

Categorical: Poisson regression

Table of contents

1	Goals	2
1.1	Goals	2
2	Counts and frequencies	2
2.1	Characteristics	2
2.2	Poisson distribution	3
3	Poisson regression	6
3.1	Poisson regression	6
3.2	Count metric	11
3.3	$\ln(\text{count})$ metric	13
3.4	Confidence intervals	14
4	Overdispersion	14
4.1	Overdispersion	14
4.2	Overdispersed Poisson regression	16
4.3	Negative binomial regression	18
5	Wrap-up	25
5.1	Choosing a model	25
6	Summary	25
6.1	Summary	25

1 Goals

1.1 Goals

1.1.1 Goals of this lecture

- Introduce models for *count outcomes*
 - Counts: *Ratio* level, but...
 - * ...discrete, lower bound, often skewed
 - Poisson (pwah-sahn) distribution
 - Two metrics for interpretation

2 Counts and frequencies

2.1 Characteristics

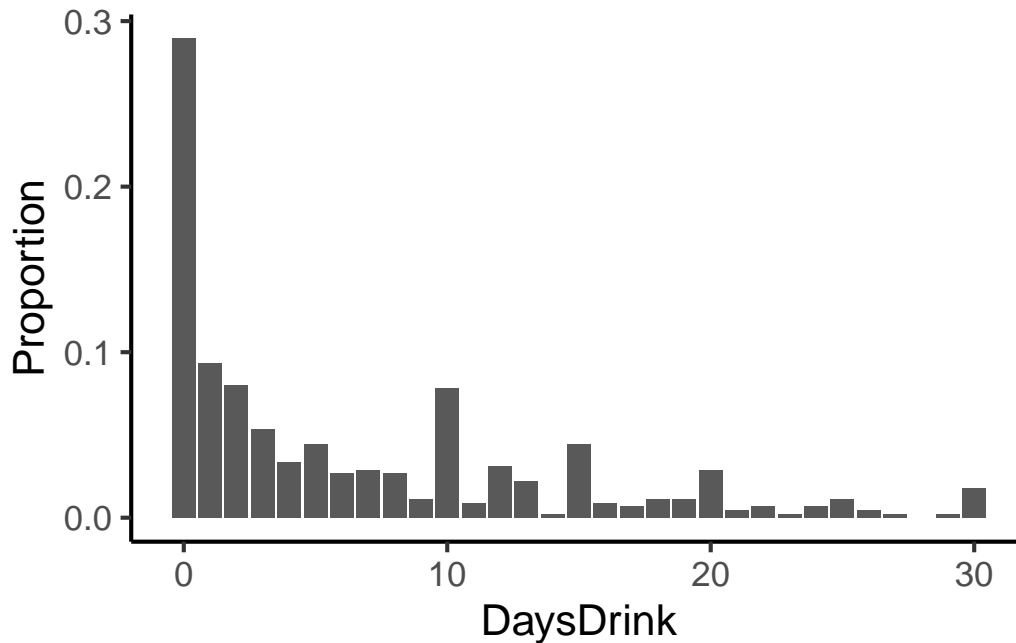
2.1.1 Count outcomes

- Discrete: Only take on whole number (integer) values
- Lower bound of 0
- Often right skewed

2.1.2 Count outcomes in psychology

- Number of depressive symptoms experienced in the past month
- Number of details you remember from a crime
- Number of times an older adult falls
- Number of aggressive acts by a child on the playground
- Number of different jobs an individual has by age 50
- And on and on...

2.1.3 Example: Number of days drinking in last 30 days



2.2 Poisson distribution

2.2.1 Poisson distribution: Simeon Denis Poisson

- Any time you *count* the *number of times* something happens, when that thing has a relatively *low probability* of happening
 - Number of soldiers killed by horse-kicks each year in each corps in the Prussian cavalry (Bortkiewicz, The Law of Small Numbers)
 - 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

2.2.2 Poisson distribution

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

- Probability of *exactly* k events

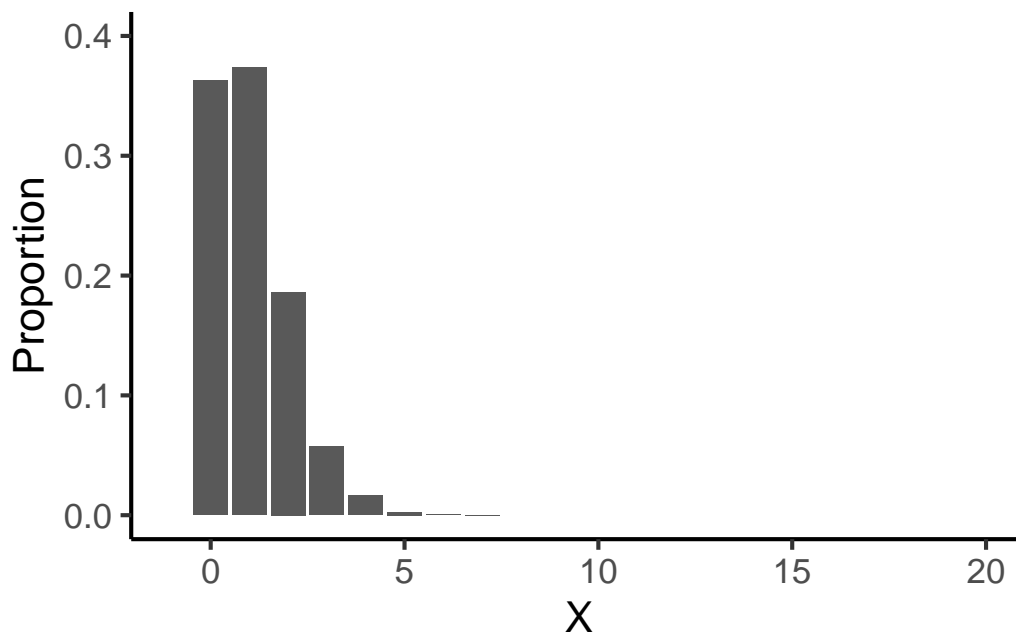
- Exactly $k = 5$ typographical errors on a page when the mean number of errors is $\lambda = 0.1$
- Exactly $k = 12$ earthquakes in a day when the mean number of earthquakes is $\lambda = 20$

