INTERNATIONAL STANDARD

ISO 12745

Second edition
2008-10-01

Copper, lead and zinc ores and concentrates — Precision and bias of mass measurement techniques

Minerais et concentrés de cuivre, de plomb et de zinc — Justesse et erreurs systématiques des techniques de pesée

Reference number
ISO 12745:2008(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO 12745 was prepared by Technical Committee ISO/TC 183, *Copper, lead, zinc and nickel ores and concentrates*.

This second edition cancels and replaces the first edition (ISO 12475:1996), which has been technically revised.

Copper, lead and zinc ores and concentrates — Precision and bias of mass measurement techniques

1 Scope

This International Standard provides guidelines to test for bias over a wide range of mass measurement techniques, to estimate the precision for each technique and to calculate the precision for wet mass when estimated by applying one of those techniques.

The guidelines are based on the application of statistical tests to verify that a mass measurement technique is unbiased, to estimate the variance as the most basic measure for its precision and to check the linearity of a static scale over its working range. Calibration methods and performance tests for compliance with applicable regulations generate test results that can be used to quantify precision and bias for each of these mass measurement techniques and to verify linearity for static weighing devices.

The guidelines apply to mass measurement techniques used to estimate the wet mass for cargoes or shipments of mineral concentrate as the basis for freight and insurance charges and for preliminary payments or for final settlements between trading partners.

The application of static scales requires that at least one certified weight with a mass of no less than one (1) tonne be either available on location or brought in for calibration purposes, and that this certified weight be applicable to the scale in accordance with the manufacturer's recommendations. A set of certified weights covering the entire working range of a weighing device simplifies the process of verifying its state of calibration, estimating its precision as a function of applied load and testing its linearity over the working range.

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.

ISO 3534-1:2006, *Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability*

ISO 3534-2:2006, *Statistics — Vocabulary and symbols — Part 2: Applied statistics*

ISO 5725-1:1994, *Accuracy (trueness and precision) of measurement methods and results — Part 1: General principle and definitions*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

NOTE 1 In authoritative textbooks on applied statistics the use of the sigma squared (σ^2) symbol is restricted to unknown population variances for which a measurement procedure gives an estimate only. By contrast, the symbol s^2 applies to variances of samples, and thus to finite sets of measurements. Standard methods on sampling of bulk materials often apply sigma-symbols (σ^2 or σ) indiscriminately.

NOTE 2 Following are definitions for the most relevant concepts and terms in mass measurement technology. They are presented to clarify the difference between this standard method, which quantifies the risk of losing and the probability of gaining in commercial transactions, and other methods that deal with mass measurement techniques from the perspective of regulatory agencies.

3.1
accuracy
generic term that implies closeness of agreement between an observed mass and its unknown true value

NOTE Accuracy is an abstract concept that cannot be quantified, but a lack of accuracy can be measured and quantified in terms of a bias or systematic error.

3.2
bias
difference between the expectation of the test result and an accepted reference value

NOTE This definition is only valid if the accepted reference value is known with absolute certainty (International Units of Mass and Length). Given that most accepted reference values are known within finite confidence limits, the difference between the expectation of a test result and an accepted reference value is only a bias if the expectation of the test result falls outside the confidence limits of an accepted reference value. [1)]

3.3
belt scale
mass measurement device that continuously integrates and records as a cumulative mass, the load on a belt while it passes the suspended scale section in a conveyor belt

NOTE Belt scales are continuous mass measurement devices that are calibrated by applying a load such as a calibrated chain on the belt above the scale section (dynamic), or a certified weight suspended from the scale's frame (static), for a specified integration period, or by measuring with the belt scale a quantity of material whose mass is measured with a static scale (material-run method).

3.4
bias detection limit
BDL
measure for the power or sensitivity of Student's t-test to detect a bias or systematic error between applied and observed loads

3.5
coefficient of variation
CV
measure for random variations in a mass measurement technique, numerically equal to the standard deviation as a percentage of the observed mass

3.6
confidence interval
CI
interval within which a predetermined percentage of the differences between all possible measurements and their mean is expected to cluster

3.7
confidence range
CR
range within which a predetermined percentage of all possible measurements is expected to cluster

NOTE In science and engineering 95 % confidence intervals and ranges are most frequently used.

---

1) For example, the mass of the lot is generally determined once only so that the measured value is not the expectation of the test result. In this International Standard a bias is the statistically significant difference between independent estimates of the wet mass of the lot (loading versus discharge, static versus dynamic scales) and mass measurements should be traceable to National Prototype Kilograms, and thus to the International Unit of Mass, through the shortest possible calibration hierarchy.

3.8
correlation coefficient
r
measure for the degree of association or interdependence between a set of certified weights and observed loads

3.9
draft survey
mass measurement technique that is based on converting the difference between a vessel's displacement under different loads into a mass on the basis of its draft tables while taking into account the density and temperature of water and ballast, and changes in ballast and supplies

NOTE Draft surveys are based on Archimedes's Principle which states that a floating body displaces its own mass. The wet mass of a cargo or shipment can be measured by converting changes in draft, trim, ballast and consumable supplies into mass on the basis of the vessel's draft table.

3.10
precision
generic term for the cumulative effect of random variations in a mass measurement technique

NOTE Precision is a generic qualifier, e.g. "a high degree of precision", "the precision is poor or low" or "the precision characteristics are excellent", are valid statements albeit without quantitative implications.

3.11
probable bias range
PBR
limits within which a measured bias is expected to fall at predetermined probabilities, either for a type I risk only or for type I and II risks

3.12
relative standard deviation
s_r
measure for random variations in a mass measurement technique, numerically equal to the standard deviation divided by the observed mass

3.13
standard deviation
s
measure for random variations in a mass measurement technique, numerically equal to the square root of the variance

3.14
static scale
mass measurement device that converts into a mass a static load on a weighbridge or on a platform, inside a hopper or suspended from a gantry scale

NOTE Static scales are batch mass measurement devices that are calibrated either with a single certified weight or with a set, and less frequently with a calibrated hydraulic press. Static scales may have automatic zero adjustment so that the sum of the differences between tare and gross loads can be used to generate a cumulative mass. Dual hopper scales allow a virtually continuous mass flow during loading and discharge operations without sacrificing the accuracy and precision characteristics of the static scale.

3.15
Student's *t*-value
t
ratio between the difference for the means for sets of applied and observed loads and the standard deviation for the mean difference

3.16
type I risk
α
risk of rejecting the hypothesis that the means for sets of applied and observed loads are compatible when their mean difference is, in fact, statistically identical to zero

3.17
type II risk
β
risk of accepting the hypothesis that the means for sets of applied and observed loads are compatible when their mean difference is, in fact, statistically different from zero

3.18
variance
s^2
measure for random variations in a mass measurement technique, numerically equal to the sum of squared deviations from the mean for a set of measurements divided by the number of measurements in the set minus 1 (divided by the degrees of freedom)

NOTE In textbooks on applied statistics the term “mean squared deviation from the mean” is often used in reference to the variance.

4 General remarks

International and national handbooks on weighing devices define uncertainties in mass measurement techniques in different ways. In some handbooks the use of the term “error” is restricted to a bias or systematic error while others refer to “maximum permissible risks”, which appears synonymous with “tolerances”, as a measure for random variations in a mass measurement technique.

Unless “maximum permissible errors” or “tolerances” are, by definition, equal to 95 % or 99 % confidence intervals, neither can be converted into a variance as the most basic measure for the precision of a measurement process. However, an unbiased estimate for the variance of the wet mass of a cargo or shipment of mineral concentrate is required before the precision for its dry mass and the masses of contained metals can be calculated and reported in terms of 95 % confidence intervals and ranges as a measure for the risk that trading partners encounter.

Annex D provides information for a step-by-step procedure for the testing of static scales.

4.1 Draft surveys

The difference between a vessel’s displacements, either before and after loading or before and after discharge, is converted into a wet mass on the basis of its draft table. Corrections are applied for changes in ballast and consumables such as fuel, potable water and supplies. Average densities of water, in ballast tanks and in proximity to the vessel during draft surveys, are measured and taken into account when converting a difference between the vessel’s displacements under different load conditions into a mass.

External factors, such as wind velocity and stratified salinity, limit the precision of draft surveys. Deformation of vessels, while in a partially loaded condition, adds another element of uncertainty that may translate into a bias. Displacement surveys for single cargo spaces are invariably less precise than displacement surveys for full cargoes. The highest degree of precision can be obtained when a vessel is surveyed at loading in a light (without ballast) and completely loaded condition, or at discharge in a completely loaded and light (without ballast) condition.

Moisture migration during the voyage would cause discrepancies between surveys at loading and discharge if drained water were removed with the bilge pumps. In such cases the wet mass measured at discharge may well be significantly lower than the wet mass at loading but the dry masses at loading and discharge are expected to be compatible. Oxidation often causes a small increase in mass that is difficult to estimate due to the highly variable degree of precision for draft surveys.

Generally, precision estimates in terms of coefficients of variation range from a low of 0,5 % to a high of 2,5 %. The lowest coefficients of variation were observed by comparing draft surveys at loading and discharge. If the marine surveyor at discharge has knowledge of the vessel's bill of lading (B/L), the draft surveys at the ports of discharge and loading are no longer statistically independent [1].

Draft surveys at loading are based on consensus between an officer of the vessel, a marine surveyor representing the shipper, and sometimes a marine surveyor representing the buyer. Under such conditions the precision of the draft surveys at loading cannot possibly be estimated. Only in the case that two or more qualified marine surveyors each complete their own draft surveys for the vessel, at the same time but independently, can the precision of this mass measurement technique be estimated in an unbiased manner.

The precision for a draft survey can also be estimated if the wet mass of a cargo or shipment is measured with a static scale with known precision characteristics, provided that it be located in close proximity to the vessel to ensure that loss of moisture and mechanical loss do not cause a bias. Unlike linearity for static mass measurement devices linearity for draft surveys cannot be defined in a meaningful manner due to the differences in the deformation of vessels over a wide range of loading conditions.

Annex C provides an example of a displacement calculation for a draft survey.

4.2 Belt scales

A belt scale is a continuous (dynamic) mass measurement device that integrates the variable load on a suspended belt section over long periods of time. Precision and bias for belt scales depend on numerous factors not the least of which is the environment in which they operate. A belt scale can be calibrated with a chain that is trailed on the belt over the scale's mechanism with a static weight that is suspended from the scale's frame, or with a quantity of material whose wet mass is measured with a static scale. Despite its relatively short time basis, the material-run test is the most reliable calibration procedure for dynamic scales [2].

A belt scale in series with a hopper scale integrated in a conveyor belt system can be calibrated, and its precision estimated, by comparing paired wet masses (static versus dynamic). Many applications would benefit from a pair of belt scales in series. Particles that become wedged between the conveyor's frame and the suspended frame of a belt scale cause discrepancies between paired measurements. Identification of anomalous differences permits corrective action to be taken. Removal of spillage from a belt scale's mechanism at regular intervals reduces drift, and thus the probability of a bias occurring.

A precision of 0,4 % in terms of a coefficient of variation has been observed for advanced belt scales under optimum conditions but under adverse conditions the coefficient of variation may well exceed 3,5 %. Reliable and realistic estimates for the precision of belt scales under routine conditions are obtained by measuring and monitoring variances between observed spans prior to each calibration. Frequent calibrations ensure that belt scales will generate unbiased estimates for wet mass. The central limit theorem implies that continuous weighing with dynamic scales gives a significantly lower precision for wet mass than batch weighing with static scales does.

Under routine conditions the linearity of belt scales is difficult to measure. Manufacturers of load cells test the linearity of response over 4 mA to 20 mA ranges. However, linearity under test conditions does not necessarily ensure linear responses to applied loads under routine conditions. Nonetheless, deviations from linearity are not likely to add more uncertainties to this mass measurement technique than other sources of variability such as belt tension and stiffness, stickiness of wet material or wind forces.

4.3 Weighbridges

The wet mass of cargoes or shipments of mineral concentrate is often measured by weighing trucks or wagons in empty and loaded condition at mines or ports, and in loaded and empty condition at ports or smelters. The precision for wet mass that is measured with a static scale such as a weighbridge, is perfectly acceptable for settlement purposes. The variance component that the measurement of wet mass contributes to the variance for contained metal is significantly lower than those for the measurement of moisture and metal contents [3].

