CS555 - April 6th, 2023

Zuowen Tang

Assignment 5

# Part One

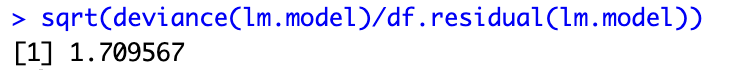
**(1) Obtain the regression equation relating feedlot time to the three diet variables.**

Table

Description automatically generated

The regression equation is 102.7 - 0.83(protein) - 4(antibio) - 1.375(supplement)

**(2) Find the value of Residual Standard Deviation 𝑆.**



**(3) Find the 𝑅2 value.**A picture containing text

Description automatically generated

**(4) How much of a collinearity problem is there with these data (using vif function in car package in R)?**

Text

Description automatically generated with medium confidence

There is the total lack of collinearity, because the independent variables are balanced. The correlation between the pairs of independent variables is zero for each and VIF is 1.00. Thus, there is no collinearity problem prevalent with the given data.

**(5) Predict the feedlot time required for a steer fed 15% protein, 1.5% antibiotic concentration, and 5% supplement.**A picture containing graphical user interface

Description automatically generated

Based on former equation, we have 102.7 - 0.83(protein) - 4(antibio) - 1.375(supplement). So 102.7 - 0.83(15) - 4(1.5) - 1.375(5) = 77.333

**(6) Do these values of the independent variables represent a major extrapolation from the data?**

No, since the given values are too much far from the given values for the independent variables, the data do not represent a major extrapolation.

**(7) Give a 95% confidence interval for the mean time predicted in part (5).**

The confidence interval for the mean time predicted is (76.47, 78.19)

**(8) Analyze the data using a regression model with only protein content as an independent variable.**  
Text

Description automatically generated

**(8a) Obtain the regression equation.**

The regression equation is 89.9 - 0.83(protein)

**(8b) Find the 𝑅2 value.**

A picture containing logo

Description automatically generated

**(8c) Test the null hypothesis that the coefficients of ANTIBIO and SUPPLEM are zero at 𝛼 = .05.**

The general regression equation for the given data is:

Time = B0 – B1(protein) – B2(antibio) – B3(supplem)

So we have:

H0: B2 = B3 = 0

Ha: At least one of the coefficients B2, B3 is not = 0.

From above data we can see that the p-value of both B2 and B3 are lesser than a = 0.05. Thus, the null hypothesis gets rejected, which means that at least one of the coefficients of ANTIBIO and SUPPLEM is not equal to zero.

# Part Two

**(1) A Matrix Scatterplot of all 13 variables of chemical concentration.**

Diagram

Description automatically generated

**(2) Summary Statistics (mean, standard deviation) of all 13 variables of chemical concentration using sapply() function.**

Table

Description automatically generated with low confidence

**(3) After standardizing the data, use the cor() function from the corrr package to calculate the correlation matrix, and uset he ggcorrplot() to generate a visualization.**  
A picture containing Word

Description automatically generated

**(4) Use princomp() to obtain PCA results, use summary table and scree plot to decide how many Principal Components to retain, and explain why.**

Text

Description automatically generated with medium confidence

**(5) Make a biplot combined with cos2 (attributes importance), and combined with (4) to discuss the contributions by important variables of chemical contributions in the new space spanned by the retained Principal Components.**