

Handout for three day Learning Curve Workshop

Unit and Cumulative Average Formulations

DAUMW

(Credits to Professors Steve Malashevitz, Bob Williams, and prior faculty. Blame to Dr. Roland Kankey, roland.kankey@dau.mil)

Introduction to Learning or Cost Improvement Curves

Some of you at one time or another may have purchased unassembled furniture kits. The better kits will often provide you with the average assembly time. This time is based on the number of parts, steps, and/or operations associated with that item, let's say a cabinet. Now what if you had bought four cabinet kits and were planning to assemble the kits one-after-another. The average assembly time is based on building one cabinet, not a series of cabinets. If the time to build the first cabinet was one hour we would not expect to take four hours to build all four, but less than four hours since we would probably get more proficient with each cabinet we built. You might say that the average time does not take into account the effect of quantity.

When we estimate the cost¹ or price of an item, whether it is based on a detailed cost build-up, an analogy, catalog price, or a cost estimating relationship, the cost or price may also not address the effect of quantity. One well-known approach to modeling the quantity effect has been called the learning curve, cost improvement curve, or experience curve. This technique was first discussed in the journals of the 1930's and continues as an industry standard today both in commercial and non-commercial (government) applications.

The general learning curve theory is that people and organizations learn to do things better and more efficiently when performing repetitive tasks, and that under certain conditions there is a usable pattern to the learning. If the conditions of the learning are different, the pattern will likely be different. For that reason a number of different learning curve formulations have been identified and used when appropriate.

In this workshop we will address two of the more widely used formulations, the unit formulation and the cumulative average formulation, and we will discuss some alternative formulations. We will also assess the impact on cost improvement due to changes in the product or the process, and the impact of an interruption or break in the production process by using the Anderlohr and retrograde techniques. This handout deals with the unit and cumulative average formulations.

In what areas of cost should we see "improvement"?

We generally associate cost improvement or learning with repetitive actions or processes such as those directly tied to producing an item over and over. We might categorize these costs as the recurring or variable costs that are a direct result of the quantity of units being produced. Elements of manufacturing (material, tooling, fabrication and assembly) are well suited for

¹ Cost is used generically here to represent the expenditure of resources which may be dollars or hours.

improvement curve consideration in cost analysis. However, it is also useful in price analysis, particularly when quantity buys are under consideration.

Why do we see Cost Improvement?

There are a number of areas in which we gain efficiencies as we produce items in quantity. Probably the most commonly recognized area of improvement is in worker learning. Improvement occurs because as you repeat a process you become more adept at the process both physically and mentally. The supervisors, in addition to realizing these gains, also become more efficient in using their people as they learn their strengths and weaknesses. And while we are talking about people, improvements in the work environment also translate into worker and supervisory improvement. Studies show that changes in climate, lighting, and general working conditions are perceived by employees as management concern. This perception motivates people to improve.

So we can see areas where both worker and supervisory improvement might take place. However, cost improvement is broader than just worker learning, in fact, studies have shown that “learning”, as we tend to think of it, is not the largest driver for improvement. It is the production *process* where we see the greatest gains. Optimization of the production line takes place through: changes in scheduling; placement of tools, bins, stock; simplification of tasks; better tooling; debugged shop instructions; and changes to the end item which make it more efficient to manufacture. Organizational changes can lead to lower recurring costs in areas such as inventory practices (just-in-time), decentralization of tasks (manufacturing cells) or centralization of tasks (heat and chemical treatment processes, tool bin, etc.). One manufacturer identified reductions in scrap rates to be the result of improved vendor relationships. The company gave “preferred vendor” status to suppliers who were able to limit defective parts to some percentage. The reduction in defective parts directly translated into savings in scrap rates, quality control hours, recurring manufacturing labor, etc.

What conditions facilitate Cost Improvement?

