

Control Charts for Attributes: Some Variations

Chin-liang Hung

1 Introduction

Control charts are one of the most widely used tools for monitoring the quality of a process. They are an integral part of quality improvement. A control chart consists of a plot of values of some quality characteristic, or subgroup statistic, over time together with limits that indicate the extent of common cause or system variation. From a control chart we can see if a particular process is stable and predictable over time. Such a process is often described as “in-control” or as being in a state of statistical control. A control chart that exhibits a random scatter of points around a center line and within limits of system or common cause variation, corresponds to a process that is said to be stable. If plotted points fall outside the limits set by system or common cause variation or an identifiable pattern over time exists, the process is said to be unstable. This instability is thought to be due to some special cause that then must be discovered and eliminated in order to create a more stable process.

Control charts can be constructed using measurement data, or subgroup statistics of measurement data. So called, variables control charts are very common in industry today. There are, however, many processes that have quality characteristics that are not measured. Instead, items are classified as “good” or “bad”, “defective” or “nondefective”, “conforming” or “nonconforming”. Such classifications result in count data which can be charted using so called attributes control charts. Variables control charts are often preferred because measurements carry more information than simple classifications of “conforming” and “nonconforming”. This shows up in the fact that variables charts can be quite effective with individual measurements or

small subgroups of measurements. Attributes control charts, in general, require larger subgroup sizes to be effective. Still, control charts for attributes are very popular and widely used.

This paper looks at control charts for attributes. In section 2 we discuss the standard control charts for count data. Sections 3 and 4 look at variations of these standard charts and give examples of their use. We end with some recommendations both for the use of attributes control charts and areas for future investigation.

2 Standard Control Charts for Attributes

Control charts for attributes are plots of counts (proportions) of items having certain attributes, conforming or nonconforming, or of counts (rates) of nonconformities per item. The distinction between number of nonconforming items and number of nonconformities per item is an important one. If we are interested in the number (proportion) of **items** that are nonconforming, we would use the np-chart (p-chart). If we are interested in the number (rate) of **nonconformities per item**, we would use the c-chart (u-chart). We will discuss each of these charts separately.

2.1 Number (proportion) of nonconforming items

The np-chart, or equivalently the p-chart, is used for data that consist of the number (proportion) of nonconforming items relative to the number of items inspected. Conceptually, as a process produces items over time subgroups consisting of n_t items are selected and inspected. Each item in a subgroup is classified as conforming or nonconforming according to some operational definition. For the np-chart, the number of nonconforming items in subgroup t , X_t , is the quantity plotted. For the p-chart, the proportion of nonconforming items $\hat{p}_t = \frac{X_t}{n_t}$ is plotted.

Theoretically, the number of nonconforming items, X_t can be thought of as a binomial random variable. This assumes that whether one item is con-

forming or nonconforming is independent of the condition of other items and that the true process proportion of nonconforming items is a constant value p . Under these assumptions, X_t has an expected value of $n_t p$ and a standard deviation of $\sqrt{n_t p(1-p)}$. For a process that is stable, or "in-control", the values of X_t should be randomly scattered around a center line set at the expected value and within ± 3 standard deviations of that center line. Since the process proportion of nonconforming items is not known, we must estimate it using the available data. The usual estimate is given by:

$$\bar{p} = \frac{\sum n_t \hat{p}_t}{\sum n_t} = \frac{\sum X_t}{\sum n_t}$$

Substituting this estimate into the expectation and standard deviation formulas above and using the conventional ± 3 "sigma" limits we have the following formulas for the np-chart:

$$\begin{aligned} \text{Center Line:} & \quad n_t \bar{p} \\ \text{Control Limits:} & \quad n_t \bar{p} \pm 3\sqrt{n_t \bar{p}(1-\bar{p})} \end{aligned}$$

A similar argument results in the following formulas for the p-chart:

$$\begin{aligned} \text{Center Line:} & \quad \bar{p} \\ \text{Control Limits:} & \quad \bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n_t}} \end{aligned}$$

These formulas allow for unequal subgroup sizes, n_t . When possible, equal subgroup sizes are preferred since they give constant values for the center line and control limits on the charts. This makes the charts easier to construct and interpret.

2.2 Number (rate) of nonconformities per item

The c-chart, or equivalently the u-chart, is used for data that consist of the number (rate) of nonconformities per item. Conceptually, items produced by a process are selected and inspected. The number of nonconformities, there should be an operational definition of what constitutes each type of

nonconformity, per item are counted. Let c_t be the number of nonconformities for the t^{th} item. If items are of different sizes, let k_t be the size of the t^{th} item. For the c-chart, c_t is plotted while for the u-chart, $u_t = \frac{c_t}{k_t}$ is plotted.

Theoretically, the number of nonconformities, c_t can be thought of as a Poisson random variable. This assumes that nonconformities are generated with intensity λ at random on the item according to a Poisson process. Essentially this means that occurrences of nonconformities are independent and that at most one nonconformity can occur in any sufficiently small area with positive probability. Under these assumptions c_t has expectation λ and standard deviation $\sqrt{\lambda}$. For a process that is stable, or "in-control", the values of c_t should be randomly scattered around a center line set at the expected value and within ± 3 standard deviations of that center line. Since the intensity with which nonconformities are produced is not known, we must estimate it using the available data. Assuming one has m items, the usual estimate is given by:

$$\hat{\lambda} = \frac{\sum c_t}{m}$$

Substituting this estimate into the expectation and standard deviation formulas above and using the conventional ± 3 "sigma" limits we have the following formulas for the c-chart:

$$\begin{aligned} \text{Center Line:} & \quad \hat{\lambda} \\ \text{Control Limits:} & \quad \hat{\lambda} \pm 3\sqrt{\hat{\lambda}} \end{aligned}$$

For the u-chart of the rate of nonconformities, the estimated average rate is:

$$\bar{u} = \frac{(\sum c_t)}{(\sum k_t)}$$

Again using the conventional ± 3 "sigma" limits, the following formulas are used for the u-chart:

$$\begin{aligned} \text{Center Line:} & \quad \bar{u} \\ \text{Control Limits:} & \quad \bar{u} \pm 3\sqrt{\frac{\bar{u}}{k_t}} \end{aligned}$$

3 Dealing with Large Subgroup Sizes

With any control chart for attributes the subgroup size has an effect on the values of the control limits. For the p-chart, control limits are inversely proportional to the subgroup size. Thus large subgroup sizes produce narrow control limits. This can give the impression that the process is wildly out of control. For example, a student ¹ in an Applied Statistics for Industry class conducted a study on the performance and reliability of an Internet access system. The study focused on the question: "How often do users get the requested answer from a company's Internet access system?" Specifically, proxy servers for a company handle all Internet access requests and count the number of successful (and unsuccessful) attempts. Table 1 gives the daily number of Internet access attempts and the number of such attempts that resulted in an error message for 20 consecutive days.

