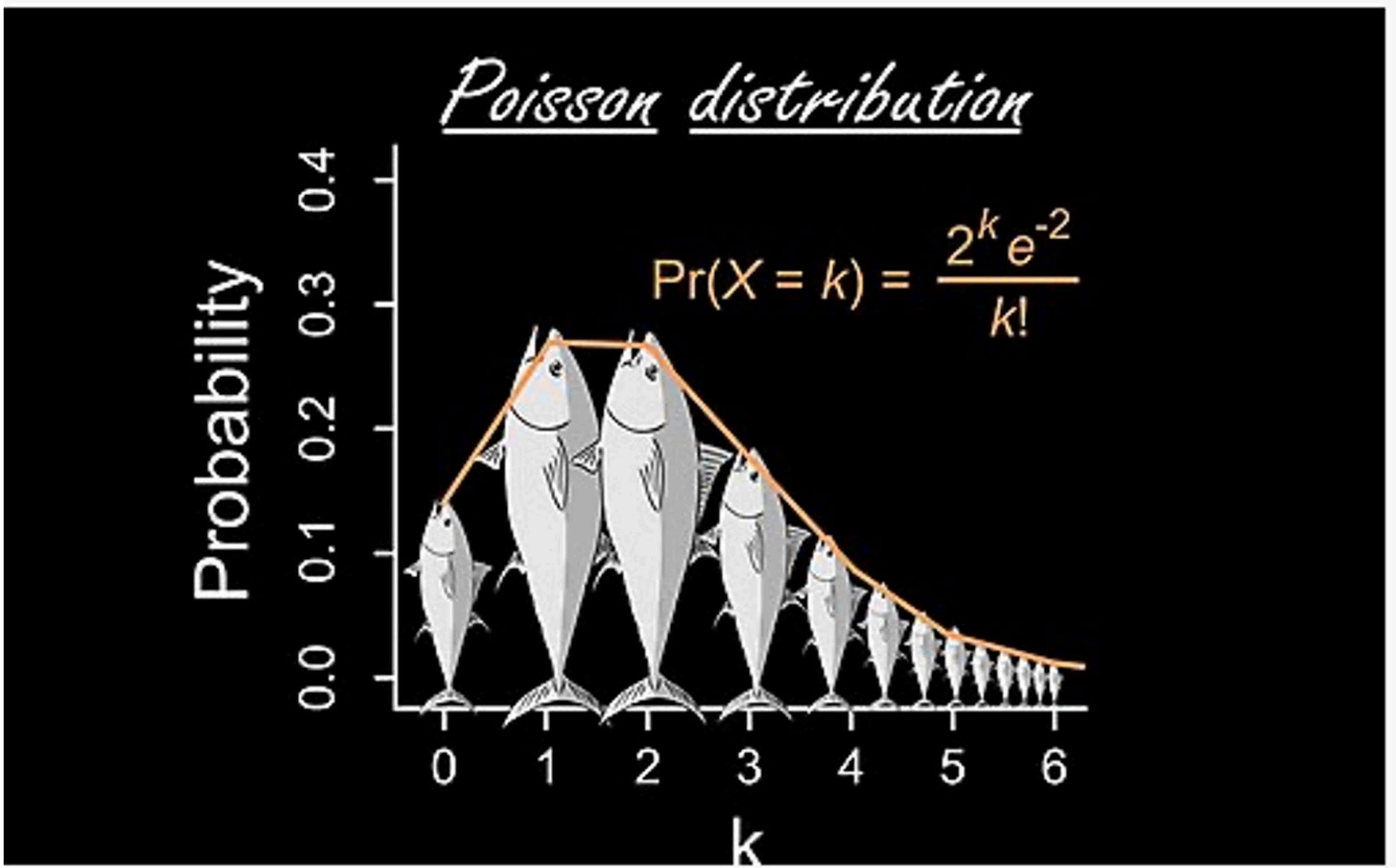
Poisson Regression - Basic concepts

Mabel Carabali

EBOH, McGill University

updated: 2024-11-27


Poisson distribution
Probability
0.0
0.1
0.2
0.3
0.4
$\Pr(X = k) = \frac{2^k e^{-2}}{k!}$
0
1
2
3
4
5
6
k

Expected competencies:

- Knowledge about linear, logistic and binomial regressions.
- Knowledge about count data and Poisson distribution.

Objectives

- Revise Poisson distribution and model assumptions
- Provide tools to estimate Risk Ratios, Risk Difference and Rate Ratios using Poisson regressions

Poisson distribution

The Poisson distribution gives the probability mass function (discrete) of all possible numbers of new cases, from 0 to ∞.

If Y_i= # of new cases with λ = expected number of cases in a given time period and

$$Y_i \sim Poisson(\lambda)$$

then the probability mass function that $Y = k$, is:

$$P(Y = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

for $k = 0, 1, 2,$ ... and $\lambda > 0$, where λ is both the mean and the variance of K; where λ = expected number of cases in a given time period.

Poisson distribution

This can be also presented as:

