Introduction to Generalised Linear Models (GLMs)

A simple linear regression model... in a new light

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

A simple linear regression model... in a new light

Intercept

Coefficient

Explanatory Variable

Response

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

A simple linear regression model... in a new light

Intercept

Coefficient

Explanatory Variable

Response

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

Error Term

$$\epsilon_i \sim \mathrm{Normal}(0, \sigma^2)$$

Normally Distributed

Mean of 0

Equal Variance

A simple linear regression model... in a new light

We can attribute the randomness directly to the response variable instead:

Response

$$Y_i \sim \text{Normal}(\underbrace{\alpha + \beta_1 x_i}_{\substack{\text{Mean is a linear} \\ \text{combination of the} \\ \text{explanatory terms}}}, \sigma^2)$$

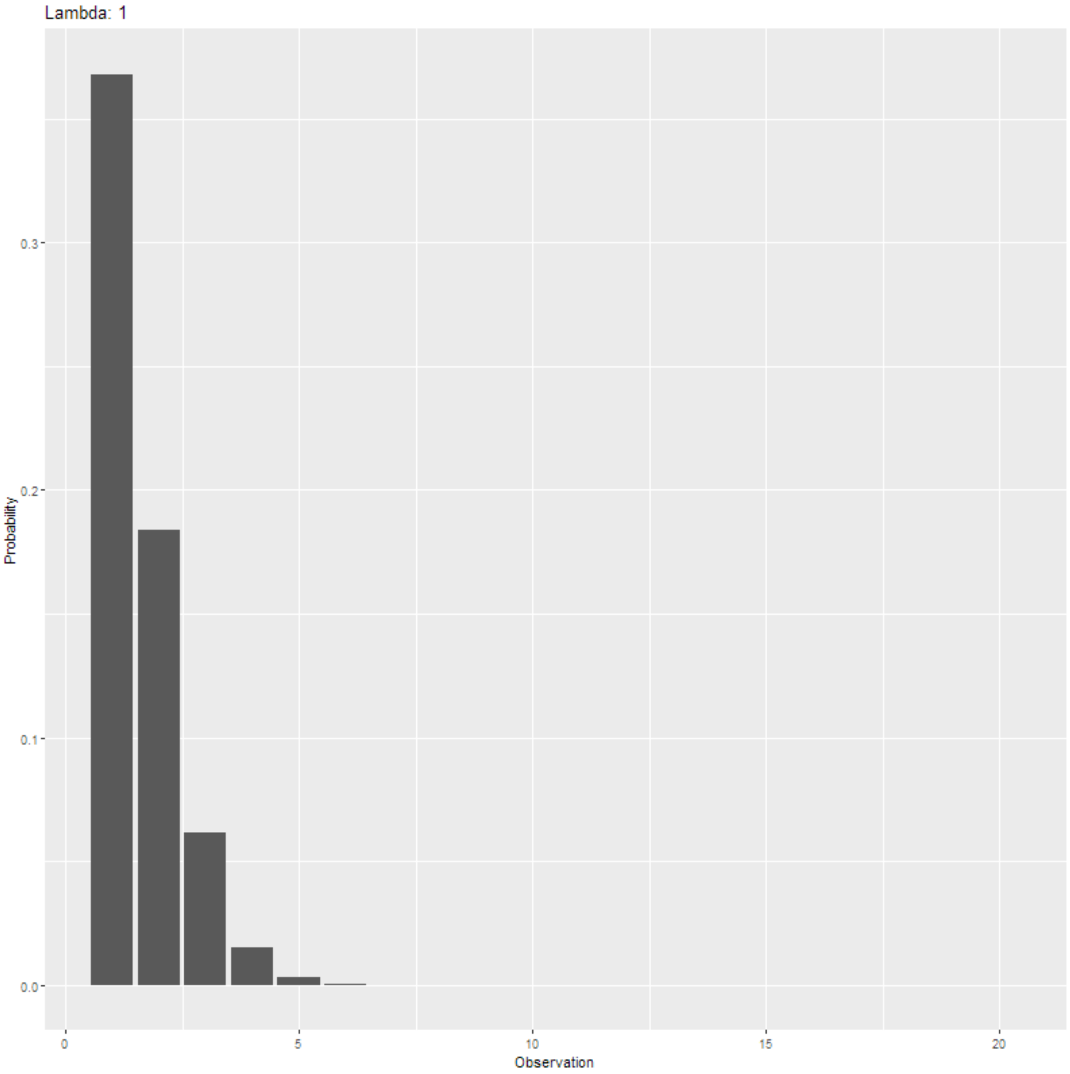
Normally Distributed

Variance, σ^2

Assumptions

- The i^{th} response, Y_i, comes from a normal distribution
- The mean of Y_i is a **linear** combination of the explanatory terms
- The variance of Y_i, σ^2, is the same for all observations
- Each observation's response is independent of all others

