

STAT 210
Applied Statistics and Data Analysis
Second Exam

November 27, 2021

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 `minions.txt`. Read the data onto a data frame. There are nine variables in the set:

- `height`, the height in cm,
- `diam`, the diameter of the minion at mid-height in cm,
- `weight`, the weight in kgs,
- `arm.ln`, the average length for the arms in cm,
- `leg.ln`, the average length for the legs in cm,
- `arm.rch`, the distance between the tips of the outstretched arms,
- `eyes`, whether the minion has one or two eyes,
- `meanness`, the meanness index for the minion, and
- `nasty`, the nastiness index.

Question 1 (30 points)

In this question, we want to explore the relation between the weight (`weight`) and the diameter (`diam`) of minions.

- (i) Start by plotting a graph of `weight` as a function of `diam`. Add a regression line.
- (ii) Fit a simple regression model. What is the R^2 for this model? Write down an equation for the model and give an interpretation of the parameters.
- (iii) What are the assumptions on which this model is based? Using graphical methods or hypothesis tests, check whether these assumptions are satisfied for this model. Use the function `residualPlots` in the `car` package and interpret the graphs and results of the hypotheses test. What do these results suggest?
- (iv) Fit a new model including the term(s) suggested by the tests in (iii), if any. Look at the summary table. What is the adjusted R^2 for this model?
- (v) Check whether the assumptions for linear regression are satisfied for the new model. Write an equation for the model.

Question 2 (30 points)

The variable **nasty** represents an accepted index for the nastiness of minions. We want to explore if this index is related to **height**.

- (i) Graph a scatterplot of **nasty** as a function of **height**. Add the regression line for these variables and comment.
- (ii) 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 for the parameters. Predict the nastiness of a minion with height 60 cm.
- (iii) Use graphical methods to check the assumptions on which the model is based. What are your conclusions?
- (iv) There are two types of minions, one and two-eyed, and this characteristic is available in the categorical variable **eyes**. 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 0.1, starting with the complete model, and using the anova table, select a minimal adequate model.
- (v) 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 index for one and two-eyed minions of height 60 cm. Compare with the previous prediction.

Question 3 (40 points)

This question is about developing a model for **meanness** as a function of the numerical variables in the set, excluding **nasty**.

- (i) Do a scatterplot matrix for the numerical variables in the data set, excluding **nasty**. Calculate and graph the correlation matrix for these variables. Comment on the results.
- (ii) Fit a regression model for **meanness** as a function of the variables mentioned in (i). Using a critical α of 0.2 and a threshold for the variance inflation factor of 2, obtain a minimal adequate model. Comment the steps that you take.
- (iii) Fit a model using the BIC criterion and compare it with the result of (ii).
- (iv) Among the two models produced in (ii) and (iii), choose the one with a bigger adjusted R^2 . Write an equation for the model, and predict the meanness index for a minion with the following covariates. Include confidence intervals at the 99% level.

Table 1: Covariates for prediction

height	diam	weight	arm.ln	leg.ln	arm.rch
50	30	17.5	20	16	50