L12. Logistic Regression

- Inferences of logistic regression rely on large sample size.

1. Wald test for individual coefficient.

   - When sample size is large, the MLE

     $\hat{\beta}_K \overset{Appr}{\sim} N(\beta_K, Var(\hat{\beta}_K))$

   - Let G denote the matrix of second order partial derivatives of $l(\vec{\beta}) = \ln L(\vec{\beta})$

$$G = \begin{bmatrix} \frac{\partial^2 l}{\partial \beta_0^2} & \frac{\partial^2 l}{\partial \beta_0 \partial \beta_1} & \cdots & \frac{\partial^2 l}{\partial \beta_0 \partial \beta_{p-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l}{\partial \beta_{p-1} \partial \beta_0} & \frac{\partial^2 l}{\partial \beta_{p-1} \beta_1} & \cdots & \frac{\partial^2 l}{\partial \beta_{p-1}^2} \end{bmatrix}$$

then $Var(\vec{b}) = [-G]^{-1} \Big|_{\vec{\beta} = \vec{b}}$

when $\vec{b}$ is the MLE of $\vec{\beta}$.

Then, $\frac{\hat{\beta}_K - \beta_K}{se(\hat{\beta}_K)} \overset{Appr}{\sim} N(0,1)$, $k = 0, 1, \cdots, p-1$

— Then a large sample test for β_K can be constructed :

$$H_0: \beta_K = 0 \quad v.s. \quad H_1: \beta_K \neq 0$$

with test stat $Z^*_{stat} = \frac{\hat{\beta}_K}{se(\hat{\beta}_K)}$

and rejection rule : $|Z^*_{stat}| \geq Z_{\alpha/2}$.

— C.I. for β_K: $\hat{\beta}_K \pm Z_{\alpha/2}\, se(\hat{\beta}_K)$

C.I. for e^{β_K}: $\exp[\hat{\beta}_K \pm Z_{\alpha/2}\, se(\hat{\beta}_K)]$

II. Deviance and Likelihood Ratio Test for reduced and full models.

If to compare two models ~~use~~ with different number of parameters : H_0: reduced model

H_1: reduced model $+ \beta_{K+1} X_{K+1} + \cdots$

or

$H_0: \beta_{k+1} = \beta_{k+2} = \cdots = \beta_n = 0$

H_1: at least one of them is not zero.

This test is carried out by comparing the log likelihood of models under H_0 and H_1.

- Deviance is defined as the difference of likelihood between the fitted model and the saturated model.

$$\text{Deviance} = \cancel{-2 \ln l(\vec{\beta})}$$

$$-2\, l(\vec{b}) + 2\, \underbrace{l(\text{saturated})}_{\text{can be proved } = 1}$$

Deviance is simplified as $-2\, l(\vec{b})$, denoted as G^2

- Then for the test above, define the test stat to be

$$\Delta G^2 = G^2(\text{reduced model}) - G^2(\text{full model})$$

$$= 2\, l(\vec{b})_{\text{-full}} - 2\, l(\vec{b})_{\text{-reduced}}$$

Rejection rule: If ΔG^2 is significantly large:

$$\Delta G^2 \geq \chi^2_{\alpha,\, df = P_{full} - P_{reduced}}$$

then the full model is significantly improved in the likelihood.

- Recall in logistic regression,

$$\ell = \sum_{i=1}^{n}\left[y_i \cdot \ln\left(\frac{\pi_i}{1-\pi_i}\right) + \ln(1-\pi_i)\right]$$

so the deviance of any fitted model is

$$G^2 = -2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right) + \ln(1-\hat{\pi}_i)\right]$$

III. Deviance Residuals

The OLS residuals don't apply in the logistic case:

Observed response value: $y_i = \begin{cases} 0 \\ 1 \end{cases}$

Predicted response value: $\hat{P}_r(Y_i = 1) = \hat{\pi}_i$

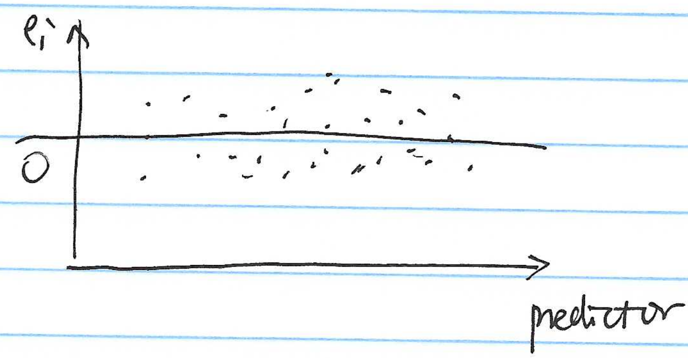
If we still want evaluate the "residuals", there are two modified versions:

(i) Pearson Residuals — similar to OLS studentized residuals.

$$e_i = \frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)}}$$

If $Y_i = 1$, $\hat{\pi}_i \to 1$.

We would observe the Pearson residuals to be distributed horizontally across the 0 line.



(ii) Deviance Residuals.

recall the deviance $G^2 = -2\ln \ell(\vec{b})$

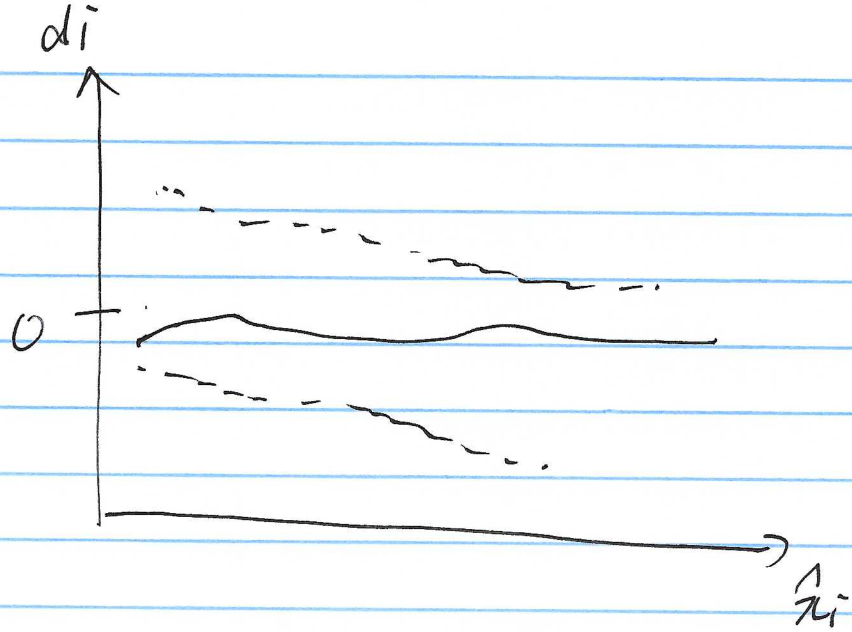
$$= -2\sum_{i=1}^{n} y_i * \ln\left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right) + \ln(1-\hat{\pi}_i)$$

~~deviance residual~~ $\cancel{d_i = \begin{cases} \sqrt{G^2} & \text{when } y_i=1 \\ -\sqrt{G^2} & \text{when } y_i=0 \end{cases}}$

$$d_i = \begin{cases} \sqrt{-2\left(y_i \ln \frac{\hat{\pi}_i}{1-\hat{\pi}_i} + \ln(1-\hat{\pi}_i)\right)} & y_i=1 \\ -\sqrt{-2\left(y_i \ln \frac{\hat{\pi}_i}{1-\hat{\pi}_i} + \ln(1-\hat{\pi}_i)\right)} & y_i=0 \end{cases}$$

Therefore, the square of each deviance residual measures the contribution of each response to the deviance of the fitted model.

we can again check the center of d_i's — lowess smooth curve. (Not required)



- Pseudo R^2: because there is ~~on~~ no OLS principle behind the logistic model, we don't have the regular R^2 to "explain the variance" either.

  There're some "Pseudo R^2) that can suggest the goodness of fit:

Efron's pseudo $R^2 = 1 - \dfrac{\Sigma(y_i - \hat{\pi}_i)^2}{\Sigma(y_i - \bar{y})^2}$

McFadden's pseudo $R^2 = 1 - \dfrac{l\,(\text{Full})}{l\,(\text{Null})}$

Note: Model selection.

Recall AIC and BIC are defined based on likelihood and number of parameters, they can still be used (prefered) as criteria for model selection in logistic regression.