4.4 Word Problem

1. (i) Null Hypothesis:

$$H_0 : B_3 = 0$$

The size of the student body to population has no effect on monthly rent.

Alternative hypothesis:

$$H_1 : B_3 \neq 0$$

There's an effect of student population on monthly rent.

(ii) Since B1 signifies the relationship between city population and log of rent, and we assume B1 to be positive, so if there's an increase in city population, there's an increase in rent.
Since B2 signifies the relationship between avg city income and log of rent, and we assume B2 to be positive, so by ceteris paribus, an increase in avg city income will also increase rent.

(iii) It's incorrect based on provided regression results. Since the coefficient for log(pop) is 0.066 percent, then it means that a 1 percent increase in population is associated with a 0.066 percent increase in rent, and not 6.6 percent.

(iv) Given t_stat = 3.29 and our degree of freedom = 60, after using the two-tailed test, we compare t-stat with t_critical. Since 2.66 < 3.29, it fails to reject the null hypothesis at 1 percent level.

(v) 90% Confidence Interval for B1: (0.0008446942786268785, 0.13115530572137313)

(vi) 95% Confidence Interval for B2: (0.34497587649428085, 0.6690241235057192)

(vii) We are 90.0% confident that the true value of B1 is greater than 0.066.
The probability that B1 is greater than 0.066 is approximately 55.00 percent.

(viii) From code, we found that the p-value for the two-sided test of H0: B1 = 0 is approximately: 0.096

2. t-statics is calculated by estimated coefficient divide by its standard error. With more variable comparison in multi-regression, we get to standardize the coefficients. In addition, different variables to see significance of these variables in comparing their variations to the dependent variables.

Computer based problems output:

#1 (i)
Number of single-person households is 2017

#1(ii)

OLS Regression Results

Dep. Variable:	nettfa	R-squared:	0.119
Model:	OLS	Adj. R-squared:	0.118
Method:	Least Squares	F-statistic:	136.5
Date:	Tue, 17 Oct 2023	Prob (F-statistic):	2.63e-56
Time:	14:12:40	Log-Likelihood:	-10524.
No. Observations:	2017	AIC:	2.105e+04
Df Residuals:	2014	BIC:	2.107e+04
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-43.0398	4.080	-10.548	0.000	-51.042	-35.038
inc	0.7993	0.060	13.382	0.000	0.682	0.916
age	0.8427	0.092	9.158	0.000	0.662	1.023

Omnibus:	3738.399	Durbin-Watson:	1.976
Prob(Omnibus):	0.000	Jarque-Bera (JB):	6761967.347
Skew:	13.359	Prob(JB):	0.00
Kurtosis:	285.393	Cond. No.	209.

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

coefficient for inc indicates that the relationship between age and net finical wealth is positive. Also, the coefficient for age indicates that there's a positive relationship between age and net financial wealth, which makes sense.
The surprise is the number for intercept, -43.04. We're not sure why it's a negative value, and may need more information to interpret it.

#1 (v)

OLS Regression Results

Dep. Variable:	nettfa	R-squared:	0.083
Model:	OLS	Adj. R-squared:	0.082
Method:	Least Squares	F-statistic:	181.6
Date:	Tue, 17 Oct 2023	Prob (F-statistic):	1.08e-39
Time:	14:12:40	Log-Likelihood:	-10565.
No. Observations:	2017	AIC:	2.113e+04
Df Residuals:	2015	BIC:	2.115e+04
Df Model:	1		

Covariance Type: nonrobust

	coef	std err	t	P>|t|	[0.025	0.975]
const	-10.5710	2.061	-5.130	0.000	-14.612	-6.530
inc	0.8207	0.061	13.476	0.000	0.701	0.940

Omnibus:	3687.885	Durbin-Watson:	1.966
Prob(Omnibus):	0.000	Jarque-Bera (JB):	6061860.886
Skew:	12.998	Prob(JB):	0.00
Kurtosis:	270.308	Cond. No.	68.7

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

The coefficient for inc in part 2 is 0.7993 and the new inc value is 0.8207. Since the difference is small, then
additional variable age didn't change the estimated relationship between inc and nettfa. Also, coefficients for
both cases are postive so there's a positive relationship between income and new financial wealth.

#1 (iv)

Since p value is les than 0.01, we reject the null hypothesis at 1 percent significance level.
age is statistically siginicant.

