

# Zastosowanie metod taksonomicznych w estymacji wskaźników ubóstwa

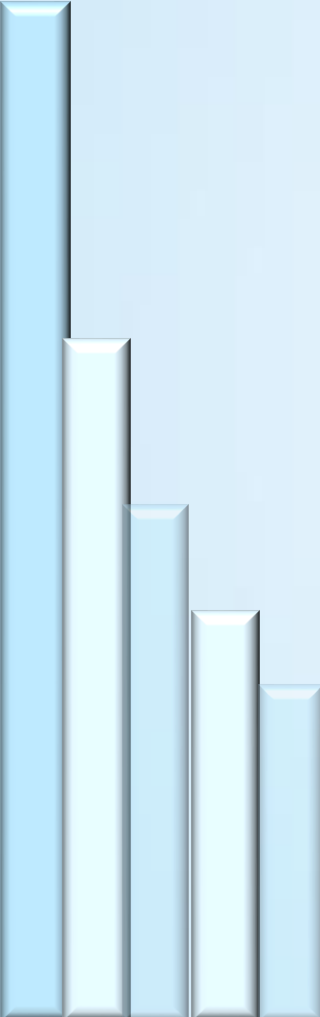
Andrzej Młodak, Łukasz Wawrowski, Tomasz Józefowski



Urząd Statystyczny w Poznaniu  
Ośrodek Statystyki Małych Obszarów

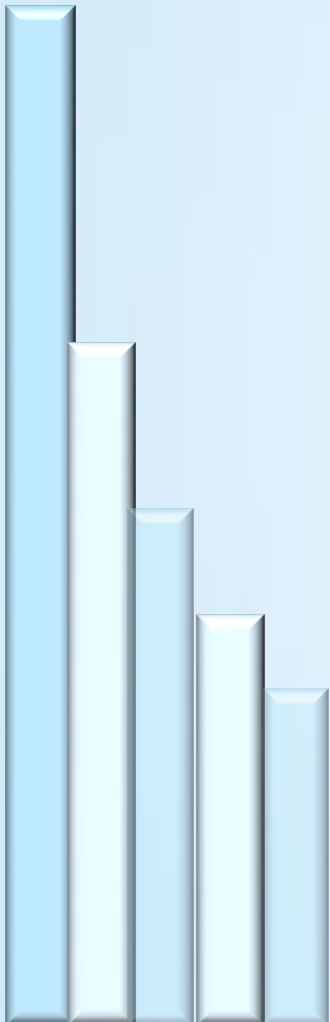


# Cele analizy

- 
- Zbadanie możliwości konstrukcji syntetycznego miernika taksonomicznego jako potencjalnej zmiennej objaśniającej dla estymacji wskaźników bardzo niskiej intensywności pracy i pogłębionej deprywacji materialnej w województwach, w tym:
    - stworzenie metod weryfikacji zmiennościowo – korelacyjnej dla danych panelowych uwzględniających oba komponenty: czasowy i przestrzenny,
    - analiza użyteczności zastosowania taksonomii wielokryterialnej
    - ocena efektywności i jakości estymacji wykorzystującej ów miernik jako zmienną pomocniczą w dynamicznym modelu Faya – Herriota.
  - Źródła danych:
    - wielokryterialne dane panelowe za lata 2005 – 2012 z BDL
    - alternatywne dane panelowe za lata 2005 – 2012 z BDL.

# Etapy konstrukcji miernika syntetycznego

- Ustalenie zestawu cech wskaźnikowych opisujących ubóstwo
  - dobór pod kątem istotności i wyczerpania zakresu zjawiska złożonego, proporcjonalności ujęcia zjawisk cząstkowych, precyzji definicyjnej i logiczności wzajemnych powiązań
  - cechy muszą być mierzalne z kompletnymi danymi.
- Weryfikacja zmiennościowo – korelacyjna:
  - eliminacja cech o zmienności mniejszej co do modułu niż 10%
  - wykluczenie cech nadmiernie skorelowanych z pozostałymi (w oparciu o elementy diagonalne odwróconej macierzy korelacji większe od 10) z uwzględnieniem korelacji z docelowymi wskaźnikami.
- Zamiana cech diagnostycznych (cech pozostałych po weryfikacji) będących destymulantami na stymulanty poprzez zmianę znaków ich wartości na przeciwne



# Etapy konstrukcji miernika syntetycznego

- Normalizacja cech z wykorzystaniem mediany Webera  $\Theta$

$$z_{ij} = \frac{x_{ij} - \theta_j}{1,4826 \cdot \operatorname{med}_{i=1,2,\dots,n} |x_{ij} - \theta_j|}, \quad i = 1, 2, \dots, n, j = 1, 2, \dots, m.$$

- Wyznaczenie taksonomicznego wzorca rozwoju

$$\psi_j = \max_{i=1,2,\dots,n} z_{ij}, \quad j = 1, 2, \dots, m$$

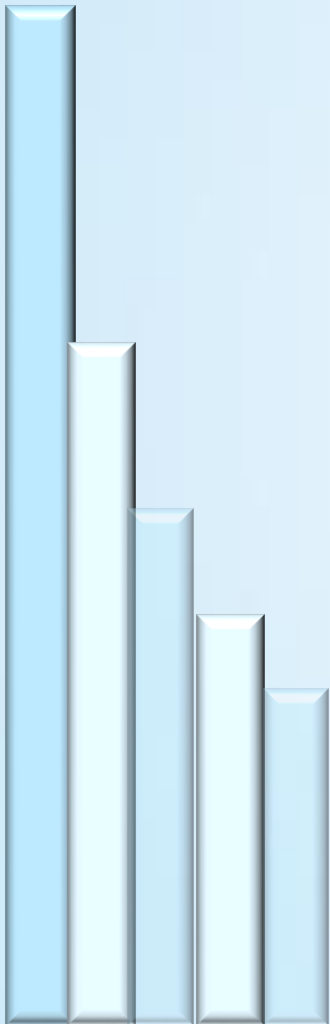
- Wyznaczenie odległości obiektów od niego

$$d_i = \operatorname{med}_{j=1,2,\dots,m} |z_{ij} - \psi_j|, \quad i = 1, 2, \dots, n.$$

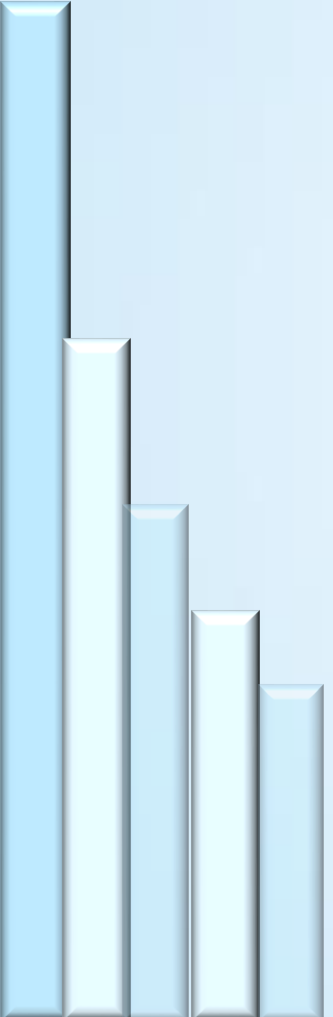
- Wyznaczenie wartości miernika kompleksowego

$$\mu_i = 1 - \frac{d_i}{\operatorname{med}(\mathbf{d}) + 2,5 \cdot \operatorname{mad}(\mathbf{d})},$$

$$i = 1, 2, \dots, n, \mathbf{d} = (d_1, d_2, \dots, d_n), \operatorname{mad}(\mathbf{d}) = \operatorname{med}_{i=1,2,\dots,n} |d_i - \operatorname{med}(\mathbf{d})|.$$

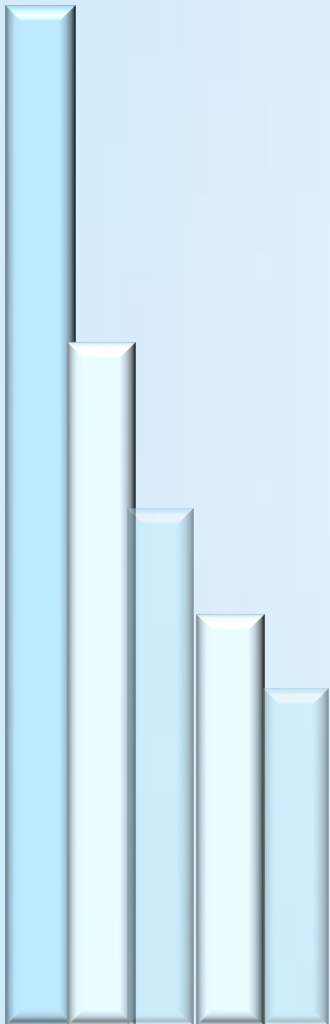


# Grupowanie w oparciu o wartości miernika

- 
- ▶ Metoda trzech median ( $\text{med}_1(\mu)$  – mediana wartości miernika większych od jego mediany,  $\text{med}_2(\mu)$  – nie mniejszych)
    - ▶ grupa 1:  $\{i \in \{1, 2, \dots, n\}: \mu_i > \text{med}_1(\mu)\}$ ,
    - ▶ grupa 2:  $\{i \in \{1, 2, \dots, n\}: \text{med}(\mu) < \mu_i \leq \text{med}_1(\mu)\}$ ,
    - ▶ grupa 3:  $\{i \in \{1, 2, \dots, n\}: \text{med}_2(\mu) < \mu_i \leq \text{med}(\mu)\}$ ,
    - ▶ grupa 4:  $\{i \in \{1, 2, \dots, n\}: \mu_i \leq \text{med}_2(\mu)\}$ .
  - ▶ Metoda progowa
    - ▶ grupa 1:  $\{i \in \{1, 2, \dots, n\}: \mu_i \geq \text{med}(\mu) + 2,5 \cdot \text{mad}(\mu)\}$ ,
    - ▶ grupa 2:  $\{i \in \{1, 2, \dots, n\}: \text{med}(\mu) \leq \mu_i < \text{med}(\mu) + 2,5 \cdot \text{mad}(\mu)\}$ ,
    - ▶ grupa 3:  $\{i \in \{1, 2, \dots, n\}: \text{med}(\mu) - 2,5 \cdot \text{mad}(\mu) \leq \mu_i < \text{med}(\mu)\}$ ,
    - ▶ grupa 4:  $\{i \in \{1, 2, \dots, n\}: \mu_i < \text{med}(\mu) - 2,5 \cdot \text{mad}(\mu)\}$ .

