DOCUMENT RESUME

ED 306 285 TM 013 146

AUTHOR	Blumberg, Carol Joyce
TITLE	Applying Statistical Process Quality Control Methodology to Educational Settings.
PUB DATE	Mar 89
NOTE	40p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Charts; Data Analysis; Educational Research; Evaluation Methods; Graphs; *Hypothesis Testing; *Quality Control; *Research Methodology; *Statistical Analysis; Statistical Inference; Statistical Studies
IDENTIFIERS	*Statistical Process Control

ABSTRACT

A subset of Statistical Process Control (SPC) methodology known as Control Charting is introduced. SPC methodology is a collection of graphical and inferential statistics techniques used to study the progress of phenomena over time. The types of control charts covered are the null X (mean), R (Range), X (individual observations), MR (moving range), p (measure of successes), and U (number of events per unit). Uses of charts in the binomial case and the Poisson case are considered. These charts provide the researcher with a combined graphical and hypothesis testing procedure for assessing the stability of various types of data collected across time. The details of constructing each of these charts and the methods for hypothesis testing are described. Examples drawn from the field of education illustrate chart construction. SPC procedures described are understandable by researchers who are not mathematically oriented; whereas other procedures often recommended in the statistically-oriented literature on single-subject design are not easily understood by these researchers. All of the SPC techniques introduced can be modified to allow the analysis of data from one-way multiple group designs and factorial designs. Although these charts do not represent all the control charting techniques used in process control, they are the ones most applicable to educational research. Five tables present sample data, and seven graphs illustrate SPC charts. (SLD)

Applying Statistical Process Quality Control Methodology to Educational Settings

Carol Joyce Blumberg
Department of Mathematics and Statistics
Winona State University
Winona, MN 55987

Paper presented at the 1989 Annual Meeting of the American Educational Research Association, San Francisco, March 27-31, 1989

ABSTRACT

The purpose of this paper is to introduce the reader to some of the control-charting techniques used in Statistical Process Control. The types of control charts covered are the $\bar{X}$-, R-, X-, MR-, p- and u-charts. These charts provide the researcner with a combined graphical/hypothesis testing procedure for assessing the stability of various types of data collected across time. The paper describes the details of constructing each of these charts and the methods for hypothesis testing. Educational examples are given throughout.

Introduction

Statistical Process Control (SPC) methodology is a collection of graphical and inferential statistics techniques used to study the progress of phenomena (e.g., the output of a particular machine or a specific behavior for an emotionally impaired youngster) over time. SPC began in the 1930's with the development of control charting techniques by Shewhart (1931), but it has not been until the last 15 years that SPC has become popular in the United States, although it has been used continuously in some American industries (e.g., Western Electric) since the 1930's. It should be noted that the popular use of SPC in Japan began approximately 40 years ago. Further, the use of SPC techniques both in Japan and the United States seems to have been restricted in the past, as well as the present, almost entirely to industrial settings.

The popularity of the SPC methodology in industrial settings can be seen by examining introductory statistics textbooks written for business and engineering students. Almost all of these textbooks have from a section to one or more chapters dealing with SPC techniques. To the best of the author's knowledge, statistics textbooks in education, the behavioral sciences and the social sciences do not, however, mention these techniques at all. This even includes books with chapters on the analysis of longitudinal data. Further, a computer-assisted search of the education literature (via the ERIC system) revealed only two published articles and no unpublished literature that dealt with any aspect of SPC. Each of the articles only dealt with a very narrow application of only the same small piece of SPC methodology. Hence, there is a need for an expository paper that introduces SPC methodology to the educational research community.

The main purpose of this paper is to introduce the reader to a large subset of Statistical Process Control methodology known as Control Charting. Some of the more popular types of control charts used presently in Statistical Process Control will be described in detail and ways in which these charts might be used to analyze educational data will be suggested. Where appropriate, the theoretical statistical underpinnings for the control charts will be discussed. It should be noted up front that the author has only recently learned the details of control charting. Hence, this paper is more of a beginning rather than a definitive review of the methodology. It is hoped, however, that others will still find this paper useful.

$\bar{X}$- and R-charts

The most commonly used SPC technique is the combination of the $\bar{X}$-chart and the R-chart. The data collected for use with this technique are of the following form. A constant number, n, of observations, with $n \geq 2$, are collected on some process at a number of time points, k, usually with $k \geq 25$. The time points, may or may not be, equally spaced. The observations at each time point must arise from a well-defined process and must be a sample from the total (i.e., population) output of the process. This sample may be a simple random, systematic, cluster, or stratified sample or arise from some combination of these sampling methods. The dependent variable measured on each unit (e.g., a part or a child) should have an underlying continuous distribution or, at the minimum, should be measured on an interval scale. Some examples of data sets where the technique of $\bar{X}$- and R-charting is appropriate are:

1. The weights of or number of pieces in a box of candy.
2. The thickness of a metallic coating.
3. Daily/weekly quiz scores for a class.
4. Weights of people using some behavioral modification technique.

As the data are collected, the mean (denoted $\bar{X}_i$) and the Range (R_i) are computed for each sample, $i = 1, 2, \ldots, k$.
After a sufficient number of baseline samples (usually $k \geq 25$) are collected an estimator of the population standard deviation is computed based on the R_i's. Although estimating the standard deviation using the ranges may seem strange at first, it is actually an excellent estimation technique. First, $\bar{X}$- and R-charts are often compiled by workers on production lines as part of their duties each shift. The computing of the range takes a

few seconds, while the computing of a standard deviation (even with a calculator with statistical capabilities) takes much longer. Second, the range can be computed even without a calculator, if absolutely necessary, by most production workers. In educational settings, as will be discussed more below, one of the more obvious applications of control charting is in the analysis of data arising from single-subject designs implemented in a school or other institutional setting. The data are often collected by teachers or other professional personnel to help them evaluate the behavior(s) of a student or small group of students. For these professionals, computing the range instead of the standard deviation advantageous time-wise because of the multiple demands of their time. Third, the concept of a range is easier for people to understand than is the concept of a standard deviation. Besides its simplicity, the use of ranges leads to estimators of the population standard deviation that have desirable statistical properties. For the following discussion and for much of the remainder of this paper, the focus will be on the analysis of data collected when no treatments are being applied (i.e., when baseline data are being collected). The question of how to detect treatment effects will be addressed later in this paper.

Statistical Background

Let σ_X denote the population standard deviation for the dependent variable of interest. Construct a new variable, W_i , with $W_i = R_i/\sigma_X$. Interestingly, when the X_i's are distributed independently as normal random variables with identical means and variances (i.e., they are i.i.d.), the expected value (i.e., population mean) of each of the W_i's is solely a function of the sample size, n (Montgomery, 1985). Denote this expected value by the symbol d_2. So, $d_2 = E(W_i) = E(R_i/\sigma_X) = E(R_i)/\sigma_X$. Hence, $\sigma_X = E(R_i)/d_2$. Summing over the k samples, $\sigma_X = (E(\sum_{i=1}^{k} R_i/k))/d_2$. Let $\bar{R}$ equal the mean of the R_i's (i.e., $\bar{R} = \sum_{i=1}^{k} R_i/k$). Then $\sigma_X = E(\bar{R})/d_2 = E(\bar{R}/d_2)$. Hence, $\bar{R}/d_2$ is an unbiased estimator of σ_X.

