Count Data Models

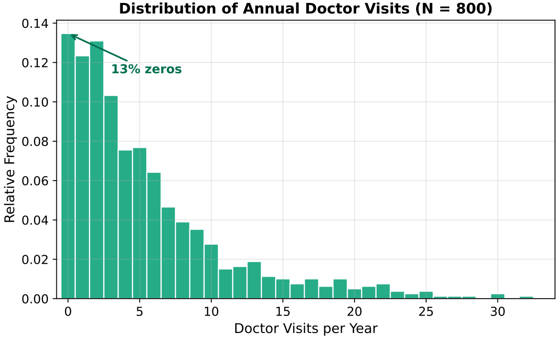
Jake Anderson

May 16, 2026

Outline

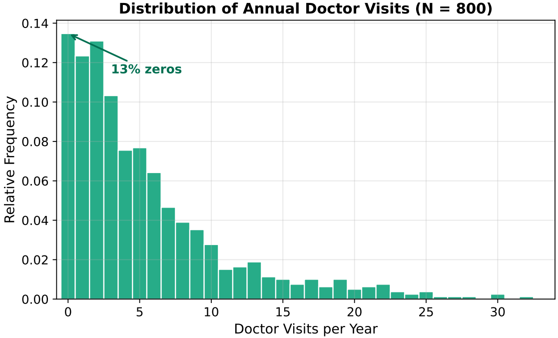
The Data

A health economist surveys **800 individuals** and records their **annual doctor visits**. Covariates include age, insurance status, and a health index (centered near 0; higher = healthier).



The Data

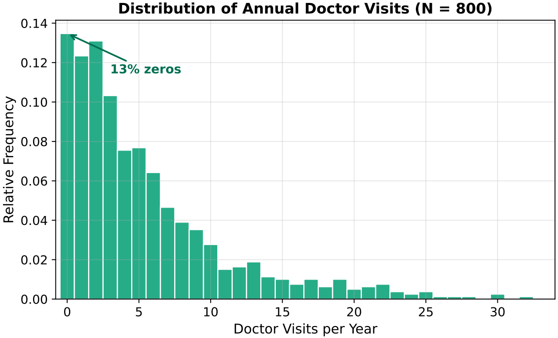
A health economist surveys **800 individuals** and records their **annual doctor visits**. Covariates include age, insurance status, and a health index (centered near 0; higher = healthier).



What do you notice about this distribution?

The Data

A health economist surveys **800 individuals** and records their **annual doctor visits**. Covariates include age, insurance status, and a health index (centered near 0; higher = healthier).



What do you notice about this distribution?

The outcome is a **count**: non-negative integers (0, 1, 2, . . .). Right-skewed with a spike at zero. Mean = 5.7, but 13% have zero visits.

OLS Predictions on Count Data

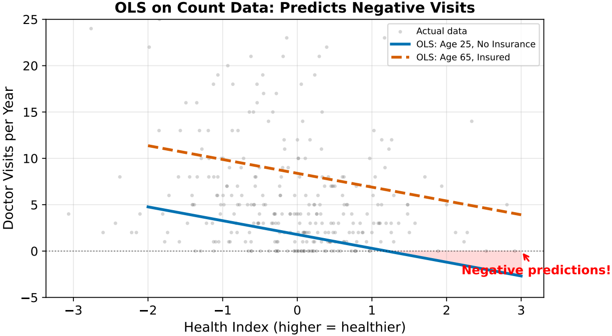
Treat doctor visits as a continuous variable and regress on covariates:

$$\text{Visits}_i = \beta_0 + \beta_1 \,\text{Age}_i + \beta_2 \,\text{Insurance}_i + \beta_3 \,\text{Health}_i + \varepsilon_i$$

OLS Predictions on Count Data

Treat doctor visits as a continuous variable and regress on covariates:

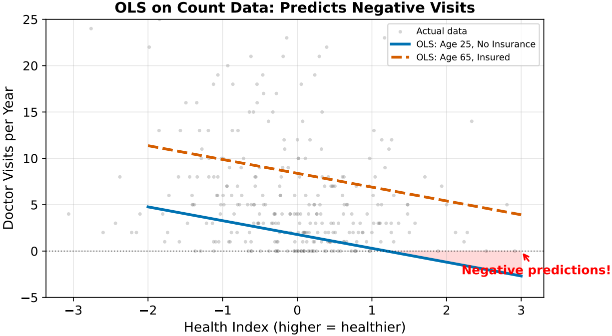
$$\text{Visits}_i = \beta_0 + \beta_1 \, \text{Age}_i + \beta_2 \, \text{Insurance}_i + \beta_3 \, \text{Health}_i + \varepsilon_i$$



OLS Predictions on Count Data

Treat doctor visits as a continuous variable and regress on covariates:

$$\text{Visits}_i = \beta_0 + \beta_1 \,\text{Age}_i + \beta_2 \,\text{Insurance}_i + \beta_3 \,\text{Health}_i + \varepsilon_i$$



For a 25-year-old without insurance, OLS predicts **negative visits** once the health index exceeds about 1.5. Doctor visits cannot be negative.

Three Failures of OLS on Counts

The plot reveals the first problem, but there are two more:

Three Failures of OLS on Counts

The plot reveals the first problem, but there are two more:

1. **Negative predictions.** OLS can predict -2.4 visits for a young, healthy, uninsured person

Three Failures of OLS on Counts

The plot reveals the first problem, but there are two more:

1. **Negative predictions.** OLS can predict -2.4 visits for a young, healthy, uninsured person
2. **Non-constant variance.** People who average 10 visits have far more variation than those who average 1

Three Failures of OLS on Counts

The plot reveals the first problem, but there are two more:

1. **Negative predictions.** OLS can predict -2.4 visits for a young, healthy, uninsured person
2. **Non-constant variance.** People who average 10 visits have far more variation than those who average 1
3. **Non-normal residuals.** Count data is right-skewed and discrete; OLS assumes symmetric, continuous errors

Three Failures of OLS on Counts

The plot reveals the first problem, but there are two more:

1. **Negative predictions.** OLS can predict -2.4 visits for a young, healthy, uninsured person
2. **Non-constant variance.** People who average 10 visits have far more variation than those who average 1
3. **Non-normal residuals.** Count data is right-skewed and discrete; OLS assumes symmetric, continuous errors

$\Longrightarrow$ We need a model built for count outcomes from the start.

What Would a Better Model Need?

A model for count outcomes should:

What Would a Better Model Need?

A model for count outcomes should:

1. **Guarantee non-negative predictions.** $\hat{y}_i \geq 0$ for all covariate values

What Would a Better Model Need?

A model for count outcomes should:

1. **Guarantee non-negative predictions.** $\hat{y}_i \geq 0$ for all covariate values
2. **Handle the variance-mean relationship.** Individuals with higher expected visits naturally have more spread

What Would a Better Model Need?

A model for count outcomes should:

1. **Guarantee non-negative predictions.** $\hat{y}_i \geq 0$ for all covariate values
2. **Handle the variance-mean relationship.** Individuals with higher expected visits naturally have more spread
3. **Accommodate the spike at zero.** Many people never visit the doctor; the model should not be surprised by this

What Would a Better Model Need?

A model for count outcomes should:

1. **Guarantee non-negative predictions.** $\hat{y}_i \geq 0$ for all covariate values
2. **Handle the variance-mean relationship.** Individuals with higher expected visits naturally have more spread
3. **Accommodate the spike at zero.** Many people never visit the doctor; the model should not be surprised by this

$\Longrightarrow$ Where can we find a probability distribution designed for non-negative integers?

