STA130H1S – Fall 2022

Problem Set 8

() and STA130 Professors

Instructions

Complete the exercises in this .Rmd file and submit your .Rmd and .pdf output through Quercus on Thursday, November 17th by 5:00 p.m. ET.

```
library(tidyverse)
```

Part 1: Multivariate Linear Regression

Question 1: More Markiokart

In this question, you will revisit the Mario Kart data we looked at in this week's class. This data set contains eBay sales of the game Mario Kart for Nintendo Wii in October 2009 and is available in the `openintro` R package (as loaded in the code chunk below).

```
mariokart <- read_csv("mariokart.csv")
mariokart2 <- mariokart %>% filter(total_pr < 100)
# data set documentation indicates that these very high-priced items were
# bundles of several games, not just the Mario Kart game.
```

(a) Sellers on eBay have the option to include a stock photo as the illustration of the product for sale. Does this choice affect the selling price?

Carry out a regression analysis and predict the mean selling price of the `total_pr` variable for sellers who do and do not use stock photos.

```
# code you answer here
```

(b) Sellers are rated by buyers on eBay, captured in the variable `seller_rating`. To simplify our analysis, we will categorize sellers by whether their rating is low, medium or high. Using `mutate()` and `case_when()`, create a new variable called `seller_rating_tier` that is "low" if `seller_rating` is less than or equal to 200, "medium" if it is greater than 200 but less than or equal to 4500, and "high" if it is greater than 4500. Carry out a regression analysis to predict `total_pr` for the "low", "medium", and "high" levels of the new `seller_rating_tier` variable.

```
# code you answer here
```

i. How many indicator variables are in the model? Describe these indicator variables.

REPLACE THIS TEXT WITH YOUR ANSWER

ii. Which seller rating group is `R` treating as the baseline category?

REPLACE THIS TEXT WITH YOUR ANSWER

iii. What is the estimate from the fitted regression line for the mean `total_pr` for sellers with low ratings? What is the estimate from the fitted regression line for the mean `total_pr` for sellers with medium ratings? What is the estimate from the fitted regression line for the mean `total_pr` for sellers with high ratings?

REPLACE THIS TEXT WITH YOUR ANSWER

iv. Create boxplots of `total_pr` for each category of seller. Is this visualization consistent with your estimates in (iv)?

REPLACE THIS TEXT WITH YOUR ANSWER

```
# code you answer here
```

(c) Now produce an appropriate plot and fit an appropriate regression line to examine whether `seller_rating_tier` has an effect on the relationship between `total_pr` and `duration`.

The regression model is

$$\texttt{total\_pr}_i = \beta_0 + \beta_1 \texttt{seller\_tier\_low}_i + \beta_2 \texttt{seller\_rating\_tier\_medium}_i + \beta_3 \texttt{duration}_i + \beta_4 \texttt{seller\_rating\_tier\_low}_i \times \texttt{duration}_i + \beta_5 \texttt{seller\_rating\_tier\_medium}_i \times \texttt{duration}_i + \epsilon_i$$

```
# code you answer here
```

i. What is the equation of the fitted regression line for sellers with low ratings?

REPLACE THIS TEXT WITH YOUR ANSWER

ii. What is the equation of the fitted regression line for sellers with medium ratings?

REPLACE THIS TEXT WITH YOUR ANSWER

iii. What is the equation of the fitted regression line for sellers with high ratings?

REPLACE THIS TEXT WITH YOUR ANSWER

(d) Does the seller rating tier modify the association between duration and total price? Write 1-2 sentences explaining your answer.

REPLACE THIS TEXT WITH YOUR ANSWER

(e) Divide the data into testing and training datasets and fit the linear regression models for total price, with the following variables as predictors (using the training dataset):

i. `stock_photo`

ii. `stock_photo`, `duration`, and their interaction

iii. `seller_rating`

iv. `stock_photo`, `seller_rating`, and their interaction

v. `stock_photo`, `seller_rating`, `duration`, and all interaction terms

```
set.seed(130) # use this seed to make your analysis reproducible
# code you answer here
```

(f) Calculate the RMSE for each of the five models from part (e), for both the training and testing datasets. Which model would you prefer to use for future predictions, and why (in 1-2 sentences).

REPLACE THIS TEXT WITH YOUR ANSWER

```
# code you answer here
```

Part 2: Preparation for the Weeks' Tutorial

In tutorial this week, you will (in groups) present the analysis of Ross *et al.* (Nature 2022) "Women are Credited Less in Science than are Men", available on quercus and online through the UT library. In preparation for that, we'll here explore the story of Rosalind Franklin which is noted by Ross *et al.* to frame their work.

