Diagnostics in Poisson Regression

Models - Residual Analysis
Outline

• Diagnostics in Poisson Regression Models - Residual Analysis

• Example 3: Recall of Stressful Events continued
Residual Analysis

• Residuals represent variation in the data that cannot be explained by the model.
• Residual plots useful for discovering patterns, outliers or misspecifications of the model. Systematic patterns discovered may suggest how to reformulate the model.
• If the residuals exhibit no pattern, then this is a good indication that the model is appropriate for the particular data.
Types of Residuals for Poisson Regression Models

• **Raw residuals:**

\[ O_i - E_i \]

• **Pearson residuals:** (or standardised)

\[ \frac{O_i - E_i}{\sqrt{E_i}} \]

• **Adjusted residuals** (preferred):

\[ \frac{O_i - E_i}{\text{SD}(O_i - E_i)} \]

\[ \text{SD}(O_i - E_i) = \sqrt{E_i(1 - E_i/n)} \]
• **If** $H_0$ **is true**, the adjusted residuals have a standard normal $N(0,1)$ distribution for large samples.

• Back to the “Recall of Stressful Events” example...
## Example 3: Recall of Stressful Events Data

<table>
<thead>
<tr>
<th>month</th>
<th>adjusted residual</th>
<th>month</th>
<th>adjusted residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.46</td>
<td>10</td>
<td>0.66</td>
</tr>
<tr>
<td>2</td>
<td>1.02</td>
<td>11</td>
<td>-0.42</td>
</tr>
<tr>
<td>3</td>
<td>2.10</td>
<td>12</td>
<td>0.30</td>
</tr>
<tr>
<td>4</td>
<td>3.18</td>
<td>13</td>
<td>1.02</td>
</tr>
<tr>
<td>5</td>
<td>-1.14</td>
<td>14</td>
<td>-1.86</td>
</tr>
<tr>
<td>6</td>
<td>1.02</td>
<td>15</td>
<td>-0.78</td>
</tr>
<tr>
<td>7</td>
<td>0.66</td>
<td>16</td>
<td>-2.58</td>
</tr>
<tr>
<td>8</td>
<td>-1.50</td>
<td>17</td>
<td>-2.58</td>
</tr>
<tr>
<td>9</td>
<td>-0.06</td>
<td>18</td>
<td>-1.50</td>
</tr>
</tbody>
</table>
Example 3: Recall of Stressful Events

- If the adjusted residuals follow $N(0,1)$, we expect $18 \times 0.05 = 0.9 \approx 1$ adjusted residual larger than 1.96 or smaller than -1.96

- Months 1, 3, 4 positive adjusted residuals
  Months 16, 17 negative adjusted residuals

- More likely to report recent events (positive residual: means observed is larger than expected, i.e. more likely to report a stressful event in a months immediately prior to interview)
Example 3: Recall of Stressful Events

- Plot of adjusted residuals by month shows downward trend.
Normal Q-Q Plot

• Probability plot, which is a graphical method for comparing two probability distributions by plotting their quantiles against each other (Q stands for Quantiles, i.e. quantile against quantile)

• Plots observed quantiles against expected quantiles, hence plot quantiles of adjusted residuals against the quantiles of the standard normal.

• Points should lie close to $y = x$ line, if adjusted residuals are $N(0,1)$. 
Conclusions

• Divergence in the tails from straight line
• Strong evidence that equiprobable model does not fit the data.
• **More likely to report recent events.** Such a tendency would result if respondents were more likely to remember recent events than distant events.
• So, use another model. Which one?
• Let us explore a Poisson time trend model (a Poisson model with a covariate)
Poisson Regression Model with a Covariate –

Poisson Time Trend Model
Outline

• Poisson Time Trend Model – Poisson regression model with a covariate

• Example 3: Recall of Stressful Events continued
Poisson Time Trend Model – including a covariate

- $y_i =$ number of events in month $i$.

- The log of the expected count is a linear function of time before interview:

\[
\log(\mu_i) = \alpha + \beta i \quad i = 1, \cdots, C
\]

where $\mu_i = E(y_i)$ is the expected number of events in month $i$.

