ΕΡΓΑΣΙΑ 1

ΠΟΛΥΜΕΤΑΒΛΗΤΗ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

ΔΗΜΗΤΡΙΟΣ ΓΚΑΒΕΡΑΣ
ΑΛΕΞΑΝΔΡΟΣ ΣΚΟΝΔΡΑΣ

Δημήτριος Γκαβέρας 1112201500042
Αλέξανδρος Σκόνδρας 1112201500206

ΕΡΓΑΣΙΑ 1

ΠΡΩΤΗ ΑΣΚΗΣΗ

1)

Σ1=

3.0000	0.6000	1.2000	2.4000
0.6000	3.0000	1.5000	1.8000
1.2000	1.5000	3.0000	2.1000
2.4000	1.8000	2.1000	3.0000

Σ2=

1.0000	0.2828	0.6928	1.6000
0.2828	2.0000	1.2247	1.6971
0.6928	1.2247	3.0000	2.4249
1.6000	1.6971	2.4249	4.0000

Με τη βοήθεια της MATLAB υπολογίζουμε τους πίνακες συνδιακύμανσης για τις περιπτώσεις i) και ii) αντίστοιχα.
Ο X* στο i) ακολουθεί πολυδιάστατη κανονική κατανομή με μέσο (0 0 0 0) και πίνακα συνδιακύμανσης Σ1.
Ο X* στο ii) ακολουθεί πολυδιάστατη κανονική κατανομή με μέσο (0 0 0 0) και πίνακα συνδιακύμανσης Σ2.
Ο $\tilde{X}$ στα i) και ii) ακολουθεί πολυδιάστατη κανονική κατανομή με μέσο (0 0 0 0) και πίνακα συνδιακύμανσης P .

2)

Πίνακες συσχέτισης για τη μεταβλητή Y

PYi =

1.0000	-0.6951
-0.6951	1.0000

PYii =

1.0000	-0.7069
-0.7069	1.0000

Πίνακες συσχέτισης για τη μεταβλητή Z

PZi =

1.0000	-0.0000
-0.0000	1.0000

PZii =

1.0000	-0.1814
-0.1814	1.0000

Χρησιμοποιώντας το γνωστό τύπο της συσχέτισης $\frac{Cov(Y_1, Z_2)}{\sigma_{Y_1} * \sigma_{Z_2}}$ και ιδιότητες του Cov καταλήγουμε με τις εντολές

```
covY1Z2=(5*Si(1,3)-2*Si(1,1)-15*Si(2,3)+6*Si(2,1)-
25*Si(3,4)+10*Si(1,4))/(sqrt(SYi(1,1))*sqrt(SZi(2,2)))

covY1Z2=(5*Sii(1,3)-2*Sii(1,1)-15*Sii(2,3)+6*Sii(2,1)-
25*Sii(3,4)+10*Sii(1,4))/(sqrt(SYii(1,1))*sqrt(SZii(2,2)))
```

στα αποτελέσματα για το i) -0.5210 και ii) -0.6009.

(το Cov(X_1, X_3) είναι το στοιχείο (1,3) στον πίνακα Si και Sii αντίστοιχα κι ομοίως για τα υπόλοιπα Cov που εμφανίζονται στον αριθμητή του τύπου συσχέτισης. Για τον παρονομαστή χρησιμοποιούμε τους πίνακες SYi, SYii και SZi, SZii.)

Παρατηρούμε ότι οι συσχετίσεις δεν είναι ίδιες. Αυτό συμβαίνει, διότι χρησιμοποιούμε διαφορετικούς πίνακες συνδιακύμανσης.

3)

wbar =

-47.4297	5.0117	-2.0276	21.0036

DS =

142.3623	-13.5631	7.2667	-53.2818
-13.5631	2.6895	-0.6203	12.2068
7.2667	-0.6203	0.9715	-1.1584
-53.2818	12.2068	-1.1584	58.7171

DR =

1.0000	-0.6931	0.6179	-0.5828
-0.6931	1.0000	-0.3837	0.9714
0.6179	-0.3837	1.0000	-0.1534
-0.5828	0.9714	-0.1534	1.0000

4)

wbar =

-47.1860	5.0186	-2.0183	21.0564

DS =

195.8930	-20.0380	12.1542	-75.8814
-20.0380	4.0788	-1.1229	18.1480
12.1542	-1.1229	1.9968	-1.6209
-75.8814	18.1480	-1.6209	87.4981

DR =

1.0000	-0.7089	0.6145	-0.5796
-0.7089	1.0000	-0.3935	0.9606
0.6145	-0.3935	1.0000	-0.1226
-0.5796	0.9606	-0.1226	1.0000

5)

