MODELE Z INTERAKCJAMI I WIELOMIANAMI ZM. OBJAŚNIANEJ.

ZAJĘCIA NR 9.

Wszystkie ćwiczenia pochodzą z podręczników Wooldridge lub Stock & Watson. Pod tym linkiem znajduje się dokument z opisem zmiennych zbiorów danych, potrzebnych do rozwiązywania zadań.

I. Modele z wielomianami.

1. Skorzystaj z danych WAGE1 w tym ćwiczeniu. (W, C2 s. 202)

   (i) Oszacuj metodą MNK poniższe równanie:

   $$log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + u$$

   Zapisz wynik w formie równania.

   Rozwiązanie:

   $$\widehat{\log(wage)} = 0.1279975 + 0.0903658educ + 0.0410089exper - 0.0007136exper^2$$

   (ii) Używając przybliżenia, znajdź (przybliżony) zwrot z piątego roku doświadczenia. Jaki jest przybliżony zwrot z dwudziestego roku doświadczenia?

   Rozwiązanie: Przybliżony zwrot z doświadczenia możemy obliczyć biorąc pochodną z równania regresji względem $exper$:

   $$\frac{\partial \widehat{\log(wage)}}{exper} = 0.0410089 - 2 * 0.0007136exper$$

   Zwrot z piątego roku doświadczenia wynosi:

   $$(\frac{\partial \widehat{\log(wage)}}{\partial exper})100\% = (0.0410089 - 2 * 0.0007136 * 4) * 100\% = 3.53\%$$

   Zwrot z piątego roku doświadczenia wynosi:

   $$(\frac{\partial \widehat{\log(wage)}}{\partial exper})100\% = (0.0410089 - 2 * 0.0007136 * 19) * 100\% = 1.39\%$$

   Zwroty z każdego następnego roku doświadczenia są coraz mniejsze.

   (iii) Przy jakiej wartości *exper*, zwiększenie doświadczenia obniża przewidywany logarytm płacy? Ile osób ma więcej (niż ta wartość) doświadczenia w próbie?

   Rozwiązanie: Według modelu, logarytm płacy jest kwadratową funkcją doświadczenia. Zatem należy sprawdzić od jakiej wartości krzywa obniża się (innymi słowy znaleźć punkt w którym funkcja ma maksimum). w tym celu należy przyrównać warunek pierwszego rzędu do zera:

   $$\frac{\partial \widehat{\log(wage)}}{exper} = 0.0410089 - 2 * 0.0007136exper = 0$$

   $$exper^\star = 28.73$$

   Według oszacowań, 29-ty rok doświadczenia zacznie zmniejszać płace. Jest 121 takich osób w próbie, a próba liczy ok. 520 osób.

2. Użyj danych BWGHT2 w tym ćwiczeniu (W, C10 p. 223).

 (i) Oszacuj równanie MNK:

$$\log(bwght) = \beta_0 + \beta_1 npvis + \beta_2 npvis^2 + u$$

 zapisz wyniki w formie równania.

 Rozwiązanie:

$$\widehat{\log(bwght)} = 7.9578827 + 0.0189167 npvis - 0.0004288 npvis^2$$

 (ii) Pokaż korzystając z oszacowania z (i), że liczba wizyt prenatalnych która maksymalizuje *log(bwght)* wynosi około 22. Jak wiele kobiet miało przynajmniej 22 wizyty w próbie?

 Rozwiązanie: Korzystamy z warunku pierwszego rzędu względem $npvis$:

$$\frac{\partial \widehat{\log(bwght)}}{npvis} = 0.0189167 - 2 * 0.0004288 npvis = 0$$

$$npvis^\star = 22.06 \approx 22$$

 Przy 22 wizytach, waga dziecka przy urodzeniu zaczyna spadać. Jest 21 kobiet w próbie z co najmniej 22 wizytami.

 (iii) Czy ma to sens, że waga urodzenia spada po przekroczeniu 22 wizyt prenatalnych? Wyjaśnij. Rozwiązanie: Może to być związane z tym, że matki które często odwidzają lekarza mają problemy z ciążą. W takiej sytuacji kształt oszacowanej krzywej wydaje się mieć sens.

 (iv) Dodaj wiek matki do równania, używając funkcji kwadratowej. Przy *npvis* ustalonym, czy jakim wieku matki waga przy narodzinach dziecka jest największa? Ile jest w próbie kobiet starszych niż obliczony optymalny wiek?

 Rozwiązanie:

$$\widehat{\log(bwght)} = 7.5837127 + 0.0180374 npvis - 0.0004079 npvis^2 + 0.0253920 mage - 0.0004119 mage^2$$

 aby znaleźć optymalny wiek matki, skorzystajmy warunek pierwszego rzędu względem $mage$:

$$\frac{\partial \widehat{\log(bwght)}}{npvis} = 0.0253920 mage - 2 \times 0.0004119 mage = 0$$

$$mage^\star = 30.82546$$

 Optymalny wiek matki maksymalizujący wagę dziecka przy urodzeniu wynosi ok. 31 lat. W próbie są 746 kobiety, starsze niż 31 lat.

