



# Trip Generation and Trip Distribution Analysis of Commuter Line

Ega Cahyani Putri<sup>1</sup>(✉), Nanda Ahda Imron<sup>1</sup>, and Nanda Sani Mitra Imron<sup>2</sup>

<sup>1</sup> Indonesian Railway Polytechnic, Tirta Raya Street, Manguharjo, Madiun 63161, Indonesia  
ega.mtp1911@taruna.ppi.ac.id

<sup>2</sup> PT. Kereta Api Indonesia (KAI), Jl. Perintis Kemerdekaan No.1Kec. Sumur Bandung Jawa Barat, Babakan Ciamis, Kota Bandung 40117, Indonesia

**Abstract.** Greater Bandung Commuter line (CL) is expected to operate after the double-track construction and electrification in Padalarang-Cicalengka route in 2024 is completed. Among the route, there is Gedebage Station, classified as “operational station” (with no passenger that board or alight from), in which its function will be enhanced as passenger station. In order to meet the public’s need of transport service, it is important to analyse the trip generation and trip distribution. In March 2022, home interview survey was conducted in residential area near Gedebage Station to obtain the origin-destination data and combined with origin-destination data of the existing trains. ARIMA method is carried out to forecast trip generation, as Fratar method is to analyze its trip distribution. The study concludes, the best model for forecasting is AR (1) with biggest R-square value, altogether with smallest Aikake Information Criterion (AIC) and Schwarz Criterion (SC). The highest trip distribution obtained is trip from Bandung Station to Cicalengka Station.

**Keywords:** Trip Generation · Trip Distribution · ARIMA · Fratar

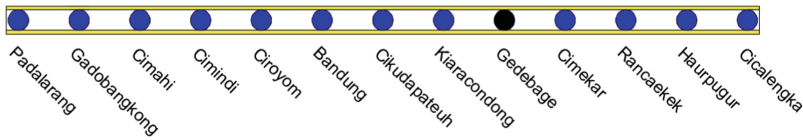
## 1 Introduction

The constant needs of transportation will be increasing transportation users and be raising congestion [1]. As an alternative to solve traffic problems, CL is going to be operated in Bandung area, referring to Indonesia Railway Masterplan (RIPNas) 2030 in the matter of Double-Track Construction and Electrification in the route of Padalarang-Bandung-Cicalengka [2] that is expected to complete by the year 2024.

In the aforementioned route, Gedebage Station as a operational station is going to be revitalized into a modern passenger station [2]. So that the transportation service be able to fulfill community needs, it is important to obtained forecast on the sum of trip generation and its trip distribution.

## 2 Research Method

Trip data is collected from data recorded by the Operation Area (DAOP) II Bandung, combined with the results of home interview survey in the area of Gedebage Station,



**Fig. 1.** CL Route Padalarang-Cicalengka

within 500 m radius. It is considered so as Gedebage Station is designated, in the spatial planning, as a centralized area for integrated intermodal, covering in range of 400–800 m from the mass transportation transit node [3]. In home interview survey, the number of populations is calculated through manual count on the roof number in *Google Earth Pro* apps and it resulted a total of 1.520 houses. Using purposive sampling, the obtained sample is 210 respondents. The indicators used during the interview including the number of the family member(s), vehicle ownership, family identity, and the family member(s)' trip. From DAOP II, the data obtained including passengers' trip (origin) who accessed Padalarang-Cicalengka route train during January 2019-March 2022.

Interview instrument is analyzed using Lawshe's Content Validity Ratio (CVR) that has been used to examine contents' validity [3]. Each respondent answered questions for every item with three answer options, namely (1) necessary, (2) less necessary, and (3) unnecessary [4]. CVR is calculated with this following formula [3]:

$$\text{CVR} = \frac{(n_e - \frac{N}{2})}{\frac{N}{2}} \quad (1)$$

CVR: Content Validity Ratio

$n_e$  : The number of respondents indicating an item as necessary

$N$  : The total number of respondents

The data obtained from the survey then accumulated with the trip generation data from DAOP II, to be analyzed using ARIMA method through *EViews* software. ARIMA model calculates AR (autoregressive) and MA (moving average), with I (integrated) that is either stationary or differential. ARIMA model using order  $p$  to indicate the value of AR, order  $d$  to indicate the differential leveling, and order  $q$  to indicate the value of MA [6].

