

# Research on Determinants of Rail Transit Ridership

*CHEN Qi*

Graduate School of Human-Environment Studies,  
Kyushu University, Japan

November 27, 2018



九州大学  
KYUSHU UNIVERSITY

# *Chapter 1*

---

## *Introduction*

---

## 1.1 Background

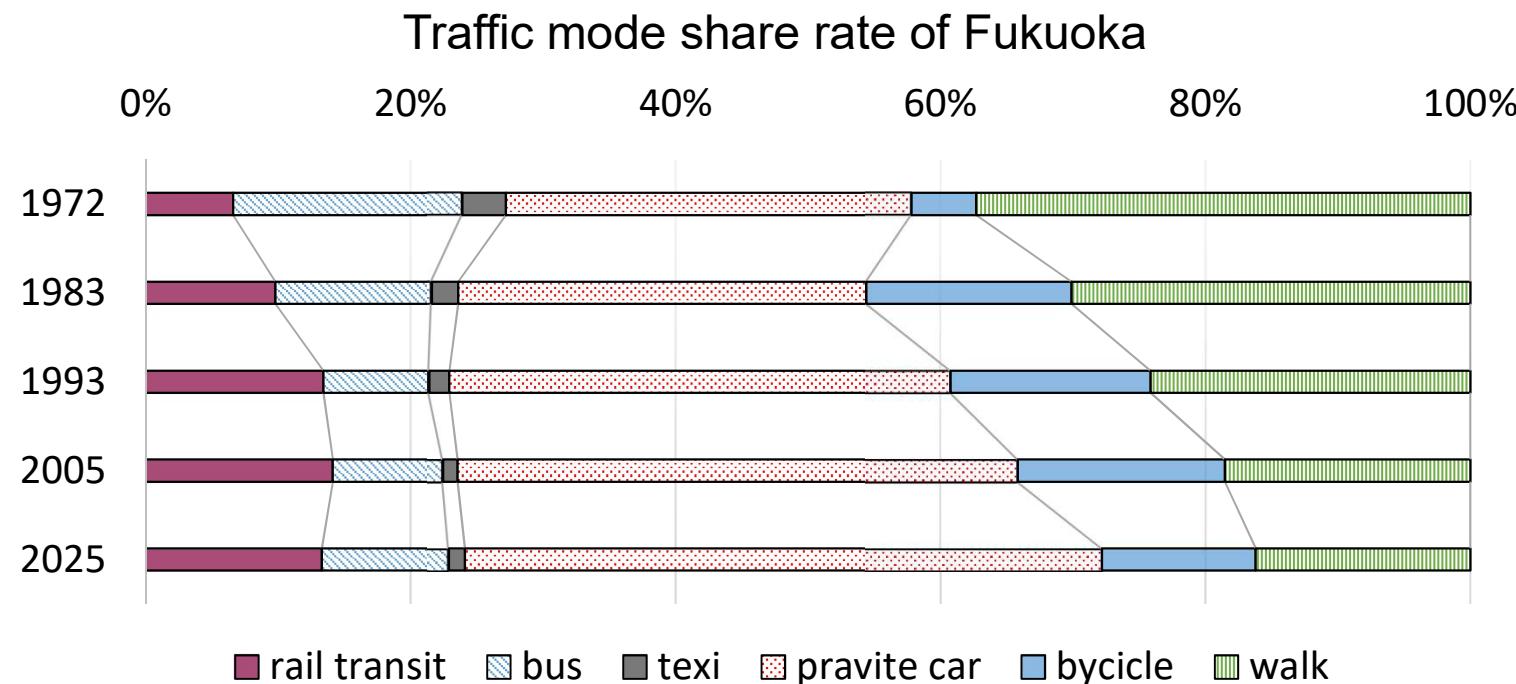
### ■ Current situation

1

Reduction in share  
rate of public transit

2

Traffic congestion



## 1.1 Background

### ■ Current situation

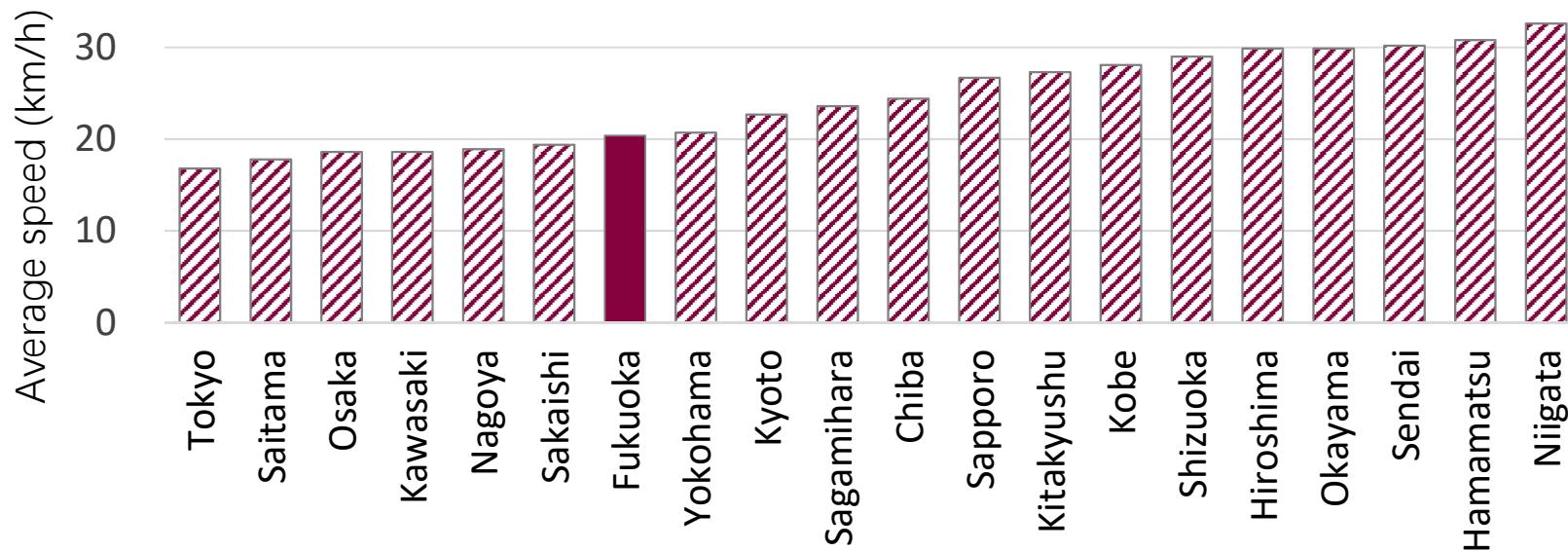
1

Reduction in share  
rate of public transit

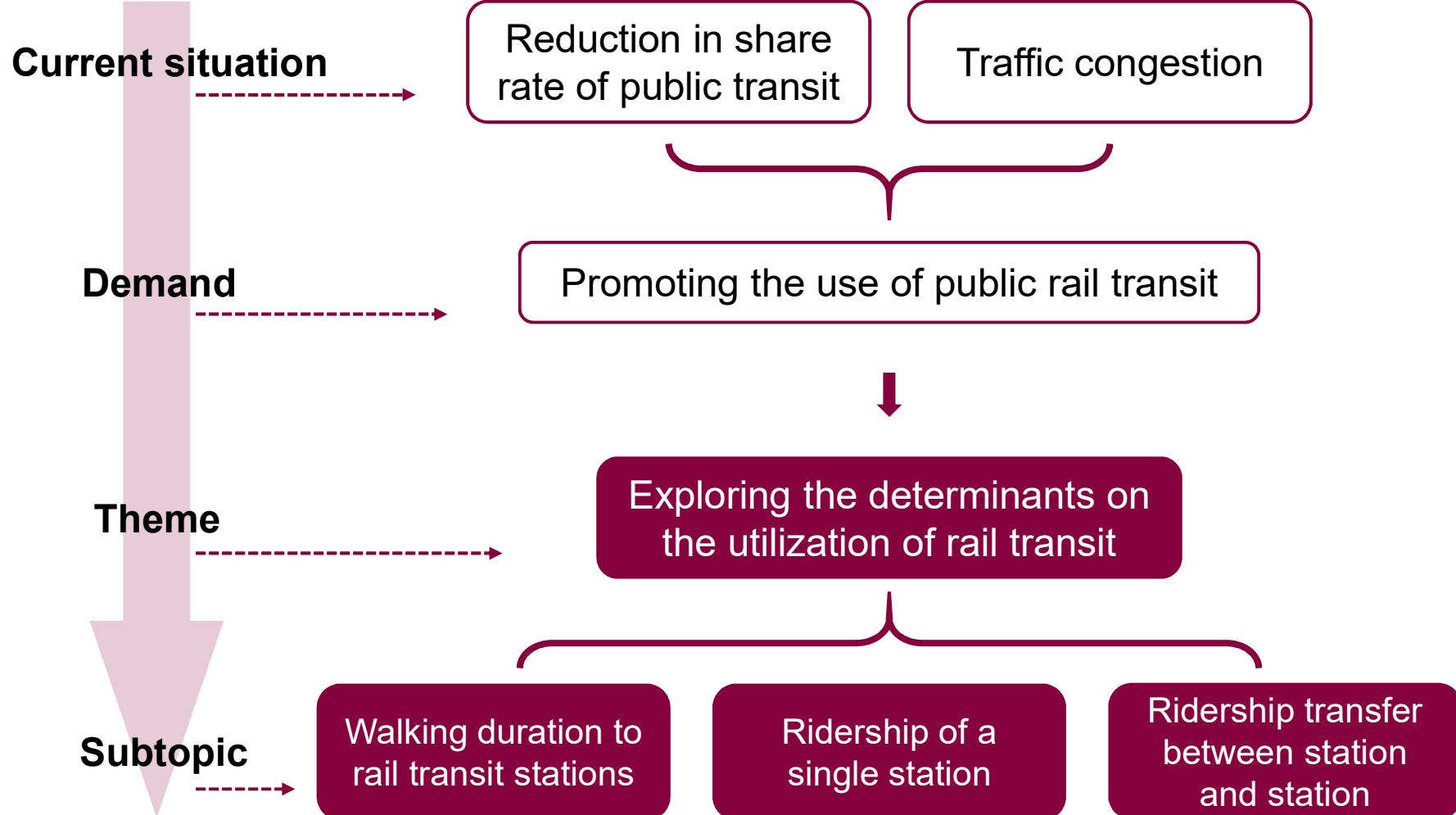
2

Traffic congestion

Average travel speed during crowded time in major cities



## 1.1 Background



## 1.1 Background

---

### ■ Problems

- 1 Walking is the most important travel modes in accessing transit stations, also one of the most key determinants on the utilization of rail transit. A quantitative method for describing walking duration preference to rail transit stations is necessary.
- 2 The rail transit ridership, land use, and population have their own distribution characteristics in space. These characteristics in the case of Fukuoka City need to be further and objectively understood.
- 3 Rail transit ridership is considered to be mainly affected by the environmental factors surrounding the station. For the rail transit station in Fukuoka City, the influencing factors need to be explored and estimated.
- 3 The rail transit station is in a network. When the ridership is affected by the environmental factors around the station, it is also influenced by the environmental factors around the other stations in the same network. How to describe and estimate the ridership transfer between station and station is an important issue.

## 1.2 Purpose

---

- Main purpose

Exploring the factors influencing rail transit ridership from multiple perspectives and estimating the influence.

- Secondary purposes

Exploring the factors influencing  
**Walking duration**

**Clarifying the characteristics of**  
ridership and land use

---

**Problem 2**

Chapter 3

**Problem 1**

Chapter 2

**Utilization of**  
rail transit

Chapter 5

**Problem 4**

---

Chapter 4

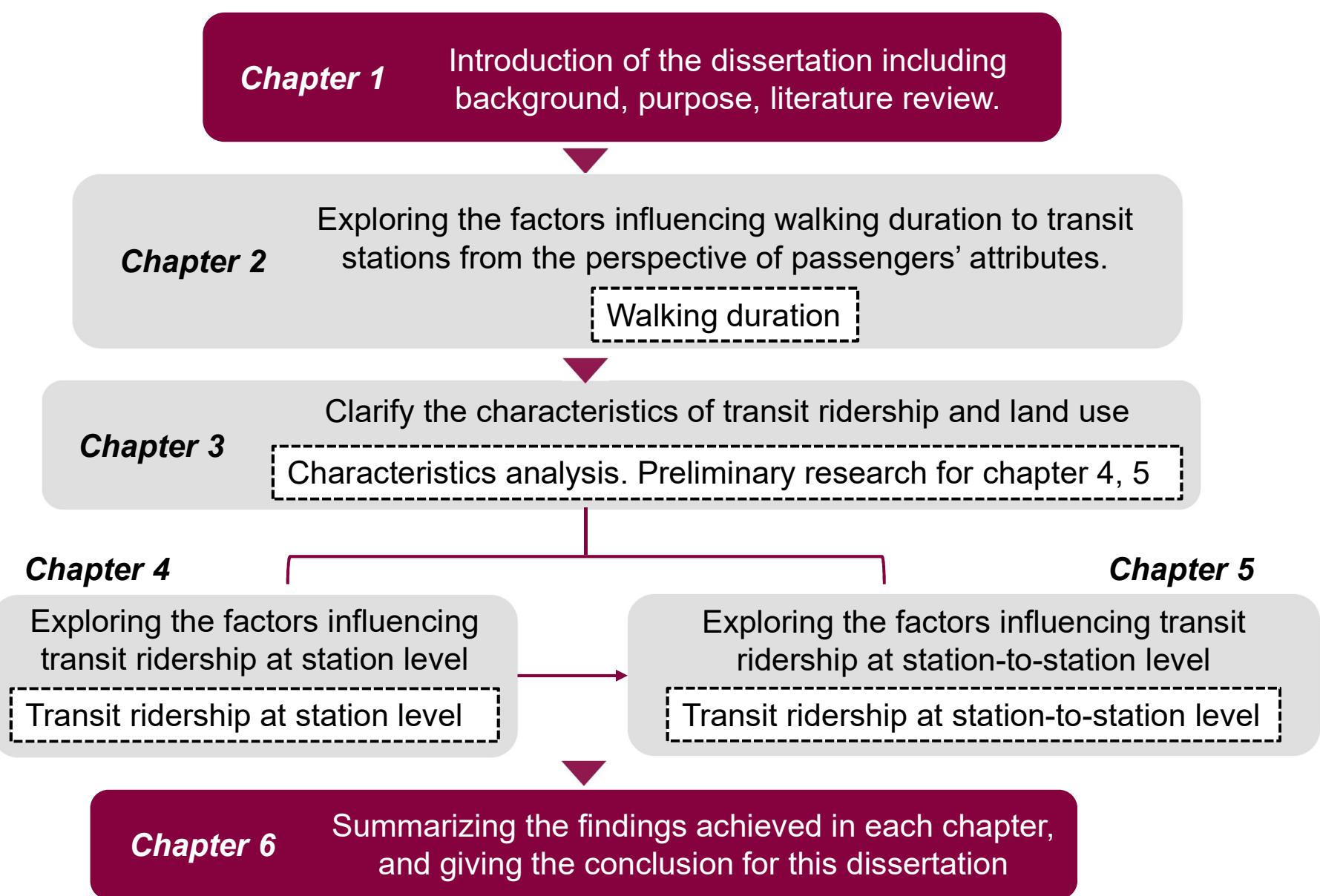
**Problem 3**

---

Exploring the factors influencing the  
**ridership at station level**

Exploring the factors influencing the  
**ridership transfer between station**  
and station

## 1.2 Organization



## *Chapter 2*

---

*Analyzing Willingness of Walking Duration to Transit  
Stations Using Socio-Demographic Characteristics*

---

## 2.1 Introduction

---

### ■ Background

- Walking is the most important travel modes in accessing transit stations.
- Walking duration to transit station is widely accepted to be an important determinants on the utilization of rail transit.
- Walking duration to rail transit stations is mainly obtained by survey statistics, which is limited by the individual and regional characteristics of survey objects that the preference of walking duration cannot be generalized to other cases.