Let’s say that you purchased a TV and a VCR. After operating each 20 or 30 times, in which task do you think you will see the most improvement from the first time you operated it, turning on your TV or in programming your VCR? You would probably say the VCR. Why? Well, there is only one step in turning on the TV, but there may be a dozen or more steps in programming your VCR. Each one of those steps represents an opportunity for improvement. The more opportunities - the more potential for improvement. This suggests that we are more likely to benefit from improvement on more complex end-items or tasks than simple ones.

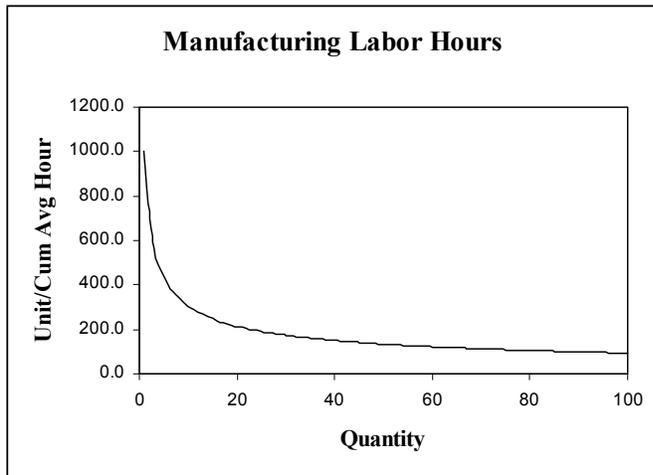
Another aspect of improvement deals with continuity of production. Have you ever felt like no sooner do you learn to use the software on your computer than someone comes along and upgrades it and now you’re back at square one? Then it makes sense that if we want to continue to benefit from improvements we need to limit the changes to the end-items or processes. It also follows that a continuous or uninterrupted process would be helpful. Otherwise we run into the “I remember doing that before, but it’s been awhile” problem.

And last, but certainly not least, there needs to be some pressure or motivation to improve. In competitive business environments, market forces require suppliers to improve

efficiency or they will cease to exist. Suppliers may even competitively price their initial product release at a loss betting on future cost improvements in order to discourage competitors from entering even new markets. These improvements must materialize for the supplier to turn a profit and survive.

What does “cost improvement” look like and how do we model it?

The term “learning curve slope” and equations like “ $y = ax^b$ ” are probably very familiar to you, but where did we get these concepts? Well, as early analysts observed production data (e.g. manufacturing labor hours) they noted that the labor hours per unit or the cumulative average labor hours decreased as illustrated in the following diagram:

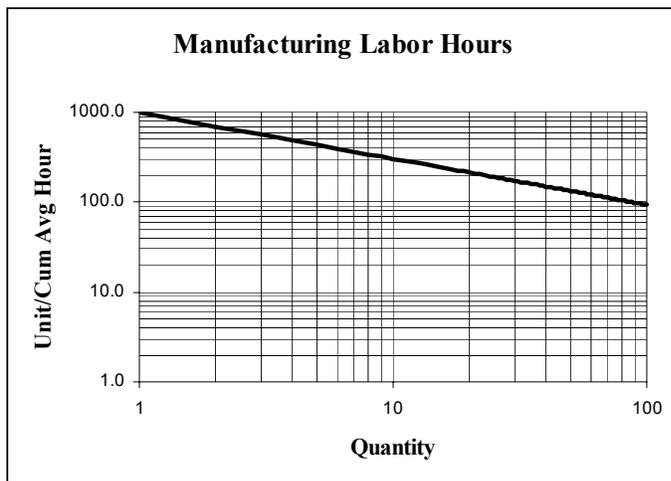


Well, if the trend between hours and quantity is of this form, where does the idea of a constant learning curve “slope” come from and what does the slope represent? And why do people refer to the intercept as the value at unit one when the intercept is usually the value at zero?

To answer these questions we must first concern ourselves with the shape of the curve. One method of dealing with such data is to transform the data so that the transformed values are linear. Another

approach is to use a model form other than the linear model ($y = b_0 + b_1x$) such as the multiplicative model ($y = b_0 x^{b_1}$). We will start off with a transformation that will eventually lead us to the multiplicative model.

