Jurnal Komputer, Informasi dan Teknologi Vol: 4, No 1, 2024, Page: 1-12

Comparison Of Population Prediction Of Bengkulu Selatan Regency With Python And SPSS

Heri Sulasono*, Dedy Abdullah, RG. Guntur Alam, Gunawan

Universitas Muhammadiyah Bengkulu

DOI: https://doi.org/10.53697/jkomitek.v4i1.1801
*Correspondence: Ryo Ricardo
Email: sulasonoheri@gmail.com

Received: 04-06-2024
Accepted: 06-06-2024
Published: 14-06-2024

Abstract: Bengkulu Selatan Regency is one of the oldest regencies in Bengkulu Province, with an area of 1,186.09 km² consisting of 11 sub-districts, 142 villages, and 16 urban-villages. Based on aggregate population data from the Population and Civil Registration Service of Bengkulu Selatan Regency, in 2023, the population was 174,936 people, consisting of 88,749 men and 86,142 women, with a population growth of 1,810 people. The population aspect is an important variable in determining the carrying capacity and capacity related to development and regional planning. In this study, researchers compared the population predictions of Bengkulu Selatan Regency using Python and the SPSS application. Population predictions use multiple linear regression analysis and are packaged in a web-based application using Python. This study uses the variables of birth, death, in-migration, and out-migration as independent variables and population as the dependent variable.The prediction results using Python and SPSS applications produce the same linear regression equation model to predict population size, namely $Y = 160036.652 - 1.346$ births + 2.048 deaths + 6.371 in-migration + 1.618 out-migration. The coefficient of determination is 0.982, indicating that the in-migration variable has a significant positive effect, while the other three variables do not have a significant effect.

Keywords: Prediction, Population, Linear Regression, Python, Spss

Introduction

South Bengkulu Regency is one of the oldest regencies in Bengkulu Province which was formed in 1949 since the division of Bengkulu Province from South Sumatra Province, currently South Bengkulu Regency is led by Gusnan Mulyadi, SE as the Regent of South Bengkulu(Mildawani et al., 2022). Based on aggregate population data from the Population and Civil Registry Office of South Bengkulu Regency in 2023, the population of South Bengkulu Regency is 174,936 people consisting of 88,749 men and 86,142 women with a population growth of 1,810 people or 0.29%. The amount of population growth can affect regional development planning so that appropriate analysis is needed (Spooner et al., 2020).

The regional medium-term development plan is carried out every 5 (five) years and evaluated every 1 (one) year. The evaluation of the development plan is based on socio-population analysis, in this case the population. Population surges can cause problems in

the future if not handled properly(Schafer et al., 2021). The population also affects the capacity and carrying capacity of an area. Thus, the p o p u l a t i o n data can be used as the main reference in considering the preparation of population projections that can provide an overview to the Regional Government in preparing the next planned and quality development plan(Mostafavi et al., 2020).

The results of previous research as support for this study include (Prawidana et al 2022) stating that the results of analysis and discussion of data mining using multiple linear regression methods about(Park et al., 2019).

The prediction of population growth rate in South Jakarta has an accurate calculation level as indicated by the Root Mean Square Error value of 0.43 < 1.0. According to (Widia et al, 2022) the results of estimating the amount of population growth using multiple linear regression methods are the right method for predicting population with a high level of accuracy and hope that the results of this study can be input to the Gunung Malela Sub-district Office in anticipating population growth rates(Boo & Choi, 2022). In a study conducted by Widia, et al, there was an increase in population in Gunung Malela District of 496 people throughout 2021- 2025. (Candra et al, 2023) state that the multiple linear regression method can provide an estimate of population increase up to an accuracy level of 94.5%. (Ayu Wulandari, et al 2022) stated that the multiple linear regression method can provide population estimation results with a high level of accuracy(Abdelaal et al., 2019), and suggested using the multiple linear regression method to be further used in calculations that make it easier for BPS to estimate the population rate(Islam et al., 2020). (Purwadi, et al 2019) stated that the analysis results obtained from data mining using multiple linear regression methods regarding population rate estimation can assist BPS in knowing what criteria can affect population growth rates and the multiple linear regression method can be implemented in predicting population growth rates with fairly accurate results(Rimal et al., 2019).

