Quality Control and Home Runs: An Examination of Appropriate Metrics for Quality Control

Dr. Joseph Williams, Computer Information Systems, Colorado State University
Dr. Rich Edgeman, Computer Information Systems, Colorado State University

Abstract

The issue of appropriate quality control metrics is discussed in the context of the production of baseballs. Specifically, Major League Baseball players stroked record numbers of home runs in 1987, generating what is commonly known as the "lively ball" explanation for the unusual power output. The Commissioner's office released a study "proving" that the 1987 baseball was not unusually lively. However, after examination of the quality control metrics used by MLB, we demonstrate that (1) it is quite possible that baseballs were indeed "lively" and (2) that acceptance sampling alone does not properly address the quality control problem.

Introduction

Major League Baseball (MLB) experienced a record number of home runs stroked in 1987. Most analysts attributed the power barrage to baseballs that were somehow "livelier" than those used in previous years. MLB officials denied this explanation, citing an independent study employing acceptance in which it was demonstrated that the 1987 baseballs were within their production specifications.

In this paper, we demonstrate that MLB's conclusion regarding the lively ball was unsupported by the evidence they provided. Using simulation, we show that baseballs manufactured within production specifications could still readily prove to be "livelier" than baseballs produced in earlier years. Moreover, we show that development and maintenance of appropriate process control charts would have allowed the issue to be addressed definitively and with clarity.

The Lively Ball Controversy

In 1987 major league batters assaulted the record books for home runs. In total, 4,458 home runs were hit by National and American League players -- surpassing by 20% the previous combined record for home runs in a season (17). By all measures of home run productivity, 1987 was a record-setting year (Figure 1). Eight of the major league teams set franchise records for total home runs. Of 226 major league veterans with at least three years experience, 50 set career highs for home runs in a season.

Many experts and baseball fans attributed the phenomenal rise in home run productivity to a baseball that was "livelier" than those used before and after 1987. Other theories offered to explain the power surge included a weakened ozone layer of the atmosphere, bats made from harder wood, pitchers of lesser ability than usual, and bigger and stronger players than in the past. Two factors tended to tarnish the luster of the latter theories: (1) batting averages did not reach levels commensurate with home run production (Figure 2), which they should have, had pitchers or batters materially declined or improved, respectively; and (2) combinations of the same ozone, bats, pitchers and batters resulted in more "normal" home run productivity in 1988 and 1989. Thus, of all the theories, the lively ball hypothesis continues to receive the most attention and support.

Baseballs used in the major leagues are manufactured by the Rawlings Sporting Goods Company. A scientifically valid random sample of balls from each consignment is subjected to a battery of tests at Rawlings' testing laboratory at Licking, Missouri. One such test is the "rebound" test. Balls are fired out of cannons at a velocity of 85 m.p.h. against a wooden wall eight feet away. League specifications stipulate that a ball must rebound at 54.6 percent of the original velocity, plus or minus 3.2 percent.

During the height of the lively ball controversy in 1987, Major League Baseball published the results of an independent test of the baseballs used that season. After sampling and rebound-testing several dozen 1987 balls and making

FIGURE 1

FIGURE 2

a comparison with balls produced in 1985, it was concluded that "the test results indicated that the 1987 baseballs are totally within the parameters of major league standards" and, further, that "1987 baseballs are no livelier than those of past years"(2).

