Régression linéaire : au-delà de la corrélation et du test t

BY NC SA

Pr. Bruno Falissard
UNIVERSITÉ PARIS SUD

$$\text{durée} = a + b \times \text{age} + \text{bruit}$$

```
> mod1 <- lm(dur.interv~age,data=smp.1)
> summary(mod1)

Call:
lm(formula = dur.interv ~ age, data = smp.1)

Residuals:
    Min      1Q  Median      3Q     Max
-62.470 -14.402  -1.712  12.341  60.055

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 57.04091    2.22028  25.691   <2e-16 ***
age          0.12625    0.05375   2.349   0.0191 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.57 on 745 degrees of freedom
  (52 observations deleted due to missingness)
Multiple R-squared:  0.00735,   Adjusted R-squared:  0.006018
F-statistic: 5.516 on 1 and 745 DF,  p-value: 0.0191
```

$b \neq 0$?

CC BY NC SA

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 57.04091    2.22028  25.691   <2e-16 ***
age          0.12625    0.05375   2.349   0.0191 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.57 on 745 degrees of freedom
  (52 observations deleted due to missingness)
Multiple R-squared:  0.00735,   Adjusted R-squared:  0.006018
F-statistic: 5.516 on 1 and 745 DF,  p-value: 0.0191

> cor.test(smp.l$dur.interv,smp.l$age)

        Pearson's product-moment correlation

data:  smp.l$dur.interv and smp.l$age
t = 2.3487, df = 745, p-value = 0.0191
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.01408787 0.15650345
sample estimates:
       cor
0.08573358
```

CC BY NC SA

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 57.04091    2.22028  25.691   <2e-16 ***
age          0.12625    0.05375   2.349   0.0191 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.57 on 745 degrees of freedom
  (52 observations deleted due to missingness)
Multiple R-squared:  0.00735,   Adjusted R-squared:  0.006018
F-statistic: 5.516 on 1 and 745 DF,  p-value: 0.0191

> cor.test(smp.l$dur.interv,smp.l$age)

        Pearson's product-moment correlation

data:  smp.l$dur.interv and smp.l$age
t = 2.3487, df = 745, p-value = 0.0191
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.01408787 0.15650345
sample estimates:
       cor
0.08573358
```

$b \neq 0$?

CC BY NC SA

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 57.04091    2.22028  25.691   <2e-16 ***
age          0.12625    0.05375   2.349   0.0191 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 19.57 on 745 degrees of freedom
  (52 observations deleted due to missingness)
Multiple R-squared:  0.00735,   Adjusted R-squared:  0.006018
F-statistic: 5.516 on 1 and 745 DF,  p-value: 0.0191

> cor.test(smp.l$dur.interv,smp.l$age)

        Pearson's product-moment correlation

data:  smp.l$dur.interv and smp.l$age
t = 2.3487, df = 745, p-value = 0.0191
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.01408787 0.15650345
sample estimates:
      cor
0.08573358
```

CC BY NC SA

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 57.04091    2.22028  25.691   <2e-16 ***
age          0.12625    0.05375   2.349   0.0191 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.57 on 745 degrees of freedom
  (52 observations deleted due to missingness)
Multiple R-squared:  0.00735,   Adjusted R-squared:  0.006018
F-statistic: 5.516 on 1 and 745 DF,  p-value: 0.0191

> cor.test(smp.l$dur.interv,smp.l$age)

        Pearson's product-moment correlation

data:  smp.l$dur.interv and smp.l$age
t = 2.3487, df = 745, p-value = 0.0191
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.01408787 0.15650345
sample estimates:
      cor
0.08573358
```

