SIXTH PART

QUANTITATIVE APPROACH

Answer to question **« How Much ? »**
First case : POPULATIONS

HOMOGENEITY AND HETEROGENEITY

According to our definitions, a **SET** of **UNITS** is said to be :

- **HOMOGENEOUS** : when ALL units are STRICTLY IDENTICAL.

- **HETEROGENEOUS** : when this condition is not fulfilled.

Homogeneity can be defined but can never be observed. Example of pure water made of molecules H_2O, ions H^+, OH^-, O^{2-} to say nothing of all possible isotopic combinations of H and O.

3

The sampling of a homogeneous set by selection of entire units would, by definition of homogeneity, be a …

STRICTLY EXACT OPERATION

All sampling errors, THEREFORE, stem from one form or another of heterogeneity

HETEROGENEITY IS THE UNIQUE SAMPLING ERROR GENERATOR

A theory of sampling is therefore logically derived from a theory of heterogeneity.

THE TWO MAIN FORMS OF HETEROGENEITY

We will distinguish between two main forms of heterogeneity :

- **CONSTITUTIONAL HETEROGENEITY**
The unit is a **single** constitutive element.

- **DISTRIBUTIONAL HETEROGENEITY**
The unit is a **group** of adjoining constitutive elements possibly correlated to one another.

DESCRIPTION OF LOT L

Lot **L** is regarded as a **POPULATION** of

N_U unspecified units U_m : m = 1, 2, … N_U

a_m : proportion of component A in unit U_m

a_L : proportion of component A in lot L

M_m : mass of unit U_m

M_L : mass of lot L ◆ $M_L \equiv N_U \, M_{m*}$ with …

M_{m*} : mass of the average unit U_{m*} of lot L

HYPOTHESIS OF HOMOGENEITY

By definition : $a_m \equiv a_L$ irrespective of m.

Constitutional and distributional homoge-neities can be mathematically defined but are **INACCESSIBLE LIMITS** never to be obser-ved in the real world.

Any hypothesis of homogeneity is the-refore unrealistic / dangerous. It amounts to solving a sampling problem by juggling it out. Lots of money have been lost as the result of this hypothesis. Human lives ?

HETEROGENEITY h_m of UNIT U_m WITHIN LOT L

h_m should logically be proportional to $(a_m - a_L)$ and weighted by the mass M_m

$$K (a_m - a_L) M_m \quad \text{(dimension of a mass)}$$

We found it easier to deal with relative, dimensionless quantities, hence the definition we have retained for h_m :

$$h_m = \frac{(a_m - a_L)}{a_L} \times \frac{M_m}{M_{m*}} \quad \text{(dimensionless)}$$

A thorough definition of h_m is …

« **CONTRIBUTION of UNIT U_m to the HETEROGENEITY of lot L** ».

To shorten this lengthy definition, we will, from now on, call it :

« **The HETEROGENEITY of U_m [in lot L]** »

● The heterogeneity h_m , a function of both grade a_m and mass M_m , is a convenient and thorough descriptor of unit U_m .

10

The **heterogeneity** h_m had attracted our attention when we developed the « **Equiprobable sampling model** » because the « **sampling variance** » was found to be proportional to $\sigma^2(h_m)$. This was confirmed when we developed the much more sophisticated « **PROBABILIST MODEL** » in 1979.

PARAMETER h_m CAN THEREFORE BE REGARDED AS THE TRUE « VECTOR » OF STRUCTURAL SAMPLING ERRORS.

- It is easy to show that, for the N_L units of L :

$$\Sigma_m \, h_m \equiv 0 \longrightarrow m(h_m) \equiv 0$$

- Heterogeneities are additive (like masses)

- Heterogeneity h_n of a group G_n is …

$$h_n \equiv \Sigma_m \, h_m \quad \text{where …}$$

Σ_m is extended to the N_G units U_m of G_n

G_n can be an Increment I or a Sample S

- Thanks to its properties, h_m is a powerful tool in our study of the structural sampling errors expressed by the mathematical model.

