Multiple Regression:
General things

Tim Frasier

General Idea

- Multiple regression
  - Influence of multiple variables on a single other variable (how does one change in relation to the others?)
  - Predict values of one parameter based on values of the others
- **Allows us to assess the affect of one variable on the predicted variable while accounting for the effects of other predictor variables**

Creates Problems!

1. Proper interpretation of coefficients
2. Deciding what predictor variables are important

Neither of these are trivial!!!

Problem #1: Interpretation of Coefficients

- Differs depending on if interaction term(s) is included
  - Two variables

no interaction $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

with interaction $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

What do each of these coefficients represent?

Problem #1: Interpretation of Coefficients

- Differs depending on if interaction term(s) is included
  - Two variables

no interaction $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

with interaction $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

The intercept (what y would be if both x values were zero)

s represent?

Problem #1: Interpretation of Coefficients

- Differs depending on if interaction term(s) is included
  - Two variables

no interaction $$y = \beta_0 + \boxed{\beta_1 x_1} + \beta_2 x_2$$

with interaction $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

nt?

The independent effect of `x₁` on `y`, once you've considered the effects of `x₂`. How much `y` increases when `x₁` increases by 1 unit.

Problem #1: Interpretation of Coefficients

- Differs depending on if interaction term(s) is included
  - Two variables

no interaction	$y = \beta_0 + \beta_1 x_1 + \boxed{\beta_2 x_2}$
with interaction	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$

The independent effect of x_2 on y, once you've considered the effect of x_1. How much y increases when x_2 increases by 1 unit.

nt?

Problem #1: Interpretation of Coefficients

- Differs depending on if interaction term(s) is included
  - Two variables

no interaction $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

with interaction $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

What do each of these coefficients represent?

Problem #1: Interpretation of Coefficients

- Differs depending on if interaction term(s) is included
  - Two variables

no interaction $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

with interaction $$y = \boxed{\beta_0} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

The intercept (what y would be if both x values were zero)

represent?

Problem #1: Interpretation of Coefficients

- Differs depending on if interaction term(s) is included
  - Two variables

no interaction $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

with interaction $$y = \beta_0 + \boxed{\beta_1} x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

How much y increases when x_1 increases by 1 and $x_2 = 0$. represent?

Problem #1: Interpretation of Coefficients

- Differs depending on if interaction term(s) is included
  - Two variables

no interaction $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

with interaction $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

sent?

How much y increases when x_2 increases by 1 and $x_1 = 0$.

Problem #1: Interpretation of Coefficients

- Differs depending on if interaction term(s) is included
  - Two variables

no interaction $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

with interaction $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

How much $\mathbf{y}$ increases when the product of $\mathbf{x}_1$ and $\mathbf{x}_2$ increases by 1. Should only be included if it is thought that $\mathbf{x}_1$ and $\mathbf{x}_2$ have a multiplicative interaction that effects $\mathbf{y}$ above and beyond the additive variation covered in β_1 and β_2 (i.e., it should not be included "just because").

Political Analysis (2005) 13:1–20
doi:10.1093/pan/mpi014

Understanding Interaction Models: Improving Empirical Analyses

Thomas Brambor
New York University, Department of Politics,
726 Broadway, 7th Floor, New York, NY 10003
e-mail: thomas.brambor@nyu.edu

William Roberts Clark
University of Michigan, Center for Political Studies,
ISR 4202 Box 1248, 426 Thompson Street, Ann Arbor, MI 48106-1248
e-mail: wrclark@umich.edu

Matt Golder
Florida State University, Department of Political Science,
531 Bellamy Building, Tallahassee, FL 32306-2230
e-mail: matt.golder@nyu.edu (corresponding author)

Hypothesis Testing and Multiplicative Interaction Terms

Bear F. Braumoeller
Associate Professor

Harvard University
Department of Government
Littauer Center, North Yard
Cambridge, MA 02138
bfbraum@fas.harvard.edu*

August 4, 2004

MOLECULAR ECOLOGY RESOURCES
Molecular Ecology Resources (2016) **16**, 382–387 doi: 10.1111/1755-0998.12499

NEWS AND VIEWS

OPINION

A note on the use of multiple linear regression in molecular ecology

TIMOTHY R. FRASIER
Department of Biology, Saint Mary's University, 923 Robie Street, Halifax, Nova Scotia, Canada B3H 3C3

Problem #1: Interpretation of Coefficients

- Differs depending on how many variables are included
    - Three variables

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3 + \beta_7 x_1 x_2 x_3$$

Problem #1: Interpretation of Coefficients

- Differs depending on how many variables are included
  - Three variables

$$y = \boxed{\beta_0} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3 + \beta_7 x_1 x_2 x_3$$

Intercept. Value of `y` when `x1`, `x2`, and `x3` are all `0`

Problem #1: Interpretation of Coefficients

- Differs depending on how many variables are included
  - Three variables

$$y = \beta_0 + \boxed{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3} + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3 + \beta_7 x_1 x_2 x_3$$

Individual effects of each predictor variable when all others are 0.

Problem #1: Interpretation of Coefficients

- Differs depending on how many variables are included
  - Three variables

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \boxed{\beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3} + \beta_7 x_1 x_2 x_3$$

Multiplicative effects of each possible pair of predictor variables when the third is 0.

Problem #1: Interpretation of Coefficients

- Differs depending on how many variables are included
  - Three variables

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3 + \boxed{\beta_7 x_1 x_2 x_3}$$

Multiplicative effects of all three predictor variables.

Problem #1: Interpretation of Coefficients

- Differs depending on how many variables are included
  - Three variables

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3 + \beta_7 x_1 x_2 x_3$$

- Again, multiplicative effects should only be included if you have a good hypothesis for why (and how) they are important.
- If you include a higher-level effect, must include all lower-level ones as well.
  - If you want to estimate β7, you must also estimate β1 - β6

Your Model Is Like An Oracle

- It will answer the exact question you ask it (**not** the question you meant to ask)

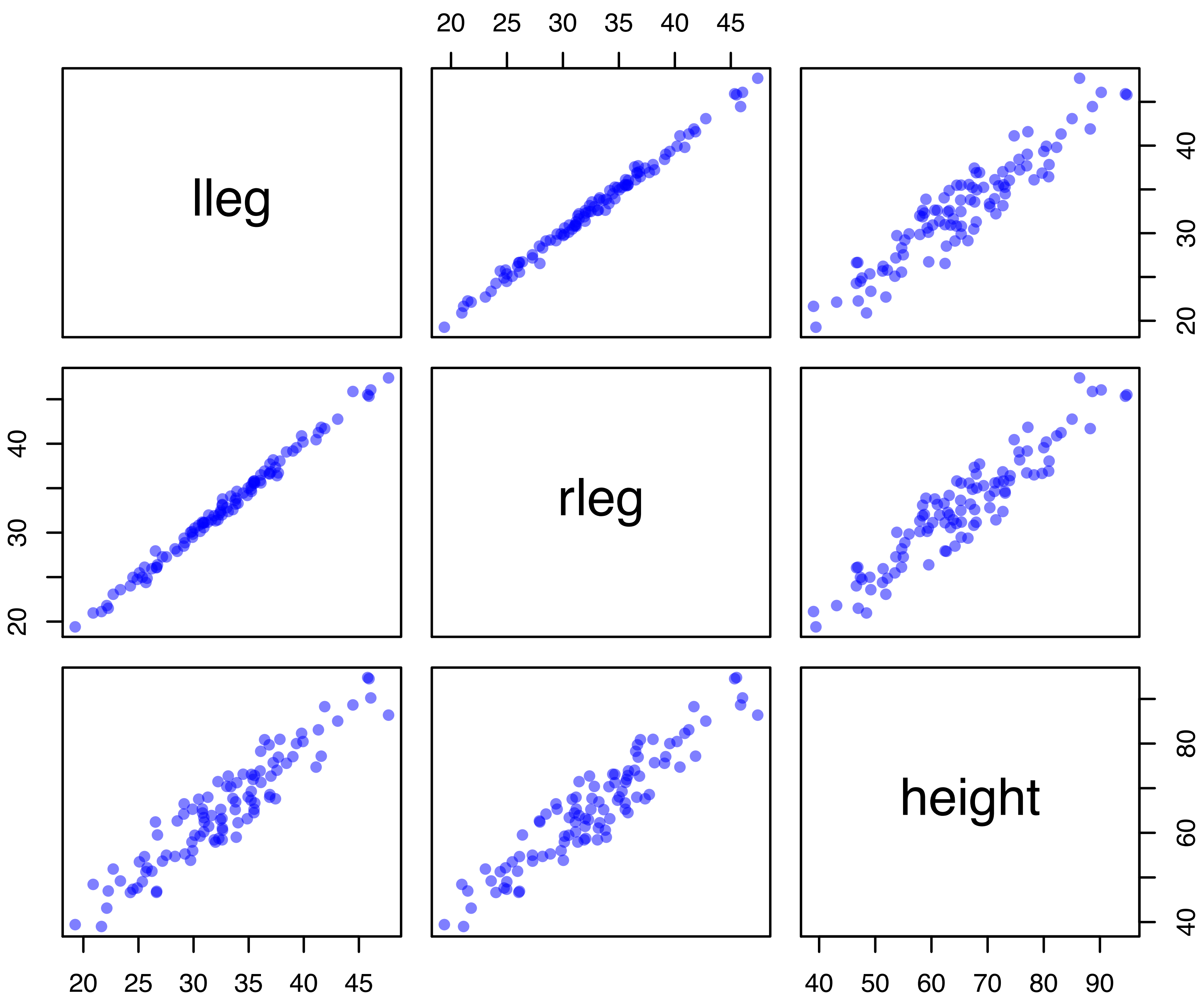
- If you don't interpret this correctly, your empire may burn down

Example: Effect of leg length on height

- Data: "`legs.csv`"

lleg	rleg	height
32.29583	32.03853	58.74574
29.90474	29.48348	65.29385
38.43200	39.07677	75.58189
45.92355	45.33978	94.53371
30.11240	30.52035	59.46205
39.01288	39.18316	77.06575
30.93059	30.56523	63.40671
45.80013	45.51175	94.79874
26.63675	26.05267	46.65168
33.78126	33.63838	65.19896

