Agenda

1. ❤ Happy Valentine's day! ❤
2. IMPORTANT correction from Tuesday's code
3. Interpreting coefficients from Poisson regressions
4. Over-dispersed Poisson regressions
5. Zero-inflated Poisson regressions
6. Over-dispersion and zero-inflation in R

Correction

All three models specified in Tuesday's lab omitted the link function (log)

Please use *corrected* lab 0212.html (online now)

In [8]:
```
# build the model
# (NOTE this does not make any reference to our data `d`)
m1 <- alist(
    hours_games ~ dpois(lambda),
    lambda <- alpha,
    alpha ~ dnorm(3,1)
)
```

In [8]:
```
# build the model
# (NOTE this does not make any reference to our data `d`)
m1 <- alist(
    hours_games ~ dpois(lambda),
    log(lambda) <- alpha,
    alpha ~ dnorm(3,1)
)
```

Interpreting coefficients

Model from last time

$$H_i \sim \text{Pois}(\lambda_i)$$
$$\log(\lambda_i) = \alpha + \beta_M M_i + \beta_G G_i$$

$$\alpha \sim \text{Norm}(3, 1)$$
$$\beta_M \sim \text{Norm}(0, 0.5)$$
$$\beta_G \sim \text{Norm}(0, 0.3)$$

H_i
Hours of video games played last week

M_i
Student gender (indicator for boys)

G_i
Student's grade, centered at 10th

	Mean	*90% credible interval*		exp(Mean)
α	0.27	0.25	0.30	1.32
β_M	1.11	1.09	1.14	3.05
β_G	-0.14	-0.15	-0.13	0.87

Interpreting coefficients

Model from last time

$$H_i \sim \text{Pois}(\lambda_i)$$
$$\log(\lambda_i) = \alpha + \beta_M M_i + \beta_G G_i$$

$$\alpha \sim \text{Norm}(3, 1)$$
$$\beta_M \sim \text{Norm}(0, 0.5)$$
$$\beta_G \sim \text{Norm}(0, 0.3)$$

	Mean	exp(Mean)
α	0.27	1.32
β_M	1.11	3.05
β_G	-0.14	0.87

α (baseline)
A student who is not a boy (M_i=0) and is in grade 10 (G_i=0) is predicted to play about 1.32 hours of games per week.

β_M (gender)
Boys (M_i=1) are expected to spend about 3.05 times more time than non-boys (M_i=0) playing games.

β_G (grade)
A one-year increase in grade is associated with playing 0.87 times as much. This is a decrease of 13% each year.

