Extreme Value Charts and Anom Based on Inverse Rayleigh Distribution

B. Srinivasa Rao
R.V.R & J.C College of Engineering
Chowdavaram Guntur-522 019
Andhra Pradesh, India
boyapatisrinu@yahoo.com

J. Pratapa Reddy
St.Ann's College for Women
Gorantla, Guntur-522035
Andhra Pradesh, India
jpstatistics@yahoo.com

K. Rosaiah
Acharya Nagarjuna University
Guntur-522 010
Andhra Pradesh, India
Rosaiah1959@gmail.com

Abstract

A measurable quality characteristic is assumed to follow Inverse Rayleigh distribution. Variable control charts based on the extreme values of each subgroup are constructed. The technique of analysis of means (ANOM) is adopted to work out the decision lines of Inverse Rayleigh distribution. The preferability of the proposed ANOM decision lines over that of Ott (1967) [11] is illustrated by some examples.

Keywords: Extreme Value charts, In Control, Q-Q plot.

Acronyms

IRD: Inverse Rayleigh Distribution
f(x) : probability density function (pdf)
F(x): cumulative distribution function (cdf)
σ : scale parameter
ANOM: analysis of means

1. Introduction

The probability density function (pdf) of a Inverse Rayleigh Distribution (IRD) with scale parameter σ is given by

$$f(x) = \frac{2\sigma^2}{x^3} e^{(-\sigma^2 / x^2)}; \quad x > 0, \ \sigma > 0 \tag{1.0.1}$$

Its cumulative distribution function (cdf) is

$$F(x) = \mathrm{e}^{-\sigma^2 / x^2}; \quad x > 0, \ \sigma > 0 \tag{1.0.2}$$

In order to construct a control chart using the extreme observations of a subgroup drawn from the production process with the quality variate following IRD we need the percentiles of extreme order statistics from IRD. 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 a sample depending solely on the extreme values namely x_1 (sample minimum) and x_n (sample maximum) of the sample. Because of this, the chart is called extreme value control chart.

Analysis of Means (ANOM) is a technique originally developed by Ott (1967) [11] for comparing a group of treatment means to see if any one of them differs significantly from the overall mean. This procedure is carried out by comparing the sample mean values to the overall grand mean, about which decision lines have been constructed. If a sample mean lies outside these decision lines it is declared significantly different from the grand mean. An ANOM chart, conceptually similar to a control chart, portrays decision lines so that statistical significance as well as practical significance of samples may be assessed simultaneously.

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. We make an attempt to develop the ANOM procedure of Ott (1967) when the data variate is suppose to follow IRD.

If $(1-\alpha)$ is taken as the confidence coefficient we should have the probability of all the subgroup means to fall within the control limits is $(1-\alpha)$. Assuming independent of subgroups the above probability statement becomes n^{th} power of the probability of a subgroup mean to fall within the limits should be equal to $(1-\alpha)$. 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 $(1-\alpha)^{1/n}$. The same principle is adopted in the rest of this paper through IRD.

Because of this paper aims at exploring ANOM using control limits of extreme value statistics we have considered only the control chart aspects but not any recently developed ANOM tables or techniques. However, a detailed literature about ANOM is available in Rao (2005) and some related works in this direction are Enrick (1976), Schilling (1979), Ohta (1981), Ramig (1983), Mason *et al.* (1989), Bakir (1994), Bernard and Wludyka (2001), Wludyka *et al.* (2001), Montgomery (2001), Nelson and Dudewicz and Nelson (2003), Farnum (2004), Guirguis and Tobias (2004) and references therein. The rest of the paper is organized as follows. The basic exposure to extreme value control chart is given in Section 2. ANOM applied to IRD using extreme value control charts of IRD is given in Section 3 followed by illustrative examples in Section 4. Summary and conclusions are given in section 5.

