

Analiza porównawcza danych rzeczywistych z wykorzystaniem metod statystyki opisowej

Paulina Mruczek (277428), Maciej Szczerczowski (277487)

Spis treści

1	Wstęp	2
1.1	Opis zbioru danych	2
1.2	Cel analizy	3
1.3	Oznaczenia	3
1.4	Zaokrąglanie	3
2	Analiza rozkładu przy pomocy metryk statystycznych	3
2.1	Miary położenia	4
2.2	Miary zmienności	6
2.3	Miary kształtu	8
2.4	Miary zależności między zmiennymi	9
3	Prezentacja rozkładów na wykresach	11
3.1	Histogram	11
3.2	Boxplot	13
3.3	Dystrybuanta	15
4	Podsumowanie	17

1 Wstęp

1.1 Opis zbioru danych

W niniejszym sprawozdaniu przedstawiona zostanie analiza dwóch wybranych zbiorów danych rzeczywistych. Pochodzą one ze zbioru danych "Budget Food", który został opracowany na podstawie badania przeprowadzonego w Hiszpanii w 1980 roku. Dane dotyczą wydatków gospodarstw domowych na żywność.

Zbiór danych składa się z 23 972 obserwacji, z których każda odpowiada jednemu gospodarstwu domowemu. Dane obejmują zmienne takie jak:

- **wfood** – procent całkowitych wydatków przeznaczonych na żywność,
- **totexp** – całkowite wydatki gospodarstwa domowego,
- **age** – wiek osoby ankietowanej,
- **size** – liczba osób w gospodarstwie,
- **town** – kategoria wielkości miejscowości zamieszkania (od 1 – małe miejscowości, do 5 – duże miasta),
- **sex** – płeć osoby ankietowanej.

L.p.	wfood	totexp	age	size	town	sex
1	0.467699	1290941	43	5	2	man
2	0.313023	1277978	40	3	2	man
3	0.376482	845852	28	3	2	man
4	0.439691	527698	60	1	2	woman
5	0.403615	1103220	37	5	2	man
...

Do analizy porównawczej wybraliśmy gospodarstwa znajdujące się w małych i dużych miastach, oznaczone odpowiednio cyframi 1 i 5. W zbiorach tych danych znajdowało się odpowiednio 2903 i 2838 obserwacji, jednak dla odpowiednio 3 i 5 obserwacji w tych grupach wartość **totexp** wynosiła zero. W związku z tym, by uniknąć problemów w obliczeniach, na przykład przy średniej geometrycznej i harmonicznej, nie uwzględniamy tych odpowiednio 3 i 5 obserwacji ograniczając nasze dane do 2900 obserwacji dla małych miast i 2833 obserwacji dla dużych miast, co nie powoduje znaczącej różnicy w wynikach przedstawianych przez nas statystyk. Źródłem danych jest artykuł A. Delgado i J. Mora (1998) „Testing non-nested semiparametric models: an application to Engel curves specification”, opublikowany w *Journal of Applied Econometrics* ([link](#)).

1.2 Cel analizy

Celem analizy jest porównanie całkowitych wydatków i udziału wydatków na żywność obu zbiorów poprzez:

- obliczenie i interpretację podstawowych miar statystycznych,
- przedstawienie danych w formie graficznej,
- wyciągnięcie wniosków dotyczących podobieństw i różnic pomiędzy zbiorami.

1.3 Oznaczenia

- n - rozmiar próby,
- x_i - wartość i -tej obserwacji,
- x_{max} - największa wartość w zbiorze danych,
- x_{min} - najmniejsza wartość w zbiorze danych,
- k — liczba wartości, które odrzucamy (średnia ucinana) albo zamieniamy (średnia winsorowska) na początku i na końcu uporządkowanego zbioru danych,
- d_i - różnica między rangami odpowiadających sobie wartości,
- X i Y - zmienne losowe (dwa zestawy danych).

1.4 Zaokrąglanie

Wszystkie wyniki w całej pracy zaokrąglone są do ostatniej cyfry, do jakiej są one zapisane. W większości przypadków dla wyników dotyczących udziału wydatków na żywność zaokrąglenia są do trzech miejsc po przecinku, a dla wyników dotyczących całkowitych wydatków do jedności.

2 Analiza rozkładu przy pomocy metryk statystycznych

Dla wszystkich definicji w tym rozdziale wyniki obliczeń dla badanych danych zostały przedstawione w tabelkach bezpośrednio poniżej, poza ostatnim podrozdziałem o miarach zależności między zmiennymi, gdzie dane zostały przedstawione zbiorczo dla całego podrozdziału. Na koniec każdego podrozdziału podsumowujemy uzyskane w nim wyniki.

