# **OPIM 5603 — Statistics in Business Analytics**

# **Spring 2021**

# **Homework 5**

Instructions: Please complete the following questions and submit them as an R Notebook script (as an .RMD file) via the submission link on HuskyCT. You must submit the assignment by the time and due date listed in HuskyCT. Failure to submit a file by the deadline will result in a score of 0 on the assignment.

As with all course material, the problems appearing in this homework assignment are taken from a variety of sources including from other courses taught at the University of Connecticut.

Note that R code submitted should work independent of the data that sits in the data structure. Your R code should be sufficiently general to work with other data sets than the one given or other parameters or variables defined in your notebook.

If you have any questions, please submit them via email to the instructor and/or the graders prior to submitting your solution.

**Deliverables**

Each homework has to be submitted through HuskyCT before the deadline. If submitted after the deadline, a late penalty may be deducted. For each homework using the R package, you need to submit:

* The code in your R notebook (an RMD file in plain text)
* An HTML page containing the output of your code (“Knitted to html”).

**Problem (30 points)**

**Oil Plus Inc.**

OilPlus, Inc. provides heating oil to residential customers in Lancaster, Pennsylvania. OilPlus would like to predict heating oil consumption among its customers in order to plan its own fuel purchases and to budget its revenue flows appropriately.

Table 1 shows heating oil consumption of OilPlus’ customer base and average monthly temperature for 55 consecutive months from August, 1989 through February, 1994.

1. Make a scatterplot of the data in Table 1. Why does a linear regression model seem appropriate? Run a regression model that predicts oil consumption as a function of temperature. Make a plot with the regression line added.
2. OilPlus would like to predict the heating oil consumption for next December. According to the temperature data in Table 1, the average December monthly temperature over the years 1988-1993 is 35.2 degrees. Use the regression model to predict the heating oil consumption for next December based on this average temperature.
3. According to the heating oil consumption data in Table 1, the average heating oil consumption for Decembers (1989-1993) was 75.92 thousand gallons for the month. It might therefore seem reasonable to use this number as a prediction for next December’s heating oil consumption. What advantage does prediction based on the regression model have over this simple averaging method?
4. Evaluate the regression model by examining the regression model output. Check the R2 , the confidence intervals, the P-values and the scatter plot of the regression residuals (both vs. predictor and vs. time). What do you notice, i.e., is this a good model?
5. Run the model with a quadratic term for Temperature added to the model. What do you conclude? Is there evidence of a non-linear effect or not? How to explain?
6. Would you recommend using this regression model to predict heating oil consumption at OilPlus? Why or why not?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Month*** | ***Heating Oil Consumption (1,000 gallons)*** | ***Average Temperature (degrees Fahrenheit)*** | ***Month*** | ***Heating Oil Consumption (1,000 gallons)*** | ***Average Temperature (degrees Fahrenheit)*** |
| August-89 | 24.32 | 72 | December-91 | 80.68 | 37 |
| September-89 | 27.33 | 69 | January-92 | 77.14 | 31 |
| October-89 | 15.68 | 53 | February-92 | 69 | 38 |
| November-89 | 57.43 | 40 | March-92 | 62.27 | 35 |
| December-89 | 104.4 | 31 | April-92 | 57.73 | 53 |
| January-90 | 48.83 | 40 | May-92 | 32.36 | 57 |
| February-90 | 55.4 | 34 | June-92 | 21 | 67 |
| March-90 | 59.58 | 50 | July-92 | 28.69 | 72 |
| April-90 | 54.48 | 52 | August-92 | 36.57 | 75 |
| May-90 | 26.29 | 60 | September-92 | 25.92 | 61 |
| June-90 | 24.22 | 74 | October-92 | 41.25 | 51 |
| July-90 | 23.84 | 74 | November-92 | 51.49 | 48 |
| August-90 | 20.07 | 71 | December-92 | 53.12 | 36 |
| September-90 | 29.52 | 70 | January-93 | 58.46 | 30 |
| October-90 | 22.83 | 61 | February-93 | 71.46 | 32 |
| November-90 | 54.95 | 53 | March-93 | 40.46 | 34 |
| December-90 | 75.18 | 37 | April-93 | 21.03 | 56 |
| January-91 | 84.7 | 32 | May-93 | 25.55 | 65 |
| February-91 | 42.81 | 36 | June-93 | 8.78 | 70 |
| March-91 | 42.39 | 45 | July-93 | 20.92 | 82 |
| April-91 | 31.11 | 59 | August-93 | 19.39 | 70 |
| May-91 | 29.49 | 73 | September-93 | 22.34 | 63 |
| June-91 | 16.99 | 72 | October-93 | 44.41 | 57 |
| July-91 | 14.78 | 74 | November-93 | 38.13 | 43 |
| August-91 | 25.27 | 77 | December-93 | 66.24 | 35 |
| September-91 | 19.96 | 67 | January-94 | 101.23 | 23 |
| October-91 | 26.8 | 55 | February-94 | 102.98 | 33 |
| November-91 | 38.09 | 45 |  |  |  |

**Table 1.**  Consumption of Heating Oil and Average Monthly   
Temperature for 55 consecutive months