2. Extreme Value Charts

The given sample observations are assumed to follow IRD model. The control lines are determined by the theory of extreme order statistics based on IRD. 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 $(1-\alpha)^{1/n}$ within the limits. This can be formulated as a probability inequality in the following way. $P(x_1 \le L) = \alpha/2$ and $P(x_n \ge U) = \alpha/2$. 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 the desired at 0.9973 then α would be 0.0027. Taking $F(x)$ as the CDF of a standard IRD model (σ = 1), 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 is given in Table 2.1 denoted as $Z_{(1)0.00135}$ and $Z_{(n)0.99865}$.

Table 2.1: Control Limits of Extreme value charts

n	$Z_{(1)0.00135}$	$Z_{(n)0.99865}$
2	0.31540	58.38606
3	0.29881	71.50803
4	0.28806	82.57036
5	0.28024	92.31647
6	0.27415	101.12762
7	0.26921	109.23032
8	0.26508	116.77212
9	0.26153	123.85554
10	0.25844	130.55520

The values of Table 2.1 indicates the following probability statement:

$$P(Z_{(1)0.00135} < Z_i < Z_{(n)0.99865}, \ \forall \ i=1,2,...,n) = 0.9973 \quad (2.0.3)$$

$$P(\sigma Z_{(1)0.00135} < x_i < \sigma Z_{(n)0.99865}, \ \forall \ i=1,2,...,n) = 0.9973 \quad (2.0.4)$$

Taking $\bar{x}$ 1.7724 as an unbiased estimate of σ, the above equation becomes

$$\Rightarrow P(D_3^* \bar{x} < x_i < D_4^* \bar{x}, \quad \forall \ i=1,2,...,n) = 0.9973$$

Where $D_3^* = \frac{Z_{(1)0.00135}}{1.7724}$ and $D_4^* = \frac{Z_{(n)0.99865}}{1.7724}$. Thus D_3^* and D_4^* would constitute the control chart constants for the extreme value charts. These are given in Table 2.2 for n = 2 (1) 10.

Table 2.2: Constants of Extreme value charts

n	D_3^*	D_4^*
2	0.17795	32.94181
3	0.16859	40.34531
4	0.16253	46.58675
5	0.15811	52.08557
6	0.15468	57.05689
7	0.15189	61.62848
8	0.14956	65.88362
9	0.14756	69.88013
10	0.14581	73.66012

3. Analysis of Means (ANOM) - Inverse Rayleigh Distribution

Suppose $\bar{x}_1, \bar{x}_2, \ldots\ldots, \bar{x}_k$ are arithmetic means of k subgroups of size 'n' each drawn from an IRD 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 sub group means fall within the control limits. Otherwise we say the process lacks control. If α is the level of significance of the above decisions we can have the following probability statements.

$$P\{LCL < \bar{x}_i, \quad \forall\ i = 1\ to\ k < UCL\} = 1 - \alpha \qquad (3.0.6)$$

Using the notion of independent subgroups (3.0.6) becomes

$$P\{LCL < \bar{x}_i < UCL\} = (1 - \alpha)^{1/k} \qquad (3.0.7)$$

With equi–tailed probability for each subgroup mean, we can find two constants say L^* and U^* such that

$$P\{\bar{x}_i < L^*\} = P\{\bar{x}_i > U^*\} = \frac{1 - (1 - \alpha)^k}{2}$$

In the case of normal population L^* and U^* satisfy $U^* = -L^*$. For the skewed populations like IRD we have to calculate L^*, U^* separately from the sampling distribution of $\bar{x}_i$. Accordingly these depend on the subgroup size 'n' and the number of subgroups 'k'. We make us of the equations (3.07) and (3.08) for specified 'n' and 'k' to get L^* and U^* for α = 0.01 and α = 0.05. These are given in Tables 3.2 and 3.3.

A control chart for averages giving 'In Control 'conclusion indicates that all the subgroup means though vary among themselves are homogeneous in some sense. This is exactly the null hypothesis in an analysis of variance technique. Hence the constants of Tables 3.2 and 3.3 can be used as an alternative to analysis of variance technique. For a normal population one can use the tables of Ott (1967) [11]. For an IRD our tables can be used. We therefore present below some examples for which the goodness of fit of IRD model 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.

4. Illustrative Examples

Example 1: Wadsworth (1986): Consider the following data of 25 observations on "A manufactures of metal products that suspected 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.

