☐ Syllabus

☐ Paper Guidelines

☐ LM (1), GLM (2), GLIM (3), generalized GLIM (4):

$\forall$: y_i, $i = 1, \ldots, n$ dependent var's

$\underset{\sim}{x}_i$, $i = 1, \ldots, n$, $\underset{\sim}{x}_i \in \mathbb{R}^p$ – explanatory

☐ (1) LM:

$$y_i = \underbrace{\underset{\sim}{x}_i^t \underset{\sim}{\beta}}_{\mu_i} + \varepsilon_i , \quad \varepsilon_i \begin{cases} E(\varepsilon_i) = 0, \ var(\varepsilon_i) = \sigma^2 \\ \varepsilon_i \sim N(0, \sigma^2) \end{cases}$$

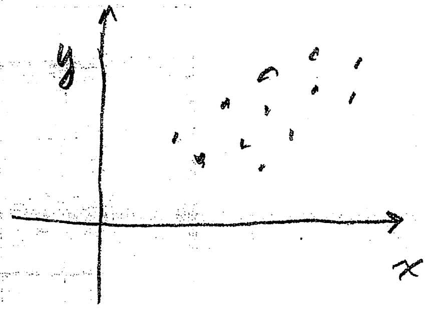
$$\Rightarrow \quad \underset{\sim}{y} = \underset{n \times p}{X} \cdot \underset{\sim}{\beta} + \underset{\sim}{\varepsilon}, \quad \begin{cases} E(\underset{\sim}{\varepsilon}) = 0, \ cov(\underset{\sim}{\varepsilon}) = \sigma^2 I \\ \underset{\sim}{\varepsilon} \sim N(\underset{\sim}{0}, \sigma^2 I) \end{cases}$$

$$\hat{\underset{\sim}{\beta}} : \begin{cases} \text{minimize } SSE = \| X\hat{\underset{\sim}{\beta}} - X\underset{\sim}{\beta} \|^2 \ (OLS) \\ MLE \end{cases}$$

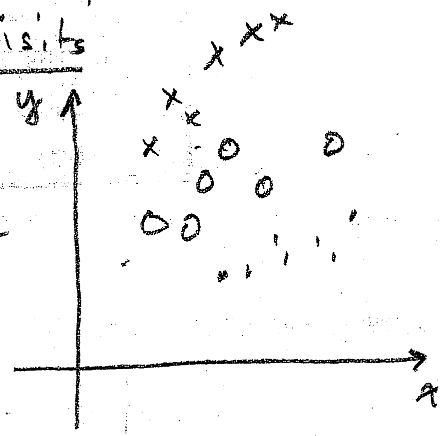
□ (2) GLM :

- y_i = Blood Pressure , x_i = cholest. , n patients

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \Leftarrow$$



- m patients , repeated visits



$i = 1, \ldots, m$ patients

$j = 1, 2, \ldots, k_i$ obs / patient

$\Rightarrow$

$$y_{ij} = (\beta_0 + \delta_i) + (\beta_1 + \gamma_i)\, x_{ij} + \varepsilon^*_{ij}$$

LM w̄ random effects

□ Model? Estimate?

Simplest case: $y_{ij} = \beta_0 + \delta_i + \beta_1 x_{ij} + \varepsilon^*_{ij}$

(same slope) , $i = 1, \ldots, m$; $j = 1, \ldots, k$

$$\delta_i \underset{ind}{\sim} N(0, \sigma^2_\delta) \;,\; \varepsilon^*_{ij} \underset{ind}{\sim} N(0, \sigma^2_{\varepsilon^*})$$

(ind)

□ Then

$$\underset{\sim}{y} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1k} \\ \vdots \\ y_{m1} \\ \vdots \\ y_{mk} \end{bmatrix} = \begin{bmatrix} 1 & x_{11} \\ \vdots & \vdots \\ 1 & x_{1k} \\ \vdots & \vdots \\ 1 & x_{m1} \\ \vdots & \vdots \\ 1 & x_{mk} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \delta_1 + \varepsilon^*_{11} \\ \vdots \\ \delta_1 + \varepsilon^*_{1k} \\ \vdots \\ \delta_m + \varepsilon^*_{m1} \\ \vdots \\ \delta_m + \varepsilon^*_{mk} \end{bmatrix} \quad \left\{ \underset{\sim}{y} = X \underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \right\}$$

