



Ποσοτικές Μέθοδοι Ανάλυσης στις Κοινωνικές Επιστήμες

Ενότητα 1: Εισαγωγή.

Θεόδωρος Χατζηπαντελής
Τμήμα Πολιτικών Επιστημών



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «Ανοικτά Ακαδημαϊκά Μαθήματα στο Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.





Έννοιες από την Στατιστική

Ορισμοί και Παραδείγματα



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΣΠΑ
2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Έννοιες από τη Στατιστική

- Πείραμα τύχης: Γνωρίζουμε εκ των προτέρων ένα σύνολο τιμών από το οποίο είναι δυνατόν να προέρχονται τα αποτελέσματα μετρήσεων για κάποιο χαρακτηριστικό γνώρισμα ή κάποια ιδιότητα ενός υποκειμένου (σε μέτρηση).
- Το σύνολο αυτό είναι το πεδίο ορισμού της μεταβλητής (όλα τα γεγονότα που μπορεί να συμβούν).



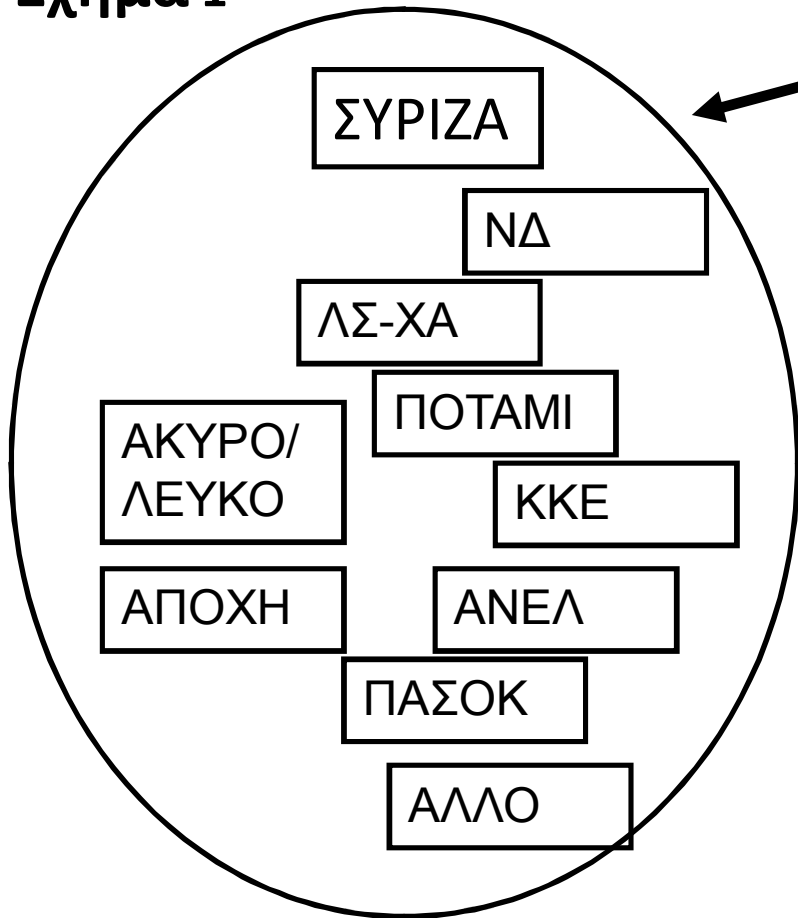
Συνέχεια Έννοιες από την Στατιστική

- Για κάθε μια τιμή του πεδίου ορισμού μπορεί να οριστεί ή να υπολογιστεί μια πιθανότητα εμφάνισης της τιμής αυτής. Τότε έχουμε μια συνάρτηση πιθανότητας της τυχαίας μεταβλητής. Και το σύνολο αυτό είναι το σύνολο τιμών της συνάρτησης πιθανότητας.