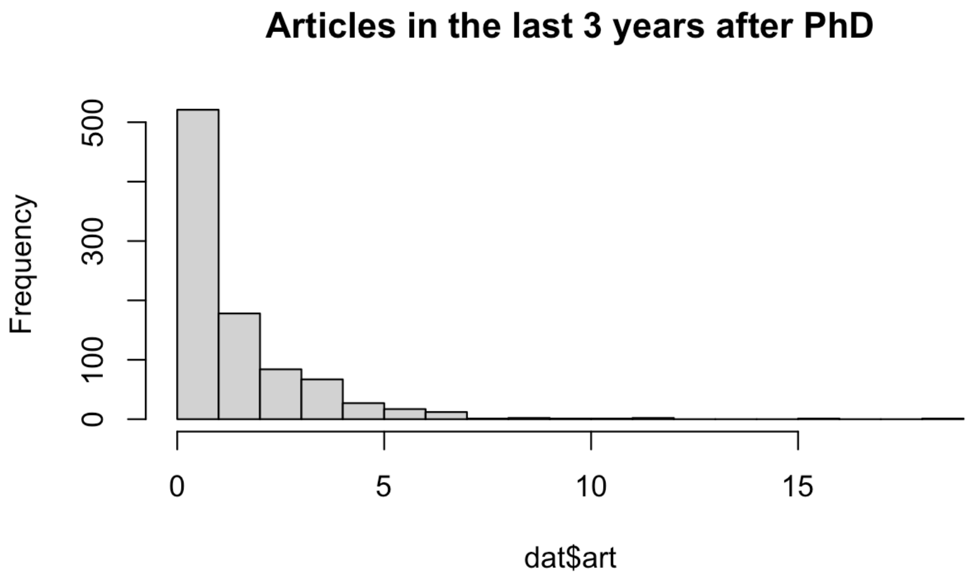
$$y_i \sim Poisson(e^{X_i\beta})$$

where the linear predictor $X_i\beta$ is the logarithm of the expected value of measurement y_i, or $ln(\lambda)$.

- In count-data regressions, each unit i corresponds to a setting (typically a spatial location or a time interval) in which y_i events are observed.
- Under the Poisson model, $sd(y_i) = \sqrt{E}(y_i)$; thus if the model accurately describes the data, we also have a sense of how much variation we would expect from the fitted curve.

For a Poisson random variable, the variance = mean = E(Y) = λ

Poisson distribution



Checking distribution

```
mean(dat$art)
```

```
## [1] 1.692896
```

```
sd(dat$art); var(dat$art)
```

```
## [1] 1.926069
```

```
## [1] 3.709742
```

Poisson considerations

- As λ increases Poisson distribution approximates to a normal distribution
- When n is large and p is very small, Poisson distribution approximates the binomial distribution.

Example: Consider binomial distribution with probability = 0.01. What is the probability of 6 occurrences in 1200 tries?

```r
# binomial
dbinom(6,1200,0.01)
```

```
## [1] 0.02516172
```

```r
# poisson
dpois(6,1200*0.01) # lambda = n*p
```

```
## [1] 0.02548128
```

To approximate the true (binomial) results:

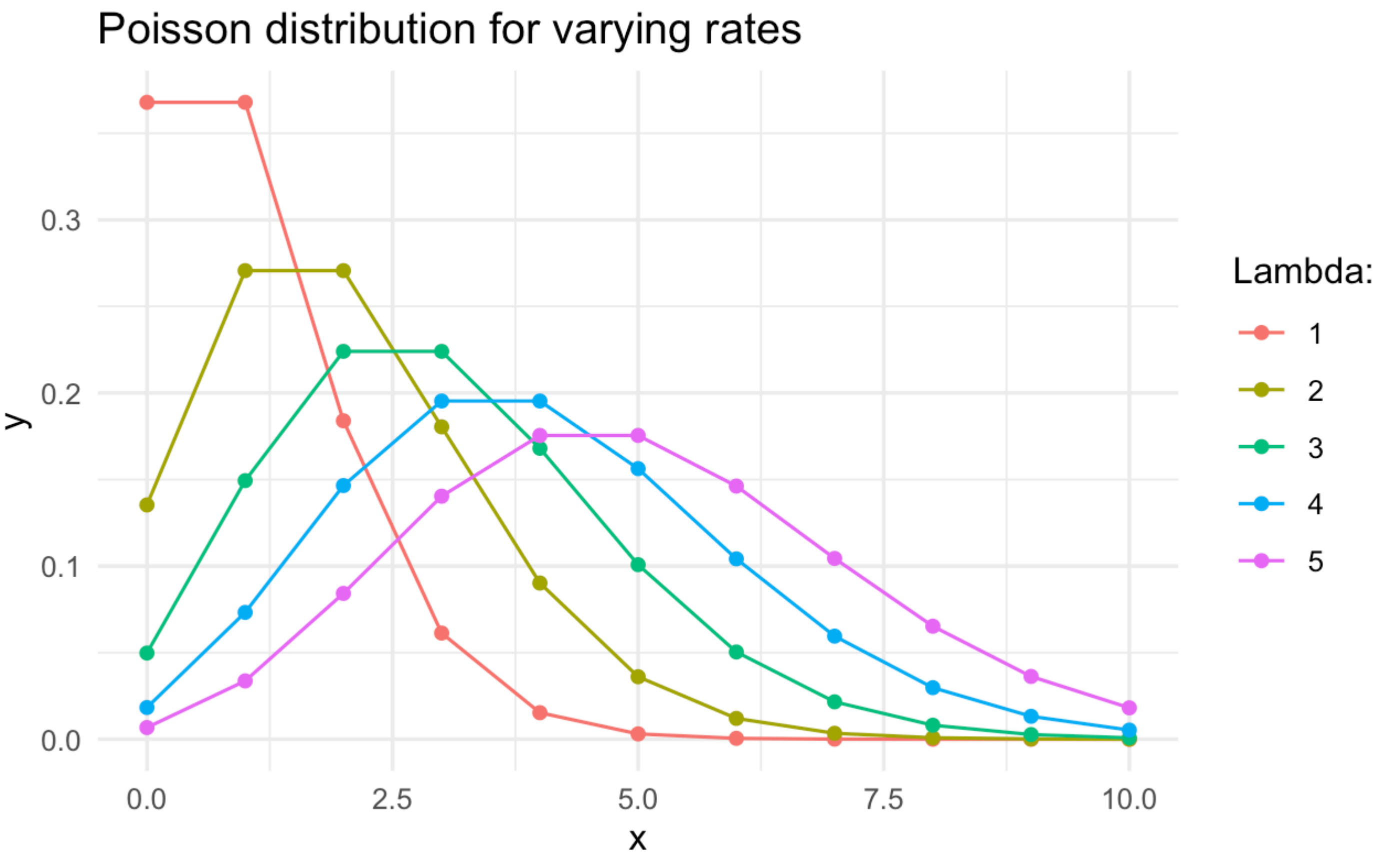
To approximate the mean

- $\mu = E(X) = \lambda = np$

To approximate the standard deviation.

