MODE BEHAVIOUR IN RELATION TO BIN SIZE AND DATA DISTRIBUTION

***OPALEYE, Adepeju A*[1] *and Charles-Owaba, O.E*[2]**

Department of Industrial &Production Engineering
University of Ibadan,
Ibadan, Nigeria

[2] Department of Industrial &Production Engineering,
University of Ibadan,
Ibadan, Nigeria

ABSTRACT: *The "mode" has been proposed as an appropriate statistic to improve estimate especially in situations when data distributions are skewed or contain outliers such as activity duration in project scheduling. Since the underlying distribution of activity duration may be unknown and different modes can be obtained using different bin sizes of the histogram method, this paper, investigates the effect of varying histogram bin width and data distribution on the behaviour of the mode. Random numbers were generated from five distributions commonly used to model project activity duration at five different levels and varying sample sizes. Each set of sample is then binned using varying histogram bin width, Sturges'rule and Scott's rule. The grand mode for all levels per classification is recorded and analyzed. It was found that bin width does not significantly affect the behaviour of the mode but the value of the mode is significantly dependent on the data distribution and sample size.*

KEYWORDS: *S*tatistical mode, Bin size, Statistical distribution, Uncertainty, Estimate

INTRODUCTION

Project scheduling has been described as one of the critical areas of project management. Morgenshtern and Dvir [1] empirically found that greater uncertainty in project scheduling is associated with increasingly erroneous project duration estimation. This may be due to inability to accurately estimate project activity duration. One of the earliest methods of quantifying uncertainty in project activity duration by Malcolm *et al* [2] emphasized the importance of knowing the 'most likely' duration experienced during stable work conditions. Statistically, actual activity duration values experienced at instances of this type are those with the highest likelihood of occurrence. Subsequent works by Ginzburg [3], Ravi Shankar *et al [4]*, Marounek [5] emphasized likewise. Hall and Johnson [6] in their study, sought and applied field data on "most of the times", "pessimistic" durations and the likelihood of overrun per critical activity in their effort to estimate expected activity duration. Various distributions which have also been used to model project activity duration [2, 7-12] are continuous, positively skewed and unimodal. For data distribution of such characteristics, Hedges and Shah [13] considered the 'mode' an appropriate statistic to represent their center of location.This shows the importance of the mode in project scheduling and other fields such as the medical domain, biology, astronomy [14-16] where it has been used to improve accuracy of estimates.

Despite the importance of the mode and its recognition as a natural measure of central tendency, there are few methods of estimating the mode [13, 17-21]. Hedges and Shah carried out a comparative study on the available mode estimating techniques, they pointed out that different mode estimators perform better in different conditions. Thus, in general circumstance in which the underlying data distribution is unknown, one should employ a nonparametric method for the inference of the mode [22]. Among them, the histogram seems to be the most basic, easily applicable non parametric method which does not involve too much computational overhead for mode estimation. Although, the method has been widely applied, its major shortcoming is that different modes can be obtained using different bin width.

The first and widely criticized histogram bin width selection rule by Sturges [23] led to the work of Scott [24].He derived formulae for the optimal bin width by minimising the integrated mean square error (IMSE) of the histogram model of the true density, based on the assumption that data is normally distributed. The bin width which minimizes the IMSE is chosen as the optimal bin width. Legg *et al*, [14] however, argued that in the context of image registration, Sturges' Rule consistently outperformed Scott's Rule and since the underlying distribution of data may be unknown, it may be inappropriate to use an optimization criterion that relies on the error between the density model and the true density [25]. Instead, Knuth[25] considered the histogram to be a piecewise-constant model of the underlying probability density. Wand[26] extended Scot's formula in order to give better consistency properties but it requires too much computational effort, therefore 'defeating' the simplicity of the histogram technique. Other techniques for the selection of histogram bin based on similar assumption are presented in [27-29]. These reveal that there is no general consensus on the determination of optimal bin width and in the context of project scheduling where distribution of project activity duration varies depending on project circumstance, there is a need to investigate the behavior of the mode with respect to varying histogram bin width and varying distribution of project activity duration.