# Taksonomia wielokryterialna

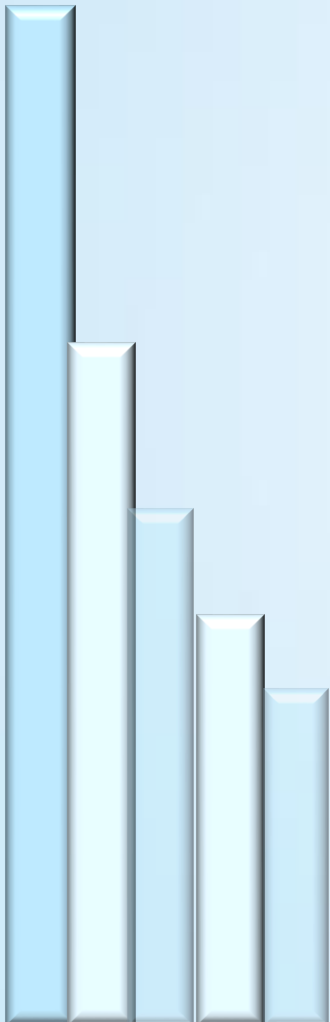
- Stosuje się ją gdy występuje znaczna wszechstronność zbiorów danych – tzn. gdy opisują one rozmaite dziedziny życia społeczno – gospodarczego, które w mniejszym bądź większym stopniu są powiązane z rozpatrywanym zjawiskiem
- Dokonuje się wówczas podziału zestawu owych zmiennych na podzestawy opisujące owe dziedziny (np. demografia, rynek pracy, warunki życia, itp.).
- Każda z tychże dziedzin jest traktowana jako odrębne zjawisko złożone. Zatem do każdej z nich stosuje się procedurę konstrukcji miernika syntetycznego (tutaj zwanego *cząstkowym*)
- W oparciu o zestaw mierników cząstkowych konstruuje się miernik kompleksowy (tą samą metodą, bez weryfikacji).



# Taksonomia dla danych panelowych

- ▶ Dane panelowe – dane łączące w sobie zarówno komponent przekrojowy jak czasowy:  $\mathbf{X} = [x_{ijt}]$ ,  $i = 1, 2, \dots, n$ ,  $j = 1, 2, \dots, m$ ,  $t = 1, 2, \dots, \tau$ ,  $n$  – liczba obiektów,  $m$  – liczba zmiennych,  $\tau$  – liczba okresów.
- ▶ Pierwszy problem: kwestia weryfikacji zmiennościowej i korelacyjnej uwzględniającej oba wymiary.
- ▶ Weryfikacja zmiennościowa:
  - ▶ wyznaczenie median Webera  $\Theta_1, \Theta_2, \dots, \Theta_\tau \in \mathbb{R}^m$  dla kolejnych okresów czasu (np. lat),  $\Theta_t = (\theta_{t1}, \theta_{t2}, \dots, \theta_{tm})$ ,  $t = 1, 2, \dots, \tau$ .
  - ▶ zdefiniowanie macierzy kompleksowej:  $\Theta = [\Theta_1^T \ \Theta_2^T \ \dots \ \Theta_\tau^T]$  rozmiaru  $m \times \tau$  i wyznaczenie jej mediany Webera  $\Theta^* = (\theta_1^*, \theta_2^*, \dots, \theta_m^*) \in \mathbb{R}^m$
  - ▶ obliczenie kompleksowych wskaźników zmienności cech:

$$CV_j = \frac{\text{mãd}(\tilde{\Theta}_j)}{\theta_j^*}; \quad \text{gdzie} \quad \text{mãd}(\tilde{\Theta}_j) = \text{med}_{t=1,2,\dots,\tau} |\theta_{tj} - \theta_j^*|, \quad j = 1, 2, \dots, m$$



# Taksonomia dla danych panelowych

## ► Weryfikacja korelacyjna

- wyznaczenie macierzy korelacji Pearsona dla każdego z badanych okresów

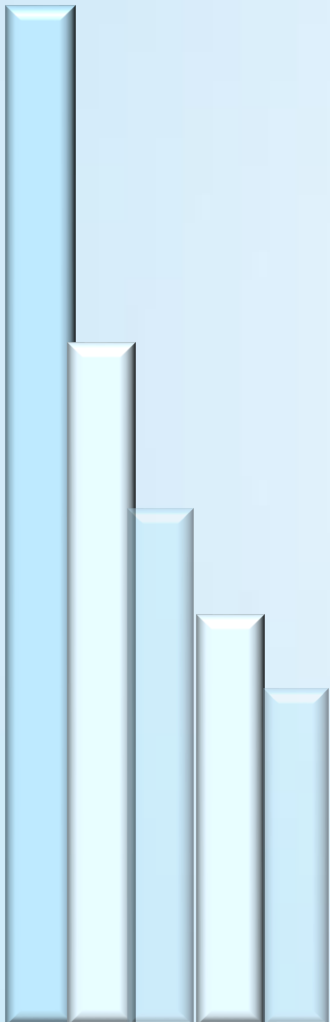
$$\mathbf{R}_t = [r_{jkt}],$$

$r_{jkt}$  – współczynnik korelacji Pearsona zmiennych  $X_j$  i  $X_k$  w okresie  $t, j, k = 1, 2, \dots, m, t = 1, 2, \dots, \tau$ .

- stworzenie kompleksowej macierzy korelacji z tych współczynników, które są maksymalne co do wartości bezwzględnej

$$\mathbf{R} = [r_{jk}^*], \quad \text{gdzie} \quad r_{jk}^* = r_{jkt^*} \quad \text{jeśli} \quad |r_{jkt^*}| = \max_{t=1,2,\dots,\tau} |r_{jkt}|$$

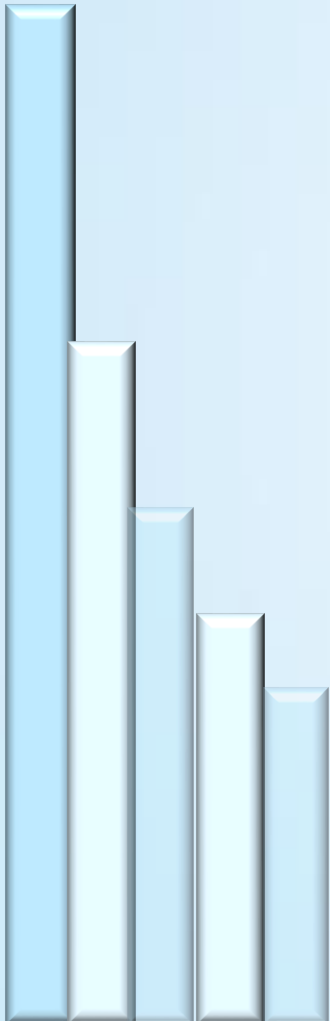
- dalsze postępowanie zgodne z metodą odwróconej macierzy korelacji uwzględniającej poziom skorelowania ze zmienną objaśnianą
  - w przypadku niejednoznaczności eliminacji w zakresie wzajemnego skorelowania, eliminowano zmienne mniej skorelowane ze zmienną docelową





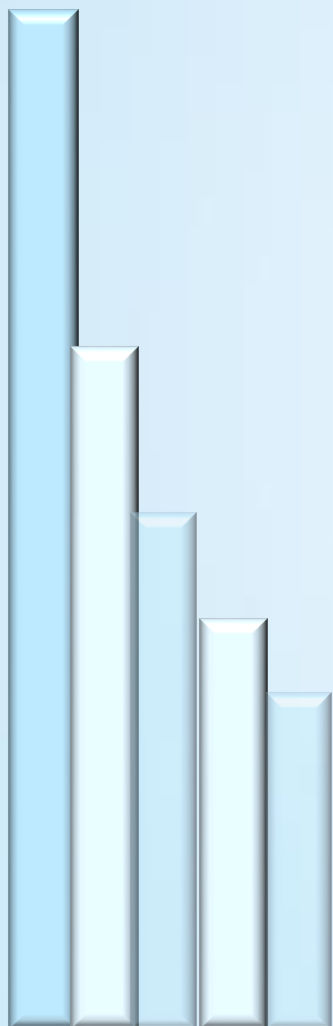
# Wielokryterialne dane panelowe

- ▶ 140 zmiennych wskaźnikowych dla lat 2005 – 2012
  - ▶ cztery dziedziny:
    - ▶ demografia
    - ▶ gospodarka mieszkaniowa i komunalna
    - ▶ rynek pracy
    - ▶ warunki życia
  - ▶ każde województwo w zakresie każdej zmiennej było opisane przy pomocy ośmiu liczb dla poszczególnych lat
  - ▶ zastosowano taksonomię wielokryterialną i weryfikację zmiennych dla danych panelowych
  - ▶ charakteru cech diagnostycznych ustalono w oparciu o ich wpływ na wykluczenie społeczne, mierzone głównie wskaźnikiem niskiej intensywności pracy



