

Wilhelm Gaus<sup>1</sup>  
Rainer Muche  
Benjamin Mayer

## STATISTICAL CONSIDERATIONS FOR LOT-BY-LOT ACCEPTANCE/REJECTION SAMPLING WITH AN ATTRIBUTE

**Article info:**  
Received 24.08.2017  
Accepted 02.11.2017

UDC – 005.6  
DOI – 10.18421/IJQR11.04-05

**Abstract:** *The standard procedure for lot-by-lot acceptance/rejection sampling described in literature makes always a decision. If there are more invalid items in the sample than specified, then the lot is rejected, otherwise the lot is accepted. However, if the actual proportion of invalid items in the lot is close to the accepted quality level then it is – in principle – impossible to control the customer's and the producer's risk with a reasonable sample size. The published standard procedure can result in acceptance of the lot even if there is insufficient information. We think this is unfair to the customer. Three statistical methods with a common acceptable quality limit are proposed and discussed: A simple procedure, confidence intervals, and finally a statistical test. The idea of a statistical test, its application in lot-by-lot acceptance/rejection sampling, and the proposed statistical test are described in detail and with examples. Additionally, determination of sample size and multiple step sampling plans are characterized.*

**Keywords:** *confidence interval, one-sided versus two-sided hypothesis, sample size, sampling plan, statistical test*

### 1. Introduction

The standard procedure for lot-by-lot acceptance/rejection sampling for attributes is described in literature, e.g. by Montgomery (2013), Kreyszig (2011), Uhlmann (1982), and Woodall (1997). A similar procedure is given by ISO 2859 (2011), by ANSI/ASQ Z1.4 (2008) and by MIL-STD 105E (1989). This most commonly used procedure uses an operation characteristic (OC) and delivers sample size and acceptance number. If the actual number of invalid items in the sample is larger than

the acceptance number then the lot is rejected, otherwise the lot is accepted.

This standard procedure always makes a decision. It is based on the customer's and the producer's risks, the producer's acceptable quality level (AQL), the customer's rejectable quality level (RQL), and the relationship  $AQL < RQL$ . The relationship  $AQL < RQL$  is demanded by the way the statistical test is employed in the standard procedure, it is not desired by needs of quality control.

Assume that a customer wants to achieve high quality. The relationship  $AQL < RQL$  of the standard procedure forces him to accept a quality level worse than the quality level established by the producer. We think

---

<sup>1</sup> Corresponding author: Wilhelm Gaus  
email: [wilhelm.gaus@uni-ulm.de](mailto:wilhelm.gaus@uni-ulm.de)

this is unfair.

The difference  $RQL - AQL$  of the standard procedure has a decisive effect on the required sample size. If  $(RQL - AQL)$  approaches zero then the sample size approaches infinity. For practitioners of quality control it is unintelligible that a similar (or even the same) acceptable quality limit of producer and customer provoke an “explosion” of the sample size. Another drawback of the standard procedure is that the customer’s risk is not always controlled if the lot is accepted.

We think that the way a statistical test is used in the standard procedure does not meet the requirements of quality control, especially for customers. Therefore we propose a modified application of a statistical test for quality control. Our proposal can be applied for any AQL and any RQL. To simplify the paper we concentrate on a common acceptable quality limit (CAQL). Further we think that the customer’s risk is as important as the producer’s risk.

Our paper describes statistical procedures ranging from a simple technique to a statistical test where the customer’s risk (if the lot is accepted), and respectively the producer’s risk (if the lot is rejected), are truly controlled. However, if the proportion of invalid items in the lot is close to CAQL, then a statistical decision with controlled risks needs an extremely large sample or a decision with controlled risks is impossible. As statisticians we are convinced that we have to tell our users not only our methods, but their limitations as well.

Only investigation of a single lot is covered by this paper. Investigation of subsequent lots done with control cards is another issue. However, it is possible to enlarge the proposed methods to encompass the quality control of subsequent lots.

Industrial production and statistics often use different terminology. In the appendix we give an overview of the terms, variables, and symbols used.

Lot-by-lot acceptance/rejection sampling is a decision if  $Q \leq CAQL$  and therefore quality is sufficient so, as a consequence, the lot is accepted or if  $Q > CAQL$ , therefore quality is insufficient, and the lot is rejected. If this decision is based on a sample it is a statistical decision. Statistical decisions are often, but not always, correct, which is why the probability for a wrong decision should be controlled.

## 2. What has influence on the sample size n?

Large samples deliver more information than small ones. On the one hand a sample should be large enough to enable a decision on the quality of the lot with acceptable customer’s and producer’s risks. On the other hand a sample should be as small as possible to save effort, time, and costs. Therefore, sample size is a crucial issue.

CAQL has most influence on n, because it varies extremely. It may range from, for example, 0.20 to  $10^{-5}$ . In contrast, the customer’s risk and the producer’s risk vary much less, typically between 10% and 1%. The smaller the CAQL, the less invalid items we expect in the sample. But a reasonable number of invalid items has to be expected in the sample (Section 3). Thus, the smaller the CAQL the bigger n has to be.

If Q, the unknown proportion of invalid samples in the lot, is close to the CAQL then a decision is difficult. If small deviations of Q from CAQL should be detected then a large sample is needed. Or vice versa: If the sample size is small then only large deviations of Q from CAQL will give a significant result and a safe decision.

The selected  $\alpha_{\text{customer}}$  and the selected  $\alpha_{\text{producer}}$  also have influence on the required sample size n. If the risks for wrong decisions have to be small then they are difficult to control. Thus, the smaller  $\alpha_{\text{Customer}}$  and the smaller  $\alpha_{\text{Producer}}$  the bigger n has to be.

Does the lot size N influence the sample size n? Obviously, it is easier to draw a large

sample from a large lot than from a small lot. Conversely, if a large sample is drawn from a small lot then the advantages of sampling and statistics disappear. Moreover, a wrong decision – regardless of whether the disadvantage is for the customer or the producer – is more important for a large lot than for a small lot. Thus, for large lots a smaller  $\alpha_{Customer}$  and a smaller  $\alpha_{Producer}$  are recommended compared to a small lot.  $N$  appears in none of the formulas used in this paper except for the hypergeometric distribution which is mentioned in Section 7. Thus, the lot size has no influence on the sample size as long as the hypergeometric distribution is not used.

Finally, we have to acknowledge that if the sample size is reduced, then larger risks must be accepted. If a higher CAQL is accepted then  $n$  may be smaller. A larger  $\alpha_{Customer}$  and a larger  $\alpha_{Producer}$  also allows for a smaller  $n$ .

### 3. Expected number of invalid items in the sample

The actual number  $x$  of invalid items in the sample is an observed frequency. The expected number of invalid items in the sample –  $E(x)$  – is an expected frequency and can be computed beforehand. The observed frequency  $x$  is expressed by a natural number (0, 1, 2, 3, etc.), but the expected frequency  $E(x)$  is expressed by a real number (e.g. 0.67, 1.23, 4.50).

The expected number of invalid items in the sample is  $E(x) = n * Q$ . This doesn't help because  $Q$  is usually unknown. But

acceptance/rejection sampling wants to find out if  $Q$  is less or bigger than CAQL. Therefore,  $n * CAQL$  can serve as an expected value for  $E(x)$  on which to base the decision about the quality of the lot. If  $Q = CAQL$  is assumed then  $E(x) = n * CAQL$  invalid items in the sample are expected. For example, if  $n = 250$  and  $CAQL = 0.01 = 1\%$  then  $250 * 0.01 = 2.5$  invalid items in the sample are expected.

How large should  $E(x)$  be? It is easy to understand that  $E(x)$  should be a reasonable, not too small number. If e.g.  $E(x) = 0.1$  we have no realistic chance to observe an invalid item in the sample even if  $Q$  is somewhat larger than CAQL. If  $E(x)$  is approximately 1.0, often there will actually be no invalid item in the sample, i.e. we observe  $x = 0$ . The larger  $E(x)$  is, the safer is the decision to reject or to accept the lot, respectively, regardless of which method is used for the decision. As a rule of thumb  $E(x)$  should be at least  $\geq 2.5$ , even better at  $\geq 4.5$ , and good at  $\geq 8.5$ . The reason for this recommendation is given in Section 7.

From the equation  $E(x) = n * CAQL$  we get the formula:

$$n = E(x) / CAQL$$

This is a first and simple approach to determine the sample size  $n$ . Example: If CAQL is set to be 0.01 and  $E(x) = 2.5$  is selected, then  $n = 2.5 / 0.01 = 250$ . One may use Table 1 to find some basic sample size determinations based on this formula.

