Poisson Regression

ehsan.karim@ubc.ca

Nov 5, 2020

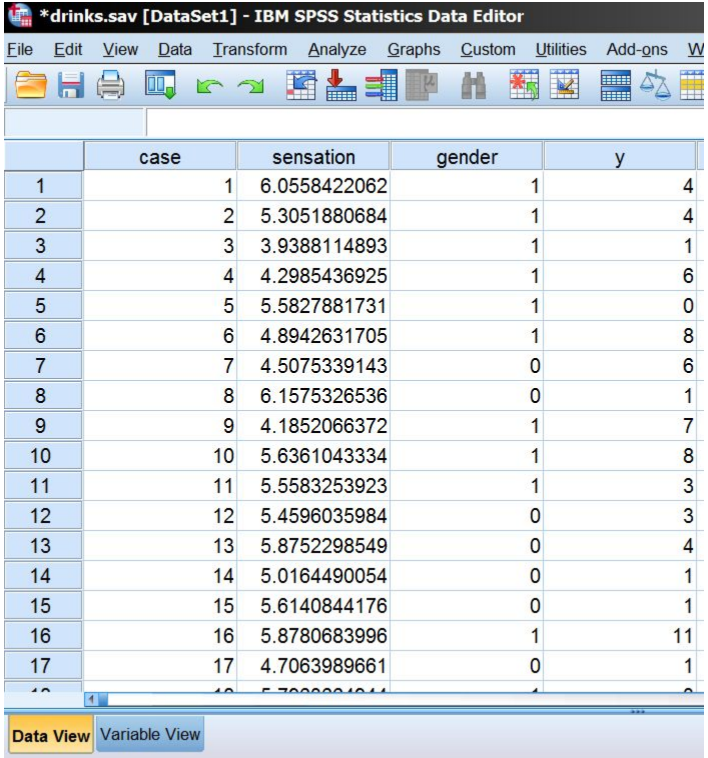
SPPH 504/007

Reference

- Stefany Coxe , Stephen G. West & Leona S. Aiken (2009) The Analysis of Count Data: A Gentle Introduction to Poisson Regression and Its Alternatives, Journal of Personality Assessment, 91:2, 121-136, DOI: 10.1080/00223890802634175

Drinks Data

- Outcome = y = count variable
- *A count variable* is a variable that takes on discrete values
    - 0,
    - 1,
    - 2,
    - . . .
- A count variable can only take on positive integer values or zero because an event cannot occur a negative number of times.



	case	sensation	gender	y
1	1	6.0558422062	1	4
2	2	5.3051880684	1	4
3	3	3.9388114893	1	1
4	4	4.2985436925	1	6
5	5	5.5827881731	1	0
6	6	4.8942631705	1	8
7	7	4.5075339143	0	6
8	8	6.1575326536	0	1
9	9	4.1852066372	1	7
10	10	5.6361043334	1	8
11	11	5.5583253923	1	3
12	12	5.4596035984	0	3
13	13	5.8752298549	0	4
14	14	5.0164490054	0	1
15	15	5.6140844176	0	1
16	16	5.8780683996	1	11
17	17	4.7063989661	0	1

Why not use Multiple Linear Regression?

- MLR Assumption:
  - Residuals need to be **normally distributed**.
  - **Homoscedasticity** is required.
  - Residuals need to be **independent**.
- Count outcome variables can violate the first two assumptions of MLR in several ways.

Generalized Linear Model

- GLM generalizes MLR for use with many different types of error structures and dependent variables.

- The GLM family of analyses can provide accurate results for data sets having
  - binary,
  - ordered categorical,
  - count, and
  - time to failure (or success) dependent variables.

- Poisson regression is a member of GLM family.

GLM

The GLM introduces two major modifications to the MLR framework.

(1) It allows transformations of the predicted outcome, which can linearize a potentially nonlinear relationship between the Y and X.

In GLM, there is a special transformation function called the link function.

- In Poisson regression, the link function is the **natural log** (i.e., log e or ln).

GLM

The GLM introduces two major modifications to the MLR framework.

(2) GLM is flexible in error structure: MLR assumes a normal error structure, whereas GLM allows for a variety of other error structures.

