Colored Candies and SAS/QC®: A Sweet Way to Produce Chi-Square Control Charts

Melvin T. Alexander, Westinghouse, Baltimore, MD
SherriJoyce King, King Information Company, Rockville, MD

Colored candies are easy-to-understand tools that can illustrate Statistical Quality Control principles. Each candy bag should have about the same piece counts of each color as the other bags. A multinomial model is useful in characterizing the color piece counts of bags.

Proportion charts (p-charts) and count charts (c/u charts) help assess the stability of the process that fills bags for each color. Maintaining separate p- or c/u-charts to monitor each color, however, is cumbersome and costly. A chart that monitors the colors jointly would be more cost-effective and desirable.

The Chi-square (χ^2) control chart serves to monitor several attributes simultaneously. The χ^2 can test: (a) goodness-of-fit of probability distributions; (b) homogeneity of proportions; and (c) independence of two or more cross-classified variables in contingency tables.

This paper will review the use of the χ^2 control chart to check the stability of piece counts for each color from bag to bag. The chart has other applications as well, such as monitoring defects in manufacturing circuit boards.

King and Alexander (1994) examined the use of Base SAS to produce χ^2 control charts. It was found that the FREQ procedure was unable to output cell χ^2, so the DATA step was used instead. This paper deals with the more statistical aspects of χ^2 and describes how SAS/QC® produces a χ^2 control chart with the DATA step that provides the parameters for PROC SHEWHART.

PURPOSE AND PROBLEM

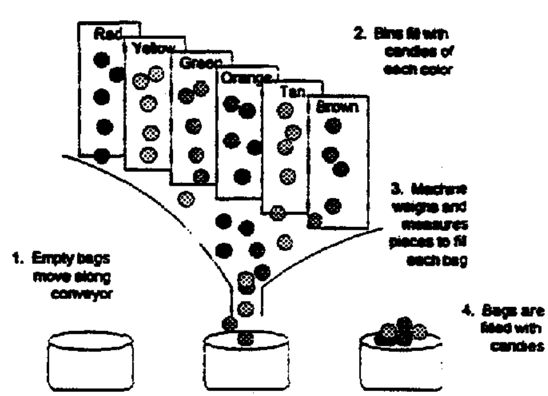
The purpose of this paper is to show the use of the Chi-square (χ^2) control chart as a way of monitoring processes like the bag-filling operations of colored candies. At SUGI 19 we described the bag-filling process and expained how control charts can be used to check the stability of that process. A typical bag of colored candies should have about the same proportional count of pieces of six different colors. To find out what the color distributions for colored candies should be, contact the candy companies (e.g., Mars Inc., Hershey Foods Corporation, etc.). Suppose the colors in each bag are red, yellow, green, orange, light (tan) and dark brown. Monitoring the process for each color requires six separate control charts, filling walls with costly paper. Others have used single-color pieces to demonstrate statistical process control, e.g., Kohn and Dorner (1992), Kullberg (1992), and Wortham (1992). These examples focused on one color and ignored the other colors. The motivation for considering the χ^2 control chart is to form a composite chart that monitors all the colors while preserving the information of these colors. See Duncan (1950) and Marcucci (1985) for more information. The χ^2 is an old statistical tool used to test (a) goodness-of-fit of probability distributions, (b) homogeneity of proportions, and, (c) independence of (two or more) cross-classified variables in contingency tables. Schlotzhauer (1991) described various χ^2 options within the SAS® System that could be used to do goodness-of-fit testing. The same χ^2 used in the above instances can check the stability of piece counts of all colors across bags as they are filled. The χ^2 control chart has other applications such as monitoring many types of defects during placement of components on complex products like printed circuit boards (PCBs) or printed wiring assemblies (PWAs); mounting MicroElectronics Devices (MEDs) on substrates; and comparing Pareto analyses of the same categories taken over different periods.

