Guide to quality control and performance improvement using qualitative (attribute) data —

Part 4: Attribute inspection performance control and improvement

ICS 03.120.30

British Standards

Committees responsible for this British Standard

The preparation of this British Standard was entrusted to Technical Committee SS/4, Statistical process control, upon which the following bodies were represented:

Association for Road Traffic Safety and Management (ARTSM)
BAE Systems
British Standards Society (BSS)
Clay Pipe Development Association (CPDA)
Federation of Small Businesses (FSB)
Gauge and Tool Makers Association (GTMA)
General Domestic Appliances Ltd.
Institute of Quality Assurance
National Physical Laboratory
Royal Statistical Society

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 31 October 2003

First published as BS 5701 February 1980

Amendments issued since publication

Amd. No.	Date	Comments

The following BSI references relate to the work on this British Standard:
Committee reference SS/4
Draft for comment 02/400887 DC

ISBN 0 580 42737 4

Contents

Foreword

BS 5701-4:2003 partially supersedes BS 5701:1980 and BS 2564:1955 and all four parts of BS 5701 together supersede BS 5701:1980 and BS 2564:1955, which are withdrawn.

BS 5701-1, 2 and 3 demonstrated the business benefits, versatility and usefulness of a very simple, yet powerful, pictorial control chart method for monitoring and interpreting qualitative data.

This qualitative data can range from overall business figures such as percentage profit to detailed operational data, such as percentage absenteeism, individual process parameters and product/service features. The data can either be expressed sequentially in yes/no, good/bad, present/absent, success/failure format, or as summary measures (e.g. counts of events and proportions). For the control charting of measured data, refer to BS 5702-1.

The treatment of charting of qualitative data in BS 5701-1 is essentially at appreciation level. However, it is intended to provide adequate information for a gainful first application, by a typical less statistically inclined user, in many everyday situations. BS 5701-2 and BS 5701-3 provide a more rigorous, statistical-based, approach to process control and improvement using qualitative data.

BS 5701-4 deals with measuring and improving the quality of decision-making in the classification process itself.

With the monitoring of measurement data, decisions are made with calibrated instruments against objective criteria. A different situation prevails with attribute data. Decisions are frequently made by uncalibrated people against subjective criteria.

Much is made of the need for determining the ongoing capability of measuring instruments. It is standard practice to have stringent ongoing calibration requirements for instruments. Considerable attention is given to the need for measurement systems analysis in terms of resolution, accuracy and precision to verify, for example, conformance with exacting uncertainty, repeatability and reproducibility criteria. Yet, surprisingly, relatively little similar thrust takes place with the intrinsically much less reliable subjective decisions taken by people. Consequently, data used to construct control charts to monitor attribute processes frequently lack credibility. A similar situation occurs with acceptance sampling by attributes. How reliable are the decisions made? How valid are they? The statisticians classify such errors as Type 1 or Type 2. A Type 2 error may be caused by excessive caution, e.g. a jury or court who let a guilty defendant go free, whereas the Type 1 error can result from excessive zeal, e.g. the conviction of an innocent person. Similar conditions prevail in business: faulty items are sometimes accepted and good items are sometimes incorrectly rejected. The situation in business is compounded by the presence of subjective standards against which to make a judgement. In turn, this is exacerbated by a number of factors, such as time and commercial pressures, workplace environment and inspector aptitude, faculties, skills and knowledge.

BS 5701-4 takes the view that much can, and should, be done to improve decision-making in business when subjective judgements are involved. It provides guidelines by which this can be achieved.

Summary of pages

This document comprises a front cover, an inside front cover, pages i and ii, pages 1 to 21 and a back cover.

The BSI copyright notice displayed in this document indicates when the document was last issued.

1 Scope

BS 5701-4 describes methods for enhancing the quality of attribute data used for monitoring, controlling and improving business processes.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

BS EN ISO 9000:2000, Quality management systems — Fundamentals and vocabulary.

BS ISO 3534-1, Statistics — Vocabulary and symbols — Part 1: Probability and general statistical terms.

BS ISO 3534-2, Statistics — Vocabulary and symbols — Part 2: Applied statistics.

3 Terms, definitions and symbols

For the purposes of this part of BS 5701, the terms, definitions and symbols given in BS ISO 3534-1, BS ISO 3534-2 and BS EN ISO 9000:2000, Clause **3** apply.

4 Attribute inspection performance control and improvement

4.1 Overview

Effective monitoring of discrete data processes is highly dependent on the validity of judgements made on whether a particular attribute, signal or event is present or not.

Sometimes the decision is obvious, as objective criteria are present; for example, the number of incoming calls per minute and whether or not calls are answered within six rings.

Frequently, however, subjective judgements are involved, such as the determination whether a particular flaw or imperfection is present or not, or the need to discriminate between acceptable and non-acceptable products. In the case of subjective judgements it is particularly important that:

a) the decision makers, the assessors, supervisors, inspectors, checkers, viewers, verifiers, auditors, and the like, have the appropriate aptitudes, faculties, knowledge and skills;

b) "operational definitions" exist for the particular attributes or events in question;

c) the influence of social and organizational factors and work-place design are taken into account;

d) assessor calibration and assessment performance audits are conducted where applicable;

e) attribute gauges have the required repeatability and accuracy.

4.2 Defining reference standards (establishing operational requirements)

Discrete data often arises in situations where some form of judgement is applied. To minimize variations in judgement, it is essential that clearly defined conformance criteria are established.

These criteria are often expressed in terms of "operational definitions" of characteristics. Such an operational definition:

a) consists of a criterion to be applied, an examination and a decision: yes or no;

b) requires a prescribed examination, a record of the result and a comparison of the result with the agreed criterion;

c) should be communicable with the same meaning to one person as to another, yesterday, today and tomorrow.

An operational definition should not:

i) quote exact values, e.g., round, flat, clean, pucker-free, 20 % fat (minimum, maximum, average?), 40 % wool (how distributed?);

ii) use adjectives that have no communicable meaning e.g. uniform, safe, good or reliable.

Written standards for attributes are frequently difficult to develop and to apply without interpretations. In such cases, other methods are used, such as reference samples, photography, line diagrams and stereoscopic pictures. Often these are graded, e.g. “preferred”, “minimum acceptable” and “unacceptable”, or graded on a scale e.g. 1 to 5.

4.3 Influence of social and organizational factors and work-place design

4.3.1 *Social and organizational factors*

An inspector can be subjected to a number of social pressures from those providing him with products to inspect and those receiving them. Sometimes, these pressures balance out. Organizational structures should be such that rational objective criteria prevail. These should always aim to support rather than undermine the technical performance of the inspector.

Methods of payment can also influence inspector’s performance. Payment based on total throughput tends to give rise to a tendency to skimp inspection. Payment based on the number of faults found can give rise to a possible increase in commission errors.

A further source of potential bias is the reject “norm” for a particular job, or in a particular department. It has been shown that it is possible for these norms to have a greater influence on the proportion of the parts rejected than the actual part quality level. These kinds of pressures can be alleviated by the provision of clear, objective acceptance standards, adequate instructions and training, effective inspection audit procedure and a sound organizational structure.

4.3.2 *Inspection workstation design*

4.3.2.1 *General*

Virtually all manual industrial inspection tasks rely heavily on vision, whether one is dealing with surface texture, colour, electrical or mechanical dimensions, correctness of assembly, etc. It is essential, therefore, if one is to ensure effective operating efficiency in terms of speed and accuracy, to take account of the significant factors affecting visual performance.

