19 November

Thursday, November 19, 2020 11:08 AM

Used for modelling the count of things as a function of covariates. Counts are non-negative integers

Examples

Vehicles in line
# of failures

Why special method?

- OLS can/will predict negative and non-integer values
- Poisson and negative binomial are popular
- Zero inflated model can work

Poisson models

$$Pr(Y=y_i)=\frac{EXP^{-\lambda_i}\lambda_i^{y_i}}{y_i!}\ ;\ y=0,1,2\ldots$$

where; $E(y_i)=\lambda_i=EXP(\beta X_i)$ $\quad\ln(\lambda_i)=\beta' X_i$

Expression b/w response and predictors

log-linear form
Exp portion of the expression constrains to positive

Poisson Models

by substituting $E(y)=EXP(\beta X)$ we get

$$LL(\beta)=\sum_{i=1}^{n}\left[-EXP(\beta X_i)+y_i\beta X_i-LN(y_i!)\right]$$

Poisson model Elasticities

Elasticity = estimate of effect of a change in independent variable on the dependent variable

Elasticity on count of individual i for k^{th} independent variable is given as

$$E_{x_{ik}}^{\lambda_i}=\frac{\partial\lambda_i}{\lambda_i}\times\frac{x_{ik}}{\partial x_{ik}}=\beta_k x_{ik}$$

if $E=-5.4$, then 1% increase $=-5.4\%$ expected frequency

Pseudo-elasticity

used for discrete variables

$$E_{x_{ik}}^{\lambda_i}=EXP(\beta_k)-1/EXP(\beta_k)$$

Poisson models, GOF measures

Log-likelihood ratio test to compare restricted + unrestricted models

$$-2[LL(\beta_R)-LL(\beta_U)]\approx\chi^2(\alpha,df_u-df_R)$$

Sum of model deviances, G-square = zero for perfect fit

$$G^2=2\sum_{i=1}^{n}y_i LN(y_i/\hat{\lambda}_i)$$

$$0\le G^2\le ?$$

Measure similar to r^2 is

$$R_p^2=1-\left(\sum_{i=1}^{n}\left[y_i-\hat{\lambda}_i/\sqrt{\hat{\lambda}_i}\right]^2\Big/\sum_{i=1}^{n}\left(y_i-\bar{y}/\sqrt{\bar{y}}\right)^2\right)$$

$rho^2=1-LL(\beta)/LL(0)$ → overall fit

No true r^2 equivalent

Example

Intersection accidents at 2 lane rds in CA and MI

Vars = State	Accident	AADT1	AADT2	Median	Drive
CA = 0, MI = 1	Count	major rd	minor rd	width of median	#driveways in 250 ft

$$E[y_i]=\lambda_i=EXP(\beta x_i)$$

$$=EXP[-.83+.0008AADT1+.0005AADT2-.06Median+.07Drive]$$

$$=EXP^{-.83}EXP^{.0008AADT1}\ldots EXP^{.07Drive}$$

$$=(.436)(AADT1\,Factor)\ldots(Drive\,Factor)$$

model is additive in the exponent or multiplicative on expected value of Y

Poisson model restriction

λ represents distribution mean and variance

Often variance ≠ mean and Poisson is bad

Over-dispersion

$Var(y)>E(x)$ occurs when:
1) Poisson process over interval has random length
2) Inter-subject variability
   $E(z)=\lambda$ and $Var(z)=\lambda/\phi$ and $\phi\neq 1$

When occurs, change model to negative binomial (gamma distributed) assuming over dispersion is normally distributed

Poisson and Negative binomial models

$\lambda_i=EXP(\beta' X_i+\varepsilon_i)$ → EXP^{ε_i} is gamma distributed
mean = 1 variance = α

Test for overdispersion

assume model has mean zero

$H_0=Var[y_i]=E[y_i]$

$H_a=Var[y_i]=E[y_i]+\alpha g(E(y_i))$