

STATS 330

Handout 1

A synthesis of generalised linear models

Department of Statistics, University of Auckland

Linear regression

You were introduced to the simple linear regression model in STATS 20x:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

$$\epsilon_i \stackrel{iid}{\sim} \text{Normal}(0, \sigma^2)$$

Where, for the i th observation,

- ▶ Y_i is the value of the response variable
- ▶ x_i is the value of an explanatory variable
- ▶ ϵ_i is the error; the difference between Y_i and its expectation

We wish to estimate

- ▶ β_0 and β_1 , which define the straight-line relationship between the explanatory variable and the expectation of the response
- ▶ σ^2 , the variance of ϵ_i

LSD effect on maths analysis

In a 1968 study, a group of volunteers were injected with LSD and completed arithmetic tests at seven time points. The data frame contains the following variables:

score: The average score of the participants at a particular time point

lsd: The predicted LSD tissue concentration of the participants at a particular time point, based on the time since injection

Note that there an observation for each of seven time points, not for each participant.

LSD effect on maths analysis

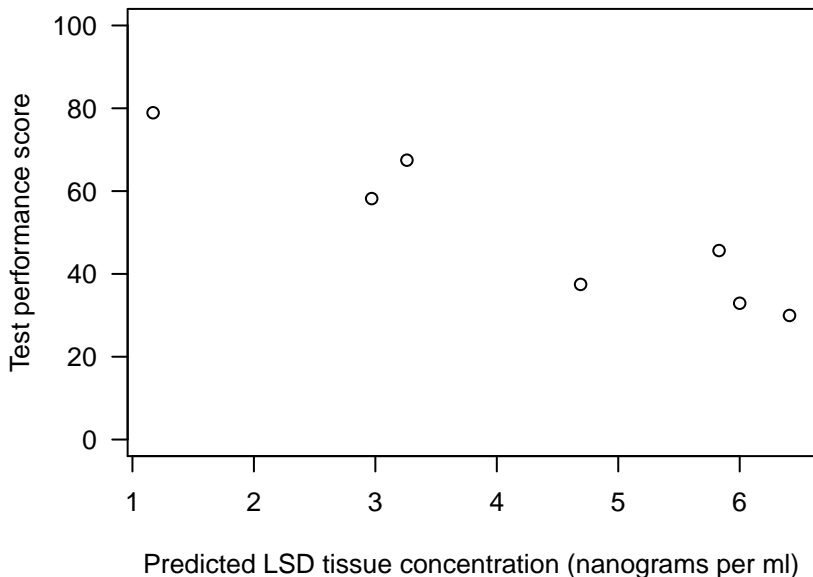
Here are the data:

```
lsd.df
```

```
##      lsd score
## 1 1.17 78.93
## 2 2.97 58.20
## 3 3.26 67.47
## 4 4.69 37.47
## 5 5.83 45.65
## 6 6.00 32.92
## 7 6.41 29.97
```

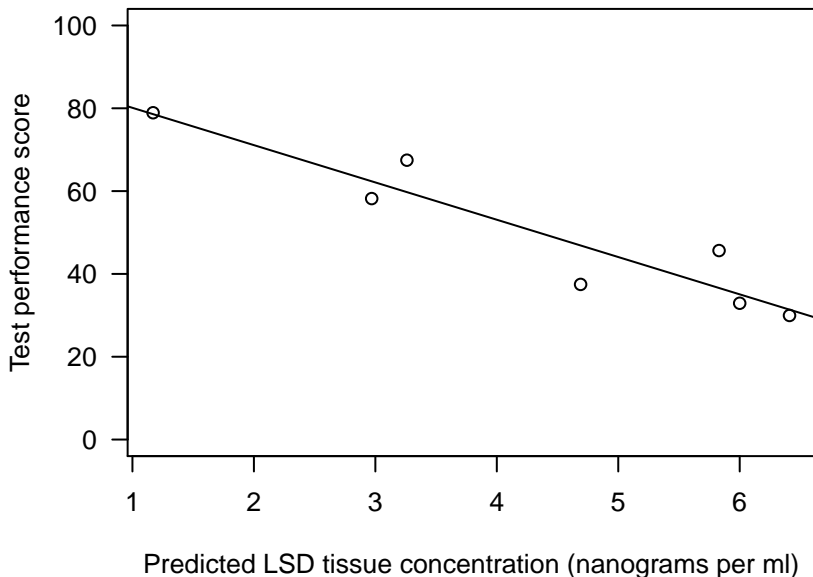
LSD effect on maths analysis

The data



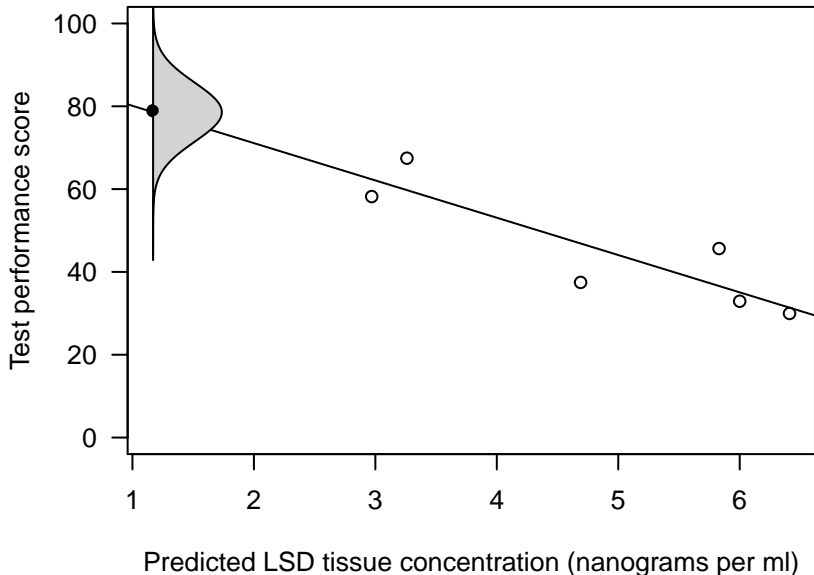
LSD effect on maths analysis

Linear regression



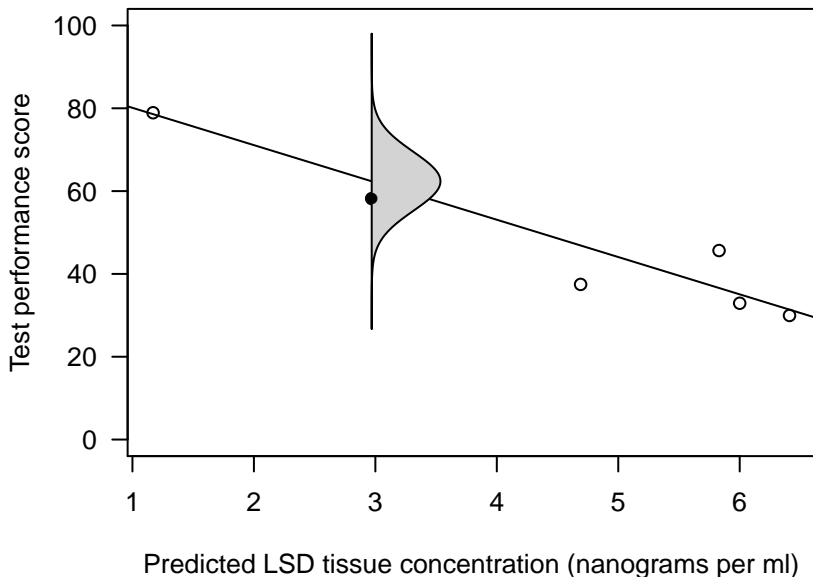
LSD effect on maths analysis

Linear regression



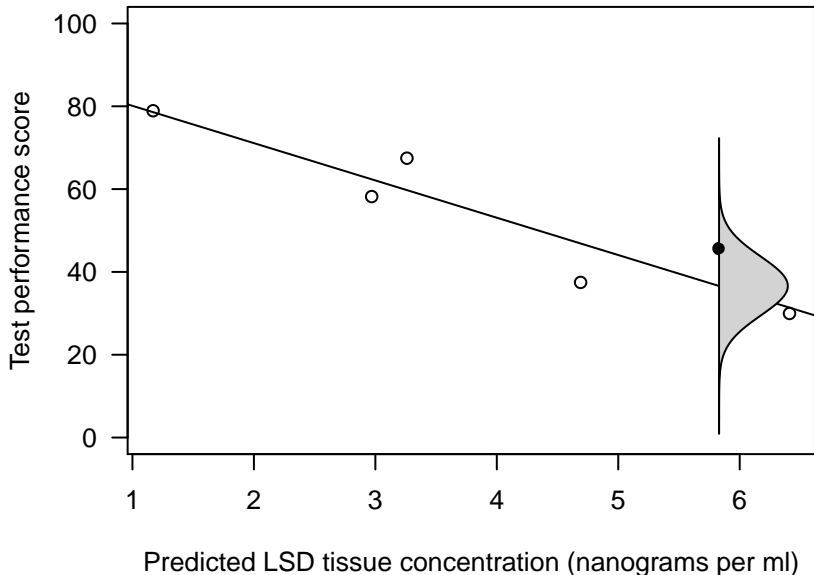
LSD effect on maths analysis

Linear regression



LSD effect on maths analysis

Linear regression



Linear regression

There is an alternative, equivalent way of specifying the linear regression model:

$$\begin{aligned}\mu_i &= \beta_0 + \beta_1 x_i \\ Y_i &\sim \text{Normal}(\mu_i, \sigma^2)\end{aligned}$$

Instead of specifying the model as random error, ϵ_i , about some relationship, $\beta_0 + \beta_1 x_i$, we attribute the randomness directly to the response variable.

We assume the i th observation's response, Y_i , comes from a normal distribution with mean $\beta_0 + \beta_1 x_i$ and variance σ^2 .

Linear regression

A linear regression model has the following assumptions:

1. The i th observation's response, Y_i , comes from a normal distribution
2. Its mean, μ_i , is a linear combination of the explanatory terms
3. Its variance, σ^2 , is the same for all observations
4. Each observation's response is independent of all others

However, in some cases, we may wish to unshackle ourselves from a model with normal errors

Generalised linear models provide the answer

Macrorhabdus ornithogaster chicken data

M. ornithogaster is a bacterial disease that affects birds. The drug Amphotericin B is often used for treatment, but its efficacy has not yet been established. It was of interest to determine if the Amphotericin B dose is related to the number of *M. ornithogaster* organisms shed in chickens' faeces, while correcting for the size of the chicken. An experiment was conducted on 23 infected chickens.

