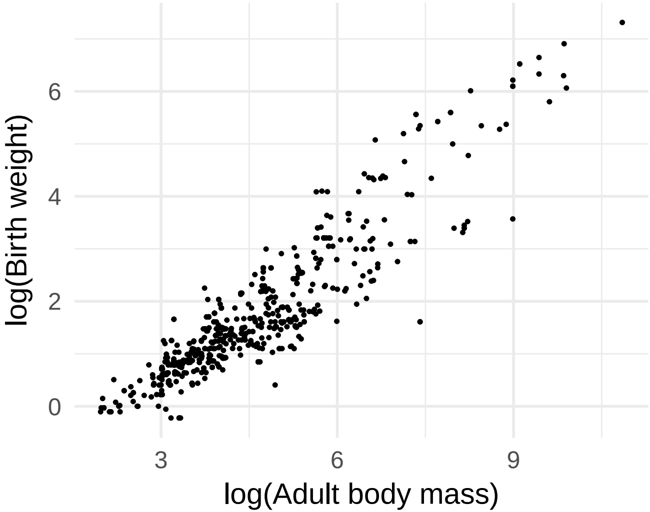
Parameter uncertainty and model criticism for linear regression

Max Joseph

March 09, 2017

Recap

How to predict birth weight?



Assumptions

1. Linear relationship between x and y:

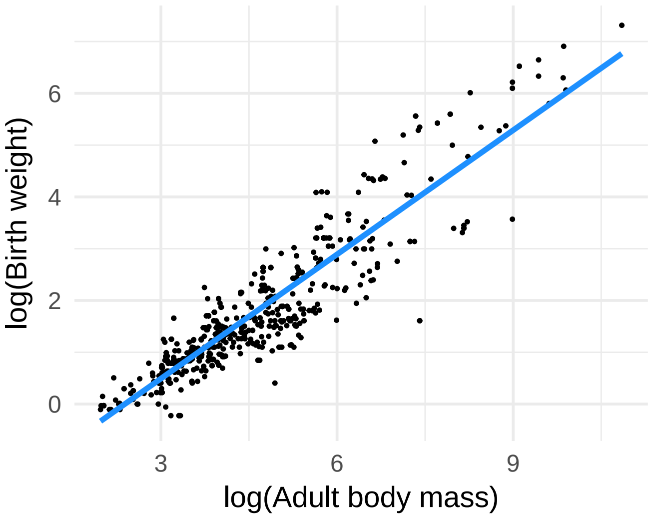
$$\hat{y}_i = \beta_0 + \beta_1 x_i$$

2. Normal error in y (variation around line):

$$y_i \sim \text{Normal}(\hat{y}_i, \sigma)$$

3. Homoskedasticity (errors have constant variance)
4. Observations are independent:

$$-\log(L(y; \beta_0, \beta_1, \sigma)) = -\sum_{i=1}^{n} \log(p(y_i; \beta_0, \beta_1, \sigma))$$


log(Birth weight)
6
4
2
0
3
6
9
log(Adult body mass)

Today: parameter uncertainty and model criticism

1. How uncertain are our estimates?
2. How much can the model explain?

Evaluating uncertainty

Standard error

$$SE = \sqrt{\text{diag}(H^{-1})}$$

where

- H: Hessian matrix and
- diag() extracts diagonal terms in a matrix

Hessians

18th century German auxiliaries

- contracted for military service by the British government
- $\approx$ 25% of British forces in American Revolutionary War
- from the German state of Hesse-Kassel

But what is a Hessian really?