Table 1: Internet Access Attempts and Errors

Day	Attempts	Errors	Day	Attempts	Errors
1	412,670	42,104	11	446,823	40,405
2	395,736	40,286	12	431,661	44,198
3	401,765	35,399	13	434,353	39,047
4	395,422	97,981	14	406,232	39,455
5	422,223	45,346	15	402,454	48,292
6	433,234	43,699	16	403,312	47,720
7	396,788	24,752	17	387,782	53,173
8	411,383	45,391	18	355,500	49,474
9	423,348	39,179	19	372,441	45,222
10	474,053	48,680	20	415,813	40,583
			Total	8,222,993	910,386

Using the formulas given in Section 2.1, one calculates $\bar{p} = \frac{910,386}{8,222,993} = 0.11071$ which is the center line for the p-chart. Since the number of attempts varies with each day, the control limits will vary similarly. Table 2 gives the lower and upper control limits for each day.

¹Bertrum Carroll, Fall semester 1996

Table 2: p-chart Control Limits

Day	LCL	UCL	Day	LCL	UCL
1	0.10924	0.11218	11	0.10930	0.11212
2	0.10921	0.11221	12	0.10928	0.11214
3	0.10922	0.11220	13	0.10928	0.11214
4	0.10921	0.11221	14	0.10928	0.11219
5	0.10926	0.11216	15	0.10923	0.11219
6	0.10928	0.11214	16	0.10923	0.11219
7	0.10922	0.11220	17	0.10923	0.11222
8	0.10924	0.11218	18	0.10920	0.11229
9	0.10926	0.11216	19	0.10913	0.11225
10	0.10934	0.11208	20	0.10925	0.11217

As you can see from Tables 1 & 2, although the number of attempts varies considerably from day to day, the control limits are not that much different. They are, however, very close to the center line of $\bar{p} = 0.11071$. Figure 1 gives the p-chart as constructed by SAS/QC. Figure 2 gives the same p-chart as constructed by Minitab. Because of the very large subgroup sizes (attempts) virtually all of the plotted points fall outside the very narrow control limits.

Heimann (1996) discusses this situation and proposes the individual “measurements” chart as an alternative to the p-chart. The rationale behind this alternative comes from the variation present in the data. Heimann indicates that there are two kinds of variation. One is the sampling variation and the other is variation of the underlying process. The sampling variation σ_S^2 is that variation attributable to random sampling. Since we are dealing with nonconforming units, the binomial distribution can be used to describe this sampling variation. In particular, when n items are selected from a process with a probability p of producing a nonconforming item, then:

$$\sigma_S^2 = \frac{p(1-p)}{n}$$

There is always a sampling variance (unless n is infinite) and that variance becomes smaller with larger subgroup sizes. The sampling variance measures only the variation introduced by random sampling.

Figure 1: p-chart of fraction of errors
SAS/QC

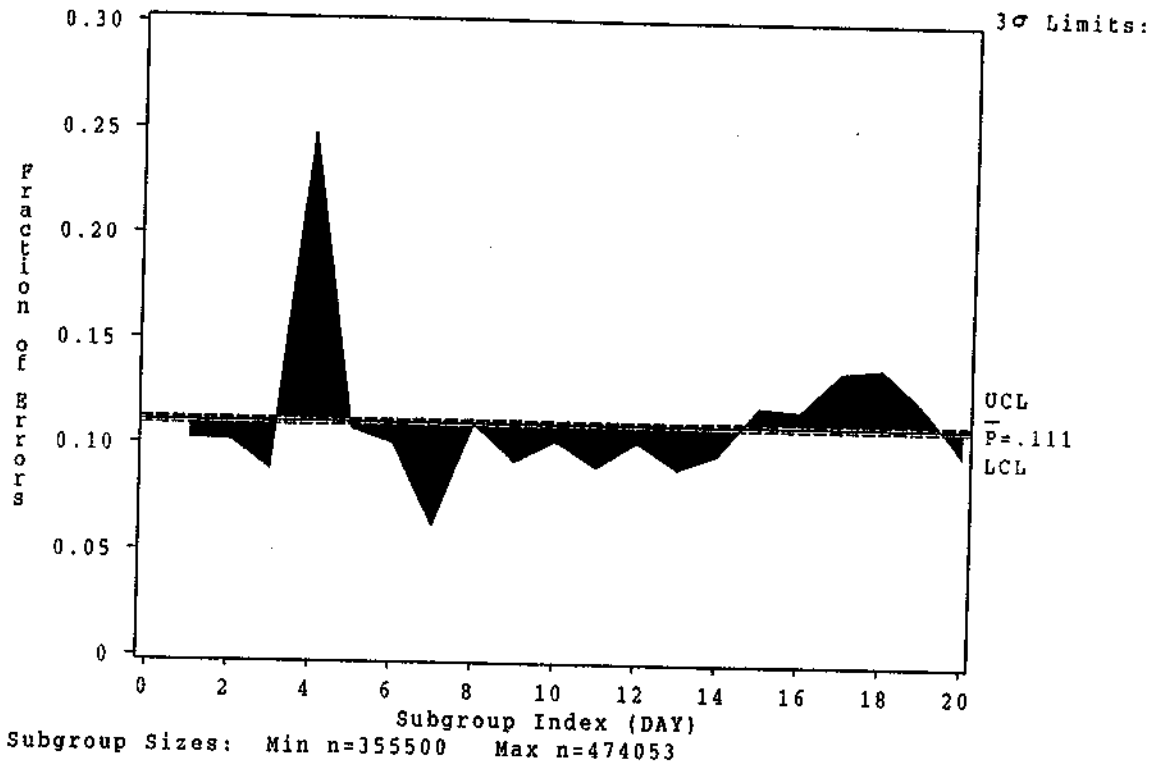
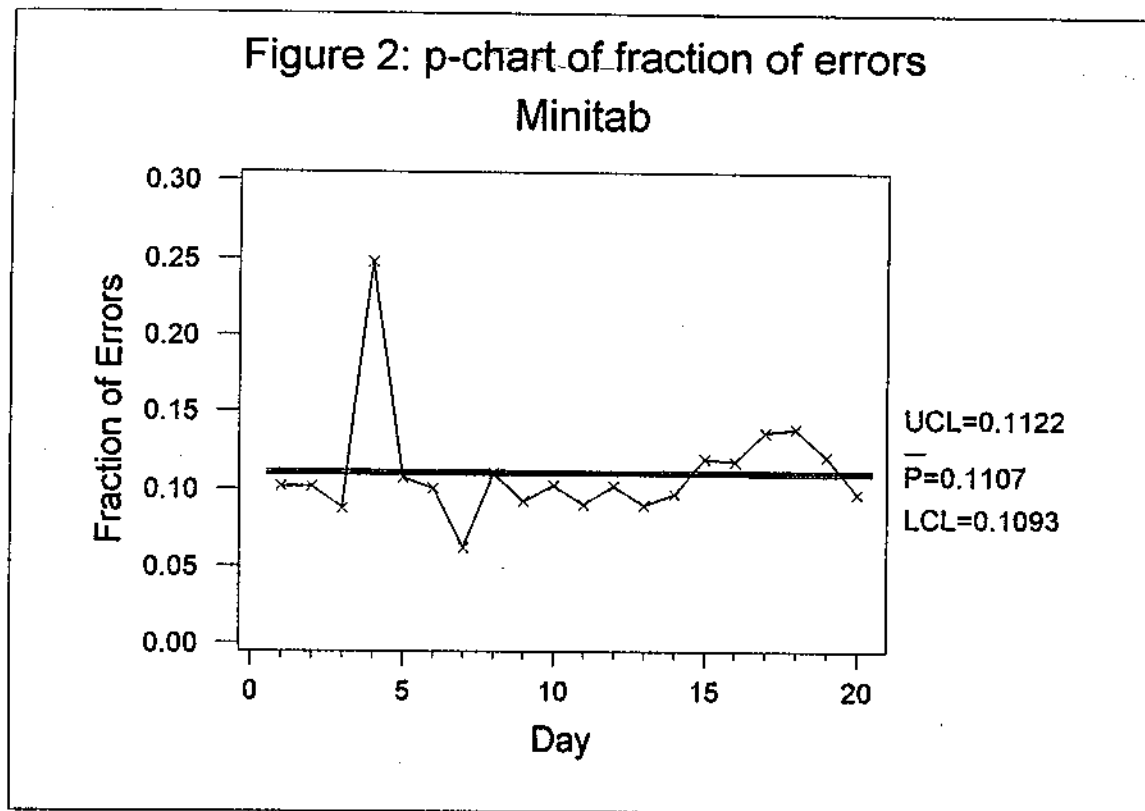