The suspended mass of the scale's beam and its support structure is only a small part of gross loads. As a result, the variance for tare loads is significantly lower than the variance for gross loads which implies that the variance for the net wet mass of a single unit is largely determined by the variance for its gross load. After each cycle the weighbridge is zero adjusted, either automatically or manually, to eliminate drift.

Regulatory agencies may use one or more wagons of certified weight to calibrate weighbridges. Each wagon gives only one calibration point so that deviations from linearity are impossible to detect. By placing two wagons on a weighbridge a set of three [3] calibration points is obtained to provide useful but limited information on its linearity. The most effective test for linearity is based on addition or subtraction of a set of certified weighs that covers the working range of a weighbridge. Equally effective but more time consuming is alternately adding a single certified weight with a mass of 1 t to 2 t and a quantity of material until the weighbridge is tested in increments of 5 t to 10 t over its working range.

Precision parameters for weighbridges can be measured and monitored by weighing in duplicate once per shift, a truck or a wagon. After the gross weight of a randomly selected truck or wagon is measured in the usual manner, it is removed from the weighbridge. Next, the zero is checked and adjusted if required, and then the unit is moved on to the weighbridge and weighed again. The mean for sets of four or more absolute differences between duplicates can be used to calculate the variance for a single test result at gross loads. In terms of a coefficient of variation the precision for a weighbridge at gross loads generally ranges from 0,1 % up to 0,5 %.

The precision can also be estimated by placing on the weighbridge, in addition to the gross load, a test mass of five times up to ten times the scale's readability or sensitivity. Measurements with and without this test mass are recorded and the variance for gross loads calculated from a set of six data points up to 12 data points. Such estimates tend to be marginally but not significantly lower than the precision between duplicates that are generated by first weighing, and then removing and reweighing a loaded truck or wagon.

This procedure can be repeated without a load on the scale. A test mass is placed on the scale and its mass recorded. Next, the test mass is removed, and the zero adjusted if required. This process is repeated no less than six times, and the variance at near-zero loads calculated.

4.4 Hopper scales

The wet mass of cargoes or shipments can also be determined with a single hopper scale or with a pair of parallel hopper scales. Upon completion of each discharge cycle a hopper scale is often automatically zero adjusted so that a bias caused by build-up of wet material and dislodgement at random times is eliminated. Otherwise, tare loads for each weighing cycle should be recorded to allow for changes in accumulated mass.

A hopper scale is calibrated by suspending from its frame a set of certified weights with a mass of 1 t to 2 t each to cover its entire working range. It is possible but more time-consuming to calibrate a hopper scale with a single certified weight of 1 t to 2 t by alternatively adding a quantity of material, recording the applied mass, suspending the certified weight and recording the applied load again.

The precision can be estimated by placing on the hopper scale a test mass of five times up to ten times a scale's readability or sensitivity, recording measurements with and without this test mass, and calculating the variance for a single weighing cycle from six test results up to 12 test results. This check can be repeated after the discharge cycle to determine whether the precision is a function of load. In terms of a coefficient of variation the precision at gross loads generally ranges from 0,1 % up to 0,25 %.

Even though the hopper's suspended mass in the loaded condition adds most to the variance for net wet mass, its suspended mass in the empty condition is large enough to add to the variance for the net wet mass measured during each weighing cycle.

4.5 Gantry scales

The wet mass of cargoes or shipments of concentrates in bulk can be determined with a gantry scale. This mass measurement device is also zero adjusted, either manually or automatically, after each load is discharged. The wet mass contained in a fully loaded clamshell bucket is of the same order of magnitude as its suspended mass and support structure so that the variances for tare and gross loads both contribute to the variance for the net wet mass of each weighing cycle.

Only a single certified weight is required on location to maintain a gantry scale in a proper state of calibration. The precision of a gantry scale can be estimated by placing on the loaded clamshell a test mass of five times up to ten times its readability or sensitivity, recording measurements with and without this test mass and calculating the variance for single weighing cycles from sets of six test results up to 12 test results. It is possible to estimate the precision of a gantry scale with partially loaded clamshells. However, only during removal of the lowest stratum in a cargo space will partial loads be encountered so that neither the precision for partial loads nor the linearity of the gantry scale are matters of much concern.

In terms of a coefficient of variation the precision of gantry scales at gross loads generally ranges from 0,15 % up to 0,4 %. The variance for the net wet mass of single grabs is equal to the sum of the variances at gross and tare loads.

4.6 Platform scales

The wet mass of shipments of contained mineral concentrate can be measured by weighing bulk bags or other containers on a platform scale, either in the empty and the loaded condition at mines, or in the loaded and the empty condition at smelters. Platform scales are often used to measure the wet mass of valuable mineral concentrates so that a proper state of calibration is extremely important.

The suspended mass of the scale’s beam and its support structure is only a small part of the suspended mass at gross loads. As a result, the variance for the tare mass is significantly lower than the variance for the gross mass. The variance for the net wet mass of a container is equal to the sum of the high variance for the gross mass and the low variance for the tare mass which implies that the variance for the wet mass of a shipment is largely determined by the variance for the gross mass of containers. Unless gross masses differ substantially from the certified weight required to calibrate a platform scale, the linearity of this mass measurement device is not a matter of concern.

The precision of platform scales (near zero and at rated capacity) can be estimated by placing a test mass of five times up to ten times its readability or sensitivity on its platform, recording measurements with and without this test mass and calculating the variance for single weighing cycles from sets of six replicate test results up to 12 replicate test results. In terms of a coefficient of variation the precision for platform scales ranges from 0,05 % up to 0,2 % at gross loads. The variance for the net wet mass is equal to the sum of the variances at gross and tare loads.

5 Certified weights

The traceability of certified weights to the International Unit of Mass through National Prototype Kilograms and a hierarchy of verifiable calibrations is of critical importance. The integrity of certified weights can be ensured by storing them in a clean and dry environment, preferably on platforms or pallets, by covering them with tarpaulins to avoid corrosion and accumulation of dirt and by handling them carefully to avoid mechanical damage.

Based on how a traceable mass is compared with a draft survey or a measurement with a belt scale, or how a certified weight is compared with test results for a static mass measurement device, calibration methods can be divided into four categories, namely:

— a single certified weight of appropriate mass;

— a set of certified weights to cover a typical working range;

— a single, but preferably two wagons of certified weight;

— a mass traceable to a properly calibrated static scale.

Weighbridges (including in-motion and coupled-in-motion weighing devices) can also be calibrated with hydraulic pressure gauges. The use of a hydraulic pressure gauge adds to the calibration hierarchy a link that is based on a completely different technology.

6 Methods of operation

6.1 General

Precision and bias for mass measurement devices and techniques can be estimated and monitored as a function of time. Calibration data for static and dynamic scales not only generate information on bias but also reliable precision estimates for mass measurements. Calibrations require more time than simple precision checks with a test mass, therefore a case can be made that precision checks be carried out at regular intervals, and that precision be monitored on control charts. Sudden changes in precision may be indicative of mechanical failures or malfunctioning electronics, and require testing for conformance with the manufacturer’s specifications.

Testing for bias, estimating precision and checking linearity are based on applied statistics, and in particular on Student’s t-test, Fisher’s F-test (analysis of variance) and correlation-regression analysis.

Annex B reviews tests and formulae required to calculate relevant parameters.

6.2 Draft surveys

Precision and bias of draft surveys can be estimated and monitored by comparing wet masses that are determined at loading and discharge, by comparing wet masses determined by draft survey (either at loading or at discharge) or with a properly calibrated static weighing device in close proximity to the port of loading or discharge. The vessel’s bill of lading, which is almost invariably based on a draft survey at the port of loading, should not be disclosed to the marine surveyor at discharge until the draft survey is completed. Otherwise, the precision between draft surveys at loading and discharge cannot be estimated in an unbiased manner.

6.2.1 Draft surveys at loading and discharge

An example of draft surveys at loading and discharge can be found Table A.1, which lists a set of ten paired wet masses that are determined by draft surveys at loading and discharge. Each shipment was loaded into a single cargo space so that these results are typical for draft surveys of partially loaded vessels. Table 1 lists the statistical parameters for this paired data set.

Table 1 — Precision and bias between draft surveys

Parameter	Symbol	Value
Mean – load (t)	$\bar{x}\ (L)$	4 111,2
Mean – discharge (t)	$\bar{x}\ (D)$	4 106,9
Mean difference (t)	$\Delta \bar{x}$	− 4,3
Mean difference (%)	$\Delta \bar{x}$	− 0,1
Variance of differences (t^2)	$s^2(\Delta x)$	1 410,92
Coefficient of variation (%)	CV	0,91
Student’s t-value	t	0,361
Bias detection limits:		
Type I risk only (%)	BDL(I)	± 0,7
Type I & II risks (%)	BDL(I & II)	± 1,2

The variance of differences of 1 410,92 t^2 is the most basic measure for the precision between draft surveys at loading and discharge, while the coefficient of variation of 0,91 % is a more transparent measure for precision. The question is whether this estimate for the precision between draft surveys is unbiased, and thus whether draft surveys at loading and discharge are statistically independent.

If the marine surveyor at the port of discharge were to have prior knowledge of the vessel's bill of lading, the draft survey at discharge would no longer be statistically independent which implies that the coefficient of variation of 0,91 % is not expected to be an unbiased estimate for the precision between draft surveys at loading and discharge. Therefore, the vessel's bill of lading should be kept confidential until the draft survey at discharge is completed to ensure that the wet mass measured at the port of discharge is also an unbiased estimate for the unknown true mass.

If the draft surveys at loading and discharge were equally precise, the variance for a single draft survey would be:

$$\frac{1\,410,92}{2} = 705,46 \text{ t}^2$$

for standard deviation of:

$$\sqrt{705,46} = 26,56 \text{ t}$$

and a coefficient of variation of:

$$\frac{26,56 \times 100}{\left[(4\,111,2 + 4\,106,9)/2\right]} = 0,65\%$$

Means of 4 111,2 t and 4 106,2 t are used to calculate the coefficient of variation. In this case the means are statistically identical but the mean of statistically different means can still be used to calculate the coefficient of variation. However, numerically it is not the most reliable precision estimate.

Because such a large set of variables interact in this mass measurement technique, the probability that displacement surveys at loading and discharge are equally precise is remote. Subclause 6.2.2 shows that this variance of differences of 1 410,92 t^2 is not an unbiased estimate for the precision between draft surveys at loading and at discharge.

The calculated t-value of 0,361 for a mean difference of 4,3 t does not exceed the tabulated value of $t_{0,95;9}$ = 2,262 which implies that means of 4 111,2 t at loading and 4 106,9 t at discharge are statistically identical. Hence, each draft survey appears to generate an unbiased estimate for the unknown true wet mass of the shipment in question. The probability of this t-value of 0,361 being caused by random variations falls between 20 % and 30 % so that the closeness of agreement is not suspect.

BDLs of ± 0,7 % or ± 27 t for the type I risk only, and ± 1,2 % or ± 49 t for type I and II risks, are different measures for the sensitivity or power of Student's t-test to detect a bias. BDLs are also measures for symmetrical risks of losing and probabilities of gaining if the settlements between trading partners were based on measuring the wet mass of shipments by draft surveys.

Based on a standard deviation of 26,56 t^2 for a single displacement survey and a tabulated t-value of: $t_{0,95;9}$ = 2,262, the 95 % confidence interval (95 % CI) for a cargo or shipment with a wet mass of 4 109 t is:

$$2,262 \times 26,56 = \pm 60 \text{ t}$$

For a 95 % confidence range (95 % CR) from 4 109 − 60 = 4 049 t up to 4 109 + 60 = 4 169 t. Table 2 lists precision estimates based on the mean of means of 4 109 t and a variance of 705,46 t^2.

Table 2 — Precision for wet mass by draft survey

Parameter	Symbol	Value
Mean (t)	M_w	4 109
Variance (t^2)	$s^2(M_w)$	705,46
Standard deviation (t)	$s(M_w)$	26,56
Coefficient of variation (%)	CV	0,65
95 % Confidence interval (t) [a]	95 % of CI	± 60,1
95 % Confidence interval (%)	95 % of CI	± 1,5
95 % Confidence range:		
lower limit (t)	95 % of CRL	4 049
upper limit (t)	95 % of CRU	4 169

[a] Based on $t_{0,95;9} \times s(M_w)$.

