

Inference and Diagnostics for Poisson Regression

Poisson regression – Stat 230

In this class we'll explore how to conduct inference for Poisson regression models. What you'll see is that the same inferential procedures (Wald test/intervals and likelihood ratio tests) are used. In addition, we'll check whether the assumptions of the model are reasonable.

To begin, we'll revisit the cancer data set from last class:

```
cancer <- read.csv("http://aloy.rbind.io/data/CancerCluster.csv")
```

Fitting multiple Poisson regression models

Last class we fit a simple Poisson regression model (i.e., we used only a single predictor variable) using the below code:

```
cancer_mod <- glm(Cases ~ MedianAge, data = cancer, family = "poisson",  
                  offset = log(Person.years))
```

To begin the day, you'll fit a Poisson regression model with multiple predictor variables. To do this, you adapt the formula to add additional predictor and interaction terms, just like you did for multiple linear and logistic regression models.

Task 1. Complete activity 22 in Chapter 8.

Task 2. Complete activity 23 in Chapter 8.

Task 3. Complete activity 27 in Chapter 8.

Task 4. Complete activity 28 in Chapter 8.

Task 5. Complete activity 29 in Chapter 8.

Conducting inference

The good news is that inference for Poisson regression uses the same tools (and code!) as inference for logistic regression, it's just the interpretations (of intervals) that differ a bit due to the link function. So, we'll review the use of Wald's test, Wald intervals, and likelihood ratio test (LRTs).

Task 6. Complete activity 30 in Chapter 8. To get the summary table for Wald's test, you can either use the `summary()` function or the `tidy()` command from `{broom}`.

Task 7. Now, let's construct confidence intervals for each coefficient. To do this in R, you can use the `confint()` function and pass in the fitted model. You can also add the intervals to your coefficient summary table by adding the argument `conf.int = TRUE` to the `tidy()` command.

Task 8. Complete activity 31 in Chapter 8. To answer whether the LRT is valid here, consider looking at the fitted values (i.e., the Poisson means, $\hat{\lambda}_i$) via `fitted(model)`. (Substitute in the name of your model.)

Model checking

The model-checking toolkit is also very similar for Poisson regression and binomial logistic regression. Specifically, we'll use deviance residual plots and the deviance goodness-of-fit test.

Task 9. Complete activity 34 in Chapter 8. To construct deviance residual plot, we can run the command

```
residualPlots(model, type = "deviance")
```

from the `{car}` package, substituting your model's name in for `model`. The deviance goodness-of-fit test statistic is reported in the same place on the `summary()` as it was in logistic regression, and you'll need to use the chi-square distribution to find the p-value.