

# Lab Assignment: *MA660E Quantitative Data Analysis*

Dr. Yuanji Cheng  
Martin Svensson

October 19, 2024

## Part One

In this part, you will work with job survey data from the book *Quantitative Data Analysis with SPSS*. The dataset is available as either an SPSS file (`jss13_ht22.sav`) or an Excel file (`Data_source.xlsx`).

The variables used are as follows:

- **ethnicgp** (ethnic group): 1 = White, 2 = Asian, 3 = West Indian, 4 = African, 5 = Other
- **gender**: 1 = Male, 2 = Female
- **income**: Gross annual income before tax (in £1000)
- **age**: Age in years
- **years**: Number of years working at this company
- **commit**: Organizational commitment (scale 1 to 5)
- **satis**: Job satisfaction
- **autonom**: Job autonomy
- **routine**: Job routine
- **attend**: Attendance at meetings (1 = Yes, 2 = No)
- **skill**: Skill level (1 = Unskilled, 4 = Highly skilled)
- **prody**: Productivity rating (1 = Very poor, 5 = Very good)
- **qual**: Quality rating (1 = Very poor, 5 = Very good)
- **absence**: Number of days absent in the last 12 months

You can use any software for calculations (MATLAB, Python, Excel, SPSS, R, etc.).

### Exercise 1.1

- a) Create a bar chart for gender and a pie chart for ethnic group.
- b) Summarize the **age** data using the five-number summary: minimum, maximum, median, 1st quartile, and 3rd quartile, then generate a box plot.
- c) Compute the mean and standard deviation of **income** and create a histogram of it.

### Exercise 1.2

- a) Create a scatter plot to visualize the relationship between **income** and **absence**.
- b) Build a simple linear regression model with **income** as the dependent variable and **absence** as the independent variable. Report the determination coefficient ( $R^2$ ).

### Exercise 1.3

Study a multiple regression model where **satis** (job satisfaction) is the dependent variable, and the following are independent variables: **commit**, **autonom**, **income**, **skill**, **qual**, **age**, and **years**.

- a) Identify which variables do not have a significant impact on **satis**.
- b) Simplify the regression model by removing non-significant variables.

### Exercise 1.4

Find the confidence interval for **job satisfaction**, and the confidence interval for the difference in **job satisfaction** between men and women.

### Exercise 1.5

Use the Mann-Whitney-Wilcoxon test to check if there is a significant difference in **skill** levels between men and women. Compare the result with the confidence interval for the difference.

### Exercise 1.6

Use the Kruskal-Wallis test to determine if there is a significant difference in **absence** among different ethnic groups. Compare this with the results from a One-Way ANOVA.

### Exercise 1.7

Recode the **income** variable into an **income class** using the following class limits:

- Low income class: [Min, Q1]
- Middle income class: (Q1, Q3]
- High income class: (Q3, Max]

Investigate if there is a significant relationship between **income class** and **skill**.

## Part Two

For this part, you will work with your own dataset. You have the freedom to choose the topic and the variables for analysis.

### Exercise 2.1

Perform a descriptive statistics analysis on at least two qualitative and two quantitative variables.

### Exercise 2.2

Compute the confidence interval for one quantitative variable, and for the difference between two groups.

### Exercise 2.3

Perform a T-test to check if there is a significant difference between two groups, or conduct an ANOVA to see if all groups have the same mean value for a specific characteristic.

### Exercise 2.4

Conduct a non-parametric test for the same variable as in Exercise 2.3 and compare the conclusions with those from the ANOVA.

### Exercise 2.5

Carry out a correlation analysis. Identify the strongest correlations and any statistically insignificant relationships.

### Exercise 2.6

Perform a multiple linear regression analysis.

## Optional: COVID-19 Analysis

If you're interested in COVID-19 data, you can obtain relevant datasets from the following sources:

- World Health Organization (WHO)
- Johns Hopkins University Coronavirus Resource Center

## Lab Report Guidelines

You may work individually or in teams of up to two students. Your lab report must include both tables/figures and written interpretations for each question. Submitting only tables or figures without explanation will be considered incomplete.

**Deadline:** Submit the lab report via Canvas by **Sunday, October 27, 2024**.