



Virtual Event 15-18 June 2020
**2020 Asia-Pacific
Statistics Week**

Leaving no one and nowhere behind

Estimation of SDGs Indicator for Non-Sampled Area Using Cluster Information

C: Integrated statistics for integrated analysis (SC2)
Journey towards integrated statistics

Presenter:

Rizky Zulkarnain & Dwi Jayanti
BPS-Statistics Indonesia

#apstatsweek2020



#apstatsweek2020



Virtual Event 15-18 June 2020
2020 Asia-Pacific
Statistics Week

Leaving no one and nowhere behind

Outline of Presentation

- Introduction
- Methodology
- Result and Discussion
- Conclusion



#apstatsweek2020



Virtual Event 15-18 June 2020

2020 Asia-Pacific Statistics Week

Leaving no one and nowhere behind

User Needs



1

More disaggregated and timely data

2

Reliable indicators to monitor SDGs achievement
at district level

3

Integrated approach is required to reduce
respondent burden and to increase cost-
efficiency

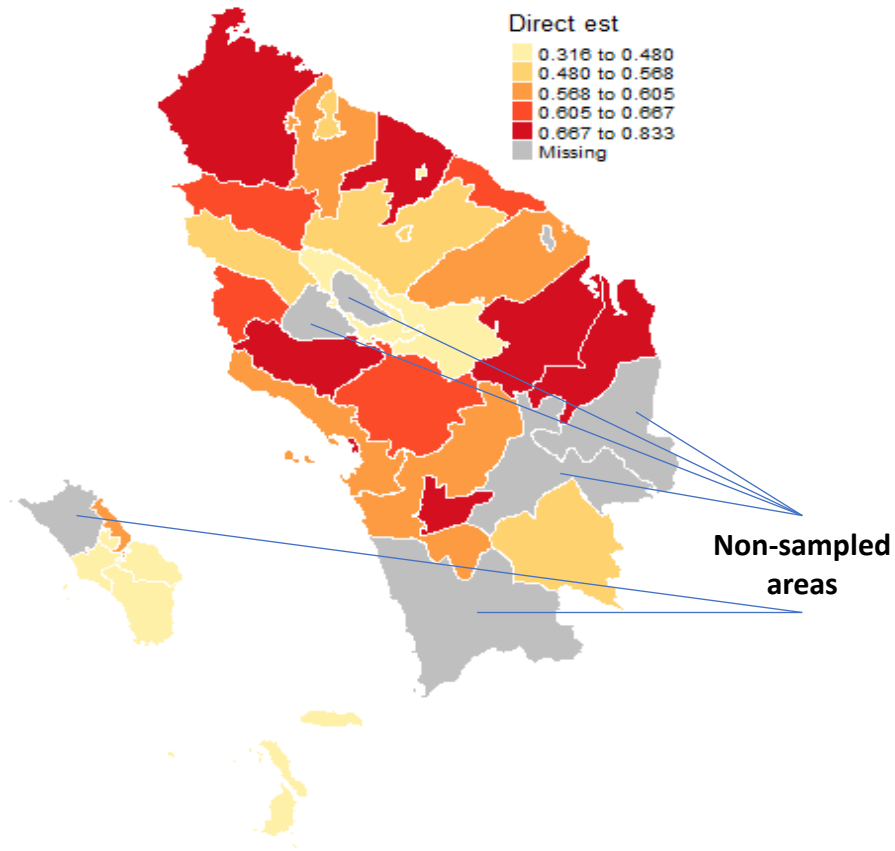


Virtual Event 15-18 June 2020

2020 Asia-Pacific Statistics Week

Leaving no one and nowhere behind

Problems



01

Some surveys such as IDHS produces estimate only for national and provincial levels.

02

SAE techniques are usually considered to resolve the problems. But, estimations are reliable for **sampled areas** only.

03

Roughly estimating **non-sampled areas** using synthetic model merely produces considerable bias.



