Test Exercise 2 – GPA

(a)

```
# import packages for the analysis
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

import statsmodels.formula.api as smf
```

```
GPA = pd.read_table('TestExer2-GPA.txt')
```

```
# create a fitted model
lm = smf.ols(formula='FGPA ~ SATV', data=GPA).fit()
```

```
lm.summary()
```

OLS Regression Results

Dep. Variable:	FGPA	**R-squared:**	0.008
Model:	OLS	**Adj. R-squared:**	0.007
Method:	Least Squares	**F-statistic:**	5.201
Date:	Sun, 22 Sep 2019	**Prob (F-statistic):**	0.0229
Time:	10:57:44	**Log-Likelihood:**	-388.44
No. Observations:	609	**AIC:**	780.9
Df Residuals:	607	**BIC:**	789.7
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	2.4417	0.155	15.747	0.000	2.137	2.746
SATV	0.0631	0.028	2.280	0.023	0.009	0.117

(i) So for the model *FGPA* = a+b**SATV*,

b = 0.063, standard error of b is 0.028, P-value is 0.023

(ii) 95% CI of SATV's increasing by 1 on FGPA is [0.009, 0.117]

(b)

```
# create a fitted model
mlm = smf.ols(formula='FGPA ~ SATM+SATV+FEM', data=GPA).fit()

# print the coefficients
mlm.summary()
```

OLS Regression Results

Dep. Variable:	FGPA	**R-squared:**	0.083
Model:	OLS	**Adj. R-squared:**	0.078
Method:	Least Squares	**F-statistic:**	18.24
Date:	Sun, 22 Sep 2019	**Prob (F-statistic):**	2.41e-11
Time:	12:04:45	**Log-Likelihood:**	-364.67
No. Observations:	609	**AIC:**	737.3
Df Residuals:	605	**BIC:**	755.0
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	1.5570	0.216	7.205	0.000	1.133	1.981
SATM	0.1727	0.032	5.410	0.000	0.110	0.235
SATV	0.0142	0.028	0.507	0.612	-0.041	0.069
FEM	0.2003	0.037	5.358	0.000	0.127	0.274

(i) So for the model *FGPA* = a+b1**SATM*+b2**SATV*+b3**FEM* ,

b2 = 0.014, standard error of b2 is 0.028, P-value is 0.612

(ii) 95% CI of SATV's increasing by 1 on FGPA is [-0.041, 0.069]

(c) The correlation Matrix:

```
GPA.iloc[:,1:].corr()
```

	FGPA	SATM	SATV	FEM
FGPA	1.000000	0.195040	0.092167	0.176491
SATM	0.195040	1.000000	0.287801	-0.162680
SATV	0.092167	0.287801	1.000000	0.033577
FEM	0.176491	-0.162680	0.033577	1.000000

It shows that the correlation between SATV and FGPA is less than 10% and much smaller than the other two pairs. So when stronger predictors introduced into the model, the influence of SATV will become insignificant.

(d)

H_0: b2 = 0

H_1: b2 $\neq$ 0

Analyze the restricted model:

```
# create a restricted fitted model without SATV
rmlm = smf.ols(formula='FGPA ~ SATM+FEM', data=GPA).fit()

R0_2 = rmlm.rsquared
R1_2 = mlm.rsquared
F = (R1_2-R0_2)*605/(1-R1_2)
F
```

```
0.25715514284401925
```

We get F = 0.257155 < 3.9. So, the effect of SATV on FGPA is insignificant.

As we know from (b) that t^2 of b2 is 0.507105^2 = 0.257155

Clearly, F = t^2