ORIGIN ≡ 0

Coded Factors and Contrast Matrices

W. Stein

"Qualitative predictor" is a term used for independent variables in multiple regression that have discrete states as values. The discrete states may be "ordered" in the sense of "ordinal" data implying that some states of the variable are closer to others, or "unordered" implying no information about a state's relative value. In R and many statistical packages, qualitative predictors are called "factors" - a term with identical meaning in multiple regression and ANOVA designs since both are fundamentally based on the Linear Model approach. In using factors, certain rules in coding the states must be observed. A binary factor involving only two states is easily coded using 0 for one state and 1 for the other (and it doesn't matter which is which). For multistate factors (i.e., 3 or more states), the rules are more complex in order to assure that the design matrix (consisting of 1 & X's) is not over-parameterized (it must be of full rank to allow finding of the inverse of X'X). In general, for p states of a factor, (p-1) "dummy variables" are created. There are multiple ways to code dummy variables, but fortunately, R and other statistical packages have powerful built-in functions largely insulating the user from the tedious paperwork. Typically, factor coding is handled internally by converting "indicator-coded variables" (with p columns of 1's or 0's for p states) or single factor with non-numeric state names into "dummy codes" using a "contrast matrix". Examples below comes from Kuter et al. (KNNL) A*pplied Linear Statistical Models* 5th Edition.

Worked Example:

Insurance Innovation Example KNNL Table 8.2

Reading & Analyzing Data using R:

```
#QUALITATIVE PREDICTORS
# READ STRUCTURED DATA TABLE
K=read.table("c:/2008LinearModelsData/InsInnovationR.txt")
K
attach(K)
options(digits=7)

#FITTING FULL MODEL WITH FACTOR
FM=lm(Y~X+FAC)
summary(FM)
anova(FM)
```

```
> K
    Y   X FAC
1  17 151   0
2  26  92   0
3  21 175   0
4  30  31   0
5  22 104   0
6   0 277   0
7  12 210   0
8  19 120   0
9   4 290   0
10 16 238   0
11 28 164   1
12 15 272   1
13 11 295   1
14 38  68   1
15 31  85   1
16 21 224   1
17 20 166   1
18 13 305   1
19 30 124   1
20 14 246   1
```

```
> summary(FM)
Call: lm(formula = Y ~ X + FAC)
Residuals:
    Min      1Q  Median      3Q     Max
-5.6915 -1.7036 -0.4385  1.9210  6.3406
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 33.874069   1.813858  18.675 9.15e-13 ***
X           -0.101742   0.008891 -11.443 2.07e-09 ***
FAC          8.055469   1.459106   5.521 3.74e-05 ***
```

```
> anova(FM)
Analysis of Variance Table
Response: Y
          Df  Sum Sq Mean Sq F value    Pr(>F)
X          1 1188.17 1188.17 114.515 5.683e-09 ***
FAC        1  316.25  316.25  30.480 3.742e-05 ***
Residuals 17  176.39   10.38
```

1188.17 + 316.25 = 1504.42

^ all values verified KNNL p. 317

More Interesting Multi-State Example: Surgical Unit Example KNNL Table 9.1

Reading & Analyzing Data using R - Numeric Codes for Factor levels:

Column X7c contains numeric codes that must be converted to factor levels using R's function factor()

```
#LOOKING AT FACTORS

#READ STRUCTURED DATA TABLE WITH
#NUMERIC CODED FACTOR
K1=read.table("c:/2008LinearModelsData/SurgicalUnit1R.txt")
K1
attach(K1)

#SETTING VARIABLE X7 AS A FACTOR
**FOR NUMERIC CODES**
X7=factor(X7c)

#VIEWING CONTRAST & DESIGN MATRIX X
FM=lm(Y~X1+X2+X3+X4+X5+X6+X7)
contrasts(X7)
model.matrix(FM)
summary(FM)
anova(FM)
```

