1. Upload the "employment.csv" data set.

2. Build the plot to look at the relationship between the variables.

What will the dependent variable (outcome), what will be the independent variable (predictor)?

3. Perform linear regression analysis (fit a simple linear regression model between the variables).

Draw the best-fit regression line.

4. Check the main assumptions of the model, use the four main plots for checking:

Plot 1. Linearity of the data, independence of residuals

Plot 2. Normality of residuals using Q-Q plot

Plot 3. Constant variance of residuals

Plot 4. No influential outliers

5. Check the assumption "Normality of residuals" using histogram and normality tests; and "Zero mean of residuals".

Don't forget to look at the Q-Q plot from the previous question.

6. Obtain parameters of the regression (the intercept, the slope of the line, the 95% confidence intervals).

7. Obtain criteria for the model evaluation (Adjusted R-squared, AIC).

8. After checking all the assumptions, what conclusion can you make?

9.Repeat all the steps for the "bpa_age_data.csv" data set.

10.Repeat all the steps for the "employment_1.csv" data set.

11. Repeat all the steps for the "age.csv" data set.

Check list

	"employment.csv"	"bpa_age_data.csv"	"employment_1.csv"	"age.csv"
Assumptions after Linear regression:				
Plot 1: Linearity of the data, independence of residuals				
Plot 2: Normality of residuals +histogram + normality tests				
Zero mean of residuals				
Plot 3: Constant variance of residuals				
Plot 4: No influential outliers				
Parameters of the regression: - intercept (α) - slope of the line ($\beta 1$) -95% CI				
Criteria for the model evaluation: -Adjusted R^2 -AIC				
Conclusion and formula (if relevant):				