EM 618: Optimization for Data Science

Take-Home Exam

Optima Foods – Multi-Week Supply & Demand Case Study (Messy Data)

Instructor: Udit Bhatia
Time Allowed: Till May 2 (11:59 PM)

Instructions:

- This is an individual take-home exam. Work independently.
- You may use any programming language/library (e.g. Python with `PuLP`/`ortools`, R with `lpSolve`, Excel Solver, etc.) to clean the data and solve the optimization.
- Your submission should include:
  - A concise written report (roughly 3–5 pages) that documents your *data-cleaning choices*, *model formulation*, *optimization results*, and *interpretations*. Only single PDF file needs to be submitted.
  - Key code snippets or an appendix showing how you handled data, built the model, and performed the analysis. (Raw solver output is not enough; please interpret the results.)
  - Answers to each question in the structure below.
- The data provided below has *missing values*, *inconsistent min–max ranges*, and *multiple weeks* with varying parameters. Real-world data is rarely perfect, so part of this exam is demonstrating how you handle that messiness.

Scenario Overview

You have joined *Optima Foods*, a farm-to-door delivery service that supplies weekly boxes of vegetables (Spinach, Tomatoes, Carrots) to three regional hubs (A, B, and C). You have data for a 3-week planning horizon. Each week has:

- Different production capacities at the farm (often with inconsistent or uncertain numbers).
- Estimated demand ranges at each hub (min–max).
- Varying transport costs (and carbon footprints) to get produce from the farm to hubs.

Your objective is to *minimize total costs* while respecting production limits, meeting hub demands, and possibly considering carbon emissions if you choose. You may solve each week separately or combine them into a single 3-week model.

Data Provided

(1) Production Capacities (Weekly)

Vegetable	Week 1	Week 2	Week 3
Spinach	500	450?	??
Tomatoes	650	580	720 (or 680?)
Carrots	400	400	420?

Notes on Production:

- For Spinach in Week 3, there is no official record (some logs say 480, others say 460).
- Tomatoes in Week 3 could be 720 or 680, depending on weather.
- Carrots in Week 3 could be 420 or 400, depending on farm manager input.

(2) Demand Forecast (Min–Max) for Hubs A, B, C (Weekly)

Hub A (Weeks 1–3):

Vegetable	Week 1	Week 2	Week 3
Spinach	100–200	110–??	100–180
Tomatoes	150–300	160–310	140–290
Carrots	50–100	60–120	80–110

Hint: For Week 2 Spinach at Hub A, the upper bound is missing. You might guess 190 or 210 based on historical range.

Hub B (Weeks 1–3):

Vegetable	Week 1	Week 2	Week 3
Spinach	80–150	90–150	100–140
Tomatoes	100–200	120–220	120–200
Carrots	70–140	70–160	**160–140** (*)

Note: The last row for Carrots in Week 3 has a reversed range (160–140). This is a data-entry error. You need to fix it.

Hub C (Weeks 1–3):

Vegetable	Week 1	Week 2	Week 3
Spinach	50–90	55–95	??–85
Tomatoes	80–160	100–170	110–??
Carrots	40–90	40–??	50–100

Note: Several ?? appear. Students must decide how to fill or handle them.

(3) Transport Costs and Carbon Footprint (Per kg) – Varying by Week

		Hub A	Hub B	Hub C	
Week	**Vegetable**	**Cost**	**Cost**	**Cost**	**Carbon**
3*Week 1	Spinach	0.30	0.35	0.45	0.20
	Tomatoes	0.25	0.40	0.45	0.20
	Carrots	0.20	0.25	0.45	0.20
3*Week 2	Spinach	0.32	0.36	??	0.22
	Tomatoes	0.28	0.42	0.48	0.20
	Carrots	0.20	0.26	0.47?	0.19
3*Week 3	Spinach	0.30	0.38	0.46	0.21
	Tomatoes	0.26	0.39	0.46	??
	Carrots	0.21	0.25	0.45	0.19

Notes:

- Week 2, Spinach to Hub C is missing. Possibly around 0.45–0.48.
- Week 2, Carrots to Hub C says 0.47 with a question mark.
- Week 3, Tomatoes to Hub C has a missing carbon figure (?), presumably around 0.20–0.22.

