

Κεφάλαιο 20

Ανάλυση σε Κύριες Συνιστώσες και Διαχωριστική Ανάλυση

20.1 Ανάλυση σε Κύριες Συνιστώσες

Η μέθοδος των Κυρίων Συνιστωσών είναι μία τεχνική ανάλυσης δεδομένων με σκοπό τη δημιουργία καινούργιων μεταβλητών, οι οποίες είναι γραμμικοί συνδυασμοί των αρχικών μεταβλητών, έτσι ώστε να είναι ασυσχέτιστες μεταξύ τους και να περιέχουν όσο το δυνατόν μεγαλύτερο μέρος της διακύμανσης των αρχικών μεταβλητών. Οι νέες μεταβλητές που παράγονται ονομάζονται *Κύριες Συνιστώσες*. Το τι επιτυγχάνεται από τη μέθοδο αυτή είναι ότι από ένα σύνολο συσχετισμένων μεταβλητών καταλήγουμε σε ένα σύνολο ασυσχέτιστων μεταβλητών, το οποίο είναι χρήσιμο για αρκετές στατιστικές μεθόδους. Επίσης, οι κύριες συνιστώσες που προκύπτουν μπορούν να ερμηνεύσουν το μεγαλύτερο ποσοστό της διακύμανσης, που σημαίνει πως καταλήγουμε σε ένα πιο μικρό αριθμό μεταβλητών από ότι είχαμε αρχικά, με κόστος ότι χάνουμε ένα μικρό ποσοστό της συνολικής μεταβλητότητας. Αυτό είναι πολύ σημαντικό ιδιαίτερα στις περιπτώσεις που έχουμε λίγες παρατηρήσεις και πολλές μεταβλητές. Συνεπώς, αν σε μια τέτοια περίπτωση θέλαμε να εφαρμόσουμε ένα (γενικευμένο) γραμμικό μοντέλο, η υπερπαραμετροποίηση του μοντέλου μπορεί να ξεπεραστεί χρησιμοποιώντας την παραπάνω μέθοδο.

Για να δούμε πως εφαρμόζεται η μέθοδος αυτή στην R, θα χρησιμοποιήσουμε

το πλαίσιο δεδομένων `possum` το οποίο βρίσκεται στη βιβλιοθήκη `DAAG` και αναφέρεται σε εννέα μορφομετρικές μετρήσεις για 104 οπόσσοι (δενδρόβιο μαρσιποφόρο ζώο). Στην R υπάρχουν δύο εντολές οι οποίες μπορούν να εφαρμόσουν την ανάλυση σε κύριες συνιστώσες. Προτιμητέα είναι η εντολή `prcomp()` με την οποία ο υπολογισμός γίνεται με την διάσπαση ιδιάζουσας τιμής του πίνακα δεδομένων (πιθανώς κανονικοποιημένος) και όχι υπολογίζοντας τις ιδιοτιμές με φασματική διάσπαση του πίνακα συνδιακύμανσης (ή συσχετίσεων) όπως γίνεται εφαρμόζοντας την εντολή `princomp()`. Το πρώτο βήμα είναι να ελέγξουμε τη μεταβλητότητα των μορφομετρικών μεταβλητών με τη βοήθεια του πίνακα συνδιακύμανσης, ο οποίος υπολογίζεται με την εντολή `cov()`.

```
> library(DAAG)
> possum.dat<-na.omit(possum[,6:14])
> cov(possum.dat)
```