Virtual Event 15-18 June 2020

2020 Asia-Pacific Statistics Week

Leaving no one and nowhere behind

Research Purpose



This paper utilizes similarities among areas to estimate SDGs indicator for non-sampled districts in North Sumatera. This paper uses CPR as indicator to be estimated.

#apstatsweek2020





Quality Dimensions



This study aims to contribute for improving Accuracy dimension of official statistics. This study refers to Data Integration dimension of Integrated Statistics.

Process
Integration

Data
Integration

Conceptual
Integration

Disciplinary
Integration



Quality Dimensions

| | |
|-------------|---------------|
| Relevance | Accessibility |
| Accuracy | Clarity |
| Reliability | Coherence |
| Timeliness | Comparability |
| Punctuality | |

Small Area Estimation considering cluster information (SAE-cluster) is intended to improve availability and quality of disaggregated statistics



Virtual Event 15-18 June 2020

2020 Asia-Pacific Statistics Week

Leaving no one and nowhere behind

Data

- Data for CPR in North Sumatera province is acquired from 2017 IDHS. Auxiliary variables are obtained from Family Planning Coordinating Board of North Sumatera.

Table 1. Lists of auxiliary variables

| No. | Variable | Description | Unit of measure |
|-----|----------|--|---------------------|
| 1 | Z_1 | Number of active acceptors | person |
| 2 | Z_2 | Number of family planning clinics | unit of clinics |
| 3 | Z_3 | Number of family planning institution | unit of institution |
| 4 | Z_4 | Number of pre-prosperous and 1 st prosperous family | unit of family |

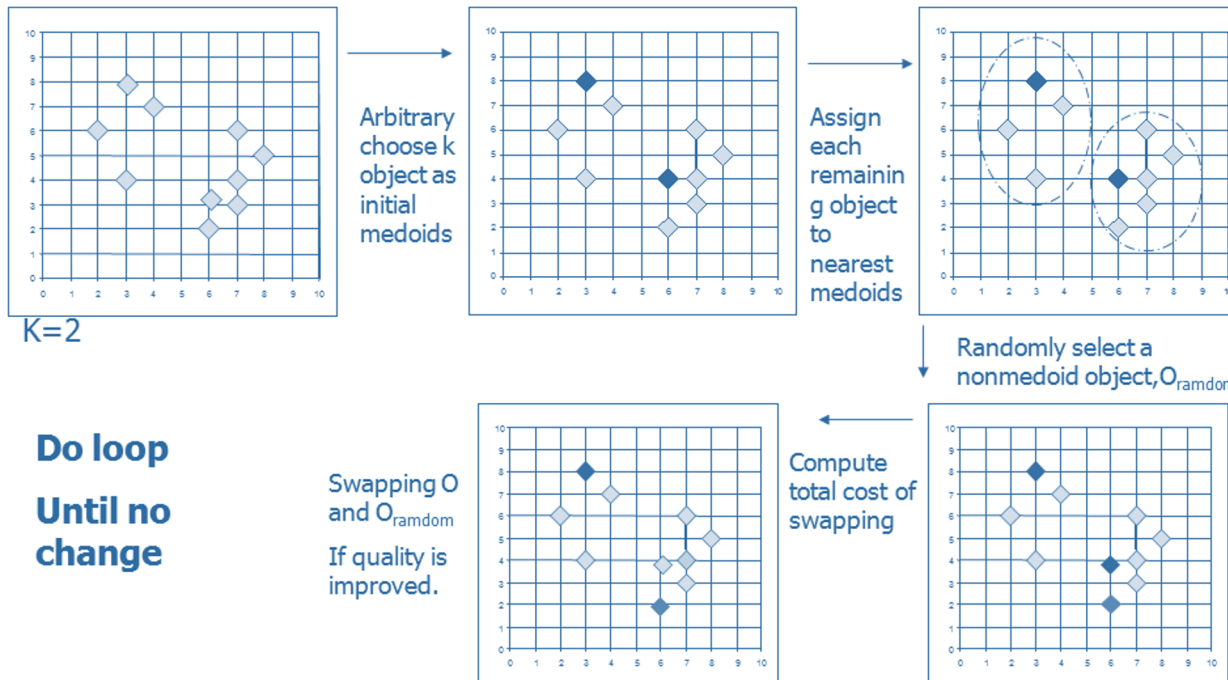


Virtual Event 15-18 June 2020

2020 Asia-Pacific Statistics Week

Leaving no one and nowhere behind

PAM Clustering





Model

- Prediction model for **sampled area**:

$$\ln \left(\frac{\hat{p}_{ik}}{1 - \hat{p}_{ik}} \right) = \hat{\beta}_0 + \sum_{p=1}^P \hat{\beta}_p Z_{ikp} + \sum_{d=1}^{K-1} \hat{\delta}_d D_d + \hat{v}_i$$

- Prediction model for **non-sampled area**:

$$\ln \left(\frac{\hat{p}_{i^*k}}{1 - \hat{p}_{i^*k}} \right) = \hat{\beta}_0 + \sum_{p=1}^P \hat{\beta}_p Z_{i^*kp} + \sum_{d=1}^{K-1} \hat{\delta}_d D_d + \bar{\hat{v}}_{(k)}$$

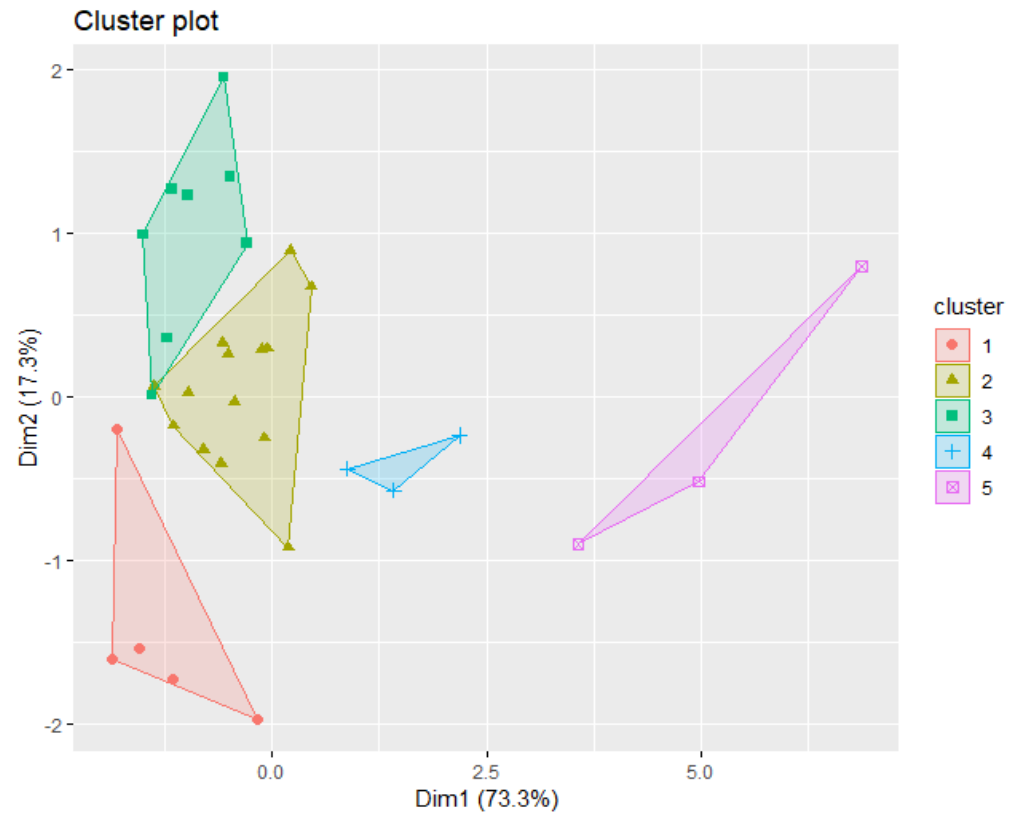
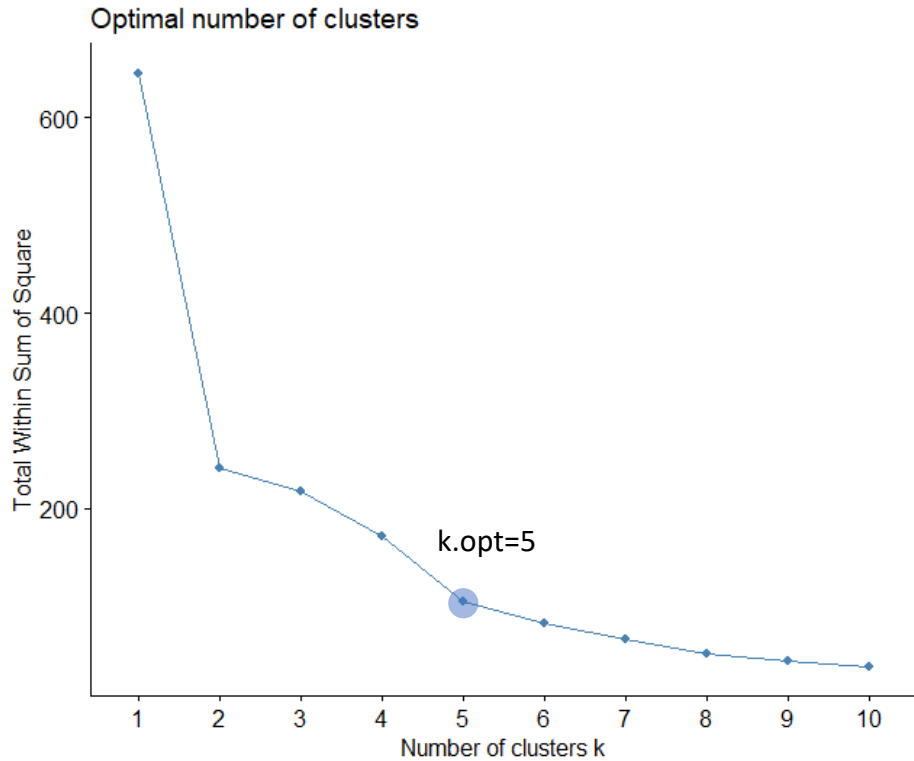


