

Multivariate: Poisson regression

Table of contents

1	Goals	1
1.1	Goals	1
2	Counts and frequencies	2
2.1	Count outcomes	2
3	Poisson regression	6
3.1	Model	6
3.2	Interpretation	11
4	Overdispersion	13
4.1	Overdispersion	13
4.2	Overdispersed Poisson regression	15
4.3	Negative binomial regression	16
5	Miscellaneous	23
5.1	Choosing a model	23
5.2	Too many or too few zeroes	24
5.3	Evaluation and comparison	26
5.4	Nested models	27
5.5	Conclusion	28

1 Goals

1.1 Goals

1.1.1 Goals of this lecture

- My outcome variable **isn't normally distributed**

- It's a **count** of how often something happened!!!
- Probably **right skewed** (among other things)
- May also have **a lot of zeroes**
- Use **Poisson regression** (or a related model) to analyze the outcome
 - It's an **extension** of linear regression
 - Many of the **same concepts** still apply

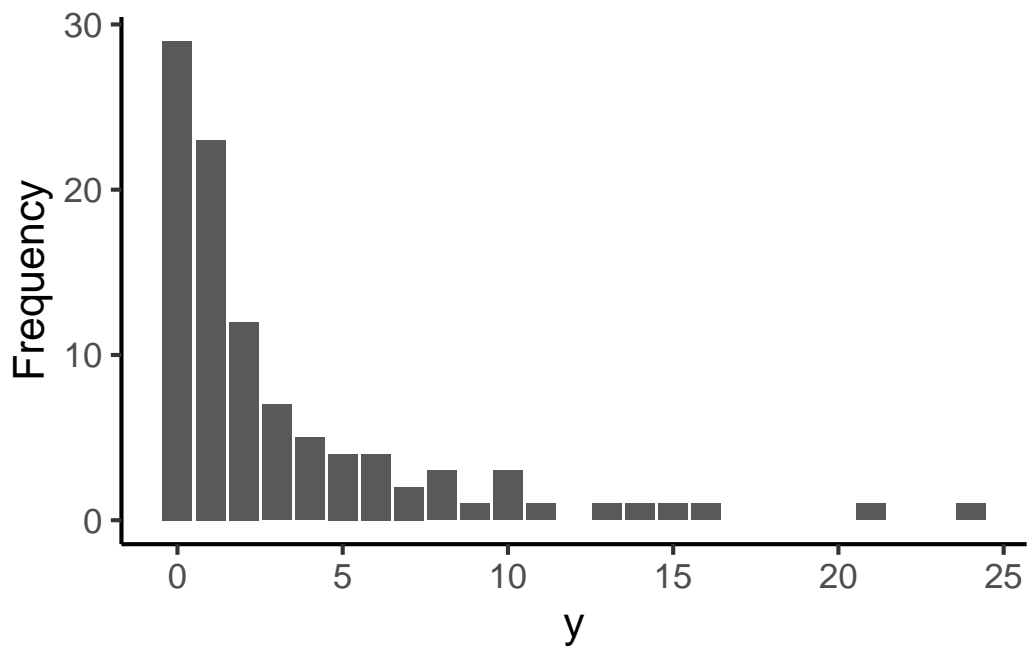
2 Counts and frequencies

2.1 Count outcomes

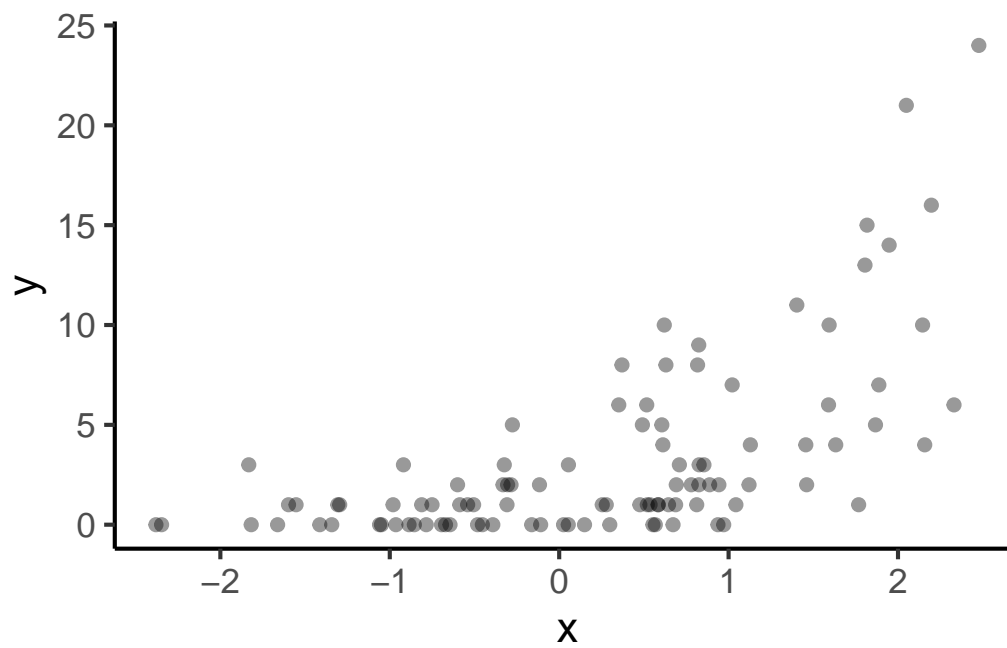
2.1.1 Count outcomes

- Easy to miss that linear regression assumptions are violated
- Count variables are
 - **Discrete**: only take on whole number (integer) values
 - **Lower bound of 0**
 - Typically **right skewed**

