Summary of Model Construction on the Dynamic Detection Accuracy of Railway Lines

Junxi Wang[1,#], Chengyi Chen[2,#], Ruiya Qi[1], Yihui Tan[3]
[1]Information Engineering College, Minzu University of China, Beijing, China
[2]Science college, Minzu University of China, Beijing, China
[3]School of Philosophy and Religious Studies, Minzu University of China, Beijing, China
[#]These authors contributed equally

Abstract: This paper for the railway line dynamic detection data analysis, the integrated use of entropy weight method, Topsis evaluation, statistical theorem, RIS rank sum ratio model knowledge, build the quantitative analysis of the reliability of the instrument, the accuracy of the real-time detection data and evaluation interval algorithm model, better solve the problem of railway line dynamic detection data analysis. In view of problem 1, since the line equipment of the dynamic detector is affected by the geometric size of the track, according to the analysis of the relevant literature, the existing data is cleaned and divided into several sub-tables according to the "line-model-vehicle number". In the statistics, the average and median of vertical acceleration and horizontal acceleration are measured by the corresponding dynamic detector under different combination of "line number + model + vehicle number", and four fractions are introduced to reduce the influence of abnormal data. Then calculate the standard deviation of two types of acceleration, using the knowledge of the normal curve, the neighborhood (x, δ), statistics under the range of abnormal data frequency, combined with the entropy weight method and Topsis distance method to build a mathematical model, a comprehensive score, and finally the sensitivity analysis. Finally, 10 unreliable detectors and 3 data errors 1001-101-1230,100-101-101-1356,100102-1580, 2001-101-1259,2001-101-1215,2001-102-1502, 4001-105-1155. For problem 2, following the problem 1 system, the unreliable detector measurement data will not participate in the data analysis. After trying to make the greedy algorithm combined with linear fitting method and Bayesian formula, the Grubbs criterion was used to build the mathematical model, and then verify the 3-level alarm data after March 1,2024 by using the server to process the data after March 1,2024. In the detection process, the alarm level is above three levels for further analysis. Build a normal data judgment model and write two judgment codes with a test date after March 1,2024 and contain data with test level 3 or above. If either of the two judgment models is normal, the data is normal detection data; otherwise, it is abnormal data. Finally, 55 normal measurement data and 24 abnormal data were obtained. In view of problem 3, the data was processed first, and the algorithm was designed to screen and divide the data. The four lines were divided into the whole line according to the 1 km interval respectively, and the original data set was established. The rank sum ratio model of the whole rank is constructed, and after the variance test and multiple comparisons, the best interval and the worst interval of each line are obtained.

Keywords: Traversal; Algorithm; Topsis-entropy; Weight Method; Normal Curve Grubbs Criterion; RSR Rank Sum Ratio Model

1. Restatement of the Problem

Railway line equipment is the basic equipment of railway transportation industry. Under the influence of the outside world, the geometric size of the track is constantly changing, the subgrade and the road bed are constantly deformed, the rail, connecting parts and sleepers are constantly worn, and the technical state of the line equipment is constantly changing. In order to ensure the safety of traffic, the railway department must timely detect the state of the line, grasp the change law of the line equipment, and timely maintain and dispose the condition before the deterioration of the line equipment to affect the traffic safety. There are two main methods to detect railway lines, namely, static inspection and dynamic detection.

The detection data in this topic comes from a railway line dynamic detection equipment-line dynamic inspection instrument, which reflects the state of the line equipment by detecting the vertical (vertical) and horizontal (horizontal) acceleration of the train body during operation. When the geometric size of the line equipment deteriorates at a certain location, when the train runs through the location, the car body will produce a relatively large shaking, and the line dynamic detector can detect a large value. In addition, the interference of electrical equipment on trains, improper operation of drivers and sudden natural disasters may also cause the detector to produce large value, which is called abnormal large value. Therefore, each large value is required to be combined by the previous multiple detection value analysis.

The acceleration detection data of the dynamic detector is not only affected by the geometric size of the rail and the line equipment, but also affected by the installed train models and vehicle conditions. The alarm level of a detection data depends on the maximum alarm level in both the vertical and horizontal detection directions. In order to avoid the deterioration of the line, it is generally necessary to intervene the low-level alarm data in advance to avoid deterioration, and the alarm needs on-site maintenance. In general, if multiple large values are detected at 50 meters before and after the same mileage point within 3 days before the occurrence of a large value, or if multiple detection values at the same mileage point gradually increase with time, it can be judged that the field line of the large value warning has problems.

