International Journal of Statistics and Applied Mathematics 2018; 3(2): 351-354

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2018; 3(2): 351-354

www.mathsjournal.com
Received: 25-01-2018
Accepted: 28-02-2018

B Srinivasarao
Associated Professer of Statistics, Department of Mathematics & Humanities, R.V.R & J.C College of Engineering, Guntur, Andhra Pradesh, India

P Sricharani
Assistant Professor, Department of Basic Science, Shri Vishnu Engineering College for Women, Bhīmavaram, Andhra Pradesh, India

Correspondence
P Sricharani
Assistant Professor, Department of Basic Science, Shri Vishnu Engineering College for Women, Bhīmavaram, Andhrapradesh, India

Extreme value charts and analysis of means based on Dagum distribution

B Srinivasarao and P Sricharani

Abstract
The probability model of a quality characteristic is assumed as a Dagum distribution. Variable control charts based on the extreme values of each subgroup are constructed. The control chart constants depend on the probability model of the extreme order statistics of each subgroup and the size of the subgroup. Accordingly, the proposed chart is known as extreme value chart. The technique of analysis of means for a skewed population is applied with respect to Dagum distribution. The results are illustrated by examples on live data.

Keywords: ANOM, Incontrol, Equitailed, Q-Q plot.

Introduction
The probability density function (pdf) of Dagum distribution (DD) is given by

$$f(x,b,a,p)=\frac{ap}{x}\left(\frac{\left(\frac{x}{b}\right)^{ap}}{\left(\left(\frac{x}{b}\right)^{a}+1\right)^{p+1}}\right), \qquad x>0, a>0, b>0, p>0. \tag{1.1}$$

Its cumulative distribution function (cdf) is

$$F(x,b,a,p)=\left(1+\left(\frac{x}{b}\right)^{-a}\right)^{-p}, x>0, a>0, b>0, p>0. \tag{1.2}$$

Dagum distribution (DD) is a skewed, uni model distribution on the positive real line. The mean, median and variance of Dagum distribution are respectively.

$$Mean=-\frac{b}{a}\frac{\Gamma\left(-\frac{1}{a}\right)\Gamma\left(\frac{1}{a}+p\right)}{\Gamma p} \quad if \ a>1 \tag{1.3}$$

$$Median=b(-1+2^{\frac{1}{p}})^{-\frac{1}{a}} \tag{1.4}$$

$$\text{variance} = \left\{-\frac{b^2}{a^2}\left(2a\frac{\Gamma\left(-\frac{2}{a}\right)\Gamma\left(\frac{2}{a}+p\right)}{\Gamma p}+\left(\frac{\Gamma\left(-\frac{1}{a}\right)\Gamma\left(\frac{1}{a}+p\right)}{\Gamma p}\right)^2\right) \quad if \ a>2\right. \tag{1.5}$$