- $\sigma = sqrt(\lambda) = sqrt(np)$

Poisson considerations



Poisson Model Assumptions

Poisson Response: The response variable is a count per unit of time or space.

Independence: The observations must be independent of one another.

Mean = Variance: Conditional means equal the conditional variances.

Linearity: The log of the mean rate, log(λ), must be a linear function of x.

- the mean values of Y at each level of X, $\lambda_{Y|X}$, fall on a curve, not a line, although the logs of the means should follow a line.

Poisson Model

- A generalized linear model with a log link and a Poisson distribution,
- Used to model count variables
- A natural model for rates in data sets with person time (e.g: Mortality).

- Under predicts zero counts and over predicts the first set of counts
- Assumes same rate of outcome among the individual observations but incorporating "observed heterogeneity" relax this assumption.

Illustration using the articles and PhD years after graduation

Characteristic	1">N = 915[1]
art	
Mean; Median (Q1, Q3)	1.69; 1.00 (0.00, 2.00)
fem	
Men	494 / 915 (54%)
Women	421 / 915 (46%)
mar	
Married	606 / 915 (66%)
Single	309 / 915 (34%)
phd	
Mean; Median (Q1, Q3)	3.10; 3.15 (2.26, 3.92)

[1] n / N (%)

Characteristic	1">N = 915[1]
profage	
Mean; Median (Q1, Q3)	33.2; 33.0 (31.0, 36.0)
kid5	
0	599 / 915 (65%)
1	195 / 915 (21%)
2	105 / 915 (11%)
3	16 / 915 (1.7%)
ment	
Mean; Median (Q1, Q3)	9; 6 (3, 12)

[1] n / N (%)

Illustration using the articles and PhD years after graduation

```r
pois0 <- glm(art ~ 1, family= "poisson", data = dat)
```

Exponentiated coefficient

```
##             exp(Est.) 2.5% 97.5% z val. p
## (Intercept)      1.69 1.61  1.78  20.72 0
```

Summary

```
## 
## Call:
## glm(formula = art ~ 1, family = "poisson", data = dat)
## 
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)    
## (Intercept)  0.52644    0.02541   20.72   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for poisson family taken to be 1)
## 
##     Null deviance: 1817.4  on 914  degrees of freedom
## Residual deviance: 1817.4  on 914  degrees of freedom
## AIC: 3487.1
## 
## Number of Fisher Scoring iterations: 5
```

Illustration using the articles and PhD years after graduation

Observations	915
Dependent variable	art
Type	Generalized linear model
Family	poisson
Link	log

$\chi^2(0)$	-0.00
Pseudo-R^2 (Cragg-Uhler)	0.00
Pseudo-R^2 (McFadden)	0.00
AIC	3487.15
BIC	3491.97

	exp(Est.)	**2.5%**	**97.5%**	**z val.**	**p**
(Intercept)	1.69	1.61	1.78	20.72	0.00

Standard errors: MLE

Illustration using the articles and PhD years after graduation

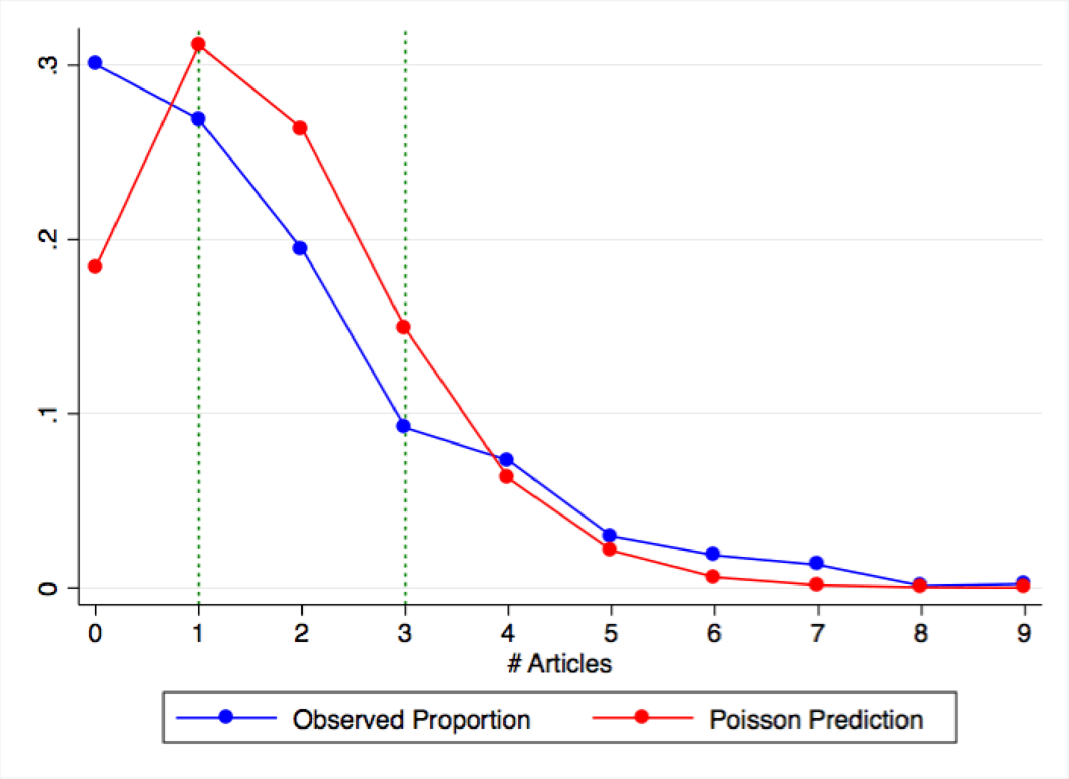


Illustration using the articles and PhD years after graduation

If the follow-up time is not specified, the software assumes that person time is one for everyone, and Poisson regression actually produces the RISK ratio:

```
pois1 <- glm(art ~ fem + mar,
             poisson(link="log"),
             data = dat)
round(jtools::summ(pois1, confint= T,
                   exp = T)$"coeftable", 2)
```

```
##               exp(Est.) 2.5% 97.5% z val.    p
## (Intercept)        1.89 1.77  2.03  18.19 0.00
## femWomen           0.79 0.71  0.87  -4.49 0.00
## marSingle          0.97 0.87  1.09  -0.50 0.62
```

