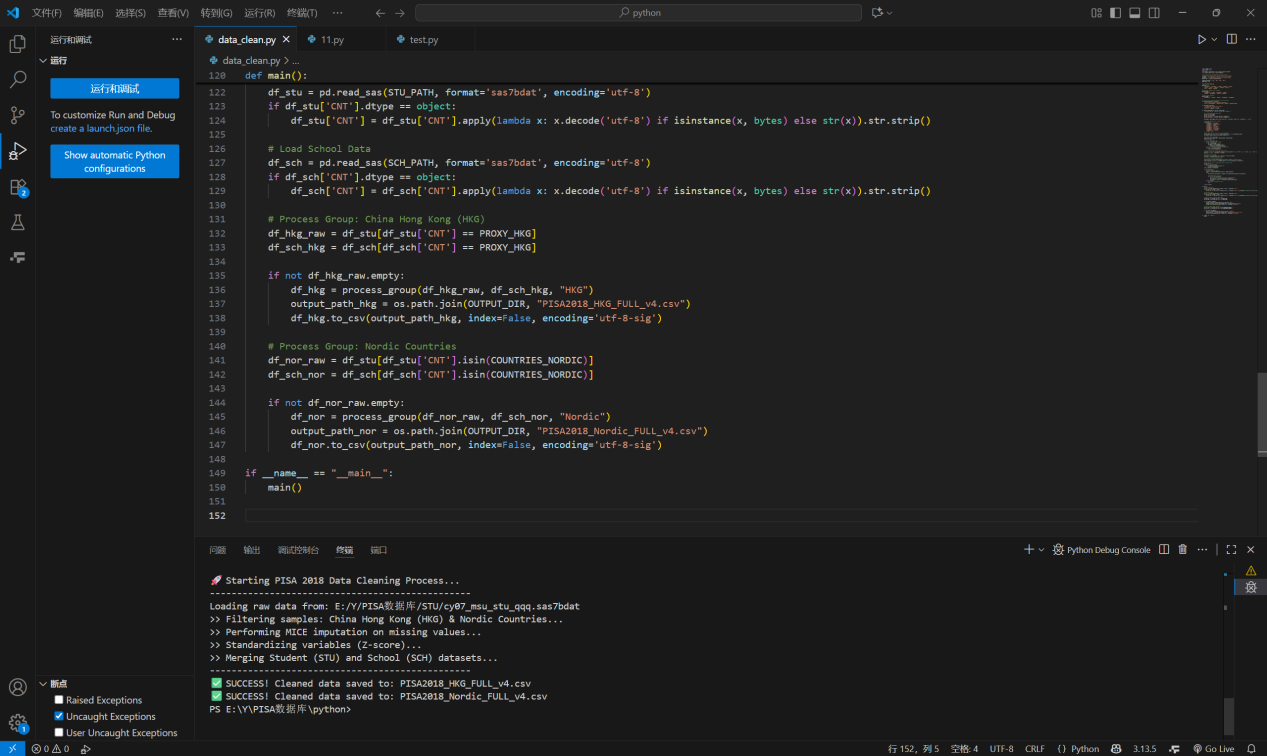
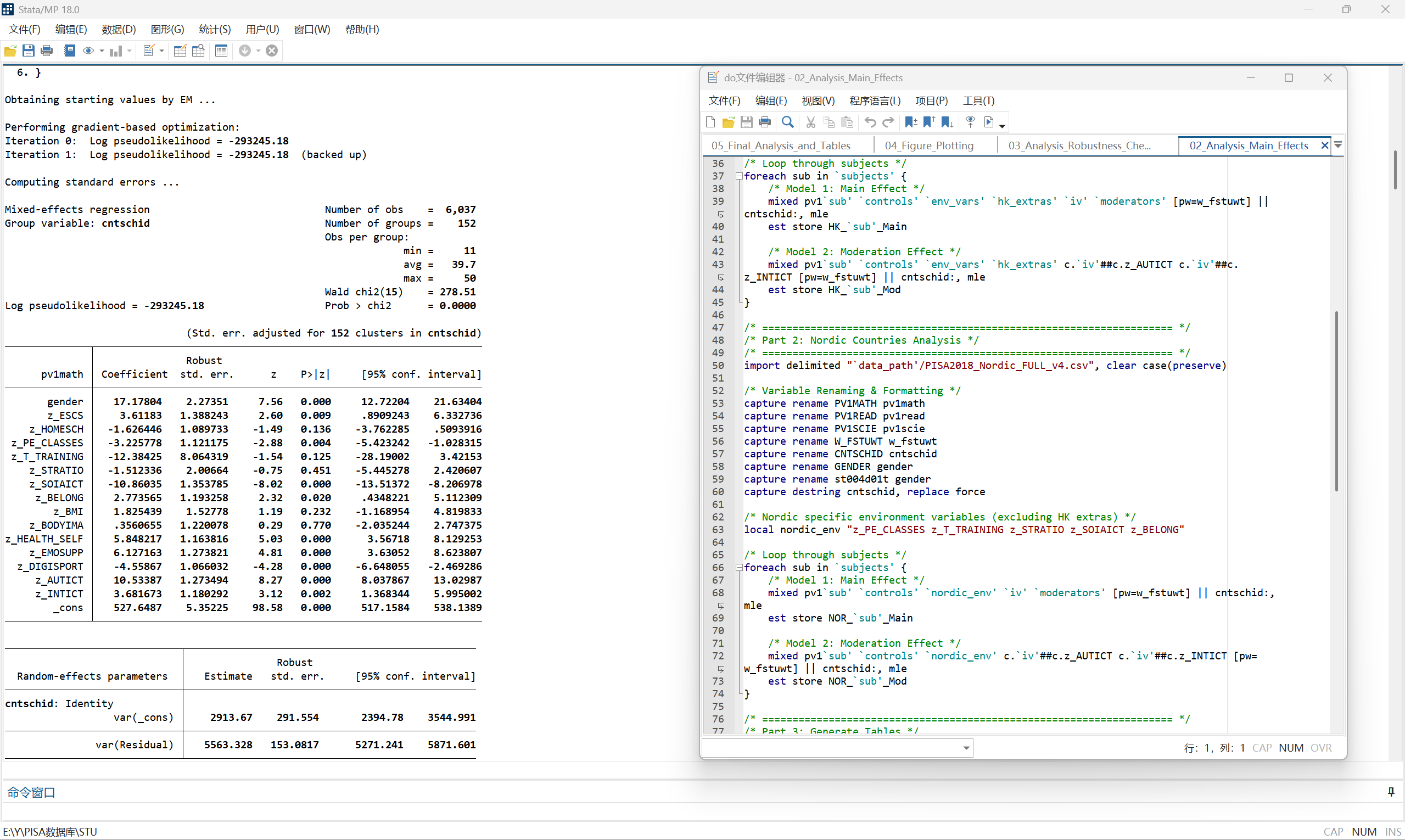
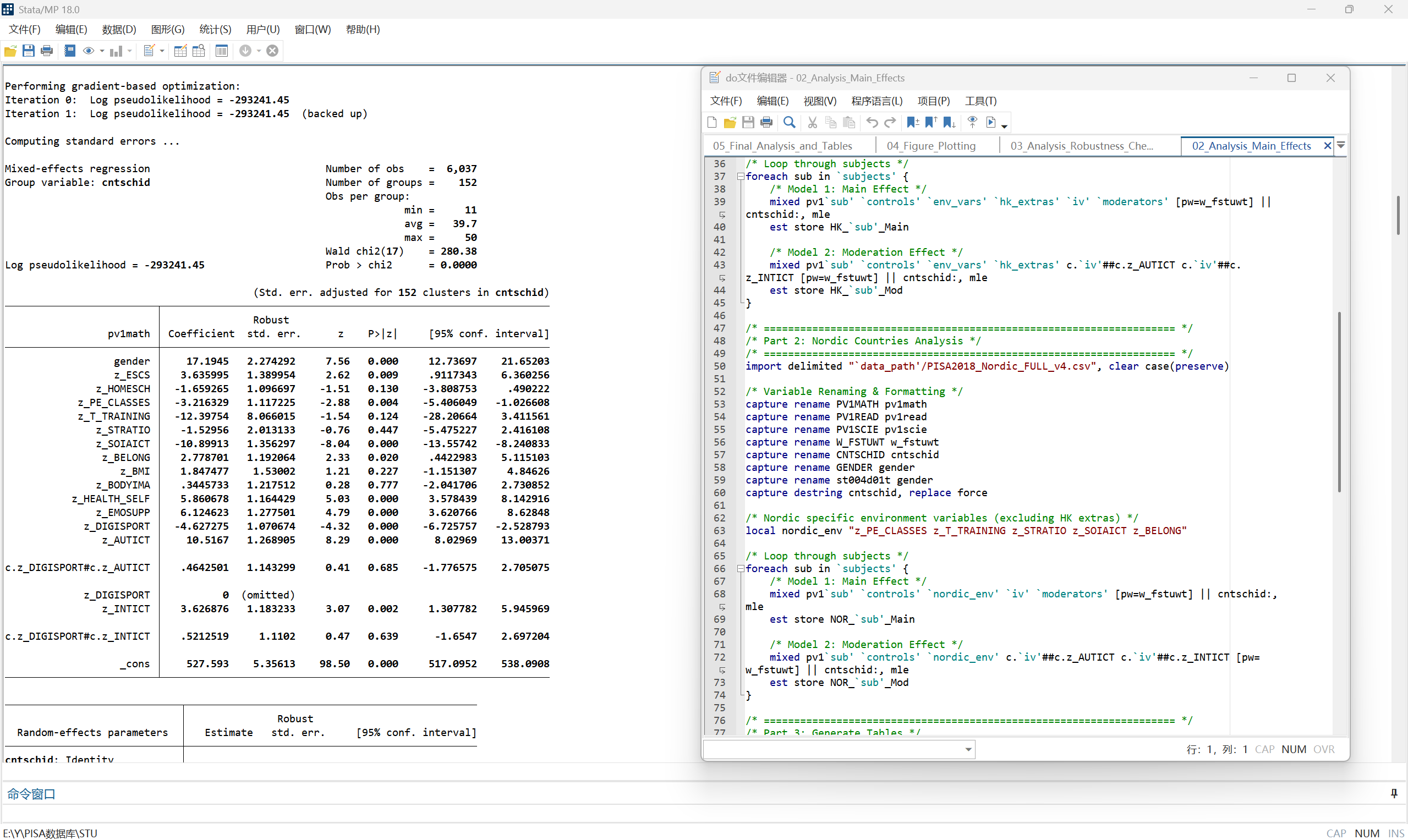
This technical appendix documents the complete data analysis workflow for the study "Digital Sports Participation and Academic Performance." It includes raw software output screenshots demonstrating the data cleaning process using Python and the Hierarchical Linear Modeling (HLM) analysis performed in Stata 17.0. These execution logs provide verifying evidence of the operational steps, model specifications, and statistical results reported in the main manuscript, ensuring the transparency and reproducibility of the findings.