4.3.2.2 *Lighting*

4.3.2.2.1 *Aspects of lighting*

Aspects of lighting of particular relevance to visual inspection are:

a) light level;

b) luminance;

c) contrast;

d) colour;

e) glare.

4.3.2.2.2 *Light level*

The rate of flow of luminous energy, (namely light) from a lamp is given in terms of lumens (lm).

For example, the light output from:

a) A 100 W tungsten filament bulb is 1 200 lumens;

b) An 80 W fluorescent lamp is 4 000 lumens.

Not all the light output from a lamp can be transferred to the viewing plane. The illumination of a work surface is expressed in terms of “lumens per unit area”. Lumens per square metre is termed “lux”. The order of service illumination proposed for different classes of inspection work is shown in Table 1.

Table 1 — Minimum light levels for different types of inspection

Type of inspection	Minimum service value of illumination lux
Minute detail (e.g. watches)	2 000 to 3 000
Small detail (e.g. gauging)	1 500 to 2 000
General viewing	500 to 1 000

4.3.2.2.3 *Luminance*

Luminance is the light emitted or reflected to the eye from an object, as opposed to illumination, which is the light falling onto it. Luminance is what is seen when inspecting an object, not illumination. The brightness (or luminance) of the object is important when considering the amount of light required for a particular viewing task. This luminance depends both on the level of illumination of the object and the proportion of that incident light which is reflected from the surface of the object. Two situations arise:

a) non-specular surface – a perfectly diffusing surface, e.g. blotting paper that has the same luminance from all angles of view. Here:

Luminance = Illumination × Reflecting ratio

$B = ER$

This relationship is often used for surfaces that are not perfect diffusers:

b) specular surfaces – bright polished surface (e.g. mirror, bright metal). Here, the luminosity will depend not only on the proportion of light reflected by the surface itself but by the image of a concentrated light source itself seen reflected in the surface. This luminance will have a directional property depending on the relative positions of the light source, observer and object being viewed.

Here, luminance is expressed as intensity in a specified direction/projected area as seen from that direction.

It should be appreciated, however, that specular surfaces also have reflectivity ratios in the same way as perfect diffusers. However, they introduce an additional complication, e.g. if one is inspecting sheets of polished aluminium, the light source itself or surroundings can be reflected in the surface at certain angles of view, thus impairing viewing efficiency.

4.3.2.2.4 *Contrast*

Regardless of the level of illumination, an object can only be seen by virtue of contrast with its background. There are two forms of contrast: brightness and colour. It is possible for two quite distinct colours to reflect the same amount of incident light – brightness contrast – yet be clearly distinguishable purely by their difference in hue.

4.3.2.2.5 *Colour*

A sensation of light in the human eye is caused by electromagnetic radiation within the wavelength band 400 nm to 750 nm. The eye is not equally sensitive to all wavelengths but has peak sensitivity at about 550 nm. Where inspection involves colour discrimination or comparison, it is essential that correct sources of illumination are used. If, for example, one inspects a red material under a fluorescent lamp deficient in red, the material will tend to grey. Colour, too, can be used in the sense of increasing contrast between an object e.g. pale blue cotton on a yellow blouse. Here, where both object and background have similar reflectivity and relatively low colour contrast, viewing conditions can be improved by using an orange light source. In general, if it is required to increase the contrast between two coloured surfaces, light which is deficient in the wavelength of the darker colour should be employed.

4.3.2.2.6 *Glare*

Glare, unwanted incident light in the visual field, reduces contrast and so the threshold of visual acuity.

Glare arises from natural (e.g. sun) or artificial (e.g. lamp) sources either directly or by reflection from the task itself or from surroundings. Glare is often classified as:

a) *discomfort glare*: that which causes physical discomfort;

b) *disability glare*: that which reduces visual effectiveness.

Glare can be reduced or avoided by:

— correct positioning of the display relative to light sources;

— elimination of reflected glare from bright images in the field of view;

— reducing possible phototropic effects. The eye tends to move to the brightest part of a visual field. This is known as phototropic effect – the involuntary turning towards light. It is distracting as well as being visually disadvantageous. If the adaption level of the eye is raised by the glare the eye becomes less sensitive to small differences in brightness.

4.3.2.3 *Complexity of static visual display*

Visual acuity falls off rapidly either side of the line of sight. This means that the eye is much more efficient in seeing a detail it is looking at than one just to the side. If the angle between a detail and the line of sight is just 20° visual acuity is decreased by 90 %. In other words, such a detail has to be 10 times as large to be seen as well.

Large complex stationary objects thus require to be scanned during the visual inspection cycle. Defect detection is basically a two-stage operation:

- a possible defect is seen some distance away from the line of sight;
- the eyes then home in on the suspect detail.

Visual inspection performance in detecting in stationary displays has been shown to be dependent on:

- the complexity of the display: where possible this should be simplified as much as possible – no information should be presented which is not essential;
- the method of scan: investigations have shown that performance variation detected between inspectors is often considerably influenced by the manner in which the scan is made;
- the peripheral acuity of the viewer: it has been demonstrated on a number of occasions that there is a tendency for peripherally located defects to be missed more often than centrally located ones. Inspectors with the best peripheral acuity tend to perform better in this respect.

4.3.2.3.1 *Moving display*

With a moving display:

- it is necessary for the eye to scan at the rate of movement of the display otherwise blurring will occur due to lack of visual acuity each side of centre. Hence good central rather than peripheral acuity is required particularly as rate of movement increases;
- there is usually less time for defect search. Hence, there is a need for both ensuring that scanning is done systematically and efficiently and that the span of the field at right angles to the line of travel (e.g. of the conveyor) is appropriate to the feed speed.

4.4 Inspector: faculties, aptitude and acquisition of skills and knowledge

4.4.1 *Introduction*

The generic term “inspector” embraces many job functions. For example:

a) supervisory: organizational and administrative abilities;

b) skilled inspection: technical knowledge and judgement;

c) viewing: good vision: ability to maintain vigilance and make sound and quick decisions;

d) gauging: tactile and visual ability: systematic;

e) patrol/roving: intimate knowledge of process and operatives.

Selection procedures for inspectors should be based on a job analysis. They should include assessment of faculties, “skills”, aptitude and “temperament”. The first two can usually be readily assessed objectively using standard tests (e.g. visual acuity, reaction time, mechanical ability, manual dexterity). The latter two depend to an extent on subjective judgement. It is not an uncommon occurrence for persons to be engaged as inspectors without prior experience or training and then be looked upon as infallible.

In practice, if quantitative systematic measurements have not been taken of a particular operation, it is highly likely that the actual efficiency of inspection is much less than what it is considered to be. Speed and accuracy can often be significantly improved by initial instruction and training, performance monitoring and retraining as necessary.

4.4.2 *Inspector selection*

4.4.2.1 *Aptitude tests*

Test results showing considerable variation between inspectors with comparable training and experience in a controlled environment indicates the profound effect on detection efficiency of varying aptitude for inspection activity.

This points to the desirability of applying aptitude tests to potential inspectors. Many types of aptitude test are available. It is essential that such tests are related to the viewing process. For example, the use of pegboard tests, intended to evaluate manual dexterity are largely irrelevant to the assessment of the potential of prospective viewers of shirts.

4.4.2.2 *Knowledge tests*

It is desirable too, in selection tests to assess the applicant's knowledge in relation to the job analysis.

4.4.2.3 *Vision tests*

Vision screeners enable one to rapidly assess the visual abilities of an actual or prospective inspector. For example, a Vision Screener takes only minutes to test an inspector for:

a) visual acuity;

b) eye coordination;

c) muscle balance;

d) depth perception;

e) colour vision.

