143

Some theory on the influence of the inspector and environmental conditions, with an example

by

W. EDWARDS DEMING

and

MORRIS H. HANSEN

Reprinted from Statistica Neerlandica Vol. 26/3, 1972 pp. 101–112

Some theory on the influence of the inspector and environmental conditions

by W. Edwards Deming *

GENERAL BACKGROUND

Purpose of this paper

The purpose here is to present some theory by which to measure the influence of the inspector, and to show how sensitive the results of inspection are to this influence. The theory also indicates what records to keep in order to improve performance of inspectors and the supervision of inspection, whether inspection be carried out by visual inspection, or by use of instruments, or by automatic recording devices.

All results are conditional

Any result is the end-product of a chain of operations, and are dependent on a host of conditions. Alteration of any of the conditions may affect the result. A record of the conditions under which a test is made is accordingly as important as the numerical result itself. Some examples of conditions are listed below.

Description of the particular material that was subjected to sampling and testing (in statistical language, the frame, the lot, source, date manufactured, and other information).

The method of sampling; whether it was
- by use of random numbers or equivalent
- by judgment
- by convenience

How the sample was prepared for test
The test-method
How the test-method was used
- the particular apparatus used
- its condition
- who used it, and how
- the randomness of successive results

Temperature
Humidity
Date of test
What went wrong with the experiment? What could be some of the effects on the results?

* Consultant in Statistical Surveys, Washington, 20016.

In many cases, a complete record of the tests may be necessary. The mean and standard deviation of a set of measurements is not an acceptable substitute for the measurements themselves except in the rare circumstance when the recorded measurements are in statistical control.

The more we know about the conditions under which a result was obtained, and how these conditions could affect it, the more useful the result becomes. It is not always obvious that some piece of information, such as the name of the observer, or the date, may be very important data of the experiment. Failure to record some piece of information about the environmental conditions of the test may greatly reduce the usefulness of the test.

Source of influence of the inspector

The problem of achieving uniform performance and valid results in testing is a general one, and of great importance. The stamp of quality or of rejection, placed on an item or on an assembly, is too often more dependent on inspection than on production.

The work of an inspector, in the absence of supervision with the aid of statistical methods, will exhibit built-in similiarity between items inspected. As the total variance between all items is fixed, this influence raises the variance between inspectors in excess of expectation on a random basis. The problem is complex, being the result of interaction of the inspector, the item inspected, the instruments used, and the so-called test-method and the criteria that he is supposed to follow. The exact mechanism of these interactions is not well understood. The point of this paper is that this influence exists, whether we understand it or not, and that it is important in the control of quality to measure this influence and to try to reduce it by training and re-training.

It is too often assumed that the results of inspection, or the results of an assay, are absolute, not subject to question; that there is a correct value for a test, such as measurement of a dimension, or of hardness, viscosity, or for visual inspection or use of a gauge to determine whether an item is defective. Such an assumption at the end of the production-line, or at the beginning in the receipt of materials, may well undo what is otherwise a good program of quality-control in the production-line. It is better to say that there are, in many instances, acceptable methods for testing, and that these methods must be properly used and continually subjected to statistical criteria of stability.

Replacement of man by automatic testing and recording devices does not eliminate the problems of uniformity and validity, nor possibly even reduce them. The truth is that all we ever have as a result of inspection, whether by man or by machine, is marks on paper. We must presume that a change of inspectors, or replacement of some of them, or a change of apparatus, will give different results, even on the same items.

These dreary truths do not mean that data of inspection are no good nor that the situation is hopeless. Rather, they mean that to use data one must understand the process that produces them. To obtain greater reliability in the method of test, and more useful results, is as important a problem in the statistical control of quality as any other.

Theory

Let there be

N	items in the frame
n	items to be inspected
m	number of inspectors on the job
$g = n/m$	items per inspector

Let

σ^2	be the variance between the $x =$ characteristic of the N items in the frame
ϱ	the average intraclass correlation between the g items within inspectors.