12

CONSTITUTIONAL HETEROGENEITY CH_L of LOT L

We first consider lot L as a population of N_F constituents F_i with i = 1, 2, … N_F . Those constituents can be fragments (hence the F), molecules or ions. The heterogeneity of F_i is h_i . We now define …

CH_L : Constitutional Heterogeneity of lot L

$$CH_L \equiv \sigma^2(h_i) \equiv \frac{1}{N_F} \times \Sigma_i \, h_i{}^2 \text{ since } \Sigma_i \, h_i \equiv 0$$

HETEROGENEITY INVARIANT HI_L of LOT L

CH_L suffers from a serious shortcoming. It can usually NOT be computed. For this reason we define the easy to compute HI_L :

HI_L : Heterogeneity Invariant of lot L

$$HI_L \equiv CH_L \times M_{m*} \equiv CH_L \times \frac{M_L}{N_F} \quad \text{(mass)}$$

Use of CH_L will be limited to theoretical developments, that of HI_L to practical issues.

DISTRIBUTIONAL HETEROGENEITY DH_L of LOT L

We now consider lot **L** as a population of N_G groups G_n of adjoining constituents with $n = 1, 2, \dots N_G$. The heterogeneity of G_n is h_n We now define …

DH_L : Distributional Heterogeneity of lot L

$$DH_L \equiv \sigma^2(h_n) \equiv \frac{1}{N_G} \sum_n h_n{}^2 \text{ since } \sum_n h_n \equiv 0$$

15

PROPERTIES OF THE CONSTITUTIONAL HETEROGENEITY CH_L of LOT L

The constitutional heterogeneity CH_L is an intrinsic property of the population of N_L constituents F_i .

It is unaffected by blending mixing or gravity segregation.

The same holds true for the (Constitutional) Heterogeneity Invariant HI_L

16

PROPERTIES OF THE DISTRIBUTIONAL HETEROGENEITY DH_L OF LOT L

On the contrary the **Distributional Heterogeneity DH_L** is a function of :

- The **Constitutional Heterogeneity CH_L**
- The **Size** (number of constituents) of the Groups G_n (the future increments I)
- The **Constituent DISTRIBUTION** throughout lot **L**, which accounts for its name.

17

ZERO-DIMENSIONAL MODEL OF THE SAMPLING PROCESS

Errors / Erreurs ♦ New / Nouvelles Notations

GEE	: Global Estimation Error	EGE	Erreur Globale d'Estimation
TSE	: Total Sampling Error	ETE	Erreur Totale d'Echantillonnage
TAE	: Total Analytical Error	ETA	Erreur Totale d'Analyse
TSE_i	: Total Sampling Error (stage i)	ETE_i	Erreur Totale d'Echant. (étage i)
TSE_1	: Primary Sampling Error	ETE_1	Erreur Primaire d'Echantillonnage
TSE_2	: Secondary Sampling Error	ETE_2	Erreur Secondaire d'Echantillonnage
CSE	: Correct Sampling Error	EEC	Erreur d'Echantillonnage Correct
ISE	: Incorrect Sampling Error	EEI	Erreur d'Echantillonnage Incorrect
FSE	: Fundamental Sampling Error	EFE	Erreur Fondamentale d'Echantillonnage
GSE	: Grouping and Segregation Error	ESG	Erreur de Ségrégation et Groupement
PSE	: Point Selection Error	ESP	Erreur de Sélection de Points
PME	: Point Materialization Error	EMP	Erreur de Matérialisation de Points
IDE	: Increment Delimitation Error	EDP	Erreur de Délimitation des Prélèvements
IXE	: Increment Extraction Error	EXP	Erreur d'Extraction des Prélèvements
IPE	: Increment and Sample Preparation Error	EPP	Erreur de Préparation des Prélèvements et de l'Echantillon

19

Relationships between Errors

Global Estimation Error :

$$GEE \equiv TSE + TAE \equiv (TSE_1 + TSE_2) + TAE \qquad \text{General relationship}$$
$$TSE_i \equiv CSE_i + ISE_i \qquad \text{Total Stage i Sampling Error}$$

Zero-dimensional model (populations) :

$$TSE \equiv CSE + ISE \equiv (FSE + GSE) + ISE$$

One-dimensional model (time series) :

$$TSE \equiv PSE + PME$$
$$PME \equiv CSE + ISE \equiv (FSE + GSE) + ISE$$
$$ISE \equiv IDE + IXE + IPE$$

General Case

$$TSE \equiv (PSE + FSE + GSE) + (IDE + IXE + IPE)$$
$$\text{(structural errors)} + \text{(circumstancial errors that can be cancelled)}$$
$$\text{(cannot be cancelled)} + \text{(can be cancelled)}$$

The total sampling error TSE is the sum of six components

PROBABILIST SAMPLING

The sampling of manageable lots **CAN EASILY**, therefore **MUST** be …

PROBABILIST

Excludes all non-probabilist methods/devices reviewed in the qualitative approach. There can be no theory of non-probabilist sampling.

