

**ΓΕΩΠΟΝΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΦΥΤΙΚΗΣ ΠΑΡΑΓΩΓΗΣ
ΕΡΓΑΣΤΗΡΙΟ ΒΕΛΤΙΩΣΗΣ ΦΥΤΩΝ ΚΑΙ ΓΕΩΡΓΙΚΟΥ ΠΕΙΡΑΜΑΤΙΣΜΟΥ**



**ΑΝΑΛΥΣΕΙΣ ΔΕΔΟΜΕΝΩΝ ΓΕΩΡΓΙΚΟΥ ΠΕΙΡΑΜΑΤΙΣΜΟΥ
ΜΕ ΤΟ ΣΤΑΤΙΣΤΙΚΟ ΠΑΚΕΤΟ R**

ΚΑΤΣΙΛΕΡΟΣ ΑΝΑΣΤΑΣΙΟΣ

ΑΘΗΝΑ 2013

Περιεχόμενα

Πρόλογος	2
Εντελώς Τυχαιοποιημένο Σχέδιο	3
Πολλαπλές Συγκρίσεις	7
Ιεραρχική Ανάλυση Διασποράς	13
Τυχαιοποιημένες Πλήρεις Ομάδες.....	15
Λατινικό Τετράγωνο.....	18
Συσχέτιση – Παλινδρόμηση	21
Ανάλυση Συνδιασποράς	24
Πολλαπλή Παλινδρόμηση.....	27
Παραγοντικά Πειράματα (2 παράγοντες)	30
Παραγοντικά Πειράματα (3 παράγοντες)	36
Υποδιαιρεμένα Τεμάχια.....	39
Ιεραρχική Ανάλυση συστάδων	41
Διακριτική Ανάλυση	43
Ανάλυση Κυρίων Συνιστωσών	46
Παραπομπές.....	48

Πρόλογος

Το R-project (<http://www.r-project.org/>) αποτελεί μια ανοικτή γλώσσα προγραμματισμού και ένα περιβάλλον για στατιστικές αναλύσεις και γραφικές απεικονίσεις. Είναι ελεύθερα διαθέσιμο σε περιβάλλον Linux, Windows και MacOS. Το βασικό πακέτο R περιλαμβάνει τις κυριότερες στατιστικές τεχνικές και γραφήματα αλλά μπορεί να επεκταθεί με πακέτα που είναι διαθέσιμα μέσω του δικτύου CRAN (<http://cran.r-project.org/>) και καλύπτουν ένα ευρύ φάσμα των σύγχρονων στατιστικών αναλύσεων. Παρότι το πακέτο R δεν διαθέτει το φιλικότερο περιβάλλον, ο χρήστης του όμως επωφελείται από την υποστήριξη που υπάρχει από μια σημαντική κοινότητα ερευνητών και ότι η ενασχόληση του τον βοηθάει σταδιακά να κατανοήσει το θεωρητικό υπόβαθρο των στατιστικών αναλύσεων.

Το παρόν φυλλάδιο παρακάμπτοντας τις βασικές εισαγωγικές έννοιες και εντολές της γλώσσας R, επικεντρώνεται στην ανάλυση κλασικών παραδειγμάτων γεωργικού πειραματισμού. Έχει γίνει προσπάθεια οι στατιστικές αναλύσεις να γίνονται με τον πιο απλό και φιλικότερο τρόπο. Τα παραδείγματα προέρχονται από τα βιβλία “Απλά Πειραματικά Σχέδια” (1997) και “Γεωργικός Πειραματισμός Παραγοντικά Πειράματα” (1989) (Π.Ι. Καλτσίκης – Εκδόσεις Σταμούλη) στα οποία παραπέμπονται όσοι επιθυμούν να εμβαθύνουν τις γνώσεις στις στατιστικές αναλύσεις. Για τις αναλύσεις χρησιμοποιήθηκε η έκδοση R 2.15.3-win και τα πακέτα `agricolae_1.1-3` και `MASS_7.3-23`.

1. Εντελώς Τυχαιοποιημένο Σχέδιο (Completely Randomized Design)

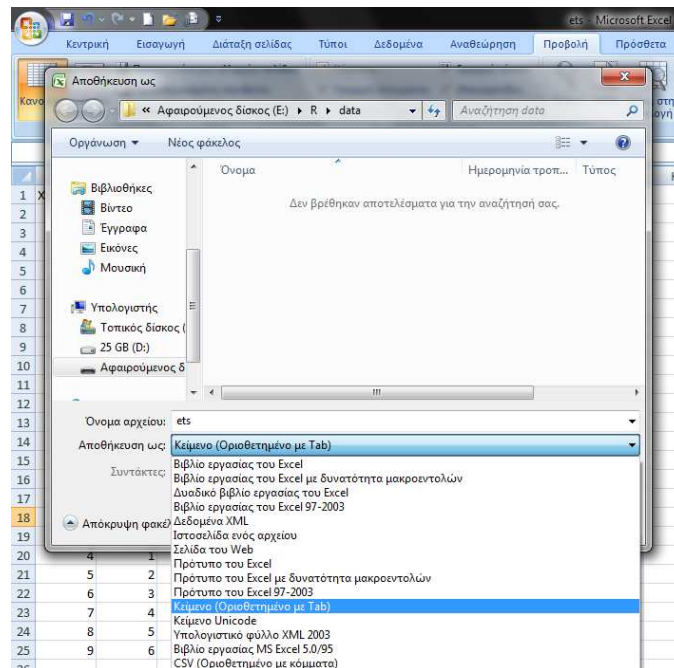
Τα δεδομένα προέρχονται από πείραμα λίπανσης σιταριού με τέσσερις επεμβάσεις και έξι επαναλήψεις. Οι επεμβάσεις τοποθετήθηκαν στα πειραματικά τεμάχια σύμφωνα με το εντελώς τυχαιοποιημένο σχέδιο. Οι μετρήσεις στον πίνακα, αντιπροσωπεύουν τα κιλά ανά πειραματικό τεμάχιο.

ΠΑΡΑΤΗΡΗΣΗ						
ΕΠΕΜΒΑΣΗ	1	2	3	4	5	6
Μάρτυρας	99	40	61	72	76	84
$K_2O + N$	96	84	82	104	99	105
$K_2O + P_2O_5$	63	57	81	59	64	72
$N + P_2O_6$	79	92	91	87	78	71

Οι επεμβάσεις, οι επαναλήψεις και τα δεδομένα αντιγράφονται σε ένα υπολογιστικό αρχείο Excel κατά στήλες. Για λόγους ευκολίας μετονομάζουμε τους τίτλους των επεμβάσεων ως X, των επαναλήψεων ως Rep και των δεδομένων ως Y, ενώ τις επεμβάσεις ως 1, 2, 3 και 4.

	A	B	C	D	E	F
1	X	Rep	Y			
2	1	1	99			
3	1	2	40			
4	1	3	61			
5	1	4	72			
6	1	5	76			
7	1	6	84			
8	2	1	96			
9	2	2	84			
10	2	3	82			
11	2	4	104			
12	2	5	99			
13	2	6	105			
14	3	1	63			
15	3	2	57			
16	3	3	81			
17	3	4	59			
18	3	5	64			
19	3	6	72			
20	4	1	79			
21	4	2	92			
22	4	3	91			
23	4	4	87			
24	4	5	78			
25	4	6	71			
26						

Στην συνέχεια τα δεδομένα από το υπολογιστικό φύλλο Excel αποθηκεύονται σε αρχείο κειμένου (ets.txt), οριοθετημένο με Tab.



και τα δεδομένα στο αρχείο κειμένου εμφανίζονται με την παρακάτω μορφή.

Αρχείο	Επεξεργασία	Μορφή	Προβολή	Βοήθεια
X	Rep	Y		
1	1	99		
1	2	40		
1	3	61		
1	4	72		
1	5	76		
1	6	84		
2	1	96		
2	2	84		
2	3	82		
2	4	104		
2	5	99		
2	6	105		
3	1	63		
3	2	57		
3	3	81		
3	4	59		
3	5	64		
3	6	72		
4	1	79		
4	2	92		
4	3	91		
4	4	87		
4	5	78		
4	6	71		

Στη κονσόλα R οι εντολές γράφονται μετά το σύμβολο `>` και εμφανίζονται με κόκκινο χρώμα. Η πρώτη εντολή είναι η `ets=read.table("E:\\R\\data\\ets.txt", header=TRUE, dec=",")`.

Η εργασία ορίζετε ως `ets`, ενώ με την εντολή `E:\\R\\data\\ets.txt` ορίζετε η διεύθυνση στην οποία έχουμε αποθηκεύσει το αρχείο κειμένου με τα δεδομένα. Η εντολή `header=TRUE` χαρακτηρίζει την πρώτη γραμμή των δεδομένων ως τίτλος - επικεφαλίδα και η εντολή `dec=","` δηλώνει ότι η δεκαδική υποδιαστολή είναι το κόμμα. Η επισυνάψη των δεδομένων γίνεται με την εντολή `attach(ets)` και δίνοντας την εντολή `ets` εμφανίζονται τα δεδομένα.

```

> ets=read.table("E:\\R\\data\\ets.txt",header=TRUE,dec=",")
> attach(ets)
> ets
  X Rep  Y
1  1  1  99
2  1  2  40
3  1  3  61
4  1  4  72
5  1  5  76
6  1  6  84
7  2  1  96
8  2  2  84
9  2  3  82
10 2  4 104
11 2  5  99
12 2  6 105
13 3  1  63
14 3  2  57
15 3  3  81
16 3  4  59
17 3  5  64
18 3  6  72
19 4  1  79
20 4  2  92
21 4  3  91
22 4  4  87
23 4  5  78
24 4  6  71

```

Χαρακτηρίζουμε ως παράγοντες (factor) τις επεμβάσεις (X) και τις επαναλήψεις (Rep), `X=factor(X); Rep=factor(Rep)`. ΠΡΟΣΟΧΗ τα γράμματα στις εντολές πρέπει να είναι ακριβώς ίδια με τα γράμματα των επικεφαλίδων (κεφαλαία-μικρά, ελληνικά-αγγλικά). Εισάγουμε το μοντέλο (δεδομένα Y: επεμβάσεις X) `fit=aov(Y~X)` και με την εντολή `summary(fit)` εμφανίζεται ο πίνακας ανάλυσης διασποράς.

```

> summary(fit)
              Df Sum Sq Mean Sq F value  Pr(>F)
X                3   2940   980.0    5.99 0.00439 **
Residuals       20   3272   163.6
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Πηγή Παραλλακτικότητας	BE	AT	MT	F	F πίνακα	Prob>F
Επέμβαση	3	2940	980,0	5,99	3,098	0,00439 **
Υπόλοιπο	20	3272	163,6			
Σύνολο	23	6212				

Από την δοκιμασία του F (Prob>F 0,00439) απορρίπτουμε την μηδενική υπόθεση H_0 . Για να ελέγξουμε πως διαφέρουν οι μέσοι των επεμβάσεων προχωρούμε σε συγκρίσεις των μέσων. Ο έλεγχος στο βασικό πακέτο R γίνεται με τη μέθοδο της Έντιμης Σημαντικής Διαφοράς (Honest Significance Difference, HSD) του Tukey, με την εντολή `TukeyHSD(fit)`.

```

R Console
> TukeyHSD(fit)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Y ~ X)

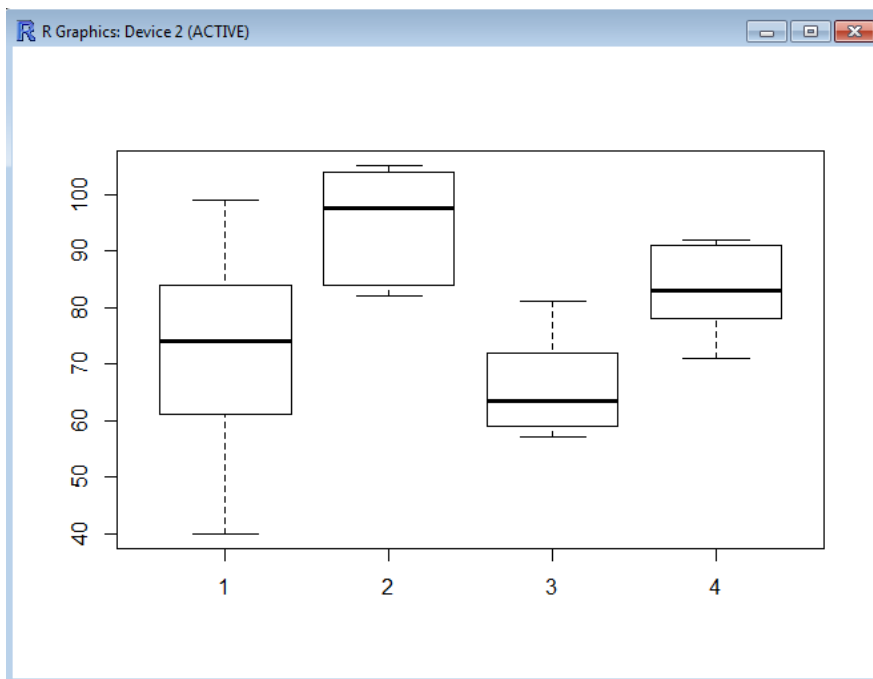
$X
      diff      lwr      upr    p adj
2-1    23  2.330784 43.669216 0.0258181
3-1   -6 -26.669216 14.669216 0.8478353
4-1    11 -9.669216 31.669216 0.4618957
3-2  -29 -49.669216 -8.330784 0.0042663
4-2  -12 -32.669216  8.669216 0.3878484
4-3    17  -3.669216 37.669216 0.1310864

```

Επομένως από την δοκιμασία του Tukey σε επίπεδο σημαντικότητας 0,05, έχουμε τις εξής διαφορές:

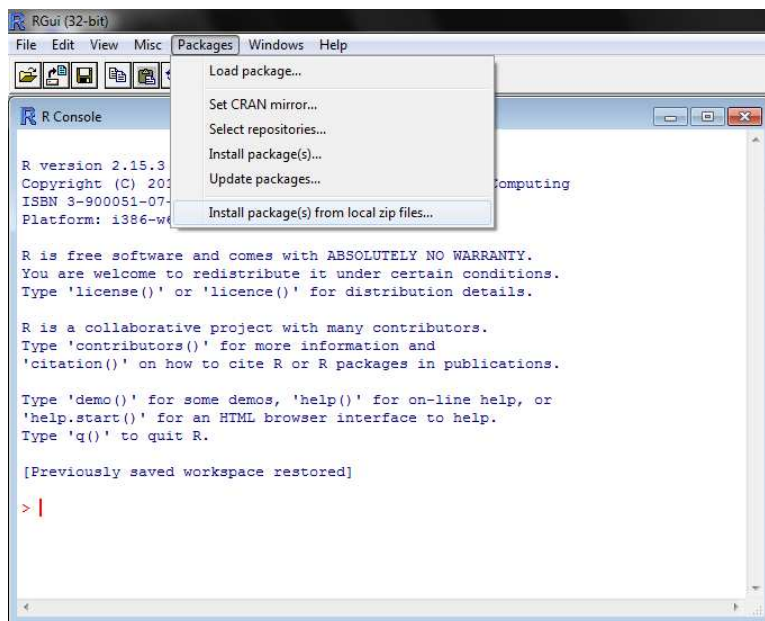
Κωδικοποίηση	Επέμβαση	Μέσος όρος	
2	K ₂ O + N	95	a
4	N + P ₂ O ₆	83	a b
1	Μάρτυρας	72	b
3	K ₂ O + P ₂ O ₅	66	b

Η εντολή `boxplot(Y~X)` αποδίδει το θηκόγραμμα.

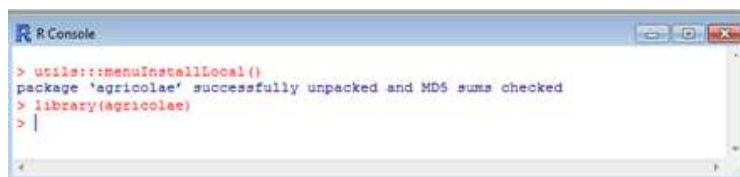


2. Πολλαπλές Συγκρίσεις (Multiple Comparisons)

Για την πραγματοποίηση επιπλέον και πιο λεπτομερών μεθόδων πολλαπλών συγκρίσεων, είναι απαραίτητη η λήψη και εγκατάσταση του πακέτου *Agricolae* (<http://cran.at.r-project.org/web/packages/agricolae/index.html>). Η εγκατάσταση του συμπιεσμένου πακέτου (.zip) γίνεται από το μενού **Packages** → **Install package(s) from local zip files...**



Δίνοντας την εντολή `library(agricolae)` στην κονσόλα R φορτώνουμε το πακέτο *Agricolae*.



2.1 Η μέθοδος της Ελάχιστης Σημαντικής Διαφοράς (Least Significance Difference)

Δίνοντας την εντολή `LSD.test(fit, "X", alpha=0.05)`, όπου $\alpha=0.05$ δηλώνουμε το επίπεδο σημαντικότητας, προχωρούμε στις συγκρίσεις των μέσων των επεμβάσεων με την μέθοδο της ΕΣΔ.

