



**ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ**  
**ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ «ΕΦΑΡΜΟΣΜΕΝΗΣ ΟΙΚΟΝΟΜΙΚΗΣ ΚΑΙ  
ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗΣ»

ΚΑΤΕΥΘΥΝΣΗ

**«ΔΙΟΙΚΗΣΗ, ΑΝΑΛΥΤΙΚΗ ΚΑΙ ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ ΕΠΙΧΕΙΡΗΣΕΩΝ»**

Master of Science in  
Business Administration, Analytics and Information Systems

**Data Analysis (Business Statistics)**

**Assignment 2 : Data sampling**

**Κωνσταντίνος Κουτσομπίνας**

Αθήνα 01/10/2024

## Ερμηνευτικά σχόλια

Σημείωση : Εφόσον οι υπολογισμοί έχουν δοθεί ως λύσεις , παραθέτω μόνο τα ερμηνευτικά σχόλια και τις απαντήσεις στις ερμηνευτικές ερωτήσεις της εργασίας.

- Παρατηρούμε ότι τα data μας έχουν αρκετά μεγάλη διασπορά (από το range , το variance και το st.dev. ) .
- Ο μέσος όρος είναι σχετικά χαμηλά.
- Υπάρχει σημαντική λοξότητα των δεδομένων προς τα δεξιά
- Τα μέτρα αυτά μαρτυρούν ότι μερικοί δήμοι παρουσιάζουν μεγάλη διαφορά προς τα πάνω , ως προς το δείκτη ανεργίας . Μπορούμε λοιπόν να σημειώσουμε ότι η πολιτεία πρέπει να διερευνήσει τους λόγους για τους οποίους παρουσιάζεται αυτή η διαφοροποίηση στους εν λόγω δήμους και να βοηθήσει στην ένταξη των κατοίκων τους στην αγορά εργασίας.
- Τα ομαδοποιημένα data φανερώνουν μερικές διαφοροποιήσεις σχετικά με τα αρχικά . Ωστόσο τα κύρια χαρακτηριστικά είναι αρκετά κοντά στα αρχικά , όπως φανερώνεται και στα διαγράμματα. Ο λόγος αυτής της εμφάνισης διαφοροποιήσεων , είναι ότι τα ομαδοποιημένα δεδομένα επικεντρώνονται στα μέσα των κλάσεων , στο μέγεθος και στις συχνότητές τους , και όχι σε όλες τις τιμές των αρχικών δεδομένων. Ουσιαστικά όσο πιο μεγάλο εύρος δίνουμε στις κλάσεις , τόσο αποκρύπτεται ή «χάνεται» πληροφορία από τα αρχικά data. Ταυτόχρονα πρέπει ο αριθμός των κλάσεων να είναι μικρός αριθμός , ούτως ώστε να επιτρέπεται η ευκολότερη ανάγνωση , ανάλυση και διερμηνευση των data. Συνεπώς πρέπει κατά την ομαδοποίηση να βρούμε των αριθμό των κλάσεων που θα μας επιτρέψει να ομαδοποιήσουμε τα data ισορροπώντας το tradeoff αυτό.