα οποία πρέπει να καλύπτουν τα χαρακτηριστικά που παρουσιάζουν ενδιαφέρον. Ειδικότερα κατά το σχεδιασμό του ερωτηματολογίου πρέπει να λαμβάνονται υπόψη τα εξής :

α) Τα μακροσκελή ερωτηματολόγια είναι κουραστικά και για τον ερευνητή και για τον ερευνώμενο. Γι' αυτό πρέπει να τίθενται μόνο τα απαραίτητα ερωτήματα για την κάλυψη του επιδιωκόμενου σκοπού.

β) Τα ερωτήματα πρέπει να είναι πρακτικά. Δεν πρέπει δηλαδή να ζητάμε τη γνώμη κάποιου ατόμου για γεγονότα που συνέβησαν πριν πολύ καιρό, οπότε δεν θα θυμάται να δώσει σωστή απάντηση. Επίσης, δεν πρέπει να κάνουμε ερωτήσεις για προσωπικές υποθέσεις που αφορούν κάποιο άτομο, οπότε είναι πιθανό να δώσει μεροληπτική απάντηση.

γ) Το περιεχόμενο των ερωτημάτων πρέπει να είναι διατυπωμένο κατά τρόπο ώστε να γίνεται κατανοητό από τον ερευνώμενο.

δ) Κατά τη σύνταξη των ερωτημάτων πρέπει να λαμβάνεται υπόψη το επίπεδο εκπαίδευσης των ερευνωμένων, έτσι ώστε να έχουμε την ανάλογη επιλογή της γλώσσας για τη διατύπωση των ερωτημάτων, προς διευκόλυνση της επικοινωνίας με τον ερευνώμενο. Π.χ. όταν το ερωτηματολόγιο μιας έρευνας απευθύνεται στο ευρύ κοινό, πρέπει να λαμβάνονται υπόψη τα άτομα που έχουν το μικρότερο επίπεδο εκπαίδευσης και να μη γίνεται χρήση εξειδικευμένων όρων οι οποίοι είναι άγνωστοι στο ευρύ κοινό.

ε) Όταν σε κάποια ερωτήματα χρησιμοποιούνται όροι όπως «νοικοκυριό», «οικογένεια», «κατάστημα», «επιχείρηση», κ.λπ. πρέπει να δίνεται και ο σχετικός ορισμός, διότι κάθε ένας από τους όρους αυτούς έχει και διαφορετική έννοια. Π.χ. το νοικοκυριό μπορεί να αποτελείται από άτομα που δεν έχουν καμία συγγένεια μεταξύ τους, αρκεί να διαμένουν στην ίδια κατοικία, να προμηθεύονται από κοινού τα απαραίτητα για τη συντήρησή τους και να τρώνε κατά κανόνα μαζί. Αντίθετα ο όρος «οικογένεια» προϋποθέτει την ύπαρξη συγγένειας μεταξύ των ατόμων. Επίσης, ο όρος «επιχείρηση» έχει ευρύτερη έννοια από τον όρο «κατάστημα», δεδομένου ότι μια επιχείρηση μπορεί να αποτελείται από ένα ή δύο και περισσότερα καταστήματα.

στ) Η σειρά των ερωτημάτων μπορεί να επηρεάσει το βαθμό ανταπόκρισης σε μια έρευνα. Στην αρχή πρέπει να τίθενται τα ερωτήματα που ενδιαφέρουν άμεσα τον ερευνώμενο και αυτά που προϋποθέτουν εύκολες απαντήσεις.

ζ) Για τα ερωτήματα τα οποία χρειάζονται περιορισμένη απάντηση, πρέπει να γίνεται προκωδικογράφηση. Π.χ. στο ερώτημα «οικογενειακή κατάσταση» του ερωτηματολογίου της απογραφής πληθυσμού 1991, η απάντηση πρέπει να δοθεί με ένα Χ σε μια από τις παρακάτω περιπτώσεις:

Άγαμος	□ 1
Έγγαμος	□ 2
Χήρος (-α)	□ 3
Διαζευγμένος (-η)	□ 4

Είναι φανερό ότι με την προκωδικογράφηση των απαντήσεων διευκολύνεται σημαντικά η επεξεργασία του ερωτηματολογίου, αφού δεν χρειάζεται να τεθούν εκ των υστέρων οι αντίστοιχοι κωδικοί. Εάν για παράδειγμα δεν είχαν γραφεί οι κωδικοί 1, 2, 3, 4 δίπλα από τα αντίστοιχα τετραγωνίδια, θα έπρεπε κατά το στάδιο της επεξεργασίας να γραφούν τόσες φορές, όσος είναι και ο αριθμός των ατόμων που είναι άγαμα, έγγαμα, κ.ο.κ.

η) Τέλος, κατά την κατάρτιση του ερωτηματολογίου πρέπει κάτω από κάθε ερώτημα να δίνονται συνοπτικές οδηγίες, οι οποίες διευκολύνουν τον ερευνητή να μεταβιβάσει στον ερευνώμενο τι ακριβώς ζητείται μέσω των ερωτημάτων. Π.χ. στο ερώτημα «οικογενειακή κατάσταση» του ερωτηματολογίου της απογραφής πληθυσμού 1991, αν δε γράφεται η διευκρίνιση : «για τα άτομα που βρίσκονται σε διάσταση, αλλά δεν έχουν πάρει διαζύγιο, να δοθεί απάντηση στην περίπτωση έγγαμος», είναι πιθανό να δοθεί εσφαλμένη απάντηση.

Οδηγίες για τον τρόπο συμπλήρωσης των ερωτημάτων που δίνονται συνήθως σε ξεχωριστά φυλλάδια και ειδικότερα όταν αυτά είναι πολυσέλιδα δεν διαβάζονται από τους ερευνητές, με συνέπεια την αύξηση των μη δειγματοληπτικών σφαλμάτων.

Δοκιμαστική έρευνα

Η διενέργεια μιας δοκιμαστικής έρευνας (pilot survey) εξυπηρετεί πολλούς σκοπούς. Ο κυριότερος σκοπός είναι η υποβολή του ερωτηματολογίου, σε δοκιμή προκειμένου να γίνουν οι τελικές διορθώσεις, πριν από την έναρξη της κύριας έρευνας. Μεταξύ των άλλων σκοπών είναι η δοκιμή της επάρκειας των οδηγιών που έχουν δοθεί στους ερευνητές, η δοκιμή της εποπτείας και της οργάνωσης στην πράξη, καθώς και ο υπολογισμός του απαιτούμενου χρόνου για τη συμπλήρωση των ερωτηματολογίων και την επεξεργασία αυτών.

Επίσης από τη δοκιμαστική έρευνα έχουμε τις πρώτες ενδείξεις σχετικά με το ποσοστό μη ανταπόκρισης.

Επισημαίνεται ότι η επιτυχία μιας έρευνας εξασφαλίζεται όταν η δοκιμαστική έρευνα γίνεται με συνθήκες όσο το δυνατόν παρόμοιες με την κύρια έρευνα.

Μέθοδοι συλλογής των στοιχείων στις δειγματοληπτικές έρευνες

Οι μέθοδοι συλλογής των στοιχείων είναι η προσωπική συνέντευξη και η ταχυδρομική αποστολή.

- **Η προσωπική συνέντευξη**

Η προσωπική συνέντευξη (personal interview) είναι ο καλύτερος τρόπος συλλογής στατιστικών στοιχείων και χρησιμοποιείται σήμερα πάρα πολύ στις δειγματοληπτικές έρευνες.

Σύμφωνα με τη μέθοδο αυτή, η συμπλήρωση του ερωτηματολογίου γίνεται από ειδικά εκπαιδευμένα άτομα που ονομάζονται ερευνητές ή ερευνήτριες.

Οι εργασίες των ερευνητών είναι οι εξής :

- Η μελέτη των οδηγιών της έρευνας
- Ο εντοπισμός των μονάδων του δείγματος
- Η πραγματοποίηση των συνεντεύξεων
- Ο έλεγχος των ερωτηματολογίων για τυχόν λάθη ή παραλείψεις

Η μέθοδος της προσωπικής συνέντευξης έχει τα εξής πλεονεκτήματα:

α) Συγκεντρώνονται στοιχεία, ποιοτικά καλύτερα από ότι με τη μέθοδο της ταχυδρομικής αποστολής, διότι δίνονται διευκρινήσεις στους ερευνώμενους, έτσι ώστε να δίνουν σωστές απαντήσεις

β) Έχουμε μεγαλύτερο ποσοστό ανταπόκρισης των ερευνωμένων, πολλές φορές μέχρι και 100%

γ) Τα στοιχεία μπορεί να συγκεντρωθούν απ' ευθείας με κομπιούτερ, τα λεγόμενα portable computer, όπου ο ερευνητής κάνει εισαγωγή των στοιχείων απ' ευθείας σε ερωτηματολόγιο που εμφανίζεται με ειδικό πρόγραμμα στην οθόνη του κομπιούτερ. Με τον τρόπο αυτό κερδίζουμε χρόνο, διότι δεν χρειάζεται να γίνει εισαγωγή των στοιχείων σε μαγνητικά μέσα από το ερωτηματολόγιο που θα συμπλήρωνε ο ερευνητής με το χέρι. Επίσης επιτυγχάνεται καλύτερη ποιότητα στοιχείων, διότι περιορίζονται τα σφάλματα επεξεργασίας (κωδικογράφηση και εισαγωγή στοιχείων σε μαγνητικά μέσα).

Τα μειονεκτήματα της προσωπικής συνέντευξης είναι ότι αυτή απαιτεί μεγαλύτερο κόστος σε σύγκριση με την ταχυδρομική αποστολή και ότι σε ορισμένες ερωτήσεις των ερωτηματολογίων μπορεί να δοθούν μεροληπτικές απαντήσεις λόγω της παρουσίας του ερευνητή. Ειδικά σε ερωτήσεις που θίγουν τον εγωισμό ενός ατόμου.

- **Ταχυδρομική αποστολή του ερωτηματολογίου**

Η μέθοδος της ταχυδρομικής αποστολής του ερωτηματολογίου (mail questionnaire), δηλαδή αποστολή του ερωτηματολογίου στον ερευνώμενο, η συμπλήρωση από αυτόν και η επιστροφή του με το ταχυδρομείο, θεωρείται ο πιο εύκολος τρόπος συλλογής στατιστικών στοιχείων. Η εφαρμογή της μεθόδου αυτής είναι αρκετά περιορισμένη λόγω των μειονεκτημάτων που παρουσιάζει και τα οποία αναφέρονται παρακάτω. Σε ορισμένες όμως περιπτώσεις είναι ο μόνος τρόπος συλλογής των στοιχείων, όπως συμβαίνει όταν ο ερευνώμενος πληθυσμός είναι διασκορπισμένος σε μεγάλες εκτάσεις, ο διαθέσιμος χρόνος και οι χρηματικοί πόροι περιορισμένοι.

Τα πλεονεκτήματα της μεθόδου αυτής είναι το μικρό κόστος, η ταχεία διεξαγωγή της έρευνας και η δυνατότητα συγκέντρωσης εμπιστευτικών πληροφοριών ειδικότερα όταν το ερωτηματολόγιο είναι ανώνυμο. Τα μειονεκτήματα της μεθόδου αυτής είναι πρώτον ότι μόνο ένα μικρό ποσοστό απαντά, συνήθως κάτω του 30% και δεύτερον ότι τα μη δειγματοληπτικά σφάλματα κατά κανόνα είναι σημαντικά.

Μέθοδοι διενέργειας της δειγματοληψίας

Για τη διενέργεια των δειγματοληπτικών ερευνών χρησιμοποιούνται κυρίως οι εξής μέθοδοι :

- Απλή τυχαία δειγματοληψία
- Συστηματική δειγματοληψία
- Στρωματοποιημένη δειγματοληψία
- Δειγματοληψία κατά ομάδες
- Δισταδιακή δειγματοληψία
- Τρισταδιακή δειγματοληψία
- Άλλοι μέθοδοι δειγματοληψίας (panels, επιφανειακή δειγματοληψία κ.λπ.)

Η εφαρμογή της μεθόδου αποφασίζεται με βάση τα παρακάτω κριτήρια:

- Από το ζητούμενο βαθμό ακριβείας των αποτελεσμάτων
- Από τα χρονικά και χρηματικά περιθώρια της έρευνας
- Από τη μεταβλητότητα των μονάδων του ερευνώμενου πληθυσμού
- Από την ύπαρξη δειγματοληπτικών πλαισίων των μονάδων του πληθυσμού
- Από τη δυνατότητα που υπάρχει για να μπορεί να διαιρεθεί ο ερευνώμενος πληθυσμός σε υποπληθυσμούς με μεγάλη ομοιογένεια

Ενημερωτική επιστολή

Η ενημερωτική επιστολή (informative latter) αποτελεί βασικό παράγοντα στην επιτυχία μιας δειγματοληπτικής έρευνας.

Η ενημερωτική επιστολή έχει σαν σκοπό να προετοιμάσει τα ερευνώμενα άτομα, ώστε να συνεργασθούν με προθυμία και ειλικρίνεια με τον ερευνητή που θα τα επισκεφθεί. Έτσι επιτυγχάνεται μεγαλύτερο ποσοστό ανταπόκρισης και βελτίωσης της ποιότητας των στοιχείων. Ειδικά όταν πρόκειται για συμπλήρωση του ερωτηματολογίου από τον ίδιο τον ερευνώμενο (self – completion), η ενημερωτική επιστολή είναι περισσότερο απαραίτητη.

ΚΕΦΑΛΑΙΟ 4^{Ο}

ΚΑΤΑΝΟΜΕΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

4.1 ΕΙΣΑΓΩΓΗ

Στη στατιστική ενδιαφερόμαστε να εκτιμήσουμε κάποιους άγνωστους δείκτες του πληθυσμού μέσα από δείκτες του δείγματος. Αυτοί οι δείκτες (δειγματοσυναρτήσεις ή δειγματικοί δείκτες ή εκτιμήτριες συναρτήσεις-sample statistics) είναι και οι ίδιοι τυχαίες μεταβλητές Έστω X1, X2, … , X n τυχαίες μεταβλητές που δείχνουν την τιμή ενός χαρακτηριστικού (π.χ. ύψος) για κάθε άτομο στο δείγμα Εφόσον ο δειγματικός μέσος= (X1+X2+ … +Xn)/n είναι μια συνάρτηση τυχαίων μεταβλητών, δε θα είναι και αυτός τυχαία μεταβλητή.

Άρα λοιπόν οι εκτιμήτριες συναρτήσεις χρησιμοποιούνται για να υπολογίσουμε τις πληθυσμιακές παραμέτρους έτσι λοιπόν ο δειγματικός μέσος X εκτιμάει τον πραγματικό μέσο μ.

ΠΡΟΒΛΗΜΑ: Διαφορετικά δείγματα δίνουν διαφορετικές εκτιμήσεις:

Μεγάλα δείγματα δίνουν καλύτερες εκτιμήσεις αλλά κοστίζουν περισσότερο.

Πως ξέρουμε αν είναι καλή η εκτίμηση (και η εκτιμήτρια συνάρτηση που επιλέξαμε);

ΛΥΣΗ: στα παραπάνω μπορούμε να βρούμε μελετώντας τη συνάρτηση κατανομής της κάθε εκτιμήτριας συνάρτησης.

Η κατανομή δειγματοληψίας μιας εκτιμήτριας συνάρτησης μπορεί να υπολογιστεί αν πάρουμε όλα τα πιθανά δείγματα (που προκύπτουν από τον πληθυσμό) του ίδιου μεγέθους.

