



# **Detecting changes in X-ray exposure using quality control methods**

**Author: Aijiao Gu**

**Supervisor: Lars Rönnegård**

D-level essay in statistics , June, 2010

School of Technology & Business Studies, Dalarna University

---

# Detecting changes in X-ray exposure using quality control methods

Author: Aijiao Gu

Supervisor: Lars Rönnegård

June, 2010

## Abstract

Radiography is a very popular test method in modern medicine. X-ray diagnostics provide X-ray images of the inside structure of some human body parts for doctor to diagnose. The digital does level of X-ray machines can be fixed by doctor or by the machine automatically. As time goes by, the X-ray machine may show abnormal results, which are reflected in the changes of exposure index. The changes of exposure index will affect the quality of the X-ray images if too low and too high doses may harm the patients. The aim of this thesis is to use quality control methods to detect changes in exposure index caused by errors in the X-ray machines.

The paper consists of six sections: introduction, model and method, data, analysis and result, conclusion and advice, and discussion. The background part introduces the history and development of radiology, how the X-ray images produced and the previous study of technical information in digital X-ray images. In second section, it shows the theory of quality control. In section 3, I describe the data and explain how the data processed and selected. Model results and control charts are presented in section 4. In the fifth section, the practical significance of the results is explained and some practical advice is given. In the final section, the need of future research is discussed. I get the conclusion that relatively small changes can be detected by using the rule of “9 consecutive points on the same side of the centerline” in a control chart. And doctor need to make periodic and adjust for the machines to detect the change in time.

Keywords: Quality Control, Sensitivity, Linear Model, X-ray images

---

## Contents

1. Introduction .....	3
1.1 Background .....	3
1.2 Aim .....	4
2. Data .....	4
2.1 The Description of Data .....	4
2.2 Data Process and Selection .....	6
3. Model and Method .....	8
3.1 Quality Control .....	8
3.2 Linear Model .....	11
3.3 Autocorrelation and ACF .....	12
4. Analysis Process and Result .....	12
4.1 Preliminary Analysis .....	12
4.2 Analysis for One machine .....	14
5. Conclusion and Practical Advice .....	18
5.1 Conclusion .....	18
5.2 Practical Advice .....	18
6. Discussion .....	18
6.1 About the Chosen Data .....	18
6.2 About the Linear Model .....	19
6.3 About the Period .....	19

---

# 1. Introduction

## 1.1 Background

Radiography started as X-rays was discovered by Wilhelm Conrad Röntgen in November 1885. X-ray was found to belong to a kind of electromagnetic radiation (Mould, 1993). Soon after that, radiography was first put into use in the medical field. As early as the First World War, radiography had been used to treat wounded soldiers by Marie Curie. X-rays diagnostic was begun so shortly after its discovery partly because the danger of radiation had not been discovered. As development and request of technology has increased, the application of radiography has been extended to many fields, such as medicine, non-destructive testing, food inspection, security and archeology and so on. Nowadays, X-rays is more and more popular used as a radiology technology for medical application in the world and X-rays diagnostic has become the second most commonly used medical tests.<sup>1</sup>

Radiography is a technique, taking advantage of X-rays to view inside objects. X-ray, produced by an X-ray generator, is toward the objects, or some part of the body. Because the density and composition of the different body parts is different, some X-rays stop and some X-rays pass through. The X-rays passing through are captured by the detector behind the object. Then the “shadow” of the exposing object is present on the detector surface. The “shadow” can be converted into visible light, which is sent out from the detector material by scanning a red laser over the detector surface. The visible light is collected by a photomultiplier tube through a light guide. Finally, the light is detected by a photo diode captured on photographic film in some ways. Then the X-rays image is formed. (Personal communication, Hans-Erik. Kallman)

The radiology departments in the county of Dalarna provide service to 280,000 people from 4 hospitals and 7 primary healthcare units. Digital patient records were introduced in 1993 and have been fully digital since May 2001. Since then, a lot of effort has been spent to implement a countywide RIS (Radiology Information System) which plays an important role in providing a big database for statistical analysis.

---

<sup>1</sup> <http://en.wikipedia.org/wiki>

The thesis Statistical Analysis of Technical Information in Digital X-ray Images (Lu & Sun, 2009), also investigated digital X-ray images. That thesis adopted linear regression to get the relationship between the dose exposure index (or equivalently the “sensitivity”) and other variables such as study date, age group, series description and the equipment. It showed how the sensitivity was affected by those variables.

## 1.2 Aim

As time goes by, some changes could happen to some parts of the machine. For example, the important part is the structural detector material. The detector consists of a chemical substance that suffers structural deformation with every photon that hits. And a decrease in detector dose is represented by an increase in the sensitivity value. According to the image generating process, besides the detector material, there are other parts, such as the laser diode, opacification of the light guide, the photomultiplier tube and so on.

The aim of this thesis is to use quality control methods to detect if changes in exposure index are caused by changes in the X-ray machines.

## 2. Data

### 2.1 The description of data

I use the same data as in the previous study of Lu & Sun (2009), which is from the digital image plate (Fuji) system at the radiology department of Avesta hospital. The data contains 133416 observations from 2005-09-02 to 2009-03-29. The variables are summarized in Table 1.

Table 1: Summary of the Fuji digital image plate data from Avesta hospital. Numbers of observations are given for each variable.

<b>studydate</b>	<b>studydescription</b>	<b>seriesdescription</b>	<b>patientsbirthdate</b>	<b>stationname</b>
2006-04-24: 339	LUNG: 26284	FRONTAL: 41532	1921-04-14: 100	ru16220071: 29003
2006-10-09: 286	LUMBAR: 11460	SIDE: 32885	1933-05-06: 91	ru16220086: 76971
2006-01-09: 278	KNEE-JOINT DX: 8012	INSIDE WRIGGLE: 5364	1929-06-20: 86	ru16220089: 27441
2006-06-26:	CERVICAL SPINE:	TORSION:	1925-09-13:	wmorrngmd004:

---

277	7605	3919	84	1
2007-03-28:	KNEE-JOINT SIN:	STANDING	1933-01-14:	
252	7548	HEAD: 3179	80	
2006-02-27:	UROGRAPHY:	L5 FRONTAL:	1944-02-09:	
251	5463	2877	74	
(Other):	(Other):	(Other):	(Other):	
131733	67044	43660	132901	