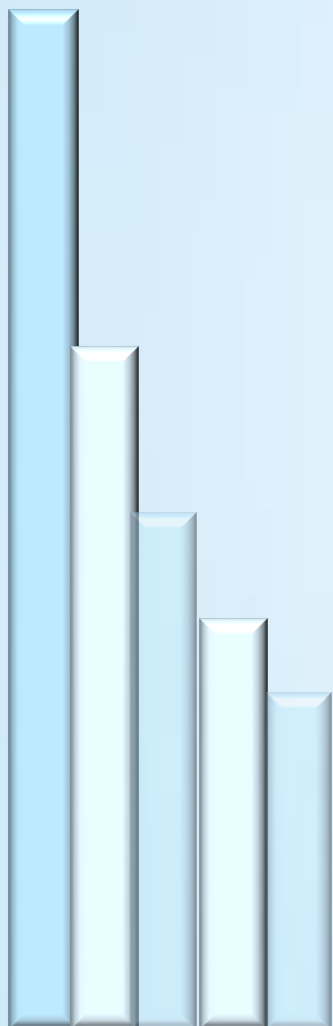
# Wielokryterialne dane panelowe

- ▶ Zestaw cech diagnostycznych (S–stymulanta, D–destymulanta):
  - ▶ demografia:
    - ▶ saldo migracji wewnętrznych – kobiety na 1000 kobiet ogółem (D),
    - ▶ zmiana liczby ludności na 1000 mieszkańców (D),
    - ▶ udział zgonów osób w wieku 20–59 lat w liczbie zgonów ogółem (S),
  - ▶ gospodarka mieszkaniowa i komunalna:
    - ▶ kwota wypłaconych dodatków mieszkaniowych na jednego mieszkańca (S),
    - ▶ korzystający z instalacji kanalizacyjnej w % ogółu ludności na wsi (S),
    - ▶ korzystający z instalacji gazowej w % ogółu ludności na wsi (D),
  - ▶ rynek pracy:
    - ▶ udział bezrobotnych kobiet zarejestrowanych w wieku 55 lat i więcej w liczbie bezrobotnych kobiet ogółem (S),



# Wielokryterialne dane panelowe

- Zestaw cech diagnostycznych (S–stymulanta, D–destymulanta):
  - rynek pracy (dok.):
    - udział kobiet poszkodowanych w wypadkach przy pracy w liczbie pracujących kobiet ogółem (S),
    - zatrudnieni w warunkach zagrożenia ogółem (S),
    - stopa bezrobocia rejestrowanego, Polska=100 (S),
  - warunki życia:
    - jednostki wykreślone z rejestru REGON na 10 tys. ludności (S),
    - podmioty 10-49 na 10 tys. mieszkańców w wieku produkcyjnym (D),
    - uczniowie przypadający na 1 komputer z dostępem do Internetu przeznaczony do użytku uczniów – gimnazja (S),
    - uczniowie przypadający na 1 komputer z dostępem do Internetu przeznaczony do użytku uczniów – szkoły zasadnicze zawodowe (D).



# Wielokryterialne dane panelowe

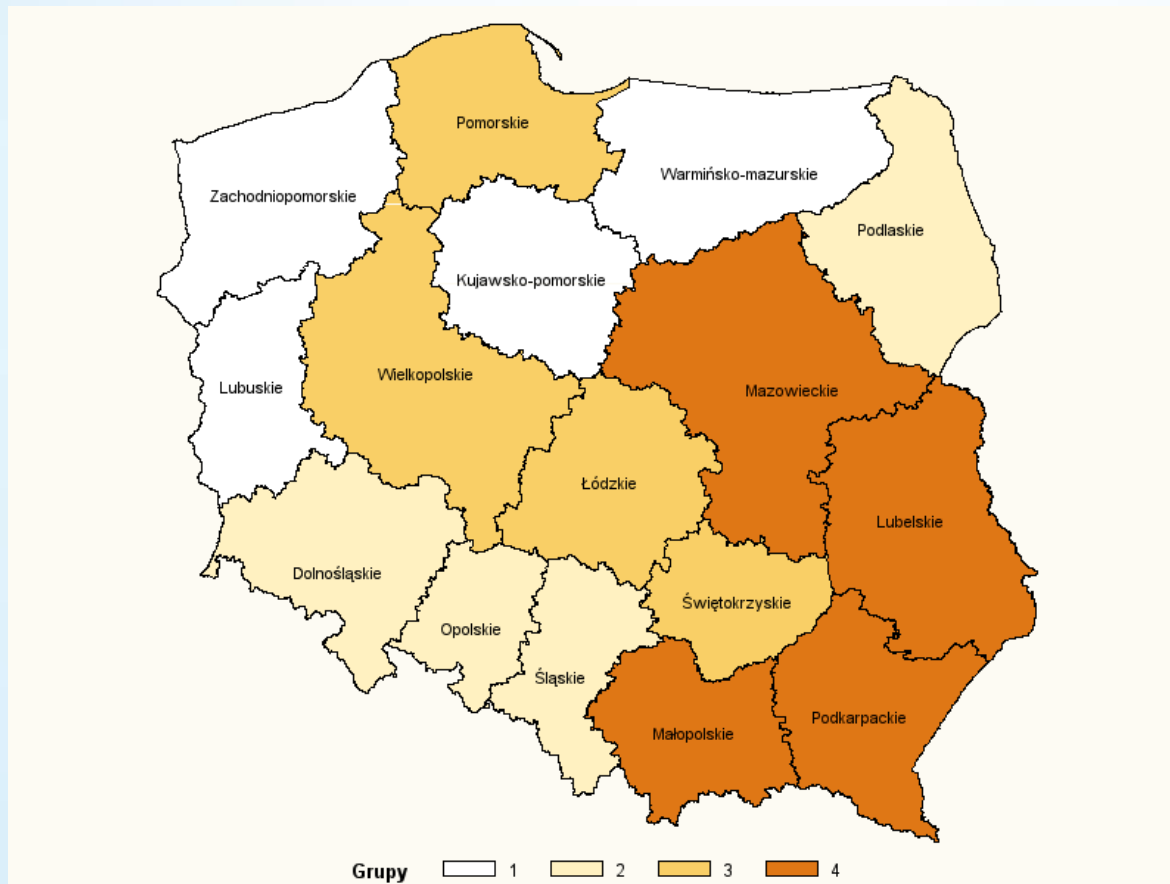
## Wyniki analizy

- podstawowe statystyki opisowe dla mierników cząstkowych i miernika kompleksowego

Wyszczególnienie	Średnia arytmetyczna	Odchylenie standardowe	Współczynnik zmienności w %	Minimum	Kwartył dolny	Mediana	Kwartył górny	Maksimum
Demografia	0,3994	0,2174	54,4194	-0,2371	0,2584	0,4129	0,5826	0,9476
Gospodarka mieszkaniowa i komunalna	0,4439	0,2249	50,6635	0,0454	0,2610	0,4367	0,6025	1,0000
Rynek pracy	0,3082	0,1649	53,5031	-0,0297	0,2075	0,2796	0,4340	0,7033
Warunki życia	0,2518	0,1626	64,5869	-0,1597	0,1440	0,2605	0,3495	0,6117
Miernik kompleksowy	0,3354	0,2174	64,8087	-0,1674	0,2134	0,3305	0,4775	0,8602

