

STA303/1002 - Methods of Data Analysis II

(Week 10 lecture note)

Wei (Becky) Lin

Mar 14/16, 2017



Topics learned last week

- Show that the Binomial log-likelihood equals Bernoulli log-likelihood up to a constant
- Deviance for Bernoulli model

$$D = -2\{\ell(\hat{\beta}) - \ell_S(\tilde{\beta})\} = 2 \sum_{i=1}^n [Y_i \log \frac{Y_i}{\hat{\pi}_i} + (1 - Y_i) \log \frac{1 - Y_i}{1 - \hat{\pi}_i}]$$

log-likelihood for saturated model
for current model

- Deviance for Binomial model

$$D = -2\{\ell(\hat{\beta}) - \ell_S(\tilde{\beta})\} = 2 \sum_{i=1}^c \left\{ y_i \log \frac{y_i}{\hat{y}_i} + (n_i - y_i) \log \frac{n_i - y_i}{n_i - \hat{y}_i} \right\}, \quad \hat{y}_i = n_i \hat{\pi}_i$$

- Null and residual deviance in summary output
- Likelihood ratio test (LRT): inference for several $\beta_k = 0$ simultaneously

$$G^2 = -2(\ell_R - \ell_F) = \text{Deviance}_R - \text{Deviance}_F \sim_{H_0} \chi^2_{df_R - df_F}$$

- Goodness-of-fit
 - for bernoulli data: Hosmer-Lemeshow Test.
 - For binomial data:
 - Pearson chisquare test
 - Drop in deviance (LRT)

*when $F = \text{saturated}$
 $R = \text{current model}$*
 $\Rightarrow G^2 = \text{Deviance}_c$
 $\approx \chi^2_{n-p}$

Topics learned last week

↳ Pseudo- R^2

- R^2 statistics for logistic regression

- McFadden's R^2

$$R_{McF}^2 = 1 - \frac{\ln(L_m)}{\ln(L_0)}$$

- The Cox and Snell R^2 is

$$R_{CS}^2 = 1 - (L_0/L_m)^{2/n}$$

- Overdispersion (the variance exceeds the variance of the binomial distribution) or underdispersion

- Overdispersion if $\hat{\phi} = \frac{\text{sum of squared Pearson residual}}{n-p} >> 1$

- Underdispersion if $\hat{\phi} = \frac{\text{sum of squared Pearson residual}}{n-p} << 1$

- Why it occurs: (1) correlated data; (2) clustered data

- Comparing nested models in presence of overdispersion

- How to get summary which accounts for dispersion:

- `summary(fit, dispersion=phi.hat)`

- `glm(cbind(y,n-y)~x,data=dd, family=quasibinomial())`

✖ • How to use logistic regression to do Classification

- TPR (sensitivity), TNR(specificity)
- ROC: plot of TPR vs 1-TNR
- AUC

$$\chi^2 = \sum_{i=1}^n r_{i,p}^2$$

$$\hat{\phi} = \frac{\chi^2}{n-p}$$

$$\tilde{F} = F / \hat{\phi}_F = \frac{df_R - df_F}{df_R + df_F} / \hat{\phi}_F$$

Poisson distribution

- Y : counts of events that occur randomly over fixed interval of time or space.
 - No upper limit for the number of events $\gamma = 0, 1, 2, \dots$
 - Approximates $\text{Bin}(n, \pi)$ distribution when n is large and π is small. Then $\lambda = n\pi$.
 - Approaches Normality as λ increases.

Probability mass function:

$$P(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!}, \quad y = 0, 1, 2, \dots$$

*

- $E(Y) = \text{Var}(Y) = \lambda$
- If $Y_i \sim \text{iid Pois}(\lambda_i)$, $i=1, 2, \dots$, then $\sum_i Y_i \sim \text{Pois}(\sum_i \lambda_i)$

$$\Rightarrow Y_i \sim \text{iid Poisson}(1) \Rightarrow \sum_{i=1}^k Y_i \sim \text{Poisson}(k)$$

$$\Rightarrow \frac{\sum_{i=1}^k Y_i - k}{\sqrt{k}} = \text{CLT} \rightarrow N(0, 1)$$

Poisson Regression

- Model: $Y_i \sim iid Pois(\lambda_i)$

$$\log(\lambda_i) = \log(\mu_i) = X_i\beta = \eta$$

- also called a **log-linear model**

$$\log\{E(Y_i)\} = \log\{\lambda_i\} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

- 2 ways** }
 to interpret }
 β }
 - β_i : increases $\log E(Y)$ by β_i as one unit change in predictor X_i while fix all the others predictors.

$$E(Y_i) = \lambda_i = \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)$$

- β_i : increasing X_i by one unit with other predictors held constant, increases mean of response by a factor e^{β_i}

- Show Log-likelihood is

$$\ell(\beta) = \sum_{i=1}^n \{y_i X_i \beta - \exp(X_i \beta) + \text{constant}\}$$

Proof: $Y_i \sim Poisson(\lambda_i) \Rightarrow p(Y_i=y_i) = e^{-\lambda_i} \frac{\lambda_i^{y_i}}{y_i!}, \lambda_i = e^{X_i \beta}$

For Y_1, \dots, Y_n : $\Rightarrow \ell(\beta) = \log L(\beta) = \log P(Y_i=y_i) = y_i \log(\lambda_i) - \lambda_i + C$

$$\Rightarrow \ell(\beta) = \log L(\beta) = \log \prod_{i=1}^n P(Y_i=y_i) = \sum_{i=1}^n P(Y_i=y_i)$$

$$= \sum_{i=1}^n \{y_i \log(\lambda_i) - \lambda_i + C\} = \sum_{i=1}^n \{y_i X_i \beta - e^{X_i \beta} + C\}$$

| Y | X |
|----------|----------|
| y_1 | x_1 |
| y_2 | x_2 |
| \vdots | \vdots |
| y_n | x_n |

$\rightarrow \lambda_1 = x_1 \beta$
 $\rightarrow \lambda_2 = x_2 \beta$
 \vdots
 $\rightarrow \lambda_n = x_n \beta$

Poisson Regression: Goodness-of-fit

- Estimation of parameter: MLE by IRLS algorithm.

$$\log(\hat{\lambda}_i) = X_i \hat{\beta} \rightarrow \hat{\lambda}_i = \exp\{X_i \hat{\beta}\}$$

- Pearson X^2

$$X^2 = \sum_{i=1}^n \frac{(y_i - \hat{\lambda}_i)^2}{\hat{\lambda}_i} = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

$O_i = \text{observation}$
 $= Y_i$

$E_i = \hat{Y}_i$

= expected
counts

- the numerator is the squared difference between observed and fitted values
- the denominator is the variance of the observed value $V(Y_i) = E(Y_i) = \lambda_i$
- The X^2 statistic has the same form for Poisson and binomial data, namely a sum of squared observed minus expected over expected

- Deviance

$$D = 2 \sum_{i=1}^n \left(y_i \log \frac{y_i}{\hat{\lambda}_i} - (y_i - \hat{\lambda}_i) \right)$$

- A measure of discrepancy between observed and fitted values is the deviance.
- the first term is identical to the binomial deviance, representing twice a sum of observed times log of observed over fitted
- the second term, a sum of differences between observed and fitted values
- one advantage of the deviance over Pearson X^2 is that it can be used to compare nested models, as what we have in logistic regression model.

Poisson Regression: Deviance

$$D = 2 \sum_{i=1}^n \left(y_i \log \frac{y_i}{\hat{\lambda}_i} - (y_i - \hat{\lambda}_i) \right)$$

For saturated model
 n obs y_i and n unkns
 $\downarrow \beta \Rightarrow \hat{\lambda}_i = y_i$

Recall:

- $D = -2 \{ l_c(\hat{\beta}) - l_s(\tilde{\beta}) \}, \quad \hat{\lambda}_i = e^{x_i \hat{\beta}}, \quad \tilde{\lambda}_i = e^{x_i \tilde{\beta}} = y_i$

- $l_c(\beta) = \sum_{i=1}^n \{ y_i \log \lambda_i - \lambda_i + c \}$

$$\Rightarrow ① \quad l_c(\hat{\beta}) = \sum_{i=1}^n \{ y_i \log \hat{\lambda}_i - \hat{\lambda}_i + c \}$$

$$② \quad l_s(\tilde{\beta}) = \sum_{i=1}^n \{ y_i \log y_i - y_i + c \}$$

$$① - ② : \quad l_c(\hat{\beta}) - l_s(\tilde{\beta}) = \sum_{i=1}^n \{ y_i \log \frac{\hat{\lambda}_i}{y_i} + y_i - \hat{\lambda}_i \}$$

$$\Rightarrow D = -2 \{ l_c(\hat{\beta}) - l_s(\tilde{\beta}) \}$$

$$= 2 \sum_{i=1}^n \{ y_i \log \frac{y_i}{\hat{\lambda}_i} - (y_i - \hat{\lambda}_i) \}$$

c for current model
 s for saturated model

Poisson Regression: model Assessment

- Can plot $\log(y)$ vs x to see if linear model seems appropriate
 - use $\log(y+k)$ if zeroes, $k>0$

- Pearson/Deviance residuals for outliers
- Deviance GoF test

- Compare fitted model to saturated (ANOVA)

- Large p-value

- model adequate

- or underpowered (not enough data)

- Small p-value

- model incorrect: missing predictors, polynomial/interaction terms...

- Poisson model inappropriate

- Variance increases faster than linear with mean

- Severe outliers present.

From summary output.

\Rightarrow residual deviance = #
on $n-p$ d.f.