□ So $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$

where $\varepsilon_{ij} = \delta_i + \varepsilon^*_{ij}$, $\left.\begin{array}{l}\delta_i \underset{ind}{\sim} N(0, \sigma^2_\delta) \\ \varepsilon^*_{ij} \underset{ind}{\sim} N(0, \sigma^2_{\varepsilon^*})\end{array}\right\}$ ind

Thus

$$\underset{\sim}{\varepsilon} = \begin{bmatrix} \begin{bmatrix} \delta_1 + \varepsilon^*_{11} \\ \vdots \\ \delta_1 + \varepsilon^*_{1k} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \delta_m + \varepsilon^*_{m1} \\ \vdots \\ \delta_m + \varepsilon^*_{mk} \end{bmatrix} \end{bmatrix} = k\left\{\begin{bmatrix} \bar{\underset{\sim}{\varepsilon}}_{(1)} \\ \bar{\underset{\sim}{\varepsilon}}_{(2)} \\ \vdots \\ \bar{\underset{\sim}{\varepsilon}}_{(m)} \end{bmatrix}\right. , \quad E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$$

□ cov(

$$cov(\underset{\sim}{\varepsilon}) = \overbrace{}^{k} k\left\{\begin{bmatrix} \boxed{\begin{matrix} (\sigma^2_\delta + \sigma^2_{\varepsilon^*}) & \sigma^2_\delta \cdots & \sigma^2_\delta \\ \vdots & \ddots & \\ \sigma^2_\delta & \cdots & (\sigma^2_\delta + \sigma^2_{\varepsilon^*}) \end{matrix}} & & 0 & & 0 \\ & \boxed{\begin{matrix} (\sigma^2_\delta + \sigma^2_{\varepsilon^*}) & \cdots \sigma^2_\delta \\ \vdots & \ddots \\ \sigma^2_\delta & (\sigma^2_\delta + \sigma^2_{\varepsilon^*}) \end{matrix}} & & \\ 0 & & \ddots & \\ 0 & & & \boxed{} \end{bmatrix}\right.$$

$= V(\sigma^2_\varepsilon, \rho)$, block-diagonal $\bar{w}$ intra-class

cov matrices $\sigma^2_\varepsilon \begin{bmatrix} 1 & \rho \cdots & \rho \\ \rho & \ddots & \rho \\ \rho & & 1 \end{bmatrix}$ on diagonal

□ MLE of $\underset{\sim}{\beta}$:

If $\underset{\sim}{y} \sim N(X\underset{\sim}{\beta}, V)$ then

$$f = lik = \frac{1}{(2\pi)^{n/2}} \cdot \frac{1}{|V|^{1/2}} e^{-\frac{1}{2}(y - X\beta)^t V^{-1}(y - X\beta)}$$

$\hat{\beta}$ minimizes $(y - X\beta)^t V^{-1}(y - X\beta)$ (*)

$\Rightarrow$ {see page ③}

$$\hat{\beta} = (X^t V^{-1} X)^{-1} X^t V^{-1} y$$

□ Remarks:

(1) $V = \sigma^2 I \Rightarrow$ $\hat{\beta}$ minimizes $\frac{1}{\sigma^2}(y - X\beta)^t(y - X\beta)$

$$\Rightarrow \hat{\beta} = (X^t X)^{-1} X^t y$$

{don't need σ^2 to compute $\hat{\beta}$}

$\hat{\beta}$ is OLS estimate

(2) $V = V(\sigma^2, \sigma^2_\delta) \Rightarrow$ need to know V to minimize (*).

$\hat{\beta}$ is WLS (weighted LS) estimate

□ Weights? If $V = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ then

(*) says: minimize $\sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{\sigma_i^2}$, so are giving more weight to $(y_i - \mu_i)^2$ when y_i is more accurate (σ_i^2 is smaller)

{ To min: $(y - X\beta)^t V^{-1} (y - X\beta)$:

$U = \begin{bmatrix} u_1 & \cdots & u_p \\ | & & | \end{bmatrix}$ eig vectors of V

$\Rightarrow \quad U^t V U = \Lambda = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \\ 0 & & \lambda_p \end{bmatrix}, \qquad U^t U = I$

$\Rightarrow \quad U^{-1} V^{-1} (U^t)^{-1} = \Lambda^{-1}$

$\Rightarrow \quad V^{-1} = U \Lambda^{-1} U^t$

$= T \cdot T^t, \qquad T = \begin{bmatrix} \lambda_1^{-1/2} \cdot \underset{\sim}{u}_1 & \cdots & \lambda_p^{-1/2} \cdot \underset{\sim}{u}_p \\ | & & | \end{bmatrix}$

