<minimum notes + assignments only>

3.9.1 → Binomial response

→ Link functions

→ In previous sections, we use a link function to connect the mean response μ to the regression coefficients β. While that is a useful generic approach (i.e. not needing to specify a distribution), it is not strictly required. At minimum, we need a link function that connects the parameter of interest to β, i.e. $g(\theta) = x^T\beta$.

For a binomial response, recall that $\theta = p$. Therefore, the norm in this context is to consider link functions that are functions of p rather than $\mu = np$

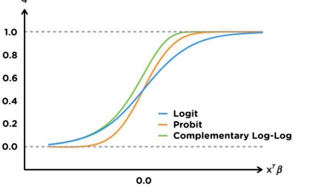
→ The following link functions are suitable candidates for a binomial response, given $0 \le p \le 1$

→

Function name	$g(p)$
logit link	$\ln\left(\frac{p}{1-p}\right)$
Probit link	$\Phi^{-1}(p)$
Complementary log-log link	$\ln[-\ln(1-p)]$

→ These are suitable b/c they restrict the range of $g^{-1}(x^T\beta) = p$ to between 0 + 1, where $x^T\beta$ is any real #.

$g(p) = x^T\beta$
$g^{-1}(g(p)) = g^{-1}(x^T\beta)$
$p = g^{-1}(x^T\beta)$



→ Model

→ Let $Y_1, \dots, Y_n \overset{ii}{\sim}$ Binomial (n_i, p_i), $\hookrightarrow = x_i^T\beta$, then we obtain the following simplified expressions:

→ $\mu_i = n_i p_i$
$\downarrow = n_i g^{-1}(x_i^T\beta)$

$\hat{\mu}_i = n_i g^{-1}(x_i^T\hat{\beta})$

→ Log likelihood function → $\ell(\beta) = \sum_{i=1}^{n}\left[y_i \ln\left(\frac{p_i}{1-p_i}\right) + n_i \ln(1-p_i) + \ln\binom{n_i}{y_i}\right]$

→ Deviance → $D = 2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) + (n_i - y_i)\ln\left(\frac{n_i - y_i}{n_i - \hat{\mu}_i}\right)\right]$

→ Pearson residual → $e_i^P = \frac{y_i - n_i\hat{p}_i}{\sqrt{n_i\hat{p}_i(1-\hat{p}_i)}}$

→ Pearson chi-square statistic

$\sum_{i=1}^{n}(e_i^P)^2 = \sum\left(\frac{y_i - n_i\hat{p}_i}{n_i\hat{p}_i(1-\hat{p}_i)}\right)$

→ A logistic regression model uses the logit link function

→ $g(p_i) = \ln\left(\frac{p_i}{1-p_i}\right)$
$\downarrow = x_i^T\beta \implies p_i = g^{-1}(x_i^T\beta)$
$\downarrow = \frac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}}$

→ New score function b/c have $g(p)$ rather than $g(\mu)$. Just memorize them.

(U + J)

$U_j = \sum_{i=1}^{n}(y_i - \mu_i)x_{i,j}$

$J = \sum_{i=1}^{n} n_i p_i(1-p_i)x_i x_i^T$

$$\downarrow = \begin{bmatrix} \sum_{i=1}^{n} m_i q_i(1-q_i) & \sum_{i=1}^{n} m_i q_i(1-q_i)x_{i,1} & \cdots & \sum_{i=1}^{n} m_i q_i(1-q_i)x_{i,p} \\ \sum_{i=1}^{n} m_i q_i(1-q_i)x_{i,1} & \sum_{i=1}^{n} m_i q_i(1-q_i)x_{i,1}^2 & \cdots & \sum_{i=1}^{n} m_i q_i(1-q_i)x_{i,1}x_{i,p} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} m_i q_i(1-q_i)x_{i,p} & \sum_{i=1}^{n} m_i q_i(1-q_i)x_{i,1}x_{i,p} & \cdots & \sum_{i=1}^{n} m_i q_i(1-q_i)x_{i,p}^2 \end{bmatrix}$$

weird notation
($m_i = n_i$ + $q_i = p_i$)

→ Interpretation of parameter estimates

→ model: $\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ (x_1: continuous, x_2: dummy var) $\xrightarrow{\beta} \ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = -1.7834 - 0.0108x_1 + 2.3628$

→ odds ratio is the ratio of the odds of an event w/ the presence of a characteristic to the odds of the same event w/o the presence of that characteristic. Intuitively, an odds ratio is expressed as a factor

$(\widehat{\text{odds}} \text{ w}/x_2 = 1) = \exp(\hat{\beta}_2)(\widehat{\text{odds}} \text{ w}/x_2 = 0)$

→ w/ is $\exp(\hat{\beta}_2)$ times larger vs w/o

→ for every 1 unit increase in X_1, the predicted odds of response change by a factor of $e^{-\hat{\beta}_1}$, assuming x_2 is the same

OR regular slope coefficient interpretation, but w/ ln(odds) instead.