Ο κώδικας που υπολογίζει τη λογαριθμική συνάρτηση πιθανοφάνειας για τυχαίο δείγμα από την πολυδιάστατη κανονική κατανομή είναι ο παρακάτω:

```
function lp=lognormpdf1(X,m,S)
[n,p]=size(X);
lp=0;
for j=1:n
lp=lp -0.5*p*log(2*pi)+0.5*log(det(inv(S)))-
0.5*((X(j,:)-m)*inv(S)*((X(j,:)-m)'));
end
end
```

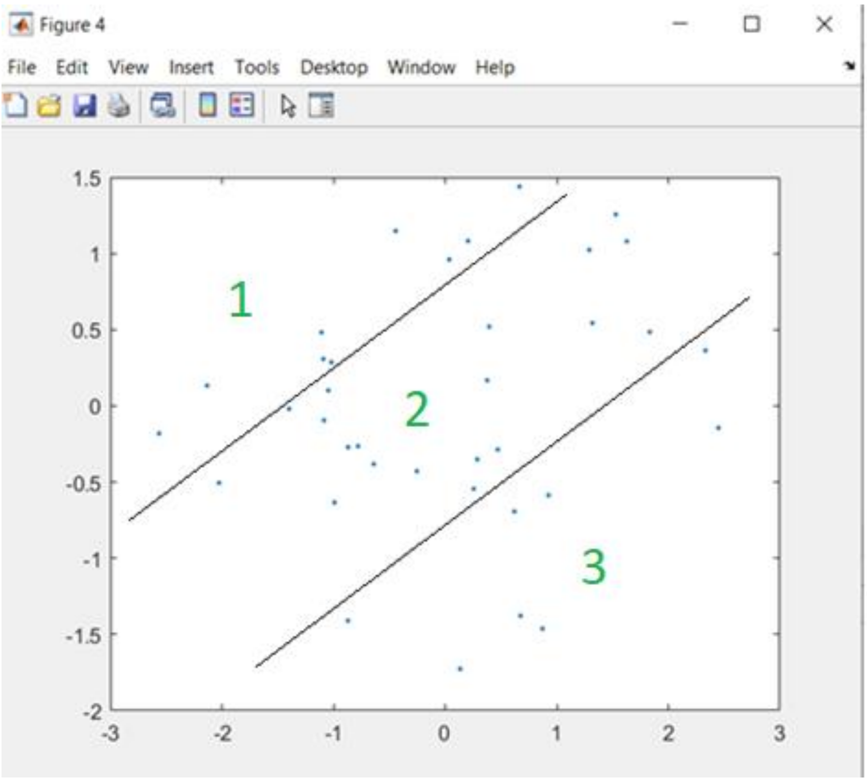
Η τιμή της λογαριθμικής πιθανοφάνειας για τα δείγματα στα ερωτήματα 3) και 4), κάθε φορά στις τιμές των ΕΜΠ των παραμέτρων, είναι αντίστοιχα:

lp =-5.8404e+03 , lp = -6.7237e+04

ΔΕΥΤΕΡΗ ΑΣΚΗΣΗ

Περίληψη Αποτελεσμάτων

Σε μια πανευρωπαϊκή έρευνα που πραγματοποιήθηκε το 1988 ζητήθηκε από φοιτητές να βαθμολογήσουν το πανεπιστήμιό τους σε διάφορα θέματα σχετικά με τη φοίτησή τους. Συγκεκριμένα βαθμολόγησαν το πανεπιστήμιό τους σε μια κλίμακα από το 1 μέχρι το 6 (1 το καλύτερο, 6 το χειρότερο) ως προς τους 8 τομείς: επαφή με έρευνα, πρακτική εξάσκηση, προετοιμασία εξετάσεων, προσωπική συμβουλευτική, κλίμα μαθητείας, εξοπλισμός υπολογιστών, διαθέσιμη βιβλιογραφία και διαθέσιμος ειδικός τύπος. Ξεκινώντας την ανάλυσή μας βρίσκουμε ότι η μέση βαθμολογία των πανεπιστημίων στα προαναφερθέντα κριτήρια είναι αντίστοιχα 3.6, 3.6, 2.8, 2.7, 3.4, 2.9, 2.4 και 2.2. Χρησιμοποιώντας μια μαθηματική μέθοδο, γνωστή ως PCA (Principal Components Analysis), κωδικοποιούμε τα αποτελέσματα της έρευνας (το δείγμα μας) σε 2 δείκτες αντιπροσωπευτικούς αυτού. Ο πρώτος δείκτης παρουσιάζει μία γενική εικόνα αξιολόγησης του κάθε πανεπιστημίου. Ο δεύτερος φανερώνει την κλίση ή μη του κάθε πανεπιστημίου στο θεωρητικό τομέα ή στον πρακτικό/ερευνητικό τομέα. Ειδικότερα, παρατηρούμε ότι τα πανεπιστήμια της Βόρειας Ευρώπης έχουν καλύτερη συνολική αξιολόγηση και κλίνουν στον πρακτικό/ερευνητικό τομέα, αντίθετα με αυτά της Νότιας Ευρώπης, που έχουν κακή συνολική αξιολόγηση και κλίνουν στο θεωρητικό τομέα. Τα πανεπιστήμια της Κεντρικής Ευρώπης δεν παρουσιάζουν ομοιογένεια που να μας επιτρέπει να τα ομαδοποιήσουμε. Όλα αυτά εξηγούνται στο παρακάτω διάγραμμα. Κάθε κουκίδα αντιπροσωπεύει ένα πανεπιστήμιο. Όσο πιο αριστερά βρίσκεται κάθε πανεπιστήμιο, τόσο καλύτερη είναι η γενική του εικόνα, όσο πιο πάνω βρίσκεται, τόσο μεγαλύτερη είναι η κλίση του στον ερευνητικό/πρακτικό τομέα και όσο πιο κάτω βρίσκεται, τόσο μεγαλύτερη είναι η κλίση του στον θεωρητικό τομέα.



