



**IPB University**  
— Bogor Indonesia —

Inspiring Innovation with Integrity  
in Agriculture, Ocean and Biosciences for a sustainable World

# Small Area Estimation



**Anang Kurnia**  
**Departemen Statistika, IPB University**  
**Padang, 7-8 November 2019**



## ***Outline :***

1. Pengantar: SDGs
2. *Small Area Estimation*
3. EBLUP (Area dan Unit Level)
4. Ilustrasi dan Diskusi
5. EBLUP dan Tantangan Pengembangannya
6. Penutup

# Pengantar: *Sustainable Development Goals*



- Meliputi 3 (tiga) dimensi pembangunan berkelanjutan, yaitu lingkungan, sosial, dan ekonomi

- Agenda internasional, merupakan kelanjutan dari Tujuan Pembangunan Milenium atau *Millennium Development Goals* (MDGs).
- SDGs disusun oleh Perserikatan Bangsa-Bangsa (PBB) dengan melibatkan 193 negara, *civil society*, dan berbagai pelaku ekonomi dari seluruh penjuru dunia
- SDGs ditetapkan pada 25 September 2015
- 17 tujuan global
- 169 target
- 241 indikator

# Pengantar: *Sustainable Development Goals*



## TUJUAN GLOBAL

Untuk Pembangunan Berkelanjutan





## Tantangan :

### 1. Ketersediaan Data

- Level Data : individu, rumah tangga, desa/kelurahan, kecamatan, kabupaten/kota, provinsi.
- Data : administrasi-sensus-survey.
- *Integrated Data*

### 2. Pemda – Otonomi Daerah

- Politik dan kapasitas birokrasi
- Level pengukuran : kab/Kota, kecamatan, desa/Kelurahan





# *Small Area Estimation*

- Kata kunci dalam pendugaan area kecil (*small area estimation, SAE*)
  - “*small in sample*” → ukuran contoh yang kecil, bukan “*area = wilayah*” yang kecil.
  - “*area*” tidak selalu identik dengan wilayah administratif/geografis, lebih umum merupakan *sub-domain*/himpunan bagian yang tidak beririsan dari populasi.
- *Direct vs Indirect*
- *Design-Based vs Model-Based*



- *Design-Based*

Menggunakan model untuk menyusun penduga (motivasi, *model assisted*), tetapi bias dan ragam penduga dievaluasi berdasarkan sebaran acak contoh.

- *Model-Based*

Pendugaan dan inferensi berdasarkan pemodelan, dengan asumsi anggota contoh terpilih yang sudah diketahui, melalui pendekatan *frequenties* atau Bayes (atau gabungan : *empirical Bayes*).

- Peubah penyerta (*auxiliary variable*) → level area, level unit  
“upaya meningkatkan efektifitas ukuran contoh”



# *SAE: Design Based Method*

## Review komprehensif - Lehtonen dan Veijanen (2009)

- Misal parameter yang menjadi perhatian :  $\bar{y}_i = N_i^{-1} \sum_j y_{ij}$
- Tidak terdapat peubah penyerta  $\rightarrow$  *direct design unbiased est.*

$$\text{SRSWOR} \Rightarrow \hat{y}_i^D = n_i^{-1} \sum_j y_{ij}$$

$$V\left(\hat{y}_i^D \mid n_i\right) = n_i^{-1} S_i^2 \left(1 - N_i^{-1} n_i\right)$$

Peluang penarikan contoh berbeda :

$$\hat{y}_i^{HT} = \left(\sum_j^{n_i} w_{ij}\right)^{-1} \sum_j^{n_i} w_{ij} y_{ij}$$

$$w_{ij} = 1 / \pi_{ij}$$

- Terdapat peubah penyerta  $\rightarrow$  *indirect est. = synthetic est.*

$$\text{SRSWOR} \Rightarrow \hat{y}_i^{\text{Syn}} = \bar{x}_i^T \hat{\beta} = N_i^{-1} \sum_{j=1}^{N_i} x_{ij}^T \hat{\beta}$$

$$\hat{\beta} = \left[ \sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij} x_{ij}^T \right]^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij} y_{ij}$$

Secara “umum”  $\hat{\beta}$  merupakan penduga tak bias dan konsisten bagi  $\beta$ .

- Peluang penarikan contoh tidak sama  $\rightarrow$  *survey regression est.*

$$\hat{\beta}_w = \left[ \sum_{i=1}^m \sum_{j=1}^{n_i} w_{ij} x_{ij} x_{ij}^T \right]^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} w_{ij} x_{ij} y_{ij}$$

$$w_{ij} = 1 / \pi_{ij}$$

$$\begin{aligned} \hat{y}_i^{SR} &= \bar{x}_i^T \hat{\beta}_w + N_i^{-1} \sum_{j=1}^{n_i} w_{ij} \left( y_{ij} - x_{ij}^T \hat{\beta}_w \right) \\ &= \hat{y}_i^{HT} + \left( \bar{x}_i - \bar{x}_i^{HT} \right) \hat{\beta}_w \end{aligned}$$

- Suatu kompromi antara kemungkinan bias yang besar pada *synthetic est* dan ragam yang besar pada *survey regression est*. → *composite est*.

$$\hat{y}_i^C = \delta_i \hat{y}_i^{SR} + (1 - \delta_i) \hat{y}_i^{Syn} ; 0 < \delta_i < 1$$

Review lebih lanjut tentang *composite est* bisa dilihat di Rao dan Molina (2015).

## Model Kalibrasi → *generalized regression (GREG)*

- Bentuk umum :  $\hat{y}_i^{Cal} = N_i^{-1} \sum_j^{n_i} w_{ij}^c y_{ij} ; t_x = \sum_j^{n_i} w_{ij}^c x_{ij}$ 
  - Estevao dan Särndal (2004, 2006) → *instrumental variable*.
  - Lehtonen, Särndal dan Veijanen (2003, 2005) → GLM & GLMM untuk *generalized GREG*.
  - Chandra dan Chambers (2009) → EBLUP untuk *model-based direct estimator* (MBDE).

Review lebih jauh tentang model kalibrasi bisa dilihat di Kott (2009).



- Kelebihan
  - Sedikit tergantung pada asumsi model (pemodelan) → model assisted.
  - Penduga bersifat (aproksimasi) tidak berbias dan konsisten
- Kekurangan
  - Contoh kecil, ragam besar.
  - Survey regression, aproksimasi tidak berbias namun (mungkin) perlu peubah penyerta yang banyak.
  - Synthetic, ragam kecil namun berbias.
  - Composite, penentuan delta optimum yang tidak mudah.



- Pendugaan area kecil berbasis model
  - Model level area.
  - Model level unit.
  
- Model dasar : Empirical Best Linear Unbiased Prediction (EBLUP)
  - Linear Mixed Model.
  - Generalized Linear Mixed Model.



