

# West Java Province Regency/City Grouping Based on Average Vegetable Type Expenditure Per Capita a Week Using Cluster Analysis

Marcelena Vicky Galena<sup>1</sup>, Dessy Nur Aisyah<sup>2</sup>, Irma Ayu Indrasta<sup>3</sup>, Marshanda Aprilia<sup>4</sup>, Dita Amelia<sup>5</sup>, M. Fariz Fadillah Mardianto<sup>6</sup>, Elly Ana<sup>7</sup>

<sup>1,2,3,4,5,6,7</sup>Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Surabaya, Indonesia

Email : <sup>1</sup>[marcelena.vicky.galena-2021@fst.unair.ac.id](mailto:marcelena.vicky.galena-2021@fst.unair.ac.id), <sup>2</sup>[dessy.nur.aisyah-2021@fst.unair.ac.id](mailto:dessy.nur.aisyah-2021@fst.unair.ac.id), <sup>3</sup>[irma.ayu.indrasta-2021@fst.unair.ac.id](mailto:irma.ayu.indrasta-2021@fst.unair.ac.id), <sup>4</sup>[marshanda.aprilias-2021@fst.unair.ac.id](mailto:marshanda.aprilias-2021@fst.unair.ac.id), <sup>5</sup>[dita.amelia@fst.unair.ac.id](mailto:dita.amelia@fst.unair.ac.id), <sup>6</sup>[m.fariz.fadillah.m@fst.unair.ac.id](mailto:m.fariz.fadillah.m@fst.unair.ac.id), <sup>7</sup>[elly-a@fst.unair.ac.id](mailto:elly-a@fst.unair.ac.id)

Corresponding Author : Dita Amelia

**Abstract:** Based on the results of the 2018 Basic Health Research (Riskesdas) shows that 95.5% of the population aged  $\geq 5$  years consume less vegetables and fruits. Meanwhile, in West Java Province, 98.2% of the population aged  $\geq 5$  years does not consume vegetables and fruits. The proportion is even higher when compared to the proportion nationally. Lack of consumption of vegetables and fruits can affect the supply of fiber, vitamins, and minerals that are needed by the body. Although the interest in vegetable consumption of Indonesians is still relatively low, in Indonesia vegetables are still one of the mandatory menus in daily consumption. Quoting from the Central Statistics Agency (BPS), the average per capita expenditure a month for vegetable consumption tends to increase every year. The peak occurred in March 2021 with a per capita expenditure of IDR 53,864 per month. This clustering process was carried out for 27 regencies/cities in West Java. Grouping into two clusters is selected based on the highest Pseudo – F value that has been calculated. From the results of the analysis, the first cluster found that the per capita income a week according to vegetables in this cluster is relatively high. In the second cluster is a group of regencies/cities in West Java with a per capita income a week according to low vegetables. If every region in Indonesia can optimize the quality of vegetables, it is expected that the consumption interest of the Indonesian people will also increase and the country's per capita income from vegetable commodities will also increase.

**Keywords—** Vegetable, Consumption, Expenditure

## 1. INTRODUCTION

Vegetables are foodstuffs derived from plant parts such as leaves, stems, and flowers [1]. In vegetables there are various kinds of nutrients including fiber, water, vitamins, and minerals. In general, vegetables derived from roots such as tubers and potatoes contain a lot of starch and carbohydrates, while green vegetables have a lower sugar content, and vegetables in the form of leaves contain a lot of vitamins, water, and minerals [2].

Based on the results of Basic Health Research on 2018 shows that 95.5% of the population aged  $\geq 5$  years consume less vegetables and fruits [3]. The proportion is even higher when compared to the proportion nationally. Lack of consumption of vegetables and fruits can affect the supply of fiber, vitamins, and minerals that are needed by the body. Sources of fiber, vitamins, and minerals found in vegetables and fruits are indispensable for the body to achieve a healthy diet in accordance with balanced nutrition guidelines [4].

