Donald Day

MSDS 5350

Regression Assignment

Use the acupuncture data set for this assignment. Read the introduction and the data dictionary before beginning the assignment.

Fill in the table below with the appropriate correlation values:

|  |  |  |
| --- | --- | --- |
|  | Age | chronicity |
| pk1 | 0.013 | 0.0783 |
| pk5 | 0.0881 | 0.133 |
| f1 | -0.0579 | 0.0318 |
| f5 | 0.101 | 0.0894 |
| age | 1.0 | 0.488 |
| chronicity | 0.488 | 1.0 |

Generate scatterplots for age and pk1, age and pk5, age and f1, age and f5, chronicity and pk1, chronicity and pk5 chronicity and f1, chronicity and f5 and age and chronicity

Interpret the relationships between the variables based on the correlations and the scatterplots.

**Age explaining pk1, pk5, and f1 all have a weak correlation as seen by the scatterplots.**

**Chronicity explaining pk1, f1 and f5 all have a weak correlation as seen by the scatterplots.**

**Age explaining chronicity has a moderate strength in correlation and a linear model can be used.**

**Although age explaining f5, and chronicity explaining pk5 have a p-value resulting in a significant slope, both have a weak correlation, and this can be seen with how much random scatter is on the plot.**

In the table below, enter the slope and p-value for the regression model with age or chronicity as the independent variable.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | age |  |  | chronicity |
|  | Slope | p-value | Slope | p-value |
| pk1 | 0.01908 | 0.7945 | 0.09196 | 0.1174 |
| pk5 | 0.13226 | 0.1274 | 0.14921 | 0.02066 |
| f1 | -0.03513 | 0.247 | 0.01550 | 0.5249 |
| f5 | 0.0775 | 0.04267 | 0.05491 | 0.0738 |
| Chronicity/age | 0.60743 | 2.2e-16 | 0.39138 | 2.2e-16 |

Which of the slopes are significant? Does this agree with what you see in the scatterplot?

**Age and f5, age and chronicity, chronicity and pk5 have significant slopes. Age and chronicity, there seems to be a linear pattern being formed. Age and f5 and chronicity and pk5 do not seem to show any relationship when looking at the scatterplot.**

Create a scatterplot using group and pk5 (pk5 is the outcome). Is the scatterplot informative?

Regress pk5 on group (pk5 is the outcome). Is the slope significant? How would you interpret the slope?

**A linear model would not be appropriate for this scatterplot. Even though the groups are numbered 0 and 1, it should be treated as categorical. The plot does give us some information as group 1 has a smaller range and variability towards the lower end than group 0. The slope also shows that this is true. When we perform a regression, we get the LSRL pk5 = 22.343 – 6.097(group). The negative slope shows us an overall trend going down from the control group to the treatment group. Since the p-value, 0.00064 < 0.05, the null hypothesis would be rejected, making this slope significant, remembering that a linear model is not appropriate. As the group increases by 1, the pain score decreases by 6.097 units.**