

Anderson-Darling: A Goodness of Fit Test for Small Samples Assumptions

Table of Contents

- Introduction
- Some Statistical Background
- Fitting a Normal Using the Anderson Darling GoF Test
- Fitting a Weibull Using the Anderson Darling GoF Test
- A Counter Example
- Summary
- Bibliography
- About the Author
- Other START Sheets Available

Introduction

Most statistical methods assume an underlying distribution in the derivation of their results. However, when we assume that our data follow a specific distribution, we take a serious risk. If our assumption is wrong, then the results obtained may be invalid. For example, the confidence levels of the confidence intervals (CI) or hypotheses tests implemented [2, 7] may be completely off. Consequences of mis-specifying the distribution may prove very costly. One way to deal with this problem is to check the distribution assumptions carefully.

There are two main approaches to checking distribution assumptions [2, 3, and 6]. One involves empirical procedures, which are easy to understand and implement and are based on intuitive and graphical properties of the distribution that we want to assess. Empirical procedures can be used to check and validate distribution assumptions. Several of them have been discussed at length in other RAC START sheets [8, 9, and 10].

There are also other, more formal, statistical procedures for assessing the underlying distribution of a data set. These are the Goodness of Fit (GoF) tests. They are numerically convoluted and usually require specific software to perform the lengthy calculations. But their results are quantifiable and more reliable than those from the empirical procedure. Here, we are interested in those theoretical GoF procedures specialized for small samples. Among them, the Anderson-Darling (AD) and the Kolmogorov-Smirnov (KS) tests stand

out. This START sheet discusses the former of the two; the latter (KS) is discussed in [12].

In this START sheet, we provide an overview of some issues associated with the implementation of the AD GoF test, especially when assessing the Exponential, Weibull, Normal, and Lognormal distribution assumptions. These distributions are widely used in quality and reliability work. We first review some theoretical considerations to help us better understand (and apply) these GoF tests. Then, we develop several numerical and graphical examples that illustrate how to implement and interpret the GoF tests for fitting several distributions.

Some Statistical Background

Establishing the underlying distribution of a data set (or random variable) is crucial for the correct implementation of some statistical procedures. For example, both the small sample t test and CI, for the population mean, require that the distribution of the underlying population be Normal. Therefore, we first need to establish (via GoF tests) whether the Normal applies before we can correctly implement these statistical procedures.

GoF tests are essentially based on either of two distribution elements: the cumulative distribution function (CDF) or the probability density function (pdf). The Chi-Square test is based on the pdf. Both the AD and KS GoF tests use the cumulative distribution function (CDF) approach and therefore belong to the class of "distance tests."

We have selected the AD and KS from among the several distance tests for two reasons. First, they are among the best distance tests for small samples (and they can also be used for large samples). Secondly, because various statistical packages are available for both AD and KS, they are widely used in practice. In this START sheet, we will demonstrate how to use the AD test with the one particular software package - Minitab.

To implement distance tests, we follow a well-defined series of steps. First, we assume a pre-specified distribution (e.g., Normal). Then, we estimate the distribution parameters (e.g., mean and variance) from the data or obtain them from prior experiences. Such a process yields a distribution hypothesis,

also called the null hypothesis (or H_0), with several parts that must be jointly true. The negation of the assumed distribution (or its parameters) is the alternative hypothesis (also called H_1). We then test the assumed (hypothesized) distribution using the data set. Finally, H_0 is rejected whenever any one of the elements composing H_0 is not supported by the data.

In distance tests, when the assumed distribution is correct, the theoretical (assumed) CDF (denoted F_0) closely follows the empirical step function CDF (denoted F_n), as conceptually illustrated in Figure 1. The data are given as an ordered sample $\{X_1 \leq X_2 \leq \dots \leq X_n\}$ and the assumed (H_0) distribution has a CDF, $F_0(x)$. Then we obtain the corresponding GoF test statistic values. Finally, we compare the theoretical and empirical results. If they agree (probabilistically) then the data supports the assumed distribution. If they do not, the distribution assumption is rejected.

