Example of Fitting Log-Linear Models

November 20, 2017

1 Fitting a Model to a 4-way Table

Our data consists of reponses by 607 individuals on a survey about different categories of government spending. All responses have the values 1, 2, or 3, with 1 meaning "Too Low", 2 means "About Right", and 3 means "Too High." The 4 categories of government spending are labelled Environment (E), Health (H), Cities (C), and Law (L). There are $3^4 = 81$ possible outcomes which can be summarized in a $3 \times 3 \times 3 \times 3$ table. The data appear on the website for the book and are discussed in Exercise 9.5.

We fit 3 successive log-linear models starting with main effects only (`fit1`), main effects plus all 2-way interactions (`fit2`), and added in all 3 way interactions (`fit3`). Based on these results, we fit a model which includes all main effects and some second order interactions (`fit2.2`).

The results are given below with some discussion afterwards.

```
> data = scan("data.txt")
Read 405 items
> data=matrix(data,ncol=5,byrow=T)
> apply(data,2,range)
     [,1] [,2] [,3] [,4] [,5]
[1,]    1    1    1    1    0
[2,]    3    3    3    3   90
> colnames(data) = c("Environment","Health","Cities","Law","Count")
> data1 = data.frame(data)
> for(j in 1:4) data1[,j] = as.factor(data1[,j])
> # seems to be necessary to do the columns separately
> fit1 = glm(Count ~ Environment + Health + Cities + Law,family=poisson,data=data1)
> summary(fit1)

Call:
glm(formula = Count ~ Environment + Health + Cities + Law, family = poisson,
    data = data1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.5722  -0.9335  -0.2423   0.7653   4.3062
```

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.77602    0.09860  38.296  < 2e-16 ***
Environment2 -1.24379    0.10032 -12.398  < 2e-16 ***
Environment3 -2.54048    0.17556 -14.470  < 2e-16 ***
Health2      -1.14579    0.09773 -11.724  < 2e-16 ***
Health3      -2.51770    0.17570 -14.329  < 2e-16 ***
Cities2       0.58280    0.10784   5.404 6.50e-08 ***
Cities3       0.55320    0.10842   5.102 3.35e-07 ***
Law2         -0.69315    0.08909  -7.781 7.22e-15 ***
Law3         -2.24601    0.16627 -13.508  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1370.46  on 80  degrees of freedom
Residual deviance:  124.34  on 72  degrees of freedom
AIC: 349.18

Number of Fisher Scoring iterations: 5

> fit2 = glm(Count ~ (Environment + Health + Cities + Law)^2,
+  family=poisson,data=data1)
> summary(fit2)

Call:
glm(formula = Count ~ (Environment + Health + Cities + Law)^2,
    family = poisson, data = data1)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.62783  -0.47669  -0.07639   0.33622   1.49280

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)           4.13581    0.11860  34.873  < 2e-16 ***
Environment2         -1.87605    0.25914  -7.240 4.50e-13 ***
Environment3         -3.35321    0.49608  -6.759 1.39e-11 ***
```

