

Ćwiczenie nr 4 – regresja wieloczynnikowa, regresja wielomianowa

1. Wczytaj dane z pliku `zarthan.txt`.

Dane przedstawiają wielkość sprzedaży (*sales*) firmy The Zathran Company w 15-tu regionach oraz populację docelową (*population*) oraz średni dochód (*income*) w danym regionie.

Narysuj wykresy zależności pomiędzy poszczególnymi zmiennymi.

Przeprowadź regresję zmiennej zależnej *sales* od zmiennych niezależnych *population* i *income*.

```
reg<-lm(sales~.,zarthan)
```

Narysuj wykresy diagnostyczne i oceń poprawność modelu.

Na wykresach regresji częściowej zbadaj czy zmienna zależna nie wykazuje krzywoliniowej zależności.

```
library(car)  
avPlots(reg)
```

Zbadaj wpływ składnika interakcji pomiędzy zmiennymi zależnymi na model.

Narysuj wykres zależności reszt od składnika interakcji.

```
plot(reg$res~I(population*income))
```

Przeprowadź regresję z dodanym składnikiem interakcji:

```
reg1<-lm(sales~.+I(population*income))  
anova(reg1)
```

Czy składnik interakcji jest potrzebny w modelu?

Przetestuj czy sprzedaż jest zależna od populacji docelowej i dochodu w regionie (test hipotezy $H_0 : \beta_1 = \beta_2 = 0$).

```
reg0<-lm(sales~1,zarthan)  
anova(reg0,reg)
```

Oblicz jednoczesne 90-procentowe przedziały ufności Bonferroniego dla parametrów β_1 i β_2 (współczynników przy *population* i *income*).

```
confint(reg,c("population","income"),level=0.95)
```

Narysuj eliptyczny 90-procentowy przedział ufności dla tych parametrów.

```
plot(ellipse(reg,c("population","income"),level=0.9),type="l")
```

Nanieś na niego przedział Bonferroniego

```
rect(0.482813482,0.007089742,0.50919647,0.01130842)
```

Wyestymuj średnią oczekiwaną odpowiedź dla regionu o populacji docelowej 220 i dochodzie 2500 wraz z 95-procentowym przedziałem ufności:

```
predict(reg,data.frame(population=220,income=2500),interval="confidence",level=0.95)
```

Wyznacz jednoczesny 90-procentowy przedział ufności dla nowych regionów: *population* = 220, *income* = 2500 i *population* = 375, *income* = 3500:

```
predict(reg,data.frame(population=c(220,375),income=c(2500,3500)),  
interval="prediction",level=0.95)
```

2. Wczytaj dane z pliku cafeteria.txt.

Dane zawierają ilość sprzedanej kawy (*sales*) oraz ilość automatów do kawy (*dispensers*) w 14-tu kafeteriach.

Narysuj wykres zależności sprzedaży kawy od ilości automatów.

Zbadaj korelację między zmienną zależną *dispensers* a jej kwadratem.

Wycentruj zmienną *dispensers*.

```
d<-dispensers-mean(dispensers)
```

Jak teraz wygląda korelacja między tą zmienną a jej kwadratem?

Przeprowadź regresję kwadratową zmiennej zależnej *sales* od zmiennej scentrowanej *d*.

```
reg<-lm(sales~d+I(d^2))
```

Zbadaj poprawność modelu oraz istotność czynnika kwadratowego.

Czy czynnik sześcienny wniósł by coś nowego do modelu?

Jakie będą współczynniki regresji dla oryginalnych zmiennych?