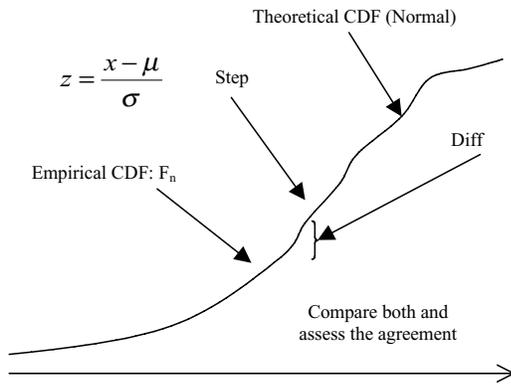


Figure 1. Distance Goodness of Fit Test Conceptual Approach

The test has, however, an important caveat. Theoretically, distance tests require the knowledge of the assumed distribution parameters. These are seldom known in practice. Therefore, adaptive procedures are used to circumvent this problem when implementing GoF tests in the real world (e.g., see [6], Chapter 7). This drawback of the AD GoF test, which otherwise is very powerful, has been addressed in [4, 5] by using some implementation procedures. The AD test statistics (formulas) used in this START sheet and taken from [4, 5] have been devised for their use with parameters estimated from the sample. Hence, there is no need for further adaptive procedures or tables, as does occur with the KS GoF test that we demonstrate in [12].

Fitting a Normal Using the Anderson-Darling GoF Test

Anderson-Darling (AD) is widely used in practice. For example, MIL-HDBKs 5 and 17 [4, 5, and 2], use AD to test Normality and Weibull. In this and the next section, we develop two examples using the AD test; first for testing Normality and then, in the next section, for testing the Weibull assumption. If there is a

need to test for Lognormality, then log-transform the original data and use the AD Normality test on the transformed data set.

The AD GoF test for Normality (Reference [5] Section 8.3.4.1) has the functional form:

$$AD = \sum_{i=1}^n \frac{1-2i}{n} \left\{ \ln(F_0[Z_{(i)}]) + \ln(1 - F_0[Z_{(n+1-i)}]) \right\} - n; \dots \quad (1)$$

where F_0 is the assumed (Normal) distribution with the assumed or sample estimated parameters (μ, σ); $Z_{(i)}$ is the i th sorted, standardized, sample value; “ n ” is the sample size; “ \ln ” is the natural logarithm (base e) and subscript “ i ” runs from 1 to n .

The null hypothesis, that the true distribution is F_0 with the assumed parameters, is then rejected (at significance level $\alpha = 0.05$, for sample size n) if the AD test statistic is greater than the critical value (CV). The rejection rule is:

$$\text{Reject if: } AD > CV = 0.752 / (1 + 0.75/n + 2.25/n^2)$$

We illustrate this procedure by testing for Normality the tensile strength data in problem 6 of Section 8.3.7 of [5]. The data set, (Table 1), contains a small sample of six batches, drawn at random from the same population.

Table 1. Data for the AD GoF Tests

338.7	308.5	317.7	313.1	322.7	294.2
-------	-------	-------	-------	-------	-------

To assess the Normality of the sample, we first obtain the point estimations of the assumed Normal distribution parameters: sample mean and standard deviation (Table 2).

Table 2. Descriptive Statistics of the Prob 6 Data

Variable	N	Mean	Median
Data Set	6	315.82	315.40

Under a Normal assumption, F_0 is normal ($\mu = 315.8$, $\sigma = 14.9$).

We then implement the AD statistic (1) using the data (Table 1) as well as the Normal probability and the estimated parameters (Table 2). For the smallest element we have:

$$P_{\mu=315.8, \sigma=14.8}(294.2) = \text{Normal}\left(\frac{294.2 - 315.8}{14.8}\right)$$

$$= F_0(z) = F_0(-1.456) = 0.0727$$

Table 3 shows the AD statistic intermediate results that we combine into formula (1). Each component is shown in the corresponding table column, identified by name.

Table 3. Intermediate Values for the AD GoF Test for Normality

i	X	F(Z)	ln F(Z)	n+1-i	F(n+1-i)	1-F(n+1-i)	ln(1-F)
1	294.2	0.072711	-2.62126	6	0.938310	0.061690	-2.78563
2	308.5	0.311031	-1.16786	5	0.678425	0.321575	-1.13453
3	313.1	0.427334	-0.85019	4	0.550371	0.449629	-0.79933
4	317.7	0.550371	-0.59716	3	0.427334	0.572666	-0.55745
5	322.7	0.678425	-0.38798	2	0.311031	0.688969	-0.37256
6	338.7	0.938310	-0.06367	1	0.072711	0.927289	-0.07549

The AD statistic (1) yields a value of $0.1699 < 0.633$, which is non-significant:

$$AD = 0.1699 < CV = \frac{0.752}{1 + 0.75/6 + 2.25/36} = \frac{0.752}{1 + 0.125 + 0.0625} = 0.6333$$

Therefore, the AD GoF test does not reject that this sample may have been drawn from a Normal (315.8, 14.9) population. And we can then assume Normality for the data.