> K1

```
    X1 X2  X3   X4 X5 X6 X7c    Y   lnY
1  6.7 62  81 2.59 50  0   2  695 6.544
2  7.8 65 115 4.30 45  0   3 2343 7.759
3  5.7 46  63 1.91 49  1   3  518 6.250
4  5.8 38  72 1.42 65  1   2  348 5.852
5  7.7 62  67 3.40 58  0   3  702 6.554
6  3.7 68  81 2.57 69  1   2  749 6.619
7  5.6 57  87 3.02 63  0   3  838 6.731
8  6.0 67  93 2.50 58  0   2 1056 6.962
9  6.7 26  68 2.10 30  0   3  599 6.395
...
46 7.3 68  74 3.56 59  1   1  550 6.309
47 5.2 52  76 2.85 39  0   1  359 5.883
48 5.2 54  56 2.71 44  1   1  477 6.167
49 2.6 74  86 2.05 45  0   1  678 6.519
50 4.3  8 119 2.85 65  1   1  362 5.893
51 5.4 52  88 1.81 40  1   1  705 6.558
52 5.2 49  72 1.84 46  0   1  536 6.283
53 5.1 67  77 2.86 66  1   1  581 6.365
54 8.8 78  72 3.20 56  0   1  651 6.478
```

> summary(FM)

```
Call: lm(formula = Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7)
Residuals:
     Min       1Q   Median       3Q      Max
-285.360 -132.747   -9.998   89.477  790.119
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1148.823    242.328  -4.741 2.17e-05 ***
X1             62.390     24.470   2.550 0.014258 *
X2              8.973      1.874   4.788 1.86e-05 ***
X3              9.888      1.742   5.677 9.39e-07 ***
X4             50.413     44.959   1.121 0.268109
X5             -0.951      2.649  -0.359 0.721231
X6             15.874     58.475   0.271 0.787269
X72             7.713     64.956   0.119 0.906007
X73           320.697     85.070   3.770 0.000474 ***
```

> contrasts(X7)

```
  2 3
1 0 0
2 1 0
3 0 1
```

> anova(FM)

```
Analysis of Variance Table
Response: Y
          Df  Sum Sq Mean Sq   F value    Pr(>F)
X1         1 1005152 1005152   24.7723 9.925e-06 ***
X2         1 1278496 1278496   31.5089 1.167e-06 ***
X3         1 3442172 3442172   84.8334 6.407e-12 ***
X4         1   57862   57862    1.4260 0.2386707
X5         1   33032   33032    0.8141 0.3717186
X6         1       1       1 1.423e-05 0.9970070
X7         2  726901  363450    8.9573 0.0005313 ***
Residuals 45 1825906   40576
```

> model.matrix(FM)

```
   (Intercept)   X1 X2  X3   X4 X5 X6 X72 X73
1            1  6.7 62  81 2.59 50  0   1   0
2            1  7.8 65 115 4.30 45  0   0   1
3            1  5.7 46  63 1.91 49  1   0   1
4            1  5.8 38  72 1.42 65  1   1   0
5            1  7.7 62  67 3.40 58  0   0   1
6            1  3.7 68  81 2.57 69  1   1   0
7            1  5.6 57  87 3.02 63  0   0   1
8            1  6.0 67  93 2.50 58  0   1   0
9            1  6.7 26  68 2.10 30  0   0   1
...
46           1  7.3 68  74 3.56 59  1   0   0
47           1  5.2 52  76 2.85 39  0   0   0
48           1  5.2 54  56 2.71 44  1   0   0
49           1  2.6 74  86 2.05 45  0   0   0
50           1  4.3  8 119 2.85 65  1   0   0
51           1  5.4 52  88 1.81 40  1   0   0
52           1  5.2 49  72 1.84 46  0   0   0
53           1  5.1 67  77 2.86 66  1   0   0
54           1  8.8 78  72 3.20 56  0   0   0
```

Reading & Analyzing Data using R - Text Codes for Factor levels:

Column X7 contains text codes that are automatically inerpreted by R to be unordered factor levels, although it never hurts to specify this explicitly.

```
#READ STRUCTURED DATA TABLE WITH TEXT CODED FACTOR
K2=read.table("c:/2008LinearModelsData/SurgicalUnit2R.txt")
K2
detach(K1)
attach(K2)

#VIEWING CONTRAST & DESIGN MATRIX X
FM=lm(Y~X1+X2+X3+X4+X5+X6+X7)
contrasts(X7)
model.matrix(FM)
summary(FM)
anova(FM)
```

