Chapters to Go

The New Science of Retailing: How Analytics Are Transforming the Supply Chain and Improving Performance

by Marshall Fisher and Ananth Raman
Harvard Business Press. (c) 2010. Copying Prohibited.

Appendix: Section A-1—Using the Gamma Function in Excel to Create Table 3-6 and Using This Table to Calculate an Optimal Buy Quantity

Overview

Table 3-6 and table A-1 below can be created using Excel. We used the GAMMADIST(d,a,b,true) function in Excel, where *d* is a demand value, *a* and *b* are two parameters that define the Gamma, and *true* indicates that we want the probability that demand will be less than or equal to *d* . The parameters *a* and *b* can be computed from the average and standard deviation using the formulas a = (average / standard deviation) ^{2} and b = (standard deviation) 2 / average.

For the navy turtleneck with an average of 94.75 and a standard deviation of 7.3272, we get a = (94.75 / 7.3272) 2 = 167.219, and b = (7.3272) ^{2} / 94.75 = 0.567.

The Gamma distribution is defined for all values, including fractional values, but since demand needs to be an integer, we use the probability for values plus or minus 0.5 of a demand value as the probability of that demand. For example, the probability of .0054 for a demand of 95 is computed as GAMMADIST(95.5,167.218,0.567,true) – GAMMADIST (94.5,167.218,0.567,true).

The cumulative probability in table 3-6 is simply the sum of all probabilities up to and including that demand value.

This shows how to create table 3-6 using Excel. Next we'll explain how to create table A-1 using the table 3-6 Excel model as a basis.

First we show how to find the probability-weighted cost for a particular buy quantity—for example, 95. Enter this in a cell somewhere in the table 3-6 model; create a column next to the "Demand," "Probability," and "Cumulative probability" columns, and enter in this the cost for the trial buy quantity and the demand value, using the fact that the cost of overbuying is $22 and of underbuying $20, for the navy turtleneck. For example, for the trial buy of 95 and a demand of 75, this cost would be (95 – 75) units × $22 = $440; for a demand of 105, the cost is (105 – 95) units – $20 = $200.

Table A-1: Finding an optimal buy for the navy turtleneck using a Gamma

Buy quantity	Probability-weighted cost	Buy quantity	Probability-weighted cost
70	$495.01	96	126.18
71	475.01	97	131.54
72	455.02	98	139.03
73	435.05	99	148.52
74	415.08	100	159.87
75	395.15	101	172.92
76	375.25	102	187.47
77	355.42	103	203.36
78	335.69	104	220.40
79	316.11	105	238.43
80	296.72	106	257.29
81	277.62	107	276.83
82	258.90	108	296.93
83	240.68	109	317.48
84	223.10	110	338.39
85	206.33	111	359.57
86	190.54	112	380.97
87	175.93	113	402.53
88	162.70	114	424.22
89	151.04	115	445.99

90	141.14	116	467.84
91	133.19	117	489.73
92	127.31	118	511.65
93	123.61	119	533.60
94	122.18	120	555.57
95	123.04	121	577.54

Now in a column next to this, record the product of the cost for each demand value times the probability for that demand value, and sum these products over all rows to get the probability-weighted cost for that trial buy quantity.

This shows how to compute one entry in table A-1—that is, the probability-weighted cost for a particular buy quantity. We could laboriously create table A-1 by changing the trial buy quantities one by one, observing the probability-weighted cost, and recording this value in table A-1. Fortunately, the Data Table function in Excel will do this automatically for us.

To use the Data Table function to create table A-1, first enter the possible buy quantities 70 through 121 in an unused portion of the table 3-6 Excel model.

Suppose you entered 70 through 121 in cells F2–F53. Then enter in cell G1 the label of the cell containing the probability-weighted cost for a trial buy quantity, highlight cells F1–G53, go to the Data tab, click the What If icon, then Data Table and when you see the command box, enter the cell containing the trial buy quantity in the 'Column input cell:' box.[1] Then click OK, and Excel will create table A-1 for you, by substituting one by one the values 70 through 121 as trial buy quantities and recording the associated probability-weighted costs in column G next to the buy quantity. Table A-1 shows the expected cost of all buys from 70 to 121. You can see from the table that a buy of 94 minimizes expected cost.[2]

[1]This instruction is for Microsoft Excel 2007. If you have a different version, you may need to consult the help command for how to construct a data table.

[2]We created tables 3-6 and A-1 using demand values ranging from 0 to 222, but only show demand from 70 to 121 in table 3-6, since other demand values have negligible probability.

Section A-2: Creating Store Clusters Based on Sales Mix

Table A-2 shows a simplified hypothetical example designed to illustrate how we clustered stores based on sales. The top section in this table shows unit sales of four products in three stores. In the middle, we have computed the percentage that each product's sales constituted of total store sales. And on the bottom, we have computed a *mix difference score* for each pair of stores by summing the absolute differences in product sale percentages for the pair of stores. For example, the sales mix difference between stores 1 and 2 is computed as $|18 − 2| + |71 − 6| + |4 − 31| + |7 − 61| = 162$. Note that the difference score of 35 between stores 1 and 3 is small relative to the differences between 2 and 1 or 2 and 3. Thus we concluded that stores 1 and 3 sold a similar mix of products, while store 2 was an outlier. If we were forming two store clusters, stores 1 and 3 would be in one cluster, and store 2 in the other.