	hdlngth	skullw	totlngth	taill	footlgth	earconch	eye	chest	belly
hdlngth	12.719151	7.990574910	10.579248	2.0111222	6.13856368	1.475288407	1.351369	4.661275	5.520779
skullw	7.990575	9.788319056	7.138732	1.5772178	3.78241767	0.002080716	1.062334856	4.050000000	3.925904245
totlngth	10.579248	7.138732153	18.597533	4.8039977	8.43164953	2.446375405	1.167453	5.145588	6.160689
taill	2.011122	1.577217780	4.803998	3.8747858	-1.09253760	-3.171178374	0.4161098	0.7083333	1.5989197
footlgth	6.138564	3.782417666	8.431650	-1.0925376	19.31871312	13.957338664	0.02410813	4.07107843	3.68086808
earconch	1.475288	0.002080716	2.446375	-3.1711784	13.95733866	16.445541595	0.600737674	1.691176471	0.600737674
eye	1.351369	1.062334856	1.167453	0.4161098	0.02410813	-0.653397106	1.10688559	0.3240196	0.7081715
chest	4.661275	4.050000000	5.145588	0.7083333	4.07107843	1.691176471	0.32401961	4.2254902	3.4583333
belly	5.520779	3.925904245	6.160689	1.5989197	3.68086808	0.600737674	0.70817152	3.4583333	7.6600514

```
> diag(cov(possum.dat))
```

hdlngth	skullw	totlngth	taill	footlgth	earconch	eye	chest
12.719151	9.788319	18.597533	3.874786	19.318713	16.445542	1.106886	4.225490

belly
7.66005

Η διαγώνιος του πίνακα συνδιακύμανσης παρουσιάζει τη διασπορά και όπως είναι φανερό, οι μεταβλητές έχουν διαφορετική μεταβλητότητα. Ο λόγος είναι ότι το κάθε μέρος του μαρσιποφόρου ζώου το οποίο μετρήθηκε έχει διαφορετικό μέγεθος. Συνεπώς, για να δοθεί η ίδια βαρύτητα στις μεταβλητές, αυτές πρέπει να κανονικοποιηθούν. Αυτό γίνεται χρησιμοποιώντας τον πίνακα συσχετίσεων στην ανάλυση κυρίων συνιστωσών, αφού στη διαγώνιό του υπάρχουν μόνο μονάδες. Στην R αυτό γίνεται προσθέτοντας το όρισμα `scale=T` στην `prcomp()`.

```
> cor(possum.dat)
      hdlngth      skullw  totlngth      taill      footlgth      earconch
hdlngth 1.0000000 0.7161350133 0.6878569 0.2864742 0.391605342 0.1020055445
skullw   0.7161350 1.0000000000 0.5291019 0.2561025 0.275059297 0.0001639965
totlngth 0.6878569 0.5291019496 1.0000000 0.5659151 0.444831755 0.1398850635
taill    0.2864742 0.2561024623 0.5659151 1.0000000 -0.126276764 -0.3972580707
footlgth 0.3916053 0.2750592967 0.4448318 -0.1262768 1.000000000 0.7830498840
earconch 0.1020055 0.0001639965 0.1398851 -0.3972581 0.783049884 1.0000000000
eye      0.3601583 0.3227423866 0.2573122 0.2009244 0.005213424 -0.1531446219
chest    0.6358242 0.6297416468 0.5804553 0.1750553 0.450590285 0.2028739514
belly    0.5593137 0.4533877714 0.5161617 0.2934857 0.302583550 0.0535235532
      eye      chest      belly
hdlngth 0.360158311 0.6358242 0.55931366
skullw   0.322742387 0.6297416 0.45338777
totlngth 0.257312171 0.5804553 0.51616171
taill    0.200924441 0.1750553 0.29348570
footlgth 0.005213424 0.4505903 0.30258355
earconch -0.153144622 0.2028740 0.05352355
eye      1.000000000 0.1498240 0.24320430
chest    0.149823996 1.0000000 0.60787245
belly    0.243204300 0.6078724 1.00000000
```

```
> possum.prc<-prcomp(possum.dat,scale=T)
```

```
> summary(possum.prc)
```

Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Standard deviation	1.986	1.407	0.944	0.859	0.7508	0.5563	0.5165	0.4048	0.380

Proportion of Variance	0.438	0.220	0.099	0.082	0.0626	0.0344	0.0296	0.0182	0.016
Cumulative Proportion	0.438	0.658	0.757	0.839	0.9018	0.9362	0.9658	0.9840	1.000

