**Working Paper Series**

**Núcleo de Normalização e Qualimetria (NNQ),**

 **Federal University of Santa Catarina, Brazil**

February 2016

WP16-001

LOT QUALITY ASSURANCE SAMPLING (LQAS) FROM A NEYMAN-PEARSON PERSPECTIVE

ROBERT WAYNE SAMOHYL, RETIRED,

FEDERAL UNIVERSITY OF SANTA CATARINA,

INDUSTRIAL ENGINEERING,

samohyl@yahoo.com

ABSTRACT:

The standards and procedures for LQAS, usually based on simple random sampling (SRS) plans, can be improved by taking into account the correct usage of hypothesis testing and, when necessary, the application of the hypergeometric distribution instead of the binomial and Poisson approximations. For the area of public health, both service provider and the target population are concerned with the health of the target population as measured by the percentage of healthy persons in the population, sometimes called coverage. Following the original tenets of Neyman and Pearson, the null hypothesis should be defined as the hypothesis that, if wrongly judged, causes relative suffering and other high costs. From the target population point of view therefore, the null hypothesis is that the population is unhealthy. Rejecting this null when it is true has extremely high costs for the population. In similar fashion but from the service provider point of view, the null hypothesis is that the population is healthy, because rejecting this null when it is true has extremely high costs for the service provider, who would have to repeat the health service when repetition is not necessary, drawing down its budget of scarce resources. While it has been emphasized that the primary error to the service provider is to incorrectly classify a healthy population as sickly, there is a secondary error to the service provider based on the false null that the population is actually sickly. On the other hand, the worst error that the target population could suffer is to be judged healthy when in fact the population is sickly. In the presence of the null that the population is unhealthy, a random sample may erroneously produce a sample result showing the population as healthy, which means that a type I error is committed from the viewpoint of the target population, a false negative (FN). The avoidance of this error should be a priority for the target population. Likewise the target population suffers from a secondary error classifying a healthy population as sickly which is of minor importance to the population but is nonetheless introduced into the LQAS procedures. The procedure proposed here allows that upper and lower quality limits can be assigned by each party (target population and service provider) independently, without assuming the domination (and submission) of either. Furthermore each party assigns its own values for primary and secondary risk. Examples, tables, R code, and figures are presented to facilitate understanding and practice.

KEY WORDS: Acceptance sampling, LQAS, hypergeometric, OCC, ROC, hypothesis test, R CRAN

JEL: C83

[https://drive.google.](https://drive.google.com/file/d/0Bzz0QCLFzzTQZ1dHZW1qWHJlRmc/view?usp=sharing)

The views expressed in the NNQ Working Paper Series are those of the author(s) and do not necessarily reflect those of the Federal University of Santa Catarina. Working Papers have not undergone formal review and approval. Papers are included in this series to elicit feedback and to encourage debate. Copyright belongs to the author(s). **Cite Working Papers only with author approval**.

LOT QUALITY ASSURANCE SAMPLING (LQAS) FROM A NEYMAN-PEARSON PERSPECTIVE

ROBERT WAYNE SAMOHYL, RETIRED,

FEDERAL UNIVERSITY OF SANTA CATARINA,

INDUSTRIAL ENGINEERING,

samohyl@yahoo.com

abstract:

The standards and procedures for LQAS, usually based on simple random sampling (SRS) plans, can be improved by taking into account the correct usage of hypothesis testing and when necessary the application of the hypergeometric distribution instead of the binomial and Poisson approximations. For the area of public health, both service provider and the target population are concerned with the health of the target population as measured by the percentage of healthy persons in the population, sometimes called coverage. Following the original tenets of Neyman and Pearson, the null hypothesis should be defined as the hypothesis that, if wrongly judged, causes relative suffering and other high costs. From the target population point of view therefore, the null hypothesis is that the population is unhealthy. Rejecting this null when it is true has extremely high costs for the population. In similar fashion but from the service provider point of view, the null hypothesis is that the population is healthy, because rejecting this null when it is true has extremely high costs for the service provider, who would have to repeat the health service when repetition is not necessary, drawing down its budget of scarce resources. While it has been emphasized that the primary error to the service provider is to incorrectly classify a healthy population as sickly, there is a secondary error to the service provider based on the false null that the population is actually sickly. On the other hand, the worst error that the target population could suffer is to be judged healthy when in fact the population is sickly. In the presence of the null that the population is unhealthy, a random sample may erroneously produce a sample result showing the population as healthy, which means that a type I error is committed from the viewpoint of the target population, a false negative (FN). The avoidance of this error should be a priority for the target population. Likewise the target population suffers from a secondary error classifying a healthy population as sickly which is of minor importance to the population but is nonetheless introduced into the LQAS procedures. The procedure proposed here allows that upper and lower quality limits can be assigned by each party (target population and service provider) independently, without assuming the domination (and submission) of either. Furthermore each party assigns its own values for primary and secondary risk. Examples, tables, R code, and figures are presented to facilitate understanding and practice.

KEY WORDS: Acceptance sampling, LQAS, hypergeometric, OCC, ROC, hypothesis test, R CRAN

JEL: C83

INTRODUCTION

Since the pioneering work of Dodge and Romig in the 1920’s, the risk of type I and type II error has been used universally for the construction of sampling plans for producers and consumers, or more recently for monitoring and evaluating public health programs from the point of view of the service provider and the target population. This paper evaluates and extends LQAS, and therefore acceptance sampling generally, in two ways:

First and most important, we elaborate acceptance sampling in terms of hypothesis testing rigorously following the original concepts of Neyman and Pearson (NP). It is common in the literature (for example, Shmueli, 2011 and Hund, 2014) to associate consumer and producer risk with the concepts of the probability of type II (**β)** and type I (**α)** error. We will show throughout this article that the decision making process may be impaired by this conceptual misspecification. By offering a common theoretical structure, hypothesis testing from NP can produce a better understanding of the application of sampling procedures and their results.[[1]](#footnote-1) In a series of examples, we show that sample size will necessarily be altered but that measures of risk will be more reliable. Second, by using the hypergeometric distribution to calculate the parameters of the sampling plans we avoid unnecessary use of approximations such as the binomial or Poisson distributions. We show that, under usual conditions, discrepancies can be large. The conclusion is that the hypergeometric distribution should have priority in the area of acceptance sampling. Finally, the construction of receiver operating characteristics (ROC) curves can secure a less conflicting result between the independent perspectives of the service provider and the target population.