---

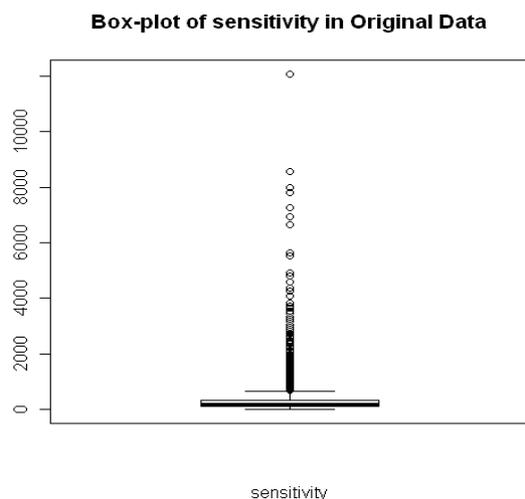


Fig. 1: Box-plot of the variable sensitivity in the Original Data

Here I explain the variables used in the analysis:

$y$ =sensitivity. Sensitivity= $c$ /Exposure, where  $c$  is a constant and Exposure is radiation exposure level reached the detector. The radiation exposure level is fixed manually or automatically, which depends on the machine, the patient, the body part, the exposure angle and so on. Sensitivity is an indicator of radiation exposure and is the response variable of interest

$x_1$ =studydate. This variable is defined as the date of each examination for the patient, in other words, it is the recording date of each observation. The range of this variable is from 2005-09-02 to 2009-03-29.

$x_2$ =studydescription. This variable is the description of the body part of the patients for examination, including LUNG, LUMBAR, CERVICAL SPINE, and so on.

$x_3$ =seriesdescription. This variable describes the X-rays exposure angles, from which the patient was examined. For example, seriesdescription FRONTAL means that the X-rays are in front of the patient and the detector is at the back. Seriesdescription includes FRONTAL, SIDE, STANDING HEAD, TORSION, and so

on.

$x_4$ =patientbirthdate. This variable, the birth date of patient, is in form of year-month-date.

$x_5$ =stationname. This variable is the name of the different equipment. There are ru1620071, ru1620086, ru1620089 and wmorrtgmd004, in which wmorrtgmd004 only has one observation so I delete it. And define  $x_{51}$ = ru1620071,  $x_{52}$ = ru1620086,  $x_{53}$ = ru1620089

## 2.2 Data process and Selection

In my thesis, I quote the estimated result of linear model in previous study (Lu & Sun, 2009), so the beginning of data process and selection is the same. I just make a simply summary of main points.

From the whole data, because the sensitivity is concentrate in [0, 1500], restricting the sensitivity in this interval will avoid some outlier's effects. Since the data is more stable after 2007-03-01, the data from 2007-03-01 to 2008-12-31 is taken as Reference Data for estimating linear model, while the data from 2009-01-01 to 2009-03-31 is taken as New Data. To get a relatively stable series data with a better normal distribution, the doctor suggests that the sensitivity interval [200,800] for reference data and the sensitivity interval [200,900] is a better choice, because the values of sensitivity concentrate in those intervals.

In the analysis, the study date  $x_1$  is the number of days since 2009-01-01 in Reference Data, while for New Data it is the number of days since 2009-04-01.

The variable,  $x_4$  = (patientbirthdate), is divided into four groups according to the age of the patient. The groups are baby (two years or younger), child 3 to 15 years old, adult 16 to 65 years old, elder (more than 65 years old). So I define  $x_{41}$ =baby,  $x_{42}$ =child,  $x_{43}$ =adult,  $x_{44}$ =elder.

In order to simplify the analysis, I choose one group of variables out of the whole data which the doctor is more interested in.

The chosen variables are as follows:

$x_1$ =studydate, from 2007-03-01 to 2009-03-29

$x_2$ =LUNG

$x_{31}$ =FRONTAL,  $x_{32}$ =SIDE

$x_{41}$ =Baby,  $x_{42}$ =Child,  $x_{43}$ =Adult,  $x_{44}$ =Elder

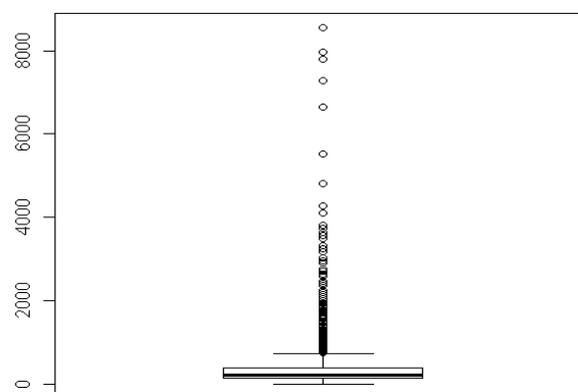
$x_{51}$ = ru1620071,  $x_{52}$ = ru1620086,  $x_{53}$ = ru1620089

The data of the chosen explanatory variables was analyzed in two parts. The data from 2007-03-01 to 2008-12-31 will be referred to as the “Reference Data” and the data from 2009-01-01 to 2009-03-29 will be referred to as the “New Data”. Reference Data contains 11428 observations (Table 2), whereas the New Data contains 1636 (Table 3).

Tab. 2: Reference Data of Chosen Explanatory Variables

<b>studydate</b>	<b>patients birthdate</b>	<b>stationname</b>	<b>series description</b>	<b>study description</b>
2007-4-25	1921-4-14	ru16220071	FRONTAL	LUNG:
59	57	48	6046	11428
2008-5-6	1932-3-26	ru16220086	SIDE:	
46	25	9089	5382	
2008-5-12	1959-1-28	ru16220089		
46	24	2291		
2007-3-27	1922-3-26			
43	22			
2007-5-31	1950-9-14			
43	22			
2008-2-25	1923-6-27			
43	21			
(Other):	(Other):			
11148	11257			

