Grouping District / Cities in Papua Based Poverty Indicators

R Y Endartyana¹, E O Permatasari²

¹Department of Statistics, Faculty of Mathematics, Computing, and Data Science, Sepuluh Nopember Institute of Technology, Teknik Mesin road 175, Keputih, Sukolilo, Surabaya, East Java

²Department of Statistics, Faculty of Mathematics, Computing, and Data Science, Sepuluh Nopember Institute of Technology, Teknik Mesin road 175, Keputih, Sukolilo, Surabaya, East Java

*ratihendartyana@gmail.com, *erma.oktania@gmail.com

Abstract. Welfare of the people is a very important thing in a country. One measure of the welfare of society is seen from indicators of poverty. The government conducts a long-term agreement in terms of encouraging change through SDG's. One of the government's target is less poverty in all its forms by 2030. The area in Indonesia which has the highest percentage of poor people in the provinces of Papua. Hence, the groups based Indicators of Poverty in Papua province using C-Means Clustering and Fuzzy C-Means Clustering. The indicators used in this analysis are an indicator of Health, Education Indicators and Household Characteristics indicator. Based on the analysis results obtained Fuzzy C-Means formed 5 group is the best method to classify District / Cities in Papua Province by minimum icdrate value. Areas that need to be the focus of the government to reduce poverty in Papua is the Puncak District.

1. Introduction
Sustainable Development Goals is one of the government's efforts to deal long-term development in terms of encouraging changes to shift towards sustainable development in the social sector, economy, and environment [1]. SDGs has 17 objectives in order to continue the efforts to achieve the Millennium Development Goals (MDGs), one of which is to ended the poverty in all its forms wherever (Without Poverty). One of the SDGs targets in terms of poverty itself is alleviating extreme poverty for all income people who are currently less than 1.25 dollars per day in 2030. To achieve the target of SDGs in 2030 in terms of poverty, the government should strive to to less the number of poverty in Indonesia.

A person can be categorized as poor if it has an average expenditure per capita per month below the poverty line, one of which is the food poverty line which is equivalent to 2100 kilo calory per capita per day [2]. Based on data from the Asian Development Bank (2018), Indonesia ranked fourth in Southeast Asia with the percentage of poverty by 10.6% in 2017. The government through the Ministry of Social Affairs is targeting the poverty rate in Indonesia in 2019 is 9.8%. Provinces that have the highest percentage of poverty in Indonesia is Papua province that is equal to 27.74%. That is three times bigger than the poverty rate target set by the Government. Based on the 29 regencies / cities in the province of Papua, none of the regions have poverty rates below the Government's target. So, we need to do an analysis of indicators of what local governments need to be considered in order to reduce poverty in Papua.
Factor analysis about poverty in Papua Province by Anandanisa [3]. She said that the efforts of the government can do to reduce poverty is to increase the GDP and government spending in the form of subsidies, grants schools, and health. It is also said by Ekasari [4] that the quality of the economic impact on poverty. However, Ekasari added that the higher the quality of health will affect the economy and lower the high quality of poverty in the area.

This research aimed to perform grouping District / cities in Papua Province by indicators of poverty. Criteria for measuring poverty by BPS conducted through the Social Economic Registration Population in 2005 and better known as PSE05. In 2008 BPS updating PSE05 later renamed the Social Protection Program Data Collection (PPLS) in 2008 [5]. In this study the variables used by the measurement PPLS 2008 of indicator housing consists of, the percentage of households and primary building material house wall made from bamboo / wood, the percentage of households and facilities where defecation is used together with other households, the percentage of households the source of drinking water from wells / springs unprotected / river / rainwater, The percentage of households with main lighting source instead of electricity, and the percentage of households with a primary fuel firewood / charcoal / kerosene. Education indicators is the percentage of the population 15 years and older based on the latest education attained. Health indicators are the percentage of people sick and unable to pay for medical expenses.

The method used in this study is a non-hierarchical clustering which will compare between C-Means method and Fuzzy C-Means. The advantages of C-Means clustering method is to partition the data into homogeneous groups that are expected to determine which indicators group led to high levels of poverty in the area. While the Fuzzy C-Means method is a modified version of the C-Means method by adding a fuzzy nature (weighting named w) so that the fuzzy method is very robust to data outliers. The best method would then be used to classify the District / Cities in Papua Province by Indicators of Poverty.