Συνάρτηση

Σχήμα 1

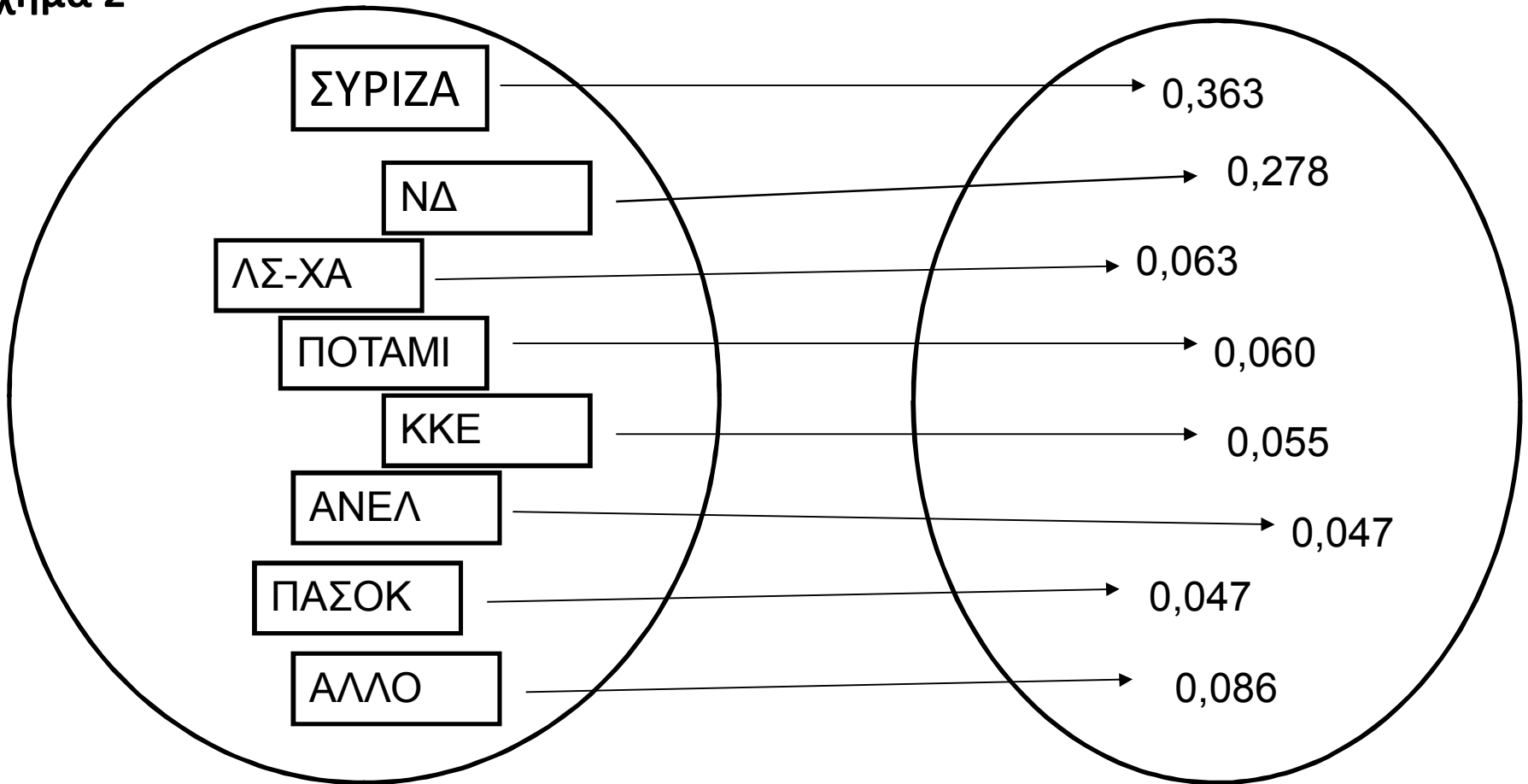


- **Πεδίο ορισμού:** είναι το σύνολο τιμών της μεταβλητής (δηλαδή το σύνολο όλων των δυνατών αποτελεσμάτων του πειράματος τύχης ή του φαινομένου).
- **Πεδίο Τιμών:** το σύνολο από το οποίο προέρχονται οι τιμές της συνάρτησης.



Συνάρτηση - έγκυρα

Σχήμα 2



Ορισμοί

- Μεταβλητή: είναι το χαρακτηριστικό γνώρισμα ή ιδιότητα.
- Πληθυσμός: το σύνολο των υποκειμένων (σε μέτρηση) ως προς τη μεταβλητή. Συνήθως επιλέγουμε ένα υποσύνολο των υποκειμένων που ονομάζουμε **δείγμα**.



Παράδειγμα

- Έστω X η τυχαία μεταβλητή που αντιστοιχεί στη στάση που θα ακολουθήσει κάποιος στις Βουλευτικές εκλογές 2007.
- Ο πληθυσμός είναι όλοι οι Έλληνες πολίτες ηλικίας 18+.
- Το πεδίο ορισμού είναι το σύνολο ενδεχομένων {ΝΔ, ΠΑΣΟΚ, ΚΚΕ, ΣΥΡΙΖΑ, ΛΑΟΣ, Άλλο, Άκυρο-Λευκό, Αποχή}.
- Το πεδίο τιμών της X είναι το πλήθος των ατόμων που είναι δυνατό να επιλέξουν κάθε μια από τις στάσεις. Κάθε περίπτωση αντιστοιχεί σε μια δαδα τιμών.



