Logistic Regression (part II)

To perform inference in Logistic Reg, we must rely on asymptotic results (i.e. using a large sample approximation).

$\boxed{\text{I}}$ Wald test for indiv coefficients.

<u>Main Idea</u>: When n is large, the MLE

$$\hat{\beta}_k \overset{\text{(approx)}}{\dot{\sim}} N(\beta_k, Var(\hat{\beta}_k))$$

where $Var(\hat{\beta}_k)$ is defined as follows:

Let G denote the Hessian matrix of the loglikelihood:

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

then the variance of the MLE is:

$$Var(\hat{\beta}) = G^{-1} \Big|_{\beta = \hat{\beta}}$$

&

$$Var(\hat{\beta}_k) = \left[G^{-1} \Big|_{\beta = \hat{\beta}} \right]_{k+1, k+1}$$

where $\hat{\beta}$ is the MLE of β.

OK! So then

$$\frac{\hat{\beta}_k - \beta_k}{SE(\hat{\beta}_k)} \dot{\sim} N(0,1) \quad \&$$

a large sample test for β_k can be constructed as:

$H_0: \beta_k = 0$ vs $H_1: \beta_k \neq 0$ w/

The test statistic:

$$Z = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)}$$

& the decision rule:

if $|Z| > Z^*_{1-\alpha/2}$, reject H_0.

VII Deviance & Likelihood Ratio Tests for Reduced & Full Models.

If we want to compare two models w/ a different # of predictors, for ex:

$H_0: \beta_r = \beta_{r+1} = \cdots = \beta_{p-1} = 0$

vs.

$H_1:$ at least one β_j, $j = r, \ldots, p-1$, not zero

$\Updownarrow$

H_0: Reduced model $\Rightarrow \text{logit}(\hat{\pi}_i) = \beta_0 + \sum_{j=1}^{r-1} \beta_j X_{ji}$

H_1: Full model $\Rightarrow \text{logit}(\hat{\pi}_i) = \beta_0 + \sum_{j=1}^{p-1} \beta_j X_{ji}$

Our test is based on comparing the log-likelihoods across H_0 & H_1.

Deviance

Def: Deviance $= -2\ell(\beta)$

Then a likelihood-based test stat is defined as:

$$\Delta D = \text{Deviance (Reduced)} - \text{Deviance (Full)}$$

Decision Rule:

If

$$\Delta D > \chi^2_{df,\,1-\alpha}\text{, reject } H_0$$

Here $df = P_{Full} - P_{Red}$.

If we reject H_0, we have evidence that the full model is a significantly better fit according to likelihood.

Recall, for logistic reg,

$$l = \sum_{i=1}^{n} \left(Y_i \log\left(\frac{\pi_i}{1-\pi_i}\right) + \log(1-\pi_i) \right)$$

So deviance of a fitted model is:

$$Dev(\hat{\beta}) = -2\sum_{i=1}^{n} \left(Y_i \log\left(\hat{\pi}_i / 1-\hat{\pi}_i\right) + \log(1-\hat{\pi}_i) \right)$$

$\boxed{\text{III}}$ Deviance Residuals

OLS residuals don't apply to logistic reg.

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

predicted response: $\hat{\pi}_i = \hat{P}(Y_i = 1)$

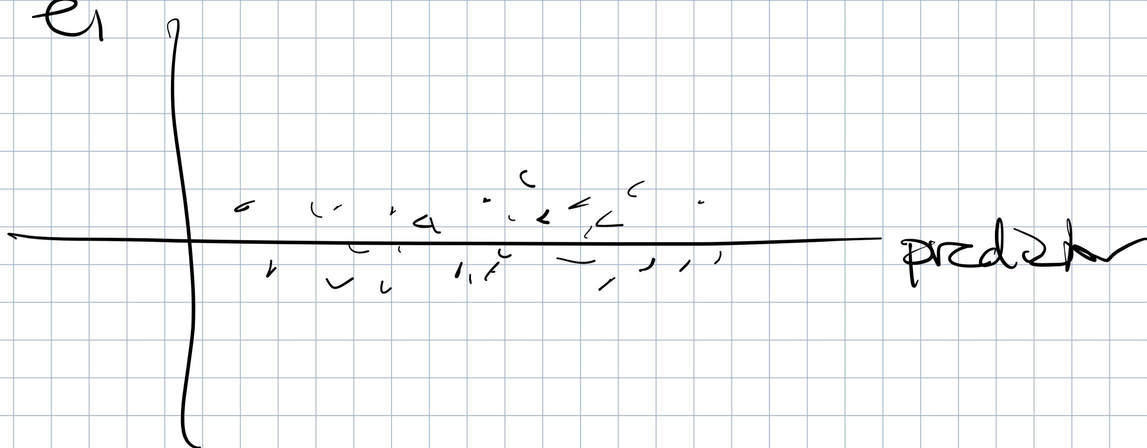
If we want to have an analog of "residuals" there are 2 alternatives:

① "Pearson Residuals"

— similar to studentized resid in OLS

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

We would like to observe Pearson resids scattered randomly & evenly across the x-axis:



↳ interpret like usual residual plots!!