**Table 1.** Sample Size  $n$  for a Desired Expected Number of Invalid Items in the Sample  
 $E(x)$  = Expected number of invalid items in the sample

|              | 1.5 | 2.5 | 3.5 | 4.5 | 6.5 | 8.5 |
|--------------|-----|-----|-----|-----|-----|-----|
| CAQL = 0.100 | 15  | 25  | 35  | 45  | 65  | 85  |
| CAQL = 0.063 | 24  | 40  | 56  | 71  | 103 | 135 |
| CAQL = 0.040 | 38  | 63  | 88  | 113 | 163 | 213 |
| CAQL = 0.025 | 60  | 100 | 140 | 180 | 260 | 340 |

**Table 1.** Sample Size  $n$  for a Desired Expected Number of Invalid Items in the Sample (continued)

|                 | E(x) = Expected number of invalid items in the sample |        |        |        |        |        |
|-----------------|---|--------|--------|--------|--------|--------|
|                 | 1.5   | 2.5    | 3.5    | 4.5    | 6.5    | 8.5    |
| CAQL = 0.016    | 94  | 156    | 219    | 281    | 406    | 531    |
| CAQL = 0.010    | 150   | 250    | 350    | 450    | 650    | 850    |
| CAQL = 0.006 3  | 238   | 397    | 556    | 714    | 1 032  | 1 349  |
| CAQL = 0.004 0  | 375   | 625    | 875    | 1 125  | 1 625  | 2 125  |
| CAQL = 0.002 5  | 600   | 1 000  | 1 400  | 1 800  | 2 600  | 3 400  |
| CAQL = 0.001 6  | 938   | 1 563  | 2 188  | 2 813  | 4 063  | 5 313  |
| CAQL = 0.001 0  | 1 500   | 2 500  | 3 500  | 4 500  | 6 500  | 8 500  |
| CAQL = 0.000 63 | 2 381   | 3 968  | 5 556  | 7 143  | 10 317 | 13 492 |
| CAQL = 0.000 40 | 3 750   | 6 250  | 8 750  | 11 250 | 16 250 | 21 250 |
| CAQL = 0.000 25 | 6 000   | 10 000 | 14 000 | 18 000 | 26 000 | 34 000 |
| CAQL = 0.000 16 | 9 375   | 15 625 | 21 875 | 28 125 | 40 625 | 53 125 |
| CAQL = 0.000 10 | 15 000  | 25 000 | 35 000 | 45 000 | 65 000 | 85 000 |

**Reading example:** If CAQL is set to 4% = 0.04 and 3.5 invalid items wanted to be expected in the sample, then a sample size of  $n = 88$  is needed.

#### 4. Lot acceptance/rejection with a simple procedure

The simplest procedure to decide whether to accept or reject a lot is to compare  $x$  with  $E(x)$ .  $E(x)$  is computed under the assumption that the actual proportion of invalid items in the lot is equal to CAQL. In Sections 6 and 7 this assumption will be called “null hypothesis”. If the sample actually includes less invalid items than expected  $x < E(x)$ , then we assume  $Q < CAQL$  and accept the lot. Otherwise, if  $x > E(x)$  we assume that  $Q > CAQL$  and reject the lot. However,  $x$  stems from a sample and is therefore an estimate. Due to sampling errors it is sometimes too small and sometimes too large. This can lead to a wrong decision on the investigated lot. But in the long run (many investigated lots) the decisions are unbiased.

No decision is possible if  $x = E(x)$ . To prevent this situation the sample size can be established in a way that  $E(x) = n * CAQL$  is +0.5 larger than an integer. Then a decision if  $x > E(x)$  or if  $x < E(x)$  is always possible

because the observed number of invalid items in the sample  $x$  is always an integer. For example, CAQL is established as 0.025. If  $n = 220$  then  $E(x) = 5.5$  (instead of  $E(x) = 5.0$  with  $n = 200$ ) and a decision is always possible.

The decisions of the simple procedure are unbiased and therefore as correct as decisions proposed by more complicated procedures. The larger the deviation  $x$  from  $E(x)$  the safer is the decision. Readers interested in the customer’s or the producer’s risk should read Section 7.

The simple procedure can be modified in a way that if the deviation of  $x$  from  $E(x)$  is small, say  $<1$ ,  $<2$  or  $<3$ , then no decision on acceptance or rejection of the lot is taken but the sample size is increased (see Section 8).

#### 5. Lot acceptance/rejection with confidence interval

Each sample taken from a population has a sampling error. The sample error is sometimes larger, sometimes smaller, but it always exists. Therefore, a sample can only

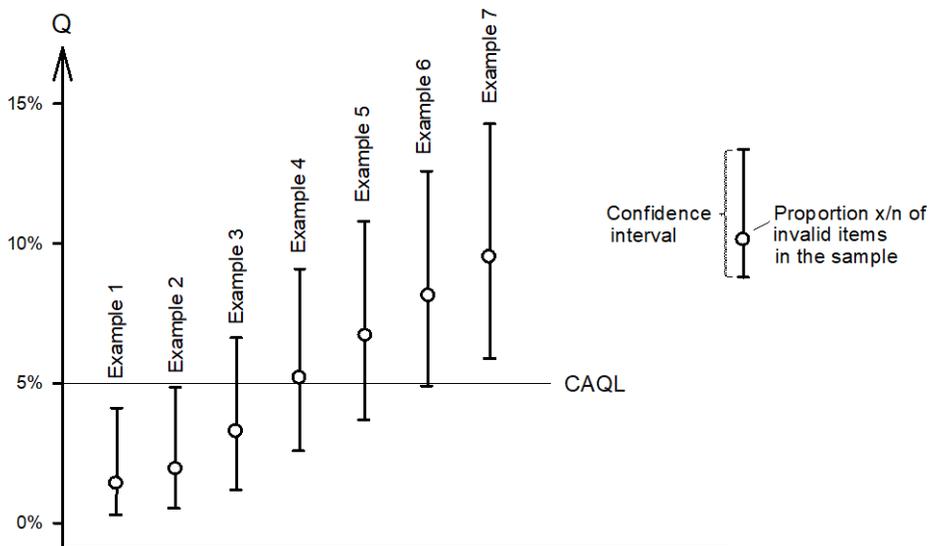
provide an estimator for the true value of the population. The sample error is the deviation of the value estimated from the sample and the true value in the population. It is always unknown (unless the whole population is inspected), but small errors occur more often than big errors.

The statistical method to estimate how large the sample error is maximally with a given probability, is called confidence interval. The true value of the population lies somewhere within the confidence interval with a predefined probability. This probability is called “confidence probability” and has to be established, often 0.95, 0.90 or 0.80 is selected.

A confidence interval covers underestimation as well as overestimation. Therefore statisticians say that confidence intervals are two-sided. This is analogous to a two-sided hypothesis mentioned in the

following section. However, producers will look at the upper bound of a confidence interval to see if Q is likely to be less than CAQL and therefore the lot has to be accepted. In contrast, consumers will look at the lower bound to see if Q is likely to be bigger than CAQL and therefore the lot has to be rejected.

A sample of size n is taken, the number of invalid items in the sample x is determined, and then the confidence interval for Q is computed from x, n, and the selected confidence probability. Now it can be checked, whether the confidence interval is below CAQL (lot will be accepted) or above CAQL (lot will be rejected) or if CAQL is within the confidence interval. Figure 1 gives a more detailed look at 7 examples. Confidence intervals for the proportion of invalid items in the lot are given in Tables 2(a, b, c).



**Figure 1.** Confidence Intervals for Q, the Proportion of Invalid Items in the Lot

**Legend to Fig.1**

These 7 examples were computed with the following indications:

- Sample size n = 210
- $\alpha_{customer} = \alpha_{producer} = 2.5\%$ . Thus, confidence probability = 0.95
- CAQL = 0.05 = 5%
- Therefore  $E(x) = n * CAQL = 210 * 0.05 = 10.5$

**Example 1:**  $x = 3$ , i.e. the proportion of invalid items in the sample is  $3/210 = 1.43\%$ : The 95% confidence interval ranges from 0.29% to 4.12%. The upper limit of the confidence interval is below CAQL. It is obvious that  $Q < CAQL$ . Decision: Lot is accepted. The customer's risk  $< 2.5\%$ .

**Example 2:**  $x = 4$ , i.e. the proportion of invalid items in the sample is  $4/210 = 1.90\%$ : The 95% confidence interval ranges from 0.52% to 4.80%. The upper limit of the confidence interval is very close to CAQL. It is obvious that  $Q \leq CAQL$ . Decision: Lot is accepted. The customer's risk  $\approx 2.5\%$ .