Συνέχεια Παραδείγματος

- Δεδομένου ότι οι εγγεγραμμένοι εκλογείς είναι $N=9918917$ πριν από τις εκλογές έχουμε τα εξής δυνατά αποτελέσματα (κόμμα, ψήφοι):
- **ΝΔ** ψ_1 , **ΠΑΣΟΚ** ψ_2 , **ΚΚΕ** ψ_3 , **ΣΥΡΙΖΑ** ψ_4 , **ΛΑΟΣ** ψ_5 , **Άλλο** ψ_6 , **Άκυρο-Λευκό** ψ_7 , **Αποχή** ψ_8
- Ισχύει ότι: $\psi_1 + \psi_2 + \psi_3 + \psi_4 + \psi_5 + \psi_6 + \psi_7 + \psi_8 = N$ για κάθε 8αδα πιθανού αποτελέσματος.
- Με τις εκλογές από όλες τις πιθανές 8αδες προκύπτει ένα αποτέλεσμα **{2994979, 2727279, 583750, 361101, 271809, 220088, 196020, 2563891}**.



Πιθανότητα κάθε Γεγονότος

- Μετά τις εκλογές μπορούμε να υπολογίσουμε την πιθανότητα κάθε γεγονόςτος (από το πεδίο ορισμού).

Πίνακας 1

Πεδίο ορισμού	Τιμή	Πιθανότητα
ΝΔ	2.994.979	$P(X=NΔ)=0,302$
ΠΑΣΟΚ	2.727.279	0,275
ΚΚΕ	583.750	0,059
ΣΥΡΙΖΑ	361.101	0,036
ΛΑΟΣ	271.809	0,027
ΆΛΛΟ	220.088	0,022
Άκυρο-Λευκό	196.020	0,020
Αποχή	2.563.891	0,259



Μετά τις εκλογές

- Μετά τις εκλογές μπορούμε να υπολογίσουμε την πιθανότητα να έχει ακολουθήσει κάθε στάση κάποιος εγγεγραμμένος.
- Ο υπολογισμός αυτός μπορεί να μας βοηθήσει σε μια επόμενη έρευνα.
- Ο υπολογισμός αυτός δεν μπορεί να γίνει – πριν από τις εκλογές- χωρίς κάποια στατιστική διαδικασία (μια έρευνα γνώμης).



Αντιστοιχία

- Ενώ χωρίς κάποια στατιστική διαδικασία μπορούμε να υπολογίσουμε το παρακάτω:
- Αν ρίξουμε ένα νόμισμα μία φορά το αποτέλεσμα (πεδίο ορισμού) θα είναι το $\{Κ, Γ\}$. Ξέρουμε ότι η $P(Κ)=1/2$ και η $P(Γ)=1/2$.
- Αν επαναλάβουμε το πείραμα περισσότερες (πχ 3) φορές έχουμε το πεδίο ορισμού είναι:

$\{ΓΓΓ, ΚΓΓ, ΓΚΓ, ΓΓΚ, ΓΚΚ, ΚΓΚ, ΚΚΓ, ΓΓΓ\}$.

Αν μετράμε τον αριθμό εμφανίσεων $Γ$ τότε το σύνολο τιμών είναι $\{3, 2, 1, 0\}$.

Και οι αντίστοιχες πιθανότητες $\{1/8, 3/8, 3/8, 1/8\}$.



Στην πράξη

- Μας ενδιαφέρει η αντιστοιχία από το σύνολο ορισμού, δηλαδή το σύνολο των τιμών που μπορεί να πάρει κάποια μεταβλητή στο σύνολο πιθανοτήτων για κάθε μια από τις τιμές.
- Θέλουμε δηλαδή να υπολογίσουμε πόσο πιθανή είναι κάθε δαδα τιμών στο παράδειγμα με τις εκλογές και κατά συνέπεια πόσο πιθανό είναι κάθε ενδεχόμενο.



Αν γνωρίζουμε

- Τότε με υπολογισμούς εκ των προτέρων μπορούμε να βρούμε τις πιθανότητες αυτές. Όπως στο νόμισμα.



Αν δεν γνωρίζουμε

- Με υπολογισμούς εκ των υστέρων –μετά από μια δειγματοληπτική έρευνα- μπορούμε να προσεγγίσουμε τις πιθανότητες. Όπως στις προεκλογικές έρευνες.
- Μπορούμε να υπολογίσουμε τις πιθανότητες αυτές εκ των υστέρων (δηλαδή: μετά από την επανάληψη ενός πειράματος τύχης).
- Μια δειγματοληπτική έρευνα είναι προσομοίωση ενός πειράματος τύχης.



Συχνότητα

- Κατά την υλοποίηση ενός πειράματος τύχης καταγράφουμε **ενδεχόμενα** δηλαδή στοιχεία από το σύνολο ορισμού. Για να υπολογίσουμε την πιθανότητα εμφάνισης κάθε ενδεχομένου (που το ονομάζουμε τιμή της μεταβλητής) υπολογίζουμε τη **συχνότητα εμφάνισης του** στο δείγμα τιμών που έχουμε καταγράψει.



