

# Generalized Linear Model

PSYC 573

---

University of Southern California

March 22, 2022

# Regression for Prediction

One outcome  $Y$ , one or more predictors  $X_1, X_2, \dots$

E.g.,

- What will a student's college GPA be given an SAT score of  $x$  ?
- How long will a person live if the person adopts diet  $x$ ?
- What will the earth's global temperature be if the carbon emission level is  $x$ ?

# Keep These in Mind

1. Likelihood function is defined for the outcome  $Y$
2. Prediction is probabilistic (i.e., uncertain) and contains error

# Generalized Linear Models (GLM)

# GLM

Three components:

- Conditional distribution of  $Y$
- Link function
- Linear predictor

# Some Examples

Outcome type	Support	Distributions	Link
continuous	$[-\infty, \infty]$	Normal	Identity
count (fixed duration)	$\{0, 1, \dots\}$	Poisson	Log
count (known # of trials)	$\{0, 1, \dots, N\}$	Binomial	Logit
binary	$\{0, 1\}$	Bernoulli	Logit
ordinal	$\{0, 1, \dots, K\}$	categorical	Logit
nominal	$K$ -vector of $\{0, 1\}$	categorical	Logit
multinomial	$K$ -vector of $\{0, 1, \dots, K\}$	categorical	Logit

# Mathematical Form (One Predictor)

$$Y_i \sim \text{Dist}(\mu_i, \tau)$$

$$g(\mu_i) = \eta_i$$

$$\eta_i = \beta_0 + \beta_1 X_i$$

- **Dist**: conditional distribution of  $Y \mid X$  (e.g., normal, Bernoulli, . . .)
  - I.e., distribution of **prediction error**; not the marginal distribution of  $Y$
- $\mu_i$ : mean parameter for the  $i$ th observation
- $\eta_i$ : linear predictor
- $g(\cdot)$ : link function
- ( $\tau$ : dispersion parameter)

# Illustration

Next few slides contain example GLMs, with the same predictor  $X$

```
num_obs ← 100  
x ← runif(num_obs, min = 1, max = 5) # uniform x  
beta0 ← 0.2; beta1 ← 0.5
```



