**1930026143 薛劭杰**

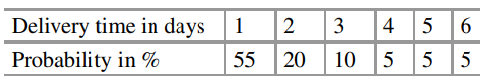
**EBIS 3103 Introduction to Business Data Analytics -**

**Individual Assignment 2**

Instructor: Dr. Jingzhi Zhang

\*For all the questions below, please use R to help you answer the questions if necessary.

1. A company has at most 6 days of delivery of a good. The fastest delivery time is 1 day, and the probabilities for the different delivery times are shown in the table below. Let *X* be the number of days of delivery. [15 marks]



(a) Find the cumulative distribution *F(x)* of *X*. How can we interpret *F(x)* in this case?

(b) Find the expected days of delivery *E[X]=µ.*

Ans:

1. Initially, we know the of this data, then we can calculate the of it.

2. Let X denote the income (in USD) of a randomly selected person. We have made 25 independent observations and found



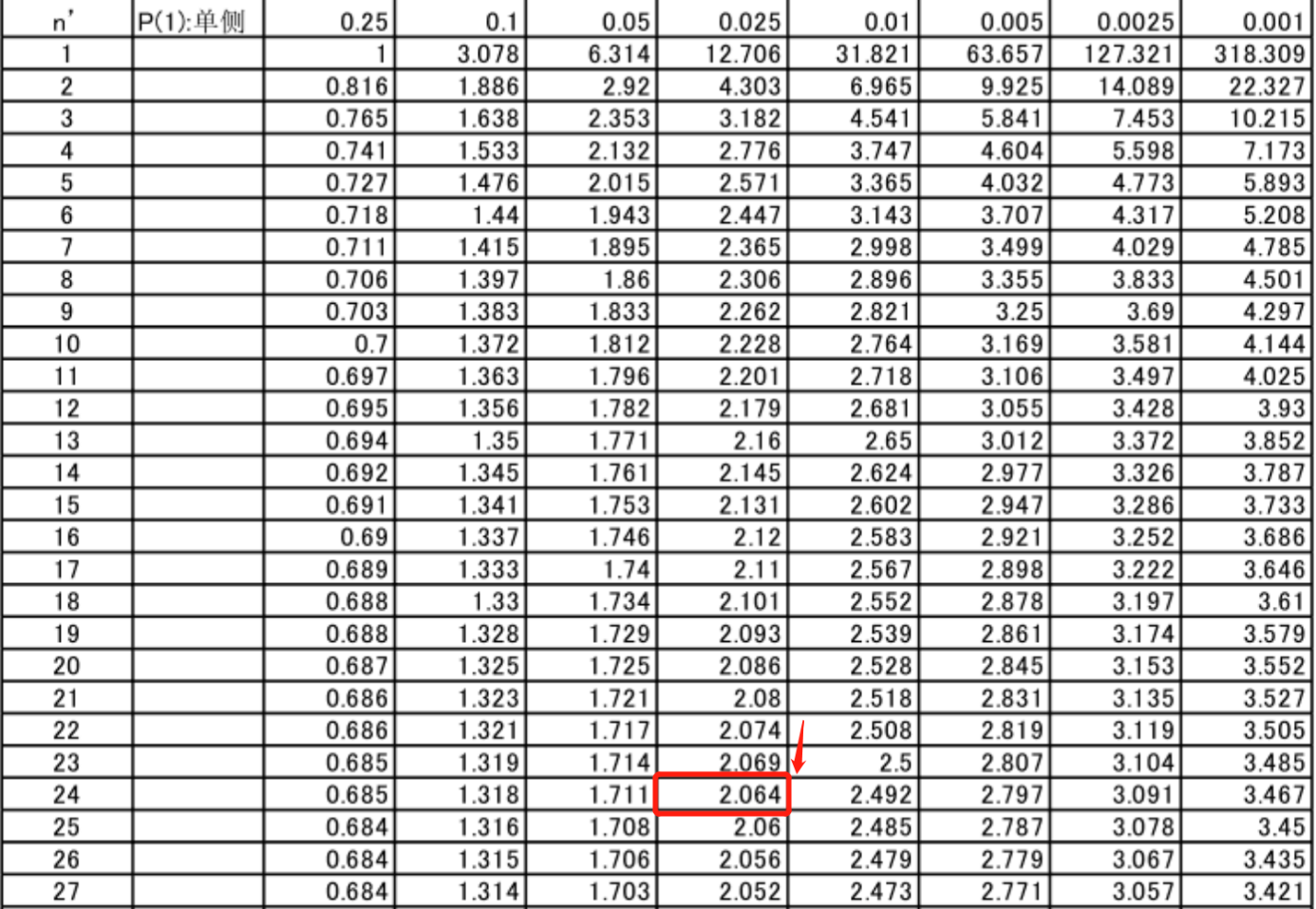
1. Assume that X is approximately normal and find a 95% confidence interval for *E[X]=µ*.