# *SAE: Model Based Method*

- Produktivitas Padi di Kab. Seruyan tahun 2016

Kode	Kecamatan	Jumlah contoh plot ubinan	Luas panen total (ha)	Produktivitas padi (ku/ha), BPS
010	Seruyan Hilir	14	1,090.4	-
011	Seruyan Hilir Timur	16	1,295.3	-
020	Danau Sembuluh	3	1,086.2	-
021	Seruyan Raya	-	100.9	-
030	Hanau	-	643.4	-
031	Danau Seluluk	-	499.8	-
040	Seruyan Tengah	20	1,511.90	-
041	Batu Ampar	11	455.9	-
050	Seruyan Hulu	10	2,209.0	-
051	Suling Tambun	15	1,232.9	-
<b>Total</b>		<b>89</b>	<b>10,125.7</b>	<b>26.68</b>

- Produktivitas Padi di Kab. Seruyan tahun 2016

Kode	Kecamatan	Jumlah contoh plot ubinan	Luas panen total (ha)	Produktivitas padi (ku/ha), BPS	Produktivitas padi Geoadditive SAE (ku/ha)
010	Seruyan Hilir	14	1,090.4	-	31.32
011	Seruyan Hilir Timur	16	1,295.3	-	34.58
020	Danau Sembuluh	3	1,086.2	-	32.56
021	Seruyan Raya	-	100.9	-	27.19
030	Hanau	-	643.4	-	30.92
031	Danau Seluluk	-	499.8	-	32.64
040	Seruyan Tengah	20	1,511.90	-	26.72
041	Batu Ampar	11	455.9	-	22.69
050	Seruyan Hulu	10	2,209.0	-	19.93
051	Suling Tambun	15	1,232.9	-	20.30
<b>Total</b>		<b>89</b>	<b>10,125.7</b>	<b>26.68</b>	<b>26.97</b>



- Linear Mixed Model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$$

$$E[\boldsymbol{\gamma}] = \mathbf{0}, \quad \text{Var}[\boldsymbol{\gamma}] = \mathbf{G}$$

$$E[\boldsymbol{\varepsilon}] = \mathbf{0}, \quad \text{Var}[\boldsymbol{\varepsilon}] = \mathbf{R}$$

$$\text{cov}(\boldsymbol{\varepsilon}, \boldsymbol{\gamma}) = \mathbf{0}$$

$$E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$$

$$\text{Var}[\mathbf{y}] = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}$$

- Linear Mixed Model

$$\boldsymbol{\gamma} \sim N(\mathbf{0}, \mathbf{G}), \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{R})$$

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$$

$$\begin{pmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{y} \end{pmatrix}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}$$

$$\hat{\boldsymbol{\gamma}} = \mathbf{G} \mathbf{Z}^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}})$$



- Model Level Area

- Fay dan Herriot (1979).
- Peubah penyerta hanya tersedia pada level area.

$$\text{Model : } \left. \begin{array}{l} \tilde{y}_i = \bar{y}_i + e_i \\ \bar{y}_i = x_i^T \beta + v_i \end{array} \right\} \tilde{y}_i = x_i^T \beta + v_i + e_i$$

$\tilde{y}_i$  : penduga langsung dari  $\bar{y}_i$

$e_i$  : sampling error  $\sim (0, \sigma_{ei}^2)$

$v_i$  : pengaruh acak area  $\sim (0, \sigma_v^2)$

- Model Level Area
  - Fay dan Herriot (1979).
  - Peubah penyerta hanya tersedia pada level area.

Penduga :  $\hat{y}_i^{BLUP} = \gamma_i \tilde{y}_i + (1 - \gamma_i) x_i^T \hat{\beta}$

$$\begin{aligned}\hat{y}_i^{EBLUP} &= \hat{\gamma}_i \tilde{y}_i + (1 - \hat{\gamma}_i) x_i^T \hat{\beta} \\ &= x_i^T \hat{\beta} + \hat{\gamma}_i (\tilde{y}_i - x_i^T \hat{\beta}) \\ &= x_i^T \hat{\beta} + \hat{u}_i\end{aligned}$$

$$\hat{\gamma}_i = \hat{\sigma}_v^2 / (\hat{\sigma}_v^2 + \sigma_{ei}^2)$$

- Model Level Unit = Nested error unit level model
  - Battese, Harter and Fuller (1988).
  - Peubah penyerta tersedia pada level unit.

$$\text{Model : } y_{ij} = x_{ij}^T \beta + v_i + e_{ij}; v_i \sim N(0, \sigma_v^2), e_{ij} \sim N(0, \sigma_e^2)$$

Penduga EBLUP bagi  $\bar{y}_i = N_i^{-1} \left( \sum_{j \in s} y_{ij} + \sum_{j \in r} y_{ij} \right)$  adalah :

$$\begin{aligned} \hat{y}_i^{EBLUP} &= N_i^{-1} \left( \sum_{j \in s} y_{ij} + \sum_{j \in r} \hat{y}_{ij}^{EBLUP} \right) \\ &= N_i^{-1} \left\{ \sum_{j \in s} y_{ij} + \sum_{j \in r} \left[ x_{ij}^T \hat{\beta} + \hat{\gamma}_i \left( \bar{y}_{is} - \bar{x}_{is}^T \hat{\beta} \right) \right] \right\} \end{aligned}$$

$$\hat{\gamma}_i = \hat{\sigma}_v^2 \left( \hat{\sigma}_v^2 + \hat{\sigma}_e^2 / n_i \right)^{-1}$$

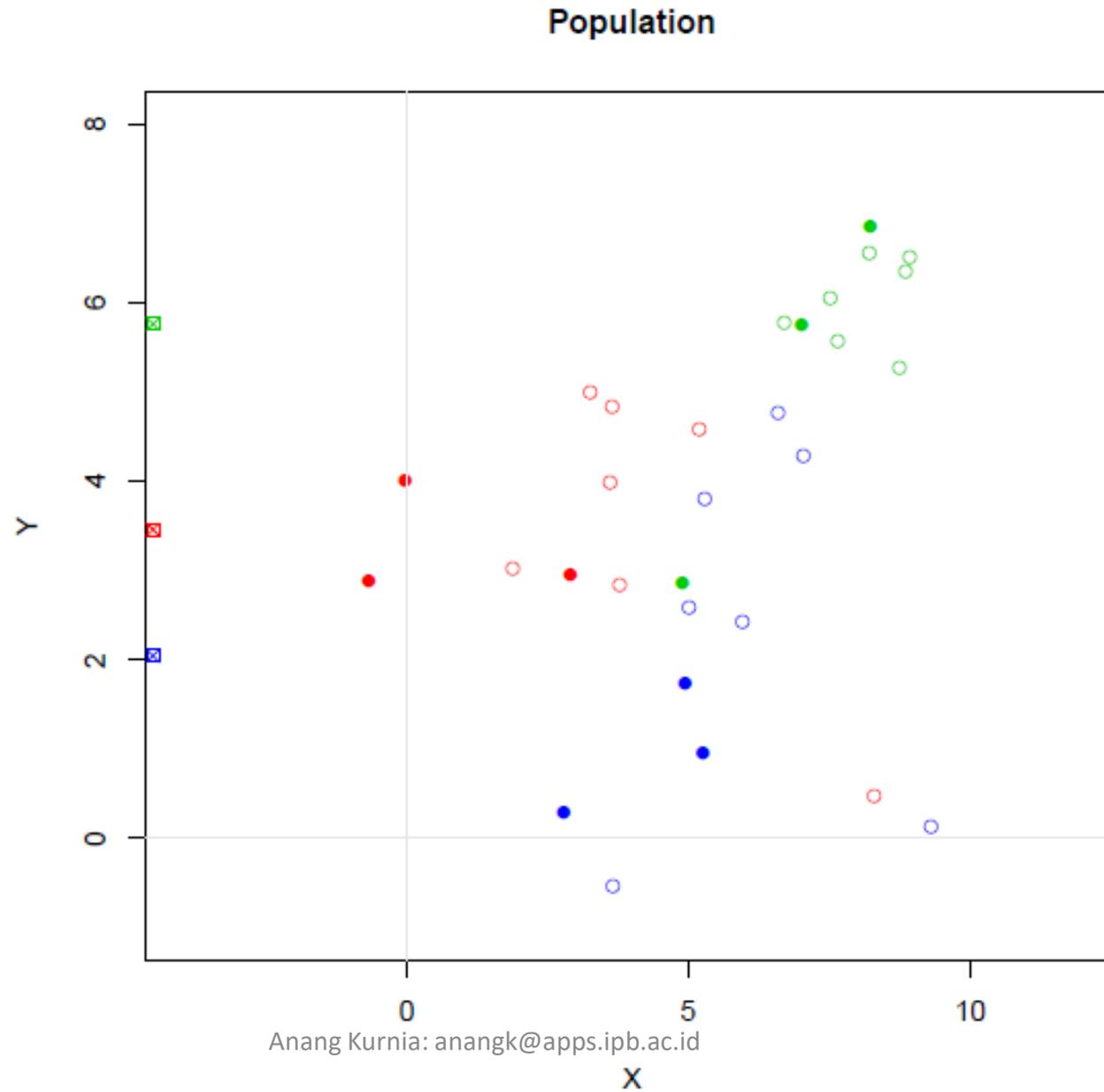
- Model Level Unit = Nested error unit level model
  - Battese, Harter and Fuller (1988).
  - Peubah penyerta tersedia pada level unit.

