An Attempt to Model Distributions of Machined Component Dimensions in Production

Can ÇOGUN*, Bünyamin KILINÇ
Mechanical Engineering Department, Gazi University, Maltepe, 06570 Ankara, Turkey

In this study, normal, log-normal, triangular, uniform, Weibull, Erlang and unit beta probability density functions are tried to represent the behaviour of frequency distributions of workpiece dimensions collected from various manufacturing firms. Among the distribution functions, the unit beta distribution function is found to be the best fit using the chi-square test of fit. An attempt is made for the adoption of the unit beta model to x-bar charts of quality control in manufacturing. In this direction, upper and lower control limits (UCL and LCL) of x-bar control charts of dimension measurements are estimated for the beta model, and the observed differences between the beta and normal model control limits are discussed for the measurement sets.

Key Words : Part Dimension, Statistical Modelling, Beta Distribution Function, x-Bar Control Charts

1. Introduction

Workpart dimensions are random variables and statistical frequency distributions of their dimensions vary from process to process. Quality control is a professional field that deals with these variations in an effort to provide quality production at minimum cost. The point of quality control is to study ongoing processes, which involves analysis of the characteristics of the population output by inference of the sample output. The detected trends result from assignable causes as opposed to random causes which are inherent in the manufacturing processes. The main tools used for identifying assignable causes of variation are control charts, of which the x-bar chart and the p-chart are prominent.

The control charts or quality control technique always assumes normal dispersion, or distribution of dimensions. Some of the studies conducted in the field have shown that the normal distribution, which assumes the symmetry of distributions, may not properly model the dimension and tolerance distributions due to the existence of skewness in their shape. For the dimension and tolerance distributions limited number of researchers proposed different models than normal, like right-skewed normal (Shainin, 1949), semi circle (Gibson, 1951), uniform (Crafts, 1952 ; Fortini, 1956), triangular (Doyle, 1951 ; Mansoor, 1960 ; 1964), moving normal (Gladman, 1959), beta (He, 1991) and sinus (Mansoor, 1960 ; 1964 ; Burr, 1958). In all of these works, very few distribution functions, mostly only one, are tested with a rather limited number of dimension or tolerance frequency distributions, i.e. samples. Some others have also raised the need for a different model than the normal to reflect the behaviour of dimension and tolerance distributions (Bjorke, 1978 ; Bennet, 1964 ; Gladman, 1980 ; Zhang and Huq, 1992 ; Fortini, 1967 ; Nelson, 1984). Although, some attempts have been made to represent the dimension distributions by another model than the normal, no attempt has been made to reconstruct x-bar control charts, according to the new proposed

* Corresponding Author,
E-mail : cogun@mikasa.mmf.gazi.edu.tr
TEL : +90-312-2317400 Ext 2452; **FAX** : +90-312-2320425
Mechanical Engineering Department, Gazi University, Ankara 06570, Turkey. (Manuscript **Received** April 2, 2001; **Revised** October 12, 2001)

models. Some published works (Bai and Chois, 1995 ; Balakrishnan and Kocherlakota, 1986 ; Borror et al., 1999 ; Chan et al., 1988 ; Choobineh and Ballard, 1987 ; Haridy and El-shabrawy, 1996 ; Lashkari and Rahim, 1982 ; Nelson, 1979 ; Padget and Spurrier, 1990 ; Shilling and Nelson, 1976 ; Shore, 1998) have provided some motivation for normal-based control charts that deal with data that is not symmetric. The authors of these works believe that it is not desirable in practice to have control limit factors for specific distributions since standard control charts based on the normal model are sufficiently robust to non-normality or can be made so with some modifications.

Formerly, an investigation was conducted by the authors of this work (Kılınç, 1999) to propose some other statistical probability density functions, or models than the normal, which would reflect the statistical behaviour of the frequency distribution of dimensions better than the normal model. The beta model, or distribution function is found to be the best model among the proposed seven distribution functions to reflect the shape behaviour of dimension frequency. The changes that should be made in the construction of x-bar charts in the use of the beta model are proposed.

2. Research model

2.1 Statistical modelling of dimension distributions

2.1.1 Data sets used in the study

The data sets (set of dimensions) are collected from parts, which are produced by well-known machine component producers in Turkey. Special attention is paid to choose functionally different parts with different sizes, shapes, tolerances and manufacturing processes to eliminate concerns that could be raised from results due to functionally similar workparts produced by similar manufacturing techniques and dimensions.

Although, a large number of sets of dimensional measurements, i.e. data sets are collected from various workparts, more than 100 sets of data, a limited number of them is presented in this paper. The information on the data sets used in the study is given in Table 1. The first 5 characters of the code of the data set is for the short description of manufacturing process, part name and dimension information. The letters after the dot (.) in the code, namely, HMA, ORS, TS, ASE, MKE, HE and MAN, are the abbreviations for HEMA Gear Company (gear manufacturer), ORS Bearing Company (bearing manufacturer), Konya Trigger and Valve Company (valve manufacturer), ASELSAN (Military Electronics Industries), MKEK Machinery and Chemicals Industry and data sets taken from published works of He (1991) and Mansoor (1964), respectively. The data sets used in other published works for reviewing the probability density functions are not included in this study due to lack of information about the dimensions and parts.

Micrometers and dial gages with different accuracies (Table 1) were used in the measurements. The measurements were performed by the quality control personnel of the companies.

2.1.2 Distribution functions used in this study

In this study, normal, log-normal, triangular, uniform, Weibull, Erlang and beta probability density functions are tried for the fit of behaviour of frequency distributions of part dimensions collected from various manufacturing companies. Weibull, Erlang, beta and log-normal distributions could be symmetric or non-symmetric (right- or left-skewed) in shape depending on the values of model parameters. The distribution functions, estimation of model parameters and shape variations of the models could be found in statistics books (Bain, 1978 ; Bury, 1975). The shapes of distribution functions of these models are given in Fig. 1. The use of the beta distribution function requires long computations involving its four model parameters. A useful and practical form of the beta distribution which requires less and simple computations is the unit beta distribution, and is used in this study. Brief information about the unit beta model is given in Appendix.

Table 1 Summary of the information on the measurement sets (Data Sets)

