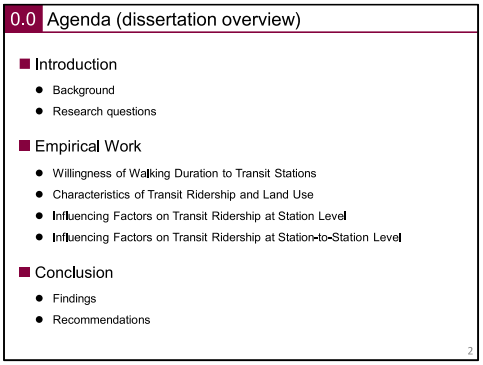
I'm Chen Qi, from Zhao laboratory. Thanks for everyone coming to my dissertation defense.

let me begin my presentation. The title of my dissertation is Research on Determinants of Rail Transit Ridership, taking Fukuoka, Japan as a Study Case.



Here a little bit about the agenda for quickly running into the presentation. This dissertation is mainly constructed by 3 parts, the introduction, the empirical works, and the conclusion.

Chapter 1

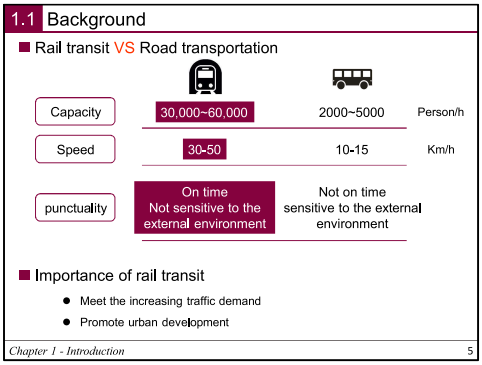
Introduction

The first chapter, introduction

1.1 Background

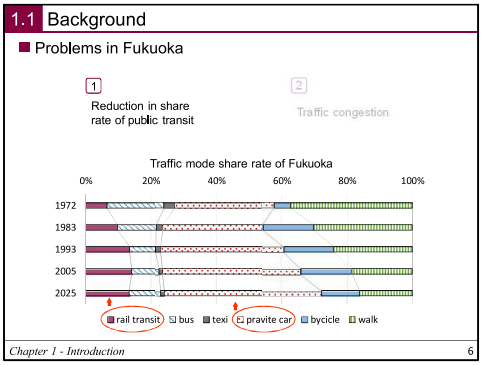
- Beginning

I'd like to begin my topic with a photograph. We can see a quite serious traffic congestion in this slide. The express way is also blocked due to the snow. If you are blocked on the way home after work, how do you feel about this situation? Of course, I think it should not be a good experience.



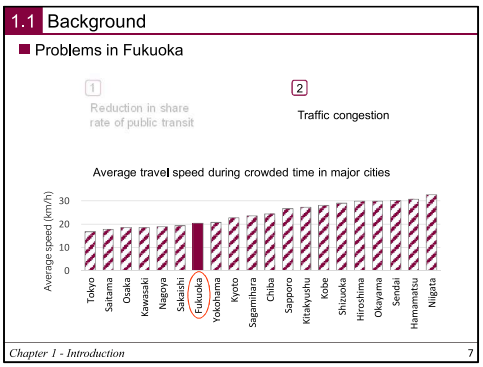
Is there any way to deal with this congestion problem?

The rail transit seems to be a good choice, since it has the features of large capacity, high speed, and punctuality. Because of these advantages, rail transit now is playing a more and more important role in the urban transportation system

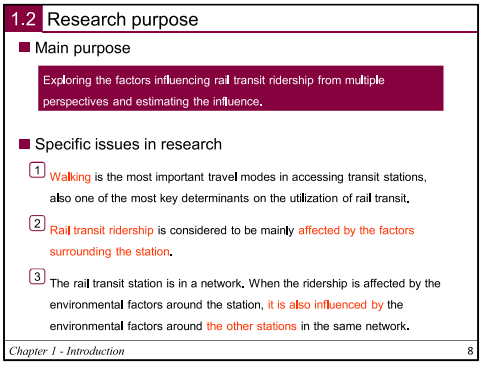


Let's take a look at Fukuoka, there is also some problems in the urban transit system.

One is the reduction in the share rate of public transit. As we can see from this chart, the share rate of rail transit tends to decrease in the future, while the share rate of private car keeps on increasing. It's not a good phenomena.



Another problem is the traffic congestion that we are now facing every day. We can see from this chart the average travel speed during the crowed time in Fukuoka now is at a very low level.

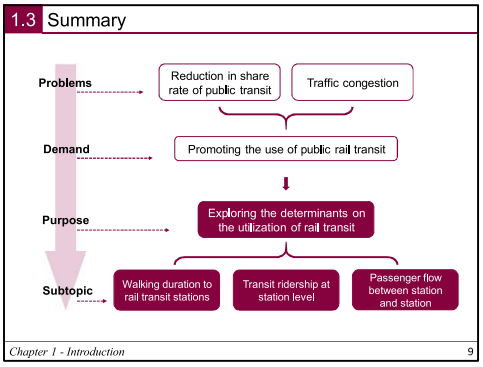


So is there anything we can do to help Fukuoka to promote the utilization of rail transit and to improve the traffic condition?

To increase the utilization of rail transit, obviously, knowing the influencing factors of transit ridership should come the first.

So, the main purpose could be, to explore the factors influencing rail transit ridership and estimate the influence.

For this purpose, there are some specific issues. The first is walking access to transit stations, second is the influencing factors of transit ridership at a single station, third is influencing factors of transit ridership at station-to-station level.

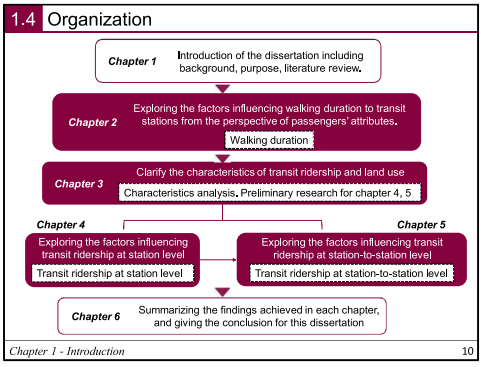


I will give a summary of the introduction.

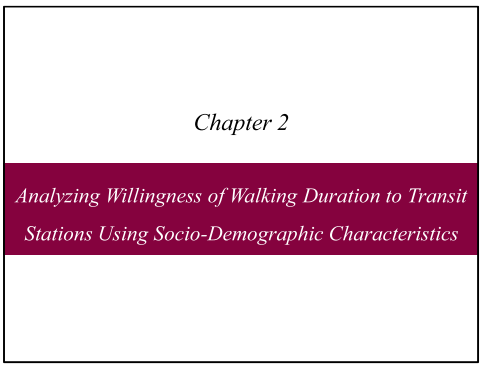
Now Fukuoka is facing the problem of reduction in share rate of public transit, and traffic congestion. So, promoting the use of public rail transit seems to be a good choice for Fukuoka.

With this background and demand, the purpose of exploring the determinants on the utilization of rail transit is proposed.

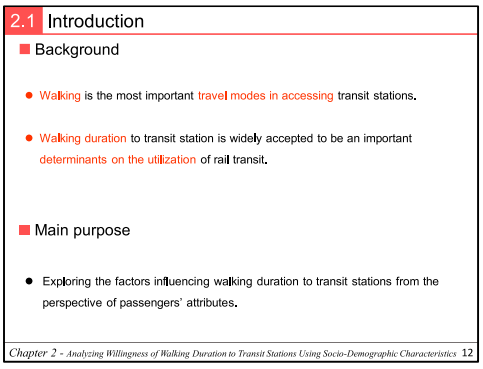
Around this main purpose, the research contents are extended to 3 subtopics walking duration to rail transit stations, Transit ridership at station level, passenger flow between station and station.



This research is organized into 6 chapters. Chapter 2, 4, 5 correspond to the three subtopics, chapter 3 is a preliminary study for supporting chapter 4 and 5.

