SAS Optimization Challenge: Case Study

Title: HELP YOUR OPTIMIZATION PROFESSOR!!!!!

TIER 5 (of 5)

Tier 5: Modeling Details

Your professor has always made the Practice exam optional to his students. Even though students who take the Practice exam tend to perform better on the Live exam, historically only about half of his students have chosen to take the Practice exam.

Your professor is reluctant to make the Practice exam mandatory since it doesn't count towards their final grade. Instead, he wants to find a way to incentivize more of his students to voluntarily take the Practice exam. To do this, he wants to allow between 10% and 20% of the questions to appear on both the Practice and Live exams.

His idea is that if students know ahead of time that between 5-10 questions on the Practice exam will also be on the Live exam, perhaps more students will opt to take the Practice exam.

One of the requirements you built into the Tier 4 model was that a question can appear on, at most, one exam. In other words, the Practice and Live exams generated from your Tier 4 model do not contain any overlapping questions.

Hint (V.) in the Tier 4 case study PDF suggested you create a conflict constraint to prevent exam questions from appearing on both exams. Relaxing this constraint to accommodate overlapping questions is the primary objective for Tier 5.

The phrase *relaxing a constraint* in decision science refers to making a constraint less strict, and involves loosening its restrictions, which can lead to a new optimal solution that might not have been feasible in the original problem. The conflict constraint from Tier 4 is quite strict. It allows a question to appear in either the Practice exam, the Live exam, or neither exam. Those are the only three possibilities.

In Tier 5, you need to relax that constraint so that a question may appear in the either the Practice exam, the Live exam, both the Practice and Live exams, or neither exam, while

ensuring the total number of overlapping questions is between 10% and 20% of the total exam questions.

The Tier 5 objective is to find the two ***most difficult***, ***balanced*** exams from the question bank, while ensuring the following requirements are satisfied ***for each exam***:

I. Each *topic objective* contains the correct number of questions for each exam.
II. No more than one question from each frenemy group is assigned to an exam.
III. The eight (8) questions in the **Force_Status** column are in the Live exam.
IV. Between 10% and 20% of exam questions are contained in both the Practice and Live exams.
    - In other words, since each exam contains fifty (50) questions, between 5 and 10 questions must appear on both exams.

The eight (8) questions forced into the Live exam are allowed to appear on the Practice exam as well, although they aren't required to.

Use the `create data` statement to generate one (1) output SAS data set containing only the questions your model assigned to the Practice or Live exams. Questions assigned to both exams will appear twice, once for the Practice exam, and once for the Live exam. Include the following columns in the output data set: **Exam, Question**, **Frenemy_Group**, **Topic**, **Objective**, **Force_Status**, **Correct**, and **Overlap**. Questions that were not chosen should not appear in your output data set. The **Overlap** column is a binary column indicating whether a question appears on both exams (1 = yes, 0 = no).

Additionally, report the *average percentage correct* for both exams (i.e., take the average of the **Correct** column) across all fifty (50) questions chosen by your model for each exam. To be balanced, these values must be identical, and can be calculated by any means necessary (e.g., in SAS, Excel, calculator, etc.).

To do this in SAS, copy and paste the following SQL procedure code at the end of your program, and replace *<insert output data set name>* with the name of your output data set that you created in the OPTMODEL procedure.

```
proc sql;
 select distinct exam,
                 count(distinct Question) as Questions,
                 avg(Correct) as Avg_Correct
  from <insert output data set name>
   group by exam
    order by exam;
quit;
```

Tier 5: Getting Started

Copy and paste your Tier 4 solution (i.e., DATA step and datalines code creating the two data sets, the DATA step to create the **Force_Status** column, and OPTMODEL code) into a new programming editor to begin.

Modifications to the two data sets are not needed.

Tier 5: Hints

The hints below are provided to help overcome relatively nuanced programming syntax and model formulation barriers that may arise in Tier 5.

I. Remove the conflict constraint from your Tier 4 solution. You can either delete it or comment it out in your program using /* and */. Green text inside of the SAS programming editor indicates commented code.