Στο χωρίο 1 βρίσκονται κυρίως πανεπιστήμια της Β. Ευρώπης, στο 2 της Κεντρικής και στο 3 της Νότιας. Τέλος, κατατάξαμε τα πανεπιστήμια από το καλύτερο στο χειρότερο (βλ. ερώτημα 6).

1)

n =36 το πλήθος των πανεπιστημίων

p =8 το πλήθος των κριτηρίων

ο δειγματικός μέσος:

xbar = (3.6556 3.6611 2.8444 2.7972 3.4222 2.9917 2.4528 2.2556)

Ο πίνακας συνδιακύμανσης:

S =

0.4471	0.2168	0.0875	0.2702	-0.0893	0.2525	-0.0030	0.0788
0.2168	0.3110	0.1043	0.2059	0.0243	0.3048	0.1018	0.1171
0.0875	0.1043	0.1328	0.1887	0.0516	0.1452	0.0904	0.0815
0.2702	0.2059	0.1887	0.4734	0.0295	0.3591	0.1062	0.1633
-0.0893	0.0243	0.0516	0.0295	0.2692	0.0268	0.1602	0.1044
0.2525	0.3048	0.1452	0.3591	0.0268	0.6791	0.1447	0.2188
-0.0030	0.1018	0.0904	0.1062	0.1602	0.1447	0.4214	0.2470

0.0788   0.1171   0.0815   0.1633   0.1044   0.2188   0.2470   0.2694

Οι διασπορές:

VUniv = (0.4471   0.3110   0.1328   0.4734   0.2692   0.6791   0.4214   0.2694)

Ο πίνακας συσχέτισης:

R =

1.0000	0.5814	0.3589	0.5872	-0.2573	0.4582	-0.0069	0.2271
0.5814	1.0000	0.5134	0.5366	0.0840	0.6633	0.2813	0.4045
0.3589	0.5134	1.0000	0.7525	0.2726	0.4836	0.3823	0.4306
0.5872	0.5366	0.7525	1.0000	0.0826	0.6334	0.2377	0.4573
-0.2573	0.0840	0.2726	0.0826	1.0000	0.0626	0.4757	0.3878
0.4582	0.6633	0.4836	0.6334	0.0626	1.0000	0.2706	0.5115
-0.0069	0.2813	0.3823	0.2377	0.4757	0.2706	1.0000	0.7330
0.2271	0.4045	0.4306	0.4573	0.3878	0.5115	0.7330	1.0000

2)

Αφού κάναμε ανάλυση σε κύριες συνιστώσες (PCA), έχουμε ότι οι συντελεστές των κύριων συνιστωσών είναι τα ιδιοδιανύσματα του πίνακα S, τα οποία είναι οι στήλες του πίνακα T, αφού έχουμε ταξινομήσει τις στήλες του, ώστε να αντιστοιχούν στις ιδιοτιμές κατά φθίνουσα σειρά.

