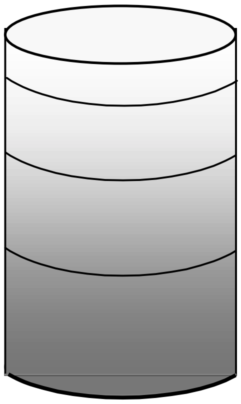
Esimerkkejä ositetun otannan käytöstä

Esimerkki 1

Suklaatehtaalla otetaan suklaamassasta rasvapitoisuuden keskiarvon määrittämiseksi koontanäyte, jonka koko on 10 kg. Kun sula jäähtyy, ehtivät suklaamassan kiintoaineet sedimentoitua ja kerrostua siten on odotettavissa, että myös rasvapitoisuus on eri kerroksissa erilainen. Tällöin on ositettu otanta näytteenotossa on varteenotettava vaihtoehto. Jos pla jaetaan yläosasta alkaen, 1kg:n, 2 kg:n, 3 kg:n ja 4 kg:n kokoisiksi osanäytteiksi alla esitetyllä tavalla ja osien sisäiset keskihajonnat tunnetaan, niin tutkitaan näytteenvalinnan vaikutusta keskiarvon epävarmuuteen. Rasvapitoisuus määritetään 1 g:n suuruisesta näytteestä.



$M_1 := 1\cdot kg,\quad s_1 = 0.032\ ,\qquad (s_1)^2 = 1\times 10^{-3}$

$M_2 := 2\cdot kg,\quad s_2 = 0.045\ ,\qquad (s_2)^2 = 2\times 10^{-3}$

$M_3 := 3\cdot kg,\quad s_3 = 0.022\ ,\qquad (s_3)^2 = 5\times 10^{-4}$

$M_4 := 4\cdot kg,\quad s_4 = 0.045\ ,\qquad (s_4)^2 = 2\times 10^{-3}$

Ilman ositusta yhden satunnaisnäytteen varianssi, $s^2 = 4.14\times 10^{-3}$ ja $s = 0.064$

Rasvapitoisuus määritetään 1 g:n suuruisesta näytteestä. Osien suhteellinen koko voidaan määritellä esim. siten, montako näytettä niistä voitaisiin valmistaa. Koot ovat siten:

$$N_1 := \frac{1000\,g}{1\cdot g} \quad ==> \quad N_1 = 1\times 10^3 \qquad N_2 := \frac{2000\,g}{1\cdot g} \quad ==> \quad N_2 = 2\times 10^3$$

$$N_3 := \frac{3000\,g}{1\cdot g} \quad ==> \quad N_3 = 3\times 10^3 \qquad N_4 := \frac{4000\,g}{1\cdot g} \quad ==> \quad N_4 = 4\times 10^3$$

$$N_t = \sum_i N_i = 1\times 10^4$$

Tutkitaan, miten keskiarvon luotettavuus riippuu näytteen valinnasta:

A Otetaan 15 osanäytettä satunnaisesti:

$n := 15$

$$s_x := \frac{s}{\sqrt{n}} \quad = \quad \frac{0.064}{\sqrt{15}} = 0.0165$$

Keskiarvon kokonaisepävarmuus 95 % luottamustasolla on

$$P_{.95} := 2 \cdot s_x \quad = \quad 2 \cdot 0.0165 = 0.033$$

B Otetaan 15 osanäytettä siten, että ositus on verrannollinen osajoukon kokoon:

$$n_i := \frac{N_i}{N_t} \cdot n \qquad ==> \qquad n_i = \begin{bmatrix} 1.5 \\ 3 \\ 4.5 \\ 6 \end{bmatrix}$$

Koska näytemäärä ei voi olla murtoluku, pyöristetään tulokset kokonaisluvuiksi, niin että näytteiden yhteismääräksi tulee 15

$$n_i := \begin{bmatrix} 2 \\ 3 \\ 4 \\ 6 \end{bmatrix}$$

$$\sum n = 15$$

Tämä ositus antaa keskiarvon keskihajonnaksi (oletetaan $N_i >> n_i$)

$$s_x := \sqrt{\sum_i \left(\frac{N_i}{N_t}\right)^2 \cdot \frac{(s_i)^2}{n_i}} \qquad ==> \qquad s_x = 9.81 \times 10^{-3}$$

ja keskiarvon kokonaisepävarmuudeksi 95 % luottamustasolla

$$P_{.95} := 2 \cdot s_x \quad = \quad 2 \cdot 0.00981 = 0.0196$$

Ositus ja osajoukon kokoon verrannollinen otanta johti tässä tapauksessa huomattavasti parempaan tulokseen kuin satunnaisotanta.

C Kustannusten suhteen optimoitu ositettu otanta

Jos oletetaan, että yhden rasvamäärityksen hinta on 20 mk, niin otanta voidaan optimoida.
Kumpikin edellisistä määrityksistä maksoi siten 15 x 20 mk = 300 mk. Suoritetaan optimointi siten, että maksimikustannukset saavat olla enintään 300 mk.

