Accuracy and precision

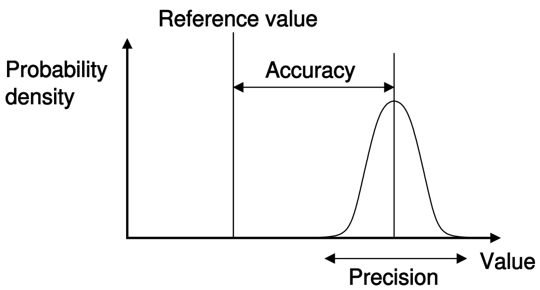
"Accuracy" redirects here. For the song by The Cure, see Three Imaginary Boys.

Precision is a description of a level of measurement that yields *consistent* results when repeated. It is associated with the concept of "*random error*", a form of observational error that leads to measurable values being inconsistent when repeated.

Accuracy has two definitions:

- The more common definition is that accuracy is a level of measurement with no inherent limitation (ie. free of *systemic error*, another form of observational error).
- The ISO definition is that accuracy is a level of measurement that yields true (no systemic errors) *and* consistent (no random errors) results.

1 Common definition



Accuracy is the proximity of measurement results to the true value; precision, the repeatability, or reproducibility of the measurement

In the fields of science, engineering and statistics, the accuracy of a measurement system is the degree of closeness of measurements of a quantity to that quantity's true value.[1] The precision of a measurement system, related to reproducibility and repeatability, is the degree to which repeated measurements under unchanged conditions show the same results.[1][2] Although the two words precision and accuracy can be synonymous in colloquial use, they are deliberately contrasted in the context of the scientific method.

A measurement system can be accurate but not precise, precise but not accurate, neither, or both. For example, if an experiment contains a systematic error, then increasing the sample size generally increases precision but does not improve accuracy. The result would be a consistent yet inaccurate string of results from the flawed experiment. Eliminating the systematic error improves accuracy but does not change precision.

A measurement system is considered *valid* if it is both *accurate* and *precise*. Related terms include *bias* (non-random or directed effects caused by a factor or factors unrelated to the independent variable) and *error* (random variability).

The terminology is also applied to indirect measurements—that is, values obtained by a computational procedure from observed data.

In addition to accuracy and precision, measurements may also have a measurement resolution, which is the smallest change in the underlying physical quantity that produces a response in the measurement.

In numerical analysis, accuracy is also the nearness of a calculation to the true value; while precision is the resolution of the representation, typically defined by the number of decimal or binary digits.

Statistical literature prefers to use the terms *bias* and *variability* instead of *accuracy* and *precision. Bias* is the amount of *inaccuracy* and *variability* is the amount of *imprecision.*

1.1 Quantification

See also: False precision

In industrial instrumentation, accuracy is the measurement tolerance, or transmission of the instrument and defines the limits of the errors made when the instrument is used in normal operating conditions.[3]

Ideally a measurement device is both accurate and precise, with measurements all close to and tightly clustered around the true value. The accuracy and precision of a measurement process is usually established by repeatedly measuring some traceable reference standard. Such standards are defined in the International System of Units (abbreviated SI from French: Système international d'unités) and maintained by national standards organizations such as the National Institute of Standards and Technology in the United States.

This also applies when measurements are repeated and averaged. In that case, the term standard error is properly applied: the precision of the average is equal to the known standard deviation of the process divided by the square root of the number of measurements averaged. Further, the central limit theorem shows that the probability distribution of the averaged measurements will be closer to a normal distribution than that of individual measurements.

With regard to accuracy we can distinguish:

- the difference between the mean of the measurements and the reference value, the bias. Establishing and correcting for bias is necessary for calibration.
- the combined effect of that and precision.

A common convention in science and engineering is to express accuracy and/or precision implicitly by means of significant figures. Here, when not explicitly stated, the margin of error is understood to be one-half the value of the last significant place. For instance, a recording of 843.6 m, or 843.0 m, or 800.0 m would imply a margin of 0.05 m (the last significant place is the tenths place), while a recording of 8,436 m would imply a margin of error of 0.5 m (the last significant digits are the units).