**Example 3:**  $x = 7$ , i.e. the proportion of invalid items in the sample is  $7/210 = 3.33\%$ : The 95% confidence interval ranges from 1.35% to 6.75%. The upper limit of the confidence interval is above the CAQL but most of the confidence interval is below CAQL. Thus, no save decision is possible, but there is a trend that  $Q$  is less than CAQL. If, nevertheless, the lot is accepted, then the customer's risk  $> 2.5\%$ .

**Example 4:**  $x = 11$ , i.e. the proportion of invalid items in the sample is  $11/210 = 5.24\%$ : The 95% confidence interval ranges from 2.64% to 9.18%. The confidence interval is nearly

symmetrical to CAQL. Maybe  $Q < CAQL$ , maybe  $Q > CAQL$ . A decision is not possible.

**Example 5:**  $x = 14$ , i.e. the proportion of invalid items in the sample is  $14/210 = 6.67\%$ : The 95% confidence interval ranges from 3.69% to 10.93%. The lower limit of the confidence interval is lower than the CAQL but most of the confidence interval is above CAQL. Thus, no save decision is possible, but there is a trend that  $Q$  is bigger than CAQL. If, nevertheless, the lot is rejected, then the producer's risk  $> 2.5\%$ .

**Example 6:**  $x = 17$ , i.e. the proportion of invalid items in the sample is  $17/210 = 8.10\%$ : The 95% confidence interval ranges from 4.79% to 12.64%. The lower limit of the confidence interval is very close to CAQL. It is obvious that  $Q \geq CAQL$ . Decision: Lot is rejected. The producer's risk  $\approx 2.5\%$

**Example 7:**  $x = 20$ , i.e. the proportion of invalid items in the sample is  $20/210 = 9.52\%$ : The 95% confidence interval ranges from 5.91% to 14.33%. The lower limit of the confidence interval is above CAQL. It is obvious that  $Q > AQL$ . Decision: Lot is rejected. The producer's risk  $< 2.5\%$ .

**Table 2a.** Confidence Intervals for  $Q$  with Confidence Probability = 0.95

|          | $x = 0$<br>$x = 5$            | $x = 1$<br>$x = 6$            | $x = 2$<br>$x = 7$            | $x = 3$<br>$x = 8$            | $x = 4$<br>$x = 9$             |
|----------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|--------------------------------|
| $n = 10$ | 0.00 - 30.85<br>18.71 - 81.29 | 0.25 - 44.50<br>26.24 - 87.84 | 2.52 - 55.61<br>34.75 - 93.33 | 6.67 - 65.25<br>44.39 - 97.48 | 12.16 - 73.76<br>55.50 - 99.75 |
| $n = 16$ | 0.00 - 20.59<br>11.02 - 58.66 | 0.16 - 30.23<br>15.20 - 64.57 | 1.55 - 38.35<br>19.75 - 70.12 | 4.05 - 45.65<br>24.65 - 75.35 | 7.27 - 52.38<br>29.88 - 80.25  |
| $n = 25$ | 0.00 - 13.72<br>6.83 - 40.70  | 0.10 - 20.35<br>9.36 - 45.13  | 0.98 - 26.03<br>12.07 - 49.39 | 2.55 - 31.22<br>14.95 - 53.50 | 4.54 - 36.08<br>17.97 - 57.48  |
| $n = 40$ | 0.00 - 8.81<br>4.19 - 26.80   | 0.06 - 13.16<br>5.71 - 29.84  | 0.61 - 16.92<br>7.34 - 32.78  | 1.57 - 20.39<br>9.05 - 35.65  | 2.79 - 23.66<br>10.84 - 38.45  |
| $n = 63$ | 0.00 - 5.69<br>2.63 - 17.56   | 0.04 - 8.53<br>3.58 - 19.59   | 0.39 - 11.00<br>4.59 - 21.56  | 0.99 - 13.29<br>5.65 - 23.50  | 1.76 - 15.47<br>6.75 - 25.39   |

**Table 2a.** Confidence Intervals for Q with Confidence Probability = 0.95 (continued)

|         | x = 0        | x = 1        | x = 2        | x = 3        | x = 4        |
|---------|--------------|--------------|--------------|--------------|--------------|
|         | x = 5        | x = 6        | x = 7        | x = 8        | x = 9        |
| n =100  | 0.00 – 3.62  | 0.03 – 5.45  | 0.24 – 7.04  | 0.62 – 8.52  | 1.10 – 9.93  |
|         | 1.64 – 11.28 | 2.23 – 12.60 | 2.86 – 13.89 | 3.52 – 15.16 | 4.20 – 16.40 |
| n =160  | 0.00 – 2.28  | 0.02 – 3.43  | 0.15 – 4.44  | 0.39 – 5.38  | 0.69 – 6.28  |
|         | 1.02 – 7.14  | 1.39 – 7.98  | 1.78 – 8.81  | 2.18 – 9.61  | 2.60 – 10.41 |
| n =250  | 0.00 – 1.46  | 0.01 – 2.21  | 0.10 – 2.86  | 0.25 – 3.47  | 0.44 – 4.05  |
|         | 0.65 – 4.61  | 0.89 – 5.15  | 1.13 – 5.68  | 1.39 – 6.21  | 1.66 – 6.72  |
| n =400  | 0.00 – 0.92  | 0.01 – 1.38  | 0.06 – 1.79  | 0.15 – 2.18  | 0.27 – 2.54  |
|         | 0.41 – 2.89  | 0.55 – 3.24  | 0.71 – 3.57  | 0.87 – 3.90  | 1.03 – 4.23  |
| n =630  | 0.00 – 0.58  | 0.00 – 0.88  | 0.04 – 1.14  | 0.10 – 1.39  | 0.17 – 1.62  |
|         | 0.26 – 1.84  | 0.35 – 2.06  | 0.45 – 2.28  | 0.55 – 2.49  | 0.66 – 2.69  |
| n =1000 | 0.00 – 0.37  | 0.00 – 0.56  | 0.02 – 0.72  | 0.06 – 0.87  | 0.11 – 1.02  |
|         | 0.16 – 1.16  | 0.22 – 1.30  | 0.28 – 1.44  | 0.35 – 1.57  | 0.41 – 1.70  |

n = sample size      x =number of invalid items in the sample  
 The body of the table gives the confidence interval for Q in percent.

**Reading example:** If a confidence probability of 95% is selected (i.e. customer’s risk of 2.5% and producer’s risk of 2.5%), the sample size is n = 160 and x = 0 invalid items were in the sample then Q can range from 0.00% to 2.28%. If the sample has x = 2 invalid items then Q can range from 0.15% to 4.44%. If the sample has x = 6 invalid items then Q can range from 1.39% to 7.98%. These statements on the size of Q are correct with a confidence probability of 0.95.

These exact confidence intervals were computed with SAS version 9.4, proc freq, option binomial. They are based on Clopper and Pearson (1934).

**Table 2b.** Confidence Intervals for Q with Confidence Probability = 0.90

|        | x = 0         | x = 1         | x = 2         | x = 3         | x = 4         |
|--------|---------------|---------------|---------------|---------------|---------------|
|        | x = 5         | x = 6         | x = 7         | x = 8         | x = 9         |
| n =10  | 0.00 – 25.89  | 0.51 – 39.42  | 3.68 – 50.69  | 8.73 – 60.66  | 15.00 – 69.65 |
|        | 22.24 – 77.76 | 30.35 – 85.00 | 39.34 – 91.27 | 49.31 – 96.32 | 60.58 – 99.49 |
| n =16  | 0.00 – 17.07  | 0.32 – 26.40  | 2.27 – 34.38  | 5.31 – 41.66  | 9.03 – 48.44  |
|        | 13.21 – 54.83 | 17.78 – 60.90 | 22.67 – 66.66 | 27.86 – 72.14 | 33.34 – 77.33 |
| n =25  | 0.00 – 11.29  | 0.20 – 17.61  | 1.44 – 23.10  | 3.35 – 28.17  | 5.66 – 32.96  |
|        | 8.23 – 37.54  | 11.01 – 41.95 | 13.95 – 46.22 | 17.03 – 50.36 | 20.24 – 54.39 |
| n =40  | 0.00 – 7.22   | 0.13 – 11.32  | 0.90 – 14.92  | 2.08 – 18.26  | 3.49 – 21.44  |
|        | 5.06 – 24.50  | 6.74 – 27.47  | 8.51 – 30.37  | 10.36 – 33.20 | 12.27 – 35.98 |
| n =63  | 0.00 – 4.64   | 0.08 – 7.31   | 0.57 – 9.66   | 1.31 – 11.85  | 2.20 – 13.94  |
|        | 3.18 – 15.97  | 4.23 – 17.93  | 5.33 – 19.86  | 6.48 – 21.75  | 7.66 – 23.61  |
| n =100 | 0.00 – 2.95   | 0.05 – 4.66   | 0.36 – 6.16   | 0.82 – 7.57   | 1.38 – 8.92   |
|        | 1.99 – 10.23  | 2.64 – 11.50  | 3.33 – 12.75  | 4.04 – 13.97  | 4.78 – 15.18  |
| n =160 | 0.00 – 1.85   | 0.03 – 2.93   | 0.22 – 3.88   | 0.51 – 4.77   | 0.86 – 5.63   |
|        | 1.24 – 6.46   | 1.65 – 7.27   | 2.07 – 8.06   | 2.51 – 8.84   | 2.97 – 9.61   |
| n =250 | 0.00 – 1.19   | 0.02 – 1.88   | 0.14 – 2.50   | 0.33 – 3.07   | 0.55 – 3.62   |
|        | 0.79 – 4.16   | 1.05 – 4.68   | 1.32 – 5.19   | 1.60 – 5.70   | 1.89 – 6.20   |