2.1.2 Figure: histogram of count



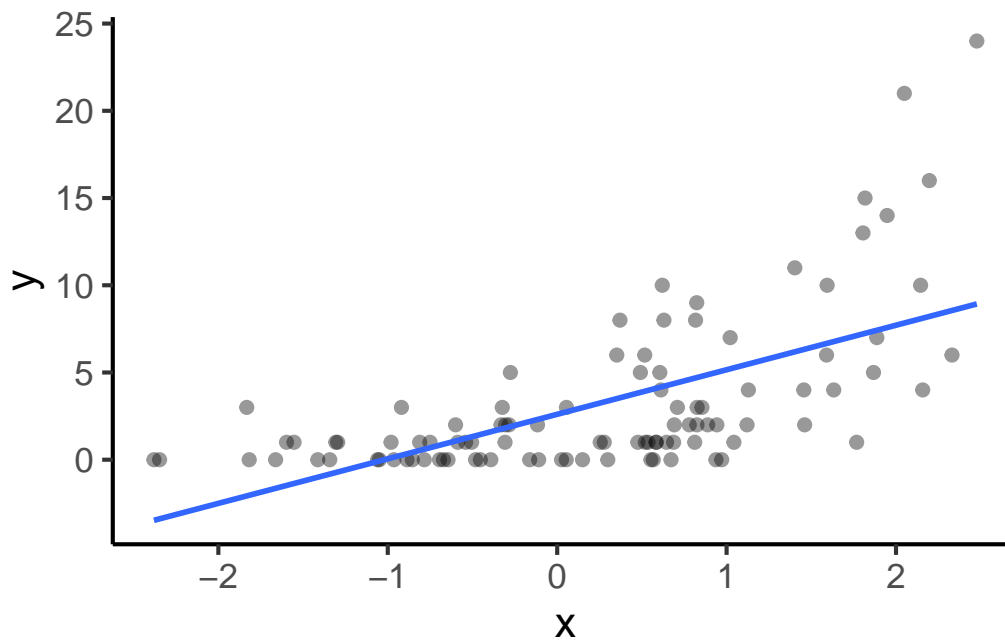
2.1.3 Figure: count as outcome



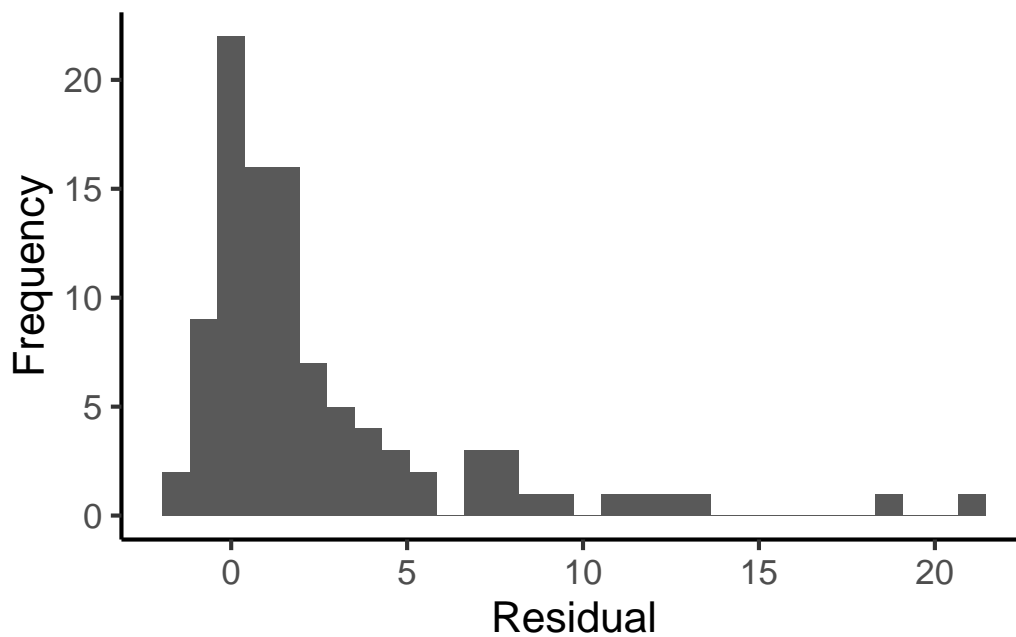
2.1.4 GLM with count outcomes

- Count outcomes violate assumptions of GLM (regression and ANOVA)
 - Residuals are not conditionally normally distributed
 - Residuals do not have a constant variance (heteroscedasticity)
- GLM also results in out-of-bound (i.e., less than 0) predicted values

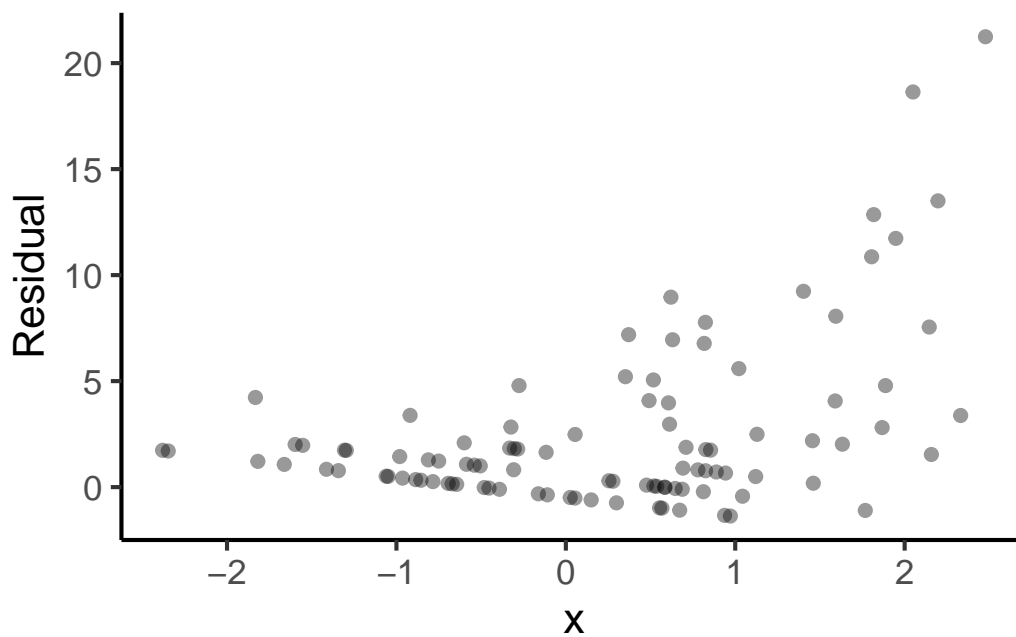
2.1.5 Figure: using GLM, out-of-bounds prediction



2.1.6 Figure: using GLM, histogram of count residuals



2.1.7 Figure: Using GLM, X versus count residuals



3 Poisson regression

3.1 Model

3.1.1 Poisson regression

- Outcome: count in a *fixed period of time*
 - Integer (whole number) values greater than or equal to 0
- Residuals: Poisson distribution
- Link function: natural log (\ln)

3.1.2 Poisson distribution

- Named after Simeon Denis Poisson (1781 - 1840)
- I had to delete the pictures because they were too large
 - Jim Morrison
 - Jim Morrison's grave
 - Pere Lachaise cemetery sign
 - Pere Lachaise cemetery

3.1.3 Poisson distribution

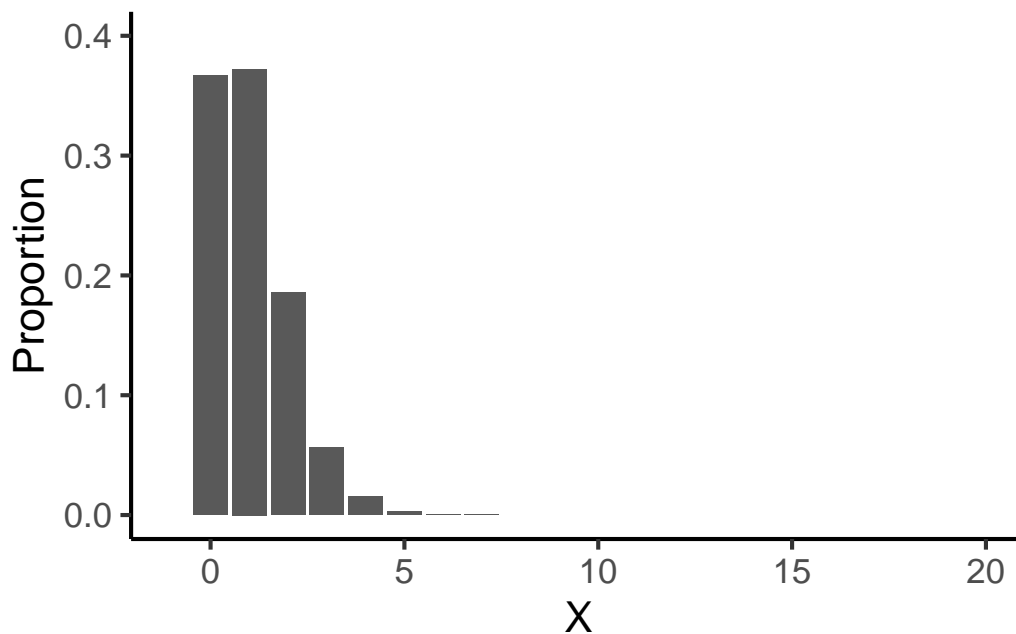
- **Number of times a low probability** event happens
 - Number of soldiers killed by horse-kicks each year in each corps in the Prussian cavalry (Bortkiewicz)
 - Number of yeast cells used when brewing Guinness beer (Gosset, "Student" of "Student's t ")
 - Number of typographical errors on each page of a book
 - Number of bombs falling in London in each square block
 - Number of earthquakes per day in Southern California