Probably the most flexible and convenient transformation is the LOG (base 10) or LN (base e) of hours and quantity. This can be accomplished either by mathematical calculation of the logs or by graphing the values on log-log paper as is seen in this graph.



We can now see a line that could be characterized as having a ‘constant slope’. We usually define the slope as the change in Y given a change in X. However, in log space the slope equates to a constant *percent* change in Y given a

Unit	Hours
1	1000.0
2	700.0
3	568.2
4	490.0
5	436.8
6	397.7
7	367.4
8	343.0
9	322.8
10	305.8

percent”. So we have a constant percent decrease in Y (cost) and a

Learning Curve Workshop

constant percent change in X (quantity, which is doubling or changing by 100%).

Perhaps the best way to illustrate what the slope represents is to look at some of the data that was used to generate the above graph. Since we are talking about the change in hours as quantity doubles let's look at the change in hours from unit 1 to unit 2. The hours decreased from 1000 to 700 or a decrease of 30%. This is called the rate of learning (ROL) or the rate of improvement. Another way of looking at this is to take a ratio of the hours at some quantity (Y_x) versus the hours at twice that quantity (Y_{2x}).

$$\frac{Y_2}{Y_1} = \frac{700}{1000} = .70 \text{ or } 70\% \qquad \frac{Y_4}{Y_2} = \frac{490}{700} = .70 \text{ or } 70\%$$

The 70% is the learning curve slope. Unit 2 is 70% of the value of unit 1, unit 4 is 70% of unit 2, and so on with each doubling of quantity. Keep in mind that the learning curve is a trend in the data and that in this case the 70% slope represents an average improvement (decrease) of 30%. It can also be calculated as $1.00 - \text{ROL} = 1 - .3 = .7$, or $100\% - 30\% = 70\%$.

Just as we've seen that the learning curve slope is not the usual mathematical slope, the same is true about the intercept. The intercept, which is usually associated with the point where $X = 0$ in "unit" space, is the point where $X = 1$ in log space. The value of Y at $X = 1$ is the first unit cost, sometimes referred to as T_1 or as "A" in many learning curve formulas.

Let's take another look at our graph along with its associated trend line. As you can see, the relationship can be portrayed by a multiplicative equation often written as $Y = AX^b$,

where:

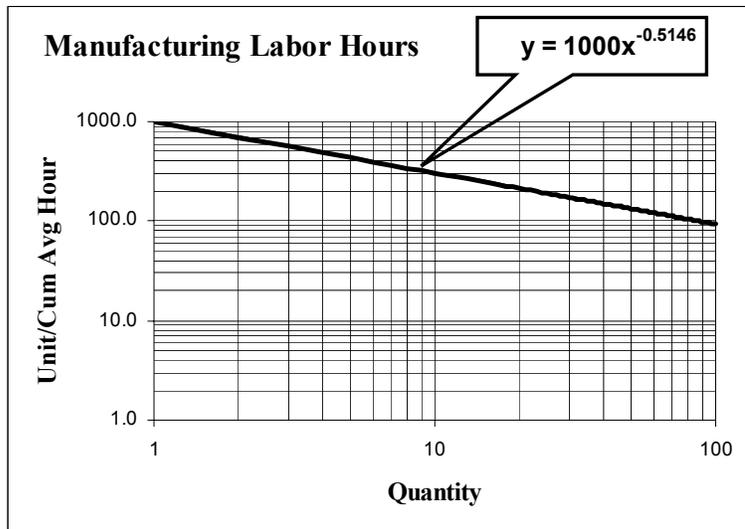
Y: the cost (or avg. cost) at unit x

A: the first unit (T_1) cost (1000)

X: the unit number

b: the slope coefficient which is:

$$\frac{\text{Log(slope)}}{\text{Log}(2)} \text{ or } \frac{\text{Ln(slope)}}{\text{Ln}(2)} = \frac{\text{Ln}(.70)}{\text{Ln}(2)} = -.5146$$



Where do we get the slope and the intercept (first unit or T_1)?

