ASYMETRIA

W METODACH ZAOKRĄGLANIA

STOSOWANYCH PRZEZ GUS PRZY WYZNACZANIU GĘSTOŚCI ZALUDNIENIA GMIN

Autor:

dr Bogdan Stępień

LUTY 2005

1. WSTĘP

W opracowaniu omówiony jest problem zaokrąglania liczb i zaprezentowane są skutki, jakie mogą wynikać z zaokrąglania zaokrąglonej wcześniej liczby. W pierwszej części opracowania przedstawiona jest teoria zaokrągleń. Zdefiniowana jest funkcja zaokrąglająca i wprowadzone jest pojęcie wskaźnika asymetrii błędu zaokrąglania. Omówione jest zaokrąglanie wcześniej zaokrąglonych liczb oraz zaprezentowany jest wpływ takiego zaokrąglania na wskaźnik asymetrii błędu zaokrąglania. Postawiono twierdzenie, że GUS przynajmniej w przypadku wyznaczania gęstości zaludnienia dla gmin, określanej i podawanej w liczbach całkowitych, dokonuje dwukrotnego zaokrąglania. Na przykładzie Gminy Pajęczno przedstawiono skutek dwukrotnego zaokrąglania gęstości zaludnienia przez GUS w naliczaniu kwoty uzupełniającej wyrównawczej części subwencji ogólnej przez Ministerstwo Finansów. W ostatniej części opracowania podana jest recepta jak ustrzec się asymetrii błędu zaokrąglania.

2. WSTEP DO TEORII Z ZAKRESU ZAOKRĄGLEŃ LICZB

Funkcja $f_n(x)$ zaokrąglająca liczbę x do n miejsca po przecinku, po funkcji dodawania/odejmowania i mnożenia/dzielenia to najbardziej powszechnie stosowana funkcja.

Ludzie się umówili, (dla rozumienia siebie, o czym mówią - to następowało i następuje w procesie nauczania od szkoły podstawowej aż po studia i doświadczenie życiowe), że zawsze jak mówią, że dokonują zaokrąglenia liczby lub mówią o podawaniu liczby z dokładnością do n–go miejsca po przecinku przez to się rozumie, że:
- jeżeli na n+1 miejscu po przecinku w zaokrąglanej liczbie jest cyfra 5, 6, 7, 8 lub 9 wtedy w zaokrąglanej liczbie zastępuje się wszystkie cyfry zerami od n+1 (włącznie) miejsca po przecinku wzwyż oraz dodaje się do zaokrąglanej liczby, liczbę równą 1×10^{-n},
- w przeciwnym przypadku zaokrąglona liczba to liczba wyjściowa z samymi zerami od n+1 (włącznie) miejsca po przecinku wzwyż.

Taka jest umowa pomiędzy ludźmi i tak należy rozumieć funkcję zaokrąglającą $f_n(x)$.

Stosują ją i naukowcy i Kowalscy, wszyscy (no prawie wszyscy) się świetnie rozumieją.

Jeżeli ktoś podaje nam liczbę y i mówi nam, że liczba ta powstała w wyniku zaokrąglenia do n-go miejsca po przecinku to wiemy, że błąd wynikający z zaokrąglenia wynosi

$$\pm5\times10^{-n-1}.$$

Liczba $-5\times10^{-n-1}$ - nazywana jest dolną granicą błędu δ_d liczby y a liczba $+5\times10^{-n-1}$ - górną granicą błędu δ_g liczby y.

Zdefiniujmy teraz pojęcie wskaźnika asymetrii błędu zaokrąglania W_{ABZ} jako

$$\mathrm{W}_{\mathrm{ABZ}} = \left|\delta_g + \delta_d\right|.$$

Wskaźnik ten określa asymetrię pomiędzy górną i dolną granicą błędu zaokrąglania.
Z oczywistych powodów $\mathrm{W}_{\mathrm{ABZ}}$ dla funkcji zaokrąglającej $f_n(x)$ wynosi 0 i można powiedzieć, że jeżeli wskaźnik $\mathrm{W}_{\mathrm{ABZ}}$ jest różny od 0, nawet dowolnie mało, to wtedy mamy do czynienia z wypaczoną funkcja zaokrąglającą, która może dawać wyniki niezrozumiałe i dla naukowca i dla Kowalskiego.