→ Probit regression + complementary log-log regression

→ Prediction example → $g(p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$; $x_1 = 37$ + $x_2 = 1$

→ under logistic model → $\ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = -1.78 - 0.01(37) + 2.362(1)$
$\downarrow = 0.73$

$\implies \hat{p} = \frac{e^{0.73}}{1+e^{0.73}}$
$\downarrow = 0.6756$

→ under probit model → $\Phi^{-1}(p) = -0.75 - 0.01(37) + 1.43(1)$
$\downarrow = 0.4567$
$\implies p = \Phi(0.456)$
$\downarrow = 0.6772$

→ under comp log-log model → $\ln[-\ln(1-\hat{p})] = -1.54 - 0.01(37) + 1.89(1)$
$\downarrow = 0.12$

$\implies p = 1 - \exp[-\exp(0.12)]$
$\downarrow = 0.6762$

→ Example → Given outputs, find the following:
Pearson residual, deviance, estimated st error of intercept, LRT statistic
for $i = 3$ of model

→ $e_3^P = \frac{e_3}{\sqrt{V(Y_3)}} = \frac{Y_3 - n_3\hat{p}_3}{\sqrt{n_3\hat{p}_3(1-\hat{p}_3)}} = \frac{1 - 0.3222}{\sqrt{(1)0.3222(1-0.3222)}} = 1.451$

→ $D = 2\left[\ell(\hat{\beta}_{max}) - \ell(\hat{\beta})\right]$ → saturated model, use y_i as $\hat{\mu}_i$

$=$ <memorize formula> OR

$= 2\sum_{i=1}^{10}\left[y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) + (n_i - y_i)\ln\left(\frac{n_i - y_i}{n_i - \hat{\mu}_i}\right)\right]$

$\hookrightarrow = 1$ b/c Bernoulli

$=$ <...>
$= 11.262$

→ $\hat{\sigma}_{\hat{\beta}_0} = \sqrt{J^{-1}} \to \begin{bmatrix} \circ & \\ & \end{bmatrix}$

→ $LRT = 2\left[\ell(\hat{\beta}) - \ell_{null}\right]$ → intercept only
$\hookrightarrow$ fitted model

→ $\ell(\beta) = \sum_{i=1}^{10}\left[y_i \ln\left(\frac{p_i}{1-p_i}\right) + 1 \cdot \ln(1-p_i) + \cancel{\ln\binom{1}{y_i}}\right]$ <$n_i = 1$>
$= 1$

→ $\ell(\hat{\beta}) = \sum_{i=1}^{10}\left[y_i \ln\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right) + \ln(1-\hat{p}_i)\right]$
$= $ <...>
$= -3.631$

→ $\ell_{null} = $ <... above w/ $\hat{p}_i = 3/10 = \bar{y}$ for all i>
$\downarrow = -6.109$

→ Assignment

→ Q3) $D = 2\left[\ln(\hat{p}_{max}) - \ln(\hat{p})\right]$, $\ell(p) = \sum\left[\ln\binom{n_i}{y_i} + y_i \ln\left(\frac{p_i}{1-p_i}\right) + n_i \ln(1-p_i)\right]$

→ $\ell(\hat{p}_{max}) = \sum\left[\ln\binom{n_i}{y_i} + y_i \ln\left(\frac{y_i/n_i}{1-y_i/n_i}\right) + n_i \ln(1 - y_i/n_i)\right]$

$\hookrightarrow$ use $\hat{p}_i = y_i/n_i$
$=$ <excel>
$= -13.105$

→ $\ell(\hat{p}) = \sum\left[\ln\binom{n_i}{y_i} + y_i \ln\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right) + n_i \ln(1-\hat{p}_i)\right]$

$\hookrightarrow$ use $y_i = $ round$(n_i\hat{p}_i, 0)$
$=$ <excel>
$= -18.7187$

$\implies D = 2\ln(-13.105 - (-18.7187))$
$\downarrow = 11.2393$

?

(come back to SF + SW)
nominal / ordinal regression

G