Stat 8053, Fall 2013: Models for Contingency Tables (Faraway, 4)

Association in a 2×2 table

Given a 2×2 table,

y_{11}	y_{12}
y_{21}	y_{22}

association is estimated by the odds-ratio, $(y_{11}/y_{21})/(y_{12}/y_{22}) = (y_{11}y_{22})/(y_{12}y_{21})$, or perhaps by its logarithm. The odds-ratio equals zero for independence, and is positive for dependence.

```
data <- data.frame(
   sex=c("m", "m", "f", "f", "m", "m", "f", "f"),
   citizen=c("yes", "no", "yes", "no", "yes", "no", "yes", "no"),
   type = c("I", "I", "I", "I", "IV", "IV", "IV", "IV"),
   y= c(219, 209, 55, 54, 71, 122, 54, 105)
   )
ftable(xtabs(y ~ sex + citizen + type, data))
```

```
            type   I  IV
sex citizen
f   no            54 105
    yes           55  54
m   no           209 122
    yes          219  71
```

```
odds_ratio <- function(a) (a[1,1]*a[2,2])/(a[1,2]*a[2,1])
```

Complete Independence

```
p1 <- glm(y ~ sex + citizen + type, poisson, data)
ftable(t1 <- xtabs(round(predict(p1, type="response"), 2) ~ sex + citizen + type, data))
```

```
            type      I     IV
sex citizen
f   no            89.23  58.49
    yes           72.66  47.63
m   no           206.76 135.53
    yes          168.36 110.36
```

```
odds_ratio(t1[1, ,])  # table for females only
```

```
[1] 1
```

```
odds_ratio(apply(t1, c(2, 3), sum)) # add over sex
```

```
[1] 1
```

Independence (in the fitted values) holds for any conditional (e.g., $\hat{y}_{ij1}$) or any marginal (e.g. $\hat{y}_{ij+}$) table .

Joint Independence

Looking at one of the three versions of this model:

```
p2 <- glm(y ~ sex*citizen + type, poisson, data)
(f2 <- ftable(t2 <- xtabs(round(predict(p2, type="response"), 2) ~ sex + citizen + type, data)))
```

```
            type      I     IV
sex citizen
f   no            96.04  62.96
    yes           65.84  43.16
m   no           199.94 131.06
    yes          175.17 114.83
```

```
chisq.test(f2)
```

```
        Pearson's Chi-squared test

data:  f2
X-squared = 0, df = 3, p-value = 1
```

Independence holds for the 4×2 table with "rows" `sex*citizen` and "columns" `type`. It holds in a marginal table adding over either `sex` or `citizen`:

```
odds_ratio(apply(t2, c(1, 3), sum))
```

```
[1] 0.9999
```

```
odds_ratio(apply(t2, c(2, 3), sum))
```

```
[1] 1
```

but independence does not hold adding over `type`

```
odds_ratio(apply(t2, c(1, 2), sum))
```

```
[1] 1.278
```

Conditional Independence

Again, one of three versions.

```
p3 <- glm(y ~ sex*citizen + sex*type, poisson, data)
ftable(t3 <- xtabs(round(predict(p3, type="response"), 2) ~ sex + citizen + type, data))
```

```
            type      I     IV
sex citizen
f   no            64.67  94.33
    yes           44.33  64.67
m   no           228.13 102.87
    yes          199.87  90.13
```

According to this model, (`citizen` $\perp\!\!\!\perp$ `type`)|`sex`:

```
odds_ratio(t3[1, , ])
```

```
[1] 1
```

```
odds_ratio(t3[2, , ])
```

```
[1] 1
```

Averaging over `sex` results in non-zero X^2:

```
odds_ratio(apply(t3, c(2, 3), sum))
```

```
[1] 0.9412
```

Homogeneous Association

This is the model of no three-factor interaction:

```
p4 <- glm(y ~ (sex + citizen + type)^2, poisson, data)
round(ftable(t4 <- xtabs(predict(p4, type="response") ~ sex + citizen + type, data)), 2)
```

```
            type      I     IV
sex citizen
f   no            54.99 104.01
    yes           54.01  54.99
m   no           208.01 122.99
    yes          219.99  70.01
```

There are no independence relationships here, but rather constant conditional association

```
odds_ratio(t4[1, , ])
```

```
[1] 0.5383
```

```
odds_ratio(t4[2, , ])
```

```
[1] 0.5383
```

The same would be true for any other two-way tables.

