Régression linéaire multiple
Analyse de variance

Pr. Bruno Falissard
UNIVERSITÉ PARIS SUD

- La durée d'entretien est liée
  - À l'âge
  - À l'existence d'une dépression

- La durée d'entretien est liée
  - À l'âge
  - À l'existence d'une dépression
  - À l'existence d'un abus de substance
  - À l'existence d'un trouble schizophrénique
- Comment faire la part des choses ?

- La durée d'entretien est liée
  - À l'âge
  - À l'existence d'une dépression
  - À l'existence d'un abus de substance
  - À l'existence d'un trouble schizophrénique
- Comment faire la part des choses ?

$$\text{Durée} = a + b \times \text{age} + c \times \text{dep} + d \times \text{subst} + e \times \text{scz} + \text{bruit}$$

- La durée d'entretien est liée
  - À l'âge
  - À l'existence d'une dépression
  - À l'existence d'un abus de substance
  - À l'existence d'un trouble schizophrénique
- Comment faire la part des choses ?

$$\text{Durée} = a + b \times \text{age} + c \times \text{dep} + d \times \text{subst} + e \times \text{scz} + \text{bruit}$$

$$b \neq 0 \ ?$$

- La durée d'entretien est liée
  - À l'âge
  - À l'existence d'une dépression
  - À l'existence d'un abus de substance
  - À l'existence d'un trouble schizophrénique
- Comment faire la part des choses ?

$$\text{Durée} = a + b \times \text{age} + c \times \text{dep} + d \times \text{subst} + e \times \text{scz} + \text{bruit}$$

$$b \neq 0 \ ?$$

```
> mod3 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons,data=smp.l)
> summary(mod3)

Call:
lm(formula = dur.interv ~ age + dep.cons + subst.cons + scz.cons,
    data = smp.l)

Residuals:
    Min      1Q  Median      3Q     Max
-63.654 -14.522  -1.193  11.482  62.482

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 48.90105    2.62213  18.649  < 2e-16 ***
age          0.22096    0.05708   3.871 0.000118 ***
dep.cons     7.38932    1.44783   5.104 4.24e-07 ***
subst.cons   5.25157    1.74318   3.013 0.002678 **
scz.cons     2.27256    2.52323   0.901 0.368062
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.1 on 742 degrees of freedom
  (52 observations deleted due to missingness)
Multiple R-squared:  0.05833,   Adjusted R-squared:  0.05325
F-statistic: 11.49 on 4 and 742 DF,  p-value: 4.692e-09
```

```
> mod3 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons,data=smp.l)
> summary(mod3)

Call:
lm(formula = dur.interv ~ age + dep.cons + subst.cons + scz.cons,
    data = smp.l)

Residuals:
    Min      1Q  Median      3Q     Max
-63.654 -14.522  -1.193  11.482  62.482

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 48.90105    2.62213  18.649  < 2e-16 ***
age          0.22096    0.05708   3.871 0.000118 ***
dep.cons     7.38932    1.44783   5.104 4.24e-07 ***
subst.cons   5.25157    1.74318   3.013 0.002678 **
scz.cons     2.27256    2.52323   0.901 0.368062
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.1 on 742 degrees of freedom
  (52 observations deleted due to missingness)
Multiple R-squared:  0.05833,   Adjusted R-squared:  0.05325
F-statistic: 11.49 on 4 and 742 DF,  p-value: 4.692e-09
```

Durée = a + b × age + c × dep + d × subst + e × scz + bruit

```
> mod3 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons,data=smp.l)
> summary(mod3)

Call:
lm(formula = dur.interv ~ age + dep.cons + subst.cons + scz.cons,
    data = smp.l)

Residuals:
    Min      1Q  Median      3Q     Max
-63.654 -14.522  -1.193  11.482  62.482

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 48.90105    2.62213  18.649  < 2e-16 ***
age          0.22096    0.05708   3.871 0.000118 ***
dep.cons     7.38932    1.44783   5.104 4.24e-07 ***
subst.cons   5.25157    1.74318   3.013 0.002678 **
scz.cons     2.27256    2.52323   0.901 0.368062
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.1 on 742 degrees of freedom
  (52 observations deleted due to missingness)
Multiple R-squared:  0.05833,   Adjusted R-squared:  0.05325
F-statistic: 11.49 on 4 and 742 DF,  p-value: 4.692e-09
```

