Teaching Note

Introducing the Basics of Spreadsheet Modeling

Thin-Yin Leong, SIM University

tyleong@unisim.edu.sg

Abstract

This teaching note explains an exercise that interactively introduces basic principles and best practices of spreadsheet model construction. This exercise is designed for use in the first lesson of spreadsheets modeling courses. Concepts elaborated include the difference between exploratory and computational modeling; and spreadsheet FILL operation, cell referencing and formatting, and table construction. It also shows ways to audit a spreadsheet and how to avoid making common errors found in spreadsheet models.

Key words: exploratory modeling; FILL; cell formula, referencing, functions, and formatting; spreadsheet errors; hard-coding; data-model-view; Fermi problems; and extrapolation

Introduction

The Spreadsheet Basics workbook is designed to help introduce basic principles and best practices of spreadsheet model construction. This interactive exercise is employed in the first lesson of our spreadsheet modeling courses. It covers ideas and concepts given in Notes 1 to 3, Appendix A; and Tips 1 to 5, Chapter 1 of Leong & Cheong 2015. Typically, this exercise is promptly followed by the Alex Processing exercise (Leong & Cheong 2015), in which ideas and concepts learned in Spreadsheet Basics are reinforced and demonstrated in a real-world application. Students have been primed to explore the bigger world of Fermi Problems and Extrapolation (see Wikipedia notes) as additional pre-class readings.

The context and value of a business modeling with spreadsheet course is explained in greater details in Leong & Cheong 2009 and Leong & Cheong 2008. In them, we tell of the success we have in getting students to learn the art of modeling and how to use spreadsheets as business consulting tools. We encourage you to read at least one of these articles before proceeding on: the first is a short version of the second. If you are already teaching a spreadsheet course, do note that we go beyond spreadsheet technical skills to help students learn how to discover and diagnose business challenges, and design and deliver workable recommendations and solutions. We also ask our students to read these articles to let them know what to expect and encourage them to join the ranks of more than 12,000 students in the past decade who have gone before and are reaping the benefits of the intensive learning effort spent in this course.

Computers and Data

To introduce the instructor and help set up an interesting environment for the class, we make the computer read out the welcome message in the Explain worksheet. For this, select the text and use the Convert Text to Speech (or Speak Text) feature of your personal computer operating system. In MS Excel™, the required feature icon has to be first added to the Quick Access toolbar (the mini-toolbar on the top-left, just above the main menu). Detailed instructions for this can be found in the Internet on the relevant support websites. In Mac OSX, just select the text and key Alt+Esc. The idea here is to help the class understand that computers are technology tools that are supposed to assist us. Their features are increasingly meant to be intuitive, often with
no accompanying user instruction manuals. Our competency comes from repeated use and learning from the Internet and each other.

An issue we often highlight to students at this point is that with technology things are often more than what meets the eye. For example, names of the instructor and teaching assistants, and even time of day reflected in the introductory greeting are not entered directly but generated by computer instructions. Try not to spend too much time here and leave the technical details for interested students to explore after class on their own. We left Prologue and Epilogue notes at the bottom of the Explain worksheet. They summarize our pedagogical philosophy and approach. Use them as you see fit.

From year to year, different teasers have been used to initiate the class on the value of data and problem discovery. They are used to “break the ice” and get the class into an interactive mood. Last year, we posed the question: “how much toothpaste do you put on your toothbrush when you brush your teeth?” A close-up advertisement picture of a thick, luscious strip of toothpaste over the whole toothbrush length is shown as backdrop. A short survey is then taken of the class to get their responses, followed by a brief discussion of why the responded amounts are used and then what would be considered optimal. The class is then queried on how do we know what is optimal. Obviously, someone would suggest Google-searching for the answer. We will let you and the students figure out how to proceed from here. Suffice to say that our prior assumptions may all well be incorrect!

We have also set up a separate exercise for students to collate various ways they and others use spreadsheets, and to discuss which are the more appropriate and effective ways to employ them. This secondary exercise can be done if there is available class time, but we typically leave this as a nice-to-have topic that students can do on their own.

**Modeling vs Model**

The most common use of spreadsheets is to organize data, followed by using them to produce models to compute numerical results. In the first, data depository use of spreadsheets, all tables are “dead” in that their cell values are obtained by external calculations. The second, using spreadsheets as computational model, starts with given (usually mathematical) models that require complex calculations, to do “numbers-in, numbers-out” operations. It treats the spreadsheets as nothing more than an advance calculator with archived results. The problem with such models is that the given model needs to be evaluated and justified first, and textbook models commonly adopted do not really fit the real-world situation it is applied for.