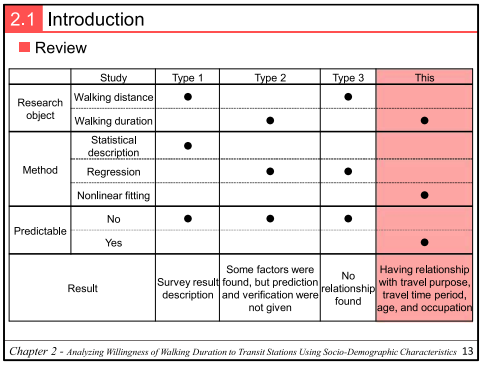


Chapter 2, Analyzing Willingness of Walking Duration to Transit Stations Using Socio-Demographic Characteristics.



Walking is the most important travel modes of accessing transit stations. The Walking duration to transit station is also widely thought to be an important determinants on the use of rail transit.

With this background, the purpose of this study can be proposed. That is to explore the factors influencing walking duration to transit stations from the perspective of passengers' attributes.

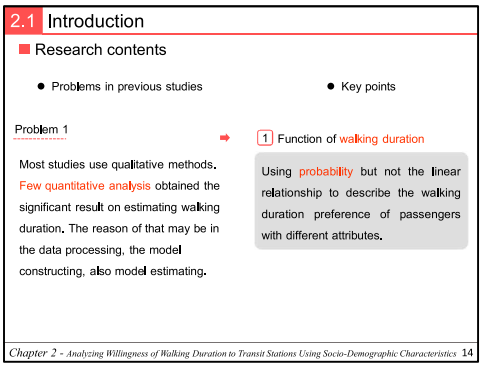


Before running the analysis, let's take a look at how this study differs from the others.

The research object in this study is the walking duration to transit stations.

As to the methods, studies based on statistical description account for the majority, while some studies tried to use regression model to find the relationship between walking duration and personal attributes, but they didn't get an expected result.

This study uses the nonlinear fitting method to find the relationship of walking duration and personal attributes, then try to make a prediction for the walking duration.

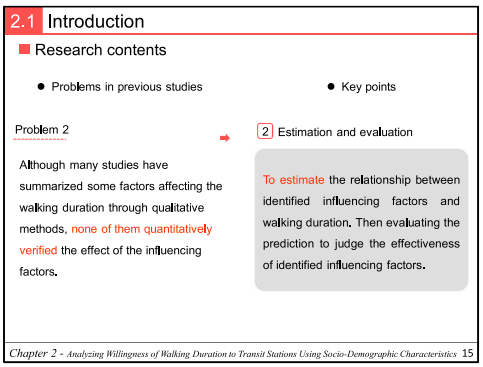


Ok, now let's have a look at the research contents.

According to the review of previous studies, some existing problems are summarized. This study will focus on some key points based on these problems.

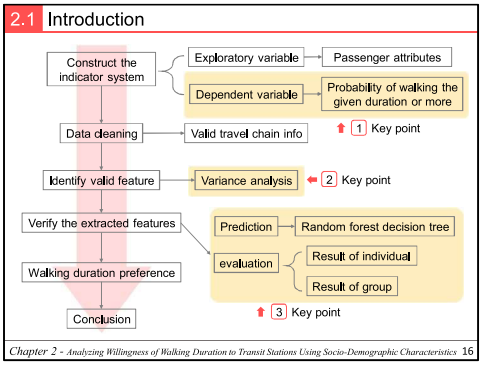
The first problem is few quantitative analyses obtained a significant result, it means they didn't find a proper function to describe the relationship between walking duration and personal attributes.

For this problem, two key points will be addressed in this study. one, construct a proper function to describe the relationship of walking duration and the indicators. two, identify the effective indicators that indeed have influences on the preference to walking duration.



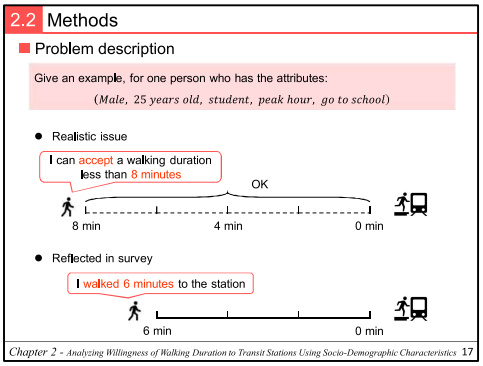
Another problem is the effect of indicators cannot be quantified. It means the result cannot be used for prediction.

So the third key point in this study is to estimate the relationship between identified influencing factors and walking duration. Then use this obtained relationship to make a prediction.



This is the flowchart of this study. It is organized into 6 parts. They are constructing the indicator system, cleaning data, identifying valid features, verifying the valid features, and estimating walking duration preference. The conclusion comes the last.

The key points of this studies are marked with yellow block.

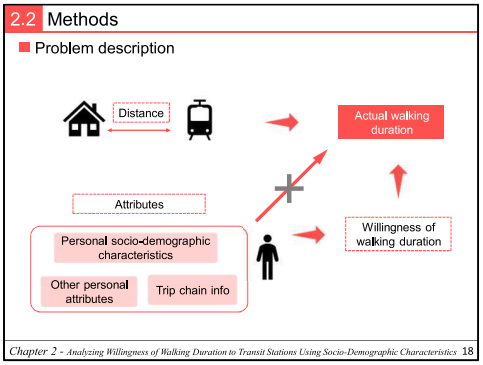


Next, I will give a simple example to explain how I deal with the walking duration.

Think about a people who has the attributes like this, we suppose this passenger can accept a walking duration less than 8 minutes for accessing transit stations.

But if it is reflected in a questionnaire, the answer may be like "I walked 6 minutes to the station". Obviously, it cannot represent how long he is willing to walk to the station. This 6 minutes is just the distance between his starting point and the station.

Here the deviation appeared. The next slide, I will continue to explain how to deal with this problem.

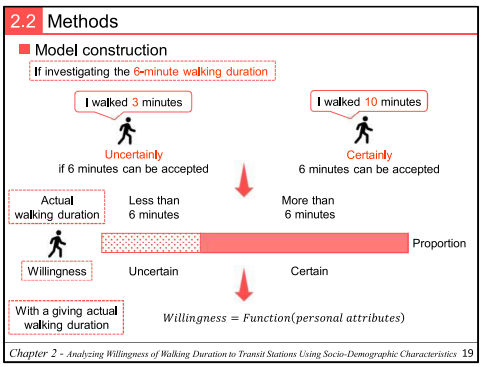


Here is a relation map of distance, attributes, willingness, and actual walking duration.

The actual walking duration is the original data that we can obtain from the questionnaires. As shown in this relation map, the actual walking duration should be directly determined by the distance between the starting point and the station.

We assume that people with different attributes should have different willingness of walking duration, and this willingness can affect the final actual walking duration as well. But, the attributes should never affect the actual walking duration directly.

So, what I want to do is to build a bridge between the attributes and actual walking duration through the willingness.

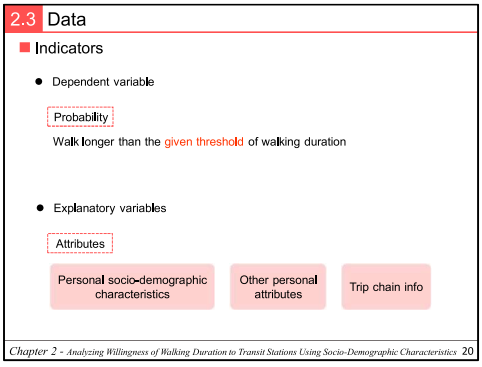


And this is the bridge. I'd like to raise another example to explain it.

We can take the 6-minute walking duration as the investigation object, if the respondent answered that "I walked 10 minutes to the station", it not only means this passenger walked 10 minutes, but also tells us the 6-minute walking duration is certain to be accepted by this passenger. While if someone said "I walked 3 minutes", it means the 6-minute walking duration is uncertain to be accepted.

Well, we can get a relationship between the actual walking duration and the willingness.
And obviously, the willingness of walking duration can be expressed as the function of personal attributes.

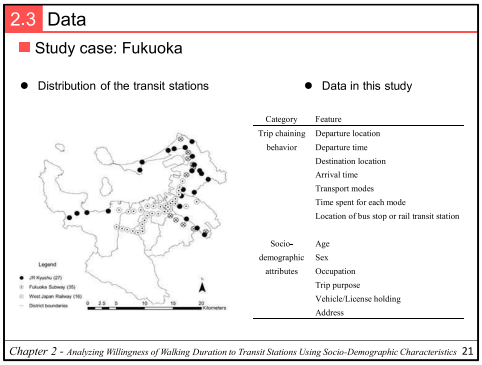
Here the bridge between personal attributes and the actual walking duration can be established.