$$\hat{y}_i^{EBLUP} = f_i \bar{y}_{is} + (1 - f_i) \hat{y}_{ir}$$

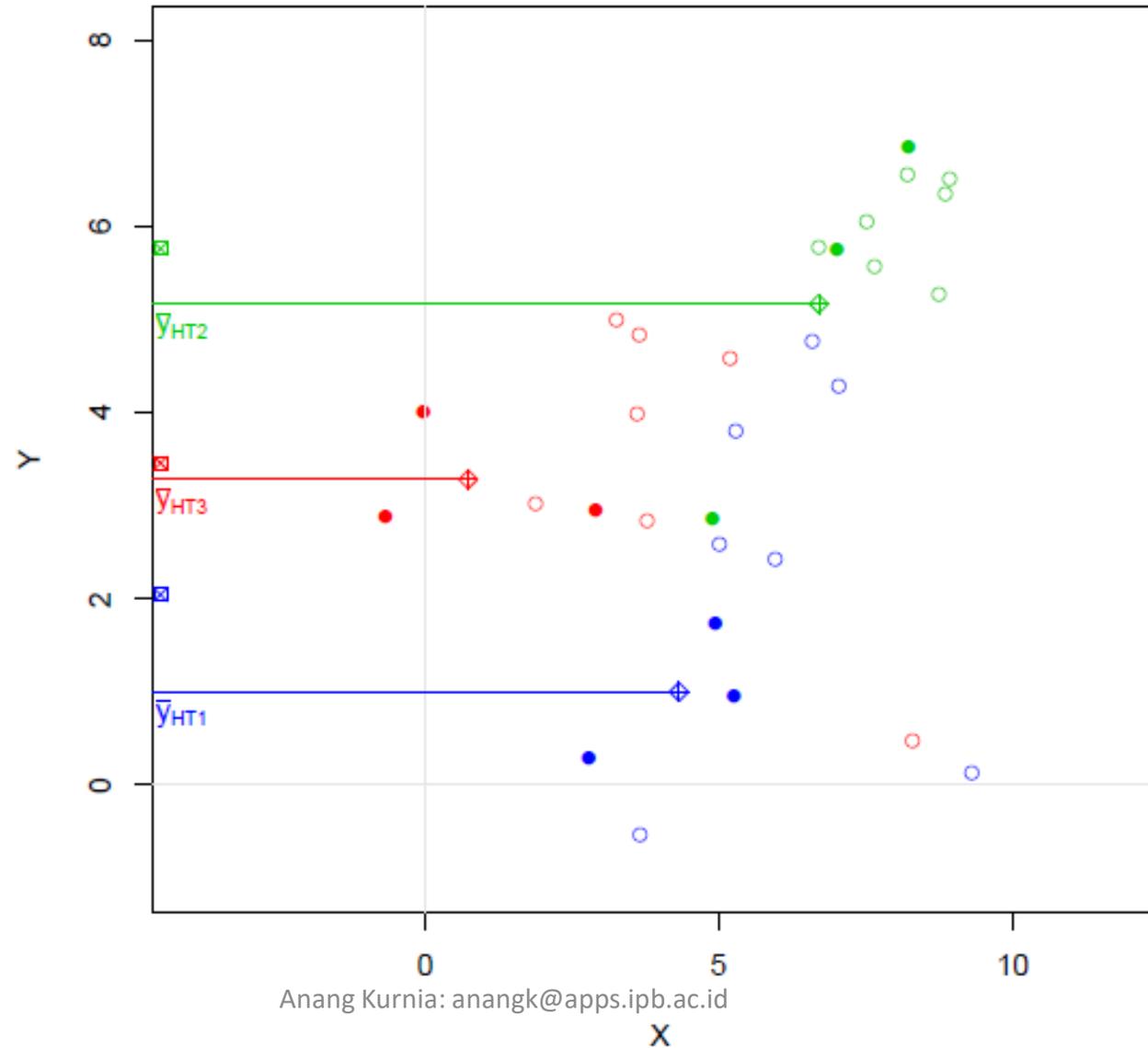
jika  $n_i \rightarrow 0, f_i \rightarrow 0$

$$\hat{y}_i^{EBLUP} = \bar{x}_i^T \hat{\beta} + \hat{\gamma}_i \left( \bar{y}_{is} - \bar{x}_{is}^T \hat{\beta} \right)$$