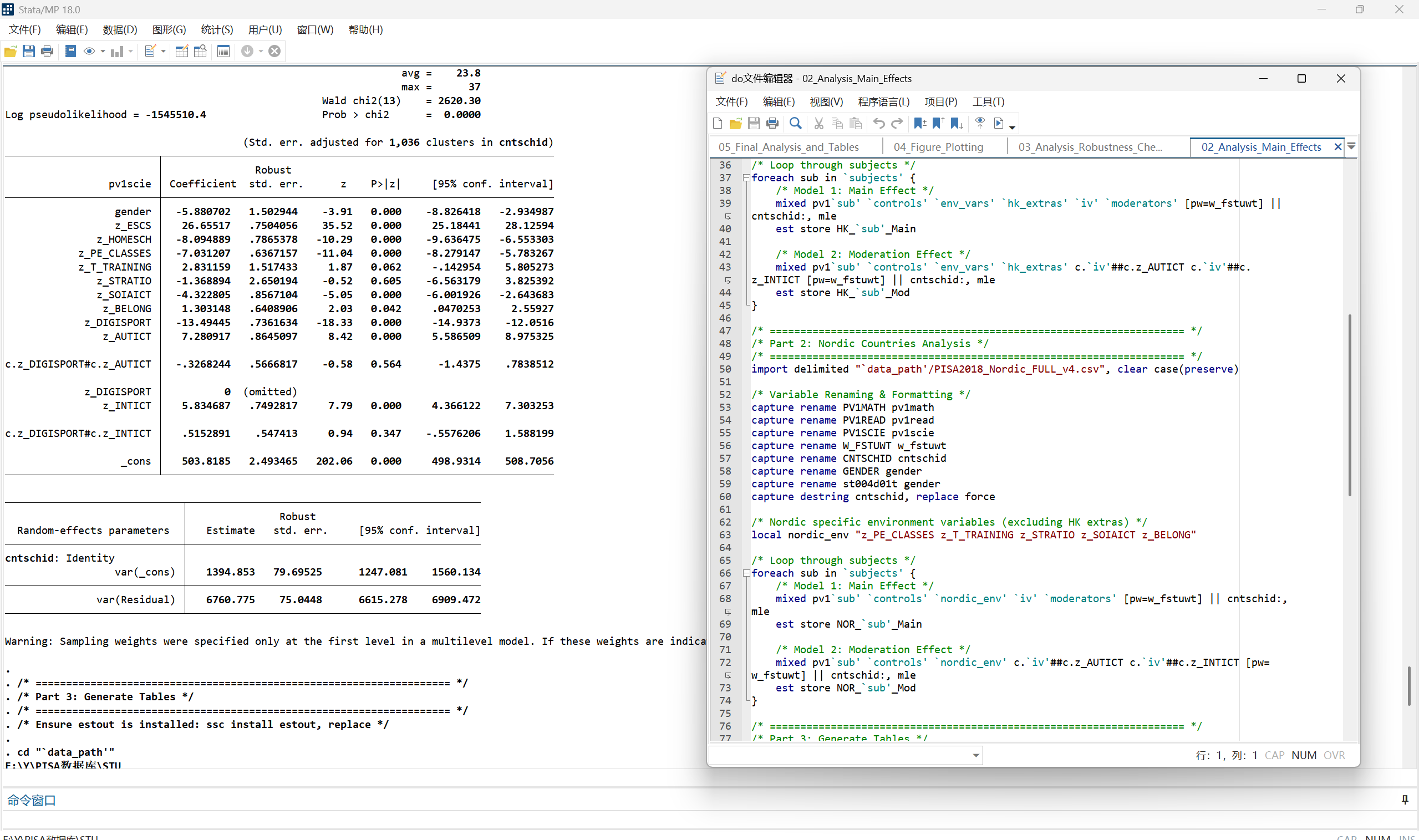
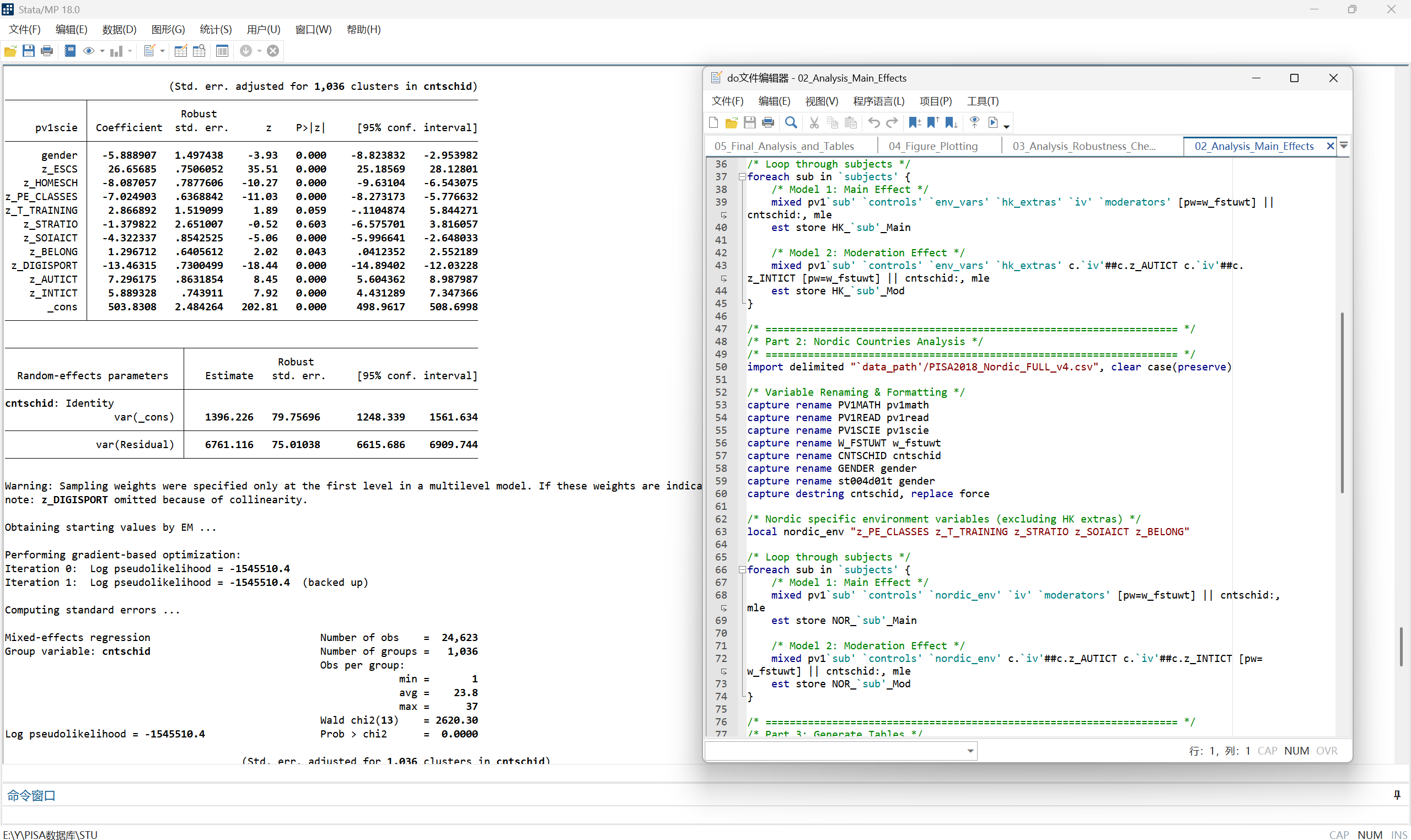
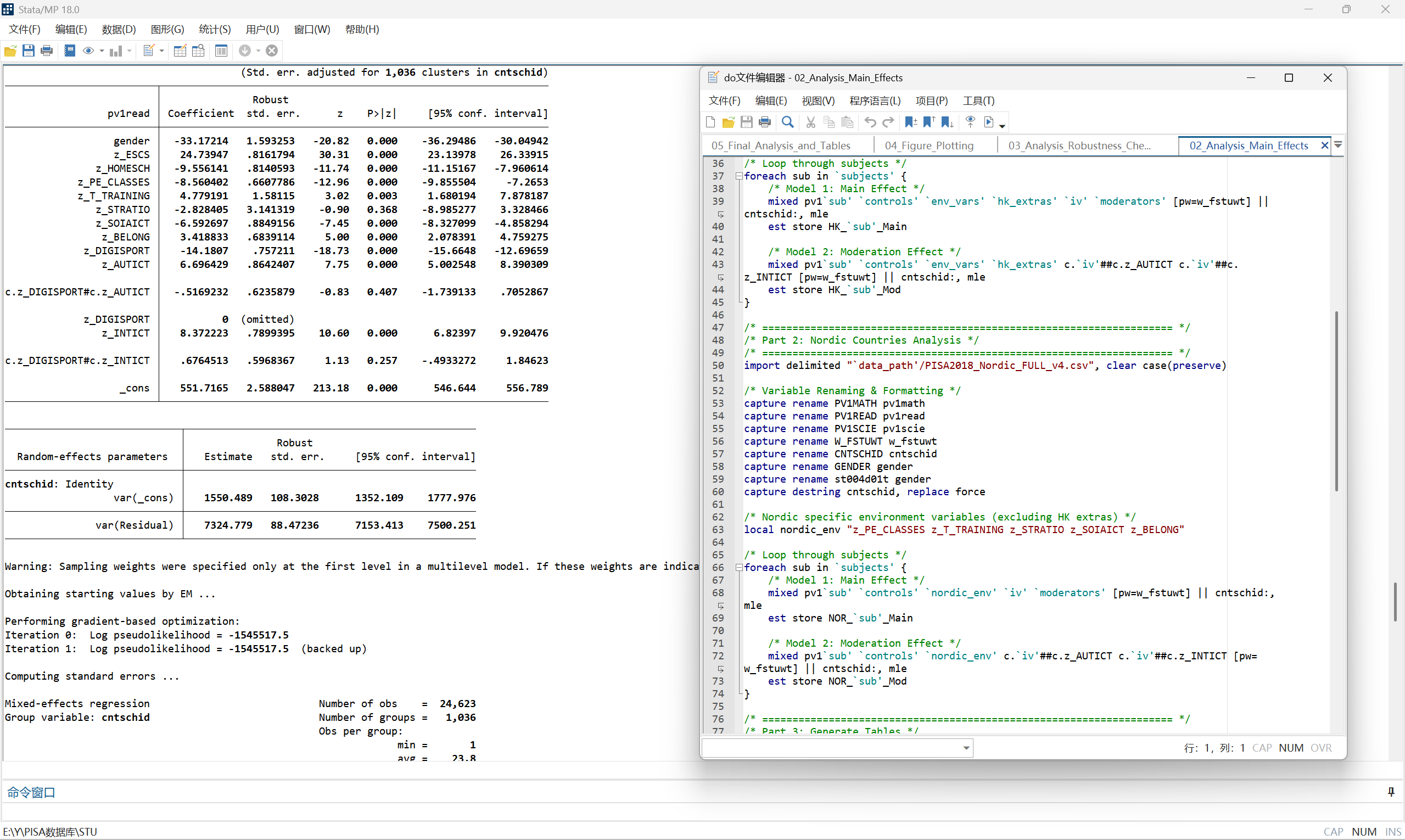
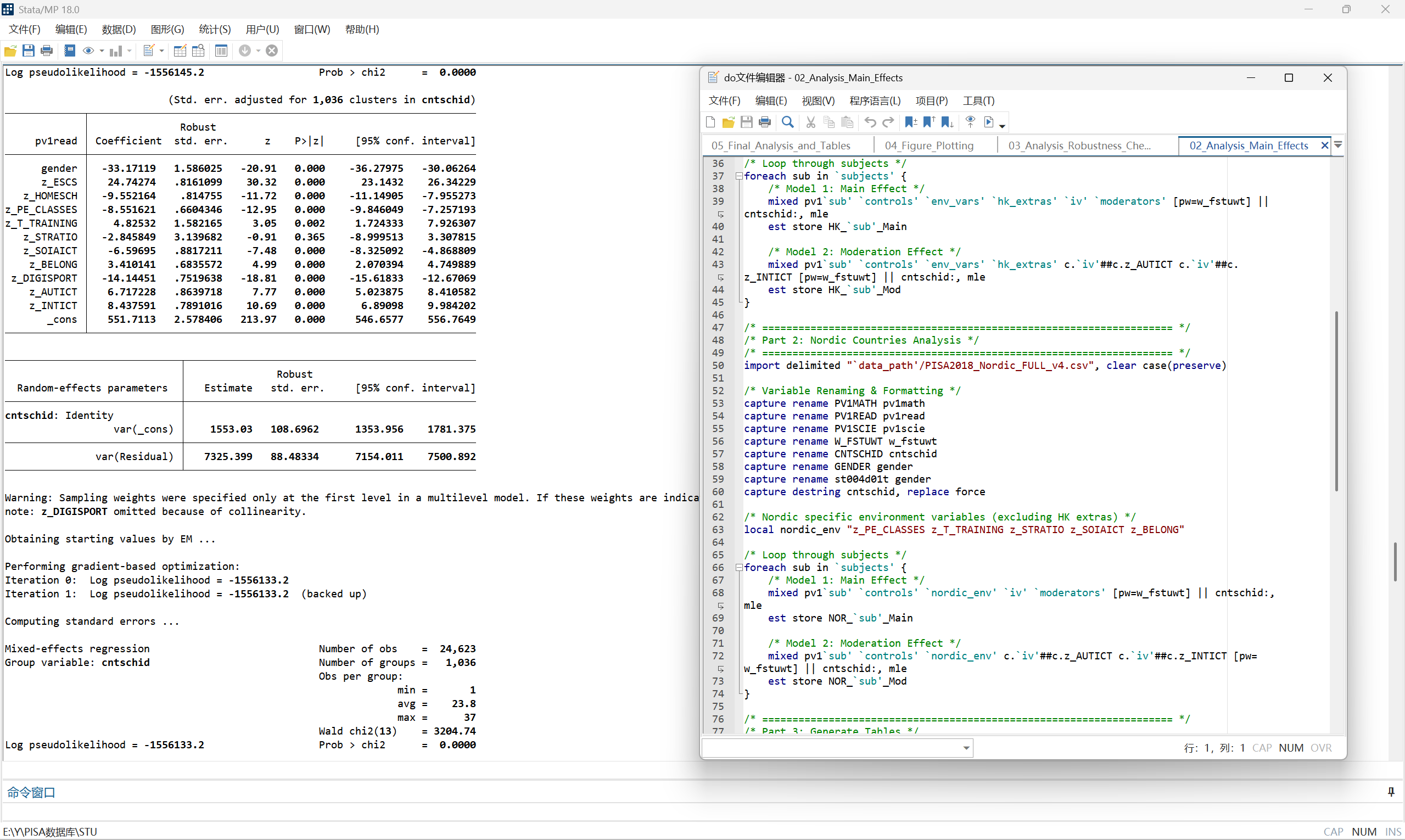
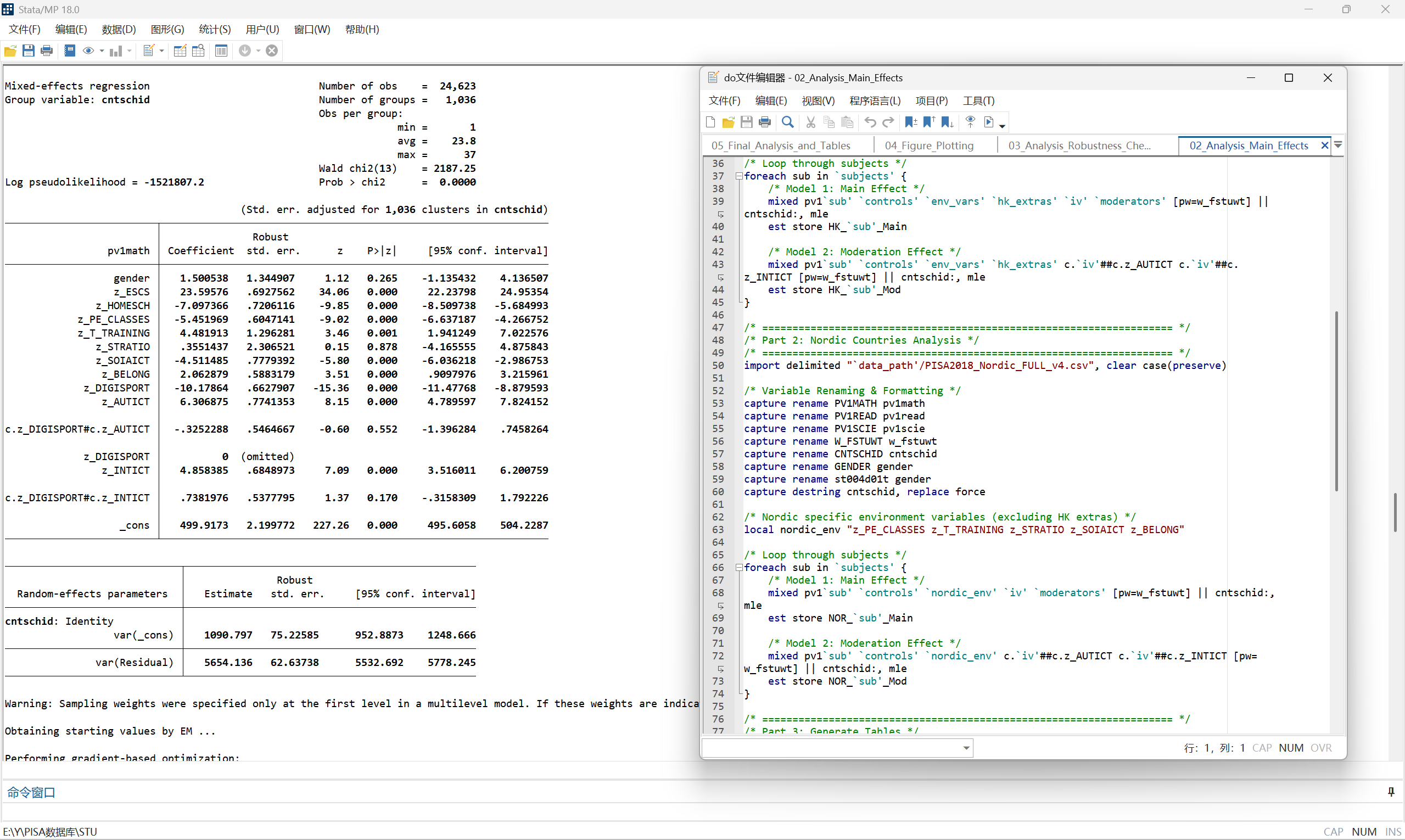
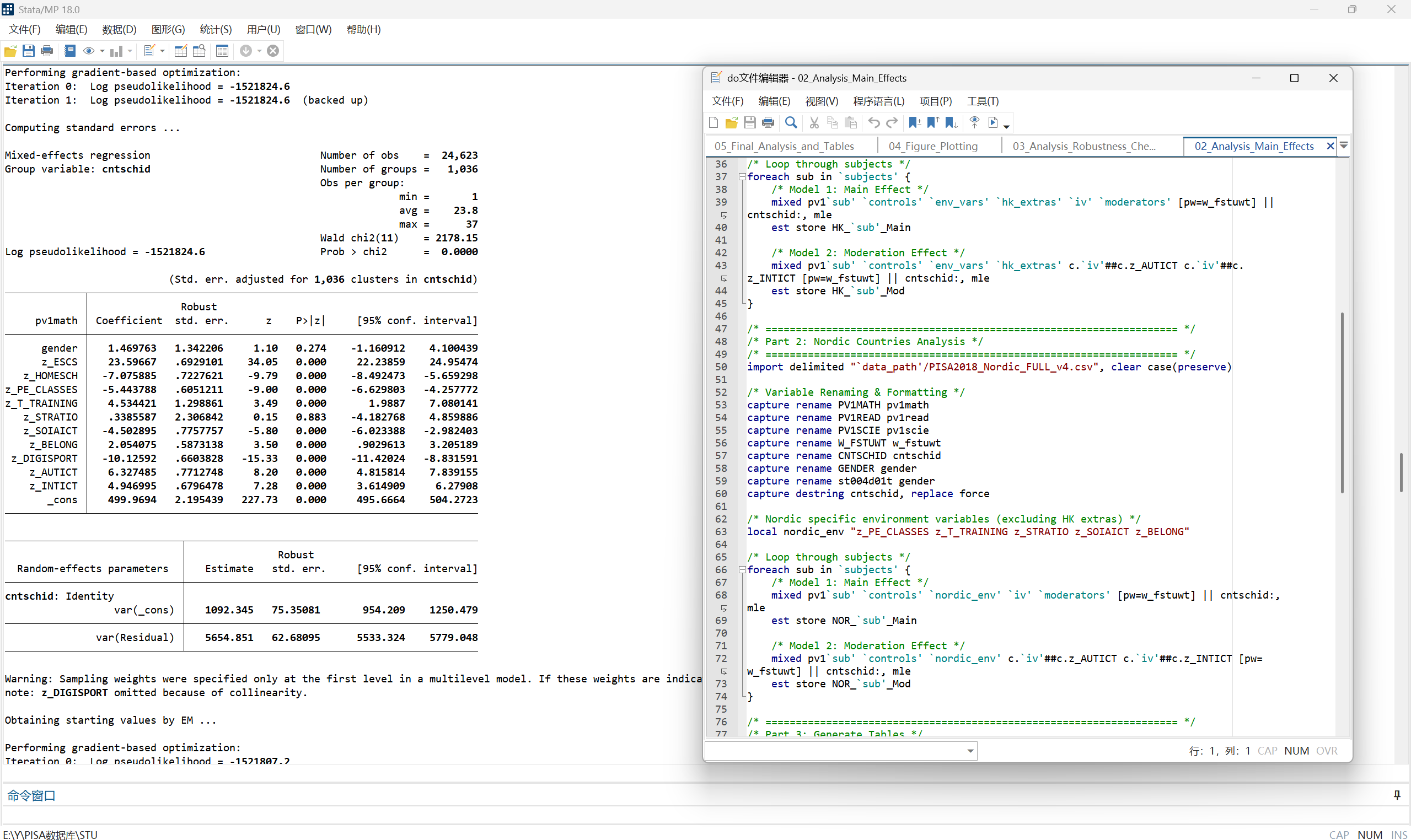
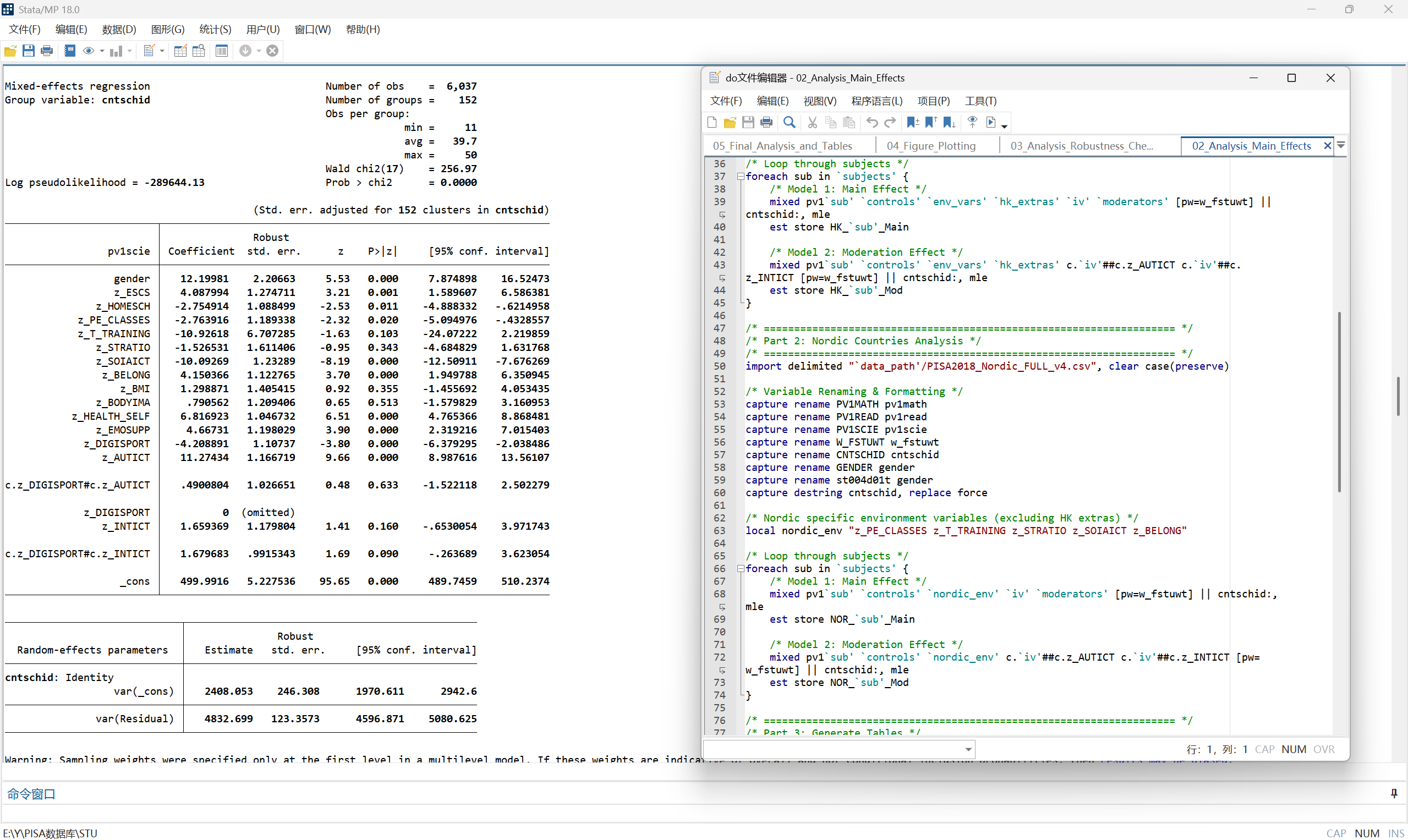
