[Volume 1 issue 7 Aug 2013]
Page No.189-196 ISSN :2320-7167

INTERNATIONAL JOURNAL OF MATHEMATICS AND COMPUTER RESEARCH

Inter-Sample-Disparity Proportions Used as a Quality-Control Method

Bisi Alabi-Labaika

Department of Business Administration, University of Lagos, Akoka, Yaba, Lagos State, Nigeria

e- mail: annafisa1960@gmail.com

Abstract: Inter-sample-disparity proportions were derived and applied to data on the number of defective water bottles in a factory. The result showed that the proportion of the first-sample observations that were numerically greater than the second -sample observations was 84/100 and the proportion of the observations in the first sample that were less than the observations in the second sample was 16/100 . The difference between the two proportions, 68/100, was tested significant with an adapted normal test of the difference between two proportions. This result implied that the population contains heterogeneous elements confirming the unstable position of the production process.

Key words: statistical quality control, manufacturing, product-homogeneity, disparity-proportions, sampling, two-sample test, ordered data, indicator variables

1.0 INTRODUCTION

1.1 **A General Explanation of the Research Area: Statistical Quality Control**

Statistical quality control involves the use of various statistical techniques to decide whether or not a manufactured product has met a pre-stated quality level and may be delivered to the customer. It is also classified as parametric statistical quality control and non-parametric statistical quality control. The latter, non-parametric statistical quality control, based mainly on order statistics and indicator variables, is the focus of this study, order statistics being a set of observations arranged in an increasing order of magnitude, that is, arranged by starting from the smallest going through to the highest. Parametric statistics is based on the parameters of a probability distribution such as the population mean (μ) and the population standard deviation (σ) of the normal probability distribution. Nonparametric (distribution-free) statistics does not depend on such parameters of a probability distribution as many situations may not necessarily follow a normal distribution as assumed. Further, in an effort to explain statistical quality control, Stevenson (2005) divides it into two categories: statistical process control during manufacturing and acceptance sampling of the finished products. Statistical process control is for manufacturing a product. Acceptance sampling is for finished products. This emphasis of this study is on finished products. Quality control is the use of techniques and activities to achieve, sustain and improve the quality of a product or service.

Neave and Wheeler (1996) stated that Shewhart created the control chart with 3-sigma limits. Shewhart's use of 3-sigma (population standard deviation) limit, as opposed to any other multiple of sigma, did not stem from any specific mathematical computation. Rather, he said that 3.0 seemed to be an acceptable economic value, and that the choice of 3.0 was justified by empirical evidence that it worked. This pragmatic approach is markedly different from the more strictly mathematical approach commonly seen in the journals of today. In fact, in order to have a practical and sensible approach to the construction of control charts, Shewhart deliberately avoided overdoing the mathematical detail. The objective is to give the user a guide for taking appropriate action-to look for assignable causes when the data display uncontrollable variation, and to avoid looking for assignable causes when the data display controllable variation.

The Objective of the Study

The objective of this study, again, is testing for the significance or otherwise of the disparity (one-to-many-correspondence ordinal difference) between two samples using order statistics and binary-digit type of indicator variables and probability and the concept of statistical quality control.

2.0 Literature Review

Establishing the difference between two populations remains a major concern in statistics Several attempts have been made to develop appropriate statistical techniques for estimating such a difference. In this study inter-sample disparity proportions and not the usual one-to-one correspondence difference is focused. Disparity is a one-to-many correspondence ordinal difference. That is, it is the difference between two sets of an application of inequalities based on ordered sample points.

Between two real numbers, x and y, there are three relationships:

$x = y$, $x>y$, or $x<y$.

This is one of the order axioms or the law of trichotomy (Adepoju, 2000).

It is this law that has been applied to formulate the following indicator variables:

IV_1: dij = {1 if $x<y_j$
{0 if $x_i> y_j$

IV_2: dij = {1 if $x_i> y_j$
{0 if $x_i<y_j$

where x_i belongs to sample vector $X_n = (x_1, x_2, \ldots,x_n)$ and y_j belongs to vector $Y_n = (y_1, y_2, \ldots, y_n)$, $i<j$, i and j are positive integers. Observations x_i and y_j are ordered. The arrangements of X_n and Y_n are sequential.

It is the objective of the present study to derive the probability of the disparity between two independent samples based on the two selected indicator variables .

Chakraborti and Wiel (2008)'s "A nonparametric control chart based on the Mann-Whitney statistic" presented the Mann-Whitney (MW) control chart this way.

A reference sample or (a training sample) denoted by $X=(x_1, x_2, \ldots, x_m)$ is from an in-control manufacturing process. A test sample denoted by $Y^h =(y_1, y_2, \ldots,y_n)$ is also taken subsequently at an appointed time interval(h) from the same process for h=1, 2, …

Assumption: all the samples involved are independent of one another.