### ■ Main purpose

- Exploring the factors influencing walking duration to transit stations from the perspective of passengers' attributes.

## 2.1 Introduction

### ■ Review

	Study	Type 1	Type 2	Type 3	This
Research object	Walking distance	●		●	
	Walking duration		●		●
Method	Statistical description	●			
	Regression		●	●	
	Nonlinear fitting				●
Method	Qualitative	●			
	Quantitative		●	●	●
Result		Survey result description	Some factors were found, but prediction and verification were not given	No relationship found	Having relationship with travel purpose, travel time period, age, and occupation

## 2.1 Introduction

### ■ Research focus

- Problems in previous studies

#### Problem 1

Most studies use qualitative methods. Few quantitative analysis obtained the significant result on estimating walking duration. The reason of that may be in the data processing, the model constructing, also model estimating.

- Research focus

#### 1 Describe walking duration

Using probability but not the linear relationship to describe the walking duration preference of passengers with different attributes.

#### 2 Identify valid indicators

Using variance analysis to identify the valid influencing factors.

## 2.1 Introduction

### ■ Research focus

- Problems in previous studies

#### Problem 2

Although many studies have summarized some factors affecting the walking duration through qualitative methods, **none of them quantitatively verified** the effect of the influencing factors.

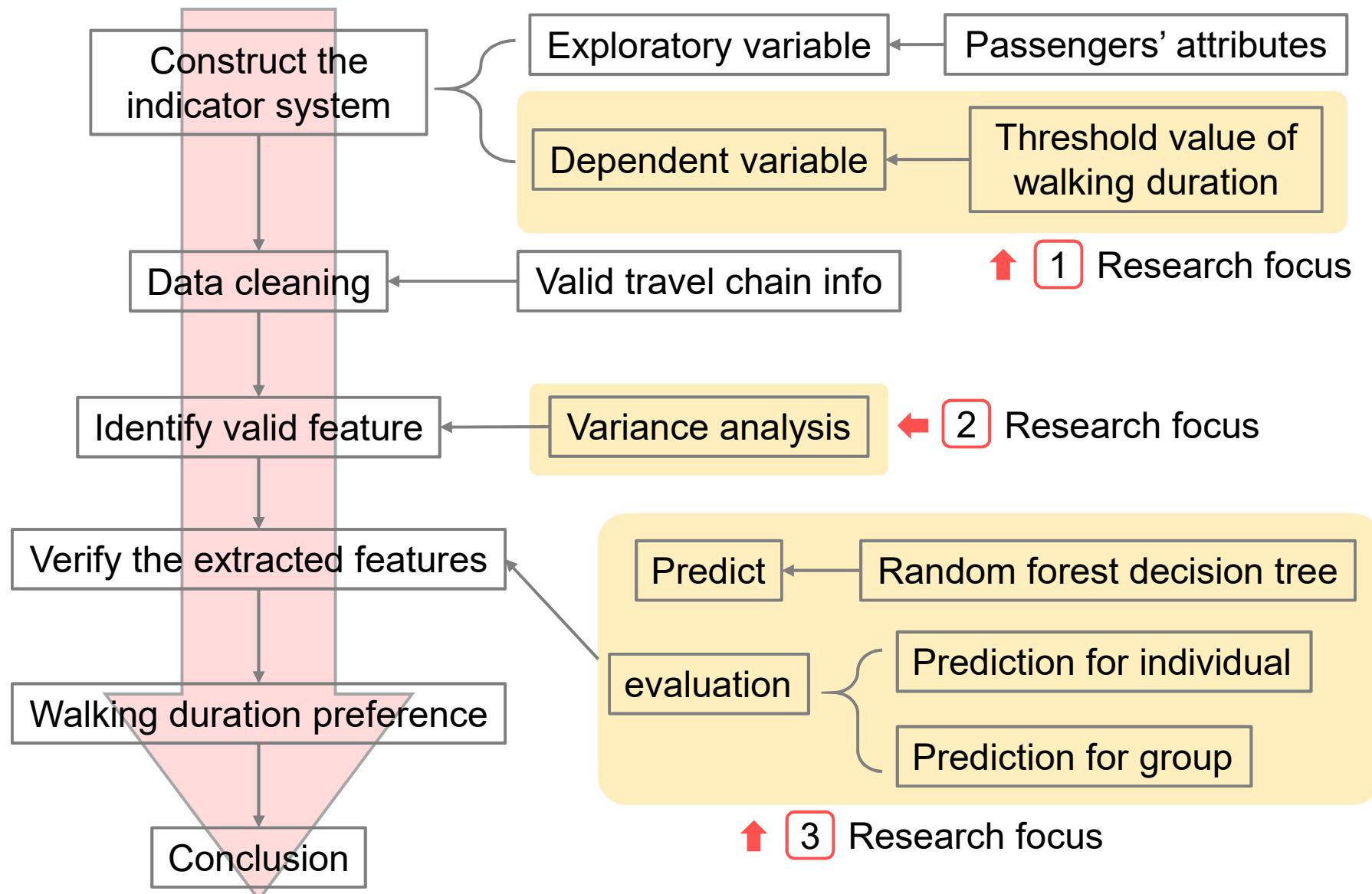


- Research focus

#### 3 Predict and evaluate

To estimate and predict the relationship between identified influencing factors and walking duration. Then evaluating the prediction to judge the effectiveness of identified influencing factors.

## 2.1 Introduction



## 2.2 Data

---

### ■ Indicators

- Dependent variable

Probability

---

Walk longer than the **given threshold** of walking duration

- Explanatory variables

Behavior

---

Personal socio-demographic characteristics

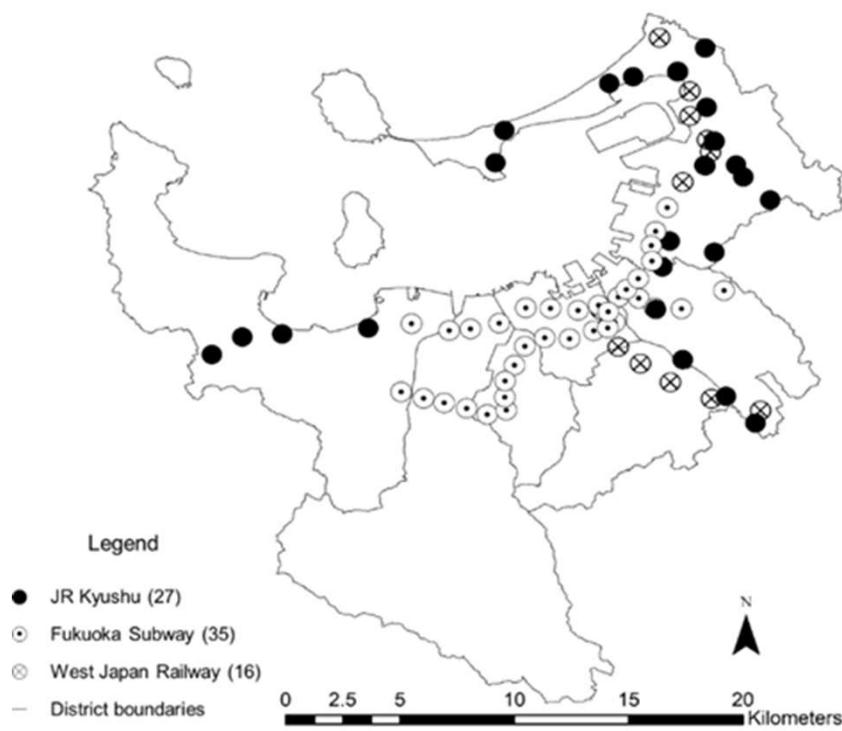
Other personal attributes

Trip chain info

## 2.2 Data

### ■ Study case: Fukuoka

- Distribution of the transit stations



- Data in this study

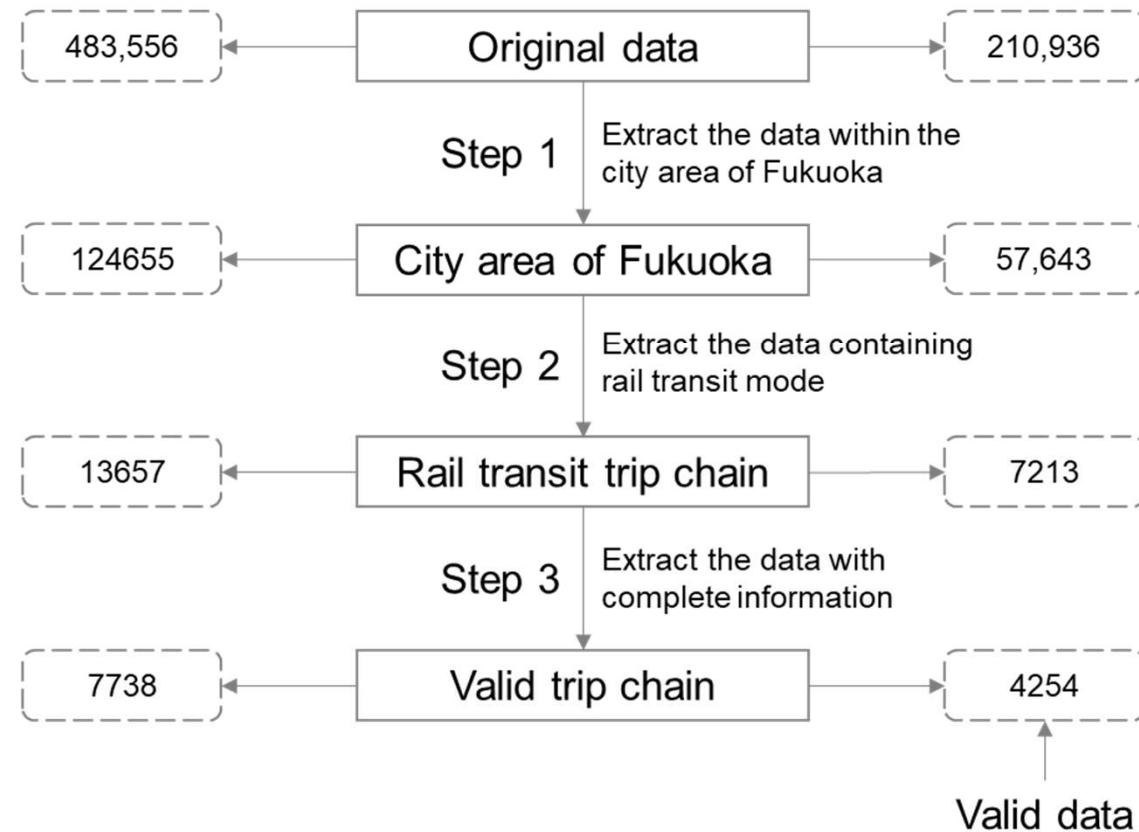
Category	Feature
Trip chaining behavior	Departure location
	Departure time
	Destination location
	Arrival time
	Transport modes
	Time spent for each mode
	Location of bus stop or rail transit station
Socio-demographic attributes	Age
	Sex
	Occupation
	Trip purpose
	Vehicle/License holding
	Address

## 2.2 Data

### ■ Data processing

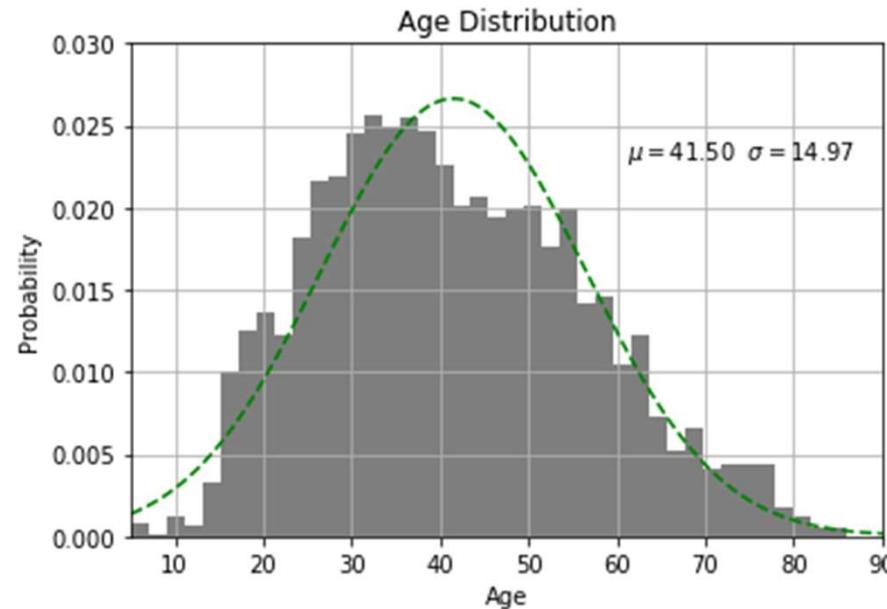
- Source: Person Trip Survey

The amount of records  
for trip chain                      The amount of records  
for individual



## 2.2 Data

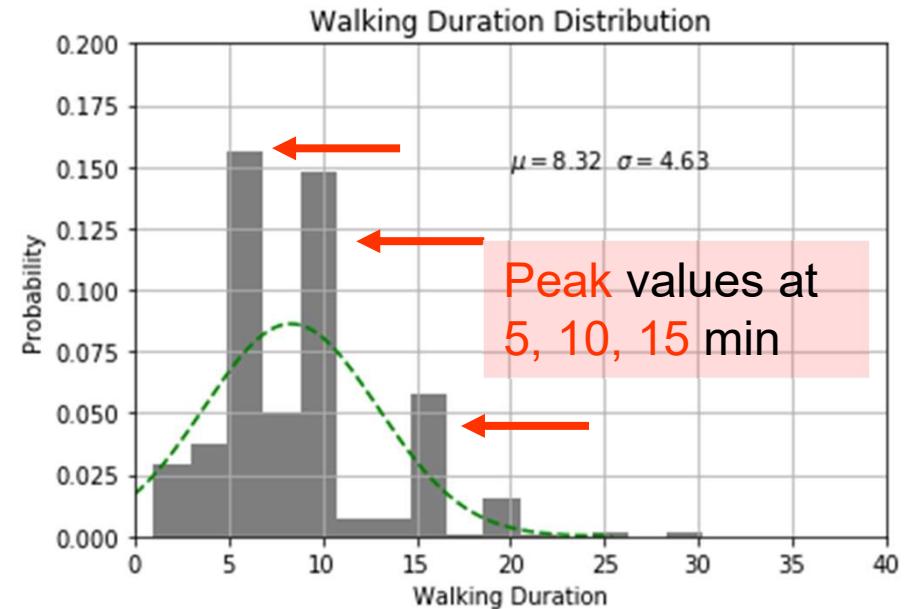
### Distribution verification



- The age distribution of passengers



The passengers aged between **25-55** who are still **at production age** account for the majority



- Distribution of walking duration



People tend to reply a **looser answer** when they are being asked some questions about details.