2.2.3 Poisson distribution

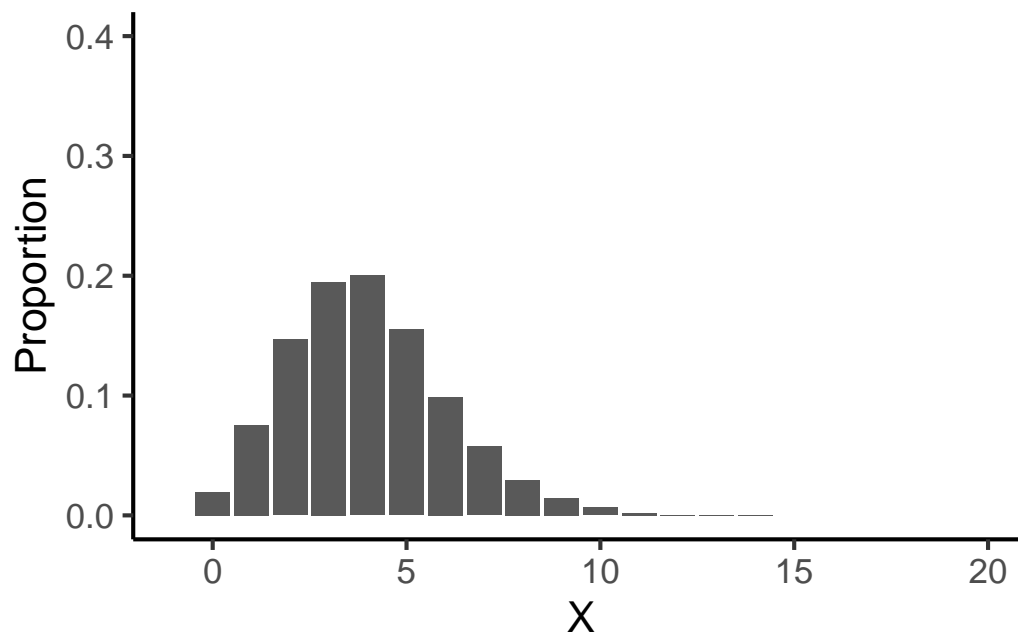
$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