```

R Console
> LSD.test(fit, "X", alpha=0.05)

Study:

LSD t Test for Y

Mean Square Error: 163.6

X, means and individual ( 95 %) CI

  Y  std.err r    LCL      UCL Min. Max.
1 72 8.234076 6 54.82402 89.17598 40 99
2 95 4.033196 6 86.58690 103.41310 82 105
3 66 3.669696 6 58.34515 73.65485 57 81
4 83 3.396076 6 75.91591 90.08409 71 92

alpha: 0.05 ; Df Error: 20
Critical Value of t: 2.085963

Least Significant Difference 15.40415
Means with the same letter are not significantly different.

Groups, Treatments and means
a      2      95
ab     4      83
bc     1      72
c      3      66
> |

```

Η Ελάχιστη Σημαντική Διάφορα υπολογίζεται από τον τύπο:

$$ΕΣΔ = t[\alpha, a(n - 1)] * \sqrt{\frac{2 * MT_{\text{τυπ}}}{n}} = t[0,05, 4(6 - 1)] * \sqrt{\frac{2 * 163,6}{6}} = 2,086 * \sqrt{54,5} = 15,40$$

Κωδικοποίηση	Επέμβαση	Μέσος όρος	
2	K ₂ O + N	95	a
4	N + P ₂ O ₆	83	a b
1	Μάρτυρας	72	b c
3	K ₂ O + P ₂ O ₅	66	c

2.2 Η μέθοδος Duncan

Δίνοντας την εντολή `duncan.test(fit, "X")`, προχωρούμε στις συγκρίσεις των μέσων των επεμβάσεων με την μέθοδο Duncan. Αν δεν επιλεγεί επίπεδο σημαντικότητας (α) τότε το πρόγραμμα θεωρεί δεδομένο το επίπεδο σημαντικότητας 0,05.

```

R Console
> duncan.test(fit, "X")

Study:

Duncan's new multiple range test
for Y

Mean Square Error: 163.6

X, means

  Y  std.err r Min. Max.
1 72 8.234076 6  40  99
2 95 4.033196 6  82 105
3 66 3.669696 6  57  81
4 83 3.396076 6  71  92

alpha: 0.05 ; Df Error: 20

Critical Range
      2      3      4
15.40415 16.16918 16.65538

Means with the same letter are not significantly different.

Groups, Treatments and means
a      2      95
ab     4      83
bc     1      72
c      3      66
> |
  
```

Η Ελάχιστη Σημαντική Περιοχή υπολογίζεται από τον τύπο:

$$ΕΣΠ = \sqrt{\frac{ΜΤυπ}{n}} * (\text{κρίσιμες τιμές } R)$$

Η κρίσιμη τιμή R υπολογίζεται από τον αντίστοιχο Πίνακα της δοκιμασίας Duncan (αριθμός μέσων προς σύγκριση n, $ΒΕ_{\text{υπολοίπου}}$ και $\alpha=0,05$).

Υπολογισμός Ελάχιστης Σημαντικής Περιοχής			
Αριθμός μέσων προς σύγκριση (n)	2	3	4
Κρίσιμη τιμή από Πίνακα (R)	2,95	3,097	3,19
Κρίσιμη Περιοχή	15,40	16,16	16,65

Κωδικοποίηση	Επέμβαση	Μέσος όρος	
2	K ₂ O + N	95	a
4	N + P ₂ O ₆	83	a b
1	Μάρτυρας	72	b c
3	K ₂ O + P ₂ O ₅	66	c

2.3 Η μέθοδος Student-Newman-Keul (SNK)

Με την εντολή `SNK.test(fit, "X")`, προχωρούμε στις συγκρίσεις των μέσων των επεμβάσεων με την μέθοδο Student-Newman-Keul.

```

R Console
> SNK.test(fit, "X")

Study:

Student Newman Keuls Test
for Y

Mean Square Error: 163.6

X, means

  Y  std.err r Min. Max.
1 72 8.234076 6  40  99
2 95 4.033196 6  82 105
3 66 3.669696 6  57  81
4 83 3.396076 6  71  92

alpha: 0.05 ; Df Error: 20

Critical Range
      2      3      4
15.40415 18.68308 20.66922

Means with the same letter are not significantly different.

Groups, Treatments and means
a      2      95
ab     4      83
b      1      72
b      3      66
> |

```

Η Κρίσιμη Περιοχή υπολογίζεται από τον τύπο:

$$ΚΠ = \sqrt{\frac{MT\upsilon\pi}{n}} * q_{\alpha}(p, BE\upsilon\pi)$$

Η κρίσιμη τιμή q_{α} υπολογίζεται από τον αντίστοιχο Πίνακα της δοκιμασίας Student-Newman-Keul και Tukey (αριθμός μέσων προς σύγκριση p , $BE_{\text{υπολοίτου}}$ και $\alpha=0,05$).

Υπολογισμός Κρίσιμης Περιοχής			
Αριθμός μέσων προς σύγκριση (p)	2	3	4
Κρίσιμη τιμή από Πίνακα (q_{α})	2,95	3,58	3,96
Κρίσιμη Περιοχή	15,40	18,68	20,66

Κωδικοποίηση	Επέμβαση	Μέσος όρος	
2	K ₂ O + N	95	a
4	N + P ₂ O ₆	83	a b
1	Μάρτυρας	72	b
3	K ₂ O + P ₂ O ₅	66	b

2.3 Η μέθοδος της Έντιμης Σημαντικής Διαφοράς (HSD Tukey)

Με την εντολή `HSD.test(fit, "X")`, προχωρούμε στις συγκρίσεις των μέσων των επεμβάσεων με την μέθοδο Tukey.

```

R Console
> HSD.test(fit, "X")

Study:

HSD Test for Y

Mean Square Error: 163.6

X, means

  Y  std.err r Min. Max.
1 72 8.234076 6  40  99
2 95 4.033196 6  82 105
3 66 3.669696 6  57  81
4 83 3.396076 6  71  92

alpha: 0.05 ; Df Error: 20
Critical Value of Studentized Range: 3.958293

Honestly Significant Difference: 20.66922

Means with the same letter are not significantly different.

Groups, Treatments and means
a      2      95
ab     4      83
b      1      72
b      3      66
> |
  
```

Η Κρίσιμη Περιοχή υπολογίζεται από τον τύπο:

$$ΚΠ = \sqrt{\frac{ΜΤυπ}{n}} * q_{\alpha}(p_{max}, ΒΕυπ) = \sqrt{\frac{163.6}{6}} * 3.958 = 20.669$$

Η κρίσιμη τιμή q_{α} υπολογίζεται από τον αντίστοιχο Πίνακα της δοκιμασίας Student-Newman-Keul και Tukey (μέγιστος αριθμός μέσων προς σύγκριση p_{max} , $ΒΕ_{υπολοίπου}$ και $\alpha=0,05$).

Κωδικοποίηση	Επέμβαση	Μέσος όρος	
2	K ₂ O + N	95	a
4	N + P ₂ O ₆	83	a b
1	Μάρτυρας	72	b
3	K ₂ O + P ₂ O ₅	66	b

2.4 Η Μέθοδος Scheffe

Με την εντολή `scheffe.test(fit, "X")`, προχωρούμε στις συγκρίσεις των μέσων των επεμβάσεων με την μέθοδο Scheffe.

```
R Console
> scheffe.test(fit, "X")

Study:
Scheffe Test for Y
Mean Square Error : 163.6

X, means
  Y  std.err r Min. Max.
1 72 8.234076 6  40  99
2 95 4.033196 6  82 105
3 66 3.669696 6  57  81
4 83 3.396076 6  71  92

alpha: 0.05 ; Df Error: 20
Critical Value of F: 3.098391

Minimum Significant Difference: 22.51437

Means with the same letter are not significantly different.

Groups, Treatments and means
a      2      95
ab     4      83
b      1      72
b      3      66
> |
```

Η Κρίσιμη Περιοχή υπολογίζεται από τον τύπο:

$$ΚΠ = \sqrt{(a-1)[F(\alpha-1), a(n-1)]} * \sqrt{\frac{2 * ΜΤυπ}{n}} = \sqrt{(4-1)(3,10)} * \sqrt{\frac{2 * 163,6}{6}} = 22,52$$

Κωδικοποίηση	Επέμβαση	Μέσος όρος	
2	K ₂ O + N	95	a
4	N + P ₂ O ₆	83	a b
1	Μάρτυρας	72	b
3	K ₂ O + P ₂ O ₅	66	b

3. Ιεραρχική Ανάλυση Διασποράς (Nested Analysis of Variance)

Δώδεκα ροδακινιές ψεκάστηκαν με τρία διαφορετικά παρασκευάσματα λιπασμάτων που υποτίθεται αυξάνουν την περιεκτικότητα των ροδακινιών σε ζάχαρο. Για κάθε παρασκεύασμα χρησιμοποιήθηκαν τέσσερα δένδρα ροδακινιάς. Μία εβδομάδα μετά τον ψεκασμό μετρήθηκε η περιεκτικότητα σε ζάχαρο σε έξι ροδάκινα, που πάρθηκαν τυχαία από κάθε δένδρο ροδακινιάς.

Λίπασμα	Ροδάκινο	Δέντρο			
		1	2	3	4
1	1	4,5	5,78	13,22	11,59
1	2	7,04	7,69	15,05	8,96
1	3	4,98	12,68	12,67	10,95
1	4	5,48	5,89	12,42	9,87
1	5	6,54	4,07	10,03	10,48
1	6	7,2	4,08	13,5	12,79
2	1	15,32	14,53	10,89	15,12
2	2	14,97	14,51	10,27	13,79
2	3	14,81	12,61	12,21	15,32
2	4	14,26	16,13	12,77	11,95
2	5	15,88	13,65	10,45	12,56
2	6	16,01	14,78	11,44	15,31
3	1	7,18	6,7	5,94	4,08
3	2	7,98	6,68	5,78	5,46
3	3	5,51	6,99	7,59	5,4
3	4	7,48	6,4	7,21	6,85
3	5	7,55	4,96	6,12	7,74
3	6	5,64	7,03	7,13	6,81

Εισάγουμε τα δεδομένα σε ένα υπολογιστικό φύλλο Excel, τοποθετώντας τα σε τέσσερις στήλες. Η πρώτη στήλη περιλαμβάνει τις επεμβάσεις X (παρασκευάσματα), η δεύτερη τα δένδρα (TR), η τρίτη τα ροδάκινα (PE) και η τέταρτη τα ζάχαρα (Y). Στη συνέχεια αποθηκεύουμε τα δεδομένα σε αρχείο κειμένου (.txt) και τα εισάγουμε στην κονσόλα R δίνοντας με την σειρά τις εντολές `nest=read.table("G:\\R\\data\\nest.txt", header=TRUE, dec=",")`, `attach(nest)` και `nest`. Εναλλακτικά με το σύμβολο ; έχουμε την δυνατότητα να ενώσουμε τις παραπάνω εντολές σε μία.

```
nest=read.table("G:\\R\\data\\nest.txt", header=TRUE, dec=",") ; attach(nest) ; nest
```

```

R Console
> nest=read.table("G:\\R\\data\\nest.txt", header=TRUE, dec=",")
> attach(nest)
> nest
  X TR PE  Y
1  1  1  1  4.50
2  1  1  2  7.04
3  1  1  3  4.98
4  1  1  4  5.48
5  1  1  5  6.54
6  1  1  6  7.20
7  1  2  1  5.78
8  1  2  2  7.69
9  1  2  3 12.68
10 1  2  4  5.89

```

Χαρακτηρίζουμε ως παράγοντες (factor) τις επεμβάσεις (X), τα δένδρα (TR) και τα ροδάκια (PE), $X=\text{factor}(X)$; $TR=\text{factor}(TR)$; $PE=\text{factor}(PE)$, εισάγουμε το μοντέλο (δεδομένα Y: επεμβάσεις X και το υπόλοιπο X:TR) $\text{fit}=\text{aov}(Y\sim X+\text{Error}(X:TR))$ και με την εντολή $\text{summary}(\text{fit})$ εμφανίζεται ο πίνακας ανάλυσης διασποράς.