## 2.3 Methods

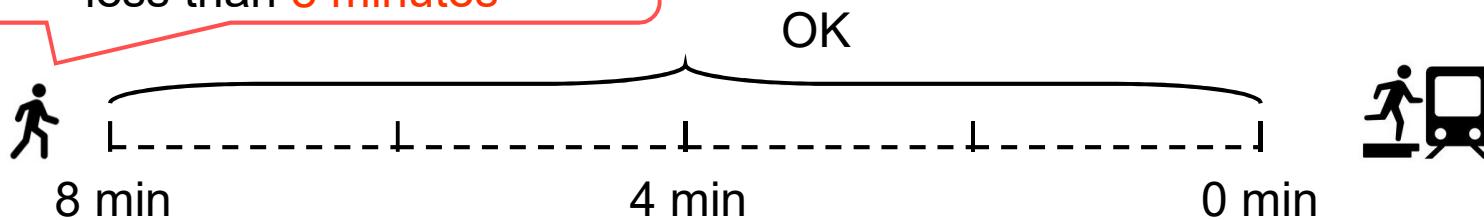
### ■ Problem description

Give an example, for one person who has the attributes:

*(Male, 25 years old, student, peak hour, go to school)*

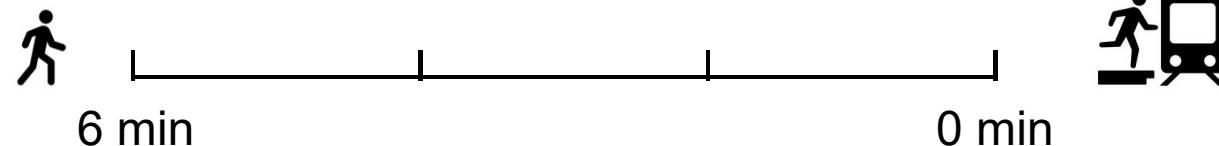
- Realistic issue

I can **accept** a walking duration less than 8 minutes



- Reflected in person trip survey

I **walked** 6 minutes to the station



## 2.3 Methods

### ■ Mathematical description

The general expression can be:

$$\text{Walking duration} = F(\text{factor}_1, \text{factor}_2, \text{factor}_3, \dots \dots)$$

- Real Walking duration obtained from person trip survey

$$Y_1 \min = F(\text{factor}_1, \text{factor}_2, \text{factor}_3, \dots \dots)$$

Where the  $Y_1$  is the reflection of the walking duration from departures to stations

Real  $\neq$  Acceptable

$Y_1 \sim Y_2$   
Deviation is appeared

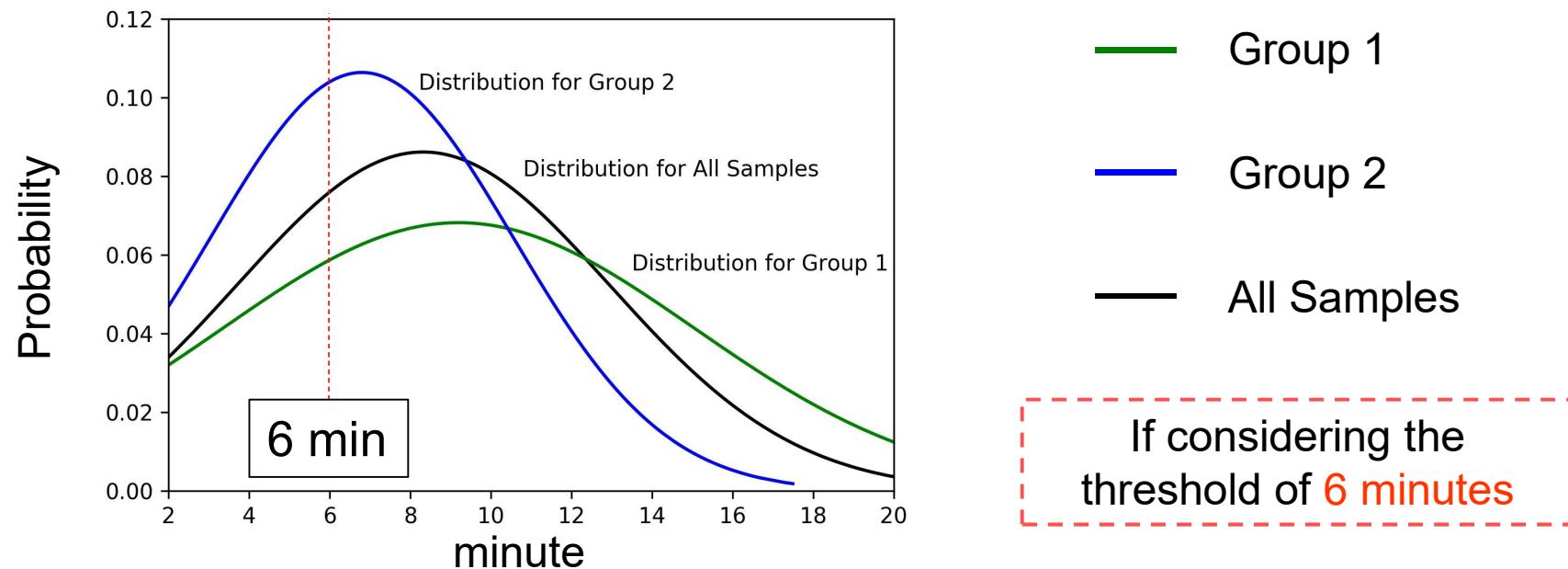
- Acceptable walking duration

$$Y_2 \min = F(\text{factor}_1, \text{factor}_2, \text{factor}_3, \dots \dots)$$

Where the  $Y_2$  is the acceptable walking duration for the people with those factors

## 2.3 Methods

### ■ Important assumption



- Assuming the distribution of walking duration of the passengers with the **same attributions** subject to the **normal distribution**.
- If the survey result obtained from the questionnaire showed one group of passengers with specific attributes **tend to walk longer**, it means this group of passengers can **accept a longer walking duration**.
- In this graph it can be considered that **Group 1** can accept a **longer walking duration than Group 2**.

## 2.3 Methods

### ■ Model construction

- Abstract form of the model

$$\text{Walking duration} = F(\text{factor}_1, \text{factor}_2, \text{factor}_3, \dots \dots)$$



Continuous value of Walking duration  $t_1 \sim t_2$

Threshold of walking duration  $t_1, t_2, t_3 \dots \dots$

$$D_{y \min} = F(\text{factor}_1, \text{factor}_2, \text{factor}_3, \dots \dots), \quad D_{y \min} \in (0, 1)$$

Where:

$y$  the given threshold of walking duration

$D_{y \min}$  the probability of accepting or not at the threshold of  $y$  minutes.

## 2.4 Selection of valid features

### ■ Selection of thresholds

- According to the existing studies, the acceptable walking duration mainly range from **5 – 13 minutes**.
- In this study, the average walking duration is about **8 minutes**.

The thresholds selected in this study is **5, 8, 13 minutes**.

### ■ Analysis of variance

L : tend to spend **Less** than the threshold  
M: tend to spend **more** than the threshold

5 min threshold		8 min threshold		13 min threshold	
Features	Effect*	Features	Effect*	Features	Effect*
Age over 65	L	Female	M	Age 45-64	M
Peak hour	M	Age 25-44	M	Peak hour	M
O_Null	L	Peak hour	M	P_commuting to work	M
P_commuting to work	M	P_commuting to work	M	P_private purpose	L
P_official business	L	P_official business	L	Valid Features at each threshold	
P_private purpose	L	P_private purpose	L		
P_going home	L	P_going home	L		

## 2.5 Prediction

### ■ Decision tree and forest random

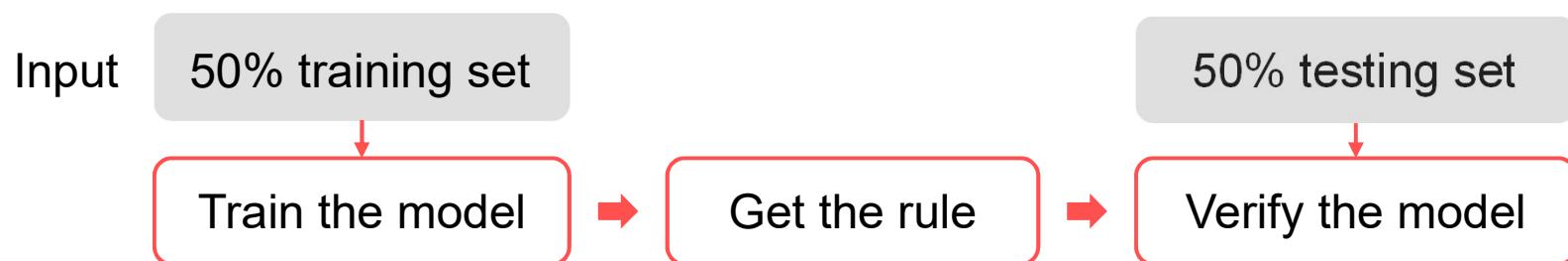
- Reason for choosing decision tree

Used as an **exploratory method** for the dataset **with unknown feature distribution**

- Reason for adding forest random

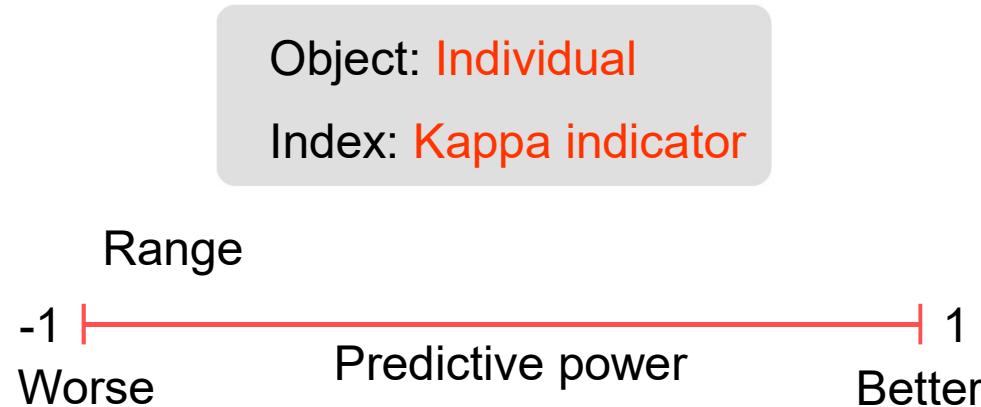
**Reduce the possibility of over-fitting** of data which is easily occurred in the decision tree model.

### ■ Flow

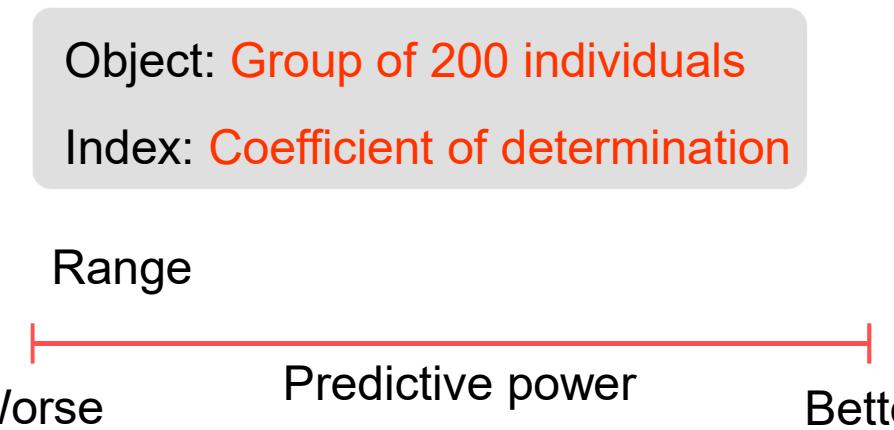


## 2.6 Evaluation

- Predictive power of individual



- Predictive power of group trends



## 2.6 Evaluation

5 min threshold		8 min threshold		13 min threshold	
Features	Effect*	Features	Effect*	Features	Effect*
Age over 65	L	Female	M	Age 45-64	M
Peak hour	M	Age 25-44	M	Peak hour	M
O_Null	L	Peak hour	M	P_commuting to work	M
P_commuting to work	M	P_commuting to work	M	P_private purpose	L
P_official business	L	P_official business	L	Valid Features at each threshold	
P_private purpose	L	P_private purpose	L		
P_going home	L	P_going home	L		

- Kappa index

0.284

0.033

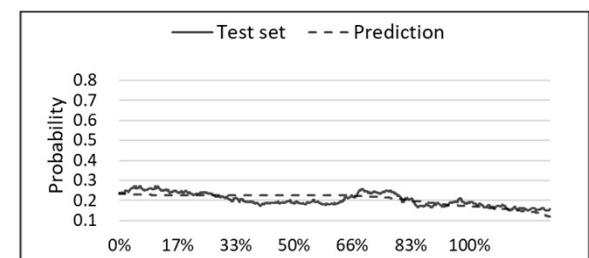
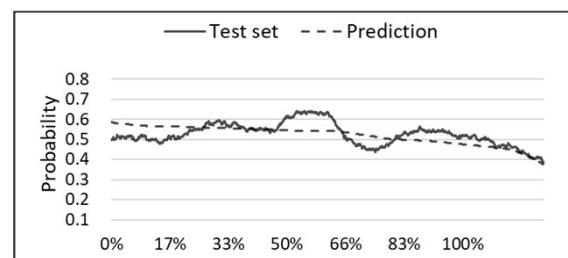
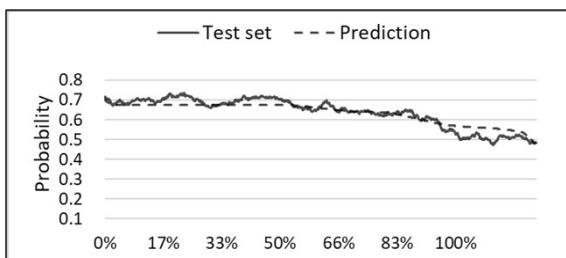
0.118

- Coefficient of determination

0.843

0.221

0.426



## 2.7 Conclusion

### ■ Summary

- Problems

- 1 Little quantitative relationship between walking duration and passengers' attributes was found.

- Achievements

Describe passengers' walking duration preferences by examining the probability that passengers with different attributes walk longer than a given threshold.

