Using Goldvarb for multivariate analysis (Getting Factor Weights)

Here is a sample Results file, with my comments and explanations in Bold (and green). It shows you how to Recode a Token file until you get rid of the KnockOuts and Singletons. Then you can run the Binomial (Multivariate) analyses to produce Factor Weights.

For more info, consult: Tagliamonte, S. 2006. *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press.

At the end of this document, I've put in a table explaining what the codes actually stand for, but it's not really important for demonstration purposes.

This is data regarding presence/absence of coda R in New England speakers.

Distributional Analysis

```
• CELL CREATION • 03/18/09 14:29:46
••••••••••••••••••••••••••••••••••••••••••
    Name of token file: R_for_351.tkn
Name of condition file: Untitled.cnd
(
; Identity recode:  All groups included as is.
(1)
(2)
(3)
(4)
(5)
(6)
(7)
(8)
(9)
(10)
(11)
(12)
(13)
)
       Number of cells:  1394
  Application value(s):  01X
  Total no. of factors:  49
```

Get rid of the "X"column by erasing the "X" in the "Choose application value(s)" window that pops up after you say "Load Cells to Memory." (You cannot Exclude anything from the Dep. variable in Recode setup.)

Group		0	1	X	Total	%
1 (2)		0	1	X		
f	N	529	445	20	994	44.8
	%	53.2	44.8	2.0		
i	N	436	377	6	819	36.9
	%	53.2	46.0	0.7		
m	N	261	142	4	407	18.3
	%	64.1	34.9	1.0		
Total	N	1226	964	30	2220	
	%	55.2	43.4	1.4		
2 (3)		0	1	X		
a	N	220	221	5	446	20.1
	%	49.3	49.6	1.1		
i	N	107	72	2	181	8.2
	%	59.1	39.8	1.1		
2	N	520	401	11	932	42.0
	%	55.8	43.0	1.2		
o	N	219	132	6	357	16.1
	%	61.3	37.0	1.7		
e	N	122	74	3	199	9.0
	%	61.3	37.2	1.5		
u	N	36	56	3	95	4.3
	%	37.9	58.9	3.2		

Exclude a factor to get rid of this KnockOut. In the Recode set-up, choose "Exclude" and click on the "x" before you copy over this group.)

x	N	2	8	0	10	0.5
	%	20.0	80.0	0.0		* KnockOut *
Total	N	1226	964	30	2220	
	%	55.2	43.4	1.4		
3 (4)		0	1	X		
w	N	397	231	13	641	28.9
	%	61.9	36.0	2.0		

```
   c    N      473      365        4      842   37.9
        %     56.2     43.3      0.5

   v    N       41      138        6      185    8.3
        %     22.2     74.6      3.2
```

Collapse 2 or more factors together to get rid of these KnockOuts. (In the Recode set-up, highlight this indep. variable and then choose "Recode." Put in new letters to make fewer groups, such as replacing "k" with "c" so that the original "c" and the original "k" tokens will be combined as "c".)

```
   k    N       41       38        0       79    3.6
        %     51.9     48.1      0.0           * KnockOut *

   n    N      204      144        5      353   15.9
        %     57.8     40.8      1.4

   p    N       10       16        0       26    1.2
        %     38.5     61.5      0.0           * KnockOut *

   f    N       60       32        2       94    4.2
        %     63.8     34.0      2.1

 Total  N     1226      964       30     2220
        %     55.2     43.4      1.4
-------------------------------------------------
 4 (5)            0        1        X
   f    N      169      117        9      295   13.3
        %     57.3     39.7      3.1

   l    N     1057      847       21     1925   86.7
        %     54.9     44.0      1.1

 Total  N     1226      964       30     2220
        %     55.2     43.4      1.4
-------------------------------------------------
 5 (6)            0        1        X
   1    N      515      505       15     1035   46.6
        %     49.8     48.8      1.4

   0    N      482      281       13      776   35.0
        %     62.1     36.2      1.7

   w    N      229      178        2      409   18.4
        %     56.0     43.5      0.5

 Total  N     1226      964       30     2220
```



```
        %     55.2     43.4      1.4
-------------------------------------------------
 6 (7)            0        1        X
   1    N      514      503       16     1033   46.5
        %     49.8     48.7      1.5

   2    N      494      335        8      837   37.7
        %     59.0     40.0      1.0

   3    N      175      106        4      285   12.8
        %     61.4     37.2      1.4

   4    N       43       20        2       65    2.9
        %     66.2     30.8      3.1

 Total  N     1226      964       30     2220
        %     55.2     43.4      1.4
-------------------------------------------------
 7 (8)            0        1        X
   5    N       67       34        6      107    4.8
        %     62.6     31.8      5.6

   1    N      225      128        3      356   16.0
        %     63.2     36.0      0.8

   2    N      267      193        1      461   20.8
        %     57.9     41.9      0.2

   4    N      226      200        6      432   19.5
        %     52.3     46.3      1.4

   3    N      389      350        8      747   33.6
        %     52.1     46.9      1.1

   0    N       52       57        5      114    5.1
        %     45.6     50.0      4.4
```

