



# ***Wykorzystanie imputacji masowej w badaniu przedsiębiorstw***

**Paweł Lańduch**

*Ośrodek Statystyki Krótkookresowej Dział Metodologii i Programowania*

## Uzasadnienie podjęcia problemu

- ❑ Rosnące wskaźniki braków odpowiedzi w badaniach.
- ❑ Prace dotyczące metod estymacji na ogół dotyczą badań społecznych.
- ❑ Rosnąca ilość zbiorów danych, które można by wykorzystać w badaniach statystycznych.

## Definicje ogólne

**Imputacja masowa** – imputacja, w której uzupełnia się stosunkowo dużą ilość braków danych.

W tym wypadku zastosowano ją jako metodę integracji.

Wykorzystywane modele braków danych:

**MAR** - (ang. missing at random) rodzaj braków danych, które można zignorować przy estymacji.

**MNAR** – (ang. missing not at random) – rodzaj braków danych, które nie mogą zostać zignorowane przy estymacji.

## Wybrane tło literaturowe dotyczące wykorzystania źródeł nielosowych i badań przedsiębiorstw

- Yang, S., Kim, J. K., Hwang, Y., (2021). Integration of data from probability surveys and big found data for finite population inference using mass imputation, *Survey Methodology*, Vol. 47, No. 1, pp. 29-58, *Statistics Canada, Catalogue No. 12-001-X*. – teoretyczne podstawy zastosowania imputacji masowej jako integracji źródeł danych (losowego i nielosowego),
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581-592.– mechanizmy braków danych,
- Cox, B. G., & Chinnappa, B. N. (1995). Unique features of business surveys. *Business survey methods*, 1-17.
- Riviére, P. (2002). What makes business statistics special ?. *International Statistical Review*, 70(1), 145-159. – specyfika badań przedsiębiorstw,
- Meng, X. L. (2018). Statistical paradises and paradoxes in big data (i) law of large populations, big data paradox, and the 2016 us presidential election. *The Annals of Applied Statistics*, 12(2), 685-726.– ograniczenia dot. nielosowych źródeł danych.

# Procedura badawcza

## ❑ Zakres badania:

- Integracja dwóch źródeł danych za pomocą imputacji masowej. Jest to ten rodzaj integracji, w którym łączy się próbę losową i nielosową za pomocą wspólnego wektora zmiennych. Imputacja masowa została potraktowana jako przypadek integracji danych, w którym analizowana zmienna jest obserwowana w próbie nielosowej, a do próby losowej jej wartości są imputowane. W próbie losowej wartości dla analizowanej zmiennej w ogóle nie występują.
- Wykonane czynności:
  - ✓ Analizowano trzy metody wykorzystujące imputację masową.
  - ✓ Uzyskane rezultaty porównano z efektami zastosowania innego estymatora opartego na modelu.
  - ✓ Dla każdej z użytych metod zrealizowano stosowny eksperyment symulacyjny.
  - ✓ Wyznaczono odpowiednie mierniki jakościowe.

# Procedura badawcza

## ❑ Źródło danych:

- Wygenerowano sztuczną populację przedsiębiorstw.
- Wzorcową strukturą do jej wygenerowania był operat ciągłego badania statystycznego prowadzonego od wielu lat przez Główny Urząd Statystyczny – Meldunek o działalności gospodarczej o symbolu DG-1, przeprowadzanego w cyklu miesięcznym dla sektora przedsiębiorstw (wybrano operat i zbiór badania dla lipca 2022 r.).
- Analizę ograniczono do przemysłu (sekcje: B, C, D, E klasyfikacji PKD 2007)
- Do modelowania wybrano dwie zmienne z formularza badania:
  - Sw\_1b – Przychody netto ze sprzedaży produktów (wyrobów i usług własnej produkcji) (wiersz pierwszy formularza DG-1),
  - Wb\_1b – Wynagrodzenia brutto osób wykazanych w wierszu 07 (tzn. dla przeciętnej liczby zatrudnionych) w tys. zł (wiersz dziewiąty formularza DG-1).
- Jako zmiennej objaśniającej użyto zmiennej lps – liczba pracujących.