- 2 Few of the existing studies quantitatively verified the effect of influencing factors

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

## 2.7 Conclusion

### ■ Findings

- 5 minutes threshold

Individual: Slight

Group trend: High

Walk more than  
the threshold

- Business purpose
- Private purpose
- Going home purpose
- Older than 65
- No occupation

Walk less than  
the threshold

- Peak hour
- Commuting to work

- 8 minutes threshold

Individual: Low

Group trend: Low

- Business purpose
- Private purpose
- Going home purpose

- Age between 25-44
- Female
- Peak hour
- Commuting to work

- 13 minutes threshold

Individual: Low

Group trend: Moderate

- Private purpose

- Age between 45-64
- Peak hour
- Commuting to work

## 2.7 Conclusion

---

### ■ Recommendations

- Encourage office buildings to move close to stations
  - ➡ Results from the factors of business purpose and going home purpose.  
Office workers tend to access rail transit station in a less time.
- The individual prediction is still not accurate enough for predicting individual behavior.
  - ➡ Results from the evaluation of individual prediction
- The group prediction can be used for mastering the overall trend of rail transit utilization in a block.
  - ➡ Results from the evaluation of group prediction

## *Chapter 3*

---

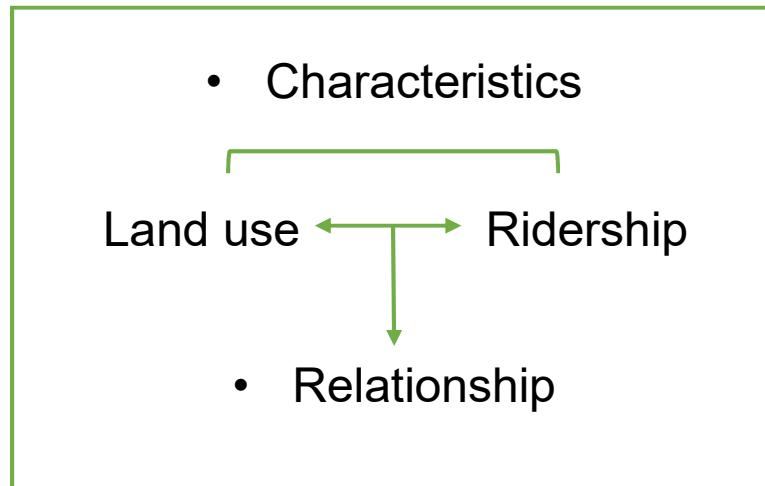
*Analysis on the characteristics of transit ridership and  
land use*

---

## 3.1 Introduction

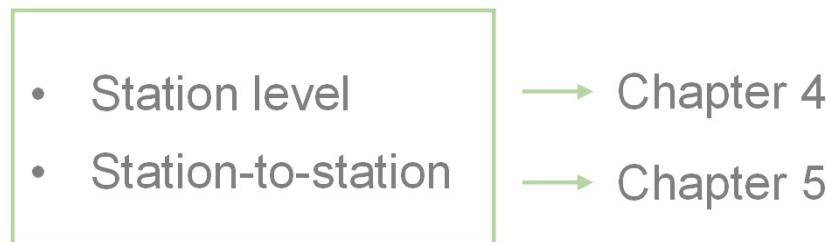
### ■ Research position

- Preliminary study



Chapter 3

- Ridership Estimation



### ■ Purpose

- To grasp the characteristics of transit stations in terms of both transit ridership and land use in Fukuoka.

## 3.1 Introduction

- Problems in previous studies

### Problem 1

The characteristics of rail transit ridership and land use in Fukuoka needs to be clarified.

- Research focus

→ 1 Characteristics of transit ridership

Summary the characteristics of rail transit ridership in spatial and temporal distribution

→ 2 Characteristics of land use

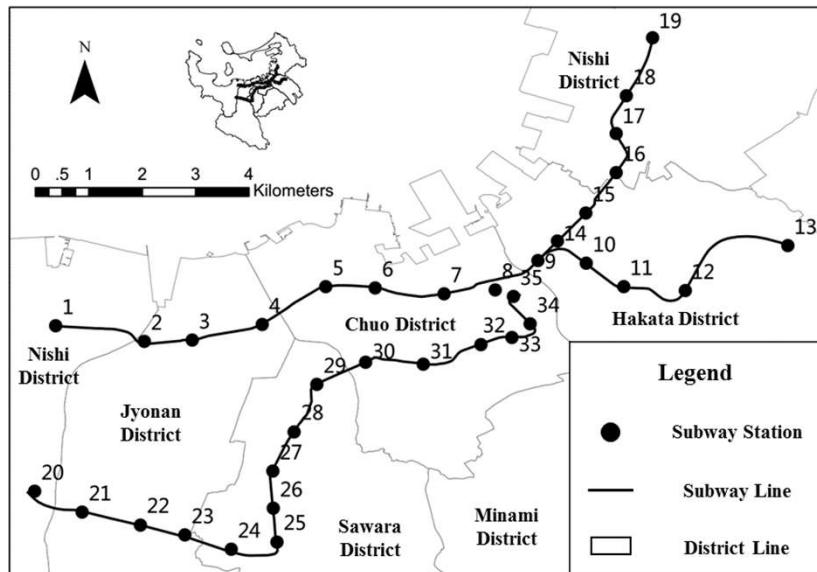
Analyze the characteristics of land use around the subway in terms of the type of land use

→ 3 Characteristics of land use

Preliminary explore the relationship between rail transit ridership and land use

## 3.2 Data

### Study case



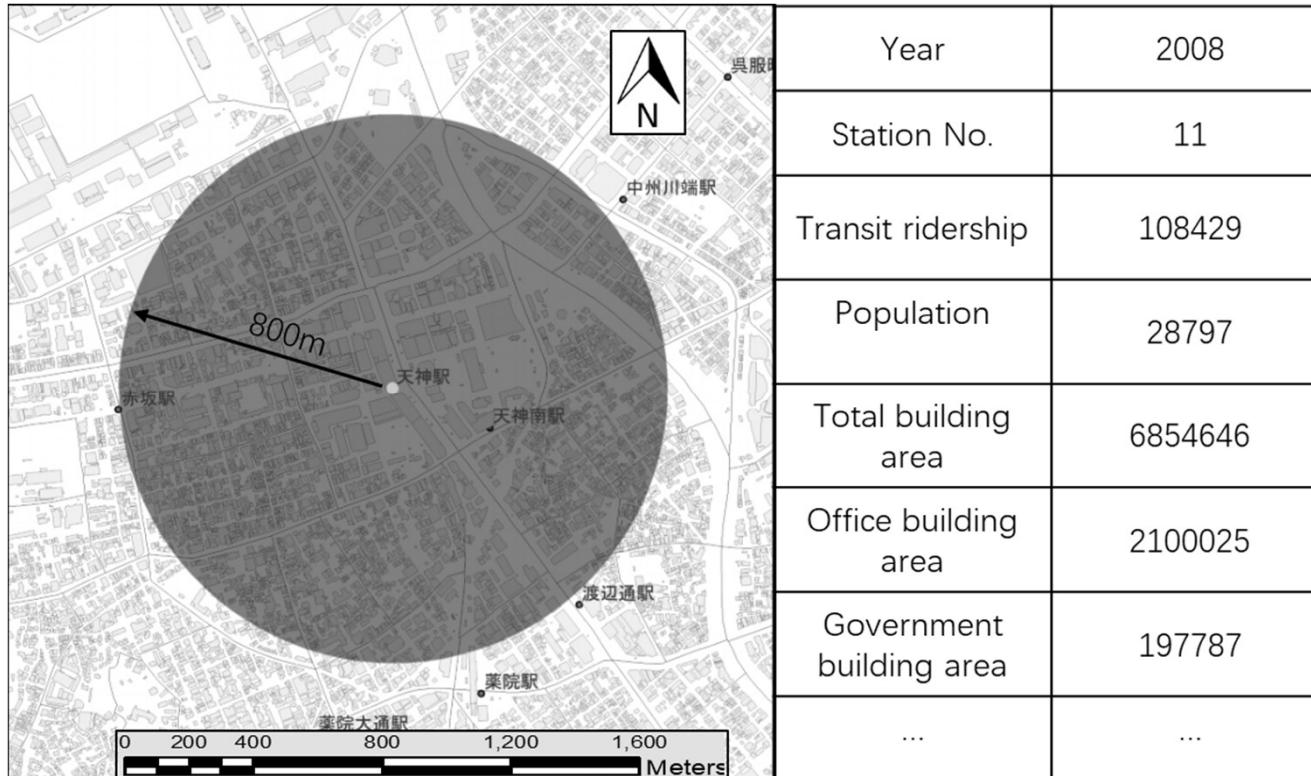
- Fukuoka subway
- 35 subway stations
- 3 operating subway lines
- 29.8 km operating mileage
- 20% share of motorized travel

### Data source

Item	Source	Data accuracy	Time point
Subway transit ridership	Fukuoka Traffic Bureau	Station	2005-2014
Population	Resident Basic Account	Town-chome	2005-2014
Land use	Urban Planning Basic Survey	Building	2003, 2008, 2012

## 3.2 Data

### ■ Data extraction

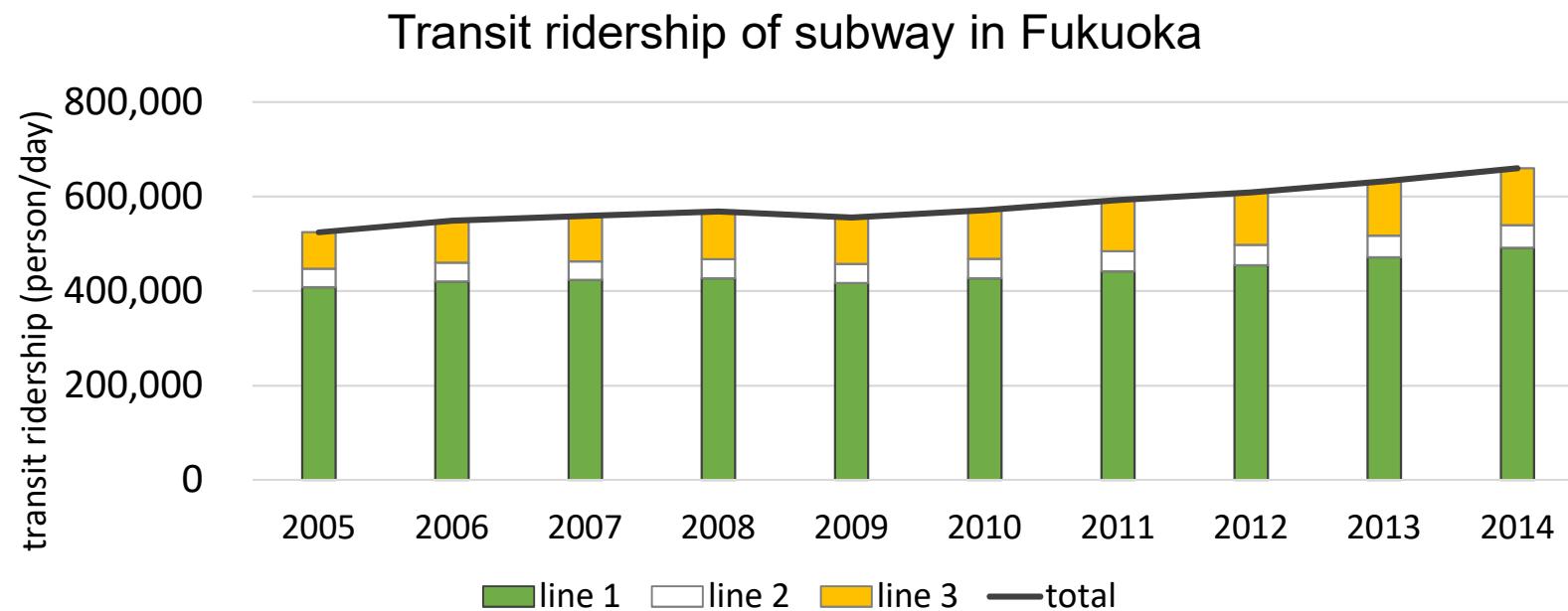


1. Matching the **region** of data.
2. Matching the **time points** of data.
3. Extracting the **data within the catchment area**.

### 3.3 Characteristics of transit ridership and land use

#### ■ Transit ridership characteristics

- Trends of variation in transit ridership

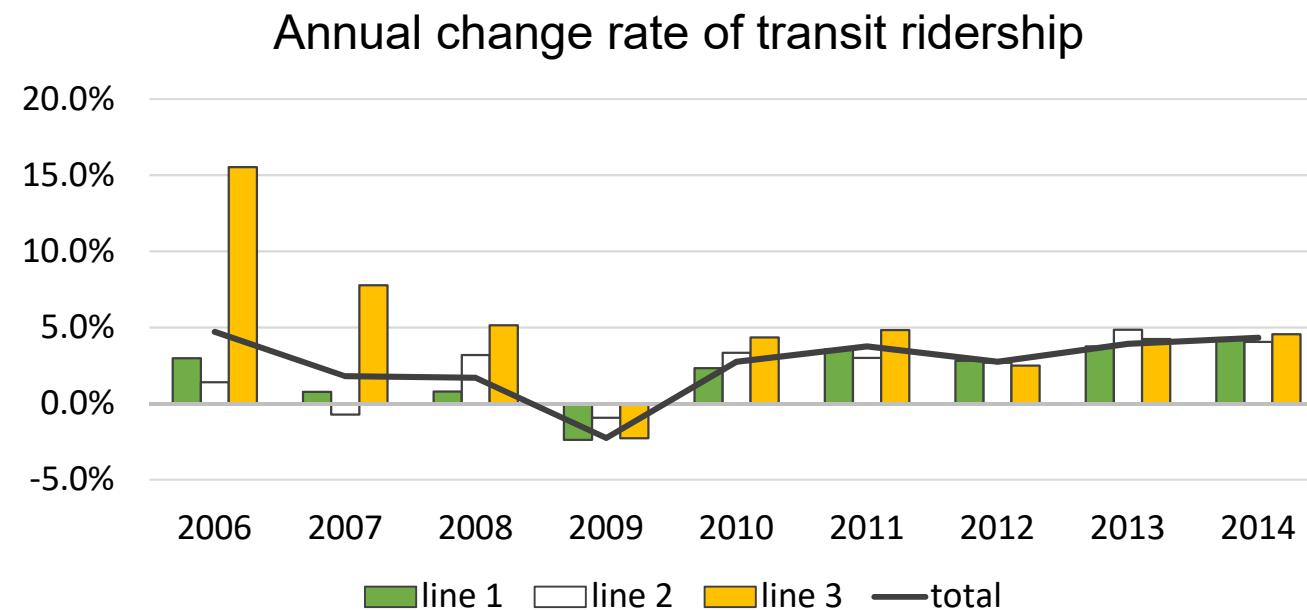


- Growth trend.
- More than 600,000 passengers per day.
- Significant difference in ridership among the three lines.

