M. Prewarski

The IBM SELECTRIC Composer

Statistical Evaluation of Printing Alignment

Abstract: This paper describes the measurement and evaluation of print alignment for the SELECTRIC Composer, and discusses the suitability of statistical techniques for achieving both. This paper deals extensively with techniques and procedures for collecting data; however, it is not intended to define a generalized method for statistical evaluation. Printing alignment is shown to be described objectively and precisely by the distribution of measured misalignment; misalignment is precisely defined.

Introduction

In any useful printing system, the alignment of printed characters, in relation to each other and to the writing line, must meet certain standards of accuracy. For the IBM SELECTRIC Composer, the total standard of print quality was set exceptionally high and therefore rigid control of alignment accuracy was especially necessary. Absolute accuracy was desirable but recognized as not obtainable because of the mechanical limitations inherent in the design of any system. Furthermore, absolute accuracy was not essential to satisfy the needs of the user. Therefore, the first problem was to determine the lowest acceptable quality level. This was set after consideration of the user's need, the cost and time requirements in development necessary to attain various quality levels, and the field effort that would be necessary to maintain any given quality level over the lifetime of the machine.

The second major problem was to develop objective procedures for evaluating the quality of printwork on any sample and to reliably compare any sample with the standard. The actual measurement of the relative position of one character with respect to another posed no problem since many good optical systems were available. However, it was not practical to measure the total population of characters and character combinations that could be composed into a sample. Therefore, it was necessary to develop and validate a procedure for sampling and measuring only a small part of the total number of combinations and inferring by statistical means the quality of the sample.

A third problem was that of relating the changes in printwork quality to the several generations of Composer development models and to changes in various mechanisms as development progressed. Again, statistical methods were used so that samples could be objectively evaluated.

This paper describes the methods which were used in evaluating the printwork of the SELECTRIC Composer and the rationale behind many of the decisions that were required before a workable procedure was achieved. The use of these methods has enabled the designers of this system to predict and control alignment quality.

Alignment quality

To define alignment quality, three premises were required. First, it was assumed that the specified character location (that is, the combination of horizontal spacing and vertical position on the writing line as specified by the type designer) was optimal and correct. Second, it was assumed that *any* measurable departure from that specified location was an alignment deviation. (This does not necessarily mean that such a departure was an alignment *error*, nor does it mean that alignment quality was necessarily reduced if such a departure existed. This will be discussed below.) Third, it was assumed that deviations of one kind rank equally with those of another, e.g., that characters spaced too closely together were as significant a deviation as characters spaced too far apart.

With these assumptions there was a basis for defining alignment quality; in an idealized system alignment is inversely proportional to the number of deviations and to the magnitude of the displacement from the specified locations. In a practical system, however, additional factors must be

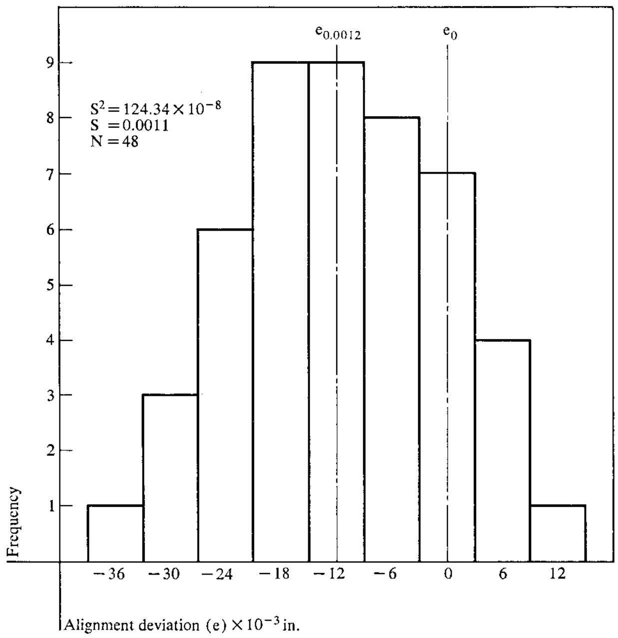


Figure 1 Frequency distribution of measured horizontal deviations in a machine output sample.

considered. Dissimilarities among characters, special (or "designed-in") aberrations, the constant bias (or tare factor) inherent in a measurement system, and the mechanical limitations of the printing system each cause discernible deviation which cannot be called error. So-called "ideal" alignment is thus defined as the smallest possible deviation that takes these factors into account.