If there have been several production lots produced of an item we could derive the slope from the trend in the data. Another approach would be to look at company history for similar efforts and calculate the slopes from those efforts. Or perhaps we could use the slope from an

analogous program. A similar effort would be to look at slopes for that particular industry (e.g. aircraft, electronics, and shipbuilding). These slopes are sometimes reported in organizational studies, research reports, or estimating handbooks. Slopes can be specific to functional areas such as manufacturing, tooling, and engineering or they may be composite slopes calculated at the system level such as an aircraft, radar, tank, or missile.

The first unit cost might be arrived at through any of a variety of means such as a factor, an analogy, a cost estimating relationship, a grassroots estimate, or by fitting the actual data. In some cases you may not have the first unit cost. The work measurement standards may provide the hours for the 5th unit. A cost estimating relationship may predict the 100th unit cost. This is not a problem as long as you have the value at some unit and have chosen a learning curve slope because you can then solve for the 1st unit cost.

The Unit Formulation

The first technique we will cover is the unit formulation, which is often associated with names such as James Crawford, Stanford Research Institute, and the Boeing Company. The unit formulation states that as the quantity of units doubles, the **unit** cost decreases by a constant percentage. It is represented by the following formula²:

$$Y_x = AX^b \quad \text{where: } Y_x: \text{ the cost of the } X^{\text{th}} \text{ unit}$$
$$A: \text{ the first unit (T}_1\text{) cost}$$
$$X: \text{ the unit number}$$
$$b: \text{ the slope coefficient where } \frac{\text{Log(slope)}}{\text{Log}(2)} \text{ or } \frac{\text{Ln(slope)}}{\text{Ln}(2)} = \frac{\text{Ln}(.70)}{\text{Ln}(2)} = -.5146$$

Example 1: If it took 25,000 manufacturing hours to build the first APG –180 radar, how many hours should it take to build the 50th unit on a 90% unit curve?

$$A = 25,000$$
$$b = \frac{\text{Ln(slope)}}{\text{Ln}(2)} = \frac{\text{Ln}(.90)}{\text{Ln}(2)} = -.152003$$
$$Y_{50} = AX^b$$
$$= (25000)(50)^{-.152003}$$
$$= 13794 \text{ hours}^3$$

Note: We automatically convert a slope in percent to the equivalent decimal value to use in the formulas. (90% is converted to .90)

² There is no distinction or meaning intended by using $Y = AX^b$ versus $y = ax^b$ or any combination therein.

³ Answers will vary slightly due to rounding. When dealing with exponents (e.g. “b”) you should carry the number to at least four decimal places. Even better, just carry the full value in the memory of your calculator.

Example 2: If the 100th Humvee cost \$70,000, what did unit one cost on an 87% unit curve?

$$Y_{100} = 70,000$$

$$b = \frac{\text{Ln}(.87)}{\text{Ln}(2)} = -.200913$$

$$Y_{100} = AX^b$$

$$70000 = (A)(100)^{-.200913}$$

$$A = \$176,573$$

Example 3: What is the slope when $b = -.074$?

$$\text{Slope} = 2^b = 2^{-.074} = .95 \text{ or } 95\%$$

Example 4: If the first Tommachop missile cost \$127,500 and you have a 91% unit curve, which unit would cost \$78,600?

$$A = 127,500$$

$$b = \frac{\text{Ln}(.91)}{\text{Ln}(2)} = -.136062$$

$$Y_X = 78,600$$

$$X = ?$$

$$Y_X = AX^b$$

$$\left(\frac{Y_X}{A}\right)^{\frac{1}{b}} = X$$

$$\left(\frac{78600}{127500}\right)^{\frac{1}{-.136062}} = 35^{\text{th}} \text{ unit}$$

