Name:

Collaborators:

Instructions:

> You must submit your worksheet individually by end-of-class or end-of-day. Your name must exist in your worksheet and the names of your collaborators.

Worksheets are marked mostly on completion, and partially on correctness. It will be marked either pass or fail, there will no detailed feedback on worksheets, and no opportunities for revisions and make-up.

Assessing Residuals of a Linear Model

1. A Best fit Line

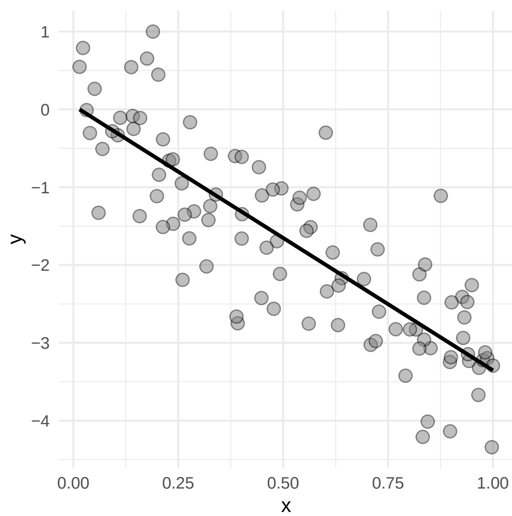

Figure 1: Scatterplot of variables x and y. The black line is the best fit linear model of the data.

a. The correlation coefficient of the two numerical variables x and y is $r = -0.8355$. The standard deviation of x is $s_x = 0.3061$. The standard deviation of y is $s_y = 1.2469$. Compute and interpret the slope estimate of the linear fit.

b. The mean of x is $\bar{x} = 0.5240$. The mean of y is $\bar{y} = -1.7337$. Apply the point-slope equation using the means $(\bar{x}, \bar{y})$ and the slope to write the equation of the red line shown in the scatter plot. What is the equation of the linear model? What is the intercept of the linear model?

2. Residual Analysis

Continued from Problem (1).

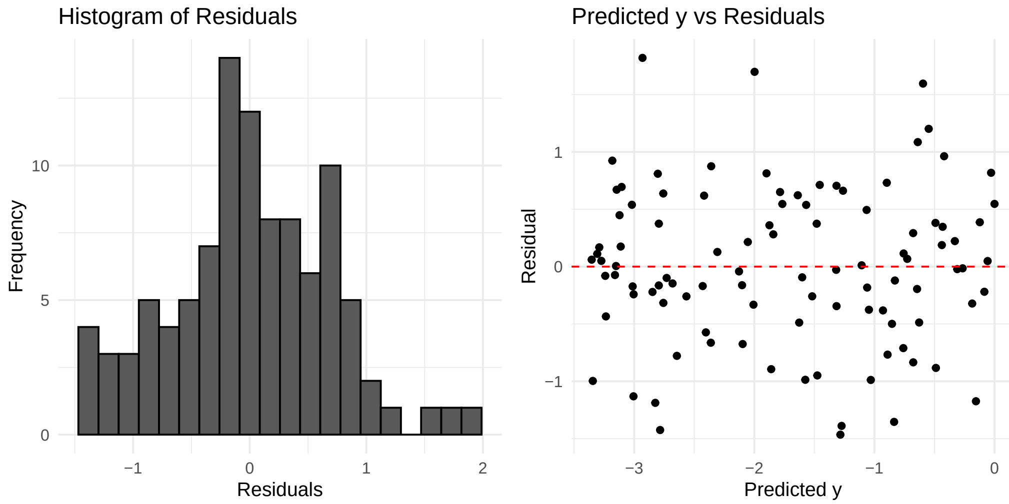

Figure 2: Histogram and scatterplot of the residuals.

a. The SST of the model is 153.9204 and the SSE of the model is 46.48605. Compute the coefficient of determination (R^2). Interpret this value in the context of the model.

b. Consider the histogram of residuals and the scatterplot of residuals. Explain why a linear regression model is appropriate for the data it represents.

References

1. Speegle, Darrin and Clair, Bryan (2021) Probability, statistics, and data: A fresh approach using r, Chapman; Hall/CRC.
2. Diez DM, Barr CD, Çetinkaya-Rundel M (2012) OpenIntro statistics, OpenIntro.