# Class overview and linear regression

Ciaran Evans

# Welcome to STA 711!

Agenda:

- Brief course overview and syllabus highlights
- Sketch plan for first couple weeks
- Linear regression

# Core course content

- Estimation
  - How do we estimate unknown parameters?
  - How do we assess uncertainty in our estimates?
  - What makes a "good" estimator?
- Asymptotics
  - What happens to our estimators as the sample size gets large?
- Hypothesis testing
  - How do we assess competing hypotheses about the data generating process?

# Course motivation

- Regression models are a natural setting for many of the key topics in 711
- We are already familiar with the idea of regression, and these models provide a good motivation for much of our course content
- Regression models are naturally multivariable
- STA 712 (GLMs) builds on our 711 material; I view 711/712 as a two-course sequence
- So: Regression (particularly linear and logistic regression) will be used as motivation throughout

# Course structure

- Class participation and seminar attendance
- Regular HW assignments (due most weeks)
- Homework presentations (3)
- Two midterm exams
- Final exam

# Plan for the first couple weeks

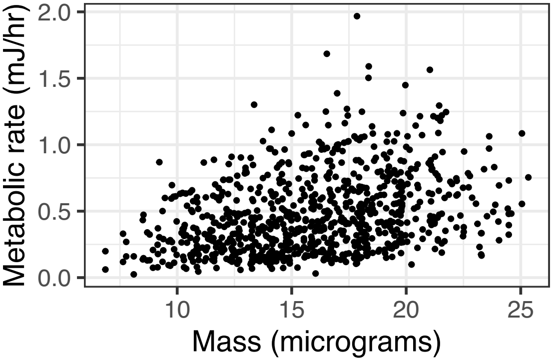
- Brief overview/review of linear regression
- Introduction to logistic regression and parameter estimation
- Maximum likelihood estimation

# Motivating example: mass and metabolic rate

- Researchers are interested in the relationship between an organism's mass and metabolic rate (how much energy it consumes)
- To study this relationship, they collected data on 568 individuals from two species of marine bryozoan

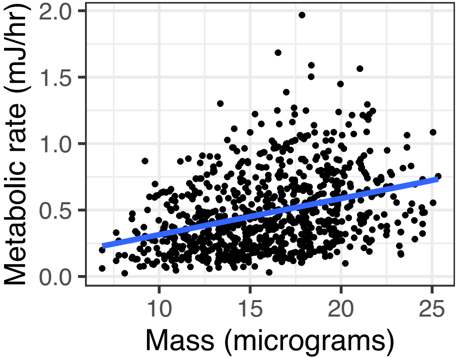
# Motivating example: mass and metabolic rate

For each individual, the researchers recorded the mass and metabolic rate:




**Question:** What do you notice about this relationship?

# Motivating example: mass and metabolic rate




If I wanted to write down a model for the relationship between mass and metabolic rate, what would it look like?

$$\text{Metabolic}_i = \beta_0 + \beta_1 \text{Mass}_i + \varepsilon_i$$

metabolic rate for $i^{th}$ observation

intercept

slope

variability around line

might assume $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

# Linear regression model

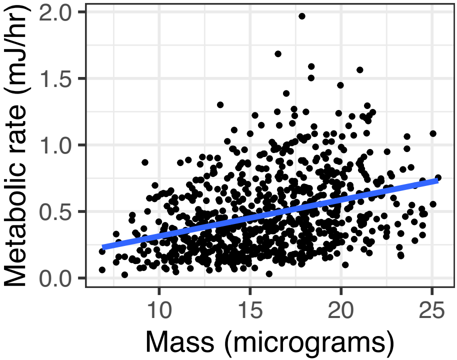
$$\text{Metabolic}_i = \beta_0 + \beta_1 \text{Mass}_i + \varepsilon_i$$

What do the parameters $\beta_0$ and $\beta_1$ represent?

$\beta_0$: average metabolic rate when Mass = 0

$\beta_1$: average change in metabolic rate for a one-mg change in mass

# Linear regression model




$$\text{Metabolic}_i = \beta_0 + \beta_1 \text{Mass}_i + \varepsilon_i$$

The true $\beta_0$ and $\beta_1$ are **unknown**. How do I estimate them from data?

# Fitting a linear regression model

intuition: want
$\hat{Y}_i$ close to $Y_i$
$\Rightarrow (Y_i - \hat{Y}_i)^2$ small

$$\text{Metabolic}_i = \beta_0 + \beta_1 \text{Mass}_i + \varepsilon_i$$

Choosen $\beta_0, \beta_1$ to minimize

$$SSE(\beta_0, \beta_1) = \sum_{i=1}^{n} (\text{Metabolic}_i - \beta_0 - \beta_1 \text{Mass}_i)^2$$

**Question:** How would I solve this minimization problem?

$$\frac{\partial}{\partial \beta_0} SSE = -2 \sum_i (\text{Metabolic} - \beta_0 - \beta_1 \text{Mass}_i) \overset{\text{set}}{=} 0$$

$$\frac{\partial}{\partial \beta_1} SSE = -2 \sum_i (\text{Metabolic} - \beta_0 - \beta_1 \text{Mass}_i) \text{Mass}_i \overset{\text{set}}{=} 0$$

solve the system!

