

# Assignment 02

## *Simple Linear Regression*

The file *teengamb.csv* contains data collected on a study of teenage gambling in Britain. The data consists of five variables and 47 observations. The variables are:

- **sex:** 0 = male; and 1 = female
- **status:** Socioeconomic status score based on parents' occupation
- **income:** Income in British pounds per week
- **verbal:** verbal score in words out of 12 correctly defined
- **gamble:** expenditure on gambling in British pounds (per year)

These source of these data is: Ide-Smith, S. G., & Lea, S. E. G. (1988). Gambling in young adolescents. *Journal of Gambling Behavior*, 4(2), 110–118.

Use the data in *teengamb.csv* to answer each of the following questions. In this homework assignment, you will be focusing on the relationship between income (predictor) and expenditures on gambling (outcome). This assignment is worth 14 points. Each question is worth 1 point unless otherwise noted.

1. Create a publication quality plot of the marginal distribution of expenditures on gambling. Be sure to format the figure according to APA (e.g., appropriate figure numbering and caption). Also be sure to appropriately size the figure so as not to take up too much page space. **(2pts.)**
2. Create a single (publication quality) table to present numerical summary information (descriptive statistics) for the marginal distributions of expenditures on gambling and income. Be sure to format the table according to APA (e.g., appropriate table numbering, caption, cell alignment, etc.). **(2pts.)**
3. Create a publication quality plot of the distribution of expenditures on gambling conditioned on income (i.e., a scatterplot). Be sure to format the figure according to APA (e.g., appropriate figure numbering and caption). Also be sure to appropriately size the figure so as not to take up too much page space. **(2pts.)**
4. Describe the relationship between income and expenditures on gambling. Be sure to comment on the form, direction and strength of the relationship. Also comment on any potential observations that deviate from following this relationship (unusual observations or clusters of observations). **(2pts.)**
5. Compute and report the Pearson correlation coefficient between income and expenditures on gambling.
6. Based on your answer to the Question #4, is the Pearson correlation coefficient an appropriate summary measure of the relationship? Explain.
7. Fit a regression model using income to predict expenditures on gambling (i.e., regress expenditure on gambling on income). Write the regression equation using Equation Editor (or some other program that correctly types mathematical expressions). Be sure the equation is labeled and numbered according to the APA format. **(2pts.)**
8. Interpret the value of the intercept from the regression equation using the context of the data.
9. Interpret the value of the slope from the regression equation using the context of the data.