START
Selected Topics in Assurance Related Technologies
Volume 9, Number 5

Graphical Comparisons of Two Populations

Table of Contents

Introduction

This START sheet addresses the problem of graphically assessing two populations based on two samples. Such comparisons arise, for example, when analyzing two processes (or two devices) whose performance (e.g., reliability or lives) we want to compare. Representative items of each process are placed on test to determine whether their performances follow the same distribution. For, the performances of two devices are equal only if their distributions and parameters are the same.

To make our point, we present in Figure 1 two different distributions with similar means and standard deviations. One is Normal (9, 2.75) and the other is Weibull (3.6, 10). They look similar in the body (the centered 95%) of the distribution. But they do differ in the "tails" (extreme upper/lower values) as can be observed in the (Min) statistics of Table 1.

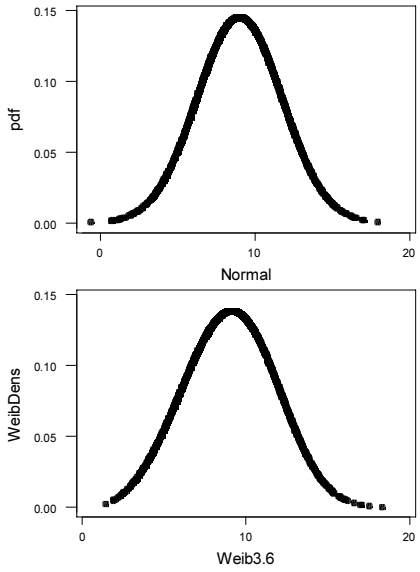

Figure 1. Two Apparently Similar Distributions: Normal (9, 2.7) and Weibull (3.6,10)

Table 1. Descriptive Statistics for a Sample of 2000 from These Two Distributions

Variable	N	Mean	Median	StDev	Min	Max
Weib3.6	2000	8.9625	8.9452	2.7605	1.4230	18.2940
Norm9	2000	8.9396	8.8778	2.7824	**-0.4785**	19.1298

For example, the probability of a Weibull taking negative values is zero. The Normal distribution, however, has a non-zero probability of taking negative values, since:

$$P_{\mu=9;\sigma=2.75}\{\text{Normal} < 0\} = \text{PNormal}\left(\frac{0-9}{2.75}\right) =$$

$$\text{PNormal}(-3.27) = 0.0005 > 0.0$$

The practical importance of such small probability, depends on the application. For example, if producing 1 million widgets and assuming their lives are Normal (9, 2.7) instead of the Weibull (3.6, 10), we will end up with 5000 "negative lives"!

In the rest of this START sheet, two samples are compared using simple graphical and analytical methods. We assess whether their two underlying distributions are equal. If they are, we conclude that both populations (performances) are equal. We illustrate the use of such graphical procedures by progressively comparing two samples. First we analyze similar and then different, well-known distributions, frequently used in reliability. We do it for two large samples of the same size, then for two small samples of different sizes. We end with a real-life, practical example that illustrates the procedure application.

Large Sample Procedures

To start our discussion, let's consider two identical Normal populations, N ($\mu = 100$, $\sigma = 10$) denoted N100-1 and N100-2. Let's assume we have two large samples of, say, 40 points each. We first obtain their Q-Q (Quantile-Quantile) Plot by (1) sorting both samples in ascending order and then (2) plotting one sorted sample vs. the other (Figure 2).

