Poisson Regression with R

glm function output

- The output begins with echoing the function call. Then the information on deviance residuals is displayed.
- Deviance residuals are approximately normally distributed if the model is specified correctly.
- Here it shows a little bit of skeweness since median is not quite zero.

glm function output

- The Poisson regression coefficients for each of the variables along with the standard errors, z-scores, p-values and 95% confidence intervals for the coefficients.
- The coefficient for math is 0.07.
- This means that the expected log count for a one-unit increase in math is 0.07.

Poisson Regression with R

```
Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept)      -5.2471     0.6585   -7.97  1.6e-15 ***
progAcademic      1.0839     0.3583    3.03   0.0025 **
progVocational    0.3698     0.4411    0.84   0.4018
math              0.0702     0.0106    6.62  3.6e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
```

Poisson Regression with R

glm function output

- The indicator variable **progAcademic** compares between **prog = Academic** and **prog = “General”** , the expected log count for **prog = Academic** increases by about 1.1.
- The indicator variable **prog.Vocational** is the expected difference in log count (≈ 0.37) between **prog = ”Vocational”** and the reference group (**prog = ”General”**).

Poisson Regression with R

- The output above indicates that the incident rate for **prog** = **"Academic"** is 2.96 times the incident rate for the reference group (**prog** = **"General"**).
- Likewise, the incident rate for **prog** = **"Vocational"** is 1.45 times the incident rate for the reference group holding the other variables at constant.

Poisson Regression with R

- The percent change in the incident rate of **num_awards** is by 7% for every unit increase in math.

Deviance

- In statistics, deviance is a quality of fit statistic for a model that is often used for statistical hypothesis testing.
- It is a generalization of the idea of using the sum of squares of residuals in ordinary least squares to cases where model-fitting is achieved by maximum likelihood.

Poisson Regression with R

glm function output

- The information on deviance is also provided.
- We can use the residual deviance to perform a goodness of fit test for the overall model.

Poisson Regression with R

glm function output

- The residual deviance is the difference between the deviance of the current model and the maximum deviance of the ideal model where the predicted values are identical to the observed.
- Therefore, if the residual difference is small enough, the goodness of fit test will not be significant, indicating that the model fits the data.

Poisson Regression with R

glm function output

- If the test had been statistically significant, it would indicate that the data do not fit the model well.
- We could try to determine if there are omitted predictor variables, if our linearity assumption holds and/or if there is an issue of over-dispersion.

Comparing Candidate Models

- We can also test the overall effect of prog by comparing the deviance of the full model with the deviance of the model excluding prog.
- The two degree-of-freedom chi-square test indicates that prog, taken together, is a statistically significant predictor of **num_awards**.

Comparing Models

```
# update m1 model dropping prog
m2 <- update(m1, . ~ . - prog)

# test model differences with chi square test
anova(m2, m1, test="Chisq")
```

Poisson Regression with R

```
Analysis of Deviance Table

Model 1: num_awards ~ math
Model 2: num_awards ~ prog + math
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       198        204
2       196        189  2     14.6  0.00069 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

Incident Rate Ratios

- Sometimes, we might want to present the regression results as **incident rate ratios** (IRRs) and their standard errors, together with the confidence interval.
- To compute the standard error for the incident rate ratios, we will use the **Delta method** (Numerical Computation Method).
- To this end, we make use the function `deltamethod` implemented in R package **msm**.

Incident Rates

Incidence rate is the occurrence of an event over person-time, for example person-years.

$$\text{Incidence Rate} = \frac{\text{events}}{\text{Person Time}}$$

Note: the same time intervals must be used for both incidence rates.

Incident Rate Ratios

A **rate ratio** (sometimes called an incidence density ratio) in epidemiology, is a relative difference measure used to compare the incidence rates of events occurring at any given point in time.

$$\text{Incidence Rate Ratio} = \frac{\text{Incidence Rate 1}}{\text{Incidence Rate 2}}$$

Poisson Regression with R

Delta Method

```
s <- deltamethod(list(~ exp(x1), ~ exp(x2), ~ exp(x3),
      ~ exp(x4)), coef(m1), cov.m1)

#exponentiate old estimates dropping the p values

rexp.est <- exp(r.est[, -3])

# replace SEs with estimates
# for exponentiated coefficients

rexp.est[, "Robust SE"] <- s
```

Poisson Regression with R

```
rexp.est

                Estimate Robust SE       LL      UL
 (Intercept)    0.005263   0.00340 0.001484 0.01867
 progAcademic   2.956065   0.94904 1.575551 5.54620
 progVocational 1.447458   0.57959 0.660335 3.17284
 math           1.072672   0.01119 1.050955 1.09484
```

Poisson Regression with R

- Sometimes, we might want to look at the expected marginal means.
- For example, what are the expected counts for each program type holding math score at its overall mean?
- To answer this question, we can make use of the predict function.
- First off, we will make a small data set to apply the predict function to it.

Poisson Regression with R

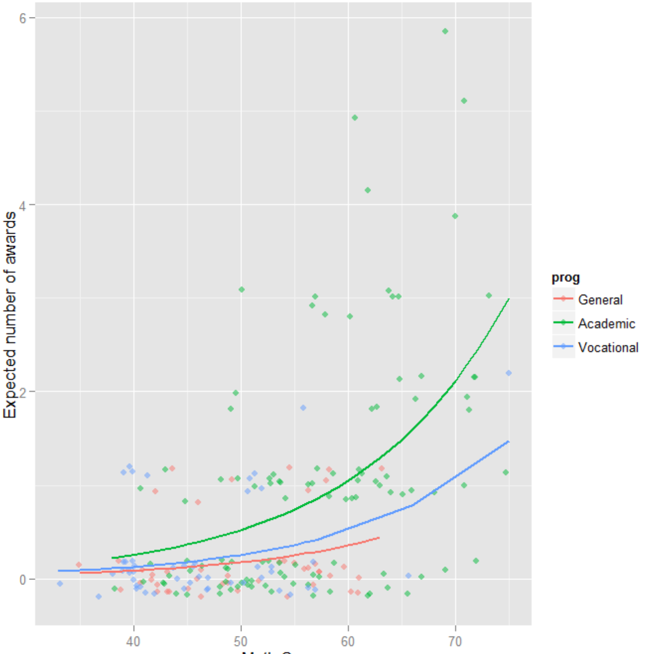
```
(s1 <- data.frame(math = mean(p$math),
  prog = factor(1:3, levels = 1:3,
  labels = levels(p$prog))))

    math       prog
 1 52.65    General
 2 52.65   Academic
 3 52.65 Vocational
```

Poisson Regression with R

```
predict(m1, s1, type="response", se.fit=TRUE)

 $fit
      1      2      3
 0.2114 0.6249 0.3060

 $se.fit
       1       2       3
 0.07050 0.08628 0.08834

 $residual.scale
 [1] 1
```

Poisson Regression with R

- In the output above, we see that the predicted number of events for level 1 of prog is about 0.21, holding math at its mean.
- The predicted number of events for level 2 of prog is higher at 0.62, and the predicted number of events for level 3 of prog is about .31.
- The ratios of these predicted counts ($\frac{0.625}{0.211} = 2.96$, $\frac{0.306}{0.211} = 1.45$) match what we saw looking at the IRR.



Expected number of awards
0
2
4
6
40
50
60
70
prog
General
Academic
Vocational