

RESEARCH ON SAMPLING INSPECTION OF NATURAL RESOURCES IN THE BIG DATA ERA

LUO Fujun¹

¹National Quality Inspection and Testing Center for Surveying and Mapping Products, 15811019392, No.28 Lianhuachi West Road, Haidian District, Beijing-1254020931@qq.com

KEY WORDS: Natural Resources Products, Sampling Inspection, Sampling Scheme, Probability of Acceptance, Operating Characteristic Curve, Producer's Risk, Consumer's Risk

ABSTRACT:

Technologies such as satellite remote sensing, global positioning, laser scanning, airborne radar, tilt photography, and drones are rapidly developing. Spatial data is exploding at an unprecedented rate every day, in a short period of time.

The elements of natural resources products vary, and sampling inspection is generally used to evaluate the quality level of a lot of products. The sampling inspection is based on the rigorous probability theory. Samples are randomly taken from a lot of products according to the sampling scheme for inspection, and the quality of the lot is represented by the quality status of a small number of samples. In the small data era, it can achieve quality inspection of natural resources products with the least labor cost and the smallest number of samples. However, how to select a good sample is a difficult problem. In theory, using any set of sample data, we can not get the exact total truth value, and the sampling error is inevitable.

This paper gives an overview of the basic rules of sampling inspection, including the basics of mathematics, basic principles and selection of sampling schemes. It introduces in detail the parameters, characteristics and methods of a sampling inspection, and uses the surveying and sampling scheme adopted for Natural Resources Products inspection. For example, it analyzes the incompatibility of the risk of the producer, the user, the inspector and the sample size selection, and puts forward suggestions for improving the sampling scheme of natural resources products.

1. GENERAL INSTRUCTIONS

Driven by a powerful engine of social and economic development, technologies such as satellite remote sensing, global positioning, laser scanning, airborne radar, tilt photography, and drones are rapidly developing. Spatial data is exploding at an unprecedented rate every day, in a short period of time. In the past ten years, natural resource data from TB to PB level has been accumulated. For example, the amount of data in the western mapping project reached 13.4TB (LI Deren, 2016), and the total data volume of "Map World" was about 30TB. Space sensor resources, the United States has 185 satellites, China has 91 satellites, Quick-Bird, WorldView-1/2/3, GeoEye-1 and other satellites can collect high-resolution images of about 200,000 square kilometers in China every day. It can collect more than one billion square kilometers of global images

every year. By 2020, China will have more than 200 satellites, and the data transmitted by satellites every day can reach the PB level.

The quality status of the product before it is untested is unknown, and the quality status of the product can only be determined after inspection. A quality inspection is an inspection of one or more quality characteristics of a product and comparing the results to specified quality requirements to determine the activity of each quality characteristic in accordance with the specified quality standards (ISO 2859-1, 1999). Economic and reasonable selection of quality inspection methods has important practical significance for improving the timeliness, accuracy and economy of quality inspection work.

Usually, in the case of low production level, if the rate of non-conforming product exceeds the standard and the missed inspection will cause significant consequences, it must be tested

one by one or the whole, which is generally suitable for products with very unstable process quality; Large products; single-piece, small-lot products; expensive, high-precision products; products with special requirements; products with low inspection costs; products capable of applying automatic inspection methods. Its advantage is that the judgment is more reliable, can provide more complete inspection data, and obtain more reliable and reliable quality information. The full test object is a single product. To get 100% qualified products, the only way is to test all. The disadvantage is that the inspection workload is large, the cycle is long, the cost is high, and the inspectors are prone to fatigue, resulting in a large error detection rate and missed detection rate. In this case, sampling inspection has obvious advantages.

Sampling inspection is to obtain samples by random sampling of a lot of products, and to speculate on the overall quality of the lot products based on the inspection results of the samples. The sampling test is a scientific test method based on probability theory and mathematical statistics. Under the premise of sufficient professionalism in the test process, it can greatly save time and labor costs. The sampling test can calculate the deviation of the sampling method in advance, and take certain measures to control its range, fully eliminate the influence of subjective factors, so that the sample taken is fully representative, so as to ensure sampling. The result is inferred to a certain degree of reliability.

