

CZĘŚĆ 2

Wybrane metody uzupełniania braków danych

Prezentacja na temat pracy magisterskiej

Marta Warzecha

Spis treści

1. Algorytm PMM
2. Podsumowanie



ALGORYTM PMM

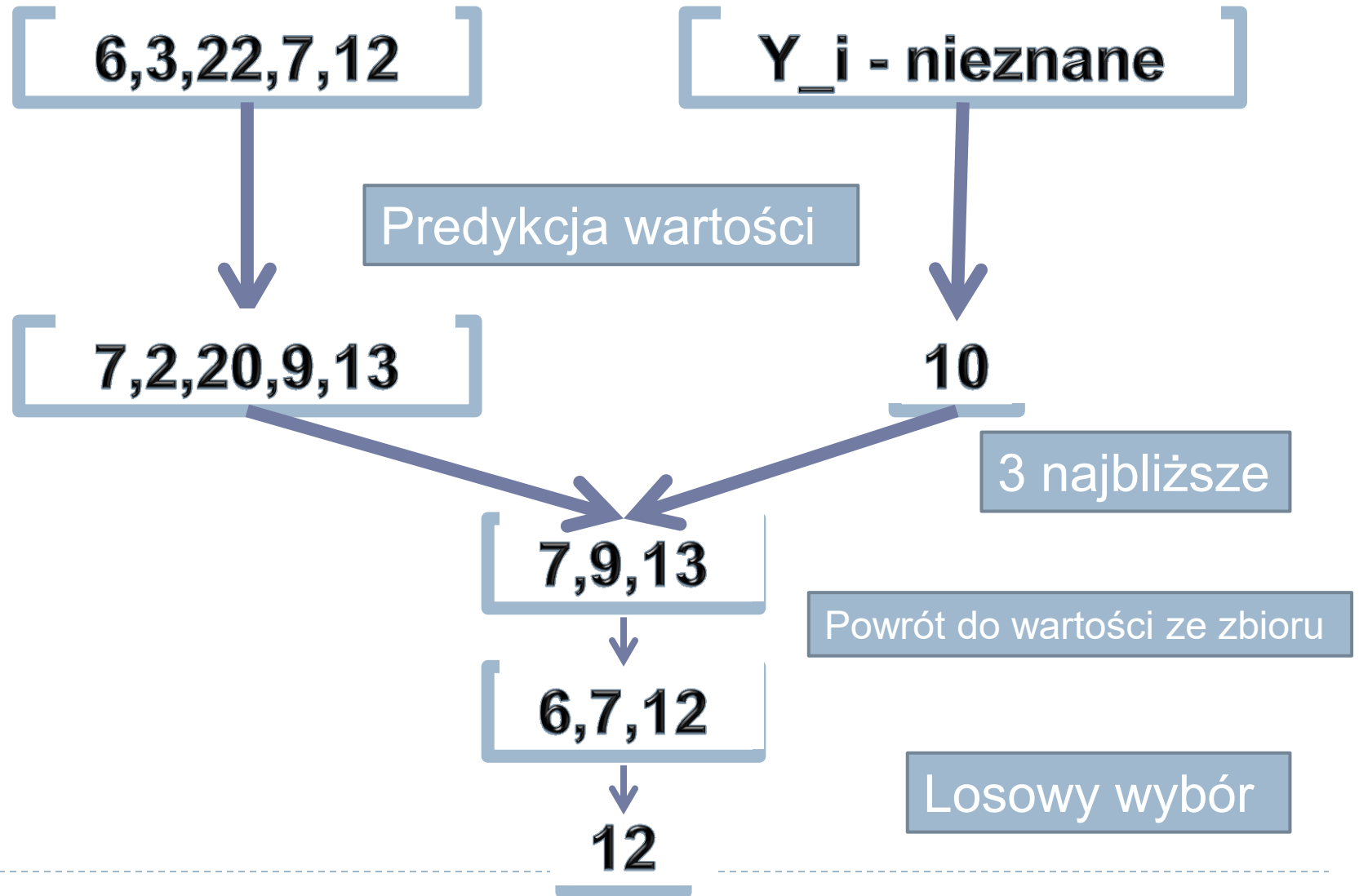
Predykcyjne dopasowywanie średniej

Ang. Predictive Mean Matching

Algorytm 1 (PMM)