2.1 Miary położenia

Definicja 2.1. Średnia arytmetyczna to suma wszystkich wartości w zbiorze danych podzielona przez liczbę tych wartości:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

	Udział wydatków na żywność	Całkowite wydatki
Małe miasta	0,460	643 674
Duże miasta	0,326	1 038 887

Definicja 2.2. Średnia geometryczna to pierwiastek n-tego stopnia z iloczynu n wartości ze zbioru danych:

$$\bar{x}_g = \left(\prod_{i=1}^n x_i \right)^{1/n}.$$

	Udział wydatków na żywność	Całkowite wydatki
Małe miasta	0,416	489 310
Duże miasta	0,286	848 230

Definicja 2.3. Średnia harmoniczna to odwrotność średniej arytmetycznej odwrotności danych:

$$\bar{x}_h = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}.$$

	Udział wydatków na żywność	Całkowite wydatki
Małe miasta	0,323	359 843
Duże miasta	0,211	664 918

Definicja 2.4. Średnia ucinana to średnia arytmetyczna obliczana po odrzuceniu określonej liczby skrajnych wartości, używana, aby zmniejszyć wpływ danych odstających:

$$\bar{x}_{\text{trim}} = \frac{1}{n - 2k} \sum_{i=k+1}^{n-k} x_i.$$

	k	Udział wydatków na żywność	Całkowite wydatki
Małe miasta	10	0,460	630 770
Duże miasta	10	0,324	1 025 726
Małe miasta	100	0,459	595 146
Duże miasta	100	0,321	982 517
Małe miasta	500	0,457	541 741
Duże miasta	500	0,313	923 486
Małe miasta	1000	0,454	520 077
Duże miasta	1000	0,310	896 839

Definicja 2.5. Średnia winsorowska to średnia arytmetyczna obliczana po zastąpieniu określonej liczby skrajnych wartości najbliższymi wartościami wewnętrznymi:

$$\bar{x}_{\text{wins}} = \frac{1}{n} \left[(k+1)x_{k+1} + \sum_{i=k+2}^{n-k-1} x_i + (k+1)x_{n-k} \right].$$

	k	Udział wydatków na żywność	Całkowite wydatki
Małe miasta	10	0,460	639 137
Duże miasta	10	0,325	1 036 443
Małe miasta	100	0,460	618 994
Duże miasta	100	0,324	1 008 495
Małe miasta	500	0,459	569 349
Duże miasta	500	0,317	953 967
Małe miasta	1000	0,455	526 044
Duże miasta	1000	0,310	906 885

Definicja 2.6. Kwartyle to wartości dzielące uporządkowany zbiór danych na cztery równe części. Pierwszy i trzeci kwartył nazywamy odpowiednio kwartylem dolnym (Q_1) i kwartylem górnym (Q_3). Drugi kwartył to mediana (Q_2), którą obliczamy według wzoru:

$$x_{\text{med}} = \begin{cases} x_{\frac{n+1}{2}}, & \text{n nieparzyste} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}), & \text{n parzyste} \end{cases}.$$

	Kwartył	Udział wydatków na żywność	Całkowite wydatki
Małe miasta	Q1	0,336	297 195
Duże miasta	Q1	0,214	577 090
Małe miasta	Q2	0,454	516 162
Duże miasta	Q2	0,309	889 815
Małe miasta	Q3	0,586	825 578
Duże miasta	Q3	0,419	1 310 004

Definicja 2.7. Moda to wartość lub wartości, które występują najczęściej w zbiorze danych.

W zestawie danych nie występują wartości powtarzające się więcej niż dwukrotnie, dlatego liczenie mody w tym przypadku jest niezasadne i nie wnosi żadnych istotnych informacji. Dane są bardzo zróżnicowane, ponieważ każdy respondent podawał własne, indywidualne wartości dotyczące wydatków, co prowadzi do braku powtarzalności wyników.

Miary położenia - wnioski

Dla wszystkich średnich udział wydatków na żywność jest istotnie mniejszy w dużych miastach niż w małych miastach. Natomiast całkowite wydatki są mniejsze w małych miastach. Wskazuje to na fakt, że mieszkańcy dużych miast, mając większy budżet na wydatki, mniejszą jego część muszą przeznaczyć na jedzenie. Wyraźnie widać to też w porównaniu kwartyli. Zarówno dla udziału wydatków na żywność, jak i całkowitych wydatków średnia ucinana i windsorowska już od małych k są stosunkowo bliskie medianie i wraz ze wzrostem k zbliżają się do mediany w dość powolnym tempie, co oznacza, że pośród obserwacji nie ma zbyt wiele bardzo skrajnych danych, które bardzo znacznie zakłócałyby wynik.