The Dagum Distribution is positively skewed and mesokurtic. In order to construct à control chart using the extreme observations of a subgroup drawn from the production process with the quality variate following Dagum Distribution. we need the percentiles of extreme order statistics in samples from Dagum Distribution. Specifically, the test statistic on extreme value control chart is the original sample vector X=(x_1; x_2; x_n) from the ongoing production. In this chart all the individual sample observations are plotted into control chart without calculating any statistic out of them. A corrective action is taken after one or either of the extreme values namely x_1 (sample minimum) and x_n (sample maximum) of the sample respectively fall below or above two specified lines (limits). Because of this, the chart is called extreme value control chart. The Shewart control chart is a common tool of statistical quality control for many practitioners. When these charts indicate the presence of an assignable cause, an adjustment of the process is made in the remedy is known. Otherwise the suspected presence of assignable cause is regarded to be an indication of heterogeneity of the subgroup statistic for which the control chart is developed. For instance if the statistic is sample mean, this leads to heterogeneity of process mean indicating departures from target mean. Such an analysis is generally carried out with the help of means to divide a collection of a given number of subgroup means into categories such that means within a category are homogenous and those between categories are heterogeneous and the procedure is called Analysis of Means (ANOM) given by Ott (1967) [7]. For using the ANOM technique the concept of the control chart for means is viewed in a different way- grouping of plotted means to fall within the control limits or some outside the control limits. For the homogeneity of all the means, it is necessary that all the means should fall within the control limits. If (1-α) is taken as the confidence coefficient, we should have the probability of all the subgroup means to fall within the control limits is (1-α). Assuming independence of subgroups the above probability statement becomes n^{th} power of the probability of a subgroup mean to fall within the limits. i. e In the sampling distribution of $\bar{x}$ the confidence interval for $\bar{x}$ to lie between two specified limits should be equal to $(1\text{-}\alpha)^{1/n}$. The same principle is adopted in the rest of this paper through Dagum distribution. Because this paper aims at exploring ANOM using control limits of extreme value statistics we have considered only the control chart aspects. However, a detailed literature about ANOM is available in Rao (2005) [9] and some related works in this direction are Ramig (1983) [8], Bakir (1994) [1], Bernard and Wludyka (2001) [2], Montgomery (2001) [5], Nelson and Dudewicz (2002) [6], Rao and Prankumar (2002) [10], Farnum (2004) [3], Guirguis and Tobias (2004) [4]. The recently developed ANOM tables or techniques based on various distributions are Srinivasa Rao and Kantam (2012a) [11], Srinivasa Rao *et al*. (2012b) [12], Srinivasa Rao *et al*. (2012c) [13] and references there in. The rest of the paper is organized as follows. The basic exposure to extreme value control charts is given in Section 2. ANOM applied to Dagum distribution using extreme value control charts of DD is given in Section 3 followed by numerical examples. Summary and conclusions are given in Section 4.

2. Extreme Value Charts

The given sample observations are assumed to follow Dagum distribution. The control lines are determined by the theory of extreme order statistics based on Dagum distribution model The control lines are to be determined in such a way that an arbitrarily chosen x_i of X=(x_1; x_2; x_n)lies with probability $(1-\alpha)^{\frac{1}{n}}$ within the limits. This can be formulated as a probability inequality in the following way $P(x_1 \le L) = \frac{\alpha}{2}$ and $P(x_n \ge U) = \frac{\alpha}{2}$. The theory of order statistics say that the cumulative distribution function of the least and highest order statistics in a sample of size n from any continuous population are $[F(x)]^n$ and $1-[1-F(x)]^n$ respectively. where F(x) is the cumulative distribution function of the population. If $1-\alpha$ is desired at 0.9973 then would be 0.0027. Taking F(x) as the CDF of dagum modal we can get solutions of the two equations $1-[1-F(x)]^n$ = 0:00135 and $[F(x)]^n$ = 0:99865 which in turn can be used to develop the control limits of extreme value chart. The solutions of the above two equations for n=2 (1) 10 are given in Table 1 denoted $Z_{(1)0.00135}$ and $Z_{(n)0.99865}$.

Table 1: Control limits of extreme value charts

N	$Z_{(1)\ 0.00135}$	$Z_{(n)\ 0.99865}$
2	1.37E-06	57.67655
3	6.08E-07	70.65498
4	3.42E-07	81.59453
5	2.19E-07	99.94373
6	1.52E-07	91.23162
7	1.12E-07	107.955
8	8.55E-08	115.4116
9	6.76E-08	122.4148
10	5.47E-08	129.0384

The values of Table. 1 indicate the following probability statement

P ($Z_{(1)0:00135}$ < Z_i < $Z_{(n)\ 0:99865}$; $\forall_I = 1,2..n$) = 0:9973 (2.1)

P (σ $Z_{(1)0:00135}$ < x_i < σ $Z_{(n)\ 0:99865}$; $\forall_I = 1,2..n$) = 0:9973 (2.2)

Taking mean of the Dagum distribution as an unbiased estimate of σ, the above equation becomes

$$P\left(D_3^*\bar{x} < x_i < D_4^*\bar{x}, \forall_{i=1,2,\ldots n}\right) = 0.9973 \quad (2.3)$$

$$where \ \ D_3^* = \frac{Z_{(1)(0.00135)}}{mean} \ and D_4^* = \frac{Z_{(n)(0.99865)}}{mean}$$

Thus D_3^*, D_4^* would constitute the control charts contain for the etreme valu charts These are given in Table 2. for n=2(1)10.

