

Week 7 Linear Regression Exercises

Simple Regression

Research Question: Does the number of hours worked per week (*workweek*) predict family income (*income*)?

Using Polit2SetA data set, run a simple regression using Family Income (*income*) as the outcome variable (Y) and Number of Hours Worked per Week (*workweek*) as the independent variable (X). When conducting any regression analysis, the dependent (outcome) variable is always (Y) and is placed on the y-axis, and the independent (predictor) variable is always (X) and is placed on the x-axis. Please submit the questions and answers only, no SPSS output. You do not need a APA title or reference page. make sure to save your document with the correct title as directed in the submission area.

Follow these steps when using SPSS:

1. Open Polit2SetA data set.
2. Click on **Analyze**, then click on **Regression**, then **Linear**.
3. Move the dependent variable (*income*) in the box labeled "Dependent" by clicking the arrow button. The dependent variable is a continuous variable.
4. Move the independent variable (*workweek*) into the box labeled "Independent."
5. Click on the **Statistics** button (right side of box) and click on **Descriptives**, **Estimates**, **Confidence Interval** (should be 95%), and **Model Fit**, then click on **Continue**.
6. Click on **OK**.
7. Check your SPSS output.

Assignment: Through analysis of the SPSS output, answer the following questions. Make sure to place the number of the question next to your answer.

1. What is the total sample size?
378
2. What is the mean income and mean number of hours worked?
Mean income: USD 1,485.49
Mean number of hours worked: 33.52
3. What is the correlation coefficient between the outcome and predictor variables?
Is it significant? How would you describe the strength and direction of the relationship?

The correlation coefficient is 0.300. The coefficient is weak but positive hence the hours worked per week in current job and family income prior month have a weak but positive relationship.

4. What is the value of R squared (coefficient of determination)? Interpret the value.

The R squared is 0.090 meaning that 9% of the variability in the outcome data cannot be explained by the model

5. Interpret the standard error of the estimate? What information does this value provide to the researcher?

The standard error of the estimate (SEE) is USD 907.877. The SEE shows the estimate of the standard deviation of the error term

6. The model fit is determined by the ANOVA table results (F statistic = 37.226, 1,376 degrees of freedom, and the p value is .001). Based on these results, does the model fit the data? Briefly explain. (Hint: A significant finding indicates good model fit.)

The F statistic is greater than the f obtained from the table hence the model fits the data since the f is significant

7. Based on the coefficients, what is the value of the y-intercept (point at which the line of best fit crosses the y-axis)?

The y-intercept is 711.651

8. Based on the output, write out the regression equation for predicting family income.

Family income = $23.083 \times \text{hours worked per week} + 711.651$

9. Using the regression equation, what is the predicted monthly family income for women working 35 hours per week?

Family income = $(23.083 \times 35) + 711.651 = \text{USD } 1,519.556$

10. Using the regression equation, what is the predicted monthly family income for women working 20 hours per week?

Family income = $(23.083 \times 20) + 711.651 = \text{USD } 1,173.311$

Multiple Regression

Assignment: In this assignment we are trying to predict CES-D score (depression) in women. The research question is: How well do age, educational attainment, employment, abuse, and poor health predict depression?

Using Polit2SetC data set, run a multiple regression using CES-D Score (*cesd*) as the outcome variable (Y) and respondent's age (*age*), educational attainment (*educatn*), currently employed (*worknow*), number, types of abuse (*nabuse*), and poor health (*poorhlth*) as the independent variables (X). When conducting any regression analysis, the dependent (outcome) variable is always (Y) and is placed on the y-axis, and the independent (predictor) variable is always (X) and is placed on the x-axis.

Follow these steps when using SPSS:

1. Open Polit2SetC data set.
2. Click on **Analyze**, then click on **Regression**, then **Linear**.
3. Move the dependent variable, CES-D Score (*cesd*) into the box labeled "Dependent" by clicking on the arrow button. The dependent variable is a continuous variable.
4. Move the independent variables (*age*, *educatn*, *worknow*, and *poorhlth*) into the box labeled "Independent." This is the first block of variables to be entered into the analysis (block 1 of 1). Click on the bottom (top right of independent box), marked "Next"; this will give you another box to enter the next block of independent variables (block 2 of 2). Here you are to enter (*nabuse*). **Note:** Be sure the Method box states "Enter".
5. Click on the **Statistics** button (right side of box) and click on **Descriptives**, **Estimates**, **Confidence Interval** (should be 95%), **R square change**, and **Model Fit**, and then click on **Continue**.
6. Click on **OK**.
7. Check your SPSS output.

Assignment: (When answering all questions, use the data on the coefficients panel from Model 2).

1. Analyze the data from the SPSS output and write a paragraph summarizing the findings. (Use the example in the SPSS output file as a guide for your write-up.)
The correlation between CES-D score and Age at first birth, Educational attainment, current employment and poor health self rating are 0.040, -0.167, -0.213 and 0.285 respectively. CES-D score is therefore positively correlated with age at first birth, negatively correlated with educational attainment, negatively correlated with current employment and positively correlated with poor health self rating.
2. Which of the predictors were significant predictors in the model?

a **predictor** that has a low p-value is likely to be a **meaningful** addition to your model hence significant predictors were educational attainment, current employed, and poor health self-rating. Age at first birth had a p-value of 0.212 (>0.05) hence not a **significant** predictor of the model.

3. Which of the predictors was the most relevant predictor in the model?
The most relevant predictors had a p-value of 0.000 which included educational attainment, current employed, and poor health self-rating
4. Interpret the unstandardized coefficients for educational attainment and poor health.

Unstandardized coefficients are used to **interpret** the effect of each independent variable on the outcome. All other variables held constant, an increase of 1 unit in Educational attainment leads to a change in CES-D score by -2.851. All other variables held constant, an increase of 1 unit in Poor Health leads to a change in CES-D score by 11.614.

5. If you wanted to predict a woman's current CES-D score based on the analysis, what would the unstandardized regression equation be? Include unstandardized coefficients in the equation.

CES-D score = $0.125 * \text{Age at first birth} - 2.851 * \text{Educational attainment} - 3.762 * \text{Currently employed} + 11.614 * \text{Poor health self rating} + 21.325$