ISSN 2616-5783 Vol.5, Issue 8: 84-89, DOI: 10.25236/AJHSS.2022.050814

Graduation planning analysis using WILCOXON rank sum test

Shuling Deng[1,*], Jianxuan Qin[2]

[1]*Faculty of Mathematics and Statistics, Changsha University of Science and Technology, Changsha, Hunan, 410000*, China
[2]*Faculty of Mathematics and Computer Science, Guangdong Ocean University, Zhanjiang, Guangdong, 524088*, China
**Corresponding author*

Abstract: *In order to discuss the different factors that influence students' graduation planning, this paper use the Wilcoxon rank sum test to analyze the collected data. The rank sum test does not depend on the specific form of the overall distribution, and can be applied without regard to what kind of distribution the subject is ,and whether the distribution is known, thus making it more practical. Therefore, this paper mainly uses R language and SPSS software to program and analyze the data.Further,by comparing the analysis results of different methods, the results show that the most important factor influencing students to choose further education is the desire to have a decent job,Under this results,we provide reference for schools to offer related courses.*

Keywords: *Non-parametric statistics, Wilcoxon, Rank sum test, Difference, R language.*

1. Introduction

The Wilcoxon rank sum test is a nonparametric test. It doesn't depend on the specific form of the overall distribution and can be applied regardless of what distribution the subject is and whether the distribution is known, thus making it more practical. The Wilcoxon rank sum test has been applied in various aspects of life and production, such as the application of product quality control or debt rating models for industrial enterprises, and even in Meteorology, such as tropical cyclone intensity forecasting.

2. Data and Methods

2.1. Nonparametric test

In cases where the overall variance is unknown or little is known, the method of using sample data to make inferences about the overall distribution pattern, etc[1].is widely used in life, and its methodological choices for specific occasions are shown in Figure 1.

ISSN 2616-5783 Vol.5, Issue 8: 84-89, DOI: 10.25236/AJHSS.2022.050814

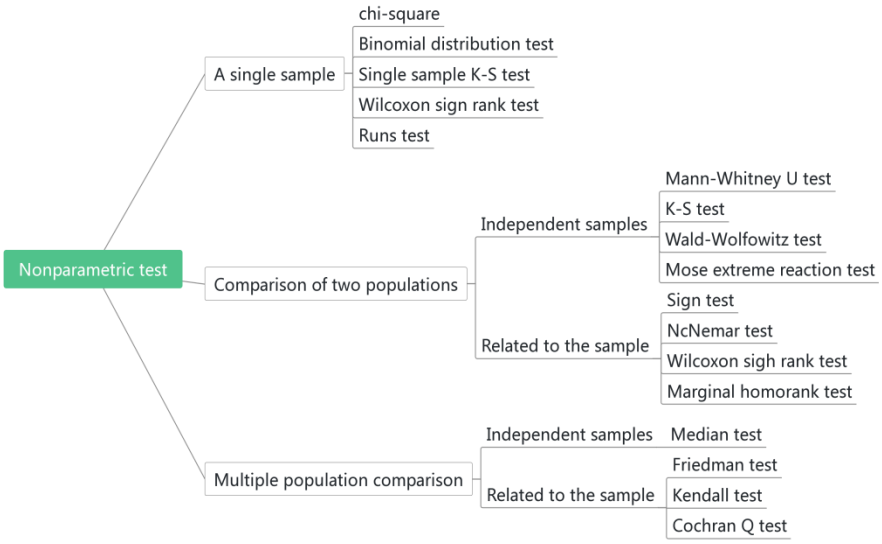


Figure 1: different types of Nonparametric test

2.2. Wilcoxon rank sum test

The Wilcoxon rank sum test is used to determine whether two independent samples are from the same or equal totals The basic idea is that if the test hypothesis holds, the rank sum of the two groups should not be too different The Wilcoxon rank sum test is based on the sample data rank sum[2]. The Wilcoxon signed rank test uses the X center of symmetry of the total as M the center of the distribution, and the X distribution of the total with $F(x)$ respect M to symmetry, M satisfying.

