Name:_______________________________

ECON 453
In-Class Exercise 4
September 26, 2023

ANSWER KEY

Please download the file "IC4 Session.gretl", a gretl Session file. This dataset comes from the 2019 World Bank Development Indicators and contains information on variables measured at the country level. This is similar to a dataset we have worked with before but contains only the lower income countries (countries below $15,000 in GDP per capita). The dataset contains basic descriptions of each of the variables.

While working on this exercise, please practice making a model table and using this to compare results across models.

1. Regress **Life Expectancy** on GDP per capita (**GDP_1000s**) Save this model to your session as an icon. Add this model to the model table (this happens in the icon view – right click on Model 1 and choose **Add to Model Table**)

2. Run the same model but add dummy variables for continents. Use Oceania as your reference category. Save this to your session as an icon, then add to your model table.
   a. Summarize briefly what we learn from the continent dummies.

Here is what my model table looks like with the first two regressions:

```
OLS estimates
Dependent variable: Life_Expectancy

                          (1)          (2)

const                 61.3250***   65.1918***
                      (0.8274)     (1.4446)

GDP_1000s              0.9498***    0.6778***
                      (0.1063)     (0.1196)

Asia                                0.5290
                                   (1.5715)

Europe                             -0.4169
                                   (2.7784)

Africa                             -4.6790***
                                   (1.4260)

North_America                       0.9759
                                   (1.9287)

South_America                      -0.7310
                                   (2.3263)

             n             93           93
     Adj. R**2         0.4613       0.5798
           lnL         -268.9       -254.7

Standard errors in parentheses
*   significant at the 10 percent level
**  significant at the 5 percent level
```

From the continent dummies we learn that life expectancy in African countries is significantly lower (about 4.7 years) than in the reference category (Oceania). We also learn that none of the other continents differ significantly from Oceania. Overall, it seems we are learning that African countries have lower life expectancy, and there appear to be no other major differences by continent.

   b. Comment on how the coefficient on GDP per capita changed when you added the continent dummies.

The magnitude of the GDP coefficient decreased pretty substantially between the first and second regressions. In the first (simpler) model, we estimate that a $1,000 increase in GDP per capita will increase life expectancy in a country by about 0.95 years. When we control for the continents, the impact of GDP is down to about 0.68 years. This tells us that we were likely overstating the impact of GDP per capita on health outcomes by not accounting for regional differences.

c. Test whether the model indicates life expectancy is significantly different for low-income countries in Europe and North America. In other words, is the coefficient on Europe significantly different from the coefficient on North America? From your regression results window, choose **Tests -> Linear restrictions.** Write out the null hypothesis, the p-value from the test, and the conclusion.

Here is how I entered the linear restriction:

gretl: linear restrictions

Specify restrictions:
(Please refer to Help for guidance)
right-click for some shortcuts

```
b[Europe]-b[North_America]=0
```

And here are my results:

```
Restriction:
 b[Europe] - b[North_America] = 0

Test statistic: F(1, 86) = 0.265083, with p-value = 0.60797

Restricted estimates:

                  coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------------
  const            65.2662       1.43131      45.60     1.60e-062 ***
  GDP_1000s         0.663771     0.116006      5.722    1.47e-07  ***
  Asia              0.570999     1.56270       0.3654   0.7157
  Europe            0.660125     1.82084       0.3625   0.7178
  Africa           -4.69311      1.41972      -3.306    0.0014    ***
  North_America     0.660125     1.82084       0.3625   0.7178
  South_America    -0.633141     2.30871      -0.2742   0.7845

  Standard error of the regression = 3.87477
```

The test here is whether there is a significant difference between the coefficients on Europe and North America. Formally, the null would be that the coefficients are equivalent, so we would write as:

H_0: $\beta_{Europe} = \beta_{North\ America}$

The test is looking to see if we have enough evidence to reject that null. A low p-value tells us the null is unlikely given the data. Our p-value from the test is 0.61, so we will fail to reject the null. Overall, we conclude that there are no differences across these regions in terms of life expectancy.