- Poisson regression:

  distribution of the errors ~ Poisson distribution

1. GLM link

Common distributions with typical uses and canonical link functions

Distribution	Support of distribution	Typical uses	Link name	Link function	Mean function
Normal	real: $(-\infty, +\infty)$	Linear-response data	Identity	$\mathbf{X}\boldsymbol{\beta} = \mu$	$\mu = \mathbf{X}\boldsymbol{\beta}$
Exponential	real: $(0, +\infty)$	Exponential-response data, scale parameters	Inverse	$\mathbf{X}\boldsymbol{\beta} = -\mu^{-1}$	$\mu = -(\mathbf{X}\boldsymbol{\beta})^{-1}$
Gamma					
Inverse Gaussian	real: $(0, +\infty)$		Inverse squared	$\mathbf{X}\boldsymbol{\beta} = -\mu^{-2}$	$\mu = (-\mathbf{X}\boldsymbol{\beta})^{-1/2}$
Poisson	integer: $0, 1, 2, \ldots$	count of occurrences in fixed amount of time/space	Log	$\mathbf{X}\boldsymbol{\beta} = \ln(\mu)$	$\mu = \exp(\mathbf{X}\boldsymbol{\beta})$
Bernoulli	integer: $\{0, 1\}$	outcome of single yes/no occurrence	Logit	$\mathbf{X}\boldsymbol{\beta} = \ln\left(\frac{\mu}{1-\mu}\right)$	$\mu = \frac{\exp(\mathbf{X}\boldsymbol{\beta})}{1+\exp(\mathbf{X}\boldsymbol{\beta})} = \frac{1}{1+\exp(-\mathbf{X}\boldsymbol{\beta})}$
Binomial	integer: $0, 1, \ldots, N$	count of # of "yes" occurrences out of N yes/no occurrences			
Categorical	integer: $[0, K)$	outcome of single K-way occurrence			
	K-vector of integer: $[0, 1]$, where exactly one element in the vector has the value 1				
Multinomial	K-vector of integer: $[0, N]$	count of occurrences of different types (1 .. K) out of N total K-way occurrences			

2. GLM error distribution

- The probability density of the **normal distribution** depends on two parameters, the mean μ and SD σ.

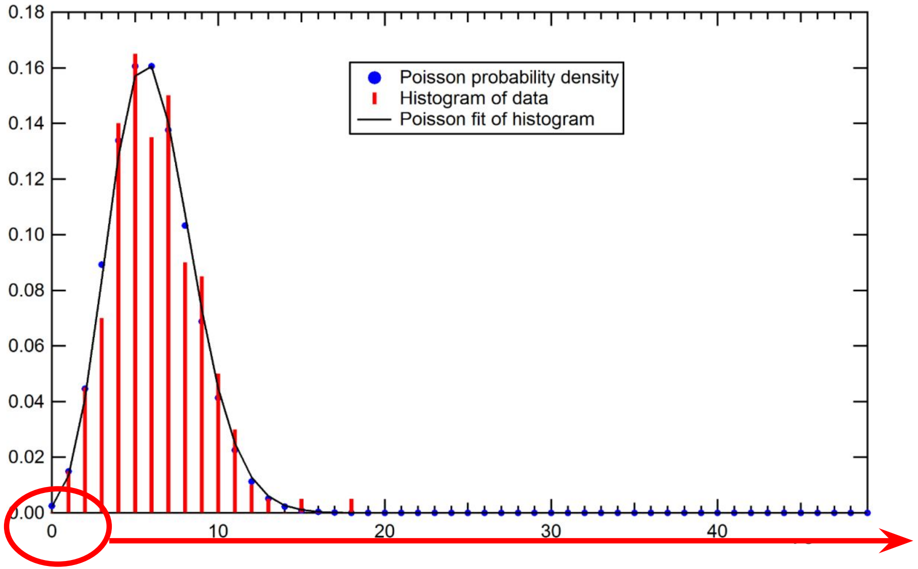
$$f(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/2\sigma^2}.$$

- **Poisson distribution** is specified by only one parameter μ. The parameter μ defines both **the mean and the variance** of the distribution.