3.1.4 Poisson distribution

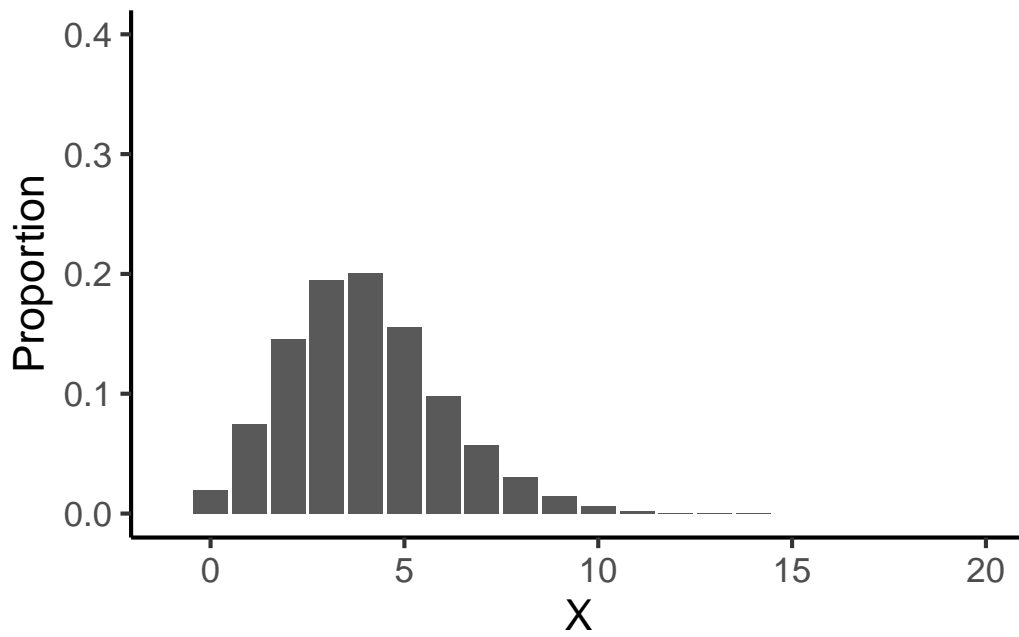
$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

- Poisson distribution properties:
 - **Mean** of a Poisson distribution = λ
 - **Variance** of a Poisson distribution = λ
 - **Discrete** distribution, so only defined for *integers*
 - Undefined below **zero**