Over-dispersion

$$H_i \sim \text{Pois}(\lambda_i)$$

$$\log(\lambda_i) = \alpha + \beta M_i$$

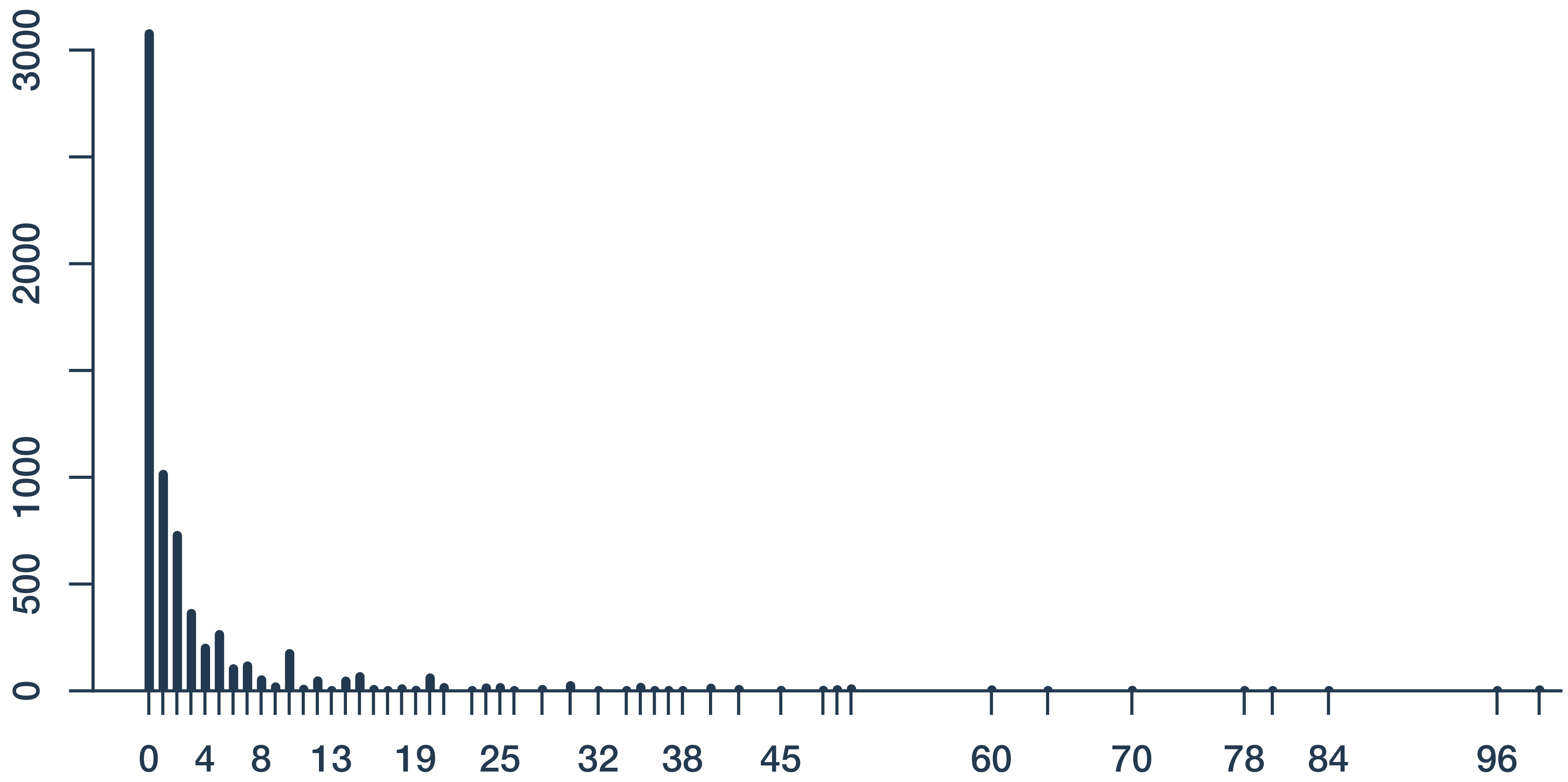
$$\alpha \sim \text{Norm}(3, 0.5)$$

$$\beta \sim \text{Norm}(0, 0.3)$$

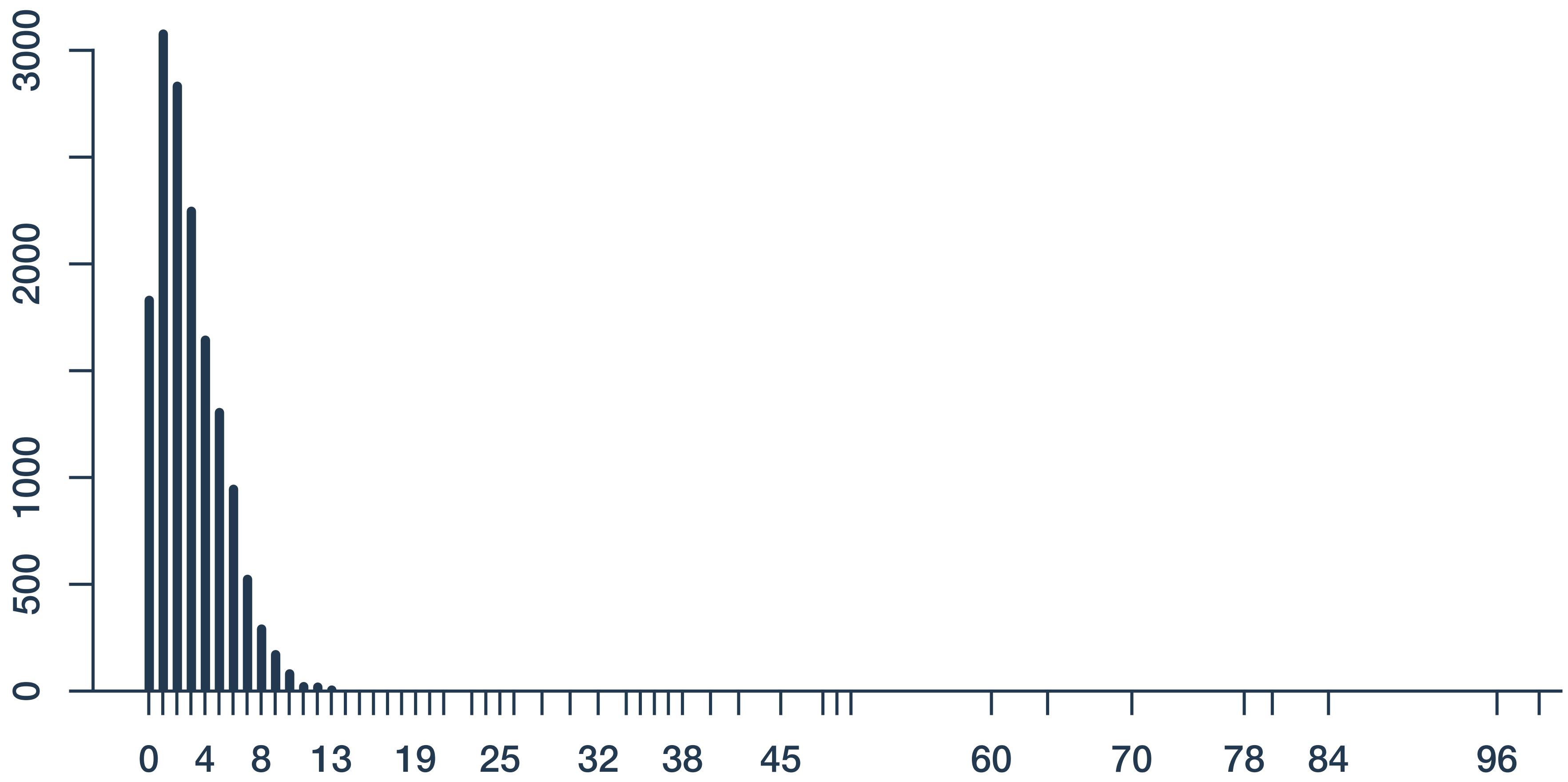
	Mean	exp(Mean)
α	0.27	1.32
β	1.11	3.05

Over-dispersion

Actual distribution



Posterior predicted distribution



Over-dispersion

Poisson regression

$$H_i \sim \text{Pois}(\lambda_i)$$

$$\log(\lambda_i) = \alpha + \beta M_i$$

$$\alpha \sim \text{Norm}(3, 0.5)$$

$$\beta \sim \text{Norm}(0, 0.3)$$

Gamma-Poisson regression

$$H_i \sim \text{Pois}(\lambda_i)$$

$$\log(\lambda_i) \sim \text{Gamma}(\mu_i, \theta)$$

$$\mu_i = \alpha + \beta M_i$$

$$\alpha \sim \text{Norm}(3, 0.5)$$

$$\beta \sim \text{Norm}(0, 0.3)$$

$$\theta \sim \text{Unif}(0, 10)$$

Extra "dispersion" from gamma

Two students who look identical based on covariates can have different Poisson rates λ_i.