PROPOSED METHOD: THE MULTINOMIAL MODEL

Figure 1 depicts the colored candies bag-filling operation.

Figure 1

Bag Filling Operation



Empty bags move along a conveyor and stop at a bag filling station. Colored pieces fill bins for each color. Machines weigh, measure, and fill the bags with colored pieces. Operators or machines pull samples of bags, count the pieces, then enter or plot the data onto control charts. Usually, the sample consists of 20-30 bags according to some continuous sampling plan. Bags showing statistical control on the charts mean the color counts and color mix are stable. These bags may be sealed, packed, and shipped to stores for sale. Charts that lack statistical control mean that some bags may be under or over filled. An investigation should be made to find, fix, or correct the assignable cause(s), e.g., bins empty of pieces, machines needing recalibration, fixture setups, etc.

Let $j=1,2,3,4,5,6$ index the set of red, yellow, green, orange, tan, and brown color categories of pieces, respectively. The six colors correspond to the $C=6$ response profiles associated with the specific characteristics of interest. Similarly, let $i=1,2,\dots, s=20\text{-}30$ index the set of subpopulations (or subgroups) defined with pertinent independent variables. In other words, s is the sample of colored candy bags filled. The subpopulation s could also represent placed components on substrates or the number of periods that the same categories are compared. Finally, let n_{i+} denote total number of colored candies in each bag i ($i=1,2,\dots,s$), where

$n_{i+} = n_{i1} + n_{i2} + n_{i3} + n_{i4} + n_{i5} + n_{i6}$, and n_{ij} = no. of pieces of color j in bag i.

The resulting data can be summarized in Table 1:

Table 1. (S x C) Response Profile of Color Categories of Counts

	COLOR						
BAG	1	2	3	4	5	6	SUM
1	n_{11}	n_{12}	n_{13}	n_{14}	n_{15}	n_{16}	n_{1+}
2	n_{21}	n_{22}	n_{23}	n_{24}	n_{25}	n_{26}	n_{2+}
⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓
i	n_{i1}	n_{i2}	n_{i3}	n_{14}	n_{i5}	n_{i6}	n_{i+}
⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓
s	n_{s1}	n_{s2}	n_{s3}	n_{s4}	n_{s5}	n_{s6}	n_{s+}
SUM	n_{+1}	n_{+2}	n_{+3}	n_{+4}	n_{+5}	n_{+6}	n_{++}

The counts (n_{ij}'s) of Table 1 can be also represented as proportions p_{ij}, where p_{ij} denotes the expected proportion for bag i having color j and is represented by Table 2:

Table 2. (S x C) Response Profile of Color Categories of Proportions

	COLOR						
BAG	1	2	3	4	5	6	SUM
1	p_{11}	p_{12}	p_{13}	p_{14}	p_{15}	p_{16}	p_{1+}
2	p_{21}	p_{22}	p_{23}	p_{24}	p_{25}	p_{26}	p_{2+}
⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓
i	p_{i1}	p_{i2}	p_{i3}	p_{14}	p_{i5}	p_{i6}	p_{i+}
⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓
s	p_{s1}	p_{s2}	p_{s3}	p_{s4}	p_{s5}	p_{s6}	p_{s+}
SUM	p_1	p_2	p_3	p_4	p_5	p_6	1.0

The null hypothesis of the χ^2 test of homogeneity is:

$H_0: p_{1j} = p_{2j} = \dots = p_{sj} = p_j$ where

$$p_j = \sum_{i=1}^{s} n_{ij} \Big/ \Big(\sum_{i=1}^{s} \sum_{j=1}^{c} n_{ij}\Big) = n_{+j}/n_{++}$$