$c_i := 20 \cdot mk = c'$ ja kokonaiskulut $c_t := 300 \cdot mk$

Kun kustannukset on määrätty, saadaan optimoitu ositus kaavasta

$$n_i := \frac{c_t}{c'} \cdot \frac{N_i \cdot s_i}{\sum_i (N_i \cdot s_i)} \quad ==> \quad n_i = \begin{bmatrix} 1.292 \\ 3.655 \\ 2.742 \\ 7.311 \end{bmatrix} \quad \text{valitaan} \quad n_i := \begin{bmatrix} 1 \\ 4 \\ 3 \\ 7 \end{bmatrix}$$

$$\sum n = 15$$

Keskiarvon keskihajonnaksi saadaan nyt:

$$s_x := \sqrt{\sum_i \left(\frac{N_i}{N_t}\right)^2 \cdot \frac{(s_i)^2}{n_i}} \quad ==> \quad s_x = 9.52 \times 10^{-3}$$

ja keskiarvon kokonaisepävarmuudeksi 95 % luottamustasolla

$$P_{.95} := 2 \cdot s_x = 2 \cdot 0.00952 = 0.019$$

Tässä tapauksessa kustannusten suhteen optimoitu otanta ei parantanut paljonkaan tulosta verrannolliseen otantaan verrattuna. Joissakin toisissa tapauksissa ero voi olla merkittävä, erityisesti, jos osien sisäiset keskihajonnat poikkeavat toisistaan enemmän kuin tässä esimerkissä ja näytteitä joudutaan tutkimaan enemmän.

D Optimoidaan otanta, kun luottamustaso keskiarvolle on määrätty

Optimoidaan otanta, kun halutaan, että keskiarvon kokonaisepävarmuus 95 %:n luottamustasolla on enintään 0.010

$$P_{.95} := 2 \cdot s_x \quad ==> \quad \text{Tavoitekeskihajonta } s_T := \frac{0.01}{2} \quad ==> \quad s_T = 5 \times 10^{-3}$$

$$s_T^2 = 2.5 \times 10^{-5}$$

Kun keskiarvon tavoitekeskihajonta on määrätty saadaan optimaalinen ositus seuraavasta kaavasta, kun määrityksen hinta on sama jokaisessa osassa.

$$n_i := \frac{s_i \cdot N_i \cdot \sum_i (N_i \cdot s_i)}{N_t^2 \cdot s_T^2} \quad ==> \quad n_i = \begin{pmatrix} 4.643 \\ 13.131 \\ 9.849 \\ 26.263 \end{pmatrix} \quad \text{valitaan} \quad n_i := \begin{pmatrix} 5 \\ 13 \\ 10 \\ 26 \end{pmatrix}$$

$$\sum n = 54$$

Tarkistetaan vielä, että saatu tulos toteuttaa tavoitteen:

$$s_x := \sqrt{\sum_i \left(\frac{N_i}{N_t}\right)^2 \cdot \frac{(s_i)^2}{n_i}} \quad ==> \quad s_x = 5 \times 10^{-3} \; = \; s_T$$

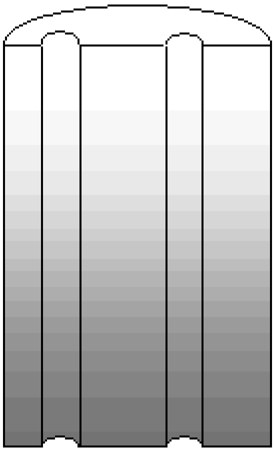
Määrityksen kokonaiskulut ovat

$$c_t := \left(\sum n\right) \cdot c' \; = \; 54 \cdot 20 \cdot mk = 1080mk$$

Kuten havaitaan vertaamalla tätä tulosta edelliseen kohtaan, niin optimoidun näytteenottoketjun kokonaisepävarmuuden pienentäminen 0.019:stä 0.01:een lisäsi kustannuksia 300 mk:sta 1080 mk:ksi, t.s. tuloksen paraneminen on verrannollinen vain panostuksen neliöjuureen:

$$\sqrt{\frac{300 \cdot mk}{1080 \cdot mk}} \cdot 0.019 = 0.01$$

Edellä kuvattu ositus ei silti välttämättä ole edullisin näytteenoton strategia. Jos sedimentoituminen on tapahtunut tasaisesti, niin suklaan vaakasuorat kerrokset ovat suhteellisen homogeenisia. Jos näytteet otetaan oheisessa kuvassa esitetyllä tavalla pystysuunnassa kaikkien kerrosten läpi, eliminoituu segregoitumisen vaikutus näytteissä, koska tällöin kaikki kerrokset (eli potentiaaliset osanäytteet vaakasuunnassa) ovat edustettuina näytteessä oikeassa suhteessa. Luonnollisesti nämä näytteet on homogenisoitava huolellisesti ennen analyysinäytteen ottamista niistä.



Moniulotteisissa kohteissa parasta näytteenottostrategiaa harkittaessa onkin pyrittävä selvittämään, missä suunnassa heterogeenisuus on suurinta. Tämä suunta on sitten otettava mahdollisimman kattavasti näytteeksi.