Assignment 2

Directions

Submit two files on Moodle before the deadline:

[1] your **solution set** as a *.pdf [2] the **R script** as a *.R type file

For both, the first line should include your student ID and the assignment number

If you worked with one or two other students write "group:" and then their student ID numbers

[1] the **solution set** should include for each answer:

- the question number/part
- your answer in words
- the command in R that was used (if relevant)
- the output from Rstudio (if relevant)

[2] the **R script** should include:

- all the commands you used to generate the solutions
- individually written comments that explain your code to the grader

Part A Interpreting R output

Questions 1-2 are based on R outputs using the `newdata` object from **HW1.**

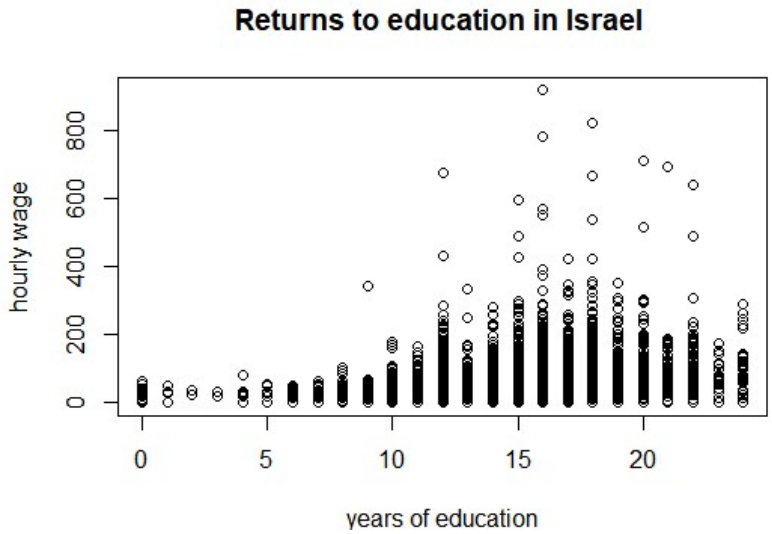
1. **Level-Level model.** Consider the output of the regression of `wage_per_hour` against `schooly` below.

```
Call:
lm(formula = wage_per_hour ~ schooly, data = newdata)

Coefficients:
(Intercept)      schooly
    -17.577        5.097
```

   a. Report the estimation results as a formula and interpret the slope parameter

   b. The scatterplot below includes the estimated level-level model. Do you think the relationship looks linear?



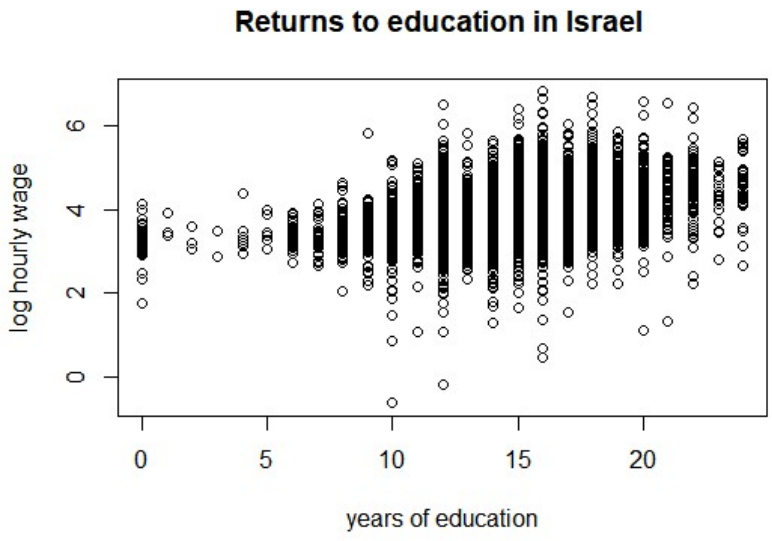
   c. For the estimated level-level model, what is the predicted hourly wage for an individual with 11 years of education? What about 17 years? What about no education?

2. Log-Level model.

a. The new variable `ln_wage` is defined below as the natural log of hourly wage. What does the R code in the box below do? Why is it necessary to include it before estimating the log-level model?

```
> newdata$ln_wage = log(newdata$wage_per_hour)
> summary(newdata$ln_wage)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   -Inf   3.310   3.735    -Inf   4.240   6.823    2427
> newdata$ln_wage[ which(newdata$ln_wage == -Inf) ] = NA
> summary(newdata$ln_wage)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 -0.638   3.470   3.838   3.908   4.312   6.823    3321
```

b. The scatterplot below shows log hourly wage (as defined in a) against years of education. Does the relationship look linear?



c. The regression of `ln_wage` against `schooly` is reported below. Interpret the slope.

```
Call:
lm(formula = ln_wage ~ schooly, data = newdata)

Coefficients:
(Intercept)      schooly
     2.7368       0.0823
```

d. For the regression in (c) what is the predicted hourly wage for an individual with 11 years of education? What about 17? What about no education?

e. For the regression in (c) what is the marginal effect of a year of education on hourly wage at the sample average hourly wage? (This question is tricky. *hint*: recall that the regression passes through the sample averages, also that $\overline{\log(x)} \neq \log(\bar{x})$)

3. Sampling distribution of the OLS estimator

a. Show that the variance of the OLS intercept estimator is equal to

$$Var(\hat{\beta}_0) = \sigma_u^2 \left[\frac{1}{N} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right].$$

Make sure to explain each step.

Part B Modeling the money supply in R

For this part of the assignment use R studio. Download the file money_v2026.xslx from the course website. It contains data on the following monthly data for the US between the years 1982-2024:

- `MONEY` – total amount of currency in circulation during one month in current prices (billions of dollars)
- `INTEREST` – the index rate (federal funds effective rate)
- `CPI` – consumer price index (a number between 0 and 100)

4. Preliminaries

a. Import the data and summarize the variables in the dataset.

b. Create the variable `MONEY_REAL` which measures the currency in circulation in units of real billions of dollars (hint: for each year divide MONEY by CPI to convert the nominal money supply to the real money supply) and summarize the new variable.

c. Plot the interest rate (x-axis) against the real money supply (y-axis).

d. Does the relationship look linear? Explain.

5. The Log-Log Model

According to theory of money supply the following model holds

$$\ln(M_i) = \alpha + \beta \ln(r_i) + u_i$$

where M is currency in circulation adjusted to real dollars (billions) and r is the interest rate.

a. Create the variables for the money supply model:

`ln_M` (the natural log of the variable from b)

`ln_r` (the natural log of the interest rate)

b. Estimate the money supply model parameters. Report the results as a formula.

c. What is the interpretation of the slope estimate?

d. What is the predicted amount of real money in circulation when the interest rate is 10%?

e. What is the predicted amount of real money in circulation when the interest rate is 1%?

f. What is the marginal effect of the interest rate on the money supply (at the point where both are at their sample average)? Interpret this result.

g. What is the rate of return at the point where both interest rates and real money supply are at their sample average? Interpret this result.

6. Classical assumptions of OLS

a. Plot the log real money supply against the log interest rate and add the estimated regression line. Does the relationship look linear? Explain.

b. Based on your plot, what (if anything) can you say about assumption A2? Explain.

c. Based on your plot, what (if anything) can you say about assumption A4? Explain.