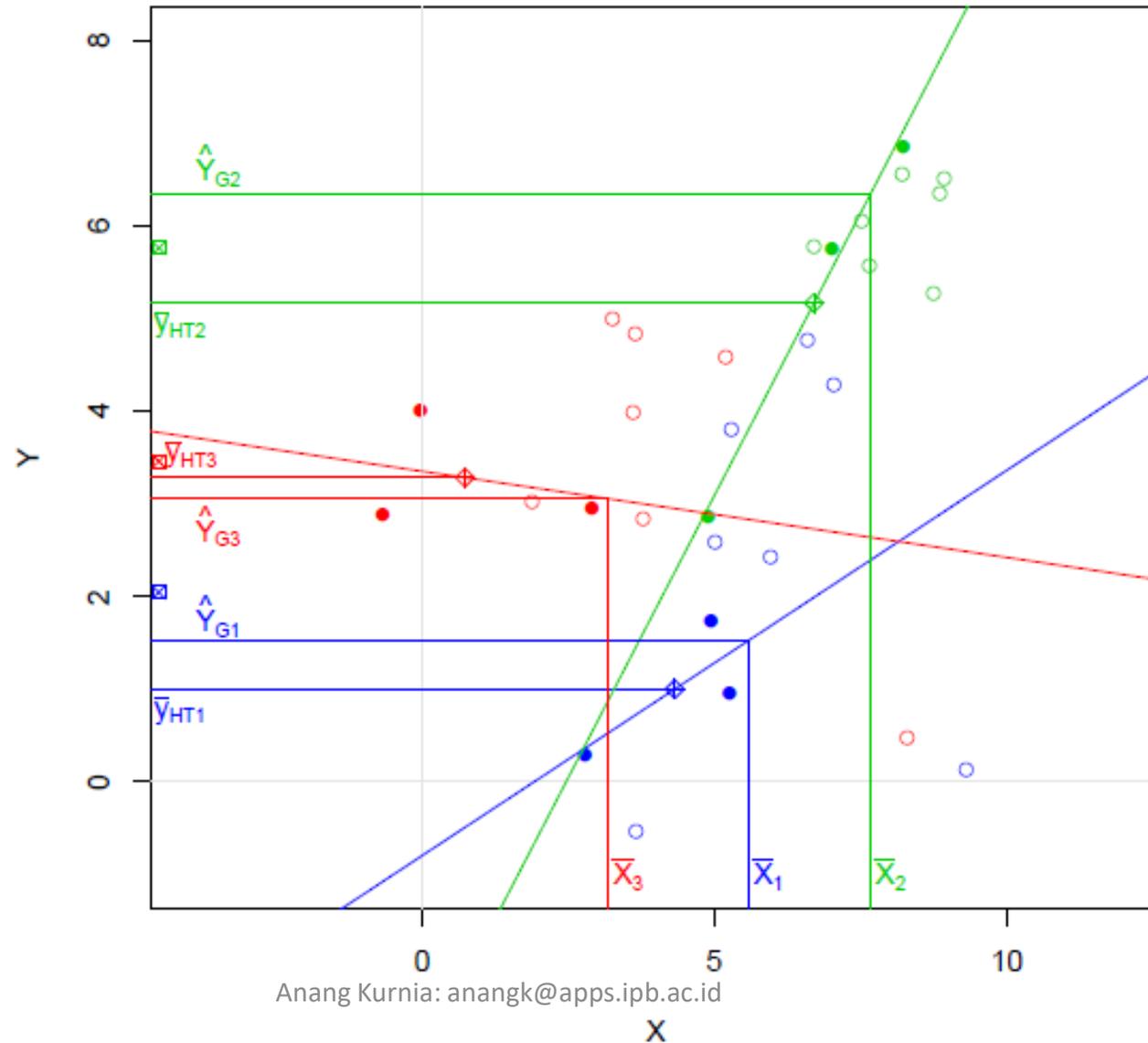


## Horvitz-Thompson



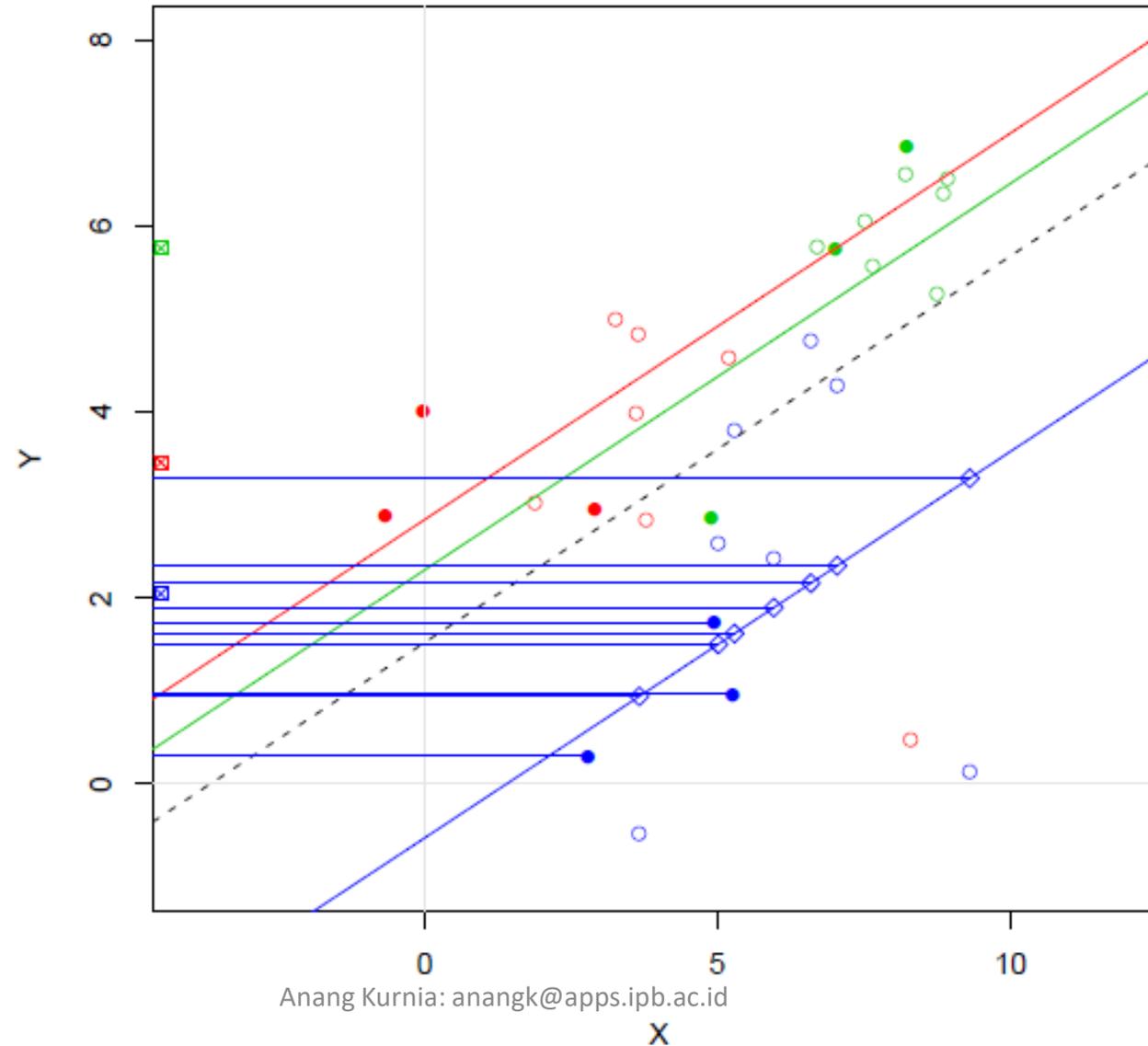
$$\hat{Y}_{Gi} = \alpha_i + \bar{X}_i \beta_i$$

