Business Analytics: Chapter 1

Navid Sabbaghi

August 2012

1 Introduction

This book introduces the field of business analytics and the application of those tools to a variety of business scenarios. In any business setting, when dealing with a new problem, there are a variety of *decisions* that can be made, a business *objective* of interest, *data* that must be gathered to assess the costs and benefits of those decisions, and company *constraints* that need to be respected. Business analytics is a set of tools and *models* centered around decisions, objectives, data, and constraints that address how to make the optimal or best decision when faced with a business problem.

Since the field is very mathematical, most students have a certain amount of anxiety when approaching this field. However, relax. This book is practical in nature and will guide you in managing and making better decisions rather than training you to become a mathematician focused on mathematical derivations and manipulations.

How? Decades ago, only trained mathematicians would have been able to apply the techniques. However, technology has advanced at such a pace that user-friendly technologies exist that enable managers to construct, use, analyze, and iterate on these models for the purposes of decision making. For smaller models, spreadsheets are often used to construct models and iterate on them and have *solver* modules that allow one to find and analyze the recommendations. For larger models that require automation, specialized software now exists.

We begin by focusing on a mantra that will become indispensable as we approach the topic of building and doctoring models: DDOC (Decisions, Data, Objective Function, Constraints).

Decisions, Data, Objective Function, Constraints

In any firm, there are *decisions* that need to be made for the products and services they produce. For example,

- When developing new products for a customer segment, what is the best production mix?
- How should we allocate our advertising budget among various advertising channels?
- How much should in invest in various stocks when deciding an optimal stock portfolio?
- How much inventory should we order this week?
- Which environmental projects should we take on to achieve a certain emission reduction?

For all of those problems and the feasible decisions, the firm first needs to collect *data*:

- For a new product decision making problem, it would help to know the production resources available and the profit margins.

- For an advertising decision making problem, it would help to know the budget, the effectiveness of the various advertising medium, and the costs of advertising in various medium.

The data provides relevant contextual information that is needed for writing down the *objective function* that is important to the firm, and the *constraints* that the decision must satisfy.

The objective function for a firm might be to maximize profit, or minimize cost, or minimize emissions, or some combination thereof. The objective function is one way for a firm to describe what factor is important in its decision making. The constraints are another means to describe the important factors. For example, in the new product development example, the firm may have a limited production capacity and would want to make a decision that does not require uses beyond the available capacity. In the advertising problem, the firm may require certain sales revenue to be met as a result of the decision.

Our goal is to help you in becoming proficient in building both mathematical and spreadsheet decision-making models. When building a decision-making model for a business problem, there are four questions you should always ask yourself to guide you in efficiently building, and iterating on a model:

1. What are the *decision variables* or knobs over which we have control?
2. What are the relevant *data*? Relevant for the purposes of describing the business objective and business constraints.
3. What is the business goal or *objective function* of the firm? Is the firm trying to maximize profit, minimize cost, minimize emissions, or some combination thereof?
4. Which business *constraints* is the firm facing? Do they have a limited resource, e.g., budget? Or, for example, do we have a limited capacity in our factories or call center? Is there a required amount of benefit? For example, do we have a sales requirement that must be met?

Asking yourself these questions will guide you in building a model without getting stuck and frustrated. Let's illustrate DDOC in more detail.

2 Case: Innovation Engines Co. Product-Mix

Tim Falcon is manager of new product development at Innovation Engines Company. The company is a small but innovative company that produces an assortment of energy producing devices, ranging from diesel engines to wind turbines.

A year ago, Tim was tasked with leading the development of a new diesel engine and a new wind turbine with innovative features that the industry was missing. Tim's team has succeeded and now the company is tasked with commencing the production of these products.

Decision variables

The main question for the Innovation Engines Company is how many diesel engines to produce per week and how many wind turbines to produce per week. Those are the decision variables or knobs that the company has at its disposal. We will denote these decision variables by D and W, respectively.

However, in order to decide the correct product-mix, Innovation Engines needs to bring together the leaders from manufacturing, marketing, and new product development to pinpoint the data relevant for finding the best decisions. A business analytics team has been formed to work with these leaders in gathering the data and building the relevant model to make recommendations to management.

Data

The head of manufacturing notes that there are three factories with spare capacity capable of meeting the tolerances needed for these two products. For simplicity let's refer to the factories as Factory 1, Factory 2, and Factory 3. In order to produce one diesel engine, we need 1 hour from Factory 1 and 3 hours from Factory 3. In order to produce one wind turbine, we need 2 hours from Factory 2, and 2 hours from Factory 3. The head of manufacturing also provides the hours available per week at each of the factories.

The head of marketing after consulting with his team of analysts feels confident that they will be able to sell the diesel engines at a profit margin of $300 and the wind turbines at a profit margin of $500. Furthermore, since the production capacity, the marketing head feels confident that the company will be able to sell as many units as it is able to produce at these prices.

