

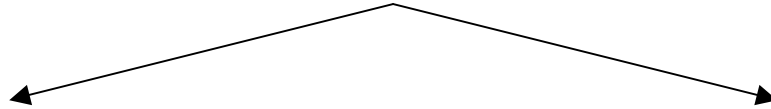
## Analiza regresji prostej – podstawowe wiadomości

Model zależności liniowej:  $Y = \beta_0 + \beta_1 X + \varepsilon$ ,  $\varepsilon \sim N(0, \sigma^2)$

Czy istnieje związek liniowy między  $Y$  i  $X$ ?

**H:  $\beta_1 = 0$  (nie zachodzi związek liniowy między  $Y$  i  $X$ )**

**K:  $\beta_1 \neq 0$  (zachodzi związek liniowy między  $Y$  i  $X$ )**



przyjmujemy H



koniec analizy regresji

odrzucaamy H



- budowa modelu z oszacowanymi parametrami  $\beta_0$  i  $\beta_1$
- miara siły dopasowania  $R^2$
- prognozowanie
- sprawdzenie założenia o normalności reszt

Analizuj >> Regresja >> Liniowa

Role zadania:  $Y$  >> zmienne zależne,  $X$  >> zmienne objaśniające

Model: Dopasowanie całego modelu

Wykresy przewidywane: Obserwowane a niezależne

Wykresy reszta: Zwyczajne a niezależne

Prognozy: Próba pierwotna i Dane dodatkowe ‘podać ścieżkę do zbioru z prognozą’, Reszta, Granice prognozy, Zapisz dane wynikowe: Prognozy ‘podać ścieżkę do zbioru, w którym zapisane będą reszty i wynik prognozy, wyświetl wynik

URUCHOM

Uwagi:

1) Przed przystąpieniem do analizy regresji prostej warto ocenić ogólny charakter zależności za pomocą wykresu rozproszenia (punktowego) i współczynnika korelacji.

- H:  $\rho = 0$  (nie zachodzi związek liniowy między  $Y$  i  $X$ )

K:  $\rho \neq 0$  (zachodzi związek liniowy między  $Y$  i  $X$ )

- wielkość współczynnika  $\rho$  informuje o sile zależności.

Analizuj >> Korelacje

Role zadania:  $Y$  >> zmienne analizowane,  $X$  >> koreluj z

Opcje: korelacja Pearsona

Rezultaty: utwórz wykresy punktowe

URUCHOM

2) W przypadku, gdy nie mamy podanego rodzaju zależności, proponujemy postać modelu na podstawie wykresu rozproszenia (używając transformacji zmiennych).

3) W przypadku regresji prostej używanie testu  $t$  w celu zbadania istotności zmiennej  $X$  jest równoważne z zastosowaniem testu  $F$  do weryfikacji hipotezy o istnieniu związku liniowego.

4) Dane na podstawie, których prognozuje się daną wartość zapisujemy w oddzielnym zbiorze danych lub razem z danymi zadania (wtedy zaznaczamy tylko Próbę pierwotną).

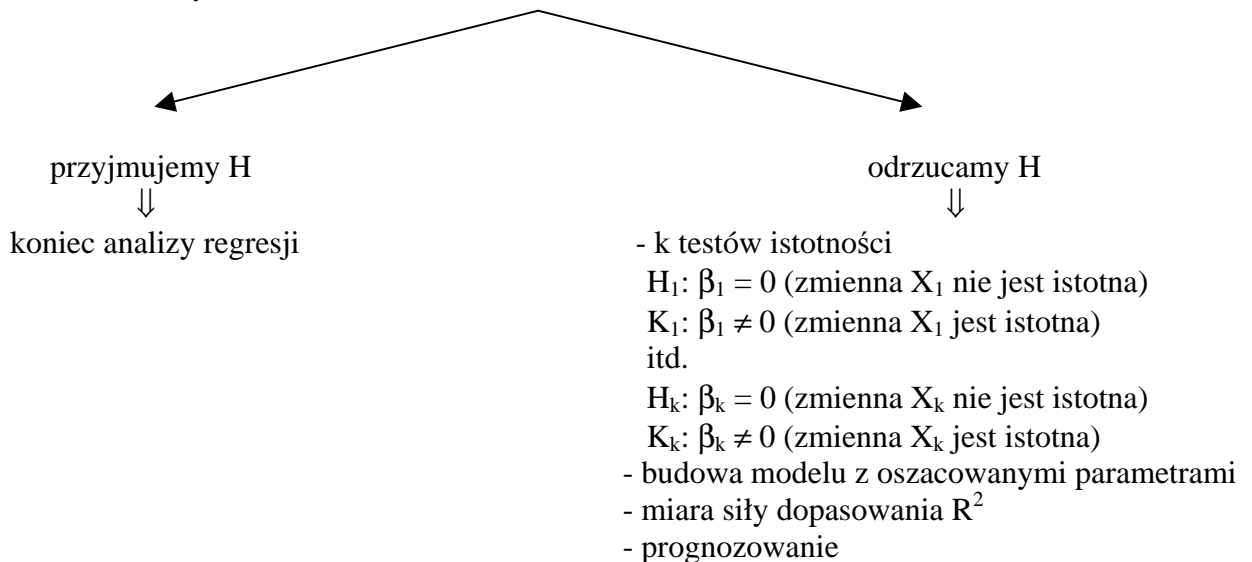
## Analiza regresji wielorakiej – podstawowe wiadomości

