KM 15(1984)
pag 110-122

The ratio of true to estimated values when determining the number of items in a batch by means of weighing

by Ir. P.M. Upperman *)

Abstract

When weighing is used as a means of determining the number n of items in a batch, we have three sources of error as regards the estimate $\hat{n}$ of n.

- weighing errors,
- an error in the average weight of an item,
- sampling errors as a result of weight variation between items.

A formula is derived giving limits for $n/\hat{n}$ as a function of the errors.

In section 4 of this report a procedure for weighing, counting and calculating is proposed.

---

*) Nederlandse Philips Bedrijven B.V., Centre for Quantitative Methods, Building VN-706, 5600 MD Eindhoven, The Netherlands. Tel. 040 - 735923.

1. Introduction

It often happens in mass production that small components are made in very large quantities. It is not uncommon for a batch to contain more than 1,000,000 items. It is then out of the question to count the number of items in each batch. It is then common practice to estimate the number of items by weighing.
It is, however, unavoidable that differences occur between the estimated and the true number of items. The allowable difference is a matter of negotiation between buyer and vendor.

In order to arrive at agreement, it is necessary to know the relation of weighing and sampling errors with differences between the estimated and true number of items.

This relation will be derived in section 6. It will be shown that this relation can be used to prescribe a correct weighing procedure ensuring that allowable differences between estimated and true number will rarely be exceeded.

2. Acknowledgements

The author wishes to thank Mr. A.T.J. Slaets, Quality Department PMF Eindhoven, for drawing his attention to the problem in question and Dr. J.P.M. de Kroon, Centre for Quantitative Methods, for his useful remarks, which improved my original notes considerably.

3. Definitions and assumptions

g : the weight of one item. We assume that for each batch the variable g has a normal distribution.

μ : the population mean of g.

σ : the population standard deviation of g.

σ/μ : the coefficient of variation of g.

m : the number of items taken to determine an estimate of μ.

$\bar{g} = \frac{1}{m} \sum_{i=1}^{m} g_i$: an unbiased estimate of μ.

n : the true unknown number of items in a batch.

w : the weight of n items in a batch.

ε : a random weighing error due to the weighing instrument. The value ε has a normal distribution with mean zero and variance σ_ε^2.

σ_ε/w : The coefficient of variation as regards random measuring error. The value of σ_ε/w is inversely proportional to the quality of the weighing instrument, i.e. the smaller σ_ε/w, the better the instrument.

$\hat{n}$: an estimate of n.

$E(x)$: the expected or mean value of a random variable x.

$Var(x)$: the variance of a random variable x.

4. The proposed weighing, counting and calculating procedure

It is assumed that σ_ε/w and σ/μ are known before weighing starts. The value of σ_ε/w can usually be obtained from specifications, whereas σ/μ could be estimated by weighing a number of items separately.

Formula (11), section 6, can now be calculated to ensure acceptable limits for $n/\hat{n}$.

We then decide on the value of w which is determined by the batch size and the capacity of the weighing instrument.

One takes so many items that a value w is read on the instrument. We then choose m items from those which have just been weighed and measure the total weight $\sum_{i=1}^{m} g_i$ of these m items. We then calculate:

$$\bar{g} = \frac{1}{m} \sum_{i=1}^{m} g_i \qquad (1)$$

and

$$\hat{n} = \frac{w}{\bar{g}} \qquad (2)$$

It is worth noting that, if w and $\bar{g}$ are found using the same weighing instrument, the use of formula (2) completely eliminates a systematic weighing error which is a fraction of the reading on the weighing instrument.

5. The variance-covariance structure of w and $\bar{g}$ and the expected values of w, $\bar{g}$ and $\hat{n}=w/\bar{g}$

We consider batches containing an unknown number of n items. We wish to estimate n. If we consider the weight of these batches we have a random variable

$$w = \sum_{i=1}^{n} g_i + \varepsilon$$

from a normal distribution with:

mean: $E(w) = n\mu$ (3)

variance: $Var(w) = n\sigma^2 + \sigma_\varepsilon^2$ (4)

From (3) it appears that w is an unbiased estimator of $n\mu$.