# Normal, Identity Link

aka linear regression

---

Model

Simulation

$$Y_i \sim N(\mu_i, \sigma)$$

$$\mu_i = \eta_i$$

$$\eta_i = \beta_0 + \beta_1 X_i$$

# Poisson, Log Link

aka poisson regression

---

Model

Simulation

$$Y_i \sim \text{Pois}(\mu_i)$$

$$\log(\mu_i) = \eta_i$$

$$\eta_i = \beta_0 + \beta_1 X_i$$

# Bernoulli, Logit Link

aka binary logistic regression

---

Model

Simulation

$$Y_i \sim \text{Bern}(\mu_i)$$

$$\log\left(\frac{\mu_i}{1 - \mu_i}\right) = \eta_i$$

$$\eta_i = \beta_0 + \beta_1 X_i$$

# Binomial, Logit Link

aka binomial logistic regression

---

Model

Simulation

$$Y_i \sim \text{Bin}(N, \mu_i)$$

$$\log\left(\frac{\mu_i}{1 - \mu_i}\right) = \eta_i$$

$$\eta_i = \beta_0 + \beta_1 X_i$$

# Remarks

Different link functions can be used

- E.g., identity link or probit link for Bernoulli variables

Linearity is a strong assumption

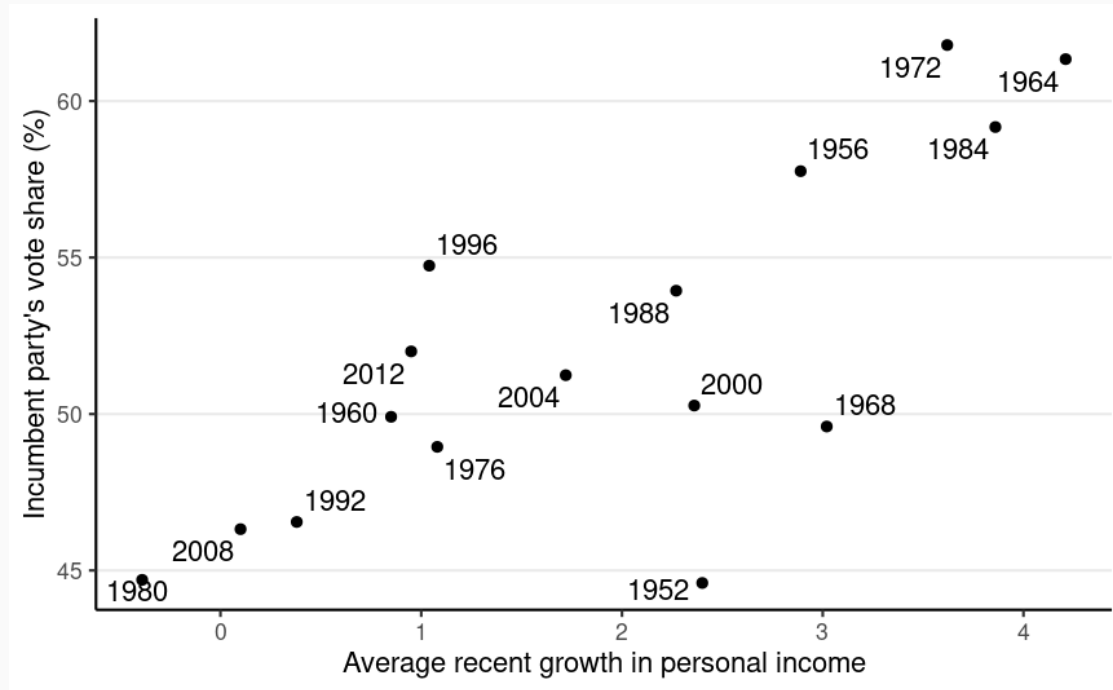
- GLM can allow  $\boldsymbol{\eta}$  and  $\boldsymbol{X}$  to be nonlinearly related, as long as it's linear in the coefficients
  - E.g.,  $\eta_i = \beta_0 + \beta_1 \log(X_i)$
  - E.g.,  $\eta_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2$
  - But not something like  $\eta_i = \beta_0 \log(\beta_1 + x_i)$

# Linear Regression

Many relations can be approximated as linear

But many relations cannot be approximated as linear

# Example: "Bread and peace" model





# Linear Regression Model

Model:

$$\text{vote}_i \sim N(\mu_i, \sigma)$$
$$\mu_i = \beta_0 + \beta_1 \text{growth}_i$$

$\sigma$ : SD (margin) of prediction error

Prior:

$$\beta_0 \sim N(45, 10)$$
$$\beta_1 \sim N(0, 10)$$
$$\sigma \sim t_4^+(0, 5)$$

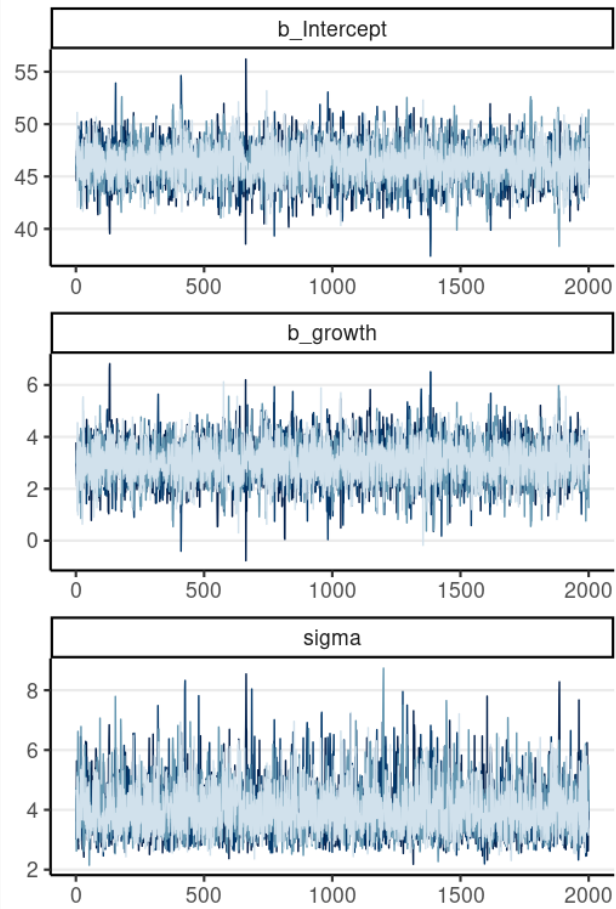
---

Stan

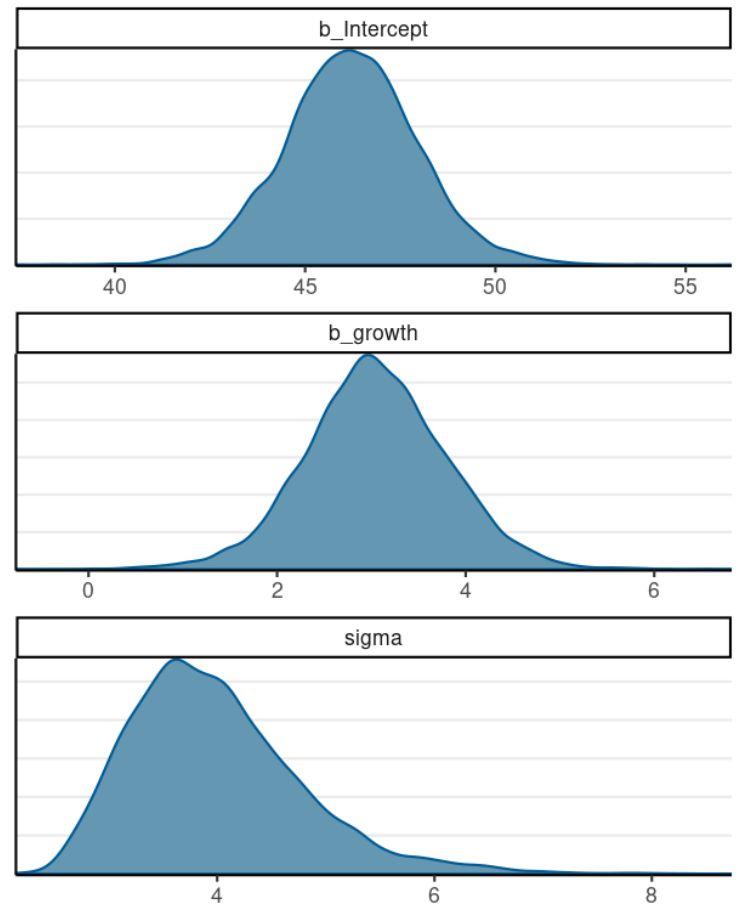
brms

brms results

```
data {  
  int<lower=0> N; // number of observations  
  vector[N] y; // outcome;  
  vector[N] x; // predictor;  
}  
parameters {  
  real beta0; // regression intercept  
  real beta1; // regression coefficient  
  real<lower=0> sigma; // SD of prediction error  
}  
model {  
  // model  
  y ~ normal(beta0 + beta1 * x, sigma);  
  // prior  
  beta0 ~ normal(45, 10);  
  beta1 ~ normal(0, 10);  
  sigma ~ student_t(4, 0, 5);  
}  
generated quantities {  
  vector[N] y_rep; // place holder  
  for (n in 1:N)  
    y_rep[n] = normal_rng(beta0 + beta1 * x[n], sigma);  
}
```