Several comments regarding these statements are in order. First, the statements are based solely on an application of acceptance sampling. Second, the sample size of "several dozen" balls is grossly inadequate for a capability assessment based on a single sample. This is due to the following considerations. A capability assessment based on a single sample inherently assumes not only that the process is in statistical control, but also that it is in control at the level given by the sample mean. To make such an assessment, the process variability must also be in statistical control. It is usually assumed that the measured characteristic is normally distributed. According to noted quality control authority Acheson J. Duncan (6), this assumption must be accepted with great caution when the estimate of the process proportion defective is very small, say less than 0.01. In such cases Duncan recommends that distribution-free methods such as Wilks' statistical tolerances be applied although samples of the order of 1,500 items may be required. While Duncan specifically states that 0.01 is a "very small proportion defective", it is worth noting that many manufacturers are measuring this proportion on a parts per million level and that a few manufacturers have made more significant headway still, measuring this proportion on a parts per billion scale. The use of a single sample from a given year implies that the process being sampled from is stable or "in-control" and does little to capture the dynamic nature of the production process. Consequently, the process variation will be over-estimated and the tests comparing 1987 baseballs to 1985 baseballs will have low resolution. A comparison of 1987 baseballs with 1985 baseballs does not logically lead to the conclusion that 1987 baseballs were no more lively than balls of past years; such a conclusion is predicated on the assumption that the process used to produce baseballs was stable over a period of years. In fact, it is a rare production process that is "in-control" for days, weeks or months, let alone years.

Before discussing problems with the MLB study, however, we review the logic contained therein.

Specification Limits and Process Capability

As described above, Major League Baseball specifications stipulate a rebound parameter of 54.6 percent of original velocity, plus or minus 3.2 percent. Consequently, baseballs must rebound at a velocity between 43.69 mph and 49.13 mph and the average rebound velocity of tested baseballs should be 46.41 mph. The lower limit, 43.69 mph, is an example of a lower specification limit (LSL). Similarly, the upper specification limit (USL) is 49.13 mph. The tolerance (TOL) for rebound speed is 3.2

percent of 85 mph, or 2.72 mph.

The majority of information contained in the remainder of this section and the whole of the next section follows from elementary statistical and quality control concepts and numerous references are available (1)(12)(13). This material is reviewed here for the sake of completeness. A process is generally considered to be minimally capable if the tolerance of the process is $TOL = 3\sigma$, where σ is the process standard deviation (1). With respect to rebound speed, the process used to manufacture baseballs for use in the major leagues would be minimally capable if the standard deviation was $\sigma = TOL/3 = .91$ mph.

If process output is normally distributed, well-centered and minimally capable, then approximately one out of every 400 baseballs produced will fall outside of the specification limits. Under these conditions, an acceptance strategy of $C = 0$ or $C = 1$ from a sample of 36 baseballs may well be effective given MLB's quality control goals.

However, if the process is not well-centered and/or there has been an increase in variation, then an increased proportion of baseballs will fail to meet specifications. As an example, consider a process that is minimally capable and well-centered so that the in-control state of the process is characterized by a normal distribution with mean μ and standard deviation σ . We can examine the impact of a shift in the process mean from μ to $\mu_1 = \mu + \delta\sigma_1$ and an increase in the standard deviation from σ to $\sigma_1 = k\sigma$. The resulting proportions of baseballs from the production process failing to meet specifications are easily computed as follows:

$$P(\text{specifications unsatisfied}) = 1 - P(\text{specifications satisfied})$$

$$= 1 - P = P(LSL \leq X \leq USL \mid \mu_1, \sigma_1) \quad [1]$$

$$= 1 - P([LSL - \mu_1]/\sigma_1 \leq Z \leq [USL - \mu_1]/\sigma_1). \quad [2]$$

Recalling that $LSL = \mu - 3\sigma$, $USL = \mu + 3\sigma$, $\mu_1 = \mu + \delta\sigma_1$, $\sigma_1 = k\sigma$, equation [2] becomes

$$= 1 - P\{(-3 - \delta k)/k \leq Z \leq (3 - \delta k)/k\} \quad [3]$$

$$= 1 - (\Phi\{(3 - \delta k)/k\} - \Phi\{(-3 - \delta k)/k\}) \quad [4]$$

where $\Phi(z)$ is the standard normal cumulative distribution function evaluated at z . Table 1 reports the proportions of baseballs from the production process failing to satisfy specifications for various values of $|\delta|$ and k .