The head of new product development emphasizes that we must aim to produce both products so that we move ahead in both sectors and are not left behind by the competition.

We summarize the data in the following table.

Plant	Diesel Engines	Wind Turbines	Available hours per week
1	1 hour	0	4 hours
2	0	2 hours	12 hours
3	3 hours	2 hours	18 hours
Unit Profit	$300	$500	

Objective Function

When confronted with the business objective, the firm's president guides the business analytics team and tasks the team to find the decisions that will maximize profit. Since each diesel engine sold attains a profit of \$300, the total profit from diesel engines is $300D$, and since each wind turbine sold attains a profit of \$500, the total profit from wind turbines is $500W$. As a result, the business objective is to

$$\text{maximize} \quad 300D + 500W.$$

Constraints

The main constraints that the firm faces are the resource constraints at each of the factories. It is important to describe the constraint verbally before trying to describe them algebraically. In particular for each factory we have constraints of the form:

$$\text{hours used at factory} \leq \text{hours available at factory.}$$

For example, for factory 1, since each diesel engine uses 1 hour from factory 1 and wind turbines don't use any hours, we would have the following constraint:

$$D \leq 4.$$

For factory 2, since each wind turbine uses 2 hours from factory 2 and diesel engines don't use any hours, we would have the following constraint:

$$2W \leq 12.$$

For factory 3, since each diesel engine uses 3 hours from factory 3 and each wind turbines uses 2 hours from factory 3, we would have the following constraint:

$$3D + 2W \leq 18.$$

Furthermore, we can not produce a negative number of diesel engines and wind turbines, so we also have the following constraints:

$$D \geq 0, W \geq 0.$$

3 Algebraic model

Having reasoned through our DDOC mantra for the case, we are now able to put together everything and say that we want to find the best values for the decision variables D and W (i.e., the product-mix) that solves the following problem:

$$\begin{aligned} \text{maximize} \quad & 300D + 500W \\ \text{subject to} \quad & D \leq 4 \\ & 2W \leq 12 \\ & 3D + 2W \leq 18 \\ & D, W \geq 0. \end{aligned}$$

We refer to this algebraic description of the problem as the *algebraic model*.

4 LP models

Notice that the objective function $300D + 500W$ is a *linear* function in the decision variables. By that we mean that the objective function is composed as the sum of a bunch of terms, each of which is a constant times a decision variable. Terms such as D^2 or DW or $\sqrt{D}$ are not linear terms.

Furthermore, notice that for each of the constraints we have a left hand side and a right hand side. And each side of those constraints is also a linear function in the decision variables. When the objective function and the constraints are linear in that way, we say that the algebraic model we have created is a *linear programming (LP)* problem or model. The term programming in linear programming has nothing to do with computer programming but rather is synonymous with the term planning or optimization.

5 Simplex algorithm

Why do we care about linear programming problems? Because we have technologies that can solve big ones very quickly. How do we characterize big? In terms of the number of decision variables and constraints. In particular, an algorithm known as the *simplex algorithm* was discovered in the mid-20th century that can solve LP problems with millions of decision variables and millions of constraints on the order of seconds.

In this chapter's case, we have a LP with only 2 decision variables and 5 constraints. But we can understand how the simplex algorithm works by considering how we would solve this simpler problem by hand.

6 Graphical method

This problem involves finding the best values of D and W that maximize the objective function and yet obey the constraints. Since this is a two dimensional problem, we can solve it visually. The process we are about to described is referred to as the *graphical method*. The first step is to visualize the set of possible (D, W) pairs that satisfy all the constraints.

Feasible region

In order to visualize the possible pairs of (D, W) that satisfy the constraints, let's plot the constraints in two dimensions (i.e., with D along the horizontal axis and W along the vertical axis). The non-negativity constraints $D \geq 0$ and $W \geq 0$ are easy to visualize. We can visualize that in two dimensions as the shaded region in Figure 1.

