START
Selected Topics in Assurance Related Technologies
Volume 12, Number 2

Understanding Binomial Sequential Testing

Table of Contents

Introduction

Determining the sample size "n" required in testing and confidence interval (CI) derivation is of importance for practitioners, as evidenced by the many related queries that we receive at the RAC in this area. Therefore, we have addressed this topic in a number of START sheets. For example, we discussed the problem of calculating the sample size for deriving a general CI in Reference 1, and for the case of acceptance testing, in Reference 2. The problem of censored sampling has been treated in Reference 3. These cases, however, only treat the situation where fixed samples of predetermined sizes are taken, all at one time.

But sampling is both expensive and time consuming. Hence, there are situations where it is more efficient to take samples sequentially, as opposed to all at one time, and to define a stopping rule to terminate the sampling process. The case where the entire sample is drawn at one instance is known as "single sampling." The case where samples are taken in successive stages, according to the results obtained from the previous samplings, is known as "multiple sampling."

Taking samples sequentially and assessing their results at each stage allows the possibility of stopping the process and reaching an early decision. If the situation is clearly favorable or unfavorable (for example, if the sample shows that a widget's quality is definitely good or poor), then terminating the process early saves time and resources. Only in the case where the data is ambiguous do we continue sampling. Only then, do we require additional information to take a better decision.

The preceding discussion justifies the need to overview multi-stage sampling plans, and we will do so in a sequence of two START sheets. In this first START sheet, we start by exploring double sampling plans. From there, we proceed to higher dimension sampling plans, namely sequential tests. We illustrate their discussion via numerical and practical examples of sequential tests for attributes (pass/fail) data that follow the Binomial distribution. Such plans can be used for Quality Control as well as for Life Testing problems. We conclude with a discussion of the ASN or "average sample number," a performance measure widely used to assess such multi-stage sampling plans. In the second START sheet, we will discuss sequential testing plans for continuous data (variables), following the same scheme used herein.

Double Sampling (Two Stage) Testing Procedures

In hypothesis testing, we often define a Null (H_0) that expresses the desired value for the parameter under test (e.g., acceptable quality). Then, we define the Alternative (H_1) as the unacceptable value for such parameter. We take a sample of pre-determined size "n" and, based upon the result obtained from drawing such a single fixed-size sample, we make a decision regarding these two hypotheses. This process is a single stage sampling procedure.

In a two-stage testing procedure (double sampling) one first draws a sample of size n_1 and compares the number "X" of items "of interest" (e.g., "non compliant") with two integers: c_1, c_2. If $X < c_1$, we accept the null hypotheses (H_0) that the batch is of acceptable quality; and if $X > c_2$, we accept the alternative hypotheses (H_1) that the batch quality is unacceptable. However, if $c_1 \leq X \leq c_2$, we draw a second sample of size n_2 and compare the total number X' (of items "of interest" found in the combined sample of size $n_1 + n_2$) with a third integer c_3 (which may be the same as c_2). We take the final decision based upon whether such combined X' is greater than, or equal to c_3. We thus describe double sampling plans via the integers n_1, n_2, c_1, c_2, and c_3, and denote them S (n_1, n_2, c_1, c_2, c_3).

The logic behind double sampling schemes is that, if things are clearly good or bad ($X < c_1$ or $X > c_2$) then we make a decision based on the first sample only. If there are some doubts ($c_1 \leq X \leq c_2$), then we draw a second sample, collecting additional information before reaching a decision. This method lowers the risk of making the wrong decision, when initially things do

not appear to be so clear, at the cost of a longer and more expensive process (i.e., drawing the second sample).

Consider the following example. Assume that we have a device with mission time "T" (say, 20 hours), which requires (H_0) a reliability of say, 0.9, with confidence 0.95. To test this hypothesis, we place "n" (say 20) items on test for the "T" hours, and then count how many items survive. The number of survivals "X" is distributed as a Binomial (n = 20, p), where "p" is the reliability of the device; that is, the probability that such device survives beyond mission time "T." We then express such a Binomial probability model as:

$$P\{X > c\} = 1 - P\{X \le c\} = 1 - \sum_{x=1}^{c} C_x^n \, p^x \, (1-p)^{n-x}$$

Now, assume that a reliability of 0.8 or less is unacceptable (H_1) and define the double sampling Plan S ($n_1 = 20$, $n_2 = 20$, $c_1 = 14$, $c_2 = 15$, $c_3 = 33$), as described. We draw a first sample of size $n_1 = 20$ and count the number of survivals "X." If $X > 15$, we don't reject H_0. If $X < 14$ we reject H_0 (reliability is unacceptable). If X is 14 or 15, we draw a second sample of size $n_2 = 20$ and count the number of survivals (Y). Then, only if $X + Y < 33$ we reject H_0 (and decide that the device reliability is unacceptable).