4.2 ΘΕΩΡΗΤΙΚΕΣ ΚΑΤΑΝΟΜΕΣ

Όταν γνωρίζουμε το χαρακτηριστικά ενός στατιστικού πληθυσμού μπορούμε, με μαθηματική συλλογιστική, να βγάλουμε ορισμένα συμπεράσματα για τα χαρακτηριστικά του δείγματος. Αυτή είναι μια διαδικασία αναγωγικής συμπερασματικής (deduotive inference) με την οποία από το ολικό αποκτούμε γνώση το μερικό, από το καθολικό σύνολο για το υποσύνολο, από τον πληθυσμό για το δείγμα.

Η αντίθετη διαδικασία γίνεται στην επαγωγική συμπερασματική κατά την οποία η γνώση για το μερικό γενικεύεται στο ολικό, ΟΛΟ το υποσύνολο στο καθολικό σύνολο και από το δείγμα στον πληθυσμό.

Η αναγωγική συμπερασματική χρησιμοποιείται στις θεωρητικές επιστήμες για αποδείξεις. Η επαγωγική συμπερασματική χρησιμοποιείται στις εμπειρικές επιστήμες για την. παραγωγή νέας γνώσης. Στις επιστήμες αυτές η διαδικασία παραγωγής γνώσης συνίσταται στην εκτέλεση ενός πειράματος, στην εξαγωγή συμπερασμάτων με βάση τα δεδομένα του πειράματος αυτού και τέλος, στη γενίκευση των συμπερασμάτων από το συγκεκριμένο πείραμα σε όλα τα όμοια πειράματα, η οποία θα μπορούσαν να είχαν γίνει. Η γενίκευση της μερικής γνώσης η οποία γίνεται με την επαγωγική συμπερασματική συνεπάγεται αβεβαιότητα και. όπως είδαμε η αβεβαιότητα μετριέται με την πιθανότητα.

Για τον ευκολότερο υπολογισμό πιθανοτήτων σε οποιοδήποτε δειγματικό χώρο θα ήταν χρήσιμο να ορισθεί μία συνάρτηση η οποία να αντιστοιχεί έναν πραγματικό αριθμό σε κάθε στοιχείο του δειγματικού χώρου. Η συνάρτηση αυτή ονομάζεται **τυχαία μεταβλητή.** Έτσι τα στοιχεία του δειγματικού χώρου (με άλλα λόγια τα απλά ενδεχόμενα) μεταφέρονται μέσω της τυχαίας μεταβλητής σε πραγματικούς αριθμούς. Οι πιθανότητες επάγονται στο πεδίο τιμών της τυχαίας μεταβλητής μέσω των πιθανοτήτων που έχουν ορισθεί στο δειγματικό χώρο.

Οι τυχαίες μεταβλητές διακρίνονται σε **διακριτές** και **συνεχείς**. Μια τυχαία μεταβλητή ονομάζεται διακριτή όταν μπορεί να πάρει μόνο διακεκριμένες τιμές σε κάποιο διάστημα, ενώ συνεχής όταν μπορεί να πάρει οποιαδήποτε τιμή σε κάποιο διάστημα. Από τον ορισμό της συνάρτησης πιθανότητας και τα αξιώματα των πιθανοτήτων είναι φανερό ότι

i. $P(X = x) \geq 0$ για κάθε x.

ii. $\sum P(X = x) = 1$

Οι ιδιότητες αυτές δίνουν τις ικανές και αναγκαίες συνθήκες που πρέπει να πληρεί μία συνάρτηση για να είναι συνάρτηση πιθανότητας.

Το αντίστοιχο συμπέρασμα για μια συνεχή τυχαία μεταβλητή Χ εκφράζεται ως:

i. $f(x) \geq 0$

ii. $\int_{-\infty}^{+\infty} f(x)dx = 1$

η συνάρτηση πυκνότητας πιθανότητας (σ.π.π.) της τυχαίας μεταβλητής.

Κάθε τυχαία μεταβλητή έχει τη δική της κατανομή πιθανότητας αλλά σε πολλές περιπτώσεις οι κατανομές αυτές έχουν μεγάλες ομοιότητες. Μπορούμε λοιπόν να δημιουργήσουμε κάποιες βασικές μορφές κατανομών που τις ονομάζουμε **θεωρητικές κατανομές** και τις οποίες χρησιμοποιούμε αντί των πραγματικών κατανομών.

Η επιλογή της κατάλληλης θεωρητικής κατανομής μας επιτρέπει να μελετήσουμε με μεγαλύτερη ευκολία αλλά και ακρίβεια την κατανομή της τυχαίας μεταβλητής που εξετάζουμε.

Λόγω ακριβώς της μεγάλης τους χρησιμότητας τα χαρακτηριστικά και οι ιδιότητες των κατανομών αυτών έχουν μελετηθεί διεξοδικά και τα αποτελέσματα διαφόρων υπολογισμών που χρησιμοποιούνται συχνά έχουν συγκεντρωθεί σε εύχρηστους πίνακες.

Οι σπουδαιότερες από τις κατανομές αυτές, τόσο διακριτές όσο και συνεχείς θα παρουσιαστούν στη συνέχεια.

4.3 ΔΙΑΚΡΙΤΕΣ ΚΑΤΑΝΟΜΕΣ

Στην ενότητα αυτή θα μελετήσουμε δύο διακριτές κατανομές οι οποίες παρουσιάζουν ιδιαίτερο ενδιαφέρον κυρίως λόγω των πολλών εφαρμογών τους, τη Διωνυμική και την Poisson

4.3.1 ΔΙΩΝΥΜΙΚΗ ΚΑΤΑΝΟΜΗ

Η **Διωνυμική κατανομή** συνδέεται με ένα πολύ απλό πείραμα τύχης. Ίσως το απλούστερο! Πρόκειται για τη **δοκιμή Bernoulli**[1], ένα πείραμα τύχης με δύο μόνο- αμοιβαίως αποκλειόμενα - δυνατά αποτελέσματα. Το ένα αποτέλεσμα ονομάζεται **επιτυχία** και το άλλο **αποτυχία**.

Το πιο «δημοφιλές» στη βιβλιογραφία παράδειγμα δοκιμής Bernoulli είναι η ρίψη ενός νομίσματος. Τα δυνατά αποτελέσματα είναι προφανώς μόνο δύο: «κεφαλή» ή «γράμματα». Αν μας ενδιαφέρει η ένδειξη «κεφαλή» χαρακτηρίζουμε επιτυχία το αποτέλεσμα «κεφαλή» και αποτυχία το αποτέλεσμα «γράμματα» ενώ αν μας ενδιαφέρει η ένδειξη «γράμματα» χαρακτηρίζουμε επιτυχία το αποτέλεσμα «γράμματα» και αποτυχία το αποτέλεσμα «κεφαλή». Ας δούμε μερικά ακόμη παραδείγματα: Μας ενδιαφέρει να διατυπώσουμε πιθανοθεωρητικά συμπεράσματα **α)** για τον αριθμό των ζώων μιας κτηνοτροφικής μονάδας που έχουν προσβληθεί από μια συγκεκριμένη ασθένεια. Η εξέταση ενός ζώου για το αν έχει προσβληθεί ή όχι από την ασθένεια είναι δοκιμή Bernoulli γιατί τα δυνατά αποτελέσματα είναι μόνο δύο: το ζώο είτε έχει προσβληθεί (επιτυχία) είτε δεν έχει προσβληθεί (αποτυχία).

β) για τον αριθμό αλλοιωμένων προϊόντων που είναι αποθηκευμένα στις εγκαταστάσεις μιας βιομηχανικής μονάδας επεξεργασίας αγροτικών προϊόντων. Ο έλεγχος ενός προϊόντος για το αν είναι αλλοιωμένο ή όχι

---

1 .

είναι δοκιμή Bernoulli γιατί το προϊόν είτε έχει αλλοιωθεί (επιτυχία) είτε δεν έχει αλλοιωθεί (αποτυχία).

γ) για τον αριθμό των φυτών μιας καλλιέργειας που η ξηρή φυτική μάζα τους ξεπερνάει τα 150gr. Ο έλεγχος ενός φυτού για το αν η ξηρή μάζα του ξεπερνάει ή όχι τα 150gr είναι δοκιμή Bernoulli γιατί το φυτό έχει ξηρή μάζα είτε μεγαλύτερη από 150gr (επιτυχία) είτε το πολύ 150gr (αποτυχία).

δ) για τον αριθμό των φυτών μιας καλλιέργειας που έχουν λιγότερα από 6 φύλλα. Ο έλεγχος ενός φυτού για το αν έχει λιγότερα από 6 φύλλα είναι δοκιμή Bernoulli γιατί το φυτό έχει είτε λιγότερα από 6 (επιτυχία) είτε τουλάχιστον 6 φύλλα (αποτυχία).

Σημείωση: Παρατηρείστε ότι οι δυνατές τιμές της ξηρής μάζας ενός φυτού στο παράδειγμα (γ) προφανώς δεν είναι μόνο δύο. Το ίδιο ισχύει και για τον αριθμό των φύλλων ενός φυτού στο παράδειγμα (δ). Όμως, με βάση το τι ενδιαφέρει στην αντίστοιχη έρευνα, ταξινομήσαμε τις δυνατές τιμές σε δύο κατηγορίες-αποτελέσματα και οδηγηθήκαμε έτσι σε δοκιμές Bernoulli. Δοκιμή Bernoulli έχουμε επίσης σε περιπτώσεις όπου, ενώ τα δυνατά εξαγόμενα είναι περισσότερα από δύο (όπως και στα παραδείγματα (γ) και (δ)), μας ενδιαφέρει εάν συμβαίνει ή όχι, μόνο ένα συγκεκριμένο. Για παράδειγμα, ένα παιδί που επιλέγεται τυχαία από τα n παιδιά μιας οικογένειας μπορεί να έχει το γονότυπο ΑΑ ή όχι[2]. Επίσης, από μία τράπουλα επιλέγουμε τυχαία ένα παιγνιόχαρτο το οποίο μπορεί να είναι ή να μην είναι «άσσος».

2 *Ο γονότυπος μπορεί να είναι ΑΑ, Αα ή αα.*

Σε μια δοκιμή Bernoulli είναι φανερό ότι ο αριθμός των επιτυχιών είναι ή 1 ή 0. Αν η πιθανότητα επιτυχίας είναι p προφανώς η πιθανότητα αποτυχίας είναι $1-p$ την οποία συμβολίζουμε με q. Ας συμβολίσουμε επίσης τον αριθμό των επιτυχιών σε μια δοκιμή Bernoulli με X. Η κατανομή της τυχαίας μεταβλητής X ονομάζεται **κατανομή Bernoulli** με παράμετρο p και έχει συνάρτηση πιθανότητας

$f(x) = P(X = x) = p^x \cdot (1-p)^{1-x},\ x = 0, 1.$ ή $f(x) = P(X = x) = \begin{cases} p, & x = 1 \\ 1-p, & x = 0. \end{cases}$

Η μέση τιμή m της X είναι $m = E(X) = 0 \cdot f(0) + 1 \cdot f(1) = p$ και η διασπορά της $s^2 = E(X^2) - [E(X)]^2 = 0^2 \cdot f(0) + 1^2 \cdot f(1) - p^2 = p - p^2 = p \cdot (1-p) = p \cdot q$.

Παρότι, όπως αναφέραμε, η δοκιμή Bernoulli είναι ένα πολύ απλό πείραμα, εντούτοις (ή μήπως γι' αυτό) βρίσκεται στον πυρήνα πολλών πραγματικών προβλημάτων με ευρύτατο φάσμα εφαρμογών που παρουσιάζουν μάλιστα ιδιαίτερο ενδιαφέρον και σε αρκετές περιπτώσεις ιδιαίτερες δυσκολίες στην αντιμετώπισή τους. Πιο συγκεκριμένα, πολλά πραγματικά προβλήματα αναλύονται σε μια σειρά–ακολουθία δοκιμών Bernoulli και τα τελικά ερωτήματα που απορρέουν από αυτά σχετίζονται με τον αριθμό των επιτυχιών που συμβαίνουν. Για παράδειγμα, το τελικό ερώτημα μετά την ανάλυση ενός τέτοιου προβλήματος μπορεί να είναι: «πόσες επιτυχίες συμβαίνουν σε n επαναλήψεις μιας δοκιμής Bernoulli» ή «πόσες επαναλήψεις μιας δοκιμής Bernoulli απαιτούνται μέχρι την πρώτη επιτυχία» ή «πόσες επαναλήψεις μιας δοκιμής Bernoulli απαιτούνται μέχρι την r επιτυχία» ή «πόσες επαναλήψεις μιας δοκιμής

Bernoulli απαιτούνται μέχρι να συμβεί μια ροή k συνεχόμενων επιτυχιών».

Στη συνέχεια θα ασχοληθούμε με προβλήματα που αναλύονται σε μια ακολουθία n ανεξάρτητων επαναλήψεων μιας δοκιμής Bernoulli και που τα πιθανοθεωρητικά συμπεράσματα που μας ενδιαφέρουν συνδέονται με το ερώτημα: «πόσες επιτυχίες συμβαίνουν σε n επαναλήψεις μιας δοκιμής Bernoulli». Ας δούμε ένα τέτοιο πρόβλημα.

<u>Πρόβλημα 1</u>

Ο γεωπόνος ενός φυτώριου ισχυρίζεται ότι ποσοστό 90% ενός είδους φυτών που παράγει δίνει περισσότερους από 5 καρπούς/φυτό. Ένας αγρότης που προμηθεύτηκε από το συγκεκριμένο φυτώριο μεγάλο αριθμό φυτών αυτού του είδους για να τα καλλιεργήσει, θέλησε να ελέγξει τον ισχυρισμό του γεωπόνου. Για το σκοπό αυτό, κατά τη συγκομιδή, επέλεξε τυχαία 20 φυτά και μέτρησε τους καρπούς κάθε φυτού. Από τα 20 φυτά που εξέτασε, περισσότερους από 5 καρπούς είχαν μόνο τα 12, γεγονός που δημιούργησε αμφιβολίες στον αγρότη για τον ισχυρισμό του γεωπόνου. Είναι, άραγε, δικαιολογημένες οι αμφιβολίες του αγρότη;

Είναι προφανές ότι η τυχαία επιλογή και εξέταση ενός φυτού, για το αν έχει ή όχι περισσότερους από 5 καρπούς, είναι μια δοκιμή Bernoulli με δύο μόνο δυνατά αποτελέσματα: το φυτό είτε έχει περισσότερους από 5 καρπούς είτε δεν έχει περισσότερους από 5 καρπούς. Η επιλογή επομένως και η εξέταση 20 φυτών είναι μια σειρά-ακολουθία 20 δοκιμών Bernoulli. Επειδή ο αγρότης ενδιαφέρεται για τα φυτά που έχουν περισσότερους από πέντε καρπούς είναι λογικό να ονομάσουμε επιτυχία το αποτέλεσμα «το φυτό έχει περισσότερους από πέντε καρπούς» και αποτυχία το αποτέλεσμα «το φυτό δεν έχει περισσότερους από πέντε καρπούς».

Το ερώτημα που, λογικά, προκύπτει είναι: αν αποδεχθούμε τον ισχυρισμό του γεωπόνου, πόσο λογικό-πιθανό είναι το αποτέλεσμα του ελέγχου που έκανε ο αγρότης; Δηλαδή, με δεδομένο ότι το ποσοστό των φυτών που παράγουν περισσότερους από 5 καρπούς είναι 90%, πόσο πιθανό είναι από τα 20 τυχαία επιλεγμένα φυτά να βρεθούν 12 τα οποία να έχουν περισσότερους από 5 καρπούς; Ή αλλιώς, ποια είναι η πιθανότητα, σε 20 επαναλήψεις μιας δοκιμής Bernoulli να συμβούν 12 επιτυχίες, με δεδομένο ότι σε κάθε επανάληψη η πιθανότητα επιτυχίας p είναι ίση με 0.9.

