

Prosta regresja liniowa.

Sebastian Zalas

s.zalas@uw.edu.pl

Zajęcia nr 4.

Wszystkie ćwiczenia pochodzą z podręczników Woolridge oraz Stock & Watson.

1. Niech *kids* oznacz liczbę dzieci urodzonych przez kobietę, i *educ* oznacza lata edukacji kobiety. Oto prosty model, wiążący dzietność z latami edukacji:

$$kids = \beta_0 + \beta_1 educ + u$$

gdzie u to nieobserwowany składnik losowy.

- (i) Jakie czynniki znajdują się w u ? Czy są one skorelowane z poziomem edukacji?
 - (ii) Czy na podstawie prostej regresji możemy ocenić efekt *ceteris paribus* edukacji na dzietność? Wyjaśnij.
2. Tabela zawiera wyniki *ACT* oraz *GPA* (grade point average) dla 8 studentów. *GPA* bazuje na 4-stopniowej skali i zostało zaokrąglone do jednego miejsca dziesiętnego.

Student	GPA	ACT
1	2.8	21
2	3.4	24
3	3.0	26
4	3.5	27
5	3.6	29
6	3.0	25
7	2.7	25
8	3.7	30

- (i) Oszacuj zależność między *GPA* a *ACT* za pomocą OLS, tzn. poszukaj oszacowań wyrazu wolnego i nachylenia. Skomentuj kierunek relacji. Czy stała ma sensowną interpretację? Wyjaśnij. O ile wzrośnie *GPA* jeśli wynik *ACT* wzrośnie o 5ptk.?
 - (ii) Oblicz wartości teoretyczne oraz reszty dla każdej obserwacji, sprawdź czy suma reszt jest równa 0.
 - (iii) Jaka będzie wartość teoretyczna *GPA* kiedy *ACT* = 20
 - (iv) Ile zmienności w *GPA* dla tych ośmiu studentów jest wyjaśnione przez *ACT*. Wyjaśnij.
3. Pokaż, że prosta regresji przechodzi przez punkt (\bar{x}, \bar{y}) .
 4. Zbiór danych BWGHT.RAW zawiera dane o urodzeniach oraz kobietach w USA. Zmienna zależna to waga noworodka w uncjach (*bwght*), a zmienną objaśniającą jest średnia liczba wypalonych przez matkę w ciągu dnia podczas ciąży (*cigs*). Poniższe równanie regresji zostało oszacowane, używając danych o 1,388 urodzeniach:

$$\hat{bwght} = 119.77 - 0.514 \times cigs$$

- (i) Jaka jest teoretyczna liczba urodzeń, przy $cigs = 0$? Co w przypadku gdy $cigs = 20$ (jedna paczka dziennie)? Skomentuj różnicę.
- (ii) Czy na podstawie powyższej regresji można powiedzieć coś o *przyczynowym* związku pomiędzy wagą noworodka a przywyczajeniem do palenia matki? Wyjaśnij.
- (iii) Przy jakiej wartości zmiennej $cigs$, waga noworodka wyniesie 125 uncji? Skomentuj.
- (iv) Odsetek kobiet w próbie, które nie palą podczas ciąży wynosi 0.85. Czy to zmienia twój pogląd na odpowiedź w punkcie (iii)?

5. W liniowej funkcji konsumpcji

$$c\hat{o}ns = \hat{\beta}_0 + \hat{\beta}_1 \times inc$$

(wyestymowana) krańcowa skłonność do konsumpcji (*Marginal propensity to consume, MPC*) to nachylenie prostej, czyli współczynnik β_1 , podczas gdy średnia skłonność do konsumpcji (*APC*) to $\frac{c\hat{o}ns}{inc} + \hat{\beta}_1$. Używając danych o rocznym dochodzie stu rodzin, oraz ich konsumpcji (obie zmienne mierzone w dolarach), otrzymano poniższe oszacowania:

$$c\hat{o}ns = -124.84 + 0.853 \times inc$$

$$n = 100, R^2 = 0.692.$$

- (i) Zinterpretuj wartość stałej w powyższym równaniu oraz skomentuj jej znak oraz wielkość.
 - (ii) Jaka jest teoretyczna wartość konsumpcji gdy dochód rodziny wynosi \$30,000?
 - (iii) Narysuj wykres wyestymowaną funkcję *MPC* w zależności *APC*, odłóż na osi x zmienną inc .
6. Badacz używając danych o rozmiarze klasy (CS) oraz średnim wyniku testu ze 100 klas, wyestymował poniższe równanie:

$$Test\hat{S}core = 520.4 - 5.82 \times CS$$

$$R^2 = 0.08, SER = 11.5.$$

- (i) Klasa liczy 22 uczniów. Jaka jest wartość przewidywana średniego wyniku testu?
 - (ii) W ostatim roku klasa liczyła 19 uczniów w tym roku liczy ona 23 studentów. Jaka jest wartość teoretyczna średniego wyniku testu?
 - (iii) Średni rozmiar klasy w próbie 100 klas wynosi 21.4. Jaki jest średni wynik testu w próbie 100 klas?
 - (iv) Jakie jest odchylenie standardowe wyniku testu w próbie 100 klas? (Przypomnij sobie wzory na R^2 and SER .)
7. Wylosowano próbę 200-stu dwudziestoletnich mężczyzn z populacji. Zebrano dane o wzroście i wadze. Otrzymano poniższe oszacowanie:

$$w\hat{a}ga = -99.41 + 3.94 \times wzrost$$

$$R^2 = 0.81, SER = 10.2, \text{ gdzie } waga \text{ jest mierzona w funtach oraz } wzrost \text{ jest mierzony w calach.}$$

- (i) Jaka jest przewidywana waga według regresji dla kogoś kto ma 70 cali wzrostu? 65 cali? 74 cali?
- (ii) Jeden z mężczyzn rośnie 1.5 cala w ciągu roku. Jaki jest przewidywany przyrost wagi?

- (iii) Zamiast zmierzenia wagi w funtach i wzrostu w calach, zmienne zostały zmierzone w centymetrach oraz kilogramach. Jakie są oszacowania z nowej regresji? (Podaj oszacowania parametrów, R^2 oraz SER.)
8. (i) Pokaż, że jeżeli oszacowanie regresji liniowej $\beta_1 = 0$. Pokaż, że wtedy $R^2 = 0$.
(ii) R^2 wynosi 0. Czy wtedy oszacowany parametr regresji β_0 musi być równy 0?
9. Mamy model regresji liniowej: $y = \beta_0 + \beta_1 x + u$, oraz założmy że $\mathbb{E}(u) = 0$. Pokaż, że nowy model w sytuacji gdy $\mathbb{E}(u) = \alpha_0$, będzie miał to samo nachylenie, ale inną stałą oraz składnik losowy, który ma wartość oczekiwaną równą zero.
10. Pokaż, że jeżeli $\mathbb{E}(u|x) = 0$, to zachodzi $\mathbb{E}(y|x) = \beta_0 + \beta_1 x$.