Figure 2: p-chart of fraction of errors
Minitab



Heimann argues that there is no perfect process and so the assumption of a constant probability of nonconforming items is never entirely correct. Instead, we can see day-to-day variation underlying the process. That is, the probability of nonconforming items varies around some mean value with variance σ_p^2 . We can not compute this mean or variance directly. However, we can use the total observed variance (sampling variation and the variation underlying the process) to get σ_p^2 indirectly.

As mentioned above, for attributes data we can not measure the variation underlying the process directly. We can observe the process output and use that to estimate the total variation, σ_T^2 . This total observed variation can be calculated by:

$$\hat{\sigma}_T^2 = \sum_{i=1}^m \frac{(\hat{p}_i - \bar{p})^2}{m-1} \quad (1)$$

Assuming that the sampling variation is independent of the underlying process variation, we have (at least approximately):

$$\hat{\sigma}_T^2 \doteq \sigma_S^2 + \sigma_p^2$$

The usual p-chart considers only sampling variation in setting the the control limits. With large subgroup sizes this produces the narrow limits and frequent alarms. As an alternative, Heimann suggests using the total observed variation to set control limits for the fraction nonconforming. The control limits would thus be:

$$\bar{p} \pm 3\hat{\sigma}_T \quad (2)$$

where $\hat{\sigma}_T$ is the square root of the quantity given by (1). Alternatively, the average moving range can be used to estimate σ_T^2 . This allows the practitioner to use standard computer programs to construct the individual measurements chart using the fraction nonconforming as the "measurements".

Returning to our example of the Internet access system, Table 3 gives the fraction of errors for the 20 days. Note that the number of attempts is no longer a factor and so some information is hidden in this alternative approach.

Table 3: Internet Access Fraction of Errors

Day	Fraction of Errors	Day	Fraction of Errors
1	0.10203	11	0.09043
2	0.10180	12	0.10239
3	0.08811	13	0.08990
4	0.24779	14	0.09712
5	0.10740	15	0.11999
6	0.10087	16	0.11832
7	0.06238	17	0.13712
8	0.11034	18	0.13917
9	0.09255	19	0.12142
10	0.10269	20	0.09760
		mean	0.11147

Heimann suggests that an individual measurements chart should be used, instead of the p-chart, if the ratio $r = \frac{\sigma_T^2}{\sigma_S^2}$ is greater than 1.357. This recommendation arises from the fact that the p-chart's narrower limits allow a "significant" proportion of points to fall outside those limits but those same points would fall within the control limits of an individual measurements chart. Recall, that the control limits of a p-chart are $\bar{p} \pm 3\sigma_S$ and those of an individual measurements chart are $\bar{p} \pm 3\sigma_T$. If the individual measurements chart limits are at the quantiles $z = \pm 3$ then the quantiles of the corresponding p-chart are at $z = \pm \left(\frac{\sigma_S}{\sigma_T}\right) = \pm \frac{3}{\sqrt{r}}$. If $Q(z)$ is the area under the standard normal curve between x and ∞ , then the probability of falling outside the limits of the p-chart is given by

$$\alpha = 2Q(z) = 2Q\left(\frac{3}{\sqrt{r}}\right)$$

This leads to the following relationship between the "significant" proportion outside the narrow limits of the p-chart and the ratio r .

1. For $\alpha = 0.05$, the value of r should be less than 2.343 to have fewer than 5% of the points outside the control limits of the p-chart.
2. For $\alpha = 0.01$, the value of r should be less than 1.357 to have fewer than 1% of the points outside the control limits of the p-chart.

3. For $\alpha = 0.0027$, the value of r should be less than 1.000 to have fewer than 0.27% of the points outside the control limits of the p-chart.

Heimann suggests that the criterion using $\alpha = 0.01$ and, therefore using an individual measurements chart whenever r is greater than 1.357.

The value of r is not known but it can be estimated by first estimating σ_T^2 and σ_S^2 . For our example, we will use $\bar{p} = 0.11147$ and an average value of $n = 411,150$.

$$\hat{\sigma}_T^2 = \sum_{i=1}^m \frac{(\hat{p}_i - \bar{p})^2}{m-1} = 0.001333$$

$$\hat{\sigma}_S^2 = \frac{\bar{p}(1-\bar{p})}{n} = 0.000000241$$

$$\hat{r} = \frac{\hat{\sigma}_T^2}{\hat{\sigma}_S^2} = 5,531$$

According to Heimann's criterion, an individual measurements chart is appropriate.

