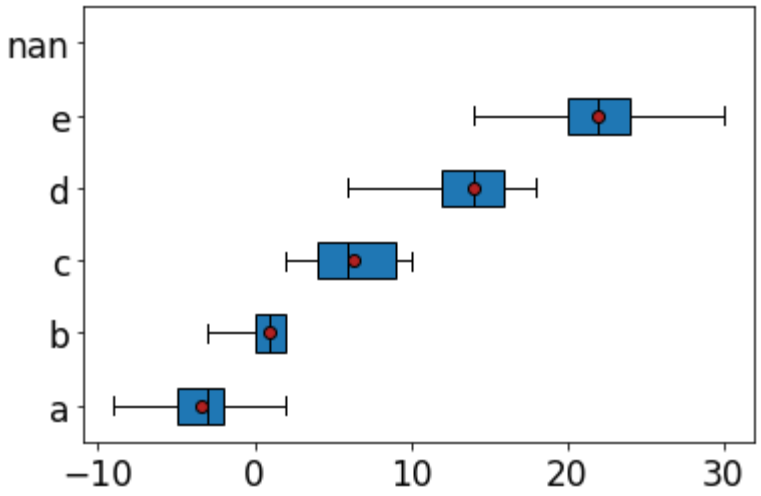
Dépendance entre le nutriscore lettre et le nutriscore point (test de Fisher)

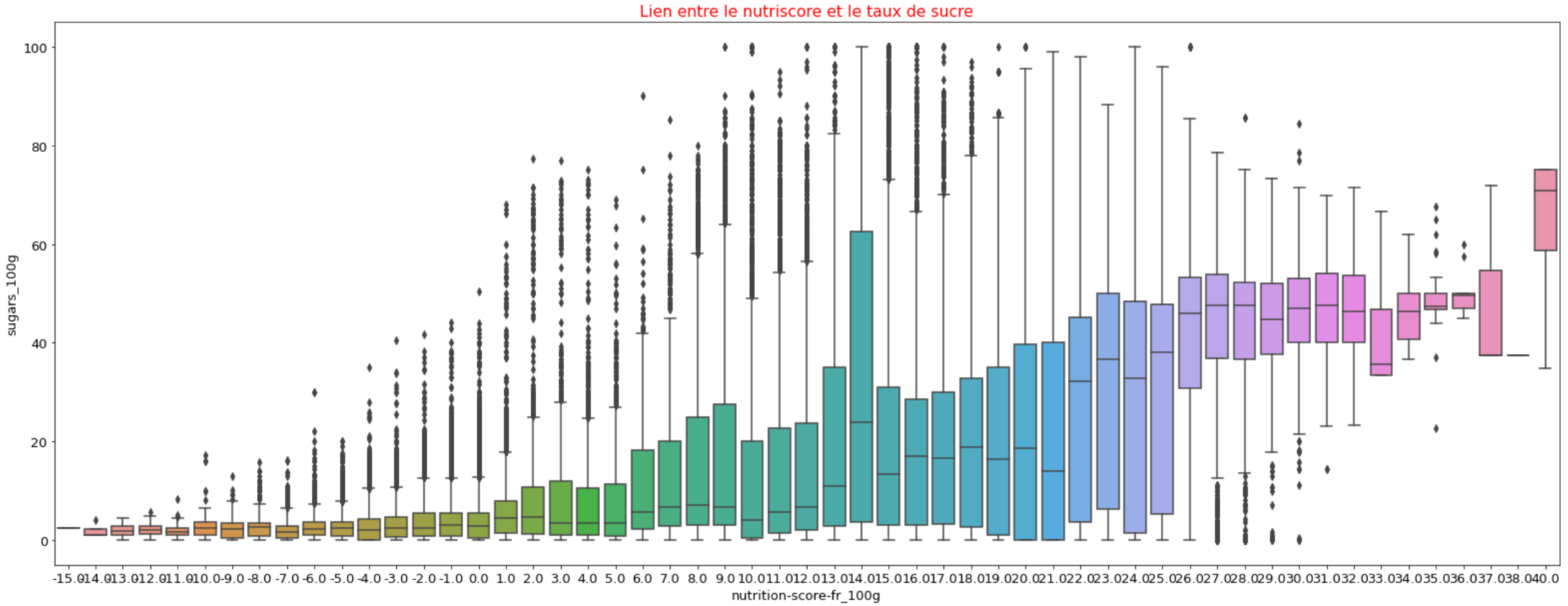



```
anova_nutricore = smf.ols('NutritionScoreFr_100g~NutritionGradeFr', data=df1).fit()
print(anova_nutricore.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     NutritionScoreFr_100g   R-squared:                       0.930
Model:                               OLS   Adj. R-squared:                  0.930
Method:                    Least Squares   F-statistic:                 7.400e+05
Date:                   Wed, 04 May 2022   Prob (F-statistic):               0.00
Time:                           06:51:58   Log-Likelihood:            -5.0643e+05
No. Observations:                 221210   AIC:                         1.013e+06
Df Residuals:                     221205   BIC:                         1.013e+06
Df Model:                              4
Covariance Type:               nonrobust
================================================================================================
                                   coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------------
Intercept                       -3.4076      0.013   -269.372      0.000      -3.432      -3.383
NutritionGradeFr[T.b]            4.3143      0.018    238.746      0.000       4.279       4.350
NutritionGradeFr[T.c]            9.7752      0.017    578.773      0.000       9.742       9.808
NutritionGradeFr[T.d]           17.4746      0.016   1103.231      0.000      17.444      17.506
NutritionGradeFr[T.e]           25.3699      0.017   1483.256      0.000      25.336      25.403
==============================================================================
Omnibus:                     9618.472   Durbin-Watson:                   1.491
