Lesson 07 Poisson Regression

Lusine Zilfimian

March 30 (Monday), 2020

Contents

- Quiz

Contents

- Quiz
- About Poisson Distribution

Contents

- Quiz
- About Poisson Distribution
- Assumptions

Contents

- Quiz
- About Poisson Distribution
- Assumptions
- Model Formulation

Contents

- Quiz
- About Poisson Distribution
- Assumptions
- Model Formulation
- Model Evaluation

Contents

- Quiz
- About Poisson Distribution
- Assumptions
- Model Formulation
- Model Evaluation
- Interpretation of Parameters

Contents

- Quiz
- About Poisson Distribution
- Assumptions
- Model Formulation
- Model Evaluation
- Interpretation of Parameters
- Inference

Contents

- Quiz
- About Poisson Distribution
- Assumptions
- Model Formulation
- Model Evaluation
- Interpretation of Parameters
- Inference
- Accuracy of model

Contents

- Quiz
- About Poisson Distribution
- Assumptions
- Model Formulation
- Model Evaluation
- Interpretation of Parameters
- Inference
- Accuracy of model
- Overdispersion

Last Lecture ReCap

- How to deal with categorical response variable with more than 2 levels in logistic regression?

Last Lecture ReCap

- How to deal with categorical response variable with more than 2 levels in logistic regression?
- How to check the goodness of fit in logistic regression?

Poisson Distribution

Poisson distribution is a *discrete* probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known *constant mean* rate and independently of the time since the last event.

- The number of problems you solve per unit of time.

Poisson Distribution

Poisson distribution is a *discrete* probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known *constant mean* rate and independently of the time since the last event.

- The number of problems you solve per unit of time.
- The number of phone calls to the call center per hour.

Poisson Distribution

Poisson distribution is a *discrete* probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known *constant mean* rate and independently of the time since the last event.

- The number of problems you solve per unit of time.
- The number of phone calls to the call center per hour.
- The number of goals scored per football game.

Poisson Distribution

Poisson distribution is a *discrete* probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known *constant mean* rate and independently of the time since the last event.

- The number of problems you solve per unit of time.
- The number of phone calls to the call center per hour.
- The number of goals scored per football game.
- Number of network failures per day.

Poisson Distribution

- The probability of observing k events within a given interval is calculated with the following formula.

Poisson Distribution

- The probability of observing k events within a given interval is calculated with the following formula.
- $$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Poisson Distribution

- The probability of observing k events within a given interval is calculated with the following formula.
- $$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$
- One of the most important characteristics for Poisson distribution and Poisson Regression is **equidispersion**. {Show this property by yourself.}

Poisson Distribution

- The probability of observing k events within a given interval is calculated with the following formula.
-
$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$
- One of the most important characteristics for Poisson distribution and Poisson Regression is **equidispersion**. {Show this property by yourself.}
- Only one parameter λ is needed to determine the probability of an event.

Poisson Distribution

- The probability of observing k events within a given interval is calculated with the following formula.
-
$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$
- One of the most important characteristics for Poisson distribution and Poisson Regression is **equidispersion**. {Show this property by yourself.}
- Only one parameter λ is needed to determine the probability of an event.
- For the large means the normal distribution is a good approximation for the Poisson distribution.

Assumptions

Poisson regression is used when

- The dependent variable is positive integer *count*.

Assumptions

Poisson regression is used when

- The dependent variable is positive integer *count*.
- Dependent variable must follow a Poisson distribution.

Assumptions

Poisson regression is used when

- The dependent variable is positive integer *count*.
- Dependent variable must follow a Poisson distribution.
- Observations must be independent.

Model Formulation

- We suppose that the Poisson λ is determined by a set of k regressor variables (the X's).

Model Formulation

- We suppose that the Poisson λ is determined by a set of k regressor variables (the X's).
- We could assume the following formula:

Model Formulation

- We suppose that the Poisson λ is determined by a set of k regressor variables (the X's).
- We could assume the following formula:
- $$\lambda = \beta_0 + \beta_1 * X$$

Model Formulation

- We suppose that the Poisson λ is determined by a set of k regressor variables (the X's).
- We could assume the following formula:
- $$\lambda = \beta_0 + \beta_1 * X$$
- However this model can predict negative mean values

Model Formulation

- We suppose that the Poisson λ is determined by a set of k regressor variables (the X's).
- We could assume the following formula:
- $$\lambda = \beta_0 + \beta_1 * X$$
- However this model can predict negative mean values
- Thus let's assume the following model:

Model Formulation

- We suppose that the Poisson λ is determined by a set of k regressor variables (the X's).
- We could assume the following formula:
- $$\lambda = \beta_0 + \beta_1 * X$$
- However this model can predict negative mean values
- Thus let's assume the following model:
- $$ln(\lambda) = \beta_0 + \beta_1 * X$$

Model Formulation

- We suppose that the Poisson λ is determined by a set of k regressor variables (the X's).
- We could assume the following formula:
- $$\lambda = \beta_0 + \beta_1 * X$$
- However this model can predict negative mean values
- Thus let's assume the following model:
- $$ln(\lambda) = \beta_0 + \beta_1 * X$$
- Simple Poisson Regression Model

Model Formulation

- We suppose that the Poisson λ is determined by a set of k regressor variables (the X's).
- We could assume the following formula:
- $$\lambda = \beta_0 + \beta_1 * X$$
- However this model can predict negative mean values
- Thus let's assume the following model:
- $$ln(\lambda) = \beta_0 + \beta_1 * X$$
- Simple Poisson Regression Model
- $$\lambda = e^{\beta_0 + \beta_1 * X}$$

Model Evaluation

- Usually the regression coefficients are estimated using the method of maximum likelihood.

Model Evaluation

- Usually the regression coefficients are estimated using the method of maximum likelihood.
- The logarithm of the likelihood function is

Model Evaluation

- Usually the regression coefficients are estimated using the method of maximum likelihood.
- The logarithm of the likelihood function is
- $l(\beta_0, \beta_1 | y_i, x_i) = \sum_{i=1}^{n} (y_i(\beta_0 + \beta_1 * x_i) - e^{\beta_0 + \beta_1 * x_i} - ln(y_i!))$

Model Evaluation

- Usually the regression coefficients are estimated using the method of maximum likelihood.
- The logarithm of the likelihood function is
- $l(\beta_0, \beta_1 | y_i, x_i) = \sum_{i=1}^{n} (y_i(\beta_0 + \beta_1 * x_i) - e^{\beta_0 + \beta_1 * x_i} - ln(y_i!))$
- The likelihood equations may be formed by taking the derivatives with respect to each regression coefficient and setting the result equal to zero.

Model Evaluation

- Usually the regression coefficients are estimated using the method of maximum likelihood.
- The logarithm of the likelihood function is
- $l(\beta_0, \beta_1 | y_i, x_i) = \sum_{i=1}^{n}(y_i(\beta_0 + \beta_1 * x_i) - e^{\beta_0+\beta_1 * x_i} - ln(y_i!))$
- The likelihood equations may be formed by taking the derivatives with respect to each regression coefficient and setting the result equal to zero.
- Doing this leads to a set of nonlinear equations that admits no closed-form solution.[1]

[1] The explicit formula for coefficients cannot be calculated. We can start with the initial guess of what the parameters might be and how they might vary. Various combinations of parameters can be tried and we can see which one gives the

Model Evaluation

- Usually the regression coefficients are estimated using the method of maximum likelihood.
- The logarithm of the likelihood function is
- $l(\beta_0, \beta_1 | y_i, x_i) = \sum_{i=1}^{n}(y_i(\beta_0 + \beta_1 * x_i) - e^{\beta_0 + \beta_1 * x_i} - ln(y_i!))$
- The likelihood equations may be formed by taking the derivatives with respect to each regression coefficient and setting the result equal to zero.
- Doing this leads to a set of nonlinear equations that admits no closed-form solution.[1]
- Thus, an iterative algorithm must be used to find the set of regression coefficients that maximum the loglikelihood.

[1] The explicit formula for coefficients cannot be calculated. We can start with the initial guess of what the parameters might be and how they might vary. Various combinations of parameters can be tried and we can see which one gives the

Model Evaluation

- Usually the regression coefficients are estimated using the method of maximum likelihood.
- The logarithm of the likelihood function is
- $l(\beta_0, \beta_1|y_i, x_i) = \sum_{i=1}^{n}(y_i(\beta_0 + \beta_1 * x_i) - e^{\beta_0+\beta_1 * x_i} - ln(y_i!))$
- The likelihood equations may be formed by taking the derivatives with respect to each regression coefficient and setting the result equal to zero.
- Doing this leads to a set of nonlinear equations that admits no closed-form solution.[1]
- Thus, an iterative algorithm must be used to find the set of regression coefficients that maximum the loglikelihood.
- **Note** that some statistical packages ignore the last term since it does not involve the regression parameters. Thus the calculated log-likelihoods will be different.