$$F(M-x)-1-F(x-M),\forall x\in R \tag{1}$$

The test statistic for the Wilcoxon signed rank test is shown in Equation (2).

$$W^{+}=\sum_{i=1}^{n}u_{i}R_{i} \tag{2}$$

Among them:

$$u_{i}=\begin{cases}1, & X_{i}-M_{0}>0\\ 0, & others\end{cases},\quad i=1,2,...,n \tag{3}$$

Where R_i is $|X_i|$ the rank in the $|X_1|,|X_2|,|X_3|...$ absolute value of the sample.

The specific steps are shown below.

1) Establish the original hypothesis: test $H_0: M=M_0$,Under the null hypothesis, the W^+ sum W^- should be about the same. Here, the test statistic is taken $W=\min(W^+,W^-)$

2) Calculate the difference $|X_i-M_0|$ and mark the sign, noting $sign(X_i-M_0)\begin{cases}+, X_i-M_0>0\\ 0, X_i-M_0=0\\ -, X_i-M_0<0\end{cases}$

3) Calculate the rank $|X_i-M_0|$ and calculate the positive W^+ and negative rank, respectively W^-.

ISSN 2616-5783 Vol.5, Issue 8: 84-89, DOI: 10.25236/AJHSS.2022.050814

$$\begin{cases} w^{+} = \sum_{i=1}^{n} R_i, sign(X_i - M_0) = "+" \\ w^{-} = \sum_{i=1}^{n} R_i, sign(X_i - M_0) = "-" \\ w^{0} = \sum_{i=1}^{n} R_i, sign(X_i - M_0) = "0" \end{cases} \tag{4}$$

4) Make $W = \min(W^{+}, W^{-})$ that the positive and W^{+} negative ranks W^{-} should be approximately the same when the original hypothesis holds, and if one of them is very small, the original hypothesis should be doubted.

5) Based on the obtained value of W, use statistical software or check the distribution table of Wilcoxon signed-rank test to obtain the p-value under the null hypothesis. If the p-value is less than or equal to the given significance level, the null hypothesis can be rejected.

2.3. Data Presentation

The data in the Table 1 shows that 80.77% of undergraduates choose to pursue further education after graduation. We can see why the choice has been made in Table 2. Therefore, we analyze and study the three main reasons for choosing further studies which shown in Table 2 and obtain the influence of the reasons that affecting students' choice to pursue further studies. Finally, we show the school's proposal for the school to offer relevant courses in Table 3.

Table 1: Main graduation planning of students from different colleges

major / planning	Mathematics	Computer and science	Electrical engineering
Examinations	158	255	197
Jobs	30	21	20
take a public examination	8	13	1

Table 2: Reasons for student planning

major / Reasons	Mathematics and statistics	Computer and Science	Electrical engineering
Academic Interest	21	42	23
Decent work	98	151	89
Stable work	75	84	92
Avoiding work	6	3	16

Table 3: Students' opinions on how to open relevant courses

major / advice	Mathematics and Statistics	Computer and Science	Electrical engineering
Career Guidance	88	99	91
College Guidance	62	162	53
Goal Planning	50	28	76

2.4. Description of symbols

The meaning of the symbols used in the text is shown in Table 4.

Table 4: Parameters adopted in this paper and their meanings

symbol	meaning
X_1	Strong academic interest
X_2	Pursuit decent job
X_3	Avoiding work
X_4	Pursuit a steady job

In this paper, the specific definitions of the above symbols are explained as follows:

1) Students who want to continue their research in that field or that direction and obtain certain

ISSN 2616-5783 Vol.5, Issue 8: 84-89, DOI: 10.25236/AJHSS.2022.050814

academic results, and whose research orientation is of a theoretical type rather than technological invention or entrepreneurial economy related.

