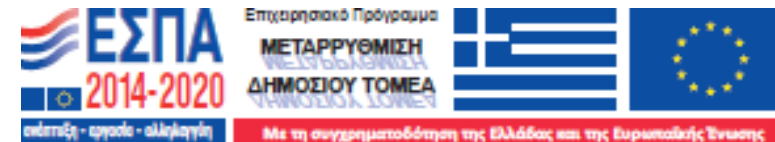




ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΥΠΟΥΡΓΕΙΟ ΕΣΩΤΕΡΙΚΩΝ

ΕΘΝΙΚΗ ΣΧΟΛΗ ΔΗΜΟΣΙΑΣ ΔΙΟΙΚΗΣΗΣ & ΑΥΤΟΔΙΟΙΚΗΣΗΣ



Η Διαχείριση Δεδομένων και η Στατιστική Επεξεργασία τους

Συντονιστής Δημήτριος Τσιμάρας
Αξιολογητής Ηλίας Μαραγκός
Συγγραφέας Κωνσταντίνα Ατσάρου

Αναλυτική περιγραφή

1. Εισαγωγή σε βασικές έννοιες της Στατιστικής
2. Δειγματοληψία
3. Περιγραφική Στατιστική
 - Πίνακες Συχνοτήτων και Βασικά στατιστικά μέτρα για την περιγραφή των δεδομένων
 - Γραφήματα για την απεικόνιση των δεδομένων
4. Βασικές έννοιες της Επαγωγικής Στατιστικής και του Ελέγχου Υποθέσεων
 - Συσχέτιση μεταβλητών και Παλινδρόμηση.

Μαθησιακοί Στόχοι

Όταν θα έχετε ολοκληρώσει την μελέτη της ενότητας θα είστε σε θέση να

- ▷ Ερμηνεύετε και να εφαρμόζετε στατιστικές έννοιες και μεθόδους
- ▷ Να αναγνωρίζετε την σημασία της στατιστικής ανάλυσης στην λήψη αποφάσεων
- ▷ Να υπολογίζετε και να ερμηνεύετε διάφορα στατιστικά μέτρα
- ▷ Να εφαρμόζετε τις αποκτηθείσες γνώσεις στην Δημόσια Διοίκηση

1. Εισαγωγή στην Στατιστική

1. Βασικές Στατιστικές έννοιες
2. Συλλογή δεδομένων
3. Τεχνικές δειγματοληψίας

Μαθησιακοί στόχοι

Όταν θα έχετε ολοκληρώσει την μελέτη της ενότητας

- ▷ Θα έχετε εξοικειωθεί με βασικές στατιστικές έννοιες
- ▷ Θα έχετε κατανοήσει τις διάφορες μεθόδους συλλογής δεδομένων
- ▷ Θα έχετε κατανοήσει τις διάφορες μεθόδους δειγματοληψίας

Στατιστική και εφαρμογές

Η Στατιστική είναι η επιστήμη που ασχολείται με τον **σχεδιασμό μιας έρευνας** και την συλλογή, την επεξεργασία, την παρουσίαση και την ανάλυση στοιχείων με στόχο την εξαγωγή συμπερασμάτων.

Βασικό αντικείμενο

- ✓ Συλλογή δεδομένων
- ✓ Οργάνωση δεδομένων
- ✓ Ομαδοποίηση δεδομένων
- ✓ Ανάλυση δεδομένων
- ✓ Ερμηνεία δεδομένων
- ✓ Εξαγωγή συμπερασμάτων



Εφαρμογές

- ✓ Ιατρική
- ✓ Διοίκηση
- ✓ Οικονομία
- ✓ Βιομηχανία

Παραδείγματα εφαρμογών Στατιστικής στον Δημόσιο τομέα

- ✓ Ανάλυση και πρόβλεψη οικονομικών στοιχείων (π.χ. Εκτίμηση ύψους εσόδων από επιβολή νέου φόρου ή ύψους δημοσίων εξόδων λόγω των επικείμενων συνταξιοδοτήσεων)
- ✓ Ανάλυση γεωγραφικών δεδομένων και διαμόρφωση πλαισίων ενίσχυσης της επιχειρηματικότητας.

Κλάδοι της Στατιστικής

Δειγματοληψία

Περιγραφική Στατιστική

Ασχολείται με την:

- ✓ Συνοπτική παρουσίαση δεδομένων μέσω συγκεντρωτικών πινάκων, στατιστικών μεγεθών ή διαγραμμάτων.
- ✓ Στατιστική Επεξεργασία πχ, Ομαδοποίηση, Ανακωδικοποίηση, Ταξινόμηση δεδομένων, διαχείριση ελλειπουσών τιμών, Επικύρωση Δεδομένων, έλεγχο εγκυρότητας κλπ.
- ✓ Εξασφάλιση προϋποθέσεων για την ανάλυση δεδομένων.
- ✓ Διερεύνηση περιγραφικών σχέσεων

Επαγωγική Στατιστική

Ασχολείται με:

- ✓ Την εκτίμηση των τιμών παραμέτρων του πληθυσμού
- ✓ Τον έλεγχο υποθέσεων για τις παραμέτρους του πληθυσμού
- ✓ Την πρόβλεψη των τιμών παραμέτρων του πληθυσμού
- ✓ Τη διερεύνηση σχέσεων/εξαρτήσεων/αλληλεπιδράσεων μεταξύ μεταβλητών.

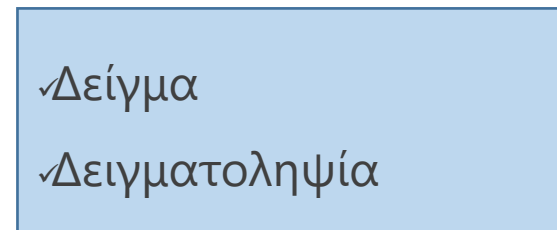
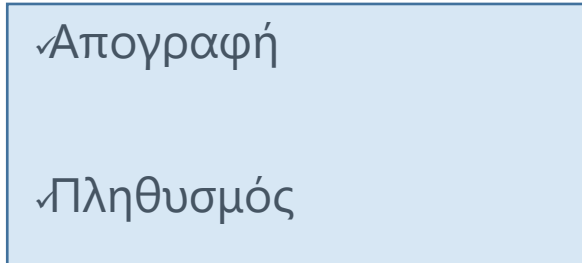
Βασικές Στατιστικές έννοιες

✓Απλό στοιχείο είναι το άτομο, νοικοκυριό, φορέας, πράξη κλπ. δηλαδή η οντότητα για την οποία θέλουμε να μετρήσουμε ή απαριθμήσουμε κάποιο **Χαρακτηριστικό ή Ιδιότητα**. Μπορεί να είναι **Υποκείμενο, αντικείμενο, έννοια, γεγονός και γενικά οντότητα**

✓**Δειγματοληπτική Μονάδα**

Τιμές χαρακτηριστικού -Μεταβλητή

Δεδομένα



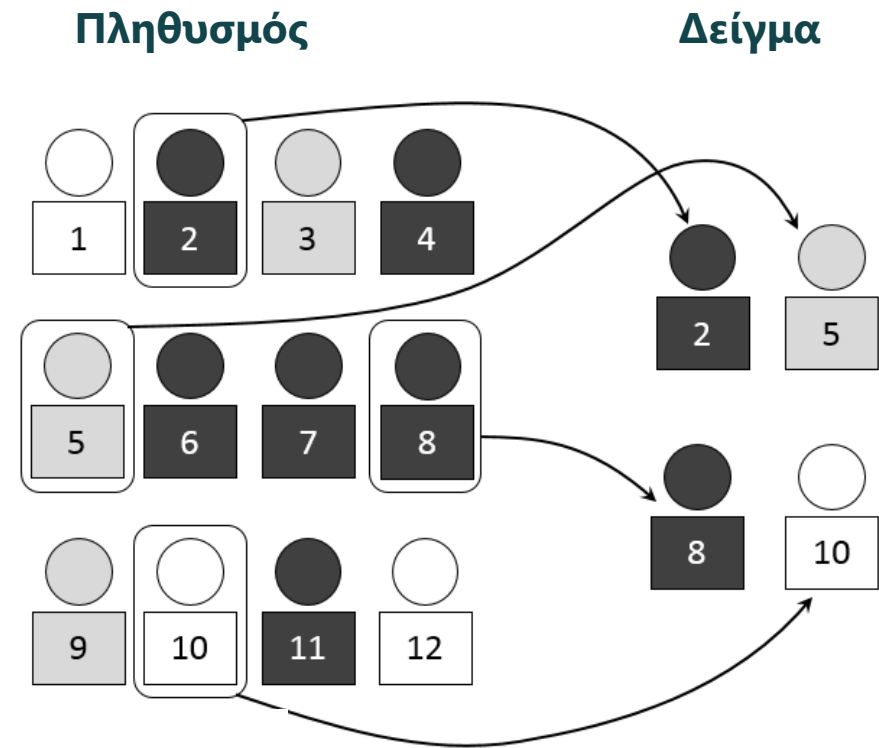
Πληθυσμός- Δείγμα Πληθυσμού

Πληθυσμός

Είναι **το σύνολο των μονάδων** που μας ενδιαφέρει να μελετήσουμε κάποια χαρακτηριστικά του.

Δείγμα

Ένα **υποσύνολο** του πληθυσμού. Είναι ένα μέρος του πληθυσμού που επιλέγεται προκειμένου να εκτιμήσουμε από αυτό τα χαρακτηριστικά του πληθυσμού

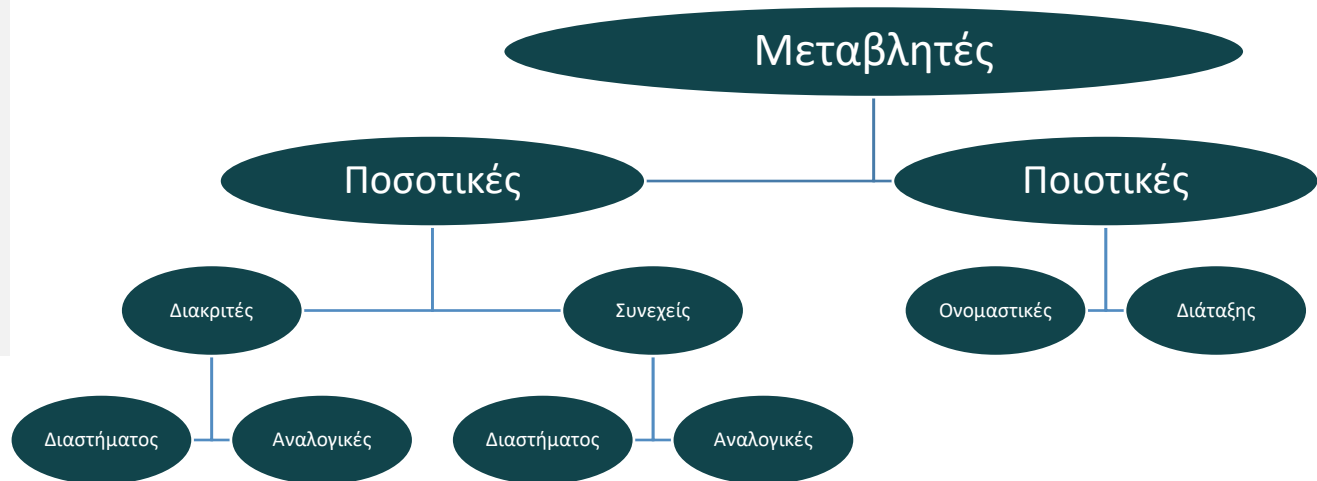


Μεταβλητή- Δεδομένα

Μεταβλητές ονομάζουμε τις αναπαράστασεις των διάφορων χαρακτηριστικών των μονάδων ενός πληθυσμού ή ενός δείγματος. Με βάση τον τρόπο μέτρησης τους διακρίνονται σε:

- ο Ποσοτικές
- ο Ποιοτικές

Δεδομένα είναι τα αποτελέσματα της μέτρησης κάποιου χαρακτηριστικού των μονάδων μιας έρευνας



ΕΚΠΑΙΔΕΥΤΙΚΟ ΕΠΙΠΕΔΟ ISCED

ΔΕ

ΔΕΙ/ΑΤΕΙ

ΥΕ

ΜΕΤΑΠΤΥΧΙΑΚΟ

ΔΙΔΑΚΤΟΡΙΚΟ

Ονομαστικές Καταμέτρηση/ Επικρατούσα Τιμή

Διατάξιμες Ιεράρχηση

Ποσοτικές

Διαστήματος αφαίρεση /πρόσθεση

Αναλογικές όλες οι πράξεις

Άσκηση

Να προσδιοριστεί η κατηγορία στην οποία ανήκουν οι παρακάτω μεταβλητές με βάση την κλίμακα μέτρησής τους.

ύψος σε εκατοστά

βάρος σε κιλά

χαρακτηρισμός ύψους(κοντός μέτριος ψηλός)

άποψη για την πολιτική κατάσταση της χώρας (1:κακή 2:καλή 3:πολύ καλή)

αριθμός απουσιών από την εργασία

επάγγελμα

Μισθός σε Ευρώ

Οι διευθύνσεις ενός Υπουργείου

Συλλογή δεδομένων

- **Απογραφή** είναι η μελέτη όλων των μονάδων του πληθυσμού
- **Δειγματοληψία** είναι η μελέτη ενός υποσυνόλου του πληθυσμού.

