Indonesian Journal of Contemporary Multidisciplinary Research (MODERN)
Vol.2, No.5, 2023: 839-858

Sampling Design for Car Survey Using Stratified Random Sampling

Valeno Glenedias Widodo[1]*, Yusep Suparman[2], Gumgum Darmawan[3]
Departemen Statistika FMIPA UNPAD
Corresponding Author: Valeno Glenedias Widodo
valeno19001@mail.unpad.ac.id

A R T I C L E I N F O

Keywords: Time Between Arrivals, Regression, Stratified Sampling

Received : 2 August
Revised : 14 August
Accepted: 18 September

©2023 Widodo, Suparman, Darmawan: This is an open-access article distributed under the termsof the Creative Commons Atribusi 4.0 Internasional.

A B S T R A C T

Perumda Pasar Juara wants to know the number of cars that enter the Kosambi Market parking area for one month. However, limited access to available data means that Perumda has to count vehicles manually. Therefore, an appropriate sampling design is needed to estimate the number of cars. The data used is the result of observations based on preliminary sampling on October 14-20, 2022 at 08:00-17:00 WIB. The variables of the data used are car arrival time, day dummy, and hour dummy. The method used is dummy regression analysis and stratified sampling design. The regression analysis results show that there are four strata where each stratum has a three-parameter Weibull distribution. Based on the results, the minimum sample size required with a 5% error rate is 116 hours per day and is allocated according to the strata

DOI: https://doi.org/10.55927/modern.v2i5.5872
E-ISSN: 2985-6108
https://journal.formosapublisher.org/index.php/modern

INTRODUCTION

The population of Bandung City continues to increase every year. Based on the results of the 2020 Population Census, the population of Bandung City experienced a growth of 0.21% per year when compared to 2010. In addition, population density also increased by 0.004% from the previous year. This number has not been added by foreign nationals domiciled in the city of Bandung and migrants who arrive as tourists at certain times. Population growth results in denser activities that require high mobility. Therefore, the use of motorized vehicles also increases every year. The number of motorized vehicles in 2010 increased by an average of 11% per year, with 859,411 two-wheeled vehicles and 134,654 four-wheeled vehicles (Bandung City Transportation Office, 2016). The increase in the number of motorized vehicles makes parking spaces mandatory in every public building, one of which is a traditional market.

Kosambi Market is one of the traditional markets managed by Perumda Pasar Juara Kota Bandung. Perumda Pasar Juara as the market manager wants to count the number of cars that enter the parking area for one month. The calculation of the number of cars can be done easily if Perumda Pasar Juara has access to see the available data. However, Perumda Pasar Juara does not have access to data on vehicles entering the parking area, so they have to count incoming cars manually. Manual calculations require large costs and energy so that sampling design is needed so that vehicle surveys can be carried out effectively efficiently.

LITERATURE REVIEW

Time Between Arrivals

One of the characteristics of the queuing system is the arrival rate (Listiyani et al, 2019). The arrival pattern of customers is usually calculated through the time between arrivals, which is the time between the arrival of two consecutive customers at a service facility. When the pattern of customer arrivals arrives one by one, then the arrival follows a process with a certain probability distribution. A commonly used probability distribution is the Poisson distribution, where the arrival is free, having no effect on the arrival before or after. The Poisson Distribution assumption suggests that customer arrivals are random. The number of customers arriving in a unit of time is the arrival rate λ and the length of the time interval between two customer arrivals is the time between arrivals $1/\lambda$ (Pardede et al., 2014).

Sampling Techniques

There are three sampling methods that can be used to obtain a representative sample, namely sober sampling, consideration or purposive sampling, and opportunity sampling.

In this study, the sampling technique used is opportunity sampling because it pays attention to the opportunities of each member of the population. Here are some sampling techniques included in opportunity sampling.

1. *Simple Random Sampling*
   Simple random sampling (SRS) is a method that selects sample units from a population with equal chances of selection (Cochran, 1997). SRS requires all units of observation in a population called a sampling frame (Lohr, 2010). Simple random sampling can be done if the population is homogeneous.
2. *Cluster Random Sampling*
   In cluster random sampling, individuals in a population can only be sampled if they belong to the cluster selected as a sample (Lohr, 2010). Cluster random sampling is done by taking clusters randomly using SRS and taking all or part of the sample from the selected cluster (Lohr, 2010). Cluster random sampling can be done if between clusters or groups are homogeneous.
3. *Stratified Random Sampling*
   Stratified random sampling is a sampling process that divides the population into non-overlapping subpopulations (strata) and then simple random sampling is carried out on each stratum (Cochran, 1997). Stratified random sampling can be done if the strata are heterogeneous and within the strata are homogeneous.