The data frame contains the following variables:

mo: The number of *M. ornithogaster* in a sample of ten faecal slides from the chicken

dose: The dose of Amphotericin B given to the chicken; either control, low, or high

weight: The weight of the chicken in grams

First, we'll just investigate the relationship between `weight` and `mo`

Macrorhabdus ornithogaster chicken data

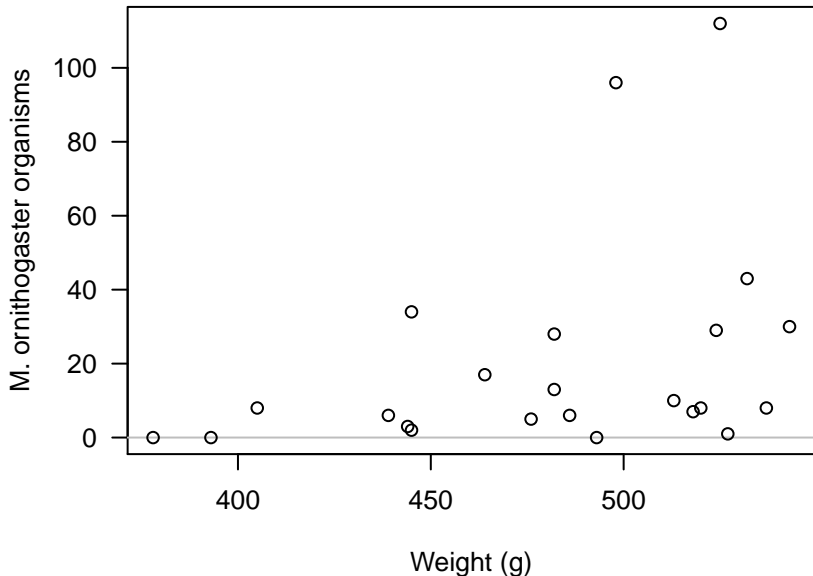
Here are the first ten rows of the data:

```
head(chickens.df, 10)
```

##	mo	weight	dose
## 1	0	378	High
## 2	0	393	Low
## 3	8	405	Low
## 4	6	439	Low
## 5	3	444	High
## 6	2	445	Low
## 7	34	445	Low
## 8	17	464	Control
## 9	5	476	Low
## 10	13	482	High

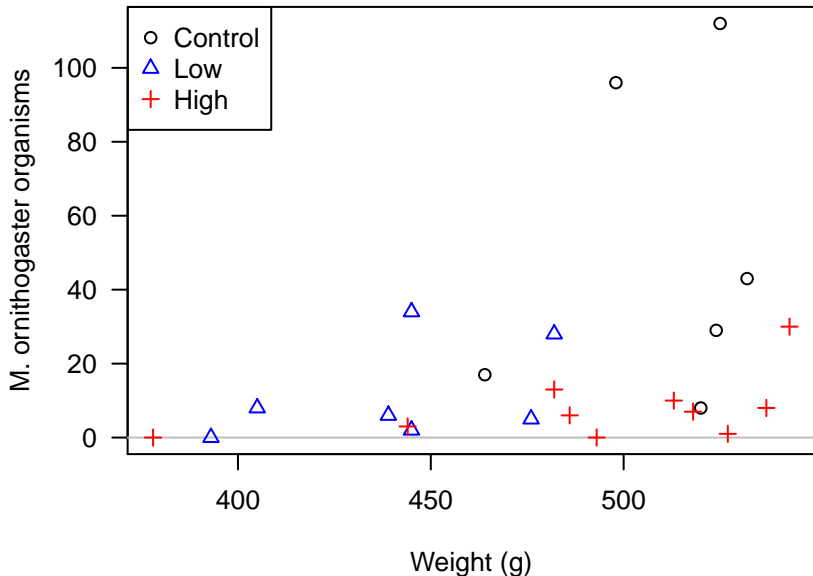
Macrorhabdus ornithogaster chicken data

The data, without dose groups



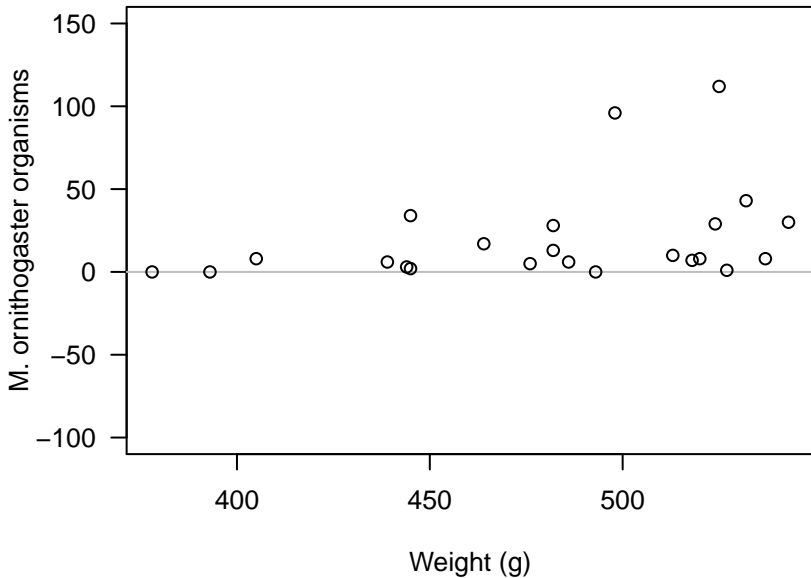
Macrorhabdus ornithogaster chicken data

The data, with dose groups



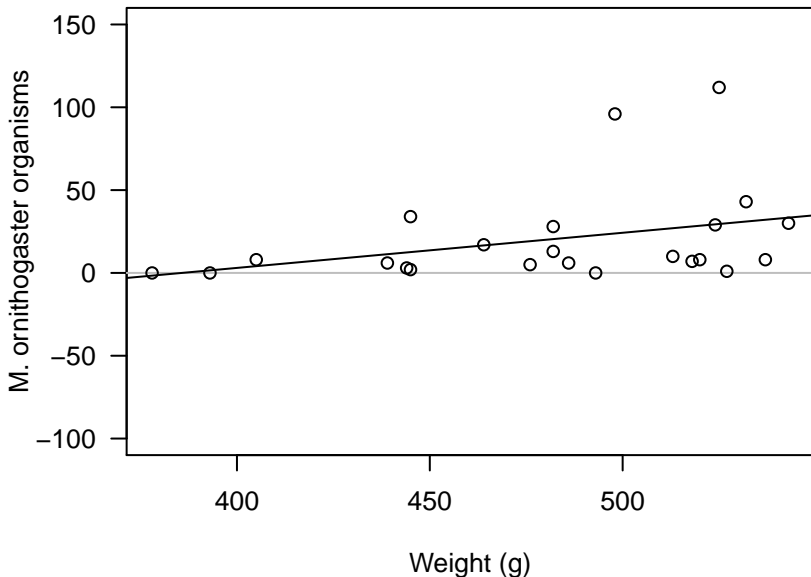
Macrorhabdus ornithogaster chicken data

The data, with wider y-axis limits



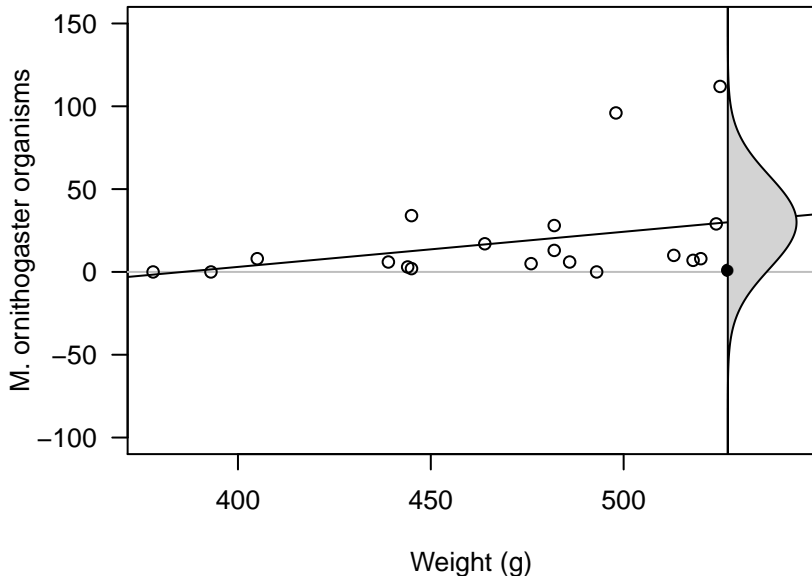
Macrorhabdus ornithogaster chicken analysis

Linear regression



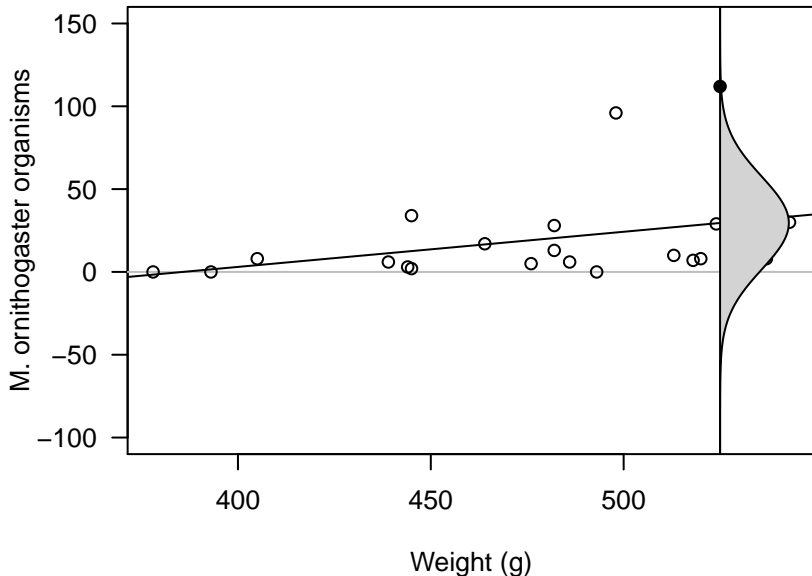
Macrorhabdus ornithogaster chicken analysis

Linear regression



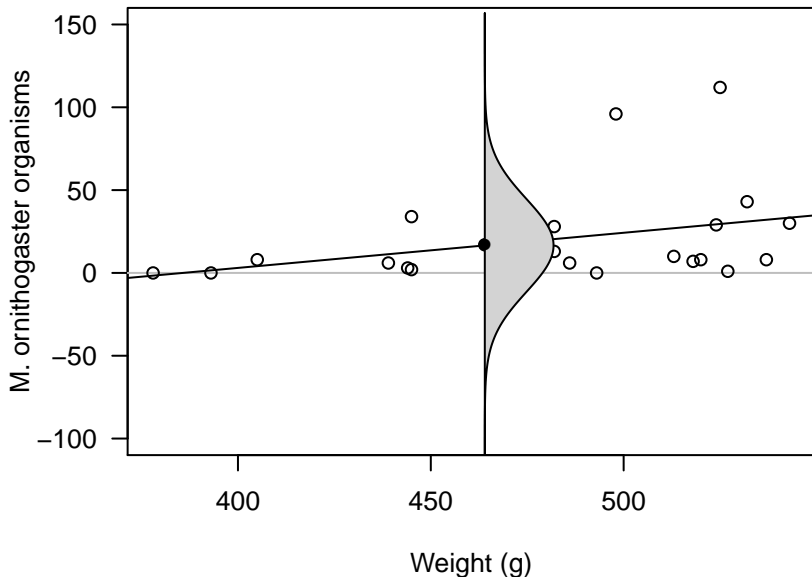
Macrorhabdus ornithogaster chicken analysis

Linear regression



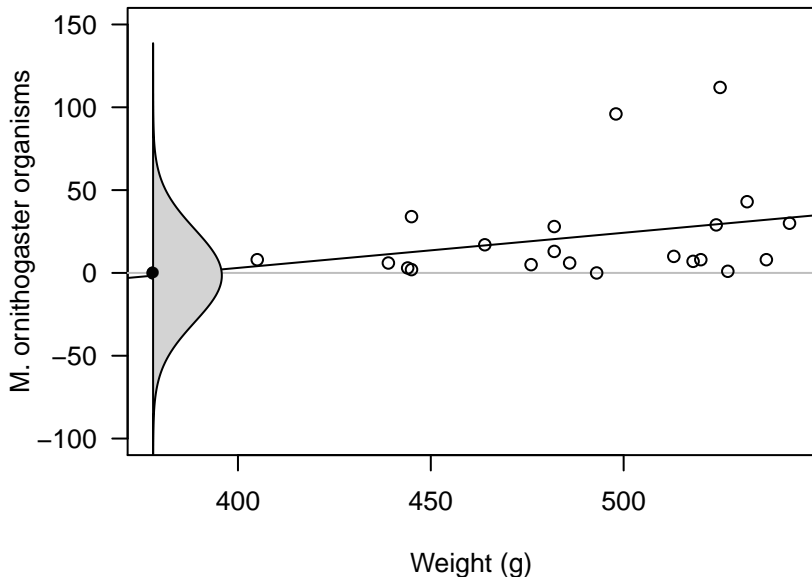
Macrorhabdus ornithogaster chicken analysis

Linear regression



Macrorhabdus ornithogaster chicken analysis

Linear regression



Macrorhabdus ornithogaster chicken analysis

Linear regression

A normal distribution does not adequately describe the response, the number of organisms on a slide:

- ▶ It is a continuous distribution, but the response is discrete
- ▶ It is symmetric, but the response looks right skewed
- ▶ It is unbounded, and assumes it is plausible for the response to be negative

Also, a linear regression model typically assumes constant variance, but here the variance appears to increase with the mean.

