

Στατιστική είναι η επιστήμη που (1)

ασχολείται με την συλλογή, την παρουσίαση και την επεξεργασία αριθμητικών και τη στοιχείων με απώτερο σκοπό την εξαγωγή συμπερασμάτων για τους ευρύτερους πληθυσμούς στους οποίους αφορούν τα στοιχεία αυτά

Υποδιαιρείται σε

ΜΑΘΗΜΑ

1

Περιγραφική Στατιστική: Αποσκοπεί στην

ατάλληλη περιγραφή του υπό μελέτη πληθυσμού στο σύνολο είτε του κατάλληλου δείγματος

Εξαγωγική Στατιστική: Αποσκοπεί στην

εξαγωγή συμπερασμάτων που αφορούν στον γεννιότερο πληθυσμό και επιτρέπει την γενίκευση συμπερασμάτων που έχουν προκύψει από ένα δείγμα για το σύνολο του πληθυσμού.

Υποδιαιρείται σε

Ελέγχος υπόθεσεων (hypothesis testing):
Εξαγωγή υποθέσεων σχετικά με την τιμή των παραμέτρων του πληθυσμού.

Εκτιμητική (inference)

- Σμφείου: εκτίμηση αριθμού τιμής
- Διαστήματος: προσδιορισμός διαστήματος που περιέχει την τιμή.

Συγκριτικές έννοιες

Πληθυσμός : Ένα σύνολο N οντοτήτων ή αντικειμένων τα χαρακτηριστικά των οποίων επιθυμούμε να μελετήσουμε.

→ Η τιμή του N μπορεί να είναι πεπεραμένη ή άπειρη

→ Αυτά τα N αντικείμενα καλούνται γονάδες δείγματοληψίας

Στοιχειώδεις γονάδες ονομάζονται όλες οι γονάδες του πληθυσμού

Πρωταρχικές γονάδες είναι οι στοιχειώδεις γονάδες ή σύνολα στοιχειωδών γονάδων

ΔΕΙΓΜΑ ονομάζεται για συλλογή στοιχειωδών ή πρωταρχικών γονάδων δείγματοληψίας

Η επιλογή του δείγματος υπαγορεύεται από τη θεωρία δείγματοληψίας.

Τα χαρακτηριστικά του δείγματος περιγράφονται από μεταβλητές.

- Ποσοτικές επιδέχονται μέτρηση και οι τιμές τους εκφράζονται με αριθμούς και μονάδες (Συνεχείς και διακριτές)
- Ποιοτικές εκφράζονται με λέξεις ή συμβόλα και περιγράφουν με αριθμητικές ιδιότητες ή χαρακτηριστικό επίπεδο.

ΚΛΙΜΑΚΕΣ ΥΕΤΡΗΣΗΣ

(3)

Ονομαστική, διατάξης, διασυντάτος, λοφού

Πηγές δεδομένων

Πρωτογενείς: τα δεδομένα επιλέγονται από τον ερευνητή μέσω υετρήσεων (πχ υετρήση βάρους, υγούς κλπ) ή υεσω συλλογή, απαντήσεων και ερωτηματολόγια που υεμσιφοποιούνται για τη διαφύνηση των αλογών των ερωτώυων επί συυερισμένων υεγμάτων.

Δευτερογενείς: τα υεδομένα ανιλονται από υεγεις υεδομένων υημρυοίων στατιστικής υηυροφορής (ΕΛΣΤΑΤ, ΟΗΕ, ΟΟΣΑ, υραυέ)...

ΣΤΑΤΙΣΤΙΚΗ ΕΡΕΥΝΑ

Οριζεται ως η επιστηυονική υεθοδολογία που εχει υυοπό να συλλεξει, επεξευασει, ανάλυγει, παρυσιασει, και ερμηνυσει υεδομένα αριθμητική υυοφως που υυορουν σε ένα ή υερισσότερα συυατικά του πλυυοφου.

Στατιστική Ανάλυγη είναι το υυμφο της

στατιστικής ερευνας που υυορδ στων ανάλυγη των υεδομένων.

Η στατιστική έρευνα είναι δύο ειδών

(4)

- Απογραφική Συλλέγονται πληροφορίες από όλες τις μονάδες του πληθυσμού
- Δειγματοληπτική Συλλέγονται πληροφορίες μόνο από ένα δείγμα του πληθυσμού και στη συνέχεια τα συμπεράσματα που προκύπτουν γενικεύονται για ολόκληρο τον πληθυσμό.

Συνήθως η έρευνα είναι δειγματοληπτική επειδή ολοκληρώνεται πιο γρήγορα με συλλογή και με επεξεργασία των αποτελεσμάτων και επειδή πρακτικά είναι αδύνατο να συλλεχθούν πληροφορίες από ολόλο του πληθυσμού.

ΜΕΛΕΤΕΣ

Παρατηρητικές ή μη παρεμβατικές
Πειραματικές ή παρεμβατικές.

Παρατηρητικές (observational)

1) Συγχρονικές (cross sectional): λαμβάνονται μετρήσεις από όλα των πληθυσμό ή από τμήμα υποομάδα του σε κάποια συγκεκριμένη χρονική στιγμή.

2) Διαμήκειες ή διαχρονικές (longitudinal): 5
Περιλαμβάνουν μετρήσεις επί των ίδιων μελε-
τών για μεγάλο χρονικό διάστημα.

3) Αναδρομικές (retrospective): αφορούνται
σε γεγονότα που έχουν συμβεί στο παρελθόν.

4) Προοπτικές (prospective): παρακολουθείται
διαχρονικά για οριστά άτομα με στόχο την
αξιολόγηση κάποιων παραγόντων.

Θεωρία Δειγματοληψίας

Προσδιορίζει τον σωστό τρόπο επιλογής ενός
δείγματος του πληθυσμού.

Τυχαιο (random) δείγμα: λαμβάνεται με
τυχαίο τρόπο με κάποια τεχνική δειγματοληψία.

