**Homework Due Day Nov 30**

The summary of three different multiple regression models are given below. The training dataset – the same dataset you worked on Monday in the lab assignment.

The test dataset for this Homework will consist from ONE entry only (partial data is listed below):

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| name | Rating | Sugars | Calories | Fiber | Protein | Fat |
| Froot\_Loops | 32.21 | 13 | 110 | 1 | 2 | 1 |

For each Model:

1. Write which independent variables are used in the model and list their significance
2. Write a multiple linear regression formula to predict Rating
3. Find an estimated Rating for the Froot\_Loops cereal which was not included in the training set
4. List R^2 and adjusted R^2

Compare these three models and answer the following questions:

1. By looking on R^2 and adjusted R^2 which model has a potential to perform the best on the training dataset?
2. Compare predicted and actual values for rating in each model. Which model gave the best prediction? Does it match your answer on the previous question? Explain

**Model 1:**

Coefficients: Estimate Std. Error t value Pr(>|t|)

(Intercept) 89.82318 6.53560 13.744 < 2e-16 \*\*\*

Calories -0.32116 0.06698 -4.795 1.09e-05 \*\*\*

Sugars -1.81077 0.28304 -6.398 2.49e-08 \*\*\*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.302 on 61 degrees of freedom

Multiple R-squared: 0.6796, Adjusted R-squared: 0.6691

F-statistic: 64.71 on 2 and 61 DF, p-value: 8.35e-16

**Model 2:**

Coefficients: Estimate Std. Error t value Pr(>|t|)

(Intercept) 87.94394 6.94122 12.670 < 2e-16 \*\*\*

Calories -0.29477 0.07446 -3.959 0.000202 \*\*\*

Sugars -1.81262 0.28381 -6.387 2.75e-08 \*\*\*

Fat -0.95431 1.16213 -0.821 0.414800

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.324 on 60 degrees of freedom

Multiple R-squared: 0.6832, Adjusted R-squared: 0.6674

F-statistic: 43.13 on 3 and 60 DF, p-value: 5.497e-15

**Model 3:**

Coefficients: Estimate Std. Error t value Pr(>|t|)

(Intercept) 66.99257 5.14769 13.014 < 2e-16 \*\*\*

Calories -0.22154 0.05366 -4.129 0.000116 \*\*\*

Sugars -1.53052 0.21384 -7.157 1.47e-09 \*\*\*

Protein 2.17105 0.84303 2.575 0.012542 \*

Fiber 2.05246 0.38067 5.392 1.29e-06 \*\*\*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.534 on 59 degrees of freedom

Multiple R-squared: 0.8623, Adjusted R-squared: 0.853

F-statistic: 92.37 on 4 and 59 DF, p-value: < 2.2e-16