#### 1. Autoregressive AR ( $p$ ) model

$$X_t = \mu' + \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + \dots + \emptyset_p X_{t-p} + e_t \quad (2)$$

$\mu'$ : A constant.

$\emptyset_p$ : Autoregressive parameter at  $p$ .

$X_{t-1}$ : Error value at  $t$ .

$X_t$ : General average.

#### 2. Moving Average MA ( $q$ ) model

$$X_t = \bar{x} + e_t - \emptyset_1 e_{t-1} - \emptyset_2 e_{t-2} - \dots + \emptyset_q e_{t-q} \quad (3)$$

$\hat{\imath}$ : Constant.

$\hat{e}_1 - \hat{e}_1$ : Moving average parameter.

$e_{t-k}$ : Error value at  $t-k$ .

$X_t$ : General average.

### 3. Autoregressive moving average ARMA ( $p, q$ ) model

$$X_t = \hat{\imath} + \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + \dots + \emptyset_p X_{t-p} + e_t + \hat{e}_1 e_{t-1} - \hat{e}_2 e_{t-2} - \dots + \hat{e}_q e_{t-q} \quad (4)$$

$X_t$ : General average.

$\emptyset_p$ : Autoregressive parameter at  $p$ .

$\hat{e}_q$ : Autoregressive parameter at  $q$ .

$e_t$ : Error value given at  $t$ .

$e_{t-q}$ : Error value given at  $t-q$ .

$\hat{\imath}$ : A constant.

After trip generation value on the planned year generated, the trip distribution then analyzed through Fratar method. The Fratar method is analyzed using repetition process or iteration using this following equation [7].

$$Ti - j = \frac{Ti(G).ti - j.Ej}{ti - j.Ej + ti - kEk + \dots + ti - nEn} \quad (5)$$

$Ti - j$ : Trip generation forecast from  $i$  to  $j$ .

$Ti(G)$ : The sum of trip in the target year.

$ti - j \dots ti - k \dots ti - n$ : The sum of trip in the current period from the origin to destination zone.

$Ej \dots Ek \dots En$ : Growth factor.

## 2.1 Home Interview

Following interview to the community lived near Gedebage Station, it is identified that the average number of family members is 3,6 per family, 54% of the respondents are male and the 46% other are female, and social activity became the dominant reason to take a train trip. Data of trip destination collected from the interview was daily sample data (trip/day), thus it has to be multiplied with expansion factor (FE) to convert it into daily population data (trip/day). Expansion factor is calculated using this following formula [8]:

$$FE = \frac{\left\{ A - \frac{A}{B} * (C + D) \right\}}{B - C - D} \quad (6)$$

$$FE = \frac{\left\{ 1520 - \frac{1520}{210} * (1520 - 210) + 0 \right\}}{210 - (1520 - 210) - 0} = 7,238095238$$

A refers to the sum of all buildings or families, B refers to the number of buildings chosen as sample, C refers to the remaining buildings that are not included as sample, and D refers to the number of buildings that were empty and did not respond or refused being sample. Daily population data (trip/day) then is converted into monthly population data (trip/month) and is resulted in 638,617 trip for its total trip generation.

### 3 Trip Generation Forecasting Using ARIMA Method

Trip generation forecasting of Greater Bandung CL passengers using time series data of trip during Januari 2019 – March 2022 (as collected from DAOP II) that is deriving out of combined origin-destination (O/D) matrix from survey result and data collected by DAOP II. It is known that during 2020–2021, the world is facing COVID-19 pandemic and travel restriction was implemented, resulting in the significant decrease on trip. Even in year 2022, travel requirement is still applied. These factors, thus, will affect the result of forecasting.

#### 3.1 Stationarity Test

As mentioned in the graph, the data tends to avoid the average value thus it can be concluded that the data has not been stationer. The result on correlogram test and unit root test is shown that the data has also not been stationer to the Level stage, so it demands differencing [5]. On the first stage of correlogram differencing test, cited that the value of Q-Statistic Ljung-Box on the last lag is 16,975 smaller that the chi square ( $\chi^2$ ) critical value while the df value is 39 in alpha ( $\alpha$ ) 5% that is 54.572; thus, the data in the first phase differential is considered stationer.