Απογραφή

Πλεονεκτήματα

- Συγκέντρωση πληροφοριών από όλα τα μέλη του πληθυσμού.
- Υπολογισμός των στατιστικών παραμέτρων του πληθυσμού

Μειονεκτήματα

1. Υψηλό κόστος
2. Χρονοβόρα
3. Δυσκολία πρόσβασης σε κάποια μέλη του πληθυσμού
4. Σε κάποιες περιπτώσεις αδύνατη

Δειγματοληψία

Πλεονεκτήματα

1. Μείωση κόστους
2. Μείωση διάρκειας
3. Χρήση σε περιπτώσεις που η πρόσβαση στις μονάδες του πληθυσμού είναι δύσκολη ή αδύνατη
4. Χρήση σε περιπτώσεις που η έρευνα απαιτεί την καταστροφή της μονάδας του πληθυσμού

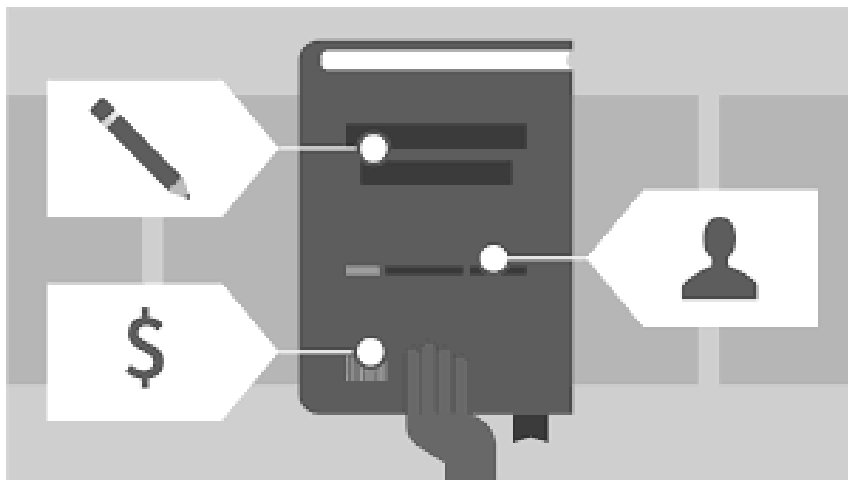
Μειονεκτήματα

1. Αδυναμία δείγματος να συμπεριλάβει όλα τα χαρακτηριστικά του πληθυσμού
2. Εσφαλμένες εκτιμήσεις των στατιστικών παραμέτρων του δείγματος, σε περίπτωση που η μέθοδος δειγματοληψίας δεν έχει επιλεγεί προσεκτικά.

Μεταδεδομένα

Είναι δεδομένα τα οποία αφορούν αντικείμενα ή άλλα δεδομένα. Συνοδεύουν αδιάρρηκτα τα αντικείμενα ή τα δεδομένα στα οποία αναφέρονται, δίνοντας ζωτικές πληροφορίες για αυτά.

Παράδειγμα μεταδεδομένων βιβλίου



- ✓Ο τίτλος
- ✓Οι λέξεις κλειδιά
- ✓Ο συγγραφέας
- ✓Ο εκδότης
- ✓Οι πίνακες περιεχομένων
- ✓Πεδία με τα δικαιώματα χρήσης

Συνήθη μεταδεδομένα Στατιστικής Έρευνας



- ✓ Η υπηρεσία που πραγματοποιεί την έρευνα
- ✓ Στοιχεία υπεύθυνου της έρευνας
- ✓ Η περιγραφή του τρόπου μέτρησης των μεταβλητών
- ✓ Η περιγραφή του πληθυσμού της έρευνας
- ✓ Θέματα εμπιστευτικότητας

Άσκηση

- Επισκεφθείτε τον υπερσύνδεσμο [ΕΛΣΤΑΤ](#)
- Στην εν λόγω τοποθεσία υπάρχουν τα συγκεντρωτικά δεδομένα της έρευνας σε xls μορφή
- Στην εν λόγω τοποθεσία υπάρχει ένα αρχείο που αφορά τα μεταδεδομένα για την έρευνα με τίτλο:
- **Έρευνα Οικογενειακών Προϋπολογισμών, 2019**
- Προσδιορίστε τον/την υπεύθυνο/η για τη δημιουργία του σχετικού ερωτηματολογίου.
- Ποιος είναι ο βασικός σκοπός και στόχοι της έρευνας;
- Η έρευνα είναι σε επίπεδο πληθυσμού (απογραφή) ή αφορά δειγματοληψία; Αν πρόκειται για δειγματοληψία προσδιορίστε με ποιο τρόπο οι υπεύθυνοι της έρευνας ισχυρίζονται ότι εξασφαλίζουν την αντιπροσωπευτικότητα του δείγματος;
- Προσδιορίστε τις στατιστικές ή δειγματοληπτικές μονάδες της έρευνας καθώς και την περιοχή αναφοράς. Είναι απαραίτητο να προσδιορίζεται η περιοχή αναφοράς και γιατί;
- Ποια είναι η πολιτική εμπιστευτικότητας που έχει υιοθετηθεί στην έρευνα αυτή; Είναι απαραίτητο να συνοδεύεται μια έρευνα από πολιτική εμπιστευτικότητας; Δικαιολογείστε.
- Αναφέρετε κάποιες από τις μεταβλητές που αποτυπώνονται από την έρευνα και προσδιορίστε την κλίμακα μέτρησής τους.

Συνήθεις τεχνικές δειγματοληψίας

Δειγματοληψία πιθανότητας

1. Απλή τυχαία Δειγματοληψία
2. Στρωματοποιημένη Δειγματοληψία
3. Συστηματική Δειγματοληψία
4. Κατά Συστάδες Δειγματοληψία

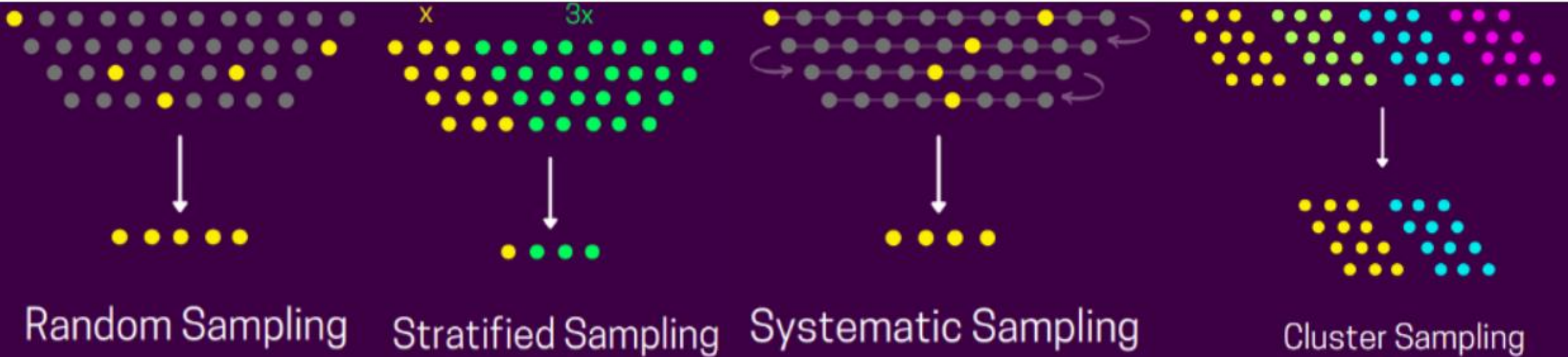
Δειγματοληψία μη πιθανότητας

1. Δειγματοληψία με προκαθορισμένα ποσοστά
2. Δειγματοληψία ευκολίας
3. Δειγματοληψία χιονοστιβάδας
4. Εθελοντική
5. Δειγματοληψία κρίσης

Τεχνικές δειγματοληψίας

Τεχνικές	Πλεονεκτήματα	Μειονεκτήματα
Απλή τυχαία Δειγματοληψία είναι η μέθοδος κατά την οποία όλες οι μονάδες του πληθυσμού έχουν την ίδια πιθανότητα να περιληφθούν στο δείγμα	Αμεροληψία , Αξιοπιστία, Έλεγχος σφάλματος. Δεν απαιτείται προηγούμενη γνώση του πληθυσμού	Υπάρχει περίπτωση υπό ή υπέρ εκπροσώπησης κάποιων ομάδων, δύσκολη η προσέγγιση των στοιχείων του δείγματος
Στρωματοποιημένη Δειγματοληψία είναι η μέθοδος κατά την οποία ο πληθυσμός χωρίζεται σε ομάδες (στρώματα) και στη συνέχεια επιλέγεται ένα απλό τυχαίο δείγμα από κάθε στρώμα	Αντιπροσωπεύονται όλες οι ομάδες βάσει συγκεκριμένων κριτηρίων	Κατάλογοι για κάθε στρώμα
Συστηματική Δειγματοληψία χρησιμοποιείται συνήθως σε περιπτώσεις που η επιλογή του δείγματος γίνεται από μια λίστα.	Εύκολος τρόπος επιλογής δείγματος	Ανάγκη καταλόγων Η σταθερή απόσταση μεταξύ των στοιχείων που λαμβάνονται αυξάνει το σφάλμα σε περίπτωση περιοδικότητας

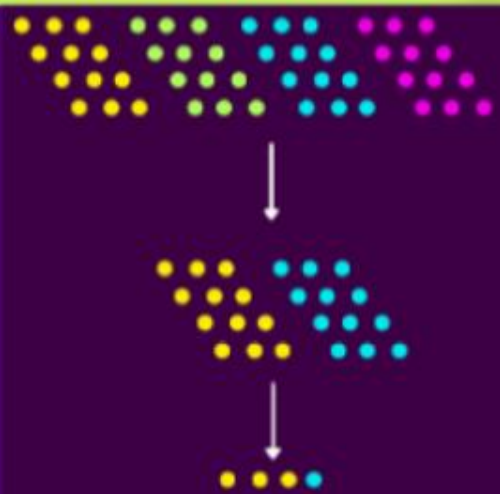
Δειγματοληψία Πιθανότητας



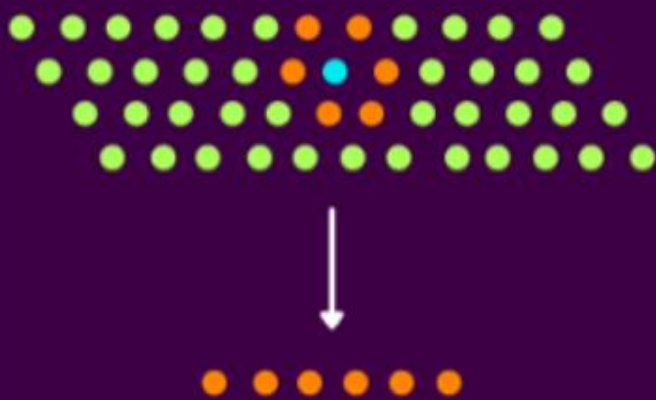
Sampling Methods

Δειγματοληψία μη Πιθανότητας

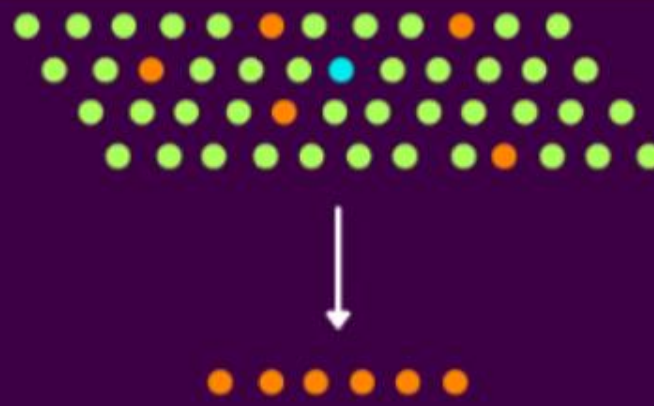
Sampling Methods



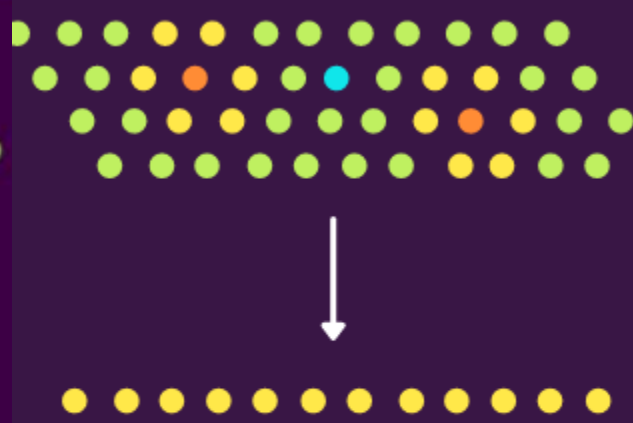
Multistage Sampling



Convenience Sampling



Voluntary Sampling



Snowball Sampling

Δειγματοληψία

Μέθοδοι Δειγματοληψίας

https://repository.kallipos.gr/bitstream/11419/1297/1/02_chapter_01.pdf

Μέθοδοι Δειγματοληψίας σε Ελέγχους

https://ec.europa.eu/regional_policy/sources/docgener/informat/2014/guidance_sampling_method_el.pdf

http://ebooks.edu.gr/ebooks/v/html/8547/4704/Mathimatika-kai-Stoicheia-Statistikis-G-EPAL-html-apli/index2_2.html

Εργαλεία Υπολογισμού Δείγματος

<https://www.abs.gov.au/websitedbs/d3310114.nsf/home/sample+size+calculator>

- <https://www.surveysystem.com/sscalc.htm>
- <http://www.raosoft.com/samplesize.html>
- <https://www.calculator.net/sample-size-calculator.html>

Περιγραφή Δεδομένων-Περιγραφική Στατιστική

1. Πίνακες Συχνοτήτων – Σχετικών Συχνοτήτων
2. Μέτρα θέσης, Μέτρα μεταβλητότητας,
3. Διαγράμματα

2. Αριθμητική Περιγραφή Δεδομένων

1. Πίνακες Συχνοτήτων – Σχετικών Συχνοτήτων
2. Μέτρα θέσης
3. Μέτρα μεταβλητότητας

Μαθησιακοί Στόχοι

Όταν θα έχετε ολοκληρώσει την μελέτη της ενότητας θα είστε στην θέση να

- ▷ Υπολογίζετε μέσω Η/Υ περιγραφικά μέτρα
- ▷ Να επιλέγετε τα κατάλληλα μέτρα σε σχέση με το είδος των δεδομένων
- ▷ Να ερμηνεύετε τα διάφορα στατιστικά μέτρα

Πίνακες Συχνοτήτων – Σχετικών Συχνοτήτων

Συχνότητες

Η Συχνότητα μιας τιμής της μεταβλητής X είναι ένας αριθμός που μας δείχνει πόσες φορές έχει εμφανισθεί η τιμή αυτή στα δεδομένα μας.

Σχετικές συχνότητες

Η Σχετική Συχνότητα μιας τιμής της μεταβλητής X είναι το πηλίκο της συχνότητας της τιμής αυτής προς το άθροισμα των συχνοτήτων όλων των τιμών.