4.4.3 *Inspector training*

4.4.3.1 *Skill and knowledge requirements of task*

The level and type of training required for an inspector will depend on the nature of the inspection task. It ranges from training in the use of complex measuring instruments (skill development) on highly technical products (product knowledge acquisition) to a single simple repetitive viewing operation. An essential prerequisite to the formulation of a training programme is a job description and analysis to identify skill and knowledge requirements.

This calls for:

— *Step 1: Job specification*: a simple concise statement defining purpose, scope, duties, accountability and responsibilities.
— *Step 2: Job description*: a breakdown of the job into components.
— *Step 3: Job analysis*: a breakdown of the job components into task elements and skill requirements.

4.4.3.2 *Level of skill and knowledge of inspectors*

The need for a training programme arises from the difference between the skill and knowledge requirements of the task and the existing skill and knowledge of the inspector. Such a gap in skill and knowledge can manifest itself on:

a) first appointment of potential inspector;

b) change of job of existing inspector;

c) performance evaluation of existing inspector.

4.4.3.3 *Training programme*

For the training programme to be cost-effective (the development of skills and/or the acquisition of knowledge in the minimum time at least cost), certain features should be taken into account:

a) inspection task skill and knowledge requirements;

b) existing levels of skill and knowledge of inspector concerned;

c) inspectors often learn little from "on the job" experience alone due in large measure to the lack of timely feedback of levels of performance (errors of omission and commission);

d) training is a formal, or semiformal, process of learning;

e) in the case of inspectors, training is unlikely to take the formal traditional classroom form;

f) the most effective inspection training often takes the form of programmed experience based on the module system.

The programmed experience or module approach to training permits an overall transferable skill/knowledge acquisition approach in a particular discipline whilst at the same time:

a) catering for the individual needs of an inspector in a particular job situation;

b) permitting one to select the most effective training method for example, self-tuition, formal "off the job" or semi-formal "on the job" instruction.

4.5 Inspector calibration and pre- and post-inspection audits

4.5.1 *Introduction*

There is often a belief that assessors are virtually infallible. However, when tested scientifically, this belief in assessor accuracy is often shown to be untrue. The continuing achievement of a high standard of inspection demands constant calibration of inspectors. The fact that this is not often done, or done on a very ad hoc basis probably arises from the over-estimation of the true quality of inspection.

Calibration is particularly important where the consequence of a fault or event slipping through the net is severe. Calibration can be done in several ways as shown in **4.5.2**, **4.5.3** and **4.5.4**.

4.5.2 *Post-inspection quality audit*

Here each inspector's work is audited. Marked improvements can arise purely from awareness that inspected work is re-inspected soon afterwards. Additionally, if the type and number of faults missed are reported back, individual propensities to miss certain faults are quickly corrected. It is desirable that each inspector should be told, not only how many defects he/she has missed (errors of omission) but also how many good items he/she has rejected (errors of commission). If omission errors only are fed back, this could give rise to an increase in commission errors: protection of the customer at the expense of the producer. Knowledge of results is an effective means of enhancing inspection performance both from a motivational and training/learning point of view.

What is the relative effectiveness of random audits on in-process inspection (post operation, roving or patrol), immediately after final inspection or on boxed stock in the warehouse? In-process inspectors should be sufficiently close to the process and operators to develop an awareness of the expected quality performance of the various plant and personnel. If he/she is engaged for a set amount of time on this activity, then he/she will usually bias his/her density of inspection accordingly. Often, however, post-final inspection quality audit is a more remote ad hoc affair. It frequently involves random checks on an unspecified and often variable proportion of production. This is the worst of all worlds. The conclusion is that, with audits of in-process inspection, it is desirable and, with audits of final inspection, it is essential to operate to systematic procedures. Such procedures are discussed.

4.5.3 *Introduction of selected pseudo defects into normal production*

Here, defects are "planted" in batches to be submitted to inspection. Such "plants" are often marked invisibly e.g., by fluorescent dye. Inspection efficiency is then assessed using the "controlled sample technique". Under certain conditions, this method is neither effective nor feasible. The inspector can get to know these "plants", particularly where production is date or serial marked. Where production is varied and a number of styles or designs are concerned, numerous "plants" need to be stored, indexed and maintained in an "as new" condition.

4.5.4 *Pre-inspection quality audit*

It has been proposed by some investigators that audits be conducted prior to 100 % final inspection. This recommendation is based on the premise that if inspectors are acquainted with the kind of faults to be expected in a batch, then the proportion of omission errors decreases significantly.

Whatever the method of calibration, an inspector should uniquely identify each item inspected. One method is to issue inspectors with numbered stamps, which are applied to all "passed" work. This tends to make the inspector feel more responsible and ensures that information on work subsequently found defective can be fed back to the appropriate inspector.

4.6 Assessment of attribute gauge repeatability and accuracy

4.6.1 *Overview*

When measured data characteristics are monitored using attribute gauges, gauge performance characteristics should be checked for acceptability. Features that need quantifying are both random and systematic errors.

4.6.2 *Random and systematic errors*

A measuring instrument can have one of two reasons for giving a reading that is inaccurate.

a) The instrument is out of calibration: a series of readings made on a single unit gives an average that differs from the true value by an amount greater than that specified. This is a measure of the systematic error, termed *bias*.

Bias is the difference between a test result (observed mean of several measurements) and the accepted reference value.

b) Irrespective of the state of calibration, the instrument will not give identical values when making a series of readings on a single unit. This is a measure of the random error, termed *precision*.

Precision is the closeness of agreement between independent test results obtained under stipulated conditions.

The difference between bias and precision is illustrated in Figure 1.

Repeatability is defined as precision under repeatability conditions. Repeatability conditions are observation conditions where independent test results are obtained with the same method on identical test items in the same test facility by the same operator using the same equipment within short intervals of time" (BS ISO 3534-2).

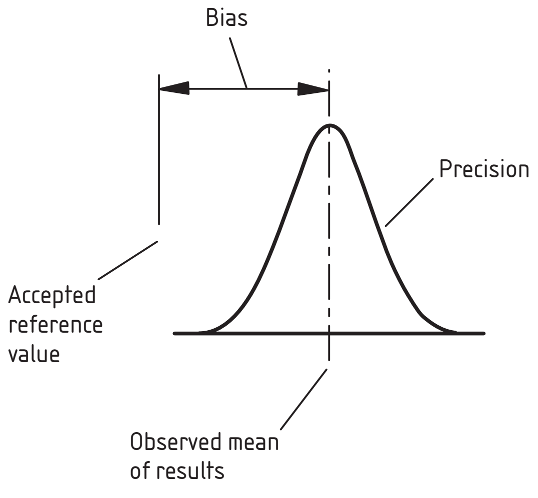


Figure 1 — Bias and precision

4.6.3 *Method*

A typical method of conducting an attribute gauge study against an upper or lower limit is shown. Stages are:

— *Stage 1*: Select eight parts with measurement values of the appropriate feature equally spaced (as far as practicable) throughout the range for which the gauge performance characteristics are to be assessed.

— *Stage 2*: Run the eight parts through the gauge $m = 20$ times and record the number of gauge "accepts" (A) against either the upper or lower limit as appropriate.

— *Stage 3*: Assess the results of Stage 2 against the following criteria:

a) Does the smallest part have a value of $A = 0$?

b) Does the largest part have a value of $A = 20$?

c) Do six of the parts have values of A between 1 and 19?

If all criteria are satisfied go to Stage 5, otherwise go to Stage 4.