3. Run a regression with life expectancy as the dependent variable and two regressors: GDP_1000s and the Africa dummy variable. Save these results to the session as an icon and add it to your model table.
    a. Compare the results from this model to the ones from the previous model in question 2. Which should be our preferred model?

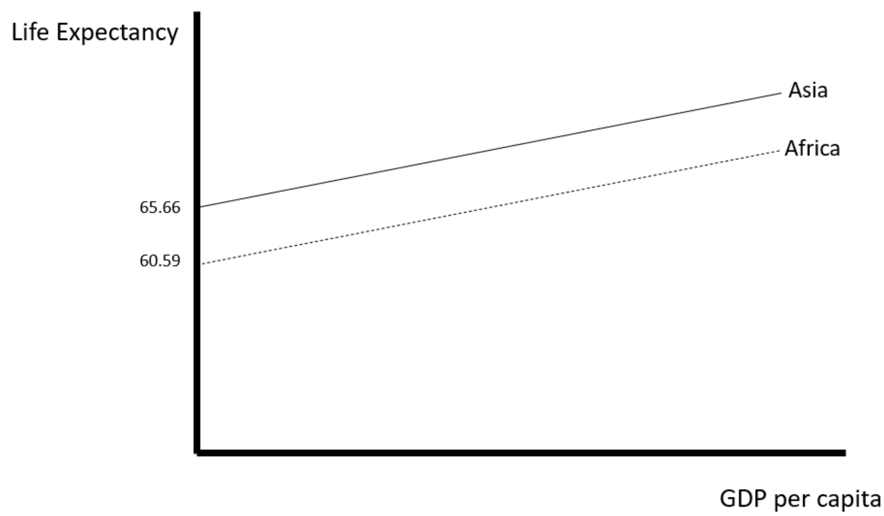
My updated model table:

	(1)	(2)	(3)
const	61.3250***	65.1918***	65.6573***
	(0.8274)	(1.4446)	(1.0587)
GDP_1000s	0.9498***	0.6778***	0.6600***
	(0.1063)	(0.1196)	(0.1059)
Asia		0.5290	
		(1.5715)	
Europe		-0.4169	
		(2.7784)	
Africa		-4.6790***	-5.0680***
		(1.4260)	(0.9106)
North_America		0.9759	
		(1.9287)	
South_America		-0.7310	
		(2.3263)	
n	93	93	93
Adj. R**2	0.4613	0.5798	0.5948
lnL	-268.9	-254.7	-255.1

We should prefer the third model, the one with just the African dummy variable. This is because the adjusted-R^2 value for this model is higher than the model with all of the continent dummies included. The other regional dummies were insignificant, so we can make a more efficient model by dropping them.

b. Sketch a simple graph that shows the relationship between GDP per capita and life expectancy in Africa and GDP and life expectancy in Asia (on the same graph).

Our equation can be written as: y = 65.66 + 0.66*GDP – 5.07*Africa. This means the prediction equation for countries not in Africa (for example, in Asia) will be y = 65.66 – 0.66*GDP (since the African dummy will be 0). For countries in Africa, we could write the equation as: y = 65.66 – 0.66*GDP – (5.07*1) = 60.59 – 0.66*GDP. This means we could make a graph that looks like:



The dummy variable allows the intercept to be different, but not the slope of the relationship between GDP per capita and life expectancy. If we think that differs by continent we will need to either: (1) run separate regressions by continent, (2) use interaction terms, or (3) give up on econometrics (and our dreams) and settle on being an accountant.