Συχνότητα

- Έτσι η **σχετική συχνότητα** θεωρείται ότι προσεγγίζει την **πιθανότητα**.
- Προσέξτε ότι χρησιμοποιούμε τα ενδεχόμενα (αποτελέσματα) ενός πειράματος κατά μία επανάληψη του για να χαρακτηρίσουμε τη μεταβλητή που μετράμε. Το σύνολο τιμών της όπως το περιγράψαμε (όλες οι δυνατές δαδες τιμών) έχει μικρή αξία για την εφαρμογή.



Δειγματοληπτική έρευνα

- Όταν για παράδειγμα ρωτάμε κάποιον «ποια στάση κράτησε στις προηγούμενες βουλευτικές εκλογές» καταγράφουμε κάποια τιμή από το σύνολο {**ΝΔ, ΠΑΣΟΚ, ΚΚΕ, ΣΥΡΙΖΑ, ΛΑΟΣ, Άλλο, Άκυρο-Λευκό, Αποχή**}.
- Στο παράδειγμα αυτό γνωρίζουμε τις πιθανότητες κάθε τιμής. Το πείραμα δεν γίνεται για να εκτιμήσουμε κάτι που ξέρουμε. Αλλά αυτή η ερώτηση μπορεί να μας βοηθήσει για να διαπιστώσουμε αν το δείγμα αυτό αντιπροσωπεύει τον πληθυσμό ως προς το χαρακτηριστικό αυτό (στάση στις Β2007).



Δειγματοληπτική έρευνα

- Αν δεν γνωρίζουμε; Πχ θέλουμε να προσεγγίσουμε τη μεταβλητή {**στάση στις επόμενες βουλευτικές εκλογές**}. Τότε μέσω του πειράματος τύχης εκτιμούμε τις πιθανότητες εμφάνισης κάθε τιμής με τη χρήση των συχνοτήτων.



Μεταβλητές

- Κάθε μεταβλητή καθορίζεται από το σύνολο ενδεχομένων. Έτσι έχουμε (ως προς την κλίμακα μέτρησης):
- Ποσοτικές μεταβλητές: **τα ενδεχόμενα είναι αριθμητικές τιμές.** Όπως το ύψος, το βάρος, η απόσταση, ο αριθμός των Γ όταν ρίχνουμε Ν φορές ένα νόμισμα
- Ποιοτικές μεταβλητές: **τα ενδεχόμενα δεν είναι αριθμητικές τιμές.** Όπως η στάση στις εκλογές, το φύλο, το χρώμα των ματιών, ο τόπος γέννησης, ο βαθμός συμφωνίας σε μια πρόταση, το εξάμηνο σπουδών.



Ποιοτικές μεταβλητές

- Επιπλέον οι ποιοτικές μεταβλητές διακρίνονται σε μεταβλητές διάταξης και κατηγορίας (ή ονομαστικές).
- Μεταβλητές διάταξης είναι μεταβλητές που οι τιμές της μεταβλητής (τα ενδεχόμενα) μπορεί να τοποθετηθούν στη σειρά χωρίς τη χρήση κάποιου κριτηρίου. Αν ρωτήσουμε «πόσο συμφωνείτε με ...» και οι τιμές είναι {**απόλυτα, αρκετά, σχετικά, λίγο, καθόλου**} το σύνολο αυτό μπορεί να διαταχθεί (δηλαδή οι τιμές του να μπουν στη σειρά).



Ποιοτικές Μεταβλητές

- Μεταβλητές κατηγορίας ή ονομαστικές είναι εκείνες που οι τιμές τους δεν είναι δυνατό να διαταχθούν (να μπουν στη σειρά). Όπως ή μεταβλητή **στάση στις Β2007** ή η μεταβλητή **χρώμα ματιών**. Προσέξτε ότι αναφερόμαστε στις τιμές και όχι στη συχνότητα εμφάνισης τους.



Η σύγκριση τιμής και συχνότητας

- Ενώ μετά τις εκλογές γνωρίζουμε ποιο κόμμα ήρθε πρώτο, ποιο δεύτερο κλπ η διάταξη αναφέρεται στις ιδιότητες των τιμών πριν από οποιαδήποτε στατιστική ή άλλη διαδικασία. Δηλαδή στο σύνολο {**ΝΔ, ΠΑΣΟΚ, ΚΚΕ, ΣΥΡΙΖΑ, ΛΑΟΣ, Άλλο, Άκυρο-Λευκό, Αποχή**} – το πεδίο ορισμού της μεταβλητής. Οι τιμές αυτές δεν μπορεί να μπουν στη σειρά .