- 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**

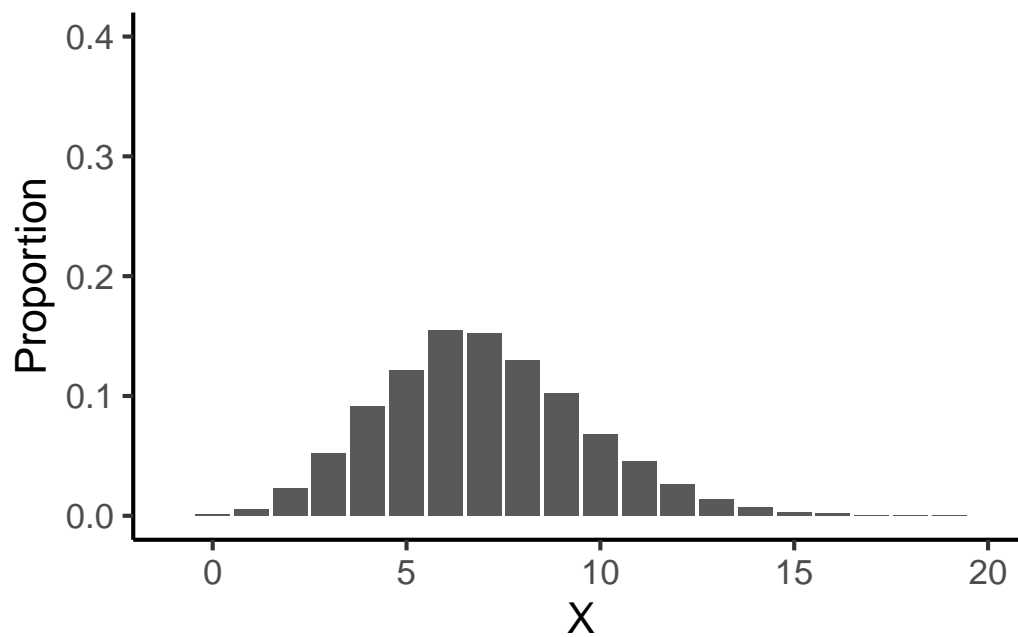
2.2.4 Figure: mean = variance = 1



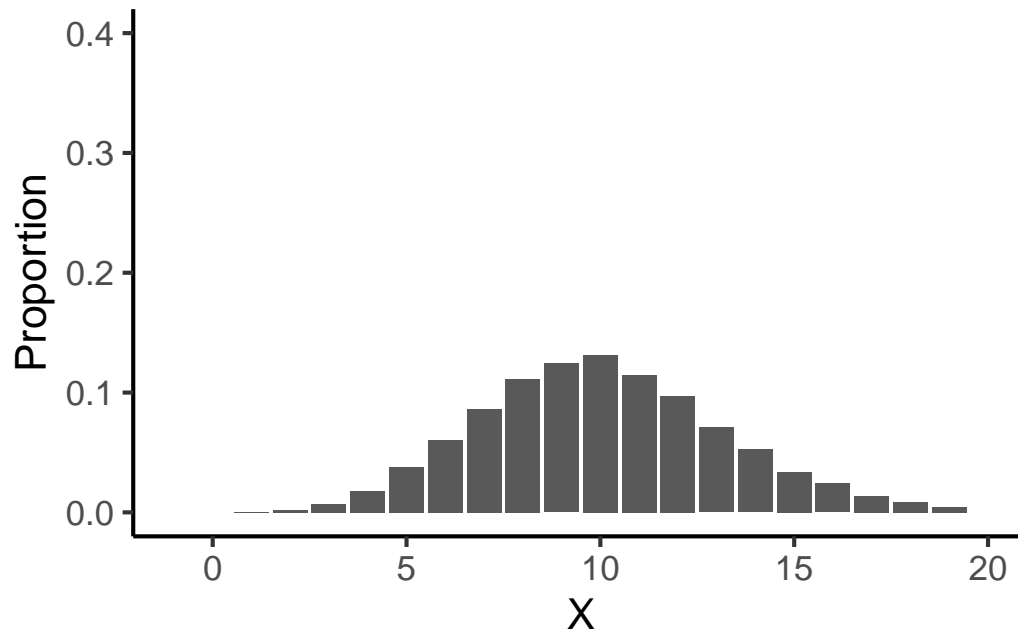
2.2.5 Figure: mean = variance = 4



2.2.6 Figure: mean = variance = 7



2.2.7 Figure: mean = variance = 10



3 Poisson regression

3.1 Poisson regression

3.1.1 Poisson regression

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

3.1.2 Two metrics for 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 \ln form is a straight line

3.1.3 Example data

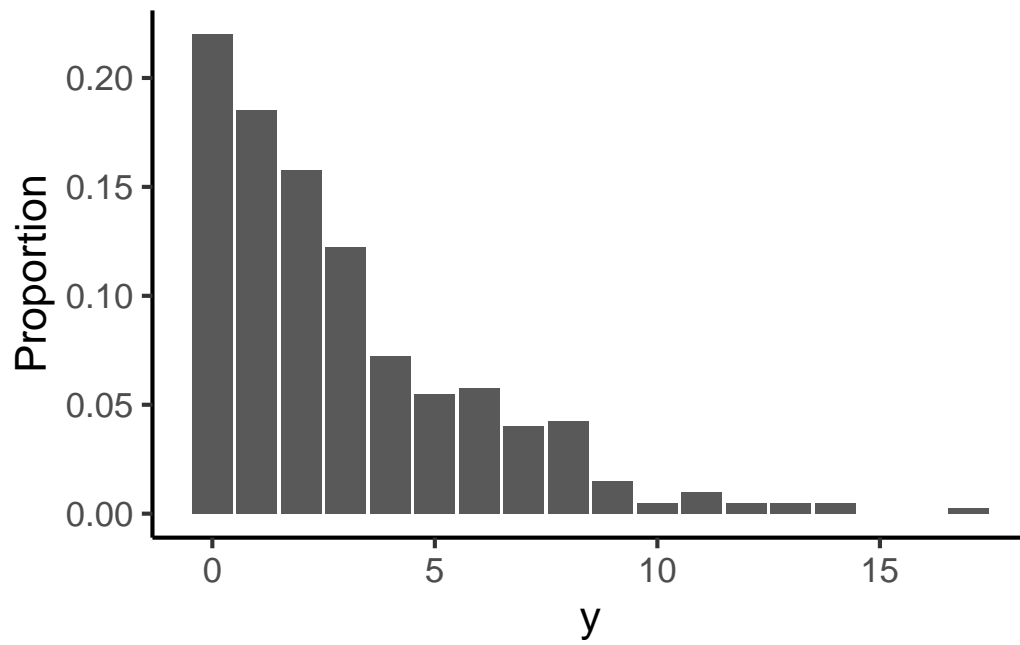
- Simulated data
 - **case**: Subject ID
 - **sensation**: Sensation seeking (1 to 7)
 - **gender**: 0 = female, 1 = male
 - **y**: Number of alcoholic beverages consumed on Saturday night

Coxe, S., West, S. G., & Aiken, L. S. (2009). The analysis of count data: A gentle introduction to Poisson regression and its alternatives. *Journal of Personality Assessment*, 91(2), 121-136.

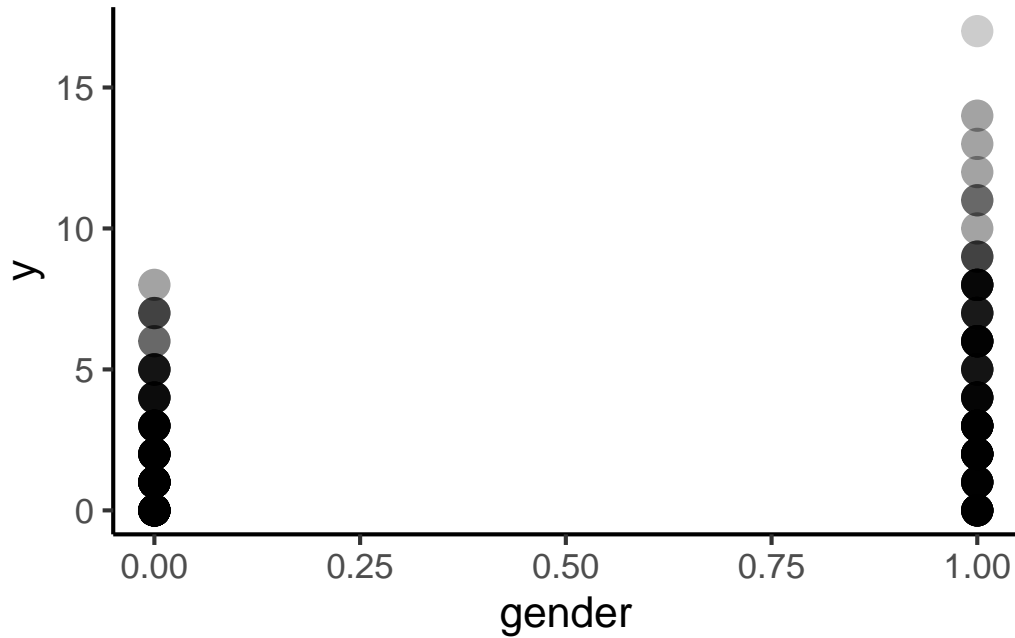
3.1.4 Example data

```
# A tibble: 12 x 4
  case sensation gender    y
  <dbl>      <dbl> <dbl> <dbl>
1     1         6.06     1     4
2     2         5.31     1     4
3     3         3.94     1     1
4     4         4.30     1     6
5     5         5.58     1     0
6     6         4.89     1     8
7     7         4.51     0     6
8     8         6.16     0     1
9     9         4.19     1     7
10    10         5.64     1     8
11    11         5.56     1     3
12    12         5.46     0     3
```