A Fishy Regression: Poisson



A Fishy Regression: Poisson

The Poisson distribution is a *discrete* distribution (of positive values only) and we expect the variance to increase with the mean.

$$Y_i \sim \mathrm{Poisson}(??)$$

A Fishy Regression: Poisson

The Poisson distribution is a *discrete* distribution (of positive values only) and we expect the variance to increase with the mean.

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

A Fishy Regression: Poisson

The Poisson distribution is a *discrete* distribution (of positive values only) and we expect the variance to increase with the mean.

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

X $\mu < 0$

$$\alpha + \beta_1 x_i$$

A Fishy Regression: Poisson

The Poisson distribution is a *discrete* distribution (of positive values only) and we expect the variance to increase with the mean.

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

$$\log(\mu_i) = \alpha + \beta_1 x_i$$

Link Function

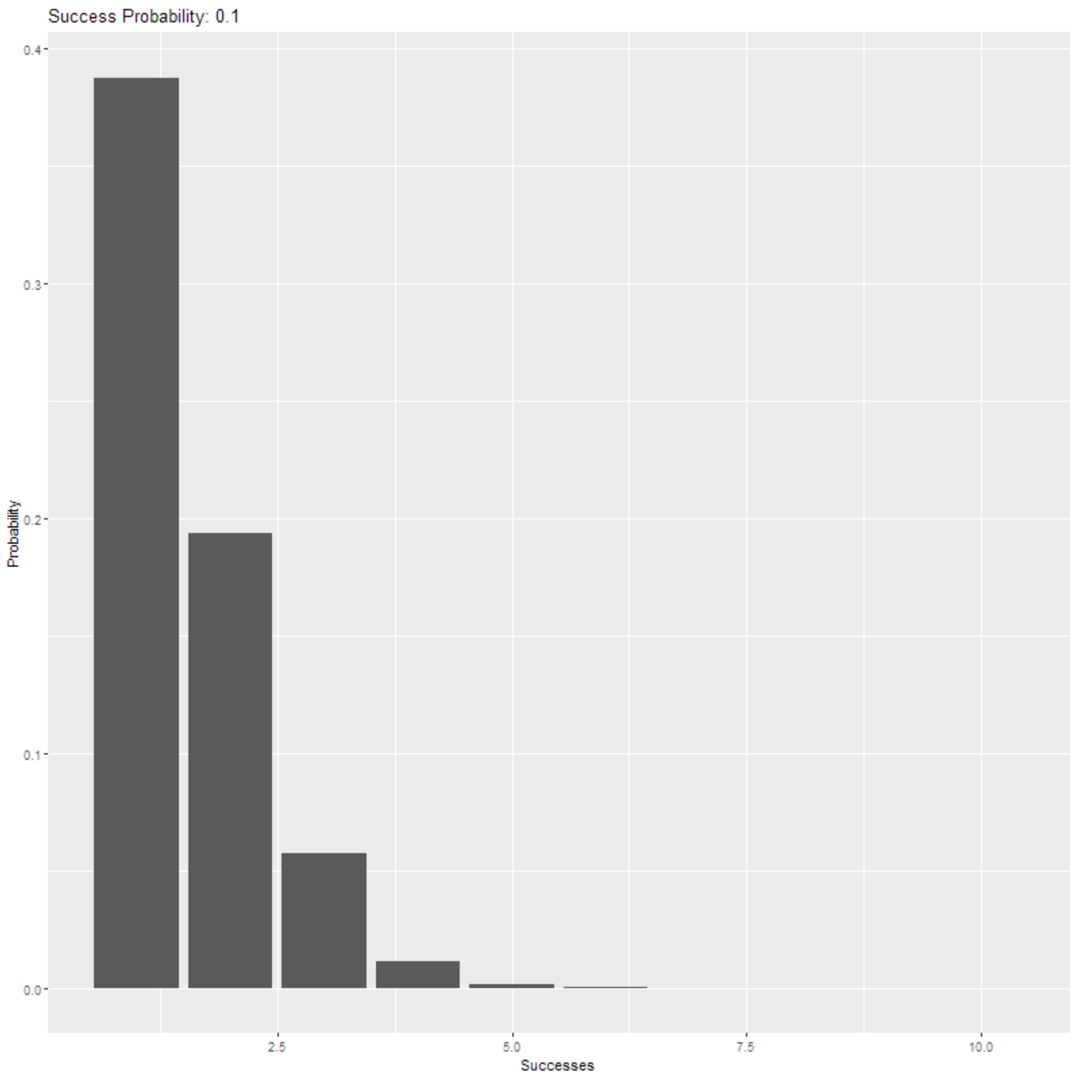
A Fishy Regression: Poisson

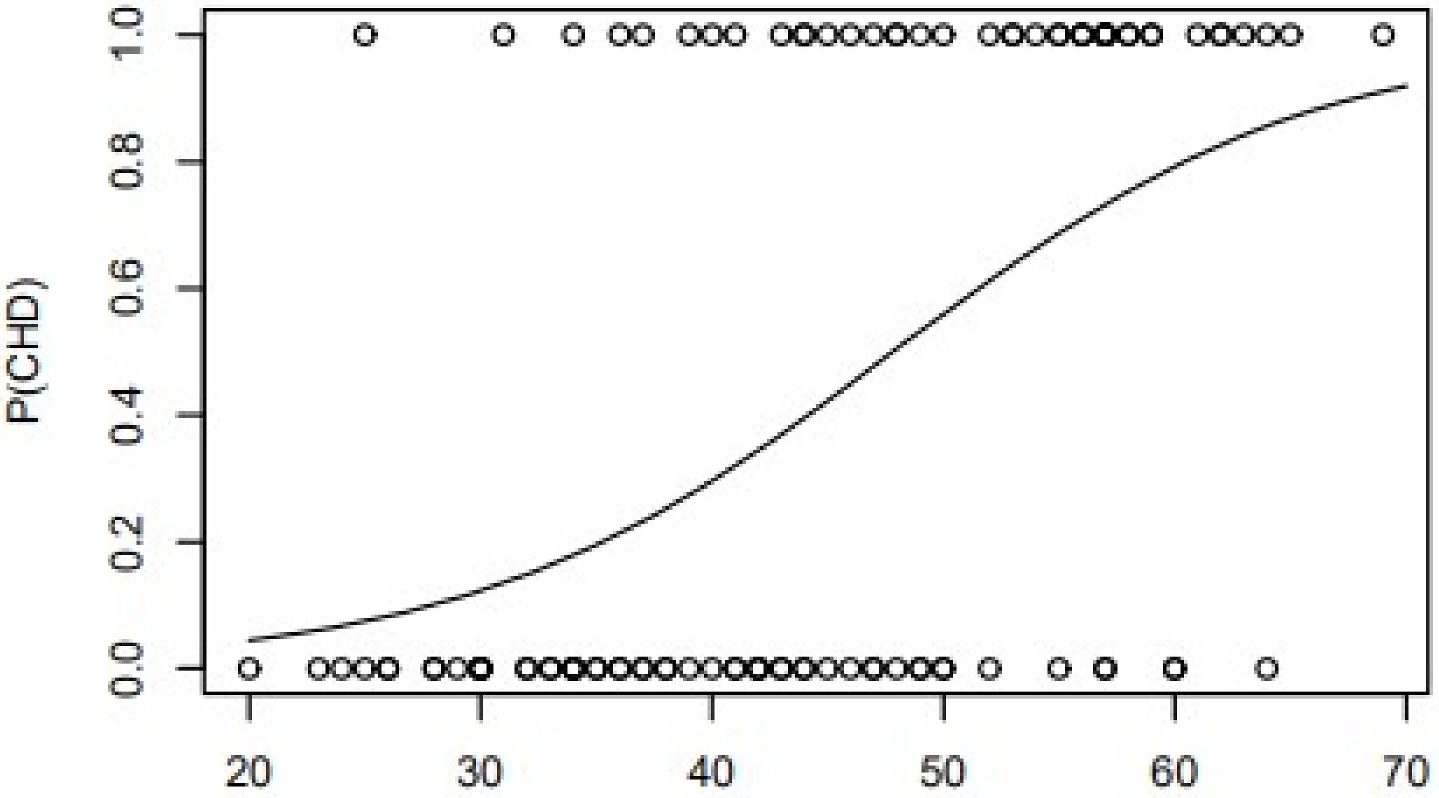
The Poisson distribution is a *discrete* distribution (of positive values only) and we expect the variance to increase with the mean.