Exploratory modeling (see Figure 1) deals with real-world problems head-on with their ill-defined, unstructured challenges, using the spreadsheet as a workspace to collate data, information and ideas from domain experts and users. It develops the model from scratch, or adapts and combines small portions of different models. The contrasts explained here are between modeling and model, and between exploratory and computational.

As suggested by its name, exploratory modeling utilizes exploration as a process to find the truth. It takes modeling as the approach and the resultant model as a by-product, whereas the other approach starts with a model and try to force-fit it to the real world. Unfortunately, our primary and secondary schools usually teach models and not modeling. Even in high-level courses that are modeling in nature, such as Mathematics, Statistics and Computer programming, instructors tend to teach standardized models, rather than serendipitous modeling. It is not surprising that most business domain courses discuss and apply the classic models of their fields. They therefore do not adequately prepare students to face the real world, where reality does not always quite fit the clinical assumptions on which these models stand.
Figure 1: Exploratory Modeling

Computational Model

As example of computational models, we ask the class to construct multiplication tables. The first and simpler version is a one-dimensional table, and then later as an in-class self-assessment, students construct the two-dimensional table. These are computational models because we already know what the models are; all that is expected of us now is to lay them out and examine the output values for each set of input values entered.

Figure 2: One-dimensional Multiplication Table

In the *Proto* worksheet of *Spreadsheet Basics*, we start by clearing away the no-longer-required instructional texts and put up the table headers: *a*, *b* and *c*. The end result expected is given in the *Table* worksheet (shown here as Figure 2). Since the last header offers no meaning to
method, place. You can discuss with the class why this works and if it can be applied to cells B7:B18 in the first

Better still, double-click on the bottom right corner of the D7:D8 selection to AutoFill the table. You can discuss with the class why this works and if it can be applied to cells B7:B18 in the first place. As a basic principle, we require students to always attempt first before asking if any method, operation or feature works. They are to check the spreadsheet Help feature, and search

For table formatting, we prefer worksheet gridlines to be switched off and border lines added individually to tables to clearly identify each table. This is particularly pertinent when there is more than one table present, closely packed. Other formatting preferences include the use of medium grey lines in cell borders in tables and pastel colors shaded header cells. (See modeling tips scattered in the various chapters of Leong & Cheong 2015 that explain these suggestions and their purpose). Encourage students to read ahead and test out the ideas there. Ask them to bring comments and further suggestions to class for discussion.

We deliberately enter values by hand into cells B7:B18, one number at a time. This is to "test the class' patience", devised to draw students to "ask for or propose better ways" to construct spreadsheet models. Students are told here that the instructor will continue to do this kind of provocation throughout the course, and their class participation contribution is to catch him at it, to keep asking, "Is there a better way?" This is also a question they can ask themselves when attempting the exercises on their own, and whilst doing the required course projects and assignments. At this point, we would liberally try out different things in front of students and repeatedly use the Undo and Redo features. We let students know that the value of using the electronic worksheet is the flexibility to keep trying without fear of messing up or losing valuable effort.

The consequence of this first series of interactions is to introduce and test out the FILL operation. On a separate area further right of the table in the Proto worksheet, we explore how various number series (including "somewhat trending" random numbers) can be extended using Fill. Students would then have to discover what Fill does, and how and when to use Fill. Here, we usually use the scatter chart to visualize the numbers generated by Fill, as a demonstration of chart construction and learning about linear extrapolation. Students are then prodded to try the various charting possibilities and chart formatting features on their own.

The implicit generation of values by Fill presents another problem: how do we know which values in the table are inputs and which are outputs. We will use font colors to help us distinguish between inputs and outputs, and between different kinds of inputs. This is mention here only in passing; the topic will be further examined in future class exercises. As you have noticed by now, we point students to new issues and spreadsheet techniques, and require them to learn most of them on their own. In fact, we adopt the "learn to learn" philosophy in our courses, reserving class time for more conceptual and deeper learning.

Fill can be used for example to enter values into cells C7:C18 and D7:D18. Students may not yet notice that the answers in D7:D18 in the Table worksheet are incorrect. So as you work in the Proto worksheet, just enter 2 and 4 in cells D7 and D8, and apply Fill to complete the column. Better still, double-click on the bottom right corner of the D7:D8 selection to AutoFill the table. You can discuss with the class why this works and if it can be applied to cells B7:B18 in the first place. As a basic principle, we require students to always attempt first before asking if any method, operation or feature works. They are to check the spreadsheet Help feature, and search
the Internet. A particularly useful set of self-learning videos on practically all spreadsheet operations and feature is found in Bill Jelen (Mr Excel)’s YouTube channel.