(e.g., `/* my commented Tier 4 conflict constraint */`)

II. In Tier 4, you should've declared a binary decision variable indexed by both QUESTIONS and EXAMS. Each of these binary decision variables were used to determine whether a question should be assigned to an exam. For example, should `ITEM_01.02.04` be assigned to the Practice exam: yes or no? Should `ITEM_01.02.04` be assigned to the Live exam: yes or no? Etcetera for all questions and exams. Suppose you named this set of decision variables `Assign` (i.e., `var Assign{QUESTIONS,EXAMS} binary;`).

To accommodate overlapping questions in Tier 5, you could introduce a new binary decision variable for each question called **Overlap** (i.e., `var Overlap{QUESTIONS} binary;`), and enforce the following constraint:

```
con overlap_q{<q,f,t,o> in QUESTIONS}:
  Assign[q,f,t,o,'Practice'] * Assign[q,f,t,o,'Live']
   = Overlap[q,f,t,o];
```

In other words, for each question, **Overlap** equals 1 if and only if that question is assigned to both the Practice and Live exams. Otherwise, **Overlap** equals 0.

While these are valid and logically correct constraints, the problem is that the product of two decision variables causes this set of constraints to be non-linear.

For example, if you ran this using the mixed-integer linear programming solver (i.e., `solve with milp;`), you would receive the following error: `ERROR: The specified optimization technique does not allow nonlinear constraints.`

If you tried to accommodate these constraints by calling the non-linear programming solver (i.e., `solve with nlp;`), you would receive the following error: `ERROR: The NLP solver does not allow integer variables.`

Fortunately, this set of constraints can be reformulated as the linearized equivalent to the product of two binary decision variables. Once you reformulate the model to linearize this constraint, the mixed-integer linear programming solver can easily accommodate the new Tier 5 requirements and solve the problem.

The question you need to find the answer to (and incorporate into your OPTMODEL program) is, "How do you linearize the product of two binary decision variables"?

Happy Googling and ChatGPT'ng.

III. Include a new constraint limiting the total number of overlapping questions to be between 10% and 20% of the total number of exam questions, fifty (50).

You can formulate the lower and upper bounds of this constraint using 5 and 10, or 10% and 20%. It doesn't matter as long as the constraint works as intended.

IV. For the new constraint mentioned in (III.) above, feel free to hard key those lower and upper bound values directly into the constraint. If you've made it this far, I'm confident you know how to declare them as parameters, read them in from SAS data sets, etc. Instead, just hard key them in and use the time you saved to focus on the more important parts of Tier 5.

V. In the Tier 4: Hints section of the Tier 4 document, hint (VII.) suggested formulating a new constraint and testing different tolerance values until you found the ***most difficult***, ***balanced*** exams.

Now that your model formulation has changed to allow questions to overlap, you will need to find new tolerance values to generate the new ***most difficult***, ***balanced*** exams.

Since between 10% and 20% of questions are allowed to overlap, exam difficulty should increase as a result.

Similar to Tier 4, re-tuning the tolerance parameter should be **the very last thing you do** once all other model reformulations have been made.

VI. Once the model has been linearized and solved, you can multiply the optimal values of two binary decision variables together inside of the `create data` statement to create a new, more interpretable **Overlap** column in your output data set.

(e.g., `Overlap = (Assign[q,f,t,o,'Practice']*Assign[q,f,t,o,'Live'])`)

Recall that **Overlap** is one of the required columns in the output table described in Tier 5: Modeling Details above.

You're allowed to define **Overlap** this way because it's being defined *outside* of the model formulation, whereas the constraint shown above in (II.) attempted to define **Overlap** *within* the model formulation, which violated the linear constraint requirement of the mixed-integer linear programming solver.

VII. If you made it this far, congratulations! Stay focused and finish strong. It's been an privilege guiding you through the SAS Optimization training and case study.

You are the next generation of applied decision scientists. Embrace it. Continue to learn and push yourselves, and never stop asking questions.