```

R Console
> X=factor(X); TR=factor(TR); PE=factor(PE)
> fit=aov(Y~X+Error(X:TR))
Warning message:
In aov(Y ~ X + Error(X:TR)) : Error() model is singular
> summary(fit)

Error: X:TR
      Df Sum Sq Mean Sq F value Pr(>F)
X       2  643.8   321.9   11.78 0.00307 **
Residuals  9   246.0    27.3
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: Within
      Df Sum Sq Mean Sq F value Pr(>F)
Residuals 60  126.2    2.104
> |

```

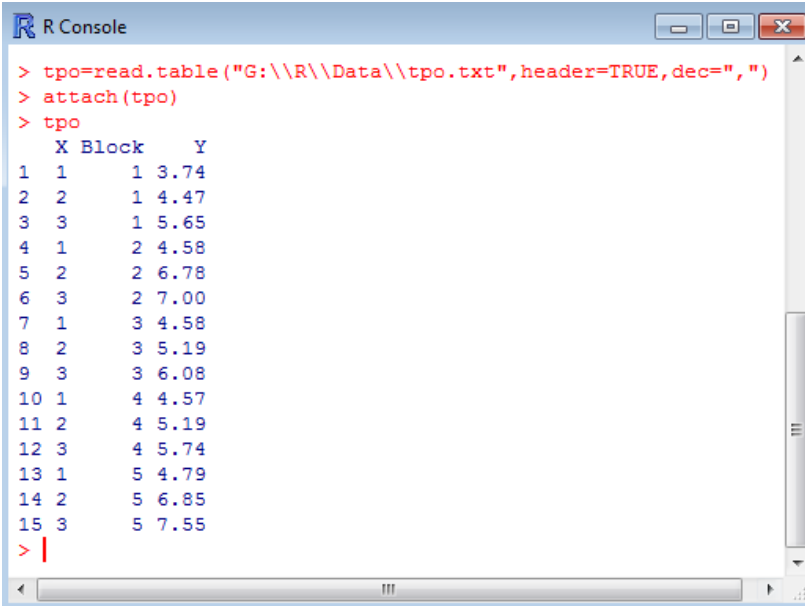
Πηγή Παραλλακτικότητας	BE	AT	MT	F	F πίνακα	Prob>F
Δένδρα	11	889,8				
Λιπάσματα	2	643,8	321,9	11,78	4,26	0,00307 **
Υπόλοιπο (Λιπάσματα/Δέντρα)	9	246	27,3			
Υπόλοιπο	60	126	2,104			
Σύνολο	73	1015,8				

4. Τυχαιοποιημένες Πλήρεις Ομάδες (Randomized Complete Block Design)

Τα δεδομένα προέρχονται από πείραμα με τρεις επεμβάσεις-σιτηρέσια που δοκιμάστηκαν σε πέντε χοιροστάσια (ομάδες). Μετρήθηκε η αύξηση του βάρους σε κιλά κατά χοίρο.

	Χοιροστάσια				
Σιτηρέσιο	1	2	3	4	5
1	3,74	4,58	4,58	4,57	4,79
2	4,47	6,78	5,19	5,19	6,85
3	5,65	7	6,08	5,74	7,55

Τα δεδομένα τοποθετούνται σε τρεις στήλες σε ένα υπολογιστικό φύλλο Excel. Η πρώτη στήλη περιλαμβάνει τις επεμβάσεις - σιτηρέσια (X), η δεύτερη τις ομάδες - χοιροστάσια (Block) και η τρίτη τα κιλά (Y). Αποθηκεύουμε τα δεδομένα σε αρχείο κειμένου (tpo.txt) και τα εισάγουμε στην κονσόλα R δίνοντας με την σειρά τις εντολές `tpo=read.table("G:\\R\\data\\tpo.txt", header=TRUE, dec=",", attach(tpo)` και `tpo`.



```
> tpo=read.table("G:\\R\\Data\\tpo.txt", header=TRUE, dec=",")
> attach(tpo)
> tpo
  X Block  Y
1  1     1 3.74
2  2     1 4.47
3  3     1 5.65
4  1     2 4.58
5  2     2 6.78
6  3     2 7.00
7  1     3 4.58
8  2     3 5.19
9  3     3 6.08
10 1     4 4.57
11 2     4 5.19
12 3     4 5.74
13 1     5 4.79
14 2     5 6.85
15 3     5 7.55
> |
```

Χαρακτηρίζουμε ως παράγοντες (factor), τις επεμβάσεις (X) και τις ομάδες (Block), `X=factor(X); Block=factor(Block)`, εισάγουμε το μοντέλο (δεδομένα Y: επεμβάσεις X και ομάδες Block) `fit=aov(Y~X+Block)` και με την εντολή `summary(fit)` εμφανίζεται ο πίνακας ανάλυσης διασποράς.


```

R Console
> X=factor(X);Block=factor(Block)
> fit=aov(Y~X+Block)
> summary(fit)
          Df Sum Sq Mean Sq F value    Pr(>F)
X           2  9.765   4.883  24.192 0.000405 ***
Block       4  6.358   1.590   7.876 0.007048 **
Residuals   8  1.615   0.202
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |

```

Πηγή Παραλλακτικότητας	BE	AT	MT	F	F πίνακα	Prob>F
Επεμβάσεις	2	9,765	4,883	24,192	4,46	0,000405***
Ομάδες	4	6,358	1,590	7,879	3,84	0,007048**
Υπόλοιπο	8	1,615	0,202			
Σύνολο	14	17,738				

Εισάγοντας το πακέτο Agricolae με την εντολή `library(agricolae)` προχωράμε σε συγκρίσεις μέσω με τη μέθοδο της Ελάχιστης Σημαντικής Διαφοράς `HSD.test(fit, "X")`

```

R Console
> HSD.test(fit, "X")

Study:

HSD Test for Y

Mean Square Error: 0.2018283

X, means

      Y  std.err r Min. Max.
1 4.452 0.1827403 5 3.74 4.79
2 5.696 0.4754955 5 4.47 6.85
3 6.404 0.3730228 5 5.65 7.55

alpha: 0.05 ; Df Error: 8
Critical Value of Studentized Range: 4.041036

Honestly Significant Difference: 0.8118931

Means with the same letter are not significantly different.

Groups, Treatments and means
a      3      6.404
a      2      5.696
b      1      4.452
> |

```

Συγκρίσεις Μέσων (ΕΣΔ = 0,819 και α = 0,05)		
Επέμβαση	Μέσος	
3	6,404	a
2	5,696	a
1	4,452	b

Για το έλεγχο της ομοιογένειας των διασπορών χρησιμοποιούμε την δοκιμασία Bartlett δίνοντας την εντολή `bartlett.test(Y~X)`. Για τον έλεγχο της κανονικότητας πραγματοποιούμε τη δοκιμασία Shapiro-Wilk με την εντολή `shapiro.test(Y)`

```
R Console
> bartlett.test(Y~X)

      Bartlett test of homogeneity of variances

data:  Y by X
Bartlett's K-squared = 2.8924, df = 2, p-value = 0.2355

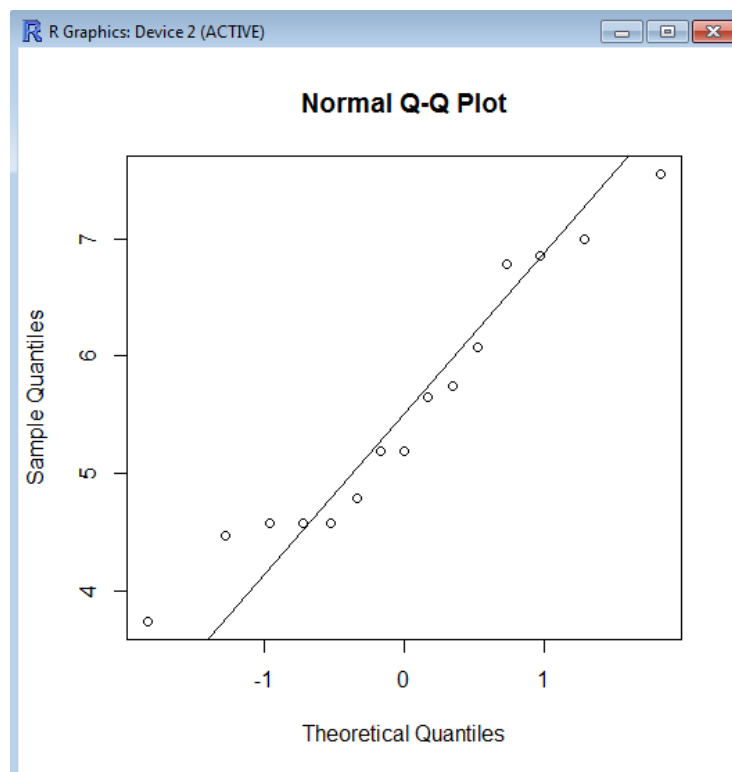
> shapiro.test(Y)

      Shapiro-Wilk normality test

data:  Y
W = 0.9418, p-value = 0.4053

> qqnorm(Y)
> qqline(Y)
> |
```

Ο γραφικός έλεγχος της κατανομής γίνεται με τις εντολές `qqnorm(Y)` και `qqline(Y)`



5. Λατινικό Τετράγωνο (Latin Square)

Έχουμε απομονώσει τέσσερις συγγενείς χημικές ουσίες και θέλουμε να εξετάσουμε κατά πόσο διαφέρουν όσον αφορά την περιεκτικότητα τους σε ένα στοιχείο, που για τον προσδιορισμό της χρειάζεται να χρησιμοποιήσουμε ως όργανο ολόκληρη τη μέρα. Το εργαστήριο διαθέτει τέσσερα τέτοια όργανα. Έτσι κάθε μέρα, επί τέσσερις μέρες, αναλύουμε μια ουσία μια μόνο κάθε μέρα και μόνο μια φορά σε κάθε όργανο. Αν και με την ανάλυση αυτή μπορούμε να μελετήσουμε τις διαφορές από μέρα σε μέρα και από όργανο σε όργανο, ενδιαφερόμαστε κυρίως για τις διαφορές των ουσιών όσον αφορά το στοιχείο αυτό.

Σειρά (ημέρες)	Στήλη (όργανα)			
	1	2	3	4
1	8,7 I	7,5 II	14 III	11,3 IV
2	9,2 II	12,7 III	9,2 IV	8,7 I
3	11,6 III	4,6 IV	5,1 I	4 II
4	9,1 IV	7,3 I	6,7 II	12,9 III

Τα δεδομένα τοποθετούνται σε τέσσερις στήλες. Η πρώτη στήλη περιλαμβάνει τις επεμβάσεις χημικές ουσίες (X, λατινικά στοιχεία), η δεύτερη τις σειρές-ημέρες (R), η τρίτη τις στήλες-όργανα (C) και η τέταρτη τα δεδομένα (Y). Στη συνέχεια αποθηκεύουμε τα δεδομένα σε αρχείο κειμένου (slt.txt) και τα εισάγουμε στην κονσόλα R δίνοντας με τη σειρά τις εντολές `slt=read.table("j:\\R\\Data\\slt.txt", header=TRUE, dec=",", attach(slt))` και `slt`.

H22					
	A	B	C	D	E
1	X	R	C	Y	
2		1	1	1	8,7
3		1	2	4	7,3
4		1	3	3	5,1
5		1	4	2	8,7
6		2	1	2	9,2
7		2	2	1	7,5
8		2	3	4	6,7
9		2	4	3	4
10		3	1	3	11,6
11		3	2	2	12,7
12		3	3	1	14
13		3	4	4	12,9
14		4	1	4	9,1
15		4	2	3	4,6
16		4	3	2	9,2
17		4	4	1	11,3

```

R Console
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> slt=read.table("j:\\R\\Data\\slt.txt",header=TRUE,dec=",")
> attach(slt)
> slt
  X R C  Y
1  1 1 1 8.7
2  1 2 4 7.3
3  1 3 3 5.1
4  1 4 2 8.7
5  2 1 2 9.2
6  2 2 1 7.5
7  2 3 4 6.7
8  2 4 3 4.0
9  3 1 3 11.6
10 3 2 2 12.7
11 3 3 1 14.0
12 3 4 4 12.9
13 4 1 4  9.1
14 4 2 3  4.6
15 4 3 2  9.2
16 4 4 1 11.3

```

Χαρακτηρίζουμε ως παράγοντες (factor), τις επεμβάσεις (X), τις σειρές (R) και τις στήλες (C) `X=factor(X); R=factor(R); C=factor(C)`, εισάγουμε το μοντέλο (δεδομένα Y: επεμβάσεις X, σειρές R και στήλες C) `fit=aov(Y~X+R+C)` και με την εντολή `summary(fit)` εμφανίζεται ο πίνακας ανάλυσης διασποράς.

```

R Console

> X=factor(X);R=factor(R);C=factor(C)
> fit=aov(Y~X+R+C)
> summary(fit)
          Df Sum Sq Mean Sq F value    Pr(>F)
X           3   86.55   28.849   42.687 0.000193 ***
R           3    5.82    1.941    2.872 0.125712
C           3   39.67   13.224   19.567 0.001683 **
Residuals   6    4.06    0.676

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Πηγή Παραλλακτικότητας	BE	AT	MT	F	F πίνακα	Prob>F
Επεμβάσεις	3	86,55	28,849	42,687	4,76	0,000193***
Σειρές	3	5,82	1,941	2,872	4,76	0,125712
Στήλες	3	39,67	13,224	19,567	4,76	0,001683**
Υπόλοιπο	6	4,06	0,676			
Σύνολο	15	136,1				

Με τις εντολές `library(agricolae)` και `HSD.test(fit, "X")` προχωράμε σε συγκρίσεις μέσω της μεθόδου της Ελάχιστης Σημαντικής Διαφοράς.

```

R Console
> library(agricolae)
> HSD.test(fit, "X")

Study:

HSD Test for Y

Mean Square Error: 0.6758333

X, means

      Y  std.err r Min. Max.
1  7.45 0.850000 4  5.1  8.7
2  6.85 1.083590 4  4.0  9.2
3 12.80 0.491596 4 11.6 14.0
4  8.55 1.410969 4  4.6 11.3

alpha: 0.05 ; Df Error: 6
Critical Value of Studentized Range: 4.895599

Honestly Significant Difference: 2.012314

Means with the same letter are not significantly different.

Groups, Treatments and means
a      3      12.8
b      4      8.55
b      1      7.45
b      2      6.85
> |

```

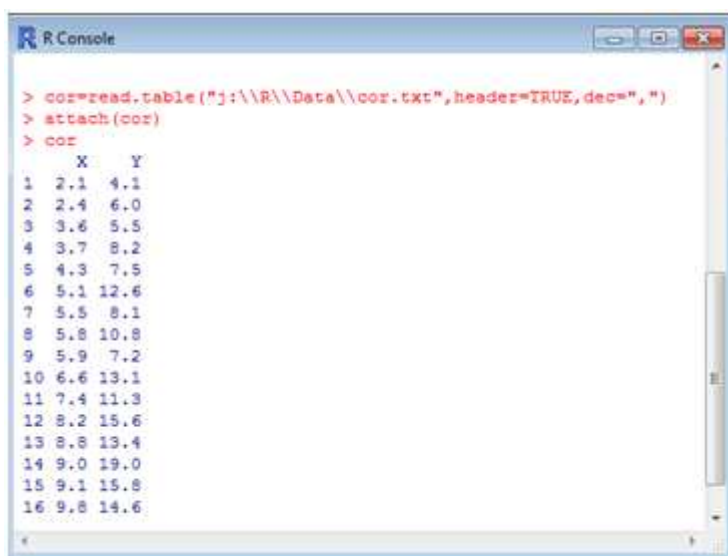
Συγκρίσεις Μέσων (ΕΣΔ = 2,012 και α = 0,05)		
Επέμβαση	Μέσος	
3	12,8	a
4	8,55	b
1	7,45	b
2	6,85	b

6. Συσχέτιση – Παλινδρόμηση (Correlation - Regression)

Τα δεδομένα αφορούν το ξερό και νωπό βάρος, σε γραμμάρια, φυταρίων σιταριού που καλλιεργήθηκαν σε τεχνητό θρεπτικό διάλυμα.