The Poisson Distribution for Counts

You already know the binary case: we replaced OLS with logit/probit to keep predictions in $[0, 1]$.

The Poisson Distribution for Counts

You already know the binary case: we replaced OLS with logit/probit to keep predictions in $[0, 1]$.

Same logic here: we need a **distribution for counts** to replace the normal distribution.

The Poisson Distribution for Counts

You already know the binary case: we replaced OLS with logit/probit to keep predictions in $[0, 1]$.

Same logic here: we need a **distribution for counts** to replace the normal distribution.

The simplest count distribution is the **Poisson**: it assigns probabilities to 0, 1, 2, 3, ... and has one parameter that controls both the mean and the variance.

The Poisson Distribution for Counts

You already know the binary case: we replaced OLS with logit/probit to keep predictions in $[0, 1]$.

Same logic here: we need a **distribution for counts** to replace the normal distribution.

The simplest count distribution is the **Poisson**: it assigns probabilities to 0, 1, 2, 3, ... and has one parameter that controls both the mean and the variance.

$\Longrightarrow$ Let's build a regression model on top of the Poisson distribution, just as logit builds on the logistic distribution.

Outline

The Poisson Distribution

A random variable Y follows a Poisson distribution with parameter $\mu > 0$ if:

$$P(Y = k) = \frac{e^{-\mu} \mu^k}{k!}, \qquad k = 0, 1, 2, \ldots$$

The Poisson Distribution

A random variable Y follows a Poisson distribution with parameter $\mu > 0$ if:

$$P(Y = k) = \frac{e^{-\mu}\,\mu^k}{k!}, \qquad k = 0, 1, 2, \ldots$$

Properties:

- $\mathrm{E}[Y] = \mu$
- $\mathrm{Var}(Y) = \mu \quad \Longrightarrow$ **equidispersion**: the variance equals the mean
- As μ increases, the distribution shifts right and spreads out

The Poisson Distribution

A random variable Y follows a Poisson distribution with parameter $\mu > 0$ if:

$$P(Y = k) = \frac{e^{-\mu}\,\mu^k}{k!}, \qquad k = 0, 1, 2, \ldots$$

Properties:

- $\text{E}[Y] = \mu$
- $\text{Var}(Y) = \mu \quad \Longrightarrow$ **equidispersion**: the variance equals the mean
- As μ increases, the distribution shifts right and spreads out

Example: if $\mu = 6$, then $P(Y = 0) = e^{-6} \approx 0.0025$ and $P(Y = 6) \approx 0.16$.

Poisson Regression: The Log Link

To build a regression, we let the Poisson parameter μ_i depend on covariates. But $\mu_i > 0$, so we need to keep predictions positive.

Poisson Regression: The Log Link

To build a regression, we let the Poisson parameter μ_i depend on covariates. But $\mu_i > 0$, so we need to keep predictions positive.

The log link: model the log of the conditional mean as a linear function:

$$\ln(\mu_i) = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Insurance}_i + \beta_3 \text{Health}_i$$

Poisson Regression: The Log Link

To build a regression, we let the Poisson parameter μ_i depend on covariates. But $\mu_i > 0$, so we need to keep predictions positive.

The log link: model the log of the conditional mean as a linear function:

$$\ln(\mu_i) = \beta_0 + \beta_1 \,\mathrm{Age}_i + \beta_2 \,\mathrm{Insurance}_i + \beta_3 \,\mathrm{Health}_i$$

Equivalently:

$$\mu_i = \mathrm{E}[\mathrm{Visits}_i \mid \mathrm{covariates}] = e^{\beta_0 + \beta_1 \,\mathrm{Age}_i + \beta_2 \,\mathrm{Insurance}_i + \beta_3 \,\mathrm{Health}_i}$$

Poisson Regression: The Log Link

To build a regression, we let the Poisson parameter μ_i depend on covariates. But $\mu_i > 0$, so we need to keep predictions positive.

The log link: model the log of the conditional mean as a linear function:

$$\ln(\mu_i) = \beta_0 + \beta_1 \,\text{Age}_i + \beta_2 \,\text{Insurance}_i + \beta_3 \,\text{Health}_i$$

Equivalently:

$$\mu_i = \text{E}[\text{Visits}_i \mid \text{covariates}] = e^{\beta_0 + \beta_1 \,\text{Age}_i + \beta_2 \,\text{Insurance}_i + \beta_3 \,\text{Health}_i}$$

Since $e^{(\cdot)} > 0$ for any input, **predicted counts are always positive**. This solves the negative-prediction problem.

Estimation: Maximum Likelihood

Poisson regression is estimated by maximizing the log-likelihood:

$$\ell = \sum_{i=1}^{N} \left[y_i \ln(\mu_i) - \mu_i - \ln(y_i!) \right]$$

Estimation: Maximum Likelihood

Poisson regression is estimated by maximizing the log-likelihood:

$$\ell = \sum_{i=1}^{N} \left[y_i \ln(\mu_i) - \mu_i - \ln(y_i!) \right]$$

where $\mu_i = e^{\beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Insurance}_i + \beta_3 \text{Health}_i}$.

Estimation: Maximum Likelihood

Poisson regression is estimated by maximizing the log-likelihood:

$$\ell = \sum_{i=1}^{N} \left[y_i \ln(\mu_i) - \mu_i - \ln(y_i!) \right]$$

where $\mu_i = e^{\beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Insurance}_i + \beta_3 \text{Health}_i}$.

No closed-form solution $\implies$ solved numerically (same as logit). Software reports coefficients, standard errors, and predicted counts $\hat{\mu}_i$.

Estimation: Maximum Likelihood

Poisson regression is estimated by maximizing the log-likelihood:

$$\ell = \sum_{i=1}^{N} \left[y_i \ln(\mu_i) - \mu_i - \ln(y_i!) \right]$$

where $\mu_i = e^{\beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Insurance}_i + \beta_3 \text{Health}_i}$.

No closed-form solution $\implies$ solved numerically (same as logit). Software reports coefficients, standard errors, and predicted counts $\hat{\mu}_i$.

$\implies$ The structure is identical to binary logit/probit MLE, just with a different distribution (Poisson instead of Bernoulli).

Interpreting Coefficients: Semi-Elasticities

Take the log-link equation:

$$\ln(\mu_i) = \beta_0 + \beta_1 \,\text{Age}_i + \beta_2 \,\text{Insurance}_i + \beta_3 \,\text{Health}_i$$

Interpreting Coefficients: Semi-Elasticities

Take the log-link equation:

$$\ln(\mu_i) = \beta_0 + \beta_1 \, \mathsf{Age}_i + \beta_2 \, \mathsf{Insurance}_i + \beta_3 \, \mathsf{Health}_i$$

Recall: a difference in logs approximates a percent change. A one-unit increase in Age_i, holding everything else fixed:

$$\ln(\mu_i^{\text{new}}) - \ln(\mu_i^{\text{old}}) = \beta_1 \quad \Longleftrightarrow \quad \frac{\mu_i^{\text{new}} - \mu_i^{\text{old}}}{\mu_i^{\text{old}}} \approx \beta_1$$

Interpreting Coefficients: Semi-Elasticities

Take the log-link equation:

$$\ln(\mu_i) = \beta_0 + \beta_1 \,\text{Age}_i + \beta_2 \,\text{Insurance}_i + \beta_3 \,\text{Health}_i$$

Recall: a difference in logs approximates a percent change. A one-unit increase in Age_i, holding everything else fixed:

$$\ln(\mu_i^{\text{new}}) - \ln(\mu_i^{\text{old}}) = \beta_1 \quad \Longleftrightarrow \quad \frac{\mu_i^{\text{new}} - \mu_i^{\text{old}}}{\mu_i^{\text{old}}} \approx \beta_1$$

$\Longrightarrow$ Each coefficient is a **semi-elasticity**: a one-unit increase in x_k changes the expected count by approximately $\beta_k \times 100\%$.