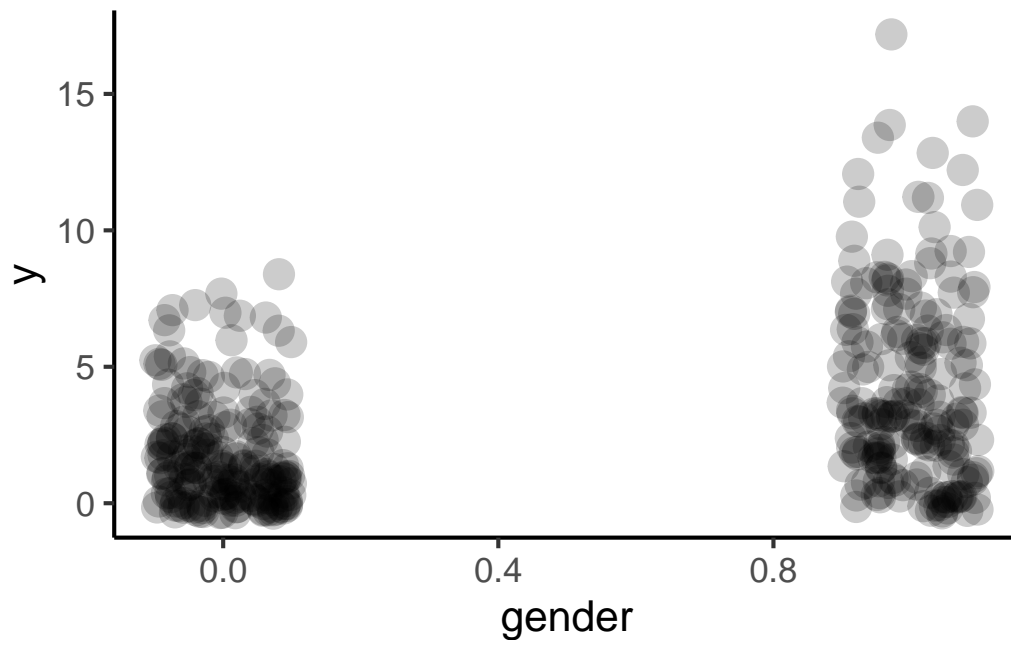
3.1.5 Example data: Outcome (y)



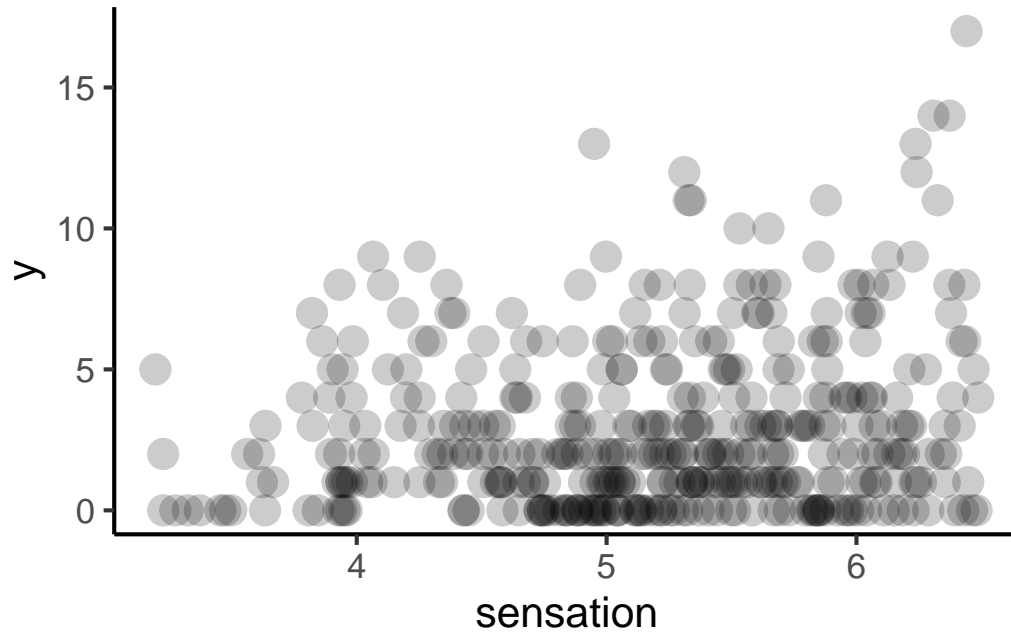
3.1.6 Example data: Predictor (gender)



