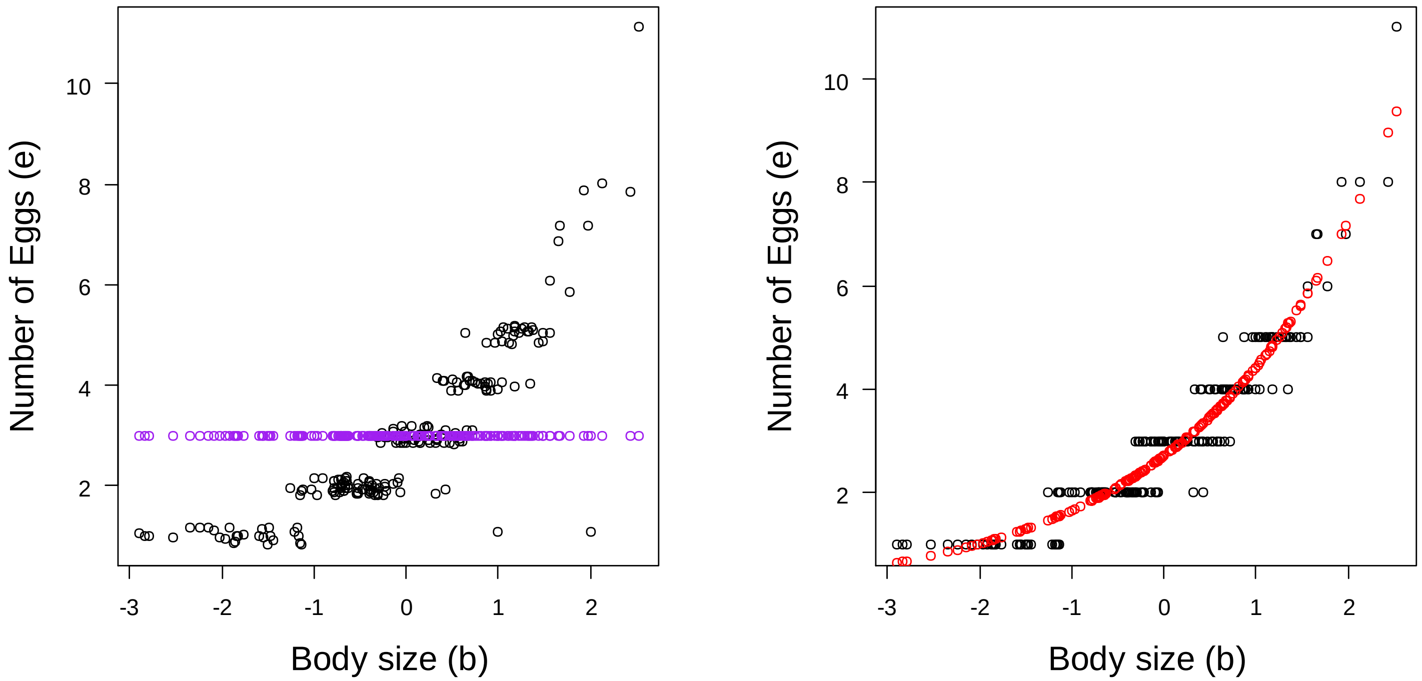
Goodness-of-fit testing (posterior predictive checks)



Also some Bayesian model selection, comparison to AICc

Wrapping up Tuesday

1. *What is a 'meaningful' difference in AICc?* **2, 3, 4, 7?**

AICc (for linear models)

$$\mathrm{AICc} = 2k + n \times \left(\ln\left(2 \times \pi \times \frac{\mathrm{RSS}}{n}\right) + 1\right) + \frac{2k^2+2k}{n-k-1}$$

Wrapping up Tuesday

1. *What is a 'meaningful' difference in AICc?* **2, 3, 4, 7?**

    - **This is somewhat dependent on n and k.**

$$\mathrm{AICc} = 2k + n \times \left(\ln\left(2 \times \pi \times \frac{\mathrm{RSS}}{n}\right) + 1\right) + \frac{2k^2+2k}{n-k-1}$$

Wrapping up Tuesday

1. *What is a 'meaningful' difference in AICc?* **2, 3, 4, 7?**

    - **This is somewhat dependent on n and k.**

2. *What does a 'meaningful' difference in AICc mean?*

$$\mathrm{AICc} = 2k + n \times \left(\ln\left(2 \times \pi \times \frac{\mathrm{RSS}}{n}\right) + 1\right) + \frac{2k^2+2k}{n-k-1}$$

Wrapping up Tuesday

1. *What is a 'meaningful' difference in AICc?* **2, 3, 4, 7?**

    - **This is somewhat dependent on n and k.**

2. *What does a 'meaningful' difference in AICc mean?*

The reduction in RSS…

$$\mathrm{AICc} = 2k + n \times \left(\ln\left(2 \times \pi \times \frac{\mathrm{RSS}}{n}\right) + 1\right) + \frac{2k^2+2k}{n-k-1}$$

offsets the penalties for adding complexity.

Wrapping up Tuesday

1. *What is a 'meaningful' difference in AICc?* **2, 3, 4, 7?**

    - **This is somewhat dependent on n and k.**

2. *What does a 'meaningful' difference in AICc mean?*

    - **This is somewhat dependent on n and k.**

3. *Remember, these are based on probabilities and likelihoods. Sometimes, <u>unlikely things happen</u> (Type 1 error: false positives, Type 2 error: false negatives)*

Announcements

1. Next week we'll start in-class work sessions

2. Presentations and final reports (in a format of your preference) are coming up!
   - Presentations Nov 26th, Dec 3rd, and Dec 5th (< 10m)
   - Reports due Dec 5th

3. Final report formatting is extremely flexible*
   - Code and data
   - Conceptual diagram(s)
   - Methods section (w/results?)
   - Could also be in the form of a dissertation chapter or manuscript draft

***You don't have to use SEM. If you don't, just explain why in ~1 page and give me the same info as above for a non-SEM model structure**

Guidance/thoughts on ‘in-class’ time

- Let’s take advantage of November!

- Doesn’t have to be SEM related (i.e., I am happy to help with other stuff if I can).

- If you don’t have data (or it’s very preliminary), we can simulate data quickly.

- **Do not** feel like it’s necessary to fully understand concepts before talking to me.
    - I’m still learning many of these concepts (and have been surprised this term!)
    - There is a decent chance I may not know the answer?
    - I almost certainly want to know the answer!

Today’s goals

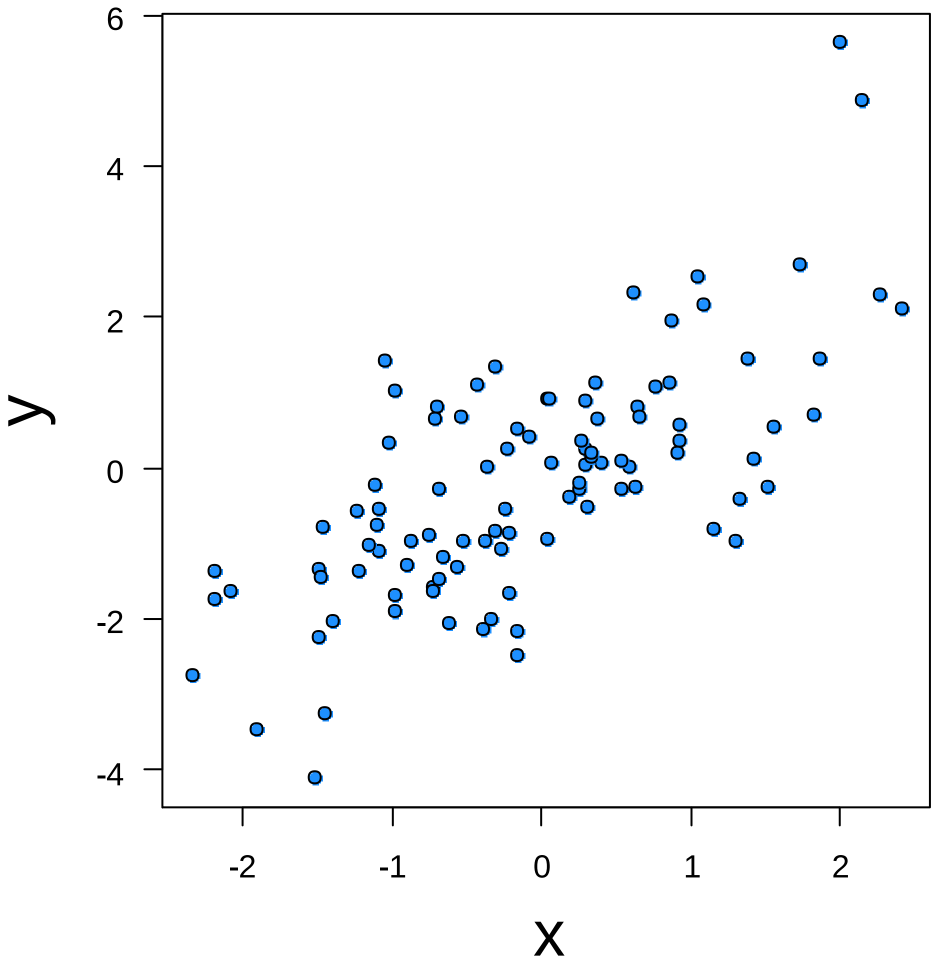
1. A simple univariate lm() example in R and JAGS

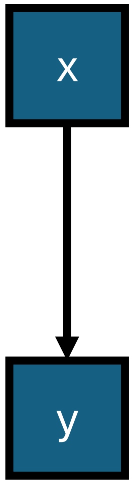
2. A simple, but also complicated, Poisson regression example in JAGS

Let’s start with a simple motivating example

(Normal_log_lik_Bayes_p.R)

Data simulation



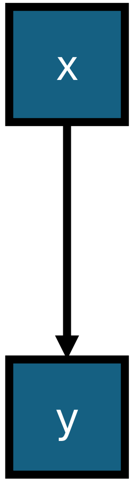


$$\boldsymbol{x} \sim \mathrm{Normal}(0,1)$$

$$\boldsymbol{y} \sim \mathrm{Normal}(\boldsymbol{x}, 1)$$

$$\mathrm{n} = 100$$

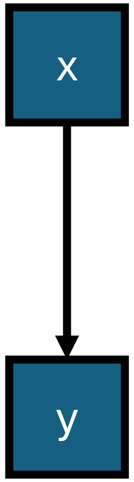
Run two models



$$\boldsymbol{y} \sim \mathrm{Normal}(\beta \boldsymbol{x}, \sigma^2)$$
$$\beta \sim \mathrm{Normal}(0,1)$$
$$\sigma \sim \mathrm{Gamma}(1,1)$$

```
glm(y ~ 0 + x)
```

And calculate a bunch of stuff



glm

- The mean of β
- The p-value for β
- AICc

JAGS

- The median of β (q50)
- The P(β > 0) (f-value)
- DIC (and pD)
- WAIC (and p_{WAIC})
- Bayesian p-value

I don’t expect y’all to memorize these today!