2.2 Miary zmienności

Definicja 2.8. IQR (rozstęp międzykwartyłowy) to różnica pomiędzy trzecim a pierwszym kwartyłem. Informuje o rozproszeniu środkowych 50% danych:

$$IQR = Q_3 - Q_1.$$

	Udział wydatków na żywność	Całkowite wydatki
Małe miasta	0,251	528 382
Duże miasta	0,205	732 914

Definicja 2.9. Rozstęp to różnica między największą a najmniejszą wartością zbioru danych:

$$R = x_{\max} - x_{\min}.$$

	Udział wydatków na żywność	Całkowite wydatki
Małe miasta	0,966	8 297 623
Duże miasta	0,893	7 085 127

Definicja 2.10. Wariancja to średnia kwadratowa odległość wartości od średniej arytmetycznej, mierzy, jak bardzo wartości w zbiorze danych różnią się od średniej:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

	Udział wydatków na żywność	Całkowite wydatki
Małe miasta	0,032	291 699 045 814
Duże miasta	0,022	489 053 842 575

Definicja 2.11. Odchylenie standardowe to pierwiastek z wariancji:

$$s = \sqrt{s^2}.$$

	Udział wydatków na żywność	Całkowite wydatki
Małe miasta	0,179	540 091
Duże miasta	0,150	699 323

Definicja 2.12. Odchylenie przeciętne to średnia wartość bezwzględnych różnic między obserwacjami a ich średnią, mierzy przeciętną odległość wartości od średniej:

$$d = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|.$$

	Udział wydatków na żywność	Całkowite wydatki
Małe miasta	0,146	361 525
Duże miasta	0,120	496 712

Definicja 2.13. Współczynnik zmienności to stosunek odchylenia standardowego do średniej arytmetycznej, wyrażony w procentach:

$$V = \frac{s}{\bar{x}} \cdot 100\%.$$

	Udział wydatków na żywność	Całkowite wydatki
Małe miasta	38,9%	83,9%
Duże miasta	46,0%	67,3%

Miary zmienności - wnioski

Wszystkie nieznormalizowane wartości dla udziału wydatków w żywności są większe dla małych miast niż dla dużych miast, natomiast dla całkowitych wydatków sytuacja jest odwrotna. Jednak znormalizowana miara, jaką jest współczynnik zmienności, pokazuje, że tak naprawdę zmienność udziału wydatków na żywność jest mniejsza w małych miastach, a zmienność wydatków jest w nich większa. Możemy stwierdzić, interpretując ten współczynnik, że zmienność udziału wydatków na żywność jest umiarkowana zarówno dla małych jak i dużych miast, natomiast zmienność całkowitych wydatków jest duża (powyżej 60%) w szczególności w małych miastach.

2.3 Miary kształtu

Definicja 2.14. Skośność to miara asymetrii rozkładu danych względem średniej. Może wskazywać na przewagę wartości mniejszych lub większych od średniej:

$$\alpha = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^3.$$

	Udział wydatków na żywność	Całkowite wydatki
Małe miasta	0,098	3,761
Duże miasta	0,538	2,265

Definicja 2.15. Kurtoza to miara, która opisuje "spiczastość" rozkładu danych. Wskazuje na koncentrację wartości wokół średniej lub w ogonach rozkładu:

$$K = \frac{n-1}{(n-2)(n-3)} \left[(n+1) \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2} - 3(n-1) \right] + 3.$$

	Udział wydatków na żywność	Całkowite wydatki
Małe miasta	2,599	35,290
Duże miasta	3,063	12,924

Miary kształtu - wnioski

Współczynnik skośności dla udziału wydatków na żywność wskazuje na niewielką skośność w prawo dla małych miast i trochę większą skośność w prawo dla dużych miast. Dla całkowitych wydatków współczynnik ten wskazuje na dużą skośność w prawo. Możemy powiedzieć, że rozkłady całkowitych wydatków zarówno w małych, jak i w dużych miastach są ciężkoogonowe. Kurtoza dla udziału wydatków na żywność w dużych, jak i w małych miastach jest zbliżona do 3, czyli rozkłady te mają "spiczastość" zbliżoną do rozkładu normalnego. Dla całkowitych wydatków w dużych, jak i w małych miastach kurtoza jest zdecydowanie większa od 3 co oznacza, że są to rozkłady leptokurtyczne. W szczególności warto zwrócić uwagę na bardzo wysoką wartość dla całkowitych wydatków w małych miastach, co wskazuje, że stosunkowo bardzo duża liczba osób ma bardzo zbliżone całkowite wydatki.