### 3.3 Characteristics of transit ridership and land use

#### ■ Transit ridership characteristics

- Annual growth rate of transit ridership

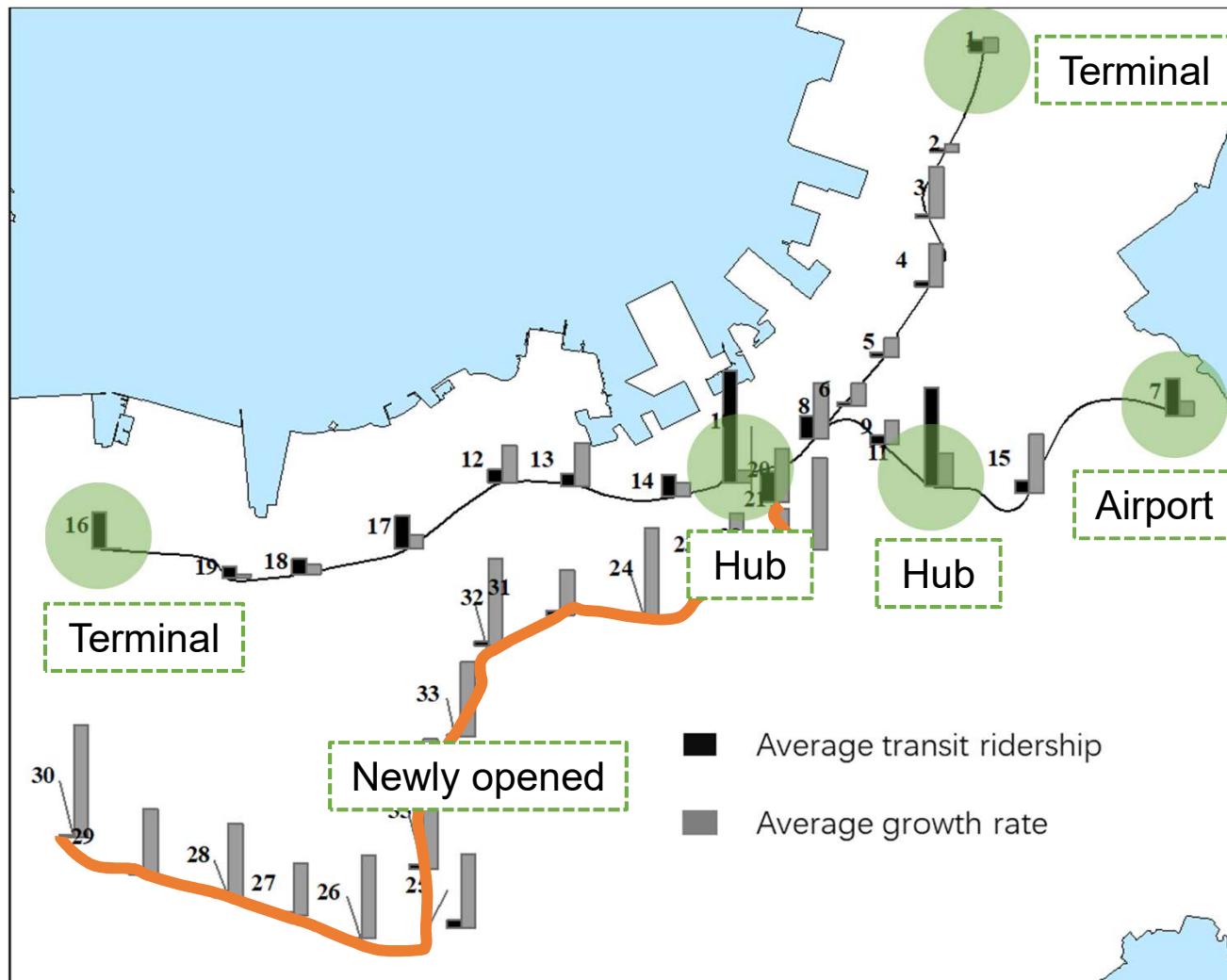


- Line 3 had a much **higher growth** rate at the first few years of the opening
- The growth rate of three lines tend to be **stable** after the year of 2009

### 3.3 Characteristics of transit ridership and land use

#### ■ Transit ridership characteristics

- Spatial distribution of transit ridership



### 3.3 Characteristics of transit ridership and land use

#### ■ Transit ridership characteristics

- Classification based on transit ridership

Transit ridership	Total	Line 1	Line 2	Line 3
Hub	2	2	0	0
Large	4	3	0	1
Medium	11	8	1	2
Small	18	0	4	4

Growth rate	Total	Line 1	Line 2	Line 3
0-2.5	12	8	4	0
2.5-4.5	11	5	2	4
4.5-6.0	6	0	0	6
6.0-	6	0	0	6

Line 1

- Large scale
- Low growth rate

Line 2

- Small scale
- Low growth rate

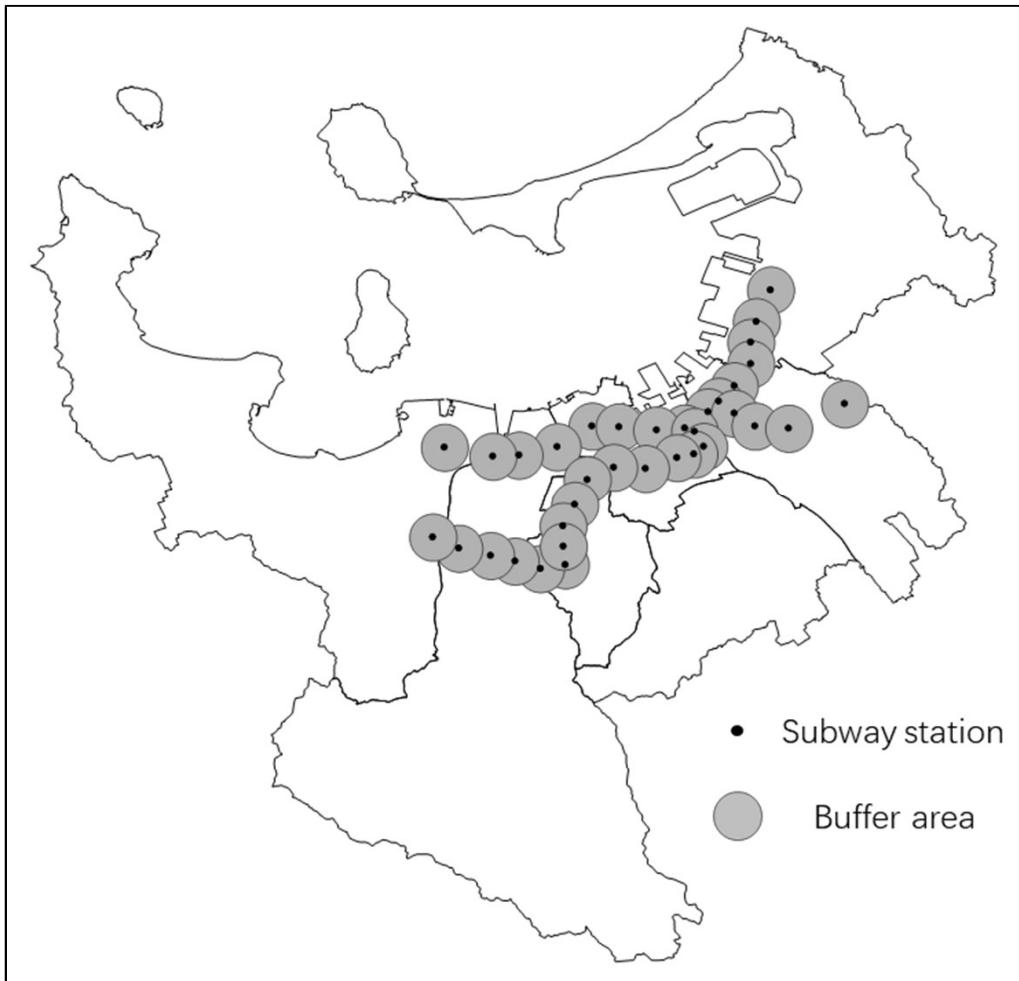
Line 3

- Small scale
- High growth rate

### 3.3 Characteristics of transit ridership and land use

#### ■ Land use characteristics

- Data extraction



#### ArcGIS

- 800m-radius region
- Extracting data

#### Main land use

- Business
- Commerce
- Hotel
- Entertainment
- Residence
- Apartment house
- Dwelling with shop
- Government
- Education
- Culture

### 3.3 Characteristics of transit ridership and land use

#### ■ Land use characteristics

- Correlation analysis for indicators

Indicator	1	2	3	4	5	6	7	8	9	10
1. Business	1.000	0.848	0.987	0.775	-0.385	0.378	0.848	0.615	0.006	0.656
2. Commerce	0.848	1.000	0.828	0.765	-0.293	0.345	0.734	0.645	0.065	0.530
3. Hotel	0.987	0.828	1.000	0.716	-0.361	0.382	0.815	0.585	0.000	0.619
4. Entertainment	0.775	0.765	0.716	1.000	-0.322	0.281	0.663	0.439	-0.018	0.552
5. Residence	-0.385	-0.293	-0.361	-0.322	1.000	0.205	-0.287	-0.164	-0.175	-0.063
6. Apartment house	0.378	0.345	0.382	0.281	0.205	1.000	0.649	0.408	0.087	0.729
7. Dwelling with shop	0.848	0.734	0.815	0.663	-0.287	0.649	1.000	0.641	0.038	0.808
8. Government	0.615	0.645	0.585	0.439	-0.164	0.408	0.641	1.000	0.395	0.549
9. Education	0.006	0.065	0.000	-0.018	-0.175	0.087	0.038	0.395	1.000	-0.012
10. Culture	0.656	0.530	0.619	0.552	-0.063	0.729	0.808	0.549	-0.012	1.000



High correlation ( >0.7)

### 3.3 Characteristics of transit ridership and land use

#### ■ Land use characteristics

- Factor analysis

Purpose

- To deal with the **strong collinearity** among indicators
- Further explore the **internal relationship**

- Test for factor analysis

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		
	0.756	
<hr/>		
Bartlett's Test of Sphericity	Chi-Square	346.086
	df	45
	Sig.	0

Suggested value  
> 0.7

Suggested value  
< 0.05

PASSED the test

### 3.3 Characteristics of transit ridership and land use

#### ■ Land use characteristics

- Factor extraction

Indicator	Component Matrix		
	Component		
	1	2	3
Office	0.962	0.122	0.060
Hotel	0.934	0.123	0.050
Commerce	0.878	0.105	0.131
Entertainment	0.850	0.044	-0.030
Dwelling with shop	0.834	0.417	0.125
Apartment house	0.301	0.860	0.134
Residence	-0.521	0.620	-0.234
Education	-0.066	-0.039	0.958
Culture	0.627	0.644	0.057
Government	0.570	0.305	0.582

#### Factor 1: Office & commerce

- Office
- Commerce
- Dwelling with shop
- Hotel
- Entertainment
- Culture

#### Factor 2: Residence

- Apartment house
- Residence
- Culture

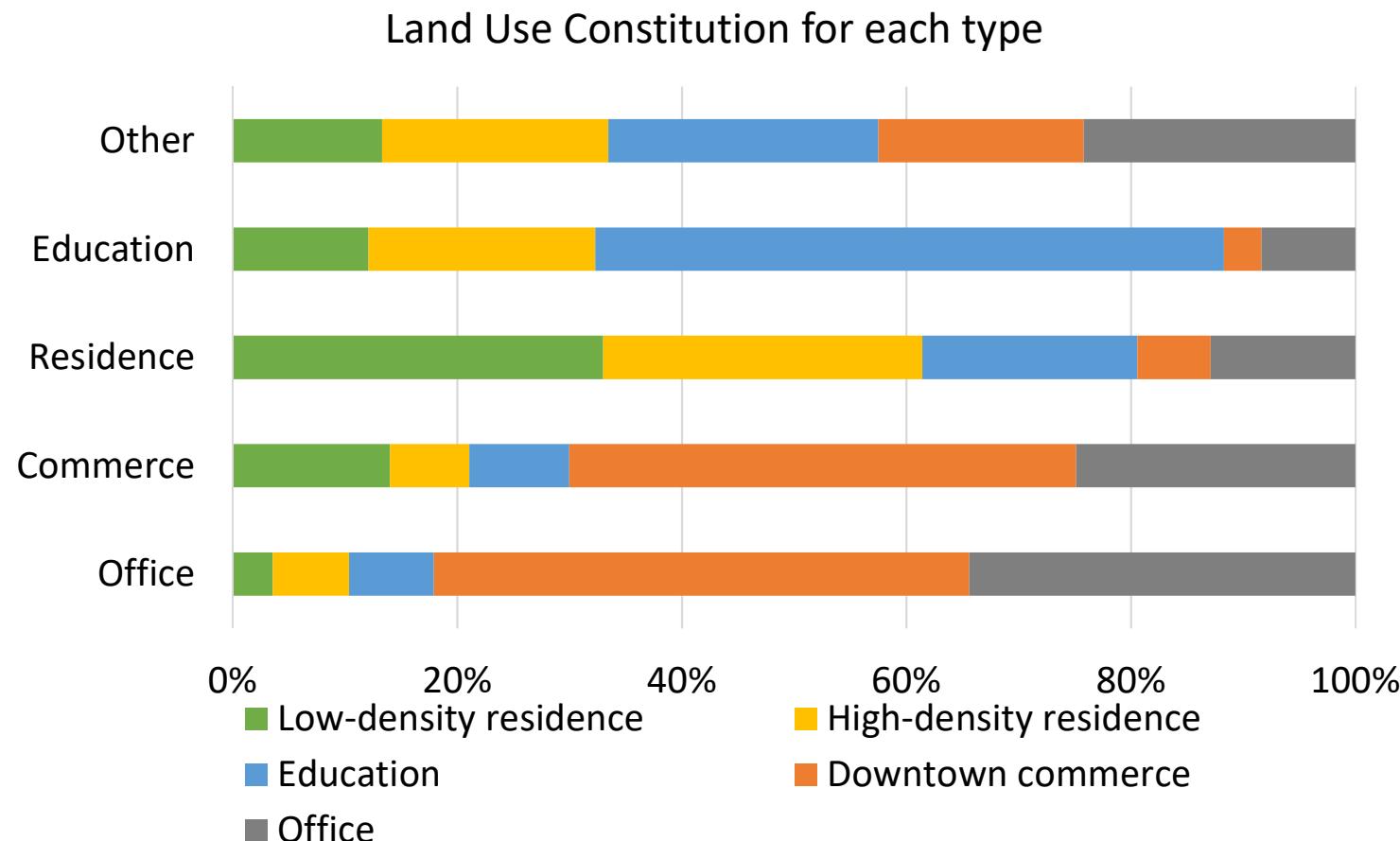
#### Factor 3: Education

- Education
- Government

## 3.4 Exploration the influence on transit ridership

### ■ Land use characteristics

- Station classification based on land use



## 3.4 Exploration the influence on transit ridership

### ■ General relationship

- Summary for each type of station

Type	Office	Commerce	Residence	Education	Transit ridership
Low-density residence	2.85%	7.99%	69.23%	4.47%	3416
High-density residence	5.43%	4.03%	59.66%	7.47%	15633
Education	6.03%	5.08%	40.22%	20.71%	7391
Downtown commerce	38.08%	25.76%	13.68%	1.24%	52300
Office	27.48%	14.20%	27.09%	3.10%	11038

- Cross table of land use and transit ridership

Type	Hub	Large-scale	Medium-scale	Small-scale
Low-density residence	0	0	0	9
High-density residence	0	2	4	3
Education	0	0	1	5
Downtown commerce	2	1	2	0
Office	0	0	3	2