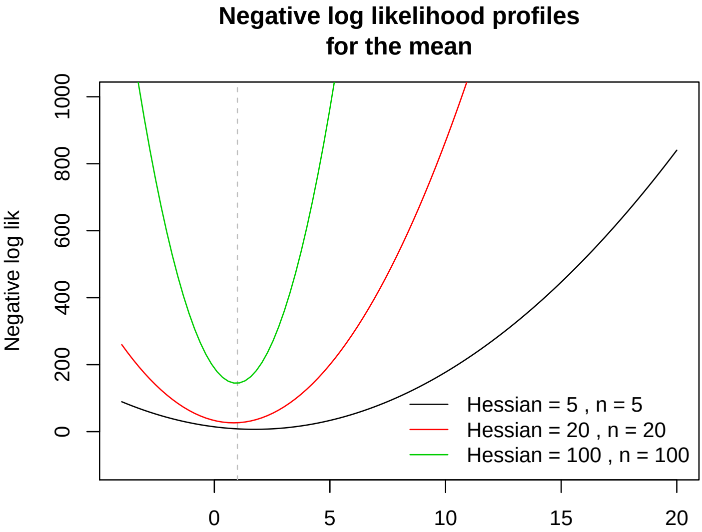
A square matrix of second-order partial deriviatives

$$H = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1\,\partial x_p} \\ \dfrac{\partial^2 f}{\partial x_2\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2\,\partial x_p} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_p\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_p\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_p^2} \end{bmatrix}$$

What is a Hessian graphically?

Curvature of the log likelihood

$\rightarrow$ how much information do we have?



Recap

Hessian tell us about curvature

SE tells us about uncertainty

Computing confidence intervals from SE

Approximation:

$$\text{CI}(\theta) = \theta_{\text{MLE}} \pm t_{\alpha/2}\text{SE}(\theta)$$

```
# e.g., 95% confidence interval
params <- fit$par

sample_size <- length(y)
n_params <- length(params)

crit_t <- qt(0.025, sample_size - n_params)
lower_ci <- params[1] + crit_t * SE[1]
upper_ci <- params[1] - crit_t * SE[1]
```

Assumption (big one)

The log likelihood profile is quadratic

- more reliable for large n

Confidence interval demo

Questions?

$$SE = \sqrt{\mathrm{diag}(H^{-1})}$$

$$\mathrm{CI}(\theta) = \theta_{\mathrm{MLE}} \pm t_{\alpha/2}\mathrm{SE}(\theta)$$

Deep connections to least squares

Minimizing this:

$$-\log(L(y; \beta_0, \beta_1, \sigma)) = -\sum_{i=1}^{n} \log(p(y_i; \beta_0, \beta_1, \sigma))$$

Is almost the same as minimizing this:

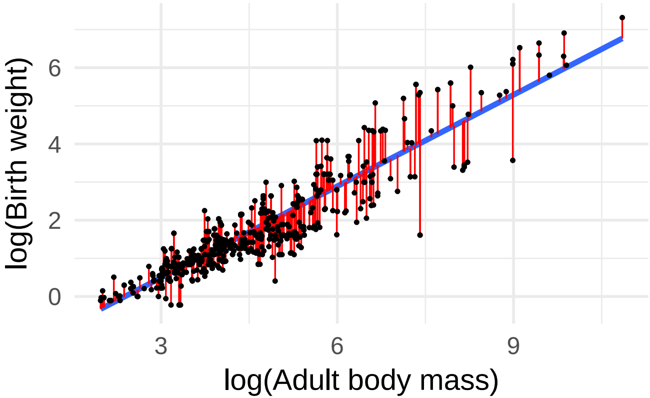
$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Comparing our estimates

Residual sum of squares

Variation in y **around the line**

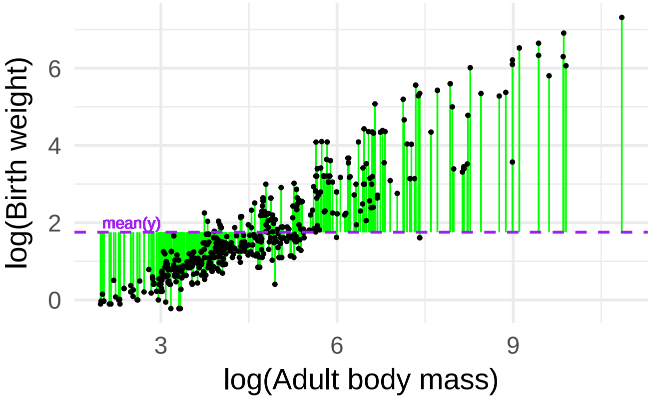
$$SS_R = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$



