**HOMEWORK 6**

*Multiple Regression and Interactions*

Reading: This assignment focuses on content from your textbook, *STAT2: Modeling with Regression and ANOVA*, Chapter 3, Sections 1-4. Read these sections of your textbook.

Notes:

* For questions requiring you to use JMP, you must provide a copy of your output. 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 must use your own words to answer all homework questions. You cannot copy information from the book or other sources.
* Round to 2 decimal places.
* For all questions requiring calculations, show your work to receive credit.

1. **Fruit Flies.** Answer the following questions using the FruitFlies dataset. The study includes 125 male fruit flies. We will focus on the following variables:

* Longevity (response variable; lifetime of a fruit fly in days)
* Thorax (length of thorax in mm)
* Type:
  + The male fruit flies were either paired with a female pregnant fruit fly (pregnant), paired with a female virgin fruit fly (virgin), or not paired with a female fruit fly (none).
  + Use “none” as the reference/baseline group.
  1. Create a graph using Graph Builder to appropriately display the 3 variables.

Chart

Description automatically generated

* 1. Based on the graph, does there appear to be an interaction between the thorax length and type? Explain.

Yes, there seems to be an interaction between the thorax length and type. For the type of none (our baseline), there is a strong positive linear relationship between thorax and longevity. Same applies to type of pregnant and virgin.

* 1. Fit an additive multiple regression model using thorax and type to predict longevity. Include the following JMP output:
     1. ANOVA
     2. Parameter Estimates with confidence intervals for the slopes
     3. Graphs to check conditions (residual by predicted plot and normal quantile plot)

Hints:

* Make sure to create indicator variables for the pregnant and virgin categories.
* Use Fit Model.
* Chart, scatter chart

  Description automatically generatedTable

  Description automatically generatedAn additive model means the analysis should have no interactions (no cross-products).

Graphical user interface, chart

Description automatically generated

* 1. What is the multiple least squares regression equation?

y^ = -56.521 + 143.638\*x + -13.349(virgin) + 3.450(pregnant)

where y^ = predicted longevity and x = thorax

A picture containing text

Description automatically generated

* 1. Check to see if the 4 conditions for regression are met for the additive model.

- Linearity: There is a linear relationship between the predictor and response variable

- Independence: met because there is a random sample.

- Constant variance: The data is plotted in the shape of a fan in the Residual by Predicted Plot so the condition is not met.

- Normality of errors: The data falls within the range of the boundaries suggested by JMP. This condition is met.

Chart, histogram

Description automatically generated

* 1. Interpret all 3 slopes from the additive model.

Graphical user interface, text, table

Description automatically generated

Virgin: The slope is -13.349. The predicted longevity in days for when the male fruit fly is paired with a virgin female fruit fly is 13.349 days lower than the male fruit flies that were not paired with a female fruit fly when thorax length is held constant

Pregnant: The slope is 3.450. The predicted longevity in days for when the male fruit fly is paired with a pregnant female fruit fly is 3.450 days higher than the male fruit flies that were not paired with a female fruit fly when thorax length is held constant.

Thorax: The slope is 143.639. There is a predicted 143.638 day increase in longevity for the male fruit fly for every unit increase in length of the thorax in mm when type is held constant.

* 1. Interpret all 3 confidence intervals for slopes from the additive model.

Thorax: We are 95% confident that a predicted longevity value will be between 117.774 to 169.501 when type is held constant.

Virgin: We are 95% confident that when a male fruit fly is paired with a virgin female fruit fly, the male’s longevity in days will be between 18.804 to 7.894 days less than the male fruit flies that were not paired with a female fruit fly when thorax length is held constant.

Pregnant: We are 95% confident that when a male fruit fly is paired with a pregnant female fruit fly, the male’s longevity in days will be between 2.012 to 0 days less or 0 to 8.912 days more than the male fruit flies that were not paired with a female fruit fly when thorax length is held constant.

* 1. What is the predicted longevity for male fruit flies with a thorax length of 0.8 mm and with a pregnant female based on the additive model?

y^ = -56.521 + 143.638\*(0.8) + -3.450

= 54.9394 days

* 1. Conduct a hypothesis test for the overall model fit for your additive model. Make sure to show all steps, including your hypothesis statements, test statistic, *p*-value, and conclusion.

H0: β1 = β2 = β3 = 0

Ha: at least 1 βi ≠ 0 where i = 1,2,3

Test statistic = 10.99

p-value = <0.0001

Conclusion: There is strong evidence of a linear relationship between Longevity and Thorax for the population of all None, Virgin and Pregnant types.

1. **Nursing.** Answer the following questions using the Nursing dataset. Data was collected on nursing homes in New Mexico. We will focus on the following variables:

* Beds (response variable; number of beds in the nursing home)
* AllPatientDays (total number of days all patients spent in the nursing home per year in 100s)
* Rural:
  + The nursing home is either in a rural area or a non-rural area
  + Use “non-rural” as the reference/baseline group.
  1. Create a graph using Graph Builder to appropriately display the 3 variables.

Chart

Description automatically generated

* 1. Based on the graph, does there appear to be an interaction between all patient days and rural? Explain.

Yes, there seems to be an interaction between AllPatientDays and rural. There is a strong positive linear relationship between AllPatientDays and beds for the a rural and non-rural setting.

* 1. Fit a multiple regression model with an interaction between all patient days and rural to predict the number of beds. Include the following JMP output:
     1. Parameter Estimates

Hints:

* Make sure to create indicator variables for rural.
* Use Fit Model.
* Table

  Description automatically generatedInclude a cross-product in the analysis.

Chart

Description automatically generated

Chart

Description automatically generated

* 1. What is the multiple least squares regression equation?

y^ = 10.843 + 0.312\*x + x\*(rural \* -0.0594) + 7.639\*rural

where y^ = predicted beds and x = all patient days

Graphical user interface, text, application

Description automatically generated

* 1. What is the predicted number of beds for a nursing home with 385 hundred all-patient days in a non-rural location based on the interaction model?

y^ = 10.843 + 0.312\*(385)

= 130.963 beds

* 1. Conduct a hypothesis test to see whether there is evidence of an interaction. Make sure to show all steps, including your hypothesis statements, test statistic, *p*-value, and conclusion.

H0: B1 = B2 = 0

Ha: at least 1 Bi ≠ 0 where i = 1,2

Test statistic = 5.21

p-value = <0.0001

Conclusion: There is strong evidence of a linear relationship between Beds and AllPatientDays for the population of non-rural and rural types.