Virtual Event 15-18 June 2020

2020 Asia-Pacific Statistics Week

Leaving no one and nowhere behind

Results (1)





Virtual Event 15-18 June 2020

2020 Asia-Pacific Statistics Week

Leaving no one and nowhere behind

Results (2)

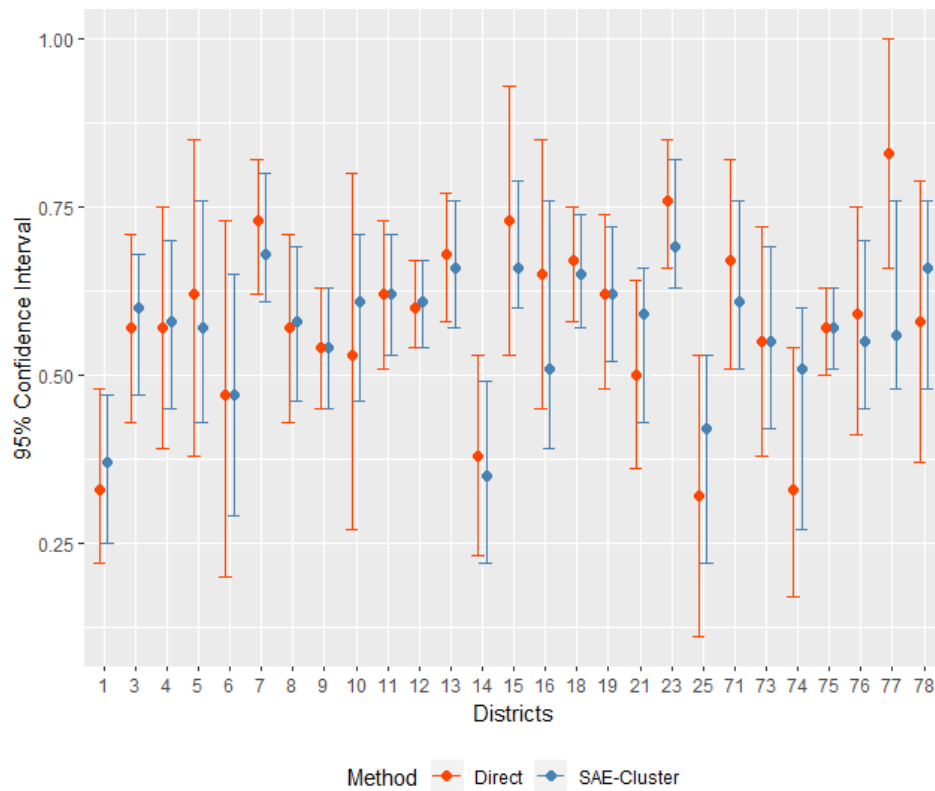
Table 2. Cluster of Districts in North Sumatera Province

