Today – Overdispersion in Poisson regression 3/24/22
– Residuals + checking model fit
– Model selection

Overdispersion – Overdispersion parameter estimate:

$$\hat{\phi} = \frac{\sum_{i=1}^{n} e_i^2}{n-p} \qquad e_i = \text{Pearson residual}$$

If $\phi \neq 1$, all standard errors are off by $\sqrt{\phi}$

$\Rightarrow$ New (corrected) SE = (original SE) $\times \sqrt{\hat{\phi}}$

If we want to test LRT comparing two models:

$$\text{New (corrected) LR test stat} = \frac{LRT}{\hat{\phi}}$$

Other option: $Y \sim \text{Negbinomial}(r, p)$

$$P(Y=k) = \begin{cases} \binom{k+r-1}{r-1} p^r (1-p)^k & k = 0,1,2,\ldots \\ 0 & \text{else} \end{cases}$$

$$\log(\mu) = \underbrace{\beta_0 + \beta_1 x + \ldots}_{\text{linear predictor}} \qquad \mu = E(Y)$$

$$E(Y) = \frac{r(1-p)}{p} = \mu \qquad Var(Y) = \frac{r(1-p)}{p^2}$$

$$= \left[\mu + \left(\frac{1}{r}\right)\mu^2\right] \cdot \phi$$

$\phi = 1$

Model Selection & Goodness of Fit –

Residual deviance → shows up :

(1) Analysis of deviance test:

H_0 : Model_0
H_a : Model_1

Model_0 nested in Model_1

Test statistic = $Dev_{M_0} - Dev_{M_1}$

$$df = df_{M_0} - df_{M_1} = (\#\text{ coef } M_1) - (\#\text{ coef } M_0)$$

"Large sample"

Test stat $\dot\sim \chi^2(df)$ under H_0

(2) "Overall" Deviance test :

H_0 : $g(\mu) = \beta_0$
H_1 : our model

Test stat =
"null deviance" – "residual deviance"
df = null df – resid df

Under H_0, large samples, Test stat $\dot\sim \chi^2(df)$

(3) Goodness-of-fit test :

H_0 : our model
H_a : perfect model (saturated model)

Goal: large p-value

⟹ No evidence the perfect model is better than ours.

Test statistic = Residual deviance $\dot\sim \chi^2(n-p)$
very limited situations

(a) grouped binomial data > 5 obs. in each group

(b) Poisson data w/ relatively large (> 5) counts for each covariate pattern

→ continuous predictor → cond. not met