There is explicit code to calculate these values, I do want to discuss it because it’s useful

What is DIC?

What is DIC?

Deviance = -2 x log-likelihood

Effective number of parameters

$$\mathrm{DIC} = \boxed{\bar{D}} + \boxed{pD}$$

$$\mathrm{AICc} = \boxed{-2\log\mathrm{L}} + \boxed{2k + \frac{2k^2 + 2k}{n - k - 1}}$$

Deviance = -2 x log-likelihood

Parameter penalty

They're very similar!

How do we get the log-likelihood?

$$\text{DIC} = \overline{D} + pD$$

$$\text{pD} = \left(\frac{\sigma_D^2}{2}\right)$$

We do this for every iteration

$$\boxed{\text{logL} = -\frac{n}{2} \times \ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2} \times \sum (\boldsymbol{y} - \beta\boldsymbol{x})^2}$$

$$\text{D} = -2 \times \text{logL}$$

Well, JAGS (and Stan and Nimble) calculates it for us!

What is WAIC?

It’s a lot like DIC and AICc with some improvements

We do this for every point at every iteration

$$\boxed{\log\mathrm{L} = -\frac{1}{2} \times \ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2} \times \sum (y - \beta x)^2}$$

It’s a lot like DIC and AICc with some improvements

1) We then take the mean of the likelihood (not log-likelihood) for each point across all iterations
2) We take the log of those means
3) And then we sum across points to get log pointwise predictive density (lppd)

It’s a lot like DIC and AICc with some improvements

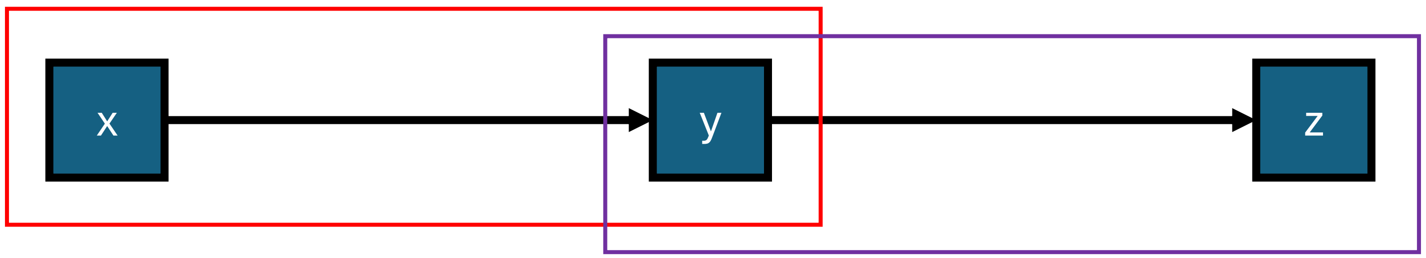
1) To get the effective number of parameters, we sum the variances of the log-likelihoods for each point

$$\text{WAIC} = -2 \times \text{lppd} + 2 \times p_{\text{WAIC}}$$

$$\text{DIC} = \bar{D} + pD$$

$$\text{AICc} = -2\text{logL} + 2k + \frac{2k^2 + 2k}{n - k - 1}$$

There are some really nice things here too



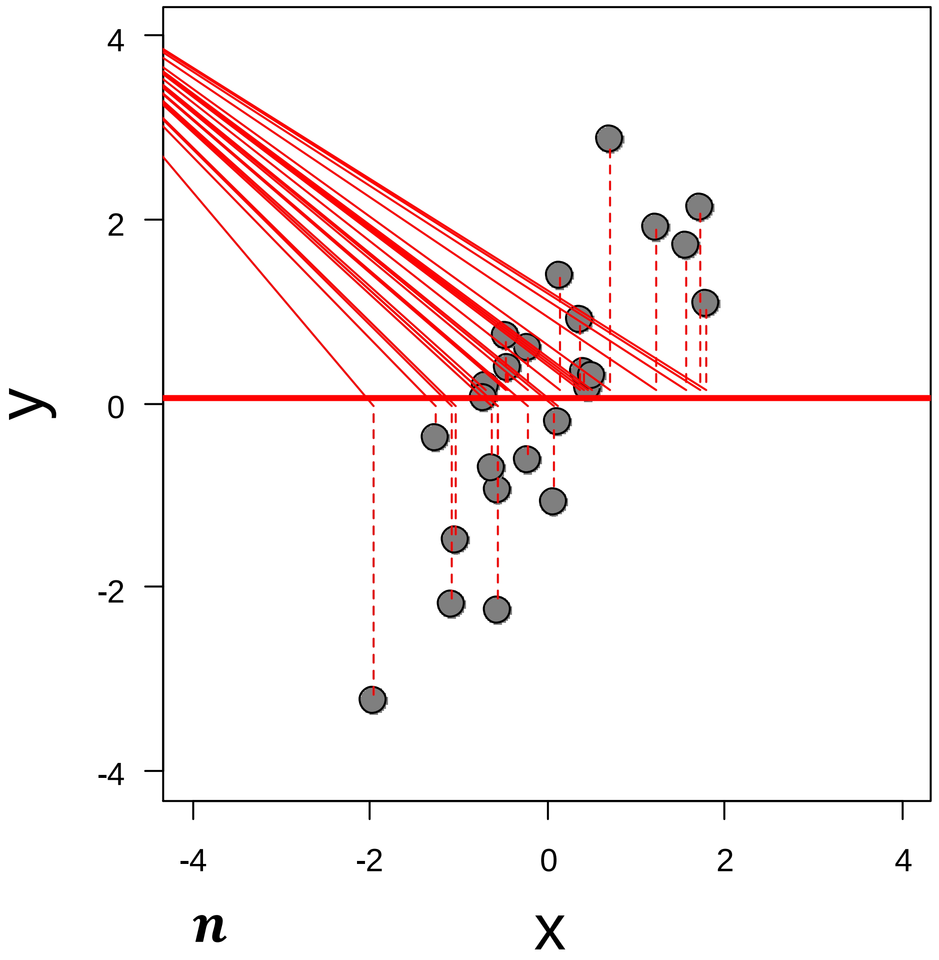
$$\text{WAIC for full model} = \sum \text{WAIC of sub} - \text{models}$$

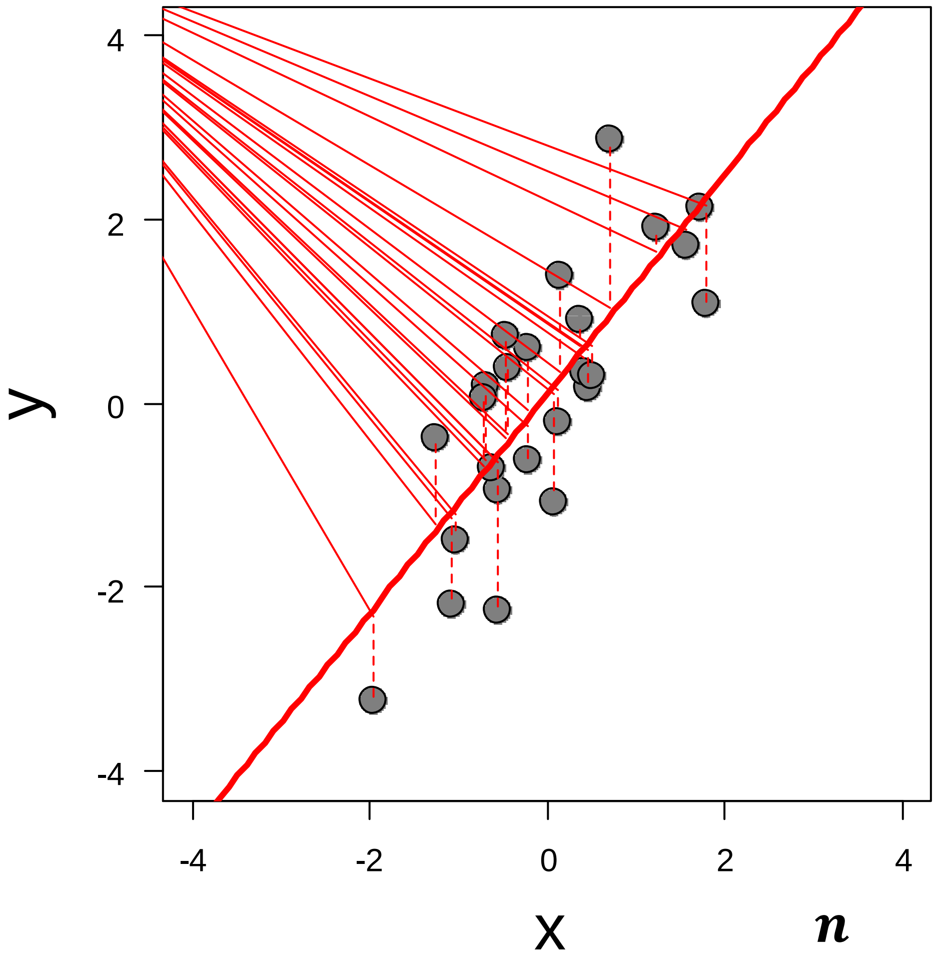
What’s up with ‘effective parameters’?