Example 5: If you estimated that the first GPS transceiver would require 1000 hours to manufacture and that you would expect to see a 90% unit slope, how many hours would be required to produce the first five units (CT₅)? With the unit cost formula already presented you could estimate the hours for each unit and then sum the results as follows:

$$CT_5 = A(1)^b + A(2)^b + A(3)^b + A(4)^b + A(5)^b$$

or

$$CT_5 = A[1^b + 2^b + 3^b + 4^b + 5^b]$$

$$CT_5 = 1000[1^{-.152003} + 2^{-.152003} + 3^{-.152003} + 4^{-.152003} + 5^{-.152003}]$$

$$CT_5 = 1000[1.000 + .9000 + .8462 + .8100 + .7830]$$

$$CT_5 = 1000[4.3392] = 4339 \text{ hours for the first 5 units}$$

Now while this provides the desired calculation, it's a somewhat tedious process. What if we

wrote the expression as $CT_X = A \sum_{i=1}^X i^b$ and had a table for $\sum_{i=1}^X i^b$ which represents the

cumulative value of the units 1 through X for a given “b” (slope coefficient)? The following is an example of a 90% cumulative progress curve table (a.k.a. Boeing table) where rows represent units of 10 and columns represent units of 1:

	90% CUMULATIVE PROGRESS CURVE TABLE 90%									
	0	1	2	3	4	5	6	7	8	9
0		1.000000	1.900000	2.746206	3.556206	4.339193	5.100778	5.844726	6.573726	7.289791
1	7.994479	8.689031	9.374458	10.051596	10.721149	11.383717	12.039817	12.689899	13.334357	13.973540
2	14.607759	15.237293	15.862390	16.483278	17.100162	17.713230	18.322654	18.928592	19.531190	20.130582
3	20.726893	21.320240	21.910730	22.498464	23.083538	23.666039	24.246052	24.823653	25.398918	25.971917
4	26.542714	27.111373	27.677953	28.242509	28.805097	29.365766	29.924565	30.481540	31.036736	31.590195
5	32.141956	32.692059	33.240541	33.787437	34.332781	34.876606	35.418944	35.959825	36.499278	37.037331

To use this table to solve the above problem we would select the cumulative value for units 1 - 5 by going to row 0 (0 * 10) and column 5 (1 * 5) to arrive at 4.339193 and apply as follows:

$$CT_X = A \sum_{i=1}^X i^b$$

$$CT_5 = 1000 \sum_{i=1}^5 i^b$$

$$= 1000 (4.339193)$$

$$= 4,339$$

If we wanted the total hours for units 1 through 37, we would go to row 3 and column 7:

$$CT_x = A \sum_{i=1}^x i^b$$

$$CT_{37} = 1000 \sum_{i=1}^{37} i^b$$

$$= 1000(24.823653)$$

$$= 24,824$$

Example 6: How many hours would be required to build units 16 through 50? Solve by taking the total hours for 50 units (units 1 - 50) minus the total hours for 15 units (units 1 - 15).

