THIRD PART

DEFINITIONS AND NOTATIONS

NECESSITY OF STRICT AND OBJECTIVE DEFINITIONS

Most authors who have written on measurement and related subjects have underlined this necessity.

- **Hendrik Colijn (weighing)** : « It is unfortunate that so much confusion and misuse of terms (precision, bias, accuracy) exists »
- **Mandel mentions** « two contradictory definitions of the concept of accuracy » …

MATERIAL BATCHES

- **LOT** : Batch of finite, well defined size.
- **MATERIALS** : We will distinguish :
  - Compact materials : e.g. mineral deposits, subject of « Geostatistics » a science founded by the French Georges Matheron.
  - Particulate materials : Matter is discrete by essence, its elements can be molecules, atoms, ions, fragments.

Any **LOT** of particulate material can be regarded as a **SET of UNITS**. We will study …

- **Two kinds of SETS**, namely …
  - **Populations of non-ordered units,**
  - **Series of ordered units**.
- **Two kinds of UNITS** according as they are made of …
  - A **single** constituent (fragment, etc.),
  - A **group** of adjoining constituents which may be correlated with one another.

CONSTITUENTS OF A LOT OF PARTICULATE MATERIAL

- **CONSTITUENTS** (or constitutive elements) **smallest physical elements** that we assume to **remain unaltered during the sampling** operations. Those constituents are …
  - **Particulate solids** : fragments,
  - **Liquids and gases** : molecules and ions

- **COMPONENTS : chemical elements** that are usually assayed.

INCREMENT, SAMPLE, SPECIMEN

- **INCREMENT** : Quantity of material taken in a single move of the sampling device : e.g. a scoop or a cutter. An **increment** is a group of adjoining constituents, **a sub-set of L** .
- **SAMPLE** : A sample is usually obtained by gathering a certain number of increments. When certain conditions, that will be specified, are fulfilled it is a « **reliable sample** ».
- **SPECIMEN** : A specimen is obtained when these conditions are not fulfilled.

- **SAMPLES** are RELIABLE while…
- **SPECIMENS** are UNRELIABLE .
- **SAMPLING** is a **mass reduction** which is achieved by **SELECTION** of several constituents. **Sampling proper** is an operation or sequence of operations that are supposed to provide a reliable **SAMPLE.** The word sampling is often misused for …
- **SPECIMEN-TAKING** is an operation or sequence of operations that provide an unreliable **SPECIMEN**.

MAJOR QUALITY REQUIRED OF A SELECTION

Selection must respect the composition of the material to be sampled (proportion of various components) **« as best as possible ».** Sampling errors must be controlled. Obviously

A THEORY OF SAMPLING IS FIRST OF ALL A THEORY OF THE SAMPLING ERRORS

Qualitative and Quantitative approaches.

DEFINITIONS and NOTATIONS

- **GRADE a_X** of object **X** : it is defined as the actual, always unknown, proportion of physical component **A** in object **X** specified by the subscript of a_X . A grade is always a **dimensionless** quantity.

$$a_X = \frac{\text{Mass of component A in object X}}{\text{Mass total of object X}}$$

- **LOT L** : Object whose proportion a_L of component **A** is to be estimated. Grade a_L is the **unknown** of our problem. **Mass reduction of L must be progressive**. We successively obtain …
- **PRIMARY SAMPLE S_1** :
  Its actual unknown grade is a_{S1}
- **SECONDARY SAMPLE S_2** :
  Its actual unknown grade is a_{S2}
- **RESULT a_R** of the analysis of S_2 :
  - $a_R \equiv$ Estimation of grade a_{S2}
  - $a_R \equiv$ Estimator of the unknown a_L

SPECIFICITIES OF PRIMARY AND SECONDARY SAMPLING

- **PRIMARY SAMPLING** : Its purpose is to generate « laboratory samples » small enough to be transported to the lab. It usually takes place in « **dirty** » industrial surroundings **and is performed by unqualified personnel**.