 (v) Czy wiek matki oraz liczba wizyt prenatalnych wyjaśnia dużo zróżnicowania w *log(bwght)*?

 Rozwiązanie: $R^2 = 2.5\%$ - raczej niewiele.

3. Użyj danych `GPA2` do tego zadania (W, C4 p. 221).

   (i) Oszacuj model

   $$sat = \beta_0 + \beta_1 hsize + \beta_2 hsize^2 + u$$

   gdzie *hsize* to rozmiar klasy absolwentów (w setkach), oraz zapisz model w postaci równania.

   Rozwiązanie:

   $$\widehat{sat} = 997.981 + 19.814hsize - 2.131hsize^2$$

   (ii) Używając oszacowań z (i) powiedz jaki jest optymalny rozmiar klasy? Uzasadnij. Rozwiązanie: Korzystamy z warunku pierwszego rzędu względem $hsize$:

   $$\frac{\partial \widehat{sat}}{hsize} = 19.814 - 2 \times 2.131hsize$$
   $$hsize^\star = 4.649958$$

   Pamiętając że $hsize$ jest podana w setkach można stwierdzić że, optymalny rozmiar klasy wynosi ok. 465 osób.

   (iii) Czy ta analiza jest reprezentatywnie przedstawia poziom osiągnięć akademickich wśród starszych uczniów liceum? Wyjaśnij.

   Rozwiązanie: W próbie mamy najprawdopodobniej tylko tych uczniów, który podeszli do testu. Zatem aby analiza była reprezentatywna, należałoby analizować losową próbę ze wszystkich uczniów który pisaliby ten sam test.

   (iv) Znajdź optymalny poziom klasy, używając *log(sat)* jako zmienną zależną. Czy jest różny od tego otrzymanego w (iii)?

   Rozwiązanie: Gdy powtórzymy obliczenia dla *log(sat)* jako zmiennej zależnej, otrzymamy bardzo podobny wynik, mianowicie 469 uczniów.

II. Modele z interakcjami

1. Użyj danych GPA2 (W, 3 p. 259).

   (i) Oszacuj równanie

$$sat = \beta_0 + \beta_1 hsize + \beta_2 hsize^2 + \beta_3 female^2 + \beta_4 black + \beta_5 female \times black + u$$

   Zmienna *sat* to łączny wynik testu SAT, *hsize* to wielkość klasy absolwentów (w setkach), *female* to zmienna zero-jedynkowa oraz *black* to zmienna zero-jedynkowa oznaczająca rasę.
   Korzystając z równania wyznacz optymalną wielkość klasy.
   Rozwiązanie:

$$\widehat{sat} = 1028.097 + 19.297 hsize - 2.195 hsize^2 - 45.091 female^2 - 169.813 black + 62.306 female \times black$$

   Korzystamy z waruku pierwszego rzędu, aby wyznaczyć optymalny rozmiar klasy:

$$\frac{\partial \widehat{sat}}{hsize} = 19.297 hsize - 2 \times 2.195 hsize = 0$$

$$hsize^{\star} = 4.39603$$

   Optymalny rozmiar klasy wynosi około 440 osób.

   (ii) Przy niezmienionym *hsize*, jaka jest szacowana różnica w wyniku SAT między nieczarnymi kobietami i czarnymi mężczyznami?
   Rozwiązanie:

$$\mathbb{E}[sat \mid female = 1, black = 0] - \mathbb{E}[sat \mid female = 0, black = 1] = 124.7212$$

   Białe kobiety uzyskują średnio lepszy wynik o 124 punkty niż czarni mężczyźni, przy innych czynnikach niezmienionych.

   (iii) Jaka jest szacowana zmiana w wyniku SAT między czarnymi mężczyznami a pozostałymi mężczyznami?
   Rozwiązanie:

$$\mathbb{E}[sat \mid female = 0, black = 1] - \mathbb{E}[sat \mid female = 0, black = 0] = -169.8126$$

   Czarni mężczyźni uzyskują średnio gorszy wynik od białych mężczyzn o ok. 170 punktów, przy innych czynnikach niezmienionych.

   (iv) Jaka jest szacowana różnica między czarnymi kobietami a pozostałymi kobietami?
   Rozwiązanie:

$$\mathbb{E}[sat \mid female = 1, black = 1] - \mathbb{E}[sat \mid female = 1, black = 0] = -107.5063$$

   Czarne kobiety uzyskują gorszy wynik od białych kobiet średnio o 107.5 punktu, przy innych czynnikach niezmienionych.

