假设检验

Demo

2022-11-26

# T-test

H0a: The average price of second-hand housing in the Central Region is not smaller than those in Outside Central Region.

H1a: The average price of second-hand housing in the Central Region is smaller than those in Outside Central Region.

data <- read.csv("h\_data.csv")

Filter out houses in Outside Central Region

CR <- subset(data,data$type != "OCR")

Filter out houses in Central Region

OCR <- subset(data,data$type == "OCR")

PCR=CR$price\_pre\_sqm  
POCR=OCR$price\_pre\_sqm  
x1=mean(PCR)  
x1

## [1] 6001.862

x2=mean(POCR)  
x2

## [1] 4187.315

s1=var(PCR)  
s1

## [1] 2209270

s2=var(POCR)  
s2

## [1] 609594.8

t=(x1-x2)/sqrt((s1/length(PCR))+(s2/length(POCR)))  
t

## [1] 107.3671

t\_base=qt(0.05,length(PCR)+length(POCR)-2)  
t\_base

## [1] -1.64489

Result: Because t>-t\_base=1.64489, we can not reject h0.

Therefore, we have 95% confident that the average price of second-hand housing in the Central Region is not smaller than those in Outside Central Region.

# Wilcoxon Rank Sum Test

Wilcoxon秩和检验，即”秩和检验”。 先将两样本看成是单一样本（混合样本）然后由小到大排列观察值统一编秩。如果原假设两个独立样本来自相同的总体为真，那么秩将大约均匀分布在两个样本中，即小的、中等的、大的秩值应该大约被均匀分在两个样本中。如果备选假设两个独立样本来自不相同的总体为真，那么其中一个样本将会有更多的小秩值，这样就会得到一个较小的秩和；另一个样本将会有更多的大秩值，因此就会得到一个较大的秩和。

A popular nonparametric test to compare outcomes between two independent groups is the Mann Whitney U test. The Mann Whitney U test, sometimes called the Mann Whitney Wilcoxon Test or the Wilcoxon Rank Sum Test, is used to test whether two samples are likely to derive from the same population (i.e., that the two populations have the same shape). Some investigators interpret this test as comparing the medians between the two populations. Recall that the parametric test compares the means (H0: μ1=μ2) between independent groups.

In contrast, the null and two-sided research hypotheses for the nonparametric test are stated as follows:

H0: The two populations are equal versus. (The average price of second-hand housing in the Central Region is equal to those in Outside Central Region.)

H1: The two populations are not equal. (The average price of second-hand housing in the Central Region is not equal to those in Outside Central Region.)

# help(wilcox.test)  
wilcox.test(price\_pre\_sqm~CR\_OCR,data=data)

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: price\_pre\_sqm by CR\_OCR  
## W = 244739953, p-value < 2.2e-16  
## alternative hypothesis: true location shift is not equal to 0

Since p-value is smaller than 0.01, we reject H0.

We have 99% confident that the average price of second-hand housing in the Central Region is different from those in Outside Central Region.

# 思路

1.检验是否满足基本假设？ （正态分布 etc）

是否要做个F检验，看看方差齐不齐

2.最基本的比较方法

3.升级方法？（解决样本规模不太一样的原因）