- Meaning: model assumes response variable $Y$ has a Poisson distribution, and assumes the logarithm of its expected value can be modelled by a linear combination of unknown parameters (here 2).
Poisson Time Trend Model – including a covariate

- For model \( \log(\mu_i) = \alpha + \beta_i \quad i = 1, \ldots, C \)
- \( \beta = 0 \rightarrow \) **equiprobable model** (no effect of the covariate).
- \( \beta > 0 \rightarrow \) the expected count \( \mu_i \) **increases** as month (i) increases.
- \( \beta < 0 \rightarrow \) the expected count \( \mu_i \) **decreases** as month (i) increases.
Poisson Model – with one covariate

- In general, the form of a Poisson model with one explanatory variable (X) is:

\[ \log(m) = a + b \times x \]

- Explanatory variable X can be categorical or continuous.
- In this example ‘month i’ is the explanatory variable X.
Poisson Model: Parameter Estimates
Predicted Values and Residuals

- Parameters estimated via maximum likelihood estimation → and $\hat{\alpha}$ $\hat{\beta}$

- These estimates are used to compute **predicted values:**

$$\log(\hat{\mu}_i) = \hat{\alpha} + \hat{\beta} i \quad \rightarrow \quad \hat{\mu}_i = \exp(\hat{\alpha} + \hat{\beta} i)$$

- The **adjusted residuals** are

$$\frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i (1 - \hat{\mu}_i / n)}} \quad \text{where} \quad n = \sum_{i=1}^{c} y_i$$
Poisson Model: Goodness of Fit Statistics

• The **goodness-of-fit test statistics** are

**Pearson Chi-squared Test Statistic**

\[ X^2 = \sum_{i=1}^{c} \left( \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i}} \right)^2 \]

**Likelihood Ratio Test Statistics (Deviance)**

\[ L^2 = 2 \sum_{i=1}^{c} y_i \log \left( \frac{y_i}{\hat{\mu}_i} \right) \]
• The **null hypothesis** is

\[ H_0: \text{The model} \]
\[ y_i \sim \text{Poisson}(\mu_i) \text{ with } \log(\mu_i) = \alpha + \beta i \]
\[ \text{holds.} \]

• Returning to our example expected counts \( \hat{\mu}_i \) (under the model) are ...
Example 3: Recall of Stressful Events

<table>
<thead>
<tr>
<th>Month</th>
<th>Count Obs</th>
<th>Count Exp</th>
<th>Month</th>
<th>Count Obs</th>
<th>Count Exp</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15</td>
<td>15.1</td>
<td>10</td>
<td>10</td>
<td>7.1</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>13.9</td>
<td>11</td>
<td>7</td>
<td>6.5</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>12.8</td>
<td>12</td>
<td>9</td>
<td>6.0</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>11.8</td>
<td>13</td>
<td>11</td>
<td>5.5</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>10.8</td>
<td>14</td>
<td>3</td>
<td>5.1</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>9.9</td>
<td>15</td>
<td>6</td>
<td>4.7</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>9.1</td>
<td>16</td>
<td>1</td>
<td>4.3</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>8.4</td>
<td>17</td>
<td>1</td>
<td>4.0</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>7.7</td>
<td>18</td>
<td>4</td>
<td>3.6</td>
</tr>
</tbody>
</table>

Expected value calculated using the model: $\hat{m}_i = \exp(a^\hat{a} + b^\hat{b}_{i})^{20}$
• **If \( H_0 \) is true**

\[
\chi^2 \sim \chi^2_{df} \quad \quad \quad L^2 \sim \chi^2_{df}
\]

where \( df = \) no. of cells – no. of model parameters

\[= C - 2 = 18 - 2 = 16\]

• \( \chi^2 = 22.71 \) with 16 df.

\[ p\text{-value} = 0.1216 \rightarrow \text{do not reject} \ H_0. \]

• \( L^2 = 24.57 \) with 16 df.

\[ p\text{-value} = 0.0778 \rightarrow \text{do not reject} \ H_0. \]
Conclusion

• The Poisson time trend model is consistent with the data.

