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**) $\dfrac{O_i - E_i}{\sqrt{E_i}}$

- **Adjusted residuals** (preferred): $\dfrac{O_i - E_i}{SD(O_i - E_i)}$

$$SD(O_i - E_i) = \sqrt{E_i(1 - E_i/n)}$$

- **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

month	adjusted residual	month	adjusted residual
1	**2.46**	10	0.66
2	1.02	11	-0.42
3	**2.10**	12	0.30
4	**3.18**	13	1.02
5	-1.14	14	-1.86
6	1.02	15	-0.78
7	0.66	16	**-2.58**
8	-1.50	17	**-2.58**
9	-0.06	18	-1.50

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

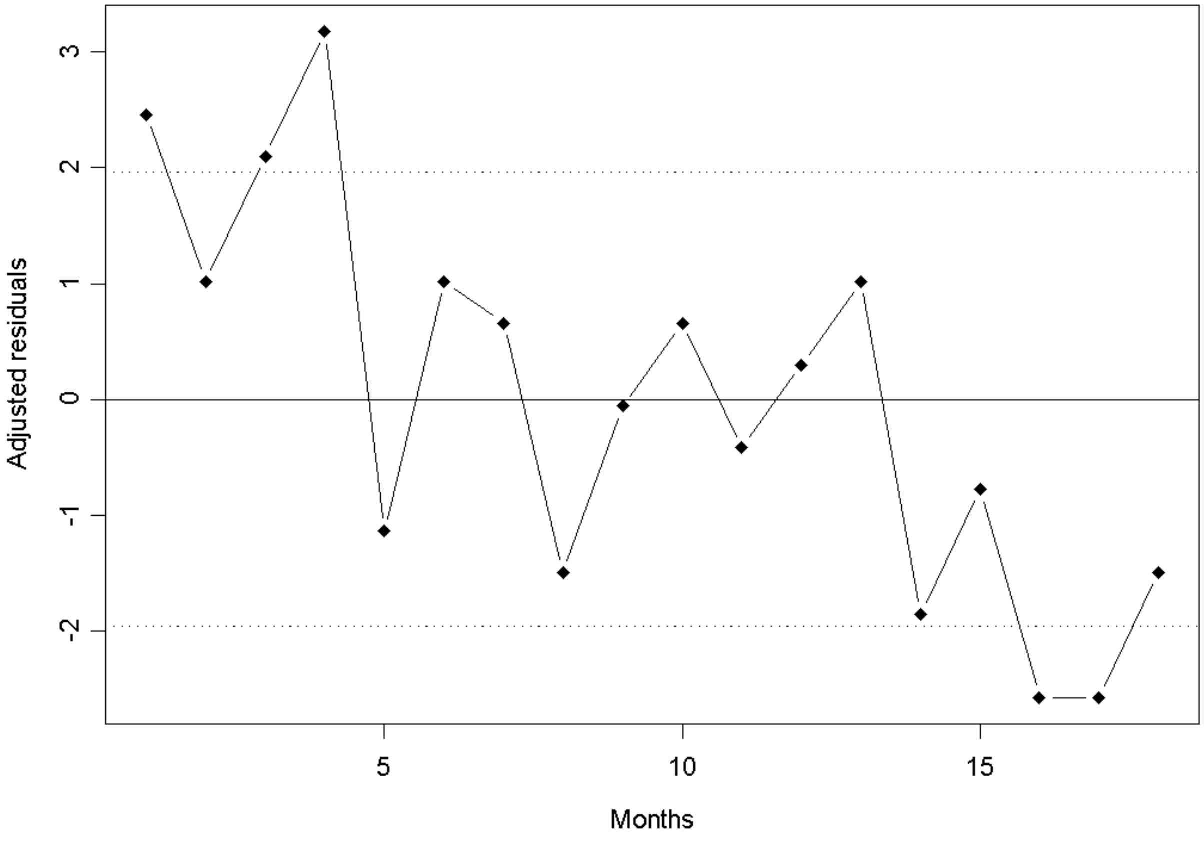
- Months 1, 3, 4 positive adjusted residuals