# Procedura badawcza

## □ Źródło danych:

### ▪ I etap :

- $\log(Y_{h,i}) = \beta_{0,h} + \beta_{1,h}\log(X_{h,i}) + \varepsilon_{hi},$
- $\log(\text{sw\_}1b_{h,i}) = \beta_{0,h} + \beta_{1,h}\log(\text{lp}s_{h,i}) + \varepsilon_{h,i},$
- $\log(\text{wb\_}1b_{h,i}) = \beta_{0,h} + \beta_{1,h}\log(\text{lp}s_{h,i}) + \varepsilon_{h,i},$
- gdzie  $h = 1, \dots, 25$ , są to działy,  $i = 1, 2, \dots, n_h$ , przy czym  $n_h$  jest liczbą jednostek w warstwie  $h$ ,  $h = 1, \dots, 25$ , a  $\varepsilon_{h,i} \sim N(0, 1)$ .

### ▪ II etap:

- $\exp(\text{wartość z modelu} + \varepsilon_{h,i}), \varepsilon_{h,i} \sim N(0, 0,5)$

# Procedura badawcza

## ❑ Wykonanie badania symulacyjnego:

1. Wybranie z populacji próby losowej według schematu losowania warstwowego. Z uwagi na zróżnicowanie jednostek wykorzystano losowanie systematyczne z prawdopodobieństwem proporcjonalnym do wielkości jednostki, bez zwracania.
2. Wygenerowanie próby nielosowej. W generowanej populacji jednostki średnie to takie gdzie  $lps \leq 250$  oraz duże -  $lps > 250$ .

| L.p. | Wariant łączenia zbiorów                                                                                                        |
|------|---------------------------------------------------------------------------------------------------------------------------------|
| 1    | Bez podziału na klasy wielkości, w próbie nielosowej jest 80% jednostek w populacji, z zastosowaniem mechanizmu MAR (losowy)    |
| 2    | Z podziałem na klasy wielkości, w próbie nielosowej jest 60% jednostek w populacji, z zastosowaniem mechanizmu MAR (losowy)     |
| 3    | Z podziałem na klasy wielkości, w próbie nielosowej jest 50% jednostek w populacji, z zastosowaniem mechanizmu MAR (losowy)     |
| 4    | Z podziałem na klasy wielkości, w próbie nielosowej jest 60% jednostek w populacji, z zastosowaniem mechanizmu NMAR (nielosowy) |
| 5    | Z podziałem na klasy wielkości, w próbie nielosowej jest 50% jednostek w populacji, z zastosowaniem mechanizmu NMAR (nielosowy) |

# Procedura badawcza

## ❑ Wykonanie badania symulacyjnego (cd.):

3. Estymacja wartości globalnych w oparciu o próbę losową.
4. Wyznaczenie parametrów jakościowych. W tym celu użyto następujących mierników:
  - a) przybliżona symulacyjnie wartość względnego obciążenia danego estymatora: (ang. *Relative Bias*):

$$RB^{(sym)}(\hat{\theta}) = \frac{\frac{1}{B} \sum_{b=1}^B \hat{\theta}^{(b)} - \theta}{|\theta|} 100 \%,$$

- gdzie:

$B$  – liczba powtórzeń w badaniu symulacyjnym,

$\hat{\theta}^{(b)}$  – wartość estymatora wartości globalnej w b-tym kroku badania symulacyjnego,

$\theta$  – szacowany parametr,

# Procedura badawcza

## ❑ Wykonanie badania symulacyjnego (cd.):

4. Wyznaczenie parametrów jakościowych. W tym celu użyto mierników:

- b) przybliżona symulacyjnie wartość względnego błędu RMSE estymatora (ang. *relative mean squared error*):

$$\bullet RRMSE^{(sym)}(\hat{\theta}) = \frac{\sqrt{\frac{1}{B} \sum_{b=1}^B (\hat{\theta}^{(b)} - \theta)^2}}{|\theta|} 100 \text{ \%}.$$

# Procedura badawcza

- ❑ Wykonanie badania symulacyjnego (dok.):
  - ❑ Zastosowane estymatory wartości globalnej w podpopulacjach wykorzystujące estymację masową:
    - ❑ Imputacja oparta na najbliższym sąsiedzie wybranym losowo, w przypadku kiedy liczba jednostek najbliższych jest większa niż 1.
    - ❑ Imputacja średnią oparta na k najbliższych sąsiadach.
    - ❑ Imputacja wykorzystująca regresję liniową.
  - ❑ Dla celów porównawczych wykorzystano:
    - ❑ Estymator Horvitz-Thompsona dla podpopulacji.
    - ❑ Estymator podejścia modelowego typu BLU.
- Dla każdej z zastosowanych metod wykonano 1000 powtórzeń w badaniu symulacyjnym.