In addition, we present the AD plot and test results from the Minitab software (Figure 2). Having software for its calculation is one of the strong advantages of the AD test. Notice how the Minitab graph yields the same AD statistic values and estimations that we obtain in the hand calculated Table 3. For example, A-Square (= 0.17) is the same AD statistic in formula (1). In addition, Minitab provides the GoF test p-value (= 0.88) which is the probability of obtaining these test results, when the (assumed) Normality of the data is true. If the p-value is not small (say 0.1 or more) then, we can assume Normality. Finally, if the data points (in the Minitab AD graph) show a linear trend, then support for the Normality assumption increases [9].

The AD GoF test procedures, applied to this example, are summarized in Table 4.

Finally, if we want to fit a Lognormal distribution, we first take the logarithm of the data and then implement the AD GoF procedure on these transformed data. If the original data is Lognormal, then its logarithm is Normally distributed, and we can use the same AD statistic (1) to test for Lognormality.

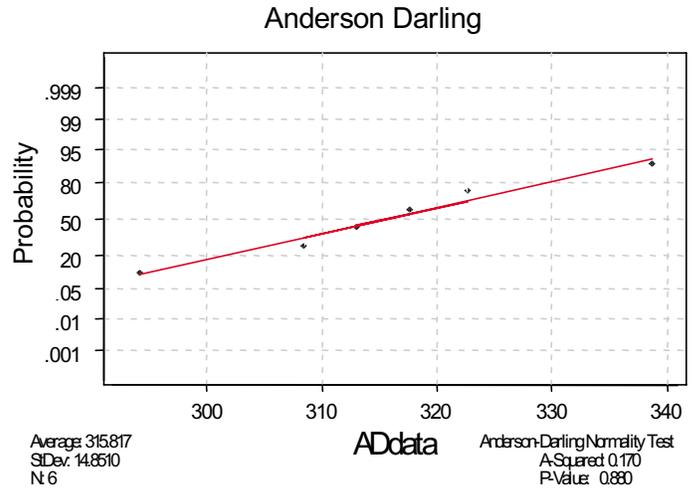


Figure 2. Computer (Minitab) Version of the AD Normality Test

Table 4. Step-by-Step Summary of the AD GoF Test for Normality

- Sort Original (X) Sample (Col. 1, Table 3) and standardize: $Z = (x - \mu)/\sigma$
- Establish the Null Hypothesis: assume the Normal (μ, σ) distribution
- Obtain the distribution parameters: $\mu = 315.8; \sigma = 14.9$ (Table 2)
- Obtain the F(Z) Cumulative Probability (Col. 2, Table 3)
- Obtain the Logarithm of the above: $\ln[F(Z)]$ (Col. 3)
- Sort Cum-PROBS F(Z) in descending order ($n - i + 1$) (Cols. 4 and 5)
- Find the Values of $1 - F(Z)$ for the above (Col. 6)
- Find Logarithm of the above: $\ln[1 - F(Z)]$ (Col. 7)
- Evaluate via (1) Test Statistics $AD = 0.1699$ and $CV = 0.633$
- Since $AD < CV$ assume distribution is Normal (315.8, 14.9)
- When available, use the computer software and the test p-value

Fitting a Weibull Using the Anderson-Darling GoF Test

We now develop an example of testing for the Weibull assumption. We will use the data in Table 5, which will also be used for this same purpose in the implementation of the companion Kolmogorov-Smirnov GoF test [12]. The data consist of six measurements, drawn from the same Weibull ($\alpha = 10; \beta = 2$) population. In our examples, however, the parameters are unknown and will be estimated from the data set.

Table 5. Data Set for Testing the Weibull Assumption

11.7216	10.4286	8.0204	7.5778	1.4298	4.1154
---------	---------	--------	--------	--------	--------

We obtain the descriptive statistics (Table 6). Then, using graphical methods in [1], we get point estimations of the assumed Weibull parameters: shape $\beta = 1.3$ and scale $\alpha = 8.7$. The parameters allow us to define the distribution hypothesis: Weibull ($\alpha = 8.7; \beta = 1.3$).