**Table 2b.** Confidence Intervals for Q with Confidence Probability = 0.90 (continued)

|          | x = 0 |        | x = 1 |        | x = 2 |        | x = 3 |        | x = 4 |        |
|----------|-------|--------|-------|--------|-------|--------|-------|--------|-------|--------|
|          | x = 5 |        | x = 6 |        | x = 7 |        | x = 8 |        | x = 9 |        |
| n = 400  | 0.00  | - 0.75 | 0.01  | - 1.18 | 0.09  | - 1.57 | 0.20  | - 1.93 | 0.34  | - 2.27 |
|          | 0.49  | - 2.61 | 0.66  | - 2.94 | 0.82  | - 3.26 | 1.00  | - 3.58 | 1.18  | - 3.89 |
| n = 630  | 0.00  | - 0.47 | 0.01  | - 0.75 | 0.06  | - 1.00 | 0.13  | - 1.23 | 0.22  | - 1.45 |
|          | 0.31  | - 1.66 | 0.42  | - 1.87 | 0.52  | - 2.08 | 0.63  | - 2.28 | 0.75  | - 2.48 |
| n = 1000 | 0.00  | - 0.30 | 0.01  | - 0.47 | 0.04  | - 0.63 | 0.08  | - 0.77 | 0.14  | - 0.91 |
|          | 0.20  | - 1.05 | 0.26  | - 1.18 | 0.33  | - 1.31 | 0.40  | - 1.44 | 0.47  | - 1.57 |

n = sample size      x = number of invalid items in the sample  
 The body of the table gives the confidence interval for Q in percent.

**Table 2c.** Confidence Intervals for Q with Confidence Probability = 0.80

|         | x = 0 |         | x = 1 |         | x = 2 |         | x = 3 |         | x = 4 |         |
|---------|-------|---------|-------|---------|-------|---------|-------|---------|-------|---------|
|         | x = 5 |         | x = 6 |         | x = 7 |         | x = 8 |         | x = 9 |         |
| n =10   | 0.00  | - 20.57 | 1.05  | - 33.68 | 5.45  | - 44.96 | 11.58 | - 55.17 | 18.76 | - 64.58 |
|         | 26.73 | - 73.27 | 35.42 | - 81.24 | 44.83 | - 88.42 | 55.04 | - 94.55 | 66.32 | - 98.95 |
| n =16   | 0.00  | - 13.40 | 0.66  | - 22.22 | 3.37  | - 29.96 | 7.10  | - 37.12 | 11.38 | - 43.89 |
|         | 16.06 | - 50.35 | 21.04 | - 56.54 | 26.29 | - 62.50 | 31.78 | - 68.22 | 37.50 | - 73.71 |
| n =25   | 0.00  | - 8.80  | 0.42  | - 14.69 | 2.15  | - 19.91 | 4.49  | - 24.80 | 7.17  | - 29.47 |
|         | 10.06 | - 33.97 | 13.12 | - 38.33 | 16.32 | - 42.58 | 19.62 | - 46.73 | 23.03 | - 50.80 |
| n =40   | 0.00  | - 5.59  | 0.26  | - 9.38  | 1.34  | - 12.76 | 2.79  | - 15.94 | 4.43  | - 19.00 |
|         | 6.21  | - 21.96 | 8.07  | - 24.85 | 10.01 | - 27.67 | 12.01 | - 30.45 | 14.06 | - 33.18 |
| n =63   | 0.00  | - 3.59  | 0.17  | - 6.03  | 0.85  | - 8.23  | 1.76  | - 10.30 | 2.80  | - 12.29 |
|         | 3.91  | - 14.23 | 5.08  | - 16.12 | 6.29  | - 17.98 | 7.54  | - 19.81 | 8.81  | - 21.61 |
| n =100  | 0.00  | - 2.28  | 0.11  | - 3.83  | 0.53  | - 5.23  | 1.11  | - 6.56  | 1.76  | - 7.83  |
|         | 2.45  | - 9.08  | 3.18  | - 10.29 | 3.94  | - 11.49 | 4.71  | - 12.67 | 5.50  | - 13.84 |
| n =160  | 0.00  | - 1.43  | 0.07  | - 2.41  | 0.33  | - 3.29  | 0.69  | - 4.13  | 1.09  | - 4.93  |
|         | 1.53  | - 5.72  | 1.98  | - 6.49  | 2.45  | - 7.25  | 2.93  | - 7.99  | 3.42  | - 8.73  |
| n =250  | 0.00  | - 0.92  | 0.04  | - 1.55  | 0.21  | - 2.11  | 0.44  | - 2.65  | 0.70  | - 3.17  |
|         | 0.98  | - 3.68  | 1.27  | - 4.17  | 1.56  | - 4.66  | 1.87  | - 5.15  | 2.18  | - 5.62  |
| n =400  | 0.00  | - 0.57  | 0.03  | - 0.97  | 0.13  | - 1.33  | 0.28  | - 1.66  | 0.44  | - 1.99  |
|         | 0.61  | - 2.31  | 0.79  | - 2.62  | 0.98  | - 2.93  | 1.17  | - 3.23  | 1.36  | - 3.53  |
| n =630  | 0.00  | - 0.36  | 0.02  | - 0.62  | 0.08  | - 0.84  | 0.18  | - 1.06  | 0.28  | - 1.26  |
|         | 0.39  | - 1.47  | 0.50  | - 1.67  | 0.62  | - 1.86  | 0.74  | - 2.05  | 0.86  | - 2.25  |
| n =1000 | 0.00  | - 0.23  | 0.01  | - 0.39  | 0.05  | - 0.53  | 0.11  | - 0.67  | 0.17  | - 0.80  |
|         | 0.24  | - 0.93  | 0.32  | - 1.05  | 0.39  | - 1.17  | 0.47  | - 1.30  | 0.54  | - 1.42  |

n = sample size      x = number of invalid items in the sample  
 The body of the table gives the confidence interval for Q in percent

If no invalid item is found in the sample this is not a proof that there are no invalid items in the lot. The maximal proportion of invalid items in the lot is given by the upper bound of the confidence interval for  $x = 0$ . Of course,  $x/n$ , the observed proportion of

invalid items in the sample, is within the confidence interval. However, it is not in the centre of the confidence interval, but instead closer to its lower bound. The reason for this is that the true proportion of invalid items in the lot cannot be less than zero.

In quality control typically producer's risk and customer's risk are communicated. But in statistics there is a long tradition to give for confidence intervals the probability for correctness instead of risks. Further, with

confidence intervals, the probability for overestimation and the probability for underestimation are usually the same. The relationship between confidence probability and error probability is simply:

$$\begin{aligned} \text{Confidence probability} + \text{error probability} &= 1.0 \\ \text{Confidence probability} &= 1.0 - (\alpha_{\text{customer}} + \alpha_{\text{producer}}) \end{aligned}$$

In literature several formulas to compute confidence intervals for proportions are published, e.g. Altman et al. (2000), Gaus and Muche (2017). However, nearly all of these formulas are not qualified for small risks, i.e. for small CAQLs. The procedure for manual computation of confidence intervals proposed by Clopper and Pearson (1934) is graphical, or needs a very detailed F-Table which is only computationally available.

If  $\alpha_{\text{customer}} \neq \alpha_{\text{producer}}$  we propose to compute the confidence interval twice: One confidence interval with confidence probability  $(1 - 2 * \alpha_{\text{customer}})$  and a second confidence interval with confidence probability  $(1 - 2 * \alpha_{\text{producer}})$ . Then the lower bound of the confidence interval computed with  $\alpha_{\text{producer}}$  is used for rejection of the lot, the upper bound of the confidence interval computed with  $\alpha_{\text{customer}}$  is used for acceptance (Figure 1).