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

$$\mu_i = \exp(\alpha + \beta_1 x_i)$$

Success or No Success? Logistic Regression




P(CHD)
0.0
0.2
0.4
0.6
0.8
1.0
20
30
40
50
60
70

Success or No Success? Logistic

Logistic Regression uses a binomial distribution where the number of successes from a number of independent trials, *n*, each have the same probability of success, *p*.

$$Y_i \sim \mathrm{Binomial}(??)$$

Success or No Success? Logistic

Logistic Regression uses a binomial distribution where the number of successes from a number of independent trials, *n*, each have the same probability of success, *p*.

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

Success or No Success? Logistic

Logistic Regression uses a binomial distribution where the number of successes from a number of independent trials, *n*, each have the same probability of success, *p*.

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

X $p < 0, p > 1$

$$p_i = \alpha + \beta_1 x_i$$

Success or No Success? Logistic

Logistic Regression uses a binomial distribution where the number of successes from a number of independent trials, *n*, each have the same probability of success, *p*.

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

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

Success or No Success? Logistic

Logistic Regression uses a binomial distribution where the number of successes from a number of independent trials, *n*, each have the same probability of success, *p*.

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

$$p_i = \frac{\exp(\alpha + \beta_1 x_i)}{1 + \exp(\alpha + \beta_1 x_i)}$$

Just Three Examples of Many

Linear regression: $Y_i \sim \mathrm{Normal}(\mu_i, \sigma^2)$ where $\mu_i = \alpha + \beta_1 x_i$

Poisson regression: $Y_i \sim \mathrm{Poisson}(\mu_i)$ where $\log(\mu_i) = \alpha + \beta_1 x_i$

Logistic regression: $Y_i \sim \mathrm{Binomial}(n_i, p_i)$ where $\mathrm{logit}(p_i) = \alpha + \beta_1 x_i$

Fitting Generalised Linear Models

```
glm(formula, family = “my choice”, data = my_data, …)
```

Building a GLM

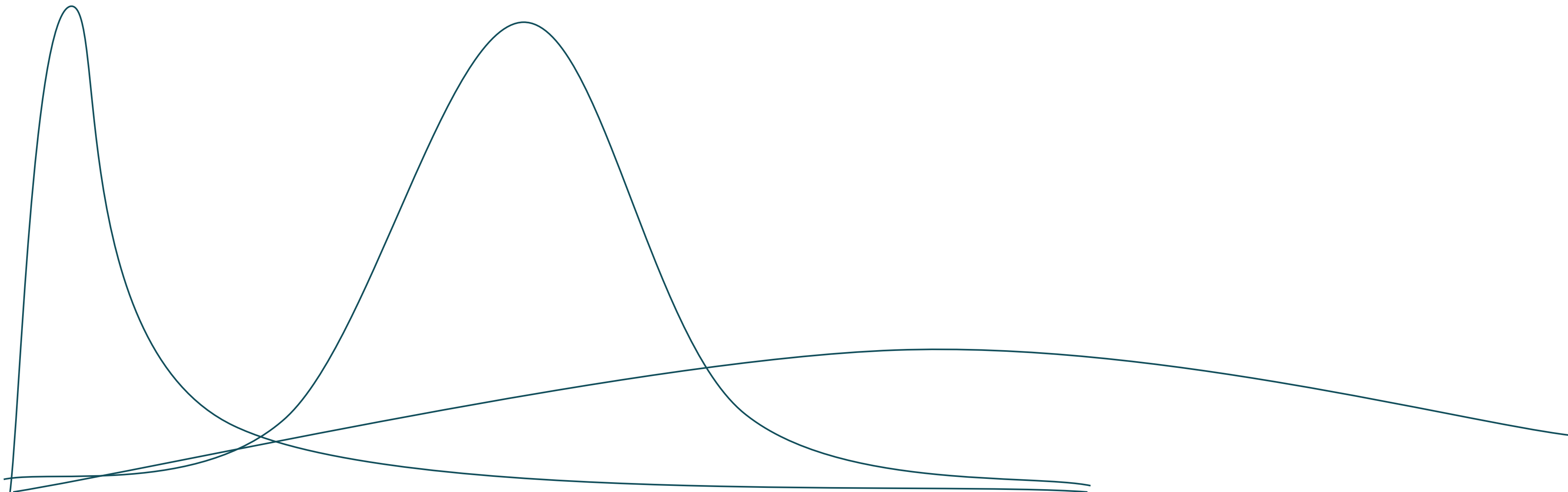
1. Assume the observations are independent of one another
2. Choose a distribution for the response
3. Choose a parameter to relate to explanatory terms
4. Choose a link function
5. Choose explanatory terms
6. Estimate additional parameters