Παράδειγμα πίνακα Συχνοτήτων–Σχετικών Συχνοτήτων

Ημέρες άδειας	Συχνότητα	Σχετική συχνότητα
1	10	50,00%
2	7	35,00%
5	3	15,00%
Σύνολο	20	100,00%

Αθροιστική & Σχετική Αθροιστική Συχνότητα

Αθροιστική Συχνότητα τιμής x

Είναι το άθροισμα των συχνοτήτων των τιμών που είναι μικρότερες ή ίσες από κάποιο επίπεδο τιμής της εξεταζόμενης μεταβλητής.

Σχετική αθροιστική συχνότητα μιας τιμής x

Είναι το άθροισμα των σχετικών συχνοτήτων των τιμών, που είναι μικρότερες ή ίσες από κάποιο επίπεδο τιμής της εξεταζόμενης μεταβλητής.

Πίνακες Αθροιστικής & Σχετικής Αθροιστικής Συχνότητας

Ημέρες άδειας	Συχνότητα	Σχετική συχνότητα	Αθροιστική συχνότητα	Σχετική αθροιστική συχνότητα
1	10	50,00%	10	50,00%
2	7	35,00%	17	85,00%
5	3	15,00%	20	100,00%
Σύνολο	20	100,00%		

Πίνακες Συχνοτήτων για Συνεχή Δεδομένα

Μισθός

1100

1120

1100

1113

1130

1140

1150

1130

1145

1145

1135

1100

1200

1250

1230

1330

1300

1360

1400

1200

Υπολογισμός Πλήθους Κλάσεων $k = 1 + 3.32 * \text{LOG}_{10}(n)$ με τον τύπο του Sturges

Παράδειγμα Μισθών $k = 1 + 3.32 \text{LOG}_{10}(20) = 5,32$

Εύρος = $1400 - 1100 = 300$

Πλάτος κλάσης $300 / 5,32 = 56,4$ ή στην πράξη περίπου 60

Εναλλακτικά μπορεί να υπολογισθούν οι κλάσεις:

- Η τετραγωνική ρίζα του n αν $n < 50$

- Από τον τύπο : $n = 2^c$ όπου c το πλήθος των κλάσεων. Δίνει μία κλάση λιγότερη από τον Sturges

Κλάσεις	Μέγιστη τιμή	Συχνότητες
1100-1160	1160	12
1161-1220	1220	2
1221-1280	1280	2
1281-1340	1340	2
1341-1400	1400	2

Πίνακες Συνάφειας (Διασταύρωσης)

Ένας πίνακας συνάφειας (ή αλλιώς, διασταύρωσης) αποτελείται από γραμμές και στήλες οι οποίες ορίζονται από τις κατηγορίες των μεταβλητών. Στα κελιά του πίνακα μπορεί να εμφανίζονται οι συχνότητες, τα ποσοστά γραμμής ή στήλης.

	Καπνιστής		
Φύλο	Ναι	Όχι	Σύνολο
Άνδρας	3	4	7
Γυναίκα	7	6	13
Σύνολο	10	10	20

Πίνακες Συνάφειας (Διασταύρωσης)

Ποσοστά επί των γραμμών

		Καπνιστής		Σύνολο
		Ναι	Όχι	
Φύλο	Άνδρας	42,86%	57,14%	100,00%
	Γυναίκα	53,85%	46,15%	100,00%
	Σύνολο	50,00%	50,00%	100,00%

Μέτρα θέσης

1. Αριθμητικός μέσος
2. Διάμεσος
3. Επικρατούσα τιμή
4. Άθροισμα
5. Ποσοστιαία Σημεία

Αριθμητικός Μέσος


Ο Αριθμητικός Μέσος ή απλά Μέσος ενός συνόλου παρατηρήσεων ισούται με το άθροισμα των παρατηρήσεων δια του πλήθους των παρατηρήσεων. Όταν υπολογίζουμε το μέσο του δείγματος αναφερόμαστε στον δειγματικό μέσο ενώ όταν υπολογίζουμε το μέσο του πληθυσμού αναφερόμαστε στη μέση τιμή του πληθυσμού.

Παράδειγμα

Οι μηνιαίοι μισθοί 5 εργαζομένων είναι:

350, 500, 500, 500, 4000, άρα ο μέσος μισθός είναι:

$$\begin{aligned}\text{Μεσος Μισθος} &= \frac{350 + 500 + 500 + 500 + 4000}{5} \\ &= \frac{5850}{5} = \mathbf{1170}\end{aligned}$$



Ο μέσος δεν είναι υποχρεωτικό να ισούται με κάποια από τις παρατηρήσεις.

Το μεγαλύτερο μέρος των παρατηρήσεων δεν βρίσκεται απαραίτητα κοντά στον μέσο.

Ο μέσος επηρεάζεται από την ύπαρξη ακραίων τιμών, δηλαδή τιμών με μικρή συχνότητα εμφάνισης και μακριά από τις τιμές του κύριου όγκου των δεδομένων

Περιομμένος ή Τετριμμένος (trimmed mean) Μέσος

Ο μέσος όρος των παρατηρήσεων που προκύπτουν , αν από τις αρχικές παρατηρήσεις εξαιρέσουμε το $\alpha\%$ των μεγαλύτερων και το $\alpha\%$ των μικρότερων τιμών ονομάζεται Τετριμμένος ή αλλιώς Περιομμένος Μέσος.

Οι βαθμολογίες 20 σπουδαστών της Σχολής στο μάθημα της στατιστικής είναι: 1, 2, 75, 75, 89, 87, 90, 90, 92, 93, 94, 94, 94, 95, 97, 100, 100, 100, 100, 100.

Μέσος : 83,4

10% Περιομμένος μέσος 87,06

Ο περιομμένος μέσος επηρεάζεται λιγότερο από τις ακραίες τιμές των παρατηρήσεων από ότι ο αριθμητικός μέσος.

Διάμεσος

Η **Διάμεσος** ενός συνόλου παρατηρήσεων είναι η τιμή εκείνη για την οποία ισχύει ότι το πολύ το 50% των παρατηρήσεων είναι μικρότερες από αυτήν και το πολύ το 50% των παρατηρήσεων είναι μεγαλύτερο από αυτήν.

Μεσαία παρατήρηση = 2

Νοικοκυριό	A	B	Γ	Δ	E	Z	H	Θ	I
Αριθμός μελών	1	1	2	2	2	3	4	5	5

Διάμεσος

Ημιάθροισμα δύο μεσαίων παρατηρήσεων
= $(2+3)/2=2,5$

Νοικοκυριό	A	B	Γ	Δ	E	Z	H	Θ	I	K
Αριθμός μελών	1	1	2	2	2	3	4	5	5	6

Επικρατούσα τιμή

Είναι η τιμή με την μεγαλύτερη συχνότητα εμφάνισης

Ημέρες άδειας	Συχνότητα
0	10
2	7
5	3
Σύνολο	20

Συνοψίζοντας σχετικά με...

Μέσο

Χρησιμοποιείται μόνο σε ποσοτικά δεδομένα

Είναι ευαίσθητος στην επίδραση ακραίων τιμών

Είναι κατάλληλος για θεωρητική ανάπτυξη

Χρησιμοποιούνται όλες οι τιμές για τον υπολογισμό του

Διάμεσο

Χρησιμοποιείται σε ποσοτικά δεδομένα και σε δεδομένα διάταξης

Δεν είναι ευαίσθητη στην επίδραση ακραίων τιμών
Δεν είναι εύκολη η θεωρητική ανάπτυξη

Δεν χρησιμοποιούνται όλες οι τιμές για τον υπολογισμό της

Επικρατούσα Τιμή

Χρησιμοποιείται σε ποσοτικά και ποιοτικά δεδομένα

Δεν είναι ευαίσθητη στην επίδραση ακραίων τιμών

Δεν είναι εύκολη η θεωρητική ανάπτυξη

Δεν χρησιμοποιούνται όλες οι τιμές για τον υπολογισμό της

Ποσοστιαία σημεία Τεταρτημόρια

Θα λέμε ότι μια τιμή x είναι το **α ποσοστιαίο σημείο** της κατανομής αν το $\alpha\%$ το πολύ των παρατηρήσεων είναι μικρότερες από το x και το πολύ το $(100-\alpha)\%$ των παρατηρήσεων είναι μεγαλύτερες από το x .

1^ο τεταρτημόριο Q1

Ονομάζεται το σημείο που το πολύ το 25% των παρατηρήσεων είναι μικρότερες από αυτό και το πολύ το 75% των παρατηρήσεων είναι μεγαλύτερες από αυτό

3^ο τεταρτημόριο Q3

Ονομάζεται το σημείο για το οποίο ισχύει ότι το πολύ το 75% των παρατηρήσεων είναι μικρότερες από αυτό και το πολύ το 25% των παρατηρήσεων είναι μεγαλύτερες από αυτό.

Σπουδαστής/ια	Βαθμολογία
Σπ1	1
Σπ2	2
Σπ3	75
Σπ4	75
Σπ5	89
Σπ6	87
Σπ7	90
Σπ8	90
Σπ9	92
Σπ10	93
Σπ11	94
Σπ12	94
Σπ13	94
Σπ14	95
Σπ15	97
Σπ16	100
Σπ17	100
Σπ18	100
Σπ19	100
Σπ20	100

10^ο ποσοστιαίο σημείο=67,7

1^ο τεταρτημόριο Q1= 88,5

2^ο τεταρτημόριο(διάμεσος)=93,5

3^ο τεταρτημόριο Q3=97,75

Μέτρο	Συνάρτηση
Αριθμητικός μέσος	=AVERAGE(B2:B21)
Περιοκομμένος μέσος Τετριμμένος μέσος	=TRIMMEAN(B2:B21;0,2)
Διάμεσος	=MEDIAN(B2:B21)
Επικρατούσα τιμή	=MODE.SNGL(B2:B21)
Τεταρτημόρια 1=1ο,2=2ο,3=3ο)	=QUARTILE.INC(B2:B21;1)
Κ-ποσοστημόριο	=PERCENTILE.INC(E7:E26;0,1)
Μέγιστο	=MAX(B2:B21)
Ελάχιστο	=MIN(B2:B21)

Υπολογισμός με το Ανάλυση Δεδομένων του EXCEL

- 1.Ενεργοποιήστε το Ανάλυση Δεδομένων του EXCEL :Αρχείο /Επιλογές/ Πρόσθετα / Πακέτο Ανάλυσης Δεδομένων
2. Επιλογή από το Ανάλυση Δεδομένων το Περιγραφικά Στατιστικά Στοιχεία

Μέσος	83,4
Τυπικό σφάλμα	6,466269
Διάμεσος	93,5
Επικρατούσα τιμή	100
Μέση απόκλιση τετραγώνου	28,91803
Διακύμανση	836,2526
Κύρτωση	5,809712
Ασυμμετρία	-2,58068
Εύρος	99
Ελάχιστο	1
Μέγιστο	100
Άθροισμα	1668
Πλήθος	20

Μέτρα μεταβλητότητας

Εύρος

Είναι η διαφορά της μικρότερης από την μεγαλύτερη τιμή των δεδομένων.

Ενδοτεταρτημοριακό Εύρος

Είναι η διαφορά του πρώτου τεταρτημορίου από το τρίτο, δηλαδή $IQR=Q3-Q1$.

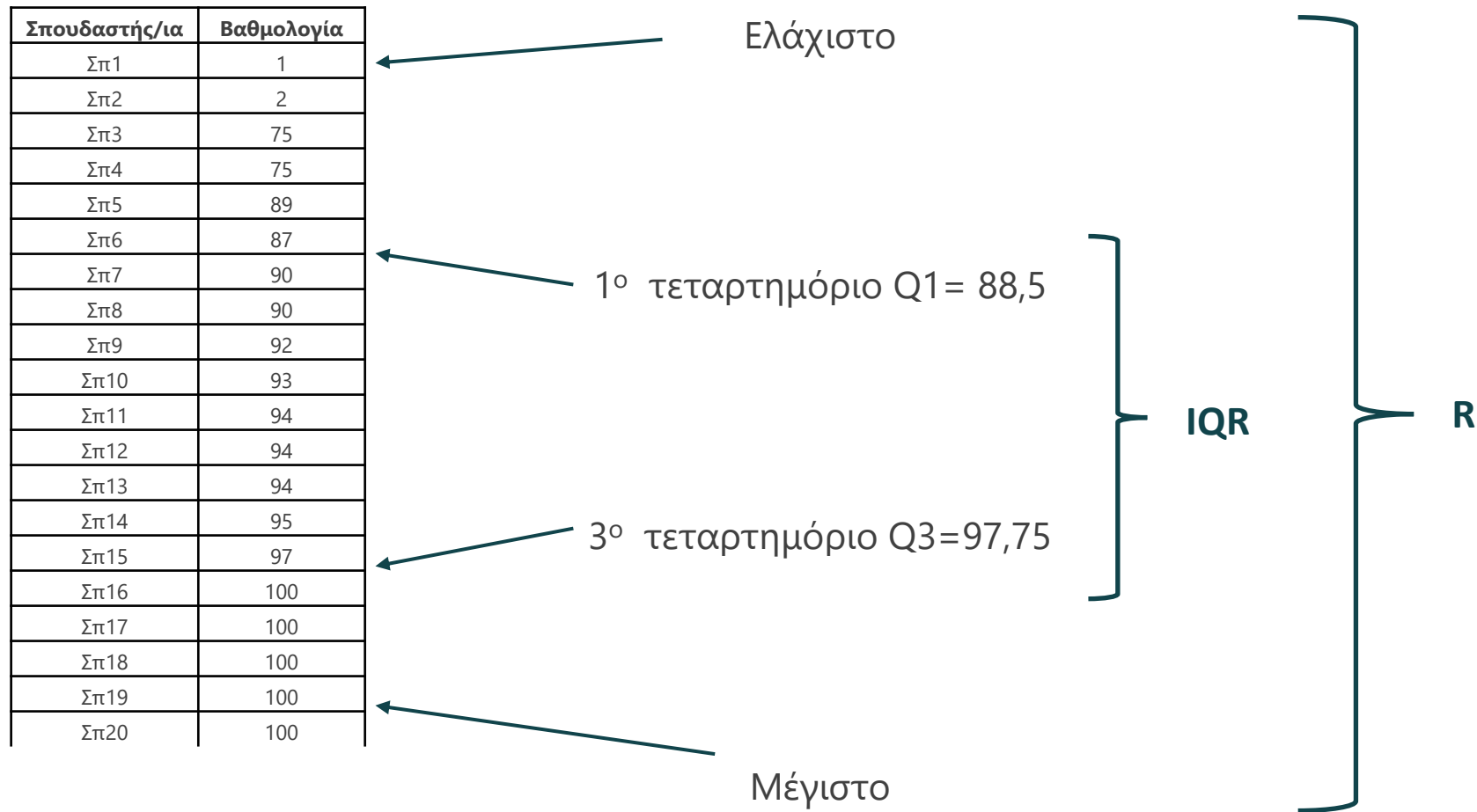
Διακύμανση/ Διασπορά

Είναι ο μέσος όρος των τετραγώνων των αποκλίσεων από τον μέση τιμή

Τυπική απόκλιση

Η τετραγωνική ρίζα της διακύμανσης

Εύρος (R) Ενδοτεταρτημοριακό Εύρος (IQR)



Συνοψίζοντας σχετικά με...