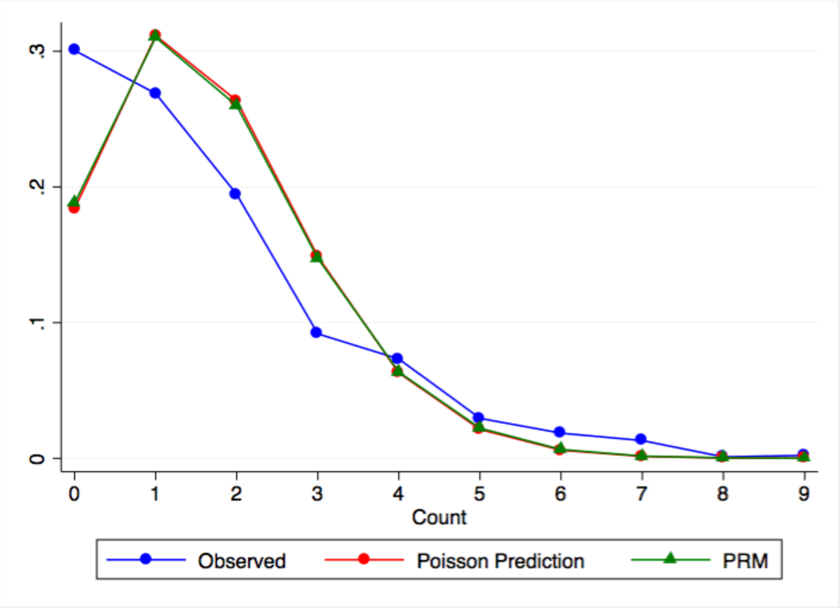


Illustration using the articles and PhD years after graduation

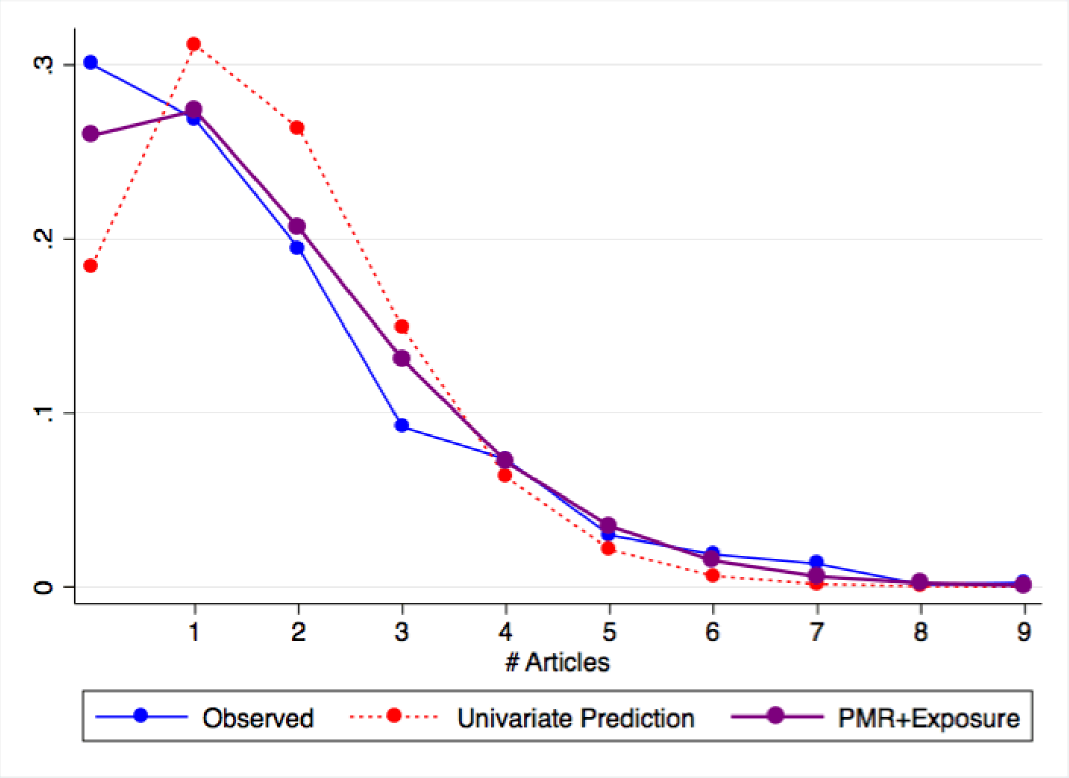
To get the RATE ratio instead of RISK ratio, we need to tell the software where to find the follow-up time:

```
dat$logprofage <- log(dat$profage)
#summary(dat$profage)
pois2a <- glm(art ~ fem + mar + offset(logprofage),
             poisson(link="log"), data = dat)
```

```
##               exp(Est.) 2.5% 97.5% z val.   p
## (Intercept)        0.06 0.05  0.06 -81.49 0.0
## femWomen           0.79 0.72  0.88  -4.30 0.0
## marSingle          0.97 0.87  1.08  -0.52 0.6
```

The exponentiated coefficient give then the Rate Ratio.