\Rightarrow H_0 : current model
 H_a : saturated M

\leftarrow p.val < 0.05, Rej H_0

\Rightarrow Test stat

$= D_c - D_s / H_0$

$= G^2$

$\sim \chi^2_{n-p}$

Poisson Regression: inference about individual β_j - Wald Test

- Same as for binomial response.
- Test $H_0: \beta_j = 0$ vs $H_a: \beta_j \neq 0$
- Test statistic

$$Z^* = \frac{b_j - 0}{\text{se}(b_j)} \mid_{H_0} \sim N(0, 1)$$

- Approximate variance $\text{se}(b)$

b is a vector

$\text{Var}(b)$: var-cov matrix $= s^2(b) = \left(-\frac{\partial^2 \ell(\beta)}{\partial \beta_j \partial \beta'_j} \Big|_{\beta=b} \right)^{-1}$: pxp matrix, the diagonal elements gives $\hat{\sigma}^2(b_j), j=1..p$

- Approximate distribution of Z^*

- $Z^* \stackrel{H_0}{\sim} N(0, 1)$, alternatively, $(Z^*)^2 \sim \chi_1^2$
- reject H_0 if $|z^*| > Z_{1-\alpha/2}$
- CI for β_j : $b_j \pm Z_{1-\alpha/2} \text{se}(b_j)$

Example: Galapagos Plants (Faraway ch3)

- Study the number of plant species on 30 Galapagos islands as function of geographic variables

```
library(faraway)
data(gala)
gala <- gala[, -2] # drop 2nd column from data
str(gala)

## 'data.frame': 30 obs. of 6 variables:
## $ Species : num 58 31 3 25 2 18 24 10 8 2 ...
X1 ## $ Area   : num 25.09 1.24 0.21 0.1 0.05 ...
X2 ## $ Elevation: num 346 109 114 46 77 119 93 168 71 112 ...
X3 ## $ Nearest : num 0.6 0.6 2.8 1.9 1.9 8 6 34.1 0.4 2.6 ...
X4 ## $ Scrub  : num 0.6 26.3 58.7 47.4 1.9 ...
X5 ## $ Adjacent: num 1.84 572.33 0.78 0.18 903.82 ...

dim(gala)

## [1] 30 6

head(gala)

##          Species  Area Elevation Nearest  Scrub  Adjacent
## Baltra      58 25.09      346     0.6    0.6     1.84
## Bartolome   31  1.24      109     0.6   26.3    572.33
## Caldwell     3  0.21      114     2.8   58.7     0.78
## Champion    25  0.10       46     1.9   47.4     0.18
## Coamano     2  0.05       77     1.9    1.9    903.82
## Daphne.Major 18  0.34      119     8.0    8.0     1.84
```

Example: Galapagos Plants (Faraway ch3)

- Study the number of plant species on 30 Galapagos islands as function of geographic variables

```
mod1 <- glm(Species ~ ., family = poisson, data = gala)
summary(mod1)
```

↳ put all x's in the model.

```
##  
## Call:  
## glm(formula = Species ~ ., family = poisson, data = gala)  
##  
## Deviance Residuals:  
##      Min        1Q     Median        3Q       Max  
## -8.2752  -4.4966  -0.9443   1.9168  10.1849  
##  
## Coefficients:  
##             Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 3.155e+00 5.175e-02 60.963 < 2e-16 ***  
## Area        -5.799e-04 2.627e-05 -22.074 < 2e-16 ***  
## Elevation    3.541e-03 8.741e-05 40.507 < 2e-16 ***  
## Nearest      8.826e-03 1.821e-03  4.846 1.26e-06 ***  
## Scrub        -5.709e-03 6.256e-04 -9.126 < 2e-16 ***  
## Adjacent     -6.630e-04 2.933e-05 -22.608 < 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: 3510.73 on 29 degrees of freedom  
## Residual deviance: 716.85 on 24 degrees of freedom  
## AIC: 889.68
```

Deviance for model with only intercept, $\Rightarrow d.f. = n - 1 = 30 - 1 = 29$ sample size

Residual deviance: 716.85 on 24 degrees of freedom

AIC: 889.68 $\Rightarrow d.f. = n - p = 30 - 6 = 24$

Example: Galapagos Plants

H_0 : Current Model
 H_a : Saturated model

Goodness of Fit Tests

• Residual Deviance = $D_c - D_s = D_c - 0 = D_c$, $G^2 = D_c - D_s = D_c \sim \chi_{n-p}^2$

- Test H_0 : current model is a good fit H_a : Saturated model
- Distribution poorly approximated by a χ_{n-p}^2 , $D_{H_0} \sim \chi_{n-p}^2$

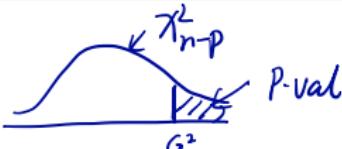
① LRT : $G^2 = D_c - 0 \sim \chi_{n-p}^2$

```
pchisq(mod1$deviance, mod1$df.residual, lower.tail = FALSE)
```

$\frac{\text{Obs. } G^2}{n-p}$

```
## [1] 7.073157e-136
```

Apply LRT



② Pearson $\chi^2 = \sum_{i=1}^n \frac{(O_i - \hat{E}_i)^2}{\hat{E}_i}$

- Test H_0 : current model is a good fit H_a : Saturated model
- Distribution better approximated by χ_{n-p}^2

```
pchisq(sum(residuals(mod1, type = "pearson")^2), mod1$df.residual, lower.tail = FALSE)
```

```
## [1] 2.18719e-145
```

- Better approximation of χ_{n-p}^2 when Poisson approaches Normal
- Can look more into model diagnostics to find possible reasons for poor fit

Poisson Model Diagnostics: Residual Plots

四 3 types residual in Poisson model

- ① • Response residual: $r_i = y_i - \hat{\lambda}_i$
- of limited use: variance is function of the expected value

- ② • Pearson residuals: $r_{i,P} = (y_i - \hat{\lambda}_i) / \sqrt{\lambda_i}$
- $\sum_i r_{i,P}^2 = X^2$
 - Account for differences in variance

- ③ • Deviance residuals

$$r_{i,D} = \text{sign}(y_i - \hat{\lambda}_i) \sqrt{2 \left(y_i \log(y_i/\hat{\lambda}_i) - (Y_i - \hat{\lambda}_i) \right)}$$

- Define $d_i = r_{i,D}^2$, then Deviance= $D = \sum_i r_{i,D}^2 = \sum_i d_i$
- Account for differences in variance

Example: Galapagos Plants - Residual Plots

- a: As in linear regression, can plot response residuals r_i vs $\hat{E}(Y_i) = \hat{\lambda}_i$
 - Of limited use: variance is function of the expected value; Most predicted expected values are small
- b: Plot response residuals r_i vs link $\hat{\eta} = X\hat{\beta}$
 - The points on x axis are better separated; but variance still function of the expected value
- c: Plot deviance residuals $r_{i,D}$ vs link $\hat{\eta} = X\hat{\beta}$
- d: Plot Pearson residuals $r_{i,p}$ vs link $\hat{\eta} = X\hat{\beta}$

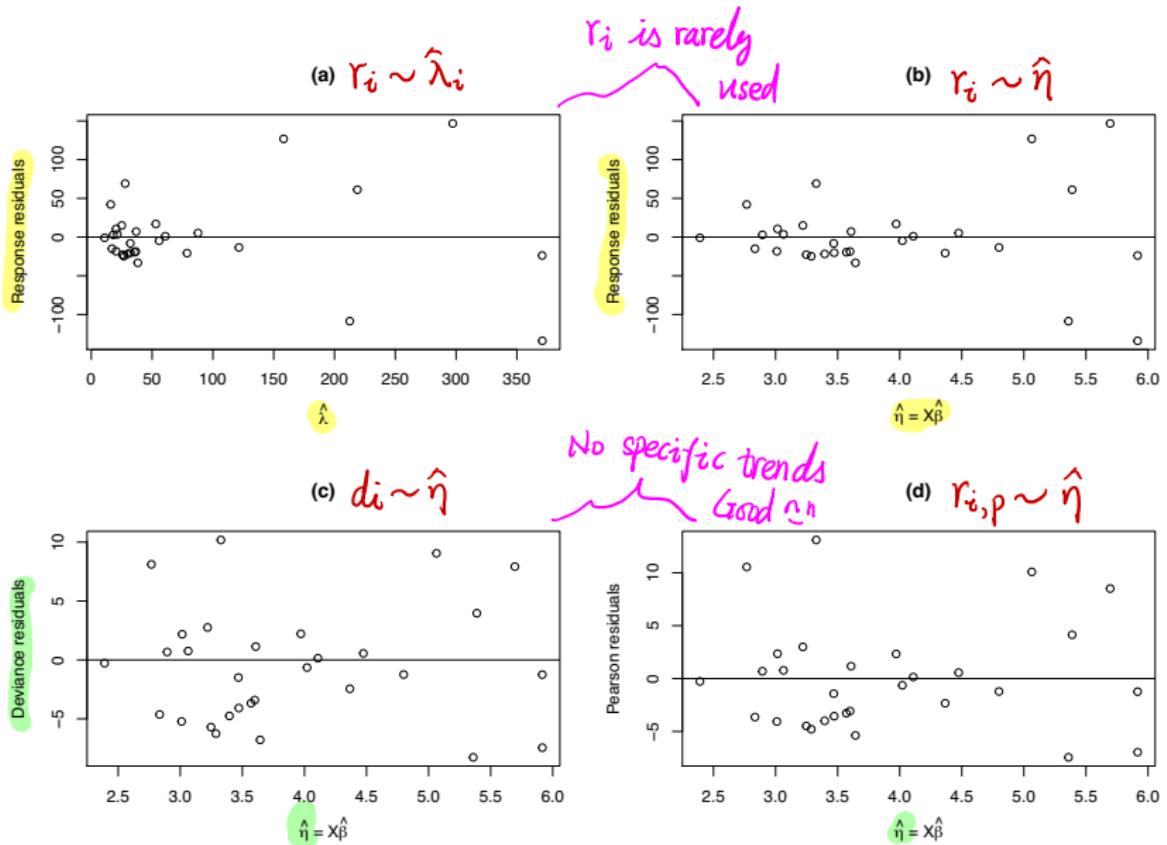
```
par(mfrow=c(2,2))
plot(residuals(mod1, type="response") ~ predict(mod1, type="response"),
      xlab=expression(hat(lambda)), ylab="Response residuals", main="(a)")
abline(h=0)
a: (
```

```
plot(residuals(mod1, type="response") ~ predict(mod1, type="link"),
      xlab=expression(paste(hat(eta), " = X", hat(beta))), 
      ylab="Response residuals", main="(b)")
abline(h=0)
b: (
```

```
plot(residuals(mod1) ~ predict(mod1, type="link"),
      xlab=expression(paste(hat(eta), " = X", hat(beta))), 
      ylab="Deviance residuals", main="(c)")
abline(h=0)
c: (
```

```
plot(residuals(mod1, type="pearson") ~ predict(mod1, type="link"),
      xlab=expression(paste(hat(eta), " = X", hat(beta))), 
      ylab="Pearson residuals", main="(d)")
abline(h=0)
d: (
```

Example: Galapagos Plants - Residual Plots



Example: Galapagos Plants - Residual Plots

Slide 12: LRT, χ^2 -test: both reject current model

- need further investigation to improve the model fitting
→ eg. transf. of X .