Some comparison of this estimator to an estimator based on the usual least squares estimators, S^2, of σ_X^2 seems appropriate here. Let $S_i^2 = (\sum_{j=1}^{n}(X_{ij} - \bar{X}_i))/(n-1)$ denote the usual least squares estimator of σ_X^2 for sample i. Then S_i^2 is an unbiased estimator of σ_X^2, but $S_i = \sqrt{S_i^2}$ is not an unbiased estimator of σ_X. When the X_i's are independent and are distributed with identical normal distributions then the bias of each of the S_i's is well known, and an unbiased estimator of σ_X is given by S_i/c_4, where

$$c_4 = \left(\frac{2}{n-1}\right)^{\frac{1}{2}} \cdot \frac{\Gamma(n/2)}{\Gamma((n-1)/2)}.$$

Let $\bar{S} = \sum_{i=1}^{k} S_i/k$. Then $\bar{S}/c_4$ is an unbiased estimator of σ_X. Even though $\bar{R}/d_2$ and $\bar{S}/c_4$ are both unbiased estimators of σ_X, $\bar{R}/d_2$ is not as efficient as $\bar{S}/c_4$. Montgomery (1985, pg. 174) gives a table comparing the relative efficiency of $\bar{R}/d_2$ to $\bar{S}/c_4$ for various values of n. Montgomery (1985) and others also give tables of the values of d_2 and c_4. None of these textbooks, however, gives a formula for d_2 or a reference to the origin of the table. Table 1 of this paper reproduces the values of d_2, c_4, and the relative efficiencies for $n \leq 10$. Since the relative

---

INSERT TABLE 1 ABOUT HERE

---

efficiency of $\bar{R}/d_2$ (when compared to $\bar{S}/c_4$) is approximately equal to 1 for $n \leq 5$, and since it is rare to have samples of size $n > 5$, it is common practice to use $\bar{R}/d_2$ instead of $\bar{S}/c_4$ because of its simplicity of computation. Hence, an unbiased estimator of $\sigma_{\bar{X}}$, the mean of population standard deviation of the sampling distribution of each of the $\bar{X}_i$'s, is given by $\hat{\sigma}_{\bar{X}} = \bar{R}/(d_2\sqrt{n})$ and it is this estimator that is used when constructing control charts.

Finally, an estimator is needed of σ_R, where σ_R is the common population standard deviation of the distributions of the R_i's. The population standard deviation of W_i, where $W_i = R_i/\sigma_X$, is solely a function of n (Montgomery, 1985) and is usually denoted by the symbol d_3. Hence, $d_3 = \text{Var}(W_i) = \text{Var}(R_i/\sigma_X) = \text{Var}(R_i)/\sigma_X^2 = \sigma_R^2/\sigma_X^2$. Solving for σ_R gives, $\sigma_R = d_3\sigma_X$. Thus, an unbiased estimator of σ_R is given by $(d_3\bar{R})/(d_2)$, since d_3 is a constant and $\bar{R}/(d_2)$ is an unbiased estimator of σ_X. The values of d_3 are given in last column of Table 1.

Hypotheses Tested

The null and alternative hypotheses tested by those using $\bar{X}$- and R-charts are:

H_0: The process is in control (stable)

H_a: The process is NOT in control (NOT stable)

where "in control" means that the $\bar{X}_i$'s are identically independently distributed (i.e., i.i.d.) normal and the R_i's are i.i.d. also. In educational settings these same hypotheses are often of interest with respect to the baseline observations on a single-subject or a same group of subject because, if the baseline is not stable, then there is no known statistical methodology that can always separate treatment effects from what would happen naturally in the absence of treatments (Huitema, 1986). It should be noted that there are, of course, a variety of methods other than the methods to be described in this paper that have been developed for ascertaining whether a baseline is stable (see e.g., Hersen & Barlow, 1976; Huitema, 1986). To test the null hypothesis versus the alternative hypothesis, two separate charts, the $\bar{X}$- and the R-chart are constructed. Once the charts are constructed, a series of rules is then applied. The construction of these two charts will be discussed next and will be followed by a discussion of the series of rules used for hypothesis testing. It should be pointed out to the reader that it is rare to find the idea of hypothesis testing mentioned when this series of rules is applied. In reality, however, it is a hypothesis test that is being performed.

Construction of the $\bar{X}$- and R-charts

Two charts are constructed simultaneously in order to test the hypotheses stated above. The first chart, the $\bar{X}$ chart, is constructed by plotting ordered pairs with the sample number (i.e., the i's) as the first coordinate and the $\bar{X}_i$'s as the second coordinate. Seven parallel horizontal lines are then put on the chart. These lines are:

(1) The Center Line, $\bar{\bar{X}}$, where $\bar{\bar{X}}$ represents the grand mean (i.e., $\bar{\bar{X}} = \sum_{i=1}^{k} \bar{X}_i/k$);

(2) & (3): $\bar{\bar{X}} - \hat{\sigma}_{\bar{X}}$ and $\bar{\bar{X}} + \hat{\sigma}_{\bar{X}}$;

(4) & (5): $\bar{\bar{X}} - 2\hat{\sigma}_{\bar{X}}$ and $\bar{\bar{X}} + 2\hat{\sigma}_{\bar{X}}$;

(6) & (7): $\bar{\bar{X}} - 3\hat{\sigma}_{\bar{X}}$ and $\bar{\bar{X}} + 3\hat{\sigma}_{\bar{X}}$,

where $\hat{\sigma}_{\bar{X}} = \bar{R}/(d_2\sqrt{n})$. The $\bar{\bar{X}} - 3\hat{\sigma}_{\bar{X}}$ and the $\bar{\bar{X}} + 3\hat{\sigma}_{\bar{X}}$ lines are usually called the Lower Control Limit (LCL) and the Upper Control Limit (UCL), respectively. They are often also referred to together as the 3-σ limits. The $\bar{\bar{X}} \pm 2\hat{\sigma}_{\bar{X}}$ lines are often referred to as the 2-σ limits, while the $\bar{\bar{X}} \pm 1\hat{\sigma}_{\bar{X}}$ lines are often referred to as the 1-σ limits. It also seems to be customary to draw in the Center, the LCL and the UCL lines heavily and draw in the 1-σ and 2-σ lines either lightly or not at all.

Table 2 presents a data set that will be used to construct an example of an $\bar{X}$-chart (as well as an R-chart). Each sample in the data set is composed of the measurements taken every half-hour of the voltage produced by four different supply units. If the reader prefers an educational example, these data can be thought of as the number of words read per minute by four different individuals as measured on 25 consecutive Monday mornings. For these data, $k = 25$, $n = 4$, $\bar{\bar{X}} = 349.808$ and $\bar{R} = 2.512$. From Table 1, $d_2 = 2.059$. Hence, $\hat{\sigma}_{\bar{X}} = 2.512/(2.059\sqrt{4}) = .6100$. This yields an $UCL_{\bar{X}}$ of 351.638, a value of 351.028 for $\bar{\bar{X}} + 2\hat{\sigma}_{\bar{X}}$, a value of 350.418 for $\bar{\bar{X}} + \hat{\sigma}_{\bar{X}}$, a Center Line of 349.808, a value of 349.198 for $\bar{\bar{X}} - \hat{\sigma}_{\bar{X}}$, a value of 348.588 for $\bar{\bar{X}} - 2\hat{\sigma}_{\bar{X}}$, and a $LCL_{\bar{X}}$ of 347.978. The $\bar{X}$-chart is given in the upper portion of Figure 1.

