

## Negative Binomial Regression with R

Negative binomial regression is for modeling count variables, usually for over-dispersed count outcome variables.

This page uses the following packages. Make sure that you can load them before trying to run the examples on this page. If you do not have a package installed, run:

`install.packages("packagename")`, or if you see the version is out of date, run: `update.packages()`.

```
require(foreign)
require(ggplot2)
```

## Negative Binomial Regression with R

Please note: The purpose of this page is to show how to use various data analysis commands. It does not cover all aspects of the research process which researchers are expected to do. In particular, it does not cover data cleaning and checking, verification of assumptions, model diagnostics or potential follow-up analyses.

## Negative Binomial Regression with R

Examples of negative binomial regression Example

1. School administrators study the attendance behavior of high school juniors at two schools.

Predictors of the number of days of absence include the type of program in which the student is enrolled and a standardized test in math.

Example 2. A health-related researcher is studying the number of hospital visits in past 12 months by senior citizens in a community based on the characteristics of the individuals and the types of health plans under which each one is covered.

## Negative Binomial Regression with R

Description of the data Let's pursue Example 1 from above.

We have attendance data on 314 high school juniors from two urban high schools in the file **nb\_data**.

The response variable of interest is days absent, `daysabs`. The variable `math` gives the standardized math score for each student. The variable `prog` is a three-level nominal variable indicating the type of instructional program in which the student is enrolled.

## Negative Binomial Regression with R

Let's look at the data. It is always a good idea to start with descriptive statistics and plots.

```
dat <- read.dta("http://www.ats.ucla.edu/sta
dat <- within(dat, {
prog <- factor(prog, levels = 1:3, labels =
id <- factor(id)
})
```

# Negative Binomial Regression with R

```
summary(dat)
```

##	id	gender	math
##	1001 : 1	female:160	Min. : 1.0
##	1002 : 1	male :154	1st Qu.:28.0
##	1003 : 1		Median :48.0
##	1004 : 1		Mean :48.3
##	1005 : 1		3rd Qu.:70.0
##	1006 : 1		Max. :99.0
##	(Other):308		
##	prog		
##	General : 40		

## Negative Binomial Regression with R

```
ggplot(dat, aes(daysabs, fill = prog)) + geom_
., margins = TRUE, scales = "free")
```

Histogram plots showing distribution of the data  
 Each variable has 314 valid observations and their  
 distributions seem quite reasonable. The  
 unconditional mean of our outcome variable is much  
 lower than its variance.