Το **Εύρος** εξαρτάται μόνο από τις δύο πιο ακραίες τιμές , δεν χρησιμοποιείται για συμπερασματολογία και είναι δύσκολη η θεωρητική ανάπτυξη

Η **Τυπική Απόκλιση** είναι λιγότερο ευαίσθητη στην επίδραση των ακραίων τιμών από το εύρος. Χρησιμοποιούνται όλες οι τιμές για τον υπολογισμό της, χρησιμοποιείται για συμπερασματολογία και είναι κατάλληλη για θεωρητική ανάπτυξη.

Συνδυασμός Μέσης Τιμής –Τυπικής Απόκλισης

Θεώρημα του Chebyshev

Για οποιοδήποτε σύνολο δεδομένων με **μέσο μ** και **τυπική απόκλιση σ** και για οποιονδήποτε αριθμό $k > 1$, τουλάχιστον το $(1-1/k^2)*100$ των δεδομένων, βρίσκονται στο διάστημα $(\mu-k\sigma, \mu+k\sigma)$

Παράδειγμα

Η μέση βαθμολογία των σπουδαστών μιας σχολής είναι 90 με τυπική απόκλιση 2

- ▷ Τουλάχιστον το 0% των σπουδαστών είχε βαθμολογία στο διάστημα (88 , 92)
- ▷ Τουλάχιστον το 75% των σπουδαστών είχε βαθμολογία στο διάστημα (86 , 94)
- ▷ Τουλάχιστον το 88,9% των σπουδαστών είχε βαθμολογία στο διάστημα (84 , 96)

Μέτρα σχετικής Θέσης - Μεταβλητότητας

Τυποποιημένες τιμές (Z - τιμές)

Μας δείχνουν πόσες τυπικές αποκλίσεις απέχει μία συγκεκριμένη τιμή από τον μέσο όρο.

Συντελεστής μεταβλητότητας (CV)

Ισούται με το πηλίκο της τυπικής απόκλισης των τιμών προς την απόλυτη τιμή του μέσου.

Παράδειγμα

Η μέση βαθμολογία της τάξης στο μάθημα Α: ήταν 17 με τυπική απόκλιση 1

Β: η μέση βαθμολογία ήταν 10 με τυπική απόκλιση 1 , Γ: η μέση βαθμολογία ήταν 13 με τυπική απόκλιση 1

$$Z_A = \frac{16 - 17}{1} = -1$$

$$Z_B = \frac{12 - 10}{1} = 2$$

$$Z_\Gamma = \frac{14 - 13}{1} = 1$$

$$CV = \frac{1}{17} = 5,9\%$$

$$CV = \frac{1}{10} = 10\%$$

$$CV = \frac{1}{13} = 7,7\%$$

Άσκηση

Στον ακόλουθο υπερσύνδεσμο [EE](https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=lfsi_emp_a&lang=en) θα βρείτε στοιχεία που αφορούν τα ποσοστά απασχόλησης στις χώρες της ΕΕ. https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=lfsi_emp_a&lang=en

Κατεβάστε το αρχείο των δεδομένων σε μορφή xls.

- Υπολογίστε τα συνήθη στατιστικά μέτρα θέσης και μεταβλητότητας για κάθε εγγραφή του αρχείου που αφορούν την επίδοση της για την περίοδο από 2011 έως και 2020. Σε περίπτωση που εμφανίζονται πολλαπλές ελλείπουσες τιμές στα δεδομένα μιας εγγραφής εξαιρέστε την εγγραφή αλλιώς αντικαταστήστε την ελλείπουσα τιμή από τη μέση τιμή των δύο γειτονικών ετών.
- Υπολογίστε την μέση επίδοση όλων των μελών της ΕΕ για το 2020 καθώς και την μέση επίδοση για τις χώρες-μέλη της ΕΕ που ανήκουν στη Βαλκανική. Αξιοποιείστε μόνο τις χώρες που προσφέρονται δεδομένα. Εμφανίζεται διαφορά μεταξύ της στατιστικής παραμέτρου (δηλ. σε επίπεδο ΕΕ) και της εκτίμησης μέσω του δείγματος της Βαλκανικής; Πόσο τις εκατό είναι η διαφορά αυτή; Θεωρείτε ότι θα ήταν αντιπροσωπευτικό δείγμα του πληθυσμού οι εν λόγω χώρες; Δικαιολογήστε την απάντησή σας.
- Μελετήστε τα μεταδεδομένα του αρχείου δεδομένων.

3. Γραφήματα- Κατανομές

1. Γραφήματα Ποιοτικών Μεταβλητών
2. Γραφήματα Ποσοτικών Μεταβλητών
3. Κατανομές
4. Μέτρα σχηματικής μορφής

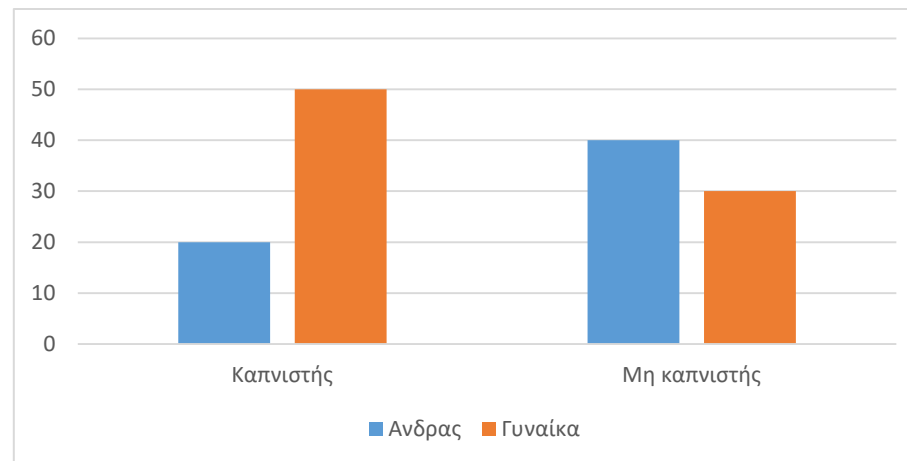
Μαθησιακοί στόχοι

Όταν θα έχετε ολοκληρώσει την μελέτη της ενότητας

- ▷ Να κατασκευάζετε γραφήματα
- ▷ Να επιλέγετε το κατάλληλο γράφημα σε σχέση με τα δεδομένα
- ▷ Θα έχετε κατανοήσει την έννοια της κατανομής

Ραβδόγραμμα

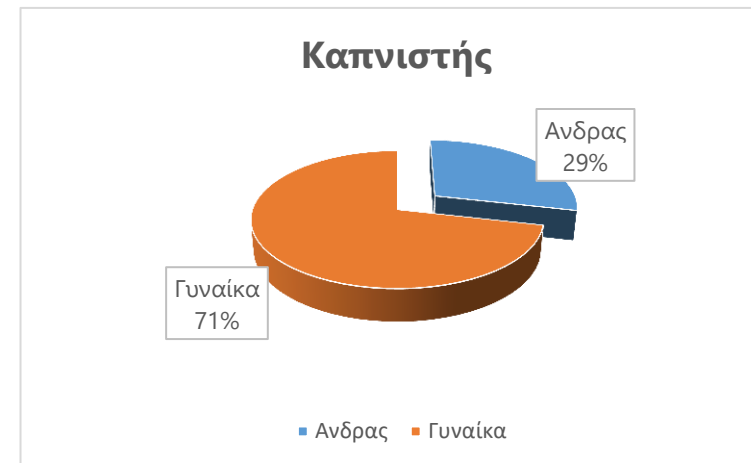
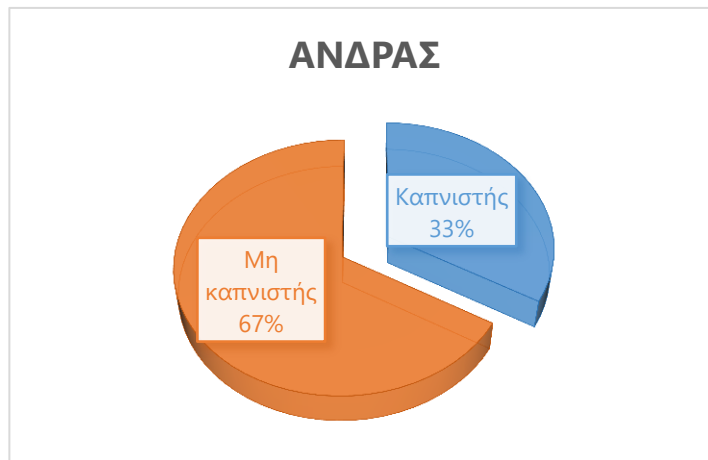
Αποτελείται από μία στήλη για κάθε κατηγορία της μεταβλητής ή των μεταβλητών που περιγράφει. Οι στήλες είναι κάθετες στον οριζόντιο ή στον κατακόρυφο άξονα και το ύψος τους ισούται με την συχνότητα ή την σχετική συχνότητα κάθε κατηγορίας (τιμές της μεταβλητής).



Κυκλικό Γράφημα (Πίτα)

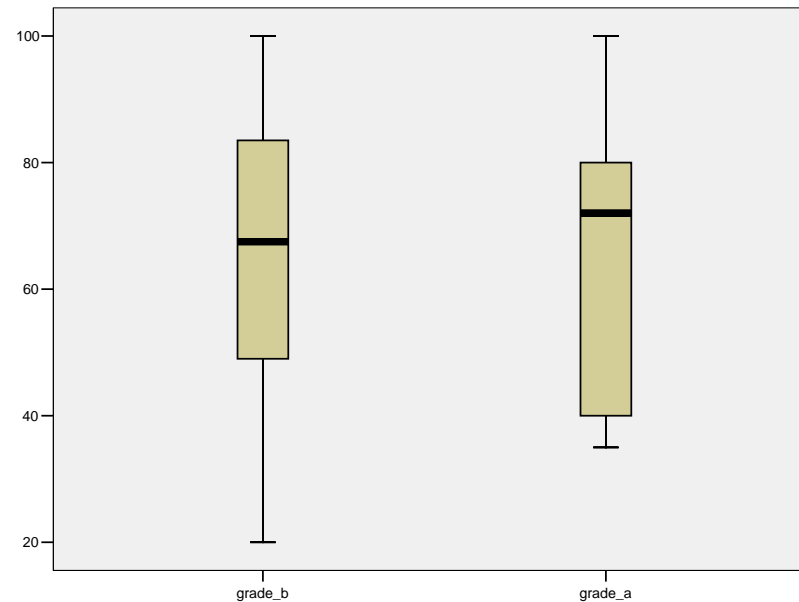
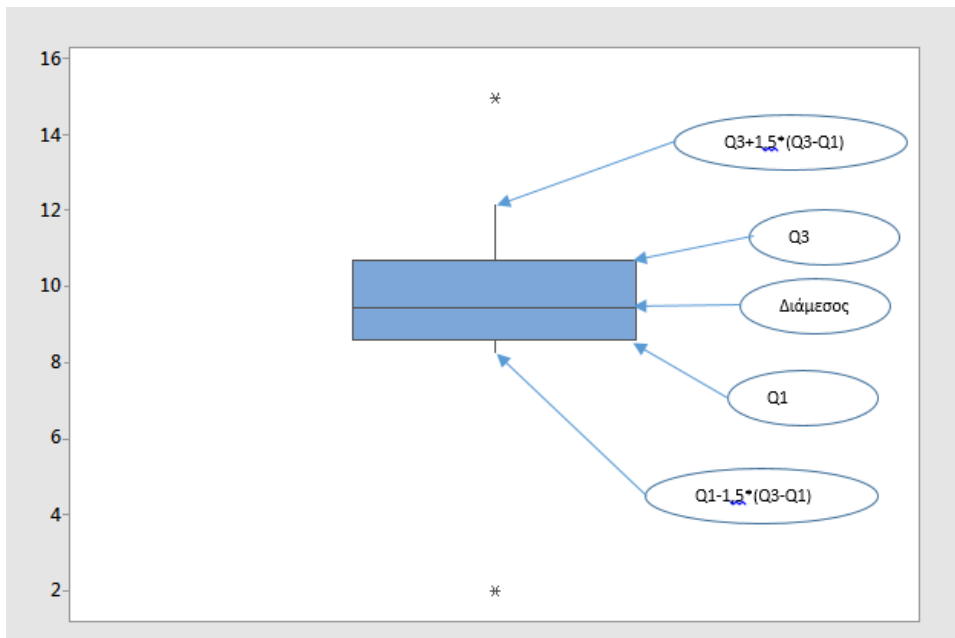
Είναι ένας κυκλικός δίσκος ο οποίος είναι χωρισμένος σε τόσα τμήματα όσες και οι κατηγορίες της μεταβλητής. Τα εμβαδά των τμημάτων στα οποία έχει χωριστεί ο κυκλικός δίσκος είναι ανάλογα των συχνοτήτων των κατηγοριών της μεταβλητής.

Το κυκλικό διάγραμμα χρησιμοποιείται για την παρουσίαση της κατανομής των τιμών μιας ποιοτικής ή μιας ποσοτικής διακριτής μεταβλητής.