For the practitioner who is looking for sampling plans readily available, table 3 and table 6 should contain sufficient information to proceed without reading this article. However, the usual caveats apply: beware of practice (theory) without theory (practice).

DODGE and ROMIG

Considering that the pioneering work of Dodge and Romig pre-dates the formalization of hypothesis testing by Neyman and Pearson (NP), having survived decades of academic debate and practice, the question is why bring hypothesis testing into the discussion at all? We attempt to show that if used appropriately it can offer a common structure that generalizes the application of acceptance sampling to areas other than industry and commerce, specifically public health. **Our intention is to show how simple random sampling (SRS) plans like the standards for LQAS can be improved by taking into account the correct usage of hypothesis testing and when necessary the hypergeometric distribution.**

Historically, the work of Dodge and Romig (1929) appeared before the concepts of hypothesis testing (and confidence intervals) received wide acceptance in practice. Their presentation depends exclusively on probability functions, and the interpretation of the concepts of supplier risk and consumer risk as the probability of error, some years before Neyman and Pearson (1933) offered their seminal interpretation of type I and type II error.

Dodge and Romig formally introduced inspection sampling in 1929 and in fact only from the viewpoint of the consumer who is in our case the population under study. They mention producer risk only marginally. In 1944, they emphasize even more clearly their position that consumer risk is their first priority (Dodge and Romig, 1944). ‘The first requirement for the method will therefore be in the form of a definite assurance against passing any unsatisfactory lot that is submitted for inspection. [...] For the first requirement, there must be specified at the outset a value for the lot tolerance percent defective (LTPD) as well as a limit to the probability of accepting any submitted lot of unsatisfactory quality. The latter has, for convenience, been termed the Consumer’s Risk ...’.The priority given to the consumer (population under study) will be an important ingredient for the discussion of hypothesis testing.

For the area of public health, both service provider and the target population are concerned with the health of the target population as measured by the percentage **p** of healthy persons in the population, sometimes called coverage. Values of p close to one means that the population is healthy, whereas values of p close to zero would indicate that the population is sickly[[2]](#footnote-2). In traditional LQAS and acceptance sampling in general , it is natural to assume that the service provider requires a high value for p (pu for upper) to guarantee that the population is in fact healthy. The value of pu is a minimum requirement for the service provider to define a population as healthy. Another tenet of traditional LQAS literature is that the population in question will allow for a value of p high enough that the population to feels satisfied with its health condition (pl for lower). For the target population, pl is a maximum value for defining a sickly population. Any value of p greater than pl signifies that the population passes the criteria for healthiness.

The classification rule from traditional LQAS which indicates the status of the population is relatively simple: A population is considered likely to be healthy if in a **sample of size n** the **number of healthy persons (x)** is greater than a predetermined cutoff value (d). The inequality (x > d) signifies that it is very likely that the population possesses an acceptable level of health. On the other hand, if x is less than or equal to d, (x ≤ d), then it is likely that the population is not healthy. The application of LQAS determines the values of d and n. When the hypergeometric distribution is used, the population size N is added as one of the determining parameters.

A succinct explanation of traditional LQAS procedure is offered by Hedt, et al. (2008):

“To use LQAS, health system managers need to identify two thresholds. The upper threshold coverage target (e.g., 80%), which is the proportion of the community that health workers wish to reach during a predetermined period, such as one year. The lower threshold is an unacceptably low level of coverage (e.g., 50%) that should provoke managers to identify the problem causing the failed service delivery and to resolve it with a focused investment of time and resources.”

At the moment of sampling inspection, counting **the number x of healthy persons** in the sample, the intention of the researcher is to minimize the probability of wrongly classifying the population either as healthy or as sickly. Equation (1) represents the probability of the sample indicating sickliness when, in fact, the population is healthy. As long as the sample includes only a small number of healthy persons (defined as x ≤ d), the population is judged as sickly and the authorities should intervene. Of course, the population could still be healthy even though the sample result produced the contrary.

1. P(x ≤ d / p > pu) = P(FP) -- the probability of a false positive (FP)[[3]](#footnote-3)

This illustrates the important point that the false positive depends above all on the chosen values of d and pu.

When the sample includes a large number of healthy persons (defined as x > d), the population is judged as healthy and the authorities should feel no need to intervene. However, the population could be sickly, incorrectly classified as healthy:

1. P(x > d / p ≤ pl) = P(FN) – the probability of a false negative (FN)

Dodge and Romig labeled equation (2) as consumer risk since acquiring bad inputs would complicate assembly lines or retail with low quality merchandise. For health services, this is target population risk: being classified as healthy when in fact the population is sickly, the false negative (FN). In traditional LQAS, equation (2) is often referred to as **β**, the probability of type 2 error. Equation (1) is labeled as producer risk by Dodge and Romig. The producer who rejects good product is creating a problem that in fact does not exist, perhaps even stopping the assembly line to find solutions to difficulties only imagined. For health services, this is service provider risk: the target population being classified as sickly when in fact it is healthy, the false positive (FP). In LQAS, equation (1) is often referred to as **α**, the probability of type 1 error.[[4]](#footnote-4) The values of α and β are predetermined in the LQAS method along with pu and pl and contribute to the solution for n and d, called a sampling plan (or sample design) PL(n, d). All of this information is summarized in table 1. It is important to emphasize that Dodge and Romig, coming before Neyman and Pearson, did not elaborate acceptance sampling as a hypothesis test in terms of α and β but rather as strictly probability calculations of consumer and producer risk.