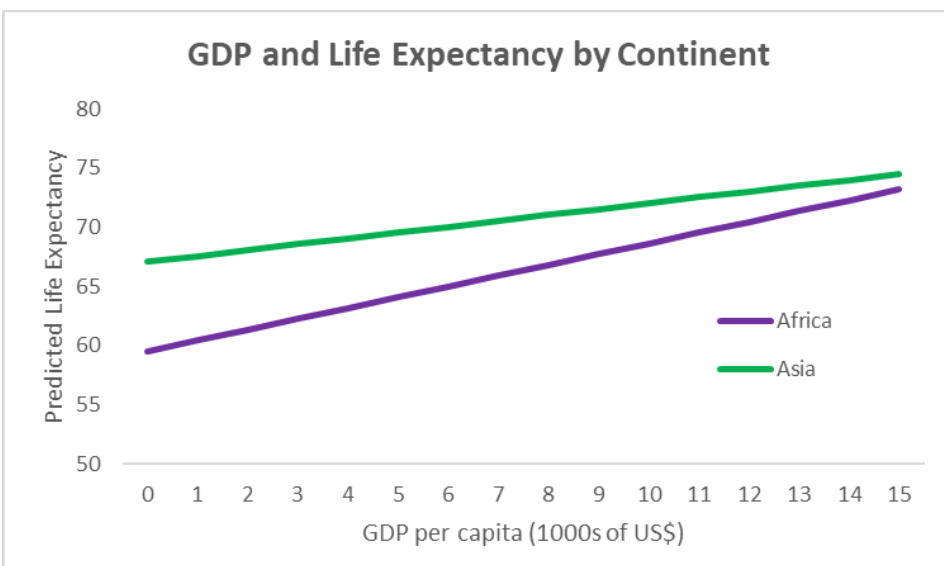
4. Restrict the sample to just the countries in Africa. Run the regression with life expectancy as the dependent variable and GDP per capita as the regressor. Save the results to your session as an icon and add it to your model table. Next, restrict the sample to just the countries **not** in Africa. Run the same regression, save these results to the session and add to your model table.
   a. Summarize what we learn about the relationship between economic performance and life expectancy in Africa and whether the relationship differs on other continents. Sketch a simple graph that shows the relationship between GDP per capita and Life Expectancy in Africa and the relationship in Asia (on the same graph).

My updated model table, with the regressions for this question in the last two columns:

OLS estimates
Dependent variable: Life_Expectancy

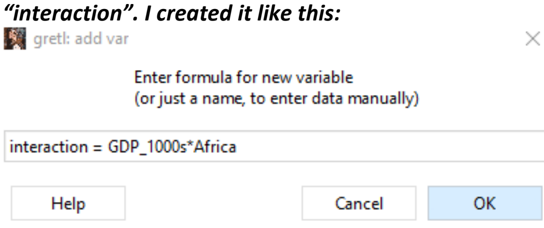
	Simple	Model 2	Model 3	Africa	Not Africa
const	61.3250***	65.1918***	65.6573***	59.5021***	67.0552***
	(0.8274)	(1.4446)	(1.0587)	(1.0944)	(0.9647)
GDP_1000s	0.9498***	0.6778***	0.6600***	0.9125***	0.4962***
	(0.1063)	(0.1196)	(0.1059)	(0.2001)	(0.1021)
Asia		0.5290			
		(1.5715)			
Europe		-0.4169			
		(2.7784)			
Africa		-4.6790***	-5.0680***		
		(1.4260)	(0.9106)		
North_America		0.9759			
		(1.9287)			
South_America		-0.7310			
		(2.3263)			
n	93	93	93	45	48
Adj. R**2	0.4613	0.5798	0.5948	0.3104	0.3248
lnL	-268.9	-254.7	-255.1	-130.8	-117.7

From this, we can see that the relationship between GDP per capita and life expectancy has a much larger magnitude for African countries. A $1,000 increase in GDP per capita will increase life expectancy by 0.91 years in Africa, but on other continents the same increase is only estimated to increase life expectancy by about 0.5 years. Now the slope and intercept will differ, and our graph looks like:



5. Restore the data to the full sample. Create an interaction term between GDP per capita (GDP_1000s) and the Africa dummy variable. Go to **Add -> Define new variable.** An interaction term is the two variables multiplied by each other. Run a regression with life expectancy as the dependent variable and the regressors of GDP_1000s, Africa, and the interaction term. Save these results to the session and add to your model table.
    a. What is the estimated impact of a $1000 increase in GDP per capita on life expectancy in Africa? In Asia? How does this compare to your estimates in question 4?