Then I will give a brief introduction for the indicators used in this study.

The definition of the dependent variable is the key point to bridge the actual walking duration and the personal attributes. Just as I interpreted in the last slide, here the dependent variable is defined as the probability of walking longer than the given threshold of walking duration.

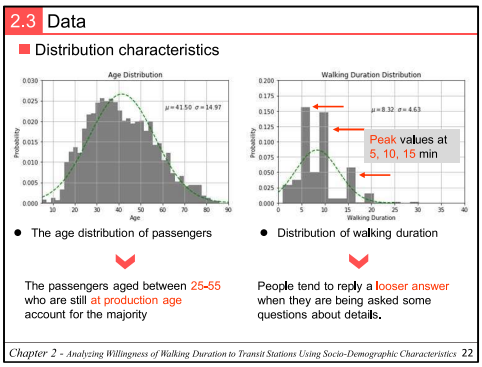
The explanatory variables are passengers' attributes including socio-demographic characteristics, some other personal attributes, and the trip chain information.



The study case and data source.

This study uses the rail transit stations of Fukuoka, including 78 stations in total.

The main data source is the person trip survey, it mainly includes the information about trip chaining behavior and some socio-demographic attributes.

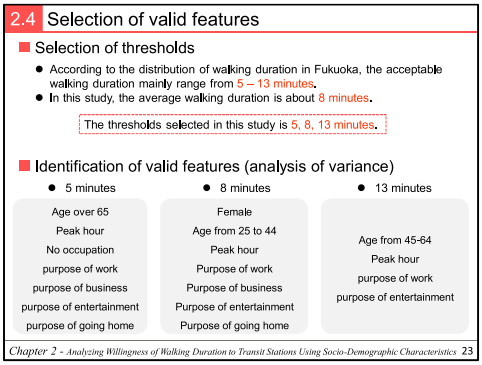


A brief description for the data. The left part is the age distribution of passengers.

As we can see in this figure, the passengers aged between 25-55 who are still at production age account for the majority. The right part is the distribution of walking duration, we can clearly see 3 peak values in this figure, they are 5, 10, 15 minutes.

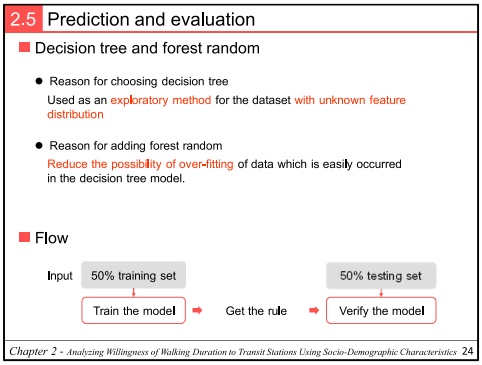
Here I want to have a try to give some explanations to it. It can be easily imagined that if people were asked some questions about details like how much did you spend, how long did you take, for such kinds of questions, people tend to reply a looser answer, some general numbers like 5, 10, or something. Maybe this is the reason that the peak values appeared at 5, 10, 15 minutes.

These peak values may be one of the reasons that the previous studies didn't get the expected result. But these peak values are considered not to lead deviation in this study. Because the walking duration that passengers answered in the questionnaires represents this walking duration has already been accepted by the passengers, even it is not accurate.



For the threshold of walking duration we are going to investigate, according to the distribution of walking duration in Fukuoka, the 5, 8, 13 minutes are selected as the threshold to be investigated.

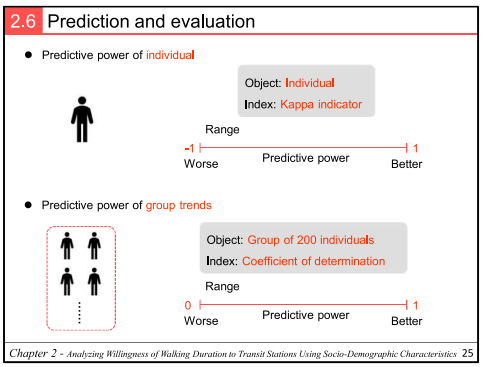
Then I use the analysis of variance to extract the valid features at each threshold of walking duration. We can see that the features vary from different walking duration.



Now we have got the valid features, the next step is to estimate the model to find the relationship between walking duration and the indicators.

With the consideration of the dataset and the research purpose, the forest random decision tree model is introduced to this study.

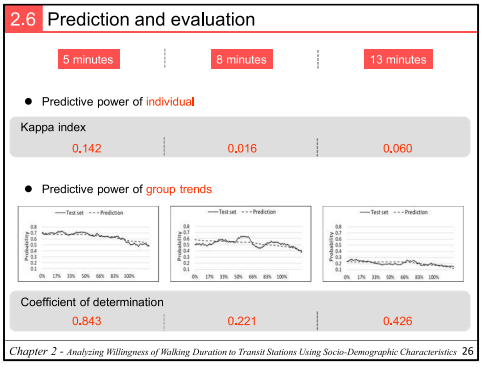
The procedure has two steps, training the model, and verifying the model. The dataset is split into the training set and testing set, for 50 percent respectively.



The prediction is evaluated from two ways, for the individual and group.

The predictive power of individual is evaluated by the Kappa index. This index ranges from minus one to one, the higher value means a better predictive power.

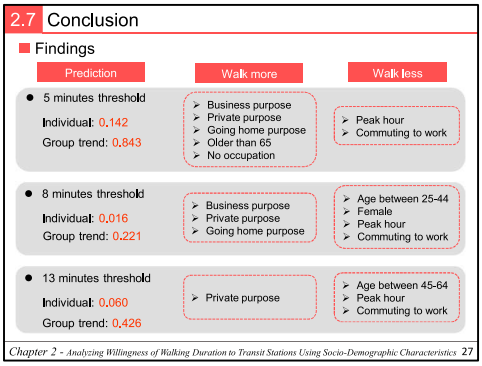
The predictive power of group is evaluated by the coefficient of determination, it ranges from 0 to 1. Higher value represent a better predictive power.



Let's take a look at the result.

We can see that the result of 5-minute walking duration is better than the other two walking duration.

The results at 8-minute are less accurate.

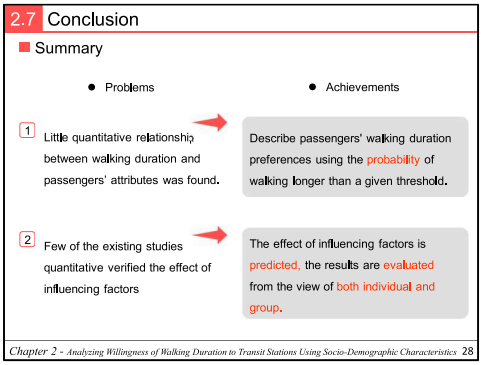


I will give a summary of the results.

In this table, the left column is the predictive power, the right part is the personal attributes identified at each threshold of walking duration. The column of walk more means passengers with these attributes tend to walk longer than the given walking duration, the attributes in the column of walk less means passengers tend to walk less than the given walking duration.

Passengers with the private travel purpose tend to accept the longest walking duration. Passengers with the commuting travel purpose and during the peak hour tend to arrive at transit station as soon as possible.

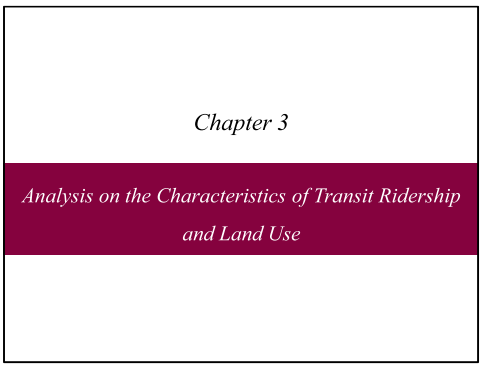
Something also can be inferred from this result. The 5 minutes walking duration can be widely accepted, some passengers who cannot accept this walking duration should have some significant characteristics, so this 5-minute walking duration has a higher predictive power. The 8 minute walking duration is a little ambiguous in passenger's behaviors, therefore it is a little difficult to predict.