Είναι άρρηκτα συνδεδεμένη με τη θεωρία
πιθανοτήτων.

Δειγματοληπτικό πλάνο

Κατάλογος δειγμάτων που διασφαλίζει τη σωστή
επιλογή των δειγμάτων υπό την έννοια που αυτά
αντικαθιστούν οντως στον υπό εξέταση πληθυσμό.
(επιτόκιοι (τηλεφωνικοί κλήσεις, έμμεσα φοιτητών)

Σχέδιο δειγματοληψίας

6

Πώς επιλέγονται τα δείγματα?

Απλή τυχαία δειγματοληψία: τα δείγματα επιλέγονται στην τύχη και κάθε μονάδα πληθυσμού έχει την ίδια πιθανότητα να συμπεριληφθεί στο δείγμα.

- με επανάληψη: γλορεί να επιλεγεί περισσότερες από μία φορές
- χωρίς επανάληψη: επιλέγεται αμοιβάς για φορές.

Συστηματική δειγματοληψία

- 1) Επιλογή m δειγμάτων από έναν πληθυσμό N ατόμων.
- 2) Υπολογισμός του βήματος δειγματοληψίας
$$k = \frac{N}{m}$$
- 3) Επιλογή ενός τυχαίου δείγματος α
- 4) Επιλογή των δειγμάτων
$$B_i = \alpha + i k \quad (i=1, 2, 3, \dots)$$

ΤΥΧΑΙΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑ (7)

Ο πληθυσμός χωρίζεται σε ομάδες βάσει κάποιων κριτηρίων (φύλλο, ηλικία, φύση κτλ) και μετά από κάθε ομάδα επιλέγεται ένα δείγμα με τυχαίο τρόπο. (random stratified)

ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΣΥΣΤΑΔΑΣ (cluster sampling).

Ο πληθυσμός διαιρείται σε ομάδες αλλά τα μέλη μέσα σε κάθε ομάδα θα πρέπει να διαφοροποιούνται όσο το δυνατόν περισσότερο, ενώ οι διαφορές μεταξύ των συστάδων θα πρέπει να είναι όσο το δυνατόν πιο γιγαντιαίες.

ΜΗ ΤΥΧΑΙΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑ

ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΚΡΙΣΗΣ (Judgment sampling)

Σχεδιάζεται σύμφωνα με την κρίση του ερευνητή έτσι ώστε να εκπαισθεί ένα όσο το δυνατόν πιο αντιπροσωπευτικό δείγμα. Η πιθανότητα επιλογής ή γονάδας στο δείγμα δεν μπορεί να υπολογιστεί.

ΚΑΤΕΥΘΥΝΘΕΝΤΗ ΔΕΙΓΜΑΤΟΛΗΨΙΑ (directed sampling)

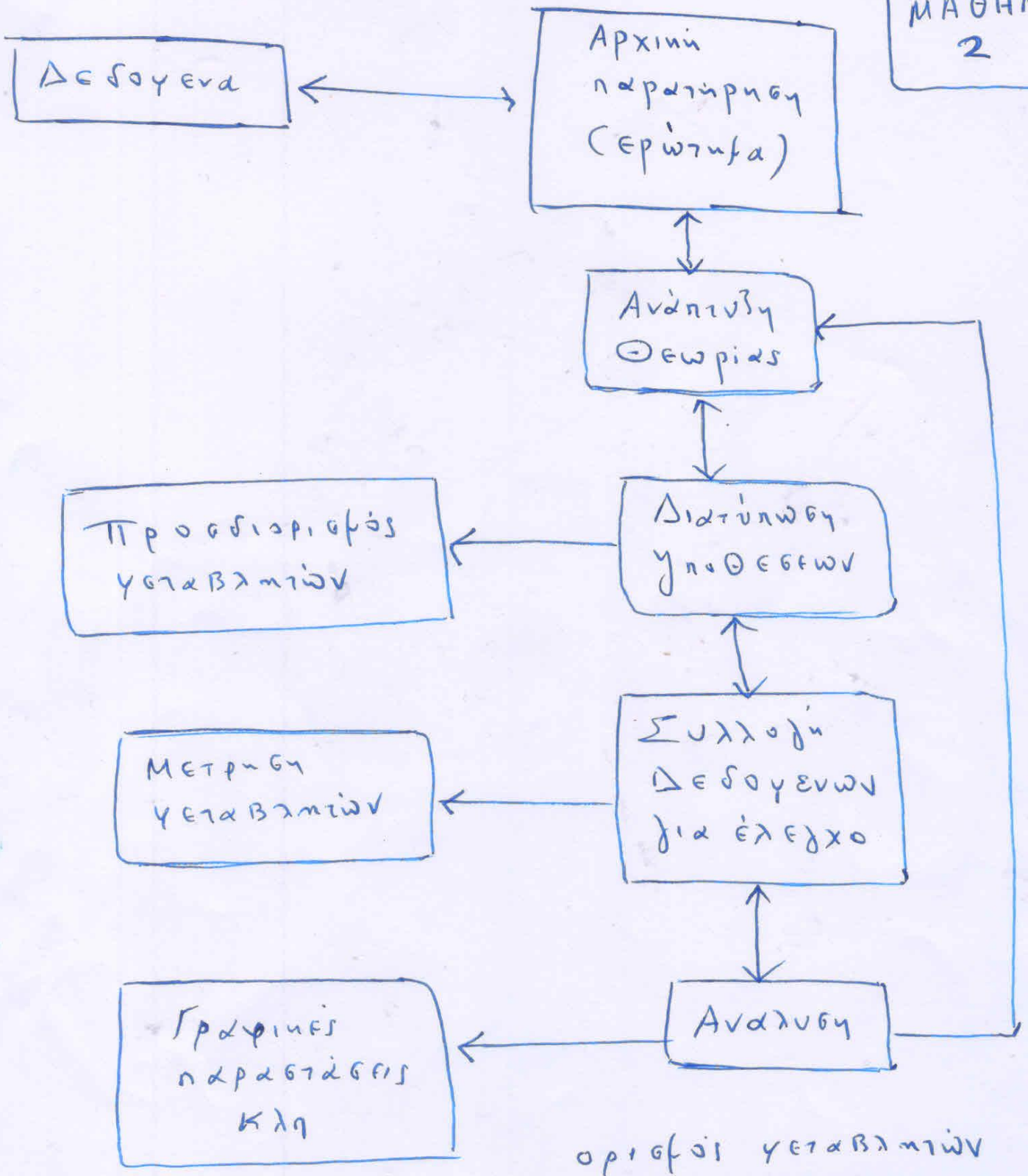
Στηρίζεται σε προαπορισμένα χαρακτηριστικά έτσι ώστε να επιλεγεί δείγμα ή ίδια συνθήκη ή αλλη του πληθυσμού.