# Fitting a linear regression model

**Model**:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i$$

**Observed data**: $(X_{11}, X_{12}, ..., X_{1k}, Y_1), ..., (X_{n1}, X_{n2}, ..., X_{nk}, Y_n)$

**Estimation**: choose $\beta_0, \beta_1, ..., \beta_k$ to minimize

$$SSE(\beta_0, ..., \beta_k) = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_k X_{ik})^2$$

How would I solve this minimization problem?

$$\frac{\partial}{\partial \beta_0} SSE = \ldots \overset{set}{=} 0$$

$$\frac{\partial}{\partial \beta_1} SSE = \ldots \overset{set}{=} 0$$

$$\vdots$$

$$\frac{\partial}{\partial \beta_k} SSE = \ldots \overset{set}{=} 0$$

k+1 equations

k+1 unknowns

# Improving our life with linear algebra

$$
\begin{aligned}
Y_1 &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \cdots + \beta_k X_{1k} + \varepsilon_1 \\
Y_2 &= \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \cdots + \beta_k X_{2k} + \varepsilon_2 \\
&\vdots \\
Y_n &= \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \cdots + \beta_k X_{nk} + \varepsilon_n
\end{aligned}
$$

$$
\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & & & & \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}
$$

# Improving our life with linear algebra

$$
\begin{aligned}
Y_1 &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \cdots + \beta_k X_{1k} + \varepsilon_1 \\
Y_2 &= \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \cdots + \beta_k X_{2k} + \varepsilon_2 \\
&\vdots \\
Y_n &= \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \cdots + \beta_k X_{nk} + \varepsilon_n
\end{aligned}
$$

$$Y = X \beta + \varepsilon$$

$$
\underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}}_{\text{vector of responses } Y} = \underbrace{\begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}}_{\text{design matrix } X} \underbrace{\begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}}_{\text{vector of coefficients } \beta} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}}_{\text{vector of noise } \varepsilon}
$$

# Rewriting the optimization problem

$v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \quad v^T v = \sum_i v_i^2$

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon$$

vector

$$SSE(\beta) = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_{i1} - \beta_k X_{ik})^2$$

$$= v^T v \qquad v = \begin{bmatrix} Y_1 - \beta_0 - \cdots - \beta_k X_{1k} \\ Y_2 - \beta_0 - \cdots - \beta_k X_{2k} \\ \vdots \\ Y_n - \beta_0 - \cdots - \beta_k X_{nk} \end{bmatrix}$$

$$= Y - X\beta$$

$$= (Y - X\beta)^T (Y - X\beta)$$

# Solving the optimization problem

$$SSE(\beta) = (Y - X\beta)^T (Y - X\beta)$$

Same idea:

$$\frac{\partial}{\partial \beta_0} SSE(\beta) = \ldots \overset{set}{=} 0$$

$$\vdots$$

$$\frac{\partial}{\partial \beta_k} SSE(\beta) = \ldots \overset{set}{=} 0$$

$$\begin{bmatrix} \frac{\partial}{\partial \beta_0} SSE \\ \vdots \\ \frac{\partial}{\partial \beta_k} SSE \end{bmatrix} : \quad \underline{\text{gradient}} \text{ of } SSE$$

$$= \frac{\partial}{\partial \beta} SSE(\beta) \overset{set}{=} 0$$

vector

So: need to find $\frac{\partial}{\partial \beta} SSE(\beta)$

# Solving the optimization problem

$$SSE(\beta) = (Y - X\beta)^T (Y - X\beta)$$

Matrix derivatives:

univariate calculus: $\frac{\partial}{\partial x} x^2 = 2x$

$$u = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \qquad \frac{\partial}{\partial u} u^T u = 2u$$

vector $x$, matrix $A$: $\frac{\partial}{\partial x} Ax = A^T$

not a function of $x$

chain rule: $\frac{\partial}{\partial \beta} u^T u = \left( \frac{\partial u}{\partial \beta} \right) \left( \frac{\partial u^T u}{\partial u} \right)$

$$\frac{\partial}{\partial \beta} \underbrace{(Y - X\beta)^T}_{u} (Y - X\beta) = \left( \frac{\partial}{\partial \beta} (Y - X\beta) \right) \left( \frac{\partial}{\partial Y - X\beta} (Y - X\beta)^T (Y - X\beta) \right)$$

$$= (-X^T)(2(Y - X\beta))$$

$$= -2X^T(Y - X\beta) \overset{set}{=} 0$$