Horseshoe Crabs Example (cont) 3/22/22

$\hat{\mu}$ = estimated mean # of satellites

Color = LM

Color = M

$\log(\hat{\mu}) = 3.571 - 0.0806\,x - 6.133 \cdot M - 8.515 \cdot DM - 10.543 \cdot D$

$+ 0.219 \cdot M \cdot x + 0.300\, DM \cdot x + 0.379\, D \cdot x$

x = width (cm)

$M = \begin{cases} 1 & \text{medium color} \\ 0 & \text{else} \end{cases}$

$DM = \begin{cases} 1 & \text{dark medium color} \\ 0 & \text{else} \end{cases}$

$D = \begin{cases} 1 & \text{dark color} \\ 0 & \text{else} \end{cases}$

Helpful: Write out fitted model for each color.

Interpret:

Mx: 0.219 → exponentiate: $e^{0.219} = 1.24$

Option 1 — how does color affect the effect of width on est. mean # of satellites?

The estimated change in mean # of satellites for a one cm increase in width is 24% higher for medium colored crabs compared to light medium.

$e^{-0.08087} \cdot e^{0.21942} = 1.149$

$e^{-0.08087} = 0.92$

For light medium crabs, we estimate the mean # of satellites decreases by about $\boxed{8\%}$ per cm increase in width.

Whereas, for medium crabs, the estimated mean # of satellites increases by 15% per cm increase in width.

Color FD : -10.5435

		Int.	Slope width
$\frac{\hat{\mu} \mid D}{\hat{\mu} \mid LM}$	LM	3.571	-0.081
	M		
	DM		
	D	$3.571 - 10.543$	$-0.081 + 0.379$

For crabs with width equal to zero cm, the estimated mean # of satellites for dark colored crabs is $e^{-10.543}$ times smaller than for light medium crabs.

Overdispersion = Actual variance $Var(Y)$ exceeds the specified GLM variance.

① Binomial data: $Y \sim Bin(n, \pi)$

$\Rightarrow E(Y) = n\pi = \mu$

$Var(Y) = n\pi(1-\pi)$ < actual variance?

② Poisson data: $Y \sim Pois(\mu)$

$\Rightarrow E(Y) = \mu$

$Var(Y) = \mu$ < actual $Var(Y)$

Why? ① Haven't accounted for important predictors.

② Correlated data

Signs of overdispersion?

① Compare sample means & sample variances in grouped data → Sample variances > Sample means

② Large residual deviance (goodness of fit) that can't be accounted for by other lack of model fit.

Adjust for overdispersion -

✗ ① Quasi-likelihood (quasi-binomial or quasi-Poisson)

② Negative binomial regression (two parameters - mean & variance can vary independently)

Quasi-Poisson: Assume $E(Y) = \mu$

- Goal: Estimate ϕ.

$Var(Y) = \phi \mu$ for some ϕ

"overdispersion parameter"

Pearson residuals: $e_i = \dfrac{Y_i - \hat{\mu}_i}{\sqrt{\hat{Var}(Y_i)}} = \dfrac{Y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i}}$

if Poisson

If $\phi = 1$ (no overdispersion),

$$E(e_i^2) \approx 1$$

If $\phi > 1$ (overdispersion), $E(e_i^2) \approx \phi$

→ Plot squared Pearson residuals against $\hat{\mu}_i$ - is average around 1?

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

p = # of coefs.