Suppose that we draw at random g items from the frame and hand them to Inspector 1; g more items and hand them to Inspector 2; g more items to Inspector 3; etc., these inspectors being drawn at random from a pool of inspectors. Compute

$\bar{x}_i$ the mean for Inspector i

$\bar{x}$ the overall mean for all m inspectors

Then it is a fact that

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{mg}\{1+(g-1)\varrho\} \tag{1}$$

It is amazing to observe that some slight influence of the inspector and of the environmental conditions may contribute more variance to the result than any other source of uncertainty – more perhaps than all other sources combined. Thus, let $\varrho = .04$ and $g = 51$ items per inspector. Then by Eq. 1,

$$\begin{aligned} \sigma_{\bar{x}} &= \frac{\sigma^2}{mg}\{1+(51-1)\times .04\} \\ &= \frac{\sigma^2}{mg}\{1+2.0\} = 3\frac{\sigma^2}{mg} \end{aligned} \tag{2}$$

$\bar{x}$ being an estimate of the fraction defective in the frame.

Now σ^2/mg would be the variance of $\bar{x}$ if mg inspectors were to inspect 1 item each. Obviously, an intraclass correlation as low as $\varrho = .04$, with $g = 51$ items per inspector, has trebled the variance of $\bar{x}$.

Intraclass correlation as low as $\varrho = .04$ is an achievement in the training of inspectors. It is seldom realized. It is in fact so small that it would be in most practice deemed inconsequential. Yet the above theory tells us that the inspector himself, even with $\varrho = .04$, could be the greatest single contributing factor to the total variance. In my own experience, this is unfortunately too often a pretty good description of exactly what is taking place.

We may thus observe from such simple theory how important it is to regard inspection as a process, and how important it is to be able to keep continuous records on the variance between inspectors, and to use statistical measures and methods to improve uniformity of performance.

Calculation of the variance between inspectors as a tool of supervision

It is common practice to set up elaborate programs of training and supervision. How effective are they? What the purchaser of product should pay for is results, not for hopes nor for expenses. What we need is statistical measures of uniformity, and statistical aids to learning and to supervision.

It is a fairly simple matter in some kinds of work to lay it out in replicated subsamples, and to allot to each inspector a random portion of each subsample. Then each inspector has a valid sample of the frame. The average variance between items within inspectors can be calculated from the average variance between subsamples. The variance between inspectors would be calculated in the usual way. Then altogether we have

σ_b^2 the variance between the m average items for the m inspectors

σ_w^2 the average variance between items within inspectors

Then

$$F = g\sigma_b^2/\sigma_w^2 \tag{3}$$

provides a measure of the effectiveness of the training and supervision. A small value of F signifies that the training and supervision were effective – at least that there was no statistical evidence to the contrary. A large value of F signifies that something went wrong: the performance of some inspector is out of line with the others, being in the nature of an outlier. Steps should be taken to discover what went wrong and to institute corrective action.

A still better plan – simpler and statistically more efficient, is to use a control chart. The results of Inspector i over a given interval of time (one hour, one day) furnish $\bar{x}_i$. Then we may use $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m$ as observed values; plot the mean $\bar{x}$ and the chart for standard deviation or for range calculated in the usual way.

Still better, plot each inspector's results, hour by hour. Observe whether any inspector is perpetually high, or low, or too uniform.

Control items

An auxiliary statistical tool is the use of control items. One item in (e.g.) 100 might be selected at random as a control item. If there are no more than 4 or 5 inspectors, every inspector would test every control item. The comparison of the results, and study of relative positions of the inspectors, alerts the supervisor to trends and need of re-training. (Only the result of the inspector that belonged to a control item is fed into the calculation of the process average.) We assume here that the test is non-destructive.

The increase in cost, for 5 inspectors and 1 control item per 1000, is about 5%. The benefits constitute huge dividends on the investment.