METHODOLOGY

To pursue this goal, one may adopt the following approach. Distributions commonly used to estimate project activity durations may be selected and a set of respective parameter values representing very low, low, medium, moderately large and very high values of activity duration specified. Based on these, samples of durations may be generated randomly from each distribution using an authenticated random number generator. Each sample is then binned into 'n' classes from which the mode per bin width is derived. Considering all samples, the modes are statistically examined for any significance differences due to number of bins and data distribution. Indeed, this is the approach adopted in this study.

Generating Random Variates

Accordingly, the uniform, normal, lognormal, triangular and beta distributions which have been used to estimate project activities duration will be used to generate the data for this study. Detail information on the distributions follows:

The uniform distribution has two parameters (minimum value, a; and maximum, b) with the following as the probability density function (PDF):

$$f(x) = \begin{cases} \frac{1}{b-a} & , \ a \le x \le b; \\ 0 \ , \ elsewhere \end{cases}$$

and the cumulative,

$$F(x) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a} & , \ a \le x \le b \\ 0, b \le x \end{cases}$$

while the associated activity duration (x) to be generated randomly is given by the expression:

$$x = a + (b - a)\, R_i; \qquad [30] \qquad (1)$$

where $R_i \sim unif(0,1)$.

Similarly, for the normal distribution, the pdf is:

$$f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \ ;$$

where 'μ' is the location parameter equal to the mean and 'σ' is the standard deviation and $-\infty<x<\infty$, $-\infty<\mu<\infty$, $\sigma>0$.

The cumulative function is given as

$$F(x) = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{dx} \qquad [31]$$

Because its inverse does not exist, R ~ normal(0,1), standard normal distribution may be generated using -Muller method or Polar-Marsaglia method and random activity duration, x generated using

$$x = \mu + \sigma R \qquad [32] \qquad (2)$$

If the distribution of activity duration is assumed to be lognormal distributed, the pdf is given as

$$y = f(x|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}}\, e^{\frac{-(lnx-\mu)^2}{2\sigma^2}}$$

The cumulative function is

$$F(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^x \frac{e^{\frac{-(\ln(t)-\mu)^2}{(2\sigma)^2}}}{t}\, dt$$

where the variable x > 0 and the parameters μ and σ > 0 all are real numbers. A lognormal distribution with mean (m) and variance (v) has parameters

$\mu = \log(m^2 / \sqrt{v + m^2}$) and

$\sigma = \log\sqrt{(v/m^2 + 1)}$ (3)

If x is lognormally distributed, then log(x) is normally distributed[33].Note that (μ, σ^2) are not the arithmetic mean and standard deviation of the lognormal distribution.

The asymmetric triangular distribution may also be used to model activity duration when an expert is able to give an upper (b) and lower bound (a) on the possible activity duration with an inclusion of an estimate of the most likely duration (m). The probability density function is given as

$$f(x|a,m,b) = \begin{cases} \frac{2(x-a)}{(b-a)(m-a)}, for\ a \le x \le b \\ \frac{2(b-x)}{(b-a)(b-m)}, for\ m \le x \le b \\ 0,\ elsewehere \end{cases}$$

where $a < b$ and $a \le m \le b$.

The cumulative function is given as

$$F(x) = \begin{cases} \left(\frac{m-a}{b-a}\right)\left(\frac{x-a}{m-a}\right)^2,\ for\ a \le x \le m \\ 1 - \left(\frac{b-m}{b-a}\right)\left(\frac{b-x}{b-m}\right)^2, for\ m \le x \le b \end{cases}$$

and the associated activity duration can be generated using the following expressions;

$$F^{-1}(y|a,m,b,n) = \begin{cases} a + \sqrt{y(m-a)(b-a)},\ for\ 0 \le y \le \frac{m-a}{b-a} \\ b - \sqrt{(1-y)(b-m)(b-a)},\ for\ \frac{m-a}{b-a} \le y \le 1 \end{cases}$$

where y ~ unif(0,1) [34] (4)

Equation (4) allows for straightforward sampling from a triangular distribution with given [35].

