Reliability and Efficiency of Data Processing in Agricultural Logistics Matching

Deng-bing ZHANG
School of Economics, Yancheng Institute of Technology, Jiansu Yancheng 224051, China

Keywords: Data processing, Subjective orientation, Validity, Information.

Abstract. In the processing of Logistics matching information of agricultural products, there is a serious subjective tendency affecting the reliability of matching. In order to generate useful information for matching decisions, it is necessary to adopt appropriate and effective data processing method. Data processing, information access, effectiveness of logistics of agricultural products are different stages which are closely related and distinguished with each other. Data value is determined by the amount of information that gained by decision makers, and the value of information is determined by its utility. The basic methods of data processing is data deformation, deformation is not only a improving process of data validation, to a certain extent that it deviates from the original objective data, thereby reducing the reliability of the data. Data processing of agricultural product logistics should response to objective facts as much as possible, also take into account the subjective orientation. We cannot emphasis on subjectivity too much and set aside specific environments. Reliability and availability of data processing must be considered to ensures validity of the logistics match.

Introduction

Inefficient logistics matching is the key factor to cause agricultural product logistics problems, and inefficient logistics matching is essentially caused by the lack of reliability of information processing. Information generated by data processing is an important basis for decision-making. In order to generate useful information for decision making, it is necessary to adopt appropriate and effective data processing method. To this end, a wide variety of data processing methods came into being, become an important means of effective decision making. However, data processing is also a double-edged sword, improper use will bring negative effects for information generating, which prevents effective decision making. From data processing based on the subjective orientation, through the analysis of data, this article studies data reliability and validity in the logistics matching process.

Objectivity of the Data

Data is a symbol of objective facts, including numbers, characters, images, and so on.

Objective Data, Non-objective Data

According to the consistency of data and facts, data can be divided into objective data, non-objective data. Objective data refers to the indicators related to specific research statistics based on data related indicators, and it accurately reflects the research objects. For example, in the finite population, data of production of agricultural products are objective data. But in many cases, the data acquisition and processing are under the influence of the subjective, and the data cannot reflect accurately the properties of study object, so that the data is called non-objective data.

Raw Data, Deformation Data

According to the stage, the data can be classified as raw data, deformation data. Raw data is the data obtained from objective phenomenon in certain ways. When the original data cannot provide a basis for policy makers, data can then be "deformed" to the deformed data. Deformation processing, there
are two basic ideas: objective, subjective deformation. When deformation is not subject to the subjective intervention, we call it objective deformation, such as the range, mean and variance of the data set. Whether it is subjective or objective deformation, deformation data will deviate from the original objective data to some extent. Therefore, objective deformation is relative, subjective deformation is absolute, any deformation is carried out in a person's subjective opinion. Some scholars believe that, completely objective fact is difficult to access, and sometimes processed data are much truer than objective data after processing. This refers to, when the original data is lack of credibility, data after processing may be closer to the objective fact. Raw data discussed in this article are credible, but there are some "insufficient information" problems.

**Data Processing, Information Retrieval and Utility of Agricultural Product Logistics**

**Information Is People's Perception and Understanding of Phenomena or Data**

Some scholars point out that, when symbols are used to indicate something, it becomes information. Information is strictly different from data, but the information is always based on a specific representation of the data exists. Information access is highly subjective, and has strong decision-making relevance in use.

**The Basic Principles and Methods of Data Processing.** With the development of statistical techniques, various kinds of new statistical indicators, methodologies and tools emerge, statistical method reform is faced by challenges. Data processing methods should be scientific, practical and operable. In my opinion, the basic principles of data processing are to provide users with more information. Data collection, classification, analysis and data deformation are designed to allow users to get more information. In order to make the decision-making more scientific and effective, decision makers need to process data for the object, to make it easier for decision makers to obtain the required information. Data processing includes data collection, classification, distortion, transmission, analysis. Data can make the data more understandable to the managers, to help users get more information. For example, in order to obtain information of the agricultural products in a certain area, the investigation method is used to obtain data. But information provided to managers is based on individual, isolated and scattered, not easy for managers to access to effective information. Therefore, some technique methods are used for the original data processing, such as histograms, variance, mean, and so on. In this way, information existing in the original data which is hard for managers to obtain, will emerge, and will assist effective decision-making. Data deformation is the basic methods of data processing, statistical analysis is the process of data deformation. Data deformation is the process of data from the low level to the advanced level, and is the process of data validation improved. Data deformation fundamental tasks is to make decision makers better grasp the hidden information in data for scientific decision making. To some extent, data processing reduces the reliability of the data. As data is bound to lose some of the original data or produce some new data, which will result in missing of the original objective information or some departure from the original objective of the information. However, compared to the original data, deformation data is easier to use in decision-making, and improves decision-making effectiveness.