Box-plot of sensitivity in Reference Data



sensitivity

Fig 2: Box-plot of sensitivity in Reference Data of Chosen Explanatory Variables

Tab. 3: New Data of Chosen Explanatory Variables

<b>sensitivity</b>	<b>studydate</b>	<b>patients birthdate</b>	<b>stationname</b>	<b>series description</b>	<b>study description</b>
Min.: 200.0	2009-1-12 40	1936-3-14 18	ru16220071: 15	FRONTAL: 855	LUNG: 1636
1st Qu.: 449.0	2009-1-15 38	1936-8-2 12	ru16220086: 1325	SIDE: 781	
Median: 565.0	2009-2-11 38	1921-5-23 8	ru16220089: 296		
Mean: 555.5	2009-1-20 37	1923-11-25 8			
3rd Qu.: 664.0	2009-1-26 37	1947-9-10 8			
Max.: 895.0	2009-2-24 35	1933-6-7 7			
	(Other): 1411	(Other): 1575			

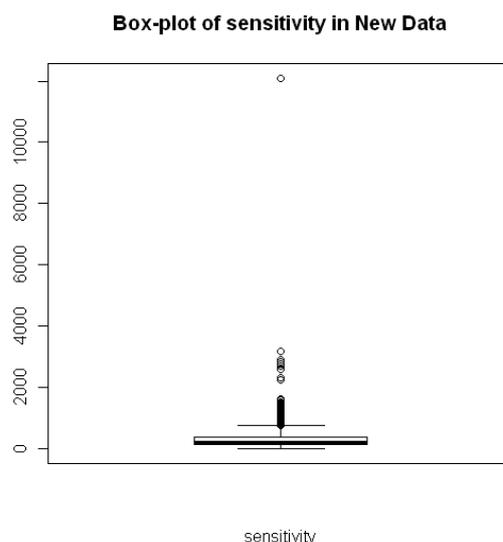


Fig. 3: Box-plot of sensitivity in New Data of Chosen Explanatory Variables

### 3. Model and Method

#### 3.1 Quality control

##### Quality

Definition1: Quality means fitness for use.

The fitness for use includes two aspects: quality of design and quality of conformance (Montgomery, 2001). Quality of design is an appropriate technical term to intend the

product produced in various levels of quality. Quality of conformance reflects how the product conforms to the specification required.

Definition2: Quality is inversely proportional to variability.

Here, the variability is only referred to harmful variability, which will bring bad effect. If the variability in the important characteristics of a product happened or increased, the quality of the product will decrease.

### **Quality control**

Quality characteristics are parameters, which describe the elements of the product quality (Montgomery, 2001). There are three types of quality characteristics: physical, sensory, time orientation. In this article, I focus on the third one, time orientation, which means reliability, durability and serviceability. And try to find when the product, here refer specially to X-rays machine; will be not reliable and serviceable any more.

Quality control is a process, in which all factors have effect on the quality of products in the production, and will be detected to make sure the process in control.

Here, I use a statistical tool, control charts, in quality control to solve the problems above.

### **Control chart**

Definition: Control charts are statistical tools that monitor a process and alert us when the process has been disturbed so that it is now out of control (Moore & McCabe, 2005). Plot control charts assume identical and independent observations. This is a signal to find and correct the cause of the disturbance.

$\bar{x}$  Control chart

Given mean  $\mu$ , standard deviation  $\sigma$  and sample size  $n$ , to evaluate the control of a process, plot a  $\bar{x}$  control chart is defined as follows:

Plot the means  $\bar{x}$  against time.

Draw a horizontal center line at  $\mu$

Draw horizontal control limits at UCL and LCL,  $UCL = \mu + 3\sigma/\sqrt{n}$  and  $LCL = \mu - 3\sigma/\sqrt{n}$

Any  $\bar{x}$  that falls between control limits means the process is in control. This is an example of control charts:

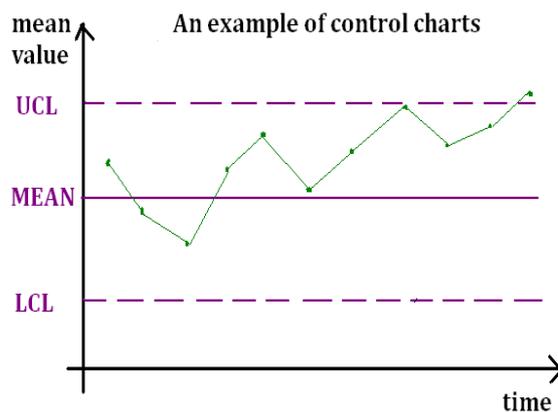
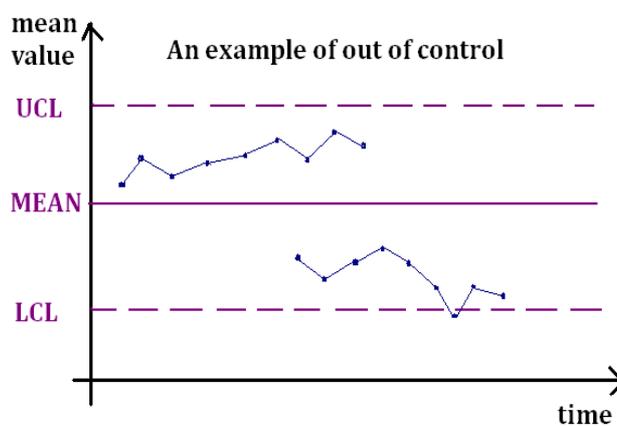


Fig. 4: Simple example of control charts

The last point of Fig4 shows the process is out of control. Besides according to individual points beyond the control limits to reflect out of control, there are other evidences for lack of control, for example, a run of 9 consecutive points on the same side of centerline. The probability of a false signal of this method is  $2\left(\frac{1}{2}\right)^9 = 0.004$ , when actually the process is in control. Because the probabilities of the run of 9 points are equal, which are above or below the centerline, and the probability of the run above the centerline is  $\left(\frac{1}{2}\right)^9 = 0.002$ . This method indicates the trend of the out of control process. When the values of observations slowly shift away from the centerline of mean, the run signal method will give the sign sooner than the individual points out of control limits. That is why it is common practice to use the two methods simultaneously (Moore & McCabe, 1998). An example is shown in Figure 5.

Fig. 5: A control chart of out of control



### Control Limits and Three -sigma rule

Firstly, I introduce two types of error:

Type 1<sup>st</sup> error: The points fall out of the control limits when the process is in control.

Type 2<sup>nd</sup> error: The points fall between the control limits when the process is out of control.