T =

0.3558	-0.4561	-0.5035	0.3562	0.1470	-0.4339	-0.2554	0.0991
0.3556	-0.0826	0.0458	0.2783	0.6098	0.5257	0.3354	-0.1624
0.2078	0.0532	-0.1685	-0.3071	0.0401	0.3202	-0.0937	0.8476
0.4671	-0.1270	-0.3186	-0.6223	-0.2621	0.1392	0.0852	-0.4262
0.0681	0.4783	-0.0270	-0.3748	0.6368	-0.4460	-0.1294	-0.0638
0.5838	-0.0817	0.7408	0.0161	-0.1217	-0.1702	-0.2370	0.0589
0.2433	0.6386	-0.2377	0.3681	-0.2232	0.2600	-0.4482	-0.1508
0.2844	0.3506	-0.0919	0.1948	-0.2568	-0.3434	0.7297	0.1832

Το ποσοστό αρχικής μεταβλητότητας που εξηγεί η κάθε συνιστώσα παρουσιάζεται στον παρακάτω πίνακα (η πρώτη κύρια συνιστώσα εξηγεί το 50.86% τη αρχικής μεταβλητότητας κ.ο.κ.)

prop =

0.5086

0.2111

0.0880

0.0714

0.0531

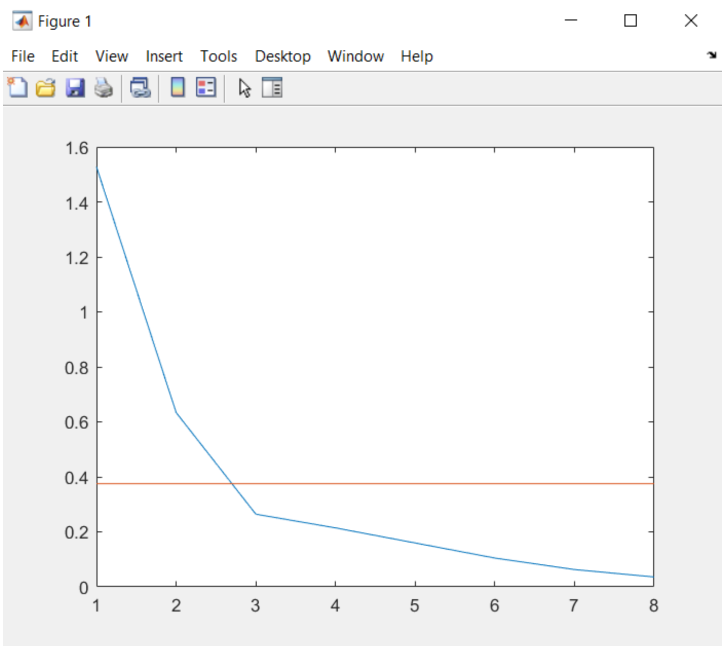
0.0349

0.0210

0.0120

3)

Σύμφωνα με το κριτήριο της μέσης ιδιοτιμής συμπεραίνουμε ότι θα κρατήσουμε τις δύο πρώτες κύριες συνιστώσες πράγμα το οποίο έχουμε ελέγξει και αλγεβρικά και με το plot.



cum_prop =	0.5086
	0.7197
	0.8077
	0.8791
	0.9322
	0.9671
	0.9880
	1.0000

Βλέποντας το παραπάνω διάνυσμα, διαπιστώνουμε ότι οι δύο πρώτες κύριες συνιστώσες εξηγούν το 71.97% της αρχικής μεταβλητότητας.

score =

-2.1313	-0.1814
-1.1060	0.1320
-1.0907	-0.5052
-1.0466	-0.0202
-2.5623	0.4814
-1.0846	0.3075
-2.0245	-0.0957
-0.9908	0.1003
-1.3967	0.2857
-0.6395	-0.6347
-0.8704	-1.4096
-1.0182	-0.2716
-0.8685	-0.2646
-0.4423	-0.3826
-0.7801	1.1471
-0.2533	-0.4285
0.0372	0.9605
0.6753	-1.7262
0.9268	1.0814
0.3766	-0.5446
0.1344	-0.3517
0.8717	0.1671
0.2563	0.5194
0.2881	-0.2868
0.2053	-0.6940
0.6662	1.4384
2.3317	-1.3771
1.2919	-1.4630
1.6265	-0.5870
1.5282	1.0225
0.4727	0.5433
1.3191	1.2559
0.6206	1.0798
0.3952	0.4840

2.4488	0.3626
1.8332	-0.1444

4)

u1	u2
0.3558	-0.4561
0.3556	-0.0826
0.2078	0.0532
0.4671	-0.1270
0.0681	0.4783
0.5838	-0.0817
0.2433	0.6386
0.2844	0.3506

Όπως παρατηρούμε, οι συντελεστές της πρώτης κύριας συνιστώσας είναι όλοι θετικοί, άρα αυτή παρουσιάζει μία συνολική εικόνα/αξιολόγηση των πανεπιστημίων.

