Module 2 Worksheet

Prepared by: Katherine Daignault and STA302H1 Fall 2025 Teaching Team

Worksheet Information

Goal of the worksheet:

The Module 2 worksheet is an opportunity to practice fitting multiple linear models in R and interpreting the coefficients when using different types of predictors. By completing this worksheet, each student will be developing the skills to achieve the following weekly learning objectives:

- Correctly interpret the estimated coefficients of a multiple linear model in the context of the dataset.

- Apply multiple linear models on various datasets using R statistical software.

- Differentiate the relationships modelled using qualitative predictors, interactions between predictors, and continuous predictors.

This worksheet is important practice for completing questions on the term test and final exam, and for your final project proposal.

Preparation assumed:

For hybrid sections: As part of the flipped design of the course, it is assumed that each student is attending this lecture having completed the following pre-class preparation:

- Watched the Module 2 Videos, attempted the Pre-Class Quiz, and accessed the code provided on the Guided Practice page

- (Optionally) attempted the corresponding LearnR module

For in person sections:

- Please complete this worksheet after attending your in-person lectures for each week. If you did not attend class, please review the annotated slides posted on Quercus which will be posted on the Module 2 Quercus Page.

- It is also recommended that you attempt the corresponding LearnR module, or have it open as a reference.

- Please see the Office Hour Schedule for drop in coding help office hours.

How to complete this worksheet:

- Students may work in groups of 2-3 if desired. However **each student** must submit their worksheet to MarkUs to receive their completion credit. It is recommended that each student work on their own copy of the assignment.

- All the code and course knowledge needed to complete this worksheet has been provided in the pre-class materials. It may help to have these open while working on this document.

- Follow the instructions provided in each question to complete the code.

- DO NOT change the names of the variables that store your final answers.

- When in doubt about a question in the worksheet or your code, ask a TA or the instructor.

- It is recommended you submit your file to MarkUs after you complete 2-3 steps of the worksheet to check your answers.

Steps for submitting to MarkUs:

1. Go to MarkUs and log in using your UofT credentials.

2. Select Worksheet 2 from the assignment list.

3. Under Submissions, upload your Rmd file and select the file name from the list.

4. Go to Automated Testing and select Run Tests to check your worksheet answers.

What to do if a test fails on MarkUs

1. Don't panic. Your work won't really be graded until the deadline, so start early to make sure you have lots of time to resolve issues before the deadline.

2. Read the message to get hints about what the problem is. For example "variable X not present" means that you may have a typo in your variable name.

3. Search on the discussion board to see if other classmates have encountered a similar error (and if not, consider posting a screenshot of the error message).

4. Come to TA or instructor office hours and post on the discussion board with your issue.

**The due date for MarkUs Worksheet 2 is Tuesday, September 16, at 11:59pm*

1) Load the dataset and explore

Load in the dataset `MichelinNY.csv` into R. This dataset contains customer ratings (out of 30) on the Food, Decor and Service characteristics of 164 restaurants, as well as the Price of dinner and an indicator of whether the restaurant has a Michelin Star.

By fitting an appropriate linear model, answer the following questions by replacing the value `NULL` in the variables `Michelin_model1` and `Not_Michelin_model1` with numeric values from the model output rounded to 1 decimal place, with no additional symbols:

1. What is the average price of a dinner in a Michelin star restaurant?

2. What is the average price of a dinner in a non-Michelin star restaurant?

```
# load the data
nyc <- read.csv(file="MichelinNY.csv", header=T)

# fit an appropriate linear model
model1 <- NULL
model1
```

```
## NULL
```

```
# What is the average price of a dinner in a Michelin star restaurant?
Michelin_model1 <- NULL

# What is the average price of a dinner in a non-Michelin star restaurant?
Not_Michelin_model1 <- NULL
```

2) Fit a multiple linear model without Michelin star variable

Fit a multiple linear model using `Price` as a response, and `Food`, `Decor`, and `Service` as predictors. Display your estimated model coefficients, and replace `NULL` in the variable `Decor_model2` with the value for the coefficient of `Decor` (rounded to 1 decimal place).

```
# fit the model with response Price and predictors Food, Service and Decor
model2 <- NULL
model2
```

```
## NULL
```

```
# What is the estimated coefficient for Decor?
Decor_model2 <- NULL
```

Question: What is the interpretation of this value in the context of the data?