ΣΤΟΧΟΣ Αντιπροσωπευτικού δείγματος.

Ερευνητική Διαδικασία

1

ΜΑΘΗΜΑ
2



Υπόθεση

Εικόνα με
οποια πρέπει να επαληθευτεί ή
να απορριφθεί

- Πραγματοποιείται συλλογή δεδομένων προς ενάτι θεώρηση ή διαψευδή της υπόθεσης.
- Εάν τα δεδομένα επαληθεύουν την ύπωση έχουμε επαληθευσιμότητα (verification) ενώ εάν αντίθετα κερνήτωση έχουμε διαψευδισιμότητα (falsification)
- Ο έλεγχος της υπόθεσης (hypothesis testing) συμπίπτει στη χρήση μεταβλητών πχ, ηλικία, φάρμακο φαίλεδο, τροφή / πως κλπ

Υψι γράζει σχέση αιτίου / αποτέλεσμα
 πχ στη δόλωση

Ο κωφός φέρνει αυτιά

Ο "κωφός" είναι το αίτιο ενώ η "αυτιά" είναι το αποτέλεσμα

Το αίτιο ονομάζεται ανεξάρτητη μεταβλητή (independent variable) ή μεταβλητή πρόβλεψης (predictor variable)

Το αποτέλεσμα ονομάζεται εξαρτημένη μεταβλητή (dependent variable) ή μεταβλητή αποτελέσματος (outcome variable)

Για την αλληλοεξάρτηση μεταβλητών (3)
μεταβλητών χρησιμοποιούμε γεωμετρικές σχέσεις

• Κλίμακα μέτρησης (level of measurement)

Η σχέση που υφίσταται ανάμεσα σε αυτό που μετράμε και στους αριθμούς που το εκφράζουν.

Τύποι μεταβλητών

ΚΑΤΗΓΟΡΙΚΗ ΜΕΤΑΒΛΗΤΗ (categorical variable)

Εκφράζει κατηγορίες πραγμάτων (πχ ανδρικοί)
Μπορεί να είναι

1) Διτιμή (binary) αν παίρνει μόνο δύο
αμοιβαία αποκλειόμενες τιμές
(πχ ανδρικοί / γυναικείο ή ΝΑΙ / ΟΧΙ)

2) Ονομαστική (nominal) αν παίρνει περι-
σότερες από δύο τιμές.

Δεν είναι αριθμητικές φέρει και δεν φαίνεται
υπόμνηση πράξης. Αν n είναι το πλήθος των
τιμών τους
τότε



• Διατάκτινή / διατάξιμη ή ιεραρχική
μεταβλητή (ordinal): επιτρέπει τη
 διατάξη των αντικειμένων γφ βάση κάποια
 λογική πχ σε έναν διαγωνισμό το Α
 βγήκε πρώτο, το Β δευτερο και το Γ τρίτο.

ωστόσο ΔΕΝ γας πληροφορούν για την διαφορά
 που υφίσταται ανάμεσα τους δηλαδή ποσοί
 ψήφοι διαφέρουν ήταν ανάμεσα στο Α και στο
 Β ή ανάμεσα στο Β και στο Γ.

• Συνεχής μεταβλητή (continuous variable)
 λαμβάνει οποιαδήποτε τιμή πχ βάρος 71.5K
 Διακρίνονται σε :

Μεταβλητές διαστήματος (interval variable)
 πχ ηλικία 5-10 ή 10-15 ή 15-20

Μεταβλητές λόγου (ratio variables)
 εάν δεν υφίσταται η ηγετικότητα προς
 γείγνησι λαμβάνουν τιμή μηδέν

• Διακριτή μεταβλητή (discrete variable)
 λαμβάνει μόνο διακριτές τιμές συνήθως
 ακέραιες πχ (ηλικία) 18, 25 κλπ

Σφάλμα μέτρησης (measurement error)

(5)

ορίζεται ως η διαφορά ανάμεσα στην μετρούμενη τιμή και στην πραγματική της τιμή

Σχετικοί ορισμοί:

• Εγκυρότητα (validity): Αξιολογεί κατά πόσο τα ερωτήρια μέτρα που έχει κατασκευαστεί για να μετρήσει αυτό για το οποίο δημιουργήθηκε με βάση κάποια κριτήρια

1) Συντρέχουσα εγκυρότητα (concurrent validity)

Η τιμή που αποτιμάται συσφώνει με εκείνη που προκύπτει από την πραγματικότητα.

2) Εγκυρότητα πρόβλεψης (predictive validity)

Αφορά σε μελλοντικές προβλέψεις

3) Εγκυρότητα περιεχομένου (content validity)

Διερευνάται ο βαθμός στον οποίο τα επιμέρους ερωτήματα (πχ κάποιον ερωτηματολόγιου) μπορούν να εκφράσουν ή να περιγράψουν πλήρως το υπό μετρηση στοιχείο

1) ΕΡΕΥΝΕΣ ΣΥΣΧΕΤΙΣΗΣ (CORRELATIONAL)

5

ή ΓΩΣΤΙΚΕΣ ΔΕΙΓΜΑΤΟΛΗΠΤΙΚΕΣ ΕΡΕΥΝΕΣ

(cross / sectional research) παρατηρούμε
αλλά τι συμβαίνει στο πραγματικό περιβάλλον
χωρίς να παρεμβαίνουμε στις μεταβλητές.