One more prior

A.K.A.

Negative-binomial regression

Over-dispersed Poisson regression

Over-dispersion

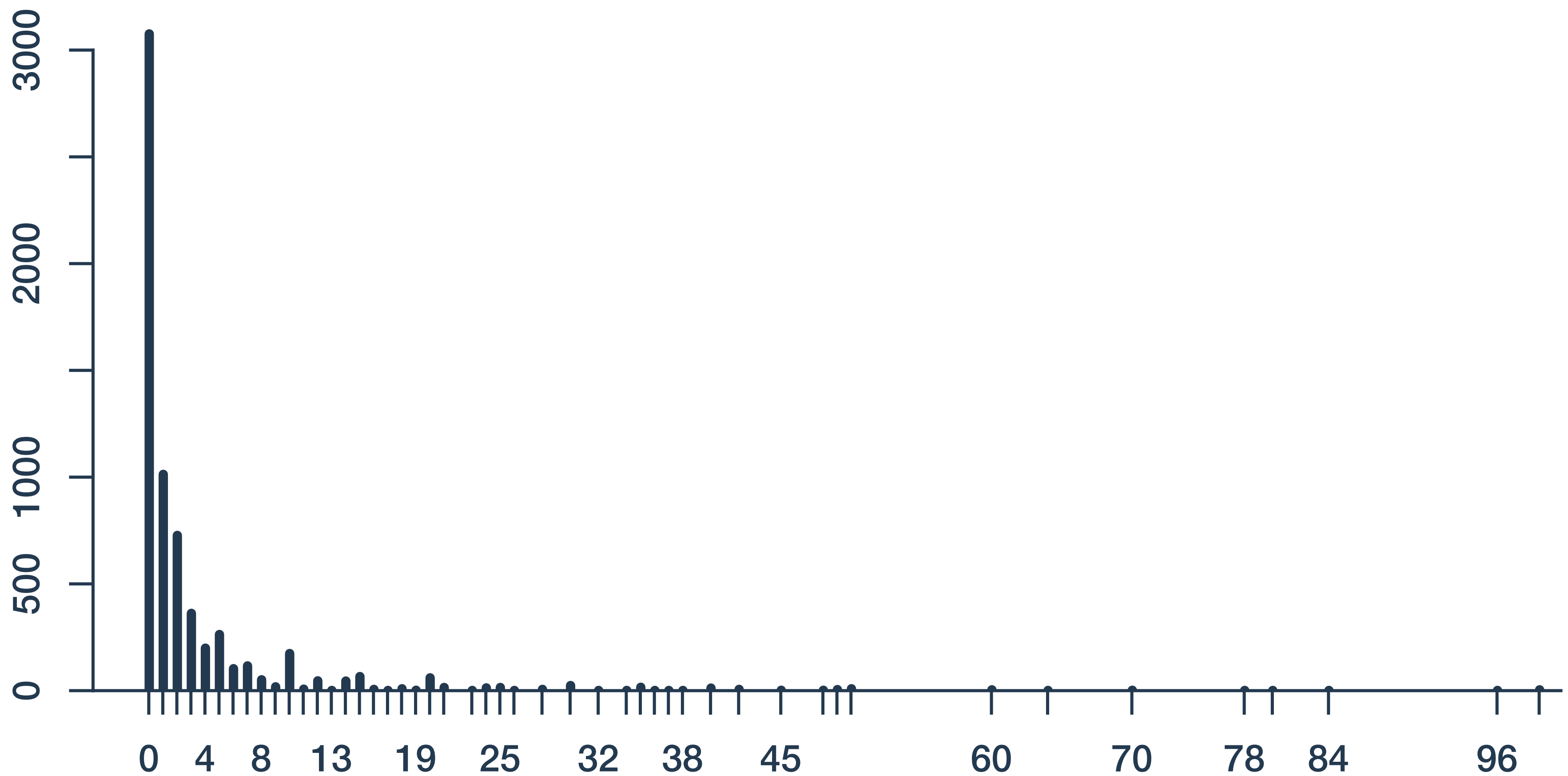
$$
\begin{aligned}
H_i &\sim \text{GammaPois}(\lambda_i, \theta) \\
\log(\lambda_i) &= \alpha + \beta M_i \\
\\
\alpha &\sim \text{Norm}(3, 0.5) \\
\beta &\sim \text{Norm}(0, 0.3) \\
\theta &\sim \text{Unif}(0, 10)
\end{aligned}
$$