|  |  |
| --- | --- |
|  | classifying healthy and sickly populations |
| Real states of population | x > d | x ≤ d |  |
| p > pu | P(x > d / p > pu) correct | P(x ≤ d / p > pu) = αFP error | ∑P = 1 |
| p ≤ pl | P(x > d / p ≤ pl) = β FN error | P(x ≤ d / p ≤ pl)correct | ∑P = 1 |

Table 1 Traditional approach to target population classification

The area between pu and pl is often called a “grey area”. In the area of health management, the term ‘grey region’ is frequently used. “Areas with coverage between pl and pu are in the ‘grey region’…...the classification procedure is not designed to accurately distinguish between areas with true coverages lying in the grey region”, (Hund, 2013). From the text, it should be clear that our position differs. Following the null hypothesis paradigm, p is a true value that exists in the population but is unknown. There is no uncertainty related to p itself. The contents of a sample of the population indicate or suggest regions of values for p in light of a sampling result that is not uncertain, x is either less than d or equal or greater than d. On the other hand, all values of p are always unknown. The value of d/x in the sample and its relation to pl and pu does not indicate a specific value for p but only the clue (as a likelihood and not a probability) that p is less than pl or greater than pu, or somewhere in between. To color the area in between as grey signifies that the area outside is white or black when in fact all areas are different shades of grey.

A second concept also unclear in LQAS and acceptance sampling in general is the necessity of linking pl exclusively to target population risk and pu exclusively to service provider risk. In reality, the population should also be preoccupied with a value of pu as should the service provider with pl. These concepts are new to the literature and will be elaborated in the remainder of this article.

NULL AND ALTERNATIVE HYPOTHESIS

The traditional presentation of LQAS and acceptance sampling in general, extremely useful in practical situations, is incomplete and suffers from inconsistencies that become more apparent when the correct interpretation of hypothesis testing is taken into account.

Our review of hypothesis testing is at most a simple skeleton of the area of scientific methodology, which is better elaborated in works like Rice (chapter 9, 1995) and the original work of Neyman and Pearson (1933). Nevertheless, our interpretation of acceptance sampling and LQAS in light of hypothesis testing is new to the literature. Here we will concentrate on the nature and definition of the null hypothesis.

Simply stated, a hypothesis is a clear statement of a characteristic (a characteristic has a certain value) or relation among variables (something happens associated with something else) that may or may not be true. It carries with itself a doubt that calls for evaluation. Hypotheses are not unique but come in pairs, a dual set, of exclusive statements. When the decision maker judges one of the hypotheses as true then the other hypothesis is necessarily judged as false. The lot is conforming or nonconforming. Children are vaccinated or not. Your candidate is winning the election campaign or is not winning. The accused is either innocent or guilty.

Incorrectly rejecting one of the hypotheses usually carries more severe consequences than incorrectly rejecting the other. As we have seen above, for the target population for instance, incorrectly classifying the sickly population as healthy, committing a false negative (FN), is a serious error for the population. **The hypothesis, that, if wrongly judged, causes relative suffering or other high costs should be chosen as the null hypothesis, (Rice, 1995).** This is the correct procedure in the area of hypothesis testing and serves to organize relevant social or industrial questions. The null carries the symbol Ho, the alternative hypothesis Ha. From the target population point of view therefore, the null hypothesis is that the population is sickly. HoPOP: p ≤ plPOP. The alternative hypothesis is that HaPOP: p > puPOP. Rejecting this null when it is true has extremely high costs for the population. In similar fashion but from the service provider point of view, the null hypothesis is that the population is healthy, HoSP: p ≥ puSP, because rejecting this null when it is true has extremely high costs for the service provider, who would have to repeat the health service when repetition is not necessary, drawing down its budget of scarce resources. The defining values plPOP and puSP are chosen by the two decision makers independently from each other.

Once recognizing the difference in the null for the separate points of view, the LQAS method requires a reformulation of equations (1) and (2).[[5]](#footnote-5)

SERVICE PROVIDER PERSPECTIVE

While it has been emphasized that the primary error to the service provider is to incorrectly classify a healthy population as sickly, there is a secondary error to the service provider based on the false null that the population is actually sickly. What we propose here is that the incorrect rejection of the false null (true alternative hypothesis) should carry some weight in LQAS for choosing values for **n and d**. In other words, the service provider suffers from two types of error, the false positive (FP) with probability αSP and the false negative (FN) with probability βSP. As explained above, the FP error is primary to the service provider and should carry more weight than secondary FN error. The contingency table 2 completes and summarizes the four alternatives LQAS defines, the combination of the true state of the population in terms of the null and the option chosen by the decision unit based on SRS. Two alternatives are errors and the other two alternatives are correct.

|  |  |
| --- | --- |
|  | Service provider chooses between states of the null hypothesis HoSP: population is healthy -- p ≥ puSP |
| Reality of null HoSP: population is healthy | accept HoSPwhen xSP > dSP | reject HoSPwhen xSP ≤ dSP |  |
|  | true HoSP: p ≥ puSP | TN correct P(xSP > dSP / p ≥ puSP) | FP type I error: **αSP** = P(xSP ≤ dSP / p ≥ puSP)  | ∑P = 1 |
| false HoSP: p < plSP | FN type II error: **βSP** = P(xSP > dSP / p < plSP)  | TP correct P(xSP ≤ dSP / p < plSP) | ∑P = 1 |