$$P(Y = y|\mu) = \frac{\mu^y}{y!} e^{-\mu}$$

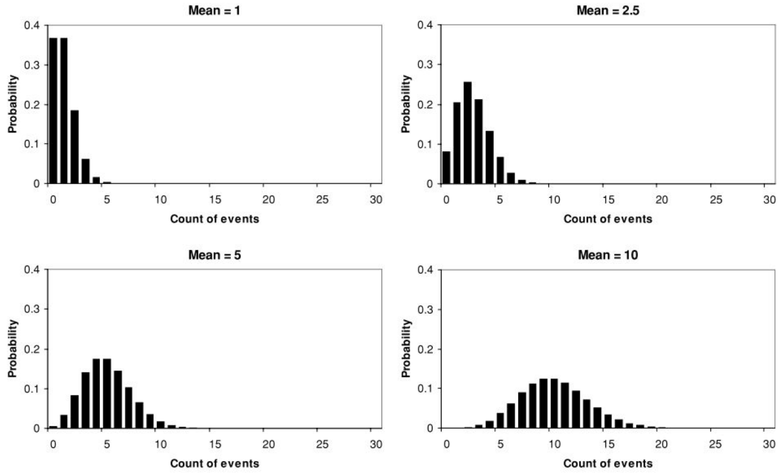
Poisson Distribution Properties

- Poisson distribution is a **discrete distribution** that takes on a probability value only for nonnegative integers;
  - **0 or greater**.



Poisson Distribution Properties

- Poisson distribution with an **expected value greater than 10** approaches a <u>normal distribution</u> in shape and symmetry.



Review of Logistic Regression

$$logit(\hat{\pi}) = \ln\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = b_0 + b_1X_1 + b_2X_2 + \cdots + b_pX_p,$$

- Logistic regression equation is also linear; specifically, logistic regression is "linear in the logit".

Poisson Regression

$$\ln(\hat{\mu}) = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_p X_p$$

- μ-hat is the predicted count on the outcome variable given the specific values on the predictors X_1, X_2,..., X_p.
- linear relationship between each predictor and the predicted score just as in MLR.
- "linear in the logarithm."
- The **residuals** of a Poisson regression model are assumed to be Poisson distributed.

Poisson Regression Interpretation

- Interpretation: For a 1-unit increase in *X1*, the predicted count (*μ*-hat) is *multiplied by exp(b1)* , holding all other variables constant (multiplicative).

- Here, the unstandardized (raw) regression coefficient *b1* is exponentiated.

- effect of a <u>1-unit change in *X1* on the outcome</u>

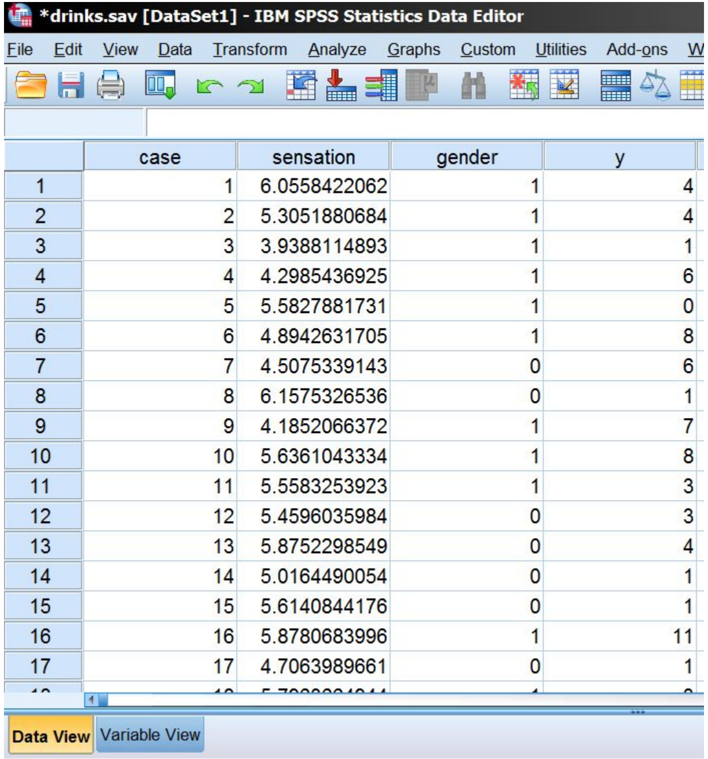
$$e^{b_1(X_1+1)} = e^{b_1 X_1 + b_1} = e^{b_1 X_1} e^{b_1}.$$