Relating the proportions from Table 1 to the simplest acceptance sampling implementation where n items are sampled from a lot and the lot is accepted if C or fewer of the sampled items are defective, suppose that $n = 36$ baseballs ("several dozen") are sampled from a lot with proportion P failing to satisfy specifications. The probabil-

ity of accepting the lot is reported in Table 2 for both $C = 0$ and $C = 1$ for a variety of values of P .

Functionally, use of the $C = 0$ or $C = 1$ acceptance criteria would be proper so long as the production process was capable and in control about desired mean (46.41 mph). This situation is depicted in Figure 3 for a series of samples. Consider, however, the situation in which the production process is capable and in control at a mean that is $1/2$ standard deviation (.455 mph) above the desired mean (Figure 4). In this instance, the balls produced are consistently livelier than desired; however, the percentage of lots that will be accepted in a single sample of 36 baseballs is 69.24% and 97.74% for $C = 0$ and $C = 1$, respectively! Therefore, Major League Baseball's test of 1987 baseballs is hardly conclusive. Indeed, the test results become little more than an additional piece of circumstantial evidence in the lively ball controversy.

Control Charts and Process Control

Prior to the start of the 1987 season, an engineer for Rawlings was quoted as saying "We (Rawlings) definitely think we're producing the most consistent ball ever" (4). However, consistency is a shallow goal if the production process is consistently producing items that fail to conform to product specifications. The circumstantial evidence favoring a lively ball is more convincing than claims to the contrary based on the results of the previously cited MLB study. The issue of the lively ball is better approached using statistical quality control techniques such as the ones that will now be discussed.

Control charts are graphical tools used to monitor a process and to test the hypothesis that the process is in control. If a process is "out-of-control," it may be so with regard to either process average or process variability, or both. Hence, two control charts—one for process average and one for variability—must be maintained. The chart used to monitor shifts

TABLE 1
Combinations of $|\delta|$ and k

	$ \delta $					
	0	.25	.50	.75	1.00	1.25
k 1.00	.00270	.00356	.00644	.01231	.02278	.04007
1.25	.01640	.01980	.05029	.08109	.12520	.18410
1.50	.04550	.05228	.07302	.10863	.16001	.22720
1.75	.08648	.09631	.12572	.17432	.24085	.32274
2.00	.13360	.14571	.18141	.23885	.31475	.40427

TABLE 2
Probability of Accepting the Baseballs Lot
 $n = 36$

Proportion Defective	$P(\text{Accept} \mid C = 0)$	$P(\text{Accept} \mid C = 1)$
.0025	.9512	.9988
.0050	.9046	.9859
.00644	.7924	.9774 ($\delta = .5, k = 1$)
.0100	.6964	.9497
.0150	.5804	.8985
.0164	.5515	.8824 ($\delta = 0, k = 1.25$)
.0200	.4832	.8382

FIGURE 3

Production Process is Capable and in Control at the Mean

FIGURE 4

Production Process is Capable and in Control at the Mean + 1/2 sd

in the process average is called an $\bar{X}$ (X-bar) chart and the chart used to monitor changes in process variability is called an R (for range) chart. Numerous widely available software packages provide control chart support and some offer a standard deviation or S chart that may be used instead of an R chart to monitor process variability. Our choice of to use the R chart rather than the S chart in this paper is based both on its simplicity and on the fact that it is much more widely used in practice. Estimates of the true process average, μ , and the true process standard deviation, σ , must be obtained before these control charts can be constructed.

Estimation of μ and σ is usually based on the collective information of M preliminary samples of N items each. These samples are selected when the process is considered to be operating in a state of statistical control. A process is considered to be in-control when differences in process output can be attributed solely to chance causes (i.e., solely to the inherent noise in the process). Typically, the value of M should be at least 20 to 25 (12). Selection of N and frequency of sampling are usually guided by economic considerations (6), but it is not unusual for N to be quite small, frequently in the four to 10 range. Estimates of the process standard deviation, σ , based on the range are very nearly as statistically efficient as estimates of σ based on the sample standard deviation when the sample size is small. This provides additional motivation for the use of R rather than S charts to monitor process variability.