If the long-term coefficient of variation were 0,8 %, the 95 % confidence interval for a wet mass of 4 109 t would be:

$$\frac{1{,}96 \times 4\,109 \times 0{,}8}{100} = \pm\,64{,}4\ \text{t}$$

for a 95% confidence range from 4 109 – 64,4 = 4 045 t up to 4 109 + 64,4 = 4 173 t. The z-value of 1,96 from the normal or Gaussian distribution is often rounded to 2 which would change the 95 % confidence interval from ± 64 t to ± 66 t, a difference that is well within the precision of this mass measurement technique.

The precision estimates in Table 2 are only valid if the variance of differences is unbiased and if the draft surveys at loading and discharge are equally precise. The question whether the draft surveys at loading and discharge are indeed equally precise could be solved by estimating the precision at loading and at discharge from statistically independent draft surveys. In other words, were two or more marine surveyors to measure independently a vessel's draft in the light and loaded condition, a set of no less than four duplicate or replicate draft surveys, on similar vessels and under comparable conditions, would be required to estimate the precision of draft surveys at a particular port.

The question whether a variance of differences is an unbiased estimate for the precision between draft surveys at loading and discharge can be solved by comparing the results of draft surveys with wet masses measured with a static scale. In draft surveys wet masses measured with a static scale at discharge are compared with wet masses estimated with a weighbridge at discharge.

6.2.2 Draft survey versus weighbridge

A comparison of wet masses by draft surveys and with a weighbridge can be found in Table A.2, which lists a set of ten pairs of wet masses for the same shipments that were also reported in Table A.1. In this case wet masses that were measured by draft surveys at the port of discharge are compared with wet masses that were measured with a weighbridge for trucks at the smelter.

The set of paired mass measurements is tested for bias by calculating the t-value for the mean difference, the variance of differences and the number of paired data in the set. In this example the variance of differences is a measure for the precision between mass measurement techniques with vastly different precision characteristics. Under such conditions the variance of difference is virtually identical to the variance for the least precise mass measurement technique (draft surveys at discharge).

Table 3 lists the most relevant statistics for this set.

Table 3 — Precision and bias between different techniques

Parameter	Symbol	Value
Mean – draft survey (t)	$\bar{x}\ (D)$	4 106,9
Mean – weighbridge (t)	$\bar{x}\ (W)$	4 134,3
Mean difference (t)	$\Delta \bar{x}$	+ 27,4
Mean difference (%)	$\Delta \bar{x}$	+ 0,7
Variance of differences (t²)	$s^2(\Delta x)$	13 243
Coefficient of variation (%)	CV	2,8
Student's t-value	t	0,753
Bias detection limits:		
Type I risk only (%)	BDL(I)	± 2,0
Type I & II risks (%)	BDL(I & II)	± 3,6

The coefficient of variation of 2,8 % is a measure for the precision between draft surveys at discharge and wet masses determined with a weighbridge at the smelter. In 6.2.1 the precision between draft surveys at loading and discharge in terms of a coefficient of variation came out at 0,91 %. The question whether coefficients of variation of 2,8 % and 0,91 % are compatible can be solved by comparing the calculated F-ratio of

$$\frac{13\ 243}{1\,410,92} = 9,39$$

(the variance between draft surveys at discharge and wet masses measured with a weighbridge at a smelter, divided by the variance between draft surveys at loading and discharge) with tabulated values of $F_{0,95;9,9}$ = 3,18 and $F_{0,99;9,9}$ = 5,35. The calculated value of 9,39 exceeds tabulated values at the 95 % and 99 % probability levels. Hence, the probability that coefficients of variation of 2,8 % and 0,91 % are statistically identical is much less than 1 %.

Thus it would appear that knowledge of the vessel's bill of lading before the draft survey at discharge is completed, results in statistical dependencies between draft surveys at loading and discharge. Therefore, the coefficient of variation of 0,91 % is a biased estimate for the precision between draft surveys and the coefficient of variation of 2,8 % is a better estimate for the precision of single draft surveys for partially loaded vessels.

The weighbridge's precision is expected to add significantly less than

$$\frac{1\ 410,92}{2} = 705,46\ \text{t}^2$$

to the variance of differences of 13 243 t² so that a variance of 13 243 − 705,46 ≈ 12 500 t² would be a better estimate for the precision of a single draft survey than the variation of 705,46 t². In terms of a coefficient of variation the precision for draft surveys for a single cargo space would then be

$$\frac{\sqrt{12\ 500} \times 100}{\left[(4\ 106,9 + 4\ 134,3)/2\right]} = 2,7\ \%$$

A calculated t-value of 0,753 for a mean difference of 27,4 t does not exceed the tabulated value of $t_{0,95;9}$ = 2,262 which implies that means of 4 106,9 t at loading and 4 134,3 t at discharge are statistically identical. Hence, the draft survey at discharge and the weighbridge at discharge apparently generate unbiased estimates for the unknown true wet mass of each shipment. Nonetheless, the precision of a static scale such as a weighbridge installs a significantly higher degree of confidence in a cumulative wet mass of 4 134,4 t than the precision of draft surveys does.

Bias Detection Limits of ± 2,0 % or ± 82 t for the type I risk only, and ± 3,6 % or ± 149 t for type I and type II risks, are measures of the power or sensitivity of this test to detect a bias. Generally, Bias Detection Limits are also estimates for the risk of one trading partner to losing and an identical probability of the other trading partner to gaining. In this case, however, the settlements were based on wet masses determined with the weighbridge so that the risk was much less than BDLs of ± 2,0 % and ± 3,6 % imply.

Precision estimates for the wet mass of a single cargo space or a complete cargo, and for the cumulative mass of a set, are calculated in the same manner. For example, a variance of 12 500 t^2 and a single wet mass of 4 107 t for draft surveys at discharge are equivalent to a 95 % confidence interval of:

$$2 \times \sqrt{12\,500} = \pm 224 \text{ t}$$

for a 95 % confidence range from 4 107 − 224 = 3 883 t up to 4 107 + 224 = 4 331 t.

Table 4 lists precision estimates that are based on a single wet mass of 4 107 t, a cumulative wet mass of 41 343 t, a variance of 12 500 t^2 for the single wet mass, and the sum of variances of 125 000 t^2 for the cumulative wet mass.

The coefficient of variation of 2,7 %, when divided by $\sqrt{10}$, becomes:

$$\frac{2,7}{3,16} = 0,9\%$$

This relationship is based on the central limit theorem, an important theorem in mathematical probability and applied statistics.

6.3 Belt scales

An example of how to calculate the prevision of wet masses measured with belt scales can be found in Table A.3, which table lists a set of 12 chain spans, recorded at weekly intervals prior to calibration and a similar set of spans that were obtained immediately following its calibration. Table 5 lists the basic statistical parameters for each moving data base.

Coefficients of variation of 0,39 % and 0,11 % are both measures for the precision of this belt scale. However, the calculated F-ratio of

$$\frac{0,1976}{0,0152} = 13,00$$

between the variances before and after calibration exceeds the tabulated values of $F_{0,95;11;11}$ = 2,82 and $F_{0,99;11;11}$ = 4,64, which implies that these variances differ significantly. The long-term variance of 0,197 6 between chain spans prior to calibration more truly reflects the magnitude of random variations in mass measurement with this belt scale as a function of time. Therefore, the coefficient of variation of 0,39 % is the more reliable estimate for its precision under routine conditions.

The question whether the belt scale generates unbiased estimates for wet mass can be solved by applying Student's t-test to the difference between the required span (115,25 for the belt scale in this example) and the mean of observed spans for a set that constitutes a moving data base. Table 6 lists the results of this test.

Table 4 — Precision for wet mass by draft survey

Parameter	Symbol	Single	Cumulative
Mean (t)	M_w	4 107	41 343
Variance (t^2)	$s^2(M_w)$	12 500	125 000
Standard deviation (t)	$s(M_w)$	111,8	353,6
Coefficient of variation (%)	CV	2,7	0,9
95 % Confidence interval (t) [a]	95 % of CI	± 224	± 707
95 % Confidence interval (%)	95 % of CI	± 5,4	± 1,7
95 % Confidence range:			
lower limit (t)	95 % of CRL	3 883	40 636
upper limit (t)	95 % of CRU	4 331	42 050

[a] Based on $z_{0,95} \times s(M_w)$, or $z_{0,95} \times s(\Sigma M_w)$.

Table 5 — Precision of a belt scale

Parameter	Symbol	Before	After
Mean (scale units)	$\bar{x}$	115,12	115,36
Variance (scale units)^2	$s^2(x)$	0,197 6	0,015 2
Standard deviation (scale units)	$s(x)$	0,444 6	0,123 4
Coefficient of variation (%)	CV	0,39	0,11

Table 6 — Testing a belt scale for bias

Parameter	Symbol	Value
Mean – required span	$\bar{x}\ (R)$	115,25
Mean – observed span	$\bar{x}\ (O)$	115,12
Mean difference (span units)	$\Delta \bar{x}$	− 0,13
Student's t-value	t	1,013
Significance	—	ns [a]
Bias detection limits:		
Type I risk only (span units)	BDL(I)	± 0,28
Type I & II risks (span units)	BDL(I & II)	± 0,51

[a] ns = not significant.

The difference of − 0,13 scale units between the required span of 115,25 and the mean of 115,12 for all test data, results in a calculated t-value of 1,013 which is below the tabulated value of $t_{0,95;11}$ = 2,201 so that the belt scale is in a proper state of calibration. BDLs of ± 0,28 for the type I risk, and ± 0,51 for type I and II risks, indicate that a mean difference of 0,13 span units is most probably due to random variations.

A belt scale need not be adjusted if the difference between the required span and the moving average of the running data base does not exceed the BDL for type I and II risks. Upon completion of the chain test the observed span is added to the data base while the first observed span in the running data base is removed. The number of test data to be retained in the moving data base depends on the required BDLs and ranges from 8 to 16.

For a wet mass of 25 000 t the coefficient of variation of 0,39 % gives a variance of

$$\left(\frac{0{,}39 \times 25\,000}{100}\right)^2 = 9\,506\ \text{t}^2$$

a standard deviation of

$$\frac{0{,}39 \times 25\,000}{100} = 97{,}5\ \text{t}$$

a 95 % confidence interval of 2 × 97,5 = ± 195 t and a 95 % confidence range from 25 000 − 195 = 24 805 t to 25 000 + 195 = 25 195 t.

Table 7 lists precision parameters for a wet mass of 25 000 t based on a coefficient of variation of 0,39 %.

Although belt scales have found wide application in mining and mineral processing, a wet mass determined with a belt scale contributes a large component to the variances for metals contained in concentrates. Therefore, wet masses of concentrate shipments on which settlements between mines and smelters are based should not be determined with belt scales.

Table 7 — Precision for wet mass with a belt scale

Parameter	Symbol	Value
Mean (t)	M_w	25 000
Coefficient of variation (%)	CV	0,39
Standard deviation (t)	$s(M_w)$	97,5
95 % Confidence interval (t)	95 % of CI	± 195
95 % Confidence interval (%)	95 % of CI	± 0,8
95 % Confidence range:		
lower limit (t)	95 % of CRL	24 805
upper limit (t)	95 % of CRU	25 195

6.4 Weighbridges

Table A.4 presents an example of how to check the state of calibration for a weighbridge and how to estimate its precision. It lists a set of paired test data for a weighbridge that was calibrated with a pair of wagons with certified mass of 31 890 kg and 70 810 kg respectively. Two subsets of four test data were generated by determining the mass of each wagon while the third subset was obtained by weighing both wagons simultaneously.

Table 8 summarizes the most important statistical parameters for this set of calibration data.

Table 8 — Precision and bias for a weighbridge

Parameter	Symbol	Value
Mean – applied loads (kg)	$\bar{x}\ (A)$	68 467
Mean – observed loads (kg)	$\bar{x}\ (O)$	68 451
Mean difference (kg)	$\Delta\bar{x}$	− 16
Number of test data	n	12
Variance of differences (kg²)	$s^2(\Delta x)$	445
Coefficient of variation (%)	CV	0,03
Student's t-value	t	2,601
Significance	—	[a]

[a] Significant at 95 % probability.

The mean difference of − 16 kg results in a calculated t-value of 2,601 which exceeds a tabulated value of $t_{0,95;11}$ = 2,201 but is still below $t_{0,99;11}$ = 3,106. This mean difference falls between BDLs of ± 13 kg for the type I risk only, and ± 24 kg for type I and II risks. Hence, this weighbridge is in a proper state of calibration if type I and II risks are both taken into account.

