# Mikroekonometria 2

Mikołaj Czajkowski
Wiktor Budziński

# Klasyczny Model Regresji Liniowej (KMRL)

- Postać modelu regresji liniowej: $y_i = \mathbf{X}_i\boldsymbol{\beta} + \varepsilon_i$
  - Modelujemy liniową zależność $y$ od zmiennych objaśniających $\mathbf{X}$
  - Parametry modelu $\boldsymbol{\beta}$ są nieznane, ale możemy je oszacować za pomocą posiadanych danych $y, \mathbf{X}$
  - Domyślnie szacujemy parametry skupiając się na średnich $y$ względem $\mathbf{X}$, ale można też skupiać się na medianie lub innych kwantylach
- Estymator – reguła stosowana w celu znalezienia oszacowania jakiejś wielkości na podstawie danych
- Estymacja parametrów – Metoda Najmniejszych Kwadratów (MNK)
  - Metoda szukania parametrów $\boldsymbol{\beta}$, polegająca na tym, żeby nasza liniowa funkcja dawała jak najmniejsze kwadraty różnic między przewidywanymi a obserwowanymi wartościami
  - Wizualizacja

(cc) BY czaj.org

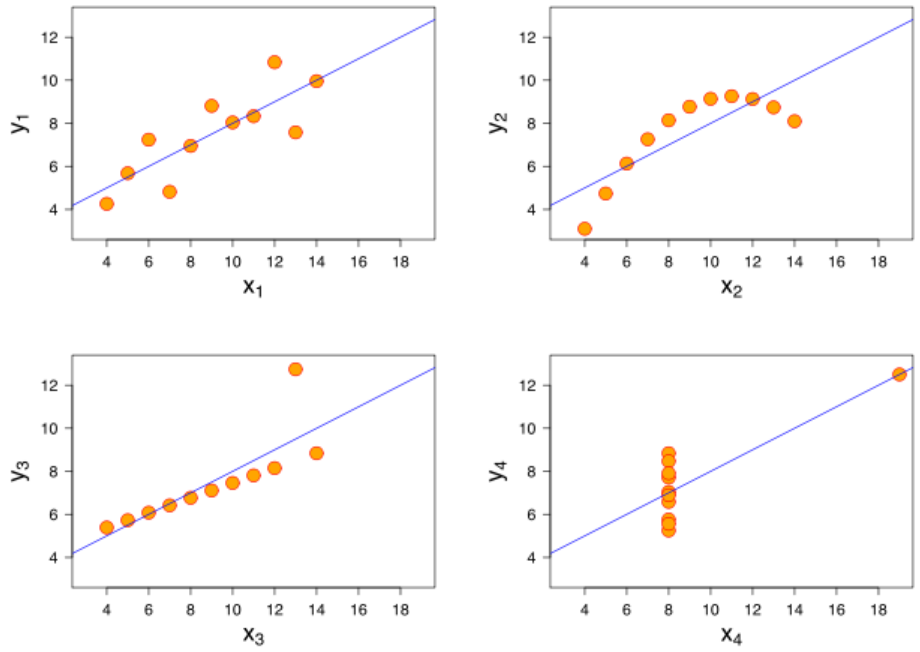
# Metoda Najmniejszych Kwadratów (MNK)

- Jeśli spełnione są określone założenia (wymogi) to estymator MNK
  1. Nieobciążony (wartość oczekiwana szacowanych parametrów będzie równa prawdziwej wartości)
  2. Zgodny (im większa próba, tym szacowane parametry będą bliższe prawdziwej wartości)
  3. Efektywny (estymator jest efektywny gdy ma najmniejszą wariancję ze wszystkich estymatorów)

(cc) BY czaj.org

# Założenia (wymogi) KMRL

1. Liniowa postać funkcji jest faktycznie prawidłowa




- Jeśli tak nie jest
  - Można estymować model nieliniowy (innymi metodami estymacji niż MNK)
  - Można spróbować dokonać jakiejś transformacji zmiennych objaśnianej lub objaśniających (np. logarytm, kwadrat), która da zależność bliższą liniowej

(cc) BY czaj.org

# Zadanie 1. Nieliniowa postać funkcyjna

1. Wygeneruj sztuczne dane, w których zmienna objaśniana zależy nieliniowo od jakiejś zmiennej objaśniającej
2. Sprawdź czy stosując KMRL i zakładając liniową specyfikację można odzyskać parametry założonego PGD
3. Zastosuj KMRL uwzględniający odpowiednią transformację

(cc) BY czaj.org

# Założenia (wymogi) KMRL

2. Zmienne objaśniające są liniowo niezależne
   - Żadnej ze zmiennych objaśniających nie da się wyrazić za pomocą liniowej kombinacji pozostałych
   - Zwykle zakłada się także, że ich średnie i wariancje są skończone
   - Jeśli jakieś zmienne objaśniające są współliniowe – nie da się dla nich określić jednoznacznych parametrów

(cc) BY czaj.org

# Zadanie 2. Współliniowe zmienne objaśniające

1. Wygeneruj sztuczne dane, w których zmienna objaśniana zależy od jakiejś zmiennej objaśniającej rozkodowanej na zmienne zerojedynkowe (w zależności od jej wartości)
2. Sprawdź co się stanie, jeśli w modelu uwzględnić wszystkie zmienne zerojedynkowe.
    1. Co się stanie jeśli pominąć jedną lub więcej zmiennych?
    2. Co się stanie jeśli pominąć stałą?

