



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 τον αριθμό των κλάσεων που θα μας επιτρέψει να ομαδοποιήσουμε τα data ισορροπώντας το tradeoff αυτό.

Ερώτημα 11

Αν υποθέσουμε ότι η κατανομή της ανεργίας ακολουθάει κανονική κατανομή με $\text{mean}=9.404\%$ / $\text{st.dev.} = 2.595\%$

Τότε για να βρούμε τον αριθμό των δήμων με ποσοστό ανεργίας μεγαλύτερο από 12.73% θα υπολογίσουμε :

- Το Z που αντιστοιχεί στο 12.73% της κανονικής κατανομής

$$Z = \frac{X - \mu}{\sigma} = \frac{0.1273 - 0.09404}{0.02595} \approx 1.281$$
- Το $\Phi(Z)$ από τον πίνακα Fisher-Yates ή από το normsdist() function του excel:

$$\Phi(Z) = 0.100010846866978 \approx 10\%$$
- Και τελικά τον αριθμό των δήμων το οποίο είναι το 10% του συνόλου δηλαδή

$$10\% \times 123 = 12.3 \approx 12 \text{ δήμοι .}$$

Παρατηρούμε ότι στην πραγματικότητα οι δήμοι που έχουν ποσοστό ανεργίας πάνω από 12.73% είναι 14 δηλαδή το 11.4% επί του συνόλου και όχι το 10% . Αυτό είναι αποτέλεσμα του μικρού δείγματος , καθώς και του γεγονότος ότι η κατανομή του δείγματος δεν είναι κανονική (όπως παρατηρήσαμε έχει λοξότητα προς τα δεξιά).