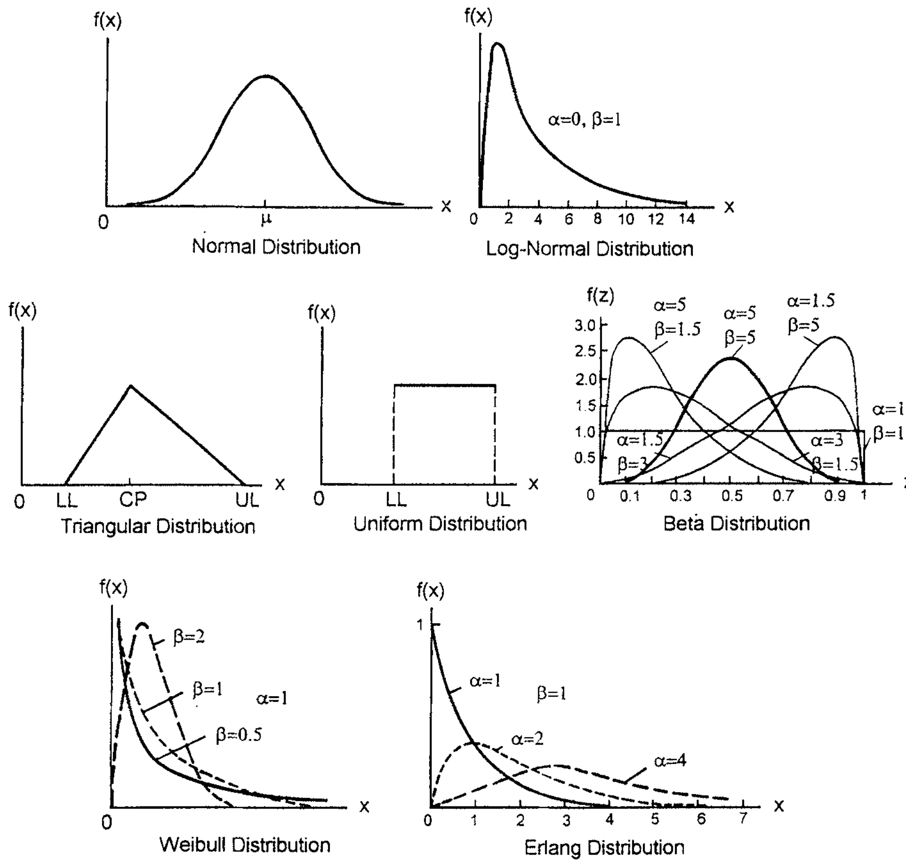
Code of the Data Set	Part Definition	Production Method	Nominal Size-Size Limits USL/LSL [mm or inch(+)]	Sample Size	Number of Samples	Number of Data in the Set	Measurement Device	Reading Accuracy [mm]
TAM1F.HMA	475 Massey Ferguson Engine gearbox outlet flange diameter	grinding	53,955/53,995	5	9	45	Micrometer	0.001
TAM2F.HMA	Gearbox outlet flange length	〃	28,53/28,63	5	9	45	〃	0.01
FRTSD.HMA	Rear axis gear tooth thickness	milling	6,82/6,86	5	25	125	〃	0.01
TOISD.HMA	Flatness of Maltese cross	turning	0.07	5	24	120	Dial Gage	0.01
TOKMC.HMA	FIAT Engine rear axis outlet dia.	〃	37,25/37,45	5	15	75	Micrometer	0.01
TAPDK.HMA	Gearbox rear axis oil sealent dia.	grinding	39,662/39,713	5	10	50	〃	0.001
TAR1D.ORS	Roller bearing(6000 series) outer ring diameter	〃	25,994/25,999	5	20	100	〃	0.001
TAR3D.ORS	Roller bearing(6308 series) outer ring diameter	〃	89,991/89,998	5	20	100	〃	0.001
TAR6Y.ORS	Roller bearing(6002 series) inner ring rollway diameter	〃	13,233/13,245	5	20	100	Dial Gage	0.001
TAR2E.ORS	Roller bearing(6307 series) outer ring width	〃	20,920/20,980	5	20	100	Micrometer	0.001
TOR4C.ORS	Roller bearing(6202 series) outer ring inner diameter	turning	29,00/29,10	5	18	87*	〃	0.01
TOREN.ORS	Roller bearing(6202 series) outer ring width	〃	11,50/11,75	5	16	80	〃	0.01
TAEKB.TS	UAZ engine exhaust valve length	grinding	116,978/117,000	3	10	30	〃	0.001
TAEMO.TS	UAZ engine inlet valve seat length	〃	4,07/4,57	3	11	32*	〃	0.01
TAEKM.TS	UAZ engine exhaust valve cam side diameter	〃	8,921/8,930	3	11	33	〃	0.001
DESA1.ASE	Selenoid valve center axis hole center	〃	0,115/0,135(+)	5	14	70	〃	0.001
DESA2.ASE	Selenoid valve center axis distance	〃	0,115/0,135(+)	5	28	140	〃	0.001
DESA3.ASE	Selenoid valve center axis distance	〃	0,415/0,435(+)	5	14	70	〃	0.001
GKT1C.MKE	Galvanized wire diameter	coating	1,93/2,07	5	24	120	〃	0.01
GKT2C.MKE	Galvanized wire diameter	〃	2,43/2,57	5	19	95	〃	0.01
TAEMM.HE	Electric motor rotor diameter	grinding	0,22/0,24	-	-	72	〃	0.001
TOSBB.MAN	Valve ring diameter(+)	turning	A:1,000/1,002 B:1,065/1,070	-	-	300	〃	0.005
DIYPB.MAN	Oil pump vane clearence(+)	drilling	A:0,3775/0,3825 B:0,3675/0,3725	-	-	240	〃	0.005

(*) insufficient number of measurements

Since the Weibull, Erlang and unit beta distributions start from zero (0) and unit beta function outside the interval (0, 1) gives probability density zero, all the collected measurements are normalized by using the formula

$$z_i=(x_i-a)/(b-a) \qquad (1)$$

where the a and the b are the smallest and the largest measurements, i.e. lower and upper limits of the distribution, respectively. The xi is the measured dimension variable and zi is the normalized value of xi (unit dimension). After normalization, the measured variables are distributed between the values 0 and 1. For normalized Weibull and Erlang distributions, the model parameter estimators give more accurate results than non-normalized variables. Unit dimensions and their distributions can be used successfully in normal, log-normal, triangular and uniform distribution functions since the frequency and shape characteristics of the distributions are not affected by normalisation. In this study, Eq. (1) is used for normalising the collected data sets. The



LL= Lower Limit, UL= Upper Limit, CP= Center Point, α,β = Model Parameters

Fig. 1 The distribution functions used in the study

first letter 'Z' in the code of the data set of Table 3 is used to differentiate the normalised data sets from the non-normalised data sets.

2.1.3 Validity test of proposed distribution functions

In order to asses the applicability of the proposed models to describe the behaviour of frequency distributions of dimensions, statistical test is necessary to see if proposed probability distributions with estimated parameters actually fits the measured data sets. There are various statistical tests to check the goodness of fit. One of the commonly used tests in statistics is the Chi-square (χ^2) goodness of fit test and it was also used in this study. The summary of the procedure applied in this study is given below.

i) The sample data x^*, here the set of dimensions, is grouped into a proper number of equal width intervals or cells (l) varying between 5 and 30. There must be at least five measurements in each cell. For the manufacturing applications, Ishikawa (1976) suggests that the test give reliable results for 5-7 intervals if the measurements are less than 50. He recommends 6-10 intervals for 50-100 measurements, 7-12 intervals for 100-250 measurements and 10-20 intervals for measurements above 250 for the Chi-square test.

ii) From the available family of statistical distributions (in this study, normal, log-normal, triangular, uniform, Weibull, Erlang and beta probability density functions), a model distribution function $F_0(x)$ is hypothesised to represent and fit the sample data

iii) The parameters of the hypothesised model, $F_0(x)$, are estimated from the data by using estimation techniques.

iv) The following statistic is calculated from the observed and expected model frequencies.