Συλλογή δεδομένων

- Από ένα **πληθυσμό** σχηματίζουμε ένα **δείγμα** υποκειμένων, κάνουμε μετρήσεις **χαρακτηριστικών** (μεταβλητών) και σχηματίζουμε ένα δείγμα **τιμών**.
- Η συλλογή αυτή γίνεται έτσι ώστε το δείγμα υποκειμένων και είναι αντιπροσωπευτικό του πληθυσμού (ως προς το χαρακτηριστικό ή τα χαρακτηριστικά) που μετράμε.



Συλλογή δεδομένων

- Προκύπτει ένα σύνολο δεδομένων που τοποθετούμε συνήθως σε ένα πίνακα δύο διαστάσεων ως εξής:
- Ο πίνακας έχει τόσες γραμμές όσα τα υποκείμενα. Αν έχουμε ρωτήσει N άτομα τότε ο πίνακας έχει N γραμμές (+1 για τις επικεφαλίδες ονόματα μεταβλητών)
- Ο πίνακας έχει τόσες στήλες όσα τα χαρακτηριστικά-μεταβλητές. Αν έχουμε καταγράψει k χαρακτηριστικά έχει k στήλες (+1 για τον αύξοντα αριθμό κάθε περίπτωσης).



Πίνακας δεδομένων 1

- Αν έχουμε ρωτήσει 100 άτομα 5 ερωτήσεις και έχουμε καταγράψει και άλλα 2 χαρακτηριστικά έχει προκύψει ένας πίνακας με $N+1$ ($100+1$) γραμμές και $k+1$ ($7+1$) στήλες.
- Στην πρώτη στήλη καταγράφουμε τον αύξοντα αριθμό του ατόμου (ή κάποιο άλλο κωδικό που είναι μοναδικός για το άτομο αυτό πχ ΑΔΤ ή ΑΦΜ ή ΑΕΜ).



Πίνακας δεδομένων 2

- Στη πρώτη γραμμή γράφουμε τα ονόματα των χαρακτηριστικών. Αν πχ έχουμε καταγράψει το φύλο και την ηλικία και έχουμε κάνει 5 ερωτήσεις η πρώτη γραμμή θα έχει τα εξής:
- [ΑΑ, ΦΥΛΟ, ΗΛΙΚΙΑ, ΕΡ1, ΕΡ2, ΕΡ3, ΕΡ4, ΕΡ5].
Προσέξτε την επικεφαλίδα ΑΑ στην πρώτη στήλη.



Πίνακας δεδομένων 3

- Έτσι κάθε στοιχείο x_{ij} (κελί) του πίνακα δεδομένων περιγράφεται με τρία στοιχεία
- Το x , το i και το j .
- Το i είναι δείκτης της γραμμής στην οποία ανήκει το κελί.
- Το j είναι δείκτης της στήλης στην οποία ανήκει το κελί.
- Το x_{ij} είναι η τιμή της j μεταβλητής για το i υποκείμενο.



Πίνακας δεδομένων

Πίνακας 2: Πίνακας Δεδομένων

ΑΑ	ΦΥΛΟ	ΗΛΙΚΙΑ	ΕΡ1	ΕΡ2	ΕΡ3	ΕΡ4	ΕΡ5
1	Α	25	ΝΑΙ	ΑΡΚΕΤΑ	ΛΙΓΟ	ΚΟΚΚΙΝΟ	800
2	Γ	45	ΌΧΙ	ΛΙΓΟ	ΑΡΚΕΤΑ	ΜΠΛΕ	500
3	Γ	23	ΌΧΙ	ΚΑΘΟΛΟΥ	ΚΑΘΟΛΟΥ	ΑΣΠΡΟ	1100
.
.
.
100	Γ	57	ΟΧΙ	ΑΡΚΕΤΑ	ΚΑΘΟΛΟΥ	ΜΑΥΡΟ	400



Πεδίο Ορισμού

- Για κάθε μεταβλητή ΠΡΙΝ από την δειγματοληψία γνωρίζουμε το πεδίο ορισμού:
- Φύλο {Α, Γ}.
- Ηλικία {συμπληρωμένα χρόνια}.
- ΕΡ1 {ΝΑΙ, ΌΧΙ}.
- ΕΡ2, ΕΡ3 {Απόλυτα, Αρκετά, Λίγο, Καθόλου}.
- ΕΡ4 {αγαπημένο χρώμα}.
- ΕΡ5 {ποσό που ξοδεύει ανά μήνα για διασκέδαση}.



Έτσι για κάθε μεταβλητή έχουμε

- Φύλο - Ονομαστική (δίτιμη).
- Ηλικία - Ποσοτική.
- EP1 - Ονομαστική (δίτιμη).
- EP2 - Διάταξης.
- EP3 - Διάταξης.
- EP4 - Ονομαστική.
- EP5 - Ποσοτική.