Example: Effect of leg length on height



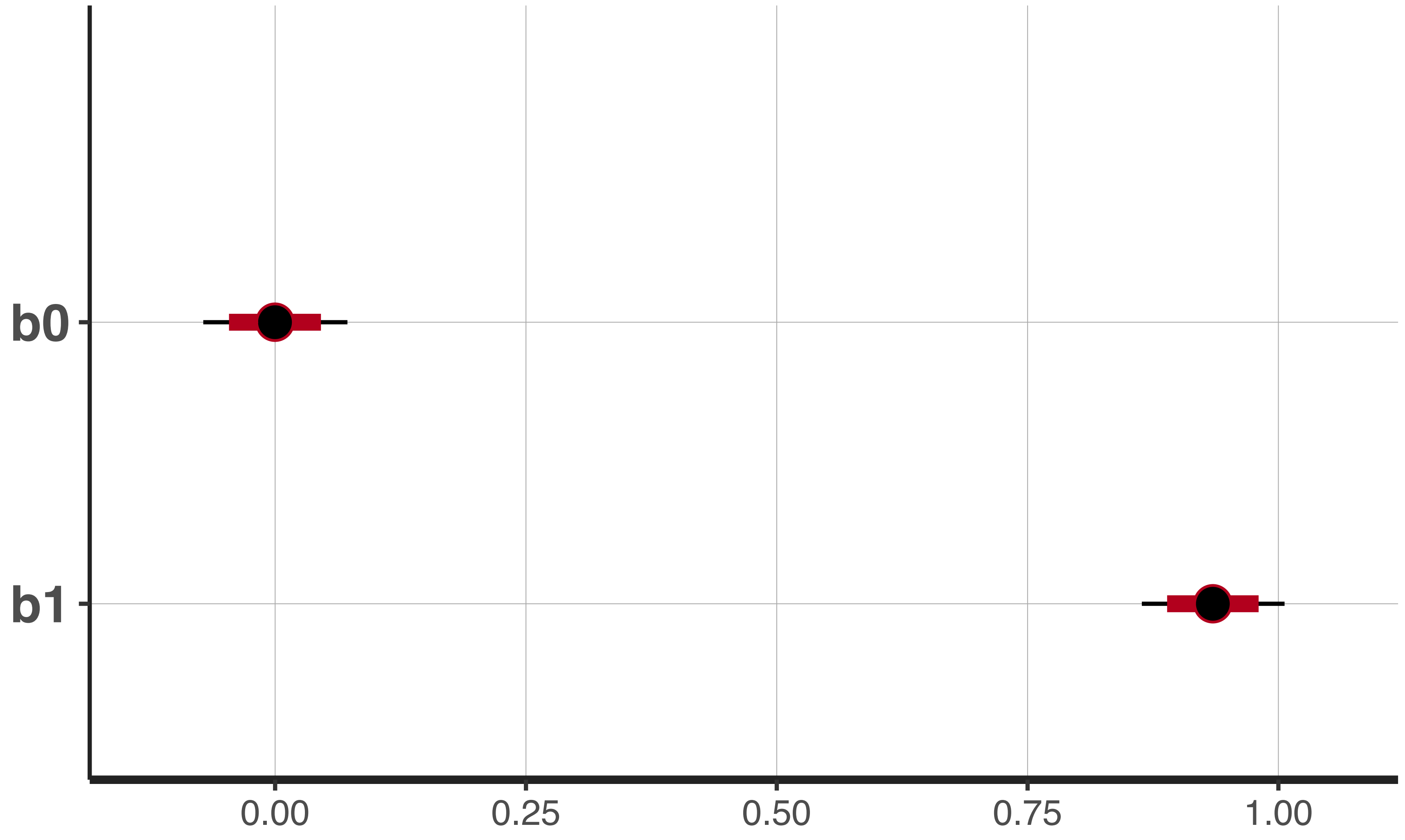
Example: Effect of leg length on height

```
cor(legs)

            lleg      rleg    height
lleg   1.0000000 0.9962084 0.9358207
rleg   0.9962084 1.0000000 0.9363314
height 0.9358207 0.9363314 1.0000000
```

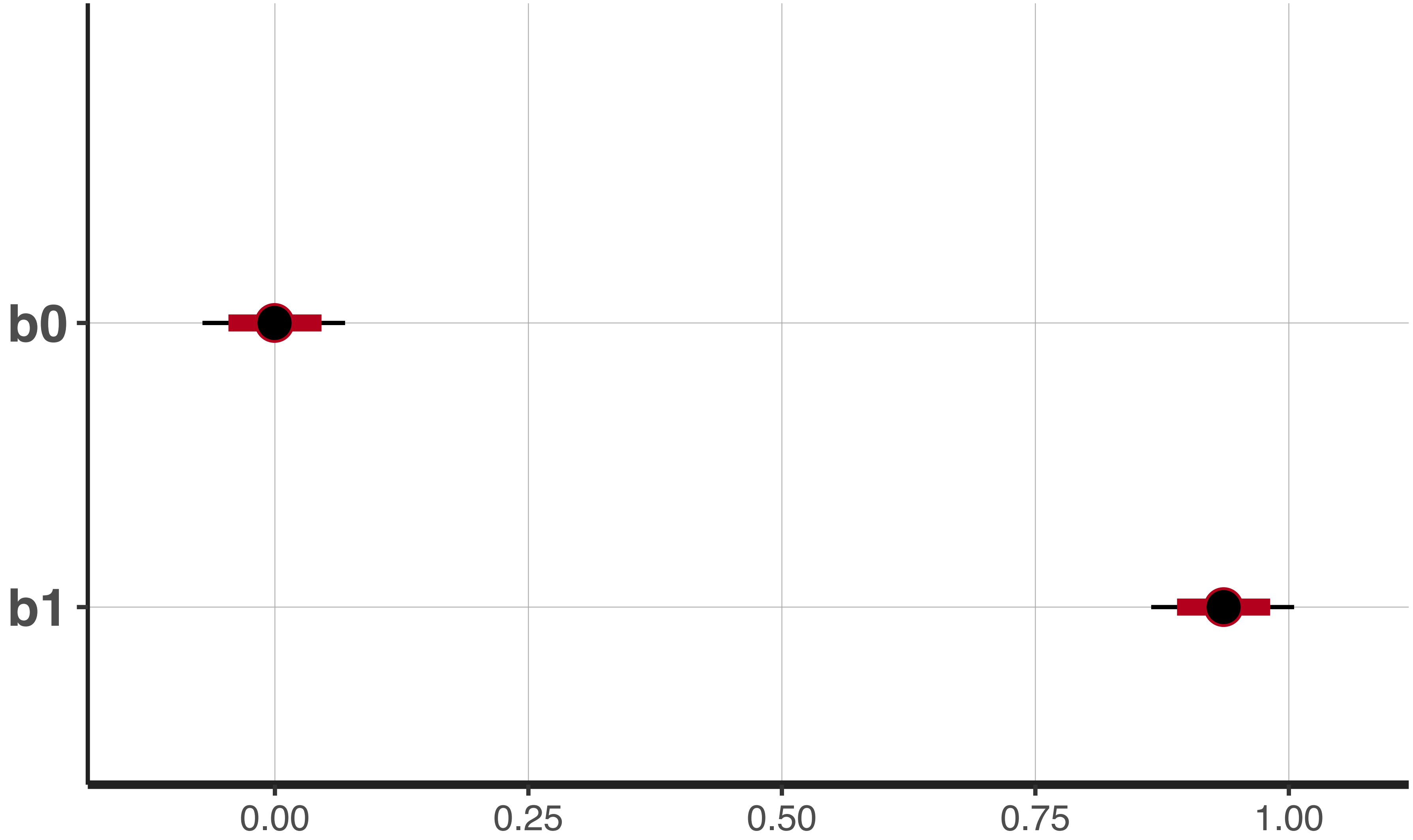
Model With Just Left Leg

$$height \sim \mathrm{normal}(mu, sd)$$
$$mu = \beta_0 + \beta_1 \mathrm{lleg}$$



Model With Just Right Leg

$$height \sim \text{normal}(mu, sd)$$
$$mu = \beta_0 + \beta_1 \text{rleg}$$



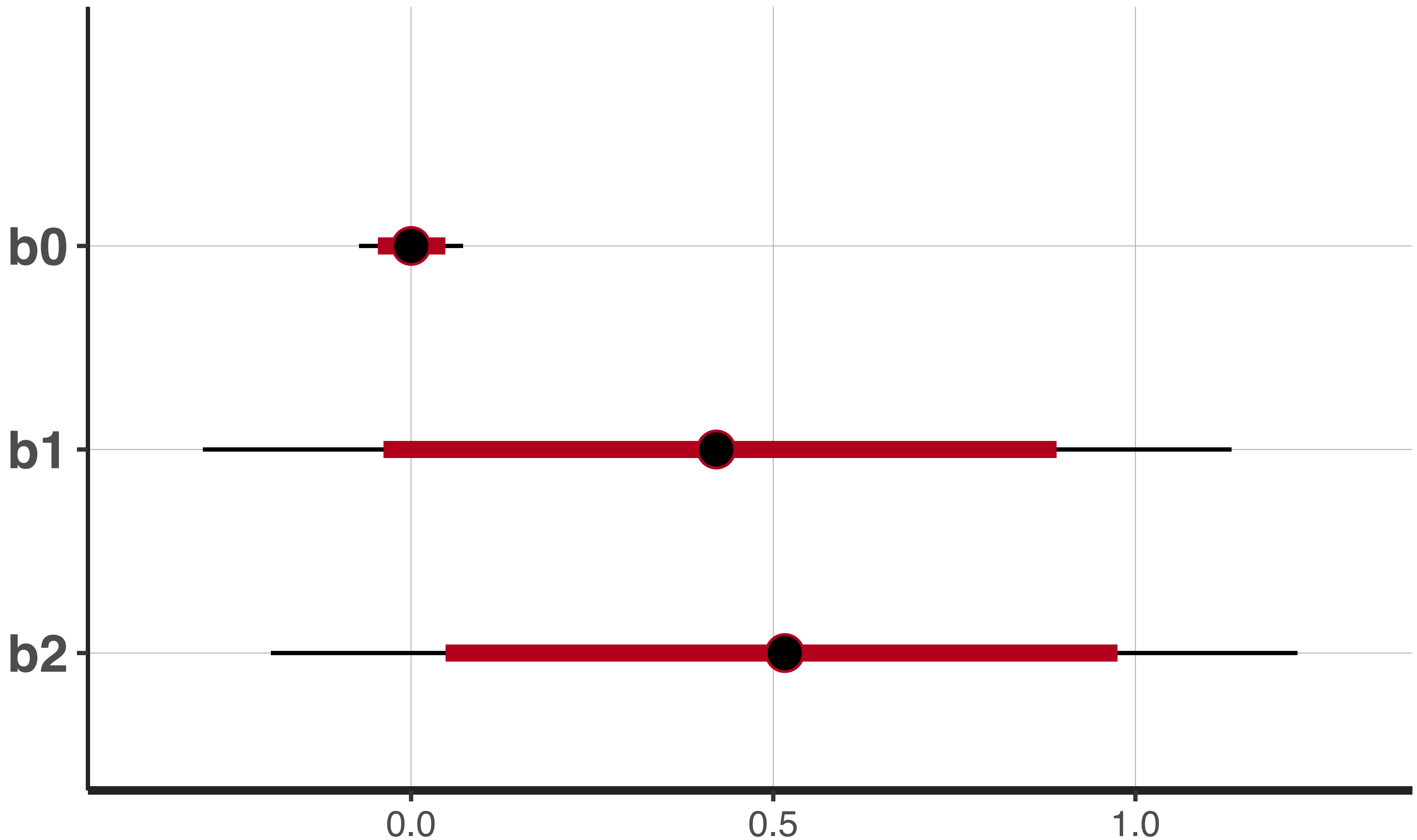
Model With Both Legs

$$height \sim \mathrm{normal}(mu, sd)$$
$$mu = \beta_0 + \beta_1 \mathrm{lleg} + \beta_2 \mathrm{lleg}$$

- What do you think will happen?