Plan S ($n_1 = 20$, $n_2 = 20$, $c_1 = 14$, $c_2 = 15$, $c_3 = 33$) is constructed in the following way. For the first sample (n_1) we selected, from the Binomial (n = 20, p = 0.9) tables, number "c_2" = 15, because the probability of acceptance (survivals $X > 15$) of a "good" batch (reliability ≥ 0.9) is $P_{0.9}\{X \ge 16\} = 0.957$. We then selected number "c_1" = 14 because the probability of rejecting a "good" batch $P_{0.9}\{X \le 13\} = 0.002$. Then, from the Binomial (n = 20, p = 0.8) table, the probability of "accepting a bad batch" (reliability 0.8 and $X \ge 16$) is $1 - P_{0.8}\{X \le 15\} = 1 - 0.3704 = 0.630$, and that of "rejecting a bad batch" is $P_{0.8}\{X \le 13\} = 0.087$. If the number of survivals is X = 14 or 15, results are considered inconclusive. The probability of an inconclusive result, when the batch is good, is 0.041, and when the batch is bad, is 0.284. If so, we take a second sample of $n_2 = 20$ and define number $c_3 = 33$. We accept the batch if the total survivals are 33 or more; and reject it if the survivals are less than 33. We don't claim that plan S is optimal; but it provides a good illustration of the construction approach.

The probability of acceptance for such double sampling plan S, for any p, is given by the following equation.

$$\begin{aligned} &P\{\text{Accepting Batch}\} = P\{\text{Accepting Initially}\} \\ &+ P\{\text{Initially Inconclusive Then Accept at 2nd}\} \\ &= P\{\text{First Successes} \ge 16\} + P\{\text{First Successes} = 14 \\ &\text{or 15 and Combined Successes} \ge 33\} \\ &= \sum_{x=16}^{20} \text{Bin}(x; n = 20, p) + \text{Bin}(x = 14; n = 20, p) \end{aligned}$$

Table 1. Individual and Cum. Probs. for Binomial (n = 20; p = 0.9 and 0.8)

Surv.	P = 0.9	P = 0.8	Cum. For P = 9	Cum. For P = 8
10	0.000006	0.002031	0.00001	0.00259
11	0.000053	0.007387	0.00006	0.00998
12	0.000356	0.022161	0.00042	0.03214
13	**0.001970**	**0.054550**	**0.00239**	**0.08669**
14	0.008867	0.109100	0.01125	0.19579
15	0.031921	0.174560	0.04317	0.37035
16	**0.089779**	**0.218199**	**0.13295**	**0.58855**
17	0.190120	0.205364	0.32307	0.79392
18	0.285180	0.136909	0.60825	0.93082
19	0.270170	0.057646	0.87842	0.98847
20	0.121577	0.011529	1.00000	1.00000

$$\begin{aligned} &\text{x [Bin(x = 19; n = 20, p) + Bin(x = 20; n = 20, p)]} \\ &\text{+ Bin(x = 15; n = 20, p) x [Bin(x = 18; n = 20, p)]} \\ &\text{+ Bin(x = 19; n = 20, p) + Bin(x = 20; n = 20, p)} \end{aligned}$$

Notice that, to obtain a result of 33 survivals or more, in the combined first and second samples, we have to obtain 14 or 15 successes (inconclusive results) in the first sample and then, enough successes in the second (say 18, 19, or 20) to add up to 33 or more. For the case where reliability p = 0.9, the double sampling plan acceptance is 0.982, instead of just 0.957 for a fixed sample test with n = 20 and c = 16:

$$\begin{aligned} &P\{\text{Accepting Batch}\} = P\{\text{Accepting Initially}\} \\ &+ P\{\text{Initially Inconclusive Then Accept at 2nd}\} \\ &= 0.957 + 0.0089 \text{ x } (0.2701 + 0.1215) + 0.0319 \\ &\text{x } (0.2851 + 0.2701 + 0.1215) = 0.982 \end{aligned}$$