2.4 Miary zależności między zmiennymi

Definicja 2.16. Kowariancja to miara, która mierzy kierunek i siłę współzależności dwóch zmiennych losowych:

$$\text{cov}(X, Y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

Definicja 2.17. Korelacja Pearsona to miara, która mierzy siłę i kierunek liniowej zależności pomiędzy dwiema zmiennymi:

$$r = \frac{\text{cov}(X, Y)}{s_X s_Y}.$$

Definicja 2.18. Korelacja Spearmana to miara, która mierzy monotoniczną zależność pomiędzy dwiema zmiennymi, opiera się na rangach zamiast na wartościach bezwzględnych:

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}.$$

W poniższej tabeli przedstawione zostały kowariancja, korelacja Pearsona i korelacja Spearmana między udziałem wydatków na żywność, a kwotą całkowitych wydatków odpowiednio dla małych i dużych miast.

	Małe miasta	Duże miasta
Kowariancja	0,0321	-54 447
Korelacja Pearsona	-0,500	-0,519
Korelacja Spearmana	-0,551	-0,575

Miary zależności między zmiennymi - wnioski

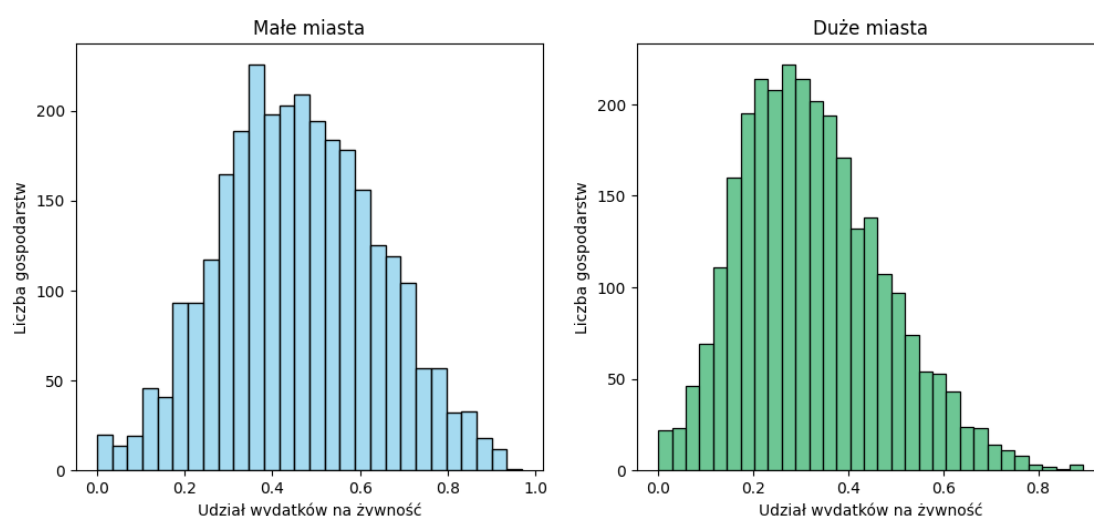
Zarówno współczynnik korelacji Pearsona, jak i współczynnik korelacji Spearmana między udziałem wydatków na żywność a całkowitymi wydatkami są zbliżone do -0,5 zarówno dla małych, jak i dużych miast. Oznacza to, że niezależnie od wielkości miasta korelacja między udziałem wydatków na żywność i całkowitymi wydatkami występuje umiarkowana korelacja ujemna. Oznacza to, że gospodarstwa domowe mające im większe mają całkowite wydatki, tym mniejsza ich część jest wydawana na jedzenie, ale nie jest to zależność wprost proporcjonalna. Możemy domyślać się, że wynika to z faktu, że bogatsi ludzie kupują lepszej jakości, droższe jedzenie, jednak zależność taka występuje tylko w umiarkowanym stopniu.

3 Prezentacja rozkładów na wykresach

3.1 Histogram

Definicja 3.1. Histogram to graficzna reprezentacja rozkładu zmiennej, w której dane grupuje się na przedziały i przedstawia w postaci prostokątów. Wysokość każdego prostokąta odzwierciedla liczbę lub częstość obserwacji przypisanych do danego przedziału.

