**HOMEWORK 4**

*Conditions, Natural Log Transformations, and Inference*

Reading: This assignment focuses on content from your textbook, *STAT2:Modeling with Regression and ANOVA*, Sections 1.2-1.4, 2.1-2.2. Read these sections of your textbook, if you haven’t already.

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.
* Round all numbers to **3 digit places** unless otherwise specified. This means keeping 3 digits staring from the first nonzero digits. e.g. 14123 is rounded to 14100, 0.00056789 is rounded to 0.000568, 0.5022 is rounded to 0.502.
* All answers will be numbers so we can quickly grade the assignment. Please notice the “be able to:” questions. You do not need to answer these on the homework, but you will be expected to able to answer them on the upcoming exam.

You will need the textbooks.jmp and sparrow.jmp data sets.

Part 1: Consider the data in textbooks.jmp. This contains prices charged for various textbooks. uclaNew is the price charged by the UCLA bookstore; amazNew is the price charged the same semester by Amazon.com.

1. Create a scatterplot where X = amazNew and Y = uclaNew. Also fit a linear regression model predicting the UCLA price (uclaNew) using the Amazon price (amazNew) then look at the residual vs predicted value plot.

Using these two plots, assess the regression conditions:

a: linearity (yes / no)

b: equal variance (yes / no)

c: normality (yes / no)

2. Based on this (and only this) assessment of the regression conditions should you:

Multiple choice: Leave variables untransformed / Transform only X (amazNew) / Transform only Y (uclaNew) / Transform both variables

Note: be able to explain your assessments of regression conditions and support your choice of transformation.

3. Can you use either of these two plots to assess the assumption of independence?

Your answer: Yes/No

Note: be able to explain why or why not.

Natural log transform the uclaNew variable. Fit a linear regression predicting log uclaNew using the Amazon price (amazNew). Don’t worry if this regression does not satisfy the usual regression conditions. What are the intercept and slope coefficients of this regression?

4. Intercept:

5. Slope:

6. You are comparing prices of two textbooks. On Amazon, textbook A costs $25 more than textbook B. Fill in the blank: The predicted UCLA price of textbook A is \_\_\_\_ percent more than textbook B.

Your answer:

You decide it would be better to natural log transform both variables, i.e. both uclaNew and amazNew. Fit the linear regression predicting log uclaNew using X=log amazNew. What are the intercept and slope coefficients of this regression?

7. intercept

8. slope

9. Again, you want to compare prices of two textbooks, A and B, where on Amazon, textbook A costs 10% more than textbook B. The predicted UCLA price of textbook A is \_\_\_\_ times that of textbook B.

Your answer:

10. You want to compare prices of two different textbooks, C and D. All you know is that on Amazon, textbook C costs 25$ more than textbook D. Using the log-log model, do you have sufficient information to say anything about the predicted UCLA bookstore prices?

Your answer: (yes / no)

Note: be able to explain why or why not.

Part 2: Questions 2.18 and parts of 2.25 from the book (Sparrows). Simple linear regression predicting Y = weight (in grams) from wing length (in mm).

11. (2.18 part a) Report the p-value for the two-sided test of slope = 0

p-value:

Note: be able to write a one-sentence conclusion based on this p-value.

(2.18 part b) Report the 95% confidence interval for the slope

12. lower bound:

13. upper bound:

Note: be able to write a sentence with an appropriate interpretation of the slope.

14. (2.18 part c) Does your 95% confidence interval include 0?

Yes / No

Note: be able to explain how including (or not including) 0 relates to the p-value for the test of slope = 0.

16. (2.25 part b) Percent variation:

(2.25 part c) ANOVA table: Only report two numbers from that table:

17. MS error:

18.Df error:

19. (2.25 part d) Compare the square root of the F statistic and the T statistic testing whether the slope = 0.

Multiple choice: T statistic is larger / they are the same / T statistic is smaller

Note: be able to explain your answer

Part 3: New data set. Details not relevant. Here is an ANOVA table with some values filled in:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source | degrees of freedom | sum-of-squares | mean square | F statistic |
| Model | 1 |  | 125.1 |  |
| Error |  | 10.3 |  |  |
| Total | 52 | 135.4 |  |  |

20. Calculate the error degrees of freedom:

Your answer:

21. Calculate the Model sum-of-squares:

Your answer:

22. Calculate the F statistic. There is only one, for the Model:

Your answer:

23. Upload your JMP output as pdf, word or html in the question.