Interpreting Coefficients: Semi-Elasticities

Take the log-link equation:

$$\ln(\mu_i) = \beta_0 + \beta_1 \,\text{Age}_i + \beta_2 \,\text{Insurance}_i + \beta_3 \,\text{Health}_i$$

Recall: a difference in logs approximates a percent change. A one-unit increase in Age_i, holding everything else fixed:

$$\ln(\mu_i^{\text{new}}) - \ln(\mu_i^{\text{old}}) = \beta_1 \quad \Longleftrightarrow \quad \frac{\mu_i^{\text{new}} - \mu_i^{\text{old}}}{\mu_i^{\text{old}}} \approx \beta_1$$

$\implies$ Each coefficient is a **semi-elasticity**: a one-unit increase in x_k changes the expected count by approximately $\beta_k \times 100\%$.

For small $|\beta_k|$ (say < 0.1), this approximation is accurate. For larger coefficients, use the exact formula: $100 \times (e^{\beta_k} - 1)\%$.

Interpreting Coefficients: Semi-Elasticities

Take the log-link equation:

$$\ln(\mu_i) = \beta_0 + \beta_1 \, \text{Age}_i + \beta_2 \, \text{Insurance}_i + \beta_3 \, \text{Health}_i$$

Recall: a difference in logs approximates a percent change. A one-unit increase in Age_i, holding everything else fixed:

$$\ln(\mu_i^{\text{new}}) - \ln(\mu_i^{\text{old}}) = \beta_1 \quad \Longleftrightarrow \quad \frac{\mu_i^{\text{new}} - \mu_i^{\text{old}}}{\mu_i^{\text{old}}} \approx \beta_1$$

$\Longrightarrow$ Each coefficient is a **semi-elasticity**: a one-unit increase in x_k changes the expected count by approximately $\beta_k \times 100\%$.

For small $|\beta_k|$ (say < 0.1), this approximation is accurate. For larger coefficients, use the exact formula: $100 \times (e^{\beta_k} - 1)\%$.

Example (Insurance, a dummy variable): if $\hat{\beta}_2 = 0.54$, then $e^{0.54} - 1 = 0.72$, so insured individuals have about 72% more visits.

Numeric Example: Predicted Visits

Suppose the Poisson estimates are $\hat{\beta}_0 = 0.50$, $\hat{\beta}_{\text{age}} = 0.017$, $\hat{\beta}_{\text{ins}} = 0.54$, $\hat{\beta}_{\text{health}} = -0.27$.

Numeric Example: Predicted Visits

Suppose the Poisson estimates are $\hat{\beta}_0 = 0.50$, $\hat{\beta}_{\text{age}} = 0.017$, $\hat{\beta}_{\text{ins}} = 0.54$, $\hat{\beta}_{\text{health}} = -0.27$.

Person A: 45 years old, insured, average health (Health $= 0$):

$$\ln(\hat{\mu}_A) = 0.50 + 0.017 \times 45 + 0.54 \times 1 + (-0.27) \times 0 = 1.805$$

$$\hat{\mu}_A = e^{1.805} \approx 6.1 \text{ visits per year}$$

Numeric Example: Predicted Visits

Suppose the Poisson estimates are $\hat{\beta}_0 = 0.50$, $\hat{\beta}_{\text{age}} = 0.017$, $\hat{\beta}_{\text{ins}} = 0.54$, $\hat{\beta}_{\text{health}} = -0.27$.

Person A: 45 years old, insured, average health (Health $= 0$):

$$\ln(\hat{\mu}_A) = 0.50 + 0.017 \times 45 + 0.54 \times 1 + (-0.27) \times 0 = 1.805$$

$$\hat{\mu}_A = e^{1.805} \approx 6.1 \text{ visits per year}$$

Person B: 25 years old, uninsured, healthy (Health $= 1.5$):

$$\ln(\hat{\mu}_B) = 0.50 + 0.017 \times 25 + 0.54 \times 0 + (-0.27) \times 1.5 = 0.520$$

$$\hat{\mu}_B = e^{0.520} \approx 1.7 \text{ visits per year}$$

Numeric Example: Predicted Visits

Suppose the Poisson estimates are $\hat{\beta}_0 = 0.50$, $\hat{\beta}_{\text{age}} = 0.017$, $\hat{\beta}_{\text{ins}} = 0.54$, $\hat{\beta}_{\text{health}} = -0.27$.

Person A: 45 years old, insured, average health (Health $= 0$):

$$\ln(\hat{\mu}_A) = 0.50 + 0.017 \times 45 + 0.54 \times 1 + (-0.27) \times 0 = 1.805$$

$$\hat{\mu}_A = e^{1.805} \approx 6.1 \text{ visits per year}$$

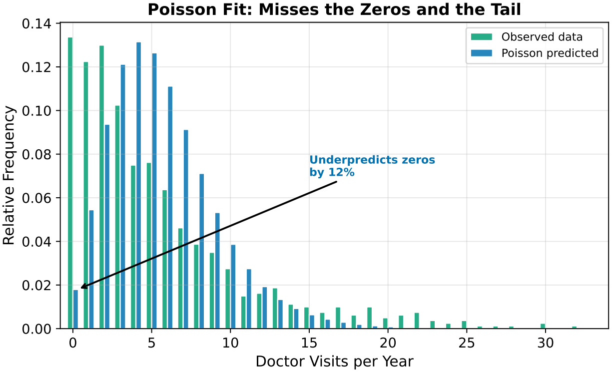
Person B: 25 years old, uninsured, healthy (Health $= 1.5$):

$$\ln(\hat{\mu}_B) = 0.50 + 0.017 \times 25 + 0.54 \times 0 + (-0.27) \times 1.5 = 0.520$$

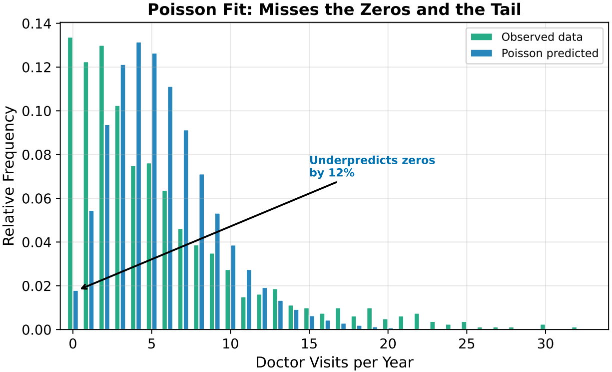
$$\hat{\mu}_B = e^{0.520} \approx 1.7 \text{ visits per year}$$

$\Longrightarrow$ Both predictions are positive. Compare to OLS, which predicted negative visits for Person B.

Poisson Fit to Our Data



Poisson Fit to Our Data



The Poisson predicts only 2% zeros; the data has 13%. It underpredicts zeros and underpredicts the right tail, concentrating too much mass in the middle. Why?