Another place where you should Exclude this "garbage" factor.

```
   x    N        0        2        1        3    0.1
        %      0.0     66.7     33.3           * KnockOut *

 Total  N     1226      964       30     2220
        %     55.2     43.4      1.4
-------------------------------------------------
 8 (9)            0        1        X
   0    N     1096      868       28     1992   89.7
        %     55.0     43.6      1.4
```

```
   b   N      100       63        1      164   7.4
       %     61.0     38.4      0.6

   a   N       30       33        1       64   2.9
       %     46.9     51.6      1.6

 Total N     1226      964       30     2220
       %     55.2     43.4      1.4
-------------------------------------------------
 9 (10)         0        1        X
   1   N      136       79        5      220   9.9
       %     61.8     35.9      2.3

   2   N      115      287        4      406  18.3
       %     28.3     70.7      1.0

   3   N      262      342        5      609  27.4
       %     43.0     56.2      0.8

   4   N      308      112        3      423  19.1
       %     72.8     26.5      0.7

   5   N      284      117        5      406  18.3
       %     70.0     28.8      1.2

   7   N      121       27        8      156   7.0
       %     77.6     17.3      5.1

 Total N     1226      964       30     2220
       %     55.2     43.4      1.4
-------------------------------------------------
```

This is a "Singleton Group," because it only has 1 factor in it. There's no possible variation. Therefore, exclude the whole group by either:
(a) not copying it over in the Recode Setup, or
(b) erasing the group ("(11)") from the .Cnd file and re-saving it.

```
10 (11)         0        1        X
   F   N     1226      964       30     2220 100.0
       %     55.2     43.4      1.4          * Singleton Group *

 Total N     1226      964       30     2220
       %     55.2     43.4      1.4
-------------------------------------------------
11 (12)         0        1        X
   1   N      484      279       19      782  35.2
       %     61.9     35.7      2.4
```



```
   3   N      297      508        7      812  36.6
       %     36.6     62.6      0.9

   2   N      445      177        4      626  28.2
       %     71.1     28.3      0.6

 Total N     1226      964       30     2220
       %     55.2     43.4      1.4
-------------------------------------------------
12 (13)         0        1        X
   l   N      438      326       18      782  35.2
       %     56.0     41.7      2.3

   u   N      470      553        9     1032  46.5
       %     45.5     53.6      0.9

   m   N      318       85        3      406  18.3
       %     78.3     20.9      0.7

 Total N     1226      964       30     2220
       %     55.2     43.4      1.4
-------------------------------------------------
 TOTAL N     1226      964       30     2220
       %     55.2     43.4      1.4


 Name of new cell file: .cel
```

If you do the recoding as I describe above, you should get a .Res file with no KnockOuts and no Singletons, like this:

```
• CELL CREATION • 03/18/09 16:57:50
•••••••••••••••••••••••••••••••••••••••••••
    Name of token file: R_for_351.tkn
Name of condition file: Untitled.cnd
(
(1 (NIL (COL 8 x))
   (NIL (COL 3 x)))
(2)
(3)
(4 (w (COL 4 w))
   (c (COL 4 c))
   (v (COL 4 v))
   (c (COL 4 k))
   (n (COL 4 n))
   (p (COL 4 p))
```

```
   (f (COL 4 f)))
(5)
(6)
(7)
(8)
(9)
(10)
(12)
(13)
)
```

```
       Number of cells:  1361
  Application value(s):  10
  Total no. of factors:  45
```

