ABSTRACT

A SAS® macro is used to perform a quality control analysis of an advanced ceramics production line for Imperial Chemical Industries (ICI) Ceramics in Auburn, California. XBAR and RANGE charts were used to identify possible problems with process performance. The supplier gave certain bounds for the lower and upper limits of the ¹⁰Boron (¹⁰B) isotope within the boron carbide loadings (B₄C). It was these ¹⁰B loadings that were of interest, since it was questionable whether or not impurities had caused the percentage of the ¹⁰Boron Isotope to be outside the limits specified by the supplier. Boron has two isotopes ¹⁰B and ¹¹B. Capability indices are calculated. A second SAS macro is provided to estimate the out-of-spec probability. The Bootstrap is the simulation approach used to estimate this 'out-of-spec' probability. This is the probability that the control process is not performing within the lower and upper control limit bounds. The limits are either specified by the user or set equal to 3 standard deviations below and above the process mean. The macros require base SAS, SAS/STAT, SAS/QC and SAS/GRAPH software to run.

INTRODUCTION

A statistical quality control and capability analysis is carried out (SAS/QC, 1992). The data used is from an advanced ceramics production line of ICI Ceramics and consists of a set of measurements from five assemblies from the manufacturing line. The ¹⁰B loadings are calculated. The specification limits are set equal to:

\[ \text{LSL} = \bar{x} - 3s / \sqrt{n} \]

\[ \text{USL} = \bar{x} + 3s / \sqrt{n} \]

The bootstrap method (see Efron and Tibshirani, 1993) calculates bootstrap probabilities of the process falling outside lower and/or upper control limit bounds. The data used for the bootstrap estimate of the out-of-spec probability is a vector of 50 observations generated from a Beta distribution with parameters \( r = 1 \) and \( s = 3 \) (see Ciarlini et al., 1999).

PROCEDURE

USL and LSL condition

These are the upper and lower specified limits and either are:

a) set equal to specified limits (see Johnson, 2000); or
b) set equal to: \( \bar{x} \pm 3s / \sqrt{n} \) (see SAS Macro #1 for code).

For the ICI data analyzed, \( n = 5 \). In general 'n' is the number of observations obtained for each sample run. The standard deviation is estimated using the MSE from an ANOVA based upon the \( n \) subgroups (see SAS/STAT, 1993). A Runs test plot (see SAS/Graph, 1991) is produced (see SAS Macro #1 for code). A Runs test is a way to check the control chart for unnatural patterns.

Note for the Zone A, B and C charts of a Runs test, 1-s, 2-s and 3-s are used for the lower and upper control limit bounds.

Computational Formulae and Constants Used:

\[ d = (\text{USL} - \text{LSL}) / 2 \]

\[ m = (\text{USL} + \text{LSL}) / 2 \]

\[ T = \text{target value based on engineering determination and specified by the user} \]

\[ \bar{x} = \text{overall mean of process} \]

\[ s = \text{standard deviation of process, estimated 3 ways as above} \]

\[ g = \left[ \bar{x} - m \right] d \]

\[ C_p = d / 3s \]

\[ C_m = C_p \left\{ 1 + \left( \frac{\bar{x} - T}{s} \right) \right\}^{1/2} \]

\[ C_{pm} = C_p \left\{ 1 - \frac{1}{2} \left( \frac{\bar{x} - T}{s} \right) \right\} \]

XBAR Chart:

\[ \text{CL} = \bar{x} \pm \sum \frac{x_l}{(n \times n_y - \#)} \]

Where \( n = 5 \), \( n_y \) is the number of sample runs (= 50), and \# is the number of missing observations (= 2 in the data example provided).

Either LSL and USL are specified (see Johnson, 2000) or set equal to LSL = \( \bar{x} - 3s / \sqrt{n} \) and USL = \( \bar{x} + 3s / \sqrt{n} \) (see SAS Macro #1 for code).