**Wartości przybliżonego symulacyjnie obciążenia analizowanych estymatorów w przypadku, gdy w populacji (źródle nielosowym) występują braki danych typu losowego w wysokości 40% jednostek średnich – wynagrodzenia brutto.**

| Dział PKD                                                                           | Zmienne obserwowane w próbie losowej | Metoda estymacji oparta na estymatorze typu BLU | Metoda imputacji średnią opartą na $k$ -najbliższych sąsiadach | Metoda imputacji masowej oparta na najbliższym sąsiedzie wybranym losowo | Metoda imputacji masowej oparta na modelu regresji liniowej |
|-------------------------------------------------------------------------------------|--------------------------------------|-------------------------------------------------|----------------------------------------------------------------|--------------------------------------------------------------------------|-------------------------------------------------------------|
| 5 górnictwo i wydobywanie                                                           | 0,6                                  | 2,9                                             | 0,6                                                            | 0,6                                                                      | 2,4                                                         |
| 10 produkcja artykułów spożywczych                                                  | 0,1                                  | 0,2                                             | -0,2                                                           | -0,2                                                                     | -10,3                                                       |
| 11 produkcja napojów                                                                | 0,1                                  | -10                                             | 0,1                                                            | 0,2                                                                      | -6,0                                                        |
| 12 produkcja wyrobów tytoniowych                                                    | -0,3                                 | 24,7                                            | -0,7                                                           | -0,7                                                                     | -0,9                                                        |
| 13 produkcja wyrobów tekstylnych                                                    | 0,2                                  | 2,1                                             | 0,0                                                            | 0,1                                                                      | -14,4                                                       |
| 30 produkcja pozostałego sprzętu transportowego                                     | -0,2                                 | -8,7                                            | 0,1                                                            | 0,2                                                                      | -7,7                                                        |
| 31 produkcja mebli                                                                  | 0,0                                  | 2,1                                             | -0,8                                                           | -0,8                                                                     | -8,3                                                        |
| 35 wytwarzanie i zaopatrywanie w energię elektryczną, gaz, parę wodną i gorącą wodę | 0,1                                  | -7,1                                            | -0,2                                                           | -0,2                                                                     | -5,6                                                        |
| 36 pobór, uzdatnianie i dostarczanie wody                                           | 0,1                                  | -10,4                                           | -0,5                                                           | -0,6                                                                     | -14,4                                                       |
| 38 gospodarka odpadami; odzysk surowców                                             | 0,1                                  | -12,2                                           | 2,5                                                            | 2,3                                                                      | -12,3                                                       |

**Wartości przybliżonego symulacyjnie obciążenia analizowanych estymatorów w przypadku, gdy w populacji (źródle nielosowym) występują braki danych typu nielosowego w wysokości 40% jednostek średnich – wynagrodzenia brutto.**