If there are more than 5 inspectors, one may allot 4 or 5 inspectors to each control point in randomized balanced blocks.

An example

by MORRIS H. HANSEN and W. EDWARDS DEMING

Description of the study

This example deals with inspection of pipe buried in the ground for the transmission of gas over a large area. The total property inspected is valued at about $ 640,000,000. Inspection took place at 640 points. The aim of the inspection was to estimate the overall physical condition of the pipe. The procedures was to dig a hole to expose a 3-foot section of pipe at each point of inspection, and to clean the pipe with a sand blast. Inspection consisted of measuring the depth of the 10 deepest pits in an exposed section; then using tables of corrosion and visual observation to arrive at the per cent condition of the pipe at each point of inspection.

The engineering-drawings maintained by the Company show for each piece of pipe its location, date of placement, size and type of pipe, type of covering, and estimated cost of installation at today's prices.

The sampling unit was a dollar. A random dollar indicated, with great accuracy, the position of a hole. (The effect of any random failure to dig holes at the points indicated is included and reflected in the standard error.)

Every dollar of investment in the pipe had the same chance of selection as any other. This was a great convenience in the analysis of results, as the weights of all inspections are equal.

The work was laid out in zones of 2 holes in each section of pipe of value $ 2,000,000, one hole randomly designated to Subsample 1, and one hole to Subsample 2 in each zone. There were 4 inspectors, A, B, C, D, allotted at random to the 4 holes in each subsample in each successive 4 zones. Three possible allocations of the many possible random allocations within sets of 4 zones are shown below.

Three examples of allotment of inspectors by subsample

1	2	1	2	1	2
A	A	B	C	D	B
B	C	A	B	C	A
C	D	C	D	A	D
D	B	D	A	B	C

One hole in 20 was a control-hole at which all four inspectors worked independently. (Only the result of the inspector allotted to that hole went into the compilation of final results.) There were 29 control-holes. The purpose of the control-holes was to provide the supervisor with a statistical control, to hold the inspectors in line so that their work could be combined. The supervisor had the privilege and duty to call a halt and to re-train his men or to hold a conference at any time, whether at a control-hole or in between.

The theory adopted here is that the four inspectors were drawn from a supply of a large number of men competent and willing to undertake the special training and discipline necessary for this work. The inspection was in charge of a supervisor of repute. There was, of course, the requirement that the inspectors in their training would make inspections of a variety of kinds of pipe under various stages of deterioration.

We pause here to ponder on what we should do if the variance σ_b^2 between inspectors had turned out to be large. The most usual cause of a large value of σ_b^2 is one inspector out on a limb, differing greatly from the other three. One might contemplate the necessity, under this painful circumstance, to throw out the work of the one man that is out of line, and to base the final results on the work of the other three.

However, exclusion of the work of one inspector may not be made purely on statistical grounds. A statistical test can only call our attention to the possibility that this man may be far out of line. We require cause, based on engineering grounds, such as the man's experience and his record of ability to subject himself to the special training and discipline required for this job.

Fortunately, as we have seen, there was no need for difficult decisions of this kind, undoubtedly because of help to the supervisor from continual study of the control-points.

Notation

We now make some calculations of the variance of the overall results. First, some definitions and notation.

i	the hole
j	the inspector
k	the inspection made by Inspector j at Hole i. Here, k takes on only the value 1 and will often be omitted. We regard the result of a particular inspection (the k-th) as a sample of one from many such possible results by an inspector.
x_{ijk}	the result of this inspection.
$\bar{x}_j$	the average result of the work of Inspector j over the sample of holes that he inspected.
$a = Ex_{ijk}$	the average of all possible inspections over all possible samples.
$a_{ij} = \underset{ij}{E}\, x_{ijk}$	the average result of all possible inspections by Inspector j at Hole i. $\underset{ij}{E}$ is the conditional inspected value holding i and j constant.
$a_i = \underset{i}{E}\, x_{ijk}$	the average result at Hole i of all possible inspections by all possible inspectors. $\underset{i}{E}$ is the conditional inspected value holding i fixed.