2) Decent work refers to legal or unincorporated entities with specific institutions or business units, such as educational and medical institutions, that are independently accounted for the purpose of making profits.

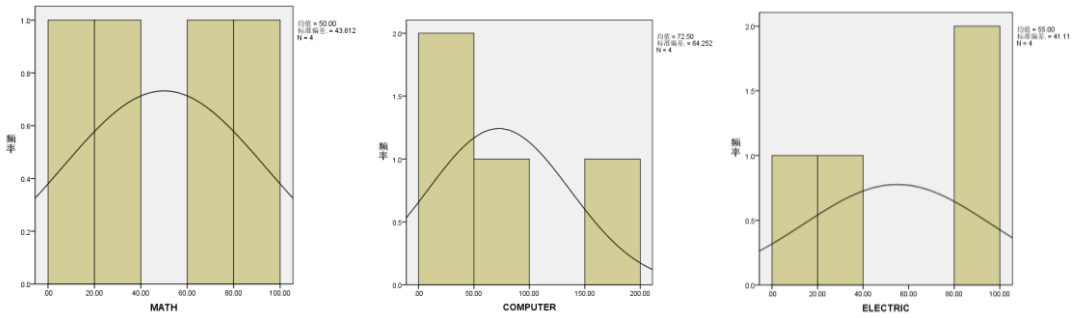
3) Work avoidance refers to a student's choice to pursue further education (in this paper, specifically, to obtain a graduate degree) to some extent in order to avoid working too early.

4) Stable jobs refer to jobs within the system or institutions, such as civil servants and other occupations that do not easily face unemployment or some public welfare units or non-public welfare functions with government functions and public welfare services as their main purpose.

3. results

3.1. Analysis of the problem using non-parametric tests

The frequency distribution of each influencing factor was plotted to test whether it conforms to the normal distribution, that is, whether it can be tested using a purely nonparametric test, and the results were obtained as shown in the Figure 1 and the SPSS software was used to analyze them, with 1, 2, 3, and 4 representing the different influencing factors, from which conclusions can be obtained, which are shwon in Figure 2(A)-(C)



(A) **(B)** **(C)**

Figure 2: (A)-(C)respectively indicates the frequency distribution of the School of Mathematics and Statistics,Computer Science and Electrical Engineering.

3.2. Modeling using Wilcoxon method

The ranking scores of the factors[3] that influence students' choice to pursue further education are shown in Table 5.

Table 5: Ranking score of each influencing factor

	A strong interest in academics	Want a decent job	Avoiding work	Stable work
1	123	72	210	175
2	170	182	219	88
3	187	194	187	116
Sort Score	480	448	616	379

Assuming that ranking is considered as a score situation, ranking is negatively correlated with higher ranking score, and higher total score indicates lower overall ranking and lower importance. According to the results of the above data, the combination of qualitative data and quantitative data analysis, the influence size can be divided into four major parts as shown in the Table 5, and it can be seen that wanting a stable job has the lowest score, which indicates that wanting a stable job has the greatest influence on most students' choice of further study, followed by wanting a decent job, while the scores of other factors, except for avoiding work, are relatively close, so the three factors were further analyzed by quantitative data analysis and tested using the Wilcoxon rank sum test.

First, the R software is used to draw the box line diagrams on X1 with X2; X2 with X3 as shown in Figure 3(A) and (B).