Moving the control limits closer leads to the risk of type 1<sup>st</sup> error increased and the risk of type 2<sup>nd</sup> error decreased. On the contrary, moving the control limits farther, leads to the risk of type 1<sup>st</sup> error decreased and the risk of type 2<sup>nd</sup> error increased. So choosing the proper control limits is very important. It has been justified that three-sigma control limits can give good results in practice.

Three-sigma rule is related in the control charts. The rule is usually used to get a rough estimate of probability and test outliers. For a normal distribution, nearly 99.73% of values fall into the interval  $[\mu - 3\sigma, \mu + 3\sigma]$ . If the process is in control, the probability of the next point beyond the control limits is 0.27%. With such a point, out of control is significant. This criterion is good enough to test outliers in control charts. The details of three-sigma rule are shown in Figure 6.

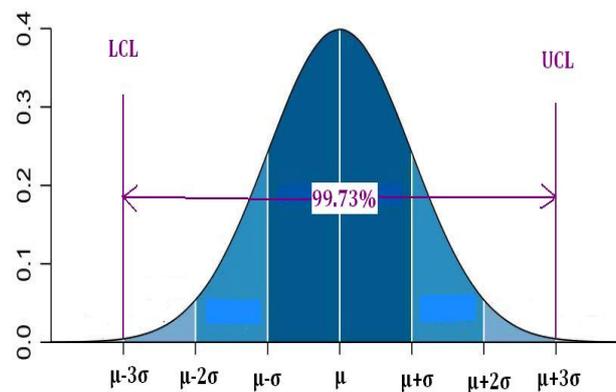


Fig. 6: The Three-Sigma Rule

### 3.2 Linear Model

Definition (Faraway, 2005): in a general linear model, the observed value of the dependent variable  $y$  for observation number  $i$  ( $i=1,2,\dots,n$ ) is modeled as a linear function of  $(p-1)$  so called independent variables  $x_1, x_2, \dots, x_{p-1}$  as

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{i(p-1)} + \varepsilon_i$$

Where  $y_i$  is the response variable,  $x_i$  are explanatory variables,  $\beta_0$  is an intercept

term,  $\beta_i$  is a regression coefficient,  $\varepsilon_i$  is a random error term. Generally, we assume that the random error term an identical and independent (i.i.d) normal distribution. So it is important to check that the data is approximately normal, that there is no autocorrelation and no heteroscedasticity. To study these assumptions I use QQ plot, estimate an autocorrelation function, and fit a model with heteroscedasticity using a double generalized linear model with the `dglm` function in R.

### 3.3 Autocorrelation and ACF

Autocorrelation, also called “lagged correlation”, refers to the correlation of values at different time points in time series.

ACF, Autocorrelation Function, is one of the tools for assessing the autocorrelation of a time series.

Definition (Shumway & Stoffer, 2006):

The autocorrelation function (ACF) is defined as  $\rho(s, t) = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(s, t)}}$ , where autocovariance function  $\gamma(s, t) = E[(x_s - \mu_s)(x_t - \mu_t)]$ .

It is obvious that  $-1 \leq \rho(s, t) \leq 1$ . The ACF measures the linear predictability of the series at time  $t$ , say,  $x_t$ , using only the value  $x_s$ .

## 4. Analysis Process and Result

The control chart is obtained by plotting the sensitivity against study date. So the first step is to eliminate the effect of variables. The solution is the true sensitivity minus the predicted sensitivity as Corrected Sensitivity based on the New Data from 2009-01-01 to 2009-03-31. The predicted sensitivity is calculated by estimated linear model with Reference Data from 2007-03-01 to 2008-12-31.

Define

$y_n$  = the observed sensitivity of New Data

$\widehat{y}_n$  = the predicted sensitivity of New Data using estimates from Reference Data

CS = Corrected Sensitivity with New Data

i.e.  $CS = y_n - \widehat{y}_n$

### 4.1 Preliminary Analysis

### The estimated linear model

This model part I quote the model and the result of the previous study thesis (Lu & Sun, 2009). The model form with sensitivity as response is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon$$

The model was fitted using R with the Reference Data (Faraway, 2005), and then the estimated linear model was:

$$y = 429.19 - 0.07x_1 + 89.99x_{42} + 175.90x_{43} + 167.10x_{44} - 93.49x_{32} + 16.26x_{52} - 60.02x_{53}$$

$x_1$  = studydate, from 2007-03-01 to 2008-12-13

$x_{32}$  = SIDE

$x_{42}$  = Child,  $x_{43}$  = Adult,  $x_{44}$  = Elder

$x_{52}$  = ru1620086,  $x_{53}$  = ru1620089

Using this model, a time series of CS is obtained with New Data.

### Check assumption: Heteroscedasticity

Plot control charts assume identical observations, to investigate this assumption I fit a generalized linear model with a log link for the dispersion (McCullagh & Nelder, 1989). Using dglm method in R, I get the estimated result (Tab. 4)

Tab. 4: the Estimated Result of Dispersion Model

Variable	Estimate Coefficient	Standard Error	Z-value	P-value
Intercept	10.93	0.38	28.58	$1.08 \times 10^{-179}$
Studydata	$3.17 \times 10^{-6}$	$6.77 \times 10^{-5}$	0.05	0.96
Child	-0.02	0.34	-0.06	0.95
Adult	-0.81	0.33	-2.49	0.01
Elder	-0.83	0.33	-2.55	0.01
FRONTAL	-0.67	0.03	-25.44	$8.60 \times 10^{-143}$

ru16220086	-0.34	0.20	-1.70	0.09
ru16220089	-0.85	0.20	-4.23	$2.38 \times 10^{-5}$

Most p-values are smaller than the criteria except the variables of a station name and the child group. It indicates serious heteroscedasticity in this model. To eliminate the effect of heteroscedasticity, so I decide to do analysis with one series description and each machine, and focus on the ru1620086 machine and the FRONTAL description to do detailed analysis which includes more observations.

## 4.2 Analysis for One machine

### The estimated linear model

The model linear form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

was estimated using R (Verzani, 2005) with Reference Data. For the chosen data for machine ru1620086 and the FRONTAL description, the estimated linear model was:

$$y = 316.01 - 0.085x_1 + 145.15x_{42} + 296.15x_{43} + 326.24x_{44}$$

$x_1$ =studydate, from 2007-03-01 to 2008-12-13

$x_{42}$ =Child,  $x_{43}$ =Adult,  $x_{44}$ =Elder

From the estimated result of linear model, the residual standard error is 93.06. Define  $\sigma = \widehat{\sigma}_e = 93.06$

Using this model, a time series of CS is obtained with the New Data.