Macrorhabdus ornithogaster chicken analysis

Linear regression

So why assume a normal distribution?

$$\mu_i = \beta_0 + \beta_1 x_i,$$

$$Y_i \sim \text{Normal}(\mu_i, \sigma^2),$$

Macrorhabdus ornithogaster chicken analysis

Poisson regression

So why assume a normal distribution? Let's use a Poisson distribution instead:

$$\mu_i = \beta_0 + \beta_1 x_i,$$

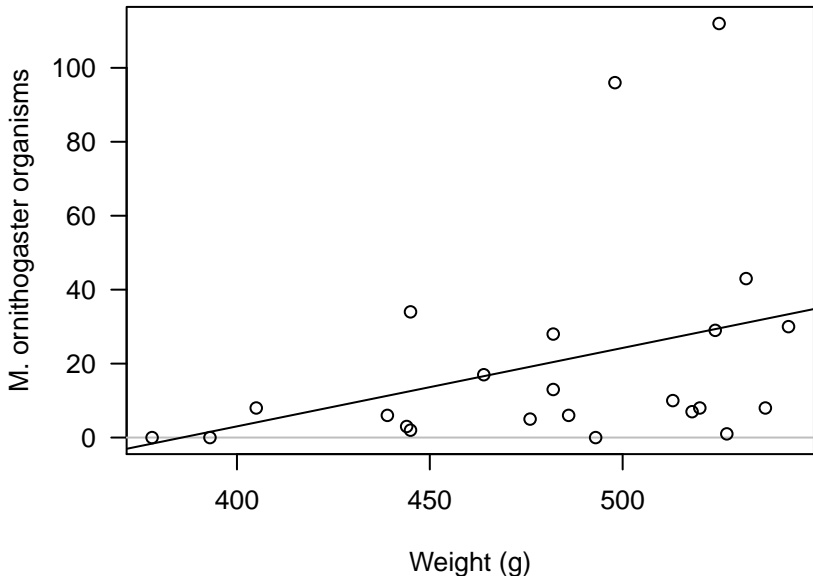
$$Y_i \sim \text{Poisson}(\mu_i),$$

You will have seen the Poisson distribution in STATS 20x, and perhaps other statistics courses. It is commonly used as a general-purpose distribution for counts.

For the Poisson distribution, $\text{Var}(Y_i) = \mu_i$, so we expect the variance to increase with the mean

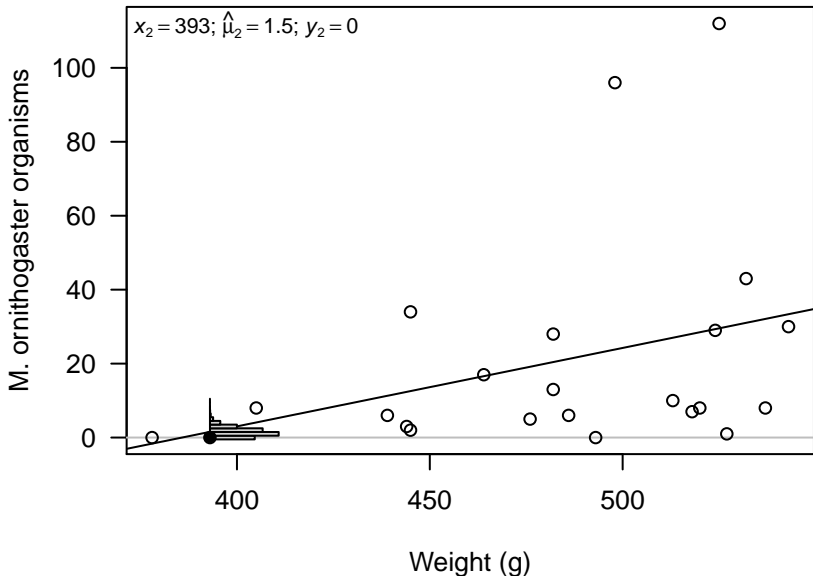
Macrorhabdus ornithogaster chicken analysis

Poisson regression



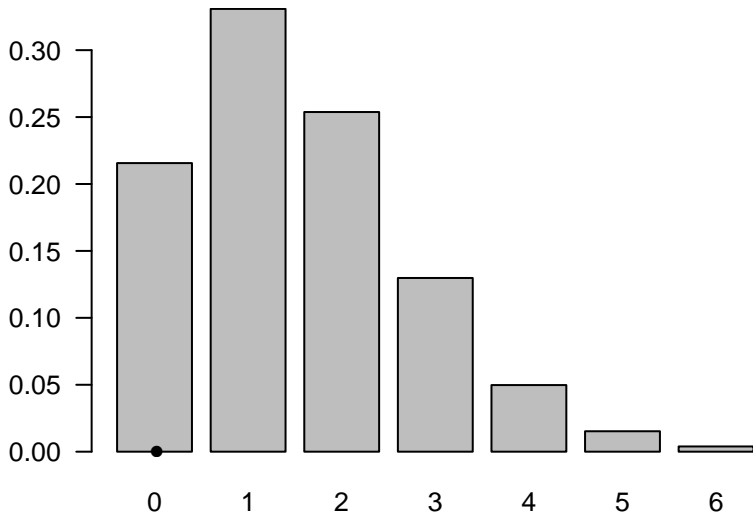
Macrorhabdus ornithogaster chicken analysis

Poisson regression



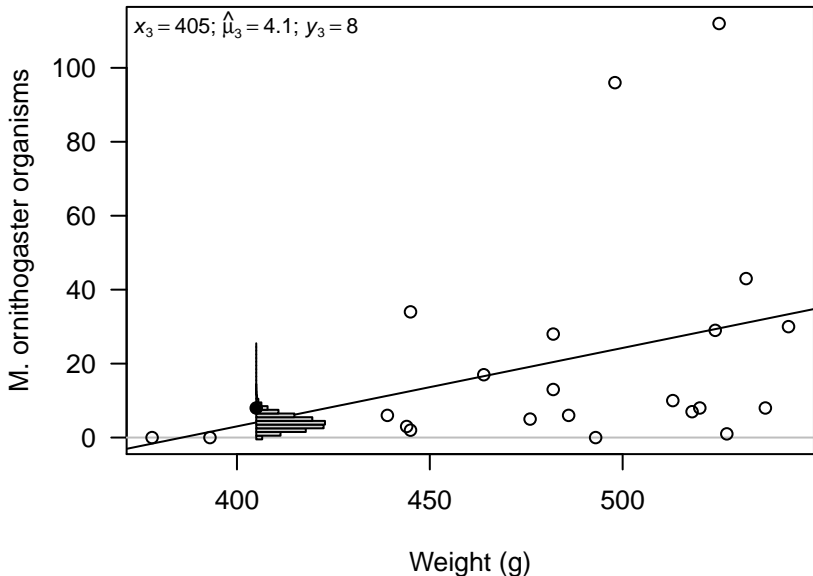
Macrorhabdus ornithogaster chicken analysis

Fitted distribution for observation 2



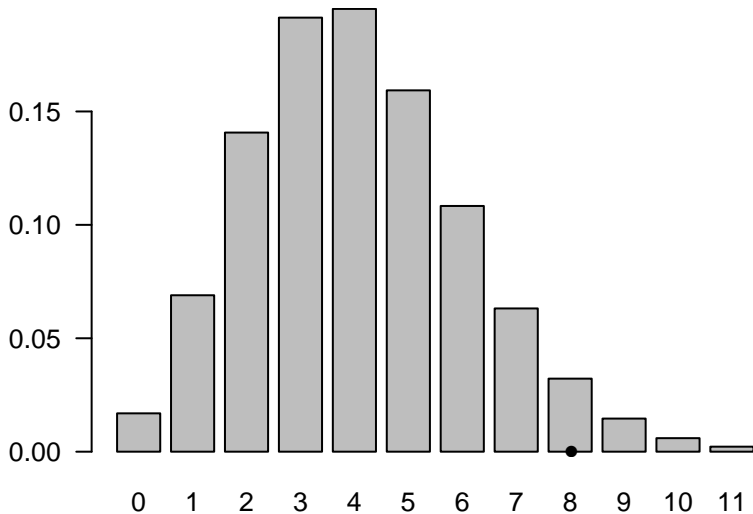
Macrorhabdus ornithogaster chicken analysis

Poisson regression



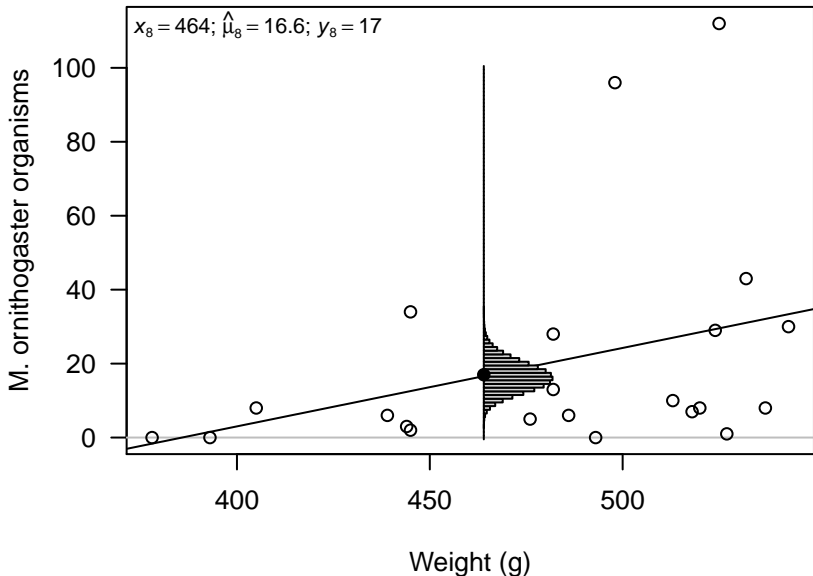
Macrorhabdus ornithogaster chicken analysis

Fitted distribution for observation 3



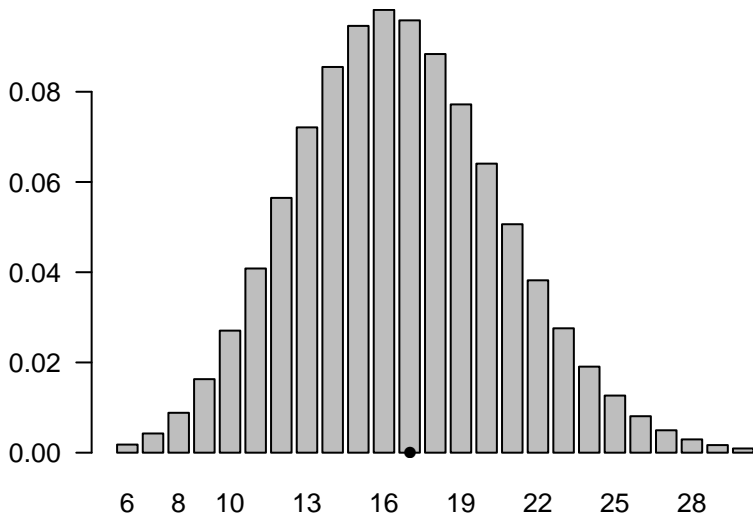
Macrorhabdus ornithogaster chicken analysis

Poisson regression



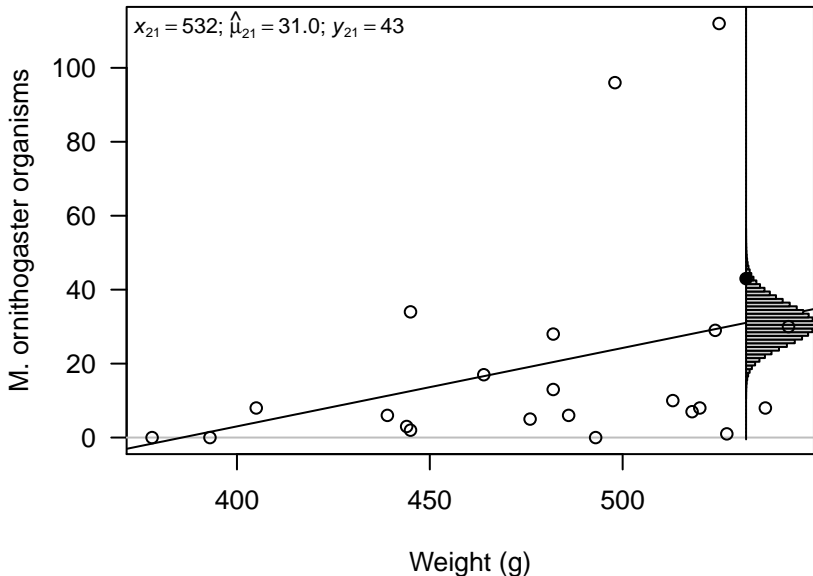
Macrorhabdus ornithogaster chicken analysis