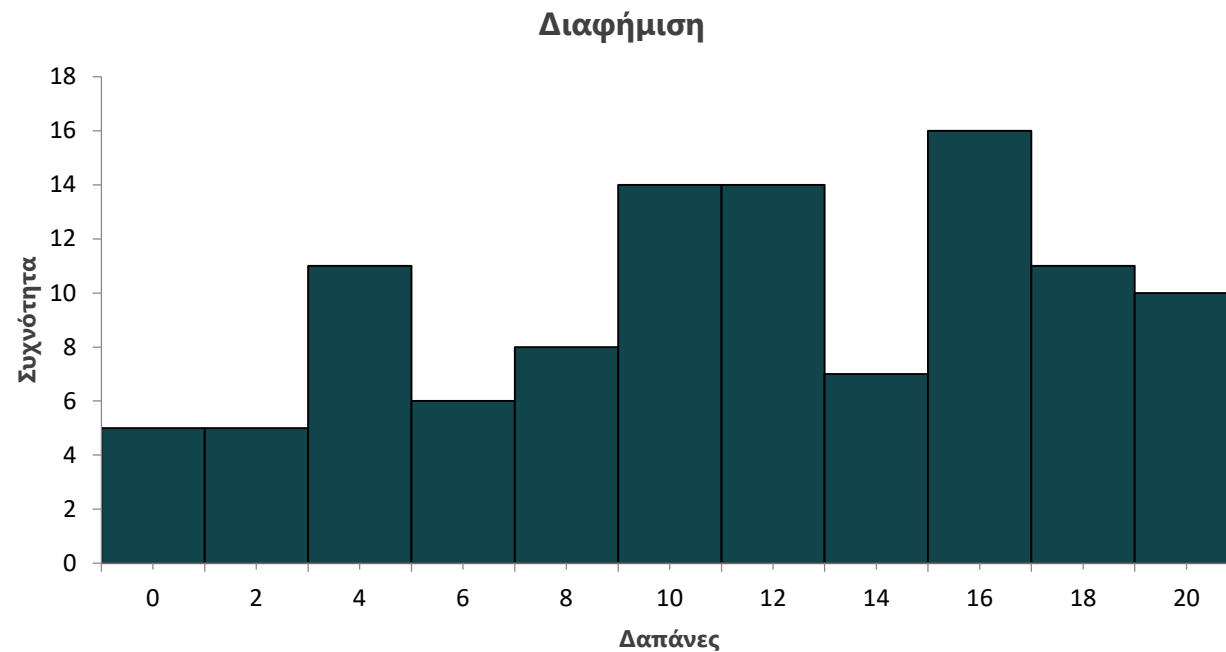
Θηκόγραμμα

Το Θηκόγραμμα αποτελείται από 5 βασικά σημεία, το πρώτο (Q1) και το τρίτο(Q3) τεταρτημόριο, την διάμεσο, το κάτω οριακή τιμή ($Q1-1,5*(Q3-Q1)$) και την άνω οριακή τιμή ($Q3+1,5*(Q3-Q1)$). Τιμές μικρότερες από την κάτω οριακή τιμή ή μεγαλύτερες από την πάνω οριακή τιμή θεωρούνται ακραίες.



Ιστόγραμμα

Με το Ιστόγραμμα απεικονίζονται γραφικά ομαδοποιημένα δεδομένα, τα οποία αναφέρονται σε συνεχείς μεταβλητές.



Διάγραμμα Αράχνης

Όταν τα δεδομένα παρουσιάζονται σε γραμμές και στήλες όπως στον παρακάτω πίνακα, ένα γράφημα για την παρουσίαση και την σύγκριση των τιμών του πίνακα είναι το αραχνοειδές.

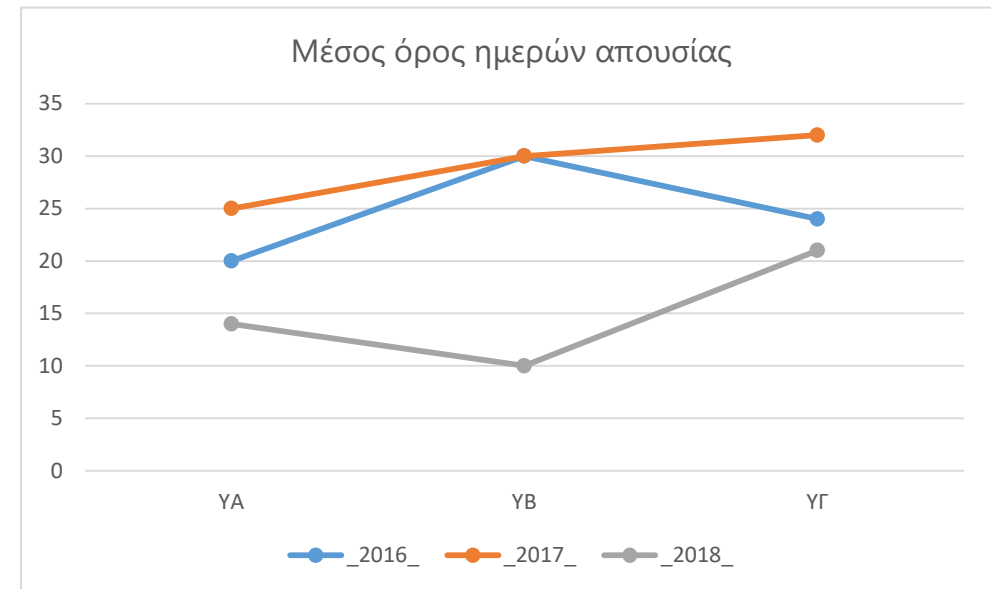
Μέσος όρος ημερών απουσίας υπαλλήλου από την υπηρεσία			
Έτη	YA	YB	YΓ
2016	20	30	24
2017	25	30	32
2018	14	10	21



Γράφημα Γραμμής

Με το γράφημα γραμμής μπορούμε να αποτυπώσουμε την εξέλιξη ενός χαρακτηριστικού, καθώς επίσης να συγκρίνουμε την εξέλιξη του σε διαφορετικές ομάδες.

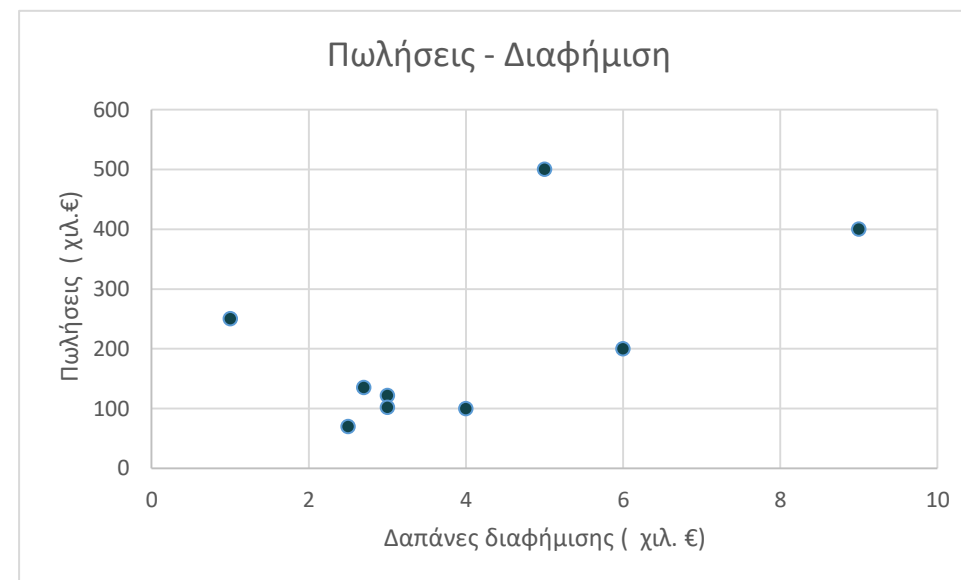
Έτη	Υπ.Α	Υπ.Β	Υπ.Γ
2016	20	30	24
2017	25	30	32
2018	14	10	21



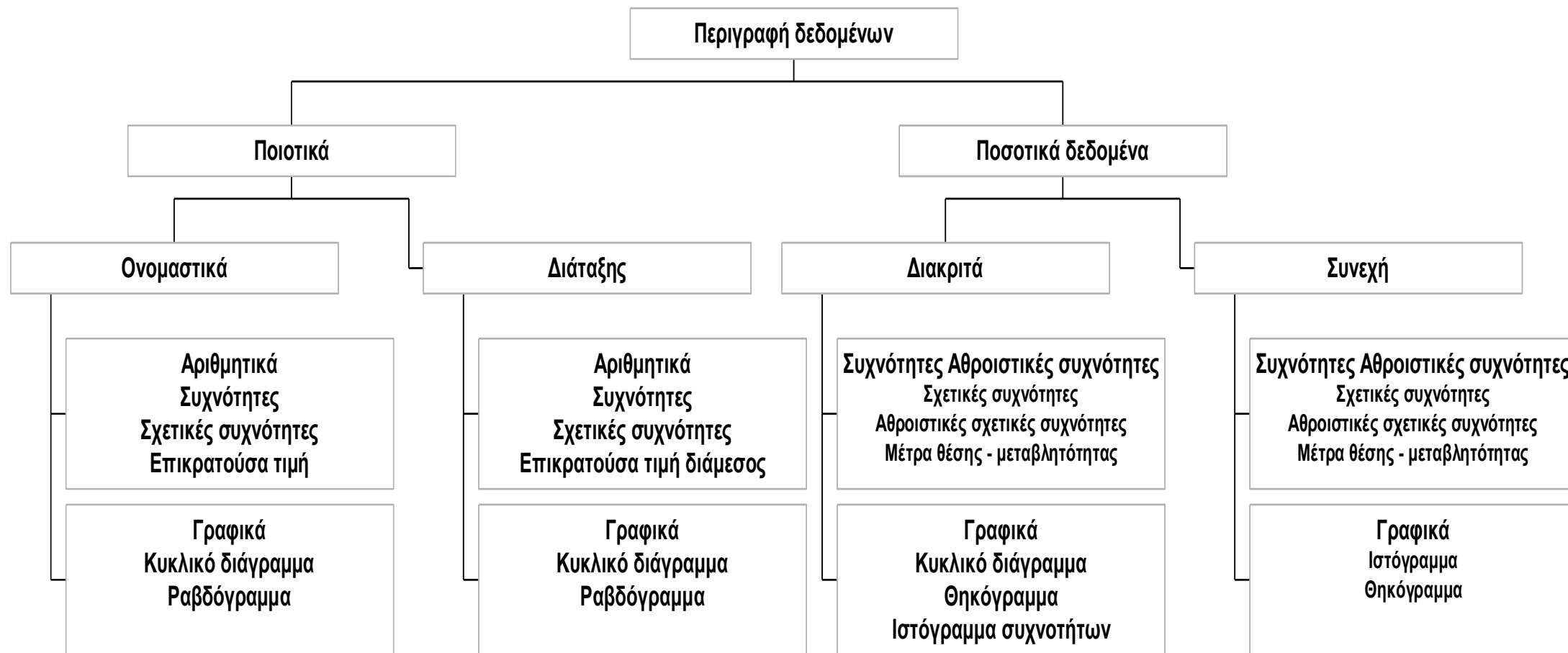
Γράφημα διασποράς

Με το γράφημα διασποράς μπορούμε να αποτυπώσουμε την σχέση μεταξύ δύο ποσοτικών μεταβλητών.

Έτος	Δαπάνες διαφήμισης (σε χιλ. €)	Πωλήσεις (σε χιλ.€)
2010	1	250
2011	3	122
2012	2,7	135
2013	6	200
2014	9	400
2015	5	500
2016	4	100
2017	3	102
2018	2,5	70



Συνοψίζοντας



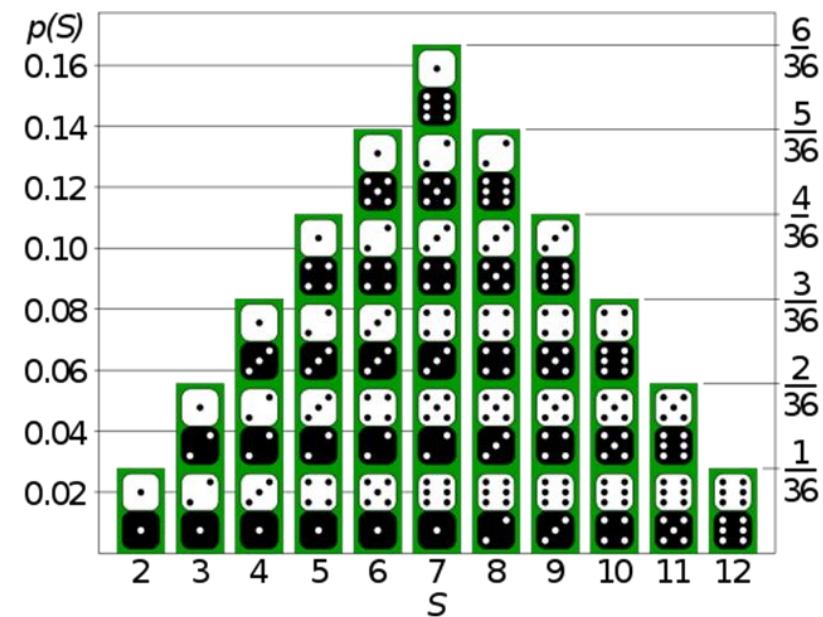
Κατανομή

Η Κατανομή μιας μεταβλητής μας δείχνει όλες τις πιθανές τιμές (ή διαστήματα) των τιμών της μεταβλητής και πόσο συχνά αυτές εμφανίζονται

Παράδειγμα

Ρίχνουμε δύο ζάρια, ένα μαύρο και ένα άσπρο και σημειώνουμε το άθροισμα των ενδείξεων,

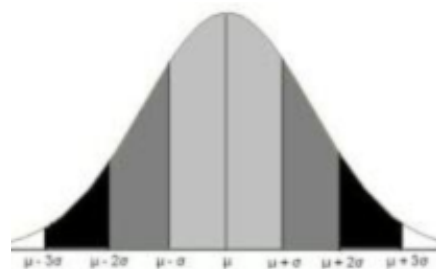
τότε η Κατανομή των τιμών της μεταβλητής 'Άθροισμα ενδείξεων', απεικονίζεται στο διπλανό γράφημα.