Πρέπει να σημειώσουμε ότι σωστά θεωρήσαμε ότι σε κάθε επανάληψη της δοκιμής Bernoulli η πιθανότητα επιτυχίας παραμένει σταθερή και ίση με 0.9 γιατί ο αριθμός των φυτών που εξετάζονται είναι πολύ μικρός σε σχέση με τον αριθμό των φυτών της καλλιέργειας και επομένως, παρότι, η δειγματοληψία γίνεται χωρίς επανάθεση η πιθανότητα αυτή πρακτικά δεν αλλάζει. Είναι επίσης λογικό να θεωρήσουμε ότι το αποτέλεσμα οποιασδήποτε δοκιμής δεν επηρεάζεται-εξαρτάται από το αποτέλεσμα των προηγούμενων δοκιμών.

Το γενικότερο επομένως ερώτημα που τίθεται είναι:
Αν X ο αριθμός των επιτυχιών σε μια ακολουθία n ανεξάρτητων δοκιμών Bernoulli με σταθερή πιθανότητα επιτυχίας p σε όλες τις δοκιμές, πώς μπορούμε να υπολογίσουμε τις πιθανότητες: $P(X = x), \quad x = 0,1,2,\ldots, n.$

Εύκολα μπορεί να αποδειχθεί[3] ότι οι πιθανότητες αυτές δίνονται από τον τύπο:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0,1,2,\ldots,n.$$

Έτσι, αν συμβολίσουμε με X τον αριθμό των φυτών (από τα 20 που επελέγησαν τυχαία) καθένα από τα οποία έχει περισσότερους από 5 καρπούς, τότε:

$$P(X = 12) = \binom{20}{12} 0.9^{12} (1-0.9)^8 = 0.0003 .$$

Δηλαδή, με την υπόθεση ότι αληθεύει ο ισχυρισμός του γεωπόνου, η πιθανότητα να συμβεί αυτό που συνέβη στον έλεγχο που έκανε ο αγρότης είναι πολύ μικρή (σχεδόν μηδενική). Άρα οι αμφιβολίες του αγρότη έχουν βάση[4]. (Φυσικά, ο αγρότης δεν υπολόγισε την πιθανότητα $P(X = 12)$, απλώς υπολόγισε το 90% του 20 που είναι 18 και περίμενε να βρει περίπου 18 δένδρα με περισσότερους από 5 καρπούς.).

---

3 .
4

Ας δώσουμε τώρα τον ορισμό της **Διωνυμικής Κατανομής**[5]:

Ορισμός

Αν X ο αριθμός των επιτυχιών σε μια ακολουθία n ανεξάρτητων δοκιμών Bernoulli με σταθερή πιθανότητα επιτυχίας p σε όλες τις δοκιμές, τότε η κατανομή της τυχαίας μεταβλητής X καλείται **διωνυμική κατανομή** με παραμέτρους n και p και συμβολίζεται με $B(n,\ p)$.

Όπως ήδη αναφέραμε, η συνάρτηση πιθανότητας της διωνυμικής κατανομής με παραμέτρους n και p δίδεται από τον τύπο:

$$f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0,1,2,\ldots,n.$$

Παρατηρείστε ότι οι πιθανότητες των τιμών 0, 1, 2, ..., n, της X είναι οι όροι του διωνυμικού αναπτύγματος (τύπος του διωνύμου του Νεύτωνα[6]) $[p + (1-p)]^n$. Έτσι εξηγείται και το όνομα της διωνυμικής κατανομής.

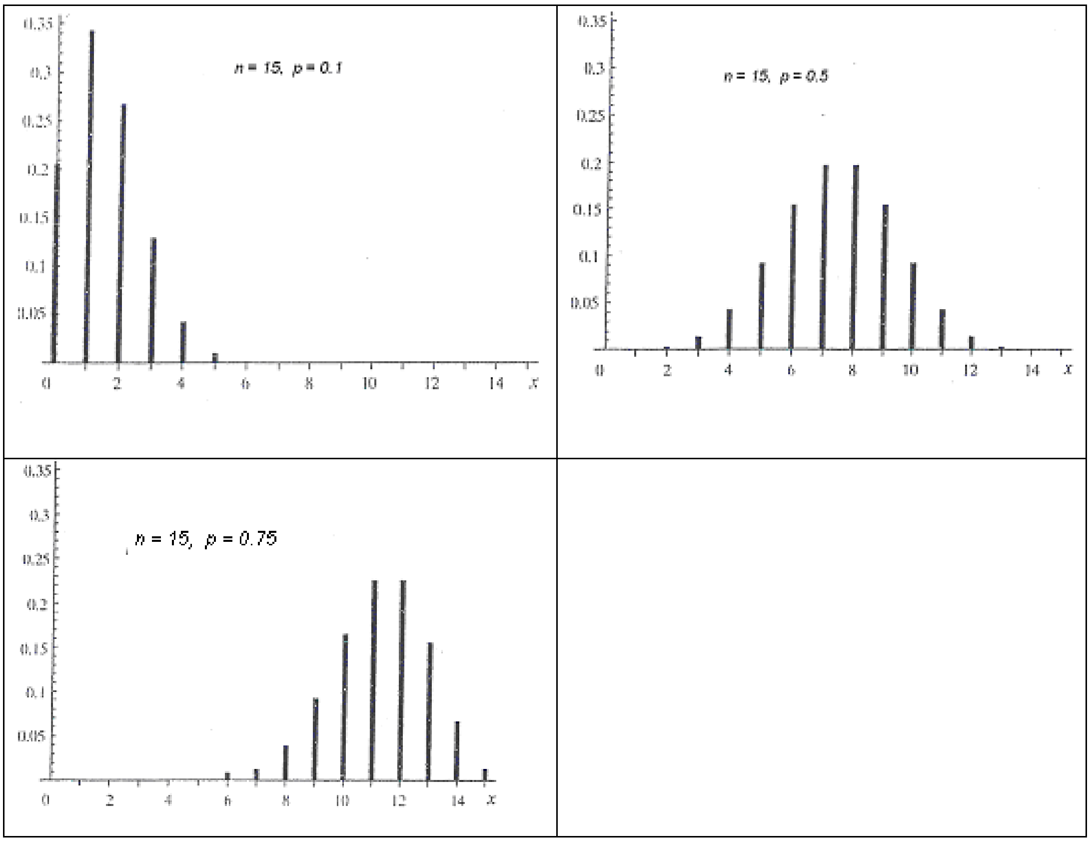
Επίσης, αποδεικνύεται ότι η μέση τιμή και η διασπορά της διωνυμικής τυχαίας μεταβλητής δίνονται από τους τύπους: $m = E(X) = np$ και $s^2 = V(X) = np(1-p) = npq$ αντίστοιχα. Έτσι, στο Πρόβλημα 1, η μέση τιμή και η διασπορά του αριθμού X των φυτών που έχουν περισσότερους από 5 καρπούς είναι $m = 20 \cdot 0.9 = 18$ φυτά και $s^2 = np(1-p) = 20 \cdot 0.9 \cdot 0.1 = 1.8$ (ή τυπική απόκλιση $s = \sqrt{1.8} = 1.34$ φυτά) αντίστοιχα.

5

6

Σχόλιο

Αν δούμε τη διασπορά, $s^2 = np(1-p)$, της διωνυμικής τυχαίας μεταβλητής ως συνάρτηση του p, $g(p) = np(1-p)$, παρατηρούμε ότι αν $0 \le p \le \frac{1}{2}$, η g είναι αύξουσα συνάρτηση του p ενώ αν $\frac{1}{2} \le p \le 1$ είναι φθίνουσα συνάρτηση του p και παρουσιάζει μέγιστο για $p = \frac{1}{2}$. Δηλαδή, η διασπορά ελαττώνεται όσο το p πλησιάζει το 0 ή το 1 και μεγιστοποιείται για $p = \frac{1}{2}$. Στα παρακάτω σχήματα, φαίνεται η συνάρτηση πιθανότητας της διωνυμικής κατανομής για $n = 15$ και $p = 0.1,\ 0.5,\ 0.75$.



Ας δούμε ένα ακόμη παράδειγμα υπολογισμού πιθανοτήτων μιας διωνυμικής τυχαίας μεταβλητής.

Πρόβλημα 2

Μια βιομηχανία κατασκευάζει μεταλλικά ελάσματα για να αντέχουν σε συγκεκριμένη καταπόνηση. Σύμφωνα με τις προδιαγραφές παραγωγής, κάθε τέτοιο έλασμα αντέχει στη συγκεκριμένη καταπόνηση με πιθανότητα 0.8. Επιλέγουμε τυχαία 9 τέτοια ελάσματα και τα υποβάλλουμε στη συγκεκριμένη καταπόνηση. Ποια είναι η πιθανότητα να αντέξουν α) το πολύ 2 ελάσματα, β) περισσότερα από 7 ελάσματα, γ) τουλάχιστον 2 ελάσματα και δ) λιγότερα από 6 και τουλάχιστον 4 ελάσματα.

Απάντηση

Έστω Χ ο αριθμός των ελασμάτων που θα αντέξουν την καταπόνηση (από τα 9 που θα ελεγχθούν). Προφανώς[7], $X \sim B(9,\ 0.7)$ και επομένως για τις ζητούμενες πιθανότητες έχουμε:

α) $P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2) =$

$$= \binom{9}{0} \cdot 0.8^0 0.2^9 + \binom{9}{1} \cdot 0.8^1 0.2^8 + \binom{9}{2} \cdot 0.8^2 0.2^7 = 0.0003$$

β) $P(X > 7) = P(X = 8) + P(X = 9) = \binom{9}{8} \cdot 0.8^8 0.2^2 + \binom{9}{8} \cdot 0.8^8 0.2^2 = 0.436207$

γ) $P(X \ge 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1) = ... = 1 - 0.000019 = 0.999981$

δ) $P(4 \le X < 6) = P(X = 4) + P(X = 5) = \binom{9}{4} \cdot 0.8^4 0.2^5 + \binom{9}{5} \cdot 0.8^5 0.2^4 = 0.0825754$

---

7 ;

Παρατηρήσεις

1. Είναι προφανές, ότι όταν το n είναι μεγάλο και το x όχι πολύ κοντά στο 0 ή το n, οι διωνυμικοί συντελεστές $\binom{n}{x}$ που εμφανίζονται στον τύπο

$$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}, \quad x=0,1,2,\ldots,n$$

παίρνουν μεγάλες τιμές με συνέπεια να γίνεται προβληματικός ο υπολογισμός των πιθανοτήτων. Ένας τρόπος αντιμετώπισης του προβλήματος είναι ο υπολογισμός των πιθανοτήτων μέσω αναδρομικού τύπου. Πράγματι, αποδεικνύεται[8] ότι:

> Για τις πιθανότητες $P(X=x),\ x=1,2,\ldots,n$, της τυχαίας μεταβλητής $X \sim B(n,p)$ ισχύει, $P(X=x)=\frac{n-x+1}{x}\cdot\frac{p}{1-p}P(X=x-1)$
> με αρχική συνθήκη: $P(X=0)=(1-p)^n$.

Έτσι, αν υπολογίσουμε μόνο την πιθανότητα $P(X=0)=(1-p)^n$, μπορούμε μέσω αυτού του αναδρομικού τύπου να υπολογίσουμε όλες τις πιθανότητες $P(X=1),\ P(X=2),\ \ldots,\ P(X=n)$ χωρίς να χρειασθεί να υπολογίσουμε τους αντίστοιχους διωνυμικούς συντελεστές. Ας δούμε ένα παράδειγμα.

8 .

Έστω $X \sim B(5,\ 0.3)$. Η πιθανότητα $P(X=3)$ μέσω του αναδρομικού τύπου υπολογίζεται ως εξής:

$$P(X=0) = 0.7^5 = 0.168$$

$$P(X=1) = \frac{5-1+1}{1}\frac{0.3}{1-0.3}0.168 = 0.360$$

$$P(X=2) = \frac{5-2+1}{2}\frac{0.3}{1-0.3}0.360 = 0.309$$

$$P(X=3) = \frac{5-3+1}{3}\frac{0.3}{1-0.3}0.309 = 0.132$$

Είναι όμως φανερό ότι αυτή η μέθοδος υπολογισμού των πιθανοτήτων έχει το εξής μειονέκτημα. Για να υπολογίσουμε μια συγκεκριμένη πιθανότητα $P(X=x)$ που μας ενδιαφέρει, πρέπει να προηγηθεί ο υπολογισμός των πιθανοτήτων $P(X=0),\ P(X=1),\ \ldots,\ P(X=x-1)$. Έτσι, όσο αυξάνει το x, αυξάνει (…δυστυχώς) και ο αριθμός των πιθανοτήτων που πρέπει να υπολογισθούν.

Στη συνέχεια, όταν θα μιλήσουμε για την κατανομή Poisson και την κανονική κατανομή, θα δούμε και άλλο τρόπο αντιμετώπισης των δυσκολιών υπολογισμού (στις περιπτώσεις που υπάρχουν) των πιθανοτήτων της διωνυμικής τυχαίας μεταβλητής[9].

2. Αν $X \sim B(n,\ p)$, άραγε από τις τιμές 0, 1, 2, ...,n της X, ποια είναι η πιο πιθανή. Παρατηρείστε στα σχήματα της σελίδας 76 ότι όσο οι τιμές της X αυξάνουν από το 0 στο n, οι αντίστοιχες πιθανότητες $P(X=x),\ \ x=0,1,2,\ldots,n$, μέχρι ένα σημείο (ας το συμβολίσουμε x_0)

9

αυξάνουν και στη συνέχεια φθίνουν. Πράγματι έτσι είναι. Ας το αποδείξουμε και ας προσδιορίσουμε αυτή την τιμή x_0 της X. Από τον αναδρομικό τύπο που αποδείξαμε προηγουμένως προκύπτει ότι:

$$\frac{P(X=x)}{P(X=x-1)}=\frac{n-x+1}{x}\cdot\frac{p}{1-p}.$$

Δηλαδή, $P(X=x)>P(X=x-1)$ αν και μόνο αν $(n-x+1)p>x(1-p)$ ή $x<(n+1)p$. Αυτό σημαίνει ότι: **α)** Όταν το γινόμενο $x<(n+1)p$ δεν είναι ακέραιος αριθμός, η πιο πιθανή τιμή x_0 της X είναι το ακέραιο μέρος του, δηλαδή, $x_0=[(n+1)p]$ διότι οι πιθανότητες $P(X=x)$ αυξάνουν γνησίως για $x\le[(n+1)p]$ και φθίνουν γνησίως για $x\ge[(n+1)p]+1$. **β)** Όταν το γινόμενο $(n+1)p$ είναι ακέραιος αριθμός, οι πιθανότητες $P(X=x)$ αυξάνουν γνησίως για $x\le(n+1)p-1$ και φθίνουν γνησίως για $x\ge(n+1)p+1$. Και επειδή $P[X=(n+1)p]=P[X=(n+1)p-1]$, στην περίπτωση αυτή, οι τιμές της X με τη μεγαλύτερη πιθανότητα είναι δύο: η $x_0=(n+1)p$ και η $x_0'=(n+1)p-1$.