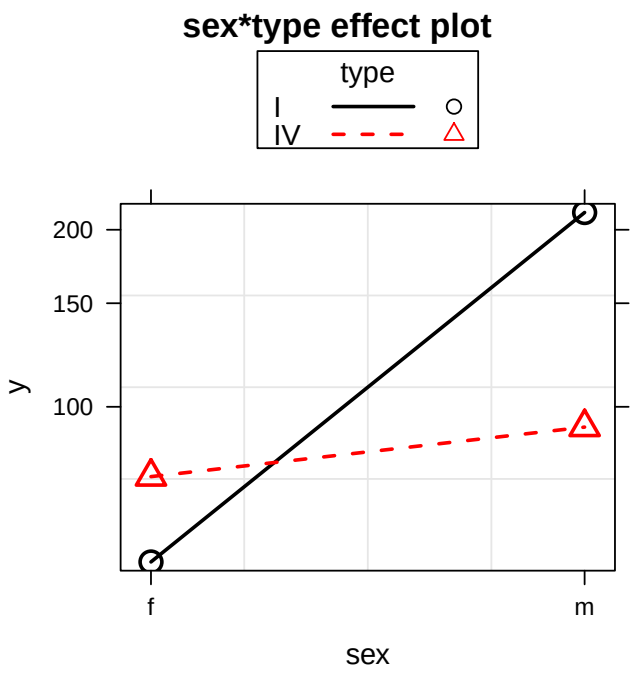
Checking All the Models at Once

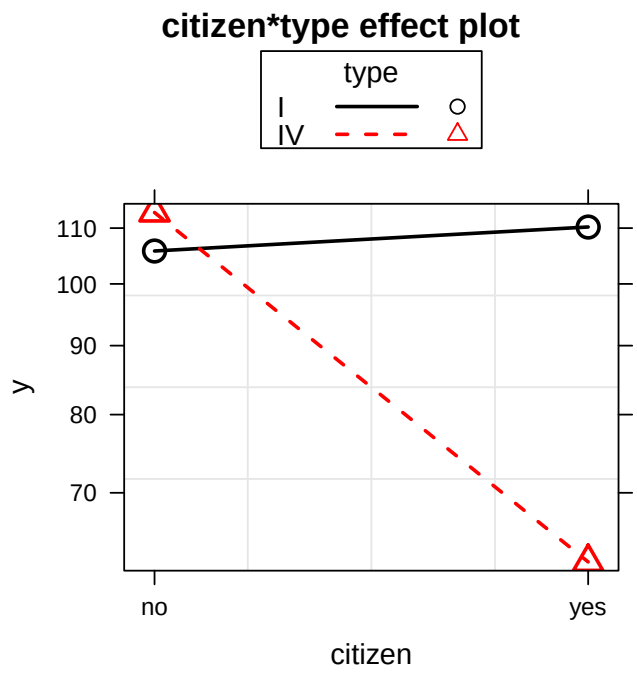
```
m1 <- glm(y ~ sex * citizen * type, poisson, data)
library(alr4)
Anova(m1)
```

```
Analysis of Deviance Table (Type II tests)

Response: y
                 LR Chisq Df Pr(>Chisq)
sex                 144.1  1    < 2e-16
citizen               9.3  1     0.0023
type                 38.8  1    4.7e-10
sex:citizen           0.2  1     0.6333
sex:type             59.3  1    1.4e-14
citizen:type         18.3  1    1.8e-05
sex:citizen:type      0.1  1     0.7589
```

```
m2 <- update(m1, ~ . - citizen:sex:type - citizen:sex)
plot(allEffects(m2), multiline=TRUE, grid=TRUE)
```





The Full Four-way Table

```
head(AMS1 <- reshape(AMSsurvey, varying=c("count", "count11"), v.names="y",
               direction="long", times=c("08-09", "11-12"), timevar="year"))
```

```
         type    sex citizen  year   y id
1.08-09 I(Pu)   Male      US 08-09 132  1
2.08-09 I(Pu) Female      US 08-09  35  2
3.08-09 I(Pr)   Male      US 08-09  87  3
4.08-09 I(Pr) Female      US 08-09  20  4
5.08-09    II   Male      US 08-09  96  5
6.08-09    II Female      US 08-09  47  6
```

```
m3 <- glm(y ~ (type + sex + citizen + year)^4, poisson, AMS1)
Anova(m3)
```

```
Analysis of Deviance Table (Type II tests)
```