For activity duration that is beta distributed, the continuous probability distribution is defined on the interval (a, b) parameterized by two positive shape parameters, typically denoted by α and β. The pdf is given as

$$f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{((x-a)^{\alpha-1})(b-x)^{\beta-1}}{(b-a)^{\alpha+\beta-1}}$$

where 'a' and 'b' are the minimum and maximum values of activity duration [35].The beta cumulative density function is given as

$$F(x|a,b) = \frac{1}{B(a,b)} \int_0^x t^{a-1}(1-t)^{(b-1)}\mathrm{dt}$$

x ~ beta(α,β). If ' y_m ' and ' y_n ' are two independent variables distributed according to the chi-squared distribution with m and n degrees of freedom, respectively, then the ratio

$$\frac{y_m}{y_n + y_m}$$ [30] (5)

is beta-distributed, See also:[36].

Using these distributions, two different project planning situations are examined in this study.

Case I: A project situation in which observed data is actually uniformly distributed (could have been any other distribution) but project managers erroneously assumed different types. In this case, the assumed may be normal, lognormal, beta and triangular distributions. Indeed, this has been the case by many practitioners as noted by Trietsch *et al* [7]. It will be useful information to know the impact this will have on mode behaviour.

Case II: A project planning situation where the nature of the observed distribution is known. The data may be of uniform, normal, lognormal, triangular or beta distribution etc. representing independent and different project activity duration. The data in Table 1 are the parameter values for simulating both cases considering projects of very low durations, low, medium, high and very high values of activity duration.

Case	I	Range of parameter values	Uniform		Normal		Lognormal		Triangular			Beta			
	I		A	B	μ	Σ	μ*	σ*	a	M	B	α	β	a	B
Case I	1	very low	1	30	15.1	8.5	2.58	0.52	1	10	30	2.77	4.65	1	30
	2	Low	31	60	45.3	8.3	3.80	0.18	31	42	60	3.26	4.51	31	60
	3	medium	61	90	76	8.6	4.32	0.11	61	76	90	4.09	3.91	61	90
	4	High	91	120	105	8.9	4.65	0.08	91	100	120	2.77	4.65	91	120
	5	very high	121	150	134	8.7	4.90	0.07	121	146	150	4.38	1.53	121	150
Case2	1	very low	12	30	16	6	2.11	0.42	10	16	40	**4.7**	2.39	36	67
	2	Low	46	87	89	9	4.02	0.12	59	70	104	**1.6**	4.45	82	140
	3	medium	105	232	104	7	5.11	0.07	126	176	343	**4.6**	1.88	167	215
	4	High	289	445	156	15	5.29	0.08	400	478	506	**3.7**	4.25	295	335
	5	very high	505	969	456	23	6.11	0.07	633	700	888	**4.4**	3.51	480	540

Table 1: Sets of parameter values

Data Generation and Classification

To summarize our approach, the necessary steps are given as follows.

Step 1: Specify probability distribution, $\tilde{F}$.

Step 2: Set i =0 where 'i' represent range of parameter values; very low, low, medium, high and very high (As given on table 1.0).

Step 3: Generate 'N' random sample from $\tilde{F}$ and 'i' to form one data set, where N =50,100,150,200,250……….1000.

Step 4: Classify the data set into 'n' bins where n=5, 10, 15 ,20, 25.Also, compute mode using Sturges and Scot's rule.

Step 5: For range 'i' obtain the mode for bin 'n'($m_{i,n}$).

Step 6: Repeat steps 2-4 for all 'i'

Step 7: Calculate the grand mode '$\bar{M}_n$' for bin ' n' ; that is $\bar{M}_n = \frac{\sum_{i=1}^{5} m_{i,n}}{5}$

Step 8: Repeat steps 2-7 for all N.