### Check assumption: Heteroscedasticity

Plot control charts assume identical observations, to investigate this assumption I fit a generalized linear model with a link-linear model for the dispersion. Using dglm method in R, I get the estimated result (Tab. 5)

Tab. 5: the Estimated Result of Dispersion Model

Variable	Estimate	Standard	Z-value	P-value
	Coefficient	Error		
Intercept	9.1314	0.5358	17.0441	$3.8678 \times 10^{-65}$

Studydata	$7.8024 \times 10^{-5}$	0.0001	0.7421	0.4580
Child	0.3532	0.5529	0.6388	0.5230
Adult	-0.0301	0.5353	-0.0562	0.9552
Elder	-0.138	0.535	-0.2577	0.7966

All p-values are larger than the criteria 0.05 , which indicates no heteroscedasticity.

### Check assumption: Normality

When plot control charts, three-sigma rule is under normal distribution. To investigate this assumption a QQ plot was made for Corrected Sensitivity with New Data (Fig. 7).

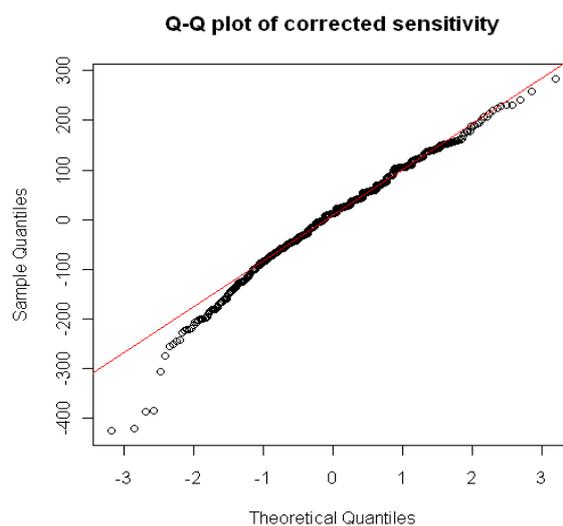


Fig. 7: Q-Q Plot of Corrected Sensitivity

On the whole, CS obeys the normal assumption (Fig. 7). But there are 4 points at the ends of the line far from the line obviously. The 4 outliers do not follow the whole normal distribution maybe because of random effect.

### Check assumption: Autocorrelation

With R(Shumway & Stoffer, 2006), plot control charts assume independent observations, to investigate this assumption I make the plot of the ACF and PACF of CS series for New Data using R (Shumway & Stoffer, 2006) (Fig. 8).

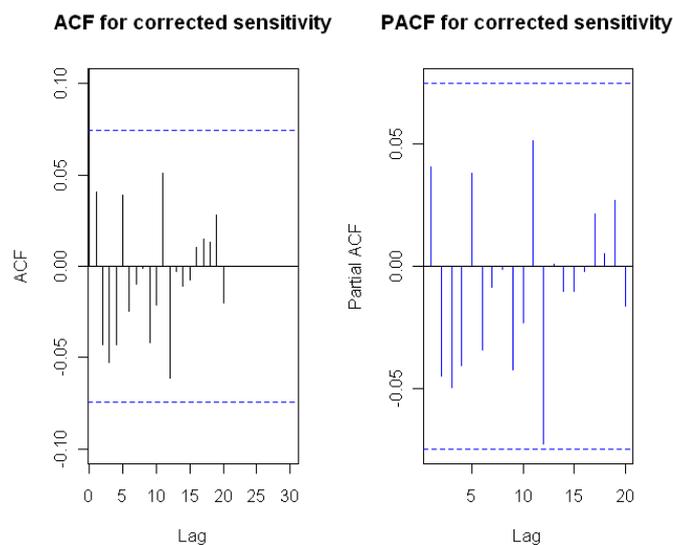


Fig. 8: The ACF and PACF of Corrected Sensitivity

The autocorrelation does not exist in Corrected Sensitivity series, which obeys independent assumption well.

### Control Chart

This is the control chart of Corrected Sensitivity for New Data.

Mean (CS) =5.9704 , UCL= Mean (CS) + $3\sigma$  = 285.1504, LCL= Mean (CS)- $3\sigma$ = -273.2096, where  $\sigma$  I used to plot control chart here is the residual standard error of linear model with Reference Data.