• *Character-to-character dissimilarity*

Because of the variety of shapes involved in the character set, and the number of combinations occurring in context, horizontal alignment is a function not only of a single character's location but also of the spacing between characters. There is no absolute correct spacing between any two characters. The type designer attempts to space his characters so as to achieve good balance among all the possible combinations. The number of permutations is very high; therefore, to attain the desired balance some compromise is necessary.

There is also a correction for vertical alignment: The type designer intentionally moves some characters off the writing line so that they may seem, optically, to be on the writing line when viewed next to other characters. (This will be more fully covered in the discussion of measurement.)

• *Variations caused by a measuring system*

Depending on how measurements are taken, the characters may appear to be grouped closer together or spread farther apart than the assumed correct spacing. Alignment quality is not affected, however, because this deviation is the same, on the average, for all combinations measured under the same set of conditions. The *re ative* spacing is unchanged.

As an example of this effect, consider the distribution in Fig. 1. The data are taken from the output of a photocomposing system which employs optics to project character images onto light-sensitive paper. This method is susceptible to deviations due to variations in image distance, focus, etc., and output characters are usually fitted either too close together or too far apart. This spacing of characters is an example of mechanical limitation producing measurable deviation that is not an alignment error. It is not error because for any given sample the relative spacing between characters is unchanged. The measurements indicate that the characters are, on the average, 0.0012 in. closer together than specified. In statistical terms, this represents a negative bias in the data of -0.0012 in., and for this sample alignment error should be judged by the distribution of deviations about $\bar{x} = -0.0012$ and not about $\bar{x} = 0$.

• *Mechanical limits affecting alignment*

In addition to the effects of character dissimilarity and measurement techniques, there is a class of deviations that are actually errors. These are deviations that are caused by the printing system itself. They can be caused by a badly located character on the type element or typebar, for example, or by looseness or inaccuracy in any of the mechanical operations involved in printing the character. It is these that the designer attempts to eliminate, within the limits of mechanical accuracy and the extent to which that accuracy can be maintained.

Alignment quality, then, has been defined as the *distribution of errors about a mean which is regarded as the perfectly aligned mode.* Inherent deviations and bias that are permissible in the judgment of the designer do not affect this quality.

Using this definition as the governing rule and statistical techniques as the means, the alignment quality of the SELECTRIC Composer was determined and tested. The sections which follow describe the particulars and review the results obtained.

Measuring methods

The selection of character combinations for measurement was completely random and unbiased. Prose text was used since tabular material, poetry, or any other "artificially" disciplined copy could have introduced an unexpected machine bias and caused a very large experimental error. Character combinations were selected with the aid of a table of random numbers. (A computer or other means which makes random selection possible without experimenter or system bias could have been used.) The selected combinations were in numerical proportion to their common usage because ordinary copy and random selection were used.

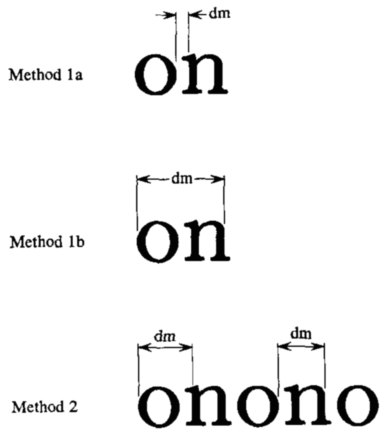


Figure 2 Methods for measuring horizontal deviation.

The selection scheme was the same for both vertical and horizontal alignment; in fact, the same sample was used for both purposes. (Ideally, the sample should be large enough to serve for any general study of alignment, but a much smaller sample can be composed and completely measured if one specific problem is to be examined.)

