Agus Sukmana and Dharma Lesmono

Using Statistical Techniques for Exploring Parking Data on Hoogstraat Rotterdam

Intisari

Masalah parkir dapat merupakan masalah bagi administrator Rotterdam. Berdasarkan data wilayah parkir jalan Hoogstraat digunakan beberapa teknik statistik untuk memperoleh beberapa kesimpulan dan saran. Ditemukan bahwa distribusi lognormal cukup representatif bagi masa parkir nyata, bahwa jumlah tempat parkir lebih dari memadai, bahwa tiada korelasi linear antara waktu parkir nyata dan waktu yang direncanakan, dan dengan rata-rata tertentu orang memarkir 14,91 menit

Abstract

The parking problem is an issue possibly faced by city authorities of Rotterdam. Given the data set of the parking bays on Hoogstraat, we want to explore the data by using statistical techniques to solve some problems and offer some suggestions. We find that the lognormal distribution is a suitable continuous distribution for actual parking times, that the number of the parking bays is more than sufficient, that there is no linear correlation between parking ticket's time and actual parking time, and that a certain average people park 14.91 minutes longer than planned.

I. Introduction

The basis of this study is a set of parking data in Hoogstraat, Rotterdam (See table 1). The aim of this study is to investigate parking behaviour and to suggest improving measures. The data were taken on Wednesday December 17, 1997 from 10.00 to 17.00 hours. The information contained in the data set are place (parking bay numbers; 1-29), arrival time, type of ticket payment (buy ticket, no ticket, and licence), length of parking time on ticket, departure time, and the police regristration number of the car. "Licence" means that the owner of the car pays a parking ticket monthly. The actual situation for the parking event can be described as follows. Consider a car entering a parking area to find an empty parking bay. The arrival time of that car is recorded manually. After parking the car, the driver walks to the parking machine to get a parking ticket (if he decides to pay). At the parking machine he has to decide how long he will park his car, and then he puts coin(s) in that machine and gets the ticket. On the ticket is also recorded the latest time he has to depart from the parking area. When he actually departs from the parking area, his departure time is again recorded manually. The police regristration number of the car is also recorded to make sure that the parking bay used by the car, arrival time, planned departure time on the ticket, and ac-

tual departure time indeed belong to the right car.

We want to explore the data set to solve the problems below, subject to the particular day when the data were taken.

II. The Problems

The problems we want to solve are the following:

- Modelling the probability distribution of the actual parking time.
  We define X (a random variable) as the actual parking time, i.e. the difference between actual departure time and arrival time. By using methods for summarizing and exploring the data, we propose a continuous probability distribution for X and check the fit of the model.
- Describing the proportion of used parking bays
  The proportion of used parking bays is the number of occupied parking bays divided by the total number of parking bays. We want to show whether there are enough parking bays and at which time most of the parking bays are used for parking. We also want to describe how the proportion of used parking bays varies over time.
- Describing the relation between actual parking time and the ticket (planned) parking time.
  The relation between the ticket parking time (the parking time one has paid for) and the actual parking time will be explored using descriptive methods.

III. Modelling the Probability Distribution of the Actual Parking Time

Based on the given data, we first calculate the actual parking time (Table 1, Actual1). In this observation there are 10 data missing. This is because the departure time of cars that arrive at 10.00 a.m. in ten different parking bays were not recorded. Then we treat the actual parking time as a random variable X. Before we can propose a continuous probability distribution for X, we first look at the histogram of the random variable X (Figure 1). This histogram tends not to follow normal distribution. Also from the Q-Q plot for actual parking data (Figure 2) we see that the plot deviates from the expected normal values for the lower and larger quantile. [1]