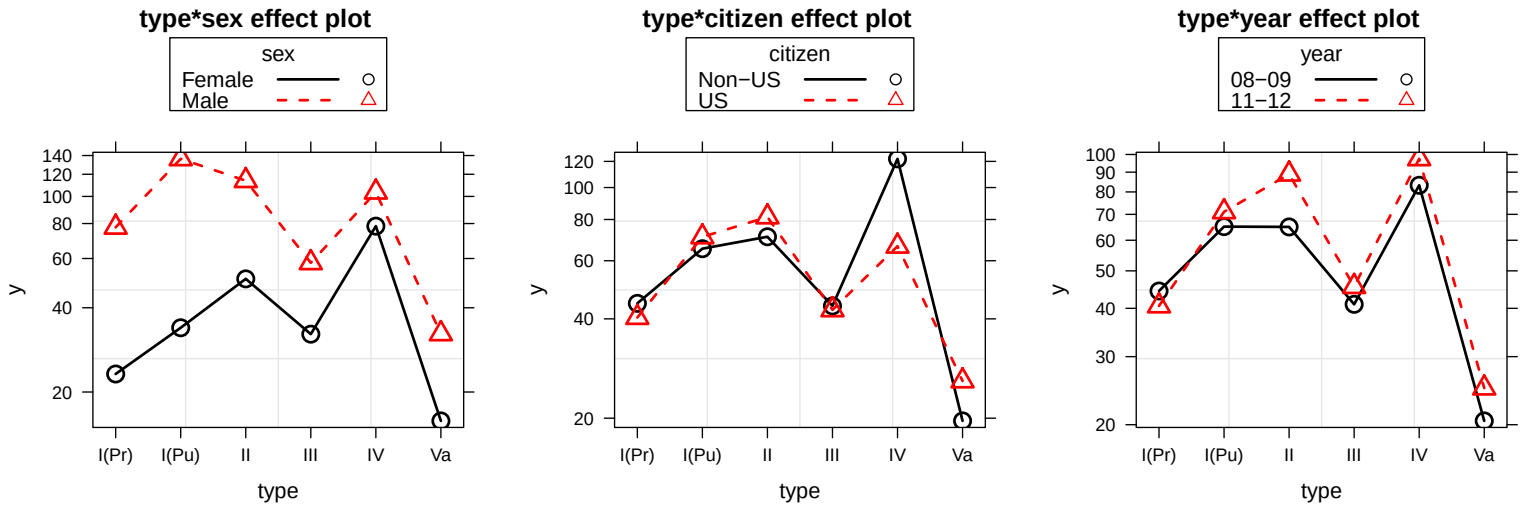
```
Response: y
                      LR Chisq Df Pr(>Chisq)
type                       534  5    < 2e-16
sex                        449  1    < 2e-16
citizen                      7  1    0.00681
year                        15  1    0.00011
type:sex                   104  5    < 2e-16
type:citizen                65  5    1.1e-12
type:year                   12  5    0.03886
sex:citizen                  2  1    0.12841
sex:year                     2  1    0.12185
citizen:year                 0  1    0.51845
type:sex:citizen             3  5    0.70774
type:sex:year                8  5    0.18084
type:citizen:year            4  5    0.50807
sex:citizen:year             0  1    0.68396
type:sex:citizen:year        1  5    0.92579
```

```
Anova(m4 <- update(m3, ~ type*(sex + citizen + year)))
```

```
Analysis of Deviance Table (Type II tests)

Response: y
             LR Chisq Df Pr(>Chisq)
type              534  5    < 2e-16
sex               449  1    < 2e-16
citizen             7  1    0.00681
year               15  1    0.00011
type:sex          108  5    < 2e-16
type:citizen       70  5    1.2e-13
type:year          11  5    0.05056
```

```
plot(allEffects(m4), multiline=TRUE, grid=TRUE, rows=1, cols=3)
```



Log-linear vs. Logistic

View `year` as a response variable.

```
logistic2 <- glm(cbind(count,count11) ~ type + sex + citizen, binomial, AMSsurvey)
loglinear2 <- glm(y ~ year + year:(type+sex+citizen) + type*sex*citizen, poisson, AMS1)

compareCoefs(logistic2, loglinear2)
```

```
Call:
1:"glm(formula = cbind(count, count11) ~ type + sex + citizen, family = binomial, data = AMSsurvey)"
2:c("glm(formula = y ~ year + year:(type + sex + citizen) + type * sex * citizen, family = poisson, ",
 "    data = AMS1)")
                            Est. 1    SE 1  Est. 2    SE 2
(Intercept)                 0.2092  0.1217  3.3378  0.1503
typeI(Pu)                  -0.1712  0.1258  0.0987  0.1986
typeII                     -0.4102  0.1270  0.5269  0.1815
typeIII                    -0.2096  0.1455  0.2029  0.1972
typeIV                     -0.2780  0.1247  1.3277  0.1661
typeVa                     -0.2924  0.1754 -0.7061  0.2485
sexMale                    -0.1245  0.0796  1.0919  0.1648
citizenUS                  -0.0507  0.0734 -0.2172  0.2110
year11-12                                  -0.2092  0.1217
year11-12:typeI(Pu)                         0.1712  0.1258
year11-12:typeII                            0.4102  0.1270
year11-12:typeIII                           0.2096  0.1455
year11-12:typeIV                            0.2780  0.1247
year11-12:typeVa                            0.2924  0.1754
year11-12:sexMale                           0.1245  0.0796
year11-12:citizenUS                         0.0507  0.0734
typeI(Pu):sexMale                           0.3177  0.2145
typeII:sexMale                             -0.5027  0.2005
typeIII:sexMale                            -0.6540  0.2216
typeIV:sexMale                             -0.9351  0.1845
typeVa:sexMale                             -0.5186  0.2803
typeI(Pu):citizenUS                         0.3986  0.2705
typeII:citizenUS                            0.1307  0.2508
typeIII:citizenUS                           0.0517  0.2731
typeIV:citizenUS                           -0.5116  0.2391
typeVa:citizenUS                            0.3785  0.3263
sexMale:citizenUS                           0.1218  0.2373
typeI(Pu):sexMale:citizenUS                -0.2787  0.3055
typeII:sexMale:citizenUS                    0.1610  0.2907
typeIII:sexMale:citizenUS                   0.0509  0.3236
typeIV:sexMale:citizenUS                    0.0373  0.2827
typeVa:sexMale:citizenUS                    0.0120  0.3882
```