X	Y
2,1	4,1
2,4	6
3,6	5,5
3,7	8,2
4,3	7,5
5,1	12,6
5,5	8,1
5,8	10,8
5,9	7,2
6,6	13,1
7,4	11,3
8,2	15,6
8,8	13,4
9	19
9,1	15,8
9,8	14,6

Τα δεδομένα τοποθετούνται σε δύο στήλες. Η πρώτη στήλη περιλαμβάνει τη μεταβλητή (X) και τη μεταβλητή (Y). Στη συνέχεια αποθηκεύουμε τα δεδομένα σε αρχείο κειμένου (cor.txt) και τα εισάγουμε στην κονσόλα R δίνοντας τις εντολές `cor=read.table("j:\\R\\Data\\cor.txt", header=TRUE, dec=",")` και `attach(cor)`.



```
> cor=read.table("j:\\R\\Data\\cor.txt", header=TRUE, dec=",")
> attach(cor)
> cor
  X  Y
1 2.1 4.1
2 2.4 6.0
3 3.6 5.5
4 3.7 8.2
5 4.3 7.5
6 5.1 12.6
7 5.5 8.1
8 5.8 10.8
9 5.9 7.2
10 6.6 13.1
11 7.4 11.3
12 8.2 15.6
13 8.8 13.4
14 9.0 19.0
15 9.1 15.8
16 9.8 14.6
```

Για υπολογίσουμε τον συντελεστή συσχέτισης και για τον έλεγχο της σημαντικότητας, δίνουμε αντίστοιχα τις εντολές `cor(X,Y,method = "pearson")` και `cor.test(X,Y, method = "pearson")`. Στην μέθοδο μπορούμε να επιλέξουμε `pearson` ή `spearman`.

```
R Console
> cor(X,Y,method = "pearson")
[1] 0.8917818
> cor.test(X,Y, method = "pearson")

Pearson's product-moment correlation

data: X and Y
t = 7.3746, df = 14, p-value = 3.486e-06
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7098920 0.9621556
sample estimates:
      cor
0.8917818

> |
```

Ο συντελεστής συσχέτισης (r) είναι 0,89 και είναι στατιστικά σημαντικός ($t=7,374$, $p\text{-value}= 3,486e-06$ για $BE= 16$).

Η ανάλυση παλινδρόμησης πραγματοποιείται εισάγοντας το μοντέλο (εξαρτημένη μεταβλητή Y : ανεξάρτητη μεταβλητή X) `fit=lm(Y~X)` και με τις εντολές `summary(fit)` και `anova(fit)` εμφανίζονται το σημείο αποκοπής (intercept), ο συντελεστής παλινδρόμησης b και ο έλεγχος της σημαντικότητας του με τις δοκιμασίες t και F .

```
R Console
> fit=lm(Y~X)
> summary(fit)

Call:
lm(formula = Y ~ X)

Residuals:
    Min       1Q   Median       3Q      Max
-3.317 -1.579 -0.099  1.210  3.645

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.3097     1.3825   0.947   0.36
X            1.5606     0.2116   7.375 3.49e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.021 on 14 degrees of freedom
Multiple R-squared:  0.7953,    Adjusted R-squared:  0.7807
F-statistic: 54.38 on 1 and 14 DF,  p-value: 3.486e-06

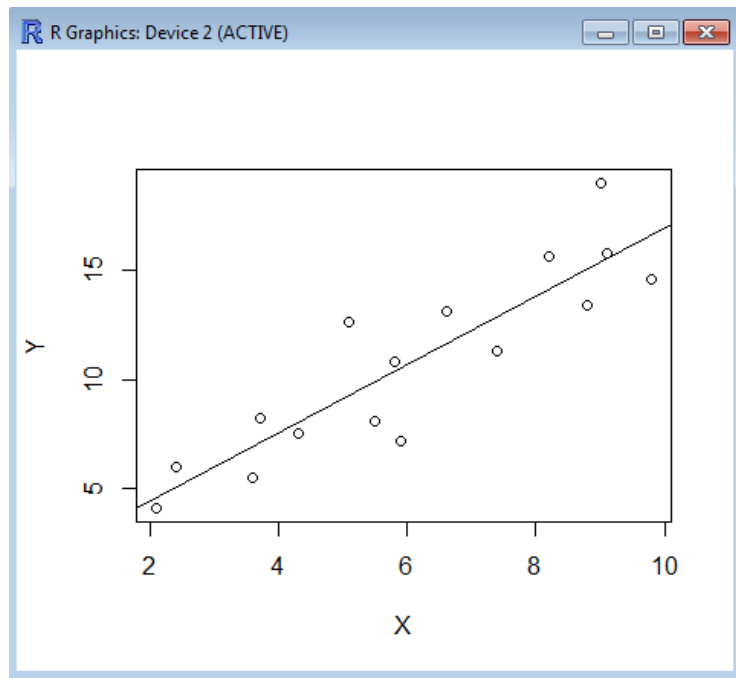
> anova(fit)
Analysis of Variance Table

Response: Y
      Df Sum Sq Mean Sq F value    Pr(>F)
X       1  222.025  222.025  54.384 3.486e-06 ***
Residuals 14  57.155    4.083
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> |
```

Πηγή Παραλλακτικότητας	BE	AT	MT	F	F πίνακα	Prob>F
Παλινδρόμηση	1	222,025	222,025	54,384	4,60	3,486e-06***
Υπόλοιπο	14	57,155	4,083			
Σύνολο	15	279,18				

Η εξίσωση είναι η εξής: $Y = 1,3097 + 1,5606 \cdot X$ και η γραφική παράσταση δημιουργείται με τις εντολές `plot(X,Y)` και `abline(lm(Y~X))`.



7. Πολλαπλή Παλινδρόμηση (Multiple Regression)

Τα δεδομένα αφορούν την απόδοση σε σπόρο (g) ανά φυτό και τα συστατικά της απόδοσης, σε 22 φυτά μιας ποικιλίας σκληρού σιταριού.

Απόδοση Y	Στάχια ανά φυτό X1	Σταχύδια ανά στάχυ X2	Σπόροι ανά σταχύδιο X3	Βάρος 100 κόκκων X4
10	14,4	9,8	1,9	4
6,8	9	14,5	1,8	2,6
18,8	15,6	12,5	2,9	3
10	10,6	20,1	1,9	2,7
13,6	13,3	13,2	2,3	3,6
4,1	5	22,3	1,4	2,9
17,4	19,4	11,7	2,2	3,5
12	12,5	18,4	2,1	3,2
11	12,8	14,2	1,9	3
12,3	13,4	14,7	1,8	3,6
7,6	8,3	17,8	1,9	2,9
13,6	15,9	13,2	2	3,5
16,2	15,3	16	2,4	3,2
14,1	15	12,5	2	3,8
11,2	12,2	14,8	2,4	3
15,3	16,3	12,5	2,6	3
12	12,4	16,4	1,8	3,3
10,1	9,6	16,1	2,1	3,1
10,9	13,8	13,3	2	2,8
9,5	9,6	16,4	1,9	3,4
10,7	14,7	11	1,8	3,8
10,9	12,3	14,3	2,1	3,1

Εισάγουμε τα δεδομένα σε ένα φύλλο Excel και τα αποθηκεύουμε σε αρχείο κειμένου (multres.txt). Στην κονσόλα R δίνουμε τις εντολές `multres=read.table("G:\\R\\Data\\cor.txt", header=TRUE, dec=",")`, `attach(multres)` και `multres`.

```

R Console
> multres=read.table("G:\\R\\Data\\multres.txt",header=TRUE,dec=",")
> attach(multres)
> multres
      Y   X1  X2  X3  X4
1  10.0 14.4  9.8  1.9  4.0
2   6.8  9.0 14.5  1.8  2.6
3  18.8 15.6 12.5  2.9  3.0
4  10.0 10.6 20.1  1.9  2.7
5  13.6 13.3 13.2  2.3  3.6
6   4.1  5.0 22.3  1.4  2.9
7  17.4 19.4 11.7  2.2  3.5
8  12.0 12.5 18.4  2.1  3.2
9  11.0 12.8 14.2  1.9  3.0
10 12.3 13.4 14.7  1.8  3.6
11  7.6  8.3 17.8  1.9  2.9
12 13.6 15.9 13.2  2.0  3.5
13 16.2 15.3 16.0  2.4  3.2
14 14.1 15.0 12.5  2.0  3.8
15 11.2 12.2 14.8  2.4  3.0
16 15.3 16.3 12.5  2.6  3.0
17 12.0 12.4 16.4  1.8  3.3
18 10.1  9.6 16.1  2.1  3.1
19 10.9 13.8 13.3  2.0  2.8
20  9.5  9.6 16.4  1.9  3.4
21 10.7 14.7 11.0  1.8  3.8
22 10.9 12.3 14.3  2.1  3.1

```

Εισάγουμε το μοντέλο (εξαρτημένη μεταβλητή Y: ανεξάρτητες μεταβλητές X1, X2, X3 και X4) $fit=lm(Y\sim X1+X2+X3+X4)$ και με τις εντολές `summary(fit)` και `anova(fit)` εμφανίζονται οι συντελεστές παλινδρόμησης, η σημαντικότητάς τους με την δοκιμασία t και ο πίνακας ανάλυσης διασποράς.

```

R Console
> fit=lm(Y~X1+X2+X3+X4)
> summary(fit)

Call:
lm(formula = Y ~ X1 + X2 + X3 + X4)

Residuals:
    Min       1Q   Median       3Q      Max
-1.52187 -0.41411 -0.01575  0.53231  1.50907

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -19.8262     4.2326  -4.684 0.000213 ***
X1           0.8463     0.1140   7.423 9.96e-07 ***
X2           0.3924     0.1090   3.598 0.002217 **
X3           5.1882     0.8627   6.014 1.40e-05 ***
X4           1.3215     0.7223   1.830 0.084898 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9032 on 17 degrees of freedom
Multiple R-squared:  0.9432,    Adjusted R-squared:  0.9299
F-statistic: 70.64 on 4 and 17 DF,  p-value: 2.311e-10

```

```

R Console
> anova(fit)
Analysis of Variance Table

Response: Y
      Df Sum Sq Mean Sq F value    Pr(>F)
X1     1 189.227  189.227  231.9422 2.432e-11 ***
X2     1  10.480   10.480  12.8459 0.002286 **
X3     1  28.081   28.081  34.4196 1.866e-05 ***
X4     1   2.731    2.731   3.3476 0.084898 .
Residuals 17  13.869    0.816
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |

```

Το μοντέλο δίνεται ως εξής:

$$Y = -19,82 + 0,84 \cdot X_1 + 0,39 \cdot X_2 + 5,18 \cdot X_3 + 1,32 \cdot X_4$$

Πηγή Παραλλακτικότητας	ΒΕ	ΑΤ	ΜΤ	F	F πίνακα	Prob>F
Μοντέλο	4	230,518	57,629	70,638	2,96	
X1	1	189,227	189,227	231,942	4,45	2,432e-11***
X2	1	10,480	10,480	12,846	4,45	0,002286**
X3	1	28,081	28,081	34,419	4,45	1,866e-05***
X4	1	2,731	2,731	3,347	4,45	0,084898
Υπόλοιπο	17	13,869	0,816			
Σύνολο	21	244,384				

Με τις εντολές `step(fit, direction='forward')` ή `step(fit, direction='backward')` προχωράμε σε σταδιακή ανάλυση παλινδρόμησης (stepwise regression) με τις μεθόδους της προοδευτικής επιλογής (forward) ή της εκ των υστέρων αποκλεισμού (backward).

```
R Console

> step(fit, direction='forward')
Start: AIC=-0.15
Y ~ X1 + X2 + X3 + X4

Call:
lm(formula = Y ~ X1 + X2 + X3 + X4)

Coefficients:
(Intercept)      X1      X2      X3      X4
-19.8262      0.8463      0.3924      5.1882      1.3215

> step(fit, direction='backward')
Start: AIC=-0.15
Y ~ X1 + X2 + X3 + X4

  Df Sum of Sq  RSS  AIC
<none>                13.869 -0.1502
- X4    1    2.731 16.600  1.8043
- X2    1   10.563 24.432 10.3067
- X3    1   29.506 43.376 22.9348
- X1    1   44.953 58.822 29.6364

Call:
lm(formula = Y ~ X1 + X2 + X3 + X4)

Coefficients:
(Intercept)      X1      X2      X3      X4
-19.8262      0.8463      0.3924      5.1882      1.3215

> |
```

και στις δύο μεθόδους επιλέγονται τελικά για το μοντέλο όλες οι ανεξάρτητες μεταβλητές.

8. Ανάλυση Συνδιασποράς (Analysis of Covariance)

Τα δεδομένα προέρχονται από πείραμα με τρεις επεμβάσεις (υβρίδια καλαμποκιού) των οχτώ επαναλήψεων. Μετρήθηκε από κάθε πειραματικό τεμάχιο η απόδοση Y (εξαρτημένη μεταβλητή) καθώς επίσης και το ύψος της βροχόπτωσης X (ανεξάρτητη μεταβλητή - συμμεταβλητή).

Tr	X	Y
1	10	15
1	6	1
1	5	4
1	8	6
1	9	10
1	4	0
1	9	7
1	12	13
2	4	6
2	8	13
2	8	5
2	8	18
2	6	9
2	11	7
2	10	15
2	9	15
3	7	14
3	8	9
3	7	16
3	3	7
3	6	13
3	8	18
3	6	13
3	8	6

Εισάγουμε τα δεδομένα σε ένα φύλλο Excel, τοποθετώντας σε στήλες. Στη συνέχεια τα αποθηκεύουμε σε αρχείο κειμένου (cov.txt) και τα εισάγουμε στην κονσόλα R δίνοντας τις εντολές `cov=read.table("G:\\R\\data\\cov.txt", header=TRUE, dec=",")`, `attach(cov)` και `cov`.

```

R Console
> cov=read.table("J:\\R\\data\\cov.txt", header=T, dec=",")
> attach(cov)
> cov
  Tr X Y
1  1 10 15
2  1  6  1
3  1  5  4
4  1  8  6
5  1  9 10
6  1  4  0
7  1  9  7
8  1 12 13
9  2  4  6
10 2  8 13
11 2  8  5
12 2  8 18
13 2  6  9
14 2 11  7
15 2 10 15
16 2  9 15
17 3  7 14
18 3  8  9
19 3  7 16
20 3  3  7
21 3  6 13
22 3  8 18
23 3  6 13
24 3  8  6

```

Χαρακτηρίζουμε ως παράγοντες (factor) τις επεμβάσεις (Tr), $Tr=factor(Tr)$, εισάγουμε το μοντέλο (δεδομένα Y: επεμβάσεις Tr) $fit=lm(Y\sim Tr)$, χωρίς να λάβουμε υπόψη τη συμμεταβλητή X και με την εντολή $anova(fit)$ εμφανίζεται ο πίνακας ανάλυσης διασποράς.

```

R Console
> Tr=factor(Tr)
> fit=lm(Y~Tr)
> anova(fit)
Analysis of Variance Table

Response: Y
      Df Sum Sq Mean Sq F value Pr(>F)
Tr      2    112   56.000   2.3614 0.1188
Residuals 21    498   23.714

```

Αν λάβουμε υπόψη τη συμμεταβλητή X, τότε το μοντέλο γράφεται ως εξής (δεδομένα Y: συμμεταβλητή X και επεμβάσεις Tr) $fit=lm(Y\sim X+Tr)$, (ΠΡΟΣΟΧΗ πρώτα η συμμεταβλητή X και μετά οι επεμβάσεις Tr) και με την εντολή $anova(fit)$ εμφανίζεται ο πίνακας ανάλυσης συνδιασποράς, με τα διορθωμένα αθροίσματα τετραγώνων.

```

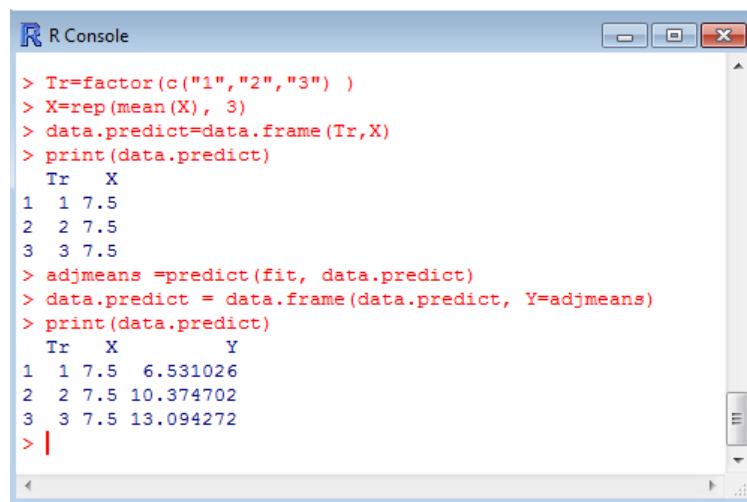
R Console
> Tr=factor(Tr)
> fit=lm(Y~X+Tr)
> anova(fit)
Analysis of Variance Table

Response: Y
      Df Sum Sq Mean Sq F value Pr(>F)
X      1  110.04  110.035   6.5855 0.01842 *
Tr      2  165.79   82.897   4.9613 0.01780 *
Residuals 20  334.17   16.709
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |

```

Για τον υπολογισμό των διορθωμένων μέσων, οι εντολές είναι οι εξής:

```
> Tr=factor(c("1","2","3")) # Τα επίπεδα της επέμβασης
> X=rep(mean(X), 3) # Ο γενικός μέσος της συμμεταβλητής
> data.predict=data.frame(Tr,X)
> print(data.predict)
> adjmeans =predict(fit, data.predict)
> data.predict = data.frame(data.predict, Y=adjmeans)
> print(data.predict)
```



```
R Console
> Tr=factor(c("1","2","3"))
> X=rep(mean(X), 3)
> data.predict=data.frame(Tr,X)
> print(data.predict)
  Tr  X
1  1 7.5
2  2 7.5
3  3 7.5
> adjmeans =predict(fit, data.predict)
> data.predict = data.frame(data.predict, Y=adjmeans)
> print(data.predict)
  Tr  X  Y
1  1 7.5 6.531026
2  2 7.5 10.374702
3  3 7.5 13.094272
> |
```

Επέμβαση	Αρχικός Μέσος	Διορθωμένος Μέσος
1	7	6,53
2	11	10,37
3	12	13,09

9. Παραγοντικά Πειράματα (Factorial Experiments)

9.1. Εντελώς Τυχαιοποιημένο Σχέδιο

Σε μια ποικιλία σιταριού χρησιμοποιήθηκαν δύο διαφορετικές δόσεις μιας χημικής ουσίας (A παράγοντας, 2 επίπεδα) για να μελετηθεί η επίδρασή της σε συνάρτηση με τη θερμοκρασία (B παράγοντας, 3 επίπεδα 15°, 20° και 25°C) στη σύζευξη των χρωμοσωμάτων. Το πείραμα έγινε μέσα σε θαλάμους στους οποίους η θερμοκρασία μπορούσε να καθοριστεί με ακρίβεια. Για κάθε συνδυασμό θερμοκρασίας-ποικιλίας χρησιμοποιήθηκαν 4 φυτά, το καθένα σε ξεχωριστή γλάστρα. Το πείραμα ακλούθησε το Εντελώς Τυχαιοποιημένο σχέδιο.

A		
B	50	200
15	14,3	12,6
15	14,5	11,2
15	11,5	11
15	13,6	12,1
20	18,1	10,5
20	17,6	12,8
20	17,1	8,3
20	17,6	9,1
25	17,6	15,7
25	18,2	17,5
25	18,9	16,7
25	18,2	16,6

Εισάγουμε τα δεδομένα σε ένα φύλλο Excel, τοποθετώντας σε στήλες όπου η πρώτη στήλη περιλαμβάνει την επέμβαση X1, η δεύτερη την επέμβαση X2 και η τρίτη τα δεδομένα (Y). Στη συνέχεια αποθηκεύουμε τα δεδομένα σε αρχείο κειμένου (dets.txt) και τα εισάγουμε στην κονσόλα R δίνοντας με τη σειρά τις εντολές `dets=read.table("G:\\R\\Data\\dets.txt", header=TRUE, dec=",")`, `attach(dets)` και `dets`.

```
R Console
> dets=read.table("G:\\R\\Data\\dets.txt",header=TRUE,dec=",")
> attach(dets)
> dets
  X1  X2  Y
1  15  50 14.3
2  15  50 14.5
3  15  50 11.5
4  15  50 13.6
5  20  50 18.1
6  20  50 17.6
7  20  50 17.1
8  20  50 17.6
9  25  50 17.6
10 25  50 18.2
11 25  50 18.9
12 25  50 18.2
13 15 200 12.6
14 15 200 11.2
15 15 200 11.0
16 15 200 12.1
17 20 200 10.5
18 20 200 12.8
19 20 200  8.3
20 20 200  9.1
21 25 200 15.7
22 25 200 17.5
23 25 200 16.7
24 25 200 16.6
```

Χαρακτηρίζουμε ως παράγοντες (factor) τις επεμβάσεις X1 και X2 $X1=$ factor(X1); $X2=$ factor(X2), εισάγουμε το μοντέλο (δεδομένα Y: επέμβαση X1, επέμβαση X2 και αλληλεπίδραση $X1*X2$) $fit=$ aov($Y\sim X1+X2+X1*X2$) και με την εντολή $summary(fit)$ εμφανίζεται ο πίνακας ανάλυσης διασποράς.

```
R Console
> X1=factor(X1);X2=factor(X2)
> fit=aov(Y~X1+X2+X1*X2)
> summary(fit)
          Df Sum Sq Mean Sq F value    Pr(>F)
X1          2   99.87    49.94   40.86 2.03e-07 ***
X2          1   77.40    77.40   63.34 2.64e-07 ***
X1:X2       2   44.11    22.05   18.05 5.00e-05 ***
Residuals  18   22.00     1.22
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Πηγή Παραλλακτικότητας	BE	AT	MT	F	F πίνακα	Prob>F
X1	2	99,87	49,94	40,86	3,55	2,03e-07***
X2	1	77,40	77,40	63,34	4,41	2,64e-07***
X1*X2	2	44,11	22,05	18,05	3,55	5,00e-05***
Υπόλοιπο	18	22,00	1,22			
Σύνολο	23	243,38				

Από τον πίνακα ανάλυσης διασποράς παρατηρούμε ότι υπάρχει στατιστικά σημαντική επίδραση των παραγόντων X1 και X2 αλλά και στατιστικά σημαντική αλληλεπίδραση.

Με τις εντολές `library(agricolae)` και `LSD.test (fit, c ("X1", "X2"))` προχωράμε σε συγκρίσεις μέσω των έξι συνδυασμών των επιπέδων των δύο παραγόντων, με τη μέθοδο της Ελάχιστης Σημαντικής Διαφοράς.

```

R Console
> library(agricolae)
> LSD.test (fit, c ("X1", "X2"))

Study:

LSD t Test for Y

Mean Square Error: 1.222083

X1:X2, means and individual ( 95 %) CI

      Y  std.err r      LCL      UCL Min. Max.
15:200 11.725 0.3772157 4 10.932499 12.51750 11.0 12.6
15:50  13.475 0.6860211 4 12.033723 14.91628 11.5 14.5
20:200 10.175 0.9860485 4  8.103389 12.24661  8.3 12.8
20:50  17.600 0.2041241 4 17.171151 18.02885 17.1 18.1
25:200 16.625 0.3682730 4 15.851287 17.39871 15.7 17.5
25:50  18.225 0.2657536 4 17.666672 18.78333 17.6 18.9

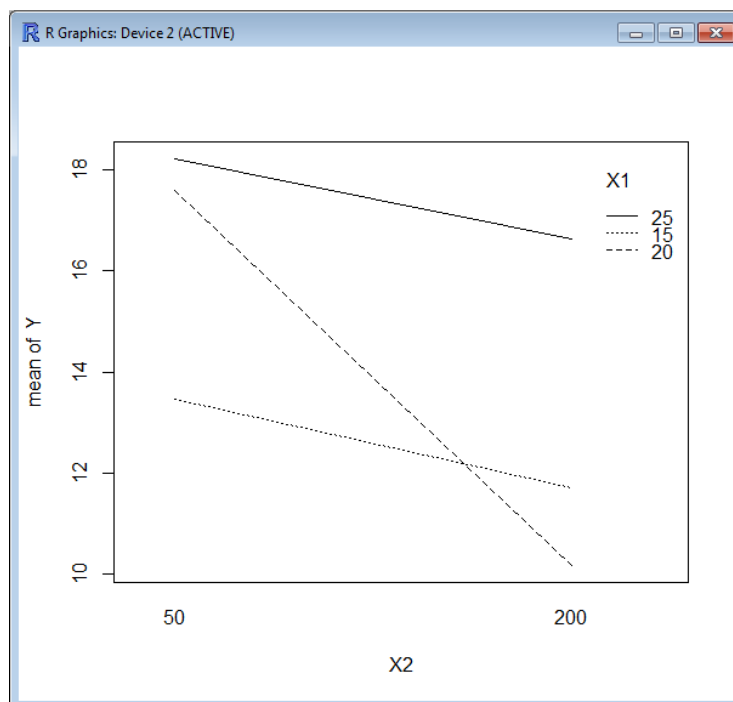
alpha: 0.05 ; Df Error: 18
Critical Value of t: 2.100922

Least Significant Difference 1.642273
Means with the same letter are not significantly different.

Groups, Treatments and means
a      25:50      18.23
a      20:50      17.6
a      25:200     16.62
b      15:50      13.48
c      15:200     11.72
c      20:200     10.18
> |

```

Η γραφική απεικόνιση δίνεται με την εντολή `interaction.plot(X2, X1, Y)`

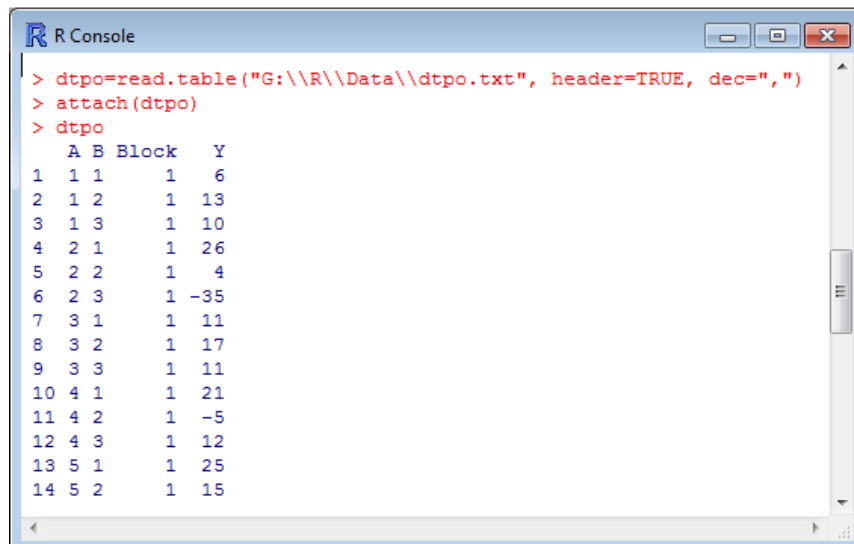


9.2. Τυχαιοποιημένες Πλήρεις Ομάδες

Ένας πειραματιστής ενδιαφέρεται για την επίδραση του είδους του ψεκαστικού υλικού και του τύπου μπεκ στην ποσότητα υγρού που ρέει από τα μπεκ αυτά. Διάλεξε τρία ορισμένα μεγέθη μπεκ και πέντε ψεκαστικά υλικά. Κάθε συνδυασμός υλικού-μπεκ δοκιμάστηκε με τυχαία σειρά σε κάθε μία από τις τρεις μέρες που κράτησε το πείραμα. Οι μέρες αποτελούν τις ομάδες, τα υλικά τον παράγοντα A, με $a=5$ επίπεδα και τα μπεκ τον παράγοντα B με $b=3$ επίπεδα.

	A	a ₁	a ₁	a ₁	a ₂	a ₂	a ₂	a ₃	a ₃	a ₃	a ₄	a ₄	a ₄	a ₅	a ₅	a ₅
	B	b ₁	b ₂	b ₃	b ₁	b ₂	b ₃	b ₁	b ₂	b ₃	b ₁	b ₂	b ₃	b ₁	b ₂	b ₃
Ομάδες	1	6	13	10	26	4	-35	11	17	11	21	-5	12	25	15	-4
	2	6	6	10	12	4	0	4	10	-10	14	2	-2	18	8	10
	3	-15	13	-11	5	11	-14	4	17	-17	7	-5	-16	25	1	24

Τα δεδομένα τοποθετούνται σε στήλες όπου η πρώτη στήλη περιλαμβάνει την επέμβαση A, η δεύτερη την επέμβαση B, η τρίτη τις ομάδες Block και η τέταρτη τη μεταβλητή (Y) και αποθηκεύονται σε αρχείο κειμένου (dtpro.txt). Στην κονσόλα R δίνουμε τις εντολές `dtpro=read.table("G:\\R\\Data\\dtpro.txt", header=TRUE, dec=",")`, `attach(dtpro)` και `dtpro`.



```

> dtpro=read.table("G:\\R\\Data\\dtpro.txt", header=TRUE, dec=",")
> attach(dtpro)
> dtpro
  A B Block  Y
1  1 1     1   6
2  1 2     1  13
3  1 3     1  10
4  2 1     1  26
5  2 2     1   4
6  2 3     1 -35
7  3 1     1  11
8  3 2     1  17
9  3 3     1  11
10 4 1     1  21
11 4 2     1  -5
12 4 3     1  12
13 5 1     1  25
14 5 2     1  15

```

Χαρακτηρίζουμε ως παράγοντες (factor) τις επεμβάσεις A, B και ομάδες Block `A=factor(A);B=factor(B);Block=factor(Block)`, εισάγουμε το μοντέλο (δεδομένα Y: επέμβαση A, επέμβαση B, αλληλεπίδραση A*B και ομάδες Block) `fit=aov(Y~A+B+A*B+Block)` και με την εντολή `summary(fit)` εμφανίζεται ο πίνακας ανάλυσης διασποράς.

```

R Console
> A=factor(A);B=factor(B); Block=factor(Block)
> fit=aov(Y~A+B+A*B+Block)
> summary(fit)
          Df Sum Sq Mean Sq F value Pr(>F)
A           4  798.8   199.7    2.064 0.11244
B           2 1427.0   713.5    7.374 0.00268 **
Block       2   328.8   164.4    1.699 0.20112
A:B         8 1821.5   227.7    2.353 0.04483 *
Residuals  28 2709.2    96.8

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>

```

Πηγή Παραλλακτικότητας	BE	AT	MT	F	F πίνακα	Prob>F
A	4	798,8	199,7	2,064	2,71	0,11244
B	2	1427,0	713,5	7,374	3,34	0,00268**
Ομάδα	2	328,8	164,4	1,699	3,34	0,20112
A*B	8	1821,5	227,7	2,353	2,29	0,04483*
Υπόλοιπο	28	2709,2	96,8			
Σύνολο	44	7085,2				

Από τον πίνακα ανάλυσης διασποράς παρατηρούμε ότι υπάρχει στατιστικά σημαντική επίδραση του παράγοντα B αλλά και στατιστικά σημαντική αλληλεπίδραση. Με τις εντολές `library(agricolae)` και `LSD.test(fit, c("A", "B"))` προχωράμε σε συγκρίσεις μέσω των δέκα πέντε συνδυασμών των επιπέδων των δύο παραγόντων, με τη μέθοδο της Ελάχιστης Σημαντικής Διαφοράς.

