**Πολυµεταβλητή Ανάλυση**

**Εργασία 2**

**Μέρος Α: Παραγοντική Ανάλυση**

Θα αναλύσετε με παραγοντική ανάλυση τα δεδομένα του αρχείου dietStudy (περιγραφή στην εκφώνηση της εργασίας 1).

Να απαντηθούν τα παρακάτω ερωτήματα (η μεταβλητή death να μην συμπεριληφθεί στην παραγοντική ανάλυση. Μπορεί να χρησιμοποιηθεί περιγραφικά.):

1. Να βρεθούν και να σχολιαστούν τα μονομεταβλητά και πολυμεταβλητά περιγραφικά μέτρα για κάθε μεταβλητή.

2. Χρησιμοποιώντας τον πίνακα συσχέτισης (τυποποιημένα δεδομένα), να γίνει παραγοντική ανάλυση χρησιμοποιώντας τη μέθοδο των κυρίων συνιστωσών, με αριθμό παραγόντων ίσο με δύο. Να υπολογισθούν οι κοινές διασπορές (communalities) των μεταβλητών χωρίς στροφή των παραγόντων.

3. Να επαναληφθεί το ερώτημα 2 χρησιμοποιώντας στροφή παραγόντων. Να γίνουν διαγράμματα φορτίων για κάθε στροφή και να επιλέξετε την στροφή εκείνη που κατά τη γνώμη σας επιτρέπει την καλύτερη ερμηνεία των δύο παραγόντων. Πώς ερμηνεύετε του παράγοντες;

4. Για την στροφή που επιλέξατε στο ερώτημα 3 να υπολογίσετε τα σκορ των δύο παραγόντων. Να γίνει γράφημα διασποράς (scatter) των δύο factor scores. Με βάση το διάγραμμα και την ερμηνεία των παραγόντων, υπάρχει κάποια ομαδοποίηση των συμμετεχόντων στη μελέτη; Συνδέεται αυτή με τη μεταβλητή death;

5. Δοκιμάστε όποιες άλλες μεθόδους εκτίμησης και αριθμό παραγόντων θέλετε για το παραγοντικό μοντέλο και παρουσιάστε συνοπτικά τα αποτελέσματα για το βέλτιστο (κατά τη γνώμη σας) μοντέλο-μέθοδο αν είναι διαφορετικό από αυτό στο οποίο καταλήξατε στα ερωτήματα 1-4.

**Μέρος Β: Διακρίνουσα Ανάλυση**

**Περιγραφή δεδομένων**

Για τη δημιουργία μιας διαδικασίας ανίχνευσης πιθανών φορέων αιμορροφιλίας Α, αναλύθηκαν δείγματα αίματος από δύο ομάδες γυναικών και καταγράφηκαν οι μετρήσεις σε δύο μεταβλητές (με AHF συμβολίζουμε τον αντιαιμοφιλικό παράγοντα):

* **Χ1:** log ( δραστηριότητα AHF )
* **Χ2:** log ( αντιγόνο τύπου-AHF )

Η πρώτη ομάδα επιλέχθηκε από έναν πληθυσμό γυναικών που δεν ήταν φορείς του γονιδίου της αιμορροφιλίας και ονομάστηκε «κανονική». Η δεύτερη ομάδα επιλέχθηκε από έναν πληθυσμό γυναικών γνωστών φορέων αιμορροφιλίας Α (κόρες αιμορροφιλικών, μητέρες με πάνω από έναν αιμορροφιλικό γιο και άλλους αιμορροφιλικούς συγγενείς). Η ομάδα αυτή ονομάστηκε «φορείς».

Χρησιμοποιήστε τα δεδομένα του παρακάτω πίνακα για να απαντήσετε τα ακόλουθα ερωτήματα.

