

Data File Used in this Analysis:

```
# M 3070-1          Scleroderma Data          May 17, 2011
# Treibergs
#
# Dataset from Elliott, "Learning SAS in the computer lab. 2nd ed.,"
# Brooks/Cole 2000
#
# 1989 study to determine safety and efficacy of a particular compound
# in the treatment of scleroderma, a skin disease characterized by
# thickening of the skin. Mobility scores skins ability to be stretched.
#
# Variable
#
# clinic  clinic number
# id      patient id
# treat   1= drug / 2 = placebo
# thick1  skin thickening at first visit: higher numbers represent worse thickening
# thick2  skin thickening at 2nd visit: higher numbers represent worse thickening
# mob1    skin mobility at 1st visit: higher numbers represent better mobility
# mob2    skin mobility at 2nd visit: higher numbers represent better mobility
# ass1    patient assessment at 1st visit: the higher the number the worse the patient
# ass2    patient assessment at 2nd visit: the higher the number the worse the patient
#
clinic id treat thick1 thick2 mob1 mob2 ass1 ass2
1 01 2 NA NA 298 418 8 5
1 02 1 NA NA 390 498 3 3
1 03 2 NA NA 543 658 2 2
1 04 1 NA NA 526 535 2 3
2 05 1 30 28 227 260 6 2
2 06 2 18 21 375 409 6 5
8 07 2 27 25 308 306 6 3
8 08 2 19 25 376 360 2 3
8 09 1 18 19 281 377 3 3
8 10 1 25 21 264 314 4 3
10 11 2 20 5 280 340 3 1
10 12 1 15 13 186 132 6 3
10 13 2 17 6 281 242 4 3
10 14 2 29 13 200 434 4 5
10 15 1 34 21 82 326 9 4
10 16 1 9 3 602 800 3 3
10 17 1 7 8 936 796 3 3
18 18 2 21 25 326 444 4 5
18 19 2 19 15 434 NA 2 3
18 20 2 29 24 241 198 3 5
18 21 2 17 17 194 166 5 5
40 22 2 25 10 403 248 6 4
40 23 2 15 16 587 485 6 6
40 24 1 16 13 518 503 NA NA
```

40	25	2	26	26	231	155	6	4
41	26	1	24	27	173	95	3	6
41	27	2	36	35	224	229	7	4
41	28	1	7	7	744	629	1	1
41	29	1	15	17	329	299	5	1
41	30	1	31	37	154	128	7	8
41	31	2	29	29	240	229	5	4
42	32	2	21	18	550	595	5	5
45	33	2	23	NA	424	388	5	8
45	34	1	38	NA	184	195	7	6
45	35	2	24	NA	472	537	2	1
45	36	1	26	NA	255	211	5	2
45	37	2	29	24	247	232	7	7
45	38	1	32	25	196	196	6	6
45	39	1	16	NA	200	367	4	4
46	40	2	9	9	347	309	5	2
46	41	2	14	13	329	260	5	5
46	42	1	28	23	175	148	5	9
46	43	1	7	7	440	444	4	3
46	44	1	17	18	191	192	7	5
46	45	2	20	22	269	242	4	5
46	46	2	5	5	461	480	3	3
46	47	2	12	15	313	373	9	5
46	48	2	9	11	359	395	5	3
46	49	2	7	11	355	378	4	4
46	50	1	17	23	296	326	4	6
47	51	1	28	28	191	NA	9	8
47	52	2	7	7	383	NA	5	2
48	53	2	18	7	444	563	6	4
48	54	2	13	28	468	460	5	6
48	55	1	20	14	481	490	6	6
48	56	1	22	25	402	306	6	6
48	57	2	23	30	363	336	6	6
48	58	2	25	20	299	444	3	3
48	59	1	26	30	211	310	9	7
48	60	2	26	34	193	156	8	9
48	61	2	24	29	514	417	2	5
48	62	2	10	12	438	440	3	3
49	63	2	18	18	378	373	3	2
49	64	2	23	23	286	163	5	5
49	65	2	9	11	489	485	2	3
49	66	2	7	7	675	636	3	1
49	67	1	11	9	343	378	3	3
49	68	2	8	8	442	440	1	1
49	69	1	6	6	531	521	2	1
49	70	1	5	5	502	574	1	1
49	71	2	36	36	41	56	8	8
49	72	1	28	27	92	173	6	6
49	73	1	9	9	521	614	4	5
50	74	2	22	9	390	382	2	2
50	75	1	19	9	432	398	5	1
50	76	2	17	9	510	448	3	1

R Session:

R version 2.11.1 (2010-05-31)
Copyright (C) 2010 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.34 (5589) i386-apple-darwin9.8.0]

```
> # We do the data processing following Elliott, "Learning  
> # SAS in the computer lab. 2nd ed.," Brooks/Cole 2000  
> # Example 2.1.  
> #  
> # Read the data file. Print to see what you got.  
> #  
> tt <- read.table("M3074ScleroData2.txt",header=TRUE)  
> tt
```

	clinic	id	treat	thick1	thick2	mob1	mob2	ass1	ass2
1	1	1	2	NA	NA	298	418	8	5
2	1	2	1	NA	NA	390	498	3	3
3	1	3	2	NA	NA	543	658	2	2
4	1	4	1	NA	NA	526	535	2	3
5	2	5	1	30	28	227	260	6	2
6	2	6	2	18	21	375	409	6	5
7	8	7	2	27	25	308	306	6	3
8	8	8	2	19	25	376	360	2	3
9	8	9	1	18	19	281	377	3	3
10	8	10	1	25	21	264	314	4	3
11	10	11	2	20	5	280	340	3	1
12	10	12	1	15	13	186	132	6	3
13	10	13	2	17	6	281	242	4	3
14	10	14	2	29	13	200	434	4	5
15	10	15	1	34	21	82	326	9	4
16	10	16	1	9	3	602	800	3	3
17	10	17	1	7	8	936	796	3	3
18	18	18	2	21	25	326	444	4	5
19	18	19	2	19	15	434	NA	2	3

20	18	20	2	29	24	241	198	3	5
21	18	21	2	17	17	194	166	5	5
22	40	22	2	25	10	403	248	6	4
23	40	23	2	15	16	587	485	6	6
24	40	24	1	16	13	518	503	NA	NA
25	40	25	2	26	26	231	155	6	4
26	41	26	1	24	27	173	95	3	6
27	41	27	2	36	35	224	229	7	4
28	41	28	1	7	7	744	629	1	1
29	41	29	1	15	17	329	299	5	1
30	41	30	1	31	37	154	128	7	8
31	41	31	2	29	29	240	229	5	4
32	42	32	2	21	18	550	595	5	5
33	45	33	2	23	NA	424	388	5	8
34	45	34	1	38	NA	184	195	7	6
35	45	35	2	24	NA	472	537	2	1
36	45	36	1	26	NA	255	211	5	2
37	45	37	2	29	24	247	232	7	7
38	45	38	1	32	25	196	196	6	6
39	45	39	1	16	NA	200	367	4	4
40	46	40	2	9	9	347	309	5	2
41	46	41	2	14	13	329	260	5	5
42	46	42	1	28	23	175	148	5	9
43	46	43	1	7	7	440	444	4	3
44	46	44	1	17	18	191	192	7	5
45	46	45	2	20	22	269	242	4	5
46	46	46	2	5	5	461	480	3	3
47	46	47	2	12	15	313	373	9	5
48	46	48	2	9	11	359	395	5	3
49	46	49	2	7	11	355	378	4	4
50	46	50	1	17	23	296	326	4	6
51	47	51	1	28	28	191	NA	9	8
52	47	52	2	7	7	383	NA	5	2
53	48	53	2	18	7	444	563	6	4
54	48	54	2	13	28	468	460	5	6
55	48	55	1	20	14	481	490	6	6
56	48	56	1	22	25	402	306	6	6
57	48	57	2	23	30	363	336	6	6
58	48	58	2	25	20	299	444	3	3
59	48	59	1	26	30	211	310	9	7
60	48	60	2	26	34	193	156	8	9
61	48	61	2	24	29	514	417	2	5
62	48	62	2	10	12	438	440	3	3
63	49	63	2	18	18	378	373	3	2
64	49	64	2	23	23	286	163	5	5
65	49	65	2	9	11	489	485	2	3
66	49	66	2	7	7	675	636	3	1
67	49	67	1	11	9	343	378	3	3
68	49	68	2	8	8	442	440	1	1
69	49	69	1	6	6	531	521	2	1
70	49	70	1	5	5	502	574	1	1
71	49	71	2	36	36	41	56	8	8

```

72  49 72  1  28  27  92 173  6  6
73  49 73  1   9   9 521 614  4  5
74  50 74  2  22   9 390 382  2  2
75  50 75  1  19   9 432 398  5  1
76  50 76  2  17   9 510 448  3  1
>
> # Note that missing data is recorded as "NA" = not available.
> # Make the inputted data set (collection of records) available to R.
> attach(tt)
>
> # To see the variable names:
> names(tt)
[1] "clinic" "id"      "treat"  "thick1" "thick2" "mob1"   "mob2"   "ass1"   "ass2"
>
> # Define a new vector of assessed improvement (higher numbers are worse.)
> improve <- ass1-ass2
>
>
> # Suppose we wish make a new data frame that has another column.
> tt4 <- data.frame(tt,improve)
> names(tt4)
[1] "clinic" "id"      "treat"  "thick1" "thick2" "mob1"   "mob2"   "ass1"
[9] "ass2"   "improve"
>
>
> # Extract from the data those records of individuals that had
> # positive improvement and clinic number from 45 to 50.
> tt5 <- subset(tt4, (improve > 0) & (45 <= clinic) & (clinic <= 50))
> tt5
  clinic id treat thick1 thick2 mob1 mob2 ass1 ass2 improve
34    45 34   1    38    NA  184  195   7   6         1
35    45 35   2    24    NA  472  537   2   1         1
36    45 36   1    26    NA  255  211   5   2         3
40    46 40   2     9     9  347  309   5   2         3
43    46 43   1     7     7  440  444   4   3         1
44    46 44   1    17    18  191  192   7   5         2
47    46 47   2    12    15  313  373   9   5         4
48    46 48   2     9    11  359  395   5   3         2
51    47 51   1    28    28  191   NA   9   8         1
52    47 52   2     7     7  383   NA   5   2         3
53    48 53   2    18     7  444  563   6   4         2
59    48 59   1    26    30  211  310   9   7         2
63    49 63   2    18    18  378  373   3   2         1
66    49 66   2     7     7  675  636   3   1         2
69    49 69   1     6     6  531  521   2   1         1
75    50 75   1    19     9  432  398   5   1         4
76    50 76   2    17     9  510  448   3   1         2
>
> # Alternatively, the elements in the frame can be
> # thought of as a matrix tt[row,col]
> # In R, it is possible to specify the columns desired
> # by putting in a logical condition for which column