Three 'typical' distributions

Linear regression: $Y_i \sim \mathrm{Normal}(\mu_i, \sigma^2)$ where $\mu_i = \alpha + \beta_1 x_i$

Poisson regression: $Y_i \sim \mathrm{Poisson}(\mu_i)$ where $\log(\mu_i) = \alpha + \beta_1 x_i$

Logistic regression: $Y_i \sim \mathrm{Binomial}(n_i, p_i)$ where $\mathrm{logit}(p_i) = \alpha + \beta_1 x_i$

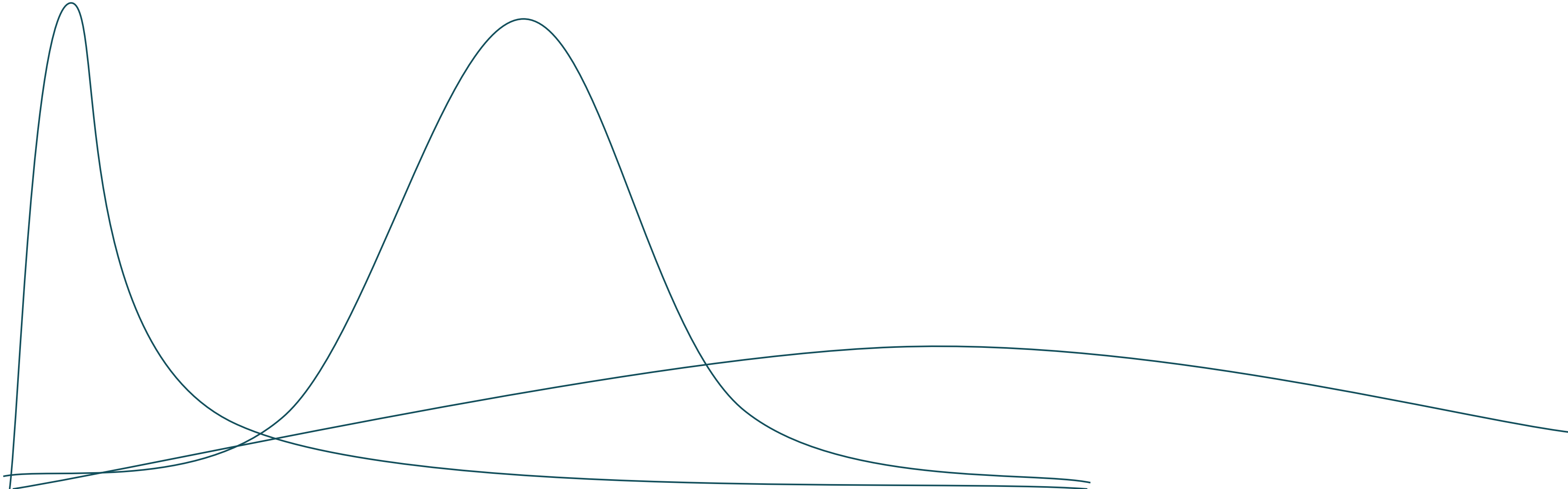


Three ‘typical’ distributions

Linear regression: $Y_i \sim \text{Normal}(\mu_i, \sigma^2)$ where $\mu_i = \alpha + \beta_1 x_i$

Poisson regression: $Y_i \sim \text{Poisson}(\mu_i)$ where $\log(\mu_i) = \alpha + \beta_1 x_i$