# Wielokryterialne dane panelowe

- Grupowanie województw metodą trzech median – mediany miernika



**grupa 1:**

$$\{i \in \{1, 2, \dots, n\}: \mu_i > 0,4789\}$$

**grupa 2:**

$$\{i \in \{1, 2, \dots, n\}: 0,3081 < \mu_i \leq 0,4789\}$$

**grupa 3:**

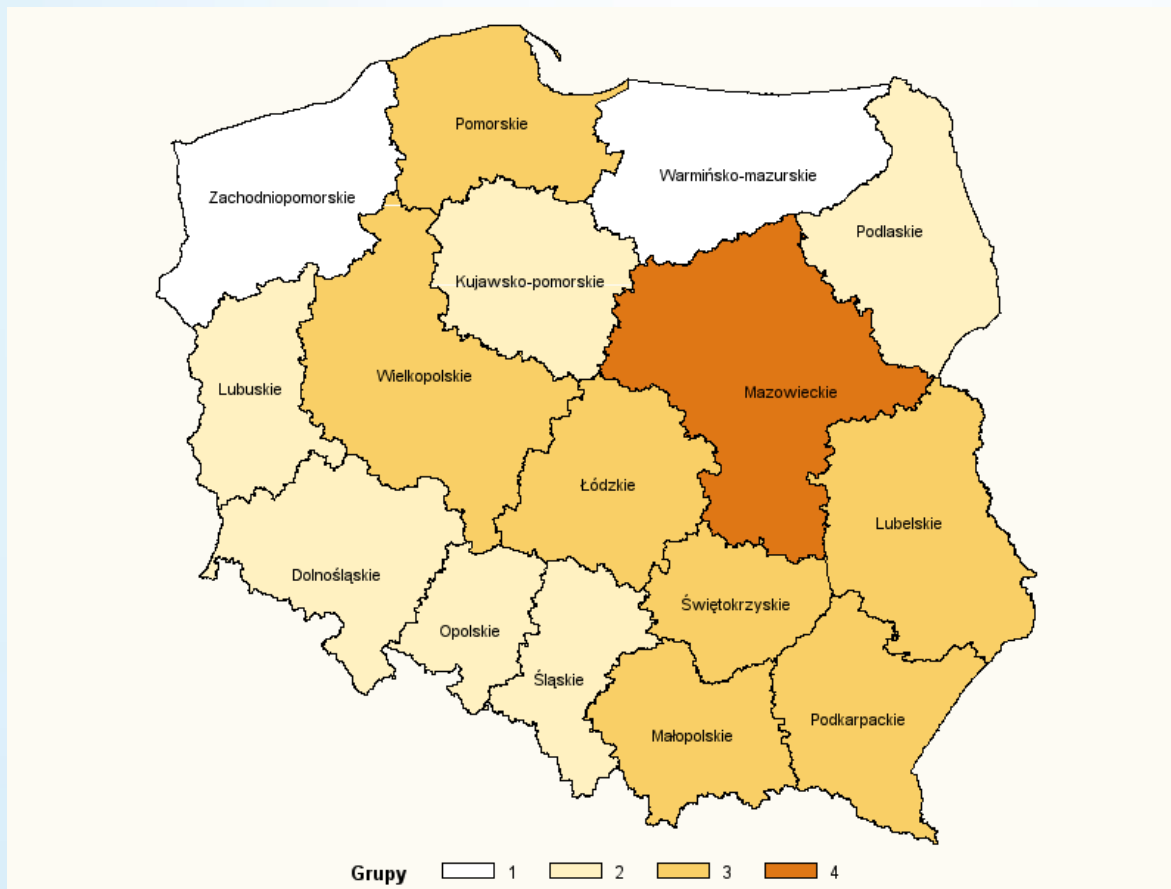
$$\{i \in \{1, 2, \dots, n\}: 0,2594 < \mu_i \leq 0,3081\}$$

**grupa 4:**

$$\{i \in \{1, 2, \dots, n\}: \mu_i \leq 0,2594\}.$$

# Wielokryterialne dane panelowe

➔ Grupowanie województw metodą progową – mediany miernika



**grupa 1:**

$$\{i \in \{1, 2, \dots, n\}: \mu_i \geq 0,6204\}$$

**grupa 2:**

$$\{i \in \{1, 2, \dots, n\}: 0,3081 \leq \mu_i < 0,6204\}$$

**grupa 3:**

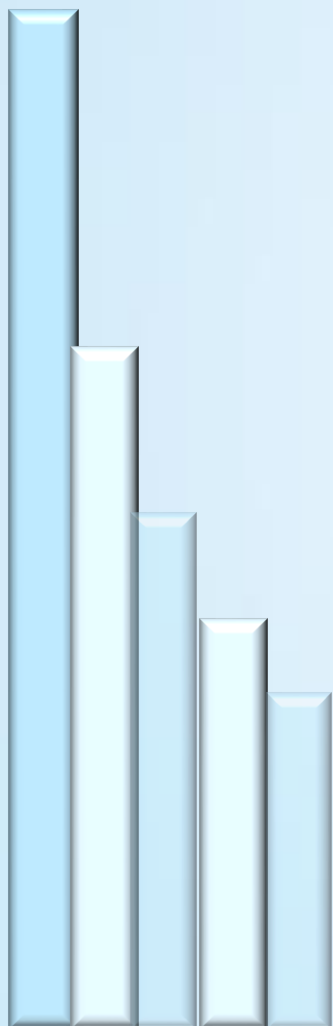
$$\{i \in \{1, 2, \dots, n\}: -0,0042 \leq \mu_i < 0,3081\}$$

**grupa 4:**

$$\{i \in \{1, 2, \dots, n\}: \mu_i < -0,0042\}.$$

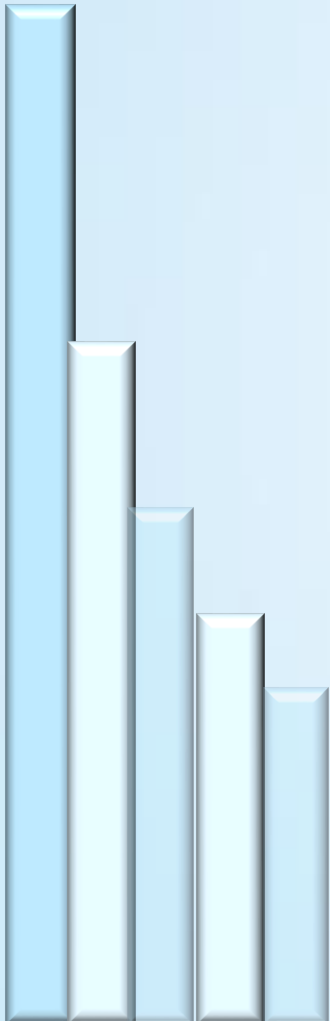
# Wielokryterialne dane panelowe

- ▶ Przeprowadzono także podobne analizy dla regionów i podregionów:
  - ▶ zestawy cech diagnostycznych różniły się wówczas od zestawu dla województw, aczkolwiek miały też elementy wspólne:
    - ▶ demografia: saldo migracji wewnętrznych – kobiety na 1000 kobiet ogółem
    - ▶ gospodarka mieszkaniowa i komunalna: kwota wypłaconych dodatków mieszkaniowych na jednego mieszkańca oraz korzystający z instalacji kanalizacyjnej w % ogółu ludności na wsi,
    - ▶ rynek pracy: stopa bezrobocia rejestrowanego (dla województw *Polska* = 100),
  - ▶ zróżnicowanie wartości mierników cząstkowych i miernika syntetycznego było największe w podregionach, tam też korelacja z bezpośrednimi oszacowaniami docelowych wskaźników była najniższa.



# Alternatywne dane panelowe

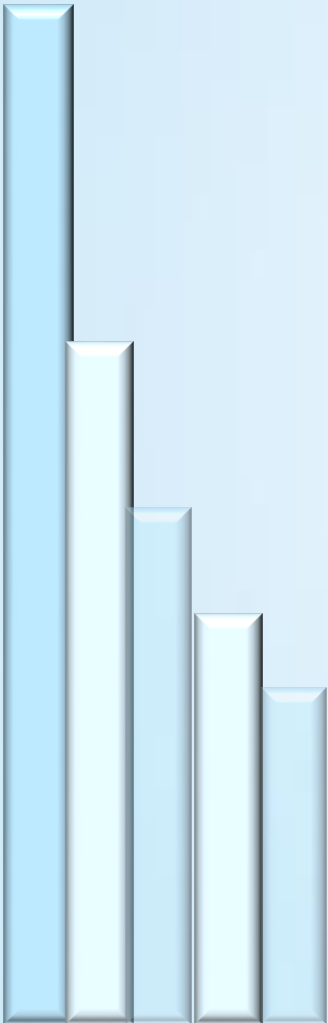
- ▶ 66 zmiennych wskaźnikowych dla lat 2005 – 2012,
- ▶ Zmienne dotyczą m.in. demografii, ochrony zdrowia i opieki społecznej, bezrobocia, warunków pracy czy wychowania przedszkolnego dzieci.
- ▶ Z uwagi na mniejszą liczbę zmiennych i ich strukturę tematyczną nie stosowano taksonomii wielokryterialnej
- ▶ W wyniku weryfikacji zmiennościowo – korelacyjnej (w tym korelacji z bezpośrednimi oszacowaniami wskaźników docelowych) stworzono dwa zestawy cech diagnostycznych:
  - ▶ dla wskaźnika bardzo niskiej intensywności pracy
  - ▶ dla wskaźnika pogłębionej deprivacji materialnej.