Durée = a + b × age + c × dep + d × subst + e × scz + bruit

```
> mod3 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons,data=smp.l)
> summary(mod3)

Call:
lm(formula = dur.interv ~ age + dep.cons + subst.cons + scz.cons,
    data = smp.l)

Residuals:
    Min      1Q  Median      3Q     Max
-63.654 -14.522  -1.193  11.482  62.482

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 48.90105    2.62213  18.649  < 2e-16 ***
age          0.22096    0.05708   3.871 0.000118 ***
dep.cons     7.38932    1.44783   5.104 4.24e-07 ***
subst.cons   5.25157    1.74318   3.013 0.002678 **
scz.cons     2.27256    2.52323   0.901 0.368062
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.1 on 742 degrees of freedom
  (52 observations deleted due to missingness)
Multiple R-squared:  0.05833,   Adjusted R-squared:  0.05325
F-statistic: 11.49 on 4 and 742 DF,  p-value: 4.692e-09
```

Durée = a + b × age + c × dep + d × subst + e × scz + bruit

Variable à expliquer = a + b × variable explicative 1 + … + bruit

Durée = a + b × age + c × dep + d × subst + e × scz + bruit

Durée = a + b × age + c × dep + d × subst + e × scz + f × prof + bruit

Durée = a + b × age + c × dep + d × subst + e × scz + f × prof + bruit

« agriculteur »
« artisan »
« cadre »
« profession intermédiaire »
« ouvrier »
« employé »
« autre »
« sans profession »

Durée = a + b × age + c × dep + d × subst + e × scz + f × prof + bruit

1 : « agriculteur »
2 : « artisan »
3 : « cadre »
4 : « profession intermédiaire »
5 : « ouvrier »
6 : « employé »
7 : « autre »
8 : « sans profession »

Durée = a + b × age + c × dep + d × subst + e × scz + f × prof + bruit

	« agriculteur »
f1 : artisan=1 *vs* 0	« artisan »
f2 : cadre=1 *vs* 0	« cadre »
f3 : prof inter=1 *vs* 0	« profession intermédiaire »
f4 : ouvrier=1 *vs* 0	« ouvrier »
f5 : employé=1 *vs* 0	« employé »
f6 : autre=1 *vs* 0	« autre »
f7 : sans prof=1 *vs* 0	« sans profession »

```
> mod4 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons+prof,data=smp.1)
> summary(mod4)

Call:
lm(formula = dur.interv ~ age + dep.cons + subst.cons + scz.cons +
    prof, data = smp.1)

Residuals:
    Min      1Q  Median      3Q     Max
-63.280 -14.164  -1.337  10.959  63.184

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               62.79202   10.20779   6.151 1.26e-09 ***
age                        0.21289    0.05884   3.618 0.000317 ***
dep.cons                   7.36792    1.45840   5.052 5.53e-07 ***
subst.cons                 5.34589    1.76902   3.022 0.002599 **
scz.cons                   2.50439    2.54734   0.983 0.325863
profartisan              -11.48515    9.82936  -1.168 0.243005
profautre                -10.28748   10.33482  -0.995 0.319862
profcadre                -19.29636   10.38568  -1.858 0.063574 .
profemployé              -13.55809    9.76340  -1.389 0.165358
profouvrier              -14.01270    9.72111  -1.441 0.149880
profprof.intermédiaire   -13.01926    9.96911  -1.306 0.191977
profsans emploi          -14.27866    9.71782  -1.469 0.142174
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.11 on 731 degrees of freedom
  (56 observations deleted due to missingness)
Multiple R-squared:  0.06595,   Adjusted R-squared:  0.05189
F-statistic: 4.692 on 11 and 731 DF,  p-value: 5.825e-07
```