Imagine you use an incredibly strong prior that drives inference. Should that count as a parameter?

What are Bayesian p-values?

Remember residual sums of squares?





$$50.156 = \sum_{i=1}^{n} (y_i - \eta_i)^2$$

$$19.448 = \sum_{i=1}^{n} (y_i - \eta_i)^2$$

What are Bayesian p-values?

- We'll simulate new data (y') from our parameter estimates at every iteration

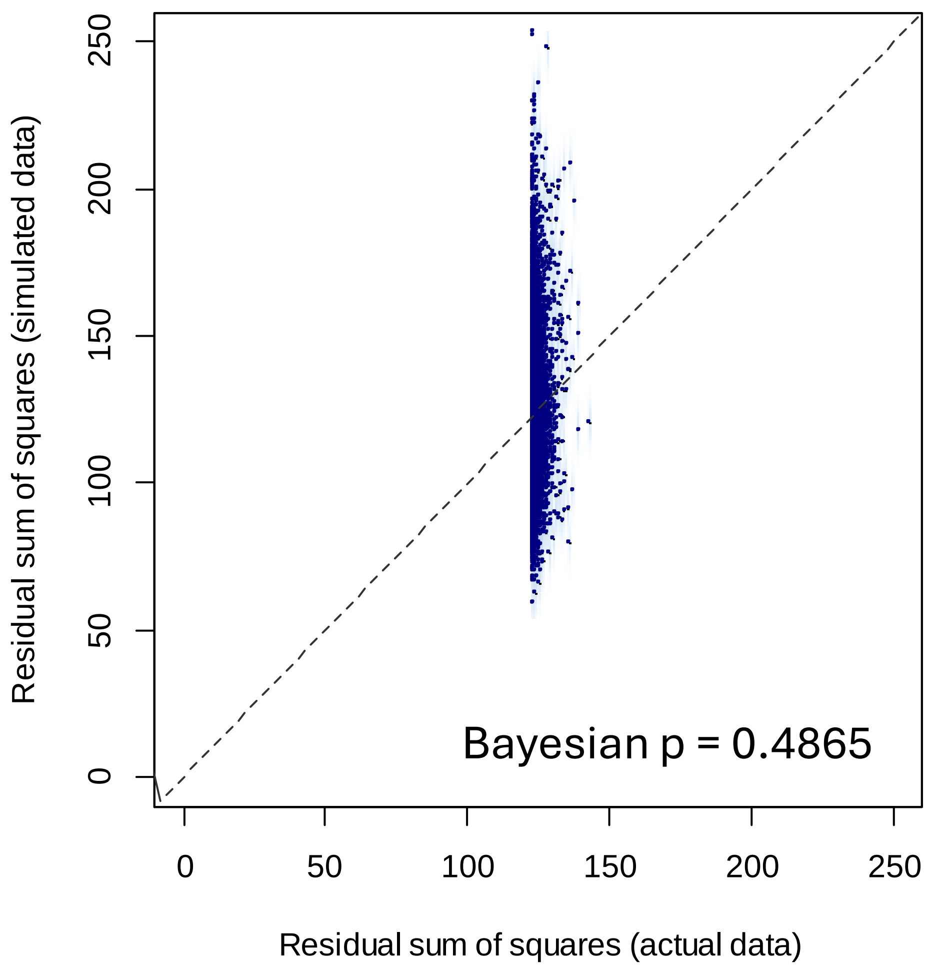
$$\boldsymbol{y} \sim \text{Normal}(\beta \boldsymbol{x}, \sigma^2)$$

$$\boldsymbol{y}' \sim \text{Normal}(\beta \boldsymbol{x}, \sigma^2)$$

- We'll calculate RSS for the new data (and the real data, note we often use other discrepancy statistics for other model types)

- We'll compare the RSS from simulated and real data. The p-value is the proportion of times that the RSS from the real data is greater than the RSS from the simulated data

What are Bayesian p-values?

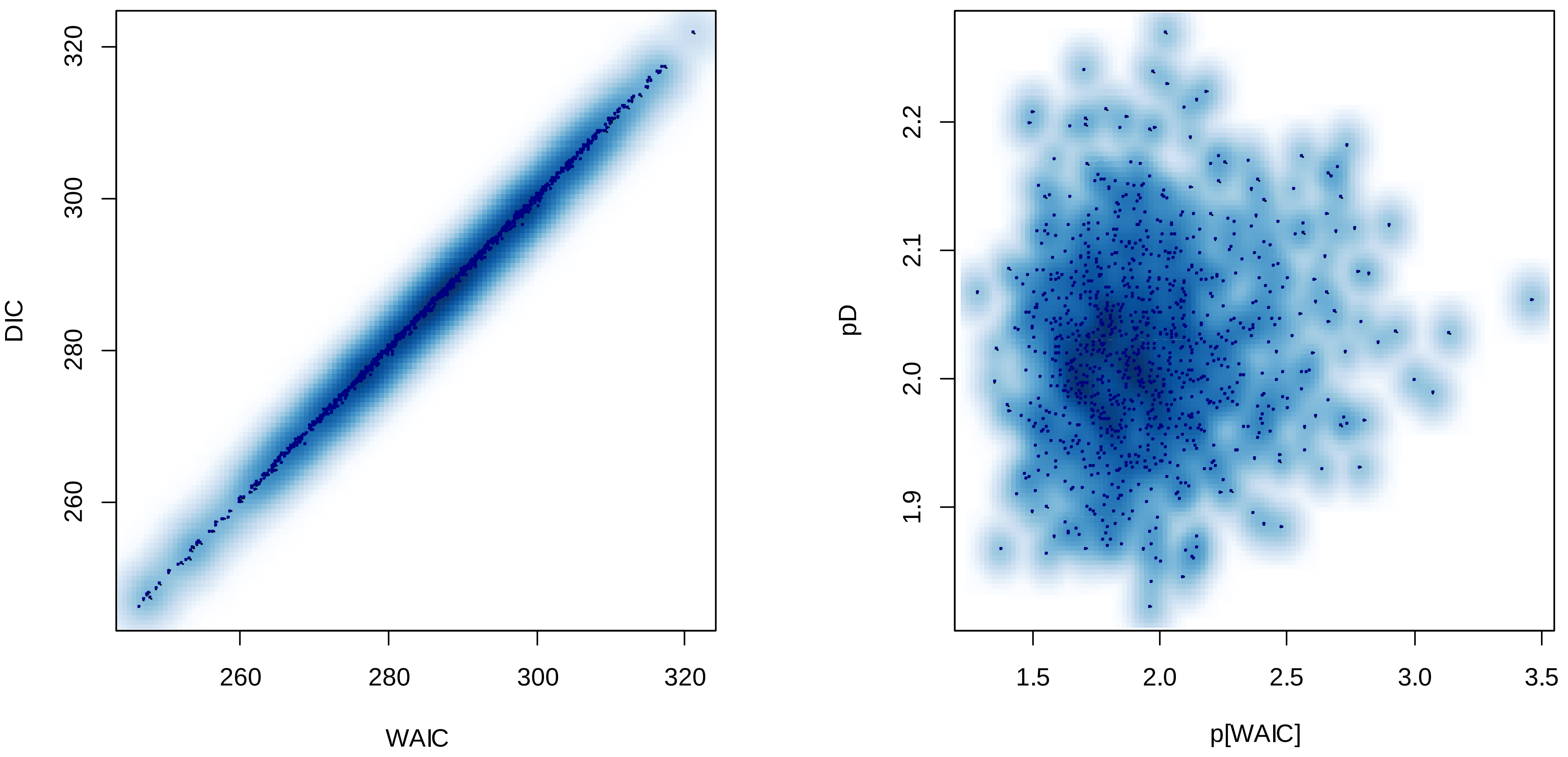


What are acceptable Bayesian p-values?

