Frisch Waugh Lovell Theorem

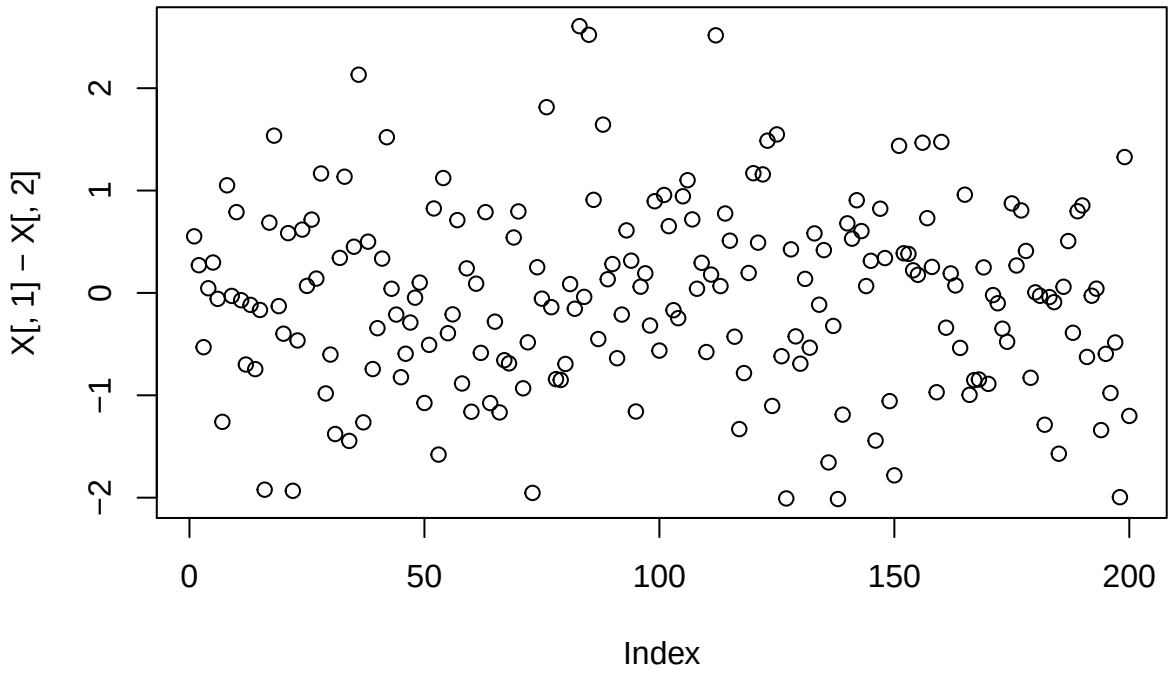
Nicolas Reigl

7 Oct 2015

Draw sampel variable for multivariate normal distribution. Sigma, correlation matrix at 0.6

```
rm(list=ls())

library(mvtnorm)
X<-rmvnorm(200, mean = c(0,0), sigma= matrix(data = c(1,0.6,0.6,1), nrow=2, ncol = 2))
plot(X[,1]-X[,2])
```




```
cor(X[,1], X[,2]) # correlation of x1 and x2
```

```
## [1] 0.5489145
```

```
X<-cbind(X, rep(rep(1,200))) # bind a constant variable to x (constant term in the regression)
b<-c(-3, 1, 0.5) # line with a slope of 1
y<-X%*%b+rnorm(200, sd = 1)


summary(lm(y~X-1))
```

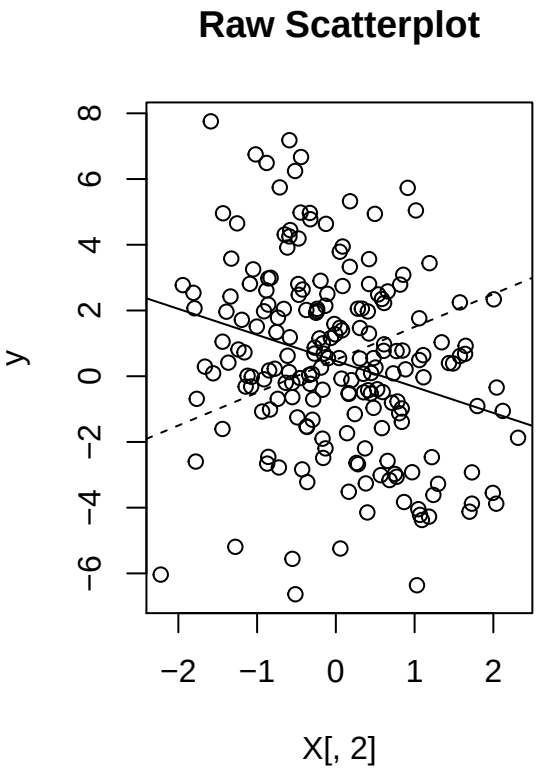
```
##
## Call:
## lm(formula = y ~ X - 1)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3279 -0.6945  0.0236  0.7088  2.3668
##
## Coefficients:
##    Estimate Std. Error t value Pr(>|t|)
## X1 -3.12215    0.09054 -34.485  < 2e-16 ***
## X2  0.96970    0.09305  10.421  < 2e-16 ***
## X3  0.37793    0.07224   5.232 4.27e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.021 on 197 degrees of freedom
## Multiple R-squared:  0.8713, Adjusted R-squared:  0.8693
## F-statistic: 444.5 on 3 and 197 DF,  p-value: < 2.2e-16
```

```
# First FWL plot
par(mfrow = c (1, 2))
plot (y~ X[,2], main = "Raw Scatterplot") # plot x2 against y
summary(lm(y~X[,2]))
```

```
##
## Call:
## lm(formula = y ~ X[, 2])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.2629 -1.7012  0.1377  1.6697  6.2480
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.4662     0.1910   2.441 0.015545 *
## X[, 2]       -0.7917     0.2058  -3.847 0.000161 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.701 on 198 degrees of freedom
## Multiple R-squared:  0.06954,    Adjusted R-squared:  0.06484
## F-statistic:  14.8 on 1 and 198 DF,  p-value: 0.0001613
```

```
abline(coefficients(lm(y~X[,2])))
abline(0.5, 1, lty= 2) # create a line with a slope of 1
```