Drinks Data

- Outcome = y = the number of alcoholic drinks that an individual consumes on one particular Saturday night during the study

- X1 = sensation = an eight-item subscale of sensation seeking (excitement seeking), potentially ranging from 1 (low) to 7 (high).

- X2 = Gender:
  - 0 = female and
  - 1 = male



	case	sensation	gender	y
1	1	6.0558422062	1	4
2	2	5.3051880684	1	4
3	3	3.9388114893	1	1
4	4	4.2985436925	1	6
5	5	5.5827881731	1	0
6	6	4.8942631705	1	8
7	7	4.5075339143	0	6
8	8	6.1575326536	0	1
9	9	4.1852066372	1	7
10	10	5.6361043334	1	8
11	11	5.5583253923	1	3
12	12	5.4596035984	0	3
13	13	5.8752298549	0	4
14	14	5.0164490054	0	1
15	15	5.6140844176	0	1
16	16	5.8780683996	1	11
17	17	4.7063989661	0	1

Drinks Data

- Outcome as scale

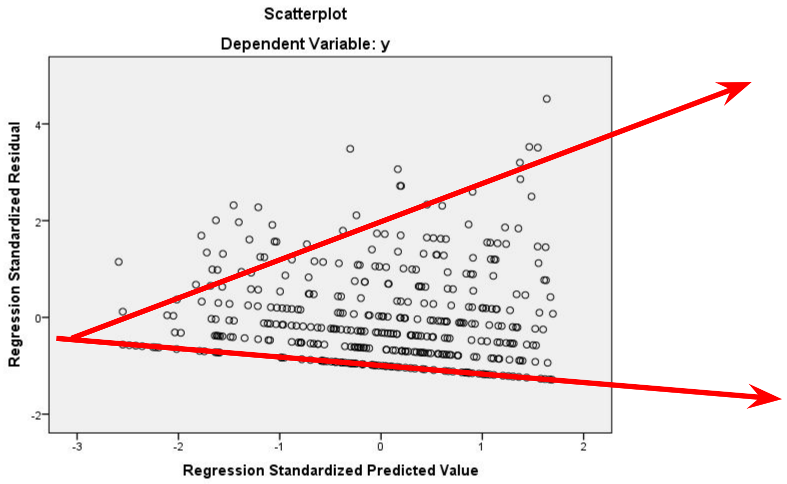
Continuous Variable Information						
		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	y	400	0	17	2.93	2.972
Covariate	sensation	400	3.194107179	6.487512912	5.183779405	.7682502918

- Mean of y not close to 3. Variance is close to 10.
- Mean and the variance (not SD) are not similar.

Drinks Data Analysis (1)

- Linear regression fit

Heterogeneity of variance indicating that MLR is not appropriate for these data.



Drinks Data Analysis (2)

- Poisson regression fit
- Equation for predicted number of alcoholic drinks consumed: μ-hat = *exp*(–0.140) x *exp*(0.231) sensation.
- *Intercept*: exp(0.140) = 0.87 is the predicted number of alcoholic drinks consumed by a person who has a score of zero on the sensation-seeking measure; (+ve value as opposed to –ve value from MLR).

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-.140	.2128	-.557	.277	.435	1	.510
sensation	.231	.0397	.154	.309	34.065	1	.000
(Scale)	1[a]						

Dependent Variable: y
Model: (Intercept), sensation

a. Fixed at the displayed value.

Drinks Data Analysis (2)

- Poisson regression fit
- Equation for predicted number of alcoholic drinks consumed: μ-hat = $exp(-0.140) \times exp(0.231)$ sensation.
- *Slope:* The exponentiation of the regression coefficient for sensation seeking, $exp(0.231) = 1.26$, is the predicted multiplicative effect of a 1-unit change in sensation seeking on number of alcoholic drinks consumed.

