Week 4: Binomial & Proportional Response

CHAPTER 3 EXERCISE 2: BINOMIAL FITS

Binomial Response

Y distribution

$Y|X \sim Bin(m, p)$ — m given

$y = 0, 1, 2, \ldots, m \leftarrow p \in [0, 1]$

Link Function

$p = \frac{e^{\eta}}{1 + e^{\eta}}$

$p \in [0, 1] \leftarrow \eta \in \mathbb{R}$

Linearity in X

$\eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \ldots$

$\eta \in \mathbb{R} \leftarrow X \in \mathbb{R}$

Does the model Fit?

THEORETICAL

If $Y|X \sim Bin(m, p)$ iid ③

where $p = \frac{e^{\eta}}{1 + e^{\eta}}$ and $\eta = \beta X$ ②

then $E[Y|X] = mp$ and ③
$Var(Y|X) = mp(1-p)$

If m is "sufficiently" large ①

⇓ THEN

$D \overset{approx}{\sim} \chi^2_{n-q}$

and

$E[D] = n - q$

Can google "chi-squared expected value proof"

Misspecification Slide 6

③ Check Overdispersion

② See if missing covariates, non-linear terms or interactions

① Check m_i is not too Small

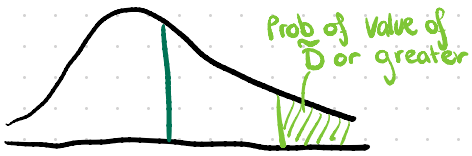
SAMPLE

Calculate $\tilde{D}$ using Data

⇓ TEST

Is $\tilde{D}$ a sample from χ^2_{n-q}?

TEST ⟹



R: 1 - Pchisq($\tilde{D}$, n-q)

Misspecification: A Closer Look

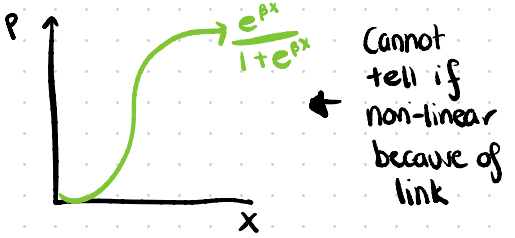
① Is the data Sparse?

- m is too small
- Texbook suggests $m_i \geq 5 \ \forall i$ (pg 53)

③ Outliers?

Can remove them and recheck the fit.

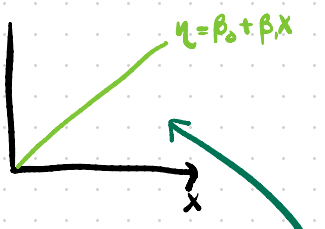
② Is $\eta = X\beta$ Correct?



Cannot tell if non-linear because of link

Instead plot X vs η

$\eta = \log\left(\frac{p}{1-p}\right)$



If this is non-linear then include quadratic X

③ Is there Overdispersion?

- Can come from a violation of the iid assumption
  - ① Trials are identical
  - ② Trials are independent
  - ③ We have assumed the correct distribution
- In the event of overdispersion we have

$$Var(Y_i) > mp(1-p)$$

Observed Variance > theoretical (expected) Variance

- How to check?

Above is equivalent to $Var(p_i) > p(1-p)/m$

Observed variance = $Var(p_i|X)$ (Calculate from data)

theoretical variance $\frac{\bar{p}(1-\bar{p})}{m}$ ← $\bar{p}$ is the mean of p from your data