

Creating a Depth-of-Repeat Sales Summary Using Excel

Peter S. Fader
www.petefader.com

Bruce G. S. Hardie[†]
www.brucehardie.com

May 2004

1. Introduction

Central to diagnosing the performance of a new product is the decomposition of its total sales into trial, first repeat, second repeat, and so on, components. More formally, we are interested in creating a summary of purchasing that tells us for each unit of time (e.g., week), the cumulative number of people who have made a trial (i.e., first-ever) purchase, a first repeat (i.e., second-ever) purchase, a second repeat purchase, and so on. We let $T(t)$ denote the cumulative number of people who have made a trial purchase by time t , and $R_j(t)$ denote the number of people who have made at least j repeat purchases of the new product by time t ($j = 1, 2, 3, \dots$).

With such a data summary in place, standard new product performance metrics such as “percent triers repeating” and “repeats per repeater” (Clarke 1984; Rangan and Bell 1994) as easily computed from these data. At any point in time t , percent triers repeating is computed as $R_1(t)/T(t)$, while repeats per repeater is computed as $R(t)/R_1(t)$, where $R(t)$ is the total number of repeat purchases up to time t :

$$R(t) = \sum_{j=1}^{\infty} R_j(t).$$

Furthermore, a simple new product sales forecasting model can easily be built around such a data summary (Fader and Hardie 2004).

In this note, we describe how to create such a sales summary from raw customer-level transaction data (typically collected via a consumer panel) using standard spreadsheet software.

[†]© 2004 Peter S. Fader and Bruce G. S. Hardie. This document and the associated spreadsheet can be found at <<http://brucehardie.com/notes/006/>>.

2. Preliminaries

A consumer panel is formed by selecting a representative sample of individuals or households (from the population of interest) and recording their complete behaviour (e.g., purchasing of FMCG products) over a fixed period of time.¹

For a given product category, we can construct a dataset that reports the timing of each purchase, along with details of the product purchased. A stylized representation of this is given in Figure 1 for a total of n households, in which we consider an observation period starting with the launch of a new product and ending at time t_{end} . We let \diamond denote a purchase of the new product and \times denote the purchase of any other product in the category.

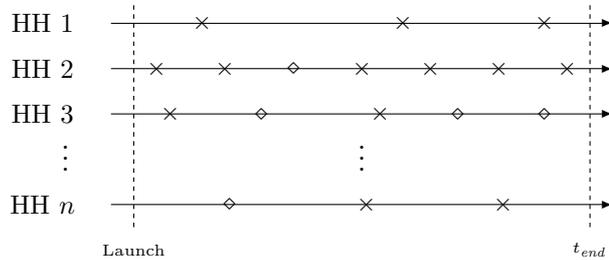


Figure 1: Purchase Histories: Total Category

We see that HH 1 made three category purchases over the observation period but never purchased the new product. HH 2 made seven category purchases; the third category purchase represents a trial purchase of the new product, and no repeat purchasing activity was observed over the remainder of the observation period. HH 3 made a trial purchase of the new product and two repeat purchases. And so on.

In many analysis situations, where we are focusing on a particular product, purchase records not associated with the focal product are removed to yield a simpler (and smaller) dataset. A stylized representation of this is given in Figure 2. As HH 1 never bought the new product, there is no explicit record of this household in the resulting dataset.

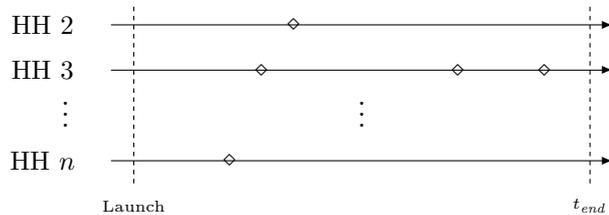


Figure 2: Purchase Histories: New Product Only

¹See Parfitt (1986) and Sudman and Wansink (2002) for background information on consumer panels.

“Kiwi Bubbles” is a masked name for a shelf-stable juice drink, aimed primarily at children, which is sold as a multipack with several single-serve containers bundled together. Prior to national launch, it underwent a year-long test conducted in two of IRI’s BehaviorScan markets. The file `kiwibubbles_tran.txt`² contains purchasing data for the new product, drawn from 1300 panelists in Market 1 and 1499 panelists in Market 2.