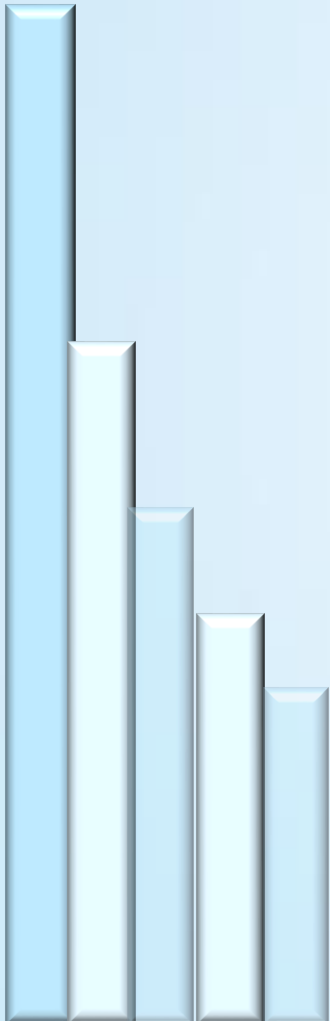
# Alternatywne dane panelowe

- ▶ Zestaw cech diagnostycznych dla wskaźnika bardzo niskiej intensywności pracy (S – stymulanta, liczba to wartość kompleksowego współczynnika korelacji z bezpośrednim oszacowaniem wskaźnika):
  - ▶ liczba osób z chorobą nowotworową na 1000 osób (S, 0,6416),
  - ▶ liczba wypłaconych dodatków mieszkaniowych na 100 osób (S, 0,7036),
  - ▶ udział liczby bezrobotnych nowo zarejestrowanych w liczbie bezrobotnych zarejestrowanych ogółem (S, 0,6077),
  - ▶ udział poszkodowanych w wypadkach przy pracy kobiet w liczbie pracujących kobiet ogółem (S, 0,7439),



# Alternatywne dane panelowe

- ▶ Zestaw cech diagnostycznych dla wskaźnika pogłębionej deprivacji materialnej (S – stymulanta, D – destymulanta; liczba to wartość kompleksowego współczynnika korelacji z bezpośrednim oszacowaniem wskaźnika):
  - ▶ udział liczby bezrobotnych kobiet zamieszkałych w mieście w liczbie bezrobotnych kobiet zarejestrowanych ogółem (D, -0,7067),
  - ▶ udział liczby bezrobotnych kobiet z prawem do zasiłku w liczbie bezrobotnych kobiet zarejestrowanych ogółem (S, 0,5228),
  - ▶ udział zatrudnionych w warunkach zagrożenia przez jedną grupę czynników związanych z uciążliwością pracy w liczbie zatrudnionych w warunkach zagrożenia ogółem (D, -0,6568),
  - ▶ liczba dzieci w wieku 3–5 lat przypadająca na jedno miejsce w przedszkolu (D, -0,5843),

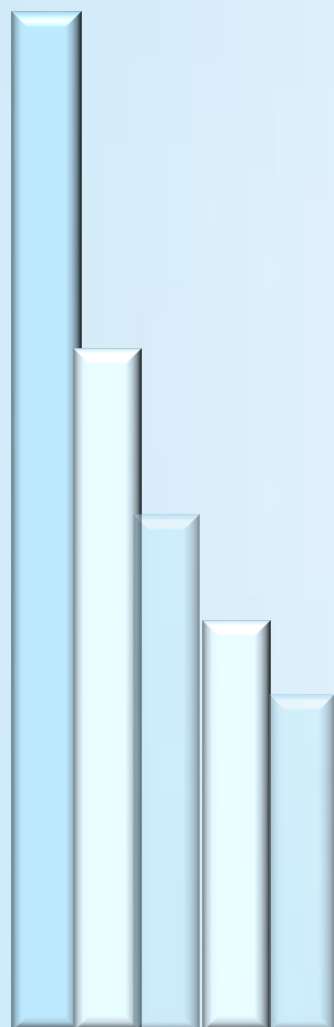


# Alternatywne dane panelowe

## ► Wyniki analizy

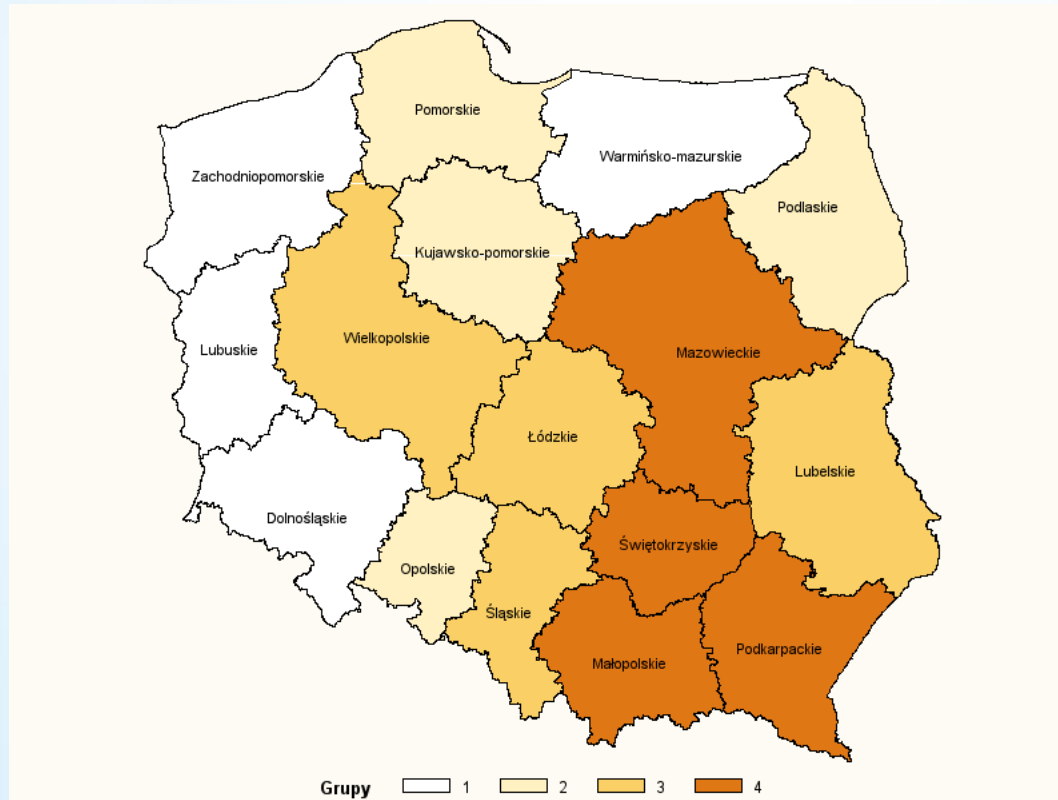
- podstawowe statystyki opisowe dla miernika syntetycznego

Wyszczególnienie	Wskaźnik	
	niskiej intensywności pracy	pogłębionej deprivacji materialnej
Średnia arytmetyczna	0,3158	0,4236
Odchylenie standardowe	0,1746	0,2345
Współczynnik zmienności w %	55,2974	55,3759
Minimum	-0,0059	-0,1395
Dolny kwartył	0,1915	0,2906
Mediana	0,3193	0,4359
Górny kwartył	0,4427	0,6225
Maksimum	0,7102	0,7814



# Alternatywne dane panelowe

- ▶ Grupowanie województw metodą trzech median – mediany miernika (dla wskaźnika bardzo niskiej intensywności pracy)



**grupa 1:**

$$\{i \in \{1, 2, \dots, n\}: \mu_i > 0,3987\}$$

**grupa 2:**

$$\{i \in \{1, 2, \dots, n\}: 0,3422 < \mu_i \leq 0,3987\}$$

**grupa 3:**

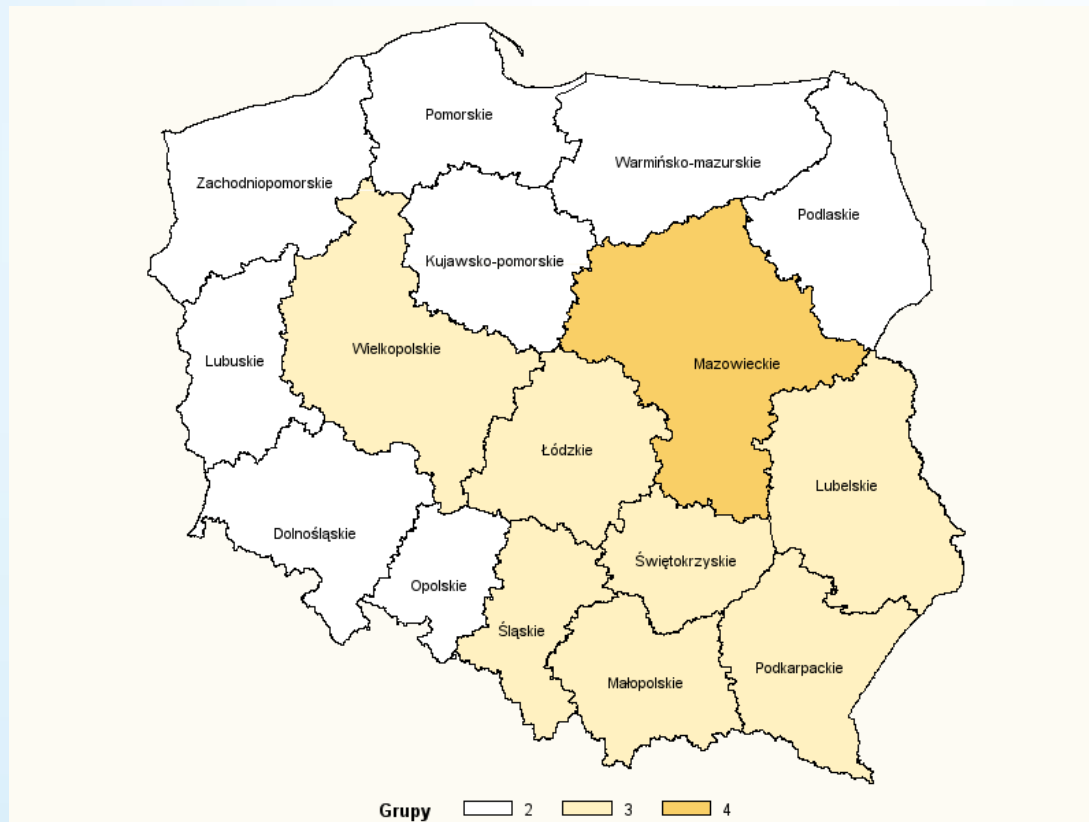
$$\{i \in \{1, 2, \dots, n\}: 0,1735 < \mu_i \leq 0,3422\}$$