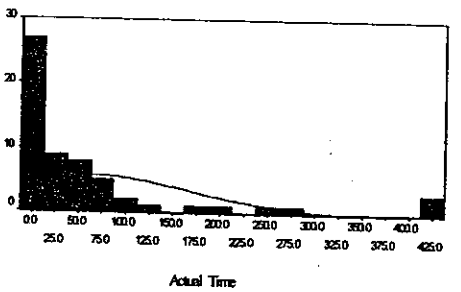

Figure 1. Histogram The Actual Parking Time Rotterdam Parking Bay Hoogstraat

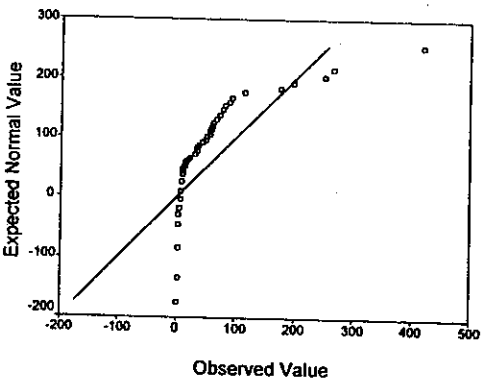

Figure 2. Normal Q-Q of Plot ACTUAL1

From the histogram, we see that the data seems to follow certain continuous distributions like exponential, gamma, and lognormal. The Q-Q plots for these continuous distributions (Figure 3, 4, 5) tend to stay close to the expected values. Then we will determine which continuous distribution is suitable for the actual parking time on December 17, 1997. We use the Kolmogorov-Smirnov (KS) test for this purpose. In general the KS Test procedure is as follows [2]:

- Arrange the data in ascending order i.e.

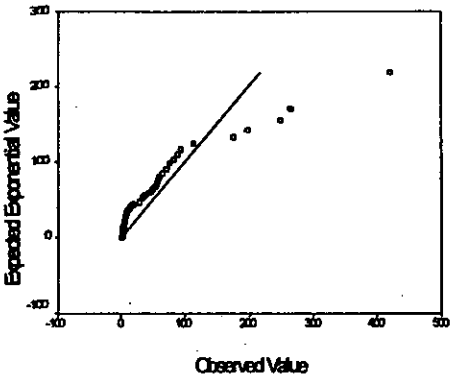

Figure 3. Exponential Q-Q Plot of ACTUAL1

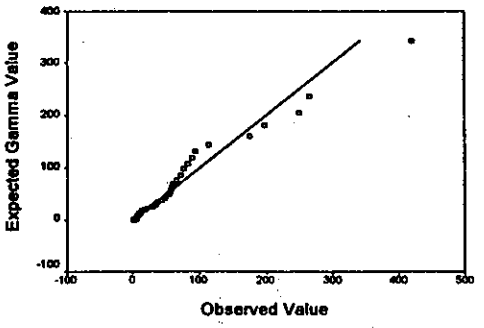

Figure 4. Gamma Q-Q Plot of ACTUAL1

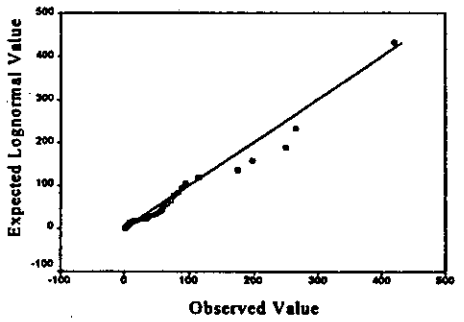

Figure 5. Lognormal Q-Q Plot of ACTUAL1

$$x1 < x2 < \ldots < xn.$$

- Calculate the cumulative distribution function S(x), where Sn(x) = t/n, t=1,2,n.
- Set the hypotheses; null hypothesis is that the actual parking data follow the expected cumulative distribution function (cdf) F(x) like the exponential, gamma, or lognormal, and conversely is the alternative hypothesis.
- Choose the significance level α.
- Calculate the statistics

$$Dmax = \max | Sn(x) - Fn(x) |$$

- Determine the critical point D (n, 1- α) from the relevant table.
- Reject the null hypothesis if

$$Dmax > D (n, 1- \alpha).$$

The calculation of the KS Test for actual parking times (Actual1) is shown in table 3. In this table, the first column stands for t that counts from 1 up to the number of observations n (59 observations). Column 2 represents in ascending order the Actual1. The Sn(x) values are in column 3. Column 4, 5 and 6 state the values of exponential, gamma and lognormal cdf. Parameters for those cdf are obtained through the output from SPSS for the Q-Q plots. Those parameters are tabulated in table 2. The absolute value of the differences between exponential, gamma, lognormal cdf and Sn(x) are placed in column 7, 8, and 9 respectively and denoted by D1, D2, and D3.