— *Stage 4*:

a) If, for the smallest part, $A \neq 0$, then select and run progressively smaller parts through the gauge until $A = 0$.

b) If, for the largest part, $A \neq 20$, then select and run progressively larger parts through the gauge until $A = 20$.

c) If six of the parts do not have values between 1 and 19, select additional parts to run through the gauge. These parts should have values mid-way between those parts already selected: starting from the appropriate end $A = 0$ or $A = 20$ and working in towards the middle of the part range.

Continue this procedure until all criteria are met, then proceed to Stage 5.

— *Stage 5*: Estimate the probabilities of acceptance of each part.

The number of accepts (A) for each part form a discrete binomial distribution with parameters m and p'_A, where p'_A is the probability of accepting a part of a particular size, say X. The gauge error is taken to have a continuous normal distribution and the estimated probability of acceptance for X, $p'_A = A/m$, is adjusted by a step or continuity factor of 0.5 thus:

$$p'_A = \frac{A + 0.5}{m}, \text{ if } A < 0.5m \text{ and } A \neq 0;$$

$$p'_A = \frac{A - 0.5}{m}, \text{ if } A > 0.5m \text{ and } A \neq 20;$$

$$p'_A = 0.5, \text{ if } A = 0.5m;$$

$$p'_A = 1.0, \text{ if } A = 20 \text{ except for the smallest negative numerical value with } A = 20 \text{ when } p'A = 0.975;$$

$$p'_A = 0, \text{ if } A = 0, \text{ except for the largest negative numerical value with } A = 0 \text{ when } p'A = 0.025.$$

Tabulate, X, A and p'_A.

— *Stage 6*: Assess gauge performance curve in terms of bias and repeatability.

This is best achieved graphically using normal probability paper. This method has the advantages of avoiding extensive calculations and, at the same time, does both a visual check for normality and for any peculiarities in the readings.

Estimate:

Bias = appropriate limit – X (at $p'_A = 0.50$)

Repeatability = X (at $p'_A = 0.995^{a}$) – X (at $p'_A = 0.005^{a}$)

[a] or other probability values stipulated in a particular requirement or contract.

EXAMPLE

Task

An attribute gauge used to monitor a dimensional characteristic that has a specified ±0.2 mm tolerance is required to quantify the bias and repeatability of the gauge.

Solution

This is a double limit gauge, however only the gauge performance relating to the lower limit will be estimated. The assumption here is that the results relating to the upper limit will be a mirror image of that analyzed in respect of the lower limit with purely a correction for the bias.

Eight parts with measurements, X_1, at intervals of 0.04 from –0.32 to –0.04 were each run through the gauge 20 times. The number of accepts (A) for each part are tabulated as A_1 in Table 2.

Table 2 — Example

X_1	A_1	X_2	A_2	X_3	A_3
–0.32	0				
–0.28	3	–0.30	1		
–0.24	9	–0.26	4		
–0.20	20	–0.22	14		
–0.16	20			–0.21	18
–0.12	20				
–0.08	20				
–0.04	20				

As only two of the parts have values of "A" between 1 and 19 further parts, X_2, were introduced with values midway between the four lowest values. Results are shown as A_2.

Five parts now have acceptance value with A between 1 and 19. A further part, X_3, was now run through the gauge with the result shown as A_3.

The probabilities of acceptance are now calculated using the continuity factor of +0.5 with the result shown in Table 3.

Table 3 — Gauge estimated probabilities of acceptance

X	A	p_A
–0.32	0	0.025
–0.30	1	0.075
–0.28	3	0.175
–0.26	4	0.225
–0.24	9	0.475
–0.22	14	0.675
–0.21	18	0.875
–0.20	20	0.975
–0.16	20	1.000

The probability (p_A) is now plotted against the item measurement (X) to obtain the performance profile of the gauge. This is done on normal probability paper as this is likely to produce a straight line graph. This permits extrapolation to the 0.995 and 0.005 probability values. The result is shown in Figure 2.

Gauge bias is given by the difference between the master measurement and the observed average.

Here, gauge bias = –0.20 – (–0.243) = 0.043 mm.

A frequent measure of gauge variation or repeatability is given by 5.16 standard deviations (where 5.16 standard deviations embrace 99 % of the measurements for a normal distribution).

Here, gauge repeatability = 5.16 × 0.051 5 = 0.266 mm.

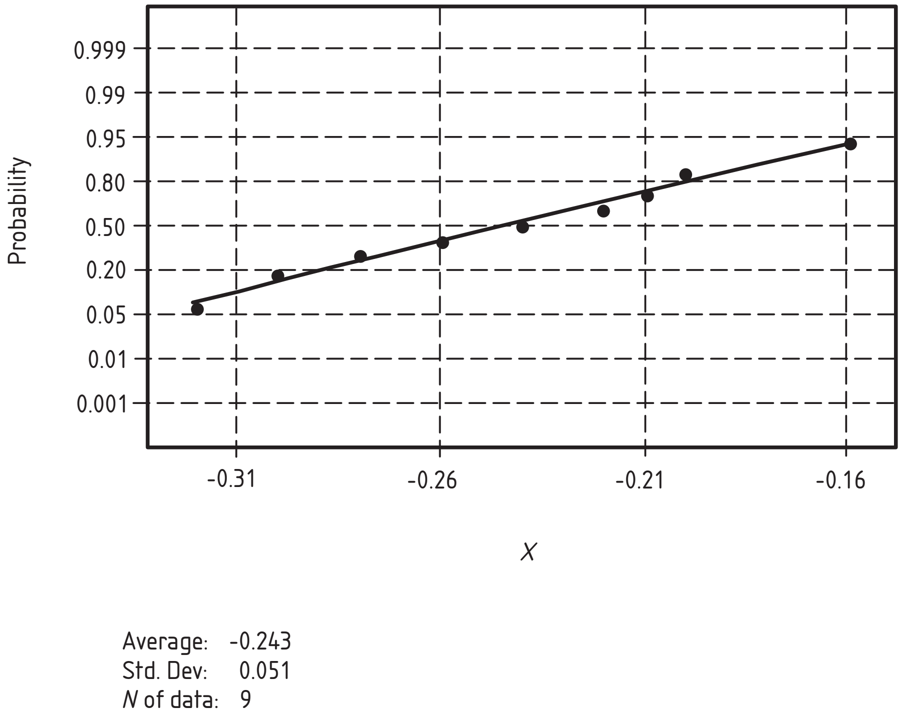


Figure 2 — Gauge performance profile

5 Inspector error in the detection of discrete events

5.1 Introduction

Classification situations abound in many fields where people are required to detect events that are near the threshold of perception. In the medical sector, for example, judgements are made on whether, or not, an abnormality is present arising from the scrutiny of X-rays and smears. In the administrative area, for instance, election votes can be misclassified. Other examples are:

— Will the security guard detect abnormal movement in one of a bank of television screens?

— Is a defendant guilty or innocent?

— In transport, will the sound of the horn of an approaching train be detected by a driver of a vehicle approaching an open level-crossing?

— Will a driver detect the vehicle or obstacle in front during conditions of poor visibility?

In industry and commerce, it is commonplace to inspect product to determine whether or not non-conformities are present or whether an item conforms, or non-conforms, to specification. Multiple attributes are often scrutinized on each individual, or item, and sometimes the result can be graded into one of a number of categories, say, according to the signal magnitude or degree of abnormality. However, for convenience and brevity, the situation discussed is restricted to whether or not a signal, event, attribute or abnormality is present.