Based on previous research that has been presented, the researcher took the research title on the comparison of population prediction South Bengkulu Regency with python and SPSS applications. The title was taken as a form of continuation of previous research(Rasjid et al., 2021). This research will implement multiple linear regression analysis packaged in an application-based(Kammar-García et al., 2021).

Methodology

1. Problem Analysis

   Researchers identified problems related to the comparison of population prediction in South Bengkulu Regency using python and SPSS applications(Xu et al., 2021). The

results of problem identification at this stage are a description of the root of the problem regarding the population prediction system in South Bengkulu Regency and comparing the results of population prediction using python and SPSS applications (Alakus & Turkoglu, 2020). In addition, researchers made a description of the proposed solutions offered in this study(Wang et al., 2023).

2. Data Collection
   Researchers collected data related to population prediction. The data collection methods used are interviews, surveys, and literature studies(Vásquez-Morales et al., 2019).

3. Data Processing
   Researchers processed the data needed in further analysis, namely multiple linear regression tests on independent variables and related variables (Duffey & Zio, 2020).

4. Analysis
   The researcher conducted a linear regression analysis of the independent and dependent variables related to the prediction of population in South Bengkulu Regency. After conducting a linear regression test, variables that have a significant and insignificant effect on population prediction were obtained (Lebaz & Sheibat-Othman, 2019).

5. Inference
   The researcher compiled the results and discussion of the comparison of the prediction of the population of South Bengkulu Regency using python and SPSS applications (Yan et al., 2020).

6. Documentation
   This stage produces a research report that includes all activities that have been carried out by researchers(Ayati et al., 2021).

System Design

At this stage, researchers design web-based applications that implement the linear regression method in predicting population. The system design developed is flowchart and mock up. The system design is as follows.

Flowchart

The system design flow diagram is shown in Figure 3.1.

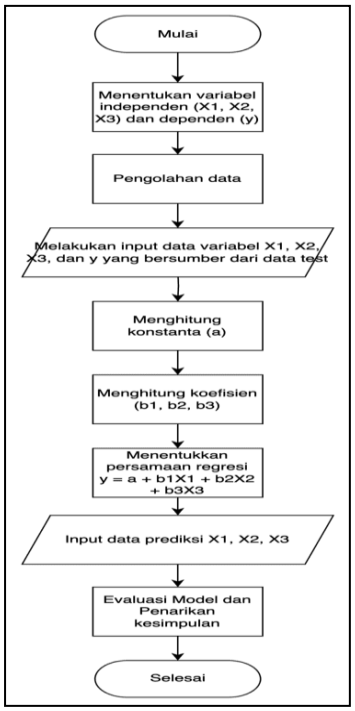

Figure 0.1. Linear Regression Flowchart

Result and Discussion

Implementation

In the implementation subchapter, there are several stages carried out, namely, data collection, data processing, regression modeling, testing results, and drawing conclusions. The stages that occur in this study are as follows(Afzal et al., 2023).

1) Data Collection

In the initial stage, it starts from preparing data that will be used as training data and test data in the formation of linear regression models. Population data to be analyzed is shown in Table 2.1. There are data columns for year, semester, population, birth, death, in-migration, and out-migration(Wannigamage et al., 2020). However, the data that will be used in multiple linear regression analysis is the total population (y) as the dependent variable, births (X1), deaths (X2), in-migration (X3), and out-migration (X4). Variable X is the independent variable(Hannon et al., 2023).