In addition, the working life of the dynamic detector is about 7 to 10 years. In the long-term operation process, with the aging of the device or the loosening of the fixed device, the reliability of the detection data will change, and the detection value will be systematically large or small, which will affect the accuracy of the detection data.

This paper intends to solve the following problems:

1) The design algorithm quantified the reliability of each detector and selects out the abnormal detector.

2) The dynamic monitor uploads a package of detection data to the server every 3 minutes or so. The server program needs to timely judge the reliability of the current latest detection data and judge whether the current data alarm level is correct. If the detection data is a large value above level 3 (including), it is further necessary to judge whether this large value is the normal detection data caused by line variation or the abnormal data caused by equipment failure or interference. Please design the algorithm to process the data in real time according to the existing data server, and check the accuracy of the algorithm. Provide the judgment results on the level 3 alarm data appearing in the detection data after March 1,2024.

3) On a line, the line condition of each section varies greatly, and usually a large acceleration will be detected in the poor interval of the line. The measurement result of the acceleration value can comprehensively reflect the overall condition of the different line sections. According to the test data in the attachment, please divide the whole line at 1 km (such as 100.000-100.999), evaluate the quality of the line by each km, and give the 10 best quality and 10 worst sections of the line.

2. The assumptions of the Model

Ignoring human factors such as improper driver operation and the impact of electrical equipment interference on detection.

Ignoring the differences in vibration characteristics between individual trains.

Ignoring the effect of four months on the monitor in the attached data.

It is assumed that the different combinations of "line number + model + train number" in the attached data correspond to the unique dynamic detector respectively.

3. Symbol Description

The first symbol is "avc", which represents "normal acceleration"; the second symbol is "a hc", which represents "horizontal acceleration"; the third symbol is "σ", which represents "standard deviation".

4. Problem analysis and Model Building

4.1 Problem 1 Analysis and Model Establishment

4.1.1 Analysis of Problem 1

The is proposed to design an algorithm to quantify the reliability of each detector and screen out the abnormal detector. According to the attached data, the driving data sheet of each dynamic detector can be obtained.

According to the topic information, the acceleration detection value of the detector is mainly affected by the interference of electrical equipment, the improper operation of the driver, the working life of the detector (individual differences), vehicle condition, speed, geometric size of the rail and line equipment. The function of the dynamic detector is to dynamically detect the railway lines, and no one in the data given in this question is the relevant part, so the hypothesis is made without considering the human factors such as the improper driver operation and the uncertain factors such as the interference of electrical equipment. Among the remaining four main factors, reduce the influence of vehicle condition on the analysis results by grouping "line number + model + vehicle number"; minimize the effect of individual differences on the analysis results, calculate the acceleration statistics, compare the standard deviation with the overall acceleration statistics and take the interval as the detection value to eliminate the influence of the speed. The main factors affecting the data after processing are the geometric size of the rail and the line equipment, and then the entropy weight method and Topsis quantify the reliability score of each dynamic detector, and select the abnormal dynamic detector.

4.1.2 Model establishment and solution of Problem 1

(1) Determination of data of each dynamic detector.

There are 1001,2001,2001,3001 and 4001. After observation, only the line number 3001 is up, and the other three are single lines, which are control variables to make the model more accurate, so the line numbers 1001,2001 and are selected

All dynamic data in 4001. Assuming that the combination of different "line number + model + train number" in the attached data corresponds to the unique dynamic detector, write the program with Matlab (see appendix A for the source code), and obtain the form of the data of each dynamic detector (2) Determine the influence of vehicle speed factors

Through the literature [1] The division of the speed interval is determined:

[40, 60], (60, 70], (70, 80], (80, 100], (100, 120], (120, 140], (140, 160].

After averaging the stratified vertical acceleration, the whole value is averaged. By comparison, the acceleration value close to the true value is obtained. The standard deviation of the average of the vertical acceleration and the whole is <0.01. To sum up, the influence of the speed on the dynamic detector detection can be negligible under the accuracy of practical application.