Best Practices

The multiplication table produced thus far, with computations done externally (using manual value entries) or implicitly (using Fill), is a “dead” table. We all can appear smarter than that, just by letting the spreadsheet do the calculations instead. For this, we enter into cell D7 the formula “=1*2”, D8 with “=2*2” and so on. Get students to work along with you. Again, let students catch you do these rather stupid and tedious steps. A possible alternative approach the class may propose would be to enter “=B7*C7” into cell D7. Now, Copy cell D7 and Paste to cells D8:D18, or equivalently Fill cell 7 to cells D8:D18. Note that Fill can extend values or formulas.

This Copy/Paste or Fill operation does not just rigidly copy the formula over to other cells. It is intelligent enough to adapt the formulas for their new cell locations. So the formula “=B7*C7” in cell D7 becomes “=B8*C8” in cell D8, “=B9*C9” in cell D9, and so on, when they are pasted to cells D8:D18. In general, copying a formula to another cell will change its column and row cell references by the exact number of column and row differences between origin and destination cells. Relative cell referencing is possibly the most important feature of the modern spreadsheets. With it, we can produce massive tables from just one row of correctly formulated cells.

We next discuss the “no hardcoding of formulas” concept. Hardcoding refers to the presence of values in formulas. The key idea is to keep the spreadsheet “alive”, in that formulas should be used to capture relationships among cells, with cells behaving as variables. Therefore, all values only reside in cells and users can explicitly update the variable values in these cells. Cells are referenced by their row and column addresses, thus saving us the need to deliberately name them, though this feature is available in spreadsheet. We do not recommend using the named cell feature, as name cells are not relatively referenced. The other permissible contents of cell formulas are mathematical operators (+, −, /, *, and ^) and functions (e.g., SUM, MAX, MIN, and AVERAGE).

Let’s revisit cell D7 again. Actually, we can do better than entering its formula by hand. In cell D7, key “=”, click cell B7, key “*”, click cell C7 and key Enter. The outcome formula is still the same, but very critically, the approach of selecting cells directly instead of entering the cell references is an important step in avoiding potential problems of incorrect entry. Why are we so obsessed with such trivialities? The answer lies with realizing how often errors are found in spreadsheet in actual business-critical use: 90 over percent. After we discussed various ways to avoid making spreadsheet errors, we next move on to how to spot errors that others made.

After entering the formula, leave your cursor in the formula bar and note that cell references are colored. Also observe that the referenced cells themselves now have graphic boxes of the same colors around them. Move your mouse cursor to one of these boxes and drag it away to another location. You will see the cell reference of the affected cell changed in the formula as well. We alert students that this graphic inspection and alteration of cell formulas is only one of three ways to inspect cells We go on to use Formula Auditing and Ctrl + ` features. Do learn and try these out first and as usual, get the class to practice along with you. Spreadsheets are unique among analytical software application and programming languages in that results of computations are shown and not computer codes. Use Ctrl+` to toggle between the two modes: results only and codes only. Note that the ‘ key is not the apostrophe or single quote key. Remember to use the key just left of the 1 key, in the top-left hand corner of your computer keyboard.

For the spreadsheet to be more idiot-proof, we required all data that appeared multiple times to
be entered only once. The other occurrences have to be referenced to the single entry cell. This way, any update is done at one place and all other places are also updated. The case in violation of this is in cells C7:C18. Hence, we set up cell D4 to hold the variable \( b \) value. To complete the column C7:C18, key “=D4” in cell C7 and fill this down. It will be obvious that cells C8:C18 are in error. We ask students to suggest what next to do and in the process identify the cause of error to be the effect of relative cell referencing. At this juncture, we explain and demonstrate the absolute and mixed referencing concepts, and the use of F4 key (or key \( \# + T \) in Mac OSX) to switch between the four modes of cell referencing.

**System Perspective and Application**

The *Table* worksheet provides the suggested solution for the one-dimensional multiplication table. Notes in cells G7:H17 there summarize our learning thus far. We round up the discussion, urging students not to treat spreadsheets as disposable software. To properly construct and use spreadsheet models as enterprise resources, we need to learn about system perspective from Software Engineering. In this, we introduce the Data-Model-View concept (see Tip 4 of Leong & Cheong 2015). Separating Data from Model (which comprises cells with formulas) was deliberated when we posed the “no hardcoding” rule. View, often ignored in spreadsheet model construction, refers the way the table is laid out, cells formatting and use of charts. It is important that students learn how to present their models simply, and also use them as opportunities to impress clients.

However, the View aspect improperly addressed can lead to errors. This will be illustrated next; we casually ask students if there is anything in *Table* that can be improved. Sharp-eyed students should realize by now of that the answers in cells D7:D18 are incorrect. Learning is more than just definitions recall. The student must know how and when to apply. They must know how to, taking from what they have assimilated and coming up without prompting, provide the steps to diagnose and resolve the identified problem. Hint: Try using formula auditing tools to inspect the multiplication table.

