Chapter 4 Part 6 Case study

05/03/2021

Residuals and Deviance Residuals

- Residual plots may provide some insight into Poisson regression models, especially linearity and outliers, although the plots are not quite as useful here as they are for linear least squares regression.
- There are a few options for computing residuals and predicted values. Residuals may have the form of residuals for LLSR models or the form of deviance residuals which, when squared, sum to the total deviance for the model.

Residuals and Deviance Residuals

- Predicted values can be estimates of the counts, $e^{\beta_0+\beta_1 X}$, or log counts, $\beta_0 + \beta_1 X$. We will typically use the deviance residuals and predicted counts.
- The residuals for *linear least squares regression* have the form:

$$\begin{aligned}\text{LLSR residual}_i &= \text{obs}_i - \text{fit}_i \\ &= Y_i - \hat{\mu}_i \\ &= Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i) \end{aligned} \tag{1}$$

- Residual sum of squares (RSS) are formed by squaring and adding these residuals, and we generally seek to minimize RSS in model building.

Residuals and Deviance Residuals

- We have several options for creating residuals for Poisson regression models.
- One is to create residuals in much the same way as we do in LLSR. For Poisson residuals, the predicted values are denoted by $\hat{\lambda}_i$; they are then standardized by dividing by the standard error, $\sqrt{\hat{\lambda}_i}$.
  - These kinds of residuals are referred to as **Pearson residuals**.

$$\text{Pearson residual}_i = \frac{Y_i - \hat{\lambda}_i}{\sqrt{\hat{\lambda}_i}}$$

- Pearson residuals have the advantage that you are probably familiar with their meaning and the kinds of values you would expect.
- For example, after standardizing we expect most Pearson residuals to fall between -2 and 2.

Deviance Residuals

- **Deviance residuals** have some useful properties that make them a better choice for Poisson regression.
- First, we define a **deviance residual** for an observation from a Poisson regression:

$$\text{deviance residual}_i = \text{sign}(Y_i - \hat{\lambda}_i)\sqrt{2\left[Y_i log\left(\frac{Y_i}{\hat{\lambda}_i}\right) - (Y_i - \hat{\lambda}_i)\right]}$$

where $\text{sign}(x)$ is defined such that:

$$\text{sign}(x) = \begin{cases} 1 & \text{if} \quad x > 0 \\ -1 & \text{if} \quad x < 0 \\ 0 & \text{if} \quad x = 0 \end{cases}$$

Deviance Residuals

- As its name implies, a deviance residual describes how the observed data deviates from the fitted model.
- Squaring and summing the deviances for all observations produces the **residual deviance** $= \sum(\text{deviance residual})_i^2$.
- Relatively speaking, observations for good fitting models will have small deviances; that is, the predicted values will deviate little from the observed.
- However, you can see that the deviance for an observation does not easily translate to a difference in observed and predicted responses as is the case with LLSR models.

Deviance Residuals

- A careful inspection of the deviance formula reveals several places where the deviance compares Y to $\hat{\lambda}$
- The sign of the deviance is based on the difference between Y and $\hat{\lambda}$, and under the radical sign we see the ratio $Y/\hat{\lambda}$ and the difference $Y - \hat{\lambda}$.
- When $Y = \hat{\lambda}$, that is, when the model fits perfectly, the difference will be 0 and the ratio will be 1 (so that its log will be 0).
- So like the residuals in LLSR, an observation that fits perfectly will not contribute to the sum of the squared deviances.
- This definition of a deviance depends on the likelihood for Poisson models. Other models will have different forms for the deviance depending on their likelihood.
- Let's use the deviance residuals in R

Deviance Residuals

- What do we see in this plot?
- We want there to be no pattern as with linear regresson regular residuals
- Incorrect link function

Goodness of Fit Again (Connects better now)

- Click HERE

Overdispersion

- **Overdispersion** suggests that there is more variation in the response than the model implies.
- Under a Poisson model, we would expect the means and variances of the response to be about the same in various groups.
- Without adjusting for overdispersion, we use incorrect, artificially small standard errors leading to artificially small p-values for model coefficients.
  - We may also end up with artificially complex models.

Overdispersion

- Two ways to try handeling this
  - We can use an estimated dispersion factor to inflate standard errors.
  - We can use a negative-binomial regression model.

