**Πολυµεταβλητή Ανάλυση**

**Εργασία 1**

**Μέρος Α: Εισαγωγή στην Πολυμεταβλητή ανάλυση με την R**

Έστω η τυχαία μεταβλητή η οποία ακολουθεί πολυδιάστατη κανονική κατανομή με μέσο και πίνακα συσχετίσεων , με τις διασπορές των να είναι (i) όλες ίσες με 2 και (ii) 2, 3, 4, 5, αντίστοιχα.

1. Ορίζουμε νέες μεταβλητές , όπου , , και , με και . Να βρεθούν με την βοήθεια της R οι πίνακες συσχετίσεων των μεταβλητών και καθώς και η συσχέτιση μεταξύ των και για τις περιπτώσεις (i) και (ii). Είναι ίδιες οι συσχετίσεις στις δύο περιπτώσεις;
2. Να προσομοιώσετε ένα δείγμα μεγέθους 10000 από την τυχαία μεταβλητή για την περίπτωση (ii). Να βρεθεί η δειγματική μέση τιμή, ο δειγματικός πίνακας συνδιακύμανσης και ο δειγματικός πίνακας συσχέτισης της τυχαίας μεταβλητής . Να σχολιάσετε τα αποτελέσματα σε σχέση με τα αποτελέσματα του ερωτήματος 1.

**Μέρος Β: Ανάλυση σε Κύριες Συνιστώσες**

**Περιγραφή δεδομένων**

Το αρχείο dietStudy αφορά μια μελέτη για την συσχέτιση διαφόρων παραγόντων κινδύνου με το ενδονοσοκομειακό θάνατο έπειτα από οξύ έμφραγμα του μυοκαρδίου. Περιέχει τις παρακάτω μεταβλητές:

* **legumes:** Κατανάλωση οσπρίων (φορές / εβδομάδα)
* **vegetabl:** Κατανάλωση λαχανικών (φορές / εβδομάδα)
* **salads:** Κατανάλωση σαλάτας (φορές / εβδομάδα)
* **meat:** Κατανάλωση κρέατος (φορές/εβδομάδα)
* **chicken:** Κατανάλωση κοτόπουλου ( φορές/εβδομάδα)
* **fish:** Κατανάλωση ψαριού: (φορές./εβδομάδα)
* **death:** ενδονοσοκομειακός θάνατος (1=ναι, 0=όχι)

Να απαντηθούν τα παρακάτω ερωτήματα (η μεταβλητή death να μην συμπεριληφθεί στην ανάλυση κυρίων συνιστωσών):

1. Ελέγξτε αν υπάρχουν ικανές συσχετίσεις μεταξύ των μεταβλητών ώστε να ικανοποιούνται οι προϋποθέσεις εφαρμογής ανάλυσης σε κύριες συνιστώσες.
2. Να πραγματοποιηθεί ανάλυση σε κύριες συνιστώσες χρησιμοποιώντας τον πίνακα συνδιακύμανσης.
   1. Ποιες είναι οι ιδιοτιμές του πίνακα συνδιακύμανσης;
   2. Ποιο είναι το ποσοστό της μεταβλητότητας των αρχικών μεταβλητών που ερμηνεύει η κάθε συνιστώσα; Πώς υπολογίστηκε;
   3. Ποια είναι τα ιδιοδιανύσματα που αντιστοιχούν στις ιδιοτιμές του πίνακα συνδιακύμανσης; Τι αντιπροσωπεύουν αυτά τα ιδιοδιανύσματα;
   4. Πόσες από τις κύριες συνιστώσες που προέκυψαν θα κρατήσετε εφαρμόζοντας: i) τον κανόνα του ποσοστού συνολικής διακύμανσης που εξηγούν οι συνιστώσες (π.χ 80%), ii) το κριτήριο του Kaiser και iii) το scree plot.
   5. Ερμηνεύσετε κάθε μία από τις κύριες συνιστώσες τις οποίες θα διατηρούσατε στην ανάλυση.
   6. Δημιουργήστε τόσες νέες μεταβλητές όσες και οι κύριες συνιστώσες που θα διατηρηθούν και ελέγξτε τη συσχέτιση κάθε μιας από αυτές με την ενδονοσοκομειακή θνητότητα των ασθενών της μελέτης. Σχολιάστε και ερμηνεύσετε τα αποτελέσματα.