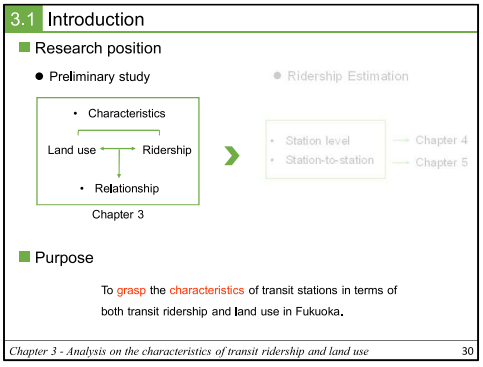
To the summary, the main achievements of this study:

1. This study described passengers' walking duration preferences using the probability of walking longer than a given threshold.

2. The effect of influencing factors is predicted, the results are evaluated from the view of both individual and group.

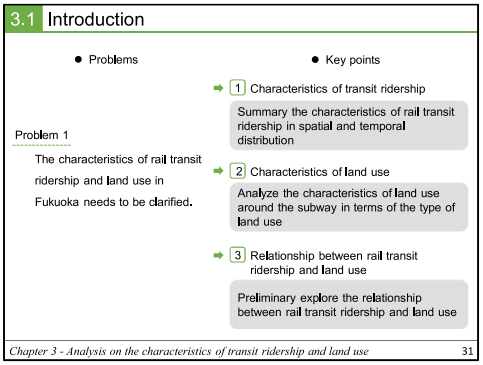


This is the chapter 2. Next, chapter 3, Analysis on the Characteristics of Transit Ridership and Land Use

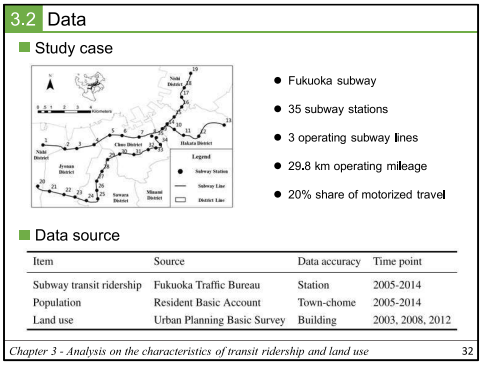


This chapter is a preliminary study for estimating the influence of transit ridership, it mainly focus on the characteristics of transit ridership and land use, also give a simple exploration on the relationship between them.

The purposes of chapter 3 is to grasp the characteristics in terms of both transit ridership and land use.



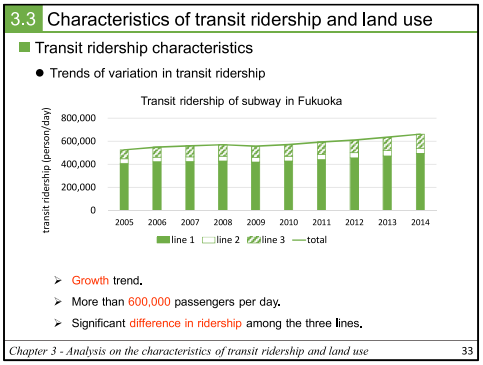
To make a detailed understanding of the rail transit system in Fukuoka, this chapter mainly focused on three key points. 1. to summarize the characteristics of rail transit ridership. 2. To analyze the characteristics of land use around the subway stations. 3. To make a preliminary exploration on the relationship between rail transit ridership and land use.



This is the overview of study case and data source.

The study case is Fukuoka subway station.

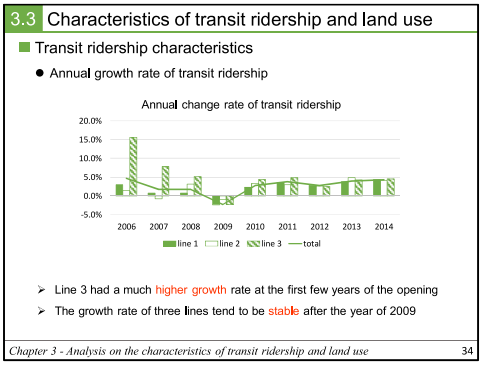
The main data source in this study is the subway transit ridership, the statistics of population and land use.



This is the trends of variation in transit ridership.

The ridership keeps a growth trend. Currently, the subway system carries 6 hundred thousands passengers per day.

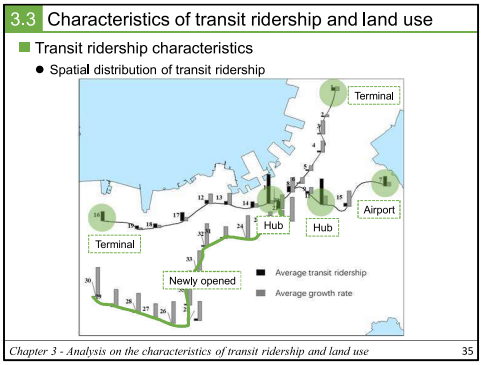
And the three lines showed significant difference in ridership.



The annual growth rate of transit ridership.

We can see from the figure that line 3 had a much higher growth rate at the first few years of the new opening.

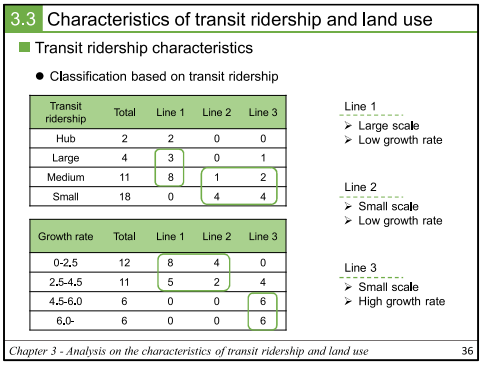
The growth rate of three lines tend to be stable after the year of 2009.



This is the spatial distribution of transit ridership.

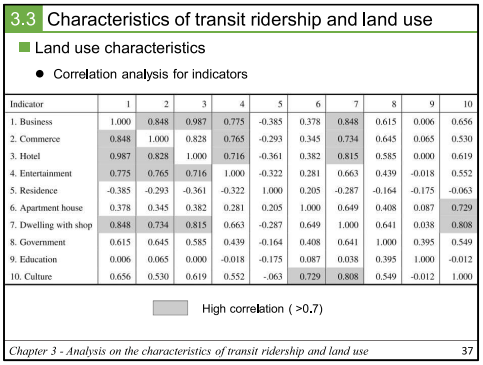
Fukuoka subway has two terminal stations, two hub stations, and one station serving for the airport.

We can see that, most stations located in the central area have larger transit ridership.



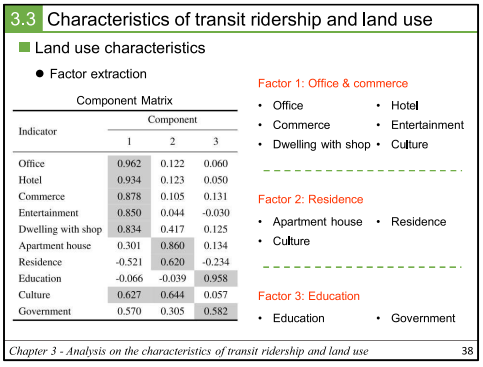
Two cross tables about the transit ridership and the growth rate are given to help to understand the characteristics.

Both line 1 and line 2 have low growth rates, line 1 has a larger transit ridership. For line 3, although the scale of transit ridership is still small, it keeps on a high growth rate.



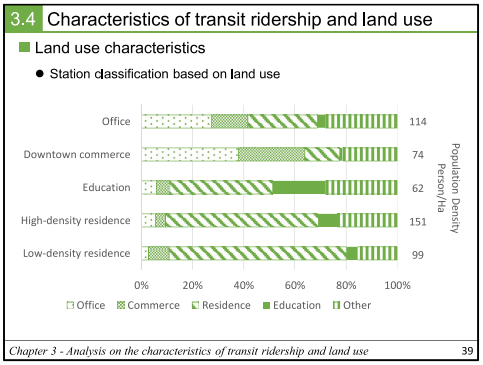
Because there are so many types of land use, we need to master the relationship among all the kinds of land use at first.