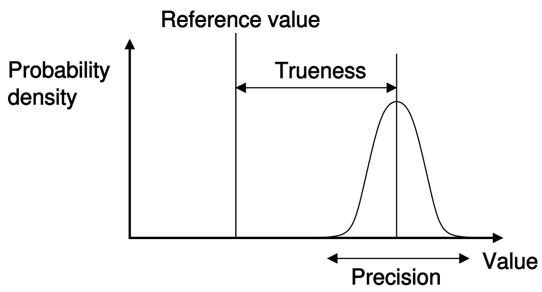
A reading of 8,000 m, with trailing zeroes and no decimal point, is ambiguous; the trailing zeroes may or may not be intended as significant figures. To avoid this ambiguity, the number could be represented in scientific notation: 8.0×10^3 m indicates that the first zero is significant (hence a margin of 50 m) while 8.000×10^3 m indicates that all three zeroes are significant, giving a margin of 0.5 m. Similarly, it is possible to use a multiple of the basic measurement unit: 8.0 km is equivalent to 8.0×10^3 m. In fact, it indicates a margin of 0.05 km (50 m). However, reliance on this convention can lead to false precision errors when accepting data from sources that do not obey it.

Precision is sometimes stratified into:

- *Repeatability* — the variation arising when all efforts are made to keep conditions constant by using the same instrument and operator, and repeating during a short time period; and
- *Reproducibility* — the variation arising using the same measurement process among different instruments and operators, and over longer time periods.

2 ISO Definition (ISO 5725)

A shift in the meaning of these terms appeared with the publication of the ISO 5725 series of standards in 1994, which is also reflected in the 2008 issue of the "BIPM International Vocabulary of Metrology" (VIM), items 2.13 and 2.14.[1]



According to ISO 5725-1, Accuracy consists of Trueness (proximity of measurement results to the true value) and Precision (repeatability or reproducibility of the measurement)

According to ISO 5725-1,[4] the general term "**accuracy**" is used to describe the closeness of a measurement to the true value. When the term is applied to sets of measurements of the same *measurand*, it involves a component of random error and a component of systematic error. In this case **trueness** is the closeness of the mean of a set of measurement results to the actual (true) value and **precision** is the closeness of agreement among a set of results.

ISO 5725-1 and VIM also avoid the use of the term "bias", previously specified in BS 5497-1,[5] because it has different connotations outside the fields of science and engineering, as in medicine and law.

3 In binary classification

Main article: Evaluation of binary classifiers

Accuracy is also used as a statistical measure of how well a binary classification test correctly identifies or excludes a condition.

That is, the accuracy is the proportion of true results (both true positives and true negatives) among the total number of cases examined.[6] To make the context clear by the semantics, it is often referred to as the "Rand accuracy" or "Rand index".[7][8][9] It is a parameter of the test.

$$\text{accuracy} = \frac{\text{positives true of number} + \text{negatives true of numbe}}{\text{positives true of number} + \text{positives false} + \text{negatives false} + \text{r}}$$

An accuracy of 100% means that the measured values are exactly the same as the given values.

On the other hand, precision or positive predictive value is defined as the proportion of the true positives against all the positive results (both true positives and false positives)

$$\text{precision} = \frac{\text{positives true of number}}{\text{positives true of number} + \text{positives false}}$$

Also see Sensitivity and specificity.

Accuracy may be determined from sensitivity and specificity, provided prevalence is known, using the equation:

$$\text{accuracy} = (\text{sensitivity})(\text{prevalence}) + (\text{specificity})(1 - \text{prevalence})$$

The accuracy paradox for predictive analytics states that predictive models with a given level of accuracy may have greater predictive power than models with higher accuracy. It may be better to avoid the accuracy metric in favor of other metrics such as precision and recall. In situations where the minority class is more important, F-measure may be more appropriate, especially in situations with very skewed class imbalance.

Another useful performance measure is the *balanced accuracy* which avoids inflated performance estimates on imbalanced datasets. It is defined as the arithmetic mean of sensitivity and specificity, or the average accuracy obtained on either class:

$$\text{accuracy balanced} = \frac{\text{sensitivity} + \text{specificity}}{2}$$

$$= \frac{0.5 * \text{positives true}}{\text{positives true} + \text{negatives false}} + \frac{0.5 * \text{negatives true}}{\text{negatives true} + \text{positives false}}$$

If the classifier performs equally well on either class, this term reduces to the conventional accuracy (i.e., the number of correct predictions divided by the total number of predictions). In contrast, if the conventional accuracy is above chance *only* because the classifier takes advantage of an imbalanced test set, then the balanced accuracy, as appropriate, will drop to chance.[10] A closely related chance corrected measure is:

$$\text{Informedness} = \text{sensitivity} + \text{specificity} - 1 = 2 * \text{accuracy balanced} - 1$$ [11]

A direct approach to debiasing and renormalizing Accuracy is Cohen's kappa, whilst Informedness has been shown to be a Kappa-family debiased renormalization of Recall.[12] Informedness and Kappa have the advantage that chance level is defined to be 0, and they have the form of a probability. Informedness has the stronger property that it is the probability that an informed decision is made (rather than a guess), when positive. When negative this is still true for the absolutely value of Informedness, but the information has been used to force an incorrect response.[11]

4 In psychometrics and psychophysics

In psychometrics and psychophysics, the term *accuracy* is interchangeably used with validity and *constant error*. *Precision* is a synonym for reliability and *variable error*. The validity of a measurement instrument or psychological test is established through experiment or correlation with behavior. Reliability is established with a variety of statistical techniques, classically through an internal consistency test like Cronbach's alpha to ensure sets of related questions have related responses, and then comparison of those related question between reference and target population.

5 In logic simulation

In logic simulation, a common mistake in evaluation of accurate models is to compare a logic simulation model to a transistor circuit simulation model. This is a comparison of differences in precision, not accuracy. Precision is measured with respect to detail and accuracy is measured with respect to reality.[13][14]

6 In information systems

The concepts of accuracy and precision have also been studied in the context of databases, information systems and their sociotechnical context. The necessary extension of these two concepts on the basis of theory of science suggests that they (as well as data quality and information quality) should be centered on accuracy defined as the closeness to the true value seen as the degree of agreement of readings or of calculated values of one same conceived entity, measured or calculated by different methods, in the context of maximum possible disagreement.[15]

Further information: Precision and recall

7 See also

- Accepted and experimental value
- Engineering tolerance
- Exactness
- Experimental uncertainty analysis
- F-score
- Measurement uncertainty
- Precision (statistics)
- Probability
- Random and systematic errors
- Sensitivity and specificity
- Significant figures
- Statistical significance

8 References

[1] *JCGM 200:2008 International vocabulary of metrology — Basic and general concepts and associated terms (VIM)*

[2] Taylor, John Robert (1999). *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books. pp. 128–129. ISBN 0-935702-75-X.

[3] Creus, Antonio. *Instrumentación Industrial*

[4] BS ISO 5725-1: "Accuracy (trueness and precision) of measurement methods and results - Part 1: General principles and definitions.", p.1 (1994)

[5] BS 5497-1: "Precision of test methods. Guide for the determination of repeatability and reproducibility for a standard test method." (1979)

[6] Metz, CE (October 1978). "Basic principles of ROC analysis" (PDF). *Semin Nucl Med.* **8** (4): 283–98.

[7] http://www.alta.asn.au/events/altss_w2003_proc/altss/courses/powers/ALTSS2003-Val+Eval-L3.pdf

[8] http://arxiv.org/pdf/1503.06410.pdf

[9] http://www.anthology.aclweb.org/E/E12/E12-1035.pdf

[10] Brodersen, K.H.; Ong, C.S.; Stephan, K.E.; Buhmann, J.M. (2010). "The balanced accuracy and its posterior distribution" (PDF). *Proceedings of the 20th International Conference on Pattern Recognition*: 3121–24.

[11] Powers, David M W (2011). "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation" (PDF). *Journal of Machine Learning Technologies* **2** (1): 37–63.

[12] Powers, David M. W. (2012). *The Problem with Kappa*. *Conference of the European Chapter of the Association for Computational Linguistics (EACL2012) Joint ROBUS-UNSUP Workshop*.