$$\chi^2=\sum_{i=1}^{l}\frac{(O_i-E_i)^2}{E_i} \tag{2}$$

where O_i is the number of observed data in cell i, E_i is the number of expected data in cell i and l is the number of cells. Values of Ei are calculated by using the postulated distribution function, $F_0(x)$. It is known from the probability theory that the statistic χ^2 is distributed as a Chi-square variable with $\nu=1-k-1$ degrees of freedom. Here k is the number of model parameters, ν is the degrees of freedom of the χ^2 distribution. The k value for triangular distribution is three, and for the other distribution functions used in this study, is two.

v) The terms of the χ^2 statistic above measure the discrepancy between the observed and postulated theoretical class frequencies. Smaller χ^2 values indicate better fit of the distribution. From the standard χ^2 tables, the level of significance (α) can be found by using χ^2 value and degrees of freedom (ν). For a given significance level α and degrees of freedom α, the critical value ($\chi^2{}_c$) is obtained from theoretical chi-square tables. The postulated model $F_0(x)$ and the sample data give rise to single value (χ^2) of the test statistic. If $\chi^2{}_c>\chi^2$, the hypothesis that $F_0(x)$ is the underlying measurement is accepted; otherwise it is rejected. If two different hypothesised models are tested for the same experimental data, the model which gives the smaller chi-square value or bigger level of significance, for the same degree of freedom, indicates the better fit.

In this study, for the same set of dimensions, chi-square test is applied to normal, log-normal, triangular, uniform, Weibull, Erlang and unit beta distribution functions, and χ^2 values and level of significances are found for the same degrees of freedom. The STATGRAF software package (Cooke, 1979) is used for the χ^2 test of the models.

2.2 Results and discussion

In the study, normal, log-normal, uniform, triangle and Weibull models are tested for the 23 sets of data given in Table 1. The Erlang and beta models are only tested for TOISD. HMA set, which has variables between 0 and 1. Although DESA1. ASE, DESA2. ASE, DESA3. ASE, TAEMM. HE, TDSBB. MAN and DIYPB. MAN sets have variables in between 0 and 1, the parameter estimates of beta and Erlang distributions were not computable due to the very close values of the distribution variables. The chi-square test parameters and results (l, k, ν, χ^2 α) for sample hypothesised models are given in Table 2 for the 23 sets of dimensions. In the last column of the Table 1, the result of the χ^2 test is summarized by giving the first letter of the names of the models in the order of increasing values of χ^2, or decreasing level of significance. For TAM1F. HMA set in Table 2, the N, L, U, T^+ listing is given which indicates that the model which gives the minimum χ^2 value is 'normal' (shortly N) and triangular (shortly T) is in the 4th position. The '+' sign placed as a superscript on the letter is the indication of significance level less than 0.05 for the model. The best 4 distribution functions given in Table 2 are plotted on the distributions of the measurement sets. Sample plots are given in Fig. 2. The figure indicates clearly that it is impossible to decide the best-fit model by visual inspection. This verifies the strong need for the chi-square test.

The results obtained from the chi-square tests for measurement sets (non-normalised) (Table 2) can be summarized as follows:

1) In 14 of the 23 measurement sets, the normal model gives the best fit. For 7 of these 14 sets of data, the level of significance is less than 0.05.

2) In the 15 measurement sets log-normal model gives the same level of significance with normal and in 5 sets log-normal gives better fit (higher significance level) than normal model. So, the log-normal model gives much better fit than normal model when all the measurement sets are considered.

Table 2 Chi-square test results of sample distribution functions for measurement sets

Code of the Data Set	Number of Data	BETA							NORMAL							WEIBULL							FITNESS OF THE MODELS
		Model Parameters		Chi-square Test					Model Parameters		Chi-square Test					Model Parameters		Chi-square Test					
		$\bar{x}$	s	l	k	v	χ^2	SL	$\bar{x}$	s	l	k	v	χ^2	SL	LL	UL	l	k	v	χ^2	SL	
AM1F.HMA	45	53.981	0.0033	6	2	3	2.03	0.564	53.981	0.0033	6	2	3	2.03	0.564	53.975	53.988	6	2	3	5.21	0.156	N, L, U, T+
TAM2F.HMA	45	28.587	0.0011	5	2	2	15.9	0.0003	28.587	0.0011	5	2	2	15.93	0.0003	28.57	28.61	5	2	2	7.78	0.0203	U+, N+, L+, T+
FRTSD.HMA	125	6.837	0.010	7	2	4	75.3	1.66E-1	6.837	0.010	7	2	4	75.23	1.8E-15	6.82	6.86	7	2	4	78.13	0.00000	U+, N+, L+, T+
TOISD.HMA	120	0.046	0.023	8	2	5	5.52	0.354	0.0471	0.0276	8	2	5	23.04	0.0003	0.01	0.14	8	2	5	158.1	0.0	N, W, B+, L+
TOKMC.HMA	75	37.312	0.0322	6	2	3	17.0	0.0007			6	2	3	17.04	0.00069	37.25	37.38	6	2	3	21.68	0.00007	N+, L+, U+, T+
TAPDK.HMA	50	39.693	0.0035	6	2	3	17.9	0.0004	39.693	0.0035	6	2	3	17.98	0.0004	39.685	39.701	6	2	3	30.0	1,3E-06	N+, L+, T+, U+
TAR1D.ORS	100	25.996	0.00075	6	2	3	8.09	0.044	25.9958	0.00075	6	2	3	8.09	0.044	25.993	25.9975	7	2	4	49.62	4.3E-10	N+, L+, T+, U+
TAR3D.ORS	100	89.994	0.00091	7	2	4	25.5	0.00003			7	2	4	25.53	0.000039	89.992	89.996	7	2	4	35.46	3,7E-07	N+, L+, T+, U+
TAR6Y.ORS	100	13.23	0.0027	8	2	5	65.7	7.8E-13	13.23	0.0027	7	2	4	17.23	0.000005	13.233	13.243	8	2	5	47.07	5.4E-09	L+, T+, U+, N+
TAR2E.ORS	100	20.973	0.0023	8	2	5	18.6	0.0023			8	2	5	18.62	0.0023	20.970	20.979	8	2	5	43.71	2E-08	N+, L+, T+, U+
TOR4C.ORS	87	29.063	0.016	7	2	4	6.98	0.136	29.063	0.016	7	2	4	6.99	0.136	29.02	29.10	7	2	4	27.57	0.000015	N, L, T+, U+
TOREN.ORS	80	11.731	0.0847	6	2	3	26.7	0.00000	11.731	0.0847	6	2	3	27.18		11.5	11.9	6	2	3	43.8	1.6E-09	N, U+, L+, T+
TAEKB.TS	30	116.991	0.0075	5	2	2	1.50	0.472	116.991	0.0075	5	2	2	1.50	0.472	116.97	117.0	5	2	2	5.86	0.053	N, L, T, U
TAEMO.TS	32	4.367	0.149	5	2	2	1.24	0.536	4.367	0.149	5	2	2	1.225	0.541	4.15	4.62	5	2	2	2.43	0.296	L, N, U, T+
TAEKM.TS	33	8.925	0.0042	5	2	2	0.49	0.779	8.925	0.0042	5	2	2	0.500	0.778	8.916	8.932	5	2	2	4.526	0.104	N, L, U, T+
DESA1.ASE	70	0.129	0.0019	6	2	3	5.84	0.119	0.129	0.0019	6	2	3	5.90	0.116	0.125	0.134	6	2	3	15.99	0.001	N, L, T, U+
DESA2.ASE	140	0.130	0.0039	7	2	4	12.9	0.011	0.130	0.0039	7	2	4	13.21	0.0102	0.121	0.137	8	2	5	24.87	0.00015	N+, L+, T+, U+
DESA3.ASE	70	0.428	0.0024	7	2	4	10.1	0.038	0.428	0.0024	7	2	4	10.26	0.036	0.422	0.433	7	2	4	29.42	0.000006	N+, L+, T+, W+
GKT1C.MKE	120	2.006	0.0298	7	2	4	30.5	0.00000	2.0055	0.0299	7	2	4	30.74	0.000003	1.95	2.05	7	2	4	16.24	0.0027	U+, L+, N+, T+
GKT2C.MKE	95	2.501	0.032	6	2	3	1.64	0.65	2.501	0.032	6	2	3	1.54	0.67	2.40	2.57	7	2	4	36.86	1,6E-07	L, N, T+, U+
TAEMM.HE	72	0.228	0.0050	8	2	5	4.76	0.444	0.228	0.0050	8	2	5	4.51	0.477	0.22	0.24	8	2	5	14.20	0.014	L, N, T, U+
TOSBB.MAN	300	0.011	0.00097	9	2	6	14.9	0.0205	0.011	0.00097	10	2	7	11.86	0.105	0.009	0.014	10	2	7	134.2	0	L, N+, T+, W+
DIYPB.MAN	240	0.0062	0.00107	9	2	6	10.1	0.116	0.00625	0.00107	9	2	6	10.55	0.103	0.004	0.009	9	2	6	120.0	0	U, N, L, W+