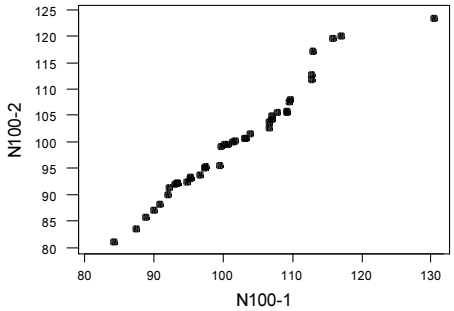

Figure 2. Q-Q Plot for Two Normal Distributions With Same Parameters

When two populations are equal, their Q-Q Plot follows an upward linear trend, with unit slope and similar range. Hence, the linear regression of these two (Q-Q plot) sorted pairs of variables, must reflect this one-to-one relation in their ranges (80:125 vs. 80:130):

N100-2 = -6.40 + 1.04 N100-1

Predictor	Coef	StDev	T	P
Constant	-6.398	3.454	-1.85	0.072
N100-1	1.04450	0.03380	30.90	0.000

R-Sq = 96.2%

Notice how the regression Index of Fit (R^2 = 96%) is very high (close to 100%). Also, the P value (0.00) for the regression coefficient T-Test (30.9) is smaller than 0.05, denoting a linear trend. The regression coefficient (slope) itself (1.0445) is close to unity, indicating a strong similarity between these two samples and their statistical distributions. A slope of unity (regression coefficient) serves as the "litmus test" of this graphical approach.

If we add and subtract three times the standard deviation (0.0338) from the regression coefficient, we obtain an approximate 99% confidence interval (CI) for the coefficient true value. In the present case this CI (0.9431, 1.145) covers unity, the value that the regression coefficient should have when the two populations or distributions are equal.

For contrast, we now compare a sample from the Normal (100, 10) with another sample, also from a Normal (denoted N70 - 20) but having different parameters: $\mu = 70$ and $\sigma = 20$. We present their Q-Q plot in Figure 3.

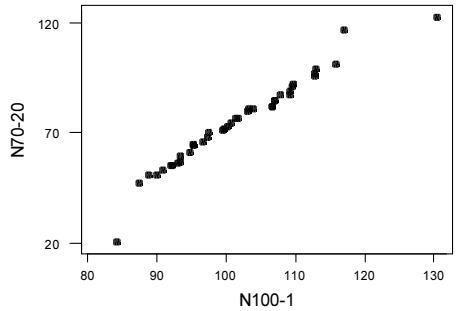

Figure 3. Q-Q Plot of Two Different Normal Distributions

Compare Q-Q plots in Figures 2 and 3. The latter differs in the variable ranges (20:120 vs. 80:130) that point toward these distribution differences. We therefore implement a linear regression to assess how the slope reflects this 2:1 ratio difference, in the ranges between the two variables and obtain:

N70-20 = -132 + 2.03 N100-1

Predictor	Coef	StDev
Constant	-132.066	7.325
N100-1	2.03484	0.07169

R-Sq = 95.5%

The regression coefficient is now 2.03, instead of unity, previously obtained when the two populations were identical. The approximate 99% CI is now (1.82, 2.25) and does not cover unity. All three samples were Normal, but the third had different parameters. Q-Q plots were able to effectively pick up such similarities and differences in distributions.

Now, consider a different distribution: an Exponential (but with the same mean 100). We perform the same Q-Q Plot graphical analysis to compare a large sample of 40 data points (denoted Exp100) from this distribution with the Normal (100, 10) sample.

The Q-Q plot of the Exponential vs. Normal samples (Figure 4) appears rather S-curved, with differing ranges (0:250 vs. 80:130). Its corresponding regression yields:

Exp100 = -636 + 7.15 N100-1

Predictor	Coef	StDev
Constant	-636.06	37.91
N100-1	7.1527	0.3710

R-Sq = 90.7%

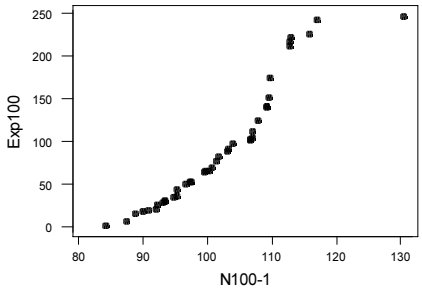

Figure 4. Q-Q Plot for the Normal and Exponential Distributions

The regression of the sorted data yields a coefficient 7.15, and an approximate 99% CI of (6.039, 8.265). These results (differing from unity) suggest both, a difference in variable ranges and in statistical distributions. Hence, if these samples are performances from two devices, we can state that the performances are not the same.

