

ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL SCIENCES

DEPARTMENT OF STATISTICS

Project
Econometrics-Stat4061

1. Data Set: Micro- and Small Enterprises (MSE)

Consider the data on several characteristics of a random sample of Micro- and Small Enterprises (MSE) in Ethiopia. The variables considered are:

Variable	description	Coding
business_registered	Is the business registered?	0 = No, 1 = Yes
TIN	Do you have TIN?	0 = No, 1 = Yes
business_license	Do you have business license or certificate?	0 = No, 1 = Yes
MSE_type	Type of MSE	1 = Micro, 2 = Small
loan	Have you obtained loan from any micro finance institution?	0 = No, 1 = Yes
land_premise	Have you ever received land or working premises for your enterprise?	0 = No, 1 = Yes
one_stop_service	Did you receive one-stop service at the nearest administration?	0 = No, 1 = Yes
util_rate	Enterprise's current capacity utilization rate	Actual data
current_capital	Current capital (in birr)	Actual data
initial_capital	Initial capital (in birr)	Actual data
years_establish	Years since establishment	Actual data
years_schooling	Years of schooling of owner(s)	Actual data
no_paid_workers	Number of paid workers	Actual data
gender	Gender of owner	0 = Male, 1 = Female
highest_education	Highest level of education of owner(s)	1 = Illiterate, 2 = Read and Write, 3 = Primary, 4 = High school, 5 = Vocational training, 6 = University degree
subsector	Sub-sector of MSE	1 = Metal and wood work 2 = Construction 3 = Agro-processing 4 = Textile and garment 5 = Leather and footwear 6 = Retail 7 = Urban agriculture subsector Sub-sector of MSE 8 = Others

Your study will be focused on assessing the relation by regressing current capital on initial capital, years since establishment, years of schooling of owner(s), number of paid workers, gender, having business license, access to loan from micro finance institutions and access to land/working premises.

2. General Instructions

- Form a group with not more than 5 students.

- Write a scientific report with **not more than 8 pages** (without preamble pages and appendix). Your report should include at least **Introduction, objectives, methods, analysis results, discussion, conclusion and references**.
- The analysis should include primarily:
 - doing appropriate descriptive statistical analysis for the variables as well as exploratory data analysis.
 - * Plot:
 - i) A multiple bar chart for highest level of education of owner(s) by gender and describe it.
 - ii) A pie-chart for sub-sectors of MSE and describe it.
 - iii) A histogram for current capital by superimposing the normal curve. Comment on the shape of the histogram.
 - * Find the first, second and third quartiles for current capital and then recode current capital as:

Original data	coding	Name
Less than first quartile	1	Low
Between first and second quartile	2	Low-medium
Between second and third quartile	3	High-medium
Greater than third quartile	4	High

Then apply appropriate tests to see if (a) gender and current capital and (b) highest level of education of owner(s) and current capital are associated or not. Briefly discuss the results.

- fitting the appropriate multiple linear regression model with the possible interactions.
- Checking the validity of all the assumptions using the best tools that you have learned in the course.
- Testing the significance of the regression relation.
- Proposing and testing a sensible hypotheses from the regression line.

Tip: Correctly addressing all the concepts from the course will increase your mark.

- Discuss the results in a professional way.
- You can use any software you want, such as R and/or SAS. (if you need help regarding software usage, feel free to contact the instructor)
- **Maximum Mark:** 10(presentation) + 20(Quality of Paper) = 30.
- Due date: **January, 2026-After your Exam.**