[15 marks]

Ans:

1. We know that the number of independent observations , then we can use the because we the is unknown and is approximately normal.

And the degree of freedom , the confidence interval is 95%. According to the , we can get the of .



We can get , so the interval of 95% confidence is:

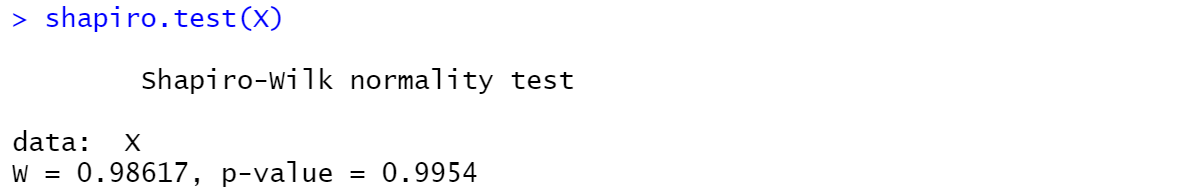
3. A tele-communication company’s past records indicate that individual customers pay on average $220 per month for local data usage. A random sample of 15 customers’ local data usage bills during a particular month produced the following amounts:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 260 | 180 | 290 | 170 | 300 | 210 | 320 | 240 | 280 | 250 | 150 | 270 | 350 | 230 |  |

1. Comment on the distribution of the above sample data. Does it appear to follow normal distribution?[15 marks]
2. At 5% level of significance, is there enough evidence to reveal that the average amount of bills for local data usage per month is high than $250 using both the critical value approach and the p-value approach to test the hypotheses. What assumption has to be made? [15 marks]

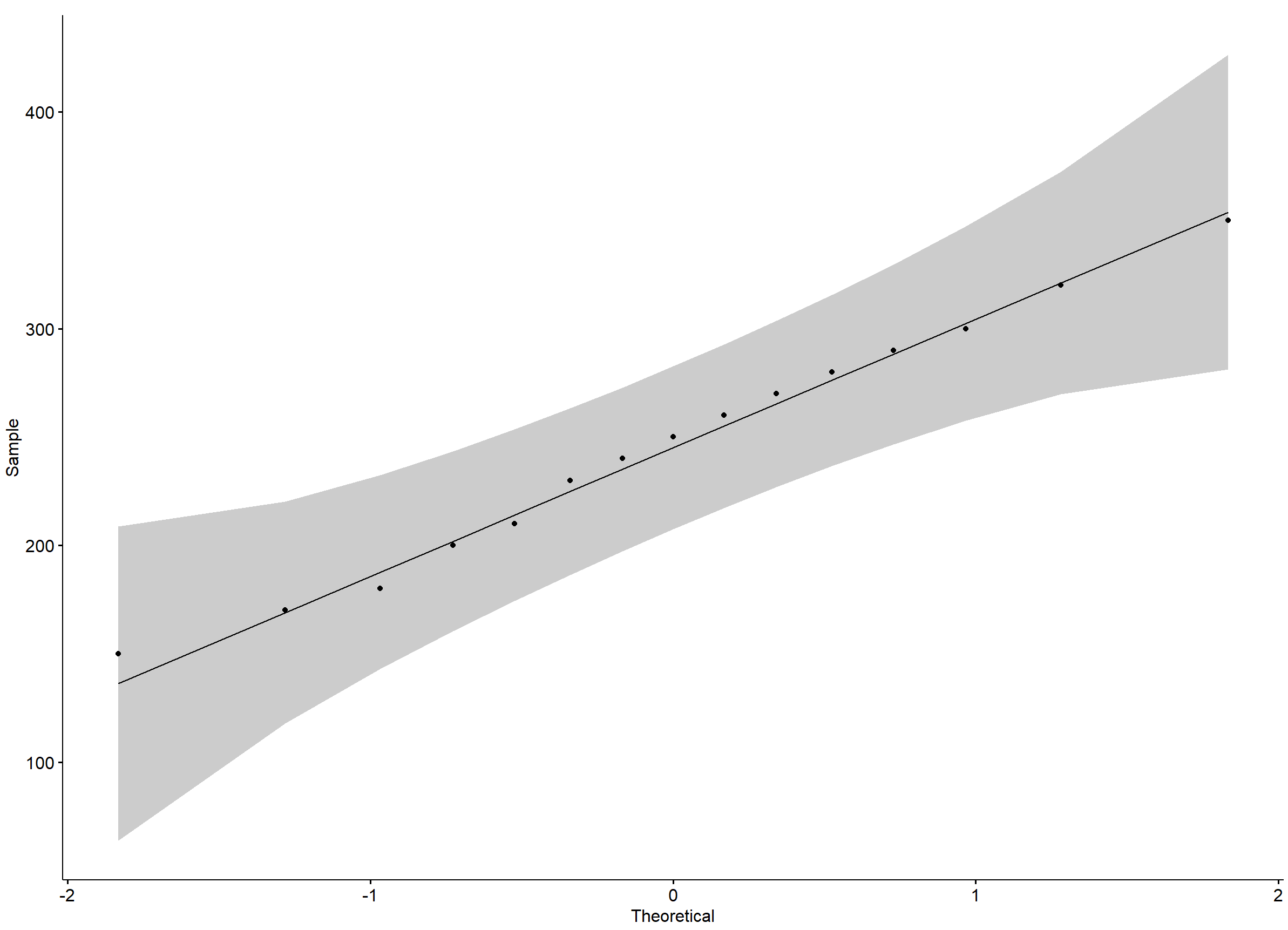
Ans:

1. Firstly, we can use the Shapiro-Wilk test to check whether it follow the normal distribution by using r language:



We can see that the p-value is close to 1 which means that it is hard to reject the null hypothesis and we have strong evidence that it is normal distribution.

In addition, we can use visual method to judge the distribution by , we can find that all of points roughly on that reference line which means that it have strong evident that it follows the normal distribution.



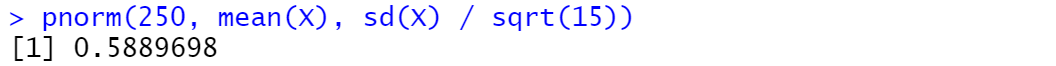
1. 1. Critical Value: From the data, we can get the mean and variance of it.

So we can set : and calculate the T distribution:

At 5% level of significance (), we can find the value of .

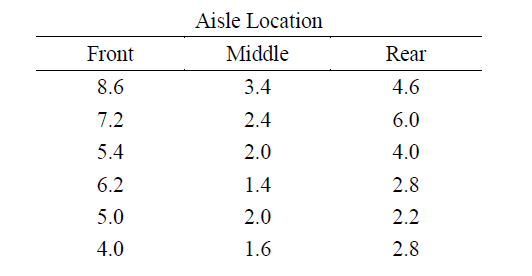
Since , we cannot reject the null hypothesis ().

2. we calculate the p-value by the function :



We can find the result is larger than 0.05 which means that we cannot reject the . We have no evidence to support that the average amount of bills is larger than 250.

4. The retail manager of a supermarket chain wants to determine whether product location has an effect on the sale of pet toys. Three different aisle locations are considered: front, middle, and rear. A random sample of 18 stores is selected with 6 stores randomly assigned to each aisle location. The size of the display area and price of the product are constant for all stores. At the end of a 1-month trial period, the sales volumes (in thousands of dollars) of the product in each store were reported as follows:



(a) At the 0.05 level of significance, is there evidence of a significant difference in mean sales among the various aisle locations? [20 marks]

1. What should the retail manager conclude? Fully describe the retail manager’s options with respect to aisle locations. [20 marks]

Ans:

1. We set the

Initially, we can set the type of location number be , and the number of sample in each group be . Then we can calculate the mean and variance of each group:

Front: , .

Middle: .