Finally, let's compare a fourth sample, of also 40 data points, from a Weibull distribution with Shape parameter 2, and Scale 120 (yielding an approximate mean of 100). Let's first consider the Q-Q plot of the Normal versus the Weibull samples (Figure 5).

The S-shaped Q-Q plot in Figure 5 also differs in ranges. The corresponding regression:

Weib-100 = -449 + 5.34 N100-1

Predictor	Coef	StDev
Constant	-449.19	19.15
N100-1	5.3424	0.1874

R-Sq = 95.5%

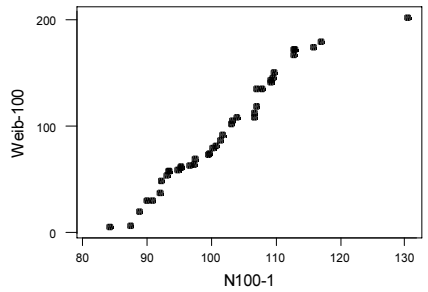

Figure 5. Q-Q Plot of the Weibull and the Normal Samples

The regression coefficient (5.34) and its approximate 99% CI (4.78, 5.91) are far greater than unity, as expected when the Q-Q plot of two different distribution are compared. The Q-Q plot has again picked up a difference between these distributions.

In Figure 6, we now present the Q-Q plots comparing two large (n = 40) Exponential and the Weibull data sets. Recall that an Exponential can be considered a Weibull with shape parameter equal to 1.

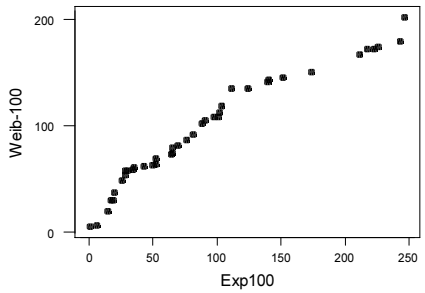

Figure 6. Q-Q Plot for the Weibull and the Exponential Data

The Q-Q plot in Figure 6 appears S-shaped and the data ranges (0:200 vs. 0:250) also differ. The regression on these two sorted data sets, implemented to assess the slope, is:

Weib-100 = 29.5 + 0.707 Exp100

Predictor	Coef	StDev
Constant	29.518	3.222
Exp100	0.70736	0.02783

R-Sq = 94.4%

Notice how the regression coefficient (0.707) and its approximate 99% CI (0.624, 0.791) also differ from unity. These results suggest that both data sets come from two different statistical distributions (as they actually do). The Q-Q Plot again picked up the difference.

To further illustrate how Q-Q plots help assess whether two distributions are similar or different, we generate two additional data sets of 40 observations, with parameters as before. One data set comes from the Exponential and the second from the Weibull:

Variable	N	Mean	Median	StDev	Min	Max
Exp100	40	91.7	73.0	71.5	0.7	246.8
Exp100-2	40	86.9	64.5	90.2	0.7	405.9
Weib100	40	94.42	83.92	52.03	4.22	202.7
Weib100-2	40	110.99	110.54	58.75	17.86	245.4

We show that the descriptive statistics for the pairs of data sets (means, medians, and standard deviations) of the two Exponentials and the two Weibulls, are close (as they should be). The maximum values do differ, since these are quite skewed distributions. The Q-Q plots reflect this.

Because Weibull and Exponential are highly skewed distributions, we trim their upper 5% tail to compensate for the existence of large outliers in their tails. Trimming is done by discarding the (0.05 x 40 = 2) two largest observations from each data set (Figure 7). This helps to improve the respective ranges.