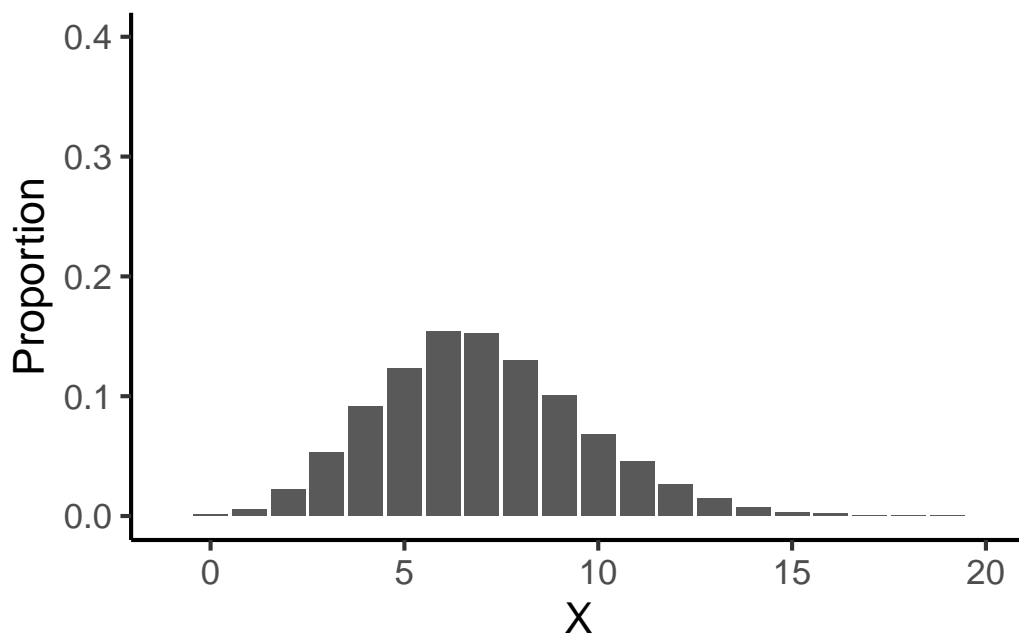
3.1.5 Figure: mean = variance = 1



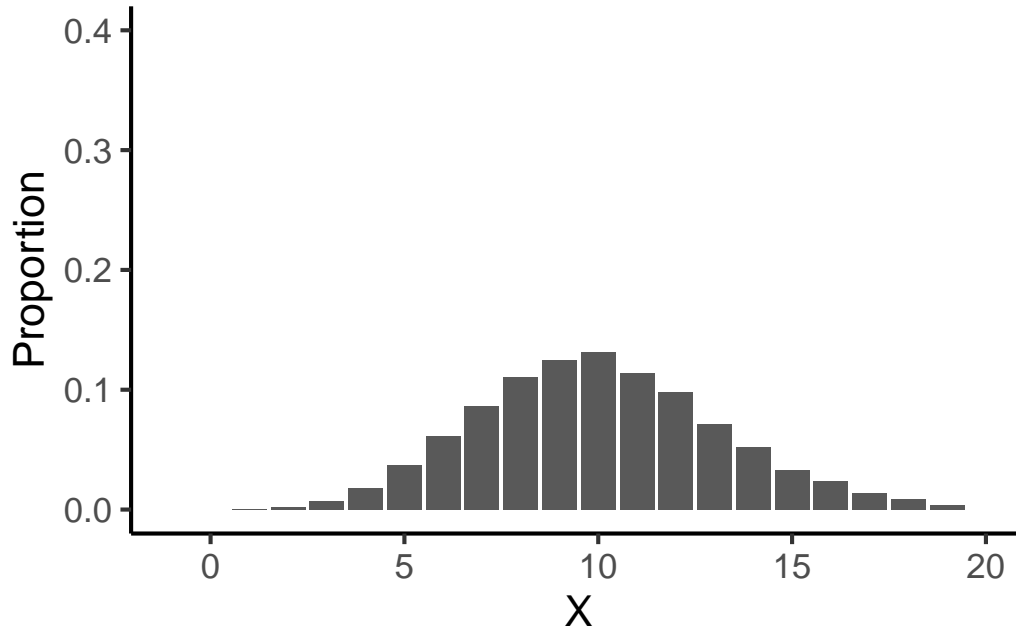
3.1.6 Figure: mean = variance = 4



3.1.7 Figure: mean = variance = 7



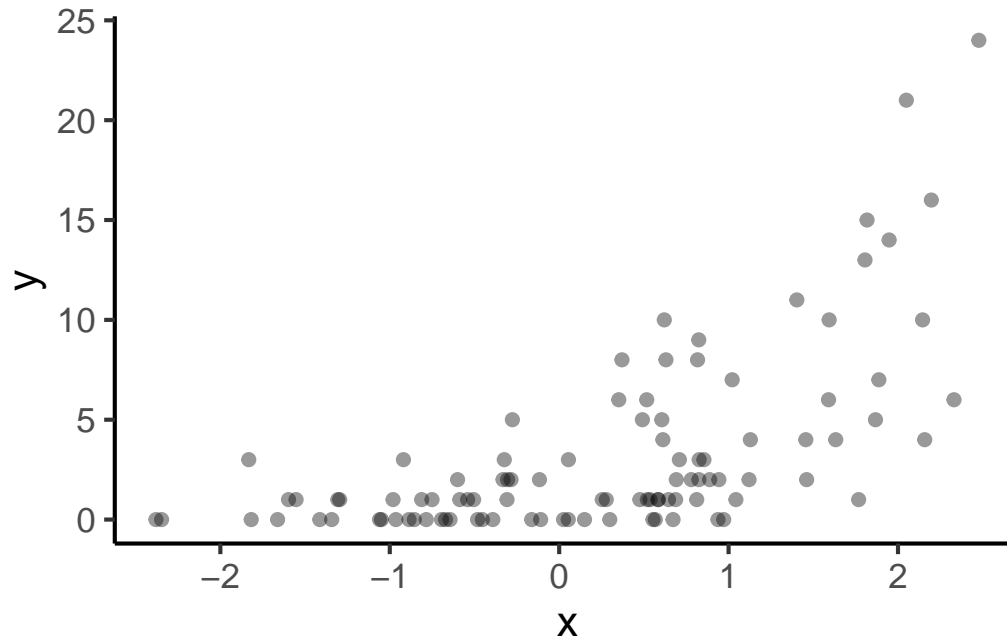
3.1.8 Figure: mean = variance = 10



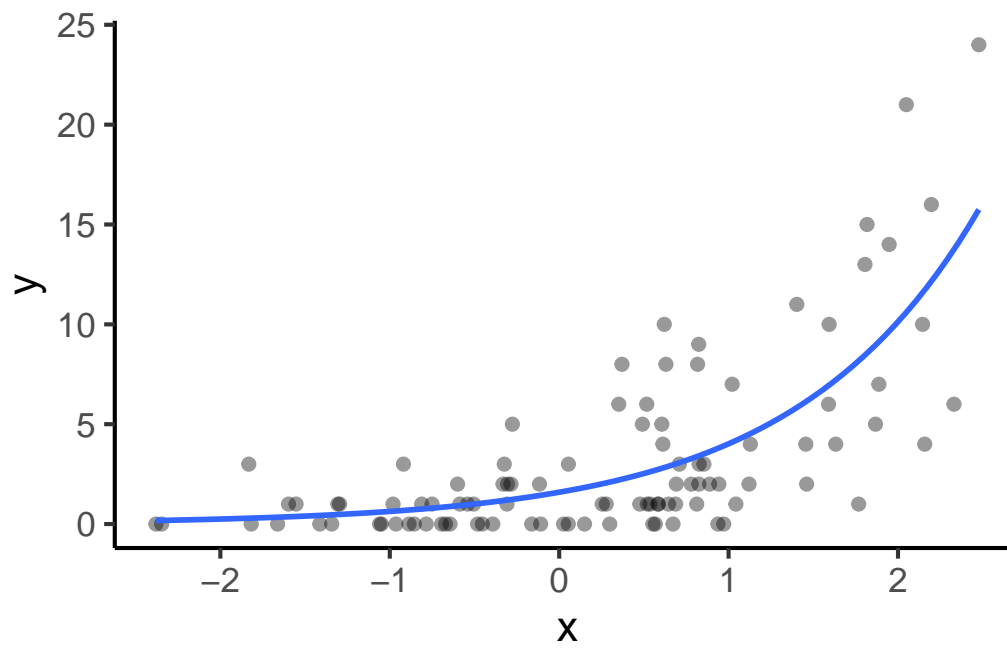
3.1.9 Poisson regression

- Outcome: count in a *fixed period of time*
 - Integer (whole number) values greater than or equal to 0
- Residuals: Poisson distribution
 - Heteroscedasticity is a **feature** of Poisson distribution
 - Mean and variance are related
- Link function: natural log (\ln)
 - \ln is only defined for positive values, so no predicted values < 0

3.1.10 Figure: Poisson regression graphically



3.1.11 Figure: Poisson regression graphically



3.2 Interpretation

3.2.1 Two forms of Poisson regression

- In terms of the **predicted count**, $\hat{\mu}$:

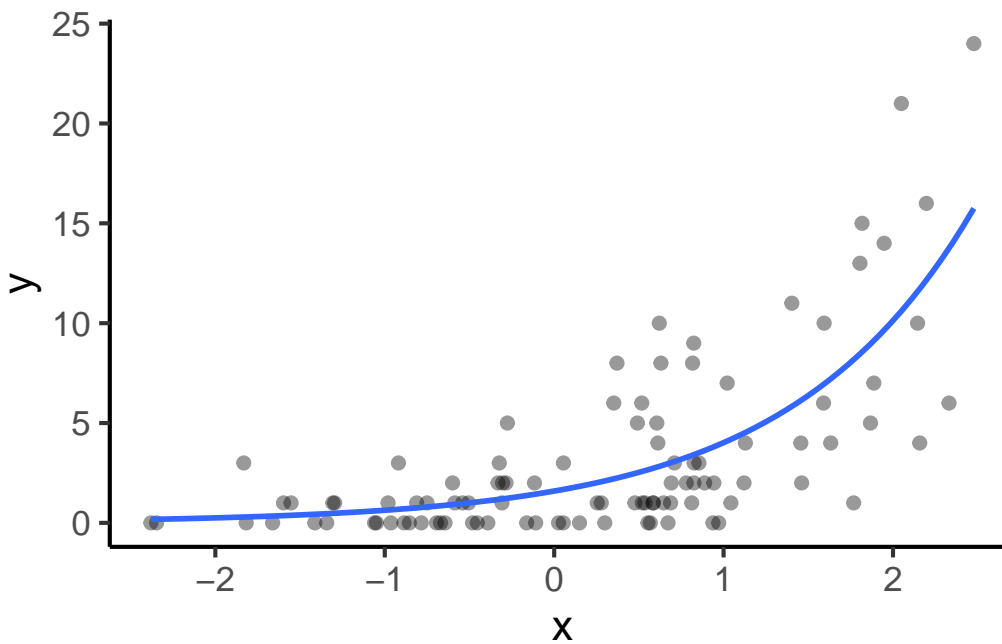
$$\hat{Y} = \hat{\mu} = e^{b_0 + b_1 X_1 + b_2 X_2 + \dots + b_p X_p}$$

- In terms of the **natural log of the predicted count**, $\ln(\hat{\mu})$:

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

- Poisson regression is sometimes called the **log-linear model** because the log (i.e., \ln) form is a straight line

3.2.2 Count metric interpretation: $\hat{\mu} = e^{0.466 + 0.925X}$



3.2.3 Count metric interpretation

- Intercept:
 - e^{b_0} is the predicted count when $X = 0$
- Slope:
 - e^{b_1} is the **multiplicative** effect of X on the predicted count
 - * For a 1 unit increase in X , the predicted count is *multiplied* by e^{b_1}
 - * This is sometimes called the *rate ratio* (RR)