[1] The explicit formula for coefficients cannot be calculated. We can start with the initial guess of what the parameters might be and how they might vary. Various combinations of parameters can be tried and we can see which one gives the

Interpretation of Parameters

Intercept only model

- $ln(\lambda) = \beta_0, \ \lambda = e^{\beta_0}$

Interpretation of Parameters

Intercept only model

- $ln(\lambda) = \beta_0, \ \lambda = e^{\beta_0}$
- The average value of dependent variable.

Interpretation of Parameters

- Consider two values for x, $x_1 = 1, x_2 = 2$, thus we have 1 unit increase in x.

Interpretation of Parameters

- Consider two values for x, $x_1 = 1, x_2 = 2$, thus we have 1 unit increase in x.
- $\frac{\lambda_2}{\lambda_1} = \frac{e^{\beta_0+\beta_1*1}e^{\beta_1}}{e^{\beta_0+\beta_1*1}} = e^{\beta_1}$

Interpretation of Parameters

- Consider two values for x, $x_1 = 1, x_2 = 2$, thus we have 1 unit increase in x.
- $\frac{\lambda_2}{\lambda_1} = \frac{e^{\beta_0 + \beta_1 * 1} e^{\beta_1}}{e^{\beta_0 + \beta_1 * 1}} = e^{\beta_1}$
- $\beta_1 > 0, e^{\beta_1} > 1,\ \lambda_2 \textit{ is } e^{\beta_1}$ times larger than λ_1

Interpretation of Parameters

- Consider two values for x, $x_1 = 1, x_2 = 2$, thus we have 1 unit increase in x.
- $\frac{\lambda_2}{\lambda_1} = \frac{e^{\beta_0 + \beta_1 * 1} e^{\beta_1}}{e^{\beta_0 + \beta_1 * 1}} = e^{\beta_1}$
- $\beta_1 > 0, e^{\beta_1} > 1,\ \lambda_2$ *is* e^{β_1} times larger than λ_1
- $\beta_1 < 0, e^{\beta_1} < 1,\ \lambda_2$ *is* e^{β_1} times smaller than λ_1.

Overdispersion

- What if the observed variance is much higher, i.e. if the data are overdispersed?

Overdispersion

- What if the observed variance is much higher, i.e. if the data are overdispersed?
- When variance is greater than mean, that is called over-dispersion.

Overdispersion

- What if the observed variance is much higher, i.e. if the data are overdispersed?
- When variance is greater than mean, that is called over-dispersion.
- We want to hold true the following equidispersion condition:

Overdispersion

- What if the observed variance is much higher, i.e. if the data are overdispersed?
- When variance is greater than mean, that is called over-dispersion.
- We want to hold true the following equidispersion condition:
- $Var(Y|X) = \mathbb{E}(Y|X)$

Overdispersion

- What if the observed variance is much higher, i.e. if the data are overdispersed?
- When variance is greater than mean, that is called over-dispersion.
- We want to hold true the following equidispersion condition:
- $Var(Y|X) = \mathbb{E}(Y|X)$
- Suppose we do not know whether the condition is being held or not, and we want to test the following hypothesis

Overdispersion

- What if the observed variance is much higher, i.e. if the data are overdispersed?
- When variance is greater than mean, that is called over-dispersion.
- We want to hold true the following equidispersion condition:
- $Var(Y|X) = \mathbb{E}(Y|X)$
- Suppose we do not know whether the condition is being held or not, and we want to test the following hypothesis
- H_0 : There is no overdispersion (there is equidispersion).

Overdispersion

- What if the observed variance is much higher, i.e. if the data are overdispersed?
- When variance is greater than mean, that is called over-dispersion.
- We want to hold true the following equidispersion condition:
- $Var(Y|X) = \mathbb{E}(Y|X)$
- Suppose we do not know whether the condition is being held or not, and we want to test the following hypothesis
- H_0 : There is no overdispersion (there is equidispersion).
- H_1 : There is overdispersion.

Overdispersion

- The alternative hypothesis assumes that there is a relationship between variance and mean in the following form:

Overdispersion

- The alternative hypothesis assumes that there is a relationship between variance and mean in the following form:
- $var(y_i|x_i) = \lambda_i + \alpha h(\lambda_i)$

Overdispersion

- The alternative hypothesis assumes that there is a relationship between variance and mean in the following form:
- $var(y_i|x_i) = \lambda_i + \alpha h(\lambda_i)$
- h is positive transformation function

Overdispersion

- The alternative hypothesis assumes that there is a relationship between variance and mean in the following form:
- $var(y_i|x_i) = \lambda_i + \alpha h(\lambda_i)$
- h is positive transformation function
- If $\alpha > 0$ then overdispersion exists