Έτσι, στο Πρόβλημα 1, επειδή το γινόμενο $(n+1)p=21\cdot 0.9=18.9$ δεν είναι ακέραιος, η πιο πιθανή τιμή του αριθμού X των φυτών που έχουν περισσότερους από 5 καρπούς είναι η $x_0=[18.9]=18$ [10]. Στο Πρόβλημα 2, επειδή το γινόμενο $(n+1)p=10\cdot 0.8=8$ είναι ακέραιος, οι τιμές με τη μεγαλύτερη πιθανότητα είναι δύο: η $x_0=(n+1)p=8$ και η $x_0'=(n+1)p-1=7$.

10

Ο ορισμός και η μελέτη της Διωνυμικής Κατανομής προϋποθέτει.

$$P(X=k)=\binom{n}{k}p^k q^{n-k}=\frac{n!}{k!(n-k)!}p^k q^{n-k} \quad ,0\le k\le n,\quad 0<p<1,\quad q=1-p$$

Για την Διωνυμική κατανομή αποδεικνύεται ότι ισχύουν

$$E(X)=np \qquad V(X)=npq$$

Παράδειγμα Έστω ότι ρίχνουμε ένα ζάρι 5 φορές και ενδιαφερόμαστε αν το αποτέλεσμα κάθε ρίψης ήταν «1» ή «όχι 1». Να απαντηθούν τα ακόλουθα ερωτήματα:

1. Ποια η πιθανότητα να μην έρθει ούτε μια φορά στις 5 προσπάθειες το «1»?

2. Ποια η πιθανότητα να έρθει ακριβώς τρεις φορές στις 5 προσπάθειες το «1»?

3. Ποια η πιθανότητα να έρθει τουλάχιστο δύο φορές στις 5 προσπάθειες το «1»?

Λύση

Ας θεωρήσουμε ως επιτυχία (Ε) το αποτέλεσμα της ρίψης να είναι το “1” και ως αποτυχία (Α) το αποτέλεσμα να είναι «όχι 1» (κοινώς το «όχι 1» σημαίνει ότι το αποτέλεσμα μπορεί να είναι 2,3,4,5 ή 6).

Η πιθανότητα να έρθει «1» όταν ρίχνουμε ένα ζάρι είναι 1/6. Άρα η πιθανότητα επιτυχίας είναι p=1/6 και επομένως η πιθανότητα αποτυχίας θα είναι q=1-p=5/6.

Αν με X συμβολίσουμε το συνολικό αριθμό επιτυχιών στις 5 επαναλήψεις του πειράματος, τότε X~B(10,1/6). Έχουμε λοιπόν:

1. $$P(X=0)=\frac{5!}{0!(5-0)!}\left(\frac{1}{6}\right)^{0}\left(\frac{5}{6}\right)^{5-0}\cong 0,4019$$

2. $$P(X=3)=\frac{5!}{3!(5-3)!}\left(\frac{1}{6}\right)^{3}\left(\frac{5}{6}\right)^{5-3}=\frac{5\times 4\times 3!}{3!2!}\left(\frac{1}{6}\right)^{3}\left(\frac{5}{6}\right)^{2}=10\times\frac{5^2}{6^5}\cong 0,03215$$

3. $$P(X\geq 2)=1-P(X<1)=1-\left[P(X=0)+P(X=1)\right] \qquad (1)$$

όμως η πιθανότητα $P(X=0)$ υπολογίστηκε στο πρώτο ερώτημα ενώ για την $P(X=1)$ έχουμε:

$$P(X=1)=\frac{5!}{1!(5-1)!}\left(\frac{1}{6}\right)^{1}\left(\frac{5}{6}\right)^{5-1}=5\times\frac{5^4}{6^5}\cong 0,4019$$

άρα η σχέση (1) γίνεται:

$$P(X\geq 2)\cong 1-[0,4019+0,4019]=0,1962$$

4.3.2 ΚΑΤΑΝΟΜΗ POISSON

Η **κατανομή Poisson** χρησιμοποιείται για να περιγράψει φαινόμενα τα οποία μπορούν να θεωρηθούν ως διαδικασία παραγωγής τυχαίων εμφανίσεων ενδεχομένων κατά διαστήματα χρόνου ή χώρου και τα οποία ικανοποιούν τις παρακάτω συνθήκες :

i. Σε κάθε διάστημα χρόνου ή χώρου ένα ενδεχόμενο μπορεί να συμβεί ή να μη συμβεί. Η εμφάνιση του ενδεχομένου καλείται συνήθως επιτυχία ενώ η μη εμφάνιση του αποτυχία.

ii. Οι τυχαίες εμφανίσεις ενδεχομένων είναι ανεξάρτητες δηλαδή η εμφάνιση ενός ενδεχομένου σε ένα διάστημα χρόνου ή χώρου δεν επηρεάζει την πιθανότητα εμφάνισης του στο επόμενο διάστημα χρόνου ή χώρου.

iii. Η πιθανότητα εμφάνισης (ή μη εμφάνισης) ενός ενδεχομένου σε ένα διάστημα χρόνου ή χώρου παραμένει σταθερή για όλη τη διάρκεια του φαινομένου.

Η κατανομή Poisson συμβολίζεται με P(λ) και χαρακτηρίζεται από μία παράμετρο, το μέσο αριθμό επιτυχιών σε ένα διάστημα χώρου ή χρόνου λ. Αν λοιπόν θεωρήσουμε ότι η τυχαία μεταβλητή ακολουθεί την P(λ) [X~P(λ)] έχουμε τη συνάρτηση πιθανότητας:

$$P(k) = P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!},\ k = 0,\ 1,\ 2,\ \ldots$$

Όπου $k = 0, 1, 2, \ldots$ ο αριθμός των επιτυχιών που θέλουμε να εμφανιστεί.

Για την κατανομή Poisson αποδεικνύεται ότι ισχύουν:

$$E(X) = \lambda \qquad V(X) = \lambda$$

Παράδειγμα

Μια υπάλληλος η οποία εισάγει δεδομένα στον Η/Υ κάνει κατά μέσο όρο τρία λάθη ανά σελίδα. Να υπολογιστεί η πιθανότητα σε τυχαία επιλεγμένη σελίδα να βρεθούν δύο λάθη.

Λύση

Έστω Χ ο αριθμός των λαθών ανά σελίδα. Σύμφωνα με τα δεδομένα Χ~P(λ=3). . Ζητάμε την πιθανότητα P(X=2)

$$P(X=2)=\frac{e^{-3}3^2}{2!}=\frac{\frac{1}{e^3}*3^2}{1*2}=\frac{(0{,}050)*9}{2}=0{,}225$$

4.4 ΣΥΝΕΧΕΙΣ ΚΑΤΑΝΟΜΕΣ

Στην ενότητα αυτή θα μελετήσουμε μια συνεχή κατανομή η οποία παρουσιάζει ιδιαίτερο ενδιαφέρον κυρίως λόγω των πολλών εφαρμογών της

Κανονική Κατανομή

Η **κανονική κατανομή** είναι η σημαντικότερη όχι μόνο από τις συνεχείς αλλά και από όλες τις κατανομές πιθανότητας και αποτελεί τη βάση της σύγχρονης στατιστικής θεωρίας. Η σπουδαιότητα της οφείλεται σε τρεις κυρίως λόγους :

i. Πολλά πειράματα μπορούν να εκφραστούν μέσω τυχαίων μεταβλητών που ακολουθούν κανονική κατανομή.

ii. Μπορεί να χρησιμοποιηθεί ως προσέγγιση πολλών άλλων κατανομών.

iii.Αποτελεί τη βάση για πολλές τεχνικές που χρησιμοποιούνται στη στατιστική συμπερασματολογία.

Η κανονική κατανομή μελετήθηκε διεξοδικά από το μαθηματικό Κ. Gauss και για το λόγο αυτό είναι γνωστή και ως κατανομή Gauss.

Έστω Χ συνεχής τυχαία μεταβλητή που μπορεί να πάρει τιμές σε ολόκληρη την ευθεία των πραγματικών αριθμών. Λέμε ότι η Χ ακολουθεί την κατανομή με παραμέτρους μ και σ^2 (και το συμβολίζουμε Χ~Ν(μ, σ^2)) αν η συνάρτηση πυκνότητας πιθανότητας δίνεται από τον τύπο:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

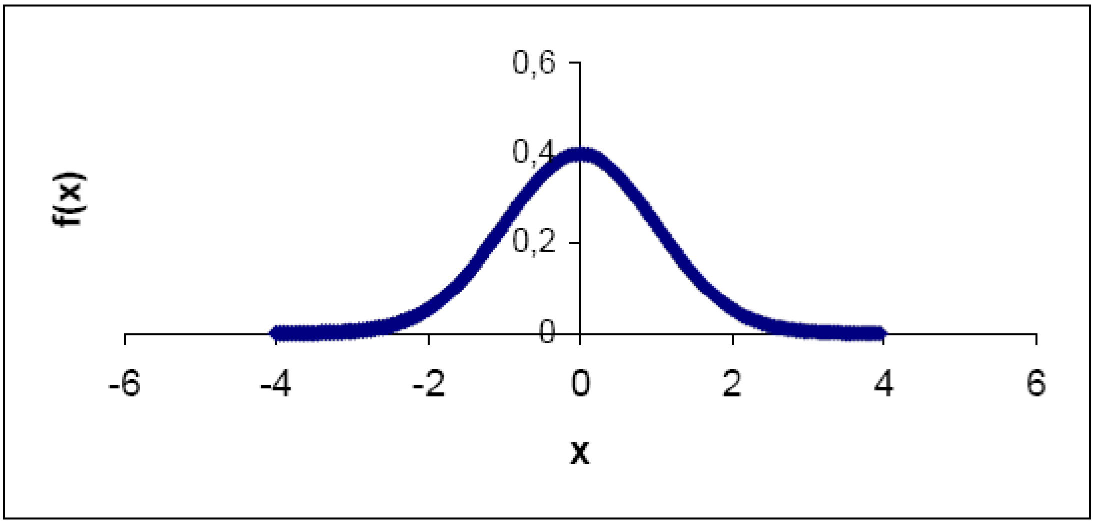
όπου π = 3,1416

e = 2,7183

Για την κανονική κατανομή αποδεικνύεται ότι ισχύουν:

E(X) = μ , V(X) = σ^2

Από τη μελέτη της συνάρτησης προκύπτει ότι η συνάρτηση πυκνότητας πιθανότητας της κανονικής κατανομής έχει μορφή παρόμοια με αυτήν που απεικονίζεται στο παρακάτω Σχήμα .



Καμπύλη συνάρτησης πυκνότητας πιθανότητας της κανονικής κατανομής

Όπως φαίνεται από το παραπάνω Σχήμα η καμπύλη της συνάρτησης πυκνότητας πιθανότητας της κανονικής κατανομής :

i. Είναι μονοκόρυφη, με κωδωνοειδές σχήμα και συμμετρική γύρω από το μ.

ii. Η μέση τιμή (μ), και η διάμεσος (Μ) και η επικρατούσα τιμή (T_0) συμπίπτουν.

iii. Είναι ασύμπτωτη ως προς τον άξονα των Χ.

Επίσης, παρατηρούμε ότι από τις δύο παραμέτρους της το μ προσδιορίζει τη θέση της κατανομής ως προς τον άξονα των Χ το δε σ^2 το σχήμα της.

Σύμφωνα με τον ορισμό της κανονικής κατανομής ο υπολογισμός μιας πιθανότητας της ισοδυναμεί με τον υπολογισμό του αντίστοιχου ολοκληρώματος της, δηλαδή

$$P(a<X<b)=\int_a^b f(x)dx$$

όπου f(x) η συνάρτηση πυκνότητας πιθανότητας .

Λόγω όμως της πολύπλοκης μορφής της ολοκληρωτέας συνάρτησης f(x), ο υπολογισμός του ολοκληρώματος είναι δύσκολος και χρονοβόρος. Ειδικά όμως για την κανονική κατανομή αποδεικνύεται ότι

Αν η τ.μ $X \sim N(\mu, \sigma^2)$, τότε η τ.μ $Z=\dfrac{X-\mu}{\sigma} \sim N(0,1)$

Προφανώς η συνάρτηση πυκνότητας πιθανότητας της Z δίνεται από τον τύπο:

$$f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}, \quad -\infty \langle\ z\ \langle +\infty$$

και αποδεικνύεται ότι ισχύουν

$$E(X) = 0 \qquad V(X) = 1.$$

Με βάση τα παραπάνω, ο υπολογισμός οποιασδήποτε πιθανότητας για μία κανονική κατανομή, ανάγεται στον υπολογισμό μιας αντίστοιχης πιθανότητας για την ειδική κανονική κατανομή N(0,1). Η κατανομή αυτή λέγεται **τυπική κανονική** (ή **τυποποιημένη κανονική**) και ο μετασχηματισμός λέγεται τυποποίηση.

Για την τυπική κανονική κατανομή υπάρχουν πίνακες που δίνουν τις τιμές των πιθανοτήτων για τις διάφορες τιμές της Z.

Παράδειγμα

Έστω ότι η εσωτερική διάμετρος των δακτυλίων μετάλλου («παξιμάδια»), που παράγει κάποιο εργοστάσιο, ακολουθεί κανονική κατανομή με μέσο 0,373 εκ. και τυπική απόκλιση 0,002 εκ. Οι δακτύλιοι αυτοί προορίζονται, μεταξύ των άλλων χρήσεών τους και για τη συναρμολόγηση ορισμένων μερών, ενός συγκεκριμένου τύπου μοτοποδηλάτου, όπου σύμφωνα με τις τεχνικές προδιαγραφές απαιτούνται δακτύλιοι εσωτερικής διαμέτρου ($0,375 \pm 0,003$) εκ. Ποια η πιθανότητα ένας τέτοιος δακτύλιος να πληρεί τις τεχνικές προδιαγραφές;

Λύση

Ζητάμε ουσιαστικά την πιθανότητα:

$P(0,375 – 0,003 < X < 0,375 + 0,003)$ ή $P(0,372 < X < 0,378)$

όπου X η τυχαία μεταβλητή που εκφράζει την εσωτερική διάμετρο του «παξιμαδιού» και για την οποία μας δίνεται ότι: $X \sim N(0,373, 0,002^2)$.

Για τον υπολογισμό της πιθανότητας αυτής θα έχουμε:

$$P(0,372 < X < 0,378) = P\left(\frac{0,372-0,373}{0,002} < \frac{X-\mu}{\sigma} < \frac{0,378-0,373}{0,002}\right)$$

$$P(-0,5 < Z < 2,5) =$$

$$P(Z < 2,5) - P(Z < -0,5) =$$

$$P(Z < 2,5) - (1 - P(Z < 0,5)) =$$

$$0,9938\text{-}1 + 0,6915 =$$

$$1,6853\text{-}1 = 0,6853$$

Παράδειγμα Δειγματοληψίας με το στατιστικό πακέτο SPSS

Η δειγματοληψία που πραγματοποιήθηκε είχε ως υποκείμενο τους εκπαιδευτικούς της Δευτεροβάθμιας Εκπαίδευσης που εργάζονται στα σχολεία της πόλης του Ηρακλείου και αφορούσε στη διερεύνηση των αντιλήψεων των εκπαιδευτικών σχετικά με τα πιθανά μέτρα που μπορούν να ληφθούν σε επιχειρηματικό επίπεδο αναφορικά με το ενεργειακό πρόβλημα και τη σωστή διαχείριση της ενέργειας. Τα ερωτηματολόγια που δόθηκαν ήταν 114.

