Homework 9 (ex. 1-4)

Federico Veronesi

2023-12-20

Exercise 1b

The breakdown point of the sample variance is 0, since in the formula the sample mean $\overline{X}$ is involved. It's sufficient to have one point arbitrarily far in order to break down the mean (as explained in the slides), thus the quantities $\left(X_i - \overline{X}\right)^2$ can be arbitrarily large with no boundaries. For this reason, the breakdown point of the sample variance is equal to the one of the sample mean and it's equal to 0.

Exercise 4

Data Preparation

```
library(tidyverse)
library(robustbase)
setwd("C:/Users/Veronesi/Desktop/uniBo/Magistrale/Modern Statistics and Big Data
Analytics")
unicef <- read.csv("unicef97.dat", sep = " ")
colnames(unicef) <- c("y", "x1", "x2", "x3", "x4", "x5", "x6", "x7")
## The outlier is Sao TP
unicefnooutl <- unicef[rownames(unicef)!="SaoTP",]
```

Regression and Robust regression

```
lm2 <- lm(formula = y~x1+x2+x3+x4+x5+x6+x7, data = unicefnooutl)
lmrob2 <- lmrob(formula = y~x1+x2+x3+x4+x5+x6+x7, data = unicefnooutl)
lm1 <- lm(formula = y~x1+x2+x3+x4+x5+x6+x7, data = unicef)
lmrob1 <- lmrob(formula = y~x1+x2+x3+x4+x5+x6+x7, data = unicef)
summary(lm1)$coefficients

summary(lm2)$coefficients

summary(lmrob1)$coefficients

summary(lmrob2)$coefficients
```

Removing the outlier doesn't change the estimates for the coefficients in robust regression. The sign of the coefficients is the same and the t-test p-value is almost equal. Additionally, there are two coefficients (β_1 and β_6) for which the p-value is really near to 0.05 and, thus, removing one observation makes possible to modify our opinion on H0. This doesn't happen in robust regression because removing the outlier we obtain the same conclusions as before in terms of keeping or dropping variables. The vectors of estimatates are very similar: the coefficient for x2 switches sign in lmrob2, but it's a value very close to zero and x2 is a variable with a clear not-significant effect. We can think that the use of a very small weight for the observation "SaoTP" in lmrob1 makes this model very similar to lmrob2, a model in which "SaoTP" is missing.

We can't really make the same statements for "traditional" linear regression, since in lm1 (a model including SaoTP) the variables with significant effect are x1, x3, x4 and x6. In lm2 (a model which was fitted excluding the observation SaoTP) the variables with relevant effect are x1, x3, x4 and x7. After an eventual variable selection we would come up with different regressors in our model, depending on the type (robust or not robust) of regression. The vector $\hat{\beta}$ changes a lot from lm1 to lm2. The magnitude of coefficients is quite different.