Example: GLM, Inference; Cell differentiation

We keep using the cell differentiation dataset.

Outcome: cell count

Covariates: TNF (tumor necrosis factor), IFN (interferon)

(a) Consider the following poisson regression model:

$log(\lambda_i) = \beta_0 + \beta_1 TNF + \beta_2 IFN + \beta_3 TNF \times IFN$

Obtain following statistics and save them in a SAS Dataset: (Standarized) Pearson residuals, (Standarized) Deviance residuals, Leverages and Cook's Distances.

- Pearson Residuals:

$$\widehat{r}_i^P = \frac{Y_i - \widehat{\mu}_i}{\widehat{V}(Y_i)^{1/2}} = \frac{Y_i - \widehat{\lambda}_i}{\lambda_i^{1/2}}$$

- Deviance Residuals:

$$\begin{aligned} D_i &= 2\left[Y_i(\widetilde{\theta}_i - \widehat{\theta}_i) - \{b(\widetilde{\theta}_i) - b(\widehat{\theta}_i)\}\right] \\ &= 2\left[Y_i \ \log\frac{Y_i}{\widehat{\lambda}_i} - (Y_i - \widehat{\lambda}_i)\right] \end{aligned}$$

$$\widehat{r}_i^D = sign(Y_i - \widehat{\lambda}_i)\sqrt{|D_i|}$$

- Leverage :
  - Hat matrix: $H = V^{1/2}X(X^TVX)^{-1}X^TV^{1/2}$
  - Leverage h_{ii} is the ith diagonal element of H

- Standarized Pearson and Deviance residuals:

$$\widehat{r}_i^{PS} = \frac{\widehat{r}_i^P}{\sqrt{1-h_{ii}}}; \quad \widehat{r}_i^{DS} = \frac{\widehat{r}_i^D}{\sqrt{1-h_{ii}}}$$

- Cook's distance:

$$D_i = \frac{1}{q}\left(\frac{h_{ii}}{1-h_{ii}}\right)(r_i^{PS})^2$$

(b) In this time, obtain them using proc genmod.

- See the SAS code

(c) Obtain deviance and Pearson's X^2, and assess the GoF of the model.

- Deviance: 5.09
- Pearson's X^2: 4.94
- Both Deviance/ DF and Pearson's X^2 / DF are not hugely different from one.

- GOF test
  - H_0: Model fits the data well
  - Test statistics (use the Pearson's X^2): D= 4.94
  - Under the NULL, D follows χ^2_{12}
  - Since $D <= \chi^2_{12} = 21.026$, we cannot reject H_0 at level 0.05.

(d) Plot the leverages and Cook's distances. Are there any high leverage or high influence observations?

- Observation 16 has $h_{ii} = 0.97 > 2q/n = 0.5$ and Cook's distance $= 1.49 > 1$
- High leverage and high influence point.

(e) Obtain VIF for each covariate.

- To get VIF, you first need to obtain the weights (diagonal element of V) using proc genmod (or IML) and use them in proc reg for the weighted least squares. Please see the SAS code.
- VIF
  - β_1 VIF: 1.68
  - β_2 VIF: 2.11
  - β_3 VIF: 2.81

  Some covariates are moderately correlated. But no serious multicollinearity problem.

(f) Carry out the LRT test for $H_0 : \beta_2 = \beta_3 = 0$ using Deviance

- H_0: $\beta_2 = \beta_3 = 0$ vs H_1: $\beta_2 \neq 0$ or $\beta_3 \neq 0$
- From the proc genmod, $D_0^* = 9.4281$ and $D_1^* = 5.0915$. LRT test statistic is

$$X_L^2 = D_0^* - D_1^* = 4.3366$$

- P-value is 0.1143.