In such a situation, only two possible errors in observation and classification exist. An error of commission is termed a "false alarm". This occurs when the observer reports a signal when it does not exist. An error of omission is called a "miss". This happens when a signal is present but it is not detected. A "correct" decision is made if the response is no when no signal exists. Another right decision is made when there is a yes response to a signal that is present. This is called a "hit". Perfect performance is achieved only when there are no misses or false alarms.

A decision matrix representing the four possible classifications is shown in Table 4.

NOTE 1 Inspector is used as an all embracing term for a verifier, auditor, checker, scrutineer, observer, decision maker and the like.

NOTE 2 Signal is used as an all embracing term for event, attribute, abnormality, non-conformity and the like.

Table 4 — Four way classification decision matrix

Judgement		**Response**	
Event		**No**	**Yes**
Signal	**No**	Correct	False alarm
	Yes	Miss	Hit

5.2 Role of signal detection theory in the assessment and improvement of inspector performance

5.2.1 *Overview*

In the presence of uncertainty, often referred to as noise, a response or judgement on whether, or not, a particular event, attribute or signal is present will depend on two components:

— the *sensitivity* of the decision maker to the event, attribute, signal or abnormality: how well the signal can be discriminated from the background noise;

— the *decision criterion*: the propensity, or bias, towards saying something is there, or not, when they are in doubt.

Signal detection theory provides simple numerical benchmarks for both *sensitivity* and decision criteria. These are:

— d' (pronounced "dee prime").The index d' provides a measure of the ability of an observer to discriminate between a signal and its background noise. The larger the value of d', the better the ability of the observer to discriminate between an event and a non-event. A value of zero indicates that it is impossible to distinguish between when a signal, or attribute, is present and when it is not;

— c. The index c relates to the bias of the decision maker or decision criteria adopted. It marks the boundary between "present" (yes) and "absent" (no) responses.

5.2.2 *Signal discriminatory index, d'*

Signal detection theory is based on the premise that there is a level of activation in the brain of the observer even when the signal to be detected is not present. This is termed noise. Noise can exist in both external form (environmental conditions) and internal form (in the mind of the observer).

An example of noise arises in the case of the viewing of an X-ray film. Here, the film itself can be somewhat fuzzy or have a smudge or a bad spot on it. Perhaps something observed on it is actually fine but looks somewhat like an abnormality. This noise, or stimulation of the visual system of the observer, in the absence of a real signal can be represented, in signal detection theory, as a normal distribution with a standard deviation of one.

When a signal, e.g. an abnormality, is actually present, the observer's general level of activation is increased by an amount equal to the sensitivity of the observer to that signal. This, again, is shown as a normal distribution but now with a displaced mean d' of from the pure noise distribution but still with a standard deviation of one as the background noise level has not changed. These distributions are shown pictorially in Figure 3.

Summarizing, d' is a measure of how well a person is able to discriminate a signal in the presence of noise. d' is the distance between the means of the "no signal" and "signal" distributions divided by the standard deviation of the distributions. The larger the difference between the means, and/or the smaller the standard deviation, the better the ability to discriminate.

NOTE 1 The distributions have a standard deviation = 1.

NOTE 2 d' = (distance between the means of the two distributions divided by their standard deviation).

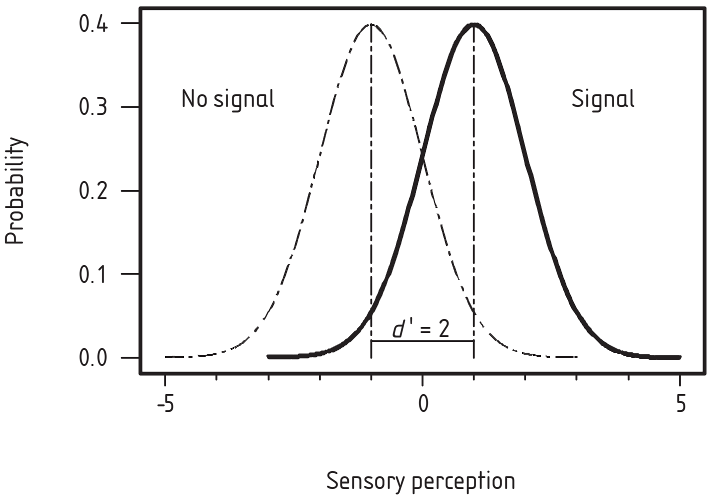


Figure 3 — Role of the value of *d′* in determining the ease of discrimination of a signal

Figure 3 indicates that:

- — when *d′* is zero, the two superimposed curves would be identical and it would be impossible for an observer to distinguish a signal;
- — when *d′* = 2, there is considerable difficulty in distinguishing the presence of a signal because of the high degree of overlap of the two distributions;
- — the larger the value of *d′*, the better able the observer is to detect a signal;
- — when *d′* = 4, it is fairly easy to distinguish the presence of a signal.

Improvements in discrimination are achieved by ensuring observers having appropriate aptitude, faculties, skill and knowledge; enhanced workplace design and better defined operational requirements.

5.2.3 *Decision criteria, c*

The decision criterion, *c*, is the second important component of signal detection theory. Many decisions on whether, or not, a signal exists are based on the judgement of the observer. See Annex A for methods of calculating *c*.

A particular decision criterion is indicated by superimposing a vertical line on the internal response distributions. The various associated hit, miss, correct and false alarm decisions, with a given discriminatory index and decision criteria are shown in Figure 4a) and Figure 4b).

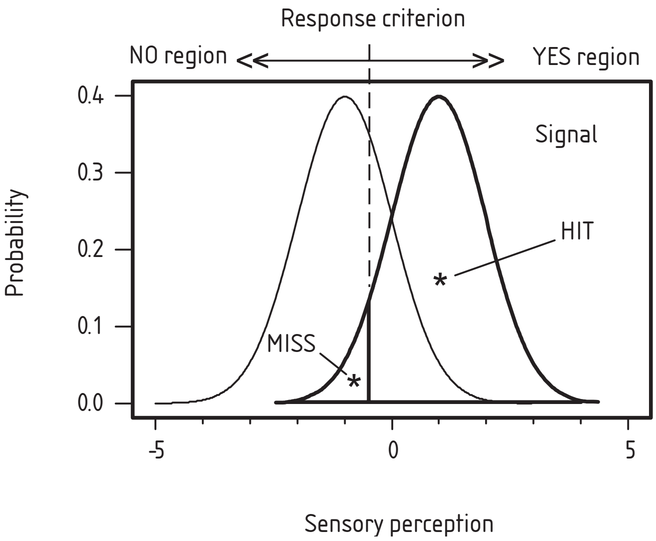
NOTE 1 In a fault detection process, "no signal" refers to a good item and "signal" relates to non-conforming item distribution. Also the 'no region' refers to "accept" and the "yes region" to "reject".

NOTE 2 The area bounded by the "signal" distribution is equal to 1. The areas bounded by "hit"' and by "miss" are proportionate to the relative probabilities. If, for example, there is a 0.9 hit probability, then the miss probability = 0.1.

NOTE 3 The area bounded by the "no" signal distribution is equal to 1. The areas bounded by "correct" and by "false alarm" are proportionate to their relative probabilities. If, for example, there is a 0.8 "correct" probability then the "false alarm" probability is 0.2.