Overdispersion

- The alternative hypothesis assumes that there is a relationship between variance and mean in the following form:
- $var(y_i|x_i) = \lambda_i + \alpha h(\lambda_i)$
- h is positive transformation function
- If $\alpha > 0$ then overdispersion exists
- If $\alpha = 0$ then equidispersion exists

Overdispersion

- The alternative hypothesis assumes that there is a relationship between variance and mean in the following form:
- $var(y_i|x_i) = \lambda_i + \alpha h(\lambda_i)$
- h is positive transformation function
- If $\alpha > 0$ then overdispersion exists
- If $\alpha = 0$ then equidispersion exists
- If $\alpha < 0$ then underdispersion exists

Overdispersion

- There are different ways to define the transormation function

Overdispersion

- There are different ways to define the transormation function
- $h(\lambda) = \lambda$, $(1 + \alpha)$ is called dispersion parameter. Then $var(y_i|x_i) = \lambda_i * dispersion$

Overdispersion

- There are different ways to define the transormation function
- $h(\lambda) = \lambda$, $(1 + \alpha)$ is called dispersion parameter. Then $var(y_i|x_i) = \lambda_i * dispersion$
- $h(\lambda) = \lambda^2$

Overdispersion

- There are different ways to define the transormation function
- $h(\lambda) = \lambda$, $(1 + \alpha)$ is called dispersion parameter. Then $var(y_i|x_i) = \lambda_i * dispersion$
- $h(\lambda) = \lambda^2$
- For Poisson regression we assume that dispersion parameter is 1

Overdispersion

- There are different ways to define the transormation function
- $h(\lambda) = \lambda$, $(1 + \alpha)$ is called dispersion parameter. Then $var(y_i|x_i) = \lambda_i * dispersion$
- $h(\lambda) = \lambda^2$
- For Poisson regression we assume that dispersion parameter is 1
- We need to check if there are overdispersion (we will do it in R).

Overdispersion: Models

- Estimated means for Poisson will be correct, but the standard errors (standard deviation) are wrong and unaccounted for by the model.

Overdispersion: Models

- Estimated means for Poisson will be correct, but the standard errors (standard deviation) are wrong and unaccounted for by the model.
- The main types of models for overdispersion are

Overdispersion: Models

- Estimated means for Poisson will be correct, but the standard errors (standard deviation) are wrong and unaccounted for by the model.
- The main types of models for overdispersion are
- QuasiPoisson Regression. It uses Quasi-likelihood estimation. So, to have a more correct standard error we can use a quasi-poisson model.

Overdispersion: Models

- Estimated means for Poisson will be correct, but the standard errors (standard deviation) are wrong and unaccounted for by the model.
- The main types of models for overdispersion are
- QuasiPoisson Regression. It uses Quasi-likelihood estimation. So, to have a more correct standard error we can use a quasi-poisson model.
- Negative Binomial Regression. It can be considered as a generalization of Poisson regression since it has the same mean structure as Poisson regression and it has an extra parameter to model the over-dispersion. If the conditional distribution of the outcome variable is over-dispersed, the confidence intervals for Negative binomial regression are likely to be narrower as compared to those from a Poisson regression.

Overdispersion: Models Comparison

- We will see that the coefficients for the poisson and quasipoissn models are the same, while coefficients for the negative binomial model are different.

Overdispersion: Models Comparison

- We will see that the coefficients for the poisson and quasipoissn models are the same, while coefficients for the negative binomial model are different.
- So the result of Poisson and Quasi-Poisson model will be different in terms of

Overdispersion: Models Comparison

- We will see that the coefficients for the poisson and quasipoissn models are the same, while coefficients for the negative binomial model are different.
- So the result of Poisson and Quasi-Poisson model will be different in terms of
- p-values for the significance test

Overdispersion: Models Comparison

- We will see that the coefficients for the poisson and quasipoissn models are the same, while coefficients for the negative binomial model are different.
- So the result of Poisson and Quasi-Poisson model will be different in terms of
- p-values for the significance test
- confidence intervals for the estimates

Accuracy of the model

- One way to look at the goodness of fit of the model is to look at the significance of the estimated parameters.