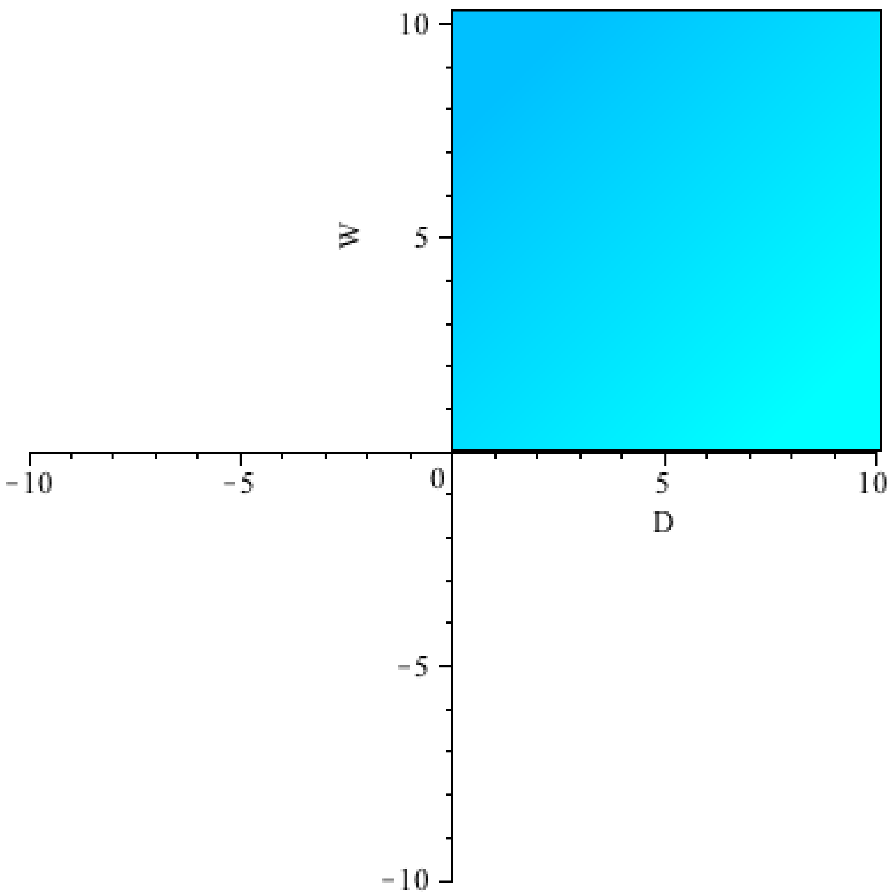


Figure 1: Visualizing the non-negativity constraints

We can visualize the capacity constraint for each factory in isolation. For example, to visualize factory 1's constraint $D \leq 4$, we first draw the line $D = 4$

and then figure out which side of the line corresponds to the inequality $D \leq 4$. We note the relevant region is to the left of the line. Similarly, to visualize factory 2's constraint $2W \leq 12$, we first draw the line $2W = 12$ and then figure out which side of the line corresponds to the inequality $2W \leq 12$. We note the relevant region is below the line. Finally, to visualize factory 3's constraint $3D + 2W \leq 18$, we first draw the line $3D + 2W = 18$ and then figure out which side of the line corresponds to the inequality $3D + 2W \leq 18$. We note the relevant region is the region that includes the origin (0,0). We then plot the intersection of all these constraints or regions. Since this is the set of (D, W) pairs that simultaneously satisfies all constraints, we refer to this region as the *feasible region*. We have drawn the feasible region for this problem in Figure 2.

Iso-profit lines

The second step of the graphical method is to visualize the objective function when it equals a constant, any constant. We refer to the equation $300D+500W =$ constant value as an *iso-profit* line because that line characterizes the set of (D, W) pairs that attain the same profit (hence the prefix iso).

So for example, we could plot the line $300D + 500W = 1500$ and the line would visually depict the set of (D, W) pairs (i.e., the product-mix) that attains a profit of \$1500. If this line intersected the feasible region, that would mean that there are feasible production rates (feasible from the standpoint of our capacity constraints) that attain a profit of \$1500. This particular iso-profit line does indeed intersect the feasible region.

Optimal solution

In order to find the optimal production rate or the optimal solution for our LP, we can then try to increase the profit requirement (i.e., increase the constant value) when we draw a iso-profit line and see if the new iso-profit line still intersects the