Histogramy, które przedstawiono na poniższym rysunku, ilustrują rozkład udziału wydatków na żywność w małych i dużych miejscowościach. Analiza ich kształtu pozwala na ocenę symetrii oraz skośności danych.

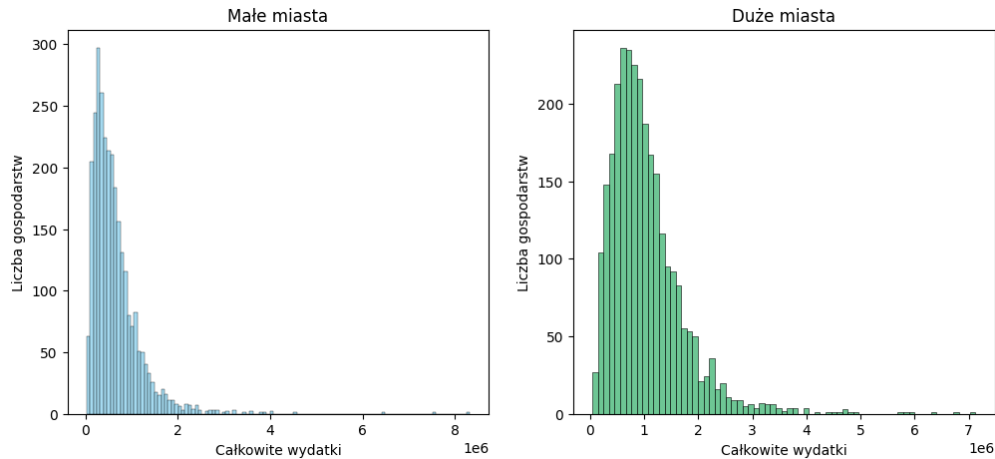


Rysunek 1: Histogram udziału wydatków na żywność

Dla małych miejscowości rozkład udziału wydatków na żywność jest w przybliżeniu symetryczny i dzwonowaty. Najwięcej gospodarstw domowych wydaje około 30%–60% swojego budżetu na jedzenie. Oznacza to, że w mniejszych miejscowościach żywność stanowi znaczącą część domowych wydatków. Występuje niewiele gospodarstw, które wydają bardzo mały lub bardzo duży procent.

Dla dużych miast rozkład jest asymetryczny – przesunięty w lewo. Większość gospodarstw przeznacza na jedzenie około 20%–40% całkowitych wydatków. Oznacza to, że w dużych miastach żywność zajmuje relatywnie mniejszy udział w budżecie, co sugeruje wyższe dochody lub większe wydatki na inne potrzeby (np. usługi, mieszkanie).

Histogramy, które znajdują się na poniższym rysunku, ilustrują rozkład całkowitych wydatków gospodarstw domowych w małych i dużych miejscowościach. Analiza ich kształtu pozwala na ocenę symetrii oraz skośności danych, co jest istotne w kontekście zrozumienia poziomu życia w różnych lokalizacjach.



Rysunek 2: Histogram całkowitych wydatków

W małych miejscowościach rozkład wydatków jest bardzo skośny w prawo (długi ogon). Większość gospodarstw domowych ma niskie całkowite wydatki (poniżej 1 mln jednostek), z jedynie pojedynczymi wyjątkami bardzo wysokich wydatków. Najwięcej osób wydaje około 500 tys. jednostek w skali roku.

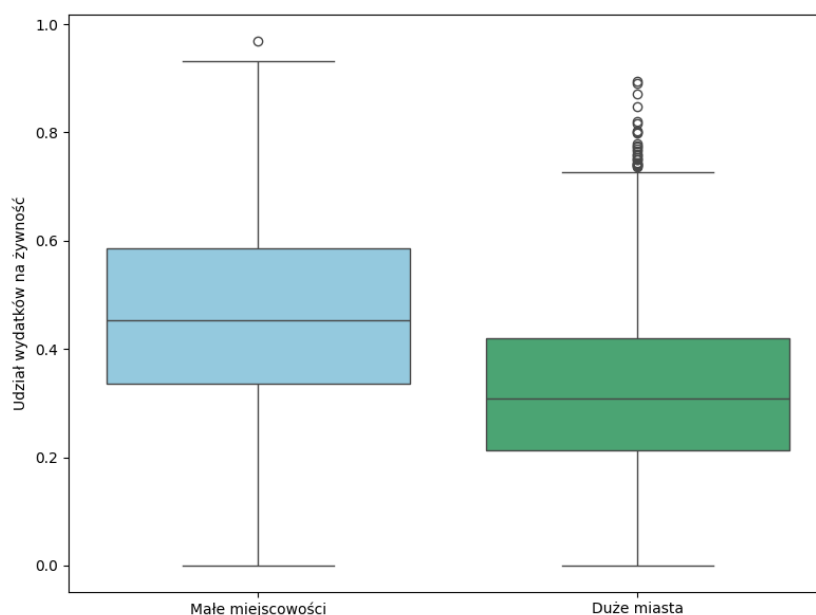
W dużych miastach również można dostrzec skośność w prawo, ale szczyt rozkładu jest przesunięty w stosunku do małych miejscowości – większość gospodarstw ma wydatki w okolicach 600 tys. – 1,2 mln jednostek. Widać, że gospodarstwa w dużych miastach wydają więcej, co jest zgodne z ogólnym trendem wyższych kosztów życia i dochodów w większych aglomeracjach.