Table 2 Contingency table for service provider; FN = false negative, FP = false positive

TP = true positive, TN= true negative

Equation (1) and (2) should be rewritten specifying the service provider as the type I and type II errors from table 2:

1. αSP = P(xSP ≤ dSP / p ≥ puSP) = P(FP)
2. βSP = P(xSP > dSP / p < plSP) = P(FN)

This emphasizes that these equations are constructed from the point of view of the service provider. Both risks are from the service provider, however αSP is considered primary and more important than βSP. Sampling plans (sometimes called designs in the LQAS literature) are constructed for sample size nSP and cutoff value dSP - PL(nSP, dSP) - by resolving the two equations for given values of αSP, βSP, puSP, plSP. In this paper, the plans are calculated based on the hypergeometric probability function. This function was not used by Dodge and Romig because of practical considerations as seen below, substituted by the Poisson or binomial distributions, much easier to compute via adding machines and manual calculations.

In practice, a sample of size nSP is drawn and the number of healthy individuals xSP is counted. The contingency table prescribes that there are only two exclusive possibilities, xSP > dSP meaning that the population is likely to be healthy, or xSP ≤ dSP pointing to the likelihood of the population being sickly. At the end of the article we present an economical procedure for drawing the sample and reaching a conclusion.

The most relevant characteristic of the hypergeometric distribution is that population size (N) is not assumed as infinite but rather enters directly into its formulation. The hypergeometric mass function produces an estimate of the probability of x healthy persons appearing in the sample, and depends on the **percentage healthy in the population p = X/N** and the sample size (n). Note that the number of **healthy persons in the population is X**, and the number of **healthy persons in the sample is x**. See Gonin, H. T. (1936) for the mathematical development of the following equations.





*Ph(p) = h{ N, n, x, p }*

For LQAS and acceptance sampling in general, probability should be calculated as a cumulative sum involving x = 0, 1, 2…d, the sample characterizing a sickly population when x ≤ d. Specifically, in a sample of size n, from a population of size N, a population may actually be healthy even though the sample may indicate otherwise with relatively few healthy persons in the sample x ≤ d. The probability PH(0,1,2,...d), or simply PH(x ≤ d), depends upon the parameters of the sampling plan PL(N, n, x ≤ d) and the health of the population p. Consequently, with H representing the cumulative hypergeometric distribution *PH(p)=H{PL(N, n,* x ≤ d*), p}*. Equations (3) and (4) can be rewritten emphasizing the use of the hypergeometric:

1. αSP = *H{ PL(N, n,* x ≤ d*), p = puSP }*
2. βSP = 1 - *H{ PL(N, n,* x ≤ d*), p = plSP }*

For some trial values of N = 1000, n = 70, d = 60, a plot of PH(p) also called the operating characteristic curve (OCC) would look like the following figure, the probability of getting between zero and d = 60, inclusive, healthy persons in the sample, remembering that x ≤ d characterizes the population as sickly. As p approaches 1.0 the probability of finding less than d healthy persons in the sample diminishes. Likewise, as p approaches zero, the probability of getting less than d healthy people in the sample increases. In fact, these probability statements are the basis for calculating type I and type II error as shown in the discussion around table 2. Figure 1 OCC is a graphical representation of these relationships.[[6]](#footnote-6) The vertical axis represents the probability of obtaining d or less healthy persons in the sample, indicating the likelihood that the population is sickly. The horizontal axis shows the value of p, the healthiness of the population.



Figure 1 OCC PH(p) for PL(1000,70,60)

R CODE

p=seq(0,1,.01);

N=1000; n = 70; x =60;

plot(p,phyper(x, p\*N, (1-p)\*N, n),type="b",

ylab="P(x<=d)",xlab="p = X/N percent healthy in population", xlim=c(.7,.95))

abline(v=c(.91,.85),h=c(.088,.624),col="green")

round(data.frame(p,phyper(x, p\*N, (1-p)\*N, n)),3)

Recall that type I error for the service provider is declaring the population as sickly when, in fact, it is healthy, as shown in equation (3) or equivalently in the NE quadrant of table 2. In other words when the population is healthy, we want a sampling plan that will produce x ≤ d **only rarely** thus avoiding type I error for the service provider. The present plan PL(1000, 70, 60) produces a P(x ≤ d) of 8.8% for type I error if the service provider defines puSP as 91%. For a plSP of (say) 85% the value of P(x ≤ d) is about 62.4%, giving a P(x > d) of 37.6%, the probability of type II error.

Another illustrative way of plotting the hypergeometric distribution is as PH(d). Here we assume the same sampling plan and fix puSP at 90%.



Figure 2 Relationship between the **cutoff d** and the probability of the sample indicating sickliness of the population

R CODE

p=.9; N=1000; n = 70; d =seq(0,70,1);

plot(d,phyper(d, p\*N, (1-p)\*N, n),type="b",

 ylab="P(x<=d)",xlab="d cutoff value", xlim=c(55,70))

data.frame(p,phyper(d, p\*N, (1-p)\*N, n))

As the **cutoff d** increases it becomes more likely that samples will occur with x ≤ d, characterizing populations as sickly and consequently expediting type I error for the service provider.

Sampling plans – PL(N, n, d) -- may be constructed with different values for the probability of primary αSP and secondary error βSP and different values for puSP and plSP. We have taken advantage of the R package Acceptance Sampling for these calculations (Kiermeier, 2008). The principal algorithm in the R package begins with given values for plSP and puSP along with αSP and βSP (and N for the hypergeometric), searching for a sampling plan that has the **smallest sample size n** and still obeys the limits of the given values. Even though population size in LQAS is usually considered large enough to allow for the use of the simplified binomial distribution and not the less convenient hypergeometric, the option chosen here is to use the computationally more complicated distribution since it is readily available in the R package and whose speed of calculation is basically equivalent to the binomial by normal computing standards. Of course, if the population is small, it would be inappropriate to use the binomial distribution.