The probability of rejection of our double sampling plan S is obtained just by substituting "acceptance" for "rejection," in the equations. For a true reliability p = 0.09, the probability of incorrect rejection is 0.018, instead of just $P\{X < 16\} = 1 - 0.957 = 0.043$, which is the corresponding probability for a fixed sample plan with n = 20 and c = 16:

$$\begin{aligned} &P\{\text{Rejecting Batch}\} = P\{\text{Rejecting Initially}\} \\ &+ P\{\text{Initially Inconclusive Then Reject at 2nd}\} \\ &= P\{\text{First Successes} \le 13\} + P\{\text{First Successes} \\ &= 14 \text{ or } 15 \text{ and Combined Successes} < 33\} \\ &= \sum_{x=0}^{13} \text{Bin}(X; n = 20, p) + \text{Bin}(x = 14; n = 20, p) \end{aligned}$$

$$\begin{aligned} &\text{x } \{1 - [\text{Bin}(x = 19; n = 20, p) + \text{Bin}(x = 20; n = 20, p)]\} \\ &+ \text{Bin}(x = 15; n = 20, p) \text{ x } \{1 - [\text{Bin}(x = 18; n = 20, p) \\ &+ \text{Bin}(x = 19; n = 20, p) + \text{Bin}(x = 20; n = 20, p)]\} \\ &= 0.0024 + 0.0089 \text{ x } (1 - (0.2701 + 0.1215)) \\ &+ 0.0319 \text{ x } (1 - (0.2851 + 0.2701 + 0.1215)) = 0.018 \end{aligned}$$

The preceding shows how the double sampling scheme, when compared with a single sample test, not only increases the probability of accepting a good batch but also reduces the probability of rejecting a good one, even if the initial test results are inconclusive. This characteristic is their strongest advantage that, in many cases, far outweighs the extra cost and effort involved in implementing such double sampling schemes.

The Sequential Probability Ratio Test (SPRT)

Assume now that we want to test that the acceptable mean life of a device (MTTF) is 200 hours or more (null hypothesis H_0) versus that it is 100 hours or less (H_1). If the device lives are distributed Exponential (with mean μ) the distribution (CDF) and density (pdf) of the random variable (r.v.) "life of a device," denoted "X," under H_i: μ_i, for i = 0, 1 are:

$$F_{\mu_i}(X) = 1 - e^{\frac{-X}{\mu_i}}; f(X) = \frac{1}{\mu_i} e^{\frac{-X}{\mu_i}}; i = 0.1; \mu_0 = 200; \mu_1 = 100$$

Assume that we place such devices, sequentially on test, one at a time, for a test duration of T = 20 hours, after which we assess whether each device is still working (Pass) or not (Fail). The probabilities of such results, under the two above hypotheses H_i, for i = 0, 1 are:

$$p_i = P\{\text{Success Under } H_i\} = P_{\mu_i}\{\text{Device Outlives } T = 20\}$$

$$= P_{\mu_i}\{X > 20\} = e^{\frac{-20}{\mu_i}}; i = 0, 1$$

$$P\{\text{Success Under } H_0\} = e^{\frac{-20}{200}} = e^{-0.1} = 0.905$$

$$P\{\text{Success Under } H_1\} = e^{\frac{-20}{100}} = e^{-0.2} = 0.819$$

We can transform these relationships into the respective equivalent hypotheses: H_0: $p_0 = 0.905$ and H_1: $p_1 = 0.819$. In such a setting, the test of each device for 20 hours is assumed to be an independent trial, with identical probability of success p_i. Therefore "y," the cumulative number of successes, out of "n" trials, is distributed as a Binomial (n, p_i), for i = 0, 1.

Define the Probability Ratio (PR) as that of the Binomial distributions, under H_0 and H_1:

$$\frac{P\{\text{"y" Successes Under } H_1\}}{P\{\text{"y" Successes Under } H_0\}} = \frac{\text{Binomial}(y; n, p_1)}{\text{Binomial}(y; n, p_0)}$$

$$= \frac{Kp_1^y (1-p_1)^{n-y}}{Kp_0^y (1-p_0)^{n-y}} = \frac{p_1^y (1-p_1)^{n-y}}{p_0^y (1-p_0)^{n-y}}$$