Model With Both Legs

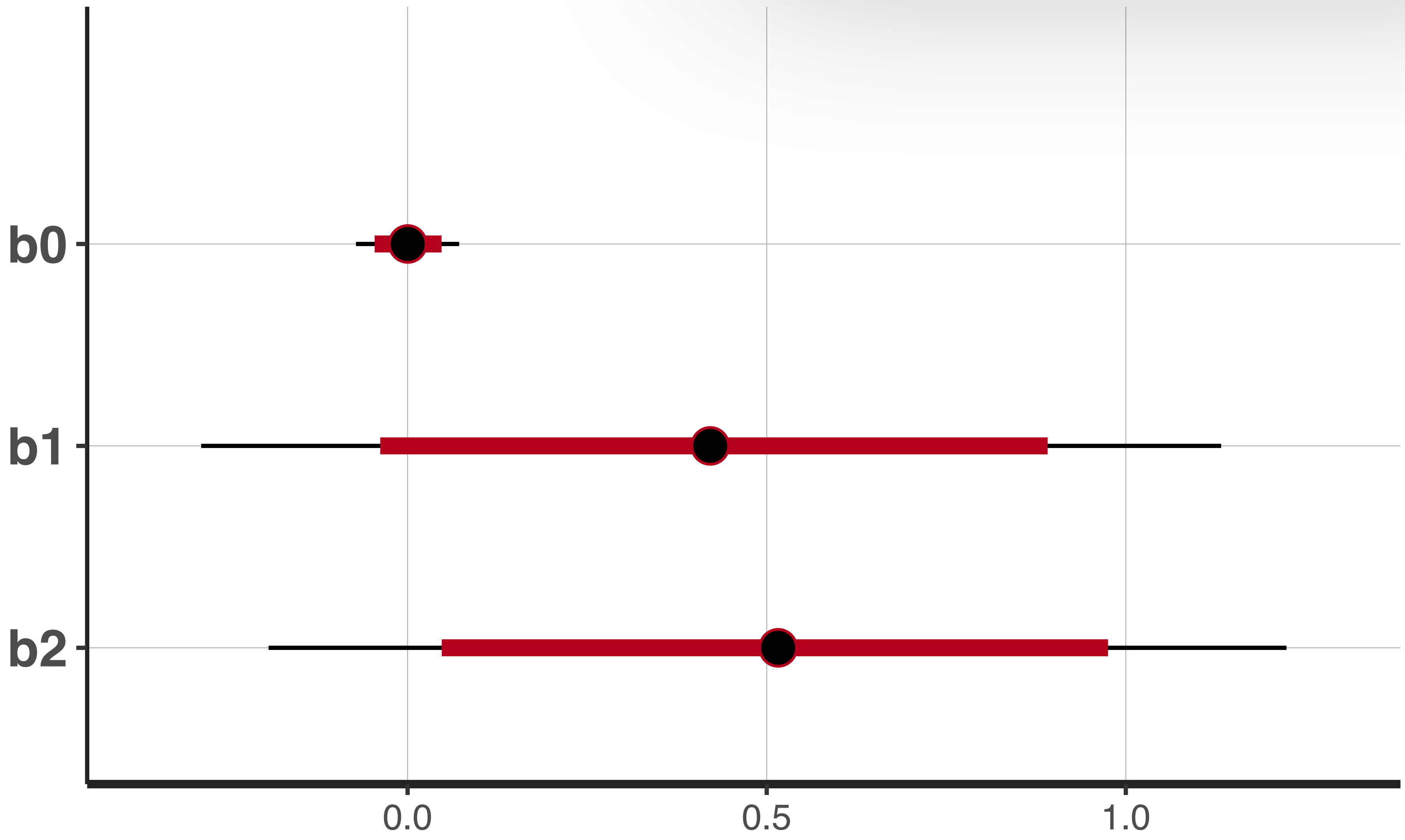
$$height \sim \mathrm{normal}(mu, sd)$$
$$mu = \beta_0 + \beta_1 \mathrm{lleg} + \beta_2 \mathrm{lleg}$$



Model With Both Legs

$$height \sim \text{norr}$$
$$mu = \beta_0 + \beta_1$$

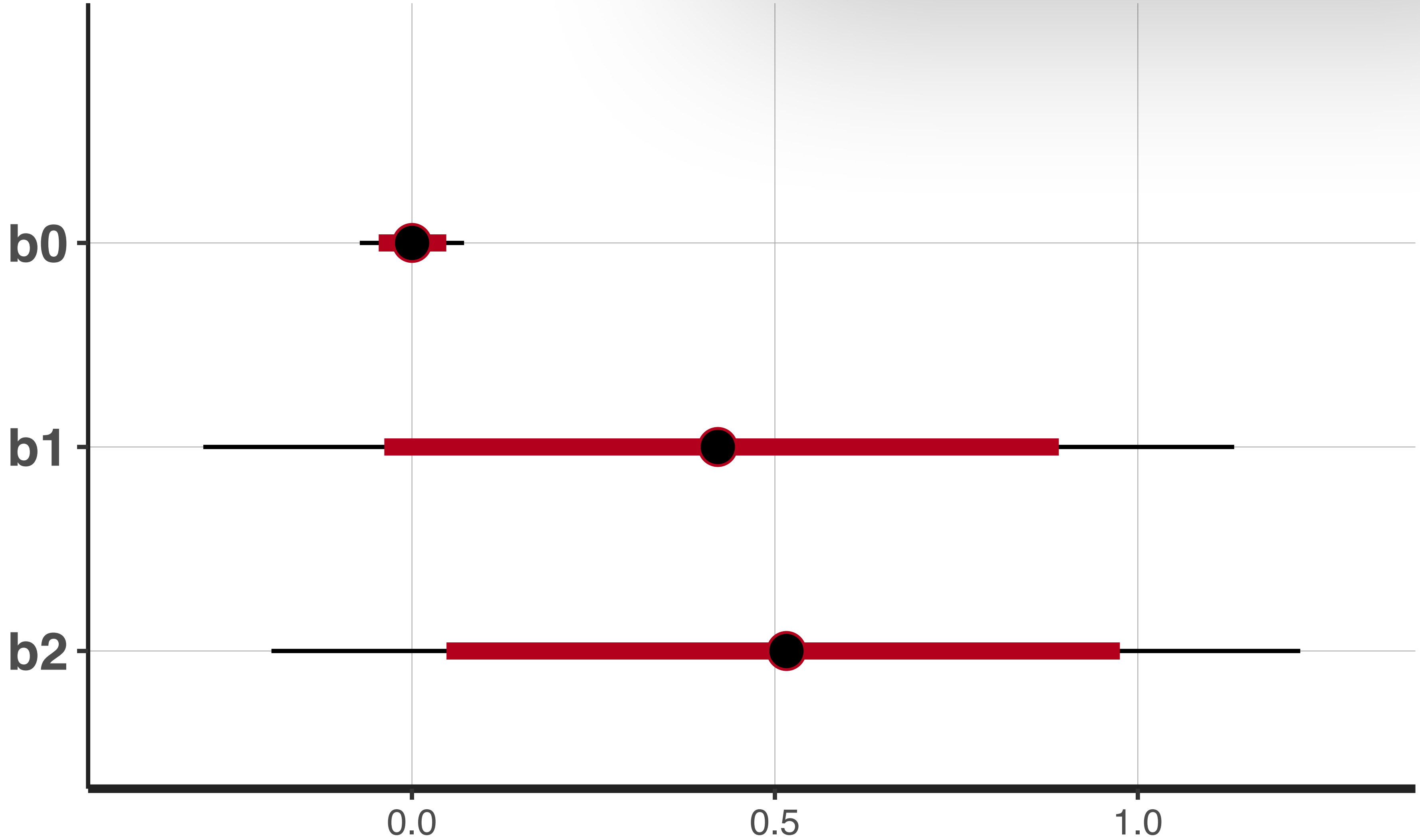
What happened?!!



Model With Both Legs

$$height \sim norm$$
$$mu = \beta_0 + \beta_1$$

Very large ranges are a sign you have a problem

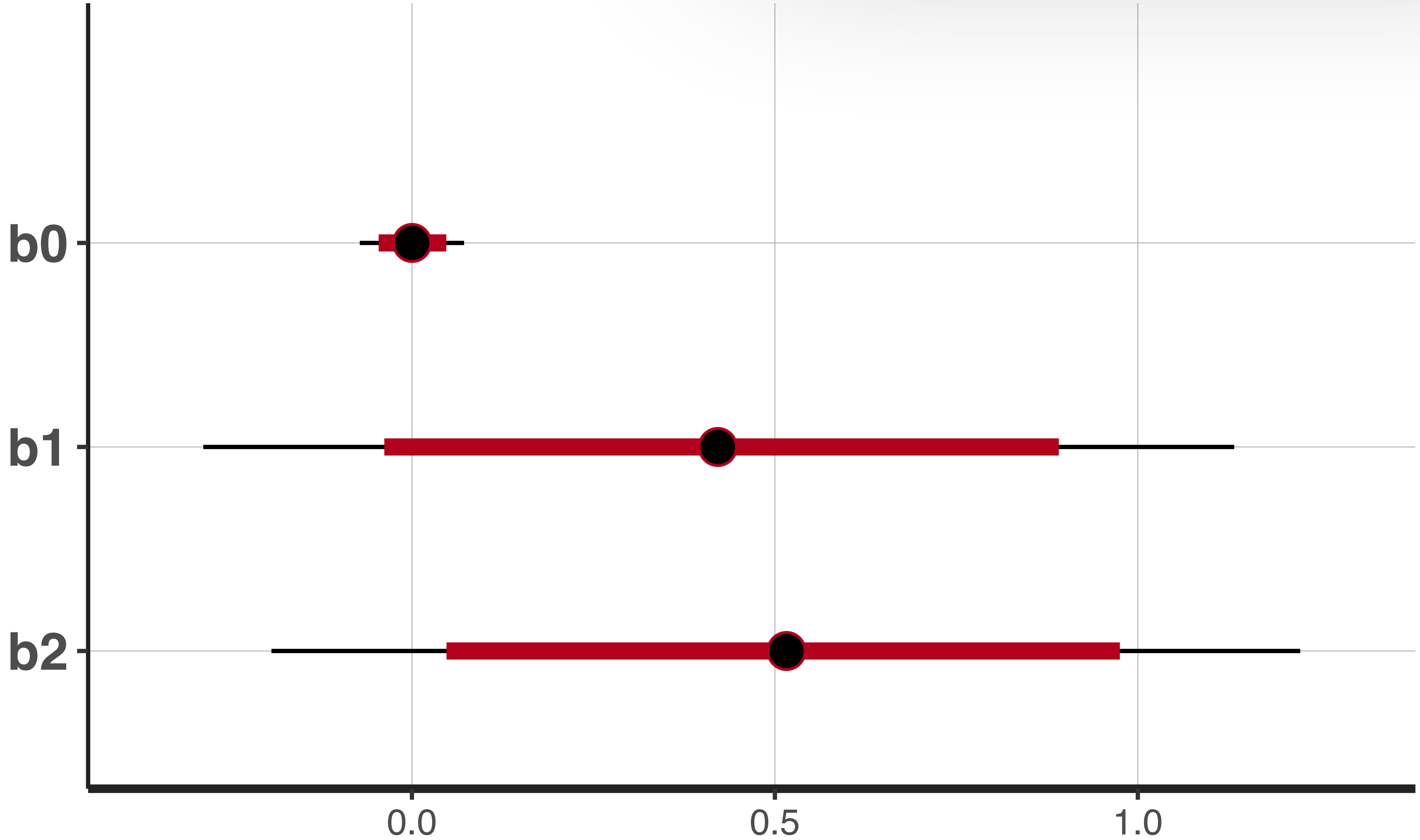


Model With Both Legs

$height \sim \text{norr}$

$mu = \beta_0 + \beta_1 \text{l}$

How would these typically been interpreted?!



Model With Both Legs

$$height \sim \text{norr}$$
$$mu = \beta_0 + \beta_1\text{l}$$

How would these typically been interpreted?!

```
model = lm(zheight ~ zlleg + zrleg)
summary(model)

Call:
lm(formula = zheight ~ zlleg + zrleg)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.70416 -0.23767 -0.00562  0.23696  0.66119 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -6.622e-16  3.530e-02   0.000    1.000
zlleg        4.016e-01  4.077e-01   0.985    0.327
zrleg        5.363e-01  4.077e-01   1.315    0.192

Residual standard error: 0.353 on 97 degrees of freedom
Multiple R-squared:  0.8779,Adjusted R-squared:  0.8754 
F-statistic: 348.8 on 2 and 97 DF,  p-value: < 2.2e-16
```

What Can You Do?

- **Always** plot data first, to see if any potential correlations
- **Always** check for correlations among predictor variables before fitting your model
    - Be wary of correlations > ~0.7
    - Check model performance with each, then with both

Creates Problems!