Η εντολή `summary()`, με όρισμα το αντικείμενο της ανάλυσης σε κύριες συνιστώσες, δίνει τις ιδιοτιμές για κάθε κύρια συνιστώσα οι οποίες αντιστοιχούν στις τυπικές αποκλίσεις τους. Αυτές δίνονται σε αύξουσα σειρά. Συνεπώς, στην πρώτη κύρια συνιστώσα (PC1) αντιστοιχεί η μεγαλύτερη ιδιοτιμή, στην δεύτερη (PC2) η δεύτερη μεγαλύτερη κ.ο.κ. Επίσης, στη δεύτερη γραμμή παρουσιάζεται το ποσοστό της μεταβλητότητας που εξηγείται από κάθε κύρια συνιστώσα, ενώ στην τρίτη γραμμή παρουσιάζεται το αθροιστικό ποσοστό μεταβλητότητας.

Στη συνέχεια, επιλέγουμε τις k πρώτες κύριες συνιστώσες που εξηγούν ένα σημαντικό ποσοστό της ολικής μεταβλητότητας. Στη βιβλιογραφία υπάρχουν πολλά κριτήρια τα οποία χρησιμοποιούνται για τον σκοπό αυτό. Πιο κάτω αναφέρονται τα τρία πιο δημοφιλή:

1. Ποσοστό συνολικής διακύμανσης που εξηγούν οι κύριες συνιστώσες
2. Κριτήριο του Kaiser
3. Τεχνική του αγκώνα - scree plot.

Σύμφωνα με το πρώτο κριτήριο, διαλέγουμε τόσες συνιστώσες ώστε αθροιστικά να εξηγούν μεγαλύτερο ποσοστό μεταβλητότητας από τον στόχο που θέσαμε στην αρχή (συνήθως μεγαλύτερο του 80%). Άρα, στο παραπάνω παράδειγμα επιλέγονται οι πρώτες τέσσερις συνιστώσες που εξηγούν αθροιστικά το 83.9% της συνολικής μεταβλητότητας των δεδομένων. Το κριτήριο του Kaiser λέει, αν χρησιμοποιείται ο πίνακας συνδιακύμανσης, να θεωρήσουμε τόσες συνιστώσες όσες η ιδιοτιμή τους είναι μεγαλύτερη από τη μέση τους τιμή, ενώ αν χρησιμοποιείται ο πίνακας συσχετίσεων, τόσες όσες έχουν ιδιοτιμή μεγαλύτερη του 1 (γιατί:). Συνεπώς, στο παράδειγμα επιλέγονται οι δύο πρώτες κύριες συνιστώσες. Τέλος, η τεχνική του αγκώνα είναι μια γραφική μέθοδος για την επιλογή του αριθμού των κυρίων συνιστωσών. Το γράφημα αυτό έχει στον οριζόντιο άξονα των x τη σειρά και στον κάθετο άξονα των y την τιμή κάθε ιδιοτιμής. Προτείνεται να επιλεγούν τόσες συνιστώσες μέχρι το γράφημα αρχίζει να δημιουργεί «αγκώνα», δηλαδή να αλλάζει κλίση. Το scree plot του παραδείγματος παρουσιάζεται στο Σχήμα 20.1. Το πρώτο παρουσιάζεται ως ραβδόγραμμα, ενώ το δεύτερο ως πολύγωνο, αφού προσθέσαμε το όρισμα `type="line"`. Είναι φανερό ότι η κλίση αλλάζει από την τρίτη ιδιοτιμή και μετά, και άρα επιλέγονται οι δύο πρώτες συνιστώσες. Συμπερασματικά, για