2. Method
2.1. Descriptive Statistic
Descriptive statistics are contain useful methods to help organize and summarize the information on the data [8]. Kaizer Mayer Olkin is used to determine the adequacy of the correlation of variables in the multivariate analysis [9]. While Bartlett test is used to determine the correlation between the variables used are described by equation (2) which is said to have a correlation between variables if

Factor analysis is intended to do summarize variable becomes smaller. So that is intended to reduce the dimension of data / variables into small homogeneous groups. Results of factors formed in the factor analysis can then be followed in conducting non-hierarchical clustering method C-Means and Fuzzy C-Means in this study. So that is intended to determine whether the analysis of factors influence the grouping observation.

2.2. Clustering Analysis
Factor analysis is intended to do summarize variable becomes smaller. So that is intended to reduce the dimension of data / variables into small homogeneous groups. Results of factors formed in the factor analysis can then be followed in conducting non-hierarchical clustering method C-Means and Fuzzy C-Means in this study. So that is intended to determine whether the analysis of factors influence the grouping observation.

2.2.1 C-Means Method
C-Means method is a method of non-hierarchical group that partitioning the data into the same group so the group of data will have similar characteristics. Objective analysis using C-Means is to minimize the variation within a group, and maximize the variation between groups [11].

The algorithm used in the C-Means method is as follows.
1. Dividing the items into k-group
2. Calculating the value of the centroid
3. Grouping based on the average value of nearby using euclidian distance
4. Recalculating the center point of the group's members (item) if there are observation which enter and close
5. Iterating the 2nd and 3rd until no item into or out so that the convergence criteria are met.

2.2.2 Fuzzy C-Means Method

Fuzzy C-Means is a grouping method that was developed from c-means by applying the fuzzy as the basis of the weighting in the membership system [11]. The concept is applied to the fuzzy c-means is determining the center of the group for the first time that will mark the location of the average for each group. The algorithm used in FCM method is as follows.

1. Defining the matrix $X$ sized $n \times p$ where $n$ is the number of data and $p$ is the number of variables
2. Determine the group to be set up ($c$)
3. Determine the rank of a weight ($w > 1$)
4. Specifies the maximum iteration / loop (max.iter = 100)
5. Determining the tolerance limits or criteria looping is a very small positive value ($\varepsilon = 10^{-6}$)
6. Forming the initial partition matrix $U$ (degree of membership in a group), which is the beginning partition matrix which is random number between 0 and 1.
7. Calculating the value of the centroid of fuzzy, $c_i, i = 1, \ldots, c$ which subsequently formed matrix centroid value to calculate the objective function
8. Calculate the objective function ($P_1$)
9. Defining the membership of the new $U$ matrix is used for determining the cluster on each observation.
10. Iteration will stop when $|P_1 - P_0|$ smaller than ($\varepsilon$) but if it is not returned in step 5.

2.3. Pseudo-F Statistics

Pseudo-F Statistics is used to indicate that the number of groups formed have the optimum, where the variation in a group has a homogeneous while the variance between groups was heterogeneous.

2.4. Internal Cluster Dispersion Rate (Icdrate)

Method to choose the best clustering and assessing homogeneity in the group can use icdrate. This method can measure the average calculation of internal distribution to partition the overall cluster.

2.5. MANOVA One-Way

MANOVA analysis is used to determine the similarity between groups that have been formed. Before the MANOVA test, the assumptions that must be met are as follows.

1. The population data must meet multivariate normal distribution.
2. Variance-covariance matrix between treatments should be homogeneous views of box-M test if the value of $F$ is smaller than $F$ table or p value greater than $\alpha$.