Comments on (c) and (d)

- Y axis accounts for differences in variance
- X axis better separates the predicted values
- No obvious problematic trends in (c) or (d)
 - Systematic nonlinear relationship → lack of fit
 - No apparent problems for this dataset
 - Large/nonconstant variance → violations of the assumed Poisson distribution
 - Potential overdispersion

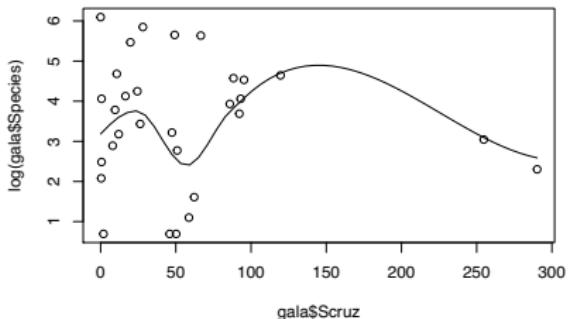
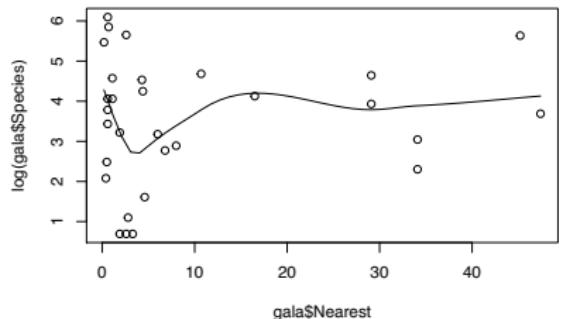
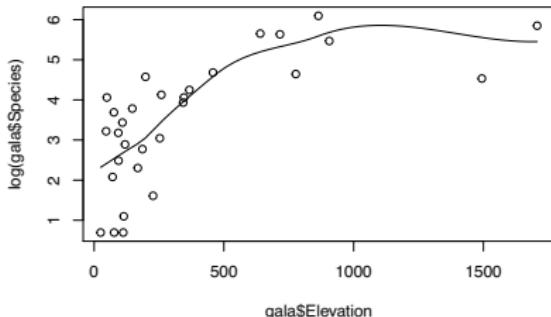
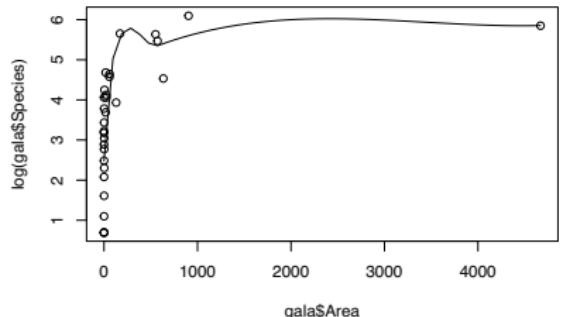
Model Diagnostics: Functional Form of Predictors

- can plot $\log(Y)$ against each predictor X_j

```
par(mfrow = c(2, 2))
scatter.smooth(gala$Area, log(gala$Species))
scatter.smooth(gala$Elevation, log(gala$Species))
scatter.smooth(gala$Nearest, log(gala$Species))
scatter.smooth(gala$Scruz, log(gala$Species))
```

- Transformations can be helpful: look for transformations for predictors even if there is no obvious problems in residual plots

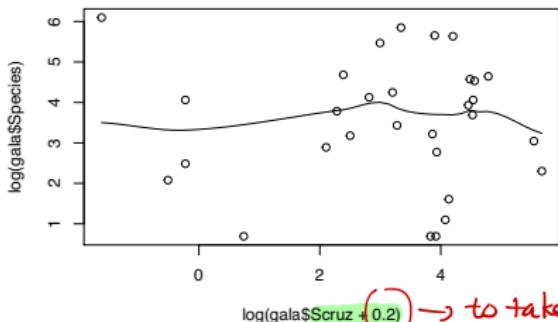
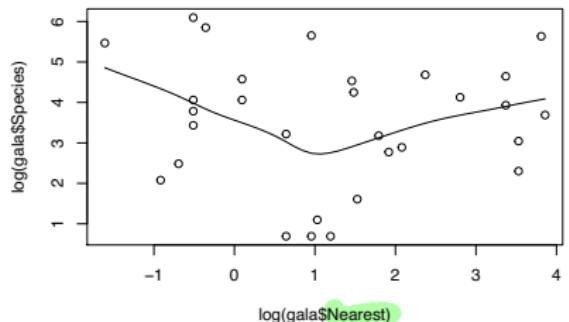
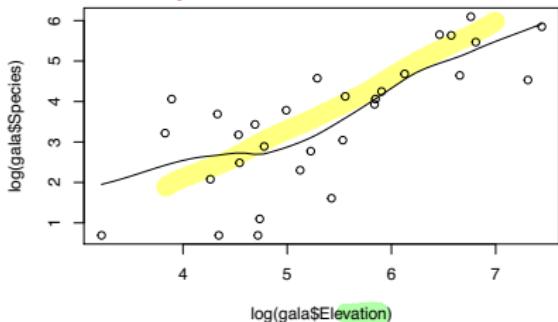
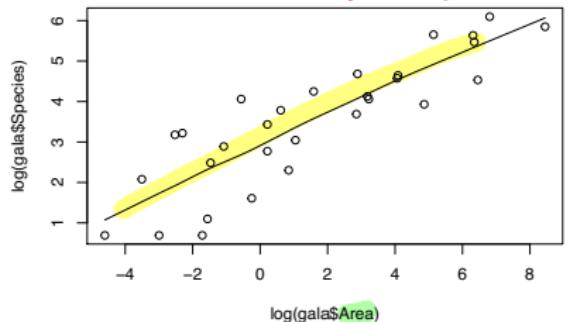
Model Diagnostics: Functional Form of Predictors



- Most predictors indicate some non-linearity $\notin \text{log } Y \sim X$

Gala data: Functional Form of Predictors

plots below: log-transform on all $X \Rightarrow \log Y \sim \log X$



- Log transformation appears appropriate, especially for Area and Elevation.

$\curvearrowleft \log Y \sim \log(X)$

$\circlearrowright \rightarrow$ to take care
 $\text{Scrub} = 0$

Gala Data: Log-Transformed Predictors

- Fit model with $\log(x)$ as predictor

```
mod2 <- glm(Species ~ log(Area) + log(Elevation) + log(Nearest) + log(Scruz +  
0.2) + log(Adjacent), family = poisson, data = gala)  
  
anova(mod1, mod2)
```

```
## Analysis of Deviance Table
```

```
##  
## Model 1: Species ~ Area + Elevation + Nearest + Scruz + Adjacent  
## Model 2: Species ~ log(Area) + log(Elevation) + log(Nearest) + log(Scruz +  
##     0.2) + log(Adjacent)  
##   Resid. Df Resid. Dev Df Deviance  
## 1      24    716.85  
## 2      24    359.59  0    357.25
```

Deviance for $\log Y \sim X := 716.85$

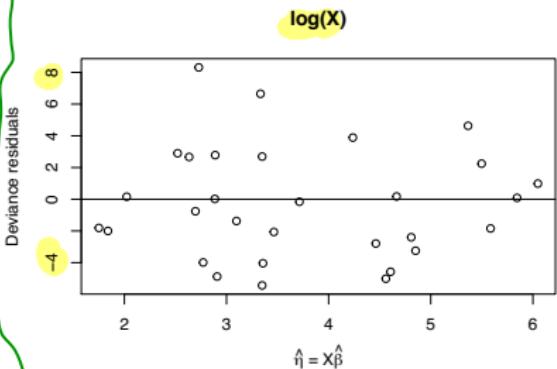
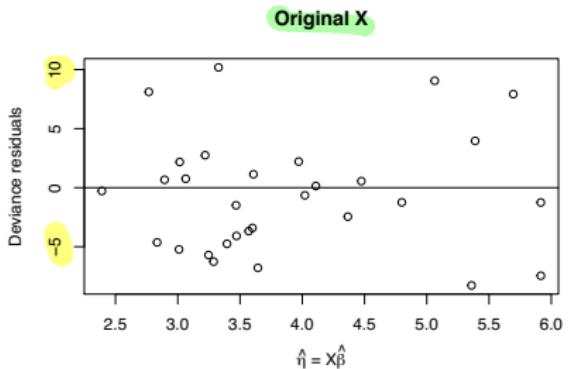
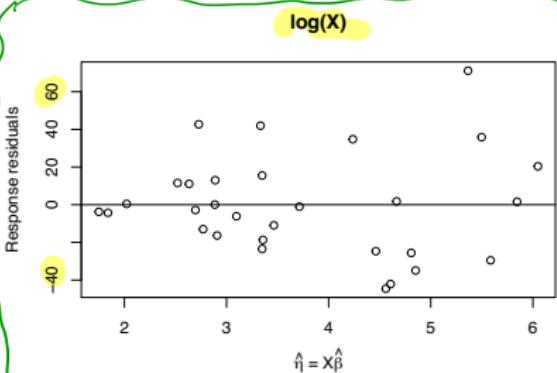
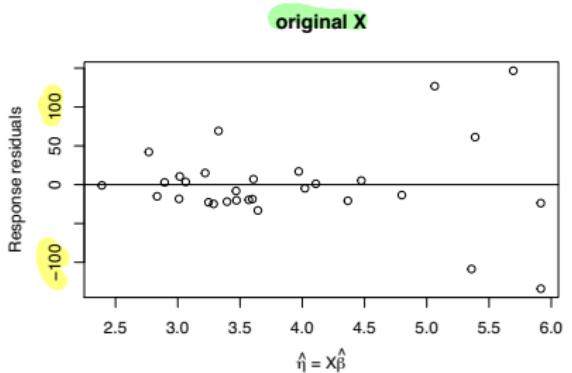
Deviance for $\log Y \sim \log X := 359.59$