If $\bar{X}_j$ is the j^{th} sample mean, then the estimate used for μ in control chart settings is given by

$$\bar{\bar{X}} = (\bar{X}_1 + \bar{X}_2 + \dots + \bar{X}_M) / M. \quad [5]$$

The range of a random sample is the numerical difference between its largest and smallest elements. Let R_j be the range of the j^{th} sample and let the average range be denoted by $\bar{R}$ where

$$\bar{R} = (R_1 + R_2 + \dots + R_M) / M. \quad [6]$$

The estimator of σ used in control chart settings is $\hat{\sigma}$ where

$$\hat{\sigma} = \bar{R}/d_2. \quad [7]$$

The value of d_2 depends only on the sample size, N , and is widely tabled (6)(9)(14). The $\bar{X}$ chart is constructed by serially plotting sample means on the vertical axis of the graph against the corresponding sample number on the horizontal axis. In its most basic form, the $\bar{X}$ chart has three horizontal lines: a center line (CL) positioned at $\bar{\bar{X}}$, an upper control limit (UCL), and a lower control limit (LCL). the UCL and LCL are plotted as horizontal lines at

$$\text{UCL} = \bar{\bar{X}} + 3\hat{\sigma}/N = \bar{\bar{X}} + A_2\bar{R} \quad [8]$$

and

$$\text{LCL} = \bar{\bar{X}} - 3\hat{\sigma}/N = \bar{\bar{X}} - A_2\bar{R}. \quad [9]$$

There are some who advocate that $\bar{\bar{X}}$ be replaced by μ in the formulas for the CL, UCL, and LCL where, in such cases, μ is the design specification. It is our recommendation that this not be done since specifications do not relate to the issue statistical control. Sensitivity of control charts to out-of-control signals can be increased by placing upper and lower "warning" limits as horizontal lines at

$$\bar{\bar{X}} + 2A_2\bar{R}/3 \quad [10]$$

and

$$\bar{\bar{X}} - 2A_2\bar{R}/3 \quad [11]$$

respectively. Similarly, upper and lower one-sigma lines may be positioned at a distance $A_2\bar{R}/3$ above and below $\bar{\bar{X}}$, respectively. If the process is in control about the CL value, then a value of $\bar{X}$ will plot outside the control limits with a probability approximately equal to .0027. Any value of $\bar{X}$ plotting outside of these control limits is subject to scrutiny and is associated with an out-of-control signal from the process. Control chart goals are to separate out-of-control signals emitted from the process from the background noise of the process and, subsequently, to identify the source of the signal and take needed corrective action.

Supplementary criteria aimed at increasing the sensitivity of control charts to out-of-control signals are report in Juran and Gryna (11). These criteria are probabilistic in nature and are based on the identification of non-random patterns on the control charts that are probabilistically rare if the process is in control, and informative if the process is out-of-control. Three of these supplementary criteria are listed below:

- A. two out of three consecutive points plot beyond the warning limits on the same side of the CL value;
- B. four out of five consecutive points plot at a distance of at least $A_2\bar{R}/3$ beyond the same side of the CL value; and
- C. eight consecutive points plot on one side of the CL value.

Process centrality (average) and process variability are usually monitored simultaneously. Variability may be monitored by an R chart. An estimate of the standard deviation of the sample range is given by

$$\hat{\sigma}_R = d_3\hat{\sigma} = d_3R/d_2. \quad [12]$$

Upper and lower control lines for the R control chart for process variability are given by

$$\text{UCL} = \bar{R} + 3d_3\bar{R}/d_2 = (1 + 3d_3/d_2)\bar{R} = D_4\bar{R} \quad [13]$$

and

$$LCL = \bar{R} - 3d_3\bar{R}/d_2 = (1 - 3d_3/d_2)\bar{R} = D_3\bar{R}. \quad [14]$$

If the lower control limit is negative then it is adjusted to zero. The center line on an R chart is positioned at $\bar{R}$. Warning lines and one-sided lines may be added to the R chart in a manner analogous to the $\bar{X}$ chart.