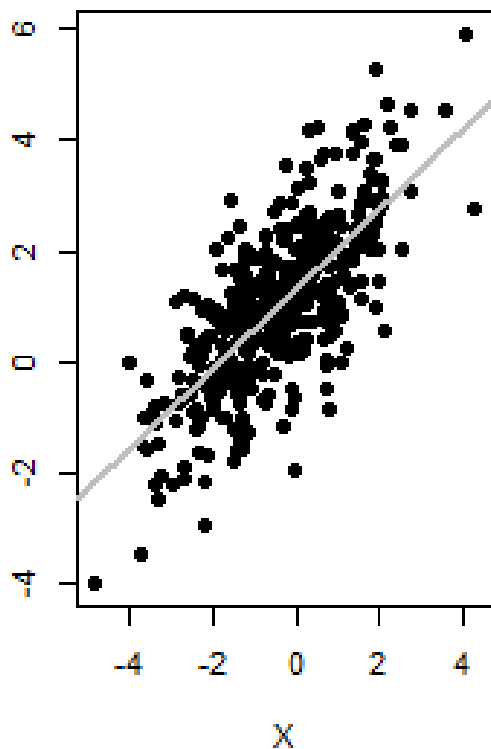
1. *Estymujemy model liniowy wyjaśniający Y za pomocą X na podstawie tylko obserwacji z pełną informacją. Rekordy, w których występuje jakiś brak pomijamy. Dostajemy w ten sposób wektor estymatorów współczynników modelu $\hat{\beta}$.*
2. *Korzystając z predykcyjnego modelu posteriori $\hat{\beta}$ produkujemy nowy zbiór współczynników modelu $\tilde{\beta}$. Standardowo powinna to być próbka losowa z wielowymiarowego rozkładu normalnego $\tilde{\beta} \sim N_n(\hat{\beta}, \text{cov}(\hat{\beta}))$. Ten krok zapewnia nam wystarczającą zmienność wartości imputowanych.*
3. *Następnie zgodnie z wyznaczonymi estymatorami współczynników modelu obliczamy predykcje:*
 - Y^{missing} dla obserwacji z brakami używając $\tilde{\beta}$.
 - Y^{full} dla pełnych obserwacji korzystając z $\hat{\beta}$.
4. *Dla każdego przypadku Y_i^{missing} , w którym występuje brak szukamy najbliższej mu wartości wśród predykowanych Y^{full} . Algorytm zazwyczaj wybiera 3 najbliższe wartości.*
5. *Losowo wybieramy jedną z tych wartości i imputujemy brak wartością zmiennej Y , dla której predykcja została wylosowana.*