3.1.7 Example data: Predictor (gender) with jitter



3.1.8 Example data: Predictor (sensation)



3.1.9 Output

Call:

```
glm(formula = y ~ sensation4, family = poisson("log"), data = alcohol)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.7912	-1.5001	-0.4624	0.8418	4.9122

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.78560	0.05977	13.144	< 2e-16 ***
sensation4	0.23148	0.03966	5.837	5.33e-09 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

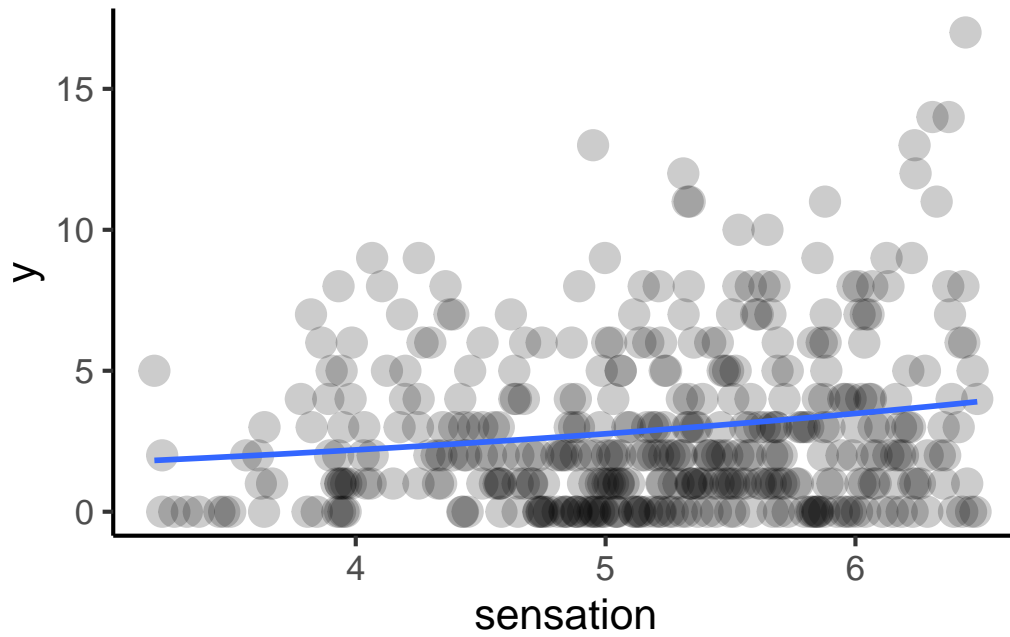
(Dispersion parameter for poisson family taken to be 1)

Null deviance: 1186.8 on 399 degrees of freedom
Residual deviance: 1151.7 on 398 degrees of freedom
AIC: 2079

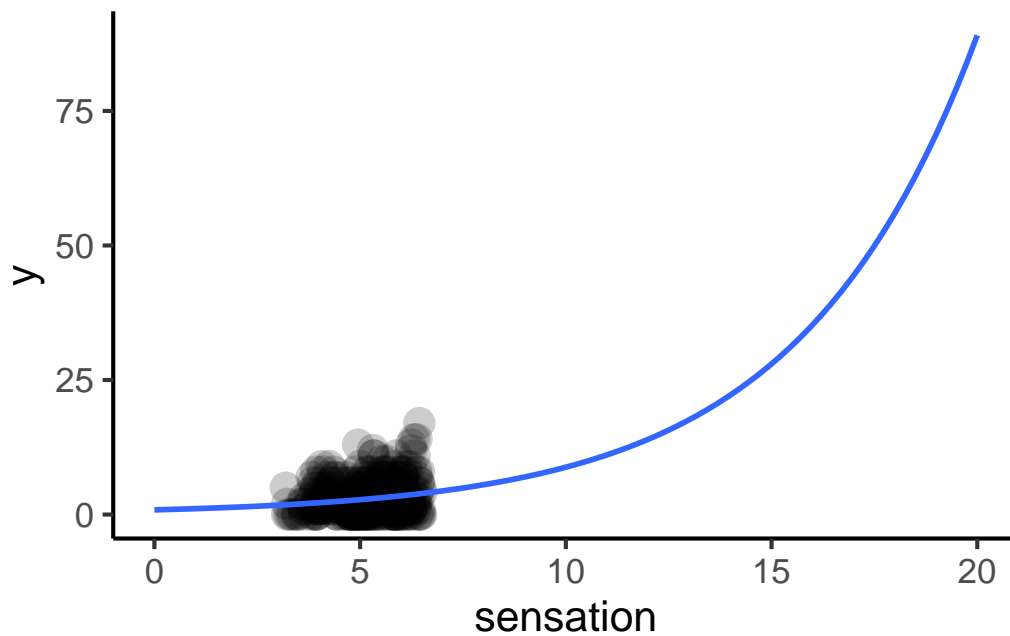
Number of Fisher Scoring iterations: 5

3.2 Count metric

3.2.1 Count metric interpretation: $\hat{\mu} = e^{0.786+0.231X}$



3.2.2 Count metric interpretation: $\hat{\mu} = e^{0.786+0.231X}$



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

- $e^{b_0} = e^{0.785603} = 2.194$
 - Predicted count when $X = 0$
 - * When **sensation** = 4
- $e^{b_1} = e^{0.2314809} = 1.26$
 - *Multiplicative* effect of X on the predicted count
 - For a 1 unit increase in X , the predicted count is **multiplied** by 1.26
 - * You can interpret this as a 26 **percent** increase

3.2.5 Count metric interpretation

X	Predicted count
-3	1.095
-2	1.381
-1	1.740
0	2.194
1	2.765
2	3.485
3	4.393

3.3 ln(count) metric

3.3.1 Log of predicted count metric interpretation $\ln(\hat{\mu}) = 0.786 + 0.231X$

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

3.3.2 Log of predicted count metric interpretation

- $b_0 = 0.7856$
 - $\ln(\hat{\mu})$ when $X = 0$
 - * When `sensation` = 4
- $b_1 = 0.2315$
 - *Additive* effect of x on the $\ln(\hat{\mu})$
 - For a 1 unit increase in X , the $\ln(\hat{\mu})$ increases by 0.2315

3.3.3 Log of predicted count metric interpretation

X	Predicted count	ln(Predicted count)
-3	1.095	0.091
-2	1.381	0.323
-1	1.740	0.554
0	2.194	0.786
1	2.765	1.017
2	3.485	1.249
3	4.393	1.480

3.4 Confidence intervals

3.4.1 Confidence intervals

- Results from software are in $\ln(\hat{\mu})$ metric
 - Linear metric
 - Significant effect if CIs do not include 0

	b	Lower CL	Upper CL
(Intercept)	0.786	0.667	0.901
sensation4	0.231	0.154	0.310

3.4.2 Confidence intervals

- Exponentiate **each value** (e^x) to get $\hat{\mu}$ metric and **rate ratio**
 - Nonlinear metric
 - Significant effect if CIs do not include 1

	exp.b.	Lower.CL	Upper.CL
(Intercept)	2.194	1.948	2.462
x	1.260	1.167	1.363

4 Overdispersion

4.1 Overdispersion

4.1.1 Equidispersion

- Poisson distribution
 - Mean = λ
 - Variance = λ
- Poisson regression
 - (Conditional) mean and (conditional) variance are *equal*
 - This is called **equidispersion**

4.1.2 Overdispersion

- In practice: *Actual* variance of Y is typically *larger* than expected by the Poisson distribution
 - This is called **overdispersion**
- Variance can be smaller: Underdispersion
 - Less common

4.1.3 Overdispersion

- Why does overdispersion happen?
 - Sometimes, just because
 - * The variance is just larger than Poisson distribution says
 - Due to omitted predictors
 - * Mean and variance are related, so anything that impacts the mean also impacts the variance
 - “Excess” zeroes
 - * More zeroes than expected for a Poisson distribution with that mean

4.1.4 Overdispersion

- Why do we care about overdispersion?
 - Variance estimated in the model (λ) helps determine **standard errors** for regression coefficients
 - Also plays a role in calculating **deviance**
- If the variance in our data is larger than λ
 - *Wrong* standard errors
 - *Wrong* deviance
 - *Wrong*: Tests of significance, pseudo- R^2 , LR tests

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

- Simplest way to deal with overdispersion
 - Include “fudge factor” to account for overdispersion
 - Variance is larger, so multiply variance by something
- 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
 - R: Scale value in output is ψ
 - SPSS: 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 Output: Overdispersed Poisson regression

Call:

```
glm(formula = y ~ sensation4, family = quasipoisson(link = "log"),
     data = alcohol)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-2.7912	-1.5001	-0.4624	0.8418	4.9122

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.78560	0.10086	7.789	5.91e-14 ***
sensation4	0.23148	0.06692	3.459	0.000601 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 2.847168)

Null deviance: 1186.8 on 399 degrees of freedom
Residual deviance: 1151.7 on 398 degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 5

4.2.6 Comparison

- Poisson regression:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.7856030	0.05977108	13.143532	1.853685e-39
sensation4	0.2314809	0.03966037	5.836581	5.328294e-09

- Overdispersed Poisson regression ($\psi = 2.85$):

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.7856030	0.10085504	7.789428	5.906205e-14
sensation4	0.2314809	0.06692113	3.459011	6.007424e-04

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

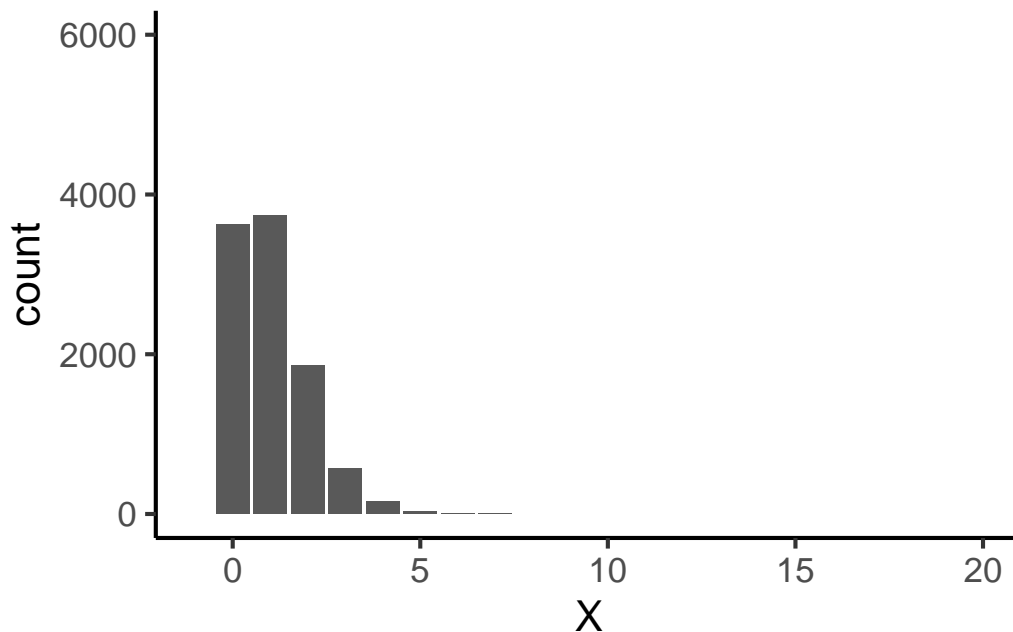
Negative binomial regression allows people with the **same predictor value(s)** to come from **different Poisson distributions**

- **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**

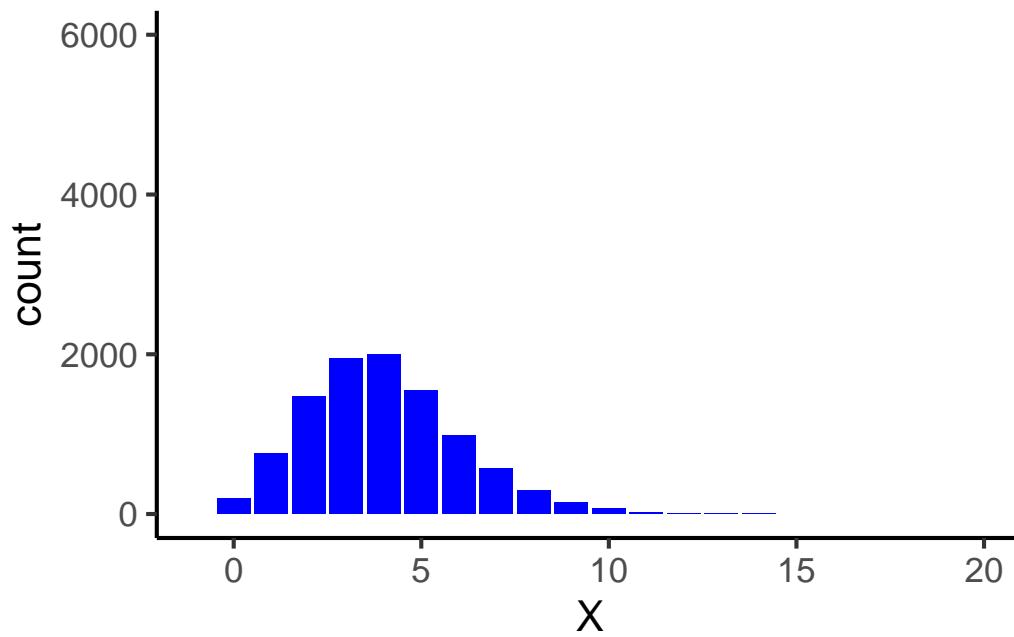
The means of the Poisson distributions follow a **gamma distribution**

- The *average* of those distributions is the same as Poisson regression
- But negative binomial has **higher variance** than Poisson