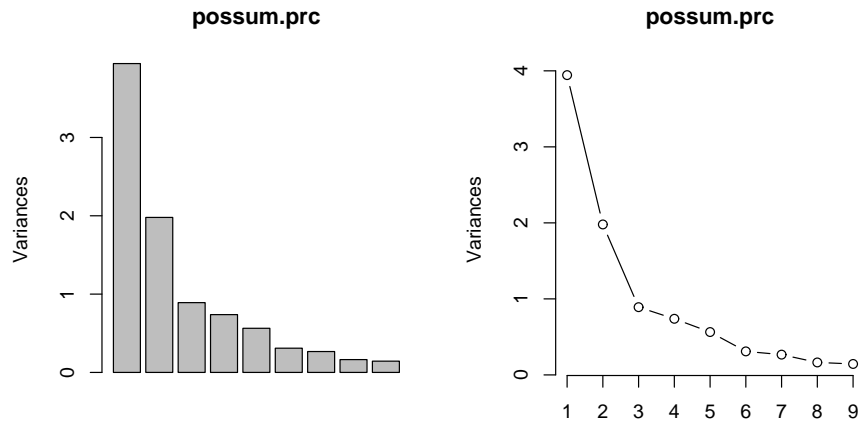
το παραπάνω παράδειγμα είναι κατάλληλη η επιλογή των δύο πρώτων κυρίων συνιστωσών PC1 και PC2.

Το πιο δύσκολο κομμάτι της ανάλυσης σε κύριες συνιστώσες είναι ίσως η ερμηνεία τους. Αυτή βασίζεται στους συντελεστές των κυρίων συνιστωσών (principal component loadings) που συνιστούν το γραμμικό συνδυασμό των αρχικών μεταβλητών. Οι συντελεστές αυτοί αντιστοιχούν στα στοιχεία των ιδιοδιανυσμάτων των ιδιοτιμών των συνιστωσών. Αυτοί δίνονται στην R μαζί με τις αντίστοιχες ιδιοτιμές καλώντας το αντικείμενο της ανάλυσης σε κύριες συνιστώσες που δημιουργήθηκε χρησιμοποιώντας την εντολή `prcomp()`. Θα δοθεί η ερμηνεία για τις πρώτες δύο συνιστώσες αφού αυτές επιλέγηκαν με βάση τα πιο πάνω κριτήρια. Η πρώτη κύρια συνιστώσα μπορεί να ερμηνευθεί ως ένας σταθμισμένος μέσος όρος των μεταβλητών, αφού όλοι οι συντελεστές τους έχουν το ίδιο πρόσημο. Η δεύτερη κύρια συνιστώσα κάνει σύγκριση του μήκους της ουράς (`tail1`) και του μεγέθους του ματιού (`eye`) με το μήκος του ποδιού (`footlgth`) και της κόγχης του αυτιού (`earconch`) των οπόσσομι.

```

> par(mfrow=c(1,2))
> screeplot(possum.prc)
> screeplot(possum.prc,type="line")

```



Σχήμα 20.1: Scree plot.

```

> possum.prc
Standard deviations:
[1] 1.9858729 1.4068546 0.9439907 0.8591514 0.7508320 0.5562584 0.5164833
[8] 0.4047894 0.3795100

```

Rotation:

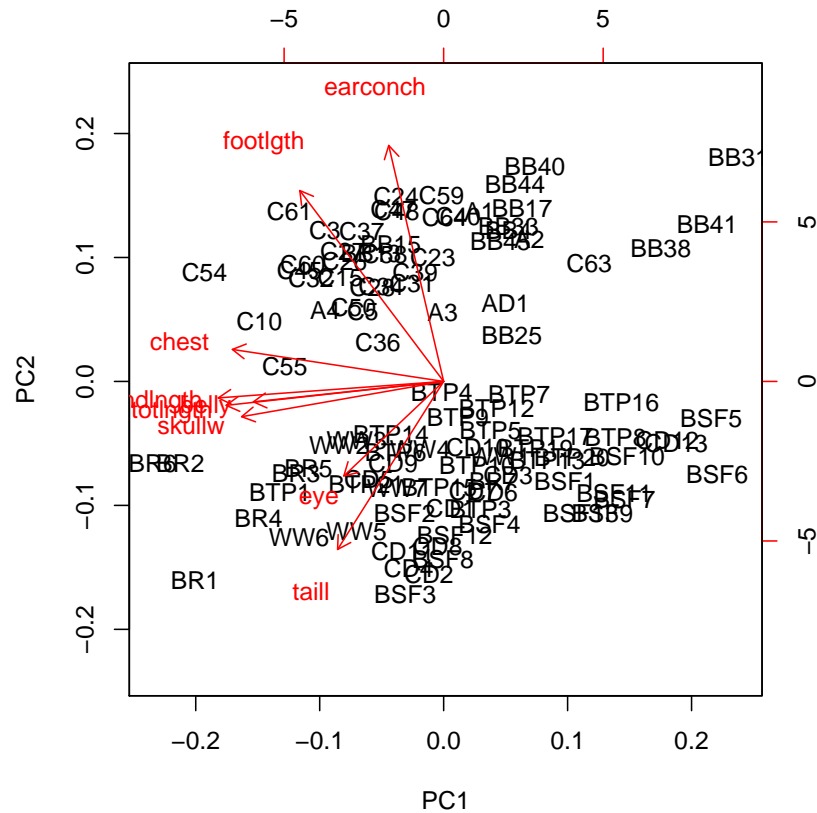
	PC1	PC2	PC3	PC4	PC5
hdlngth	-0.4380078	-0.04519844	0.138623007	-0.07495574	0.238115223
skullw	-0.3927120	-0.09678900	0.223585881	-0.32364347	0.501176778
totlngth	-0.4227082	-0.06564460	-0.312750646	0.33593765	0.120510988
taill	-0.2061992	-0.46100652	-0.514351991	0.38374763	0.026090130
footlngth	-0.2796247	0.52384020	-0.062416712	0.27857386	0.002538413
earconch	-0.1067739	0.64749223	-0.004813129	0.26143915	-0.017773149
eye	-0.1940729	-0.25932151	0.748023649	0.49876897	-0.192020543
chest	-0.4106589	0.08780572	-0.037517222	-0.41050426	-0.058210391
belly	-0.3699061	-0.05604043	-0.059633425	-0.25745316	-0.797707163

	PC6	PC7	PC8	PC9
hdlngth	-0.567980976	0.447545311	-0.439147036	0.09393289
skullw	-0.028120923	-0.617005313	0.216587933	0.04581025
totlngth	-0.023209818	0.329000800	0.693643612	-0.04355347
taill	0.149454731	-0.307513951	-0.429684531	0.17585120
footlngth	0.026742546	-0.203737843	-0.229883084	-0.68616711
earconch	0.004502918	-0.142151245	-0.018410561	0.69289065
eye	0.221844208	0.006576962	-0.007914748	0.02419833
chest	0.720163882	0.313950090	-0.170795256	0.06299215
belly	-0.291796339	-0.236383757	0.109774529	0.02794910

Όπως έχει αναφερθεί στην αρχή του κεφαλαίου, αφού έγινε η επιλογή των κυρίων συνιστωσών με τις οποίες χάνεται κάποια πληροφορία, αυτές μπορούν να χρησιμοποιηθούν στη συνέχεια σε ένα (γενικευμένο) γραμμικό μοντέλο. Σε αυτό το μοντέλο θα χρησιμοποιηθούν οι τιμές των κυρίων συνιστωσών, οι οποίες υπολογίζονται για κάθε γραμμή των δεδομένων από τον γραμμικό συνδυασμό των τιμών των αρχικών μεταβλητών, με συντελεστές τους συντελεστές από τις κύριες συνιστώσες (principal components loadings). Αυτές οι τιμές ονομάζονται principal components scores και στην R υπολογίζονται από την εντολή `predict()` με όρισμα το αντικείμενο της ανάλυσης σε κύριες συνιστώσες. Έτσι, οι τιμές των πρώτων δύο κυρίων συνιστωσών στο πιο πάνω παράδειγμα δίνονται από:

```
> predict(possum.prc)[,1:2]
      PC1      PC2
C3    -1.93077  1.7406
C5    -1.31261  0.8120
C10   -2.99606  0.6969
C15   -1.68351  1.2040
C23   -0.17734  1.4358
.      .      .
.      .      .
.      .      .
BTP16  2.88827 -0.2431
BTP17  1.81839 -0.6301
BTP19  1.49279 -0.7675
BTP20  2.08842 -0.8750
BTP21 -1.25482 -1.1833
```