- Large differences in different types
- Low density = Low ridership
- Commercial center = High ridership

## 3.4 Exploration the influence on transit ridership

### ■ Statistical relationship - quantification method I

#### ● Part 1

Factor category	Category	Number	Score	Range
population density (person/ha)	0-40	3	4644	
	40-80	13	2283	
	80-120	8	164	13004
	120-160	8	-2481	13.84%
	160-	3	-8360	
Commerce & Office (m <sup>2</sup> )	0-100,000	16	-6249	
	100,000-400,000	9	-6617	43288
	400,000-1,000,000	5	-4765	46.07%
	1,000,000-	5	36671	
Residence (m <sup>2</sup> )	0-300,000	4	-1450	
	300,000-800,000	20	-5575	16240
	800,000-	11	10664	17.28%
Government (m <sup>2</sup> )	0-1,000	13	-2957	
	1,000-10,000	9	-5501	12267
	10,000-	13	6766	13.06%
Education (m <sup>2</sup> )	0-10,000	5	4842	
	10,000-50,000	11	993	9159
	50,000-100,000	11	-4317	9.75%
	100,000-	8	1544	
Independent variable			Sample size	35
Transit ridership			Coefficient of determination	0.513



#### Dependent variable

- Transit ridership per day

#### Explanatory variables

- Population density
- Commerce & office
- Residence
- Education

#### Coefficient of determination

- 0.513

#### Key factors

- Commerce area
- Office area

## 3.4 Exploration the influence on transit ridership

### ■ Statistical relationship - quantification method I

#### ● Part 2

Factor category	Category	Number	Score	Range
population density (person/ha)	0-40	3	0.0155	0.0265 28.44%
	40-80	13	0.0020	
	80-120	8	-0.0001	
	120-160	8	-0.0110	
Commerce & Office (m <sup>2</sup> )	160-	3	0.0056	0.0097 10.44%
	0-100,000	16	-0.0016	
	100,000-400,000	9	0.0006	
	400,000-1,000,000	5	0.0069	
	1,000,000-	5	-0.0028	
Residence (m <sup>2</sup> )	0-300,000	4	-0.0183	0.0226 24.31%
	300,000-800,000	20	0.0043	
	800,000-	11	-0.0012	
Government (m <sup>2</sup> )	0-1,000	13	0.0111	0.0228 24.54%
	1,000-10,000	9	0.0008	
	10,000-	13	-0.0117	
Education (m <sup>2</sup> )	0-10,000	5	-0.0078	0.0114 12.27%
	10,000-50,000	11	-0.0025	
	50,000-100,000	11	0.0034	
	100,000-	8	0.0036	
Independent variable		Sample size		35
Growth rate of transit ridership		Coefficient of determination		0.537



#### Dependent variable

- Growth rate of transit ridership

#### Explanatory variables

- Population density
- Commerce & office
- Residence
- Government
- Education

#### Coefficient of determination

- 0.537

#### Key factors

- Population density

## 3.5 Conclusion

---

### ● Summary

- Through the summary and analysis, we got a **comprehensive understanding** of rail transit utilization and land use status in the case of Fukuoka City.
- **Provided foundation and references** for the further exploration on the determinants on transit ridership

### ● Implications

- **Pedestrian area** should be considered.
- **Explanatory variables** should be enriched.
- **Model** should be improved.
- The approach of dealing with **small sample case** should be considered.

# *Chapter 4*

---

## *Influencing Factors on Transit Ridership at Station Level*

---

## 4.1 Introduction

---

### ■ Background

- 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. Few studies focus on such a city like Fukuoka with only tens rail transit stations.

### ■ Main purpose

- Explore and explain the factors influencing subway ridership **at station level** using a **small sample case**

## 4.1 Introduction

### ■ Review

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

## 4.1 Introduction

### ■ Research focus

- Problems in previous studies

#### Problem 1

Index system need to be enriched



#### Problem 2

Few studies work on the selection of valid indicators using small sample case.



#### Problem 3

The consideration of spatial correlation among the explanatory variables is inadequate.



- Research focus

#### 1 Improve index system

Propose and reconstruct indicators to improve the index system

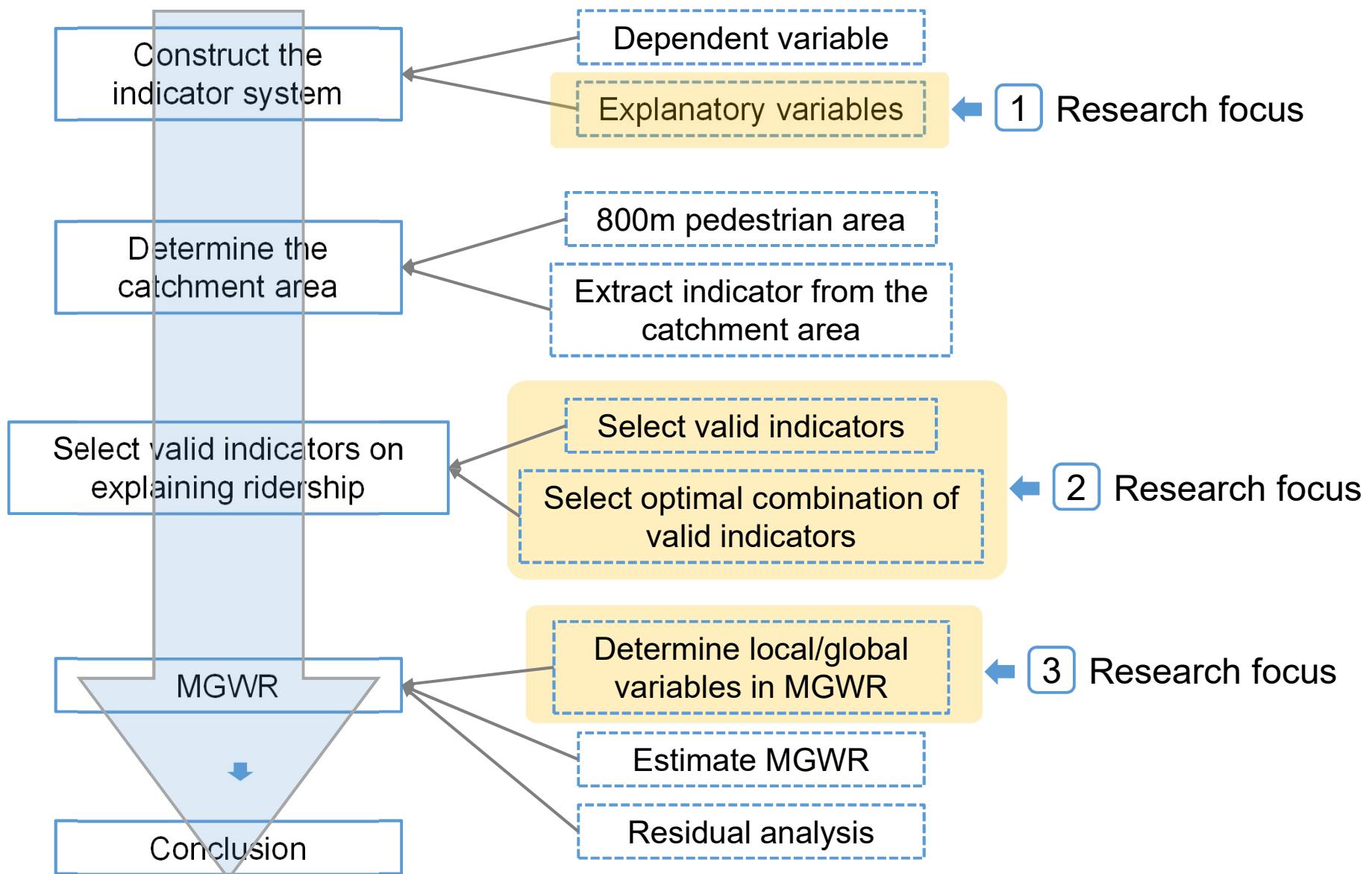
#### 2 Identify valid indicators

Propose effective index screening method for small samples

#### 3 Improve accuracy in estimation

Consider the different spatial distribution of indicators in the spatial regression model by identifying global/local variables.

## 4.1 Introduction



## 4.1 Introduction

### Catchment area of stations

- In general

- Catchment area: walking distance accessing the subway station.  
(Pedestrian Catchment Area, PCA)
- Pedestrian area range: **400m ~ 1000m**
- Most adopted: **800m**

- In the case of Fukuoka

- Walking speed: about **4.8km/h**
- **Average** walking distance: **about 600m**

	Walking		Bicycle	
	Proportion	Ave time	Proportion	Ave time
AVG	78%	7.2	9%	9.0
Mid	85%	7.0	8%	9.5
Max	22%	4.8	0%	0.0
Min	100%	9.2	33%	20.0

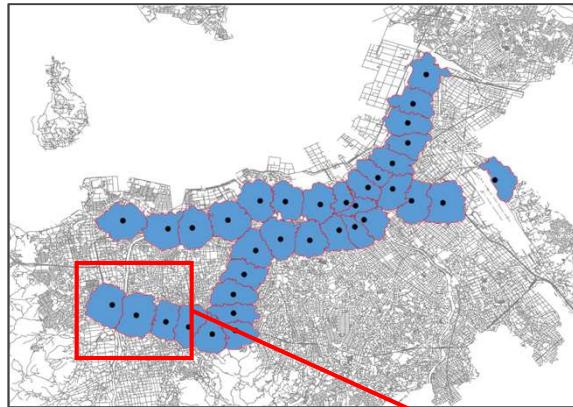


1. Walking is the main method of accessing transit station.

2. Most walking distances are within 800m

## 4.1 Introduction

### Catchment area of stations



Separated along the river



The walking distances along the real road network are equal from any point on the boundary line to both the station on both sides.

## 4.2 Indicators system

### ■ Variable summary

Category	Variable	Expected sign	Min Value	Max Value	Average	Unit
Built environment	Commerce	+	2,921	811,281	114,353	m <sup>2</sup>
	Office	+	2,614	839,956	167,088	m <sup>2</sup>
	Residence	+	110,748	1,067,523	528,533	m <sup>2</sup>
	Education	+	294	305,559	59,691	m <sup>2</sup>
	Government	+	-	128,471	20,878	m <sup>2</sup>
	Transportation Facility Area	+	197	132,777	21,204	m <sup>2</sup>
	Land use Aggregation	+	0.09	0.75	0.31	-
Transportation Accessibility	Transfer Dummy	+	1	4	1.34	-
	Bicycle Parking	Unknown	64	4,375	778	-
	Bus Capacity	Unknown	3	260	58.48	-
	Bus Accessibility	Unknown	4	455	89.71	-
Demographic and Socioeconomic Environment	Road Density	-	191	479	299	m/ha <sup>2</sup>
	Population	+	1,908	19,393	9,813	-
	House Member	-	1.86	2.79	2.18	-
	Population/Job Balance	Unknown	1.27	2.61	1.80	-
	Tenant Proportion	-	0.15	0.65	0.43	%

■ Dependent variable: average daily subway ridership

## 4.2 Indicators system

### ■ Variable interpretation

- Land-use Aggregation

Describing the variety of land use

$$A = \sqrt{\sum(p_i - P_i)^2}$$



$i$  : the type of land use (respectively government, commercial, residence and education)

$p_i$ : the floor area of land use type  $i$  within catchment area of subway station

$P_i$ : the average proportion of  $i$  type land use.

Range of PCA	Residence	Office	Commerce	Education	Total
800	55.3%	17.5%	12.0%	6.2%	90.9%

→ Average  $P_i$

## 4.2 Indicators system

### ■ Variable interpretation

- Indicators of bus system

Supposition

Bus system has both **positive and negative effect** on rail transit ridership



**Share** ridership from rail transit

VS

**Bring** ridership to rail transit

- Bus capacity

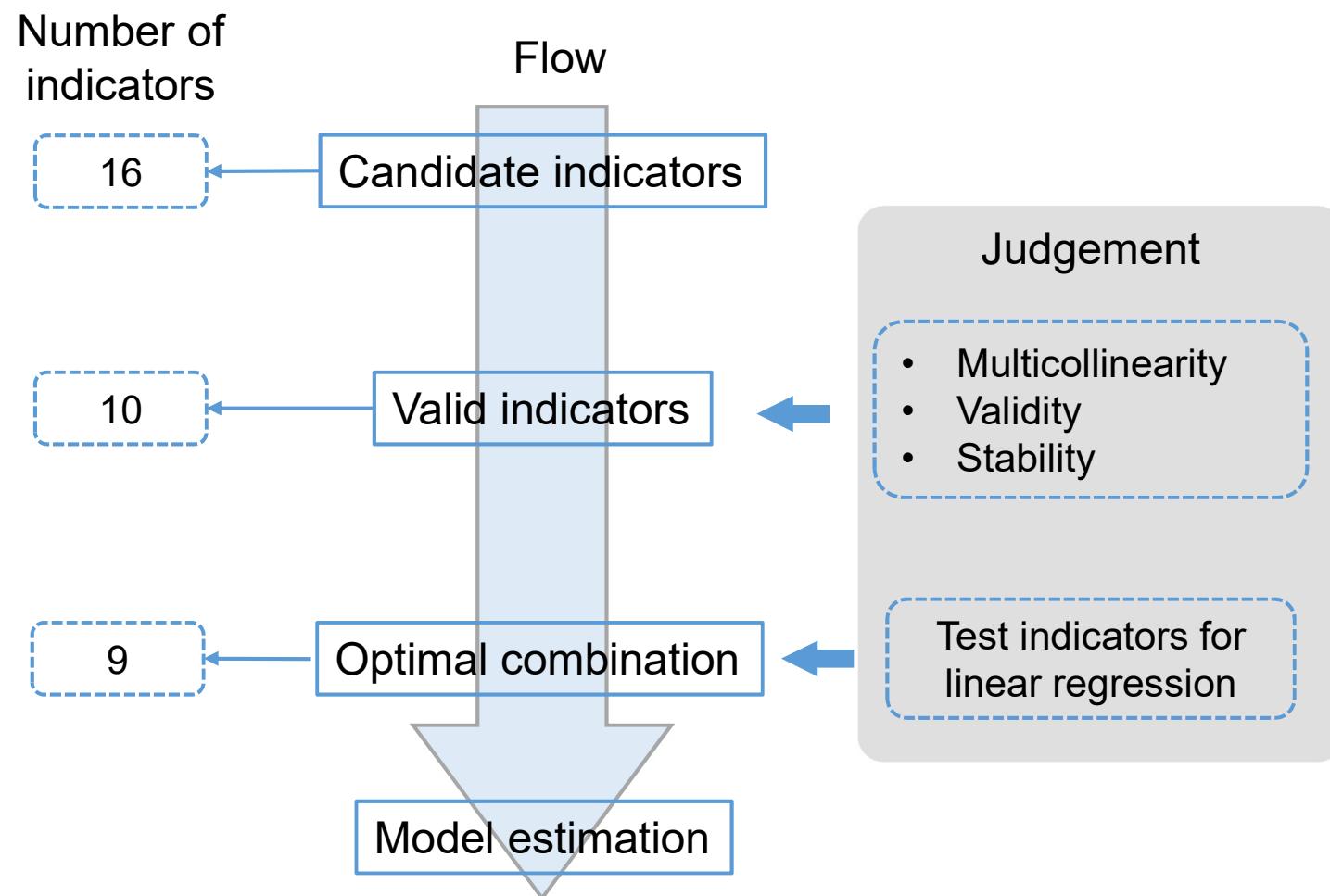
**Share passengers** from rail transit. **Negative effect** on rail transit ridership

- Bus Accessibility

**Bring passengers** to rail transit. **Positive effect** on rail transit ridership

## 4.3 Identification of valid factors

### ■ Flowchart of screening valid factors with small sample



## 4.3 Identification of valid factors

### ■ Valid factors

Category	Variable	Suggested value VIF< 7.5; Validity > 10%; Stability >90%					
		VIF	Validity	stability	times	+	-
	Total				367	253	114
Built environment	Commerce Area	8.8	32.1%	100.0%	25	25	0
	Office Area	10.4	9.0%	71.4%	7	5	2
	Residence Area	27.5	12.8%	70.0%	10	7	3
	Education Area	1.8	7.7%	100.0%	6	6	0
	Government Area	1.8	42.3%	100.0%	33	33	0
	Transportation Facility	3.0	67.9%	100.0%	53	53	0
	Land use Aggregation	2.5	50.0%	100.0%	39	39	0
Transportation Accessibility	Transfer Dummy	3.0	34.6%	100.0%	27	27	0
	Bicycle Parking	4.5	44.9%	100.0%	35	35	0
	Bus Capacity	6.6	37.2%	96.6%	29	1	28
	Bus Accessibility	6.8	19.2%	100.0%	15	15	0
	Road Density	2.3	1.3%	100.0%	1	1	0
De Socioeconomic Environment	Population	23.9	12.8%	60.0%	10	6	4
	Household Members	3.0	39.7%	100.0%	31	0	31
	Population/Job Balance	3.0	37.2%	100.0%	29	0	29
	Tenant Proportion	2.2	21.8%	100.0%	17	0	17

■ The cell marked with dark color means that it doesn't meet the judgment.