The MW test statistic is

$$M_{xy} = \sum_{i=1}^{m} \sum_{j=1}^{n} I(X_i<Y_j)$$

$$= \sum I(Y_j>X_1 +Y_j>X_2 +Y_j>X_3+ \ldots+Y_j>X_m)$$

Where $I(X_i<Y_j)$ is the indicator variable (indicator function, dummy variable, binary digits, uniform distribution) for the event $I(X_i<Y_j)$.

$$0< Mxy < mn$$

Large values of Mxy indicate a positive shift and small values of it signal a negative shift. The process under test is out of control

If

$$M^h_{xy} < L_{mn}\text{, the lower control limit}$$

Or if

$$M^h_{xy} > U_{mn}\text{, the upper control limit}$$

The distribution of M_{xy} is known to be symmetrical about mn/2 when the process is in-control. At that point of control $L_{mn} = mn - U_{mn}$

In this project we probabilistically compare two samples in all possible pairs and decide to reject or accept the null hypothesis that there is no significant disparity between the two samples.

3.0 Methods

This section provides a new approach to assess the disparity (ordered difference) between two independent samples in statistical quality control based on order statistics and the proposed indicator variables and how to collect data with which the method is to be tested.

When analysing the difference between two samples, the data are taken as given. However, in this study, it is proposed that each sample should have its observations ordered before any analysis of that nature can be carried out. For, it is by doing that that corresponding values in terms of sizes or positions are considered. It is reasonable to do that.

Recall that

$$IV_1: d_{ij} = \begin{cases} 1 \text{ if } x<y_j \\ 0 \text{ otherwise} \end{cases}$$

$$IV_2: d_{ij} = \begin{cases} 1 \text{ if } x_i \geq y_j \\ 0 \text{ otherwise} \end{cases}$$

Derivations

In manufacturing, in taking samples for the purpose of constructing a control chart, it is necessary to know the relationship between a sample (except the first sample) and its predecessor(s) and between the same sample (except the last one) and its successor(s). The reason for this is to find out whether or not the observations in the sample under consideration are dominating or being dominated by the observations in the other samples. In that case there are two relationships: the relationship to the preceding samples (Backward Pass Relationship) and relationship to the succeeding samples (Forward Pass Relationship).

Let m samples each of size n have their observations ordered and symbolized as column vectors. Apply an iv to the ordered observations in pairs (x_i, y_j) for all possible pairs

Let the current reference column vector be denoted by X_i and let every one of the subsequent column vectors be denoted Y_j.

Further, let $d_{ij} = \sum\sum I(x_i<y_j)$ for every x_i belonging to X and every y_j belonging to Y for i =1, 2, 3, …, m-1 and j =2, 3, 4, …, m for m samples.
Sample sizes are each equal to n. Hence, mn in Chakrabort,et al. (2003, 2008) becomes n^2 for this work.
Then we shall have the following array of values for m samples each of n observations with ordered paired relationships:

$$
\begin{array}{llllll}
X_1 = x_{11} & & & & & \\
\quad x_{12} & & & & & \\
\quad x_{13} & & & & & \\
\quad . & & & & & \\
\quad . & & & & & \\
\quad . & & & & & \\
\quad x_{1n} & d_{1j} & & & & \\
X_2 = x_{21} & n & & & & \\
\quad x_{22} & n & & & & \\
\quad x_{23} & n & & & & \\
\quad \ldots\ldots & & & & & \\
X_m = x_{m1} & n \ldots & & & & \\
\quad x_{m2} & n & n \ldots & & & \\
\quad x_{mn} & & n & n & n \ldots & n
\end{array}
$$

Consider the following iv scores :

$$
d_{ij} = 1 \text{ if } X_i < y_j
$$
$$
\quad = 0 \text{ otherwise}
$$

that has been applied on the last array(table) of values
It is being proposed here that the relationship between the scores, d_{ij}, and the maximum score, n^2t, be symbolized as p,

$$
p = \frac{\text{observed score (actual score)}}{\text{maximum score}} = \frac{\sum d_{ij}}{n^2(w)} \quad \ldots\ldots\ldots\ldots\ldots\ldots(1)
$$

where n is the sample size, w is the number of test samples compared to the current reference sample.

Proof: From the fore-going table of values the maximum value of d_{ij} for every $X_i < Y_j$ for i =1, 2, 3, …, m-1 and j = 2, …, m is n.
For all $Xi < Y_j$ for i =1, 2, 3,…, m-1 , the maximum score $\sum\sum I(Xi < Y_j)$ is n^2. That is, all x_i's are less than all y_j's. Let this happen in w samples.
Putting the results together gives the maximum score as $n^2(w)$.