```
Health2             -1.71829    0.24541  -7.002 2.53e-12 ***
Health3             -3.32730    0.43918  -7.576 3.56e-14 ***
Cities2              0.35994    0.15091   2.385 0.017077 *
Cities3              0.18281    0.15530   1.177 0.239138
Law2                -1.22106    0.21381  -5.711 1.12e-08 ***
Law3                -2.82469    0.42119  -6.706 1.99e-11 ***
Environment2:Health2  0.30916    0.24086   1.284 0.199289
Environment3:Health2  0.72034    0.42138   1.709 0.087363 .
Environment2:Health3  1.41304    0.41061   3.441 0.000579 ***
Environment3:Health3  2.14249    0.55665   3.849 0.000119 ***
Environment2:Cities2  0.49541    0.29000   1.708 0.087576 .
Environment3:Cities2 -0.18961    0.62730  -0.302 0.762446
Environment2:Cities3  0.50830    0.29291   1.735 0.082676 .
Environment3:Cities3  1.20002    0.51769   2.318 0.020448 *
Environment2:Law2     0.18129    0.22199   0.817 0.414120
Environment3:Law2    -0.50679    0.43985  -1.152 0.249249
Environment2:Law3     0.13019    0.41894   0.311 0.755978
Environment3:Law3    -0.13285    0.63780  -0.208 0.835001
Health2:Cities2       0.26396    0.27714   0.952 0.340864
Health3:Cities2      -0.93284    0.53780  -1.735 0.082819 .
Health2:Cities3       0.28104    0.28199   0.997 0.318940
Health3:Cities3      -0.18648    0.45472  -0.410 0.681730
Health2:Law2          0.72344    0.20826   3.474 0.000513 ***
Health3:Law2          0.83749    0.42130   1.988 0.046826 *
Health2:Law3         -0.06301    0.47762  -0.132 0.895044
Health3:Law3          1.87407    0.50792   3.690 0.000225 ***
Cities2:Law2          0.42931    0.24677   1.740 0.081907 .
Cities3:Law2          0.30279    0.25359   1.194 0.232469
Cities2:Law3         -0.20576    0.54351  -0.379 0.705005
Cities3:Law3          0.87351    0.46044   1.897 0.057811 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1370.458  on 80  degrees of freedom
Residual deviance:   31.669  on 48  degrees of freedom
AIC: 304.5

Number of Fisher Scoring iterations: 5
```

```
> fit3 = glm(Count ~ (Environment + Health + Cities + Law)^3,
+  family=poisson,data=data1)
Warning message:
glm.fit: fitted rates numerically 0 occurred
> summary(fit3)

Call:
glm(formula = Count ~ (Environment + Health + Cities + Law)^3,
    family = poisson, data = data1)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.08934  -0.11186  -0.00002   0.11730   0.73809

Coefficients:
                               Estimate Std. Error z value Pr(>|z|)
(Intercept)                   4.141e+00  1.254e-01  33.020  < 2e-16 ***
Environment2                 -1.827e+00  3.252e-01  -5.618 1.93e-08 ***
Environment3                 -3.043e+00  5.908e-01  -5.150 2.60e-07 ***
Health2                      -1.830e+00  3.283e-01  -5.573 2.50e-08 ***
Health3                      -3.436e+00  6.704e-01  -5.125 2.97e-07 ***
Cities2                       3.532e-01  1.631e-01   2.165   0.0304 *
Cities3                       1.573e-01  1.701e-01   0.924   0.3553
Law2                         -1.327e+00  2.687e-01  -4.938 7.90e-07 ***
Law3                         -2.653e+00  4.839e-01  -5.482 4.21e-08 ***
Environment2:Health2          1.625e-01  6.642e-01   0.245   0.8067
Environment3:Health2          7.311e-01  1.200e+00   0.609   0.5425
Environment2:Health3          1.097e+00  1.009e+00   1.087   0.2771
Environment3:Health3          2.337e+00  1.335e+00   1.751   0.0800 .
Environment2:Cities2          3.913e-01  3.940e-01   0.993   0.3206
Environment3:Cities2         -6.708e-01  8.783e-01  -0.764   0.4450
Environment2:Cities3          5.534e-01  3.988e-01   1.388   0.1652
Environment3:Cities3          9.207e-01  6.874e-01   1.340   0.1804
Environment2:Law2             2.117e-01  5.773e-01   0.367   0.7139
Environment3:Law2            -2.139e+01  1.453e+04  -0.001   0.9988
Environment2:Law3            -2.256e-01  1.240e+00  -0.182   0.8557
Environment3:Law3            -2.071e+01  7.635e+03  -0.003   0.9978
Health2:Cities2               4.625e-01  3.944e-01   1.173   0.2410
Health3:Cities2              -5.460e-01  9.565e-01  -0.571   0.5681
```