x	the average result over all inspections as carried out.
$a_j = E_j x_j$	the conditional expected value for Inspector j for his particular work assignment of holes in this sample.
n	the total number of holes in the sample, 640.
$\bar{n} = \frac{1}{4}n$	the work-load of each inspector
$\sigma^2 = E(x_{ijk}-a)^2$	the total variance
$\sigma_w^2 = E(x_{ijk}-a_{ij})^2$	the variance within inspectors
$\sigma_b^2 = E(a_j-a)^2$	the variance between inspectors in this study, in general dependent on the size and type of work-assignment.
$\sigma_B^2 = E(a_{ij}-a_i)^2$	a second definition of the variance between inspectors.

σ_B^2 is, in general, different from σ_b^2. The two variances between inspectors would be equal if each inspector at every hole had a constant expected difference from the expected average of all inspectors. This equality need not be assumed. In this study, as will be seen, the estimated effect is obtained for each variance, and it turns out that neither one makes an important contribution to the total variance, wherefore we need not be concerned about the difference between them. We define also

$\sigma_s^2 = \underset{h}{E}\,E(a_i-a)_h^2$ the sampling variance for a simple random sample of holes within zones, where $\underset{h}{E(a_i-a)_h^2}$ is the conditional expected value for the holes within Zone h. E signifies expectation over all holes.

Then

$$\sigma^2 = \sigma_s^2+\sigma_b^2+\sigma_w^2 \tag{4}$$

The estimate x from the study of the per cent condition of the plant is the simple average of the per cent conditions over all holes.

To obtain an estimate of the variance of x that properly reflects the variance between and within inspectors as well as the sampling variance, one may divide up the whole job into the inspections made by A, those by B, those by C, and those by D, compute the average result for each inspector, and then calculate the variance between the 4 averages. This procedure will give a valid estimate of σ_x^2 because each inspector had a separate valid sample of the whole job.

Results

Results by inspector

Inspector j	Average result x_j
A	83.15
B	84.12
C	86.83
D	83.36
Average $x = 84.37$	

With this approach the whole job thus gives the estimates

$$\hat{\sigma}_x^2 = \frac{1}{4(4-1)}\sum_j^4 (x_j - x)^2 = .72 \quad \text{(3 d.f.)} \tag{5}$$

for the variance, wherefore

$$\hat{\sigma}_x = .85 \tag{6}$$

for the standard error of x.

The estimate of Var x just made is valid for the total variance, but it is based only on 3 degrees of freedom. We seek a better estimate. The total variance is composed of the sampling variance plus the variance between inspectors and the variance within inspectors, which we may write out in full as follows:

$$\hat{\sigma}_x^2 = \frac{\hat{\sigma}_b^2}{4} + \frac{\hat{\sigma}_w^2}{n} + \frac{\hat{\sigma}_s^2}{n} \quad (n \text{ is the total number of holes, 640}) \tag{7}$$

We now make a further calculation. Let

$A_1 - A_2$ be the difference recorded by the same inspector (A, or B, or C, or D) on two different holes in the same zone of 2 holes. 92 usable zones.

$A_1 - B_2$ be the difference recorded by two different inspectors (A and B, A and C, A and D, B and C, etc.) on two different holes in the same zone of 2 holes. 213 usable zones.

The reader may note that the total number of usable zones for these two types of difference is $92+213=305$, which lacks 30 zones of being half of 640. The reason is that there were 30 cases of no access, points designated for inspection where, as it turned out, no hole for inspection could be dug, because of some special problem. Examples: permission not granted by civic authority to open the street, pipe under a tree, pipe under a reservoir, and a few points where the pipe was removed and replaced with new pipe before the inspector arrived (in which case the per cent condition of the pipe replaced was arbitrarily written off as 0).

