**UDC 378.14: 519.8**

# OPTIMAL QUALITY CONTROL PROVIDING USING THE NON-CENTRAL STUDENT'S DISTRIBUTION

Kateryna Vadymivna Stiepanova, Ph.D.,
Simon Kuznets Kharkiv National University of Economics, Kharkiv, Ukraine

***Abstract*** **— The aim of this brief communication is to shown that decision on just such attributes $X_i < L$ or $X_i \geq L$ is much less effective than using the values $X_i$ in their entirety to estimate the underlying normal population and from that get a better idea about $p$ for much smaller sample size.**

***Keywords*** **— non-central Student's distribution, quality control.**

Quality control deal with variables acceptance sampling plans (VASP). In a VASP the quality of items in a given sample is measured on a quantitative scale. An item is judged defective when its measured quality exceeds a certain threshold. The samples are drawn randomly from population of items. The objective is to make inferences about the proportions of defectives in the population. This leads either to an acceptance or a rejection of population quality as a whole.

In various applications the term "population" can have different meanings. It represents that collective of items from which the sample is drawn. Thus it could be a shipment, a lot of a batch of any other collective entity. Ultimately, any batch, lot or shipment is comprised of items that come from a certain process. If that process were to run indefinitely it would produce an infinite population of such items.

Thus the sampled items from the batch, lot or shipment could be considered as a sample from that larger conceptual population. Clearly, if the sample indicates that something is wrong the producer would presumably adjust the process appropriately.

Speaking about a VASP it is usually assumes that measurements $X_1,...,X_n$ for a random sample of $n$ items from a population are available and that defectiveness for any given sample item $i$ is equivalent to $X_i < L$, where $L$ is some given lower specification limit (see, for example, [1]). Assume that we deal with a random sample from a normal population with mean $\mu$ and standard deviation $\sigma$. The following note will be it terms of a lower specification limit $L$. Put

$$p = p(\mu, \sigma, L) = P_{\mu,\sigma}(X < L) = \Phi\left(\frac{L-\mu}{\sigma}\right),$$

which represents the probability that a given individual item in the population will be defective. Here $\Phi(x)$ denotes the normal distribution function, and $p$ can be interpreted as the proportion of defective items in the population. It is in the consumer's interest to keep the probability $p$ or proportion $p$ of defective items in the population below a tolerable value $p_1$. The producer will try to keep $p$ only so low as to remain cost effective. Hence the producer will aim for keeping $p \leq p_0$, where $p_0$ typically is somewhat smaller than $p_1$, in order to provide a sufficient margin between producer and consumer interest.

The consumer's demand that $p \leq p_1$ does not specify how that has to be accomplished in terms of $\mu$ and $\sigma$. The producer can control $p \leq p_1$ by either increasing $\mu$ sufficiently or by reducing $\sigma$, provided $\mu > L$. Reducing $\sigma$ is usually more difficult since various sources of variation have to be controlled more tightly. Increasing $\mu$ is

mainly in some way and is usually easier to accomplish. For normal data the standard VASP consist in computing $\bar{X}$ (sampling mean) and $S$ (standard deviation) from the obtained sample of $n$ items and in comparing $\bar{X} - kS$ with $L$ for an appropriately chosen constant $k$. If $\bar{X} - kS \geq L,$ then the consumer accepts the population from which the sample was drawn and otherwise it is rejected.

Note that rejection or acceptance is not based on the sample proportion of items with $X_i < L.$ Such classification would ignore how far above or below $L$ each measurement $X_i$ is. Basing decision on just such attributes $X_i < L$ or $X_i \geq L$ is much less effective than using the values $X_i$ in their entirety to estimate the underlying normal population and from that get a better idea about $p$ for much smaller sample size. Attribute data should only be used when the direct measurements are not available or not feasible. In that case one needs to employ attribute sampling plans based on the binomial distribution, requiring typically much higher sample size.

Before discussing the choice of $k$ in the acceptance criterion $\bar{X} - kS \geq L,$ it is appropriate to define the two notions of risk for such a VASP. Due to the random nature of the sample there is some chance that the sample misrepresents the population at least to some extent and thus may induce us to take incorrect action. The consumer's risk is probability of accepting the population when in fact the proportion $p$ of defectives in population is greater than the acceptable limit $p_1$. The producer's risk is the probability of rejecting the population when in fact the proportion $p$ of defectives in the population is less or equal $p_0$.

In [2] turns out that the probability of acceptance for a given VASP can be expressed in terms of test statistic as follows: $P_{\mu,\sigma}(\bar{X} - kS \geq L) =$

$$= P_{\mu,\sigma}\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} + \frac{\sqrt{n}(\mu - L)}{\sigma} \geq k\sqrt{n}\frac{S}{\sigma}\right) =$$

$$= P_{\mu,\sigma}\left(\frac{\sigma}{S}\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} + \delta(p)\right) \geq k\sqrt{n}\right),$$

where

$$\delta(p) = \frac{\sqrt{n}(\mu - L)}{\sigma} = -\sqrt{n}\Phi^{-1}(p) = -\sqrt{n}z_p. \quad (1)$$

According to [1] the non-centrality parameter $\delta(p)$ is a decreasing function of $p$, from [2] the operating characteristic curve of the VASP in given case $\gamma(p) = 1 - G_{n-1,\delta(p)}(k\sqrt{n})$ is decreasing in $p$ too.