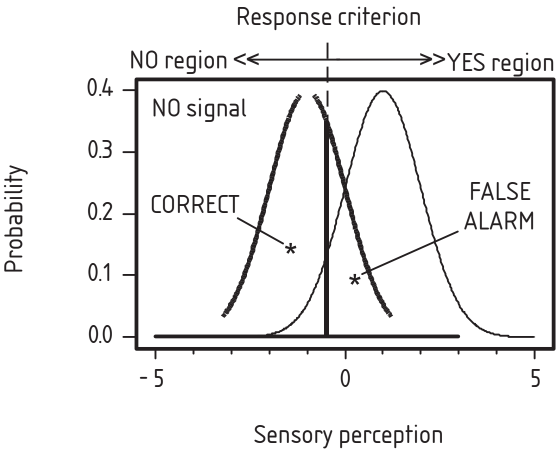
NOTE 4 Moving the response criteria to the right decreases both the hit rate and the rate of false alarms. Moving the response criteria to the left increases the hit rate at the expense of raising the false alarm rate.

Summarizing, for a given discriminatory value of *d′*, the larger the value of *c* the more likely the observer is to say "no". The smaller the value of *c*, the more likely the observer is to say "yes".



NOTE d' = 2 and c = –0.5 in the figure.

Figure 4a) — Relationship between criterion response, *c*, and a hit and a miss for a given value of discriminatory index, *d′*



NOTE d' = 2 and c = –0.5

Figure 4b) — Relationship between criterion response, *c*, and a correct decision and false alarm for a given value of discriminatory index, *d′*

5.2.4 *Case study*

Different observers can feel differently about the same detected abnormality. For instance, two doctors can have similar training and experience and have the same information about an abnormality placed before them, have identical discriminatory powers (equal d') and yet reach a different decision.

One doctor can feel that early diagnosis can mean the difference between life and death and a false alarm would result only in a routine biopsy. He would tend to be biased towards "abnormality present" decisions.

The other doctor can feel that unnecessary surgeries, even minor ones, are unwelcome because of existing waiting lists, cost, stress and other factors and, in any case, any abnormality that might be present will probably be picked up at the next check-up. He would tend to be biased towards "abnormality absent" decisions. Figure 5 shows the results of checking for a specific medical condition by two doctors, A and B, with similar training and experience having identical information presented to them. The difference in results indicate that they had a different decision criterion in their mind.

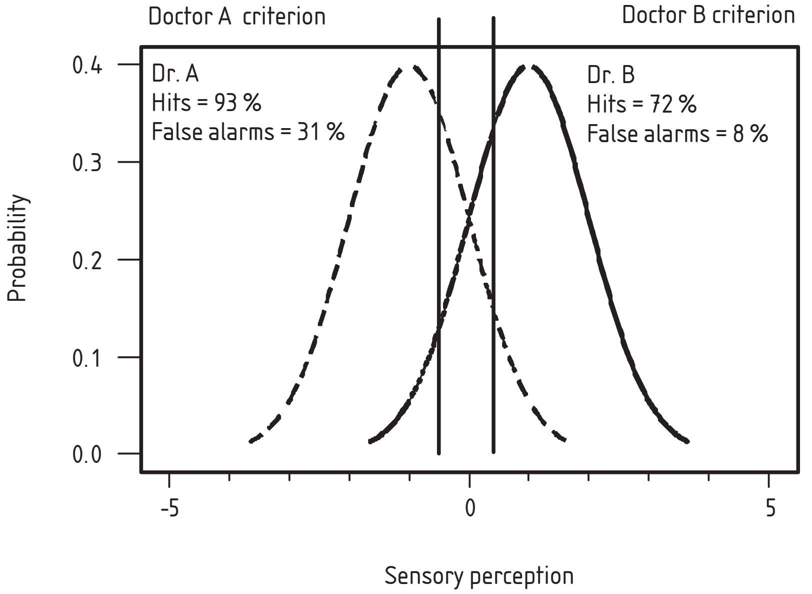


Figure 5 — Comparison of the results from two observers having identical discriminatory powers but different decision criterion

5.2.5 *Operating characteristics*

An operating characteristic curve consists of a plot of hits (on the vertical axis) against false alarms (on the horizontal axis) for a given value of d'.

For a given degree of signal discrimination, namely a constant d' value, the effect of a different decision criterion is to change the proportion of hits and false alarms. This is indicated in the operating characteristic for a particular observer on a specific discrimination task in Figure 6.

A comparison of the discriminatory performance of a number of observers on a given task can be made from their operating characteristics. This calls for prior calculation of their d' values.

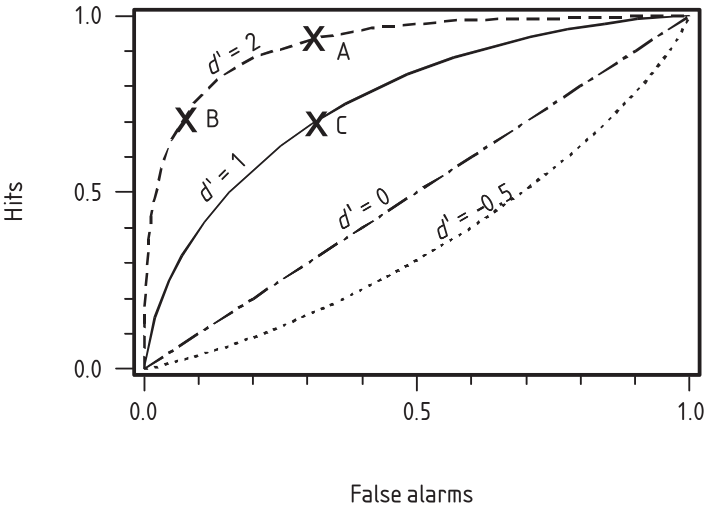


Figure 6 — Operating characteristics showing the effect on performance of changes in the signal discriminatory index, *d′*, through the range of decision criterion, *c*

Knowing, too, the *c* values provides information on what part of the operating characteristic they are working; namely, where they have positioned their decision criteria. For example in Figure 6:

— observer A has a better discriminatory performance than observer C;

— observer A has an identical discriminatory performance (*d′* value) to observer B. However because of their different decision criteria (*c* values) their rates of hits and false alarms differ. Observer A has a ratio of hits to false alarms of 93 % to 31 % whilst B has a ratio of 72 % to 8 %;

— observer C has an identical false alarm rate to that of observer A at 31%. However, C’s hit rate is much less than that of A at 72 % rather than 93 %. This is due to C’s inferior discriminatory performance, having a *d′* = 1 rather than the *d′* = 2 of A;

A *d′* = 0 means that the choice of decision is purely random: a negative *d′* indicates that the decision is perverse. This has been known to happen in a particular factory where a group of final product inspectors were in a dispute over pay. Product audits found that stock in the warehouse had higher fault rates than that prior to final inspection. A similar incident occurred in another area where a particular inspector acted perversely following criticism of her work.

5.3 Inspector performance audit

5.3.1 *Overview*

The value of inspector performance audit is discussed in terms of faulty products. The findings are equally relevant to other areas. Here a sample of product is to be inspected for non-conforming items. The two possible errors in classification are:

— a non-conforming (bad) item can be classified as conforming (good);

— a conforming (good) item can be classified as non-conforming (bad).

If the inspector:

— accepts a good item: this is a “correct decision”;

— rejects a good item: this is an incorrect decision and is considered to be a “false alarm”;

— accepts a bad item: this is an incorrect decision and is considered to be a “miss”;

— rejects a bad item: this is a correct decision and is considered to be a “hit”.

These various decisions are illustrated in Table 5.