Τέλος, ένα χρήσιμο γράφημα που χρησιμοποιείται στην ανάλυση σε κύριες συνιστώσες είναι το biplot. Είναι η γραφική παράσταση των τιμών των κυρίων συνιστωσών (principal components scores) και των συντελεστών (principal components loadings) ταυτόχρονα στο ίδιο γράφημα. Οι μεταβλητές παριστάνονται με βέλη που ξεκινούν από την αρχή των αξόνων. Αν τα βέλη είναι κάθετα μεταξύ τους τότε οι αντίστοιχες μεταβλητές δε συσχετίζονται μεταξύ τους, ενώ αντίθετα, αν τα βέλη δείχνουν προς την ίδια (ή αντίθετη) κατεύθυνση τότε οι μεταβλητές συσχετίζονται ισχυρά θετικά (ή αρνητικά). Η αντίστοιχη εντολή στην R είναι η `biplot()` και στο Σχήμα 20.2 παρουσιάζεται το γράφημα για το πιο πάνω παράδειγμα.



Σχήμα 20.2: Biplot.

20.2 Διαχωριστική Ανάλυση

Η βασική ιδέα της διαχωριστικής ανάλυσης (discriminant analysis) είναι να κατατάξει παρατηρήσεις (συνήθως πολυδιάστατες) σε γνωστούς πληθυσμούς με γνωστές κατανομές για κάθε πληθυσμό. Η διαχωριστική ανάλυση αποτελεί μια μέθοδο με πλήθος εφαρμογών σε πολλές επιστήμες.

Έστω ότι υπάρχουν k υπο-πληθυσμοί (ομάδες), $\Pi_1, \Pi_2, \dots, \Pi_k$ με $k \geq 2$. Για τον κάθε υπο-πληθυσμό Π_k υπάρχει και μία κατανομή, f_k . Η διαχωριστική ανάλυση έχει 2 στόχους :

1. Τη διαχώριση ενός πληθυσμού σε ευδιάκριτα σύνολα (υπο-πληθυσμούς) και
2. Την ταξινόμηση παρατηρήσεων στους προηγούμενους γνωστούς πληθυσμούς με γνωστές κατανομές για κάθε πληθυσμό, με τη βοήθεια ενός κανόνα.

Αυτό που θα εξεταστεί εδώ είναι το πως οι επεξηγηματικές μεταβλητές συνεισφέρουν στην σωστή ταξινόμηση των ατόμων, των οποίων η ιδιότητα είναι ήδη γνωστή (supervised classification). Για τη διαχώριση σε k ομάδες χρειάζονται $k - 1$ διαχωριστές (discriminators).