(+) : SL less than 0.05

SL : Significance Level

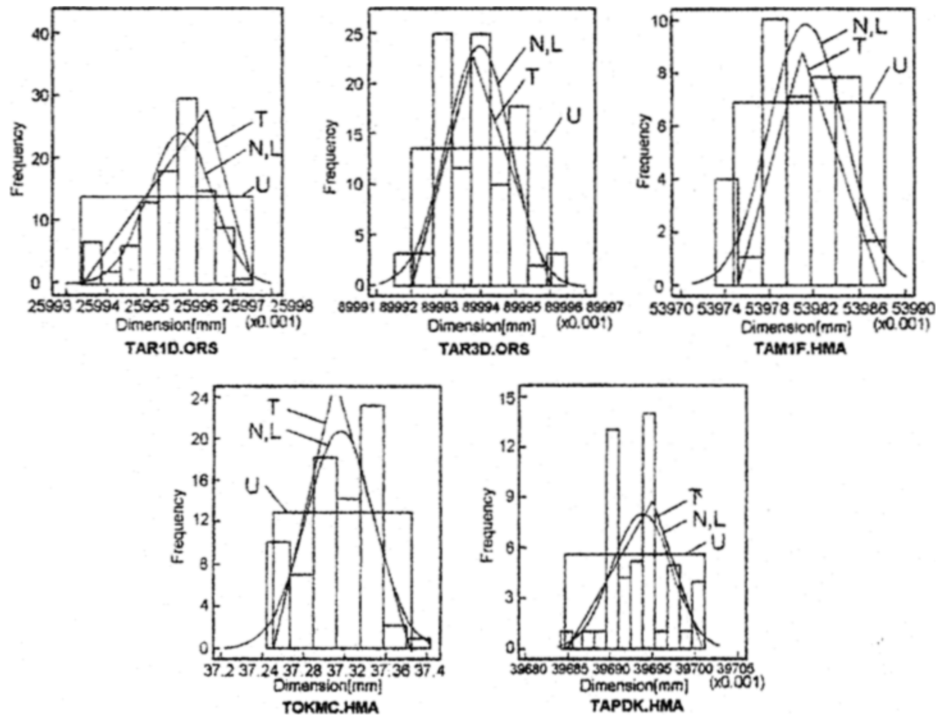


Fig. 2 Frequency distributions of sample measurement sets and the plot of the best three distribution functions (N=Normal, L=Log-normal, T=Triangular, U-Uniform, W-Weibull, E=Erlang, B=Bata)

3) Uniform distribution function models 3 data sets better than the other models. This model is generally placed in the 3rd or 4th places after normal and log-normal models.

4) Triangular distribution models only one data set better than the other distribution functions. In all the data sets, it is placed in the 3rd and 4th position.

5) Weibull model parameters (estimated) and χ^2 values are calculated only for 7 sets of measurements. In these 7 sets, the model is in the 2nd place once and 4th place three times.

According to the above results, it is found that the log-normal and normal distribution functions model the measurement sets better than the other distribution functions. In modelling, the log-normal distribution gives better results than normal due to the advantage of modelling left-skewness of the data sets. Uniform distribution could be considered as the 3rd best model in modelling the measurement sets.

In the above analysis Erlang, Weibull and beta distribution functions cannot be used in modelling due to the non-normalised data sets. In the second part of the study, 23 data sets are normalised according to Eq. (1). The chi-square test results (l, k, ν, χ^2, α) for sample hypothesised models are given in Table 3 for the 23 sets of normalised dimension measurements. The best 4 distribution functions given in Table 3 are plotted on the frequency distributions of the normalised measurement sets. Sample plots showing the best three distributions are given in Fig. 3. When Tables 2 and 3 are analysed together, it can be easily deduced that the goodness of fit of the models (last column of the Table), except Erlang, Weibull and beta distributions, are same. As an example for the set, TAM1F. HMA the goodness of fit list is in the order of N, L, U, T in Table 2. In the normalised set, ZTAM1F. HMA the relative positions of N, U, T are not changed but a new model (beta) is included in the list i.e. B, N, U, T. Some small changes in the orders of models in Tables 2 and 3 are due to the small variations in

Table 3 Chi-square test results of sample hypothesized distribution functions for normalized distributions