**grupa 4:**

$$\{i \in \{1, 2, \dots, n\}: \mu_i \leq 0,1735\}.$$

# Alternatywne dane panelowe

- ▶ Grupowanie województw metodą progową – mediany miernika (dla wskaźnika bardzo niskiej intensywności pracy)



**grupa 1:**

$$\{i \in \{1, 2, \dots, n\}: \mu_i \geq 0,6292\}$$

**grupa 2:**

$$\{i \in \{1, 2, \dots, n\}: 0,3422 \leq \mu_i < 0,6292\}$$

**grupa 3:**

$$\{i \in \{1, 2, \dots, n\}: 0,0551 \leq \mu_i < 0,3422\}$$

**grupa 4:**

$$\{i \in \{1, 2, \dots, n\}: \mu_i < 0,0551\}.$$

# Alternatywne dane panelowe

- ▶ Grupowanie województw metodą trzech median – mediany miernika (dla wskaźnika pogłębionej deprywacji materialnej)



**grupa 1:**

$$\{i \in \{1, 2, \dots, n\}: \mu_i > 0,6202\}$$

**grupa 2:**

$$\{i \in \{1, 2, \dots, n\}: 0,5144 < \mu_i \leq 0,6202\}$$

**grupa 3:**

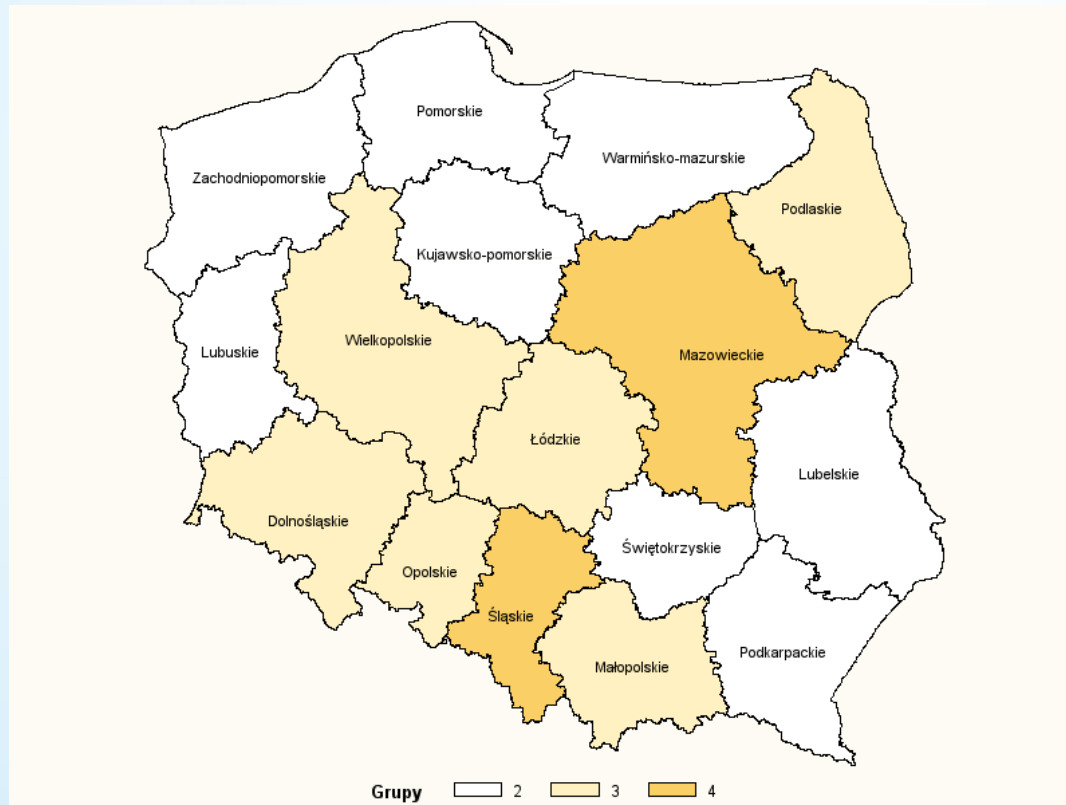
$$\{i \in \{1, 2, \dots, n\}: 0,3072 < \mu_i \leq 0,5144\}$$

**grupa 4:**

$$\{i \in \{1, 2, \dots, n\}: \mu_i \leq 0,3072\}.$$

# Alternatywne dane panelowe

- ▶ Grupowanie województw metodą progową – mediany miernika (dla wskaźnika pogłębionej deprywacji materialnej)



**grupa 1:**

$$\{i \in \{1, 2, \dots, n\}: \mu_i \geq 0,9135\}$$

**grupa 2:**

$$\{i \in \{1, 2, \dots, n\}: 0,5144 \leq \mu_i < 0,9135\}$$

**grupa 3:**

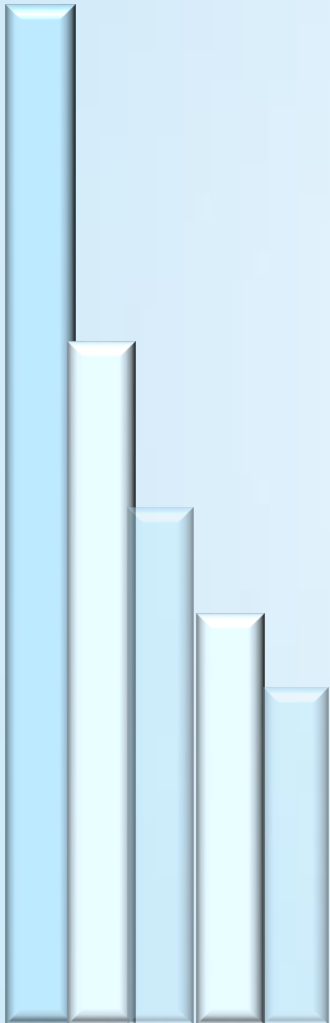
$$\{i \in \{1, 2, \dots, n\}: 0,1152 \leq \mu_i < 0,5144\}$$

**grupa 4:**

$$\{i \in \{1, 2, \dots, n\}: \mu_i < 0,1152\}.$$

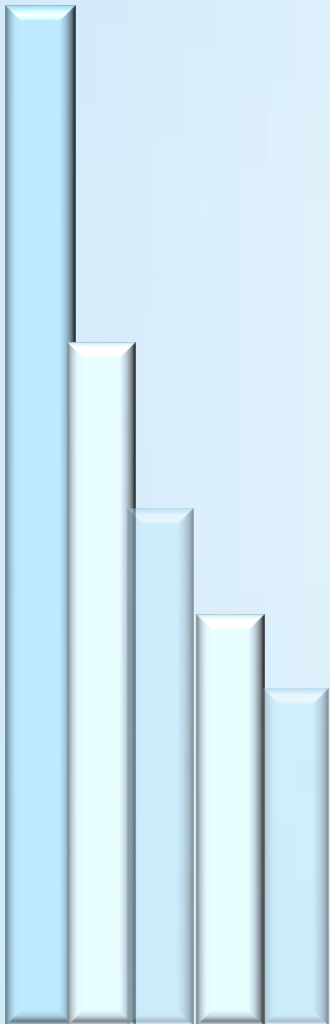
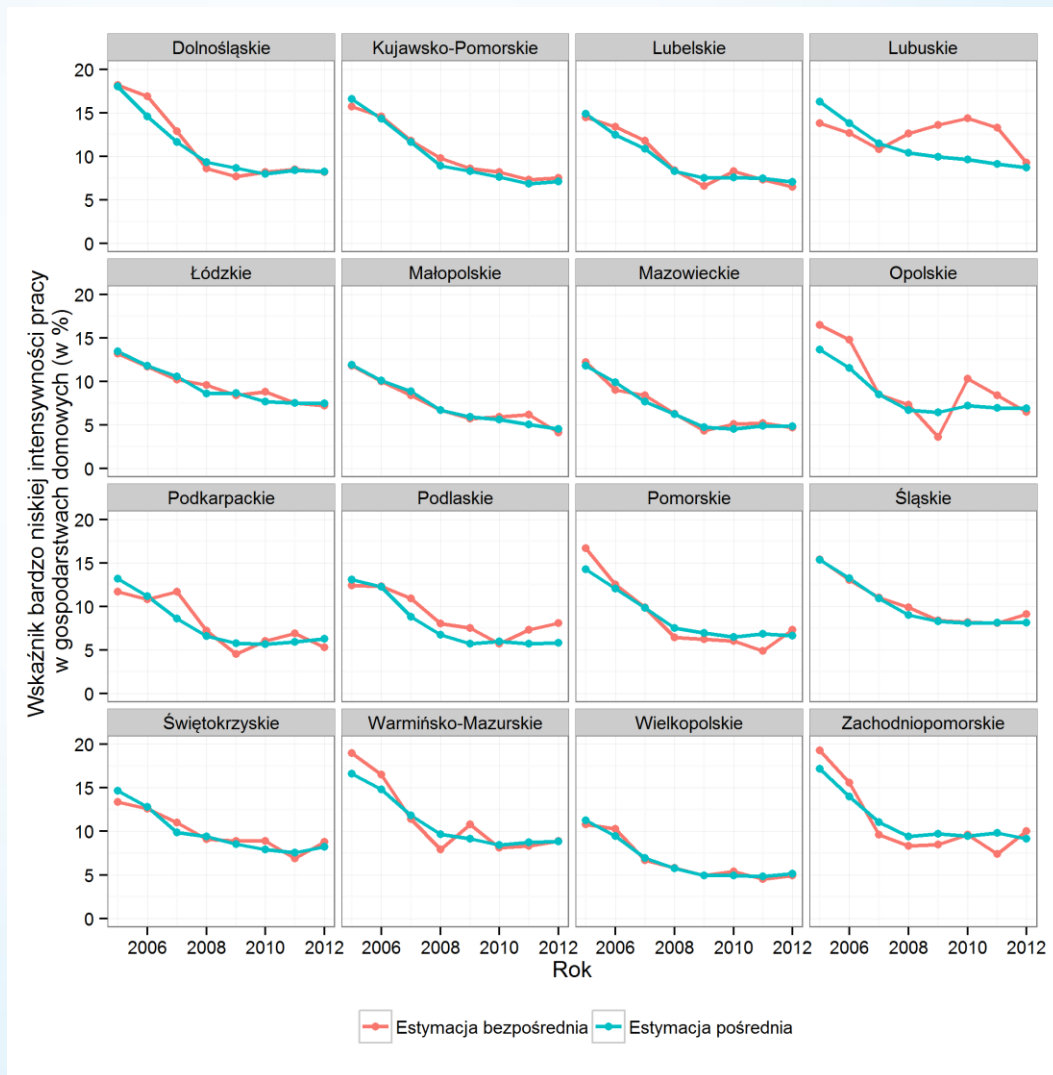
# Alternatywne dane panelowe