Table 1. Population Data to be Processed

Year	Smt	Total Population	Birth	Death	Incoming Movement	Move Out
2018	1	167003	396	42	1134	120
2018	2	168371	409	353	1156	647
2019	1	168936	418	366	1202	671

2019	2	169519	437	422	1218	724
2020	1	170028	523	462	1246	997
2020	2	170338	576	542	1260	1192
2021	1	170440	598	573	1284	1203
2021	2	170546	902	599	1296	1214
2022	1	170931	1234	612	1484	1231
2022	2	171806	1495	629	1690	1236
2023	1	173126	1719	643	1975	1241
2023	2	174936	2124	663	2165	1278

Source: Population and Civil Registration Office of South Bengkulu Regency (2024)

In this research, there are manual calculations, SPSS calculations, and using python library calculations so that the data is made in (He et al., 2021). csv format which contains birth, death, in-migration, out-migration, and total population variables(Belsher et al., 2019). Furthermore, researchers can use this data to conduct multiple linear regression analysis using several methods of manual calculation, SPSS application assistance, and using the python library(Jaidka et al., 2020).

2) Analysis

Calculation Using Manual or Ms. Excel

In the initial stages of manual calculation using Ms. Excel, researchers(Rabby et al., 2019) made a data filling table as an initial stage of creating a matrix in finding the regression equation shown in Table 2.2. The data that has been prepared is shown in Table 2.3(Keuning et al., 2020).

Table 2. Matrix Preparation

Y	**X1Y**	**X2Y**	**X3Y**	**X4Y**	**X1X2**	**X1X3**	**X1X4**	**X2X3**	**X2X4**	**X3X4**	**$X1^2$**	**$X2^2$**	**$X3^2$**	**$X4^2$**
							.....							
$\sum$ **Y**	$\sum$X1Y	$\sum$X2Y	$\sum$X3Y	$\sum$X4Y	$\sum$X1X2	$\sum$X1X3	$\sum$X1X4	$\sum$X2X3	$\sum$X2X4	$\sum$X3X4	$\sum X1^2$	$\sum X2^2$	$\sum X3^2$	$\sum X4^2$

Table 2 Matrix Preparation Data Filling

jml_population (Y)	birth (X1)	death (X2)	move_in (X3)	move_out (X4)
167003	396	42	1134	120
168371	409	353	1156	647
168936	418	366	1202	671
169519	437	422	1218	724
170028	523	462	1246	997
170338	576	542	1260	1192

170440	598	573	1284	1203
170546	902	599	1296	1214
170931	1234	612	1484	1231
171806	1495	629	1690	1236
173126	1719	643	1975	1241
174936	2124	663	2165	1278
∑Y	∑X1	∑X2	∑X3	∑X4
2045980	10831	5906	17110	11754

X1Y	X2Y	X3Y	X4Y
66133188	7014126	189381402	20040360
68863739	59434963	194636876	108936037
70615248	61830576	203061072	113356056
74079803	71537018	206474142	122731756
88924644	78552936	211854888	169517916
98114688	92323196	214625880	203042896
101923120	97662120	218844960	205039320
153832492	102157054	221027616	207042844
210928854	104609772	253661604	210416061
256849970	108065974	290352140	212352216
297603594	111320018	341923850	214849366
371564064	115982568	378736440	223568208
∑X1Y	∑X2Y	∑X3Y	∑X4Y
1859433404	1010490321	2924580870	2010893036

X1X2	X1X3	X1X4	X2X3	X2X4	X3X4
16632	449064	47520	47628	5040	136080
144377	472804	264623	408068	228391	747932
152988	502436	280478	439932	245586	806542
184414	532266	316388	513996	305528	881832
241626	651658	521431	575652	460614	1242262
312192	725760	686592	682920	646064	1501920
342654	767832	719394	735732	689319	1544652
540298	1168992	1095028	776304	727186	1573344
755208	1831256	1519054	908208	753372	1826804
940355	2526550	1847820	1063010	777444	2088840
1105317	3395025	2133279	1269925	797963	2450975
1408212	4598460	2714472	1435395	847314	2766870
∑X1X2	∑X1X3	∑X1X4	∑X2X3	∑X2X4	∑X3X4
6144273	17622103	12146079	8856770	6483821	17568053