Logistic regression: $Y_i \sim \text{Binomial}(n_i, p_i)$ where $\text{logit}(p_i) = \alpha + \beta_1 x_i$

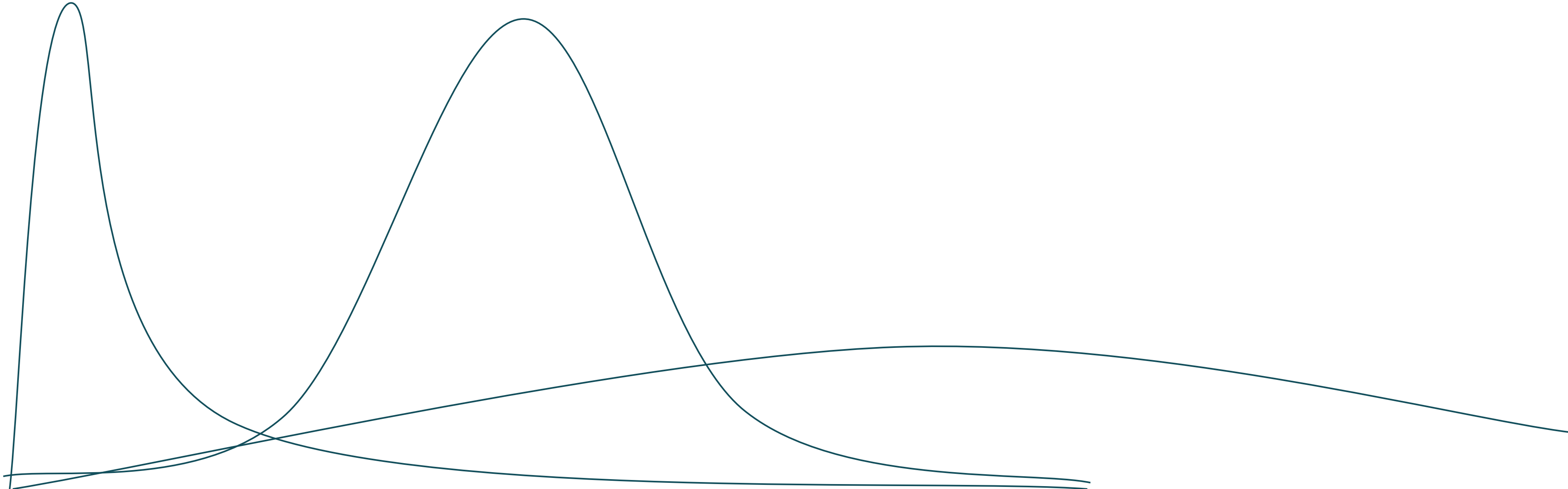


Three ‘typical’ distributions

Linear regression: $Y_i \sim \mathrm{Normal}(\mu_i, \sigma^2)$ where $\mu_i = \alpha + \beta_1 x_i$

Poisson regression: $Y_i \sim \mathrm{Poisson}(\mu_i)$ where $\log(\mu_i) = \alpha + \beta_1 x_i$

Logistic regression: $Y_i \sim \mathrm{Binomial}(n_i, p_i)$ where $\mathrm{logit}(p_i) = \alpha + \beta_1 x_i$

