

```

#!/usr/bin/python3

import pickle

# Lecture de La base de données.
#
# Il faudra vraisemblablement adapter Le chemin vers Le fichier wooldridge.pkl.

fid = open('C:/Users/claire.loupias/Desktop/wooldridge23.pkl', 'rb')
alldatasets = pickle.load(fid)
fid.close()

dataset = alldatasets['discrim']

#
# Question 1.
#

import statsmodels.formula.api as smf
import numpy as np

dataset['logged_psoda'] = np.log(dataset['psoda'])
dataset['logged_income'] = np.log(dataset['income'])

# Estimation du modèle avec rbisyr
results = smf.ols('logged_psoda ~ prpblick + logged_income + prppov', data=dataset).fit()
print(results.summary())

#
# Question 2.
#

from scipy.stats import pearsonr

nonan = ~np.isnan(dataset['prppov']) # À cause d'une variable manquante (ne pose pas de problème
# utilisée pour calculer la corrélation n'est pas robuste c
corr, _ = pearsonr(dataset.logged_income[nonan], dataset.prppov[nonan])
print('La corrélation entre log(income) et prppov est ', corr)

#
# Question 3.
#

dataset['logged_hseval'] = np.log(dataset['hseval'])

results_nc = smf.ols('logged_psoda ~ prpblick + logged_income + prppov + logged_hseval', data=dat
print(results_nc.summary())

#
# Question 4.
#

results_c = smf.ols('logged_psoda ~ prpblick + logged_hseval', data=dataset).fit()

from scipy import stats

F = (results_c.ssr-results_nc.ssr)/2/(results_nc.ssr/results_nc.df_resid)
pval = 1-stats.f.cdf(F, 2, results_nc.df_resid)

print('La pvalue du test de nullité jointe des coefficients associés à log(income) et prppov est

```