Laws of logarithms + exponentials (log → natural log)

$e^x \cdot e^y = e^{x+y}$

$e^{\log(x)} = x$ and $\log(e^x) = x$

$\log(x/y) = \log(x) - \log(y)$ and $\log(x \cdot y) = \log(x) + \log(y)$

$\log(1) = 0$

Wald approximation for calculating confidence intervals of predicted values (with offset):

$$\exp(X_i\hat{\beta} + \log(n_i)) \pm 1.96 \cdot \text{std. error})$$

95% CIs — std. error of the fitted value ⇒ se.fit in R

note: ignores uncertainty about $\hat{\beta}$ (?)

alternative: `rockchalk` package in R ↳ 'predictOMatic' function

estimated expected count given X_i:

$$\widehat{E[y_i|X_i]} = \exp(X_i\hat{\beta}) + \log(n_i)$$

AER library → 'dispersiontest' function

dispersiontest(m1, trafo=2)
↑ model object

equidispersion
↓
mean + variance of the dispersion are the same

interpretation of coefficients:

$\lambda_1 = \lambda_0 \exp(\beta_1)$ → mean of Y is multiplied by $\exp(\beta_1)$ when X is increased by 1 unit

(negative binomial has same link function so same interpretation)

model visualization ⇨ 'jtools' package

marginal results ↳ predictions

'emmeans' → estimate marginal means (may not allow for offset) most suited for factor variables

'ggeffects' ⇨ visualize estimated marginal means

slide 2D slice

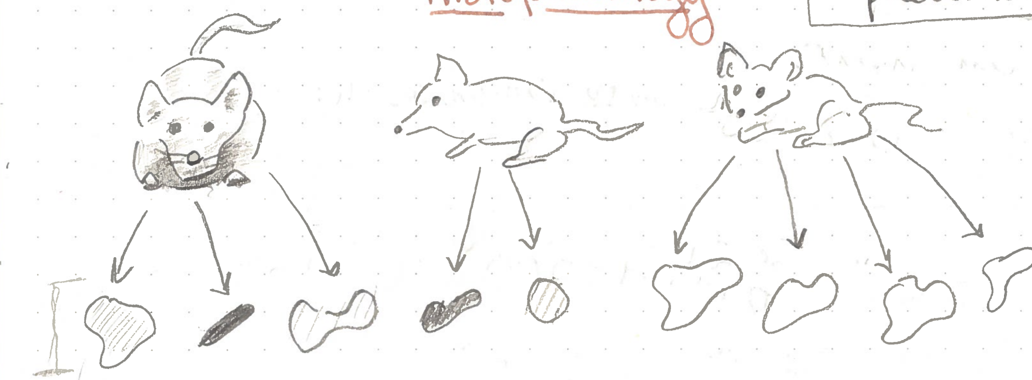
granuloma 3D

github.com/sarahcooper22/pathology-data

Histopathology

TB-associated pneumonia

multiple mice



each mouse has multiple lesions

lesion measures:

size:
- area
- # host cells
- # TB cells
- total cell count

host cell content:
- CD8 cells
- CD4 cells
- B cells

arrangement of host cells:
- clustered by type
- edge vs. center of lesion

overdispersion
↓
potentially use negative binomial
↳ glm.nb
↳ 'MASS' package

Poisson

$$E[y_i|X_i] = \lambda_i = e^{X_i\beta} = \exp(X_i\beta)$$

↑ ensures λ_i will always be positive

# of events — expected number of events, drawn from a Poisson distribution

GLM with "log link":

$$\log(E[y_i|X_i]) = X_i\beta$$

estimates expected count (predict type="link" se.fit=TRUE then exp(pred ± 1.96 se))

$$\mu_i = \widehat{E[y_i]} = \exp(X_i\hat{\beta})$$

expected # of events — estimated model coefficients

Rate model (adds offset)

$$\lambda_i = e^{X_i\beta + \log(n_i)}$$

offset (log of the "exposure") no coefficient fit

$$\underbrace{E[y_i/n_i]}_{\text{rate}} = e^{X_i\beta} \Rightarrow \frac{1}{n_i}E[y_i] = e^{X_i\beta} \Rightarrow E[y_i] = e^{X_i\beta} n_i$$

⇩ $x = e^{\log(x)}$

$$E[y_i] = e^{X_i\beta} e^{\log(n_i)}$$

⇩ $e^x \cdot e^y = e^{x+y}$

$$E[y_i] = e^{X_i\beta + \log(n_i)}$$

expected value is a linear operator: for any H, $E[Hy_i] = HE[y_i]$