

PARTIE 2: Dummy variables.

Exemple: Voir cours

	X	Y
Groupe A	8	5.3
	0	0.9
	12	7.1
	2	2.4

	X	Y
Groupe B	9	5.1
	7	4.4
	8	5.2
	6	3.8

Définissons la "dummy variable" z qui vaut 0 si on est dans le groupe A et 1 sinon. La droite de régression qui représente ces données est donc sous la forme

$$Y = \beta_0 + \beta_1 X + \alpha_0 z + \alpha_1 Xz + \varepsilon,$$

avec

$Y = (5.3, 0.9, 7.1, 2.4, 5.1, 4.4, 5.2, 3.8)$, $X = (8, 0, 12, 2, 9, 7, 8, 6)$ et $z = (0, 0, 0, 0, 1, 1, 1, 1)$.

Ci-dessous le CODE R.

```
#Create covariables
Y <- c(5.3, .9, 7.1, 2.4, 5.1, 4.4, 5.2, 3.8); X <- c(8, 0, 12, 2, 9, 7, 8, 6)
z <- c(0, 0, 0, 0, 1, 1, 1, 1); t <- X*z

#linear regression model with all these covariables
out <- lm(Y ~ X + z + t)
summary(out)

##
## Call:
## lm(formula = Y ~ X + z + t)
##
## Residuals:
##      1      2      3      4      5      6      7      8
## 0.1099 -0.2418 -0.1143  0.2462 -0.2300  0.0100  0.3400 -0.1200
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.14176    0.21826   5.231  0.00638 **
## X              0.50604    0.02998  16.879 7.22e-05 ***
```

```
## z          -0.04176    0.99412  -0.042  0.96851
## t          -0.03604    0.13137  -0.274  0.79738
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.286 on 4 degrees of freedom
## Multiple R-squared:  0.9873, Adjusted R-squared:  0.9777
## F-statistic: 103.5 on 3 and 4 DF,  p-value: 0.0003022
```

On pourrait remarquer que les covariables z et Xz prises individuellement ne sont pas significatives pour le modèle.