

Module 5 Exercise

1. Preliminaries

a. Create a Project and RMarkdown document

Either in a preexisting RStudio Project folder, or in a new Project, create a RMarkdown document for the exercise.

b. Initialize the Ecdat library

For the first part of the exercise, you will be working with the [Ecdat](#) package, which has many economics datasets.

- Install and initialize the package.

2. Extramarital Affairs Exercise

b. Create an “affairs” tibble from the Fair data

For the first part of the exercise, you will be working the extramaritals data from Fair (Econometrica 1977).

- Create an “affairs” data frame from the Fair data in Ecdat as follows:

```
affairs <- Fair
```

- Turn the data set into a tibble
- Select the following variables: `nbaffairs`, `ym`, `child`, `rate`, `age`.
- Rename the variables as follows:
 - Change “nbaffairs” to “affairs”
 - Change “ym” to “yrsmarr”
 - Change “rate” to “mrating”

c. Convert `mrating` to a factor

d. Perform regression

- Create a regression object that regresses `affairs` on `age`, `child`, `yrsmarr` and $yrsmarr^2$.
- Then view the summary output of the regression.

f. Perform model diagnostics

- Test for heteroskedasticity.
- Test for missing polynomial terms.

Write the statistical decision from each test in your RMarkdown report.

g. Create a bar plot of average number of affairs by marital rating

- Create a summary data frame of number of affairs by marital rating
- Create a bar plot of the results, adding a graph title, axis labels, and style it like a Stata graph using `theme_stata()` from the `ggthemes` package.

3. Inequality in the World Development Indicators Example

For the next part of the exercise, let's revisit the inequality indicators from the World Development Indicators dataset.

- Import "wdi_data", which is just the data from Exercise 2, summarized by country and year from 2000.

a. Create scatter plots

- Create two scatterplots: each with secondary school completion as the x-axis.
 - For the first scatterplot, choose female-to-male employment ratio as the y-variable.
 - For the second scatterplot, make unemployment among those with basic education the y-variables:
- Color the points based on the value of GDP.
- Add titles and axis labels.

b. Perform regression

Perform a pooled OLS regression of GDP average on secondary school completion and view summary information from the regression.

c. Test for autocorrelation

d. Fixed effects regression

Now instead perform the regression of GDP on secondary school completion as a fixed effects regression (including fixed effects for country and year).

e. Autocorrelation robust inference

Test the coefficients of the model using the Stata-style HC_1 estimation of Newey-West heteroskedasticity and autocorrelated (HAC) robust standard errors.

- Use the function `vcovNW()` for specifying the variance method inside of `coefest()`.