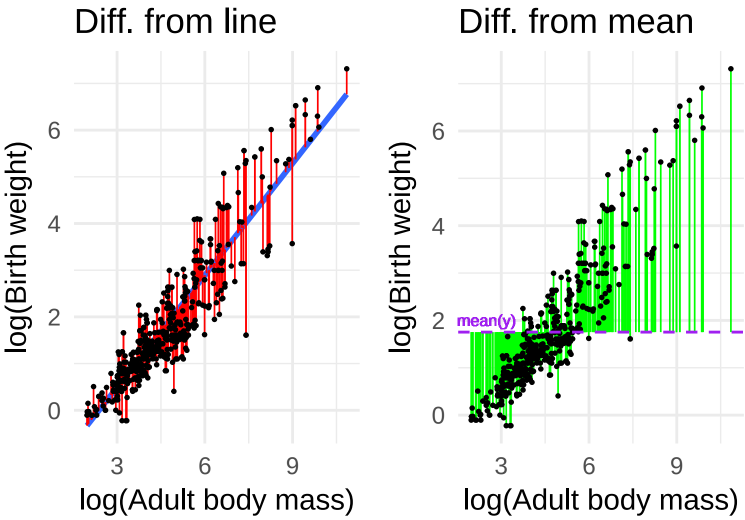
Total sum of squares

Variation in y **relative to the sample mean**

$SS_T = \sum_{i=1}^{n}(y_i - \bar{y})^2$, where $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$



Residual vs. total sum of squares



Residual vs. total sum of squares

$$SS_R = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$SS_T = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

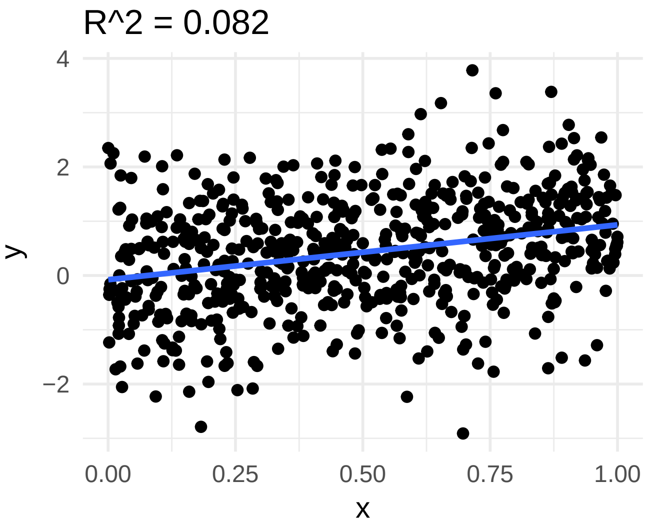
*demo

How much can our model explain?