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-.140	.2128	-.557	.277	.435	1	.510
sensation	.231	.0397	.154	.309	34.065	1	.000
(Scale)	1[a]						

Dependent Variable: y
Model: (Intercept), sensation

a. Fixed at the displayed value.

Drinks Data Analysis (2)

- Poisson regression fit

$$R^2_{\text{deviance}} = 1 - \frac{\text{deviance(fitted\_model)}}{\text{deviance(intercept\_only)}}.$$

Deviance is a measure of lack of fit; this **measure is reduced by adding predictors** to the intercept-only model **if the predictors have some accuracy in accounting for the outcome**.

- Poisson regression (like logistic) has no direct analogue to R-square. *Deviance* value for the model can be used to assess fit of the model. For comparison, we need **nested** model.

Drinks Data Analysis (2)

- Poisson regression fit
  - ✔ Poisson distribution has **one parameter μ**, which characterizes both the mean and the variance of the distribution.
  - ✔ The Poisson model assumes that the conditional **mean and variance are equal**, a condition known as *equidispersion*.
  - ✔ The situation in which the variance is larger than the mean is known as *overdispersion*.

Drinks Data Analysis (3)

- Overdistersed Poisson regression fit (Alternative 1)
  - The simplest adjustment for overdispersion that can be made to the Poisson regression model is the **overdispersed Poisson model**.
  - This model includes a **second parameter** that is used in the estimation of the conditional variance known as the *overdispersion scaling parameter*, φ.
  - The model estimated with this correction now essentially assumes an error distribution that is Poisson with **mean μ and variance $\varphi\mu$**.
    - The scaling parameter φ will be greater than 1 if **overdispersion** is present in the data;
    - φ will be equal to 1 if there is **equidispersion**, and the resulting model is equivalent to the standard Poisson regression model.
    - φ will be less than 1 if the data are **underdispersed**.

Drinks Data Analysis (4)

- Negative binomial regression fit (Alternative 2)
  - The negative binomial model also accounts for overdispersion.
    - An *α* parameter (*dispersion parameter; but different than scale parameter in poisson*) greater than 0 indicates that overdispersion is present;
    - Larger values indicate more overdispersion.
  - **Interpretation** of regression coefficients for the negative binomial model is identical to that for the standard Poisson model.
  - A **pseudo-*R*-square** (such as was calculated for Poisson regression) **cannot be calculated** for negative binomial models when dispersion parameter is introduced (will not be nested).

Drinks Data Analysis (4)

- (Overdispersed) Negative binomial regression fit
  - The exponentiation of the regression coefficient for sensation seeking, $exp(0.220) = 1.25$, is the multiplicative effect of a 1-unit change in sensation seeking on number of alcoholic drinks consumed, **allowing for heterogeneity between individuals**.

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-.078	.3350	-.735	.578	.054	1	.816
sensation	.220	.0635	.095	.344	11.969	1	.001
(Scale)	.722[a]						
(Negative binomial)	1[b]						

Dependent Variable: y
Model: (Intercept), sensation

a. Computed based on the Pearson chi-square.

b. Fixed at the displayed value.

Drinks Data Analysis (4)

- (Overdispersed) Negative binomial regression fit
  - The estimate of α for this model is 0.722, which is greater than 0, indicating that there is overdispersion in the data.

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-.078	.3350	-.735	.578	.054	1	.816
sensation	.220	.0635	.095	.344	11.969	1	.001
(Scale)	.722[a]						
(Negative binomial)	1[b]						

Dependent Variable: y
Model: (Intercept), sensation

a. Computed based on the Pearson chi-square.

b. Fixed at the displayed value.

Drinks Data Analyses

	MLR	Poisson	Overdispersed Poisson	(Overdispersed) Negative Binomial
Sensation	.652	**.231**	**.231**	**.220**
SE	.191	.0397	.0669	.0635
P-value	.001	<.0001	.001	0.001
Scale		1	2.847	1
Alpha (scale)		0	0	.722

Summary

- Poisson regression is often used for modeling **count data**. It has a number of extensions useful for count models.
- 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.

Thanks!

ehsan.karim@ubc.ca