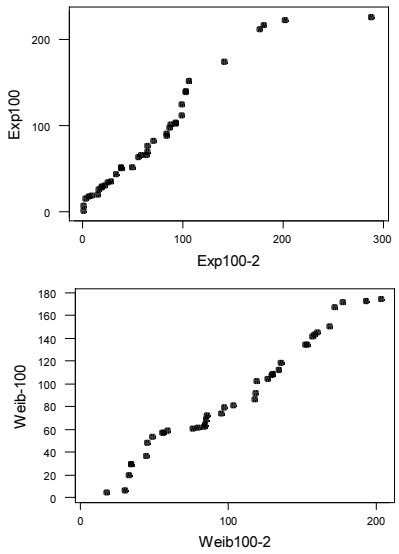

Figure 7. Trimmed Q-Q Plots for Two Exponentials and Two Weibull

To assess whether the slope is unity, we again implement the two regression models. The regression equation for comparing the two Exponential data sets is:

Exp100 = 12.8 + 0.986 Exp100-2

Predictor	Coef	StDev	T	P
Constant	12.802	4.178	3.06	0.004
Exp100-2	0.98586	0.04415	22.33	0.000

S = 16.74 R-Sq = 93.3% R-Sq(adj) = 93.1%

The regression equation for comparing the two Weibull data sets is:

Weib-100 = -6.75 + 0.924 Weib100-2

Predictor	Coef	StDev
Constant	-6.753	3.056
Weib100-2	0.92384	0.02642

R-Sq = 97.1%

In both cases, the regression Index of Fit values (93% and 97%) are close to 100% (the higher, the better). The regression coefficients (0.985 and 0.923) are very close to unity and their approximate 99% CI, (0.853, 1.11700) and (0.845, 1.003), also cover unity. Q-Q plots and regression results support the similarities between these two pairs of data sets.

All above examples show how, with the aid of Q-Q plots, we can assess the similarity or lack thereof, between two distributions. However, Q-Q plots are qualitative rather than quantitative methods. So, once a distribution similarity has been established, we can then test their parameters for equality. For example, in the case of the two large Normal data sets (N100 - 1 and N100 - 2) Q-Q plots established that both these samples follow the same distribution. We should now compare the Normal parameters mean and variance via the two-sample t-test and the two sample F-test. The specifics of these test procedures will be part of a separate paper. The results show that both means and variances also agree.

Hence, two Normal distributions that have the same mean and variances (i.e., the same parameters) are equal. Therefore, the performance of the two devices under comparison, exhibiting such lives distributions as reflected in both data sets, is identical.

Small Sample Case

Some times the data sets are small, or unbalanced (of different size). The above Q-Q plot procedures can still be adapted in the manner illustrated with the following two small, unbalanced but equally Exponential samples (of sizes $n_1 = 10$; $n_2 = 7$):

Exp100-s1 185.169 33.238 21.551 80.269 170.875 130.811 114.070 69.359 40.923 16.306

Exp100-s2 21.468 16.743 71.948 59.081 1.018 98.723 202.553

First, combine ($n = n_1 + n_2 = 17$) the two samples and then sort the data in ascending order as in Table 2.

The relative frequencies (F_1; F_2) are obtained by dividing $F_j = i/n_j$ for I = 1, .. n_j and j = 1,2 (by each data set). Then, we obtain the absolute values of the differences F_1-F_2 (AbsDif).