On the unit root test developed by Dicky-Fuller, it has  $H_1$  hypothesis that if  $|t\text{-count}| > |t\text{-table}|$  meaning the data is stationer, while  $H_0$   $|t\text{-count}| < |t\text{-table}|$  meaning the data has not been stationer [9].

The value of *Augmented Dickey-Fuller* (ADF) is absolute, *t-statistic* value in -4.679782 is bigger than value shown in each t-table that reached -3.621023 (1% level), -2.943427 (5% level), and -2.610263 (10% level). It is concluded that the data has been stationer in the first stage of the different [9].

#### 3.2 Determining the Model

There are several methods to achieve ARIMA parameters, including the following:

1. The order chosen based on the highest value and  $d$  value refers to its differential stage [5].
2. Looking to the ACF and PACF pattern that reached the lag peak and  $d$  value is in accordance with the differential stage [10].
3. Trial and error, by testing a number of different values then choosing one value (or several values) that will be minimizing the sum of square residual [11].
4. Looking at the ACF and PACF pattern to find out the tally on correlogram that is cut off or is out of the bartlett, with  $d$  value that is in accordance with the differential stage.

Table 1. Matrix of combined existing trip data and survey result

O/D	PDL	GK	CMI	CMD	CIR	BD	CTH	KAC	GDB	CMK	RCK	HRP	CCL	Pi
PDL	-	78	1,714	2,886	2,406	21,848	3,029	7,569	2,917	2,616	4,018	405	11,747	61,233
GK	102	-	39	165	457	4,519	830	2,309	1,346	598	992	540	5,252	17,149
CMI	1,919	93	-	49	250	5,153	978	4,334	2,468	1,061	2,756	423	9,318	28,802
CMD	4,526	124	110	-	37	954	417	2,688	1,795	924	2,696	391	9,517	24,179
CIR	5,609	360	295	227	-	489	295	1,112	1,346	499	2,780	456	12,671	26,139
BD	23,059	2,985	4,369	636	56	-	507	2,735	9,873	3,983	18,980	2,213	50,924	120,320
CTH	4,590	827	1,090	396	84	91	-	863	3,814	1,071	5,968	938	11,934	31,666
KAC	11,739	1,641	4,334	1,928	698	5,514	77	-	6,507	956	9,948	1,538	26,774	71,654
GDB	4,263	1,346	2,917	2,693	1,795	9,648	3,814	6,956	-	2,917	3,366	1,795	3,590	45,101
CMK	2,957	574	1,138	938	695	5,744	1,025	2,234	1,795	-	598	197	2,714	20,609
RCK	5,408	868	3,004	2,484	2,590	32,512	4,244	8,433	2,019	2,290	-	228	1,483	65,563
HRP	2,251	123	469	557	1,048	8,092	605	2,510	1,795	348	1,483	-	675	19,956
CCL	13,517	2,714	6,622	5,252	7,002	43,987	6,011	14,422	3,366	2,246	913	193	-	106,245
Aj	79,940	11,733	26,101	18,211	17,118	138,551	21,832	56,165	39,042	19,509	54,498	9,317	146,599	638,617

Notes:

- PDL = Padalarang Station

GK = GadoBangkong Station

CMI = Cimahi Station

CMD = Cimindi Station

RCK = Rancaekek Station
- CIR = Ciroyom Station

BD = Bandung Station

CTH = Cikudapateuh Station

HRP = Haurpugur Station
- KAC = Kiaracondong Station

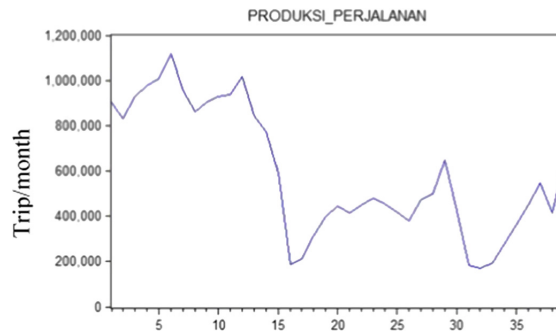
GDB = Gedebage Station

CMK = Cimekar Station

CCL = Cicalengka Station

**Table 2.** Trip Generation

Month	Period	Trip production	Month	Period	Trip production	Month	Period	Trip production
1	Jan-19	900,224	14	Feb-20	771,944	27	Mar-21	472,618
2	Feb-19	829,511	15	Mar-20	589,606	28	Apr-21	499,841
3	Mar-19	929,241	16	Apr-20	189,370	29	May-21	647,295
4	Apr-19	977,357	17	May-20	208,614	30	Jun-21	423,325
5	May-19	1,005,657	18	Jun-20	310,890	31	Jul-21	185,231
6	Jun-19	1,117,691	19	Jul-20	395,282	32	Aug-21	171,262
7	Jul-19	952,989	20	Aug-20	445,695	33	Sep-21	191,827
8	Aug-19	859,791	21	Sep-20	412,239	34	Oct-21	277,725
9	Sep-19	902,224	22	Oct-20	447,829	35	Nov-21	362,195
10	Oct-19	928,501	23	Nov-20	480,087	36	Dec-21	449,154
11	Nov-19	934,576	24	Dec-20	452,721	37	Jan-22	546,723
12	Dec-19	1,015,798	25	Jan-21	420,595	38	Feb-22	413,057
13	Jan-20	844,921	26	Feb-21	379,640	39	Mar-22	638,617

**Fig. 2.** Trip generation pattern (monthly)

By applying the four aforementioned methods, the best model is acquired through trial and error method. The most suitable method is chosen by considering the probability value that is smaller than alpha ( $\alpha$ ) 5% and based on goodness of fit of each parameter. Goodness of fit has several criteria; those are the biggest R-square value and the smallest value of Akaike Information Criterion (AIC) and Schwarz Criterion (SC) [5].

**Table 3.** First stage of ADF Test

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.569750	0.4879
Test critical values:	1% level	-3.615588	
	5% level	-2.941145	
	10% level	-2.609066	

\* MacKinnon (1996) one-sided p-values.

**Table 4.** Parameter estimate

Model	Variable	Coeff.	AIC	SC	R <sup>2</sup>	Prob.
ARIMA(1,0,0)	C AR(1)	645069.9 0.897424	26.40444	26.53241	0.818782	0.0013 0.0000

ARIMA model (1,0,0) or AR (1) is the most suitable model in forecasting trip generation of Greater Bandung CL passengers. AR model (1) is chosen with the following criteria: the biggest R-square value at 0.818782; Akaike Information Criterion (AIC) value at 26.40444 and Schwarz Criterion (SC) value at 26.53241, which are the smallest amongst the other model.

**3.3 Diagnostic Test/Model Evaluation**

The test on diagnose residual is held to examine whether the estimated residual result has been qualified as white noise or random, by analyzing its residual value through correlogram ACF and PACF [5]. This following table displays the result of residual model test.

During the last lag, the value of Q-Statistic Ljung-Box shown at 14.167, that is smaller than the critical value of *chi square* ( $\chi^2$ ) with df at 39 in alpha ( $\alpha$ ) 5% that is given at 54.572. This means, the residual value estimated has been qualified as white noise or random; thus, the model is considered good [5].

**Table 5.** Diagnostic test

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
.  *.	.  *.	1	0.183	0.183	1.4069
.  .	.  .	2	0.030	-0.003	1.4467
.  .	.  .	3	-0.054	-0.061	1.5760
**  .	**  .	4	-0.253	-0.241	4.5077
.  .	.  *.	5	-0.004	0.093	4.5083
.  .	.  .	6	-0.018	-0.027	4.5239
.  .	.  .	7	-0.008	-0.024	4.5272
.  *.	.  .	8	0.090	0.040	4.9414
.  .	.  .	9	0.004	0.003	4.9423
.  .	.  .	10	-0.042	-0.064	5.0393
.  *.	.  *.	11	-0.144	-0.144	6.2172
.  *.	.  *.	12	-0.179	-0.105	8.1128
**  .	**  .	13	-0.226	-0.208	11.262
.  *.	.  *.	14	0.096	0.167	11.851
.  *.	.  .	15	0.143	0.049	13.206
.  *.	.  .	16	0.118	0.018	14.167

### 3.4 Forecasting

In the following step, the forecasting is done using *EViews* 9.0 software for the period when Greater Bandung CL will be operating on December 2024, that is given at 645,187 trips and forecasting for the following five-years period or up to December 2027 at 645,072 trips. The forecasting results are displayed in the following tables. The decrease of trip generation on December 2027 period happens, since the trip data that shown significant decrease during pandemic as a consequence of the travel restriction.