Η δεύτερη κύρια συνιστώσα έχει θετικούς συντελεστές στις γραμμές 3,5,7,8 που αντιστοιχούν στα κριτήρια αξιολόγησης: Εξοπλισμός Υπολογιστών, Προσωπική Συμβουλευτική, Πρακτική Άσκηση κι Έρευνα (πρακτικά κριτήρια), ενώ έχει αρνητικούς συντελεστές στις γραμμές 1,2,4,6 που αντιστοιχούν στα: Διαθέσιμος Ειδικός Τύπος, Διαθέσιμη Βιβλιογραφία, Κλίμα Μαθητείας και Προετοιμασία Εξετάσεων (θεωρητικά/exam-oriented κριτήρια). Άρα η δεύτερη κύρια συνιστώσα μπορούμε να πούμε ότι είναι ένας δείκτης κλίσης των πανεπιστημίων στον θεωρητικό τομέα.

Η ομαδοποίηση των κριτηρίων αξιολόγησης που επιτυγχάνεται με τη δεύτερη κύρια συνιστώσα είναι ικανοποιητικά καλή, όπως φαίνεται και στο διάγραμμα παρακάτω (4 κριτήρια αξιολόγησης κάτω από το 0 και 4 πάνω από το 0).

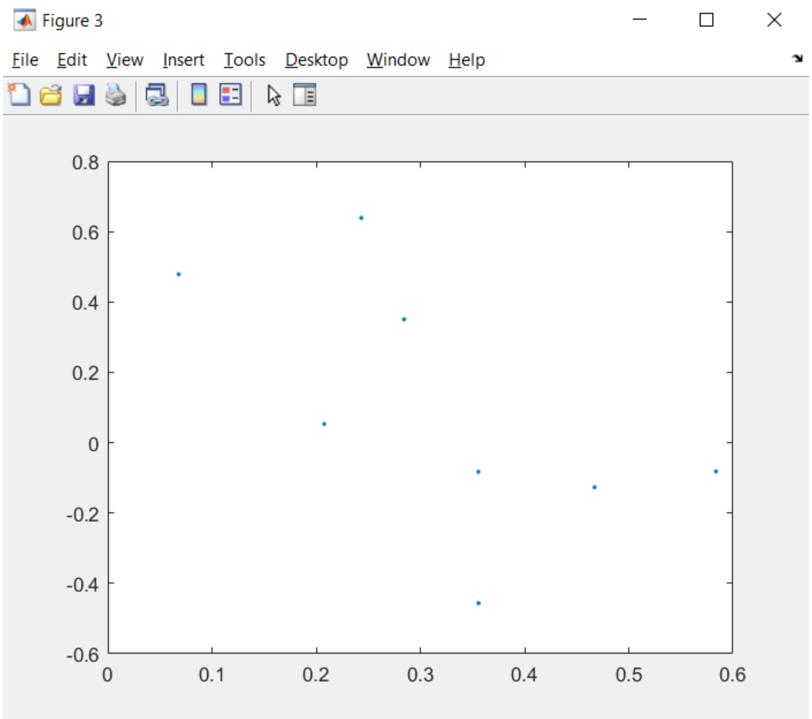
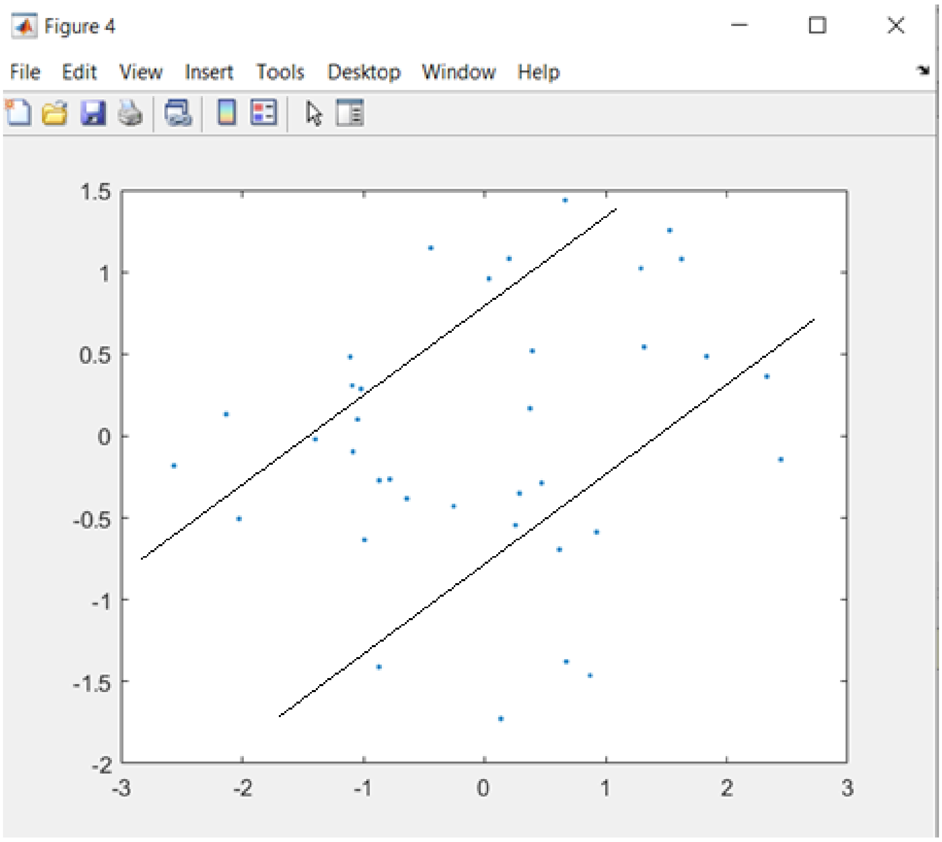