| Dział PKD                                                                           | Zmienne obserwowane w próbie losowej | Metoda estymacji oparta na estymatorze typu BLU | Metoda imputacji średnią opartą na $k$ -najbliższych sąsiadach | Metoda imputacji masowej oparta na najbliższym sądzie wybranym losowo | Metoda imputacji masowej oparta na modelu regresji liniowej |
|-------------------------------------------------------------------------------------|--------------------------------------|-------------------------------------------------|----------------------------------------------------------------|-----------------------------------------------------------------------|-------------------------------------------------------------|
| 5 górnictwo i wydobywanie                                                           | 0,8                                  | 2,9                                             | 0,8                                                            | 0,8                                                                   | 2,1                                                         |
| 10 produkcja artykułów spożywczych                                                  | 0,0                                  | 0,2                                             | -1,0                                                           | -1,0                                                                  | -10,9                                                       |
| 11 produkcja napojów                                                                | 0,0                                  | -10,0                                           | 0,2                                                            | 0,3                                                                   | -5,1                                                        |
| 12 produkcja wyrobów tytoniowych                                                    | 0,5                                  | 24,9                                            | 0,0                                                            | 0,0                                                                   | -2,1                                                        |
| 13 produkcja wyrobów tekstylnych                                                    | 0,2                                  | 2,1                                             | -2,1                                                           | -2,0                                                                  | -15,8                                                       |
| 30 produkcja pozostałego sprzętu transportowego                                     | 0,2                                  | -8,5                                            | -2,4                                                           | -2,3                                                                  | -9,8                                                        |
| 31 produkcja mebli                                                                  | 0,0                                  | 2,1                                             | -1,7                                                           | -1,6                                                                  | -9,7                                                        |
| 35 wytwarzanie i zaopatrywanie w energię elektryczną, gaz, parę wodną i gorącą wodę | 0,0                                  | -7,1                                            | -1,1                                                           | -1,1                                                                  | -8,2                                                        |
| 36 pobór, uzdatnianie i dostarczanie wody                                           | 0,0                                  | -10,4                                           | -3,5                                                           | -3,5                                                                  | -16,9                                                       |
| 38 gospodarka odpadami; odzysk surowców                                             | 0,1                                  | -12,2                                           | -1,8                                                           | -1,8                                                                  | -13,7                                                       |

**Wartości przybliżonego symulacyjnie obciążenia analizowanych estymatorów w przypadku, gdy w populacji (źródle nielosowym) występują braki danych typu losowego w wysokości 50% jednostek średnich – wynagrodzenia brutto.**

| Dział PKD                                                                           | Zmienne obserwowane w próbie losowej | Metoda estymacji oparta na estymatorze typu BLU | Metoda imputacji średnią opartą na $k$ -najbliższych sąsiadach | Metoda imputacji masowej oparta na najbliższym sąsiedzie wybranym losowo | Metoda imputacji masowej oparta na modelu regresji liniowej |
|-------------------------------------------------------------------------------------|--------------------------------------|-------------------------------------------------|----------------------------------------------------------------|--------------------------------------------------------------------------|-------------------------------------------------------------|
| 5 górnictwo i wydobywanie                                                           | -0,1                                 | 3,0                                             | -0,1                                                           | -0,1                                                                     | 1,7                                                         |
| 10 produkcja artykułów spożywczych                                                  | 0,0                                  | 0,2                                             | -0,3                                                           | -0,3                                                                     | -10,2                                                       |
| 11 produkcja napojów                                                                | -0,2                                 | -10                                             | 0,0                                                            | 0,1                                                                      | -6,4                                                        |
| 12 produkcja wyrobów tytoniowych                                                    | -0,2                                 | 24,8                                            | -0,9                                                           | -0,9                                                                     | -1,7                                                        |
| 13 produkcja wyrobów tekstylnych                                                    | 0,1                                  | 2,1                                             | -0,5                                                           | -0,5                                                                     | -14,8                                                       |
| 30 produkcja pozostałego sprzętu transportowego                                     | -0,2                                 | -8,6                                            | -0,5                                                           | -0,5                                                                     | -7,4                                                        |
| 31 produkcja mebli                                                                  | 0,1                                  | 2,2                                             | -1,0                                                           | -1,0                                                                     | -8,5                                                        |
| 35 wytwarzanie i zaopatrywanie w energię elektryczną, gaz, parę wodną i gorącą wodę | 0,0                                  | -7,1                                            | -0,3                                                           | -0,3                                                                     | -5,2                                                        |
| 36 pobór, uzdatnianie i dostarczanie wody                                           | -0,2                                 | -10,5                                           | -0,4                                                           | -0,4                                                                     | -14,0                                                       |
| 38 gospodarka odpadami; odzysk surowców                                             | -0,1                                 | -12,3                                           | 3,7                                                            | 3,6                                                                      | -12,6                                                       |