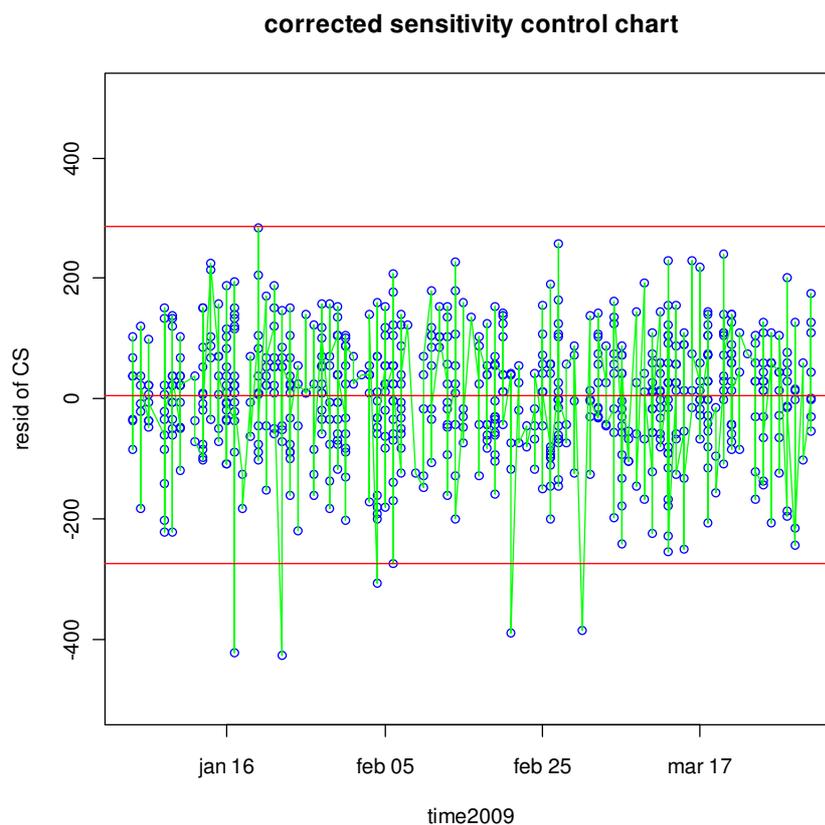


Fig. 9: The Control Chart of Corrected Sensitivity

### Final Analysis Result

There are 6 observations beyond 3-sigma limits (Fig. 9) 692 observations. There are 4 observations out of 6 points far away from the lower control limit. The study date of observation below the LCL are "2009-01-17" "2009-01-23" "2009-02-21" and "2009-03-02", these observations are the same as the 4 points in the left bottom corner of the Q-Q plot (Figure 7). The proportion of points out of control is 0.867% which is only larger than 0.27%.

There was no chain of 9 consecutive points on the same side of centerline and it is concluded that the process is not out of control.

In summary, with the analysis in two above methods, the process is in control, in other words, the machine is normal with no changes.

### Check the Method

To check how good the method is in detecting changes, a value of 20 was added to Corrected Sensitivity and analyze the control chart of new CS. Define  $CS1=CS+20$ .

The control chart of the new CS is obtained.

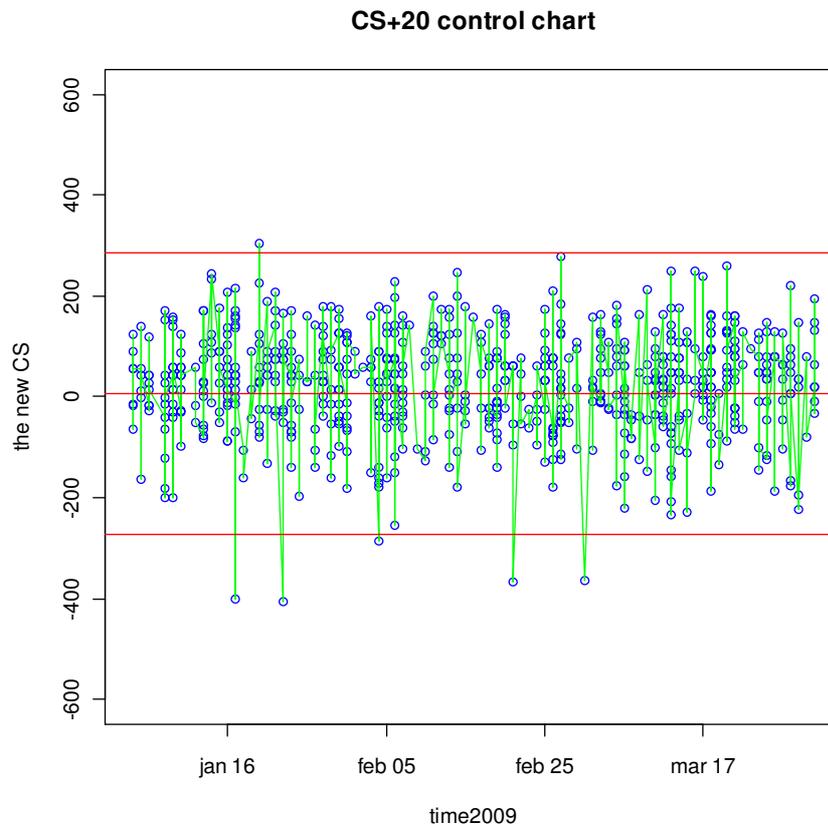


Fig. 10: The Control Chart of CS+20

There are still 6 points of the new residual out of 3-sigma control limits (Fig. 10), which are the same as the previous. The proportion of points out of control is 0.867% which is only larger than 0.27%. But there is a run of 9 consecutive points on the same side of centerline. In this case, a process out of control was simulated and it can be detected with sufficient evidence. The run of 9 consecutive points appears from "2009-03-17" to "2009-03-18", and New Data is recorded from "2009-01-01". So for this specific scenario it takes the doctors more than two months and a half to detect the change of the machine. The change of the machine can be detected with changing by 20, which means that this is a good method to resolve this problem.

## 5. Conclusion and Practical Advice

### 5.1 Conclusion

This thesis presents a method to detect X-ray machines giving exposure results out of

---

control.

I focus on detecting the machine ru1620086. With the estimated linear model and New Data, a relatively stable series data Corrected Sensitivity is obtained, which obeys all the assumptions. There was no chain of 9 consecutive points on the same side of centerline in the control chart. The conclusion from the analysis is that there was nothing wrong with the machine ru1620086 during the first three months of 2009. Relatively small changes can be detected by using the rule of “9 consecutive points on the same side of the centerline” in a control chart. Even though quality control chart is a good method to detect the changes, it still takes more than two month to discover the changes.

## **5.2 Practical Advice**

According to the analysis above, my advice is make periodic detect and adjust the departments of the machine, which change as time goes by. Such as some departments I have referred at the thesis part of aim, the structural detector material, the laser diode, opacification of the light guide, the photomultiplier tube and so on. It is absolutely required to make a proper periodic detect and adjust the machines, so that the changes can be detected and get the machine in control in time. This is the final aim of the quality control method. As the machine gets older, the probability of changes gets larger, so the fixed period depends on the individual machine.

## **6. Discussion**

### **6.1 About the Chosen Data**

Because I adopt the same method of chosen data as the master thesis of Lu & Sun (2009), we study the same problem. The patients are grouped by age, but, there are some differences in the character of the patients' body part in the same group. For example, adult's bone density is very different between the age of 16 and the age of 65. And the patients at the limit ages from the different groups are similar in the character of body part. Adult's bone character are similar to elders' when they are 65-year-old. Because the girls are relatively earlier than boys, their body part structure and density are different even though at the same group. Thus, if we could have a

---

more appropriate and better group method, there will be a more accurate result.