The smaller the confidence interval the easier a decision can be made. The larger the sample size and the larger both  $\alpha_{\text{customer}}$  and  $\alpha_{\text{producer}}$ , the smaller the confidence interval. Table 2 gives confidence intervals for Q for various sample sizes and various numbers of invalid items in the sample.

The main advantage of a decision by confidence interval is – compared to the simple procedure described in Section 4 – that a decision with controlled  $\alpha_{\text{customer}}$  and controlled  $\alpha_{\text{producer}}$  is possible. However, often CAQL will be within the confidence interval and a decision with controlled risks is not possible. But a confidence interval gives at least a feeling of how likely it is that the decision is correct. If CAQL is within the

confidence interval, then the sample size can be enlarged (see Section 8) or a decision without controlled customer's risk or without controlled producer's risk is always possible (see Section 4). Another advantage of confidence intervals is that it is possible that the customer and the producer make a CAQL, or instead a common AQL, their own acceptable quality limits ( $AQL_{\text{customer}} \neq AQL_{\text{producer}}$ ) and use the same risks ( $\alpha_{\text{customer}} = \alpha_{\text{producer}}$ ) or different risks ( $\alpha_{\text{customer}} \neq \alpha_{\text{producer}}$ ).

## 6. Lot acceptance/rejection with a statistical test

A statistical test is a formal decision founded on observed data, Altman (1991), Bland (2000), and Ilakovac (2009). It is based on the so called "null hypothesis", i.e. the effect investigated is null or non-existent. The exact antithesis to the null hypothesis is called the "alternative". All possibilities have to be covered by either the null hypothesis or the alternative.

A null hypothesis and its alternative may be one-sided or two-sided. In a one-sided hypothesis the alternative has only one predisposed direction. A two-sided hypothesis allows the alternative to demonstrate differences in one or the other direction. Assume that S is a sample and P is the reference population. The question is, if the sample stems from the reference population or not. If a sample is compared with a population then:

Two-sided hypothesis: Null hypothesis:  $S = P$ ; Alternative:  $S \neq P$   
 One-sided hypothesis: Null hypothesis:  $S \leq P$ ; Alternative:  $S > P$   
 or  
 Null hypothesis:  $S \geq P$ ; Alternative:  $S < P$

The test computes the probability that the observed data – or more extreme data – originate from the null hypothesis with a purely random process. If this probability (p-value) is less than or equal to an established probability, called “level of significance”, then the test is significant, otherwise the test is non-significant.

The standard procedure (Montgomery, 2013; Kreyszig, 2011; Uhlmann, 1982; Woodall, 1997; ISO 2859, 2011; ANSI/ASQ Z1.4, 2008; and MIL-STD-105E, 1989) uses a statistical test for a one-sided hypothesis. But we think that in quality control the hypothesis is two-sided: Null hypothesis:  $Q = CAQL$ , alternative:  $Q \neq CAQL$ . According to the null hypothesis we expect  $E(x) = n \cdot CAQL$  invalid items in the sample.

The test has to find out if  $x$  deviates significantly from  $E(x)$ . If the test is significant and  $Q > CAQL$  then we decide to reject the lot, if the test is significant and  $Q < CAQL$  then we decide to accept the lot.

“Significant” means that the deviation of the observed from the expected number of invalid items in the sample cannot be by chance. More precisely, if the null hypothesis is actually true then the probability that the observed number of invalid items in the sample deviates only at random from the expected number is smaller than  $\alpha_{customer}$  or  $\alpha_{producer}$ .

Every decision can be correct or incorrect. The decision taken by a statistical test has the possibilities described in Table 3.

**Table 3.** Possibilities of a Decision Based on a Statistical Test

|                         | In reality null hypothesis is correct, $S = P$ | In reality the alternative is correct, $S > P$ or $S < P$ |
|-------------------------|--|---|
| Test is significant     | Type I mistake, $\alpha$ -risk                 | Correct decision  |
| Test is not-significant | Correct decision                               | Type II mistake, $\beta$ -risk                            |

The outcome of a statistical test comparing a sample  $S$  with a population  $P$  with a two-sided hypothesis is described by a fourfold scheme. Traditionally wrong decisions are called type I mistakes and type II mistakes. The probability for a type I mistake is  $\alpha$ -risk, the probability for a type II mistake a  $\beta$ -risk.

To prevent misunderstanding we must point out that if a test is significant, then the probability that its decision is incorrect is not the  $\alpha$ -risk, and if a test is not-significant, then the probability that its decision is

incorrect is not the  $\beta$ -risk. This becomes clear by using conditional probabilities (event | condition).

prob (test significant | null hypothesis is actually correct) =  $\alpha$ -risk

prob (null hypothesis is actually correct | test is significant) is not known but typically small.

prob (test not-significant | alternative is actually correct) =  $\beta$ -risk

prob (alternative is actually correct | test is not-significant) is not known.

A statistical test controls the  $\alpha$ -risk directly. If a statistical test is significant we can decide to reject the null hypothesis and to accept the alternative.

The power of a statistical test is the probability to achieve significance if the alternative is actually correct.  $Power = (1 - \beta\text{-risk})$ . The power depends on several influences, but mainly on the sample size and the actual (but unknown) effect size, i.e. the actual deviation from the null hypothesis. In quality control the effect size is the deviation of Q from CAQL. For computation of the power, the effect size and n have to be specified among other specifications (McHugh 2008).

Non-significance can have two reasons: (1<sup>st</sup>) the null hypothesis is actually true or, (2<sup>nd</sup>) the power of the test was too low for the actual effect size and the given sample size. Thus only a significant result gives a “statistical proof” with a controlled risk.

Assume a lot with  $Q = CAQL$  and that  $E(x)$  is sufficient large. From this population many random samples – say 100, each with sample size n – are drawn. From each of these samples the number of invalid items x is determined. We expect for half of the samples  $x < E(x)$  and for the other half  $x > E(x)$ .

This experiment in thought demonstrates that for  $Q = AQL$  no substantiated decision on

acceptance/rejection of the lot is possible. The more Q approaches CAQL the more this decision is dominated by randomness.

### 7. Proposal for a statistical test for lot acceptance/rejection

A basic rule of statistics is that there is no statistical test without descriptive statistics. Hence, descriptive statistics is the first part, and a statistical test the second part of data analysis. First we compare x with  $E(x)$ . If  $x > E(x)$  then the customer will tend to reject the lot. If  $x < E(x)$  then the producer will tend to accept the lot. The statistical test decides if the difference between expected and observed frequencies may be due to chance (test is not significant) or not (test is significant).

If  $x > E(x)$  and the test is significant, then the lot is rejected. If  $x < E(x)$  and the test is significant, then the lot is accepted. If the test is not significant then the deviation of x from  $E(x)$  may be due to randomness and a decision is insecure.

There are three mathematical methods to compute the probability to observe x occurrences for a given sample size n and a given probability for a simple event (Table 4).

**Table 4.** Three mathematical methods

| Method                      | Computational Work | Prerequisites   |
|-----------------------------|--------------------|---|
| Poisson distribution        | small              | CAQL is small, i.e. $CAQL \leq 0.05$<br>$20 \leq n \ll N$ , i.e. $20 \leq n \leq 0.1 * N$ |
| Binomial Distribution       | medium             | $n \ll N$ , i.e. $n \leq 0.1 * N$   |
| Hypergeometric Distribution | large              | none  |

For all examples we used the Poisson distribution to enable highly interested readers to re-compute the examples with a simple, hand-held pocket computer. But there is no problem in using the other methods with a programmed computer.

The probability that a single item of a sample is invalid (simple event) is Q in this context. However, Q is usually unknown. Acceptance/rejection sampling wants to find out if  $Q > CAQL$  or if  $Q \leq CAQL$ . Thus, Q is replaced by CAQL.

Adapted to quality control the formula of the Poisson distribution is:

$$\text{prob}(x) = (n * \text{CAQL})^x * e^{-(n * \text{CAQL})} / x!$$

where  $x!$  is defined as  $1*2*3*4* \dots *x$ , and

$(n * \text{CAQL})$  is the expected number of invalid items in the sample  $E(x)$ .

There is no specific statistical test procedure

for the problem discussed here. But it is rather simple to compute the probability that the observed data – or more extreme data – originate from the null hypothesis with a purely random process. Tables 5(a, b) describes the procedure to compute the customer’s risk and the producer’s risk.