TYPE YOUR ANSWER BELOW:

3) Including Michelin star variable as a main effect

Now, update your `model2` by including `InMichelin` as the fourth predictor (alongside `Food`, `Service` and `Decor`). Using this new model (i.e. `model3`), answer the following questions by replacing the value `NULL` in the variables `model3_interpretation1` and `model3_interpretation2` with values from the model output (rounded to 1 decimal place):

1. How much more does a dinner cost on average if a restaurant's Decor rating changes from 25 to 26 but all other ratings and Michelin star status stay the same?

2. What is the average price of a dinner at a non-Michelin star restaurant who has no customer ratings for Food, Service and Decor?

```
# fit the model with response Price and predictors Food, Service, Decor and InMichelin
model3 <- NULL
model3
```

```
## NULL
```

```
# How much more does a dinner cost on average if a restaurant's Decor rating changes
# from 25 to 26 but all other ratings and Michelin star status stay the same?
model3_interpretation1 <- NULL

# What is the average price of a dinner at a non-Michelin star restaurant who has
# no customer ratings for Food, Service and Decor?
model3_interpretation2 <- NULL
```

Question: What is the interpretation of the coefficient of `InMichelin` in the context of the data? Write your answer in a way such that someone who has never taken a statistics course can understand.

TYPE YOUR ANSWER BELOW:

4) Including Michelin star variable as interaction

Alternatively, update your `model2` by including `InMichelin` as an interaction with `Decor` alongside `Food`, `Service` and `Decor` main effects. Using this new model (i.e. `model4`), replace `NULL` in the variable `model4_interpretation` with the value from the model output (rounded to 1 decimal place) representing the average Price of dinner for a non-Michelin star restaurant resulting from a one-unit increase in Decor rating when all other ratings stay the same.

```
# fit the model with response Price, predictors Food and Service, and interaction of Decor and InMichelin
model4 <- NULL
model4
```

```
## NULL
```

```
# value representing the average Price of dinner for a non-Michelin star restaurant resulting
# from a one-unit increase in Decor rating when all other ratings stay the same
model4_interpretation <- NULL
```

Question: What is the interpretation of the coefficient of `Decor:InMichelin` in the context of the data? Write your answer in a way such that someone who has never taken a statistics course can understand.

TYPE YOUR ANSWER BELOW:

5) Making a prediction

Lastly, update your `model3` to include `InMichelin` as both a main effect **and** as an interaction with `Decor` alongside `Food`, `Service` and `Decor` main effects.

Use this model (i.e. `model5`) to determine the mean dinner Price for a non-Michelin star restaurant with Food, Service, and Decor ratings of 15, 20, and 10 respectively. Enter your answer, rounded to 1 decimal place, by replacing `NULL` in the variable `prediction`.

To answer the poll question at the end, replace `NULL` in the variable `interaction_column` with an appropriate vector representing the column of the **X** matrix corresponding to the interaction term in your `model5`. Do this by extracting the **X** matrix (the design matrix) directly from the model using the `model.matrix()` function and selecting the relevant column.

```
# fit the model with response Price, predictors Food, Decor,
# Service, InMichelin, and interaction of Decor and InMichelin
model5 <- NULL
```

```r
# What is the mean dinner Price for a non-Michelin star restaurant with
# Food, Service, and Decor ratings of 15, 20, and 10 respectively?
model5_prediction <- NULL

# Store the column of X corresponding to the interaction term below
model5_interaction_column <- NULL

# Consider the X matrix for `model5` and the column representing the interaction
#term that you stored above. How many values in this column will be non-zero?

model5_interaction_nonzero <- NULL
```

END OF WORKSHEET - BE SURE TO SUBMIT YOUR WORKSHEET ON MARKUS TO RECEIVE COMPLETION CREDIT