Figure 3
File Edit View Insert Tools Desktop Window Help
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
0
0.1
0.2
0.3
0.4
0.5
0.6

5)



Παρατηρώντας το διάγραμμα των scores παραπάνω και αντιστοιχίζοντας την κάθε κουκίδα με το πανεπιστήμιο της, διαπιστώνουμε ότι μπορούμε να κάνουμε μια ομαδοποίηση στα πανεπιστήμια ανάλογα με το κομμάτι της Ευρώπης στο οποίο βρίσκονται, Βόρεια, Κεντρική και Νότια Ευρώπη. (βλέπε μαύρες γραμμές στο διάγραμμα). Τα πανεπιστήμια της Βόρειας Ευρώπης έλαβαν σχετικά υψηλή βαθμολογία από τους φοιτητές και είναι καλύτερα στον τομέα της έρευνας και πρακτικής άσκησης, ενώ αντίθετα τα πανεπιστήμια της Νότιας Ευρώπης έλαβαν σχετικά χαμηλή βαθμολογία από τους φοιτητές και διαπρέπουν στο θεωρητικό τομέα (βιβλιογραφία κλπ), είναι πιο exam-oriented. Τα πανεπιστήμια της Κεντρικής Ευρώπης διαφέρουν μεταξύ τους σε συνολική εικόνα και κλίση σε κάποιον συγκεκριμένο τομέα.

6)

J =

- 5 Πανεπιστήμιο Γουώργουικ
- 1 Πανεπιστήμιο Λωζάνης
- 7 Πανεπιστήμιο Κέιμπριτζ
- 9 Πανεπιστήμιο Λισαβώνας
- 2 Πανεπιστήμιο Μπαιρόιτ
- 3 Πανεπιστήμιο Πασάου
- 6 ΤΕΙ Οικονομικών Ελσίνκι

- 4 Πανεπιστήμιο Γκρόνινγκεν
- 12 Οικονομικό Πανεπ. Μιλάνου
- 8 Πανεπιστήμιο Οξφόρδης
- 11 Πανεπιστήμιο Παρισίων IX
- 13 Οικονομ. Πανεπ. Στοκχόλμης
- 15 School of Economics Λονδίνου
- 10 Καθολικό Πανεπιστήμιο Λεβέν
- 14 Πανεπιστήμιο Κοπεγχάγης
- 16 Πανεπιστήμιο Στρασβούργου
- 17 University College Λονδίνου
- 21 Πανεπιστήμιο Τουλούζης
- 25 Πανεπιστήμιο Αθηνών
- 23 Πανεπ. Carlos III Μαδρίτης
- 24 Πανεπιστήμιο Λυών
- 20 Πανεπιστήμιο Μαδρίτης
- 34 Πανεπιστήμιο Βιέννης
- 31 Πανεπιστήμιο Παρισίων I
- 33 Πανεπιστήμιο Μονάχου
- 26 University College Δουβλίνου
- 18 Πανεπιστήμιο Σιένας
- 22 Πανεπιστήμιο Μανχάιμ
- 19 Πανεπιστήμιο Άμστρενταμ
- 28 Πανεπιστήμιο Μπολώνιας
- 32 Πανεπιστήμιο Κολονίας
- 30 Πανεπιστήμιο Μίνστερ
- 29 Πανεπιστήμιο Βενετίας
- 36 Πανεπιστήμιο Βαρκελώνης
- 27 Πανεπιστήμιο Ρώμης
- 35 Πανεπιστήμιο Βόνης

Εφαρμόζοντας την PCA στη MATLAB και χρησιμοποιώντας τα scores της πρώτης κύριας συνιστώσας, η οποία δίνει μια συνολική εικόνα/αξιολόγηση των πανεπιστημίων, μπορούμε να sortάρουμε τα πανεπιστήμια από το καλύτερο σε γενική βαθμολογία προς το χειρότερο (βλ. παραπάνω).