```

R Console
> library(agricolae)
> LSD.test(fit, c("A", "B"))

Study:

LSD t Test for Y

Mean Square Error: 96.75556

A:B, means and individual ( 95 %) CI

      Y  std.err r      LCL      UCL Min. Max.
1:1 -1.000000  7.000000 3 -15.3388500 13.338850 -15 6
1:2 10.666667  2.333333 3  5.8870500 15.446283  6 13
1:3  3.000000  7.000000 3 -11.3388500 17.338850 -11 10
2:1 14.333333  6.173420 3  1.6876563 26.979010  5 26
2:2  6.333333  2.333333 3  1.5537167 11.112950  4 11
2:3 -16.333333 10.170764 3 -37.1671994  4.500533 -35  0
3:1  6.333333  2.333333 3  1.5537167 11.112950  4 11
3:2 14.666667  2.333333 3  9.8870500 19.446283 10 17
3:3 -5.333333  8.412953 3 -22.5664863 11.899820 -17 11
4:1 14.000000  4.041452 3  5.7214611 22.278539  7 21
4:2 -2.666667  2.333333 3 -7.4462833  2.112950 -5  2
4:3 -2.000000  8.082904 3 -18.5570778 14.557078 -16 12
5:1 22.666667  2.333333 3 17.8870500 27.446283 18 25
5:2  8.000000  4.041452 3 -0.2785389 16.278539  1 15
5:3 10.000000  8.082904 3 -6.5570778 26.557078 -4 24

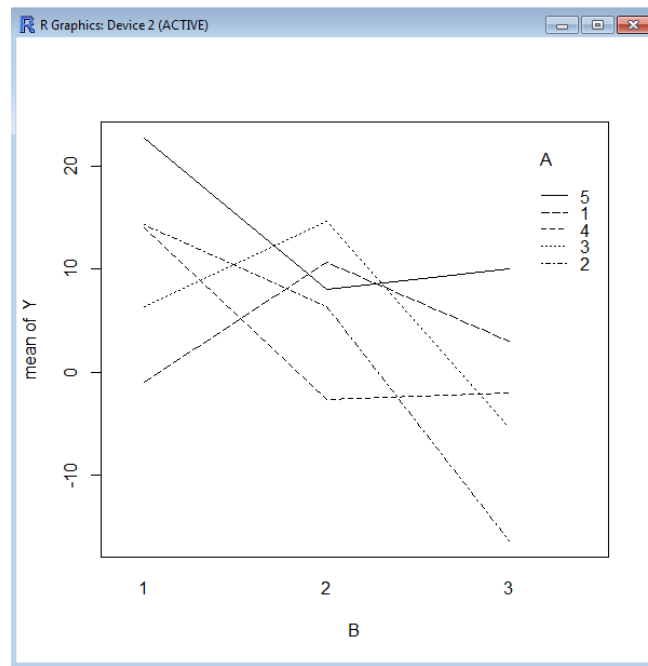
alpha: 0.05 ; Df Error: 28
Critical Value of t: 2.048407

Least Significant Difference 16.45162
Means with the same letter are not significantly different.

Groups, Treatments and means
a      5:1      22.67
ab     3:2      14.67
abc    2:1      14.33
abc    4:1       14
abcd   1:2      10.67
abcd   5:3       10
abcd   5:2       8
abcd   2:2      6.333
abcd   3:1      6.333
bcd    1:3       3
bcde   1:1      -1
cde    4:3      -2
de     4:2     -2.667
de     3:3     -5.333
e      2:3     -16.33
>

```

Η γραφική απεικόνιση δίνεται με την εντολή `interaction.plot(B, A, Y)`

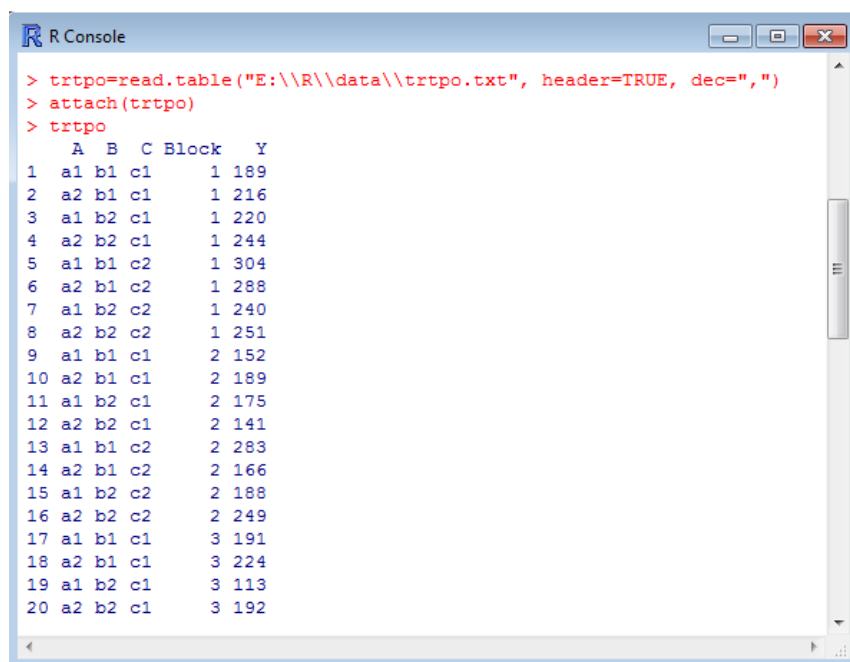


10. Παραγοντικό με 3 παράγοντες

Ένας ειδικός επί των γεωργικών μηχανημάτων μελέτησε τη δύναμη που χρειάζεται για όργωμα σε υγρό χωράφι όταν ο ελκυστήρας έχει μια σταθερή ταχύτητα. Τα δεδομένα αφορούν τη θέση του δεξιού τροχού (A1 ίσια και A2 υπό γωνία), το μέγεθος του ελαστικού (B1 6,5x16 και B2 17,5x16) και το ύψος έλξης (C1 5 εκ. και C2 10 εκ.). Το πείραμα ακολούθησε το σχέδιο των Τυχαιοποιημένων Πλήρων Ομάδων με 8 επαναλήψεις.

Παράγοντες			Ομάδες							
A	B	C	1	2	3	4	5	6	7	8
a ₁	b ₁	c ₁	189	152	191	196	189	167	110	139
a ₂	b ₁	c ₁	216	189	224	218	140	104	187	230
a ₁	b ₂	c ₁	220	175	113	223	192	145	201	180
a ₂	b ₂	c ₁	244	141	192	171	168	186	159	121
a ₁	b ₁	c ₂	304	283	184	157	101	149	136	198
a ₂	b ₁	c ₂	288	166	266	255	218	277	268	177
a ₁	b ₂	c ₂	240	188	205	140	171	223	219	174
a ₂	b ₂	c ₂	251	249	210	224	168	225	175	191

Η πρώτη στήλη στο αρχείο Excel περιλαμβάνει την επέμβαση A, η δεύτερη την επέμβαση B, η τρίτη την επέμβαση C, η τέταρτη τις επαναλήψεις και η πέμπτη τα δεδομένα (Y). Στη συνέχεια αποθηκεύουμε τα δεδομένα σε αρχείο κειμένου (trtpro.txt) και τα εισάγουμε στην κονσόλα R δίνοντας με τη σειρά τις εντολές `trtpro=read.table("E:\\R\\data\\trtpro.txt", header=TRUE, dec=",")`, `attach(trtpro)` και `trtpro`.



```
> trtpro=read.table("E:\\R\\data\\trtpro.txt", header=TRUE, dec=",")
> attach(trtpro)
> trtpro
  A B C Block  Y
1 a1 b1 c1    1 189
2 a2 b1 c1    1 216
3 a1 b2 c1    1 220
4 a2 b2 c1    1 244
5 a1 b1 c2    1 304
6 a2 b1 c2    1 288
7 a1 b2 c2    1 240
8 a2 b2 c2    1 251
9 a1 b1 c1    2 152
10 a2 b1 c1   2 189
11 a1 b2 c1   2 175
12 a2 b2 c1   2 141
13 a1 b1 c2   2 283
14 a2 b1 c2   2 166
15 a1 b2 c2   2 188
16 a2 b2 c2   2 249
17 a1 b1 c1   3 191
18 a2 b1 c1   3 224
19 a1 b2 c1   3 113
20 a2 b2 c1   3 192
```

Χαρακτηρίζουμε ως παράγοντες (factor) τις επεμβάσεις A, B, C και τις ομάδες Block `A=factor(A); B=factor(B); C=factor(C); Block=factor(Block)`, εισάγουμε το μοντέλο (δεδομένα Y: ομάδες, επέμβασεις A, B, C και αλληλεπιδράσεις A*B, B*C, A*C και A*B*C) `fit=aov(Y~Block+A+B+C+A*B+A*C+B*C+A*B*C)` και με την εντολή `summary(fit)` εμφανίζεται ο πίνακας ανάλυσης διασποράς.

```
R Console
> A=factor(A); B=factor(B); C=factor(C); Block=factor(Block)
> fit=aov(Y~Block+A+B+C+A*B+A*C+B*C+A*B*C)
> summary(fit)
      Df Sum Sq Mean Sq F value Pr(>F)
Block  7 29882    4269   2.778 0.01630 *
A       1  6480    6480   4.217 0.04539 *
B       1   529     529   0.344 0.56010
C       1 15876   15876  10.330 0.00232 **
A:B     1  4096    4096   2.665 0.10898
A:C     1  2862    2862   1.862 0.17858
B:C     1   420     420   0.273 0.60339
A:B:C   1    12      12   0.008 0.92922
Residuals 49 75306    1537
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

Πηγή Παραλλακτικότητας	BE	AT	MT	F	F πίν.	Prob>F
Ομάδα (R)	7	29882	4269	2,778	2,20	0,01630 *
A	1	6480	6480	4,217	4,04	0,04539 *
B	1	529	529	0,344	4,04	0,56010
C	1	15876	15876	10,330	4,04	0,00232 **
A*B	1	4096	4096	2,665	4,04	0,10898
A*C	1	2862	2862	1,862	4,04	0,17858
B*C	1	420	420	0,273	4,04	0,60339
A*B*C	1	12	12	0,008	4,04	0,92922
Υπόλοιπο	49	75306	1537			
Σύνολο	63	135463				

Από τον πίνακα ανάλυσης διασποράς παρατηρούμε ότι υπάρχει στατιστικά σημαντική επίδραση του παράγοντα A και C. Με τις εντολές `library(agricolae)`, `LSD.test (fit, "A")` και `LSD.test (fit, "C")` προχωράμε σε συγκρίσεις μέσω των δύο επιπέδων των παραγόντων A και C, με τη μέθοδο της Ελάχιστης Σημαντικής Διαφοράς

```
R Console
> library(agricolae)
> LSD.test (fit, "A")

Study:

LSD t Test for Y

Mean Square Error: 1536.857

A, means and individual ( 95 %) CI

      Y std.err r      LCL      UCL Min. Max.
a1 182.9375 7.965264 32 166.9307 198.9443 101 304
a2 203.0625 8.159565 32 186.6652 219.4598 104 288

alpha: 0.05 ; Df Error: 49
Critical Value of t: 2.009575

Least Significant Difference 19.69523
Means with the same letter are not significantly different.

Groups, Treatments and means
a      a2      203.1
b      a1      182.9
> LSD.test (fit, "C")

Study:

LSD t Test for Y

Mean Square Error: 1536.857

C, means and individual ( 95 %) CI

      Y std.err r      LCL      UCL Min. Max.
c1 177.25 6.524612 32 164.1383 190.3617 104 244
c2 208.75 8.830734 32 191.0040 226.4960 101 304

alpha: 0.05 ; Df Error: 49
Critical Value of t: 2.009575

Least Significant Difference 19.69523
Means with the same letter are not significantly different.

Groups, Treatments and means
a      c2      208.8
b      c1      177.2
> |
```

11. Υποδιαιρεμένα Τεμάχια (Split plot)

Τα δεδομένα προέρχονται από πείραμα έλεγχου μια ασθένειας στο σπόρο του κριθαριού. Οι επεμβάσεις ήταν δύο υδραργυρούχα παρασκευάσματα και μάρτυρας και έγιναν σε έξι κατηγορίες σπόρων (ανάλογα με την ευαισθησία τους στην ασθένεια). Η κατηγορία σπόρου θα αποτελέσει την κύρια μονάδα (A), η οποία υποδιαιρείται σε τρεις υπομονάδες (B). Η μια υπομονάδα είναι ο μάρτυρας D1 και οι άλλες δύο τα παρασκευάσματα D2 και D3. Το πείραμα σχεδιάστηκε για έξι επαναλήψεις.

		Επαναλήψεις (R)					
A	B	1	2	3	4	5	6
1	d1	58,6	64,8	63,2	58,5	60,1	63,4
1	d2	54,6	53,8	54,4	57,6	63,6	58,5
1	d3	50,6	52,1	55,2	51,2	57,2	58,8
2	d1	54,2	52,8	54	58,6	59,2	59,4
2	d2	58,9	58,8	58,2	58,9	62,6	56,4
2	d3	51,2	52,2	51,7	60,3	57,8	56,3
3	d1	54,2	51,2	54	55,4	54	53,8
3	d2	57	60,8	55,4	58,8	58	56
3	d3	49,7	51,6	52,9	57,7	59,8	52,2
4	d1	48	50,4	54	53,4	50,4	45,6
4	d2	50,4	60,4	58,2	57	63,6	60
4	d3	47,7	53,4	52,7	54,8	57,3	55,7
5	d1	37,2	38,2	49	47,8	46,2	48
5	d2	47	62,8	60,2	55	66,4	55,2
5	d3	47,1	55,8	58,4	55,6	62,4	54,9
6	d1	33,4	38,2	40,2	50	49	42
6	d2	55,4	52,8	51,4	56	60,2	53,8
6	d3	46,1	49,7	50,3	51,1	57,9	53,3

Η πρώτη στήλη στο αρχείο Excel περιλαμβάνει την επέμβαση A (κύρια τεμάχια), η δεύτερη την επέμβαση B (υποτεμάχια), η τρίτη τις ομάδες R και η τέταρτη τα δεδομένα (Y). Στη συνέχεια αποθηκεύουμε τα δεδομένα σε αρχείο κειμένου (.txt) και τα εισάγουμε στην κονσόλα R δίνοντας με τη σειρά τις εντολές `γρο=read.table("G:\\R\\data\\γρο.txt", header=TRUE, dec=",", attach(γρο)` και `γρο`.


```

R Console
> ypo=read.table("G:\\R\\data\\ypo.txt",header=TRUE,dec=",")
> attach(ypo)
> ypo
  R A B Y
1  1 1 d1 58.6
2  1 1 d2 54.6
3  1 1 d3 50.6
4  1 2 d1 54.2
5  1 2 d2 58.9
6  1 2 d3 51.2
7  1 3 d1 54.2
8  1 3 d2 57.0
9  1 3 d3 49.7
10 1 4 d1 48.0