K is the number of "SPRT-feasible" ways that one can obtain "y" successes out of "n" trials. Now, define the two hypothesis test errors: α (producer's risk, or probability of rejecting a device with acceptable life) and β (consumer's risk, or probability of accepting a device with unacceptable life) and let $\alpha = \beta = 0.128$. Then, we can find two values A and B such that, at any stage "n," that is having tested "n" devices sequentially (one at a time) and having obtained "y" cumulative successes, the PR fulfills the probability:

$$P\{PR > A\} = P\left\{\frac{p_1^y (1-p_1)^{n-y}}{p_0^y (1-p_0)^{n-y}} > A\right\} = \beta$$

$$P\{PR < B\} = P\left\{\frac{p_1^y (1-p_1)^{n-y}}{p_0^y (1-p_0)^{n-y}} < B\right\} = 1 - \alpha$$

Thus, we define S (B,A), the "Sequential Probability Ratio Test" (SPRT), by the preceding equations, as one that compares PR with values A and B at every stage "n", and decides: (i) to accept H_0 if PR < B; (ii) accept H_1 if PR > A; or (iii) continue testing if B < PR < A.

For example, at any given stage "n," say after placing the tenth (n = 10) device on test, we obtain "y" successes (say y = 6). We thus obtain the Binomial SPRT test result:

$$\frac{P\{y = 6 \text{ Successes Under } H_1\}}{P\{y = 6 \text{ Successes Under } H_0\}} = \frac{\text{Binomial}(6; 10, p_1)}{\text{Binomial}(6; 10, p_0)}$$

$$= \frac{0.819^6 (1-0.819)^{10-6}}{0.905^6 (1-0.901)^{10-6}} = \frac{3.24E-4}{4.47E-5} = 7.23$$

Then, we compare value 7.23 with adequate values for A and B and decide: (i) to stop testing and accept (H_1) that probability of success is 0.819, and hence the Exponential mean is 100 hours or less, if the PR value 7.23 is greater than A; (ii) to stop and accept (H_0) that the probability of success is 0.905, and hence the Exponential mean is 200 hours or more, if the PR value 7.23 is smaller than B; or (iii) to take another sample and repeat the the process, if 7.23 is between the values of B and A.

We can simplify the process and equations, by taking the Logarithms in the PR inequality below, which defines the region leading to the continuation of the test:

$$B < \frac{p_1^y (1-p_1)^{n-y}}{p_0^y (1-p_0)^{n-y}} = \left(\frac{p_1}{p_0}\right)^y \left(\frac{1-p_1}{1-p_0}\right)^{n-y} < A$$

The result produces a linear equation that is a function of the number of successes "y", out of a given number of trials (stage) "n", and is bounded by the Logarithms of values A and B:

$$\ln(B) < y \ln\left(\frac{p_1}{p_0}\right) + (n - y) \ln\left(\frac{1 - p_1}{1 - p_0}\right)$$

$$= n \ln\left(\frac{1 - p_1}{1 - p_0}\right) + y\left\{\ln\left(\frac{p_1}{p_0}\right) - \ln\left(\frac{1 - p_1}{1 - p_0}\right)\right\} < \ln(A)$$

$$\ln(B) < an + by < \ln(A); \text{ with } : a = \ln\left(\frac{1 - p_1}{1 - p_0}\right)$$

$$b = \left\{\ln\left(\frac{p_1}{p_0}\right) - \ln\left(\frac{1 - p_1}{1 - p_0}\right)\right\}$$

The coefficients "a" and "b" of these equations are, obviously, functions of p_i, i = 0, 1. It can be shown (References 4, 5, and 6) that the constants A and B are approximated by:

$$A \cong \frac{(1 - \beta)}{\alpha}; B \cong \frac{\beta}{1 - \alpha}$$

In our example: $p_0 = 0.905$, $p_1 = 0.819$, n =10, y = 6 and $\alpha = \beta = 0.128$. Hence, the SPRT coefficients "a" and "b" can be calculated and values "A" and "B" can be approximated:

$$a = \ln\left(\frac{1 - p_1}{1 - p_0}\right) = \ln\frac{1 - 0.819}{1 - 0.905} = \ln(1.906) = 0.6446;$$

$$b = \left\{\ln\left(\frac{p_1}{p_0}\right) - \ln\left(\frac{1 - p_1}{1 - p_0}\right)\right\} = \ln\frac{0.819}{0.905} - \ln\frac{1 - 0.819}{1 - 0.905}$$

$$= -0.0998 - 0.6446 = -0.7444;$$

$$A \cong \frac{(1 - \beta)}{\alpha} = \frac{(1 - 0.128)}{0.128} = 6.813; \Rightarrow \ln(A) = \ln(6.813) = 1.919$$

$$B \cong \frac{\beta}{1 - \alpha} = \frac{0.128}{1 - 0.128} = 0.147; \Rightarrow \ln(B) = \ln(0.147) = -1.919$$