To test the average similarity vector by finding the sum of squared matrix and product for the treatment and leftovers. In equivalent, will be obtained relationship relative size of the error (residual) and total (correction) the sum of the squares. Statistical tests used in MANOVA One-Way is as follows.

\[
F = \frac{SS_g / (g - 1)}{SS_n / (\sum n_i - g)}
\]

However, if the homogeneity test can not be met then used the test statistic pillai's trace as in equation 2.

\[
P = \text{tr} \left[ B (B + W)^{-1} \right]
\]

To test the difference in average group that has only one dependent variable used One-Way ANOVA test. Which hypothesis is as follows. The test statistic for ANOVA One-Way can be described by equations 3.
3. Results and Discussion
3.1. Description of Poverty Indicators
The characteristics of each of the indicators of poverty will be measured through centralization and dissemination of data. Centralization of power is described by the average value while the spread of data over the minimum and maximum.

The variables that explain the health indicator is the percentage of BAB facilities used by other households and the percentage of the population can not afford the cost of treatment. The average population who do not have their own facilities BAB is 40% which is the district with the highest percentage and the lowest Supiori district. While the average rat population in Papua own awareness of treatment is indicated by the average percentage of 0.82 with 16 regency / city that has a value of 0, which means they can always seek treatment when ill.

Education indicator used is the percentage of the population 15 years and over the last education elementary / No School. Jayapura is an area with a fairly good education is shown by the low percentage of the population with last education elementary / no school. While the Puncak District which has the highest percentage of 91.45%. That is still a lot of people in the District that low education level.

Presidential Decree No. 131 in 2015 set 122 regions in Indonesia are disadvantaged [13]. Puncak District is an area that still use traditional roof compared to the other regions. Neither the Peak District residents still use the lighting source instead of electricity. While there are 2 Regency / City which already use such kind of decent roof Nabire and Mimika regency. As for some of the figures also show that the Mimika district has the lowest number also in the category of Information Resources used and the type of Fuel used. Papua province, there are some areas that using this type of house walls are less suitable for bamboo / wood of which is Paniai, Nduga, Yalimo, Intan Jaya, and Deiyai. Mamberano Tengah District is an area that all of society is still spring water from wells / springs are not protected. Inversely proportional to the circumstances in Asmat, where all the people have to use water resources eligible for the household.

3.2. Testing Assumptions On Multivariate Analysis
In multivariate analysis, the assumptions that must be met is to meet the multivariate normal distribution, correlation between variables using Bartlett test and see adequacy correlation variable based on the value of KMO.

3.2.1 Multivariate Normal Assumptions
Testing multivariate normal distribution assumption can be seen through the scatterplot. Figure 2 shows that visually qq plot tends to form a straight line and reinforced the value of the proportion of 0.517 or 51.7%, it can be said that the data have a multivariate normal distribution.

\[
F = \frac{\sum i \sum j (Y_i - \bar{Y}_j)^2}{r-1 \sum i \sum j (Y_i - \bar{Y})^2/n-r}
\]
While based on a correlation of 0.983 is greater than 0.962 is also said that the data has met the normal distribution assumption.

1. Bartlett Correlation Test and Adequacy Correlation Variables

Testing bartlett and the adequacy of the data variables for indicators of poverty in Papua can be seen KMO value above 50% is 81.5% and the value of Chi-Square test bartlett 143.211 greater than 41.33. Thus concluded indicator variable on Poverty in Papua province are correlated and have enough multivariate analysis.

Factor analysis is used to describe covariance relationships between several variables that affect the indicators of poverty in Papua province in 2018. determining the number of factors that are formed based on the value of eigen value in Table 1.

| Component | Total | Variance Percentage | Cumulative Percentage |
|-----------|-------|----------------------|-----------------------|
| 1         | 4.750 | 59.369               | 59.369                |
| 2         | 0.986 | 12.326               | 71.695                |
| 3         | 0.817 | 10.208               | 81.903                |
| 4         | 0.592 | 7.398                | 89.301                |
| 5         | 0.371 | 4.640                | 93.941                |
| 6         | 0.238 | 2.979                | 96.920                |
| 7         | 0.150 | 1.878                | 98.798                |
| 8         | 0.096 | 1.202                | 100.00                |

Table 1 explains that many factors formed is one factor because only the main component of which has a value above 1, which is a factor of 1 with eigenvalues value equal to 4.750. However Rencher (2002) said that in addition to the value of eigen value principle component values can be seen from the magnitude of the cumulative percentage of 80%. So many factors that formed was 3. Variable-forming factors can be explained through Table 7. Which one factor consists of 6 variables, factor 2 consists of one variable, and the third factor is composed of one variable.

