Poisson Regression Models for Count Data

Dr Gabriele Durrant

Outline

- Review
- Introduction to Poisson regression
- A simple model: equiprobable model
- Pearson and likelihood-ratio test statistics
- Residual analysis
- Poisson regression with a covariate (Poisson time trend model)

Review of Regression

You may have come across:

Dependent Variable	Regression Model
Continuous	Linear
Binary	Logistic
Multicategory (unordered) (nominal variable)	Multinomial Logit
Multicategory (ordered) (ordinal variable)	Cumulative Logit

Regression

In this session:

Dependent Variable	Regression Model
Continuous	Linear
Binary	Logistic
Multicategory (unordered) (nominal variable)	Multinomial Logit
Multicategory (ordered) (ordinal variable)	Cumulative Logit
Count variable	**Poisson Regression (Log-linear model)**

Data

Data for this session are assumed to be:

- A **count** variable Y (e.g. number of accidents, number of suicides)
- One **categorical** variable (X) with C possible categories (e.g. days of week, months)
- Hence Y has C possible outcomes $y_1, y_2, \ldots, y_C$

Introduction: Poisson regression

- Poisson regression is a form of regression analysis **model count data** (if all explanatory variables are categorical then we **model contingency tables (cell counts))**.
- The model models expected frequencies
- The model specifies how the count variable depends on the explanatory variables (e.g. level of the categorical variable)

Introduction: Poisson regression

- Poisson regression models are generalized linear models with the logarithm as the (canonical) link function.

- Assumes response variable Y has a Poisson distribution, and the logarithm of its expected value can be modelled by a linear combination of unknown parameters.

- Sometimes known as a **log-linear model**, in particular when used to model contingency tables (i.e. only categorical variables).

Example: Suicides (count variable) by Weekday (categorical variable) in France

Mon	1001	15.2%
Tues	1035	15.7%
Wed	982	14.9%
Thur	1033	15.7%
Fri	905	13.7%
Sat	737	11.2%
Sun	894	13.6%
Total	**6587**	**100.0%**

Introduction: Poisson regression

- Let us first look at a simple case: the equiprobable model (here for a 1-way contingency table)

Equiprobable Model

- An **equiprobable model** means that:
  - All outcomes are equally probable (equally likely).
  - That is, for our example, we assume a uniform distribution for the outcomes across days of week (Y does not vary with days of week X).

Equiprobable Model

- The equiprobable model is given by:

$$\boxed{P(Y=y_1) = P(Y=y_2) = \ldots = P(Y=y_C) = 1/C}$$

i.e. we expect an equal distribution across days of week.

- Given the data we can test if the assumption of the equiprobable model (H_0) holds

Example 1: Suicides by Weekday in France

Mon	1001	15.2%
Tues	1035	15.7%
Wed	982	14.9%
Thur	1033	15.7%
Fri	905	13.7%
Sat	737	11.2%
Sun	894	13.6%
Total	**6587**	**100.0%**

H_0: Each day is equally likely for suicides (i.e. the expected proportion of suicides is 100/7 = **14.3%** each day)

Example 2: Traffic Accidents by Weekday

Mon	11	11.8%
Tues	9	9.7%
Wed	7	7.5%
Thur	10	10.8%
Fri	15	16.1%
Sat	18	19.4%
Sun	23	24.7%
Total	**93**	**100.0%**

H_o: Each day is equally likely for an accident (i.e. the expected proportion of accidents is 100/7 = **14.3%** each day)

Hypothesis Testing

- H_0: Each day is equally likely for an accident.

- Alternative null hypotheses are:
  - H_0: Each working day equally likely for an accident.
  - H_0: Saturday and Sunday are equally likely for an accident.

- Omitted variables? For example, distance driven each day of the week.

Poisson regression – without a covariate

- We can express this equiprobable model more formally as a Poisson regression model (without a covariate), which models the expected frequency

Poisson regression

- We assume a **Poisson** distribution with parameter μ for the random component, i.e. y_i ~ Poisson(μ), i.e.

- Y is a random variable that takes only positive integer values
- Poisson distribution has a single parameter (μ) which is both its mean and its variance.

$$P(Y_i = y_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} \quad \text{where } y_i = 1, 2, 3$$

Poisson regression: Simple Model (No Covariate)

- We aim to model the expected value of Y. It can be shown that this is the parameter μ, hence we aim to model μ.

- We can write the **equiprobable model** defined earlier as a simple Poisson model (no explanatory variables), i.e. mean of Y does not change with month:

$$\mathrm{E}(y_i) = \boldsymbol{m}_i = 1/C$$

$$\log(\boldsymbol{m}_i) = \boldsymbol{a} \qquad \mathrm{i} = 1,\cdots,\mathrm{C}$$

where $\boldsymbol{a} = \log(1/C)$ is a constant.