2) Πειραματική Έρευνα (Experimental Research)

παρεμβαίνουμε στις μεταβλητές προκειμένου
να παρατηρήσουμε την επίδραση της παρέμβασης
στις μεταβλητές.

Στην έρευνα συσχέτισης γέτουμε πραγματικά
γεγονότα ή τα ελπίζα ρυθάνγης σε ένα
ποτάφι, έχοντας τη δυνατότητα να προβούμε σε
έναν πολύ φυσικό τρόπο καταφρής με
άνάντμγης του ερευνητικού ερωτήματος.

ΜΕΘΟΔΟΛΟΓΙΑ ΠΕΙΡΑΜΑΤΙΚΗΣ ΕΡΕΥΝΑΣ

ηροσδιορισμός της αιτιατής συσχέτισης γειάζύ
των εγληκόγενων γειάβλητων η οποία γλορει
να είναι περισσότερο ή λιγότερο προφανής.

Σε ποιο βαθμό το αιτιό σχετίζεται με το αποτέλεσμα?

(7)

Θεωρία David Hume (1711-1776)

προκειμένου να υπάρχει σχέση μεταξύ αιτίου και αποτελέσματος θα πρέπει:

- 1) Το αιτιό και το αποτέλεσμα εμφανίζονται σε κοντινά χρονικά διαστήματα
- 2) Το αιτιό προηγείται του αποτελέσματος
- 3) Το αποτέλεσμα δεν εμφανίζεται αν δεν υπάρχει το αιτιό

Το αιτιό συνδέεται με υψηλό βαθμό συσχέτισης με τα εναλλακτικά βασικά συβάντα

Σε περιπτώσεις κατά τις οποίες η συσχέτιση δεν είναι εμφανής επηρεάζεται και ενδεώς περισσότεροι λόγοι ή κρυφοί παράγοντες το ενοσφαισθενό ΤΡΙΤΟΝ ΤΙ

Αυτοί οι εξωτερικοί παράγοντες σχετίζονται με μεταβλητές συσχέτισης

(confounding variables)

Πώς συλλέγουμε δεδομένα?

(8)

1) Σχεδιασμός μεταβυ ογώνων:

Χειρ/ογαστε την ανεξαρτητη μεταβλητη χρησιμοποιωντας διαφορετικα ατογα (πχ συλληψη ερωτηματολογιων από ογώνων φοιτητών)

2) Σχεδιασμός γειρήσεων εντός των υποκειμένων

Χρησιμοποιουμε τα ίδια ατογα μεταβαλλοντας τις συνθκες του περιβάτους.

ΤΥΠΟΙ ΜΕΤΑΒΛΗΤΟΤΗΤΑΣ

Συστηματική μεταβλητότητα (systematic variation): η αποκλιση οφειλεται αποκλειστικά σε κάποια ενεργεια που πραγματοποιεί εκείνος που εκτελεί το πείραμα και η οποία πραγματοποιείται στη ίδια κατάσταση αλλά όχι στην άλλη.

Μη συστηματική μεταβλητότητα (unsystematic variation): η αποκλιση οφειλεται σε τυχαίους παράγοντες οι οποίοι υφίστανται και διαφθοροποιούνται μεταξύ των διαφορετικών πειραμάτων και καταστάσεων.

Αυτή η fm συστημάτων μεταβλητότητα (9)
⊖ α πρέπει να ηφριερίεται στο ελάχιστο
εργαστείτε να έχουν ένα αυριβεβέτορο γέτρο
για τη συστημική παρτεβ αγγ.

Τυχαιοποίηση (Randomization)

Διερεύνηση των πημών συστημικής και
μη συστημικής μεταβλητότητας και
συνδυασμό γέτρο τους.

Εναλλακ (Βανογέρες) γέτρος

- 1) επίδραση εβοικαιωσ (practice)
- 2) επίδραση πλίζης (boredom)

Οι επίδρασεις γειωνονται γέ εναλλαγή της
γέτρος ολοκληρωσ των ηφρατατιών
συν θύων.

οφοίως και στις αυφβέρτες γέτρος

① ΣΧΕΔΙΑΣΜΟΣ ΥΕΛΕΥΣ

Τι θέλω να εξετάσω?

ΜΑΘΗΜΑ

3

ΤΕΧΝΙΚΕΣ ΣΥΛΛΟΓΗΣ ΣΤΑΙΧΕΙΩΝ

- Προφορική συνέντευξη
Προσωπική (face 2 face)
Τηλεφωνική
Οταβική (οχι αξιολογική)
- Ερωτηματολόγια
- Φυσική παρατήρηση

Για τα ερωτηματολόγια πριν δοθεί απαιτείται προεργασία γνωστή ως πιλοτική έρευνα (pilot survey) έτσι ώστε να διασφαλιστεί ότι καλύπτει τις ανάγκες της έρευνας

Διαδικασία διεξαγωγής έρευνας

- Προσδιορισμός Βάσεων στόχων
- Προσδιορισμός πληθυσμού - στόχου
- Επιλογή δείγματος
- Καθορισμός τρόπου συλλογής δεδομένων
- Ευλόγηση δεδομένων
- Κατάχωρηση δεδομένων / λογισμικό ελέγχου
- Στατιστική ανάλυση
- Συδράση ή εκθέσεις αποτελεσμάτων

2

Συνέντευξη

Προσωπική, ομαδική τηλεφωνική
Εδώ το ερωτηματολόγιο συμπληρώνεται από τον
ερευνητή.