Step 9: Tabulate the result as shown in table 2

SAMPLE SIZE =N	I	Range of parameter values	5 bins	10 bins	15 bins	20 bins	25 bins	Sturges ruls	Scot's rule
	1	very low	$m_{1,5}$	$m_{1,10}$	$m_{1,15}$	$m_{1,20}$	$m_{1,25}$	$m_{1,St.}$	$m_{1,Sc.}$
	2	Low	$m_{2,5}$	$m_{2,10}$	$m_{2,15}$	$m_{2,20}$	$m_{2,25}$	$m_{2,St.}$	$m_{2,Sc.}$
	3	Medium	$m_{3,5}$	$m_{3,10}$	$m_{3,15}$	$m_{3,20}$	$m_{3,25}$	$m_{3,St.}$	$m_{3,Sc.}$
	4	High	$m_{4,5}$	$m_{4,10}$	$m_{4,15}$	$m_{4,20}$	$m_{4,25}$	$m_{4,st.}$	$m_{4,Sc.}$
	5	very high	$m_{5,5}$	$m_{5,10}$	$m_{5,15}$	$m_{5,20}$	$m_{5,25}$	$m_{5st.}$	$mm_{5Sc.}$
		Grand mode	$\bar{M}_{5.}$	$\bar{M}_{10}$	$\bar{M}_{15}$	$\bar{M}_{20.}$	$\bar{M}_{25.}$	$\bar{M}_{St.}$	$\bar{M}_{Sc.}$

Table 2 Mode values for specified distribution

Hypothesis Testing

Based on the above described situations, our main hypothesis is to evaluate the effect of the different bin width of histogram on the mode and to determine whether there is a significant difference in the modes obtained due to the distribution of data. In addition, we need to determine the effect of increasing sample size on the obtained mode also investigate if there is a significant difference between the mode obtained by Sturges rule and Scot's rule. In order to investigate these statistical variations, the analysis of variance test is used[37, 38].The test statistics is carried out at 0.05 significance level (α), a probability threshold below which the null hypothesis will be rejected. In specific term the hypothesis is stated as thus:

Hypothesis 1
H_{10}: *There is no significant difference in the behaviour of the mode with respect to varying bin width.*
H_{11}: *There is a significant difference in the behaviour of the mode with respect to varying bin width.*
Hypothesis 2
H_{20}: *There is no significant difference between the mode obtained using varying bin width and Sturges rule*
H_{21}: *There is a significant difference between the mode obtained using varying bin width and Sturges rule.*
Hypothesis 3
H_{30}: *There is no significant difference between the mode obtained using varying bin width and Scot rule.*
H_{31}: *There is a significant difference between the mode obtained using varying bin width and Scot rule.*
Hypothesis 4
H_{40}: *There is no significant difference between the mode obtained using Sturges and Scot's rule.*
H_{41}: *There is a significant difference between the mode obtained using Sturges and Scot's rule.*
Hypothesis 5
H_{50}: *There is no significant difference in the behaviour of the mode with respect to data distribution.*
H_{51}: *There is a significant difference in the behaviour of the mode with respect to data distribution*.
Hypothesis 6
H_{60}: *There is no significant difference in the mode across the four sample sizes.*
H_{61}: *There is a significant difference in the mode across the four sample sizes.*

PRESENTATION AND DISCUSSION OF RESULTS

In the previous section we presented the methodology for demonstrating the utility of the histogram method for determining the mode of several different data sets from five different distributions. A sample result of the grand mode for the case of uniformly distributed data at 50-1000 sample sizes is presented in table 3.