Fitted distribution for observation 2



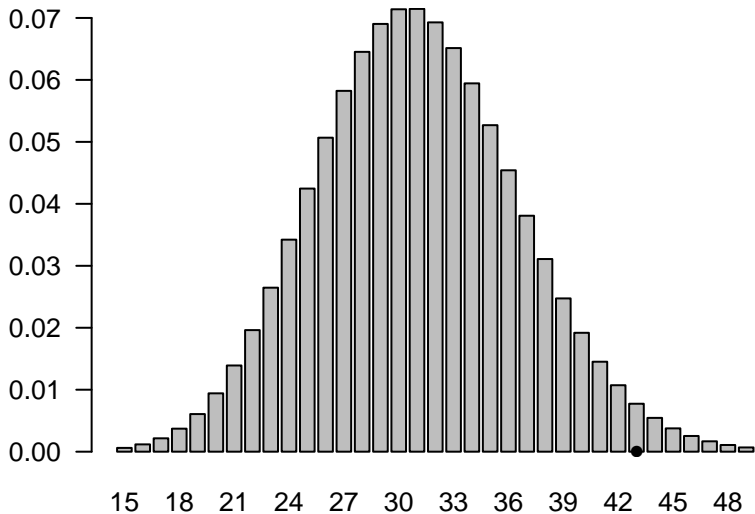
Macrorhabdus ornithogaster chicken analysis

Poisson regression



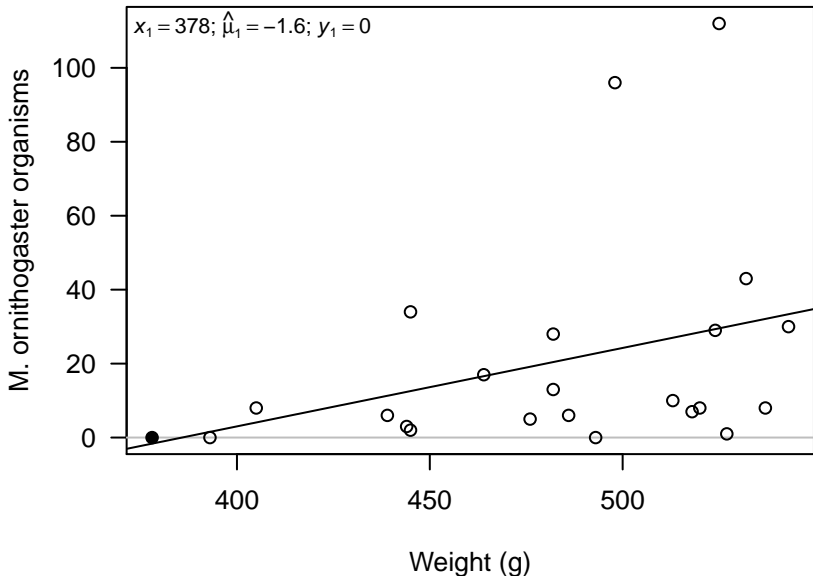
Macrorhabdus ornithogaster chicken analysis

Fitted distribution for observation 2



Macrorhabdus ornithogaster chicken analysis

Poisson regression



Macrorhabdus ornithogaster chicken analysis

Poisson regression

The Poisson distribution better reflects the nature of the response:

- ▶ It is a discrete distribution
- ▶ It is right skewed (but almost symmetric when μ is large)
- ▶ It does not allow the response to be negative
- ▶ It assumes that the variance increases with the mean

... But our linear relationship allows $\mu < 0$ for light chickens.

Macrorhabdus ornithogaster chicken analysis

Poisson regression

We use a link function to map our parameter of interest, μ_i to the real number line:

$$\log(\mu_i) = \beta_0 + \beta_1 x_i;$$

in other words, $\mu_i \geq 0$, but $-\infty < \log(\mu_i) < +\infty$.

Equivalently, we have

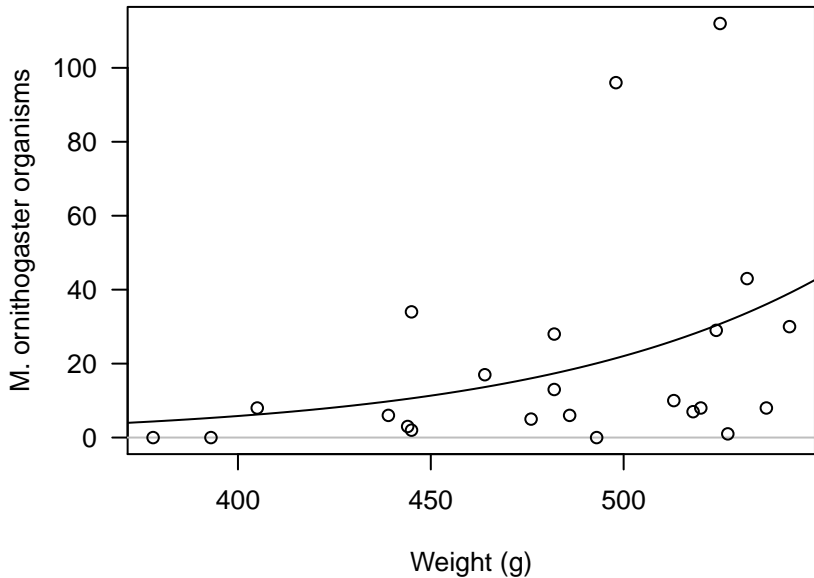
$$\mu_i = \exp(\beta_0 + \beta_1 x_i),$$

$$Y_i \sim \text{Poisson}(\mu_i).$$

No matter how negative the linear predictor $\beta_0 + \beta_1 x_i$ is, μ_i will always be positive

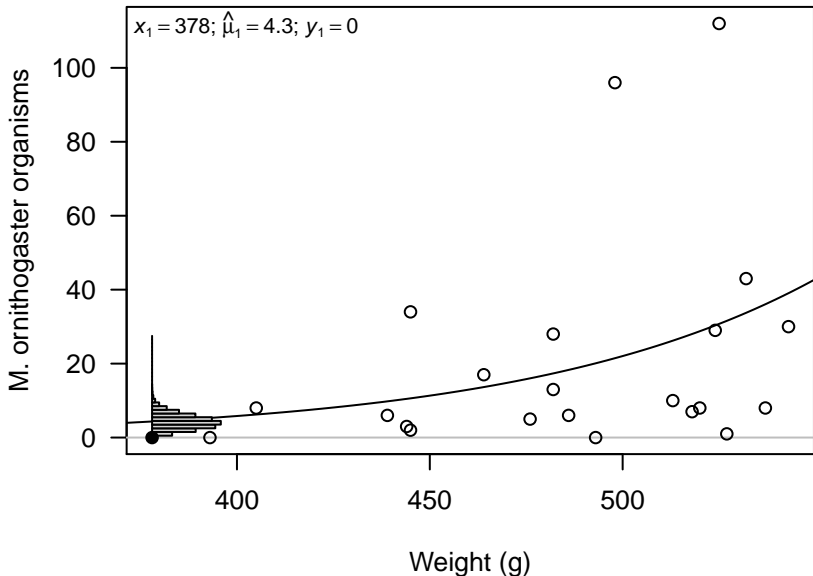
Macrorhabdus ornithogaster chicken analysis

Poisson regression



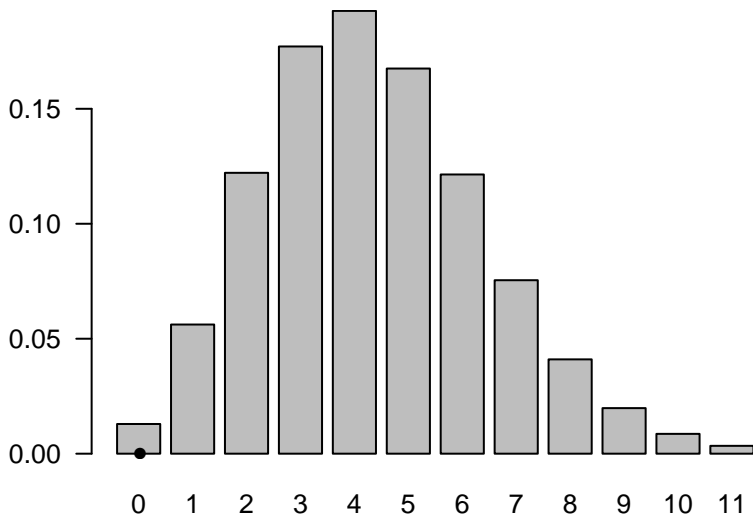
Macrorhabdus ornithogaster chicken analysis

Poisson regression



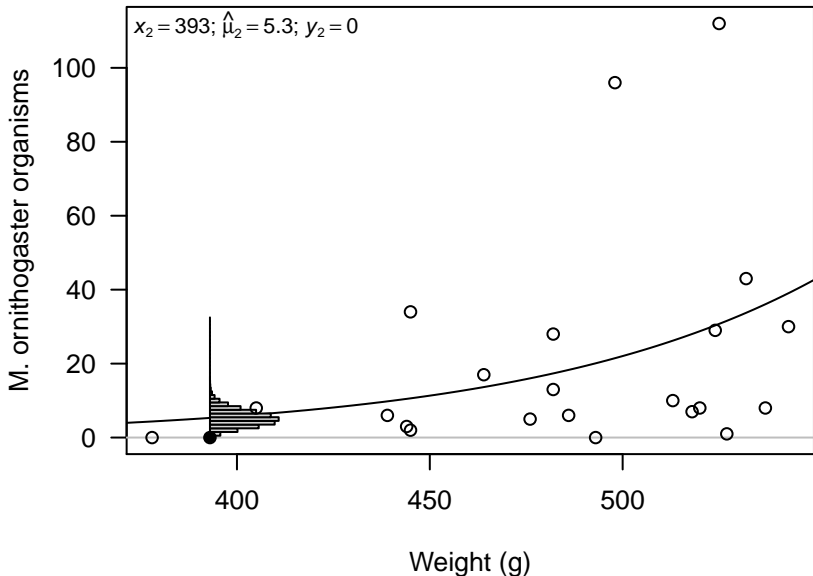
Macrorhabdus ornithogaster chicken analysis

Poisson regression



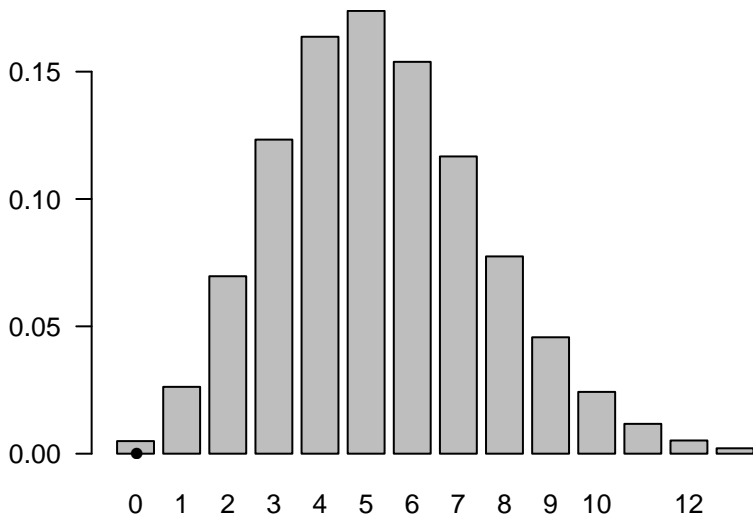
Macrorhabdus ornithogaster chicken analysis