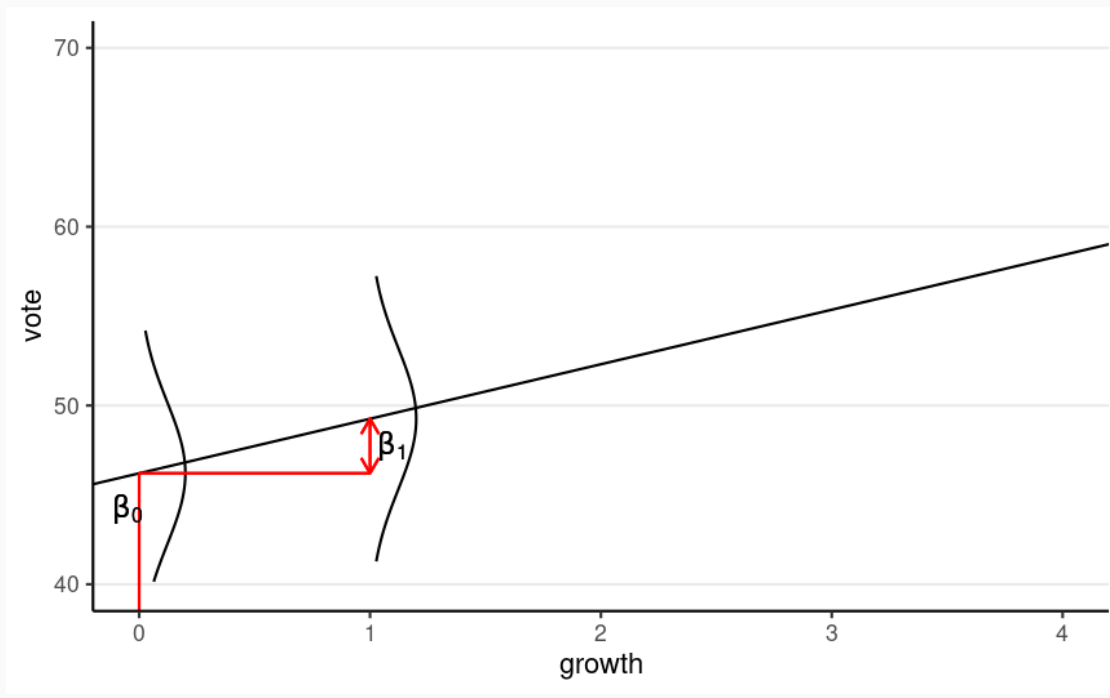
Chain  
— 1  
— 2  
— 3  
— 4



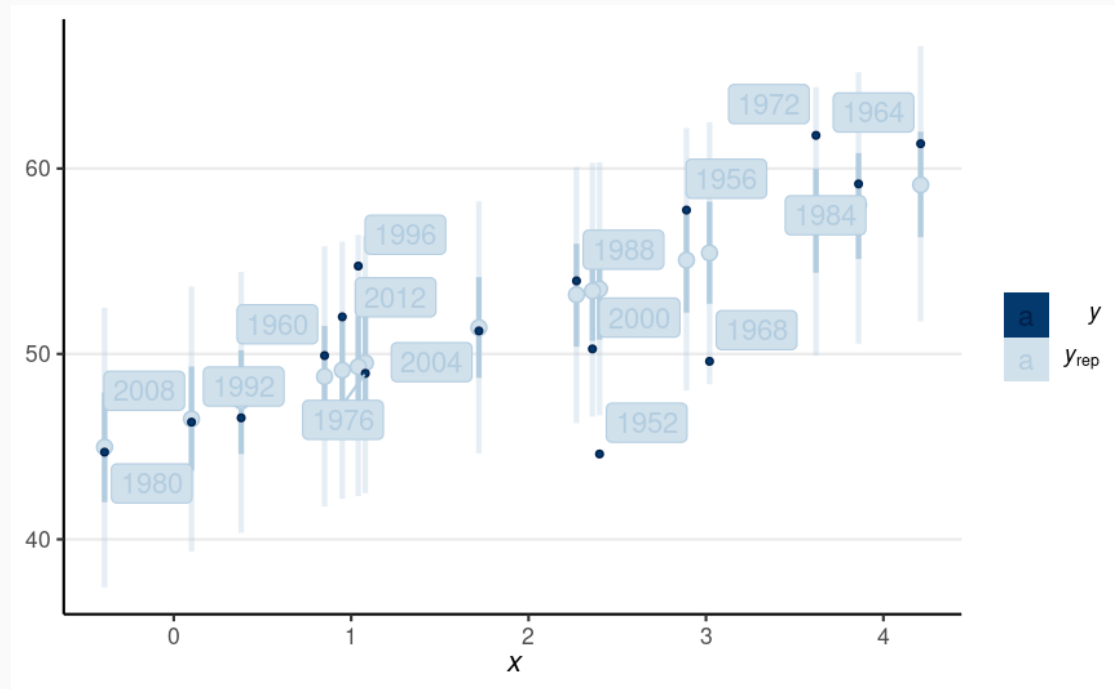
# Meaning of Coefficients

When growth = 0,  $\text{vote} \sim N(\beta_0, \sigma)$

When growth = 1,  $\text{vote} \sim N(\beta_0 + \beta_1, \sigma)$



# Posterior Predictive Check

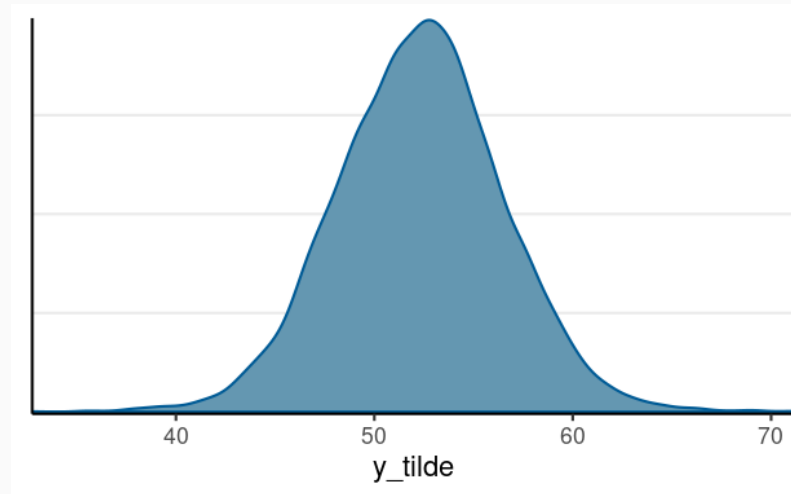


The model fits a majority of the data, but not everyone. The biggest discrepancy is 1952.

# Prediction

Predicted vote share when growth = 2:  $\tilde{y} \mid y \sim N(\beta_0 + \beta_1 \times 2, \sigma)$

```