$\Delta D = 357.25$



- obtain a substantial reduction in deviance
- the model with all the transformed predictors yields a better overall fit

Gala Data: visualize the change before and after Log-transformation



- More compact residuals after the transform

check Y-axis

Poisson model: overdispersion

- Suppose $Y \sim \text{Pois}(\lambda)$: as for Binomial response, overdispersion arises in case of correlated or clustered observations.

example • suppose λ itself is a r.v. with $E(\lambda) = \theta$. Then can be shown that

$$\begin{aligned} E(Y) &= E(E(Y|\lambda)) = \theta; V(Y) = E(V(Y|\lambda)) + V(E(Y|\lambda)) = \theta + V(\lambda) > \theta \\ &= E(\lambda) + V(\lambda) \end{aligned}$$

- Quasi-Poisson distribution

- $E(Y) = \lambda, V(Y) = \phi\lambda$
- estimate $\hat{\phi} = \frac{1}{n-p} X^2 = \frac{1}{n-p} \sum_i^n (y_i - \hat{\lambda}_i)^2 / \hat{\lambda}_i$
 - $\hat{\phi} \gg 1$: overdispersion
 - $\hat{\phi} \ll 1$: underdispersion (rare)

- Comparing nested models in presence of overdispersion

- Test H_0 : reduced model vs H_a : full model
- $F = \frac{D_R - D_F}{df_R - df_F} / \hat{\phi}_F \sim F_{df_R - df_F, df_F}$

{ From reduced model $\rightarrow \hat{\phi}_R$
From full model $\rightarrow \hat{\phi}_F$

Gala data: overdispersion

```
mod3 <- glm(Species ~ log(Area) + log(Adjacent), family = poisson, data = gala)

# log-log plot of mean vs variance
lam.hat <- predict(mod3, type = "response")
plot(log(lam.hat), log((gala$Species - lam.hat)^2), xlab = expression(hat(lambda)),
     ylab = expression((y - hat(lambda))^2))
abline(0, 1)
```

$$Y \sim \text{pois}(\lambda)$$

$$\Rightarrow \textcircled{1} E(Y) = \lambda$$

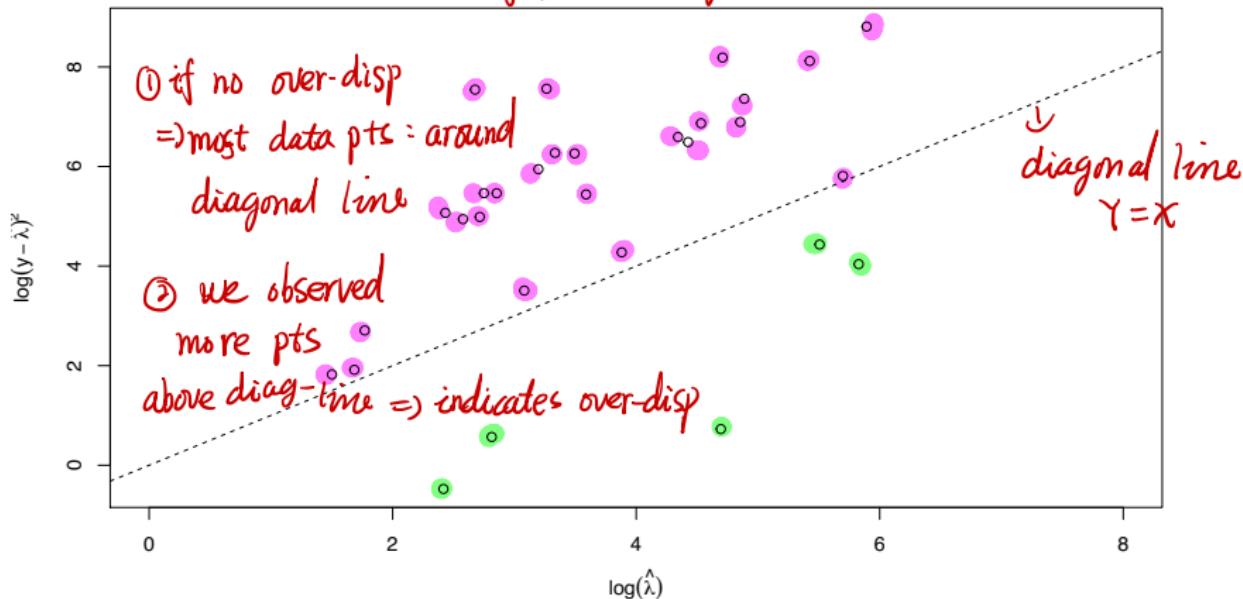
$$\textcircled{2} \text{Var}(Y) = \lambda = E\{(Y-\lambda)^2\}$$

$(Y-\lambda)^2 \sim \lambda$ could tell us about overdisp.

$\log(Y-\lambda)^2 \sim \log \lambda$

Gala data: overdispersion

$$\log(Y - \lambda)^2 \sim \log(\lambda)$$



- X axis: $\log(\widehat{E}(Y))$ of each observation
- Y axis: $\log(\widehat{V}(Y))$ of each observation
- The variance is increasing with the mean
- The variance is larger than the mean → indicates overdispersion

Gala data: account for overdispersion

① ✓
②

```
# Estimate dispersion parameter
( phi = sum(residuals(mod3, type="pearson")^2) / mod3$df.res ) ←  $\hat{\phi} = \frac{\chi^2}{n-p}$ 
## [1] 16.52745
```

①

```
# Summary output accounts for overdispersion
summary(mod3, dispersion=phi)
```

```
##
## Call:
## glm(formula = Species ~ log(Area) + log(Adjacent), family = poisson,
##      data = gala)
##
## Deviance Residuals:
##      Min        1Q     Median        3Q       Max
## -6.3760   -2.9527   -0.8288    2.9144    8.5786
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 3.27668  0.17942  18.26 < 2e-16 ***
## log(Area)   0.37503  0.03262  11.50 < 2e-16 ***
## log(Adjacent) -0.09575  0.02487 -3.85 0.000118 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 16.52745)
##
## Null deviance: 3510.73 on 29 degrees of freedom
## Residual deviance: 395.54 on 27 degrees of freedom
## AIC: 562.37
```

$\hat{\phi} > 1$: won't affect $\hat{\beta}$
 but affect $se(\hat{\beta})$
 and \hat{z}^* , p-value.

$$SE_{\hat{\phi}}(\hat{\beta}_i) = \sqrt{\hat{\phi}} SE(\hat{\beta}_i)$$

Gala data: account for overdispersion

{
①
②} ✓

(2)

```
# Using quasi-poisson model to account for overdispersion
mod3qs = glm(Species ~ log(Area)+log(Adjacent), family=quasipoisson, data=gala)
summary(mod3qs)
```

```
## 
## Call:
## glm(formula = Species ~ log(Area) + log(Adjacent), family = quasipoisson,
##      data = gala)
## 
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max 
## -6.3760 -2.9527 -0.8288  2.9144  8.5786 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 3.27668   0.17942  18.26 < 2e-16 ***
## log(Area)   0.37503   0.03262  11.50 6.56e-12 ***
## log(Adjacent) -0.09575  0.02487 -3.85 0.000658 ***  
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for quasipoisson family taken to be 16.52745)
## 
## Null deviance: 3510.73 on 29 degrees of freedom
## Residual deviance: 395.54 on 27 degrees of freedom
## AIC: NA
## 
## Number of Fisher Scoring iterations: 5
```

← compare highlighted columns
with slide 25: same

$$\hat{\beta}_j \rightarrow \sqrt{\hat{\phi}} SE(\hat{\beta}_j)$$

- Same estimates for parameters, but larger SE, weaker significance

Log-linear Models for Contingency Tables

- 2-way C-table : Table involves 2 variables
 - 3-way C-table : Table involves 3 variables
- - -

Log-linear Models for Two-way Contingency Tables

↳ two categorical variables: one for row, one for col.