INSERT TABLE 2 AND FIGURE 1 ABOUT HERE

The R-chart is constructed in a manner similar to the $\bar{X}$-chart. The points graphed are the ordered pairs (i, R_i). The Center Line is given by $\bar{R}$. The UCL is given by $\bar{R} + 3\hat{\sigma}_R$, where $\hat{\sigma}_R = (d_3\bar{R})/d_2$, and the LCL is given by the maximum of $\{0, \bar{R} - 3\hat{\sigma}_R\}$. This definition of the LCL is advantageous since, by definition, the R_i's are always nonnegative. The other four lines are given by $\bar{R} + 2\hat{\sigma}_R$, $\bar{R} + 1\hat{\sigma}_R$, $Max\{0, \bar{R} - \hat{\sigma}_R\}$ and $Max\{0, \bar{R} - 2\hat{\sigma}_R\}$. For the data set presented in Table 2, $\bar{R} = 2.512$. From Table 1, $d_2 = 2.059$ and $d_3 = .8798$. Hence $\hat{\sigma}_R = (.8798 \cdot 2.512)/2.059 = 1.0734$. Thus, $UCL_R = 5.7322$, $\bar{R} + 2\hat{\sigma}_R = 4.6583$, $\bar{R} + \hat{\sigma}_R = 3.5854$, the Center Line = 2.5120. $\bar{R} - \hat{\sigma}_R = 1.4386$, $\bar{R} - 2\hat{\sigma}_R = .3652$, and $LCL_R = 0$. The bottom portion of Figure 1 gives the R-chart for these data. It should be noted that some authors (e.g., Wheeler & Chambers, 1986) prefer to say that LCL_R "does not exist" when the value of $\bar{R} - 3\hat{\sigma}_R < 0$.

Hypothesis Testing

The null hypothesis that a process is stable is rejected if any of the following 12 conditions are met (Western Electric, 1956, pp. 25-26). The first six conditions are:

(i) One of more points on the $\bar{X}$-chart are beyond the 3-σ limits. That is, for <u>some</u> i, $|\bar{X}_i - \bar{\bar{X}}| > 3\hat{\sigma}_{\bar{X}}$.

(ii) Two out of three successive points on the X-chart are beyond the same 2-σ limit. That is, for <u>some</u> i, at least two of $\{\bar{X}_{i-2}, \bar{X}_{i-1}, \bar{X}_i\}$ are less than $\bar{\bar{X}} - 2\hat{\sigma}_{\bar{X}}$ or at least two of $\{\bar{X}_{i-2}, \bar{X}_{i-1}, \bar{X}_i\}$ are greater than $\bar{\bar{X}} + 2\hat{\sigma}_{\bar{X}}$.

(iii) Four out of five successive points are beyond the same 1-σ limit.

(iv) Eight successive points fall on the same side of the Center Line.

(v) A run up (i.e., nondecreasing $\bar{X}$'s) or a run down (i.e., nonincreasing $\bar{X}$'s) occurs for seven out of eight successive points.

(vi) An unusual or non-random pattern emerges for the $\bar{X}$'s. For example, if the observations were taken daily and Mondays were always high and Fridays were always low.

Conditions (vii)-(xii) are analogous to conditions (i)-(vi) except that $\bar{X}$ is replaced by R throughout.

The probabilities of violating conditions (i) through (iv) when H_0 is true (i.e., the probabilities of a Type I error) can be easily calculated when one assumes that the $\bar{X}_i$'s are i.i.d. normal, although none of the sources consulted for this paper derived these probabilities or even simply gave the probabilities without proof. For condition (i), the probability of a Type I error for any particular value of i is equal to $P(|\overline{X}_i - \overline{\overline{X}}| > 3\hat{\sigma}_{\overline{X}}) = P(|z| > 3) = .00270$, where z is a standard normal random variable. For condition (ii), the probability of a Type I error for any particular i is equal to $P(\text{Two of } \{\overline{X}_{i-2}, \overline{X}_{i-1}, \overline{X}_i\} > \overline{\overline{X}} + 2\hat{\sigma}_{\overline{X}}) + P(\text{Two of } \{\overline{X}_{i-2}, \overline{X}_{i-1}, \overline{X}_i\} < \overline{\overline{X}} - 2\hat{\sigma}_{\overline{X}}) = \binom{3}{2}(P(z > 2))^2 + \binom{3}{2}(P(z < -2))^2 = 2(3P(z > 2))^2 = .0031054$. Similarly, for condition (iii) the probability of a Type I error for any particular i is equal to $2(\binom{5}{4}(P(z > 1))^4) = .0063368$. Finally, for condition (iv) the probability of a Type I error for any particular i is equal to $2(P(z > 0)^8) = .0078125$. The probabilities of a Type I error associated with conditions (v) and (vi) were not discussed in any of the references used for this paper, nor does this author know of a good way of estimating them. Under H_0, however, these errors logically seem to be rarer than those associated with conditions (i) to (iv). Hence, a rough approximation is available to the probability of one or more Type I errors occurring when the family consisting of conditions (i) to (vi) is applied. This approximation is given by $\alpha = 1 - \prod_{j=i}^{iv} (1 - P(\text{Type I error associated with condition (j)}))$. This approximation assumes that conditions (i) to (iv) are independent and ignores conditions (v) and (vi). To the extent that conditions (i) to (iv) are not independent, the approximation overestimates α and to the extent that conditions (v) and/or (vi) hold, this approximation underestimates α. Using this approximation, a value of .0198151 is obtained. Hence, the probability of a Type I error using this set of conditions on the $\bar{X}$-chart is approximately .02 . None of the references used discussed the probability of a Type I error associated with the application of conditions (vii) to (xii) to the R-chart. The derivation of the probabilities would be much more complicated than for conditions (i) to (vi) because the sampling distribution of the R's is positively skewed. Further, no mention was made in any of the references studied by the author of the power associated with this hypothesis testing procedure; but, this is not all that surprising since this procedure is rarely thought of as being a hypothesis testing method in Quality Control textbooks.

Some Examples

Returning to the data from Table 2 (and whose corresponding control charts are portrayed in Figure 1), it can determined from visual inspection of the $\bar{X}_i$'s or of the $\bar{X}$-chart that none of the conditions (i) to (v) is violated. Condition (vi) is subjective, but in this author's opinion is not violated in this data set. Conditions (vii) to (xi) are also not violated here, as can be determined by visual inspection of either the Ranges in Table 2 or the R-chart. Condition (xii) is also subjective and in this author's opinion does not seem to be violated. Hence, since none of the 12 conditions is violated, H_o is not rejected and it can be concluded that there is not yet enough evidence to conclude that the process is not stable. If the data from Table 2 are thought of as reading rates for four children measured over a 25 week period, then it would be concluded that there is not yet enough evidence to conclude that the students' reading rates are unstable.

As a second example, consider the data presented in Table 3.

INSERT TABLE 3 ABOUT HERE

These data are the means and ranges, for samples of size $n = 2$, for the weights of a particular product for $k = 30$ observation times. Similar to the above example, the reader can think of reading rates as the dependent variable instead of weights and the observations as being rates for two randomly selected passages read by a single individual at 30 different times. For these data, $\bar{\bar{X}} = 278.3333$, $\bar{R} = 31.333$, $d_2 = 1.128$ and $d_3 = .8525$. Hence, $\hat{\sigma}_{\bar{X}} = \bar{R}/(d_2\sqrt{n}) = 19.6418$ and $\hat{\sigma}_R = d_3\bar{R}/d_2 = 23.6805$. So, $UCL_{\bar{X}} = 337.2587$, $\bar{\bar{X}} + 2\hat{\sigma}_{\bar{X}} = 317.6169$, $\bar{\bar{X}} + \hat{\sigma}_{\bar{X}} = 297.9751$, the Center Line = 278.3333, $\bar{\bar{X}} - \hat{\sigma}_{\bar{X}} = 258.6915$, $\bar{\bar{X}} - 2\hat{\sigma}_{\bar{X}} = 239.0497$, and $LCL_{\bar{X}} = 219.4079$.