- *Horizontal alignment*

Figure 2 shows the two principal methods for measuring the distance between characters. In Method 1a, *dm* is the closest distance between body boundaries. Serifs and projections are ignored. In Method 1b, *dm* is the farthest distance between body boundaries. In Method 2, *dm* is the distance between the extreme left or extreme right sides of the body boundaries. Again, serifs and projections are ignored. In Method 1 (a and b), *dm* will be excessively large or small due to such factors as ink spread, character shrinkage during molding of the type element, edge loss caused by plating of the element, the impact velocity of the printhead, optical or gear-train magnification, etc. In Method 1, it is necessary to compute the mean of the frequency distribution for the sample and correct all indicated deviations by $\bar{x}$ which accounts for any shift caused by the biasing factors (note the example of Fig. 1). The indicated deviation for any pair of characters is given by

$$x_i = ds - dm\,, \tag{1}$$

where *ds* is the specified distance between characters and *dm* is the measured distance, and

$$\bar{x} = \left(\sum_{i=1}^{n} x_i\right)/n$$

where *n* is the number of measurements. The actual deviation for any single measurement is then

$$x_i - \bar{x}\,. \tag{2}$$

Figure 3 Method 1 for measuring vertical deviation.

Method 2 will exclude all of the biasing factors except optical magnification (which will be negligible) and the correction using $\bar{x}$ is therefore unnecessary.

Both of the principal methods were satisfactory, but Method 2 required less computation. The latter is an advantage when the number of samples or the number of measurements per sample is large, but the choice is otherwise arbitrary. (Once chosen, of course, one method or the other should be used consistently.)

- *Vertical alignment*

There are also two methods for measuring vertical alignment. The first of these measures the vertical deviation of one character relative to an adjacent character, while the second determines the vertical displacement of any character with respect to the writing line.

Figure 3 shows an example of Method 1. The location of the first character is used as a reference and, since only deviations between characters are measured, it is not necessary to assume that the reference character is positioned correctly with respect to the writing line. The vertical deviation of the adjacent (right-hand) character as compared to the reference character is measured using a sign convention. If the right-hand character is higher than the reference character, the distance *dm* is treated as a positive deviation.

Certain characters are specified by the type designer to be positioned off the writing line. If either of the characters being measured is one of these, a correction *c* is added to *dm* to obtain the true deviation. Figure 4 shows how *c* is obtained. The line x–x is the center of curvature of all characters of all type fonts. The distance from x–x to the writing line is the same for all characters. For the SELECTRIC Composer this distance is arbitrarily set at 0.035 in. A lower case serif-bottomed character, such as the "n", is positioned exactly on the writing line. The distance from x–x to the bottom of "n" is then $c_1 = 0.035$ in. The lower case "o" is positioned off the writing line so that in the type designer's best judgment the character appears to be optically on the writing line when viewed by the observer and referenced to a square or serif-bottomed character. The distance from x–x to the bottom of "o" can be $c_2 = 0.0365$ in. For the example in Fig. 4 the correction is:

$$c = c_2 - c_1 = 0.0365 - 0.0350 = 0.0015.$$

c can be negative, of course, when $c_2 \leq c_1$. This correction factor remains constant for all occurrences within a given

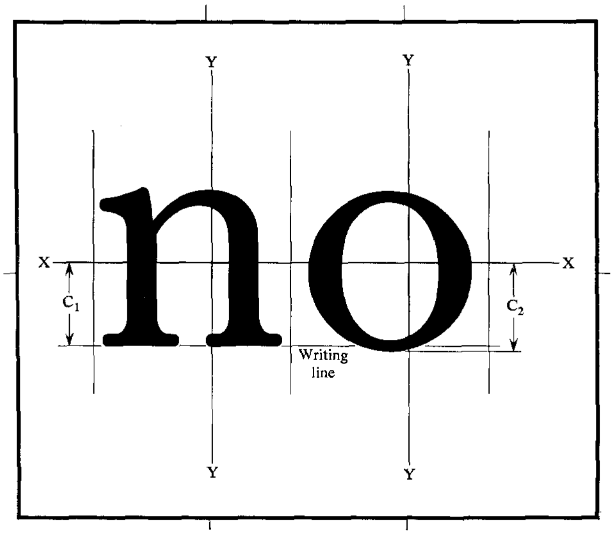
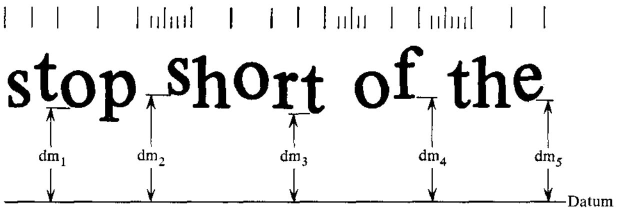