Table 2. Combined, Sorted Sample With the Relative Frequencies and Absolute Differences

Row	Combined	F1	F2	AbsDif(F1-F2)
1	1.018	0/10	1/7	0.14
2	16.306	1/10	1/7	0.04
3	16.743	1/10	2/7	0.19
4	21.468	1/10	3/7	**0.33**
5	21.551	2/10	3/7	0.23
6	33.238	3/10	3/7	0.13
7	40.923	4/10	3/7	0.03
8	59.081	4/10	4/7	0.17
9	69.359	5/10	4/7	0.07
10	71.948	5/10	5/7	0.21
11	80.269	6/10	5/7	0.11
12	98.723	6/10	6/7	0.26
13	114.070	7/10	6/7	0.16
14	130.811	8/10	6/7	0.06
15	170.875	9/10	6/7	0.04
16	185.169	10/10	6/7	0.14
17	202.553	10/10	7/7	0.00

Table 2 is the tableau of a Two-Sample Kolmogorov-Smirnov (K-S) GoF test, that uses the maximum of the Absolute Differences (0.33). K-S will be the topic of another paper. Here we will just assess the samples qualitatively by plotting (and regressing) F1 vs. F2 and by plotting the time series of these differences (Figure 8).

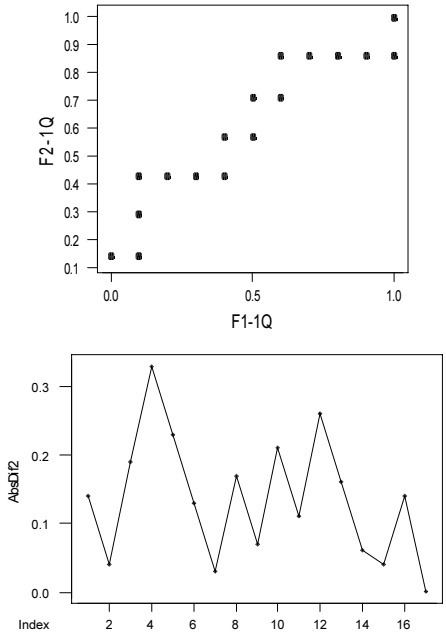

Figure 8. Plot of F2 vs. F1 (Rel. Frequencies) and the Series of Their Absolute Differences

In either plot, discrepancies between the two distribution functions are not large: F1 vs. F2 Plot remains close to the diagonal; and Max of absolute differences, 0.31, is not large for such small sample size. This lends support to assume that the two distributions are similar (in fact, both sets are Exponential with mean

100). We then proceed to implement the linear regression of the two frequencies (F1, F2) and obtain the results:

F2 = 0.222 + 0.778 F1

Predictor	Coef	StDev
Constant	0.22167	0.04262
F1	0.77825	0.07400

R-Sq = 88.1%

Notice how the regression coefficient (0.778) is close to unity and its approximate 99% CI (0.556, 1.00025) covers unity. Just as in the previous case, we conclude that, the performance of two devices that have identically distributed lives, must be equivalent.

Practical Example

In Table 3 we show two data sets, each comprised of the lives of 30 ball bearings. The bearings correspond to two different manufacturers and we want to assess whether their lives are equivalent, or not. We use the approach presented above, to do this.

Table 3. Ball Bearing Data

Row	Manufacturer A	Manufacturer B
1	24120	21488
2	26912	27457
3	34990	28504
4	42960	34755
5	49983	36060
6	50802	38591
7	52776	49274
8	55821	50319
9	57584	54215
10	63092	58773
11	67395	61264
12	72380	61391
13	76403	62051
14	77488	64603
15	77895	65459
16	79237	66767
17	90281	67277
18	91357	69991
19	95504	73721
20	96382	74523
21	97896	78817
22	112249	79351
23	112319	79553
24	122917	81159
25	130287	85262
26	133941	86695
27	137000	97396
28	148427	114649
29	153284	116291
30	206365	117350

First, we implement the Q-Q Plot for these two sorted sets of ball bearing lives (see Figure 9).