ISSN 2616-5783 Vol.5, Issue 8: 84-89, DOI: 10.25236/AJHSS.2022.050814

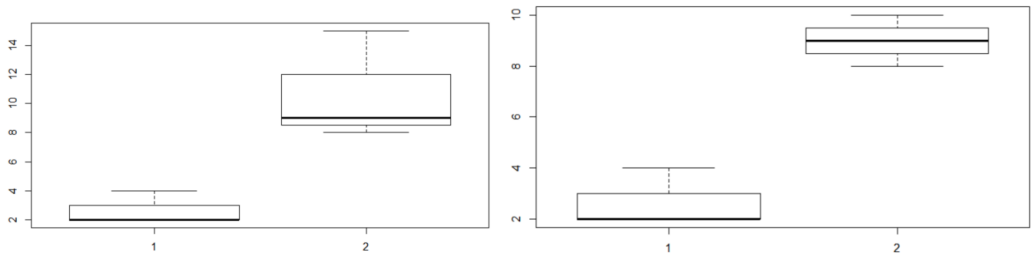


Figure 3: (A)-(B)respectively indicates denotes the box line plot of X1 and X2,denotes the box line plot of X2 and X3.

Wilcoxon rank sum test with a p value less than 0.01 rejected the original hypothesis , hypothesis X1 : have a great interest in academics, X2 : want to have a decent job, X3: want to escape from work. The Wilcoxon rank-sum test was used to compare the magnitude of the influence of these three factors on students' graduation planning[7] (R software).

3.3. Comparison of X1 with X2 and X2 with X3

Based on the above analysis, we can model and analyze the problem using the Wilcoxon rank sum test, as shown in the following steps in Figure 4:

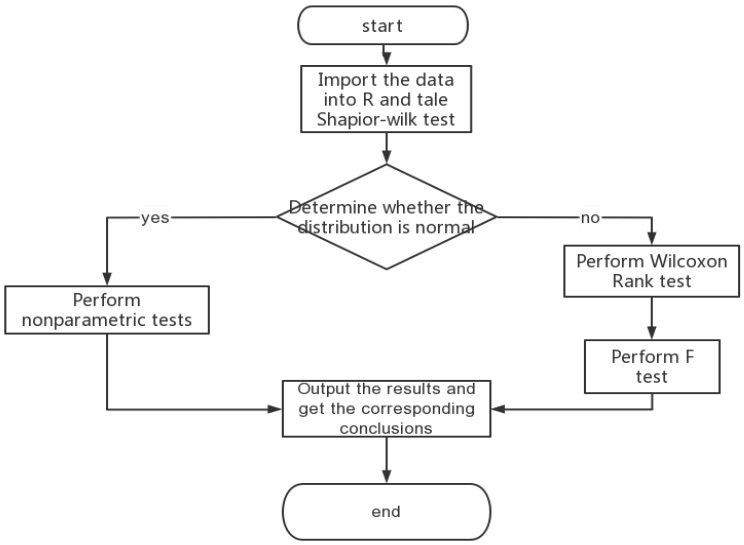


Figure 4: Flow chart of model building

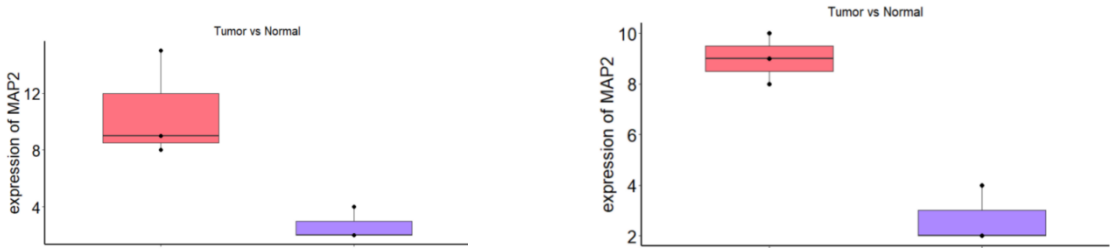
In the S-W test output, the first p set of values is less than the significance level of 0.05, indicating that the distribution of the two data sets is significantly different from the normal distribution. In the F-test output result median p value = 0.5552, which is greater than the significance level 0.05.Above all, we cannot use the Student-t test. Wilcoxon test needs to be used.

