

# ASSIGNMENT 1: Multiple Linear Regression

## First order Model with Interaction Term (Quantitative Variables)

**Problem 1.** (From Exercise 1) The amount of water used by the production facilities of a plant varies. Observations on water usage and other, possibly related, variables were collected for 250 months. The data are given in **water.csv file**. The explanatory variables are

TEMP= average monthly temperature (degree celsius)

PROD=amount of production (in hundreds of cubic)

DAYS=number of operationing day in the month (days)

HOURL=number of hours shut down for maintenance (hours)

The response variable is USAGE=monthly water usage (gallons/minute)

- Fit the model containing all four independent variables. What is the estimated multiple regression equation?
- Test the hypothesis for the full model i.e the test of overall significance. Use significance level 0.05.
- Would you suggest the model in part b for predictive purposes? Which model or set of models would you suggest for predictive purposes? Hint: Use Individual Coefficients Test (t-test) to find the best model.)
- Use Partial  $F$  test to confirm that the independent variable (removed from part c) should be out of the model at significance level 0.05.
- Obtain a 95% confidence interval of regression coefficient for TEMP from the model in part c. Give an interpretation.
- Use the method of Model Fit to calculate  $R^2_{adj}$  and RMSE to compare the full model in part a and the model in part c. Which model or set of models would you suggest for predictive purpose? For the final model, give an interpretation of  $R^2_{adj}$  and RMSE.
- Build an interaction model to fit the multiple regression model from the model in part f. From the output, which model would you recommend for predictive purposes?

---

**Problem 2.** A collector of antique grandfather clocks sold at auction believes that the price received for the clocks depends on both the age of the clocks and the number of bidders at the auction. Thus, (s)he hypothesizes the first-order model

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

where

$y$  = Auction price (dollars)

$X_1$  = Age of clock (years)

$X_2$  = Number of bidders

A sample of 32 auction prices of grandfather clocks, along with their age and the number of bidders, is given in data file **GFCLOCKS.CSV**

- Use the method of least squares to estimate the unknown parameters  $\beta_0, \beta_1, \beta_2$  of the model.
- Find the value of SSE that is minimized by the least squares method.
- Estimate  $s$ , the standard deviation of the model, and interpret the result.
- Find and interpret the adjusted coefficient of determination,  $R^2_{Adj}$ .
- Construct the ANOVA table for the model and test the global F-test of the model at the  $\alpha = 0.05$  level of significance.
- Test the hypothesis that the mean auction price of a clock increases as the number of bidders increases when age is held constant (i.e., when  $\beta_2 \neq 0$ ). (Use  $\alpha = 0.05$ )
- Find a 95% confidence interval for  $\beta_1$  and interpret the result.
- Test the interaction term between the 2 variables at  $\alpha = .05$ . What model would you suggest to use for predicting  $y$ ? Explain.

---

**Problem 3. Cooling method for gas turbines.** Refer to the Journal of Engineering for Gas Turbines and Power (January 2005) study of a high pressure inlet fogging method for a gas turbine engine. The heat rate (kilojoules per kilowatt per hour) was measured for each in a sample of 67 gas turbines augmented with high pressure inlet fogging. In addition, several other variables were measured, including cycle speed (revolutions per minute), inlet temperature (degree celsius), exhaust gas temperature (degree Celsius), cycle pressure ratio, and air mass flow rate (kilograms persecond). The data are saved in the **TURBINE.CSV** file.

| RPM   | CPRATIO | INLET-TEMP | EXH-TEMP | AIRFLOW | HEATRATE |
|-------|---------|------------|----------|---------|----------|
| 27245 | 9.2     | 1134       | 602      | 7       | 14622    |
| 14000 | 12.2    | 950        | 446      | 15      | 13196    |
| 17384 | 14.8    | 1149       | 537      | 20      | 11948    |
| 11085 | 11.8    | 1024       | 478      | 27      | 11289    |
| 14045 | 13.2    | 1149       | 553      | 29      | 11964    |
| .     | .       | .          | .        | .       | .        |
| 18910 | 14.0    | 1066       | 532      | 8       | 12766    |
| 3600  | 35.0    | 1288       | 448      | 152     | 8714     |
| 3600  | 20.0    | 1160       | 456      | 84      | 9469     |
| 16000 | 10.6    | 1232       | 560      | 14      | 11948    |
| 14600 | 13.4    | 1077       | 536      | 20      | 12414    |

*Source: Bhargava, R., and Meher-Homji, C. B. "Parametric analysis of existing gas turbines with inlet evaporative and overspray fogging," Journal of Engineering for Gas Turbines and Power, Vol. 127, No. 1, Jan. 2005.*

The first and last five observations are listed in the table.

- Write a first-order model for heat rate ( $y$ ) as a function of speed, inlet temperature, exhaust temperature, cycle pressure ratio, and air flow rate.
- Test the overall significance of the model using  $\alpha = 0.01$
- Fit the model to the data using the method of least squares. (Suggestion! check both models with and without a predictor that has p-value close to 0.05, and propose the best model.)
- Test all possible interaction terms for the best model in part (c) at  $\alpha = .05$ . What is the final model would you suggest to use for predicting  $y$ ? Explain.
- Give practical interpretations of the  $\beta_i$  estimates.
- Find RMSE,  $s$  from the model in part (d)
- Find the adjusted-R<sup>2</sup> value from the model in part (d) and interpret it.
- Predict a heat rate ( $y$ ) when a cycle of speed = 273,145 revolutions per minute, inlet temperature= 1240 degree celsius, exhaust temperature=920 degree celsius, cycle pressure ratio=10 kilograms persecond, and air flow rate=25 kilograms persecond.