In this KS Test, we take α = 5% and the critical value for D is 0,1771. In Table 3, we see that the maximum value of D1, D2, and D3 is 0.31, 0.19, and 0.13 respectively. Because only the maximum value of D3 is less than the critical value D, we conclude that the distribution of the actual parking time is lognormal with scale parameter 18.5725 and shape parameter 1.6387.

IV. Describing the Proportion of Used Parking Bays

The proportion of used parking bays is calculated by dividing the number of used parking bays at a certain time with the total number of parking bays (29 parking bays). What we mean by "certain time" is the time when there is a change in the number of cars in the parking area. The result is shown in figure 6.

Because the proportion of used parking bays is less than one at each time, we can conclude that there are still enough parking bays. On average there are 6.2463 cars/minute. This

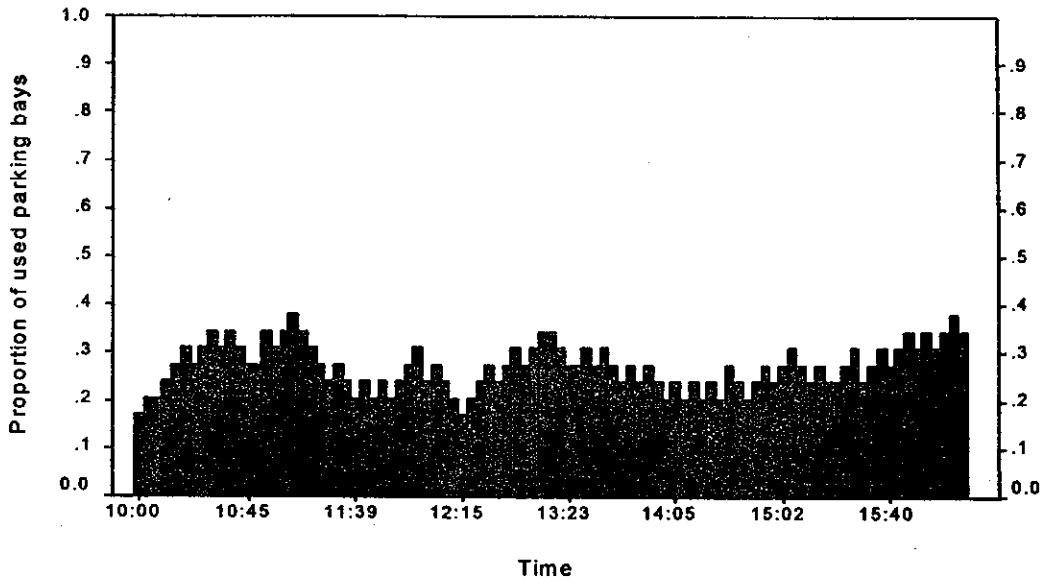

Figure 6. Bar Chart of Proportion of Used Parking Bays

values is calculated by summing up the area under the bar charts in figure 6 divide by the total time available for parking (10.00-17.00).

The proportion of used parking bays varies over time day from 0.17 up to 0.38 during the day considered.

V. Describing the Relation between Actual Parking Time and the Ticket Parking Time

Ticket parking time, which is the parking time one has paid for, is defined as the difference between time on ticket and arrival time. Because we want to describe the relation between actual parking time and ticket parking time, we are only concerned with data where one has paid in coins for the parking ticket. There are 17 data valid for this observation. First of all, we calculate from the data, the ticket parking time as shown in column 9, table 4. Column 8 in table 4 states the difference between departure time and time on ticket that will be used in further analysis.

For the second step, we make a scatter plot of ticket parking time and actual parking time (Figure 7). If we draw a line with actual parking time equal to ticket parking time, we can separate the data into two categories. The first category is for the data lying below the line, which means that the actual parking time is greater than the ticket parking time. These data show that people park their car longer than they should do. The second category is data that lie above the line, which means people leave the parking area earlier than the time on the ticket. From the scatter plot we see that 11 data lie below the line and 12 data lie above the line.