Table 6. Descriptive Statistics

Variable	N	Mean	Median	StDev	Min	Max	Q1
Data Set	6	7.22	7.80	3.86	1.43	11.72	3.44

The Weibull version of the AD GoF test statistic is different from the one for Normality, given in the previous section. This Weibull version is explained in detail in [2, 5] and is defined by:

$$AD = \left[\sum_i \frac{1-2i}{n} \left\{ \ln(1 - \exp(-Z_{(i)})) - Z_{(n+1-i)} \right\} - n \right]$$

and $AD^* = (1 + 0.2/\sqrt{n})AD$ (2)

where $Z_{(i)} = [x_{(i)}/\theta^*]^{\beta^*}$ and where the asterisks (*) in the Weibull parameters denote the corresponding estimations. The OSL (observed significance level) probability (p-value) is now used for testing the Weibull assumption. If $OSL < 0.05$ then the

Weibull assumption is rejected and the error committed is less than 5%. The OSL formula is given by:

$$OSL = 1/\{1 + \exp[-0.1 + 1.24 \ln(AD^*) + 4.48 (AD^*)]\}$$

To implement the AD GoF test, we first obtain the corresponding Weibull probabilities under the assumed distribution H_0 . For example, for the first data point (1.43):

$$P_{\alpha=8.7; \beta=1.3}(1.43) = 1 - \exp(-z_1) = 1 - \exp\left\{-\left(\frac{x_1}{\alpha}\right)^\beta\right\}$$

$$= 1 - \exp\left\{-\left(\frac{1.43}{8.7}\right)^{1.3}\right\} = 1 - 0.909 = 0.091$$

Then, we use these values to work through formulas AD and AD^* in (2). Intermediate results, for the small data set in Table 5, are given in Table 7.

Table 7. Intermediate Values for the AD GoF Test for the Weibull

Row	DataSet	Z(i)	WeibProb	Exp-Z(i)	Ln(1-Ez)	Zn-i+1	ith-term
1	1.430	0.09560	0.091176	0.908824	-2.39496	1.47336	0.64472
2	4.115	0.37789	0.314692	0.685308	-1.15616	1.26567	1.21092
3	7.578	0.83566	0.566413	0.433587	-0.56843	0.89967	1.22342
4	8.020	0.89967	0.593296	0.406704	-0.52206	0.83566	1.58401
5	10.429	1.26567	0.717949	0.282051	-0.33136	0.37789	1.06387
6	11.722	1.47336	0.770846	0.229154	-0.26027	0.09560	0.65242

The AD GoF test statistics (2) values are: $AD = 0.3794$ and $AD^* = 0.4104$. The value corresponding to the OSL, or probability of rejecting the Weibull (8.7; 1.3) distribution erroneously with these results, is $OSL = 0.3466$ (much larger than the error $\alpha = 0.05$).

Hence, we accept the null hypothesis that the underlined distribution (of the population from where these data were obtained) is Weibull ($\alpha = 8.7; \beta = 1.3$). Hence, the AD test was able to recognize that the data were actually Weibull. The GoF procedure for this case is summarized in Table 8.

Table 8. Step-by-Step Summary of the AD GoF Test for the Weibull

• Sort Original Sample (X) and Standardize: $Z = [x(i)/\theta^*]^{\beta^*}$ (Cols. 1 & 2, Table 7).
• Establish the Null Hypothesis: assume Weibull distribution.
• Obtain the distribution parameters: $\alpha = 8.7; \beta = 1.3$.
• Obtain Weibull probability and $\exp(-Z)$ (Cols. 3 & 4).
• Obtain the Logarithm of $1 - \exp(-Z)$ (Col. 5).
• Sort the Z(i) in descending order (n-i+1) (Col. 6).
• Evaluate via (1): $AD^* = 0.4104$ and $OSL = 0.3466$.
• Since $OSL = 0.3466 > \alpha = 0.05$, assume Weibull ($\alpha = 8.7; \beta = 1.3$).
• Software for this version of AD is not commonly available.

Finally, recall that the Exponential distribution, with mean α , is only a special case of the Weibull ($\alpha; \beta$) where the shape parameter $\beta = 1$. Therefore, if we are interested in using AD GoF test to assess Exponentiality, it is enough to estimate the sample mean (α) and then to implement the above Weibull procedure for this special case, using formula (2).

There are not, however, AD statistics (formulas) for all the distributions. Hence, if there is a need to fit other distributions than the four discussed in this START sheet, it is better to use the Kolmogorov Smirnov [12] or the Chi Square [11] GoF tests.

A Counter Example

For illustration purposes we again use the data set ‘prob6’ (Table 1), which was shown to be Normally distributed. We will now use the AD GoF procedure for assessing the assumption that the data distribution is Weibull. The reader can find more information on this method in Section 8.3.4 of MIL-HDBK-17 (1E) [5] and in [2].

