Week 5.QuantStat

March 28, 2020

1 Handling response variable: how often did you skip full day in school without an excuse?

Bare in mind, that originally it's a string variable, because of "Skip" answer, which is marked as "
"

2 Statistical check

2.0.1 Null-hypothesis is: religious adolescents skip the same amount of days in school as non-religious adolescents

Religion amongst adolescents is a categorical variable with 4 levels Amount of skipped days is quantative variable

2.0.2 Mean and Standard deviations for amount of skipped days grouped by religiosity

```
Mean value for amount of skipped days by religiosity level
        SCHSKIP
H1RE4
1.0    1.171201
2.0    2.182361
3.0    3.140992
4.0    3.434951
standard deviations for amount of skipped days by religiosity level
         SCHSKIP
H1RE4
1.0     5.672074
2.0     7.696166
3.0     9.995861
4.0    10.046981
```

2.0.3 Ordinary least squared mathod for categorical variable with more than 2 levels

```
                      OLS Regression Results
==============================================================================
Dep. Variable:             SCHSKIP   R-squared:                      0.013
Model:                         OLS   Adj. R-squared:                 0.012
Method:              Least Squares   F-statistic:                    26.96
```

```
Date:                Sat, 28 Mar 2020   Prob (F-statistic):           2.54e-17
Time:                      19:36:57     Log-Likelihood:               -21817.
No. Observations:             6347       AIC:                          4.364e+04
Df Residuals:                 6343       BIC:                          4.367e+04
Df Model:                        3
Covariance Type:          nonrobust
========================================================================
===
                      coef     std err        t       P>|t|      [0.025
0.975]
------------------------------------------------------------------------
---
Intercept           1.1712      0.143      8.168      0.000      0.890
1.452
C(H1RE4)[T.2.0]     1.0112      0.216      4.684      0.000      0.588
1.434
C(H1RE4)[T.3.0]     1.9698      0.411      4.798      0.000      1.165
2.775
C(H1RE4)[T.4.0]     2.2638      0.275      8.233      0.000      1.725
2.803
========================================================================
Omnibus:                      8753.169   Durbin-Watson:                  1.944
Prob(Omnibus):                   0.000   Jarque-Bera (JB):        1962376.468
Skew:                            8.109   Prob(JB):                       0.00
Kurtosis:                       87.601   Cond. No.                       4.89
========================================================================
```

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.

Based on p = 2.54e-17 Null-hypothesis can be rejected, which means, that adolescents, for whom
religion is more important tend to skip less days in school without explanations.

2.0.4 Multiple comparisons

```
Multiple Comparison of Means - Tukey HSD, FWER=0.05
==================================================
group1 group2 meandiff p-adj   lower   upper  reject
--------------------------------------------------
   1.0    2.0   1.0112  0.001  0.4564 1.5659   True
   1.0    3.0   1.9698  0.001  0.9147 3.0249   True
   1.0    4.0   2.2638  0.001  1.5572 2.9703   True
   2.0    3.0   0.9586 0.0986 -0.1135 2.0307  False
   2.0    4.0   1.2526  0.001  0.5209 1.9843   True
   3.0    4.0    0.294     0.9 -0.864  1.4519  False
--------------------------------------------------
```

Multiple comparisons shows that there's significant difference between amount of skipped days for

those adolescents, who consider religion to be a very important important part of their lifes from all the other groups. And also there's a difference between those, for whom religion is fairly important from those, for whom it is not important at all.