1. Proper interpretation of coefficients
2. Deciding what predictor variables are important

Frequentist Approach

- Which model should we use/trust?
- “Stargazing”
  - Fit a bunch of predictor variables (and perhaps interactions) and keep those that are significant

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     1057.8982    44.3287  23.865   <2e-16 ***
guber$StuTeaRat   -4.6394     2.1215  -2.187   0.0339 *
guber$Salary       2.5525     1.0045   2.541   0.0145 *
guber$PrcntTake   -2.9134     0.2282 -12.764   <2e-16 ***
```

Frequentist Approach

- Which model should we use/trust?
- "Stargazing"
  - Fit a bunch of predictor variables (and perhaps interactions) and keep those that are significant

- Not all "significant" predictors improve the model
- Some "non-significant" predictors may have important effects
- Plays havoc with interpretation of coefficients (with interactions)

Seque On Goals of Models

- Common diagnostics (*P*-values on parameters and R^2) tell us about fit to current data
  - Surprise - this isn't what we're interested in!!!

Seque On Goals of Models

- Common diagnostics (*P*-values on parameters and R^2) tell us about fit to current data
  - Surprise - this isn't what we're interested in!!!

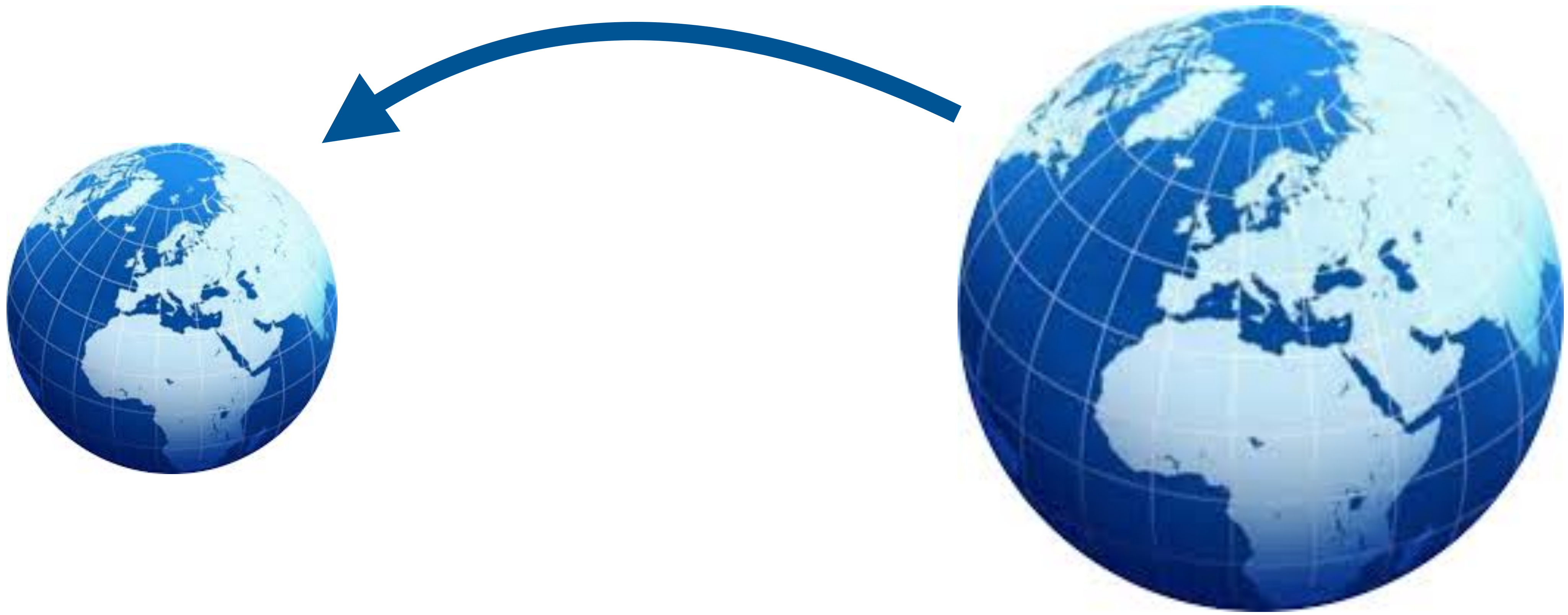
- All data sets contain a combination of information and noise