[13] Acken, John M. (1997). *Encyclopedia of Computer Science and Technology* **36**: 281–306.

[14] Glasser, Mark; Mathews, Rob; Acken, John M. (June 1990). "1990 Workshop on Logic-Level Modelling for ASICS". *SIGDA Newsletter* **20** (1).

[15] Ivanov, Kristo (1972). "Quality-control of information: On the concept of accuracy of information in data banks and in management information systems".

9 External links

- BIPM - Guides in metrology, *Guide to the Expression of Uncertainty in Measurement (GUM) and International Vocabulary of Metrology (VIM)*
- "Beyond NIST Traceability: What really creates accuracy", *Controlled Environments* magazine
- Precision and Accuracy with Three Psychophysical Methods
- Appendix D.1: Terminology, *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*
- Accuracy and Precision
- Accuracy vs Precision — a brief video by Matt Parker
- What's the difference between accuracy and precision? by Matt Anticole at TED-Ed

10 Text and image sources, contributors, and licenses

10.1 Text

- **Accuracy and precision** *Source:* https://en.wikipedia.org/wiki/Accuracy_and_precision?oldid=687082213 *Contributors:* XJaM, SJK, Patrick, Michael Hardy, Dcljr, Rethunk, Ahoerstemeier, Den fjättrade ankan~enwiki, BigFatBuddha, Reddi, Ike9898, Riddley, Robbot, Altenmann, Rholton, Cutler, Adam78, Ancheta Wis, Giftlite, Barbara Shack, BenFrantzDale, Tom harrison, Macrakis, Bobblewik, Beland, Neutrality, Urhixidur, Tsemii, Danarmak, ELApro, Julian Mehnle~enwiki, AliveFreeHappy, Discospinster, Vsmith, Smyth, Bender235, Cyclopia, Calair, CDN99, BrokenSegue, Davidruben, Zetawoof, Nicolasbock, Musiphil, Alansohn, Diego Moya, Wtmitchell, Velella, SidP, Stephan Leeds, Oleg Alexandrov, Zntrip, Simetrical, Linas, Mindmatrix, Uncle G, Armando, Benbest, Jeff3000, Ryan Reich, BD2412, NebY, Rui Silva, FlaBot, TheMidnighters, Glenn L, Chobot, Wavelength, Bovineone, Wimt, Sanguinity, Charlie.zender, NawlinWiki, Welsh, Moe Epsilon, Garrepi, Zwobot, Zvika, Capitalist, A bit iffy, SmackBot, InverseHypercube, Unyoyega, Brick Thrower, RedSpruce, Bobzchemist, Boris Barowski, Gilliam, Hraefen, Slo-mo, Saros136, Can't sleep, clown will eat me, RedHillian, Memming, Wen D House, Cybercobra, DMacks, Lambiam, Harryboyles, Archimerged, Mgiganteus1, 16@r, TastyPoutine, RichardF, Peter Horn, Joseph Solis in Australia, Twas Now, GDallimore, AbsolutDan, Jamamala, Jackzhp, BoH, Pgr94, Tex, Gregbard, Animator18, Father Chaos, Diogobruno, Epbr123, Headbomb, DarkRyder, James086, Wikidenizen, Rosalien, KrakatoaKatie, AntiVandalBot, Jayron32, Altamel, Ioeth, Leuko, Herr blaschke, Xact, Magioladitis, Wdflake, STBot, El Krem, Rettetast, Pekaje, Agogin, El0i, Galahaad, Ooga Booga, J.delanoy, Rlsheehan, Nigholith, Mlonguin~enwiki, Ian.thomson, Iida-yosiaki, AKA MBG, Afluegel, Mikael Häggström, Cmelsheimer, Iulus Ascanius, Taifunbrowser, Jamesontai, Qqzzccdd, Jacob J. Walker, Wikieditor06, Jeskec, ABF, Kevinkor2, Qxz, Stafusa, Corvus cornix, Wingedsubmariner, Ostrouchov, Jamelan, Softtest123, Dirkbb, IndulgentReader, Schoolproject2, Dharnan, Nbanic, Dogbyter, Flyer22, Paolo.dL, JSpung, Iain99, Khvalamde, Dodger67, Mygerardromance, Melcombe, Hoplon, ClueBot, Binksternet, Winsteps, MATThematical, Gregcaletta, Mild Bill Hiccup, Kenbrey, Sv1xv, XenonX3, Excirial, PixelBot, Leonard^Bloom, Brews ohare, Versus22, Qwfp, Johnuniq, Fhsjrm, Dthomsen8, WikHead, Dododerek, Addbot, Fgnievinski, Jncraton, Bob K31416, RW Dutton, Tide rolls, Bartledan, LuK3, Margin1522, Ben Ben, Yobot, Panek, Fr6, Mceike, AnomieBOT, Kangohenry, Jim1138, Beheaderchaseman, Faridkandil, Eumolpo, Capricorn42, GrouchoBot, Fernzy3, WissensDürster, Rb88guy, Erik9, FrescoBot, Whoosit, Oashi, Pinethicket, A8UDI, Serols, SpaceFlight89, Pianoplonkers, Trappist the monk, Elizabeth Gordon, Lotje, Catalin Bogdan, Jzana, Rentzepopoulos, Specs112, Johnmainstreamamerican, DeanBrettle, Mean as custard, RjwilmsiBot, 2clap, Dmwpowers, Becritical, Usna71, DASHBot, EmausBot, Racerx11, JustinTime55, Tommy2010, Wikipelli, Mz7, HiW-Bot, ZéroBot, Grunny, Lagaffe, Boulbass, Donner60, ClueBot NG, Jack Greenmaven, Lochlan1, Selvasivagurunathan m, Delusion23, Widr, MerlIwBot, Electriccatfish2, SchroCat, GuySh, BG19bot, BendelacBOT, Hallows AG, AwamerT, Mark Arsten, Toccata quarta, Lanceculnane, Ankur3020, Whym, KnowItAll'sKnowItAll, Khazar2, Deltahedron, Theethamizh26, Lugia2453, Aerowikicycle, Eyesnore, Agyeyaankur, 2000grad, Stamptrader, 32RB17, Hdjvdhb, Navstar55, Manteater, Bigdaddyw, Parthkavi24696, Boruch Baum and Anonymous: 325