According to equation (2) the control limits should be calculated using the estimate of σ_T . If the process is in a state of statistical control, this should be similar to using the average moving range to establish the control limits. In our example, there is an unusual day, day 4, that could be the result of a special cause. Table 4 summarizes the various center lines and control limits depending on whether one uses $\hat{\sigma}_T$ or the average moving range and whether or not one includes day 4 in the calculations.

Table 4: Control limits for an individual measurements chart

Variance estimate	All data	Excluding day 4
$\hat{\sigma}_T$	UCL=0.22098	UCL=0.15795
	CL =0.11147	CL =0.10430
	LCL=0.00196	LCL=0.05064
\overline{MR}	UCL=0.19067	UCL=0.14587
	CL =0.11147	CL =0.10430
	LCL=0.03227	LCL=0.06272

One can see that the one unusual day, day 4, inflates the estimate of σ_T more than it affects the value of the average moving range. When day 4 is removed, the two ways of estimating variance give much more comparable values for the upper and lower control limits of the individual measurements chart.

Figure 3 gives the individual measurements chart as constructed by SAS/QC using the moving range to establish the control limits. Figure 4 gives the individual measurements chart as constructed by Minitab using the estimated value of $\hat{\sigma}_T$ to establish the control limits. Both charts show relative stability over time except for day 4. This is quite a different picture than the one given in Figures 1 & 2. The large subgroup sizes in the original p-chart create control limits that are too narrow. The individual measurements chart essentially ignores the information on the size of each subgroup and treats the subgroup fractions as individual "measurements".

4 Dealing with Clusters of Nonconformities

There are often situations where nonconformities are not randomly dispersed but tend to appear in clusters. In these situations the Poisson model assumed for the c- and u-chart is inappropriate. A more appropriate model is given by the negative binomial distribution. There are several theoretical situations that lead to the negative binomial distribution. In one situation, nonconformities occur according to a Poisson distribution however the parameter λ is itself a random variable with a gamma distribution. Another situation views nonconformities occurring in clusters with the number of clusters having a Poisson distribution and the number of nonconformities in each cluster following a logarithmic series distribution. Both of these situations, as we will see below, give us a random variable for the number of nonconformities that follows a negative binomial distribution. A negative binomial random variable can also be generated by summing independent geometric random variables. It is this latter relationship that leads Kaminsky, Benneyan, Davis and Burke (1992) to propose the g- and h-chart for monitoring counts that follow a negative binomial distribution.

Figure 3: Individuals chart
SAS/QC

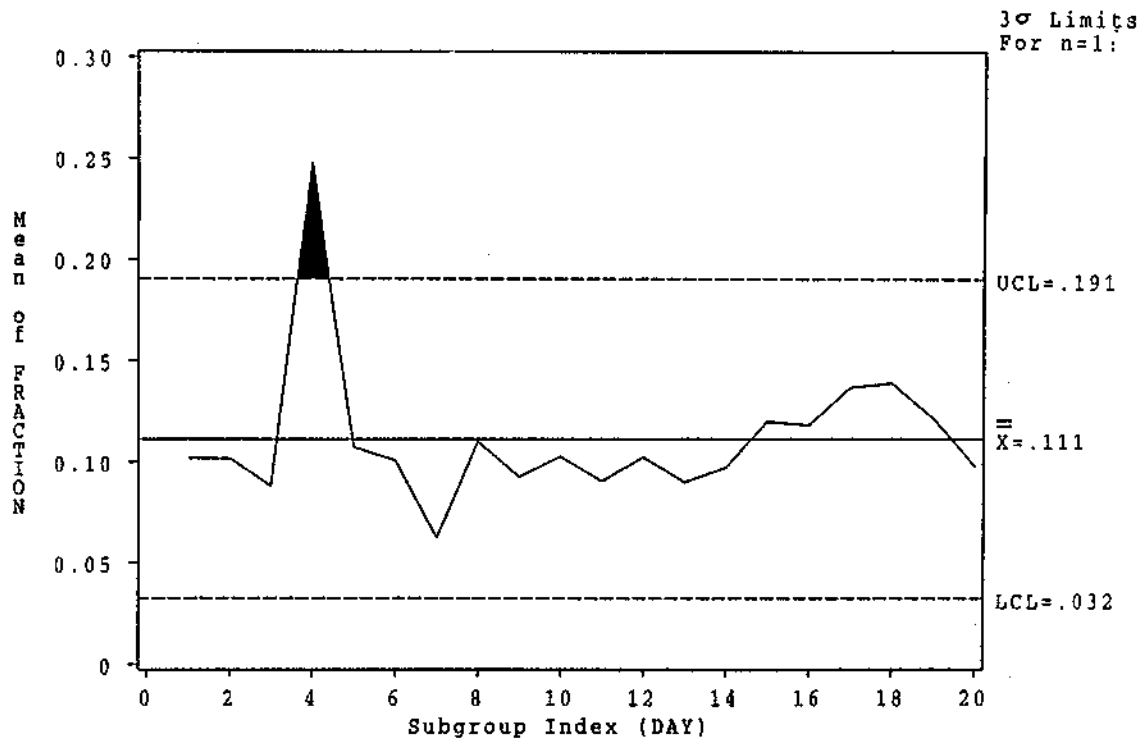
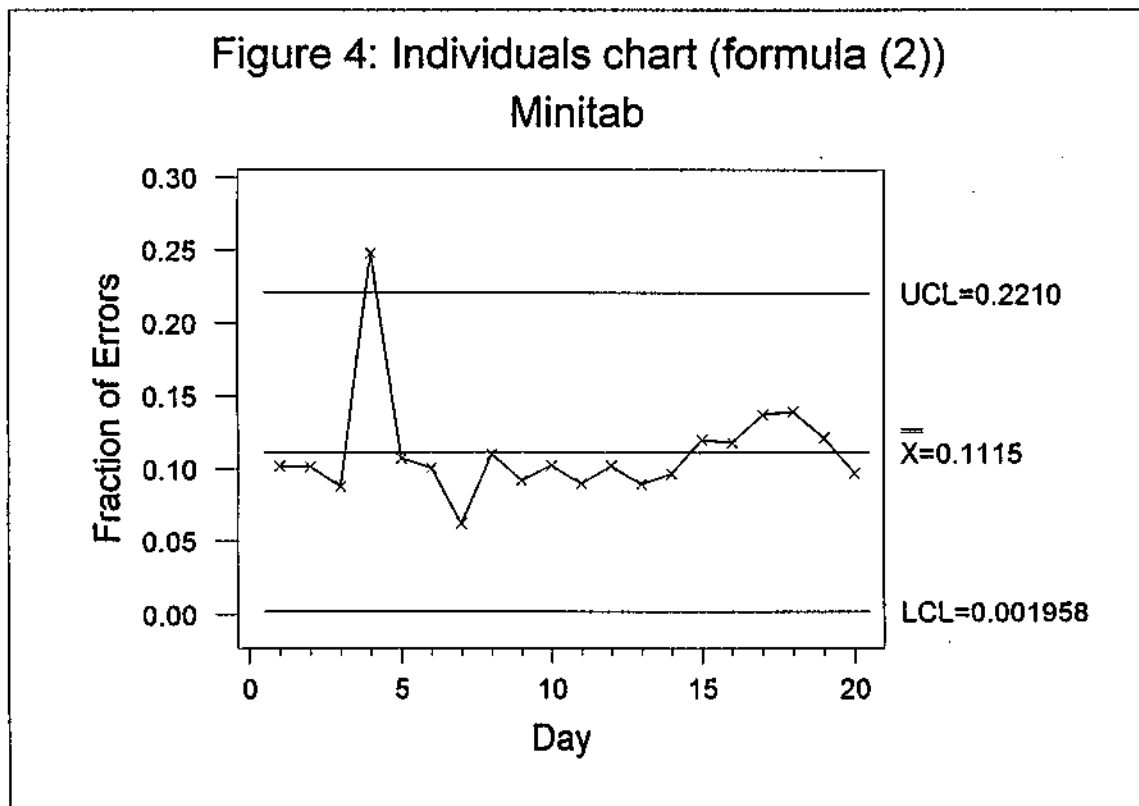