```

```

> # to take. "comma blank", i.e., [something,]
> # means there is no condition on the row.
>
> tt6 <- tt5[order(tt5[,3]),]
> tt6
  clinic id treat thick1 thick2 mob1 mob2 ass1 ass2 improve
34     45 34     1     38     NA  184  195     7     6     1
36     45 36     1     26     NA  255  211     5     2     3
43     46 43     1     7      7  440  444     4     3     1
44     46 44     1    17     18  191  192     7     5     2
51     47 51     1    28     28  191   NA     9     8     1
59     48 59     1    26     30  211  310     9     7     2
69     49 69     1     6     6  531  521     2     1     1
75     50 75     1    19     9  432  398     5     1     4
35     45 35     2    24     NA  472  537     2     1     1
40     46 40     2     9     9  347  309     5     2     3
47     46 47     2    12    15  313  373     9     5     4
48     46 48     2     9    11  359  395     5     3     2
52     47 52     2     7     7  383   NA     5     2     3
53     48 53     2    18     7  444  563     6     4     2
63     49 63     2    18    18  378  373     3     2     1
66     49 66     2     7     7  675  636     3     1     2
76     50 76     2    17     9  510  448     3     1     2
> >
> # Let us now sort these records by treat. (drug=1, placebo=2)
>
> tt6 <- tt5[order(tt5[,3]),]
> tt6
  clinic id treat thick1 thick2 mob1 mob2 ass1 ass2 improve
34     45 34     1     38     NA  184  195     7     6     1
36     45 36     1     26     NA  255  211     5     2     3
43     46 43     1     7      7  440  444     4     3     1
44     46 44     1    17     18  191  192     7     5     2
51     47 51     1    28     28  191   NA     9     8     1
59     48 59     1    26     30  211  310     9     7     2
69     49 69     1     6     6  531  521     2     1     1
75     50 75     1    19     9  432  398     5     1     4
35     45 35     2    24     NA  472  537     2     1     1
40     46 40     2     9     9  347  309     5     2     3
47     46 47     2    12    15  313  373     9     5     4
48     46 48     2     9    11  359  395     5     3     2
52     47 52     2     7     7  383   NA     5     2     3
53     48 53     2    18     7  444  563     6     4     2
63     49 63     2    18    18  378  373     3     2     1
66     49 66     2     7     7  675  636     3     1     2
76     50 76     2    17     9  510  448     3     1     2
>
>
> # This works because, the third column and the relative
> # position in the column is
>
> tt5[,3]

```

```

[1] 1 2 1 2 1 1 2 2 1 2 2 1 2 2 1 1 2
> order(tt5[,3])
[1] 1 3 5 6 9 12 15 16 2 4 7 8 10 11 13 14 17

>
> # Suppose that we wish to select of these that have clinic = 49
>
> clinic49 <- subset(tt6, clinic==49)
> clinic49
  clinic id treat thick1 thick2 mob1 mob2 ass1 ass2 improve
69     49 69     1      6      6 531  521    2    1         1
63     49 63     2     18     18 378  373    3    2         1
66     49 66     2      7      7 675  636    3    1         2
>

```