This table is the result of correlation analysis, the cells marked with grey color means there is a strong linear correlation between the two indicators.



To deal with this linear correlation, the factor analysis is conducted to make a in-depth exploration on the internal relationship.

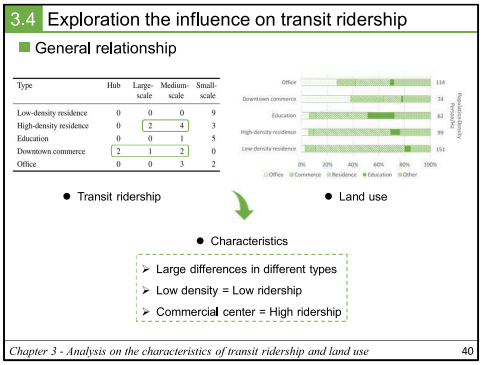
As a result, three factors are extracted, they are office & commerce, Residence, and education.



According to the analysis of land use indicators, all the 35 stations are classified into several types.

They are office type, downtown commerce type, education type, high-density residence type, and low-density residence type.

This is the summary for each type of stations.

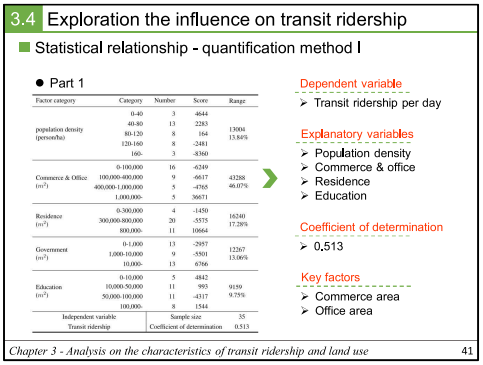


Combining with the cross table of land use and transit ridership, we can get some information.

That is, different land use characteristics have different scales in transit ridership.

The density of both population and land use have significant influence on transit ridership.

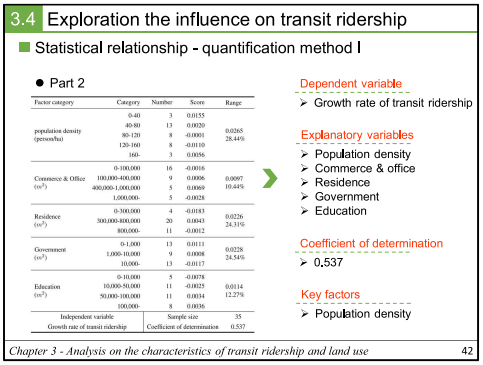
Commercial land use produces more ridership than other land use types



After a general understanding of the relationship between ridership and land use, the quantification method one is used to make a deeper exploration.

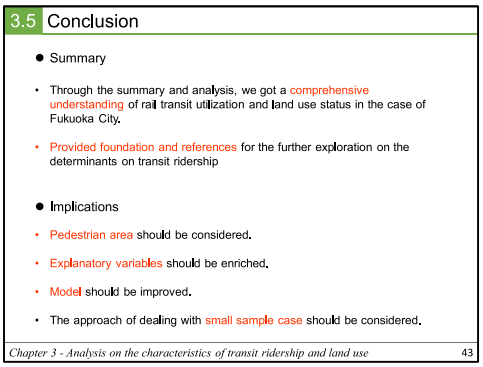
In the first part of quantification method one, the dependent variable is transit ridership per day.

But the estimation did not show a good fitness



In part 2, the same indicators are used for estimating the growth rate of transit ridership.

The same with part 1, the result did not show a good fitness as well.

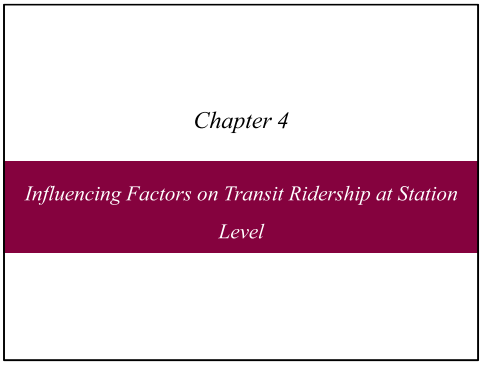


This is the contents of chapter 3.

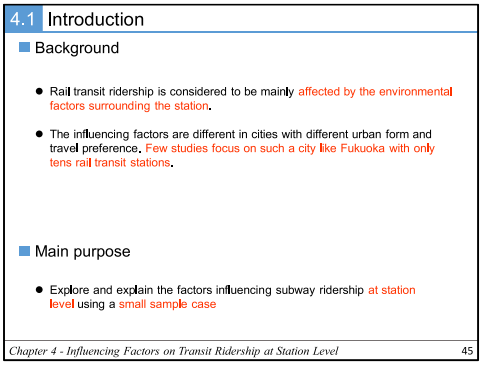
Through this study, we got a comprehensive understanding of transit ridership and land use status in the case of Fukuoka City.

This study also provided the foundation and references for the next exploration of the determinants on transit ridership

Some implications can be got from this chapter.
1. Pedestrian area should be considered in depicting the catchment area.
2. Explanatory variables should be enriched.
3. The Model should be improved.
4. The approach of dealing with small sample case should be considered.



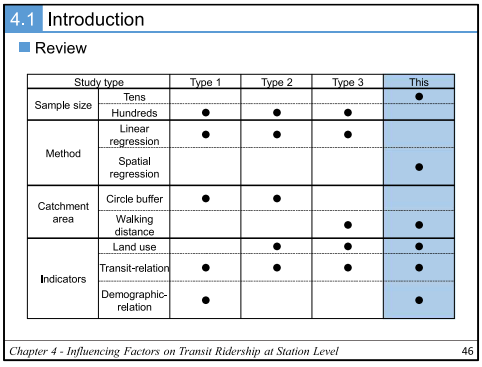
On the base of chapter 3, chapter 4 aims to make a deeper exploration on factors influencing subway ridership at station level.



As we know, rail transit ridership is considered to be mainly affected by the environmental factors surrounding the station.

The influencing factors are different in cities with different urban form and travel preference. There is a problem in the existing studies that few studies focus on such a city like Fukuoka with only tens stations.

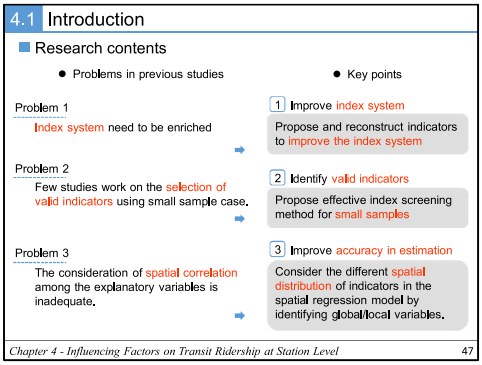
So, the purpose of this study is to explore the factors influencing subway ridership at station level using a small sample case



Study type		Type 1	Type 2	Type 3	This
Sample size	Tens				●
	Hundreds	●	●	●	
Method	Linear regression	●	●	●	
	Spatial regression				●
Catchment area	Circle buffer	●	●		
	Walking distance			●	●
Indicators	Land use		●	●	●
	Transit-relation	●	●	●	●
	Demographic-relation	●			●

At first, let's take a look at how this study differs from the others. The biggest difference in this study is the sample size. Few studies deal with a case just having tens of the samples. Because this kind of studies is commonly based on statistics, a sufficient sample size is necessary.

Moreover, with the consideration of the spatial distribution of stations, the spatial regression will be introduced to this study. Also, to define the catchment area more accurately, the walking distance along the real road network is used to determine the catchment area.

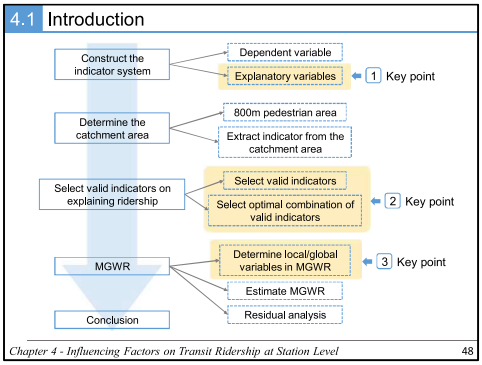


According to the review, this study will focus on three points to deal with the problems existing in the previous studies.