Dispersion Parameter

- We can estimate a dispersion parameter, ϕ, by dividing the model deviance by its corresponding degrees of freedom; i.e., $\hat{\phi} = \frac{\sum(\text{Pearson residuals})^2}{n-p}$ where p is the number of model parameters.
- It follows from what we know about the χ^2 distribution that if there is no overdispersion, this estimate should be close to *one*.
- It will be larger than *one* in the presence of overdispersion.

Dispersion Parameter - Quasilikelihood

- We inflate the standard errors by multiplying the variance by ϕ, so that the standard errors are larger than the likelihood approach would imply; i.e., $SE_Q(\hat{\beta}) = \sqrt{\hat{\phi}} * SE(\hat{\beta})$, where Q stands for "quasi-Poisson" since multiplying variances by ϕ is an ad-hoc solution.
- Our process for model building and comparison is called **quasilikelihood**—similar to likelihood but without exact underlying distributions.
- If we choose to use a dispersion parameter with our model, we refer to the approach as quasilikelihood.
- Let's illustrate in R

Comments on our Example

- In the absence of overdispersion, we expect the dispersion parameter estimate to be 1.0.
- The estimated dispersion parameter here is much larger than 1.0 (4.447) indicating overdispersion (extra variance) that should be accounted for.
- The larger estimated standard errors in the quasi-Poisson model reflect the adjustment.
- For example, the standard error for the West region term from a likelihood based approach is 0.7906, whereas the quasilikelihood standard error is $\sqrt{4.47} * 0.7906$ or 1.6671. This term is no longer significant under the quasi-Poisson model.

Comments on our Example

- In fact, after adjusting for overdispersion (extra variation), none of the model coefficients in the quasi-Poisson model are significant at the .05 level! This is because standard errors were all increased by a factor of 2.1 ($\sqrt{\hat{\phi}} = \sqrt{4.447} = 2.1$), while estimated coefficients remain unchanged.

t-Tests

- The tests for individual parameters are now based on the t-distribution rather than a standard normal distribution, with test statistic $t = \frac{\hat{\beta}}{SE_Q(\hat{\beta})}$ following an (approximate) t-distribution with $n - p$ degrees of freedom if the null hypothesis is true ($H_O : \beta = 0$).

Drop in Deviance

- Drop-in-deviance tests can be similarly adjusted for overdispersion in the quasi-Poisson model.
- In this case, you can divide the test statistic (per degree of freedom) by the estimated dispersion parameter and compare the result to an F-distribution with the difference in the model degrees of freedom for the numerator and the degrees of freedom for the larger model in the denominator.
- That is, $F = \frac{\text{drop in deviance}}{\text{difference in df}} / \hat{\phi}$ follows an (approximate) F-distribution when the null hypothesis is true (H_0: reduced model sufficient).
- Now let's test for an interaction between region and type of institution after adjusting for overdispersion (extra variance):

Our Results

- After adjusting for overdispersion, we still have statistically significant evidence ($F = 4.05, p = .0052$) that the difference between colleges and universities in violent crime rate differs by region.

No Dispersion vs. Overdispersion

This table summarizes the comparison between Poisson inference (tests and confidence intervals assuming no overdispersion) and quasi-Poisson inference (tests and confidence intervals after accounting for overdispersion).

No Dispersion vs. Overdispersion

Table 1: Comparison of Poisson and quasi-Poisson inference.

	Poisson	quasi-Poisson
Estimate	$\hat{\beta}$	$\hat{\beta}$
Std error	$SE(\hat{\beta})$	$SE_Q(\hat{\beta}) = \sqrt{\hat{\phi}}SE(\hat{\beta})$
Wald-type test stat	$Z = \hat{\beta}/SE(\hat{\beta})$	$t = \hat{\beta}/SE_Q(\hat{\beta})$
Confidence interval	$\hat{\beta} \pm z^{'}SE(\hat{\beta})$	$\hat{\beta} \pm t^{'}SE_Q(\hat{\beta})$
Drop in deviance test	$\chi^2 = \text{resid dev(reduced) - resid dev(full)}$	$F = (\chi^2/\text{difference in df})/\hat{\phi}$

Practice

- We need to practice this but also do a little review based on exams
- Let us start with exercise 14 conceptual in chapter 4
- Next the first two parts of 15
- Then in guided 2, do I