(3) The design algorithm quantified the reliability of each dynamic detector

A. manipulation data

In statistics, means are usually used to measure the average level of the data, while quartiles are used to understand the distribution and dispersion of the data, and to understand the degree of variation of the data. Therefore, the vertical acceleration and horizontal acceleration of each dynamic detector are taken as mean and quartiles respectively with the help of Matlab. The range of acceleration is set as (ahc σ, ahc + σ) and (avc σ, avc + σ) to calculate the proportion of the outgoing data of each instrument to eliminate the influence of the relative value on the quantitative evaluation.

B. The entropy weight

Entropy weight method is a kind of multi-attribute decision analysis method, through the use of entropy weight method to determine the weight, the weight of each index is based on the statistical characteristics of the data itself, can ensure that the next comprehensive evaluation more objective and scientific, this question is used to determine the vertical acceleration beyond the proportion, the horizontal acceleration and three quartile weight, in order to consider the importance of each index in the decision-making process. The calculation results of information entropy are used to determine the weight of each index, so as to make the weight allocation more objective and scientific.

Comprehensive evaluation of c.Topsis

Topsis The method can comprehensively consider the performance of multiple indicators, and rank the results of each indicator.

D. partial results output

According to the comprehensive score, the abnormal detector and part of the detector are selected as follows (the closer to 0, the better the reliability). A total of 10 abnormal dynamic detectors were selected, two of which were five-digit car numbers. According to the official data, the data of five car numbers were wrong, and they can be screened out. The rest is a negative number, which is verified in the schedule of sensitivity analysis and is also incorrect data. So finally, there are only 7 cases.

(4) Sensitivity analysis Using the algorithm in Topsis comprehensive evaluation, change the data, get the change of comprehensive score, and compare with the original comprehensive score line chart. Due to the large amount of data, we selected 1001-101 groups from the Topsi s comprehensive evaluation for analysis. According to the comprehensive consideration of multiple line charts, this model has high sensitivity.

4.2 Problem 2 Analysis and Model Building

4.2.1 Analysis of Problem 2

In view of problem 2, using the algorithm of problem 1, we can find out the abnormal dynamic detector, whose measurement data does not have reference significance, and do not participate in the analysis of the real-time processing data.

For the first small question (see question 2 flowchart 1), design the server real-time processing algorithm to determine the reliability of the latest data and whether the alarm level is correct. First, the frequency of the same vertical addition and water addition being judged as different levels was counted. Due to the large sample size, the frequency can be expressed as a probability. Several groups of data with the same number are randomly selected, multiplying the vertical and water added by the corresponding probability of different levels respectively, and the corresponding level of acceleration is determined by the Bayesian formula. Also try to fit the function to estimate the threshold, and directly judge the alarm level of acceleration. The above two tests performed well in the sections suitable for a large amount of data, but considering the practical application, the Grubbs criteria were used to screen outliers.

For the second small question (see question 2), the large value of level 3 or above should be further analyzed. According to the title information " Generally, if a large value is detected at 50 meters before and after the same mileage point, or multiple detection values at the same mileage point gradually increase with time, it can be judged that the large value warning field line has problems."After the algorithm of the previous question, it is concluded that the normal detection data caused by line deterioration or the abnormal data caused by equipment failure or interference.

4.2.2 Model building and solution of Problem 2

(1) Design the algorithm for the server processing data in real time

a. bayesian inference

$$P(A|B)=P(B|A)*P(A)/P(B) \quad (1)$$

among:

P (A | B) is the probability of an event (A) after observing (B), which is called the posterior probability. P (B | A) is the matter

The probability of observing (B) of piece (A). P (A), is the prior probability of an event (A) before observing (B). P (B) is the probability of observing (B).

Based on Bayes' theorem, we first count the frequency of each acceleration (vertical acceleration and horizontal acceleration respectively) in each line. Because this stratified sample size is large, the probability can be estimated by frequency, and the full probability formula can be applied. The specific principle is as follows:

First, Matlab is used to find out the frequency of each vertical acceleration value and each horizontal acceleration value in each line, respectively, in two Excel files.