**Wartości przybliżonego symulacyjnie obciążenia analizowanych estymatorów w przypadku, gdy w populacji (źródło nielosowym) występują braki danych typu nielosowego w wysokości 50% jednostek średnich – wynagrodzenia brutto.**

| Dział PKD                                                                           | Zmienne obserwowane w próbie losowej | Metoda estymacji oparta na estymatorze typu BLU | Metoda imputacji średnią opartą na $k$ -najbliższych sąsiadach | Metoda imputacji masowej oparta na najbliższym sąsiedzie wybranym losowo | Metoda imputacji masowej oparta na modelu regresji liniowej |
|-------------------------------------------------------------------------------------|--------------------------------------|-------------------------------------------------|----------------------------------------------------------------|--------------------------------------------------------------------------|-------------------------------------------------------------|
| 5 górnictwo i wydobywanie                                                           | 0,2                                  | 3,0                                             | 0,0                                                            | 0,0                                                                      | 1,8                                                         |
| 10 produkcja artykułów spożywczych                                                  | 0,0                                  | 0,2                                             | -1,2                                                           | -1,1                                                                     | -10,9                                                       |
| 11 produkcja napojów                                                                | -0,2                                 | -10,0                                           | 0,0                                                            | 0,2                                                                      | -5,6                                                        |
| 12 produkcja wyrobów tytoniowych                                                    | -0,5                                 | 24,7                                            | -1,2                                                           | -1,2                                                                     | -2,8                                                        |
| 13 produkcja wyrobów tekstylnych                                                    | 0,2                                  | 2,1                                             | -2,3                                                           | -2,2                                                                     | -15,9                                                       |
| 30 produkcja pozostałego sprzętu transportowego                                     | 0,1                                  | -8,5                                            | -2,4                                                           | -2,4                                                                     | -9,2                                                        |
| 31 produkcja mebli                                                                  | -0,1                                 | 2,1                                             | -1,9                                                           | -1,9                                                                     | -9,8                                                        |
| 35 wytwarzanie i zaopatrywanie w energię elektryczną, gaz, parę wodną i gorącą wodę | 0,0                                  | -7,1                                            | -1,2                                                           | -1,2                                                                     | -7,6                                                        |
| 36 pobór, uzdatnianie i dostarczanie wody                                           | -0,1                                 | -10,4                                           | -3,7                                                           | -3,7                                                                     | -16,4                                                       |
| 38 gospodarka odpadami; odzysk surowców                                             | -0,1                                 | -12,3                                           | -2,0                                                           | -2,0                                                                     | -14,1                                                       |

**Wartości przybliżonego symulacyjnie obciążenia analizowanych estymatorów w przypadku, gdy w populacji (źródle nielosowym) występują braki danych typu losowego w wysokości 20%, ale bez podziału na wielkość jednostki. – wynagrodzenia brutto.**

| Dział PKD                                                                           | Zmienne obserwowane w próbie losowej | Metoda estymacji oparta na estymatorze typu BLU | Metoda imputacji średnią opartą na $k$ -najbliższych sąsiadach | Metoda imputacji masowej oparta na najbliższym sąsiedzie wybranym losowo | Metoda imputacji masowej oparta na modelu regresji liniowej |
|-------------------------------------------------------------------------------------|--------------------------------------|-------------------------------------------------|----------------------------------------------------------------|--------------------------------------------------------------------------|-------------------------------------------------------------|
| 5 górnictwo i wydobywanie                                                           | -1,1                                 | 2,5                                             | -19,8                                                          | -19,8                                                                    | -19,6                                                       |
| 10 produkcja artykułów spożywczych                                                  | 0,0                                  | 0,2                                             | -12,2                                                          | -12,1                                                                    | -12,8                                                       |
| 11 produkcja napojów                                                                | -0,1                                 | -10                                             | -16,5                                                          | -16,5                                                                    | -16,2                                                       |
| 12 produkcja wyrobów tytoniowych                                                    | 1,0                                  | 24,9                                            | -11,8                                                          | -11,8                                                                    | -12,9                                                       |
| 13 produkcja wyrobów tekstylnych                                                    | 0,0                                  | 2,0                                             | -20,8                                                          | -20,8                                                                    | -21,2                                                       |
| 30 produkcja pozostałego sprzętu transportowego                                     | -0,1                                 | -8,6                                            | -9,2                                                           | -9,2                                                                     | -8,6                                                        |
| 31 produkcja mebli                                                                  | 0,1                                  | 2,1                                             | -16,1                                                          | -16,0                                                                    | -15,3                                                       |
| 35 wytwarzanie i zaopatrywanie w energię elektryczną, gaz, parę wodną i gorącą wodę | -0,1                                 | -7,1                                            | -33,3                                                          | -33,2                                                                    | -30,3                                                       |
| 36 pobór, uzdatnianie i dostarczanie wody                                           | 0,0                                  | -10,4                                           | -19,9                                                          | -20,0                                                                    | -19,7                                                       |
| 38 gospodarka odpadami; odzysk surowców                                             | 0,0                                  | -12,3                                           | -8,4                                                           | -8,3                                                                     | -8,2                                                        |