Define

$$S_1^2 = Av(A_1 - A_2)^2 = 2(\hat{\sigma}_s^2 + \hat{\sigma}_w^2) \qquad \text{(92 zones, 92 d.f.)} \tag{8}$$

$$S_2^2 = Av(A - B)^2 = 2(\hat{\sigma}_s^2 + \hat{\sigma}_w^2 + \hat{\sigma}_B^2) \qquad \text{(213 zones. 213 d.f.)} \tag{9}$$

Here, average means the average of the observations. Actual calculations on this job gave

$$S_1^2 = 36{,}521/92 = 397.0 \qquad \text{(92 zones, 92 d.f.)} \tag{10}$$

$$S_2^2 = 67{,}746/213 = 318.1 \qquad \text{(213 zones, 213 d.f.)} \tag{11}$$

whence an estimate of σ_B^2 is

$$\hat{\sigma}_B^2 = \tfrac{1}{2}(S_2^2 - S_1^2) = -39 \tag{12}$$

As this estimate of σ_B^2 is negative, we conclude that the inspectors made no important contribution to the total variance.

There is no significant difference between S_1^2 and S_2^2; hence no explanation need be found for the negative variance.

Another measure of the contribution from the variance between inspectors is obtained by comparing $\hat{\sigma}_x^2$ in Eq. 7 with $S_1^2/2n$. The only difference between the expected values of these estimates is the contribution of the variance σ_B^2 between inspectors.

We have from Eq. 7

$$\hat{\sigma}_x^2 = \frac{\hat{\sigma}_b^2}{4} + \frac{\hat{\sigma}_w^2}{n} + \frac{\hat{\sigma}_s^2}{n} = .72 \quad \text{(3 d.f.)} \tag{7}$$

and also

$$\hat{\sigma}_x^2 = \frac{S_1^2}{2n} = \frac{\hat{\sigma}_w^2}{n} + \frac{\hat{\sigma}_s^2}{n} = \tfrac{1}{2} \times 397/640 = .31 \quad \text{(92 d.f.)} \tag{13}$$

These two estimates differ only by $\hat{\sigma}_b^2/n$ and they use independent estimates of $\hat{\sigma}_w^2$ and $\hat{\sigma}_s^2$. We compare them by computing

$$F = .72/.31 = 2.3 \tag{14}$$

This value of F is not significantly different from 1, hence again there is no statistical evidence of any important contribution from the variance between inspectors. We conclude that it is safe to pool the results of the inspectors, and we accordingly calculate the following estimate of Var x, which has many more degrees of freedom than Eq. 7 has.

$$\hat{\sigma}_x^2 = \sum_h^{305} \frac{1}{305 \times 640} \frac{1}{2(2-1)} (x_{h1} - x_{h2})^2 \tag{15}$$

$$= 171/640 = .267 \qquad (h \text{ is the zone. 305 d.f.}) \tag{16}$$

whence

$$\hat{\sigma}_x = .52 \tag{17}$$

The main aim here has been to show that it is possible by appropriate lay-out to hold inspectors in line and to justify pooling of their results. We may comment that small variance between inspectors is not an accident: it is not to be expected without careful supervision with statistical control.

The illustration given here is for non-destructive tests. Appropriate modification could be made for destructive tests. Other studies on record could be cited where there has been no statistical control and where the variance between investigators, not known till afterward, have led to difficulties in interpretation of results. Other statistical tools are helpful and efficient as controls, such as a run chart and sign-test, which the supervisor may construct as the work procedures, but we do not pursue them here.

It is a pleasure to record here the privilege of working with Messrs. Francis Wright and Michael G. Bartels of the East Ohio Gas Company, Cleveland, on the engagement that furnished the data for this example. Their knowledge of engineering, and their skill in following and using the statistical controls specified in the sampling procedures, made this paper possible.