

Ćwiczenie nr 7 – regresja ważona, estymowanie brakujących obserwacji

1. Wczytaj dane z pliku *blood.txt*. Dane te zawierają ciśnienie rozkurczowe krwi (*pressure*) 54 zdrowych kobiet, oraz wiek tych kobiet (*age*).
 - Narysuj wykres zależności pomiędzy zmiennymi.
 - Przeprowadź regresję zmiennej zależnej *pressure* od zmiennej *age*. Narysuj wykres reszt od zmiennej niezależnej.
 - Podziel reszty na grupy w zależności od wieku. Policz estymator wariancji w każdej z grup.

```
grupa<-rep(1,54)
grupa[30<=age&age<40]<-2
grupa[40<=age&age<50]<-3
grupa[50<=age]<-4
v1<-var(reg$res[grupa==1])
itd
wariancja<-c(v1,v2,v3,v4)
```
 - Sprawdź czy wariancja jest proporcjonalna do zmiennej niezależnej, jej kwadratu lub pierwiastka.

```
srodek<-c(25,35,45,55)
zal<-data.frame(srodek,wariancja)
zal<-cbind(zal,v.x=wariancja/srodek,
v.x2=wariancja/srodek^2,v.sqrt.x=wariancja/sqrt(srodek))
```
 - Przeprowadź regresję ważoną z wagami będącymi odwrotnościami wariancji w grupach.

```
waga=rep(1:54)
waga[grupa==1]<-1/v1
itd
reg1<-lm(pressure~age,weight=waga)
```
 - Porównaj wyniki.
2. Wczytaj dane *chmiss* z biblioteki *faraway*. Dane te dotyczą ubezpieczeń domów w różnych regionach Chicago. Zawierają one następujące zmienne: *race* - procent mniejszości w populacji regionu; *fire* - ilość pożarów na 100 gosp. domowych regionie; *theft* - ilość kradzieży na 1000 osób zamieszkujących region; *age* - procent domów wybudowanych przed 1939 w regionie; *involact* - ilość zawartych oraz odnowionych polis typu FAIR (socjalnych, Fair Access to Insurance Requirements) na 100 gosp. domowych; *income* - mediana dochodu rodziny w danym regionie;
Część obserwacji została losowo usunięta.

- Przeglądnij dane i przeprowadź regresję ze zmienną zależną *involact* na pozostałe zmienne.
- Uzupełnij brakujące dane w zmiennych niezależnych za pomocą średniej wartości tych zmiennych. Przeprowadź ponownie regresję.

```
srednie<-apply(chmiss,2,mean,na.rm=T)
chm<-chmiss
for(i in c(1:4,6)) chm[is.na(chm[,i]),i]<-srednie[i]
```
- Uzupełnij brakujące zmienne niezależne poprzez regresję na pozostałe zmienne niezależne.

```
gr <- lm(race~fire+theft+age+income,chmiss)
predict(gr,chmiss[is.na(chmiss$race),])
```

Czy wszystko jest w porządku?
Spróbuj transformacji logitowej zmiennej *race*.

```
gr<-lm(logit(race/100)~fire+theft+age+income,chmiss)
ilogit(predict(gr,chmiss[is.na(chmiss$race),]))*100
```
- Porównaj z prawdziwymi wartościami tej zmiennej.

```
data(chicago)
chicago$race[is.na(chmiss$race)]
```