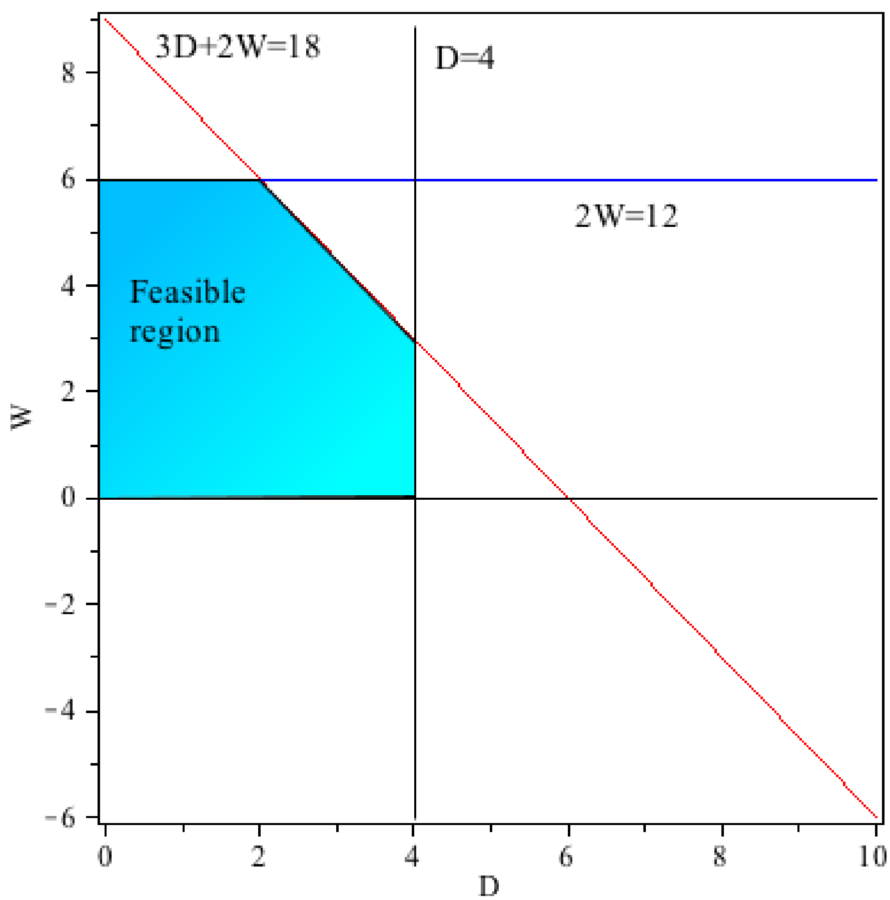


Figure 2: Visualizing the feasible region

feasible region. For example, we can draw the iso-profit line $300D + 500W = 3000$. In Figure 3, we note that this iso-profit line still intersects the feasible region signifying that there are feasible production rates that attain a profit of \$3000. Note that when we increase the constant value, the iso-profit line simply shifts because we are only changing the y-intercept and not the slope. We can continue to increase the profit requirement even further to \$3600 and when we do so, in Figure 3, we notice that the iso-profit line $300D + 500W = 3600$ keeps moving in the same direction but now barely touches the feasible region.

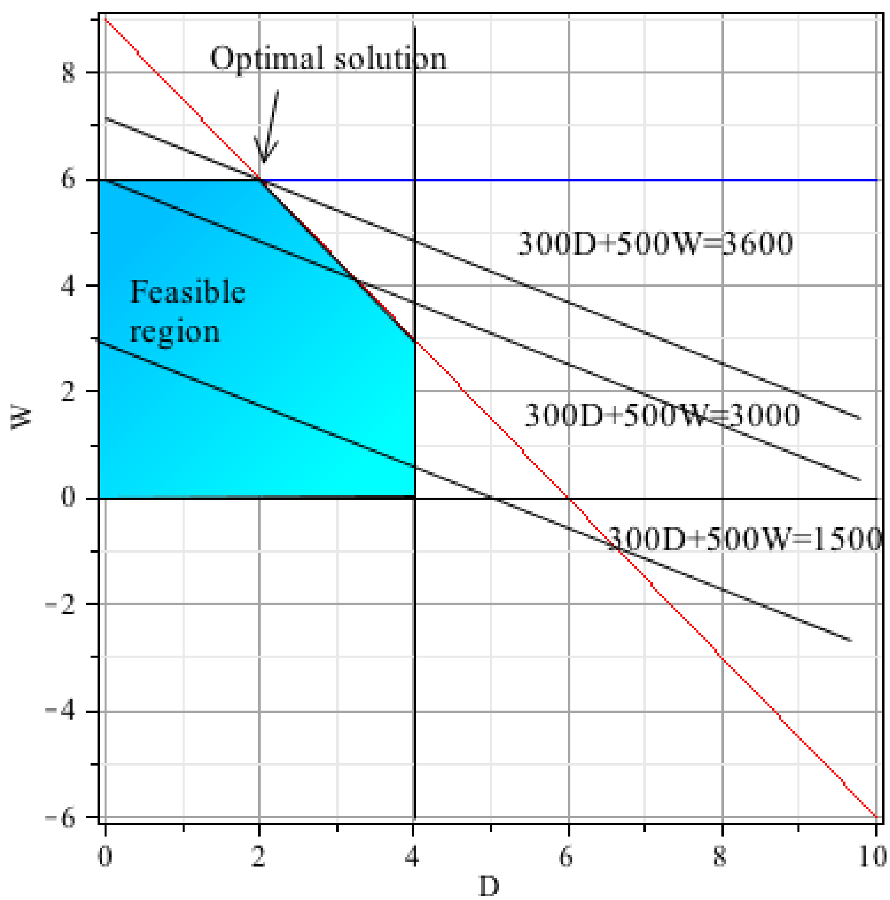


Figure 3: Visualizing the iso-profit lines

In fact it just touches the feasible region at a corner, signifying that the corner representing a production rate of $D = 2$ diesel engines per week and $W = 6$ wind turbines a week attains a profit of \$3600. If we were to increase the profit requirement just a bit more, the line would no longer intersect the feasible region, signifying that the largest attainable profit among all feasible production rates is \$3600 and the optimal production rate (i.e., optimal solution) that attains the maximum profit is $D = 2$, $W = 6$.

Intuition: Simplex algorithm

In general, for a linear programming (LP) problem, the feasible regions tend to be more general shapes called *polyhedrons* with many corners. And just as in this example, the solution that attains the maximum value, if there is one, will always occur at a corner. The simplex algorithm takes advantage of this fact and is a greedy algorithm that searches the corners in the following way.

