Example: Overdispersion

We return to the clinical trial analyzed in the first part of the course; to recap ...

A clinical trial was conducted in order to evaluate the impact of Progabide on the frequency of epileptic seizures. Patients were randomized to either receive or not receive Progabide. The data set contains information on:

- age at start of study (AGE; measured in years)
- baseline seizure count; defined as the number of seizures in the 8 weeks prior to the study's commencement (BASE)
- treatment indicator (Z; 1=treated, 0=placebo)
- seizure counts in each of 4 two-week periods (Y1, Y2, Y3, Y4)

The investigators define the outcome as total post treatment seizure count: $Y_i \equiv \sum_{j=1}^{4} Y_{ij}$.

(a) Read in the data file and calculate correlations among $Y1, Y2, Y3, Y4$

- See SAS Code
- There are very strong positive correlations among $(Y1, Y2, Y3, Y4)$. For example, $corr(Y1, Y2) = 0.87$.

(b) When $Y_i \equiv \sum_{j=1}^{4} Y_{ij}$ is used for the outcome in poisson regression, do you think there will be an overdispersion problem?

- In this data, $(Y1, Y2, Y3, Y4)$ are not independent. Dependency among them can induce overdispersion.
- Suppose $Y_{ij} \sim Poisson(\lambda_{ij})$, then

$$E(Y_i) = E(\sum_{j=1}^{4} Y_{ij}) = \sum_{j=1}^{4} \lambda_{ij}$$

$$Var(Y_i) = Var(\sum_{j=1}^{4} Y_{ij})$$

$$= \sum_{j=1}^{4} \lambda_{ij} + 2 \sum_{j<k} Cov(Y_{ij}, Y_{ik})$$

When $Cov(Y_{ij}, Y_{ik}) > 0$ for all (j, k),
$Var(Y_i) > E(Y_i)$.

(c) Estimate the covariate-adjusted treatment effect through a Poisson regression model. Do you have an evidence of overdispersion?

- Model

$$\begin{aligned} log(\lambda_i) &= \beta_0 + \beta_1 Age + \beta_2 Base + \beta_3 Z \\ Y_i &\sim Poisson(\lambda_i) \end{aligned}$$

- Since $Deviance/DF = 10.18 >> 1$, we can conclude that there exists an overdispersion problem.

(d) Re-estimate (and re-test) the treatment effect by estimating the scale parameter (quasi-likelihood)

- Number of parameters (including intercept): 4

$\widehat{\phi} = \frac{D}{59-4} = 10.18$

$\sqrt{\widehat{\phi}} = \sqrt{10.179} = 3.19$

$\widehat{\beta}_3$ is not changed, but the standard error estimate of $\widehat{\beta}_3$ is changed.

- SE of β_3:

$$\widehat{SE}(\widehat{\beta}_3) = 0.0465 * 3.19 = 0.148$$

(e) Carry out a Wald test for the age and treatment effects using quasi-likelihood.

- H_0: $\beta_1 = \beta_3 = 0$ vs H_1: $\beta_1 \neq 0$ or $\beta_3 \neq 0$
- Contrast matrix

$$C = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

- Test statistic:

$$X_w^2 = \{C\widehat{\beta}\}^T \{C\widehat{\phi} I(\widehat{\beta})^{-1} C^T\}^{-1} \{C\widehat{\beta}\},$$

- In this data, $X_w^2 = 5.34$ and the corresponding

p-value=0.069. We cannot reject the null hypothesis.

(f) Carry out a likelihood ratio test for the age and treatment effects by fitting full and reduce models. Make a comment on the validity of the test result.

- Full model:

$$log(\lambda_i) = \beta_0 + \beta_1 Age + \beta_2 Base + \beta_3 Z$$

- Reduced Model:

$$log(\lambda_i) = \beta_0 + \beta_2 Base$$

From SAS output $l_{full} = 555.85$ and $|_{reduced} = 522.99$.

$$X_l^2 = 2 * (555.85 - 522.99) = 65.72$$

- LRT test statistic is too large. Does not seem to valid.

(g) Carry out the likelihood ratio test, in this time, using the same scale parameter (estimated from the full model) for both full and reduced models.

With fixing scale=3.1905, $|_{reduced} = 553.19$.

$$X_l^2 = 2 * (555.85 - 553.19) = 5.32$$

- Similar to the Wald test statistics

(h) In this time, carry out both Wald and LRT using the contrast statement.

See the SAS code.

(e) Re-estimate (and re-test) the treatment effect using the negative binomial regression.

- $\widehat{\beta}_3 = -0.2112$
- $p - value$: 0.1681

(g) Carry out LRT for the age and treatment effects by fitting the full and reduce models.

- H_0: $\beta_1 = \beta_3 = 0$ vs H_1: $\beta_1 \neq 0$ or $\beta_3 \neq 0$
- From SAS output $l_{full} = 5846.28$ and $|_{reduced} = 5844.59$.

  $X_l^2 = 2*(5846.28-5844.59) = 3.38 < 5.99 = \chi^2_{2,0.05}$
- Cannot reject H0 at $\alpha = 0.05$

(f) Re-estimate (and re-test) the treatment effect using GEE.

- For GEE fit we use the same variance structure in poisson GLM.

$$Var(Y_i) = V_i = a(\phi)v(\mu_i) = \mu_i$$

Since $D_i = \partial\mu_i/\partial\beta_i^T = 1/g'(\mu_i)X_i^T = v(\mu_i)X_i^T$, GEE equation is

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} D_i^T V_i^{-1}(Y_i - \mu_i) = \sum_{i=1}^{n} X_i(Y_i - \mu_i)$$

- $\widehat{\beta}$ from GEE is the same as MLE

- Sandwich estimator of variance:

$$Var(\widehat{\beta}) \quad = \quad H_1(\widehat{\beta})^{-1} H_2(\widehat{\beta}) H_1(\widehat{\beta})^{-1}$$

$$
\begin{aligned}
H_1(\widehat{\beta}) &= \sum_{i=1}^{n} D_i^T V_i^{-1} D_i = \sum_{i=1}^{n} X_i X_i^T \widehat{\mu}_i \\
H_2(\widehat{\beta}) &= \sum_{i=1}^{n} D_i^T V_i^{-1} Var(Y_i) V_i^{-1} D_i \\
&= \sum_{i=1}^{n} X_i X_i^T (Y_i - \widehat{\mu}_i)^2
\end{aligned}
$$