**HOMEWORK 4**

*Natural Log Transformations*

Reading: This assignment focuses on content from your textbook, *STAT2: Building Models for a World of Data*, Chapter 1 Section 4. Read this section of your textbook.

Notes:

* For questions requiring you to use JMP, you must provide a copy of your output at the end of your assignment or embedded within your assignment. No credit will be given if you do not include your output, even if your answer is correct.
* Recall that you can download JMP to your personal computer for free. See the JMP information posted on Canvas. Problems due to not getting JMP working will not allow you to submit your assignment late. Please plan to work ahead and email your instructor questions if needed.
* You are required to use your own words in answering all homework questions. You cannot copy information from the book or other sources.
* Round all numbers to 2 decimal places unless otherwise specified.
* For all questions requiring calculations, show your work in order to receive credit.

1. 1.29a and b – Caterpillar waste versus mass. (.jmp file posted on Canvas). It is recommended that you use Graph Builder in JMP for these first 2 graphs.
   1. Include a copy of your JMP output and interpret the scatterplot. Do not forget to discuss form, direction, strength, and presence/absence of outliers.

Chart, scatter chart

Description automatically generated

There is a positive relationship between mass and wetfrass. It looks like there is some curvature as well as increasing variability for mass from 9 to 12, where the outliers are found.

* 1. Include a copy of your JMP output and interpret the scatterplot. Do not forget to discuss form, direction, strength, and presence/absence of outliers.

IMPORTANT NOTE: In the JMP file, you will see variables labeled as log’s (e.g., LogWetFrass). Those are common log’s. We are using natural log’s in our course. You will need to create the natural log’s using a formula. See the bottom of Page 7 in the JMP Guide for instructions.

Chart, scatter chart

Description automatically generated

There is a strong positive linear relation between the transformed variables: ln(Mass) and ln(WetFrass). There are many more outliers throughout the plot compared to the first graph though.

1. More Caterpillar waste versus mass questions not in your book.
   1. Fit 4 models to predict WetFrass; (a) simple linear regression model, (b) log-linear model, (c) linear-log model, and (d) log-log model. Consider the conditions of each model. In each column, rank the conditions on best (1) to worst (4).

For example, if linearity looks best for the simple linear model, rank that as a “1,” and if it looks worst for the log-linear model, rank that as a “4.” There may be some close calls. For example, if both the simple linear model and linear-log model look the best, you can rank them both as “1/2.”

You do NOT need to explain and you do NOT need to include JMP output.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Linearity | Constant variance | Normality |
| Simple linear model | **3** | **4** | **4** |
| Log-linear model | **4** | **2** | **1** |
| Linear-log model | **2** | **3** | **3** |
| Log-log model | **1** | **1** | **2** |

* 1. Neither transformation is perfect, but which would be most appropriate? Explain by commenting on the appropriate plots and conditions.

The log-log model will be the most appropriate. All following assumptions are met with this model. There is strong linearity in this model. Although the log-linear model shows better normality, there is reasonable normality in log-log. There is a bell-shaped histogram in the residual by predicted plot and the QQ plot shows linearity. There is constant variance as the residual plot has no fan pattern. Log-log would be the best model to use to most accurately predict WetFrass.

* 1. Suppose we ignored the conditions and made some predictions. For each transformation, what is the predicted WetFrass for caterpillars with a mass of 10 grams?

1. Log-linear model:

Predicted ln(WetFrass) = -4.889483 + 0.7001119\*Mass

= -4.889483 + 0.7001119\*10

Predicted ln(WetFrass) = 2.11

Predicted wetfrass = ln y = 2.11 = 8.26

1. Linear-log model:

Predicted WetFrass = 0.8837407 + 0.2191506\*ln(Mass)

= 0.8837407 + 0.2191506\*ln(10)

Predicted wetfrass = 1.39

1. Log-log model:

Predicted ln(WetFrass) = -1.700717 + 1.0536139\*ln(Mass)

= -1.700717 + 1.0536139\*ln(10)

Predicted ln(wetfrass) = 0.73

Predicted wetfrass = ln y = 0.73 = 2.07

* 1. Compare your 3 predictions in the previous question to the graph you created in Question 1a. Which model appears to have given the best prediction?

The linear log model seems to be the best when comparing to the graph in 1a. At mass 10, there is a wetfrass value of around 1.5.

* 1. Suppose we ignored the conditions and do some interpretations. For each transformation, interpret the value of the slope (or a function of the slope).

1. Log-linear model:

Eb­­­­­1= e0.7001119 =2.0139780

When mass increases by 1 unit, we predict the median wetfrass to change by a factor of 2.01.

1. Linear-log model:

b1ln(P)= 0.2191506\*ln(1.1)= 0.02088728

When mass increases by 10% we predict wetfrass to increase by 0.02

1. Log-log model:

(P)b1=(1.1)1.0536139=1.10563533

When mass increases by 10%, we predict the median wetfrass to change by a factor of 1.11.