Each record in this file comprises five fields: Panelist ID, Market, Week, Day, and Units. The value of the *Market* field is either 1 or 2. The *Week* field gives us the week number in which the purchase occurred (the product was launched at the beginning of week 1), the *Day* field tells us the day of the week (1–7) in which the purchase occurred, and the *Units* field tells us how many units of the new product were purchased on that particular purchase occasion.

We load this dataset into Excel and add a header row — see Figure 3. (The associated Excel spreadsheet is called `creating_dor_summary.xls`.) We see that there are a total of 857 transactions across the two markets during the year-long test.

	A	B	C	D	E
1	ID	Market	Week	Day	Units
2	10001	1	19	3	1
3	10002	1	12	5	1
4	10003	1	37	7	1
5	10004	1	30	6	1
6	10004	1	47	3	1
856	20137	2	13	6	1
857	20138	2	18	7	1
858	20139	2	49	3	1

Figure 3: Raw Transaction Data

To illustrate the process of creating a sales by depth-of-repeat summary from this raw transaction data using Excel, we will focus just on Market 2. We create a copy of the `kiwibubbles_tran` worksheet (let’s call it **Market 2 Master**), delete the Market 1 records (rows 2–552) and the column corresponding to the *Market* field — see Figure 4.

Let us consider the transaction history of panelist 20014 (Figure 5). We note that this panelist made his/her trial purchase *and* first repeat purchase in week 4. Similarly, this panelist’s third and fourth repeat purchases occurred in week 7. We also note that this panelist typically purchased several units of the product on any given purchase occasion.

This suggests several possible versions of the desired sales by depth-of-repeat summary:

- Our summary counts the number of trial, first repeat, second repeat, etc. *transactions* that occurred each week. The process of creating

²See <http://brucehardie.com/datasets/kiwibubbles.zip>

	A	B	C	D
1	ID	Week	Day	Units
2	20001	49	1	1
3	20002	14	7	1
4	20003	24	6	1
5	20004	49	2	1
6	20005	6	7	1
305	20137	13	6	1
306	20138	18	7	1
307	20139	49	3	1

Figure 4: Market 2 Transaction Data

	A	B	C	D
1	ID	Week	Day	Units
16	20014	4	2	1
17	20014	4	4	1
18	20014	6	6	2
19	20014	7	2	3
20	20014	7	6	3
21	20014	12	5	2
22	20014	17	6	1
23	20014	23	4	2
24	20014	47	6	2

Figure 5: Transaction History for Panelist 20014

such a summary is described in §3 below.

- Our summary reports the *sales volume* (e.g., units) associated with trial, first repeat, second repeat, etc. transactions that occurred each week. The process of creating such a summary is described in §4 below.
- We have noted that this panelist’s trial and first repeat purchases occurred in the same week, albeit on different days. Similarly, his/her fourth and fifth repeat purchases occurred in the same week. The structure of many simple models of new product sales forecasting is such that a customer can have only one transaction per unit of time. If the unit of time is one week (as it typically the case), we clearly have a problem. One solution would be change the unit of time from week to day.³ However, as such purchasing behaviour tends to be rare, Eskin

³But what happens if we observe multiple transactions on the same day? This is very rare and typically reflects bad pre-processing of the panel data. For example, as an individual’s purchases are scanned at the supermarket checkout, one six-pack of Coke could be the first item scanned with another six-pack of Coke being the last item scanned. As the raw data are “cleaned-up” these two purchases should be combined into one transaction with a quantity of two. But this doesn’t always happen. If the (very) raw panel data file contains a transaction time field, we easily determine whether the two records come from the same or different shopping trips. Even if they did come from separate shopping trips on the same day, our natural reaction would be to combine them into a single transaction with multiple units, rather than shift to an even smaller time unit (e.g., hour). The interested

(1973, p. 118, footnote 2) suggests that, “[f]or estimation purposes, second purchases within a single week are coded in the following week.” The process of creating such a “shifted” summary is described in §5 below.

3. Creating a “Raw” Transactions Summary

We start by creating a copy of the `Market 2 Master` worksheet, calling it `Market 2 ‘raw’`. The first thing we need to do is add a field that indicates the depth-of-repeat level associated with each record; i.e., is this a trial purchase ($\text{DoR} = 0$), a first repeat purchase ($\text{DoR} = 1$), a second repeat purchase ($\text{DoR} = 2$), etc.?