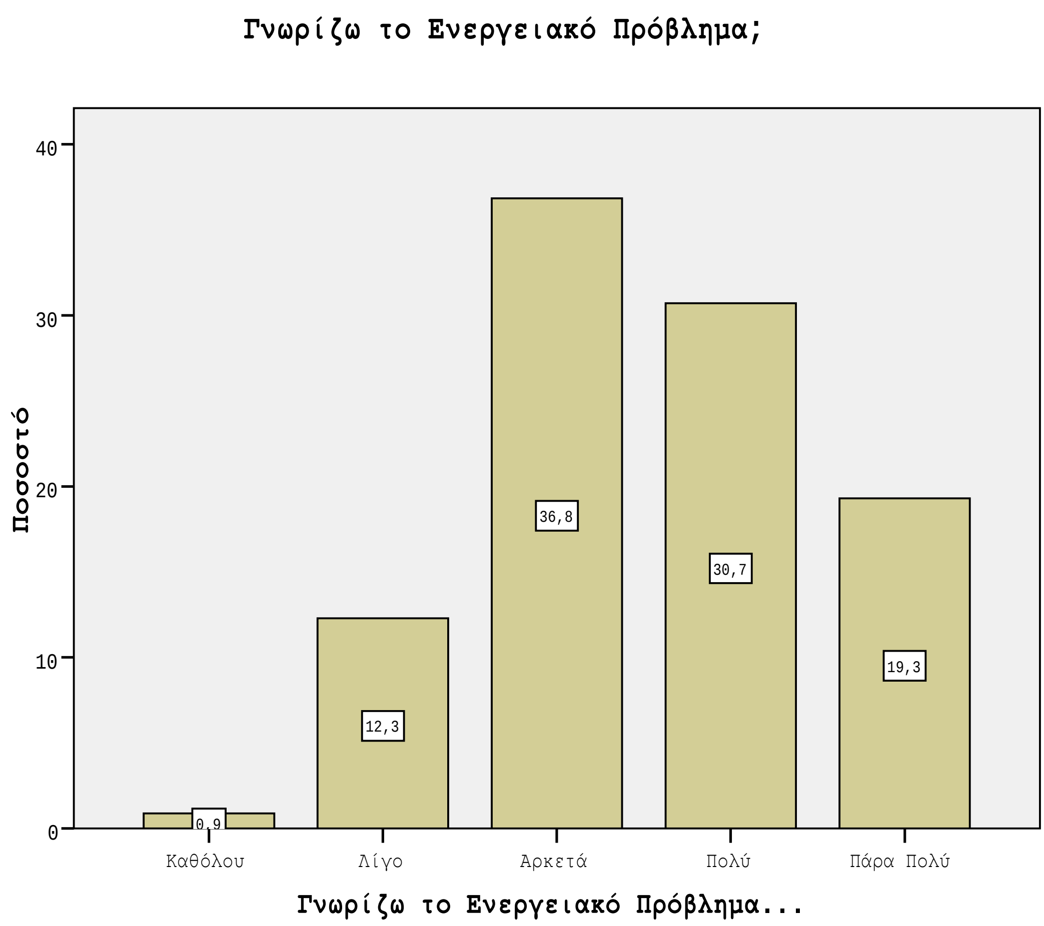
Η ειδικότητα και ο κλάδος των εκπαιδευτικών που έλαβαν μέρος στην έρευνα φαίνονται παρακάτω (Πίνακας 1).

Ειδικότητα		Συχνότητα	Σχετική Συχνότητα
ΤΕ01	Σχεδιαστές	1	0.9
ΠΕ01	Θεολόγοι	6	5.3
ΠΕ02	Φιλόλογοι	15	13.2
ΠΕ03	Μαθηματικοί	17	14.9
ΠΕ04	Φυσικοί – Χημικοί – Βιολόγοι – Γεωλόγοι	14	12.3
ΠΕ06	Αγγλικών	6	5.3
ΠΕ07	Γερμανικών	1	0.9
ΠΕ09	Οικονομολόγων	7	6.1
ΠΕ10	Κοινωνιολόγοι	2	1.8
ΠΕ11	Φυσικής Αγωγής	4	3.5
ΠΕ12	Πολιτικοί Μηχανικοί – Αρχιτεκτόνων Μηχανικών – Τοπογράφων Μηχανικών	3	2.6
ΠΕ13	Νομικών Επιστημών	3	2.6
ΠΕ14	Γεωπόνοι - Δασολόγοι	2	1.8
ΠΕ16	Μουσικοί	2	1.8
ΠΕ17	Ηλεκτρολόγοι Μηχανικοί – Ηλεκτρονικοί ΑΣΕΤΕΜ	4	3.5
ΠΕ18	Λογιστικής	14	12.3
ΠΕ19	Επιστήμης Υπολογιστών	8	7.0
ΠΕ20	Επιστήμης Υπολογιστών Τ.Ε.Ι	2	1.8
	Σύνολο	111	97.4
	Κενά	3	2.6
Τελικό Σύνολο		114	100.0

Πίνακας 1 - Η ειδικότητα των εκπαιδευτικών που έλαβαν μέρος στην έρευνα

Το 14.9% των συμμετεχόντων είναι μαθηματικοί, το 13.2% Φιλόλογοι και το 12.3% Φυσικοί – Χημικοί – Βιολόγοι – Γεωλόγοι όπως και Λογιστές. Όπως παρατηρούμε, αυτές είναι οι κυρίαρχες – εάν εξεταστούν πληθυσμιακά – ειδικότητες στο δείγμα μας.

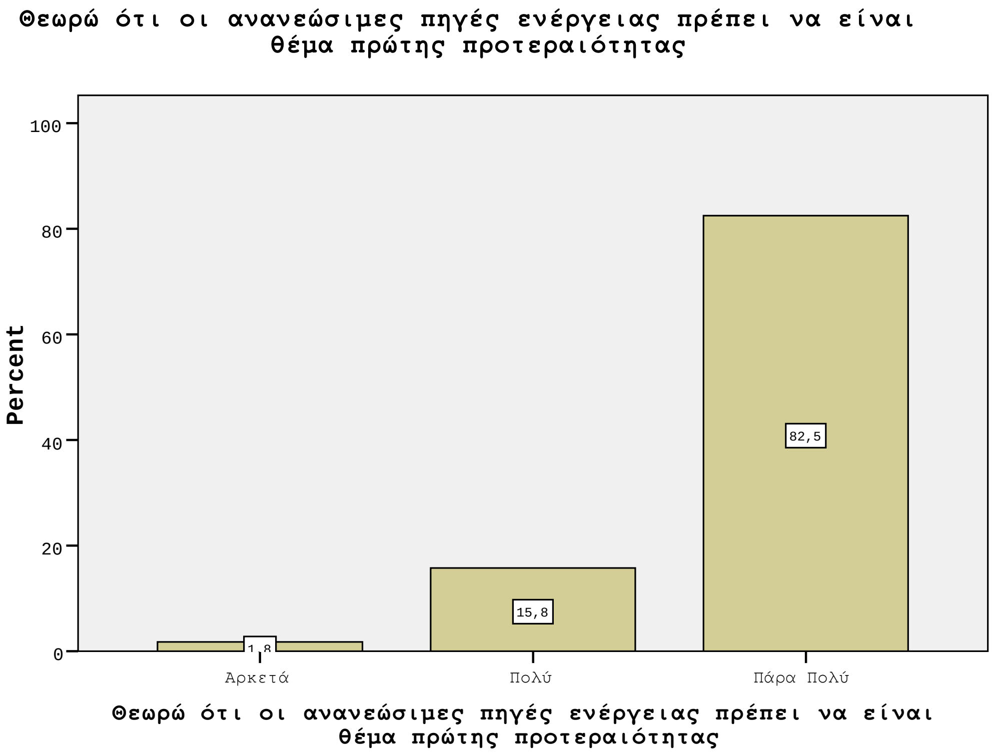
Σχετικά με το βαθμό στον οποίο πιστεύουν οι εκπαιδευτικοί ότι είναι ενημερωμένοι για την ρύπανση του περιβάλλοντος, τα αποτελέσματα έδωσαν το μεγαλύτερο ποσοστό των εκπαιδευτικών να πιστεύει ότι έχει αρκετή γνώση των ενεργειακών προβλημάτων. Το 36.8% των εκπαιδευτικών χαρακτηρίζει τον εαυτό του «αρκετά» ενημερωμένο, το 30.7% χαρακτηρίζει τον εαυτό του «πολύ» ενημερωμένο, το 19.3% θεωρεί ότι γνωρίζει «πάρα πολύ» τα ενεργειακά ζητήματα ενώ το 12.3% χαρακτηρίζει τον εαυτό του «λίγο» ενημερωμένο και το 0.9% θεωρεί ότι δεν ξέρει τίποτα για τα ενεργειακά ζητήματα (Γράφημα 1).



Γράφημα 1 – Ο βαθμό στον οποίο πιστεύουν οι εκπαιδευτικοί ότι είναι ενημερωμένοι για την ρύπανση του περιβάλλοντος

Στην ερώτηση για τον αν οι ανανεώσιμες πηγές ενέργειας (ΑΠΕ) θα πρέπει να είναι θέμα πρώτης προτεραιότητας παρατηρήθηκε ότι ένα 82.5% των συμμετεχόντων δηλώνει ότι οι ΑΠΕ θα πρέπει («πάρα πολύ») να είναι θέμα πρώτης προτεραιότητας. Κανείς δεν δήλωσε ότι δεν θεωρεί «καθόλου» είτε ότι θεωρεί «λίγο» ότι οι ΑΠΕ θα πρέπει να είναι ζήτημα πρώτης προτεραιότητας. Από αυτό συμπεραίνουμε ότι οι συμμετέχοντες αξιολογούν ιδιαίτερα τις ΑΠΕ

(Γράφημα 2, Πίνακας 2)

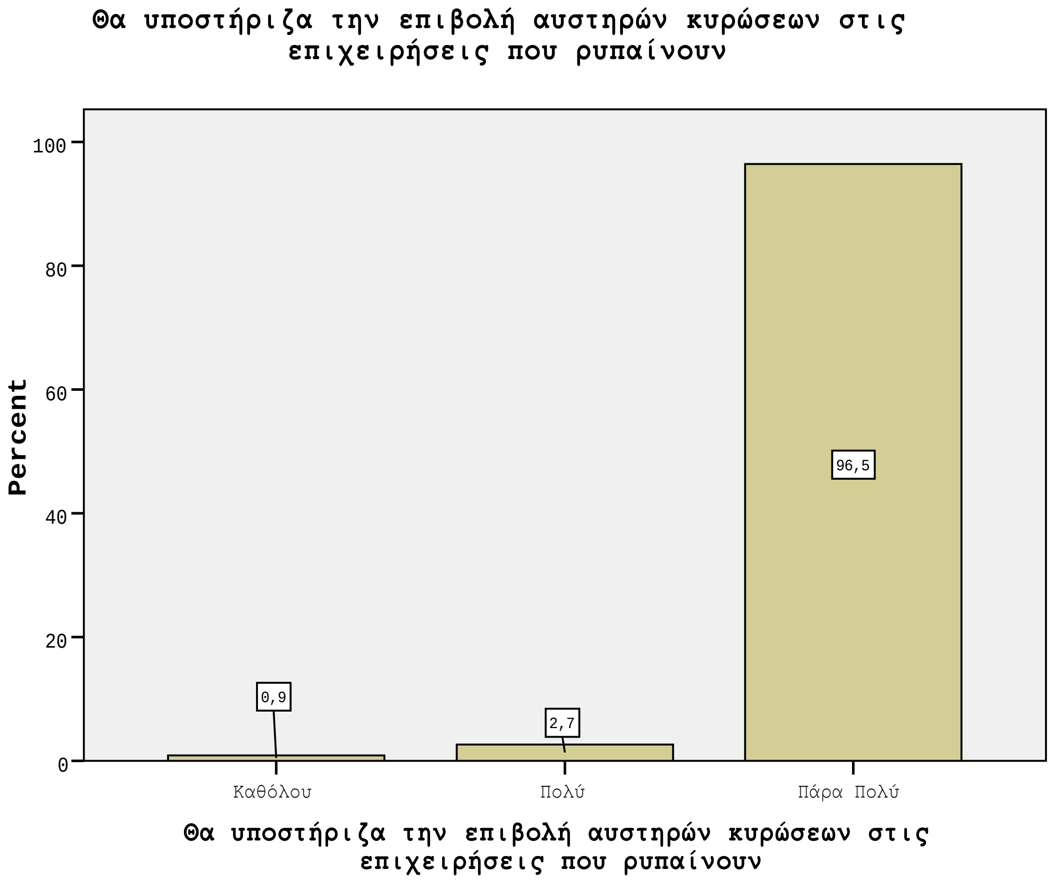


Γράφημα 2 – Θεωρώ ότι οι ανανεώσιμες πηγές ενέργειας πρέπει να είναι θέμα πρώτης προτεραιότητας

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Αρκετά	2	1,8	1,8	1,8
	Πολύ	18	15,8	15,8	17,5
	Πάρα Πολύ	94	82,5	82,5	100,0
	Total	114	100,0	100,0	

Πίνακας 2 - Θεωρώ ότι οι ανανεώσιμες πηγές ενέργειας πρέπει να είναι θέμα πρώτης προτεραιότητας

Στην ερώτηση για το εάν θα υποστήριζαν την επιβολή αυστηρών κυρώσεων στις επιχειρήσεις που ρυπαίνουν, παρατηρούμε ότι το 96.5% συμφωνεί «πάρα πολύ» με την άποψη αυτή ενώ ένας μόνο εκπαιδευτικός (ποσοστό 0.9%) δηλώνει ότι δεν συμφωνεί «καθόλου» με την επιβολή αυστηρότερων κυρώσεων. Αυτό υποδηλώνει μια σχεδόν καθολική συμφωνία με την θέση της ερώτησης (Πίνακας 3, Γράφημα 3).

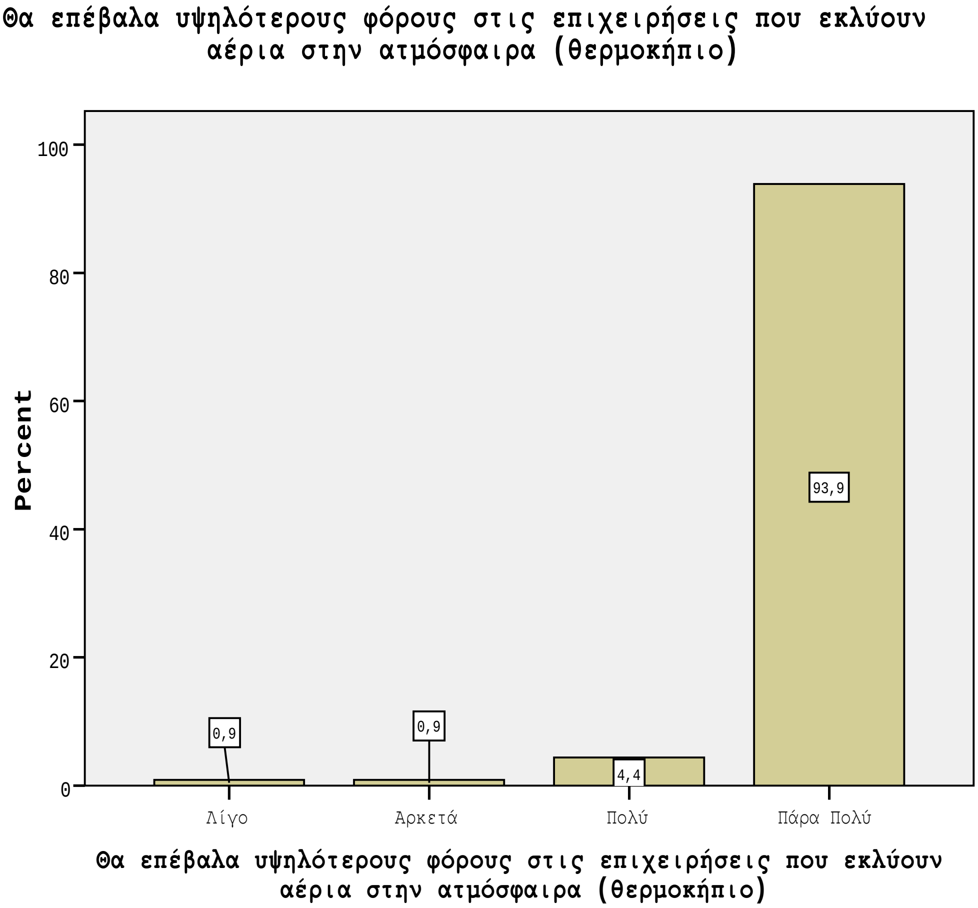


Γράφημα 3- Θα υποστήριζα την επιβολή αυστηρών κυρώσεων στις επιχειρήσεις που ρυπαίνουν

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Καθόλου	1	,9	,9	,9
	Πολύ	3	2,6	2,7	3,5
	Πάρα Πολύ	109	95,6	96,5	100,0
	Total	113	99,1	100,0	
Missing	System	1	,9		
Total		114	100,0		

Πίνακας 3- Θα υποστήριζα την επιβολή αυστηρών κυρώσεων στις επιχειρήσεις που ρυπαίνουν

Αναφορικά με την άποψη της επιβολής υψηλότερων φόρων στις επιχειρήσεις που εκλύουν αέρια στην ατμόσφαιρα, παρατηρούμε ότι το 98.2% των ερωτηθέντων απαντά ότι συμφωνεί από «πολύ» έως «πάρα πολύ», ενώ δύο μόλις εκπαιδευτικοί (ποσοστό 1.8%) δείχνουν είτε μικρή («αρκετά») είτε μικρότερη («λίγο») διαφωνία με τη θέση που τίθεται προς εξέταση (*Πίνακας 4*, *Γράφημα 4*). Αυτό το αποτέλεσμα είναι συμβατό με το γεγονός ότι το 99.1% των εκπαιδευτικών (*Πίνακας 3*) συμφωνούν με την επιβολή αυστηρών κυρώσεων σε επιχειρήσεις που ρυπαίνουν.