Group		1	0	Total	%
1 (2)		1	0		
f	N	435	529	964	44.3
	%	45.1	54.9		
i	N	377	434	811	37.2
	%	46.5	53.5		
m	N	142	261	403	18.5
	%	35.2	64.8		
Total	N	954	1224	2178	
	%	43.8	56.2		
2 (3)		1	0		
a	N	221	220	441	20.2
	%	50.1	49.9		
i	N	70	107	177	8.1
	%	39.5	60.5		
2	N	401	520	921	42.3
	%	43.5	56.5		
o	N	132	219	351	16.1
	%	37.6	62.4		
e	N	74	122	196	9.0
	%	37.8	62.2		
u	N	56	36	92	4.2

	%	60.9	39.1		
Total	N	954	1224	2178	
	%	43.8	56.2		
3 (4)		1	0		
w	N	227	397	624	28.7
	%	36.4	63.6		
c	N	402	514	916	42.1
	%	43.9	56.1		
v	N	134	41	175	8.0
	%	76.6	23.4		
n	N	144	202	346	15.9
	%	41.6	58.4		
p	N	15	10	25	1.1
	%	60.0	40.0		
f	N	32	60	92	4.2
	%	34.8	65.2		
Total	N	954	1224	2178	
	%	43.8	56.2		
4 (5)		1	0		
f	N	117	169	286	13.1
	%	40.9	59.1		
l	N	837	1055	1892	86.9
	%	44.2	55.8		
Total	N	954	1224	2178	
	%	43.8	56.2		
5 (6)		1	0		
1	N	495	515	1010	46.4
	%	49.0	51.0		
0	N	281	480	761	34.9
	%	36.9	63.1		
w	N	178	229	407	18.7
	%	43.7	56.3		

Total	N	954	1224	2178	
	%	43.8	56.2		
6 (7)		1	0		
1	N	493	514	1007	46.2
	%	49.0	51.0		
2	N	335	494	829	38.1
	%	40.4	59.6		
3	N	106	175	281	12.9
	%	37.7	62.3		
4	N	20	41	61	2.8
	%	32.8	67.2		
Total	N	954	1224	2178	
	%	43.8	56.2		
7 (8)		1	0		
5	N	34	67	101	4.6
	%	33.7	66.3		
1	N	128	225	353	16.2
	%	36.3	63.7		
2	N	193	265	458	21.0
	%	42.1	57.9		
4	N	195	226	421	19.3
	%	46.3	53.7		
3	N	350	389	739	33.9
	%	47.4	52.6		
0	N	54	52	106	4.9
	%	50.9	49.1		
Total	N	954	1224	2178	
	%	43.8	56.2		
8 (9)		1	0		
0	N	858	1094	1952	89.6
	%	44.0	56.0		
b	N	63	100	163	7.5
	%	38.7	61.3		

a	N	33	30	63	2.9
	%	52.4	47.6		
Total	N	954	1224	2178	
	%	43.8	56.2		
9 (10)		1	0		
1	N	79	136	215	9.9
	%	36.7	63.3		
2	N	280	115	395	18.1
	%	70.9	29.1		
3	N	341	261	602	27.6
	%	56.6	43.4		
4	N	111	307	418	19.2
	%	26.6	73.4		
5	N	116	284	400	18.4
	%	29.0	71.0		
7	N	27	121	148	6.8
	%	18.2	81.8		
Total	N	954	1224	2178	
	%	43.8	56.2		
10 (12)		1	0		
1	N	274	483	757	34.8
	%	36.2	63.8		
3	N	504	296	800	36.7
	%	63.0	37.0		
2	N	176	445	621	28.5
	%	28.3	71.7		
Total	N	954	1224	2178	
	%	43.8	56.2		
11 (13)		1	0		
l	N	325	437	762	35.0
	%	42.7	57.3		
u	N	544	469	1013	46.5

```
       %    53.7    46.3

   m   N      85     318     403  18.5
       %    21.1    78.9

 Total N     954    1224    2178
       %    43.8    56.2
-------------------------------------
 TOTAL N     954    1224    2178
       %    43.8    56.2

 Name of new cell file: Untitled.cel
```

For demo purposes, I'd suggest also deleting Groups (5-13) before running the Binomial analysis, so it doesn't take too long.