The Equidispersion Problem

Recall the Poisson assumption: $\text{Var}(Y_i) = \mu_i$. This restriction is called **equidispersion**: the variance must equal the mean.

The Equidispersion Problem

Recall the Poisson assumption: $\text{Var}(Y_i) = \mu_i$. This restriction is called **equidispersion**: the variance must equal the mean.

This means individuals with $\mu_i = 6$ expected visits should have variance $= 6$. But in our data:

The Equidispersion Problem

Recall the Poisson assumption: $\text{Var}(Y_i) = \mu_i$. This restriction is called **equidispersion**: the variance must equal the mean.

This means individuals with $\mu_i = 6$ expected visits should have variance $= 6$. But in our data:

	Mean visits	**Variance**
Full sample	5.7	43.8

The Equidispersion Problem

Recall the Poisson assumption: $\text{Var}(Y_i) = \mu_i$. This restriction is called **equidispersion**: the variance must equal the mean.

This means individuals with $\mu_i = 6$ expected visits should have variance $= 6$. But in our data:

	Mean visits	Variance
Full sample	5.7	43.8

The variance is **7.7 times** the mean. The Poisson model says these should be equal.

The Equidispersion Problem

Recall the Poisson assumption: $\text{Var}(Y_i) = \mu_i$. This restriction is called **equidispersion**: the variance must equal the mean.

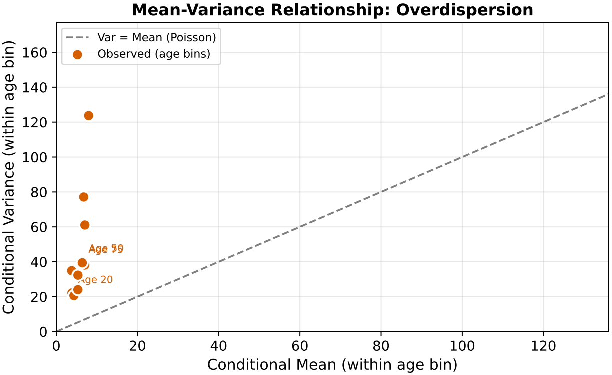
This means individuals with $\mu_i = 6$ expected visits should have variance $= 6$. But in our data:

	Mean visits	**Variance**
Full sample	5.7	43.8

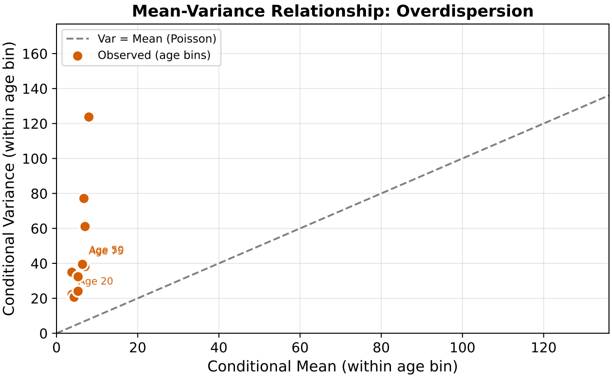
The variance is **7.7 times** the mean. The Poisson model says these should be equal.

This is called **overdispersion**: more variability in the data than the Poisson distribution allows. It is extremely common with count outcomes.

Visualizing Overdispersion



Visualizing Overdispersion



Every age bin lies **above** the 45-degree line. The variance grows faster than the mean, violating the Poisson assumption.

Consequence of Overdispersion: Standard Errors Are Wrong

What happens if we fit Poisson regression to overdispersed data?

Consequence of Overdispersion: Standard Errors Are Wrong

What happens if we fit Poisson regression to overdispersed data?

- The **coefficient estimates** are still consistent, as long as the conditional mean ($\mu_i = e^{\beta_0+\beta_1 x_1+\cdots}$) is correctly specified

Consequence of Overdispersion: Standard Errors Are Wrong

What happens if we fit Poisson regression to overdispersed data?

- The **coefficient estimates** are still consistent, as long as the conditional mean ($\mu_i = e^{\beta_0 + \beta_1 x_1 + \cdots}$) is correctly specified
- But the **model-based standard errors are too small** because they assume $\text{Var}(Y_i) = \mu_i$, while the true variance is larger

Consequence of Overdispersion: Standard Errors Are Wrong

What happens if we fit Poisson regression to overdispersed data?

- The **coefficient estimates** are still consistent, as long as the conditional mean ($\mu_i = e^{\beta_0 + \beta_1 x_1 + \cdots}$) is correctly specified
- But the **model-based standard errors are too small** because they assume $\text{Var}(Y_i) = \mu_i$, while the true variance is larger
- $\Longrightarrow$ Confidence intervals are too narrow, p-values are too small, you reject the null too often

Consequence of Overdispersion: Standard Errors Are Wrong

What happens if we fit Poisson regression to overdispersed data?

- The **coefficient estimates** are still consistent, as long as the conditional mean ($\mu_i = e^{\beta_0 + \beta_1 x_1 + \cdots}$) is correctly specified
- But the **model-based standard errors are too small** because they assume $\text{Var}(Y_i) = \mu_i$, while the true variance is larger
- $\implies$ Confidence intervals are too narrow, p-values are too small, you reject the null too often

$\implies$ With overdispersion, Poisson regression gives you the right answer with the wrong confidence.

What Poisson Gets Right, and What It Gets Wrong

What Poisson gets right:

- Positive predictions for all covariate values (the log link)
- Coefficients are semi-elasticities, easy to interpret
- Consistent coefficient estimates (even with overdispersion)

What Poisson Gets Right, and What It Gets Wrong

What Poisson gets right:

- Positive predictions for all covariate values (the log link)
- Coefficients are semi-elasticities, easy to interpret
- Consistent coefficient estimates (even with overdispersion)

What Poisson gets wrong:

- Forces $\text{Var}(Y_i) = \mu_i$, but our data has variance 7.7$\times$ the mean
- Standard errors are too small $\implies$ false confidence
- Predicted distribution misses the spike at zero and the long tail

What Poisson Gets Right, and What It Gets Wrong

What Poisson gets right:

- Positive predictions for all covariate values (the log link)
- Coefficients are semi-elasticities, easy to interpret
- Consistent coefficient estimates (even with overdispersion)

What Poisson gets wrong:

- Forces $\text{Var}(Y_i) = \mu_i$, but our data has variance $7.7\times$ the mean
- Standard errors are too small $\implies$ false confidence
- Predicted distribution misses the spike at zero and the long tail

Can we keep the Poisson's log link but relax the variance constraint?

Outline

Our Data Has Variance 7.7× the Mean

The Poisson forces $\text{Var}(Y_i) = \mu_i$. Our data violates this dramatically:

$$\frac{\text{Sample Variance}}{\text{Sample Mean}} = \frac{43.8}{5.7} = 7.7$$

Our Data Has Variance 7.7× the Mean

The Poisson forces $\text{Var}(Y_i) = \mu_i$. Our data violates this dramatically:

$$\frac{\text{Sample Variance}}{\text{Sample Mean}} = \frac{43.8}{5.7} = 7.7$$

We want a model that:

- Keeps the **same log link**: $\ln(\mu_i) = \beta_0 + \beta_1 x_1 + \cdots$ (positive predictions, semi-elasticities)
- Adds a **free variance parameter** so the variance can exceed the mean

Our Data Has Variance 7.7× the Mean

The Poisson forces $\text{Var}(Y_i) = \mu_i$. Our data violates this dramatically:

$$\frac{\text{Sample Variance}}{\text{Sample Mean}} = \frac{43.8}{5.7} = 7.7$$

We want a model that:

- Keeps the **same log link**: $\ln(\mu_i) = \beta_0 + \beta_1 x_1 + \cdots$ (positive predictions, semi-elasticities)
- Adds a **free variance parameter** so the variance can exceed the mean

$\Longrightarrow$ The **Negative Binomial** does exactly this: it generalizes the Poisson by adding one parameter.