Although the interest in vegetable consumption of Indonesians is still relatively low, in Indonesia vegetables are still one of the mandatory menus in daily consumption. As

shown by the Central Agency on Statistics [5], the average per capita expenditure on vegetable spending was IDR 48,654 per month in September 2021. However, this amount decreased by 9.67% when compared to March 2021, which was IDR 53,864 per month. Although per capita expenditure in one month of vegetable spending decreased, vegetables are still the food commodity with the highest expenditure after finished food and beverages, cigarettes, and grains, which is 8.41% of the total per capita expenditure on food in one month. If you look at the trend, the average per capita expenditure a month on vegetable consumption tends to increase every year. The peak occurred in March 2021 with a per capita expenditure of IDR 53,864 per month [6].

Based on data from the Central Agency on Statistics, the average per capita expenditure a week by vegetable type in 2021 in West Java Province can be grouped by region and type of vegetable. So that it can be known which groups have the lowest, middle, and highest per capita expenditures from both regional groupings and vegetables. The grouping will then be analyzed and then conclusions and suggestions can be obtained to optimize plantation yields so that the quality of vegetables in West Java Province is better, so as to increase the interest in vegetable consumption of the Indonesian

people and can increase the country's per capita income from vegetable commodities.

## 2. REVIEW OF LITERATURE

### 2.1 Per Capita Expenditure

Per capita expenditure is a cost incurred for the consumption of all family members divided by the number of household members that has been adjusted to purchasing power parity. Spending data can reveal household consumption patterns in general using indicators of the proportion of expenditure on food and non-food. The composition of household expenditures can be used as a measure to assess the level of economic well-being of the population. The lower the percentage of expenditure on food to total expenditure, the better the level of welfare.

### 2.2 Multivariate Analysis

The multivariate analysis method is a method in statistics with the aim of analyzing variables in a data where the variables are interconnected with  $I = 1, 2, \dots, n$  and  $j = 1, 2, \dots, p$  [7]. Multivariate data with  $n$  observations on  $p$  variables can be displayed as follows.

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2k} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{j1} & x_{j2} & \dots & x_{jk} & \dots & x_{jp} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} & \dots & x_{np} \end{bmatrix} \quad (1)$$

Matrix  $X$  contains data from all variables where  $n$  rows and  $p$  columns [8].

### 2.3 Cluster Analysis

Cluster Analysis is a multivariate method that aims to group  $n$  objects into  $k$  clusters with  $k < n$  based on  $p$  changes (variables) so that the observation units in one cluster have more homogeneous characteristics than observation units in other clusters. The objectives of this cluster analysis are:

1. Grouping objects on the basis of similarity or inequality
2. Is a method that is often used if there is no hypothesis to be tested
3. Grouping based solely on the fact of the data

There are 2 types of cluster analysis methods, namely hierarchical cluster analysis and non-hierarchical analysis. In the cluster hierarchy method, there are two basic types, namely agglomerative and divisive.

1. Agglomerative  
Each object is considered as a cluster of its own, then two similar clusters will be combined, and so on.

2. Divisive

It started from a large cluster consisting of all objects. Next, the least similar objects are separated, and so on

In agglomerative there are five methods that are quite well known, namely: Single Linkage, Complete Linkage, Average Linkage, Ward's Method, Centroid Method.

- Single Linkage  
Based on the minimum distance. Starting with two objects separated by the shortest distance, they will be placed in the first cluster, and so on. This method is also known as the closest neighbor approach.
- Average Linkage  
Average linkage treats the distance between two clusters as the average distance between all pairs of items where one member of that pair belongs to each cluster.
- Complete Linkage  
Complete linkage clustering produces many of the same ways as single linkage clustering, except for: at each stage, the distance (similarity) between the clusters is determined by the distance (similarity) between the two elements, one of each of the most distant clusters.
- Ward's Method  
The distance between the two clusters in this method is based on the total sum of square of two clusters on each variable.
- Centroid Method  
The distance between the two clusters in this method is based on the centroid distance of the two clusters in question [9].

### 2.4 Complete Linkage Method

The Complete Linkage method calculates the distance of two clusters which is referred to as the farthest distance. Before starting the calculation, first calculate the minimum value on the matrix  $d_{UV}$  and then combine the corresponding objects. It is then calculated on each cluster by the following equation:

$$d_{(UV)W} = \max\{d_{uw}, d_{vw}\} \quad (2)$$

where  $d_{uw}$  and  $d_{vw}$  is the farthest distance in members of the  $U$  and  $W$  clusters and  $V$  and  $W$  clusters.

### 2.5 Pseudo – F Statistic

To determine the optimal number of clusters using the statistical values Pseudo – F. Equations used are:

$$\text{Statistik Pseudo} - F = \frac{\left(\frac{R^2}{k-1}\right)}{\left(\frac{1-R^2}{n-k}\right)} \quad (3)$$

where,

$$R^2 = \frac{(SST - SSW)}{SST} \quad (4)$$

$$SST = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (x_{ijk} - \bar{x}_j)^2 \quad (5)$$

$$SSW = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (x_{ijk} - \bar{x}_{jk})^2 \quad (6)$$

Information:

SST (Sum Square Total) : Total sum of the squares of the sample distance against the overall average

SSW (Sum Square Within) : Total sum of the squares of the sample distance against the average of its groups

n : Number of samples

c : Number of variables

p : Number of groups

$x_{ijk}$ : Sample to  $i$  on variable to  $j$  group to  $k$

$\bar{x}_j$  : Average of all samples on variable to  $j$

$\bar{x}_{jk}$  : Sample mean on variable to  $j$  and group to  $k$

The optimal number of clusters is obtained if the number of clusters has the highest result from the Pseudo – F statistic.

### 3. METHOD

This research is a type of qualitative research. According to Saryono in 2010 [10] qualitative research is research used to investigate, discover, describe, and explain the quality or specialness of social influences that cannot be explained, measured, or described through a quantitative approach. This research was carried out in December 2022 with data sources obtained from the Central Agency on Statistics regarding the average per capita expenditure a week according to the type of vegetables in 2021 in West Java Province [11].

This study aims to analyze per capita expenditure by type of vegetables in 2021 in West Java Province using the cluster analysis method, which then the results of the analysis can provide insights to optimize plantation yields so that the quality of vegetables in West Java Province is better, so as to increase interest in vegetable consumption of the Indonesian people and can increase the country's per capita income from vegetable commodities.

Cluster analysis is a type of statistical analysis that aims to group objects based on the similarity of characteristics between these objects. The objects will be grouped into one or more groups (clusters) so that the objects in one group will have similarities with each other [12]. The complete linkage method is a method that uses the maximum distance between groups. The use of this method will provide one solution in its completion, which is based on a measure of similarity in the distance technique used.

The results of the data that have been processed and analyzed will then be discussed based on theoretical reviews so that conclusions and suggestions can be obtained to help optimize plantation yields so that the quality of vegetables in West Java Province is better, so as to increase the interest in public vegetable consumption. If every region in Indonesia can optimize the quality of vegetables, it is expected that the consumption interest of the Indonesian people will also increase and the country's per capita income from vegetable commodities will also increase.

## 4. RESULTS AND DISCUSSION

### 4.1 Determination of the Optimal Number of Clusters

Before grouping or analyzing clusters, we must first determine the optimal number of clusters using the calculation of pseudo-F statistical values for the number of clusters as many as 2 to 4 clusters and the results are displayed in the table below.

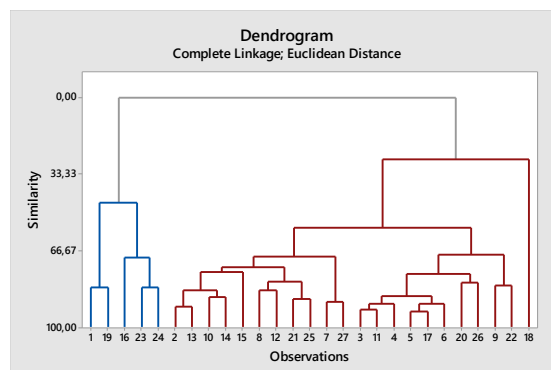
**Table 1:** Pseudo - F Values in the Complete Linkage Method

| Multiple Clusters | Pseudo – F |
|-------------------|------------|
| 2                 | 3,060      |
| 3                 | 2,410      |
| 4                 | 2,056      |

Based on Table 1, the highest Pseudo – F value is found in many clusters of 2. So that the optimal cluster is 2 clusters for cluster analysis hierarchy of complete linkage methods.

### 4.2 Analysis Results

The results of the cluster analysis of the complete linkage hierarchy in the grouping of per capita expenditure a week regarding vegetables in West Java are divided into 2 clusters. The following are the output and interpretation results of the complete linkage hierarchy cluster analysis.



**Fig. 1.** Dendrogram Output of Observation Cluster Analysis

Based on Fig 1, there are 2 different clusters marked with color differences. Blue is the first cluster containing 5 regions, red is the second cluster containing 22 regions. Here are which regions are members of each cluster when viewed based on dendrograms.

**Table 2:** Output of Final Partition Analysis of Observation Cluster

|           | Number of Observations | Average Distance from Centroid |
|-----------|------------------------|--------------------------------|
| Cluster 1 | 5                      | 827,617                        |
| Cluster 2 | 22                     | 667,548                        |

Based on Table 2, it was found that the average distance in the first cluster from the centroid value was 827,617 and the average distance in the second cluster from the centroid value was 667,548.

**Table 3:** Output Distances Between Cluster Centroids Analysis Cluster Observations

|           | Distances Between Cluster Centroids |           |
|-----------|-------------------------------------|-----------|
|           | Cluster 1                           | Cluster 2 |
| Cluster 1 | 0,00                                | 2204,67   |
| Cluster 2 | 2204,67                             | 0,00      |

Based on Table 3, it is obtained that the distance of the centroid value between the cluster to 1 and itself is 0,00. The centroid value distance between cluster to 1 and cluster to 2 (and vice versa) is 2204,67. Then the distance of the centroid value between the 2nd cluster and itself is 0,00.

**Table 4:** Results of Observation Cluster Analysis

| Cluster Type | Cluster Members                                | Characteristic                   |
|--------------|--|----------------------------------|
| Cluster 1    | Bogor Region (1), Bogor City area (19), Bekasi | The area is a relatively densely |

| Cluster Type | Cluster Members  | Characteristic  |
|--------------|--|---|
|              | Region (16), Bekasi City area (23), Depok City area (24)   | populated area so that the interest and purchasing power of the people in the cluster is also high.   |
| Cluster 2    | Sukabumi Region (2), Subang Region (13), Majalengka Region (10), Purwakarta Region (14), Karawang Region (15), Kuningan Region (8), Indramayu Region (12), Bandung City area (21), Cimahi City area (25), Ciamis Region (7), Banjar City area (27), Cianjur Region (3), Sumedang Region (11), Bandung Region (4), Garut Region (5), West Bandung Region (17), Tasikmalaya Region (6), Sukabumi City Region (20), Tasikmalaya City area (26), Cirebon Region (9), Cirebon City area (22), Pangandaran Region (18) | The areas in cluster 2 contain mostly sparsely populated areas because many residents of the region have jobs in the area that is in cluster 1 so that cluster 2 has a small interest and purchasing power in vegetables. |

The resulting interpretation is that when viewed based on centroid values in cluster 1 consisting of 5 regions, are the areas where the average per capita expenditure a week is

considered the largest. When viewed based on centroid values in cluster 2 consisting of 22 regions where the average per capita expenditure a week is considered the smallest compared to other regions. So that the purchasing power of vegetables in the 1st cluster is higher than the 2nd cluster.



**Fig. 2.** Cluster Analysis Map

Based on Fig 2, there are three areas marked in red, which include the Bekasi City area, Depok City area, and Bogor City area. This means that the three regions are included in cluster 1. Then there are areas marked with yellow, which include the Sukabumi City area, Cimahi City area, Bandung City area, Tasikmalaya City area, Banjar City area, and Cirebon City area. This means that these regions belong to cluster 2. where the two characteristics of the two clusters are as in Table 4.

## 5. COVER

### 5.1 Conclusion

Based on the analysis that has been carried out using cluster complete linkage, the results of grouping the average per capita expenditure a week based on the type of vegetables in West Java province in 2021 were obtained based on cluster observations. The results of the observation cluster analysis are based on the average expenditure of the people of West Java on vegetable types by looking at the area that has the highest average per capita. Regions in West Java are grouped into 2 clusters, namely areas where the average per capita expenditure a week is considered the largest and smallest.

### 5.2 Suggestion

The reasons that can be given to the results of the analysis that has been carried out are:

#### a. Analysis revenue share

Based on the results of the analysis, optimization of vegetable growth can be carried out throughout the West Java region by taking good care of vegetable crops, such as applying quality fertilizers. The application of fertilizer in each region is expected to have good quality

so that the vegetable harvest of each region has good quality as well. In addition, the government is expected to provide good quality fertilizer subsidy assistance to help increase vegetable yields in West Java.

#### b. For readers

The results of this cluster analysis can be explored using other clustering methods.

## 6. REFERENCES

- [1] D. Tansari. (2014). "Kualitas Es Krim Dengan Kombinasi Wortel (*Daucus Carota* L.) Dan Tomat (*Lycopersicum Esculentum* Mill.)". Available: <http://e-journal.uajy.ac.id/6537/>.
- [2] Z. Powerful. *Pengertian Sayuran dan 3 Macam Pengelompokannya* [Online]. Available: <https://www.awalilmu.com/2017/09/pengertian-sayuran-dan-3-macam-pengelompokannya.html>
- [3] O.F.Chandry, M.Haya, A.P.Sari, "Overview of Knowledge of the Importance Consumption of Fruit and Vegetables in Age Children School in Rena Long Village," 2021.
- [4] Hermina and S. P. " Gambaran Konsumsi Sayur dan buah penduduk Indonesia dalam konteks gizi seimbang: Analisis Lanjut survei Konsumsi Makanan Individu (SKMI) 2014" *Health Research Bulletin*, 44(3), 2016.
- [5] BPS. *Pendapatan Per Kapita* [Online]. Available:<https://sirusa.bps.go.id/sirusa/index.php/indikator/197#:~:text=Pengeluaran%20per%20kapita%20adalah%20biaya,disesuaikan%20dengan%20paritas%20daya%20beli>.
- [6] M. A. Rizaty. (2022, Sep. 1). *Pengeluaran Konsumsi Sayur di Indonesia turun per September 2021* [Online]. Available: <https://dataindonesia.id/bursa-finance/detail/expenditure-vegetable-consumption-in-indonesia-down-per-september-2021>.
- [7] Hidayat. (2016, Nov). *Penjelasan Tentang Analisis Multivariat Dan Jenisnya* [online]. Available:<https://www.statistikian.com/2016/11/analisis-multivariat.html>
- [8] W. Alwii and M. Hasrul, " Analisis Klaster Untuk Pengelompokan Kabupaten/Kota Di Provinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat" *MSA Journal*, 35-42, 2018.
- [9] A, Hidayat. (2014, March). *Penjelasan Lengkap Tentang Analisis Cluster* [Online]. Available:<https://www.statistikian.com/2014/03/analisis-cluster.html>.
- [10] P. A. Hidayat. (2021, April. 14). *Metode Penelitian Kualitatif*. [Online]. Available: <https://www.statistikian.com/2012/10/penelitian-kualitatif.html>.

- [11] BPS. (2021). *Rata-rata pengeluaran perkapita seminggu Menurut Sayur-sayuran (Rupiah/Kapita/Minggu), 2021 Provinsi Jawa Barat : Badan Pusat Statistik* [Online]. Available: <https://jabar.bps.go.id/indicator/5/790/1/rata-rata-pengeluaran-perkapita-seminggu-menurut-sayur--sayuran.html>.
- [12] M. Kuncoro, *Metode Riset untuk Bisnis dan Ekonomi*, Jakarta: Erlangga.