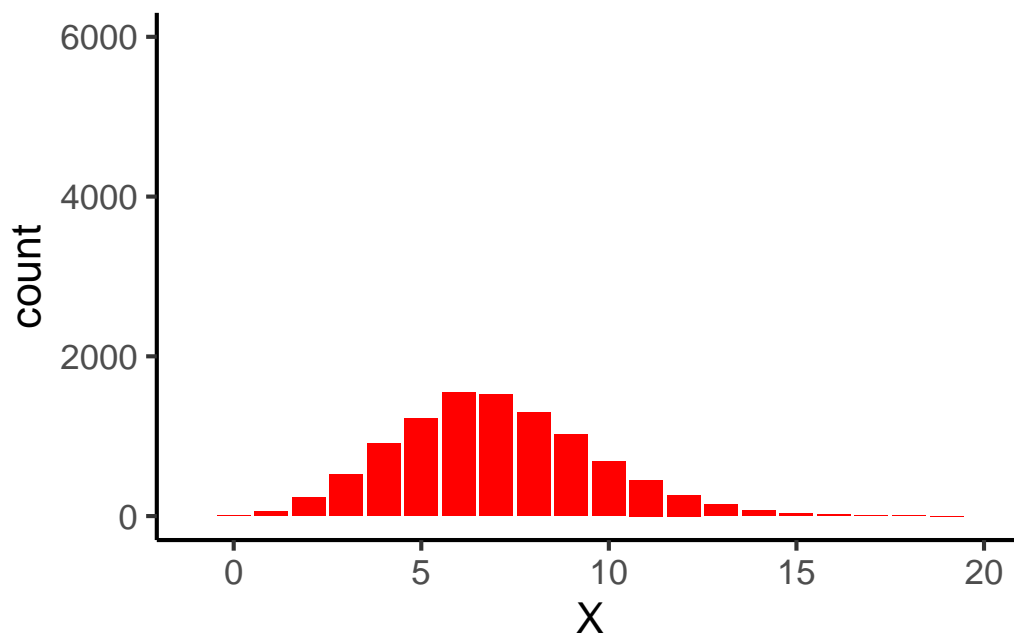
4.3.4 Person A: Y from Poisson distribution with $\lambda = 1$



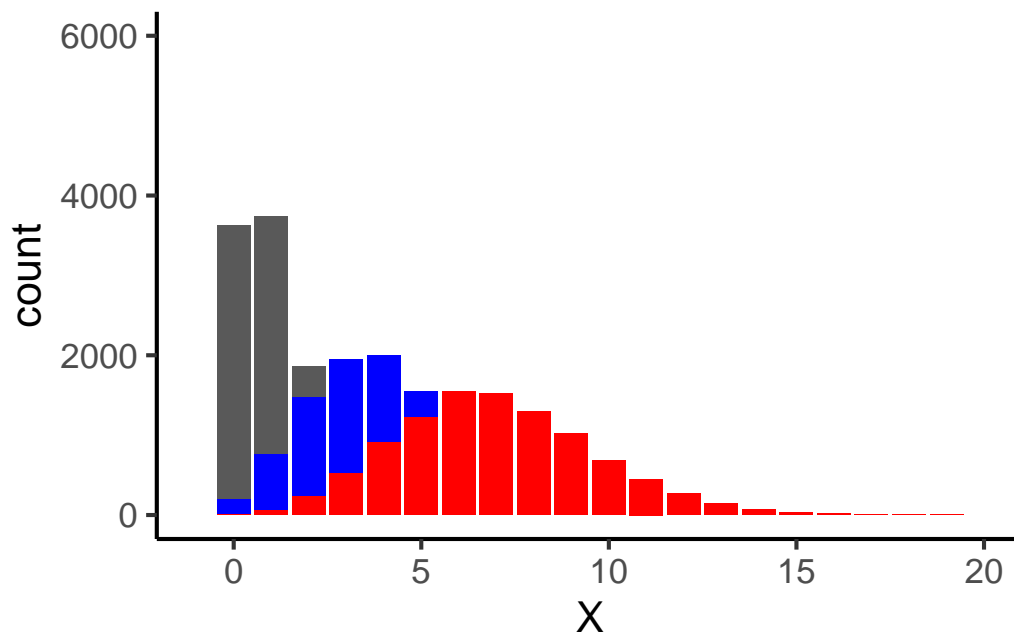
4.3.5 Person B: Y from Poisson distribution with $\lambda = 4$



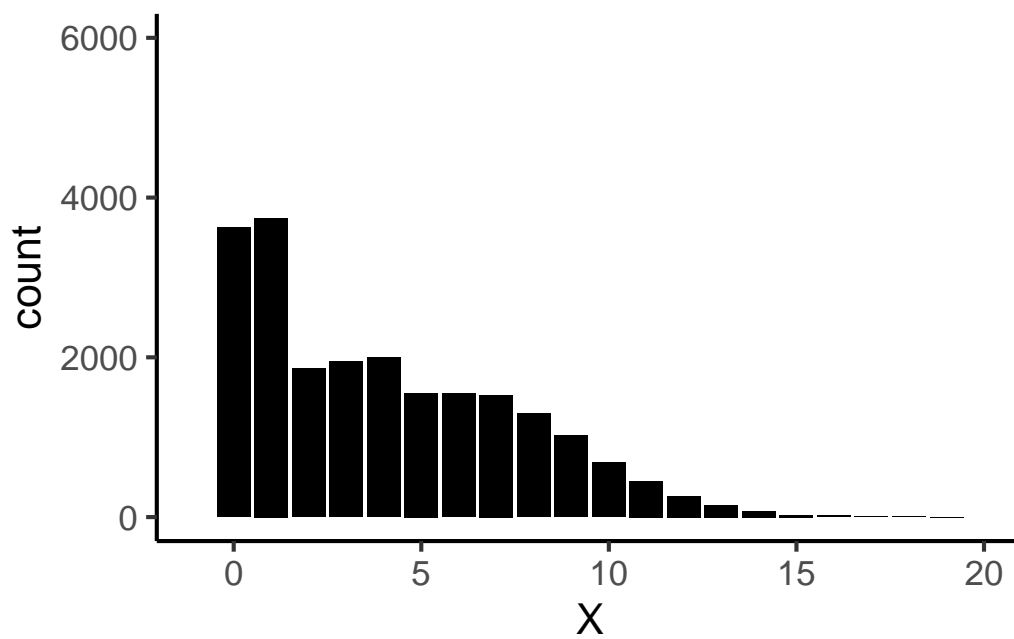
4.3.6 Person C: Y from Poisson distribution with $\lambda = 7$