Condition (i) is not violated since none of the 30 $\bar{X}_i$'s are beyond the Upper or Lower Control Limits. Condition (ii) is violated for $i = 5$, since the $\bar{X}_i$'s for both times 4 and 5 are below $\bar{\bar{X}} - 2\hat{\sigma}_{\bar{X}}$. Since, condition (ii) is violated, we can reject H_o and conclude that the process is unstable.

For the purposes of this paper it was decided to continue to investigate violations of the conditions further in order to help the reader understand the 12 conditions better. Also, it is common practice in Statistical Process Control to investigate all 12 conditions since each violation might have arisen from a different cause and the more causes that can be identified and corrected the better the final product will be. It is the author's opinion that this same argument holds for educational data; the main difference, however, is that causes are often harder to identify in educational settings than in industrial settings. But, perhaps with practice using control charts, educators will improve their abilities to identify causes, since the violations of the 12 conditions explicitly point out where and in what direction changes have occurred rather than simply indicating there has been some change.

Returning to condition (ii), it is also violated for i=17, since the means for times 15 and 17 are both above $\bar{\bar{X}} + 2\hat{\sigma}_{\bar{X}}$. Condition (iii) is violated for i=8 (since the means for times 4,5,6, and 8 are below $\bar{\bar{X}} - \hat{\sigma}_{\bar{X}}$) and for i=9 (since the means for times 5,6,8 and 9 are below $\bar{\bar{X}} - \hat{\sigma}_{\bar{X}}$. Conditions (iv) and (v) are not violated nor, by visual inspection, does condition (vi) seem to be violated. Turning to the R-chart, $UCL_R = 102.3748$, $\bar{R} + 2\hat{\sigma}_R = 78.6943$, $\bar{R} + \hat{\sigma}_R = 55.0138$, the Center Line is $\bar{R} = 31.3333$, $\bar{R} - \hat{\sigma}_R = 7.7528$, $Max\{0, \bar{R} - 2\hat{\sigma}_R\} = 0$, and $LCL_R = 0$. Since none of the R_i's are greater than 102.3748 (nor, of course, less than 0), condition (vii) is not violated. Condition (viii) is violated for i=17, since the ranges for times 15 and 17 are both greater than 78.6943. Condition (ix) is not violated. Condition (x) is violated for i=30, since the R_i's for times 23 through 30 are all below $\overline{R}$. Condition (xi) is not violated and, by visual inspection, condition (xii) does not appear to be violated.

Moving Range Charts

In many educational, as well as industrial, settings only a single observation (X_i) can be obtained on an individual or machine at each time point. The $\overline{X}$- and R-charts can not be used directly in these cases. They can be modified, however, into a useful set of charts by introducing the concept of a moving range. Some examples of situations where moving range charts are useful are:

1. An industrial process where each unit takes a long time to produce.

2. Any single-subject research (i.e., n = 1) design where the variable of interest is continuous (e.g., weight or time on task).

In most circumstances the moving ranges (denoted MR_i's) are based on sets of two consecutive observations, X_{i-1} and X_i, and are defined as the range of these two observations (i.e., $MR_i = |X_{i-1} - X_i|$). The individual X_i's then play the role of the $\overline{X}_i$'s and the MR_i's play the role of the R_i's from the $\overline{X}$- and R-chart. The charts used in this situation are referred to as the X-chart and the MR-chart. It is suggested practice (American Society for Testing and Materials, 1976) to take moving ranges based on exactly two observations. There is no difference in the statistical bases used to analyze these charts if three or more observations are used to compute the moving ranges.

The use of X- and MR-charts can be illustrated by using the data from Table 3 and thinking of the "Weekly Averages" as reading rates for one student over 30 weeks based on a single passage (rather than two passages). Since now n = 1, the $\bar{X}$- and R-charts are not appropriate and the X- and MR-charts must be used. Table 4 presents the moving ranges (and for convenience

---

INSERT TABLE 4 ABOUT HERE

---

the X_i's). The mean moving range is denoted by $\overline{MR}$ and is equal to $(\sum_{i=2}^{k}(MR)_i)/(k-1)$. For this example, $\overline{MR} = 36.2069$. By the earlier discussion of the statistical underpinnings of the $\bar{X}$-chart, $\hat{\sigma}_X = \overline{MR}/d_2$, where the value of d_2 is determined from Table 1 using n = 2 (since each of the moving ranges is based on two observations). That is, $d_2 = 1.128$. Further, $\sigma_{MR} = d_3\overline{MR}/d_2$, where d_3 is determined from Table 3 using n = 2. That is, $d_3 = .8525$. For this example, $\sigma_X = 32.0983$ and $\hat{\sigma}_{MR} = 27.3638$. Hence, for the X-chart, $UCL_X = 374.6282$, $\bar{X} + 2\hat{\sigma}_X = 342.5299$, $\bar{X} + \hat{\sigma}_X = 310.4316$, the Center Line = 278.3333, $\bar{X} - \hat{\sigma}_X = 246.2350$, $\bar{X} - 2\hat{\sigma}_X = 214.1367$, and $LCL_X = 182.0384$. For the MR-chart, $\overline{MR} + 3\hat{\sigma}_{MR} = 118.2983$, $\overline{MR} + 2\hat{\sigma}_{MR} = 90.9345$, $\overline{MR} + \hat{\sigma}_{MR} = 63.5707$, the Center

Line = 36.2069, $\overline{MR} - \hat{\sigma}_{MR}$ = 8.8431, and the $-2\hat{\sigma}_{MR}$ limit and the LCL_{MR} are both equal to zero. Figure 3 gives the X- and MR-charts for this example.

INSERT FIGURE 3 ABOUT HERE

Applying the hypothesis testing procedure discussed earlier to these charts, it can be seen that conditions (i) and (ii) are not violated. Condition (iii) is violated when i = 8, since observations 4,5,6, and 8 are all below $\overline{X} - \hat{\sigma}_X$. Hence, H_o is rejected and it is concluded that the reading rates are not stable. Continuing to investigate the violations of the 12 conditions (for the reasons cited previously), it can be seen that condition (iii) is violated also when i = 9, since observations 5,6,8, and 9 are all below $\overline{X} - \hat{\sigma}_X$. Conditions (iv) and (v) are not violated. By visual inspection, condition (vi) does not seem to be violated. Turning to the MR-chart, none of the conditions (vii) to (xi) are violated. Further, by visual inspection, condition (xii) does not seem to be violated. This example clearly shows the loss of power associated with the hypothesis testing procedure when only one observation, instead of even just two, is taken at each time point, since when the data were considered to have two observations per time point there were six violations of the conditions, while treating the data as having only one observation per time lead to only two violations.

This is an important point for educational researchers to note. Too often in single-subject research, researchers or others collect (or record) only one observation per time point, while it would be very little extra work in some cases to collect two or more observations. The reading rates example is a good example of this. The increased number of observations will lead to increased power for detecting a non-stable baseline and will thus save the researcher the time and money wasted by applying a treatment in a situation where the baseline is not stable.

Charts Based on Counts

In some cases the observations collected over time are integer-valued counts that follow approximately either a binomial or Poisson probability density function. The most common industrial situations where counts occur are when one is interested in the number of defective parts per lot (which follows a binomial density) or the number of defects per unit area, such as on a piece of material (which follows a Poisson density). Some educationally oriented variables that follow a binomial density are:

1. Daily class attendance
2. Number of students on task at any point in time
3. Number of students getting each of a series of test items correct/incorrect. Here, the longitudinal dimension is the test items rather than time.

Some educationally oriented variables that follow a Poisson density are:

1. Number of misbehaviors per day by an emotionally impaired student.

2. Number of spelling errors per 100 words in an essay.

The analyses of counts that follow either a binomial or Poisson density use the properties of the appropriate density to estimate the appropriate standard deviations.