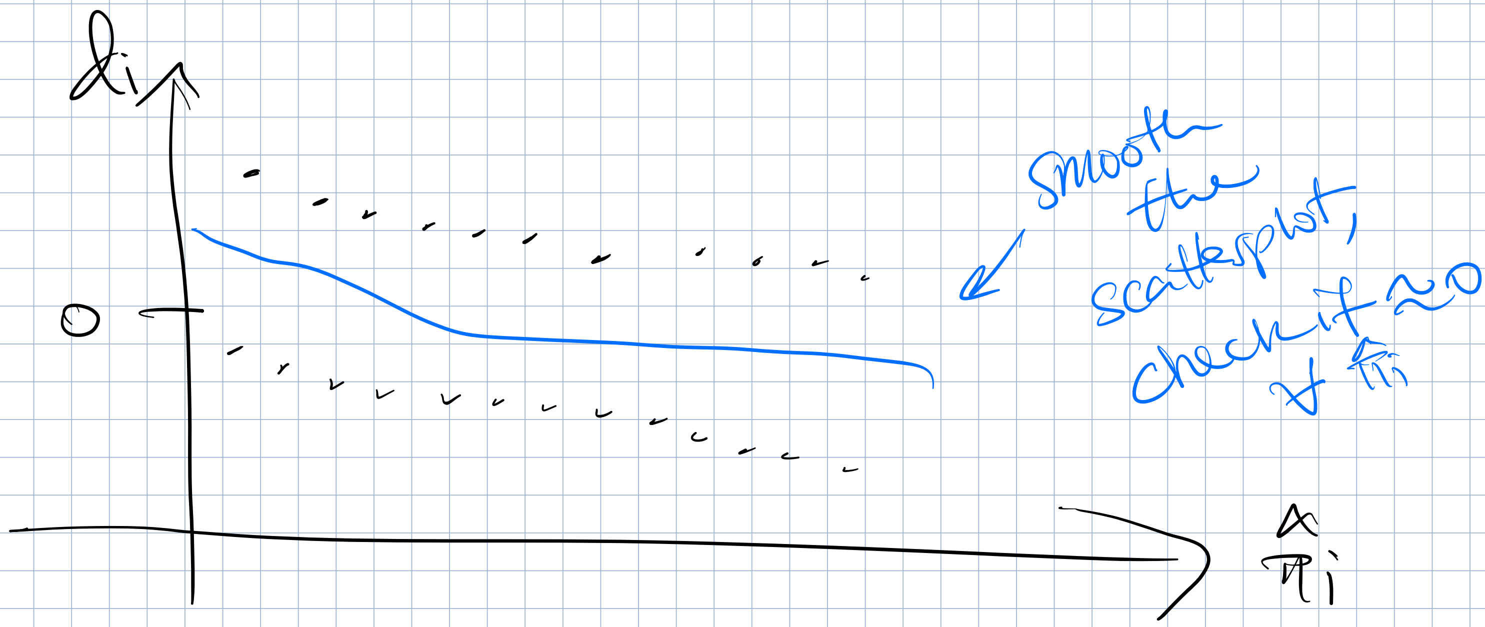
② Deviance Residuals

Define the deviance residuals as:

$$d_i = \begin{cases} \sqrt{-2\left(y_i \log\left(\hat{\pi}_i / 1-\hat{\pi}_i\right) + \log(1-\hat{\pi}_i)\right)} & \text{if } y_i = 1, \\ -\sqrt{-2\left(y_i \log\left(\hat{\pi}_i / 1-\hat{\pi}_i\right) + \log(1-\hat{\pi}_i)\right)} & \text{if } y_i = 0. \end{cases}$$

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

We can check the dev. resid scatter plot:



- Pseudo R^2:

Because there is no OLS principle, the regular R^2 doesn't exist for logistic reg. to "explain variance".

Some pseudo R^2 statistics have been invented to assess goodness-of-fit:

Efron:
$$\text{pseudo } R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{\pi}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

McFadden: pseudo $R^2 = 1 - \frac{\ell(\text{Full})}{\ell(\text{Null})}$

Note: AIC / BIC can still be used for model selection w/ logistic reg.