Range (R) Chart:

\[ \text{CL} = \bar{R} \pm \sum \frac{R_r}{n_r} \]

LSR = \( \bar{R} \times d_1 \) and USR = \( \bar{R} \times d_4 \)

Table 1. Constants Used in Constructing the R Chart

<table>
<thead>
<tr>
<th>( n )</th>
<th>( d_1 )</th>
<th>( d_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.000</td>
<td>3.267</td>
</tr>
<tr>
<td>3</td>
<td>0.000</td>
<td>2.574</td>
</tr>
<tr>
<td>4</td>
<td>0.000</td>
<td>2.282</td>
</tr>
<tr>
<td>5</td>
<td>0.000</td>
<td>2.115</td>
</tr>
<tr>
<td>6</td>
<td>0.000</td>
<td>2.004</td>
</tr>
<tr>
<td>7</td>
<td>0.076</td>
<td>1.924</td>
</tr>
<tr>
<td>8</td>
<td>0.136</td>
<td>1.864</td>
</tr>
<tr>
<td>9</td>
<td>0.184</td>
<td>1.816</td>
</tr>
<tr>
<td>10</td>
<td>0.223</td>
<td>1.777</td>
</tr>
</tbody>
</table>

Source: (see Montgomery, 1996).
When the capability index is greater than 1, the process is said to be capable. An index smaller than 1 indicates a problem. Sometimes a more conservative value of 1.33 is used since some processes produce an output that is not normally distributed (i.e., not Gaussian in distribution). Kotz and Johnson (1999) explain the relationships found between $C_p$, $C_{pk}$ and $C_{pm}$.

**Runs Test:**

A process needs to be investigated if any single point is beyond the 3-sigma control limit; or 2 of 3 successive points are found in Zone A or beyond; or 4 of 5 successive points are found in Zone B or beyond; or 8 successive points are found in Zone C or beyond; or 15 successive points are found in Zone C on either side of the centerline; or 8 successive points are found outside of Zone C.

Note for Zone A, B and C construction:

- **Zone A:** $\text{LSL}_3 = \bar{x} - 3s/\sqrt{n}$ and $\text{USL}_3 = \bar{x} + 3s/\sqrt{n}$
- **Zone B:** $\text{LSL}_2 = \bar{x} - 2s/\sqrt{n}$ and $\text{USL}_2 = \bar{x} + 2s/\sqrt{n}$
- **Zone C:** $\text{LSL}_1 = \bar{x} - s/\sqrt{n}$ and $\text{USL}_1 = \bar{x} + s/\sqrt{n}$

**Bootstrap Estimate of the Out-of-Spec Probability**

Next follows a bootstrap SAS macro (see SAS Macro #2 for code), where the bootstrap methodology is applied for the quality control of an industrial production line according to required lower (LSL) and upper (USL) specification limits. Suppose the parameter of interest is $\Theta$. A random sample, $x_1, x_2, \ldots, x_n$, is taken from the population, from which an estimate $\hat{\Theta}$, of $\Theta$, is calculated. The idea behind bootstrapping is that this sample can then be used repeatedly to find out properties (such as bias and variance) of some given estimator of $\Theta$. The original sample of $n$ observed values, each with probability $1/n$, is resampled. The sampling is done with replacement. $\hat{\Theta}^*_1$ is calculated from this new sample. Another sample is generated by resampling with replacement again from the original sample. $\hat{\Theta}^*_i$ is calculated. The procedure is repeated $B$ times, resulting in $B$ individual estimates of $\Theta$.

Of interest may be the standard bootstrap confidence interval. The standard deviation, $\sigma$, is estimated by the standard deviation of estimates of the parameter $\Theta$ that are found by bootstrap resampling of the values in the original sample of data. The interval is then $\hat{\Theta} \pm z_{1-\alpha/2}$ (Bootstrap standard deviation).

The bootstrap estimate (see Efron, 1979; 1981) of bias is calculated by subtracting the simple estimate of $\Theta$ from the overall bootstrap estimate obtained by averaging the $B$ estimates, $\hat{\Theta}^*_i$ ($i = 1, \ldots, B$), i.e.,

$$
\text{BIAS}_B^\Theta = \bar{\hat{\Theta}} - \hat{\Theta} = \frac{1}{B} \sum_{i=1}^{B} \hat{\Theta}^*_i - \hat{\Theta}.
$$

The probability of out-of-spec (see Ciartini, Gigli and Regoliosi, 1999) is defined as $P(x \notin [\text{LSL}, \text{USL}]) = F(\text{LSL}) + 1 - F(\text{USL})$, where $X$ is a random variable with probability distribution $F$.