This is a straightforward exercise. If the panelist ID associated with this record does not equal that of the previous record, we are dealing with a new panelist and we set the depth-of-repeat indicator to 0. If the panelist ID associated with this record does equal that of the previous record, we are dealing with a repeat purchase by that panelist and we increment the depth-of-repeat indicator by 1. We implement this in our Excel worksheet in the following manner: in cell E2 we type the formula `=IF(A2<>A1,0,E1+1)` and copy it down to cell E307. The corresponding records for panelist 20014 are given in Figure 6

	A	B	C	D	E
1	ID	Week	Day	Units	DoR
16	20014	4	2	1	0
17	20014	4	4	1	1
18	20014	6	6	2	2
19	20014	7	2	3	3
20	20014	7	6	3	4
21	20014	12	5	2	5
22	20014	17	6	1	6
23	20014	23	4	2	7
24	20014	47	6	2	8

Figure 6: Transaction History for Panelist 20014 with DoR Indicator

The next step is to create a table that tells us how many trial, first repeat, etc. purchases (columns) occurred in each week (rows). We want there to be 52 rows, one for each week of the test. It turns out, however, that this panel of 1499 households only purchased the test product in 49 weeks; no purchases occurred in weeks 25, 39, and 41. How can we create a table that will contain zeros in the rows corresponding to these three weeks?

At the bottom of column B, we add the numbers 1, 2, . . . , 52. For these new records, we assign a depth-of-repeat level of -1 (cells E308:E359). We

reader should reflect on how to determine whether we observe multiple transactions for an individual panelist on the same day once the raw panel data has been loaded into a spreadsheet. (Note that there are no such occurrences in the Kiwi Bubbles dataset.)

can now create the desired summary table using the “pivot tables” feature in Excel. Highlighting the cell range A1:E359, we select the *PivotTable and PivotChart ...* option under the *Data* menu. We use *Week* as the *row field*, *DoR* as the *column field*, and *ID* as the *data item*. (Make sure the “Pivot Table Field” is “Count of ID”, not sum or another summary of the ID field.) The resulting table is reported in the `Pivot Table 1` worksheet. We note that there are no entries the rows corresponding to weeks 25, 39, and 41. Over the year-long test, 139 of the 1499 panelists made at least one purchase of the new product, with a total of 306 purchase occasions. We also note that by the end of the year, one person had made eleven repeat purchases of the new product.

A cleaned-up summary that reports these weekly transactions in cumulative form (i.e., $T(t)$, $R_1(t)$, $R_2(t)$, etc.) is created in the `DoR -- ‘raw’ transactions` worksheet. (A quick examination of the formulas will reveal how this is done.)

4. Creating a “Raw” Sales Volume Summary

Having created a weekly transaction by depth-of-repeat level summary, it is extremely easy to create an equivalent sales volume (e.g., units) summary. Going back to the `Market 2 ‘raw’` worksheet, we once again highlight the cell range A1:E359 and select the *PivotTable and PivotChart ...* option under the *Data* menu. We use *Week* as the *row field* and *DoR* as the *column field*. Instead of using *ID* as the *data item*, we select *Units*. If we don’t see “Sum of Units” at the top left corner of the table, double-click on this cell and change the “summarize by” option to *Sum*. The resulting table is reported in the `Pivot Table 2` worksheet. We note that a total of 396 units of the product were purchased (across the 306 purchase occasions).

A cleaned-up summary that reports these weekly transactions in cumulative form is given in the `DoR -- ‘raw’ sales volume` worksheet. We note that a total of 161 units were purchased on the 139 trial purchase occasions, an average 1.16 units per trial purchase.

5. Creating a “Shifted” Transactions Summary

We now turn our attention to the task of creating a weekly transaction by depth-of-repeat level summary under the assumption that a customer can have only one transaction per week. In other words, a second purchase within a single week is “shifted” to the next week (i.e., coded as occurring in the following week).

Referring back to Figure 6, the field that indicates the depth-of-repeat level associated with each record (*DoR*) is correct. What we need to do is makes some changes to the week field: we want the week associated with the first repeat purchase to be 5, and the week associated with the fourth repeat

purchase to be 8. One solution would be to create a new week variable that equals the original week variable +1 if the week associated with the current record is the same as that of the previous record. But what if we have three purchases occurring in the same week?