2. Użyj danych `TWOYEAR` (W, 9 p. 261)

   (i) Oszacuj równanie

$$\log(wage) = \beta_0 + \beta_1 female + \beta_2 totcoll + \beta_3 female \times totcoll$$

   Używając oszacowań, znajdź wartości *totcoll* takie że przewidywane wartości *log(wage)* są takie same dla kobiet i mężczyzn.

   Rozwiązanie:

$$\widehat{\log(wage)} = 2.28877 - 0.35727 female + 0.04966 totcoll + 0.02991 female \times totcoll$$

   Aby znaleźć wartość $totcoll$ dzięki której płace kobiet i mężczyzn będą równe, należy rozpatrzeć przypadki oszacowań dla kobiet i mężczyzn i porównać je:

$$\mathbb{E}[\log(wage) \mid female = 0, totcoll] = 2.28877 + 0.04966 totcoll$$
$$\mathbb{E}[\log(wage) \mid female = 1, totcoll] = 2.28877 - 0.35727 + 0.04966 totcoll + 0.02991 totcoll$$

   gdy porónamy prawe strony powyższych wyrażeń, otrzymamy:

$$-0.35727 + 0.02991 totcoll = 0$$
$$totcoll = 11.94483 \approx 12$$

   Płace kobiet zrównałyby się z płacami mężczyzn gdyby kobiety studiowały prawie 12 lat dłużej.

   (ii) Korzystając z oszacowania z (i) powiedz, czy kobiety naprawdę mogą uzyskać tyle lat edukacji, aby ich zarobki dogoniły zarobki mężczyzn? Wyjaśnij.

   Rozwiązanie: Teoretycznie 12 lat edukacji więcej po stronie kobiet domknęło by lukę płacową, ale z praktycznego punkty widzenia, ten wynik jest zbyt wysoki aby miał jakieś zastosowanie.

3. Poniższy model pozwala aby zwrot z edukacji zależał od edukacji rodziców (*pareduc*) (W, 4 p. 218):

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 educ \times pareduc + \beta_3 exper + \beta_4 tenure + u.$$

(i) Pokaż że w przybliżeniu zwrot z dodatkowego roku edukacji można opisać formuła:

$$\frac{\Delta log(wage)}{\Delta educ} = \beta_1 + \beta_2 pareduc$$

Jakiego znaku spodziewasz się przy β_2?

<u>Rozwiązanie:</u> Przy innych czynnikach niezmienionych, otrzymujemy:

$$\Delta \log(wage) = \beta_1 \Delta educ + \beta_2 \Delta educ \times pareduc = (\beta_1 + \beta_2 pareduc)\Delta educ$$

dzieląc obie strony przez $\Delta educ$:

$$\frac{\Delta \log(wage)}{\Delta educ} = \beta_1 + \beta_2 pareduc$$

Znak β_2 nie jest oczywisty, gdyby $\beta_2 > 0$ to oznaczałoby to że dziecko tym bardziej korzysta z kolejnego roku edukacji, im bardziej wyedukowani są rodzice dziecka.

(ii) Używając danych *WAGE2* oszacuj model. Zinterpretuj współczynnik przy interakcji. Pomocnym może być wybranie dwóch wartości zmiennej *pareduc* np.: $pareduc = 32$ (obaj rodzice mają wykształcenie wyższe), lub $pareduc = 24$ (obaj rodzice mają wykształcenie średnie) i porównaj oszacowany zwrot z edukacji.

<u>Rozwiązanie:</u> Oszacowanie modelu:

$$\log\widehat{(wage)} = 5.646519 + 0.046752 educ + 0.000775 educ \times pareduc + 0.018871 exper + 0.010217 tenure$$

Różnica w zwrocie z edukacji dziecka, pomiędzy rodzicami mającymi wykształcenie wyższe i średnie:

$$\left(\frac{\Delta \log(wage)}{\Delta educ} \mid totcoll = 32\right) - \left(\frac{\Delta \log(wage)}{\Delta educ} \mid totcoll = 24\right) =$$
$$0.046752 + 0.000775 \times 32 - 0.046752 + 0.000775 \times 24 = -0.006199767$$

czyli zwrot z edukacji dzieci których rodzice mają wykształcenie wyższe jest średnio niższy o ok. 0.62% niż dzieci których rodzice mają wykształcenie średnie.

(iii) Dodaj *pareduc* jako oddzielną zmienną i oszacuj model. Czy teraz zależność między zwrotem z edukacji a edukacją rodziców jest pozytywna?

<u>Rozwiązanie:</u> Szacujemy model uwzględniając zmienną *pareduc* i obliczamy wpływ edukacji rodziców na zwrot z edukacji dziecka, tak jak w poprzednim podpunkcie. Teraz wynosi on 1.35%.