## 4 Trip Distribution Analysis Using Fratar Method

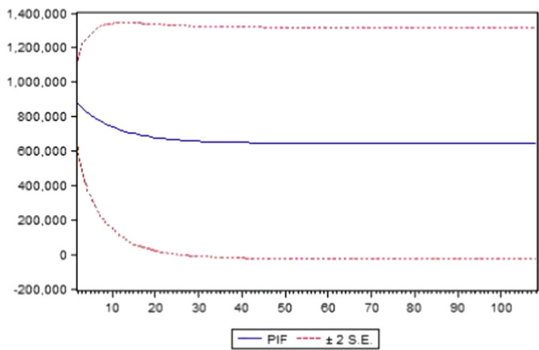
After the value of production trip in a month period obtained, then the distribution of passengers' trip for Padalarang-Cicalengka route is calculated using Fratar method [7]. The growth level must be achieved at  $E_i = 1$  through iteration so that the analysis is considered as complete [13].

The trip distribution for December 2024 period, when the CL is expected to be operated, the value of  $E_i$  has not reached 1; thus, iteration is applied in order to reach  $E_i$  value at 1.00.  $E_i$  value reached 1 during the 38th iteration, meanwhile the highest trip distribution from Bandung Station to Cicalengka Station is at 38,895 trips.

The trip distribution for CL passengers for the next five-years period or up to December 2027 is also calculated. For the aforementioned condition, the  $E_i$  value has not reached 1, so that iteration is needed.  $E_i$  reached 1 on its 37th iteration and the highest trip distribution from Bandung Station to Cicalengka Station is at 38,888 trips.

**Table 6.** Forecasting

Period	Pi	Period	Pi	Period	Pi	Period	Pi
Apr-22	648,817	Oct-23	645,604	Apr-25	645,146	Oct-26	645,081
May-22	648,433	Nov-23	645,549	May-25	645,138	Nov-26	645,080
Jun-22	648,088	Dec-23	645,500	Jun-25	645,131	Dec-26	645,079
Jul-22	647,778	Jan-24	645,456	Jul-25	645,125	Jan-27	645,078
Aug-22	647,501	Feb-24	645,416	Aug-25	645,119	Feb-27	645,077
Sep-22	647,251	Mar-24	645,381	Sep-25	645,114	Mar-27	645,076
Oct-22	647,027	Apr-24	645,349	Oct-25	645,110	Apr-27	645,076
Nov-22	646,827	May-24	645,320	Nov-25	645,106	May-27	645,075
Dec-22	646,646	Jun-24	645,295	Dec-25	645,102	Jun-27	645,074
Jan-23	646,485	Jul-24	645,272	Jan-26	645,099	Jul-27	645,074
Feb-23	Feb-23	Aug-24	645,251	Feb-26	645,096	Aug-27	645,074
Mar-23	Mar-23	Sep-24	645,232	Mar-26	645,093	Sep-27	645,073
Apr-23	Apr-23	Oct-24	645,216	Apr-26	645,091	Oct-27	645,073
May-23	May-23	Nov-24	645,201	May-26	645,089	Nov-27	645,073
Jun-23	Jun-23	Des-24	645,187	Jun-26	645,087	Des-27	645,072
Jul-23	Jul-23	Jan-25	645,175	Jul-26	645,085		
Aug-23	Aug-23	Feb-25	645,604	Aug-26	645,083		
Sep-23	Sep-23	Mar-25	645,155	Sep-26	645,082		



**Fig. 3.** Forecasting Pattern





## 5 Conclusion

1. Based on the prior time series data, this study concluded that ARIMA model (1,0,0) or AR (1) is the best model to forecast trip generation CL passengers' for Padalarang-Cicalengka route in the future
2. The number of trip generation for December 2024, when CL will be operated, is at 645,187 trips in a month, while the number of trip generation for December 2027 is at 645,072 trips in a month.
3. In December 2024, the highest trip distribution from Bandung Station to Cicalengka Station is 38,895 trips. Meanwhile, for December 2027, the highest trip distribution from Bandung Station to Cicalengka Station is at 38,895 trips.