pp_growth_eq_2 <- posterior_predict(m1_brm,  
  newdata = list(growth = 2)  
)
```



Probability of incumbent's vote share > 50% = 0.713

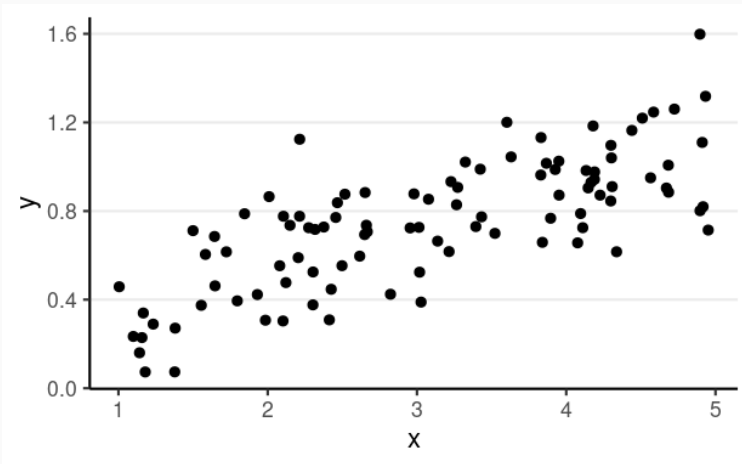
# Table

```
library(modelsummary)
msummary(m1_brm, estimate = "{estimate} [{conf.low}, {conf.high}]",
         statistic = NULL, fmt = 2)
```

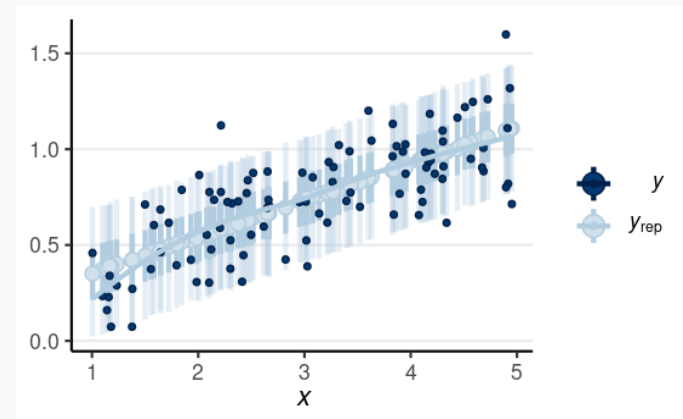
	<b>Model 1</b>
b_Intercept	46.20 [42.76, 49.75]
b_growth	3.03 [1.56, 4.56]
sigma	3.88 [2.56, 5.51]
Num.Obs.	16
ELPD	-46.1
ELPD s.e.	3.6
LOOIC	92.3
LOOIC s.e.	7.2
WAIC	92.1
RMSE	24.97