How to scientifically and effectively extract samples to achieve the purpose of inspection is a problem to be considered when conducting sampling inspection of natural resources products. It is necessary and reasonable to master the capacity of the sample to be taken. If the capacity of the sample is too large, it will inevitably increase the load of sample extraction, thus forming a phenomenon of sampling and testing waste, and even more is extracted for the purpose of pursuing the sample size. Invalid or irrelevant data, then inferring the population based on the selected sample results will result in non-conformance. At the same time, the sampling test should calculate the error occurred during the quality inspection process, and develop an appropriate sampling scheme to effectively control the error, thereby improving the reliability of the sampling inspection.

2. SAMPLING SCHEME

2.1 Mathematical Model

The proportion of nonconformities "P(d)" refers to the probability that exactly d unqualified products are drawn from the sample. The calculation formula is divided into three cases:

(1) Accurately calculate the formula using the hypergeometric distribution formula.

$$P(d) = \frac{C_{NP}^d C_{N-NP}^{n-d}}{C_N^n} \quad (1)$$

Where P - process proportion of nonconformities, %

N - lot size

n - sample size

d - The number of nonconforming items in the sample

(2) When $N \geq 20n$, the proportion of nonconformities extraction is approximated by the binomial distribution formula.

$$P(d) = C_n^d P^d (1-P)^{n-d} \quad (2)$$

(3) When $n \geq 20$ and $P \leq 0.05$, approximate calculation of proportion of nonconformities using Poisson distribution formula.

$$P(d) = \frac{(nP)^d}{d!} e^{-(nP)} \quad (3)$$

2.2 Quality Parameter of Lot

The total quantity N of the inspection lot, the sample size n and the number of accept A_c , these three parameters determine a sampling scheme (N, n, A_c). When the number of nonconforming items in the sample is determined, the lot of products is judged to be conformed and accepted; when $d > A_c$, it is determined that the lot of products is nonconformity and rejected.

For product with proportion of nonconformities P, the probability that the lot product is received according to a given sampling scheme is called the probability of acceptance, which is represented by L(P), because the number of nonconformity in

the sample is $d=0,1, 2, \dots, Ac$, the lot is judged as conformity, so

the probability of acceptance is:

$$L(P)=P(d \leq Ac) = \sum_{d=0}^{Ac} P(d) \quad (4)$$

2.3 Operating Characteristic Curve

When a lot product is inspected with a certain sampling scheme, the probability of acceptance L of the lot product is a decreasing function of the proportion of nonconformities P , and the relationship is expressed by equation (4), which can be plotted as a operating characteristic curve (OC curve, Figure 1). The OC curve and the sampling scheme are one-to-one correspondence. Changing the parameters in the sampling scheme will cause the OC curve to change.

(1) When n and Ac are fixed, N has little effect on the OC curve (Figure 2). Generally, if the sampling scheme is decided, the influence of N may not be considered. Therefore, the sampling scheme is generally simply represented by (n, Ac) . In fact, if N is too large, if the mistake is made during the sampling inspection, the product will be misjudged as unqualified and rejected, which will bring losses, so the lot cannot be oversized in order to share the inspection cost.

(2) When N and n are fixed, the larger Ac is, the curve is smooth upward, and the scheme is relaxed; the smaller Ac is, the curve becomes steeper downward and the scheme is tightened (Figure 3).

(3) When N and Ac are fixed, the larger n is, the curve becomes steeper downward, and the scheme is tightened. At this time, the probability of receiving the inferior lot and rejecting the high quality lot is steeply small; when n is decreased, the curve is smooth upward and the scheme is relaxed (Figure 4).

(4) When N is fixed, n and Ac change at the same time, n increases and Ac decreases, the scheme is tightened; n decreases and Ac increases, the scheme is relaxed; when n and Ac increase or decrease simultaneously, the influence on OC curve is compared complex, it depends on how much the change range of n and Ac are, and cannot be generalized. If n and Ac are reduced as much as possible, then the scheme is tightened; for the case where n and Ac vary in different amounts, and the respective variation ranges are appropriately selected, the

scheme can be made at $(0, P_1)$ and $(P_1, 1)$. The interval of one interval is tightened while the other interval is relaxed.

