INTERNATIONAL STANDARD

ISO 12745

First edition
1996-07-01

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 12745:1996(E)

Contents

Annexes

International Organization for Standardization
Case Postale 56 • CH-1211 Genève 20 • Switzerland

Printed in Switzerland

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.

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.

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

Annexes A to D of this International Standard are for information only.

This page intentionally left blank

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 standards contain provisions which, through reference in this text, constitute provisions of this International Standard. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this International Standard are encouraged to investigate the possibility of applying the most recent editions of the standards indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO 3534-1:1993, *Statistics — Vocabulary and symbols — Part 1: Probability and general statistical terms.*

ISO 3534-2:1993, *Statistics — Vocabulary and symbols — Part 2: Statistical quality control.*

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

3 Definitions

The terminology and symbols in this International Standard are largely compatible with other ISO Standards, see ISO 3534-1, ISO 3534-2 and ISO 5725-1. Some discrepancies are due to a lack of uniformity between ISO Standards and metrology — the science of measurement — as it applies to mining and metallurgy.

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.

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 2 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 3 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 Student's t-value; symbol t: Ratio between the difference for the means for sets of applied and observed loads and the standard deviation for the mean difference.

3.4 Bias Detection Limit; symbol 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 Type I risk; symbol 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.

3.6 Type II risk; symbol 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.

3.7 Probable Bias Range; symbol 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.

3.8 precision: Generic term for the cumulative effect of random variations in a mass measurement technique.

NOTE 4 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.

3.9 variance; symbol 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 5 In textbooks on applied statistics the term "mean squared deviation from the mean" is often used in reference to the variance.

3.10 standard deviation; symbol s: Measure for random variations in a mass measurement technique, numerically equal to the square root of the variance.

3.11 relative standard deviation; symbol s_r: Measure for random variations in a mass measurement technique, numerically equal to the standard deviation divided by the observed mass.

3.12 Coefficient of Variation; symbol CV: Measure for random variations in a mass measurement technique, numerically equal to the standard deviation as a percentage of the observed mass.

3.13 Confidence Interval; symbol CI: Interval within which a predetermined percentage of the differences between all possible measurements and their mean is expected to cluster.

3.14 Confidence Range; symbol CR: Range within which a predetermined percentage of all possible measurements is expected to cluster.

NOTE 6 In science and engineering 95 % confidence intervals and ranges are most frequently used.

3.15 correlation coefficient; symbol r: Measure for the degree of association or interdependence between a set of certified weights and observed loads.

3.16 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 7 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.

---

1) E.g. 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.

3.17 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 8 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).

3.18 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 9 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.

4 General remarks

International and national handbooks on weighing devices define the 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.

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 (2) 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-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 weights 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-2 t and a quantity of material until the weighbridge is tested in increments of 5 t-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 (4) 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 (5) times up to ten (10) 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 (6) data points up to twelve (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 (6) 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-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-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 (5) times up to ten (10) 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 (6) test results up to twelve (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 (5) times up to ten (10) 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 (6) test results up to twelve (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 (5) times up to ten (10) 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 (6) replicate test results up to twelve (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 (4) 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 (2) 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 in table A.1 of annex A. Table A.1 lists a set of ten (10) 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:

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

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. In 6.2.2 evidence will be presented to show that this variance of differences of 1 410,92 t^2 is not an unbiased estimated 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.

Bias Detection Limits 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. Bias Detection Limits 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) [1]	95 % CI	± 60,1
95 % Confidence Interval (%)	95 % CI	± 1,5
95 % Confidence Range:		
lower limit (t)	95 % CRL	4 049
upper limit (t)	95 % CRU	4 169

[1] 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 (2) or more marine surveyors to measure independently a vessel's draft in the light and loaded condition a set of no less than four (4) 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 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 of annex A. Table A.2 lists a set of ten (10) 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 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 differences 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 (t^2)	$s^2(\Delta x)$	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 t^2 so that a variance of 13 243 − 705,46 ≈ 12 500 t^2 would be a better estimate for the precision of a single draft survey than the variance of 705,46 t^2. In terms of a coefficient of variation the precision for draft surveys for a single cargo space would then be

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

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 precision of wet masses measured with belt scales can be found in table A.3 of annex A. This table lists a set of twelve (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

$$\frac{0,197\,6}{0,015\,2} = 13,00$$

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 this belt scale), 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) [1]	95 % CI	± 224	± 707
95 % Confidence Interval (%)	95 % CI	± 5,4	± 1,7
95 % Confidence Range:			
lower limit (t)	95 % CRL	3 883	40 636
upper limit (t)	95 % CRU	4 331	42 050

1) Based on $z_{0,95} \times s(M_w)$, or $z_{0,95} \times s\left(\sum M_w\right)$.

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