Adding an Overdispersion Parameter

The Poisson model forces $\text{Var}(Y_i) = \mu_i$. To allow overdispersion, we add a parameter $\alpha > 0$:

$$\text{Var}(Y_i) = \mu_i + \alpha\,\mu_i^2$$

Adding an Overdispersion Parameter

The Poisson model forces $\text{Var}(Y_i) = \mu_i$. To allow overdispersion, we add a parameter $\alpha > 0$:

$$\text{Var}(Y_i) = \mu_i + \alpha\, \mu_i^2$$

- The extra term $\alpha\, \mu_i^2$ lets the variance **exceed** the mean
- How much extra variance depends on α

Adding an Overdispersion Parameter

The Poisson model forces $\text{Var}(Y_i) = \mu_i$. To allow overdispersion, we add a parameter $\alpha > 0$:

$$\text{Var}(Y_i) = \mu_i + \alpha\,\mu_i^2$$

- The extra term $\alpha\,\mu_i^2$ lets the variance **exceed** the mean
- How much extra variance depends on α

Boundary condition: when $\alpha \to 0$, the extra term vanishes and we get $\text{Var}(Y_i) = \mu_i$. That is exactly Poisson.

Adding an Overdispersion Parameter

The Poisson model forces $\text{Var}(Y_i) = \mu_i$. To allow overdispersion, we add a parameter $\alpha > 0$:

$$\text{Var}(Y_i) = \mu_i + \alpha\,\mu_i^2$$

- The extra term $\alpha\,\mu_i^2$ lets the variance **exceed** the mean
- How much extra variance depends on α

Boundary condition: when $\alpha \to 0$, the extra term vanishes and we get $\text{Var}(Y_i) = \mu_i$. That is exactly Poisson.

$\implies$ Poisson is a special case of the Negative Binomial with $\alpha = 0$. The NB nests the Poisson.

The Negative Binomial Model

The Negative Binomial regression model specifies:

The Negative Binomial Model

The Negative Binomial regression model specifies:

1. **Same log link** as Poisson:

$$\ln(\mu_i) = \beta_0 + \beta_1 \,\text{Age}_i + \beta_2 \,\text{Insurance}_i + \beta_3 \,\text{Health}_i$$

The Negative Binomial Model

The Negative Binomial regression model specifies:

1. **Same log link** as Poisson:

$$\ln(\mu_i) = \beta_0 + \beta_1 \, \text{Age}_i + \beta_2 \, \text{Insurance}_i + \beta_3 \, \text{Health}_i$$

2. **NB probability formula** instead of Poisson. It uses a different formula to assign probabilities to each count value (software handles it)

The Negative Binomial Model

The Negative Binomial regression model specifies:

1. **Same log link** as Poisson:

$$\ln(\mu_i) = \beta_0 + \beta_1 \,\text{Age}_i + \beta_2 \,\text{Insurance}_i + \beta_3 \,\text{Health}_i$$

2. **NB probability formula** instead of Poisson. It uses a different formula to assign probabilities to each count value (software handles it)

3. **Variance:** $\text{Var}(Y_i) = \mu_i + \alpha\, \mu_i^2$, where α is estimated from the data

The Negative Binomial Model

The Negative Binomial regression model specifies:

1. **Same log link** as Poisson:

$$\ln(\mu_i) = \beta_0 + \beta_1 \,\text{Age}_i + \beta_2 \,\text{Insurance}_i + \beta_3 \,\text{Health}_i$$

2. **NB probability formula** instead of Poisson. It uses a different formula to assign probabilities to each count value (software handles it)

3. **Variance:** $\text{Var}(Y_i) = \mu_i + \alpha\, \mu_i^2$, where α is estimated from the data

We estimate $(\beta_0, \beta_1, \beta_2, \beta_3)$ and α jointly by MLE.

The Negative Binomial Model

The Negative Binomial regression model specifies:

1. **Same log link** as Poisson:

$$\ln(\mu_i) = \beta_0 + \beta_1 \, \text{Age}_i + \beta_2 \, \text{Insurance}_i + \beta_3 \, \text{Health}_i$$

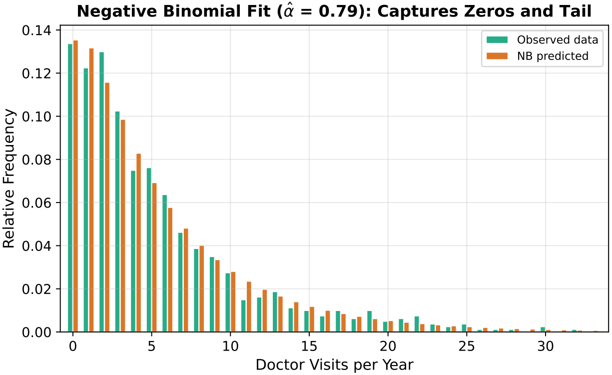
2. **NB probability formula** instead of Poisson. It uses a different formula to assign probabilities to each count value (software handles it)

3. **Variance:** $\text{Var}(Y_i) = \mu_i + \alpha \, \mu_i^2$, where α is estimated from the data

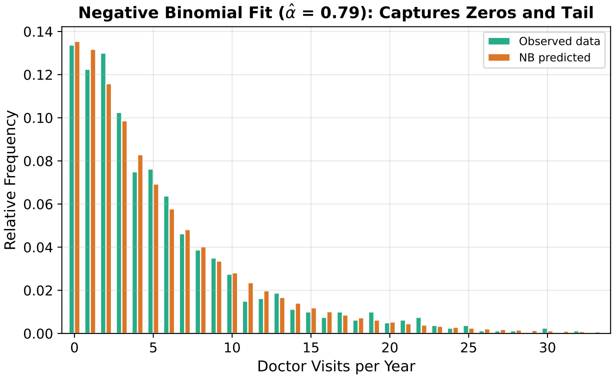
We estimate $(\beta_0, \beta_1, \beta_2, \beta_3)$ and α jointly by MLE.

$\implies$ Coefficients have the **same semi-elasticity interpretation** as Poisson. The only change is allowing more variance.