```
Health2:Cities3                4.808e-01  4.067e-01   1.182   0.2372
Health3:Cities3               -5.957e-01  9.631e-01  -0.618   0.5363
Health2:Law2                   1.083e+00  5.024e-01   2.156   0.0311 *
Health3:Law2                   1.500e+00  8.974e-01   1.671   0.0947 .
Health2:Law3                  -1.918e+01  9.446e+03  -0.002   0.9984
Health3:Law3                   2.398e+00  1.071e+00   2.239   0.0252 *
Cities2:Law2                   5.669e-01  3.226e-01   1.757   0.0789 .
Cities3:Law2                   4.767e-01  3.364e-01   1.417   0.1565
Cities2:Law3                  -5.572e-01  7.027e-01  -0.793   0.4278
Cities3:Law3                   7.479e-01  5.766e-01   1.297   0.1946
Environment2:Health2:Cities2 -1.724e-01  7.193e-01  -0.240   0.8106
Environment3:Health2:Cities2  4.521e-01  1.504e+00   0.301   0.7637
Environment2:Health3:Cities2  6.724e-01  1.230e+00   0.547   0.5846
Environment3:Health3:Cities2 -1.949e+01  1.350e+04  -0.001   0.9988
Environment2:Health2:Cities3 -9.485e-02  7.403e-01  -0.128   0.8981
Environment3:Health2:Cities3 -1.276e-01  1.348e+00  -0.095   0.9246
Environment2:Health3:Cities3  1.217e+00  1.112e+00   1.094   0.2740
Environment3:Health3:Cities3 -4.823e-01  1.836e+00  -0.263   0.7928
Environment2:Health2:Law2     5.134e-01  5.082e-01   1.010   0.3124
Environment3:Health2:Law2     5.656e-02  1.015e+00   0.056   0.9556
Environment2:Health3:Law2    -8.817e-01  1.010e+00  -0.873   0.3828
Environment3:Health3:Law2     3.972e-01  1.674e+00   0.237   0.8125
Environment2:Health2:Law3     4.775e-01  1.099e+00   0.434   0.6641
Environment3:Health2:Law3    -1.805e+01  6.693e+03  -0.003   0.9978
Environment2:Health3:Law3    -3.357e-01  1.233e+00  -0.272   0.7853
Environment3:Health3:Law3     2.046e+00  1.977e+00   1.035   0.3008
Environment2:Cities2:Law2     7.886e-02  6.447e-01   0.122   0.9026
Environment3:Cities2:Law2     2.117e+01  1.453e+04   0.001   0.9988
Environment2:Cities3:Law2    -3.892e-01  6.658e-01  -0.585   0.5588
Environment3:Cities3:Law2     2.084e+01  1.453e+04   0.001   0.9989
Environment2:Cities2:Law3     7.037e-01  1.441e+00   0.488   0.6253
Environment3:Cities2:Law3     3.266e+00  1.466e+04   0.000   0.9998
Environment2:Cities3:Law3     2.190e-01  1.347e+00   0.163   0.8708
Environment3:Cities3:Law3     2.043e+01  7.635e+03   0.003   0.9979
Health2:Cities2:Law2          -5.844e-01  5.861e-01  -0.997   0.3187
Health3:Cities2:Law2          -1.825e+00  1.447e+00  -1.262   0.2071
Health2:Cities3:Law2          -5.786e-01  6.080e-01  -0.952   0.3413
Health3:Cities3:Law2           1.918e-01  1.150e+00   0.167   0.8675
Health2:Cities2:Law3           1.879e+01  9.446e+03   0.002   0.9984
Health3:Cities2:Law3           3.365e-02  1.506e+00   0.022   0.9822
```