$$r = b \times \frac{e.t.(\hat{a}ge)}{e.t.(durée\ entretien)}$$

CC BY NC SA

durée = a + b × age + bruit

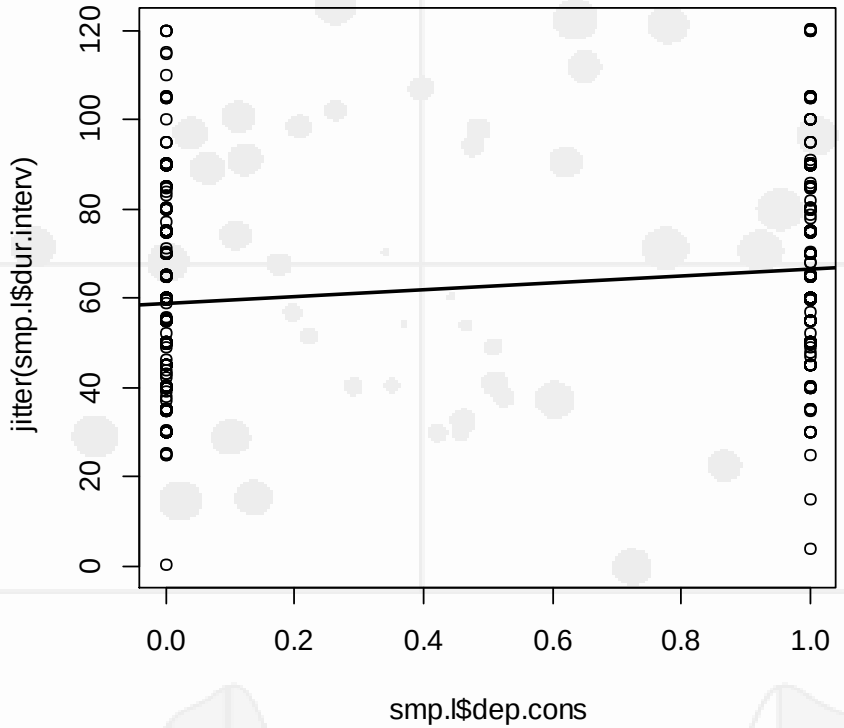
```
> mod1 <- lm(dur.interv~age,data=smp.1)
> summary(mod1)

Call:
lm(formula = dur.interv ~ age, data = smp.1)

Residuals:
    Min      1Q  Median      3Q     Max
-62.470 -14.402  -1.712  12.341  60.055

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 57.04091    2.22028  25.691   <2e-16 ***
age          0.12625    0.05375   2.349   0.0191 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.57 on 745 degrees of freedom
  (52 observations deleted due to missingness)
Multiple R-squared:  0.00735,   Adjusted R-squared:  0.006018
F-statistic: 5.516 on 1 and 745 DF,  p-value: 0.0191
```

CC BY NC SA

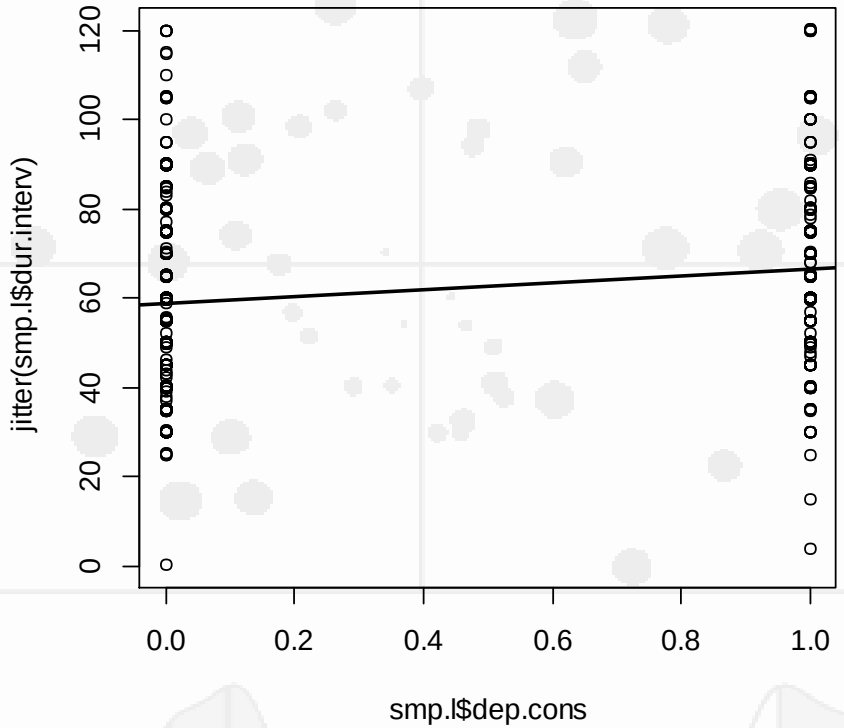
- Régression linéaire entre
  - Y = durée de l'interview
  - X = présence/absence d'une dépression