Several sampling plans are the contents of table 3.

|  |  |  |
| --- | --- | --- |
|  |  | SERVICE PROVIDER RISK |
|  |  | PRIMARY | SECONDARY | SAMPLING PLAN |
| plSP\* | puSP | αSP | βSP | n | d | d/n |
| 0.55 | 0.70 | .05 | .05 | 114 | 72 | .63 |
|  |  | .05 | .10 | 93 | 58 | .62 |
|  |  | .05 | .30 | 50 | 30 | .60 |
|  |  | .05 | .50 | 28 | 16 | .57 |
|  |  | .10 | .10 | 71 | 45 | .63 |
| 0.75 | 0.90 | .05 | .05 | 69 | 58 | .84 |
|  |  | .05 | .10 | 55 | 46 | .84 |
|  |  | .05 | .30 | 27 | 22 | .81 |
|  |  | .05 | .50 | 19 | 15 | .79 |
|  |  | .10 | .10 | 40 | 34 | .85 |
| 0.90 | 0.95 | .05 | .05 | 285 | 265 | .93 |
|  |  | .05 | .10 | 232 | 215 | .93 |
|  |  | .05 | .30 | 124 | 114 | .92 |
|  |  | .05 | .50 | 67 | 61 | .91 |
|  |  | .10 | .10 | 175 | 163 | .93 |

Table 3 Hypergeometric sampling plans (**service provider**) N=10,000

Source: R package Acceptance Sampling, Kiermeier (2008)

\*Suggested values for plSP and puSP from Hund (2015)

R CODE

library("AcceptanceSampling", lib.loc="~/R/win-library/3.0")

plSP=.75; puSP=.9

PR=0.05; #service provider risk

CR=0.5 #population risk

pop=10000

AQL=1-puSP;LTPD=1-plSP;

plan=find.plan(PRP=c(AQL, 1-PR), CRP=c(LTPD, CR),N=pop,type="b")

#b=binomial; h=hypergeometric

(n=plan$n)

(c=plan$c)

(d=n-c)

(d/n)

Where primary αSP and secondary βSP risks are 10%, the plans are comparable to the results from Hund (2015), She finds PL(71, 44), PL(40, 33) and PL(187, 173). Except for the last case, our results are practically the same. The difference in the last case, which some researchers might consider as small, should be better analyzed, but is certainly related to her use of the less exact nature of the binomial distribution.

In all the plans in table 3, the most economical plan for the service provider is the one with the smallest sample size PL(10000, 19, 15) with primary and secondary risks at .05 and 0.50. Secondary risk is relatively high but tolerable, considering its minor importance to the service provider. A ROC curve analysis can show the tradeoff between primary and secondary risk graphically, and help the service provider choose the appropriate combination.



Figure 3 Service provider ROC

sample size = 19, N = 10000, puSP =.9; plSP =.75

R CODE

size=19 #sample size

N=10000

n=size

puSP=.9;plSP=.75

d=seq(0,size, 1)

alphaSP=phyper(d,puSP\*N,(1-puSP)\*N,n)

plot(d,alphaSP,type="b",main="type I and type II error SP",

 ylab="alphaSP, betaSP",xlim=c(size-20,size))

betaSP=1-phyper(d,plSP\*N,(1-plSP)\*N,n)

lines(d,betaSP)

sumSP=alphaSP+betaSP

lines(d,sumSP)

dfSP=data.frame(n,d,plSP,puSP,alphaSP,betaSP,sumSP)

round(dfSP,3)

plot(alphaSP,betaSP,type="l",col="blue",

 xlim=c(0,max(alphaSP)-.7),

 ylim=c(0,max(betaSP)-.2),

 lwd=4, main="ROC SP")

abline(a=0,b=1)

abline(h=betaSP[which.min(sumSP)],

 v=alphaSP[which.min(sumSP)],col="green")

In figure 3, the ROC curve shows the tradeoff as the cutoff value **d** migrates from zero to 19 (the sample size). The straight line, fixed at the origin, distinguishes the relevant part of the ROC where αSP is always smaller than βSP. It might be irrational for a decision maker to use a sampling plan that accentuates primary risk and diminishes secondary risk.[[7]](#footnote-7) Another interesting result is at the intersection of the green lines. At that point the **sum of the two risks sumSP** is minimized. All of these alternative sampling plans can be appreciated in table 4.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| n | d | plSP | puSP | αSP | βSP | sumSP |
| 19 | 7 | 0.75 | 0.90 | 0.000 | 1.000 | 1.000 |
| 19 | 8 | 0.75 | 0.90 | 0.000 | 0.998 | 0.998 |
| 19 | 9 | 0.75 | 0.90 | 0.000 | 0.991 | 0.991 |
| 19 | 10 | 0.75 | 0.90 | 0.000 | 0.971 | 0.971 |
| 19 | 11 | 0.75 | 0.90 | 0.000 | 0.923 | 0.923 |
| 19 | 12 | 0.75 | 0.90 | 0.002 | 0.825 | 0.827 |
| 19 | 13 | 0.75 | 0.90 | 0.009 | 0.668 | 0.676 |
| 19 | 14 | 0.75 | 0.90 | 0.035 | 0.465 | 0.500 |
| **19** | **15** | **0.75** | **0.90** | **0.115** | **0.263** | **0.378** |
| 19 | 16 | 0.75 | 0.90 | 0.295 | 0.111 | 0.406 |
| 19 | 17 | 0.75 | 0.90 | 0.580 | 0.031 | 0.611 |
| 19 | 18 | 0.75 | 0.90 | 0.865 | 0.004 | 0.869 |
| 19 | 19 | 0.75 | 0.90 | 1.000 | 0.000 | 1.000 |

Table 4 Sampling plans for the service provider, sumSP = βSP + αSP

From the results of table 4 there are two outcomes to emphasize: first of all, primary risk αSP tends to rise along with d, as was shown in figure 2. Secondly, the best plan for the service provider would be PL(19,14) if the decision criterion is accepted that the sum of primary and secondary risks is to be minimized.