Γραπτή ατομική συμπλήρωση

Το ερωτηματολόγιο συμπληρώνεται από
τον ερωτώμενο χωρίς την παρουσία του
ερευνητή. Η κριτική είναι και αποβλεπτική
συμπληρωμένο ταχυδρομικά ή μέσω email

Σχεδιασμός ερωτηματολογίου

- Τελείως αναγκαία ερωτήματα για να φαν
που κερδίζουν τον ερευνητή
- Σαφής διατύπωση ερωτήσεων
- Ομαδοποίηση ερωτήσεων σε τμήματα
- Η σειρά των ερωτήσεων είναι σημαντική.

Τύποι ερωτήσεων

1/ Κλειστές

~~Α~~ επιδέχονται συγκεκριμένη
απάντηση

Διχοτομικές ΝΑΙ / ΟΧΙ

2/ Ανοιχτές

ελεύθερο

κείμενο.

3) Ερωτήσεις απλής επιλογής και
πολλαπλών επιλογών

3

4) Ερωτήσεις κατάταξης ΠΧ

Πολύ βυθιτικό

Σημαντικό

Αδιαφορο

Αβυθιτικό

Εντελώς αβυθιτικό

Κωδικοποίηση ερωτηματολογίου

ορίζεται ως η διαδικασία αντιστοίχησης
ενός αριθμού σε κάθε δυνατή / εναλλακτική
απάντηση ενός ερωτήματος.

ΠΧ

Ανδρας → 1

Γυναίκα → 2

Αυτοί οι αριθμοί προφανώς ΔΕΝ χρησιμοποιούνται
σε αριθμητικές πράξεις.

Βιβλίο κωδικοποίησης Code Book

Πινάκας μεταβλητών που περιλαμβάνονται στο
αρχείο δεδομένων που βοηθά στην
κατανόηση των αποτελεσμάτων.

Τα επιθυμητά χαρακτηριστικά ενός
ερωτηματολογίου περιλαμβάνουν:

4

(Αξιολόγηση)

• Διαχωριστική ικανότητα (discrimination)

Ατομά με διαφορετικές βιολογικές
σε ένα ερωτηματολόγιο θα πρέπει να
διαφέρουν ως προς την έννοια που μετρά
το ερωτηματολόγιο

• Εγκυρότητα (validity): αποτελεί

μέτρο του συστηματικού σφάλματος και
αντιαντικειμενικά κατά πόσον το εργαλείο της
μέτρησης μετρά οντως αυτό που υποτίθεται
πως μετρά.

1) Εμπυρότητα περιεχομένου: εξετάζει το

βιολογικό στον οποίο ένα εργαλείο μέτρησης
περιλαμβάνει ερωτήσεις που αντιπροσωπεύουν
όλες τις πιθανές πτυχές της έννοιας που
μετράται, δηλαδή το κατά πόσον αντιπροσω-
πευτικές είναι οι ερωτήσεις.

2) Εμπυρότητα εννοιολογικής δομής: μετρά

τη σύμφωνία του εννοιολογικού ή θεωρητικού
του λειτουργίου ορισμού των υπό έρευνα εννοιών
δεικνύει δηλαδή εάν η ψυχολογικά αντιλαμβανόμενη
είναι το εννοιολογικό πλαίσιο.

3) Εξυμνοσιότητα κριτηρίου

(5)

Διδικυμνεται σε προβλεπτική και ταυτοχρονη
υε κριτήριο το εαν υεβολαβκι αριετο διαβτη-
υα ανδρεβα στη υετηρη και στο κριτηριο.

Η πρμη περιπτωση (προβλεπτική)
αφορα στην καια σταβα καια των ολοια
ειναι δυνατη η προβλεψη γεγονοτων υε βδ βα
των υλο υελετη εννοια

Η δευτερη περιπτωση (ταυτοχρονη) αφορα
στην καια σταβα καια των ολοια υγιετα διαβκι
υε σχετιγη υεταβυ των εννοιων Α και Β η
ολοια ελαλη δευεται αλο το το ερωτηματολοιο
υια των καια σταβα Β.

Σε αφοτερες τις περιπτώσεις το
ερωτηματολοιο θεωρείται εξυμνο

Σφάλμα υέτηρης

Παρατηρηθειβα τιμή —
πραγματοική τιμή

Τυχαίο σφάλμα: επιφερει αποκλιση στις
παρατηρηθειβες τιβες υυρω αλο των
πραγματοική τιμή

Συστηματικό σφάλμα: η αποκλιση δεν
σφειλεται στην ωχη

Αξιοπιστία (reliability)

6

Αποτελεί μέτρο του πλάτους διακλίσεων και αντανακλά τη συνέπεια (consistency) με την οποία ένα εργαλείο μέτρησης μετρά ένα χαρακτηριστικό. Ένα τέτοιο εργαλείο θεωρείται αξιόπιστο όταν σε επαναλαμβανόμενες μετρήσεις δίνει αποτελέσματα που δεν διαφέρουν μεταξύ τους.

Σταθερότητα (stability): αναφέρεται σε συμφωνία διαδοχικών μετρήσεων

Ισοδυναμία (equivalence): αναφέρεται στην συμφωνία μεταξύ δύο ή η περισσότερων τμημάτων του ίδιου εργαλείου μέτρησης

Ομοιογένεια (homogeneity): αναφέρεται στη συμφωνία μεταξύ των ερωτήσεων που αποτελούν ένα ερωτηματολόγιο

Θεωρώντας ένα σύνολο γεωμετρικών που έχουν
υποδιαιρεθεί σε κάποια από τις γεωμετρικές συλλογές
στοιχείων που έχουν να παρουσιάσει, μπορούμε να
υποδιαιρέσουμε σε (Α) Φορέι Ε/φ ανήκει και
τιμή u_i να υποδιαιρεθεί σε x φορές n u_i u_i u_i
συχνοτήτων που αφιέρχουν τη συχνότητα f u_i
νίσις και την αθροιστική συχνότητα o_i
στο ενοφενό n u_i u_i (σύνολο 100 ατόμων)