7)

Χρησιμοποιώντας τον ίδιο κώδικα, αλλά για τον πίνακα R αντί για S και τα τυποποιημένα δεδομένα αντί για τα κεντρικοποιημένα, διαπιστώνουμε ότι η ανάλυσή μας δεν έχει ιδιαίτερες διαφορές. Ακόμα κι η κατάταξη των πανεπιστημίων δε διαφέρει σημαντικά από την αντίστοιχη κατάταξη με τα κεντρικοποιημένα δεδομένα.

```
G[J J1]
```

G =

5	5
1	1
7	7
9	9
2	12
3	13
6	6
4	2
12	4
8	11
11	3
13	8
15	10
10	14
14	15
16	16
17	21
21	18
25	24
23	17
24	25
20	31
34	23
31	34

33	20
26	28
18	22
22	26
19	19
28	33
32	29
30	32
29	27
36	30
27	36
35	35

Αυτό είναι λογικό, γιατί χρησιμοποιώντας τυποποιημένα δεδομένα «εξαλείφουμε» τις μονάδες μέτρησης και στη συγκεκριμένη έρευνα οι μονάδες είναι ίδιες. Ένα χαρακτηριστικό παράδειγμα που θα ήταν πολύ χρήσιμη η ανάλυση με τα τυποποιημένα δεδομένα είναι η μελέτη χαρακτηριστικών αυτοκινήτων (ταχύτητα σε Km/h, κατανάλωση βενζίνης σε λίτρα).

```
Di=diag([3 3 3 3]);
P=[1 0.2 0.4 0.8 ; 0.2 1 0.5 0.6 ; 0.4 0.5 1 0.7 ;
0.8 0.6 0.7 1];
Si=(Di^(1/2))*P*(Di^(1/2));%βρισκουμε τον πίνακα
συνδιακύμανσης μέσω του πίνακα συσχέτισης
Dii=diag([1 2 3 4])
Sii=(Dii^(1/2))*P*(Dii^(1/2));%βρισκουμε τον πίνακα
συνδιακύμανσης μέσω του πίνακα συσχέτισης

%ερωτημα2
A=[1 -3 0 -5 ; 0 0 1 0];
B=[-1 0 0 0 ; -2 0 5 0];
SYi= A*Si*A';
SYii= A*Sii*A';
SZi=B*Si*B';
SZii=B*Sii*B';
DYi=diag(diag(SYi));
DYii=diag(diag(SYii));
DZi=diag(diag(SZi));
DZii=diag(diag(SZii))
PYi=(DYi^(-1/2))*SYi*(DYi^(-1/2))
PYii=(DYii^(-1/2))*SYii*(DYii^(-1/2))
PZi=(DZi^(-1/2))*SZi*(DZi^(-1/2))
PZii=(DZii^(-1/2))*SZii*(DZii^(-1/2))
covY1Z2=(5*Si(1,3)-2*Si(1,1)-15*Si(2,3)+6*Si(2,1)-
25*Si(3,4)+10*Si(1,4))/(sqrt(SYi(1,1))*sqrt(SZi(2,2))
)
covY1Z2=(5*Sii(1,3)-2*Sii(1,1)-
15*Sii(2,3)+6*Sii(2,1)-
25*Sii(3,4)+10*Sii(1,4))/(sqrt(SYii(1,1))*sqrt(SZii(2
,2)))
%συσχέτιση Y1,Z2

%ερώτημα 3
mu=[2;3;5;8];
K=mvnrnd(mu,Sii,1000); %δείγμα 1000 παρατηρήσεων
Y1=K(:,1)-3*K(:,2)-5*K(:,4);
Y2=K(:,3);
Z1=(-1)*K(:,1);
Z2=5*K(:,3)-2*K(:,1);
Y=[Y1 Y2];
Z=[Z1 Z2];
W=[Y1 Y2 Z1 Z2];
sum(W);
std(W);
wbar= sum(W)/1000
Wbar=repmat(wbar,[1000,1]);
```

```
Wstar=W-Wbar;% Centered observations
stds=repmat(std(W),[1000,1]);
Wtilde=(W-Wbar)./stds; % Standardised observations
DS=(Wstar'*Wstar)/(1000-1)% δειγματικός πίνακας
συνδιακύμανσης
DR=(Wtilde'*Wtilde)/(1000-1)% δειμγάτικος πίνακας
συσχέστισης

%ερώτημα 4
X4
[n p]=size(X);
Y1=X(:,1)-3*X(:,2)-5*X(:,4);
Y2=X(:,3);
Z1=(-1)*X(:,1);
Z2=5*X(:,3)-2*X(:,1);
Y=[Y1 Y2];
Z=[Z1 Z2];
W=[Y1 Y2 Z1 Z2];
sum(W);
std(W);
wbar= sum(W)/n
Wbar=repmat(wbar,[n,1]);
Wstar=W-Wbar;% Centered observations
stds=repmat(std(W),[n,1]);
Wtilde=(W-Wbar)./stds; % Standardised observations
DS=(Wstar'*Wstar)/(n-1)% δειγματικός πίνακας
συνδιακύμανσης
DR=(Wtilde'*Wtilde)/(n-1)% δειμγάτικος πίνακας
