**1. Consider the following data set that describes the relationship between “velocity” of**

**an enzymatic reaction () and the substrate concentration ().** **You are asked to investigate whether this linear transformation results in a good or poor fit by doing the following steps:**

1. Generate a scatterplot of vs . Comment on the shape.

A graph with blue dots

Description automatically generated

This plot doesn’t pass the “eyeball test” for linear regression, as its clear that the data trend quickly departs from linear behavior around a concentration level of 0.2.

1. Define new variables for and in SAS and generate a scatterplot.

A graph with numbers and dots

Description automatically generated

Now linear regression seems much more reasonable, though there are two significant outliers at that could affect the appropriateness of the fit.

1. Is the distribution of different than ? Are there any points that may be more influential in determining the fit?

A graph of a distribution of concentration

Description automatically generatedA graph of a distribution of a number of curves

Description automatically generated with medium confidence

Their distributions appear similar, as evidenced by the histograms. Looking at the scatter plots above in parts a and b, it’s evident that and are very influential points.

1. Determine the least squares regression line for vs . Save the residuals and predicted values. Does the residual plot suggest any problems?

A graph with blue dots

Description automatically generatedA screenshot of a computer

Description automatically generated

The residuals increase in a conical fashion as increases, so it appears that the assumption of constant variance for linear regression is violated.

1. Convert this regression line back into the original nonlinear model and plot the predicted curve on a scatterplot of vs . Comment on the fit.

A graph with numbers and lines

Description automatically generated

A regression line seems appropriate for concentration levels up to 0.21, but quickly becomes nonviable for values beyond that threshold.

**2. Typographical errors. Shown below are the number of galleys for a manuscript () and**

**the total dollar cost of correcting typographical errors () in a random sample of recent orders handled by a firm specializing in technical manuscripts. Since involves variable costs only, an analyst wished to determine whether regression-through-the-origin model (4.10) is appropriate for studying the relation between the two variables.**

1. Fit regression model (4.10) and state the estimated regression function.

A screenshot of a computer

Description automatically generated

The estimated regression function is .

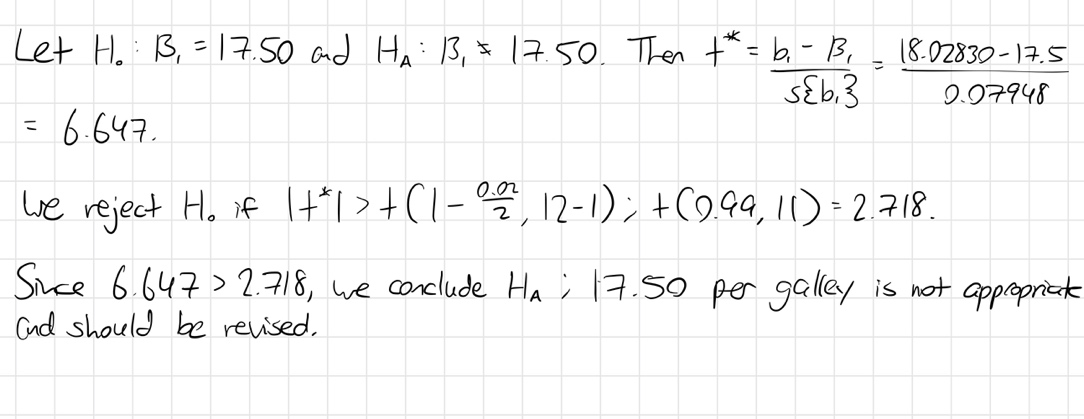
1. Plot the estimated regression function and the data. Does a linear regression function through the origin appear to provide a good fit here? Comment.

**A graph with blue lines and white text

Description automatically generated**

Yes, the data exhibits very linear behavior, as evidenced by the line-of-fit.

1. In estimating costs of handling prospective orders, management has used a standard of $17.50 per galley for the cost of correcting typographical errors. Test whether or not this standard should be revised; use = .02. State the alternatives, decision rule, and conclusion.



1. Obtain a prediction interval for the correction cost on a forthcoming job involving 10 galleys. Use a confidence coefficient of 98 percent.

**A math equations on a white sheet

Description automatically generated**

**3. When the predictor variable is so coded that and the normal error regression model (2.1) applies, are and independent? Are the joint confidence intervals for and then independent?**

When , and ’s value is indeterminate without knowing any more information, but we know is needed to find , so we know and are not independent because influences .

Likewise, the joint confidence intervals for and require and , respectively, which both require knowing the . Thus these confidence intervals are not independent either.

**4. Derive the formula for given in Table 4.1 for linear regression through the origin.**

We desire by Table 4.1. For regression through the origin, we know . Then . Because is fixed, by definition of variance we have .

Since for regression through the origin, we know . Because is an unbiased estimator of , we also have . So, we conclude,

**5. Set up the matrix and vector for each of the following regression models (assume .**

**6. (SAS Exercise) Use the brand preference data described in KNNL Problem 6.5. Run the linear regression with moisture and sweetness of the product as the explanatory variables and degree of liking as the response variable.**

1. Summarize the regression results by giving the fitted regression equation and .

A screenshot of a table

Description automatically generated

and the fitted regression equation is .

1. State the results of the significance test for the null hypothesis that the two regression coefficients for the explanatory variables are *all* zero (give null and alternative hypotheses, test statistic with degrees of freedom, p-value, and a brief conclusion in words).

A table with numbers and letters

Description automatically generated

The null and alternative hypotheses are, ; and , respectively. The test stat is with 13 degrees of freedom and the -value is . Hence, we conclude , that is, there is a relationship between liking the product and the two associated predictors: moisture and sweetness.

1. Describe the results of the hypothesis tests for the individual regression coefficients (give null and alternative hypotheses, test statistic with degrees of freedom, p-value, and a brief conclusion in words).

A screenshot of a table

Description automatically generated

For , and . Test stat is with 13 degrees of freedom and the -value is . We conclude , so there is statistical evidence to suggest that the intercept is not equal to 0.

For , and . Test stat is with 13 degrees of freedom and the -value is . We conclude , so there is statistical evidence to suggest that there is a linear relationship between moisture and degree of liking.

For , and . Test stat is with 13 degrees of freedom and the -value is . We conclude , so there is statistical evidence to suggest that there is a linear relationship between sweetness and degree of liking.

1. Give separate 95% confidence intervals for the regression coefficients of sweetness and moisture. What is the relationship between these confidence intervals and the above hypothesis results?

A screenshot of a table

Description automatically generated

For moisture, . Then , , and . So .

For sweetness, . Then , , and . So .

Neither confidence interval contains , which supports the conclusion from the above hypothesis test that neither nor are equal to .