Table 5 — Inspector decision matrix

		Inspector Decision		**Total**
		Accept	**Reject**	
Input quality	**Good**	Correct	False alarm	OK
	Bad	Miss	Hit	NOT OK
Total		Accept	Reject	TOTAL

The performance of inspectors subjected to audit can now be assessed in various ways, for example, in terms of:

a) hit rate

$$\%\ \text{hit rate} = \frac{\text{no. of bad items rejected}}{\text{total no. of bad items}} \times 100$$

b) false alarm rate

$$\%\ \text{false alarm rate} = \frac{\text{no. of good items rejected}}{\text{total no. of good items}} \times 100$$

c) miss rate

$$\%\ \text{miss rate} = \frac{\text{no. of bad items accepted}}{\text{total no. of bad items}} \times 100$$

d) probability of correct decision

$$\text{probability of correct decision} = \frac{\text{no. of accepts} + \text{no. of hits}}{\text{total inspected}}$$

e) efficiency rate

efficiency rate = proportion of bad rejection × proportion of good accepted

$$= \frac{\text{no. of hits}}{\text{no. of bad items}} \times \frac{\text{no. of corrects}}{\text{no. of good items}}$$

the rate becomes:

0	when no bad items are rejected or no good items are accepted;
1	when there is perfect discrimination;
0.25	when the decisions are made by pure chance.

f) post-inspection reject rate (after detected bad items have been removed)

$$\%\ \text{post inspection fault rate} = \frac{\text{no. of misses}}{\text{no. accepted}} \times 100\ \%$$

g) signal discriminatory index, d', where:

$$d' = z(\text{probability of a false alarm}) - z(1 - \text{probability of a hit})$$

where z is defined in Annex A.

h) decision criterion, c, where:

$$c = \frac{y(1 - \text{probability of a hit})}{y(\text{probability of a false alarm})}$$

where z is defined in Annex A.

5.3.2 ***Case Study***

An inspector performance audit was conducted on three inspectors, Tom, George and Ahmed. Each inspector was requested to inspect a batch of 200 items for a particular attribute and classify the items, against predetermined standards, as acceptable or non-acceptable.

The results of the inspection are as shown in Table 6.

Table 6 — Results of inspector performance audit

Items	Decision					
	Tom		George		Ahmed	
	Accept	Reject	Accept	Reject	Accept	Reject
Good	178	2	172	8	178	2
Bad	4	16	2	18	2	18

The performance of inspectors subjected to audit can now be assessed in various ways, for example, in terms of:

a) $$\%\ \text{hit rate} = \frac{\text{no. of bad items rejected} \times 100}{\text{total no. of bad items}}$$

b) $$\%\ \text{false alarm rate} = \frac{\text{no. of good items rejected} \times 100}{\text{total no. of good items}}$$

c) $$\%\ \text{miss rate} = \frac{\text{no. of bad items rejected} \times 100}{\text{total no. of bad items}}$$

Results of the analysis are shown in Table 7.

Table 7 — Characteristics of inspectors

Inspector	Tom	George	Ahmed
Actual fault rate %	10	10	10
Reject rate %	9	13	10
Hit rate %	80	90	90
False alarm rate %	1.1	4.4	1.1
Miss rate %	20	10	10

Table 6 shows that Ahmed is making the best decisions. Whilst George has the same hit rate as Ahmed his false alarms rate is greater. So he is rejecting a larger number of good items. Tom has a lower hit rate than the others. This leaves the customer more vulnerable. However his false alarm rate is comparable with Ahmed.

Annex A (normative)
Application of signal detection theory to the assessment of inspection performance

A.1 Table of ordinates, y, and p to z conversions for a standardized normal distribution

Table A.1 — Table of ordinates, y, and p to z conversions for a standardized normal distribution

p	z	y	$1 – p$	p	z	y	$1 – p$
0.001	3.090	0.003	0.999	0.240	0.706	0.311	0.760
0.002	2.878	0.006	0.998	0.250	0.675	0.318	0.750
0.003	2.748	0.009	0.997	0.260	0.643	0.324	0.740
0.004	2.652	0.012	0.996	0.270	0.613	0.331	0.730
0.005	2.576	0.015	0.995	0.280	0.583	0.337	0.720
0.010	2.326	0.027	0.990	0.290	0.553	0.342	0.710
0.020	2.054	0.049	0.980	0.300	0.525	0.348	0.700
0.030	1.881	0.068	0.970	0.310	0.496	0.353	0.690
0.040	1.751	0.086	0.960	0.320	0.468	0.358	0.680
0.050	1.645	0.103	0.950	0.330	0.440	0.362	0.670
0.060	1.555	0.119	0.940	0.340	0.413	0.367	0.660
0.070	1.476	0.134	0.930	0.350	0.385	0.370	0.650
0.080	1.405	0.149	0.920	0.360	0.359	0.374	0.640
0.090	1.341	0.162	0.910	0.370	0.332	0.378	0.630
0.100	1.282	0.176	0.900	0.380	0.306	0.381	0.620
0.110	1.227	0.188	0.890	0.390	0.279	0.384	0.610
0.120	1.175	0.200	0.880	0.400	0.253	0.386	0.600
0.130	1.126	0.212	0.870	0.410	0.228	0.389	0.590
0.140	1.080	0.223	0.860	0.420	0.202	0.391	0.580
0.150	1.037	0.233	0.850	0.430	0.176	0.393	0.570
0.160	0.995	0.243	0.840	0.440	0.151	0.395	0.560
0.170	0.954	0.253	0.830	0.450	0.126	0.396	0.550
0.180	0.915	0.262	0.820	0.460	0.101	0.397	0.540
0.190	0.878	0.271	0.810	0.470	0.075	0.398	0.530
0.200	0.842	0.280	0.800	0.480	0.050	0.399	0.520
0.210	0.807	0.288	0.790	0.490	0.025	0.399	0.510
0.220	0.772	0.296	0.780	0.500	0.000	0.399	0.500
0.230	0.739	0.304	0.770				

where

p is the distribution tail area probability;

z is the number of standard deviations between the mean of a normal distribution and a value of interest (here the response criterion, c);

y is the ordinate of distribution;

$1 – p = 1 –$ (distribution tail area probability).

A.2 Example of application of tables in the calculation of inspection performance

A.2.1 *Calculation of discriminatory index*

Figure 4a) shows that the probability of a hit (p_{HIT}) corresponds with the area under the "signal" distribution to the right of the response criterion, *c*. Similarly, the probability of a false alarm (p_{FA}) corresponds with the area under the "no signal" distribution to the right of *c*.

Provided:

$$p_{\mathrm{HIT}} = \frac{\text{number of hits}}{\text{number of trials with signal present}}$$

and

$$p_{\mathrm{FA}} = \frac{\text{number of false alarms}}{\text{number of trials with no signal present}}$$

When these are found from audits of inspectors then the corresponding areas are known. This leads to the calculation of the discriminatory index, *d′*. *d′* is simply the distance between the means of the "no signal" and the "signal" distribution, measured in units of standard deviations.

If the response criterion, *c* is between the means of "no signal" and "signal" distribution, then:

$d' = z_{\mathrm{FA}} + z_{(1-\mathrm{HIT})}$; or

$d' = z_{\mathrm{FA}} + z_{\mathrm{MISS};}$

where

z_{FA} = z score corresponding with the probability of a false alarm;

$z_{(1-\mathrm{HIT})}$ = z score corresponding with the probability of (1 – HIT);

$z_{(\mathrm{MISS})}$ = z score corresponding with the probability of a miss.

This is shown pictorially in Figure A.1.