Poisson regression



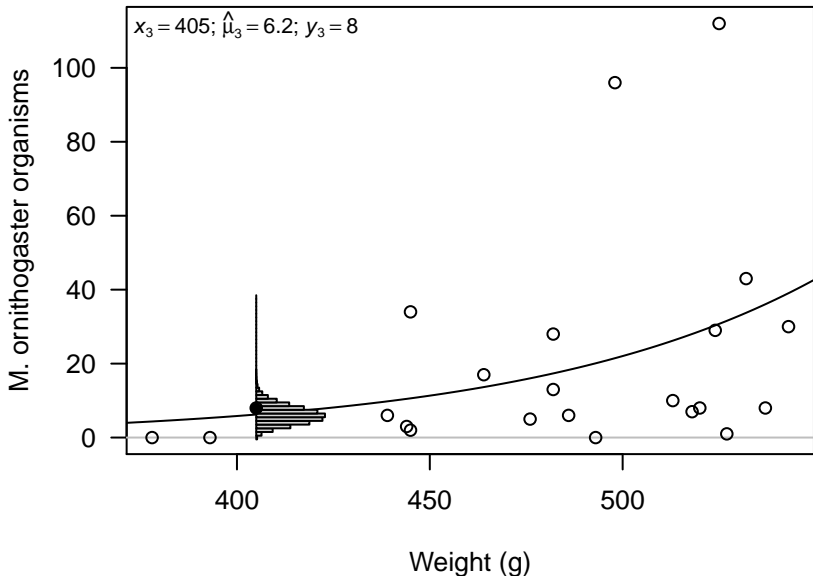
Macrorhabdus ornithogaster chicken analysis

Poisson regression



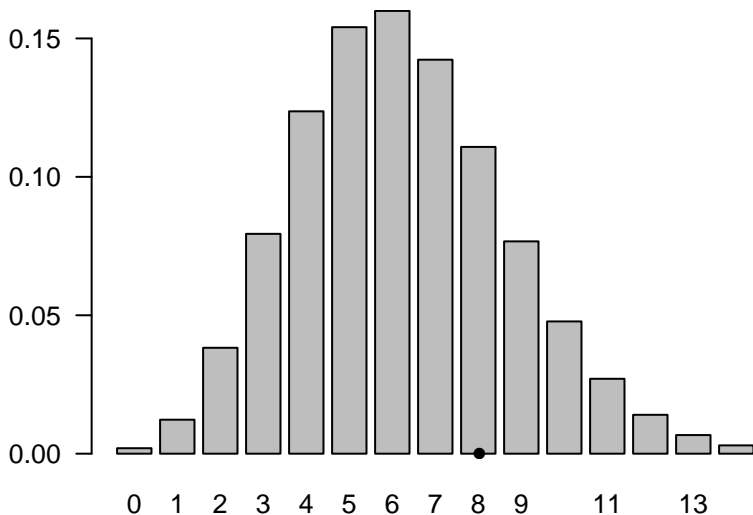
Macrorhabdus ornithogaster chicken analysis

Poisson regression



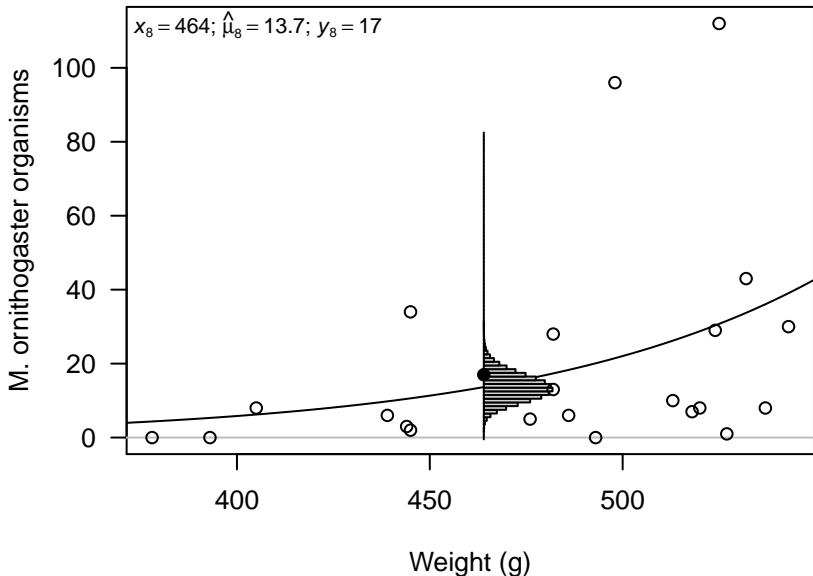
Macrorhabdus ornithogaster chicken analysis

Poisson regression



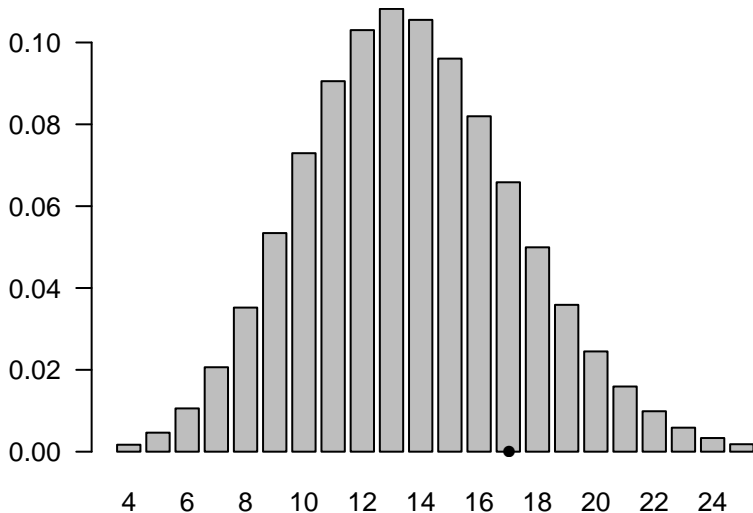
Macrorhabdus ornithogaster chicken analysis

Poisson regression



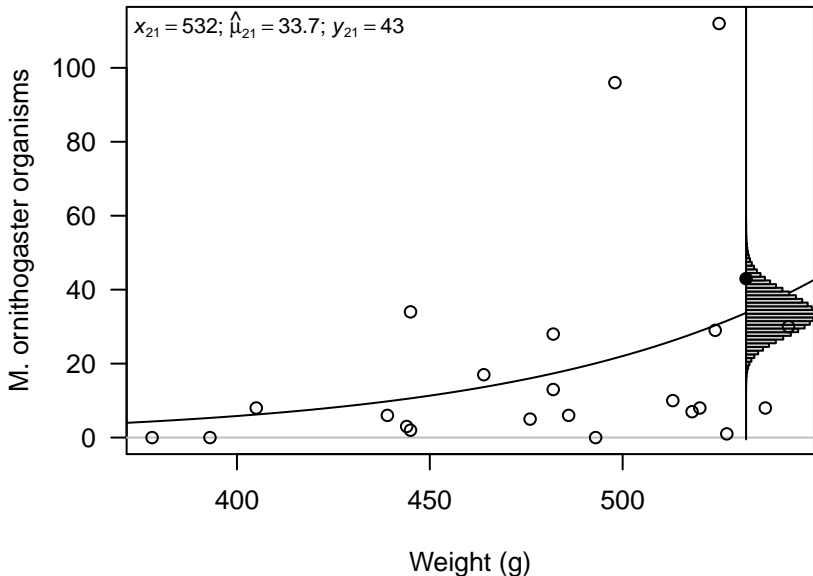
Macrorhabdus ornithogaster chicken analysis

Poisson regression



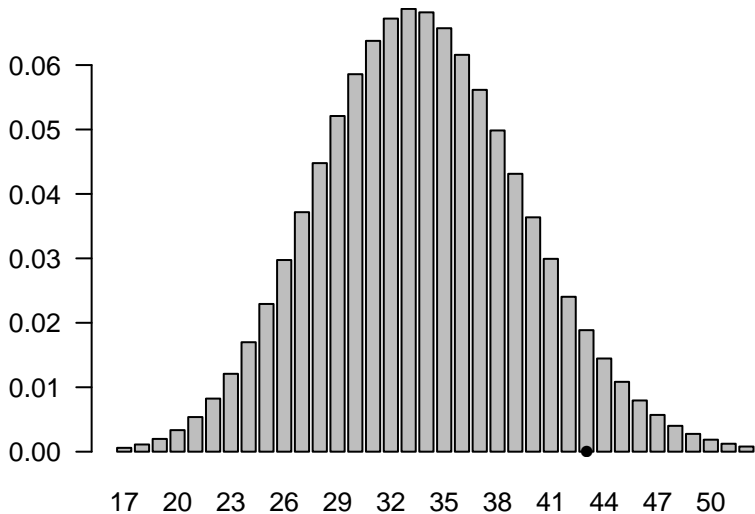
Macrorhabdus ornithogaster chicken analysis

Poisson regression



Macrorhabdus ornithogaster chicken analysis

Poisson regression



Coronary heart disease data

A sample of 100 subjects from an at-risk population were tested for presence of significant coronary heart disease (CHD).

The data were grouped prior to analysis, so that subjects of the same age form a group. The data frame contains the following variables:

- age:** The age of subjects in a particular group in years.
- n:** The number of subjects in a particular age group.
- y:** The number of subjects in a particular age group with significant CHD.
- p:** The proportion of subjects in a particular age group with significant CHD (y/n).

These data are taken from the textbook Hosmer and Lemeshow (2013), *Applied Logistic Regression*.

Coronary heart disease data

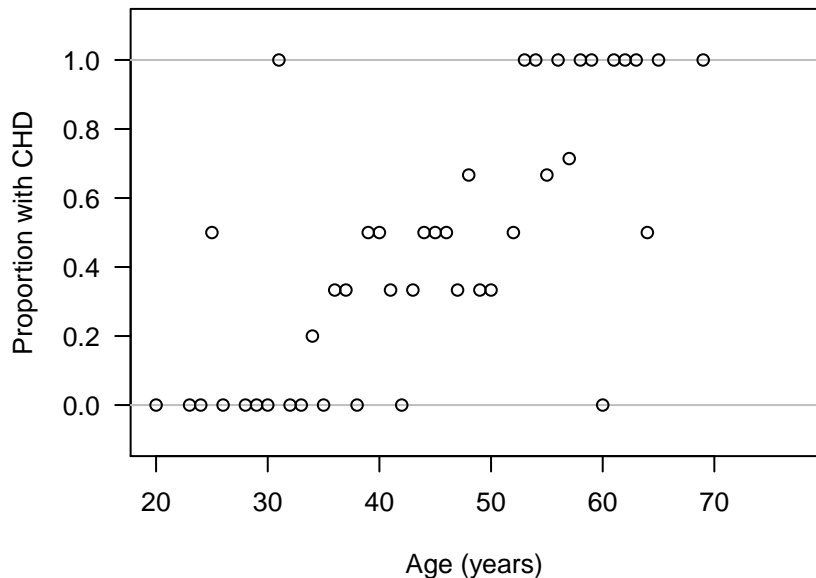
Here are the first ten rows of the data:

```
head(chd.df, 10)
```

##		age	y	n	p
## 1		20	0	1	0.0
## 2		23	0	1	0.0
## 3		24	0	1	0.0
## 4		25	1	2	0.5
## 5		26	0	2	0.0
## 6		28	0	2	0.0
## 7		29	0	1	0.0
## 8		30	0	5	0.0
## 9		31	1	1	1.0
## 10		32	0	2	0.0

Coronary heart disease data

The data



Coronary heart disease analysis

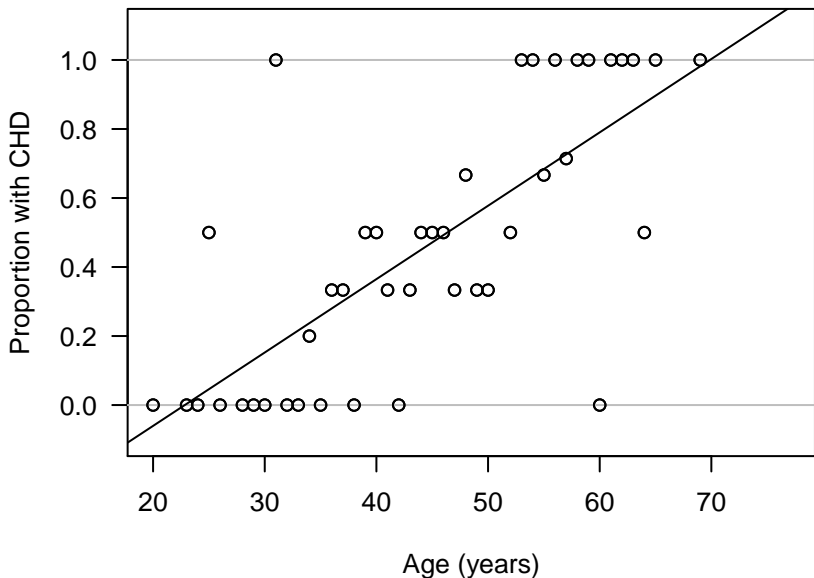
It may be tempting to fit a linear regression model, with proportion with CHD as the response and size as the explanatory variable. In fact, this is what the authors did in their published paper.