$$\text{Durée} = a + b \times \text{age} + c \times \text{dep} + d \times \text{subst} + e \times \text{scz} + f \times \text{prof} + \text{bruit}$$

```
> mod4 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons+prof,data=smp.l)
> summary(mod4)

Call:
lm(formula = dur.interv ~ age + dep.cons + subst.cons + scz.cons +
    prof, data = smp.l)

Residuals:
    Min      1Q  Median      3Q     Max
-63.280 -14.164  -1.337  10.959  63.184

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               62.79202   10.20779   6.151 1.26e-09 ***
age                        0.21289    0.05884   3.618 0.000317 ***
dep.cons                   7.36792    1.45840   5.052 5.53e-07 ***
subst.cons                 5.34589    1.76902   3.022 0.002599 **
scz.cons                   2.50439    2.54734   0.983 0.325863
profartisan              -11.48515    9.82936  -1.168 0.243005
profautre                -10.28748   10.33482  -0.995 0.319862
profcadre                -19.29636   10.38568  -1.858 0.063574 .
profemployé              -13.55809    9.76340  -1.389 0.165358
profouvrier              -14.01270    9.72111  -1.441 0.149880
profprof.intermédiaire   -13.01926    9.96911  -1.306 0.191977
profsans emploi          -14.27866    9.71782  -1.469 0.142174
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.11 on 731 degrees of freedom
  (56 observations deleted due to missingness)
Multiple R-squared:  0.06595,   Adjusted R-squared:  0.05189
F-statistic: 4.692 on 11 and 731 DF,  p-value: 5.825e-07
```

$$\text{Durée} = a + b \times \text{age} + c \times \text{dep} + d \times \text{subst} + e \times \text{scz} + f \times \text{prof} + \text{bruit}$$

```
> mod4 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons+prof,data=smp.l)
> summary(mod4)

Call:
lm(formula = dur.interv ~ age + dep.cons + subst.cons + scz.cons +
    prof, data = smp.l)

Residuals:
    Min      1Q  Median      3Q     Max
-63.280 -14.164  -1.337  10.959  63.184

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               62.79202   10.20779   6.151 1.26e-09 ***
age                        0.21289    0.05884   3.618 0.000317 ***
dep.cons                   7.36792    1.45840   5.052 5.53e-07 ***
subst.cons                 5.34589    1.76902   3.022 0.002599 **
scz.cons                   2.50439    2.54734   0.983 0.325863
profartisan              -11.48515    9.82936  -1.168 0.243005
profautre                -10.28748   10.33482  -0.995 0.319862
profcadre                -19.29636   10.38568  -1.858 0.063574 .
profemployé              -13.55809    9.76340  -1.389 0.165358
profouvrier              -14.01270    9.72111  -1.441 0.149880
profprof.intermédiaire   -13.01926    9.96911  -1.306 0.191977
profsans emploi          -14.27866    9.71782  -1.469 0.142174
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.11 on 731 degrees of freedom
  (56 observations deleted due to missingness)
Multiple R-squared:  0.06595,   Adjusted R-squared:  0.05189
F-statistic: 4.692 on 11 and 731 DF,  p-value: 5.825e-07
```

$$\text{Durée} = a + b \times \text{age} + c \times \text{dep} + d \times \text{subst} + e \times \text{scz} + f \times \text{prof} + \text{bruit}$$

Variables catégorielles à plus de deux classes

```
> smp.l$prof <- relevel(smp.l$prof,ref="ouvrier")
> mod5 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons+prof,data=smp.l)
> summary(mod5)

Call:
lm(formula = dur.interv ~ age + dep.cons + subst.cons + scz.cons +
    prof, data = smp.l)

Residuals:
    Min      1Q  Median      3Q     Max
-63.280 -14.164  -1.337  10.959  63.184

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)             48.77932    2.83938  17.180  < 2e-16 ***
age                      0.21289    0.05884   3.618 0.000317 ***
dep.cons                 7.36792    1.45840   5.052 5.53e-07 ***
subst.cons               5.34589    1.76902   3.022 0.002599 **
scz.cons                 2.50439    2.54734   0.983 0.325863
profagriculteur         14.01270    9.72111   1.441 0.149880
profartisan              2.52755    2.48989   1.015 0.310381
profautre                3.72522    3.99637   0.932 0.351567
profcadre               -5.28366    4.25567  -1.242 0.214798
profemployé              0.45460    2.12659   0.214 0.830785
profprof.intermédiaire   0.99344    2.95809   0.336 0.737089
profsans emploi         -0.26596    1.87727  -0.142 0.887375
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.11 on 731 degrees of freedom
  (56 observations deleted due to missingness)
Multiple R-squared:  0.06595,   Adjusted R-squared:  0.05189
F-statistic: 4.692 on 11 and 731 DF,  p-value: 5.825e-07
```