GREG



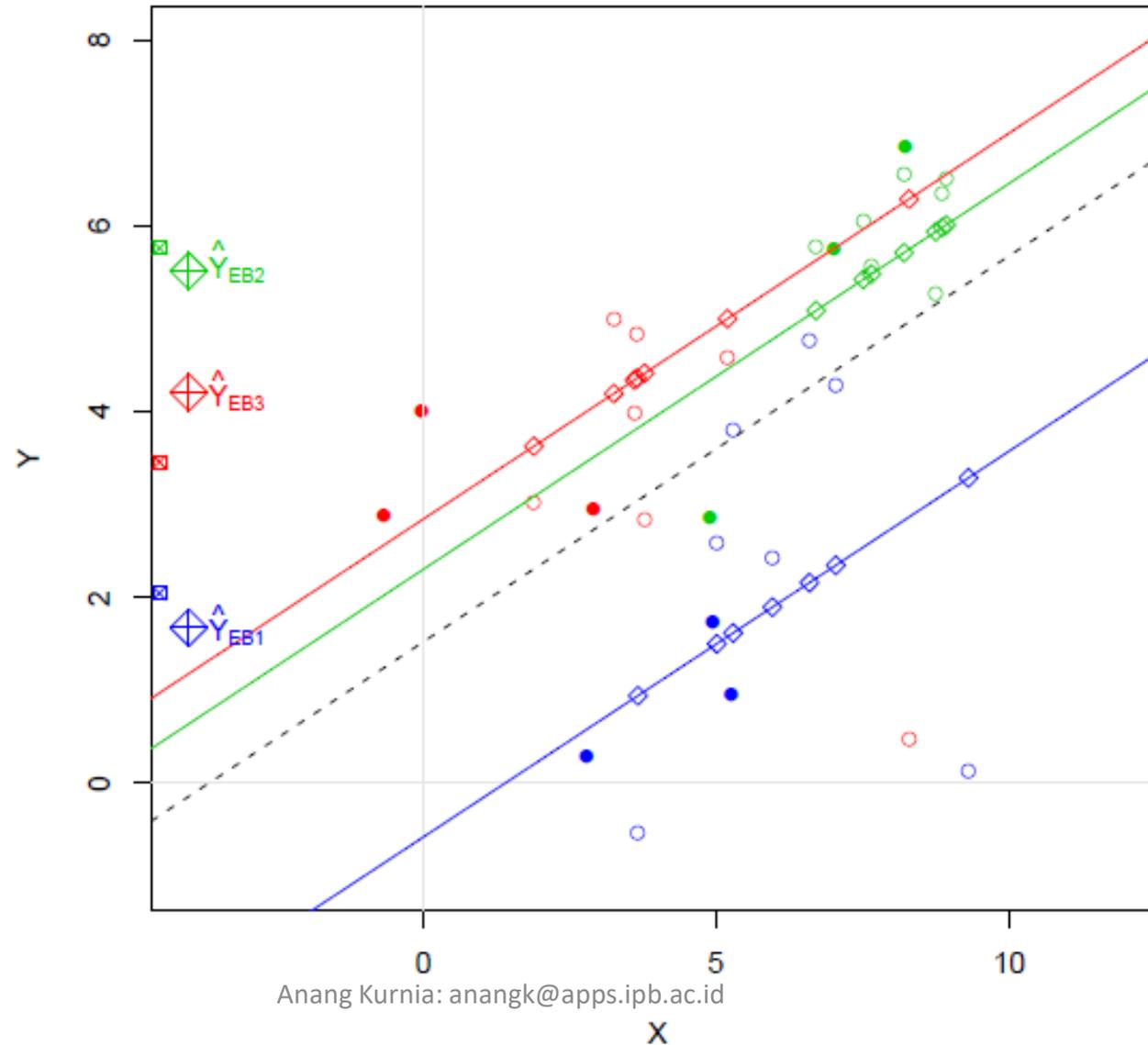
$$\hat{Y}_i = N_i^{-1} \left\{ \sum_{j \in s_i} y_{ij} + \sum_{j \in r_i} \hat{y}_{ij} \right\}$$

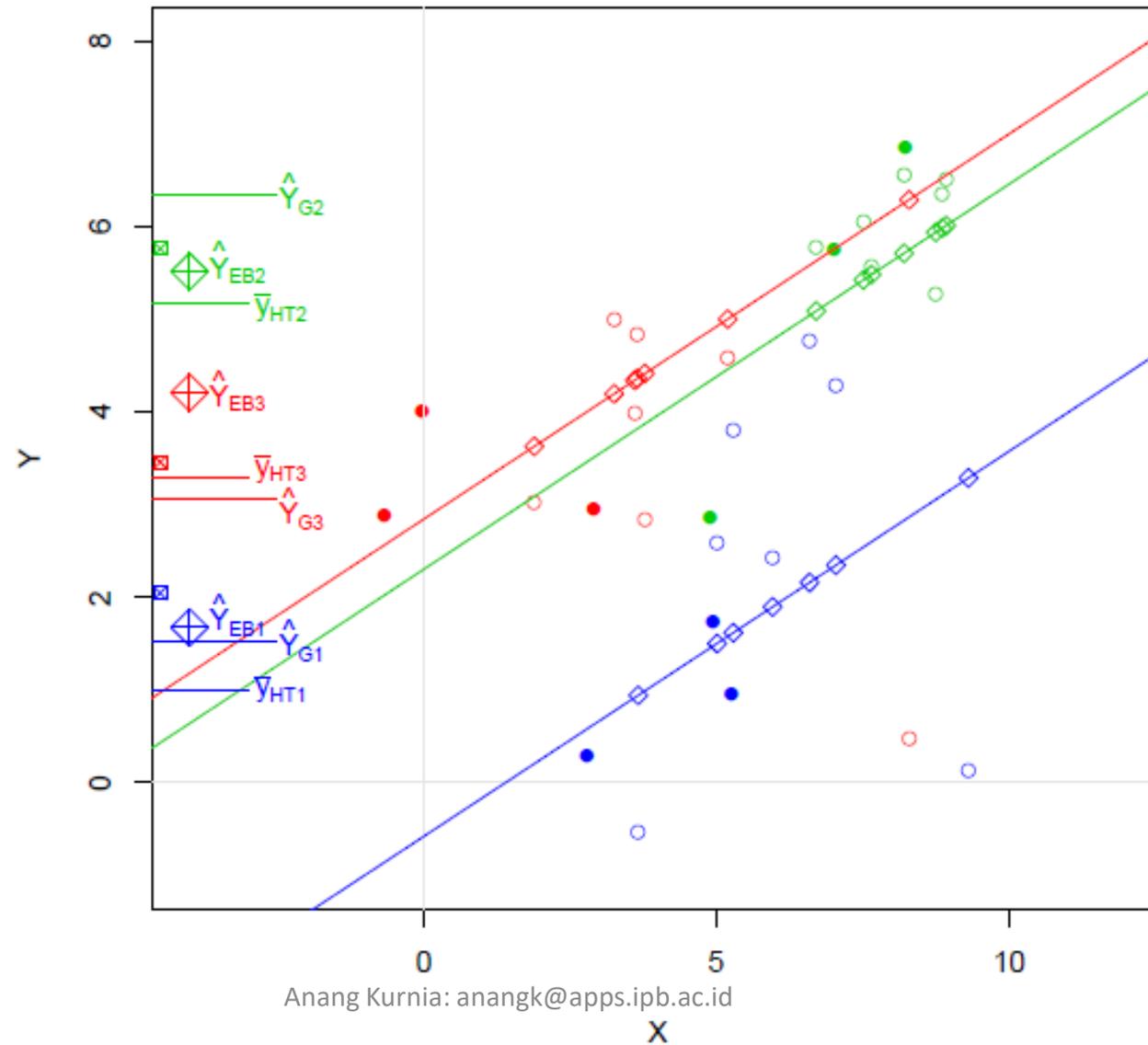
EBLUP



$$\hat{Y}_i = N_i^{-1} \left\{ \sum_{j \in s_i} y_{ij} + \sum_{j \in r_i} \hat{y}_{ij} \right\}$$

