Statistical Diagnosis(SPD) for Control of SARS Epidemic Situation of Beijing

Gongxu Zhang[1] & Jing Sun[2]

May 11 of 2003, revised on June 10

Abstract

Under the strong leadership of Chinese Government to the anti-SARS struggle, the situation has been successfully controlled. Since May 1 of 2003, the Ministry of Health of China published daily the number of newly increased SARS patient of Beijing, the authors analyzed these data using X_{cs}-R_{scs} cause-selecting control charts of Statistical Diagnosis(SPD) Theory. Data about number of newly increased SARS patient consists of two kinds of variation: random variation and tendency variation of SARS epidemic. It is concluded that SARS epidemic of Beijing was already controlled since May 9 of 2003.

Key Words : anti-SARS, Statistical Process Diagnosis(SPD), X_{cs}-R_{scs} cause-selecting control charts, controlled state.

Under the correct and strong leadership of Chinese Government to the anti-SARS struggle, China has achieved great successes. In this paper the authors showed that the SARS epidemic had been controlled from the viewpoint of Statistical Process Diagnosis(SPD) Theory according to SARS data published in Beijing Times page 02 dated on May 14 of 2003, see Table 1, column (3).

The so-called SPD is to use statistical methods to diagnose causes which induced the abnormality, SPD can be used to all trades including the control of SARS epidemic.

In this paper, we take y as the number of newly increased SARS patient everyday in Beijing expressing the SARS epidemic situation, see column (3) of Table 1.

[1] Professor, Supervisor of Ph.D. Students, Beijing University of Science & Technology, Beijing 100083, P.R. China. Tel.: 8610-62334002, Fax: 8610-62322033, E-mail: **gxzhang1@163.net**

[2] Associate Professor , School of Economics & Management, Tsinghua University, Beijing 100084, P.R. China. Tel.: 8610-6277-2087, Fax: 8610-62784555, E-mail: **sunj3@em.tsinghua.edu.cn**

Table 1. Calculating Table of X_{cs}-R_{scs} Cause-Selecting Control Chart

Group No. (1)	X (2)	Y (3)	$R_s(X)$ (4)	$R_s(Y)$ (5)	$\hat{Y}$ (6)	Y_{cs} (7)	R_{scs} (8)
1	1	122	–	–	112.486	9.514	–
2	2	96	1.00	26.00	106.598	-10.598	20.112
3	3	114	1.00	18.00	100.710	13.290	23.888
4	4	69	1.00	45.00	94.822	-25.822	39.112
5	5	98	1.00	29.00	88.934	9.066	34.888
6	6	70	1.00	28.00	83.046	-13.046	22.112
7	7	97	1.00	27.00	77.158	19.842	32.888
8	8	94	1.00	3.00	71.270	22.730	2.888
9	9	48	1.00	46.00	65.382	-17.382	40.112
10	10	54	1.00	6.00	59.495	-5.495	11.888
11	11	42	1.00	12.00	53.607	-11.607	6.112
12	12	48	1.00	6.00	47.719	0.281	11.888
13	13	48	1.00	0.00	41.831	6.169	5.888
14	14	39	1.00	9.00	35.943	3.057	3.112
					Sum:	-0.001	254.888
					Average:	-0.0001	19.60677

Note: Here, we used the linear regression model, if we use the curvilinear regression model, the error would be smaller, but it improved only a little, and thus be omitted for simplicity.

1. Model Selection

First, we make the scatter diagrams of data y, see Fig. 1(linear model) and Fig. 2(curvilinear model) respectively, evidently, the error of Fig. 2 is smaller than Fig.1, but the error was improved only a little, hence we choice the linear model for simplicity.

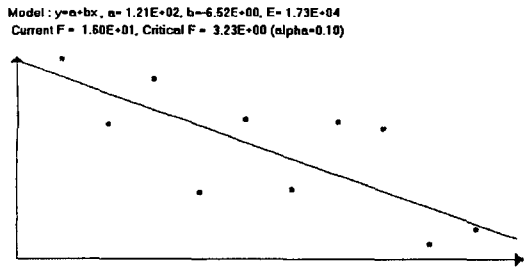

Figure. 1 Scatter Diagram for Linear model

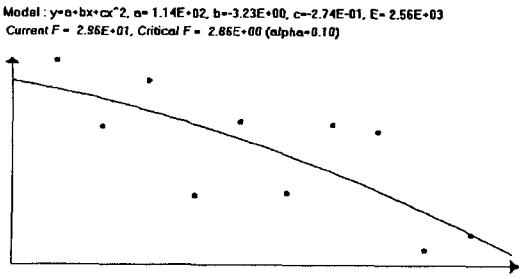

Figure. 2 Scatter Diagram for curvilinear model

2. Types of Variation of y Data

The variation of y data consists of two types, which are respectively:

① random variation

② tendency variation of SARS epidemic situation

The former reflects the chance, and is inevitable, the latter reflects the certainty and is the basis on which we can decide whether the SARS epidemic of Beijing is controlled or not. These two kinds of variation are interweaved together and is not easy to be discerned, hence we need use the SPD theory to differentiate them. Again, since the number of data is not so many especially at the beginning stage, we have to use the simplest cause-selecting X_{cs}-R_{scs} control charts to analyze the SARS epidemic.