Οι συναρτήσεις που χρειάζονται για τη διαχωριστική ανάλυση βρίσκονται στη βιβλιοθήκη MASS. Ως παράδειγμα θα αναπτυχθεί η διαχωριστική ανάλυση στο πλαίσιο δεδομένων possum, το οποίο χρησιμοποιήθηκε και στην πιο πάνω ανάλυση σε κύριες συνιστώσες. Αυτό που θα εξεταστεί είναι το αν είναι δυνατόν, βάσει των μορφομετρικών μετρήσεων, να διακριθούν τα ζώα από διάφορες περιοχές (sites). Αυτό γίνεται χρησιμοποιώντας την εντολή lda όπως πιο κάτω.

```
> library(MASS)
> here <- !is.na(possum$footlgth)
> possum.lda <- lda(site ~ hdlngth+skullw+totlngth+ taill+footlgth+
+ earconch+eye+chest+belly, data=possum, subset=here)
> possum.lda
Call:
lda(site ~ hdlngth + skullw + totlngth + taill + footlgth + earconch +
    eye + chest + belly, data = possum, subset = here)

Prior probabilities of groups:
      1      2      3      4      5      6      7
0.32039 0.11650 0.06796 0.06796 0.12621 0.12621 0.17476
```

Group means:

	hdlngth	skullw	totlngth	taill	footlgth	earconch	eye	chest	belly
1	93.72	57.20	89.71	36.38	73.00	52.58	15.02	27.88	33.27
2	89.85	55.13	81.67	34.67	70.75	52.11	14.37	26.29	31.17
3	94.57	58.90	88.07	37.21	66.60	45.26	16.07	27.57	34.86
4	97.61	61.69	92.24	39.71	68.93	45.84	15.47	29.64	34.57
5	92.18	56.23	86.92	37.65	64.73	43.87	15.38	26.65	32.04
6	89.25	54.19	84.54	37.65	63.07	43.97	15.34	25.23	31.50
7	92.63	57.23	85.69	37.69	65.74	45.86	14.47	26.14	31.92

Coefficients of linear discriminants:

	LD1	LD2	LD3	LD4	LD5	LD6
hdlngth	-0.150534	0.05833	-0.25569	-0.01238	-0.08195	-0.18713
skullw	-0.026530	0.03985	-0.24977	0.12448	-0.13648	0.14378
totlngth	0.106957	0.27924	0.30519	-0.18493	-0.13904	-0.08918
taill	-0.450631	-0.08960	-0.44579	-0.17304	0.32416	0.49507
footlgth	0.301903	-0.03601	-0.03909	0.07558	0.11910	-0.12606
earconch	0.586270	-0.04357	-0.07310	-0.08819	-0.06376	0.28015
eye	-0.056141	0.08921	0.78453	0.46439	0.28481	0.29723
chest	0.090620	0.10418	0.04985	0.12749	0.64746	-0.07871
belly	0.009966	-0.05166	0.09358	0.16715	-0.29033	0.19391

Proportion of trace:

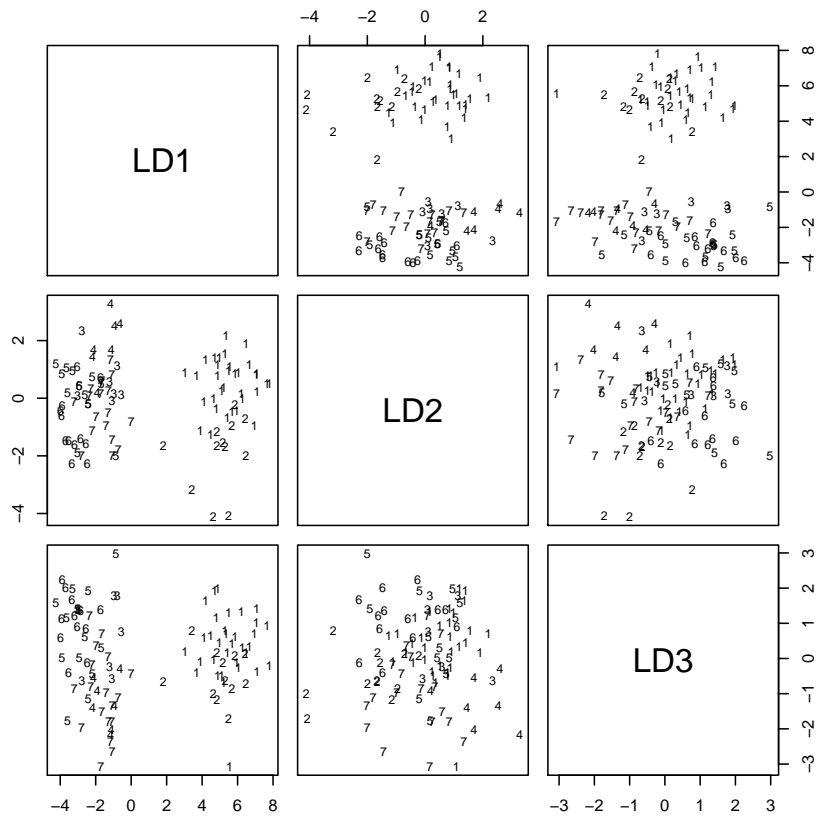
LD1	LD2	LD3	LD4	LD5	LD6
0.8927	0.0557	0.0365	0.0082	0.0047	0.0022

```
> options(digits=4)
```

```
> possum.lda$svd # Examine the singular values
```

```
[1] 15.7578 3.9372 3.1860 1.5078 1.1420 0.7772
```