The Binomial Case

For the binomial case, only one chart is constructed per data set. For both this case and the Poisson case, an unequal number of observations is allowed per time point, as opposed to the $\bar{X}$- and R- charts and the X- and MR-charts cases where the sample sizes were required to be equal across time. When the sample sizes (denoted as the n_i's, where i represents the time point and i=1 to k, with k = number of time points) are unequal for the different time points the chart used is a p-chart. When the sample sizes are equal for the different time points either a p-chart or an np-chart is used (depending upon the preference of the researcher). The details of constructing p- and -np-charts will now be described.

The data collected at each time point can be thought of as arising from n trials where each trial represents a single trial of a binomial probability experiment performed on an individual unit (e.g., a part or a student). Recall that a binomial probability experiment has the following properties (Johnson, 1988):

1. Each trial has two possible outcomes. These two outcomes are called "success" and "failure"

2. P(Success) = p, P(Failure) = q, and p+q=1 .

3. The random variable X_i represents the number of successes in the n_i trials.

Next, define the variable p_i to be equal to X_i/n_i. Finally, define $\bar{p} = (\sum_{i=1}^{k} X_i)/(\sum_{i=1}^{k} n_i)$ and $\bar{q} = 1-\bar{p}$. Recall that the variance of p_i is given by pq/n_i . An estimator of this variance is given by $\bar{p}\bar{q}/n_i$. Since the p_i's may be based on different n_i's and since the estimated variances of the p_i's are dependent on the n_i's, the control limits will be dependent on the n_i's. The UCL_{p_i} is defined as $\bar{p} + 3\sqrt{\bar{p}\bar{q}/n_i}$, the 2-σ upper limit as $\bar{p} + 2\sqrt{\bar{p}\bar{q}/n_i}$, the 1-σ upper limit by $\bar{p} + \sqrt{\bar{p}\bar{q}/n_i}$, the Center Line by $\bar{p}$, the 1-σ and 2-σ lower limits by $\bar{p} - \sqrt{\bar{p}\bar{q}/n_i}$ and $\bar{p} - 2\sqrt{\bar{p}\bar{q}/n_i}$ and the LCL_{p_i} is defined as $\bar{p} - 3\sqrt{\bar{p}\bar{q}/n_i}$. As with the R-chart, if any of the formulas for the lower 1-σ,2-σ, or control limits yield negative values, then a value of zero is used instead for these limits. Table 5 gives some data that are appropriate for this type of charting. Notice that for these data a success is defined as an "incomplete invoice" while a failure is defined as a "complete invoice". Although may seem strange at first, it is best for ease of interpretation to define a success as that value of the characteristic which is of most interest to the researcher. Figure 4 gives the p-chart and the UCL_{p_i}'s and LCL_{p_i}'s for the data in Table 5. The 1-σ and 2-σ limits were not drawn in because

---

INSERT TABLE 5 AND FIGURE 4 ABOUT HERE

---

they make the chart overwhelming to look at. The hypothesis testing procedure described in the $\bar{X}$- and R-chart section can be applied to p-charts by substituting the symbol "p" wherever the symbol "$\bar{X}$" appears in Conditions (i) to (vi). Conditions (vii) to (xii) are, of course, ignored since there is only one chart here. For the data of Table 5 and portrayed in Figure 4, it can be seen that condition (i) is violated since for i=6, p_6 is above the UCL_{p_6}. Hence, H_o is rejected and it is concluded that the process (i.e., the number of incomplete invoices) is unstable.

When all the n_i's are the same, the p-chart becomes much easier to draw since the horizontal lines (e.g., UCL_{p_i} and LCL_{p_i}) are the same for all values of i. When all of the n_i's are equal their common value is denoted by n. Further, when all the n_i's are equal, a chart called an np-chart is often drawn instead. The np-chart is almost identical to the p-chart. The only difference is that the vertical axis goes from 0 to 1 for the p-chart but from 0 to n for the np-chart. That is, the distinction between the np-chart and the p-chart is exactly the same as the distinction between a raw frequency histogram and a relative frequency histogram.

The Poisson Case

As in the Binomial case, only one chart is constructed when the characteristic of interest follows a Poisson distribution. The chart is called a u-chart, where u is the symbol for the number of events (e.g., nonconformities, misbehaviors, misspelled words) per "unit". The "unit" is defined by the researcher. For example, if one is interested in misbehaviors, then the unit could be a day. If one were interested in the number of misspelled words in an essay, then a unit could be defined as a 50-word block. At each time point data may be collected on some number of units, a_i, where a_i is some nonnegative real number (a_i need not be an integer). Let c_i denote the total number of events occurring at time i and let $u_i = c_i/a_i$. Let $\bar{u} = (\sum_{i=1}^{k} u_i)/(\sum_{i=1}^{k} a_i)$. For a Poisson density the appropriate estimated standard deviation of u_i is given by $\sqrt{\bar{u}/a_i}$. The UCL_{u_i} is given by $\bar{u} + 3\sqrt{\bar{u}/a_i}$, the upper 2-σ limit by $\bar{u} + 2\sqrt{\bar{u}/a_i}$, the Center LIne by $\bar{u}$, the LCL_{u_i} by $\bar{u} - 3\sqrt{\bar{u}/a_i}$, etc.. The statistical testing methodology developed for the $\bar{X}$- and R-charts is then applied with slight modification by replacing "$\bar{X}$" in conditions (i) to (vi) by "u" and by ignoring conditions (vii) to (xii). If all the a_i's are equal to one, then $c_i = u_i$, and the symbol c is used throughout instead of c and the chart is known as a c-chart.

It should be noted by the reader that there is a slight problem with the use of the 1-σ, 2-σ, and 3-σ limits as defined above for the Binomial and Poisson cases. The problem is that the Binomial and Poisson densities are not symmetric. Hence, when symmetric limits are used the probability of a Type I error is a function of some unknown population parameter. This parameter is p in the Binomial case and u (the population number of occurrences per unit) in the Poisson case. Some people prefer to use "probability limits" instead. The use of probability limits is discussed in Grant & Leavenworth (1988).

Assessment of Treatment Effects

For any of the above cases (i.e., $\bar{X}$'s, Moving Ranges, or Counts), the effects of some treatment can be assessed once a stable baseline has been established. To do this, the researcher computes the UCL, 2-σ and 1-σ upper limits, the Center Line, the 1-σ and 2-σ lower limits, and LCL based on the baseline data. The treatment is then applied. Additional measurements are then taken either during or after treatment (depending on the treatment). The hypothesis testing procedures discussed above for the various cases are then applied to the new measurements using the UCL, Center Line, LCL, etc. computed from the baseline data. If any of the conditions (i) to (xii) is violated, it can then be concluded that there is a treatment effect. It is realized by the author that this procedure is a conditional procedure based on the correct decision being made about the baseline. That is, if a Type II error occurs (i.e., the baseline is truly unstable but the researcher wrongly concluded it was stable) when testing the baseline data, then conclusions made about treatment effects may not be correct since the baseline was unstable. It should be noted, however, that this procedure is still better than visual inspection, which is the usual method used by many researchers. Further, the procedures described in this paper have the advantage over other procedures suggested in the literature for the analysis of single-subject data that they are understandable by non-mathematically oriented researchers. The other procedures often recommended in the statistically-oriented literature on single-subject design (e.g., time series methods and randomization tests) are not easily understood by these researchers.

Some Final Notes

1. All of the SPC techniques introduced in this paper can be modified to allow for the analysis of data from one-way multiple group designs and factorial designs.
2. This paper does not contain all of the control charting techniques used in Statistical Process Control. It does, however, cover the ones that are used heavily in Statistical Process Control. Fortunately, these are the same techniques that are useful in educational settings. There are also Statistical Process Control techniques (such as Process Capability Studies) that are not based on control charting, but the author has yet to come up with good educational applications of these techniques.
3. As the author was finishing this paper she became aware of a new textbook in Statistical Process Control by Ryan (1989).