```

Ορίζουμε ως παράγοντες (factor) τις επεμβάσεις A, B και ομάδες R $A=$ factor(A); $B=$ factor(B); $R=$ factor(R) εισάγουμε το μοντέλο (δεδομένα Y: επέμβαση A, επέμβαση B, αλληλεπίδραση A*B και υπόλοιπο (R/A) $fit=$ aov(Y~A+B+A*B+Error(R/A)) και με την εντολή `summary(fit)` εμφανίζεται ο πίνακας ανάλυσης διασποράς.

```

R Console
> A=factor(A);B=factor(B);R=factor(R)
> fit=aov(Y~A+B+A*B+Error(R/A))
> summary(fit)

Error: R
      Df Sum Sq Mean Sq F value Pr(>F)
Residuals  5  623.7   124.8

Error: R:A
      Df Sum Sq Mean Sq F value    Pr(>F)
A         5  781.1  156.22   12.55 3.71e-06 ***
Residuals 25  311.3   12.45
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: Within
      Df Sum Sq Mean Sq F value    Pr(>F)
B         2  663.3   331.7   39.24 1.27e-11 ***
A:B       10  942.6    94.3   11.15 1.84e-10 ***
Residuals 60  507.1     8.5
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |

```

Πηγή Παραλλακτικότητας	ΒΕ	ΑΤ	ΜΤ	F	F πίνακα	Prob>F
R	5	623,7	124,8			
A	5	781,1	156,22	12,55	2,60	3,71e-06***
A*R (Υπόλοιπο a)	25	311,3	12,45			
B	2	663,3	331,7	39,24	3,15	1,27e-11***
A*B	10	942,6	94,3	11,15	1,99	1,84e-10***
Υπόλοιπο (b)	60	507,1	8,5			
Σύνολο	107	3829,1				

12. Ιεραρχική Ανάλυση συστάδων (Hierarchical Cluster Analysis)

Τα δεδομένα προέρχονται από πείραμα ταυτοποίησης ποικιλιών σίτου μετά από διαχωρισμό των υποκλασμάτων της γλιαδίνης, με ηλεκροφόρηση.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
KYPERO	0	0	0	1	0	1	1	1	1	0	1	0	1	0	1	0	0	1	0	0
ARONAS	0	0	0	1	0	0	1	0	0	1	0	1	1	0	1	0	1	0	0	0
MESAORI	0	0	1	1	0	0	1	0	1	0	0	1	1	0	1	0	1	0	1	0
KARPASI	1	1	0	1	0	0	1	0	1	0	0	1	0	1	1	0	1	0	1	0
LAKOTA	0	0	0	1	1	0	1	0	1	0	1	0	1	0	1	0	1	1	1	0
DURTAL	0	1	0	1	1	1	1	0	1	0	0	0	1	0	1	1	1	0	1	1
WELLS	0	0	0	1	1	1	1	0	1	0	0	0	1	1	1	1	1	0	1	1
MONDUR	1	0	0	1	0	1	1	0	0	0	1	0	1	0	1	1	1	0	1	0
LIMNOS	1	0	1	1	1	0	1	1	0	1	0	1	1	1	1	0	1	0	1	1
METHONI	0	0	0	1	0	0	1	1	0	1	0	1	1	0	1	1	1	0	1	1
SKITI	0	0	0	1	0	0	1	1	0	0	0	0	1	0	1	0	1	0	1	0
SIPHNOS	0	1	0	1	0	0	1	0	0	0	0	0	1	0	1	0	1	0	0	0

Τα δεδομένα τοποθετούνται σε στήλες όπου η πρώτη στήλη περιλαμβάνει τις ονομασίες των ποικιλιών και οι υπόλοιπες στήλες την παρουσία (1) ή την απουσία (0) γλιαδίνης στις σχετικές ζώνες κινητικότητας. Στη συνέχεια αποθηκεύουμε τα δεδομένα σε αρχείο κειμένου (10.txt) και τα εισάγουμε στην κονσόλα R δίνοντας με τη σειρά τις εντολές `cluster=read.table("G:\\R\\data\\10.txt", header=TRUE)`, `attach(cluster)` και `cluster`.

```
R Console
> cluster=read.table("G:\\R\\data\\10.txt",header=TRUE)
> attach(cluster)
> cluster
      X1 X2 X3 X4 X5 X6 X7 X8 X9 X10 X11 X12 X13 X14 X15 X16 X17 X18 X19 X20
KYPERO 0 0 0 1 0 1 1 1 1 0 1 0 1 0 1 0 0 1 0 0
ARONAS 0 0 0 1 0 0 1 0 0 1 0 1 0 1 0 1 0 1 0 0
MESAORI 0 0 1 1 0 0 1 0 1 0 0 1 1 0 1 0 1 0 1 0
KARPASI 1 1 0 1 0 0 1 0 1 0 0 1 0 1 1 0 1 0 1 0
LAKOTA 0 0 0 1 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 1
DURTAL 0 1 0 1 1 1 1 0 1 0 0 0 1 0 1 1 1 1 0 1
WELLS 0 0 0 1 1 1 1 0 1 0 0 0 1 1 1 1 1 1 0 1
MONDUR 1 0 0 1 0 1 1 0 0 0 1 0 1 0 1 0 1 1 0 1
LIMNOS 1 0 1 1 1 0 1 1 0 1 0 1 1 1 1 0 1 0 1 1
METHONI 0 0 0 1 0 0 1 1 0 1 0 1 1 0 1 1 1 0 1 1
SKITI 0 0 0 1 0 0 1 1 0 0 0 0 1 0 1 0 1 0 1 0
SIPHNOS 0 1 0 1 0 0 1 0 0 0 0 0 1 0 1 0 1 0 0 0
> |
```

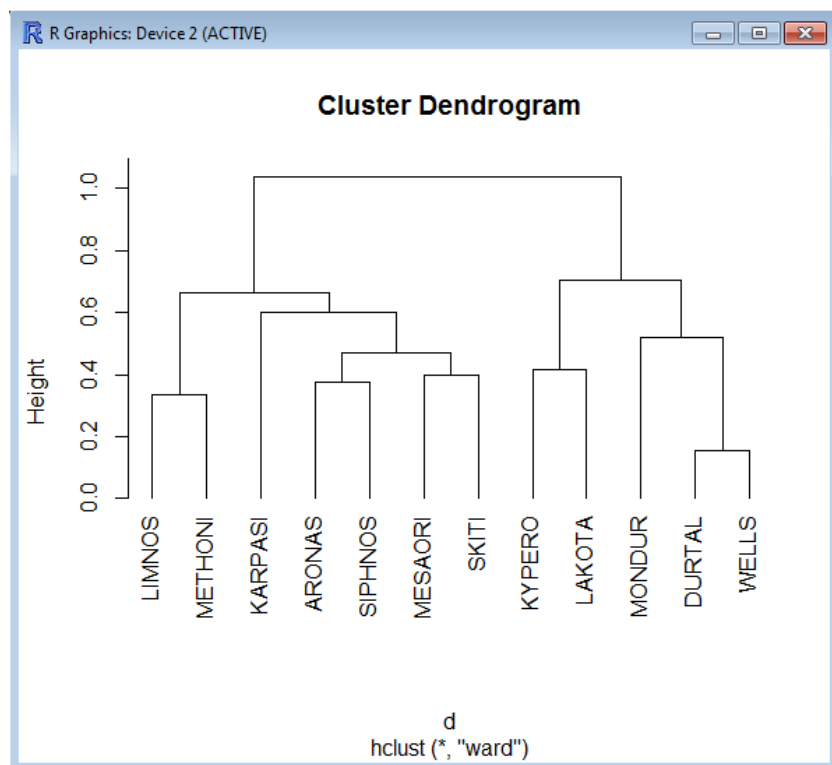
Με την εντολή `d=dist(cluster, method = "binary")` δημιουργούμε τις αποστάσεις ή τους συντελεστές ομοιότητας ή ανομοιότητας, επιλέγοντας μία από τις μεθόδους euclidean, maximum, manhattan, canberra, binary ή minkowski και με την εντολή `d` εμφανίζονται οι αντίστοιχες τιμές.

```

R Console
> d =dist(cluster, method = "binary")
> d
      KYPERO  ARONAS  MESAORI  KARPASI  LAKOTA  DURTAL  WELLS  MONDUR  LIMNOS  METHONI  SKITI
ARONAS  0.6666667
MESAORI 0.6153846 0.4000000
KARPASI 0.7333333 0.5833333 0.4166667
LAKOTA  0.4166667 0.5833333 0.4166667 0.5714286
DURTAL  0.6000000 0.6428571 0.5000000 0.5333333 0.4285714
WELLS   0.6000000 0.6428571 0.5000000 0.5333333 0.4285714 0.1538462
MONDUR  0.5384615 0.5833333 0.5384615 0.5714286 0.4615385 0.4285714 0.4285714
LIMNOS  0.7222222 0.5000000 0.4666667 0.5000000 0.5882353 0.5555556 0.4705882 0.5882353
METHONI 0.6666667 0.3636364 0.4615385 0.6000000 0.6000000 0.4666667 0.4666667 0.5000000 0.3333333
SKITI   0.5454545 0.4444444 0.4000000 0.5833333 0.4545455 0.5384615 0.5384615 0.4545455 0.5000000 0.3636364
SIPHROS 0.6363636 0.3750000 0.5000000 0.5454545 0.5454545 0.5000000 0.6153846 0.5454545 0.6666667 0.5833333 0.3750000

```

Ο υπολογισμός του δενδρογράμματος γίνεται την εντολή `fit = hclust(d, method="ward")` επιλέγοντας μία από τις μεθόδους `ward`, `single`, `complete`, `average`, `mcquitty`, `median` ή `centroid` και η εμφάνιση του δενδρογράμματος γίνεται με την εντολή `plot(fit, hang = -1)`.



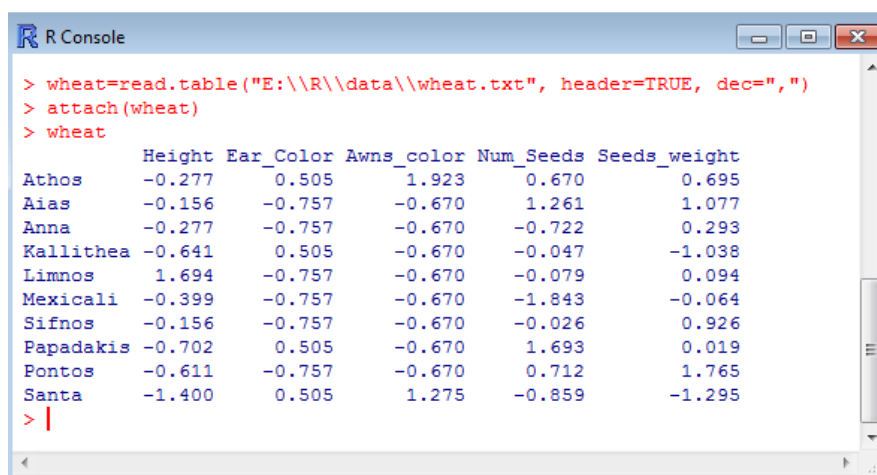
13. Ανάλυση Κυρίων Συνιστωσών (Principal Components Analysis)

Τα δεδομένα προέρχονται από πείραμα αξιολόγησης ποικιλιών σιταριού στα οποία μελετήθηκαν μια σειρά ποσοτικών και ποιοτικών χαρακτηριστικών. Τα δεδομένα παρουσιάζονται στον παρακάτω πίνακα μετά από κανονικοποίηση των τιμών.

	Height	Ear_Color	Awns_color	Num_Seeds	Seeds_weight
Athos	-0,277	0,505	1,923	0,670	0,695
Aias	-0,156	-0,757	-0,670	1,261	1,077
Anna	-0,277	-0,757	-0,670	-0,722	0,293
Kallithea	-0,641	0,505	-0,670	-0,047	-1,038
Limnos	1,694	-0,757	-0,670	-0,079	0,094
Mexicali	-0,399	-0,757	-0,670	-1,843	-0,064
Sifnos	-0,156	-0,757	-0,670	-0,026	0,926
Papadakis	-0,702	0,505	-0,670	1,693	0,019
Pontos	-0,611	-0,757	-0,670	0,712	1,765
Santa	-1,400	0,505	1,275	-0,859	-1,295

Τοποθετούνται τα δεδομένα σε στήλες, όπου η πρώτη στήλη περιλαμβάνει τις ονομασίες των ποικιλιών και οι υπόλοιπες στήλες τις μεταβλητές. Στη συνέχεια αποθηκεύουμε τα δεδομένα σε αρχείο κειμένου (wheat.txt) και τα εισάγουμε στην κονσόλα R δίνοντας με τη σειρά τις εντολές

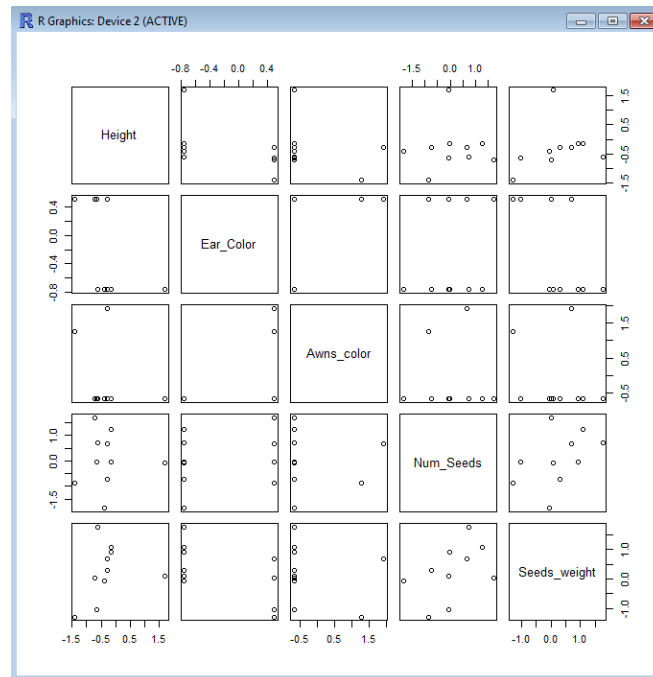
```
wheat=read.table("E:\\R\\data\\wheat.txt", header=TRUE, dec=","), attach(wheat)  
και wheat.
```



```
R Console  
> wheat=read.table("E:\\R\\data\\wheat.txt", header=TRUE, dec=",")  
> attach(wheat)  
> wheat  
      Height Ear_Color Awns_color Num_Seeds Seeds_weight  
Athos  -0.277    0.505     1.923     0.670     0.695  
Aias   -0.156   -0.757    -0.670     1.261     1.077  
Anna   -0.277   -0.757    -0.670    -0.722     0.293  
Kallithea -0.641    0.505    -0.670    -0.047    -1.038  
Limnos  1.694   -0.757    -0.670    -0.079     0.094  
Mexicali -0.399   -0.757    -0.670    -1.843    -0.064  
Sifnos  -0.156   -0.757    -0.670    -0.026     0.926  
Papadakis -0.702    0.505    -0.670     1.693     0.019  
Pontos  -0.611   -0.757    -0.670     0.712     1.765  
Santa  -1.400    0.505     1.275    -0.859    -1.295  
> |
```

Αν δεν είχε προηγηθεί η κανονικοποίηση των τιμών στα δεδομένα μας, αυτή θα γινόταν με την εντολή `scale(wheat)`.

Με την εντολή `pairs(wheat)` δημιουργούνται γραφικές παραστάσεις ανά ζεύγη μεταβλητών, οι οποίες μας βοηθούν στον εντοπισμό τυχόν συσχετίσεων.