BY NC SA

Régression linéaire et test t

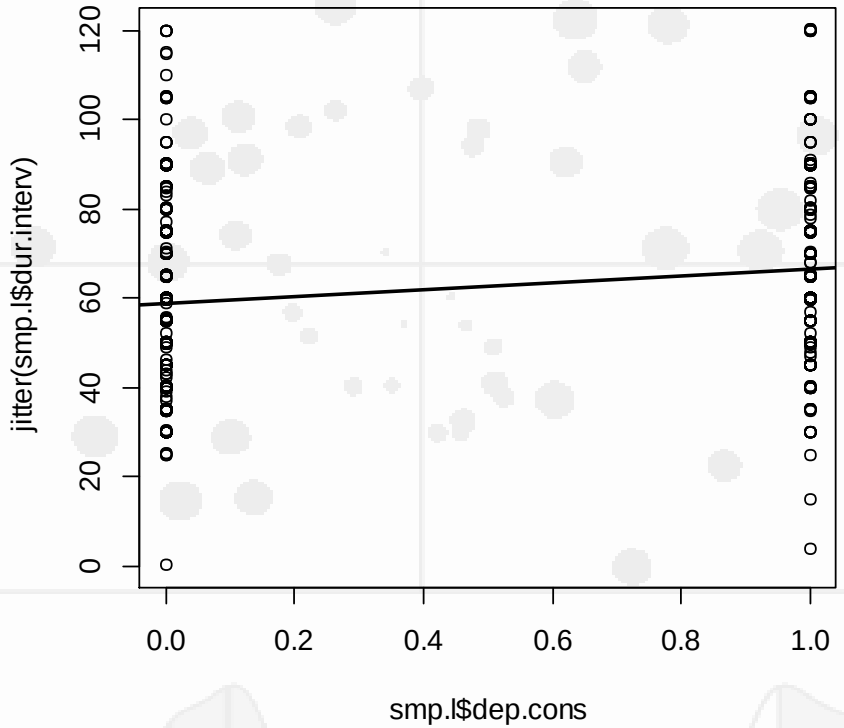


CC BY NC SA



$$\text{durée} = a + b \times \text{dep} \rightarrow b \neq 0 ?$$

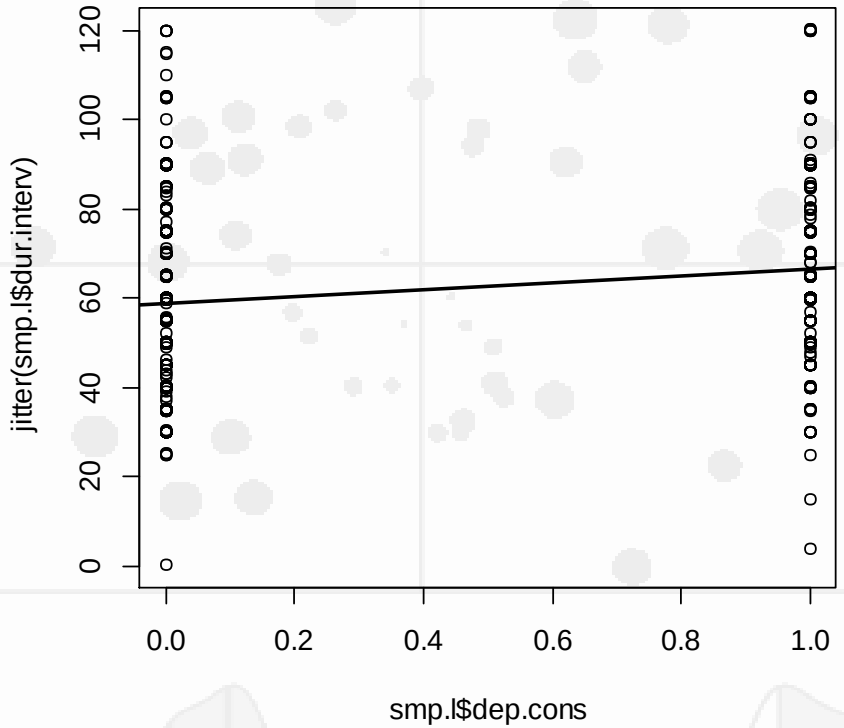
CC BY NC SA



durée = a + b × dep → b ≠ 0 ?

durée(déprimés) ≠ durée(non déprimés) ?