- **SECONDARY SAMPLING** : Its purpose is to generate one / several « assay portions ». Takes place in « **clean** » laboratory conditions **Not always performed by qualified per-sonnel, due to a lack of education**.

DEFINITION of ERRORS

All errors are relative (easier to compare) and expressed as a proportion of a_L.

- **PRIMARY SAMPLING ERROR TSE_1**
  ≡ Total Sampling Error (stage 1)
- **SECONDARY SAMPLING ERROR TSE_2**
  ≡ Total Sampling Error (stage 2)

$$TSE_1 = \frac{a_{S1} - a_L}{a_L}$$

$$TSE_2 = \frac{a_{S2} - a_{S1}}{a_L}$$

- **TOTAL SAMPLING ERROR** (stage i) TSE_i with $i = 1, 2, \ldots n$ ($S_n \equiv$ **Assay Portion**)
- **TOTAL ANALYTICAL ERROR TAE**

$$TSE_i = \frac{a_{Si} - a_{S(i-1)}}{a_L}$$

$$TAE = \frac{a_R - a_{Sn}}{a_L}$$

- **GLOBAL ESTIMATION ERROR GEE**

$$GEE = \frac{a_R - a_L}{a_L} = \sum_i TSE_i + TAE$$

ADDITIVITY OF ERRORS

Quality control is the sequence of three **GROUPS** of error-generating operations.

