STA130 Tutorial 6
Xuanqi Henry Wei

Dear Miss Lang,

Thank you for your interest in using linear regression for the analysis of our sidewalk chalk study. In our study, we are comparing the enjoyment children receive from our sidewalk chalk versus our competitor's, Mr. Bingbong. We can use linear regression to see if there is a significant difference in enjoyment between the two brands of sidewalk chalk. To do this, we will use the enjoyment rating as our outcome or dependent variable and the brand of sidewalk chalk as our explanatory or independent variable.

Linear regression is a statistical method used to examine the relationship between two variables, often used to predict the value of one variable based on the value of the other. It can be used to identify a linear association or an approximately linear relationship between the two variables. A linear association means that as one variable increases, the other variable also increases or decreases at a constant rate. An approximately linear relationship means that the relationship is not perfect but can still be modeled with a straight line.

The simple linear regression model we will use for our analysis is:

$$\text{enjoyment} = \beta_0 + \beta_1 * x_1 + \varepsilon$$

β_0 is the intercept, which represents the expected enjoyment rating when the chalk brand is zero (the enjoyment score of the competitor company). β_1 is the slope or regression coefficient, which represents the expected change in enjoyment rating when the chalk brand changes by one unit. ε is the noise or error term, which represents the random variation in enjoyment ratings that is not explained by the chalk brand. x_1 is an indicator to tell the model which company is being calculated, which in this circumstance, 0 represent the competitor company and 1 represent our company.

We can use hypothesis testing to determine if the difference in enjoyment ratings between our sidewalk chalk and Mr. Bingbong's is significant. We will test the null hypothesis that there is no difference in enjoyment ratings between the two brands of sidewalk chalk versus the alternative hypothesis that there is a significant difference.

So, to summarize, it is appropriate to use linear regression for our analysis since we are examining the relationship between two continuous variables. I hope this explanation of linear regression and its appropriateness for our study is helpful.

Sincerely,
Henry Wei