$$A = 1000$$

$$\text{Slope} = 90\%$$

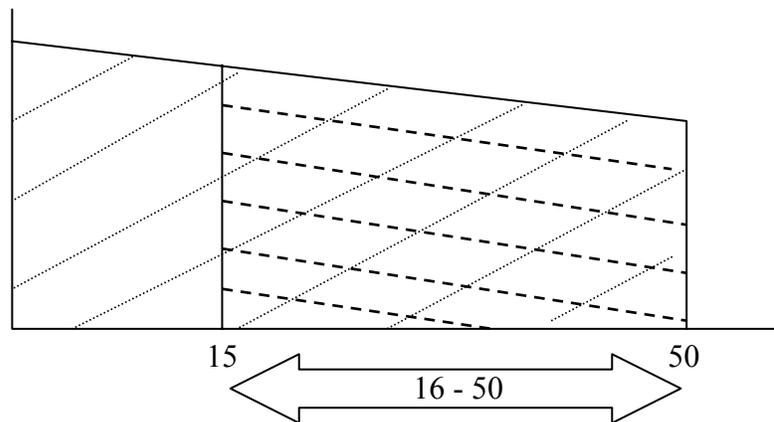
$$TC_{16,50} = [CT_{50} - CT_{15}]$$

$$= \left[A \sum_{i=1}^{50} i^b - A \sum_{i=1}^{15} i^b \right]$$

$$= A \left[\sum_{i=1}^{50} i^b - \sum_{i=1}^{15} i^b \right]$$

$$= 1000 [32.141956 - 11.383717]$$

$$= 20,758$$



Or, as an alternative, an approximation formula that does not require the use of tables:

$$\begin{aligned} TC_{F,L} &= \left[\frac{(L + .5)^{b+1} - (F - .5)^{b+1}}{b + 1} \right] * T_1 \\ &= \left[\frac{(50 + .5)^{.8480} - (16 - .5)^{.8480}}{.8480} \right] * 1000 \\ &= \left[\frac{(27.82) - (10.22)}{.8480} \right] * 1000 \\ &= 20,755 \text{ (or 20,758 if you don't round)} \end{aligned}$$

The LMP then becomes our independent variable (X) which can be transformed logarithmically and used in our simple linear regression equations to find the learning curve for our production situation.

The dependent variable (Y) to be used is the AUC (or LAC) which can be found by:

$$\text{AUC} = \frac{\text{Lot Cost}}{\text{Lot Size}}$$

Again, the dependent variable (Y) must be transformed logarithmically before use in the simple linear regression equations.

Example 7: Given the following historical production data on the turret assembly for a particular tank system, find the Unit learning curve equation which best models this production environment.

<u>Lot</u>	<u>Given</u>			<u>Calculated</u>	
	<u>Size</u>	<u>Units</u>	<u>Cost (man-hrs)</u>	<u>LMP</u>	<u>AUC</u>
1	15	1 - 15	36,750	5	2450
2	10	16 - 25	19,000	20.25	1900
3	60	26 - 85	90,000	51.25	1500
4	30	86 - 115	39,000	99.97	1300
5	50	116 - 165	60,000	139.42	1200

The LMP and AUC now become our independent and dependent variables respectively. These variables can now be transformed logarithmically and used in our simple linear regression equations to solve for A and b, as follows:

<u>LMP</u>	<u>AUC</u>	<u>Log (LMP) = X</u>	<u>Log (AUC) = Y</u>
5	2450	0.6990	3.3892
20.25	1900	1.3064	3.2788
51.25	1500	1.7097	3.1761
99.97	1300	1.9999	3.1139
139.42	1200	2.1443	3.0791

Now using the simple linear regression equations, we obtain:

$$b = -0.216779 \quad \text{slope} = 2^b = 2^{-0.216779} = .8605 \text{ or } 86.05\%$$

and,

$$\text{Log (A)} = 3.5482, \quad \text{and therefore } A = 10^{3.5482} = 3533.46$$

We can now write the learning curve equation that models this production environment as:

$$Y_x = 3533.46 (X)^{-0.216779}$$

Cumulative Average Formulation

The second learning curve formulation that we are going to discuss is the cumulative average formulation commonly associated with T.P. Wright and his 1936 article “Factors Affecting the Cost of Airplanes”. The theory is stated that “as the total quantity of units produced doubles, the **cumulative average** cost decreases by a constant percentage”. This approach uses the same functional form as the unit formulation, but it is interpreted differently.