```

Réaliser un test de Fisher.
Ce qui nous intéresse réellement, c'est le test de Fisher. La p-valeur de ce test (0.00) est inférieure à 5 %. On rejette donc l'hypothèse H0 selon laquelle $\alpha1=\alpha2=\alpha3=\alpha4=\alpha5=0$.
Il y a bien un lien (une dépendance) entre le nutrisore lettre (A, B, C, D, E) et le nutriscore point(qui échelonne entre -15 et 40) comme nous en avions l'intuition en regardant les boîtes à moustaches.

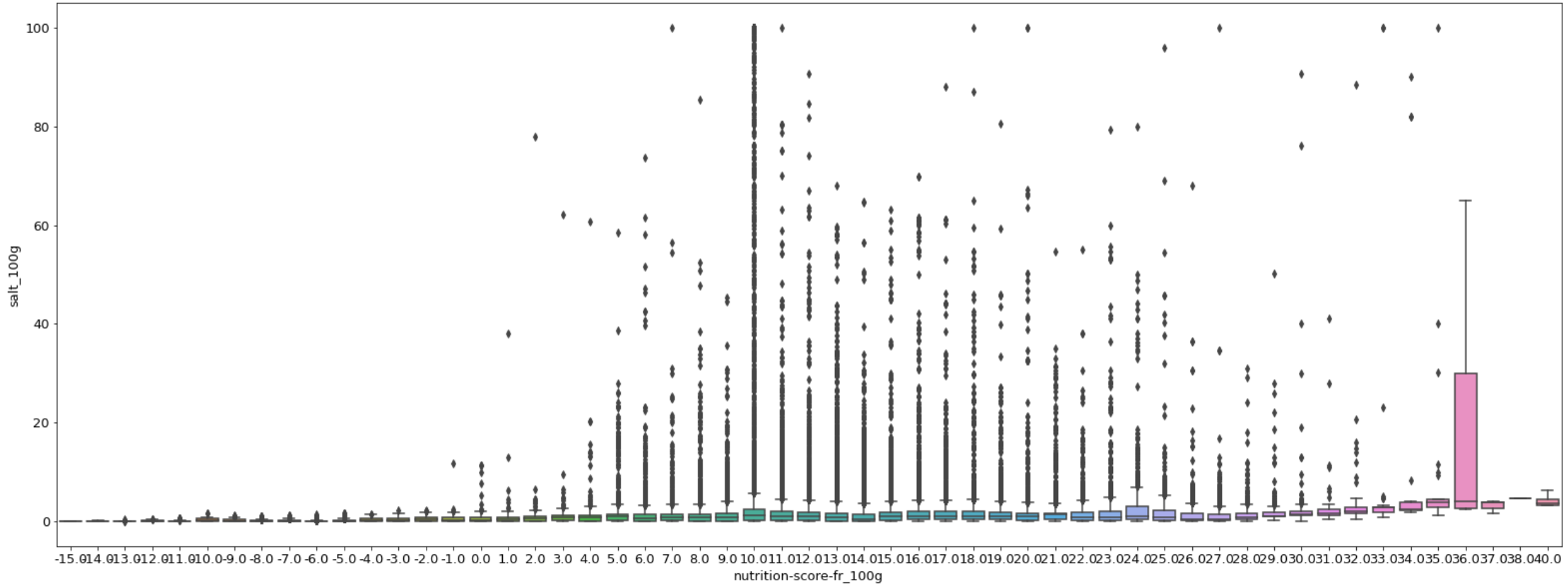
Dépendance entre le nutriscore lettre et taux de sure (test de Fisher)



```
anova_sugars = smf.ols('sugars_100g~NutritionGradeFr', data=df1).fit()
print(anova_sugars.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             sugars_100g   R-squared:                       0.198
Model:                             OLS   Adj. R-squared:                  0.198
Method:                  Least Squares   F-statistic:                 1.363e+04
Date:                 Wed, 04 May 2022   Prob (F-statistic):               0.00
Time:                         07:00:24   Log-Likelihood:            -9.5010e+05
No. Observations:               221210   AIC:                         1.900e+06
Df Residuals:                   221205   BIC:                         1.900e+06
Df Model:                            4
Covariance Type:             nonrobust
==============================================================================
                             coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept                  3.4916      0.094     37.144      0.000       3.307       3.676
NutritionGradeFr[T.b]      1.7887      0.134     13.321      0.000       1.526       2.052
NutritionGradeFr[T.c]      7.7860      0.126     62.039      0.000       7.540       8.032