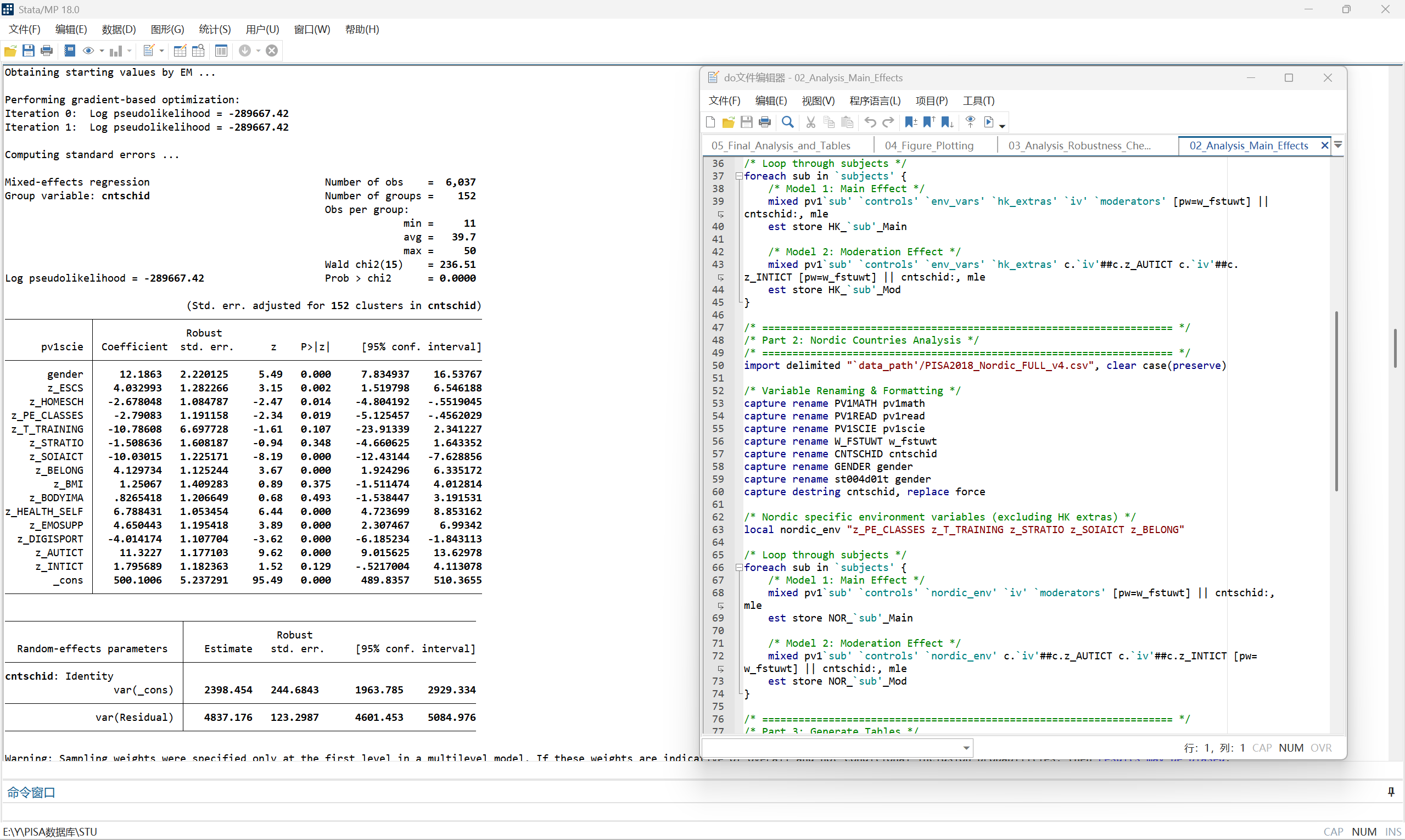
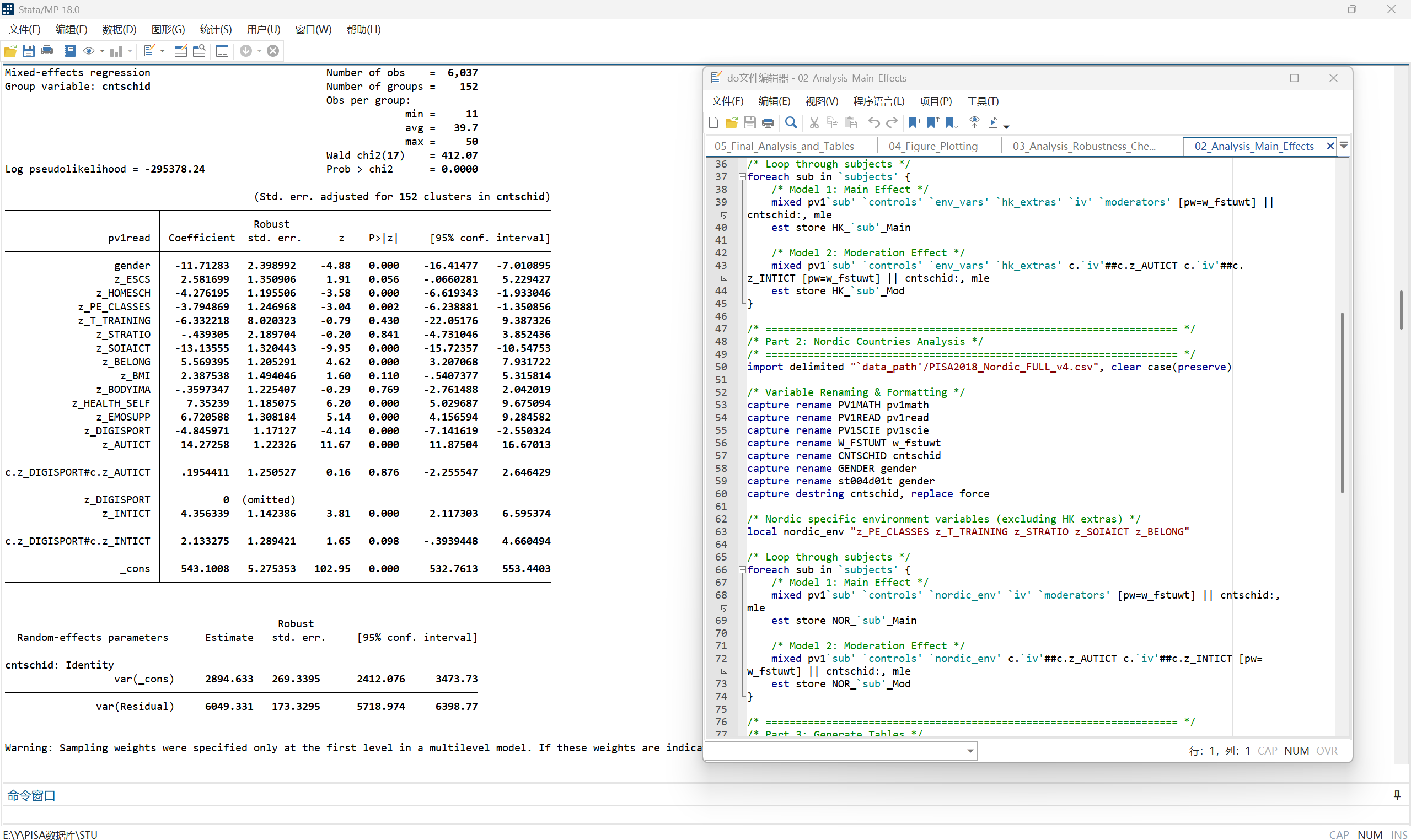
The screenshot below demonstrates the execution of the data cleaning script 01\_Data\_Cleaning\_Python.py in a Visual Studio Code environment. The script utilizes the pandas and sklearn libraries to load raw PISA 2018 SAS datasets, perform MICE imputation for missing values, and standardize key variables. The terminal output explicitly confirms the successful merging of student and school data, resulting in two cleaned CSV files (HKG\_FULL\_v4 and Nordic\_FULL\_v4) ready for statistical analysis.



