**Reproducibility report**

**Date of submission : October 21, 2025**

***Poll: Were you able to reproduce the work based only on the program (not the code)?***

Yes, entirely. However, I occasionally referred to the code since I’m not yet fully familiar with Stata.

***Questions for discussion:***

1. **Where did it go off the rails? Was there one error that compounded everything?**

The steps were very clearly described, and everything went smoothly.

1. **What pieces of the program left you guessing as to what to do?**

The guidelines provided in the README file were clear and easy to follow.

1. **Any comments or reflections about the reproducibility exercise? For instance, has it changed your mind about the benefits (or costs) of open science? For those in fields where replication materials are not readily available, could an emphasis on reproducibility disadvantage researchers who use qualitative methods or restricted (e.g., administrative) data?**

This exercise showed me how crucial reproducibility is and that it should be planned from the very beginning of an analysis. It made me think more carefully about how to organize my own workflow for my thesis. It also motivated me to read *The Workflow of Data Analysis Using Stata b*y Scott Long.

**Summary:**  
The project was fully reproducible without issues.

**Data sources**  
The dataset itself was not provided, but the **DATA\_ACCESS** file contains a detailed step-by-step guide explaining the access conditions and how to locate the codebook and additional information about the data.

**Analysis data files**  
No analysis data files were explicitly mentioned.

**Code description**

There is one Stata .do file included and one Stata .log file included.

It produces the following outputs:

**Descriptive Tables:**

* . tab dpgrsum, m → *not included* in the table file
* . tab newrace, m → *included* in the table file
* . proportion newrace [pw=weight] → *included* in the table file
* . mean totinc [pw=weight], over(newrace) → *not included* in the table file
* . svy: mean totinc, over(newrace) → *included* in the table file

**Linear Regression Table:**

* . reg lntotinc i.newrace i.sex i.agegrp i.ssgrad i.pr [pw=weight], vce(robust)  
  (sum of weights = 26,070,295.905708) → *included* in the table file

**Histogram:**

* . histogram lntotinc, normal → *included* in the plot file

**Predictive Margins:**

* . marginsplot → *included* in the plot file
* . marginsplot, xdimension(newrace) by(sex) noci → *included* in the plot file

**Stated requirements**

No requirements specified except installation of Stata 17+ (written in the README).

**Missing requirements**

No missing requirements.

**Computing environment of the replicator**

Device name LAPTOP-02ST1DTD

Processor Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz (1.99 GHz)

Installed RAM 8.00 GB (7.86 GB usable)

Device ID 8BEBC651-30DC-4506-899D-9A720C1B2FDA

Product ID 00330-80000-00000-AA959

System type 64-bit operating system, x64-based processor

Edition Windows 11 Pro

Version 25H2

Installed on ‎2025-‎01-‎24

OS build 26200.6725

Experience Windows Feature Experience Pack 1000.26100.253.0 A close-up of a computer

AI-generated content may be incorrect.

**Replication steps:**

1. *Obtain the data :* OK
2. *Prepare the environment :* OK
3. *Prepare for Analysis :* OK
   1. Renaming all variables to lowercase: completed (variables were already lowercase).
   2. Retaining only the subset of variables needed downstream: unclear.
   3. All other steps: completed successfully.
4. *Analysis steps* 
   1. All steps were clearly explained and executed successfully

**Findings**  
The reproduced tables and graphs closely matched the original. Overall, the project was highly reproducible.