STAT 210
Applied Statistics and Data Analysis
Second Exam

November 26, 2022

This exam is open notes and open book but not open internet. You are not allowed to surf the internet or look for answers to the questions

You are reminded to adhere to the academic integrity code established at KAUST.

Show complete solutions to get full credit. Label your graphs appropriately

Please, do not submit zip files and identify the files you submit with your surname

---

For this exam, we will use the data in the file `dragons.txt`. Read the data onto a data frame. There are ten variables in the set:

- `height`, the height in m,
- `length`, the length from head to tail in m,
- `weight`, the weight in tons,
- `wing.ln`, the average length for the wings in m,
- `leg.ht`, the average length for the legs in m,
- `wing.span`, the distance between the tips of the outstretched wings,
- `sp`, the species with two values, `black` and `gold`,
- `age`, in years
- `strength`, the strength index for the dragon, and
- `firepwr`, a combined measure of the caloric power, size and duration of the fire breath.

Question 1 (30 points)

In this question, you have to explore the relationship between the variables `strength` and `wing.span`.

(i) (2.5 pts) Graph a scatterplot of `strength` as a function of `wing.span`. Add the regression line for these variables and comment.

(ii) (5 pts) Fit a simple regression model for these variables and print the summary table. What is the R^2 for this model? Write down the equation for the model and give an interpretation of the parameters. Predict the strength of a dragon with a wingspan of 60 m. and include a prediction interval.

(iii) (5 pts) Use graphical methods and tests to check the assumptions on which the model is based. What are your conclusions?

(iv) (10 pts) There are two species of dragons in the file, black and gold, and this characteristic is available in the categorical variable `sp`. We want to add this variable to the regression model. If the variable was not read as a `factor`, transform it before you continue. Fit a model that includes the previous variable, the new variable, and the interaction between the two. Using a critical value for α of 0.05 and starting with the complete model, select a minimal adequate model.

(v) (7.5 pts) Check the assumptions for the final model. Compare the adjusted R^2 with the previous model. Write down the equation for the regression model and predict the value of the strength for black and gold dragons of height 60 m, including prediction intervals. Compare with the previous prediction and comment.

Question 2 (30 points)

In this question, we want to explore the relation between the weight (`weight`) and the length (`length`) of dragons.

(i) (15 pts) Start by plotting a graph of `weight` as a function of `length`. Fit a simple regression model and add a regression line to the plot. What is the R^2 for this model? Write down an equation for the model and give an interpretation of the parameters. Give a prediction of the weight of a dragon with a length of 58 m, including a confidence interval. State explicitly the assumptions on which this model is based. Check whether these assumptions are satisfied. Use the function `residualPlots` in the `car` package and interpret the graphs and results of the hypotheses test. What do these results suggest?

(ii) (15 pts) Fit a new model, including the term(s) suggested by the tests in (i), if any. Look at the summary table. What is the adjusted R^2 for this model? Check whether the assumptions for linear regression are satisfied for the new model. Write an equation for the model. Give a prediction of the weight of a dragon with a length of 58 m, including a confidence interval, and compare it with the result in part (i).

Question 3 (40 points)

This question is about developing a model for `firepwr` as a function of the numerical variables in the set, excluding `strength`.

(i) (5 pts) Do a scatterplot matrix for the numerical variables in the data set, excluding `strength`. Calculate and graph the correlation matrix for these variables. Comment on the results.

(ii) (15 pts) Fit a regression model for `firepwr` as a function of the variables mentioned in (i). Using a critical α of 0.15 and a threshold for the variance inflation factor of 2, obtain a minimal adequate model that includes an intercept. Comment on the steps that you take.

(iii) (7.5 pts) Fit a model using the BIC criterion and compare it with the result of (ii).

(iv) (7.5 pts) Write an equation for the final model and predict the `firepwr` for a dragon with the following covariates. Include confidence intervals at the 99% level.

Table 1: Covariates for prediction

height	length	weight	wing.ln	leg.ht	wing.span	age
50	55	60	30	6	70	350

(v) (5 pts) Print an anova table for the final model and find the estimated variance of the errors. Describe explicitly the sampling distribution for the estimated parameters.