| Component | 1     | 2     | 3     |
|-----------|-------|-------|-------|
| Roof Bamboo / Wood / Ijuk | 0.763 | 0.479 | 0.038 |
| Wall Wood / Bamboo | 0.737 | 0.398 | 0.233 |
| Not Have Amenities CHAPTER | 0.919 | -0.083 | 0.137 |
| Spring Water Well / Spring Not Protected | 0.166 | 0.917 | 0.144 |
| Non Electric Lighting Source | 0.676 | 0.566 | 0.190 |
| Top BB Wood / Charcoal / Kerosene | 0.318 | 0.457 | 0.612 |
| Recently diploma SD / No School | 0.755 | 0.398 | 0.355 |
| Not Able to Pay Medical Expenses | -0.112 | -0.048 | -0.923 |

3.3. Clustering Analysis

Analysis of the group used C-Means and Fuzzy C-Means. The number of groups that are formed based on the results of the analysis of the hierarchy cluster dendogram using the single linkage formed two groups so that the determination of the amount of the initial group were used in this analysis is the 2 groups.

Results grouping with C-Means method form 2 to 5 groups. The optimum group visits based on the value of the pseudo optimum f-statistics in Table 8. Which of these groupings is also compares the data on receipts of factor analysis and data using factor analysis.

Pseudo F-statistics value most optimum is the group number 2 on the data without using factor analysis with pseudo f-statistics value of 31.836. However, based on the minimum value of 0.854
icdrate, the optimum cluster C-Means method is 4 clusters using factor analysis. The number of members that are formed in the group 1 is 6 districts / cities, the number of members in the group 2 is 3 districts / cities, group 3 there are 8 districts / cities and group 4 contained 12 Kabupeten / city.

Table 3: Pseudo F-statistics Value Using C-Means Method

| Number of Groups | Pseudo F-Statistics | icdrate |
|------------------|---------------------|---------|
| Without Factor Analysis | 2 | 31,836* | 0,962 |
| Factor Analysis | 3 | 19,390 |  |
| | 4 | 31,207 |  |
| | 5 | 0,9616 |  |
| Using Factor Analysis | 2 | 9,176 |  |
| Factor Analysis | 3 | 12,468 |  |
| | 4 | 22,465 | 0,854* |
| | 5 | 19,286 |  |

The second method used is Fuzzy C-Means with the group results optimum value based Pseudo f-statistics are shown in Table 4 which shows that the group most optimum and value icdrate minimum indicated by the fuzzy c-means with 5 groups by factor analysis with icdrate value amounted to 0,806.

Table 4: Pseudo F-statistics Value Using Fuzzy C-Means Method

| Number of Groups | Pseudo F-Statistics | icdrate |
|------------------|---------------------|---------|
| Without Factor Analysis | 2 | 31,8364 |  |
| Factor Analysis | 3 | 78,4019* | 0,924 |
| | 4 | 31,2075 |  |
| | 5 | 24,5657 |  |
| Using Factor Analysis | 2 | 9,3572 |  |
| Factor Analysis | 3 | 11,8313 |  |
| | 4 | 11,8848 |  |
| | 5 | 24,9247 | 0,806* |

Furthermore, to the best methods used to compare views through icdrate minimum value between the two methods that can be explained through Table 5. Fuzzy C-Means method using factor analysis is the best method to classify the data indicators Poverty in Papua province in 2018. It such as Fuzzy C-Means has the minimum icdrate value that will give a homogenous grouping results between groups. Which amount of members in each group is formed Optimum 3 districts / cities respectively in group 1 and group 2, 6 counties / cities in group 3, group 4 there are 10 districts / cities and groups of 5 are 7 districts / cities.

Table 5: Icdrate Value for C-Means dan Fuzzy C-Means Method

| Value | C-Means | Fuzzy C-Means |
|-------|---------|---------------|
| SSW   | 22,727  | 16,279        |
| SSB   | 12,217  | 16,273        |
| icdrate | 0.854   | **0.806**    |

3.4. MANOVA One-Way
One-Way MANOVA testing is used to determine whether or not differences in each group formed on the dependent variable factor that has more than 1. Before the One-Way MANOVA test then testing the homogeneity of variances explained obtained value P-value is less than 0,05 so it is said that the variance covariance matrix is not homogeneous so that in testing MANOVA using Pillia's Trace test statistics that can be viewed by the value in Table 6.
Based on Table 6 known value of $F$ test statistic is greater than 2.0073 then it means that the group 1, group 2, group 3, group 4 and group 5 are formed each have differences.

For show the differences between groups were formed on the factors that has one dependent variable used One-Way ANOVA test in which the factor 2 and factor 3 has a value of $P$-value respectively 0.000 and 0.025 less than 0.05 then concluded between groups formed mutual differences. It can be concluded that the clustering analysis using fuzzy c-means and factor analysis between groups were formed each have differences.

### Table 7 Difference Each Group

| Group                          | 1     | 2     | 3     | 4     | 5     |
|-------------------------------|-------|-------|-------|-------|-------|
| N                             | 3     | 3     | 6     | 10    | 7     |
| Roof Bamboo / Wood / Ijuk     | 16.1% | 65.8% | 60.7% | 3.14% | 23.86%|
| Wall Wood / Bamboo            | 64.7% | 94.31%| 99.48%| 38.64%| 93.84%|
| Got Tidpak CHAPTER Facilities | 31.2% | 84.54%| 42.77%| 21.39%| 48.90%|
| Spring Water Well / Spring    | 8.66% | 10.30%| 81.64%| 21.23%| 37.59%|
| Not Protected                 |       |       |       |       |       |
| Non Electric Lighting Source   | 10.42%| 71.6% | 77.65%| 7.21% | 35.11%|
| BB Wood / Charcoal / Oil      | 94.50%| 99.87%| 99.91%| 97.86%| 99.45%|
| Tanh                          |       |       |       |       |       |
| Recently diploma SD / No      | 42.7% | 84.88%| 80.88%| 39.87%| 73.33%|
| School                        |       |       |       |       |       |
| Not Able to Pay Medical       | 6.9%  | 0.00% | 0.79% | 1.33% | 0.00% |
| Expenses                      |       |       |       |       |       |

On the average it is known that Group 4 is a District / Cities underlying variable low rate of poverty in Papua than the other groups. This can be explained by the average of the percentage of each variable were the lowest compared to other groups. Economy of people in the region can be said to have a good

Group 1 is the area with the average percentage of the poverty indicators are quite low. Figures low enough Atam shown on variable kinds of bamboo / wood, the source of drinking water, non-electric light sources, as well as the inability of the doctor. This shows that predominantly people living in the area in group 1 began much conscious to have a good occupancy and sufficient for life in a proper manner.

Group 5 had the highest percentage of poverty level indicators third after the group 2 and 3. However, the percentage is less than 50% in some of the variables in the group 5 address that from a few people who still have not been decent occupancy, yet still have the awareness to seek treatment when ill is shown through the percentage of treatment capability that shows the number 0.

Group 3 is an area that is still poverty is characterized by the lack of ability to provide for decent for everyday life. This can be demonstrated by the high percentage of water source wells / unprotected, non-electric lighting source, and fuel wood / charcoal / kerosene. Likewise decent housing through the wall type used still uses wood / bamboo reached 99.91%. This attribute that almost all people in the group 3 still uses the traditional type of wall.
Group 2 is a region whose education was the lowest compared to other groups, leading to a lack of awareness to have occupancy decent, evidenced by the high percentage of the population that uses the roof of bamboo / wood, as well as lack of awareness of hygiene as evidenced by the percentage does not have BAB facility reached 84.88%.

Based on the results of the characteristics of each group is formed it can be concluded that one of the indicators that should be the focus of the government in reducing poverty in the province of Papua on average are indicators of household characteristics. But there are more important indicators to be repaired as a benchmark for other indicator fix is education. Impact obtained by repairing education of the population in the area will provide public awareness for the care and health conscious and create shelter and good housekeeping.