## 4.3 Identification of valid factors

### ■ Optimal combination of factors

Independent variables	Test model		
	Beta	Sig	VIF
Government Area	0.10	0.02	1.36
Transportation Facility Area	0.20	0.00	2.31
Land use Aggregation	0.10	0.03	1.42
Bicycle Parking	0.46	0.00	2.60
Bus Capacity	-0.20	0.01	3.56
Bus Accessibility	0.26	0.00	4.71
Transfer Dummy	0.26	0.00	2.79
Population/Job Balance	-0.10	0.05	1.94
Tenant Proportion	-0.09	0.03	1.32
Household Members	-	-	-
Residual sum of squares	337744990		
Adjusted $R^2$	0.96		
AICc	694.39		
Jarque-Bera test (Sig)	0.61	No biased standard errors due to heteroscedasticity	
Koenker (BP) test (Sig)	0.85	Not deviating from a normal theoretical distribution	
SA test (Sig)	0.27	Not spatial autocorrelated	

## 4.4 Model estimation

### ■ Identification for global/local variable

Variable	Moran's Index	Z-score	P-value*	Pattern	DIFF of Criterion**	Type
Government Area	0.04	0.66	0.51	Random	-	Global
Transportation Facility	0.29	3.33	0.00	Clustered	-1.95	Local
Land use Aggregation	-0.01	0.20	0.84	Random	-	Global
Transfer Dummy	0.13	1.58	0.12	Random	-	Global
Bicycle Parking	-0.12	-0.92	0.36	Random	-	Global
Bus Capacity	0.70	7.57	0.00	Clustered	0.18	Local
Bus Accessibility	0.45	5.04	0.00	Clustered	0.04	Local
Population/Job Balance	0.77	7.61	0.00	Clustered	-0.17	Local
Tenant Proportion	0.24	2.54	0.01	Clustered	1.02	Local

\* Statistical significance means autocorrelation

\*\* More than or equal to 2.00, means the local term is better to be assumed as global

## 4.4 Model estimation

### ■ Estimation of Mix Geographically Weighted Regression

Variable	Type	OLS model			MGWR model		
		B	SE	t	B	SE	t
Government Area	Global	0.05	0.02	2.41	0.05	0.02	2.59
Land Use Aggregation	Global	11,832.53	5323.15	2.22	13,384.34	5,408.81	2.48
Transfer Dummy	Global	6,578.09	1216.82	5.41	5,968.65	1,198.72	4.98
Bicycle Parking	Global	7.51	0.93	8.11	7.72	0.90	8.59
Transport Area	Local	0.11	0.03	3.58	0.10	0.02	-
Bus Capacity	Local	-59.21	20.71	-2.86	-55.14	5.94	-
Bus Accessibility	Local	50.71	14.64	3.46	48.61	2.43	-
Population/Job Balance	Local	-2,762.24	1,079.33	-2.56	-2,411.17	364.44	-
Tenant Proportion	Local	-9,987.81	4,380.30	-2.28	-10,304.66	756.09	-
Best bandwidth		5.7km					
AICc		694.39			690.60		
Residual sum of squares		337,744,989			296,311,499		

## 4.5 Model estimation

### Distribution of residuals

Indicator	OLS	MGWR
Moran's index	0.08	0.03
Expected index	-0.03	-0.03
Variance	0.01	0.01
z-score	1.09	0.61
p-value	0.27	0.54

← Closer to the expected value is better

← Less is better

- Moran's index in MGWR is closer to the expected index than that in OLS
- Z-score in MGWR is lower than that in OLS

## 4.6 Conclusion

### ■ Summary

- Problems

1

Indicator system need to be enriched



- Achievements

Indicators system are constructed from 3 categories. Redefine the indicators of **land use variety**. The indicators of bus system are defined into **bus capacity** and **bus accessibility**.

2

Selection of effective indicators under small sample



To reduce the possibility of statistical 1,2 errors in small samples, use **exploratory regression to screen effective indicators** and the most metric combinations.

3

Insufficient consideration of spatial relevance of indicators



**Identified global/local indicator** in MGWR using the spatial autocorrelation of indicators.

## 4.6 Conclusion

### ■ Findings

- Increasing ridership

- Government Area
- Land Use Aggregation
- Transfer Dummy
- Bicycle Parking
- Transport Area
- Bus Accessibility

- Decreasing ridership

- Bus Capacity
- Population/Job Balance
- Tenant Proportion

## 4.6 Conclusion

### ■ Recommendation

- Increase bicycle parking
  - Results from the indicator of **bicycle parking**
- From the perspective of **bus lines and frequency**, reduce the overlap rate between bus lines and rail transit lines, **strengthen the transfer function** of bus system
  - Results from the indicator of **bus accessibility** and **bus capacity**
- Encourage new **construction of apartments** around the rail transit station
  - Results from the indicator of **tenant proportion**

# *Chapter 5*

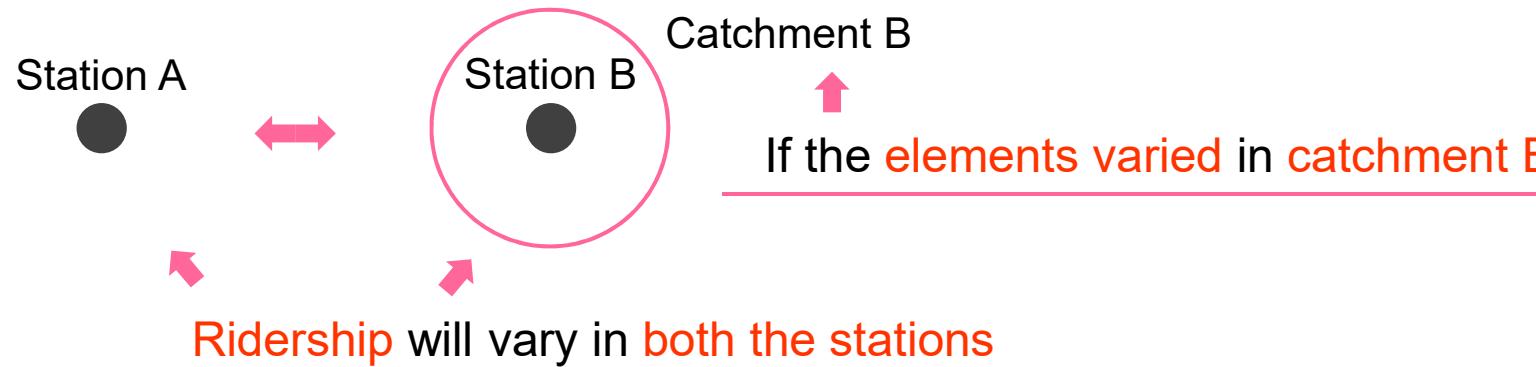
---

## *Influencing Factors on Transit Ridership at Station-to- Station Level*

---

## 5.1 Introduction

### ■ Background



- The variation of land use in the catchment B can affect the probability that passengers boarding from station A choose station B as the destination

### ■ Main purpose

- Exploring the factors influencing transit ridership at station-to-station level

## 5.1 Introduction

### ■ Review

Previous study		Type 1	Type 2	Type 3	This
Perspective	Station	●			
	Station-to-station		●	●	●
Research object	Dependent variable	Transit ridership	Transit ridership	Transportation Impedance	Destination selection probability
	Explanatory variable	Single station	Both O and D stations	Both O and D stations	Both O and D stations
Method	Linear Regression	●	●		
	Gravity model			●	
	Probability regression				●

## 5.1 Introduction

### ■ Research focus

- Problems in previous studies

#### Problem 1

On the research object, the ridership between ODs should be produced by the origin station, and then transferred to the destination station. Direct model for OD transit ridership is lack of this process on estimating the transfer between OD stations.



- Research focus

#### 1 Describe the transfer between station and station

To find an indicator to describe the passenger transfer between station and station

## 5.1 Introduction

### ■ Research focus

- Problems in previous studies

#### Problem 2

In terms of method, the issue of analyzing the transfer of ridership between station and station usually adopt the gravity model, however, the gravity model is not used for estimating the influence of indicators.



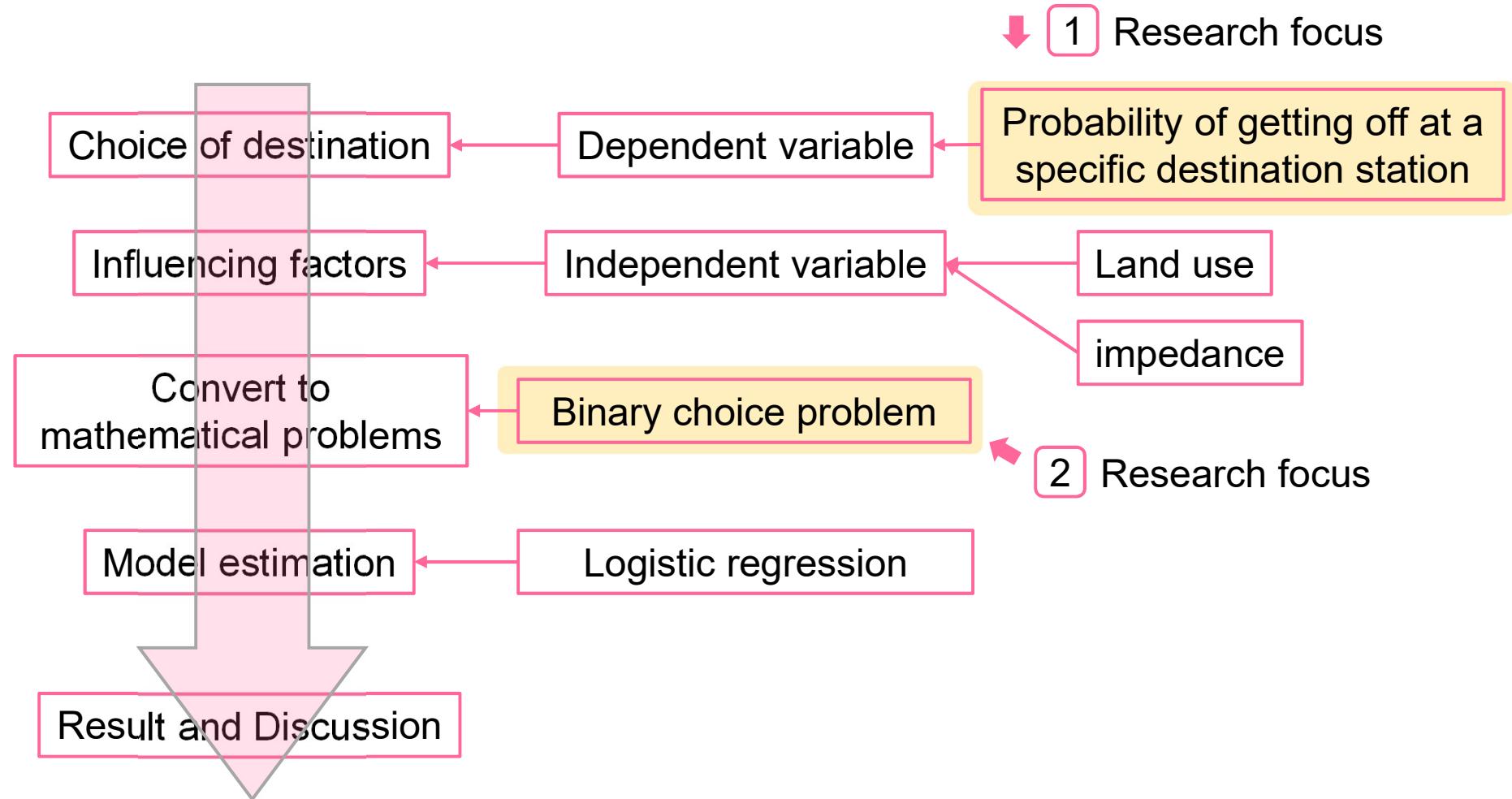
- Research focus

2 Estimate the influencing factors on the transfer of ridership between station and station

Convert this issue into the mathematical problem, then estimate the effect of influencing factors.

