**IBEHS 4C03: STATISTICAL METHODS IN BIOMEDICAL ENGINEERING**

**ASSIGNMENT #4**

**Topics:**

Simple Linear Regression

**Total marks: 100**

**Submission Instructions:** Electronic submission to the IBEHS 4C03 Avenue to Learn Assignment 3 folder.

**Assignment Submission Format**:

File Names: StudentLastNameFirstNameAssignment#.doc/pdf/jpynb etc.

Both your name and student numbers should appear at the top of the document. If separate documents are submitted, then you should submit the answers to the questions in a document including the plots you made in python. Regardless, the python file used to generate

the plots and any provided answers should also be submitted with any submission. Submit a single zip file per assignment.

**Problem 1: [50 points]**

A study was conducted to determine the relationship between birthweight (g/100) and estriol level (mg/24 hr) in pregnant women (Birthweight\_Estriol.csv). Use the data in the CSV file to answer the following questions.

1. What are the assumptions of a least-squares regression model?
2. Review each assumption based on the data and comment on the fit of the data using a least-squares regression model.
3. Construct a scatter plot of estriol level versus birthweight. Does the graph suggest anything about the nature of the relationship between these variables?
4. Estimate the covariance and correlation of estriol level in pregnant women and infant birthweight for this population of infants. Does infant birthweight tend to increase or decrease as estriol level increases?
5. Test the null hypothesis that the population correlation equals zero. What do you conclude?
6. Using birthweight as the response and estriol level as the explanatory variable, compute the least-squares regression line. Show the work by hand and verify in python. Interpret the estimated slope and y-intercept of the line; what do they mean in words?
7. At the 0.05 level of significance, test the null hypothesis that the true population slope is equal to zero. What do you conclude?
8. What is the estimated mean birthweight for the population of infants whose mother’s estriol level is 31 mg/24 hr?
9. Does the least-squares regression model seem to fit the observed data? Comment on the coefficient of determination and standard error of the regression.

**Problem 2: [20 points]**

Your coworker has been analyzing data from a gas (methane) furnace. They say that they’ve built a model to predict the output CO2 rate of the process (which you need to track for environmental reasons) based on the input gas flow rate. They’ve provided you with their ANOVA table for a second opinion on whether or not the model is suitable.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Type of Variance | Distance | Degrees of Freedom | Sum of Squares | Mean Squares |
| Regression | 𝑦̂𝑖 − 𝑦̅ | 1 | 2614.9 | 2614.9 |
| Error | 𝑦𝑖 − 𝑦̂𝑖 | 𝑛 − 2 | 409.9 | 1.39 |
| Total | 𝑦𝑖 − 𝑦̅ | 𝑛 - 1 | 3024.8 |  |

1. Calculate the R2 value from the table and interpret its meaning.
2. Calculate the standard error the regression from the table and interpret its meaning.
3. Would you recommend use of their model? Why or why not?

# Problem 3: [30 points]

This question utilizes the ReactorData.csv file. The data contains information from a lab-scale bioreactor. The outcome of interest is the yield of the bioreactor (g). Each batch was performed with a different combination of temperature (measured in C), mixing speed (measured in 100 RPM) and reactor type (one with baffles and one without). For the questions where you need to fit and interpret a regression model, you can assume all of the assumptions hold (you don’t need to verify them).

1. Build a scatterplot of temperature (on the x-axis) vs. reactor yield (on the y-axis). Describe the relationship you see in the plot.
2. Determine the model coefficients for the least squares model 𝑦𝑖 = 𝑏0 +𝑏1 ∗ 𝑇𝑒𝑚𝑝+𝑒𝑖. Don’t forget to include the units when reporting the values.
3. Interpret the 𝑅2 value of the model. Do you think the performance of the model could be improved? If so, how?
4. Provide a confidence interval for the predictions from this model. Discuss what this confidence interval means in the context of the example
5. Provide a prediction interval for the predictions from this model. Discuss what this prediction interval means in the context of the example. Based on this interval, is this a good model?