Variables catégorielles à plus de deux classes

```
> smp.l$prof <- relevel(smp.l$prof,ref="ouvrier")
> mod5 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons+prof,data=smp.l)
> summary(mod5)

Call:
lm(formula = dur.interv ~ age + dep.cons + subst.cons + scz.cons +
    prof, data = smp.l)

Residuals:
    Min      1Q  Median      3Q     Max
-63.280 -14.164  -1.337  10.959  63.184

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)             48.77932    2.83938  17.180  < 2e-16 ***
age                      0.21289    0.05884   3.618 0.000317 ***
dep.cons                 7.36792    1.45840   5.052 5.53e-07 ***
subst.cons               5.34589    1.76902   3.022 0.002599 **
scz.cons                 2.50439    2.54734   0.983 0.325863
profagriculteur         14.01270    9.72111   1.441 0.149880
profartisan              2.52755    2.48989   1.015 0.310381
profautre                3.72522    3.99637   0.932 0.351567
profcadre               -5.28366    4.25567  -1.242 0.214798
profemployé              0.45460    2.12659   0.214 0.830785
profprof.intermédiaire   0.99344    2.95809   0.336 0.737089
profsans emploi         -0.26596    1.87727  -0.142 0.887375
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.11 on 731 degrees of freedom
  (56 observations deleted due to missingness)
Multiple R-squared:  0.06595,   Adjusted R-squared:  0.05189
F-statistic: 4.692 on 11 and 731 DF,  p-value: 5.825e-07
```

```
> smp.l$prof <- relevel(smp.l$prof,ref="ouvrier")
> mod5 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons+prof,data=smp.l)
> summary(mod5)

Call:
lm(formula = dur.interv ~ age + dep.cons + subst.cons + scz.cons +
    prof, data = smp.l)

Residuals:
    Min      1Q  Median      3Q     Max
-63.280 -14.164  -1.337  10.959  63.184

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)           48.77932    2.83938  17.180  < 2e-16 ***
age                    0.21289    0.05884   3.618 0.000317 ***
dep.cons               7.36792    1.45840   5.052 5.53e-07 ***
subst.cons             5.34589    1.76902   3.022 0.002599 **
scz.cons               2.50439    2.54734   0.983 0.325863
profagriculteur       14.01270    9.72111   1.441 0.149880
profartisan            2.52755    2.48989   1.015 0.310381
profautre              3.72522    3.99637   0.932 0.351567
profcadre             -5.28366    4.25567  -1.242 0.214798
profemployé            0.45460    2.12659   0.214 0.830785
profprof.intermédiaire 0.99344    2.95809   0.336 0.737089
profsans emploi       -0.26596    1.87727  -0.142 0.887375
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.11 on 731 degrees of freedom
  (56 observations deleted due to missingness)
Multiple R-squared:  0.06595,   Adjusted R-squared:  0.05189
F-statistic: 4.692 on 11 and 731 DF,  p-value: 5.825e-07
```

```
> smp.l$prof <- relevel(smp.l$prof,ref="ouvrier")
> mod5 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons+prof,data=smp.l)
> summary(mod5)

Call:
lm(formula = dur.interv ~ age + dep.cons + subst.cons + scz.cons +
    prof, data = smp.l)

Residuals:
    Min      1Q  Median      3Q     Max
-63.280 -14.164  -1.337  10.959  63.184

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         48.77932    2.83938  17.180  < 2e-16 ***
age                  0.21289    0.05884   3.618 0.000317 ***
dep.cons             7.36792    1.45840   5.052 5.53e-07 ***
subst.cons           5.34589    1.76902   3.022 0.002599 **
scz.cons             2.50439    2.54734   0.983 0.325863
profagriculteur     14.01270    9.72111   1.441 0.149880
profartisan          2.52755    2.48989   1.015 0.310381
profautre            3.72522    3.99637   0.932 0.351567
profcadre           -5.28366    4.25567  -1.242 0.214798