Ανάλυση

- Μία μεταβλητή.
- Σχέση δύο μεταβλητών.
- Σχέση πολλών (τριών ή περισσότερων) μεταβλητών.



Ανάλυση (μία μεταβλητή)

- Θέλουμε να εκτιμήσουμε παραμέτρους του πληθυσμού με τη χρήση στατιστικών που υπολογίζουμε από το δείγμα. Ανάλογα με την κλίμακα μέτρησης της μεταβλητής έχουμε για τις παραμέτρους του πληθυσμού:



Μια εφαρμογή

Ο υπεύθυνος επιθεώρησης μίας τράπεζας υποστηρίζει ότι οι πελάτες που ασχολούνται στη βιοτεχνία έχουν καταθέσεις με **μέση τιμή 3000** ευρώ και **τυπική απόκλιση 300** ευρώ. Αν η τράπεζα έχει **1250 πελάτες** που ασχολούνται στην βιοτεχνία.

1. Να εκτιμήσετε τον αριθμό των πελατών που θα έχουν καταθέσεις από 2400 έως 3600 ευρώ.
2. Η τράπεζα αποφάσισε να στείλει μία ευχετήρια κάρτα σε όλους τους πελάτες που ασχολούνται στη βιοτεχνία και έχουν καταθέσεις από 2100 έως 3900 ευρώ. Ποιος είναι ο αριθμός καρτών που αναμένεται να στείλει η τράπεζα;



α.1

Σύμφωνα με το Θεώρημα του Chebyshev σε απόσταση k (δύο) τυπικών αποκλίσεων από τον μέσο (δηλαδή στο διάστημα από 2400 έως 3600 ευρώ) βρίσκεται τουλάχιστον το

$$\left(1 - \frac{1}{k^2}\right) = \left(1 - \frac{1}{2^2}\right) = 0,75$$

των πελατών.

Έτσι εκτιμάται ότι τουλάχιστον **75%** πελάτες θα έχουν από 2400 έως 3600 ευρώ, δηλαδή τουλάχιστον **938** πελάτες της τράπεζας.



α.2

- Κάνοντας και πάλι χρήση του **Θεωρήματος του Chebyshev** σε απόσταση τριών τυπικών αποκλίσεων από τον μέσο (δηλαδή στο διάστημα από 2100 έως 3900 ευρώ) βρίσκεται τουλάχιστον το 0,89 των πελατών.
- Έτσι εκτιμάται ότι τουλάχιστον 89% πελάτες θα έχουν από 2100 έως 3900 ευρώ, δηλαδή τουλάχιστον 1113 πελάτες της τράπεζας. Άρα η τράπεζα αναμένεται να στείλει τουλάχιστον 1113 ευχετήριες κάρτες.



Συνέχεια

- Ο διευθυντής ενός υποκαταστήματος της τράπεζας στην προσπάθειά του να ελέγξει αν τα στοιχεία που αναφέρονται παραπάνω και που αφορούν ολόκληρη την τράπεζα ευσταθούν στην περίπτωση του δικού του υποκαταστήματος, συλλέγει ένα **τυχαίο δείγμα από 60 πελάτες** του υποκαταστήματος και καταγράφει το ποσό των καταθέσεών τους (σε ευρώ, στρογγυλοποιημένο στην πλησιέστερη δεκάδα). Ο παρακάτω πίνακας περιέχει τα δεδομένα αυτά.



Δείγμα Πελατών

Πίνακας 3: Δείγμα Πελατών.

2530	2420	2870	3520	3030	3490	3110	2970	2930	2820
2870	2940	2150	3380	3010	2200	3120	2780	2720	4040
2330	2460	2550	3170	3020	3290	2980	3310	3240	3290
3230	2940	2650	3070	2810	2850	3080	2740	3560	3310
2240	2610	2980	4240	2710	2980	3740	3720	4240	4160
3830	2970	3490	2910	3380	3990	4110	2490	3870	3790



Επίσης

- Να υπολογιστούν ο αριθμητικός μέσος, το εύρος, το ενδοτεταρτημοριακό εύρος, η τυπική απόκλιση των δεδομένων καθώς και ο συντελεστής ασυμμετρίας (β_3) ο οποίος και να ερμηνευθεί. (Οι υπολογισμοί να γίνουν χρησιμοποιώντας και τις αντίστοιχες συναρτήσεις του Excel).



β

- Αριθμητικός μέσος.
- Τυπική απόκλιση.
- Εύρος.
- Τεταρτημόρια.
- Συντελεστής ασυμμετρίας.