1. Reconstruct indicators to improve the index system.

2. Select the effective indicators for explaining transit ridership in a small sample case.

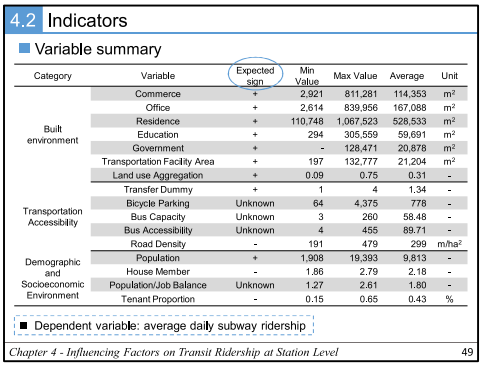
3. Identify the global/local indicators in the estimation of the spatial regression.



This is the organization of this study.

It is constructed by 5 parts, constructing the indicator system, determining the catchment area, selecting the valid indicators, estimating the model, and drawing the conclusion.

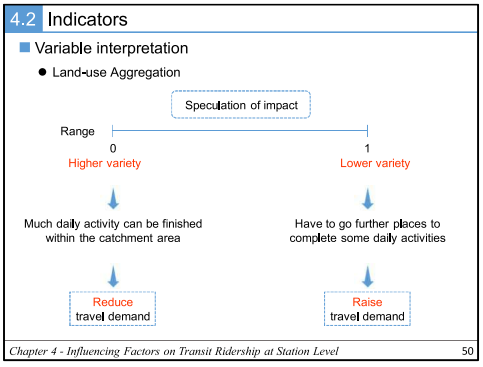
The parts marked with yellow blocks are the key points in this study.



The introduction of the indicator system.

In this study, the dependent variable is the average daily transit ridership.

The candidate independent variables are summarized into 3 categories: 1. built environment; 2. transportation accessibility; 3. demographic and socioeconomic environment.

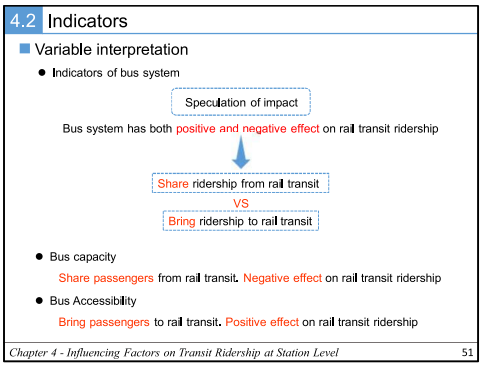


Then I will introduce some of the indicators in details.

First one is the indicator of land-use aggregation.

This indicator is used to describe the variety of land use.

The land use aggregation ranges from 0 to 1. A higher variety of land use infers much daily activity can be finished within the catchment area, it is supposed to lead a reduction in transit ridership. On the contrary, a lower variety of land use is thought to raise the transit ridership



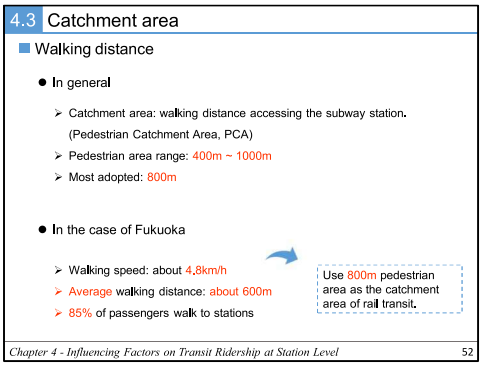
Another one is about the bus system.

In this study, the indicator of bus system is considered from two aspects.

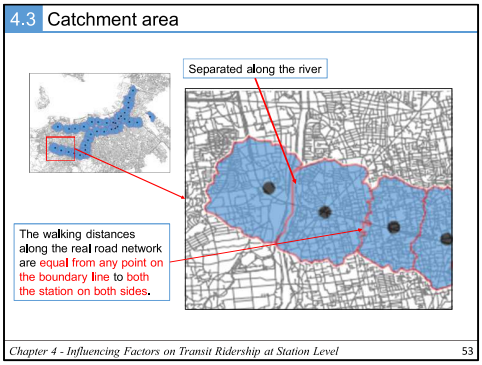
As shown in the figure, bus is supposed to have double effects on the rail transit, one is sharing passengers from rail transit, another one is bringing passengers to rail transit.

The indicator of bus capacity represent the ability of sharing passengers from rail transit, which is supposed to have negative effect on rail transit ridership.

The indicator of bus accessibility represent the ability of bringing passengers to rail transit, which is supposed to have positive effect on the rail transit ridership.



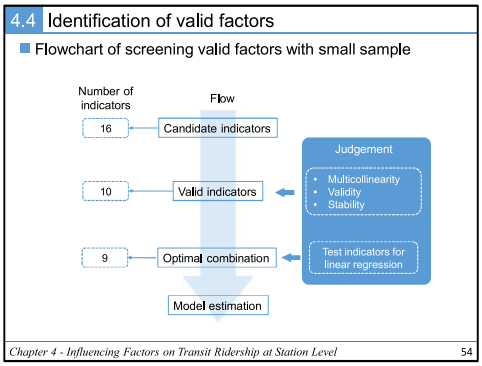
As to the catchment area, according to the previous studies, and the subway survey report of Fukuoka, this study uses the 800 meters real walking distance.



The catchment area is depicted as shown in this figure.

Any point on the boundary between two stations has the equivalent distance to both the stations.

Here, there is a river between the two stations that the river separated the road network. As a result, the two catchment areas are split along this river.

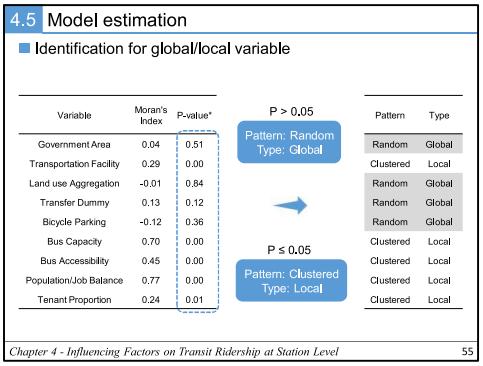


This part is identifying the valid factors in the small sample case.

This is the flowchart of screening valid factors. The identification is conducted by two steps.

The first step is to identify the valid variables, using the judgement of multicollinearity, validity, and stability.

In the second step, the linear regression is conducted to select an optimal combination of valid explanatory variables.

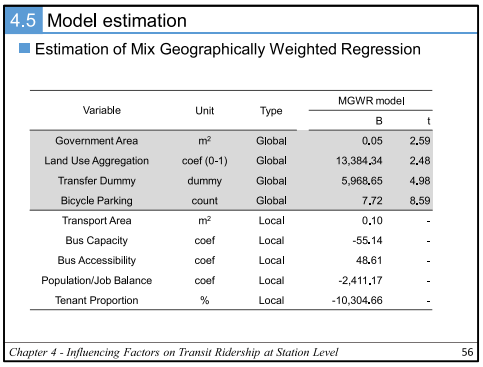


Now we have got the valid explanatory variables.

Before estimating the model, we should determine whether the indicator is a local or global indicator.

This procedure is conducted by using Moran's index to examine if the variable is spatial autocorrelation or not.

As the result, the indicators of clustered pattern are selected as the local variable, while the random pattern are judged as the global variable.



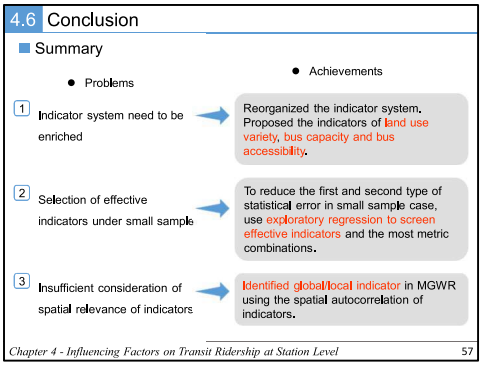
This is the result of estimation.

As a result, the indicators of building area have a positive effect on the transit ridership.