Figure 4: Individuals chart (formula (2))
Minitab



4.1 Poisson/gamma Model

Let X_1, X_2, \dots, X_n be independent discrete random variables where X_i represents the number of nonconformities on a single unit. One method of deriving the negative binomial model is to assume that X_i has a Poisson distribution with parameter λ but λ is itself a random variable having a gamma distribution. Then the marginal distribution of X_i is a negative binomial distribution. To see this situation, we have the probability function of X_i given λ as:

$$p_i(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \quad x = 0, 1, 2, \dots$$

with $\lambda \sim \text{gamma}(\alpha, \beta)$, the probability density function for λ is given by:

$$p(\lambda) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \lambda^{\alpha-1} e^{-\lambda/\beta} \quad 0 \leq \lambda < \infty$$

The marginal distribution of X_i is given by

$$\begin{aligned} p_i(x) &= \int_0^\infty p(x|\lambda)p(\lambda)d\lambda \\ &= \int_0^\infty \frac{\lambda^x e^{-\lambda}}{x!} \frac{1}{\Gamma(\alpha)\beta^\alpha} \lambda^{\alpha-1} e^{-\lambda/\beta} d\lambda \\ &= \frac{1}{x!\Gamma(\alpha)\beta^\alpha} \int_0^\infty \lambda^{(x+\alpha)-1} e^{-\lambda(1+\frac{1}{\beta})} d\lambda \\ &= \frac{\Gamma(x+\alpha)}{x!\Gamma(\alpha)\beta^\alpha \left(1+\frac{1}{\beta}\right)^{x+\alpha}} \int_0^\infty \frac{1}{\Gamma(x+\alpha)} \left(1+\frac{1}{\beta}\right)^{x+\alpha} \lambda^{(x+\alpha)-1} e^{-\lambda(1+\frac{1}{\beta})} d\lambda \end{aligned} \quad (3)$$

The value of the integral is one since it is the integral of a probability density function, *e.g.* $\text{gamma}(x+\alpha, \frac{\beta}{1+\beta})$.

$$\begin{aligned} p_i(x) &= \frac{\Gamma(x+\alpha)}{x!\Gamma(\alpha)\beta^\alpha \left(1+\frac{1}{\beta}\right)^{x+\alpha}} \\ &= \frac{\Gamma(x+\alpha)}{x!\Gamma(\alpha)} \beta^{-\alpha} \left(\frac{\beta}{1+\beta}\right)^{x+\alpha} \\ &= \binom{x+\alpha-1}{x} \left(\frac{\beta}{1+\beta}\right)^x \left(\frac{1}{1+\beta}\right)^\alpha \quad x = 0, 1, 2, \dots \end{aligned} \quad (4)$$

This is the probability function of a negative binomial random variable with parameters α and β . (See Johnson and Kotz (1969), Chapter 5.) One can easily get the expected value and variance of this negative binomial random variable either directly or as follows:

$$E(X) = E(E(X|\lambda)) = E(\lambda) = \alpha\beta$$

$$Var(X) = Var(E(X|\lambda)) + E(Var(X|\lambda)) = Var(\lambda) + E(\lambda) = \alpha\beta(1 + \beta)$$

4.2 Poisson/logarithmic series Model

Now, if we assume that the number of clusters, N , follows a Poisson distribution and that the number of nonconformities in each cluster, X_i , has a logarithmic series distribution, then we are interested in the random variable $H = X_1 + X_2 + \dots + X_N$ where $N \sim Poisson(\lambda)$ and

$$p_i(X = x) = \frac{-1}{\log(p)} \frac{(1-p)^x}{x} \quad x = 1, 2, \dots$$

We will show below that H has a negative binomial distribution by computing the moment generating function of H .

$$\begin{aligned} E(e^{Ht}) &= E(E(e^{Ht}|N)) \\ &= E(E(e^{(X_1+X_2+\dots+X_N)t}|N)) \\ &= E([E(e^{X_1t})]^N) \end{aligned} \tag{5}$$

The last step in (5) follows from the fact that the moment generating function of a sum of independent random variables is equal to the product of the individual moment generating functions. Now,

$$\begin{aligned} E(e^{X_1t}) &= \sum_{x_1=1}^{\infty} e^{x_1t} \frac{-1}{\log(p)} \frac{(1-p)^{x_1}}{x_1} \\ &= \frac{-1}{\log(p)} \sum_{x_1=1}^{\infty} \frac{(e^t(1-p))^{x_1}}{x_1} \\ &= \frac{-1}{\log(p)} (-\log[1 - e^t(1-p)]) \end{aligned} \tag{6}$$

The last step uses the fact that the summation is that of the logarithmic series, hence the name for the distribution. Now, replacing (6) into (5) we

have:

$$\begin{aligned}
E(e^{Ht}) &= E\left(\left[\frac{\log[1 - e^t(1-p)]}{\log(p)}\right]^N\right) \\
&= \sum_{n=0}^{\infty} \left(\frac{\log[1 - e^t(1-p)]}{\log(p)}\right)^n \frac{e^{-\lambda} \lambda^n}{n!} \\
&= e^{-\lambda} e^{\frac{\lambda \log[1 - e^t(1-p)]}{\log(p)}} \sum_{n=0}^{\infty} \frac{e^{\frac{-\lambda \log[1 - e^t(1-p)]}{\log(p)}} \left(\frac{\lambda \log[1 - e^t(1-p)]}{\log(p)}\right)^n}{n!}
\end{aligned} \tag{7}$$