Μορφωτικό Επίπεδο	Συχνότητα Ε/φ νίσις	Αθροιστική Συχνότητα
Δημοτικό	5	5
Γυμνάσιο	10	15
Λυκείο	25	40
ΑΕΙ	47	87
Μεταπτυχιακό	10	97
Διδακτορικό	3	100

Αν και αυτοί οι πίνακες είναι ιδιαίτερα (2)
υδατοφιλικόι ωστόσο τα φράγματα
είναι επίσης πολύ εξυπηρετικά διότι συνοψίζουν
σε φραγμένο τρόπο τα βασικά χαρακτηριστικά
των δεδομένων και τα παρουσιάζουν ως
εικόνες

Τύποι φραγμάτων

Ραβδόφραγματα

Ισοφραγματα

Κυκλικά διαφραγματα

Διαφραγματα γραμμής

Διαφραγματα διασποράς

Θύλοφραγματα

Τριεδιστάτα διαφραγματα

Συμμετρικά ραβδόφραγματα

Χρονοδιαφραγματα

ΣΤΑΤΙΣΤΙΚΑ ΜΕΤΡΑ

3

• Μέτρα Κεντρικής Τάσης

Αποτελούν για τον ενδιαφερόμενο επιχειρηματία να περιγράψει ένα σύνολο δεδομένων για τον εντοπισμό της κεντρικής θέσης του συνόλου δεδομένων

- Αριθμητικός μέσος
- Διαμέσος
- Σημείο μέγιστης συχνότητας

• Μέτρα Θέσης

Χρησιμοποιούνται για:

- 1) Τον εντοπισμό της σχετικής θέσης μιας τιμής σε ένα σύνολο δεδομένων
- 2) Τη συσχέτιση των τιμών των δεδομένων που προέρχονται από διαφορετικά σύνολα δεδομένων
- 3) Τη συσχέτιση τιμών μέσα στο ίδιο σύνολο δεδομένων
- 4) Τον καθορισμό των άκρων τιμών σε ένα σύνολο δεδομένων.

Στην υατρηγορια αυτη ανηκουν

(4)

- Τετρατηγορια
- Δευατηγορια
- Εηατοβτηγορια

• Μετρα διαθορας

① Προσδιοριζουν το βαθμο συμμετρως η διαθορας των τιων και χρησιμοποιουνται για αναλογισενεις υατανογες

② Εξαρτωνται απο τη διαφορα των τιων και οχι απο τη θεση τους ενω μεταβαλλονται αυτιστροφα γε τη συμμετρως των τιων ηνω απο για παραγετρο θεως.

- Ευρος
- Ενδοτετρατηγοριακο ευρος
- Μεση απολυτη απολιση
- Διχυνηση
- Τυλινη απολιση
- Σχετινη διαθορα.

• Μετρα ασυμμετρως

μας πληροθορουν για το εαν οι τιες της μεταβλητης ειναι τοποθετηενεις συμμετρικα

γύρω από το μέσο όρο ή όχι, και πόσο.

(5)

Συντελεστής συνχρησίας Pearson
Συντελεστής συνχρησίας Bowley

• Μέτρα κέντρωσης

Η κέντρωση χαρακτηρίζεται κατά πόσον η
κατανομή $f(x)$ κεντρικώς είναι
πληθυσμική ή όχι, σύμφωνα με
ακρονύχια κέντρωσης αυτής της ακρονύχια
κατανομής (Gauss)

Ανάλυση αλυσών μεταβλητών

Αριθμητικός μέσος (αριθμητικός και
αριθμητικός μέσος), βάρος $C=1$

πληθυσμικός N ατόμων, δείγμα n ατόμων

Μέσος όρος πληθυσμικός

$$\mu = \frac{1}{N} \sum_{i=1}^N X_i$$

Μέσος όρος δείγματος $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

6

Εάν οι παρατηρήσεις είναι διάφορες
 γέφυρες και έχει κατάσχεση και τις
 κατανομή συχνοτήτων γέφυρες X_i να
 κατασκευαστεί ένα σύστημα επ' αυτών F_i

Τότε

$$\bar{X} = \frac{F_1 X_1 + F_2 X_2 + \dots + F_N X_N}{F_1 + F_2 + \dots + F_N} =$$

$$\left(\sum_{i=1}^N F_i X_i \right) / \left(\sum_{i=1}^N F_i \right)$$

Παράδειγμα

Αριθμός μαθητών X_i	Αριθμός ομάδων F_i
0	4
1	7
2	9
3	5
4	2
5	3

Από

$$\bar{X} = \frac{0 \times 4 + 1 \times 7 + 2 \times 9 + 3 \times 5 + 4 \times 2 + 5 \times 3}{4 + 7 + 9 + 5 + 2 + 3} = \frac{63}{30} = 2.1 \approx 2 \text{ μαθητ.}$$

7

Αστέθμητος αριθμητικός

μέσος ομαλοποιημένων παρατηρήσεων

Ομαλοποιημένες παρατηρήσεις σε ομαλές με-
ρίδες ή άνισο πλάτος. Εάν a_i και b_i είναι το
άνω και το κάτω όριο της i -αδης ομαλότητας,
υπολογίζουμε για την i -αδα των κεντρι-
κή τιμή

$$\bar{x}_i = \frac{a_i + b_i}{2}$$

η οποία θεωρείται αντιπροσωπευτική της διάστη-
ματος και αντιυπόλογιστά των οφθαλμομέτρων υπολογιστούς

Στη συνέχεια υπολογίζουμε το μέσο οριώς

$$\bar{X} = \frac{\sum_{c=1}^v F_i \bar{x}_i}{\sum_{c=1}^v F_i}$$

Ευελιξιώντας πως η διάφορα μέτρων συνυψη-
τερο υπολογιστού δεν είναι πολύ μεγάλη.

