Linear Regression $\longrightarrow$ Generalized Linear model

$$\boxed{y_i = x_i \beta + \varepsilon_i}$$

i : $1, \dots, n$ (observations)

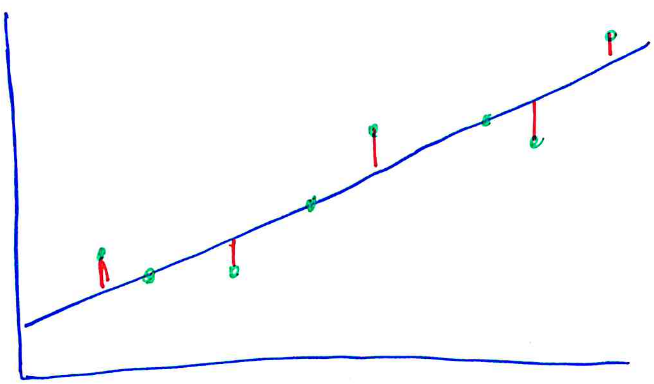
y_i : k outcome values (DV)

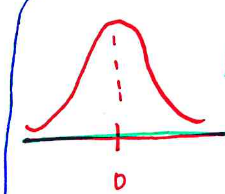
x_i : vector of k IVs / predictors

β : $k \times 1$ vector of unknown parameters

ε_i : zero mean stochastic disturbances (error)

ε_i independent across observations, constant variance σ^2, $\sim N$





$\varepsilon_i \quad N(\mu, \sigma^2)$

1 — stochastic component

$y_i \overset{iid}{\sim} N(\mu_i, \sigma^2)$

② systematic component

covariates x_i
combine linearly with
coefficients to form
linear predictor – $\boxed{\eta_i = X_i \beta}$

$$x_0 \beta_0 + x_1 \beta_1 + x_2 \beta_2 + \ldots + x_n \beta_n + \varepsilon_i = y_i$$

"linear" regression
because systematic component
is additive

③ link function

linear predictor

$\boxed{X_i\beta = \eta_i}$ is a function of mean parameter $\boxed{\mu_i}$ via a link function,

$$\boxed{\eta_i = g(\mu_i)}$$

in the case of the normal linear model (linear regression):

$$g(\mu_i) = \mu_i \leftarrow \text{"Identity function"} \quad f(x) = x$$

$$\mu_i = \eta_i$$

Generalized Linear Models

1. stochastic component
2. systematic component
3. link function

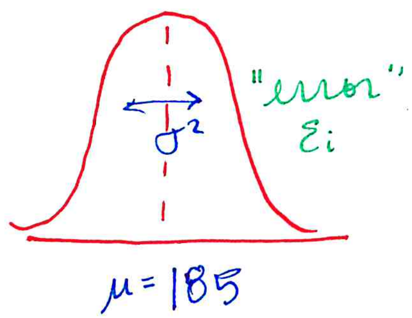
stays the same

Stochastic components (besides $y_i \overset{iid}{\sim} N(\mu, \sigma^2)$) and link functions (besides $\eta_i = g(\mu_i) = \mu_i$) vary to create other types of GLMs.

Stochastic component

for a given level of $X_i\beta$, what is the distribution of y_i?

for men 5'11", what is distribution of weight?



Linear regression (general linear model) - stochastic component is the normal / Gaussian distribution

Other distributions:

"exponential family"

- beta
- binomial
- Dirichlet
- Pareto
- Poisson
- Bernoulli
- Exponential
- ...

NOT exponential?

- uniform
- Student's t

Link functions

could be: any monotonic, differentiable function

in practice: $g(\mu_i)$ functions are used if inverse link is easy to calculate

Link: $g(\mu_i) = \eta_i$

inverse link: $\mu_i = g^{-1}(\eta_i)$

and, g^{-1} maps $x_i\beta = \eta_i \in \mathbb{R}$ into set of admissible values for μ_i

"canonical links"

Poisson model $\rightarrow$ log link

Binomial $\rightarrow$ logit

$\rightarrow$ probit

$\rightarrow$ complementary

Logistic Regression

1. Stochastic component

Binary outcome variable

$y \in \{0, 1\}$

$\downarrow$

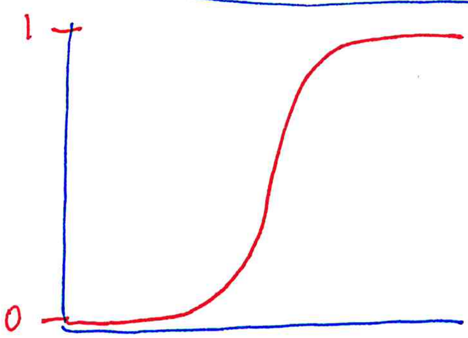
Bernoulli distribution

$$p(y|\pi) = \pi^y (1-\pi)^{1-y}$$

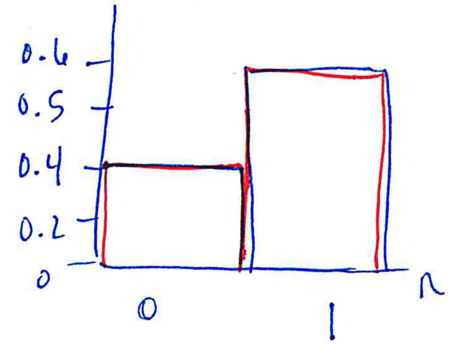
$$\eta = \psi(\pi) = \log\left(\frac{\pi}{1-\pi}\right) \leftarrow \text{log odds ratio}$$

$$\pi = \psi^{-1}(\eta) = \frac{1}{1+e^{-\eta}}$$

$$\boxed{\pi = \frac{1}{1+e^{-\eta}}}$$ — logistic function "logit link"



P(n) for p=0.6



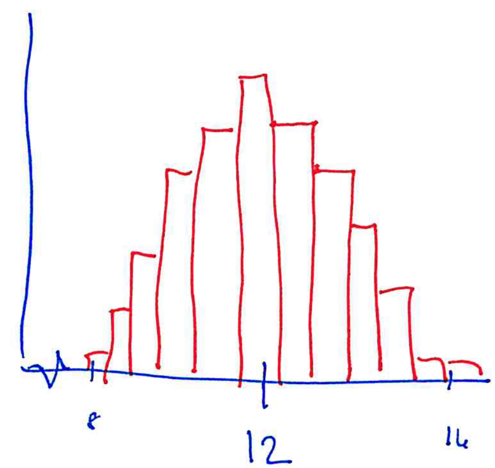
Bernoulli

$X_1, \ldots X_n$ iid rand vars,

Bernoulli $P(n) = p$:

$$Y = \sum_{k=1}^{n} X_k \sim B(n, p)$$

Binomial



$B(0.6, 20)$

$\mu = 12$

$\sigma = 20 \times .6 \times .$

$= 4.8$