Values of A_2 , d_3 , D_3 and D_4 depend only on the sample size, N , and are tabled in the references cited previously for d_2 . Out-of-control signals for process variability are identified in a manner analogous to that for $\bar{X}$ charts. A thorough development of the theory behind control charts as well as numerous numerical examples are comprehensively developed elsewhere (1)(12).

Quality Control and Lively Baseballs

Had control charts been maintained over the years on key characteristics of baseballs manufactured for use in the major leagues, the lively ball issue could be addressed with clarity. An increase or decrease of any specified magnitude in the liveliness of baseballs could be detected by routinely selecting random samples of N baseballs at regular production intervals and comparing computed values of certain statistical measures to "control limits." If any of the measures were outside of the control limits or if any of the supplementary criteria were satisfied, then we would say that the process had emitted an out-of-control signal and it would be necessary to identify the source of the signal to apply corrective action(s).

Absent available data, we developed control charts based on a simulation of the rebound speed of baseballs under three sets of conditions using a simulation package by Edgeman and Scott (7). Initially, $M = 25$ random samples of $N = 6$ items each were simulated based on the assumptions that rebound speed is normally distributed, centered between the specification limits discussed above, and with a PCR value equal to 1.00. Process control was established using the information contained in the initial 25 samples, resulting in the $\bar{X}$ and R charts of Figures 5 and 6, respectively. The control limits recorded on these charts are the standards to which values of $\bar{X}$ and R from future samples are compared to assess the state of the production process. The dashed lines in Figures 5 and 6 are the warning lines and the one-sigma lines.

Figures 5 and 6 depict a process that is in a state of statistical control. Both display random patterns with no points plotting beyond the control limits. None of the 150 simulated rebound speeds in the initial 25 samples failed to conform to the specifications for rebound speed.

For the next 25 samples of $N = 6$ items each, an increase in liveliness was introduced into the process by increasing the average rebound speed by .455 mph (one-

half the value of the standard deviation used for the initial 25 samples), resulting in Figures 7 and 8. Process variability was left unchanged at this point. Based on these changes, any out-of-control signals detected on the R chart of Figure 8 would erroneous, but any out-of-control signals detected on the $\bar{X}$ chart of Figure 7 should, presumably, indicate an increase in liveliness.

No out-of-control signals with respect to process variation are detectable on the R chart of Figure 8. The $\bar{X}$ chart of Figure 7 provides multiple out-of-control signals with respect to process centrality (i.e., average rebound speed) and all of these signals are consistent with an increase in liveliness. The earliest detectable out-of-control signal comes when, according to Supplementary Criterion A of Juran and Gryna discussed above, the second and fourth sample means plot above the upper warning limit. Early detection of an out-of-control condition is desirable since it allows for earlier corrective action and leads to the production of fewer baseballs that fail to conform to specifications. Of the 150 simulated rebound speed that went into the construction of Figures 7 and 8, only two failed to conform to specifications--yet clearly the process was out-of-control and producing livelier baseballs.

For the last 25 samples of $N = 6$ items each, the process output was again centered between the specifications but a 25 percent increase in process variation was introduced, resulting in Figures 9 and 10. Hence, any out-of-control signals detected by the R chart of Figure 10 should be consistent with increased variation. Although the process was in control with respect to centrality during this last portion of the simulation study, it would not be unusual for the X chart to indicate lack of control in the process, given the increase in variation.

An out-of-control signal is detected on the $\bar{X}$ chart of Figure 9 when, according to Supplementary Criterion A above, the means of the seventh and eight samples both plot below the lower warning limit. No other indications of an out-of-control condition are evident on the $\bar{X}$ chart. The R chart of Figure 10 gives multiple out-of-control signals with respect to process variation. The earliest detectable signal occurs when the R value for the ninth sample plots above the UCL, indicating increased process variation. It is at this point that process intervention would occur. Of the 150 simulated rebound speed used in this portion of the simulation only one failed to meet specifications; yet, again, the process was clearly out-of-control.