# Diagnostics

## Linearity (functional form)

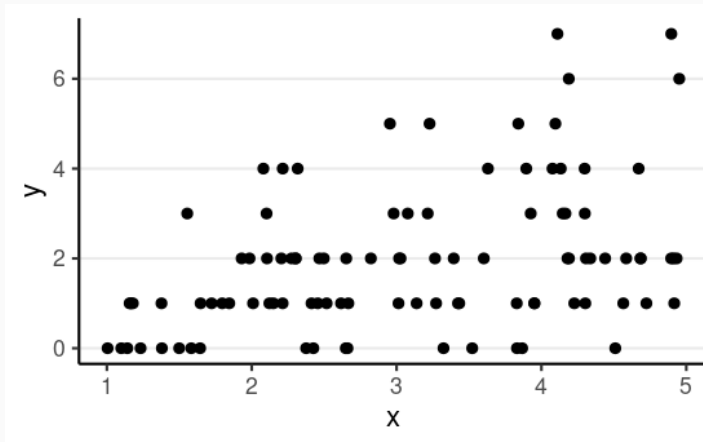


```
pp_check(m_lin, type = "intervals", x = "  
geom_smooth(aes(x = x, y = y), se = FAL
```

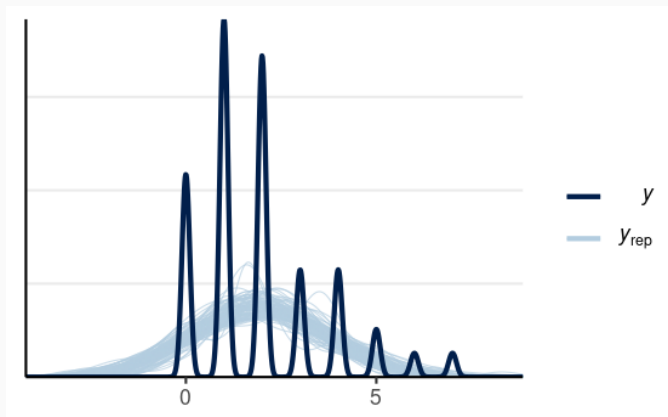




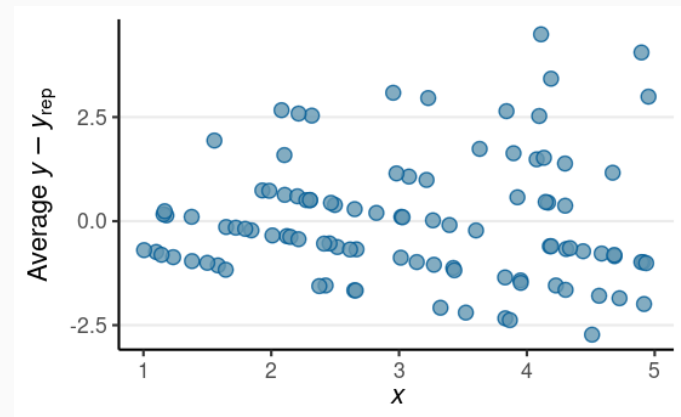
# Residual Plots



```
pp_check(m_lin_norm, ndraws = 100, bw = "
```



```
pp_check(m_lin_norm, type = "error_scatter"
```



# Prediction vs. Explanation

Is personal income growth a reason a candidate/party got more vote share?

If so, what is the mechanism?

If not, what is responsible for the association?

