POLS 6481. Research Design and Quantitative Methods II
Lecture 16. Poisson Regression for Counts

Readings: Wooldridge, *Introductory Econometrics 5e*, 17.3

Outline

1. Poisson Distribution
2. Poisson Regression for Count Data
3. Example: *Elephant Mating Behavior*
4. Poisson Regression Predicted Values
5. Example: *Extramarital Affairs*
6. Zero·Inflated Poisson Regression

* note: we return to Poisson regression in Lecture 19 on dummy independent variables

POLS 6481. Research Design and Quantitative Methods II
Lecture 16. Poisson Regression for Counts
Readings: Wooldridge, *Introductory Econometrics 5e*, 17.3

1.

Poisson distribution: $P(y = h) = e^{-\mu} \frac{\mu^h}{h!}, h = 0, 1, 2, \ldots$

L. Bortkiewicz's example: examined 10 Prussian Army cavalry units for 20 years
collected data on soldiers' deaths from horse/mule kicks

Deaths	Observed Units	Predicted Units
0	109	108.7
1	65	66.3
2	22	20.2
3	3	4.0
4	1	0.6
5+	0	0.2

Compute predicted *relative* frequencies based on: 122 deaths ÷ 200 unit-years = μ = 0.61
Compute predicted frequencies by multiplying *relative* frequency x 200 unit-years

Code for frequencies: `200*dpois(c(0:4), .61)`
Code for cumulative frequencies: `200*ppois(c(0:4), .61, lower.tail = TRUE)`

2. Poisson Regression for Count Data

Poisson distribution: $P(y = h) = e^{-\mu}\dfrac{\mu^h}{h!}$, $h = 0, 1, 2, \ldots$

Exponential regression: $\mu(\mathbf{x}) = \exp(\boldsymbol{x\beta}) = \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k) > 0$

Combine the two: $P(y = h|\mathbf{x}) = \exp\left[-\exp(\boldsymbol{x\beta})\right]\left[\exp(\boldsymbol{x\beta})\right]^h / h!,\ \ h = 0, 1, 2, \ldots$

Marginal effects (uses chain rule): $\dfrac{\partial\mu(\mathbf{x})}{\partial x_j} = \exp(\boldsymbol{x\beta})\beta_j = \mu(\mathbf{x})\beta_j$

$$\Rightarrow \beta_j = \frac{\frac{\partial\mu(\mathbf{x})}{\mu(\mathbf{x})}}{\partial x_j}$$

Exponential regression and marginal effects calculations look almost identical to log-linear regression in Lecture 15, *except* focus is on change in μ not y

To interpret β_j: $100 \cdot [\exp(\hat{\beta}_j \Delta x_j) - 1]$ tells you the percent or proportional change in $\mu(x)$ from a Δx_j–unit change in x_j

POLS 6481. Research Design and Quantitative Methods II
Lecture 16. Poisson Regression for Counts
Readings: Wooldridge, *Introductory Econometrics 5e*, 17.3

2. Poisson Regression for Count Data in R

Poisson model: `pmodel <- glm( y ~ x, data=mydata, family=poisson)`

Compare to: `model.ols <- lm(y~x, data=mydata)`

Robust std. errors: `install.packages("sandwich"); library(sandwich)`

Longer version:

```
robvar <- vcovHC(pmodel, type="HC0")
robse <- sqrt(diag(cov.m))
r.est <- cbind("Estimate" = coef(pmodel), "Robust SE" = robse,
        "p value" = 2*pnorm(abs(coef(pmodel)/robse), lower.tail=FALSE),
        LL = coef(pmodel) - 1.96*robse, UL = coef(pmodel) + 1.96*robse)
r.est
```

Shorter version:

```
install.packages("lmtest"); library(lmtest)
coeftest(pmodel, vcov = sandwich)
```

Readings: Wooldridge, *Introductory Econometrics 5e*, 17.3

3. Example: *Elephant Mating Behavior*

Load dataset: elephants.csv

Inspect data:

```
plot(Matings ~ Age, pch=19, cex=.75)
abline(lm(Matings~Age), lwd=2, col="blue")
```

Poisson model:

```
pmodel <- glm(Matings ~ Age, poisson)
summary(pmodel)
```

Interpret coefficent:

```
beta  <- pmodel$coef
exp(beta[2])
exp(confint(model,2))
```

Interpretation is that a 1 year increase in age yields 7.1 % increase in mean number of mates:

$$\widehat{\beta_{Age}} = 0.06869$$

$$\exp(\widehat{\beta_{Age}}) = 1.071107$$

$$\%\ \Delta y = 100 \cdot [\exp(\widehat{\beta_{Age}}) - 1] = 100 \cdot [.071107] = 7.1107\ \%$$

See script (Lecture 16 elephants.R); includes plot with Poisson regression curve

4. Poisson Regression Predicted Values

Recall from earlier: $P(y = h|\mathbf{x}) = \exp[-\exp(\boldsymbol{x\beta})][\exp(\boldsymbol{x\beta})]^h / h!, \; h = 0, 1, 2, \ldots$

where $\exp(\boldsymbol{x\beta}) = \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k) > 0$

Use $\widehat{\beta_0} = -1.5820$ and $\widehat{\beta_{Age}} = 0.0687$

If *Age* = 39, then: $x\beta = 1.097$
$\exp(x\beta) = 2.995$

A 39-year old elephant is expected to have roughly 3.0 matings.

