

MIEF Skills Workshop: Modeling in R

Session 1

January 30, 2026

Course Purpose

What this course will do:

- Introduce you to iterative regression methods in R
- Provide you with a template to conduct statistical modeling, analysis, and visualization in R

What this course WON'T do:

- Teach you econometrics

Today

This course will involve less theory than the previous course. You should refer to the introductory R course if you're fuzzy on basic R details.

This course will be far more **practical**. I will be building an R script alongside you, as you build your own R script to "learn by doing."

Today, we will be:

- Building a simple statistical model to answer a very simple research question: one dependent variable, one independent variable
- Fleshing the statistical process out more with control variables and fixed effects
- Iterating our statistical process to encompass multiple dependent variables
- Exploring a simple method to output regression tables
- Iterating that method to work for multiple dependent variables

Quick Refresh: Tidyverse

Base R Layout

Tidyverse Layout

```
names(iris)

## [1] "Sepal.Length" "Sepal.Width"    "Petal.Length"  "Petal.Width"   "Species"

str_replace(str_to_lower(names(iris)), "\\.", "_")

## [1] "sepal_length" "sepal_width"   "petal_length" "petal_width"  "species"
```

Quick Refresh: Tidyverse

Base R Layout

Tidyverse Layout

Tidyverse functions introduce a 'cleaner' method to write code out, using what is called the 'pipe operator': `%>%`. It's almost like writing a recipe, step by step.

```
names(iris)
```

```
## [1] "Sepal.Length" "Sepal.Width"    "Petal.Length"  "Petal.Width"   "Species"
```

```
iris %>%  
  names() %>%  
  str_to_lower() %>%  
  str_replace("\\"., "_")
```

```
## [1] "sepal_length" "sepal_width"   "petal_length" "petal_width"  "species"
```

(Note — you can also write the pipe operator as `|>`)

Quick Refresh: Using `map()`

The `map()` function from the `purrr` package allows us to **apply the same function(s) to a set of objects**.

```
library(purrr)
purrr::map(
  .x = list(
    c(-10, 0, 10), c(1, 2, 3, 4, 5)
  ),
  .f = function(x)
    (x - min(x, na.rm = TRUE)) /
    (max(x, na.rm = TRUE) - min(x, na.rm = TRUE))
)
```

```
## [[1]]
## [1] 0.0 0.5 1.0
##
## [[2]]
## [1] 0.00 0.25 0.50 0.75 1.00
```

Quick Refresh: Using `map()`

`map()` has a shorthand to simplify its use: instead of writing the function out, you can replace `function() {}` with `\(argument)` and the function's argument with `argument` ("argument" can be replaced with whatever you want).

```
c(1, 3) %>%  
  map(  
    function(x) {  
      x + 10  
    }  
  ) %>%  
  as.numeric()  
  
## [1] 11 13
```

```
c(1, 3) %>%  
  map(  
    \vlu vlu + 10  
  ) %>%  
  as.numeric()  
  
## [1] 11 13
```

Regression in R

The most basic regression function in R is `lm()`. It performs a linear regression, with the possibility to add additional control variables.

```
Syntax: lm(data = ds, formula = x ~ y)
```

Today, we will be using both `lm()` and a function, `felm()` from the `lfe` package, which allows us to add fixed effects and cluster-robust standard errors to our regression. This is just a peek into how to perform more advanced econometrics in R, although it is not the focus of this course.

```
Syntax: lfe::felm(data = ds, formula = x ~ y | fe | 0 | fe)
```

The `felm()` formula allows us to include (1) the basic formula with independent and dependent variable(s), (2) our fixed-effect variable(s), (3) instrumental variable(s), and (4) variables within which we want to cluster standard errors.

Today's Practical Component

- You will be following my live coding and then replicating it on your script.
- After each coding chunk, you will get 5-10 minutes to apply it to your script:
 - If you are not feeling confident in the live setting, you can exactly replicate what I am doing
 - If you would like to get a head start on your course assignment, you can apply the process I am following to **a different set of variables within the dataset**.