So scatterplots are not very good explanation tools in cases of strongly correlated variables with effects that go into different directions x2 is sucking up the negative correlation with x1. scatterplot is missleading

```
# FWL corrected scatterplot
z<-X[,c(1,3)] # extract all the rows of x2

y.res<-y-predict(lm(y~z-1)) #  regression of x1 and a constant without x2
x.res<-X[,2]-predict(lm(X[,2]~z-1)) # residuals of the y.reg


summary(lm(y.res~x.res-1)) # run a y residuals vs. the x residuals
```

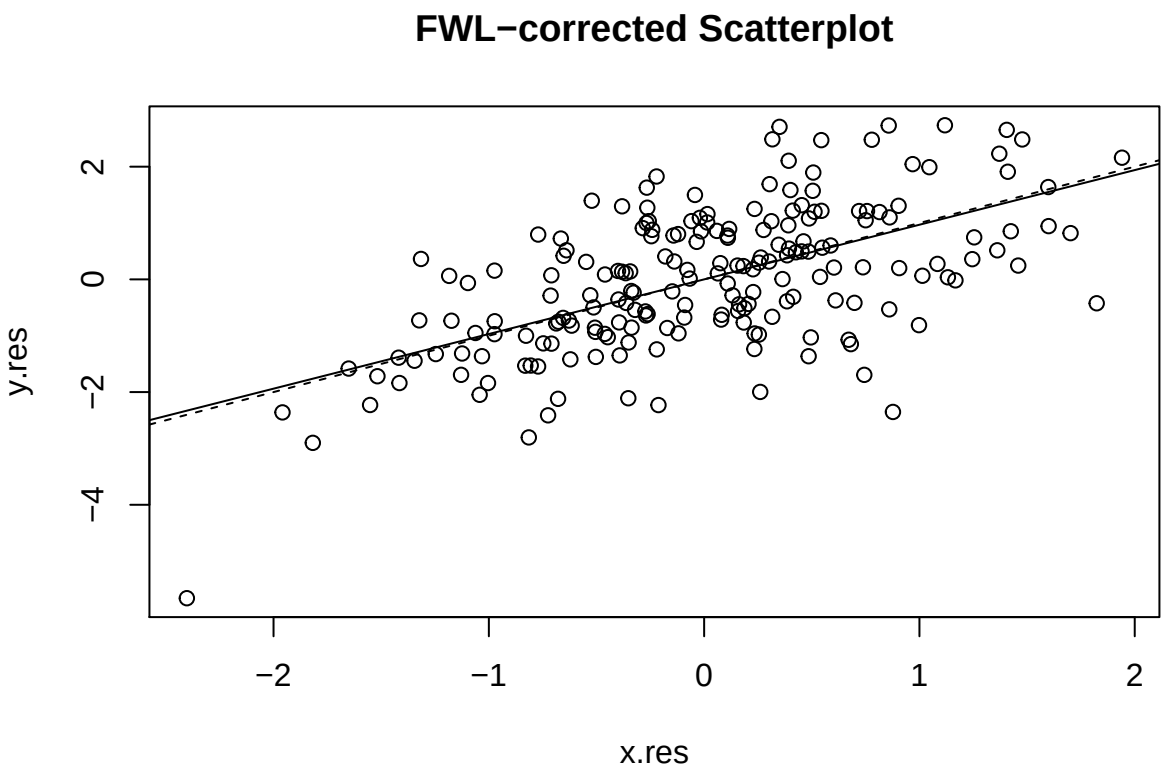
```
##
## Call:
## lm(formula = y.res ~ x.res - 1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3279 -0.6945  0.0236  0.7088  2.3668
##
## Coefficients:
##       Estimate Std. Error t value Pr(>|t|)
## x.res  0.96970    0.09258   10.47   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.016 on 199 degrees of freedom
## Multiple R-squared:  0.3554, Adjusted R-squared:  0.3521
## F-statistic: 109.7 on 1 and 199 DF,  p-value: < 2.2e-16
```

```r
# the beta is is the same as in the first regression

# fwl corrected scatterplot
plot(y.res~x.res, main= "FWL-corrected Scatterplot")
summary(lm(y.res~x.res-1)) #
```

```
##
## Call:
## lm(formula = y.res ~ x.res - 1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3279 -0.6945  0.0236  0.7088  2.3668
##
## Coefficients:
##       Estimate Std. Error t value Pr(>|t|)
## x.res  0.96970    0.09258   10.47   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.016 on 199 degrees of freedom
## Multiple R-squared:  0.3554, Adjusted R-squared:  0.3521
## F-statistic: 109.7 on 1 and 199 DF,  p-value: < 2.2e-16
```

```r
abline(coefficients(lm(y.res~x.res)))
abline(0, 1, lty= 2) # create a line with a slope of 1
```



```
## creates a correct scatter plot when spurious regression is prevalent
```

FWL is very close to the correct line