In short, if acceptance sampling was used to determine the liveliness of baseballs, it is extremely unlikely that any consignments would have been refused or even that any alarm would have been aroused. The use of control charts, however, clearly indicates that a capable process can be frequently out-of-control and producing livelier baseballs. We do not wish to give the indication that there

FIGURE 5

FIGURE 6

FIGURE 7

FIGURE 8

FIGURE 9

X-BAR CHART FOR REBOUND SPEED: INCREASE IN VARIATION

FIGURE 10

R CHART FOR REBOUND SPEED: INCREASE IN VARIATION

is anything magical about the use of control charts. In general control charts can be thought of as repeated application of a test of hypothesis which is in itself acknowledgement that the process under surveillance is dynamic. Application of supplementary rules, such as those given in Juran and Gryna (11), allow for detection of process trends - something that is not standard fare for either one-time or repeated application of standard tests of hypothesis.

Summary

In determining whether a consignment of baseballs is acceptable for use in the major leagues, a random sample of N baseballs from each consignment is subjected to a battery of tests and, if C or fewer fail to conform to specifications, the consignment is accepted. The simulation study of the previous section produced only three rebound speeds among the 450 simulated values that failed to conform to specifications. If acceptance sampling (rather than control chart methods) were used to determine the acceptability of the baseballs, it is extremely unlikely that any consignments of baseballs would have been refused, yet the control charts clearly indicated a production process that was frequently out-of-control and sometimes producing livelier baseballs.

Quality control, when limited to capability assessment and acceptance sampling, is not sufficient. Whether the process yields baseballs or medication or boxes of cereal, process control in the form of control charts and additional SPC techniques must be implemented if the complete quality picture is to emerge.

Bibliography

1. Burr, I.W., *Statistical Quality Control Methods*, Marcel Dekker, New York, 1976.
2. *Dallas Times Herald*, July 31, 1987, p. D1.
3. Deming, W.E., "On Some Statistical Aids Toward Economic Production," *Interfaces*, Vol. 5, No. 4, p. 5, 1975.
4. *Denton Record-Chronicle*, February 17, 1989, C-1.
5. Duncan, A.J., *Process Capability Studies*, Proceedings, Middle Atlantic Conference of the American Society for Quality Control, February 1958.
6. Duncan, A.J., *Quality Control and Industrial Statistics*, Irwin, Homewood, IL, 5th ed., 1986.
7. Edgeman, R.L. and R.C. Scott, "Generator: A Program for Simulating Non-Uniform Random Variates," *The American Statistician*, Vol. 40, No. 1, p. 54, 1986.
8. Feigenbaum, A.V., *Total Quality Control*, McGraw-Hill, New York, 1983.
9. Grant, E.L. and R.S. Leavenworth, *Statistical Quality Control*, McGraw-Hill, New York, 5th ed., 1980.
10. Guenther, W.C., *Sampling Inspection in Statistical Quality Control*, Macmillan, New York, 1977.
11. Juran, J.M. and F.M. Gryna, *Quality Planning and Analysis*, McGraw-Hill, New York, 2nd ed., 1980.
12. Montgomery, D.C., *Statistical Quality Control*, Wiley, New York, 2nd ed., 1991.
13. Ryan, T.P., *Statistical Methods for Quality Improvement*, Wiley, New York, 1989.
14. Schilling, E.G., *Acceptance Sampling in Quality Control*, Marcel Dekker, New York, 1982.
15. Shafer, E., "Baseballs Take Major Beatings," *A.P. Wire Story*, February 17, 1987.
16. Tarver, M.G., "Process Capability Studies," in L. Walsh, R. Wurster & R. Kimber (Eds.), *Quality Management Handbook*, Marcel Dekker, New York, 1986.
17. Thorn, J. and P. Palmer (Eds.), *Total Baseball*, Warner Books, New York, 1989.
18. Wolak, J., "Motorola Revisited," *Quality*, Vol. 26, No. 5. pp. 18-23, 1987.