## References

1. Alfani, A. F., Mujib, M. A., & Ikhsan, F. A. (2020). Tingkat Kemacetan Dan Realita Transportasi Di Jalan Letjen Suprpto (*The Congestion Rate and Transportation Reality at Jalan Letjen Suprpto*). Sosearch: Social Science Educational Research, 1(1).
2. Sujadi, A. (2021). *Menanti kehadiran KRL Bandung Raya (Waiting for Greater Bandung CL)*. 25 Juli 2021, accessed from <https://gpsindonesia.co/opini/read/1642/Menanti-Kehadiran-KRL-Bandung-Raya.html> pada 3 April 2022.
3. Fatmadiwi A, Hairida H, Sartika RP, Melati HA, Rasmawan R. Pengembangan Video Pembelajaran pada Konsep Asesmen Autentik untuk Mahasiswa (*Developing Learning Video for Authentic Assessment Concept for Scholar*). *Edukatif J Ilmu Pendidik*. 2021;4(1):266-277. doi:<https://doi.org/10.31004/edukatif.v4i1.1322>.
4. Nengsih NR, Yusmaita E, Gazali F. Evaluasi Validitas Konten dan Konstruksi Bahan Ajar Asam Basa Berbasis REACT (*The Evaluation on Content and Learning Material Construct Validity on Acid-Alkali Material Based on REACT*). *EduKimia*. 2019;1(1):1-10. doi:<https://doi.org/10.24036/ekj.v1i1.104017>
5. Teapon RRH. Pembentukan Model ARIMA Untuk Peramalan Inflasi Kelompok Bahan Makanan Di Kota Ternate (*Developing ARIMA Model to Forecast Inflation for Groups of Food and Ingredients in Ternate*). *Ekonomi*. 2015;04(01):13-30.
6. Djami RJ, Nanlohy YWA. Peramalan Indeks Harga Konsumen Di Kota Ambon Menggunakan Autoregressive Integrated Moving Average (ARIMA) Dan Double Exponential Smoothing (*Forecasting Index of Consumer Price in Ambon Using Autoregressive Integrated Moving Average (ARIMA) and Double Exponential Smoothing*). 2022;4:1–14.
7. Miro F, Mukhtim VP. Analisis Sebaran Perjalanan Pada Empat Zona Dalam Wilayah Administrasi Kota Solok (*The Analysis of Trip Distribution in Four Zones in Administrative Region of Solok*). 2022;19(1):1-10.
8. Sekolah Tinggi Transportasi Darat. Pedoman Pelaksanaan Praktek Kerja Lapangan (*The Guideline of Professional Placement Practice*). 2011: III-9.
9. Anggraini TB, Yefriza Y. Nilai Tukar Rupiah Dan Net Ekspor Indonesia 2000 – 2017 (Granger Causality Test) (*Exchange Rate of Rupiah and Indonesia Net Export 2000 – 2017 (Granger Casuality Test)*). *Converg J Econ Dev*. 2020;1(1):9-24. Doi:<https://doi.org/10.33369/Convergence-Jep.V1i1.10854>.
10. Wigati Y, Rais R, Utami IT. Pemodelan Time Series Dengan Proses ARIMA Untuk Prediksi Indeks Harga Konsumen (Ihk) Di Palu – Sulawesi Tengah (*Time Series Modeling Using Arima Process to Forecast Consumer Price Index (CPI) in Palu – Central Sulawesi*). *J Ilm Mat Dan Terap*. 2017;12(2):149-159. Doi:<https://doi.org/10.22487/2540766x.2015.V12.I2.7908>.

11. Hendro Waryanto Dan Dwi Aprisa Wanti. (Jurnal Statistika Dan Matematika). 2019;1(1):88–102.
12. La Murdani AI, Nanlohy YWA. Implementasi Model Autoregressive Integrated Moving Average (ARIMA) Untuk Peramalan Jumlah Penumpang Kapal Laut Di Pelabuhan Ambon (*Implementing Autoregressive Integrated Moving Average (ARIMA) to Forecast The Number of Ship Passenger in Ambon Harbor*). *Var J Stat Its Appl*. 2022;3(2):81-90. Doi:<https://doi.org/10.30598/Variancevol3iss2page81-90>.
13. Dahriansyah R, Novita N, Agustiah P. Permodelan Pengembangan Angkutan Pelajar Khusus SMP Negeri di Kota Pontianak (*Modeling and Developing Public Junior High School Students' Transport in Pontianak*). Published online 2018:1–10.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