TARGET POPULATION PERSPECTIVE

The worst error that the target population could suffer is to be judged healthy when in fact the population is sickly. If the null is established to be that the population is unhealthy HoPOP: p ≤ plPOP, but the sample erroneously produces x > d, then a type I error is committed from the viewpoint of the target population, a false negative (FN). The avoidance of this error should be a priority for the target population, and consequently should demand that this error be small.[[8]](#footnote-8) In other words, the population is better off assuming a pessimistic outlook concerning its health. Table 5 is a representation of the contingency decisions available to the population.

|  |  |
| --- | --- |
|  | Target population chooses between states of the null hypothesis HoPOP: population is unhealthy -- p ≤ plPOP |
| Reality of null HoSP: population is unhealthy | accept HoSPwhen xPOP ≤ dSPOP | reject HoSPwhen xPOP > dPOP |  |
| true HoPOP: p ≤ plPOP | TP correct P(xPOP ≤ dSPOP / p ≤ plPOP.) | FN type I error: **αPOP**=P(xPOP > dPOP/p ≤plPOP)  | ∑P = 1 |
| false HoPOP: p > puPOP | FP type II error: **ΒPOP**=P(xPOP ≤ dSPOP/p>puPOP)  | TN correct P(xPOP > dPOP / p > puPOP) | ∑P = 1 |

Table 5 Contingency table for target population; FN = false negative, FP = false positive

TP = true positive, TN= true negative

In comparison with the previous section on the service provider perspective, type I and type II errors have suffered an alteration to take into account the perspective of the target population. Since the null has been altered to the statement that the population is unhealthy, type I error is now FN and type II error is FP.

1. **αPOP** = P(xPOP > dPOP/p ≤ plPOP) = P(FN)
2. **βPOP** = P(xPOP ≤ dSPOP/p > puPOP) = P(FP)

Several sampling plans are the contents of table 6. The most economical plan for the target population is the one with the smallest sample size PL(23,21) with primary and secondary risks at .05 and 0.50. Secondary risk is very high as expected but tolerable, considering its minor importance. However, the population is not necessarily worried about the minimum expenditure of a sampling plan that respects the minimum size of the sample but rather with a strong assurance that the population is healthy or needs attention, expenditure being a less important consideration. Therefore the population may demand higher levels for plPOP and puPOP.

|  |  |  |
| --- | --- | --- |
|  |  | TARGET POPULATION RISK |
|  |  | PRIMARY | SECONDARY | SAMPLING PLAN |
| plPOP | puPOP | αPOP | βPOP | n | d | d/n |
| 0.55 | 0.70 | .05 | .05 | 114 | 72 | .63 |
|  |  | .05 | .1 | 92 | 59 | .64 |
|  |  | .05 | .3 | 52 | 35 | .67 |
|  |  | .05 | .5 | 34 | 24 | .71 |
|  |  | .1 | .1 | 71 | 45 | .63 |
| 0.75 | 0.90 | .05 | .05 | 69 | 58 | .84 |
|  |  | .05 | .1 | 55 | 47 | .85 |
|  |  | .05 | .3 | 34 | 30 | .88 |
|  |  | .05 | .5 | 23 | 21 | .91 |
|  |  | .1 | .1 | 40 | 34 | .85 |
| 0.90 | 0.95 | .05 | .05 | 285 | 265 | .93 |
|  |  | .05 | .1 | 238 | 222 | .93 |
|  |  | .05 | .3 | 141 | 133 | .94 |
|  |  | .05 | .5 | 89 | 85 | .95 |
|  |  | .1 | .1 | 175 | 163 | .93 |

Table 6 Hypergeometric sampling plans (**target population**) N=10,000

Source: R package Acceptance Sampling, Kiermeier (2008)[[9]](#footnote-9)

Suggested values for plSP and puSP from Hund (2015)

It may choose 90 and 95 per cent as adequate, whose corresponding values are represented at the bottom of table 6. The ROC analysis that appears in figure 4 (and table 7) uses sample size of 89, and shows the risk tradeoff graphically, which may help the target population choose the appropriate combination of type I and type II risk. The minimum value of the sum of the risks (0.357) is from PL(10000, 89, 82) but the problem appears that this sampling plan reflects a primary risk that is greater than secondary risk. The plan PL(10000,89,83) is more appropriate and reduces primary risk from 0.201 to 0.108.



Figure 4 Population ROC

sample size = 89, N = 10000, puPOP =.95; plPOP =.90

R CODE

size=89 #sample size

N=10000

n=size

puPOP=.95;plPOP=.9

d=seq(0,size, 1)

betaPOP=phyper(d,puPOP\*N,(1-puPOP)\*N,n)

alphaPOP=1-phyper(d,plPOP\*N,(1-plPOP)\*N,n)

plot(d,alphaPOP,type="b",ylab="alphaPOP betaPOP ",

 xlim=c(size-10,size))

lines(d,betaPOP)

sumPOP=alphaPOP+betaPOP

lines(d,sumPOP,lwd=2)

dfPOP=data.frame(n,d,plPOP,puPOP,

 alphaPOP,betaPOP,sumPOP)

round(dfPOP,3)

plot(alphaPOP,betaPOP,type="l",col="blue",

 xlim=c(0,max(alphaPOP)),

 ylim=c(0,max(betaPOP)-.2),

 lwd=4)

abline(a=0,b=1)

abline(h=betaPOP[which.min(sumPOP)],

 v=alphaPOP[which.min(sumPOP)],col="green")

