marketing insights

Data, but No Information

Presentation really is everything—or close to it.

By Andrew Ehrenberg

Marketing data are mostly poorly presented and difficult to understand. Exhibit 1 from an article in the *Journal of Marketing Research* (*JMR*) illustrates this.

The authors reported certain simulated repeat-buying and switching probabilities for four leading brands of U.S. paper towels and private labels, in three market segments that they had constructed. But it's not clear what the table says. What patterns does it show? It just seems a jumble of undigested data. Whoever produced it either did not know what his data was saying or was not letting on.

The same data communicates much better when displayed as in Exhibit 2, having been subjected to five traditional "Data Reduction" rules or guidelines:

- Order by some aspect of size (here separating off the large Repeat-buying figures that were in the diagonals in Exhibit 1 from the small Switching off-diagonals)
- Round drastically to one, or at most two, effective digits (effective digits are ones that vary in that part of the data)
- Use averages to provide a visual focus as well as a summary
- Use layout and labeling to guide the eye
- Give a brief verbal summary

Thus, with Exhibit 2 the reader can see that the Repeat figures are almost all high, averaging at about .7 and that the Switching ones are low, averaging at .1.

Interestingly, this "high/low" result in Exhibit 2 is close to one that my colleague, Neil Barnard, obtained in a 1993 unpublished consultancy analysis on paper towels based on disaggregated scanner-panel data for more than 13,000 households. If Exhibit 1 had been based on real data, the accompanying *JMR* paper could have been less speculative, simpler, and shorter.

Faced with Exhibit 1 we all lack numeracy. Faced with Exhibit 2 most readers can, I think, cope. It fulfils the criterion of a good table, namely that we can see what it says, at least once we know what that is (i.e., high Repeats, low Switching). Yet in Exhibit 1 this message is still not clear, even though we now know it.

The five rules of data reduction work wonders in turning data into information. They let us eyeball and then check our data by enabling our fragile short-term memory to cope in taking in and relating the different numbers.

To illustrate the rules further, I consider some data on performance measures for functional product variants of fabric conditioners. This illustrates the decision-related use of a descriptive model as discussed in my Fall 2001 "Marketing Insights" column in this magazine.

Unlike the hypothetical simulations in Exhibit 1, Exhibit 3 on page 38 is based on the real-life

EXHIBIT 1 Simulated segment-level transition probablilties

Segment	Size (%)	Percentage TAE Error	Brand	Transition Probabilities: PL	Bounty	Brawny	Scott	Viva
1	23.41	3.9	PL	.2076	.1009	.1835	.4489	.0591
			Bounty	.0192	.9555	.0091	.0110	.0052
			Brawny	.0995	.0495	.6382	.1706	.0422
			Scott	.0456	.0394	.0382	.8518	.0250
			Viva	.0520	.0091	.1045	.0443	.7901
2	41.05	3.7	PL	.8687	.0278	.0555	.0444	.0036
			Bounty	.1100	.8689	.0137	.0063	.0011
			Brawny	.1807	.1121	.6797	.0100	.0175
			Scott	.0411	.0010	.0091	.9213	.0275
			Viva	.0394	.0008	.0356	.0099	.9143
3	35.54	6.9	PL	.4412	.1699	.1900	.0983	.1006
			Bounty	.0523	.9307	.0099	.0007	.0064
			Brawny	.2110	.0577	.6789	.0199	.0325
			Scott	.3299	.0345	.0445	.4999	.0912
			Viva	.1445	.1134	.0466	.1933	.5022

Source: Bockenholt U. and W.R. Dillon (2000), "Inferring Latent Brand Dependencies," *Journal of Marketing Research*, 37 (February) 85.

empirical IRI Behaviorscan data. Exhibit 3 gives the observed O values of six performance measures like market share, penetration, and average purchase frequency, together with theoretical Dirichlet predictions T based on the market shares.

Exhibit 3 is set out in the traditional journal format—in alphabetical order of the product formulae from Light to Unscented, the columns equally spaced across the page, with ample precision to four digits. It seems hard to fathom. What patterns, if any, are there?

This is where the five rules again help. (See Exhibit 4.)

Rule 1: Order by size. In Exhibit 4 the rows have been ordered by a measure of size, namely market share. (If there are many such tables, one should generally stick to the same order.)

Exhibit 4 lets us readily see not only how the market shares decrease from 72% to 3%, but also whether or not the other columns largely follow the same order. They in fact do so—for Penetration, Average Purchase Frequency, the incidence of Light Buyers (in reverse order, naturally!), and the others.