- ▶ Przeprowadzono także podobne analizy dla regionów i podregionów:
  - ▶ dla żadnego ze wskaźników docelowych nie było cech wspólnych we wszystkich zestawach
  - ▶ dla wskaźnika pogłębionej deprivacji materialnej kwota wypłaconych dodatków mieszkaniowych przypadająca na sto dodatków w przypadku regionów była stymulantą (korelacja 0,9554), zaś dla podregionów – destymulantą (-0,4287)
  - ▶ podregiony wykazywały największe zróżnicowanie mierników syntetycznych w obu przypadkach, województwa – najmniejsze
  - ▶ korelacja z bezpośrednimi oszacowaniami docelowych wskaźników jest zdecydowanie najmniejsza dla podregionów – tylko najwyższa przekroczyła 50%, podczas gdy w innych obszarach są nawet takie >90%

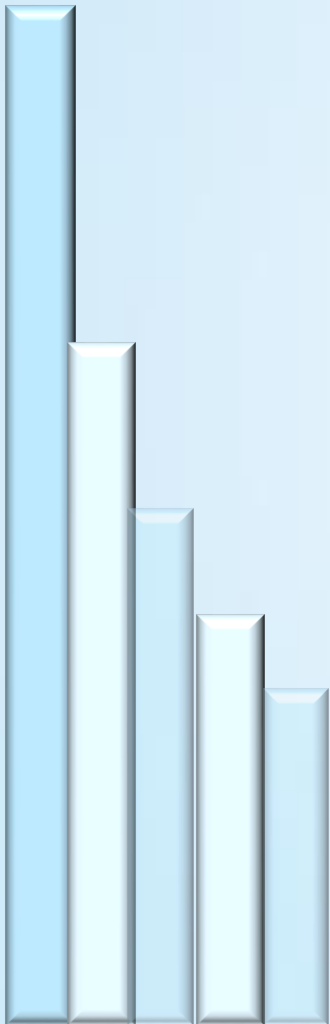
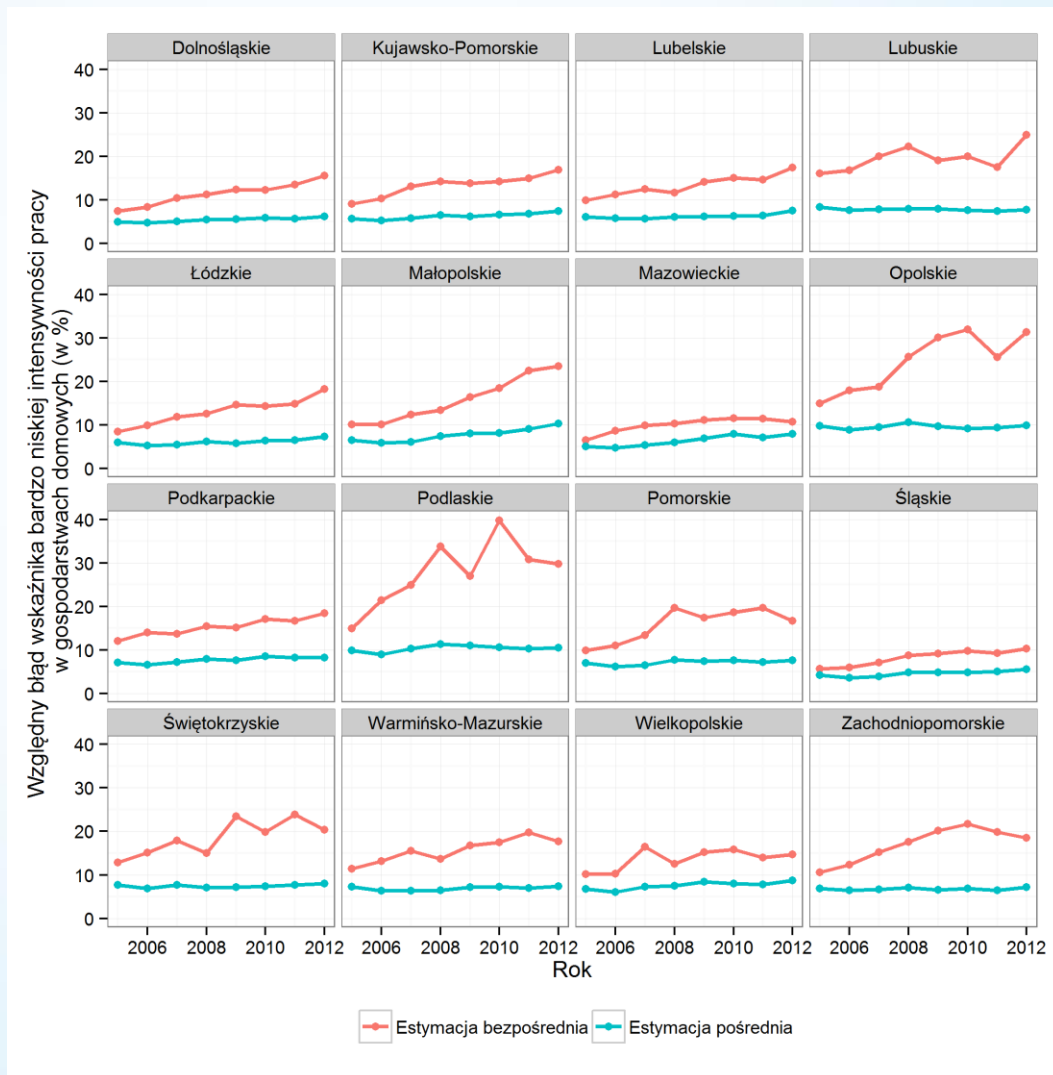