Illustration using the articles and PhD years after graduation



comparing "goodness-of-fit" using AIC (Akaike Information Criteria)

```
##         Empty     Risk     Rate
## [1,] 3487.147 3467.868 3470.165
```

Illustration using the articles and PhD years after graduation

```r
pois2b <- glm(art ~ fem + mar + offset(profage),
              poisson(link="log"), data = dat)
round(jtools::summ(pois2b, confint= T, exp = T)$"coeftable", 2)
```

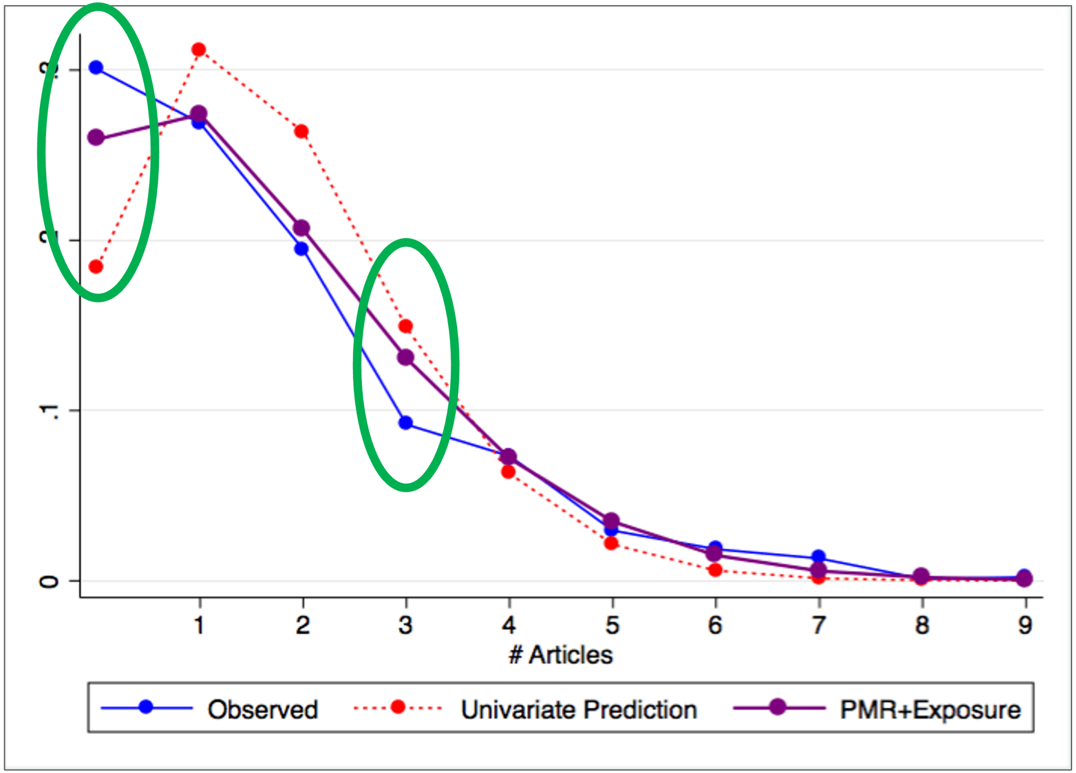
```
##             exp(Est.) 2.5% 97.5%   z val.    p
## (Intercept)      0.00 0.00  0.00 -997.79 0.00
## femWomen         1.04 0.94  1.15    0.76 0.45
## marSingle        0.71 0.64  0.79   -6.15 0.00
```

Using the variable age without the log-transformation changes the estimates and the likelihood is affected as evidenced by the AIC

comparing "goodness-of-fit" using AIC (Akaike Information Criteria)

```
##         Empty     Risk     Rate Rate_nolog
## [1,] 3487.147 3467.868 3470.165   14060.01
```

Illustration using the articles and PhD years after graduation



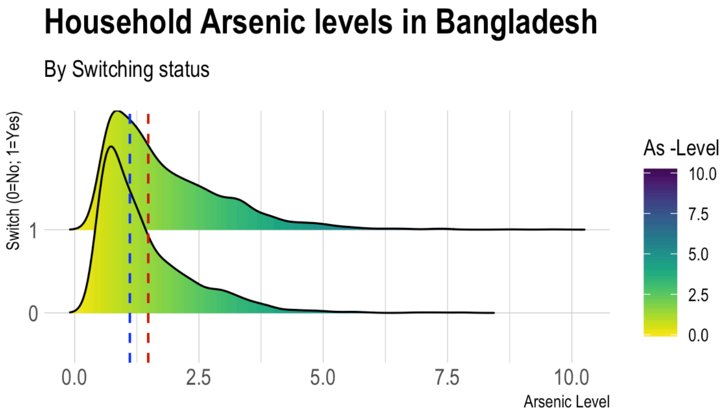
- Under predicts zero counts and over predicts the first set of counts
- Assumes same rate of outcome among the individual observations but incorporating "observed heterogeneity" relax this assumption.

Poisson Models for Risk Ratios and Risk Differences

Recall our example: **Wells in Bangladesh**

Characteristic	[1]">***N = 3,020***[1]
switch	1,737 (58%)
arsenic	
Median (Q1, Q3)	1.30 (0.82, 2.20)
Min, Max	0.51, 9.65
dist100	
Median (Q1, Q3)	0.37 (0.21, 0.64)
Min, Max	0.00, 3.40
assoc	1,277 (42%)

[1] n (%)



Recall our example: Wells in Bangladesh

Model "switching" to a safer well (outcome) as a function of whether the owner belongs to a community association and As. level.