Therefore, with the target determination data, there are three data sets:

The vertical acceleration and horizontal acceleration value and alert level for the current group data are extracted. Write to a, b, And c. retrieving all rows with the same a value of the current group data in the vertical acceleration numerical probability set, Calculate the sum of the product of probabilities in these rows, Obobtain sum 1. All rows with the same b value as the current group data, Calculate the sum of the products of the second column in these rows, sum 2. Calculate the total grade grade. judge the data reliability, If the absolute difference between the c value of the current group data and the total grade grade is less than 0.5, Then the reliability reliability is 1. Otherwise, 0 is false.

$$sum1=(data1(i,2)),data1(i,1)==a \quad (2)$$

$$sum2=(data2(i,2)),data2(i,1)==b \quad (3)$$

$$grade=(sum1+sum2)/200 \quad (4)$$

If the distance between the grade and the alarm level data is less than 0.5 levels, the data can be considered reliable.

After importing all historical data and checking the accuracy, it is found that the dynamic detector with large data volume is in good condition with great limitations.

And b. Fitting function method

The plane rectangular coordinate system is established according to the known upper and lower bounds of grade 1. The x-axis of the coordinate system is marked, which has no practical significance and is only used to distinguish different y values. The y value of the coordinate system is the difference between the lower bound of the separated classes, and let m be the difference between the lower bound of rank 1 and 0, where the upper bound of each rank is equal to the lower bound of the latter rank.

According to the difference m between the lower bound of grade 1 and 0 of the threshold model based on the Bayesian formula, and the difference n between the lower bound of grade 2 and the lower bound of grade 1, which is expressed in the coordinate system as points (1, m) and (2, n).

For the lower bound of rank 3, combining the greedy algorithm and linear fitting, let an be the first term 0, the tolerance 0.1, and the upper bound 1. Calculate (1, m), (2, n), (3, ai) the slope of the three-point fit and the line respectively, and find an ak in an, so that the slope of the fitted line is closest to 0, that is, the difference between the lower boundary of grade 3 and the lower bound of grade 2. From the previous threshold model obtained based on the Bayesian formula, the lower bound of grade 1 is h, and the lower bound of grade 3 is (h + n + ak).

For the lower bound of grade 4, step 2 is repeated, that is, the arithmetic column with bn 0, tolerance 0.1 and upper bound 1.

Calculate (1, m), (2, n, n), (3, ak), (4, bi) and the slope of the line, and find a term b k in b n, so that the slope of the fitted line is closest to 0, which is the difference between the lower boundary of grade 4 and the lower bound of grade 3. And so on, the lower bound value of level 4 is (h + n + ak + bk).

After importing all historical data and checking the accuracy, it is found that the dynamic detector with large data volume is in good condition with great limitations.

The c. Grubbs guidelines

To improve the model accuracy, we reviewed the literature [2], With a small processing sample size, the outliers can be removed using the Grubbs criterion.

The solution steps are performed as follows:

Step 1: Calculate the mean of acceleration of each line $\bar{x}$ And standard deviation s: Mean calculation formula:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} Xi \quad (5)$$

Standard deviation calculation formula:

$$S = \sqrt{\frac{1}{n-1}}\sqrt{\sum i = 1^n (Xi - \overline{X})^2} \quad (6)$$

Step 2: Calculate the deviation between each data point from the mean, and calculate the Grubbs value G:

$$G = \frac{|Xi - \overline{X}|}{S} \quad (7)$$

Step 3: Determine the critical value. Formula formula of critical value:

$$\text{G critical} = \frac{k * s}{\sqrt{n}} \quad (8)$$

Here is reduced to 3 times the standard deviation range. Step 4: Compare the Grubbs value G to the critical value G critical: If G <G critical, the data point xi is considered an outlier.

(2) The establishment and solution of the inference model 1

A. According to the analysis of the problem, judge whether the large value is the normal monitoring data caused by the line variation or the abnormal data caused by equipment failure or interference. Through the design algorithm, judge whether there are multiple detection values at the same mileage point gradually increasing with time, and judge the situation of the target data. (According to the analysis in the process of problem 1 model establishment, it can be concluded that the effect of speed on vertical and water addition is small, so that the effect of speed as the detection value is ignored, and the detection value is divided into vertical and water addition).

b. situation analysis

There are three scenarios for any level 3 alert data that appears after March 1,2024:

Situation 1: This mileage point has only experienced one measurement of the dynamic detector, and there is no more data at the same mileage point. At this time, method 2 cannot judge whether this group of data is normal or abnormal, then this group of data is determined to be impossible to judge.