profemployé          0.45460    2.12659   0.214 0.830785
profprof.intermédiaire  0.99344    2.95809   0.336 0.737089
profsans emploi     -0.26596    1.87727  -0.142 0.887375
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.11 on 731 degrees of freedom
  (56 observations deleted due to missingness)
Multiple R-squared:  0.06595,   Adjusted R-squared:  0.05189
F-statistic: 4.692 on 11 and 731 DF,  p-value: 5.825e-07
```

```
> drop1(mod5,.~.,test="F")
Single term deletions

Model:
dur.interv ~ age + dep.cons + subst.cons + scz.cons + prof
           Df Sum of Sq    RSS    AIC F value    Pr(>F)
<none>                  266846 4395.6
age         1    4778.4 271624 4406.8 13.0899 0.0003173 ***
dep.cons    1    9317.1 276163 4419.1 25.5233 5.527e-07 ***
subst.cons  1    3333.6 270180 4402.8  9.1322 0.0025992 **
scz.cons    1     352.8 267199 4394.6  0.9666 0.3258633
prof        7    2295.5 269142 4388.0  0.8983 0.5071556
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> drop1(mod5,.~.,test="F")
Single term deletions

Model:
dur.interv ~ age + dep.cons + subst.cons + scz.cons + prof
           Df Sum of Sq    RSS    AIC F value    Pr(>F)
<none>                  266846 4395.6
age         1    4778.4 271624 4406.8 13.0899 0.0003173 ***
dep.cons    1    9317.1 276163 4419.1 25.5233 5.527e-07 ***
subst.cons  1    3333.6 270180 4402.8  9.1322 0.0025992 **
scz.cons    1     352.8 267199 4394.6  0.9666 0.3258633
prof        7    2295.5 269142 4388.0  0.8983 0.5071556
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> drop1(mod5,.~.,test="F")
Single term deletions

Model:
dur.interv ~ age + dep.cons + subst.cons + scz.cons + prof
           Df Sum of Sq    RSS    AIC F value    Pr(>F)
<none>                  266846 4395.6
age         1    4778.4 271624 4406.8 13.0899 0.0003173 ***
dep.cons    1    9317.1 276163 4419.1 25.5233 5.527e-07 ***
subst.cons  1    3333.6 270180 4402.8  9.1322 0.0025992 **
scz.cons    1     352.8 267199 4394.6  0.9666 0.3258633
prof        7    2295.5 269142 4388.0  0.8983 0.5071556
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> mod3 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons,data=smp.l)
> summary(mod3)

Call:
lm(formula = dur.interv ~ age + dep.cons + subst.cons + scz.cons,
    data = smp.l)

Residuals:
    Min      1Q  Median      3Q     Max
-63.654 -14.522  -1.193  11.482  62.482

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 48.90105    2.62213  18.649  < 2e-16 ***
age          0.22096    0.05708   3.871 0.000118 ***
dep.cons     7.38932    1.44783   5.104 4.24e-07 ***
subst.cons   5.25157    1.74318   3.013 0.002678 **
scz.cons     2.27256    2.52323   0.901 0.368062
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.1 on 742 degrees of freedom
  (52 observations deleted due to missingness)
Multiple R-squared:  0.05833,   Adjusted R-squared:  0.05325
F-statistic: 11.49 on 4 and 742 DF,  p-value: 4.692e-09
```

```
> mod3 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons,data=smp.l)
> summary(mod3)

Call:
lm(formula = dur.interv ~ age + dep.cons + subst.cons + scz.cons,
    data = smp.l)

Residuals:
    Min      1Q  Median      3Q     Max
-63.654 -14.522  -1.193  11.482  62.482

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 48.90105    2.62213  18.649  < 2e-16 ***
age          0.22096    0.05708   3.871 0.000118 ***
dep.cons     7.38932    1.44783   5.104 4.24e-07 ***
subst.cons   5.25157    1.74318   3.013 0.002678 **
scz.cons     2.27256    2.52323   0.901 0.368062
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.1 on 742 degrees of freedom
  (52 observations deleted due to missingness)
Multiple R-squared:  0.05833,   Adjusted R-squared:  0.05325
F-statistic: 11.49 on 4 and 742 DF,  p-value: 4.692e-09
```