Model	Frequentist Estimates
Linear Model [1]	RD= -0.036; RR= N/A
Logit Model (Predicted probs.) [2]	RD= -0.033; RR= 0.939
AME (prediction `StdGLM`) [2]	RD= -0.031; RR= 0.948
GLM: Log-Binomial [2]	RD= N/A; RR= 0.928

[1] Single predictor; [2] Two predictors

Poisson Model and Risk Ratios

–our example- **Wells in Bangladesh**

```r
mod.glm.p.rr <- glm(switch ~ assoc + arsenic, data = wells,
                    family=poisson(link="log"))
```

```
#> 
#> Call:
#> glm(formula = switch ~ assoc + arsenic, family = poisson(link = "log"), 
#>     data = wells)
#> 
#> Coefficients:
#>             Estimate Std. Error z value Pr(>|z|)    
#> (Intercept)  -0.7489     0.0473  -15.83  < 2e-16 ***
#> assoc        -0.0559     0.0488   -1.14     0.25    
#> arsenic       0.1258     0.0192    6.55  5.6e-11 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#> 
#> (Dispersion parameter for poisson family taken to be 1)
#> 
#>     Null deviance: 1921.5  on 3019  degrees of freedom
#> Residual deviance: 1880.2  on 3017  degrees of freedom
#> AIC: 5360
#> 
#> Number of Fisher Scoring iterations: 5
```

Poisson Model and RRs –our example- **Wells in Bangladesh**

Observations	3020
Dependent variable	switch
Type	Generalized linear model
Family	poisson
Link	log

$\chi^2(2)$	41.25
Pseudo-R^2 (Cragg-Uhler)	0.02
Pseudo-R^2 (McFadden)	0.01
AIC	5360.21
BIC	5378.25

	exp(Est.)	**2.5%**	**97.5%**	**z val.**	**p**
(Intercept)	0.47	0.43	0.52	-15.83	0.00
assoc	0.95	0.86	1.04	-1.14	0.25
arsenic	1.13	1.09	1.18	6.55	0.00
Standard errors: MLE					

Poisson Model and Risk Ratios, our example- Wells in Bangladesh

```
#>              exp(Est.)  2.5% 97.5% z val.        p
#> (Intercept)      0.473 0.431 0.519 -15.83 1.82e-56
#> assoc            0.946 0.859 1.041  -1.14 2.52e-01
#> arsenic          1.134 1.092 1.178   6.55 5.61e-11
```

- Poisson models provides RISK ratios with biased confidence intervals.
- The model assumes a Poisson distribution when in fact the outcomes are binary.
- To overcome issues with the variance, we could use a variance estimator that is robust to the misspecification of the model.

Poisson Model and RRs (Robust Variance)

–our example- **Wells in Bangladesh**

```
jtools::summ(mod.glm.p.rr, confint=T, exp=T,
             robust= "HC0")$"coeftable"
```

```
#>               exp(Est.)  2.5% 97.5% z val.         p
#> (Intercept)       0.473 0.445 0.502 -24.48 2.63e-132
#> assoc             0.946 0.889 1.006  -1.77  7.74e-02
#> arsenic           1.134 1.110 1.159  11.50  1.33e-30
```

Quasi-Poisson Risk Ratio

```
mod.glm.qp.rr <- glm(switch ~ assoc + arsenic, data = wells,
                     family= quasipoisson(link="log"))
jtools::summ(mod.glm.qp.rr, confint=T, exp=T)$"coeftable"
```

```
#>               exp(Est.)  2.5% 97.5% t val.         p
#> (Intercept)       0.473 0.445 0.502 -24.31 2.19e-119
#> assoc             0.946 0.888 1.006  -1.76  7.90e-02
#> arsenic           1.134 1.107 1.162  10.06  1.87e-23
```

*Note: Pseudo-R2 for quasibinomial/quasipoisson families is calculated by refitting the fitted and null models as binomial/poisson.

Poisson Model and Risk Differences

–our example- **Wells in Bangladesh**

```
mod.glm.p.rd <- glm(switch ~ assoc + arsenic, data = wells,
                    family=poisson(link="identity"))
```

```
#>
#> Call:
#> glm(formula = switch ~ assoc + arsenic, family = poisson(link = "identity"),
#>     data = wells)
#>
#> Coefficients:
#>             Estimate Std. Error z value Pr(>|z|)
#> (Intercept)   0.4388     0.0278   15.80  < 2e-16 ***
#> assoc        -0.0268     0.0275   -0.97     0.33
#> arsenic       0.0891     0.0138    6.47  9.8e-11 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> (Dispersion parameter for poisson family taken to be 1)
#>
#>     Null deviance: 1921.5  on 3019  degrees of freedom
#> Residual deviance: 1876.2  on 3017  degrees of freedom
#> AIC: 5356
#>
#> Number of Fisher Scoring iterations: 5
```

Poisson Model and Risk Differences

```r
jtools::summ(mod.glm.p.rd, confint=T)$"coeftable"
```

```
#>                 Est.    2.5%  97.5% z val.        p
#> (Intercept)  0.4388  0.3844 0.4932 15.800 3.09e-56
#> assoc       -0.0268 -0.0806 0.0271 -0.974 3.30e-01
#> arsenic      0.0891  0.0621 0.1161  6.470 9.83e-11
```

Quasi-Poisson Risk Difference (Robust Variance)

```r
mod.glm.qp.rd1 <- glm(switch ~ assoc + arsenic, data = wells,
                      family= quasipoisson(link="identity")) #
jtools::summ(mod.glm.qp.rd1, confint=T)$"coeftable"
```

```
#>                 Est.    2.5%   97.5% t val.         p
#> (Intercept)  0.4388  0.4033 0.47429  24.23 1.19e-118
#> assoc       -0.0268 -0.0619 0.00837  -1.49  1.36e-01
#> arsenic      0.0891  0.0715 0.10674   9.92  7.54e-23
```

Why and When do we use Poisson for Risk Ratios and Risk Differences?

Poisson regression

- Used for rare outcomes
- When the follow-up has different lengths

Logistic regression: Used with same length of follow-up time

Poisson approximation to Binomial

- Cohort studies where all patients have equal follow-up times.
- Poisson regression can be used similarly as logistic regression, with a time-at-risk value specified as 1 for each subject.
- If the model adequately fits the data, this provides a correct estimate of the AdjRR

Chen et al.: Comparison of robustness to outliers between robust poisson models and log-binomial models when estimating relative risks for common binary outcomes: a simulation study. BMC Medical Research Methodology 2014 14:82.

Comparing Results

Model "switching" to a safer well (outcome) as a function of whether the owner belongs to a community association and As. level.