For our example, at the SPRT 10th stage (or n = 10 trials), with y = 6 successes, we get:

$$\ln(B) = -1.919 < 0.6446n - 0.7444y < \ln(A) = 1.919$$

But: 0.64446 x 10 - 0.7444 x 6 = 1.979 > 1.919

Since 1.979 is above bound 1.919 (PR > A) we stop testing and accept H_1 that $p_1 = 0.819$ (hence, that device MTTF $\leq$ 100 hours). Had the equation 0.6446n - 0.7444y yielded a value below -1.919, we would have accepted H_0 that $p_0 = 0.905$ (and hence, that device MTTF $\geq$ 200 hours). Had the equation results been between -1.919 and 1.919 (the continuation region), we would have proceeded with the sampling. Figure 1 represents these choices.

The described procedure, even when accurate and correct, is difficult to follow. To better track the SPRT test values, we need an equation comparing "y," the number of "successes," directly with stage boundaries that are a function of the number of trials "n":

$$\ln(B) < an + by < \ln(A)$$

$$\frac{\ln(B)}{b} - \frac{a}{b}n = h_1 + sn > y > \frac{\ln(A)}{b} - \frac{a}{b}n = h_0 + sn$$

For our current example the values are: ln(B) = -1.919; ln(A) = 1.919; b = -0.7444; a = 0.6446.

$$h_1 + sn = \frac{-1.919}{-0.7444} - \frac{0.6446}{-0.7444}n > y > \frac{1.919}{-0.7444} - \frac{0.6446}{-0.7444}n = h_0 + sn$$

$$r_n = h_1 + sn = 2.578 + 0.866n > y > -2.578 + 0.866n = h_0 + sn = a_n$$

By letting "n," the number of "stages," run from 1, 2, …, we obtain (a_n; r_n) the SPRT decision "boundaries" (or acceptance and rejection numbers) for our example are shown in Figure 2.

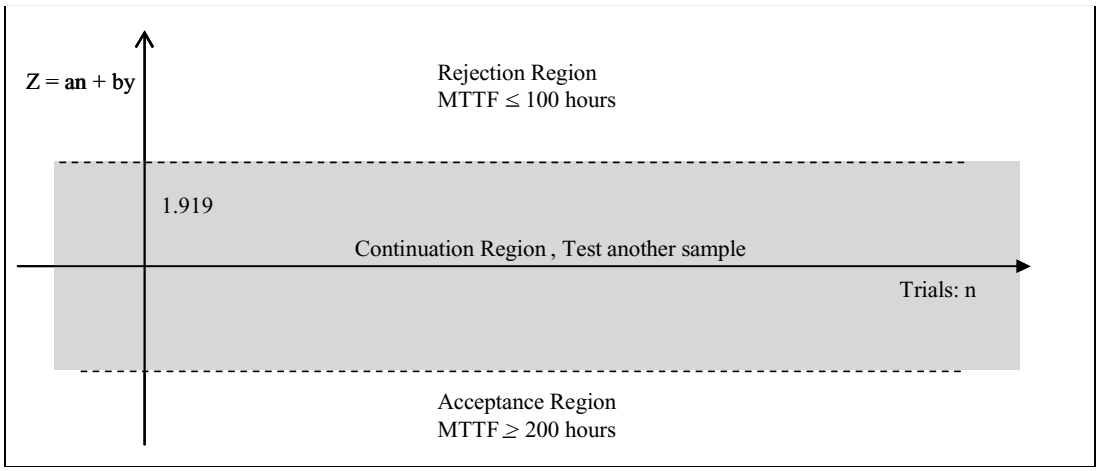

Figure 1. Representation of the SPRT Test for Horizontal Regions

Row	Stage	Reject	Accept
1	1	-1.712	3.444
2	2	-0.846	4.310
3	3	0.020	5.176
4	**4**	**0.886**	**6.042**
5	5	1.752	6.908
6	6	2.618	7.774
7	7	3.484	8.640
8	8	4.350	9.506
9	9	5.216	10.372
10	10	6.082	11.238
11	11	6.948	12.104
12	12	7.814	12.970
13	13	8.680	13.836
14	14	9.546	14.702
15	15	10.412	15.568
16	16	11.278	16.434
17	17	12.144	17.300
18	18	13.010	18.166
19	19	13.876	19.032
20	**20**	**14.742**	**19.898**
21	21	15.608	20.764
22	22	16.474	21.630
23	23	17.340	22.496
24	24	18.206	23.362
25	25	19.072	24.228

