Les tests statistiques en pratique : comparaison de deux pourcentages

Pr. Bruno Falissard

- Comparer deux pourcentages : le test du « Chi-2 »
- Attention aux faibles effectifs et aux petits pourcentages → test exact de Fisher

```
> smp.c$ed.b <- ifelse(smp.c$ed>2,1,0)
> str(smp.c)
'data.frame':   799 obs. of  10 variables:
 $ age      : int  31 49 50 47 23 34 24 52 42 45 ...
 $ prof     : Factor w/ 8 levels "agriculteur",..: 3 NA 7 6 8 6 3 ...
 $ dep.cons : int  0 0 0 0 1 0 1 0 1 0 ...
 $ scz.cons : int  0 0 0 0 0 0 0 0 0 0 ...
 $ grav.cons: int  1 2 2 1 2 1 5 1 5 5 ...
 $ n.enfant : int  2 7 2 0 1 3 5 2 1 2 ...
 $ rs       : int  2 2 2 2 2 1 3 2 3 2 ...
 $ ed       : int  1 2 3 2 2 2 3 2 3 2 ...
 $ dr       : int  1 1 2 2 2 1 2 2 1 2 ...
 $ ed.b     : num  0 0 1 0 0 0 1 0 1 0 ...
```

```
> table(smp.c$ed.b,smp.c$dep.cons,deparse.level=2,useNA="always")
          smp.c$dep.cons
smp.c$ed.b    0   1 <NA>
       0    335 135    0
       1     96 126    0
       <NA>  51  56    0
```

```
> table(smp.c$ed.b,smp.c$dep.cons,deparse.level=2,useNA="always")
          smp.c$dep.cons
smp.c$ed.b    0    1 <NA>
      0     335  135    0
      1      96  126    0
      <NA>   51   56    0
```

```
> table(smp.c$ed.b,smp.c$dep.cons,deparse.level=2,useNA="always")
          smp.c$dep.cons
smp.c$ed.b    0    1 <NA>
      0     335  135    0
      1      96  126    0
      <NA>   51   56    0
```

```
> table(smp.c$ed.b,smp.c$dep.cons,deparse.level=2,useNA="always")
          smp.c$dep.cons
smp.c$ed.b    0    1 <NA>
      0     335  135    0
      1      96  126    0
      <NA>   51   56    0
> tab <- table(smp.c$ed.b,smp.c$dep.cons,deparse.level=2)
```

```
> table(smp.c$ed.b,smp.c$dep.cons,deparse.level=2,useNA="always")
          smp.c$dep.cons
smp.c$ed.b    0    1 <NA>
      0     335  135    0
      1      96  126    0
      <NA>   51   56    0
> tab <- table(smp.c$ed.b,smp.c$dep.cons,deparse.level=2)
> prop.table(tab,1)
          smp.c$dep.cons
smp.c$ed.b         0         1
         0 0.7127660 0.2872340
         1 0.4324324 0.5675676
```

```
> table(smp.c$ed.b,smp.c$dep.cons,deparse.level=2,useNA="always")
          smp.c$dep.cons
smp.c$ed.b    0    1 <NA>
      0     335  135    0
      1      96  126    0
      <NA>   51   56    0
> tab <- table(smp.c$ed.b,smp.c$dep.cons,deparse.level=2)
> prop.table(tab,1)
          smp.c$dep.cons
smp.c$ed.b         0         1
         0 0.7127660 0.2872340
         1 0.4324324 0.5675676
```

```
> table(smp.c$ed.b,smp.c$dep.cons,deparse.level=2,useNA="always")
          smp.c$dep.cons
smp.c$ed.b    0    1 <NA>
      0     335  135    0
      1      96  126    0
      <NA>   51   56    0
> tab <- table(smp.c$ed.b,smp.c$dep.cons,deparse.level=2)
> prop.table(tab,1)
          smp.c$dep.cons
smp.c$ed.b         0         1
         0 0.7127660 0.2872340
         1 0.4324324 0.5675676
```

```
> table(smp.c$ed.b,smp.c$dep.cons,deparse.level=2,useNA="always")
          smp.c$dep.cons
smp.c$ed.b    0    1 <NA>
      0     335  135    0
      1      96  126    0
      <NA>   51   56    0
> tab <- table(smp.c$ed.b,smp.c$dep.cons,deparse.level=2)
> prop.table(tab,1)
          smp.c$dep.cons
smp.c$ed.b         0         1
         0 0.7127660 0.2872340
         1 0.4324324 0.5675676
> prop.table(tab,2)
          smp.c$dep.cons
smp.c$ed.b         0         1
         0 0.7772622 0.5172414
         1 0.2227378 0.4827586
```

```
> table(smp.c$ed.b,smp.c$dep.cons,deparse.level=2,useNA="always")
          smp.c$dep.cons
smp.c$ed.b    0    1 <NA>
      0     335  135    0
      1      96  126    0
      <NA>   51   56    0
> tab <- table(smp.c$ed.b,smp.c$dep.cons,deparse.level=2)
> prop.table(tab,1)
          smp.c$dep.cons
smp.c$ed.b          0         1
         0 0.7127660 0.2872340
         1 0.4324324 0.5675676
> prop.table(tab,2)
          smp.c$dep.cons
smp.c$ed.b          0         1
         0 0.7772622 0.5172414
         1 0.2227378 0.4827586
```

```
> chisq.test(smp.c$ed.b,smp.c$dep.cons,correct=FALSE)

        Pearson's Chi-squared test

data:  smp.c$ed.b and smp.c$dep.cons
X-squared = 50.4416, df = 1, p-value = 1.228e-12
```

```
> chisq.test(smp.c$ed.b,smp.c$dep.cons,correct=FALSE)

        Pearson's Chi-squared test

data:  smp.c$ed.b and smp.c$dep.cons
X-squared = 50.4416, df = 1, p-value = 1.228e-12
```

```
> chisq.test(smp.c$ed.b,smp.c$dep.cons,correct=FALSE)

        Pearson's Chi-squared test

data:  smp.c$ed.b and smp.c$dep.cons
X-squared = 50.4416, df = 1, p-value = 1.228e-12
```

```
Message d'avis :
In chisq.test(x) : l'approximation du Chi-2 est peut-être incorrecte
```

```
> fisher.test(smp.c$ed.b,smp.c$dep.cons)

        Fisher's Exact Test for Count Data

data:  smp.c$ed.b and smp.c$dep.cons
p-value = 2.033e-12
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 2.303664 4.603460
sample estimates:
odds ratio
  3.250819
```

```
smp.c$ed.b <- ifelse(smp.c$ed>2,1,0)
str(smp.c)
table(smp.c$ed.b,smp.c$dep.cons,deparse.level=2,useNA="always")
tab <- table(smp.c$ed.b,smp.c$dep.cons,deparse.level=2)
prop.table(tab,1)
prop.table(tab,2)
chisq.test(smp.c$ed.b,smp.c$dep.cons,correct=FALSE)
fisher.test(smp.c$ed.b,smp.c$dep.cons)
```