3.2.4 Count metric interpretation

- $e^{b_0} = e^{0.4659581} = 1.594$
 - Predicted count when $X = 0$
- $e^{b_1} = e^{0.9247665} = 2.521$
 - *Multiplicative* effect of X on the predicted count
 - For a 1 unit increase in X , the predicted count is **multiplied** by 2.521

3.2.5 Count metric interpretation

X	Predicted count
-2	0.251
-1	0.632
0	1.594
1	4.018
2	10.130

3.2.6 Log of predicted count metric interpretation $\ln(\hat{\mu}) = 0.466 + 0.925X$

- Intercept:
 - b_0 is the $\ln(\hat{\mu})$ when $X = 0$
- Slope:
 - b_1 is the **additive** effect of X on the $\ln(\hat{\mu})$

3.2.7 Log of predicted count metric interpretation

- $b_0 = 0.466$
 - $\ln(\hat{\mu})$ when $X = 0$
- $b_1 = 0.9248$
 - *Additive* effect of x on the $\ln(\hat{\mu})$
 - For a 1 unit increase in X , the $\ln(\hat{\mu})$ increases by 0.9248

3.2.8 Log of predicted count metric interpretation

X	Predicted count	$\ln(\text{Predicted count})$
-2	0.251	-1.384
-1	0.632	-0.459
0	1.594	0.466
1	4.018	1.391
2	10.130	2.315

3.2.9 Confidence intervals

- Results from software are in $\ln(\hat{\mu})$ metric (linear metric)

	b	Lower CL	Upper CL
(Intercept)	0.466	0.277	0.643
x	0.925	0.806	1.047

- To get **predicted count metric**, exponentiate **each value** (e^b)

	exp.b.	Lower.CL	Upper.CL
(Intercept)	1.594	1.320	1.902
x	2.521	2.239	2.848

4 Overdispersion

4.1 Overdispersion

4.1.1 Equidispersion

- Remember the Poisson distribution:

- Mean = Variance = λ
- This is called **equidispersion**

4.1.2 Overdispersion

- In Poisson regression, **equidispersion** means that the (conditional) mean and (conditional) variance of Y are equal
 - In practice, the actual variance of Y is typically larger than expected by the Poisson distribution
 - This is called **overdispersion**
 - (Can technically be smaller, uncommon, *underdispersion*)

4.1.3 Overdispersion

- Why does overdispersion happen?
 - Sometimes, just because
 - Due to omitted predictors
 - “Excess” zeroes

4.1.4 Overdispersion

- Why do we care about overdispersion?
 - The variance estimated in the model (λ) helps determine the **standard errors** for regression coefficients
 - It also plays a role in calculating **deviance**
 - **If the variance in our data is larger than λ , we are using the wrong standard errors and the deviance is wrong**
 - * Test of significance, pseudo- R^2 , model comparisons are wrong

4.1.5 What do we do?

- There are two main alternative models for overdispersion
 - Overdispersed Poisson regression
 - Negative binomial regression
- Both handle overdispersion, but in slightly different ways

4.2 Overdispersed Poisson regression

4.2.1 Overdispersed (or quasi-) Poisson regression

- The simplest way to deal with overdispersion is to include a “fudge factor” to account for overdispersion
 - The variance is larger than the model says, so we just multiply it by something to make it larger
- The fudge factor is provided by the scale parameter (ψ)
 - If equidispersion holds, $\psi = 1$
 - If there is overdispersion, $\psi > 1$
 - (ψ will show up in your output, no need to calculate)

4.2.2 Overdispersed (or quasi-) Poisson regression

- Compared to Poisson regression:
 - Each standard error in the OD Poisson model is multiplied by $\sqrt{\psi}$
 - If $\psi > 1$, this will make the standard errors **larger**
- Software
 - SPSS: Scale value in output is ψ
 - R: Scale value in output is ψ
 - SAS: Scale value in output is $\sqrt{\psi}$

4.2.3 Estimating the scale parameter

- The scale parameter can be estimated in two similar ways
 - Pearson χ^2 : $\psi = \frac{\chi^2_{Pearson}}{df}$
 - Deviance: $\psi = \frac{Deviance}{df}$
- The results are typically very close, but **Pearson** is preferred

4.2.4 Overdispersed Poisson regression interpretation

- Interpretation for the overdispersed Poisson regression model is **identical** to interpretation for the Poisson regression model
 - Interpret coefficients the same
 - * Predicted count metric or log of predicted count metric
- Regression coefficients **same** as coefficients from Poisson regression
 - Only standard errors (and therefore potentially significance) will change

4.2.5 Comparison

- Poisson regression:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.466	0.093	5.000	0
x	0.925	0.061	15.063	0

- Overdispersed Poisson regression (from R, $\psi = 2.45$):

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.466	0.146	3.196	0.002
x	0.925	0.096	9.628	0.000

4.3 Negative binomial regression

4.3.1 Negative binomial distribution

- Related to Poisson distribution
 - But that's really **not obvious**
 - **Mixture** of *Poisson* and *gamma* distributions
 - Addition of gamma adds in **extra variance** to make **variance** > **mean**

$$P(X = k) = \binom{k-r-1}{k} (1-p)^r p^k$$

4.3.2 Negative binomial in words 1

Poisson regression assumes that **every person** with the **same predictor value(s)** comes from the **same Poisson distribution**

- **Person A with $X = 1$:** Y from Poisson distribution w mean = $e^{b_0+b_1(1)}$
- **Person B with $X = 1$:** Y from Poisson distribution w mean = $e^{b_0+b_1(1)}$
- **Person C with $X = 1$:** Y from Poisson distribution w mean = $e^{b_0+b_1(1)}$

But what if there's actually **more variability** among subjects than captured by the predictors? (i.e., overdispersion)

- For example, what if we have omitted a predictor?

4.3.3 Negative binomial in words 2

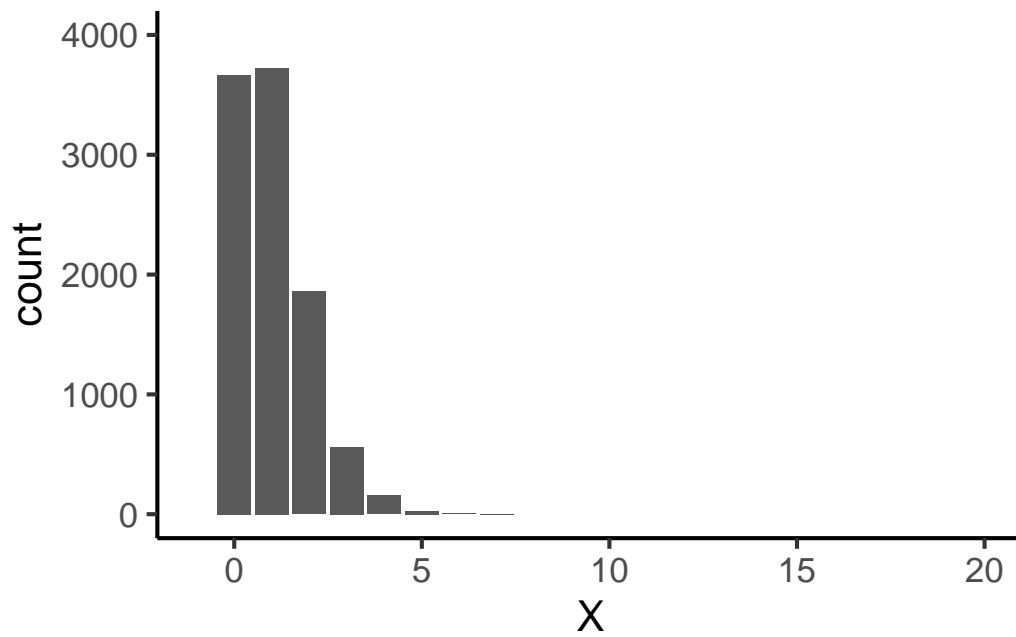
Negative binomial regression allows people with the **same predictor value(s)** to come from **different Poisson distributions**

