

QUANTITATIVE QUESTIONS

QUESTION 1: The Long-Term Impact of the Slave Trade [30 Points]

This part uses data from the following paper,

Nathan Nunn (2008). “The Long-Term Effects of Africa’s Slave Trades.” *Quarterly Journal of Economics* 123 (1): 139-176.

To help answer this question, first read the paper. Part of your task is to replicate and extend some of Nunn’s results, which he produces using instrumental variables. If you are unable to exactly reproduce Nunn’s results, report your best effort to do so. Whether or not you can exactly replicate the paper’s findings, **ensure that both your write-up and your R script clearly indicate how you obtained your results**. The dataset for this question is `nunn.Rda`. It contains these variables:

Variable name	Variable description
<i>country</i>	Country name
<i>ln_realgdp2000</i>	Log real per capita GDP in 2000, also called “ln y” in the paper
<i>ln_export_area</i>	Log total number of slaves exported, divided by land area
<i>atlantic_dist</i>	Sailing distance to nearest destination of Atlantic slave trade
<i>indian_dist</i>	Sailing distance to nearest destination of Indian slave trade
<i>saharan_dist</i>	Overland distance to nearest port of export for Saharan slave trade
<i>redsea_dist</i>	Overland distance to nearest port of export for Red Sea slave trade
<i>colonial_power</i>	Name of colonizer, if any, prior to independence
<i>equator_dist</i>	Distance from equator
<i>longitude</i>	Longitude
<i>rain_min</i>	Minimum monthly rainfall
<i>humid_max</i>	Average maximum humidity
<i>low_temp</i>	Average minimum temperature
<i>ln_coastline_area</i>	Log coastline divided by land area
<i>low_distance</i>	=1 if situated at a low distance from a major slave destination
<i>high_slavery</i>	=1 if country had a high level of slave exports

Note that *low_distance* and *high_slavery* do not feature in Nunn’s paper: they have been created for this question. The other variables are identical to those used in the paper.

Answer the following questions:

- a) **[7 points]** Begin by examining the relevance of Nunn’s instruments:
 - i. Replicate the first stage results from the first column of Table IV on p.162. Report how you estimated them, and your results
Note: You only need to produce the four coefficients and four standard errors
 - ii. Conduct a statistical test to replicate the reported F statistic of 4.55 for these first-stage results, shown in Table IV
 - iii. Using the results from (i) and (ii), how relevant do you think that the instruments are? Explain your answer
- b) **[5 points]** Do you think that the instruments can be considered to be as-if randomly assigned? Explain your answer

- c) **[6 points]** Replicate the second-stage coefficients and standard errors for $\ln(\text{exports/area})$ in columns (1), (2) and (3) of Table IV on p.162. Report how you estimated them, and your results. **Note:** You only need to produce the three coefficients and three standard errors.
- d) **[2 points]** Why do you think that Nunn estimated the additional models in columns (2) and (3) of Table IV that include colonizer fixed effects and geographic controls?
- e) **[10 points]** Now, you will re-do Nunn's analysis of the impact of slavery on GDP with the Wald Estimator and binary variables. Use the single binary instrumental variable *low_distance*, the binary treatment variable *high_slavery*, and the outcome variable *ln_realgdp2000*.
- Explain, in this case, what type of country is a complier and what type of country is an always-taker
 - Calculate and report the proportion of compliers and the intent-to-treat effect
 - Use your answers from (ii) to calculate and report the Complier Average Causal Effect (CACE) of high slavery on GDP.
 - Using an appropriate method, calculate and report the p-value for this CACE estimate.
 - What do these results suggest about the relationship between slave exports and GDP?

QUESTION 2: An Experiment with Missing Data [20 Points]

This question analyses a simulated experimental dataset contained in the file "experiment_essay.Rda." It includes 100 units and the following five variables:

Variable name	Variable description
<i>y1</i>	Potential outcome under treatment
<i>y0</i>	Potential outcome under control
<i>r</i>	Missingness, under both treatment and control (=1 if missing, 0 otherwise)
<i>x</i>	A covariate (=0 or 1)
<i>d</i>	Treatment assignment (=1 if in treatment group, 0 if in control group)

Answer the following questions:

- [4 points]** Using two t-tests, assess whether units in the experiment can be considered to be missing at random. Explain your answer.
- [4 points]** Calculate the true average treatment effect for all units regardless of missingness, and the average treatment effect for always-reporters. Do these two treatment effects differ?
- [5 points]** Using appropriate tests, assess whether the units in this experiment can be considered to be missing independently of potential outcomes given *x*. Explain your answer
- [7 points]** Using only non-missing data, calculate a weighted average treatment effect that re-weights missing units appropriately, given their value of *x*. How does it compare to the two average treatment effects that you calculated in (b)?
Hint: You might find the `weighted.mean()` function helpful