For the Binomial one-level (1st) and the Binomial up-and-down (2nd), here's what you might see:

Binomial, One-Level

```
• BINOMIAL VARBRUL, 1 step • 03/18/09 17:03:09
•••••••••••••••••••••••••••••••
Name of cell file: Untitled.cel

Averaging by weighting factors.
One-level binomial analysis...

Run # 1, 54 cells:
Convergence at Iteration 17
Input 0.438
```

In the table below the first column of numbers are the Factor Weights. The 2nd column of numbers is the proportion of [r]'s out of the total (the % [r]) – it matches the percentages for the application value reported in the distributional analysis above. You can ignore the final column.

```
Group Factor Weight App/Total Input&Weight

   1:    f    0.543    0.45       0.48
         i    0.507    0.46       0.45
         m    0.386    0.35       0.33

   2:    a    0.545    0.50       0.48
         i    0.503    0.40       0.44
         2    0.499    0.44       0.44
         o    0.403    0.38       0.35
```



```
         e    0.496    0.38       0.43
         u    0.665    0.61       0.61

   3:    w    0.369    0.36       0.31
         c    0.540    0.44       0.48
         v    0.778    0.77       0.73
         n    0.501    0.42       0.44
         p    0.633    0.60       0.57
         f    0.376    0.35       0.32
```

You can ignore this long list.

```
  Cell  Total    App'ns      Expected      Error
   muc    11         6         5.865      0.007
   mon    11         4         2.743      0.768
   mof     1         1         0.166      5.017
   moc    31        11         8.669      0.870
   min    29         8         9.628      0.412
   mic    33        12        12.125      0.002
...
   f2p    17         9        10.432      0.509
   f2n    10         4         4.805      0.260
   f2f    59        20        21.059      0.083
   f2c    26        12        13.506      0.349

Total Chi-square = 116.5183
 Chi-square/cell = 2.1577
Log likelihood = -1422.679
```

These numbers at the bottom of the one-level analysis are used for comparing how how good one model is to another. Don't need to worry about them now.

Binomial, Up and Down

```
• BINOMIAL VARBRUL • 03/18/09 17:02:16
••••••••••••••••••••••••••••••••••••••
Name of cell file: Untitled.cel

Averaging by weighting factors.
Threshold, step-up/down: 0.050001

Stepping up...

---------- Level # 0 ----------

Run # 1, 1 cells:
Convergence at Iteration 2
Input 0.438
```

```
Log likelihood = -1492.896


---------- Level # 1 ----------

Run # 2, 3 cells:
Convergence at Iteration 4
Input 0.437
Group # 1 -- f: 0.514, i: 0.528, m: 0.412
Log likelihood = -1485.245  Significance = 0.000

Run # 3, 6 cells:
Convergence at Iteration 5
Input 0.437
Group # 2 -- a: 0.564, i: 0.457, 2: 0.498, o: 0.437, e: 0.438,
u: 0.666
Log likelihood = -1479.044  Significance = 0.000

Run # 4, 6 cells:
Convergence at Iteration 5
Input 0.439
Group # 3 -- w: 0.422, c: 0.500, v: 0.806, n: 0.476, p: 0.657,
f: 0.405
Log likelihood = -1443.611  Significance = 0.000

Add Group # 3 with factors wcvnpf

---------- Level # 2 ----------

Run # 5, 17 cells:
Convergence at Iteration 19
Input 0.439
Group # 1 -- f: 0.551, i: 0.498, m: 0.384
Group # 3 -- w: 0.375, c: 0.534, v: 0.777, n: 0.511, p: 0.619,
f: 0.367
Log likelihood = -1434.937  Significance = 0.000

Run # 6, 32 cells:
Convergence at Iteration 6
Input 0.439
Group # 2 -- a: 0.566, i: 0.481, 2: 0.491, o: 0.419, e: 0.478,
u: 0.660
Group # 3 -- w: 0.414, c: 0.508, v: 0.803, n: 0.469, p: 0.670,
f: 0.417
Log likelihood = -1430.810  Significance = 0.000

Add Group # 2 with factors ai2oeu
```



```
---------- Level # 3 ----------

Run # 7, 54 cells:
Convergence at Iteration 17
Input 0.438
Group # 1 -- f: 0.543, i: 0.507, m: 0.386
Group # 2 -- a: 0.545, i: 0.503, 2: 0.499, o: 0.403, e: 0.496,
u: 0.665
Group # 3 -- w: 0.369, c: 0.540, v: 0.778, n: 0.501, p: 0.633,
f: 0.376
Log likelihood = -1422.679  Significance = 0.000

Add Group # 1 with factors fim

Best stepping up run:  #7
-------------------------------------------

Stepping down...

---------- Level # 3 ----------

Run # 8, 54 cells:
Convergence at Iteration 17
Input 0.438
Group # 1 -- f: 0.543, i: 0.507, m: 0.386
Group # 2 -- a: 0.545, i: 0.503, 2: 0.499, o: 0.403, e: 0.496,
u: 0.665
Group # 3 -- w: 0.369, c: 0.540, v: 0.778, n: 0.501, p: 0.633,
f: 0.376
Log likelihood = -1422.679

---------- Level # 2 ----------

Run # 9, 32 cells:
Convergence at Iteration 6
Input 0.439
Group # 2 -- a: 0.566, i: 0.481, 2: 0.491, o: 0.419, e: 0.478,
u: 0.660
Group # 3 -- w: 0.414, c: 0.508, v: 0.803, n: 0.469, p: 0.670,
f: 0.417
Log likelihood = -1430.810  Significance = 0.000

Run # 10, 17 cells:
Convergence at Iteration 19
Input 0.439
Group # 1 -- f: 0.551, i: 0.498, m: 0.384
```

```
Group # 3 -- w: 0.375, c: 0.534, v: 0.777, n: 0.511, p: 0.619,
f: 0.367
Log likelihood = -1434.937  Significance = 0.000

Run # 11, 18 cells:
Convergence at Iteration 5
Input 0.437
Group # 1 -- f: 0.509, i: 0.530, m: 0.420
Group # 2 -- a: 0.546, i: 0.476, 2: 0.504, o: 0.425, e: 0.451,
u: 0.667
Log likelihood = -1473.116  Significance = 0.000

All remaining groups significant
```