- **Person A with $X = 1$:** Y from **one Poisson distribution**
- **Person B with $X = 1$:** Y from **another** Poisson distribution
- **Person C with $X = 1$:** Y from **yet another** Poisson distribution

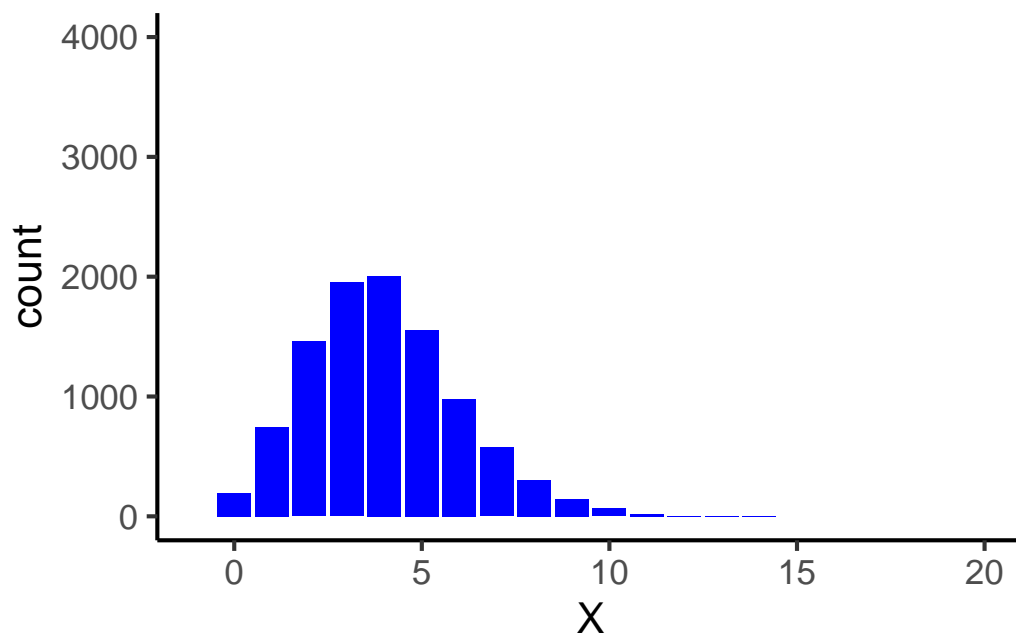
The means of the Poisson distributions follow a **gamma distribution**

- The *average* of those distributions is the same as Poisson regression
- But **variability** of several **different** Poisson distributions > variability of **single** Poisson distribution

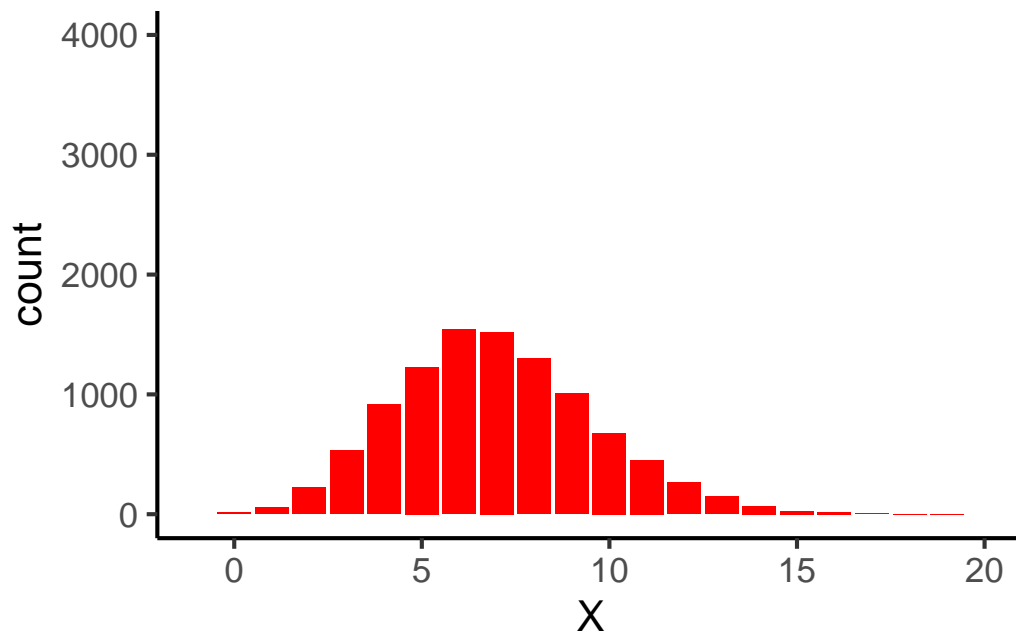
4.3.4 Person A: Y from Poisson distribution with $\lambda = 1$



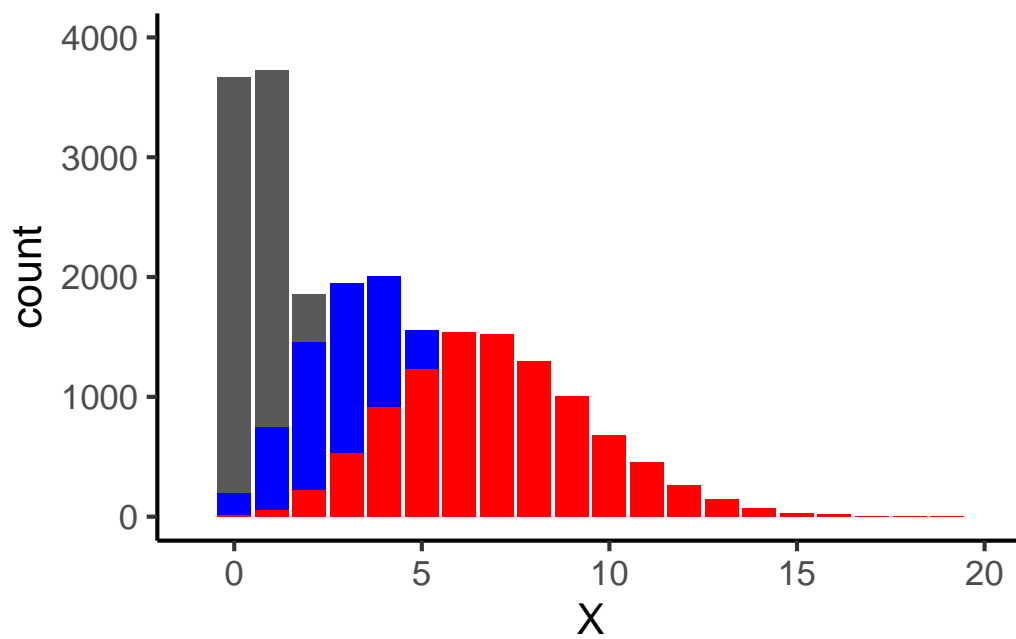
4.3.5 Person B: Y from Poisson distribution with $\lambda = 4$