**Data Processing, Information Acquisition and Utility Producing.** In the Logistics matching process, there is transformation of data-information-utility, which includes data processing, information acquisition, and utility producing, which are different stages and is closely associated with each other. Data processing is the base, information obtaining is the means, and utility is the goal. Data value is determined by the amount of information got by decision makers, and the value of information is determined by its utility. Relationships between them as shown in Figure 1.
Subjective Orientation of Agricultural Product Logistics Data Processing

The data sources of agricultural products logistics is wide with high updating speed and low consistency, the original data must be processed. There are two reasons for subjective process: one is that raw data is insufficient to directly provide the scientific information required for decision-making, the second is that there is a certain gap between the original data and decision makers expectation on the new data distribution. These two reasons are essentially closely related to data users’ subjective, that is, users subjectively consider that only after a certain processing on raw data can it provide the required information or the correct information. For example, through a survey of quality of agricultural products, may get a set of data X that "is relatively concentrated, discrimination is not obvious". Based on the purpose of "improved discrimination", how to process data to get deformation data with high distinction? A number of measures can be taken, but all of these methods are clearly subjective orientation.

Min-max Normalization Method

To enlarge data gaps and reflect the difference between raw data, the most appropriate method of minimum-maximum normalization method.

The subjective orientation of min-max using a standardized method is: "the raw data is not reasonable, there is a clear expectation".

The original data set X Interval is \([x_{\min}, x_{\max}]\), Standardized expectations after the interval is\([x'_{\min}, x'_{\max}]\), The data Standardization of data x as follows:

\[
x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} (x'_{\max} - x'_{\min}) + x'_{\min}
\]  
(1)

The Isometric Ranking Methodology

Subjective orientation of using isometric ranking method is: "the value difference of the two neighbors data is equivalent, and contribution ability is related only to the ranking, regardless of specific values with the raw data; there is a clear expectation". If you want to open the gap between data, you can determine the normalization X interval is \([x_{\min}, x_{\max}]\), all data decreases according to certain values

\[
d = \frac{Max - Min}{M - 1}
\]  
(2)

The standard value for the nth data is

\[
x_n = x_{\min} + (n - 1)d
\]  
(3)
Decreasing Normalization with a Given End Boundary

Decision-maker may think: the top data should give greater differentiation, and there is an expected interval of maximum value. We can set a maximum score Max (e.g. 90), to a descending sequence 5, 4, 3, 2, 1, the ranking score is 90, 85, 81, 78, 76, 75.

The Decline Normalization With Given Boundaries on Both Ends

Decision-maker thinks: top data should give greater differentiation, have clear expectations. We set standard interval \([x_{\text{min}}, x_{\text{max}}]\), the descending sequence is \(B_n\), followed by data whose intervals will become more and more dense. The standard value for the \(n\)th data is:

\[
x'_n = x'_{n-1} - b_{n-1} = x_{n-1} - [b_1 - (n - 2)d]. \quad (n \geq 2)
\]

Of which:

\[
b_n = b_1 - (n - 1)d
\]

\[
a_M (\text{Min}) \quad a_5 \quad a_4 \quad a_3 \quad a_2 \quad a_1 (\text{Max})
\]

\[
b_{M-1} \quad \ldots \quad b_5 \quad b_4 \quad b_3 \quad b_2 \quad b_1
\]

Because,

\[
a_M - a_1 = b_1 + b_2 + \ldots + b_{M-1} = \frac{(M - 1)(b_1 + b_{M-1})}{2}
\]

so,

\[
\frac{(M - 1)[2b_1 - (M - 2)d]}{2} = \text{Max} - \text{Min}
\]

With a given \(b_1\), \(d\) can be calculated.

(Here \(b_1\) should be a value greater than \(\frac{\text{Max} - \text{Min}}{M - 1}\), in order to ensure \(b_1\) decreasing)

Normal Distribution Method

If you think the data should be in line with normal distribution, you can set five sections according to the ranking (such as: excellent, good, middle, pass, poor). Here we give each parts with proportion 2.5%, 22.5%, 50%, 22.5%, 2.5%.

<table>
<thead>
<tr>
<th>number of All data(M)</th>
<th>The proportion</th>
<th>Scope of standardization</th>
<th>Method of Standardization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>2.5%</td>
<td>-2.5 ~ 1.96</td>
<td>Given limits at both ends &quot;decreases&quot; normalization.</td>
</tr>
<tr>
<td>Good</td>
<td>20%</td>
<td>-1.96+2/M ~ -0.67</td>
<td></td>
</tr>
<tr>
<td>Middle</td>
<td>50%</td>
<td>-0.6744/M ~ 0.67</td>
<td></td>
</tr>
<tr>
<td>Pass</td>
<td>20%</td>
<td>0.6744/M ~ 1.96</td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>2.5%</td>
<td>1.96+2/M ~ 2.5</td>
<td></td>
</tr>
</tbody>
</table>

Using these standardized data \(b_n\) as independent variable substituted into the Normal equation:
Deformation data based on normal distribution can be obtained. Depending on the circumstances, $\sigma$, $\mu$ can be flexibly selected to get different results.