- Men and women were asked whether they believed in the after life (1991 General Social Survey).
- Results:

| X=sex \ Y=belief | Yes | No | row.tot |
|------------------|-----|-----|---------|
| M | 435 | 147 | 582 |
| F | 375 | 134 | 509 |
| col.tot | 810 | 281 | 1091 =n |

- Question: is belief in afterlife independent of gender? \Leftrightarrow is $X \perp\!\!\!\perp Y$?

```
afterlife <- matrix(c(435,147,375,134),nrow=2,byrow=TRUE)
dimnames(afterlife)<-list(c("Female","Male"),c("Yes","No"))
names(dimnames(afterlife))<- c("Gender","Believer")
afterlife
```

```
##           Believer
## Gender   Yes  No
##   Female 435 147
##   Male    375 134
```

Fisher's exact test (permutation method)

| X=sex \ Y=belief | Yes | No | row.tot |
|------------------|-----|-----|---------|
| M | a | b | a+b |
| F | c | d | c+d |
| col.tot | a+c | b+d | n |

Fisher showed that the probability of obtaining any such set of values was given by the **hypergeometric distribution**:

T if $X \perp\!\!\!\perp Y$

$$p = \frac{\binom{a+b}{a} \binom{c+d}{c}}{\binom{n}{a+c}}$$

* Fisher's test assumes
row total is fixed

(A)

```
# fisher's exact test  
fisher.test(afterlife)
```

```
##  
## Fisher's Exact Test for Count Data  
##  
## data: afterlife  
## p-value = 0.7287  
## alternative hypothesis: true odds ratio is not equal to 1  
## 95 percent confidence interval:  
## 0.7979867 1.4004250  
## sample estimates:  
## odds ratio  
## 1.057334
```

Model for Two-way Contingency Tables

| X=sex \ Y=belief | Yes | No | row.tot |
|------------------|-----|-----|---------|
| M | 435 | 147 | 582 |
| F | 375 | 134 | 509 |
| col.tot | 810 | 281 | 1091 |

- The joint PDF of (X,Y): $P(X = x_i, Y = y_i) = \pi_{ij}$
- The marginal PDF of X: $P(X = x_i) = \pi_{i..}$, i=M, F
- The marginal PDF of Y: $P(Y = y_j) = \pi_{.j}$, j=Yes, No
- Expected cell counts: $\mu_{ij} = n\pi_{ij}$, $n = n_{++}$ is the total count.
- $N = (\# \text{ of rows}) * (\# \text{ of columns}) = rc$ is the effective sample size = number of observations
- Poisson rate: $\lambda_{ij} = \mu_{ij} = n\pi_{ij}$
- Log-linear model on $\log \mu_{ij}$

Independence model for Two-way Tables

- If X and Y are independent, then

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$$

- and the expected count is

$$\mu_{ij} = n\pi_{ij} = n\pi_{i.}\pi_{.j}$$

- this implies that the log-linear model satisfies

$$\log(\mu_{ij}) = \log(n) + \log(\pi_{i.}) + \log(\pi_{.j}) = \lambda + \lambda_i^X + \lambda_j^Y$$

they are NOT 1:1 corresponded.

This is in terms of
model form

glm(counts ~ X + Y, poisson)

Row, col variables

Independence model for Two-way Tables

- Under independence

$$\log(\mu_{ij}) = \underbrace{\chi\beta}_{\text{C}} + \underbrace{\lambda_i^X}_{\downarrow} + \underbrace{\lambda_j^Y}_{\downarrow}$$
$$\boxed{\log(\mu_{ij}) = \log(n) + \log(\pi_{i\cdot}) + \log(\pi_{\cdot j}) = \lambda + \lambda_i^X + \lambda_j^Y}$$

- What about variance? Because the data is Poisson, $V(Y_{ij}) = E(Y_{ij}) = \lambda_{ij}$,
the variance scales with the mean.
- Why Poisson?

$$Pois(k) = \sum_{i=1}^k Pois(1)$$

$$CLT : \frac{Pois(k) - k}{\sqrt{k}} \rightarrow N(0, 1)$$

- Thus “large poissons (large k) are like Gaussians”. But small Poissons (small k) are quite different.

Model for Two-way Contingency Tables

eg1

- $Y_{11} = 435$,
- $\hat{E}_{11} = \frac{n_1 \cdot n_{1\cdot}}{n} = \frac{582 \times 810}{1091} = 432.099$

2x2 table

$$\begin{cases} r=2 \\ c=2 \end{cases}$$

| X=sex \ Y=belief | Yes | No | row.tot |
|------------------|-----|-----|---------|
| M | 435 | 147 | 582 |
| F | 375 | 134 | 509 |
| col.tot | 810 | 281 | 1091 |

eg2

- $Y_{21} = 375$,
- $\hat{E}_{21} = \frac{n_2 \cdot n_{1\cdot}}{n} = \frac{509 \times 810}{1091} = 377.901$

- Model: $Y_{ij} \sim Pois(\lambda_{ij})$

H_0 : independence. i.e., $\lambda_{ij} = \lambda \alpha_i \beta_j \Leftarrow \alpha_i = \lambda_i^X, \beta_j = \lambda_j^Y$

H_a : λ_{ij} arbitrary

Pearson's χ^2 Test

$$\chi^2 = \sum_{i,j} \frac{(Y_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} \sim \chi^2_1 \text{ under } H_0$$

- $E(Y_{ij}) = \mu_{ij} = n\pi_i \cdot \pi_{j\cdot}, \hat{\pi}_{i\cdot} = n_{i\cdot}/n, \hat{\pi}_{j\cdot} = n_{j\cdot}/n, \Rightarrow \hat{\mu}_{ij} = \hat{n}_{i\cdot} \cdot \hat{n}_{j\cdot} = \hat{\gamma}_{ij} = \frac{n_{i\cdot} \cdot n_{j\cdot}}{n}$
- $\hat{\mu}_{ij} = n_{i\cdot} \cdot n_{j\cdot} / n$
- why df=1?

- In total, we have $2 \times 2 = 4$ observations, we need to estimate one α_i , one β_j , and the common λ , so the left over is: $(r-1)(c-1) = (2-1)(2-1) = 1$

Contingency table: Pearson test of independence

Suppose that we have k by m table. After life example: $k=m=2$. We call this a $k \times m$ contingency table.

- Model

$$Y_{ij} \sim Pois(\lambda_{ij})$$

- Mean function

$$\log(\lambda_{ij}) = \log(n) + \log(\pi_{i\cdot}) + \log(\pi_{\cdot j}) = \lambda + \lambda_i^X + \lambda_j^Y$$

- Pearson test for independence

$$\chi^2 = \sum_{i,j} \frac{(Y_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} \sim \chi^2_{(k-1)(m-1)} \text{ under } H_0$$

- $df=(k-1)(m-1)$: in total we have $n=km$ observations, we need to estimate $(k-1)$ of $\pi_{i\cdot}$, $(m-1)$ of $\pi_{\cdot j}$, and the common λ , so the $df=km-(k-1)-(m-1)-1=km-k-m+1=(k-1)(m-1)$

$$\sum_i^K \pi_{i\cdot} = 1$$

↑
need to estimate only $k-1$

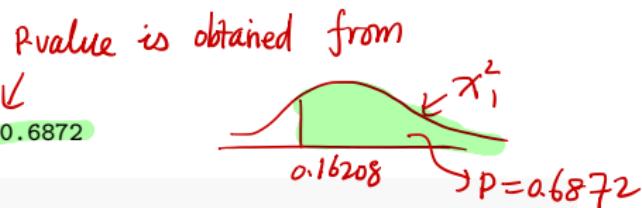
Model for Two-way Contingency Tables

B Pearson χ^2 test of independence

```
# Pearson's Chi-square test  
chisq.test(afterlife, correct=F)
```

```
##  
## Pearson's Chi-squared test  
##  
## data: afterlife  
## X-squared = 0.16208, df = 1, p-value = 0.6872
```

```
chisq.test(afterlife, correct=T)
```



```
##  
## Pearson's Chi-squared test with Yates' continuity correction  
##  
## data: afterlife  
## X-squared = 0.11103, df = 1, p-value = 0.739
```

```
chisq.test(afterlife, simulate.p.value=TRUE, B=10000) # default B=2000
```

/ obtained with correction

improve accuracy

```
##  
## Pearson's Chi-squared test with simulated p-value (based on 10000  
## replicates)  
##  
## data: afterlife  
## X-squared = 0.16208, df = NA, p-value = 0.732
```

obtained
by simulation

Model for Two-way Contingency Tables

C Deviance test (LRT)

- The saturated model for the two-way table can be written

$$\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY} \quad \begin{matrix} \leftarrow \text{rc observations} \\ \text{rc parameters} \end{matrix}$$

• $\hat{\mu}_{ij} = y_{ij} = \lambda_{ij}$

- This leads to the deviance of a contingency model having the form

$$D = \sum_{i,j} 2y_{ij} \log \frac{y_{ij}}{\hat{\mu}_{ij}} = \sum 2O_{ij} \log \frac{O_{ij}}{E_{ij}} \sim \chi^2_{(r-1)(c-1)}$$

- for the independence model $df = (r - 1)(c - 1)$

```
Gender <- as.factor(c("Female", "Female", "Male", "Male"))
Believer <- as.factor(c("Yes", "No", "Yes", "No"))
Count <- c(435, 147, 375, 134)
afterlife2 <- data.frame(Gender, Believer, Count)
afterlife2
```

| | X ₁ | X ₂ | Y | |
|------|----------------|----------------|-------|--|
| ## | Gender | Believer | Count | |
| ## 1 | Female | Yes | 435 | |
| ## 2 | Female | No | 147 | |
| ## 3 | Male | Yes | 375 | |
| ## 4 | Male | No | 134 | |

↙ rewrite 2x2 table
as a data frame for analysis

Model for Two-way Contingency Tables

Chi-square testing of independence using a Poisson model

```
modaf <- glm(Count~Gender+Believer, family=poisson, data=afterlife2)
# chisquare test using deviance
pchisq(deviance(modaf), df.residual(modaf), lower.tail=F)
```

[1] 0.6873263 ← C

```
# chisquare using pearson residuals
pearsonres = sum(resid(modaf,type="pearson")^2)
pchisq(pearsonres, df.residual(modaf), lower.tail=F) ← chisq.test(afterlife, correct=F)
```

[1] 0.6872451 ← B



In summary, to test the independence of column variable and row variable

- (A) • Fisher exact test.
- Run a Poisson model (or so called log-linear model)
- (B) • Pearson χ^2 test.
- (C) • χ^2 test using deviance. = LRT

↓
Three methods

- (A) Fisher-Exact
- (B) Pearson χ^2
- (C) LRT < Deviance
- (D) : slide 39 37/60



Take a break, and see you on Thursday

Review 2-way Table

For row, column variables

$$\log(E(Y_{ij})) = \mu + \alpha_i + \beta_j, (*) \quad \alpha_1 = \beta_1 = 0$$

Parameterization (Agresti Ch.8)

- $\mu = \log\{E(Y_{11})\}$
- α_i and β_j are deviations of $E(Y_{ij})$ from the reference cell (1,1) due to row i and column j.
- Total parameters to estimate: $1 + (r-1) + (c-1) \rightarrow df = rc - 1 - (r-1) - (c-1) = (r-1)(c-1)$

ML estimation with Poisson likelihood

- $\hat{\lambda}_{ij} = e^{\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j} = \frac{n_{i,j}}{n}$
- $\hat{\pi}_{ij} = \frac{\hat{\lambda}_{ij}}{\sum_{i,j} \hat{\lambda}_{ij}} = \frac{e^{\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j}}{\sum_{i,j} e^{\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j}} = \frac{e^{\hat{\alpha}_i}}{\sum_i e^{\hat{\alpha}_i}} \frac{e^{\hat{\beta}_j}}{\sum_j e^{\hat{\beta}_j}} = \hat{\pi}_i \hat{\pi}_j = \frac{n_{i,j}}{n^2} = \hat{\lambda}_{ij}/n$

As in ANOVA, all parameterizations produced identical estimates of probs. and counts.