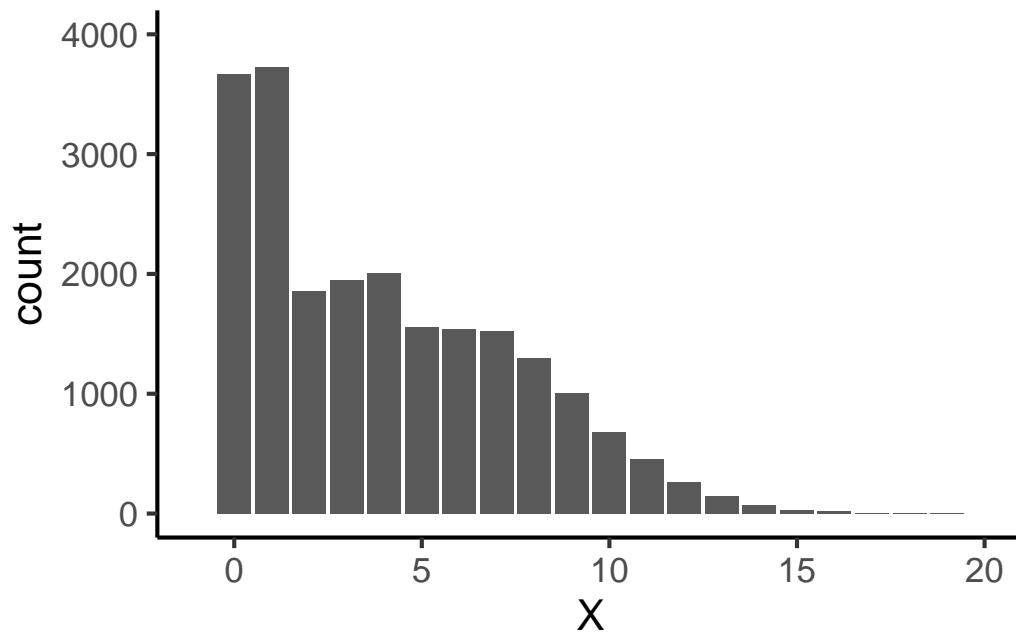
4.3.6 Person C: Y from Poisson distribution with $\lambda = 7$



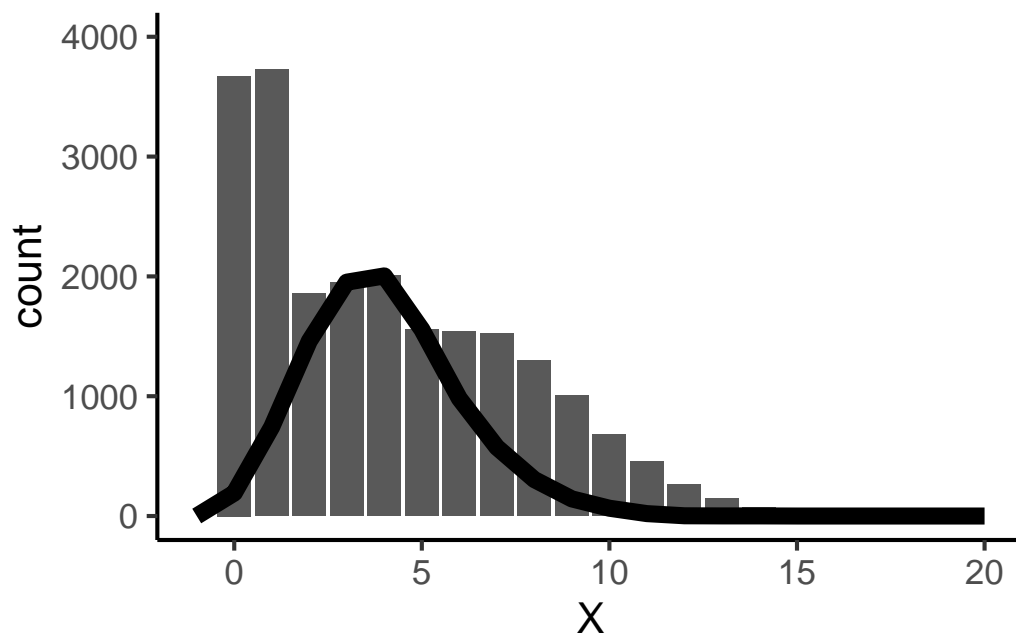
4.3.7 Negative binomial distribution



4.3.8 Negative binomial distribution



4.3.9 Negative binomial distribution



4.3.10 Negative binomial scale parameter

- Negative binomial regression estimates a different scale parameter
 - Usually called α (SPSS, SAS, STATA)
 - Sometimes called θ which is $1/\alpha$ (R)

4.3.11 Negative binomial scale parameter

- Conditional residual distribution based on negative binomial
 - Mean = μ
 - Variance = $\mu + \alpha\mu^2$
 - * If there is overdispersion, $\alpha > 0$
 - * If there is not, $\alpha = 0$ and this reduces to Poisson
- Standard errors don't need to be adjusted
 - They are calculated using variance = $\mu + \alpha\mu^2$ to begin with

4.3.12 Negative binomial regression interpretation

- Interpretation for the negative binomial regression model is **identical** to interpretation for the Poisson regression model
 - Interpret coefficients the same
 - * Predicted count metric or log of predicted count metric
- Regression coefficients for negative binomial will be **different**
 - Now that the model isn't constrained to “mean = variance” anymore, the **mean may shift**
 - * Regression coefficients change
 - * Standard errors will also change

4.3.13 Comparison

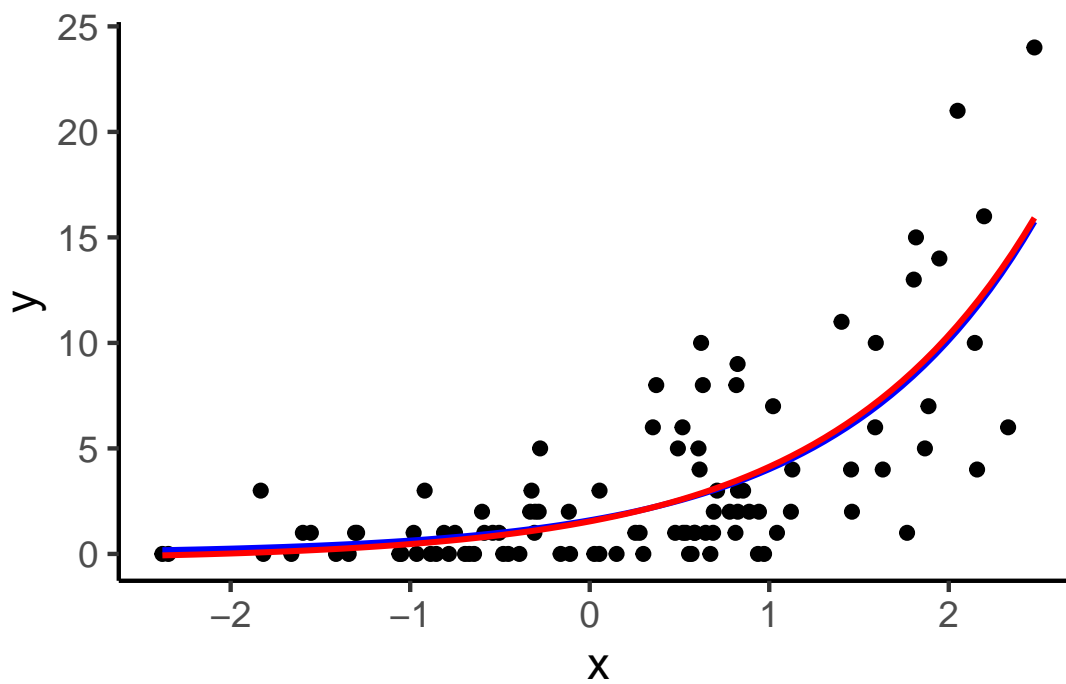
Poisson regression:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.466	0.093	5.000	0
x	0.925	0.061	15.063	0