The following assumptions are made and paired box line diagrams are drawn based on the examples.

H_{00}: a greater impact X_2 than X_1 ; H_{10}: a greater impact X_3 than X_1

Based on the data from the survey results obtained in the basic assumptions, the following paired box line plots were obtained by programming the analysis in the R software.

ISSN 2616-5783 Vol.5, Issue 8: 84-89, DOI: 10.25236/AJHSS.2022.050814



(A) **(B)**

Figure 5: (A)-(B) respectively denotes the Paired box-plot of X1 and X2 and the Paired box-plot of X2 and X3.

In the Figure5(A), the red part is the academic box line diagram, and the blue part is the box line diagram of decent work impact;

In the Figure5(B), the red part is the decent work box line diagram, and the blue part is the box line diagram of the influence of stable work.

The output from the R programming language gives the following results:In the comparison X1 with X2 and X2 with X3 : the W value is 97775 and 164938 ,the p-value is 0.01032 and 0.0004818, obviously, the two p value is greater than the significance level α = 0.01 and we are to accept the original hypothesis H_{00} and H_{10}

4. Conclusion

When the problem was analyzed by non-parametric statistical methods, the data did not conform to the normal distribution and no conclusions could be drawn. However, the Wilcoxon rank sum test was used to obtain more accurate results and also proved that the Wilcoxon rank sum test was more applicable to the actual problem. The above analysis shows that of the three main factors influencing students' choice of continuing education, the greatest influence is wanting a decent job, and having a strong interest in academics is the least influential. Therefore, schools can use the results of the above analysis to provide career guidance classes or similar courses to help students have a better planing.

References

[1] U.M. Okeh; I. Onyeagu Sidney.Comparison of Two Diagnostic Test Procedures Using Modified Wilcoxon Signed Rank Test[J].Asian Journal of Mathematics & Statistics, 2020(1-13): 14-20.
[2]Cyrille Joutard.Asymptotic tail approximations for some nonparametric test statistics[J].Journal of Nonparametric Statistics, 2020(3-32),756-768.
[3] Masato Kitani;Hidetoshi Murakami,The limiting distribution of combining the t and Wilcoxon rank sum tests[J].Statistics, 2020(4-54).871-884.
[4] Xuan Li; Yunqiao Wu; Mengting Wei; Yiyun Guo; Zhenhua Yu; Haixian Wang; Zhanli Li; Hui Fan.A novel index of functional connectivity: phase lag based on Wilcoxon signed rank test[J].Cognitive Neurodynamics,2020(15):1-16.
[5] Mel éndez Rafael; Giraldo Ram ón; Leiva V ctor.Sign, Wilcoxon and Mann-Whitney Tests for Functional Data: An Approach Based on Random Projections[J].Mathematics, 2020(1-9):.44.
[6]Kai Xu; Yeqing Zhou.Maximum-type tests for high-dimensional regression coefficients using Wilcoxon scores[J].Journal of Statistical Planning and Inference,2021(211).221-240.
[7]Lin Tuo; Chen Tian; Liu Jinyuan; Tu Xin M.Extending the Mann-Whitney-Wilcoxon rank sum test to survey data for comparing mean ranks[J].Statistics in Medicine,2021(7-40).1705-1717.
[8]Hakan Polat; Song ül Karabatak.Effect of flipped classroom model on academic achievement, academic satisfaction and general belongingness[J].Learning Environments Research,2021.1-24.
[9]Fay Michael P; Brittain Erica H; Shih Joanna H; Follmann Dean A; Gabriel Erin E.Causal estimands and confidence intervals associated with Wilcoxon-Mann-Whitney tests in randomized experiments[J].Statistics in Medicine,2018(20-37).2923-2937.
[10]Malela Majika Jean Claude.New distribution-free memory-type control charts based on the Wilcoxon rank-sum statistic[J].Quality Technology and Quantitative Management,2021(2-18).135-155.