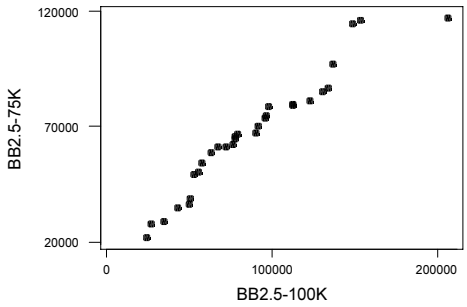

Figure 9. Q-Q Plot for the Lives of Two Sets of Ball Bearing From Different Manufacturers

We immediately visualize that the ranges are not the same (0:200K vs. 0:120K). Even if we trim the last pair of values, the ranges remain different (0:180K vs. 0:120K). We, therefore, implement the linear regression of the Q-Q Plot sorted values:

Mfg. A = 15861 + 0.579 Mfg. B

Predictor	Coef	StDev
Constant	15861	2929
Mfg. B	0.57890	0.03014

R-Sq = 92.9%

Notice how the regression coefficient (0.5789) is less than unity and its approximate 99% CI (0.489, 0.669) does not cover unity. All indications point toward these two data sets coming from two different distributions. To verify this, we next obtain the descriptive statistics of these two data sets:

	Variable							
	N	Mean	Median	StDev	Min	Max	Q1	Q3
Mfg. A	30	66767	66113	25245	21488	117350	50058	79955
Mfg. B	30	87935	78566	42042	24120	206365	55060	114969

Notice how the ball bearings from Manufacturer A do have lower mean, median, min, and max values than those from Manufacturer B ball bearings. Quartiles Q1 and Q3 (lives of 25% and 75% percentiles of the bearings) are also lower than those of Manufacturer B's bearings. In conclusion, Manufacturer A's ball bearings have a different (obviously lower) performance than those from Manufacturer B, as shown by the difference in distribution (and in descriptive statistics). Hence, if longer lives is what we are looking for, the second manufacturer looks more promising.

Conclusions

In this START sheet we discussed the complex, challenging and interesting statistical problem of graphically comparing whether two populations are equal. We do so, by examining Q-Q Plots from two samples drawn from them. Such comparisons are examples of EDA, or Exploratory Data Analysis. The problem may require additional analyses, for Q-Q Plots are rather qualitative tools. However, EDA remains a valid and easy to apply tool for the initial analysis that helps assess whether further efforts are worthwhile or

even necessary. We have developed examples of such graphical analyses, when the sample sizes are large as well as small. And we have provided references for further readings on this topic.

A final note: some authors have used tolerance limits to assess whether two samples come or not from the same Normal distribution (tolerance intervals are discussed in another forthcoming START sheet). If two distributions are equal, then tolerance limits closely overlap. But this does not provide an overwhelming evidence, by itself (maybe they just have an overlapping distribution body but differ in the tails, as in example of Figure 1). In addition, when the two populations are not Normal, it is not always easy to obtain their tolerance intervals.

Bibliography

1. Probability and Statistics for Engineers and Scientists. Walpole and Myers. Prentice Hall, NJ, 1988.
2. Quality Control and Industrial Statistics, Duncan.,Richard Irwin, Inc., Ill, 1974.
3. A Practical Guide to Statistical Analysis of Material Property Data, Romeu, J.L. and C. Grethlein, AMPTIAC, 2000.
4. Statistical Assumptions of an Exponential Distribution, RAC START, Volume 9, Number 2.
5. Statistical Confidence, RAC START, Volume 9, Number 4.
6. Practical Statistical Tools for Reliability Engineers, Coppola, A., RAC, 1999.

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 Enterprises.

For his work in education and research and for his publications and presentations, Dr. Romeu has been elected 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 has received several international grants and awards, including a Fulbright Senior Lectureship and a Speaker Specialist Grant from the Department of State, in Mexico. He has extensive experience in international assignments in Spain and Latin America and is fluent in Spanish, English, and French.

Romeu is a senior technical advisor for reliability and advanced information technology research with IIT Research Institute (IITRI). Since joining IITRI 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.iitri.org/DATA/START>.

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.iitri.org>

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). Since its inception in 1968, the RAC has been operated by IIT Research Institute (IITRI). Technical management of the RAC is provided by the U.S. Air Force's Research Laboratory Information Directorate (formerly Rome Laboratory).