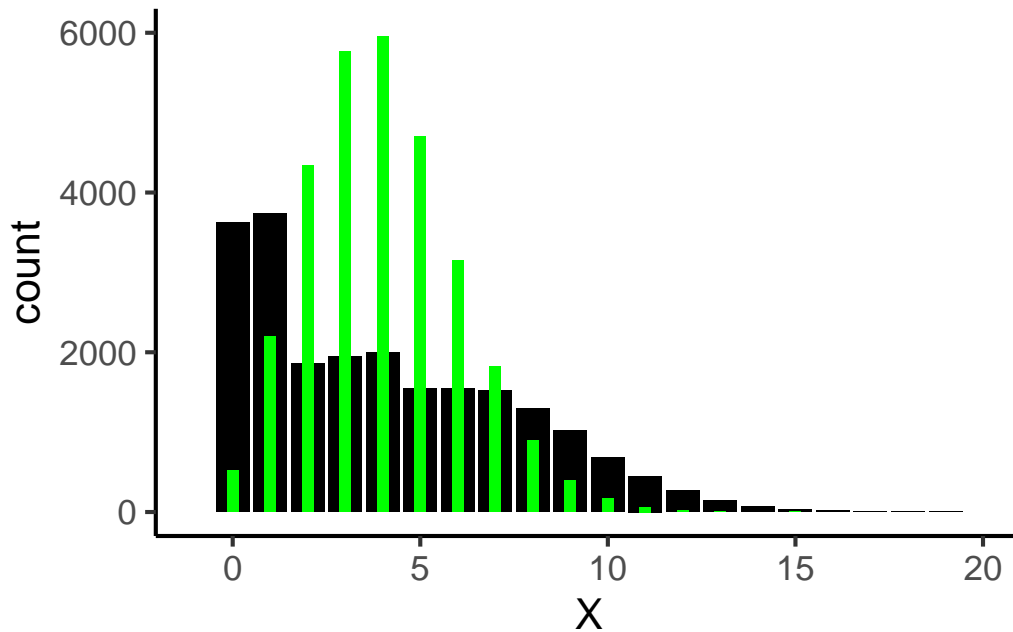
4.3.7 Combine distributions from person A, B, and C



4.3.8 Combine distributions from person A, B, and C



4.3.9 Compare to Poisson distribution with same mean



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 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 change*

4.3.13 Output: Negative binomial regression

Call:

```
glm.nb(formula = y ~ sensation4, data = alcohol, init.theta = 1.392859438,  
link = log)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.9207	-0.9158	-0.2809	0.4658	2.1349

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.79937	0.09831	8.131	4.26e-16 ***
sensation4	0.22050	0.06822	3.232	0.00123 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(1.3929) family taken to be 1)

Null deviance: 463.49 on 399 degrees of freedom
Residual deviance: 452.56 on 398 degrees of freedom
AIC: 1772.1

Number of Fisher Scoring iterations: 1

Theta: 1.393
Std. Err.: 0.163

2 x log-likelihood: -1766.091

4.3.14 Comparison

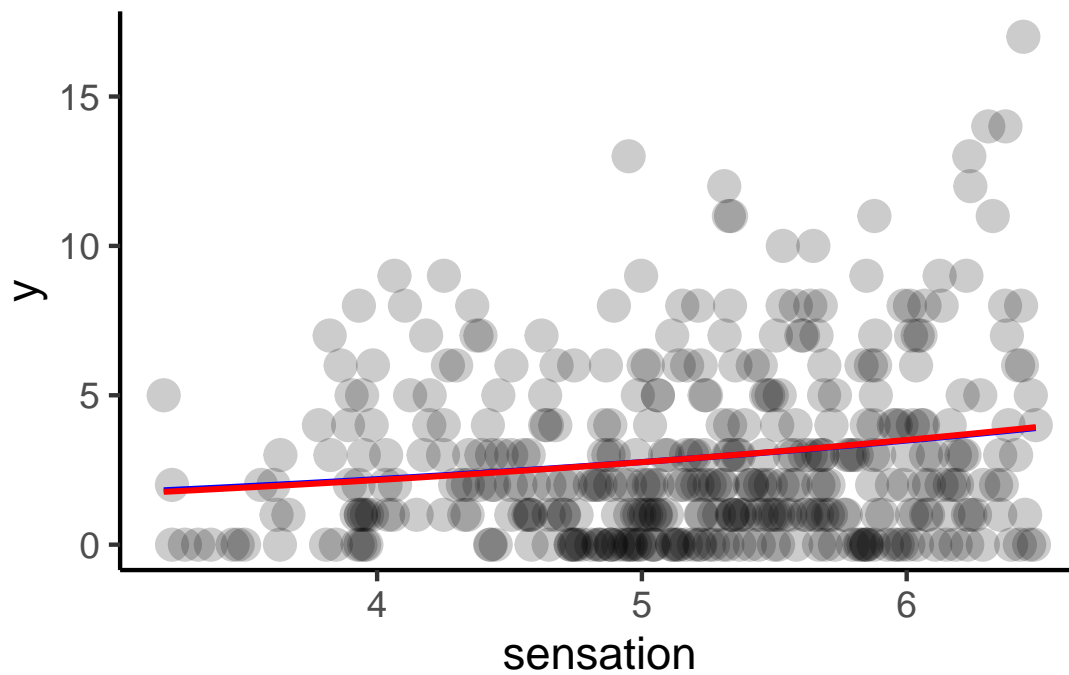
Poisson regression:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.7856030	0.05977108	13.143532	1.853685e-39
sensation4	0.2314809	0.03966037	5.836581	5.328294e-09

Negative binomial regression ($\alpha = 0.718$):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.7993708	0.09831247	8.130920	4.260448e-16
sensation4	0.2204963	0.06822183	3.232049	1.229061e-03

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



5 Wrap-up

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*

6 Summary

6.1 Summary

6.1.1 Summary of this week

- Models for counts
 - Poisson is the simplest
- Overdispersion
 - Concern for count models because Poisson: $\text{mean} = \text{variance}$

- Overdispersed Poisson
- Negative binomial

6.1.2 Next week: More count models!

- Zero is *lower bound* for a count
 - But also substantively interesting
 - * “Nothing” occurred: No drinking, no symptoms, no incidents
 - What if we have “too many” or “too few” 0s?
- Variable lengths of time
 - Poisson distribution: Number of events in a *fixed period of time*
- Model comparisons, etc.