Πηγή Wikipedia

Κανονική Κατανομή

Είναι συμμετρική ως προς το μέσο



Ο Μέσος και η Διάμεσος της Κανονικής Κατανομής είναι ίσοι μεταξύ τους

Στο $(\mu - \sigma, \mu + \sigma)$ ανήκει το 68% των παρατηρήσεων

Στο $(\mu - 2\sigma, \mu + 2\sigma)$ ανήκει το 95% των παρατηρήσεων

Στο $(\mu - 3\sigma, \mu + 3\sigma)$ ανήκει το 99,7% των παρατηρήσεων

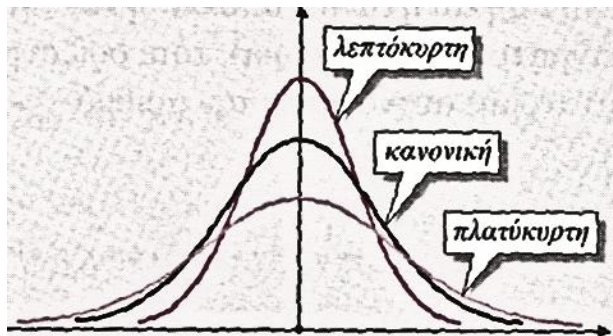
Τυποποιημένη κανονική κατανομή έχει Μέσο $\mu=0$, Τυπική απόκλιση $\sigma=1$

Μέτρα σχηματικής μορφής

Συντελεστής κύρτωσης

Ο Βαθμός Κύρτωσης δείχνει πόσο απότομη είναι η καμπύλη της κατανομής. Κατά πόσο συγκεντρώνονται οι τιμές γύρω από το μέσο.

- ✓ Συντελεστής κύρτωσης=0 μεσόκυρτη
- ✓ Συντελεστής κύρτωσης<0 πλατύκυρτη
- ✓ Συντελεστής κύρτωσης>0 λεπτόκυρτη

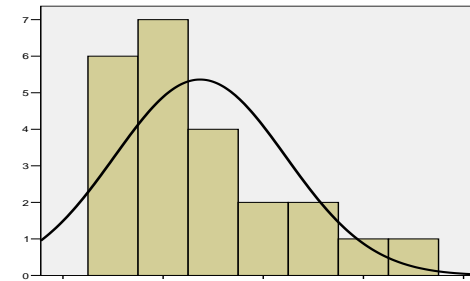


Συντελεστής ασυμμετρίας

Συντελεστής ασυμμετρίας=0 συμμετρική γύρω από το μέσο

Συντελεστής ασυμμετρίας<0 αριστερή ασυμμετρία

Συντελεστής ασυμμετρίας>0 δεξιά ασυμμετρία



Άσκηση

- Στον ακόλουθο υπερσύνδεσμο [ΕΕ](#) θα βρείτε στοιχεία που αφορούν τα ποσοστά απασχόλησης στις χώρες της ΕΕ. Να δημιουργήσετε
- Διάγραμμα γραμμής για την Ελλάδα και τη Γερμανία στο διάστημα ετών από 2008 έως και 2016 αναφορικά με το βαθμό απασχόλησης στις δύο χώρες.
- Με τις τιμές όλων των κρατών-μελών της ΕΕ για τα έτη 2008 και 2016 δημιουργείστε θηκογράμματα και συγκρίνετε τις δύο κατανομές. Εξετάστε αν εμφανίζονται ακραίες τιμές και αν οι εν λόγω κατανομές θα μπορούσαν να χαρακτηριστούν ως συμμετρικές.
- Υπολογίστε τα μέτρα ασυμμετρίας και κύρτωσης για τις ετήσιες κατανομές των τιμών της απασχόλησης στην ΕΕ , αναφορικά με τα έτη 2010, 2012 και 2014. Πως χαρακτηρίζετε τις κατανομές αυτές; Σημειώνεται κάποια σημαντική διαφοροποίηση; Συζητήστε την άποψη σας.

4. Επαγωγική Στατιστική

1. Εκτίμηση παραμέτρων
 - a. Σημειακή εκτίμηση
 - b. Διάστημα εμπιστοσύνης
2. Έλεγχος Υποθέσεων
 - a. Έλεγχος υπόθεσης για την μέση τιμή ενός πληθυσμού
 - b. Έλεγχος ανεξαρτησίας χ^2

Μαθησιακοί στόχοι

Όταν θα έχετε ολοκληρώσει την μελέτη της ενότητας

- ▷ Θα έχετε εξοικειωθεί με τις έννοιες της εκτίμησης παραμέτρων και του ελέγχου υποθέσεων
- ▷ Θα γνωρίζετε μεθόδους εκτίμησης παραμέτρων
- ▷ Θα μπορείτε να ερμηνεύετε τα αποτελέσματα ενός ελέγχου

Εκτίμηση σημείου

Παράδειγμα

Το 2014 διεξήχθη έρευνα, η οποία είχε στόχο να μελετήσει τις αιτίες και τα χαρακτηριστικά αποφοίτων Τριτοβάθμιας εκπαίδευσης ηλικίας 25 έως 40 ετών, οι οποίοι μετανάστευσαν προς εύρεση εργασίας.

Σε δείγμα 50 αποφοίτων εδόθη ερωτηματολόγιο στο οποίο εκτός άλλων χαρακτηριστικών, κατεγράφη και ο χρόνος από την αποφοίτηση μέχρι την ημέρα της μετανάστευσης.

Ο μέσος χρόνος στο δείγμα ήταν τα **8 έτη**  **Εκτίμηση πληθυσμιακού μέσου**

Εκτίμηση διαστήματος εμπιστοσύνης

Παράδειγμα

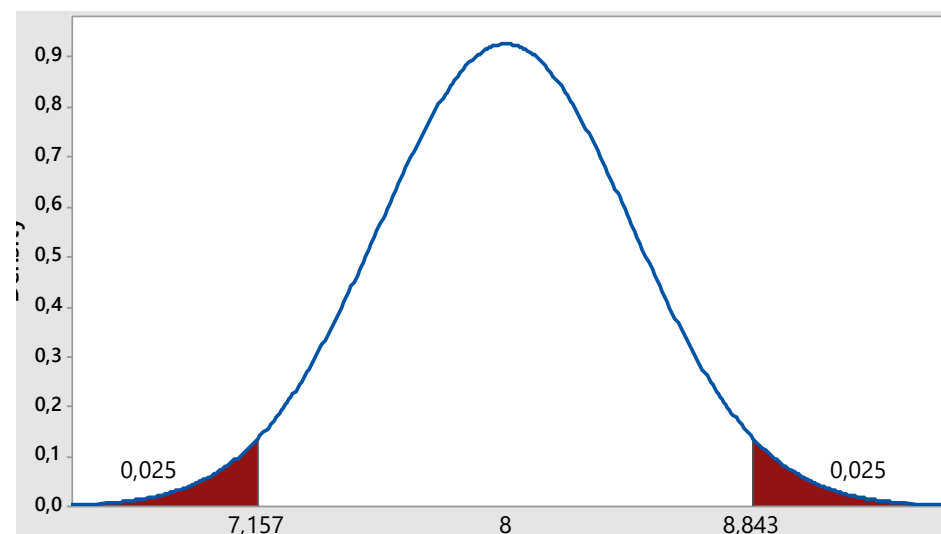
Το 2014 διεξήχθη έρευνα, η οποία είχε στόχο να μελετήσει τις αιτίες και τα χαρακτηριστικά αποφοίτων Τριτοβάθμιας εκπαίδευσης ηλικίας 25 έως 40 ετών, οι οποίοι μετανάστευσαν προς εύρεση εργασίας.

Σε δείγμα 50 αποφοίτων εδόθη ερωτηματολόγιο στο οποίο εκτός άλλων χαρακτηριστικών, κατεγράφη και ο χρόνος από την αποφοίτηση μέχρι την ημέρα της μετανάστευσης, ο οποίος ήταν τα 8 έτη.

Από τα δεδομένα του δείγματος κατασκευάστηκε το **διάστημα (7,2 , 8,8)**, το οποίο περιέχει τον μέσο του πληθυσμού με εμπιστοσύνη 95%

Ερμηνεία διαστήματος εμπιστοσύνης

με πιθανότητα 95% το διάστημα(7,2 , 8,8) θα περιέχει την μέση τιμή του πληθυσμού ή ισοδύναμα το ποσοστό των δειγμάτων των οποίων ο μέσος θα περιέχεται στο παραπάνω διάστημα είναι 95% με το 95% να ονομάζεται **επίπεδο εμπιστοσύνης**.



Υπολογισμός με το EXCEL

Δεδομένα → Ανάλυση δεδομένων → Περιγραφικά στατιστικά

Παράδειγμα

Λαμβάνουμε δείγμα 90 πελατών και σημειώνουμε τον χρόνο αναμονής στο ταμείο μιας τράπεζας. Να κατασκευασθεί ένα 95% διάστημα εμπιστοσύνης για τον μέσο χρόνο αναμονής στο ταμείο της τράπεζας.

Μέσος χρόνος αναμονής στο δείγμα → 31,66 λεπτά

Εμπιστοσύνη 95% → 3,64 λεπτά

Υπολογισμός ορίων	Κάτω όριο	=31,65-3,64=28,01
	Άνω όριο	=31,65+3,64=35,29

ο μέσος χρόνος αναμονής στο ταμείο της τράπεζας, περιέχεται στο διάστημα (28,01 , 35,29), με εμπιστοσύνη 95%.

Έλεγχοι Υποθέσεων

Βασικές έννοιες ελέγχου υποθέσεων

1. Μηδενική Υπόθεση (H_0)
2. Εναλλακτική Υπόθεση (H_1)
3. Επίπεδο σημαντικότητας
4. Παρατηρούμενο επίπεδο σημαντικότητας (p-value)
5. Λάθη τύπου I και II

Στατιστική Υπόθεση

Είναι η υπόθεση που αναφέρεται στην κατανομή μιας ή περισσότερων τυχαίων μεταβλητών.

Ο έλεγχος γίνεται από τα στοιχεία ενός δείγματος τα οποία χρησιμοποιούνται για να επιβεβαιώσουν ή να απορρίψουν την υπόθεση

Παραδείγματα

Θέλουμε να ελέγξουμε αν

- ✓ Ο μέσος χρόνος αναμονής στο ταμείο μιας εφορίας είναι μεγαλύτερος από 1 ώρα.
- ✓ Ο αριθμός υπαλλήλων που θα συνταξιοδοτηθούν θα είναι μικρότερος από 2000.

Μηδενική υπόθεση H_0 -Εναλλακτική υπόθεση H_1

H_0 : Η υπόθεση για την οποία πρέπει να υπάρχουν ισχυρές ενδείξεις για να απορριφθεί.

H_1 : Η εναλλακτική υπόθεση της H_0

Παράδειγμα

Για έναν κατηγορούμενο ο δικαστής θεωρεί

- ✓ H_0 : Ο κατηγορούμενος είναι αθώος
- ✓ H_1 : Ο κατηγορούμενος είναι ένοχος

Σφάλμα τύπου I – Σφάλμα τύπου II

	Πραγματική κατάσταση	
Απόφαση	Αληθής η H_0	Ψευδής η H_0
Απορρίπτεται η H_0	Λανθασμένη απόφαση Λάθος τύπου I	Σωστή απόφαση
Δεν απορρίπτεται η H_0	Σωστή απόφαση	Λανθασμένη απόφαση Λάθος τύπου II

Επίπεδο σημαντικότητας α

Η πιθανότητα τα στοιχεία του δείγματος να οδηγήσουν τον αποφασίζοντα, ο οποίος θεωρεί ότι η H_0 είναι αληθής σε απόρριψη της. Είναι η πιθανότητα λανθασμένης απόρριψης της H_0 που επιτρέπει ο αποφασίζων στον εαυτό του.

Παράδειγμα

Στον έλεγχο

- ✓ H_0 Ο κατηγορούμενος είναι αθώος
- ✓ H_1 Ο κατηγορούμενος είναι ένοχος

Η πιθανότητα, ενώ ο κατηγορούμενος είναι αθώος, να καταδικασθεί είναι το επίπεδο σημαντικότητας του ελέγχου

Βασικά βήματα για τον έλεγχο υπόθεσης

1. Καθορισμός της H_0
2. Επιλογή επιπέδου σημαντικότητας
3. Επιλογή μεγέθους δείγματος
4. Επιλογή της στατιστικής συνάρτησης
5. Δειγματοληψία
6. Έλεγχος αν από το συγκεκριμένο δείγμα η συνάρτηση ελέγχου παίρνει τιμή που απορρίπτει ή όχι την μηδενική υπόθεση.

Έλεγχος υπόθεσης για την μέση τιμή ενός πληθυσμού

Παράδειγμα

Λαμβάνουμε δείγμα 90 πελατών και σημειώνουμε τον χρόνο αναμονής στο ταμείο μιας τράπεζας. Να εξετάσετε αν διαφέρει ο μέσος χρόνος αναμονής στο ταμείο της τράπεζας από 34 λεπτά σε επίπεδο σημαντικότητας 5% .

Έλεγχος

$$H_0: \mu=34$$

$$H_1: \mu \neq 34$$

Μέθοδος 1

Δημιουργία 95% διάστημα εμπιστοσύνης για την μέση τιμή του χρόνου αναμονής (28,01 , 35,29) το 34 περιέχεται στο διάστημα, οπότε δεν απορρίπτουμε την μηδενική υπόθεση.

Μέθοδος 2

Δημιουργία στήλης με 90 κελιά Πληκτρολόγηση της τιμής 34 σε όλα τα κελιά της στήλης αυτής
Δεδομένα → Ανάλυση δεδομένων → *t*-test κατά ζεύγη
 $p\text{-value} = 0,20 > 0,05$, άρα δεν απορρίπτουμε την υπόθεση ότι ο μέσος χρόνος αναμονής στο ταμείο της τράπεζας είναι 34 λεπτά.