**Table 5a.** Example for Computation of the Customer’s Risk and the Producer’s Risk

|         | Step 2<br>Compute $p(x)$ | Step 3<br>Cumulate | Step 4<br>$1-p(\leq(x-1))$ | Step 5<br>$x'$ | Step 6<br>Two-Sided |
|---------|--------------------------|--------------------|----------------------------|----------------|---------------------|
| $x = 0$ | $p(0)=0.082$             | $p(\leq 0)=0.082$  |                            | 5              | $p(\leq 0)=0.190$   |
| $x = 1$ | $p(1)=0.205$             | $p(\leq 1)=0.287$  |                            | 4              | $p(\leq 1)=0.529$   |
| $x = 2$ | $p(2)=0.257$             | $p(\leq 2)=0.544$  |                            | 3              | $p(\leq 2)=1.000$   |
| $x = 3$ | $p(3)=0.214$             | $p(\leq 3)=0.758$  | $p(\geq 3)=0.456$          | 2              | $p(\geq 3)=1.000$   |
| $x = 4$ | $p(4)=0.134$             | $p(\leq 4)=0.892$  | $p(\geq 4)=0.242$          | 1              | $p(\geq 4)=0.529$   |
| $x = 5$ | $p(5)=0.067$             | $p(\leq 5)=0.959$  | $p(\geq 5)=0.108$          | 0              | $p(\geq 5)=0.190$   |
| $x = 6$ | $p(6)=0.028$             | $p(\leq 6)=0.987$  | $p(\geq 6)=0.041$          | -              | $p(\geq 6)=0.041$   |
| $x = 7$ | $p(7)=0.010$             | $p(\leq 7)=0.997$  | $p(\geq 7)=0.013$          | -              | $p(\geq 7)=0.013$   |
| $x = 8$ | $p(8)=0.002$             | $p(\leq 8)=0.999$  | $p(\geq 8)=0.003$          | -              | $p(\geq 8)=0.003$   |
| $x = 9$ | $p(9)=0.001$             | $p(\leq 9)=1.000$  | $p(\geq 9)=0.001$          | -              | $p(\geq 9)=0.001$   |

Assume that  $E(x) = 2.5$  invalid items are expected for the sample. For  $x < E(x)$  the customer’s risk is computed and for  $x > E(x)$  the producer’s risk (step 6), both for different numbers of invalid items in the sample ( $x$ ). For explanation of the steps of computation see separate box.

This table is only valid for  $E(x)=2.5$ , for other values of  $E(x)$  see Table 6.

For  $x > 9$  the probability is negligible. The differences in results compared to Table 6 are due to reduced accuracy when calculating with hand held computer.

Reading examples: For  $E(x)=2.5$  and  $x=0$  and the lot is accepted then customer’s risk is 0.19. For  $E(x)=2.5$  and  $x=6$  and the lot is rejected then producer’s risk is 0.041

**Table 5b.** Steps of Computation for the Results of Table 5a

|         |  |
|---------|--|
| Step 1  | Check if the prerequisites for Poisson distribution are fulfilled: $\text{CAQL} \leq 0.05$ AND $20 \leq n \leq 0.1 * N$ . If not then replace in step 2 Poisson distribution by binomial or hypergeometric distribution as appropriate. For the example $\text{CAQL} = 0.025$ , $n = 100$ , and therefore $E(x) = 2.5$ is the Poisson distribution valid and used. |
| Step 2: | Compute with the Poisson (or binomial or hypergeometric) distribution for $E(x) = 2.5$ the probability that $x$ invalid items are observed in the sample, i.e. $p(x)$ .  |
| Step 3: | Cumulate the probabilities. $p(\leq x) = p(x) + p(x-1) + p(x-2) + \text{etc. until } + p(0)$<br>For $x < E(x)$ this is the one-sided customer’s risk.  |
| Step 4: | For $x > E(x)$ compute the probabilities $p(\geq x) = 1.0 - p(\leq(x-1))$<br>For $x > E(x)$ this is the one-sided producer’s risk.   |

**Table 5b.** Steps of Computation for the Results of Table 5a (continued)

|         |   |
|---------|---|
| Step 5: | Determine $x'$ which is opposite to $x$ on the other side of $E(x)$ .<br>Example 1: $x = 1$ is 1.5 below $E(x) = 2.5$ . Thus, $x'$ is 1.5 above $E(x)$ , it is $x' = 4$ .<br>Example 2: $x = 4$ is 1.5 over $E(x) = 2.5$ . Thus, $x'$ is 1.5 below $E(x)$ , it is $x' = 1$ .<br>Example 3: $x = 5$ is 2.5 over $E(x) = 2.5$ . $x'$ is 2.5 below $E(x)$ , it is $x' = 0$ .<br>Example 4: For $x = 6$ the opposite $x'$ would be $-1$ . This is not existent.   |
| Step 6: | Compute the two-sided probabilities.<br>For $x < E(x)$ add $p(\leq x) + p(\geq x')$ . In particular:<br>For $x=0$ : $p(\leq 0) + p(\geq 5) = 0.082 + 0.108 = 0.190$<br>For $x=1$ : $p(\leq 1) + p(\geq 4) = 0.287 + 0.242 = 0.529$<br>For $x=2$ : $p(\leq 2) + p(\geq 3) = 0.544 + 0.456 = 1.000$<br>For $x > E(x)$ add $p(\geq x) + p(\leq x')$ . In particular:<br>For $x=3$ : $p(\geq 3) + p(\leq 2) = 0.456 + 0.544 = 1.000$<br>For $x=4$ : $p(\geq 4) + p(\leq 1) = 0.242 + 0.287 = 0.529$<br>For $x=5$ : $p(\geq 5) + p(\leq 0) = 0.108 + 0.082 = 0.190$<br>If $x'$ does not exist then the addition is not applicable. In particular:<br>For $x=6$ : $p(\geq 6) = 0.041 = 0.041$<br>For $x=7$ : $p(\geq 7) = 0.013 = 0.013$<br>etc.<br>This column gives for $x < E(x)$ the two-sided customer's risk, and for $x > E(x)$ the two-sided producer's risk. |

Table 6 gives the risk probabilities for various situations. Examples: For  $\alpha_{\text{customer}} = 0.2$  a minimal  $E(x) = 2.5$  is necessary, for  $\alpha_{\text{customer}} = 0.1$  a minimal  $E(x) = 3.5$ , and for  $\alpha_{\text{customer}} = 0.05$  a minimal  $E(x) = 4.5$ , otherwise these customer's risks are not achieved even with  $x = 0$ .

**Table 6.** The Customer's and the Producer's Risk for Expected and Observed Number of Invalid Items in the Sample

|   |  |   |   |
|---|--|---|---|
| $E(x) = 0.5$<br>prob (0) = 1.000,<br>prob ( $\geq 4$ ) = 0.002  | prob ( $\geq 1$ ) = 1.000,   | prob ( $\geq 2$ ) = 0.090,  | prob ( $\geq 3$ ) = 0.014,  |
| $E(x) = 1.5$<br>prob (0) = 0.414,<br>prob ( $\geq 4$ ) = 0.066,   | prob ( $\leq 1$ ) = 1.000,<br>prob ( $\geq 5$ ) = 0.019,                               | prob ( $\geq 2$ ) = 1.000,<br>prob ( $\geq 6$ ) = 0.004                                 | prob ( $\geq 3$ ) = 0.414,  |
| $E(x) = 2.5$<br>prob (0) = 0.191,<br>prob ( $\geq 4$ ) = 0.530,<br>prob ( $\geq 8$ ) = 0.004                                | prob ( $\leq 1$ ) = 0.530,<br>prob ( $\geq 5$ ) = 0.191,                               | prob ( $\leq 2$ ) = 1.000,<br>prob ( $\geq 6$ ) = 0.042,                                | prob ( $\geq 3$ ) = 1.000,<br>prob ( $\geq 7$ ) = 0.014,                                |
| $E(x) = 3.5$<br>prob (0) = 0.095,<br>prob ( $\geq 4$ ) = 1.000,<br>prob ( $\geq 8$ ) = 0.027,                               | prob ( $\leq 1$ ) = 0.278,<br>prob ( $\geq 5$ ) = 0.595,<br>prob ( $\geq 9$ ) = 0.010, | prob ( $\leq 2$ ) = 0.595,<br>prob ( $\geq 6$ ) = 0.278,<br>prob ( $\geq 10$ ) = 0.003  | prob ( $\leq 3$ ) = 1.000,<br>prob ( $\geq 7$ ) = 0.095,                                |
| $E(x) = 4.5$<br>prob (0) = 0.051,<br>prob ( $\leq 4$ ) = 1.000,<br>prob ( $\geq 8$ ) = 0.148,<br>prob ( $\geq 12$ ) = 0.002 | prob ( $\leq 1$ ) = 0.148,<br>prob ( $\geq 5$ ) = 1.000,<br>prob ( $\geq 9$ ) = 0.051, | prob ( $\leq 2$ ) = 0.343,<br>prob ( $\geq 6$ ) = 0.639,<br>prob ( $\geq 10$ ) = 0.017, | prob ( $\leq 3$ ) = 0.639,<br>prob ( $\geq 7$ ) = 0.343,<br>prob ( $\geq 11$ ) = 0.007, |