We can also see more readily than in Exhibit 3 how far the Observed and Theoretical figures, O and T, agree. Exceptions stand out more readily, too—certainly compared with Exhibit 3—for example, the high observed incidence of 100%-loyal buyers of Light, and the smallish number of once-only buyers of Unscented.

Rule 2: Rounding to two effective digits. Rounding numbers to just one or at most two effective digits makes the data vastly easier to read and remember than the longer numbers in Exhibit 3. It's also much easier to manipulate short numbers mentally (e.g., to take ratios or percentages). In Exhibit 3 we would mostly find it difficult to subtract say 19.82% from 72.37% in our heads, let alone to divide one by the other (without mental rounding). But we can all see that 72 is nearly four times 20. People's ability to cope with two-digit numbers, but not longer ones, has been widely established. No compromise seems possible.

Thus saying as another example that sales of product X rose from 28,942 to 94,347 and those of Y rose from 13,492 to 36,051 does not tell us that sales have trebled, unless we round the figures mentally. This could have been done for us on paper: Sales of X rose from 29,000 to 94,000, and those of Y rose from 13,000 to 36,000.

Now, many of us can even see by simple mental arithmetic that X increased just over three-fold, and Y just under three-fold. But we still would not be able to see this even with just three digits (e.g., 28, 900 and 94,300) without mentally rounding further. With long numbers we are all non-numerate. With two-digit ones, we can cope.

Rounding is the only one of the five rules where some precision is lost. But any such loss of accuracy is less than is routinely accepted in using mathematical models or graphs. One can also reassure the reader that "the full data are available on the Web" (and wait for the rush).

Marginal exceptions to such two-digit rounding occur in our decimal system with numbers near 100 (e.g., 104 and 92). Here, rounding the 104 to 100 to just two effective digits would be over-rounding. In contrast, deliberate over-rounding to just one digit can at times be effective (e.g., reporting a probability or a correlation as .6, not as .63). And variable rounding works well when figures vary greatly in size, as in much accounting data. The table below illustrates this.

	Full ($)	Variable rounding
Sales:	329,176,540	330,000,000
Profits:	6,364,179	6,400,000
Charities:	239,485	240,000

Rule 3: Give averages. Row and column averages in a table provide a visual and mental focus. In Exhibit 3 on the next page, for instance, to get a view of the market shares in the first column, we have to compare all the different four-digit shares with each other in our heads—20.02% with 72.37%, 72.37% with 3.76%, 3.76% back with 20.02%, and so on. This is quite tedious to do and to remember.

But in the first column of Exhibit 4 we can compare each entry with just the one single baseline average, the 25. We can now see which of the figures are relatively high compared with

EXHIBIT 2 The simulated repeat and switching probablilties from Exhibit 1: Reformatted

Segment	**I.** (23%)						**II.** (41%)						**III.** (36%)					
	Repeat	**Switching**					**Repeat**	**Switching**					**Repeat**	**Switching**				
Brand		PL	Bt	Bw	Sc	Vi		PL	Bt	Bw	Sc	Vi		PL	Bt	Bw	Sc	Vi
PL	.2†		.1	.2	.4	.1	.9		.0	.1	.0	.0	.4		.2	.2	.1	.1
Bounty	.9	.0		.0	.0	.0	.9	.1		.0	.0	.0	.9	.1		.0	.0	.0
Brawny	.6	.1	.0		.2	.0	.7	.2	.1		.0	.0	.7	.2	.1		.0	.0
Scott	.9	.0	.0	.0		.0	.9	.0	.0	.0		.0	.5	.3	.0	.0		.1
Viva	.8	.1	.0	.1	.0		.9	.0	.0	.0	.0		.5	.1	.1	.0	.2	
AV	**.7**	**.1**	**.0**	**.1**	**.2**	**.0**	**.9**	**.1**	**.0**	**.0**	**.0**	**.0**	**.6**	**.2**	**.1**	**.1**	**.1**	**.1**

† The leading diagonal .2076 in Exhibit 1, rounded.

25 (i.e., Regular and Light at 72 and 20) or low (Unscented and Stainguard at about 3 or 4). And we can more readily remember this. The memory strain is less.

An average need not be "typical" in order to be useful. Indeed, calculating an average lets us see whether or not it *is* typical of the constituent numbers. Averages seem to be the main tool of practical data analysis. A regression line, for example, is simply the average value of y for a given value of x.