We also provide, as self-assessment to evaluate whether students truly understand cell referencing, the challenge to do a 200 by 200 two-dimensional multiplication table (see Figure 3 below). We ask students to use the *Proto(2)* worksheet to work on this, starting with a small 12 x 12 version and enlarging it when they completed the simpler one. They are advised not look at the cell formulas in the *Table(2)* worksheet yet, but use the challenge to gauge if they can come up with answers from basic concepts, without being dependent on “model answers”. Actually, detailed notes and step by step instructions for both multiplication tables construction are already given in Notes 3 of Leong & Cheong 2015. Students have to be reminded that, like all crafts, spreadsheet skills must be learned by heart, and applied quickly and instinctively, beyond the given narrow settings. Worksheet *Table(3)* is an example of a different application scenario. This table is a preview of the *Time Value* exercise of Chapter 4 of Leong & Cheong 2015. Again, there is usually not enough time to do this in class; this is better left as something the more diligent students can investigate by themselves.
Figure 3: Two-dimensional Multiplication Table

Fermi Problem

As concluding class discussion of this exercise and a bridge to the course proper on exploratory spreadsheet modeling, we show a video (Michell 2012). Alternatively to reduce synchronous class time, students may be asked to view it on their own before coming to class. The quick estimation or “guestimation”, so succinctly demonstrated in the video give students a glimpse into the real world, in which “quick-and-dirty, back-of-envelope” answers are needed. These often provide the basis on which more funds may be requested for further deeper data collection and analytical studies. We do not go much further into this as the rest of the course, in the many exercises to follow, will guide students into this less familiar but most relevant and effective way of using spreadsheets.

We did however set up two optional Explore worksheets in Spreadsheet Basic that instructors can use, should you think doing them would be more concrete in bringing the Fermi problem and exploratory modeling concepts to your class. The first Explore worksheet is on estimating the number of piano tuners. The original city for which this problem was applied is Chicago. You can alter the city involved to your city, or another city of interest to your students. When we use this worksheet, we apply an approach more elaborate than the “orders of magnitude” method Fermi used. The appropriate starting point for any new Fermi problem would be to gather the relevant and available data. Let the class identify them. These data should include the city’s population, average family size, and fraction of families with pianos. Very soon, you and your students will notice that averages and ratios are very important.

Alternatively, you may use the “piano tuner” problem as launch pad to look at a different problem that better arouse the curiosity of your students. We propose in the second Explore worksheet the question, “Are there intelligent life beyond the earth?” The model derived is called Drake’s equation. You can use the exploratory modeling approach to help the class learn the answer to a scientific question. Again, the key issue is not in knowing what is the formula and how to apply it, but rather how to derive those knowledge by first principles, like the first person that discovers this wonderful revelation. Also, do get the class to evaluate the implications of their discovery and ask the “so-what” questions, such as “Can we make contact with other intelligent life?” “Are they expected to be friendly?” and “Will they be less or more advanced than us?”

Discussion

This article discusses the teaching of an exercise used to introduce spreadsheet modeling to a broad student audience. The instructor of such a class has to deal with a range of student
computer and spreadsheets skills competency, from those with computer-phobia to expert power-users. Competent or not, two critical well-documented issues are spreadsheet models in used are by and large, from serious research and auditing results, plagued with errors and spreadsheets are usually not appropriately employed to tap anywhere near their full capabilities.

The first issue has been attributed to the fact that most people are over-confident of their spreadsheet competency and the serious misjudgment of treating spreadsheet models as just disposable software, and the second because the lack of understanding of spreadsheets as a exploratory modeling tool, rather than a computational tool. The purpose of bringing students through such an exercise is to show them that regardless of their competency level, correctly self-assessed or not, they can learn to enjoy and apply this ubiquitous and very powerful tool, in school as well as in the real-world.

The exercise demonstrates how a studio-class room works, one in which the instructor works with student apprentices, to learn the art. Often, we “clown around” by deliberately making mistakes giving opportunities for students to pick them up. To protect the dignity of the instructor, we often explain this before detail work is done in the exercise, so that students do not hold back in “finding fault” with the instructors, knowing that the approach is set up as such. In the process, we as instructors do find ourselves making unplanned mistakes and student spotting them and providing more appropriate or faster ways. We are grateful to the many batches of students we have taken through the exercise and the many tips we learned from them that we now pass on.

We do confess to students, time permitting, of the mistakes we made. It shows that it is perfectly all right to make mistakes and admit them. The caveat and final note to leave to students in this exercise is that we are all on a learning journey. The spreadsheet is just a tool that we are only beginning to learn and will master over time. The truly important lesson is what we with this tool can learn, in exploring the world and discovering issues, understanding them and making progress in the quest for better answers to work and life challenges.

References


YouTube 2015. Bill Jelen ("Mr Excel") Channel.