## **6.2 About the Linear Model**

Radiology Information System has introduced new variables in the database nowadays, so a new estimated linear model is required to be done when this method is put into practice. And the linear model I used is the simplest model which is also the most widely used both in theory and in practice. We also can consider using generalized linear model (GLIM), linear mixed model (LMM) or generalized linear mixed model (GLMM) with some random effects ( McCulloch & Searle, 2001). Some of these methods do not require strict assumptions, so maybe they are appropriate choices to use. And some of these models with random effect model may fit better. I expect to solve the practical problems with a simple approach, so there are still a lot of further researches to do on this topic.

## **6.3 About the Period**

After the machine becomes abnormal, it takes more than two months to discover the changes, in which period doctors have to wait for enough observations to draw a conclusion. Because the machines really have something wrong, the machine may have a harmful affect on patients during that period. Therefore, considering the health of the patients and the accurateness of diagnostic, it is better to find some ways to improve this method so that the period gets shorter.

## **References**

[1] Faraway J.J. (2005), *Linear Models with R* (Texts in Statistical Science), Chapman & Hall/CRC Press.

- 
- [2] Lu X. & Sun K. (2009), *Statistical Analysis of Technical Information in Digital X-ray Images*.
- [3] Moore D.S. & McCabe G.P. (1998), *Introduction to the Practice of Statistics*, 3<sup>th</sup> ed., New York: W. H. Freeman.
- [4] Montgomery D.C. (2001), *Introduction to Statistical Quality Control*, 4<sup>th</sup> ed., New York: Wiley.
- [5] Mould R.F. (1993), *A Century of X-rays and Radioactivity in Medicine: with Emphasis on Photographic Records of the Early Years*, 2nd ed., Institute of Physics Publishing, Bristol, UK.
- [6] McCulloch C. E. & Searle S. R. (2001), *Generalized, Linear, and Mixed Models*, New York: Wiley.
- [7] McCullagh P. & Nelder J. A. (1989), *generalized linear models*, 2nd ed., London: Chapman & Hall.
- [8] Pawitan Y. (2001), *In All Likelihood: statistical modelling and inference using likelihood*, Oxford: Clarendon.
- [9] Shumway R.H. & Stoffer D.S. (2006), *Time Series Analysis and Its Applications with R examples*, 2<sup>nd</sup> ed., Springer.
- [10] Thaga K., *Control Chart for Autocorrelated Processes with Heavy Tailed Distributions*, 197-206.
- [11] Verzani J. (2005), *Using R for Introductory Statistics*, Chapman & Hall/CRC Press.

## **Appendix A**

R code:

---

```
#####DATA SUM#####
```

```
indat<-read.table(file="i:/Fuji.txt",header=T,sep = ",")
summary(indat)
indat<-indat[-c(44212,44213,112196,112288,122687),]
indat1<-indat[indat$studydescription=="LUNGOR",]
indat2<-indat1[indat1$seriesdescription=="SIDA"indat1$seriesdescription=="FRON
TAL",]
indat3<-indat2[indat2$stationname=="ru16220086"indat2$stationname=="ru162200
71"indat2$stationname=="ru16220089",]
indat4<-indat3[indat3$sensitivity<=1500&indat3$sensitivity>=0,]
indat5<-indat4[indat4$sensitivity<=800&indat4$sensitivity>=200,]
indatr<-indat5[as.Date(indat5$studydate)>"2007-3-01"&as.Date(indat5$studydate)<"
2009-1-01",]
summary(indatr)
boxplot(indatr$sensitivity,which="inner",lty="solid",xlab="sensitivity",main="Box-pl
ot of sensitivity in Reference Data")
indat6<-indat4[indat4$sensitivity<=900&indat4$sensitivity>=200,]
indatn<-indat6[as.Date(indat6$studydate)>"2009-01-01",]
summary(indatn)
boxplot(indatn$sensitivity,which="inner",lty="solid",xlab="sensitivity",main="Box-p
lot of sensitivity in New Data")
#####ESTIMATED MODEL#####
indat<-read.table(file="i:/Fuji.txt",header=T,sep = ",")
summary(indat)
indat<-indat[-c(44212,44213,112196,112288,122687),]
indatr<-indat[as.Date(indat$studydate)>"2007-3-01"&as.Date(indat$studydate)<"200
9-1-01",]
indatn<-indat[as.Date(indat$studydate)>"2009-01-01",]
pdate<-as.Date(indatr[,5])
ed<-c("2009-01-01")
```

---

```
edate<-as.Date(ed)
age<-difftime(edate,pdate,units="days")
age<-as.numeric(age)
age<-age%%/%365
indatr[,5]<-age
indatr[,5][indatr[,5]<=2]<-1
indatr[,5][indatr[,5]>2&indatr[,5]<=15]<-2
indatr[,5][indatr[,5]>15&indatr[,5]<=65]<-3
indatr[,5][indatr[,5]>65]<-4
sdate<-as.Date(indatr[,2])
ed<-c("2009-01-01")
edate<-as.Date(ed)
diffdate<-difftime(edate,sdate,units="days")
diffdate<-as.numeric(diffdate)
indatr[,2]<-diffdate
lungor<-indatr[indatr$studydescription=="LUNGOR",]
lungorfs<-lungor[lungor$seriesdescription=="SIDA"|lungor$seriesdescription=="FR
ONTAL",]
lungorfs1<-lungorfs[lungorfs$stationname=="ru16220086"|lungorfs$stationname=="r
u16220071"|lungorfs$stationname=="ru16220089",]
lungorfs1<-lungorfs1[lungorfs1$sensitivity<=1500&lungorfs1$sensitivity>=0,]
lungorfs2<-lungorfs1[lungorfs1$sensitivity<=800&lungorfs1$sensitivity>=200,]
lm<-lm(sensitivity~studydate+factor(patientsbirthdate)+seriesdescription+stationnam
e,data=lungorfs2)
summary(lm)
#####HETEROSCEDASTICITY#####

library(dglm)
lm0<-dglm(sensitivity~studydate+factor(patientsbirthdate)+seriesdescription+stationn
ame,dformula=~studydate+factor(patientsbirthdate)+seriesdescription+stationname,d
```