> K2

```
    X1 X2  X3   X4 X5 X6    X7    Y   lnY
1  6.7 62  81 2.59 50  0   mod  695 6.544
2  7.8 65 115 4.30 45  0 heavy 2343 7.759
3  5.7 46  63 1.91 49  1 heavy  518 6.250
4  5.8 38  72 1.42 65  1   mod  348 5.852
5  7.7 62  67 3.40 58  0 heavy  702 6.554
6  3.7 68  81 2.57 69  1   mod  749 6.619
7  5.6 57  87 3.02 63  0 heavy  838 6.731
8  6.0 67  93 2.50 58  0   mod 1056 6.962
9  6.7 26  68 2.10 30  0 heavy  599 6.395
...
46 7.3 68  74 3.56 59  1  none  550 6.309
47 5.2 52  76 2.85 39  0  none  359 5.883
48 5.2 54  56 2.71 44  1  none  477 6.167
49 2.6 74  86 2.05 45  0  none  678 6.519
50 4.3  8 119 2.85 65  1  none  362 5.893
51 5.4 52  88 1.81 40  1  none  705 6.558
52 5.2 49  72 1.84 46  0  none  536 6.283
53 5.1 67  77 2.86 66  1  none  581 6.365
54 8.8 78  72 3.20 56  0  none  651 6.478
```

> contrasts(X7)

```
  2 3
1 0 0
2 1 0
3 0 1
```

> summary(FM)

```
Call: lm(formula = Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7)
Residuals:
     Min       1Q   Median       3Q      Max
-285.360 -132.747   -9.998   89.477  790.119
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1148.823    242.328  -4.741 2.17e-05 ***
X1             62.390     24.470   2.550 0.014258 *
X2              8.973      1.874   4.788 1.86e-05 ***
X3              9.888      1.742   5.677 9.39e-07 ***
X4             50.413     44.959   1.121 0.268109
X5             -0.951      2.649  -0.359 0.721231
X6             15.874     58.475   0.271 0.787269
X72             7.713     64.956   0.119 0.906007
X73           320.697     85.070   3.770 0.000474 ***
```

> anova(FM)

```
Analysis of Variance Table
Response: Y
          Df  Sum Sq Mean Sq   F value    Pr(>F)
X1         1 1005152 1005152   24.7723 9.925e-06 ***
X2         1 1278496 1278496   31.5089 1.167e-06 ***
X3         1 3442172 3442172   84.8334 6.407e-12 ***
X4         1   57862   57862    1.4260 0.2386707
X5         1   33032   33032    0.8141 0.3717186
X6         1       1       1 1.423e-05 0.9970070
X7         2  726901  363450    8.9573 0.0005313 ***
Residuals 45 1825906   40576
```

> model.matrix(FM)

```
   (Intercept)   X1 X2  X3   X4 X5 X6 X72 X73
1            1  6.7 62  81 2.59 50  0   1   0
2            1  7.8 65 115 4.30 45  0   0   1
3            1  5.7 46  63 1.91 49  1   0   1
4            1  5.8 38  72 1.42 65  1   1   0
5            1  7.7 62  67 3.40 58  0   0   1
6            1  3.7 68  81 2.57 69  1   1   0
7            1  5.6 57  87 3.02 63  0   0   1
8            1  6.0 67  93 2.50 58  0   1   0
9            1  6.7 26  68 2.10 30  0   0   1
...
46           1  7.3 68  74 3.56 59  1   0   0
47           1  5.2 52  76 2.85 39  0   0   0
48           1  5.2 54  56 2.71 44  1   0   0
49           1  2.6 74  86 2.05 45  0   0   0
50           1  4.3  8 119 2.85 65  1   0   0
51           1  5.4 52  88 1.81 40  1   0   0
52           1  5.2 49  72 1.84 46  0   0   0
53           1  5.1 67  77 2.86 66  1   0   0
54           1  8.8 78  72 3.20 56  0   0   0
```

Reading & Analyzing Data using R - Indicator Coded Data:

Columns X7n,X7m,X7h contain indicator codes that are over-parameterized. These must be converted to "dummy codes" using a contrast matrix...

```
#READ STRUCTURED DATA TABLE WITH
#INDICATOR CODES FOR FACTOR
K3=read.table("c:/2008LinearModelsData/
SurgicalUnit3R.txt")
K3
detach(K2)
attach(K3)

#DIRECT CALCULATION WITH CONTRAST MATRIX
X7c=cbind(X7n,X7m,X7h)
X7c
Ct=contr.treatment(3)
Ct

#MAKING CODED FACTOR
X7=X7c%*%Ct  #MATRIX MULTIPILICATION
X7

#VIEWING CONTRAST & DESIGN MATRIX X
FM=lm(Y~X1+X2+X3+X4+X5+X6+X7)
model.matrix(FM)
summary(FM)
anova(FM)
```

```
> K3
     X1 X2  X3   X4 X5 X6 X7n X7m X7h Y  lnY
1   6.7 62  81 2.59 50  0   0   1   0  695 6.544
2   7.8 65 115 4.30 45  0   0   0   1 2343 7.759
3   5.7 46  63 1.91 49  1   0   0   1  518 6.250
4   5.8 38  72 1.42 65  1   0   1   0  348 5.852
5   7.7 62  67 3.40 58  0   0   0   1  702 6.554
6   3.7 68  81 2.57 69  1   0   1   0  749 6.619
7   5.6 57  87 3.02 63  0   0   0   1  838 6.731
8   6.0 67  93 2.50 58  0   0   1   0 1056 6.962
9   6.7 26  68 2.10 30  0   0   0   1  599 6.395
...
46  7.3 68  74 3.56 59  1   1   0   0  550 6.309
47  5.2 52  76 2.85 39  0   1   0   0  359 5.883
48  5.2 54  56 2.71 44  1   1   0   0  477 6.167
49  2.6 74  86 2.05 45  0   1   0   0  678 6.519
50  4.3  8 119 2.85 65  1   1   0   0  362 5.893
51  5.4 52  88 1.81 40  1   1   0   0  705 6.558
52  5.2 49  72 1.84 46  0   1   0   0  536 6.283
53  5.1 67  77 2.86 66  1   1   0   0  581 6.365
54  8.8 78  72 3.20 56  0   1   0   0  651 6.478
```

```
> Ct=contr.treatment(3)
> Ct
  2 3
1 0 0
2 1 0
3 0 1
```

< contrast matrix Ct

```
> X7c
      X7n X7m X7h
 [1,]   0   1   0
 [2,]   0   0   1
 [3,]   0   0   1
 [4,]   0   1   0
 [5,]   0   0   1
 [6,]   0   1   0
 [7,]   0   0   1
 [8,]   0   1   0
 [9,]   0   0   1
...
[46,]   1   0   0
[47,]   1   0   0
[48,]   1   0   0
[49,]   1   0   0
[50,]   1   0   0
[51,]   1   0   0
[52,]   1   0   0
[53,]   1   0   0
[54,]   1   0   0
```

^ indicator codes

```
> X7
      2 3
 [1,] 1 0
 [2,] 0 1
 [3,] 0 1
 [4,] 1 0
 [5,] 0 1
 [6,] 1 0
 [7,] 0 1
 [8,] 1 0
 [9,] 0 1
...
[46,] 0 0
[47,] 0 0
[48,] 0 0
[49,] 0 0
[50,] 0 0
[51,] 0 0
[52,] 0 0
[53,] 0 0
[54,] 0 0
```

^ treatment "dummy codes"

```
> model.matrix(FM)
   (Intercept)   X1 X2  X3   X4 X5 X6 X72 X73
1            1  6.7 62  81 2.59 50  0   1   0
2            1  7.8 65 115 4.30 45  0   0   1
3            1  5.7 46  63 1.91 49  1   0   1
4            1  5.8 38  72 1.42 65  1   1   0
5            1  7.7 62  67 3.40 58  0   0   1
6            1  3.7 68  81 2.57 69  1   1   0
7            1  5.6 57  87 3.02 63  0   0   1
8            1  6.0 67  93 2.50 58  0   1   0
9            1  6.7 26  68 2.10 30  0   0   1
...
46           1  7.3 68  74 3.56 59  1   0   0
47           1  5.2 52  76 2.85 39  0   0   0
48           1  5.2 54  56 2.71 44  1   0   0
49           1  2.6 74  86 2.05 45  0   0   0
50           1  4.3  8 119 2.85 65  1   0   0
51           1  5.4 52  88 1.81 40  1   0   0
52           1  5.2 49  72 1.84 46  0   0   0
53           1  5.1 67  77 2.86 66  1   0   0
54           1  8.8 78  72 3.20 56  0   0   0
```

> summary(FM)

```
Call:
lm(formula = Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7)
Residuals:
     Min       1Q   Median       3Q      Max
-285.360 -132.747   -9.998   89.477  790.119
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1148.823    242.328  -4.741 2.17e-05 ***
X1             62.390     24.470   2.550 0.014258 *
X2              8.973      1.874   4.788 1.86e-05 ***
X3              9.888      1.742   5.677 9.39e-07 ***
X4             50.413     44.959   1.121 0.268109
X5             -0.951      2.649  -0.359 0.721231
X6             15.874     58.475   0.271 0.787269
X72             7.713     64.956   0.119 0.906007
X73           320.697     85.070   3.770 0.000474 ***
```