We use Weibull probability paper, as explained in [1] to estimate the shape (β) and scale (α) parameters from the data. These esti-

mations yield 8 and 350, respectively, and allow us to define the distribution hypothesis H_0 : Weibull ($\alpha = 350$; $\beta = 8$).

We again use the same Weibull version [5] of AD and AD* GoF test statistics (2) to obtain the OSL value, as was done in the previous section. And as before, if $OSL < 0.05$ then the Weibull assumption is rejected and the error committed is less than 5%.

As an illustration, we obtain the corresponding probability (under the assumed Weibull distribution) for the first data point (294.2).

$$P_{\alpha=350;\beta=8}(294.2) = 1 - \exp(-z_1) = 1 - \exp\left\{-\left(\frac{x_1}{\alpha}\right)^\beta\right\}$$

$$= 1 - \exp\left\{-\left(\frac{294.2}{350}\right)^8\right\} = 1 - 0.7794 = 0.2206$$

Then, we use these values to work through formulas AD and AD* in (2). Intermediate results, for the small data set in Table 1, are given in Table 9.

The AD GoF test statistics (2) values are $AD = 2.7022$ and $AD^* = 2.9227$. The corresponding OSL, or probability of rejecting Weibull (8, 350) distribution erroneously, with these results is ($OSL = 6 \times 10^{-7}$) extremely small (i.e., less than $\alpha = 0.05$).

Hence, we (correctly) reject the null hypothesis that the underlined distribution (of the population from where these data were obtained) is Weibull ($\alpha = 350$; $\beta = 8$). As we verify, the AD test was able to recognize that the data were actually not Weibull. The entire GoF procedure, for this case, is summarized in Table 10.

Table 9. Intermediate Values for the AD GoF Test for the Weibull

ith	x_i	z_i	$\exp(-z_i)$	$\ln(1-\exp)$	$n+1-i$	$z(n+1-i)$	ith-term
1	294.2	0.249228	0.779402	-1.51141	6	0.769090	0.29344
2	308.5	0.364332	0.694661	-1.18633	5	0.522213	0.77533
3	313.1	0.410129	0.663565	-1.08935	4	0.460886	1.24957
4	317.7	0.460886	0.630725	-0.99621	3	0.410129	1.69995
5	322.7	0.522213	0.593206	-0.89945	2	0.364332	2.13249
6	338.7	0.769090	0.463434	-0.62257	1	0.249228	2.55137

Table 10. Step-by-Step Summary of the AD GoF Test for the Weibull

- Sort Original Sample (X) and standardize: $Z = [x_{(i)}/\theta^*]^{\beta^*}$ (Cols. 1 & 2, Table 5).
- Establish the Null Hypothesis: assume Weibull distribution.
- Obtain the distribution parameters: $\alpha = 350$; $\beta = 8$.
- Obtain the $\exp(-Z)$ values (Col. 3).
- Obtain the Logarithm of $1 - \exp(-Z)$ (Col. 4).
- Sort the $Z(i)$ in descending order of $(n-i+1)$ (Cols. 5 and 6).
- Evaluate via (1): $AD^* = 2.92$ and $OSL = 6 \times 10^{-7}$.
- Since $OSL = 6 \times 10^{-7} < \alpha = 0.05$, reject assumed Weibull ($\alpha = 350$; $\beta = 8$).
- Software for this version of AD is not commonly available.

cal and practical issues and have provided several references for background information and further readings.

The large sample GoF problem is often better dealt with via the Chi-Square test [11]. It does not require knowledge of the distribution parameters - something that both, AD and KS tests theoretically do and that affects their power. On the other hand, the Chi-Square GoF test requires that the number of data points be large enough for the test statistic to converge to its underlying Chi-Square distribution - something that neither AD nor KS require. Due to their complexity, the Chi-Square and the Kolmogorov Smirnov GoF test are treated in more detail in separate START sheets [11 and 12].