The discovery, or proposal, of the double-helix nature of DNA was published in 1953. In 1962 the Nobel Prize for Physiology or Medicine "for their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material" was awarded to three men: Wilkins, Crick, and Watson. A woman, Rosalind Franklin was not included in the award. However, Watson himself donated a commemorative statute upon which the base (i.e., foundation) read "The double helix model was supported by the work of Rosalind Franklin and Maurice Wilkins" while the helices themselves contain the statement "The structure of DNA was discovered in 1953 by Francis Crick and James Watson while Watson lived here at Clare [College, Cambridge, England]". The remainder of the inscriptions read on the statue are "The molecule of DNA has two helical strands that are linked by base pairs Adenine – Thymine or Guanine – Cytosine" and "These strands unravel during cell reproduction. Genes are encoded in the sequence of bases". Sadly, there is a technical excuse for why Rosalind Franklin is not a co-recipient on the award: the Nobel Prize cannot be awarded posthumously, and Rosalind Franklin died in 1958 (at only 37).

Question 2: personal bias and the "Appeal to Authority" fallacy

The Ross *et al.* manuscript states "[Rosalind Franklin] was wrongfully denied authorship on the original Crick and Watson [1953] paper."

This appears to be different claim than is made by the paper "Rosalind Franklin and the discovery of the structure of DNA" (Klug Nature 1968) which Ross *et al.* referenced in the context of their statement, whose abstract reads

> "In this article Dr Klug discusses Dr Franklin's contribution to the discovery of the structure of DNA in the light of accounts given by Professor Watson in his book The Double Helix and by Dr Hamilton in a recent article in Nature"

and whose concern is oriented around contribution and recognition rather than authorship.

(a) Explore your initial reaction to the Ross *et al.* statement. What are your personal biases, experiences, and preferences that lead you to feel this way? In your reflection, be as open and honest with yourself as you are able to be.

REPLACE THIS TEXT WITH YOUR ANSWER

(b) In principle, what is the difference between misrepresenting something you are reporting, versus misrepresenting what someone else has reported, if anything? Think carefully about this question *per se*, rather than as a judgemental or opinionated "loaded question" statement relative to the Ross *et al.* manuscript.

REPLACE THIS TEXT WITH YOUR ANSWER

Question 3: personal bias and the "Appeal to Authority" fallacy, part 2

STA130 professor Scott Schwartz, upon reading the abstract of the Ross *et al.* manuscript, specifically

> "women in research teams are significantly less likely than men to be credited with authorship... women are significantly less likely to be named on a given article [13.24% gap] or patent [58.40% gap] produced by their team relative to their male peers. The gender gap in attribution is present across most scientific fields and almost all career stages."

had the following to say:

> "I don't see how rightful authorship to the level reported by this paper could be possible... statistics and math papers do not have doubt as to who authors should be: authorship is not about contribution of ideas, but contributions to papers (with contributing ideas cited and referenced). I do see possible authorship bias happening from a mechanism whereby women are not invited to collaborate on a paper when they should be."

(a) What biases does the statement appear to have? Does it rush to judgement quickly, or does it appear to read the abstract honestly and charitably?

REPLACE THIS TEXT WITH YOUR ANSWER

(b) Was Professor Schwartz statement compelling to any degree? Why or why not?

REPLACE THIS TEXT WITH YOUR ANSWER

(c) Do you view Professor Schwartz as an authority figure whose opinion might matter here? Or do you think he's not an authority figure in this context?

REPLACE THIS TEXT WITH YOUR ANSWER

(d) Relative to Professor Schwartz's opinion, how important, or in what ways might it be important whether or not Professor Schwartz is an authority figure on this topic?

REPLACE THIS TEXT WITH YOUR ANSWER

Question 4: hedging

Previously you've watched this 7-minute video introduction to hedging. But now watch it again with a new perspective in mind. Yes, as previously noted the last time this video was introduced

> "Hedging is helpful whenever you can't say something is 100% one way or another, as is often the case. In statistics, hedging should always be used with respect to the limitations of data and the strength and generalizability of the conclusions."

but also, hedging is a way to overcome hesitation and defensiveness people can experience when confronted with opinions that challenge their personal biases and *a priori* beliefs. Approaching conversations with more open-minded language can keep people from "rushing to judgement" and "shutting down" and more willing to engage and listen.

(a) Do you think the statements under consideration in Questions 2 and 3 could be made more approachable and palatable by using hedging? How so?