Table 2: Constants of Extreme value charts

N	D_3^*	D_4^*
2	7.609E-07	32.08687
3	3.382E-07	39.30708
4	1.902E-07	45.39302
5	1.218E-07	50.75436
6	8.458E-08	55.60112
7	6.214E-08	60.05801
8	4.758E-08	64.20626
9	3.759E-08	68.1023
10	3.045E-08	71.78721

3. Analysis of Means(ANOM)- Dagum Distribution

When the data variate follows Dagum distribution, suppose $\bar{x}_1, \bar{x}_2, \ldots \bar{x}_k$ are arithmetic means of k subgroups of size 'n' each drawn from a Dagum model. If these subgroup means are used to develop control charts to assess whether the population from which these subgroups are drawn is operating with admissible quality variations. Depending on the basic population model, we may use the control chart constants developed by us or the popular Shewart constants given in any SQC text book. Generally we say that the process is in control if all the subgroup means fall within the control limits. Otherwise we say the process lacks control. If α is the level of signifi cance of the above decisions we can have the following probability statements.

$$P\left(LCL < \bar{x}_i, \forall_{i=1,2,\ldots k} < UCL\right) = 1 - \alpha \qquad (3.1)$$

Using the notation of independent subgroups (3.1) becomes

$$P\left(LCL < \bar{x}_i < UCL\right) = (1-\alpha)^{\frac{1}{K}} \qquad (3.2)$$

with equal-tailed probability for each subgroup mean, we can find two constants say L^* and U^* such that

$$P(\overline{x_i} < L^*) = P\left(\overline{x_i} > U^*\right) = \frac{1-(1-\alpha)^{\frac{1}{K}}}{2} \qquad (3.3)$$

In the case of normal population L^* and U^* satisfy $U^* = -L^*$. For the skewed populations like Dagum Distribution we have to calculate L^*, U^* separately from the sampling distribution of $\overline{x_i}$. Accordingly these depend on the subgroup size 'n' and the number of subgroups 'k'. The percentiles of sampling distribution of $\overline{x}$ in samples from Dagum distribution are worked out through Monte-Carlo simulation and are given in the Table 3.

Table 3: Percentiles of sample Mean in Dagum Distribution

p	**0.99865**	**0.995**	**0.99**	**0.975**	**0.95**	**0.05**	**0.025**	**0.01**	**0.005**	**0.00135**
n=2	11.468	5.1834	3.9115	2.5239	1.8169	0.0321	0.0147	0.0061	0.0035	0.0009
3	9.1636	4.7826	3.5575	2.3065	1.6691	0.0714	0.0451	0.0263	0.0162	0.0085
4	7.7031	3.9709	2.9331	2.0182	1.5083	0.1009	0.0706	0.0469	0.0292	0.0155
5	5.8986	3.6035	2.7199	1.8759	1.4330	0.1266	0.0941	0.0676	0.0513	0.0338
6	5.4368	3.4367	2.4518	1.7638	1.4001	0.1513	0.1168	0.0887	0.0653	0.0320
7	5.1056	3.2389	2.5704	1.7609	1.3269	0.1687	0.1317	0.1005	0.0798	0.0542
8	5.3257	3.0336	2.2657	1.6277	1.2457	0.1875	0.1487	0.1212	0.1040	0.0761
9	4.8355	2.6043	2.1420	1.5916	1.2802	0.1992	0.1651	0.1307	0.1091	0.0811
10	4.2427	2.5570	2.0747	1.5372	1.2430	0.2133	0.1787	0.1366	0.1185	0.0982