**Μη φορείς** **Φορείς**

|  |  |  |  |
| --- | --- | --- | --- |
| Χ1 | Χ2 | Χ1 | Χ2 |
| -0.0056 | -0.1657 | -0.3478 | 0.1151 |
| -0.1698 | -0.1585 | -0.3618 | -0.2008 |
| -0.3468 | -0.1879 | -0.4986 | -0.0860 |
| -0.0894 | 0.0064 | -0.5015 | -0.2984 |
| -0.1679 | 0.0713 | -0.1326 | 0.0097 |
| -0.0836 | 0.0106 | -0.6911 | -0.3390 |
| -0.1979 | -0.0005 | -0.3608 | 0.1237 |
| -0.0762 | 0.0392 | -0.4535 | -0.1682 |
| -0.1913 | -0.2123 | -0.3479 | -0.1721 |
| -0.1092 | -0.1190 | -0.3539 | 0.0722 |
| -0.5268 | -0.4773 | -0.4719 | -0.1079 |
| -0.0842 | 0.0248 | -0.3610 | -0.0399 |
| -0.0225 | -0.0580 | -0.3226 | 0.1670 |
| 0.0084 | 0.0782 | -0.4319 | -0.0687 |
| -0.1827 | -0.1138 | -0.2734 | -0.0020 |
| 0.1237 | 0.2140 | -0.5573 | 0.0548 |
| -0.4702 | -0.3099 | -0.3755 | -0.1865 |
| -0.1519 | -0.0686 | -0.4950 | -0.0153 |
| 0.0006 | -0.1153 | -0.5107 | -0.2483 |
| -0.2015 | -0.0498 | -0.1652 | 0.2132 |
| -0.1932 | -0.2293 | -0.2447 | -0.0407 |
| 0.1507 | 0.0933 | -0.4232 | -0.0998 |
| -0.1259 | -0.0669 | -0.2375 | 0.2876 |
| -0.1551 | -0.1232 | -0.2205 | 0.0046 |
| -0.1952 | -0.1007 | -0.2154 | -0.0219 |
| 0.0291 | 0.0442 | -0.3447 | 0.0097 |
| -0.2228 | -0.1710 | -0.2540 | -0.0573 |
| -0.0997 | -0.0733 | -0.3778 | -0.2682 |
| -0.1972 | -0.0607 | -0.4046 | -0.1162 |
| -0.0867 | -0.0560 | -0.0639 | 0.1569 |
|  |  | -0.3351 | -0.1368 |
|  |  | -0.0149 | 0.1539 |
|  |  | -0.0312 | 0.1400 |
|  |  | -0.1740 | -0.0776 |
|  |  | -0.1416 | 0.1642 |
|  |  | -0.1508 | 0.1137 |
|  |  | -0.0964 | 0.0531 |
|  |  | -0.2642 | 0.0867 |
|  |  | -0.0234 | 0.0804 |
|  |  | -0.3352 | 0.0875 |
|  |  | -0.1878 | 0.2510 |
|  |  | -0.1744 | 0.1892 |
|  |  | -0.4055 | -0.2418 |
|  |  | -0.2444 | 0.1614 |
|  |  | -0.4784 | 0.0282 |

1. Να υπολογίσετε τα κέντρα βάρους και τους δειγματικούς πίνακες συσχέτισης για τους δύο πληθυσμούς. Κάντε περιγραφική ανάλυση των δεδομένων για να ερευνήσετε την καταλληλόλητα των δύο μεταβλητών για τη διάκριση των δύο πληθυσμών.
2. Να βρεθεί η διακρίνουσα συνάρτηση του Fisher που θα χρησιμοποιηθεί για τη διάκριση των δύο πληθυσμών με βάση τα 2 χαρακτηριστικά τους.
3. Να ταξινομηθούν οι παρατηρήσεις του δείγματος σύμφωνα με τον κανόνα διάκρισης του Fisher και να υπολογιστεί το ποσοστό λανθασμένης ταξινόμησης.
4. Να ταξινομηθούν οι επόμενες νέες περιπτώσεις χρησιμοποιώντας τη διακρίνουσα συνάρτηση του ερωτήματος 2.

|  |  |  |
| --- | --- | --- |
| **Περίπτωση** | **Χ1** | **Χ2** |
| 1 | -0.112 | -0.279 |
| 2 | -0.059 | -0.068 |
| 3 | 0.064 | 0.012 |
| 4 | -0.043 | -0.052 |
| 5 | -0.123 | -0.143 |
| 6 | -0.210 | -0.090 |

1. Ερευνήστε την υπόθεση της διμεταβλητής κανονικότητας για τις δύο ομάδες.
2. Επαναλάβετε τα ερωτήματα 1-4 χρησιμοποιώντας τον κανόνα διάκρισης του Bayes, θεωρώντας ότι η a-priori πιθανότητα μια γυναίκα να μην είναι φορέας είναι ¾.