10.2 Images

- **File:Accuracy_(trueness_and_precision).svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/9/92/Accuracy_%28trueness_and_precision%29.svg *License:* CC BY-SA 3.0 *Contributors:* Own work *Original artist:* SV1XV
- **File:Accuracy_and_precision.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/3/38/Accuracy_and_precision.svg *License:* GFDL *Contributors:* Transferred from en.wikipedia to Commons. *Original artist:* Pekaje at English Wikipedia
- **File:Edit-clear.svg** *Source:* https://upload.wikimedia.org/wikipedia/en/f/f2/Edit-clear.svg *License:* Public domain *Contributors:* The *Tango! Desktop Project*. *Original artist:*

  The people from the Tango! project. And according to the meta-data in the file, specifically: "Andreas Nilsson, and Jakub Steiner (although minimally)."
- **File:High_accuracy_Low_precision.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/1/10/High_accuracy_Low_precision.svg *License:* Public domain *Contributors:* DarkEvil, based on Image:Accuracy and precision-highaccuracylowprecision.gif which is free. *Original artist:* DarkEvil
- **File:High_precision_Low_accuracy.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/3/3a/High_precision_Low_accuracy.svg *License:* Public domain *Contributors:* DarkEvil, based on Image:Accuracy and precision-highprecisionlowaccuracy.gif which is free. *Original artist:* DarkEvil
- **File:Merge-arrow.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/a/aa/Merge-arrow.svg *License:* Public domain *Contributors:* ? *Original artist:* ?
- **File:Wiktionary-logo-en.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/f/f8/Wiktionary-logo-en.svg *License:* Public domain *Contributors:* Vector version of Image:Wiktionary-logo-en.png. *Original artist:* Vectorized by Fvasconcellos (talk · contribs), based on original logo tossed together by Brion Vibber

10.3 Content license

-