Στέθμητος αριθμητικός μέσος

$$\bar{X} = \frac{x_1 c_1 + x_2 c_2 + \dots + x_v c_v}{c_1 + c_2 + \dots + c_v} = \frac{\sum_{c=1}^v x_i c_i}{\sum_{c=1}^v c_i}$$

οπου c_i ο συντελεστή Βάρους της
παρατήρησης x_i

8) Διάγροος : Αποτέλει το βήμα δαίστερο

γεία των αριθμητικώ γεγο βιαηβηκώ γετρο
κεντρικώ τάλου

Χωρίγει την κατανομή σε δύο ίσουλκ θείη ογάνει
και η τιμή της είναι η κεντρική τιμή της
γεταβλκτής.

Με άλλα λόγια, το 50% των τιμών είναι
γικροτέρω από την τιμή της διαφέρου ενώ
το άλλο 50% είναι γεγάλυτέρω από αυτή
των τιμών.

Υπολογισμός

- Εάν οι n κλήθησ κερκτρηθήσει είναι
αταβινομητες, τω τάλινοφέουσε, οπότε η
διαφέροο είναι το βείηκλ που βριγεηκεί
για θεση κη αριθμόν
$$\frac{n+1}{2} = α$$
- Εάν το n είναι κερητο η τιμή της διαφέρου
είναι το βείηκλ που βριγεηκεί για θεση $α$
ενω διαφορηκώ κερκτρηκόν κέρο από των βείηκλ-
των που βριγεηκεί για θεση η και $n+1$
όπου $n < α < n+1$.

Εάν οι παρατηρήσεις είναι ταξινομημένες και για κάθε X_i υπάρχει η συχνότητα εμφάνισής F_i , η θέση της διαμέσου είναι

$$\alpha = \frac{1}{2} \sum_{i=1}^v F_i = \frac{N}{2}$$

όπου N το πλήθος των παρατηρήσεων

Για τον υπολογισμό της διαμέσου απαιτείται η κατάταξη της σειράς των αθροισμών συχνοτήτων. Στη συνέχεια προσδιορίζουμε τις δύο συνεχόμενες αθροιστικές συχνότητες ανάμεσα στις οποίες βρίσκεται η τιμή του α και επιλέγουμε ως τιμή διαμέσου την τιμή του X_i που αντιστοιχεί στη μεγαλύτερη από αυτές.

Παράδειγμα

Κατάλογος ορθογωνίων τετραγώνων σε δρόμα γαϊδών

Πλήθος τετραγώνων X_i	Πλήθος γαϊδών F_i	Αθροιστική συχνότητα Φ_i
0	12	12
1	27	39
2	29	68
3	19	87
4	8	95
5	4	99
6	1	100

← $\frac{N}{2} = 50$

Παρατηρούμε πως ανάφεσα στο 39 και

(10)

στο 68 υπάρχει το 50, η τιμή που υπολογίζουμε

Η μεγαλύτερη τιμή από το 39 και το 68 είναι

η τιμή 68 που αντιστοιχεί σε $\chi_i = 2$ η

οποία αποτελεί και την τιμή της διαμέσου.

Επομένως το 50% των γαλακτωμάτων τουλάχιστον

2 λίτρα ενώ το υπόλοιπο 50% έχει υδρή

περιεχόμενο λιγότερο. Εάν η τιμή του $\frac{N}{2}$ ταυιστεί

επί αυτών με κάποια αθροιστική συχνότητα

Φ_i , η διάμεσος είναι το αντίστοιχο χ_i

Διάμεσος ομαλοποιημένων

παρατηρήσεων γνησίου δ

Υπολογίζεται με τον ακόλουθο τρόπο

- 1) Συμπληρώνουμε τη στήλη των αθροιστικών συχνοτήτων
- 2) προσδιορίζουμε τη θέση της διαμέσου $\frac{N}{2}$
- 3) προσδιορίζουμε δύο διαδοχικές αθροιστικές συχνοτήτες ανάφεσα στις οποίες περιέχεται το $\frac{N}{2}$
- 4) χαρτάσουμε για γραμμική ανάφεσα τους.
- 5) επιλέγουμε το χ_i της εγγράμμης πάνω από τη γραμμή
- 6) επιλέγουμε το F_i της εγγράμμης πάνω από τη γραμμή
- 7) επιλέγουμε το Φ_i της εγγράμμης πάνω από τη γραμμή

8) Υπολογίστε την διαταγή και την εκφραγή

$$M_{1/2} = X_i + \frac{\delta}{F_i} \left(\frac{N}{2} - \Phi_i \right)$$

9) ελεγχουμε εαν η τιμή που βρουμε κινηει σε οριζοντιο κατω αρα η πραξη

ΠΑΡΑΔΕΙΓΜΑ

Υατανόση τιθων φερων

<u>ΜΙΓΘΟΙ</u> <u>ΕΡΓΩΝ</u>	<u>ΠΙΝΘΟΣ</u> <u>ΕΡΓΩΝ</u>	<u>ΑΘΡΑΙΣΤΗ</u> <u>ΥΑΤΑΝΟΤΗ</u>
<u>X_i</u>	<u>F_i</u>	<u>Φ_i</u>
150-200	5	5
250-350	13	18
350-450	20	38
<u>450-550</u>	<u>35</u>	73
550-650	18	91
650-750	7	98
750-850	2	100

Φ_i
 $\frac{N}{2} = 50$

ΑΡΑ

$X_i = 450-550 \rightarrow 450$

$\delta = 100$

$M_{1/2} = 450 + \frac{100}{35} \left(\frac{100}{2} - 38 \right) = 484.29 \text{ €}$