For a weighbridge a coefficient of variation of 0,03 % is exceptionally low so that the power or sensitivity of the t-test to detect a bias is high. Therefore, the weighbridge's state of calibration and its precision are perfectly acceptable for commercial applications. The question of whether precision is a function of load, can be checked by applying correlation-regression analysis to applied loads and differences between certified weights and observed masses. The correlation coefficient of − 0,170 is statistically identical to zero. Hence, there is no evidence that the precision of this weighbridge is a function of applied load.

The variance for the wet mass of the contents of a wagon is equal to the sum of the variances for gross and tare masses. For example, if gross and tare masses of 120 000 kg and 20 000 kg respectively were measured using this weighbridge, the sum of the variances and thus the variance for the wet mass would be

$$\left(\frac{120\ 000 \times 0{,}03}{100}\right)^2 + \left(\frac{20\ 000 \times 0{,}03}{100}\right)^2 = 1\ 296 + 36 = 1\ 332\ \text{kg}^2$$

Evidently, the measurement of gross mass largely determines the variance for wet mass.

Table 9 lists precision parameters that would apply to a single wagon with a wet mass of 100 t and to a set of 250 wagons with a cumulative wet mass of 25 000 t.

Table 9 — Precision for wet mass with a weighbridge

Parameter	Single wagon	250 wagons
Wet mass (t)	100	25 000
Variance (t²)	0,000 9	0,225
Standard deviation (t)	0,03	0,47
95 % Confidence interval (t) [a]	± 0,06	± 0,95
95 % Confidence interval (%)	± 0,06	± 0,004
95 % Confidence range:		
lower limit (t)	99,94	24 999
upper limit (t)	100,06	25 001

[a] Based on $2 \times s(M_w)$, or $2 \times s(\Sigma M_w)$.

The weighbridge's linearity can be checked by applying correlation-regression analysis to the set of paired calibration data. Table A.5 summarizes the results for the complete set of paired data and for the means of each subset. The test for paired means has only one degree of freedom and is therefore much less robust than the test for paired data with ten degrees of freedom.

Table 10 summarizes correlation-regression parameters for all data and for means only.

Table 10 — Linearity of a weighbridge

Parameter	Symbol	All data	Means only
Correlation coefficient	r	1,000	1,000
Significance	—	— [a]	— [a]
Slope	m	0,999 9	0,999 9
Significance	—	— [a]	— [a]
Intercept (kg)	α	− 7,7	− 7,7
Significance	—	ns [b]	ns [b]

[a] Significant at 99,9 % probability.

[b] ns = not significant.

The t-test can be applied to the slope and intercept of the regression line. The slope usually ranges from a minimum of 0,999 8 to a maximum of 1,000 2 and the intercept should not be statistically significant.

6.5 Hopper scales

Table A.6 presents an example of how to use the differences between applied and observed loads for a hopper scale for checking its state of calibration and estimating its precision. It lists test data for a hopper scale that was calibrated using a set of certified weights with a mass of 2 000 kg each and the statistical parameters for this set of calibration data. After the scale's zero was adjusted, the first certified weight was placed on the frame underneath the hopper and the observed mass recorded. Additional weights were placed on the scale until the complete set of 12 certified weights was loaded on the scale. Table 11 lists the most relevant statistical parameters for this set of calibration data.

Table 11 — Precision and bias for a hopper scale

Parameter	Symbol	Value
Mean – applied loads (kg)	$\bar{x}\,(A)$	13 000
Mean – observed loads (kg)	$\bar{x}\,(O)$	13 003
Mean difference (kg)	$\Delta\bar{x}$	+ 3
Number of test data	n	12
Variance of differences (kg^2)	$s^2(\Delta x)$	46
Coefficient of variation (%)	CV	0,05
Student's t-value	t	1,410
Significance	—	ns [a]

[a] ns = not significant.

The mean difference of 3 kg results in a calculated t-value of 1,410 which is below the tabulated value of $t_{0,95;11}$ = 2,201 and thus below the BDL of ± 4 kg for the type I risk only. Hence, this hopper scale is in a proper state of calibration, even when only the type I risk is taken into account.

A coefficient of variation of 0,05 % for a single measurement with a hopper scale is excellent. Therefore, the hopper scale's state of calibration and its precision are acceptable for commercial applications. The question whether precision is a function of load can be checked by applying correlation-regression analysis to applied loads and differences between certified and observed masses. The correlation coefficient of − 0,114 is statistically identical to zero. Hence, there is no evidence that the precision of this hopper scale is a function of applied load.

The variance for the wet mass of a hopper load is equal to the sum of the variances for empty and loaded conditions, For example, if the hopper contained 24 000 kg, the variance in the loaded condition would be

$$\left(\frac{24000 \times 0,05}{100}\right)^2 = 144 \text{ kg}^2$$

With automatic zero adjustments between discharge cycles the variance for an empty hopper with its large suspended mass is not expected to be significantly less than 144 kg^2 so that the variances for a net wet mass of 24 000 kg would be 288 kg^2. Table 12 lists precision parameters that would apply to a single hopper load with a wet mass of 24 t and to a set of 1 000 hopper loads with a cumulative wet mass of 24 000 t.

Table 12 — Precision for wet mass with hopper scale

Parameter	Single cycle	1 000 cycles
Wet mass (t)	24	24 000
Variance (t^2)	0,000 3	0,288 0
Standard deviation (t)	0,017	0,54
Coefficient of variation	0,07	0,002
95 % Confidence interval (t) [a]	± 0,034	± 1,07
95 % Confidence interval (%)	± 0,14	± 0,004
95 % Confidence range:		
lower limit (t)	23,97	23 998,9
upper limit (t)	24,03	24 001,1

[a] Based on 2 × $s(M_w)$, or 2 × $s(\Sigma M_w)$.

The scale's linearity can be checked by applying correlation-regression analysis to the set of paired calibration data. Table A.7 lists the results for this paired data set and Table 13 summarizes the correlation-regression parameters for the set of paired means only.

Table 13 — Linearity of a hopper scale

Parameter	Symbol	Value
Correlation coefficient	r	1,000
Significance	—	— [a]
Slope	m	0,999 9
Significance	—	— [a]
Intercept (kg)	α	+ 4,1
Significance	—	ns [b]

[a] Significant at 99,9 % probability.

[b] ns = not significant.

The t-test can be applied to the slope and intercept of the regression line. The slope usually ranges from a minimum of 0,999 8 to a maximum of 1,000 2 and the intercept should not be statistically significant.

6.6 Gantry scales

Table A.8 presents an example for precision and bias for gantry scales on the basis of a set of paired test data. The first pair was obtained by zero adjusting the scale with the clamshell bucket empty, suspending a certified weight with a mass of 2 000 kg from the clamshell and recording the observed mass (1 994 kg). The next pair was obtained by recording the mass of the partially loaded clamshell bucket (2 102 kg), suspending the certified weight from the clamshell and then recording the observed mass (4 105 kg). The process of adding about 2 t of material to the clamshell bucket, recording the observed mass, and then adding the certified weight of 2 000 kg was repeated until the clamshell bucket was loaded to its rated capacity. Table 14 lists the most important statistical parameters for this set of calibration data.

Table 14 — Precision and bias for gantry scale

Parameter	Symbol	Value
Mean – applied loads (kg)	$\bar{x}\,(A)$	9 027
Mean – observed loads (kg)	$\bar{x}\,(O)$	9 026
Mean difference (kg)	$\Delta\bar{x}$	− 1
Coefficient of variation (%)	CV	0,11
Student's t-value	t	0,280
Significance	—	ns [a]

[a] ns = not significant.

The mean difference of −1 kg results in a calculated t-value of 0,28 which is below the tabulated value of $t_{0,95;7}$ = 2,365. In fact, a mean difference of −1 kg is below BDLs of ± 8 kg for the type I risk only and ± 13 kg for type I and II risks, hence this gantry scale is in a perfect state of calibration.

For a gantry scale a coefficient of variation of 0,11 % is acceptable and the power or sensitivity of the t-test to detect a bias is high. Thus the gantry scale's state of calibration and its precision are acceptable for commercial applications. The question whether its precision is a function of load can be checked by applying correlation-regression analysis to the means of, and the differences between, the applied and observed loads. The correlation coefficient of 0,085 is statistically identical to zero. Thus there is no evidence that the precision of this gantry scale is a function of applied load.

The variance for the wet mass of the content of a clamshell bucket is equal to the sum of the variances at gross and zero loads. Based on a coefficient of variation of 0,11 % the variance for a gross load of 10 000 kg is

$$\left(\frac{0{,}11 \times 10\ 000}{100}\right)^2 = 121\ \text{kg}^2$$

Because the scale is linear, the variance for the empty clamshell bucket is expected to be 121 kg^2, so that the variance for the net wet mass of 10 000 kg is 242 kg^2.

Table 15 lists precision parameters that would apply to a single clamshell load with a net wet mass of 10 000 kg and to a set of 2 500 loads with a cumulative wet mass of 25 000 t.

Table 15 — Precision for wet mass with a gantry scale

Parameter	Single load	2 500 loads
Wet mass (t)	10	25 000
Variance (t^2)	0,000 242	0,605
Standard deviation (t)	0,015 6	0,778
Coefficient of variation	0,16	0,03
95 % Confidence interval (t) [a]	± 0,03	± 1,56
95 % Confidence interval (%)	± 0,3	± 0,01
95 % Confidence range:		
lower limit (t)	9,97	24 998,5
upper limit (t)	10,03	25 001,6

[a] Based on $2 \times s(M_w)$, or $2 \times s(\Sigma M_w)$.

The gantry scale's linearity can be checked by applying correlation-regression analysis to the set of paired calibration data. Table A.9 lists the results for the set of calibration data.

Table 16 summarizes the correlation-regression parameters for the set.

Table 16 — Linearity of a gantry scale

Parameter	Symbol	Value
Correlation coefficient	r	1,000
Significance	—	— [a]
Slope	m	1,000 2
Significance	—	— [a]
Intercept (kg)	α	− 2,6
Significance	—	ns [b]

[a] Significant at 99,9 % probability.

[b] ns = not significant.

The t-test can be applied to the slope and intercept of the regression line. The slope usually ranges from a minimum of 0,999 8 to a maximum of 1,000 2 and the intercept should not be statistically significant.

6.7 Platform scales

How to check a platform scale's state of calibration and how to estimate its precision.

Table A.10 lists two sets of calibration data for a platform scale and the most important statistical parameters for each set. This type of static scale can be used to determine the wet mass of concentrate shipments in bulk bags with a capacity of approximately 2 000 kg each.

Table 17 lists the statistical parameters for each set of calibration data.

A mean difference of 5 kg for the first set of calibration data gives a calculated r-value of 6,124 which exceeds a tabulated value of $t_{0,99;5}$ = 4,032 by a considerable margin but is still below $t_{0,999;5}$ = 6,859. This mean difference exceeds the BDL of ± 2,1 kg for the type I risks only and ± 3,7 kg for type I and II risks. Hence the first data set indicates that the platform scale is not in a proper state of calibration.

The lower and upper limits of PBRs for the type I risk range from PBL(I) = 2,9 kg to PBU(I) = 7,1 kg and from PBL(I&II) = 1,3 kg to PBU(I&II) = 8,7 kg for type I and II risks. These lower and upper limits are estimates for the range within which the observed bias of 5 kg is expected to fall when either a type I risk only, or type 1 and II risks, are taken into account.

The mean difference of 0,3 kg for the second set of calibration data gives a calculated t-value of 0,42 which is far below the tabulated value of $t_{0,95;5}$ = 2,571. Nor does this mean difference exceed the BDL of ± 1,8 kg for the type I risk. Hence, the second set of calibration data shows that the scale is in a perfect state of calibration.

Table 17 — Precision and bias of a platform scale

Parameter	Symbol	First	Second
Certified weight (kg)	$\bar{x}\ (C)$	2 000	2 000
Mean – observed loads (kg)	$\bar{x}\ (O)$	2 005	2 000,3
Mean difference (kg)	$\Delta\bar{x}$	+ 5	+ 0,3
Coefficient of variation (%)	CV	0,10	0,09
Student's *t*-value	t	6,124	0,420
Significance	—	— [a]	ns [b]
Bias detection limits (kg):			
Type I risk only	BDL(I)	± 2,1	± 1,8
Type I & II risks	BDL(I&II)	± 3,7	± 3,3
Probable bias ranges			
Type I risk only:			
lower limit (kg)	PBL(I)	2,9	na [c]
upper limit (kg)	PBU(I)	7,1	na
Type I & II risks:			
lower limit (kg)	PBL(I&II)	1,3	na
upper limit (kg)	PBU(I&II)	8,7	na

[a] Significant at 99 % probability.

