EE325 STATA session II HW 4 due April 29,2022
Dummy variables and Multicollinearity
In order to receive HW points from the following questions, students must submit your answer with log file.

1. **Table 9.1** Public school teachers' salaries (Dollars) by geographical region. Consider and estimate the following model with the dummy variable:

$$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + u_i$$

Where Y = (average) salary of public school teacher in state i

D_{2i} =1 *if the state is* in *the Northeast or North Central* = 0 *otherwise*

D_{3i} =1 *if the state is* in *the South* = 0 *otherwise*

t = time
D = 1 for observations in 1970-1981

= 0 otherwise

Find the mean salary of public school teacher in the South and the mean salary of public school teacher in the Northeast or North Central.

2. **Table 8.9** Savings and Personal Disposable income (billions of dollars), United States, 1970-1995.

a) Given the data in the table, estimate the following linear savings function using personal disposable income

Time period 1970-1981: $Y_t = \lambda_1 + \lambda_2 X_t + u_{1t}$ $n_1 = 12$

Time period 1982-1995: $Y_t = \gamma_1 + \gamma_2 X_t + u_{2t}$ $n_2 = 14$

Time period 1970-1995: $Y_t = \alpha_1 + \alpha_2 X_t + u_{3t}$ $n_1 + n_2 = 26$

b) On the basis of the Chow test that there was a difference in the regression of savings on income between the two periods. Consider and estimate the following model with the dummy variable:

$$Y_t = \alpha_1 + \alpha_2 D_t + \beta_1 X_t + \beta_2 (D_t X_t) + u_t$$

Where Y = Savings
X= Personal disposable income
t = time
D = 1 for observations in 1982-1995

= 0 otherwise

i. Test the coefficients individually statistically significant at the 5 percent level? From this test, how would you describe the difference in the two regressions (coincident regression, parallel regression, concurrent regression, dissimilar regression)?
ii. Write down the mean personal savings function for 1970-1981 and the mean personal savings function for 1982-2005.

3. Table 9.4 U.S. refrigerator sales over the sample period

$$Y_t = \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 D_{4t} + u_t \quad (1)$$

$$Y_t = \beta_1 + \beta_2 D_{2t} + \beta_3 D_{3t} + \beta_4 D_{4t} + u_t \quad (2)$$

Where Y = Sales of refrigerators (in thousands)
D = 1 for relevant quarter

= 0 otherwise

Interpret the meaning. Test the coefficients individually statistically significant at the 10 percent level?

4. Table 9.3 U.S. Dishwasher sales over the sample period

$$Y_t = \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 D_{4t} + u_t \quad (1)$$

$$Y_t = \beta_1 + \beta_2 D_{2t} + \beta_3 D_{3t} + \beta_4 D_{4t} + u_t \quad (2)$$

Where Y = Sales of dishwasher (in thousands)
D = 1 for relevant quarter

= 0 otherwise

Interpret the meaning. Test the coefficients individually statistically significant at the 10 percent level?

5. **Table 9.7** provides data on a sample of 114 workers in an industrial town in southern India in 1990. The variables are defined as follows:

   WI = weekly wage income in rupees

   Age = age in years

   D_{sex} = 1 for male workers and 0 for female workers

   DE_2 = a dummy variable taking a value of 1 for workers with an education level up to primary

   DE_3 = a dummy variable taking a value of 1 for workers up to a secondary level of education

   DE_4 = a dummy variable taking a value of 1 for workers with higher than secondary education

   DPT= a dummy variable taking a value of 1 for workers with permanent jobs and a value of 0 for temporary workers

   We estimate the following regression model:

$$lnWI_{\text{i}} = \beta_1 + \beta_2 AGE_{\text{i}} + \beta_3 D_{\text{sex}} + \beta_4 DE_2 + \beta_5 DE_3 + \beta_6 DE_4 + \beta_7 DPT + u_{\text{i}} \quad (1)$$

$$lnWI_{\text{i}} = \lambda_1 + \lambda_2 AGE_{\text{i}} + \lambda_3 D_{\text{sex}} + \lambda_4 DE_2 + \lambda_5 DE_3 + \lambda_6 DE_4 + \lambda_7 D_{\text{sex}} DE_2 + \lambda_8 D_{\text{sex}} DE_3 + \lambda_9 D_{\text{sex}} DE_4 + \lambda_{10} DPT + u_{\text{i}} \quad (2)$$

$$lnWI_{\text{i}} = \gamma_1 + \gamma_2 AGE_{\text{i}} + \gamma_3 D_{\text{sex}} + \gamma_4 D_{\text{sex}} DE_2 + \gamma_5 D_{\text{sex}} DE_3 + \gamma_6 D_{\text{sex}} DE_4 + \gamma_7 DPT + u_{\text{i}} \quad (3)$$

6. **Table 10.5 Consumption expenditure in relation to income and wealth.**

a) Regress consumption expenditure on income and wealth and interpret your regression.
b) Is there evidence of multicollinearity in the data? How do you know?
c) Compute the variance inflation factor and explain.
d) If there is the multicollinearity problem, what remedial action, if any, would you take?

7. **Table 10.7 Consumption Function for United States, 1947-2000.**

a) Regress consumption expenditure on real disposable personal income and real wealth and real interest, interpret your regression.
b) Is there evidence of multicollinearity in the data? How do you know?
c) Compute the correlations and variance inflation factor and explain.
d) If there is the multicollinearity problem, what remedial action, if any, would you take?