The sum in the last step of equation (7) equals 1 since it is the sum of all the probabilities for a Poisson random variable with parameter $\frac{\lambda \log[1 - e^t(1-p)]}{\log(p)}$. Therefore,

$$\begin{aligned}
E(e^{Ht}) &= e^{-\lambda} e^{\frac{\lambda \log[1 - e^t(1-p)]}{\log(p)}} \\
&= [e^{-\log(p)}]^{\lambda/\log(p)} [1 - e^t(1-p)]^{\lambda/\log(p)} \\
&= \left[\frac{1}{p} - e^t \left(\frac{1-p}{p}\right)\right]^{\lambda/\log(p)}
\end{aligned} \tag{8}$$

Equation (8) is the moment generating function of a negative binomial random variable with parameters $-\lambda/\log(p)$ and $\frac{1-p}{p}$. Again, it is easy to show that the expected value and variance of this negative binomial random variable are, respectively,

$$\begin{aligned}
E(H) &= \frac{-\lambda(1-p)}{\log(p)p} \\
Var(H) &= \frac{-\lambda(1-p)}{\log(p)p^2}
\end{aligned}$$

4.3 Geometric Model

Kaminsky, Benneyan, Davis and Burke (1992) assume that events are generated according to a geometric distribution. That is, if X is the number of nonconformities per process unit, then the probability function of X is given by:

$$Pr[X = x] = p(1-p)^{x-a} \quad \text{for } x = a, a+1, a+2, \dots$$

where a is a known minimum number of nonconformities. For the remainder of the development we will assume $a = 0$. The total number of nonconformities is given by $T = X_1 + X_2 + \dots, X_n$ and the average number of nonconformities is given by $\bar{X} = T/n$. The sum of independent and identically distributed geometric random variables is distributed as a negative binomial. This can be shown using moment generating functions. The probability function for T is given by:

$$Pr[T = t] = \binom{n+t-1}{n-1} p^n (1-p)^t \quad \text{for } t = 0, 1, 2, \dots$$

Since the event $[\bar{X} = t/n]$ is identical to the event $[T = t]$, the distribution of \bar{X} is also a negative binomial. The expected values and variances of these statistics are:

$$\begin{aligned} E(T) &= \frac{n(1-p)}{p} \\ Var(T) &= \frac{n(1-p)}{p^2} \\ E(\bar{X}) &= \frac{(1-p)}{p} \\ Var(\bar{X}) &= \frac{(1-p)}{np^2} \end{aligned} \tag{9}$$

It is easy to see that this is just a special case of the Poisson/gamma model with $\alpha = n$ and $\beta = \frac{1-p}{p}$ or a special case of the Poisson/logarithmic series model with parameters $\frac{-\lambda}{\log(p)}$ and $\frac{1-p}{p}$. Using the expectations and variances in (9), a centerline and the 3“sigma” limits can be calculated for charting T , the g-chart, or \bar{X} , the h-chart. Table 5 summarizes these formulas.

Table 5: Control chart formulas for the g- and h-charts

	g-chart	h-chart
Centerline (CL)	$\frac{n(1-p)}{p}$	$\frac{1-p}{p}$
Upper Control Limit (UCL)	$\frac{n(1-p)}{p} + 3\sqrt{\frac{n(1-p)}{p^2}}$	$\frac{1-p}{p} + \frac{3}{\sqrt{n}}\sqrt{\frac{1-p}{p^2}}$
Lower Control Limit (LCL)	$\frac{n(1-p)}{p} - 3\sqrt{\frac{n(1-p)}{p^2}}$	$\frac{1-p}{p} - \frac{3}{\sqrt{n}}\sqrt{\frac{1-p}{p^2}}$

In practice, the value of p is not known and must be estimated. Using either the method of moments or the method of maximum likelihood, the estimator of p is given by:

$$\hat{p} = \frac{1}{\bar{\bar{X}} + 1}$$

where $\bar{\bar{X}}$ is the average number of nonconformities for r subgroups each of size n . Formulas for the sample based limits for the g- and h-charts are given in Table 6.

Table 6: Sample based formulas for the g- and h-charts

Centerline (CL)	$n\bar{\bar{X}}$	$\bar{\bar{X}}$
Upper Control Limit (UCL)	$n\bar{\bar{X}} + 3\sqrt{n\bar{\bar{X}}(\bar{\bar{X}} + 1)}$	$\bar{\bar{X}} + \frac{3}{\sqrt{n}}\sqrt{\bar{\bar{X}}(\bar{\bar{X}} + 1)}$
Lower Control Limit (LCL)	$n\bar{\bar{X}} - 3\sqrt{n\bar{\bar{X}}(\bar{\bar{X}} + 1)}$	$\bar{\bar{X}} - \frac{3}{\sqrt{n}}\sqrt{\bar{\bar{X}}(\bar{\bar{X}} + 1)}$

4.4 Example

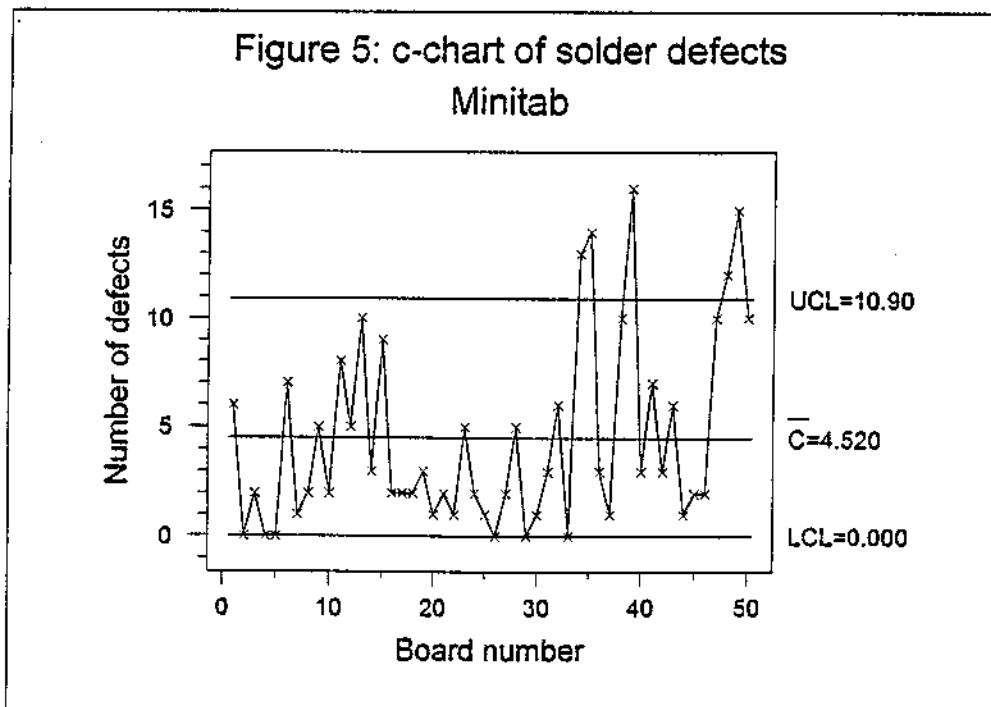
As integrated circuits (ICs) continue to grow in area and density, the issue of defect clusters becomes more important. Currently, IC board manufacturers assume a Poisson distribution for nonconformities. This approach gives accurate results for assemblies that are not very complex. However, for complex boards the Poisson model tends to underestimate the variability in the number of nonconformities due to clustering. In particular, solder defects tend to be clustered. Table 7 gives data on the number of solder defects for 50 boards. In the notation used in section 4.3, there are $r = 50$ subgroups of size $n = 1$.

