# Building a predictive model for nutritional rating for Cereal Dataset

**(a)** The MSE for linear regression is shown below:

**Training MSE:**

```
> error_train <- (1/length(y_train))*sum((y_train-y_true)^2)
> error_train
[1] 0.07677425
>
```

**Testing MSE:**

```
> error_test <- (1/length(y_test))*sum((y_test-y_true)^2)
> error_test
[1] 0.09924431
>
```

**(b) The result of Forward Subset selection is shown below:**

```
1 subsets of each size up to 12
Selection Algorithm: forward
          calories protein fat sodium fiber carbo sugars potass vitamins shelf weight cups
1  ( 1 )  " "      " "     " " " "    " "   " "   "*"    " "    " "      " "   " "    " "
2  ( 1 )  " "      " "     " " " "    "*"   " "   "*"    " "    " "      " "   " "    " "
3  ( 1 )  " "      " "     " " "*"    "*"   " "   "*"    " "    " "      " "   " "    " "
4  ( 1 )  " "      " "     "*" "*"    "*"   " "   "*"    " "    " "      " "   " "    " "
5  ( 1 )  " "      " "     "*" "*"    "*"   "*"   "*"    " "    " "      " "   " "    " "
6  ( 1 )  " "      " "     "*" "*"    "*"   "*"   "*"    " "    "*"      " "   " "    " "
7  ( 1 )  " "      "*"     "*" "*"    "*"   "*"   "*"    " "    "*"      " "   " "    " "
8  ( 1 )  "*"      "*"     "*" "*"    "*"   "*"   "*"    " "    "*"      " "   " "    " "
9  ( 1 )  "*"      "*"     "*" "*"    "*"   "*"   "*"    "*"    "*"      " "   " "    " "
10 ( 1 )  "*"      "*"     "*" "*"    "*"   "*"   "*"    "*"    "*"      " "   " "    "*"
11 ( 1 )  "*"      "*"     "*" "*"    "*"   "*"   "*"    "*"    "*"      "*"   " "    "*"
12 ( 1 )  "*"      "*"     "*" "*"    "*"   "*"   "*"    "*"    "*"      "*"   "*"    "*"
```

**The MSE for the forward subset selection is shown below:**

```
> val.f.error
 [1] 6.825561e+01 5.754333e+01 2.358048e+01 6.821211e+00 6.462474e+00 6.293982e+00 3.230211e+00 8.832327e-01 9.073826e-14
[10] 9.111452e-14 9.152292e-14 8.873154e-14
>
```

**(c) The result for exhaustive subset selection is shown below:**

```
1 subsets of each size up to 12
Selection Algorithm: exhaustive
         calories protein fat sodium fiber carbo sugars potass vitamins shelf weight cups
1  ( 1 ) " "      " "     " " " "    " "   " "   "*"    " "    " "      " "   " "    " "
2  ( 1 ) " "      " "     " " " "    "*"   " "   "*"    " "    " "      " "   " "    " "
3  ( 1 ) " "      " "     " " "*"    "*"   " "   "*"    " "    " "      " "   " "    " "
4  ( 1 ) " "      " "     "*" "*"    "*"   " "   "*"    " "    " "      " "   " "    " "
5  ( 1 ) "*"      "*"     " " "*"    "*"   "*"   " "    " "    " "      " "   " "    " "
6  ( 1 ) "*"      "*"     " " "*"    "*"   "*"   "*"    " "    " "      " "   " "    " "
7  ( 1 ) "*"      "*"     "*" "*"    "*"   "*"   "*"    " "    " "      " "   " "    " "
8  ( 1 ) "*"      "*"     "*" "*"    "*"   "*"   "*"    " "    "*"      " "   " "    " "
9  ( 1 ) "*"      "*"     "*" "*"    "*"   "*"   "*"    "*"    "*"      " "   " "    " "
10 ( 1 ) "*"      "*"     "*" "*"    "*"   "*"   "*"    "*"    "*"      " "   " "    "*"
11 ( 1 ) "*"      "*"     "*" "*"    "*"   "*"   "*"    "*"    "*"      "*"   " "    "*"
12 ( 1 ) "*"      "*"     "*" "*"    "*"   "*"   "*"    "*"    "*"      "*"   "*"    "*"
>
```

**The MSE for the exhaustive subset selection is shown below:**

```
> val.e.error
 [1] 6.825561e+01 5.754333e+01 2.358048e+01 6.821211e+00 4.552774e+00 2.665865e+00 1.677157e+00 8.832327e-01 9.073826e-14
[10] 9.111452e-14 9.152292e-14 8.873154e-14
>
```

**(d) The best model as per our results is the exhaustive subset selection model with 9 variables.**