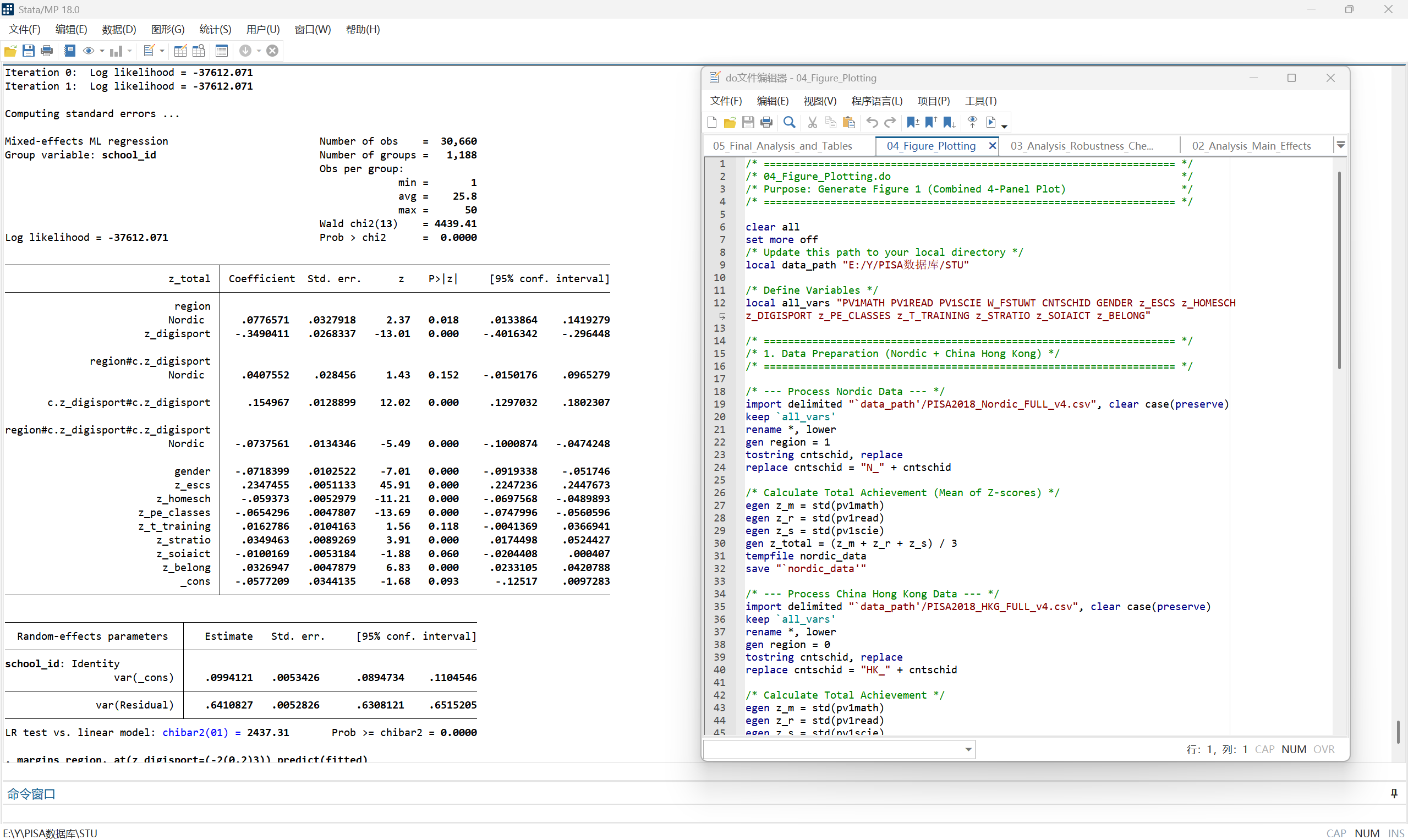
The following series of screenshots display the raw output from the main Hierarchical Linear Modeling (HLM) analysis executed in Stata 18.0 using the 02\_Analysis\_Main\_Effects.do script. The results present the mixed-effects regression coefficients for Mathematics, Reading, and Science achievement across both Hong Kong (HK) and Nordic samples. Key model specifications include the random-intercept structure to account for school-level clustering, and the inclusion of both linear (z\_DIGISPORT) and quadratic terms to test the non-linear U-shaped hypothesis. The output confirms the convergence of the models and provides the estimated coefficients, standard errors, and significance levels for all predictors.







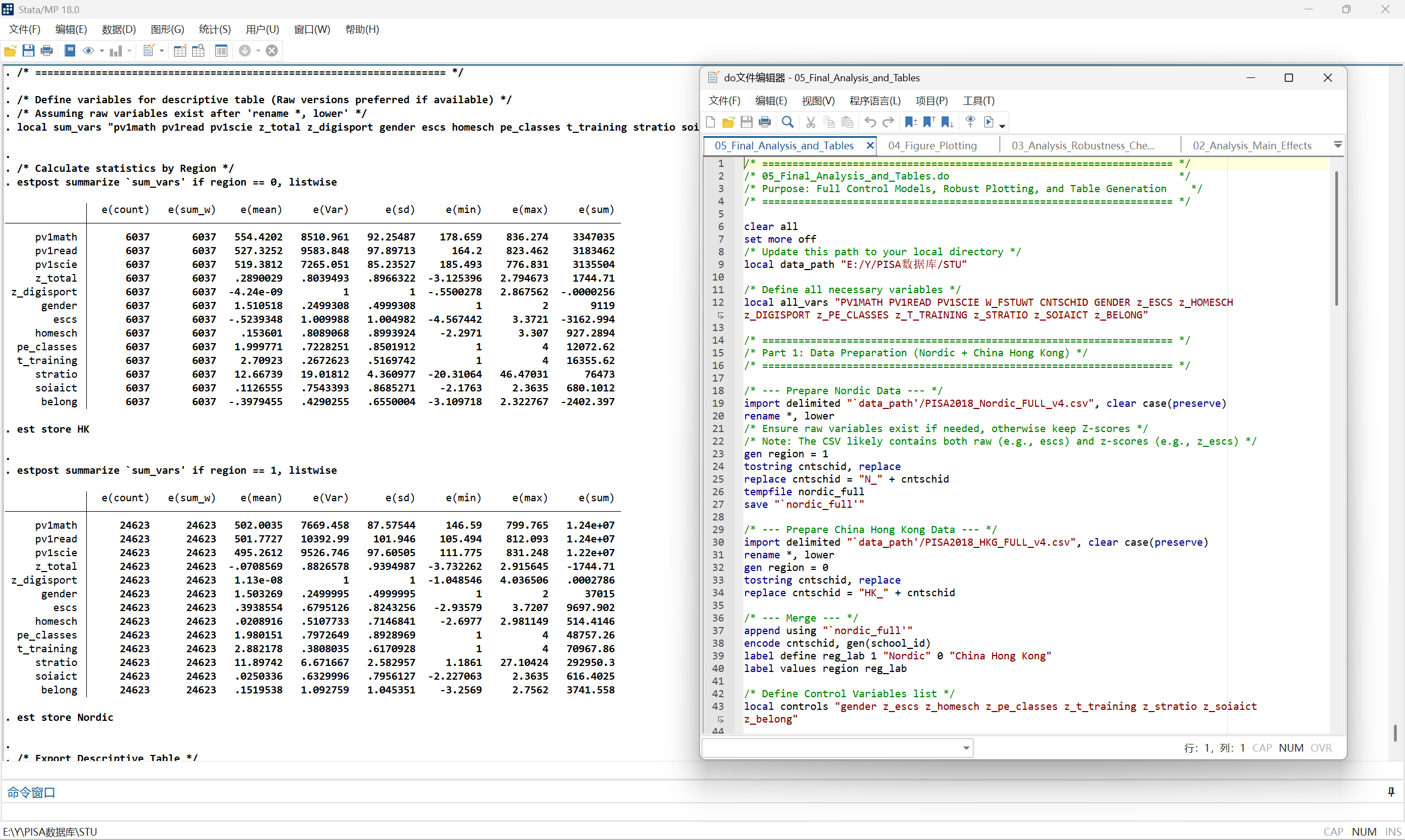
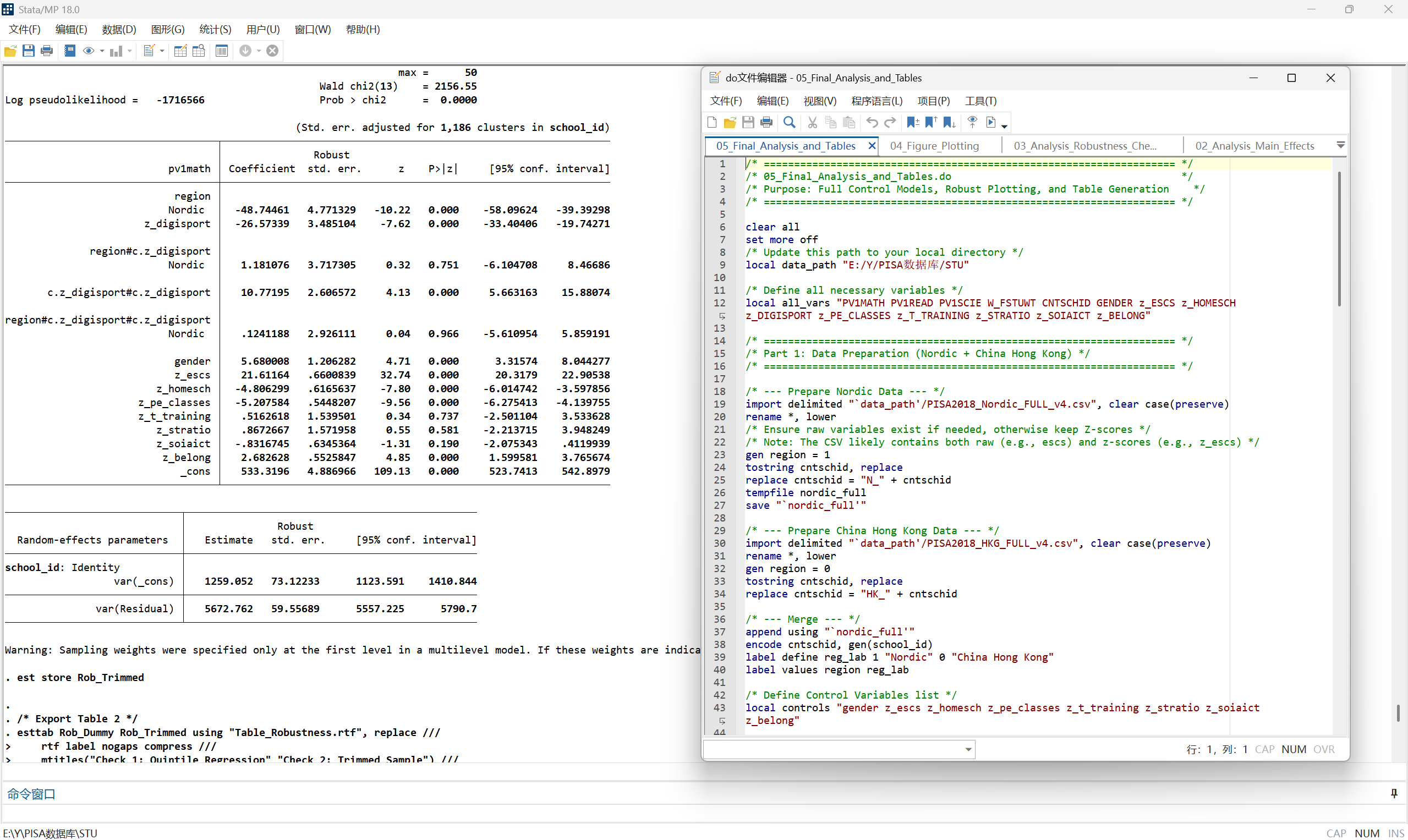