1. It starts at a corner and looks at neighboring corners to see which corner attains a higher profit,
2. It moves to the better corner, if there is one.
3. Otherwise, it stops and claims to have found the optimal corner.

Of course there is a large amount of formal mathematics that goes into proving under what conditions this algorithm finds the optimal solution. But from a high level, that is all there is to it.

This algorithm is particularly useful for large problems that have many decision variables and many constraints because 1) there is no way we can solve the problem by hand using the graphical method because visualization becomes difficult and 2) the algorithm is very fast at finding the optimal corner if there is one.

7 Spreadsheet model

Setting up the spreadsheet model is similar to setting up the algebraic model. Namely, it is easy to do if we follow the DDOC mantra. In this book we use the color coding convention that blue cells in the spreadsheet represent data, yellow cells represent decision variables and the orange cells represent the objective function value.

Decision Variables and Data

The first step is to put aside blue cells for the data and yellow cells for the decision variables.

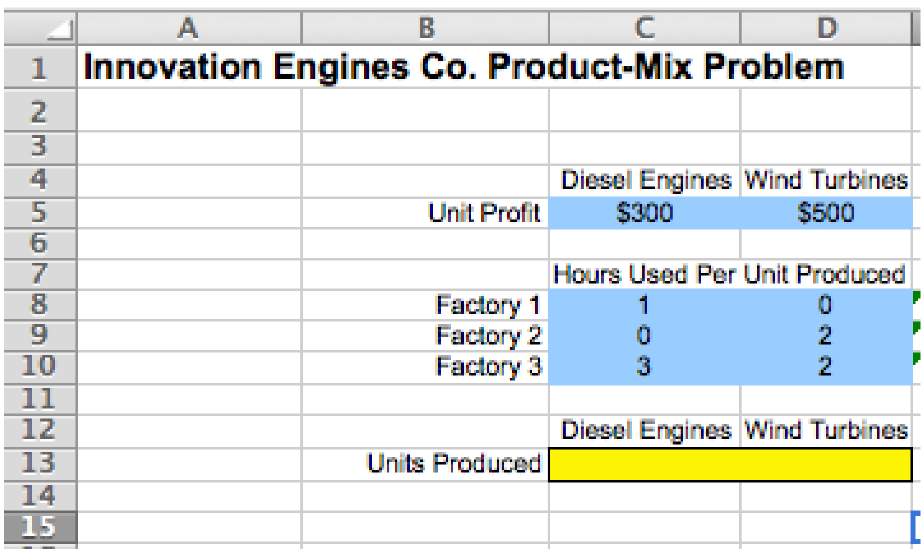


Figure 4: Creating Data and Decision Variable cells

Notice how the data is organized by the decision variables.

Objective function

The next step is to put aside an orange cell for the objective function equation.

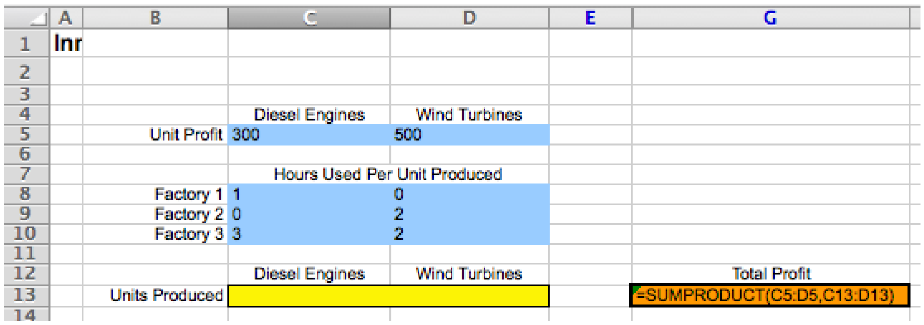


Figure 5: Objective Function

Note that the cell's value starts with an equals sign denoting that the cell is a formula or equation.

Constraints

The final step is to put aside cells for both the left-hand side and right-hand side of each constraint.

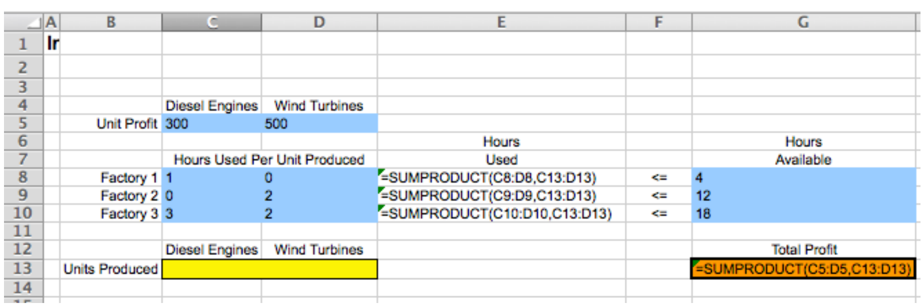


Figure 6: Constraints

Notice that the left-hand side of each constraint is a formula or equation whereas the right-hand side is data. Also note that the $\leq$ sign here is symbolic and serves purely as a reminder of the type of inequality. Finally, noticed how we created this spreadsheet moving from top to down and left to right as we recalled the DDOC mantra.