• Participants are more likely to report recent events.
## Adjusted Residuals

<table>
<thead>
<tr>
<th>month</th>
<th>adjusted residual</th>
<th>month</th>
<th>adjusted residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.05</td>
<td>10</td>
<td>1.11</td>
</tr>
<tr>
<td>2</td>
<td>-0.89</td>
<td>11</td>
<td>0.18</td>
</tr>
<tr>
<td>3</td>
<td>0.36</td>
<td>12</td>
<td>1.26</td>
</tr>
<tr>
<td>4</td>
<td>1.61</td>
<td>13</td>
<td>2.42</td>
</tr>
<tr>
<td>5</td>
<td>-1.86</td>
<td>14</td>
<td>-0.98</td>
</tr>
<tr>
<td>6</td>
<td>0.34</td>
<td>15</td>
<td>0.64</td>
</tr>
<tr>
<td>7</td>
<td>0.28</td>
<td>16</td>
<td>-1.70</td>
</tr>
<tr>
<td>8</td>
<td>-1.58</td>
<td>17</td>
<td>-1.60</td>
</tr>
<tr>
<td>9</td>
<td>-0.09</td>
<td>18</td>
<td>0.20</td>
</tr>
</tbody>
</table>
Adjusted Residuals

• If the adjusted residuals follow $N(0,1)$, we expect $18 \times 0.05 = 0.9 \approx 1$ adjusted residual larger than 1.96 or smaller than -1.96.

• So, it is not surprising that we see one residual (month 13: 2.42) exceed 1.96 (i.e. no evidence against $H_0$).

• Adjusted residuals show no obvious pattern based on plot by month.
Poisson Regression: Model Interpretation

• Fitted values help with interpretation
• Fitted values are obtained by

\[ \hat{y}_i = \hat{m}_i = \exp (\hat{a} + \hat{b}^i) \]
Poisson Regression: Model Interpretation – cont.

• For a **binary explanatory variable**, where $X=0$ means absence and $X=1$ means presence, the **rate ratio (RR)** for presence vs. absence

$$RR = \frac{E \left( y_i \mid Presence \right)}{E \left( y_i \mid Absence \right)} = \frac{E \left( y_i \mid x_i = 1 \right)}{E \left( y_i \mid x_i = 0 \right)} = \exp(\beta)$$

• For a **continuous explanatory variable** a one unit increase in $X$ will result in a multiplicative effect of $\exp(\beta)$ on $\mu$
Confidence Interval for the Slope

• Estimate of the slope is $\hat{\beta} = -0.0838$ with estimated standard error $\text{se}(\hat{\beta}) = 0.0168$.

• Therefore a 95% CI for $\beta$ is:

\[
= -0.0838 \pm 1.96 \times 0.0168 \\
= (-0.117; -0.051)
\]

(therefore negative, CI does not include 0), the expected count $\mu_i$ decreases as month (i) increases.
Estimated Change

• To add interpretation, the estimated change (from one month to the next) in proportionate terms is

\[
\frac{\hat{\mu}_{i+1} - \hat{\mu}_i}{\hat{\mu}_i} = \frac{\hat{\mu}_{i+1}}{\hat{\mu}_i} - 1
\]

\[
= \frac{\exp(\hat{\alpha} + \hat{\beta}(i+1))}{\exp(\hat{\alpha} + \hat{\beta}i)} - 1
\]

\[
= \exp(\hat{\beta}) - 1
\]
Estimated Change

• For our example “Recall of Stressful Events” this means:

\[
\frac{\hat{m}_{i+1} - \hat{m}_i}{\hat{m}_i} = \exp(-0.0838) - 1
\]

\[
= 0.920 - 1 = -0.08
\]

i.e. the estimated change is an **8% decrease** per month.
## CI for the Estimated Change

<table>
<thead>
<tr>
<th></th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>$-0.117$</td>
<td>$-0.051$</td>
</tr>
<tr>
<td>$\exp(\beta)$</td>
<td>$\exp(-0.117) = 0.89$</td>
<td>$\exp(-0.051) = 0.95$</td>
</tr>
<tr>
<td>$\exp(\beta) - 1$</td>
<td>$0.89 - 1 = -0.11$</td>
<td>$0.95 - 1 = -0.05$</td>
</tr>
</tbody>
</table>

- Estimated change is **-8%** per month with a 95% CI of **(-11% , -5%)**. The number of stressful events decreases.
References


  p. 3 for “Recall of stressful events” example.
  p. 87 for “Suicides” example.
Thank you for your attention