	Mean	*90% credible interval*		exp(Mean)
α	0.55	0.50	0.61	1.74
β	0.86	0.80	0.91	2.35
θ	8.56	8.09	9.03	—

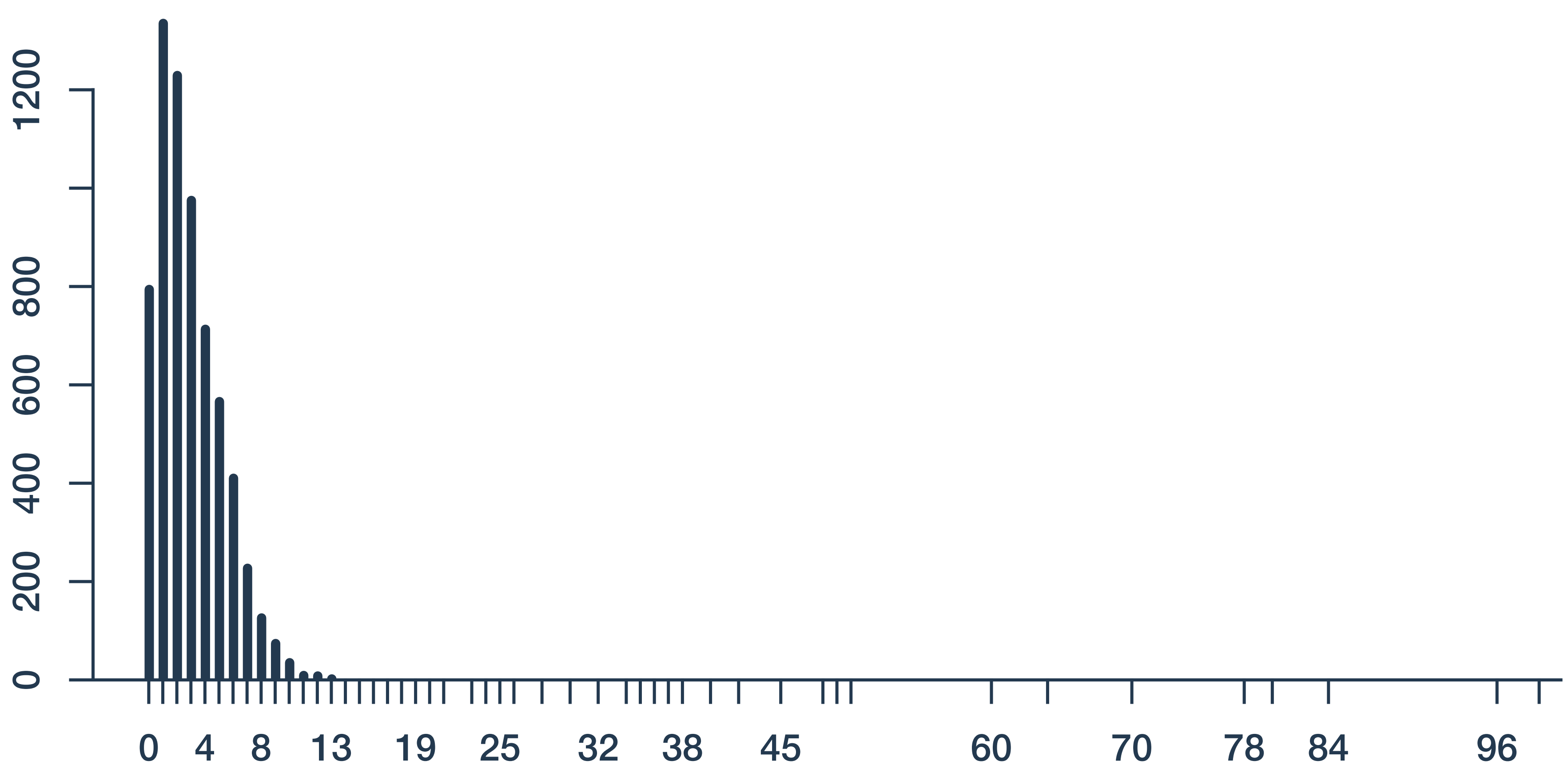
θ measures extra dispersion →

Zero-inflation

Actual distribution



Posterior predictive distribution



Zero-inflation

Outcome variable is result of one of two processes

Either the student does not own a game console ($c_i = 1$) **or** the student does own a console and plays at some rate λ_i ($c_i = 0$).

$$H_i \begin{cases} = 0 & \text{if } c_i = 1 \\ \sim \text{Pois}(\lambda_i) & \text{if } c_i = 0 \end{cases}$$

Zero-inflation

Their chance of owning a console is modelled with p_i

$$H_i \begin{cases} = 0 & \text{if } c_i = 1 \\ \sim \text{Pois}(\lambda_i) & \text{if } c_i = 0 \end{cases}$$

$$c_i \sim \text{Bern}(p_i)$$

Zero-inflation

p_i is modelled as a linear function of family income

$$H_i \begin{cases} = 0 & \text{if } c_i = 1 \\ \sim \text{Pois}(\lambda_i) & \text{if } c_i = 0 \end{cases}$$

$$c_i \sim \text{Bern}(p_i)$$

$$\text{logit}(p_i) = a_p + \beta_p W_i$$

Zero-inflation

λ_i is modelled as a linear function of gender

$$H_i \begin{cases} = 0 & \text{if } c_i = 1 \\ \sim \text{Pois}(\lambda_i) & \text{if } c_i = 0 \end{cases}$$

$$c_i \sim \text{Bern}(p_i)$$

$$\text{logit}(p_i) = \alpha_p + \beta_p W_i$$
$$\log(\lambda_i) = \alpha_\lambda + \beta_\lambda M_i$$

Zero-inflation

All four parameters need priors

$$H_i \begin{cases} = 0 & \text{if } c_i = 1 \\ \sim \text{Pois}(\lambda_i) & \text{if } c_i = 0 \end{cases}$$