References

American Society for Testing and Materials. (1976). ASTM manual on presentation of data and control chart analysis. Philadelphia: American Society for Testing and Materials.

Grant, E. L. & Leavenworth, R. S. (1988). Statistical quality control (6th ed.). New York: McGraw-Hill

Hersen, M. & Barlow, D.H. (1976). Single-case experimental designs: Strategies for studying behavior change. New York: Pergamon.

Huitema, B. E. (1986). Statistical analysis and single-subject designs: Some misunderstandings. In Poling, A., & Fuqua, R. W. (1986) , Research methods in applied behavior analysis (pp. 209-231). New York: Plenum Publishing.

Johnson, R. R. (1988). Elementary statistics (5th ed.). Boston: PWS-KENT Publishing Company.

Montgomery, D. C. (1985). Introduction to statistical quality control. New York: John Wiley & Sons.

Ryan, T. P. (1989). Statistical methods for quality improvement. New York: John Wiley & Sons.

Shewhart, W. A. (1931). Economic control of quality of manufactured product. New York: D. Van Nostrand Co.

Western Electric Co. (1956). Statistical quality control handbook. New York: Western Electric.

Wheeler, D. J. , & Chambers, D. S. (1986). Understanding statistical process control. Knoxville, TN: Statistical Process Controls, Inc.

Table 1

Values of Relative Efficiency, d_2, c_4, and d_3

n	Relative Eff.	d_2	c_4	d_3
2	1.000	1.128	.7979	.8525
3	.992	1.693	.8862	.8884
4	.975	2.059	.9213	.8798
5	.955	2.326	.9400	.8641
6	.930	2.534	.9515	.8480
7	____	2.704	.9594	.8332
8	____	2.847	.9650	.8198
9	____	2.970	.9693	.8078
10	.850	3.078	.9727	.7971

Note: The values for relative efficiency for n = 7, 8, and 9 were not reported in Montgomery (1985). The values for d are from Grant and Leavenworth (1988).

Table 2

Voltage Data

Subgroup number	DC-voltage output at 20 mA: Sample letter a	b	c	d	Average $\bar{X}$	Range R
1	348.5	350.2	348.3	350.3	349.3	2.0
2	351.3	351.2	347.1	349.7	349.8	4.2
3	348.5	350.5	348.5	349.0	349.1	2.0
4	351.4	350.4	348.6	353.2	350.9	4.6
5	349.4	348.0	349.6	351.1	349.5	3.1
6	351.1	348.1	349.2	350.1	349.7	3.0
7	348.3	349.9	350.7	348.5	349.4	2.4
8	349.9	349.1	349.0	349.6	349.4	0.9
9	349.2	348.7	348.8	350.3	349.3	1.6
10	349.2	351.6	351.9	349.2	350.5	2.7
11	350.1	350.5	351.2	347.9	349.9	3.3
12	350.4	350.8	350.3	352.6	351.0	2.3
13	347.7	349.6	348.6	349.3	348.8	1.9
14	349.0	351.1	350.2	348.0	349.6	3.1
15	350.7	349.3	349.3	350.2	349.9	1.4
16	350.0	351.8	352.3	349.8	351.0	2.5
17	350.1	349.8	349.6	349.2	349.7	0.9
18	351.1	350.6	346.9	349.8	349.6	4.2
19	351.4	349.3	349.7	349.6	350.0	2.1
20	348.8	349.6	351.3	349.2	349.7	2.5
21	349.4	350.2	350.2	351.8	350.4	2.4
22	351.7	351.6	349.9	347.1	350.1	4.6
23	350.4	349.0	349.2	349.6	349.6	1.4
24	349.4	348.7	350.3	348.8	349.3	1.6
25	349.6	349.1	349.6	351.2	349.9	2.1
Σ	.....	.....	.....	.....	8.745.2	62.8

Note: Data taken without permission from Grant and Leavenworth (1988).

Table 3

Weight Data

Subgroup No.	Average	Range
1	255	10
2	330	20
3	280	100
4	235	10
5	230	40
6	240	0
7	280	20
8	235	50
9	240	20
10	315	30
11	325	10
12	280	60
13	260	20
14	275	50
15	330	100
16	250	40
17	320	80
18	260	60
19	275	30
20	295	30
21	225	30
22	300	40
23	330	0
24	275	10
25	290	20
26	295	10
27	265	10
28	280	20
29	285	10
30	295	10

Note: Data taken without permission from Wheeler and Chambers (1986).

Table 4

Moving Ranges for Weight Data

Subgroup No.	X	MR
1	255	—
2	330	75
3	280	50
4	235	45
5	230	5
6	240	10
7	280	40
8	235	45
9	240	5
10	315	75
11	325	10
12	280	45
13	260	20
14	275	15
15	330	55
16	250	80
17	320	70
18	260	60
19	275	15
20	295	20
21	225	70
22	300	75
23	330	30
24	275	55
25	290	15
26	295	5
27	265	30
28	280	15
29	285	5
30	295	10

Table 5

Invoice Data

DAILY PROPORTIONS FOR INCOMPLETE INVOICES

Date	*Number of Incomplete Invoices*	*Total Number of Invoices*	p_i	*Date*	*Number of Incomplete Invoices*	*Total Number of Invoices*	p_i
9/27	20	98	.204	10/11	7	50	.140
9/28	18	104	.173	10/12	7	53	.132
9/29	14	97	.144	10/13	9	56	.161
9/30	16	99	.162	10/14	5	49	.102
10/1	13	97	.134	10/15	8	56	.143
10/4	29	102	.284	10/18	9	53	.170
10/5	21	104	.202	10/19	9	52	.173
10/6	14	101	.139	10/20	10	51	.196
10/7	6	55	.109	10/21	9	52	.173
10/8	6	48	.125	10/22	10	47	.213

Note: Data taken without permission from Wheeler and Chambers (1986).