Υπολογισμοί

- Μέσος = άθροισμα/ πλήθος.
- Εύρος = μεγαλύτερη - μικρότερη.
- Τεταρτημόρια: χωρίζω σε τέταρτα το σύνολο. Άρα πρέπει να βρω τρία σημεία: το 25%, το 50%, το 75%. Τύπος: $i * (n+1)/4$.

**$1 * (60+1)/4 = 15,25$. Άρα είναι στη θέση 15,25.
Τα βάζω στη σειρά.**



Υπολογισμοί 1

- Τυπική απόκλιση. Υπολογίζω τη διασπορά. Αφαιρώ από κάθε παρατήρηση το μέσο όρο, υψώνω στο τετράγωνο, προσθέτω και μετά διαιρώ με το $n-1$. Βρίσκω την τετραγωνική ρίζα.



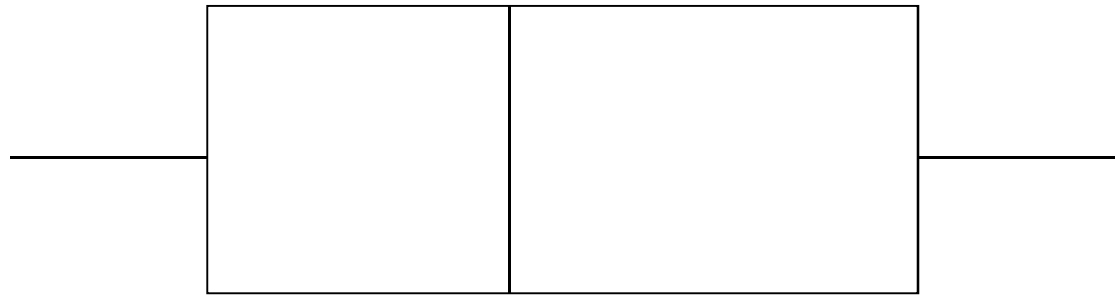
Υπολογισμοί 2

- **Συντελεστής ασυμμετρίας 1:** Αφαιρώ από το μέσο όρο την επικρατούσα τιμή και διαιρώ με την τυπική απόκλιση.
- **Συντελεστής ασυμμετρίας 3:** Αφαιρώ από κάθε παρατήρηση το μέσο όρο, υψώνω στην τρίτη δύναμη, προσθέτω και μετά διαιρώ με το n . Διαιρώ με την τρίτη δύναμη της τυπικής απόκλισης.



1.γ Box-plot 1

Min Q1 M Q3 max



2150 2787,5 3015 3462,5 4240



1.δ

Στην περίπτωση που η πολιτεία προχωρήσει σε επιβολή φορολογίας των καταθέσεων η οποία **ανέρχεται στο 10% του ύψους των καταθέσεων**, καθώς επίσης και στην επιβολή **πάγιου τέλους που ανέρχεται σε 50 ευρώ**, να υπολογισθεί ο αριθμητικός μέσος, η διάμεσος, το εύρος και η τυπική απόκλιση του ύψους των καταθέσεων στο δείγμα μετά την επιβολή της φορολογίας και του πάγιου τέλους. (Οι υπολογισμοί να γίνουν με τη χρήση των ιδιοτήτων των αντιστοιχών μέτρων).



Ιδιότητες

- Για μέσο όρο και διάμεσο: θα μου μείνουν 90% και από το καθένα πρέπει να αφαιρέσω 50.
- Για εύρος και τυπική απόκλιση: θα μου μείνουν 90%.



1.ε

Να κατασκευασθεί Πίνακας Κατανομής Συχνοτήτων για το ποσό των καταθέσεων του υποκαταστήματος της τράπεζας χρησιμοποιώντας τάξεις εύρους 500 ευρώ, με κεντρική τιμή της πρώτης τάξης τα 2000 ευρώ. (Να δοθεί η λύση **και** με το Excel, χρησιμοποιώντας την κατάλληλη συνάρτηση).



1.στ

- Χρησιμοποιώντας τα (ομαδοποιημένα) ταξινομημένα δεδομένα να υπολογιστούν ο αριθμητικός μέσος, η επικρατούσα τιμή, το πρώτο τεταρτημόριο και η διακύμανση του ποσού καταθέσεων. (Να δημιουργηθεί στο Excel κατάλληλος Πίνακας στον οποίο να γίνουν οι προαπαιτούμενοι υπολογισμοί).



1.ζ

- Σε ποια από τις δύο περιπτώσεις (προ φόρου και μετά φόρου) μπορεί να παρατηρηθεί μεγαλύτερη μεταβλητότητα σε ότι αφορά το ποσό των καταθέσεων;
- Συντελεστής μεταβλητότητας: **τυπική απόκλιση δια μέσου όρου (%)**.



Ανάλυση (μια μεταβλητή)

Πίνακας 4.1 : Ανάλυση μιας μεταβλητής.