The **first two** are subdivided into sub-stages

- **PRIMARY SAMPLING ERRORS TSE_1**
- **SECONDARY SAMPLING ERRORS TSE_2**
- **TOTAL ANALYTICAL ERROR TAE**

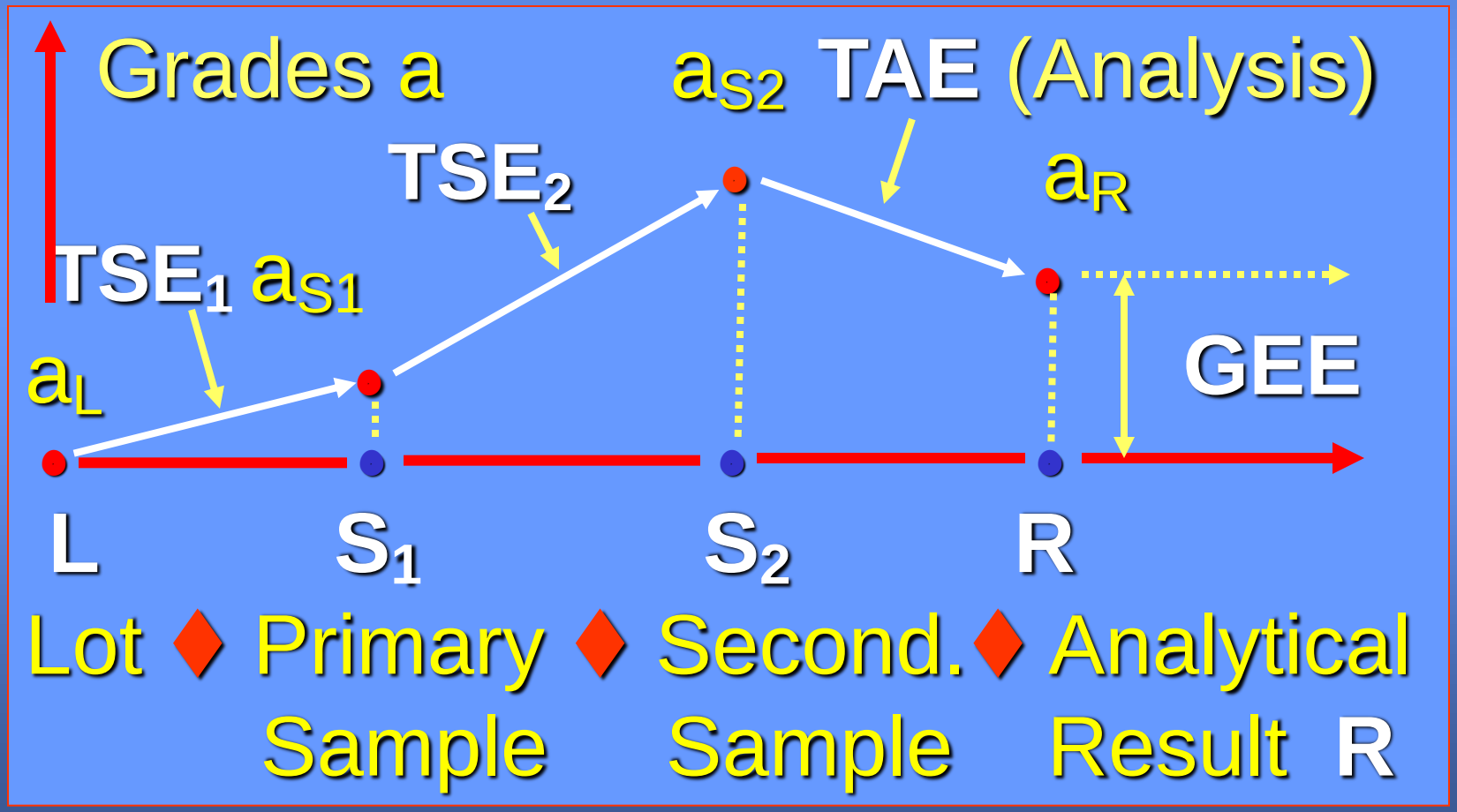
These errors are additive …

- **GLOBAL ESTIMATION ERROR GEE.**

$$GEE \equiv TSE_1 + TSE_2 + TAE$$

ADDITIVITY OF ERRORS

Usually : $TSE_1 >> TSE_2 >> TAE$



INDEPENDENCE IN PROBABILITY

- **TSE_1, TSE_2 and TAE ≡ random variables**
- **Random variables** can be characterized by their **distribution law** (often normal) and their **moments** : mean, variance, mean-squ.
- **TSE_1, TSE_2 and TAE ≡ independent in probability.** Entails that their **MEANS and VARIANCES are ADDITIVE** .

GLOBAL ESTIMATION ERROR GEE

$$GEE = \Sigma_i\, TSE_i + TAE$$

GEE is the sum of several random variables.

These variables are independent in probability with the following consequences:

- **Additivity** of biases m() and variances σ^2()

$$m(GEE) = \Sigma_i\, m(TSE_i) + m(TAE)$$
$$\sigma^2(GEE) = \Sigma_i\, [\sigma^2(TSE_i)] + \sigma^2(TAE)$$

- **USUALLY …**

$$m(TSE_1) >> m(TSE_2) >> m(TAE)$$

RECAPITULATION

- **Errors, biases, variances of sampling and analysis, add to one another**. Sampling should be treated with the same considera-tion as analysis. **In the year 2000 it is not !**
- **A THEORY OF SAMPLING DOES EXIST**.
  It has never been contested : it is simply ignored by those who should know it. Should be taught to and understood by everyone, especially **University Professors**, Analysts and Technicians in charge of Quality Control

- **FOR THE SAKE OF EFFICIENCY**, analysts should be in charge of **QUALITY CONTROL** and more specifically of **SAMPLING**.
- **DECISION MAKERS IN RESEARCH AND INDUSTRY** should understand that sufficient investments should be attributed to sampling.
- **E. ROTH (1997)** : **« Analysts should refuse to give results when they are not satisfied of the REPRESENTATIVITY of samples »** (scientific definition given below)

HOMOGENEITY AND HETEROGENEITY OF A SET OF UNITS

A **SET** of **UNITS** is said to be :

- **HOMOGENEOUS :** when it is made of **STRICTLY IDENTICAL** units .
- **HETEROGENEOUS :** when units are **NOT STRICTLY IDENTICAL**.

Homogeneity can be defined but can never be observed. Example of pure water made of H_20 molecules, H^+, OH^-, O^{2-} ions and all isotopic combinations of H & O. To say nothing of alien components.