Accuracy of the model

- One way to look at the goodness of fit of the model is to look at the significance of the estimated parameters.
- Pearson statistics

Accuracy of the model

- One way to look at the goodness of fit of the model is to look at the significance of the estimated parameters.
- Pearson statistics
- Deviance test

Accuracy of the model

- One way to look at the goodness of fit of the model is to look at the significance of the estimated parameters.
- Pearson statistics
- Deviance test
- Pseudo R^2

Goodness of fit: Pearson statistics

- $P_p = \sum \frac{(y_i - \hat{\lambda}_i)^2}{\hat{\lambda}_i}$

Goodness of fit: Pearson statistics

- $P_p = \sum \frac{(y_i - \hat{\lambda}_i)^2}{\hat{\lambda}_i}$
- H_0 : There is no significant difference between the observed and the expected value: $y_i = \hat{\lambda}_i$

Goodness of fit: Pearson statistics

- $P_p = \sum \frac{(y_i - \hat{\lambda}_i)^2}{\hat{\lambda}_i}$
- H_0 : There is no significant difference between the observed and the expected value: $y_i = \hat{\lambda}_i$
- H_A : There is a significant difference between the observed and the expected value.

Goodness of fit: Pearson statistics

- $P_p = \sum \frac{(y_i - \hat{\lambda}_i)^2}{\hat{\lambda}_i}$
- H_0 : There is no significant difference between the observed and the expected value: $y_i = \hat{\lambda}_i$
- H_A : There is a significant difference between the observed and the expected value.
- When a test is rejected, there is a significant lack of fit.

Goodness of fit: Pearson statistics

- $P_p = \sum \frac{(y_i - \hat{\lambda}_i)^2}{\hat{\lambda}_i}$
- H_0 : There is no significant difference between the observed and the expected value: $y_i = \hat{\lambda}_i$
- H_A : There is a significant difference between the observed and the expected value.
- When a test is rejected, there is a significant lack of fit.
- When a test is not rejected, there is no evidence of lack of fit.

Goodness of fit: Deviance test

- $D_p = 2\sum[y_i ln(\frac{y_i}{\hat{\lambda}_i}) - (y_i - \hat{\lambda}_i)]$

Goodness of fit: Deviance test

- $D_p = 2\sum[y_i ln(\frac{y_i}{\hat{\lambda}_i}) - (y_i - \hat{\lambda}_i)]$
- If the model fits well, the observed values y_i will be close to their predicted means $\hat{\lambda}_i$. Thus deviance will be small.

Goodness of fit: Pseudo R-squared

- In Poisson regression, the most popular pseudo-R-squared measure is the function of the log-likelihoods of three models

Goodness of fit: Pseudo R-squared

- In Poisson regression, the most popular pseudo-R-squared measure is the function of the log-likelihoods of three models
- $R^2 = \frac{LL(fit) - LL(0)}{LL(max) - LL(0)}$

Goodness of fit: Pseudo R-squared

- In Poisson regression, the most popular pseudo-R-squared measure is the function of the log-likelihoods of three models
- $R^2 = \frac{LL(fit) - LL(0)}{LL(max) - LL(0)}$
- LL(0) is the log-likelihood of the intercept-only model

Goodness of fit: Pseudo R-squared

- In Poisson regression, the most popular pseudo-R-squared measure is the function of the log-likelihoods of three models
- $R^2 = \frac{LL(fit) - LL(0)}{LL(max) - LL(0)}$
- LL(0) is the log-likelihood of the intercept-only model
- LL(fit) is the log-likelihood of the current model

Goodness of fit: Pseudo R-squared

- In Poisson regression, the most popular pseudo-R-squared measure is the function of the log-likelihoods of three models
- $R^2 = \frac{LL(fit) - LL(0)}{LL(max) - LL(0)}$
- LL(0) is the log-likelihood of the intercept-only model
- LL(fit) is the log-likelihood of the current model
- LL(max) is the maximum log-likelihood possible. The maximum log-likelihood occurs when the actual responses (the y_i) exactly equal the predicted responses (the $\hat{\lambda}_i$).

Goodness of fit: Pseudo R-squared

- In Poisson regression, the most popular pseudo-R-squared measure is the function of the log-likelihoods of three models
- $R^2 = \frac{LL(fit) - LL(0)}{LL(max) - LL(0)}$
- LL(0) is the log-likelihood of the intercept-only model
- LL(fit) is the log-likelihood of the current model
- LL(max) is the maximum log-likelihood possible. The maximum log-likelihood occurs when the actual responses (the y_i) exactly equal the predicted responses (the $\hat{\lambda}_i$).
- Note: an intercept is in the model.

Ideas for project

- Negative Binomial Regression

Ideas for project

- Negative Binomial Regression
- Quasi-Poisson Regression

Ideas for project

- Negative Binomial Regression
- Quasi-Poisson Regression
- Zero-inflated Poisson Regression

Ideas for project

- Negative Binomial Regression
- Quasi-Poisson Regression
- Zero-inflated Poisson Regression
- How to deal with underdispersion?