

Ćwiczenie nr 6 – wybór zmiennych objaśniających

1. Wczytaj dane *state*.

```
data(state)
statedata<-data.frame(state.x77,row.names=state.abb,check.names=T)
```

Dane zebrane w 50-ciu stanach zawierają nast. zmienne:

Population - przewidywana wielkość populacji na 1 lipca 1975; *Income* - dochód na głowę w 1974; *Illiteracy* - procent analfabetów w populacji w 1970; *Life.Exp* - oczekiwana długość życia w 1969-71; *Murder* - procent morderstw i nieumyślnych zabójstw na 100000 osób w 1976; *HS.Grad* - procent absolwentów szkół wyższych w 1970; *Frost* - średnia ilość dni z temperaturą poniżej 32 stopni w latach 1931-1960 w stolicach lub dużych miastach; *Area* - powierzchnia stanu w milach kwadratowych.

Przeprowadź regresję ze zmienną zależną *Life.Exp* i z pozostałymi opisanymi powyżej zmiennymi jako zmiennymi niezależnymi.

```
g<-lm(Life.Exp~.,statedata)
```

Które zmienne powinny pozostać w modelu? Przeprowadź eliminację zmiennych metodą krokową wsteczną. Zinterpretuj otrzymane wyniki.

```
summary(g)
g<-update(g,.~.-Area)
```

2. Przeprowadź regresję krokową wsteczną dla modelu pełnego z poprzedniego zadania stosując jako kryterium eliminacji Kryterium Informacyjne Akaike (AIC). Czy otrzymaliśmy ten sam model co w zadaniu pierwszym?

```
step(g,direction="backward")
```

3. Dobierz zmienne do modelu z zadania pierwszego przy wykorzystaniu kryterium C_p Mallowa.

```
library(leaps)
x<-model.frame(g)[,-1]
y<-model.frame(g)[,1]
leaps(x,y,method="Cp",names=names(statedata)[-4])>cp
Cpplot(cp)
```

4. Powtórz zadanie trzecie dla kryterium R_a^2 .
5. Na dobór zmiennych w modelu duży wpływ mogą mieć obserwacje odstające. Usuń z modelu stan o największej dźwigni. Czy teraz kryterium R_a^2 wskaże ten sam model?

```
sort(hatvalues(g))
x1<-x[state.abb!="AK",]
y1<-y[state.abb!="AK"]
```

6. Narysuj wykresy pudełkowe dla wszystkich zmiennych. Przeprowadź transformację zmiennych o dużej skośności, a nast. ponownie narysuj wykresy pudełkowe oraz dobierz zmienne do modelu przy wykorzystaniu kryterium R_a^2 .

```
par(mfrow=c(3,3))
for(i in 1:8)boxplot(statedata[,i],main=names(statedata)[i])
nx<-cbind(log(x[,1]),x[,2],log(x[,3]),x[,4:6],log(x[,7]))
par(mfrow=c(3,3))
apply(nx,2,boxplot)
```