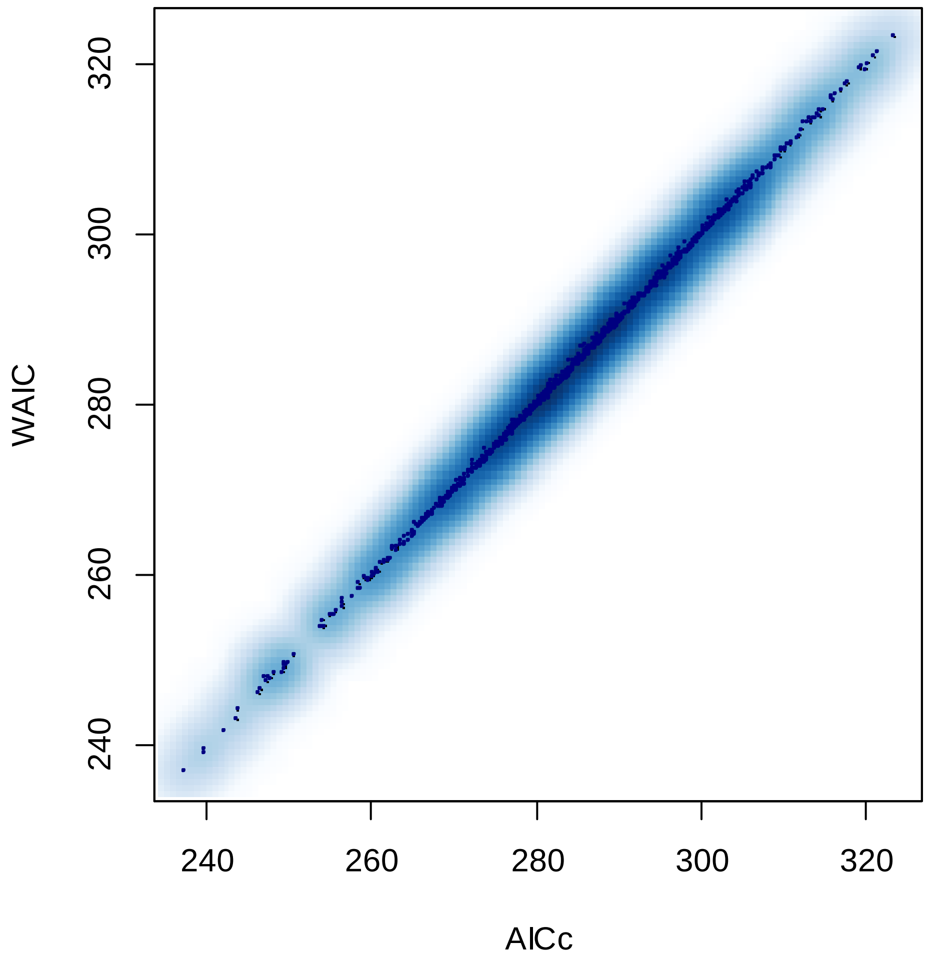
- A fitting model has a Bayesian p-value near 0.5, and values close to 0 or close to 1 suggest doubtful fit of the model (Kéry [**2010**] *Academic Press*)