In other words, the χ_i^2 checks if each bag i has the same color proportion p_j of pieces. The Pearson's χ_i^2 is computed as:

$$\chi_i^2 = \sum_{j=1}^{c} \left[(n_{ij} - n_{i+} {*} p_j)^2 / (n_{i+} {*} p_j) \right]$$

n_{ij} = observed jth-color count in bag where $i = 1,2,\dots, s$; and p_j = proportion of the jth-color. χ_i^2 has 6-1=5 degrees of freedom and approximates reasonably well so long as $n_{i+} {*} p_j$ is not very small (i.e., ≥ 5). If contingency tables have empty cells or small expected cell frequencies (i.e., < 5), then computing exact probabilities is an alternative approach to testing homogeneity of proportions. Fisher's exact test is a method of computing exact cell probabilities and more extreme-than-observed cells for given contingency tables. The method takes changing values of cell frequencies to produce distinct exact probabilities for fixed marginal color totals and bag counts. The size of the summed exact probabilities determines the acceptance or rejection of the homogeneous proportions' hypothesis. Fisher's exact test typically is used with 2X2 contingency tables. Ghent (1972) and Mehta and Patel (1983) described methods extending Fisher's exact test beyond just 2X2 contingency tables. Although their approaches involved extensive calculation, tools within the SAS system facilitate these computations. More could be said about computing exact probabilities of contingency tables. These issues are beyond the scope of this presentation, but are discussed by Mehta (1994).

The control limits for χ_i^2 can be set to warning limits (2-sigma) instead of the customary 3-sigma limits. Warning limits increase the sensitivity of control charts and signal early detection of suspicious processes if they do not perform properly. The warning control limits (denoted with the lower-left subscript w) for the χ_i^2 are:

$$
\begin{aligned}
{}_{w}UCL_{\chi}{}^2 &= \chi^2\,(1-\alpha/2, 5df) \\
CL_{\chi}{}^2 &= \chi^2\,(0.5, 5df) \\
{}_{w}LCL_{\chi}{}^2 &= \chi^2\,(\alpha/2, 5df).
\end{aligned}
$$

where $\alpha = 0.05$ for 2-sigma warning limits.
The out of control criteria are :
Any two of three consecutive
$\chi_i^2 < \chi^2\,(\alpha/2, 5df)$ or
$\chi_i^2 > \chi^2\,(1-\alpha/2, 5df)$ signals bag i was overfilled or underfilled of one or more colors. Usually more concern is given to points above the upper control limits than to points below the lower control limits. Points below the lower control limit tend to show more uniformity than expected because all colors are equally present in each bag. This means that:
the P_{ij}'s = 1/6 (j=1,2,...,6) and $\chi_i^2 = 0$. Looking for the largest or smallest color cell χ_{ij}^2 contribution tells the analyst, process engineer, or operator which jth-color count (proportion) differed from the expected. Nelson (1982) discusses this procedure in more detail.

King and Alexander (1994) showed color data and Chi-square control limit calculations of a sample of 20 colored candy bags. Besides monitoring all colors simultaneously, preserving the sequential order of production is another advantage of the χ_i^2 control chart. It allows the analyst, engineer, or operator to study runs, trends, and other patterns that might arise in the process. For more information, see Montgomery (1991) and Nelson (1987).

OTHER USES

Other extensions of the χ_i^2 control chart include comparing different Pareto analyses that have similar attribute categories. Kenett (1991) gives a good discussion of how the Chi-square compares Pareto analyses. His χ^2 method may be used to monitor categories of different product (whether they are candy bags, PCBs, or PWAs, etc.) that may span across various machines, operators, or inspectors. This tool helps technical trainers focus training efforts so those defect occurrences (or recurrences) in manufactured product can be reduced.

THE PROGRAM

We created chi-square (χ^2) control charts in SAS/QC ®using then bags' data as follows:

Create two data sets:

- To calculate χ^2 values for each bag across the color vectors.
- To calculate the χ^2 control limit parameters that will be used to construct the control chart using the SHEWHART procedure.