$X1^2$	$X2^2$	$X3^2$	$X4^2$
156816	1764	1285956	14400
167281	124609	1336336	418609
174724	133956	1444804	450241
190969	178084	1483524	524176
273529	213444	1552516	994009
331776	293764	1587600	1420864
357604	328329	1648656	1447209
813604	358801	1679616	1473796
1522756	374544	2202256	1515361
2235025	395641	2856100	1527696
2954961	413449	3900625	1540081
4511376	439569	4687225	1633284
$\sum X1^2$	$\sum X2^2$	$\sum X3^2$	$\sum X4^2$
13690421	3255954	25665214	12959726

After filling in the data for the next matrix creation, researchers used the following formula to create the A matrix and H matrix shown in Table 2.4 and Table 2.5. The n is the amount of data analyzed.

$$A = \begin{bmatrix} n & \sum X1 & \sum X2 & \sum X3 & \sum X4 \\ \sum X1 & \sum X1^2 & \sum X1X2 & \sum X1X3 & \sum X1X4 \\ \sum X2 & \sum X1X2 & \sum X2^2 & \sum X2X3 & \sum X2X4 \\ \sum X3 & \sum X1X3 & \sum X1X3 & \sum X3^2 & \sum X3X4 \\ \sum X4 & \sum X1X4 & \sum X1X & \sum X3X4 & \sum X4^2 \end{bmatrix}$$

$$H = \begin{bmatrix} \sum Y \\ \sum X1Y \\ \sum X2Y \\ \sum X3Y \\ \sum X4Y \end{bmatrix}$$

Table 3. Matrix A

Matriks A				
12	10831	5906	17110	11754
10831	13690421	6144273	17622103	12146079
5906	6144273	3255954	8856770	6483821
17110	17622103	8856770	25665214	17568053
11754	12146079	6483821	17568053	12959726

Table 1. Matrix H

H

2045980
1859433404
1010490321
2924580870
2010893036

Furthermore, the researcher forms the A1 - A5 matrix which is used to find the determinant value, coefficient value and determine the equation formula. The A1-A5 matrix is shown in Table 6.

Table 2. Matriks A1, A2, A3, A4, A5

Matriks A1				
2045980	10831	5906	17110	11754
1859433404	13690421	6144273	17622103	12146079
1010490321	6144273	3255954	8856770	6483821
2924580870	17622103	8856770	25665214	17568053
2010893036	12146079	6483821	17568053	12959726
Matriks A2				
12	2045980	5906	17110	11754
10831	1859433404	6144273	17622103	12146079
5906	1010490321	3255954	8856770	6483821
17110	2924580870	8856770	25665214	17568053
11754	2010893036	6483821	17568053	12959726
Matriks A3				
12	10831	2045980	17110	11754
10831	13690421	1859433404	17622103	12146079
5906	6144273	1010490321	8856770	6483821
17110	17622103	2924580870	25665214	17568053
11754	12146079	2010893036	17568053	12959726
Matriks A4				
12	10831	5906	2045980	11754
10831	13690421	6144273	1859433404	12146079
5906	6144273	3255954	1010490321	6483821
17110	17622103	8856770	2924580870	17568053
11754	12146079	6483821	2010893036	12959726
Matriks A5				
12	10831	5906	17110	2045980
10831	13690421	6144273	17622103	1859433404
5906	6144273	3255954	8856770	1010490321
17110	17622103	8856770	25665214	2924580870

11754	12146079	6483821	17568053	2010893036

A2 Matrix

12	2045980	5906	17110	11754
10831	1859433404	6144273	17622103	12146079
5906	1010490321	3255954	8856770	6483821
17110	2924580870	8856770	25665214	17568053
11754	2010893036	6483821	17568053	12959726

A3 Matrix

12	10831	2045980	17110	11754
10831	13690421	1859433404	17622103	12146079
5906	6144273	1010490321	8856770	6483821
17110	17622103	2924580870	25665214	17568053
11754	12146079	2010893036	17568053	12959726

A4 Matrix

12	10831	5906	2045980	11754
10831	13690421	6144273	1859433404	12146079
5906	6144273	3255954	1010490321	6483821
17110	17622103	8856770	2924580870	17568053
11754	12146079	6483821	2010893036	12959726

A5 Matrix

12	10831	5906	17110	2045980
10831	13690421	6144273	17622103	1859433404
5906	6144273	3255954	8856770	1010490321
17110	17622103	8856770	25665214	2924580870
11754	12146079	6483821	17568053	2010893036

Furthermore, researchers used the matrix to calculate the determinant value of each matrix shown in Table 2.7. The determinant value is obtained by using the MDETERM(matrix) formula in excel. The selected matrix is adjusted for each determinant value as an example det(A1) using matrix A1 and so on.