The 3 lines below show which factors should be kept in an analysis and which should be excluded. In this case, all are relevant, so none are eliminated. It says that the BEST combo of factor weights for each factor can be found be looking above at Runs #7 and #8. You'll see that they are identical, and that the weights are the same as in the one-level analyis above.

The ORDER in which the groups are added to the analysis is their order of their Log likelihood (closer to 0 = better). The first one added has the greatest explanatory value. In the above example, Group 3 was added first (see end of Level 1 in the step-up); Group 2 was added second (see end of Level 2 in the step-up); and Group 1 was added last (see end of Level 3 in the step-up).

Usually there will be a line at the end of the step-up specifying the order that the factors were added, e.g., "Groups selected while stepping up: 3 2 1."

```
Groups eliminated while stepping down: None
Best stepping  up  run: #7
Best stepping down run: #8
```


Coding Schema

I. DEPENDENT VARIABLE: Presence or Absence of [r]

Code	Description
0	[r] not present
1	[r] present

II. Linguistic Independent Variables

Position of (r)

Code	Description	Example
f	word-final	"pair"
i	morpheme-internal	"Arkansas"
m	morpheme-final	"miserly"

Preceding Segment

Code	IPA	*Description*	*example*
i	i	high front tense	"beer"
e	e	mid front tense	"bear"
a	a	low central	"bar"
o	oh	mid back round	"bore"
U	u	high back round	"boor"
2	ə	schwa	"runner"
x	j	glides or other	"your"; "p'ticipate"

Following Segment

Code	Description	Example
c	Following C in same syllable	"card game"
n	Following C in next syllable	"carton," "car talk"
w	Following C, next word	"car sounds"
v	Following V, next word	"car is"
p	Pause followed by following V	"car – is"
k	Pause followed by following C	"car – talk"
f	Utterance-final/Long pause	"car –"

Word class of (r) word

f	function word	"are", "for"
l	non-function (lexical) word	"car"

Stress/Focus

1	monosyllabic word	The car is here.
0	no stress	My brother is wrong.
w	word stress	The carburetor is broken.

Number of syllables

1	monosyllabic	"beard"
2	bisyllabic	"curtain"
3	trisyllablic	"appearance"
4	four or more syllables	"opportunities"

Lexical frequency

(from the American National Corpus, spoken text)

Code	Description	Example
0	does not appear in the corpus	snow-covered
1	appears <100 times in the corpus of spoken text	fierce
2	<1,000	appear
3	<10,000	warm
4	<100,000	our
5	<1,000,000	there

Another r in the word?

0	no other	carton
b	another /r/ before	recard
a	another /r/ after	carder

III. Social Independent Variables

Age/Decade

Code	Description
1	teenage (18-19)
2	20-29
3	30-39
4	40-49
5	50-59
6	60-69
7	70+9

Sex

F	female
M	male

Education

1	high school
2	college
3	post-college

Occupation/Income

L	Lower: <$40,000/yr
M	Middle: $40-50,000/yr
U	Upper: >$50,000/yr