BIST P8110: Applied Regression II

21. Negative Binomial Regression

Qixuan Chen

Department of Biostatistics
Columbia University

Negative Binomial Regression

- The Poisson regression model is the most basic model for count data, but it requires that the conditional mean and the conditional variance to be the same.
- Negative binomial regression is an extension of Poisson regression in which the conditional variance may exceed the conditional mean (overdispersion).

Negative Binomial Distribution

- Let $Y \sim$ NegBinom, then

$$E(Y) = \mu$$

$$Var(Y) = \mu + \alpha\mu^2$$

- α is called the dispersion parameter
- As $\alpha \to 0$, we get the Poisson distribution
- The extra parameter α in the variance expression allows the mean and the variance not need to be equal

SAS Syntax

- Negative binomial models can be fitted using PROC COUNTREG:

```
proc countreg data=one;
model y = x / dist=negbin(p=2);
run;
```

- It can also be fitted using PROC GENMOD

```
 proc genmod data=one;
model y = x / dist=negbin link=log;
run;
```

Case Study

- Table below summarizes responses of 1308 subjects to the question: “Within the past 12 months, how many people have you known personally that were victims of homicide?” (*Categorical Data Analysis*, Alan Agresti, p.561)

Response	Black	White
0	119	1070
1	16	60
2	12	14
3	7	4
4	3	0
5	2	0
6	0	1

Poisson/Negative Binomial Model

- The mean function of the Poisson/Negative binomial model:

$$\begin{aligned}\log\{E(Y_i|X_i)\} &= \beta_0 + \beta_1 X_i \\ &= \log(\mu_i)\end{aligned}$$

  - Y=number of people known personally that were victims of homicide
  - $X = 1$ if black and 0 if white
- $f(Y_i|X_i) \sim$ Poisson(μ_i) or $f(Y_i|X_i) \sim$ NegBin(μ_i, α), where α is an overdispersion parameter and the variance of Y in the NegBin distribution is $(\mu_i + \alpha\mu_i^2)$.

SAS codes

```
/*Poisson regression*/
proc countreg data=homicide;
 model response = race /dist=poisson;
 freq count;
run;

/*Negative binomial regression*/
proc countreg data=homicide;
 model response = race /dist=negbin(p=2);
 freq count;
run;
```

Poisson vs. Negative Binomial Models

- How to decide whether a Poisson or a negative binomial model is more appropriate?
  - A t-test for the overdispersion parameter α is provided in PROC COUNTREG to test $H_0 : \alpha = 0$ vs. $H_\alpha : \alpha \neq 0$ when the option “dist=negbin(p=2)" is used.
  - SAS output of the Case Study:

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	-2.383208	0.117200	-20.33	<.0001
race	1	1.733145	0.238477	7.27	<.0001
_Alpha	1	4.942862	1.000487	4.94	<.0001

Other Models

- When the number of zeros in the sample exceeds the number of zeros predicted by either the Poisson or negative binomial model
  - zero-inflated Poisson regression
  - zero-inflated negative binomial regression
- Zero-inflated regression are two-part models
  - one models zero versus non-zero counts using logistic regression
  - the other models the actual counts using a Poisson or negative binomial distribution
- Definition of ZIP: http://support.sas.com/documentation/cdl/en/etsug/60372/HTML/default/viewer.htm#etsug_countreg_sect018.htm
- Definition of ZINB: http://support.sas.com/documentation/cdl/en/etsug/60372/HTML/default/viewer.htm#etsug_countreg_sect019.htm

Summary: Key Points

- How to write Poisson regression models for counts and rates?
- What is offset?
- How to interpret intercept and slopes of a Poisson model?
- What is Deviance and how to use it for models comparison?
- How to conduct goodness of fit test?
- What is overdispersion?
- How to know if there is overdispersion?
- How to “fix” overdispersion?

Summary: Key Points (Cont.)

- How to code Poisson model using GENMOD in SAS?
- How to calculate the expected number of counts based on a Poisson model?
- How to calculate relative risk based on a Poisson model?
- What is negative binomial regression?
- How to code negative binomial models using GENMOD and COUNTREG in SAS?
- How to decide whether a Poisson or a negative binomial model is more appropriate for a data set using PROC COUNTREG in SAS?

Suggested Readings

- Chapters 8 & 9 (Dupont)
- Liu W. and Cela J. “Count Data Models in SAS". SAS Global Forum 2008.