Weibull Distribution With Three Parameters

Weibull distributions are commonly used in lifetime distributions in reliability. The two-parameter Weibull distribution can represent a decreasing, constant, or increasing failure rate. The Weibull distribution can also be added a third parameter, namely the location parameter denoted by gamma γ (Forbes et al., 2011). The Weibull distribution with three parameters has the following density function:

$$f(x) = \left[\frac{\beta(x-\gamma)^{\beta-1}}{\eta^{\beta}}\right] exp\left\{-\left[\frac{(x-\gamma)}{\eta}\right]^{\beta}\right\} \quad (1)$$

With:
x>γ
γ : parameter location, γ>0
η : parameter scale, η>0
β : parameter shape, β>0

The Weibull distribution with three parameters has the following average and variance (Forbes et al., 2011):

$$mean = \gamma + \eta\Gamma\left[\frac{(\beta+1)}{\beta}\right] \quad (2)$$

$$Variance = \eta^2\left(\Gamma\left[\frac{(\beta+2)}{\beta}\right] - \left\{\Gamma\left[\frac{(\beta+1)}{\beta}\right]\right\}^2\right) \quad (3)$$

Dummy Linear Regression

Regression analysis is a statistical technique used to investigate and model relationships between variables (Montgomery et al., 2012). Regression analysis can be used to explain the relationship between the response variable (Y) and the explanatory variable (X). In some cases, the independent/explanatory variable in

regression analysis is not only ratio-scaled but can be nominally scaled. These variables are known as categorical variables or dummy variables (Gujarati &; Porter, 2009). A regression model with dummy variables can be expressed as follows:

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} \ldots + \beta_n D_{ni} + \varepsilon_i \quad (4)$$

With:

$D_{1i}, D_{2i}, \ldots, D_{ni}$: variabel *dummy*

METHODOLOGY

The method to be used is a stratified sampling design with dummy regression analysis and wald test to form strata. The data used is time data between arrivals from observations based on a preliminary sampling design conducted on October 14-20, 2022 at 08:00-17:00 WIB. The variables from the data were car arrival time, dummy day, and dummy hour for a total of 1061 observations.

RESEARCH RESULT

Outlier Detection

Outliers or outliers are values that are very different from other values. Values greater than the upper quartile plus 3*IQR are extreme outliers. The same rule is also applied to cases below the lower quartile (Field, 2017). In this study, outlier detection will be carried out by looking for values that are outside 3*IQR with the following formula:

$$Outlier = Upper\ Quartile + 3 * \text{IQR}, \text{Lower Quartile} - 3 * \text{IQR} \quad (5)$$

Regressio Model

The regression model that will be used in this study can be written with the following equation (Gujarati &; Porter, 2009):

$$Y_{jhi} = \beta_0 + \sum_{j=1}^{6} \beta_j D_{ji} + \sum_{h=1}^{8} \beta_h D_{hi} + \varepsilon_{jhi} \quad (6)$$

With:

Y_{jhi} : Time between arrival on day i, time j, and observation i
β_0 : intercept
β_j : j-day regression coefficient
β_h : h-hour regression coefficient
D_{ji} : Day J dummy variables
D_{hi} : i-hour dummy variable
ε_{jhi} : Residual regression model on day i, time j, and observation i-day

The parameter estimation method used is OLS or the least squares method. OLS is a regression method that minimizes squared errors. The parameters β can be estimated with OLS as follows (Montgomery et al., 2012):

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T Y} \quad (7)$$

With:

$\widehat{\boldsymbol{\beta}}$: Regression coefficient estimator vector
$\boldsymbol{X}$: matrix of independent variables
$\boldsymbol{Y}$: dependent variable vector

The selection of regression models is carried out by backward elimination. The backward elimination method tries to find a model with the opposite direction of forward elimination. The backward elimination method starts by entering all K regressor candidates and removing the regressor with the smallest partial F (or t) statistic. Then the partial F statistic will be recalculated for the regression model with K-1 regressor. The procedure continues until the smallest partial F statistic (or t) is greater than the previously established limit value (Montgomery et al., 2012).