Rule 4: Layout to guide the eye. The layout of many marketing research tables hinders the reader rather than helping. Typical of many journals, Exhibit 1 is spread across the whole page. This requires us to move our eyes from left to right, which is well-known to interrupt our short-term memory, as it is easy to test. Thus when we look at the numbers on the right of Exhibit 3 (35.97, 33.24, etc.) we will have forgotten the formula names on the left. When we therefore look back at these names on the left (Light, Regular, etc.) we will have forgotten the numbers on the right.

Exhibit 4 requires much less eye movement and is therefore easier to take in. Other helpful changes in layout are more purposeful spacing of columns and some selective vertical rules to guide the eye.

Rule 5: A verbal storyline. If the author of a table knows what his data are saying, he should let readers know, rather than expecting each reader to reinvent that wheel. (If you don't know what your data are saying, *don't* show the data.) With a brief verbal storyline already in mind (e.g., "Repeats high, Switching low"), it's much easier to take in the detailed data (e.g., to see the exceptions).

IMPLEMENTATION

When, as marketing professionals, we look at tables of data and are honest with ourselves, we find most of them unreadable. The fault is not in ourselves, but in our data. Or with the producers of our data. Few producers realize their tables could easily be made to communicate much better.

As readers we are to blame only if we opt out of the challenge of bringing about improvements: We must all learn to return a bad table to its author. It then helps if we can say what the author should do with it, like giving a brief verbal summary (Rule 5). The author/analyst would then actually learn what, if anything, his data means, before passing them on to others.

The five rules of data reduction seem very simple—mere common sense. But common sense does not mean common knowledge or common practice. The rules cannot be implemented overnight. In practice, this requires effort: There will

EXHIBIT 3 Six performance measures: "Formula" variants (observed O and theoretical T predictions)

Conditioners `91 (Philadelphia)	Size-Related Market Share %	Penetration %		Loyalty-Related Purch. Freq.		Once Only %		100%-Loyals %		SCR	
FORMULAE		O	T	O	T	O	T	O	T	O	T
Light	19.82	31.96	39.74	2.96	2.44	14.32	18.69	35.97	8.31	33.24	24.02
Regular	72.37	63.41	62.10	5.03	5.35	3.94	5.23	41.32	34.84	74.01	72.74
Stainguard	3.16	8.88	10.70	1.79	4.42	40.79	56.00	9.62	4.22	19.04	10.21
Unscented	4.43	9.24	13.09	1.96	1.39	26.01	51.13	14.34	3.98	27.62	10.87

Data Source: IRI (*n*=860 households) SCR: Share of Category Requirements

EXHIBIT 4 Performance measures rounded with average

Conditioners `91 (Philadelphia)	Size-Related Market Share %	Penetration %		Loyalty-Related Purch. Frequ.		Once Only %		100% loyal %		SCR %	
FORMULAE		O	**T**	O	**T**	O	**T**	O	**T**	O	**T**
Regular	72	63	**62**	5.0	**5.4**	4	**5**	41	**35**	74	**73**
Light	20	32	**40**	3.0	**2.4**	14	**19**	36	**8**	33	**24**
Unscented	4	9	**13**	2.0	**1.4**	26	**52**	14	**4**	28	**11**
Stainguard	3	9	**11**	1.7	**1.4**	41	**56**	10	**4**	19	**10**
Average	25	28	**32**	2.9	**2.7**	21	**33**	18	**13**	39	**30**

be quite large set-up costs and up-set costs. Management will therefore need to allocate people and budgets to make tables readable. It is not enough just to say "avoid long numbers" or "use good layout." We need also to encourage, train, and monitor people (including ourselves) and their output.

One starting point can be with editors. They are usually not marketing-orientated but production-orientated, dominated by their house style: "That's how we do it." Their concern is mainly with how their competitors might react (i.e., other editors), not with their readers (the consumer). Typically, journals either allow virtually no vertical rules in their tables to help the reader (vertical rules were difficult to set in cold type 100 years ago). Or each number is put into a separate grid-like box (like was done to my Exhibit 1 in the Fall 2001 issue of this magazine). This hinders the reader's eye from moving across each of the rows of that Exhibit, as I, the author, had intended readers to be able to do.

There also is a constant worry about accuracy, mainly among people who never really use the data anyway. But a friendly commentator has pointed out that the loss in accuracy in rounding X's sales of 13,492 down to 13,000 just now was only 3.78%! (Expressing the rounding error as 4% would have made this easier to remember, to use, and to relate to other such numbers, with less strain on our memory.)