Suppliers				
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 brands 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 brands of batteries are different at 5 % level of significance.

Weeks of life		
Brand 1	Brand 2	Brand 3
100	76	108
96	80	100
92	75	96
96	84	98
92	82	100

Example 3: Four catalysts that may effect 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.

Catalyst			
1	2	3	4
58.2	56.3	50.1	52.9
57.2	54.5	54.2	49.9
58.4	57.0	55.4	50.0
55.8	55.3	54.9	51.7

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 IRD is a better model, exhibiting significance linear relation between sample and population quantiles.

	IRD	Normal
Example 1	0.9171	0.2067
Example 2	0.9580	0.4149
Example 3	0.9651	0.4447

Treating these observations in the data as a single sample, we have calculated the decision limits for the Normal population as well as inverse Rayleigh population and have given them in the Table 3.4 respectively.

Table 3.4

	(LDL, UDL)	No. of subgroups fall			
		With in the decision lines	Coverage probability	Outside the decision lines	Coverage probability
Example 1 n = 5, k = 5, α =0 .05	[3.379 , 3.517] [5.6299, 39.5272]	3 5	0.6 1.0	2 0	0.4 0.0
Example 2 n = 5, k = 3, α =0 .05	[87.82 , 95.52] [106.4446, 913.2634]	2 3	0.7 1.0	1 0	0.3 0.0
Example 3 n = 4, k = 4, α =0 .05	[26.14, 82.84] [59.0808, 632.7116]	2 4	0.5 1.0	2 0	0.5 0.0

In each cell the first row values represents the Normal distribution and second row values represents the Inverse Rayleigh distribution.

5. Summary and Conclusions

ANOM tables of Ott (1967) [11] yield that the number of homogeneous means for each data set are 3,2,2 respectively and those away from heterogeneity are 2,1,2 respectively. On the other hand when the ANOM tables of our model (IRD) are used for data sets we get the number of homogeneous means to be 5,3,4 respectively without exhibiting deviation of any mean from homogeneity. Thus usage of normal model resulted in homogeneity for some means and deviation from some other means, indicating a possible rejection of these means. This decision is valid if Normal distribution is a good fit to the data. As a comparison, we have already established by Q-Q plot that IRD is a better model than Normal as supported by the Q-Q plot correlation coefficient of each data set with Normal as well as IRD separately. Therefore, all the means to be homogeneous with the help of IRD (Table 3.5) is a better decision than some means to be away from homogeneity using Normal, ANOM procedure.

References

1. Bakir. S.T, *Analysis of means using ranks for randomized complete block designs*, Communications in Statistics- Simulation and Computation 23 (1994), 547-568.
2. Bernard. A.J. and Wludyka. P.S, *Robust i-sample analysis of means type randomization tests for variances*, Journal of Statistical Computation and Simulation 69 (2001), 57-88.
3. Dedewicz. E.J. and Nelson. P.R, *Heteroscedastic analysis of means (hanom)*, American Journal of Mathematical and Management Sciences 23 (2003), 143-181.
4. Enrick. N.L, *An analysis of means in a three-way factorial*, Journal of Quality Technology 8 (1976), 189-196.
5. Farnum. N.R, *Analysis of means tables using mathematical processors*, Quality Engineering 16 (2004), 399-405.
6. Guirguis. G.H and Tobias. R.D, *On the computation of the distribution for the analysis of means*, Communication in Statistics- Simulation and Computation 16 (2004), 861-887.
7. Gunst. R.F. Mason. R.L. and Hess. J.L, *Statistical design and analysis of experiments*, John Wiley and Sons, New York, 1989.
8. Montgomery. D.C, *Design and analysis of experiments*, fifth ed., John Wiley and Sons, 2000.
9. Nelson, P.R. Wludyka. P.S. and Silva. P.R, *Power curves for analysis of means for variances*, Journal of Quality Technology 33 (2001), 60-65.
10. Nelson. P.R. and Dudewicz. E.J, *Exact analysis of means with unequal variances*, Technometrics 44 (2002), 115-119.
11. Ohta. H, *A procedure for pooling data by analysis of means*, Journal of Quality Technology 13 (1981), 115-119.
12. Ott. E.R, *Analysis of means- a graphical procedure*, Industrial Quality Control 24 (1967).
13. Ramig. P.F, *Application of analysis of means*, Journal of Quality Technology 15 (1983), 19-25.
14. Rao. C.V and Prankumar. M, *Anom-type graphical methods for testing the equality of several correlation coefficients*, Gujarat Statistical Review 29 (2002), 47-56.
15. Rao. C.V, *Analysis of means- a review*, Journal of Quality Technology 37 (2005).
16. Schilling. E.G, *A simplified graphical grant lot acceptance sampling procedure*, Journal of Quality Technology 11 (1979), 116-127.
17. Stephens. K.S. Wadsworth. H.M. and Godfrey. A.F, *Modern methods of quality control and improvement*, John Wiley and Sons. Inc. New York, 2nd ed. 1986.

