Note statistique N° 6

Éliminer l'erreur systématique d'échantillonnage.

Il ne s'agit pas ici d'une erreur imputable à l'imperfection du mode de prélèvement, par exemple du mode systématique de fines d'une pelle-tée; Kesige a montré dans sa thèse (voir référence X note St. n° 4) que, sans avec un mode d'échantillonnage parfaitement aléatoire, ou seul fait que l'on prétend estimer la teneur d'un pannneau de volume V, du moyen d'un ou de plusieurs échantillons de volume global inférieur à V, on en introduit une erreur systématique qui conduit à sous-estimer la teneur du pannneau, si elle est basse, et à la surestimer, si elle est élevée.

En effet, reprendre les notations de la Note n° 4, désignons par la variance des pannneaux de volume V, par X la teneur moyenne du pannneau, par x l'évaluation de X faite par le mode d'échantillonnage adopté qui peut être quelconque, et v est encore que v suit une loi lognormale de moyenne $\mu_v$ et de variance $\sigma_v^2$, soient enfin $\mu_x$ la médiane des x dans le pannneau, et $\mu_x^*$ celle des x dans le glissement, $\mu_x$ étant également la médiane les $\mu_v$.

On connaît x et on cherche à estimer $\mu_v$. Déterminons donc la loi de probabilité suivie par X lorsque x est fixé. On a

\[ \mu_v = \mu_x \frac{1}{\sigma_v^2} \]

Pour x fixé, le coefficient de corrélation de $\mu_v$, étant $\frac{1}{\sigma_v^2}$, la lognormale est

\[ \begin{cases} \mathcal{N}(\mu_v; \sigma_v^2) \\ \mathcal{N}(\mu_x; \sigma_x^2) \end{cases} \]

\[ \begin{align*} \mathcal{N}(\mu_v; \sigma_v^2) & = \mathcal{N}(\mu_x; \sigma_x^2) \\ \mathcal{N}(\mu_v^*; \sigma_v^*), \mathcal{N}(\mu_x^*; \sigma_x^*), \mathcal{N}(\mu_x^*; \sigma_x^*) \end{align*} \]

Comme $\mu_v^*$ est lognormale avec

\[ \begin{cases} \mathcal{N}(\mu_v^*; \sigma_v^*), \mathcal{N}(\mu_x^*; \sigma_x^*), \mathcal{N}(\mu_x^*; \sigma_x^*) \end{cases} \]

et la valeur moyenne de $\mu_v^*$ a pour logarithme

\[ \mathbb{E}(\mu_v^*) = \mathbb{E}(\mu_x^*) + \mathbb{E}(\sigma_v^2) \mathbb{E}(\mu_x^*) + \mathbb{E}(\sigma_x^2) \mathbb{E}(\mu_x^*) / 2 \sigma_v^2 \]

\[ \begin{align*} \mathbb{E}(\mu_v^*) & = \mathbb{E}(\mu_x^*) + \mathbb{E}(\sigma_v^2) \mathbb{E}(\mu_x^*) + \mathbb{E}(\sigma_x^2) \mathbb{E}(\mu_x^*) / 2 \sigma_v^2 \end{align*} \]

C'est cette valeur qui doit être prise comme estimation de la teneur d'un pannneau pour lequel un mode d'échantillonnage est de variance $\sigma_v^2$. Donné la valeur x, on peut que pour $\sigma_v^2 = 0$ que la valeur moyenne de $\mu_v^*$ est celle à x.

Pratiquement, si $\sigma_v^2 = 0$, on n'est jamais mal, et la teneur x doit être multipliée par le coefficient correcteur $K$ (Block Plan Factor de Kesige).
K est plus grand que 1 lorsque K est inférieur à \( \gamma \). 

La valeur probable de la teneur d'un bloc de minéral pour lequel l'échantillon prélevé à un teneur \( x \) n'est pas \( x \), mais \( Kx \). 

Cela signifie, de façon précise, que la teneur moyenne d'un grand nombre de blocs échantillonnés à la valeur \( x \), n'est pas \( x \), mais \( Kx \).

Le rôle du coefficient \( K \) est de corriger chaque mesure individuelle \( x \). Mais si l'on substitue à la population des \( x \) celle des \( Kx \), on obtient une population lognormale de même valeur moyenne \( x \), mais cette fois, trop concentrée, puisque sa variance est \( \frac{\sigma^2}{K^2} < \sigma^2 \).

Il n'y a rien de naturel, puisqu'en substituant \( x \) à \( x \) ou à \( Kx \), on a négligé la variance des \( x \) pour \( x \) fixé, variance donnée par (\( x \)).

La somme de ces variances \( \sum \frac{\sigma^2}{K^2} \) est naturellement égale à la variance \( \sigma^2 \) des blocs.

On peut dire que le coefficient \( K \) permet de se faire une meilleure idée de chaque portion du gisement.

Si au contraire on cherche une représentation d'ensemble, il faut avoir recours à la loi lognormal de moyenne \( x \) et de variance \( \sigma^2 \).

Pratiquement les résultats d'analyse des échantillons permettent de déterminer \( x \) et \( \sigma \). A moins donc que le mode d'échantillonnage ne soit presque parfait (\( \chi \) très petit), on ne connait que la teneur moyenne du gisement.

Pour déterminer le coefficient corrigé \( K \), ayant bien que pour fixer, la répartition du tonnage dans les tranches de teneur, il est obligatoire de connaître la variance \( \sigma^2 \). Le mode d'échantillonnage de prélever sur un même bloc et d'analyser un nombre suffisant d'échantillons semblables. Connaissant \( \sigma^2 \) et on déduit la variance valeur de différence:

\[ \sigma^2 = \sigma^2 - \sigma^2 \]