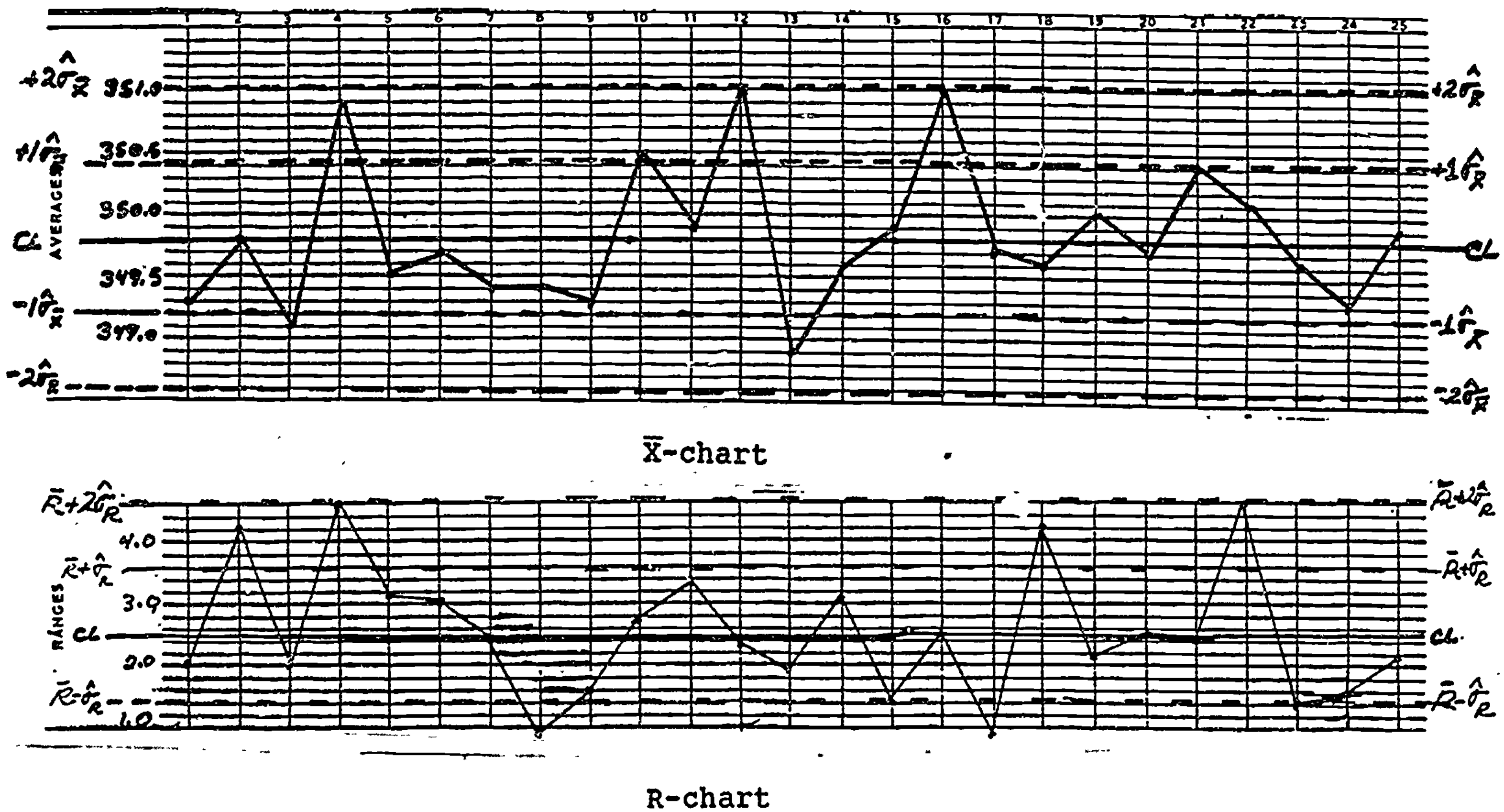


Figure 1. X̄- and R-charts for the Voltage data

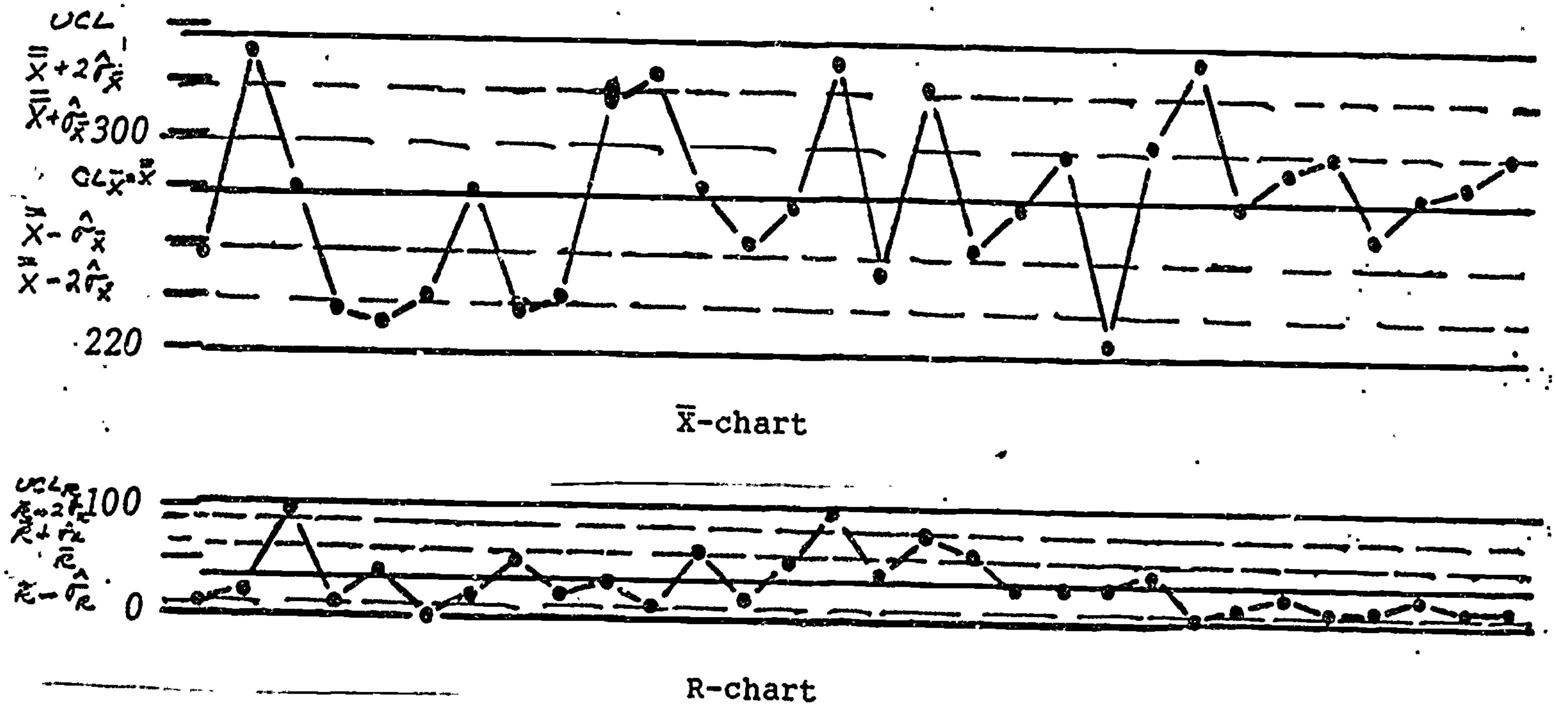
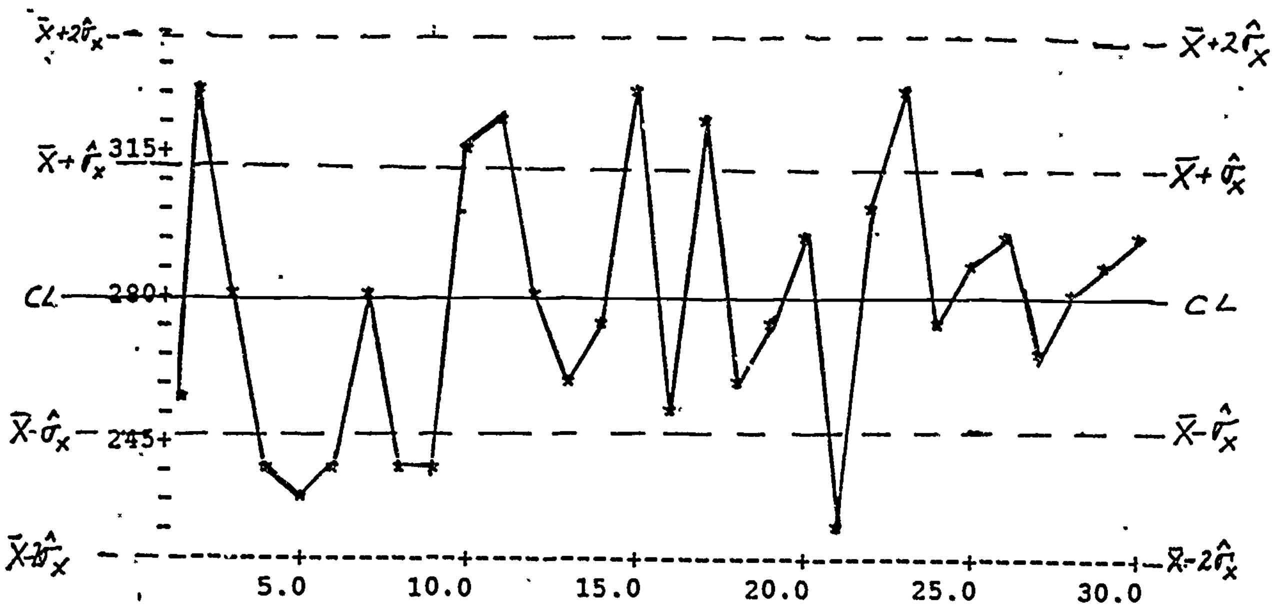
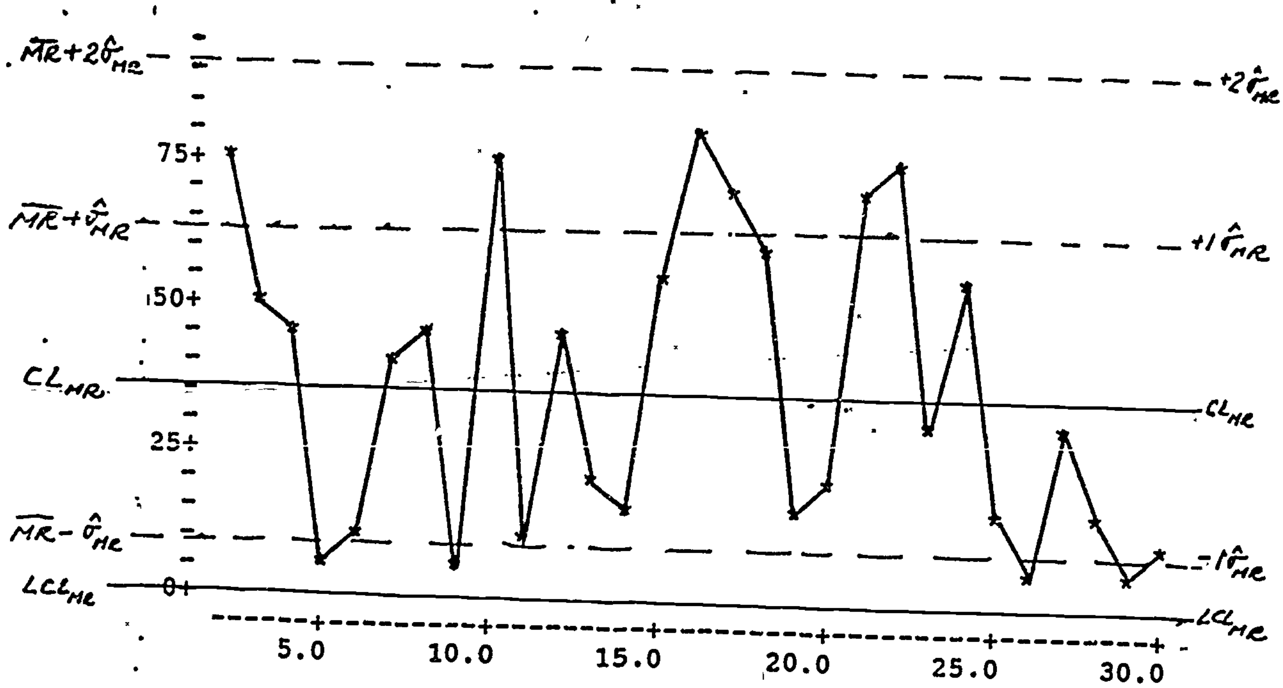