REPLACE THIS TEXT WITH YOUR ANSWER

Question 5: plagiarism (and possibly more hedging)

Previously you've watched this 8-minute video introduction to plagiarism. Plagiarism is an extremely important topic for academic integrity, and citing the reference material you rely on makes your own work come across even more convincingly. So, watch this video again, and try to pick up even more than what you remember learning last time you watched it. For very important topics, like plagiarism, returning to review the topic is a very important tool for deeply integrating the topic into your own thinking and actions.

Let us now return to Rosalind Franklin, and consider the notion of plagiarism in her context.

As noted in a commentary,

> "Alongside the Watson-Crick paper in the April 25, 1953, issue of Nature were separately published papers by scientists Maurice Wilkins and Rosalind Franklin of King's College, who worked independently of each other. The Wilkins and Franklin papers described the X-ray crystallography evidence that helped Watson and Crick devise their structure. The authors of the three papers, their lab chiefs, and the editors of Nature agreed that all three would be published in the same issue."

and New York Times science reporter Nicholas Wade, in a panel discussion with preeminent Rosalind Franklin scholar Lynne Osman Elkin, agreed, saying:

> "There were three papers published in Nature, the first by Watson and Crick. The second and the third by Wilkins and by Franklin. So Franklin got to say, in Nature, in the same issue as Watson and Crick, everything she knew about DNA, including the publication of the photographs."

Elkin, however, points out the following:

> "In the early years, only crystallographers knew that there was something wrong with the three papers. They looked at their brilliant, brilliant first paper. And I want to be clear, I think everybody deserves credit for this. I do not diminish anybody's contribution to this work. And then it's, okay, where's the data? Anybody who knew anything knew we need data for this.
>
> Then we look at Wilkin's paper and say they're as near. Then they look at Franklin's paper and there is the data but why isn't she acknowledged. And part of that was because of the snaky deal that was done between Randall, the head of Kings's, and Bragg, the head of Cavendish [Laboratory], to cover up the very awkward fact that the data had migrated from one place to the other."

In the Watson-Crick manuscript itself, the authors have the following to say:

> "We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at King's College, London."
>
> and
>
> "So far as we can tell, [our model for DNA structure] is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications [from Dr. Wilkins and Dr. Franklin]. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments."

Two more notes about these statements are made in the previously noted commentary, however:

> 'Watson and Crick say that they "were not aware of the details" of the work of King's College scientist Rosalind Franklin – a statement that marks what many consider an inexcusable failure to give Franklin proper credit. According to Lynne Elkin ... Watson and Crick were not aware of all the details of Franklin's work, but they were aware of enough of the details to discover the structure of DNA. Yet this paper does not ever formally acknowledge her, instead concealing her significant role by saying they "were not aware" of her work ... Franklin was at work on

> the DNA puzzle using X-ray crystallography, which involved taking X-ray photographs of DNA samples to infer their structure. By late February 1953, her analysis of these photos brought her close to the correct DNA model.'
>
> and
>
> 'Interestingly... a stronger acknowledgment of Franklin's work [appears] in an early draft of the paper: "We have also been stimulated by the very beautiful experimental work of Dr. M. H. Wilkins and his co-workers at Kings College, London." Elkin suggests that the phrase "very beautiful" is most likely a nod to Franklin's X-ray photograph. The same draft also acknowledged Franklin's work with the sentence: "It is known that there is much unpublished experimental material." When Maurice Wilkins read the draft, he advised Watson and Crick to delete this sentence and the phrase "very beautiful." They agreed to his suggestion.'

In your responses to the questions below, consider using hedging to try to make your arguments more approachable in a manner that invites consideration, engagement, and exploration. This will likely have a better chance of "winning someone over" as opposed to merely expressing an opinion (since, as you know, what they say about opinions is... everybody has one).

(a) What do you think? Was plagiarism involved in the seminal manuscript from Watson and Crick?

REPLACE THIS TEXT WITH YOUR ANSWER

(b) Returning to the introduction of Part 2 (describing how Rosalind Franklin was not awarded the Nobel Prize was, but Wilkins, Crick, and Watson were), and now knowing that Rosalind Franklin is universally accepted as deserving credit in the discovery of the double-helix structure of DNA, do you think she has not recieved proper acknoweledgement and credit? And if so, how?

REPLACE THIS TEXT WITH YOUR ANSWER

(c) In your final assessemnt, do you think Rosalind Franklin was a victim of plagiarism? Do you agree or disagree with the Ross *et al.* statement "[Rosalind Franklin] was wrongfully denied authorship on the original Crick and Watson [1953] paper."?