The overall mean number of solder defects is $\bar{\bar{X}} = 4.52$. Using the formulas for the center line and control limits for the c-chart given in section 2.2 we have:

$$\begin{aligned} \text{Center Line:} & \quad 4.52 \\ \text{Control Limits:} & \quad 4.52 \pm 3\sqrt{4.52} \end{aligned}$$

Table 7: Number of solder defects for 50 circuit boards

Board	Defects	Board	Defects	Board	Defects	Board	Defects
1	6	16	2	31	3	46	2
2	0	17	2	32	6	47	10
3	2	18	2	33	0	48	12
4	0	19	3	34	13	49	15
5	0	20	1	35	14	50	10
6	7	21	2	36	3		
7	1	22	1	37	1		
8	2	23	5	38	10		
9	5	24	2	39	16		
10	2	25	1	40	3		
11	8	26	0	41	7		
12	5	27	2	42	3		
13	10	28	5	43	6		
14	3	29	0	44	1		
15	9	30	1	45	2		



Examination of Figure 5 and Table 7, indicates that there are several boards that fall outside the control limits, specifically boards 34, 35, 39, 48 and 49. However one should also note that the sample variance of the fifty observations is 18.87, much greater than what would be expected under the Poisson model. Looking at the histogram of the number of solder defects (Figure 6) one sees a highly skewed distribution. Comparing this to a histogram of simulated Poisson counts with mean 4.52 (Figure 7) one sees that the number of solder defects does not match the Poisson distribution. To formalize this comparison we perform a χ^2 goodness of fit test. This test is summarized in Table 8.

Table 8: χ^2 goodness of fit test for solder data.

Interval	Poisson Probability	Expected Number	Observed Number	Contribution to χ^2
$0 \leq X < 2$	0.0601	3.005	13	33.24
$2 \leq X < 4$	0.2788	13.940	17	0.67
$4 \leq X < 6$	0.3606	18.030	4	10.92
$6 \leq X < 8$	0.2123	10.615	5	2.97
$8 \leq X < \infty$	0.0882	4.41	11	9.95
				$\chi^2 = 57.65$

The calculated value of the goodness of fit test is $\chi^2 = 57.65$ which exceeds the critical value of 7.81 based on 3 degrees of freedom and $\alpha = 0.05$. This verifies our visual comparison of the histograms in Figures 6 and 7 and suggests that something other than a Poisson distribution describes the solder defects.

Given the extra variation in the data, a negative binomial distribution may fit the data better. Since there is only one board per subgroup, a geometric distribution can be used. Using the result that

$$\hat{p} = \frac{1}{\bar{X} + 1}$$

we perform a goodness of fit test using geometric probabilities instead of Poisson. The results are summarized in Table 9.

Figure 6: Histogram of solder defects

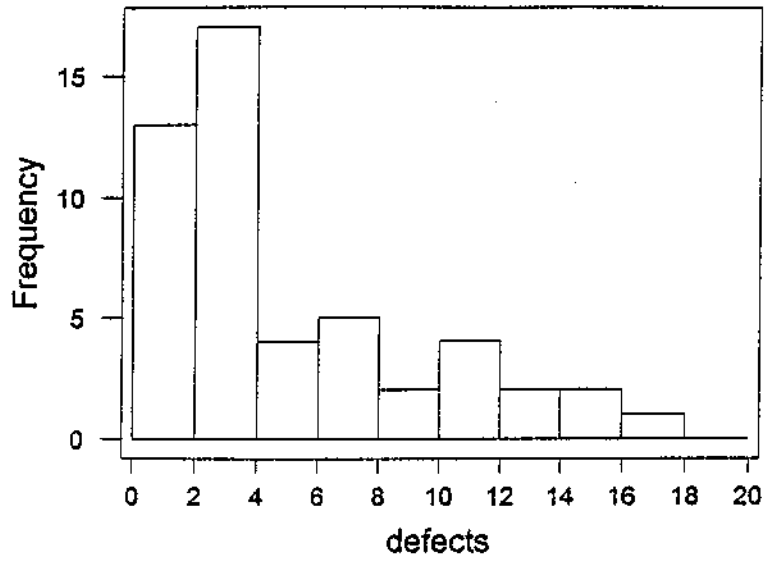


Figure 7: Histogram of simulated Poisson counts
mean=4.52

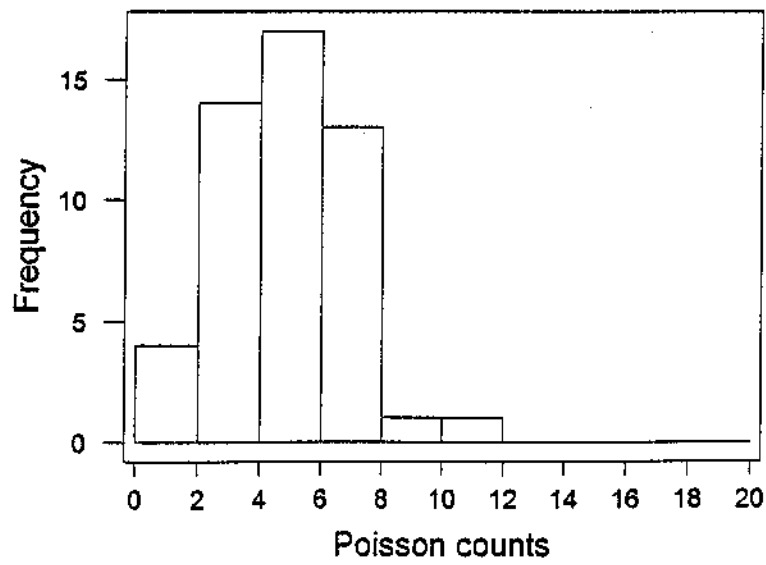


Table 9: χ^2 goodness of fit test for solder data.