The results of the mapping of the group formed from the results of the analysis using Fuzzy C-Means can be seen in Figure 2. Which at grouping the results into 5 groups that formed it can be said that the group 2 is shown by the red color in Figure 4.7 is a group with poverty high, group 3 is shown by the color orange is relatively high poverty, group 5, which is shown by the yellow color is poverty being, group 1 indicated by the green color is poverty that is quite low, and group 4 indicated by the blue color is poverty The low one.

4. Conclusion
Puncak District is an area that has had a percentage in the lowest education indicators, areas where people still use this type of roof on the fringe, still using illumination sources instead of electricity, and many who use the BAB facility along with other households. So it can be said that the Puncak District is an area of the highest poverty rates are found in all indicators of poverty. KMO and Bartlett test based on the data said to have been worth the multivariate analysis and found to meet the normal distribution assumption.

Grouping using C-Means uses 2 to 5 groups of 4 groups are among the most optimum for f value pseudo-statistics biggest and most minimum icdrate which the data to better factor analysis is used for this method. While the method of Fuzzy C-Means Optimum group is 5 begitupula group Fuzzy C-Means method most optimum is to use factor analysis. Icdrate minimum value indicates that the Fuzzy C-Means method is better than C-Means. While the analysis MANOVA One-Way and One-Way ANOVA showed that each variable in the group pick the difference.

Based on the results of grouping said that the group 2 is a group with indicators of high poverty rates, group 3 is a group with indicators of poverty rate is high enough, a group of 5 is a group with poverty indicators being, group 1 is a group in which poverty rates were quite low as well as a group of 4 a group with a low poverty rate indicator.

Suggestions for local governments is poverty is very crucial to the population problem in Indonesia which Papua is enough area left in Indonesia. So the focus of the government to reduce poverty is needed, especially in terms of improving the educational level of the population in Papua Province so
as to support other indicators in terms of population so that the poverty rate in Papua Province can be reduced.

References
[1] Bappenas. (2019). What the SDGs. Taken back from the Ministry of Planning / National Development Planning Agency (Sustainable Development Goals):http://sdgs.bappenas.go.id/
[2] BPS. (2018). Province of Papua in Jayapura, 2018. Figures: Statistics of Papua Province.
[3] Anandanisa, NP (2018). Analysis of Poverty in Papua province. Yogyakarta: Islamic University of Indonesia.
[4] Ekasari, DF (2011). SEM with Generalized Structure Modeling Component Analysis (GSCA) (Case Study of Poverty in Central Java). Surabaya: Department of Statistics Faculty of ITS.
[5] Isdijoso, W., Suryahadi, A., & Akhmadi. (2016). Determination of criteria and variables of the Poor Comprehensive Data Collection in the Context of the Protection of the Poor in the District / City. Institute Research Institute.
[6] Sadjidah, A. (2016). Grouping Province in Indonesia Based Public Welfare Indicator Method C-Means and Fuzzy C-Means Clustering. Surabaya: Sepuluh Nopember Institute of Technology.
[7] Sukim. (2011). Study on Method C-Means Cluster and Fuzzy C-Means Cluster And Its Application in Case Grouping / District Village By Status underdevelopment. Surabaya: Sepuluh Nopember Institute of Technology.
[8] Walpole, R. (2012). Introduction to Statistical Methods. Interpretation: Ir Bambang. Jakarta: Gramedia Pustaka Utama.
[9] Hair, JF, Black, WC, Babin, JB, & Anderson, ER (2001). Multivariate Data Analysis (Seventh ed.). New Jersey: Prentice Hall Inc.
[10] Johnson, RA, & Winchern, DW (2007). Applied Multivariate Statistical Analysis (6th ed.). United State of America: Pearson Prentice Hall.
[11] Agusta, Y. (2007). K-Means - Implementation, Problems and Related Methods. Journal of Information Systems, 47-60.
[12] BPS. (2019). Poverty and Inequality. Taken back from the Central Bureau of Statistics: https://www.bps.go.id/subject/23/kemiskinan-dan-ketimpangan.html.
[13] Sawabi, G. (2015). In Papua There are 26 District Otherwise Disadvantaged, NTT 18 districts.