By definition of homogeneity, the sampling of a **HOMOGENEOUS** set by **SELECTION OF WHOLE UNITS** would be a **STRICTLY EXACT OPERATION. THEREFORE …**

HETEROGENEITY
IS THE SOLE AND UNIQUE
SAMPLING ERROR GENERATOR

THEREFORE a sampling theory is naturally derived from a qualitative and quantitative theory of heterogeneity.

THE TWO MAIN FORMS OF HETEROGENEITY

Matter, any matter, is heterogeneous.

Two main forms of heterogeneity must be distinguished and defined …

- **CONSTITUTIONAL HETEROGENEITY**

The **UNIT** is a **single constitutive element.**

- **DISTRIBUTIONAL HETEROGENEITY**

The **UNIT** is a **group of adjoining elements**

These definitions will be useful in our quantitative approach.

How to qualify the

CONDITIONS

of a Selection or a Sampling

CONDITIONS and RESULTS of a SELECTION

Necessity of **STRICT** definitions in metrology. Especially re. the concepts of accuracy, precision, reproducibility and re-presentativity (no scientific definitions of the latter so far neither in English nor in French).

This necessity is underlined by numerous specialists such as Colijn (weighing) or Mandel (analysis of experimental data) …

PROBABILIST / NON-PROBABILIST

Sample / specimen **S** is a sub-set of set **L** which it is meant to represent. **S** is the result of a selection. This selection is said to be …

- **NON-PROBABILIST** (IC) : when a certain number of constituents of the set **L** have a **zero probability** of being selected.
- **PROBABILIST** (IC) : when all constituents of the set **L** have a **non-zero probability** of being selected. A probabilist selection can be correct or incorrect …

CORRECT / INCORRECT

- **CORRECT** : when simultaneously :
  - **All constituents** have an **equal probability** of being selected.
  - **The integrity of increments and sample** is duly respected.
- **INCORRECT** : when one of these conditions, at least, is not respected.

According to theoretical results, a **CORRECT** selection **ALONE** provides reliable « **SAMPLES** ».

HOW TO ACHIEVE A « CORRECT » SELECTION

Correctness is the quality of a selection that gives all constituents of **L** an equal probability of being selected. There are two ways of achieving a correct selection :

- **THE CONSTITUENTS are distributed AT RANDOM** throughout the domain of lot **L**.
- **THE INCREMENTS are distributed AT RANDOM** throughout the domain of lot **L**.

CONSTITUENTS DISTRIBUTED AT RANDOM THROUGHOUT LOT L

Definition of distributional homogeneity :

IN THIS CASE ANY FRACTION OF LOT L IS A CORRECT SAMPLE OF L

but due to gravity segregation, electrostatic, magnetic forces, or the like, such a distribution never exists **spontaneously**. It would be reached through a perfect homogenizing by mixing. **The hypothesis of spontaneous homogeneity is unrealistic.**

INCREMENTS DISTRIBUTED AT RANDOM THROUGHOUT LOT L

IN THIS CASE, THE SELECTION IS CORRECT WHATEVER THE CONSTITUENT DISTRIBUTION THROUGHOUT LOT L

- **FREQUENT MISTAKE** : the operator implicitly assumes a random distribution and takes increments from the most accessible fraction of the lot, which amounts to solving the problem by choosing to ignore its existence.

CORRECTNESS IN PRACTICE

- **DEVIATIONS FROM CORRECTNESS** may be the consequence …
  - either of **ignorance** of the personnel in charge of sampling and quality control,
  - or of **awkwardness** of the sampling operator, partly due to his own ignorance
  - or of a **deliberate will to bias the sample**.

    Sampling, weak point of the estimation chain, plays the role of « scapegoat ».

• HOW TO ACHIEVE A CORRECT SAMPLING

- **MANAGEABLE LOTS** (light enough to be handled in totality). The correct solution consists in homogenizing by mixing whenever possible. As an additional precaution, increments may be taken at random.
- **NON-MANAGEABLE LOTS**. Except with one-dimensional flowing lots, there is no correct solution.