Trial values

Now we can change the decision variable cells, i.e., the yellow cells and note whether or not the corresponding production mix satisfies the constraints as well as note the corresponding profit. However, we are not interested in just any production mix. Rather we'd like to know the optimal production mix. That is the production mix that maximizes the profit and at the same time satisfies the capacity constraints. Let's see how to use Excel's Solver add-in to complete this step for us.

8 Using Excel Solver

When we have set up our spreadsheet we are ready to ask Excel to find the best values for the decision variables. In other words, we are ready to ask Excel's Solver add-in to use the Simplex algorithm to find the optimal values for D and W. In order to activate the Solver, choose the Solver option from the Data ribbon in Excel.

In order to use the solver, complete the following steps:

1. Set the objective to be equal to the cell that holds the objective function formula.

2. Choose the max radio button since we are maximizing.

3. Set the changing variable cells to be the two decision variable cells.

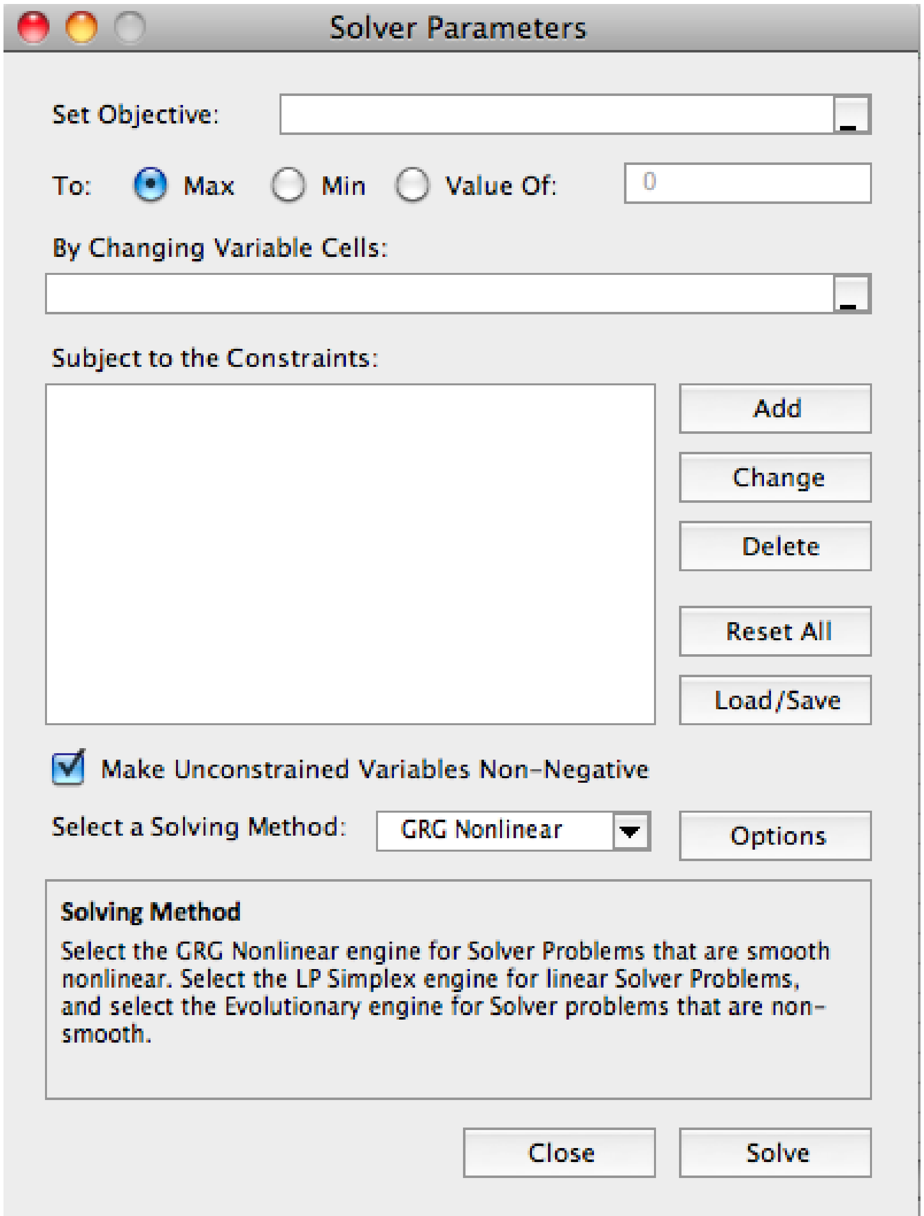


Figure 7: Solver dialogue box

4. Make sure that the Make Unconstrained Variables Non-Negative checkbox is selected. This adds the non-negativity constraints.