Interaction entre deux variables explicatives

```
> mod6 <- lm(dur.interv~age+dep.cons*subst.cons+scz.cons,data=smp.l)
> summary(mod6)

Call:
lm(formula = dur.interv ~ age + dep.cons * subst.cons + scz.cons,
    data = smp.l)

Residuals:
    Min      1Q  Median      3Q     Max
-62.032 -14.251  -1.163  11.472  62.313

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        49.51693    2.65788  18.630  < 2e-16 ***
age                 0.21728    0.05711   3.805 0.000154 ***
dep.cons            6.15780    1.69775   3.627 0.000306 ***
subst.cons          3.17244    2.29849   1.380 0.167931
scz.cons            1.97233    2.53094   0.779 0.436059
dep.cons:subst.cons 4.49688    3.24296   1.387 0.165963
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.08 on 741 degrees of freedom
  (52 observations deleted due to missingness)
Multiple R-squared:  0.06077,   Adjusted R-squared:  0.05443
F-statistic: 9.588 on 5 and 741 DF,  p-value: 7.024e-09
```

Durée = a + b × age + c × dep + d × subst + e × scz + f × (dep × subst) + bruit

Interaction ⇔ synergie

```
> mod6 <- lm(dur.interv~age+dep.cons*subst.cons+scz.cons,data=smp.l)
> summary(mod6)

Call:
lm(formula = dur.interv ~ age + dep.cons * subst.cons + scz.cons,
    data = smp.l)

Residuals:
    Min      1Q  Median      3Q     Max
-62.032 -14.251  -1.163  11.472  62.313

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        49.51693    2.65788  18.630  < 2e-16 ***
age                 0.21728    0.05711   3.805 0.000154 ***
dep.cons            6.15780    1.69775   3.627 0.000306 ***
subst.cons          3.17244    2.29849   1.380 0.167931
scz.cons            1.97233    2.53094   0.779 0.436059
dep.cons:subst.cons 4.49688    3.24296   1.387 0.165963
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.08 on 741 degrees of freedom
  (52 observations deleted due to missingness)
Multiple R-squared:  0.06077,   Adjusted R-squared:  0.05443
F-statistic: 9.588 on 5 and 741 DF,  p-value: 7.024e-09
```

Durée = a + b × age + c × dep + d × subst + e × scz + f × (dep × subst) + bruit

Interaction ⇔ synergie

Interaction entre deux variables explicatives

```
> mod6 <- lm(dur.interv~age+dep.cons*subst.cons+scz.cons,data=smp.l)
> summary(mod6)

Call:
lm(formula = dur.interv ~ age + dep.cons * subst.cons + scz.cons,
    data = smp.l)

Residuals:
    Min      1Q  Median      3Q     Max
-62.032 -14.251  -1.163  11.472  62.313

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        49.51693    2.65788  18.630  < 2e-16 ***
age                 0.21728    0.05711   3.805 0.000154 ***
dep.cons            6.15780    1.69775   3.627 0.000306 ***
subst.cons          3.17244    2.29849   1.380 0.167931
scz.cons            1.97233    2.53094   0.779 0.436059
dep.cons:subst.cons 4.49688    3.24296   1.387 0.165963
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.08 on 741 degrees of freedom
  (52 observations deleted due to missingness)
Multiple R-squared:  0.06077,   Adjusted R-squared:  0.05443
F-statistic: 9.588 on 5 and 741 DF,  p-value: 7.024e-09
```

Durée = a + b × age + c × dep + d × subst + e × scz + f × (dep × subst) + bruit

Interaction ⇔ synergie

ANalysis Of VAriance

=

ANOVA

=

régression linéaire
où les variables explicatives
sont toutes catégorielles !

```
> mod7 <- lm(dur.interv~prof,data=smp.l)
```

ANalysis Of VAriance

=

ANOVA

=

régression linéaire
où les variables explicatives
sont toutes catégorielles !

ANalysis Of VAriance

=

ANOVA

=

régression linéaire
où les variables explicatives
sont toutes catégorielles !

```
> mod7 <- lm(dur.interv~prof,data=smp.1)
> summary(mod7)

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

Residuals:
    Min      1Q  Median      3Q     Max
-61.731 -13.826  -1.731  12.947  58.912

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)             61.7315     1.3359  46.211   <2e-16 ***
profagriculteur         17.0185     9.9071   1.718   0.0863 .
profartisan              2.0941     2.5033   0.837   0.4031
profautre                2.4993     4.0755   0.613   0.5399
profcadre               -4.7750     4.3063  -1.109   0.2679
profemployé              0.3220     2.1742   0.148   0.8823
profprof.intermédiaire   1.3440     3.0096   0.447   0.6553
profsans emploi         -0.6432     1.9168  -0.336   0.7373
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.63 on 735 degrees of freedom
  (56 observations deleted due to missingness)
Multiple R-squared:  0.008295,  Adjusted R-squared:  -0.001149
F-statistic: 0.8783 on 7 and 735 DF,  p-value: 0.5231
```