Kiedy może pojawić się sytuacja, gdy wskaźnik $\mathrm{W}_{\mathrm{ABZ}}$ będzie różny od zera?
Otóż wtedy, gdy dokonamy zaokrąglenia liczby x do n-tego miejsca po przecinku a potem wynik tego zaokrąglenia poddany kolejnemu zaokrągleniu. Najbardziej skrajny przypadek dokonania tego typu zaokrągleń przedstawia Tabela 1. W tabeli tej prezentowane są kolejne kroki zaokrągleń z podanym błędem zaokrąglenia oraz wskaźnikiem asymetrii błędu zaokrąglania. Proszę zauważyć,

że z liczby 1,444…4445… po kolejnych zaokrągleniach otrzymano w efekcie liczbę 2. To jest wynik absolutnie zaskakujący oraz niezgodny z intuicją wyrobioną na podstawie doświadczenia życiowego, pobranych nauk, studiów itp. Dla porównania w Tabeli 2 przedstawiony jest sposób kolejnych zaokrągleń według metody poprawnej, zgodnej z definicją funkcji zaokrąglającej $f_n(x)$ oraz zgodny z umową między ludźmi.

Krok zaokrąglania	Zaokrąglanie do określonej liczby miejsc po przecinku	Wynik zaokrąglania liczby 1.444…4445… gdzie cyfra 5 leży na n+1 miejscu po przecinku	Wskaźnik asymetrii błędu zaokrąglania
1	**2**	**3**	**4**
1	n	$f_n(1.444...4445...) = 1.444...445$ błąd zaokrąglania: $\begin{cases} +\ 0.000\ \dots\ 0005 \\ -\ 0.000\ \dots\ 0005 \end{cases}$	0
2	n-1	$f_{n-1}(f_n(1.444...4445...)) = 1.444...45$ błąd zaokrąglania (skumulowany): $\begin{cases} +\ 0.000\ \dots\ 0045 \\ -\ 0.000\ \dots\ 0055 \end{cases}$	0.000...001
…	…	…	…
n-2	3	$f_3(\dots f_{n-1}(f_n(1.444\dots4445\dots))\dots) = 1.445$ błąd zaokrąglania(skumulow.): $\begin{cases} +\ 0.0004\ \dots\ 4445 \\ -\ 0.0004\ \dots\ 5555 \end{cases}$	0.0001...111
n-1	2	$f_2(f_3(\dots f_{n-1}(f_n(1.444\dots4445\dots))\dots)) = 1.45$ błąd zaokrąglania (skumulowany): $\begin{cases} +\ 0.004\ \dots\ 4445 \\ -\ 0.005\ \dots\ 5555 \end{cases}$	0.001...111
n	1	$f_1(f_2(f_3(\dots f_{n-1}(f_n(1.444...4445...))\dots))) = 1.5$ błąd zaokrąglania (skumulowany): $\begin{cases} +\ 0.044\ \dots\ 4445 \\ -\ 0.055\ \dots\ 5555 \end{cases}$	0.011...111
n+1	0	$f_0(f_1(f_2(f_3(\dots f_{n-1}(f_n(1.444\dots4445\dots))\dots)))) = 2$ błąd zaokrąglania (skumulowany): $\begin{cases} +\ 0.444\ \dots\ 4445 \\ -\ 0.555\ \dots\ 5555 \end{cases}$	0.111...111

Tabela 1. Do czego skrajnie może prowadzić stosowanie zaokrągleń do wcześniej zaokrąglanych liczb.

Krok zaokrąglania	Zaokrąglanie do określonej liczby miejsc po przecinku	Wynik zaokrąglania liczby liczby 1.444…4445… gdzie cyfra 5 leży na n+1 miejscu po przecinku	Wskaźnik asymetrii błędu zaokrąglania
1	**2**	**3**	**4**
1	n	$f_n(1.444\dots4445\dots) = 1.444...445$ błąd zaokrąglania: $\pm 5\times10^{-n-1}$	0
2	n-1	$f_{n-1}(1.444\dots4445\dots) = 1.444...44$ błąd zaokrąglania: $\pm 5\times10^{-n}$	0
…	…	…	0
n-2	3	$f_3(1.444\dots4445\dots) = 1.444$ błąd zaokrąglania: $\pm 5\times10^{-4}$	0
n-1	2	$f_2(1.444\dots4445\dots) = 1.44$ błąd zaokrąglania: $\pm 5\times10^{-3}$	0
n	1	$f_1(1.444\dots4445\dots) = 1.4$ błąd zaokrąglania: $\pm 5\times10^{-2}$	0
n+1	0	$f_0(1.444\dots4445\dots) = 1$ błąd zaokrąglania: $\pm 5\times10^{-1}$	0

Tabela 2. Poprawna metoda zaokrąglania liczb. Przy tej metodzie zaokrąglania wskaźnik asymetrii błędu zaokrąglania zawsze wynosi 0 i wynika to wprost z definicji funkcji zaokrąglającej $f_n(x)$.