[b] ns = not significant.

[c] na = not applicable.

The F-ratio of

$$\frac{4}{3{,}07} = 1{,}30$$

between the variances of 4 kg^2 for the first set and 3,07 kg^2 for the second set does not exceed the tabulated value of $F_{0,95;5;5}$ = 5,05. Therefore, variances of 4 kg^2 and 3,07 kg^2 are statistically identical, which implies that the precision of the platform scale remained constant during the calibration process.

For bulk bags with similar gross and tare masses, the linearity of this type of scale is not a cause for concern. The variance for the net wet mass of a single bulk bag's content is the sum of the variances for gross and tare masses.

For example, the sum of the variance of

$$\left(\frac{0{,}09 \times 2\ 050}{100}\right)^2 = 3{,}404 \text{ kg}^2$$

for a gross mass of 2 050 kg and

$$\left(\frac{0{,}09 \times 50}{100}\right)^2 = 0{,}002 \text{ kg}^2$$

for the tare mass of 50 kg, would result in a variance of 3,406 kg^2 for the net wet mass of 2 000 kg in a single bulk bag.

Table 18 lists precision parameters that would apply to the net wet mass of 2 000 kg in a single bulk bag and to a set of 500 bulk bags with a cumulative wet mass of 1 000 000 kg.

Table 18 — Precision for wet mass with a platform scale

Parameter	Single bag	500 bags
Wet mass (kg)	2 000	1 000 000
Variance (kg^2)	3,406	1 703
Standard deviation (kg)	1,85	41,3
95 % Confidence interval (kg) [a]	± 3,7	± 82,5
95 % Confidence interval (%)	± 0,18	± 0,01
95 % Confidence range:		
lower limit (kg)	1 996	999 917
upper limit (kg)	2 004	1 000 083

[a] Based on 2 × $s(M_w)$, or 2 × $s(\Sigma M_w)$.

Annex A
(informative)

Tables

Table A.1 — Precision and bias for draft surveys at loading and discharge

Ship	Loaded t	Discharged t	Difference t
1	3 675,4	3 727	+ 51,6
2	3 307,2	3 283,1	− 24,1
3	4 086,7	4 093,3	+ 6,6
4	3 867,9	3 808,7	− 59,2
5	4 002,8	4 014,6	+ 11,8
6	5 465,8	5 424,1	− 41,7
7	4 100,9	4 087,7	− 13,2
8	4 688,3	4 666	− 22,3
9	4 003,7	4 062,2	+ 58,5
10	3 913,2	3 902,3	− 10,9
Parameter		**Symbol**	**Value**
Sum-loaded (t)		$M_w(L)$	41 111,9
Sum-discharged (t)		$M_w(D)$	41 069
Difference (t)		Δx	42,9
Mean-loaded (t)		$\bar{x}\ (L)$	4 111,2
Mean-discharged (t)		$\bar{x}\ (D)$	4 106,9
Mean difference (t)		$\Delta \bar{x}$	− 4,3
Mean difference (%)		$\Delta \bar{x}$	− 0,1
Variance of differences (t^2)		$s^2(\Delta x)$	1 410,92
Standard deviation (t)		$s(\Delta x)$	37,56
Coefficient of variation (%)		CV	0,91
Number of paired data		n	10
Variance of mean difference		$s^2(\Delta \bar{x})$	141,09
Standard deviation		$s(\Delta \bar{x})$	11,88
Student's t-value		t	0,361
Significance		—	ns [a]
Bias detection limits:			
Type I risk only (t)		BDL(I)	± 27
Type 1 & II risks (t)		BDL(I&II)	± 49
Tabulated t-values		$t_{0,90;9}$	1,833
		$t_{0,95;9}$	2,262
		$t_{0,99,9}$	3,250
		$t_{0,999;9}$	4,781

[a] ns = not significant.

Table A.2 — Precision and bias for draft surveys and weighbridge at discharge

Lot	Draft survey t	Weighbridge t	Difference t
1	3 727	3 668,9	− 58,1
2	3 283,1	3 289	+ 5,9
3	4 093,3	3 991,8	− 101,5
4	3 808,7	3 835,1	+ 26,4
5	4 014,6	4 036,5	+ 21,9
6	5 424,1	5 722,4	+ 298,3
7	4 087,7	4 061,7	− 26
8	4 666	4 609,4	− 56,6
9	4 062,2	4 091,3	+ 29,1
10	3 902,3	4 036,8	+ 134,5

Parameter	Symbol	Value
Sum-draft survey (t)	$M_w(D)$	41 069
Sum-weighbridge (t)	$M_w(W)$	41 342,9
Difference (t)	Δx	+ 273,9
Mean-draft survey (t)	$\bar{x}\,(D)$	4 106,9
Mean-weighbridge (t)	$\bar{x}\,(W)$	4 134,3
Mean difference (t)	$\Delta\bar{x}$	+ 27,4
Mean difference (%)	$\Delta\bar{x}$	+ 0,7
Variance of differences (t[2])	$s^2(\Delta\bar{x})$	13 243
Standard deviation (t)	$s(\Delta\bar{x})$	115,1
Coefficient of variation (%)	CV	2,8
Number of paired data	n	10
Variance of mean difference	$s^2(\Delta\bar{x})$	1 324,3
Standard deviation	$s(\Delta\bar{x})$	36,4
Student's t-value	t	0,753
Significance	—	ns [a]
Bias detection limits:		
Type I risk only (t)	BDL(I)	± 82
Type I & II risks (t)	BDL(I&II)	± 149
Tabulated t-values	$t_{0,90;9}$	1,833
	$t_{0,5;9}$	2,262
	$t_{0,99,9}$	3,250
	$t_{0,999;9}$	4,781

[a] ns = not significant.

Table A.3 — Precision and bias for belt scales — required and observed spans before and after calibration

Test		Before scale units	After scale units
1		115,02	115,19
2		114,83	115,31
3		115,61	115,33
4		115,35	115,34
5		115,87	115,42
6		114,48	115,51
7		114,44	115,45
8		114,71	115,58
9		115,46	115,14
10		115,12	115,36
11		115,29	115,30
12		115,32	115,34
Parameter	**Symbol**	**Before**	**After**
Mean (scale units)	$\bar{x}$	115,12	115,36
Variance (scale units)^2	$s^2(x)$	0,197 6	0,015 2
Standard deviation (scale units)	$s(x)$	0,444 6	0,123 4
Coefficient of variation (%)	CV	0,39	0,11
Required chain span	$\bar{x}\ (R)$	115,25	115,25
Observed chain span	$\bar{x}\ (O)$	115,12	115,36
Mean difference	$\Delta\bar{x}$	− 0,13	+ 0,11
Number of test data	n	12	12
Variance of mean	$s^2(\bar{x})$	0,016 5	0,001 3
Standard deviation	$s(\bar{x})$	0,128 3	0,035 6
Student's t-value	t	1,013 1	3,091
Significance	—	ns [a]	— [b]
Bias detection limits:			
Type I risk only (t)	BDL(I)	± 0,28	± 0,08
Type I & II risks (t)	BDL(I&II)	± 0,51	± 0,14
Tabulated t-values	$t_{0,90;11}$	1,796	
	$t_{0,95;11}$	2,201	
	$t_{0,99,11}$	3,106	
	$t_{0,999;11}$	4,437	

[a] ns = not significant.

[b] Significant at 99 % probability.

Table A.4 — Precision and bias for weighbridges — applied loads versus observed loads

Test	Applied kg	Observed kg	Difference kg
1	31 890	31 890	0
2	31 890	31 870	− 20
3	31 890	31 900	+ 10
4	31 890	31 870	− 20
5	70 810	70 770	− 40
6	70 810	70 780	− 30
7	70 810	70 770	− 40
8	70 810	70 820	+ 10
9	102 700	102 710	+ 10
10	102 700	102 690	− 10
11	102 700	102 650	− 50
12	102 700	102 690	− 10
Parameter		**Symbol**	**Value**
Mean-applied load (kg)		$\bar{x}\ (A)$	68 467
Mean-observed loads (kg)		$\bar{x}\ (O)$	68 451
Mean difference (kg)		$\Delta \bar{x}$	−16
Variance of differences (kg^2)		$s^2(\Delta x)$	445
Standard deviation (kg)		$s(\Delta x)$	21,1
Coefficient of variation (%)		CV	0,03
Number of paired data		n	12
Variance of mean difference		$s^2(\Delta \bar{x})$	37
Standard deviation		$s(\Delta \bar{x})$	6,1
Student's t-value		t	2,601
Significance		—	— [a]
Bias detection limits:			
Type I risk only (kg)		BDL(I)	± 13
Type I & II risks (kg)		BDLI(I&II)	± 24
Tabulated t-values		$t_{0,90;11}$	1,796
		$t_{0,95;11}$	2,201
		$t_{0,99,11}$	3,106
		$t_{0,999;11}$	4,437

[a] Significant at 95 % probability.

Table A.5 — Linearity of weighbridges

Test	Applied kg	Observed kg	Difference kg
1	31 890	31 890	0
2	31 890	31 870	− 20
3	31 890	31 900	+ 10
4	31 890	31 870	− 20
Mean	31 890	31 882	− 8
5	70 810	70 770	− 40
6	70 810	70 780	− 30
7	70 810	70 770	− 40
8	70 810	70 820	+ 10
Mean	70 810	70 785	− 25
9	102 700	102 710	+ 10
10	102 700	102 690	− 10
11	102 700	102 650	− 50
12	102 700	102 690	− 10
Mean	102 700	102 685	− 15

Parameter	Symbol	Value
Correlation coefficient:		
all data points	r	1,000 0
Significance	—	— [a]
Slope	m	0,999 9
Significance	—	— [a]
Intercept	a	− 7,7
Significance	—	ns [b]
Tabulated r-values	$r_{0,95,10}$	0,576
	$r_{099,10}$	0,708
	$r_{0,99,10}$	0,823
Correlation coefficient:		
means only	r	1,000 0
Significance	—	— [a]
Slope	m	0,999 9
Significance	—	— [a]
Intercept	a	− 8,4
Significance	—	ns [b]
Tabulated t-values	$t_{0,95;1}$	0,997
	$t_{0,99;1}$	1,0
	$t_{0,999;1}$	1,0

[a] Significant at 99 % probability.

[b] ns = not significant.

Table A.6 — Precision and bias for hopper scales — applied loads versus observed loads

Test	Applied kg	Observed kg	Difference kg
1	2 000	2 004	+ 4
2	4 000	4 005	+ 5
3	6 000	6 009	+ 9
4	8 000	7 993	− 7
5	10 000	10 007	+ 7
6	12 000	12 008	+ 8
7	14 000	13 991	− 9
8	16 000	16 007	+ 7
9	18 000	18 005	+ 5
10	20 000	20 003	+ 3
11	22 000	21 992	− 8
12	24 000	24 009	+ 9
Parameter		**Symbol**	**Value**
Mean – applied loads (kg)		$\bar{x}\,(A)$	13 000
Mean – observed loads (kg)		$\bar{x}\,(O)$	13 003
Mean difference (kg)		$\Delta\bar{x}$	+ 3
Variance of differences (kg^2)		$s^2(\Delta x)$	46
Standard deviation (kg)		$s(\Delta x)$	6,8
Coefficient of variation (%)		CV	0,05
Numbered of paired data		n	12
Variance of mean difference		$s^2(\Delta\bar{x})$	4
Standard deviation		$s(\Delta\bar{x})$	2
Student's t-value		t	1,410
Significance		—	ns [a]
Bias detection limits:			
Type I risk only (kg)		BDL(I)	± 4
Type I & II risks (kg)		BDL(I&II)	± 8
Tabulated t-values		$t_{0,90;11}$	1,796
		$t_{0,95;11}$	2, 201
		$t_{0,99;11}$	3, 106
		$t_{0,999;11}$	4, 437

[a] ns = not significant.