EBLUP







- IPM mengukur pencapaian pembangunan rata-rata suatu regional berdasarkan 3 aspek utama: kesehatan, pengetahuan, dan standar hidup
- Dimensi/Indikator pembentuk IPM:
  - Kesehatan: **Angka Harapan Hidup saat Lahir (AHH)**
  - Pengetahuan: **Rata-Rata Lama Sekolah (RLS)** dan **Harapan Lama Sekolah (HLS)**
  - Standar hidup: **Pengeluaran per Kapita**

# Lab: Indeks Pembangunan Manusia

$$\Rightarrow I_{kesehatan} = \frac{AHH - AHH_{min}}{AHH_{maks} - AHH_{min}}$$

$$\Rightarrow I_{pengetahuan} = \frac{I_{RLS} + I_{HLS}}{2}$$

$$I_{RLS} = \frac{RLS - RLS_{min}}{RLS_{maks} - RLS_{min}}$$

$$I_{HLS} = \frac{HLS - HLS_{min}}{HLS_{maks} - HLS_{min}}$$

$$\Rightarrow I_{pengeluaran} = \frac{\ln(\text{pengeluaran}) - \ln(\text{pengeluaran}_{min})}{\ln(\text{pengeluaran}_{maks}) - \ln(\text{pengeluaran}_{min})}$$

$$IPM = \sqrt[3]{I_{kesehatan} \times I_{pengetahuan} \times I_{pengeluaran}}$$



# *EBLUP: Tantangan Pemodelan*

- Model transformasi logaritma – Empirical Best Predictor

$$l_{ij} = \log(y_{ij}) = x_{ij}^T \beta + v_i + e_{ij}$$

- Lognormal : Kurnia dan Chambers (2011), Berg (2011), Chandra dan Berg (2014).

$$\bar{y}_i = \frac{1}{N_i} \left[ \sum_s y_{ij} + \sum_r y_{ij} \right]$$



$$\hat{y}_i = \frac{1}{N_i} \left[ \sum_s y_{ij} + \sum_r \hat{y}_{ij} \right]$$



$$\hat{y}_i^{Karl} = N_i^{-1} \left[ \sum_{s_i} y_{ij} + \sum_{r_i} \hat{y}_{ij}^{Karl} \right]$$

$$\hat{y}_{ij}^{Karl} = \left( c_{ij}^{Karl} \right)^{-1} \exp \left\{ x_{ij}^T \hat{\beta} + \frac{1}{2} \left( \hat{\sigma}_v^2 + \hat{\sigma}_e^2 \right) \right\}$$

$$c_{ij}^{Karl} = 1 + \frac{1}{2} x_{ij}^T V \left( \hat{\beta} \right) x_{ij} + \frac{1}{8} \nabla \left( \hat{\sigma}_v^2 + \hat{\sigma}_e^2 \right)$$

- Spasial SAE

$$y_{ij} = x_{ij}^T \beta + v_i + e_{ij} \Rightarrow Y = X\beta + Zv + \varepsilon$$

$$v = \rho Wv + \nu \Rightarrow v = (I_m - \rho W)^{-1} \nu$$

$$Y = X\beta + Z(I_m - \rho W)^{-1} \nu + \varepsilon$$

- Spatial EBLUP (SAR) – Pratesi dan Salvati (2008)



$$\hat{y}_i^{SEBLUP} = N_i^{-1} \left[ \sum_{j \in s_i} y_{ij} + \sum_{j \in r_i} \hat{y}_{ij}^{SEBLUP} \right]$$

$$\hat{y}_{ij}^{SEBLUP} = x_{ij}^T \hat{\beta} + z_{ij} \hat{v}_i$$

$$\hat{v}_i = b_i^T \left( Z^T (\hat{\sigma}_e^2 I_n)^{-1} Z + \hat{D}^{-1} \right)^{-1} Z^T (\hat{\sigma}_e^2 I_n)^{-1} (Y - X \hat{\beta}) b_i$$

$$\hat{D} = \hat{\sigma}_v^2 \left[ (I_m - \hat{\rho} W)(I_m - \hat{\rho} W^T) \right]^{-1}$$

- Spasial SAE: EBP

$$y_{ij}^* = x_{ij}^T \beta + v_i + e_{ij} \Rightarrow Y^* = X \beta + Zv + \varepsilon$$

$$v = \rho Wv + v \Rightarrow v = (I_m - \rho W)^{-1} v$$

$$y_{ij}^* = \log(y_{ij})$$

- Spatial EBP (SAR) - Handayani, Kurnia, Folmer, dan Notodiputro (2018)



$$\hat{y}_i^{SEBP} = N_i^{-1} \left[ \sum_{s_i} y_{ij} + \sum_{r_i} \hat{y}_{ij}^{SEBP} \right]$$

$$\hat{y}_{ij}^{SEBP} = (c_{ij}^{SEBP})^{-1} e^{x_{ij}^T \hat{\beta} + \frac{1}{2}(z_{ij}^2 \hat{\sigma}_v^2 + \hat{\sigma}_e^2)}$$

$$\hat{\sigma}_v^2 = b_i^T \hat{D} b_i$$

$$\hat{D} = \hat{\sigma}_v^2 \left[ (I_m - \hat{\rho}W)(I_m - \hat{\rho}W^T) \right]^{-1}$$

- Spasial SAE

- GWR – Brundsdon et al. (1998); Fotheringham et al. (2002)
- GWLMM – Chandra et. al (2012)

**Idea : Lokal model**

$$y_{ij} = x_{ij}^T \beta(u_k) + v_i + e_{ij} w^{-1/2}(u_k, u_{ij})$$

$$\hat{y}_i^{GWEBLUP} = N_i^{-1} \left[ \sum_{s_i} y_{ij} + \sum_{r_i} \hat{y}_{ij}^{GWEBLUP} \right]$$

$$\hat{y}_{ij}^{GWEBLUP} = x_{ij}^T \hat{\beta}(u_{ij}) + \hat{v}_i(u_{ij})$$

- Robust SAE
  - Robust EBLUP (Sinha dan Rao, 2009)

$$\hat{y}_i^{REBLUP} = N_i^{-1} \left[ \sum_{s_i} y_{ij} + \sum_{r_i} \hat{y}_{ij}^{REBLUP} \right]$$

$$\hat{y}_{ij}^{REBLUP} = x_{ij}^T \hat{\beta}^R + \hat{v}_i^R$$

.....