Με την εντολή `pca=princomp((wheat), cor=T)` υπολογίζονται οι κύριες συνιστώσες, επιλέγοντας με συσχέτιση (`cor=TRUE`) ή με συνδιασπορά (`cor=FALSE`) και με την εντολή `summary(pca, loadings=T)` εμφανίζονται τα αποτελέσματα.

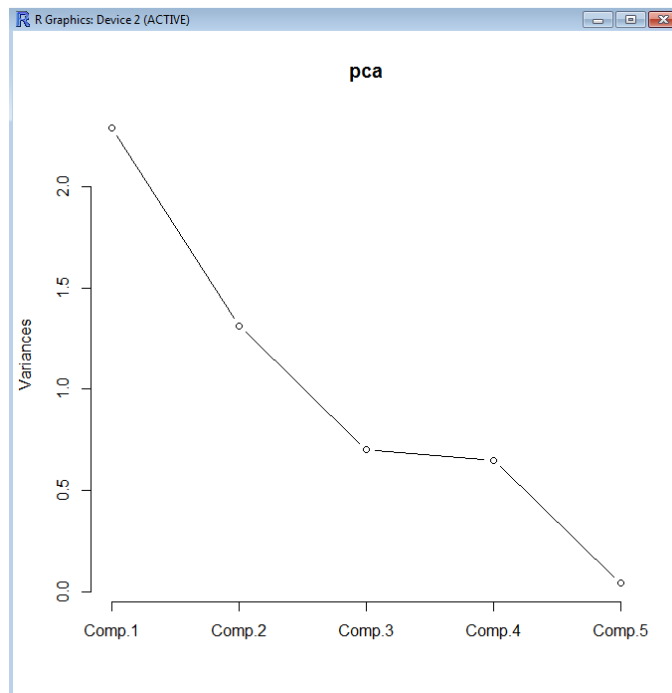
```

R Console
> pca=princomp((wheat), cor=T)
> summary(pca, loadings=T)
Importance of components:
              Comp.1   Comp.2   Comp.3   Comp.4   Comp.5
Standard deviation  1.514049 1.1461775 0.8368640 0.8067453 0.206767841
Proportion of Variance 0.458469 0.2627446 0.1400683 0.1301676 0.008550588
Cumulative Proportion 0.458469 0.7212135 0.8612818 0.9914494 1.000000000

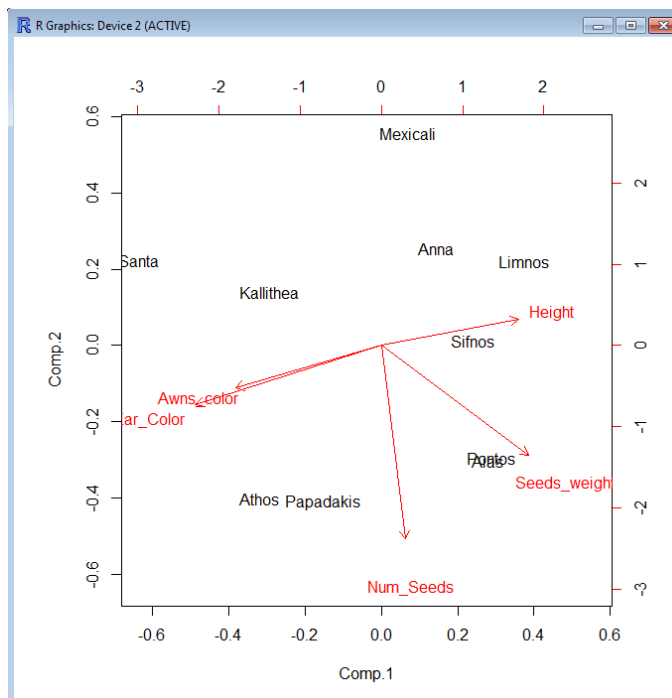
Loadings:
              Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
Height         0.440  0.111  0.845 -0.243  0.145
Ear_Color     -0.597 -0.252  0.136 -0.321  0.677
Awns_color    -0.468 -0.181  0.501  0.650 -0.272
Num_Seeds          -0.822          -0.374 -0.422
Seeds_weight    0.474 -0.465 -0.124  0.525  0.518
>

```

Με την εντολή `screplot(pca, type="lines")` εμφανίζεται το διάγραμμα ιδιοτιμών των χαρακτηριστικών ριζών (Eigenvalues).

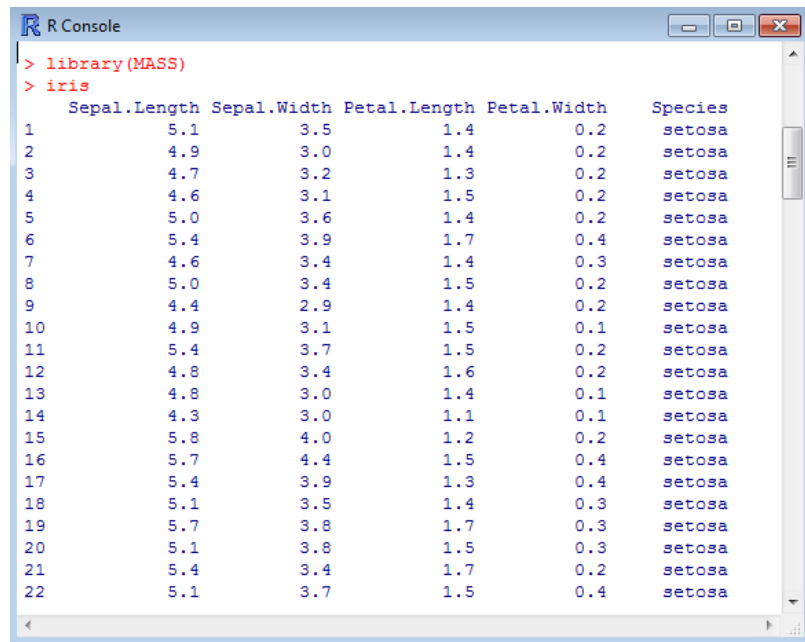


Με την εντολή `biplot(pca)` εμφανίζεται το διάγραμμα διασποράς μεταβλητών και των ποικιλιών.



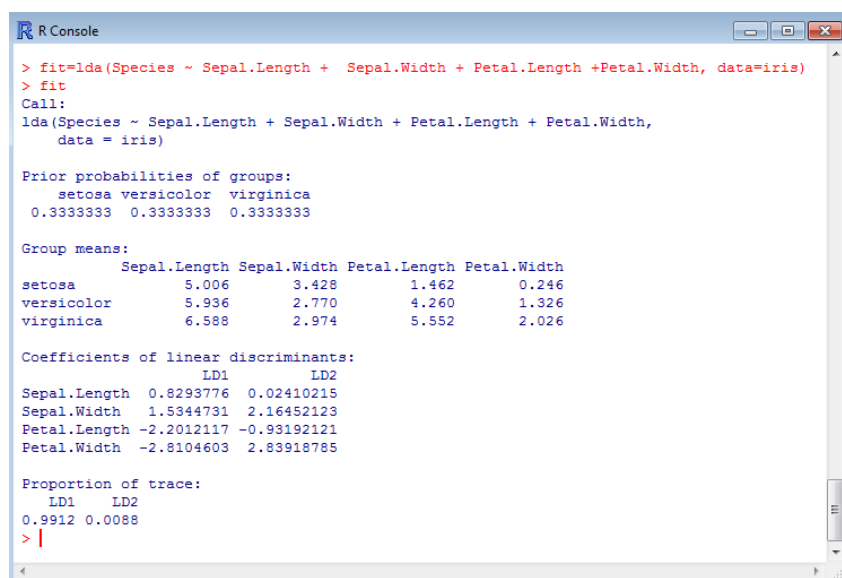
14. Διακριτική Ανάλυση (Discriminant Analysis)

Για την πραγματοποίηση της Διακριτικής Ανάλυσης είναι απαραίτητη η λήψη του πακέτου Mass (<http://cran.r-project.org/web/packages/MASS/index.html>) και η εγκατάστασή του. Τα δεδομένα προέρχονται από το κλασικό παράδειγμα του Fischer με τα τρία είδη του γένους Iris, όπου μετρήθηκαν το μήκος και πλάτος των σεπάλων και πετάλων. Τα δεδομένα υπάρχουν στο βασικό πακέτο και με την εντολή `iris` εμφανίζονται ως εξής.



```
R Console
> library(MASS)
> iris
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1           5.1           3.5           1.4           0.2   setosa
2           4.9           3.0           1.4           0.2   setosa
3           4.7           3.2           1.3           0.2   setosa
4           4.6           3.1           1.5           0.2   setosa
5           5.0           3.6           1.4           0.2   setosa
6           5.4           3.9           1.7           0.4   setosa
7           4.6           3.4           1.4           0.3   setosa
8           5.0           3.4           1.5           0.2   setosa
9           4.4           2.9           1.4           0.2   setosa
10          4.9           3.1           1.5           0.1   setosa
11          5.4           3.7           1.5           0.2   setosa
12          4.8           3.4           1.6           0.2   setosa
13          4.8           3.0           1.4           0.1   setosa
14          4.3           3.0           1.1           0.1   setosa
15          5.8           4.0           1.2           0.2   setosa
16          5.7           4.4           1.5           0.4   setosa
17          5.4           3.9           1.3           0.4   setosa
18          5.1           3.5           1.4           0.3   setosa
19          5.7           3.8           1.7           0.3   setosa
20          5.1           3.8           1.5           0.3   setosa
21          5.4           3.4           1.7           0.2   setosa
22          5.1           3.7           1.5           0.4   setosa
```

Με την εντολή `fit=lda(Species ~ Sepal.Length + Sepal.Width + Petal.Length + Petal.Width, data=iris)` δίνεται το μοντέλο για την γραμμική διακριτική ανάλυση (`lda`) και με την εντολή `fit` εμφανίζονται τα αποτελέσματα της ανάλυσης.



```
R Console
> fit=lda(Species ~ Sepal.Length + Sepal.Width + Petal.Length +Petal.Width, data=iris)
> fit
Call:
lda(Species ~ Sepal.Length + Sepal.Width + Petal.Length + Petal.Width,
    data = iris)

Prior probabilities of groups:
  setosa versicolor virginica
0.3333333 0.3333333 0.3333333

Group means:
      Sepal.Length Sepal.Width Petal.Length Petal.Width
setosa           5.006           3.428           1.462           0.246
versicolor       5.936           2.770           4.260           1.326
virginica        6.588           2.974           5.552           2.026

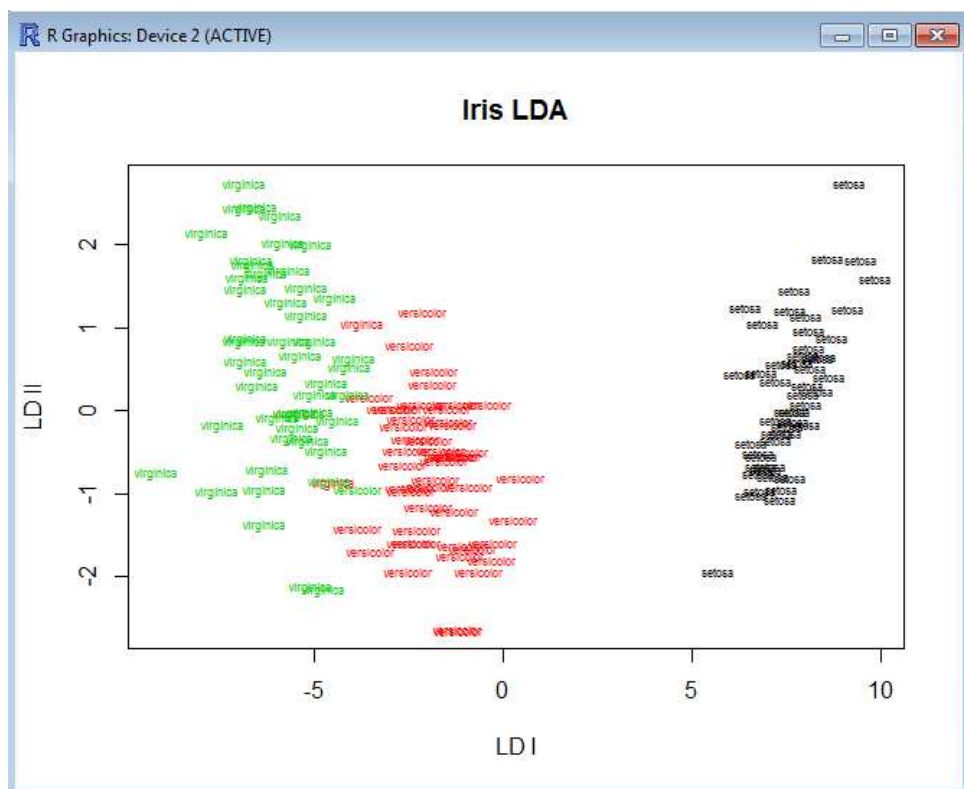
Coefficients of linear discriminants:
      LD1      LD2
Sepal.Length 0.8293776 0.02410215
Sepal.Width  1.5344731 2.16452123
Petal.Length -2.2012117 -0.93192121
Petal.Width  -2.8104603 2.83918785

Proportion of trace:
      LD1      LD2
0.9912 0.0088
> |
```

Με τις εντολές `predict(fit)$x`, `predict(fit)$posterior`, `predict(fit)$class` και `table(iris$Species, predict(fit)$class)` εμφανίζονται τελικά τα άτομα που δεν ανήκουν στον πληθυσμό όπου αρχικά είχαν καταχωρηθεί.

```
R Console
> predict(fit)$class
 [1] setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa
[11] setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa
[21] setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa
[31] setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa
[41] setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa
[51] versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor
[61] versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor
[71] virginica versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor
[81] versicolor versicolor versicolor virginica versicolor versicolor versicolor versicolor versicolor versicolor
[91] versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor
[101] virginica virginica virginica virginica virginica virginica virginica virginica virginica virginica
[111] virginica virginica virginica virginica virginica virginica virginica virginica virginica virginica
[121] virginica virginica virginica virginica virginica virginica virginica virginica virginica virginica
[131] virginica virginica virginica versicolor virginica virginica virginica virginica virginica virginica
[141] virginica virginica virginica virginica virginica virginica virginica virginica virginica virginica
Levels: setosa versicolor virginica
> table(iris$Species, predict(fit)$class)
      setosa versicolor virginica
setosa     50         0         0
versicolor  0         48         2
virginica  0          1         49
> |
```

Οι εντολές `plot(predict(fit)$x, type="n", xlab="LD I", ylab="LD II", main="Iris LDA")` και `text(predict(fit)$x, levels(predict(fit)$class)[predict(fit)$class], col=unclass(iris$Species) cex=0.5)` δίνουν το διάγραμμα διασποράς των ατόμων των τριών πληθυσμών.



Παραπομπές – Χρήσιμες Συνδέσεις

<http://www.r-project.org/>

<http://www.statmethods.net/>

<http://www.r-tutor.com/>

<http://www.inside-r.org/>

<http://www.r-bloggers.com/>

<http://manuals.bioinformatics.ucr.edu/>

Το R-project στα ελληνικά

<http://cran.r-project.org/doc/contrib/mainfokianoscharalambous.pdf>

http://stat-athens.aueb.gr/~grstats/notes/r_giagos.pdf

[http://users.sch.gr/epdiaman/images/stories/ergasies/biblia/statistics_with_cal
c_and_R_project.pdf](http://users.sch.gr/epdiaman/images/stories/ergasies/biblia/statistics_with_cal
c_and_R_project.pdf)