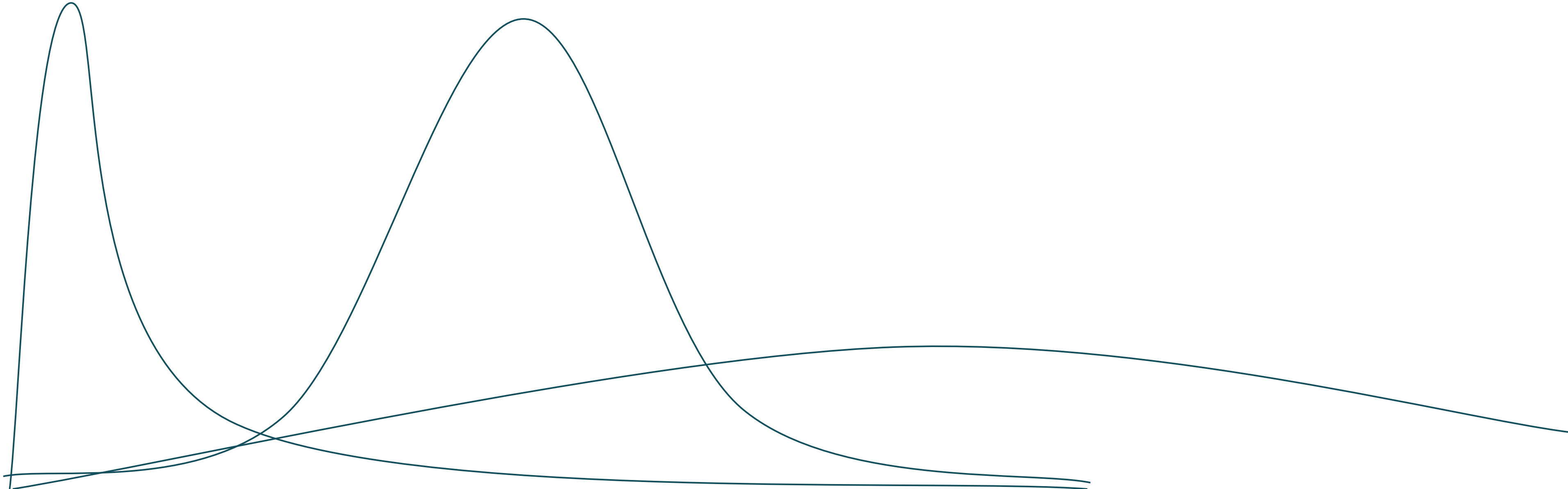


Three ‘typical’ distributions

Linear regression: $Y_i \sim \mathrm{Normal}(\mu_i, \sigma^2)$ where $\mu_i = \alpha + \beta_1 x_i$

Poisson regression: $Y_i \sim \mathrm{Poisson}(\mu_i)$ where $\log(\mu_i) = \alpha + \beta_1 x_i$

Logistic regression: $Y_i \sim \mathrm{Binomial}(n_i, p_i)$ where $\mathrm{logit}(p_i) = \alpha + \beta_1 x_i$

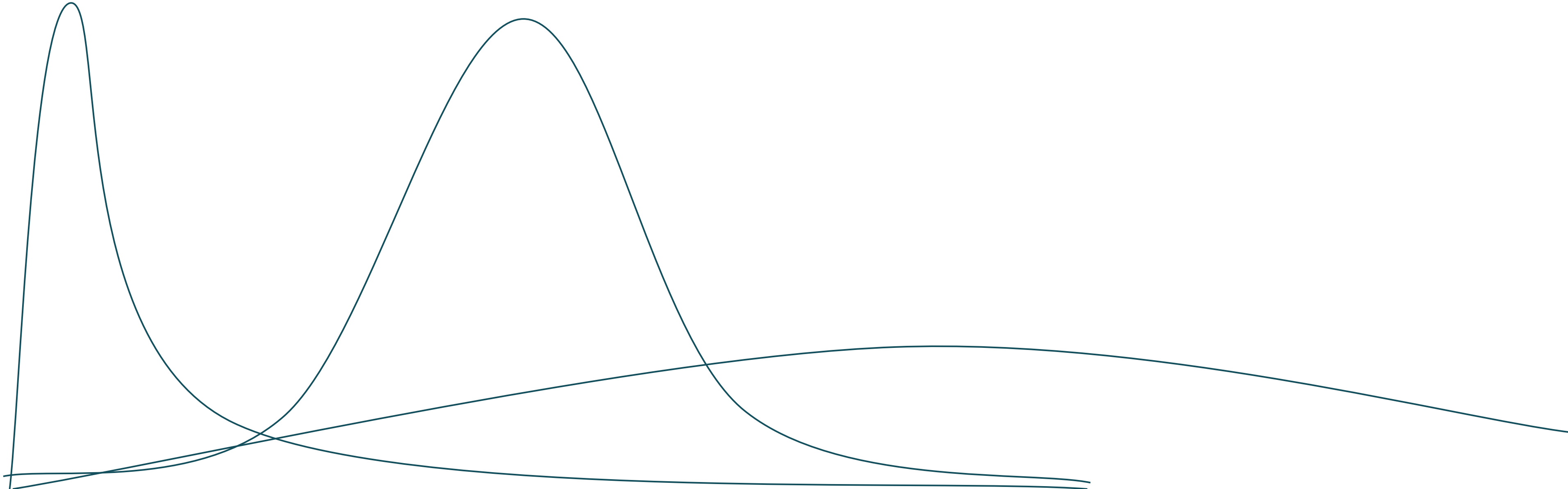


Three ‘typical’ distributions

Linear regression: $Y_i \sim \mathrm{Normal}(\mu_i, \sigma^2)$ where $\mu_i = \alpha + \beta_1 x_i$

