**Regression Analysis on Excel**

You will perform a regression analysis on data provided in an Excel spreadsheet ‘Regression Analysis Assignment Data.xlsx’. **Please read through the entire assignment before you start.**

The spreadsheet includes 41 measurements of radiation counts through increasing thicknesses of each of two different metals, aluminum (Al) and lead (Pb). The units of measurement are ‘counts per minute’ (cpm, or min-1) – that is, how many events of ionizing radiation were detected after the source radiation passed through the given thickness of material. This is not an SI measure, but is widely used in studies of radioactive particle detection. There are various techniques for inferring the disintegration rate at the source, and the effects in terms of the dose rate, but those do not concern us here. In fact, we don’t even know what the sources of the radiation are, or whether they are the same!

Your job is to calculate the absorption coefficients for each of these two materials, for the given sources of radiation. This will tell us how much more radiation is blocked or absorbed as we increase the thickness of each material. But you have to be able to demonstrate to an audience that this is meaningful (i.e. statistically significant).

1. Create a **scatter chart of the raw data** in a new worksheet. *(15 marks)*
   1. Show all of the values for each of the two dependent variables against the value of the independent variable. *(6 marks)*
   2. Assuming that the counts are accurate within ±5%, add ‘error bars’ to the values for each variable at each thickness. *(1 mark)*
   3. Add a linear trend line for each series, and make sure the equation and the coefficient of determination (R2) are both shown for each trendline. *(2 marks)*
   4. Add an exponential trend line for each series, with the equation and the coefficient of determination showing for each. *(1 mark)*
   5. With the use of appropriate formatting (titles, labels, axes, location, colour, bold face, line type and thickness, etc.), improve the chart to show what you can about these two series of data. *(3 marks)*
   6. With the use of appropriate formatting, highlight the trend line and related labels that you think better describe each series. *(2 marks)*

*As we can see, we need to manipulate this data further!*

1. **Transform the data** provided. *(5 marks)*
   1. In a new column for each material, use the LN function to calculate the natural logarithm (exponent base *e*) for each counts value – we can call this **Ln(counts)**. *(4 marks)*
   2. Use appropriate formatting (fill colour, borders, font colour/style/size), ensure the new columns are clearly marked as not original data. *(1 mark)*
2. Create a **scatter chart of the transformed data** in another new worksheet. *(5 marks)*
   1. Copy the earlier chart. For each of the two data series, change the dependent variable source from the raw data columns to the transformed data column. *(3 marks)*
   2. Examine the trend lines; keep only the best one per data series. *(1 mark)*
   3. Update the titles, axis labels and so forth to accurately represent the contents of the new chart. *(1 mark)*

*Examine the coefficients in the equations for the trend lines for each variable in each chart, and their coefficients of determination. You should see that the coefficients for the exponential trend line on the first chart are exactly the same as those for the linear trend line on the second chart. Think about why that is.*

1. Using the Data Analysis tool in Excel, **perform four regression analyses**. *(10 marks)*
   1. Perform a regression analysis for each dependent variable, with residual, probability and line fit plots: Al raw, Pb raw, Al transformed, and Pb transformed. Include each on a clearly-named separate worksheet in your spreadsheet. *(6 marks)*
   2. Fill in the following four tables *(4 marks)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Aluminum Raw** | Regression Equation |  | | |
|  | R-squared value |  | | |
|  | F-statistic |  | | |
|  | P-value |  | | |
|  | Is there a statistically significant relationship between these two variables? How do you know? | | |  |
|  | What do the plots tell you? | |  | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lead Raw** | Regression Equation |  | | |
|  | R-squared value |  | | |
|  | F-statistic |  | | |
|  | P-value |  | | |
|  | Is there a statistically significant relationship between these two variables? How do you know? | | |  |
|  | What do the plots tell you? | |  | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Aluminum Transformed** | Regression Equation |  | | |
|  | R-squared value |  | | |
|  | F-statistic |  | | |
|  | P-value |  | | |
|  | Is there a statistically significant relationship between these two variables? How do you know? | | |  |
|  | What do the plots tell you? | |  | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lead Transformed** | Regression Equation |  | | |
|  | R-squared value |  | | |
|  | F-statistic |  | | |
|  | P-value |  | | |
|  | Is there a statistically significant relationship between these two variables? How do you know? | | |  |
|  | What do the plots tell you? | |  | |

1. **Share the absorption coefficients for each material and radiation**. *(5 marks)* The absorption coefficient is the negative of the slope of the transformed data, because the reduction in the counts reflects an amount of absorption of radiation.
   1. Fill in the following chart. *(5 marks)*

|  |  |  |
| --- | --- | --- |
| **Material** | **Absorption Coefficient** |  |
| Aluminum |  | cm-1 |
| Lead |  | cm-1 |

1. Answer the following questions. *(10 pts)*

Note: Absorption is defined by the *lowest* number of counts.

* 1. Which material has the greatest overall absorption? How do you know? *(1 mark)*
  2. Which material has the greatest absorption rate change per thickness? Why? *(1 mark)*
  3. Which model (raw or transformed data) has better prediction rate for aluminum? Why? *(1 mark)*
  4. Which model (raw or transformed data) has better prediction rate for lead? Why? *(1 mark)*
  5. At approximately what thickness will the two metals have equal absorptions? Why? *(3 marks)*
  6. If you wanted approximately 200 counts of radiation, which material should you use and at what thickness? How do you know? *(3 marks)*

**Your submissions should include a filled out copy of this word document and an Excel spreadsheet that includes your modifications and output along with the original raw data.**