# Modeling Assignment 2

November 14, 2025

Pavitra Bhargavi Allamaraju, Rowan Cake, Phil Sukparangsee, and Eric Yee

`packages imported`

## 1 Problem Statement

Construct a mathematical model to estimate the **total annual income** of Loafe Café at both the Alumni Centre and Sauder locations as a function of daily temperature, daily precipitation, daily humidity and frequency of customers.

## 2 Variables and Parameters

| Symbol | Description | Type | Dimension | Units |
|---|---|---|---|---|
| $d$ | day index within a (non-leap) year, $d \in \{1, \ldots, 365\}$ | independent variable | 1 | — |
| $T_d, P_d, H_d$ | temperature, precipitation, humidity on day $d$ | independent variable (weather input) | $T, L, 1$ | °C,mm,% |
| $C_d$ | expected total customers on day $d$ (both locations) | dependent variable (model output) | 1 | customers/day |
| $s$ | average spend per customer | parameter | 1 | \$/customer |
| $R$ | total annual income | derived aggregate output | 1 | \$/year |
| $T_0, P_0, H_0$ | reference temperature, precipitation, humidity level | parameter (reference level) | $T, L, 1$ | °C,mm,% |
| $C_0$ | baseline customers at reference weather $(T_0, P_0, H_0)$ | parameter | 1 | customers/day |

| Symbol | Description | Type | Dimension | Units |
|---|---|---|---|---|
| $a$ | rate of change of customers w.r.t temperature (squared) | parameter (sensitivity) | $T^{-2}$ | customers / (°C)$^2$ |
| $b$ | rate of change of customers w.r.t precipitation | parameter (sensitivity) | $L^{-1}$ | customers / mm |
| $c$ | rate of change of customers w.r.t humidity (squared) | parameter (sensitivity) | 1 | customers / (%)$^2$ |

## 3 Assumptions and Constraints:

- Menu prices and the amount of money individual customers spend remain constant year-round.
- On average, customers at the Alumni Centre spend the same amount of money as customers at the Sauder building. This allows us to combine the customer data of the two.
- External factors such as the school year schedule do not impact total revenue.
- Weather trends from 1997-2024 apply to the present.
- A non-leap year will be used to model annual revenue.
  - Leap year data from the original datasets can still be used to construct the model since we sample weather behaviour from a month's general distribution. Therefore, we can take 365 samples to estimate the total number of customers for a non-leap year.
- Reference values are the mean values from the entire dataset, as these are the points from which deviating weather measurements would cause a difference in the expected number of customers.

## 4 Building the Solution:

We estimate total annual revenue of Loafe by:

1) Loading the provided datasets and doing some light exploratory data analysis.

2) Constructing probability density functions for temperature, precipitation, and humidity $(T, P, H)$ for each month. We achieve this with kernel density estimation over the data set using a gaussian kernel.

3) Estimate the parameters of the daily customers equation from data.
   - Estimate $a, b, c, C_0$ using the January/July customer dataset.
   - Estimate $T_0, P_0, H_0$ using the entire weather dataset.

4) Simulate a year of revenue by sampling weather conditions for each day (eg. if a day is in March then we sample $(T_d, P_d, H_d)$ off of each respective March PDF). We then use those weather samples $(T_d, P_d, H_d)$ as inputs to the daily customer function and then multiply the daily customers by an average purchase amount. We do this over all 365 days of the year.

5) Finally, we perform uncertainty analysis via Monte Carlo methods on our yearly revenue calculations to get a range of plausible values.

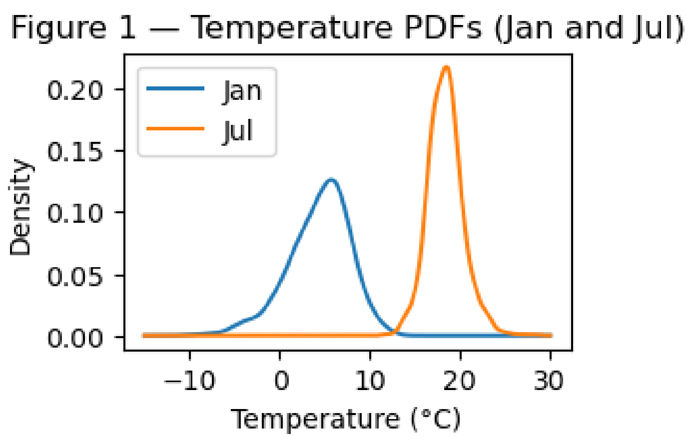
### 4.1 Data Processing

In the code cell below we combine both Loafe locations customer data into a January total and a July total respectively.