Situation 2: This mileage point has undergone multiple measurements by the dynamic detector, and the multiple detection values taken at this mileage point do not gradually increase with time increase, so this set of data is determined to be abnormal data caused by equipment fault or interference.

Situation 3: This mileage point has undergone multiple measurements by the dynamic detector, and the multiple detection values taken by this mileage point have gradually increased with the increase of time, so this group of data is determined to be the normal detection data caused by line deterioration.

C. Establish a linear regression equation for the detection value of the same mileage point

According to the question, the linear regression equation for multiple detected values of the same mileage point can be set as: $y = mx + b$

Where the dependent variable y is regarded as the corresponding vertical and water addition, the independent variable x is regarded as the mileage point, m is the slope and b is the intercept.

And d. Model solution

Step 1: Unique mileage cycle through: obtain the unique mileage and process it one by one to ensure that the corresponding data of each mileage is correctly judged.

Same mileage data extraction: in each unique mileage cycle, the same mileage is extracted by logical mileage index

All of the data.

Data screening: When processing data with the same mileage, exclude reliable data with the same vehicle type, speed, vehicle number and line number as the current data, so as to ensure the difference of data.

Step 2:

State logic judgment: trathrough all the data with the same mileage

Sort and difference processing: find the vertical and water addition of different data of the same mileage, and calculate the difference between adjacent data to ensure the consistency of the data.

Status judgment: if the difference of vertical and water are greater than 0, it means that both vertical and water gradually increase with time, that is, this set of data is the normal detection data caused by the variation of the line.

(3) Judge the establishment and solution of model 2

A. According to the occurrence of the corresponding large value of a level 3 alarm level, multiple different detectors detect multiple large values within 50 meters before and after the same mileage point, a judgment model is constructed to judge that the value is the normal detection data caused by the line variation.

B. Build one of the judgment models according to the known information, and detect multiple large values at 50 meters before and after the same mileage point within 3 days before the corresponding large value of a 3-level alarm level, which is a criterion to judge that the value is the normal detection data caused by line variation.

The c. solving steps are as follows:

Step 1: Set the parameter m to quantify the number of large values detected by 50 meters with different dynamic meters within 3 days. m 1 is the number measured in the first 50 meters of the mileage point, and m 2 is the number measured at the last 50 meters. Initial values for m, m 1, and m 2 are mean 0. M, m equals the sum of m 1 and m 2.

Step 2: judge the alarm level . If it is less than level 3, then judge the next line. If it is greater than or equal to three levels, remember this behavior in line i, and record the corresponding mileage of the line in line ai.

Step 3: Within the range of (ai- -0.05, ai), if the mileage with q before the bank falls in the range, further judge whether the interval between the measurement date of line (i- -p) and the measurement date of line i is less than or equal to 3 days. If this happens, m 1 = q, if not, judge the measurement date of line (i- -p- -1) and line i until (i- -p- -k) meets all the above conditions, remember m 1 = i- -p- -k, (k <= p).

Step 4: Repeat steps within the range of (ai, ai + 0.05). Finally, m 1, m 2, m= m 1 + m 2, that is, the number of 50 meters before and after the same mileage point in 3 days is m. Repeat the above judgment until all the rows in the form are completed.

And d. Analysis of the advantages and disadvantages:

The model has high objectivity and scientificity. All calculations in this model are closely related to the original data

Sex; accurate positive and abnormal discrimination range, which makes the results accurate and credible. The model is computationally efficient. Making it computational efficient and able to generate results in a short time. The disadvantage is that the judgment mileage range is not detailed. Suppose that the miles a of line i is the center of the range, and the miles of both adjacent rows are not in the range (a- -0.05, a + 0.05). If the difference between the mileage of rows i-1 or i +

1 and the mileage of line i is only 0.06, whether the data has reference significance or can be included into the reference range under sum conditions, so as to reduce the miscalculation of equipment failure caused by not satisfying the model one and increase the judgment accuracy. And the lack of detailed judgment criteria. The model only positions the number of measurements equal to 0 as abnormal data caused by equipment fault or interference. If, under certain conditions, the measured value of 1 or 2 can be positioned as the abnormal data caused by equipment failure or interference, and the progressive judgment standard, the misjudgment of equipment failure by model 1 can be reduced, and the judgment accuracy can be increased.