Negative binomial regression ($\alpha = 0.49$):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.501	0.122	4.123	0
x	0.880	0.101	8.719	0

4.3.14 Figure: comparison (blue = Poisson, red = NB)



5 Miscellaneous

5.1 Choosing a model

5.1.1 Which model to pick?

- **Poisson regression**
 - You will likely never satisfy the equidispersion assumption
- **Overdispersed Poisson regression**
 - Preferred to Poisson regression
 - Appropriate standard errors and relatively simple model
- **Negative binomial regression**
 - More complex, but better able to reproduce the observed counts
 - This may be preferred if you're interested in *prediction*

5.1.2 A word of caution...

Berk & MacDonald (2008)

If apparent overdispersion results from specification errors in the systematic part of the Poisson regression model, resorting to the negative binomial distribution does not help. It can make things worse by giving a false sense of security when the fundamental errors in the model remain.

- In other words, negative binomial regression doesn't *fix a bad model*

5.1.3 Variable lengths of time

- Poisson regression and variants assume a *fixed length of time*
 - number of aggressive acts committed by a child in 1 hour
 - number of cigarettes smoked per day
- Often, we measure a count over some *variable period of time*

5.1.4 Variable lengths of time

- Can extend Poisson type models to variable time periods
 - Include natural log of the measurement interval as a predictor with regression coefficient equal to 1:
 - * $\ln(\hat{\mu}) = \ln(\text{time}) + b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p$
 - In software: include time (specifically, $\ln(\text{time})$) as an “offset”

5.2 Too many or too few zeroes

5.2.1 Zeroes can be really important

- Conceptually, zeroes are meaningful
 - Lowest possible value of a count
 - Indicate “nothing”
 - Two situations:
 - * Too few zeroes (relative to a Poisson distribution)
 - * Too many zeroes (relative to a Poisson distribution)

5.2.2 Too few zeroes

- Sometimes you have a study where the outcome is a count, but it **cannot take on a value of 0**
 - Study of medical visits
 - * **Have to visit the doctor** to get involved in the study
 - Study of substance use
 - * Only interested in **substance users**
- Truncated Poisson regression model uses a *truncated Poisson distribution* that removes the probability of zeroes

5.2.3 Excess zeroes

- Counts often display “**excess**” **zeroes**
 - More values of 0 than expected for a Poisson distribution
- Even if the rest of the distribution is approximately Poisson, the “excess” zeroes lead to *overdispersion*
 - Sometimes, what looks like overdispersion is really excess zeroes
- There are specific Poisson family models to appropriately deal with excess zeroes, depending on *why* the zeroes are there

5.2.4 Why are there all these zeroes?

- This is a substantive question
 - You need to know about the outcome you’re studying
- Question: Do the **people who are responding zero** have **some probability** of responding otherwise?

5.2.5 Do they have some probability of responding other than 0?

- Yes → **zero-inflated Poisson regression**
 - Cigarettes smoked for smoker who hasn’t smoked yet today
 - Alcohol consumed for someone who hasn’t had any yet today
- No → **hurdle regression** or **with-zeroes regression**
 - Cigarettes smoked: nonsmoker always responds 0
 - Alcoholic drinks consumed: abstainer always responds 0
 - (These are called “structural zeroes”)

5.2.6 Zero-inflated Poisson regression

- There are zeroes have **some probability** to be non-zero
- Two parts modeled simultaneously:
 - Logistic regression: *is each zero structural or not?*
 - Poisson regression: non-structural zeroes and positive values
- Can use same set of predictors in both parts, but do not have to
- Can also **model non-structural zeroes and positive values** using *overdispersed* Poisson regression or *negative binomial* regression

5.2.7 Hurdle regression or with-zeroes regression

- There are zeroes that have **no probability** of being non-zero
 - Two different populations: smokers and nonsmokers, drinkers and nondrinkers
- Two parts modeled simultaneously:
 - Logistic regression: *whether zero or not*
 - Truncated Poisson regression: positive values only
- Can use same set of predictors in both parts, but do not have to
- Can also **model positive values** using **truncated overdispersed Poisson** regression or **truncated negative binomial** regression

5.3 Evaluation and comparison

5.3.1 Pseudo- R^2 values

$$R^2 = 1 - \frac{deviance_{model}}{deviance_{interceptmodel}}$$

- Theoretically bounded by 0 and 1
- Recommended for count models
- Also recommended: correlation between Y and \hat{Y}

NOTE: for count models, use **deviance** not -2LL

5.3.2 Likelihood ratio tests

$$\chi^2 = deviance_{model1} - deviance_{model2}$$

- Model 1: simpler model (fewer predictors, worse fit)
- Model 2: more complex model (more predictors, better fit)
- **Degrees of freedom** = difference in number of parameters
 - **Significant test**: Model 1 is significantly worse than Model 2
 - **NS test**: Model 1 and 2 are not significantly different, so go with simpler one (Model 1)

NOTE: for count models, use **deviance** not -2LL

5.4 Nested models

5.4.1 Nested models

- Two models are **nested** if
 - All terms in one are **included in the other**
 - You can get from one to the other by **fixing some paths to a value**
- Why do we care?
 - Likelihood ratio tests can only be used to compare *nested models*
 - * If your models are not nested, you **cannot** use the likelihood ratio tests to compare them
 - Instead, use **AIC** or **BIC** (smaller is better)

5.4.2 Nested models

1. Linear regression with predictor X_1
 2. Linear regression with predictors X_1 and X_2
- Model 1 is nested within model 2
 - Fix the regression path for X_2 to 0 and Model 2 becomes Model 1

5.4.3 Nested models

3. Poisson regression with predictors X_3 and X_4
 4. Overdispersed Poisson regression w/ predictors X_3 and X_4
- Model 3 is nested within model 4
 - Note that this is a test of the dispersion parameter, ψ

5.4.4 Complicated nesting situations 1

5. Overdispersed Poisson regression w/ predictor X_5
 6. Overdispersed Poisson regression w/ predictors X_5 and X_6
- Model 5 is NOT nested within model 6
 - In addition to the extra predictor (X_6), the *overdispersion parameter* is not the same in these models
 - Model 5 has a ψ parameter based on only X_5
 - Model 6 has a **different** ψ parameter based on both X_5 and X_6
 - You can't get from Model 6 to Model 5 by fixing a parameter to 0

5.4.5 Complicated nesting situations 2

7. Overdispersed Poisson regression w/ predictors X_7 and X_8
 8. Negative binomial regression w/ predictors X_7 and X_8
- Model 7 is NOT nested within model 8
 - The overdispersion parameters are **completely different** for the two models
 - Model 7 has a ψ parameter and variance = $\psi\mu$
 - Model 8 has a α parameter and estimates variance = $\mu + \alpha\mu^2$

5.5 Conclusion

5.5.1 Summary

- When your outcome is a **count**, use a model from the **Poisson regression** family
 - *Nonlinear, exponential model or log-linear model*
 - *Overdispersion*: overdispersed Poisson or negative binomial
 - Too many or too few zeroes

5.5.2 In class exercises

- Use the same data as last time
 - Original count outcome of **number of drinks**, not binge
- Run a variety of count models
 - Poisson, overdispersed Poisson, negative binomial