ML dan Huber's influence function :

$$X^T V^{-1} (y - X \beta) = 0 \Rightarrow X^T V^{-1} U^{1/2} \psi(r) = 0$$

$$r = U^{1/2} (y - X \beta), \text{ dst ...}$$

- Robust SAE
  - Spatial Robust EBLUP (Schmid, 2012)

$$Y = X\beta + Z(I_m - \rho W)^{-1}v + \varepsilon$$

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

$$\hat{y}_i^{SREBLUP} = N_i^{-1} \left[ \sum_{j \in s_i} y_{ij} + \sum_{j \in r_i} \hat{y}_{ij}^{SREBLUP} \right]$$

$$\hat{y}_{ij}^{SREBLUP} = x_{ij}^T \hat{\beta}^R + z_{ij} \hat{v}_i^R$$

$\hat{\beta}^R$  dan  $\hat{v}_i^R$  diperoleh berdasarkan konsep robust

- Robust SAE

- M-quantile (Chambers dan Tzavidis, 2006; Tzavidis et al., 2010; Salvati et al., 2011)

Mquantile regression :  $m_q(\mathbf{X}) = \mathbf{X}\hat{\beta}_q^{MQ} \rightarrow y_{ij} = x_{ij}^T \hat{\beta}_{q_{ij}}^{MQ}$

sehingga 
$$\hat{y}_i^{MQ} = N_i^{-1} \left[ \sum_{j \in s_i} y_{ij} + (N_i - n_i) \bar{x}_{ir}^T \hat{\beta}_{\bar{q}_i}^{MQ} \right]$$

$$\bar{q}_i = n_i^{-1} \sum_{j \in s_i} q_{ij}$$

- Robust SAE

- Winsorization Models (Kurnia et.al, 2015)

1.  $\hat{y}_{ij} = x_{ij}^T \hat{\beta} + \hat{v}_i$

2.  $\hat{y}_{ij}^* = x_{ij}^T \hat{\beta} + \hat{v}_i^*$

$$\hat{v}_i^* = \hat{v}_i + (K_v - \hat{v}_i)I_{(\hat{v}_i \geq K_v)} + (L_v - \hat{v}_i)I_{(\hat{v}_i \leq L_v)}$$

$$(K_v = c_v \hat{\sigma}_v, L_v = -c_v \hat{\sigma}_v)$$

3.  $y_{ij}^w = \hat{y}_{ij}^* + e_{ij}^*$

$$e_{ij}^* = \hat{e}_{ij} + (K_e - \hat{e}_{ij})I_{(\hat{e}_{ij} \geq K_e)} + (L_e - \hat{e}_{ij})I_{(\hat{e}_{ij} \leq L_e)} ;$$

$$(K_e = c_e \hat{\sigma}_e, L_e = -c_e \hat{\sigma}_e)$$



$$y_{ij}^w = x_{ij}^T \beta^w + v_i^w + e_{ij}^w$$

$$\hat{y}_i^w = N_i^{-1} \left[ \sum_{j \in s_i} y_{ij} + \sum_{j \in r_i} \hat{y}_{ij}^w \right]$$

- FGT Model → Foster et al (1984)

$$P_{ai} = \frac{1}{N_i} \sum_{j=1}^{N_i} P_{aij} ; i = 1, 2 \dots m ; \quad P_{aij} = \left( \frac{z - R_{ij}}{z} \right)^\alpha I(R_{ij} < z) ; j = 1, 2 \dots N_i ; \alpha = 0, 1, 2$$

- EBP (Molina and Rao, 2010)

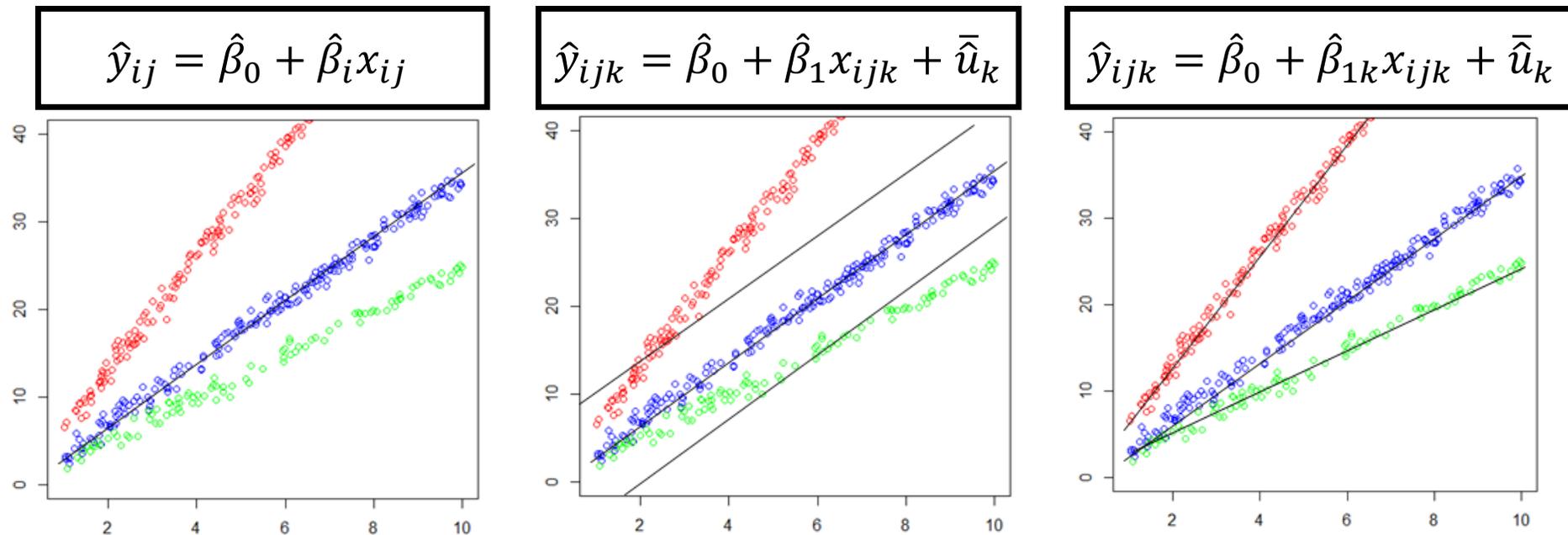
$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + v_i + e_{ij} ; j = 1, 2 \dots N_i ; i = 1, 2 \dots m$$

- Spatial EBP (Handayani, Folmer, Notodiputro, Kurnia, in progress)

$$Y = X\boldsymbol{\beta} + Z(I - \rho W)^{-1} v + \varepsilon$$

- **Non-sample problem**

- Informasi kelompok pada EBLUP (Anisa, Kurnia, Indahwati, 2015)
- Informasi kelompok pada FGT Model (Vinny, Sadik, Kurnia, 2017)



- **Panel Data SAE → Rotating Survey**

- Rao-Yu : Area level
- Muchlisoh, Kurnia, Notodiputro, Mangku (2017) : Unit level
- Kurnia, Surida, and Muchlisoh (2019) *submitted to JAPS* : pseudo panel
- Tanur, Kurnia, Notodiputro, Soleh (2019) : panel, lognormal, measurement error

$$\mathbf{y}_{itj}^* = \mathbf{x}_{ij}^T \boldsymbol{\beta}_t + \mathbf{w}_i \alpha_t + \mathbf{v}_i + \mathbf{u}_{it} + \mathbf{e}_{itj}$$

$$E(\mathbf{y}_{itj}^* | \mathbf{v}_i, \mathbf{u}_{it}) = \mathbf{x}'_{itj} \boldsymbol{\beta}_t + \mu_w \alpha_t + \mathbf{v}_i + \mathbf{u}_{it}$$

$$\hat{\mathbf{y}}_{itj} = E(\mathbf{y}_{itj} | \mathbf{v}_i, \mathbf{u}_{it}) = \exp[E(\mathbf{y}_{itj}^* | \mathbf{v}_i, \mathbf{u}_{it}) + 0.5 \text{Var}(\mathbf{y}_{itj}^* | \mathbf{v}_i, \mathbf{u}_{it})]$$

$$\begin{aligned} \theta_{it} &= \frac{1}{N_{it}} [\sum_{j \in S} y_{itj} + \sum_{j \in R} y_{itj}] \\ &= f_{it} (\bar{y}_{it}^S) + (1 - f_{it}) (\bar{y}_{it}^R) \end{aligned}$$