NB Fit to Our Data

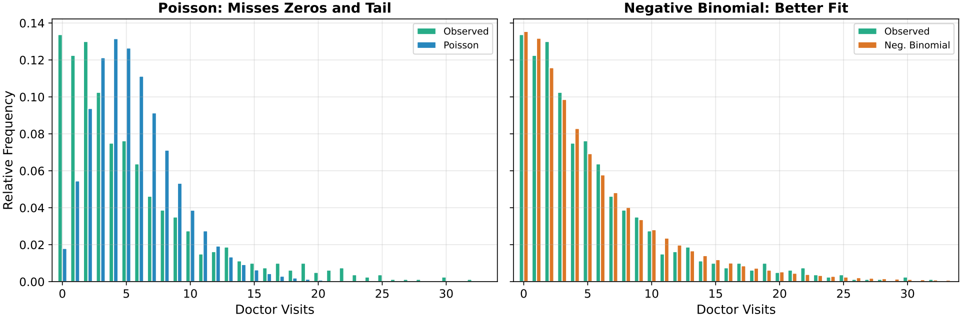


NB Fit to Our Data

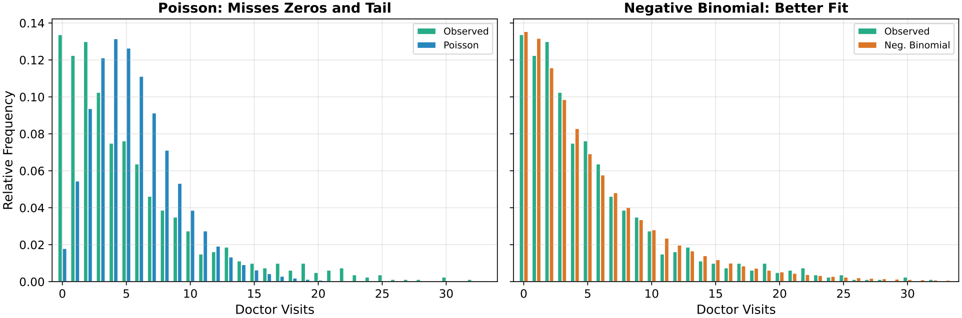


With an estimated $\hat{\alpha} = 0.79$, the Negative Binomial captures the spike at zero and the long right tail that Poisson missed.

Side-by-Side: Poisson vs. Negative Binomial



Side-by-Side: Poisson vs. Negative Binomial



The Poisson (left) squeezes too much mass into the middle. The NB (right) spreads it out to match the data.

Testing for Overdispersion

Since Poisson is nested inside NB (H_0: $\alpha = 0$), we can test directly.

Testing for Overdispersion

Since Poisson is nested inside NB (H_0: $\alpha = 0$), we can test directly.

Method 1: Cameron–Trivedi regression test.
Regress $(y_i - \hat{\mu}_i)^2 - y_i$ on $\hat{\mu}_i^2$ (no intercept). If the slope $\hat{\alpha}$ is significantly positive $\implies$ overdispersion.

Testing for Overdispersion

Since Poisson is nested inside NB (H_0: $\alpha = 0$), we can test directly.

Method 1: Cameron–Trivedi regression test.
Regress $(y_i - \hat{\mu}_i)^2 - y_i$ on $\hat{\mu}_i^2$ (no intercept). If the slope $\hat{\alpha}$ is significantly positive $\implies$ overdispersion.

Intuition: under the Poisson, $(y_i - \mu_i)^2 - y_i$ should average to zero. If it is systematically positive, there is extra variance beyond what Poisson allows.

Testing for Overdispersion

Since Poisson is nested inside NB (H_0: $\alpha = 0$), we can test directly.

Method 1: Cameron–Trivedi regression test.
Regress $(y_i - \hat{\mu}_i)^2 - y_i$ on $\hat{\mu}_i^2$ (no intercept). If the slope $\hat{\alpha}$ is significantly positive $\implies$ overdispersion.

Intuition: under the Poisson, $(y_i - \mu_i)^2 - y_i$ should average to zero. If it is systematically positive, there is extra variance beyond what Poisson allows.

Method 2: Likelihood ratio test.
$LR = 2\left[\ell_{\text{NB}} - \ell_{\text{Poisson}}\right] \sim \chi_1^2$ under H_0: $\alpha = 0$ (conservative, since $\alpha = 0$ is on the boundary of the parameter space).

Testing for Overdispersion

Since Poisson is nested inside NB (H_0: $\alpha = 0$), we can test directly.

Method 1: Cameron–Trivedi regression test.
Regress $(y_i - \hat{\mu}_i)^2 - y_i$ on $\hat{\mu}_i^2$ (no intercept). If the slope $\hat{\alpha}$ is significantly positive $\implies$ overdispersion.

Intuition: under the Poisson, $(y_i - \mu_i)^2 - y_i$ should average to zero. If it is systematically positive, there is extra variance beyond what Poisson allows.

Method 2: Likelihood ratio test.
$LR = 2\left[\ell_{\text{NB}} - \ell_{\text{Poisson}}\right] \sim \chi^2_1$ under H_0: $\alpha = 0$ (conservative, since $\alpha = 0$ is on the boundary of the parameter space).

In our data: $\hat{\alpha} = 0.79$ with $p < 0.001$.

Testing for Overdispersion

Since Poisson is nested inside NB (H_0: $\alpha = 0$), we can test directly.

Method 1: Cameron–Trivedi regression test.
Regress $(y_i - \hat{\mu}_i)^2 - y_i$ on $\hat{\mu}_i^2$ (no intercept). If the slope $\hat{\alpha}$ is significantly positive $\implies$ overdispersion.

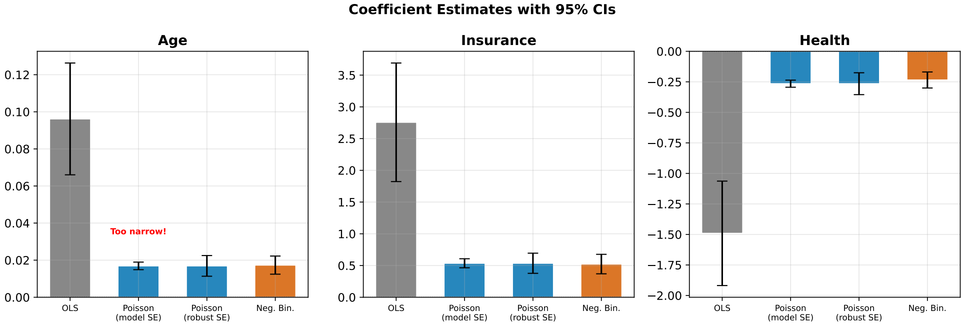
Intuition: under the Poisson, $(y_i - \mu_i)^2 - y_i$ should average to zero. If it is systematically positive, there is extra variance beyond what Poisson allows.

Method 2: Likelihood ratio test.
$LR = 2\left[\ell_{\text{NB}} - \ell_{\text{Poisson}}\right] \sim \chi^2_1$ under H_0: $\alpha = 0$ (conservative, since $\alpha = 0$ is on the boundary of the parameter space).

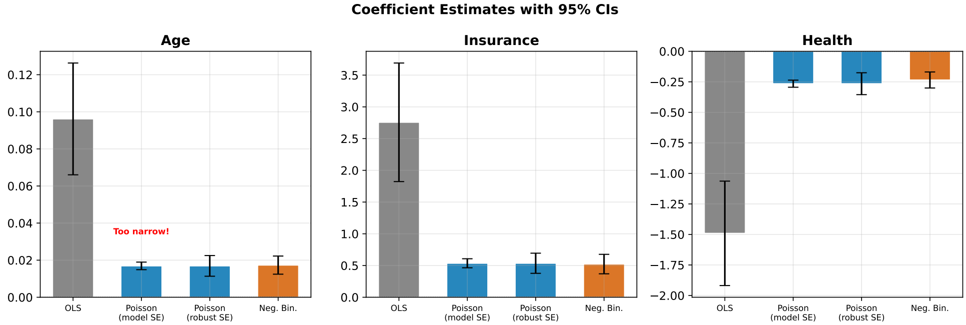
In our data: $\hat{\alpha} = 0.79$ with $p < 0.001$.

$\implies$ Strong evidence of overdispersion. The Poisson model is rejected in favor of NB.