When we take a known number of m items from the n mentioned above we have a total weight equal to

$$\sum_{j=1}^{m} g_j + \varepsilon' \quad \text{where } \varepsilon' \text{ is the weighing error.}$$

We then calculate:

$$\bar{g}' = \frac{1}{m}\left(\sum_{j=1}^{m} g_j + \varepsilon'\right)$$

or $$\bar{g}' = \bar{g} + \frac{\varepsilon'}{m}$$

Obviously:

Mean of $\bar{g}'$: $E(\bar{g}') = E(\bar{g}) = \mu$

Variance of $\bar{g}'$: $Var(\bar{g}') = \frac{\sigma^2}{m} + \frac{\sigma^2_{\varepsilon'}}{m^2} = \frac{1}{m}\left(\sigma^2 + \frac{\sigma^2_{\varepsilon'}}{m}\right)$

In practical situations we have that $\sigma^2_{\varepsilon'}/m$ is very much smaller than σ^2 and we shall therefore ignore $\sigma^2_{\varepsilon'}/m$ and consequently also ε'/m.

Therefore:

$$\bar{g}' = \bar{g} = \frac{1}{m} \sum_{j=1}^{m} g_j$$

and $E(\bar{g}') = E(\bar{g}) = \mu$ (5)

$$\mathrm{Var}(\bar{g}') = \mathrm{Var}(\bar{g}) = \frac{\sigma^2}{m} \qquad (6)$$

In section 4 it was stated that the m items where taken out of the n. We may therefore write:

$$w = g_1 + g_2 + \ldots + g_m + g_{m+1} + \ldots + g_n + \varepsilon \qquad (7)$$

$$\bar{g} = \frac{1}{m}(g_1 + g_2 + .. + g_m) \qquad (8)$$

From (7) and (8) it follows that:

Covariance $(w, \bar{g}) = \frac{1}{m}\sigma^2 + \frac{1}{m}\sigma^2 + \ldots + \frac{1}{m}\sigma^2 = \sigma^2$

We shall later see that the positive correlation between w and $\bar{g}$ decreases the bias and the width of the confidence interval of $(n/\hat{n})$.

An estimate $\hat{n}$ of n will be:

$$\hat{n} = \frac{w}{\bar{g}}$$

since w is an estimate of $n\mu$, see (3) and $\bar{g}$ is an estimate of μ, see (5).

We shall now calculate $E(\hat{n})$.

$$E(\hat{n}) = E(\frac{w}{\bar{g}}) = E\left[\frac{w}{\mu(1 + \frac{\bar{g} - \mu}{\mu})}\right] = \frac{1}{\mu} E\left[\frac{w}{1 + \frac{\bar{g} - \mu}{\mu}}\right]$$

Using the power series expansion $\frac{1}{1 + x} = 1 - x + x^2 + \ldots$ for small $x < 1$, we have:

$$E(\hat{n}) = \frac{1}{\mu} E\left[w\{1 - \frac{\bar{g} - \mu}{\mu} + (\frac{\bar{g} - \mu}{\mu})^2\}\right]$$

$$= \frac{1}{\mu} E(w) - \frac{1}{\mu} E \{(w - n\mu)(\frac{\bar{g} - \mu}{\mu}) - (w - n\mu)(\frac{\bar{g} - \mu}{\mu})^2\} +$$

$$+ \frac{1}{\mu} E \{ -n\mu (\frac{\bar{g} - \mu}{\mu}) + n\mu (\frac{\bar{g} - \mu}{\mu})^2\}$$

Ignoring small terms of order 3 we have:

$$E(\hat{n}) = \frac{1}{\mu} n\mu - \frac{1}{\mu} \{\frac{1}{\mu} \operatorname{cov}(w,\bar{g}) - 0\} + \frac{1}{\mu} \{0 + \frac{n\mu}{\mu^2} \operatorname{var}(\bar{g})\}$$

$$E(\hat{n}) = n - \frac{\sigma^2}{\mu^2} + \frac{n}{\mu^2} \frac{\sigma^2}{m} = n\{1 + (\sigma^2/\mu^2)(\frac{1}{m} - \frac{1}{n})\} \qquad (9)$$

If w and $\bar{g}$ are not correlated, we can easily prove that $E(\hat{n}) = n(1 + \frac{\sigma^2/\mu^2}{m})$

The bias has therefore decreased by an amount $\frac{\sigma^2/\mu^2}{n}$ because of the correlation between w and $\bar{g}$.