Table A.7 — Linearity of hopper scales — applied loads versus observed loads

Test	Applied kg	Observed kg	Difference kg
1	2 000	2 004	+ 4
2	4 000	4 005	+ 5
3	6 000	6 009	+ 9
4	8 000	7 993	− 7
5	10 000	10 007	+ 7
6	12 000	12 008	+ 8
7	14 000	13 991	− 9
8	16 000	16 007	+ 7
9	18 000	18 005	+ 5
10	20 000	20 003	+ 3
11	22 000	21 992	− 8
12	24 000	24 009	+ 9

Parameter	Symbol	Value
Correlation coefficient	r	1,000 0
Significance	—	— [a]
Slope	m	0,999 9
Significance	—	— [a]
Intercept	a	4,1
Significance	—	ns [b]
Tabulated r-values	$r_{0,95;10}$	0,576
	$r_{0,99;10}$	0,708
	$r_{0,999;10}$	0,823

[a] Significant at 99,9 % probability.

[b] ns = not significant.

Table A.8 — Precision and bias for gantry scales — applied loads versus observed loads

Test	Initial kg	Added kg	Observed kg	Difference kg
1	0	2 000	1 994	− 6
2	2 102	2 000	4 105	+ 3
3	4 234	2 000	6 229	− 5
4	5 975	2 000	7 983	+ 8
5	8 125	2 000	10 107	− 18
6	9 996	2 000	12 004	+ 8
7	11 880	2 000	13 891	+ 11
8	13 905	2 000	15 896	− 9

Parameter	Symbol	Value
Mean – applied loads (kg)	$\bar{x}\ (A)$	9 027
Mean – observed loads (kg)	$\bar{x}\ (O)$	9 026
Mean difference (kg)	$\Delta \bar{x}$	− 1
Variance of differences (kg[2])	$s^2(\Delta x)$	102
Standard deviation (kg)	$s(\Delta x)$	10,1
Coefficient of variation (%)	CV	0,11
Numbered of paired data	n	8
Variance of mean difference	$s^2(\Delta \bar{x})$	13
Standard deviation	$s(\Delta \bar{x})$	3,6
Student's t-value	t	0,280
Significance	—	ns [a]
Bias detection limits:		
Type I risk only (kg)	BDL(I)	± 8
Type I & II risks (kg)	BDL(I&II)	± 15
Tabulated t-values	$t_{0,90;7}$	1,895
	$t_{0,95;7}$	2,365
	$t_{0,99;7}$	3,499
	$t_{0,999;7}$	5,405

[a] ns = not significant.

Table A.9 — Linearity of gantry scales — applied loads versus observed loads

Test	Initial kg	Added kg	Observed kg	Difference kg
1	0	2 000	1 994	− 6
2	2 102	2 000	4 105	+ 3
3	4 234	2 000	6 229	− 5
4	5 975	2 000	7 983	+ 8
5	8 125	2 000	10 107	− 18
6	9 996	2 000	12 004	+ 8
7	11 880	2 000	13 891	+ 11
8	13 905	2 000	15 896	− 9

Parameter	Symbol	Value
Correlation coefficient	r	1,000 0
Significance	—	— [a]
Slope	m	1,000 2
Significance	—	— [a]
Intercept	α	− 2,6
Significance	—	ns [b]
Tabulated r-values	$r_{0,95;6}$	0,707
	$r_{0,99;6}$	0,834
	$r_{0,999;6}$	0,925

[a] Significant at 99,9 % probability.

[b] ns = not significant.

Table A.10 — Precision and bias for platform scales — certified weights versus observed loads

Test No.	Certified weight kg	Observed load First kg	 Second kg
1	2 000	2 006	1 998
2	2 000	2 002	2 001
3	2 000	2 005	2 003
4	2 000	2 008	2 000
5	2 000	2 005	1 999
6	2 000	2 004	2 001
Parameter	**Symbol**	**First**	**Second**
Certified weight (kg)	$\bar{x}\ (C)$	2 000	2 000
Mean – observed load (kg)	$\bar{x}\ (O)$	2 005	2 000,3
Mean difference (kg)	$\Delta \bar{x}$	+ 5	+ 0,3
Variance (kg^2)	$s^2(x)$	4	3,07
Standard deviation (kg)	$s(x)$	2	1,75
Coefficient of variation (%)	CV	0,10	0,09
Number of paired data	n	6	6
Variance of mean	$s^2(\bar{x})$	0,667	0,511
Standard deviation	$s(\bar{x})$	0,816	0,715
Student's t-value	t	6,124	0,420
Significance	—	— [a]	ns [b]
Bias detection limits:			
Type I risk only (kg)	BDL(I)	± 2,1	± 1,8
Type I & II risks (kg)	BDL(I&II)	± 3,7	± 3,3
Probable bias ranges			
Type I risk only:			
lower limit (kg)	PBL(I)	± 2,9	na [c]
upper limit (kg)	PBU(II)	± 7,1	na
Type I & II risks:			
lower limit (kg)	PBL(I&II)	± 1,3	na
upper limit (kg)	PBU(I&II)	± 8,7	na
Tabulated t-values	$t_{0,90;5}$	2,015	
	$t_{0,95;5}$	2,571	
	$t_{0,99,5}$	4,032	
	$t_{0,999;5}$	6,859	

[a] Significant at 99 % probability.

[b] ns = not significant.

[c] na = not applicable.

Annex B
(informative)

Statistics

B.1 Terms and symbols

In this annex are introduced statistical terms and symbols, tests and techniques required to check for bias, to estimate the precision of mass measurement techniques and to verify the degree of causality between applied and observed loads. In Table B.1 statistical terms and symbols applied in different sections are listed.

B.2 Measure for central tendency

Only the arithmetic mean is required as a measure for central tendency and thus for the most probable estimate of the unknown true mass of a quantity of mineral concentrate. This statement implies that wet mass can only be estimated with a finite degree of precision as each mass measurement is an estimate, hopefully unbiased, for the unknown true wet mass.

For application in the field of mass measurement, the arithmetic mean is an effective estimate for central tendency and thus for the unknown true value of the wet mass of a quantity of concentrate. Its formula is:

$$\bar{x} = \sum x_i / n \qquad \text{(B.1)}$$

where

$\bar{x}$ is the mean for a set of n measurements;

x_i is the i^{th} measurement;

n is the number of measurements in a set.

In practice only a single estimate for the wet mass of a quantity of concentrate is obtained. System calibrations, however, generate paired sets of applied and observed loads so that the mean differences can be tested for statistical significance.

Table B.1 — List of statistical terms and symbols

Parameter	Symbol	Parameter	Symbol
Mean	$\bar{x}$	Difference	$\Delta \bar{x}$
Variance	$s^2(\bar{x})$	Mean difference	$\Delta \bar{x}$
Standard deviation	$s(\bar{x})$	Variance of differences	$s^2(\Delta x)$
Coefficient of variance	CV	Standard deviation of differences	$s(\Delta x)$
Student's t-value	t	Variance of mean difference	$s^2(\Delta \bar{x})$
Correlation coefficient	r	Standard deviation of mean difference	$s(\Delta \bar{x})$
95 % confidence interval	95 % CI	95 % confidence range	95 % CR
Bias detection limits	BDL	Type I risk only	BDL(I)
		Type I and II risks	BDL(I&II)
Probable bias range	PBR	Type I risk only: lower limit	PBL(I)
		Type I risk only: upper limit	PBU(I)
		Type I & II risks: lower limit	PBL(I&II)
		Type I & II risks: upper limit	PBU(I&II)
NOTE Derived symbols such as $s^2(M_W)$ or $s(\Sigma M_d)$ are used in various clauses of this part of ISO 12745.			

B.3 Measures for variability

The variance is the most fundamental measure for determining variability. The variance can be calculated using the following basic formula:

$$s^2(x) = \frac{\sum\left(x_\text{i} - \bar{x}\right)^2}{n - 1} \qquad \text{(B.2)}$$

where

$\bar{x}$ is the mean for a set of n measurements;

x_i is the i^th measurement in the set;

n is the number of measurements in the set;

$n - 1$ is the degrees of freedom

This formula requires that the mean be calculated before differences are squared and added which introduces a measure of uncertainty due to rounding. The next formula is equivalent to the basic formula, but it is more precise and much faster and simpler to use in computer applications.

$$s^2(x) = \frac{\sum x_\text{i}^2 - \left(\sum x_\text{i}\right)^2 / n}{n - 1} \qquad \text{(B.3)}$$

where

$\sum x_\text{i}^2$ is the sum of squared measurements;

x_i is the sum of all measurements;

n is the number of measurements in a set;

$n - 1$ is the degrees of freedom.

The following formula is used to calculate the variance of differences between identifiably different sets of paired measurements such as applied loads and observed loads:

$$s^2(\Delta x) = \frac{\sum \Delta x_\text{i}^2 - \left(\sum \Delta x_\text{i}\right)^2 / n}{n - 1} \qquad \text{(B.4)}$$

where

$s^2(\Delta x)$ is the variance of differences;

$\sum \Delta x_i^2$ is the sum of all squared differences;

$\left(\sum \Delta x_i\right)^2$ is the sum of all differences squared;

n is the number of paired measurement;

$n - 1$ is the degrees of freedom.

The variance of differences between identifiably different paired data such as calibration data for static scales and the number of pairs in a set, determines the power or sensitivity of Student's t-test to detect a bias or systematic error. Hence, the variance of differences and the number of applied and observed loads in the set are the most important statistical parameters to test for bias and estimate precision of mass measurement techniques.

The following formula generates an estimate for the variance of a single measurement with a static scale from a set of duplicate measurements:

$$s^2(x) = (\pi/4)\left[\left(\Sigma\left|x_{i_1} - x_{i_2}\right|\right)/n\right]^2 \quad \text{(B5)}$$

where

$s^2(x)$ is the variance for a single measurement;

x_{i1} is the first measurement of i th pair;

x_{i2} is the second measurement of i th pair;

n is the number of paired measurements.

Absolute differences between simultaneous duplicates are a measure of precision only. By contrast, relative differences between identifiably different paired data (certified weights and observed masses) with their signs taken into account, generate a measure of the absolute accuracy of a mass measurement technique.

Due to its squared dimension the variance is not a useful measure to check and compare variability and precision at a glance. The standard deviation, which is the square root of the variance and has the same dimension as the variable of interest, is a more readily understood parameter for precision. Derived parameters such as the coefficient of variation (CV), confidence intervals (CIs) and confidence ranges (CRs) as measures of the precision of means and bias detection limits (BDLs) and probable bias ranges (PBRs), are more readily understood measures of precision than variances.

In science and engineering 95 % confidence intervals (95 % CIs) and 95 % confidence ranges (95 % CRs) are used most frequently. The power or sensitivity of Student's t-test to detect a bias between applied and observed loads is usually reported at 95 %, 99 % and 99,9 % probability levels.

B.4 Measures for precision

The CV is a most effective measure for quantifying the precision of a mass measurement technique. CVs can be plotted on control charts for precision. Their value is numerically equal to the standard deviation as a percentage of the mean of observed loads so that the following formula applies:

$$\text{CV} = (100/\bar{x}) \times \sqrt{s^2(x)} = (100/\bar{x}) \times s(x) \quad \text{(B.6)}$$

where

$s^2(x)$ is the variance;

$s(x)$ is the standard deviation;

$\bar{x}$ is the mean of observed loads.

A CV's dimension is represented by a percentage so CVs make it simple to compare at a glance the precision parameters for difference mass measurement techniques.

CIs and CRs are useful measures for determining precision. In mining and metallurgy 95 % CIs and 95 % CRs are used most frequently but 99 % and 99,9 % probability levels are often used in statistical tests. In the case where the number of data points on which a variance estimate is based is unknown but expected to be large, a 95 % CI is the product of the mean's standard deviation and the factor 1,96, the z-value from the normal of Gaussian distribution for a symmetrical 95 % probability. This implies that 95 % CI = $s(\bar{x}) \times z$ 0,96 $\approx - s(\bar{x}) \times 2$. In the case where the number of data points is known and small, a tabulated t-value with $n - 1$ degrees of freedom is used to compute 95 % confidence intervals and ranges so that:

$$95\ \%\ \text{CI} = s(\bar{x}) \times t_{0,95;n-1} \qquad \text{(B.7)}$$

Generally, the 95 % confidence range for the most probable value is the estimated value minus its 95 % confidence interval and the estimated value plus its 95 % confidence interval thus:

$$95\ \%\ \text{CR: lower limit} = \bar{x} - 95\ \%\ \text{CI} = \bar{x} - s(\bar{x}) \times t_{0,95\text{i},n-1} \approx \bar{x} - s(\bar{x}) \times z_{0,95} \qquad \text{(B.8)}$$

$$95\ \%\ \text{CR: upper limit} = \bar{x} + 95\ \%\ \text{CI} = \bar{x} + s(\bar{x}) \times t_{0,95;n-1} \approx \bar{x} + s(\bar{x}) \times z_{0,95} \qquad \text{(B.9)}$$

where

$\bar{x}$ is the mean;

$s(\bar{x})$ is the standard deviation of the mean;

$t_{0,95\text{i},n-1}$ is the t-value for symmetrical 95 % probability;

$z_{0,95}$ is the z-value for symmetrical 95 % probability.