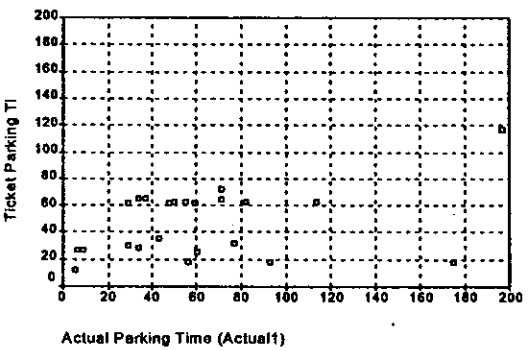

Figure 7. Scatter Plot: Parking Ticket Time vs Actual1

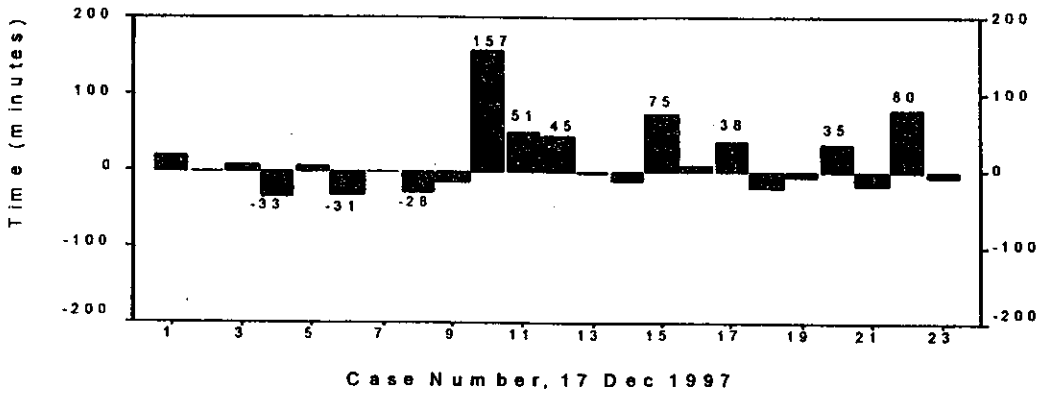

Figure 8. The Difference between departure time and time on parking ticket

In order to know the relation between actual parking time and ticket parking time, we perform the Pearson correlation and test whether there is a correlation between them. The correlation is 0.387 but this value is not significant at $\alpha = 5\%$, so there is no linear correlation between actual parking time and ticket parking time.

As mentioned before, we also calculate the difference between departure time and time on ticket. The graphical representation for these calculations is provided in figure 8. This figure gives information not only on the number of cars that park more or less than the time on the ticket, but also how long (in minutes) it differs from the time on ticket. According to the data, the number of cars that park more than the time on the ticket is quite the same with the number of cars that park less than the time on the ticket. But we take the time differences into account. It means we look at how long people leave the parking area before or after the time on the ticket.

From figure 8 we see that the differences are larger for people who leave after the time on the ticket. People who leave after the time on the ticket should pay more, but in reality they do not pay more. From point of view of the city authorities of the Rotterdam, these events should be reduced in the future. On average people park their car 14.91 minutes more than the time on ticket. This means that we can suggest to the city authorities of Rotterdam to increase the frequency of controlling the parking area in Hoogstraat.

VI. Conclusion

We can make several conclusions according to the analysis of the parking data set as follow:

- The distribution of the actual parking time on December 17, 1997 is lognormal with scale parameter 18.5725 and shape parameter 1.6387.
- The proportion of used parking bays varies over time during a day but there are still enough parking bays. This proportion varies over time day from 0.17 up to 0.38 during a day.
- There is no linear correlation between actual parking time and ticket parking time But, because there are still many people who park their cars more than the time on ticket, we suggest the city authorities of Rotterdam to improve the controlling function to reduce the occurrence of these events.

Table 1. Rotterdam Parking Bay Hoogstraat Data

Place	Arrival	Ticket	Time on ticket	No ticket	Licence	Departure	Plate	Actual (minutes)
1	10:00				1	14:10	VS-21-DK	250
1	15:16			1		15:30	GN-DJ-38	14
2	10:00				1	17:00	LL-77-LZ	420
3	10:38			1		10:45	SL-XS-16	7
3	12:59				1	13:16	PN-22-RR	17
3	15:02	1	16:05			16:24	RV-62-GV	82
4	11:29			1		11:32	LG-NR-79	3
5	10:50			1		10:52	YJ-73-BH	2
5	10:57	1	12:09			12:08	JV-XT-02	71
5	12:45	1	13:20			13:28	PZ-14-RJ	43
5	14:24			1		15:21	ZL-24-BT	57
5	15:54				1	17:00	NT-49-FV	66
6	10:05			1		10:08	VD-26-PH	3
6	11:02			1		11:09	JB-ZR-75	7
6	12:06			1		12:08	VS-VZ-18	2
6	12:39			1		12:40	NR-HN-33	1
6	13:08	1	14:10			13:37	SH-24-JL	29
6	13:56			1		13:58	RV-ZR-21	2
6	15:24	1	15:52			15:58	LG-XH-73	34
7	10:23			1		10:30	JB-ZR-75	7
7	10:45			1		10:47	LD-NL-08	2
7	10:47	1	11:52			11:21	SZ-59-YK	34
7	11:49			1		11:52	RN-RP-95	3
7	12:46			1		12:50	NH-RS-75	4
7	13:16			1		13:23	FT-SJ-24	7
7	14:28			1		14:31	DH-XG-41	3
7	14:38			1		14:40	RJ-95-VD	2
7	15:04			1		15:17	YB-02-TY	13
7	15:36			1		15:40	NJ-HJ-16	4
8	10:00				1	14:25	BN-60-XF	265
8	14:45	1	15:15			15:14	NB-BX-07	29
8	15:21			1		15:30	BY-92-JD	9
9	10:32			1		10:35	RB-HR-53	3
9	12:07			1		12:12	PD-ZS-99	5
9	15:31			1		17:00	NT-36-JS	89
10	10:00				1	17:00	VV-04-JJ	420
11	10:50	1	11:55			11:27	RD-NG-10	37
11	12:32	1	13:34			13:20	NR-01-PD	48
11	13:33			1		13:40	JN-FT-20	7
11	14:05	1	14:23			17:00	DZ-FT-29	175
12	10:08	1	11:11			12:02	XT-96-DT	114
12	12:05			1		12:10	RS-VD-61	5
12	15:43	1	16:15			17:00	NH-25-BY	77
13	10:09	1	11:11			11:08	JR-ZN-29	59
13	13:35	1	14:38			14:25	JJ-PH-03	50
13	15:27	1	15:45			17:00	RG-ZG-27	93

Table 1. Rotterdam Parking Bay Hoogstraat Data (continued)

Plate	Arrival	Ticket	Time on ticket	No ticket	Evidence	Departure	Plate	Actual (minutes)
14	10:28	1	11:32			11:39	HH-FH-57	71
14	12:09			1		12:15	PN-BL-04	6
14	12:38	1	12:56			13:34	LT-31-NK	56
14	14:38	1	15:05			14:44	ZN-30-GY	6
14	14:50			1		14:52	LV-76-KH	2
14	15:48	1	16:00			15:53	RT-89-TX	5
14	16:00	1	16:25			17:00	KB-27-BH	60
15	10:31	1	10:58			10:40	SG-LB-51	9
16	11:54	1	13:51			15:11	RK-18-KZ	197
17	16:05	1	17:08			17:00	LZ-RR-96	55
18	10:00						NO	
19	10:00						NO	
20	10:00						NO	
21	10:00						NO	
22	10:00				1	17:00	SJ-GD-87	420
22	13:47			1		13:57	TG-05-JH	10
23	10:00						NO	
24	10:00						NO	
25	10:00						NO	
26	13:28			1		13:48	RR-SP-09	20
27	10:00						NO	
28	10:00						NO	
29	10:00						NO	

Table 2. Parameters for Various Expected Cumulative Distribution Functions

CDF	Exponential	Gamma		Lognormal	
Parameters	Scale	Scale	Shape	Scale	Shape
ACTUAL	0.0164	0.00589	0.3582	18.5725	1.6387

Table 3. Calculation of KS Test for Actual Parking Time

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
t	Actual	$S_n(x)$	Exponential	Gamma	Lognormal	D1	D2	D3
1	1	0.02	0.02	0.18	0.04	0	0.16	0.02
2	2	0.03	0.03	0.23	0.09	0	0.19	0.05
3	2	0.05	0.03	0.23	0.09	0.02	0.18	0.04
4	2	0.07	0.03	0.23	0.09	0.04	0.16	0.02
5	2	0.08	0.03	0.23	0.09	0.05	0.14	0
6	2	0.1	0.03	0.23	0.09	0.07	0.13	0.01
7	2	0.12	0.03	0.23	0.09	0.09	0.11	0.03
8	3	0.14	0.05	0.26	0.13	0.09	0.13	0
9	3	0.15	0.05	0.26	0.13	0.1	0.11	0.02
10	3	0.17	0.05	0.26	0.13	0.12	0.09	0.04

Table 3. Calculation of KS Test for Actual Parking Time (Continued)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
i	Actual Time	$S_n(x)$	Exponential	Gamma	Lognormal	D1	D2	D3
11	3	0.19	0.05	0.26	0.13	0.14	0.08	0.05
12	3	0.2	0.05	0.26	0.13	0.16	0.06	0.07
13	4	0.22	0.06	0.29	0.17	0.16	0.07	0.05
14	4	0.24	0.06	0.29	0.17	0.17	0.05	0.06
15	5	0.25	0.08	0.32	0.21	0.18	0.06	0.04
16	5	0.27	0.08	0.32	0.21	0.19	0.04	0.06
17	5	0.29	0.08	0.32	0.21	0.21	0.03	0.08
18	6	0.31	0.09	0.34	0.25	0.21	0.03	0.06
19	6	0.32	0.09	0.34	0.25	0.23	0.01	0.08
20	7	0.34	0.11	0.35	0.28	0.23	0.02	0.06
21	7	0.36	0.11	0.35	0.28	0.25	0	0.08
22	7	0.37	0.11	0.35	0.28	0.26	0.02	0.1
23	7	0.39	0.11	0.35	0.28	0.28	0.04	0.11
24	7	0.41	0.11	0.35	0.28	0.3	0.05	0.13
25	9	0.42	0.14	0.39	0.33	0.29	0.04	0.09
26	9	0.44	0.14	0.39	0.33	0.3	0.05	0.11
27	10	0.46	0.15	0.4	0.35	0.31	0.06	0.1
28	13	0.47	0.19	0.44	0.41	0.28	0.04	0.06
29	14	0.49	0.21	0.45	0.43	0.29	0.04	0.06
30	17	0.51	0.24	0.48	0.48	0.27	0.03	0.03
31	20	0.53	0.28	0.51	0.52	0.25	0.02	0.01
32	29	0.54	0.38	0.57	0.61	0.16	0.03	0.06
33	29	0.56	0.38	0.57	0.61	0.18	0.01	0.05
34	34	0.58	0.43	0.6	0.64	0.15	0.02	0.07
35	34	0.59	0.43	0.6	0.64	0.17	0.01	0.05
36	37	0.61	0.45	0.62	0.66	0.16	0.01	0.05
37	43	0.63	0.51	0.64	0.7	0.12	0.02	0.07
38	48	0.64	0.54	0.67	0.72	0.1	0.02	0.07
39	50	0.66	0.56	0.67	0.73	0.1	0.01	0.07
40	55	0.68	0.59	0.69	0.75	0.08	0.01	0.07
41	56	0.69	0.6	0.7	0.75	0.09	0	0.05
42	57	0.71	0.61	0.7	0.75	0.1	0.01	0.04
43	59	0.73	0.62	0.71	0.76	0.11	0.02	0.03
44	60	0.75	0.63	0.71	0.76	0.12	0.04	0.02
45	66	0.76	0.66	0.73	0.78	0.1	0.04	0.02
46	71	0.78	0.69	0.74	0.79	0.09	0.04	0.01
47	71	0.8	0.69	0.74	0.79	0.11	0.06	0
48	77	0.81	0.72	0.76	0.81	0.1	0.06	0.01
49	82	0.83	0.74	0.77	0.82	0.09	0.06	0.01
50	89	0.85	0.77	0.78	0.83	0.08	0.06	0.02
51	93	0.86	0.78	0.79	0.84	0.08	0.07	0.03
52	114	0.88	0.85	0.83	0.87	0.04	0.05	0.02
53	175	0.9	0.94	0.9	0.91	0.04	0	0.02
54	197	0.92	0.96	0.92	0.93	0.05	0	0.01
55	250	0.93	0.98	0.95	0.94	0.05	0.01	0.01
56	265	0.95	0.99	0.95	0.95	0.04	0	0

Table 3. Calculation of KS Test for Actual Parking Time (Continued)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
i	Actual Ti	FS(x)	Exponential	Gamma	Lognormal	D1	D2	D3
57	420	0.97	1	0.98	0.97	0.03	0.02	0.01
58	420	0.98	1	0.98	0.97	0.02	0	0.01
59	420	1	1	0.98	0.97	0	0.02	0.03

Table 4. Calculation of Ticket Parking Time

Place	Arrival Time	Ticket	Time on Ticket	Departure	Plate	Actual T	Diff	Ticket P. Time
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
3	15:02:00	1	16:05:00	16:24:00	RV-62-GV	82	19	63
5	10:57:00	1	12:09:00	12:08:00	JV-XT-02	71	-1	72
5	12:45:00	1	13:20:00	13:28:00	PZ-14-RJ	43	8	35
6	13:08:00	1	14:10:00	13:37:00	SH-24-JL	29	-33	62
6	15:24:00	1	15:52:00	15:58:00	LG-XH-73	34	6	28
7	10:47:00	1	11:52:00	11:21:00	SZ-59-YK	34	-31	65
8	14:45:00	1	15:15:00	15:14:00	NB-BX-07	29	-1	30
11	10:50:00	1	11:55:00	11:27:00	RD-NG-10	37	-28	65
11	12:32:00	1	13:34:00	13:20:00	NR-01-PD	48	-14	62
11	14:05:00	1	14:23:00	17:00:00	DZ-FT-29	175	157	18
12	10:08:00	1	11:11:00	12:02:00	XT-96-DT	114	51	63
12	15:43:00	1	16:15:00	17:00:00	NH-25-BY	77	45	32
13	10:09:00	1	11:11:00	11:08:00	JR-ZN-29	59	-3	62
13	13:35:00	1	14:38:00	14:25:00	JJ-PH-03	50	-13	63
13	15:27:00	1	15:45:00	17:00:00	RG-ZG-27	93	75	18
14	10:28:00	1	11:32:00	11:39:00	HH-FH-57	71	7	64
14	12:38:00	1	12:56:00	13:34:00	LT-31-NK	56	38	18
14	14:38:00	1	15:05:00	14:44:00	ZN-30-GY	6	-21	27
14	15:48:00	1	16:00:00	15:53:00	RT-89-TX	5	-7	12
14	16:00:00	1	16:25:00	17:00:00	KB-27-BH	60	35	25
15	10:31:00	1	10:58:00	10:40:00	SG-LB-51	9	-18	27
16	11:54:00	1	13:51:00	15:11:00	RK-18-KZ	197	80	117
17	16:05:00	1	17:08:00	17:00:00	LZ-RR-96	55	-8	63

References

[1] Rice, John A., "*Mathematical Statistics and Data Analysis*", Second Edition, Duxbury Press, Belmont, California, 1995.

[2] Conover, W.J., "*Practical Nonparametric Statistics*", John Wiley, New York, 1971.

Author

Dharma Lesmana, S.Si, M.T and Agus Sukmana, S. Si are lecturers at the Department of Mathematics, Catholic University of Parahyangan.