Podsumowując, mieszkańcy dużych miast wydają więcej, lecz mniejszą część budżetu przeznaczają na jedzenie. W małych miastach gospodarstwa wydają mniej pieniędzy w ciągu roku, ale jedzenie jest dla nich większym obciążeniem.

3.2 Boxplot

Definicja 3.2. Wykresu pudełkowego (boxplot) to graficzna reprezentacja rozkładu danych, w której przedstawione są kluczowe statystyki: mediany, kwartyle oraz wartości odstające, umożliwiającą porównanie rozkładów między różnymi grupami.

Aby uzyskać pełniejszy obraz rozkładu procentowego udziału wydatków na żywność w różnych grupach gospodarstw domowych, przedstawiliśmy dane za pomocą wykresu pudełkowego.

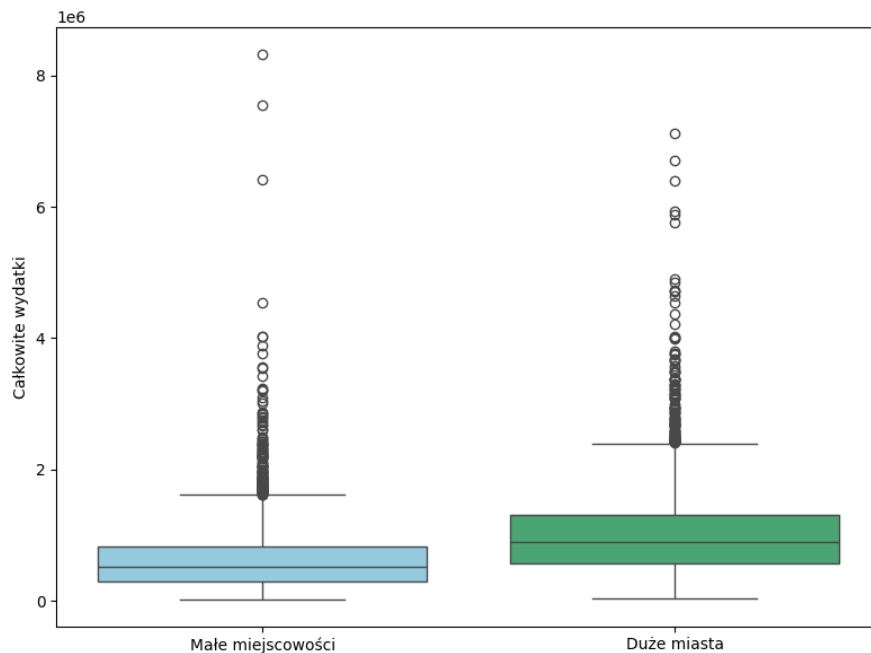


Rysunek 3: Boxplot udziału wydatków na żywność

Na ich podstawie zauważyć można istotne różnice pomiędzy gospodarstwami domowymi w małych miejscowościach oraz dużych miastach. Mediana procentowego udziału wydatków na żywność w przypadku małych miejscowości jest wyższa i wynosi około 0,45, podczas gdy dla dużych miast kształtuje się na niższym poziomie, około 0,30. Potwierdza to, że gospodarstwa w mniejszych miejscowościach przeznaczają relatywnie większą część swoich całkowitych wydatków na żywność.

Ponadto rozstęp międzykwartyłowy w przypadku małych miejscowości jest szerszy, co świadczy o większej zmienności w strukturze wydatków w tej grupie. W przypadku dużych miast zmienność ta jest mniejsza, co oznacza, że gospodarstwa w tej grupie są bardziej jednorodne pod względem udziału wydatków na żywność. Dodatkowo w obu grupach można zaobserwować wartości odstające, jednak wśród gospodarstw z dużych miast ich liczba jest wyraźnie większa.