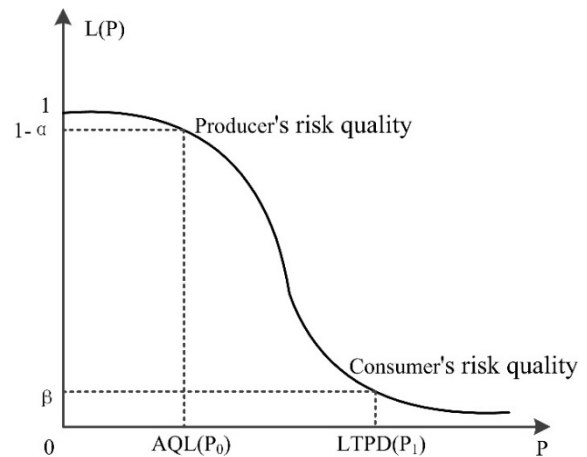


Figure 1. OC Curve

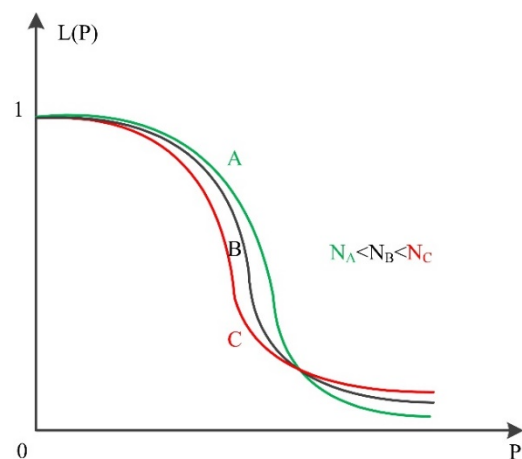


Figure 2. The effect of N on the OC curve

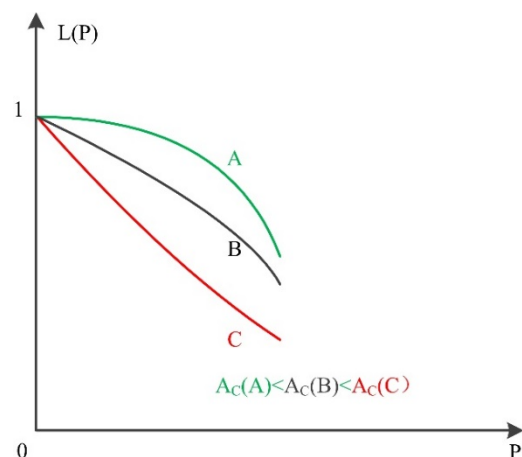


Figure 3. The effect of Ac on the OC curve

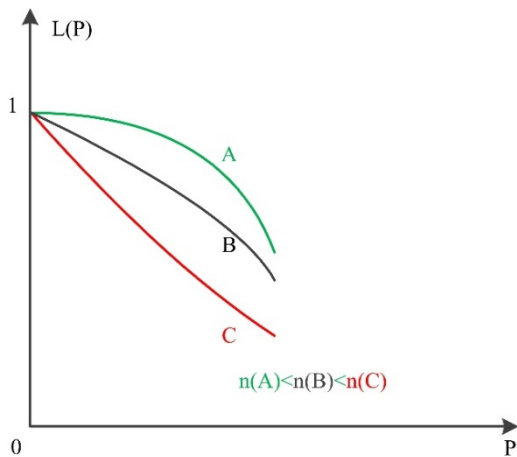


Figure 4. The effect of n on the OC curve

2.4 Discrimination Ratio

In the product quality inspection, the proportion of nonconformities is always first specified as a criterion P_0 , that is, when $P \leq P_0$, the product lot is accepted, when $P > P_0$, the product lot is rejected. Therefore, the ideal sampling scheme should satisfy: when $P \leq P_0$, the probability of acceptance $L(P) = 1$, when $P > P_0$, $L(P) = 0$. The OC curve is two horizontal lines (Figure 5). Actually, the ideal state can only be realized in the total inspection without error detection and missed detection, and does not exist in the sampling inspection.