The indicator of land-use aggregation shows land use variety can lead to decrease in transit ridership.

As the expectation, the indicator of bus capacity has a negative effect on the subway ridership, and the bus accessibility has a positive effect.

The factor of resident/job balance and tenant proportion tells that working people and tenants are more willing to use subway.

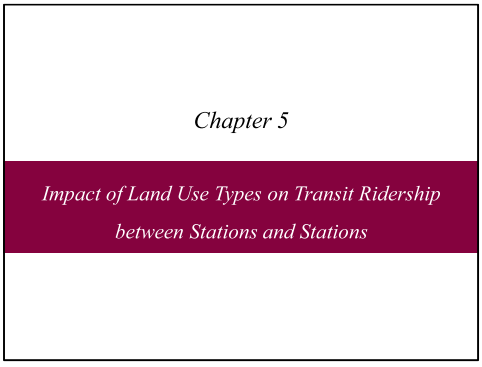


The main achievements of this study can be summarized as following

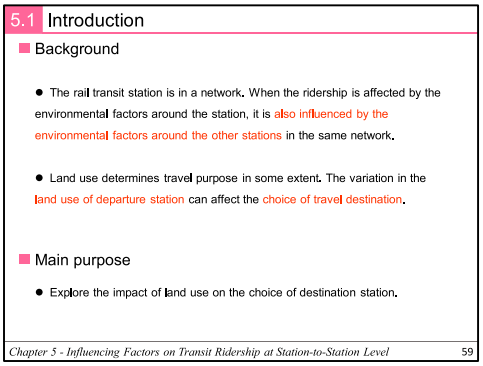
1. This study reorganized the indicator system.

2. To reduce the first and second type of statistical error in small samples case, this study used exploratory regression to screen effective indicators.

3. Identified global/local indicator in Mix Geographically Weighted Regression.



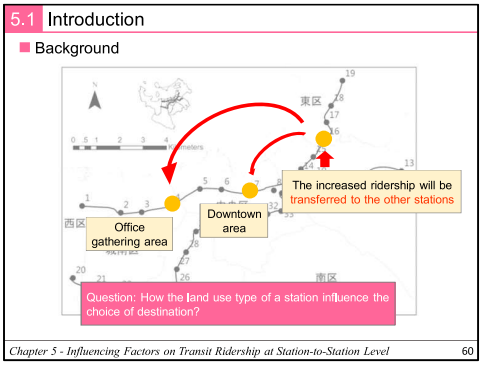
Chapter 5, Impact of Land Use Types on Transit Ridership between Stations and Stations



We know that the rail transit stations are in a network. The transit ridership is not only affected by the elements around the station, but also be affected by the other stations in the same network.

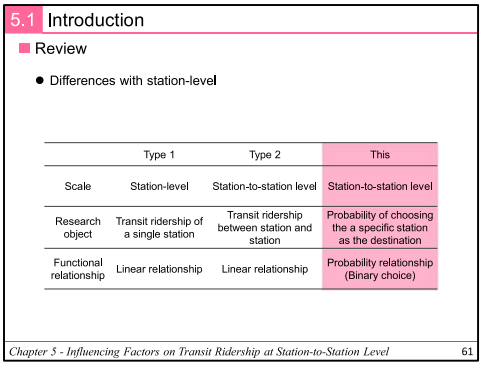
On the other hand, because land use is one of the most important determinants on travel purpose. The variation in the land use of departure station is considered to have influence on the choice of travel destination.

Therefore, the main purpose of this study I to **explore the impact of land use on the choice of destination station.**



I'd like to give a brief introduction for what I want to do in this study.

As shown in this slide, if the transit ridership increased in this station, the increased part will be transferred to the other stations in the network, Due to different travel purpose, this part of increased passengers will choose different stations as the destinations.

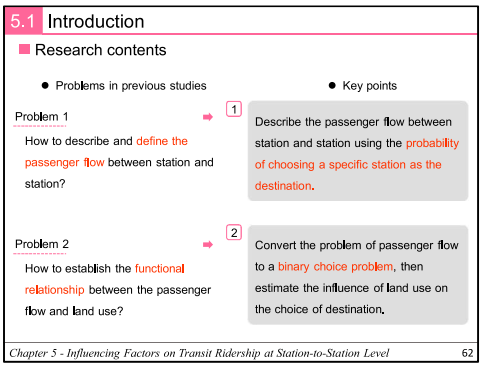


Then I will give a brief review to make clear the differences from previous studies.

For the research object, differ from station-level, the research object is not the passenger volume of a single station but the passenger flow between station and station.

Although some studies have paid attention to the station-to-station level, they just simply changed the research object to the passenger flow between station and station, but the functional relationship they examined is still the linear relationship.

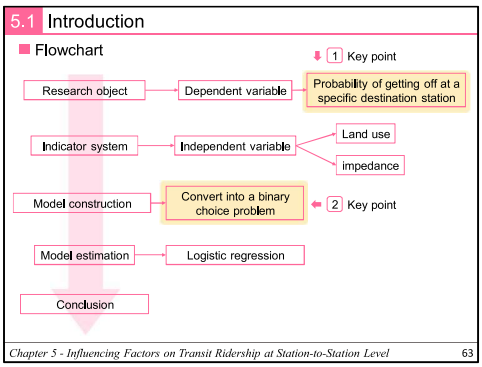
This study uses the probability of choosing a specific station as the destination to describe the passenger flow between station and station.



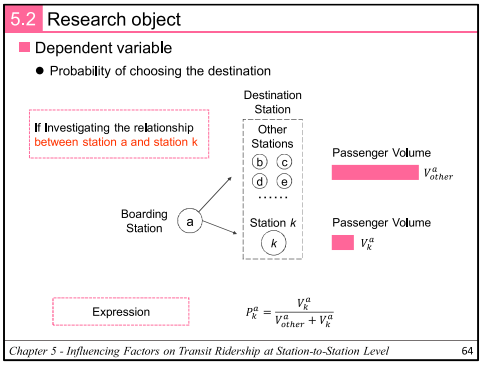
According to the review, this study will focus on 2 points to deal with the problems existing in the previous studies.

For the problem of how to describe and define the passenger flow between station and station? This studies describes it using the probability of choosing a specific station as the destination

For the second problem of how to establish the functional relationship between the passenger flow and land use? This study converts this issue to a binary choice problem, then use logistic regression to estimate the influence of land use on the choice of destination.

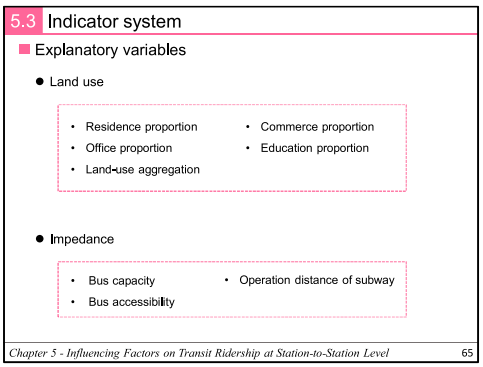


This is the general flow of this study. It contains 5 main parts. Defining the research object. Establishing the indicator system. Building the model, then estimating the model. Finally drawing the conclusion. The key points of this study is marked with yellow blocks.

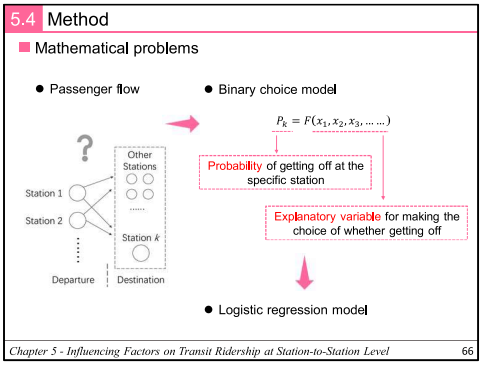


This is the definition of the dependent variable. It is defined as the probability of choosing a specific destination station. I will use a simple example to explain it.

If investigating the relationship between station a and station k, the passenger flow between station a and station k can be described as the following formula. The probability of choosing station k as the destination can be expressed as pak. This study supposes this probability should relate to the land use of both departure station a and destination station k.