For the $b$'th bootstrap $z^*$ is calculated via $z^* = \frac{x - \bar{x}}{s}$, and for the $b$'th bootstrap $F'(\text{LSL})$ and $F'(\text{USL})$ are calculated via:

$$
F'(\text{LSL}) = \frac{1}{n} \sum \left[ I\left( z^* \leq \frac{1}{s} \right) \right],
$$

and

$$
1 - F'(\text{USL}) = \frac{1}{n} \sum \left[ I\left( z^* > \frac{1}{s} \right) \right],
$$

where $I$ is the indicator function, $\bar{x}$ and $s$ are the mean and standard deviation of the sample $x_1, x_2, \ldots, x_n$ (see SAS Macro #2 for code).

The $B$ bootstraps are summed and averaged to obtain final estimates of the probabilities and standard errors.

$$
P_{\text{LSL}}^* + P_{\text{USL}}^* = F_{\text{LSL}}^* + 1 - F_{\text{USL}}^*.$$

Then a bootstrap approximation for the probability of out-of-spec is

$$
P_{\text{LSL}}' + P_{\text{USL}}' = F_{\text{LSL}}' + 1 - F_{\text{USL}}'.$$

**RESULTS**

**Figure 1. XBAR Chart (with LSL and USL)**

**Figure 2. Range (R) Chart (with LSR and USR)**

**Figure 3. RUNS Test with Zones A, B and C**
The example used for the quality control and capability analysis consists of 64 sample runs of 5 assembly lines ("lots"). The specification limits are set equal to $\bar{x} \pm 3\frac{s}{\sqrt{n}}$. The estimate for the standard deviation, $s$, described above is shown in Table 2.

Note for this example $LSL = \bar{x} - 3\frac{s}{\sqrt{n}}$ and $USL = \bar{x} + 3\frac{s}{\sqrt{n}}$

$$d = 3\frac{s}{\sqrt{n}}$$
$$C_p = d/3s = 1/\sqrt{n} = 0.44721$$ (and does not depend upon the value for $s$).

But $m = \bar{x}$ and hence $C_p = C_p = (1 - m - d)/d$.

Figures 1, 2 and 3 indicate there are two points that failed the out-of-control tests. These two points are sample point 23 and sample point 24. They failed because they fell between $\bar{x} - 3\frac{s}{\sqrt{n}}$ and $\bar{x} + 3\frac{s}{\sqrt{n}}$, i.e., they were 2 of 3 successive points found in Zone A. The next 3 sample points (25, 26 and 27) all fell within Zone C. Hence the process had either corrected itself or action had been taken to move the 10B loadings more closely towards the target value of 42.0 g/cm³. The process is considered to be in statistical control. However since the values for $C_p$, $C_{pk}$, $C_{pm}$ and $C_{pk}$ are all less than 1, this indicates the process is not 'capable'. The process needs to be changed to improve the output.

1000 bootstraps were carried out for the vector of 50 observations generated from a Beta distribution with parameters $r = 1$ and $s = 3$. The bootstrap procedure estimates the probability of out-of-spec to be approximately 6% (see Table 3).