Coefficient Estimates: OLS vs. Poisson vs. NB



Coefficient Estimates: OLS vs. Poisson vs. NB



Poisson and NB give similar coefficient estimates, but Poisson model-based SEs are far too narrow. The NB SEs properly account for overdispersion.

Why Poisson SEs Are Too Small

	Poisson (model SE)	Poisson (robust SE)	NB
Age	0.001	0.003	0.003
Insurance	0.036	0.081	0.078
Health	0.015	0.046	0.034

Why Poisson SEs Are Too Small

	Poisson (model SE)	Poisson (robust SE)	NB
Age	0.001	0.003	0.003
Insurance	0.036	0.081	0.078
Health	0.015	0.046	0.034

Poisson model SEs assume $\text{Var}(Y_i) = \mu_i$. Since the true variance is much larger, these SEs are roughly 2–3 times too small.

Why Poisson SEs Are Too Small

	Poisson (model SE)	Poisson (robust SE)	NB
Age	0.001	0.003	0.003
Insurance	0.036	0.081	0.078
Health	0.015	0.046	0.034

Poisson model SEs assume $\text{Var}(Y_i) = \mu_i$. Since the true variance is much larger, these SEs are roughly 2–3 times too small.

Two fixes:

1. **Robust (sandwich) SEs:** keep the Poisson model but correct the SEs
2. **Negative Binomial:** model the extra variance directly

Why Poisson SEs Are Too Small

	Poisson (model SE)	Poisson (robust SE)	NB
Age	0.001	0.003	0.003
Insurance	0.036	0.081	0.078
Health	0.015	0.046	0.034

Poisson model SEs assume $\text{Var}(Y_i) = \mu_i$. Since the true variance is much larger, these SEs are roughly 2–3 times too small.

Two fixes:

1. **Robust (sandwich) SEs:** keep the Poisson model but correct the SEs
2. **Negative Binomial:** model the extra variance directly

$\Longrightarrow$ Both give CIs based on a consistent estimator of the true sampling variance, so coverage is correct asymptotically (typically wider than the under-dispersion-assuming Poisson CI).

Three-Model Comparison: OLS vs. Poisson vs. NB

	OLS	**Poisson**	**Neg. Binomial**
Predicted range	$(-\infty, +\infty)$	$(0, +\infty)$	$(0, +\infty)$
Variance assumption	constant	Var $= \mu$	Var $= \mu + \alpha\mu^2$
SE reliability (model-based)	heteroskedasticity biased	too small if overdispersed	correct if α well-estimated
Coefficient interpretation	level change (Δy per unit Δx)	semi-elasticity ($\approx \%\Delta y$)	semi-elasticity ($\approx \%\Delta y$)

Three-Model Comparison: OLS vs. Poisson vs. NB

	OLS	**Poisson**	**Neg. Binomial**
Predicted range	$(-\infty, +\infty)$	$(0, +\infty)$	$(0, +\infty)$
Variance assumption	constant	Var $= \mu$	Var $= \mu + \alpha\mu^2$
SE reliability (model-based)	heteroskedasticity biased	too small if overdispersed	correct if α well-estimated
Coefficient interpretation	level change (Δy per unit Δx)	semi-elasticity ($\approx \%\Delta y$)	semi-elasticity ($\approx \%\Delta y$)

$\Longrightarrow$ Moving from OLS to Poisson solves the boundary problem; moving from Poisson to NB solves the variance problem.

Outline

Quasi-Poisson: A Quick SE Correction

Sometimes you want to keep the Poisson model structure but fix the SEs. The **Quasi-Poisson** approach:

Quasi-Poisson: A Quick SE Correction

Sometimes you want to keep the Poisson model structure but fix the SEs. The **Quasi-Poisson** approach:

- Estimates the same coefficients as Poisson MLE
- Introduces a dispersion parameter ϕ: $\text{Var}(Y_i) = \phi\,\mu_i$
- Multiplies all Poisson SEs by $\sqrt{\hat{\phi}}$, where $\hat{\phi}$ is estimated from the model residuals

Quasi-Poisson: A Quick SE Correction

Sometimes you want to keep the Poisson model structure but fix the SEs. The **Quasi-Poisson** approach:

- Estimates the same coefficients as Poisson MLE
- Introduces a dispersion parameter ϕ: $\text{Var}(Y_i) = \phi\,\mu_i$
- Multiplies all Poisson SEs by $\sqrt{\hat{\phi}}$, where $\hat{\phi}$ is estimated from the model residuals

In our data, Quasi-Poisson SEs are roughly 2–3 times larger than Poisson model SEs.

Quasi-Poisson: A Quick SE Correction

Sometimes you want to keep the Poisson model structure but fix the SEs. The **Quasi-Poisson** approach:

- Estimates the same coefficients as Poisson MLE
- Introduces a dispersion parameter ϕ: $\text{Var}(Y_i) = \phi\,\mu_i$
- Multiplies all Poisson SEs by $\sqrt{\hat{\phi}}$, where $\hat{\phi}$ is estimated from the model residuals

In our data, Quasi-Poisson SEs are roughly 2–3 times larger than Poisson model SEs.

Quasi-Poisson vs. robust SEs: Quasi-Poisson assumes $\text{Var} = \phi\mu$ (overdispersion is a linear scaling of the mean). Robust SEs make no assumption about the variance form.

Quasi-Poisson: A Quick SE Correction

Sometimes you want to keep the Poisson model structure but fix the SEs. The **Quasi-Poisson** approach:

- Estimates the same coefficients as Poisson MLE
- Introduces a dispersion parameter ϕ: $\text{Var}(Y_i) = \phi\,\mu_i$
- Multiplies all Poisson SEs by $\sqrt{\hat{\phi}}$, where $\hat{\phi}$ is estimated from the model residuals

In our data, Quasi-Poisson SEs are roughly 2–3 times larger than Poisson model SEs.

Quasi-Poisson vs. robust SEs: Quasi-Poisson assumes $\text{Var} = \phi\mu$ (overdispersion is a linear scaling of the mean). Robust SEs make no assumption about the variance form.

Approach	Variance structure	When to use
Poisson	$\text{Var} = \mu$	Mild or no overdispersion
Quasi-Poisson	$\text{Var} = \phi\,\mu$	Quick SE correction; no full likelihood
Neg. Binomial	$\text{Var} = \mu + \alpha\mu^2$	Full model; predictions, LR tests, AIC

Excess Zeros: When to Consider Zero-Inflated Models

Sometimes overdispersion comes from **excess zeros**: more zeros than even the NB can accommodate.

Excess Zeros: When to Consider Zero-Inflated Models

Sometimes overdispersion comes from **excess zeros**: more zeros than even the NB can accommodate.

Example: doctor visits. Some people *never* go (they avoid doctors entirely), while others go based on their health needs. Two different processes generate the zeros.

Excess Zeros: When to Consider Zero-Inflated Models

Sometimes overdispersion comes from **excess zeros**: more zeros than even the NB can accommodate.

Example: doctor visits. Some people *never* go (they avoid doctors entirely), while others go based on their health needs. Two different processes generate the zeros.

Zero-inflated models combine:

1. A binary model (logit) for whether someone is a "certain zero" vs. a potential visitor
2. A count model (Poisson or NB) for potential visitors

Excess Zeros: When to Consider Zero-Inflated Models

Sometimes overdispersion comes from **excess zeros**: more zeros than even the NB can accommodate.

Example: doctor visits. Some people *never* go (they avoid doctors entirely), while others go based on their health needs. Two different processes generate the zeros.