**The Subjective Orientation in Integrated Data Processing**

Single indicator evaluation tends to have higher objectivity, comparability and impartiality. But too many indicators are not beneficial to scientific judgment to the evaluated object, and indicators of the same object there are more differences.

So, the comprehensive indicator evaluation method must be used. To form integrated indicators from different dimensions, there are two general approaches.

One is the additive model, dimensionless processing is taken for different indicators (such as converting scores), and then integrated by weight. For example, on the evaluation of agricultural product logistics level, additive model can be adopt, considering the different weights of logistics time, logistics costs, logistics loss, and other factors.

Another is a multiplicative model, some indicators are multiplicative, which has a clear multiple meaning. When evaluate human input in agricultural production, we can use "number of people" and "period of time". For example, agricultural products of human input is 0.5 person / ton, while another product B  is 0.6 person / ton, which are scientific and comparability.

Subjective orientation of data processing are mainly embodied in purposeful and rational aspects.

Integrated data processing has its purpose. For example, quality of agricultural products can evaluate from indicators as nutrients, body size and appearance of the product. any of these separate indicators are meaningless. Therefore the quality of agricultural products is a comprehensive indicator. The value of this indicator is based on the subjective decisions. Due to the different purposes, each set of indicator scores and weights can vary. Therefore, the aim determines the method of comprehensive indicator.

Integrated data processing has its reasonableness. Realization of integrated indicators is based on indicators and weights. Comprehensive indicators have objectivity. It should reflect the level of the evaluated object (capacity, performance, and so on), to a certain extent, reflect the objectivity, fairness. The other hand, comprehensive indexes have its guidance. Comprehensive indicator is essentially a subjective evaluation, reflects the person's will, making synthetic indicator with a clear direction. In other words, object being evaluated adjust their behavior based on comprehensive index principle in order to achieve good overall indicator score.

**Analysis of the Effectiveness in Data Processing**

**Reliability and Availability of Data Processing**

Data processing of agricultural product logistics must adequately weigh the reliability and availability of data. Raw data is often "irrational distribution" and "not easy to use", then the data would be modified. However, any amendments to the data will inevitably affect the authenticity of the data, even the amendments for singular data may also interfere with the reliability of the data. While increasing information amount and its usability, data processing will always cause some deviation from the original data and reduce its reliability.

To a certain extent, data processing must reflect objective facts, and also guide the evaluated object, to guide them contribute more to the development of their organization. For example, after zooming in or out, the data will have significant impact on integrated indicators. The integrated indicators such as "the quality of agricultural products", "service level", due to extremely small range, the data is zoomed in and involved in calculating comprehensive index, which will have a great influence on comprehensive indicator. Alternatively, when the indicator data range is too large and we shrink it

$$ x_n = \frac{1}{\sqrt{2\pi \sigma}} e^{-\frac{(x_n - \mu)^2}{2\sigma^2}} $$ (8)
down, the comprehensive indicator is influenced little. That leads a negative impact on data availability.

**Analysis on Fairness of Data Processing**

Fairness is always a relative concept. Fairness represents not only the distribution of inputs and outputs, but reflects a subjective sense of identity. This subjective sense is closely related with the specific environment.

Original data processing often pose a challenge to the fairness of the evaluation. The same raw data using different processing methods can produce different results. Therefore, subjective choices of data processing method will directly lead to the problem of fairness.

When subjective data processing method acceptable to all as a data processing specification or system, it is fair. In statistics of agricultural product logistics, trimmed mean is often used to measure central tendency of the data. To some extent, trimmed mean is superior to mean and median because it reduces the range, increased data centralization. Also, if every set of data using this method, the groups comparison will be fair. However, to the trimmed mean, there are two hypotheses. First, the mean is a basic attribute of the data group. In other words, managers believe that data mean is the value property of an object. Secondly, the deviation from the mean of the data is not objective, and greater deviation lead worse objectivity. Precisely because of the existence of these assumptions, trimmed mean is fair and scientific. But it denies the effectiveness of endpoint data, the result may not be in line with the objective facts.

Thus, the fairness of data processing is usually a relatively fair. This fairness must as far as possible reflect the objective facts in the original data. Subjective data processing must depend on the circumstances, if we emphasis on subjectivity too much and set aside specific environment, data processing will be unreliable and unfair.

**Summary**

Data processing, information acquisition and utility producing are different stages and is closely associated with each other. Data deformation is the basic methods of data processing, statistical analysis is the process of data deformation. To some extent, data processing reduces the reliability of the data. Data processing of agricultural product logistics must adequately weigh the reliability and availability of data. The fairness must as far as possible reflect the objective facts in the original data. Subjective data processing must depend on the circumstances, if we emphasis on subjectivity too much and set aside specific environment, data processing will be unreliable and unfair.

**Acknowledgement**

This research was financially supported by the National Natural Science Foundation of China (No. 71540013), Social Science Key Projects of Jiangsu province (13EYA003) and Jiangsu Philosophy and Social Sciences in colleges (2014ZDIXM019).

**References**