W Tabeli 3 przedstawiony jest mniej skrajny przypadek w stosunku do Tabeli 1, prowadzący również do asymetrii błędu zaokrąglenia. Sytuacja wyjściowa jest taka: mamy liczbę równą 1.4995.... Dokonujemy jej zaokrąglenia do 3 miejsc po przecinku w wyniku, czego otrzymujemy liczbę równą 1.500. Następnie poddajemy wynik poprzedniego zaokrąglenia, czyli liczbę 1.500 kolejnemu zaokrągleniu tym razem do liczby całkowitej (czyli do zera miejsc po przecinku) w wyniku, czego otrzymujemy liczbę 2. Wskaźnik asymetrii błędu zaokrąglania wynosi 0.001 czyli jest różny od zera, a to świadczy o tym, że otrzymany wynik jest niezgodny z definicją funkcji zaokrąglającej. Z wyjściowej liczby 1.4995 czyli mniejszej od 1.5 otrzymano w efekcie liczbę 2.

Krok zaokrąglania	Zaokrąglanie do określonej liczby miejsc po przecinku	Wynik zaokrąglania liczby 1.4995	Wskaźnik asymetrii błędu zaokrąglania
1	**2**	**3**	**4**
1	3	$f_3(1.4995...) = 1.500$ błąd zaokrąglania: $\begin{cases} +\ 0.0005 \\ -\ 0.0005 \end{cases}$	0
2	0	$f_0(f_3(1.4445\ldots)) = 2$ błąd zaokrąglania (skumulowany): $\begin{cases} +\ 0.4995 \\ -\ 0.5005 \end{cases}$	0.001

Tabela 3. Mniej skrajny przypadek dokonywania zaokrągleń prowadzący do niezerowego wskaźnika asymetrii błędu zaokrąglania.

3. GUS DOKONUJE ZAOKRĄGLEŃ WCZEŚNIEJ ZAOKRĄGLANYCH LICZB

Teza: GUS wyznaczając gęstości zaludnienia dla gmin i podając ją w liczbach naturalnych, dokonuje w czasie obliczeń dwukrotnego zaokrąglania wyników (do drugiego zaokrąglania bierze wynik z pierwszego zaokrąglania), co w konsekwencji prowadzi do niezerowego wskaźnika asymetrii błędu zaokrąglania, to z kolei może implikować zniekształcanie powszechnie rozumianego pojęcia zaokrąglania w tym przypadku gęstości zaludnienia.

Dowód.

Założenie: do wyznaczania gęstości zaludnienia GUS stosuje techniki komputerowe oraz że do wyznaczania gęstości zaludnienia używa tej samej procedury obliczeń dla każdej gminy.

Powierzchnia Miasta i Gminy Pajęczno wg stanu na dzień 31 grudnia 2003 r. wynosiła 11 344 ha a liczba ludności faktycznie zamieszkałych wynosiła 11 741. Na podstawie tych wielkości GUS wyznaczył gęstość zaludnienia dla Miasta i Gminy Pajęczno[1] na 104 osoby/km^2.
Tymczasem gęstość zaludnienia dla tej gminy wynosi:

$$\frac{liczba\ ludnosci \times 100}{powierzchnia\ w\ ha} = \frac{11\,741 \times 100}{11\,344} = 103.4996...\ \text{osób/km}^2. \qquad (*)$$

Dla innej gminy np. dla Gminy Sawin powierzchnia wg stanu na dzień 31 grudnia 2003 r. wynosiła 19 020 ha a liczba ludności faktycznie zamieszkałych wynosiła 5 801. Na podstawie tych wielkości GUS wyznaczył gęstość zaludnienia dla Gminy Sawin[1] na 30 osób/km^2. Dokładna wartość tej gęstości to 30.4994....

Zastanówny się jak można:
(dla Miasta i Gminy Pajęczno) z liczby 103.4996… otrzymać liczbę 104,
a (dla Gminy Sawin) z liczby 30.4994… otrzymać liczbę 30.