Summary

In this START Sheet we have discussed the important problem of the assessment of statistical distributions, especially for small samples, via the Anderson Darling (AD) GoF test. Alternatively, one can also implement the Kolmogorov Smirnov test [12]. These tests can also be used for testing large samples. We have provided several numerical and graphical examples for testing the Normal, Lognormal, Exponential and Weibull distributions, relevant in reliability and maintainability studies (the Exponential is a special case of the Weibull, as is the Lognormal of the Normal). We have also discussed some relevant theoretic

Bibliography

1. Practical Statistical Tools for Reliability Engineers, Coppola, A., RAC, 1999.
2. A Practical Guide to Statistical Analysis of Material Property Data, Romeu, J.L. and C. Grethlein, AMPTIAC, 2000.
3. An Introduction to Probability Theory and Mathematical Statistics, Rohatgi, V.K., Wiley, NY, 1976.
4. MIL-HDBK-5G, Metallic Materials and Elements.
5. MIL-HDBK-17 (1E), Composite Materials Handbook.

6. Methods for Statistical Analysis of Reliability and Life Data, Mann, N., R. Schafer, and N. Singpurwalla, John Wiley, NY, 1974.
7. Statistical Confidence, Romeu, J.L., RAC START, Volume 9, Number 4, http://rac.alionscience.com/pdf/STAT_CO.pdf.
8. Statistical Assumptions of an Exponential Distribution, Romeu, J.L., RAC START, Volume 8, Number 2, http://rac.alionscience.com/pdf/E_ASSUME.pdf.
9. Empirical Assessment of Normal and Lognormal Distribution Assumptions, Romeu, J.L., RAC START, Volume 9, Number 6, <http://rac.alionscience.com/pdf/NLDIST.pdf>.
10. Empirical Assessment of Weibull Distribution, Romeu, J.L., RAC START, Volume 10, Number 3, <http://rac.alionscience.com/pdf/WEIBULL.pdf>.
11. The Chi-Square: a Large-Sample Goodness of Fit Test, Romeu, J.L., RAC START, Volume 10, Number 4, http://rac.alionscience.com/pdf/Chi_Square.pdf.
12. Kolmogorov-Smirnov GoF Test, Romeu, J.L., RAC START, Volume 10, Number 6.

Enterprises. Dr. Romeu is a Chartered Statistician Fellow of the Royal Statistical Society, Full Member of the Operations Research Society of America, and Fellow of the Institute of Statisticians.

Romeu is a senior technical advisor for reliability and advanced information technology research with Alion Science and Technology. Since joining Alion in 1998, Romeu has provided consulting for several statistical and operations research projects. He has written a State of the Art Report on Statistical Analysis of Materials Data, designed and taught a three-day intensive statistics course for practicing engineers, and written a series of articles on statistics and data analysis for the AMPTIAC Newsletter and RAC Journal.

Other START Sheets Available

Many Selected Topics in Assurance Related Technologies (START) sheets have been published on subjects of interest in reliability, maintainability, quality, and supportability. START sheets are available on-line in their entirety at <http://rac.alionscience.com/rac/jsp/start/startsheet.jsp>.

About the Author

Dr. Jorge Luis Romeu has over thirty years of statistical and operations research experience in consulting, research, and teaching. He was a consultant for the petrochemical, construction, and agricultural industries. Dr. Romeu has also worked in statistical and simulation modeling and in data analysis of software and hardware reliability, software engineering and ecological problems.

Dr. Romeu has taught undergraduate and graduate statistics, operations research, and computer science in several American and foreign universities. He teaches short, intensive professional training courses. He is currently an Adjunct Professor of Statistics and Operations Research for Syracuse University and a Practicing Faculty of that school's Institute for Manufacturing

For further information on RAC START Sheets contact the:

Reliability Analysis Center
 201 Mill Street
 Rome, NY 13440-6916
 Toll Free: (888) RAC-USER
 Fax: (315) 337-9932

or visit our web site at:

<http://rac.alionscience.com>



About the Reliability Analysis Center

The Reliability Analysis Center is a Department of Defense Information Analysis Center (IAC). RAC serves as a government and industry focal point for efforts to improve the reliability, maintainability, supportability and quality of manufactured components and systems. To this end, RAC collects, analyzes, archives in computerized databases, and publishes data concerning the quality and reliability of equipments and systems, as well as the microcircuit, discrete semiconductor, and electromechanical and mechanical components that comprise them. RAC also evaluates and publishes information on engineering techniques and methods. Information is distributed through data compilations, application guides, data products and programs on computer media, public and private training courses, and consulting services. Located in Rome, NY, the Reliability Analysis Center is sponsored by the Defense Technical Information Center (DTIC). Alion, and its predecessor company IIT Research Institute, have operated the RAC continuously since its creation in 1968. Technical management of the RAC is provided by the U.S. Air Force's Research Laboratory Information Directorate (formerly Rome Laboratory).