συσχέστισης

%ερωτημα5
function lp=lognormpdf1(X,m,S)
[n,p]=size(X);
lp=0;
for j=1:n
lp=lp -0.5*p*log(2*pi)+0.5*log(det(inv(S)))-
0.5*((X(j,:)-m)*inv(S)*((X(j,:)-m)'));
end
end
[n,p]=size(K);
semp3=zeros(4,4);
memp3=sum(K)/n; %υπολογισμός της εμπ της μεσης τιμής
for i=1:n %με το for υπολογίζουμε την εμπ του πίνακα
συνδιακύμανσης
semp3=semp3+(((K(i,:)- memp3)')*((K(i,:)-memp3)));
end
semp3=semp3/n;
```

```
lp=lognormpdf1(K,memp3,semp3)
[n,p]=size(X);
semp4=zeros(4,4);
memp4=sum(X)/n; %υπολογισμός της εμπ της μεσης τιμής
for i=1:n %με το for υπολογίζουμε την εμπ του πίνακα
συνδιακύμανσης
semp4=semp4+(((X(i,:)- memp4)')*((X(i,:)-memp4)));
end
semp4=semp4/n;
lp=lognormpdf1(X,memp4,semp4)

%ASKHSH 2
%erwtima 1
Universities
[n,p]=size(Univ)
xbar=mean(Univ) % means
S=cov(Univ) % covariance matrix
VUniv=var(Univ) % variances
D=diag(VUniv); % diagonal matrix with variances on
the diagonal
R=D^(-(1/2))*S*(D^(-1/2)) % correlation matrix

%%erwtima 2
[T,L]=eig(S) % T: the columns of T are the
eigenvectors,
              % L: diagonal matrix, the diagonal
elements are the eigenvalues
T*L*T'
eigenvalues=diag(L);
[eigenvalues,I]=sort(eigenvalues,'descend');
T=T(:,I)%ο παλιός T με την κατάλληλη σειρά
L=L(I,I)
T*L*T'
lamda1=eigenvalues(1);
u1=T(:,1);
prop=eigenvalues/sum(eigenvalues)

%ερωτημα3
lamda_bar=mean(eigenvalues)
find(eigenvalues>lamda_bar)
cum_prop=cumsum(prop)%αθροιστικά

% select two principal components
u1
u2=T(:,2)
Xbar=repmat(xbar,[n,1]);
Xstar=Univ-Xbar; % Centered observations
```

```
scores=Xstar*T
score=[scores(:,1),scores(:,2)]
% Scree plot
figure(1)
plot(1:p,eigenvalues)
hold on
plot(1:p,lamda_bar*ones(1,p))
%erwthma 4
figure(3)
plot(T(:,1),T(:,2),'.')
%erwthma 5
figure(4)
plot(scores(:,1),scores(:,2),'.')
%ερωτημα 6
[scores(:,1),J]=sort(scores(:,1),'ascend');
J
%ΕΡΩΤΗΜΑ 7
[T1,L1]=eig(R)
eigenvalues1=diag(L1);
[eigenvalues1,I1]=sort(eigenvalues1,'descend');
T1=T1(:,I1)%ο παλιός T με την κατάλληλη σειρά
T1=-T1
L1=L1(I1,I1)
lamda11=eigenvalues1(1);
u11=T1(:,1);
prop1=eigenvalues1/sum(eigenvalues1)

%ερωτημα3
lamda_bar1=mean(eigenvalues1)
find(eigenvalues1>lamda_bar1)
cum_prop1=cumsum(prop1)%αθροιστικά

% select two principal components
u11
u21=T1(:,2)
Xbar=repmat(xbar,[n,1]);
Xstar=Univ-Xbar; % Centered observations
stds=repmat(std(Univ),[n,1]);
Xtilde=(Univ-Xbar)./stds; % Standardised observations
scoresTilde=Xtilde*T1
scoreTilde=[scoresTilde(:,1),scoresTilde(:,2)]
% Scree plot
figure(5)
plot(1:p,eigenvalues1)
hold on
plot(1:p,lamda_bar1*ones(1,p))
%erwthma 4
```

```
figure(7)
plot(T1(:,1),T1(:,2),'.')
%erwthma 5
figure(8)
plot(scoresTilde(:,1),scoresTilde(:,2),'.')
%ερωτημα 6
[scoresTilde(:,1),J1]=sort(scoresTilde(:,1),'ascend')
;
J1
```