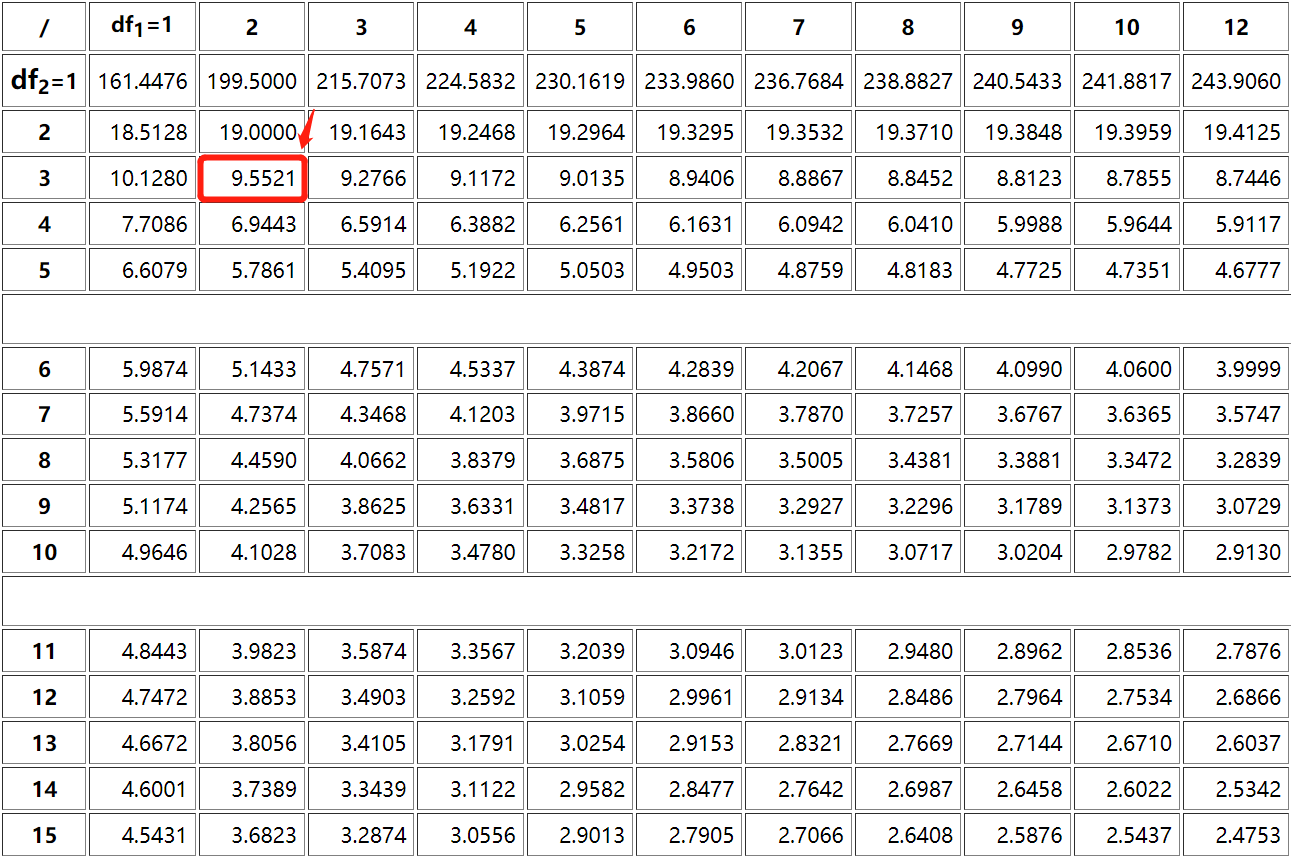
Rear: , .

And the mean of the groups: *.*

Then we can get the SSA, MSA and the SSW, MSW:

And then we can get the test:

The degree of freedom of MSA and MSW are and respectively, then we can check the :

**

which means that we can reject the and not all of the μ are same.

1. From the above analysis, we can infer that not all of the μ are same. And we can get the mean of sale for each group, it is clearly that the sale in the front location is the best, which means that we can put the most important thing to sell here such as products whose shelf life is nearing its end, products with short shelf life like foods as well as popular products. By contrast, the sale in the rear and the middle location are relatively lower, we can put some items that can be kept for a long time there.