To complicate matters, consider the transaction history of panelist 20069 (Figure 7). This person’s trial and first repeat purchases occurred in the same week. We therefore change the week associated with the first repeat purchase from 18 to 19. But this creates another problem as this person’s second repeat purchase occurred in week 19. Having shifted the first repeat purchase to week 19, we have to shift the second repeat purchase to week 20.

	A	B	C	D
1	ID	Week	Day	Units
148	20069	18	1	1
149	20069	18	5	1
150	20069	19	4	2

Figure 7: Transaction History for Panelist 20069

Our solution is to create an offset variable that can be added to the value of the week field, giving us a “shifted” week variable. Clearly the value of this offset will be zero for the trial purchase. For any repeat purchase record, if the week associated with the current record is the same as that of the previous record, we increment the offset variable by 1. (This ensures that the third purchase in a given week is shifted two weeks.) We also need to shift any purchases encroached on by the shifting of previous purchases (such as the second repeat purchase for panelist 20069).

We create this offset variable in the following manner. Making a copy of the **Market 2 ‘raw’** worksheet, and calling it **Market 2 ‘shifted’**, we type the following formula in cell F2

`=IF(A2=A1,IF(B2=B1,F1+1,MAX(0,B1+F1-B2+1)),0)`

and copy it down to cell F359 (NB: not F307). The shifted week variable is created by adding this offset to the value of the Week field. In cell G2 we type the formula `=B2+F2` and copy it down to cell G359. The corresponding records for panelists 20014 and 20069 are given in Figure 8. We observe that the desired “shifting” of purchase dates has occurred.

The next step is to create a table that tells us how many trial, first repeat, etc. purchases occurred in each week. We create a pivot table in which we use *shWeek* as the *row field*, *DoR* as the *column field*, and *ID* as the *data item*. (Don’t forget to include the records with a depth-of-repeat level of -1.) The resulting table is reported in the **Pivot Table 3** worksheet. A cleaned-up summary that reports these weekly transactions in cumulative form is given in the **DoR -- ‘shifted’ transactions** worksheet.

	A	B	C	D	E	F	G
1	ID	Week	Day	Units	DoR	Offset	shWeek
16	20014	4	2	1	0	0	4
17	20014	4	4	1	1	1	5
18	20014	6	6	2	2	0	6
19	20014	7	2	3	3	0	7
20	20014	7	6	3	4	1	8
21	20014	12	5	2	5	0	12
22	20014	17	6	1	6	0	17
23	20014	23	4	2	7	0	23
24	20014	47	6	2	8	0	47
148	20069	18	1	1	0	0	18
149	20069	18	5	1	1	1	19
150	20069	19	4	2	2	1	20

Figure 8: “Shifted” Transaction Histories for Panelists 20014 and 20069

The differences between the “raw” and “shifted” cumulative transaction counts by depth-of-repeat level are small; there are only nine depth-of-repeat/week occasions on which the two sets of numbers differ, and the maximum deviation is one transaction. Perhaps the most obvious difference is with respect to first repeat. Looking at the “raw” numbers, we observe a first repeat purchase in the first week the product was on the market. Under the assumption that a trial and first repeat purchase cannot occur in the same week, this first repeat purchase is shifted to week 2 in the “shifted” numbers.

Now, how do we create a “shifted” sales volume summary? That should be obvious by now . . . and is left as an exercise for the interested reader.

References

- Clarke, Darral G. (1984), “G. D. Searle & Co.: Equal Low-Calorie Sweetener (B),” Harvard Business School Case 9-585-011.
- Eskin, Gerald J. (1973), “Dynamic Forecasts of New Product Demand Using a Depth of Repeat Model,” *Journal of Marketing Research*, **10** (May), 115–129.
- Fader, Peter S. and Bruce G. S. Hardie (2004), “The Value of Simple Models in New Product Forecasting and Customer-Base Analysis,” *Applied Stochastic Models in Business and Industry*, forthcoming.
- Parfitt, John (1986), “Panel Research,” in Robert M. Worcester and John Downham (eds.), *Consumer Market Research Handbook*, third revised and enlarged edition, London: McGraw-Hill.
- Rangan, V. Kasturi and Marie Bell (1994), “Nestlé Refrigerated Foods: Contadina Pasta & Pizza (A),” Harvard Business School Case 9-595-035.
- Sudman, Seymour and Brian Wansink (2002), *Consumer Panels*, second edition, Chicago, IL: American Marketing Association.