There are a few problems with this:

- ▶ The proportion of subjects with CHD is bounded on the interval $[0, 1]$, but the normal distribution is not;
- ▶ The variance is unlikely to be constant. Age groups with fewer subjects and with expected proportions closer to 0.5 are likely to have more variance; and
- ▶ The expected proportion of subjects with CHD is similarly bounded to the interval $[0, 1]$, but a straight line is not.

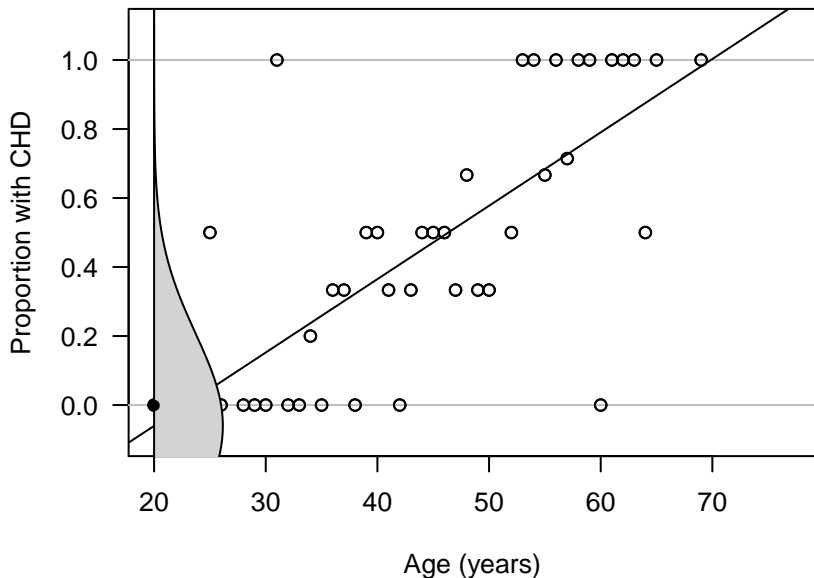
Coronary heart disease analysis

Linear regression



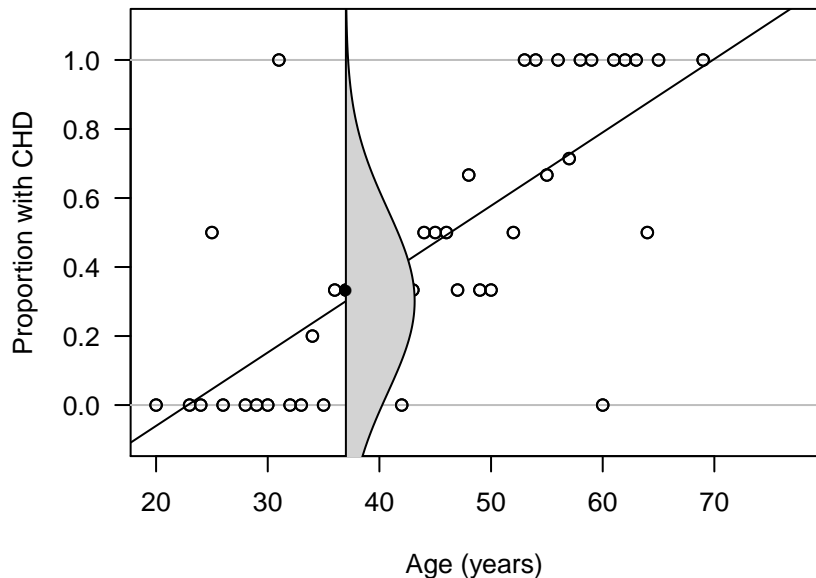
Coronary heart disease analysis

Linear regression



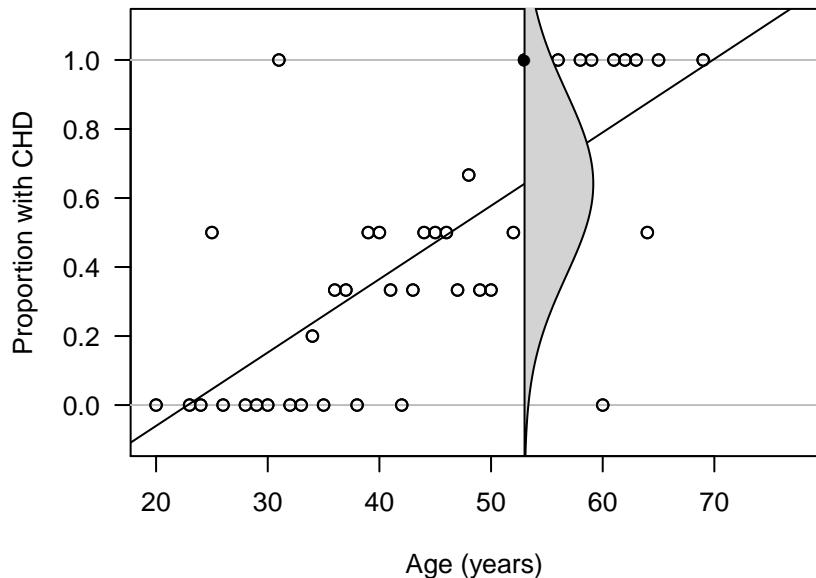
Coronary heart disease analysis

Linear regression



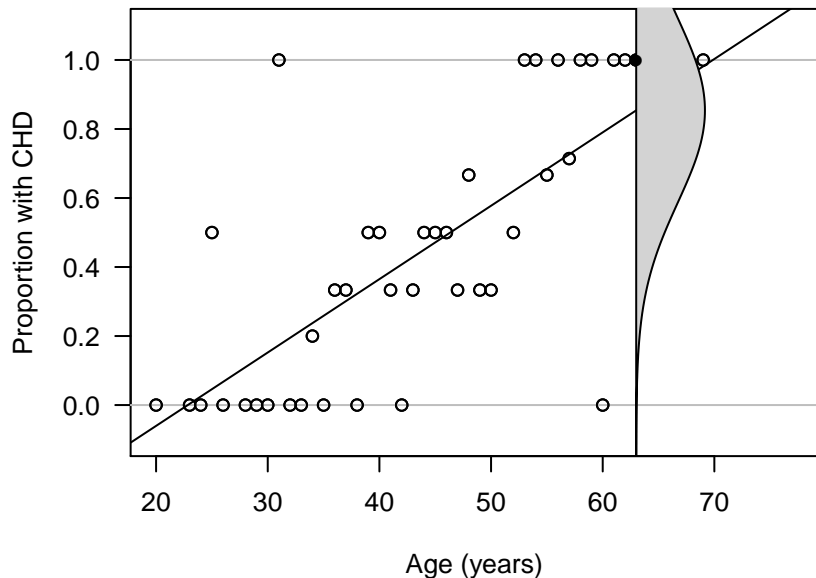
Coronary heart disease analysis

Linear regression



Coronary heart disease analysis

Linear regression



Coronary heart disease analysis

So why assume a normal distribution?

$$\mu_i = \beta_0 + \beta_1 x_i,$$

$$Y_i/n_i \sim \text{Normal}(\mu_i, \sigma^2)$$

Coronary heart disease analysis

So why assume a normal distribution? We can solve the first two problems by using a binomial distribution instead:

$$p_i = \beta_0 + \beta_1 x_i,$$

$$Y_i \sim \text{Binomial}(n_i, p_i)$$

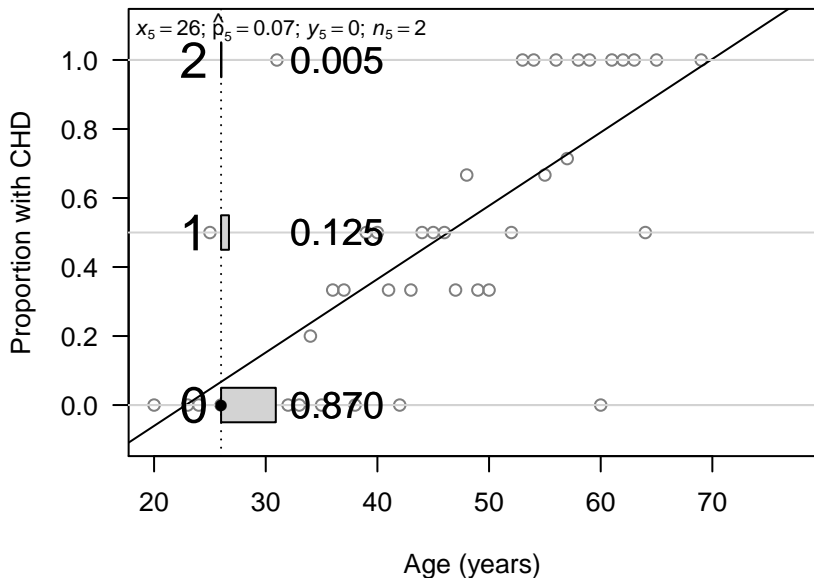
A random variable has a binomial distribution if it is the number of 'successes' from a number of independent trials, each with the same probability of success.

This builds in the nonconstant variance of Y_i and Y_i/n_i :

- ▶ $\text{Var}(Y_i) = n_i p_i (1 - p_i)$
- ▶ $\text{Var}(Y_i/n_i) = p_i (1 - p_i) / n_i$