3. When is the SARS epidemic of Beijing controlled?

We can analyze this problem through the following two viewpoints:

(1) From the X_{cs}-R_{scs} Cause-Selecting Control Chart, see Fig. 3.
From Fig. 3 we can see both the R_{scs} chart and the X_{cs} chart are shown in statistical control, these mean that the SARS epidemic of Beijing is under control from the viewpoint of SPD.

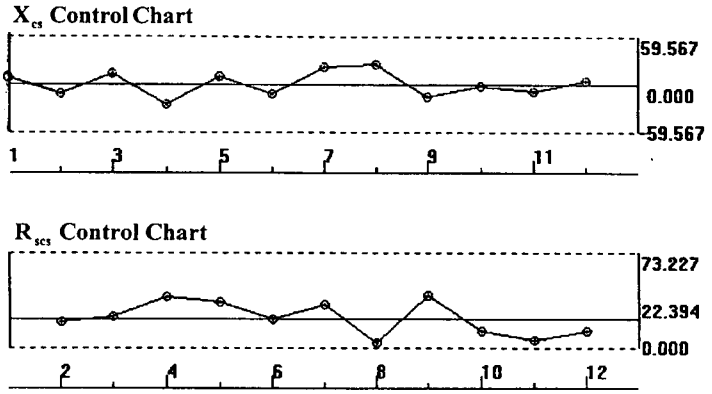

Figure. 3 The X_{cs}-R_{scs} Cause-Selecting Control Charts

(2) From the correlation of y data shown in Table 2, we can see that after May 9 of 2003, the y data always show significant of linear correlation, this fact expresses that the SARS epidemic of Beijing was controlled early since May 9.

Table 2. SPD Used for the Analysis of SARS Epidemic of Beijing

serial number of data (1)	Correlation significance (2)	Number of newly Increased SARS Patients (3)	SARS Epidemic Situation (4)
2	non	96	not Controlled
3	non	114	not Controlled
4	non	69	not Controlled
5	non	98	not Controlled
6	yes	70	controlled, but not considered
7	non	97	not Controlled
8	non	94	not Controlled
9	yes	48	controlled since May 9
10	yes	54	controlled
11	yes	42	controlled
12	yes	48	controlled
13	yes	48	controlled
14	yes	39	controlled

From the X_{cs} cause-selecting control chart shown in Fig. 3, we also can see that as far as the number of newly increased SARS patient of next day is smaller than the corresponding critical boundary 59[3], the SARS epidemic of Beijing is still under control from the viewpoint of SPD. For example, the number of newly increased SARS patient of next day (May 15), is 27, hence, the conclusion: the SARS epidemic of Beijing is still under control.

4. What is the meaning of controlled state to SARS epidemic?

Different from product/service, SARS is a kind of epidemic disease, the state controlled, of course, does not mean that we want to reserve the SARS, but expresses only that we can discern the random variation and the tendency variation. The control limit of the cause-selecting control chart is just the critical boundary to discern them. Now, the object controlled is the SARS epidemic, hence we want the critical boundary will decrease and even removed gradually as the SARS epidemic becomes weakened even disappeared(i.e. decreased to zero) gradually. Thus, we need to add the new y datum of each day together with the original data to calculate the new X_{cs}-R_{scs} Cause-Selecting Control Chart.

5. The More is the Data, the More Severe is the Critical Boundary

As the number of y data increased, the control limit will be more severe. This phenomena is reasonable, because if there are more data, its regularity will be more strong, and the control limit would be more severe as can be seen from column (4), Table 3 and Fig. 4.

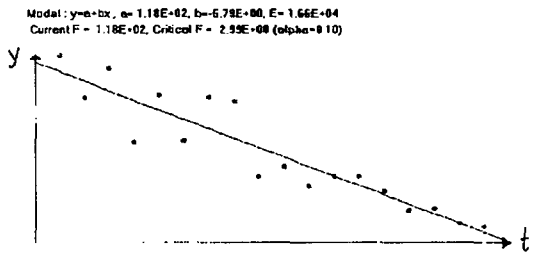

Figure. 4 Fitting of y data and the linear regression model

[3] Here, we must add the data of newly increased SARS patient of each day to correct the boundary of "59". Normally, it would be decreased gradually.

From Fig. 4, we can see that the fitting of the last 17 points would be much more better, hence we make the scatter diagram of the last 17 points (May 25 to June 10), see Fig. 5.