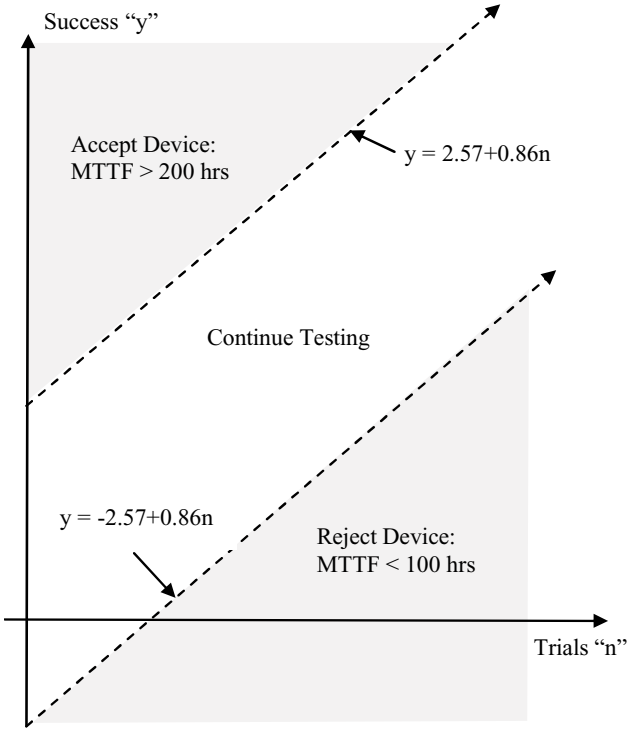

Figure 2. Representation of SPRT

Notice how it is impossible to reject the device (accept H_1: $p_1 = 0.819$, or MTTF $\leq$ 100 hrs.) until having tested at least n = 4 devices, all of which must have failed. It is also impossible to accept the device (H_0: $p_0 = 0.905$, or MTTF $\geq$ 200 hrs.) until having tested at least n = 20 devices, without observing a single failure (these critical boundary values are "bolded" in Figure 2). At any stage, the probability of decision error is $\alpha = \beta = 0.128$.

In addition, notice how the boundary equation parameters (slope and intercept) depend on SPRT errors α and β, and on probabilities p_0, p_1, which in turn depend on the device MTTFs:

$$\text{Intercepts}: h_1 = \frac{\ln(B)}{b} = \frac{\ln\left\{\frac{\beta}{1-\alpha}\right\}}{\left\{\ln\left(\frac{p_1}{p_0}\right) - \ln\left(\frac{1-p_1}{1-p_0}\right)\right\}}; h_0 = \frac{\ln(A)}{b}$$

$$= \frac{\ln\left\{\frac{(1-\beta)}{\alpha}\right\}}{\left\{\ln\left(\frac{p_1}{p_0}\right) - \ln\left(\frac{1-p_1}{1-p_0}\right)\right\}}$$

$$\text{Slope}: s = -\frac{a}{b} = -\frac{\ln\left(\frac{1-p_1}{1-p_0}\right)}{\left\{\ln\left(\frac{p_1}{p_0}\right) - \ln\left(\frac{1-p_1}{1-p_0}\right)\right\}} > 0$$

Notice how, in the preceding equations for the slope (s) and intercepts (h_1, h_0), the smaller the errors (α, β), the larger the intercepts (in absolute value). This means that, on the average, it will take the SPRT test longer (more stages) to arrive its final decision. This result is intuitive. Since we are demanding larger assurances (smaller errors) from the SPRT test procedure, the SPRT will necessarily require more information (more stages) to be able to provide a decision that fulfills such errors α, β. Some examples are shown in Table 2.

Table 2. Intercepts for the Lower/Upper Increasing Bounds of SPRT

	Lower Bound Intercept			Upper Bound Intercept		
Beta/Alpha	0.01	0.05	0.128	0.01	0.05	0.128
0.01	-6.1723	-4.0105	-2.7478	6.1723	6.1169	6.0019
0.05	-6.1169	-3.9551	-2.6924	4.0105	3.9551	3.8400
0.128	-6.0019	-3.8400	-2.5773	2.7478	2.6924	2.5773

The slope s, which is common to both boundary lines, depends on the two hypothesized probabilities p_0 and p_1, which in turn depend on the hypothesized MTTF μ_0 and μ_1 (but not on the aforementioned test errors α, β). The "continuation region," defined by the slope s and the intercepts h_1 and h_0, characterize the speed at which decisions are taken at every step. So, all other factors being equal, the further apart the probabilities p_0, p_1 (and hence the two MTTFs μ_0, μ_1), the more comfortably we can discriminate between them as shown in Table 3.

