Записки Ленинградского ордена Ленина, ордена Октябрьской Революции
и ордена Трудового Красного Знамени горного института
им. Г.В.Плеханова, т.ХУШ, вып.2, 1975 г.

ВЛИЯНИЕ ПОГРЕШНОСТИ ОПРОБОВАНИЯ НА ВЕЛИЧИНУ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ МЕЖДУ ГЕОЛОГО-ПРОМЫШЛЕННЫМИ ПАРАМЕТРАМИ

Г.С.ПОРОТОВ

Коэффициент корреляции как мера линейной взаимосвязи между геолого-промышленными параметрами (мощностью рудных тел, содержанием компонентов в рудах и др.) широко применяется на практике и имеет многочисленные приложения. Между тем значения геолого-промышленных параметров всегда измеряют с какой-то погрешностью, что, несомненно, влияет на величину коэффициента корреляции и может даже привести к ошибочному заключению о взаимосвязи геолого-промышленных параметров. Рассмотрим влияние погрешности опробования на величину коэффициента корреляции между содержаниями двух компонентов в руде.

Случайная погрешность опробования, которая обусловлена техникой отбора, обработки и анализа проб приводит к тому, что вместо истинных содержаний в основных пробах x_i и в контрольных пробах y_i устанавливают наблюдаемые содержания

$$x_н = x_i + \delta_x ; \qquad y_н = y_i + \delta_y ,$$

где δ_x и δ_y – случайные погрешности.

Если сделать вполне правдоподобные предположения о том, что δ_x и δ_y имеют симметричное распределение (положительные и отрицательные погрешности равновероятны) и не зависят от значения измеряемых величин и друг от друга, то можно показать, что погрешности влияют главным образом на дисперсии содержаний, увеличивая их так, что мы устанавливаем не истинные дисперсии σ_x^2 и σ_y^2, а наблюдаемые

$$\sigma^2_{x_н} = \sigma_x^2 + \sigma^2_{x_{опр}} ; \qquad \sigma^2_{y_н} = \sigma_y^2 + \sigma^2_{y_{опр}} ,$$

где $\sigma^2_{x_{опр}}$ и $\sigma^2_{y_{опр}}$ – дисперсии погрешностей опробования. Дисперсии погрешностей опробования можно установить контрольным опробованием и вычислением по формуле

$$\sigma^2_{\text{опр}} = \frac{1}{2n} \sum_{i=1}^{n} (x_i - y_i)^2 ,$$

где n – число контрольных проб.

Как известно, коэффициент корреляции

$$r = \frac{K_{xy}}{\sigma_x \sigma_y} , \qquad (1)$$

где K_{xy} – корреляционный момент.

При наличии погрешности опробования мы вычисляем оценку не истинной величины r, а наблюдаемой

$$r_{\text{н}} = \frac{K_{xy}}{\sigma_{x_{\text{н}}} \sigma_{y_{\text{н}}}} = \frac{K_{xy}}{\sqrt{\sigma_x^2 + \sigma^2_{x_{\text{опр}}}} \cdot \sqrt{\sigma_y^2 + \sigma^2_{y_{\text{опр}}}}} . \qquad (2)$$

Сравнение формул (1) и (2) показывает, что коэффициенты корреляции связаны соотношением

$$r = r_{\text{н}} \alpha_x \alpha_y , \qquad (3)$$

где α_x и α_y – поправки на погрешность опробования,

$$\alpha_x = \sqrt{1 + \frac{\sigma^2_{x_{\text{опр}}}}{\sigma_x^2}} ; \qquad \alpha_y = \sqrt{1 + \frac{\sigma^2_{y_{\text{опр}}}}{\sigma_y^2}} .$$

Поскольку величины $\alpha_x > 1$ и $\alpha_y > 1$, r по абсолютной величине больше, чем $r_{\text{н}}$.

Таким образом, случайная погрешность опробования снижает коэффициент корреляции. При большой погрешности снижение может оказаться настолько значительным, что зависимость между содержаниями компонентов не будет выявлена.

Формула (3) позволяет оценивать реальную величину коэффициента корреляции. Хотя формула (3) получена на примере содержаний компонентов в рудах, она имеет более общее значение при исследовании взаимосвязи между любыми геолого-промышленными параметрами.

Необходимо учесть, что значения α_x и α_y практически не зависят от числа наблюдений: увеличение числа наблюдений не повышает коэффициент корреляции (хотя надежность его вычисления растет).

Величина поправки зависит от содержания компонентов. Значение α для главного компонента руд – железа (содержание измеряется десятками процентов) – близко к единице (см. таблицу), т.е. погрешность опробования практически не влияет на коэффициент корреляции содержаний железа с содержаниями других компонентов. Для серы (содержание – несколько процентов) величина α немного больше, но тоже близка к единице. Но для кобальта – типичного попутного компонента (содержание – тысячные и сотые доли процента) – $\alpha > 2$, т.е. в данном случае погрешность опробования может снизить коэффициент корреляции в два и более раза. Действительно, коэффициент корреляции содержаний

Примеры поправок к коэффициенту корреляции

Компонент	Наблюдаемая дисперсия содержаний	Дисперсия погрешности опробования	Истинная дисперсия содержаний	Поправка на погрешность опробования
Железо	47,6	1,02	46,6	1,01
Сера	3,27	0,23	3,04	1,04
Кобальт	0,0000144	0,0000116	0,0000028	2,27

кобальта с содержаниями железа и серы редко поднимается выше 0,4, тогда как истинное значение r, согласно формуле (3), можно было бы ожидать 0,8–0,9.

Приведенный пример показывает, что погрешность опробования сильнее всего сказывается на корреляции содержаний попутных компонентов, что надо учитывать при геологических исследованиях.