Έλεγχος ανεξαρτησίας χ^2 (1)

Ο έλεγχος ανεξαρτησίας μας βοηθά να διαπιστώσουμε αν δύο κριτήρια ταξινόμησης ως προς τα οποία εξετάζουμε τον πληθυσμό μας είναι ανεξάρτητα το ένα από το άλλο.

Παραδείγματα

- ✓ Είναι το επίπεδο ικανοποίησης από τις υπηρεσίες μιας υπηρεσίας του Δημοσίου ανεξάρτητο του φύλου;
- ✓ Είναι η καπνιστική συνήθεια ανεξάρτητη του φύλου του επιπέδου εκπαίδευσης;
- ✓ Είναι η καπνιστική συνήθεια ανεξάρτητη του επιπέδου εκπαίδευσης;

H_0 : Ανεξαρτησία

H_1 : Εξάρτηση

Προϋποθέσεις ελέγχου

Καμία αναμενόμενη συχνότητα να μην είναι μικρότερη του 1
Το πολύ το 20% των κελιών του πίνακα να έχει αναμενόμενη συχνότητα μικρότερη του 5.

Παραβίαση προϋποθέσεων ελέγχου

2x2 πίνακες → Fisher's exact test
n x m πίνακες → Συγχώνευση κελιών

Έλεγχος ανεξαρτησίας χ^2 (2)

Παράδειγμα 1°

Για την μελέτη της καπνιστικής συνήθειας, επελέγη δείγμα 49 ατόμων, στα οποία εδόθη ερωτηματολόγιο, στο οποίο μεταξύ άλλων οι ερωτώμενοι σημείωναν το φύλο και αν είναι καπνιστές ή όχι.

Έλεγχος υπόθεσης

H0: Ανεξαρτησία φύλου και καπνίσματος

H1: Εξάρτηση φύλου και καπνίσματος

	Καπνιστής		
Φύλο	Ναι	Όχι	Σύνολο
Άνδρας	20	10	30
Γυναίκα	15	4	19
Σύνολο	35	14	49

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E
1	Παρατηρούμενες συχνότητες				
2		Καπνιστής			
3	Φύλο	Ναι	Όχι	Σύνολο	
4	Άνδρας	20	10	=SUM(B4:C4)	
5	Γυναίκα	15	4	=SUM(B5:C5)	
6	Σύνολο	=SUM(B4:B5)	=SUM(C4:C5)	=SUM(D4:D5)	
7					
8	Αναμενόμενες συχνότητες				
9		Καπνιστής			
10	Φύλο	Ναι	Όχι		
11	Άνδρας	=B6*D4/D6	=C6*D4/D6		
12	Γυναίκα	=B6*D5/D6	=C6*D5/D6		
13					
14					

Έλεγχος ανεξαρτησίας χ^2 (3)

Παράδειγμα 2°

Για την μελέτη της καπνιστικής συνήθειας, επελέγη δείγμα 87 ατόμων, στα οποία εδόθη ερωτηματολόγιο στο οποίο μεταξύ άλλων οι ερωτώμενοι σημείωναν το φύλο και αν καπνίζουν πολύ, αρκετά, λίγο ή καθόλου.

Έλεγχος υπόθεσης

H_0 : Ανεξαρτησία φύλου και καπνίσματος

H_1 : Εξάρτηση φύλου και καπνίσματος

Συχνότητες					Σύνολο
	Καθόλου	Λίγο	Αρκετά	Πολύ	
Γυναίκα	20	1	10	40	71
Άνδρας	10	1	0	5	16
Σύνολο	30	2	10	45	87

Αναμενόμενες συχνότητες				
	Καθόλου	Λίγο	Αρκετά	Πολύ
Γυναίκα	24,48	1,63	8,16	36,72
Άνδρας	5,52	0,37	1,84	8,28

Έλεγχος ανεξαρτησίας χ^2 (4)

Παράδειγμα 2^ο Πίνακες μετά την συγχώνευση κελιών

Παρατηρούμενες συχνότητες μετά την συγχώνευση				
	Καθόλου	Μέτρια	Πολύ	Σύνολο
Γυναίκα	20	11	40	71
Άνδρας	10	1	5	16
Σύνολο	30	12	45	87

Αναμενόμενες συχνότητες μετά την συγχώνευση			
	Καθόλου	Μέτρια	Πολύ
Γυναίκα	24,48	9,79	36,72
Άνδρας	5,517	2,21	8,28

Άσκηση

Το Υπουργείο Παιδείας και Θρησκευμάτων θέλει να ερευνήσει τον χρόνο που ξοδεύουν οι μαθητές Γυμνασίου στο Διαδίκτυο και τις συνέπειές που αυτό έχει στην επίδοσή τους. Για τον λόγο αυτό διεξήγαγε δειγματοληπτική έρευνα,, στην οποία εκτός άλλων χαρακτηριστικών, καταγράφηκε το φύλο (μεταβλητή Φύλο), αν χρησιμοποιούσαν το διαδίκτυο στο Δημοτικό (μεταβλητή Χρήση στο Δημοτικό), οι ώρες που ξοδεύει ο μαθητής στο διαδίκτυο την ημέρα (μεταβλητή Χρόνος στο διαδίκτυο), και η βαθμολογία του μαθητή σε ένα τεστ (μεταβλητή Βαθμολογία) για την μέτρηση της επίδοσής του.

Τα δεδομένα της άσκησης, βρίσκονται στο αρχείο με όνομα Δεδομένα στο φύλλο με το όνομα Άσκηση_4

1. Να υπολογισθεί ένα διάστημα εμπιστοσύνης για τον μέσο αριθμό ωρών ημερήσιας χρήσης του διαδικτύου από τους μαθητές.
2. Να ελεγχθεί αν ο μέσος χρόνος χρήσης του διαδικτύου διαφέρει από τις 2 ώρες ημερησίως
3. Να ελεγχθεί η ανεξαρτησία φύλου και χρήσης του διαδικτύου στο δημοτικό

5. Συσχέτιση-Παλινδρόμηση

1. Συσχέτιση δύο μεταβλητών
 - a. Συντελεστής συσχέτισης
 - b. Έλεγχος υπόθεσης για τον συντελεστή συσχέτισης
2. Παλινδρόμηση
 - a. Απλή γραμμική παλινδρόμηση
 - b. Έλεγχοι υποθέσεων για το μοντέλο

Μαθησιακοί στόχοι

Όταν θα έχετε ολοκληρώσει την μελέτη της ενότητας θα είστε σε θέση να

- ▷ Υπολογίζετε και ερμηνεύετε την γραμμική συσχέτιση δύο μεταβλητών
- ▷ Κατασκευάζετε ένα μοντέλο γραμμικής παλινδρόμησης
- ▷ Να κάνετε συμπερασματολογία για το μοντέλο

Συσχέτιση

Αν και η συσχέτιση είναι μια ευρεία έννοια που περιλαμβάνει και την εξάρτηση και υλοποιείται με διάφορα στατιστικά μέτρα, συνήθως εκφράζει το βαθμό στον οποίο υποκρύπτεται γραμμική σχέση μεταξύ δύο ποσοτικών μεταβλητών. Το διάγραμμα διασποράς γραμμικά συσχετισμένων ποσοτικών μεταβλητών αποτελείται από σημεία τα οποία θα τείνουν να συσσωρεύονται γύρω από μία ευθεία γραμμή.

Παραδείγματα:

- ✓ Συσχέτιση μεταξύ των αναστημάτων δύο αδελφών.
- ✓ Συσχέτιση ύψους και βάρους

Μέτρηση έντασης γραμμικής σχέσης Συντελεστής συσχέτισης $\rightarrow \rho$

Τιμές συντελεστή συσχέτισης $-1 \leq \rho \leq 1$

$\rho = -1$ τέλεια αρνητική συσχέτιση

$-1 < \rho < 0$ αρνητική συσχέτιση

$\rho = 0$ ανυπαρξία γραμμικής συσχέτισης

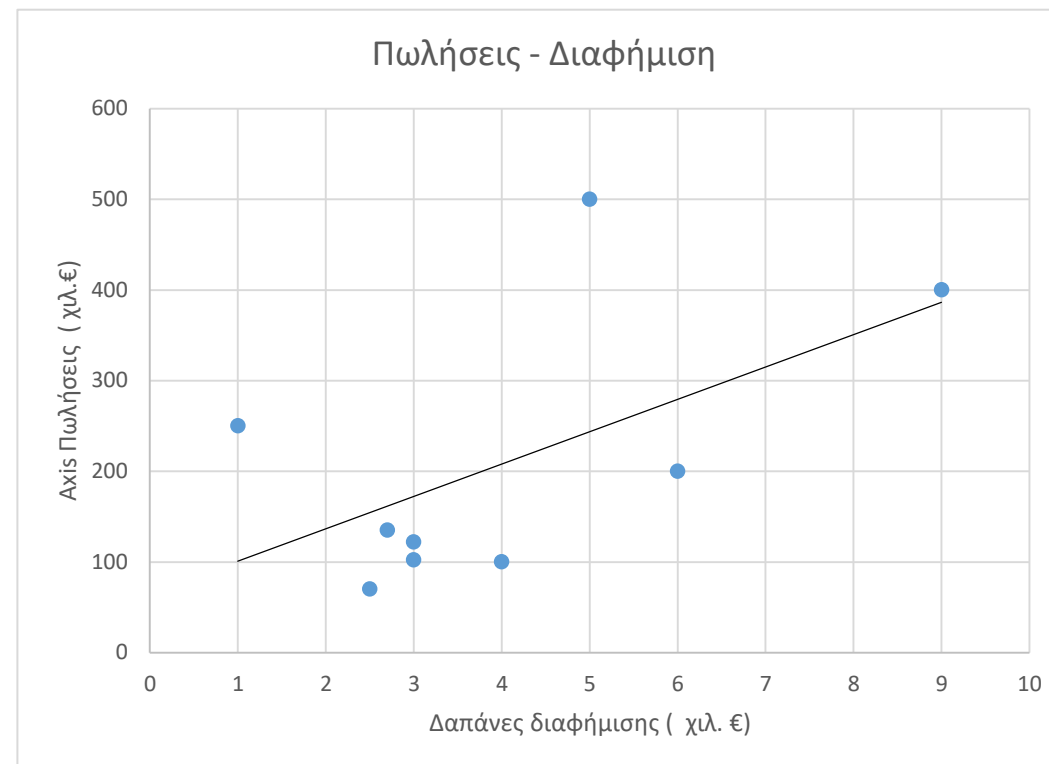
$0 < \rho < 1$ θετική συσχέτιση

$\rho = 1$ τέλεια θετική συσχέτιση

Συσχέτιση

Παράδειγμα

Έτος	Δαπάνες διαφήμισης (σε χιλ. €)	Πωλήσεις (σε χιλ.€)
2010	1	250
2011	3	122
2012	2,7	135
2013	6	200
2014	9	400
2015	5	500
2016	4	100
2017	3	102
2018	2,5	70



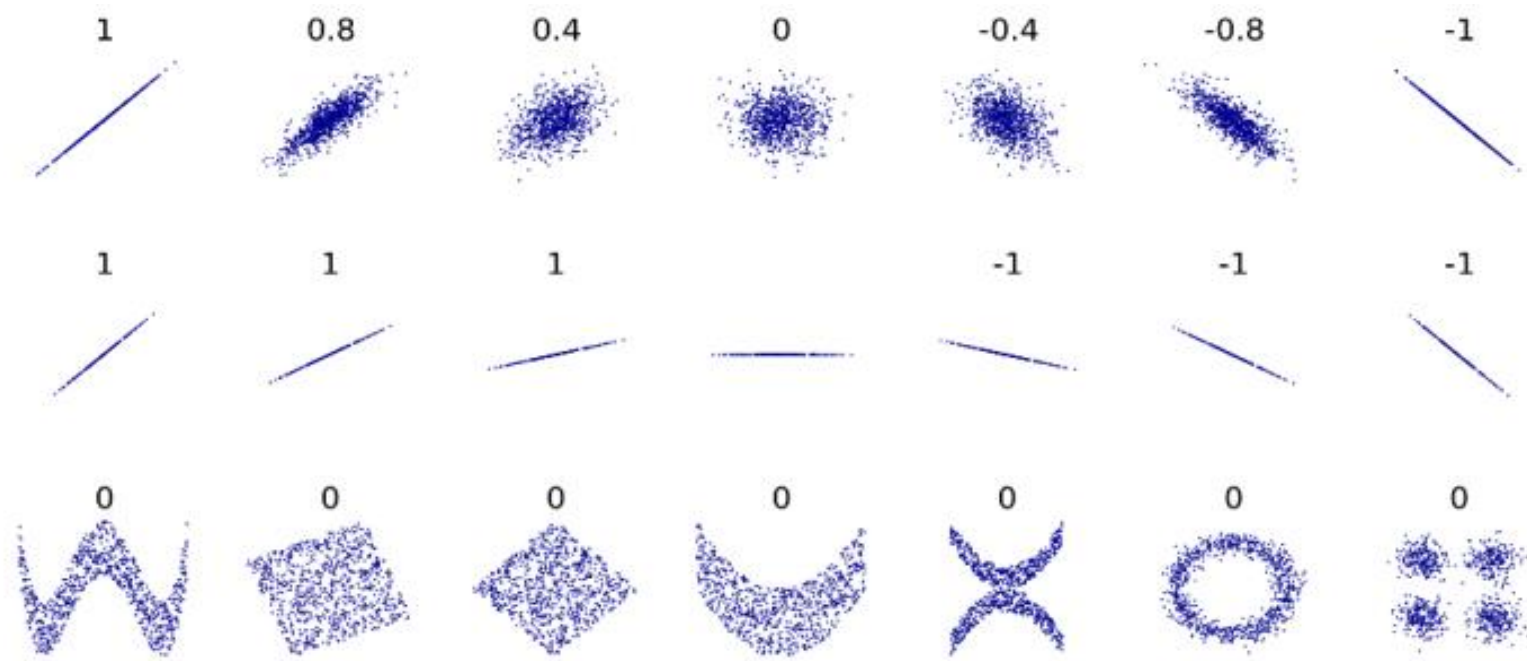
Συντελεστής συσχέτισης=0,91



Ισχυρή θετική συσχέτιση

Συσχέτιση

Όταν δύο μεταβλητές δεν έχουν γραμμική σχέση, δεν σημαίνει ότι αυτές δεν έχουν μία άλλη μορφή σχέσης.