Conclusion

Multiple linear regression analysis is the right tool to determine how much influence the independent variable has on the dependent variable. In this study, a conclusion was obtained that the linear regression equation formula obtained was feasible to use in predicting population size. Based on the coefficient of determination obtained, it is 98.2%, which means that as much as 98.2% of the four independent variables (birth, death, in-migration, and out-migration) have influenced the population variable and 1.8% is influenced by other variables. However, if tested partially or individually, it is found that the variables of births and in-movements have a significant influence on the total population variable. The linear regression equation formula obtained for the calculation of population prediction is Y = 160036.652 + -1.346 births + 2.048 deaths + 6.371 in-migration + 1,618 out-migration.

Based on the results of the comparison of the prediction of the population of South Bengkulu Regency using python and the SPSS application and t h e added manual calculation method, it is found that the resulting multiple linear regression model is the same. The difference lies in the level of detail of the numbers behind the comma. Overall, the coefficient and constant values are similar in each method both using the python library, SPSS application, and manual calculation.

References

Abdelaal, T., Michielsen, L., Cats, D., Hoogduin, D., Mei, H., & ... (2019). A comparison of automatic cell identification methods for single-cell RNA sequencing data. Genome Biology. https://doi.org/10.1186/s13059-019-1795-z

Afzal, S., Ziapour, B. M., Shokri, A., Shakibi, H., & Sobhani, B. (2023). Building energy consumption prediction using multilayer perceptron neural network-assisted models; comparison of different optimization algorithms. Energy. https://www.sciencedirect.com/science/article/pii/S0360544223018406

Alakus, T. B., & Turkoglu, I. (2020). Comparison of deep learning approaches to predict COVID-19 infection. Chaos, Solitons &Fractals. https://www.sciencedirect.com/science/article/pii/S0960077920305178

Ayati, N., Lee, S. T., Zakavi, S. R., Cheng, M., & ... (2021). Response evaluation and survival prediction after PD-1 immunotherapy in patients with non–small cell lung cancer: comparison of assessment methods. Journal of Nuclear …. https://jnm.snmjournals.org/content/62/7/926.abstract

Belsher, B. E., Smolenski, D. J., Pruitt, L. D., Bush, N. E., & ... (2019). Prediction models for suicide attempts and deaths: a systematic review and simulation. JAMA …. https://jamanetwork.com/journals/jamapsychiatry/article-abstract/2727387

Boo, Y., & Choi, Y. (2022). Comparison of mortality prediction models for road traffic accidents: an ensemble technique for imbalanced data. BMC Public Health. https://doi.org/10.1186/s12889-022-13719-3

Duffey, R. B., & Zio, E. (2020). Prediction of CoVid-19 infection, transmission and recovery rates: A new analysis and global societal comparisons. Safety Science. https://www.sciencedirect.com/science/article/pii/S0925753520302514

Hannon, M. P., Carney, D. J., Floyd, S., Parker, L. J. F., & ... (2023). Cross-sectional comparison of body composition and resting metabolic rate in Premier League academy soccer players: Implications for growth and maturation. Science and …. https://doi.org/10.4324/9781003375968-14

He, Y., Lakhani, C. M., Rasooly, D., Manrai, A. K., & ... (2021). Comparisons of polyexposure, polygenic, and clinical risk scores in risk prediction of type 2 diabetes. Diabetes …. https://diabetesjournals.org/care/article-abstract/44/4/935/138586

Islam, M. M., Haque, M. R., Iqbal, H., Hasan, M. M., & ... (2020). Breast cancer prediction: a comparative study using machine learning techniques. SN Computer …. https://doi.org/10.1007/s42979-020-00305-w