Εαν $\frac{N}{2} < \Phi_i \Rightarrow \Phi_i = 0$

Σημείο μέγιστης συχνότητας

- Εκφράζει την τιμή της μεταβλητής Y τη μεγαλύτερη συχνότητα.
- Προσδιορίζεται Y απευθείας παρατήρηση του πίνακα
- Εάν οι παρατηρήσεις είναι ομαδοποιημένες σε ομάδες Y στο ίδιο πλαίσιο, εντορίζουμε την ομάδα Y τη μεγαλύτερη συχνότητα και τη θέση των F_i και στη συνέχεια υπολογίζουμε το σημείο μέγιστης συχνότητας ως

$$M_0 = X_i + \frac{f \Delta_1}{\Delta_1 + \Delta_2}$$

όπου X_i είναι το κάτω άκρο του διαστήματος που αντιστοιχεί στη μέγιστη συχνότητα, f είναι το εύρος αυτού του διαστήματος, Δ_1 είναι η διαφορά της συχνότητας της προηγούμενης ομάδας και τη μέγιστη συχνότητα και Δ_2 είναι η διαφορά της συχνότητας της επόμενης ομάδας και τη μέγιστη συχνότητα $\frac{\sum X}{N}$

150 - 250	5
250 - 350	13
350 - 450	20
450 - 550	35 ←
550 - 650	18
650 - 750	7
750 - 850	2

$$X_i = 450 \quad f = 100$$
$$\Delta_1 = 15 \quad \Delta_2 = 17$$

Αρα (Y
 αντίστα-
 σταθμ).

$$M_0 = 496.8$$

Γεωμετρικός μέσος διτρίνομων

παρατηρήσεων

$$\bar{G} = \sqrt[n]{X_1 X_2 \dots X_N} = \left(\prod_{i=1}^n X_i \right)^{1/n}$$

Εάν οι τιμές των παρατηρήσεων είναι μεγάλες και οι παρατηρήσεις πολλές, χρησιμοποιούμε λογαρίθμους

$$\log \bar{G} = \frac{1}{n} \sum_{i=1}^n \log(X_i)$$

Τεταρτημόρια (quartiles)

Προσκαλούν τον διαχωρισμό της κατανομής σε 4 ίσα τμήματα του 25% το καθένα

- 1^ο τεταρτημόριο → 25%
- 2^ο τεταρτημόριο → 50% (διάμεσος)
- 3^ο τεταρτημόριο → 75%

$$M_{1/4} = X_i + \frac{\delta}{f_i} \left(\frac{N}{4} - \Phi_i \right) \quad \begin{matrix} \text{πρώτο} \\ \text{τεταρτημόριο} \end{matrix}$$

$$M_{3/4} = X_i + \frac{\delta}{f_i} \left(\frac{3N}{4} - \Phi_i \right) \quad \begin{matrix} \text{τρίτο} \\ \text{τεταρτημόριο} \end{matrix}$$

Με παρόμοιο τρόπο ορίζονται τα δευτερά και τα εκατοβτημόρια

$$M_{d/10} = X_i + \frac{\delta}{F_i} \left(\frac{dN}{10} - \Phi_i \right) \quad d=1,2,\dots,10$$

(decile)

$$M_{p/100} = X_i + \frac{\delta}{F_i} \left(\frac{pN}{100} - \Phi_i \right) \quad p=1,2,3,\dots,100$$

(percentile)

Μέτρα διασποράς

Ευρος: η διαφορά της ελάχιστης από τη μεγαλύτερη τιμή

$$\max(X_i) - \min(X_i)$$

Ενδοτεταρτηφορικό ευρος: η διαφορά ανάμεσα στο τρίτο και στο πρώτο τεταρτηφορικό

$$Q = M_{3/4} - M_{1/4}$$

Μέση απόλυτη απόκλιση

Απόλυτων τετράγωνων διαφορών

$$MAA = \frac{1}{N} \sum |X_i - \bar{X}|$$

Τετράγωνων τετράγωνων διαφορών

$$MAA = \frac{\sum F_i |X_i - \bar{X}|}{\sum F_i}$$

Διακύμανση

$$\sigma^2 = \frac{L}{N} \sum_{i=1}^N (X_i - \mu)^2 = \frac{1}{N} \sum_{i=1}^N X_i^2 - \mu^2$$

Τυπική απόκλιση

$$\sigma = \sqrt{\frac{L}{N} \sum_{i=1}^N (X_i - \mu)^2} = \sqrt{\frac{L}{N} \sum_{i=1}^N X_i^2 - \mu^2}$$

Εάν χρησιμοποιήσουμε ως κλάδο πληθυσμού N αντιπροσώπευσε ως μ $\mu \in \mu$ ως \bar{X} ενώ για γύρω πληθυσμού διδρούσε $\mu \in \mu$ $N-1$, οχι $\in N$

Τέλος, όταν οι παρατηρήσεις είναι συαβουλομένες και υπάρχουν οι συχνοτήτες F_i έχουμε

$$\sigma^2 = \frac{\sum_{i=1}^v F_i (X_i - \mu)^2}{\sum_{i=1}^v F_i} = \frac{\sum_{i=1}^v F_i X_i^2}{\sum_{i=1}^v F_i} - \mu^2$$

και ομοίως για τα υπολοιπα.

Εάν όλα τα X_i είναι ίσα έχουμε $\sigma = 0$.

Κύρτωση (ροπή 4ης τάξης)

$$K = \frac{L}{N} \sum_{i=1}^v F_i (X_i - \bar{X})^4$$

Γενικές παρατηρήσεις

- Μέγος όρος

Εμπραξεται από όλες τις τιμές και ιδιαίτερα από τις άραιες τιμές, γεγονός που βεβαιώνει η φεικτώγη του υαλιού ανδριόλιου. Μπορεί να πάρει τιμή που δεν ταυτίζεται με κανένα από τα χί

- Διάχυτος: Δεν εμπραξεται από τις άραιες τιμές

- Σμείο μεγιστης ευχνοτητας:

Δεν εμπραξεται από τις άραιες τιμές. Σε πολύ άσυνεπες κτηνοφει είναι πιο αντιστορωπεντιού από το πέρο ορο