Although the ideal sampling scheme does not exist, it does not hinder it as the basis for evaluating the merits of the sampling scheme. The closeness of the OC curve of a sampling scheme and the OC curve of the ideal plan is the criterion for evaluating the effectiveness of the plan inspection. In order to measure this proximity, it is usually first to specify two parameters P_0 and P_1 ($P_0 < P_1$), P_0 is the upper limit of reception, that is, it is desirable to approve the product with $P \leq P_0$ as high as possible receive (generally considered to be greater than 95%); P_1 is the lower limit of rejection, that is, it is desirable to receive the product lot with $P \geq P_1$ with the lowest possible probability (generally considered less than 10%). The discrimination ratio (OR) used probability of acceptance 10% to the quality level of the corresponding probability of acceptance 95% quantitatively measures the comprehensive ability of the sampling scheme to distinguish between good lot and bad lot.

$$OR = \frac{P_{0.10}}{P_{0.95}}$$

(5)

Where $P_{0.10}$ - the corresponding quality level when the probability of acceptance is 0.10

$P_{0.95}$ - the corresponding quality level when the probability of acceptance is 0.95

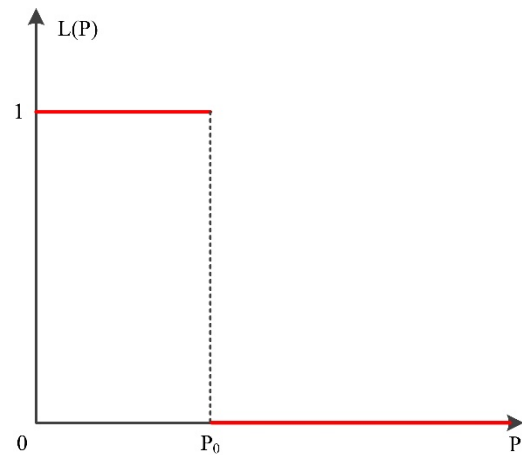


Figure 5. The Ideal OC Curve

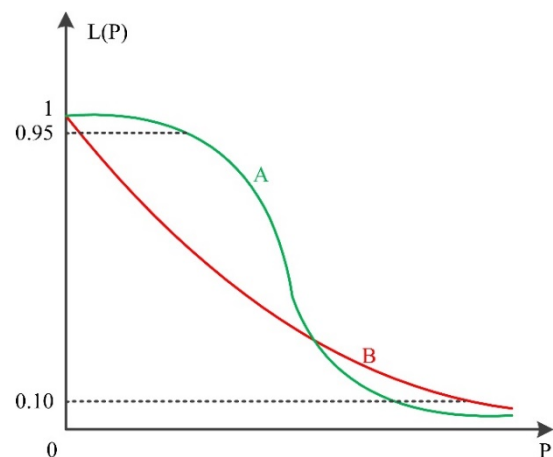


Figure 6. OR of OC Curve

2.5 P_0 and P_1

P_0 is an acceptable quality limit (AQL) and is affected by many factors, such as the category of quality characteristics, and the loss caused by product failure. P_0 is usually negotiated between the producer and the owner, and based on historical data, the average proportion of nonconformities is selected.

P_1 is the lot tolerance percent defective (LTPD), and its selection

should be separated from P_0 . When $P_1 \leq 3P_0$ will increase the sample size, when $P_1 \geq 20P_0$ can not guarantee the quality of the project, generally $4P_0 \leq P_1 \leq 10P_0$ is specified.

2.6 Risk of Sampling Inspection

Sampling inspection is based on the inspection results of some products to infer the quality of the entire lot. It is inevitable to make two types of mistakes. The first type is the risk that the quality level is acceptable but not received by the sampling scheme, called producer's risk, expressed by α ; the second is when the quality level is not satisfied. The value, but the probability of being received by the sampling scheme, is called the consumer's risk and is expressed by β (GB/T13264-2008).

3. ANALYSIS OF CURRENT SAMPLING SCHEME

Currently, natural resources products mainly refer to spatial data management and inspection. A good sampling scheme should take into account the interests of the producer and the consumer, strictly control the two types of error probabilities, and consider the cost of the inspection. The sampling scheme should be based on the proportion of nonconformities P_0 and P_1 proposed by the producer and the consumer respectively. When the proportion of nonconformities of the lot product is not greater

than P_0 , the manufacturer's risk is not to exceed α , and the calculation is started from $Ac=0$, and the formula (5) is obtained, and the sample size upper limit n_1 is obtained; when the lot product proportion of nonconformities is greater than the consumer risk P_1 , the consumer risk does not exceed β , and the formula (4) is obtained, and the lower limit of the sample size n_2 is obtained. When $n_1 \geq n_2$, n_2 is taken as the sample size, and the sampling scheme is (N, n_2, Ac) ; when $n_1 < n_2$, the number of pass judgments Ac is increased by 1, and the calculation is cycled. When Ac is large and $L(p)>0.95$, α can take 0.05; when Ac is small and $L(p)<0.1$, β can take 0.1 (CAI Xia, WU Lingyun). Due to lack of experience and basis, it is extremely difficult to accurately judge P_0 and P_1 for producer and users. In this paper, let $p_0=p_1$ (LUO Fujun, 2017), then $\alpha+\beta=1$, according to the geomatics products inspection standards. The lot upper limit and the corresponding sample size are used to calculate the production side risk under different lot (Table 1). The geomatics requirements are 0 non-conforming products, that is, $Ac=0$, and the number of unqualified products in the lot N is rounded up.