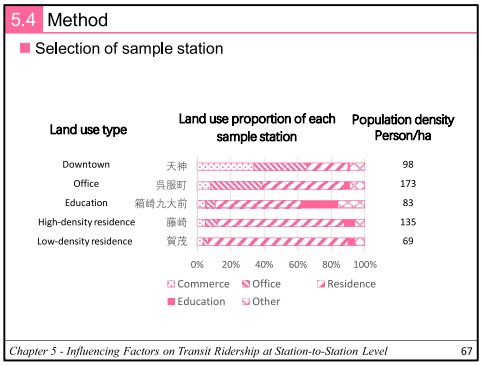


After defining the dependent variable, let's take a look at the explanatory variables. The indicator system is constructed by two categories, the land-use, and impedance.



Now we have prepared the dependent and explanatory variables, we need a model to establish the functional relationship between them.

Because the dependent variable is the probability of getting off at a specific station, this issue can be viewed as a binary choice problem. This is the abstract function, it can be estimated using logistic regression model.

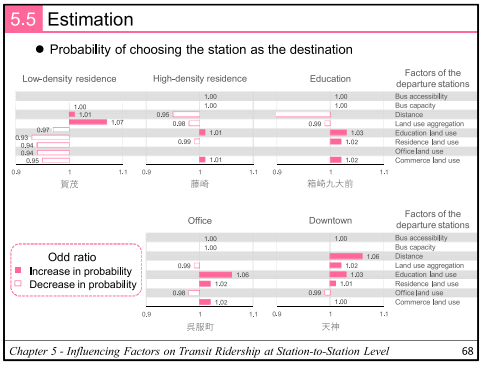


Because what we want to know is how land use affect the choice of destination, here I select some typical stations as the destination to estimate this influence.

According to the classification in the chapter 3, we select 5 stations from 5 types respectively.

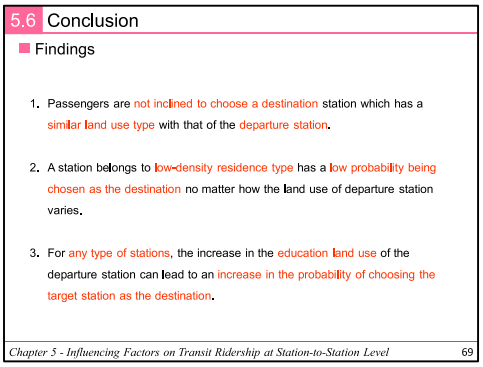
They are Tenjin station of downtown type, Gofukumachi station of office type, Hakozakikyudaimae station of education type, Fujisaki station of high-density residence type, Kamo station of medium-density residence type.

We can see the five stations have significant features in land use composition.



Now let's check how the land use of departure station can influence the probability of choosing these 5 stations as the destination.

The solid square means the corresponding indicator of the departure stations can lead an increase in the probability of choosing this target station as the destination, while the hollow square represents a decrease in this probability.

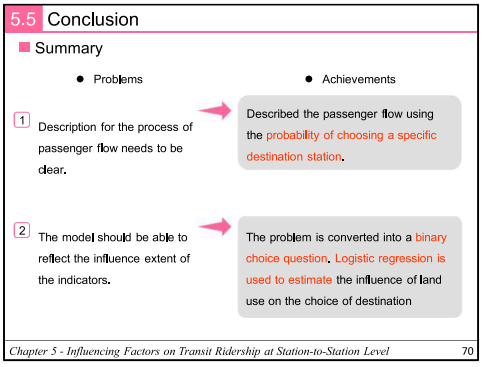


According to the result of estimation, we got some findings. They can be summarized as following:

1. Passengers are not inclined to choose a destination which has a similar land use type with that of the departure station.

2. A station belongs to low-density residence type has a low probability being chosen as the destination no matter how the land use of departure station varies.

3. For any type of stations, the increase in the education land use of the departure station can lead to an increase in the probability of choosing the target station as the destination.



The summary for the main achievements.

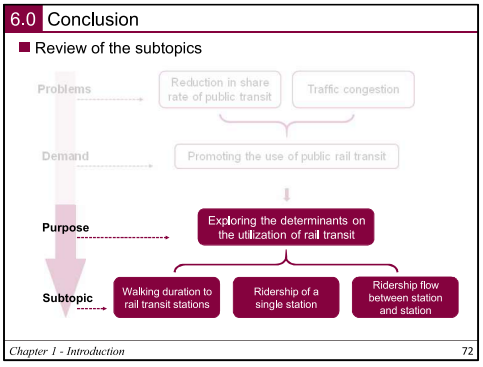
First one, this study described the passenger flow using the probability of choosing a specific destination station.

Secondly, converted this issue into a binary choice question, then estimated the influence of land use on the choice of destination using logistic regression model.

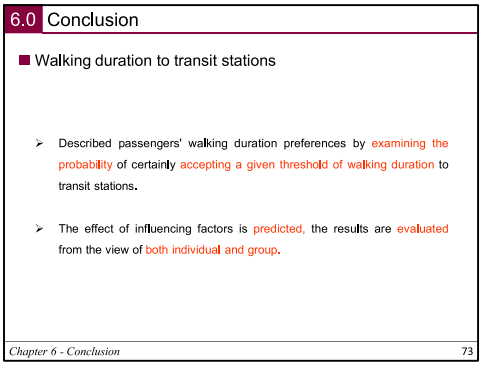
Chapter 6

Conclusion

This is the contents of chapter 5. Finally, chapter 6, the conclusion.



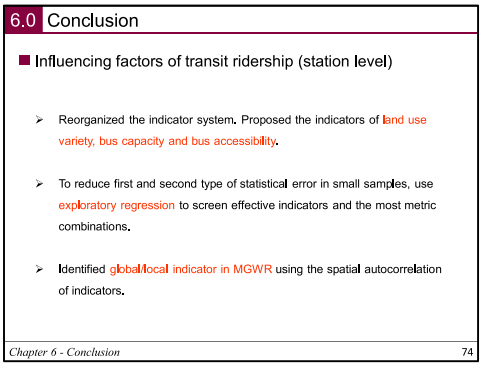
Centering the main purpose of exploring the determinants on the utilization of rail transit, this research focused on 3 subtopics, walking duration to rail transit stations, transit ridership of single station, passenger flow between station and station.



In the topic of walking duration to transit stations,

This research described passengers' walking duration preferences by examining the probability of certainly accepting a given threshold of walking duration to transit stations;

Then the preference of walking duration was predicted, and the results are evaluated in both individual and group.

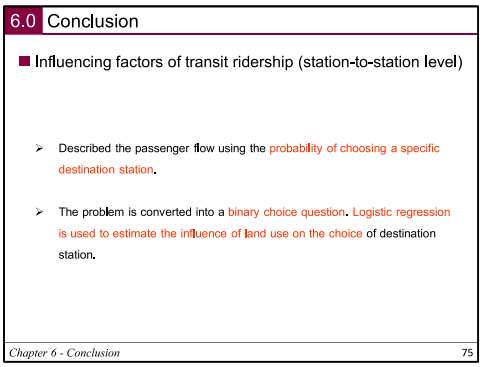


In the topic of Influencing factors of transit ridership at station level,

Firstly, this study redefined the indicators of land use variety and the indicators of bus system.

Secondly, this study used exploratory regression to screen effective indicators and the most metric combinations.

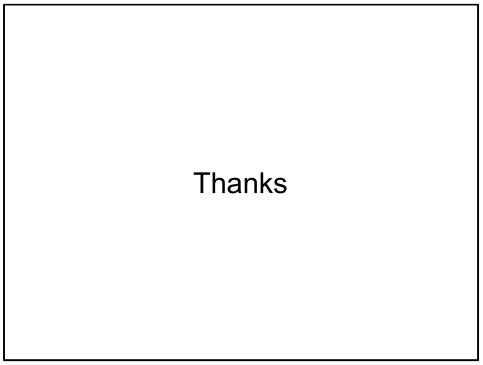
Thirdly, proposed an approach to identify the global/local indicator in MGWR.



In the topic of Influencing factors of transit ridership at station-to-station level,

this research described the passenger flow using the probability of choosing a specific destination station.

Then constructed a binary choice model to describe the passenger flow between station and station, and estimated the influence of land use on the choice of destination station.



That's all for my presentation, thanks for paying attention.