NutritionGradeFr[T.d]     18.0273      0.118    153.165      0.000      17.797      18.258
NutritionGradeFr[T.e]     23.2261      0.127    182.743      0.000      22.977      23.475
==============================================================================
Omnibus:                    49881.373   Durbin-Watson:                   0.884
```

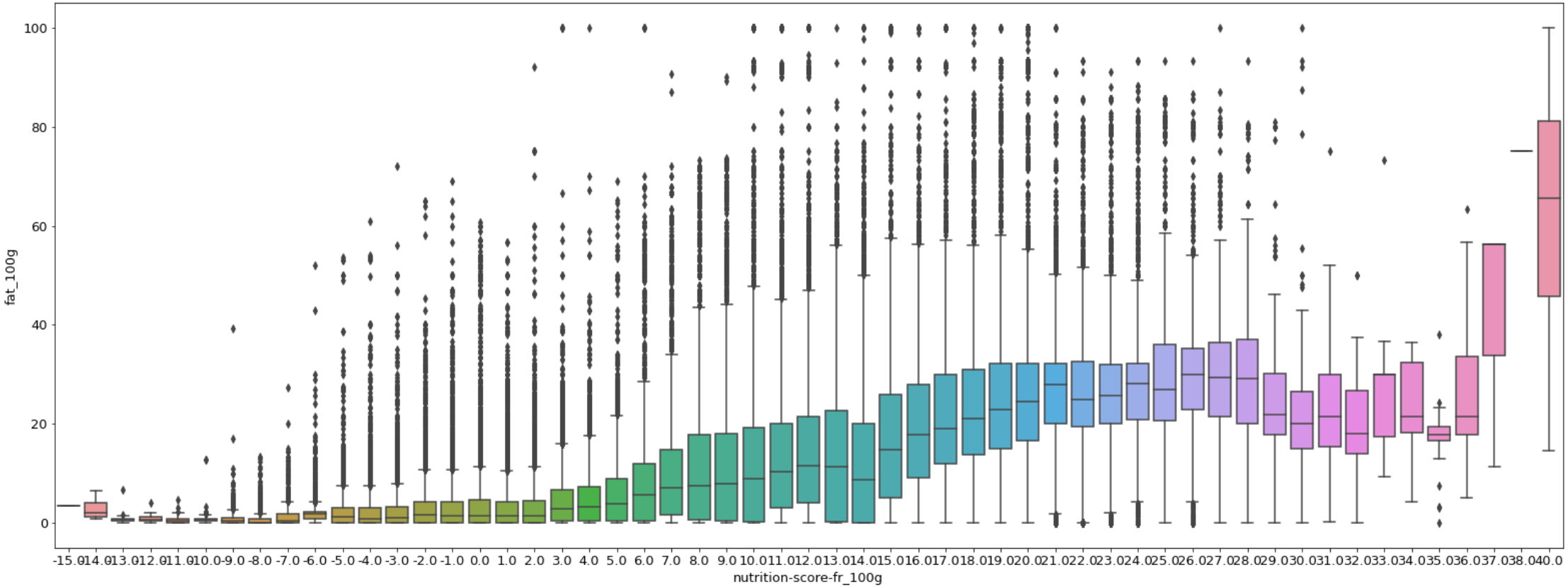
Dépendance entre le nutriscore lettre et taux de sel (test de Fisher)



```
anova_salt = smf.ols('salt_100g~NutritionGradeFr', data=df1).fit()
print(anova_salt.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              salt_100g   R-squared:                       0.020
Model:                            OLS   Adj. R-squared:                  0.020
Method:                 Least Squares   F-statistic:                     1116.
Date:                Wed, 04 May 2022   Prob (F-statistic):               0.00
Time:                        07:03:41   Log-Likelihood:            -6.1733e+05
No. Observations:              221210   AIC:                         1.235e+06
Df Residuals:                  221205   BIC:                         1.235e+06
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                             coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept                  0.3311      0.021     15.854      0.000       0.290       0.372
NutritionGradeFr[T.b]      0.1890      0.030      6.336      0.000       0.131       0.248
NutritionGradeFr[T.c]      1.3701      0.028     49.137      0.000       1.315       1.425
NutritionGradeFr[T.d]      1.2999      0.026     49.709      0.000       1.249       1.351
NutritionGradeFr[T.e]      1.1701      0.028     41.438      0.000       1.115       1.225
==============================================================================