**CODE**

```sql
/*-------------------------------------*/
| SAS Macro #1 |
|-------------------------------------*/

```sql
/*-------------------------------------*/
| SAS Macro in order to calculate various capability indices |
| to measure the performance of a process. |
| In this case LSL and USL are obtained via: |
| $m_{xbar} = 3*s/sqrt(5)$.
| |
| A Runs Test plot is produced. |
|-------------------------------------*/

```sql
OPTIONS NODATE NONUMBER PAGESIZE = 60
_LINESIZE = 80; goptions cback=white
 colors=(black cyan yellow green blue magenta);
```
data a;set a;
    xbar = mean(x1,x2,x3,x4,x5);
    range = range(x1,x2,x3,x4,x5); row=_N_; proc print data = a;
    TITLE1 'XBAR and RANGE Values for the 5 Subgroups';
    proc gplot data = a; plot xbar*row;
    symbol1 i=join v=dot l=1 color=black h=0.5;
    TITLE1 'XBAR Chart';
    proc gplot data = a; plot range*row;
    symbol1 i=join v=dot l=1 color=black h=0.5;
    TITLE1 'R Chart';
    proc means data = a noprint;var x1 x2 x3 x4 x5;
    output out = b mean = mean1 mean2 mean3 mean4 mean5;
    data a;set a;dm = 'a'; data b;set b;dm = 'a';
    proc sort data = a;by dm; proc sort data = b;by dm;
    data c;merge a b;by dm; data c;set c;
    %macro qc;
    %do i = 1 %to 5;
    mx&i = (x&i-mean&i)*(x&i-mean&i); ind=1;
    if (x&i=.) then ind = 1;else ind1=0; %end;
    %mend qc;
    %qc
    proc means data = c noprint;
    var x1 x2 x3 x4 x5 mx1 mx2 mx3 mx4 mx5 ind1 ind2 ind3 ind4
    ind5 range;
    output out = d sum = sum_x1 sum_x2 sum_x3 sum_x4 sum_x5
    sum1 sum2 sum3 sum4 sum5 sum_ind1 sum_ind2 sum_ind3 sum_ind4 sum_ind5 s_range;
    /*---------------------------------------------------------------------*/
| There are 5 subgroups. d3 and d4 are needed. These |
| constants are found in the Introduction to Statistical |
| Quality Control, 3rd ed. by Douglas C. Montgomery. | 
| The target value, t, is set equal to 42.0.   |
| Note cp=cp2 and cpk=cpk2 as expected. | 
| For the Zone A, B and C charts of the Runs test, 1*s, 2*s and |
| 3*s are used for the bounds. | 
|---------------------------------------------------------------------*/
data dd;set d;
    sse = sum1+sum2+sum3+sum4+sum5;
    dfe = (5*sum_ind1)-(5-1)-(sum_ind1+sum_ind2+sum_ind3 +
    sum_ind4+sum_ind5);
    mse = sse/dfe; s = sqrt(mse);
    m_xbar = (sum_x1+sum_x2+sum_x3+sum_x4+sum_x5)/(5*sum_ind1+
    sum_ind2+sum_ind3 + sum_ind4+sum_ind5);
    m_range = s_range/sum_ind;
    d3=0;d4=2.115;
    usl = m_xbar+3*s/sqrt(5);
    lsl = m_xbar-3*s/sqrt(5);
    if lsl<0 then lsl=0;if usl>0 then usl=0;
    t=42.0; d = (usl-lsl)/2; m=(usl+lsl)/2;
    g = abs(m_xbar-m)/d;
    cpk=cpm=cp*(1+(m_xbar-t)**2)**-0.5;
    cpq=cp*(1-1*(((m_xbar-t)**2)/2);
    zu=(usl-m_xbar)/s;
    zi=(lsl-m_xbar)/s;
    cp2 = (zu-zi)/6;
    cpk2 = min(abs(zu)/3.abs(zi)/3);
    /*---------------------------------------------------------------------*/
| There are 5 subgroups. d3 and d4 are needed. These |
| constants are found in the Introduction to Statistical |
| Quality Control, 3rd ed. by Douglas C. Montgomery. | 
| The target value, t, is set equal to 42.0.   |
| Note cp=cp2 and cpk=cpk2 as expected. | 
| For the Zone A, B and C charts of the Runs test, 1*s, 2*s and |
| 3*s are used for the bounds. | 
|---------------------------------------------------------------------*/
data e;merge a d2;set d;dm1 = 'a';
    usl3 = m_xbar+3*s/sqrt(5); lsl3 = m_xbar-3*s/sqrt(5);
    usl2 = m_xbar+2*s/sqrt(5); lsl2 = m_xbar-2*s/sqrt(5);
    usl1 = m_xbar+1*s/sqrt(5); lsl1 = m_xbar-1*s/sqrt(5);
    if lsl1<0 then lsl1 = 0; if usl1>0 then usl1 = 0; if lsl2<0 then lsl2 = 0; if lsl3<0 then lsl3 = 0;
    proc gplot data = e;
    plot xbar*row lsl*row usl*row m_xbar*row/overlay;
    symbol1 i=join l=1 color=black;
    symbol2 i=join l=2 color=black;
    symbol3 i=join l=2 color=black;
    symbol4 i=join l=1 color=black;
    TITLE1 'XBAR Chart (with LSL and USL)';
    proc gplot data = e;
    plot range*row lsr*row usr*row m_range*row/overlay;
    symbol1 i=join v=dot l=1 color=black h=0.5;
    symbol2 i=join l=2 color=black;
    symbol3 i=join l=2 color=black;
    symbol4 i=join l=1 color=black;
    TITLE1 'R Chart (with LSR and USR)';
    proc gplot data = e;
    plot xbar*row lsl3*row usl3*row m_xbar*row/overlay;
    symbol1 i=join v=dot l=1 color=black h=0.5;
    symbol2 i=join l=2 color=black;
    symbol3 i=join l=2 color=black;
    symbol4 i=join l=1 color=black;
    symbol5 i=join l=1 color=black;
    symbol6 i=join l=3 color=black;
    symbol7 i=join l=4 color=black;
    symbol8 i=join l=1 color=black;
    TITLE1 'RUNS Test with Zones A, B and C';
    run;
    /*---------------------------------------------------------------------*/
| SAS Macro #2 |
|---------------------------------------------------------------------*/
| SAS Macro in order to calculate the probability of |
| out-of-spec for samples with the use of the Bootstrap |
| for repeated sampling. | 
| In this case LSL and USL are input by the user. | 
|---------------------------------------------------------------------*/
data a; input x @@; cards;
0.31 0.19 0.08 0.35 0.49 0.10 0.72 0.39 0.04 0.19 0.22 0.05 0.09 0.14 0.59 0.09 0.14 0.12 0.10 0.25 0.10 0.11 0.54 0.42 0.06 0.19 0.32 0.66 0.59 0.20 0.07 0.07 0.12 0.45 0.43 0.04 0.18 0.02 0.24 0.36 0.06 0.03 0.12 0.14 0.17 0.15 0.48 0.33;

