问题

1. chapter 2 步行时间的研究与乘客量的预测之间有没有直接关系？

2. chapter 5 为什么站与站之间的客流量不能用线性关系描述？

3. chapter 2 为什么不直接用意愿调查？

4. chapter 5 最后的表要怎么读？

5. chapter 4 能否展示一些筛选有效变量的细节？

**1. In chapter 2, the research about walking duration, I know walking duration should relate to the transit ridership, but you didn't mention it. Can you show me some details about how the walking duration affects transit ridership? （任宇杰）**

The transit ridership is affected by the elements around the station. To estimate the influences on transit ridership, at first, we have to define this “around”, that is the catchment area. With a certain catchment area, the influencing factors can be extracted, and then use these factors to estimate the influence.

The problem in this procedure is how to define the catchment. I will show another slide to explain how can use the walking duration to determine the catchment area.

I will give an example. We have a station in the center, and three blocks around this station, the distances to this station are different. Now if we have the data of personal attributes in each block, we can input this dataset to the model we have got in chapter 2, and then get the trends of using this station in each block.

With this output, obviously, the catchment area of the station can be estimated.

**2. In chapter 5, you used probability to describe the passenger flow between station and station, but why not the linear relationship? （陈思婷）**

Because the passenger flow between two stations is not only determined by one station, it is affected by the transit ridership of both two stations.

I’d like to show an example to help explain it.

Now if we investigate two stations, A and B. Station A has a small scale of transit ridership, but the transit ridership of station B is much larger than that of station A. Limited by the transit ridership of station A, the passenger flow will not always increase with the increase of transit ridership in station B.

So, the functional relationship should not be linear

**3. In chapter 2, you showed a relation map of personal attributes, willingness, and actual walking distance. I can understand the relation of them. But why not conduct a willingness survey directly? （韩博士）**

The willingness is something that not has happened, while the actual walking duration is something that has happened. They should not be the same thing. You can imagine that, if you say, "I can accept an 8-minute walking duration", this 8-minutes is your willingness. But in fact, there are too many reasons making you accept a longer walking duration when you actually decide to walk to the station.

This is one of the reasons. There is another more important reason, that I have to say, I am unable to do such a scale of willingness investigation.

**4. In chapter 5, the slide of the result, I am not quite clear about how to read it. can you explain it once again in detail? （待定）**

In this table, the values less than 1 means Increase in this type of land use will lead to a decrease in the probability of choosing this station as the destination

The values larger than 1 means Increase in this type of land use will lead to an increase in the probability of choosing this station as the destination

For the 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

.

The blank means this value does not pass the statistical significance test.

**5. In chapter 4, the part of identifying the valid factors, can you show some details of the procedure? （新）**

This procedure is conducted by the exploratory regression.

The exploratory regression is to attempt all the combinations of explanatory variables by running the ordinary regression.

In this study, the exploratory regression ran 367 times in total, each variable entered in the models 78 times.

The validity is the proportion of showing statistical significance in all the models.

The stability means the proportion of showing positive or negative effects on the transit ridership.

The variables which have the VIF value less than 7.5, validity more than 10%, stability more than 90% are judged as the valid variable.