> anova(FM)

```
Analysis of Variance Table

Response: Y
          Df  Sum Sq Mean Sq   F value     Pr(>F)
X1         1 1005152 1005152   24.7723 9.925e-06 ***
X2         1 1278496 1278496   31.5089 1.167e-06 ***
X3         1 3442172 3442172   84.8334 6.407e-12 ***
X4         1   57862   57862    1.4260 0.2386707
X5         1   33032   33032    0.8141 0.3717186
X6         1       1       1 1.423e-05 0.9970070
X7         2  726901  363450    8.9573 0.0005313 ***
Residuals 45 1825906   40576
```

Different Kinds of Contrasts for Unordered Factors in R:

Treatment Contrasts (R default):

#TREATMENT CONTRASTS (DEFAULT)
X7=factor(K1$X7c)
contrasts(X7)
FM=lm(Y~X1+X2+X3+X4+X5+X6+X7)
model.matrix(FM)
summary(FM)
anova(FM)

> contrasts(X7)

```
  2 3
1 0 0
2 1 0
3 0 1
```

> model.matrix(FM)

```
   (Intercept)   X1 X2  X3   X4 X5 X6 X72 X73
1            1  6.7 62  81 2.59 50  0   1   0
2            1  7.8 65 115 4.30 45  0   0   1
3            1  5.7 46  63 1.91 49  1   0   1
4            1  5.8 38  72 1.42 65  1   1   0
5            1  7.7 62  67 3.40 58  0   0   1
...
46           1  7.3 68  74 3.56 59  1   0   0
47           1  5.2 52  76 2.85 39  0   0   0
48           1  5.2 54  56 2.71 44  1   0   0
49           1  2.6 74  86 2.05 45  0   0   0
50           1  4.3  8 119 2.85 65  1   0   0
51           1  5.4 52  88 1.81 40  1   0   0
52           1  5.2 49  72 1.84 46  0   0   0
53           1  5.1 67  77 2.86 66  1   0   0
54           1  8.8 78  72 3.20 56  0   0   0
```

> summary(FM)

```
Call:
lm(formula = Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7)
Residuals:
     Min       1Q   Median       3Q      Max
-285.360 -132.747   -9.998   89.477  790.119
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1148.823    242.328  -4.741 2.17e-05 ***
X1             62.390     24.470   2.550 0.014258 *
X2              8.973      1.874   4.788 1.86e-05 ***
X3              9.888      1.742   5.677 9.39e-07 ***
X4             50.413     44.959   1.121 0.268109
X5             -0.951      2.649  -0.359 0.721231
X6             15.874     58.475   0.271 0.787269
X72             7.713     64.956   0.119 0.906007
X73           320.697     85.070   3.770 0.000474 ***
```

> anova(FM)

```
Analysis of Variance Table
Response: Y
          Df  Sum Sq Mean Sq   F value     Pr(>F)
X1         1 1005152 1005152   24.7723 9.925e-06 ***
X2         1 1278496 1278496   31.5089 1.167e-06 ***
X3         1 3442172 3442172   84.8334 6.407e-12 ***
X4         1   57862   57862    1.4260 0.2386707
X5         1   33032   33032    0.8141 0.3717186
X6         1       1       1 1.423e-05 0.9970070
X7         2  726901  363450    8.9573 0.0005313 ***
Residuals 45 1825906   40576
```

Helmert Contrasts:

```
#SETTING HELMERT CONTRASTS
contrasts(K2$X7)=contr.helmert
FM=lm(Y~X1+X2+X3+X4+X5+X6+K2$X7)
contrasts(K2$X7)
model.matrix(FM)
summary(FM)
anova(FM)
```

> contrasts(K2$X7)

```
      [,1] [,2]
heavy   -1   -1
mod      1   -1
none     0    2
```

> model.matrix(FM)

```
   (Intercept)   X1 X2  X3   X4 X5 X6 K2$X71 K2$X72
1            1  6.7 62  81 2.59 50  0      1     -1
2            1  7.8 65 115 4.30 45  0     -1     -1
3            1  5.7 46  63 1.91 49  1     -1     -1
4            1  5.8 38  72 1.42 65  1      1     -1
5            1  7.7 62  67 3.40 58  0     -1     -1

46           1  7.3 68  74 3.56 59  1      0      2
47           1  5.2 52  76 2.85 39  0      0      2
48           1  5.2 54  56 2.71 44  1      0      2
49           1  2.6 74  86 2.05 45  0      0      2
50           1  4.3  8 119 2.85 65  1      0      2
51           1  5.4 52  88 1.81 40  1      0      2
52           1  5.2 49  72 1.84 46  0      0      2
53           1  5.1 67  77 2.86 66  1      0      2
54           1  8.8 78  72 3.20 56  0      0      2
```