5. Select the solving method to be equal to Simplex LP since our problem is a linear programming problem (i.e., the objective function and both sides of the constraints are linear in the decision variables).

6. Finally we need to add the constraints. For each constraint, click on the Add button in section labelled Subject to the Constraints section.

7. After pressing the add button a dialogue box opens up that asks for the cell reference for the left-hand side of the constraint, the type of constraint, and the cell reference for the right-hand side of the constraint. Fill these in for each constraint.

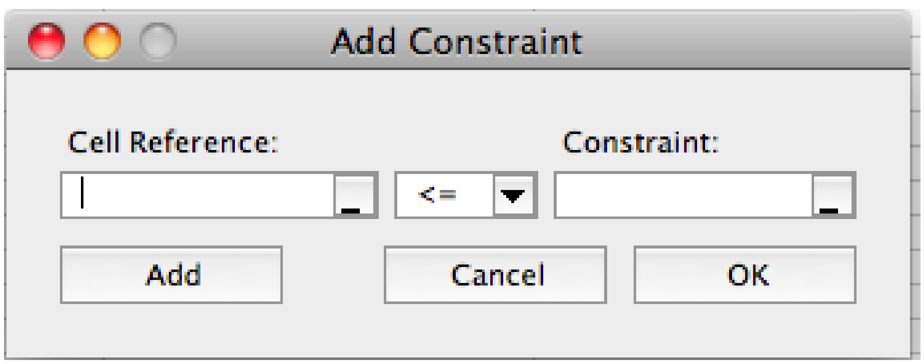


Figure 8: Dialogue box for each constraint

We now have a solver dialogue box that is filled with enough information to find the optimal solution if one exists.

After pressing the solve button, the simplex algorithm will run and if an optimal solution exists will tell us it found it.

And the excel spreadsheet will have the optimal production mix in the yellow cells as expected.

Note that the optimal solution found via solver matches the optimal solution we found via the graphical methods.