**Table 6.** The Customer’s and the Producer’s Risk for Expected and Observed Number of Invalid Items in the Sample (continued)

|                     |                     |                     |                     |
|---------------------|---------------------|---------------------|---------------------|
| E(x) = 5.5          |                     |                     |                     |
| prob ( 0) = 0.029,  | prob (≤1) = 0.080,  | prob (≤2) = 0.194,  | prob (≤3) = 0.392,  |
| prob (≤4) = 0.671,  | prob (≤5) = 1.000,  | prob (≥6) = 1.000,  | prob (≥7) = 0.671,  |
| prob (≥8) = 0.392,  | prob (≥9) = 0.194,  | prob (≥10) = 0.080, | prob (≥11) = 0.029, |
| prob (≥12) = 0.011, | prob (≥13) = 0.004, | prob (≥14) = 0.002, | prob (≥15) = 0.001  |
| E(x) = 6.5          |                     |                     |                     |
| prob ( 0) = 0.018,  | prob (≤1) = 0.045,  | prob (≤2) = 0.110,  | prob (≤3) = 0.234,  |
| prob (≤4) = 0.432,  | prob (≤5) = 0.696,  | prob (≤6) = 1.000,  | prob (≥7) = 1.000,  |
| prob (≥8) = 0.696,  | prob (≥9) = 0.432,  | prob (≥10) = 0.234, | prob (≥11) = 0.110, |
| prob (≥12) = 0.045, | prob (≥13) = 0.018, | prob (≥14) = 0.007, | prob (≥15) = 0.003  |
| E(x) = 8.5          |                     |                     |                     |
| prob ( 0) = 0.007,  | prob (≤1) = 0.016,  | prob (≤2) = 0.037,  | prob (≤3) = 0.082,  |
| prob (≤4) = 0.165,  | prob (≤5) = 0.301,  | prob (≤6) = 0.493,  | prob (≤7) = 0.733,  |
| prob (≤8) = 1.000,  | prob (≥9) = 1.000,  | prob (≥10) = 0.733, | prob (≥11) = 0.493, |
| prob (≥12) = 0.301, | prob (≥13) = 0.165, | prob (≥14) = 0.082, | prob (≥15) = 0.037, |
| prob (≥16) = 0.016, | prob (≥17) = 0.007, | prob (≥18) = 0.003, | prob (≥19) = 0.001  |

Reading examples:

If E(x) = 5.5, x = 1 and the lot is accepted then customer’s risk is 0.080.

If E(x) = 6.5, x = 3 and the lot is accepted then customer’s risk is 0.234.

If E(x) = 4.5, x = 9 and the lot is rejected then producer’s risk is 0.051.

If a sample is investigated and it turns out that the number of invalid items is already so large that the decision to reject the lot is possible, then further investigation of the sample is superfluous. Such an early halt due to many invalid items during investigating the sample is called “curtailment of a sample”. For example, if E(x) is 2.5 and  $\alpha_{\text{producer}}$  is 0.20 then according to Table 6 the lot is rejected if the number of invalid items in the sample is  $x \geq 5$ . If the sample is partly investigated and already 5 invalid items are found, then the lot is already rejected and further investigation of the sample can be saved.

## 8. Multiple steps sampling plans

As already pointed out, statistical methods – whether confidence interval or statistical test – are unable to decide (given a reasonable sample size) with controlled  $\alpha_{\text{Customer}}$  and controlled  $\alpha_{\text{Producer}}$  if  $Q \approx \text{CAQL}$ . The only two solutions to overcome this problem are (i) to enlarge sample size n or (ii) to accept increased  $\alpha_{\text{customer}}$  and  $\alpha_{\text{producer}}$ .

The customer and the producer may agree to increase sample size n if a decision with controlled  $\alpha_{\text{customer}}$  and  $\alpha_{\text{producer}}$  is not possible. Such an agreement should be established beforehand in the sampling plan. A multiple steps sampling plan should fix the maximal number of steps and the sample size of each step in addition to the usual stipulations. For evaluation of a step all previous samples are pooled.

Example: The customer and the producer agree on the following items of the sampling plan: CAQL = 0.05,  $\alpha_{\text{Customer}} = 0.10$ ,  $\alpha_{\text{producer}} = 0.10$  and sample size  $n_1 = 110$ . Further, they agree that if a decision with these conditions is not possible then the sample size is increased by 100 to  $n_2 = 210$ . Thus, E(x) is increased from 5.5 to 10.5.

In most productions such a two-step procedure will be reasonable. However, a three or more step procedure is also possible. If a decision with controlled risks is not possible after one step, then the sample size is increased for the next step.

Sample size of the first step may be rather small with e.g. E(x) = 2.5 or 3.5. In the case

that  $Q$  is rather different from CAQL then a decision is already possible by the first step. For the last step of a multiple sampling plan, we recommend giving up controlling  $\alpha_{\text{customer}}$  and  $\alpha_{\text{producer}}$ , and using the simple procedure described in Section 4 in order to come to a decision with a reasonable sample size.

In a multiple sampling plan more than one test is performed on the same topic. According to formal statistics the  $p$ -value computed for a single test is too small and has to be adjusted for multiple testing, for example, with the Bonferroni-Holm-procedure (Holm, 1979). But all such adjustments give a trend towards not significant. In quality control a decision is wanted. Therefore, we would not worry about waiving the adjustment although it is formally necessary.

## 9. Conclusion

Quality control is a serious matter (Zimon, 2017). In recent years several papers have been published to adapt the standard procedure of acceptance sampling with an attribute to specific situations. Kasprikova et al. (2015) investigated if costs can be saved by inspection of quantitative variables additionally to the inspection of an attribute. Fallanezhad (2015) analysed the consequences of inspection errors. Aslam et al. (2015) compared for the attribute “lifetime sufficient yes/no” an inspection one item after the other with an inspection of several items simultaneously. Rao et al. (2013) analyzed mean lifetime of the products with truncated life time tests. Hsu et al. (2012) presented an economic model for a sequentially sampling up to two subsequent samples. Iacobini (2000) gave a cost analysis approach for lot-by-lot sampling inspection. None of these papers discuss the principle of the standard procedure.

Inspection of a sample with controlled  $\alpha_{\text{customer}}$  and  $\alpha_{\text{producer}}$  has three possible results: (1<sup>st</sup>) lot is accepted, (2<sup>nd</sup>) lot is

rejected, and (3<sup>rd</sup>) a decision is not possible with the given sample size  $n$ . Of course, the customer and the producer may take a decision anyway, but they should be aware of uncontrolled risks.

The standard procedure for acceptance/rejection sampling for attributes (Montgomery 2013; Kreyszig, 2011; Uhlmann, 1982; Woodall, 1997; ISO 2859, 2011; ANSI/ASQ, 2008; and MIL-STD, 1989) is widely used. It is based on an operation characteristic, and delivers sample size  $n$  and acceptance number  $a$ . If  $x > a$ , then the lot is rejected, otherwise it is accepted.

The standard procedure needs two different acceptable quality limits – AQL for the producer and RQL for the customer – with the relationship  $AQL < RQL$ . This is necessary to compute the operation characteristic respective to compute the power of the test. The larger the difference  $RQL - AQL$  the smaller the sample size may be. We think that the two different acceptable quality levels and their necessary relationship  $AQL < RQL$  are illogical, or at least difficult to explain to practitioners.

From the statistical point of view the standard procedure is a test with a one-sided hypothesis (null hypothesis:  $Q \leq AQL$ , alternative  $Q > AQL$ ). If the test is significant ( $x > a$ ) then  $Q > AQL$  with controlled  $\alpha$ , and the lot is correctly rejected. If the test is not significant ( $x \leq a$ ) then the lot is accepted. But a non-significant test is not a proof that the null hypothesis is correct. A non-significant test guarantees only that  $Q \leq RCL$  with controlled  $\beta$ . If  $AQL \leq Q < RQL$  this is called an “indifferent” lot by Kreyszig (2011). Then the test is often not significant with uncontrolled  $\beta$  and acceptance of the lot is not justified. “Absence of evidence is not evidence of absence” (Altman and Bland, 1995). Thus, the standard procedure results in acceptance of the lot even though a decision with controlled customer’s risk is impossible due to an insufficient sample size

for the actual effect size. We feel that this is unfair for the customer.