NOTE Care should be taken when referring to statistical tables for t-values as many compilations use a notation different to that used in this International Standard. Many statistical reference books write the t-value shown in this International Standard $t_{0,95;n-1}$, in the alternative form, $t_{0,0\ 5;n-1}$, which explicitly shows it is the 5 % probability of $|t| > t_{0,95;n-1}$.

It is also commonly found that $t_{0,95;n-1}$ is written as $t_{0,025;n-1}$ to indicate that the probability is 2,5 % that $t > t_{0,025;n-1}$ and 2,5 % that $t < t_{0,025;n-1}$ for the symmetrical t-distribution. By comparing the tabulated t-values in this International Standard with the statistical tables being consulted, it should be clear as to what convention the tables are using for the probability.

CIs and CRs are calculated from the variances for means. Confidence intervals and confidence ranges at various probability levels are also effective control and action limits for control charts for precision and bias. The relationship between the variance for a set of measurements and the variance for the mean of the set is also based on the central limit theorem.

BDLs are measures of the power or sensitivity of Student's t-test for detecting a bias. PBRs are measures for the probable range within which an observed bias is expected to fall when preselected statistical risks are taken into account. PBRs can be reported for the type I risk alone or for type I and II risks. It only makes sense to report PBRs if the mean difference between applied loads and observed loads exceeds BDLs.

B.5 Central limit theorem

The central limit theorem is one of the most important theorems in applied statistics. It plays a fundamental role in many applications such as calculating a precision estimate for the mean of a set of measurements, testing sets of paired test results for compatibility, calculating BDLs and PBRs for bias test programs and plotting BDLs in control charts.

For this application the central limit theorem is defined as follows:

> The variance for the mean difference between a set of n paired measurements is n times smaller than the variance of differences between measurements.

A robust and sensitive test to check sets of identifiably different paired data for compatibility is Student's t-test. It is used to test for bias the mean difference between certified loads and observed loads and to estimate the precision of a mass measurement technique.

B.6 Student's *t*-test

The *t*-test is applied to check for compatibility sets of identifiably different paired data (e.g. certified weights against observed loads). It gives information on the absolute accuracy of mass measurements and on the variance of differences between identifiably different paired measurements. Student's *t*-test is applied to check whether the mean difference between a set of paired data is statistically different from zero and thus is a measure for a bias, or is statistically identical to zero and thus a measure for the cumulative effect of all random variations in a mass measurement technique.

If the mean difference between a set of certified weights and observed loads is statistically different from zero (reject null hypothesis) the mass measurement system is not in a proper state of calibration. If the mean difference between applied and observed loads is statistically identical to zero (accept null hypothesis) the scale is in a proper state of calibration. Otherwise, it should be adjusted and recalibrated before its state of calibration is acceptable for commercial applications.

Student's *t*-value is the ratio between the mean difference and its standard deviation so that the following formulae apply:

$$t = \frac{\overline{x_l} - \overline{x_2}}{s(\Delta\bar{x})}$$

$$= \frac{\Delta\bar{x}}{s(\Delta\bar{x})} \qquad \text{(B.10)}$$

where

t	is the calculated *t*-value;
$\overline{x_1}$	is the mean for first data set;
$\overline{x_2}$	is the mean for second data set;
$\Delta\bar{x}$	is the mean difference;
$s(\Delta\bar{x})$	is the standard deviation of mean difference.

The formula implies that the relationship between $s(\Delta\bar{x})$, the standard deviation of the mean difference and $s(\Delta\bar{x})$, the standard deviation of differences between paired data with their signs taken into account is based on the central limit theorem so that:

$$s(\Delta\bar{x}) = \sqrt{\left[s^2(\Delta x)/n\right]}$$

$$= \frac{s(\Delta x)}{\sqrt{n}} \qquad \text{(B.11)}$$

Hence, three variables interact and yield the calculated *t*-value.

The mean difference between a set of paired data is either statistically identical zero, or a measure for a bias. A bias, in turn, is either positive (higher than a certified weight) or negative (lower than a certified weight). The variance of differences and the number of paired data in a set determine the power or sensitivity of the *t*-test to detect a bias. The variance of differences is a function of the variability between paired test results and thus of the precision of the mass measurement technique applied to obtain the set

The *t*-test was applied to a set of calibration data for a weighbridge. Table A.4 lists the complete data set and its statistical parameters and Table B.2 summarizes the statistical parameters for this paired data.

Table B.2 — Student's *t*-test for calibration data

Parameter	Symbol	Value
Mean – applied loads (kg)	$\bar{x}\,(A)$	68 467
Mean – observed loads (kg)	$\bar{x}\,(O)$	68 451
Mean difference (kg)	$\Delta\bar{x}$	− 16
Variance of differences (kg^2)	$s^2(\Delta x)$	445
Standard deviation (kg)	$s(\Delta x)$	21,1
Coefficient of variation (%)	CV	0,03
Number of paired data	n	12
Variance of mean difference	$s^2(\Delta\bar{x})$	37
Standard deviation	$s(\Delta\bar{x})$	6,1
Student's *t*-value	t	2,601
Significance	—	— [a]

[a] Significance at 95 % probability.

The calculated *t*-value of

$$\frac{16}{6,1} = 2,601$$

exceeds the tabulated value of $t_{0,5;11}$ = 2,201 at 95 % probability but is still below the tabulated value of $t_{0,99;11}$ = 3,106 at 99 % probability. The question is then whether a mean difference of − 16 kg is statistically identical to zero. The concept of BDLs for the type I risk only and for type I and II risks, makes it simple to assess whether this mean difference of − 16 kg is a measure for a bias or due to random variations only.

The number of paired data in a set determines the BDLs. Theoretically, it is possible to prove that even a small and commercially insignificant difference is a bias if the number of paired data in a set is large enough.

B.7 Bias detection limits

BDLs are measures of the power or sensitivity of Student's *t*-test to detect a bias between a set of certified weights and observed loads. BDLs are defined either for the type I risk only or for type I and II risks. The effect of the number of paired data in the set, on the sensitivity of the *t*-test becomes evident upon realizing that each BDL is the product of the standard deviation for the mean difference and either a single tabulated *t*-value for the BDL(I), or the sum of two for the BDL (I&II). Tabulated *t*-values are also a function of the number of paired data in the set.

In science and engineering a symmetrical two-sided 5 % probability for the type I risk (reject null hypothesis when the mean difference is statistically identical to zero) and a symmetrical one-sided 5 % probability for type II risks (accept null hypothesis when the mean difference is statistically different from zero) are used to quantify the power of the *t*-test to detect a bias. Based on this convention BDLs are calculated as follows:

$$\text{BDL(I)} = s(\Delta\bar{x}) \times t_{0,95;n-1} \quad \text{(B.12)}$$

$$\text{BDL(I\&II)} = s(\Delta\bar{x}) \times [t_{0,90;n-1} + t_{0,95;n-1}] \quad \text{(B.13)}$$

where

BDL(I) = BDL for the type I risk only;

BDL(I&II) = BDL for type I and II risks;

$s(\Delta\bar{x})$ is the standard deviation of mean difference;

$t_{0,90;n-1}$ is the tabulated t-value at 90 % probability;

$t_{0,95;n-1}$ is the tabulated t-value at 95 % probability.

For the set of twelve pairs of applied and observed loads that are listed in Table A.4, the BDLs are:

$$\sqrt{(445/12)} \times 2,201 = \pm 13 \text{ kg}$$

for the type I risk only

and

$$\sqrt{(445/12)} \times (1,793 + 2,201) = \pm 24 \text{ kg}$$

for the type I and II risks

However, if the data set consisted of 50 paired test data, the BDLs would be:

$$\sqrt{(445/50)} \times 2,010 = \pm 6 \text{ kg}$$

for the type I risk only

and

$$\sqrt{(445/50)} \times (1,681 + 2,010) = \pm 11 \text{ kg}$$

for type I and II risks.

BDLs for the type I risk, and for type I and II risks, are very effective control and action limits for charts in which precision and bias of a mass measurement technique are monitored as a function of time.

B.8 Probable bias ranges

If the mean difference between a set of paired test data is statistically significant, then the BDLs for the type I risk, and for type I and II risks, can be added to and deducted from this bias. The resulting ranges, which are referred to as PBRs, are obtained with the following formulae:

$$\text{PBL(I)} = \Delta\bar{x} - \text{BDL(I)} \quad \text{(B.14)}$$

$$\text{PBU(I)} = \Delta\bar{x} + \text{BDL(I)} \quad \text{(B.15)}$$

$$\text{PBL(I\&II)} = \Delta\bar{x} - \text{BDL(I\&II)} \quad \text{(B.16)}$$

$$\text{PBU(I\&II)} = \Delta\bar{x} + \text{BDL(I\&II)} \quad \text{(B.17)}$$

For example, the mean difference of – 16 kg in Table B.2 between the means of 68 467 kg for applied loads and 68 451 kg for observed loads was found to be statistically significant at the 95 % probability level and the sensitivity of this t-test in terms of BDLs was ± 13 kg for the type I risk and ± 24 kg for type I and II risks. Thus the lower and upper limits of the PBRs for the mean difference of – 16 kg are:

$$\text{PBL(I)} = -16 - 13 = -29$$

$$\text{PBU(I)} = -16 + 13 = -3$$

$$\text{PBL(I\&II)} = -16 - 24 = -40 \text{ (na)}$$

$$\text{PBU(I\&II)} = -16 + 24 = +8 \text{ (na)}$$

Because the zero mean difference still falls between lower and upper limits of the PBR(I&II) for this set of calibration data, the mean difference of – 16 kg is not a bias if type I and II risks are both taken into account, but would be a bias if only the type I risk were considered. Taking into account the probability for type I and II risks therefore results in a more robust test.

The probability for the type I risk ranges from a lower limit of – 29 kg to an upper limit of – 3 kg while probability of encountering the type I and II risks ranges from a lower limit of – 40 kg to an upper limit of + 8 kg. The zero mean difference, which implies a proper state of calibration, falls between the lower and upper limits of the PBR(I&II). Therefore, the PBL(I&II) and the PBU(I&II) are marked "na" (not applicable). The conclusion is then that the scale's state of calibration would be acceptable if type I and II risks were both taken into account, but would be unacceptable if only the type I risk were taken into account.

B.9 Correlation-regression

Correlation is a measure for the degree of association between applied loads and observed loads while regression quantifies this relationship. Correlation coefficients or r-values are the quantitative measures for the degree of causality between applied and observed loads. A correlation coefficient of $r \approx 0$ displays a complete lack of causality, and a correlation coefficient of $r \approx \pm 1$ implies a very high degree of association. Calculated correlation coefficients are compared with tabulated r-values with appropriate degrees of freedom and at different probability levels, to determine statistical significance.

Correlation coefficients between certified weights and observed loads are invariably statistically significant to the extreme (exceed 99,9 % probability level). However, correlation coefficients between certified weights and differences between certified weights and observed loads are only significant if the scale's precision is a function of load. Due to their low suspended masses only weighbridges and platform scales may display a significant correlation between the means of applied and observed loads and their differences.

Below is a formula for calculating the correlation coefficient or r-value for a paired set of applied and observed loads from sums and products of $x(A_i)s$ and $x(O_i)s$, and the sums of squares of $x(A_i)s$ and $x(O_i)s$:

$$r = \frac{\sum \times(A_i)\times(O_i) - \frac{\sum \times(A_i)\times(O_i)}{n}}{\sqrt{\left[\sum \times^2(A_i) - \frac{\left(\sum \times(A_i)\right)^2}{n}\right]\left[\sum \times^2(O_i) - \frac{\left(\sum \times(O_i)\right)^2}{n}\right]}} \tag{B.18}$$

where

r is the correlation coefficient;

$x(A_i)$ is the i^{th} applied load;

$x(O_i)$ is the i^{th} observed load;

n is the number of paired data.