PROC SHEWHART has options to reproduce the χ^2 control charts we made using Base SAS. The first data set can be created using the SUMMARY procedure . The code was supplied in King and Alexander (1994).

The output of the SUMMARY procedure was the χ^2 values for the individual bags. These values would be combined with χ^2 control limits (CTRLLIMS) to form another data set that would construct a control chart using the PLOT or GPLOT procedures (in Base SAS® and SAS/GRAPH®, respectively) or the SHEWHART procedure (SAS/QC) for the second data set.

When using the SHEWHART procedure, two data sets are needed:

- A data set that will be used with the LIMITS= option that contains the control limits parameters. Let's call this data set CHPARMS. It has the following variables:

  VAR - the name of the process variable, e.g., BAGCHI2 (bags' χ^2 variable).

  SUBGRP - subgroup variable by bag, e.g., BAGID.

  SIGMAS - the multiple of the standard error of the moving average or EWMA. For warning limits, _SIGMAS_=2.

  MEAN - the process level or mean. In the case of the χ^2 control chart, _MEAN_=CINV(0.5,5). CINV is the SAS System's inverse chi-square function. With these arguments it calculates the chi-square value associated with a probability of .5 (50%) and 5 degrees of freedom.

  STDDEV - the process level/mean standard deviation. According to standard control charting conventions, if we let:

$\_UCLI\_ = wUCL_{\chi}{}^2 = CL_{\chi}{}^2 + 2\sigma$ be the 97.5 percentile,

$\_LCLI\_ = wLCL_{\chi}{}^2 = CL_{\chi}{}^2 - 2\sigma$ be the 2.5 percentile,

and

$CL_{\chi}{}^2$ is the process center. Then

$CL_{\chi}{}^2 = CINV(0.5,5)$,

$\_UCLI\_ = wUCL_{\chi}{}^2 = CINV(.975,5)$,

and

$\_LCLI\_ = wLCL_{\chi}{}^2 = CINV(0.025,5)$.

Therefore, _STDDEV_ is not really necessary since the standard deviation is contained in the upper and lower control limits when the CINV function was used. SHEWHART uses _STDDEV_ to draw the control limits for the moving range (_R_). The variables _STDDEV_, _R_, _LCLR_, and _UCLR_ are required by SHEWHART to produce the Chi-square values for individuals and moving range charts. However, only the individuals chart for the Chi-square will be kept, and the moving range chart omitted. The NOCHART2 option suppresses display of the moving range chart. The DATA step below creates the LIMITS= option data set (CHPARMS). The LIMITS= and READLIMITS statements specifiy that the control limits are to be read rather than computed:

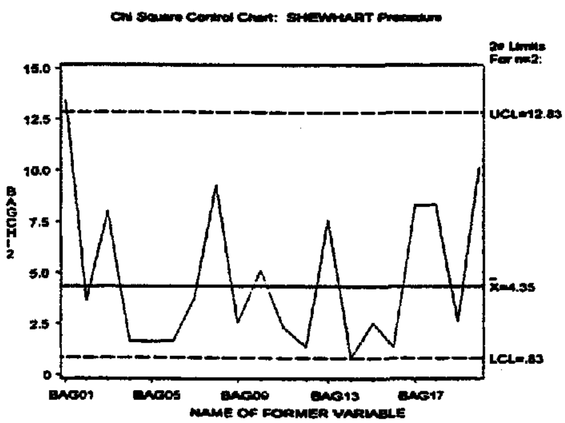
```
 data chparms;
 set ctrllims;
length _var_ _subgrp_ _type_ $ 8
_index_ $ 16;
_var_ = 'bagchi2';
_subgrp_ = 'bagid';
_sigmas_ = 2;
_type_ ='standard';
_stddev_ = (cinv(.975,5)-cinv(.025,5))/2;
_r_     = _stddev_ * 1.128;
_lclr_   = 0;
_ucl_    = _r_ * 3.267;
 run;

title 'Chi Square Control Chart:
SHEWHART Procedure';
 proc shewhart data=bagchi2
 limits=chparms graphics;
   irchart bagchi2*bagid /
 outtable=outtbl readlimits
 skiplables=3 nochart2;
run;
```