$$c_i \sim \text{Bern}(p_i)$$

$$\text{logit}(p_i) = \alpha_p + \beta_p W_i$$

$$\log(\lambda_i) = \alpha_\lambda + \beta_\lambda M_i$$

$$\alpha_p \sim \text{Norm}(0, 1)$$

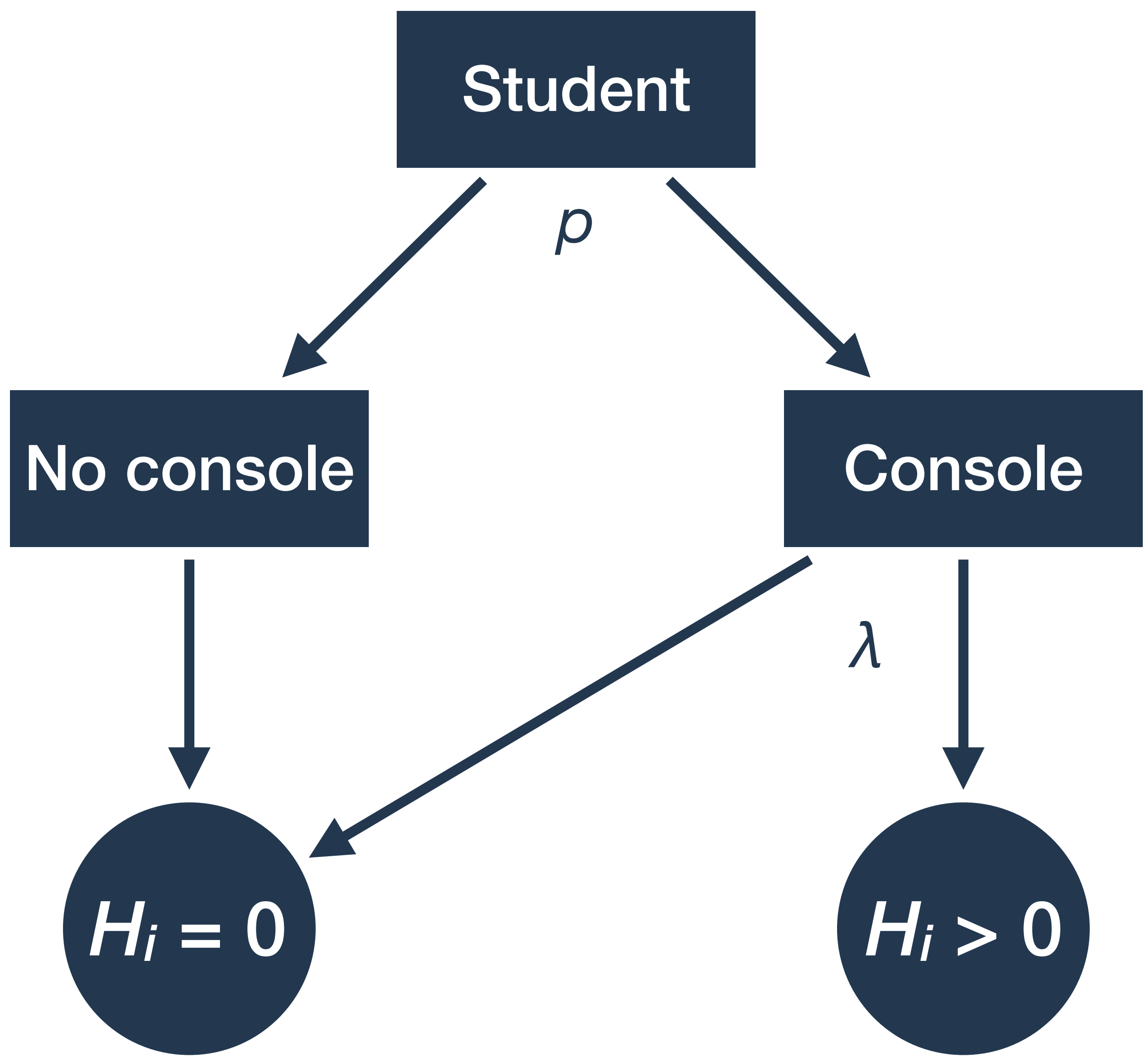
$$\beta_p \sim \text{Norm}(0, 2)$$

$$\alpha_\lambda \sim \text{Norm}(3, 0.5)$$

$$\beta_\lambda \sim \text{Norm}(0, 0.3)$$

Zero-inflation

Data story:



$$H_i \begin{cases} = 0 & \text{if } c_i = 1 \\ \sim \text{Pois}(\lambda_i) & \text{if } c_i = 0 \end{cases}$$

$$c_i \sim \text{Bern}(p_i)$$

$$\text{logit}(p_i) = \alpha_p + \beta_p W_i$$
$$\log(\lambda_i) = \alpha_\lambda + \beta_\lambda M_i$$

$$\alpha_p \sim \text{Norm}(0, 1)$$
$$\beta_p \sim \text{Norm}(0, 2)$$
$$\alpha_\lambda \sim \text{Norm}(3, 0.5)$$
$$\beta_\lambda \sim \text{Norm}(0, 0.3)$$