Table A.5 lists three means for applied and observed loads for a weighbridge. Table B.3 summarizes the terms required to calculate the correlation coefficient for this paired data set.

The slopes and intercepts of regression lines can be tested for statistical significance. In this case the slope of 0,999 89 is statistically significant to the extreme, while the intercept of – 8,4 kg is insignificant. The difference between the calculated slope of 0,999 89 and the theoretical slope of unity (1) can also be tested for statistical significance by applying the t-test.

Table B.3 — Correlation between applied and observed loads

Test	$x(A_i)$	$x(A_i)^2$	$x(O_i)$	$x(O_i)^2$	$x(A_i) \times x(O_i)$
1-4	31 890	1 016 972 100	31 882	1 016 461 924	1 016 716 980
5-8	70 810	5 014 056 100	70 785	5 010 516 225	5 012 285 850
9-12	102 700	10 547 290 000	102 685	10 544 209 000	10 545 749 000
Sum	205 400	16 578 318 200	205 352	16 571 187 149	16 574 751 830

Numerator: $$16\,574\,751\,830 - \frac{(205\,400 \times 205\,352)}{3} = 2\,514\,985\,400$$

Denominator:

first term: $$16\,578\,318\,200 - \frac{205\,400^2}{3} = 2\,515\,264\,870$$

second term: $$16\,571\,187\,149 - \frac{205\,352^2}{3} = 2\,514\,706\,080$$

r-value: $$\frac{2\,514\,985\,400}{\sqrt{2\,515\,264\,870 \times 2\,514\,706\,080}} = 1{,}000$$

Slope: $$\frac{2\,514\,985\,400}{2\,515\,264\,870} = 0{,}999\,89$$

Intercept: $$\left(\frac{205\,352}{3}\right) - \left(0{,}999\,9 \times \frac{205\,400}{3}\right) = -\,8{,}4 \text{ kg}$$

Annex C
(informative)

Draft surveys

Table C.1 — Example of displacement calculation

Drafts and densities		**Initial** m		**Final** m
Forward draft corrected		4,82		7,37
After draft corrected		6,41		7,88
Mean forward and after draft		5,615		7,625
Port midships draft		5,6		7,66
Standard midships draft		5,62		7,62
Mean midships draft		5,61		7,64
Mean of mean drafts		5,612 5		7,632 5
Double mean of drafts [a]		5,611 25		7,636 25
Density of water (t/m^3)		1,023		1,022 5
Displacements	**Code**	**Initial** t	**Code**	**Final** t
Displacement at double mean draft	I	47 327	F	65 712
Displacement corrected for trim		46 780		65 557
Density correction		– 91		– 160
Displacement corrected for density		46 689		65 397
Ballast and consumables	**Code**	**Initial** t	**Code**	**Final** t
Ballast water	i	10 225	f	6 750
Bunker oils		865		860
Potable water		285		275
Miscellaneous supplies		—		—
Mass of ballast and consumables		11 375		7 885
Mass of cargo transferred		**Code**		**Mass** t
Final mass of vessel and cargo		(F-f)		57 512
Initial mass of vessel		(I-i)		35 314
Mass of cargo loaded		(F-f)-(I-i)		22 198
Initial mass of vessel and cargo		(I-i)		na [b]
Final mass of vessel		(F-f)		na
Mass of cargo discharged		(I-i)-(F-f)		na

[a] Mean of mean drafts and mean midships draft.

[b] na = not applicable in this example.

Table C.2 — Example of a displacement or draft table

Draft m	**Mass** t	**Draft** m	**Mass** t	**Draft** m	**Mass** t
5,18	7 498	3,66	4 760	2,13	2 094
5,21	7 451	3,68	4 714	2,16	2 053
5,23	7 405	3,71	4 669	2,18	2 010
5,26	7 358	3,73	4 623	2,21	1 969
5,28	7 311	3,76	4 579	2,24	1 927
5,31	7 265	3,78	4 533	2,26	1 885
5,33	7 218	3,81	4 488	2,29	1 842
5,36	7 171	3,84	4 442	2,31	1 800
5,38	7 124	3,86	4 396	2,34	1 758
5,41	7 077	3,89	4 351	2,36	1 716
5,44	7 030	3,91	4 306	2,39	1 674
5,46	6 984	3,94	4 261	2,41	1 632
4,88	6 937	3,35	4 215	1,83	1 590
4,9	6 888	3,38	4 171	1,85	1 549
4,93	6 848	3,4	4 127	1,88	1 508
4,95	6 803	3,43	4 083	1,91	1 467
4,98	6 759	3,45	4 040	1,93	1 426
5	6 715	3,48	3 997	1,96	1 384
5,03	6 670	3,51	3 953	1,98	1 343
5,05	6 625	3,53	3 909	2,01	1 302
5,08	6 581	3,56	3 865	2,03	1 260
5,11	6 536	3,58	3 821	2,06	1 219
5,13	6 491	3,61	3 778	2,08	1 178
5,16	6 448	3,63	3 734	2,11	1 136
4,57	6 403	3,05	3 690	1,52	1 095
4,6	6 358	3,07	3 645	1,55	1 055
4,62	6 313	3,1	3 599	1,57	1 015
4,65	6 268	3,12	3 554	1,6	975
4,67	6 223	3,15	3 508	1,63	935
4,7	6 178	3,18	3 462	1,65	895
4,72	6 134	3,2	3 416	1,68	855
4,75	6 089	3,23	3 371	1,7	815
4,78	6 044	3,25	3 325	1,73	775
4,8	5 999	3,28	3 280	1,75	735
4,83	5 954	3,3	3 234	1,78	695
4,85	5 909	3,33	3 188	1,8	655
4,27	5 865	2,74	3 142	1,22	615

Draft m	**Mass** t	**Draft** m	**Mass** t	**Draft** m	**Mass** t
4,29	5 818	2,77	3 100	1,24	575
4,32	5 771	2,79	3 057	1,27	535
4,34	5 724	2,82	3 013	1,3	495
4,37	5 677	2,84	2 970	1,32	455
4,39	5 631	2,87	2 927	1,35	415
4,42	5 584	2,9	2 884	1,37	375
4,45	5 538	2,92	2 841	1,4	335
4,47	5 491	2,95	2 798	1,42	295
4,5	5 444	2,97	2 754	1,45	254
4,52	5 398	3	2 712	1,47	214
4,55	5 351	3,02	2 669	1,5	174
3,96	5 304	2,44	2 626	0,91	134
3,99	5 258	2,46	2 582	0,94	94
4,01	5 213	2,49	2 537	0,97	53
4,04	5 167	2,51	2 493	0,99	13
4,06	5 122	2,54	2 448	1,02	
4,09	5 077	2,57	2 404		
4,11	5 032	2,59	2 361		
4,14	4 986	2,62	2 316		
4,17	4 941	2,64	2 272		
4,19	4 895	2,67	2 227		
4,22	4 851	2,69	2 183		
4,24	4 805	2,72	2 138		

Annex D
(informative)

Procedure for the testing of static scales

D.1 Scope

This annex describes a step-by-step procedure for the testing of static scales.

D.2 General information

This International Standard provides a general summary of testing principles and procedures for weighbridges (6.4), hopper scales (6.5), gantry scales (6.6) and platform scales (6.7), augmented by a list of test mass categories in Clause 6. This annex provides a procedure to undertake such tests.

Weighbridges, hopper scales and platform scales represent the most common systems encountered in the context of concentrate mass determinations for commercial consignments. Although the principles of testing are similar, details vary depending on the availability and number of certified weights and the level of automation (in the case of hopper scales having built-in reference weights).

D.3 Frequency of testing

The frequency of in-house and routine scale performance tests (in addition to mandatory certification intervals stipulated by regulatory authorities) observed in practice ranges from none to as high as three times per consignment in the case of some fully-automated weighing hopper installations. The disparity reflects the absence of explicit guidelines as well as different risk perceptions and available resources. Routine scale performance checks carried out once per commercial consignment, before the loading or discharge commences, are regarded as an optimum requirement by many operators. However, it is practically impossible to carry out such checks once per commercial consignment, because the tests cause delays in loading or discharge and are very expensive. In addition, improved stability of scales in recent times should allow reduced frequency of testing. It is therefore recommended that the frequency of testing be decided by agreement between the parties concerned, based on the risk and the reliability of the scales.

D.4 Precision test procedures

D.4.1 General

Although precision tests do not require certified test weights, it is stressed that they provide no information concerning potential bias or linearity problems.

D.4.2 Determining the precision of weighbridges by replicate tests

a) Check and, if necessary, adjust the zero setting of the scale.

b) Place a truck or rail wagon (selected at random) on the weighbridge and record the gross weight W_1.

c) Remove the truck or wagon from the weighbridge and check/adjust the zero setting again.

d) Place the same truck or rail wagon [from step b)] on the weighbridge and record the gross weight W_2.

A minimum of four duplicate determinations (four data pairs W_1, W_2) are recommended in order to calculate the scale precision, in accordance with this International Standard.

D.4.3 Determining the precision of weighbridges by replicate tests

a) Check and, if necessary, adjust the zero setting of the scale.

b) Use a test mass of about five to ten times the scale's readability or sensitivity (e.g. 25 kg for a scale sensitivity of 5 kg) to produce paired measurements with and without this test mass respectively, and record the corresponding gross weights W_1 and W_2.

A minimum of six data pairs (W_1, W_2) from a single weighing cycle are recommended in order to calculate the scale precision, in accordance with this International Standard.

D.5 Calibration (bias and linearity test) procedures

D.5.1 General

Bias and linearity tests require at least one certified reference weight of suitable mass (1 t or 2 t). In the case of hopper scales, the weight, or weights, are usually suspended from the weigh frame.

The use of a large number of smaller reference weights (for example a set of 100 individually certified test weights of 20 kg each, in accordance with the requirements of National Weights and Measures authorities) is a suitable alternative, especially where the entire test sequence has to be performed manually.

D.5.2 Calibration procedure using a single certified test weight

This procedure applies to situations where only one certified test weight (or set of small weights having an equivalent total mass) covering a small part of the scale's designated range is available. Individual calibration points at increasing initial loads are generated by adding the certified test weight W_{cert} at a given load state W_0 and comparing the expected scale reading $W_1 = W_{cert} + W_0$ to the observed value W_2. The scale deviation at a given point is thus given by $W_2 - W_1$.

A minimum of three determinations is recommended and one test each at initial loads of: zero; approximately half the scale capacity; approximately full scale capacity less the certified weight.

a) Check and, if necessary, adjust the zero setting of the scale.

b) Place the certified test weight on the scale and record the first calibration point (data pair) W_1, W_2.

c) Remove the certified test weight, add a quantity of material of approximately equal to half the scale capacity, and record the exact weight.

d) Add the certified test weight to the scale and record the second calibration point W_1, W_2.

e) Remove the certified test weight, add a further quantity of material to produce a total approximately equal to half the scale capacity, and record the exact weight.

f) Add the certified test weight to the scale and record the third point W_1, W_2.

D.5.3 Calibration procedure using a set of certified test weights

The procedure applies to situations where a set of certified test weights covering the scale's entire designated range is available. Individual calibration points over the full range are generated by the incremental addition or subtraction of individual test weights, comparing the expected (certified) weight W_1 at each stage with the corresponding scale readout W_2. The scale deviation at a given point is calculated as $W_2 - W_1$ as in D.5.2.

a) Check and, if necessary, adjust the zero setting of the scale.

b) Place the first certified test weight on the scale and record the first calibration point (data pair) W_1, W_2.

c) Add the second certified test weight to the scale and record the second calibration point W_1, W_2.

d) Repeat step c) for the third and subsequent test weight in the series.

e) Remove the individual test weights, one at a time, and record the resulting data pairs, W_1, W_2, as well as the final readout at zero load.

Bibliography

[1] MERKS, J.W. and MERKS, E.A.T. *Precision and Bias for Mass Measurement Techniques*, Matrix Consultants Limited, Vancouver, 1992

[2] COLIJN, H. *Weighing and Proportioning of Bulk Solids*, 2nd Edition, Trans Tech Publications, Clausthall, Germany, 1983

[3] MERKS, J.W. *Sampling and Weighing of Bulk Solids*, Trans Tech Publications, Clausthall, Germany, 1985

ICS 73.060.99

Price based on 47 pages