Πηγή Wikipedia

Ανάλυση Παλινδρόμησης

Με την **Ανάλυση παλινδρόμησης** εξετάζουμε την σχέση μεταξύ δύο ή περισσότερων μεταβλητών με στόχο την πρόβλεψη των τιμών μιας από τις μεταβλητές μέσω των τιμών των υπολοίπων μεταβλητών.

Παραδείγματα:

Πρόβλεψη

- ✓ Ύψος γιου από το ύψος πατέρα
- ✓ Βάρους από το ύψος
- ✓ Προσέλευσης τουριστών από διάφορους παράγοντες
- ✓ Συνολικής παραγωγής ελαιόλαδου με βάση το ύψος της βροχόπτωσης και το βάρος του καταναλωθέντος λιπάσματος

Εξαρτημένη μεταβλητή

Η μεταβλητή της οποίας προβλέπουμε τις τιμές

Ανεξάρτητη/ές μεταβλητή/ές

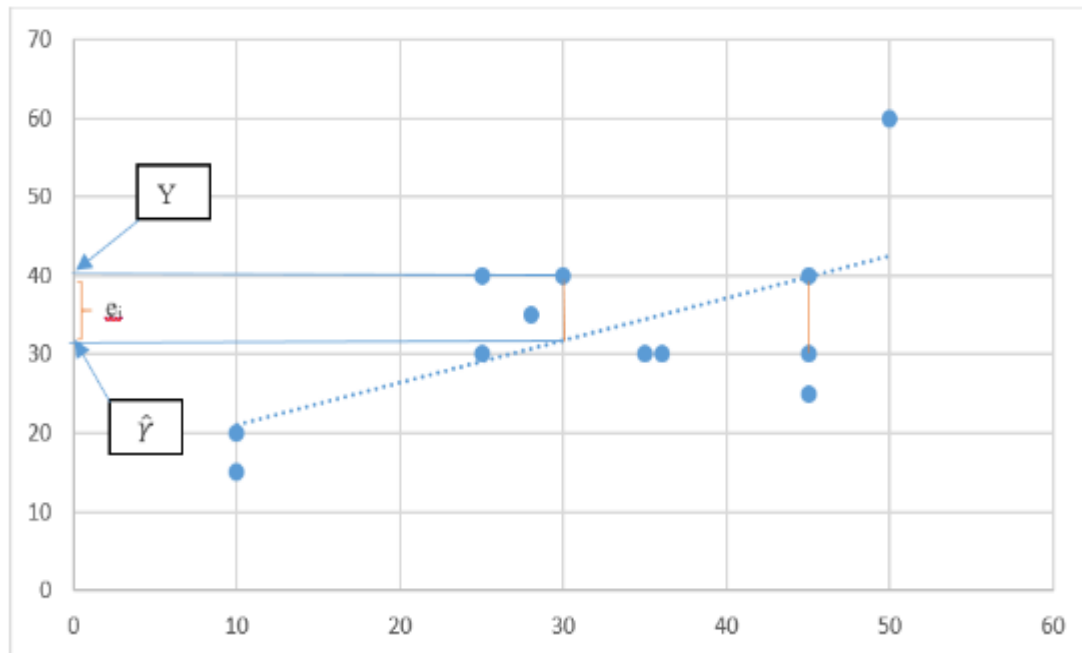
Η μεταβλητή/ες βάσει των οποίων προβλέπουμε τις τιμές της εξαρτημένης μεταβλητής

Ανάλυση Παλινδρόμησης

Μοντέλο Γραμμικής Παλινδρόμησης

$$Y_i = \alpha + \beta * X_i + e_i.$$

$$\hat{Y} = \alpha + \beta * X$$



Συντελεστής προσδιορισμού→ R^2

Μέτρο για να διαπιστώσουμε πόσο καλά προσαρμόζεται η ευθεία στα δεδομένα μας. Εκφράζει % της συνολικής μεταβλητότητας που ερμηνεύεται από το μοντέλο.

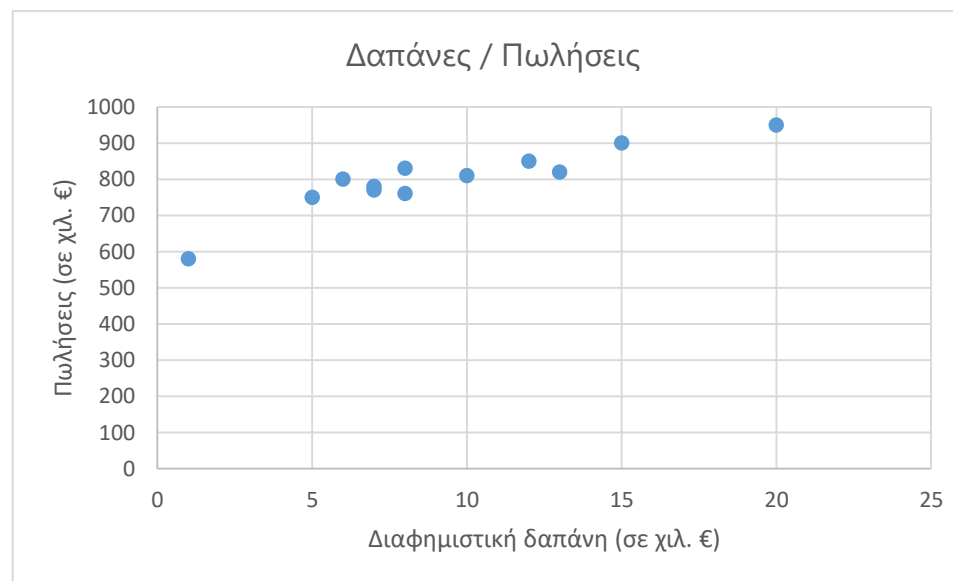
$$0 \leq R^2 \leq 1$$

Ανάλυση Παλινδρόμησης

Παράδειγμα

Λαμβάνουμε δείγμα 12 εταιριών και σημειώνουμε τις δαπάνες διαφήμισης και τις πωλήσεις τους..Να εξετάσετε αν επιδρά η διαφήμιση στις πωλήσεις

A/A	Δαπάνες διαφήμισης (σε χιλ. €)	Πωλήσεις (σε χιλ.€)
1	6	800
2	8	830
3	12	850
4	15	900
5	20	950
6	5	750
7	7	780
8	8	760
9	7	770
10	10	810
11	13	820
12	1	580



Ανάλυση Παλινδρόμησης

Υπολογισμός με το EXCEL

Δεδομένα → Ανάλυση Δεδομένων → Παλινδρόμηση

- ✓ Στην περιοχή εισόδου Y τις τιμές της εξαρτημένης μεταβλητής 'Πωλήσεις'
- ✓ Στην περιοχή εισόδου X τις τιμές της ανεξάρτητης μεταβλητής 'Δαπάνες Διαφήμισης'
- ✓ Επιλογή εμφάνισης των καταλοίπων
- ✓ Επιλογή εμφάνισης των γραφημάτων

Συντελεστής προσδιορισμού $R^2 = 0,82 = 82\%$

Άρα το 82% της ολικής μεταβλητότητας ερμηνεύεται από την παλινδρόμηση

	Συντελεστές	Τυπικό σφάλμα	t Stat	P-value
Σταθερά (α)	648,36	25,19	25,74	0,00
Δαπάνες διαφήμισης (β)	16,25	2,40	6,78	0,00

Εξίσωση της ευθείας παλινδρόμησης

$$\text{Πωλήσεις} = 16,25 * \text{Δαπάνες διαφήμισης} + 648,36$$

Αύξηση 1000€ για διαφήμιση → $16,25 * 1000 = 16250€$ αύξηση των πωλήσεων

Ανάλυση Παλινδρόμησης

Εξίσωση της ευθείας παλινδρόμησης

$$\text{Πωλήσεις} = 16,25 \cdot \text{Δαπάνες διαφήμισης} + 648,36$$

Έλεγχος υπόθεσης για την σημαντικότητα της παλινδρόμησης

$$H_0: \beta = 0$$

$$H_1: \beta \neq 0$$

Απόρριψη της H_0 → Στατιστικά σημαντική Παλινδρόμηση

Απόφαση

P-value = 0,00 < 0,05, άρα η παλινδρόμηση είναι στατιστικά σημαντική

	Συντελεστές	Τυπικό σφάλμα	t Stat	P-value
Σταθερά (α)	648,36	25,19	25,74	0,00
Δαπάνες διαφήμισης (β)	16,25	2,40	6,78	0,00

Άσκηση

Το Υπουργείο Παιδείας και Θρησκευμάτων θέλει να ερευνήσει τον χρόνο που ξοδεύουν οι μαθητές Γυμνασίου στο Διαδίκτυο και τις συνέπειές που αυτό έχει στην επίδοσή τους. Για τον λόγο αυτό διεξήγαγε δειγματοληπτική έρευνα, στην οποία εκτός άλλων χαρακτηριστικών, καταγράφηκε το φύλο (μεταβλητή Φύλο), αν χρησιμοποιούσαν το διαδίκτυο στο Δημοτικό (μεταβλητή Χρήση στο Δημοτικό), οι ώρες που ξοδεύει ο μαθητής στο διαδίκτυο την ημέρα (μεταβλητή Χρόνος στο διαδίκτυο), και η βαθμολογία του μαθητή σε ένα τεστ (μεταβλητή Βαθμολογία) για την μέτρηση της επίδοσής του.

Τα δεδομένα της άσκησης, βρίσκονται στο αρχείο με όνομα Δεδομένα στο φύλλο με το όνομα Άσκηση_4

1. Να υπολογισθεί ο συντελεστής συσχέτισης για τις μεταβλητές, Χρόνος στο διαδίκτυο και Βαθμολογία και να σχολιασθεί η συσχέτιση των δύο μεταβλητών.
2. Να κατασκευασθεί μοντέλο παλινδρόμησης για την επίδραση του χρόνου που περνούν οι μαθητές στο διαδίκτυο και την βαθμολογία τους στο τεστ.
3. Είναι η παλινδρόμηση στατιστικά σημαντική;
4. Να γραφεί η εξίσωση της παλινδρόμησης και να ερμηνευθούν οι συντελεστές.
5. Να δοθεί η ερμηνεία του συντελεστή προσδιορισμού.

Βιβλιογραφία

Καστανιά, Α. & Αποστολάκης, Ι. & Πιερράκου, Χ (2003) Στατιστική επεξεργασία δεδομένων στην υγεία, Αθήνα Παπαζήσης

Κιόχος, Π. (1985). Οικονομική Στατιστική, Πειραιάς Σταμούλης

Παπαϊωάννου, Π. & Λουκάς, Σ. (2002). Εισαγωγή στην Στατιστική, Πειραιάς Σταμούλης

Douglas Downing, Jeffrey Clark ; μετάφραση Παναγιώτης Σταυρόπουλος, Γιώργος Σταυρόπουλος.(1998), Αθήνα Κλειδάριθμος

Κουτροβέλης Ι. (1999). Πιθανότητες και Στατιστική Ι. Ελληνικό Ανοιχτό Πανεπιστήμιο.

Κουτροβέλης Ι. (1999). Πιθανότητες και Στατιστική ΙΙ. Ελληνικό Ανοιχτό Πανεπιστήμιο.

Witte, R. and Witte, J. (2016). *Statistics*. 11th ed. Wiley.

Madsen, B. (2016). *Statistics for non-statisticians*. Springer-Verlag Berlin Heidelberg.

Cox, V. (2017). *Translating statistics to make decisions*. Apress.

<https://support.office.com/el-gr/excel>

https://imegsevee.gr/wp-content/uploads/2018/02/ypologistika_baseis_dedomenwn.pdf

<https://www.youtube.com/watch?v=nDxo1gT8LGY>

<https://www.youtube.com/watch?v=x7FNlq6FgcM>

<https://www.statistics.gr/>

<https://ec.europa.eu/eurostat/statistics-explained/>

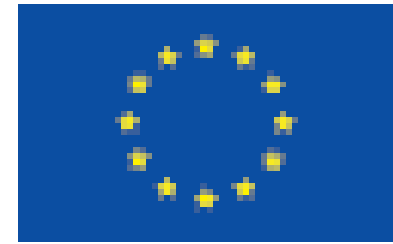
<https://www.youtube.com/watch?v=yx5KZi5QArQ>

[http://www.statistics.gr/documents/20181/995824/%CE%9C%CE%B5%CF%84%CE%B1%CE%B4%CE%B5%CE%B4%CE%BF%CE%BC%CE%AD%CE%BD%CE%B1+%CE%A3%CF%84%CE%B1%CF%84%CE%B9%CF%83%CF%84%CE%B9%CE%BA%CF%89%CE%BD+%CE%9C%CE%B5%CF%84%CE%B1%CE%BD%CE%B1%CF%83%CF%84%CE%B5%CF%85%CF%84%CE%B9%CE%BA%CE%AE%CF%82+%CE%9A%CE%AF%CE%BD%CE%B7%CF%83%CE%B7%CF%82+\(+2008+\).pdf/e8b00b6a-36f2-4cfc-b9f6-5336f0d92837?t=1445088749621](http://www.statistics.gr/documents/20181/995824/%CE%9C%CE%B5%CF%84%CE%B1%CE%B4%CE%B5%CE%B4%CE%BF%CE%BC%CE%AD%CE%BD%CE%B1+%CE%A3%CF%84%CE%B1%CF%84%CE%B9%CF%83%CF%84%CE%B9%CE%BA%CF%89%CE%BD+%CE%9C%CE%B5%CF%84%CE%B1%CE%BD%CE%B1%CF%83%CF%84%CE%B5%CF%85%CF%84%CE%B9%CE%BA%CE%AE%CF%82+%CE%9A%CE%AF%CE%BD%CE%B7%CF%83%CE%B7%CF%82+(+2008+).pdf/e8b00b6a-36f2-4cfc-b9f6-5336f0d92837?t=1445088749621)



ανάπτυξη - εργασία - αλληλεγγύη

Επιχειρησιακό Πρόγραμμα
ΜΕΤΑΡΡΥΘΜΙΣΗ
ΔΗΜΟΣΙΟΥ ΤΟΜΕΑ



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης