

Factors Affecting the Punctuality of Logistics Services Using Binary Logistic Regression

Biwei Li*

Faculty of Information Technology, Monash University, Melbourne, Australia

*Corresponding author: blii0041@student.monash.edu

Abstract. The Indian logistics industry is in a period of rapid development. The purpose of this study is to analyze the factors affecting the punctuality of trucks for logistics distribution services in India based on data analysis. The data comes from 6880 pieces of data from the VTS data platform. A binary regression model was established for data analysis, and the model was tested to ensure statistical significance. Binary logistic regression was used to examine truck types, geographic locations, commodity types, order types, and suppliers for logistics services to examine significant differences between on-time and late. Binary logistic regression concluded that the shorter the distance, the signing of the perfect contract, and the selection of a specific type of vehicle, the better development of the logistics level is easier to be on time. This study provides suggestions for improving management strategies, which can be used as a reference for the development of logistics companies.

Keywords: punctuality; logistics distribution services; VTS data platform; binary logistic regression; management strategies.

1. Introduction

1.1 Background

India is one of the most populous countries in the world, and various industries are growing very rapidly. India's logistics industry is also developing rapidly, which is the focus of attention. World-renowned logistics service companies have established logistics networks in India. Since 2017, the compound annual growth rate has remained above 10.5%, and the market size in 2020 will exceed 215 billion US dollars [1]. The industry has become a strong support for the current economic growth of India. With the development of e-commerce platforms, more industries and competitors are joining the logistics network, and the quality of logistics services and customer satisfaction are becoming more and more important.

1.2 Related research

With regard to logistics services and punctuality, scholars from various countries have conducted research from many different perspectives.

Some researchers have studied the relationship between truck drivers' services and user purchase behavior, such as Claye-Puau et al. (2019) discussed the relationship between truck drivers and work: 34% of truck drivers do not like their profession and do not want to maintain customer relationships [2]. They refuse to build better customer relationships and only keep their jobs at a minimum. This could result in consumers with longer and more complex addresses and contact details being considered last in the same circumstances. A more in-depth study of this issue provides an in-depth analysis of how truck drivers influence customer buying behavior in the supplier-customer relationship [3]. The service quality of truck drivers will directly affect consumers' purchasing behavior. The degree of influence will also weaken as the size of the buyer's enterprise increases.

Some scholars have studied and analyzed the evaluation system of logistics services. Wang (2019) proposed an improved method for existing evaluation metrics based on the SERVQUAL scale [4]. Akıl & Ungan (2021) analyzed e-commerce data and structural equation modeling and found that the accuracy and timeliness of orders have a positive impact on customer satisfaction and customer loyalty [5].

In addition to studying the influencing factors of logistics services, scholars have also studied the impact and changes of the times and the environment on the logistics industry. Mehroliia & Solaiikutty (2021) used logistic regression to analyze the threat to food delivery services during COVID-19. Combined with the theory of self-protection, user behavior is summarized and analyzed [6]. Bookbinder & Springer (2013) analyzed the opportunities and challenges faced by Asian logistics development and discussed in detail how developing countries should develop their own logistics systems [7].

The purpose of the current research is primarily to examine the influence of the entire logistics service, the relationship between customer happiness and the quality of the logistics service, and the research on the current quality evaluation system. Research on the relationships between influencing factors and the quality of logistics services is lacking. The study of the variables influencing the standard of logistics services in India mostly focuses on firm size and consumer groups. For example, Arlbjørn (2010) found that the economic level is the determinant of the development level of the express delivery industry [1]. Abraham (2011) used the value-added model analysis method to find that the age, gender, and transportation mode of the main consumer groups will have an impact on the express delivery industry [8].

Existing research has not clearly defined the influencing factors of logistics service quality, and is only limited to the exploration of influencing factors and the research on the evaluation system of logistics service. relation. Based on the research perspective of the influencing factors of logistics service quality, this paper uses the logistic regression model to explore the influencing factors of the punctuality of logistics services and ensures the accuracy of the research results through the stability test.

1.3 Objection

Analyzing the punctuality of logistics services in India in view of several contributing factors is the goal of this study. The study examines different influencing factors of logistics services, such as contract type, distance, truck type, geographic location, commodity type, and supplier.

The structure of this paper is as follows. The literature review is covered in the study's first section, which focuses in particular on truck drivers' self-awareness and the variables influencing the caliber of logistics services. The study's next section describes the data processing and research methods. Section three provides detailed data analysis. The discussion and results of the research are covered in the fourth section. In the last part, the study concludes by summarizing its shortcomings and potential research directions.

2. Methods

2.1 Data collection

The order status of delivery trucks was considered as the main study group for this study. The data mainly records the data of delivery trucks on the VTS data platform from March 28, 2019, to August 17, 2020. No personal information was collected in this study, and the dataset was obtained from Kaggle. The dataset contains 6880 pieces of data, mainly including location information, punctuality information, supplier information, truck information, and commodity information. The content covers most of India and covers as many common commodity types as possible. Therefore, this dataset can serve as an ideal research background.

2.2 Data preprocessing

Make a deep copy of the data set, view the basic data types, and describe the numerical data statistically. The statistical results of the data set are shown in Fig. 1. There are a total of 6880 data, the missing rate is 11%, and there are 4 types of values. Remove highly correlated and useless eigenvalues as required by the regression model. For example, the longitude and latitude of the coordinates and the order of location information affect each other. And columns with uniqueness,

such as the ID of the order, etc. Columns with too high missing data also need to be deleted, as shown in Fig. 2 and Fig. 3. For missing values, use the median or mean as necessary, and use 'Unknown' instead for character data. Finally, the data is case converted to ensure the correct classification, and dummy variables are set.

Dataset statistics		Variable types	
Number of variables	32	CAT	19
Number of observations	6880	DATE	6
Missing cells	25125	NUM	4
Missing cells (%)	11.4%	UNSUPPORTED	3
Duplicate rows	0		
Duplicate rows (%)	0.0%		
Total size in memory	1.7 MiB		
Average record size in memory	256.0 B		

Fig. 1 Dataset statistics (Photo credit :Original)

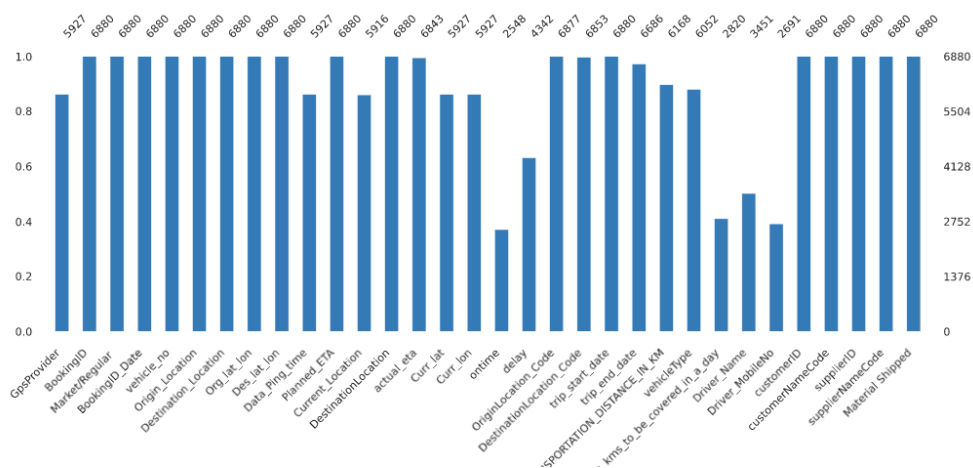


Fig. 2 Graph for missing data (Photo credit :Original)

trip_end_date has 194 (2.8%) missing values	Missing
TRANSPORTATION_DISTANCE_IN_KM has 712 (10.3%) missing values	Missing
vehicleType has 828 (12.0%) missing values	Missing
Minimum kms_to_be_covered_in_a_day has 4060 (59.0%) missing values	Missing
Driver_Name has 3429 (49.8%) missing values	Missing
Driver_MobileNo has 4189 (60.9%) missing values	Missing
BookingID is uniformly distributed	Uniform
OriginLocation_Code is an unsupported type, check if it needs cleaning or further analysis	Unsupported
DestinationLocation_Code is an unsupported type, check if it needs cleaning or further analysis	Unsupported
supplierID is an unsupported type, check if it needs cleaning or further analysis	Unsupported

Fig. 3 Missing data (Photo credit :Original)

2.3 Logistic regression model

Logistic Regression, which is essentially linear regression, maps the result of a linear function to a sigmoid function, as shown in function (1). The logistic regression model directly models the probability of classification, which can effectively avoid the problem of inaccurate distribution of assumptions. It can obtain approximate probability forecasts rather than just "categories" predictions,

which is helpful for many jobs where probability is needed to aid in decision-making; The logistic regression model can utilize several strategies to obtain the best answer and has good mathematical features. The binary logistic regression model is faster to compute and has a much smaller memory footprint than other models.

$$\pi(x) = 1 + \frac{1}{e^{-w^T x}} \quad (1)$$

3. Results

Based on the model output results are analyzed. It is decided whether the OR value of the variable is statistically significant, or whether the model as a whole is meaningful, based on the thorough evaluation of the model coefficients. The Hosmer and Lemeshow Test evaluates the model's goodness of fit and decides whether the data's information has been sufficiently extracted. Variables in the Equation determine how much different variables affect the outcome.

3.1 Omnibus Inspection

Comprehensive assessments of model coefficients are known as omnibus tests. The Model line outputs the results of the likelihood ratio test to determine whether all parameters in the Logistic Regression Model are equal to 0. $P < 0.05$ indicates that the model fitted this time has an OR value of at least one variable that is statistically significant, indicating that the model is overall meaningful. The specific results and values are shown in Table 1.

Table1. Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	4653.051	923	.000
	Block	4653.051	923	.000
	Model	4653.051	923	.000

3.2 Model Summary

To quantitatively evaluate the effect of the goodness of fit, the -2 pair likelihood value is an important index for model evaluation, which is 3098. The smaller the value, the better, and can be used to analyze the modeling effect of the model. The percentage of the dependent variable's variance that the fitted model can explain is shown in the Model Summary table. The table contains Cox & Snell R Square and Nagelkerke R Square, these two R2s are sometimes referred to as pseudo R2s, which have little meaning in logistic regression (unlike in linear regression) and can be ignored [9]. The specific results and values are shown in Table 2.

Table 2. Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	3098.351 ^a	.530	.740

Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

3.3 Hosmer and Lemeshow Test

The Hosmer and Lemeshow test is one of the longest used methods when it comes to assessing the goodness of fit of a model. The specific results and values are shown in Table 3 & 4. It is believed that the information in the present data has been thoroughly retrieved and the model has a high goodness of fit when the P value is not less than the test level (i.e., $P > 0.05$) [9]. The p-value of this study is $0.125 > 0.05$. Through the HL test, there is no need to consider reducing the dimension of variables or increasing the sample size.

Table 3. Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	12.648	8	.125

Table 4. Contingency Table for Hosmer and Lemeshow Test

Step 1						
	OnTime = 0		OnTime = 1		Total	
	Observed	Expected	Observed	Expected		
1	617	617.000	0	.000	617	
2	616	616.000	0	.000	616	
3	612	612.500	2	1.500	614	
4	582	594.534	35	22.466	617	
5	542	534.569	67	74.431	609	
6	492	486.011	123	128.989	615	
7	434	416.407	183	200.593	617	
8	254	268.282	363	348.718	617	
9	33	36.697	581	577.303	614	
10	0	.000	632	632.000	632	

3.4 Variable analysis

To test the research objectives, the binary logistic regression results are shown below (Table 5-8). In this study, whether the goods were delivered on time or not, the resulting results were treated as dependent variables (0- not delivered on time, 1- delivered on time); distance of transport, type of transport truck, region where the truck is located, supplier ID, the type of goods shipped, the status of the contract with the supplier were taken to be the predictor variables.

Table 5 also provides information on the impact of independent variables affecting on-time delivery (see odds ratio [OR]). Distance ($b=-0.001$, $p=0.001$, $p<0.01$), contract type ($b=1.958$, $p=0.044$, $0.01<p<0.05$), p -values meet the criteria and are statistically significant, contract type (Market/ Regular) has a positive regression slope, indicating that contracted suppliers are more likely to deliver goods on time. The regression slope for distance is negative, indicating that orders with greater distances are more difficult to be on time. The odds ratios of the predictors show that the odds ratio of orders being delivered on time varies by a factor of 0.999 with the increase in the raw score of the overall distance of the order, and by a factor of 7.084 with the increase in the raw score of the contract type.

Table 5. Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	
Step 1 ^a	Market/Regular (1)	1.958	.973	4.052	1	.044	7.084
	TRANSPORTATION_DISTANCE_IN_KM	-.001	.000	10.321	1	.001	.999
	vehicleType			43.297	22	.004	
	vehicleType(1)	-72.214	22766.195	.000	1	.997	.000
	vehicleType(2)	-20.651	40192.970	.000	1	1.000	.000
	vehicleType(3)	-2.304	24337.283	.000	1	1.000	.100
	vehicleType(4)	-36.036	22318.730	.000	1	.999	.000
	vehicleType(5)	-70.821	22766.195	.000	1	.998	.000
	vehicleType(6)	-41.049	22318.730	.000	1	.999	.000
	vehicleType(7)	-41.857	22318.730	.000	1	.999	.000

Variables VehicleType (DF=22, $p=0.004$, $p<0.01$), Veh_stat (DF=27, $p=0.000$, $p<0.01$), supplierID (DF=276, $p=0.714$, $p>0.05$), MaterialShipped (DF= 596, $p=1.000$, $p>0.05$) is a categorical

variable, and dummy variables are set. Among them, the p-values of the truck type and the region where the trucks are located are as expected and statistically significant. The specific values are shown in Table 6-8. Compared with the unknown model as a standard quantity, the b values of several variables such as 32FT(b=-72), 28FT(b=-71), 1MT(b=-2), and 37FT(b=-4) are more obvious. The standard quantities in the area where the truck is located are KA, UK (b=20), AP (b=-20), AS (b=-15) and other variables with obvious b values.

However, the types and suppliers of transported goods are There is no significant effect on the punctuality of commodity shipments in this model. Therefore, the type and supplier of the shipped item are not important predictors of whether the item will be delivered on time, so the type and supplier of the shipped item are not supported.

Table 6. Variables of stat

veh_stat	65.87727 .000					
veh_stat(1)	-1.004	1.075	.871	1	.351	.366
veh_stat(2)	.254	1.183	.046	1	.830	1.289
veh_stat(3)	-.010	.882	.000	1	.991	.990
veh_stat(4)	-1.509	1.791	.710	1	.399	.221
veh_stat(5)	3.053	14730.839	.000	1	1.000	21.183
veh_stat(6)	-3.118	.859	13.181	1	.000	.044
veh_stat(7)	1.513	1.555	.946	1	.331	4.540
veh_stat(8)	.026	.772	.001	1	.973	1.026
veh_stat(9)	.343	.959	.128	1	.721	1.409
veh_stat(10)	-1.991	.700	8.099	1	.004	.137

Table 7. Variables of supplier

supplierID_8	262.264276 .714					
supplierID_8(1)	25.714	39765.713	.000	1	.999	147021598986.278
supplierID_8(2)	-2.568	48845.377	.000	1	1.000	.077
supplierID_8(3)	-5.450	48845.377	.000	1	1.000	.004
supplierID_8(4)	-1.695	37705.239	.000	1	1.000	.184
supplierID_8(5)	47.653	28117.073	.000	1	.999	495700038298350800000.000
supplierID_8(6)	41.321	48845.376	.000	1	.999	881734376964027900.000
supplierID_8(7)	-1.900	48845.377	.000	1	1.000	.150
supplierID_8(8)	32.243	48845.376	.000	1	.999	100653750780288.190
supplierID_8(9)	-1.025	39643.539	.000	1	1.000	.359
supplierID_8(10)	41.004	48845.376	.000	1	.999	642720636765848700.000

Table 8. Variables of material

Material Shipped_8	330.509596 1.000					
Material Shipped_8(1)	-40.920	39848.053	.000	1	.999	.000
Material Shipped_8(2)	-7.981	40466.273	.000	1	1.000	.000
Material Shipped_8(3)	-34.886	49395.631	.000	1	.999	.000
Material Shipped_8(4)	-20.495	28370.656	.000	1	.999	.000
Material Shipped_8(5)	-42.422	49197.246	.000	1	.999	.000
Material Shipped_8(6)	-40.990	34631.552	.000	1	.999	.000
Material Shipped_8(7)	-21.573	51075.320	.000	1	1.000	.000
Material Shipped_8(8)	-21.342	28370.656	.000	1	.999	.000
Material Shipped_8(9)	-22.451	28370.656	.000	1	.999	.000
Material Shipped_8(10)	-22.455	28370.656	.000	1	.999	.000

3.5 Model accuracy analysis

The probability of the result event in this case, on-time delivery can be determined for any combination of independent variables after the logistic regression model has been fitted. The regression model will think that the courier has arrived on time (when the probability of the time occurrence is greater than 0.5); if the probability is less than 0.5, it is judged that the event has not occurred (not delivered on time). Therefore, the prediction effect of the logistic regression model can be evaluated compared with the real situation [10].

The logistic regression model in this study was able to correctly classify 88.8% of the observed samples, as expected. This metric is often referred to as percentage accuracy in classification [10]. The specific values are shown in Table 9. 95.6% of cases of late delivery were predicted by the model to be late. 74.3% of on-time situations were correctly predicted by the model to be on-time.

Table 9. Classification Table^a

Table 3. Classification Table				
Observed		Predicted		
		OnTime		Percentage Correct
		0	1	
Step 1	OnTime	0 4000	182	95.6
		1 510	1476	74.3
	Overall Percentage			
a. The cut value is .500				

a. The cut value is .500

4. Discussion and Implications

In the research, a fluky model of logistic regression was successfully developed with the main purpose of distinguishing the factors affecting the punctuality of express delivery. This study concluded that contract type, truck type, truck location, transportation distance, supplier, and commodity type were the influencing factors for group differences. In other words, on-time delivery of logistics items was associated with less distance of delivery and have a contract. In addition, the type and location of the truck also had an impact on timing. Some substantial marketing implications can be explored and studied, from result.

Research conducted by Clys-Puau et al. (2019) shows that truck drivers are reluctant to actively maintain customer relationships, and the more complex the customer conditions, the greater the negative impact [2,11]. This is consistent with the model results, when the distance is farther, the probability of the on-time arrival of the logistics is lower. Akıl & Ugan (2021) believe that the punctuality of orders has a positive relationship with customers' purchasing behavior and loyalty [5]. This is consistent with the findings that orders with cyclical contracts are significantly more punctual than those without contracts. The research of Arlbjørn (2010) shows that the level of economic development in a region will determine the level of logistics services in the region [1,12]. This conclusion is not completely consistent with the research results of the model. The per capita GDP rankings of AP, AS, and UK provinces at the extreme point are 19, 30, and 8 respectively [13]. Economically developed regions have better punctuality of logistics services than underdeveloped regions, but this is not absolute. The experimental results can only prove the existence of Positive influence trend, but only as a reference condition.

Based on these findings, the punctuality of logistics can be improved in the following aspects. Many logistics companies are already improving incentives to boost employee motivation, but it's not enough. Logistics enterprises need to build a more complete management model [14]. It is very necessary to actively change the management mode. In addition to adopting a more reasonable reward and punishment system, it is also necessary to apply the idea of quality management to truck drivers, service personnel, and work processes. Let the service personnel who directly connect with customers think more about customers and take improving service quality as the first goal. In addition, it is also necessary to strengthen the supervision of employees on the quality of logistics services, provide

better logistics services, and establish standardization of logistics services. Make the logistics services of enterprises more standardized and rationalized. Logistics service providers can provide service evaluation procedures. For the evaluation content, the logistics company should check it in real time to ensure that the service of each order is supervised. This will greatly reduce the driver's resistance to long-distance orders due to negative emotions, and increase the on-time rate of express delivery services.

With the development of the logistics industry, the competitiveness and service standards of logistics enterprises are constantly increasing. Logistics enterprises should actively use advanced equipment and technology to better meet the requirements of logistics services [6]. It can mainly be reflected in more suitable vehicle configurations, better trucks, and better query services. More efficient selection of truck types and better truck configuration for the company can improve the punctuality of logistics services. Just like Express Mail Service, through the introduction of advanced management technology and modern equipment (mainly reflected in logistics equipment), it has effectively improved the efficiency of logistics services, improved management capabilities and query capabilities. Greatly improved service quality and customer satisfaction [2, 15].

5. Conclusion

By preprocessing the data and testing the model, the overall percentage >60% represents a good fit, and the test value of $P > 0.05$ means that the established model is reasonable for the data. It means that the results reflect the true relationship between the data. Therefore, the model is statistically significant. Based on the analysis of data on logistics services in India, this study found the main influencing factors. The choice of truck drivers and equipment has a great impact on the punctuality of logistics services. Better management models and employee training, as well as better equipment, which can improve the customer satisfaction and quality of logistics services at the same time. Therefore, logistics companies can consider the potential of these factors to improve competitiveness.

The limitations of this study that can be addressed in future studies. Here, the author takes the data of the VTS platform as the target group, which does not reflect the current situation in India as a whole. We should add more data from other platforms or companies to better analyze the factors that affect the punctuality of logistics. Due to a large number of samples and the serious missing of some data, some data have been removed, and the research results cannot represent the real situation and can only be used as a reference. The model predicted the punctuality of logistics and suggested adding other considerations such as time constraints on orders, whether there are delay penalties, driver earnings and bonuses to develop more important features.

References

- [1] Arlbjørn J. (2010). Supply chain management: Business operations in India. Odense: University Press of Southern Denmark.
- [2] Claye-Puaux, Sophie, Jacquemier-Paquin, Laure, & Jeanpert, Sophie. Delivery truck drivers' professional commitment profiles: Customer facing matters. Supply Chain Forum, 2019, 20(3): 202-214.
- [3] Bode, Christoph, Lindemann, Eckhard, & Wagner, Stephan M. Driving Trucks and Driving Sales? The Impact of Delivery Personnel on Customer Purchase Behavior. Journal of Business Logistics, 2011, 32(1): 99-114.
- [4] Wang Y Y. China's Third-Party Logistics Service Quality Management Improvement. Academic Journal of Engineering and Technology Science, 2019, 2: 66-72.
- [5] Akıl, Siber, & Ungan, Mustafa Cahit. E-Commerce Logistics Service Quality: Customer Satisfaction and Loyalty. Journal of Electronic Commerce in Organizations, 2021, 20(1): 1-19.
- [6] Mehroliya, Sangeeta, Alagarsamy, Subburaj, & Solaikutty, Vijay Mallikraj. Customers response to online food delivery services during COVID-19 outbreak using binary logistic regression. International Journal of Consumer Studies, 2021, 45(3): 396-408.

- [7] Bookbinder, J., & Springer Science Business Media. (2013). Handbook of Global Logistics Transportation in International Supply Chains (1st ed. 2013. ed., International Series in Operations Research & Management Science, 181).
- [8] Abraham, Nandita. The apparel aftermarket in India - a case study focusing on reverse logistics. *Journal of Fashion Marketing and Management*, 2011, 15(2): 211-227.
- [9] Stoltzfus, Jill C. Logistic Regression: A Brief Primer. *Academic Emergency Medicine*, 2011, 18(10): 1099-1104.
- [10] Norton, Edward C, Dowd, Bryan E, & Maciejewski, Matthew L. Marginal Effects—Quantifying the Effect of Changes in Risk Factors in Logistic Regression Models. *JAMA: The Journal of the American Medical Association*, 2019, 321(13): 1304-1305.
- [11] Folinas, D., & Ebrary, Inc. (2013). Outsourcing management for supply chain operations and logistics services. Hershey, PA: Business Science Reference.
- [12] Wood, L., & IGI Global. (2018). Contemporary approaches and strategies for applied logistics.
- [13] Reserve Bank of India. (2018). Annual Report.
- [14] Ziółkowski, Jarosław, & Borucka, Anna. Markov model in logistic management of enterprise / Model Markowa w logistycznym zarządzaniu przedsiębiorstwem. *Journal of KONBiN*, 2016, 38(1): 271-290.
- [15] Mengistu Bogale Ayele, Habtamu Bekele, & Mesfin Workineh. The Influence of Express Mail Service Quality on Customer Satisfaction: Evidence from EMS Ethiopia. *International Journal of Marketing and Business Communication*, 2020, 9(1&2), 1.