Let’s do that 1k times

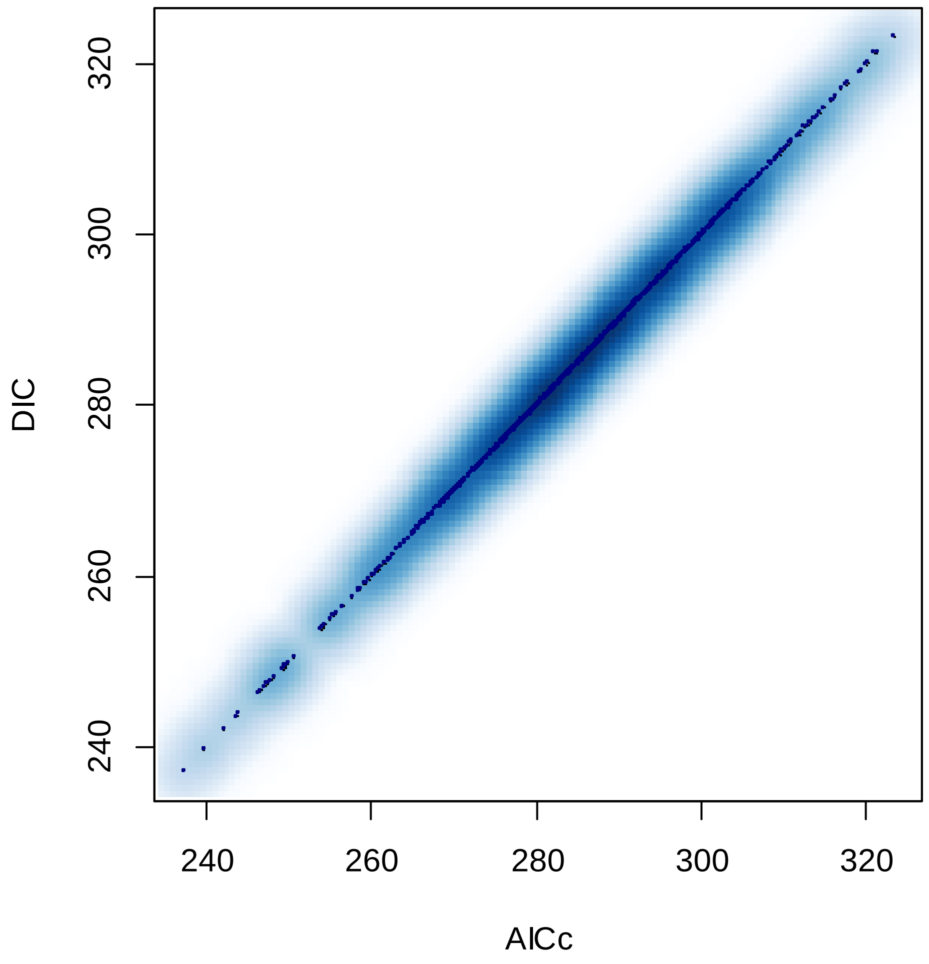
WAIC vs. DIC



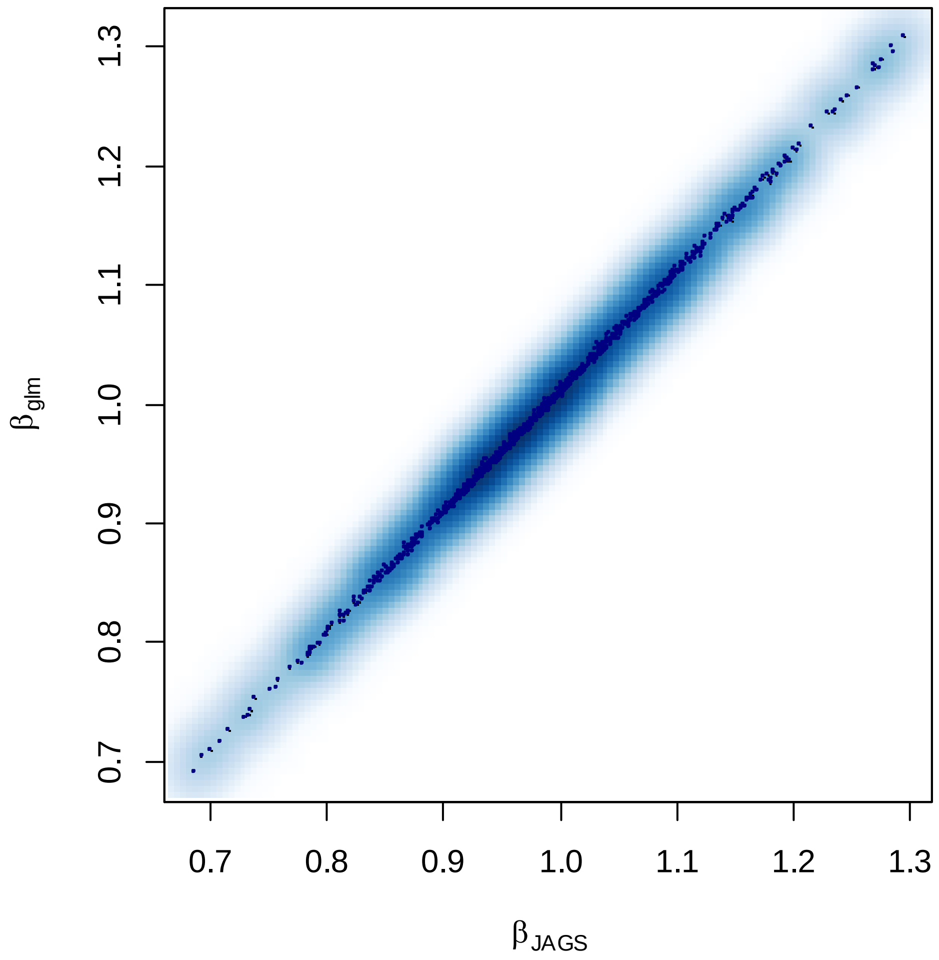
WAIC vs. AIC



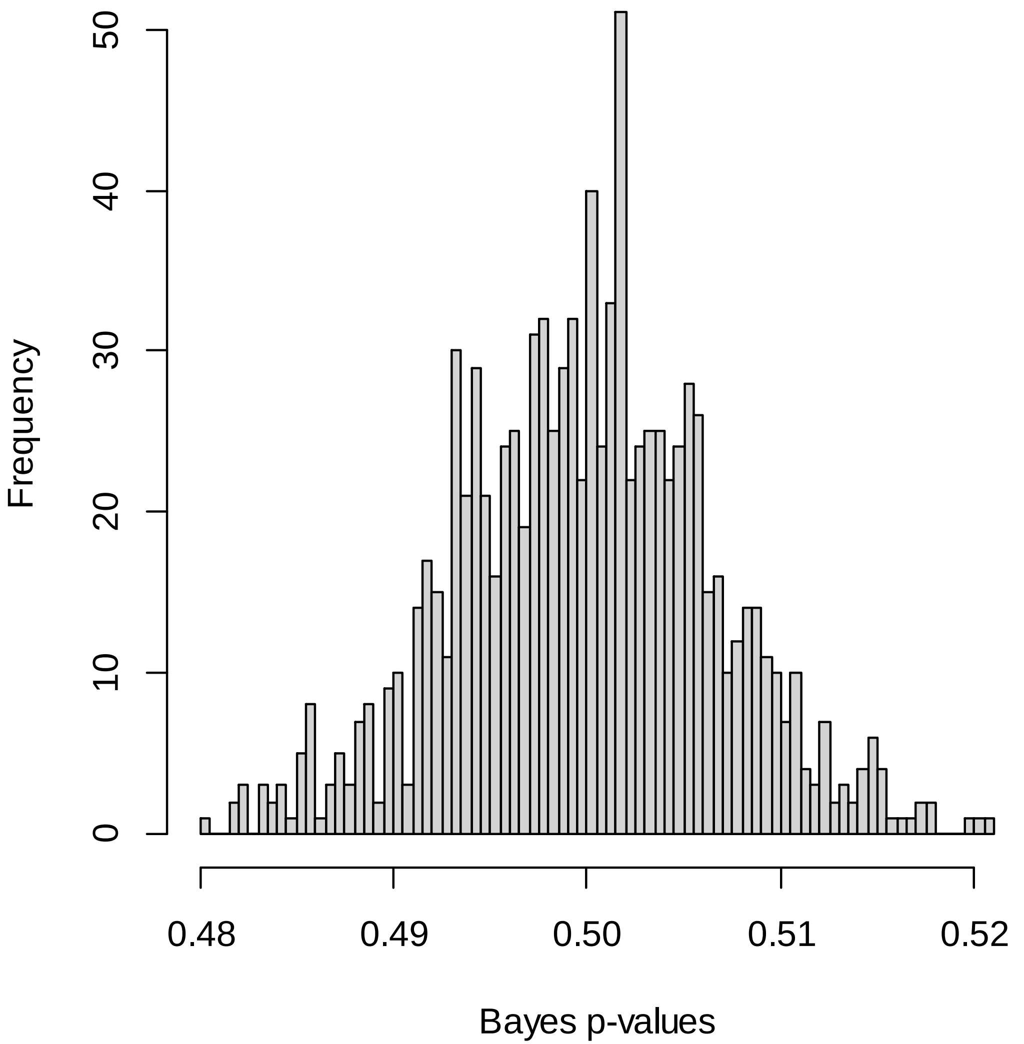
DIC vs. AIC



Parameter estimates



Bayesian p-values



Ok, let’s break some stuff again

Why not? It’ll be fun

We're going to simulate clutch size as a function of body size

$$n = 200$$

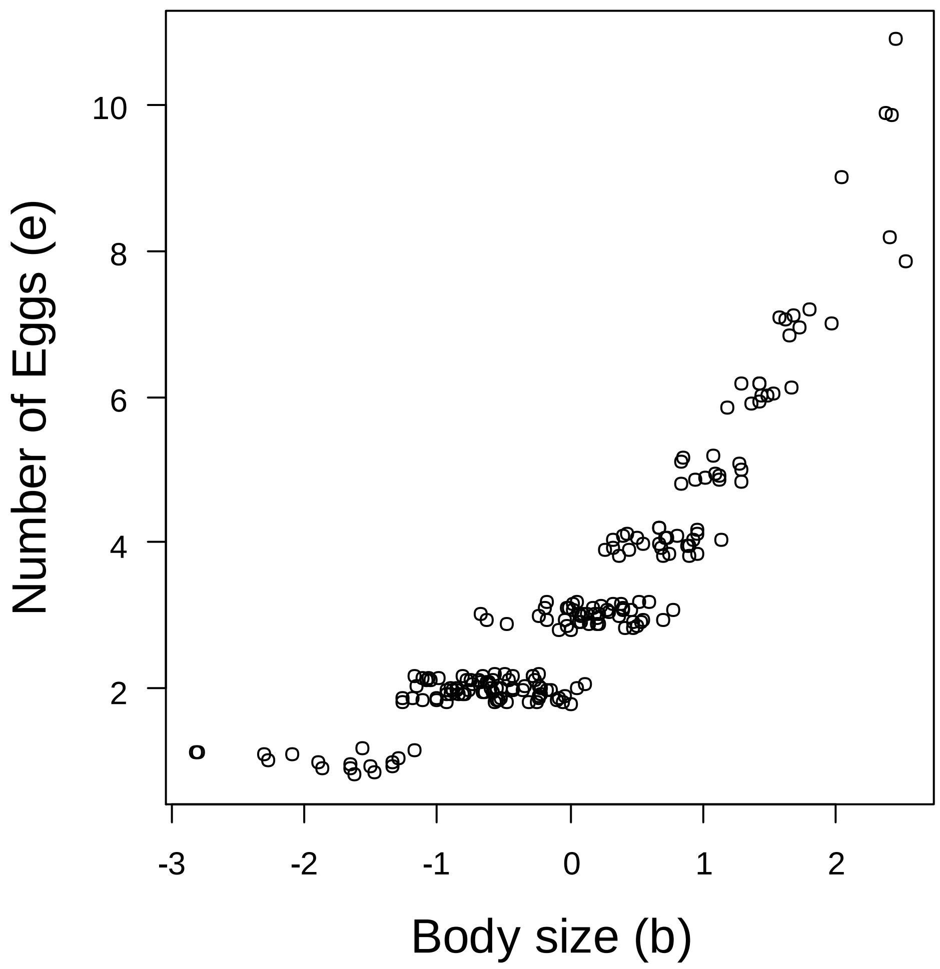
$$\boldsymbol{b} \sim \text{Normal}(0,1)$$

$$\boldsymbol{\psi} = e^{\alpha + \beta \times \boldsymbol{b}}$$

$$\boldsymbol{e} \sim \text{round}(\psi)$$

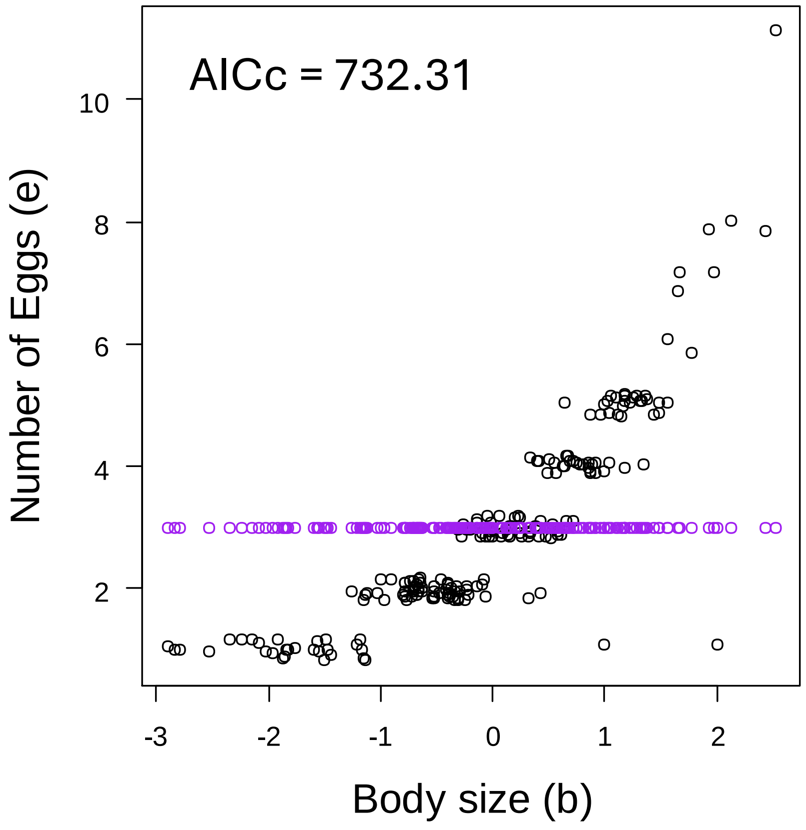
$$\alpha = 1$$

$$\beta = 0.5$$

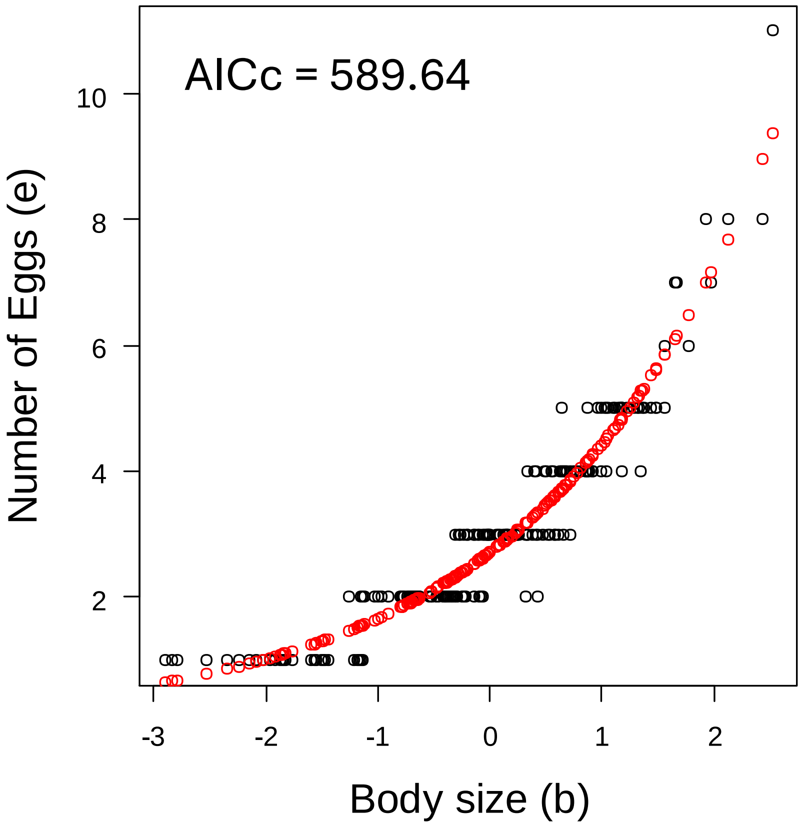


Residuals and AICc…

`glm(y ~ 1, family = 'poisson')`



`glm(y ~ b, family = 'poisson')`



Deviance (D) Information Criterion (DIC; run the same models in JAGS)