Γράφημα 4 - Θα επέβαλα υψηλότερους φόρους στις επιχειρήσεις που εκλύουν αέρια στην ατμόσφαιρα (θερμοκήπιο)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Λίγο	1	,9	,9	,9
	Αρκετά	1	,9	,9	1,8
	Πολύ	5	4,4	4,4	6,1
	Πάρα Πολύ	107	93,9	93,9	100,0
	Total	114	100,0	100,0	

Πίνακας 4 - Θα επέβαλα υψηλότερους φόρους στις επιχειρήσεις που εκλύουν αέρια στην ατμόσφαιρα (θερμοκήπιο)

4.5 ΚΑΤΑΝΟΜΕΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

Κάθε στατιστική, ως τυχαία μεταβλητή, ακολουθεί μια κατανομή πιθανότητας, η οποία καλείται κατανομή δειγματοληψίας (sampling distribution). Οι κατανομές δειγματοληψίας των στατιστικών χαρακτηρίζονται από βαθμούς ελευθερίας (degrees of freedom). Αν μια στατιστική δεν περιέχει παραμέτρους της κατανομής του πληθυσμού, ή περιέχει παραμέτρους, αλλά οι τιμές τους είναι γνωστές, τότε η κατανομή δειγματοληψίας της στατιστικής αυτής έχει πλήρεις βαθμούς ελευθερίας, οι οποίοι ισούνται με το μέγεθος των λαμβανομένων δειγμάτων, έστω n. Αν όμως μια στατιστική στη θέση κάποιων παραμέτρων της κατανομής του πληθυσμού, των οποίων οι τιμές δεν είναι γνωστές, περιέχει άλλες στατιστικές, οι οποίες εκτιμούν τις άγνωστες αυτές παραμέτρους, τότε η κατανομή δειγματοληψίας της στατιστικής αυτής χάνει τόσους βαθμούς ελευθερίας, όσες και οι άγνωστοι παράμετροι.

Αν ο πληθυσμός X ακολουθεί οποιαδήποτε κατανομή με E(X)= μ και V(X)=σ 2 και { } 1 2 , , ,n X X KX είναι τυχαίο δείγμα μεγέθους n από τον πληθυσμό αυτό, τότε η κατανομή δειγματοληψίας του δειγματικού μέσου X έχει:

$$E\left(\bar{X}\right) = \mu \text{ και } V\left(\bar{X}\right) = \frac{\sigma^2}{n}$$

Για τις συνιστώσες του τυχαίου δείγματος ισχύουν:

$E\left(X_i\right) = \mu$, $V\left(X_i\right) = \sigma^2$ με X_i, $i = 1, 2, \ldots, n$ ανεξάρτητες μεταξύ τους, οπότε

$$E\left(\bar{X}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right) = \frac{1}{n}\cdot n\mu = \mu \quad \text{και}$$

$$V\left(\bar{X}\right) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V\left(X_i\right) + 0 = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}$$

όπου το μηδέν στον αμέσως προηγούμενο τύπο αντιστοιχεί στο άθροισμα των συνδιακυμάνσεων των συνιστωσών του δείγματος ανά δύο, οι οποίες όμως είναι μηδενικές, λόγω της ανεξαρτησίας τους.

Αν πληθυσμός X ακολουθεί την κανονική κατανομή $N\left(\mu, \sigma^2\right)$ και $\left\{X_1, X_2, \ldots, X_n\right\}$ είναι τυχαίο δείγμα μεγέθους n από τον πληθυσμό αυτό, τότε ο δειγματικός μέσος $\bar{X}$ ακολουθεί την κανονική κατανομή $N\left(\mu, \frac{\sigma^2}{n}\right)$.

Έστω ότι ο πληθυσμός X ακολουθεί οποιαδήποτε κατανομή με $E(X)=\mu_1$ και $V(X)=\sigma_1^2$ και $\{X_1, X_2, \ldots, X_{n_1}\}$ είναι τυχαίο δείγμα μεγέθους n_1 από τον πληθυσμό αυτό. Έστω ότι ο πληθυσμός Y, ανεξάρτητος του πληθυσμού X, ακολουθεί οποιαδήποτε κατανομή με $E(Y)=\mu_2$ και $V(Y)=\sigma_2^2$ και $\{Y_1, Y_2, \ldots, Y_{n_2}\}$ είναι τυχαίο δείγμα μεγέθους n_2 από τον πληθυσμό αυτό. Αν $\bar{X}$ και $\bar{Y}$ είναι οι αντίστοιχοι δειγματικοί μέσοι, τότε η κατανομή δειγματοληψίας της διαφοράς $\bar{X}-\bar{Y}$ των δειγματικών μέσων έχει:

$$E\left(\bar{X}-\bar{Y}\right)=\mu_1-\mu_2 \text{ και } V\left(\bar{X}-\bar{Y}\right)=\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$$

Έστω ότι ο πληθυσμός X ακολουθεί την κανονική κατανομή $N\left(\mu_1, \sigma_1^2\right)$ και $\{X_1, X_2, \ldots, X_{n_1}\}$ είναι τυχαίο δείγμα μεγέθους n_1 από τον πληθυσμό αυτό. Έστω ότι ο πληθυσμός Y, ανεξάρτητος του πληθυσμού X, ακολουθεί την κανονική κατανομή $N\left(\mu_2, \sigma_2^2\right)$ και $\{Y_1, Y_2, \ldots, Y_{n_2}\}$ είναι τυχαίο δείγμα μεγέθους n_2 από τον πληθυσμό αυτό. Αν $\bar{X}$ και $\bar{Y}$ είναι οι αντίστοιχοι δειγματικοί μέσοι,

Έστω ότι ο πληθυσμός X ακολουθεί οποιαδήποτε κατανομή με $E(X)=\mu_1$ και $V(X)=\sigma_1^2$ και $\{X_1, X_2, \ldots, X_{n_1}\}$ είναι τυχαίο δείγμα μεγέθους n_1 από τον πληθυσμό αυτό. Έστω ότι ο πληθυσμός Y, ανεξάρτητος του πληθυσμού X, ακολουθεί οποιαδήποτε κατανομή με $E(Y)=\mu_2$ και $V(Y)=\sigma_2^2$ και $\{Y_1, Y_2, \ldots, Y_{n_2}\}$ είναι τυχαίο δείγμα μεγέθους n_2 από τον πληθυσμό αυτό. Αν $\bar{X}$ και $\bar{Y}$ είναι οι αντίστοιχοι δειγματικοί μέσοι, τότε η κατανομή δειγματοληψίας της διαφοράς $\bar{X}-\bar{Y}$ των δειγματικών μέσων ακολουθεί ασυμπτωτικά (δηλαδή για μεγάλα δείγματα, πρακτικά $n_1, n_2 > 30$) την κανονική κατανομή $N\left(\mu_1-\mu_2, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)$.

Αν πληθυσμός X ακολουθεί την κανονική κατανομή $N\left(\mu,\sigma^2\right)$ και $\{X_1, X_2, \ldots, X_n\}$ είναι τυχαίο δείγμα μεγέθους n από τον πληθυσμό αυτό, τότε η στατιστική $\dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}$ ακολουθεί την κατανομή χ_n^2.

Για τις συνιστώσες του τυχαίου δείγματος ισχύουν:

$$X_i \sim N\left(\mu,\sigma^2\right) \Rightarrow \frac{X_i-\mu}{\sigma} \sim N(0,1) \Rightarrow \left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi_1^2 \;, i=1,2,\ldots,n$$

$$\text{οπότε: } \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi_n^2$$

Αν πληθυσμός X ακολουθεί την κανονική κατανομή $N\left(\mu,\sigma^2\right)$ και $\{X_1, X_2, \ldots, X_n\}$ είναι τυχαίο δείγμα μεγέθους n από τον πληθυσμό αυτό, $\bar{X}$ ο δειγματικός μέσος και S^2 η δειγματική διακύμανση, τότε η στατιστική $\dfrac{(n-1)S^2}{\sigma^2} = \dfrac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}$ ακολουθεί την κατανομή χ_{n-1}^2.

Αν πληθυσμός X ακολουθεί την κανονική κατανομή $N\left(\mu,\sigma^2\right)$ και $\{X_1, X_2, \ldots, X_n\}$ είναι τυχαίο δείγμα μεγέθους n από τον πληθυσμό αυτό, $\bar{X}$ ο δειγματικός μέσος και S^2 η δειγματική διακύμανση, τότε η στατιστική $\dfrac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$ ακολουθεί την κατανομή t_{n-1}.

Είναι $\dfrac{\overline{X}-\mu}{\dfrac{\sigma}{\sqrt{n}}} \sim N(0,1)$ και $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$, οπότε σύμφωνα με τον ορισμό της κατανομής t_ν θα είναι

$$\frac{\dfrac{\overline{X}-\mu}{\dfrac{\sigma}{\sqrt{n}}}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2} \Big/ n-1}} = \frac{\overline{X}-\mu}{\dfrac{S}{\sqrt{n}}} \sim t_{n-1}.$$

Έστω ότι ο πληθυσμός X ακολουθεί την κανονική κατανομή $N\left(\mu_1,\sigma^2\right)$ και $\left\{X_1, X_2, \dots, X_{n_1}\right\}$ είναι τυχαίο δείγμα μεγέθους n_1 από τον πληθυσμό αυτό. Έστω ότι ο πληθυσμός Y, ανεξάρτητος του πληθυσμού X, ακολουθεί την κανονική κατανομή $N\left(\mu_2,\sigma^2\right)$ (οι δύο πληθυσμοί έχουν κοινή διακύμανση) και $\left\{Y_1, Y_2, \dots, Y_{n_2}\right\}$ είναι τυχαίο δείγμα μεγέθους n_2 από τον πληθυσμό αυτό. Αν $\overline{X}$, $\overline{Y}$ είναι οι αντίστοιχοι δειγματικοί μέσοι και S_1^2, S_2^2 είναι οι αντίστοιχες δειγματικές διακυμάνσεις, τότε η στατιστική

$$\frac{\overline{X}-\overline{Y}-\left(\mu_1-\mu_2\right)}{\sqrt{\dfrac{\left(n_1-1\right)S_1^2+\left(n_2-1\right)S_2^2}{n_1+n_2-2}\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$$

ακολουθεί την κατανομή $t_{n_1+n_2-2}$.

Εστω ότι ο πληθυσμός X ακολουθεί την κανονική κατανομή $N\left(\mu_1, \sigma_1^2\right)$ και $\left\{X_1, X_2, \ldots, X_{n_1}\right\}$ είναι τυχαίο δείγμα μεγέθους n_1 από τον πληθυσμό αυτό. Εστω ότι ο πληθυσμός Y, ανεξάρτητος του πληθυσμού X, ακολουθεί την κανονική κατανομή $N\left(\mu_2, \sigma_2^2\right)$ και $\left\{Y_1, Y_2, \ldots, Y_{n_2}\right\}$ είναι τυχαίο δείγμα μεγέθους n_2 από τον πληθυσμό αυτό. Κάτω από τις υποθέσεις αυτές η στατιστική

$$\frac{\sum_{i=1}^{n_1}\left(X_i-\mu_1\right)^2 \Big/ \sigma_1^2}{\sum_{i=1}^{n_2}\left(Y_i-\mu_2\right)^2 \Big/ \sigma_2^2}$$ ακολουθεί την κατανομή F_{n_1, n_2}.

Είναι $\dfrac{\sum_{i=1}^{n_1}\left(X_i-\mu_1\right)^2}{\sigma_1^2} \sim \chi_{n_1}^2$ και $\dfrac{\sum_{i=1}^{n_2}\left(Y_i-\mu_2\right)^2}{\sigma_2^2} \sim \chi_{n_2}^2$, οπότε ο λόγος τους θα ακολουθεί την κατανομή F_{n_1, n_2}.

Έστω ότι ο πληθυσμός X ακολουθεί την κανονική κατανομή $N\left(\mu_1, \sigma_1^2\right)$ και $\left\{X_1, X_2, \ldots, X_{n_1}\right\}$ είναι τυχαίο δείγμα μεγέθους n_1 από τον πληθυσμό αυτό. Έστω ότι ο πληθυσμός Y, ανεξάρτητος του πληθυσμού X, ακολουθεί την κανονική κατανομή $N\left(\mu_2, \sigma_2^2\right)$ και $\left\{Y_1, Y_2, \ldots, Y_{n_2}\right\}$ είναι τυχαίο δείγμα μεγέθους n_2 από τον πληθυσμό αυτό. Αν $\bar{X}$, $\bar{Y}$ είναι οι αντίστοιχοι δειγματικοί μέσοι και S_1^2, S_2^2 είναι οι αντίστοιχες δειγματικές διακυμάνσεις, τότε η στατιστική $\dfrac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2}$ ακολουθεί την κατανομή F_{n_1-1, n_2-1}.

Είναι $\dfrac{S_1^2}{\sigma_1^2} \sim \chi_{n_1-1}^2$ και $\dfrac{S_2^2}{\sigma_2^2} \sim \chi_{n_2-1}^2$, οπότε ο λόγος τους θα ακολουθεί την κατανομή F_{n_1-1, n_2-1}.

Παράδειγμα

Από πληθυσμό που ακολουθεί κατανομή με σ.π.

$$p(x) = \begin{cases} \dfrac{1}{5} &, \quad x = 0,1,2,3,4 \\ 0 &, \quad \alpha\lambda\lambda o\acute{\upsilon} \end{cases}$$

παίρνουμε τυχαία δείγματα μεγέθους $n = 50$. Να βρεθούν οι πιθανότητες:

$P(\overline{X} > 2)$ και $P(\overline{X} < 2,4)$.