The consumer's risk consist of the chance of accepting the population when in fact $p \geq p_1.$ In order to control the consumer's risk $\gamma(p)$ has to be kept at some sufficiently small level $\beta$ for $p \geq p_1.$ Since $\gamma(p)$ is decreasing in $p$ we need only insure $\gamma(p_1) = \beta$ by proper choice of $k.$ The factor $k$ is then found by solving the equation $\beta = 1 - G_{n-1,\delta(p_1)}(k\sqrt{n}),$ from which

$$k = G^{-1}_{n-1,\delta(p_1)}(1-\beta)/\sqrt{n}. \quad (2)$$

The probability of rejecting the population is $1 - \gamma(p),$ which is maximal over $p \leq p_0$ at $p_0.$ Hence the producer would want to limit this maximal risk $1 - \gamma(p_0)$ by some value $\alpha,$ customarily chosen to be 0.05. Note that $\alpha > 0$ and $\beta > 0$ must satisfy the constraint $\alpha + \beta < 1.$ Thus the producer is interested in ensuring that

$$\alpha = 1 - \gamma(p_0) = G_{n-1,\delta(p_0)}(k\sqrt{n}) \quad (3)$$

Solving (3) for $k,$ we will typically lead to a different value that obtained in (2). This conflict to having two different values of $k,$ depending on whose interest is being served, can be resolved by leaving the sample size $n$ flexible so that there are two control

parameters, $n$ and $k$, which can be used to satisfy the two conflicting goals. One slight problem is that $n$ is an integer and so it may not be possible to satisfy both equations (2) and (3) exactly.

What one can do instead the following: for a given value $n$ find $k = k(n)$ to solve (2). If that $k(n)$ also yields $\alpha \geq G_{n-1,\delta(p_0)}(k(n)\sqrt{n})$, then this sample size $n$ was possibly chosen too high and a lower value of $n$ should be tried. If we have $\alpha < G_{n-1,\delta(p_0)}(k(n)\sqrt{n})$, then $n$ was definitely chosen too small and a larger value of $n$ should be tried next. Through iteration one can arrive at the smallest sample size $n$ such that $k(n)$ and $n$ satisfy both (2) and (3). This iteration process will lead to a solution provided $p_0 < p_1$. If $p_0$ and $p_1$ are too close to each other, vary large sample sizes will be required.

In the case of an upper specification limit $U$ we accept the lot of population whenever $\bar{X} + kS \leq U$. Just by rewriting $X > U$ as $X' = -X < U = L$ we have the previous case: $\bar{X} + kS \leq U \Leftrightarrow -\bar{X} - kS \geq -U \Leftrightarrow \bar{X}' - kS' \geq L$, here $S' = S$, then $p = P(X > U) = P(X' < L)$. The same $k$ and $n$ as before suffice as solution as long as we identify $p = P(X > U)$ with $p = P(X' < L)$, i.e., specify only this risk $p$ of a population item being defective. Point out that the VASP does not say how the producer accomplishes the value $p \leq p_0$. This is usually based on extensive testing of the producer's broad experience. Also, the consumer cannot set $p_1$ arbitrarily low.

In order to understand the effect on the required sample size when all requirements are kept at the same levels. Let us compare the previously discussed here (the VASP) with a corresponding attributes acceptance sampling plan (AASP). As you known in the AASP the number $X$ of defective items is counted and the population is accepted when $X \leq k$, where $k$ and the smallest sample size $n$ are determined such that for given $p_0 < p_1$ and $\alpha > 0$, $\beta > 0$ with $\alpha + \beta < 1$ we have

$$P_{p_1}(X \leq k) \leq \beta,\ P_{p_0}(X \leq k) \geq 1 - \alpha.$$

In finding $n$ we start out with an $n$ suggested by the normal approximation to $X$ (with continuity correction)

$$P_p(X \leq k) \approx \Phi\left(\frac{k + 0.5 - np}{\sqrt{np(1-p)}}\right).$$

The requirements lead to the required sample size needed for appropriate quality control, in [2] rounded up to the next integer:

$$n = \left(\frac{z_\beta \sqrt{np_1(1-p_1)} + z_\alpha \sqrt{np_0(1-p_0)}}{p_0 - p_1}\right)^2,$$

where $z_\beta = \Phi^{-1}(\beta)$, $z_\alpha = \Phi^{-1}(\alpha)$ as defined in formula (1).

To sum up, let us compare the calculations. If $\alpha = 0.05, \beta = 0.1, p_0 = 0.01, p_1 = 0.05$, then for the VASP we had $n = 55$ whereas for the AASP $n = 132$, which is considerably higher. Thus, we can conclude that required sample size in VASP much less than in AASP when all requirements are kept at the same levels. Emphasize also that in VASP the search for the minimal sample size $n$ does not involve $L$, $\mu$ and $\sigma$.

## References

1. Hogg Robert V. Probability and Statistical Inference / R. V. Hogg, E. A. Tanis, D. L. Zimmerman. – Pearson: Pearson Higher Education, 2014. – 548 p.

2. Scholz F. Industrial Statistics / F. Scholz. – Washington: Washington EDU, 2014. – 58 p.

Author

**Stiepanova Kateryna Vadymivna,** Assistant of the Department of Higher Mathematics, Economic and Mathematical Methods, KhNUE, Kharkov, Ukraine (stepanova.katerina@hneu.net).

Тези доповіді надійшли 21 січня 2016 року.

Опубліковано в авторської редакції.