Zero-inflated models combine:

1. A binary model (logit) for whether someone is a “certain zero” vs. a potential visitor
2. A count model (Poisson or NB) for potential visitors

How to tell if you need one:

- Compare observed zero proportion to the predicted zero proportion from your NB model
- If NB already fits the zeros well, zero-inflation is unnecessary

Excess Zeros: When to Consider Zero-Inflated Models

Sometimes overdispersion comes from **excess zeros**: more zeros than even the NB can accommodate.

Example: doctor visits. Some people *never* go (they avoid doctors entirely), while others go based on their health needs. Two different processes generate the zeros.

Zero-inflated models combine:

1. A binary model (logit) for whether someone is a "certain zero" vs. a potential visitor
2. A count model (Poisson or NB) for potential visitors

How to tell if you need one:

- Compare observed zero proportion to the predicted zero proportion from your NB model
- If NB already fits the zeros well, zero-inflation is unnecessary

$\Longrightarrow$ In our data, NB captures the 13% zeros adequately. Zero-inflation would be needed if, say, 40% of the sample had zero visits.

Decision Framework: Which Count Model to Use

1. **Start with Poisson.** It is the simplest count model and gives consistent coefficient estimates even under overdispersion

Decision Framework: Which Count Model to Use

1. **Start with Poisson.** It is the simplest count model and gives consistent coefficient estimates even under overdispersion
2. **Test for overdispersion.** Cameron–Trivedi test or LR test (H_0: $\alpha = 0$)

Decision Framework: Which Count Model to Use

1. **Start with Poisson.** It is the simplest count model and gives consistent coefficient estimates even under overdispersion
2. **Test for overdispersion.** Cameron–Trivedi test or LR test (H_0: $\alpha = 0$)
3. **If overdispersion is detected:**
   - **Minimum fix:** use robust (sandwich) SEs with the Poisson model
   - **Better fix:** switch to Negative Binomial regression

Decision Framework: Which Count Model to Use

1. **Start with Poisson.** It is the simplest count model and gives consistent coefficient estimates even under overdispersion
2. **Test for overdispersion.** Cameron–Trivedi test or LR test (H_0: $\alpha = 0$)
3. **If overdispersion is detected:**
   - **Minimum fix:** use robust (sandwich) SEs with the Poisson model
   - **Better fix:** switch to Negative Binomial regression
4. **If excess zeros remain:** consider a zero-inflated Poisson (ZIP) or zero-inflated NB (ZINB)

Decision Framework: Which Count Model to Use

1. **Start with Poisson.** It is the simplest count model and gives consistent coefficient estimates even under overdispersion
2. **Test for overdispersion.** Cameron–Trivedi test or LR test (H_0: $\alpha = 0$)
3. **If overdispersion is detected:**
   - **Minimum fix:** use robust (sandwich) SEs with the Poisson model
   - **Better fix:** switch to Negative Binomial regression
4. **If excess zeros remain:** consider a zero-inflated Poisson (ZIP) or zero-inflated NB (ZINB)
5. **If the outcome has a known upper bound** (e.g., number correct out of 10):

   $\Longrightarrow$ This is not a count model problem; consider binomial regression instead

Summary: Back to Doctor Visits

1. **OLS on counts fails:** it predicted negative visits for young, healthy, uninsured individuals

Summary: Back to Doctor Visits

1. **OLS on counts fails:** it predicted negative visits for young, healthy, uninsured individuals
2. **Poisson regression** uses a log link ($\ln \mu_i = \beta_0 + \beta_1 x_1 + \cdots$) to guarantee positive predictions. Coefficients are semi-elasticities

Summary: Back to Doctor Visits

1. **OLS on counts fails:** it predicted negative visits for young, healthy, uninsured individuals
2. **Poisson regression** uses a log link ($\ln \mu_i = \beta_0 + \beta_1 x_1 + \cdots$) to guarantee positive predictions. Coefficients are semi-elasticities
3. **Equidispersion** ($\text{Var} = \mu$) almost never holds in practice. Our doctor visits data had variance $7.7\times$ the mean, so Poisson SEs were 2–3$\times$ too small

Summary: Back to Doctor Visits

1. **OLS on counts fails:** it predicted negative visits for young, healthy, uninsured individuals
2. **Poisson regression** uses a log link ($\ln \mu_i = \beta_0 + \beta_1 x_1 + \cdots$) to guarantee positive predictions. Coefficients are semi-elasticities
3. **Equidispersion** ($\text{Var} = \mu$) almost never holds in practice. Our doctor visits data had variance $7.7\times$ the mean, so Poisson SEs were 2–3$\times$ too small
4. **Negative Binomial** adds one parameter (α) that allows $\text{Var} = \mu + \alpha\mu^2$. It captured the spike at zero and the long tail that Poisson missed

Summary: Back to Doctor Visits

1. **OLS on counts fails:** it predicted negative visits for young, healthy, uninsured individuals
2. **Poisson regression** uses a log link ($\ln \mu_i = \beta_0 + \beta_1 x_1 + \cdots$) to guarantee positive predictions. Coefficients are semi-elasticities
3. **Equidispersion** ($\text{Var} = \mu$) almost never holds in practice. Our doctor visits data had variance $7.7\times$ the mean, so Poisson SEs were 2–$3\times$ too small
4. **Negative Binomial** adds one parameter (α) that allows $\text{Var} = \mu + \alpha\mu^2$. It captured the spike at zero and the long tail that Poisson missed
5. **Test for overdispersion** before reporting Poisson results. Use the Cameron–Trivedi test or a likelihood ratio test

Summary: Back to Doctor Visits

1. **OLS on counts fails:** it predicted negative visits for young, healthy, uninsured individuals
2. **Poisson regression** uses a log link ($\ln \mu_i = \beta_0 + \beta_1 x_1 + \cdots$) to guarantee positive predictions. Coefficients are semi-elasticities
3. **Equidispersion** ($\text{Var} = \mu$) almost never holds in practice. Our doctor visits data had variance $7.7\times$ the mean, so Poisson SEs were 2–$3\times$ too small
4. **Negative Binomial** adds one parameter (α) that allows $\text{Var} = \mu + \alpha\mu^2$. It captured the spike at zero and the long tail that Poisson missed
5. **Test for overdispersion** before reporting Poisson results. Use the Cameron–Trivedi test or a likelihood ratio test
6. **Zero-inflated models** are a further extension when excess zeros come from a separate process

Summary: Back to Doctor Visits

1. **OLS on counts fails:** it predicted negative visits for young, healthy, uninsured individuals
2. **Poisson regression** uses a log link ($\ln \mu_i = \beta_0 + \beta_1 x_1 + \cdots$) to guarantee positive predictions. Coefficients are semi-elasticities
3. **Equidispersion** ($\text{Var} = \mu$) almost never holds in practice. Our doctor visits data had variance $7.7\times$ the mean, so Poisson SEs were 2–$3\times$ too small
4. **Negative Binomial** adds one parameter (α) that allows $\text{Var} = \mu + \alpha\mu^2$. It captured the spike at zero and the long tail that Poisson missed
5. **Test for overdispersion** before reporting Poisson results. Use the Cameron–Trivedi test or a likelihood ratio test
6. **Zero-inflated models** are a further extension when excess zeros come from a separate process

$\Longrightarrow$ Always start with Poisson, test for overdispersion, and upgrade to NB or robust SEs as needed.

Thank you!

jakeanderson@g.ucla.edu