UNIFORM DISTRIBUTION CASE 1									UNIFORM DISTRIBUTION CASE 2						
	5Bins	**10Bins**	**15Bin**	**20Bins**	**25Bins**	**SturgesBin**	**ScotBin**		**5Bins**	**10Bins**	**15Bins**	**20Bins**	**25Bins**	**SturgesBin**	**ScotBin**
50	73	74	79	75	79	75	77		285	292	287	328	306	296	335
100	77	74	73	77	73	73	78		290	302	271	311	300	314	297
150	74	75	72	75	73	77	71		292	285	288	260	281	285	300
200	78	78	81	78	81	78	77		270	279	281	304	253	240	335
250	77	81	79	84	78	81	85		262	262	327	334	299	324	313
300	75	67	70	75	72	68	73		285	308	284	275	280	295	287
350	79	74	80	76	81	75	81		248	237	259	213	249	238	220
400	79	84	79	79	76	82	82		219	243	236	243	244	243	245
450	79	77	78	81	78	76	73		287	290	277	279	280	283	287
500	80	76	78	76	83	76	75		305	308	266	277	302	305	276
550	77	77	72	73	76	76	77		283	292	296	304	299	291	299
600	75	80	78	77	78	82	83		236	289	295	300	286	297	298
650	77	73	74	75	71	73	77		301	296	314	297	254	272	265
700	74	73	75	80	79	83	77		302	307	209	284	274	286	299
750	74	74	75	74	81	76	75		334	322	342	346	314	330	334
800	78	76	75	77	75	77	75		281	263	234	306	275	256	268
850	79	81	74	71	78	70	83		319	316	314	313	312	323	326
900	85	84	81	84	86	84	86		329	281	280	248	270	316	284
950	77	74	76	75	74	76	75		224	240	224	250	253	237	268
1000	78	77	73	80	80	82	78		254	260	308	304	288	278	294

Table 3 Mode for uniformly distributed data

The ANOVA summary result for the six hypotheses across all the distributions is shown in table 4.

Hypothesis	*Case*	*df1*	*df2*	*Fcritical*	*Fstatistics*	*Pvalue*
1	I	4	495	2.38995	0.176235	0.9506
	II	4	495	2.38995	0.02446	0.99884
2	I	5	594	2.22919	0.597461	0.70194
	II	5	594	2.22919	0.027746	0.99963
3	I	5	594	2.22919	1.231465	0.29273
	II	5	594	2.22919	0.074764	0.996
4	I	1	198	3.88885	0.329305	0.56672
	II	1	198	3.88885	0.062276	0.80319
5	I	4	695	2.38475	175.0199	1E-103
	II	4	695	2.38475	2419.925	0
6	I					
Uniform distribution		19	120	1.67388	7.906759	1.1E-13
Normal distribution		19	120	1.67388	6.282486	6.6E-11
Lognormal distribution		19	120	1.67388	1.849823	0.02447
Triangular distribution		19	120	1.67388	16.59659	3.7E-25
Beta distribution		19	120	1.67388	4.600535	9.1E-08
	II					
Uniform distribution		19	120	1.67388	10.2319	3E-17
Normal distribution		19	120	1.67388	3.866522	2.5E-06
Lognormal distribution		19	120	1.67388	4.502335	1.4E-07
Triangular distribution		19	120	1.67388	11.32565	8.9E-19
Beta distribution		19	120	1.67388	3.587229	9.2E-06

Table 4 ANOVA Summary Result

Based on the result in table 4.0, the $F_{statistic}$ and p-value at 4 and 495 degrees of freedom for the first hypothesis reveal that we do not reject the null hypothesis of no significant difference in the behaviour of the mode with respect to varying bin width for the first and second case situations. Related hypotheses 2,3 and 4 also reveal that we do not reject the null of these three hypotheses. This implies that varying histogram bin width do not significantly affect the mode of a set of data. This may be the reason for the use of rules of thumb for determining the number of bins, such as 5-20 bins usually considered as adequate (for example, Matlab uses 10 bins as a default). However, for hypothesis 5, at 4 and 695 degrees of freedom, $F_{statistic}$ and p-value reveal that the null hypothesis should be rejected. In other words, the distribution of data has a direct impact on its mode, this is expected. Finally, the sixth hypothesis reveals that the mode is significantly unequal across all the sample sizes. A plot of the modes against sample sizes reveals a significant the random behaviour of the mode for all the distributions.