The development of the probabilist model EXPLICITLY assumes that ALL constituents U_m of lot L have a certain probability $P_m \neq 0$ of being selected. Definitions …

THE PROBABILIST SAMPLING MODEL

This model, developed in 1951 and gene-ralized in 1979, is valid for

- **All 0-dim. lots of manageable size,**
- **All 1-dim. flowing streams on the scale of INCREMENTS.**

It assumes that the N_F constituents F_i of L are submitted to a selection with a probability $P_i \neq 0$ of being selected.

The mean and variance of the …
Total Sampling Error TSE generated **when P_i is not a constant** have been figured out. The most important conclusion is that …

When $P_i \neq$ constant, then m(TSE) is significantly $\neq$ 0. THE SAMPLING IS BIASED

The formulas for mean and variance can be found in our latest books but are **USELESS IN PRACTICE**, except when sampling is **CORRECT** with $P_i \equiv P \equiv$ **constant** …

PRACTICAL IMPLEMENTATION OF THE PROBABILIST MODEL

With few exceptions – such as studies on small numbers of units in artificial conditions – the formulas involving the set of values of P_i can be applied only when :

$$P_i \equiv P \equiv \text{constant}$$

i.e. when the selection is **CORRECT**. From now on we **EXPLICITELY** assume that **THIS CONDITION IS DULY FULFILLED.**

THE FUNDAMENTAL SAMPLING ERROR FSE

- **DEFINITION** :

The **Fundamental Sampling Error FSE** is the error generated when the N_F elements of lot **L** are submitted to a selection …

- ■ **With a uniform probability P** of being selected (selection is assumed to be correct)

- ■ **One by one and independently**.

FSE is the incompressible minimum of **TSE.**

- **STATISTICAL PROPERTIES OF FSE**

  ■ **Distribution** :

    ◆ normal down to 1 ppm (or 1 g/t or 10^{-6})
    ◆ log-normal below 1 ppm.

  ■ **Mean** : **in first approximation**

    m(FSE) $\cong$ 0 : **the sampling is unbiased**

In second approximation :

$$m(FSE) \cong - \frac{1 - P}{P} \times \sum_i \left[\frac{a_i - a_L}{a_L} \times \frac{M_i^2}{M_L^2} \right]$$

■ **Variance** : first approximation …

$$\sigma^2(FSE) \cong \frac{1-P}{P\,N_F} \times CH_L \cong \left[\frac{1}{M_S} - \frac{1}{M_L} \right] HI_L$$

The second approximation is given in our books. Is is useless in practice.

When the sampling ratio is low and to all intents and purposes …

When $M_S \ll M_L$ ◆ $\sigma^2(FSE) \cong HI_L / M_S$

27

- **PRACTICAL EXPRESSION OF σ^2(FSE)**

The expression of **HI$_L$** can be simplified :

$$\text{HI}_L \cong c \; \beta \; f \; g \; d^3 \quad \text{(dimension of a mass)}$$

c : **constitutional factor** (math. defined)

β : **liberation factor** : $0 < \beta \leq 1$

f : **particle shape factor** : practically $f \cong 0.5$

g : **size distribution factor** : $0 < g \leq 1$

d or **d$_{95}$** : **top particle size** (95% passing)

$$\sigma^2(\text{FSE}) \cong \left[\frac{1}{M_S} - \frac{1}{M_L} \right] c \; \beta \; f \; g \; d^3$$

- **CONSTITUTIONAL FACTOR c**

$$c \cong \frac{1 - a_L}{a_L} \times [(1 - a_L)\,\delta_A + a_L\,\delta_G]\ (\text{g.cm}^{-3})$$

δ_A and $\delta_G \equiv$ densities of A and non-A (G)

- ■ Simplified ◆ High-grade : $c \cong (1 - a_L) \times \delta_G$
- ■ Simplified ◆ Low-grade : $c \cong \delta_A \div a_L$
- ■ Simplified ◆ Uniform δ : $c \cong \delta\,(1 - a_L) \div a_L$

- **UNITS** : δ and c always in g.cm^{-3} (tons.m^{-3})
  a_L always in decimal value : 1% $\equiv$ 0.01
  1 ppm $\equiv$ 1 g/t $\equiv 10^{-6}$ (dimensionless)

- **LIBERATION FACTOR β :**

Its name is derived from the mineral industry where components are intertwined …

$$0 \leq \beta \leq 1 \quad \text{(dimensionless)}$$

Generally speaking, the only safe way to estimate β is experimental (see books).