Brown, Newcombe and Zhao (2009) introduced separation measure for the Mann-Whitney two-sample test for two distributions as

$$
\Theta = P(x>y) - P(x<y)
$$

Frank and Althoen (2002) made proportion an alternative to probability.
In this paper the difference between two proportions is used to test the significance or otherwise of the disparity (one-to-many-correspondence ordinal difference, OMCOD) between two independent samples using (1). OMCOD is not the same as the existing difference which is a one-to-one correspondence difference (OOCD). OMCOD is more embracing in that every observation in sample X is compared with all the observations in sample Y as against comparing the first observation in sample X with the first observation in sample Y and so on.

The test process

Let $P_{xi>yj}$, be the proportion of $x_i<y_j$ and $P_{xi>yj}$ be the proportion of $x_i>y_j$

$P_{xi<yj}$ + $P_{xi>yj}$ =1 if there is no $x_i= y_j$

This is a property of probability.

Recall from (1) that

$p = \sum dij/n^2$ ………………………………………………………………………….(2)

when w =1 and that $0< p \leq 1$

Ho: $P_{xi<yj-} P_{xi>yj =0}$

H_1: $P_{xi<yj-} P_{xi>yj\ j} \neq 0$

The test statistic is that of the difference between two proportions.

Assuming the normal distribution for the proportions $P_{xi<yj}$ and $P_{xi>yj}$

For two proportions let $p = \frac{(x_1+x_2)}{(n_1+n_2)}$

be an equivalent of $(dx<y +dx>y)/)/2n^2$ ……..(3)

Variance of p = $pq(1/n^2 +1/n^2))$

$= (dx<y +dx>y)/)/2n^2 (1-(dx<y +dx>y)/2n^2)(2/ n^2)$

$=(dx<y +dx>y)/2n^2 ((2n^2-(dx<y +dx>y)/ 2n^2 {}_/2/ n^2)$

$=(dx<y +dx>y) ((2n^2-(dx<y +dx>y))/ 2n^6$ …………….(4)

(With the continuity assumption)

$= n^2(2n^2-n^2)/2n^6$

$=n^2n^2/2n^6$

$=1/2n^2$ ………………………….(5)

$s = 1/\sqrt{2}n =\sqrt{2}/2n$…………………………(6)

The test statistic is

$Z =D/s= D2n/\sqrt{2}$ ……………………….(7)

Based on the foregoing derivations we have Table 1.

Table 1 : Derivation of the Disparity Between Two Independent Samples

Sample 1	Sample 2	$x_i<y_j$	$x_i>y_j$
Ordered x	Ordered y	dij	dij
$x_{(1)}$	$y_{(1)}$	d11	d21
$x_{(2)}$	$y_{(2)}$	d_{12}	d_{22}
.	.	.	.

.	.	.	.
.	.	.	.
$x_{(n-1)}$	$y_{(n-1)}$	d1n-1	d2n-1
$x_{(n)}$	$y_{(n)}$	d1n	d2n
	Total	$\sum d_{ij}$	$\sum d_{ij}$
	P	$P_{xi<yj} = \sum d_{ij}/n^2$	$P_{xi>yj} = \sum d_{ij}/n^2$
	D	$P_{xi<yj-} P_{xi>yj}$	

4.0 Results

Consider Table 2 of ten samples of defective water plastic bottles in a factory for the required testing of the new methods:

Table 2: Ten samples of ordered observations from daily output of defective plastic- bottles' factory

	1	2	3	4	5	6	7	8	9	10
Item										
1	43	9	16	32	10	36	21	20	13	20
2	76	27	17	37	34	40	30	27	17	31
3	94	32	37	46	48	54	52	38	35	32
4	202	52	39	78	62	57	59	64	46	46
5	290	149	71	259	136	116	104	140	104	53
Total	705	269	180	452	290	303	266	289	215	182
Mean	141	54	36	90	58	61	53	58	43	36

On Table 2, samples 1 and 2 of means 141 and 54 are compared for disparity proportion using

$$IV_{2:}\ d_{ij} = \{1 \text{ if } x_{i>}y_j \quad \{0 \text{ otherwise}$$

This is shown on Table 3.