Δίνεται $G(2) = 0,9772$.

Λύση:

Για τον πληθυσμό βρίσκουμε:

$$\mu = E(X) = \sum_x x p(x) = \frac{1}{5}\left(0+1+2+3+4\right) = 2 \text{ και}$$

$$E\left(X^2\right) = \sum_x x^2 p(x) = \frac{1}{5}\left(0^2+1^2+2^2+3^2+4^2\right) = 6 \text{, άρα}$$

$$\sigma^2 = V(X) = E\left(X^2\right) - \mu^2 = 2 .$$

Επειδή $n = 50 > 30$, θα είναι $\overline{X} \underset{\alpha}{\sim} N\left(2, \frac{2}{50} = \frac{1}{25}\right)$, οπότε

$$P(X > 2) \approx P\left(Z > \frac{2-2}{\frac{1}{5}}\right) = P(Z > 0) = 0,5 \text{ και}$$

$$P(X < 2,4) \approx P\left(Z < \frac{2,4-2}{\frac{1}{5}}\right) = G(2) = 0,9772$$

Από δύο ανεξάρτητους πληθυσμούς $X \sim N(3,14)$ και $Y \sim N(4,12)$ παίρνουμε τυχαία δείγματα μεγέθους $n_1 = 5$ και $n_2 = 10$ αντίστοιχα. Να υπολογισθούν οι πιθανότητες $P\left(\bar{X} > \bar{Y}\right)$ και $P\left(\left|\bar{X} - \bar{Y}\right| < 1\right)$.

Δίνεται ότι: $G(0,5) = 0,6915$ και $G(1) = 0,8413$.

Λύση:

Είναι $\bar{X} \sim N\left(3, \frac{14}{5}\right)$ και $\bar{Y} \sim N\left(4, \frac{12}{10}\right)$, οπότε $\bar{X} - \bar{Y} \sim N\left(3-4, \frac{14}{5} + \frac{12}{10}\right)$, δηλαδή $\bar{X} - \bar{Y} \sim N(-1,4)$. Άρα

$$P\left(\bar{X} > \bar{Y}\right) = P\left(\bar{X} - \bar{Y} > 0\right) = P\left(Z > \frac{0+1}{2}\right) = 1 - G(0,5) = 1 - 0,6915 = 0,3085$$ και

$$P\left(\left|\bar{X} - \bar{Y}\right| < 1\right) = P\left(-1 < \bar{X} - \bar{Y} < 1\right) = P\left(\frac{-1+1}{2} < Z < \frac{1+1}{2}\right) = G(1) - G(0) = 0,8413 - 0,5$$
$$= 0,341$$

Η κατανομή χ^2

Έστω $Z \sim N(0,1)$.

Τότε λέγεται ότι η τυχαία μεταβλητή $X = Z^2$ ακολουθεί την χ_1^2-**κατανομή (χι-τετράγωνο με ένα βαθμό ελευθερίας)**.

Η σππ της X είναι

$$f(x) = \begin{cases} \frac{1}{2\pi} x^{-1/2} e^{-x/2}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

Ιδιότητες

1. Αν $X_1, X_2, ..., X_\nu$, ν το πλήθος ανεξάρτητες μεταξύ τους τυχαίες μεταβλητές που ακολουθούν κατανομή χ_1^2, τότε λέγεται ότι η τυχαία μεταβλητή $X = \sum_{i=1}^{\nu} X_i$ ακολουθεί την **χ_ν^2-κατανομή (χι-τετράγωνο με ν βαθμούς ελευθερίας)**.

2. $E(X) = \nu$.

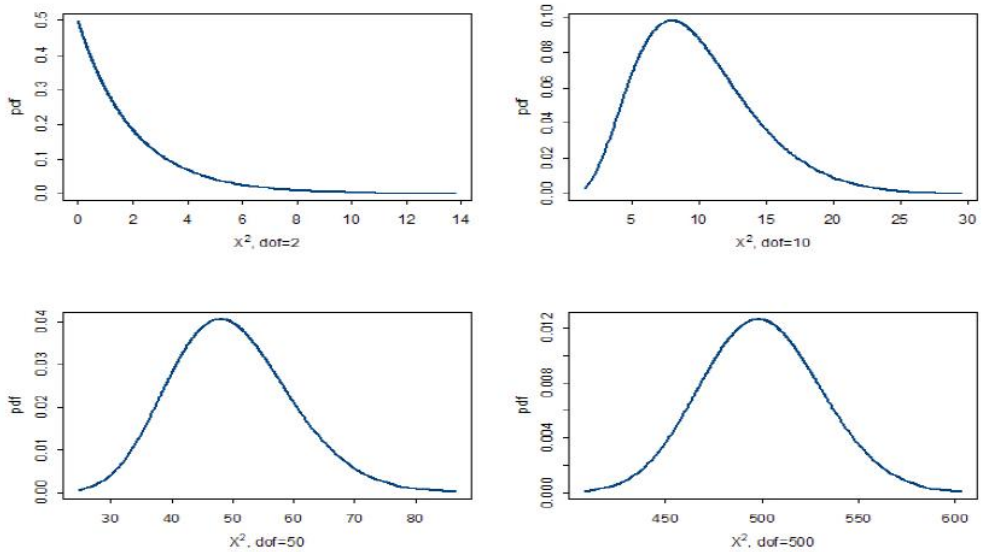
3. $\text{var}(X) = 2\nu$.

4. Για $\nu \to \infty$, η χ_ν^2 τείνει στην κανονική κατανομή.

5. Για μεγάλα ν μία καλή προσέγγιση στην χ_ν^2 προσφέρει η $Z \sim N(0,1)$, μέσω της σχέσης

$$z_p = \sqrt{2\chi_{\nu,p}^2} - \sqrt{2\nu - 1},$$

όπου $\chi_{\nu,p}^2$ και z_p τα αντίστοιχα p-ποσοστιαία σημεία.



Διάφορες κατανομές χ-τετράγωνο. Για πολλούς βαθμούς ελευθερίας πλησιάζουμε στην κανονική κατανομή.

Η κατανομή t

- Έστω $Z \sim N(0,1)$ και $V \sim \chi^2_\nu$.
- Σχηματίζουμε την τυχαία μεταβλητή T,

$$T = \frac{Z}{\sqrt{V/\nu}},$$

όπου οι Z και V είναι ανεξάρτητες.

- Η κατανομή της T λέγεται **t-κατανομή (ή κατανομή Student) με ν βαθμούς ελευθερίας** και συμβολίζεται με $\mathbf{t}_\nu$.

- Η συνάρτηση πυκνότητας πιθανότητας της T είναι

$$f(t)=\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi\nu}}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}},\ -\infty<t<\infty.$$

  - Η συνάρτηση $\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}dx$, $\alpha>0$ είναι γνωστή ως **Γ-συνάρτηση**.

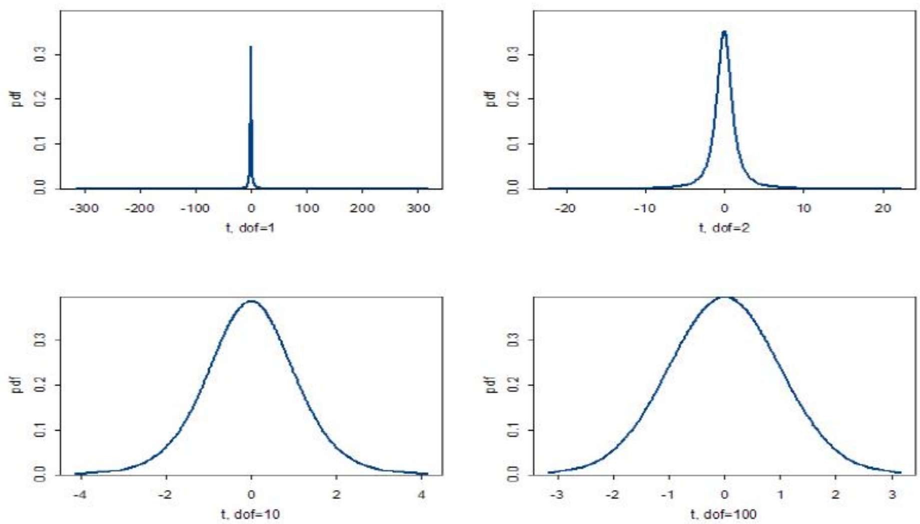
  - Για την Γ-συνάρτηση ισχύουν:
    1. $\Gamma(\alpha)=(\alpha-1)\Gamma(\alpha-1)$.
    2. $\Gamma(1/2)=\sqrt{\pi}$.
    3. $\Gamma(\alpha)=(\alpha-1)!$, όταν α θετικός ακέραιος.

- Για την t_ν κατανομή ισχύουν τα ακόλουθα:

  1. Είναι συμμετρική γύρω από το μηδέν.
  2. $E(T)=0$.
  3. $\mathrm{var}(T)=\nu/(\nu-2),\ \nu>2$.
  4. Για $\nu\to\infty,\ T\to Z\sim N(0,1)$.

  Πρακτικά, πολλοί δέχονται πως για $\nu>30$, τα σημεία $t_{\nu,p}$ και z_p είναι ίδια.

  5. Για $\nu=1$ η t_1 λέγεται **κατανομή Cauchy**, η οποία δεν έχει μέσο και διακύμανση.



σπп κατανομών t. Για μεγάλο n πλησιάζουμε την σ.π.π. της τυπικής κανονικής κατανομής.

Η κατανομή F

- Έστω τυχαίες μεταβλητές $X_1 \sim \chi^2_{\nu_1}$ και $X_2 \sim \chi^2_{\nu_2}$ ανεξάρτητες μεταξύ τους.
- Σχηματίζουμε την τυχαία μεταβλητή $X = \dfrac{X_1/\nu_1}{X_2/\nu_2}$.
- Η X λέγεται ότι ακολουθεί κατανομή F με ν_1 και ν_2, βαθμούς ελευθερίας, (βε), την οποία συμβολίζουμε με F_{ν_1,ν_2} ή $F(\nu_1,\nu_2)$.
- ν_1 είναι οι βε του αριθμητή και ν_2 του παρονομαστή. Μ'αυτή τη σειρά εμφανίζονται στο $F(\nu_1,\nu_2)$.

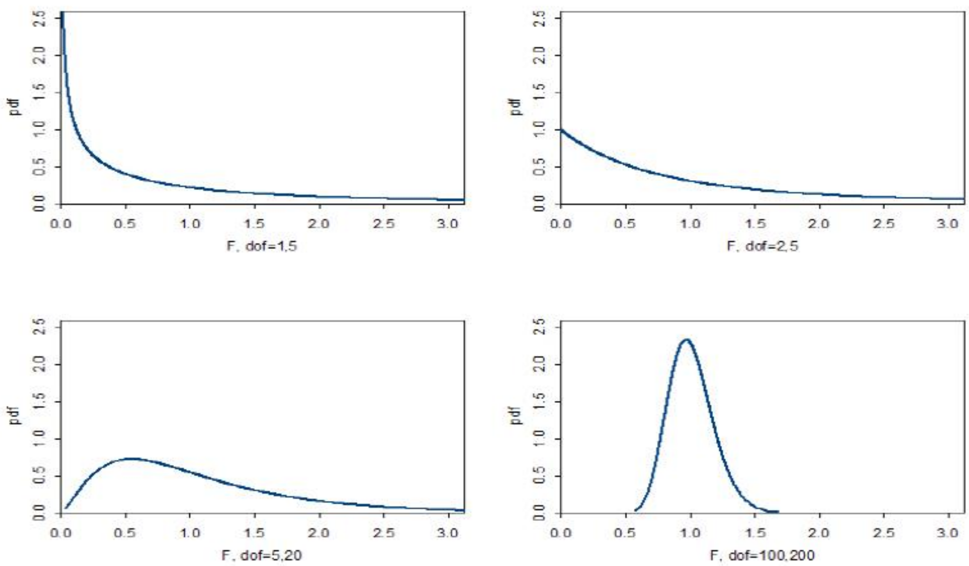
- Η αντίστοιχη σππ δίνεται από τη σχέση

$$f(x) = \frac{\Gamma\left(\left(\nu_1+\nu_2\right)/2\right)\nu_1^{\nu_1/2}\nu_2^{\nu_2/2}}{\Gamma\left(\nu_1/2\right)\Gamma\left(\nu_2/2\right)}\frac{x^{\frac{\nu_1}{2}-1}}{\left(\nu_2+x\nu_1\right)^{\frac{\nu_1+\nu_2}{2}}},$$

όπου $0 < x < \infty$.

- Για την κατανομή F ισχύουν τα εξής:

1. Είναι ασύμμετρη προς τα δεξιά.
2. Οι βε αποτελούν διατεταγμένο ζεύγος.
3. $E\left(F_{\nu_1,\nu_2}\right)=\dfrac{\nu_2}{\nu_2-2},\ \nu_2>2.$
4. $\operatorname{var}\left(F_{\nu_1,\nu_2}\right)=\dfrac{2\nu_2^2\left(\nu_1+\nu_2-2\right)}{\nu_1\left(\nu_2-2\right)^2\left(\nu_2-4\right)},\ \nu_2>4.$
5. Για ν_1 πεπερασμένο και $\nu_2\to\infty$, $F_{\nu_1,\nu_2}\sim\chi^2\left(\nu_1\right)^1.$
6. $F_{\alpha;\nu_1,\nu_2}=\dfrac{1}{F_{1-\alpha;\nu_2,\nu_1}}.$



Διάφορες F κατανομές

4.6 ΚΑΤΑΝΟΜΗ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ ΤΟΥ ΜΕΣΟΥ

Έστω χ ,χ_, ,χ ένα τυχαίο δείγμα ν. στοιχείων από έναν πληθυσμό με μέσο μ και πεπερασμένη διακύμανση. Τότε, όπως είδαμε στο προηγούμενο κεφάλαιο,

$$\bar{X} = \sum_{i=1}^{n} X_i / n,$$

κατανομή δειγματοληψίας του μέσου, έχει μέση τιμή

$$E(\bar{X}) = \mu$$

$$Var(\bar{X}) = \sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$$

και διακύμανση

Η τυπική απόκλιση σ^. της X λέγεται και τυπικό σφάλμα του δειγματικού μέσου.

Όταν η κατανομή πληθυσμού, δηλαδή η κατανομή πιθανοτήτων των τυχαίων μεταβλητών X(,X2,...,Xπ) είναι κανονική, τότε αποδεικνύεται ότι η κατανομή του δειγματικού μέσου είναι επίσης κανονική.

Όταν η κατανομή πληθυσμού δεν είναι κανονική τότε, σύμφωνα με το κεντρικό οριακό θεώρημα, η κατανομή πιθανοτήτων της τυχαίας μεταβλητής προσεγγίζει την τυπική κανονική όσο αυξάνει το η. Και όπως έχουμε δει, η προσέγγιση αυτή γίνεται σχετικά γρήγορα. Επομένως όταν γνωρίζουμε τη μέση τιμή και τη διακύμανση του πληθυσμού η κατανομή του πληθυσμού είναι κανονική ή το μέγεθος του δείγματος n είναι μεγάλο, τότε μπορούμε να υπολογίσουμε την πιθανότητα ο δειγματικός μέσος Κ να βρίσκεται σε ένα διάστημα τιμών.

$$P(\alpha \leq \bar{X} \leq b) = P\left(\frac{\alpha-\mu}{\sigma/\sqrt{n}} \leq Z \leq \frac{b-\mu}{\sigma/\sqrt{n}}\right)$$

$$= F_Z\left(\frac{b-\mu}{\sigma/\sqrt{n}}\right) - F_Z\left(\frac{\alpha-\mu}{\sigma/\sqrt{n}}\right)$$

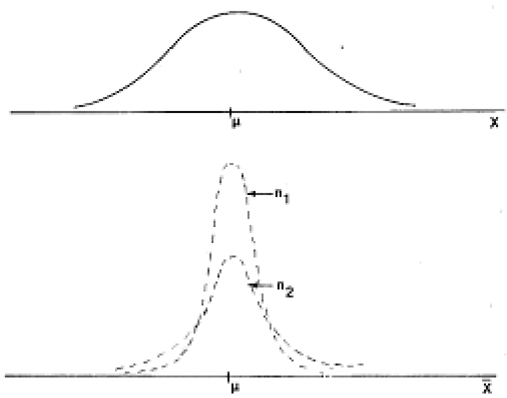
όπου F„ (z) είναι η αθροιστική πιθανότητα της τυπικής κανονικής κατανομής.