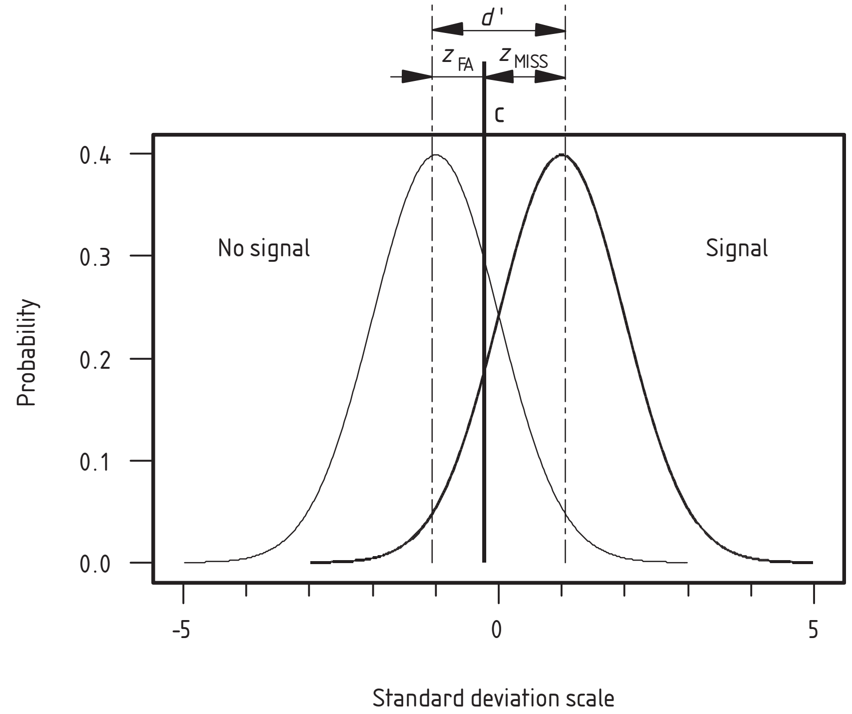


Figure A.1 — Illustration of signal and no signal relationships

EXAMPLE

In a trial of inspector performance, it was determined that the probability of a false alarm, $p_{\mathrm{FA}} = 0.2$ and that of a hit, $p_{\mathrm{HIT}} = 0.9$.

For z_{FA}, enter table at $p = 0.2$, thence
$z_{\mathrm{FA}} = 0.842$

For $z_{(1-\mathrm{HIT})}$ enter table at $1 - p = 0.9$, thence
$z_{(1-\mathrm{HIT})} = 1.282$

Then $d' = z_{\mathrm{FA}} + z_{(1-\mathrm{HIT})} = 0.842 + 1.282 = 2.124$.

A.2.2 *Calculation of response criterion*

A.2.2.1 *Criterion 1, c index*

The value of c will depend on the origin of the horizontal axis of the two distributions in Figure A.1. In this example, the origin is midway between the means of the two distributions. Hence, here:

$$c = \frac{d'}{2} - z_{\mathrm{MISS}}$$

EXAMPLE

Take $p_{\mathrm{FA}} = 0.2$;

$p_{\mathrm{MISS}} = 1 - p_{\mathrm{HIT}} = 1 - 0.9 = 0.1$; thence from previous example,

$$c = \frac{2.124}{2} - 1.282 = -0.22$$

A.2.2.2 *Criterion 2, B index*

Sometimes, B is used as an alternative response criterion index. This criterion is defined as the ratio of the ordinates (y value on vertical axis) of the standardized normal distributions at their intercept with the vertical response criterion line. This is written in equation form as:

$$B = \frac{y_{\mathrm{HIT}}}{y_{\mathrm{FA}}} = \frac{y_{\mathrm{MISS}}}{y_{\mathrm{FA}}}$$

EXAMPLE

Take $p_{\mathrm{FA}} = 0.2$;

$p_{\mathrm{MISS}} = 1 - p_{\mathrm{HIT}} = 1 - 0.9 = 0.1$

For $y_{1-\mathrm{HIT}}$, enter Table A.1 at $p = 0.1$; thence

$y_{\mathrm{MSS}} = 0.176$

For p_{FA} enter table at $p = 0.2$; thence

$Y_{\mathrm{FA}} = 0.28$; and

$$B = \frac{0.176}{0.28} = 0.63$$

A.3 Availability of interactive Java applets on the internet

Interactive Java applets are freely available on the Internet for the calculation, and sometimes visual presentation, of inspection performance using signal detection theory. An example is:

http://acad.cgu.edu/wise/sdt/sdt.html

This pictorial applet is initiated by entering "hit" and "false alarm" rates. Distributions of "signal" and "no signal" are plotted as well as the operating characteristic of hits versus false alarms.

Tags are provided to enable the user to adjust both signal sensitivity, d' and decision criterion c.

Bibliography

Standards publications

BS 5701-1, *Guide to quality control and performance improvement using qualitative (attribute) data — Part 1: Uses and value of attribute charts in business, industry, commerce and public service.*

BS 5701-2, *Guide to quality control and performance improvement using qualitative (attribute) data — Part 2: Fundamentals of standard attribute charting for monitoring, control and improvement.*

BS 5701-3, *Guide to quality control and performance improvement using qualitative (attribute) data — Part 3: Technical aspects of attribute charting: special situation handling.*

BS ISO 3534-2, *Statistics — Vocabulary and symbols — Part 2: Statistical quality control.*

BSI — British Standards Institution

BSI is the independent national body responsible for preparing British Standards. It presents the UK view on standards in Europe and at the international level. It is incorporated by Royal Charter.

Revisions

British Standards are updated by amendment or revision. Users of British Standards should make sure that they possess the latest amendments or editions.

It is the constant aim of BSI to improve the quality of our products and services. We would be grateful if anyone finding an inaccuracy or ambiguity while using this British Standard would inform the Secretary of the technical committee responsible, the identity of which can be found on the inside front cover.
Tel: +44 (0)20 8996 9000. Fax: +44 (0)20 8996 7400.

BSI offers members an individual updating service called PLUS which ensures that subscribers automatically receive the latest editions of standards.

Buying standards

Orders for all BSI, international and foreign standards publications should be addressed to Customer Services. Tel: +44 (0)20 8996 9001.
Fax: +44 (0)20 8996 7001. Email: orders@bsi-global.com. Standards are also available from the BSI website at http://www.bsi-global.com.

In response to orders for international standards, it is BSI policy to supply the BSI implementation of those that have been published as British Standards, unless otherwise requested.

Information on standards

BSI provides a wide range of information on national, European and international standards through its Library and its Technical Help to Exporters Service. Various BSI electronic information services are also available which give details on all its products and services. Contact the Information Centre.
Tel: +44 (0)20 8996 7111. Fax: +44 (0)20 8996 7048. Email: info@bsi-global.com.

Subscribing members of BSI are kept up to date with standards developments and receive substantial discounts on the purchase price of standards. For details of these and other benefits contact Membership Administration.
Tel: +44 (0)20 8996 7002. Fax: +44 (0)20 8996 7001.
Email: membership@bsi-global.com.

Information regarding online access to British Standards via British Standards Online can be found at http://www.bsi-global.com/bsonline.

Further information about BSI is available on the BSI website at http://www.bsi-global.com.

Copyright

This does not preclude the free use, in the course of implementing the standard, of necessary details such as symbols, and size, type or grade designations. If these details are to be used for any other purpose than implementation then the prior written permission of BSI must be obtained.

Details and advice can be obtained from the Copyright & Licensing Manager.
Tel: +44 (0)20 8996 7070. Fax: +44 (0)20 8996 7553.
Email: copyright@bsi-global.com.

BSI
389 Chiswick High Road
London
W4 4AL