$\bar{Y}_x = AX^b$ where: \bar{Y}_x : the *average* cost of X units
 A: the first unit (T_1) cost
 X: the cumulative number of units
 b: the slope coefficient where $b = \frac{\text{Log}(\text{slope})}{\text{Log}(2)}$ or $\frac{\text{Ln}(\text{slope})}{\text{Ln}(2)}$

The difference between the unit and cumulative average formulations can be illustrated with this table:

Unit Formulation		Cum Avg Form.
Unit(s)	Unit Cost	Cum Avg Cost
X	Y	\bar{Y}
1	1000.0	1000.0
2	700.0	700.0
3	568.2	568.2
4	490.0	490.0
5	436.8	436.8
6	397.7	397.7
7	367.4	367.4
8	343.0	343.0
9	322.8	322.8
10	305.8	305.8

Both columns were generated with the equation: $\text{Cost} = 1000(\text{Unit})^{-.514573}$ (70% slope)

Unit Formulation

$Y_x = AX^b$

Cost of *Unit X* = $1000(\text{Unit})^{-.514573}$

(∴ the 10th *unit* cost 305.8)

Cum Avg

$\bar{Y}_x = AX^b$

Average Cost of X units = $1000(\text{Unit})^{-.514573}$

(∴ the *average* cost of 10 *units* is 305.8)

So a 70% unit curve ≠ a 70% cum avg curve⁴. This is a good reason to specify both the slope and the formulation when you are documenting the use of a learning curve.

⁴ Additionally, the Cum Avg formulation does not make use of the cumulative progress curve tables as in the Unit formulation.

Example 8: If under the cum avg formulation the 305.8 represents the average cost of 10 units, then what is the cost of the 10th unit given a 70% cum avg curve? To answer that question let's start out with the basic formula.

If $\bar{Y}_x = AX^b$ is the average cost of X units,

then multiplying by X should give us the total cost of X units:

$$CT_X = AX^b * X \quad \text{or} \quad CT_X = AX^{b+1}$$

Given the first unit cost 1000 and we have a 70% curve, the total cost of 10 units is:

$$\begin{aligned} CT_{10} &= (1000)(10)^{-.514573 + 1} \\ &= (1000)(10)^{.485427} \\ &= 3058 \end{aligned}$$

Since we can calculate the total cost of X units, then we can solve for the cost of the 10th unit by taking the total cost of 10 units and subtracting the total cost of 9 units.

$$\begin{aligned} Y_{10} &= CT_{10} - CT_9 \\ &= AX^{b+1} - A(X-1)^{b+1} \\ &= A[X^{b+1} - (X-1)^{b+1}] \\ &= 1000[10^{.485427} - 9^{.485427}] \\ &= 152.5 \end{aligned}$$

Example 9: What is the lot cost for units 16 to 30? We can make use of the unit formula and replace the X's with "L" (last unit in the lot) and "F" (first unit in the lot). Starting with the unit cost formula:

$$Y_X = A[X^{b+1} - (X-1)^{b+1}]$$

we could generalize and say that the total cost (TC) of a lot containing units "F" through "L" can be calculated by:

$$\begin{aligned} TC_{F,L} &= A[L^{b+1} - (F-1)^{b+1}] \\ &= 1000[30^{.485427} - (16-1)^{.485427}] \\ &= 1489 \end{aligned}$$

Using Historical Data to find the Slope and the First Unit Cost

As we discussed earlier, under the unit formulation, when your data is in the form of the cost of production lots, you need some means of representing each lot by a specific point. In the unit form we did this by using the AUC and the LMP for each lot. Under the cum avg form this is accomplished by using the cumulative average cost and the cumulative units⁵ as follows:

<u>Units</u>	<u>Lot Cost</u>	<u>Cum Cost</u>	<u>Cum Units</u>	<u>Cum Avg Cost</u>
1 – 10	5,960	5,960	10	596.00
11 – 25	5,130	11,090	25	443.60
26 – 40	4,160	15,250	40	381.25
41 - 60	4,825	20,075	60	334.58

Example 10: Given the following historical production data on the turret assembly for a particular tank system, find the Cum Avg learning curve equation which best models this production environment.