Code of the Data Set	Number of Data	BETA							NORMAL							WEIBULL							FITNESS OF THE MODELS
		Model Parameters		Chi-square Test					Model Parameters		Chi-square Test					Model Parameters		Chi-square Test					
		α	β	l	k	v	χ^2	SL	x	s	l	k	v	χ^2	SL	α	β	l	k	v	χ^2	SL	
ZTAM1F.HMA	45	1.445	1.472	5	2	2	1.101	0.576	0.495	0.252	5	2	2	1.185	0.552	1.313	0.518	5	2	2	4.69	0.095	B, N, U, T
ZTAM2F.HMA	45	1.00	1.179	4	2	1	0.569	0.450	0.444	0.281	4	2	1	0.52	0.217	0.972	0.441	4	2	1	5.49	0.019	B, N, U, E+
ZFRTSD.HMA	125	1.00	1.298	7	2	4	90.34		0.434	0.273	7	2	4	74.81		1.034	0.437	7	2	4	107.21	0.00	**
ZTOISD.HMA	120	1.492	3.896	7	2	4	14.14	0.007	0.276	0.177	8	2	5	20.44	0.001	1.375	0.297	8	2	5	20.60	0.000	B+, N+, W+, E+
ZTOKMC.HMA	75	1.388	1.546	8	2	5	13.65	0.020	0.476	0.252	7	2	4	11.58	0.020	1.661	0.520	7	2	4	24.58		B=N, U+, T+
ZTAPDK.HMA	50	2.234	1.896	7	2	4	15.64	0.0035	0.540	0.220	7	2	4	16.59	0.0023	2.382	0.597	7	2	4	29.41		T+, B+, N+, U+
ZTAR1D.ORS	100	3.316	2.297	6	2	3	7.36	0.068	0.580	0.190	6	2	3	6.88	0.068	3.177	0.637	6	2	3	8.11	0.043	B=N, W+, T+
ZTAR3D.ORS	100	1.834	2.068	7	2	4	15.66	0.0035	0.470	0.225	7	2	4	24.14		1.910	0.516	7	2	4	23.18	0.000	B+, W+, T+, N+
ZTAR6Y.ORS	100	1.00	1.302	8	2	5	35.26		0.405	0.274	8	2	5	38.20		1.132	0.418	8	2	5	56.26		**
ZTAR2E.ORS	100	1.00	2.00	8	2	5	13.16	0.022	0.306	0.235	8	2	5	14.40	0.013	0.735	0.273	7	2	4	48.99		B+, N+, E+, W+
ZTOR4C.ORS	87	2.556	2.150	6	2	3	2.297	0.513	0.543	0.208	6	2	3	3.276	0.350	2.445	0.595	6	2	3	8.90	0.030	B, N, W+, T+
ZTOREN.ORS	80	2.573	1.882	6	2	3	27.58		0.577	0.211	6	2	3	27.40		1.761	0.608	7	2	4	58.23		B+, N+, T+, W+
ZTAEKB.TS	30	1.455	1	5	2	2	1.485	0.476	0.619	0.299	5	2	2	1.387	0.499	1.775	0.674	5	2	2	3.26	0.195	N, B, W, T
ZTAEMO.TS	32	1	1.15	5	2	2	0.936	0.626	0.461	0.317	5	2	2	1.66	0.435	1.319	0.495	5	2	2	1.467	0.480	B, N, W, E
ZTAEKM.TS	33	1.373	1.061	5	2	2	0.242	0.885	0.564	0.267	5	2	2	1.115	0.572	1.900	0.617	5	2	2	1.505	0.470	B, N, W, E
ZDESA1.ASE	70	2.34	2.15	6	2	3	2.23	0.525	0.522	0.213	6	2	3	5.91	0.115	2.431	0.578	6	2	3	51.47		B, N, T, U+
ZDESA2.ASE	140	1.918	1.431	6	2	3	5.46	0.141	0.572	0.237	6	2	3	16.33	0.001	2.52	0.638	6	2	3	18.40	0.000	B, N+, T+, W+
ZDESA3.ASE	70	2.548	1.945	7	2	4	8.60	0.071	0.567	0.211	7	2	4	5.40	0.248	2.687	0.624	7	2	4	9.39	0.052	N, B+, W+, T+
ZGKT1C.MKE	120	1.15	1.00	8	2	5	6.57	0.254	0.546	0.299	9	2	6	14.75	0.022	1.445	0.585	9	2	6	30.32		B, U, N+, W+
ZGKT2C.MKE	95	3.39	2.29	7	2	4	8.69	0.069	0.596	0.189	7	2	4	6.45	0.167	2.964	0.648	7	2	4	10.44	0.033	N, B, W+, T+
ZTAEMM.HE	72	1.128	1.644	8	2	5	2.40	0.790	0.406	0.252	8	2	5	4.95	0.421	1.547	0.448	8	2	5	3.99	0.55	B, W, N, E
ZTOSBB.MAN	300	2.297	3.172	10	2	7	14.22	0.047	0.420	0.194	10	2	7	15.01	0.035	2.232	0.471	10	2	7	14.44	0.043	B+, W+, N+, E
ZDIYPB.MAN	240	1.943	2.363	8	2	5	15.02	0.0102	0.451	0.216	9	2	6	31.69		2.134	0.506	9	2	6	9.84	0.131	W, E, T+, B+

** : Very small SL value

(+) : SL less than 0.05

Table 4 UCL and LCL values for beta and normal distributions and their differences for the sample measurement sets

Data Set	α	β	ZUCLb	ZLCLb	UCLs[mm]	LCLs[mm]	UCLb[mm]	LCLb[mm]	DUCLsb[%]	DLCLsb[%]	RCLRsb
ZTAM1F.HMA	1.445	1.472	0.9912	0.0079	53,986	53,977	53,9878	53,9751	0,0033	−0,0035	1,41
ZFRTSD.HMA	1.000	1.298	0.9933	0.0012	6,86	6,82	6,859	6,82005	−0,014	0,0007	0,97
ZTOKMC.HMA	1.398	1.546	0.9888	0.0065	37,34	37,28	37,378	37,2509	0,1	−0,078	2,11
ZTAPDK.HMA	2.234	1.896	0.9831	0.0339	39,698	39,688	39,7007	39,6855	0,0068	−0,0063	1,52
ZTAR1D.ORS	3.316	2.297	0.9765	0.0843	25,9967	25,9949	25,9974	25,9938	0,0027	−0,0042	2
ZTAR3D.ORS	1.834	2.068	0.9729	0.0160	89,9953	89,9925	89,9959	89,9921	0,0007	−0,0004	1,36
ZTOR4C.ORS	2.556	2.150	0.9765	0.0459	29,081	29,045	29,098	29,024	0,06	−0,07	2,05
ZTOREN.ORS	2.573	1.882	0.9854	0.0517	11,814	11,648	11,894	11,52	0,68	−1,09	2,25
ZTAEKB.TS	1.455	1.000	0.9990	0.0115	116,998	116,983	116,999	116,9753	0,00085	−0,0066	1,58
ZTAEMO.TS	1.000	1.150	0.9965	0.0013	4,73	4,004	4,618	4,1506	−2,36	3,66	0,64
ZTAEKM.TS	1.373	1.061	0.9984	0.0083	8,934	8,915	8,9319	8,9161	−0,023	0,012	0,83
ZDESA1.ASE	2.340	2.150	0.9747	0.0355	0,133	0,127	0,1338	0,1253	0,6	−1,33	1,42
ZDESA2.ASE	1.918	1.431	0.9941	0.0255	0,136	0,124	0,1369	0,1214	0,66	−2,09	1,29
ZDESA3.ASE	2.548	1.945	0.9834	0.0491	0,431	0,425	0,4328	0,4225	0,42	−0,59	1,72
ZGKT1C.MKE	1.150	1.000	0.9987	0.0035	2,05	1,96	2,049	1,9504	−0,49	−0,48	1,1
ZGKT2C.MKE	3.398	2.296	0.9770	0.0890	2,55	2,45	2,566	2,415	0,62	−1,42	1,51