  - Want to gain as much information as we can from the data without being tricked into thinking that noise is information

Seque On Goals of Models

- "Underfitting"
  - Learn too little from the data

- "Overfitting"
  - Learning "too much" from the data (interpreting noise as information)

Sampling



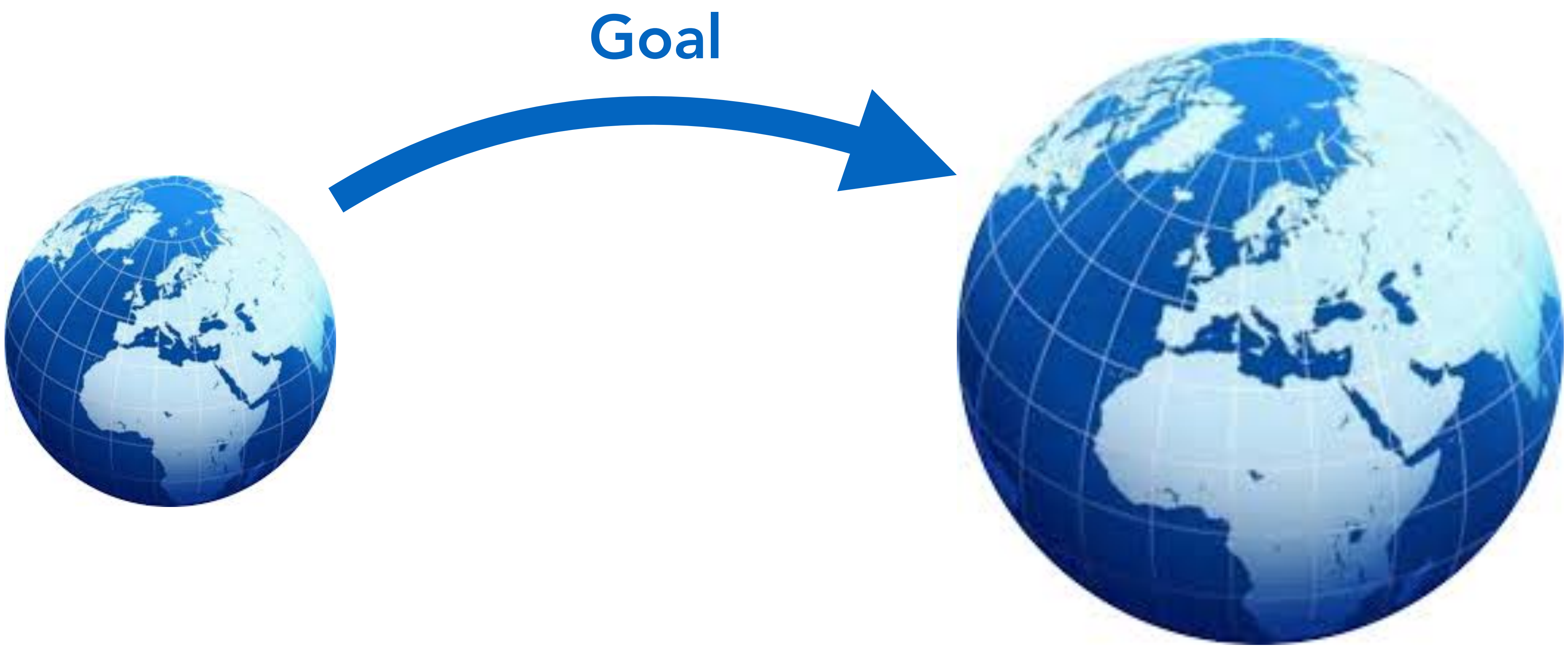
Small world

- Our data set
- Our model

Large world

- The population we are trying to learn about
- The processes we are trying to learn about

Inference



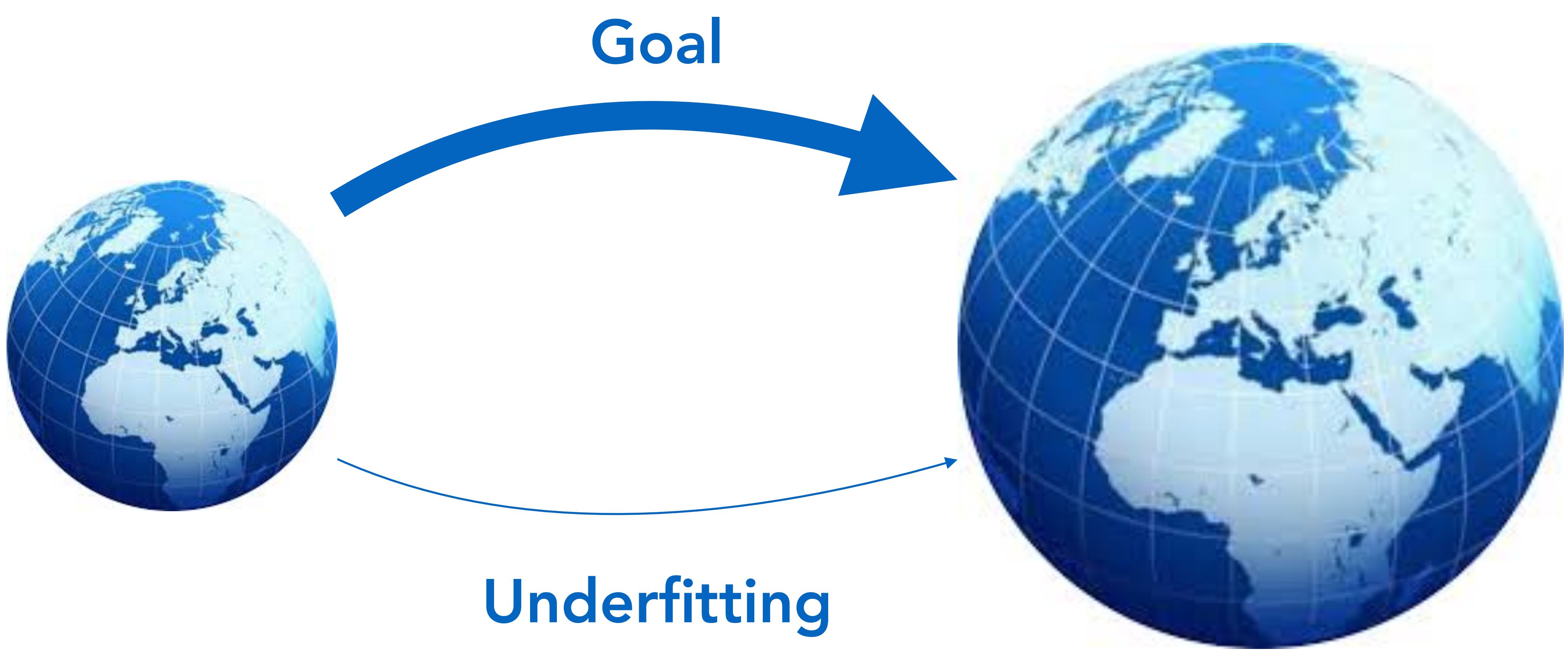
Small world

- Our data set
- Our model

Large world

- The population we are trying to learn about
- The processes we are trying to learn about

Inference



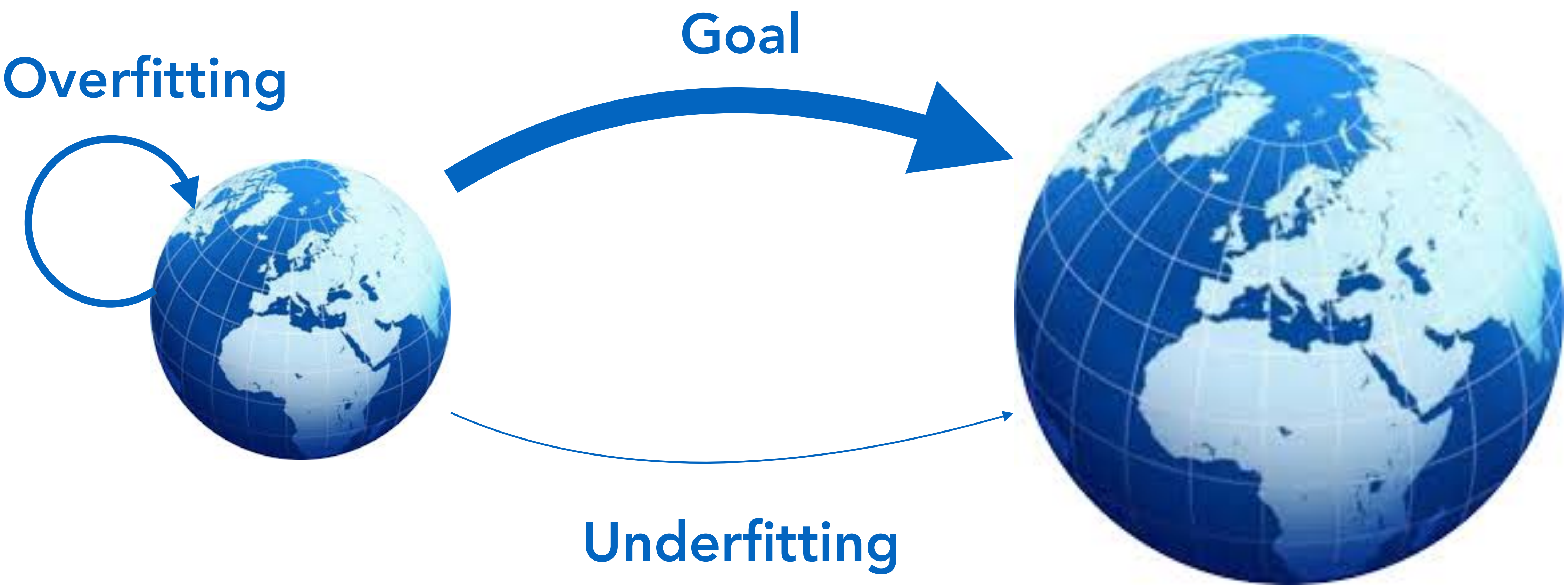
Small world

- Our data set
- Our model

Large world

- The population we are trying to learn about
- The processes we are trying to learn about

Inference



Small world

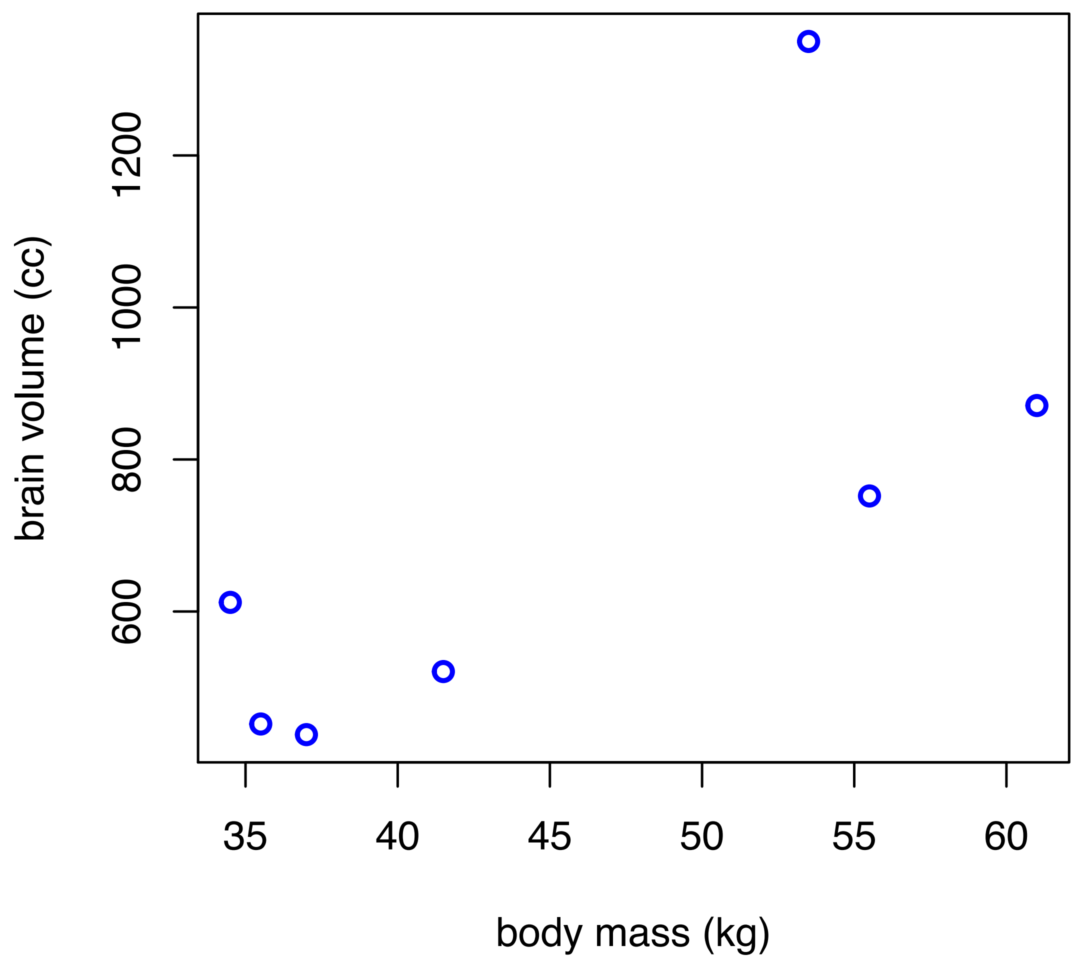
- Our data set
- Our model

Large world

- The population we are trying to learn about
- The processes we are trying to learn about

Example

- Brain volume (in cubic centimetres) and body mass (kilograms) information for 6 hominin species

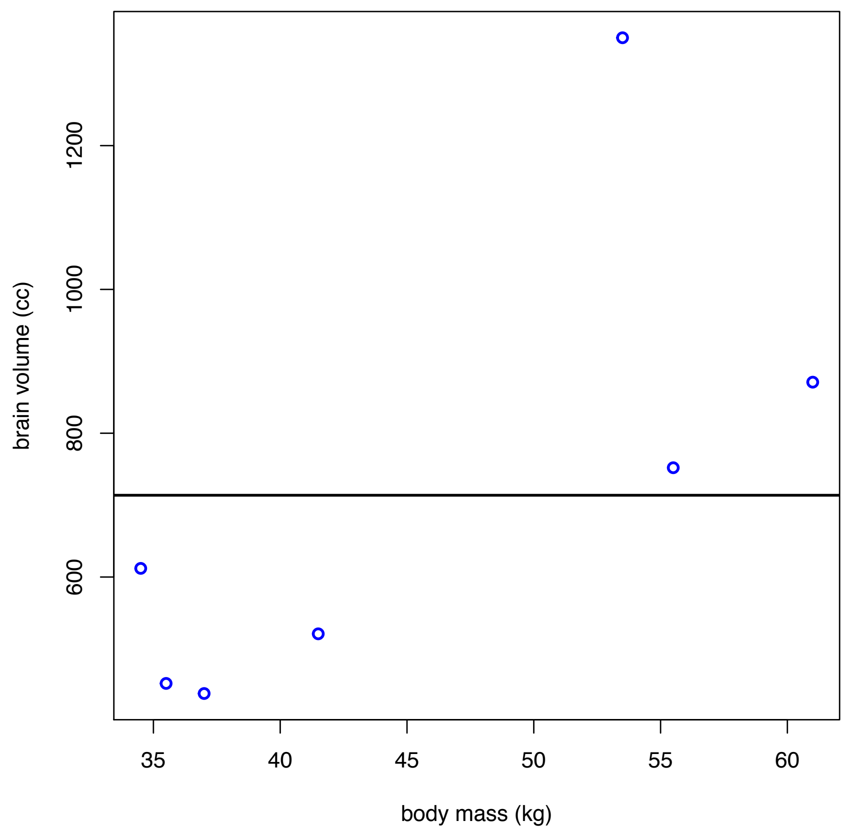


Example idea from McElreath (2016) *Statistical Rethinking*. CRC Press, Boca Raton, FL.
Data from McHenry & Coffing (2000) *Annual Review of Anthropology* **29**: 125-146.

Example