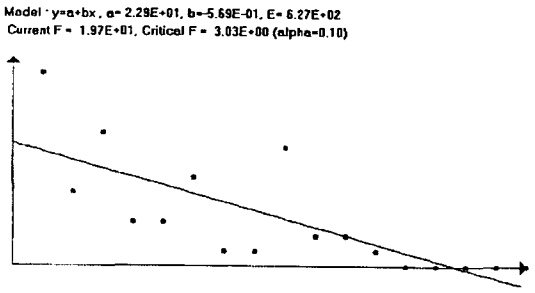

Figure. 5 Scatter diagram of last 17 points (May 25 to June 10)

6. Prediction of y data

The mathematical model adopted in this paper is a linear model, in general, we cannot use this model outside its area of linear model significance, but observe Fig. 4, we can see that this linear model and the SARS epidemic fits very well, and they fit each other better and better as the number of y data is increased, hence we can use the last 17 points to make the scatter diagram, see Fig. 5. Thus, we could make the following prediction: When will the y data could be decreased to zero?
From Fig. 5, we know that

$$\hat{Y} = 22.9 - 0.569x$$

Now, let $\hat{Y} = 0$, then

$$x=22.9/0.569=40.25 \approx 41 \text{days (since June 11)}$$

i. e. we predict that the daily number of newly increased SARS patient of Beijing would decrease to zero on July 22 of 2003.

Table 3. The Criterion for Diagnosing Beijing SARS epidemic with SPD

Date (1)	Number of data of y (2)	Linear Correlation (3)	Control Limit (4)	y data of next day (5)	Judgement of SARS Epidemic Situation of Beijing(6)	Illustration (7)
May 1	1				no judgement	
May 2	2	not significant	non		no judgement	
May 3	3	not significant	non		no judgement	
May 4	4	not significant	non		no judgement	
May 5	5	not significant	non		no judgement	
May 6	6	significant	not considered		not considered	individual significant, not considered
May 7	7	not significant	non		no judgement	
May 8	8	not significant	non		no judgement	
May 9	9	significant	71.8271	54	controlled	y data is linear model significant, i.e. controlled since May 9
May10	10	significant	67.3967	42	controlled	
May11	11	significant	62.2462	48	controlled	
May12	12	significant	60.9660	48	controlled	
May13	13	significant	55.8655	39	controlled	
May14	14	significant	52.1552	27	controlled	
May15	15	significant	52.0952	28	controlled	
May16	16	significant	47.0547	19	controlled	
May17	17	significant	45.0545	17	controlled	
May18	18	significant	42.9942	7	controlled	
May19	19	significant	41.2341	12	controlled	
May20	20	significant	38.7138	8	controlled	
May21	21	significant	37.1437	15	controlled	
May22	22	significant	37.1737	15	controlled	
May23	23	significant	36.4636	26	controlled	
May24	24	significant	35.3435	13	controlled	
May25	25	significant	35.4635	5	controlled	
May26	26	significant	35.3435	9	controlled	

May27	27	significant	34.7734	3	controlled	
May28	28	significant	33.6233	3	controlled	
May29	29	significant	32.7732	6	controlled	
May30	30	significant	32.2432	1	controlled	
May31	31	significant	31.2531	1	controlled	
June 1	32	significant	38.3838	8	controlled	
June 2	33	significant	33.4133	2	controlled	
June 3	34	significant	29.6629	2	controlled	
June10	We use the last 17 points (May 25 to June 10)as the input data of linear model.	significant	7.19≈7 (Considering the following axis translating,: x'=x y'=y+1 y is in fact 6)	1	controlled	The X_{cs}-R_{scs} charts are in control., and the control limit is 6. If the value of y on next day is smaller than 6, it is the random variation; otherwise, it is the tendency variation of SARS epidemic of Beijing, i,e. a relapse of SARS.

References

1. Zhang, Gongxu, 1984, "Cause-Selecting Control Chart, Theory and Practice", Beijing Publisher of People's Telecommunications, written in Chinese. This book had been awarded the Chinese national Scientific and Technological Prize with the scientific program of "Control Charts Diagnosis" together in 1987.
2. Zhang, Gongxu, 1992, "Cause-Selecting Control Chart and Diagnosis", The Aarhus Graduate School of Business, Aarhus, Denmark, Book No: ISBN 87-89695-08-9, written in English.
3. Zhang, Gongxu, 2001, "Diagnostic Theory with Two Kinds of Quality and Its Applications", Beijing Sciences Publisher, written in Chinese. This book is published under the support of Chinese National Fund of Natural Science.
4. Sun, Jing, 2001, "Quality Control of Near-Zero Non-conformity Process", Publisher of Tsinghua University, written in Chinese. This book is published under the support of Chinese National Fund of Natural Science.