#2 (a)

	ibm	gm	ge	jnj	dowjones
count	96.000000	96.000000	96.000000	96.000000	96.000000
mean	0.004511	-0.001077	0.000352	0.006672	0.002285
std	0.089003	0.114665	0.060587	0.053222	0.039925
min	-0.226375	-0.241359	-0.176766	-0.160301	-0.123688
25%	-0.038953	-0.072206	-0.026785	-0.020052	-0.018721
50%	-0.001212	-0.005298	0.001441	0.001630	0.003320
75%	0.046534	0.062756	0.036834	0.037881	0.026217
max	0.353782	0.260413	0.179080	0.174327	0.106047

Simple Linear Regression - Confidence Interval for IBM Slope:

	0	1
const	-0.010121	0.007719
ibm	0.243707	0.444939

Multiple Linear Regression - Confidence Interval for IBM Slope:

	0	1
const	-0.009964	0.006073
ibm	-0.005579	0.234332
dowjones	0.512430	1.047261

#2 (b)

Test for Constraints

	coef	std err	t	P>|t|	[0.025	0.975]
c0	0.7798	0.161	-1.368	0.175	0.460	1.099

The null hypothesis means that a one-unit change in the dowjones is equal to 1.
The p value is 0.175
At a 5 percent significance level, the p value of 0.175 is greater than 0.05, thus, we don't have enough evidence to reject the null hypothesis.
At a 10 percent level, the p value is still greater than 0.1, thus we fail to reject the null hypothesis

HW5/HW5.py

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm


# Chapter 2, Computer Exercise C4

# Include similar code as HW1 for the basic steps
file_location = "/Users/amyliang/Eco441K/HW5/401ksubs.dta"
f2 = "/Users/amyliang/Eco441K/HW5/stocks.dta"
df1 = pd.read_stata(f2)
df= pd.read_stata(file_location)

pd.set_option('display.max_columns', None)

# Wooldridge Computer Exercise C4.8
print("Wooldridge Computer Exercise C4.8")

print(df.describe())
print()
print("#1 (i)")

# Filter the dataset for single-person households
single_person_households = df[df['fsize'] == 1]

# Count the number of rows in the filtered dataset
num_single_person_households = single_person_households.shape[0]

# Display the result
print("Number of single-person households is", num_single_person_households)

print()

print("#1(ii)")

# Filter the dataset for single-person households
single_person_data = df[df['fsize'] == 1]

# Define the dependent variable (net financial wealth) and independent variables (income and age)
y = single_person_data['nettfa']
X = single_person_data[['inc', 'age']]

# Add a constant term to the independent variables matrix
X = sm.add_constant(X)

```

```
# Fit the OLS model
model = sm.OLS(y, X).fit()

# Print the regression results
print(model.summary())
ans = """
coefficent for inc indicates that the relatioship between age and net fincial wealth is positive.
Also, the coefficent for age indicates that there's a positive relationship between age and net financial wealth, which makes sense.
The surprise is the number for intercept, -43.04. We're not sure why it's a negative value, and may need more information to interpret it.
"""
print(ans)

print("#1 (v)")

# Filter the dataset for single-person households
single_person_data = df[df['fsize'] == 1]

# Define the dependent variable (net financial wealth) and independent variable (income)
y_simple = single_person_data['nettfa']
X_simple = single_person_data[['inc']]

# Add a constant term to the independent variable matrix
X_simple = sm.add_constant(X_simple)

# Fit the simple OLS model
model_simple = sm.OLS(y_simple, X_simple).fit()

# Print the regression results for the simple model
print(model_simple.summary())
ans_1 = """
The coefficient for inc in part 2 is 0.7993 and the new inc value is 0.8207. Since the difference is small, then
additional variable age didn't change the estimated relationship between inc and nettfa. Also, coefficients for
both cases are postive so there's a positive relationship between income and new financial wealth.
"""


print(ans_1)

print()
print("#1 (iv)")
ans_2 = """
Since p value is les than 0.01, we reject the null hypothesis at 1 percent significance level.
age is statistically siginicant.
"""

```

```
print(ans_2)

print()

print("#2 (a)")
print(df1.describe())
# Simple Linear Regression of GE on IBM
model_simple = sm.OLS(df1['ge'], sm.add_constant(df1['ibm'])).fit()
ci_simple = model_simple.conf_int(alpha=0.1)  # 90% confidence interval

# Multiple Linear Regression of GE on IBM and Dow Jones
model_multiple = sm.OLS(df1['ge'], sm.add_constant(df1[['ibm', 'dowjones']]))
.fit()
ci_multiple = model_multiple.conf_int(alpha=0.1)  # 90% confidence interval

# Display the results
print("Simple Linear Regression - Confidence Interval for IBM Slope:")
print(ci_simple)

print("\nMultiple Linear Regression - Confidence Interval for IBM Slope:")
print(ci_multiple)

print()
print("#2 (b)")

# Multiple Linear Regression of GE on IBM and Dow Jones
model_multiple = sm.OLS(df1['ge'], sm.add_constant(df1[['ibm', 'dowjones']]))
.fit()

# Hypothesis testing for dowjones coefficient
hyp_test_result = model_multiple.t_test('dowjones = 1')

# Display the results
print(hyp_test_result)

ans_3 = """
The null hypothesis means that a one-unit change in the dowjones is equal to 1.
The p value is 0.175
At a 5 percent significance level, the p value of 0.175 is greater than 0.05,
thus, we don't have
enough evidence to reject the null hypothesis.
At a 10 percent level, the p value is still greater than 0.1, thus we fail to
reject the null hypothesis
"""

print(ans_3)

```