# Additional Notes

Outlier: use  $Y_i \sim t_\nu(\mu_i, \sigma)$

Nonconstant  $\sigma$

- One option is  $\log(\sigma_i) = \beta_0^s + \beta_1^s X_i$

Check whether linearity holds

- Other options: splines, quadratic, log transform (i.e., lognormal model), etc

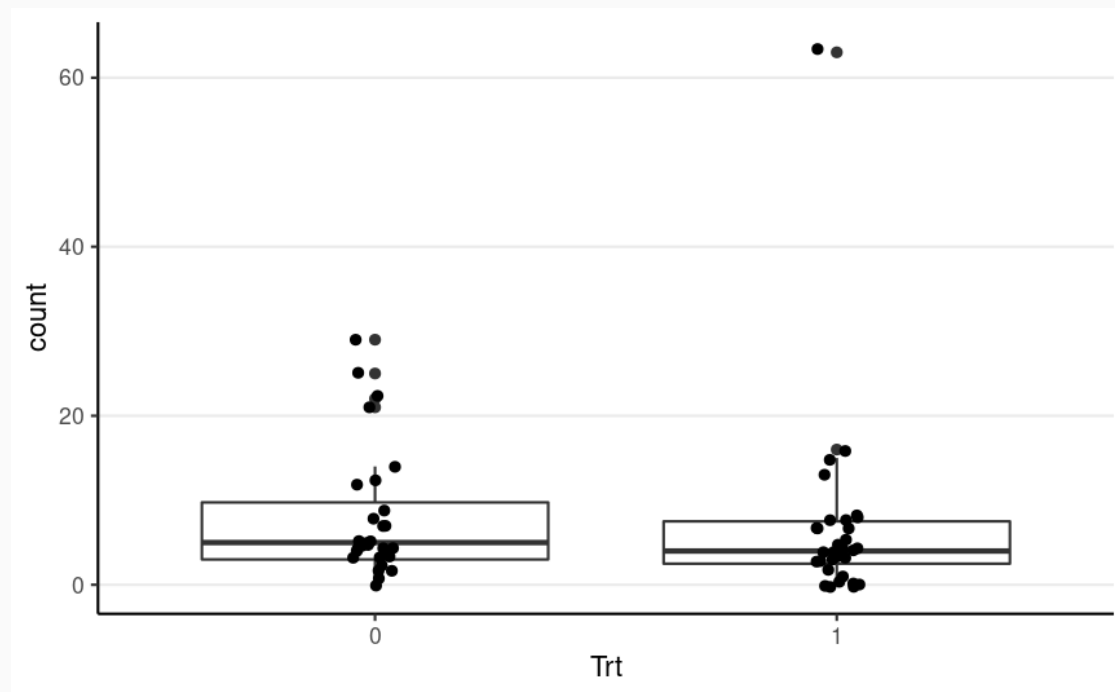
# Poisson Regression

- **count**: The seizure count between two visits
- **Trt**: Either 0 or 1 indicating if the patient received anticonvulsant therapy

$$\text{count}_i \sim \text{Pois}(\mu_i)$$

$$\log(\mu_i) = \eta_i$$

$$\eta_i = \beta_0 + \beta_1 \text{Trt}_i$$



## Poisson with log link

Predicted seizure rate =  $\exp(\beta_0 + \beta_1) = \exp(\beta_0) \exp(\beta_1)$   
for Trt = 1;  $\exp(\beta_0)$  for Trt = 0

$\beta_1$  = mean difference in **log** rate of seizure;  $\exp(\beta_1)$  = ratio in rate of seizure

```
m2 <- brm(count ~ Trt, data = epilepsy4,  
          family = poisson(link = "log"))
```

```
># Family: poisson  
># Links: mu = log  
># Formula: count ~ Trt  
># Data: epilepsy4 (Number of observations: 59)  
># Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;  
># total post-warmup draws = 4000  
>#  
># Population-Level Effects:  
># Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS  
># Intercept 2.07 0.07 1.95 2.20 1.00 3759 2833  
># Trt1 -0.17 0.10 -0.36 0.01 1.00 3082 2363  
>#  
># Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS  
># and Tail_ESS are effective sample size measures, and Rhat is the potential  
># scale reduction factor on split chains (at convergence, Rhat = 1).
```

## Poisson with identity link

In this case, with one binary predictor, the link does not matter to the fit

$$\text{count}_i \sim \text{Pois}(\mu_i)$$

$$\mu_i = \eta_i$$

$$\eta_i = \beta_0 + \beta_1 \text{Trt}_i$$

$\beta_1$  = mean difference in the rate of seizure in two weeks

```
m3 ← brm(count ~ Trt, data = epilepsy4,  
          family = poisson(link = "identity"))
```



	<b>log link</b>	<b>identity link</b>
b_Intercept	2.07	7.97
	[1.95, 2.20]	[6.94, 8.96]
b_Tr1	-0.17	-1.25
	[-0.35, 0.02]	[-2.58, 0.16]
Num.Obs.	59	59
ELPD	-343.1	-345.1
ELPD s.e.	93.8	95.7
LOOIC	686.2	690.2
LOOIC s.e.	187.7	191.3
WAIC	688.5	687.8
RMSE	10.50	10.53