  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)

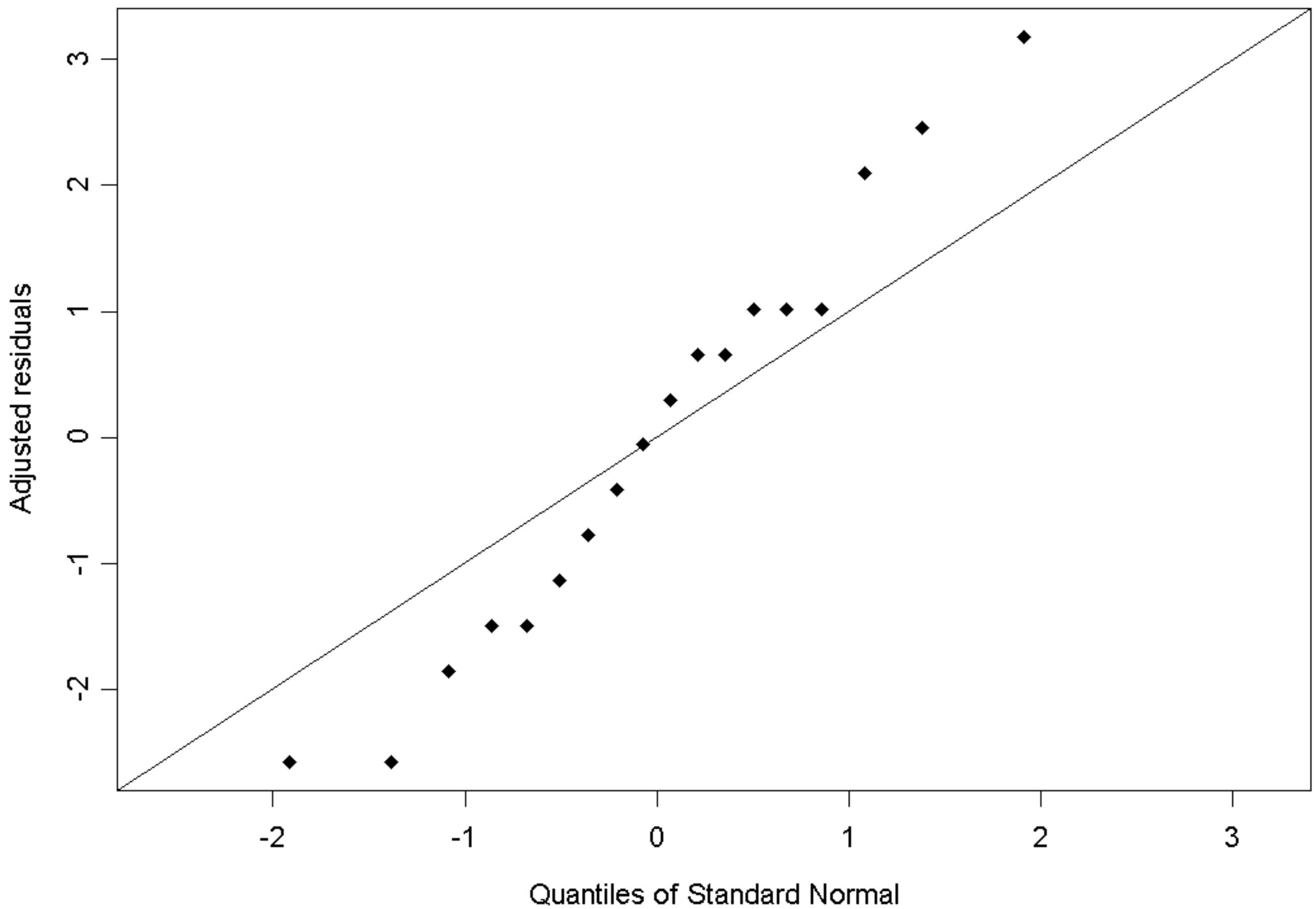
Example 3: Recall of Stressful Events

- 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).