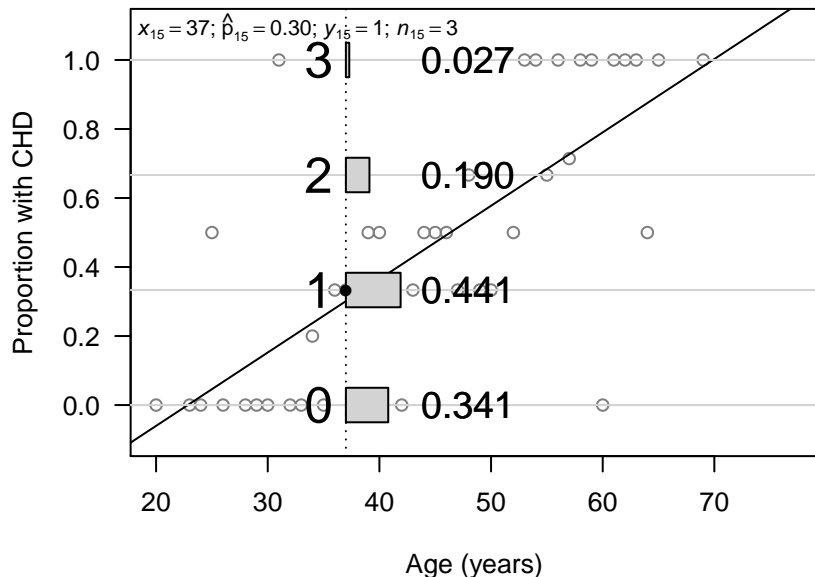
Coronary heart disease analysis

Logistic regression



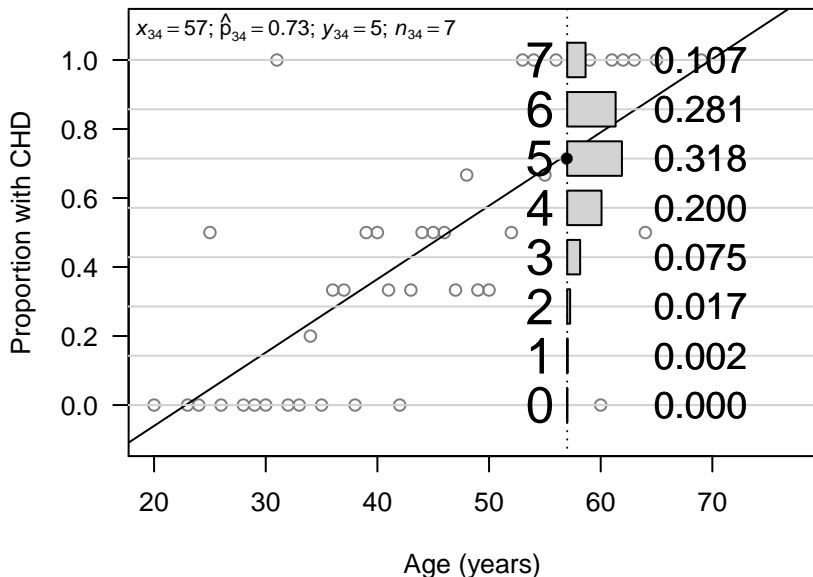
Coronary heart disease analysis

Logistic regression



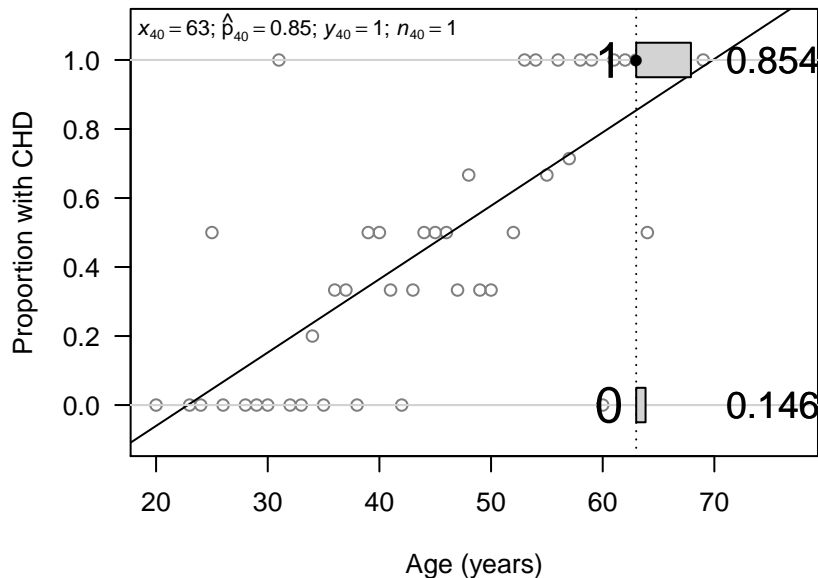
Coronary heart disease analysis

Logistic regression



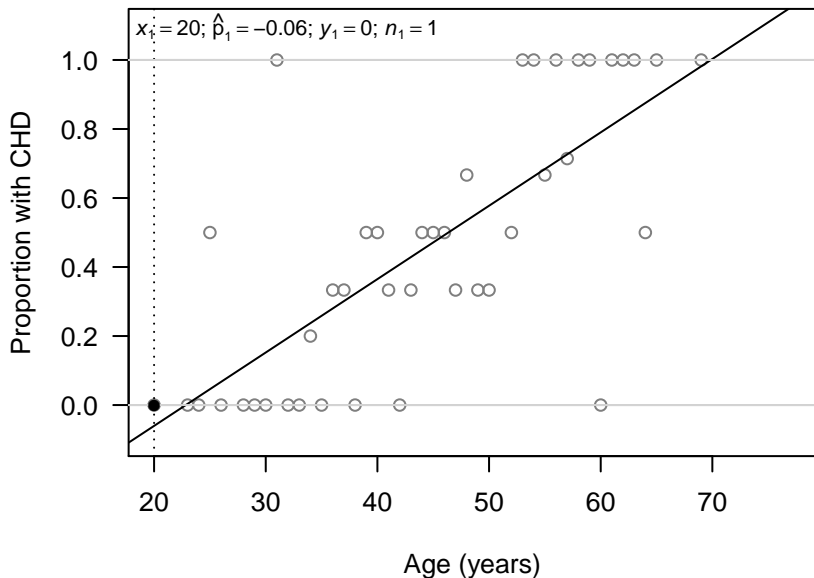
Coronary heart disease analysis

Logistic regression



Coronary heart disease analysis

Logistic regression



Coronary heart disease analysis

Logistic regression

The binomial distribution better reflects the nature of the response:

- ▶ It is a discrete distribution,
- ▶ It acknowledges that the number of successes associated with the i th observation must be an integer between 0 and n_i , and
- ▶ It accounts for nonconstant variance in a sensible way

... But our linear relationship allows a trial's probability of success to be greater than 1 (for old subjects) or less than 0 (for young subjects).

Coronary heart disease analysis

Logistic regression

For logistic regression, the standard link function is the logit function:

$$\text{logit}(p_i) = \log \left(\frac{p_i}{1 - p_i} \right) = \beta_0 + \beta_1 x_i.$$

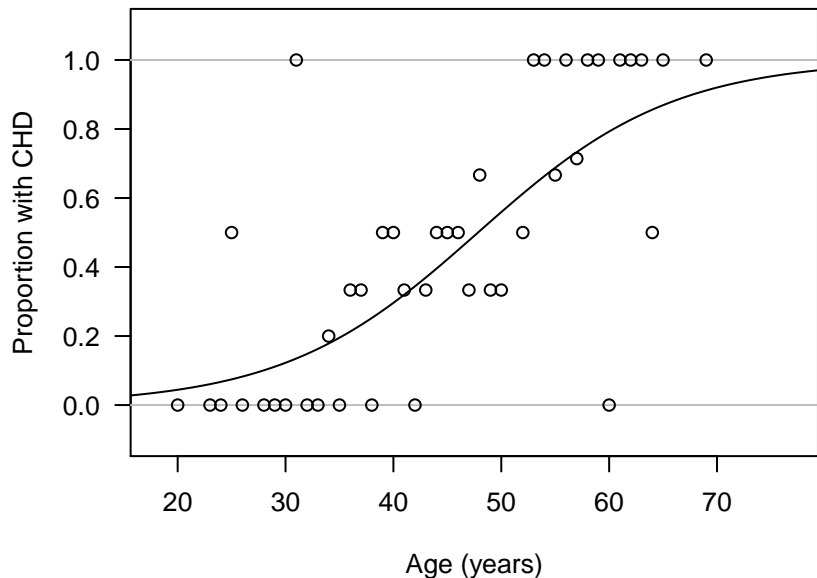
By making p_i the subject of the above equation, we obtain

$$p_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}$$
$$Y_i \sim \text{Binomial}(n_i, p_i)$$

This is the logistic function, and produces an S-shaped curve.

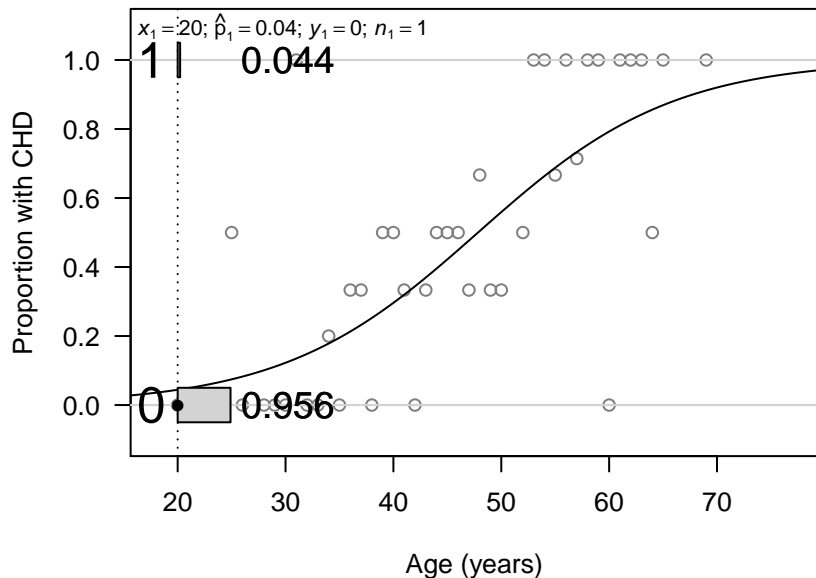
Coronary heart disease analysis

Logistic regression



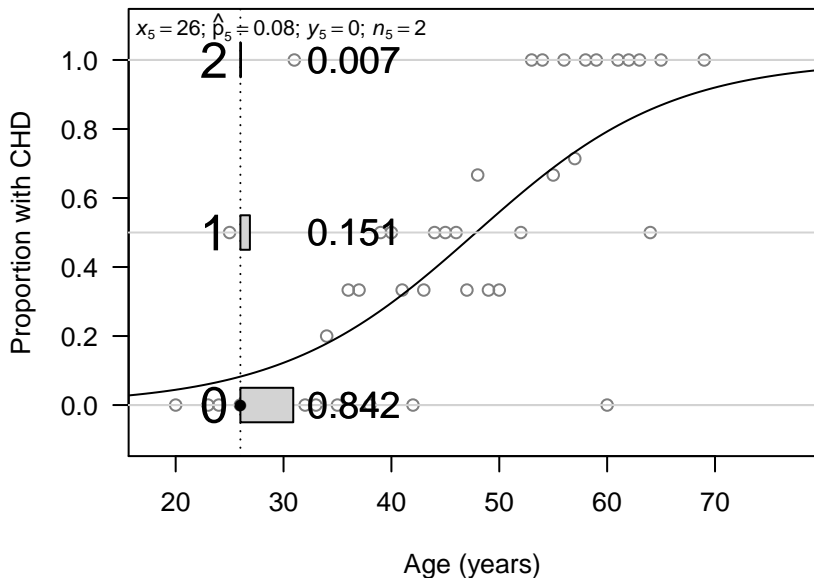
Coronary heart disease analysis

Logistic regression



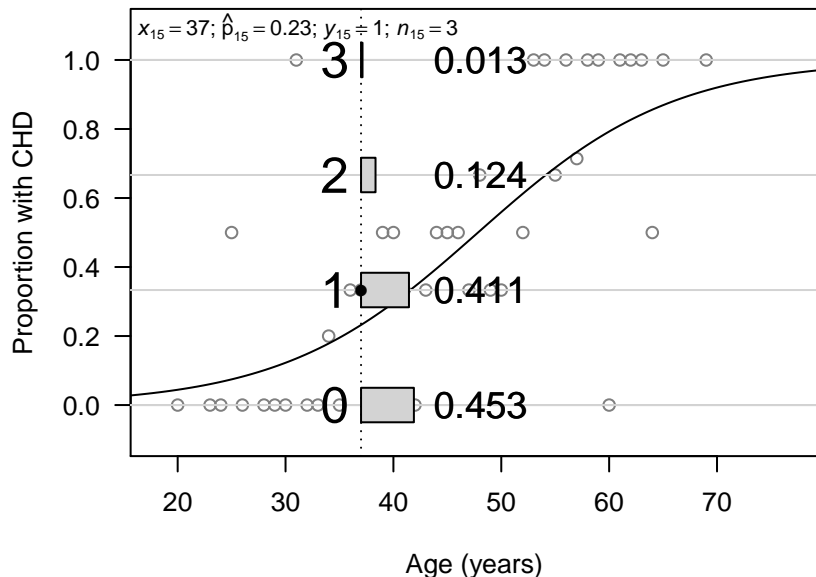
Coronary heart disease analysis

Logistic regression



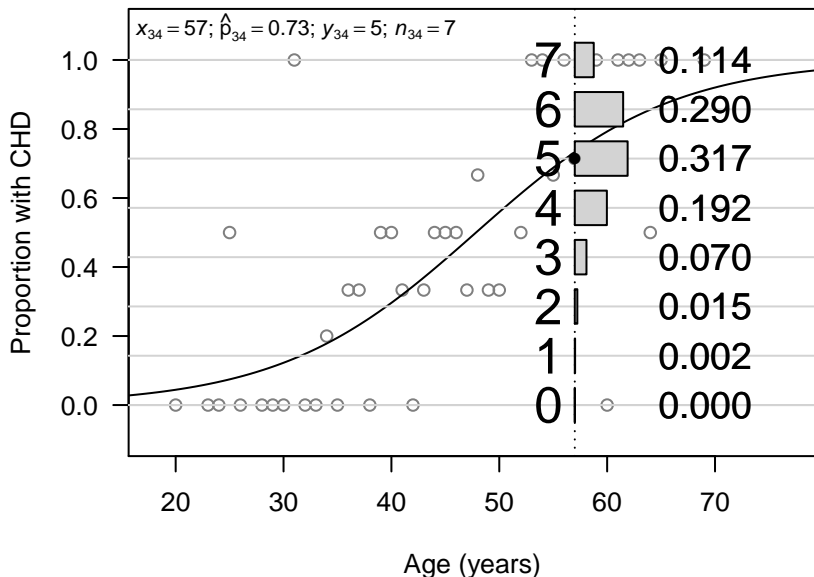
Coronary heart disease analysis

Logistic regression



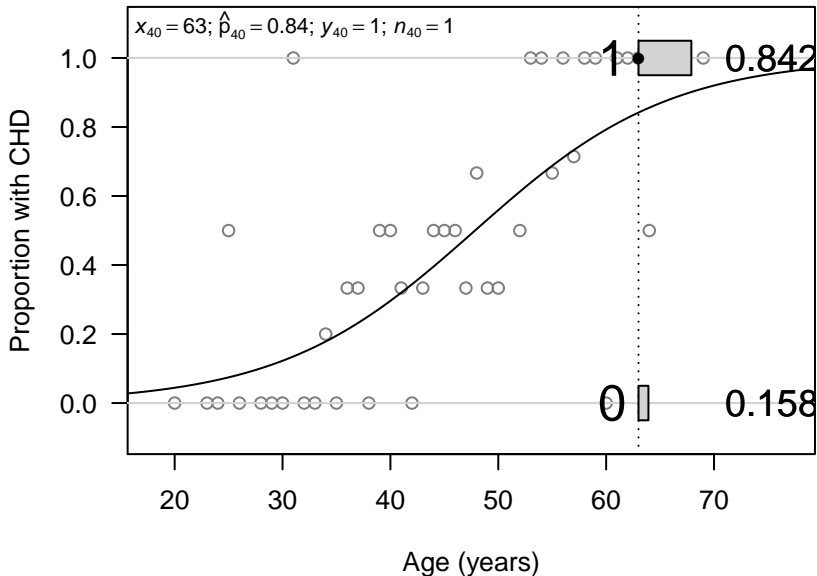
Coronary heart disease analysis

Logistic regression



Coronary heart disease analysis

Logistic regression



Generalised linear models

We can do the same things with explanatory terms for all models.