## 5.1 Introduction

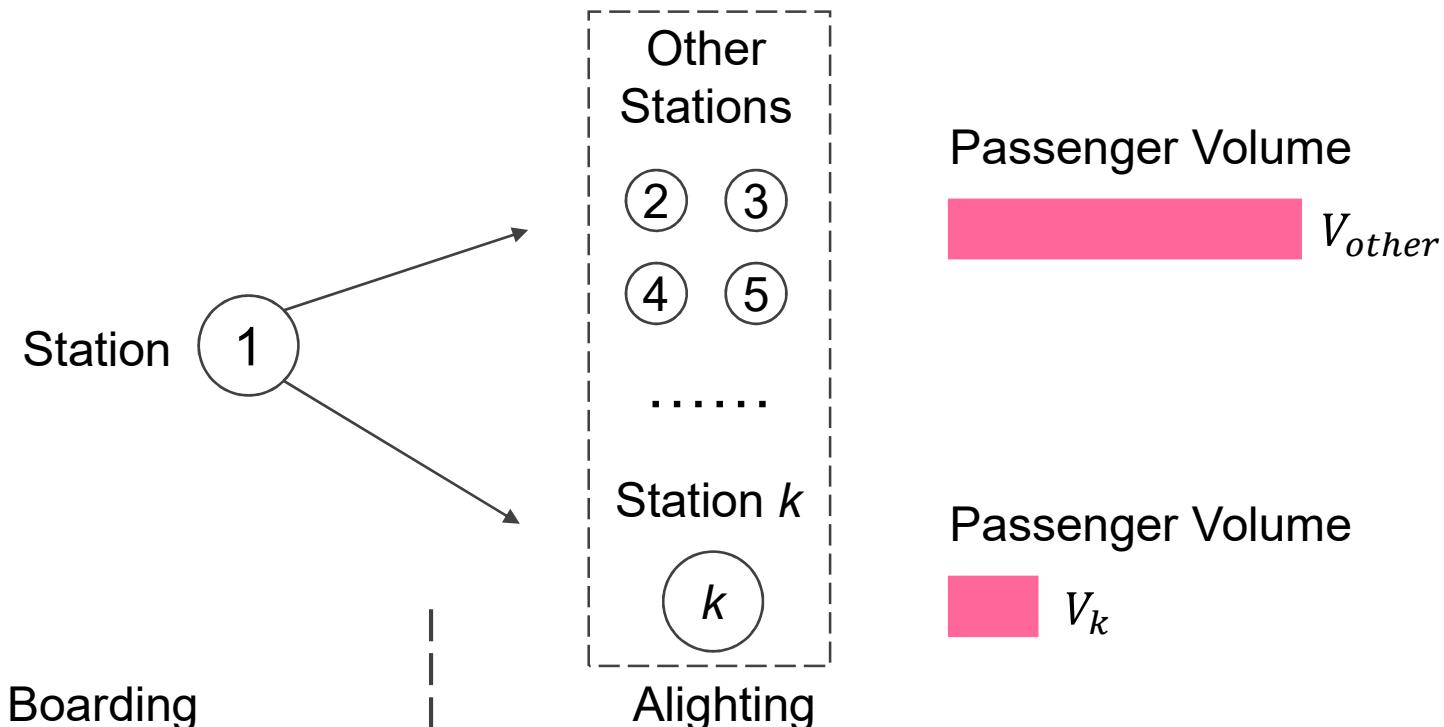
### ■ Flowchart



## 5.2 Research object

### ■ Dependent variable

- Probability of choosing the destination



For the passengers boarding from station 1  
The probability of getting off at station  $k$

$$P = \frac{V_k}{V_{other} + V_k}$$

## 5.2 Research object

### ■ Independent variables

- Land use

- Residence proportion
- Office proportion
- Commerce proportion
- Education proportion
- Land-use aggregation

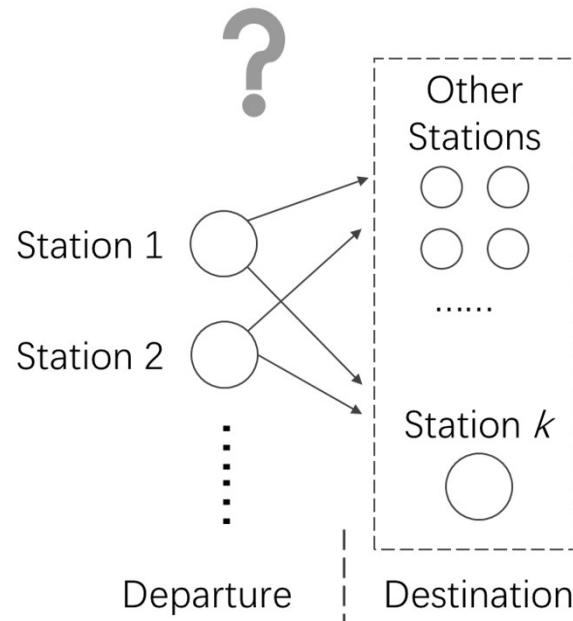
- Impedance

- Bus capacity
- Operation distance of subway
- Bus accessibility

## 5.3 Method

### ■ Mathematical problems

- Conversion



- Binary discrete model



$$P_k = F(x_1, x_2, x_3, \dots)$$

Probability of getting off at the investigated station

Explanatory variable for making the choice of whether getting off



- Logistic regression model

## 5.3 Method

### ■ Selecting sample station

Station type	Station name	Population density	Commerce proportion	Office proportion	Residence proportion	Education proportion
Low-density residence	賀茂	98 Person/Ha	4%	3%	83%	4%
High-density residence	藤崎	173 Person/Ha	5%	7%	76%	6%
Education	箱崎九大前	83 Person/Ha	5%	6%	51%	22%
Office	呉服町	135 Person/Ha	8%	31%	49%	3%
Commerce	天神	69 Person/Ha	34%	32%	24%	1%

Note:

The highlighted cells refer to the representative values that having **much higher proportion** than other land-use types.

## 5.4 Results and discussion

### Result of estimation

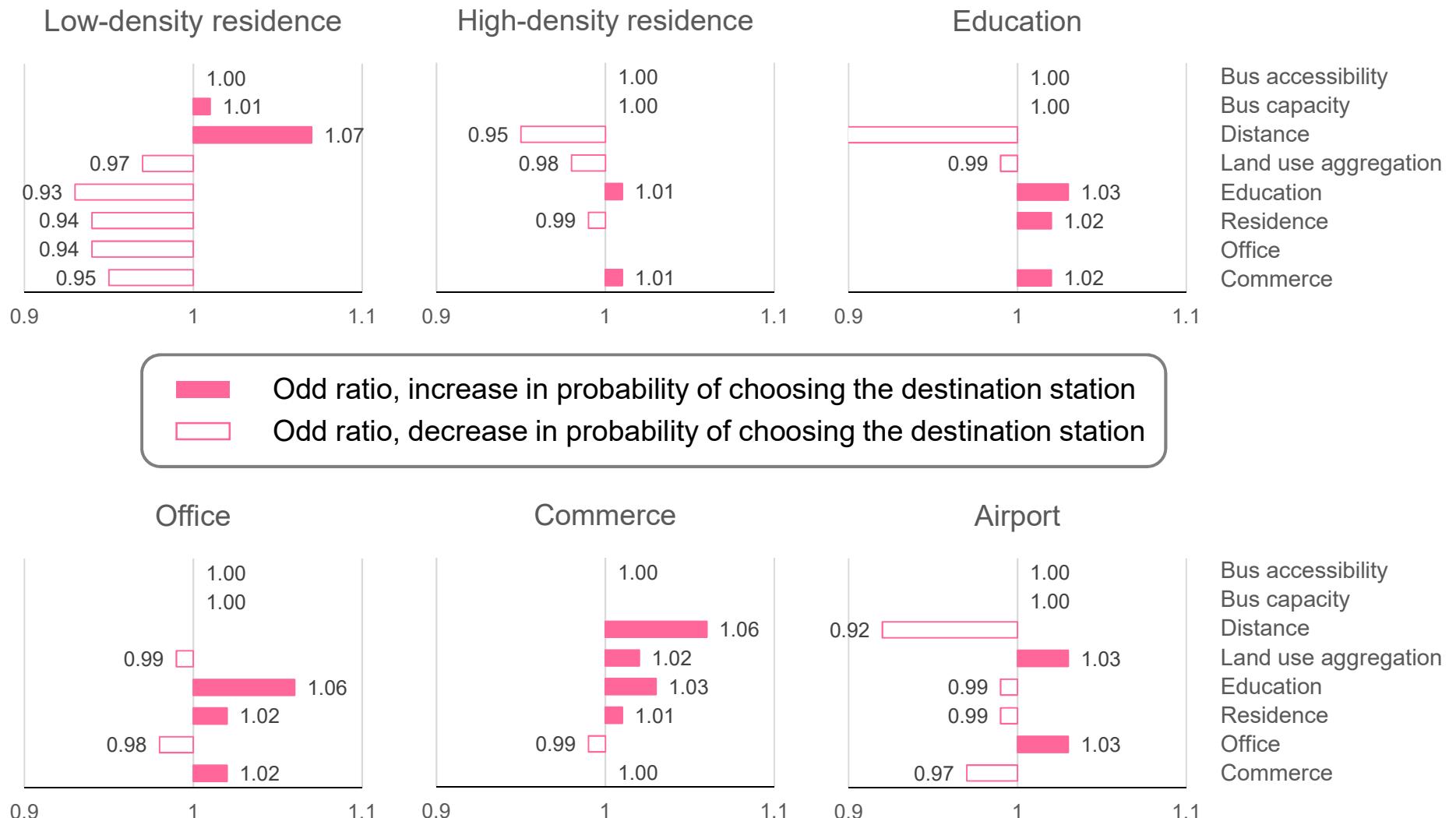
Destination Station		Variables in Departure Station							
Station Type	Station Name	Land-use					Impedance		
		C	O	R	E	L-A	D	B-C	B-A
Low-density residence	賀茂	0.95	0.94	0.94	0.93	0.97	1.07	1.01	1.00
High-density residence									1.00
Education	箱根								0.00
Office	呉服町	1.02	0.98	1.02	1.06	0.99	1.00	1.00	
Commerce	一社								1.00
Airport									1.00

- Note
- Meaning of the coefficients  
1% increase in the proportion of land use will lead to a corresponding variation in the probability of getting off at the destination
  - Full name of the abbreviation  

C: Commerce	R: Residence	L-A: Land-use Aggregation	B-C: Bus Capacity
O: Office	E: Education	D: Distance	B-A: Bus Accessibility

## 5.4 Results and discussion

### ■ Discussion for results



## 5.5 Conclusion

### ■ Summary

- Problems

- 1 Description for the process of passenger transfer needs to be clear.

- Achievements

Described the passenger transfer using the **probability of choosing a specific destination station**.

- 2 The model should be able to reflect the influence extent of the indicators.

The problem is converted into a **binary choice question**. **Logistic regression** is used to estimate the influence of land use on the probability of passenger destination selection, also taking into account the influence of inter-station impedance.

## 5.5 Conclusion

---

### ■ Findings

1. The probability of choosing a destination has a trend of **decreasing** between the **same type of land use**.
2. The probability of choosing a destination belong to **low-density residence** type **has no tend to raise** regarding to the variation of land use in the departure station.
3. Increase in the **education** land use tends to lead **increase** in the ridership between station and station.

## 5.5 Conclusion

---

### ■ Recommendations

- It is recommended to pay more attention to the location of rail transit stations while judging the site selection of a **newly planned school**.
  - Results from **finding 3**
- For **low-density residence type stations**, it is recommended that promoting the utilization of rail transit could start with other measures than land use, such as fare adjustment.
  - Results from **finding 2**
- Combined with the achievement from chapter 4, a complete forecasting model for rail transit ridership is possible to be established.
  - Results from **chapter 4 and chapter 5**

# *Chapter 6*

---

## *Conclusion*

---

## 6.1 Conclusion

---

### ■ Summary

Based on the theme of exploring the determinants on rail transit utilization in Fukuoka City, this study discussed 4 major issues.

- One
  
- Chapter 2 discussed how the walking duration to rail transit station is affected by passengers' attributes. Centering with this problem, 3 specific research questions are proposed according to the previous studies, then the answers to each question were given.

## 6.1 Conclusion

### ■ Summary

Based on the theme of exploring the determinants on rail transit utilization in Fukuoka City, this study discussed 4 major issues.

- Two
  
- Chapter 3 summarized the characteristics of rail transit ridership and land use in the case of Fukuoka, which helped to make a comprehensive understanding of the research object in this dissertation. This chapter also gave some important reference and implication for the next research.

## 6.1 Conclusion

### ■ Summary

Based on the theme of exploring the determinants on rail transit utilization in Fukuoka City, this study discussed 4 major issues.

- Three
- Chapter 4 explored and estimated the influencing factors on rail transit ridership at the station level. With the focus on the case of Fukuoka with small sample size, this chapter proposed improved approach from indicator system, valid indicator selection, and model estimation three aspects, and the results are proved effective and improved.

## 6.1 Conclusion

### ■ Summary

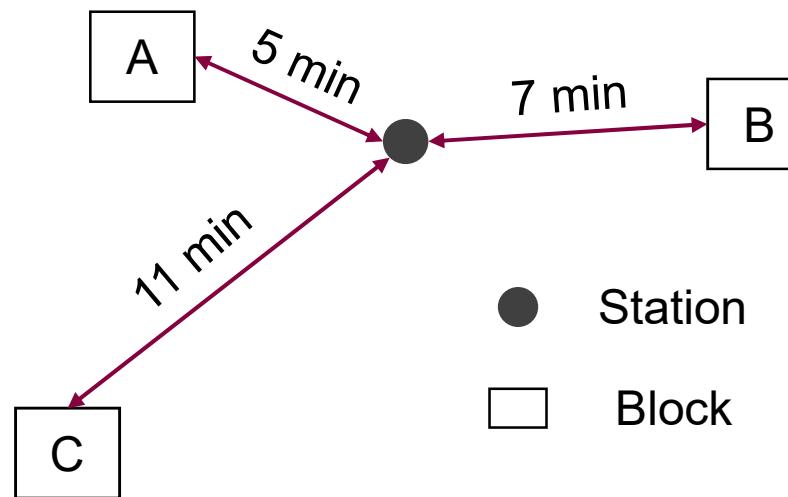
Based on the theme of exploring the determinants on rail transit utilization in Fukuoka City, this study discussed 4 major issues.

- Four
- Chapter 5 gave a description for the transfer of ridership between station and station that use the land use type around a station to explain the probability of choosing that station as the destination station. Then ran logistic regression to estimate the influence of land use on the probability of getting off at the destination station. The influence of land use on the transfer of ridership between station and station is confirmed in this chapter.

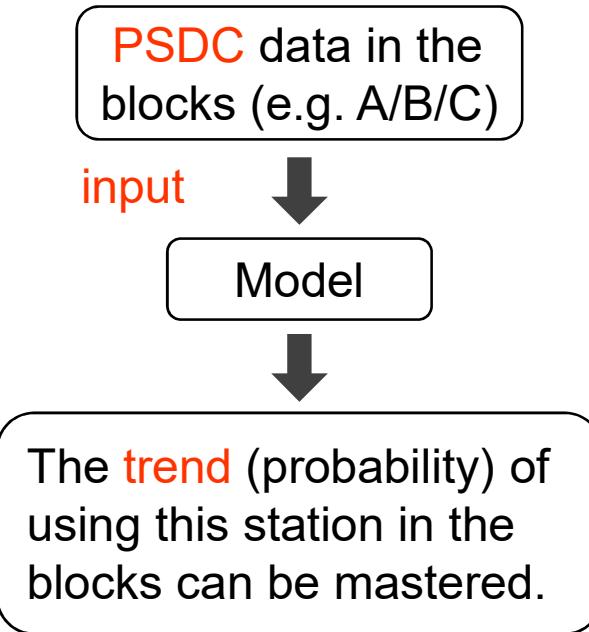
## 6.2 Recommendations

### ■ Application of the achievements

- Achievement from **Chapter 2**



Estimating the catchment area



- Combination with the achievements

- Using the estimated **catchment area** to make prediction of transit ridership
- Using the results of **ridership transfer between station and station** to correct the prediction of ridership at **single station**