To create an interaction term, you can just multiply the variables together. In this case, my interaction will be the GDP variable multiplied by the African dummy. I am not very creative, so I just called my new variable "interaction". I created it like this:



And here are my results:

```
Model 8: OLS, using observations 1-93
Dependent variable: Life_Expectancy

                 coefficient   std. error   t-ratio    p-value
  -------------------------------------------------------------
  const           67.0552       1.26530      53.00     4.17e-069 ***
  GDP_1000s        0.496167     0.133929      3.705    0.0004    ***
  Africa          -7.55314      1.55828      -4.847    5.27e-06  ***
  interaction      0.416334     0.213500      1.950    0.0543    *

Mean dependent var   67.48677   S.D. dependent var   6.003111
Sum squared resid    1260.444   S.E. of regression   3.763282
R-squared            0.619826   Adjusted R-squared   0.607011
F(3, 89)             48.36768   P-value(F)           1.23e-18
Log-likelihood      -253.1691   Akaike criterion     514.3382
Schwarz criterion    524.4686   Hannan-Quinn         518.4286
```

What we get from these results is the equation: y = 67.0552 + 0.4962*GDP – 7.5531*Africa + 0.4163*(Africa*GDP). This means we can break this into two equations, for non-African countries (Africa=0) and African countries (Africa=1).

For non-African countries:

y = 67.0552 + 0.4962*GDP – 7.5531*0 + 0.4163*(0*GDP)
y = 67.0552 + 0.4962*GDP

For African countries:

y = 67.0552 + 0.4962*GDP – 7.5531*1 + 0.4163*(1*GDP)
y = 67.0552 + 0.4962*GDP – 7.5531 + 0.4163*GDP
y = 59.5021 + 0.9125*GDP

WHAT! These are the same equations we got when we ran them separately. HOW COOL IS THAT?! Using an interaction model allows us to test whether the difference in estimated sloped between continents is significant. That is tested by looking at the significance of the interaction term. Here we have mild significance, with a p-value slightly over 0.05.

6. Run a regression that uses life expectancy as the dependent variable. Your regressors will be GDP_1000s, Immunized_DPT, Africa, and the interaction of the Africa and immunization variables.
    a. Summarize what we learn about the impact of improving immunization rates in Africa and whether this differs from the estimated impact of immunizations on other continents.

```
Q6: OLS, using observations 1-93
Dependent variable: Life_Expectancy

                  coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------------
  const            60.5098      3.39720      17.81     6.27e-031 ***
  GDP_1000s         0.522571    0.0960088     5.443    4.67e-07  ***
  Immunized_DPT     0.0717355   0.0391719     1.831    0.0704    *
  Africa          -13.7596      4.31329      -3.190    0.0020    ***
  inter2            0.105258    0.0498423     2.112    0.0375    **

Mean dependent var   67.48677   S.D. dependent var   6.003111
Sum squared resid    964.3257   S.E. of regression   3.310324
R-squared            0.709141   Adjusted R-squared   0.695920
F(4, 88)             53.63792   P-value(F)           8.13e-23
Log-likelihood      -240.7169   Akaike criterion     491.4337
Schwarz criterion    504.0967   Hannan-Quinn         496.5467
```

My results indicate that there is a positive and mildly significant relationship between the immunization rate and life expectancy in non-African countries. The estimate says that increasing the immunization rate by 1 percentage point will increase the life expectancy by 0.072 years. The interaction term tells us the impact is much larger in magnitude for African countries. In this case, the impact of immunizations is 0.072+0.105 = 0.177. Each increase in immunization rates by 1 percentage point will increase life expectancy by about 0.177 years, all else equal, in African countries. The difference in impact by continent is statistically significant (based on the p-value of the interaction term).