REPLACE THIS TEXT WITH YOUR ANSWER

Question 6: Personal Ethics and Decisions (and possibly more hedging)

Watson and Crick produced a manuscript without Rosalind Franklin as a co-author while understanding the relevance of her data to their thesis. In STA130 professor Scott Schwartz's assessment, since Watson and Crick did not present Franklin's **data** in their manuscript (which is universally accepted as having merit) and only suggested its alignment with their proposed theory, the acknowledgments they provided in their paper were acceptable and sufficient.

Professor Schwartz's personal opinion is contrary to that of preeminent Rosalind Franklin scholar Lynne Osman Elkin, however. Elkin argues that "acceptable and sufficient" is hardly an appropriate characterization, and that the Watson and Crick manuscript intentionally obfuscates the truth of Rosalind Franklin's contribution. Again from the previously noted commentary:

> 'Then in perhaps the most pivotal moment in the search for DNA's structure, Wilkins, a longtime friend of Crick, showed Watson one of Franklin's photographs without Franklin's permission. Watson recalled, "The instant I saw the picture my mouth fell open and my pulse began to race." To Watson, the cross-shaped pattern of spots in the photo meant that DNA had to have a helical structure. Franklin's photograph was critical in solving the problem, as Watson admitted in his 1968 book, The Double Helix. Watson and Crick also had access to an internal report from

the Medical Research Council, a British agency for funding life sciences, summarizing much of Franklin's unpublished work on DNA, including precise measurements of the molecule. As the Cavendish representative to the agency, scientist Max Perutz had a copy of the report, and when Crick asked to see it, Perutz obliged. While the report was not confidential, science historian Lynne Elkin contends that "showing unpublished work to an unacknowledged competitor was a questionable act which justifiably infuriated" John Randall, the head of King's. Crick later said the data in the report enabled him to reach the significant conclusion that DNA has two chains running in opposite directions. Although Franklin was listed in the acknowledgements section with other scientists, there was no specific mention of her contributions.'

In your responses to the questions below, be mindful of opportunities to use hedging to try to make your arguments more approachable in a manner that invites consideration, engagement, and exploration

(a) Should we accept the acknowlegements of the Watson and Crick manuscript as "acceptable and sufficient"? Should we be hesitant to take them at face value and trust that they are indeed "acceptable and sufficient"?

REPLACE THIS TEXT WITH YOUR ANSWER

(b) In the context of the "race to the double helix", what do you make of the choice by Watson and Crick to benefit off of the work of Rosalind Franklin without approaching or including her as a direct co-collaborator?

REPLACE THIS TEXT WITH YOUR ANSWER

Question 7: Personal Ethics, Decisions, and Hidden and Unconscious Bias

Wilkins have Rosalind Franklin both worked at King's College, but "worked independently of each other". So why did Wilkins ("longtime friend of Crick") have Rosalind Franklin's image to share with Crick "without Franklin's permission"? From the commentary:

> "... Franklin decided to take another job. As she was preparing to leave, she turned her X-ray photographs over to her colleague Maurice Wilkins."

What about Max Perutz? From wikipedia:

> Max Perutz gave Rosalind Franklin's images to Crick without consulting Rosalind Franklin. Perutz later stated that (a) the images were part of a King's College portfolio which he was told should be shared with other research groups and that (b) Rosalind Franklin gave a public talk with the same images which Watson had attended so they were already "public knowledge".

(a) Was it reasonable for Wilkins to share Rosalind Franklin's photographs with Crick without consulting Rosalind Franklin?

REPLACE THIS TEXT WITH YOUR ANSWER

(b) Was it reasonable for Perutz to share a report "summarizing much of Franklin's unpublished work on DNA, including precise measurements [and images] of the molecule." with Watson without consulting Rosalind Franklin?

REPLACE THIS TEXT WITH YOUR ANSWER

(c) Wilkins, Crick, Watson AND Franklin are now understood to deserve credit for the discovery of the double-helix structure of DNA. Watson and Crick had Wilkins review their draft manuscript, and made edits to their final manuscript as a result of Wilkins recommendations. Was it reasonable for Watson and Crick to exclude Rosalind Franklin from the review and feedback process as they finalized their manuscript?

REPLACE THIS TEXT WITH YOUR ANSWER

Question 8: Personal Ethics, Decisions, and Hidden and Unconscious Bias, part 2

John Randall, the head of King's college where Rosalind Franklin worked was reportedly furious at Perutz for "showing unpublished work to an unacknowledged competitor". Nonetheless, "The authors of the three papers [Wilkins, Crick, Watson, and Franklin], their lab chiefs, and the editors of Nature agreed that all three would be published in the same issue." While Elkin in their panel discussion contends that Rosalind Franklin "didn't know" she was being "ill-treated", fellow panelist Nicholas Wade contends that

> "... there's absolutely no evidence that she herself believed this to be the case. ... She was definitely in a position to complain if she wished. She had just arranged a new job. She was leaving the King's College department to go to Birkbeck College. We know that she complained vociferously about things she thought were unfair, like being paid less than – at the MRC – at being paid less than men who did the same job. But she never, ever, complained about this. ... Moreover, she became great close friends with Watson and with Crick. ... unlikely – if in fact she felt they had stolen her discovery. She must have known that they were using her data because there were no other data – her data are acknowledged in Crick's paper."