---

```
ata=lungorfs2)
summary(lm0)
#####ESTIMATED MODEL#####
indat<-read.table(file="i:/Fuji.txt",header=T,sep = ",")
summary(indat)
indat<-indat[-c(44212,44213,112196,112288,122687),]
indatr<-indat[as.Date(indat$studydate)>"2007-3-01"&as.Date(indat$studydate)<"200
9-1-01",]
indatn<-indat[as.Date(indat$studydate)>"2009-01-01",]
pdate<-as.Date(indatr[,5])
ed<-c("2009-01-01")
edate<-as.Date(ed)
age<-difftime(edate,pdate,units="days")
age<-as.numeric(age)
age<-age%/%365
indatr[,5]<-age
indatr[,5][indatr[,5]<=2]<-1
indatr[,5][indatr[,5]>2&indatr[,5]<=15]<-2
indatr[,5][indatr[,5]>15&indatr[,5]<=65]<-3
indatr[,5][indatr[,5]>65]<-4
sdate<-as.Date(indatr[,2])
ed<-c("2009-01-01")
edate<-as.Date(ed)
diffdate<-difftime(edate,sdate,units="days")
diffdate<-as.numeric(diffdate)
indatr[,2]<-diffdate

lungor<-indatr[indatr$studydescription=="LUNGOR",]
lungorfs<-lungor[lungor$seriesdescription=="FRONTAL",]
lungorfs1<-lungorfs[lungorfs$stationname=="ru16220086",]
```

---

```
lungorfs1<-lungorfs1[lungorfs1$sensitivity<=1500&lungorfs1$sensitivity>=0,]
lungorfs2<-lungorfs1[lungorfs1$sensitivity<=800&lungorfs1$sensitivity>=200,]
lm1<-lm(sensitivity~studydate+factor(patientsbirthdate),data=lungorfs2)
summary(lm1)
#####heteroscedasticity#####
library(dglm)
lm2<-dglm(sensitivity~studydate+factor(patientsbirthdate),dformula=~studydate+fact
or(patientsbirthdate),data=lungorfs2)
summary(lm2)
#####CONTROL CHART#####
pdate<-as.Date(indatn[,5])
ed<-c("2009-04-01")
edate<-as.Date(ed)
age<-difftime(edate,pdate,units="days")
age<-as.numeric(age)
age<-age%/%365
indatn[,5]<-age
indatn[,5][indatn[,5]<=2]<-1
indatn[,5][indatn[,5]>2&indatn[,5]<=15]<-2
indatn[,5][indatn[,5]>15&indatn[,5]<=65]<-3
indatn[,5][indatn[,5]>65]<-4
sdate<-as.Date(indatn[,2])
ed<-c("2009-04-01")
edate<-as.Date(ed)
diffdate<-difftime(edate,sdate,units="days")
diffdate<-as.numeric(diffdate)
indatn[,2]<-diffdate
lungor<-indatn[indatn$studydescription=="LUNGOR",]
lungorfs<-lungor[lungor$seriesdescription=="FRONTAL",]
lungorfs<-lungorfs[lungorfs$stationname=="ru16220086",]
```

---

```
lungorfsnew<-lungorfs[lungorfs$sensitivity<=900&lungorfs$sensitivity>=200,]
predict09<-predict(lm1,lungorfsnew)
true09<-lungorfsnew[,7]
corsensi<-true09-predict09
qqnorm(corsensi,main="Q-Q plot of corrected sensitivity")
qqline(corsensi,col="red")
#####ACF AND PACF#####
acf(corsensi,lag.max=20,type=c("correlation"),plot=FALSE,na.action=na.fail,demean
= TRUE)
pacf(corsensi,lag.max=20,plot=FALSE,na.action=na.fail)
par(mfrow=c(1,2))
acf(corsensi,lag.max=20,xlim=c(1,30),ylim=c(-0.1,0.1),type=c("correlation"),plot=T
RUE,na.action = na.fail,demean=TRUE,col="black",main="ACF for corrected
sensitivity")
pacf(corsensi,lag.max=20,plot=TRUE,na.action=na.fail,col="blue",main="PACF for
corrected sensitivity")
#####CONTROL CHARTS#####
mean(corsensi,na.rm=TRUE) ###5.970426
sigma<-93.06
length(corsensi) ###692
UCL<-mean(corsensi,na.rm=TRUE)+3*sigma
UCL###285.1504
LCL<-mean(corsensi,na.rm=TRUE)-3*sigma
LCL###-273.2096
sdate<-as.Date(lungorfsnew[,2],origin="2009-01-01")
plot(sdate, corsensi, col="blue", main="corrected sensitivity control chart",
ylim=c(-500,500), xlab="time2009",ylab="resid of CS")
lines(sdate, corsensi,col="green",main="correctedsensitivity
chart",ylim=c(-500,500),xlab="time2009",ylab="resid of CS")
abline(h=mean(corsensi,na.rm=TRUE),col="red")
```

---

```
abline(h=UCL,col="red")
abline(h=LCL,col="red")
#####INDIVIDUAL POINTS#####
sum(corsensi<LCL,na.rm = TRUE)
sum(corsensi>UCL,na.rm = TRUE)
#####NINE POINTS RULE#####
tem<-numeric(length(corsensi))
for(i in 2:length(corsensi)){
  if((corsensi[i]*corsensi[i-1])<0)
    tem[i]=0
  else
    tem[i]=tem[i-1]+1
}
max(tem)###8
#####CHECK#####
corsensi1<-corsensi+20
sdate<-as.Date(lungorfsnew[,2],origin="2009-01-01")
plot(sdate,corsensi1,col="blue",main="CS+20 control chart",ylim=c(-600,600),
xlab="time2009",ylab="the new CS")
lines(sdate,corsensi1,col="green",main="CS+20 control chart",ylim=c(-600,600),
xlab="time2009",ylab="the new CS")
abline(h=mean(corsensi,na.rm=TRUE),col="red")
abline(h=UCL,col="red")
abline(h=LCL,col="red")
sum(corsensi1<LCL,na.rm = TRUE)###5
sum(corsensi1>UCL,na.rm = TRUE)###1

tem1<-numeric(length(corsensi1))
for(i in 2:length(corsensi1)){
  if((corsensi1[i]*corsensi1[i-1])<0)
```

---

```
    tem1[i]=0
else
    tem1[i]=tem1[i-1]+1
}
max(tem1)#####10#####
sum(tem1==9)#####1#####
which(tem1==9)#####586#####
rundate<-c(sdate[578],sdate[586])
rundate#####"2009-03-17" "2009-03-18"#####
```