Assumption testing

1. *Normality Test*

   A normality test is performed to determine whether the residual value in the regression follows the normal distribution. There are several methods that can be used to detect whether residuals have a normal distribution, one of which is Kolmogorov-Smirnov. Here is a normality test using the Kolmogorov-Smirnov method:

   Hypothesis:

   $H_0 : F_0(x) = F(x)$ (normally distributed residuals)
   $H_1 : F_0(x) \neq F(x)$ (residuals are not normally distributed)

   Statistic Test:

   $$D = \underset{x}{Sup}|F_0(x) - S_n(x)| \quad (8)$$

   With:

   $F_0(x)$: cumulative frequency distribution function below H_0
   $S_n(x)$: cumulative chance function of a random sample of as many as N observations.

   Test Criteria : Reject H_0 if value $D_{hitung} > D_{tabel}$ or $p\text{-}value < \alpha$, Accept in other things.

2. *Uji Multikolinearitas*

   Multicollinearity is the existence of perfect or near-perfect linear relationships among some or all of the explanatory variables in a regression model. Detection of multicollinearity in linear regression models can be done by paying attention to the value of VIF (Variance Inflation Factor). If the VIF value > 10, then multicollinearity occurs between predictor variables (Gujarati & Porter, 2009). The formula for calculating the value of VIF is as follows:

   $$VIF_j = \frac{1}{1 - {R_j}^2} \quad (9)$$

   With:

   j : 1,2,…,p, where p is the number of predictor variables
   ${R_j}^2$: p-th independent variable coefficient of determination with other variables

3. *Homoscedasticity Test*

Homoscedasticity is the assumption of fallacy or the variable of disorder in regression models has a constant variance. Heteroscedasticity can be determined using the Breusch-Pagan (BP) test as follows (Verbeek, 2004):
Hypothesis:
$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$ (identical or no heteroscedasticity occurs)
H_1 : minimal ada satu $\sigma_i^2 \neq \sigma^2, i = 1, 2, \ldots, n$ (heteroscedasticity occurs)
Test Statistics:

$$BP = nR_{aux}^2 \quad (10)$$

With:
n : The number of observations
R_{aux}^2: value R^2 of auxiliary *regression*
Test Criteria: Reject H_0 if BP $> \chi_j^2$ or *p-value* $< \alpha$, Accept in other things.

4. *Autocorrelation Test*

Autocorrelation is a condition where there is a correlation between residuals in a certain period with residuals in the previous period. Many methods can be used to detect autocorrelation problems, one of which is Durbin-Watson (Gujarati &; Porter, 2009). Here is an autocorrelation test using Durbin-Watson (Widarjono, 2018):
Hypothesis:
$H_0 : \rho = 0$ (no autocorrelation)
$H_1 : \rho \neq 0$ (There is autocorrelation)
Test Statistics:

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \quad (11)$$

With:
T : Number of observations
e_t : Residual t-th Regression Model
Test Criteria : : Reject H_0 if *p-value* $< \alpha$, Accept in other things

Uji Wald

The Wald test is one method for testing hypotheses in regression analysis (Gujarati & Porter, 2009). The Wald test can be used to see if = so that data can be combined into one variable. Limitations in the Wald test can be written (Verbeek, 2004). Here are the testing steps of Wald's testing$\beta_i \beta_j R\beta = q$:
Hypothesis:
$H_0 : \beta_i = \beta_k$ (The I-th regression coefficient is equal to the k-th regression coefficient)
$H_1 : \beta_i \neq \beta_k$(The I-th regression coefficient is not the same as the K-th regression coefficient)
Test Statistics:

$$\xi = \frac{(\boldsymbol{Rb} - q)'\left(R(\boldsymbol{X'X})^{-1}\boldsymbol{R'}\right)(\boldsymbol{Rb} - q)}{\sigma^2} \quad (12)$$

With:
R : Dimensionless vectors contain constraints$1x\ J$
q : Scalar with a value of 0
b : Regression coefficient vector with dimensions $J\ x\ 1$
X : matrix of independent variables
σ^2 : Regression model residual variance

Test Criteria: Reject H_0 if $\xi > \chi_j^2$ or *p-value* $< \alpha$, Accept in other things.

