

Ćwiczenie nr 2 – regresja jednoczynnikowa

1. Wczytaj dane z pliku `westwood.txt`:

```
westwood<-read.table("westwood.txt");attach(westwood).
```

Dane przedstawiają wielkość produkcji *LotSize* wraz z ilością roboczogodzin potrzebną do jej osiągnięcia *ManHours* (z 10 cykli produkcyjnych). Sporządź wykres danych i przejrzyj podstawowe statystyki:

```
summary(westwood); plot(ManHours~LotSize).
```

Sporządź wykres pudełkowy dla zmiennej niezależnej oraz wykres jej zależności od czasu. Zinterpretuj wyniki.

```
boxplot(LotSize); plot(LotSize,type="o")
```

Przeprowadź regresję ze zmienną zależną *ManHours* i niezależną *LotSize*. Narysuj wykres danych wraz z linią regresji. Omów uzyskane wyniki.

```
reg<-lm(ManHours~LotSize); summary(reg); abline(reg)
```

Narysuj wykresy zależności zmiennej niezależnej od reszt, reszt od czasu, wykres pudełkowy dla reszt, normalny dla reszt, reszt od wartości teoretycznych:

```
plot(LotSize,reg$res); plot(ManHours,reg$res); abline(h=0);  
plot(reg$res,type="o"); abline(h=0); boxplot(reg$res);  
qqnorm(reg$res); qqline(reg$res); plot(reg$res~reg$fit)
```

Czy któreś z obserwacji mogą być obserwacjami odstającymi? Zidentyfikuj je.

```
identify(LotSize,reg$res)
```

2. Wczytaj dane *janka* z biblioteki *SemiPar*:

```
library(SemiPar); data(janka); janka.
```

JankaHardness jest australijską miarą twardości drewna. Test polega na pomiarze siły potrzebnej do osadzenia stalowej kuli w kawałku drewna, co daje informację o wytrzymałości drewna.

Powtórz czynności z poprzedniego przykładu dla zmiennej zależnej *hardness* i zmiennej niezależnej *dens*. Czy wszystkie założenia modelu regresji liniowej są spełnione?

Przeprowadź ponownie regresję dla zmiennej zależnej $\log(\text{hardness})$.

```
reg1<-update(reg,I(log(hardness))~.)
```

Co można teraz powiedzieć o resztach w modelu? Zaproponuj odpowiednią transformację zmienną niezależną i przeprowadź ponownie regresję. Przetestuj normalność reszt. Jakie testy normalności można użyć w tym wypadku?