DEFINITIONS

How to qualify the

RESULTS

of a Selection or a Sampling

STRUCTURAL AND CIRCUMSTANTIAL PROPERTIES

A property of a selection is said to be :

- **STRUCTURAL** : when it depends only on the properties of this selection,

- **CIRCUMSTANTIAL** : when, in addition to those, it does depend on external circumstances and especially on the composition of the material being sampled.

A selection can be qualified in terms of properties of the **total sampling error TSE**

At any sampling stage the error **TSE** is

$$TSE = \frac{a_S - a_L}{a_L}$$

TSE is a random variable characterized by its distribution law and its moments :

Distribution : $a_L > 10^{-6}$: normal law

$a_L < 10^{-6}$: log-normal law

COMPONENTS OF THE ERROR TSE

TSE is the sum of **TWO** major components

TSE ≡ CSE + ISE

- **CORRECT SAMPLING ERROR CSE.**
  Observed when probability **P** is **uniform.**

CSE ≡ incompressible minimum of **TSE.**

- **INCORRECT SAMPLING ERROR ISE .**
  Additional error observed when the selection probability **P** is **NOT uniform**.

CSE and **ISE** = Σ of several components.

The components **CSE** and **ISE** of **TSE** are independent in probability. This entails that their means and variances are additive.

$$m(TSE) \equiv m(CSE) + m(ISE)$$

$$\sigma^2(TSE) \equiv \sigma^2(CSE) + \sigma^2(ISE)$$

Unlike variances, biases do not decrease when the number of data increases.

In Chemometrics, for instance, biases are a major nuisance ! They must be eliminated whenever possible.

HOW TO MINIMIZE THE BIAS m(TSE)

m(TSE) ≡ m(CSE) + m(ISE)

- **m(CSE)** is **structural**. It is not caused by our awkwardness. We can neither nullify nor reduce it (see quantitative approach).
- **m(ISE)**, unlike **m(CSE)**, is **conjunctural**. It can be nullified **when the selection is correct**, which we know how to achieve.

The only way to minimize **m(TSE)** is to carry out a **CORRECT** selection / sampling.

PROPERTIES OF m(CSE)

Theory shows that the mean **m(CSE)** is never zero. We call it **« structural bias. »** It is **usually negligible** (except with traces).

Theory expresses **m(CSE)** in terms of the composition of the lot **L**. Useless except in experimental work.

In practice we never try to compute **m(CSE)** : we just remember that it is not zero and that **we can do nothing about it.**

CONCEPT OF EXACTNESS

Exactness is a property of the error **TSE** itself that can be defined but that is never observed. A selection **would be** said to be

- **EXACT** : **if TSE were** structurally zero, i.e.

TSE ≡ 0 whatever the circumstances

Never observed : **TSE** can never be zero.

The words « exact » and « exactness » do not belong to the sampling vocabulary.

CONCEPT OF ACCURACY

The degree of accuracy of a selection is a property of the means **m(TSE)** and **m(ISE).**

NEW DEFINITION : a selection **IS** said to be

- **ACCURATE**, when **structurally** :

m(ISE) ≡ 0 ➔ m(TSE) ≡ m(CSE)

- **BIASED** when **m(ISE) ≠ 0** :

m(ISE) ≡ SAMPLING BIAS

A selection is accurate **WHEN CORRECT and ONLY THEN.** About ACCURACY …

Kaye, Illinois Inst. of Technology (1967)

« The accuracy of many analytical results has too often been neglected with disastrous financial consequences » …

« The accuracy of many analytical data reports is a mirage because unwitting negligence and false cost consciousness have ensured that a sample of powder taken with cursory swiftness has been examined with costly precision ».