(cc) BY czaj.org

# Założenia (wymogi) KMRL

3. Błędy są sferyczne $\operatorname{var}\left(\varepsilon \middle| \mathbf{X}\right) = \sigma^2 I_n$
   - Brak autokorelacji $\underset{i \neq j}{\forall} E\left(\varepsilon_i \varepsilon_j \middle| \mathbf{X}\right) = 0$ czyli $\operatorname{cov}\left(\varepsilon_i, \varepsilon_j\right) = 0$
     - Błędy pomiędzy obserwacjami są nieskorelowane
     - Może być niespełnione np. dla szeregów czasowych, danych panelowych itp.
     - Jeśli tak nie jest – estymator MNK nadal jest nieobciążony i zgodny, ale jest nie już efektywny
       - Można zastosować np. Uogólnioną MNK
   - Homoskedastyczność
     - Składnik losowy ma tę samą wariancję dla wszystkich obserwacji $E\left(\varepsilon_i^2 \middle| \mathbf{X}\right) = \sigma^2$
     - Jeśli tak nie jest – estymator MNK nadal jest nieobciążony i zgodny, ale jest nie już efektywny
       - Można zastosować np. Ważoną MNK lub odporne macierze kowariancji
   - Czasem zakłada się dodatkowo normalność składnika losowego
     - Nie jest niezbędne, ale jest potrzebne do dowodów własności testów dla skończonych prób
     - Jeśli błędy są normalne to estymator MNK jest ekwiwalentny do estymatora Metody Największej Wiarygodności (MNW)

(cc) BY czaj.org

# Zadanie 3. Heteroskedastyczność

1. Wygeneruj sztuczne dane, w których wariancja błędu losowego zależy od wartości zmiennej objaśniającej
2. Sprawdź, czy KMRL nadal prawidłowo odzyskuje parametry PGD

(cc) BY czaj.org

# Założenia (wymogi) KMRL

4. Zmienne objaśniające są egzogeniczne
   - Są znane z pewnością, nie trzeba traktować ich jak zmienne losowe (np. z powodu błędów pomiaru)
   - Nie są skorelowane z błędem losowym $E(\varepsilon|\mathbf{X})=0$
   - Czyli (m.in.) błędy losowe mają średnią 0
   - Jeśli tak nie jest (są endogeniczne, czyli korelacja $y$ i $\mathbf{X}$ jest możliwa naturalnie, bez zależności przyczynowo-skutkowej) estymator MNK przestaje być zgodny
     - Można zastosować metodę zmiennych instrumentalnych (znaleźć zmienne, które wpływają na $\mathbf{X}$, ale nie mają wpływu na $y$)
     - Można zastosować metodę analizy wielomianowej (ang. *multivariate analysis*), w której dowolna zmienna może być traktowana jako objaśniana

(cc) BY czaj.org

# Zadanie 4. Endogeniczność

1. Wygeneruj model z dwiema zmiennymi objaśniającymi, które są ze sobą skorelowane
2. Sprawdź, czy KMRL prawidłowo odzyskuje parametry PGD, kiedy nie uwzględnimy jednej z tych zmiennych w modelu

(cc) BY czaj.org

# Case study – determinanty cen wina

- Costanigro, M., Mittelhammer, R. C., and McCluskey, J. J. (2009). ***Estimating class-specific parametric models under class uncertainty: local polynomial regression clustering in an hedonic analysis of wine markets***. Journal of Applied Econometrics, 24(7), 1117-1135.
  - Dane dotyczące cen i charakterystyk win ze stanów Kalifornia i Waszyngton w latach 1991-2000
  - Analiza hedoniczna – analiza przy pomocy regresji, w której wartość dobra (przybliżona ceną) jest rozbita na wartość poszczególnych charakterystyk danego dobra

CC BY czaj.org

## Zadanie 5. Wykorzystując dane me.wine.dta przygotuj model regresji hedonicznej dla cen wina

1. Przeprowadź regresję objaśniającą ceny wina skupiając się na wpływie liczbie wyprodukowanych skrzynek wina oraz ocenie jego smaku
   1. Zacznij od prostego modelu regresji liniowej
      - Czy założenia KMRL są spełnione?
   2. Dodaj do modelu efekty nieliniowe
   3. Dodaj do modelu interakcje między wybranymi zmiennymi

CC BY czaj.org