CONCLUSION

We investigated the behaviour of mode with respect to histogram bin width of five distributions commonly used to model duration of project activities. The results show that an estimate of mode is not directly related to the bin width. For set of historical activity duration, the choice of histogram bins between 5 and 20 as suggested in the literature or any of the optimal bin width selection technique is adequate for mode estimation. However, mode estimate for a data set may be misleading when durations in the set are from different sources. In other words, durations of project activities are unique to the project organisation and durations of similar projects from different organisations may not be mixed to obtain a set. We therefore conclude that the histogram method is an appropriate technique for mode estimation in project scheduling with the availability of historical data and yields minimal computation overhead.

RECOMMENDATION FOR FUTURE STUDIES

This investigation on the behavior of the mode with respect to the histogram bin width and distribution may provide basic information for developing a framework for estimating activity duration and quantifying schedule risk in project management. This aspect which is beyond the scope of this paper is considered further studies.

References

1. Morgenshtern, O., T. Raz, and D. Dvir, *Empirical analysis in software process simulation modeling.* . Information and Software Technology ; in press, 2006.
2. Malcolm D. G., J.H. Roseboom, and C.E. Clark, *Application of a technique of research and development program evaluation.* Operations Research Letters, 1959 **7**p. 646-669.
3. Golenko-Ginzburg, D., *" On the Distribution of Activity Time in PERT"* Journal of Operational Research Society, 1988. **39**(8): p. 767-771.
4. Shankar, N.R. and V. Sireesha, *An Approximation for the Activity Duration Distribution, Supporting Original PERT.* Applied Mathematical Sciences, 2009. **3**(57): p. 2823-2834.
5. Marounek, P., *Simplified approach to effort estimation in software maintenance.* JOURNAL OF SYSTEMS INTEGRATION, 2012. **3**(1): p. 53-62.
6. Earl, H. and J. Johnson, *Risk Analysis, Integrated Project Management*, 2003, Pearson Education, Upper saddle, New Jersey.

.
7. Trietsch, D., et al., *Modeling activity times by the Parkinson distribution with a lognormal core: Theory and validation.* European Journal of Operational Research, 2012. **216**(2): p. 386-396.
8. Cottrell, W., *Simplified program evaluation and review technique (PERT).* . J Construct Eng Mgmt 1999. **125**(1): p. 16-22.
9. Hahn, E.D., *Mixture densities for project management activity times: A robust approach to PERT.* . European Journal of Operational Research 2008. **188** p. 450–459.
10. Jannat, S. and A.G. Greenwood. *Estimating Parametewrs Of The Triangular Distribution Using Nonstandard Information*. in *Proceedings Of The 2012 Winter Simulation Conference.*. 2012.
11. Jaskowski, P. and S. Biruk, *Estimating Distribution Parameters Of Schedule Activity Duration On The Basis Of Risk Related To Expected Project Conditions* International Journal Of Business And Management Studies, 2011. **Vol 3** (No 1): p. pp 299-307 Issn: 1309-8047 (Online).
12. McCombs, E.L., M.E. Elam, and D.B. Pratt, *Estimating Task Duration in PERT using the Weibull Probability Distribution.* Journal of Modern Applied Statistical Methods, 2009. **8**(1): p. 282-288.
13. Hedges, B.S. and P. Shah, *Comparison of mode estimation methods and application in molecular clock analysis.* . BMC Bioinformatics, 2003. **4**(31): p. 1471-2105.
14. Legg, P.A., P. L. Rosin, D. Marshall, and J. E. Morgan *" Improving accuracy and efficiency of registration by mutual information using Sturges' histogram rule"* In MIUA,, 2007 p. pp26-30. .
15. Heckman, D., et al., *Molecular evidence for the early colonization of land by fungi and plants* Science 2001. **293**: p.1129-1133.
16. Markov, H., et al., *An algorithm to "clean" close stellar companions.* Astronomy and Astrophysics Supplement Series 1997. **122**: p. 193-199.
17. Mantazeri, H., *Non-parametric density and mode estimation for bounded data* 2011, ETH Zurich,Swiss Federal Institute of Technology,Zurich.
18. Agarwal, B., *Basic Statistics* 1991, New Dehli: Wiley Eastern Limited.
19. Bickel, D., *Robust and efficient estimation of the mode of continuous data: The mode as a viable measure of central tendency Journal of Statistical Computation and Simulation (in press)*
preprint: http://interstat.stat.vt.edu/interstat/articles/2001/abstracts/n01001.html-ssi 2001: p. 1-22.
20. Bickel, D., *Robust estimators of the mode and skewness of continuous data* Computational Statistics and Data Analysis., 2002. **39**: p. 153-163.
21. Grenander, U., *Some direct estimates of the mode.* Annals of Mathematical Statistics 1965: p. 131-13.
22. Koyama, S. and S. Shinomoto, *Histogram Bin Width Selection For Time-Dependent Poisson Processes,* . Institute Of Physics Publishing Journal Of Physics A: Mathematical And General J. Phys. A: Math. Gen. , 2004. **37**: p. 7255–7265.
23. Sturges, H., *The Choice Of A Class-Interval.* Journal Of Amer. Statist. Assoc., 1926. **21**: p. 65–66. .