- Simplest model (severe underfitting)

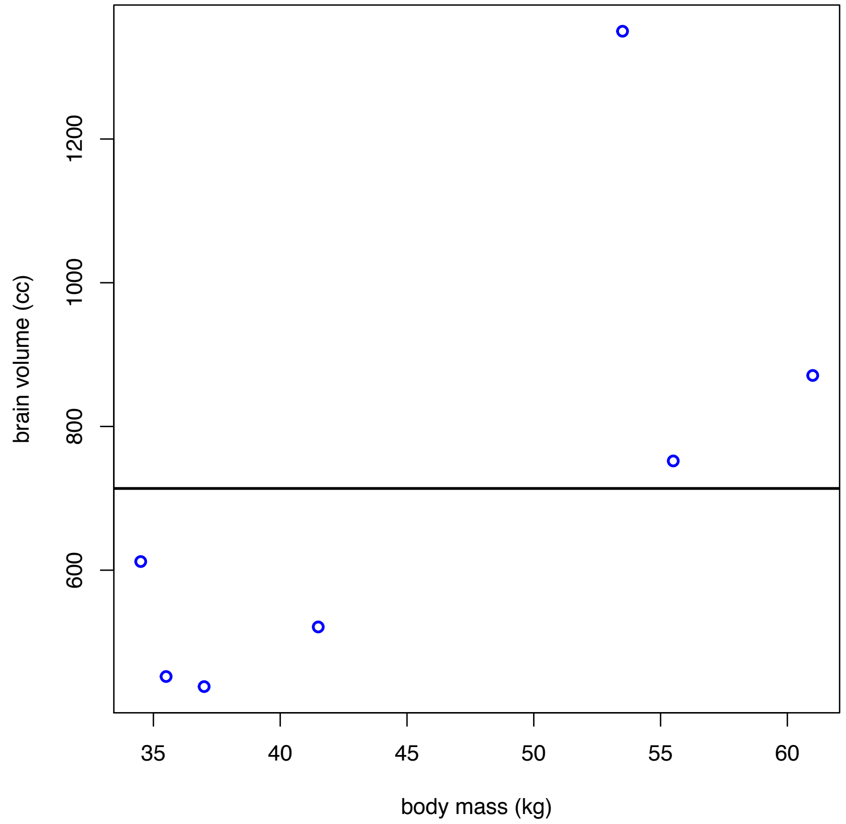
$$y = \alpha$$



Example

- Simplest model (severe underfitting)

$$y = \alpha$$



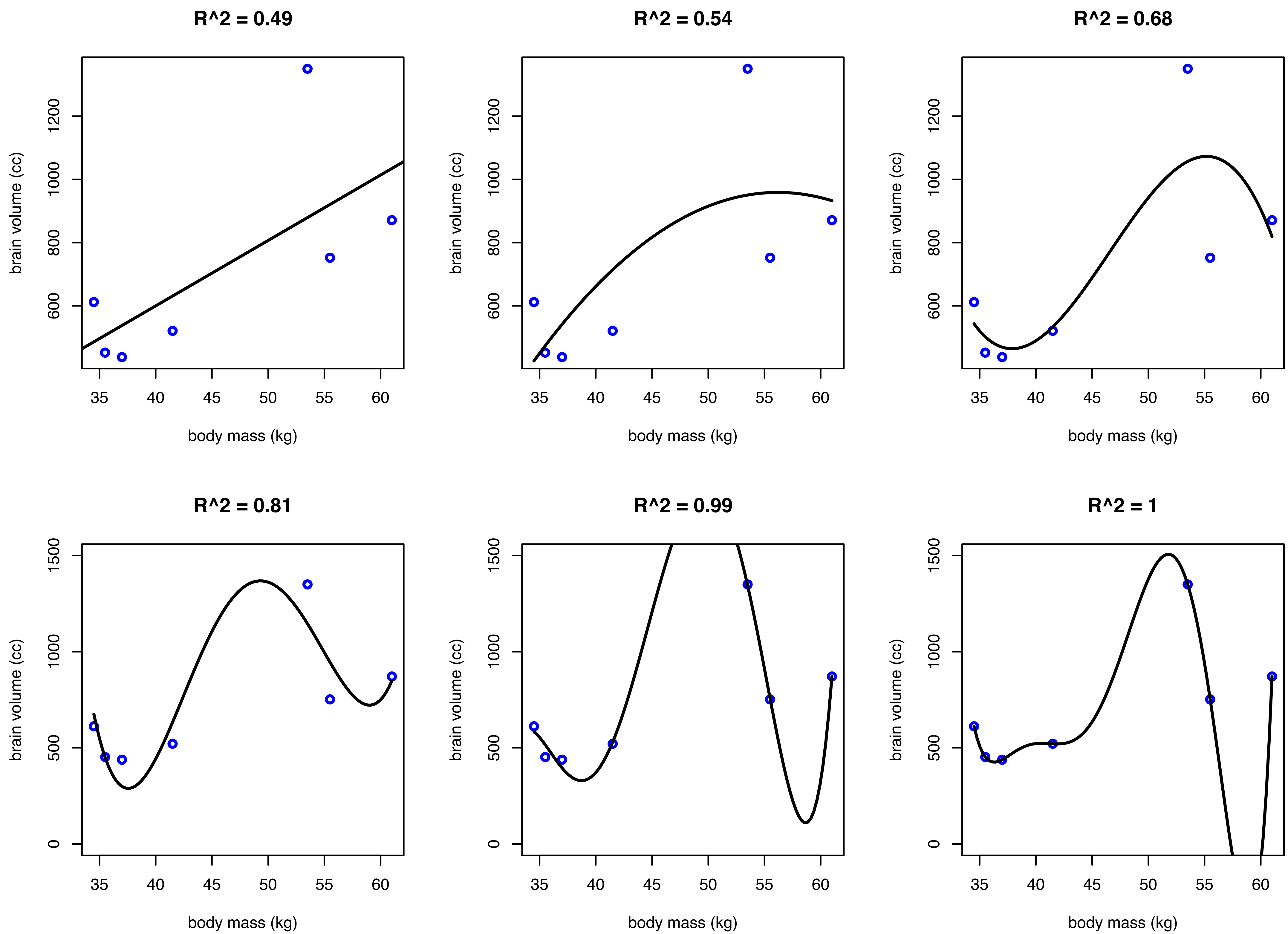
Try

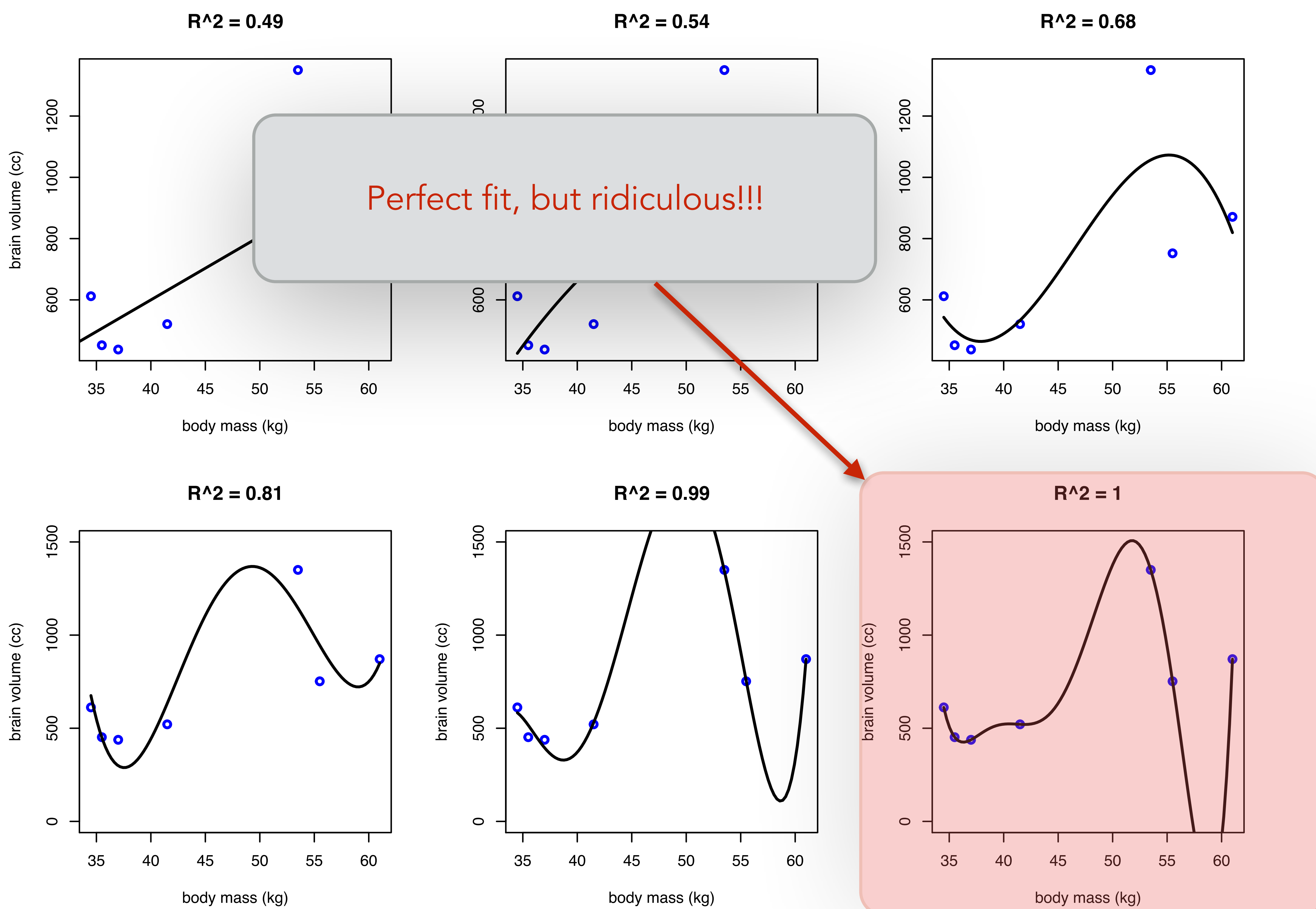
$$y = \alpha + \beta_1 x_1$$

$$y = \alpha + \beta_1 x_1 + \beta_2 x_1^2$$

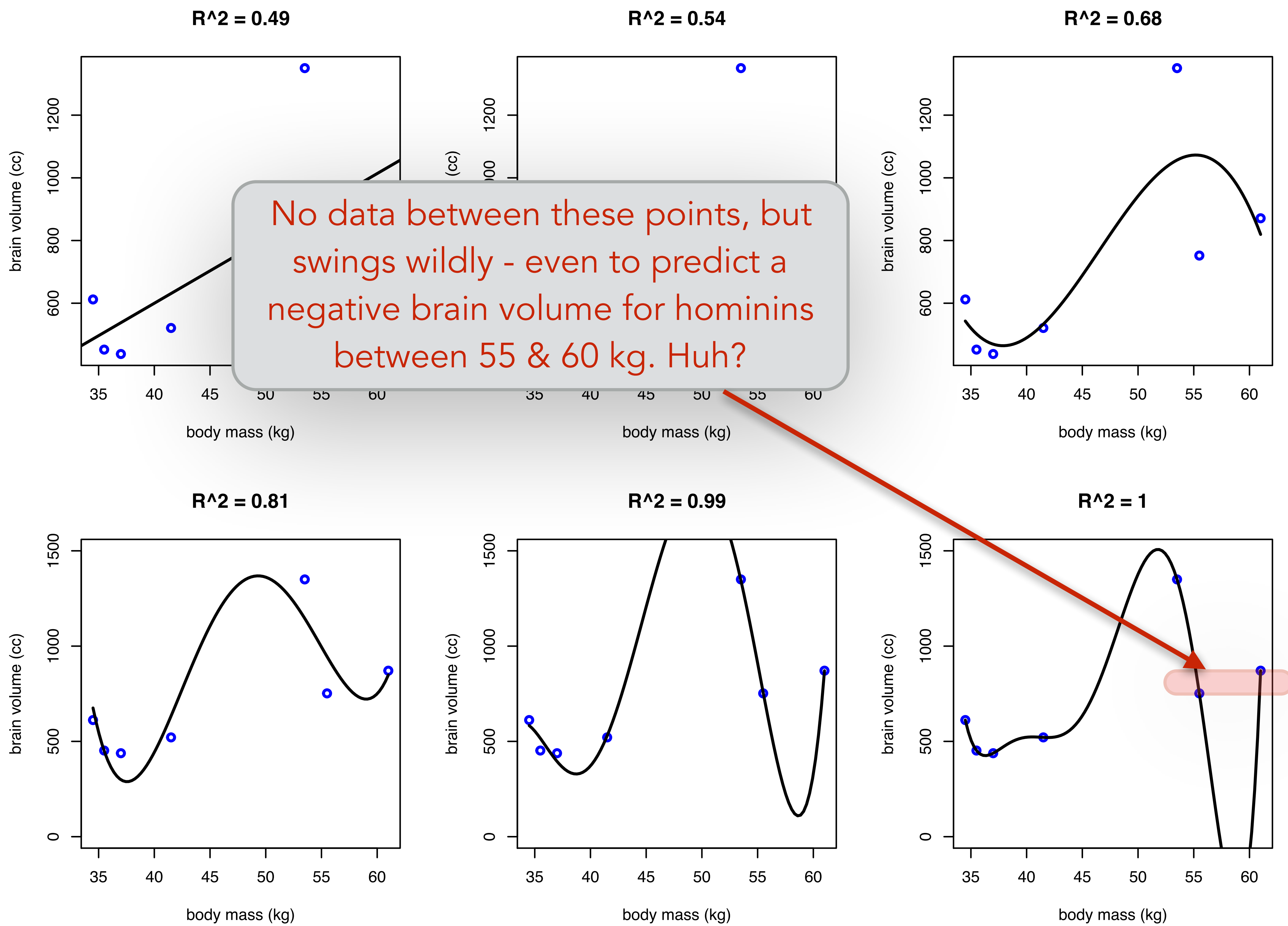
$$y = \alpha + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_1^3$$

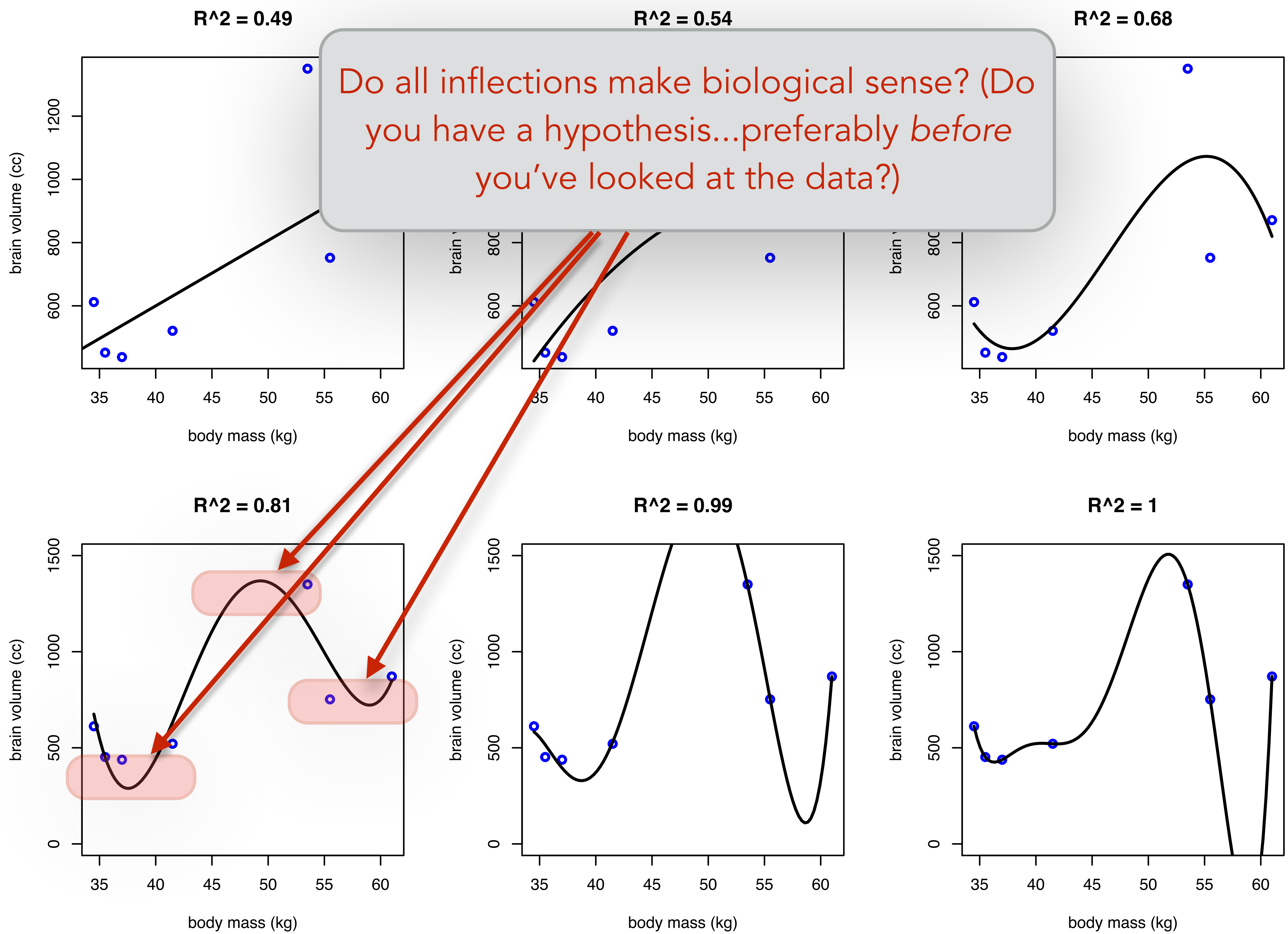
$$\ldots$$


R^2 = 0.49
R^2 = 0.54
R^2 = 0.68
R^2 = 0.81
R^2 = 0.99
R^2 = 1
brain volume (cc)
body mass (kg)
600
800
1000
1200
0
500
1000
1500
35
40
45
50
55
60

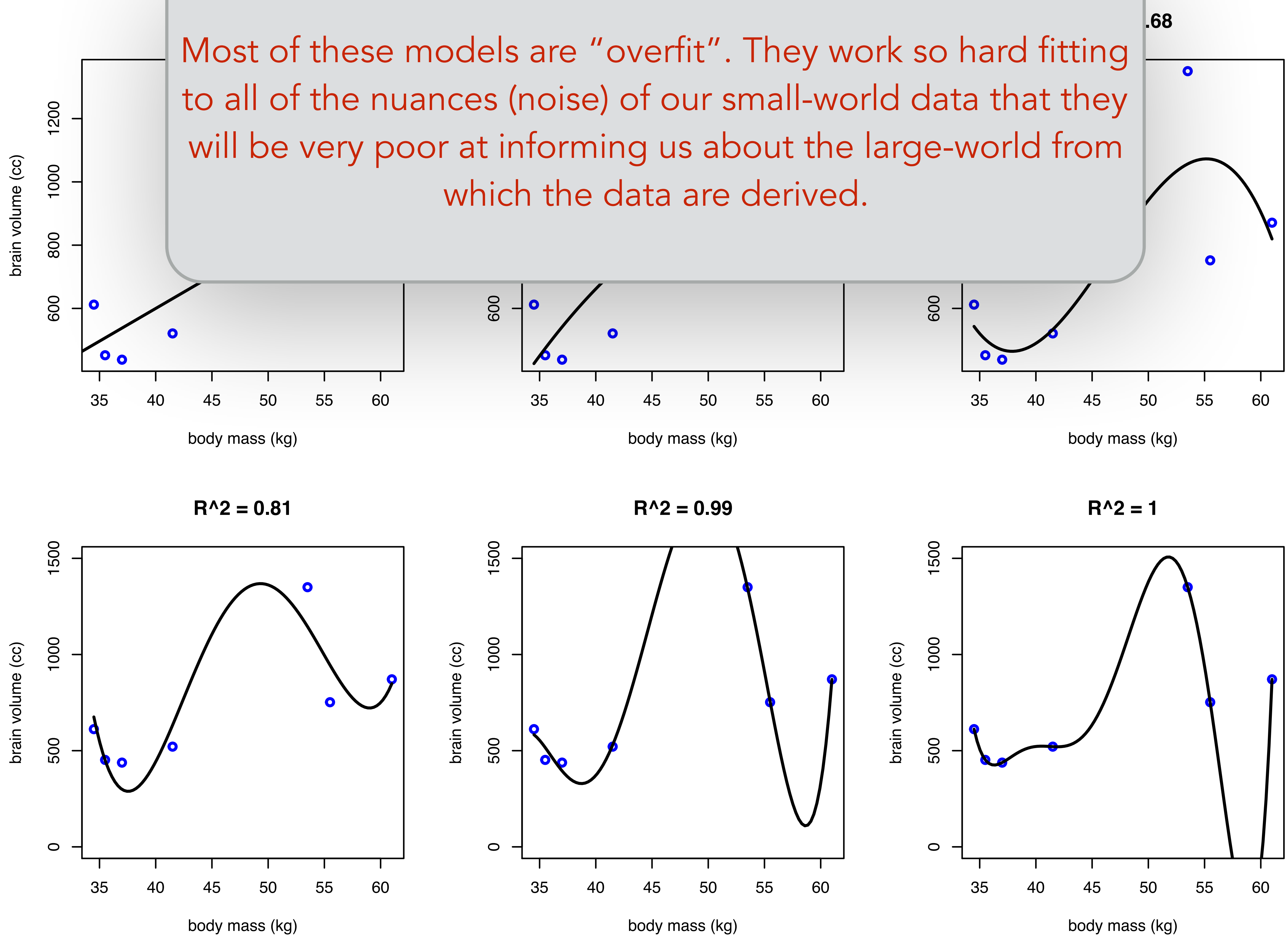


R^2 = 0.49
R^2 = 0.54
R^2 = 0.68
R^2 = 0.81
R^2 = 0.99
R^2 = 1
Perfect fit, but ridiculous!!!
brain volume (cc)
body mass (kg)

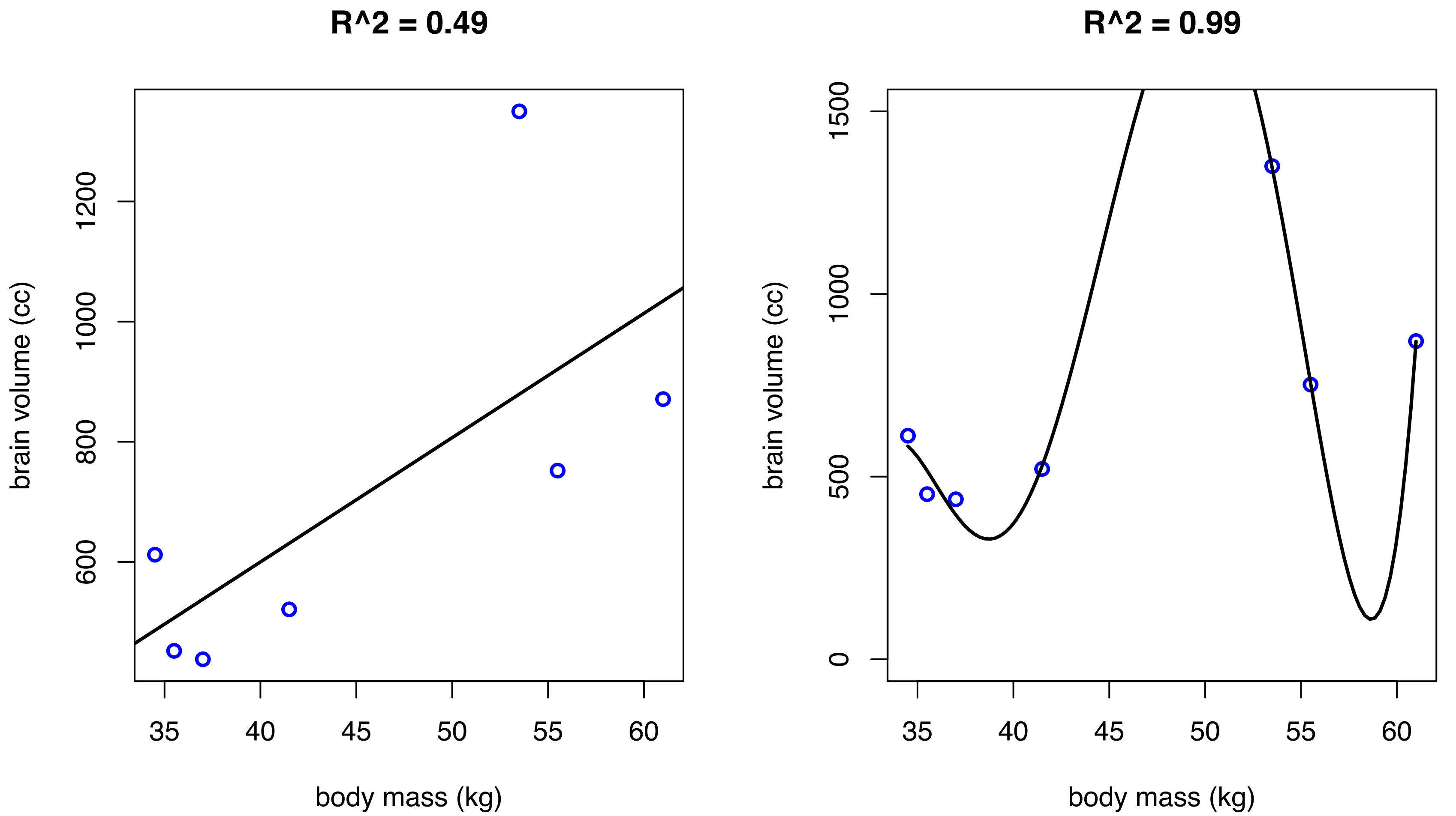

R^2 = 0.49
R^2 = 0.54
R^2 = 0.68
R^2 = 0.81
R^2 = 0.99
R^2 = 1
brain volume (cc)
body mass (kg)
No data between these points, but swings wildly - even to predict a negative brain volume for hominins between 55 & 60 kg. Huh?


R^2 = 0.49
R^2 = 0.54
R^2 = 0.68
R^2 = 0.81
R^2 = 0.99
R^2 = 1
brain volume (cc)
body mass (kg)
Do all inflections make biological sense? (Do you have a hypothesis...preferably *before* you've looked at the data?)

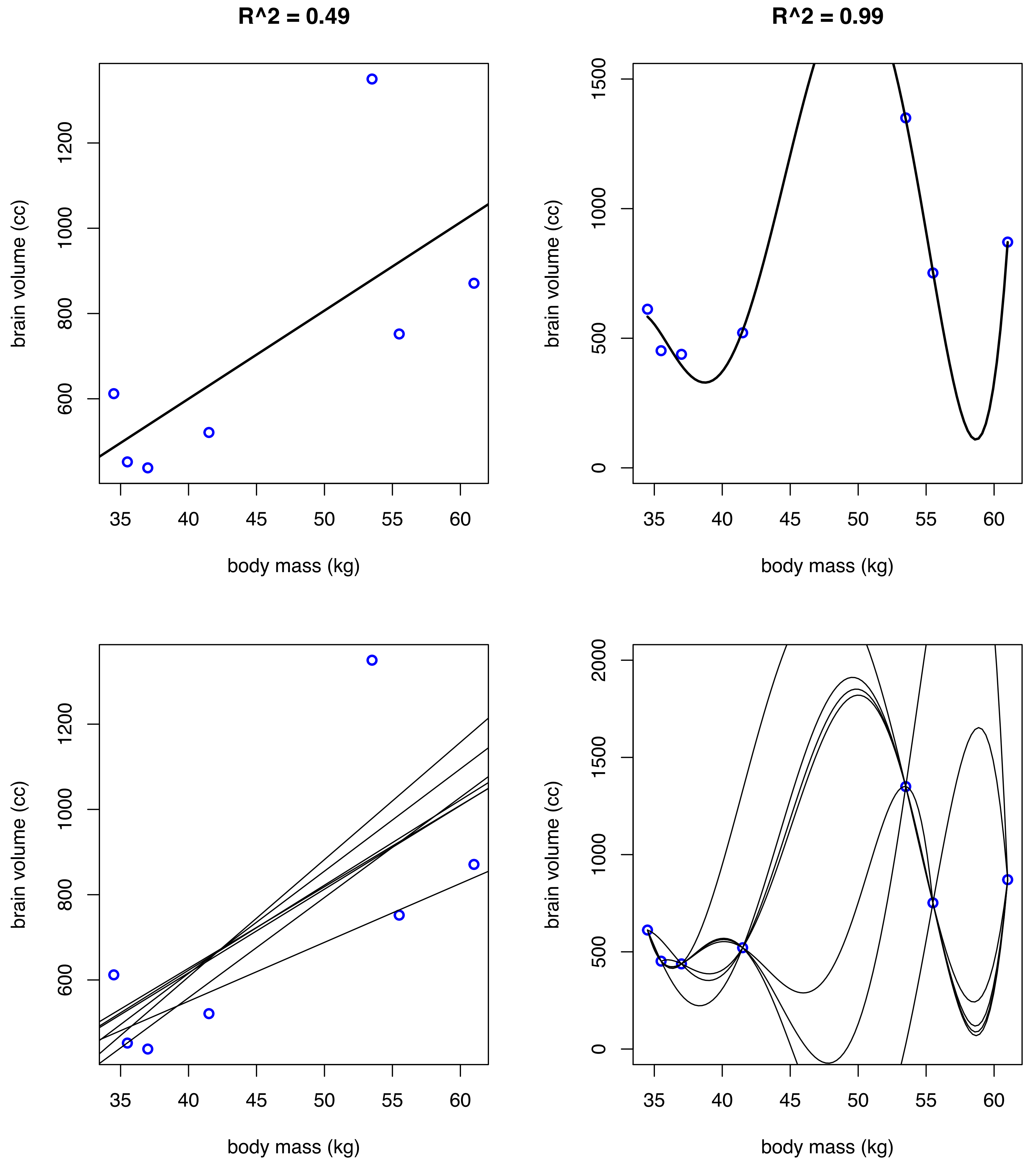