Krok pierwszy:

Zaokrąglamy uzyskane wyniki ze wzoru (*) do trzech miejsc po przecinku:

[1] Źródło- „Powierzchnia i ludność w przekroju terytorialnym w 2004 r.”, GUS. Dane na dzień 31-12-2003 r.

z liczby 103.4996… otrzymujemy liczbę 103.500,
a z liczby 30.4994… otrzymujemy liczbę 30.499.

Krok drugi:

Wyniki uzyskane w kroku pierwszym zaokrąglamy do liczb całkowitych:
z liczby 103.500 otrzymujemy liczbę 104,
a z liczby 30.499 otrzymujemy liczbę 30.

GUS twierdzi, że w pierwszym kroku „obliczeń dokonuje do dwóch miejsc po przecinku" a w kroku drugim dokonuje zaokrąglania do liczby całkowitej. Co to może oznaczać, że „obliczeń dokonuje do dwóch miejsc po przecinku"? Na pewno, nie oznacza to tego, że pracownik GUS bierze ołówek oraz kartkę papieru i dokonuje dzielenia dwóch liczb i poprzestaje na wykonaniu dzielenia po uzyskaniu dwóch cyfr po przecinku, bo gdyby tak było nie powstałoby to opracowanie. Zapewne GUS do wyznaczania gęstości zaludnienia tj. dzielenia dwóch liczb przez siebie używa wbudowanych funkcji dzielenia w programach komputerowych i w wyniku dzielenia otrzymuje liczby typu 103.4996… oraz 30.4994…. W związku z powyższym „obliczeń dokonuje do dwóch miejsc po przecinku" należy rozumieć dokonuje zaokrąglania uzyskanego wyniku z wzoru (*) do 3 miejsc po przecinku a nie 2 miejsc po przecinku, bo w przeciwnym przypadku z liczby 30.4994…, otrzymano by liczbę 30.50 w pierwszym kroku a w drugim 31 a tymczasem otrzymano w drugim kroku liczbę 30.

4. KONSEKWENCJE STOSOWANEJ PRZEZ GUS PROCEDURY WYZNACZANIA GĘSTOŚCI ZALUDNIENIA DLA GMIN

Dwukrotne zaokrąglanie gęstości zaludnienia może prowadzić w konsekwencji do zaniżania – jeżeli gęstość zaludnienia w gminie jest większa od połowy gęstości zaludnienia w kraju lub zawyżania – jeżeli gęstość zaludnienia w gminie jest mniejsza od połowy gęstości zaludnienia w kraju, subwencji uzupełniającej dla gmin naliczanej przez Ministerstwo Finansów.

Najlepszym tego przykładem może być rozważna wyżej Gmina Pajęczno. Gdyby GUS ustalił poprawnie (zgodnie z powszechnie rozumiana funkcją zaokrąglającą) gęstość zaludnienia dla tej gminy wg stanu na dzień 31 grudnia 2003 r. tzn. 103 osoby/km^2 zamiast 104 osób/km^2 to wtedy gmina ta otrzymałaby wyższą subwencję uzupełniającą na rok 2005 o kwotę około 12 200 zł.

5. WNIOSKI

Jak ustrzec się asymetrii błędu zaokrąglania?

Nie należy zaokrąglać liczby, która wcześniej była zaokrąglona. Do zaokrąglania bierzemy zawsze pierwotne źródło, w omawianym wyżej przypadku jest to iloraz liczby mieszkańców do powierzchni gminy.

W przypadku określania gęstości zaludnienia dla gmin procedura zaokrąglania powinna wyglądać następująco:
jeżeli chcemy podać gęstość zaludnienia z dwoma miejscami po przecinku to robimy to tak:

$$\rho = f_2\left(\frac{liczba\ \text{mieszkancow}}{\text{powierzchnia gminy}}\right) = f_2\left(\frac{11\,741\times 100}{11\,344\ ha}\right) = f_2(103.4996\ldots) = 103.50\ \text{osoby/km}^2,$$

a jeżeli w liczbach naturalnych (z zerem miejsc po przecinku) to tak:

$$\rho = f_0\left(\frac{liczba\ \text{mieszkancow}}{\text{powierzchnia gminy}}\right) = f_0\left(\frac{11\,741\times 100}{11\,344\ ha}\right) = f_0(103.4996\ldots) = 103\ \text{osoby/km}^2.$$