**2.2. Multipla regresija: diagnostika**

Za regresijski model brez lestvice natančnosti ( lmr <- lm(DU~TN+PRI+VZT,data=X) ) izdelajmo diagnostične slike:

plot(lmr)

* graf ostankov (rezidualni graf),
* normalni kvantilni graf ostankov,
* graf odnosa med napovedano vrednostjo in variabilnostjo,
* graf vplivnosti oseb.

Izračunajmo Cookove razdalje:

cd <- cooks.distance(lmr)

Narišimo njihovo porazdelitev:

Boxplot(cd)

Deveta oseba ima očitno največjo vplivnost (gl. sliko), kar vidimo tudi iz

which.max(cd)

Kolikšna pa je njena Cookova razdalja?

cd[9]

Podobno lahko naredimo s studentiziranimi ostanki:

sr<-rstudent(lmr)

Boxplot(sr)

describe(sr)

Največja abs. vrednost je 2,35.

**2.3. Regresija s kategorialnimi napovedniki**

Za skupino otrok in njihovih mater imamo naslednje podatke:

* govor: govorni kod (višje vrednosti -> prevladujoč razširjeni kod; nižje vrednosti -> prevladujoč omejeni kod)
* otrokova starost v mesecih
* slog: vzgojni slog staršev (1= avtoritativni, 2 = permisivni, 3 = avtoritarni)
* pismen: materin dosežek na testu funkcionalne pismenosti.

Zanima nas, v kolikšni meri lahko govorni kod pojasnimo z vzgojnim slogom in s pismenostjo.

Preberimo podatke:

X<-read.table("mr2.txt",header = T)

Definirajmo vzgojni slog kot faktor:

X$slog <- as.factor(X$slog)

Če želimo, lahko označimo vrednosti:

levels(X$slog) <- c("avtoritativen","permisiven","avtoritaren")

Naredimo regresijo samo z vzgojnim slogom (enakovredno enosmerni ANOVA!). R samodejno izdela indikatorske spremenljivke (prva kategorija je referenčna).

lmr1<-lm(govor~slog, data=X)

summary(lmr1)

Dodajmo še pismenost:

lmr2<-lm(govor ~ slog + pismen, data=X)

> summary(lmr2)

Ali smo z vključitvijo pismenosti izboljšali napoved?

anova(lmr1,lmr2)

**2.4. Nelinearni odnosi**

Poglejmo odnos med govornim kodom in otrokovo starostjo:

sp(govor ~ starost, data=X)

Zdi se, da bi odnos lahko opisali s kvadratno krivuljo:

lmr <- lm(govor ~ starost + I(starost\*\*2), data=X)

Pozor: nujno moramo uporabiti funkcijo I(…) !

summary(lmr)

plot(lmr) etc.

Primerjajmo kvadratni model z linearnim:

lmr1 <- lm(govor ~ starost, data=X)

anova(lmr1,lmr)

Nelinearnost bi lahko zaznali tudi iz grafa ostankov za linearni model:

plot(lmr1) #(gl. prvo sliko)