Quiz 8 — 3/28/2024

Instructions

This take-home quiz is due on **Thursday, March 28 at 23:59**.

You may use your own course materials, as well as any materials directly linked from the course website. **No collaboration allowed.**

Type your answers **directly in this Jupyter notebook**, and submit this notebook (just the `ipynb` file) using the submission form on the course website.

Problem 1

The dataset `Diamonds` in the `Stat2Data` library has information on several variables for 351 diamonds.

For this problem, we will focus on three variables: the total price of the diamond in dollars ($TotalPrice$), the size of the diamond in carats ($Carat$), and the depth of the cut as a percentage of the diameter ($Depth$).

Run the cell below to load and preview this data.

In [1]:
```
library(Stat2Data)
data(Diamonds)
head(Diamonds)
```

A data.frame: 6 × 6

	Carat	Color	Clarity	Depth	PricePerCt	TotalPrice
	<dbl>	<fct>	<fct>	<dbl>	<dbl>	<dbl>
1	1.08	E	VS1	68.6	6693.3	7228.8
2	0.31	F	VVS1	61.9	3159.0	979.3
3	0.31	H	VS1	62.1	1755.0	544.1
4	0.32	F	VVS1	60.8	3159.0	1010.9
5	0.33	D	IF	60.8	4758.8	1570.4
6	0.33	G	VVS1	61.5	2895.8	955.6

a.

Fit a linear regression model to predict $TotalPrice$ based on $Carat$ and $Depth$.

Provide **only** the summary output for this part.

In [ ]:
```

```

Feedback. Note that the problem asks you to predict $TotalPrice$ based only on $Carat$ and $Depth$. You did not need to include the interaction term $Carat \times Depth$.

b.

Compute the variable inflation factors for each predictor in the model you fit in part a. Using the rule of thumb we covered in class, should we be concerned about multicollinearity?

In [ ]:

Write your answer here. Double-click to edit.

> **Feedback.** See *Formal detection of multicollinearity* in Lesson 21 for details on how to compute the VIF. Make sure that you have the `car` package installed correctly on your computer.

c.

Fit the complete second-order model predicting $TotalPrice$ using $Carat$ and $Depth$.

Provide **only** the summary output for this part.

In [ ]:

> **Feedback.** See *Using one model to fit two lines with different intercepts AND different slopes* in Lesson 18 Part 2 on how to include interaction terms in R.

d.

Perform a nested F-test to compare the models you fit in parts a and c. Assume a significance level of 0.05.

In the code cell below, write R code to perform the test. You may use any R "shortcut" functions we covered in class.

In the Markdown cell below, write:

- the test statistic,
- the p-value, and
- the conclusion of the test.

In [ ]:

Write your answer here. Double-click to edit.

> **Feedback.** Make sure to correctly specify *which subset of variables* are being tested. Which variables are in the model you fit in part c that are not in the model you fit in part a?

Grading rubric

Problem	Weight
1a	1.0
1b	1.0
1c	1.0
1d	1.0
Max Score	**40**