```

Dépendance entre le nutriscore lettre et taux de gras (test de Fisher)



```
anova_fat = smf.ols('fat_100g~NutritionGradeFr', data=df1).fit()
print(anova_fat.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                fat_100g   R-squared:                       0.279
Model:                            OLS   Adj. R-squared:                  0.279
Method:                 Least Squares   F-statistic:                 2.145e+04
Date:                Wed, 04 May 2022   Prob (F-statistic):               0.00
Time:                        07:05:37   Log-Likelihood:            -8.8875e+05
No. Observations:              221210   AIC:                         1.778e+06
Df Residuals:                  221205   BIC:                         1.778e+06
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                             coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept                    2.7830      0.071     39.069      0.000       2.643       2.923
NutritionGradeFr[T.b]        1.3557      0.102     13.323      0.000       1.156       1.555
NutritionGradeFr[T.c]        6.9461      0.095     73.036      0.000       6.760       7.133
NutritionGradeFr[T.d]       14.2866      0.089    160.178      0.000      14.112      14.461
NutritionGradeFr[T.e]       23.4873      0.096    243.861      0.000      23.298      23.676
==============================================================================
```

La p-valeur de ces 3 tests (0.00) est inférieure à 5 %. On rejette donc l'hypothèse H0 selon laquelle $\alpha1=\alpha2=\alpha3=\alpha4=\alpha5=0$.
Il y a bien un lien (une dépendance) entre le nutrisore lettre (A, B, C, D, E) et les taux de sucre, de sel et de gras comme nous en avions l'intuition en regardant les boîtes à moustaches.