Sampling Design

1. Determining Research Objectives
   The purpose of sampling in general is the estimation of parameters including estimating the average, total, or proportion of a population. Sampling can also be aimed at hypothesis testing. The purpose of this study is to obtain a sampling design so that the sample obtained is representative of the population so that it can be used to estimate the number of cars entering the Kosambi Market parking area through the average time between arrivals.
2. Determining the Population
   The target population is a collection of objects that you want to observe. The determination of the target population will have an impact on the conclusions resulting from the study. In this study, the target population determined is the large number of cars that enter the parking area of Pasar Kosambi Bandung City at 08:00-17:00 WIB in a certain month for 30 days.
3. Constructing the Sampling Framework
   The unit of observation is the object to be measured. Meanwhile, the sampling unit is the unit to be sampled. The observation unit and sampling unit in this study is the time between the arrival of cars and parked cars at Kosambi Market. The sampling units are arranged in a sampling frame to be randomly selected based on the size obtained.
4. Regression Analysis
   Based on the model in equation (6), regression analysis was performed using six dummy day variables and eight dummy hour variables as independent variables. Meanwhile, the dummy variables on Wednesday and 16:00 – 17:00 are used as references. The variables that will be used for regression analysis can be seen in the following table:

Table 1. Variables Used for Regression Analysis

Variabel	Information	Variabel	Time
Y	Time between arrivals	D_8	*Dummy* for 09:00 – 10:00
D_1	*Monday dummy*	D_9	*Dummy* for 10:00 – 11:00
D_2	*Tuesday dummy*	D_{10}	*Dummy* for 11:00 – 12:00
D_3	*Thursday dummy*	D_{11}	*Dummy* for 12:00 – 13:00
D_4	*Friday dummy*	D_{12}	*Dummy* for 13:00 – 14:00
D_5	*Saturday dummy*	D_{13}	*Dummy* for 14:00 – 15:00
D_6	*Sunday dummy*	D_{14}	*Dummy* for 15:00 – 16:00
D_7	*Dummy* for 08:00 – 09:00	-	-

1. Forming strata

The results of regression analysis can be used as guidelines for strata formation. Variables that are not significant in the regression analysis will be combined with references as a group. In addition, the Wald test was performed to incorporate variables that were as significant as the rest of the group. The formation of strata is carried out by combining groups of hours and days based on a regression model, where strata will be formed according to the following Table:

Table 2. How to form strata

Strata	Group Day	Hour Group
1	Insignificant combined days with referenced days	Combination of an insignificant clock with the clock used as a reference
2	Insignificant combined days with referenced days	Significant hours
3	Significant days	Combination of an insignificant clock with the clock used as a reference
4	Significant days	Significant hours

2. Determining Sampling Techniques

The sampling technique that will be used to estimate the population average is stratified *random sampling*. The stratified random sampling method can be used when the strata are heterogeneous and the strata are homogeneous. In this study, the time data between arrivals is heterogeneous per strata so that stratified sampling methods are suitable to be used.

3. Determining Estimated Population Size
   The population size will be estimated by the following equation:

$$\widehat{N}_h = \frac{Number\ of\ hours\ of\ days\ per\ stratum}{Average\ time\ between\ arrivals\ per\ strata} \quad (13)$$

With:
$\widehat{N}_h$ = population size estimator

4. Menentukan Ukuran Sampel

$$n = \frac{(\sum W_h S_h)^2}{\left(\frac{e}{Z_{\frac{\alpha}{2}}}\right)^2 + \frac{1}{N}\sum W_h S_h^2} \quad (14)$$

With:
n : sample size
S_h : Standard deviation per stratum
S_h^2 : varians per *stratum*
W_h: weight Stratum $\left(\frac{\widehat{N}_h}{\widehat{N}}\right)$

5. Determining Sampling Allocation
   In this study, the sampling allocation to be used is Neyman's allocation. Neyman's allocation is a special case of optimal allocation that assumes the survey cost for each stratum is the same while the variance is different. Neyman's allocation has the following equation:

$$n_h = n\frac{N_h S_h}{\sum N_h S_h} \quad (15)$$

The resulting sample size states how many arrivals are needed to estimate the average time between arrivals. However, because it is not possible to randomize vehicles, randomization is carried out using day hours by multiplying the sample size by the average time between arrivals.

6. Performing the Sampling Process
   Because the sampling technique used is stratified random sampling, the sampling process is carried out randomly using simple random sampling in each strata according to the sample size. The randomization process is done by numbering each hour of the day of the month (30 days) and randomly selecting one number as the first sample. The process of retrieving numbers is carried out without return to avoid the appearance of numbers that have been selected. Next, the remaining numbers are randomly reselected as a second sample. This step is carried out as much as the minimum number of hours of days from each stratum that has been determined. Randomization is done using random numbers with the help of R *software*.