4.6.1 ΠΑΡΑΤΗΡΗΣΕΙΣ ΣΤΗΝ ΚΑΤΑΝΟΜΗ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ ΤΟΥ ΜΕΣΟΥ

Είδαμε ότι το τυπικό σφάλμα της κατανομής δειγματοληψίας του X ισούται με

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Επομένως, για δεδομένη τυπική απόκλιση του πληθυσμού, το τυπικό σφάλμα του δειγματικού μέσου μειώνεται καθώς αυξάνει το μέγεθος του δείγματος. Ειδικότερα, με τον τετραπλασιασμό του μεγέθους του δείγματος, μειώνεται στο μισό το τυπικό σφάλμα του X. Στο σχήμα παρουσιάζονται η κατανομή δειγματοληψίας του μέσου X που προέρχεται από κανονικό πληθυσμό για μεγέθη δείγματος **η**1 **και** n2, n1>n2.



Εξάλλου, για δεδομένο μέγεθος δείγματος, το τυπικό σφάλμα της κατανομής δειγματοληψίας του Χ είναι ανάλογη με την τυπική απόκλιση του πληθυσμού. Επομένως, όσο μεγαλύτερη είναι η μεταβλητότητα στον πληθυσμό, τόσο μεγαλύτερη είναι η πιθανότητα η μέση τιμή δείγματος μεγέθους να διαφέρει από τη μέση τιμή του πληθυσμού.

ΚΕΦΑΛΑΙΟ 5^{Ο}

ΓΕΝΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ

5.1 ΓΕΝΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ ΣΤΗ ΔΕΙΓΜΑΤΟΛΗΨΙΑ

Η θεωρεία της δειγματοληψίας είναι μέρος της στατιστικής επιστήμης, που εφαρμόζεται για τη σωστή διεξαγωγή ερευνών σε οικονομικά και κοινωνικά φαινόμενα.

Για τη δειγματοληψία υφίστανται δύο κύριες μέθοδοι συγκέντρωσης στατιστικών στοιχείων, η απογραφή και η δειγματοληπτική έρευνα. Κατά την απογραφή συλλέγονται όλες οι παρατηρήσεις του πληθυσμού σε μια συγκεκριμένη χρονική στιγμή, ενώ κατά τη δειγματοληπτική έρευνα οι εκτιμήσεις για τις ιδιότητες του πληθυσμού αποτελούν προσεγγίσεις που περιέχουν κάποιο σφάλμα. Το σύνολο των τεχνικών βάση των οποίων εξάγονται συμπεράσματα για τον πληθυσμό χρησιμοποιώντας τις δειγματικές πληροφορίες αναφέρεται ως στατιστική συμπερασματολογία

Κατά το σχεδιασμό της δειγματοληψίας δεν απαιτείται μόνο ο προσδιορισμός των οικονομικών και χρονικών περιορισμών, αλλά είναι απαραίτητο πριν προβούμε στην επιλογή του δείγματος, να ορίσουμε το σύνολο των μονάδων που αποτελούν τον ερευνώμενο πληθυσμό οι οποίες ονομάζονται δειγματοληπτικές μονάδες (sampling units). Η δειγματοληπτική μονάδα πρέπει να ορίζεται με σαφήνεια, ώστε να μπορούμε να γενικεύσουμε τα συμπεράσματα που προκύπτουν από το δείγμα στο σύνολο του ερευνώμενου πληθυσμού. Αυτό επιτυγχάνεται με το σχεδιασμό του κατάλληλου δειγματοληπτικού πλαισίου.

Το δειγματοληπτικό πλαίσιο αποτελεί βασική προϋπόθεση για την επιτυχία μιας δειγματοληπτικής έρευνας, δεδομένου ότι πρέπει να περιέχει όλο το δειγματοληπτούμενο πληθυσμό ώστε κάθε δειγματοληπτική μονάδα να έχει την ιδία πιθανότητα επιλογής κατά τη δειγματοληψία. Σε αντίθετη περίπτωση, όπου υφίστανται παραλείψεις ή διπλές καταχωρήσεις, το δείγμα που θα επιλεγεί, δε θα είναι αντιπροσωπευτικό και κατά συνέπεια τα στατιστικά συμπεράσματα που θα γενικευτούν δε θα είναι αξιόπιστα. Οι συνήθεις τύποι λαθών που περιέχονται στα δειγματοληπτικά πλαίσια είναι οι εξής: λανθασμένη πληροφόρηση, ελλιπής πληροφόρηση, περισσότερες πληροφορίες από ότι χρειάζεται. Για τον περιορισμό των παραπάνω λαθών είναι απαραίτητη η ενημέρωση των καταλόγων πλαισίων, πριν τη χρησιμοποίηση τους για την επιλογή των τελικών μονάδων του δείγματος, ώστε να προκύψουν αξιόπιστα αποτελέσματα. Το πόσο τυχαίο είναι το δείγμα του ερευνώμενου πληθυσμού, χαρακτηρίζει τη δειγματοληψία σε τυχαία και μη τυχαία.

Στην τυχαία δειγματοληψία (random sampling) η επιλογή των δειγματοληπτικών μονάδων γίνεται κατά τρόπο τυχαίο γεγονός που δε συμβαίνει στη μη τυχαία δειγματοληψία (non random sampling), όπου οι μονάδες του ερευνώμενου πληθυσμού δεν έχουν την ίδια πιθανότητα να συμπεριληφθούν στο δείγμα και κατά συνέπεια τα στατιστικά συμπεράσματα δεν είναι αντιπροσωπευτικά στο σύνολο του ερευνώμενου πληθυσμού. Τέλος αξίζει να αναφερθούμε και στην περίπτωση της μεροληπτικής δειγματοληψίας (biased sampling) όπου κατά την επιλογή του δείγματος από ένα πληθυσμό, δίνεται μεγαλύτερη ευκαιρία σε μερικές μονάδες να συμπεριληφθούν στο δείγμα.

Το δειγματοληπτικό σφάλμα (sampling error) ορίζεται ως η διαφορά μεταξύ της εκτίμησης μιας παραμέτρου που προκύπτει από την τυχαία επιλογή των δειγματοληπτικών μονάδων και της αντίστοιχης υπολογισθείσας πληθυσμιακής παραμέτρου που προκύπτει από την απογραφή του πληθυσμού. Όσο αυξάνει το μέγεθος του δείγματος σε μια έρευνα τόσο ελαττώνεται το δειγματοληπτικό σφάλμα. Κατά την απογραφή το δειγματοληπτικό σφάλμα είναι μηδέν.

Το μη δειγματοληπτικό σφάλμα (non sampling error) δεν οφείλεται στη δειγματοληψία, αλλά είναι σφάλμα το οποίο συναντάται σε κάθε είδους έρευνα, απογραφή ή δειγματοληπτική. Μη δειγματοληπτικά σφάλματα μπορεί να είναι μεταξύ άλλων η ακαταλληλότητα του ερωτηματολογίου, τα σφάλματα ανταπόκρισης, σφάλματα που προέρχονται από τον ερευνητή, η ακόμα και σφάλματα που έχουν δημιουργηθεί κατά την επεξεργασία των πληροφοριών.

Οι μέθοδοι που χρησιμοποιούνται για τη συλλογή στοιχείων στις δειγματοληπτικές έρευνες είναι η προσωπική συνέντευξη, η οποία χαρακτηρίζεται ως η πλέον αξιόπιστη λόγω της αμεσότητας μεταξύ του ερευνητή και του ερευνώμενου, η ταχυδρομική αποστολή του ερωτηματολόγιου, και τέλος η ενημερωτική επιστολή η οποία έχει σκοπό να προετοιμάσει τα ερευνώμενα άτομα να συνεργαστούν πρόθυμα και με ειλικρίνεια με τον ερευνητή που θα τα επισκεφτεί.

5.2 ΓΕΝΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ ΣΤΙΣ ΚΑΤΑΝΟΜΕΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

Τυχαία μεταβλητή ονομάζεται η συνάρτηση η οποία αντιστοιχεί ένα πραγματικό αριθμό σε κάθε στοιχειό του δειγματικού χώρου. Οι τυχαίες μεταβλητές διακρίνονται σε διακριτές και συνεχείς.

Μια τυχαία μεταβλητή ονομάζεται διακριτή όταν μπορεί να πάρει μονό διακεκριμένες τιμές σε κάποιο διάστημα, ενώ συνεχείς όταν μπορεί να πάρει οποιαδήποτε τιμή σε κάποιο διάστημα. Οι κυριότερες διακριτές κατανομές, που παρουσιάζουν ενδιαφέρον εξαιτίας των πολλών εφαρμογών τους είναι η διωνυμική και η Poisson.

Η διωνυμική κατανομή συνδέεται με τη δόκιμη Bernoulli, ένα απλό πείραμα τύχης, όπου το αποτέλεσμα μπορεί να είναι είτε «επιτυχία» είτε «αποτυχία». Αν Χ ο αριθμός σε μια ακολουθία n ανεξάρτητων δοκίμων Bernoulli με σταθερή πιθανότητα επιτυχίας p σε όλες τις δόκιμες, τότε η κατανομή της τυχαίας μεταβλητής Χ καλείται διωνυμική κατανομή. Η διωνυμική κατανομή εφαρμόζεται σε πειράματα τύχης που χαρακτηρίζονται από τις ακόλουθες συνθήκες:

1) Υπάρχουν δυο μόνο λογικά αποτελέσματα στο πείραμα τύχης: να πραγματοποιηθεί, για παράδειγμα, το ενδεχόμενο Α ή να μην πραγματοποιηθεί. Οπότε, αν οι αντίστοιχες πιθανότητες είναι p και q, τότε ισχύει ότι: $p+q=1$.

2) Το πείραμα τύχης εφαρμόζεται n ανεξάρτητες φορές ή εκτελούνται n ανεξάρτητα αλλά ίδια πειράματα τύχης μία φορά.

3) Η πιθανότητα p της πραγματοποίησης του ενδεχομένου A παραμένει σταθερή στις n επαναλήψεις του πειράματος τύχης ή είναι ίδια και στα n πειράματα που θα εκτελεστούν μία φορά.

Η κατανομή Poisson χρησιμοποιείται για να περιγράψει φαινόμενα τα οποία μπορούν να θεωρηθούν ως διαδικασία παράγωγης τυχαίων εμφανίσεων ενδεχομένων κατά διαστήματα χώρου ή χρόνου. Απαραίτητη προϋπόθεση για τη χρησιμοποίηση της κατανομής Poisson είναι να ισχύουν τα εξής:

1) Η πιθανότητα να πραγματοποιηθεί ένα συμβάν σε διάστημα μήκους t είναι ανάλογη προς το μήκος του διαστήματος.

2) Η πιθανότητα να πραγματοποιηθούν δυο ή περισσότερα συμβάντα σε διάστημα μήκους t είναι πολύ μικρή σε σχέση με την πιθανότητα εμφάνισης ενός μόνο συμβάντος, όταν το μήκος του διαστήματος είναι μικρό.

3) Η πιθανότητα εμφάνισης k συμβάντων σ’ ένα διάστημα είναι ανεξάρτητη από την αντίστοιχη πιθανότητα σ’ ένα δεύτερο διάστημα ξένο προς το πρώτο.

Από τις συνεχείς κατανομές εκείνη που παρουσιάζει ιδιαίτερο ενδιαφέρον είναι η κανονική κατανομή. Η κανονική κατανομή αποτελεί τη βάση για πολλές τεχνικές που χρησιμοποιούνται στη στατιστική συμπερασματολογία. Στην κανονική κατανομή η μέση τιμή, η διάμεσος και η επικρατούσα τιμή συμπίπτουν. Πολλά πειράματα μπορούν να εκφραστούν μέσω τυχαίων μεταβλητών που ακολουθούν κανονική κατανομή ενώ μπορεί να χρησιμοποιηθεί και ως προσέγγιση πολλών άλλων κατανομών. Η λεγόμενη κατανομή δειγματοληψίας είναι μια κατανομή πιθανότητας που ακολουθεί κάθε στατιστική. Οι κατανομές δειγματοληψίας των στατιστικών χαρακτηρίζονται από βαθμούς ελευθερίας. Μια στατιστική έχει πλήρεις βαθμούς ελευθέριας όταν δεν περιέχει παραμέτρους κατανομής του πληθυσμού ή περιέχει παραμέτρους, αλλά οι τιμές τους είναι γνωστές.

Όσον αφορά στην κανονική κατανομή του μέσου, αποδεικνύεται ότι για τυχαία δείγματα μεγέθους n από οποιοδήποτε πληθυσμό με μέσο μ και διακύμανση σ^2 που έχει άπειρο πλήθος στοιχείων ή από πληθυσμό που έχει πεπερασμένο πλήθος στοιχείων αλλά που η δειγματοληψία γίνεται με επανατοποθέτηση, ότι οι μέσοι των τυχαίων αυτών δειγμάτων κατανέμονται με μέσο ίσο με μ και διακύμανση ίση με σ^2/n.

ΒΙΒΛΙΟΓΡΑΦΙΑ- ΠΗΓΕΣ

- Χαλικιάς, Ιωάννης Γ, ΣΤΑΤΙΣΤΙΚΗ, Rosil, ΑΘΗΝΑ, 2003
- Ζαχαροπούλου, Χρυσούλα, Στατιστική, Σοφία, 2005
- Ζαχαροπούλου, Χρυσούλα, Ασκήσεις Στατιστικής, Σοφία, 2005
- Ιωαννίδης, Δημήτρης, Στατιστικές μέθοδοι, Ζήτη, Θεσσαλονίκη, 2005
- Χρήστου, Γεώργιος Κ, Εισαγωγή στην οικονομετρία Α και Β ΤΟΜΟΣ, Gutenberg, Αθήνα, 2003
- Γεωργίου Δ, Πιθανότητες και στατιστική, Κλειδάριθμος, Αθήνα, 2009
- Paul, Port, Sidney, Stone, Charles, Εισαγωγή στη θεωρία των πιθανοτήτων, Πανεπιστημιακές Εκδόσεις Κρήτης, 2002
- Γκαρούτσος, Γιάννης Β, Μαθήματα Θεωρίας Πιθανοτήτων, SPIN, Αθήνα
- ΦΑΡΜΑΚΗΣ ΝΙΚΟΣ, ΕΙΣΑΓΩΓΗ ΣΤΗ ΔΕΙΓΜΑΤΟΛΗΨΙΑ, Χριστοδουλίδη, 2002
- Κιόχος, Πέτρος Α., Χαρίσης, Κώστας Ι, Θεωρία δειγματοληψίας και εφαρμογές, Interbooks, Αθήνα, 1997

ΔΙΑΔΙΚΤΥΟ

- www.openarchives.gr
- www.uoa.gr
- www.aueb.gr
- www.uoc.gr
- www.uom.gr
- www.teipat.gr
- www.aegean.gr