proc means data =a noprint;var x;
output out = b mean = mean_x std = std_x;
%macro boot;
%let n_boot=1000; %let usl = 0.60; %let lsl = 0.01;
%do i=1 %to &n_boot;
data cboot1;scan:
set a end=last;n+1;
if not last then goto scan;
do j=1 to n;
seed=floor(&i+1000000000*(sqrt(time())-floor(sqrt(time()))));
k=ceil(ranuni(seed)*n); set a point=k;
if _error_ then abort; output;
end;
stop;
proc means data = cboot1 noprint;var x;
output out = b2 mean=mean_x2 std = std_x2;
data b2;set b2;dm = 'a';drop _type_ _freq_;data b;set b;dm = 'a';drop _type_ _freq_;data cboot1;set cboot1;dm = 'a';
proc sort data = cboot1;by dm;
proc sort data = b;by dm;
proc sort data = b2;by dm;
data cboot2;merge cboot1 b b2;by dm;
data cboot2;set cboot2;drop dm;
zb=(x - mean_x)/std_x;
zl=(lsl-mean_x)/std_x;
zu=(usl-mean_x)/std_x;
if zb <= zl then ind_l = 1;else ind_l = 0;
if zb >= zu then ind_u = 1;else ind_u = 0;
ind=1;
proc means data = cboot2 noprint;
var ind ind_l ind_u;
output out = c sum = sum sum_l sum_u;
data c;set c;drop _type_ _freq_; pboot_l = sum_l/sum;
pboot_u = sum_u/sum;
keep pboot_l pboot_u;
data cc&i;set c;
%end;
%do j = 2 %to &n_boot;
proc append base = cc1 data = cc&j;
%end;
proc means data = cc1 noprint; var pboot_l pboot_u;
output out=d mean = pstar_l pstar_u std = se_l se_u;
data d;set d;drop _type_ _freq_; p_out = pstar_l + pstar_u;
proc print data = d;
var pstar_l pstar_u se_l se_u p_out;
TITLE1 'Application of the Bootstrap to Estimate the
Probability of Out-of-Spec (p_out)';
TITLE2 "Output: p*(L), p*(U), their standard errors and p_out
(&n_boot Bootstraps)";
%mend boot; %boot
run;

REFERENCES
Accuracy of Quality Parameters by Means of a Monte Carlo
Simulation,” Communications in Statistics, Simulation and
Computation, 28(3): 821-848.
Efron, B., and Tibshirani, R. J. (1993). An Introduction to the
Bootstrap, New York: Chapman and Hall.
Routines,” Prodigy Internet Provider
Basic Process Capability Indices \( \hat{C}_p \), \( \hat{C}_{pk} \) and \( \hat{C}_{pm} \), and their
Modifications,” Communications in Statistics, Simulation and

ACKNOWLEDGMENTS
SAS, SAS/STAT, SAS/QC and SAS/GRAPH are registered
trademarks or trademarks of SAS Institute Inc. in the USA and
other countries. ® indicates USA registration.

CONTACT INFORMATION
Paul Johnson, P.O. Box 4146, Davis CA 95617-4146
E-Mail: P.Johnson@prodigy.net
Ling Huang, Saint Mary's College, Moraga CA 94 575
E-Mail: Lhuang@stmarys-ca.edu