We make use of its percentiles in Equation (3.3) for specified 'n' and 'k' to get L^* and U^* for α = 0:01 and 0.05. These are given in Tables 8 and Table 9. A control chart for averages giving 'In Control' conclusion indicates that all the subgroup means though vary among themselves are homogenous in some sense. This is exactly the null hypothesis in an analysis of variance technique. Hence the constants of Tables 8 and 9 can be used as an alternative to analysis of variance technique. For a normal population one can use the tables of Ott (1967) [7]. For a Dagum Distribution our tables can be used. We therefore present below some examples for which the goodness of fit Dagum model is assessed with Q-Q plot technique (strength of linearity between observed and theoretical quantiles of a model) and tested the homogeneity of means involved in each case

Example 1: Wadsworth (1986) [14]: Consider the following data of 25 observations on manufacture of metal products that suspect variations in iron content of raw material supplied by five suppliers. Five ingots were randomly selected from each of the five suppliers. The following table contains the data for the iron determinations on each ingots in percent by weight

Table 4: Supplier

1	**2**	**3**	**4**	**5**
3.46	3.59	3.51	3.38	3.29
3.48	3.46	3.64	3.40	3.46
3.56	3.42	3.46	3.37	3.37
3.39	3.49	3.52	3.46	3.32
3.40	3.50	3.49	3.39	3.38

Example 2: Three bands of batteries are under study. It is suspected that the life (in weeks) of the three brands is different. Five batteries of each brand are tested with the following results. Test whether the lives of these brand of batteries are different at 5% level of significance.

Table 5: Weeks of Life

Brand1	**Brand2**	**Brand3**
100	76	108
96	80	100
92	75	96
96	84	98
92	82	100

Example 3: Four catalysts that may affect the concentration of one component in a three- component liquid mixture are being investigated. The following concentrations are obtained. Test whether the four catalysts have the same affect on the concentration at 5% level of significance.

Table 6: Catalyst

1	**2**	**3**	**4**
58.2	56.3	50.1	52.9
57.2	54.5	54.2	49.9
58.4	57	55.4	50
55.8	55.3	54.9	51.7

Graphs indicating QQ-Plot for the data in the illustrated examples

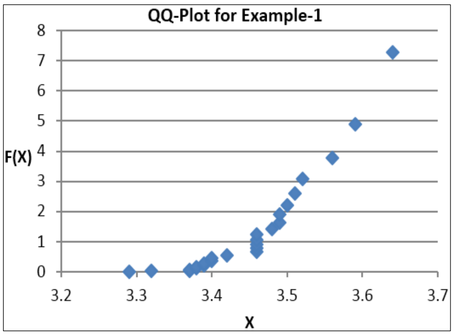

Fig 1

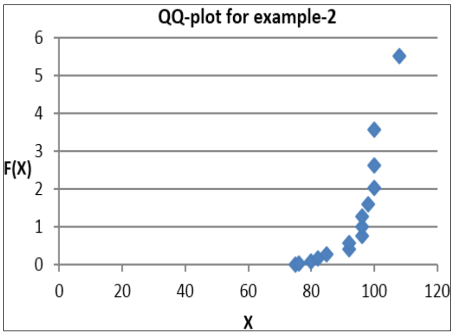

Fig 2

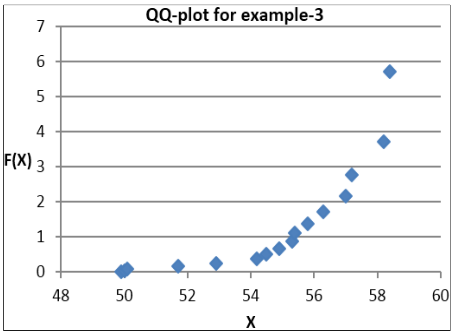

Fig 3

The goodness of fit of data in these three examples as revealed by Q-Q plot (correlation coefficient) are summarized in the following table, which shows that Dagum distribution is a better model, exhibiting significant linear relation between sample and population quantiles.