ANalysis Of VAriance

=

ANOVA

=

régression linéaire
où les variables explicatives
sont toutes catégorielles !

```
> mod7 <- lm(dur.interv~prof,data=smp.1)
> summary(mod7)

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

Residuals:
    Min      1Q  Median      3Q     Max
-61.731 -13.826  -1.731  12.947  58.912

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)            61.7315     1.3359  46.211   <2e-16 ***
profagriculteur        17.0185     9.9071   1.718   0.0863 .
profartisan             2.0941     2.5033   0.837   0.4031
profautre               2.4993     4.0755   0.613   0.5399
profcadre              -4.7750     4.3063  -1.109   0.2679
profemployé             0.3220     2.1742   0.148   0.8823
profprof.intermédiaire  1.3440     3.0096   0.447   0.6553
profsans emploi        -0.6432     1.9168  -0.336   0.7373
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.63 on 735 degrees of freedom
  (56 observations deleted due to missingness)
Multiple R-squared:  0.008295,  Adjusted R-squared:  -0.001149
F-statistic: 0.8783 on 7 and 735 DF,  p-value: 0.5231

> drop1(mod7,.~.,test="F")
Single term deletions

Model:
dur.interv ~ prof
       Df Sum of Sq    RSS    AIC F value Pr(>F)
<none>              283316 4432.1
prof    7    2369.9 285686 4424.3  0.8783 0.5231
```

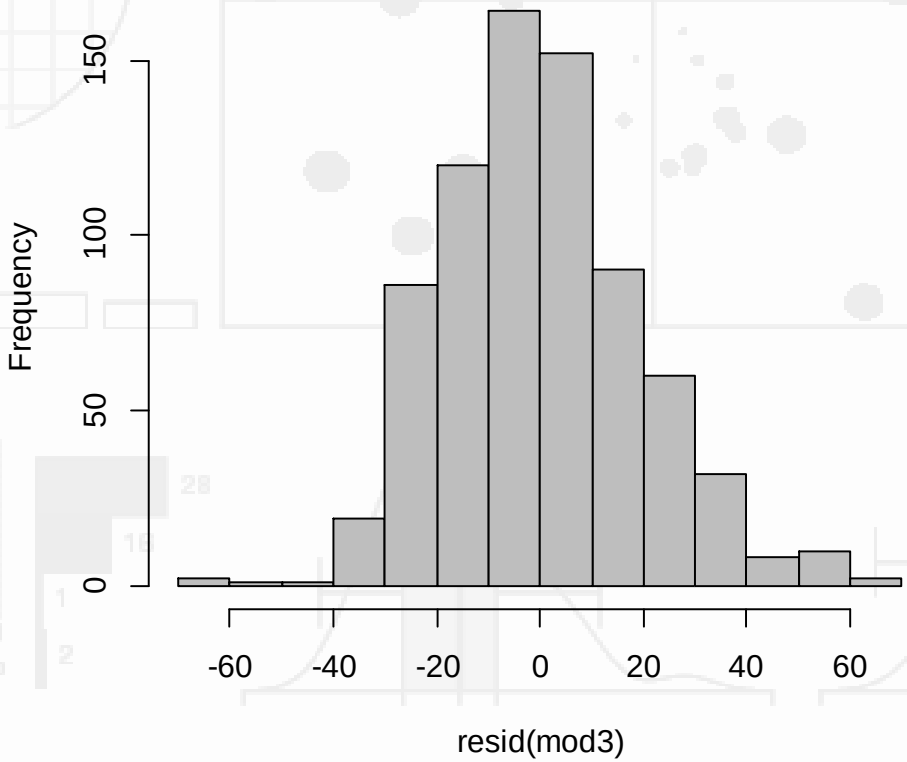
- Normalité du « bruit »

- Normalité du « bruit »
- La variance du bruit ne doit dépendre ni des valeurs de la variable à expliquer, ni des valeurs des variables explicatives

- Normalité du « bruit »
- La variance du bruit ne doit dépendre ni des valeurs de la variable à expliquer, ni des valeurs des variables explicatives
- Le bruit doit être un « vrai » bruit (pas de structure de corrélation évidente)