CC BY NC SA



durée = a + b × dep → b ≠ 0 ?

durée(déprimés) ≠ durée(non déprimés) ?

CC BY NC SA

```
> mod2 <- lm(dur.interv~dep.cons,data=smp.1)
> summary(mod2)

Call:
lm(formula = dur.interv ~ dep.cons, data = smp.1)

Residuals:
    Min      1Q  Median      3Q     Max
-62.538 -13.923   1.077  12.077  61.077

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  58.9234     0.9041  65.171  < 2e-16 ***
dep.cons      7.6143     1.4481   5.258  1.9e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.33 on 747 degrees of freedom
  (50 observations deleted due to missingness)
Multiple R-squared:  0.03569,   Adjusted R-squared:  0.0344
F-statistic: 27.65 on 1 and 747 DF,  p-value: 1.9e-07

> t.test(smp.1$dur.interv~smp.1$dep.cons,var.equal=TRUE)

        Two Sample t-test

data:  smp.1$dur.interv by smp.1$dep.cons
t = -5.2583, df = 747, p-value = 1.9e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -10.457001  -4.771515
sample estimates:
mean in group 0 mean in group 1
       58.92341        66.53767
```

CC BY NC SA

```
> mod2 <- lm(dur.interv~dep.cons,data=smp.1)
> summary(mod2)

Call:
lm(formula = dur.interv ~ dep.cons, data = smp.1)

Residuals:
    Min      1Q  Median      3Q     Max
-62.538 -13.923   1.077  12.077  61.077

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  58.9234     0.9041  65.171  < 2e-16 ***
dep.cons      7.6143     1.4481   5.258  1.9e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.33 on 747 degrees of freedom
  (50 observations deleted due to missingness)
Multiple R-squared:  0.03569,   Adjusted R-squared:  0.0344
F-statistic: 27.65 on 1 and 747 DF,  p-value: 1.9e-07

> t.test(smp.1$dur.interv~smp.1$dep.cons,var.equal=TRUE)

        Two Sample t-test

data:  smp.1$dur.interv by smp.1$dep.cons
t = -5.2583, df = 747, p-value = 1.9e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -10.457001  -4.771515
sample estimates:
mean in group 0 mean in group 1
       58.92341        66.53767
```

CC BY NC SA

```
> mod2 <- lm(dur.interv~dep.cons,data=smp.1)
> summary(mod2)

Call:
lm(formula = dur.interv ~ dep.cons, data = smp.1)

Residuals:
    Min      1Q  Median      3Q     Max
-62.538 -13.923   1.077  12.077  61.077

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  58.9234     0.9041  65.171  < 2e-16 ***
dep.cons      7.6143     1.4481   5.258  1.9e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.33 on 747 degrees of freedom
  (50 observations deleted due to missingness)
Multiple R-squared:  0.03569,   Adjusted R-squared:  0.0344
F-statistic: 27.65 on 1 and 747 DF,  p-value: 1.9e-07

> t.test(smp.1$dur.interv~smp.1$dep.cons,var.equal=TRUE)

        Two Sample t-test

data:  smp.1$dur.interv by smp.1$dep.cons
t = -5.2583, df = 747, p-value = 1.9e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -10.457001  -4.771515
sample estimates:
mean in group 0 mean in group 1
       58.92341        66.53767
```

CC BY NC SA

```
mod1 <- lm(dur.interv~age,data=smp.l)
summary(mod1)
cor.test(smp.l$dur.interv,smp.l$age)

plot(smp.l$dep.cons,jitter(smp.l$dur.interv))
abline(lm(smp.l$dur.interv~smp.l$dep.cons),lwd=2)
mod2 <- lm(dur.interv~dep.cons,data=smp.l)
summary(mod2)
t.test(smp.l$dur.interv~smp.l$dep.cons,var.equal=TRUE)
```

BY NC SA