(a) Given the often encountered representation that the Nobel Prize was "stolen" from Rosalind Franklin, and the now universal consensus of the relevance and contribution of her work, do you think it's possible that Rosalind Franklin could (or should?) have been more assertive in seeking recognition for her work?

REPLACE THIS TEXT WITH YOUR ANSWER

(b) Assuming Rosalind Franklin should have sought more recognition for her work during her lifetime, what reasons do you think might have contributed to her not doing so?

REPLACE THIS TEXT WITH YOUR ANSWER

(c) What circumstances do you think lead to neither Watson and Crick, Wilkins and nor John Randall nor Max Perutz, nor the relevant lab chiefs and the editors of Nature standing up for the acknowledgement of Rosalind Franklin at the critical points leading up to the 1953 publication? What could be happening that so many people did not think to do so?

REPLACE THIS TEXT WITH YOUR ANSWER

- *Note: New York Times science reporter Nicholas Wade and preeminent Rosalind Franklin scholar Lynne Osman Elkin in their panel discussion appear to agree that Francis Crick did in fact generally work to ensure Rosalind Franklin was acknowledged and recognized. It is generally accepted that Crick's commentaries and memoirs were crucial for raising awareness of the contribution and relevance of Rosalind Franklin's work.*

Question 9: Overt Sexism

We've seen that Rosalind Franklin was moving jobs. From the commentary:

> "What exactly was Franklin's research, and how did Watson and Crick gain access to it? While they were busy building their models, Franklin was at work on the DNA puzzle using X-ray

> crystallography, which involved taking X-ray photographs of DNA samples to infer their structure. By late February 1953, her analysis of these photos brought her close to the correct DNA model.
>
> But Franklin stopped her work on DNA because she was frustrated with a strained environment at King's, one that pitted her against her colleagues. In an institutional culture that barred women from the dining room and other social venues, she was denied access to the informal discourse that is essential to any scientist's work. Seeing no chance for a tolerable professional life at King's, Franklin decided to take another job. As she was preparing to leave, she turned her X-ray photographs over to her colleague Maurice Wilkins."

And from a nature profile:

> 'Friends and close colleagues considered Franklin a brilliant scientist and a kindhearted woman. However, she could also be short-tempered and stubborn, and some fellow scientists found working with her to be a challenge. Among them was Maurice Wilkins, the man she was to work with at King's College.
>
> A misunderstanding resulted in immediate friction between Wilkins and Franklin, and their clashing personalities served to deepen the divide. The two were to work together on finding the structure of DNA, but their conflicts led to them working in relative isolation. While this suited Franklin, Wilkins went looking for company at "the Cavendish" laboratory in Cambridge where his friend Francis Crick was working with James Watson on building a model of the DNA molecule.'

And from another article:

> "Wilkins was out of town when Rosalind Franklin was hired, and their first meeting did not go well. He mistook her for a new secretary, and it appears that they never established a good working relationship as fellow scientists. Franklin had gotten the impression she would be working alone in the lab; Wilkins saw her in a more supportive role to his work. In Wilkins' later memoir, he referred to Franklin as 'Rosy', complaining that she had an attitude and was too independent."

and the same article goes on to note a report 'that Watson "did not like Franklin" ', so all of this may shed light on why Rosalind Franklin was not involved in the apparent collegial relationship that was shared between Watson and Crick and Wilkins.

(a) Does the kind of environment Rosalind Franklin was working seem like one in which she might not be given appropriate recognition and acknowledgement? Why or why not?

REPLACE THIS TEXT WITH YOUR ANSWER

(b) How much of Rosalind Franklin's experience do you feel boils down to "personality" versus "sexism"?

REPLACE THIS TEXT WITH YOUR ANSWER

(c) How much of the fact that Rosalind Franklin was not consulted and kept "in the loop" do you feel can be explained by the timing of her change of position and research project?

REPLACE THIS TEXT WITH YOUR ANSWER

(d) What possible changes to the contextual situation in which Rosalind Franklin found herself would have been conducive to her receiving a more appropriate level of recognition and acknowlegement for her work?

REPLACE THIS TEXT WITH YOUR ANSWER