Conditions de validité

```
> mod3 <- lm(dur.interv~age+dep.cons+
  subst.cons+scz.cons,data=smp.l)
> hist(resid(mod3), col="grey", main="")
```



```
mod3 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons,data=smp.l)
summary(mod3)
mod4 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons+prof,data=smp.l)
summary(mod4)
smp.l$prof <- relevel(smp.l$prof,ref="ouvrier")
mod5 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons+prof,data=smp.l)
summary(mod5)
drop1(mod5,.~.,test="F")

mod6 <- lm(dur.interv~age+dep.cons*subst.cons+scz.cons,data=smp.l)
summary(mod6)

mod7 <- lm(dur.interv~prof,data=smp.l)
summary(mod7)
drop1(mod7,.~.,test="F")

mod3 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons,data=smp.l)
hist(resid(mod3), col="grey", main="")
```

```
mod3 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons,data=smp.l)
summary(mod3)
mod4 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons+prof,data=smp.l)
summary(mod4)
smp.l$prof <- relevel(smp.l$prof,ref="ouvrier")
mod5 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons+prof,data=smp.l)
summary(mod5)
drop1(mod5,.~.,test="F")

mod6 <- lm(dur.interv~age+dep.cons*subst.cons+scz.cons,data=smp.l)
summary(mod6)

mod7 <- lm(dur.interv~prof,data=smp.l)
summary(mod7)
drop1(mod7,.~.,test="F")

mod3 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons,data=smp.l)
hist(resid(mod3), col="grey", main="")
```

```
mod3 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons,data=smp.l)
summary(mod3)
mod4 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons+prof,data=smp.l)
summary(mod4)
smp.l$prof <- relevel(smp.l$prof,ref="ouvrier")
mod5 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons+prof,data=smp.l)
summary(mod5)
drop1(mod5,.~.,test="F")

mod6 <- lm(dur.interv~age+dep.cons*subst.cons+scz.cons,data=smp.l)
summary(mod6)

mod7 <- lm(dur.interv~prof,data=smp.l)
summary(mod7)
drop1(mod7,.~.,test="F")

mod3 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons,data=smp.l)
hist(resid(mod3), col="grey", main="")
```

```
mod3 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons,data=smp.l)
summary(mod3)
mod4 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons+prof,data=smp.l)
summary(mod4)
smp.l$prof <- relevel(smp.l$prof,ref="ouvrier")
mod5 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons+prof,data=smp.l)
summary(mod5)
drop1(mod5,.~.,test="F")

mod6 <- lm(dur.interv~age+dep.cons*subst.cons+scz.cons,data=smp.l)
summary(mod6)

mod7 <- lm(dur.interv~prof,data=smp.l)
summary(mod7)
drop1(mod7,.~.,test="F")

mod3 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons,data=smp.l)
hist(resid(mod3), col="grey", main="")
```

```
mod3 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons,data=smp.l)
summary(mod3)
mod4 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons+prof,data=smp.l)
summary(mod4)
smp.l$prof <- relevel(smp.l$prof,ref="ouvrier")
mod5 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons+prof,data=smp.l)
summary(mod5)
drop1(mod5,.~.,test="F")

mod6 <- lm(dur.interv~age+dep.cons*subst.cons+scz.cons,data=smp.l)
summary(mod6)

mod7 <- lm(dur.interv~prof,data=smp.l)
summary(mod7)
drop1(mod7,.~.,test="F")

mod3 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons,data=smp.l)
hist(resid(mod3), col="grey", main="")
```

```
mod3 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons,data=smp.l)
summary(mod3)
mod4 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons+prof,data=smp.l)
summary(mod4)
smp.l$prof <- relevel(smp.l$prof,ref="ouvrier")
mod5 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons+prof,data=smp.l)
summary(mod5)
drop1(mod5,.~.,test="F")

mod6 <- lm(dur.interv~age+dep.cons*subst.cons+scz.cons,data=smp.l)
summary(mod6)

mod7 <- lm(dur.interv~prof,data=smp.l)
summary(mod7)
drop1(mod7,.~.,test="F")

mod3 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons,data=smp.l)
hist(resid(mod3), col="grey", main="")
```

```
mod3 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons,data=smp.l)
summary(mod3)
mod4 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons+prof,data=smp.l)
summary(mod4)
smp.l$prof <- relevel(smp.l$prof,ref="ouvrier")
mod5 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons+prof,data=smp.l)
summary(mod5)
drop1(mod5,.~.,test="F")

mod6 <- lm(dur.interv~age+dep.cons*subst.cons+scz.cons,data=smp.l)
summary(mod6)

mod7 <- lm(dur.interv~prof,data=smp.l)
summary(mod7)
drop1(mod7,.~.,test="F")

mod3 <- lm(dur.interv~age+dep.cons+subst.cons+scz.cons,data=smp.l)
hist(resid(mod3), col="grey", main="")
```