12/21/2021

Danielle sherman

CTO, Blackwell Electronics

Overview:

As per your request, I explored R and RStudio. I utilized two datasets in testing R, the R tutorial Car data and the petal data from the Find the Errors task.

Installing and setting up R and RStudio was very straight forward and I encountered no issues. The tutorial was more than adequate as a jumping off point for working in that language. I would highly recommend incorporating the tutorial in R beginner’s guides. Below you will find my predictions concerning how far a certain car can travel based on speed, followed by my predictions concerning the petal length by using the petal’s width.

**Predictions concerning how far a certain car can travel based on speed**:

The Multiple R-squared score is 0.9248, which indicates a strong fit.

The p-value is 0.00000000000000022, which is less than 0.05 and therefore statistically significant.

**Predictions**:

1 2 3 4 5 6 7 8

-14.95415 -14.95415 10.41329 30.70724 30.70724 35.78073 35.78073 35.78073

9 10 11 12 13 14 15

56.07468 56.07468 61.14817 66.22166 76.36864 86.51561 86.51561

The predictions show that the higher the car’s speed, the father the distance traveled.

**Predictions concerning the petal length through using the petal’s width**:

The Multiple R-squared score is 0.0895, which does not indicate a strong fit.

The p-value is 0.08083, which is greater than 0.05 and therefore the Independent Variable has no effect on the Dependent Variable.

**Predictions**:

1 2 3 4 5 6 7 8 9 10 11

0.2204113 0.2204113 0.2666838 0.2358355 0.2204113 0.2358355 0.2358355 0.1587147 0.2204113 0.2358355 0.2204113

12 13 14 15 16 17 18 19 20 21 22

0.2049871 0.2512596 0.2204113 0.2512596 0.7294087 0.6985604 0.7602571 0.6214396 0.7139846 0.6985604 0.7294087

23 24 25 26 27 28 29 30 31 32 33

0.5134704 0.7139846 0.6060154 0.5443188 0.6522879 0.6214396 0.7294087 0.5597429 0.6831362 0.6985604 0.6368638

34 35 36 37 38 39 40 41 42 43 44

0.6985604 0.6060154 0.7448329 0.6214396 0.7602571 0.7294087 0.6677121 0.6831362 0.7448329 0.7756812 0.6985604

45 46 47 48 49 50 51 52 53 54 55

0.5443188 0.5905913 0.5751671 0.6060154 0.7911054 0.6985604 0.6985604 0.7294087 0.6831362 0.6368638 0.6214396

56 57 58 59 60 61 62 63 64 65 66

0.6831362 0.7139846 0.6214396 0.5134704 0.6522879 0.6522879 0.6522879 0.6677121 0.4671979 0.6368638 0.9299229

67 68 69 70 71 72 73 74 75 76 77

0.7911054 0.9144987 0.8682262 0.8990746 1.0224679 0.6985604 0.9761954 0.8990746 0.9453470 0.7911054 0.8219537

78 79 80 81 82 83 84 85 86 87 88

0.8528021 0.7756812 0.7911054 0.8219537 0.8528021 1.0378920 1.0687404 0.7756812 0.8836504 0.7602571 1.0378920

89 90 91 92 93 94 95 96 97 98 99

0.7602571 0.8836504 0.9299229 0.7448329 0.7602571 0.8682262 0.8990746 0.9453470 0.9916195 0.8682262 0.7911054

100 101 102 103 104 105 106 107 108 109 110

0.8682262 0.9453470 0.8682262 0.8528021 0.7448329 0.8373779 0.8682262 0.7911054 0.7911054 0.9144987 0.8836504

111 112 113 114 115

0.8065296 0.7756812 0.8065296 0.8373779 0.7911054

Based on the model results the predictions are not very reliable.

Regarding the errors in the script, I identified many errors in syntax before even running the code. I was able to fix the majority of errors by going through the tutorial. I included the pylr library in order to change the Setosa, Versicolor, and Virginica to 1, 2, and 3, respectively. After implementing these changes, I was able to generate a histogram of the Species column. The most important lesson learned was getting acquainted with RStudio and R syntax.

GRANT JOHNSON