CONCEPT OF REPRODUCIBILITY

The degree of reproducibility of a selection is a property of the variance σ^2**(TSE).**

A selection, a sampling or the resulting sample **IS** said to be :

- **REPRODUCIBLE** (or sufficiently rep.) when

$$\sigma^2(TSE) \leq \sigma_o^2 \quad \text{where …}$$

$$\sigma_o^2 \equiv \textbf{maximum variance acceptable}$$

- **NON-REPRODUCIBLE** (or insufficiently reproducible) when :

$$\sigma^2(TSE) > \sigma_o^2$$

CONCEPT OF REPRESENTATIVITY

The degree of representativity of a selection is a property of the **MEAN-SQUARE r^2(TSE).** A sampling **IS** said to be :

- **REPRESENTATIVE** (or sufficiently representative) when :

$$r^2(TSE) \leq r_o^2 \quad \text{where}$$

$r_o^2 \equiv$ maximum mean-square acceptable

- **NON-REPRESENTATIVE** (or insufficiently representative) when :

$$r^2(TSE) > r_o^2$$

ACHIEVEMENT OF REPRESENTATIVITY

Statistical definition of a mean-square

$$r^2(TSE) \equiv m^2(TSE) + \sigma^2(TSE)$$

To all intents and purposes, a sample is **REPRESENTATIVE** when it is both …

- **ACCURATE** : $m(TSE) \equiv 0$ … and …
- **REPRODUCIBLE** : $\sigma^2(TSE) \leq \sigma_o^2$

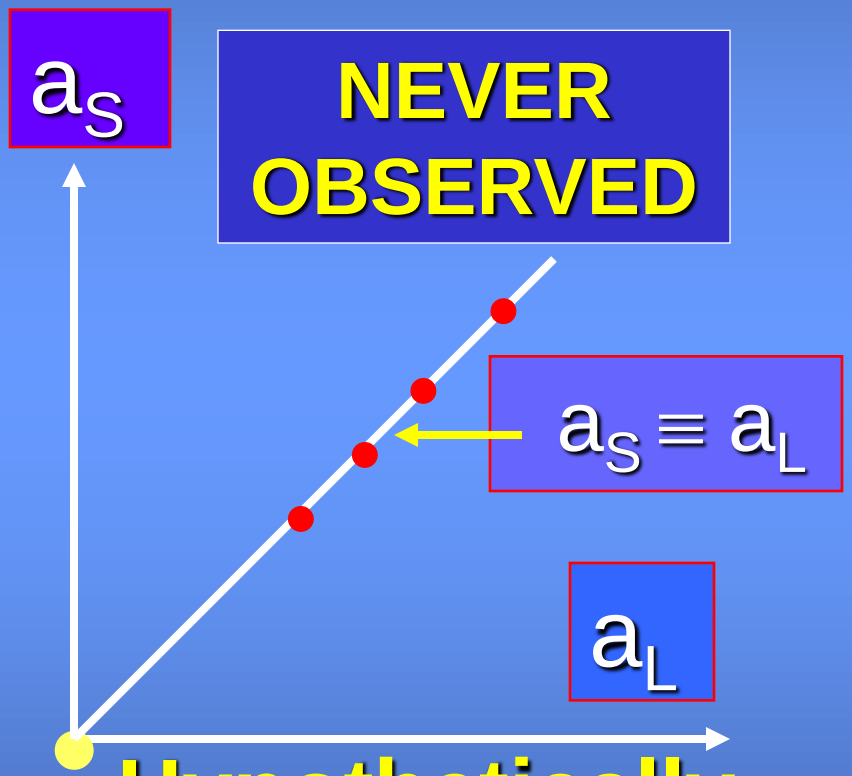
SOLE SCIENTIFIC DEFINITION OF A REPRESENTATIVE SAMPLE

NOTION OF PRECISION

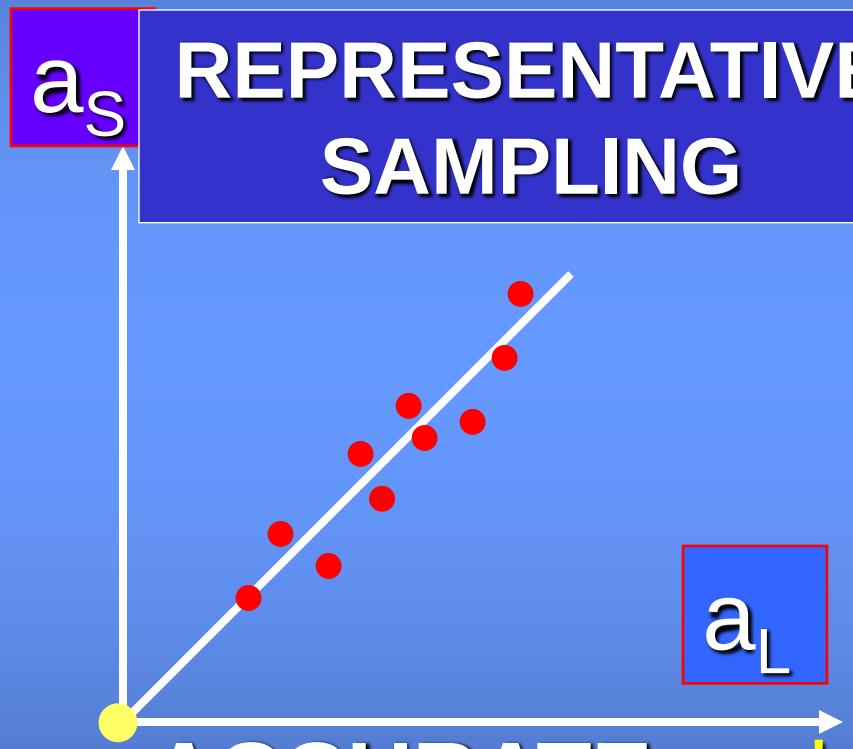
The French / English words **« précision »** and **« precision »** are used in both languages with a vague, ambiguous, non-scientific **GLOBALLY GOOD** meaning. Is it a property of the mean or one of the mean-square ? Authors disagree !

For this reason, we do not recommend their use in a scientific context. We will restrict this use to expressions such as **« precision instrument, precision scales »** or the like.