```
Health2:Cities3:Law3         1.936e+01  9.446e+03   0.002   0.9984
Health3:Cities3:Law3        -1.168e+00  1.486e+00  -0.786   0.4318
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1370.4575  on 80  degrees of freedom
Residual deviance:    8.5237  on 16  degrees of freedom
AIC: 345.36

Number of Fisher Scoring iterations: 19

> fit2.2 = glm(Count ~ Environment + Health + Cities + Law +
+   Environment*Health + Health*Law,family=poisson,data=data1)
> summary(fit2.2)

Call:
glm(formula = Count ~ Environment + Health + Cities + Law + Environment *
    Health + Health * Law, family = poisson, data = data1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9483  -0.7317  -0.2197   0.4454   2.0474

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)           3.90715    0.09960  39.227  < 2e-16 ***
Environment2         -1.41968    0.12382 -11.466  < 2e-16 ***
Environment3         -2.92376    0.24195 -12.084  < 2e-16 ***
Health2              -1.51455    0.14586 -10.384  < 2e-16 ***
Health3              -3.69507    0.34408 -10.739  < 2e-16 ***
Cities2               0.58280    0.10784   5.404 6.51e-08 ***
Cities3               0.55320    0.10842   5.102 3.35e-07 ***
Law2                 -0.91458    0.10942  -8.359  < 2e-16 ***
Law3                 -2.45788    0.20839 -11.795  < 2e-16 ***
Environment2:Health2  0.36231    0.23698   1.529 0.126298
Environment3:Health2  0.67247    0.41118   1.635 0.101951
Environment2:Health3  1.41968    0.39773   3.569 0.000358 ***
Environment3:Health3  2.23061    0.52230   4.271 1.95e-05 ***
```

```
Health2:Law2          0.73226    0.20622   3.551 0.000384 ***
Health3:Law2          0.76043    0.40832   1.862 0.062557 .
Health2:Law3         -0.02703    0.47326  -0.057 0.954457
Health3:Law3          2.01605    0.47535   4.241 2.22e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1370.458  on 80  degrees of freedom
Residual deviance:   71.422  on 64  degrees of freedom
AIC: 312.26

Number of Fisher Scoring iterations: 5
```

We note the following items in the analysis:

1. Note that the first level of each factor is set as the baseline, so it is automatically 0 in all effects. That is why we see Law2 and Law3 terms but no Law1 terms.

2. In `fit1`, all terms are significant, suggesting we can reject a null hypothesis of independence. The LRT test for independence has a chi-squared value of $1370.46 - 124.34$ with $80 - 72 = 8$ degrees of freedom and clearly has a p-value of basically 0.

3. In `fit2`, the main effects of C "lose" significance, and most of the interaction terms of C with other variables are not "very" significant.

4. In `fit3`, we see most of the second order effects with are not significant, and all of the third order terms are not significant. Also, there was a warning "`fitted rates numerically 0 occurred`", all of which suggests the third order model is unnecessary.

Based on these observations, we did the `fit2.2`, which includes all main effects and only the E*H and H*L interactions. Note that this model corresponds to independence of C with (E,H,L), and conditional independence of E,L given H. Clearly, this is a much simpler model than the full second order model, and would be easy to summarize as a simple graphical model (if you know what those are). Also, most of the terms are "very" significant.

I would hesitate to drop selective terms from different factor levels in second order interactions. So, based on the significance patterns, I conclude this is a pretty good model.

We perhaps should do a full, formal test of the null hypothesis that the fit2.2 model is valid vs. the more general fit2 model. The LRT chi-squared is $71.422 - 31.669 = 39.75$ with $64 - 48 = 16$ d.f. The p-value is .00085, so we can reject the null hypothesis if we want, but I like the model for being more parsimonious.

This analysis shows the power of log-linear models. We can quickly fit them and use the results to arrive at a parsimonious model, and often interpret the results in terms of independence or conditional independence.

2 Fitting Ordinal Predictors Using Linear Scores