The World Values Survey

"The survey, which started in 1981, seeks to use the most rigorous, high-quality research designs in each country. The WVS consists of nationally representative surveys conducted in almost 100 countries which contain almost 90 percent of the world's population, using a common questionnaire. [...] WVS seeks to help scientists and policy makers understand changes in the beliefs, values and motivations of people throughout the world."

- Social values, attitudes & stereotypes
- Societal well-being
- Social capital, trust and organizational membership
- Economic values
- Corruption
- Migration
- Post-materialist index
- Science & technology
- Religious values
- Security
- Ethical values & norms
- Political interest and political participation
- Political culture and political regimes
- Demography

Today's Practical Component

At this point, you should know how to set up a data project folder. Do this now. This folder should include:

- A "data" folder
- A "code" folder
- An "output" folder
- A **.rproj** file

Download the course data and documentation from Canvas. Once this is done, **open a new R script**. I will go through a chunk of code, and then you will be expected to replicate it.

Today's Practical Component

I am answering the research question: what relationship is there between a respondent's education level and how they raise their children? To do this, I am using Q275 (education level) as the independent variable, and Q7-17 (child values) as the dependent variables from the World Values Survey.

For the purpose of today's session, I would recommend sticking to education level as the independent variable. Here are some other dependent variables you could consider using:

- Life-Importance Subjects (Q1-6)
- Neighbor Preferences (Q18-26)
- Trust in Other People (Q58-63)
- Confidence in Institutions (Q64-89, NOT with 82_suffix options)
- Organization Membership (Q94-104, NOT with _r suffix)
- Corruption Perception (Q113-117)
- Immigration Perception (Q122-129)
- Information Source (Q201-208)
- Elections Perception (Q224-233)

Today's Practical Component

1. Setup
2. Import Data
3. Explore Data
4. Simple Regression – One Dependent Variable
5. Iterated Regressions – Multiple Dependent Variables
6. Simple Visualization (Stargazer and LateX)
7. Iterated Visualizations (Stargazer and LateX)
8. Outputting Iterated Regressions as a Dataset

Annex

Survey Variable Details

Life-Importance Subjects (Q1-6)

For each of the following aspects, indicate how important it is in your life. Would you say it is very important, rather important, not very important or not important at all?

- Dataset matching regex: "q00[1-6]_life"
- Values:
 - 0 — Not at all important
 - 1 — Not very important
 - 2 — Rather important
 - 3 — Very important
- Subjects:
 - Family
 - Friends
 - Leisure
 - Politics
 - Work
 - Religion

Child Values (Q7-17)

Here is a list of qualities that children can be encouraged to learn at home. Which, if any, do you consider to be especially important? Please choose up to five.

- Dataset matching regex: "q0(0[7-9]|1[0-7])_child"
- Values:
 - 0 — Not mentioned
 - 1 — Important
- Subjects:
 - Good manners
 - Independence
 - Hard work
 - Responsibility
 - Imagination
 - Tolerance
 - Thrift
 - Perseverance
 - Faith
 - Selflessness
 - Obedience

Neighbor Preferences (Q18-26)

On this list are various groups of people. Could you please mention any that you would not like to have as neighbors?

- Dataset matching regex: "q0(1[8-9]|2[0-6])_neighbor"
- Values:
 - 0 — Not mentioned
 - 1 — Mentioned
- Subjects:
 - Drugs addicts
 - People of a different race
 - People who have AIDS
 - Immigrants/Foreign workers
 - Homosexuals
 - People of a different religion
 - Heavy drinkers
 - Unmarried couples living together
 - People who speak a different language

Trust in Other People (Q58-63)

I'd like to ask you how much you trust people from various groups. Could you tell me for each whether you trust people from this group completely, somewhat, not very much or not at all?

- Dataset matching regex: "q0(5[8-9]|6[0-3])_trust"
- Values:
 - 0 — Do not trust at all
 - 1 — Do not trust very much
 - 2 — Trust somewhat
 - 3 — Trust completely
- Subjects:
 - Family
 - Neighborhood
 - People you know personally
 - People you meet for the first time
 - People of another religion
 - People of another nationality

Confidence in Institutions (Q64-89)

I am going to name a number of organizations. For each one, could you tell me how much confidence you have in them: is it a great deal of confidence, quite a lot of confidence, not very much confidence or none at all?