Figure 4 Method for determining the vertical correction factor, c ($c = c_2 - c_1$).

Figure 5 Method 2 for measuring vertical deviation.



sample, and it need be determined only once. If the sample is changed, however (e.g., taken from a different system or from a different size or style of type), c must be recalculated.

An example of Method 2 is shown in Fig. 5. With this method, a datum line is established anywhere convenient to the writing line and measurements are taken from the datum to the bottom edge of each character. It is important that the datum line be parallel to the writing line. The correction factor c_k for this method is just the distance, if any, between the bottom of the k'th character and the writing line specified by the type designer. For a datum line set below the writing line and according to the sign convention, c_k is positive for characters specified to be above the writing line.

The mean, or apparent, distance from the datum to the writing line is

$$\bar{y} = \left[\sum_{k=1}^{n} (dm_k - c_k) \right] / n \, .$$

The true vertical deviation from the writing line for any character k is then

$$y_k - \bar{y} \, .$$

Since the two methods differ in their employment of c and in measuring the deviation, the results of one cannot be compared with those of the other. Method 1 is normally more useful because it tends to exclude any aberrations due to paper slippage, etc. Method 1 has been especially useful in instances where one character was being evaluated in relation to other characters.

- *Measuring instruments*

Any measuring instrument that gives a sharp edge definition and is accurate within (arbitrarily) ±0.0002 in. may be used. A printer's microscope, a toolmaker's microscope, and a microscope mounted in a servo-controlled carriage with a special visual inspection device which employs a digital readout were used in this study.

Setting a quality standard

Because there has been no commonly accepted, quantitative definition of alignment quality, setting a specification for alignment has been a controversial subject. By defining alignment quality as the statistical distribution of deviations about an identifiable mean, a quantitative means for comparing samples becomes available. As pointed out earlier, the specification itself remains arbitrary, since there is still no absolute criterion by which one may judge how large a deviation is acceptable. However, though the standard is subjective in that it exhibits acceptable deviation according to the judgment of the examiner, the evaluation of samples is objective in that the distribution of the deviations is evaluated using statistical methods.

In the SELECTRIC Composer design program, the first approach toward developing a standard was to create a sample page that could be achieved by the Composer without particular care or tuning. Offset reproductions of this sample were made and inspected by a number of prospective users. When the sample was found satisfactory by most of the examiners, it was decided that this would represent the minimum alignment standard for the Composer. This standard was then measured and numerical values, to which all other printed work could be compared, were obtained (see Figs. 6 and 7).

A second approach was to obtain samples of typical work from existing commercial cold-type systems. The alignment quality of these was evaluated and compared with the alignment capability of the Composer prototype. From this information a second minimum standard was set, but in this case only for vertical alignment. (Differences in escapement, justification, design of the respective print elements, etc., made meaningful comparison for horizontal alignment impossible.)

To simplify the taking of data and because at the time of the study the design information on characters which descend below the writing line was not available, the sampling and measuring technique was modified. Only those char-