Poisson regression: $Y_i \sim \mathrm{Poisson}(\mu_i)$ where $\log(\mu_i) = \alpha + \beta_1 x_i$

Logistic regression: $Y_i \sim \mathrm{Binomial}(n_i, p_i)$ where $\mathrm{logit}(p_i) = \alpha + \beta_1 x_i$



But there are many more potential response distributions…

`family = …`

- `“binomial”`
- `“gaussian”`
- `“Gamma”`
- `“inverse.gaussian”`
- `“poisson”`
- `“quasi”`
- `“quasibinomial”`
- `“quasipoisson”`

With many more choices of link functions…

`family = …`

- `binomial(link = "logit")`
- `gaussian(link = "identity")`
- `Gamma(link = "inverse")`
- `inverse.gaussian(link = "1/mu^2")`
- `poisson(link = "log")`
- `quasi(link = "identity", variance = "constant")`
- `quasibinomial(link = "logit")`
- `quasipoisson(link = "log")`

Distribution	Notation	Mean	Variance	Linear predictor (link function)
Gaussian	$Y \sim \mathbf{Normal}(\mu, \sigma^2)$	μ	σ^2	$\mathrm{I}(\mu) = \alpha + \Sigma_{j=1}^{n_{\text{covariates}}} \beta_j x_j$
Poisson	$Y \sim \mathbf{Poisson}(\mu) \text{where} \mu = \text{rate}$	μ	μ	$\log(\mu) = \alpha + \Sigma_{j=1}^{n_{\text{covariates}}} \beta_j x_j$
Binomial	$Y \sim \mathbf{Bonomial}(\mathrm{n}, p)$ where n = number of trials and p = probability of success	$\mathrm{n}p$	$\mathrm{n}p(1-p)$	$\mathrm{logit}(p) = \alpha + \Sigma_{j=1}^{n_{\text{covariates}}} \beta_j x_j$
Gamma	$Y \sim \mathbf{Gamma}(k, \theta = \frac{1}{\text{rate}})$ where k = shape and θ = scale	$k\theta$	$k\theta^2$	$\log(E(Y)) = \alpha + \Sigma_{j=1}^{n_{\text{covariates}}} \beta_j x_j$
Beta	$Y \sim \mathbf{Beta}(a, b)$ where a = shape and b = scale	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$	$\log(E(Y)) = \alpha + \Sigma_{j=1}^{n_{\text{covariates}}} \beta_j x_j$
Negative binomial	$Y \sim \mathbf{NB}(r, p)$ where r = number of successes until the experiment is stopped and p = probability of success or $Y \sim \mathbf{NB}(k, p)$ where k = number of failures given p = probability of success	$\frac{r(1-p)}{p}$ or $\mu = k\frac{p}{1-p}$	$\frac{r(1-p)}{p^2}$ or $\mu + \frac{\mu^2}{k}$	$\log(E(Y)) = \alpha + \Sigma_{j=1}^{n_{\text{covariates}}} \beta_j x_j$
Beta-binomial	$Y \sim \mathbf{BetaBin}(\mathrm{n}, a, b)$ where n = number of trials and $p = \frac{a}{a+b}$, the probability of success	$\frac{\mathrm{n}a}{a+b} = \mathrm{n}p$	$\frac{\mathrm{n}ab(a+b+\mathrm{n})}{(a+b)^2(a+b+1)}$	$\mathrm{logit}(p) = \alpha + \Sigma_{j=1}^{n_{\text{covariates}}} \beta_j x_j$

Function and arguments	Response	Random effects
• `lm()` ○ `function (formula, data, …)`	Gaussian	No
• `glm()` ○ `function (formula, family = gaussian, …)`	Any	No
• `lme4::lmer()` ○ `function (formula, data, …)`	Gaussian	Yes
• `lme4::glmer()` ○ `function (formula, data, family = gaussian, …)`	Any	Yes
• `glmmTMB::glmmTMB()` ○ `function (formula, data, family = gaussian(), …)`	Any	Yes