RESULTS

We were able to get the consistent results obtained by King and Alexander (1994). Figure 2 produced when we defined standard deviation like this:

```
_stddev_ = (cinv(.975,5)-cinv(.025,5))/2;
```

Figure 2



CONCLUSIONS

The Chi-square control chart presented earlier is versatile, well known (to some extent), and useful for quality control practitioners. By combining the Chi-square with the SAS System, we hope to make this tool applicable, accessible, and palatable (smile)!

REFERENCES

Duncan, A.J., (1950), "A Chi-Square Chart for Controlling a Set of Percentages." *Industrial Quality Control*, 7, 11-15.

Ghent, A.W., (1972), "A Method for Exact Testing of 2X2, 2X3, 3X3, and Other Contingency Tables, Employing Binomial Coefficients," *The American Midland Naturalist*, 88, 15-27.

Kenett, R.S., (1991), "Two Methods for Comparing Pareto Charts," *Journal of Quality Technology*, 23, 27-31.

King, SJ, and Alexander, M.T. (1994), "Colored Candies and Base SAS®: A Sweet Way to Produce Chi-Square Control Charts, Part 1: Tutorial" Paper presented at the SUGI 19 Conference, Dallas, TX, April 10-13, 1994.

Kohn, R.F. and Dorner, W.W., (1992), "A Hands-On Approach to Teaching Process Control and Capability," Paper presented at the 1992 Winter ASA Conference, Louisville, KY, January 3-5, 1992.

Kullberg, K.N., (1992), "M&M's Exercise," *Advanced Quality System Training Course,* Seattle, WA: The Boeing Company.

Marcucci, M., (1985), "Monitoring Multinomial Processes," *Journal of Quality Technology*, 17, 86-91.

Mehta, C.R. and Patel, N.R., (1983), "A Network Algorithm for Performing Fisher's Exact Test in r x c Contingency Tables," *Journal of the American Statistical Association,* 78, 427-434.

___________, (1994), "Small-Sample Inference in SAS® Software: Using Exact Methods," Paper presented at the SUGI 19 Conference, Dallas, TX, April 10-13, 1994.

Montgomery, D.C., (1991), *Introduction to Statistical Quality Control,Second Edition,* New York: Wiley, 322-330.

Nelson, L.S., (1987), "A Chi-Square Control Chart for Several Proportions," *Journal of Quality Technology*, 19, 229-231.

Nelson, W., (1982), *Applied Life Data Analysis,* New York:Wiley, 454-459 .

Schlotzhauer, D. , (1991), " Computing Goodness-of-fit Chi-square Statistics with SAS® Software," *Observations,* 1, 17-24.

Wortham, B.L., (1992), "Poisson Green M&M's - Attributes Charting Simplified," *1992 ASQC Annual Quality Congress Transactions,* Milwaukee, WI: ASQC, 158-164.

ACKNOWLEDGEMENTS

Special thanks go to Bob Rodriguez, Annette Sanders, and Richard Bynum of SAS Institute, Inc. and Donna Fulenwider of Meridian Software, Inc. for their contributions to this paper.

AUTHOR CONTACT

Melvin T. Alexander
Integrated Product Development
Westinghouse Electronic Systems Group
7323 Aviation Boulevard
P.O. Box 746, MS G16
Baltimore, MD 21203-0746
(410) 993-1478, voice; 765-1485, fax
INTERNET: Alexande@OPS1.bwi.wec.com

SherriJoyce King
King Information Company
16 Scotch Mist Court
Rockville, MD 20854-2929
(301) 251-6207, voice; 738-6873, fax
INTERNET: 72764.164@compuserve.com