Producers of tables of data usually do not realize how unnecessarily incomprehensible their tables mostly are. They often also think that how their data are to be presented depends on one's purpose (e.g., either to persuade or to inform). But no. Tables that follow the five rules seem to communicate better for any purpose. Try it and see. ●

Author's Note: This article is derived from *Marketing Learning* (Oct. 1998), produced initially for participants of our *R&D Initiative* (www.sbu.ac.uk/RandDI). This was in turn based on earlier articles in the *American Statistician* (1981) and in *Admap* (1992).

Additional Reading

Ehrenberg, A.S.C. (1972), *Data Reduction: Analysis and Interpreting Statistical Data*. New York and Chichester: John Wiley (available www.empgens.com).

—— (1982, 2000). *A Primer in Data Reduction*. New York and Chichester: John Wiley.

Making Numbers Work (A 24-minute training video). London: Melrose Learning Resources (fax: 44 (0) 20 7622 0421).

Andrew Ehrenberg is chairman of South Bank University's *R&D Initiative*, a program of basic research into marketing issues supported by some 90 North American and European companies. He may be reached at ehrenba@sbu.ac.uk.

U.S. Postal Service Statement of Ownership, Management, and Circulation (Required by 39 U.S.C. 3685)

1. Title of Publication: *Marketing Research: A Magazine of Management & Applications*
2. Publication No. 1040-8460
3. Filing Date: October 1, 2001
4. Issue Frequency: Quarterly
5. No. of Issues Published Annually: 4
6. Annual Subscription Price: $70
7. Complete Mailing Address of Known Office of Publication: 311 South Wacker Dr., Suite 5800, Chicago, IL 60606-2266
8. Complete Mailing Address of the Headquarters or General Business Offices of the Publisher: 311 South Wacker Dr., Suite 5800, Chicago, IL 60606-2266
9. Full Names and Complete Mailing Address of Publisher, Editor, and Managing Editor: Publisher: Jack Hollfelder, 311 S. Wacker Dr., Ste. 5800, Chicago, IL 60606-2266; Editor: Chuck Chakrapani, Standard Research Systems Inc., 380-26 Soho St., Toronto, ON M5T 1Z7; Managing Editor: Mary Egan, 311 S. Wacker Dr., Ste. 5800, Chicago, IL 60606-2266
10. Owner: American Marketing Association, 311 South Wacker Dr., Suite 5800, Chicago, IL 60606-2266
11. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1% or More of Total Amount of Bonds, Mortgages, or Other Securities (If there are none, so state): None
12. Tax Status (for nonprofit organizations authorized to mail at special rates). The purpose, function,and nonprofit status of this organization and the exempt status for federal income tax purposes: Has Not Changed During Preceding 12 Months
13. Publication Name: *Marketing Research: A Magazine of Management & Applications*
14. Issue Date for Circulation Data Below: Fall, 2001
15. Extent and Nature of Circulation:

	Average no. copies each issue during preceding 12 months	Actual no. copies of single issue published nearest to filing date
a. Total No. Copies (Net Press Run)	4,750	4,800
b. Paid and/or Requested Circulation		
(1) Paid/Requested Outside-County Mail Subscriptions	3,612	3,455
(2) Paid/Requested In-County Mail Subscriptions	0	0
(3) Sales Through Dealers and Carriers, Street Vendors, Counter Sales & Other Non-USPS Paid Distribution	0	0
(4) Other Classes Mailed Through the USPS	0	0
c. Total Paid and/or Requested Circulation *(Sum of 15b(1), (2), (3) and (4))*	3,612	3,455
d. Free Distribution by Mail *(Samples, Complimentary, and Other Free)*		
e. Free Distribution Outside the Mail (Carriers or Other Means)	35	35
f. Total Free Distribution *(Sum of 15d and 15e)*	0	0
g. Total Distribution *(Sum of 15c and 15f)*	35	35
h. Copies Not Distributed	3,647	3,490
i. Total *(Sum of 15g and h)*	1,103	1,310
Percent Paid and/or Requested Circulation *(15c/15g x 100)*	4,750 99.0% paid circulation	4,800 99.0% paid circulation

16. This Statement of Ownership will be printed in the Winter, 2001, issue of this publication.
17. Signature and title of editor, publisher, business manager, or owner: Jack Hollfelder, Publisher. I certify that the information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions *(including fines and imprisonment) and/or civil sanctions (including multiple damages and civil penalties).*