$$\begin{cases} L(P_0) \geq 1 - \alpha \\ L(P_1) \leq \beta \end{cases} \quad (6)$$

Table 1. Consumer's Risk

n / N p	3/20	5/40	7/60	9/80	10/100	11/120	12/140	13/160	14/180	15/200
1%	15.00%	87.50%	88.40%	88.80%	90.00%	82.50%	83.60%	84.40%	85.10%	85.60%
3%	15.00%	76.30%	77.90%	69.60%	72.70%	67.80%	63.50%	65.10%	66.40%	62.30%
5%	15.00%	76.30%	68.50%	61.50%	58.40%	55.50%	52.70%	50.00%	47.50%	45.10%
10%	71.60%	57.30%	45.90%	36.80%	33.10%	29.80%	26.80%	24.10%	21.60%	19.40%
15%	59.70%	42.30%	30.00%	21.30%	18.10%	15.40%	13.10%	11.10%	9.40%	8.00%

(1) As the product with the same lot proportion of nonconformities, the producer's risk increases, the producer will consciously improve the quality of the product; the product with stable proportion of nonconformities increases with the increase of the lot quantity, and the producer's risk also increases. It shows that the current sampling scheme has the defects of "small lot is easy, large lot are difficult", and the producer will tend to submit the acceptance trend in small lot and increase the total workload of the inspection.

(2) The inspection standard used for geomatics products stipulates that $Ac=0$, which causes the risks of the producer, the

consumer to be very large.

(3) The standard stipulates that when the lot exceeds 200, the lot inspection will not be applicable to the inspection of the vast amount of remote sensing products, land coverage products and etc. in the era of big data, which will lead to huge inspection workload.

(4) The determination of the product failure rate is a difficult point. The average value of the percentage of non-conforming products in a series of inspection lot can be used to determine the AQL value, which is to average the previous inspection data and determine the AQL value based on this.

79-384.

4. CONCLUSIONS

Natural resources products have the general attributes of products, that is, there are certain non-conforming products objectively. Taking the third national land survey orthophoto product as an example, the proportion of nonconformities is about 30% in the inspection. At the same time, as a spatial data product, nonconforming products can be turned into excellent products through continuous modification and improvement. Therefore, the sampling scheme of geomatics products designed based on the principle of zero non-conforming products is directly applied to the inspection of natural resources products, which will lead to extremely high inspection risks. In order to ensure the quality of natural resources products, the producer must be required to pay attention to quality, carry out self-examination in the mode of full inspection before the inspection, and face up to the fact that there is still a gap between the current natural resources products and "zero unqualified products", and pass historical data. And the production process analysis, the development of reasonable product proportion nonconformities, acceptable quality limit, lot tolerance percent defective, the existing sampling program to improve, in order to meet the current natural resources product sampling inspection and application.

GB/T13264, 2008. Sampling procedures and tables for small lot inspection by attributes for percent nonconforming items.

CAI Xia, WU Lingyun, 2011: Improvement of Single Sampling Table and Its Application in Agricultural Production, *Journal of Anhui Agri. Sci.*, 39(32), 96-19728.

LUO Fujun, 2017: Discussion on the Use of GB/T 18316-2008. *Geomatics & Spatial Information Technology*, Vol.40, No.3, 54-56.

ACKNOWLEDGEMENTS

Founded by National Natural Science Foundation of China (416 71440).

Founded by the Quality Wing Program of National Quality Inspection and Testing Center for Surveying and Mapping Products (2018).

REFERENCES

ISO 2859-1, 1999. Sampling procedures for inspection by attributes - Part1: Sampling schemes indexed by acceptance quality limit (AQL) for lot-by-lot inspection.

LI Deren, 2016: Towards Geo-spatial Information Science in Big Data Era. *Acta Geodaetica et Cartographica Sinica*, 45(4),