Figure 2. X̄- and R-charts for the Weight data when each time point is considered to have two observations



X-chart



MR-chart

Figure 3. X- and MR-charts for the Weight data when each time point is considered to have one observation

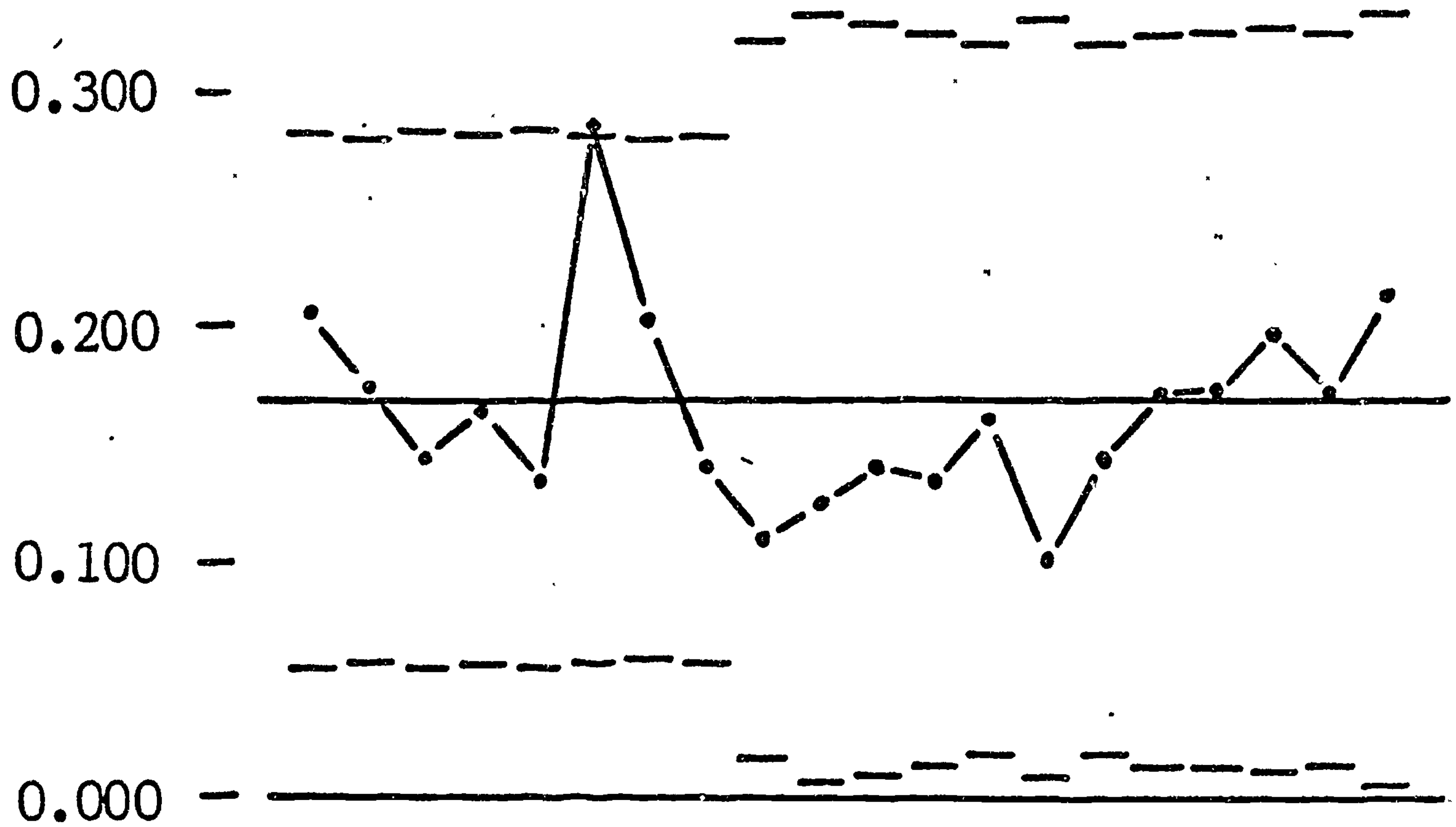


Figure 4. p-chart for the Invoice data