PMM – SCHEMAT DZIAŁANIA

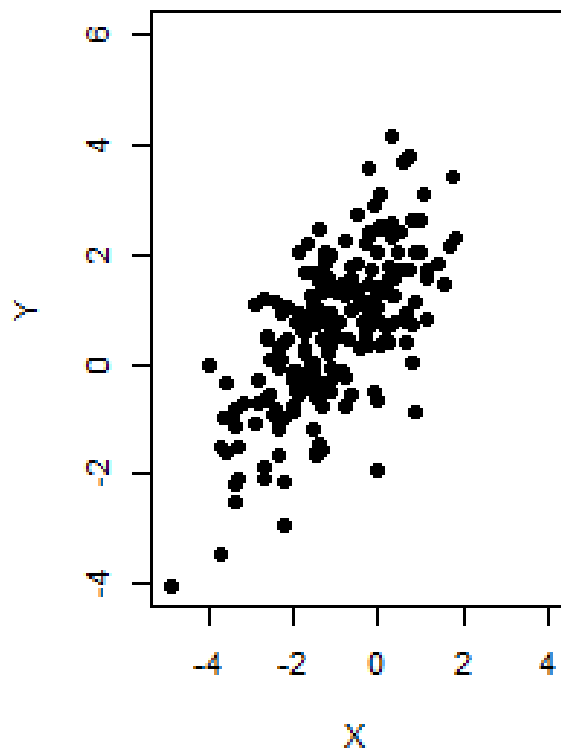


PMM – PRZYKŁAD SKUTECZNOŚCI

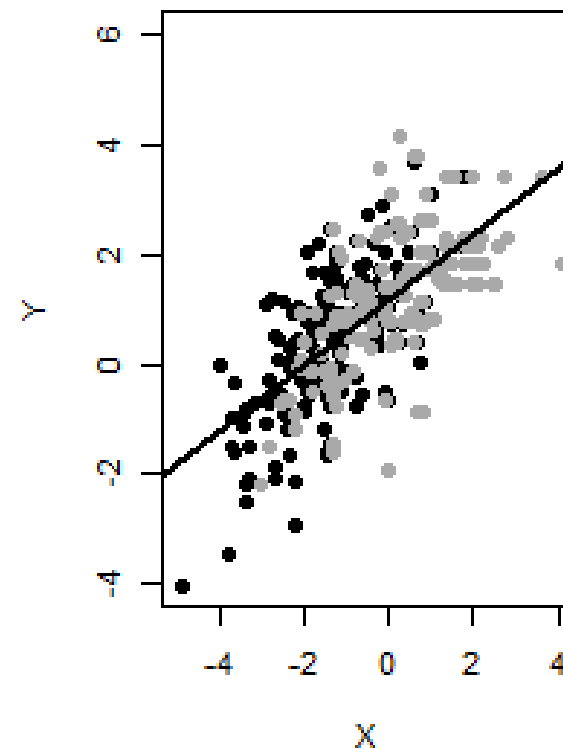
oryginalny zbiór



po usunięciu braków

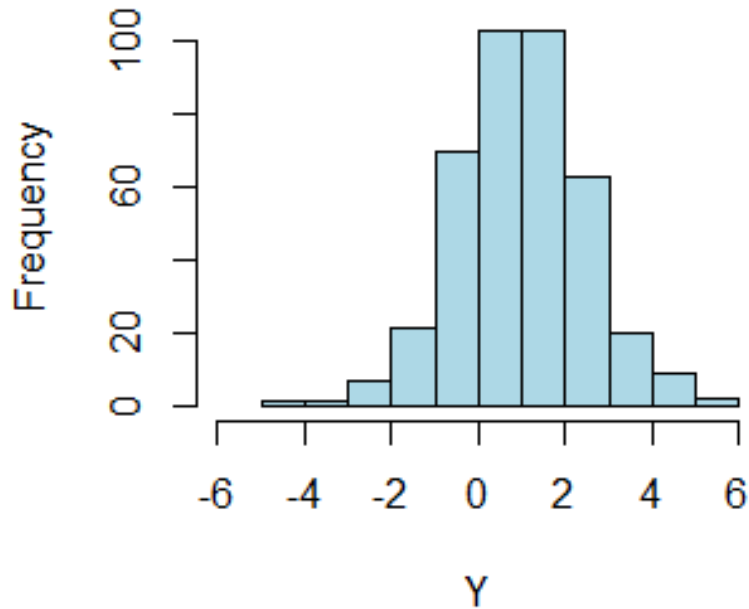


po imputacji pmm

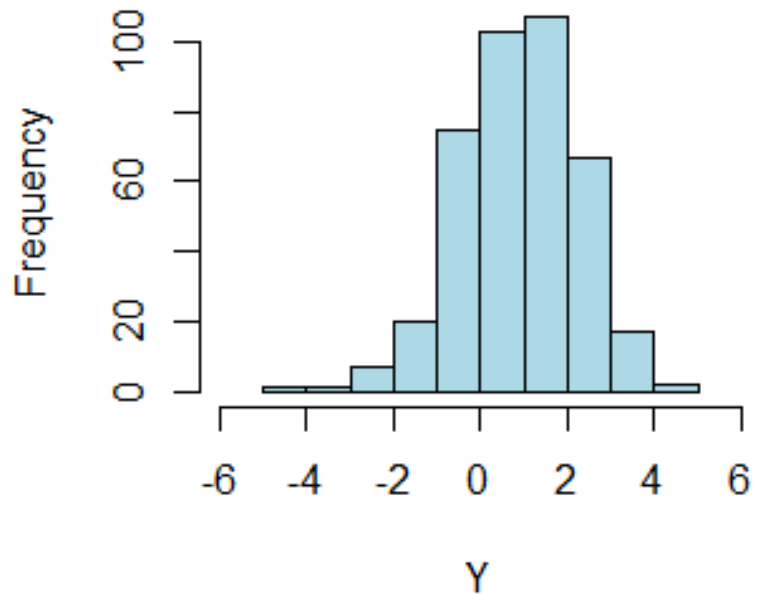


PMM – ROZKŁAD ZMIENNEJ

**HISTOGRAM Y
oryginalny**



**HISTOGRAM Y
po imputacji**



PODSUMOWANIE

PMM

Wielokrotne
imputacje

Regresja
stochastyczna

Imputacja
średnią

Imputacja
regresyjna



Dziękuję za uwagę!