Table A-2: Measuring the difference in sales mix between stores

	UNIT SALES OF FOURSKUS AT THREE STORES			
Store	A	B	C	D
1	100	400	20	40
2	8	32	160	320
3	200	300	10	30
	SALES CONVERTED TO PERCENTAGES			
Store	A	B	C	D
1	18%	71%	4%	7%
2	2%	6%	31%	61%
3	37%	55%	2%	6%
	STORE MIX DIFFERENCE SCORES			

Store	1	2	3	
1		162	38	
2	162		168	
3	38	168		

To choose ten test stores from the stores of a chain, you'd use this approach to compute a difference score between every pair of stores, and then employ standard clustering procedures to form ten clusters chosen to minimize the average difference score between stores within a cluster. With more than one thousand stores, a difference score needs to be computed for more than one thousand times one thousand store pairs, so obviously a computer is needed.

You'd then use your sales history database to create a formula that best predicts chain season sales from test sales. You can use a statistical technique called *linear regression*, which is available in Excel and many other statistical analysis packages, to do this. Recall how the retailer we worked with would have predicted its total sales by multiplying each store's test sales by 80. The approach here is similar, in that you multiply each test store's sales by a factor (the coefficient in the regression equation), except now the factor for each store can differ. To illustrate how useful this can be, imagine that you have a store with relatively low sales that is highly predictive of the sales for the chain. It would be an ideal test store, but because its test sales will be relatively low, it needs a bigger factor in a prediction formula.

Section A-3: A Forecast Model Incorporating Price and Inventory

We used the formula shown in figure A-1 and explained here to forecast sales, for a shoe retailer, of a particular style/color sandal (call it sandal j) in a particular week in the season in all stores and sizes. The formula consists of a base forecast of what you expect to sell over the entire season with an average price and enough inventory to avoid stockouts, multiplied by factors for seasonality and price and inventory adjustment. The base forecast, K_j, was simply what we had used before— that is, sales in the first eight weeks of the spring season divided by 0.107. The seasonality factor, $s(t)$, is the percentage of total-season sales we expect to occur in week t. These factors can be computed using sales from the prior year's spring sandal sales and computing the fraction that each week's sales constitute of total season sales. The price factor raises the forecast if sandal j is priced relatively lower than other sandals, and lowers the forecast if it is priced higher. The average ratio of price to cost in week t for all of the retailer's sandals is $p(t)$, and $p_j(t)$ is the ratio of price to cost for sandal j. We compare price–cost ratios rather than raw prices because the various sandals differed in quality, which caused differences in their base price. We reasoned that quality was correlated with cost, and the price–cost ratio would reflect how a sandal was priced relative to its intrinsic quality.
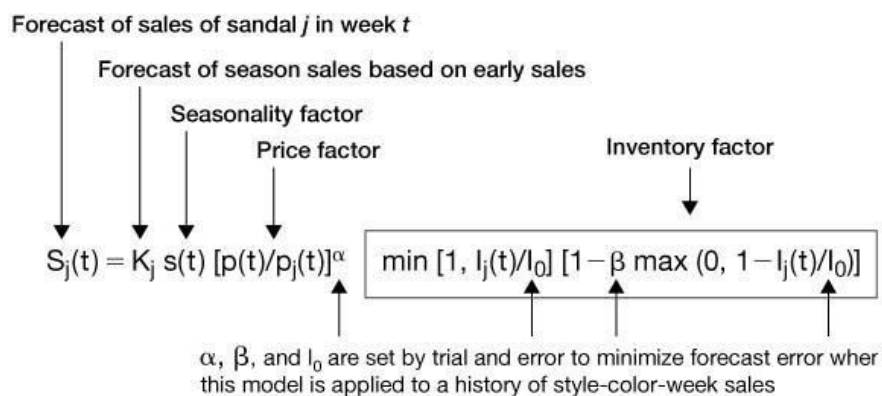

Forecast of sales of sandal j in week t

Forecast of season sales based on early sales

Seasonality factor

Price factor

Inventory factor

$$S_j(t) = K_j\, s(t)\, [p(t)/p_j(t)]^{\alpha}\ \min[1,\, I_j(t)/I_0]\, [1 - \beta \max(0,\, 1 - I_j(t)/I_0)]$$

α, β, and I_0 are set by trial and error to minimize forecast error when this model is applied to a history of style-color-week sales

FIGURE A-1: Forecast model incorporating the impact of price and inventory

If $p(t)/p_j(t)$ is less than 1 (that is, sandal j's price–cost ratio is high relative to an average sandal in week t) then the price factor $[p(t)/p_j(t)]$ is less than 1, and it lowers the forecast. Conversely, if $p(t)/p_j(t)$ is greater than 1 (that is, sandal j's price–cost ratio is low relative to an average sandal in week t) then the price factor raises the forecast.

The inventory factor is 1 if the retailer has enough inventory to avoid stockouts, and less than 1 if its inventory falls below a critical threshold, I_0. $I_j(t)$ is the inventory of sandal j in all stores in week t. We chose the parameters , , and I_0 to minimize forecast error when we applied this model to the 1997 history. We found the best value for I_0 was ten pairs per store, which makes sense because when inventory falls below this level, popular sizes are stocked out in many stores, causing lost sales.