### 4.2 Build Monthly PDFs

We split weather data by month and fit Gaussian KDEs for temperature, precipitation, and humidity. We use KDEs to build these monthly weather PDFs (for temperature, precipitation, humidity) because KDEs give a smooth probability density function without having to make any assumptions about what type of distribution the data follows.

**Figure 1** overlays **January and July** PDFs for temperature so we can visually confirm that the winter/summer weather patterns that will drive the customer model are different. We only display temperature for sake of conciseness.




Specify the daily customer amount equation as

$$C_d(T_d, P_d, H_d) = \max\{\, C_0 + a\,(T_d - T_0)^2 - b\,(P_d - P_0) - c\,(H_d - H_0)^2,\ 0\,\}$$

- We chose $T_0, P_0$, and $H_0$ to be the mean value for each of the respective weather measurements over the entire dataset. The justification for this is that from the structure of the equation if the inputted values $(T_d, P_d, H_d)$ are exactly equal to each respective parameter ($T_0, P_0$, and $H_0$) the amount of customers for that day will just be equal to $C_0$. Therefore the reference values should represent what typical weather looks like, and the mean value is the statistic that does that.
- We get **Jan/Jul mean weather** from the KDEs (numerical mean via integration on a fine grid).

- We estimate $a, b, c$ using a finite-difference idea based on the **change in average daily customers** from January to July.
- We chose $C_0$ to be the mean between the average daily number of customers for January and the average daily number of customers for July.

### 4.2.1 Estimating the coefficients:

We know the customer response model as defined above:

$$C_d(T_d, P_d, H_d) = C_0 + a(T_d - T_0)^2 - b(P_d - P_0) - c(H_d - H_0)^2$$

where $C_0$ is the baseline value at reference weather $(T_0, P_0, H_0)$, and $a, b, c$ represent sensitivities to temperature, precipitation, and humidity.

Let $\Delta C = C_{Jul} - C_{Jan}$ be the observed change in average daily customers between January and July, let $(T_{Jan}, P_{Jan}, H_{Jan})$ and $(T_{Jul}, P_{Jul}, H_{Jul})$ denote the corresponding mean weather conditions.

Coefficient $a$: Assume, for the purpose of isolating the effect of temperature, that precipitation and humidity remain at their reference levels while temperature moves from its January mean to its July mean. Then the change predicted by the model is

$$\Delta C \approx a[(T_{Jul} - T_0)^2 - (T_{Jan} - T_0)^2] \implies a = \frac{\Delta C}{(T_{Jul} - T_0)^2 - (T_{Jan} - T_0)^2}$$

Similarly for coefficient $b$ and $c$, isolating precipitation and humidity respectively gives us

$$b = \frac{\Delta C}{(P_{Jul} - P_0) - (P_{Jan} - P_0)}, \quad c = \frac{\Delta C}{(H_{Jul} - H_0)^2 - (H_{Jan} - H_0)^2}$$

For each, we treat the change in customers from January to July as if it were caused by the change in just one weather variable (temperature, precipitation, or humidity), keeping the others fixed at their typical values. Solving the model under these simplified comparisons gives the formulas for $a, b$, and $c$. This produces reasonable, interpretable estimates given the limited data.

```
Reference mean averages: T0=10.60°C, P0=3.19mm, H0=78.10%
Jan vs Jul daily customers: 350.4 → 175.4
a=-8.121, b=36.557, c=-14.826, C0=262.9
```

## 4.3 Yearly Simulation

We simulate a **typical** (non-leap) year by: - mapping day $d \in \{1..365\}$ to its **month**,
- drawing a single $(T, P, H)$ sample from that month's PDFs,
- applying $C_d(T_d, P_d, H_d)$, and
- multiplying by a constant \$ $s$ **spend per customer**.

**Note:** At the Sauder Loafe location, the average price for a food item is \$13.73 and the average price for a drink is \$5.74. If we assume that 25% of customers buy only food, 50% buy only a drink, and 25% both food and a drink, we get that the average customer would spend