Κλίμακα	ΠΛΗΘΥΣΜΟΣ (παράμετρος)	ΔΕΙΓΜΑ (στατιστικό μέτρο)
Ονομαστική	Μέση τιμή	Επικρατούσα τιμή
Διάταξης	Μέση τιμή	Διάμεσος
Ποσοτική	Μέση τιμή	Μέσος όρος
Ονομαστική	Διασπορά	Αριθμός τιμών
Διάταξης	Διασπορά	Αριθμός τιμών
Ποσοτική	Διασπορά	Διασπορά



Ανάλυση (μια μεταβλητή)

Πίνακας 4.2 : Ανάλυση μιας μεταβλητής.

Κλίμακα	ΠΛΗΘΥΣΜΟΣ (παράμετρος)	ΔΕΙΓΜΑ (στατιστικό μέτρο)
Ονομαστική	Κατανομή	Συχνότητες
Διάταξης	Κατανομή	Συχνότητες, εκατοστιαία σημεία
Ποσοτική	Κατανομή	Συχνότητες ομαδοποιημένων δεδομένων



Ανάλυση (μια μεταβλητή)

Πίνακας 4.3 : Ανάλυση μιας μεταβλητής.

Κλίμακα	ΠΛΗΘΥΣΜΟΣ (παράμετρος)	ΔΕΙΓΜΑ (στατιστικό μέτρο)
Ονομαστική	Κατανομή	Κυκλικό διάγραμμα
Διάταξης	Κατανομή	Ραβδόγραμμα
Ποσοτική	Κατανομή	Ιστόγραμμα



Σχέση δύο μεταβλητών

Πίνακας 5 : Σχέση δύο μεταβλητών.

Ποσοτική	Ποσοτική	Γραμμική παλινδρόμηση
Ποσοτική	Διάταξης	Συντελεστής Spearman
Ποσοτική (μετά από ομαδοποίηση)	Ονομαστική	Τεστ χ^2
Διάταξης	Διάταξης	Συντελεστής Spearman
Διάταξης	Ονομαστική	Τεστ χ^2
Ονομαστική	Ονομαστική	Τεστ χ^2
Ποσοτική	Ποσοτική	Γραμμική παλινδρόμηση



Επιπλέον χρήσιμα μέτρα

- Διάστημα εμπιστοσύνης.
- Ανισότητα Chebyshev.
- BOX-PLOT.
- Αθροιστική συχνότητα.



Πολλές μεταβλητές

- Ομαδοποίηση υποκειμένων (Cluster (Κεφάλαιο 2, Multidimensional scaling Κεφάλαιο 3).
- Ομαδοποίηση μεταβλητών (Principal component analysis Κεφάλαιο 5, Factor Κεφάλαιο 6).
- Από κοινού ομαδοποίηση υποκειμένων και μεταβλητών (Correspondence Analysis Κεφάλαιο 4).



Πολλές μεταβλητές

- Η σημασία της τιμής {**δεν ξέρω / δεν απαντώ**}
- Η σημασία της κλίμακας μέτρησης – η **κατάλληλη μέθοδος για τα κατάλληλα δεδομένα.**
- Ποσοτικές μέθοδοι **δεν είναι μόνο** οι στατιστικές μέθοδοι.
- Διάκριση για την ανάλυση **πληθυσμού-δείγματος. Απογραφή-δειγματοληψία.**



Σημείωμα Χρήσης Έργων Τρίτων (1/2)

- Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:
- Εικόνες/Σχήματα/Διαγράμματα/Φωτογραφίες
- Σχήμα 1-2: Παραδείγματα.



Σημείωμα Χρήσης Έργων Τρίτων (2/2)

- Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:
- Πίνακες
- Πίνακας 1-3: Παραδείγματα.
- Πίνακας 4-5: Θεωρία.



Σημείωμα Αναφοράς

Copyright Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Θεόδωρος Χατζηπαντελής. «Ποσοτικές Μέθοδοι Ανάλυσης στις Κοινωνικές Επιστήμες. Εισαγωγή». Έκδοση: 1.0. Θεσσαλονίκη 2014. Διαθέσιμο από τη δικτυακή διεύθυνση: <http://eclass.auth.gr/courses/OCRS309/>.



Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά - Μη Εμπορική Χρήση - Όχι Παράγωγα Έργα 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.

Ως **Μη Εμπορική** ορίζεται η χρήση:

- που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
- που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
- που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο

[1] <http://creativecommons.org/licenses/by-nc-nd/4.0/>





Τέλος ενότητας

Επεξεργασία: Σωτήρογλου Μαρίνα
Θεσσαλονίκη, Εαρινό Εξάμηνο 2014-2015



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ



ΑΡΙΣΤΟΤΕΛΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΟΝΙΚΗΣ

Σημειώματα

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.