> summary(FM)

```
Call:
lm(formula = Y ~ X1 + X2 + X3 + X4 + X5 + X6 + K2$X7)
Residuals:
     Min       1Q   Median       3Q      Max
-285.360 -132.747   -9.998   89.477  790.119
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1039.353    243.484  -4.269 0.000100 ***
X1             62.390     24.470   2.550 0.014258 *
X2              8.973      1.874   4.788 1.86e-05 ***
X3              9.888      1.742   5.677 9.39e-07 ***
X4             50.413     44.959   1.121 0.268109
X5             -0.951      2.649  -0.359 0.721231
X6             15.874     58.475   0.271 0.787269
K2$X71       -156.492     39.234  -3.989 0.000242 ***
K2$X72        -54.735     21.574  -2.537 0.014713 *
```

> anova(FM)

```
Analysis of Variance Table

Response: Y
          Df  Sum Sq Mean Sq   F value    Pr(>F)
X1         1 1005152 1005152   24.7723 9.925e-06 ***
X2         1 1278496 1278496   31.5089 1.167e-06 ***
X3         1 3442172 3442172   84.8334 6.407e-12 ***
X4         1   57862   57862    1.4260 0.2386707
X5         1   33032   33032    0.8141 0.3717186
X6         1       1       1 1.423e-05 0.9970070
K2$X7      2  726901  363450    8.9573 0.0005313 ***
Residuals 45 1825906   40576
```

Sum Contrasts:

```
#SETTING SUM CONTRASTS
contrasts(K2$X7)=contr.sum
FM=lm(Y~X1+X2+X3+X4+X5+X6+K2$X7)
contrasts(K2$X7)
model.matrix(FM)
summary(FM)
anova(FM)
```

> contrasts(K2$X7)

```
      [,1] [,2]
heavy    1    0
mod      0    1
none    -1   -1
```

> model.matrix(FM)

```
   (Intercept)   X1 X2  X3   X4 X5 X6 K2$X71 K2$X72
1            1  6.7 62  81 2.59 50  0      0      1
2            1  7.8 65 115 4.30 45  0      1      0
3            1  5.7 46  63 1.91 49  1      1      0
4            1  5.8 38  72 1.42 65  1      0      1
5            1  7.7 62  67 3.40 58  0      1      0

46           1  7.3 68  74 3.56 59  1     -1     -1
47           1  5.2 52  76 2.85 39  0     -1     -1
48           1  5.2 54  56 2.71 44  1     -1     -1
49           1  2.6 74  86 2.05 45  0     -1     -1
50           1  4.3  8 119 2.85 65  1     -1     -1
51           1  5.4 52  88 1.81 40  1     -1     -1
52           1  5.2 49  72 1.84 46  0     -1     -1
53           1  5.1 67  77 2.86 66  1     -1     -1
54           1  8.8 78  72 3.20 56  0     -1     -1
```

> summary(FM)

```
Call:
lm(formula = Y ~ X1 + X2 + X3 + X4 + X5 + X6 + K2$X7)
Residuals:
     Min       1Q   Median       3Q      Max
-285.360 -132.747   -9.998   89.477  790.119
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1039.353    243.484  -4.269 0.000100 ***
X1             62.390     24.470   2.550 0.014258 *
X2              8.973      1.874   4.788 1.86e-05 ***
X3              9.888      1.742   5.677 9.39e-07 ***
X4             50.413     44.959   1.121 0.268109
X5             -0.951      2.649  -0.359 0.721231
X6             15.874     58.475   0.271 0.787269
K2$X71        211.227     50.077   4.218 0.000118 ***
K2$X72       -101.757     38.753  -2.626 0.011772 *
```

> anova(FM)

```
Analysis of Variance Table
Response: Y
          Df  Sum Sq Mean Sq   F value    Pr(>F)
X1         1 1005152 1005152   24.7723 9.925e-06 ***
X2         1 1278496 1278496   31.5089 1.167e-06 ***
X3         1 3442172 3442172   84.8334 6.407e-12 ***
X4         1   57862   57862    1.4260 0.2386707
X5         1   33032   33032    0.8141 0.3717186
X6         1       1       1 1.423e-05 0.9970070
K2$X7      2  726901  363450    8.9573 0.0005313 ***
Residuals 45 1825906   40576
```