

IV. Significance test

IV-1. Statistical testing works

There are various statistical test methods depending on purpose and data type. It is impossible to explain all of them. For selection of statistical testing method, the author explains general flow of the selection. There two types of data. One is parametric data and the other is non-parametric data. This is ideological typology. All data include both quantitative information and qualitative information. There remains question what is parametric and what is non-parametric. In this text book, the author categorize data without homoscedasticity as non-parametric data even it has qualitative nature. In such cases, χ^2 test is applicable, when we can assume expectation value. F test and t test cannot be applied to qualitative data without transformation. We can consider application of t test and F test after proper transformation of data such as probit transformation or logit transformation. When the author was young, there were little computer analysis software applicable to non-parametric data, and various transformation techniques were often used. Recently there are various software applicable to non-parametric data and many researchers are using such software.

Requirements of student's t test and F test are normal distribution of parent population and homoscedasticity. Drawing frequency distribution is recommendable for the confirmation of normal distribution. Researchers can obtain various nature of the data by drawing frequency distribution. Habituation of drawing frequency distribution before statistical analysis is recommendable. In the case of limited number of data, we cannot draw frequency distribution. In those case, we need consideration of normality combining previous knowledge and character of the data. It is possible to transform the data to have normality by adequate functions in several cases. After that, we can consider application of t test or F test. Generally, t test is used for the test of average of 2 group. We can use F test for the comparison of 2 groups considering the freedom of numerator as 1, though t test is recommendable, because t test is easy to understand. In the case number of data population is more than 3, we use F test. We can consider by pick up two groups and do t test of all combination of the group. We can do this method mathematically, however, it is not recommendable. Because, when we repeat probability judgement, the probability decreases with repeat. In the case the null hypothesis is correct in the probability of 19/20, ($p=0.05$). When we repeat the judgment twice the probability of both two null hypotheses are collect is $\left(\frac{19}{20}\right)^2 = 0.9025$. Even solid betting ticket, we cannot expect continuous wins. In the case of 5 data group,

number of combinations of 2 group is ${}_{10}C_2 = 10_5C_2=10$

Probability of both 10 null hypotheses are collect is as follow

$$\left(\frac{19}{20}\right)^{10} = 0.598734$$

This means $p=0.401266$

The possibility of no error in total discussion is less than 60%. A possible solution is use probability level $p = 0.005$. Bonferroni correction is such method. However, when we increase probability level, we cannot get positive conclusion in case of small size data. F test is recommendable, significance of the ratio of variance among data populations and variance caused by randomness is tested in F test. T

Flow diagram of selection of method for significance analysis