Most of these models are “overfit”. They work so hard fitting to all of the nuances (noise) of our small-world data that they will be very poor at informing us about the large-world from which the data are derived.



- Can visualize this by removing one point at a time, and seeing how much that changes the pattern (perceived relationships)
  - In an underfit model, removing one point will not change the patterns
  - In an overfit model, removing one point will change the patterns dramatically

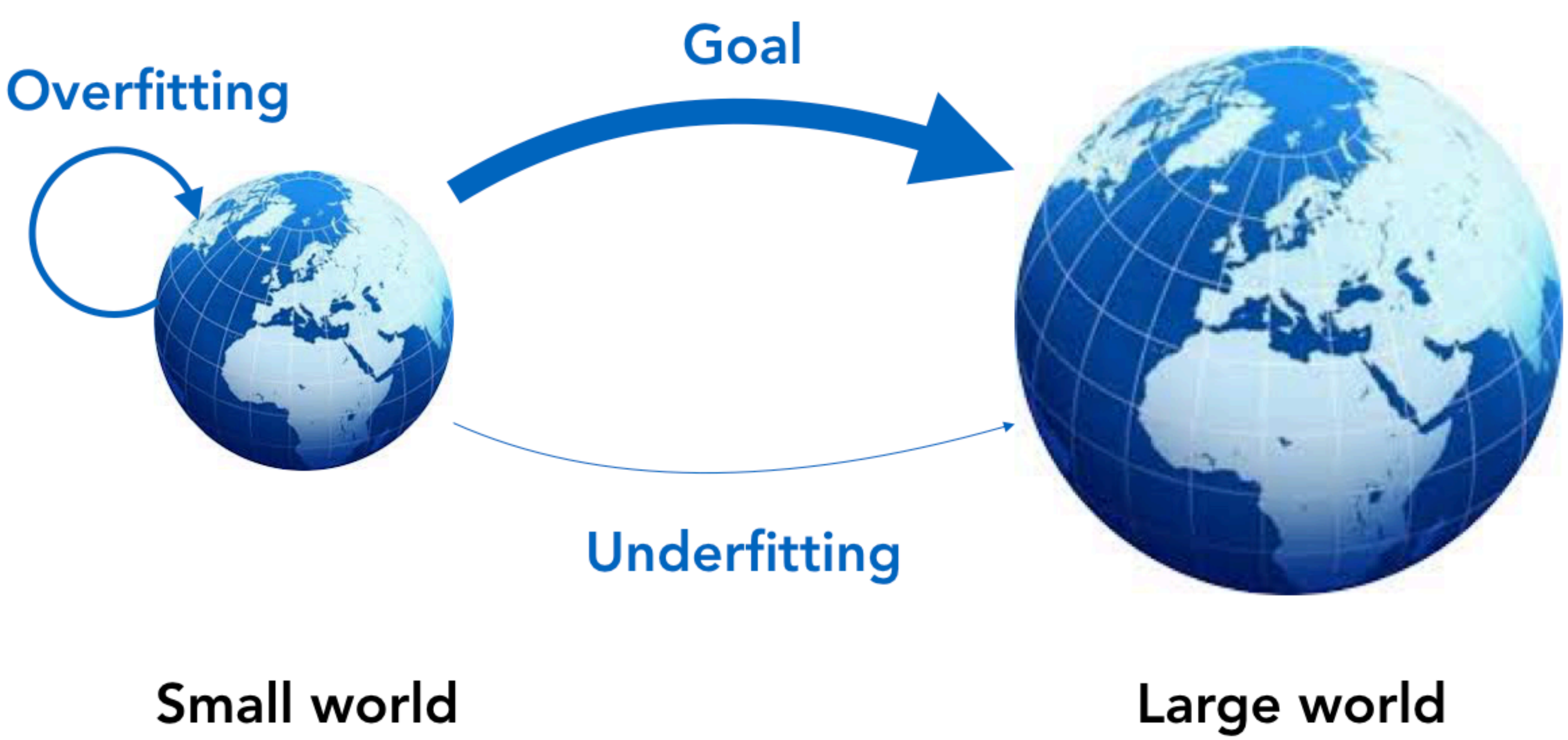


R^2 = 0.49
brain volume (cc)
600
800
1000
1200
35
40
45
50
55
60
body mass (kg)
R^2 = 0.99
brain volume (cc)
0
500
1000
1500
35
40
45
50
55
60
body mass (kg)


R^2 = 0.49
R^2 = 0.99
brain volume (cc)
body mass (kg)
600
800
1000
1200
0
500
1000
1500
2000
35
40
45
50
55
60

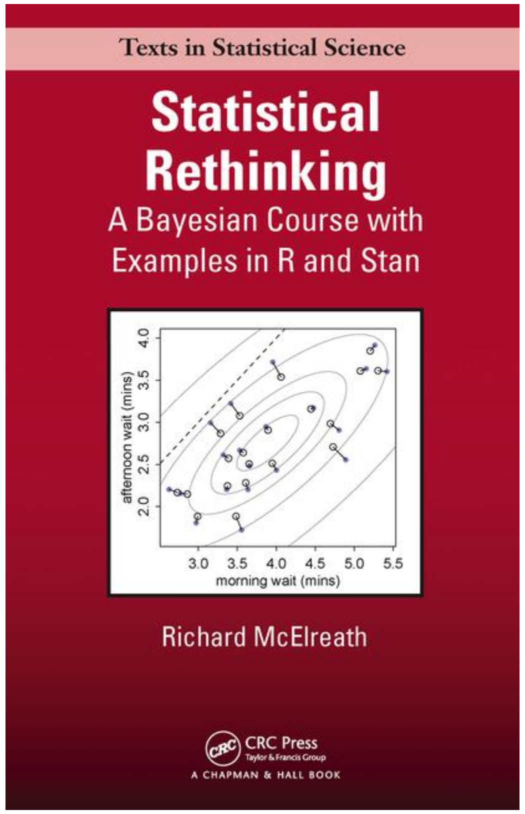
- Ideally, have a method to quantify how well our model will perform "out of sample"

- This is the idea of information criteria



Information Criteria

- Details can be found elsewhere
  - e.g., Chapter 6 in *Statistical Rethinking*

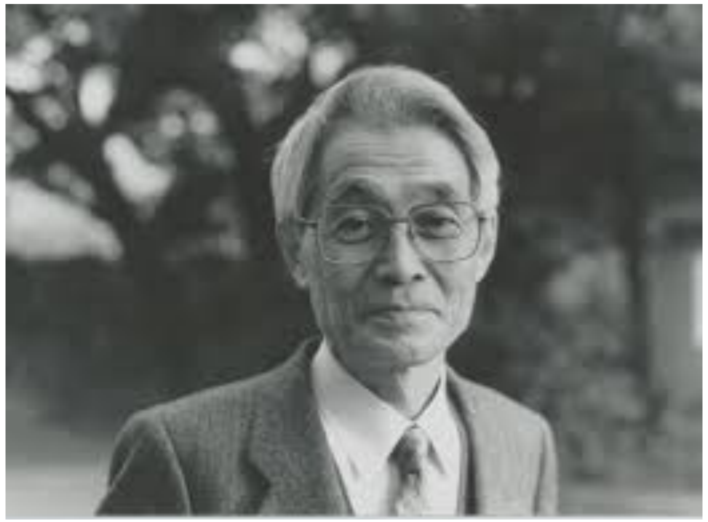


- "Score" nested models on their out-of-sample predictive abilities, with some penalty for adding parameters

A trade-off between the goodness of fit of the model and the complexity of the model.

Information Criteria

- Akaike Information Criterion (AIC)
- Deviance Information Criterion (DIC)