The bias will in practice be very small, as the following numerical example shows.

Example

$\sigma/\mu = 0.05 \qquad m = 100 \qquad n \simeq 200{,}000$

$$E(\hat{n}) = n\{1 + 0.05^2 (\frac{1}{100} - \frac{1}{200{,}000})\} = 1.000025\ n$$

6. The relation between limits for $(n/\hat{n})$ and (σ/μ), $\sigma_\varepsilon/w)$

We may write $n = \frac{n\mu}{\mu}$

We already saw that an estimate $\hat{n}$ of n equals

$\hat{n} = w/\bar{g}$ since w is an estimate of $n\mu$ and $\bar{g}$ is an estimate of μ.

Using Fieller's theorem (see [1], [2] and [3]) we introduce the variable:

$z = n\bar{g} - w$

z has a normal distribution with

$$E(z) = nE(\bar{g}) - E(w) = n\mu - n\mu = 0$$

$$Var(z) = n^2 Var(\bar{g}) + Var(w) - 2n\ cov(\bar{g},w) = \frac{n^2\sigma^2}{m} + n\sigma^2 + \sigma_\varepsilon^2 - 2n\sigma^2$$

Hence,

$$Var(z) = n^2\sigma^2/m - n\sigma^2 + \sigma^2{}_\varepsilon$$

It follows that

$$u = \frac{n\bar{g} - w}{\sqrt{(n^2\sigma^2/m - n\sigma^2 + \sigma^2{}_\varepsilon)}}$$ has a N(0,1) distribution.

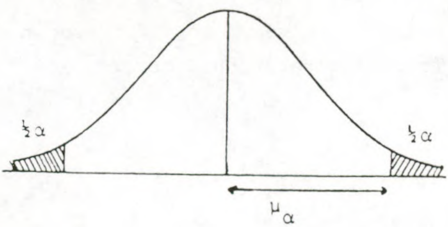

Figure 1 The N(0,1) distribution

A (1-α)% prediction interval for u is, see Figure 1,

$$-u_\alpha < \frac{n\bar{g} - w}{\sqrt{(n^2\sigma^2/m - n\sigma^2 + \sigma^2_\varepsilon)}} < + u_\alpha$$

These inequalities can, after some simple calculations, be transformed into a (1-α)% confidence interval for n with limits

$$n = \frac{\hat{n} - \frac{u_\alpha^2\sigma^2}{2\bar{g}} \pm \hat{n}\sqrt{\left[(1 - \frac{u_\alpha^2\sigma^2}{2\hat{n}\bar{g}^2})^2 - (1 - \frac{u_\alpha^2\sigma_\varepsilon^2}{w^2})(1 - \frac{u_\alpha^2\sigma^2}{m\bar{g}^2})\right]}}{(1 - \frac{u_\alpha^2\sigma^2}{m\bar{g}^2})} \quad \text{or}$$

$$\frac{n}{\hat{n}} = \frac{1 - \frac{u_\alpha^2\sigma^2}{2\hat{n}\bar{g}^2} \pm \sqrt{\left[(1 - \frac{u_\alpha^2\sigma^2}{2\hat{n}\bar{g}^2})^2 - (1 - \frac{u_\alpha^2\sigma_\varepsilon^2}{w^2})(1 - \frac{u_\alpha^2\sigma^2}{m\bar{g}^2})\right]}}{(1 - \frac{u_\alpha^2\sigma^2}{m\bar{g}^2})} \qquad (10)$$

Each of the terms in (10) behind the number 1 is very close to zero in practice. We shall therefore use the following relations for $|x|$ and $|y| < 1$ and close to zero.

$$(1 + x)^2 \approx 1 + 2x$$
$$1/(1 + x) \approx 1 - x$$
$$(1 + x)(1 + y) \approx 1 + x + y$$