Model	**Frequentist Estimates**
Linear Model ^{1}	RD= -0.036; RR= N/A
Logit Model (Predicted probs.) ^{2}	RD= -0.033; RR= 0.939
AME (prediction `StdGLM`) ^{2}	RD= -0.031; RR= 0.948
GLM: Log-Binomial ^{2}	RD= N/A; RR= 0.928
Poisson Model ^{2}	RD = -0.027; RR= 0.946

^{1} Single predictor; ^{2} Two predictors

Why and When do we use Poisson for Risk Ratios and Risk Differences?

- For common outcomes, Poisson regression is likely to compute a confidence interval(s) that is conservative, suggesting less precision than is true.

- Poisson regression produces wider confidence intervals (compared with a log-binomial model and stratified analysis) because Poisson errors are overestimates of binomial errors when the outcome is common

- Poisson errors approximately equal binomial errors when the outcome is rare

Guangyong Zou. A Modified Poisson Regression Approach to Prospective Studies with Binary Data, AJE, Vol 159: 7 (2004).702–706 (https://doi.org/10.1093/aje/kwh090).

Chen et al. Comparison of robustness to outliers between robust poisson models and log-binomial models when estimating relative risks for common binary outcomes: a simulation study. BMC Med Res Method 2014 14:82.

Robust Poisson models

- The Poisson regression uses a logarithm as the natural link function under the generalized linear model framework.
- The robust Poisson regression model uses the classical sandwich estimator under the generalized estimation equation (GEE) framework to correct the inflated variance (also known as over-dispersion) in the standard Poisson regression.
- The technique is known as modified Poisson regression or pseudo-likelihood estimation.
- This correction can be achieved by using the `robust` option form `jtools` in `R` or the `quasipoisson` from `glm`, also using the REPEATED statement in SAS Proc GENMOD; the ROBUST option in STATA's Poisson procedure.

Robust Poisson models vs. Log-Binomial

- In presence of linear confounders, the two models yield comparable relative biases.
- With the non-linear confounders, the robust Poisson model outperform the log-binomial model. Larger differences with rare outcomes.
- The robust Poisson models are more robust to outliers compared to the log-binomial models when estimating relative risks or risk ratios for common binary outcomes.

Chen et al. Comparison of robustness to outliers between robust poisson models and log-binomial models when estimating relative risks for common binary outcomes: a simulation study. BMC Medical Research Methodology 2014 14:82.

Log-Binomial vs Robust Poisson?

- Of the two methods, the log-binomial method is generally preferred due to the fact that the MLEs estimated by the log-binomial models are more efficient compared to the pseudo-likelihood estimators used by the robust Poisson models.
- Spiegelman and Hertzmark recommend using the log-binomial models over the robust Poisson models when convergence is not an issue.
- It appears that the gain in efficiency is beneficial to log-binomial models only for samples of small sizes.
- Due to the concern of lack of efficiency for the robust Poisson models for small samples, log-binomial may still be the choice when the sample size is small.

Other Count Models

- Negative Binomial (Overdispersed Poisson Models)
- Zero Inflated Models
- Zero Truncated Models

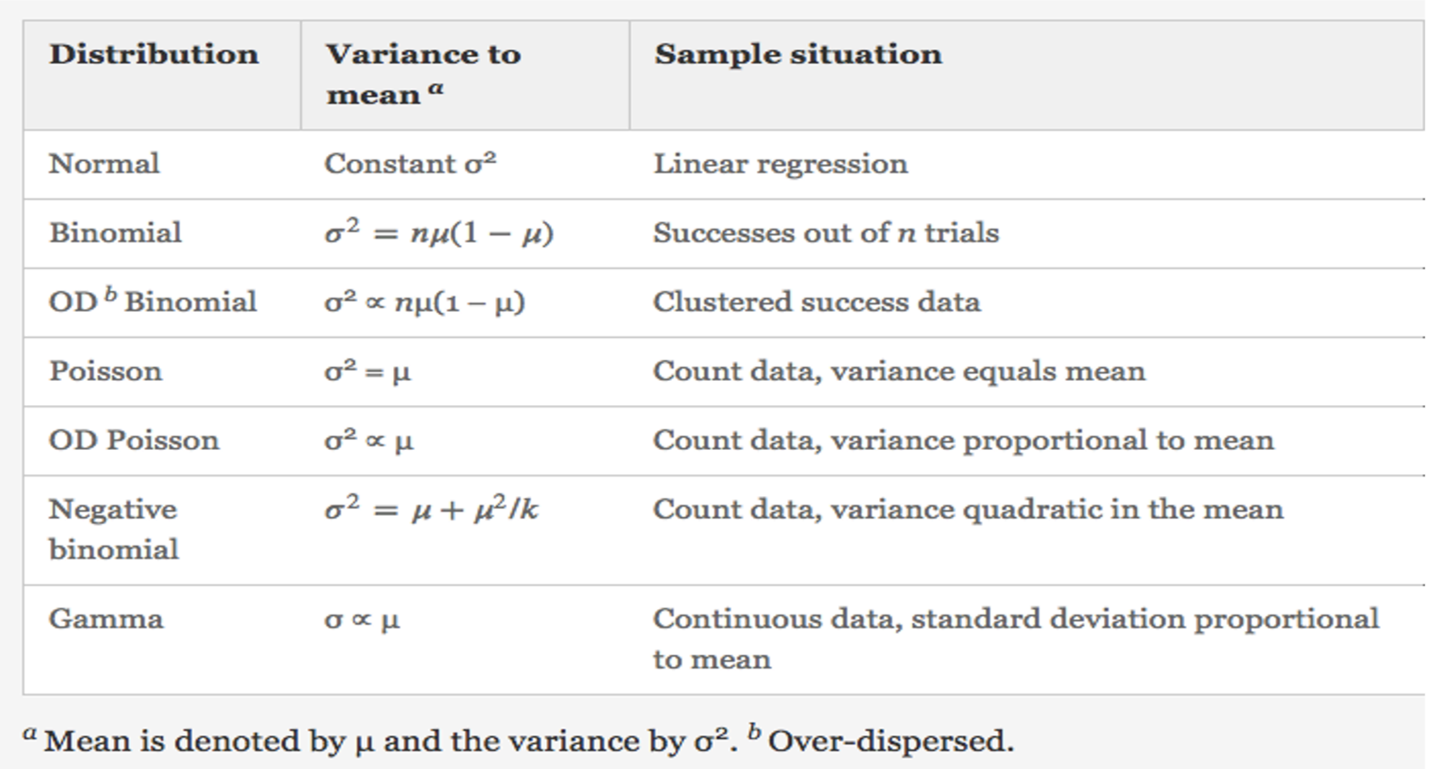
... Out of the scope for this session

QUESTIONS?