Table 3.2: Inverse Rayleigh Distribution Constants for Analysis of Means (1-α = 0.99)

n	k =1	2	3	4	5	6	7	8	9	10	15	20	30	40	50
2	0.85448 17.07730	0.82339 21.83038	0.79941 25.54116	0.79589 28.16052	0.78588 31.67590	0.78062 34.65051	0.77741 36.13404	0.76043 37.84520	0.75896 38.48962	0.75749 39.91580	0.72465 42.32666	0.71382 44.65455	0.70425 46.37594	0.64107 47.67488	0.57789 48.97382
3	0.97400 14.32808	0.92399 19.23667	0.88982 21.63157	0.87709 26.66736	0.85812 28.59329	0.85585 33.34991	0.84462 36.15921	0.84258 40.72837	0.83196 46.44379	0.82135 55.16400	0.81851 61.23094	0.81717 66.37949	0.78539 66.59434	0.78166 79.66517	0.77793 92.73600
4	1.06427 12.69503	1.01391 18.11874	0.99194 20.30165	0.98495 24.16641	0.97370 24.50200	0.96215 25.12369	0.95936 25.93983	0.95618 26.56538	0.95604 26.85261	0.95589 27.02677	0.94176 36.34138	0.93938 53.97368	0.92586 71.21388	0.91257 85.23645	0.89929 94.56236
5	1.16193 11.49139	1.10037 14.88874	1.08102 17.10290	1.06457 20.22572	1.05002 21.26451	1.02919 22.07586	1.02137 23.45721	1.01763 28.30374	1.00982 28.88048	1.00201 29.36230	0.97935 35.48009	0.96892 48.31933	0.94013 56.31998	0.91807 67.56239	0.89600 79.56237
6	1.19071 13.43964	1.14640 19.03197	1.10642 23.44173	1.09794 29.05212	1.08879 33.51589	1.07830 34.12185	1.04237 39.55799	1.03263 47.35653	1.02406 52.95726	1.01549 59.63255	1.01513 64.23568	1.01512 70.23514	0.97736 78.25639	0.92793 89.56231	0.87850 95.23659
7	1.25728 10.51804	1.19569 13.49930	1.16140 15.28878	1.14872 18.68253	1.14435 20.97097	1.12808 21.03114	1.11283 21.25473	1.10672 21.86218	1.09671 22.25846	1.08669 23.42054	1.06252 35.51519	1.05936 41.27640	0.96193 49.31925	0.95918 58.1234	0.95643 69.25634
8	1.30714 10.33636	1.23662 12.62289	1.21004 14.31379	1.20127 18.51217	1.18365 18.85951	1.16649 19.18591	1.16082 19.38574	1.15409 20.01832	1.15131 22.10203	1.14853 25.77496	1.13918 36.34584	1.13666 42.23564	1.13136 50.12478	1.11799 59.23564	1.10462 71.02315
9	1.36316 9.96866	1.31137 12.16147	1.26927 13.88278	1.25119 16.94146	1.24503 16.39588	1.23493 18.03501	1.22933 19.70967	1.22679 20.46421	1.22496 21.08694	1.22313 22.84848	1.20260 32.39181	1.19902 42.01325	1.17335 56.03698	1.16414 68.23014	1.15494 74.23109
10	1.37832 9.29947	1.34108 11.73556	1.31834 13.57333	1.30062 16.89849	1.28208 17.26595	1.27588 18.64872	1.27326 18.67806	1.27155 21.11200	1.26978 25.36958	1.26801 29.63442	1.19279 35.20369	1.17624 42.69387	1.15606 50.23698	1.13160 58.23410	1.10714 69.23001