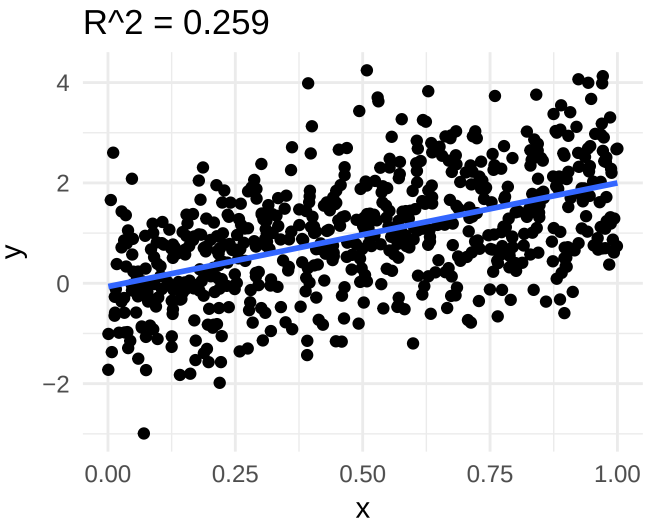
Coefficient of determination

$$R^2 = 1 - \frac{SS_R}{SS_T}$$

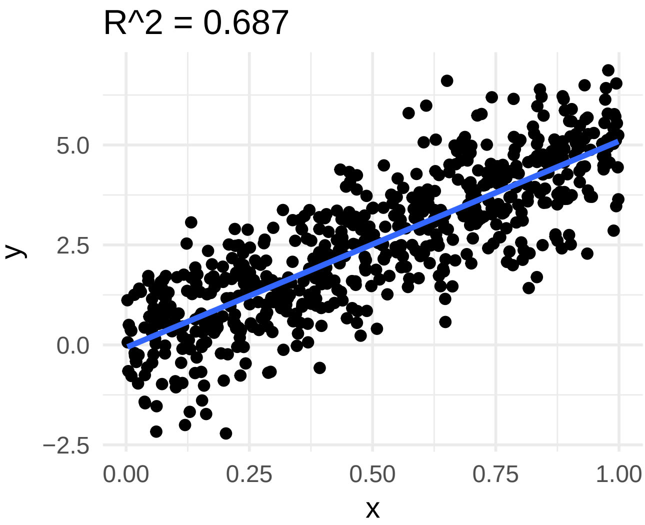
Visualizing R^2



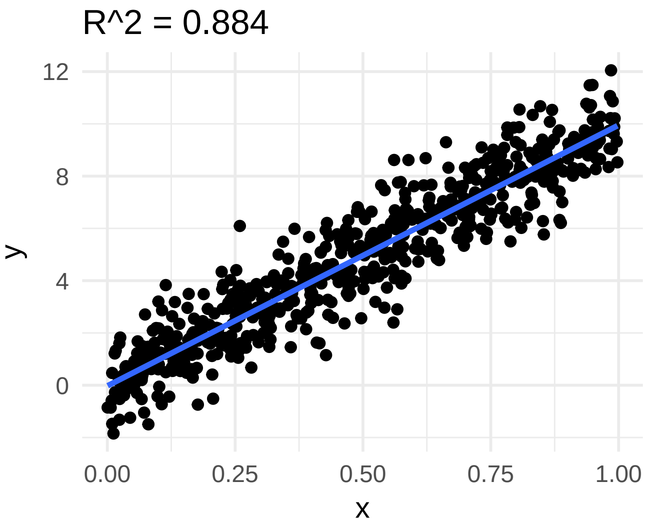
Visualizing R^2



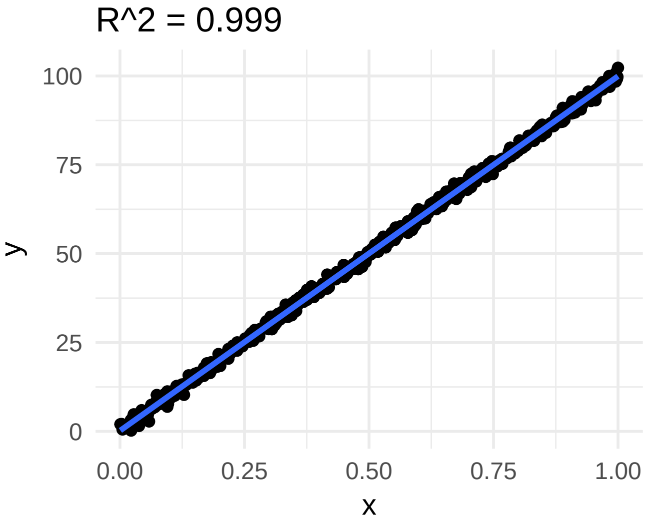
Visualizing R^2



Visualizing R^2

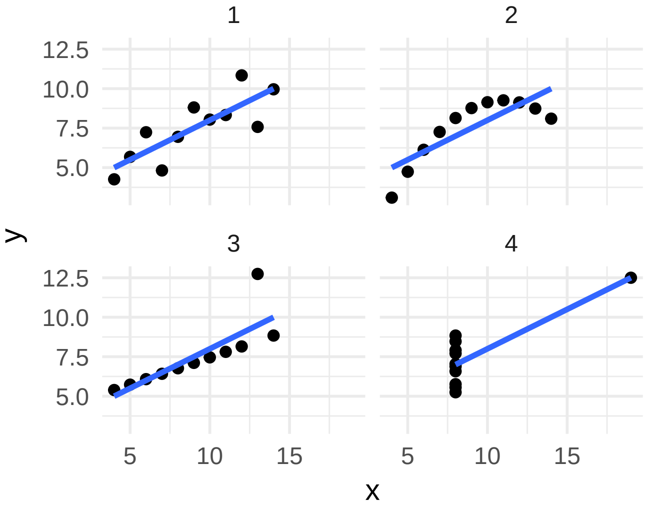


Visualizing R^2



Beware R^2

$R^2 = 0.666$ for each of these



Computing R^2 demo

Summary: R^2

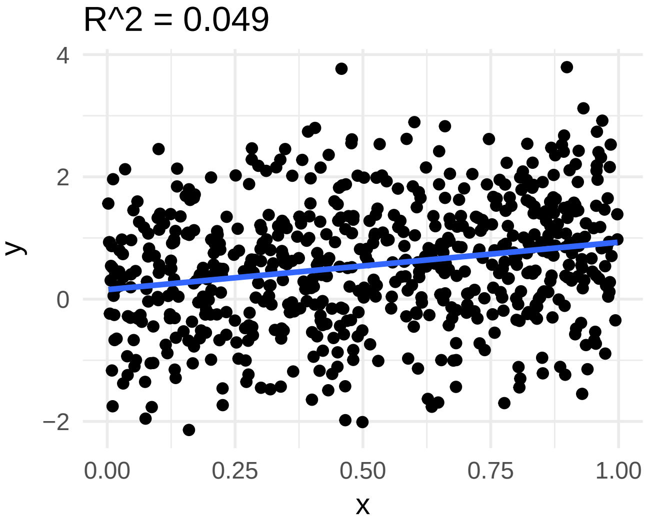
$$R^2 = 1 - \frac{SS_R}{SS_T}$$

- how well the line approximates the data
- bounded between 0 and 1

Problems with R^2

1. Sensitive to the range of x

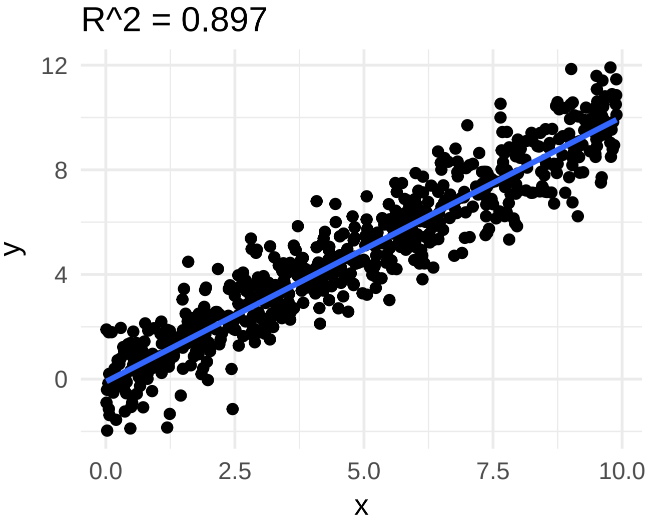
$y \sim \text{Normal}(0 + x, 1)$



Problems with R^2

$y \sim \text{Normal}(0 + x, 1)$

1. Sensitive to the range of x



Problems with R^2

1. Sensitive to the range of x
2. Does not measure prediction error

- only measures fit to the **training** data

Problems with R^2

1. Sensitive to the range of x
2. Does not measure prediction error
3. Not comparable among models with different y

Problems with R^2

1. Sensitive to the range of x
2. Does not measure prediction error
3. Not comparable among models with different y
4. We get the same R^2 by regressing x on y

Problems with R^2

1. Sensitive to the range of x
2. Does not measure prediction error
3. Not comparable among models with different y
4. We get the same R^2 by regressing x on y
5. Can calculate R^2 for silly models

- e.g., `http://tylervigen.com/spurious-correlations`

Recap for today

1. Quantifying uncertainty in parameters

- reviewed Hessians, standard error, and confidence intervals

2. Introducing R^2 (coefficient of determination)

- how to calculate from sums of squares
- some shortcomings