Test for independence of rows and columns

(A) F-Exact (B) Pearson χ^2 (C) LRT : Deviance (D) Test sig. or not
for $(\alpha\beta)_{ij}$
in saturated M.

(D)

- $H_0 : (\alpha\beta)_{ij} = 0$ vs $H_a : (\alpha\beta)_{ij} \neq 0$. This is equivalent to
- H_0 : model (*) is correct; H_a : model (*) is inappropriate

Saturated model

$$\log(E(Y_{ij})) = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}, \quad \alpha_1 = \beta_1 = (\alpha\beta)_{1j} = (\alpha\beta)_{i1} = 0$$

- Total number of parameters to estimate = $1 + (r-1) + (c-1) + (r-1)(c-1) = rc$
- (i.e. describes each cell perfectly)

Log-linear vs Logit models

$$\text{model } \lambda_{rc} = E(Y_{rc}) \Leftrightarrow \pi_{rc} = p(X=r, Y=c) \\ \Leftrightarrow \text{model } \pi_{rc} = p(X=r, Y=c)$$

- Loglinear models are of use when at least two variables are response variables. If we only have a single categorical response, it is simpler and more natural to use logit models.
- When you have two variables (e.g. Gender vs after-life belief), then logit might treat one as explanatory and the other as response, while there is an equivalent log-linear model.
- Log-linear models view data as N independent cell counts rather than individual classifications of n subjects, $n = \sum_i^N Y_i$, and do not treat the row sums as fixed.

$$rc = (\# \text{ rows}) \times (\# \text{ of columns})$$

\hookrightarrow Fisher Exact method assume so.

$$\text{model } \pi(x) = p(Y|x) \Rightarrow \log \frac{\pi(x)}{1-\pi(x)} = x\beta$$

2×2 tables

Want to show $X \perp\!\!\!\perp Y \Rightarrow \begin{cases} \textcircled{1} \text{ show } P(X=x, Y=y) = P(X=x)P(Y=y) \\ \textcircled{2} \text{ show } P(Y=y | X=x) = P(Y=y) \end{cases}$

| X=sex \ Y=belief | Yes | No | row.tot |
|------------------|-----|-----|---------|
| M | 435 | 147 | 582 |
| F | 375 | 134 | 509 |
| col.tot | 810 | 281 | 1091 |

\textcircled{1} : poisson model

(log-linear
model)

\textcircled{2} : logit model

\textcircled{1}

Model: $Y_{ij} \sim Pois(\lambda_{ij})$

\textcircled{2}

if $Y_{i1} \sim Pois(\lambda_{i1}) \perp Y_{i2} \sim Pois(\lambda_{i2})$, then conditional on their sum, each count is a binomial

Assume $K \times 2$: $Y_{i1} | Y_{i1} + Y_{i2} = n_i \sim Bin(Y_{i1} + Y_{i2}, p_i = \frac{\lambda_{i1}}{\lambda_{i1} + \lambda_{i2}})$

$\Rightarrow \begin{cases} Y_{i1} | n_i \sim Bin(\dots) \\ Y_{i2} | n_i \sim Bin(\dots) \end{cases}$

Then

$$\text{logit}(p_i) = \log \frac{p_i}{1 - p_i} = \log \frac{\lambda_{i1}}{\lambda_{i2}} = \log(\lambda_{i1}) - \log(\lambda_{i2})$$

Under $H_0: \lambda_{ij} = \lambda \alpha_i \beta_j \Leftrightarrow X \perp\!\!\!\perp Y$

$$\log(\lambda_{ij}) = \log(\lambda) + \log(\alpha_i) + \log(\beta_j)$$

$$P(Y_{i1} | \text{row}=i, \text{row sum}=n_i) = \log(\beta_1) + \log(\beta_2)$$

$$\log(\lambda_{i1}) - \log(\lambda_{i2})$$

the logit does not depend on i .

In binomial regression, we are modelling

$$\text{logit}\{P(Y_{i1}|X)\} = \beta_0 + \beta_1 X = \text{logit } P(Y|X), \text{ if } \beta_i = 0$$

Testing H_0 is equivalent to testing $\beta_1 = 0$ in logistic regression.

then $P(Y|X) = P(Y)$

2×2 tables

Testing row, col variable (X, Y) are indep.

- ① Fisher-Exact
- ② Pearson χ^2
- ③ Deviance: LRT
- ④ Check sig. or
not for interaction
term in saturated
model
- ⑤ Test. $H_0: \beta_1 = 0$
in logit model

↑
equivalent to

Compare

$M_1: Y \sim X$

with $M_0: Y \sim 1$

method

(4)

```

modpois <- glm(Count~Gender*Believer, family=poisson, data=afterlife2)
anova(modpois,test="LRT")

## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: Count
##
## Terms added sequentially (first to last)
##
##          Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL              3    272.685
## Gender            1     4.888    2    267.797  0.02704 *
## Believer          1   267.635    1     0.162 < 2e-16 ***
## Gender:Believer  1     0.162    0     0.000  0.68733
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

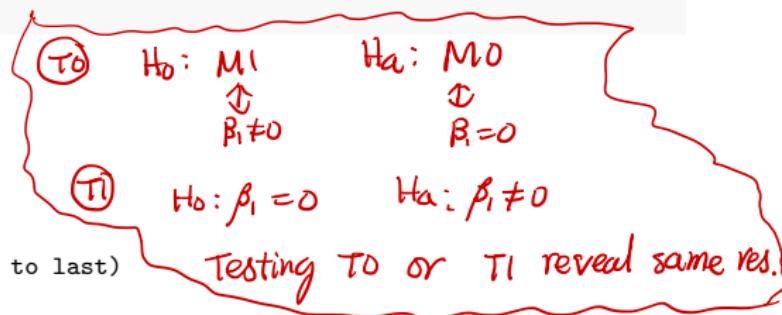
```

2×2 tables

method ⑤

```
Yes = c(rep(1,435),rep(0,147),rep(1,375),rep(0,134))
Gender = c(rep(0,(435+147)),rep(1,(375+134)))
aflife=data.frame(Yes,Gender)
modlogit = glm(Yes~Gender, family=binomial, data=aflife)
anova(modlogit,test="LRT")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Yes
##
## Terms added sequentially (first to last)
##
##          Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL           1090      1244.8
## Gender         1     0.162    1089      1244.7 0.6873
```



$\leftarrow M_0: Y \sim 1 \Leftarrow \log \pi = \beta_0$

$\leftarrow M_1: Y \sim X \Leftarrow \log \pi = \beta_0 + \beta_1 X$

\downarrow

equivalent to testing

$H_0: \beta_1 = 0 \quad \text{vs} \quad \beta_1 \neq 0$
 $\in M_0 \quad \in M_1$

3-way table: mutual independence $\leftrightarrow X \perp\!\!\!\perp Y \perp\!\!\!\perp Z$

↳ Table involves 3 categorical variables

- 3-way $I \times J \times K$ cross-classification of r.v. X, Y and Z
- Assume the count $Y_{ijk} \sim Pois(E\{Y_{ijk}\})$
- X, Y and Z are mutually independent if $\pi_{ijk} = \pi_i \pi_j \pi_k$

↳ assumption of this model

$$\log E\{Y_{ijk}\} = \log(n) + \log(\pi_i) + \log(\pi_j) + \log(\pi_k)$$

- total number of parameters: $1 + (I-1) + (J-1) + (K-1)$
- The log-linear model is

$$\log E\{Y_{ijk}\} = \mu + \alpha_i + \beta_j + \gamma_k$$

- $\mu = \log E(Y_{111})$, $\alpha_1 = \beta_1 = \gamma_1 = 0$
- $\hat{\lambda}_{ijk} = e^{\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k} = \frac{n_{i..} n_{.j.} n_{..k}}{n^2}$
- $\hat{\pi}_{ijk} = \frac{\hat{\lambda}_{ijk}}{\sum_{i,j,k} \hat{\lambda}_{ijk}} = \frac{e^{\hat{\alpha}_i}}{\sum_i e^{\hat{\alpha}_i}} \frac{e^{\hat{\beta}_j}}{\sum_j e^{\hat{\beta}_j}} \frac{e^{\hat{\gamma}_k}}{\sum_k e^{\hat{\gamma}_k}} = \hat{\pi}_i \hat{\pi}_j \hat{\pi}_k = \frac{n_{i..} n_{.j.} n_{..k}}{n^3} = \hat{\lambda}_{ijk}/n$

3-way table: Joint independence $X \perp\!\!\!\perp Y$, but $(XY) \perp\!\!\!\perp Z$