- Dataset matching regex: "q0(6[4-9]|7[0-9]|8[0-9])_confidence"
- Values:
 - 0 — None at all
 - 1 — Not very much
 - 2 — Quite a lot
 - 3 — A great deal

Confidence in Institutions (Q64-89)

- Subjects:
 - Churches
 - Armed forces
 - Press
 - Television
 - Labor unions
 - Police
 - Justice system
 - Government
 - Political parties
 - Parliament
 - Civil services
 - Universities
 - Elections
 - Major companies
 - Banks
 - Environmental organizations
 - Women's organizations
 - Charities
 - Major regional organizations (e.g. EU)
 - United Nations (UN)
 - International Monetary Fund (IMF)
 - International Criminal Court (ICC)
 - North Atlantic Treaty Organization (NATO)
 - World Bank (WB)
 - World Health Organization (WHO)
 - World Trade Organization (WTO)

Organization Membership (Q94-104)

Now I am going to read out a list of voluntary organizations; for each one, could you tell me whether you are a member, an active member, an inactive member or not a member of that type of organization?

- Dataset matching regex: "q(09[4-9]|10[0-4])_member"
- Values:
 - 0 — Don't belong
 - 1 — Inactive member
 - 2 — Active member
- Subjects:
 - Church or religion organization
 - Sport or recreational organization
 - Art, music, or educational organization
 - Labor union
 - Political party
 - Environmental organization
 - Professional organization
 - Charitable or humanitarian organization
 - Consumer organization
 - Self-help or mutual aid group
 - Women's group

Corruption Perception (Q113-117)

Among the following groups of people, how many do you believe are involved in corruption? Tell me for each group if you believe it is none of them, few of them, most of them or all of them?

- Dataset matching regex: "q11[3-7]_corr"
- Values:
 - 0 — None of them
 - 1 — Few of them
 - 2 — Most of them
 - 3 — All of them
- Subjects:
 - State authorities
 - Business executives
 - Local authorities
 - Civil service providers
 - Journalists and media

Immigration Perception (Q122-129)

From your point of view, what have been the effects of immigrants on the development of this country?

- Dataset matching regex: "q12[2-9]_immigr"
- Values:
 - 0 — Disagree
 - 1 — Hard to say
 - 2 — Agree
- Subjects:
 - Fill useful jobs in the workforce
 - Strengthen cultural diversity
 - Increase the crime rate
 - Give asylum to political refugees
 - Increase the risks of terrorism
 - Help poor people establish new lives
 - Increase unemployment
 - Lead to social conflict

Information Source (Q201-208)

People learn what is going on in this country and the world from various sources. For each of the following sources, please indicate whether you use it to obtain information daily, weekly, monthly, less than monthly or never.

- Dataset matching regex: "q20[1-8]_info_source"
- Values:
 - 0 — Never
 - 1 — Less than monthly
 - 2 — Monthly
 - 3 — Weekly
 - 4 — Daily
- Subjects:
 - Daily newspaper
 - TV news
 - Radio news
 - Mobile phone
 - Email
 - Internet
 - Social media
 - Talk with friends or colleagues

Elections Perception (Q224-233)

In your view, how often do the following things occur in this country's elections?

- Dataset matching regex: "q2(2[4-9]|3[0-3])_elect"
- Values:
 - 0 — Not at all often
 - 1 — Not often
 - 2 — Fairly often
 - 3 — Very often
- Subjects:
 - Votes are counted fairly
 - Opposition candidates are prevented from running
 - TV news favors the governing party
 - Voters are bribed
 - Journalists provide fair coverage of elections
 - Election officials are fair
 - Rich people buy elections
 - Voters are threatened with violence at the polls
 - Voters are offered a genuine choice in the elections
 - Women have equal opportunities to run the office