Inference

Poisson regression – Stat 230

Review: Poisson regression model

- If Y_i = # of events in a fixed interval of time/space, then $Y_i \sim \text{Poisson}(\lambda)$
- This leads to the Poisson regression model for counts:

$$\log(\lambda) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

or equivalently

Assumptions

1. **Linearity**: the log of the mean of Y, $\log(\lambda)$, must be a linear function of the predictors
2. **Mean = Variance**: the mean of a Poisson random variable must be equal to its variance
3. **Independence**: the observations must be independent

Types of inference

1. **Wald based**

   For large enough sample sizes, the sampling distributions for the regression coefficients are approximately normal

   $\Longrightarrow$ We can use intro stat-type tests/intervals

2. **Likelihood based**

   We can use the likelihood function (or deviance) to compare two fitted models

Wald test

Hypotheses: $H_0 : \beta_i = 0$ vs. $H_a : \beta_i \neq 0$

Test statistic:

$$z = \frac{\widehat{\beta}_i}{SE(\widehat{\beta}_i)}$$

Reference distribution: $N(0, 1)$

Poisson model for puffin nesting data

term	estimate	std.error	statistic	p.value
(Intercept)	2.120	0.473	4.479	<0.001
grass	−0.001	0.003	−0.202	0.8
soil	0.019	0.011	1.745	0.081
angle	0.008	0.013	0.609	0.5
distance	−0.027	0.015	−1.762	0.078

Wald interval

A normal-based confidence interval for β_i

$$\widehat{\beta}_i \pm z^*_{1-\alpha/2} SE(\widehat{\beta}_i)$$

But this is on the log scale, so we need to back transform!

$$\left(e^{\widehat{\beta}_i - z^*_{1-\alpha/2} SE(\widehat{\beta}_i)}, \ e^{\widehat{\beta}_i + z^*_{1-\alpha/2} SE(\widehat{\beta}_i)} \right)$$

Poisson model puffin nesting data

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	2.120	0.473	4.479	<0.001	1.185	3.041
grass	−0.001	0.003	−0.202	0.8	−0.007	0.006
soil	0.019	0.011	1.745	0.081	−0.003	0.040
angle	0.008	0.013	0.609	0.5	−0.017	0.034
distance	−0.027	0.015	−1.762	0.078	−0.057	0.003

$-0.027 \pm 1.96(0.015)$

$(e^{-0.0564} \approx 0.945, e^{-0.0024} \approx 1.002)$

Likelihood ratio test

Full model: $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \beta_{k+1} x_{k+1} + \cdots + \beta_p x_p$

Reduced model: $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$

Hypotheses: $H_0 : \beta_{k+1} = \cdots = \beta_p = 0 \quad \text{vs.} \quad H_a$: at least one $\beta_j \neq 0$

Test statistic:

$$LRT = \text{residual deviance(reduced model)} - \text{residual deviance(full model)}$$

Reference distribution: χ^2 distribution

d.f. = # βs in full model – # βs in reduced model

How big is “big enough”?

- Inference requires large enough sample sizes
- This is reasonable if n is “large”
- OR if $\widehat{\lambda}_i > 5$