$$s = 0.25 * (\$13.73) + 0.5 * (\$5.74) + 0.25 * (\$13.73 + \$5.74) = \$11.17.$$

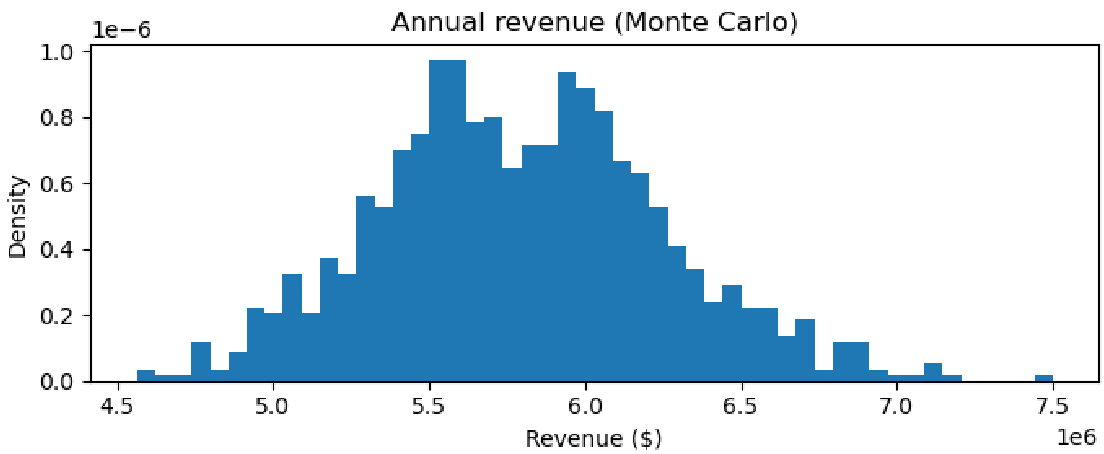
This produces a **sample** $R$ of annual revenue that follows the equation

$$R = \sum_{d=1}^{365} C(T_d, P_d, H_d) \times s.$$

```
Sample revenue/year (s=$11.17): $5,715,629
```

### 4.4 Monte Carlo Estimation

**Below is an Monte Carlo simulation of our yearly revunue:**

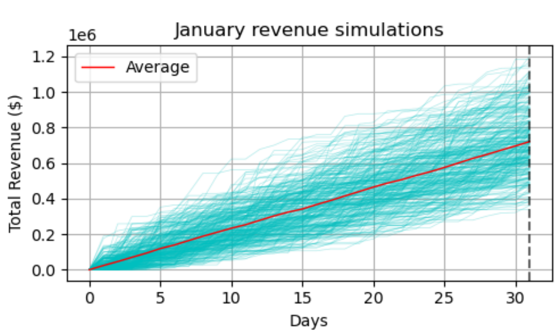


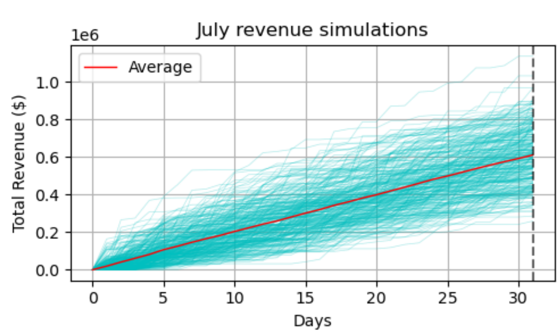

```
Mean annual revenue: $5,797,884 |Median annual revenue: $5,786,658 | 90% CI:
[$5,055,835, $6,589,732]
```

**Below is the error analysis Monte Carlo simulation:**

```
Uncertainty / Error Scales Used in Monte Carlo
--------------------------------------------------
T0 error ( _T0): 5.840  °C
P0 error ( _P0): 6.473  mm
H0 error ( _H0): 9.464  %
C0 error ( _C0): 435.738  customers/day
```

**Figure 3 and 4:** Monte Carlo simulation with all errors included for total cumulative revenue gained in January and July.



```
Std dev of final revenue (January): 218010.46542098687
Std dev of final revenue (July): 186868.48156183792
```

## 5 Analysis and Assessment:

From our Monte Carlo simulation of total annual revenue, the mean annual revenue for Loafe across all locations is $\$5,797,884$ per year with a median value of $\$5,786,658$ and a 90% confidence interval of $[\$5\,055\,835, \$6\,589\,732]$. Since the number of customers each day is dependent on random weather variables, our use of Monte Carlo analysis is justified in order to construct a range of plausible estimates. Loafe makes around $\$15,870$ per day, which implies that they have around 1356 customers per day who spend \$11.17 on average. Although Loafe operates two locations with high foot traffic, these values are unreasonably high as 1356 is much higher than our $C_0$ estimate of 262 daily customers. This indicates that something may be inaccurate in our parameter calculations for $a, b$, and $c$.

By isolating the effects of the parameters' uncertainty on total revenue in hidden exploratory plots, Monte Carlo simulations show that the model is equally sensitive to all weather features. However, when all errors are included, the standard deviation of total revenue gained is $224,624.38$ and $184,429.82$ for January and July. Therefore, the cafe is incentivized to measure each model's parameters carefully.

Concerning assumptions, the model doesn't account for the school year schedule and seasonal patterns. The number of people on campus is usually lower during the summer months, so the number of people would influence revenue more than the weather. With access to more accurate data on customer spending habits, we could better estimate average revenue per customer, which could also potentially vary by season.