**Wartości przybliżonego symulacyjnie względnego RMSE analizowanych estymatorów w przypadku, gdy w populacji (źródle nielosowym) występują braki danych typu losowego w wysokości 40% jednostek średnich – wynagrodzenia brutto.**

| Dział PKD                                                                           | Zmienne obserwowane w próbie losowej | Metoda estymacji oparta na estymatorze typu BLU | Metoda imputacji średnią opartą na $k$ -najbliższych sąsiadach | Metoda imputacji masowej oparta na najbliższym sąsiedzie wybranym losowo | Metoda imputacji masowej oparta na modelu regresji liniowej |
|-------------------------------------------------------------------------------------|--------------------------------------|-------------------------------------------------|----------------------------------------------------------------|--------------------------------------------------------------------------|-------------------------------------------------------------|
| 5 górnictwo i wydobywanie                                                           | <b>19,0</b>                          | 6,1                                             | <b>19,0</b>                                                    | <b>19,0</b>                                                              | 6,0                                                         |
| 10 produkcja artykułów spożywczych                                                  | 1,5                                  | 0,6                                             | 1,9                                                            | 2,2                                                                      | 10,4                                                        |
| 11 produkcja napojów                                                                | 7,3                                  | 10,1                                            | 7,9                                                            | 8,5                                                                      | 6,5                                                         |
| 12 produkcja wyrobów tytoniowych                                                    | <b>12,8</b>                          | 25,3                                            | <b>12,7</b>                                                    | <b>12,7</b>                                                              | 4,0                                                         |
| 13 produkcja wyrobów tekstylnych                                                    | 4,0                                  | 2,5                                             | 6,1                                                            | 6,4                                                                      | 14,6                                                        |
| 30 produkcja pozostałego sprzętu transportowego                                     | 7,9                                  | 8,9                                             | 6,9                                                            | 7,3                                                                      | 7,9                                                         |
| 31 produkcja mebli                                                                  | 1,8                                  | 2,3                                             | 2,9                                                            | 3,2                                                                      | 8,4                                                         |
| 35 wytwarzanie i zaopatrywanie w energię elektryczną, gaz, parę wodną i gorącą wodę | 1,6                                  | 7,1                                             | 2,1                                                            | 2,2                                                                      | 6,0                                                         |
| 36 pobór, uzdatnianie i dostarczanie wody                                           | 5,0                                  | 10,5                                            | 7,1                                                            | 8,2                                                                      | 14,6                                                        |
| 38 gospodarka odpadami; odzysk surowców                                             | 4,0                                  | 12,3                                            | 5,6                                                            | 6,2                                                                      | 12,4                                                        |

**Wartości przybliżonego symulacyjnie względnego RMSE analizowanych estymatorów w przypadku, gdy w populacji (źródle nielosowym) występują braki danych typu losowego w wysokości 20%, ale bez podziału na wielkość jednostki – wynagrodzenia brutto.**