DISCUSSION

Outlier Detection

Outlier detection for time data between arrivals is done with the 3*IQR rule and produces outliers as follows:

Table 3. Outlier *Detection Results*

No.	Time between arrivals	No.	Time between arrivals
1	8.850	11	9.817
2	10.300	12	9.067
3	9.767	13	13.917
4	9.467	14	11.233
5	11.500	15	17.267
6	8.717	16	9.500
7	33.283	17	8.750
8	12.117	18	9.333
9	9.517	19	8.400
10	10.667	20	9.750

Based on Table 3, there are 20 observations on inter-arrival time data that are considered extreme outliers. In this study, extreme outliers were not included in the regression analysis so that the total observations to be used were 1041.

Based on Table 3, there are 20 observations on inter-arrival time data that are considered extreme outliers. In this study, extreme outliers were not included in the regression analysis so that the total observations to be used were 1041Penaksiran parameter regresi dilakukan dengan menggunakan metode OLS. Berikut merupakan hasil taksiran parameter dari regresi OLS:

Table 4. Estimated OLS Regression Parameters

Variabel	**Estimate**	**Std. Error**	**t value**	**Pr (>|t|)**
Intercept	2.365	0.235	10.077	< 0.000
Monday	0.066	0.229	0.289	0.289
Tuesday	0.173	0.243	0.712	0.712
Thursday	-0.076	0.201	-0.378	-0.378
Friday	-0.138	0.202	-0.682	-0.682
Saturday	-0.122	0.233	-0.526	-0.526
Sunday	0.302	0.229	1.320	1.320
08:00-09:00	-0.822	0.280	-2.936	-2.936
09:00-10:00	-0.767	0.218	-3.515	-3.515
10:00-11:00	-0.626	0.241	-2.595	-2.595
11:00-12:00	-0.713	0.224	-3.186	-3.186
12:00-13:00	-0.517	0.218	-2.372	-2.372
13:00-14:00	-1.112	0.217	-5.129	-5.129
14:00-15:00	-0.693	0.232	-2.982	-2.982
15:00-16:00	-1.379	0.290	-4.749	-4.749

Based on Table 4, there are variables with *p-value* > α = 5%, so it can be concluded that there are insignificant variables in the regression model. The selection of regression models will be done by backward eliminating, which is eliminating variables one by one until all variables are significant. Here are the results of estimating parameters using OLS after removing insignificant variables:

Table 5. Estimation of OLS Regression Parameters After Removing Insignificant Variables

Variabel	**Estimate**	**Std. Error**	**t value**	**Pr (>|t|)**
Intercept	2.312	0.174	13.263	< 0.000
Sunday	0.355	0.164	2.161	0.031
08:00-09:00	-0.776	0.258	-3.011	0.003
09:00-10:00	-0.800	0.215	-3.719	0.000
10:00-11:00	-0.620	0.234	-2.653	0.008
11:00-12:00	-0.661	0.217	-3.045	0.002
12:00-13:00	-0.477	0.211	-2.258	0.024
13:00-14:00	-1.161	0.211	-5.502	0.000
14:00-15:00	-0.709	0.223	-3.178	0.002
15:00-16:00	-1.350	0.286	-4.718	0.000

Based on Table 5, the variables that are not significant and combined with reference are Friday, Saturday, Monday, Tuesday, and Thursday.

Assumption Testing

1. *Uji Normalitas*

Table 6. Kolmogorov-Smirnov Test Results

D	0.142
p-value	< 0.000

Based on the results of the residual normality test, *a p-value of* < α = 5% is obtained, so that the decision to accept H_0 or the residual is not normally distributed. However, because the data sample is large, the central limit theorem applies and the normality assumption can be ignored.

2. *Multicollinearity Test*

Table 7. Value of VIF Regression Variables

Variabel	Nilai VIF
Sunday	1.357936
08:00-09:00	1.723491
09:00-10:00	2.387504
10:00-11:00	2.039997
11:00-12:00	2.429221
12:00-13:00	2.497653
13:00-14:00	2.506897
14:00-15:00	2.262269
15:00-16:00	1.794348

Based on the results above, it can be seen that the entire variable has a VIF value of less than 10 so that it can be concluded that there is no case of multolinearity.

3. *Homoscedasticity Test*

Table 8. Breusch-Pagan Test Results

BP	11.928
df	9
p-value	0.217

Based on the results of Table 8, *a p-value of* > α = 5% was obtained, so that the decision to fail to reject H_0 was obtained which means there is no heteroscedasticity in regression.

