Lista 4

Aneta Przydróżna

Wpływ korelacji.

Generujemy X_{100x2} z niezależnymi wierszami z rozkładu $N(0, \sigma/100)$, gdzie σ to macierz 2x2 z jedynkami na przekątnej głównej i 0.9 na pozostałych.Na podstawie X liczymy $Y = \beta_1 * X_1 + \epsilon$, $\beta_1 = 3$, X_1 i ϵ $N(0, I)$.

Konstruujemy 95% przedział ufności dla β_1 i wykonujemy t-test poziomu istotności 0,05 dla H_0 : $\beta_1 = 0$ z modelem prostej regresji liniowej i dla dwóch zmiennych wyjaśniających.

```r
A<-lm(y~X[,1])
B<-lm(y~X[,1]+X[,2])
confint(A,level=0.95)
```

```
##                  2.5 %    97.5 %
## (Intercept) -0.1065448 0.2953413
## X[, 1]       1.4164432 5.0686455
```

```r
confint(B,level=0.95)
```

```
##                  2.5 %   97.5 %
## (Intercept) -0.1066309 0.296765
## X[, 1]      -0.2546027 9.158697
## X[, 2]      -6.2281123 3.510915
```

```r
summary(A)$coefficients
```

```
##               Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 0.09439824  0.1012579 0.9322554 0.3534956617
## X[, 1]      3.24254434  0.9201971 3.5237498 0.0006482675
```

```r
summary(B)$coefficients
```

```
##                Estimate Std. Error    t value   Pr(>|t|)
## (Intercept)  0.09506704  0.1016252  0.9354669 0.35187080
## X[, 1]       4.45204737  2.3714390  1.8773611 0.06347346
## X[, 2]      -1.35859878  2.4534976 -0.5537396 0.58103129
```

Generujemy 1000 iid kopi błędu.Estymujemy β_1 .Wykonujemy test na znaczenie β_1 w obu modelach (z jedną i dwiema zmiennymi objaśniającymi).

H_0 : $\beta_1 = 0$ ale P_wartość wynosi 0.23, więc przyjmujemy alternatywę i uznajemy znaczący wpływ współczynnika regresji. Moc dla podstawowj regresji wyszła : 0.025, natomiast dla rozbudowanego : 0.027. Wyestymowane dchylenie w pierwszym przypadku to : 1.07 , w drugim 2.45. Wyniki są porównywalne do wyników otrzymanych w liczeniu teoretycznym.

Wpływ wymiaru.

Generujemy macierz zmiennych losowych z rozkładu normalnego z odchyleniem standardowym 0.1 $X_{1000x950}$, następnie dopasowujemy model, gdzie $\beta = (3, 3, 3, 3, 3, 0, ..., 0)$

szacujemy wartość współczynników regresji i używamy t-testuna poziomie istotności 0.05, przy użyciu kilku kolejnych coraz dłuższych ciągów.

```
n<-1000
p <- 950
X<-matrix(rnorm(n*p,0,0.1),n,p)
b <- rep(0,times = p)
b[1:5] <- rep(3,times = 5)

Y  <- X%*%b + rnorm(n)
RSS <- c()
MSE <-c()
mu <- X%*%b
k<- c(1,2,5,10,50,100,500,950)
AIC <- c()
p_values <- c()
FD <- c(rep(0,times  = length(k)))
pval <- matrix(0,length(k),2)
for(i in 1:length(k)){
  reg <- lm(Y~X[,1:k[i]])
  sum_reg <- summary(reg)

  RSS[i]<-sum(reg$residuals^2)
  MSE[i]<- sum((mu - reg$fit)^2)
  AIC[i]<- AIC(reg)
  if(i ==1) pval[i][1] <- sum_reg$coefficients[2,4]
  else pval[i,] <- sum_reg$coefficients[2:3,4]

  if(i>3){
    FD[i]<- sum(sum_reg$coefficients[7:(k[i]+1),4]<0.05)
  }
}
```