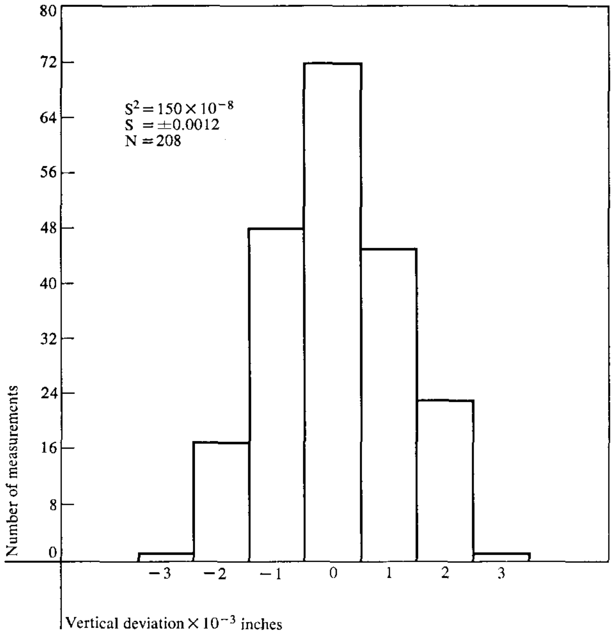


Figure 6 Vertical deviation distribution for a sample from the development standard.

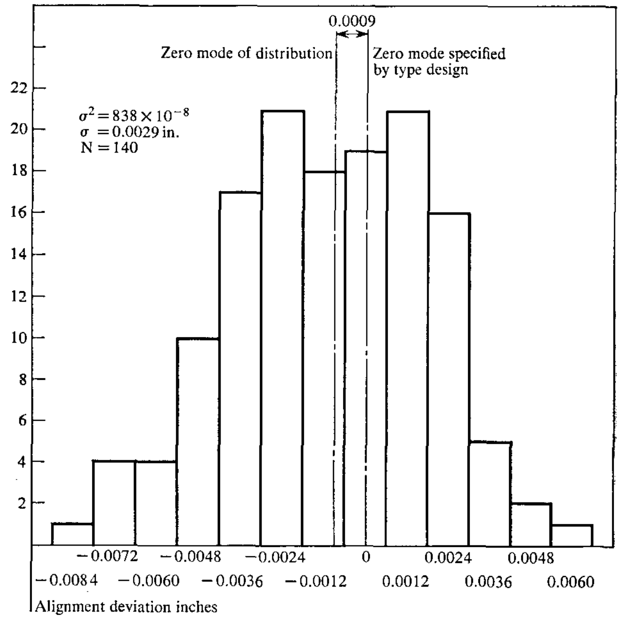


Figure 7 Horizontal deviation distribution for a sample from the development standard.

acters with a flat or square bottom edge were considered. The alignment of such characters as f, h, i, k, l, m, r, and z was measured with respect to the arbitrary reference character n, and the correction factor c was eliminated.

Although the second approach was used later in the development program, it agreed surprisingly well with the earlier standard using the first approach. Since both standards were in substantial agreement, it was concluded that, despite its origin in subjectivity, the development standard used for the SELECTRIC Composer was very nearly optimum.

Statistical prediction and comparison

The *quality* of a character set (or, in general, a system) can be determined by statistical prediction; the technique for our purposes was to compare a sample with the standard (assuming, as can be done quite safely in this case, that the character frequency and alignment deviation distributions are of the same type for both the sample and the standard) and employ the results of the comparison to predict whether the set represented by the sample was aligned as well as the standard. If the confidence of the prediction (determined from published tables) exceeded an arbitrary minimum, quality was regarded as acceptable, i.e., better than or equal to the standard.

Procedures for evaluation and comparison

In the complete evaluation process, two procedures were necessary. The sample being tested was first compared with the standard, to provide a prediction of system quality based on the sample itself. In order to establish that the sample being tested was representative, however, it was necessary to compare several samples with each other. If the agreement was poor, the confidence that was placed in the first comparison was correspondingly reduced, and the size of the original sample was increased. Therefore, for each new test (i.e., whenever a new character set or redesigned system was being evaluated for the first time), the second comparison determined if the test sample size was adequate.

- *Comparing the variance of a sample with a standard*

(*1*) A sample of 2- and 3-character combinations was randomly selected and measured. Depending upon the needs of the moment, the sample ranged from a limited group containing only one character (in combination, of course, with the other characters with which it is normally found) to a "composed" sample of the entire set, which was totally measured.

(*2*) Indicated deviations were calculated:

$x = ds - dm$, for horizontal deviations and
$y = dm$, for vertical deviations.