| Cluster | Number of cluster members | Districts |
|---------|---------------------------|--|
| 1 | 5 | Nias, Nias Selatan, Pakpak Bharat, Nias Utara* , Nias Barat |
| 2 | 14 | Mandailing Natal* , Tapanuli Selatan, Tapanuli Tengah, Tapanuli Utara, Labuhan Batu, Dairi, Karo, Humbang Hasundutan, Batu Bara, Padang Lawas Utara* , Padang Lawas, Labuhanbatu Selatan* , Labuhanbatu Utara, Gunungsitoli |
| 3 | 8 | Toba Samosir, Samosir* , Sibolga, Tanjungbalai* , Pematangsiantar, Tebing Tinggi, Binjai, Padangsidempuan |
| 4 | 3 | Asahan, Simalungun, Serdang Bedagai |
| 5 | 3 | Deli Serdang, Langkat, Medan |

Note: *non-sampled area



Results (3)



SAE-cluster revises the direct estimates upward or downward. Confidence intervals of SAE-cluster are generally shorter than the direct estimates. This indicates that SAE-cluster produces more precise estimates than that direct method.

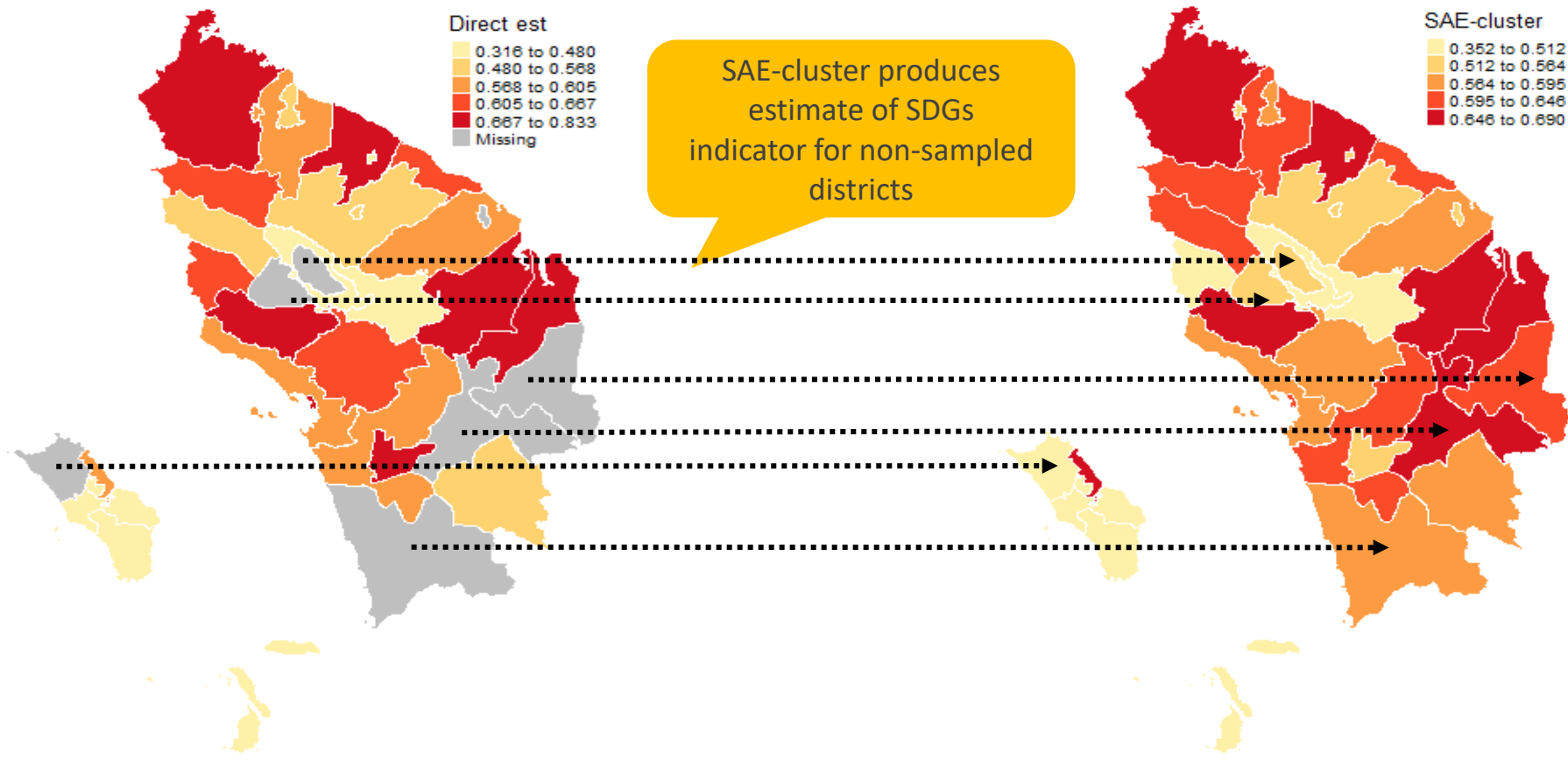


Virtual Event 15-18 June 2020

2020 Asia-Pacific Statistics Week

Leaving no one and nowhere behind

Results (4)





Virtual Event 15-18 June 2020

2020 Asia-Pacific Statistics Week

Leaving no one and nowhere behind

Conclusions

- SAE-cluster can improve the quality of disaggregated statistics, either for sampled areas or for non-sampled areas.
- SAE-cluster utilizes existing data sources to produce disaggregated statistics. Hence, respondent burden could be reduced and cost-efficiency could be increased.