$\Rightarrow \quad (*) = (y - X\beta)^t T \cdot T^t (y - X\beta)$

$= \| T^t y - X\beta \|^2 \Rightarrow$ OLS estimate when obs. are $T^t y$

$\Rightarrow \ldots$

$\hat{\beta} = (X^t V^{-1} X)^{-1} X^t V^{-1} y$ }

□ So: LM, GLM:

$$y_i = \underset{\sim}{x}_i^t \underset{\sim}{\beta} + \varepsilon_i \quad , \quad E(\varepsilon_i) = 0$$

$$\text{cov}(\underset{\sim}{\varepsilon}) = V = \begin{cases} \sigma^2 I \\ V(\sigma^2, \sigma_\delta^2) \end{cases}$$

↓

$$y_i = \mu_i + \varepsilon_i \quad , \quad E(\varepsilon_i) = 0$$

$$\mu_i = E(Y_i)$$

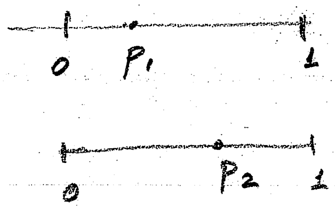
$$\mu_i = \underset{\sim}{x}_i^t \underset{\sim}{\beta}$$

□ Q: what if $y_i = \begin{cases} 1 & \text{heart disease} \\ 0 & \text{no} \end{cases}$ $\quad \underset{\sim}{x}_i = BMI_i$?

$$E(Y_i) = p_i = P(Y_i = 1)$$

Problems:

(1) $Y_i = p_i + \varepsilon_i \quad , \quad \varepsilon_i$ iid:



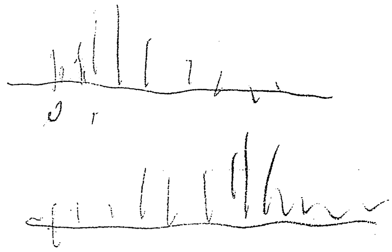
(2) $p_i = \underset{\sim}{x}_i^t \underset{\sim}{\beta}$ $\quad$? $\varepsilon_i \sim$?

where $0 < p_i < 1 \Rightarrow$ constraints on $\underset{\sim}{\beta}$

□ Q: what if y_i = # accidents/year in road segment $\quad x_i$ = speed limit ?

$$y_i = 0, 1, 2, \ldots$$

☐ $y_i \sim \text{Poisson}(\lambda_i)$



(1) $Y_i = \lambda_i + \varepsilon_i$

(2) $\lambda_i = \underset{\sim}{x}_i^t \underset{\sim}{\beta}$ $\qquad \lambda_i > 0$

☐ Q: What if y_i = life time of bulb

x_i = brand

OR

$y_i \sim N(\mu_i, \sigma^2)$

$y_i \sim \Gamma$, $E(Y_i) > 0$

☐ What to do?

- Compute MLE for each case separately

OR

- Formulate model (GLM) which includes $\forall$ cases

Y_i

- GLIM : Assume

  (1) Y_i independent, with distribution from **exponential family**.

  (2) If $\mu_i = E(Y_i)$, then

  $g(\mu_i) = \underset{\sim}{x}_i^t \underset{\sim}{\beta}$ for a **link function** g.

  (eg $g(\mu_i) = \log(\mu_i) = \underset{\sim}{x}_i^t \underset{\sim}{\beta}$)

- Then Compute MLE of β.

- **Problem** : If L = log-likelihood

  $= \log f(y_1, y_2, .., y_n)$ for independent y_i,

  $\frac{\partial L}{\partial \beta_j} = 0$ for $j = 1, .., p$ might not have closed-form solution.