Model zależności liniowej:  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$ ,  $\varepsilon \sim N(0, \sigma^2)$

Czy istnieje związek liniowy między  $Y$  a którąkolwiek z  $X_1, X_2, \dots, X_k$ ?

**H:**  $\beta_1 = \beta_2 = \dots = \beta_k = 0$  (nie zachodzi związek liniowy między Y a którąkolwiek z  $X_i$ ,  $i=1, \dots, k$ )

**K:** nie wszystkie  $\beta_i$  są równe 0 (zachodzi związek liniowy między Y a którąkolwiek z  $X_i$ )



Analizuj >> Regresja >> Liniowa

Role zadania: Y >> zmienna zależna,  $X_1, X_2, \dots, X_k$  >> zmienne objaśniające

Prognozy: Próba pierwotna i Dane dodatkowe 'podać ścieżkę do zbioru z prognozą', Reszta, Granice prognozy, Zapisz dane wynikowe: Prognozy 'podać ścieżkę do zbioru, w którym zapisane będą reszty i wynik prognozy, wyświetl wynik

URUCHOM

Uwagi:

1) W sytuacji, gdy nie wszystkie zmienne objaśniające są w modelu istotne dokonujemy selekcji zmiennych, tj. usuwamy z modelu zmienne nieistotne. Możemy to uczynić m.in. za pomocą następujących metod:

A. „ręczne” usuwanie zmiennych (*full model*; w SAS: Dopasowanie całego modelu)

- W Rolach zadania uwzględniamy tylko istotne zmienne objaśniające i dla tych zmiennych tworzymy model regresji

B. metoda dołączania (*forward selection*; w SAS: Wybór następnych)

- krok 1: Start od modelu zawierającego tylko stałą.

- krok 2: Wybierana jest ta zmienna spośród możliwych, dla której p-value odpowiadającego jej testu t jest najmniejszą p-value  $< \alpha$ .

- krok 3: Rozpatrując wszystkie możliwe zmienne nie znajdujące się w modelu powtarza się krok 2.

STOP, kiedy żadnemu z potencjalnych kandydatów na włączenie do modelu nie odpowiada p-value  $< \alpha$ .

C. metoda eliminacji (*backward selection*; w SAS: Eliminacja poprzednich)

- krok 1: Uwzględnione są wszystkie potencjalnie interesujące nas zmienne.

- krok 2: Zakładając prawdziwość tego modelu, testowane są indywidualne hipotezy o istotności poszczególnych zmiennych i usuwana zostaje ta zmienna, dla której p-value odpowiadającego testu t jest największą p-value  $> \alpha$ .

- krok 3: Dopasowywany jest mniejszy model z usuniętą zmienną i powrót do kroku 2

STOP, gdy w pewnym kroku wszystkie p-value  $< \alpha$ .

D. metoda krokowa (w SAS: Wybór krokowy)

- połączenie metody dołączeń i eliminacji; często jest to metoda najbardziej efektywna.

2) W procedurze SAS wybierając metodę dołączania, (o ile nie zostaną podane inne wartości poziomu istotności), podajemy poziom istotności wstawiania do modelu 0.05; wybierając metodę eliminacji podajemy poziom istotności pozostania w modelu 0.01; zaś wybierając metodę krokową wpisujemy odpowiednio 0.05 i 0.01.