For gold ores, D. François-Bongarçon proposes the following expression derived from his studies and experience …

$$\beta \cong (d_{lib} / d)^{1.5}$$

where d_{lib} is the liberation size of gold.

- **OTHER FACTORS f , g , d :**

f : **shape factor** is a coefficient of cubicity
Except for flat or elongated elements
practically : $f \cong 0.5$

g : **size distribution factor**

- ■ non-calibrated populations : $g \cong 0.25$
- ■ calibrated populations : $0.4 < g < 0.8$

d or **d_{95}** : opening of the square mesh that retains (or would retain) **5 %** of mass M_L.

- **UNITS :** f , β and g are dimensionless, d should be expressed in cm.

- **NULLIFICATION OF σ^2(FSE)**

$$\sigma^2(FSE) \cong \left[\frac{1}{M_S} - \frac{1}{M_L} \right] c \, \beta \, f \, g \, d^3$$

■ First mathematical solution :

$M_S \equiv M_L$: sample S $\equiv$ lot L. Useless.

■ Second mathematical solution :

$HI_L \cong c \, \beta \, f \, g \, d^3 = 0$ ♦ $\beta \, f \, g \, d^3$ always > 0

$c \equiv 0$ when $a_L \equiv 1$: L is pure component A

- **CONCLUSION :** the fundamental variance σ^2(FSE) is never zero → σ^2(FSE) > 0 32

● MINIMIZATION OF σ^2(FSE)

$$\sigma^2(\text{FSE}) \cong \left[\frac{1}{M_S} - \frac{1}{M_L} \right] c \, \beta \, f \, g \, d^3$$

■ First solution involving M_S : M_S should be as large as economically possible.

■ Second solution : $c \, \beta \, f \, g \, d^3$ as small as possible. $c \, \beta \, f \, g$ are data of the problem. The only solution is to reduce d (solid fragments).

● **CONCLUSION**. Two solutions : to increase M_S and/or to reduce **d** (with solids only). 33

• SOLVING SAMPLING PROBLEMS

Let **K** be the constant : **K = c β f g** and σ^2 be the variance σ^2**(FSE)** $\cong$ **K** d^3 / M_S

We can solve three problems :

■ knowing K , M_S and d, estimate the corresponding **"fundamental variance σ^2 "**.

■ knowing K , σ_0^2 and d, estimate the minimum sample mass $M_{S0} \cong$ K d^3 / σ_0^2

■ knowing K , M_S and σ_0^2 estimate required top particle size d. Only with particulate solids which can be crushed or ground. 34

- **PRACTICAL COMPUTATION TOOLS**

The practical expression of HI_L has been published in 1951. To solve sampling pro-blems, several tools have been proposed :

- 1955 : Pierre Gy : Charts
- 1956 : Pierre Gy : Circular Nomogram
- 1965 : Pierre Gy : Sampling Slide Rule

COMPUTER PROGRAMS

- 1986 : Pentti Minkkinen : Sampex
- 1986 : BRGM : Echant & Samp 1.0
- 1997 : BRGM : Samp 2.0

35

● **DEFINITION ♦ A REMINDER** : The **Correct Sampling Error CSE** is defined as the error generated when the N_U **units** making up the lot **L** are submitted to the selection process with a **uniform probability P** of being selected.

We must now distinguish between two cases …

● The N_F elements of lot **L** can be submitted to the selection …

■ **EITHER one by one** and **independently** this selection generates the **Fundamental Sampling Error FSE**. We can do no better…

$$\text{CSE} \equiv \text{FSE}$$

■ **OR by groups of adjoining, non-independent elements** : this generates the **ADDITIONAL Grouping and Segregation Error GSE** (most frequent case) :

$$\text{CSE} \equiv \text{FSE} + \text{GSE}$$

BEWARE OF FORMULAS

One should not forget, however, **AS MANY PEOPLE DO** that **FSE** is **NOT** the sole error involved and that it is the most important only with coarse solids. **NEVER FORGET THAT** formulas such as the formula giving σ^2(**FSE**) are valid **ONLY** if a certain number of conditions are fulfilled, namely :

• The selection is **CORRECT**,

• The elements are taken **ONE BY ONE** and **INDEPENDENTLY**.

