|  |  |  |  |
| --- | --- | --- | --- |
| **PM592: Regression Analysis for Health Data Science** |  |  |  |
| **Lab 11 – Regression for Count Outcomes**  **Data Needed:** *-* | | | |

# Outline

* Estimated Marginal Means

In today’s lab we will be using the emmeans package, which is useful when presenting model results. With this package you can:

* Perform post-hoc pairwise comparisons among groups when you find a significant dummy variable set.
* Get the estimated value of Y for certain values of an X variable, averaged over the levels of all other X variables in the data set.
* Graphically display the results of your model, highlighting the effects of particular X variables of interest.

The following vignettes are useful for learning about the emmeans package:

* <https://cran.r-project.org/web/packages/emmeans/vignettes/basics.html>
* <https://cran.r-project.org/web/packages/emmeans/vignettes/comparisons.html>
* <https://cran.r-project.org/web/packages/emmeans/vignettes/interactions.html>

Note that the estimated marginal means approach can be used with any type of GLM (not just Poisson regression).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Lab 11 Exercises** |  |  |  |  |  |  |

|  |  |
| --- | --- |
| Objective(s): | Use familiar regression techniques, applied to count outcomes |
| Datasets Required: | hospitaler.dta |

This data set contains temporal trends in LAC+USC hospital admission from the ER over a 3-year period. Suppose hospital administration asked us to determine if the rate of hospital admissions for those visiting the ER is changing over time. They want to know if, and how, this rate is changing.

|  |  |
| --- | --- |
| Variable | Description |
| month | Month number (0-35) |
| mo4 | Month, grouped into 4-month aggregate bins |
| er\_visit | # of individuals who are seen by the ER per month |
| admit | # of individuals who are seen by the ER per month, who subsequently are admitted to the hospital |
| readmit | # of individuals who are seen by the ER per month, who subsequently are admitted to the hospital, who had been discharged from the hospital in the past 30 days |
| hosp\_ed |  |

1. Perform exploratory data analysis.
   1. For each 4-month group, and overall, examine the mean and variance of er\_visit and admit. Does the mean appear to be roughly equal to the variance? (We do this by 4-month grouping so we can get an estimate of the variance and thus evaluate the Poisson distribution assumption over time.)

A screenshot of a computer screen

Description automatically generated

For ER visits and hospital admissions, the variance seems to be higher than the mean for most 4-month groups. For readmissions, the mean appears to be roughly equal to the variance.

* 1. Add a variable that indicates the rate at which individuals are admitted to the hospital from the ER.

A black background with white text

Description automatically generated

* 1. Create a scatter plot of the (log) hospital admission rate vs. month. How does this relationship look in terms of linearity?

A black background with white text

Description automatically generated

A graph showing the growth of a number of months

Description automatically generated with medium confidence

The relationship between (log) hospital admission rate and month does not look linear.

*Thoughts: we could either 1) create a very well-fit model using some higher-order polynomial terms or 2) group the data together and present the results from a dummy-variable predictor model. I’m going to proceed with the latter because it will be more interpretable when presenting the results to hospital administration.*

1. Fit the Poisson model for the effect of 4-month time period on rate of hospital admissions from the ER.
   1. Write the equation for the best fit model in terms of expected number of admissions.

A screenshot of a computer screen

Description automatically generated

The equation for the best fit model is:

* 1. Write the equation for the best fit model in terms of expected rate of admissions.

A screenshot of a computer program

Description automatically generated

* 1. Explain how rate of hospital admissions has changed over time. Provide rate ratios, 95% confidence intervals, and p-values.

A screen shot of a computer

Description automatically generated

The expected rate of hospital admissions for month group 1 is 0.988 times the rate of hospital admissions in month group 0 (CI = (0.954, 1.02), p=0.48). The expected rate of hospital admissions for month group 2 is 0.930 times the rate of hospital admissions in month group 0 (CI = (0.898, 0.963), p<0.01). The expected rate of hospital admissions for month group 3 is 0.937 times the rate of hospital admissions in month group 0 (CI = (0.905, 0.970), p=0.00025). The expected rate of hospital admissions for month group 4 is 0.860 times the rate of hospital admissions in month group 0 (CI = (0.830, 0.891), p<0.01). The expected rate of hospital admissions for month group 5 is 0.817 times the rate of hospital admissions in month group 0 (CI = (0.788, 0.830), p=0.902). The expected rate of hospital admissions for month group 6 is 0.798 times the rate of hospital admissions in month group 0 (CI = (0.770, 0.827), p<0.01). The expected rate of hospital admissions for month group 7 is 0.849 times the rate of hospital admissions in month group 0 (CI = (0.820, 0.880), p<0.01). The expected rate of hospital admissions for month group 8 is 0.946 times the rate of hospital admissions in month group 0 (CI = (0.913, 0.980), p=0.0019).

1. Compute the predicted number of hospital admissions and the predicted rate of hospital admissions. You can use either the approach from the class notes or the emmeans package.
   1. Compute the predicted number of hospital admissions in each 4-month period. Plot the predicted number of admissions, with 95% confidence intervals, in each group.

A screen shot of a computer

Description automatically generated

A graph with a line graph

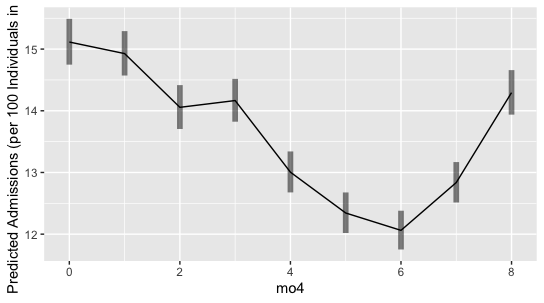
Description automatically generated

*Note: If you use emmeans, the predictions will be based on the average value of the offset across all observations. Thus this procedure provides an “adjusted” expected number of admissions holding the offset constant across months.*

* 1. Compute the predicted rate of hospital admissions per 100 individuals seen in the ER in each 4-month period by specifying the offset to be log(100). Plot the predicted rate of admissions (per 100), with 95% confidence intervals, in each group.

A screen shot of a computer

Description automatically generated



1. Assess the model goodness of fit.
   1. Provide the Pearson GOF test statistic and p-value.

A screenshot of a computer

Description automatically generated

* 1. Provide the Deviance GOF test statistic and p-value.

A screenshot of a computer program

Description automatically generated

* 1. Plot the expected number of hospital admissions vs. the observed number of hospital admissions. What are your impressions?

A graph with a line and a red line

Description automatically generated

The fit is not that good, the line of expected rate of admissions is different from the line of the observed rate of admissions.

* 1. Check for evidence of model overdispersion.

A screenshot of a computer program

Description automatically generated

There is evidence to suggest that overdispersion parameter is not equal to (greater than) 1. The dispersion parameter is 3.5.

1. **Extra Practice 1.** Upon finding evidence of model overdispersion, re-fit this model as a negative binomial model.

A computer screen with white text

Description automatically generated

1. **Extra Practice 2.** The hospital administration was also interested in whether the rate of individuals readmitted to the hospital is changing over time. Fit another Poisson model for the number of readmissions, with number of admissions as the offset.