Model 0

$$\boldsymbol{\psi} = e^{\alpha}$$

$$\boldsymbol{e} \sim \text{Poisson}(\psi)$$

$$\alpha \sim \text{Normal}(1,1)$$

Model 1

$$\boldsymbol{\psi} = e^{\alpha+\beta\times\boldsymbol{b}}$$

$$\boldsymbol{e} \sim \text{Poisson}(\psi)$$

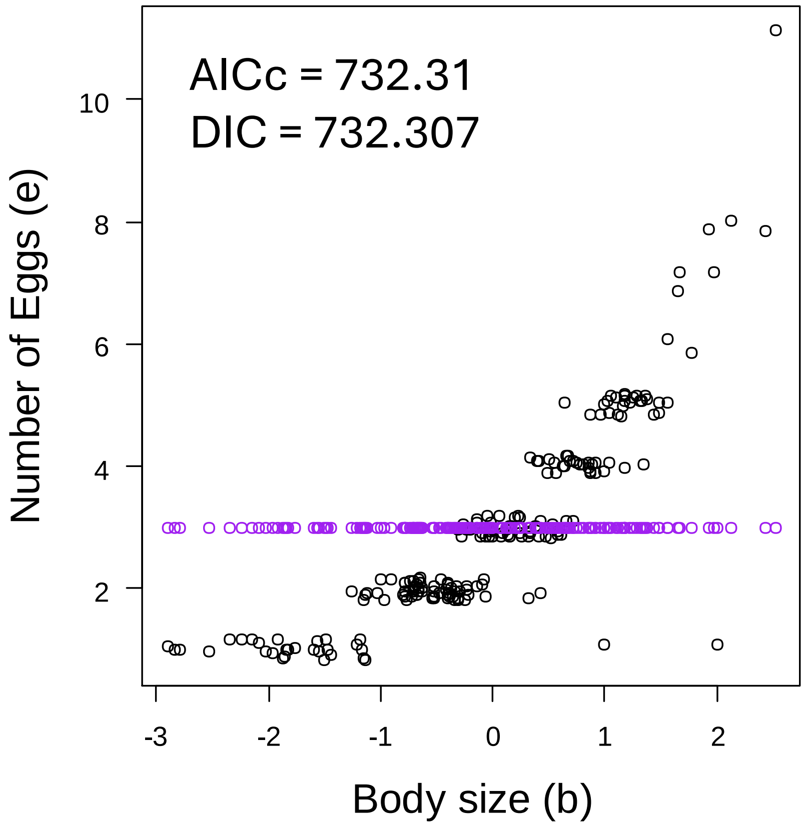
$$\alpha \sim \text{Normal}(1,1)$$

$$\beta \sim \text{Normal}(0,1)$$

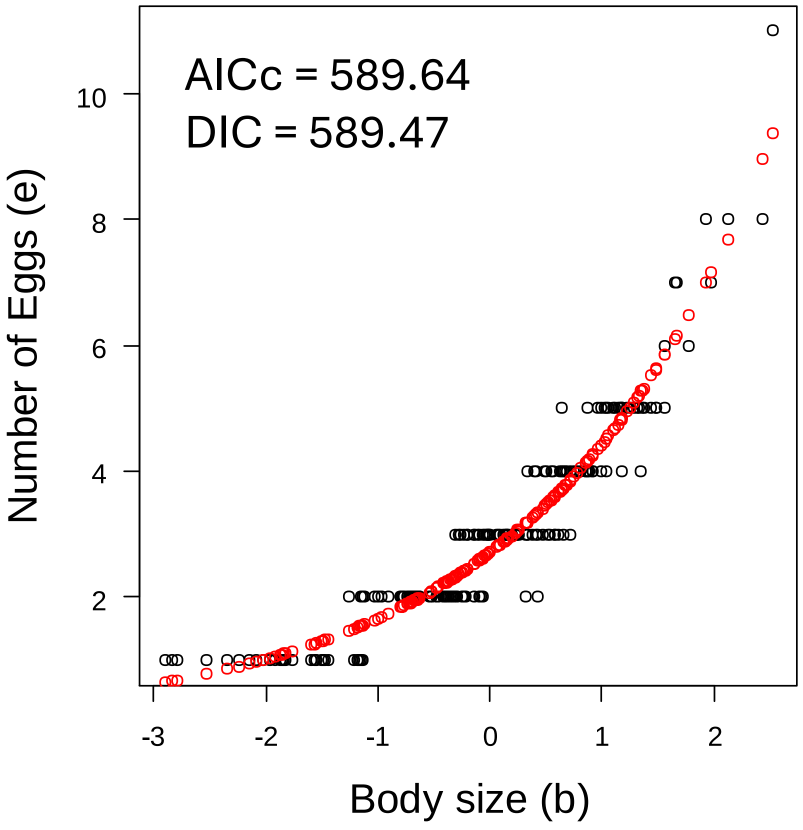
$$\mathbf{DIC} = \boldsymbol{\bar{D}} + \mathbf{2}\boldsymbol{pD}$$

$$\boldsymbol{pD} = \frac{\boldsymbol{\sigma}_{\boldsymbol{D}}^{\mathbf{2}}}{\mathbf{2}}$$

DIC

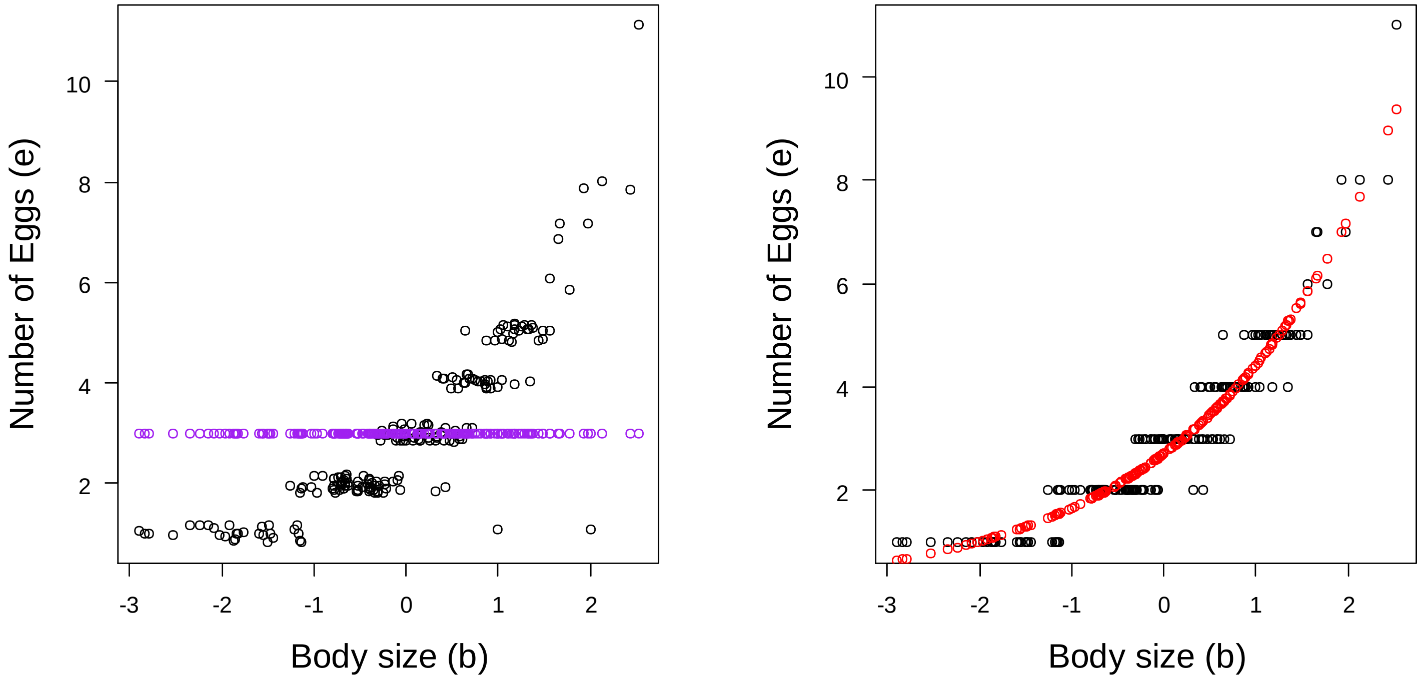


AICc = 732.31
DIC = 732.307
Number of Eggs (e)
Body size (b)



AICc = 589.64
DIC = 589.47
Number of Eggs (e)
Body size (b)

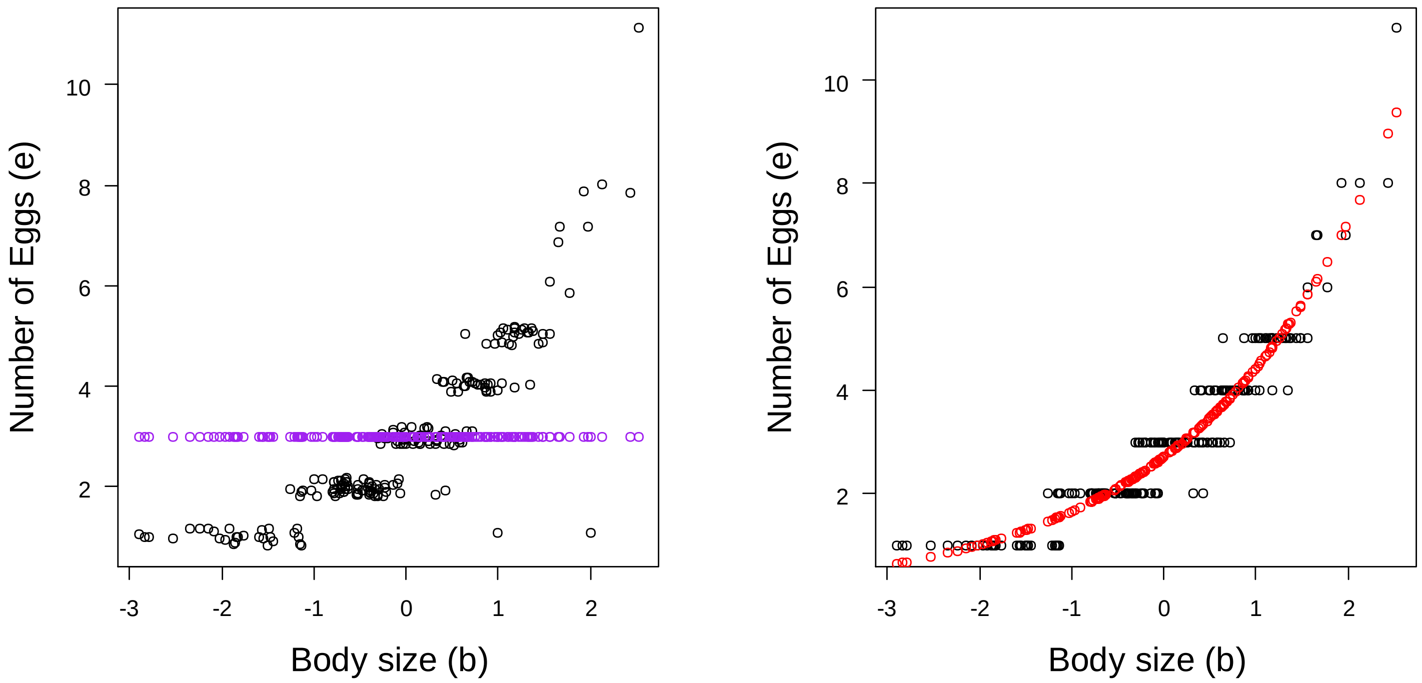
Bayesian p-values (RSS for this Poisson regression)



