```
#        TD de Stat Mass 05: regression linéaire / résidus
# Reprise du cours 05
library(MASS)  # MASS = Modern Applied Statistics with S
data(survey)   #lecture des données de  survey
survey.cc<-survey[complete.cases(survey),] #nettoyage de  survey

# 1.1 On recommence avec des noms plus simples: x, y, z dans un tableau MesDonnées
x<-survey.cc$Wr.Hnd;x  #x  prends la valeur de la première colonne de  Tailles.
y<-survey.cc$NW.Hnd  #y  prends la valeur de la deuxième colonne de  Tailles.cc
z<-survey.cc$Height  #z  prends la valeur de la troisième colonne de  Tailles.cc
MesDonnées=data.frame(x,y,z)
head(MesDonnées)

# 1.2: Scatter-plot
pairs(~x+y+z,data=MesDonnées)

# 1.3
# un nuage assez rectiligne
plot(x,y,pch=17)

# un nuage peu rectiligne
plot(y,z)

# 1.4 regression de z sur y
DteReg_z_sur_y=lm(z~y,data=MesDonnées)
coef(DteReg_z_sur_y) # ordonnée en 0 (b=Intercept) et pente (a) de la droite z=ay+b

# 1.5 Dessin
abline(coef(DteReg_z_sur_y),col="blue")

# 1.6 Le centre de gravité
points(mean(y),mean(z),col="red",pch=19)

# 1.7 les "prédites" z de y
z2y=fitted(DteReg_z_sur_y)

# 1.8 Dessin des (y,z2y)
points(y,z2y,col="green",pch=1)

# 1.9 étude des "résidus"
windows();dev.cur()   # on ouvre une fenêtre et affiche le nom
Res_z2y=z - z2y
hist(Res_z2y)
#hist(Res_z2x,add=T)
mean(Res_z2y)
sd(Res_z2y)

# 2 Simulation
# 2.1 Fabrication de l'échantillon simulé centré, sigma=8
set.seed(21420) # si salle 214 machine 20
Residu=rnorm(length(z),0,8)

# 2.2
mean(Residu) #  donne 8.455099e-16
sd(Residu)    # donne 7.737373

# 2.3 zSim
zSim=3*y+118+Residu

# 2.4
DteReg_zSim_sur_y=lm(zSim~y)
coef(DteReg_zSim_sur_y)

# 2.5
zSim2y=fitted(DteReg_zSim_sur_y)
Res_zSim2y=zSim-zSim2y
mean(Res_zSim2y)
sd(Res_zSim2y)

# 2.6
plot(y,zSim)
abline(coef(DteReg_zSim_sur_y),col="blue")
abline(c(118,3),col="red")
points(mean(y),mean(zSim),col="green")
```