Table 3. Discrimination Between Two MTTFs: Slope of the Increasing SPRT Boundaries (slopes in italics)

p_1	p_0		
	0.99	0.95	0.905
0.9	*3.0929*	*0.9311*	*0.0689*
0.819	*3.8899*	*1.7280*	*0.8659*
0.7	*4.5686*	*2.4068*	*1.5446*

Finally, the preceding can be similarly applied to an SPC/Quality Control problem. For example, assume that we are interested in assessing a batch of incoming material and that the acceptable level of quality (AQL) is defined by some maximum percent of defectives, say 10% ($p \leq 0.1$). Also, assume another percent defective defines an unacceptable quality (LTPD), beyond which we will not receive the batch (say 20% defectives or more). Then, define values $p_0 = 1 - 0.1 = 0.9$, and $p_1 = 1 - 0.2 = 0.8$, for hypotheses H_0 and H_1.

Now assume that for procedural ease, cost or other practical reason, it is decided to test the lot by taking each item sequentially, one at a time, instead of taking a single sample of fixed and predetermined size "n", all at one time.

Then, for the acceptance sampling problem described, all the previously described SPRT derivations and results are applicable, with the pertinent modifications.

The Average Sample Number (ASN)

The main advantage of multiple stage sampling plans is the reduced "long run" or average sample size, required to arrive to a good decision. For now, the random variable "sample size" is a probabilistic outcome (varies with every case). Its "Expected Value," known as ASN or "Average Sample Number," depends on the value of the real parameter under test, be it the percent defective "p," reliability, or any other parameter of interest.

The ASN is obtained following the definition of Expected Value. For double sampling:

$$ASN = E\{SN\} = \sum_{SN} SN \times P\{SN\}$$

$$= n_1 \times P(n_1) + (n_1 + n_2) \times P(n_1 + n_2)$$

In the double sampling scheme, SN (sample number) can be only n_1 or $n_1 + n_2$. $P(n_1)$ is the probability of drawing a "first" sample only, which occurs when arriving at a decision at the first sample (with probability $1 - P\{c_1 \leq Y \leq c_2\}$). The probability $P\{n_1 + n_2\}$ of having to draw a second sample, totaling a size of $n_1 + n_2$, occurs when we had an "inconclusive" outcome from the first sample (i.e., with probability: $P\{c_1 \leq Y \leq c_2\}$).

We illustrate this case using our double sampling example S ($n_1 = 20$, $n_2 = 20$, $c_1 = 14$, $c_2 = 15$, $c_3 = 33$), described earlier. Let the true reliability "p," be $p = 0.9$, and let Y be the number of survivals obtained in the first sample of size $n_1 = 20$. Then, the probability of taking "no decision" on the first sample, when $p = 0.9$, is $P(c_1 \leq Y \leq c_2) = P(14 \leq Y \leq 15) = 0.0089 + 0.0319$. This yields ASN = 20.81, barely larger than the exact n = 20 elements that would be required by a single sample plan, of fixed size $n_1 = 20$:

$$ASN = n_1 \times P(n_1) + n_2 \times P(n_2)$$

$$= 20 \times [1 - (0.0089 + 0.0319)] + 40 \times (0.0089 + 0.0319)$$

Table 4 presents the ASN values for the double sampling plan S, described in the first section, calculated at selected values of the (reliability) parameter "p."

Table 4. Comparison of ASN for Double Sampling, Given "p"

Reliability	0.95	0.9	0.85	0.8	0.75	0.7	0.65	0.6
ASN	20.05	20.82	22.96	25.67	27.42	27.41	25.96	23.98

In Figure 3, we show graphically the relationship between the reliability parameter "p" and the corresponding double sample ASN.