Interval	Geometric Probability	Expected Number	Observed Number	Contribution to χ^2
$0 \leq X < 2$	0.3295	16.475	13	0.73
$2 \leq X < 4$	0.2209	11.045	17	3.21
$4 \leq X < 6$	0.1481	7.405	4	1.57
$6 \leq X < 8$	0.0993	4.965	5	0.00
$8 \leq X < \infty$	0.2022	10.110	11	0.08

$\chi^2 = 5.59$

This time the calculated value of the test statistic, $\chi^2 = 5.59$, does not exceed the critical value of 7.81. This indicates that a geometric distribution does fit the data.

Constructing the g-chart for these data one has the same centerline, $CL=4.52$, as with the c-chart. However, the extra variation shows up in terms of wider control limits. The UCL for the g-chart is 19.505 compared to the UCL for the c-chart of 10.90. With the wider control limits none of the boards fall outside these bounds. Some may argue that this is bad since there were five boards that were identified as unusual using the c-chart but now are 'in-control' using the g-chart. More important is the interpretation of the respective charts. If the c-chart is used, line personnel will investigate the five boards with the larger numbers of defects and try to identify a special cause. Recall that special causes are not a part of the system all of the time but arise because of special circumstances. With the g-chart one views the clustering as part of the system and so contributing to the common cause variation. Unless the system is changed, clustering will continue to occur. The g-chart properly focusses attention on changing the system to avoid clustering of defects. Figure 8 gives a version of the g-chart produced using SAS Graph.

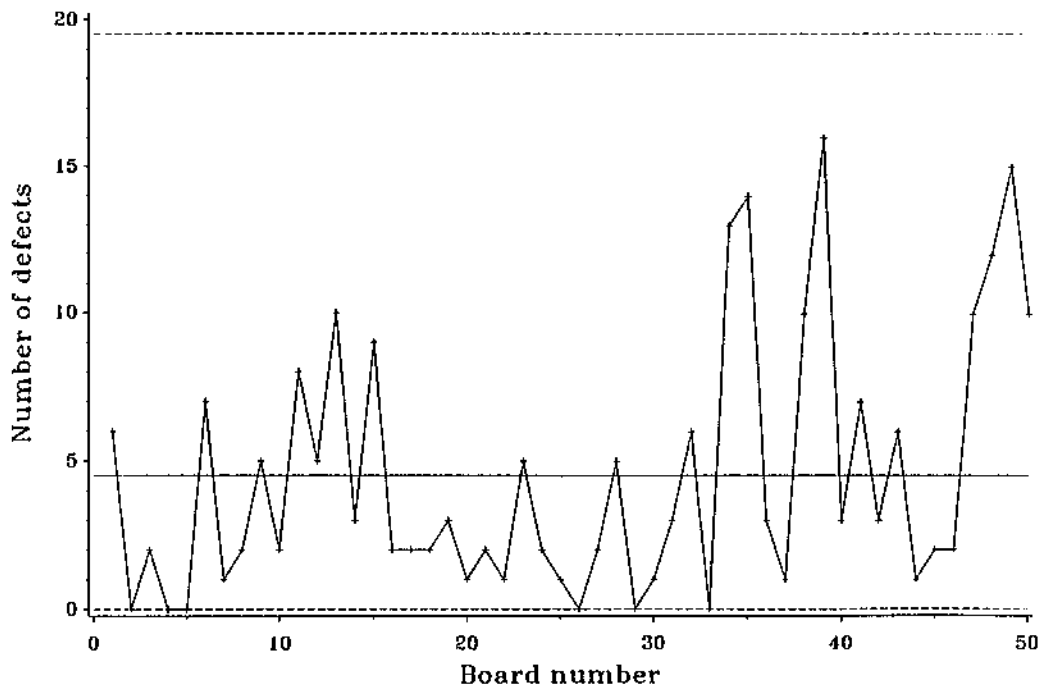
5 Conclusion

This paper has looked at several aspects of control charts for attribute data. The standard charts, p- and np-charts for the number of nonconforming items, and c- and u-charts for the number of nonconformities per item were reviewed. Two extensions to these standard charts were investigated. The first

involved the number of nonconforming items where the subgroup sizes were large. The second involved extra variation in the number of nonconformities per item. Both extensions pointed out the need for statistical practioners to carefully view the context and the nature of the data before constructing an attributes control chart. Rote application of standard methods can lead to misinterpretation of the data and incorrect decisions about special and common causes of variability.

Attributes control charts are well known and much used monitoring devices. The paper by Woodall (1997) provides an extensive bibliography of research into this area. The reader is directed to that paper for a discussion of topics for future research and other recommendations.

Figure 8: g-chart of solder defects
SAS Graph



6 References

- Bissell, A.F. (1988), "Control chart limits for attributes and events," *Journal of Applied Statistics*, 15, 97-105.
- Goh, T.M. (1987), "A Charting Technique for Control of Low-Defective Production," *International Journal of Quality & Reliability Management*, 4, 53-62.
- Heimann, P.A. (1996), "Attributes Control Charts with Large Sample Sizes," *Journal of Quality Technology*, 28, 451-459.
- Jackson, J.E. (1972), "All Count Distributions Are Not Alike," *Journal of Quality Technology*, 4, 86-92.
- Johnson, N.L. and Kotz, S. (1969), *Discrete Distributions*, New York, Wiley.
- Kaminsky, F.C., Benneyan, J.C., Davis, R.D. and Burke, R.J. (1992), "Statistical Control Charts Based on a Geometric Distribution," *Journal of Quality Technology*, 24, 63-69.
- Nelson, L.S. (1994), "A Control Chart for Parts-Per-Million Nonconforming Items," *Journal of Quality Technology*, 26, 239-240.
- Ryan, T.P. and Schwertman, N.C. (1997), "Optimal Limits for Attributes Control Charts," *Journal of Quality Technology*, 29, 86-98.
- Sheaffer, R.L. and Leavenworth, R.S. (1976), "The Negative Binomial Model for Counts in Units of Varying Size," *Journal of Quality Technology*, 8, 158-163.
- Schwertman, N.C. and Ryan, T.P. (1997), "Implementing Optimal Attributes Control Charts," 29, 99-104.
- Winterbottom, A. (1993), "Simple Adjustments to Improve Control Limits on Attribute Charts," *Quality and Reliability Engineering International*, 9, 105-109.
- Woodall, W.H. (1997), "Control Charts Based on Attribute Data: Bibliography and Review," *Journal of Quality Technology*, 29, 172-183.