Dodatkowo, aby uzyskać pełniejszy obraz rozkładu całkowitych wydatków w różnych grupach gospodarstw domowych, przedstawiliśmy dane za pomocą wykresu pudełkowego.



Rysunek 4: Boxplot całkowitych wydatków

Ponownie obserwujemy wyraźną różnicę między medianą całkowitych wydatków w analizowanych grupach. W małych miejscowościach mediana wynosi około 500 tys. jednostek, natomiast w dużych miastach osiąga poziom około 1 mln jednostek. Świadczy to o tym, że przeciętne gospodarstwo domowe w dużym mieście wydaje znacznie więcej niż to w mniejszej miejscowości.

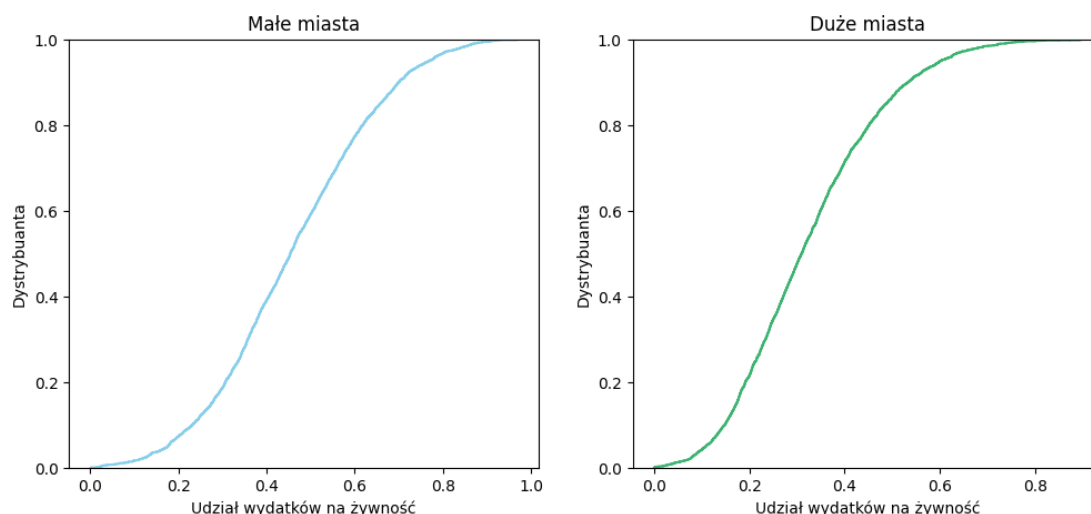
Rozstęp międzykwartyłowy jest w tym wypadku szerszy dla dużych miast, co wskazuje na większą różnorodność w całkowitych wydatkach w tej grupie. W dużych miastach zapotrzebowanie na różne dobra i usługi jest bardziej zróżnicowane, co może być wynikiem różnych poziomów dochodów oraz stylu życia mieszkańców.

3.3 Dystrybuanta

Definicja 3.3. Dystrybuanta empiryczna to funkcja, która dla każdej liczby rzeczywistej x określa, jaki odsetek obserwacji w próbie przyjmuje wartość mniejszą lub równą x . Formalnie dana jest wzorem:

$$F(x) = \frac{\#\{x_i \leq x\}}{n}.$$

Aby lepiej zrozumieć rozkład procentu wydatków na żywność w małych i dużych miejscowościach, przedstawiliśmy poniżej dystrybuanty dla obu grup gospodarstw domowych.