24. Scott, D.W., *On optimal and data-based histograms.* . Biometrika., 1979. **66**: p. 605-610.
25. Knuth, K.H., *"Optimal data-based binning for histograms,"* ArXiv Physics e-prints., 2006
26. Wand, M.P., *Statistical Computing and Graphics: Data-based choice of histogram bin-width.* American Statistical Association, 1997 **Vol 51**(No1): p. pp 59-64.
27. Rudemo, M., *Empirical choice of histograms and kernel density estimators.* Scand. J. Statist., 1982 **9**: p. 65–78.
28. Meeden(, K.H.a.G., *Selecting the number of bins in a histogram: A decision theoretic approach.* Journal of Statistical Planning and Inference 1997 **Vol 61**: p. 59-59.
29. Stone, C.J. *An asymptotically histogram selection rule.* . in *Proc. Second Berkeley Symp (ed. J. Neyman)* 1984 Berkeley: Univ. California Press.
30. Walck, C., *Hand-book on STATISTICAL DISTRIBUTIONS for experimentalists* I.R.S.P. University of Stockholm, Editor 2007.
31. Polyanin, A.D. and A.V. Manzhirov, eds. *Handbook of Integral Equations* ed. 2nd2008, Chapman and Hall/CRC,Taylor & Francis Group 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL 33487-2742.
32. Frey, C.H. and D.S. Rhodes, *Quantitative Analysis of Variability and Uncertainty in Environmental Data and Models , Volume 1. Theory and Methodology Based Upon Bootstrap Simulation*, in *Work Performed Under Grant No.: DE-FG05-95ER30250 Report No.: DOE/ER/30250--Vol. 1,Water Resources and Environmental Engineering Program Department of Civil Engineering*, North Carolina State University Raleigh, NC 27695-7908.
33. Evans, M., Hsatings,N. and Peacock B., *Statistical Distributions* N.W.-I. Hoboken, Editor 2000 p. pp102-105.
34. Kotz, S. and J.R. Van Dorp, *Beyound beta: Other continous families of distributions with bounded support and application* World Scientific Singapore, 2004
35. Biruk, P. and P. Jaskowski, *ON THE PROBLEMS OF MODELLING AND RELIABILITY ASSESSMENT OF CONSTRUCTION PROJECTS* RT&A, 2010. **Vol 04** (19): p. .pp 6-14.
36. Davis, R., *Teaching Note Teaching Project Simulation in Excel Using PERT-Beta Distributions.* Inform, 2008. **Vol. 8**(3): p. 139–148.
37. Hossen Jamal, E.H., Siyam Quddus Khan and Subrata Kumar Saha, *Analysis of cotton yarn count variation by two way Anova.* Annals Of The University Of Oradea Fascicle Of Textiles, LeatherWork,, 2012 **Vol 13**(No 2): p. pp 83-85.
38. Haque Emdadul, M.F.R., Jamal Hossen and Md. Ruhul Amin, *Effect of stitch length and yarn count on grey and finished width of 1X1 Flat knit fabric.* Annals Of The University Of Oradea Fascicle Of Textiles, LeatherWork, 2012. **No 13**(2): p. 71-76.