1. **Calculate (Pearson's) residuals at each iteration**

$$\boldsymbol{\varepsilon} = \frac{\boldsymbol{e} - e^{\beta_0 + \beta_1 \times \boldsymbol{b}}}{\sigma_e}$$

Bayesian p-values (RSS for this Poisson regression)

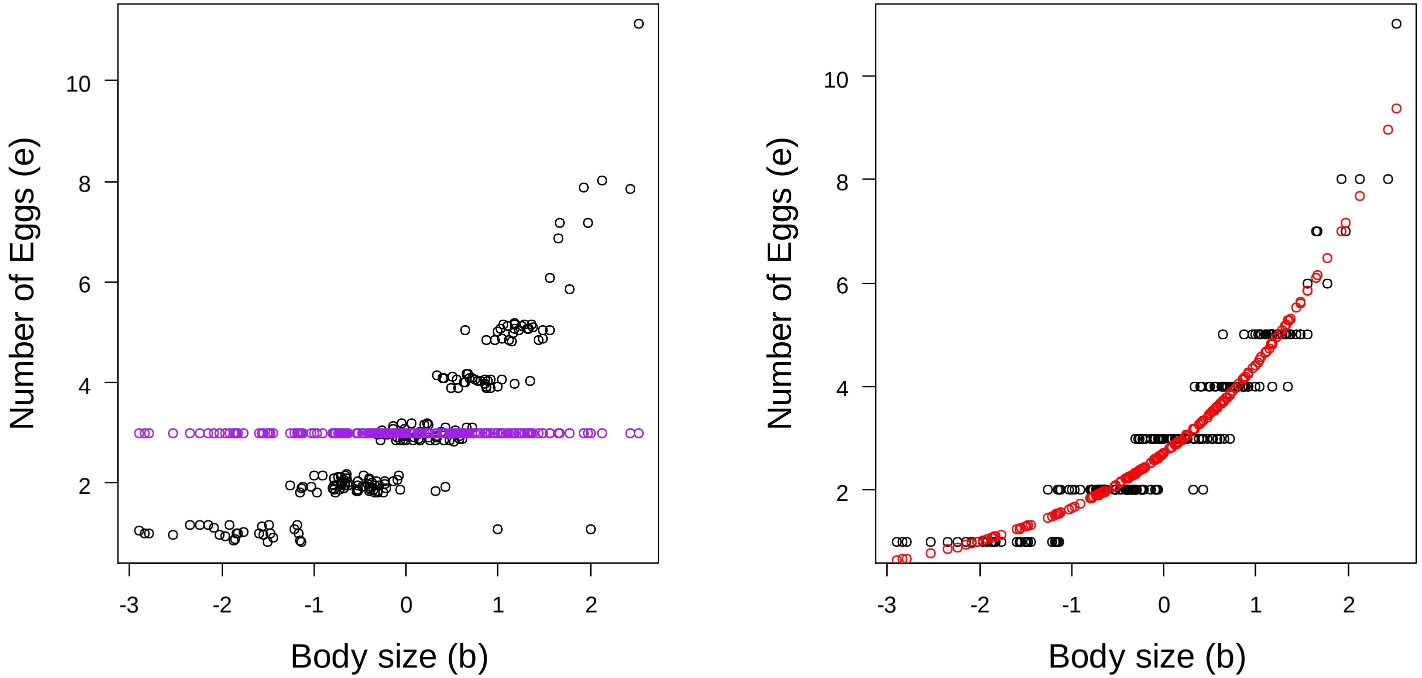


1. **Calculate (Pearson's) residuals at each iteration**
2. **Generate 'new' data**

$$\boldsymbol{e} \sim \text{Poisson}(\boldsymbol{\psi})$$

$$\boldsymbol{e}' \sim \text{Poisson}(\boldsymbol{\psi})$$

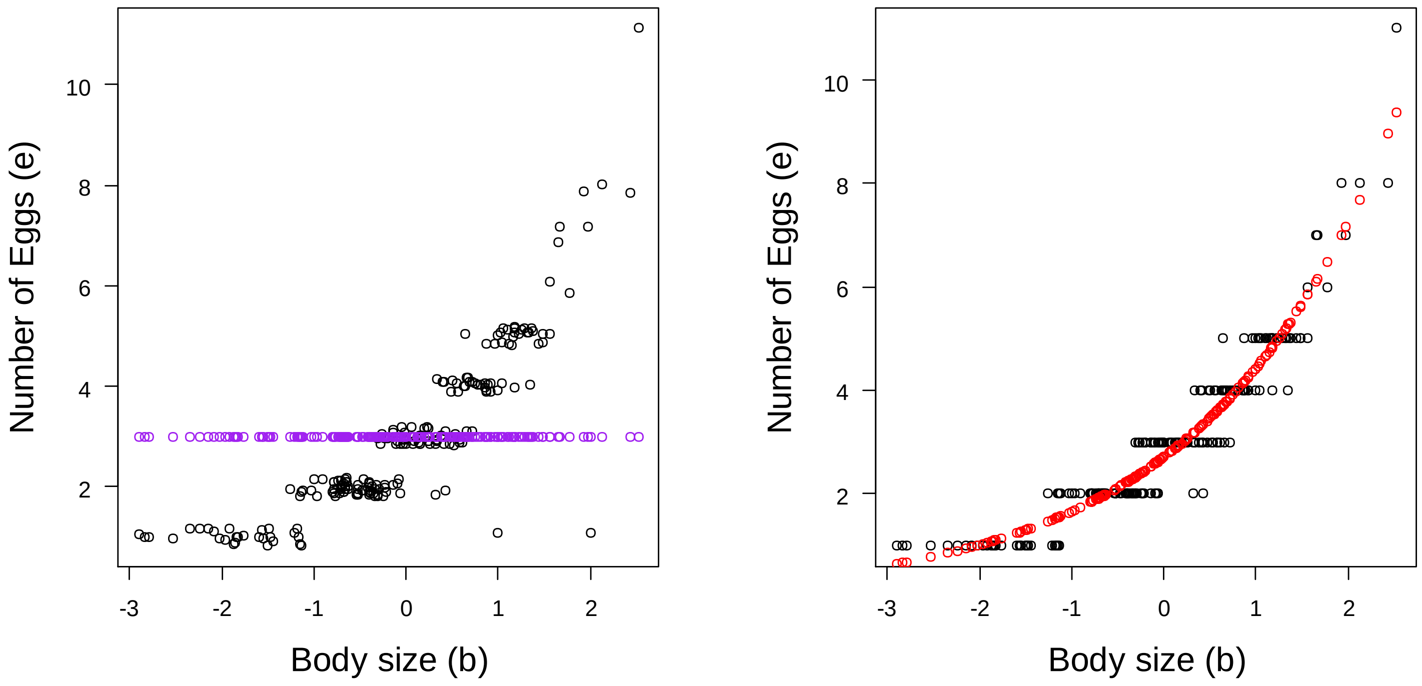
Bayesian p-values (RSS for this Poisson regression)



1. **Calculate (Pearson's) residuals at each iteration**
2. **Generate 'new' data**
3. **Calculate (Pearson's) residuals at each iteration for new data**

$$\boldsymbol{\varepsilon}' = \frac{\boldsymbol{e}' - e^{\beta_0 + \beta_1 \times \boldsymbol{b}}}{\sigma_{e'}}$$

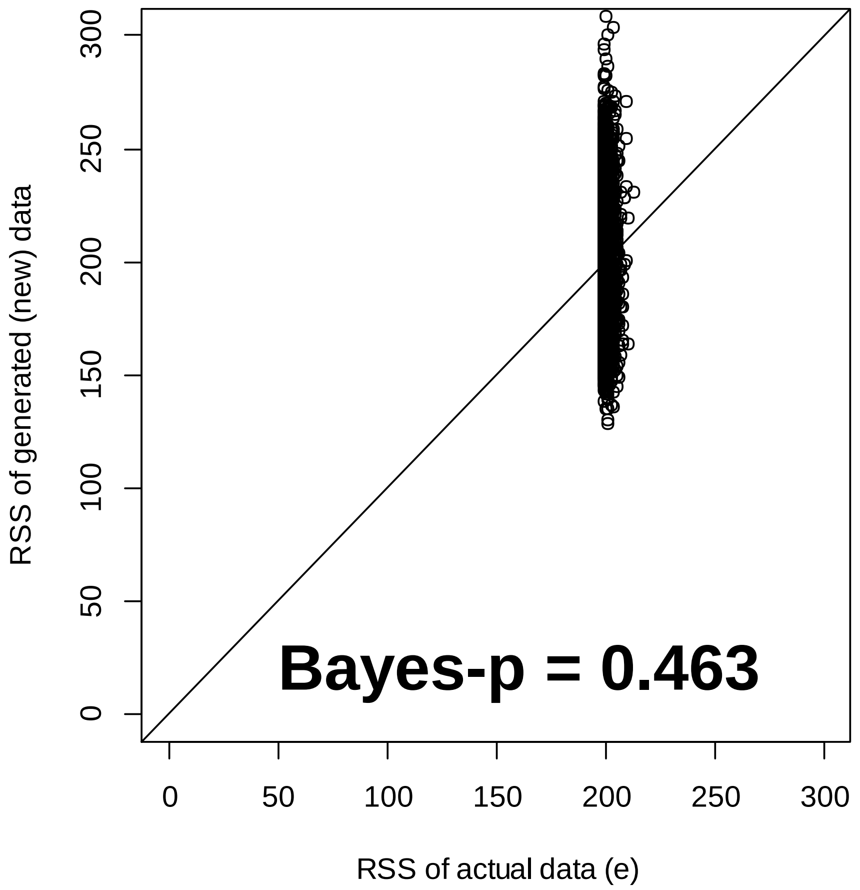
Which model will fit the data better (have a Bayes p closer to 0.5)?