COMMENTS?

RECOMMENDATIONS?

Distributions



Distribution	**Variance to mean** [a]	**Sample situation**
Normal	Constant σ^2	Linear regression
Binomial	$\sigma^2 = n\mu(1-\mu)$	Successes out of n trials
OD [b] Binomial	$\sigma^2 \propto n\mu(1-\mu)$	Clustered success data
Poisson	$\sigma^2 = \mu$	Count data, variance equals mean
OD Poisson	$\sigma^2 \propto \mu$	Count data, variance proportional to mean
Negative binomial	$\sigma^2 = \mu + \mu^2/k$	Count data, variance quadratic in the mean
Gamma	$\sigma \propto \mu$	Continuous data, standard deviation proportional to mean

[a] Mean is denoted by μ and the variance by σ^2. [b] Over-dispersed.

Vittinghoff E., Glidden D.V., Shiboski S.C., McCulloch C.E. (2012) Generalized Linear Models. In: Regression Methods in Biostatistics. Statistics for Biology and Health. Springer, Boston, MA

Log-Binomial vs Robust Poisson?

- Software utilizes iterative weighted least squares (IWLS*) approach or variations of IWLS to find MLEs for generalized linear models.
- For log-binomial models, the weights used by the IWLS approach contain the term 1/(1-p), where $p = exp(XT\beta)$ with a range from 0 to 1.
- The MLE of a log-binomial model is likely to be too sensitive to outliers because a very large p has a large influence on the weights.
- MLE and pseudolikelihood estimators are deteriorated in presence of outliers.
- The level of deterioration differed when the relationships between the confounder and the outcome was not in a simple form

Maximum likelihood estimate of lambda

$$MLE = \ell(\lambda; y_1, \dots, y_n) = \sum_{i=1}^{n} \log f(y_i; \lambda) = \sum_{i=1}^{n} \log \frac{\lambda^{y_1} exp^{-\lambda}}{y_1!}$$

Take first derivative (notation $' = \frac{d(.)}{d\lambda}$)

$$\ell'(\lambda; y_1, \dots, y_n) = \sum_{i=1}^{n} \log \frac{\lambda^{y_1} exp^{-\lambda}}{y_1!}' = \sum_{i=1}^{n} \{-\lambda + y_i \cdot \log(\lambda) - \log(y_i!)\}'$$

$$= \sum_{i=1}^{n} \left\{-1 + y_i \cdot \frac{1}{\lambda}\right\} = -n + \frac{1}{\lambda} \sum_{i=1}^{n} y_i$$

Maximum likelihood estimate of lambda

The first-order condition for maximizing the log-likelihood sets its derivative to zero

$$\ell'(\lambda; y_1, \ldots, y_n) = 0$$

$$-n + \frac{1}{\lambda} \sum_{i=1}^{n} y_i = 0$$

$$n \cdot \lambda = \sum_{i=1}^{n} y_i$$

$$\lambda = \frac{1}{n} \sum_{i=1}^{n} y_i = \bar{y}$$

Thus, the maximum likelihood estimator is simply the empirical mean $\hat{\lambda} = \bar{y}$

Poisson distribution (PD)

If instead are given the average rate, *r*, at which events occur then $\lambda = rt$, and

$$P(k\, events\, in\, interval\, t) = \frac{rt^k e^{-rt}}{k!}$$

Poisson distribution (PD)

PD a special case of the binomial, with trials n -> ∞ and P(success in any trial) -> 0
If p is small, binomial P(k successes) $\approx$ poisson P(k with λ = np)

Example: Consider binomial distribution with probability = 0.01. What is the probability of 8 occurences in 1000 tries?

```
# binomial
dbinom(8,1000,0.01)
```

```
#> [1] 0.113
```

```
# poisson
dpois(8,1000*.01) # lambda = n*p
```

```
#> [1] 0.113
```

Offset vs Exposure

In most applications of count-data regression, there is a baseline or exposure, some value such as the number of person years that the counts occurred over

We can model y_i as the number of cases in a process with rate θ_i and exposure μ_i

$$y_i \sim negative\,binomial(\mu_i \theta_i, \phi)$$

where, as before, $\theta_i = e^{X_i \beta}$ and includes Poisson regression as the special case of ϕ → ∞

The logarithm of the exposure, $log(\mu_i)$, is called the offset in GLM terminology.

The regression coefficients β now reflect the associations between the predictors and θ_i (for example, the rate of deaths)

Putting the logarithm of the exposure into the model as an offset, is equivalent to including it as a regression predictor, but with its coefficient fixed to the value 1. Another option is to include it as a predictor and let its coefficient be estimated from the data.

AIC = Akaike information criterion

$AIC = -2(log - likelihood) + kn_{par}$, where $k = 2$, equivalently

$AIC = 2k - 2\ln(\hat{L})$

k = number of estimated parameters in the model $\hat{L}$ = maximum value of the likelihood function for the model

Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models.

Hence, AIC provides a means for model selection

BIC = Bayesian Information Criterion

$BIC = -2(log - likelihood) + kn_{par},$

where $k = log(n)$,

and n is the number of observations