Table 3 : Generation of disparity between samples X and Y

X	Y	x>y	x<y
43	9	3	2
76	27	4	1
94	32	4	1
202	52	5	0
290	149	5	0
	∑dij	21	4

From Table 3 $P_{xi>yj} = \frac{\sum d_{ij}}{n^2 t} = 21/25(1) = 0.84$

From Table 3 $P_{xi<yj} = \frac{\sum d_{ij}}{n^2 t} = 4/25(1) = 0.16$

The total probability or proportion is 0.84+ 0.16 =1.00

The inter-sample disparity proportion, D= / $P_{xi<yj}$ - $P_{xi>yj}$ / = /0.84 -0.16/ = 0.68 or 68%

Since dx<y+dx>y21+4= 25= n^2 the test statistic from (7) is
Z =D/s
= D2n/√2
= 0.68(2(5))/ 1.414= 4.809052
which is significant at α=0.05 of 1.65 and at α =0.025 of 1.96 and at α = 0.1 of 1.28

Two samples of equal means
Consider Table 4.

Table 4:Disparity between two samples of approximately equal means (36, 36.4)

	X	Y	x<y	x>y
	16	20	5	0
	17	31	5	0
	37	32	2	3
	39	46	2	3
	71	53	0	5
Total	180	182	14	11
Mean	36	36.4		

On Table 4, since dx<y+dx>y=14+11= 25= n^2
the test statistic from (7) is
Z =D/s
= (14-11)/25(2n/√2)
= 0.12(2(5))/ 1.414= 0.848656

This is not significant at α=0.05 of 1.65 and at α =0.025 of 1.96 and at α = 0.1 of 1.28

From Table 4 the probability of disparity , 0.12 , is not significant.
Compare with the difference between the two sample means of 36 and 36.4 , i.e., a difference of 36.4-36=0.4 which, without a test, will have been deemed not significant.

Discussion: Two random samples have been used to determine whether or not there is a significant difference between the two populations from which the samples were taken. From Tables 3 and 4 and their analysis, the probability that the observations in sample 1 are less than the observations in sample 2 is 0.84 and that the observations in sample 1 are more than the observations in sample 2 is 0.16. The inter-sample disparity proportion is 68%. In the second pair of samples it is 12%

Conclusion : In the two examples, the observations in sample 1 are sufficiently larger than the observations in sample 2. Hence, the two samples are significantly not the same, hence, their parent populations. If the two samples were from the same population, a significant test means the parent population contains heterogeneous elements, thus justifying the rejection of the population in quality control.
The implication of this for the host factory is that the significant disparity proportion signifies that `the number of defectives is not controlled. Since this collection is on a daily basis, it means the production process is not

stable as a result of identifiable causes such as the materials, methods of handling, operators' problems and others which need to be sorted out. If the production process is stable, the two proportions should be approximately equal, and the disparity should also be approximately 0.

Recommendation: This method , **Inter-Sample-Disparity Proportions,** between two independent samples is very simple compared to the use of assumed probability distributions.

For further research, inter-sample disparity proportions may be compared with the correlation of the corresponding ordered data. There is a need to find out whether two samples of approximately equal averages have their inter-sample-disparity proportions approximately equal

References

Adepoju, J. A. (2000) A First Course in Real Analysis: Expository Texts in Undergraduate Mathematics, Uinversity of Lagos, Lagos, p. 26-27, 194pages

Brown, B. M., Newcombe, R. G. and Zhao, Y. (2009) Non-null Semi-parametric Inference for the Mann-Whitney Measure, Journal of Non-parametric Statistics, vol. 21, Issue 6, pp. 745-755

Chakraborti, S.and vande Wiel, M. A. (2008) A nonparametric Control Chart Based on the Mann-Whitney Statistic, Beyond Parametrics in Interdisciplinary Research , pp. 156-172

Chakraborti, S. and van de Wiel (2003) A Nonparametric Control Chart Based on the Mann-Whitney Statistics, Department of Mathematics and Computer Science, Eindohoven University of Technology, Eindohoven, The Netherlands,e-mail:m.a.v.d.wiel@TUE.nl and S. Chakraborti, Dept. of Information Systems, Statitstics and Management Scieence, Univ. of Alabama, Tuscaloosa, Alabama,, U.S.A.

Frank, H. and Althoen, S. C. (2002) Statistics Concepts and Applcation, Cambridge : Cambridge University Press, pp.635-647

Neave, H. N. and Wheeler, D. J. (1996) Shewhart's Charts and the Probability Approach, 2002 SPC Press, Knoxville, Tennessee, Ninth Annual Conference of the British Deming Association

Stevenson,W. J. (2005) Operations Management, (8^{th} edition), McGraw- Hill, Boston

Mason, D. M. and Schuenemeyer, J. H. (1983) A Modified Kolmogorov-Smirnov

Test Sensitive to Tail Alternatives, Annals of Statistics, Volume 11, Number 3, pp. 933-946

Siegel, S. and Castellan Jr, N. J. (1988) Nonparametric Statistics for the Behavioral Sciences, 2^{nd} edition, McGrawl Hill, New York : pp.45-51

SPC for MS EXCEL (e-mail: bill@spcforexcel.com) March 2005