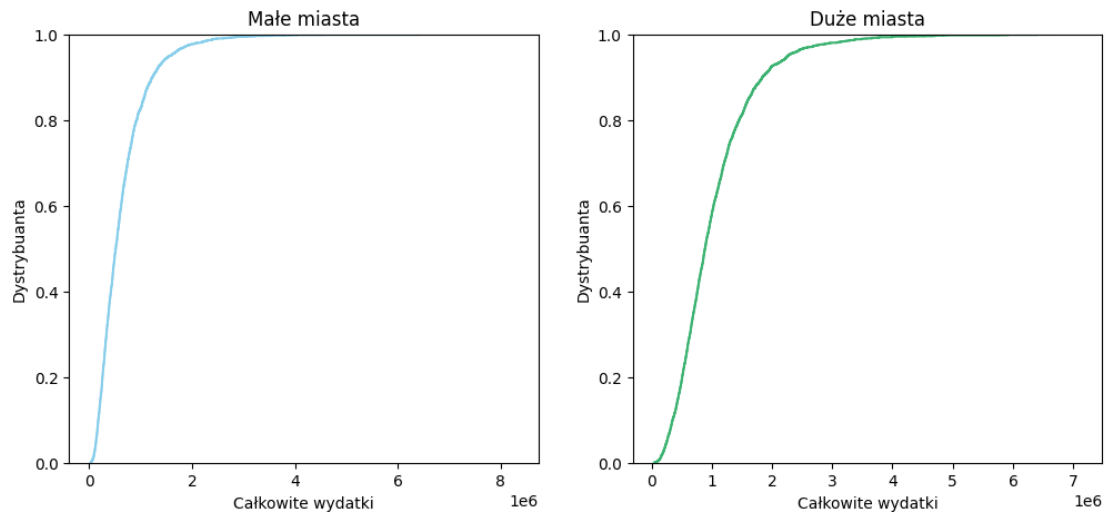
Rysunek 5: Dystrybuanta empiryczna udziału wydatków na żywność

Dla małych miejscowości dystrybuanta rośnie stosunkowo równomiernie, co potwierdza symetryczny rozkład udziału wydatków na żywność. Wartość dystrybuanty osiąga 50% dla około 45% wydatków, co jest zgodne z wcześniejszymi obserwacjami dotyczącymi mediany. Oznacza to, że połowa gospodarstw w małych miejscowościach przeznaczą na żywność nie więcej niż 45% swojego budżetu.

W przypadku dużych miast dystrybuanta rośnie szybciej dla niższych wartości, co odzwierciedla lewostronną asymetrię rozkładu. Wartość dystrybuanty osiąga 50% dla około 30% wydatków, co potwierdza, że mediana w tej grupie jest niższa niż w małych miejscowościach. Ponadto, powolny wzrost dystrybuanty dla wyższych wartości wskazuje na obecność gospodarstw, które przeznaczają na żywność relatywnie większą część budżetu, choć jest ich stosunkowo niewiele.

Podsumowując, analiza tych dystrybuant potwierdza istotne różnice w strukturze wydatków między małymi i dużymi miejscowościami.

Uzupełnieniem analizy jest wykres dystrybuant empirycznych, które ukazują, jaka część gospodarstw nie przekracza danego poziomu całkowitych wydatków.



Rysunek 6: Dystrybuanta empiryczna całkowitych wydatków

Widzimy, że krzywa dla małych miejscowości szybciej osiąga wartość bliską 1, co oznacza, że większość gospodarstw osiąga stosunkowo niskie wydatki. W przypadku dużych miast dystrybuanta rośnie wolniej – wskazując, że większy odsetek gospodarstw ponosi wyższe wydatki. Różnice w nachyleniu krzywych potwierdzają większą różnorodność w dużych miastach oraz przesunięcie całego rozkładu w kierunku wyższych wartości.

4 Podsumowanie

Z analizy porównawczej udziału wydatków na żywność i całkowitych wydatków gospodarstw domowych w dużych i małych miastach możemy wywnioskować, że:

- Średni udział wydatków na żywność jest znacząco większy w małych miastach niż w dużych miastach (średnie arytmetyczne to odpowiednio 0,460 i 0,326). Natomiast średnie wydatki całkowite są większe w dużych miastach (średnie arytmetyczne to 643 674 dla małych i 1 038 887 dla dużych miast).
- Udział wydatków na żywność niezależnie od wielkości miasta charakteryzuje się umiarkowaną zmiennością, natomiast zmienność całkowitych wydatków jest duża w szczególności w małych miastach. Wskazuje to na fakt, że w małych miastach mieszka wielu ludzi niezamożnych, ale także wielu ludzi bogatych.
- Niezależnie od wielkości miasta udział wydatków na żywność charakteryzuje się małą skośnością i jest mezokurtyczny. Natomiast rozkłady całkowitych wydatków są ciężkoogonowe i prawostronnie skośne oraz leptokurtyczne w szczególności w małych miastach.
- Zarówno w małych i w dużych miastach występuje umiarkowana korelacja ujemna między udziałem wydatków na żywność a całkowitymi wydatkami.
- Ogólnie można stwierdzić, że pod badanymi względami duże miasta są bardziej zróżnicowane od małych miast, jednak wiele osób bogatych (z dużymi wydatkami) postanawia mieszkać w małych miastach.