P(h = 0 | *Age* = 39) = 5.0 %
P(h = 1 | *Age* = 39) = 15.0 %
P(h = 2 | *Age* = 39) = 22.4 %
P(h = 3 | *Age* = 39) = 22.4 %
P(h = 4 | *Age* = 39) = 16.8 %
P(h = 5 | *Age* = 39) = 10.0 %
...

If *Age* = 40, then: $x\beta = 1.166$
$\exp(x\beta) = 3.208$

A 40-year old elephant is expected to have roughly 3.2 matings.

P(h = 0 | *Age* = 40) = 4.0 %
P(h = 1 | *Age* = 40) = 13.0 %
P(h = 2 | *Age* = 40) = 20.8 %
P(h = 3 | *Age* = 40) = 22.2 %
P(h = 4 | *Age* = 40) = 17.8 %
P(h = 5 | *Age* = 40) = 11.5 %
...

5. Example: *Extramarital Affairs*

Dataset is Affairs (dataframe in AER package)

Population is married American adults; sample is 601 respondents to 1969 *Psychology Today*

Dependent variable: *affairs* (number of affairs in past year)

Explanatory variables:

gender factor indicating gender.

age numeric variable coding age in years: 17.5 = under 20, 22 = 20–24, 27 = 25–29, 32 = 30–34, 37 = 35–39, 42 = 40–44, 47 = 45–49, 52 = 50–54, 57 = 55 or over.

yearsmarried numeric variable coding number of years married: 0.125 = 3 months or less, 0.417 = 4–6 months, 0.75 = 6 months–1 year, 1.5 = 1–2 years, 4 = 3–5 years, 7 = 6–8 years, 10 = 9–11 years, 15 = 12 or more years.

children factor. Are there children in the marriage?

religiousness numeric variable coding religiousness: 1 = anti, 2 = not at all, 3 = slightly, 4 = somewhat, 5 = very.

education numeric variable coding level of education: 9 = grade school, 12 = high school graduate, 14 = some college, 16 = college graduate, 17 = some graduate work, 18 = master's degree, 20 = Ph.D., M.D., or other advanced degree.

occupation numeric variable coding occupation according to Hollingshead classification (reverse numbering).

rating numeric variable coding self rating of marriage: 1 = very unhappy, 2 = somewhat unhappy, 3 = average, 4 = happier than average, 5 = very happy.

Readings: Wooldridge, *Introductory Econometrics 5e*, 17.3

5. Example: *Extramarital Affairs*

Dataset is Affairs (dataframe in AER package or just affairs.dta)

Load data frame:

```
install.packages("AER"); data( "Affairs", package = "AER" )
```

Inspect data*:

```
hist(Affairs$affairs)
plot(Affairs$affairs ~ Affairs$rating, pch=19, cex=.75)
abline(lm(affairs ~ rating, data=Affairs), lwd=2, col="blue")
```

OLS regression:

```
reg.B <- lm(affairs ~ age + yearsmarried + religiousness + occupation +
rating, data = Affairs)
```

Poisson regression:

```
pois.B <- glm(affairs ~ age + yearsmarried + religiousness + occupation
+ rating, data = Affairs, family=poisson)
```

Compare results:

```
library(stargazer); stargazer(reg.B, pois.B, type="text", title="",
single.row=FALSE, omit.stat=c("f", "ser"))
```

* Notice the high proportion of zeroes!

5. Example: *Extramarital Affairs*

Recall specification: affairs ~ age + yearsmarried + religiousness + occupation + rating

Interpret coefficient for **age**:

```
beta <- pois.B$coef
exp(beta[2])
exp(confint(pois.B,2))
```

Each "year" increase in **age** yields 3% decrease in mean number of affairs:

$$\beta_{\text{age}} = -0.0322553$$
$$\exp(\beta_{\text{age}}) = 0.9682594$$
$$\%\,\Delta y = 100\cdot[\exp(\beta_{\text{age}})-1] = 100\cdot[-0.03174064] = -3.17\ \%$$

Interpret coefficient for **rating**:

```
exp(beta[6])
exp(confint(pois.B,2))
```

Each unit increase in **rating** yields 34% decrease in mean number of affairs:

$$\beta_{\text{rating}} = -0.409$$
$$\exp(\beta_{\text{rating}}) = 0.664$$
$$\%\,\Delta y = 100\cdot[\exp(\beta_{\text{rating}})-1] = 100\cdot[-.336] = -33.6\ \%$$

POLS 6481. Research Design and Quantitative Methods II
Lecture 16. Poisson Regression for Counts
Readings: Wooldridge, *Introductory Econometrics 5e*, 17.3

6. Zero·Inflated Poisson Regression

Histogram for last example showed too many zeroes (which meant it was over-dispersed!)

Code to check o/u-dispersion:

```
E2 <- resid(model, type = "pearson")
N <- nrow(data)
p <- length(coef(model))
sum(E2^2) / (N–p)
```

Code for zero·inflated Poisson:

```
install.packages("pscl")
library(pscl)
zipmodel <- zeroinfl(y ~ x | z, dist = "poisson", data)
vuong(pmodel, zipmodel)
```

Note that x and z can be the same variables