Adjusted residuals
3
2
1
0
-1
-2
-2
-1
0
1
2
3
Quantiles of Standard Normal

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:

$$\log(\mu_i) = \alpha + \beta i \qquad i = 1, \cdots, C$$

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 $\log(\mu_i) = \alpha + \beta i \qquad i = 1, \cdots, C$

- $\beta = 0 \rightarrow$ equiprobable model (no effect of the covariate).

- $\beta > 0 \rightarrow$ the expected count μ_i **increases** as month (i) increases.

- $\beta < 0 \rightarrow$ the expected count μ_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:

$$\log(\boldsymbol{m}) = \boldsymbol{a} + \boldsymbol{b}\,x$$

- 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:**

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

- 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: The model

  $y_i \sim \text{Poisson}(\mu_i)$ with $\log(\mu_i) = \alpha+\beta i$

  holds.

- Returning to our example expected counts $\hat{\mu}_i$ (under the model) are ...

Example 3: Recall of Stressful Events

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

Expected value calculated using the model: $\hat{m}_i = \exp(\hat{a} + \hat{b}\, i)$

- **If H_0 is true**

$$X^2 \sim \chi^2_{df} \qquad\qquad L^2 \sim \chi^2_{df}$$

where df = no. of cells – no. of model parameters

= C – 2 = 18 – 2 = 16

- $X^2 = 22.71$ with 16 df.

p-value = 0.1216 $\rightarrow$ **do not reject H_0.**

- $L^2 = 24.57$ with 16 df.

p-value = 0.0778 $\rightarrow$ **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

month	adjusted residual	month	adjusted residual
1	-0.05	10	1.11
2	-0.89	11	0.18
3	0.36	12	1.26
4	1.61	13	**2.42**
5	-1.86	14	-0.98
6	0.34	15	0.64
7	0.28	16	-1.70
8	-1.58	17	-1.60
9	-0.09	18	0.20

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
- The coefficient is interpreted whether the explanatory variable is a categorical or a continuous variable.

$$\hat{y}_{\mathrm{i}} = \hat{\boldsymbol{m}}_i = \exp(\hat{\boldsymbol{a}} + \hat{\boldsymbol{b}} \ \mathrm{i})$$

$\hat{\boldsymbol{b}}$

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

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

- For a **continuous explanatory variable** a one unit increase in X will result in a multiplicative effect of $\exp(\beta)$ on μ

Confidence Interval for the Slope

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

- Therefore a 95% CI for β is:

= -0.0838 ± 1.96 x 0.0168

= (-0.117 ; -0.051)

(therefore negative, CI does not include 0), the expected count μ_i decreases as month (i) increases

Estimated Change

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

$$\frac{\hat{\mu}_{i+1} - \hat{\mu}_i}{\hat{\mu}_i} = \frac{\hat{\mu}_{i+1}}{\hat{\mu}_i} - 1 \qquad \hat{\mu}_i = \hat{E}(Y_i) = \exp(\hat{\alpha} + \hat{\beta} i)$$

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

$$= \exp(\hat{\beta}) - 1$$

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

	Lower Bound	**Upper Bound**
β	−0.117	−0.051
exp(β)	exp(-0.117) = 0.89	exp(-0.051) = 0.95
exp(β) - 1	0.89 - 1 = -0.11	0.95 - 1 = -0.05
	-11%	**-5%**

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

References

- Agresti, A (2013) Categorical Data Analysis, 3rd ed, New Jersey, Wiley.

- Agresti, A. (2007) *An Introduction to Categorical Data Analysis.* 2nd ed, Wiley.

- Haberman, S. J. (1978) *Analysis of Qualitative Data, Volume 1,* Academic Press, New York.

  p. 3 for "Recall of stressful events" example.

  p. 87 for "Suicides" example.

National Centre for Research Methods

Thank you for your attention