Hence

$$\frac{n}{\hat{n}} \approx 1 + \frac{u_\alpha^2\sigma^2}{\bar{g}^2}\left(\frac{1}{m} - \frac{1}{2\hat{n}}\right) \pm u_\alpha\sqrt{\left\{\frac{\sigma^2}{\bar{g}^2}\left(\frac{1}{m} - \frac{1}{n}\right) + \frac{\sigma_\varepsilon^2}{w^2}\right\}}$$

replacing $\bar{g}$ by μ gives

$$\frac{n}{\hat{n}} \approx 1 + \frac{u_\alpha^2\sigma^2}{\mu^2}\left(\frac{1}{m} - \frac{1}{2\hat{n}}\right) \pm u_\alpha\sqrt{\left\{\frac{\sigma^2}{\mu^2}\left(\frac{1}{m} - \frac{1}{\hat{n}}\right) + \frac{\sigma_\varepsilon^2}{w^2}\right\}} \qquad (11)$$

(11) now gives a $(1-\alpha)\%$ confidence interval for (n/n).

It can easily be proved that the terms

$-\frac{1}{2\hat{n}}$ and $-\frac{1}{\hat{n}}$ in (11) change into

$+\frac{1}{2\hat{n}}$ and $+\frac{1}{\hat{n}}$ respectively when w and $\bar{g}$ are not correlated.

It follows that the proposed weighing procedure which resulted in $\mathrm{cov}\,(w,\bar{g})$ being equal to σ^2 not only decreases the width of the confidence interval of $(n/\hat{n})$, but also centres it more closely around one.

7. A numerical example

We assume that we have a good quality weighing instrument with $\sigma_\varepsilon/w = 0.001$

We further have
$\sigma/\mu = 0.05$ and also

$m = 100$ and $n \approx 200{,}000$

We use a 95% confidence interval and therefore $u_\alpha = 2$. From (11) we find

$$\frac{n}{\hat{n}} = 1 + 4*0.0025\left(\frac{1}{100} - \frac{1}{400{,}000}\right) \pm 2\surd\left(0.05^2\left(\frac{1}{100} - \frac{1}{200{,}000}\right)+0.001^2\right) \quad (12)$$

$$\frac{n}{\hat{n}} = 1 + 0.0001 - 25 * 10^{-9} \pm 0.0102 \quad (13)$$

$$0.9899 < \frac{n}{\hat{n}} < 1.0103 \quad \text{or} \quad 0.9898\ n < \hat{n} < 1.0102\ n$$

We thus see that differences between n and $\hat{n}$ can therefore amount to 1%.

Using the method discussed in [4] we find $0.9902\ n < \hat{n} < 1.0098\ n$.

The practical difference between the methods is therefore negligible.

8. Recommendations and conclusions

From (12) and (13) it appears that the term $0.05^2/100$, the variance of $\bar{g}$, the estimate of μ, contributes the most to the width of the interval for $n/\hat{n}$.

It is therefore wrong practice to use a fixed value μ for each code number, when μ may vary from batch to batch. It could well happen that μ varies through wear and tear of machines when many small metal components are made on those machines.

Once it has been decided to determine an estimate of μ, this estimate can be made as accurate as required by using a sufficiently large value of m. On the face of it, it seems rather contradictory to count and weigh items in a weighing procedure which has been designed to avoid counting. However, the value of m can be so much smaller than n and counting can be greatly facilitated by using a special flat paddle with, say, 100 depressions so that it can contain only 100 items. In this way, the counting and weighing of m items hardly involves any extra costs. From (12) and (13) it appears that, for the given example, a narrow interval for $n/\hat{n}$ can most easily be obtained by increasing the value of m.

In most practical situations the value of n is much larger than m. We therefore may omit $\hat{n}$ in the right-hand side of equation (11). We then find:

$$\frac{n}{\hat{n}} = 1 + \frac{u_\alpha^2 \sigma^2}{\mu^2 m} \pm u_\alpha \sqrt{\left\{ \frac{\sigma^2}{\mu^2 m} + \frac{\sigma_\varepsilon^2}{w^2} \right\}}$$

Literature

[1] Fieller, E.C., "The distribution of the index in a normal bivariate population"
Biometrika 24 (1932), 428-40.

[2] Finney, D.J., "Probit Analysis"
Cambridge University Press, 1971.

[3] Marsaglia, G., "Ratios of normal variables and ratios of sums of uniform variables"
American Statistical Association Journal, March 1965, 193-204.

[4] Nelson, P.R., "Counting a large number of items using the weight of a small sample".
Technometrics, Vol. 25, No. 2, May 1983, 173-177.

Ontvangen: 28-11-83