- ▶ Linear regression:

$$\mu_i = \beta_0 + \beta_1 x_i,$$
$$Y_i \sim \text{Normal}(\mu_i, \sigma^2),$$

- ▶ Poisson regression:

$$\log(\mu_i) = \beta_0 + \beta_1 x_i,$$
$$Y_i \sim \text{Poisson}(\mu_i),$$

- ▶ Logistic regression:

$$\text{logit}(p_i) = \beta_0 + \beta_1 x_i,$$
$$Y_i \sim \text{Binomial}(n_i, p_i),$$

Explanatory terms: Multiple explanatory variables

We can do the same things with explanatory terms for all models.

- ▶ Linear regression:

$$\mu_i = \beta_0 + \beta_1 x_i + \beta_2 z_i,$$

$$Y_i \sim \text{Normal}(\mu_i, \sigma^2),$$

- ▶ Poisson regression:

$$\log(\mu_i) = \beta_0 + \beta_1 x_i + \beta_2 z_i,$$

$$Y_i \sim \text{Poisson}(\mu_i),$$

- ▶ Logistic regression:

$$\text{logit}(p_i) = \beta_0 + \beta_1 x_i + \beta_2 z_i,$$

$$Y_i \sim \text{Binomial}(n_i, p_i),$$

Explanatory terms: Quadratics

We can do the same things with explanatory terms for all models.

- ▶ Linear regression:

$$\mu_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2,$$

$$Y_i \sim \text{Normal}(\mu_i, \sigma^2),$$

- ▶ Poisson regression:

$$\log(\mu_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2,$$

$$Y_i \sim \text{Poisson}(\mu_i),$$

- ▶ Logistic regression:

$$\text{logit}(p_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2,$$

$$Y_i \sim \text{Binomial}(n_i, p_i),$$

Explanatory terms: Interactions

We can do the same things with explanatory terms for all models.

- ▶ Linear regression:

$$\mu_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i z_i,$$

$$Y_i \sim \text{Normal}(\mu_i, \sigma^2),$$

- ▶ Poisson regression:

$$\log(\mu_i) = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i z_i,$$

$$Y_i \sim \text{Poisson}(\mu_i),$$

- ▶ Logistic regression:

$$\text{logit}(p_i) = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i z_i,$$

$$Y_i \sim \text{Binomial}(n_i, p_i),$$

Building a generalised linear model

- In general:

$$g(\theta_i) = \beta_0 + \beta_1 x_i + \dots$$

$$Y_i \sim \text{Distribution}(\theta_i, \gamma_i)$$

1. Choose a distribution for the response
2. Choose a parameter to relate to explanatory terms
3. Choose a link function
4. Choose explanatory terms (more on this later)
5. Additional parameters may be estimated/observed

We assume the observations are independent of one another

Building a generalised linear model: Linear regression

- ▶ In general:

$$g(\theta_i) = \beta_0 + \beta_1 x_i + \dots$$

$$Y_i \sim \text{Normal}$$

1. Choose a distribution for the response: **Normal**
2. Choose a parameter to relate to explanatory terms
3. Choose a link function
4. Choose explanatory terms (more on this later)
5. Additional parameters may be estimated/observed

We assume the observations are independent of one another

Building a generalised linear model: Linear regression

► In general:

$$\mu_i = \beta_0 + \beta_1 x_i + \dots$$

$$Y_i \sim \text{Normal}(\mu_i)$$

1. Choose a distribution for the response: **Normal**
2. Choose a parameter to relate to explanatory terms: μ_i
3. Choose a link function
4. Choose explanatory terms (more on this later)
5. Additional parameters may be estimated/observed

We assume the observations are independent of one another

Building a generalised linear model: Linear regression

► In general:

$$\mu_i = \beta_0 + \beta_1 x_i + \dots$$

$$Y_i \sim \text{Normal}(\mu_i)$$

1. Choose a distribution for the response: **Normal**
2. Choose a parameter to relate to explanatory terms: μ_i
3. Choose a link function: **Identity**
4. Choose explanatory terms (more on this later)
5. Additional parameters may be estimated/observed

We assume the observations are independent of one another

Building a generalised linear model: Linear regression

► In general:

$$\mu_i = \beta_0 + \beta_1 x_i + \dots$$

$$Y_i \sim \text{Normal}(\mu_i, \sigma^2)$$

1. Choose a distribution for the response: **Normal**
2. Choose a parameter to relate to explanatory terms: μ_i
3. Choose a link function: **Identity**
4. Choose explanatory terms (more on this later)
5. Additional parameters may be **estimated**/observed: σ^2

We assume the observations are independent of one another

Building a generalised linear model

- In general:

$$g(\theta_i) = \beta_0 + \beta_1 x_i + \dots$$

$$Y_i \sim \text{Distribution}(\theta_i, \gamma_i)$$

1. Choose a distribution for the response
2. Choose a parameter to relate to explanatory terms
3. Choose a link function
4. Choose explanatory terms (more on this later)
5. Additional parameters may be estimated/observed

We assume the observations are independent of one another

Building a generalised linear model: Poisson regression

► In general:

$$g(\theta_i) = \beta_0 + \beta_1 x_i + \dots$$

$$Y_i \sim \text{Poisson}$$

1. Choose a distribution for the response: **Poisson**
2. Choose a parameter to relate to explanatory terms
3. Choose a link function
4. Choose explanatory terms (more on this later)
5. Additional parameters may be estimated/observed

We assume the observations are independent of one another

Building a generalised linear model: Poisson regression

► In general:

$$\mu_i = \beta_0 + \beta_1 x_i + \dots$$

$$Y_i \sim \text{Poisson}(\mu_i)$$

1. Choose a distribution for the response: **Poisson**
2. Choose a parameter to relate to explanatory terms: μ_i
3. Choose a link function
4. Choose explanatory terms (more on this later)
5. Additional parameters may be estimated/observed

We assume the observations are independent of one another

Building a generalised linear model: Poisson regression

► In general:

$$\log(\mu_i) = \beta_0 + \beta_1 x_i + \dots$$

$$Y_i \sim \text{Poisson}(\mu_i)$$

1. Choose a distribution for the response: **Poisson**
2. Choose a parameter to relate to explanatory terms: μ_i
3. Choose a link function: **Log**
4. Choose explanatory terms (more on this later)
5. Additional parameters may be estimated/observed

We assume the observations are independent of one another

Building a generalised linear model: Poisson regression

► In general:

$$\log(\mu_i) = \beta_0 + \beta_1 x_i + \dots$$

$$Y_i \sim \text{Poisson}(\mu_i)$$

1. Choose a distribution for the response: **Poisson**
2. Choose a parameter to relate to explanatory terms: μ_i
3. Choose a link function: **Log**
4. Choose explanatory terms (more on this later)
5. Additional parameters may be estimated/observed: **None**

We assume the observations are independent of one another

Building a generalised linear model

- In general:

$$g(\theta_i) = \beta_0 + \beta_1 x_i + \dots$$

$$Y_i \sim \text{Distribution}(\theta_i, \gamma_i)$$

1. Choose a distribution for the response
2. Choose a parameter to relate to explanatory terms
3. Choose a link function
4. Choose explanatory terms (more on this later)
5. Additional parameters may be estimated/observed

We assume the observations are independent of one another

Building a generalised linear model: Logistic regression

- In general:

$$g(\theta_i) = \beta_0 + \beta_1 x_i + \dots$$

$$Y_i \sim \text{Binomial}$$

1. Choose a distribution for the response: **Binomial**
2. Choose a parameter to relate to explanatory terms
3. Choose a link function
4. Choose explanatory terms (more on this later)
5. Additional parameters may be estimated/observed

We assume the observations are independent of one another

Building a generalised linear model: Logistic regression

► In general:

$$p_i = \beta_0 + \beta_1 x_i + \dots$$

$$Y_i \sim \text{Binomial}(p_i)$$

1. Choose a distribution for the response: **Binomial**
2. Choose a parameter to relate to explanatory terms: p_i
3. Choose a link function
4. Choose explanatory terms (more on this later)
5. Additional parameters may be estimated/observed

We assume the observations are independent of one another

Building a generalised linear model: Logistic regression

► In general:

$$\log \left(\frac{p_i}{1 - p_i} \right) = \beta_0 + \beta_1 x_i + \dots$$
$$Y_i \sim \text{Binomial}(p_i)$$

1. Choose a distribution for the response: **Binomial**
2. Choose a parameter to relate to explanatory terms: p_i
3. Choose a link function: **Logit**
4. Choose explanatory terms (more on this later)
5. Additional parameters may be estimated/observed

We assume the observations are independent of one another

Building a generalised linear model: Logistic regression

► In general:

$$\log \left(\frac{p_i}{1 - p_i} \right) = \beta_0 + \beta_1 x_i + \dots$$
$$Y_i \sim \text{Binomial}(n_i, p_i)$$

1. Choose a distribution for the response: **Binomial**
2. Choose a parameter to relate to explanatory terms: p_i
3. Choose a link function: **Logit**
4. Choose explanatory terms (more on this later)
5. Additional parameters may be estimated/**observed**: n_i

We assume the observations are independent of one another

There's more!

Other distributions:

- ▶ Gamma and inverse-Gaussian, for continuous responses on the interval $[0, \infty)$
- ▶ Beta, for continuous responses on the interval $[0, 1]$
- ▶ Negative binomial, for discrete responses on $\{0, 1, 2, \dots\}$, with $\text{Var}(Y) \geq E(Y)$
- ▶ ... And so on

How to do it in R

```
fit <- glm(y ~ x, family = "distribution")
```

Linear regression: LSD analysis

```
## Either  
lsd.fit <- lm(score ~ lsd)  
## Or  
lsd.fit <- glm(score ~ lsd, family = "gaussian")
```

Poisson regression: *M. Ornithogaster* analysis

```
chickens.fit <- glm(mo ~ weight, family = "poisson")
```

Logistic regression: CHD analysis:

```
chd.fit <- glm(cbind(y, n - y) ~ age, family = "binomial")
```

Matrix representation

Sometimes GLMs are specified in an alternative way, using matrix multiplication:

$$g(\theta) = \mathbf{X}\beta$$

$$\begin{bmatrix} g(\theta_1) \\ g(\theta_2) \\ g(\theta_3) \\ g(\theta_4) \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 x_1 \\ \beta_0 + \beta_1 x_2 \\ \beta_0 + \beta_1 x_3 \\ \beta_0 + \beta_1 x_4 \\ \vdots \end{bmatrix}$$

The matrix \mathbf{X} is called the design matrix. Its first column is full of ones. The subsequent column(s) contain the explanatory terms for each observation.

The vector β contains the coefficients we wish to estimate.