4. *Autocorrelation Test*

Table 9. Durbin-Watson Test Results

D-W	1.975501
p-value	0.524

Based on the results of Durbin Watson's test, it was found that the *p-value* > α = 5%, so that the decision failed to reject H_0 which means there is no autocorrelation.

Wald Test

1) Time 10:00 – 11:00 and 12:00 – 13:00

$H_0 : \beta_4 = \beta_6$ (The regression coefficient of 10:00 – 11:00 is the same as the regression coefficient of 12:00 – 13:00)

$H_1 : \beta_4 \neq \beta_6$(The regression coefficient of 10:00 – 11:00 is not the same as the regression coefficient of 12:00 – 13:00)

Table 10. Wald Test Results $\beta_4 = \beta_6$

ξ	7.60534
p-value	0.006

Based on the results of Wald testing for parameters $\beta 4$ and β_6, It was found that *p-value* < α = 5%, so that a decision of rejection is obtained H_0 yang berarti β_4 not the same as β_6.

2) Time 11:00 – 12:00 dan 12:00 – 13:00

$H_0 : \beta_5 = \beta_6$ (koefisien regresi jam 11:00 – 12:00 sama dengan koefisien regresi jam 12:00 – 13:00)

$H_1 : \beta_5 \neq \beta_6$ (koefisien regresi jam 11:00 – 12:00 tidak sama dengan koefisien regresi jam 12:00 – 13:00)

Table 11. Hasil Uji Wald $\beta_5 = \beta_6$

ξ	1.975501
p-value	0.524

Based on the results of Wald testing for parameters β_5 and β_6, It was found that p-value > α = 5%, So that a decision was obtained to fail to refuse H_0 which means β_5 same as β_6. Because β_5 same as β_6 Then the variables 11:00-12:00 and the variables 12:00-13:00 can be combined to produce regression as follows:

Table 12. Estimated Regression Parameters After Combining Variables 11:00-12:00 &; Variable Hours 12:00-13:00

Variabel	Estimate	Std. Error	t value	Pr (>|t|)
Intercept	1.912	0.111	17.214	< 0.000
Minggu	0.426	0.163	2.619	0.009
08:00-09:00	-0.376	0.221	-1.704	0.089
09:00-10:00	-0.408	0.170	-2.395	0.017
10:00-11:00	-0.219	0.192	-1.144	0.253
11:00-13:00	-0.109	0.159	-0.682	0.495
13:00-14:00	-0.776	0.167	-4.651	0.000
14:00-15:00	-0.308	0.178	-1.728	0.084
15:00-16:00	-1.004	0.262	-3.840	0.000

Regression analysis is continued with the backward elimination method until obtaining significant variables. The following are the results of parameter estimation after using the backward elimination method:

Table 13. Estimated Regression Parameters

Variabel	Estimate	Std. Error	t value	Pr (>|t|)
Intercept	1.710	0.056	30.631	< 0.000
Sunday	0.447	0.160	2.794	0.005
13:00-14:00	-0.579	0.137	-4.238	0.000
15:00-16:00	-0.818	0.243	-3.374	0.001

Based on Table 13, the variables that are not significant and combined with references are the variables of 08:00-09:00, 09:00-10:00, 10:00-11:00, 11:00-12:00, 12:00-13:00, 14:00-15:00. Next, the Wald test will be carried out for variables at 13:00-14:00 and 15:00-16:00.

3) Time 13:00 – 14:00 and 15:00 – 16:00

$H_0 : \beta_7 = \beta_9$ (The regression coefficient of 13:00 – 14:00 is the same as the regression coefficient of 15:00 – 16:00)

$H_1 : \beta_7 \neq \beta_9$ (The regression coefficient of 13:00 – 14:00 is not the same as the regression coefficient of 15:00 – 16:00)

Table 14. Hasil Uji Wald $\beta_7 = \beta_9$

ξ	0.8632343
p-value	0.353

Based on the results of Wald testing for parameters β_7 and β_9, It was found that *p-value* > α = 5%, So that a decision was obtained to fail to refuse H_0 which means β_7 same as β_9. Because β_7 same as β_9 then the hour variable 13:00-14:00 and hour variables 15:00-16:00 can be combined to produce regression:

Table 15. Estimated Regression Parameters

Variabel	**Estimate**	**Std. Error**	**t value**	**Pr (>|t|)**
Intercept	1.713	0.056	30.788	< 0.000
Sunday	0.396	0.150	2.636	0.008
13:00 – 14:00 &15:00 – 16:00	-0.626	0.127	-4.946	0.000