- This study contributes to Accuracy dimension of official statistics and Data Integration dimension of Integrated Statistics.



Virtual Event 15-18 June 2020

2020 Asia-Pacific Statistics Week

Leaving no one and nowhere behind

References

1. Anisa R, Kurnia A, Indahwati. (2014). Cluster Information of Non-Sampled Area in Small Area Estimation. IOSR Journal of Mathematics, 10, 15-19.
2. Anisa R, Notodiputro KA, Kurnia A. (2014). Small Area Estimation for Non-Sampled Area Using Cluster Information and Winsorization with Application to BPS Data. Proc. ICCS-13, 27, 453-462.
3. Hastie T, Tibshirani R, Friedman J. (2009). The Elements of Statistical Learning 2nd Edition. New York: Springer-Verlag.
4. Rao JNK. (2003). Small Area Estimation. New York: John Wiley & Sons.
5. Saei A and Chambers R. (2005). Empirical Best Linear Unbiased Prediction for Out of Sample Area. Working Paper M05/03, Southampton Statistical Sciences Research Institute.
6. Sarda-Espinosa A. (2019). Time-Series Clustering in R Using the dtwclust Package. The R Journal. <https://doi.org/10.32614/RJ-2019-023>.





Virtual Event 15-18 June 2020

2020 Asia-Pacific Statistics Week

Leaving no one and nowhere behind

Thank You

#apstatsweek2020



DECADE
OF
ACTION



UNITED NATIONS
ESCAP
Economic and Social Commission for Asia and the Pacific



#apstatsweek2020