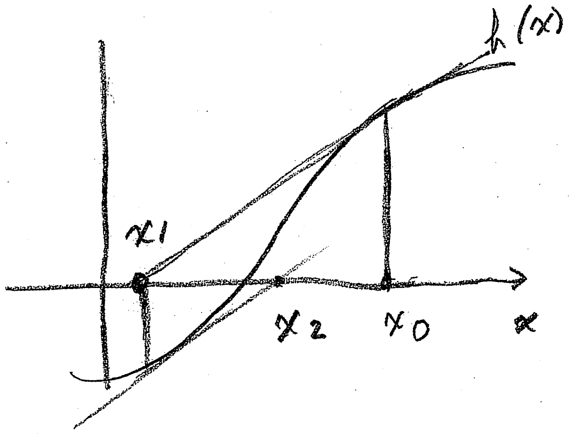
  (regression : $\nabla_{\underset{\sim}{\beta}} L_\beta = X^t(X\underset{\sim}{\beta} - \underset{\sim}{y})$

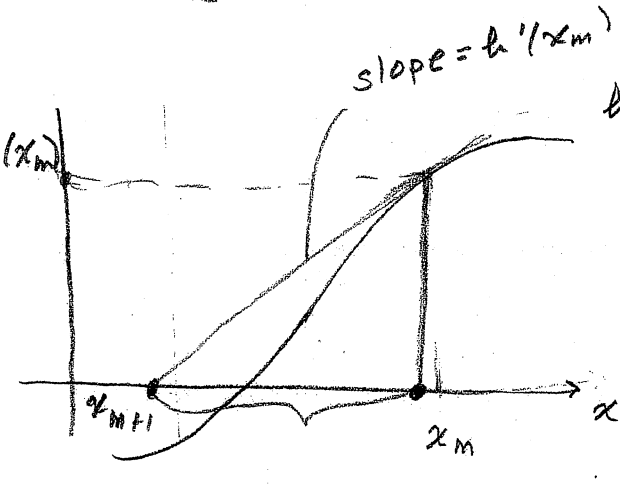
  $\Rightarrow \hat{\underset{\sim}{\beta}} = (X^tX)^{-1}X^t\underset{\sim}{y}$)

- $\Rightarrow$ Use Newton-Raphson (NR) algorithm

☐ Newton – Raphson :

To solve $h(x) = 0$, $h : \mathbb{R} \to \mathbb{R}$ when $h' \exists$





$$\frac{h(x_m)}{x_m - x_{m+1}} = h'(x_m) \Rightarrow x_{m+1} = x_m - \frac{h(x_m)}{h'(x_m)}$$

$\{\text{if } h'(x_m) \neq 0\}$

☐ To solve

$$\underset{\sim}{h}(\underset{\sim}{x}) = \begin{bmatrix} h_1(\underset{\sim}{x}) \\ \vdots \\ h_p(\underset{\sim}{x}) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \underset{\sim}{x} = (x_1, \dots, x_p)^t :$$

NR: $\underset{\sim}{x}_{m+1} = \underset{\sim}{x}_m - \left[D\underset{\sim}{h}(\underset{\sim}{x}_m) \right]^{-1} \cdot \underset{\sim}{h}(\underset{\sim}{x}_m) \Big]$ where

$$Dh_{p \times p} = \begin{cases} \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_p} \\ \vdots & & \\ \frac{\partial h_p}{\partial x_1} & \cdots & \frac{\partial h_p}{x_p} \end{bmatrix} \checkmark \\ \\ \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_p}{\partial x_1} \\ \vdots & & \\ \frac{\partial h_1}{\partial x_p} & \cdots & \frac{\partial h_p}{\partial x_p} \end{bmatrix} \end{cases}$$

($(Dh \cdot \underset{\sim}{x})$ in Taylor $h(x) = h(x_0) + (Dh)(x - x_0)$)

D Application of NR to MLE (in general)

$$h(\underset{\sim}{x}) = \underset{\sim}{\nabla} L(\underset{\sim}{\beta}) = \begin{bmatrix} \frac{\partial L}{\partial \beta_1} \\ \vdots \\ \frac{\partial L}{\partial \beta_p} \end{bmatrix}$$

$$Dh = \begin{bmatrix} \frac{\partial^2 L}{\partial \beta_1 \partial \beta_1} & \cdots & \frac{\partial^2 L}{\partial \beta_p \partial \beta_1} \\ \vdots & & \\ \frac{\partial^2 L}{\partial \beta_1 \partial \beta_p} & \cdots & \frac{\partial^2 L}{\partial \beta_p \partial \beta_p} \end{bmatrix} \equiv \text{Hessian matrix of } L$$

So

$$\underset{\sim}{\beta}_{m+1} = \underset{\sim}{\beta}_m - \left(\frac{\partial^2 L}{\partial \beta_i \partial \beta_j} \right)^{-1} (\underset{\sim}{\nabla} L)$$

evaluated at $\underset{\sim}{\beta}_m$.

E <u>Recommendation</u>: PRINT THE FILE

glm.formulas.pdf

on the website!