- **Multivariate & SEMM SAE (area level)**
  - Ubaidillah, Notodiputro, Kurnia, Mangku (2017; 2019)

$$\mathbf{Y}(-\mathbf{A}) = \mathbf{XB} + \mathbf{ZU} + \mathbf{E}$$

$$\mathbf{Y} = \mathbf{XB}(-\mathbf{A})^{-1} + \mathbf{ZU}(-\mathbf{A})^{-1} + \mathbf{E}(-\mathbf{A})^{-1}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\Pi} + \mathbf{Z}\mathbf{W} + \boldsymbol{\Xi}$$

$$\mathbf{y}^* = \mathbf{X}_A^* \boldsymbol{\beta}_A^* + \mathbf{Z}^* \mathbf{u}^* + \mathbf{e}^*$$

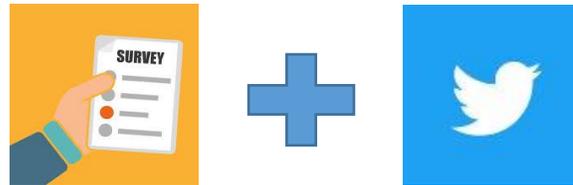
 Stacking method

Using 3SLS Method and REML :

$$\hat{\boldsymbol{\mu}} = \mathbf{X}_A^* \hat{\boldsymbol{\beta}}_{A(3SLS)}^* + \mathbf{Z}^* \left( \mathbf{G}(\hat{\boldsymbol{\delta}}) \otimes \mathbf{I}_n \right) \mathbf{Z}^{*T} \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\delta}}) \left( \mathbf{y}^* - \mathbf{X}_A^* \hat{\boldsymbol{\beta}}_{A(3SLS)}^* \right)$$

- **Big Data: Data Sosial Media**

- Muhyi, Kurnia, Sartono (2019) → Eksplorasi data twitter + measurement error



Mengkombinasikan **data survey** dengan  
**data yang diperoleh dari Twitter**

BIG  
DATA



Small Area  
Estimation

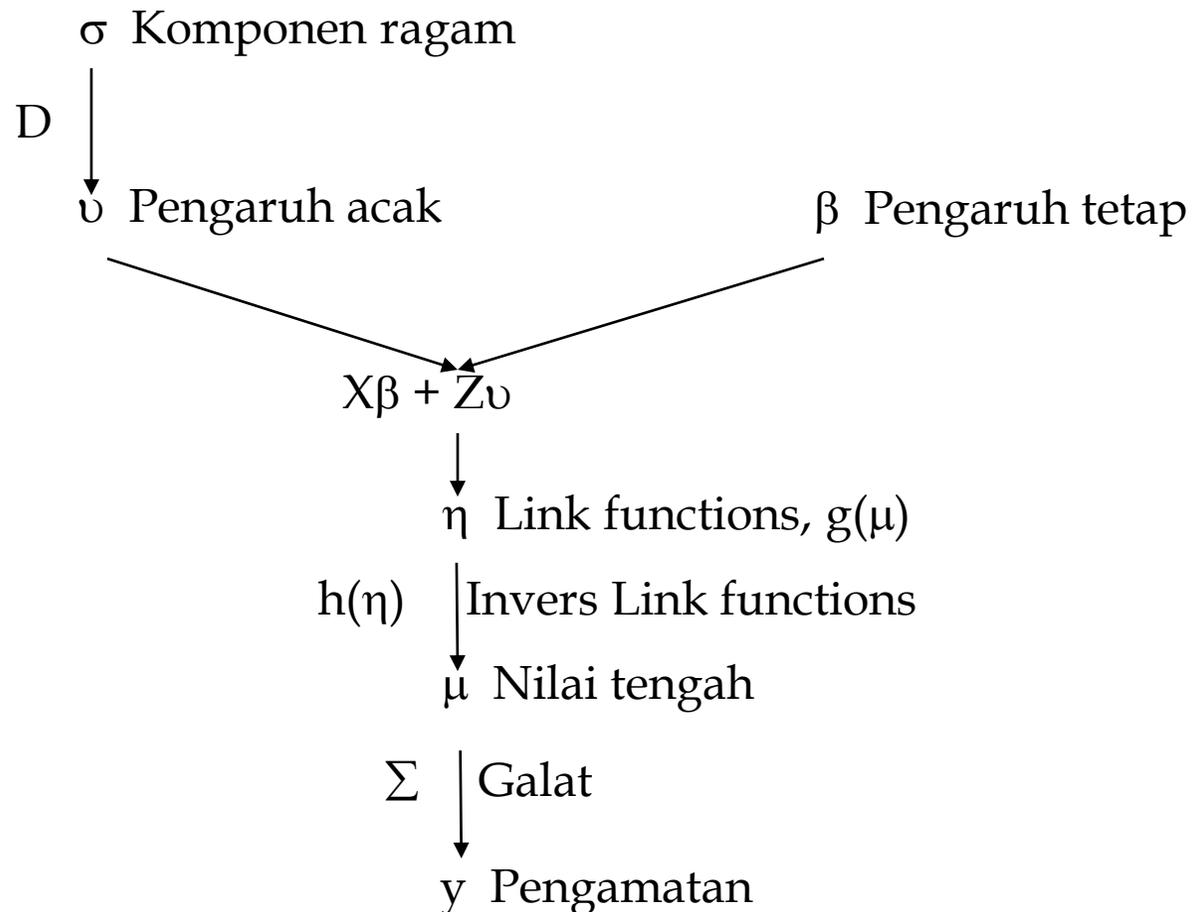
$$\log\left(\frac{p_i}{1-p_i}\right) = \alpha + \mathbf{X}_i^T \boldsymbol{\beta} + v_i, i = 1, 2, \dots, m$$

$$\mathbf{x}_i = \mathbf{X}_i + \boldsymbol{\eta}_i$$



- **Lain-lain (perhatikan):**
  - Binomial
  - Poisson
  - Zero/One Inflated
  - Tidak memenuhi asumsi sebaran data
  - Integrated Data

- **Lain-lain (perhatikan): Generalized Linear Mixed Model**



- Lain-lain (perhatikan): **Generalized Linear Mixed Model**

$$E[\mathbf{y}|\mathbf{u}] = g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) = g^{-1}(\boldsymbol{\eta})$$

$$E[\mathbf{u}] = \mathbf{0}, \quad \text{Var}[\mathbf{u}] = \mathbf{G}$$

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$$

$$\mathbf{Y} = \boldsymbol{\eta} + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$$

$$\mathbf{y}|\mathbf{u} \sim (g^{-1}(\boldsymbol{\eta}), \mathbf{R})$$

$$V(\mathbf{y}) = E[V(\mathbf{y}|\mathbf{u})] + V[E(\mathbf{y}|\mathbf{u})] = \mathbf{A}^{1/2}\mathbf{R}\mathbf{A}^{1/2} + \mathbf{Z}\mathbf{G}\mathbf{Z}^T$$



"All models are wrong but some are useful"  
George Box (1919 – 2013)



*Terima Kasih*

Anang Kurnia: [anangk@apps.ipb.ac.id](mailto:anangk@apps.ipb.ac.id)



**KEEP CALM  
AND LET THE  
DATA ANALYST  
HANDLE IT**