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| n | d | plPOP | puPOP | αPOP | βPOP | sumPOP |
| 89 | 81 | 0.90 | 0.95 | 0.323 | 0.076 | 0.399 |
| 89 | 82 | 0.90 | 0.95 | 0.201 | 0.157 | 0.357 |
| 89 | 83 | 0.90 | 0.95 | 0.108 | 0.286 | 0.394 |
| 89 | 84 | 0.90 | 0.95 | 0.049 | 0.461 | 0.510 |
| 89 | 85 | 0.90 | 0.95 | 0.018 | 0.657 | 0.675 |
| 89 | 86 | 0.90 | 0.95 | 0.005 | 0.829 | 0.834 |
| 89 | 87 | 0.90 | 0.95 | 0.001 | 0.942 | 0.943 |
| 89 | 88 | 0.90 | 0.95 | 0.000 | 0.990 | 0.990 |
| 89 | 89 | 0.90 | 0.95 | 0.000 | 1.0 | 1.0 |

Table 7 Sampling plans for the target population, sumPOP = betaPOP + alphaPOP

The two sampling plans presented here PL(10000, 19, 15) for the service provider and PL(10000,89,82) for the target population offer sample sizes and cutoff values that are substantially different. Before settling on a specific sampling plan it would be wise to negotiate a compromise between the two positions, bringing sample size and the cutoff to approximate values. There are several ways of arriving at this compromise by manipulating the definitions for a healthy population and the tolerable risk levels. Ideally, a unique sampling plan with tolerable values of risk for both parties would be the best way to avoid conflicts. The sampling plan PL(10000,150,120) might be considered as an interesting compromise given that the primary risk factors for both parties are less than 0.10 and secondary risk is practically null. This plan results from the following parameters:

For the target population

plPOP = .75, puPOP = .9

αPOP = 0.061, βPOP = 0.000,

and for the service provider

plSP = .7, puSP = .85,

αSP = .057, βSP = .002

This plan allows that the service provider work with limiting parameters for quality pu and pl slightly less rigorous than the population. The authors are currently involved in the task of writing algorithms that search out solutions of compromise.

FINAL COMMENTS

This article offers a generalization for LQAS based on the original tenets of hypothesis testing as developed by Neyman and Pearson. The procedure has been generalized to afford more respect for the desires and requirements of each of the two parties in question. Upper and lower quality limits can be assigned by each party (target population and service provider) independently, without assuming the domination (and submission) of either. Furthermore each party assigns its own values for primary and secondary risk. When pl and pu are considered as the same value for both population and service provider (plSP = plPOP and puSP = puPOP), then it is not difficult to show that βSP is equal to αPOP , and in like fashion βPOP is equal to αSP. These traditional equalities in LQAS are unduly restrictive, usually leading to the enhancement of the entitlement of one party at the expense of the weakened position of the other. The approach offered here allows each party to search for its own parameters and if necessary to find compromising positions that are transparent and impartial.

Further work in this area includes the elaboration of two open questions: first of all, Type I and II errors should be treated with differing weights to represent that errors are more or less important depending on the point of view of the decision maker. Additionally, the weighting process could be extended to the necessities of the two decision makers, perhaps giving more weight to the desires of the target population since they are the victims of the health question being studied. Secondly we are presently constructing functions in R to facilitate calculations.

PROCEDURE FOR MORE EFFICIENT SAMPLING

To illustrate an economical procedure for acceptance sampling in public opinion, we use the parameters from the last sampling plan PL(10000,150,120).

**STEP 1** Make sure that the 150 individuals from the target population for the potential sample comes from a random draw.

**STEP 2** Start interviewing individuals in the potential sample and record the hypothesized condition. For example, after the first 10 interviews the intermediate value of **x healthy individuals** might be 6. After another ten interviews, the total number of healthy individuals might be 13.

**STEP 3, part 1** Assume that the count of 121 **x healthy individuals** were reached with only 125 total sample individuals counted, x>d. At this point, in line with the sampling plan, the null of the population (sickly) is declared false (table 5) and the null of the service provider (healthy population) is accepted as true (table 2). We have sufficient evidence to conclude that coverage p is satisfactory. Sampling can stop before reaching the total number of sample individuals of 150.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| n | i interview | n-i interviews remaining | d | x(i) | ifx(i) > d, then | if(n-i) + x(i) ≤ dthen |
| 150 | 1 | 119 | 120 | x(1) | .. | .. |
| 150 | 2 | 118 | 120 | x(2) | .. | .. |
| 150 | 3 | 117 | 120 | x(3) | .. | .. |
| 150 | 4 | 116 | 120 | .. | .. | .. |
| 150 | .. | .. | 120 | .. | YES, STOPPopulation is healthy | STOPPopulation is sickly |
| 150 | .. | .. | 120 | .. |
| 150 | .. | .. | 120 | .. |

Table 8 Spreadsheet for tabulating sequential sampling results.

x(i) the number of healthy individuals counted up to and including the i-th interview.

**STEP 3, part 2** In another hypothetical case, this time assume that after 81 individuals are questioned, 50 are healthy. There are 69 (= 150 – 81) individuals remaining in the sample for classification. In this case, it would be impossible to surpass the **cutoff d** of 120. Here at i = 81 we already have sufficient evidence to accept the null of the target population (Ho: population is sickly) and reject the null of the service provider (Ho: population is healthy).