| Dział PKD                                                                           | Zmienne obserwowane w próbie losowej | Metoda estymacji oparta na estymatorze typu BLU | Metoda imputacji średnią opartą na $k$ -najbliższych sąsiadach | Metoda imputacji masowej oparta na najbliższym sąsiedzie wybranym losowo | Metoda imputacji masowej oparta na modelu regresji liniowej |
|-------------------------------------------------------------------------------------|--------------------------------------|-------------------------------------------------|----------------------------------------------------------------|--------------------------------------------------------------------------|-------------------------------------------------------------|
| 5 górnictwo i wydobywanie                                                           | <b>17,3</b>                          | 5,1                                             | <b>40,3</b>                                                    | <b>40,3</b>                                                              | <b>40</b>                                                   |
| 10 produkcja artykułów spożywczych                                                  | 1,6                                  | 0,6                                             | 13,7                                                           | 13,8                                                                     | 13,9                                                        |
| 11 produkcja napojów                                                                | 7,2                                  | 10,1                                            | 20,5                                                           | 20,8                                                                     | 19,2                                                        |
| 12 produkcja wyrobów tytoniowych                                                    | <b>13,8</b>                          | 25,6                                            | <b>26,6</b>                                                    | <b>26,6</b>                                                              | <b>24,4</b>                                                 |
| 13 produkcja wyrobów tekstylnych                                                    | 4,1                                  | 2,5                                             | 21,5                                                           | 21,6                                                                     | 21,6                                                        |
| 30 produkcja pozostałego sprzętu transportowego                                     | 8,0                                  | 8,9                                             | 14,7                                                           | 14,7                                                                     | 12,2                                                        |
| 31 produkcja mebli                                                                  | 1,8                                  | 2,3                                             | 16,8                                                           | 16,8                                                                     | 15,7                                                        |
| 35 wytwarzanie i zaopatrywanie w energię elektryczną, gaz, parę wodną i gorącą wodę | <b>1,7</b>                           | 7,1                                             | <b>36,6</b>                                                    | <b>36,6</b>                                                              | 34,2                                                        |
| 36 pobór, uzdatnianie i dostarczanie wody                                           | 4,6                                  | 10,4                                            | 21,4                                                           | 21,8                                                                     | 20,6                                                        |
| 38 gospodarka odpadami; odzysk surowców                                             | 3,9                                  | 12,3                                            | 10,0                                                           | 10,4                                                                     | 9,3                                                         |

## Wnioski dotyczące wykorzystania możliwości wykorzystania imputacji masowej

1. Imputacja masowa jako metoda integracji danych, w ujęciu łączenia próby losowej i nielosowej, może stać się użytecznym narzędziem – wspomagającym lub podstawowym – służącym uzyskaniu szacunków porównywalnej jakości w porównaniu z badaniami przedsiębiorstw prowadzonymi w oparciu o losową i warstwową próbę, kiedy zakładamy, że wartości analizowanych cech są obserwowane tylko w próbie nielosowej i stamtąd mogą być zaczerpnięte.
2. Wskaźniki jakościowe wyznaczone metodą symulacyjną dla obu wersji generowania zbioru nielosowego, wskazują na akceptowalne poziomy precyzji wyników z punktu widzenia metodologii prowadzenia reprezentacyjnych badań statystycznych.
3. Wielkość jednostki jest kluczowa w estymacji wartości globalnych w statystyce przedsiębiorstw, a także ma ogromny wpływ na dokładność szacunków i wykorzystanie estymatorów pośrednich.

## Wnioski dotyczące dalszych prac

1. Istnieje pole do ewentualnych dalszych badań symulacyjnych, których celem będzie użycie estymatorów opartych na modelach, nawet jeśli w źródłach danych będzie występować problem selekcji.
2. W pracach nad wykorzystaniem rejestrów administracyjnych w badaniach statystycznych należałoby włączyć metody oparte na imputacji masowej.
3. Należałoby poświęcić również uwagę samym metodom wykorzystującym imputację masową pod kątem estymacji wyników na niskich poziomach agregacji wzorując się na standardach wyznaczanych np. przez statystykę małych obszarów.
4. Analizę można by również rozszerzyć o możliwość zastosowania estymatorów wykorzystujących estymację masową w innych rodzajach działalności.
5. Zastosowanie metod wykorzystujących imputację masową mogłoby wzbogacić estymację wyników o tzw. jednostki małe i mikro-firmy, których statystyce krótkookresowej dotyczącej sektora przedsiębiorstw się nie uwzględnia.
6. Analizę należałoby rozszerzyć uwzględniając kontekst badań longitudinalnych.



Dziękuję za uwagę!