Based on the results of the Wald test and regression analysis in Table 13, the researcher formed four strata to be used where stratum one is the group of insignificant days and hours, stratum two is the group of insignificant days and significant hours, stratum three is the group of significant days and insignificant hours, stratum four is the group of significant days and hours. The strata results can be seen in the following Table:

Table 16. Result of Strata

Strata	Day	Time
1	Monday, Tuesday, Wednesday, Thursday, Friday, Saturday	08:00-09:00, 09:00-10:00, 10:00-11:00, 11:00-12:00, 12:00-13:00, 14:00-15:00, 16:00-17:00
2	Monday, Tuesday, Wednesday, Thursday, Friday, Saturday	13:00-14:00, 15:00-16:00
3	Sunday	08:00-09:00, 09:00-10:00, 10:00-11:00, 11:00-12:00, 12:00-13:00, 14:00-15:00, 16:00-17:00
4	Minggu	13:00-14:00, 15:00-16:00

Determining Research Objectives

The purpose of this study is to obtain a sampling design so that the sample obtained is representative of the population so that it can be used to estimate the number of cars entering the Kosambi Market parking area through the average time between arrivals.

Determining the Population

The target population in this study is cars that park at Pasar Kosambi Kota Bandung in one month (30 days) at time intervals of 08:00 – 17:00 WIB. The value of standard deviation per stratum can be calculated based on equation (3). Meanwhile, the estimated population size and weight of each stratum can be calculated based on equations (13) and (14) respectively. The results of the calculation of standard deviation, estimated population size, and stratum weight can be seen in the Table as follows:

Table 17. Average, Standard Deviation, Estimated Population Size, and Weight of Each Stratum

Strata	Standard deviation (S_h)	Estimated population size ($\widehat{N_h}$)	Weight (W_h)
1	2,006	5,493	0.620
2	2,065	2,087	0.235
3	2,041	867	0.098
4	2,561	416	0.047
Total (N)		8,863	1

Menentukan Unit Sampling

The sampling unit in this study is cars that enter the parking area of Pasar Kosambi Kota Bandung in one month (30 days) starting on Sunday with a time interval of 08:00 – 17:00 WIB.

Determining Sampling Techniques

The sampling technique that will be used in this study is stratified sampling technique.

Menentukan Ukuran Sampel

The minimum sample size that needs to be observed to estimate the average time between car arrivals into the parking area of Pasar Kosambi Kota Bandung can be seen based on equation (14). Based on the sample size formula, a minimum sample of the number of car arrivals into the parking area of Pasar Kosambi Bandung City that needs to be observed with a confidence level of 95%, is 3730 units of vehicles.

Perform the Sample Allocation Process

After determining the minimum sample size that needs to be observed, then the sample size needs to be allocated into each strata. The sample size for each stratum is calculated using equation (15) with the following results:

Table 18. Minimum Sample Size (Vehicle Unit)

Strata	Ukuran Sampel (Unit)
1	2,263
2	886
3	364
4	219

Performing the Sampling Process

The sampling process uses a simple random sampling technique at each stratum by randomizing the number of hours of observation days from each strata. The minimum number of hours of day required can be seen from the following Table:

Table 19. Minimum Sample Size (Time)

Strata	Sample Size (Time, Day)
1	73
2	22
3	15
4	6

Random sampling of day hours will use R *software* with the following results:

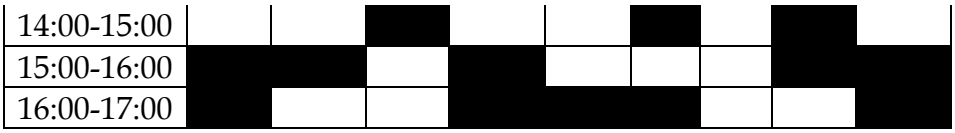
Table 20. Selected Samples

Time	Week 1						
	Sun	Mon	Tue	Wed	Thu	Fri	Sat
08:00-09:00							
09:00-10:00							
10:00-11:00							
11:00-12:00							
12:00-13:00							
13:00-14:00							
14:00-15:00							
15:00-16:00							
16:00-17:00							

Time	Week 2						
	Sun	Mon	Tue	Wed	Thu	Fri	Sat
08:00-09:00							
09:00-10:00							
10:00-11:00							
11:00-12:00							
12:00-13:00							
13:00-14:00							
14:00-15:00							
15:00-16:00							
16:00-17:00							