Table 3.3: Inverse Rayleigh Distribution Constants for Analysis of Means (1-α = 0.95)

n	1	2	3	4	5	6	7	8	9	10	15	20	30	40	50
2	1.00597 8.24548	0.94274 10.99050	0.89226 13.24835	0.87278 14.87059	0.85451 17.06327	0.84893 18.08129	0.84036 19.46764	0.82878 20.54635	0.82594 20.90586	0.82368 21.76887	0.80118 25.54116	0.79594 32.67590	0.78552 34.29048	0.76043 38.84520	0.75749 38.91580
3	1.14055 7.25130	1.07491 9.47307	1.03026 11.82874	0.99689 13.01135	0.97805 14.26081	0.96691 15.01537	0.95948 17.61906	0.94550 18.05645	0.93312 18.91424	0.92445 19.05136	0.89154 21.63157	0.88189 22.59329	0.85594 27.76890	0.84258 36.72837	0.82135 55.16400
4	1.22500 6.79626	1.14909 8.74355	1.11631 10.10721	1.08430 11.61205	1.06491 12.66264	1.05514 13.45529	1.04624 14.19294	1.03367 15.58742	1.01805 17.86069	1.01672 17.97045	0.99274 20.30165	0.98573 21.50200	0.96830 24.66198	0.95618 26.56538	0.95589 27.02677
5	1.30963 6.45101	1.23930 8.33600	1.20703 9.96287	1.18611 10.67705	1.16366 11.49046	1.15021 12.89824	1.14144 13.56349	1.11824 14.49699	1.10598 14.63227	1.10329 14.82594	1.08328 17.10290	1.07093 18.26451	1.04636 21.52091	1.01763 28.30374	1.00201 29.36230
6	1.35342 6.26251	1.27741 8.27609	1.23807 9.73974	1.20922 11.53357	1.19087 13.06461	1.17466 14.12457	1.16683 15.51316	1.16054 15.79422	1.15180 17.72442	1.14989 18.51918	1.10814 23.44173	1.10073 25.51589	1.08807 33.87675	1.03263 47.35653	1.01549 58.63526
7	1.43083 5.88863	1.35618 7.49318	1.31410 8.66073	1.28171 9.80909	1.26203 10.45695	1.23536 11.58157	1.21879 12.29849	1.21086 12.70628	1.19877 13.25381	1.19573 13.27092	1.16356 15.28878	1.14950 15.97097	1.13843 20.99200	1.10672 21.26218	1.08669 23.42054
8	1.44695 5.71590	1.38760 7.15846	1.34869 8.21998	1.32480 9.26921	1.30813 10.25907	1.29380 11.04316	1.27270 11.61821	1.25842 11.82029	1.25063 12.29202	1.24432 12.45219	1.21376 14.31379	1.20128 16.85951	1.17876 19.07109	1.15409 20.81832	1.14853 25.77496
9	1.49412 5.60920	1.42014 7.18460	1.39235 8.28981	1.37377 8.97227	1.36327 9.91334	1.34805 10.31926	1.34141 10.84661	1.33147 11.16278	1.31711 11.75150	1.31249 12.14814	1.27015 13.88278	1.25249 16.39588	1.24421 17.06145	1.22679 20.46421	1.22313 22.84848
10	1.53157 5.37731	1.46659 6.78438	1.42323 8.03952	1.39821 8.68432	1.38339 9.28536	1.36621 10.14494	1.35741 10.57363	1.35547 10.83351	1.34542 11.41635	1.34188 11.66737	1.31966 13.57333	1.30464 16.26595	1.27891 17.63707	1.27155 21.11200	1.26801 29.63442