Exam Questions

Question 1: Data Exploration and Cleaning

- **(a)** Identify each missing or contradictory entry and describe how you addressed it. (E.g., you might take an average of known values, choose a conservative capacity, or correct reversed ranges.)
- **(b)** Show a short descriptive summary: for each week, how many kg of each vegetable can be produced, and what is the plausible demand range at each hub? Summarize the cost structure.
- **(c)** Comment briefly on any assumptions (e.g., "We assume Week 3 spinach capacity is 460 kg," or "We replaced the missing cost with 0.46, matching other data.")

Question 2: Model Formulation

- **(a)** Explain your approach to modeling the 3-week horizon. Are you doing a *separate* optimization per week, or a *single* multi-period model with 3 sets of constraints (one for each week)?

- **(b)** Define decision variables (e.g., $x_{v,h,w}$ for kg of vegetable v shipped to hub h in week w) and write your objective function (minimizing total cost across all weeks, or cost + carbon if you choose a multi-objective).
- **(c)** State the constraints clearly, including production limits, demand min–max, and non-negativity. If you use or consider integer constraints, mention why.

Question 3: Solving and Key Results

- **(a)** Implement your model in a solver. Present the *optimal* allocations for each week: how many kg of each vegetable go to each hub, total cost, and any carbon summary if relevant.
- **(b)** If you run separate week-by-week optimizations, compare them. If you run a single multi-week optimization, present your consolidated results.
- **(c)** Include a short commentary on any surprising findings (e.g., "We end up sending almost no spinach to Hub C," or "Tomatoes are shipped only to Hub A because of cost differences.")

Question 4: Sensitivity or Scenario Testing

- **(a)** Choose one parameter (e.g., cost to Hub C, or production capacity of tomatoes in Week 3). Increase or decrease it by 10%. Does the optimal solution change significantly?
- **(b)** Provide a brief reflection on which parameters are most critical. Is there a certain route, week, or vegetable that strongly influences total cost?

Question 5: Strategic Extensions

- **(a)** Propose one extension to handle *uncertain demands* or *multi-objective* trade-offs (like carbon constraints). How might you incorporate these in your optimization model?
- **(b)** In a short paragraph, summarize how you would present your final recommendation to BrightGro Foods management. Mention the limitations of your model (e.g., ignoring seasonality or daily fluctuations) and how to address them in future iterations.

Hints and Tips (Not Required, But Helpful)

- **Handling Missing Values:**
  - If capacity is missing or contradictory, *consider a "worst case" or "most likely case"* to be safe.

  - For demand min–max ranges that are reversed or incomplete, *adjust them logically* (e.g., if 160–140 is reversed, fix it to something like 120–160 based on context).

- **Modeling Per Week vs. Multi-Week:**
  - *Week-by-week approach:* You solve 3 separate linear programs (one for each week).
  - *Single multi-week approach:* You keep separate constraints for each week but sum cost across all 3 weeks in the objective.
  - There's no single "right" answer; the choice might depend on whether leftover produce can be carried over or not, or if you treat each week independently.

- **Interpretation:**
  - Don't just give numeric results. *Explain* which constraints are "active" (i.e., used up capacity or fully met the min demand) and how that shaped the solution.
  - If certain routes are never used, discuss why (maybe they're too expensive or there's not enough demand).

- **Sensitivity Analysis:**
  - If you see minimal changes in the solution after adjusting a parameter by 10%, that suggests a stable solution.
  - Large changes might indicate you should refine data estimates or constraints because the system is highly sensitive to that parameter.

- **Carbon Emissions:**
  - You can treat carbon as a second objective or as a constraint (e.g., "Keep total CO_2 under X").
  - If you keep it in the main objective, decide how you weight cost vs. carbon.

- **Practical Relevance:**
  - Mention real-world issues: produce perishability, partial truckloads, daily variability.
  - *Industry context:* Typically, companies re-run such optimizations periodically (weekly, monthly) as data updates.

Submission Requirements:

- A short PDF or Word report containing your solutions, reasoning, and reflections.
- Attach or link to your code. Make sure your code is *commented* so we can see how you handled data cleaning and modeling steps.
- Clearly label your answers to each question (1 through 5).

Grading Criteria:

- **Data Handling & Consistency:** How well did you address the missing/contradictory data?
- **Model Formulation:** Are your variables, objective, and constraints correct and clearly explained?
- **Solution Interpretation:** Do you provide insights into why the solution is optimal and how changes affect it?
- **Depth of Analysis:** Did you go beyond a raw solver output? Is your sensitivity analysis informative?
- **Professional Communication:** Clarity of writing, charts, explanation for a non-technical stakeholder.