(*3*) Corrections for mean deviation were made and actual deviations were calculated:

(*a*) For horizontal measurements, a correction was made for shift in distribution mode, to obtain $x = ds - dm - \bar{x}$.

(*b*) Vertical measurements were corrected for *specified* deviations:

$y = dm - c$ (Method 1), or $y = y_k - \bar{y}$ (Method 2).

(4) The variance, s^2, of the sample was calculated:

$$s_x^2 = \left(\sum_{k=1}^{n} |x_k|^2\right)/(n-1), \quad \text{and}$$

$$s_y^2 = \left(\sum_{k=1}^{n} |y_k|^2\right)/(n-1), \tag{9}$$

where n is the number of measurements and $(n - 1) = \phi$, the number of degrees of freedom in the sample.

(5) The sample variance s^2 was compared with the variance of the standard σ^2, for ϕ degrees of freedom:

$$\chi^2 = \frac{\phi S^2}{\sigma^2} \tag{10}$$

or, combining steps 4 and 5,

$$\chi_x^2 = \sum_{k=1}^{n} |x_k|^2/\sigma_x^2, \quad \text{and}\ \chi_y^2 = \sum_{k=1}^{n} |y_k|^2/\sigma_y^2. \tag{11}$$

(6) χ^2 was compared with statistical tables, which give the value of χ^2 for various levels of confidence.

Example—The alignment quality of a recently typed Composer sample is to be compared to the quality of the standard. The determined variance of the standard is $\sigma^2 = 838.7 \times 10^{-8}$. If thirty-five measurements have been taken, say for vertical deviation, then $n = 35$ and $\phi = 34$. Let the computed sample variance be $s^2 = 289 \times 10^{-8}$. The computed χ^2 value is then

$$\chi^2 = (\phi s^2)/\sigma^2 = \frac{34 \times 289 \times 10^{-8}}{838.7 \times 10^{-8}} = 11.7.$$

Statistical tables[1] show that for $\phi = 34$ and 99.5% confidence, the value of χ^2 is 16.4. Since the calculated χ^2 is less than the tabular value, it can be concluded with more than 99.5% confidence that the (vertical) alignment quality of the population represented by the sample is better than that of the standard.

- *Comparing the test sample with another sample*

The example above assumed that the SELECTRIC Composer sample being considered was large enough and random enough to accurately represent the test population as a whole. To validate the comparison, however, this assumption was verified by examining other independent samples. This was done by following steps 1–4 above for a second (and perhaps a third, fourth, etc.) sample to obtain other estimates of s^2. The significance of the difference between the test sample (say sample a with samples b, c, etc.) was then computed as follows:

(1) The ratios $F_1 = s_a^2/s_b^2$; $F_2 = s_a^2/s_c^2$, etc., were computed (F is the ratio of two variances and is, arbitrarily, always greater than unity; hence if the test sample variance s_a^2 is the smaller of the two variances, the ratio function is inverted.)

(2) Based on the degrees-of-freedom ϕ_a and ϕ_b (or ϕ_c, etc.), and on the selected significance level, an F-ratio was obtained from statistical tables. (Significance is, in this case, the probability of significant difference between the two sample variances.)

Example—To illustrate, assume that the variances for the test sample and for another sample have been computed. Let $s_a^2 = 980 \times 10^{-8}$ for the variance of the test sample, and $s_b^2 = 750 \times 10^{-8}$ for the variance of the second sample. Then

$$F = s_a^2/s_b^2 = \frac{980}{750} = 1.30.$$

Let the number of measurements for sample a be 61, for sample b, 61; degrees-of-freedom values are then $\phi_a = 60$, $\phi_b = 60$. The tables[1] for $\phi = 60$ show:

Probability of difference being significant	*F-Ratio*
0.1	1.40
0.05	1.53
0.01	1.84

If the significance level is chosen at 0.1 or 10%, we conclude that there was no significant difference between samples a and b, since the F-ratio of 1.30 was less than the tabular F-ratio, 1.40. If the process is now repeated with the same results for samples c, d, etc., we conclude with increasing confidence that the test sample was large enough to be valid.