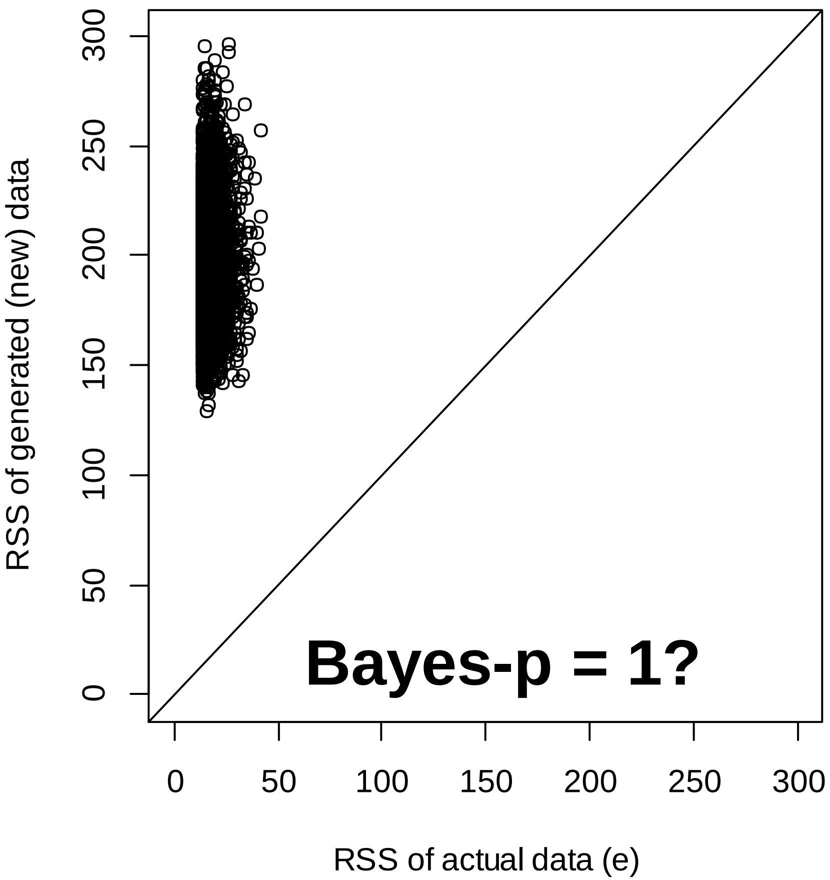


1. Calculate (Pearson's) residuals at each iteration
2. Generate 'new' data
3. Calculate (Pearson's) residuals at each iteration for new data

$$\boldsymbol{\varepsilon}' = \frac{\boldsymbol{e}' - e^{\beta_0 + \beta_1 \times \boldsymbol{b}}}{\sigma_{e'}}$$

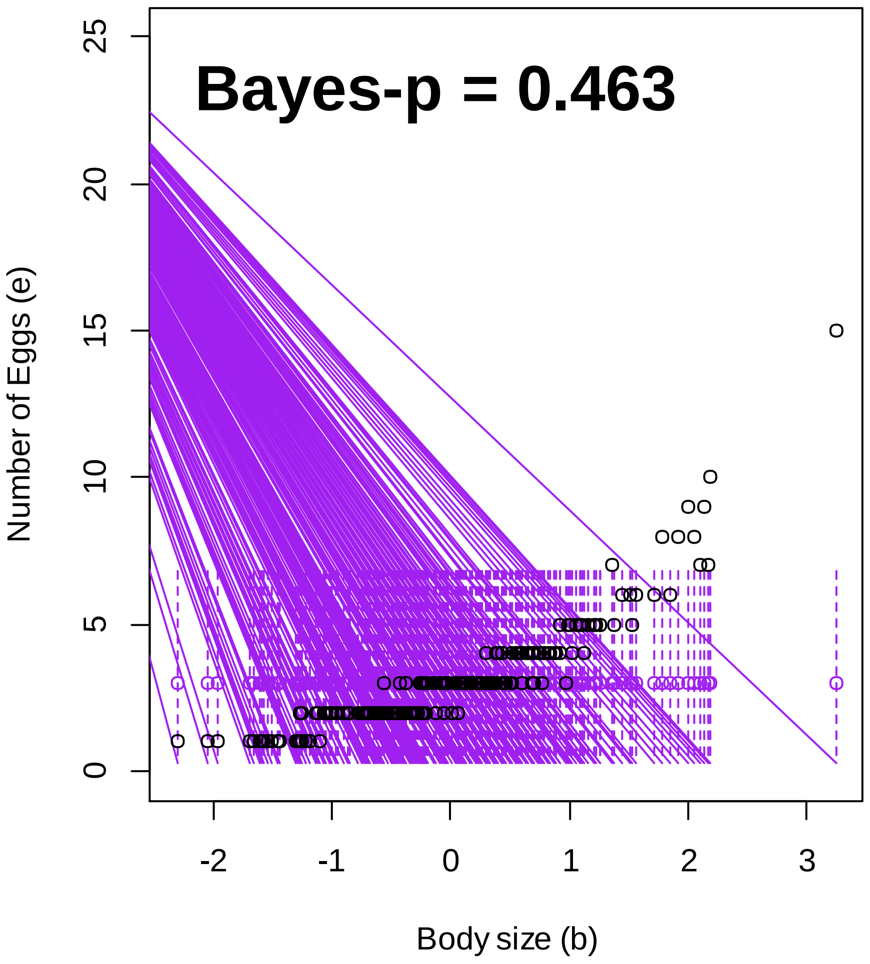
Oh for goodness sake!!

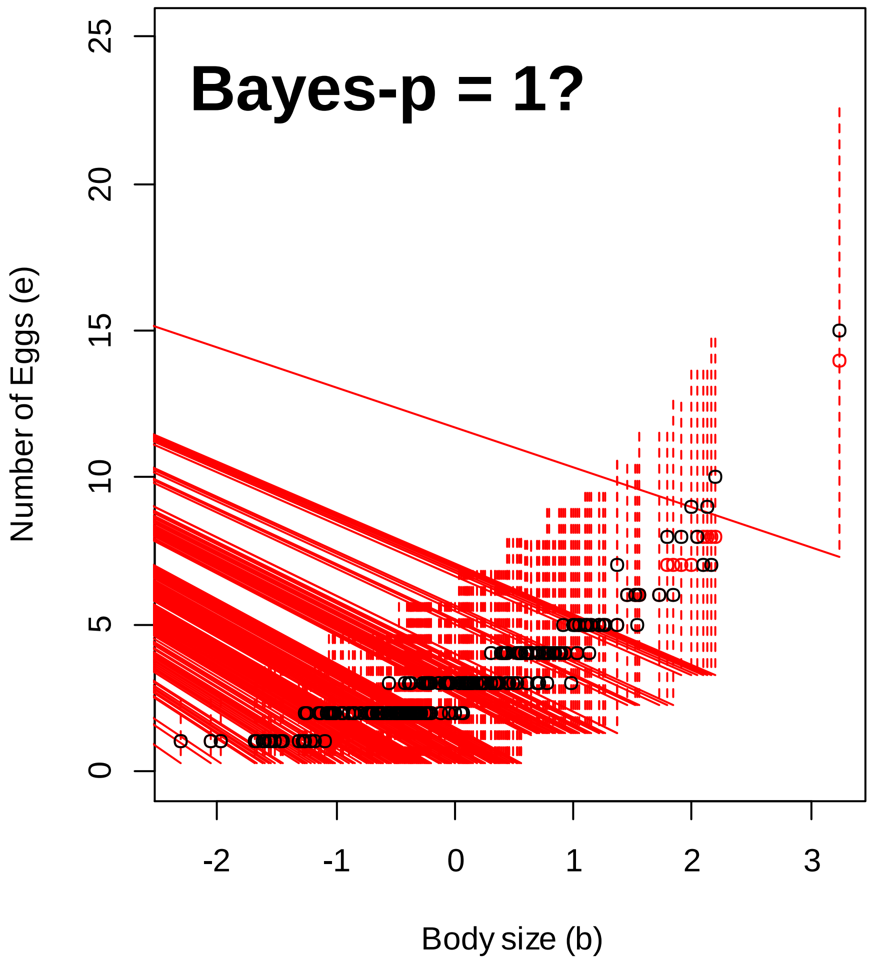




Bayes-p > 0.5 means real data fit better than simulated, and vice versa

Oh for goodness sake!!





Bayes-p > 0.5 means real data fit better than simulated, and vice versa

Let’s dig into the code and find out why!?

It will be fun.

The take-home message here is that poor model fit doesn’t necessarily mean a ‘bad’ model. Rather, it may just mean a bit more thought is necessary?