Jam	Week 3						
	Sun	Mon	Tue	Wed	Thu	Fri	Sat
08:00-09:00							
09:00-10:00							
10:00-11:00							
11:00-12:00							
12:00-13:00							
13:00-14:00							
14:00-15:00							
15:00-16:00							
16:00-17:00							

Jam	Week 4							Week 5	
	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon
08:00-09:00									
09:00-10:00									
10:00-11:00									
11:00-12:00									
12:00-13:00									
13:00-14:00									



CONCLUSIONS AND RECOMMENDATIONS

Based on the discussion, the conclusions obtained from this study are as follows:

1. The sampling technique used is a stratified sampling technique with a total of four strata, where each strata is distributed Weibull three parameters.
2. With a confidence level of 95%, the minimum sample size that must be taken is 3730 units of vehicles or 116 hours of observation are required.

This study aims to create a sampling design for one month using data from preliminary sampling results for one week so as to assume there are no differences in time characteristics between arrivals in different weeks. For future research, it is recommended to use more complete data to see if there are differences in time characteristics between arrivals in different weeks or on holidays.

REFERENCES

Cochran, william G. (1997). Sampling Techniques. John Wiley & Sons, Inc.

Daulay, I. N. (2014). Analisis Model Antrian Dan Kelayakan Layout Stasiun Pengisian Bahan Bakar Umum (SPBU) Pekanbaru. 22, 1–18.

Dinas Perhubungan Kota Bandung. (2016). Bandung Urban Mobility Project.

Djamal, N., & Maulana, H. E. (2018). Analisis Antrian Kapal di Dermaga 5 Pelabuhan Banten. Jurnal INTECH Teknik Industri Universitas Serang Raya, 4(1), 33. https://doi.org/10.30656/intech.v4i1.855

Faraway, J. J. (2002). Practical Regression and Anova using R. 16(4), 323–335. https://doi.org/10.1016/0360-1315(91)90006-D

Field, A. (2017). Discovering Statistics Using IBM SPSS Statistics (5th ed.). SAGE Publications Ltd.

Forbes, C., Evans, M., Hastings, N., & Peacock, B. (2011). Statistical Distributions, 4th ed. (4th ed.). John Wiley & Sons, Inc.

Gujarati, D. N., & Porter, D. C. (2009). Dasar-dasar Ekonometrika (5th ed.). Salemba Empat.

Hasanah, R., & Miru, S. (2021). Analisis Sistem Antrian Loket Pendaftaran Peserta Pada Badan Penyelenggaraan Jaminan Sosial Kesehatan Kota Palu. 7(2), 117–126.

Heizer, J., Render, B., & Munson, C. (2017). Operations management : sustainability and supply chain management (12th ed.). Pearson Education, Inc.

Listiyani, R., Linawati, L., & Sasongko, L. R. (2019). Analisis Proses Produksi Menggunakan Teori Antrian Secara Analitik dan Simulasi. Jurnal Rekayasa Sistem Industri, 8(1), 9–18. https://doi.org/10.26593/jrsi.v8i1.3154.9-18

Lohr, S. L. (2010). Sampling : Design and analysis. In Earth Reinforcement and Soil Structures. https://doi.org/10.1016/b978-0-408-03549-1.50013-3

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to Linear Regression Analysis -5th ed. (5th ed.). John Wiley & Sons, Inc.

Pardede, A. M. H., Mawengkang, H., & Situmorang, Z. (2014). Simulasi Antrian Kedatangan Berkelompok Dengan Pelayanan Weibull Oleh Banyak Server. Jurnal Teknologi Informasi dan Komunikasi.

Rachman, T. (2016). Simulasi Model Antrian Optimal Loket Pembayaran Parkir. Jurnal Inovisi TM, 12, 72.

Sudjana. (2013). Metoda Statistika (ketujuh). Tarsito.

Swarjana, I. K. (2022). Teknik Sampling & Bias dalam Penelitian (E. Risanto (ed.)). Andi.

Taylor, B. W. I. (2004). Introducing to Management Schience. In Journal of Manufacturing Science and Engineering, Transactions of the ASME (Vol. 83, Nomor 3). Pearson Education, Inc.

Verbeek, M. (2004). A Guide to Modern Econometrics. In Applied Econometrics (2nd ed., Vol. 8, Nomor 4). John Wiley & Sons, Ltd.

Widarjono, A. (2018). Ekonometrika (5 ed.). UPP STIM YKPN.