- *Predetermining sample size*

It is sometimes desirable to determine in advance the sample size that would be required for significance. To do this in the Composer alignment study, the validating comparison described last above was eliminated and the χ^2-test using the standard was the only procedure used. However, the sample size can be predetermined only if certain values can be prespecified. The designer must decide in advance what difference between σ^2 and s^2 will be considered significant, and what risk of failure will be tolerated (i.e., risk either of failure to detect difference or of seeming to detect difference when there is none).

In practice, significance is usually specified by establishing a limiting value for $\sigma - s$, or equivalently, by stipulating that s will be significant only if it equals (or is less than) some arbitrary fraction of σ.

Risk factors α and β are established, where

α = risk of detecting significant difference when none exists, and

β = risk of failure to detect a specified difference. The ratios

$R = s^2/\sigma^2$ where s is a specified fraction of σ, and

$R' = 1/R$ are calculated.

Using the value obtained for R' and the risk factors α and

β, a value for ϕ (degrees of freedom) and hence for required sample size ($n = \phi + 1$) can be obtained from the tables[1].

Example—Assume that a design improvement is expected to produce a 20% improvement in alignment quality, that is, a σ 20% lower than the standard, which is considered a sufficiently important alignment improvement to justify the cost. The χ^2-test will be used to determine whether the output of the new design has a σ 20% less than the standard. The number n of observations required to obtain s^2, and hence χ^2, is to be determined. The difference important to detect is 20%, or 0.2, and it is required to have $\sigma - s = 0.2\sigma$, or $s = 0.8\sigma$. The standard deviation σ is known, and the R can be determined:

$$R = \frac{s^2}{\sigma^2} = \frac{(0.8\sigma)^2}{\sigma^2} = 0.64\,,$$

$$R' = \frac{1}{R} = 1.561\,.$$

Assume that an α risk of 5% and a 10% β risk are tolerable. From the tables [1], for $\alpha = 0.05$ and $\beta = 0.1$,

R'	ϕ
1.702	60
1.457	120

and interpolating for the value 1.561, the desired numbers are $\phi = 88$ and $n = 89$. Therefore 88 combinations are required for the χ^2-test given by

$$\chi^2 = \frac{88s^2}{0.8\sigma}\,.$$

If χ^2 for the measured sample is less than the tabular value for χ^2 then the new system design is significantly better than the standard system and economically worthwhile.

Setting the confidence level

The methods for measuring and evaluating print alignment have been described, but a comment is appropriate on the question: "At what level of confidence should design decisions be made?" Setting the confidence level high sets the number of measurements also high. But if a proposed improvement, for example, involves a costly change in tooling and an appreciable cost increase for component parts, then the improvement must be both significant and verifiable if the change is to be justified. Fortunately, the cost of taking samples is relatively low, and for most practical decisions an alignment of 90% confidence or 10% significance has been adopted as a useful and practical level.

Conclusions

Statistical alignment evaluation has proved useful not only in development but also in product testing, manufacturing, and assembly. To illustrate its application throughout the Composer program, examples from three areas are given below.

Development—The concept was used to set the tolerances in the escapement system which indexes the carrier horizontally and in the tilt ring which positions the type element vertically. The same concept was used in locating the characters on the type element. High-usage characters were positioned where their alignment would be least affected by tolerances of the parts and by machine dynamics.

Manufacturing—Molding set-ups in the type element molding machines were evaluated before each production run by testing the alignment quality of a sample element.

Assembly—Alignment evaluation was used to determine the minimum time for machine tuning and as a final check on finished machines before shipping.

The concepts described here were thus generally applicable in assessing the design and operation of the entire SELECTRIC Composer system. Any function which affects alignment (and more than half of those in the system do, however indirectly), can be at the very least qualitatively judged by alignment evaluation.

Acknowledgments

The author wishes to express his appreciation to Mr. W. J. Diamond, who introduced him to the statistical methods for which these applications have been developed and whose critical evaluation of the paper was most helpful. Thanks are due also to the many colleagues in the SELECTRIC Composer program whose discussion and advice have been invaluable.

References

1. O. L. Davis, *Design and Analysis of Industrial Experiments*, Hafner Publishing Co., New York, 1956.

Received December 23, 1966.