
ZADANIA

Zadanie 1. Wykorzystując dane GPA2, oszacowano następujące równanie regresji:

$$\begin{aligned} \widehat{sat} = & 1,028.10 + 19.30 \, hsize - 2.19 \, hsize^2 - 45.09 \, female \\ & (6.29) \quad (3.83) \quad (.53) \quad (4.29) \\ & - 169.81 \, black + 62.31 \, female \cdot black \\ & (12.71) \quad (18.15) \\ & n = 4,137, R^2 = .0858. \end{aligned}$$

Zmienna *sat* to łączny wynik testu SAT, *hsize* to wielkość ukończonej przez ucznia szkoły średniej, w setkach, *female* jest zmienną jakościową opisującą płeć, a *black* opisuje rasę i jest równa jeden dla afroamerykanów i zero w przeciwnym razie.

- (i) Czy istnieją mocne dowody, że zmienna $hsize^2$ powinien zostać uwzględniony w modelu? Od to równanie, jaki jest optymalny rozmiar szkoły średniej?
- (ii) Przy ustalonym poziomie zmiennej *hsize*, jaka jest szacunkowa różnica między wynikami SAT białych kobiet i białych mężczyzn? Jak istotna statystycznie jest ta szacowana różnica?
- (iii) Jaka jest szacunkowa różnica w wynikach SAT między białymi mężczyznami i mężczyznami przedstawicielami afroamerykanów? Przetestuj hipotezę zerową, że nie ma różnicy między ich wynikami, a alternatywą jest różnica.
- (iv) Jaka jest szacunkowa różnica w wyniku SAT między afroamerykankami i białymi kobietami? Co musisz zrobić, aby sprawdzić, czy różnica jest statystycznie istotny?

Zadanie 2. W przykładzie 2 z wykładu niech *nPC* będzie zmienną jakościową równą jednej, jeśli nie posiada komputer, a zero w przeciwnym przypadku.

- (i) Jeśli w równaniu regresji *nPC* zostanie użyte w miejsce zmiennej *PC*, co stanie się z wyrazem wolnym w estymowanym równaniu? Jaki będzie współczynnik przy zmiennej *nPC*?
- (ii) Co stanie się z R^2 , jeśli zamiast *PC* zostanie użyta zmienna *nPC*?
- (iii) Czy zmienne *PC* i *nPC* mogą być uwzględnione jako zmienne niezależne w model? Uzasadnij swoją odpowiedź.

Zadanie 3. Równanie wyjaśniające wynagrodzenie dyrektora generalnego to

$$\widehat{\log(\text{salary})} = 4.59 + .257 \log(\text{sales}) + .011 \text{roe} + .158 \text{finance} \\
\begin{matrix} (.30) & (.032) & & (.004) & & (.089) \\ & & + .181 \text{consprod} - .283 \text{utility} \\ & & (.085) & & (.099) \end{matrix} \\
n = 209, R^2 = .357.$$

gdzie finance, consprod i utility są zmiennymi binarnymi wskazującymi odpowiednio sektor finansowy, sektor produktów konsumenckich i sektor usług. Pominęto usługi transportowych.

- (i) Oblicz przybliżoną różnicę procentową szacunkowego wynagrodzenia pomiędzy usługami i transportem, przy ustalonym poziomie zmiennych *sales* i *roe*. Czy różnica jest statystycznie istotna na poziomie 1%?
- (ii) Użyj równania z wykładu, $100[\exp(\hat{\beta}_1) - 1]$, aby uzyskać dokładną różnicę procentową w oszacowaniu wynagrodzenia między sektorami usług i transportem i porównaj to z odpowiedzią uzyskaną w części (i).
- (iii) Jaka jest przybliżona różnica procentowa szacunkowego wynagrodzenia między sektorem produktów konsumenckich i branżą finansową? Napisz równanie, które pozwoli ci sprawdzić, czy różnica jest istotna statystycznie.