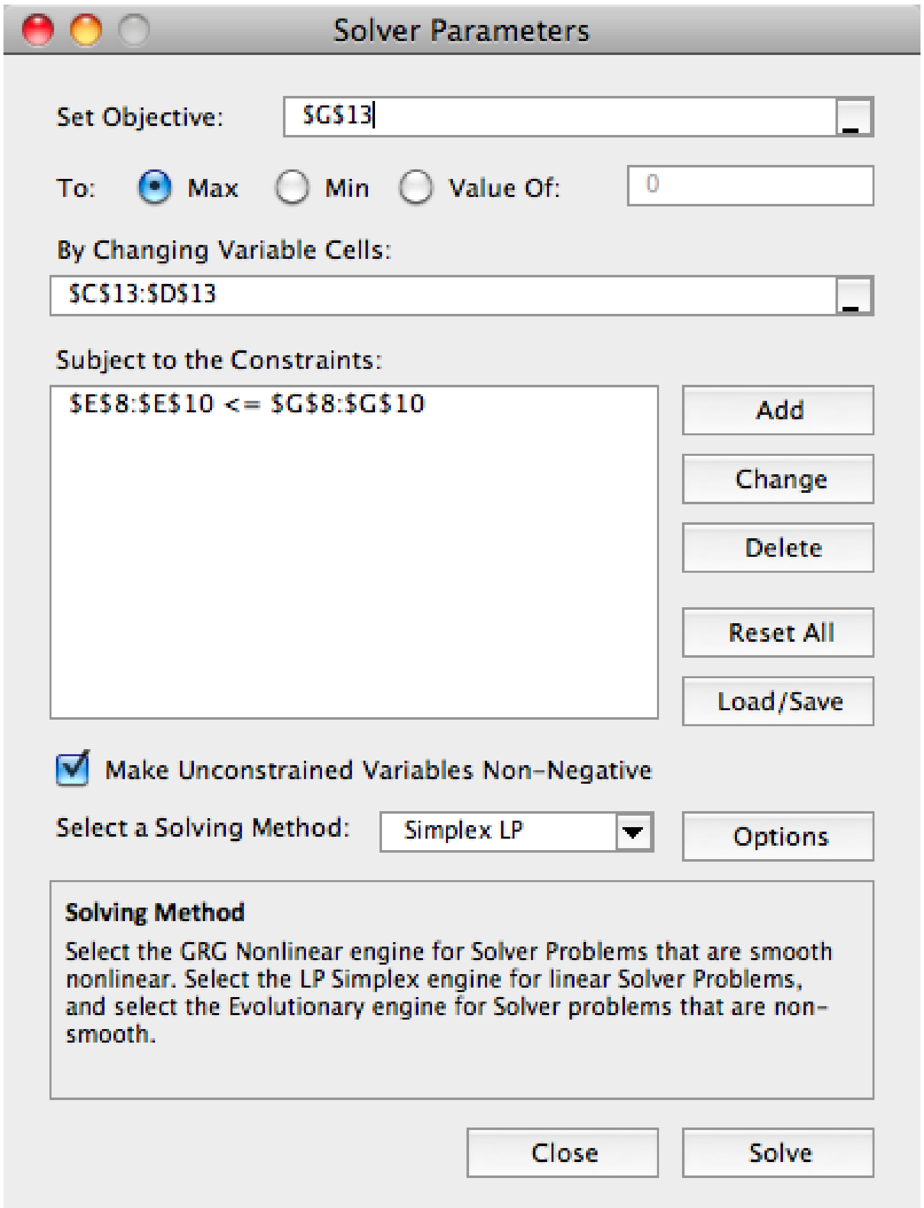


Figure 9: Solver dialogue box filled out

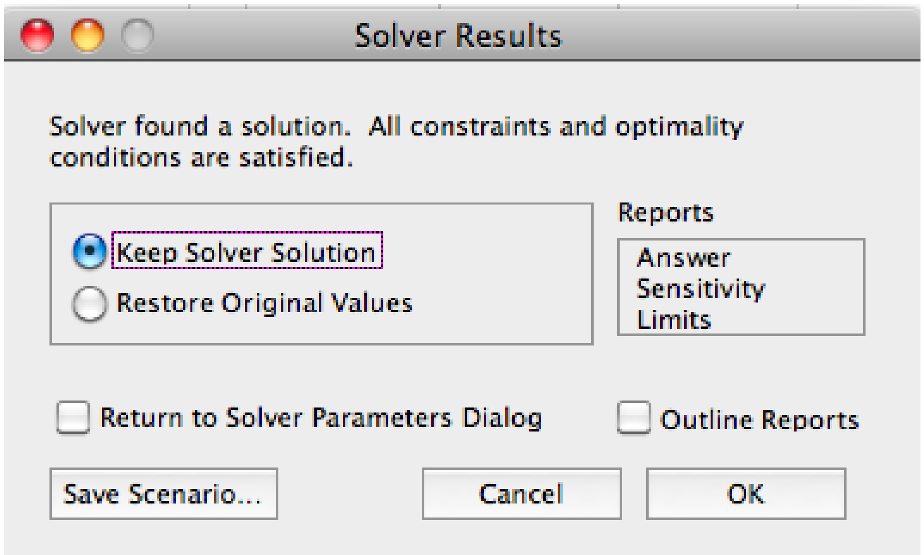
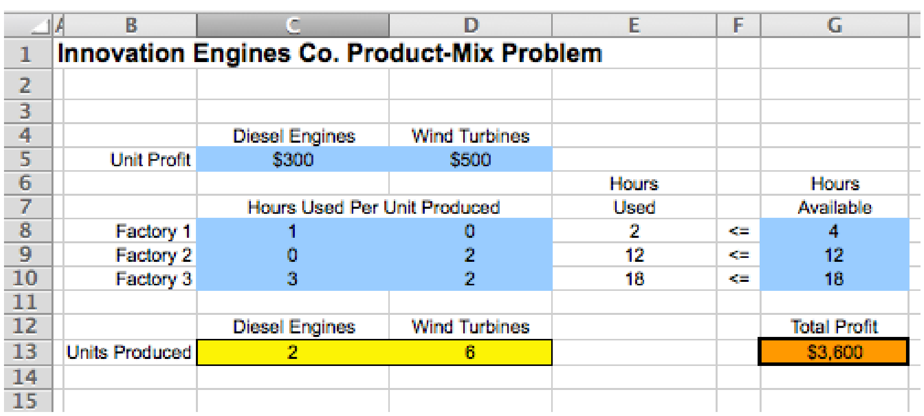


Figure 10: Solver Results window



	A	B	C	D	E	F	G
1		**Innovation Engines Co. Product-Mix Problem**					
2							
3							
4			Diesel Engines	Wind Turbines			
5		Unit Profit	$300	$500			
6					Hours		Hours
7			Hours Used Per Unit Produced		Used		Available
8		Factory 1	1	0	2	<=	4
9		Factory 2	0	2	12	<=	12
10		Factory 3	3	2	18	<=	18
11							
12			Diesel Engines	Wind Turbines			Total Profit
13		Units Produced	2	6			$3,600
14							
15							

Figure 11: Optimal Solution as found via Solver

9 Summary

This chapter introduced the model building process and the use of the DDOC mantra for building an algebraic model and spreadsheet model. We also introduced the main technology that is used to solve LP models: the simplex algorithm. We understood how it intuitively worked by considering the graphical approach to solving LPs with two decision variables. We also saw how to run the simplex algorithm from within Excel for a spreadsheet model.

The process of building a model is like the process for building any product. We do not stop after building it. Rather we question the recommendations of the model and the assumptions behind the model, and iterate in order to create models with more realistic recommendations. However, the key to managing the complexity involved in model building and model iteration is to start simple and keep your spreadsheet model organized. That is one of the reasons we color code our cells and clearly label. We hope you do the same too.