QUALITIES OF A SELECTION
♦ GRAPHICAL ILLUSTRATION ♦

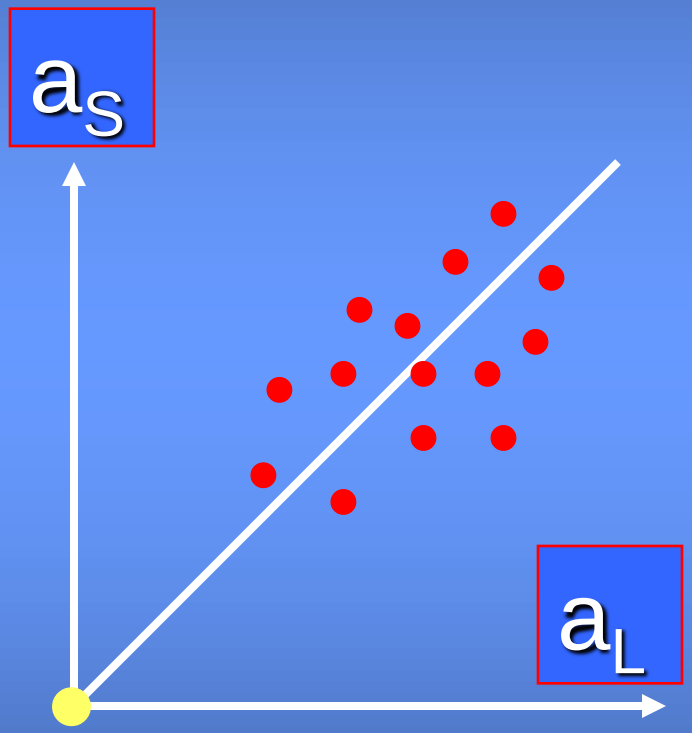


Hypothetically
EXACT SELECTION

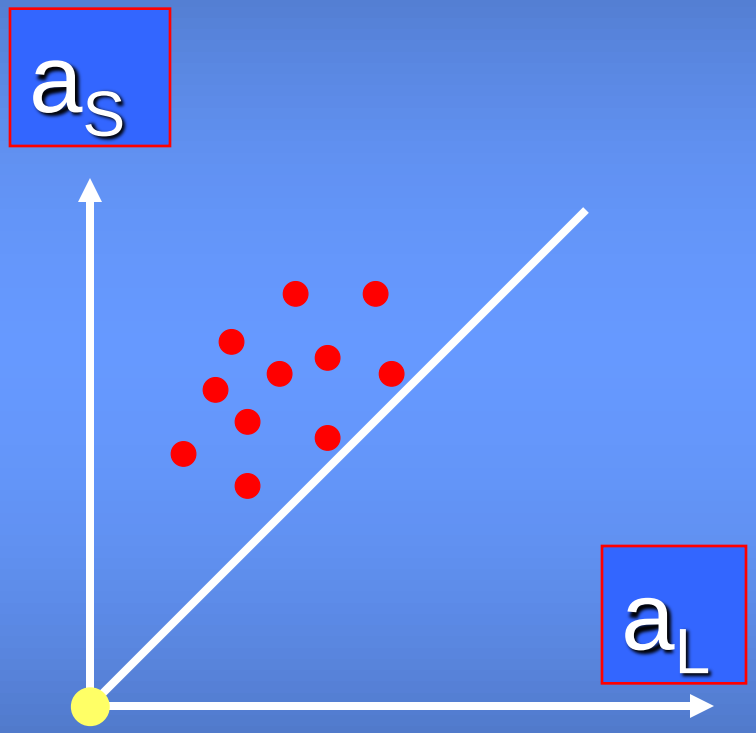


ACCURATE and
REPRODUCIBLE

QUALITIES OF A SELECTION
♦ GRAPHICAL ILLUSTRATION ♦



ACCURATE / POORLY REPRODUCIBLE



BIASED / POORLY REPRODUCIBLE

DEFINITIONS

Relationships between

CONDITIONS AND RESULTS

of a Selection or a Sampling

THE UTILIZER'S STANDPOINT

The utilizer **should be** interested above all by the results i.e. by minimizing the sampling errors. His objective, his interests, are or rather should be, to obtain …

REPRESENTATIVE SAMPLES

that is, samples that are, at the same time …

ACCURATE and REPRODUCIBLE

But he is seldom able to express it.

THE MANUFACTURER'S STANDPOINT

The sampling equipment manufacturer has no **direct** means of controlling the sampling **ERRORS**. The only thing he can do is to control the sampling **CONDITIONS**. He designs or manufactures a device that is

PROBABILIST or not, CORRECT or not

and that respects, or not, the conditions specified in the quantitative approaches to be presented shortly ...

THE THEORETICIAN'S ROLE

- Theory has built up a bridge between **selecting conditions** and **sampling results**
- It tells the utilizer what he can expect and what he must demand from a manufacturer to obtain representative samples,
- It tells the manufacturer what he should or should not do to achieve this purpose,
- Theoreticians should be asked to advise standardization in scientific matters.

THE ACHIEVEMENT OF **ACCURACY**
AS A CONSEQUENCE OF **CORRECTNESS**

is dealt with in the

QUALITATIVE APPROACH

THE ACHIEVEMENT OF **REPRODUCIBILITY**

is dealt with in the

QUANTITATIVE APPROACH

RECAPITULATION

- The achievement of a **PROBABILIST and CORRECT** selection is the only way to obtain **ACCURATE** therefore **RELIABLE SAMPLES.**

- **NON-PROBABILIST** or **PROBABILIST but INCORRECT** selection provides nothing but **BIASED, UNRELIABLE SPECIMENS** on the basis of which no safe decision, **especially with vital or financial consequences** should be made.