4.3 Problem 3 Analysis and model building

4.3.1 Analysis of Problem 3

According to the appendix, according to the test data related to acceleration, the whole line is divided into an interval of 1 km (such as 100.000-100.999). The quality of the line is evaluated according to each kilometer, and the 10 intervals with the best quality and the 10 worst intervals of each line are given. First of all, it is necessary to divide the lines first. According to the observation data, each line is 100 kilometers long, which can be divided into 100 sections. Therefore, all intervals can be divided into ten layers by stratification, and the head and tail two layers are in line with the meaning of the question. Secondly, we can first consider one line and the analogy of the other lines. This paper follows with the 1001 line as an example. Thirdly, the first two questions can see the obvious normal distribution characteristics through the data statistics, but here we are not obvious, so we should choose not mandatory to meet the normal distribution hypothesis [3] The evaluation method, considered together, we chose the RSR (Rank Sum Ratio) rank-sum ratio method, a non-parametric statistical method, to compare the location parameters of two independent samples. Finally, according to the question, the index is determined as vertical acceleration, horizontal acceleration and alarm level, and because the large acceleration will be detected in the interval with poor line, so it is a low excellent index, and each interval is taken as the sample object, and the weight is determined by Topsis method [4].

4.3.2 Model building and solution of Problem 3

(1) Model establishment

a. data handling. The program will divide the interval and average the vertical acceleration, horizontal acceleration and alarm level as the raw data. Then build the raw data matrix. Each line is 100 kilometers long, so there are 100 objects in the interval, and the size of the original data matrix is 100 * 3.

B. The whole rank transformation was performed to eliminate the dimensional effect. The combined sample data are sorted by size and replaced with the original data by rank order, where the minimum observation is recorded as 1, the second small as 2, and so on, until the maximum is 100. Then find the rank sum, calculate the rank sum of each sample object separately, and add the ranks belonging to the respective sample.

C. Calculate the good and bad index of the sample object, that is, the ratio of the rank sum (RSR value).

D. Selected boundary value. First, the rank sum ratio distribution is determined, and the frequency number is arranged from small to large, and the cumulative frequency number and average rank of each group are calculated. Recalculate the regression equation. Finally, the probability unit boundary value of ten grades is selected. The mileage range of the output head and end.

E. variance test. Find the critical value corresponding to the RSR value based on the significance level α (usually 0.05 or 0.01) and the degree of freedom F (usually the sample size of the small sample 1). Based on the calculated RSR value. The principle of variance test is based on the comparison of the two samples by testing whether the ratio significantly deviates from the expectation.

(2) The optimized full rank transformation of the model eliminates the influence of dimension, but also loses the relative size of the values in the original data, and only retains the size relationship, resulting in information difference. After observation, the numerical difference between different data of this model is large, so the rank sum ratio method is used to determine the weight of the three indicators by Topsi s objective empowerment, which makes the quantitative linear relationship between the compiled rank and the original data, and reduces the systematic error.

5. Brief Summary

This paper focuses on the actual statistical data, and analyzes the actual data with the theoretical guidance to solve the problem, so as to avoid the error caused by subjective factors, and the results are more objective and true. Considering the normal distribution of data, and solving data with different models with different sample sizes, turning multiple targets into single targets through simple step-by-step classification idea, and completing complex judgment with simple models.

As for the shortcomings, the human factors such as improper operation, actual organization and coordination of the driver are ignored in the hypothesis. Meanwhile, the time span of the data in the attachment is four months, and the possibility of replacing the dynamic detector in part of the bullet train and the statistical form is wrong, so the cleaning of historical data still has a large room for improvement in the process.

References

[1] Liu Jianshe, Zhao Junpeng and Hou Shuang. Analysis of vertical Vibration Characteristics of High-speed Metro Train. Seismic engineering and engineering vibration 43, period 4 (2023): 24550.

[2] Zhu Minren, Grabs Criteria and Its Curve Fitting, Physics and Engineering, Period Z 2 (1989): 38.

[3] Tao Zhuang. Detailed Interpretation of Classical Rank and Comparison Method. Journal of Mathematical Medicine, Issue 2 (2007): 12225.

[4] Fan Zhao. Comprehensive evaluation of TOPSIS method and rank sum ratio method. Statistics of Chinese hospitals, period 4 (2000): 21416.