- Widely Applicable Information Criterion (WAIC)
- Others…



Hirotugo Akaike
(1927-2009)

Information Criteria

- Preferred model should be the one with the lowest IC
- Provides information on the relative informativeness of compared models. Tells nothing of the potential truth or usefulness of any
    - All tested models may be poor approximations of the truth
- Think of as less for “model selection” and more as a means to tell you about parameters

Frequentist Approach

Potential solution - AIC

```
model_both = lm(legs$height ~ legs$lleg + legs$rleg)
model_lleg = lm(legs$height ~ legs$lleg)
model_rleg = lm(legs$height ~ legs$rleg)

AIC(model_both, model_lleg, model_rleg)

           df      AIC
model_both  4 578.8851
model_lleg  3 578.6527
model_rleg  3 577.8802
```

Frequentist Approach

Potential solution - AIC

```
model_both = lm(legs$height ~ legs$lleg + legs$rleg)
model_lleg = lm(legs$height ~ legs$lleg)
model_rleg = lm(legs$height ~ legs$rleg)

AIC(model_both, model_lleg, model_rleg)

           df      AIC
model_both  4 578.8851
model_lleg  3 578.6527
model_rleg  3 577.8802
```

Including just the right leg has the lowest AIC.

Should not be used to find "the best model", but rather to understand effects of parameters.

Should **not** be interpreted as "left leg had no effect". It **does** still have an effect, but we need to account for it in a different way.

Questions?