Jaidka, K., Giorgi, S., Schwartz, H. A., & ... (2020). Estimating geographic subjective well-being from Twitter: A comparison of dictionary and data-driven language methods. Proceedings of the …. https://doi.org/10.1073/pnas.1906364117

Kammar-García, A., Castillo-Martínez, L., & ... (2021). Comparison of bioelectrical impedance analysis parameters for the detection of fluid overload in the prediction of mortality in patients admitted at the emergency …. Journal of Parenteral …. https://doi.org/10.1002/jpen.1848

Keuning, B. E., Kaufmann, T., Wiersema, R., & ... (2020). Mortality prediction models in the adult critically ill: A scoping review. Acta …. https://doi.org/10.1111/aas.13527

Lebaz, N., & Sheibat-Othman, N. (2019). A population balance model for the prediction of breakage of emulsion droplets in SMX+ static mixers. Chemical Engineering Journal. https://www.sciencedirect.com/science/article/pii/S1385894718325762

Mildawani, M., Murti, T. R., Maryatmi, A. S., & Abraham, J. (2022). … model of competitive behavior: social comparison as a mediator of the critical thinking, self-efficacy, and adaptation ability prediction among college students. Heliyon. https://www.cell.com/heliyon/fulltext/S2405-8440(22)03493-4

Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., & ... (2020). Variable prediction accuracy of polygenic scores within an ancestry group. Elife. https://elifesciences.org/articles/48376

Park, K. M., Sung, J. M., Kim, W. J., An, S. K., Namkoong, K., & ... (2019). Population-based dementia prediction model using Korean public health examination data: A cohort study. PloS One. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0211957

Rabby, A., Mamata, R., Laboni, M. A., & ... (2019). Machine learning applied to kidney disease prediction: Comparison study. 2019 10th …. https://ieeexplore.ieee.org/abstract/document/8944799/

Rasjid, Z. E., Setiawan, R., & Effendi, A. (2021). A comparison: prediction of death and infected COVID-19 cases in Indonesia using time series smoothing and LSTM neural network. Procedia Computer Science. https://www.sciencedirect.com/science/article/pii/S1877050921001368

Rimal, R., Almøy, T., & Sæbø, S. (2019). Comparison of multi-response prediction methods. Chemometrics and Intelligent Laboratory …. https://www.sciencedirect.com/science/article/pii/S016974391930187X

Schafer, K. M., Kennedy, G., Gallyer, A., & Resnik, P. (2021). A direct comparison of theory-driven and machine learning prediction of suicide: A meta-analysis. PloS One. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0249833

Spooner, A., Chen, E., Sowmya, A., Sachdev, P., & ... (2020). A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction. Scientific Reports. https://www.nature.com/articles/s41598-020-77220-w

Vásquez-Morales, G. R., Martinez-Monterrubio, S. M., & ... (2019). Explainable prediction of chronic renal disease in the colombian population using neural networks and case-based reasoning. Ieee …. https://ieeexplore.ieee.org/abstract/document/8877828/

Wang, Z., Zhao, G., Li, B., Fang, Z., Chen, Q., & ... (2023). Performance comparison of computational methods for the prediction of the function and pathogenicity of non-coding variants. Genomics …. https://academic.oup.com/gpb/article-abstract/21/3/649/7588917

Wannigamage, D., Barlow, M., Lakshika, E., & ... (2020). Analysis and prediction of player population changes in digital games during the COVID-19 pandemic. … Joint Conference on …. https://doi.org/10.1007/978-3-030-64984-5_36

Xu, D., Jiang, M., Hu, W., Li, S., Pan, R., & ... (2021). An online prediction approach based on incremental support vector machine for dynamic multiobjective optimization. IEEE Transactions on …. https://ieeexplore.ieee.org/abstract/document/9546930/

Yan, B., Tang, X., Liu, B., Wang, J., Zhou, Y., Zheng, G., & ... (2020). An improved method for the fitting and prediction of the number of covid-19 confirmed cases based on lstm. ArXiv Preprint ArXiv …. https://arxiv.org/abs/2005.03446