Poisson Regression with R

Analysis methods you might consider

- Below is a list of some analysis methods you may have encountered. Some of the methods listed are quite reasonable, while others have either fallen out of favor or have limitations.
- Poisson regression - Poisson regression is often used for modeling count data. Poisson regression has a number of extensions useful for count models.

- Negative binomial regression - Negative binomial regression can be used for over-dispersed count data, that is when the conditional variance exceeds the conditional mean. 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.
- Zero-inflated regression model - Zero-inflated models attempt to account for excess zeros. In other words, two kinds of zeros are thought to exist in the data, "true zeros" and "excess zeros". Zero-inflated models estimate two equations simultaneously, one for the count model and one for the excess zeros.

Poisson Regression with R

- OLS regression - Count outcome variables are sometimes log-transformed and analyzed using OLS regression. Many issues arise with this approach, including loss of data due to undefined values generated by taking the log of zero (which is undefined) and biased estimates.

Poisson regression

At this point, we are ready to perform our Poisson model analysis using the glm function. We fit the model and store it in the object m1 and get a summary of the model at the same time.

Poisson Regression with R

```
summary(m1 <- glm(num_awards ~ prog + math,


 Call:
 glm(formula = num_awards ~ prog + math, fam

 Deviance Residuals:
     Min       1Q   Median       3Q      Max
 -2.204   -0.844   -0.511    0.256    2.680
```

Poisson Regression with R

```
Coefficients:
                 Estimate Std. Error z value
(Intercept)       -5.2471     0.6585   -7.97
progAcademic       1.0839     0.3583    3.03
progVocational     0.3698     0.4411    0.84
math               0.0702     0.0106    6.62
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*'
```

Poisson Regression with R

```
(Dispersion parameter for poisson family ta

    Null deviance: 287.67  on 199  degrees
Residual deviance: 189.45  on 196  degrees
AIC: 373.5

Number of Fisher Scoring iterations: 6
```

Poisson Regression with R

Cameron and Trivedi (2009) recommended using robust standard errors for the parameter estimates to control for mild violation of the distribution assumption that the variance equals the mean. We use R package sandwich below to obtain the robust standard errors and calculated the p-values accordingly. Together with the p-values, we have also calculated the 95

Poisson Regression with R

```
cov.m1 <- vcovHC(m1, type="HC0")
std.err <- sqrt(diag(cov.m1))
r.est <- cbind(Estimate= coef(m1), "Robust S
"Pr(>|z|)" = 2 * pnorm(abs(coef(m1)/std.err)
LL = coef(m1) - 1.96 * std.err,
UL = coef(m1) + 1.96 * std.err)
```

Poisson Regression with R

```
r.est

                   Estimate Robust SE  Pr(>|z|)
 (Intercept)       -5.24712   0.64600 4.567e-16
 progAcademic       1.08386   0.32105 7.355e-04
 progVocational     0.36981   0.40042 3.557e-01
 math               0.07015   0.01044 1.784e-11
```