SA463 • Operations Research in Action Fall 2024 • Uhan

**Project 1 – Model and Results – Feedback**

**General**

* I try my best to provide detailed feedback, but I cannot catch everything. Try to generalize my feedback to other places in your report where you think it is appropriate.

**The model section**

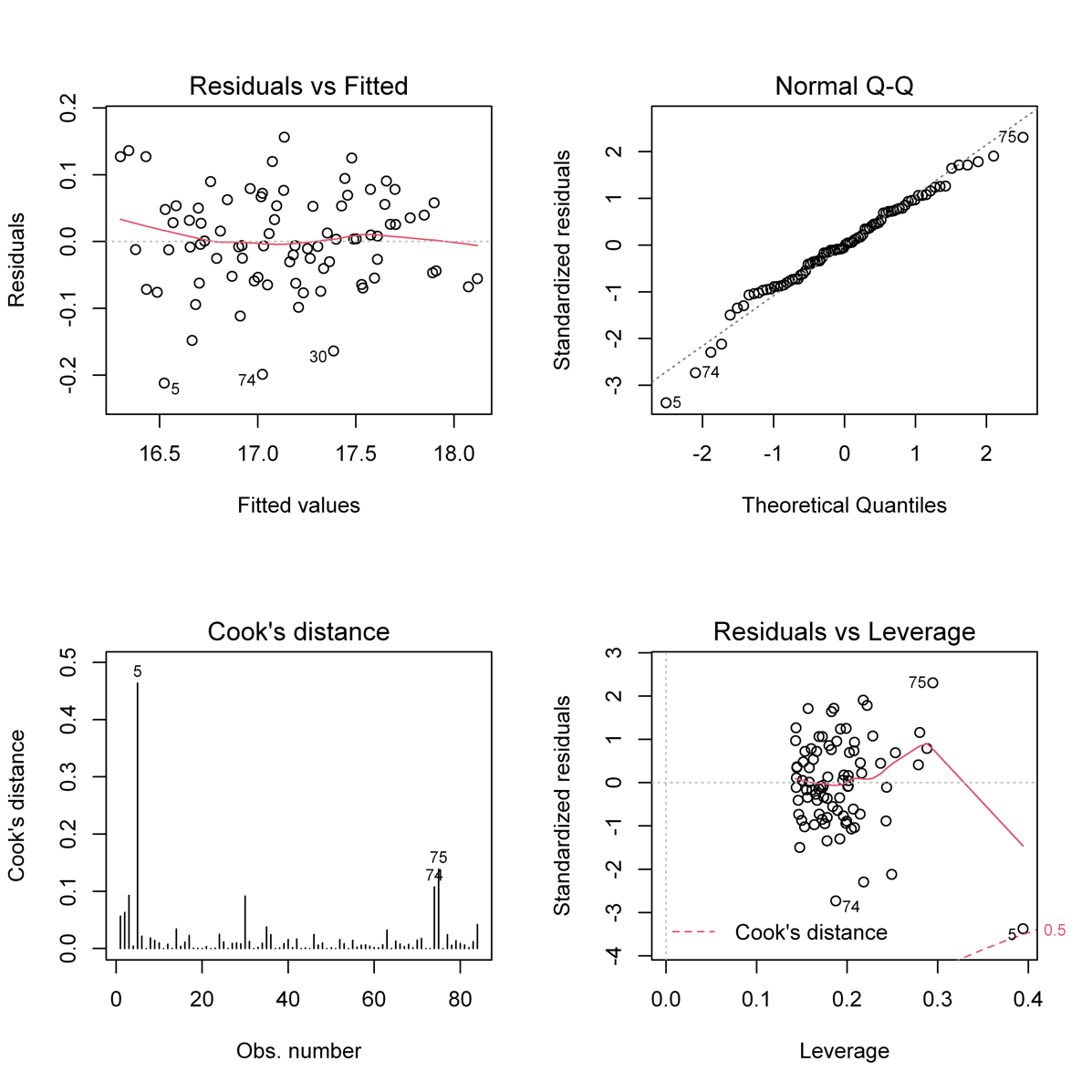
* Describing the variable selection process.
  + Even though we spent an entire Jupyter notebook on this, you should keep this description short.
  + To summarize what we did in Part 2 of the Project 1 Modeling lesson:
    - We used best subsets regression to determine which subset of explanatory variables to include in our model, using the AIC as a means of comparing different models.
    - We found that the best model from best subsets regression violates the linearity condition for linear regression.
    - To fix this, we tried using the log of monthly beer consumption as the response variable instead. (Many of you forgot to include this.)
    - We ran best subsets regression again with this new response variable.
    - This time, we found that the best model had a few statistically insignificant variables. We removed them to obtain our final model, which we describe in the results section below.

**The results section**

* Describing the final linear regression model.
  + Restate that the response variable is the *log of monthly beer consumption*. This is worth emphasizing, since we changed the response variable during the variable selection process.
  + You need to include a table that shows the estimated coefficients of your final fitted model (the values of  ) as well as the p-values from the t-tests for these coefficients, like this:

|  |  |  |
| --- | --- | --- |
|  | Estimate | t-test p-value |
| (Intercept) | -256.7000 | 0.0000 |
| AvgBeerPrice | -0.0005 | 0.0001 |
| AvgCannedSoftDrinkPrice | 0.0002 | 0.0098 |
| RamadanDays | -0.0145 | 0.0000 |
| Year | 0.1375 | 0.0000 |
| MonthFebruary | 0.0410 | 0.3508 |
| MonthMarch | 0.3540 | 0.0000 |
| MonthApril | 0.2796 | 0.0000 |
| MonthMay | 0.5993 | 0.0000 |
| MonthJune | 0.7858 | 0.0000 |
| MonthJuly | 0.9545 | 0.0000 |
| MonthAugust | 0.9922 | 0.0000 |
| MonthSeptember | 0.7667 | 0.0000 |
| MonthOctober | 0.5802 | 0.0000 |
| MonthNovember | 0.4507 | 0.0000 |
| MonthDecember | 0.3729 | 0.0000 |

Table 1. Estimated coefficients and t-test p-values for final model.

* + Note that *summary statistics* are descriptive values to summarize a set of observations, like the mean, median, minimum and maximum. The table above is *not* an example of summary statistics.
  + After discussing the diagnostics that indicate that the conditions for linear regression have been reasonably met, you also need to discuss what the p-values mean.
  + When including R graphics (or any other graphics) in your report, resize the graphics so that the font size is roughly the same size as or slightly smaller than the main text.
  + To increase the resolution of R graphics in a Jupyter notebook, use:  
      
    options(repr.plot.res = 800) # play around with the value  
      
    Here are the diagnostic plots generated with this setting:  
      
    
* Interpreting the final fitted linear regression model.
  + Use the Project 1 Results lesson as a guide on how to interpret the results. Feel free to borrow the language we wrote together in class.
  + You should include an example of how we can use the model to predict beer consumption, including computing prediction intervals with a specified level of confidence.