Table 7

	DD	Normal
Example1	0.7347	0.2067
Example2	0.8048	0.4149
Example3	0.7741	0.4447

Treating these observations in the data as a single sample, we have calculated the decision limits for the Normal population, DD population and have given them in the Tables 4 and 5 respectively.

Table 8: Normal Distribution

		No. of Counts			
	[LDL, UDL](Ott 1967)	In	P=in/k	Out	Out/k
Example 1: n=5, k=5, α = 0.05	[3.517, 3.879]	1	0.2	4	0.8
Example 2: n=5, k=3, α = 0.05	[87.82, 95.52]	2	0.7	1	0.3
Example 3: n=4, k=4, α = 0.05	[26.14, 82.84]	0	0	4	1

Table 9: Dagum Distribution

		No. of Counts			
	[LDL, UDL](Table-7)	In	P=in/k	Out	Out/k
Example 1: n=5, k=5, α = 0.05	[1.06517, 7.4039]	5	1	0	0
Example 2: n=5, k=3, α = 0.05	[29.3617, 185.2045]	3	1	0	0
Example 3: n=4, k=4, α = 0.05	[15.8977, 124.8574]	4	1	0	0

4. Summary and Conclusion

The usage of normal model resulted in homogeneity for some means and deviation for some other means, indicating a possible rejection of those means. This decision is valid if Normal distribution is a good fit to the data. As a comparision, we have alredy established by Q-Q plot that Dagum distribution is a better model than Normal as supported by the Q-Q plot correlation coefficient of each data set with Normal as well as Dagum distribution separately. We have observed that more error is likely to be associated with decision process of Normal distribution. Therefore, more number of the means to be homogenous with the help of Dagum distribution(Table-5) is a better decision than some means to be away from homogeneity using Normal, ANOM procedure.

5. References

1. Bakir ST. Means using ranks for randomised complete block design Communications in Statistics-Simulation and Computation. 1994; 23:547-568.
2. Bernard AJ, Wludyka PS. Robust I-sample Analysis of means type randomization tests for variances Journal of Statistical Computation and Simulation. 2001; 69:57-88.
3. Farnum NR. Analysis of Means Tables using mathematical processors Quality Engineering. 2004; 16:399-405.
4. Guirguis GH, Tobias RD. On the computation of the distribution for the analysis of means Communications in Statistics- Simulation and Computation. 2004; 16:861-887.
5. Montgomery DC. Design and Analysis of Experiments Fifth edition, John Wiley and Sons, New York, 2004.
6. Nelson PR, Dudewicz EJ. Exact Analysis of Means with Unequal Variances Technometrics. 2002; 44:152-160.
7. Ott ER. Analysis of Means-A graphical procedure Industrial Quality Control. 1967; 24:101-109.
8. Ramig PF. Applications of Analysis of Means Journal of Quality Technology. 1983; 15:19-25.
9. Rao CV. Analysis of Means-A review Journal of Quality Technology. 2005; 37:308-315.
10. Rao CV, Prankumar M. ANOM-type Graphical Methods for testing the Equality of Several Correlation Coeffcients Gujarat Statistical Review. 2002; 29:47-56.
11. Srinivasa Rao B, Pratapa Reddy J, Sarath Babu G. Extreme value charts and analysis of means (ANOM) based on log logistic distribution Journal of Modern Applied Statistical Methods. 2012a; 11(2):493-505.
12. Srinivasa Rao B, Kantam RRL. Extreme value charts and analysis of means based on half logistic distribution International Journal of Quality, Reliability & Management. 2012b; 29(5):501-511.
13. Srinivasa Rao B, Pratapa Reddy J, Rosaiah K. Extreme value charts and analysis of means based on inverse rayleigh distribution Pakistan Journal of Statistics and Operations Research. 2012c; 8(4):759-766.
14. Wadsworth HM, Stephens KS, Godfrey AF. Modern methods of quality control and important John Wiley and Sons. Inc. New York.