Satoshi Ido
34788706
Professor. Craig
31 January, 2023

STAT 582

HW2

1. Ideas/Content (Development)

   Score: 4

   Reasons: The report described mainly statistical modeling and we can see that motivation easily. However, the author could clarify its purpose/goal and results on top of the report so that the readers can understand the report quickly by just skimming through it. Also, the author included some unnecessary information in the middle of the report, which potentially distracts readers from reading the document.

2. Organization

   Score: 3

   Reasons: The author could have built more concise and
   The author included both methods/design and analysis parts within the same paragraph: “Modeling ignition delay for loads 20, 40, and 60”. However, the person could separate those parts in order to deliver more clear flow and information to readers. For example, the author included Table 1: Data on Diesel Engine is unnecessary and it almost distracts readers more than helps the person. He/she could put the table in an appendix as well.

3. Voice

   Score: 4

   Reasons: The author showed concise explanations alongside each plot or table. However, those include too many jargons and unclear messages and they are less engaging for any readers who have little knowledge of statistical analysis. Also, the

author did not provide a detailed analysis process for an adept reader who could reproduce the analysis if needed. He/she could include those at least in the appendix.

4. Word Choice (less technical - no stats jargon)
   Score: 4
   Reasons: As I pointed out above, the author often used statistical jargon within the report and made the report hard to understand among non-statisticians. There were also some grammatical errors in the report such as frequent use of a passive voice sentence or some mistakes of grammatical number.

5. Sentence Fluency
   Score: 4
   Reasons: As the reader who has some statistical knowledge, it was not hard to understand the author's report until he/she brought up the alternative models in the section: "Modeling ignition delay through temperature and pressure" of page 4. From my perspective, the idea came suddenly. Hence, I naturally had a skeptical stance toward the new approach.

6. Conventions (e.g., spelling, punctuation, capitalization, grammar, passive voice?)
   Score: 4
   Reasons: The data is plural and not singular. The author wrote down the phrases as "Note that no data was collected for the" in page 1, but this should be "Note that no data were collected for the." Also, in page 5, it is better to keep an active voice in a sentence so phrases "Further data in this range is needed" should be "We needed further data in this range"

7. Analytic Methods (plots output might be better than a table)
   Score: 3
   Reasons: The author provided some output from his/her exploratory data analysis. Yet, those are less intuitive and are not reader-friendly. For example, the author provided the Table 2: Full model fits for each load and timing combination, yet it would provides the readers unclear information and message because of some jargon such as R^2, Load*Alcohol^2 (which the author describes as quadratic in the sentence as well). Instead, the author could replace those terms with some reader-friendly words and

sentences. For example, we can replace R^2 with “how much the model fit the observed data,” or Load*Alcohol^2 with “a combined effect of load and alcohol with some modeling adjustment”

8. Figures and Tables (unified)

   Score: 3

   Reasons: There are multiple parts where the figures and tables are possibly unclear for any readers who do not have statistical knowledge. For example, in page 3, after including the “speed” factor and “ injection timing” factor into the model to analyze, the author simply pasted the model outputs, which the readers would find hard to interpret without the statistics background. The author instead put those results in the appendix and simplified the results by only mentioning key parts such as whether the model describes the data enough or not.

Also, I noticed that the author has not clearly figured out how to handle the lack of data for loads = 70, which he could have discussed with the client in advance.