The standard procedure ignores that if  $AQL \leq Q < RQL$  or  $Q \approx CAQL$ , respectively, then a decision with controlled risks needs a very large sample size up to the inspection of the entire lot. This holds in general, and is not a matter of a certain statistical procedure. We are convinced that statisticians should be honest and should not use insufficient information to accept a lot as it is done by the standard procedure.

In contrast to the standard procedure we (1<sup>st</sup>) enable an acceptable quality limit common for the customer and the producer, (2<sup>nd</sup>) give a first orientation on sample size by  $E(x)$ , (3<sup>rd</sup>) offer a simple but unbiased procedure for acceptance/rejection of the lot, (4<sup>th</sup>)

illustrate the situation by a confidence interval for  $Q$ , (5<sup>th</sup>) propose a statistical test with a two-sided hypothesis, and (6<sup>th</sup>) make a decision only if the test is significant. If sample size is insufficient to prove the actual effect size we acknowledge that a decision with controlled  $\alpha_{customer}$  or  $\alpha_{producer}$  is impossible. It is up to the customer and the producer to agree beforehand to increase sample size in a multiple steps sampling plan, or to take a decision without controlled  $\alpha_{customer}$  or  $\alpha_{producer}$ .

**Acknowledgement:** We thank Dr. Maximilian Swerev, PAUL HARTMANN AG, Heidenheim, Germany. Due to his questions we became sensitive to lot-by-lot acceptance/rejection sampling.

## References:

- Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman&Hall/CRC.
- Altman, D. G., & Bland, J. M. (1995). Statistics notes: Absence of evidence is not evidence of absence. *British Medical Journal*, 311, 485. <http://dx.doi.org/10.1136/bmj.311.7003.485>
- Altman, D. G., Machin, D., Bryant, T. N., & Gardner, M. J. (2000). *Statistics with confidence* (2nd ed). British Medical Journal (BMJ)-Books.
- ANSI/ASQ Z1.4 (2008). Sampling procedures and tables for inspection by attributes.
- Aslam, M., Azam, M., & Jun, C.-H. (2015). Acceptance sampling plans for multiple-stage process based on time-truncated Weibull distribution. *The International Journal of Advanced Manufacturing Technology*, 79, 1779-1785. doi: 10.1007/s00170-015-6938-0
- Bland, M. (2000). *An introduction to medical statistics*. New-York: Oxford University Press.
- Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26, 403-413.
- Gaus, W., & Muche, R. (2017). *Medizinische Statistik* (2nd ed). Stuttgart: Schattauer Verlag.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65-70.
- Hsu, L.-F., & Hsu, J.-T. (2012). Economic design of acceptance sampling plans in a two-stage supply chain. *Advances in Decision Sciences*, 1-14. doi: 10.1155/2012/359082
- Iacobini, A. (2000). Generalized cost and gain functions in lot-by-lot sampling inspection. *Journal of the Italian Statistical Society*, 1, 97-105.
- Ilakovac, V. (2009). Statistical hypothesis testing and some pitfalls. *Biochemica Medica*, 19, 10-16. <http://dx.doi.org/10.11613/BM.2009.002>
- ISO 2859 (2011). Sampling procedures for inspection by attributes, Part 1: Sampling schemes indexed by acceptance quality limit (AQL) for lot-by-lot-inspection.

- Kasprikova, N., & Klufa, J. (2015). AOQL sampling plans for inspection by variables and attributes versus the plans for inspection by attributes. *Quality Technology and Quality Management*, 12, 133-142.
- Kreyszig, E. (2011). Acceptance sampling. In *Advanced engineering mathematics* (10th ed.) (pp. 1092-1096). John Wiley & Sons.
- McHugh, M. L. (2008). Power analysis in research. *Biochemia Medica*, 18, 263-274.
- MIL-STD-105E (1989). *Sampling procedures and tables for inspection by attributes*. Washington, D.C: Department of Defense.
- Montgomery, D. C. (2013). Basic acceptance sampling procedures. In *Statistical quality control. A modern introduction* (7th ed.) (pp. 649-687) John Wiley & sons.
- Rao, B. S., Kumar, C. S., & Rosalah, K. (2013). Acceptance sampling plans from life tests based on percentiles of half normal distribution. *Journal of Quality and Reliability Engineering*, 1-7. <http://dx.doi.org/10.1155/2013/320469>
- Uhlmann, W. (1982). *Statistische Qualitätskontrolle* (2nd ed.). Stuttgart: B. G. Teubner.
- Woodall, W. H. (1997). Control charts based on attribute data: bibliography and review. *Journal of Quality Technology*, 29, 172-183.
- Zimon, D. (2017). The impact of TQM philosophy for the improvement of logistic processes in the supply chain. *International Journal for Quality Research*, 11(1), 3-16. doi: 10.18421/IJQR11.01-01

---

**Wilhelm Gaus**

University of Ulm,  
Department for  
Epidemiology and Medical  
Biometry  
89070 Ulm  
Germany  
[wilhelm.gaus@uni-ulm.de](mailto:wilhelm.gaus@uni-ulm.de)

**Rainer Mucbe**

University of Ulm,  
Department for  
Epidemiology and Medical  
Biometry  
89070 Ulm  
Germany  
[rainer.mucbe@uni-ulm.de](mailto:rainer.mucbe@uni-ulm.de)

**Benjamin Mayer**

University of Ulm,  
Department for  
Epidemiology and Medical  
Biometry  
89070 Ulm  
Germany  
[benjamin.mayer@uni-ulm.de](mailto:benjamin.mayer@uni-ulm.de)

---

## Appendix:

Symbols concerning the lot (population) are upper case letters. Symbols concerning the sample are lower case letters. Probabilities are written with Greek letters.

- N**      Number of items in the lot = size of population
- X**      Number of invalid items in the lot (usually unknown),  $0 \leq X \leq N$
- Q**      Proportion of invalid items in the lot (usually unknown),  $Q = X / N$
- AQL**    Acceptable quality limit or accepted quality level. This variable is used in the standard procedure. It is the maximal proportion of invalid items in the lot accepted by the producer. It has to be established and defined and is therefore known.  $0 < AQL < 1$
- RQL**    Rejectable quality level. This variable is used in the standard procedure. It is the maximal proportion of invalid items in the lot accepted by the customer. It has to be established and defined and is therefore known.  $0 < AQL < RQL < 1$
- CAQL**   Common acceptable quality limit. The maximal proportion of invalid items in the lot accepted by the producer as well as by the customer. It has to be established and defined and is therefore known.  $0 < CAQL < 1$
- 
- n**      Number of items in the sample = sample size,  $0 < n < N$
- x**      Number of invalid items in the sample. In statistical terms is x an observed frequency.  $0 \leq x \leq n$
- a**      Acceptance number. This variable is used in the standard procedure. If the number of invalid items in the sample is less than or equal to this number ( $x \leq a$ ) then the lot is accepted, if  $x > a$  then the lot is rejected.
- E(x)**   Expected number of invalid items in the sample. It relates to the sample and should therefore be written in lower case letters. But in statistics E for “expected” is written traditionally with upper case letter.
- 
- $\alpha$**      In general the probability that a statistical test is falsely significant. In the standard procedure it is the producer’s risk.
- $\beta$**      In general the probability that a statistical test is falsely not significant. In the standard procedure it is the customer’s risk.
- The customer’s risk is the probability that a lot is accepted although  $Q > CAQL$ . (In the standard procedure  $Q > RQL$ ) Customer’s risk = prob (lot is accepted |  $Q > AQL$ ). In statistical shorthand it is the p-value of the test for the side  $Q > CAQL$ .
- $\alpha_{customer}$**    Maximal risk accepted by the customer. Often  $\alpha_{customer} = 5\%$ ,  $10\%$  or  $20\%$  is selected. In statistical terms is  $\alpha_{customer}$  the level of significance for the side  $Q > CAQL$ .
- The producer’s risk is the probability that a lot is rejected although  $Q < CAQL$ . (In the standard procedure  $Q \leq AQL$ ) Producer’s risk = prob (lot is rejected |  $Q < AQL$ ). In statistical shorthand it is the p-value of the test for the side  $Q < CAQL$ .
- $\alpha_{producer}$**    The maximal risk accepted by the producer. Often  $\alpha_{producer} = 5\%$ ,  $10\%$  or  $20\%$  is selected. In statistical terms is  $\alpha_{producer}$  the level of significance for the side  $Q < CAQL$ .