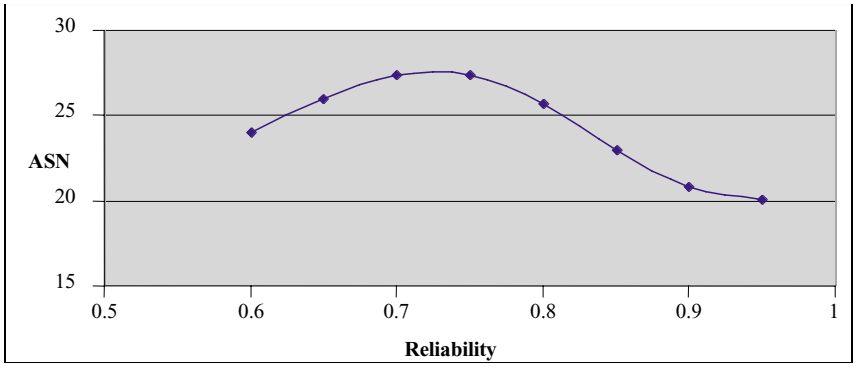

Figure 3. ASN for Double Sampling

In the general case of sequential tests (SPRTs), the ASN is obtained following the same principles described for the double sampling scheme. However, the stages "n" are extended to any number, not just two. We multiply each SPRT stage (the samples taken so far) times the probability of arriving to a decision at that stage, given the true parameter "p" and that we have not taken a decision earlier in the sequential test (i.e., that we have followed a path inside the SPRT "continuation region", up to the present stage):

$$ASN(p) = E_p\{SN\}$$

$$= \sum_{SN \geq 1} SN \times P_p\{\text{Decision at Stage SN But not before}\}$$

The ASN values for sequential (SPRT) tests, however, are in general not easy to obtain. A full treatment of this topic falls beyond the scope of this START sheet. Those readers interested in pursuing it further may want to consult References 4, 5, 6, and 7. In the second START of this two-part series, discussing SPRTs for continuous variables, we will obtain the ASN for a reduced number of values of the parameter of interest.

Conclusions

Multiple stage sampling plans can save substantial time and resources, when the case is clear-cut. Multiple stage plans provide an alternative to fixed sample plans that can help diminish the producer and consumer risks (α, β) of arriving at a wrong decision. Sequential (SPRTs) tests constitute a natural extension of the double sampling plans.

In this first START sheet, we have overviewed SPRTs for the discrete case, via the development of sequential tests for the Binomial distribution. Such an approach is useful in life testing, when devices are sequentially placed on test for a fixed time "T," and when only the number "Y" of survivals/failures from such tests is recorded. Binomial sequential tests are also useful in SPC/Quality Control acceptance sampling problems, when batch items are tested, one at a time, on a pass/fail basis, to assess the percent of defective items in a lot, instead of by taking a fixed size sample, one at a time. We have also discussed the double sampling plans in detail, and the problem of assessing these plans via the ASN.

In a second, follow-up START sheet, we will discuss the use of SPRT tests with continuous variables, specifically, when testing Exponential lives. Such tests are widely used and are included in the well-known MIL-HDBK-781-D, Reliability Testing for Engineering Development. We will use the same example of this paper and compare, using the corresponding ASN, the efficiency of the two strategies (1) of letting each device extinguish its life, versus that of (2) curtailing the test at a fixed time T, used here, and only observing whether such device passed or failed the test.

For More Information

1. Romeu, J.L., Statistical Confidence, RAC START, Volume 9, Number 4, <http://rac.alionscience.com/pdf/STAT_CO.pdf>.
2. Romeu, J.L., OC Function and Acceptance Testing, RAC START, Volume 12, Number 1. <http://rac.alionscience.com/pdf/OC_Curves.pdf>.
3. Romeu, J.L., Censored Data, RAC START, Volume 11, Number 3, <http://rac.alionscience.com/pdf/CENSOR.pdf>.
4. Duncan, A.J., Quality Control and Industrial Statistics, Fourth Edition, Irwin, IL 1974.
5. MIL-HDBK-781D and MIL-STD-781D, Reliability Testing for Engineering Development, Qualification, and Production, DoD, October 1986.
6. Wald, A. (1947), Sequential Analysis, John Wiley & Sons, New York, NY.
7. Methods for Statistical Analysis of Reliability and Life Data, Mann, N., R. Schafer, and N. Singpurwalla, John Wiley, NY, 1974.

About the Author

Dr. Jorge Luis Romeu has over thirty years of statistical and operations research experience in consulting, research, and teaching. He is a Certified Reliability Engineer (CRE). 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 Alion Science and Technology previously IIT Research Institute (IITRI). Since rejoining 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>.

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 world-wide 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, electronics, 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. Alion, and its predecessor company IIT Research Institute, have operated the RAC continuously since its creation in 1968.