These steps are summarized in table 8, a spreadsheet for implementing acceptance sampling for LQAS. As interviews are held and healthy individuals tabulated in the column x(i), interviews continue until either x(i) > d concluding that the population is healthy, or (n-i) + x(i) ≤ d in which case the population is considered to be sickly. Figure 5 is a graphical illustration from hypothetical data. The black horizontal line at x(i) = d + 1 = 121 is the boundary line for indicating the likelihood of the healthiness of the population. If the value of x(i) becomes equal to d + 1 then the population is judged as healthy and sampling can stop. The diagonal black line is a graphical representation of the last column of table 8. It is the boundary for indicating that the population is sickly. If the count of x(i) goes below the diagonal line then the researcher can accept the likelihood that the population is sickly and stop sampling. This is the case for figure 5. The count of x(i) passes below the diagonal line at i = 58 and x(i) = 28. At i = 58 there are 92 interviews remaining, and even if all these interviews were with healthy individuals the value of x(i) would not reach the d = 121 limit. In the figure, x(i) is extended out to the sample size limit of 150 only to show the economy of sample elements afforded by our procedure, but actually the sample values that come after i = 58 are in practice unnecessary.



Figure 5 Sequential sampling procedure

REFERENCES

DODGE H.F., ROMIG H.G. (1929) A Method of Sampling Inspection. The Bell System Technical Journal. 8 (398):613–631. doi: 10.1002/j.1538-7305.1929.tb01240.x

DODGE, H.F., ROMIG, H.G. (1944). Sampling Inspection Tables. New York: John Wiley & Sons.

GONIN, H. T. (1936) "The use of factorial moments in the treatment of the hypergeometric distribution and in tests for regression," Philosophical Magazine, 7, Ser. 21, 215-226.

HEDT, BETHANY L, OLIVES, C., PAGANO, M., AND VALADEZ, J. (2008). “Large Country-Lot Quality Assurance Sampling : A New Method for Rapid Monitoring and Evaluation of…”, Health, Nutrition and Population (HNP) Discussion Paper [www.worldbank.org/hnppublications](file:///C%3A%5CUsers%5Csamohyl%5CAppData%5CRoaming%5CMicrosoft%5CWord%5CACCEPTANCE%20SAMPLING1304330541670010277%5Cwww.worldbank.org%5Chnppublications)

HUND L. New Tools for Evaluating LQAS Survey Designs. Emerging themes in epidemiology. 2014; 11 (1). doi: 10.1186/1742-7622-11-2 PMID: 24528928

HUND, L., BEDRICK, E. J., PAGANO, M. (2015). Choosing a Cluster Sampling Design for Lot Quality Assurance Sampling Surveys. *PLoS ONE*, 1–15. <http://doi.org/10.1371/journal.pone.0129564>

KIERMEIER A (2008). “Visualizing and Assessing Acceptance Sampling Plans: The R Package AcceptanceSampling.” Journal of Statistical Software, 26(6). URL http://www. jstatsoft.org/v26/i06

NEYMAN, J., PEARSON, E. (1933). On the problem of the most efficient tests of statistical hypotheses.

*Philosophical Transactions of the Royal Society of London, Series A, 231*, 289–337.

PROVOST, F., FAWCETT, T. (2013). Data Science and Business. O´Reilly, Cambridge.

RICE, JOHN, (1995). Mathematical Statistics and Data Analysis, Second Edition. Duxbury Press.

RHODA, D.A|., FERNANDEZ,S.A., FITCH, D.J., LEMESHOW,S. (2010). “LQAS: User Beware.” *International Journal of Epidemiology* 39 (1) (February): 60–8. doi:10.1093/ije/dyn366. <http://www.ncbi.nlm.nih.gov/pubmed/20139433.>

SAMOHYL, R. (2013). “Audits and Logistic Regression, Deciding What Really Matters in Service Processes. A Case Study of a Government Funding Agency for Research Grants.” *Proceedings of the Fourth International Workshop on Knowledge Discovery, Knowledge Management and Decision Support*. doi:10.2991/.2013.33. <http://www.atlantis-press.com/php/pub.php?publication=eureka-13&frame=http%3A//www.atlantis-press.com/php/paper-details.php%3Fid%3D9647.>

SAMOHYL, R. (2015). “An acceptance sampling approach to polling public opinion.” Working Paper Series,

Núcleo de Normalização e Qualimetria (NNQ), Federal University of Santa Catarina, Brazil, September, WP15-002.

SHMUELI, GALIT (2011). *Practical Acceptance Sampling*. Second edition.

1. The basic ideas in this article have been applied to the polling of public opinion (Samohyl, 2015). [↑](#footnote-ref-1)
2. Throughout the paper the nomenclature of healthy and sickly is used to classify the population but other names are possible, for instance vaccinated and not vaccinated, tested and not tested. [↑](#footnote-ref-2)
3. The concepts of positive and negative are well established in the medical and data science literature, for instance, Provost and Fawcett (chap. 7, 2013). A positive result indicates the presence of sickness or other undesirable traits. The true positive rate defined as TP/(TP+FN) is called sensitivity, and the true negative rate TN/(TN+FP) is specificity. [↑](#footnote-ref-3)
4. In the LQAS literature, type I and type II error are ill defined and cause much confusion, and will be discussed in the next sections. [↑](#footnote-ref-4)
5. Rhoda, et al (2010) must be credited with emphasizing the importance of the target population viewpoint, but did not generalize the LQAS method through Neyman and Pearson as we propose in this article. [↑](#footnote-ref-5)
6. The x axis has been conveniently shortened for your viewing pleasure. [↑](#footnote-ref-6)
7. Arguments here have not been elaborated in detail. For choosing the best point along the ROC, see Samohyl (2013). Weighting schemes can be derived for accentuating type I error and diminishing the cost of type II error. In fact, the number of potential weighting schemes is practically infinite and therefore should be constructed with strict criteria. [↑](#footnote-ref-7)
8. Of course, the entire population does not demand anything directly, because it is too large to organize and manifest itself as a unique decision making entity, but its representatives may. [↑](#footnote-ref-8)
9. This is basically the same R code used for table 3. [↑](#footnote-ref-9)