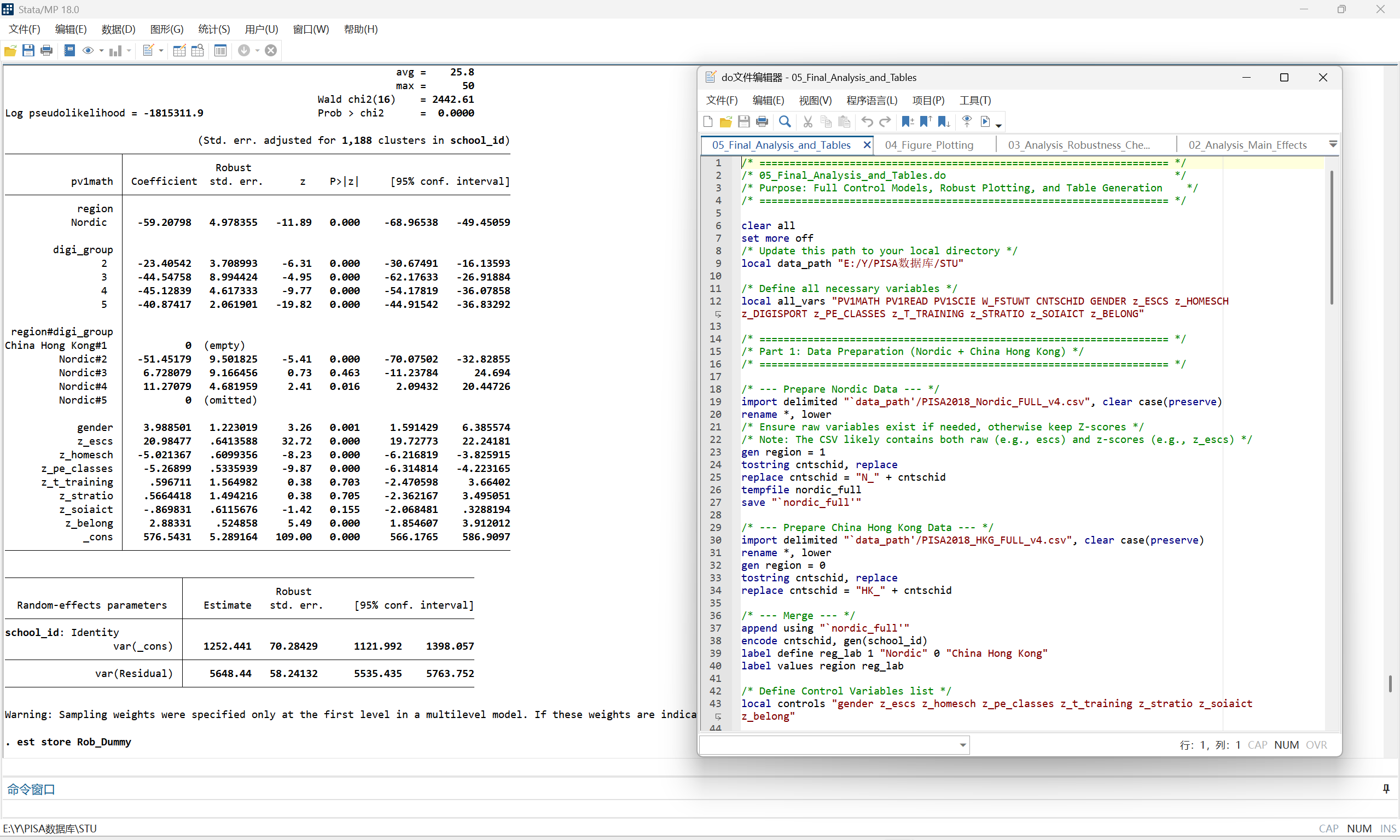
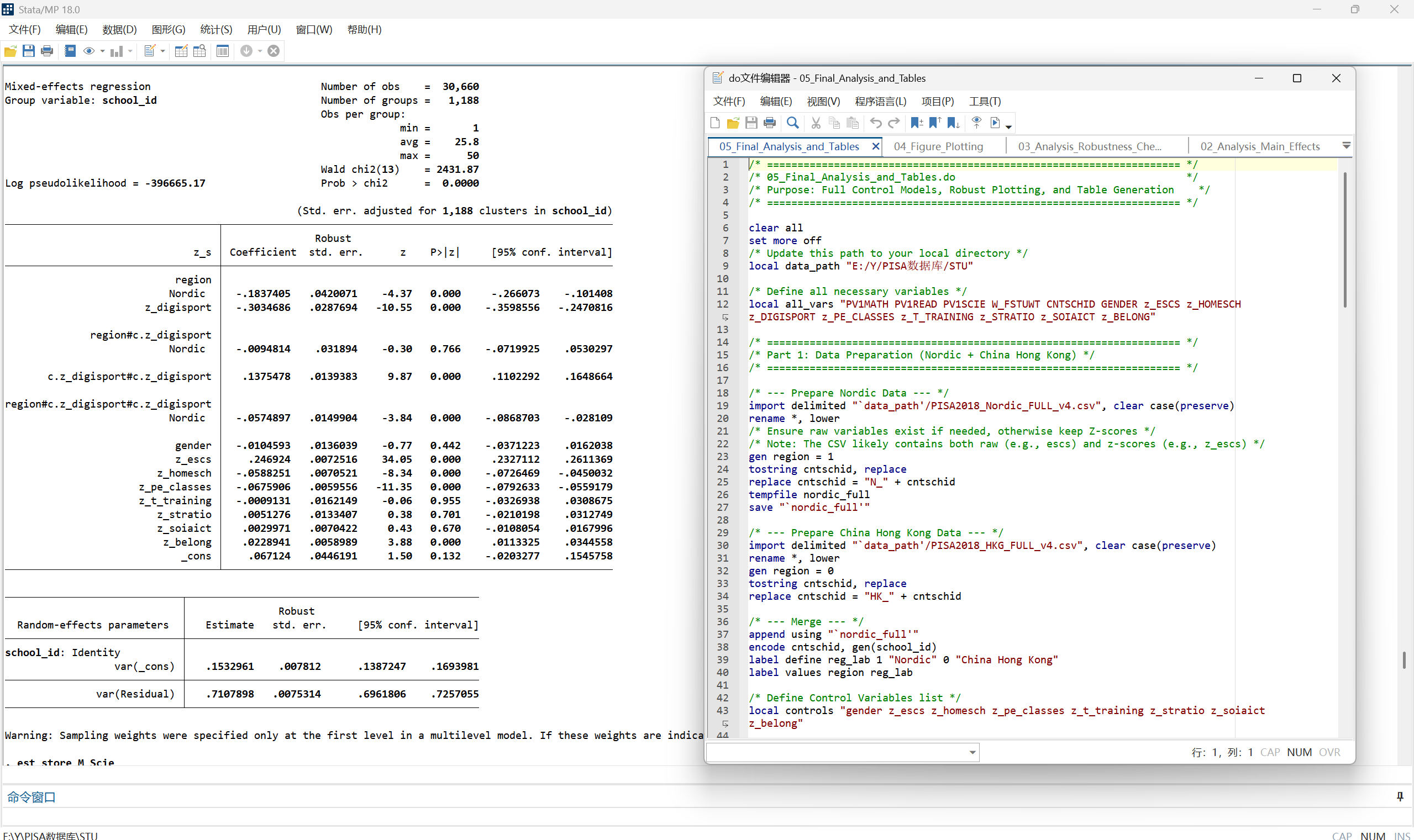
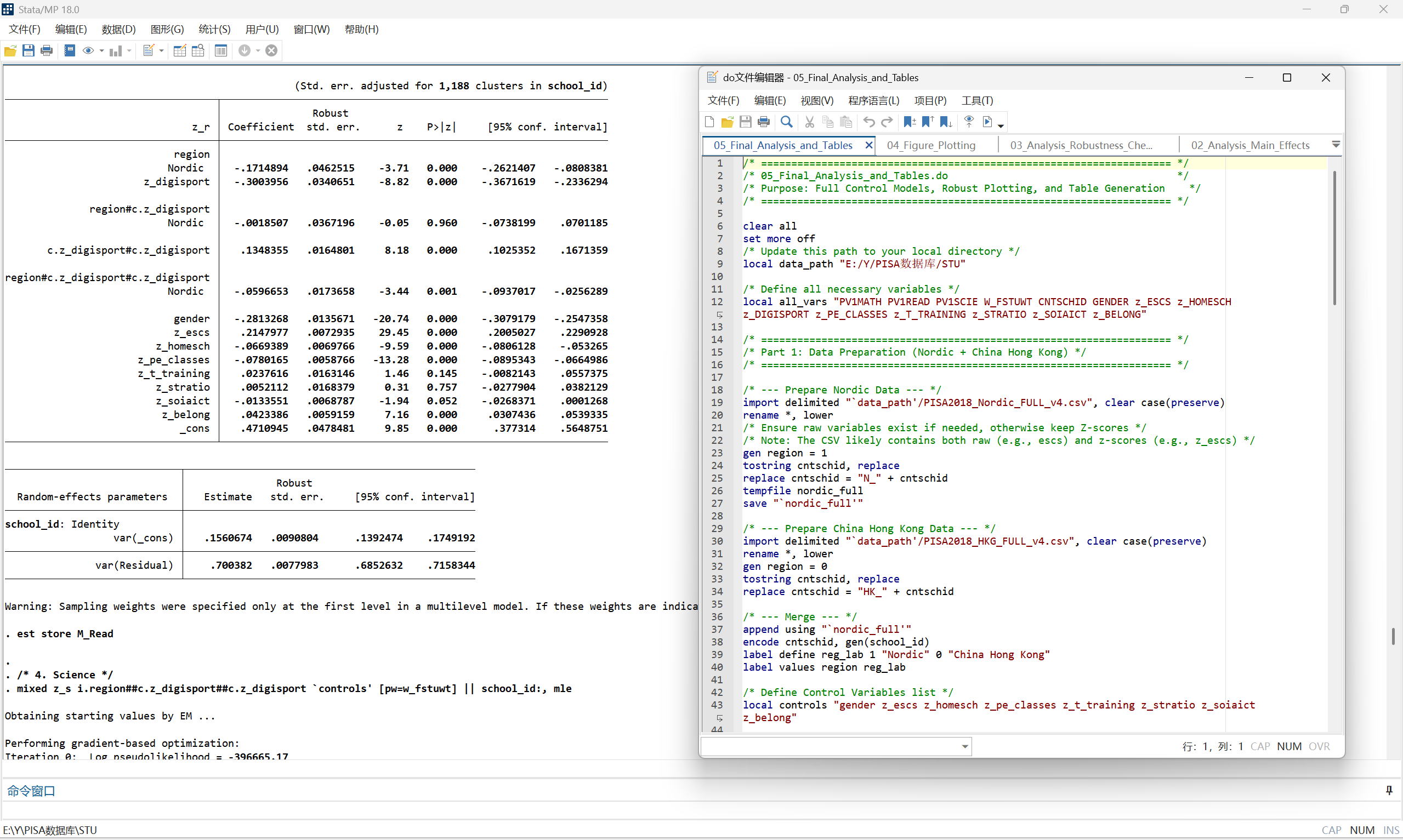
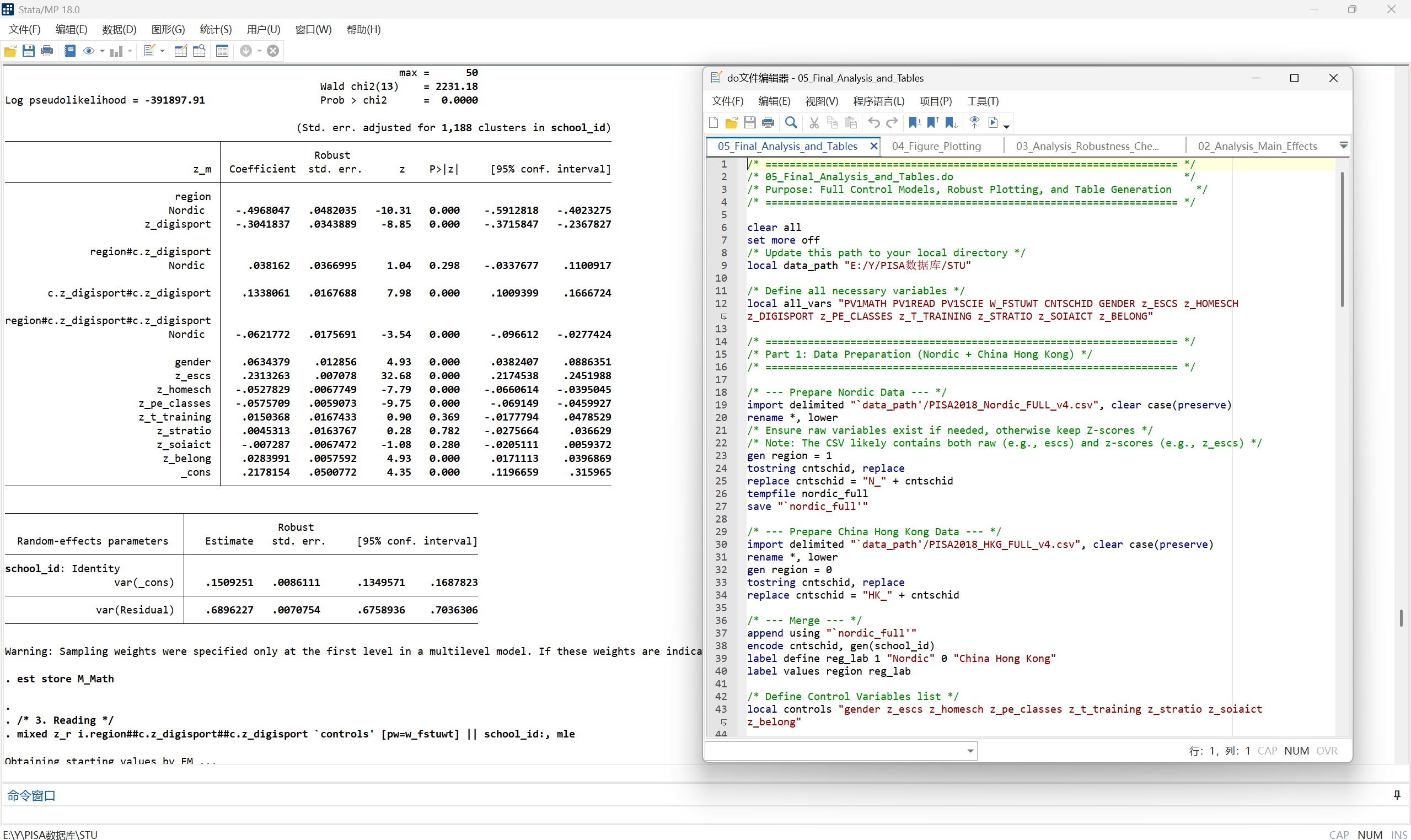
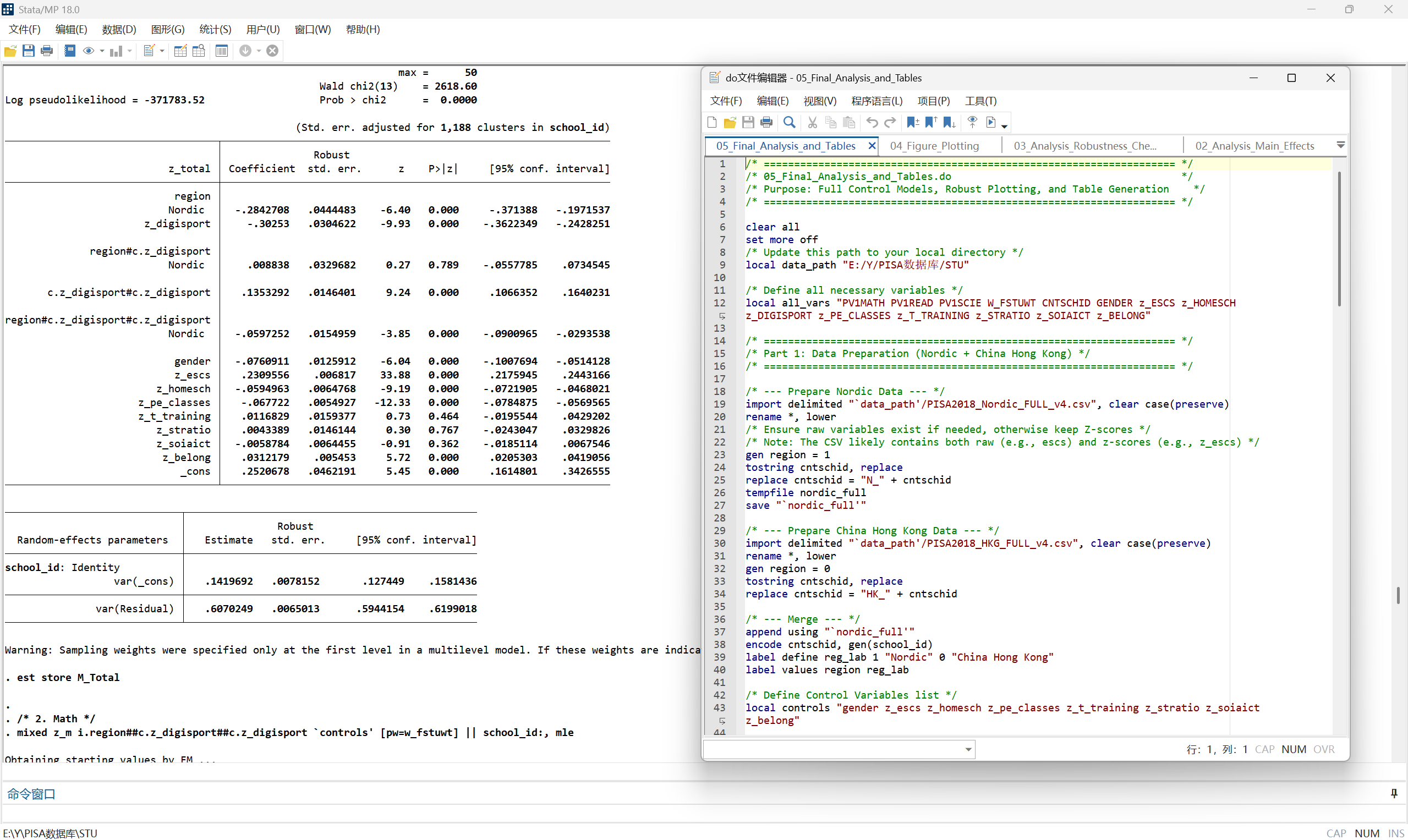
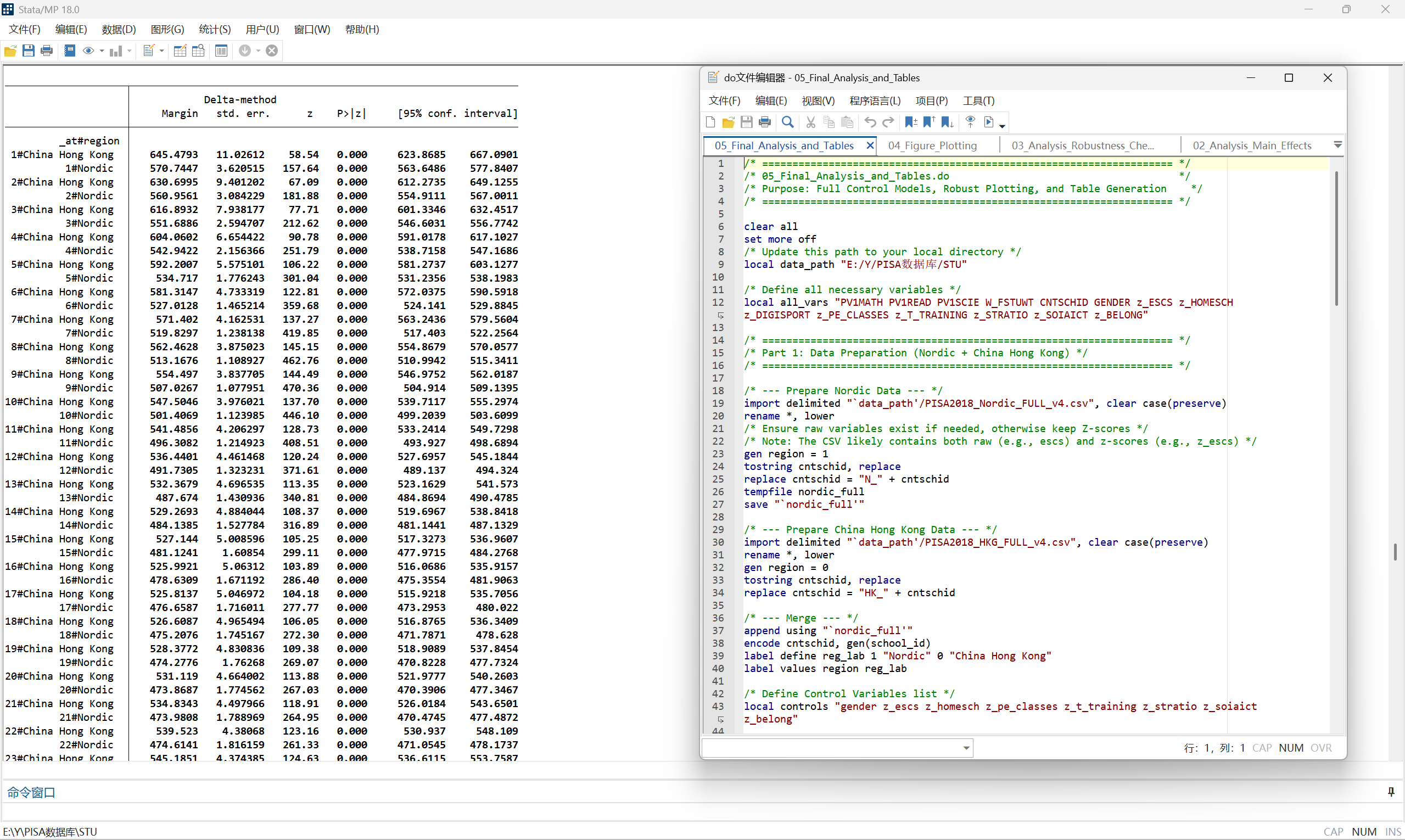
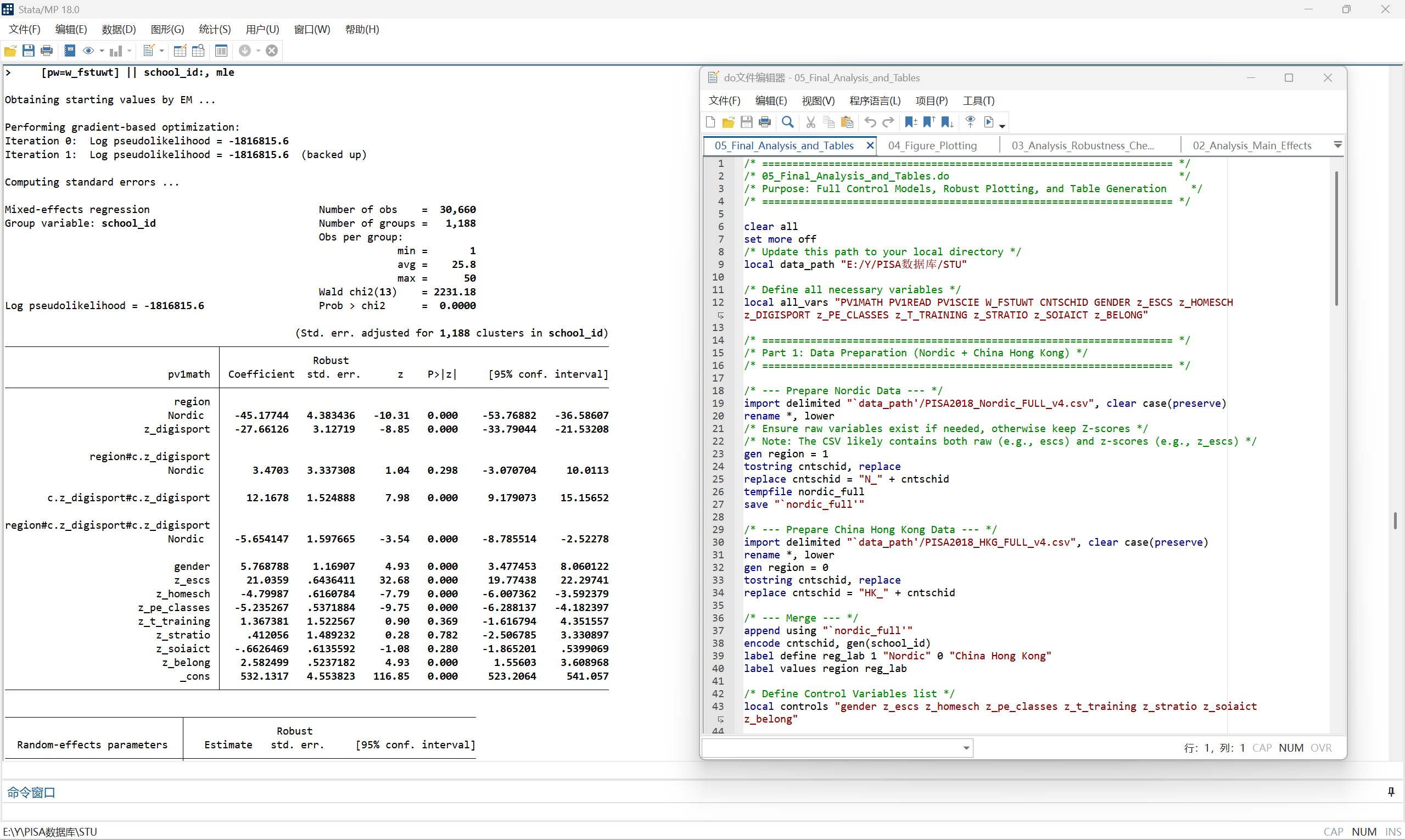
The screenshot below displays the execution of the 04\_Figure\_Plotting.do script, which visualizes the non-linear relationship between digital sports participation and academic achievement. Using the margins and marginsplot commands in Stata, the script calculates the predicted academic scores across the standardized range of digital sports frequency (Z-scores from -2 to +3). The resulting plot (Figure 1) clearly illustrates the U-shaped association, with the red solid line representing Hong Kong and the blue dashed line representing the Nordic countries, providing visual confirmation of the quadratic effects identified in the main HLM analysis.



The following set of screenshots documents the execution of the comprehensive analysis script 05\_Final\_Analysis\_and\_Tables.do. This script serves as the final integration step to generate the definitive statistics reported in the manuscript. The output logs explicitly demonstrate:

Descriptive Statistics: The calculation of means, standard deviations, and sample distributions for both Hong Kong and Nordic datasets (generating the data for Table 1).

Full Model Estimation: The execution of the fully specified Hierarchical Linear Models (HLM) for Mathematics, Reading, and Science. These outputs verify the final regression coefficients, standard errors, and significance levels after controlling for all student-level and school-level covariates.



The subsequent screenshots document the execution of robustness checks using the 03\_Analysis\_Robustness\_Checks.do script to validate the core non-linear findings. Two distinct approaches were employed:

Quintile Regression (Check 1): The sample was categorized into five frequency groups . The regression output confirms a non-linear pattern where the negative impact peaks at moderate intensity levels (Groups 3 and 4) and attenuates at high intensity (Group 5), independent of parametric assumptions.

Trimmed Sample Analysis (Check 2): Extreme outliers (|Z| > 2.5) were excluded to test model sensitivity. The results show that the quadratic term remains statistically significant (P < 0.001), corroborating the stability of the U-shaped association even after removing extreme values.