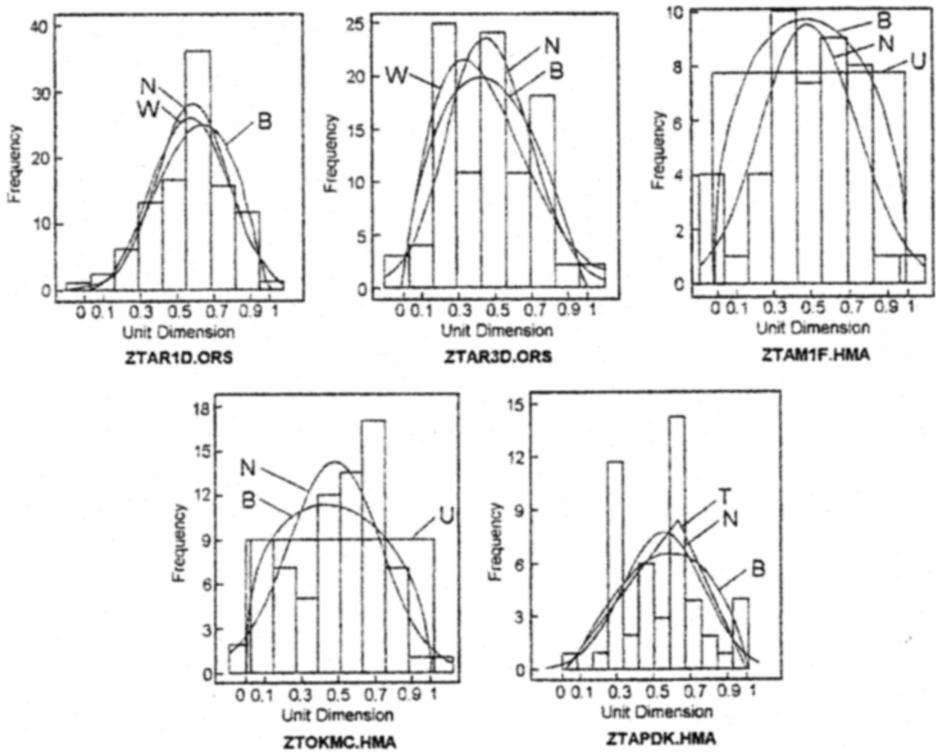


Fig. 3 Frequency distributions of normalized sample measurement sets and the plot of the best three distribution functions(N=Normal, L=Log-nomal, T=Triangular, U=Uniform, W=Weibull, E=Erlang, B=Bata)

number of intervals (cells) of the frequency distributions of the normalised and non-normalised data sets.

The results obtained from the chi-square tests for normalised measurement sets can be summarized as follows:

1) For all the proposed models the significance levels of ZFRTSD. HMA and ZTARGY. ORS measurement sets are in the order of $10^{-6}-10^{-15}$. Therefore, these two normalised data sets are discarded from the analysis, which would possibly give unreliable and misleading results.

2) In 16 of 21 normalised data sets, beta distribution function gives the best fit. In 5 of these 16 sets, the level of significance is less than 0.05.

3) In 3 of 21 normalised sets, normal distribution function gives the best fit. Normal distribution is the second model in the sets for which beta is the best.

4) Weibull and triangular distribution functions are best in only one normalised set. In normalised sets of ZTAPDK. HMA and ZDIYPB. MAN in which the triangular and Weibull distributions are the best, respectively, the second best fits are beta distribution functions.

5) Erlang model generally gives a poor fit for the normalised data sets.

6) Although the Weibull model seems to produce a better fit than the Erlang, in 14 of 16 normalised data sets, the Weibull gives a poorer fit than the beta and normal distributions.

7) Triangular and uniform models are generally placed in the 3rd or 4th place after beta and normal. Only in the ZTAPDK. HMA set does the triangular distribution give the best fit.

8) After the careful visual inspection of the normalised frequency distributions, it is observed that the right- and left-skewness characteristics of the dimension frequency distributions are best represented by the beta distribution model.

It is clear from the above given results and observations that the beta distribution function models the distribution of dimensions better than normal and the other commonly used statistical models. Since normalisation technique only changes the scale of the non-normalised distributions without changing the frequency and shape characteristics, it could be stated that the beta distribution function is the best model for reflecting the behaviour of the measurement sets. The model parameters α and β control the skewness, shape and scale of the model. Erlang and Weibull distribution functions are also known to be good in reflecting the right- or left-skewness. However, their model parameter estimators are found weak when compared with beta, which will eventually result in a poor chi-square fit.

The beta model is also proposed by He (1991) to model the dimension distributions. In his work, only the beta model is used and the superiority of the model with respect to the other distribution functions is not emphasised. In his work, only two data sets are used. One of the two sets of data is used to explain the use of the distribution for a small data set (16 measurements) and the other is taken from Bennet's (1964) study to explain the procedure for a large data set.

3. The Adoption of Beta Model to x-bar Control Charts

In this part of the paper, an attempt is made for the adoption of beta model to x-bar control charts.

3.1 Normal-distribution-based x-bar charts

16 sets of measurements (Table 1) are used in the construction of x-bar control charts. In this study, the x-bar charts conducted for the normal data are based on sample size of five. Schilling and Nelson (1976) showed that the Shewhart x-bar chart for modelling means works well with a sample size four or five. Authors of this work believe that real motivation for improving the control chart methods for skewed data occurs when a small sample size is required, such as $n=1$.

The Upper and Lower Control Limits of the charts are abbreviated as UCLs and LCLs. The letter s in the abbreviations indicates 'standard' (normal-distribution-based) control limits. Sample calculation of UCLs and LCLs for TAM1F. HMA data set is given below.

$$UCLs=x+A_2R=53.9814+0.577(0.008)=53.986\text{mm}$$
$$LCLs=x-A_2R=53.9814-0.577(0.008)=53.977\text{mm}$$

Here, x=53.9814mm is the average of 9 samples (53.979, 53.980, 53.981, 53.982, 53.982, 53.983, 53. 980, 53.983 and 53.983 mm) with 5 measurements in every sample. R (the maximum deviation of measurements in a sample) values for the samples are 0.007, 0.008, 0.010, 0.013, 0.012, 0.005, 0.004, 0.005, 0.004 mm and their mean value (R) is 0. 008 mm. A_2 chart constants are dependent on sample size and can be found in statistical quality control books in tabulated forms (Bain, 1978; Bury, 1975). For sample size (n) of 5, A_2 value is 0.577. In this study, control limit calculations are made by using the formula $CLs=x\pm A_2R$ for the simplicity of the calculations. The control limit values for standard (normal model based) procedure is given in Table 4.

3.2 Beta-distribution-based x-bar charts

For most of the manufacturing companies, '3 defectives in 1000 parts' is an acceptable limit (type I error rate of 3/1000). With the normal distribution this error rate is represented very closely by $+/-3\sigma$ distance from the distribution mean which gives symmetric tail probabilities of about 0.0015 on each side. Due to the skewed shape of the beta distribution, it is impossible to find equal symmetric distances from the distribution mean which would give the equal tail probabilities mentioned above. Therefore, it is almost impossible to find chart constants (A_1, A_2, B_1, B_2 etc. as in normal model) for the beta-model-based control charts. No references are available in the literature for adjusting control limits of a x-bar chart for skewed data by using the beta distribution and other distributions different from the normal. Upper and lower control limits of the beta model can be estimated by using the probability limit method. That is, it can be estimated by obtaining the percentiles of the beta distribution. In this study, the following procedure is applied to estimate upper and lower control limits for the beta model (UCLb and LCLb):

1. Normalise all the measurements (variables) in the data set.
2. Obtain the α and β values of the beta model by using the STATGRAPH package (If a set of measurements has the population mean μ and variance σ^2, the beta model parameters α and β are estimated from equations A2 and A3 given in the Appendix.).
3. Obtain the critical values, which would give 0.0015 and 0.9985 probabilities in the beta function. The critical value, which would give 0.0015 probability is the ZLCLb. Similarly, the critical value which would give 0.9985 is the ZUCLb. The obtained values are the normalised control limits of the beta model and the letter Z indicates that these values are calculated from normalised distributions (ie. unit dimensions).
4. Use the following formulation to convert ZUCLb and ZLCLb values to UCLb and LCLb.

$$UCLb = ZUCLb(bb - ab) + ab \quad (3)$$
$$LCLb = ZLCLb(bb - ab) + ab \quad (4)$$

Here, ab and bb are the lower and upper limits of the beta distribution population, and they are taken as the minimum and the maximum measurements (dimensions) in the set. It should be known that the maximum of original sample (especially for cases with a small original sample size) is not the maximum for the whole population. If the maximum and minimum used to normalise a new sample, are taken from the data used to generate the control limits, there is a possibility that the new sample will have a data point outside the maximum and minimum of the original sample.