- 3-way $I \times J \times K$ cross-classification of r.v. X, Y and Z
- Assume the count $Y_{ijk} \sim Pois(E\{Y_{ijk}\})$
- X and Y are dependent, but together they are independent of Z if

model
Assump. \rightarrow

$$\pi_{ijk} = \pi_{ij}\pi_k$$

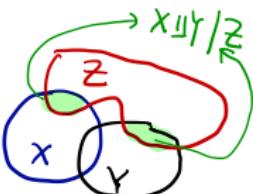
$$\log E\{Y_{ijk}\} = \log(n) + \log(\pi_{ij}) + \log(\pi_k)$$

- total number of parameters: $1 + (I-1) + (K-1)$
- The log-linear model is

$$\log E\{Y_{ijk}\} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k,$$

- $\mu = \log E(Y_{111})$, $\alpha_1 = \beta_1 = \gamma_1 = (\alpha\beta)_{1j} = (\alpha\beta)_{i1} = 0$
- $\hat{\lambda}_{ijk} = e^{\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\alpha\beta}_{ij} + \hat{\gamma}_k} = \frac{n_{ij} \cdot n_{..k}}{n}$
- $\hat{\pi}_{ijk} = \frac{\hat{\lambda}_{ijk}}{\sum_{i,j,k} \hat{\lambda}_{ijk}} = \frac{e^{\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\alpha\beta}_{ij}}}{\sum_{i,j} e^{\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\alpha\beta}_{ij}}} \frac{e^{\hat{\gamma}_k}}{\sum_j e^{\hat{\gamma}_k}} = \hat{\pi}_{ij} \hat{\pi}_k = \frac{n_{ij} \cdot n_{..k}}{n^2} = \hat{\lambda}_{ijk}/n$

3-way table: Conditional independence $\leftrightarrow X \perp\!\!\!\perp Y | Z$



- 3-way $I \times J \times K$ cross-classification of r.v. X, Y and Z
- Assume the count $Y_{ijk} \sim Pois(E\{Y_{ijk}\})$
- X and Y are independent given Z if $\pi_{ij|k} = \pi_{i|k}\pi_{j|k}$ $\Leftrightarrow \hat{\pi}_{ijk} = \hat{\pi}_{ij|k}\hat{\pi}_k$

$$\log E\{Y_{ijk}\} = \log(n) + \log(\pi_{ij}) + \log(\pi_k)$$

- weaker than mutual or joint independence
- $\pi_{ijk} = \pi_{i|k}\pi_{j|k}\pi_k = \frac{\pi_{ik}}{\pi_k} \frac{\pi_{jk}}{\pi_k} \pi_k = \frac{\pi_{ik}\pi_{jk}}{\pi_k}, \hat{\pi}_{ijk} = \frac{n_{i,k}n_{j,k}}{n_{..k}}$ $\Leftrightarrow \hat{\lambda}_{ijk} = \frac{n_{i,k}n_{j,k}}{n_{..k}}$

- The log-linear model is

$$\log E\{Y_{ijk}\} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk},$$

- $\mu = \log E(Y_{111})$
- $\alpha_1 = \beta_1 = \gamma_1 = (\alpha\gamma)_{1k} = (\alpha\gamma)_{i1} + (\beta\gamma)_{1k} = (\beta\gamma)_{j1} = 0$
- $\hat{\lambda}_{ijk} = e^{\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \hat{\alpha}\gamma_{ik} + \hat{\beta}\gamma_{jk}}$
- $\hat{\pi}_{ijk} = \frac{e^{\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \hat{\alpha}\gamma_{ik} + \hat{\beta}\gamma_{jk}}}{\sum_{i,j,k} e^{\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \hat{\alpha}\gamma_{ik} + \hat{\beta}\gamma_{jk}}} = \frac{e^{\hat{\alpha}_i + \hat{\gamma}_k + \hat{\alpha}\gamma_{ik}}}{\sum_{i,k} e^{\hat{\alpha}_i + \hat{\gamma}_k + \hat{\alpha}\gamma_{ik}}} \frac{e^{\hat{\beta}_j + \hat{\gamma}_k + \hat{\beta}\gamma_{jk}}}{\sum_{j,k} e^{\hat{\beta}_j + \hat{\gamma}_k + \hat{\beta}\gamma_{jk}}} \frac{\sum_k e^{\hat{\gamma}_k}}{e^{\hat{\gamma}_k}}$

3-way table: Uniform (homogeneous) Association

- 3-way $I \times J \times K$ of r.v. X, Y and Z, $Y_{ijk} \sim \text{Pois}(E\{Y_{ijk}\})$
- For each level of one variable, same association of the other two variables if

model
Assump → $\pi_{ijk} = \pi_{ij}\pi_{jk}\pi_{ik}$

$$\log E\{Y_{ijk}\} = \log(n) + \log(\pi_{ij}) + \log(\pi_{jk}) + \log(\pi_{ik})$$

- The log-linear model is

$$\log E\{Y_{ijk}\} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta)_{ij},$$

- $\mu = \log E(Y_{111})$
- $\alpha_1 = \beta_1 = \gamma_1 = (\alpha\gamma)_{1k} = (\alpha\gamma)_{i1} + (\beta\gamma)_{1k} = (\beta\gamma)_{j1} = (\alpha\beta\gamma)_{1j} = (\alpha\beta\gamma)_{i1} = 0$
- Not a saturated model, since no 3-way interaction.

- $\hat{\lambda}_{ijk} = e^{\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \hat{\alpha}\gamma_{ik} + \hat{\beta}\gamma_{jk} + \hat{\alpha}\beta_{ij}}$ ↪ not explicit form

- $\hat{\pi}_{ijk} = \frac{e^{\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \hat{\alpha}\gamma_{ik} + \hat{\beta}\gamma_{jk} + \hat{\alpha}\beta_{ij}}}{\sum_{i,j,k} e^{\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \hat{\alpha}\gamma_{ik} + \hat{\beta}\gamma_{jk} + \hat{\alpha}\beta_{ij}}} = \frac{e^{\hat{\alpha}_i + \hat{\gamma}_k + \hat{\alpha}\gamma_{ik}}}{\sum_{i,k} e^{\hat{\alpha}_i + \hat{\gamma}_k + \hat{\alpha}\gamma_{ik}}} \frac{e^{\hat{\beta}_j + \hat{\gamma}_k + \hat{\beta}\gamma_{jk}}}{\sum_{j,k} e^{\hat{\beta}_j + \hat{\gamma}_k + \hat{\beta}\gamma_{jk}}} \frac{e^{\hat{\alpha}_i + \hat{\beta}_j + \hat{\alpha}\beta_{ij}}}{\sum_{i,j} e^{\hat{\alpha}_i + \hat{\beta}_j + \hat{\alpha}\beta_{ij}}} = \hat{\pi}_{ij}\hat{\pi}_{jk}\hat{\pi}_{ik}$

3-way table: Interpretation of Uniform Association

- constant odds ratios between the levels of two variables, for each level of the third variable
 - e.g. for $i=1,2, j=1,2$ and a given level k
$$\log OR = \log \frac{\lambda_{11k} \lambda_{22k}}{\lambda_{12k} \lambda_{21k}} = (\alpha\beta)_{11} + (\alpha\beta)_{22} - (\alpha\beta)_{12} - (\alpha\beta)_{21}$$
 - independent of k
- no easy way to estimate $\hat{\lambda}_{ijk}$ and $\hat{\pi}_{ijk}$ based on cell counts.

3-way: loglinear models summary

- Y_{ijk} count in cell (i,j,k), $Y_{ijk} \sim Pois(\lambda_{ijk})$

- Conditional on $n = \sum_{i,j,k} n_{ijk}$

$$Y_{ijk}|n \sim Multinom(\pi_{ijk})$$

- $\mu_{ijk} = \log E(Y_{ijk})$ -reference cell
- $\alpha_i, \beta_j, \gamma_k$: deviations of $\log E(Y_{ijk})$ from reference cell, $\alpha_1 = \beta_1 = \gamma_1 = 0$
- Residual df=IJK- # of parameters.

(X)

| Model | $\log E\{Y_{ijk}\} =$ |
|--------------|---|
| Mut. Indep | $\mu + \alpha_i + \beta_j + \gamma_k$ |
| Joint Indep. | $\mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij}$ |
| Cond. Indep. | $\mu + \alpha_i + \beta_j + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$ |
| Unif. Assoc. | $\mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$ |
| Saturated | $\mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}$ |

(Y)

| Model | $\pi_{ijk} =$ | $\hat{\lambda}_{ijk} =$ |
|--------------|------------------------------|---------------------------------|
| Mut. Indep | $\pi_i \pi_j \pi_k$ | $n_{i++} n_{+j+} n_{++k} / n^2$ |
| Joint Indep. | $\pi_{ij} \pi_k$ | $n_{ij+} n_{++k} / n$ |
| Cond. Indep. | $\pi_{ik} \pi_{jk} / \pi_k$ | $n_{i+k} n_{+jk} / n_{++k}$ |
| Unif. Assoc. | $\pi_{ij} \pi_{ik} \pi_{jk}$ | Iterative |
| Saturated | π_{ijk} | n_{ijk} |

3-way Tables: ACM example

Survey asked 2276 students in their final year of high school in a nonurban area near Dayton, Ohio whether they ever used alcohol, cigarettes, or marijuana.

| Alcohol Use | Cigarette Use | Marijuana Use Yes | Marijuana Use No |
|-------------|---------------|----------------------|---------------------|
| Yes | Yes | 911 | 538 |
| No | Yes | 3 | 43 |
| | No | 2 | 279 |

← 3 categorical variables

- each has 2 level

$$\Rightarrow \# \text{ of obs } Y_{ijk} \\ = 2 \times 2 \times 2 = 8$$

This is example of a $2 \times 2 \times 2$ contingency table. Shorthand: A=alcohol, C=cigarette, M=marijuana.