Το αντικείμενο της διαχωριστικής ανάλυσης (στο παράδειγμα, `possum.lda`) περιέχει τις αρχικές αναλογίες (prior probabilities) των επιπέδων της εξαρτημένης μεταβλητής, τις μέσες τιμές κάθε επεξηγηματικής μεταβλητής για κάθε επίπεδο της εξαρτημένης μεταβλητής, τους συντελεστές των διαχωριστών και την αναλογία του ίχνους κάθε διαχωριστή. Οι διαχωριστές ταξινομούνται σε σειρά, από αυτόν με το μεγαλύτερο ίχνος σε αυτόν με το μικρότερο ίχνος. Από την αναλογία του ίχνους κάθε διαχωριστή παρατηρείται ότι οι τελευταίοι τρεις έχουν πολύ μικρή



Σχήμα 20.3: Διάγραμμα διασπορών των τιμών των τριών πρώτων κανονικών μεταβλητών.

```
> tab<-table(possum[here,]$site,predict(possum.lda)$class)
> tab
```

```
      1  2  3  4  5  6  7
1 33  0  0  0  0  0  0
2  3  9  0  0  0  0  0
3  0  0  6  1  0  0  0
4  0  0  0  6  0  0  1
5  0  0  0  0  7  3  3
6  0  0  1  0  2  9  1
7  0  0  0  0  2  0 16
```

```
> sum(diag(tab))/sum(tab)
[1] 0.835
```

Ένας τρόπος για να γίνει σταυρωτή-επικύρωση (cross-validation) είναι να αφαιρούνται παρατηρήσεις, μία κάθε φορά, και μετά χρησιμοποιώντας τα εναπομείναντα δεδομένα να γίνει εκτίμηση για την παρατήρηση που αφαιρέθηκε. Για να εκτελεστεί αυτή η μέθοδος, είναι αρκετό να προστεθεί το όρισμα $CV=T$ στην εντολή της διαχωριστικής ανάλυσης. Πρέπει να αναφερθεί ότι τώρα η πρόβλεψη της ταξινόμησης δίνεται απ' ευθείας από το αντικείμενο της διαχωριστικής ανάλυσης.

```
> possum.lda.cv <- lda(site ~ hdlngth+skullw+totlngth+ taill+footlght+
+ earconch+eye+chest+belly, data=possum, subset=here, CV=T)
> tab.cv<-table(possum[here,]$site,possum.lda.cv$class)
> tab.cv
```

```
      1  2  3  4  5  6  7
1 31  2  0  0  0  0  0
2  4  7  0  0  0  0  1
3  0  0  3  1  1  0  2
4  0  0  1  4  0  0  2
5  0  0  1  0  4  4  4
6  0  0  1  0  4  7  1
7  0  0  1  2  3  0 12
```

```
> sum(diag(tab.cv))/sum(tab.cv)
[1] 0.6602
```