Zadanie 4. Załóżmy, że zbierasz dane ankietowe na temat płac, wykształcenia, doświadczenia i płci. Ponadto poprosisz o informacje na temat używania marihuany. Pierwotne pytanie brzmi: „Ile razy w ciągu ostatniego miesiąca paliłeś marihuanę?”

- (i) Napisz równanie, które pozwoli ci oszacować wpływ używania marihuany na płace, przy jednoczesnym kontrolowaniu innych czynników. Powinieneś być w stanie wypowiedzieć takie zdanie, jak: „Szacuje się, że palenie marihuany pięć i więcej razy w miesiącu zmienia wynagrodzenie o $x\%$.”
- (ii) Podaj postać modelu, który pozwoli ci sprawdzić, czy używanie marihuany ma inny wpływ na płace mężczyzn i kobiet. Jak byś to przetestował hipotezę mówiącą, że nie ma różnic w skutkach używania marihuany w grupie mężczyzn i grupie kobiet?
- (iii) Załóżmy, że uważasz, że lepiej jest mierzyć użycie marihuany poprzez umieszczanie ludzi w jedną z czterech kategorii: nieużytkownik, lekki użytkownik (1 do 5 razy w miesiącu), umiarkowany użytkownik (6 do 10 razy w miesiącu) i intensywny użytkownik (ponad 10 razy w miesiącu). Teraz napisz model, który pozwala oszacować wpływ używania marihuany na płace.
- (iv) Korzystając z modelu w części (iii), wyjaśnij szczegółowo, jak przetestować hipotezę zerową mówiącą, że użycie marihuany nie ma wpływu na płace. Bądź konkretny i sporządź dokładną listę stopni swobody.
- (v) Jakie są potencjalne problemy z wyciąganiem wniosków przyczynowych za pomocą tak przeprowadzonej ankiety?

PROBLEMY

Problem 1. Przeanalizuj dane zawarte w pliku GPA1.

- (i) Dodaj zmienne *mothcoll* i *fathcoll* do równania (6) z wykładu i przeprowadź analizę dla tak skonstruowanego modelu. Co stanie się estymatorem efektu zmiennej opisującej posiadanie komputera? Czy *PS* jest nadal statystycznie istotna?
- (ii) Przetestuj łączne znaczenie zmiennych *mothcoll* i *fathcoll* w równaniu z części (i).
- (iii) Dodaj *hsGPA*² do modelu z części (i) i zdecyduj, czy to uogólnienie jest potrzebne.

Problem 2. Przeanalizuj dane zawarte w pliku BEAUTY.

- (i) Znajdź osobne frakcje mężczyzn i kobiet, które są sklasyfikowane jako o ponadprzeciętnym wyglądzie. Czy więcej osób zostało zaklasyfikowanych jako o powyżej średnim czy poniżej średnim wyglądzie?
- (ii) Przetestuj hipotezę zerową, że frakcje kobiet i mężczyzn, których wygląd został oceniony jako powyżej średni są takie same.
- (iii) Oszacuj teraz model

$$\log(wage) = \beta_0 + \beta_1 belavg + \beta_2 abvavg + u$$

oddzielnie dla mężczyzn i kobiet, i podaj wyniki w przejrzystej formie. W obu przypadkach interpretuj współczynnik dla *belavg*. Wyjaśnij co oznacza hipoteza postaci $H_0 : \beta_1 = 0$ przeciwko $H_1 : \beta_1 < 0$, i wyznacz *p*-wartość dla mężczyzn i kobiet.

- (iv) Czy istnieją przekonujące dowody na to, że kobiety o ponadprzeciętnym wyglądzie zarabiają więcej niż kobiety o przeciętnym wyglądzie? Wyjaśnij na czym opierasz swoją odpowiedź.
- (v) W przypadku mężczyzn i kobiet dodaj zmienne wyjaśniające *educ*, *exper* *exper*², *union*, *goodhlth*, *black*, *married*, *south*, *bigcity*, *smllcity* i *service*. Czy wpływ zmiennej "wygląd" zmienia się w istotny sposób?.