The percent difference between upper control limits of standard procedure (normal-model-based) and beta distribution populations (DUCLsb), and the percent difference between lower control limits of standard procedure and beta distribution population (DLCLsb) are calculated by using the following equations:

$$DUCLsb = (UCLb - UCLs).\ 100/UCLs \quad (5)$$
$$DLCLsb = (LCLb - LCLs).\ 100/LCLs \quad (6)$$

To compare the range between UCL and LCL values obtained both from beta model and standard (normal distribution based) procedure (RCLRsb) the following equation is used:

$$RCLRsb = (UCLb - LCLb)/(UCLs - LCLs) \quad (7)$$

The results obtained from Eqs. (3) to (7) is summarised in Table 4 for the 16 sets of measurements.

3.3 Sample case

For the ZTAM1F. HMA data set, the α and β values are found from the STATGRAPH package as 1,445 and 1.472. For the 0.0015 and 0.9985 probabilities, the critical values are 0.0079 and 0.9912 (ie. ZLCLb and ZUCLb). Equations (3) and (4) are used to find UCLb and LCLb. From Eqs. (3) and (4);

$$UCLb = 0.9912(53.988 - 53.975) + 53.975 = 53.9878\text{mm}.$$
$$LCLb = 0.0079(53.988 - 53.975) + 53.975 = 53.9751\text{mm}.$$

The percent difference between UCL and LCL values for beta and standard values are (Eqs. (5) and (6));

$$DUCLsb = (53.9878 - 53.986).\ 100/53.986 = 0.0033\%$$
$$DLCLsb = (53.9751 - 53.977).\ 100/53.977 = -0.0035\%$$

Comparison of the range between beta model control limits and standard procedure control limits is performed by using Eq. (7):

$$RCLRsb = (53.9878 - 53.9751)/(53.986 - 53.977) = 1.41$$

From numerical results obtained for TAM1F. HMA measurement set, it is clear that the UCLb is higher than the UCLs, and LCLb is lower than LCLs. The range between UCLb and LCLb values is 41% bigger than the range obtained from standard (normal-model-based) procedure.

3.4 Results and discussion

The sample analysis given in Sec. 3.3 is repeated for 16 sets of measurements and summary of the results is given in Table 4. Some sample x-bar charts showing both standard (normal-model-

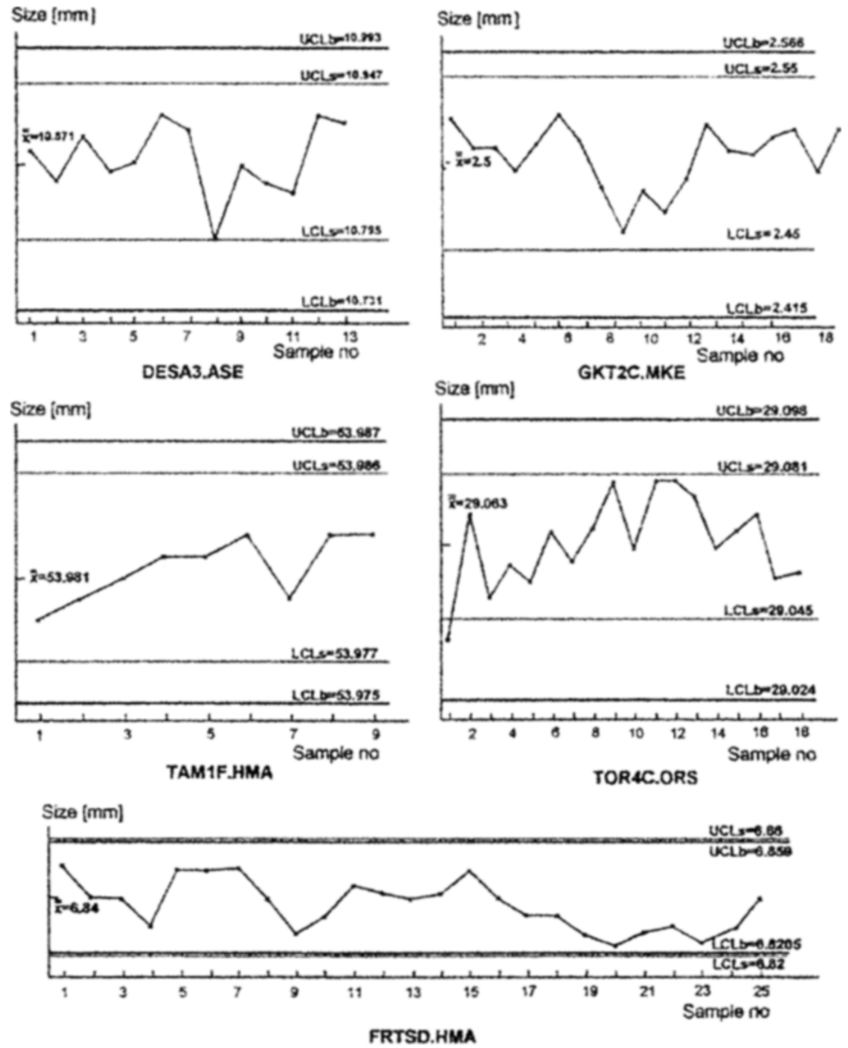


Fig. 4 Some sample $\bar{x}$ control charts for the selected measurement sets

based) and beta distribution control limits are given in Fig. 4. The following results can be deduced from Table 4:

1. For the twelve of the sixteen data sets, the UCLb is higher than UCLs, and LCLb is lower than LCLs.
2. The range between UCLb and LCLb values is generally wider than that of UCLs and LCLs. The ratio of the range of beta control limits to the range of standard procedure (RCLRsb) is between 1 and 2.

The DUCLsb and DLCLsb values indicate that the beta-model-based control limits are not symmetric with respect to x value. The control limits of the standard procedure are closer to the x value than that of beta model (narrower control zone). From the above results it can be deduced that the normal-model-control limits provides closer control over sample means than that of beta model. It is possible that some sample averages (ie. x-bar values) which fall beyond the control limits of the normal model will be considered acceptable (safe) by the beta-model-based control limits. By using the proposed, one can estimate the control limits of beta model to monitor the ongoing process.