Wyniki obliczeń:

```
RSS
```

```
## [1] 1390.79830 1307.87087  996.45412  991.13613  955.92014  903.78932
## [7]  495.27249   49.81794
```

```
MSE
```

```
## [1] 371.267400 277.309122   1.444288   6.762286  41.978275  94.109087
## [7] 502.625918 948.080476
```

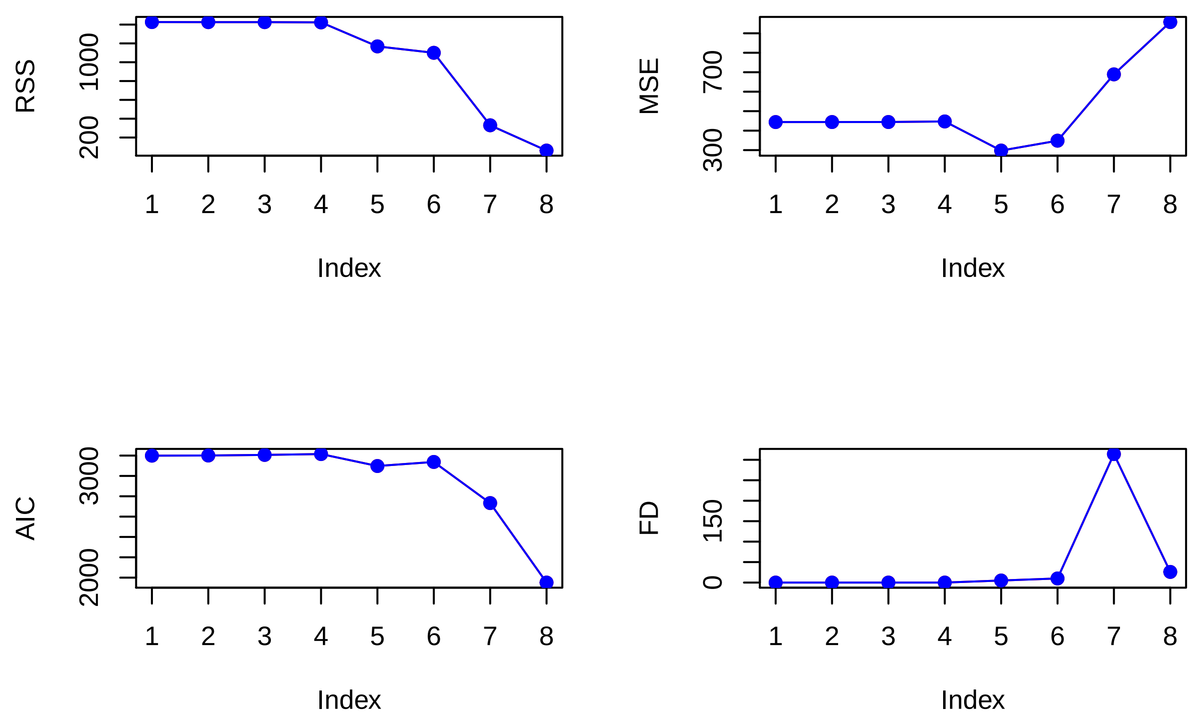
```
AIC
```

```
## [1] 3173.755 3114.278 2848.325 2852.974 2896.796 2940.718 3139.230 1742.497
```

```
FD
```

```
## [1]  0  0  0  0  0  4 20 17
```

Powtórzenie procedury dla modelu zbudowanego z maksymalnych oszacowanych współczynników regresji.



Kiedy z podchodzimy blisko n to kryterium AIC dąży do minus nieskonczoności. kryterium AIC jest w stanie zidentyfikować prawdziwy model gdy znamy sigme, ale w zadaniu estymujemy ją jakościowo słabo i kryterium zbiega do -nieskończoności.