These are VERY SELDOM FULFILLED ! 39

GROUPING AND SEGREGATION ERROR GSE

● **THEORY** : **FSE** is the component of the correct error **CSE** generated when the **N$_F$** elements of **L** are submitted to the selection **ONE BY ONE** and **INDEPENDENTLY**.

● **ACTUAL PRACTICE** : elements are extracted by **INCREMENTS** i.e. by **GROUPS OF NON-INDEPENDENT ADJOINING ELE-MENTS** which generates an additional error …

- **GROUPING AND SEGREGATION ERROR** **GSE** is the error that **adds up to** **FSE** when the elements are submitted to a selection …

  ■ With a **uniform probability P** of being selected : the selection is **CORRECT**,

  ■ **By groups of adjoining non-independent elements** forming the **INCREMENTS**.

  ■ **Adjoining elements are often correlated** as the result of differential segregation in Earth's field of gravity. Hence the name of **Grouping and Segregation Error** **GSE** .

STATISTICAL PROPERTIES of GSE

- **DISTRIBUTION OF GSE** : normal except with traces …

- **EXPECTED VALUE m(GSE)** : when the selection of groups is correct : **m(GSE) ≅ 0**

- **VARIANCE σ^2(GSE)** : in practice, can be estimated experimentally by difference, only :

$$\sigma^2(GSE) \equiv \sigma^2(CSE) - \sigma^2(FSE)$$
$$\sigma^2(GSE) \equiv Y \times Z \times \sigma^2(FSE)$$

with :
- ■ **Y ≡ Grouping factor**,
- ■ **Z ≡ Segregation factor**.

42

NULLIFICATION OF σ^2(GSE)

- **σ^2(GSE),** product of three factors is nullified when, and only when, one of them is zero :
  - ■ **Y : Grouping factor $\equiv 0$** when increments are made of **one, and only one**, element.
  - ■ **Z : Segregation factor $\equiv 0$** when, and only when **L has been well homogenized.**
  - ■ **σ^2(FSE)** : is never zero (see above).

- **CONCLUSION** : the only way to nullify **GSE would be to homogenize L prior to sampling**. Elements distributed at random.

MINIMIZATION OF σ^2(GSE)

Homogenizing, perfect or not, can be achieved only if **L is small enough** to be thoroughly mixed.

- σ^2**(GSE),** usually not zero, is minimized

  - ■ By **taking small increments**,
  - ■ By **mixing L** as thoroughly as possible,
  - ■ By **minimizing σ^2(FSE)** as seen above.

- **CONCLUSION** : we never try to estimate **GSE**. We minimize it as best as we can.

ROOTS of FSE and GSE

- **FSE** stems from the sole **CONSTITUTIONAL HETEROGENEITY** CH_L **of** L.
  Like CH_L, to which it is proportional, $\sigma^2(FSE)$ can never be nullified.

- **GSE** stems from the sole **DISTRIBUTIONAL HETEROGENEITY** DH_L **of L.**
  Like DH_L, $\sigma^2(GSE)$ can be minimized by homogenizing lot **L** prior to sampling.

PROPERTIES OF THE CORRECT SAMPLING ERROR CSE

● The **Correct Selection Error CSE** is the sum of two, and only two, components :

$$\text{CSE = FSE + GSE}$$

■ a **structural error FSE** which can never be nullified but can be estimated and …

■ a **circumstantial error GSE** which cannot be estimated theoretically.

● We know how to minimize both **FSE** and **GSE** and with them **CSE**.

46

THE INCORRECT SAMPLING ERROR ISE

So far we have assumed that the conditions of correct sampling were fulfilled and that the **Total Sampling Error TSE** was limited to the **Correct Sampling Error CSE**.

But this is not always the case and an **Incorrect Sampling Error ISE** often takes place. **ISE** is more frequently observed with one-dimensional lots. For this reason, **ISE** will be studied in the seventh part.

47