4. Conclusions

In this study, experimental and theoretical efforts are spent to model the behaviour of dimension distributions of machined workpieces. 23 sets of dimension distributions of different parts are used in this study. In order to model the behaviour of the collected data, 7 different statistical distribution functions, namely, normal, log-normal, triangular, uniform, Weibull, Erlang and unit beta distributions, were used. The beta distribution is found the best statistical distribution function in representing the frequency distributions of the measurement sets by using chi-square goodness of fit tests.

In the second stage of the work, the upper and lower control limits of the beta model are estimated by obtaining the percentiles of the distribution. The tail probabilities of 0. 0015 and 0. 9985 are used to find the critical values for beta distribution function and these values are taken as the upper and lower control limits of the control charts by using type I error rate of 3/1000. It is found that mostly the UCL of the beta model is higher than that of normal model and LCL of the beta model is lower than that of the normal. So, the normal-model-based UCL and LCL provide closer control over sample means than that of beta model. It can be inferred that the beta-based chart would result in fewer alarms, but it is difficult to know when it will detect true out-of-control points.

Many quality control engineers believe that in practice, normality is not too much of a problem in the case of x-bar charts since the errors associated with its use are relatively small. The authors believe that the real contribution for improving the control chart methods for skewed data occurs when a small sample is required, such as the case where $n=1$. Development of an approach for adjusting the control limits of a control chart for skewed data, which can be modelled by using different models than normal, will be another contribution in the field.

References

Bai, D. S. and Chois, I. S., 1995, "Xbar-Control and R-Control Charts for Skewed Populations," *Journal of Quality Technology*, 27, (2), 120~131.

Bain, L. J., 1978, "Statistical Analysis of Reliability and Life Testing," *Marcel Dekker*, New York.

Balakrishnan, N. and Kocherlakota, S., 1986, "Effects of Non-normality on Xbar Charts -Assignable Cause Model," *Sankhya-The Indian Journal of Statistics Series B*, 48, 439~444.

Bennet, G., 1964, "The Application of Probability Theory to the Allocation of Engineering Tolerances," Ph. D. Thesis, University of New South Wales, Australia.

Bjorke, Q., 1978 "Computer Aided Tolerancing," Tapir Publishers, Trondheim, Norway.

Borror, C. M., Montgomery, D. C. and Runger, G. C., 1999, "Robustness of the EWMA Control Chart to Non-Normality," *Journal of Quality Technology*, 31, (3), 309~316.

Bury, K. V., 1975, "Statistical Models in Applied Science," John Wiley, New York.

Burr, I. W., 1958, "Some Theoretical and Practical Aspects of Tolerances for Mating Parts," *Quality Control*, 15, (3), 18.

Chan, L. K., Hapuarachi, K. P. and MacPherson, B. D., 1988, "Robustness of Xbar and R-Charts," *IEEE Transactions on Reliability*, 37, (1), 117~123.

Choobineh, F. and Ballard, J. L., 1987, "Control Limits of QC Charts for Skewed Distributions Using Weighted Variance," *IEEE Transactions on Reliability*, 36, (4), 473~477.

Cooke, P., 1979, "Statistical Inference For Bounds of Random Variables," *Biometrica*, 66, 367~374.

Crafts, J. W., 1952, "Assembly Tolerance Problem," *Engineer* (London), 206, 918~928.

Doyle, L. E., 1951, "Statistical Aids for Tool Engineers," *Tool Engineer*, 26, (1), 48~51.

Fortini, E. T., 1956, "Dimension Control in Design," *Machine Design*, 28 (28), 82~89.

Fortini, E. T., 1967, *Dimensioning For Interchangeable Manufacture*, Industrial Press Inc., New York.

Gibson, J., 1951, *A New Approach to Engineering Tolerances*, The Machinery Publishing Co. Ltd., Brighton and London.

Gladman, C. A., 1959, "Techniques For Applying Probability to the Tolerancing of Machined Dimensions," C. S. I. R. D. Australian National Std. Lab. Tech. Paper No. 11.

Gladman, C. A., 1980, "Applying Probability in Tolerance Technology," *The Institution of Engineers*, Australia, 5, (2), 82~88.

Haridy, A. M. A. and El-Shabrawy, A. Z., 1996, "The Economic Design of Cumulative Sum Charts Used to Maintain Current Control of Non-Normal process Means," *Computers in Industrial Engineering*, 31, (3-4), 783~790.

He, J. R., 1991, "Estimating the Distributions of Manufactured Dimensions With the Beta Probability Density Function," *Int. J. Mach. Tools Manuf*., 31, (3), 383~396.

Ishikawa, K., 1976, "Guide to Quality Control," *Asia Productivity Organisation*, Tokyo, 9.

Kılınç, B., 1999, "Computer Aided Statistical Investigation on the Variation of Dimensions of Workpieces Machined Under the Same Machining Conditions," M. Sc. Thesis, Gazi University, Ankara.

Lashkari, R. S. and Rahim, M. A., 1982, "An Economic Design of Cumulative Sum Charts to Control Non-Normal Process Means," *Computers in Industrial Engineering*, 6, (1), 1~18.

Mansoor, E. M., 1960, "The Dimension Analysis of Engineering Designs," M. Sc. Thesis, University of Melbourne.

Mansoor, E. M., 1964, "The Application of Probability to Tolerances Used in Engineering Designs," *Proc. Instn. Mech. Engrs*., 178, (1), 29~51.

Nelson, P. R., 1979, "Control Charts for Weibull Processes with Standards Given," *IEEE Transactions on Reliability*, 28, (4), 283~288.

Nelson, L. S., 1984, "The Shewhart Control Chart: Tests For Special Causes," *J. of Quality Technology*, 16, (4), 237~239.

Padget, W. J. and Spurrier, J. D., 1990, "Shewhart Type Charts for Percentiles of Strength Distributions," *Journal of Quality Technology*, 22, (4), 283~288.

Shainin, D., 1949, "Cost-cutting Chance Laws Can Control Design Tolerances," *Machine Design*, 21 (6), 130~140.

Shilling, E. G. and Nelson, P. R., 1976, "The Effect of Non-Normality on the Control Limits of X-bar Charts," *Journal of Quality Technology*, 8, (4), 183~188.

Shore, H., 1998, "A New Approach to Analysing Non-Normal Quality data with Application to Process Capability Analysis," *Int. J. Prod. Res*., 36, (7), 1917~1933.

STATGRAF, 1993, Statistical Graphics System by Statistical Graphics Corporation (STSC Inc.) - Plus Ware Product, ver. 5.0

Zhang, H. C. and Huq, M. E., 1992, "Tolerancing Techniques: The State-of-the-Art," *Int. J. Prod. Res*., 30, (9), 2111~2135.

Appendix

A.1 Unit beta distribution function

The distribution function of the unit beta model is

$$f(x)=\begin{cases}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1} & 0<x<1,\\ \alpha>0,\ \beta>0 & \\ 0 & \end{cases} \tag{A1}$$

where α and β are the model parameters. The mean (μ) and variance (σ^2) of the model are calculated from

$$\mu=\frac{\alpha}{\alpha+\beta} \tag{A2}$$

$$\sigma^2=\frac{\alpha\cdot\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{A3}$$

The model parameter estimators (α and β) for data less than 21 and more than 21 are given in the works of Cooke (1979) and Bjorke (1978). The model parameters α and β control the skewness and shape of the distribution. For different α and β values, the shapes of the unit beta functions are given in Fig. 1.