ECON 0150 | Spring 2026 | Homework 5.4

Due: Sunday April 19, at 11:59PM

Homework is designed to both test your knowledge and challenge you to apply familiar concepts in new applications. Answer clearly and completely. You are welcomed and encouraged to work in groups so long as your work is your own. Use the provided datasets to answer the following questions. Then submit your figures and answers to Gradescope.

Q1. Causation and Model Comparison in the Recession and Health Study

Over the past several assignments, you have progressively built up a model of the relationship between county unemployment and BMI using 2011 BRFSS data. You started with a simple regression in HW 4.1, added a categorical control in HW 5.1, tested for interaction effects in HW 5.2, and added multiple numerical controls in HW 5.3. Now let's step back and evaluate the full arc.

a) Fill in the table below using your results from previous assignments for the coefficient on `unemployment_rate` changed across your models.

Model	**Specification**	$\hat{\beta}_1$ **(unemployment_rate)**
HW 4.1	BMI5 ~ unemployment_rate	
HW 5.1c	BMI5 ~ unemployment_rate + Female	
HW 5.3d	BMI5 ~ unemployment_rate + AGE	
HW 5.3e	BMI5 ~ unemployment_rate + Female + AGE + College + Married	

What happened to the coefficient as you added controls? What does this tell you about confounding in the unemployment–BMI relationship?

b) Compare the R^2 values for the simple model (HW 4.1) and the full model (HW 5.3e). You can find R^2 using `model.rsquared`. Report both values.

```
# Fit the simple model
simple_model = smf.ols('BMI5 ~ unemployment_rate', data=data).fit()
print("Simple model R²:", simple_model.rsquared)

# Fit the full model
full_model = smf.ols('BMI5 ~ unemployment_rate + Female + AGE + College + Married',
data=data).fit()
print("Full model R²:", full_model.rsquared)
```

c) Use `full_model.compare_f_test(simple_model)` to get the F-statistic and p-value. Report both. Does adding controls significantly improve the model?

```
# F-test comparing the full model to the simple model
f_stat, p_value, df_diff = full_model.compare_f_test(simple_model)
print("F-statistic:", f_stat)
print("p-value:", p_value)
```

d) Even with all these controls, can we say that county unemployment *causes* higher BMI? In 2–3 sentences, discuss at least one reason why the relationship might not be causal, even after controlling for gender, age, education, and marital status.