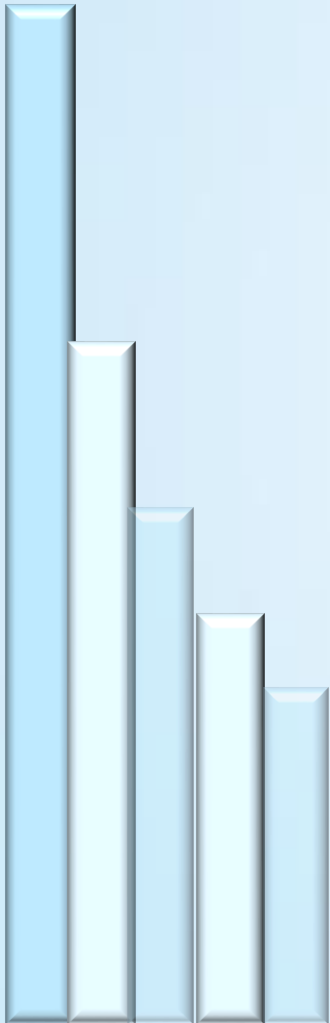
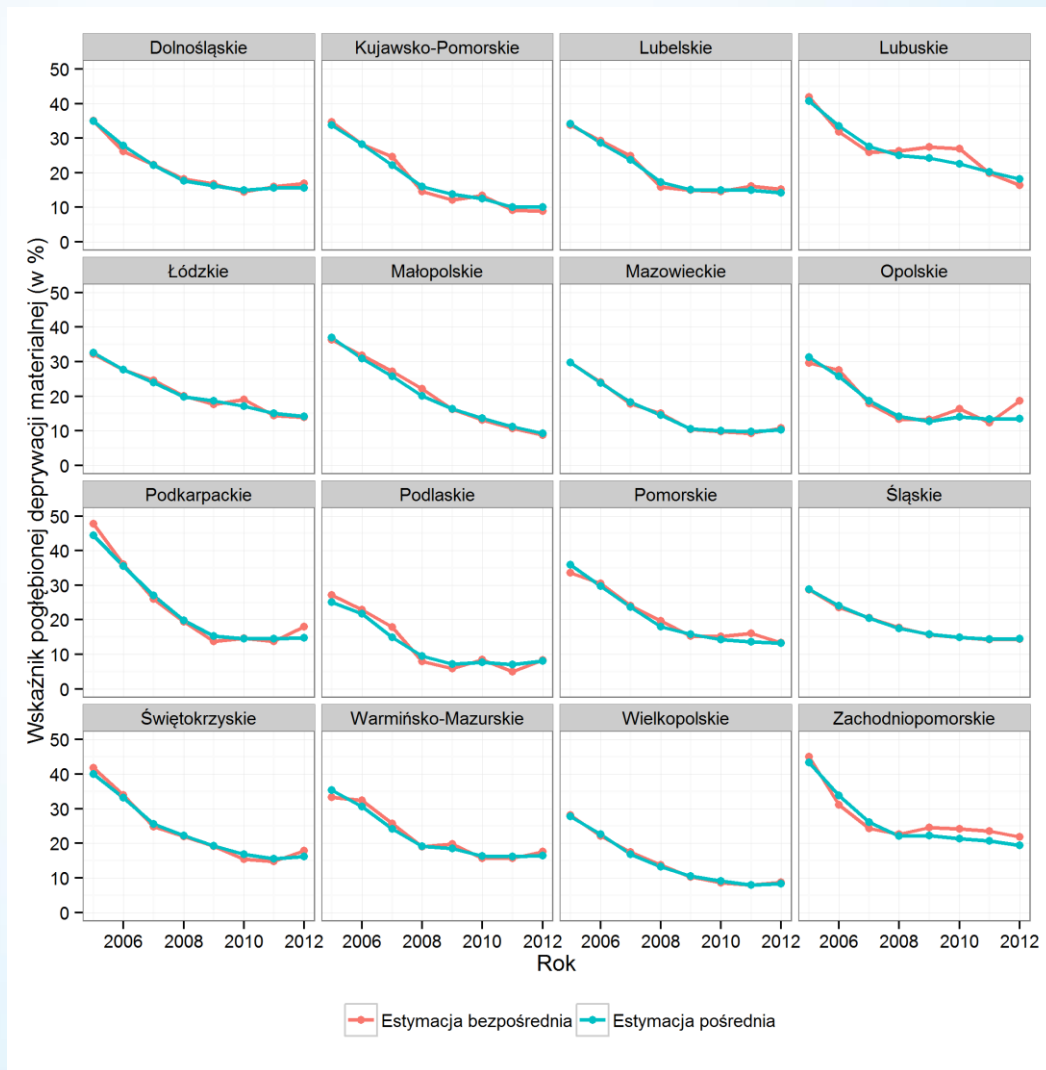
# Oszacowania wskaźnika bardzo niskiej intensywności pracy



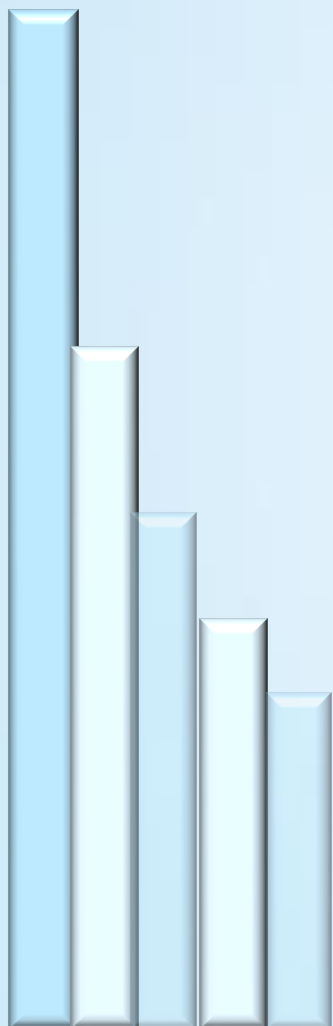
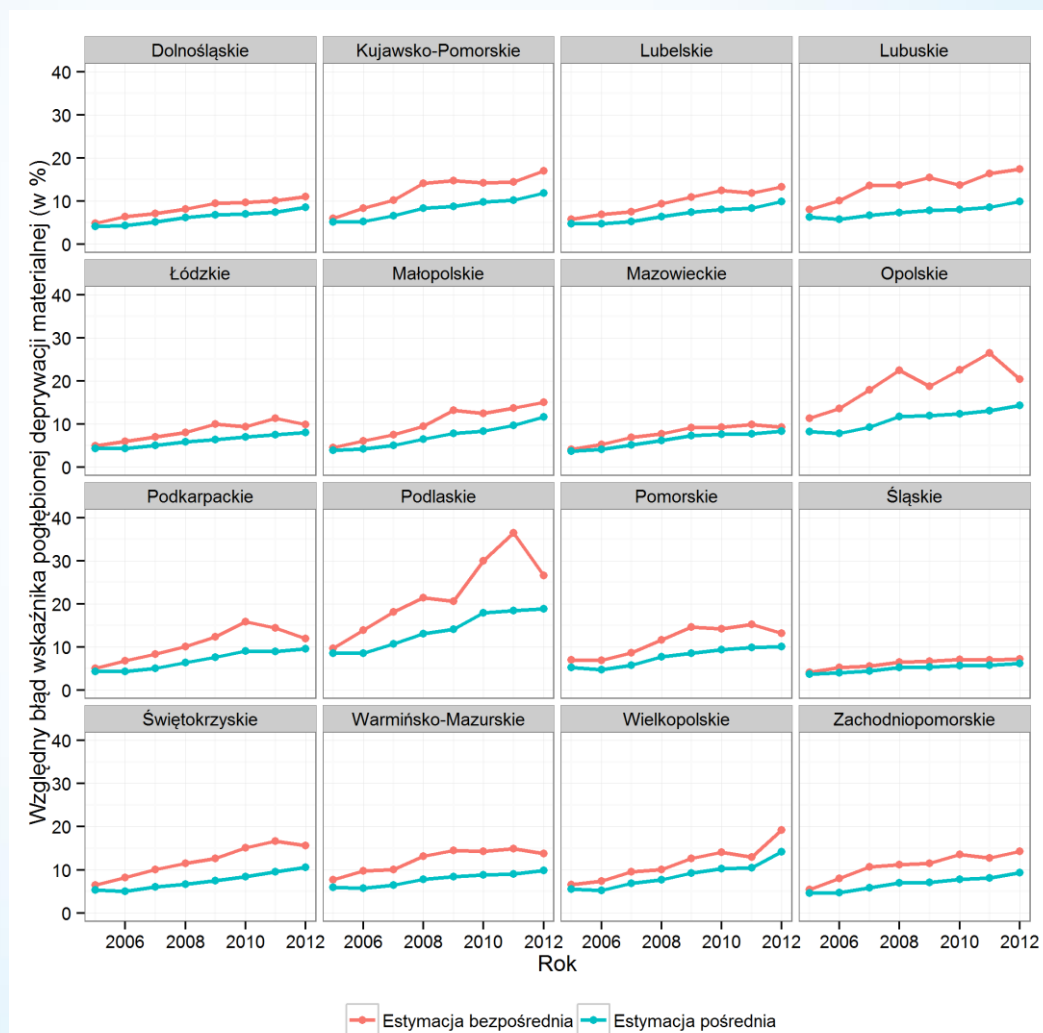
# Względne błędy oszacowań wskaźnika bardzo niskiej intensywności pracy



# Oszacowania wskaźnika pogłębionej deprivacji materialnej

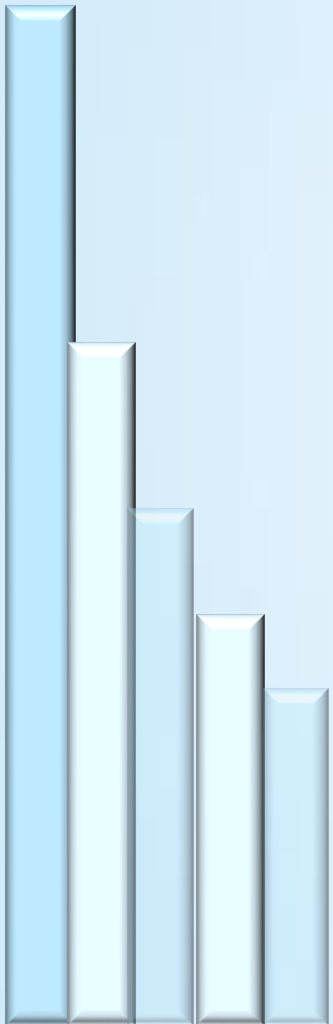


# Względne błędy oszacowań wskaźnika pogłębionej deprywacji materialnej



# Konkluzje

- Wskaźniki niskiej intensywności pracy oraz pogłębionej deprivacji materialnej są na ogół miernikami odrębnych aspektów ubóstwa i jako takie determinowane są przeważnie innymi czynnikami; im niższy poziom przestrzenny tym skala owej determinacji słabsza
- Mierniki taksonomiczne ukazują największe problemy w tym zakresie na słabo zurbanizowanych obszarach północnej i zachodniej Polski
- Mierniki okazały się przydatne w estymacji rzeczonych wskaźników ubóstwa dla województw.
  - najlepsze wyniki uzyskuje się dla wskaźnika bardzo niskiej intensywności pracy. Specyfika wskaźnika pogłębionej deprivacji materialnej znacznie utrudnia znalezienie doń efektywnych zmiennych pomocniczych.





**Dziękujemy za uwagę**