3-way Tables: ACM example

| Alcohol Use | Cigarette Use | Marijuana Use Yes | Marijuana Use No |
|----------------|------------------|----------------------|---------------------|
| Yes | Yes | 911 | 538 |
| | No | 44 | 456 |
| No | Yes | 3 | 43 |
| | No | 2 | 279 |

- model: $Y_{ijk} \sim Pois(\lambda_{ijk})$
- conditional on total (N): $Y_{ijk} \sim Multinom(N, \pi_{ijk})$
 - $\pi_{i..}$ be the prob of row A=i
 - $\pi_{j..}$ be the prob of row A=i, C=j, and etc
- Model A.C.M : A,C, and M are mutually independent

$$\log(\lambda_{ijk}) = \lambda + \lambda_i^A + \lambda_j^C + \lambda_k^M$$

$$\pi_{ijk} = \pi_{i..} \pi_{j..} \pi_{k..}$$

3-way Tables: ACM example

| Alcohol Use | Cigarette Use | Marijuana Use Yes | Marijuana Use No |
|----------------|------------------|----------------------|---------------------|
| Yes | Yes | 911 | 538 |
| | No | 44 | 456 |
| No | Yes | 3 | 43 |
| | No | 2 | 279 |

- model: $Y_{ijk} \sim Pois(\lambda_{ijk})$
- conditional on total (N): $Y_{ijk} \sim Multinom(N, \pi_{ijk})$
 - $\pi_{i..}$ be the prob of row A=i
 - $\pi_{j..}$ be the prob of row A=i, C=j, and etc
- Model AC.M: M is jointly independent of A, C $\leftrightarrow A, C$ together $\perp\!\!\!\perp M$

$$\log(\lambda_{ijk}) = \lambda + \lambda_i^A + \lambda_j^C + \lambda_k^M + \lambda_{ij}^{AC}$$

$$\pi_{ijk} = \pi_{ij.} \pi_{..k}$$

3-way Tables: ACM example

| | | Marijuana Use | |
|-----|-----|---------------|---------------|
| | | Yes | No |
| | | Alcohol Use | Cigarette Use |
| Yes | Yes | 911 | 538 |
| | No | 44 | 456 |
| No | Yes | 3 | 43 |
| | No | 2 | 279 |

- model: $Y_{ijk} \sim Pois(\lambda_{ijk})$
- conditional on total (N): $Y_{ijk} \sim Multinom(N, \pi_{ijk})$
 - $\pi_{i..}$ be the prob of row A=i
 - $\pi_{j..}$ be the prob of row A=i, C=j, and etc
- Model AC.AM : C and M conditionally independent given A

$$\log(\lambda_{ijk}) = \lambda + \lambda_i^A + \lambda_j^C + \lambda_k^M + \lambda_{ij}^{AC} + \lambda_{ik}^{AM}$$

$$\pi_{jk|i} = \pi_{j..|i} \pi_{.k|i} \Leftrightarrow \pi_{ijk} = \pi_i \pi_{jk|i}$$

$$= \pi_i \pi_{j..} \pi_{.k|i}$$

3-way Tables: ACM example

| Alcohol Use | Cigarette Use | Marijuana Use Yes | Marijuana Use No |
|----------------|------------------|----------------------|---------------------|
| Yes | Yes | 911 | 538 |
| | No | 44 | 456 |
| No | Yes | 3 | 43 |
| | No | 2 | 279 |

- model: $Y_{ijk} \sim Pois(\lambda_{ijk})$
- conditional on total (N): $Y_{ijk} \sim Multinom(N, \pi_{ijk})$
 - $\pi_{i..}$ be the prob of row A=i
 - $\pi_{ij.}$ be the prob of row A=i, C=j, and etc
- Model AC.AM.CM: Each pair of A,C, and M has homogeneous association.

(AC) $\perp\!\!\!\perp$ (AM) $\perp\!\!\!\perp$ (CM)

$$\log(\lambda_{ijk}) = \lambda + \lambda_i^A + \lambda_j^C + \lambda_k^M + \lambda_{ij}^{AC} + \lambda_{ik}^{AM} + \lambda_{jk}^{CM}$$

- e.g. the dependence relationship of A,C, does not dependent on M.

3-way Tables: ACM example

| Alcohol Use | Cigarette Use | Marijuana Use Yes | Marijuana Use No |
|----------------|------------------|----------------------|---------------------|
| Yes | Yes | 911 | 538 |
| | No | 44 | 456 |
| No | Yes | 3 | 43 |
| | No | 2 | 279 |

- model: $Y_{ijk} \sim Pois(\lambda_{ijk})$
- conditional on total (N): $Y_{ijk} \sim Multinom(N, \pi_{ijk})$ $N = \sum_i \sum_j \sum_k Y_{ijk}$
 - $\pi_{i..}$ be the prob of row A=i
 - $\pi_{ij.}$ be the prob of row A=i, C=j, and etc
- Saturated Model ACM :

$$\log(\lambda_{ijk}) = \lambda + \underbrace{\lambda_i^A}_{\text{Main effect}} + \underbrace{\lambda_j^C}_{\text{all two-way effect}} + \underbrace{\lambda_k^M}_{\text{3-way effect.}} + \lambda_{ij}^{AC} + \lambda_{ik}^{AM} + \lambda_{jk}^{CM} + \lambda_{ijk}^{ACM}$$

3-way Tables: ACM example

| Model | Interpretation |
|-------------|---------------------------------------|
| A.C.M | Mutual independence |
| AC.M | AC jointly independent of M |
| AC.AM | M,C conditionally independent given A |
| AC.AM.CM | homogeneous association of each pair |
| Saturated M | - |

```
# ---- Alcohol, Cigarette, and Marijuana Use Example. ----
ACM <- data.frame(expand.grid(marijuana=factor(c("Yes","No"),levels=c("No","Yes")),
                                cigarette=factor(c("Yes","No"),levels=c("No","Yes")),
                                alcohol=factor(c("Yes","No"),levels=c("No","Yes"))),
                                count=c(911,538,44,456,3,43,2,279))
head(ACM,5)

##    marijuana cigarette alcohol count
## 1      Yes       Yes     Yes   911
## 2      No        Yes     Yes   538
## 3      Yes       No      Yes    44
## 4      No        No      Yes   456
## 5      Yes       Yes      No     3
```

3-way Tables: ACM example

```
# mutual independence.  
A.C.M.glm<-glm(count~alcohol+marijuana+cigarette,data=ACM,family=poisson)  
# Joint independence of marijuana and alcohol and cigarette.  
AC.M.glm<-glm(count~alcohol*cigarette+marijuana,data=ACM,family=poisson)  
# Conditional independence of marijuana and cigarette given alcohol.  
AM.AC.glm<-glm(count~alcohol*cigarette+alcohol*marijuana,data=ACM,family=poisson)  
# Homogeneous association  
AC.AM.CM.glm<-glm(count~alcohol*cigarette+alcohol*marijuana+cigarette*marijuana,data=ACM,family=poisson)  
# Saturated Model  
ACM.glm<-glm(count~alcohol*cigarette*marijuana,data=ACM,family=poisson)  
aicval =round( c( AIC(A.C.M.glm),AIC(AC.M.glm),AIC(AM.AC.glm),AIC(AC.AM.CM.glm),  
           AIC(ACM.glm)),digits=4)  
aicval  
  
## [1] 1343.0634 902.8701 558.4127 63.4174 65.0434 ← AIC for all models
```

3-way Tables: ACM example

| Model | Interpretation | AIC |
|-------------|---------------------------------------|---------|
| A.C.M | Mutual independence | 1343.06 |
| AC.M | AC jointly independent of M | 902.87 |
| AC.AM | M,C conditionally independent given A | 558.41 |
| ✓ AC.AM.CM | homogeneous association of each pair | 63.42 |
| Saturated M | - | 65.04 |

AIC as model selection criteria : the smaller, the better

3-way Tables: ACM example

```
aicval =round( c( AIC(A.C.M.glm),AIC(AC.M.glm),AIC(AM.AC.glm),AIC(AC.AM.CM.glm),
AIC(ACM.glm)),digits=4)
```

```
aicval
```

```
## [1] 1343.0634 902.8701 558.4127 63.4174 65.0434
```

```
anova(A.C.M.glm,AC.M.glm,AM.AC.glm,AC.AM.CM.glm,ACM.glm, test="Chisq")
```

nested model testing

```
## Analysis of Deviance Table
```

```
##
```

```
## Model 1: count ~ alcohol + marijuana + cigarette  $\rightarrow D_1 = 1286.02$ 
```

```
## Model 2: count ~ alcohol * cigarette + marijuana  $\rightarrow D_2 = 843.83$ 
```

```
## Model 3: count ~ alcohol * cigarette + alcohol * marijuana  $D_3 = 497.37$ 
```

```
## Model 4: count ~ alcohol * cigarette + alcohol * marijuana + cigarette *
##           marijuana  $\rightarrow D_4 = 0.37$ 
```

```
## Model 5: count ~ alcohol * cigarette * marijuana
```

```
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)
```

```
## 1      4    1286.02
```

```
## 2      3     843.83  1   442.19 <2e-16 ***
```

```
## 3      2     497.37  1   346.46 <2e-16 ***
```

```
## 4      1      0.37  1   497.00 <2e-16 ***
```

```
## 5      0      0.00  1     0.37  0.5408
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
```

$$\text{eg. } 442.19 = D_1 - D_2 = 1286.02 - 843.83$$

$$497.00 = D_3 - D_4 = 497.37 - 0.37 \dots$$



$H_0: M_1$ $H_1: M_2$

$H_0: M_2$ $H_1: M_3$

$H_0: M_3$ $H_1: M_4$

$H_0: M_4$, $H_1: \text{Saturated}$

After Lecture This Week

Practice problems

- Review all the slides
- Try all the R example in slides.

Topics for next week:

- Topic in the following week after midterm
 - Introduction to linear mixed effect model