<u>Lot</u>	<u>Size</u>	<u>Cost (man-hrs)</u>	<u>Cum Units</u>	<u>Cum Avg Cost</u>
1	15	36,750	15	2450.00
2	10	19,000	25	2230.00
3	60	90,000	85	1714.71
4	30	39,000	115	1606.52
5	50	60,000	165	1483.33

The Cum Units and Cum Avg Cost now become our independent and dependent variables respectively. These variables can now be transformed logarithmically and used in our simple linear regression equations to solve for A and b, as follows:

<u>LMP</u>	<u>AUC</u>	<u>Log (LPP) = X</u>	<u>Log (AUC)= Y</u>
15	2450.00	1.1761	3.3892
25	2230.00	1.3979	3.3483
85	1714.71	1.9294	3.2342
115	1606.52	2.0607	3.2059
165	1483.33	2.2175	3.1712

Now using the simple linear regression equations, we obtain:

$$b = -0.210514 \qquad \text{slope} = 2^b = 2^{-0.210514} = .8642 = 86.42\%$$

$$\text{Log (A)} = 3.6395 \qquad A = 10^{3.6395} = 4360.14$$

The learning curve equation that models this data is: $\bar{Y}_x = 4360.14(X)^{-0.210514}$

Choosing Between Unit and Cum Avg

⁵ The cumulative units are sometimes referred to as the lot plot point (LPP).

Which formulation should you use? Choosing between the Unit and the Cum Average Formulations is not so much a science as it is an art. There are no firm rules that would cause you to select one formulation over the other, but there are some factors that we can look at to decide which might best model the actual production environment. Some of the factors to be considered when determining the formulation to use are:

- Analogous systems
- Industry standards
- Historical experience
- Expected production environment

Analogous Systems - Systems that are similar in form, function, development, or production process may provide justification for choosing one formulation over another. For example, if a service is looking to buy a modified version of a commercial jet and on previous purchases of modified commercial jets the Unit curve was proven to best model the production environment, the estimator may choose the Unit formulation based upon this analogy.

Industry Standards - Certain industries sometimes gravitate toward one formulation versus another. For example, certain types of space systems have been shown to best fit the Cum Average formulation. If an estimator were estimating the same type of space system, the Cum Average formulation might be selected based upon the industry standard.

Historical Experience - Some defense contractors have a history of using one formulation versus another. This formulation may be used because it has been shown to best model the production environment for that particular contractor. An estimator may choose to also use that formulation based on that historical experience.

Expected Production Environment - Certain production environments tend to favor one formulation versus another. For example, the Cum Average formulation has been found to best model production environments where the contractor is starting production with the use of “soft” or prototype tooling, has an inadequate supplier base established, expects early design changes, or is subject to short lead-times. This might be the result of a high degree of “concurrency” or overlap between the development and production phases. This formulation is sometimes preferred in these situations because the effect of averaging the production costs tends to “smooth out” the initial cost variations and provides a better fit to the data. Conversely, the Unit formulation tends to fit best in production environments where the contractor is well prepared to begin production in terms of tooling, suppliers, lead-times, etc. Less “smoothing” means changes are easier to see.

There are no firm rules as to choosing one formulation versus another. However, an estimator should **not** choose a formulation solely because it generates the lowest cost estimate. A formulation should be chosen based on the estimator’s ability to determine which would best model the costs for the system being produced. The estimator should also be aware that the smoothed (averaged) data used in the cumulative average formulation will most likely have a better fit statistic (example, R^2) than the unit formulation. Beware of using a fit statistic as your basis for formulation selection.

Conclusion

The general learning curve theory is that people and organizations learn (become more efficient) when performing repetitive tasks. Under certain conditions there is a usable pattern to this learning. This pattern can be very useful to the analyst performing cost or price analysis. Learning or Improvement Curves have been used for most of a century, and remain an important estimating tool. As Jack Hale indicated, we need to recognize that different situations and different environments will drive different yet usable learning curve patterns. The Unit and Cumulative Average formulations are the most common, but other forms clearly exist. The Learning Curve analyst should become familiar with these differing formulations, when they most likely would apply, their limitations and advantages, and what to expect when the situation, product, or process changes.