APPENDICES: CONTENTS

PREFACE TO THE APPENDICES

The Appendices show how to do the calculations in this book. To gain practice this can be tackled in two ways, by working through small-scale numerical examples by hand or by using a computer package. We strongly advise the reader to do both: to "get your hands dirty" and to acquire some helpful computing experience before attempting analyses of larger data sets of this kind.

The worked examples are presented in two Appendices, for an individual brand and for multi-brand buying. In Sections A. 1 to A. 11 of Appendix A, we show how to calculate the theoretical NBD/LSD repeat-buying statistics. In § A.12 to A.14, which are new to this edition, we comment briefly on the appropriate tabulations of observed data. Appendix B sets out some look-up tables of theoretical statistics of the NBD/LSD model which remain useful even in these days of rapid computing.

In Appendix C, which again is new, the calculation of the Dirichlet parameters is illustrated. This provides alternative predictions of single-brand buying behaviour and, more importantly, it gives us theoretical values for multi-brand buying. Appendix C also outlines how to tabulate the relevant observed data, and how the results can be used to check the assumptions of the Dirichlet model.

Another way to get a feel for the models is to run trials using a computer package. One version of such software is available (for a small fee) on a floppy disc designed for IBM personal computers, though it will transfer to other personal computers and mainframe systems*. On the disc are two programs, written in standard FORTRAN 77; one calculates all the theoretical single-brand statistics and the other is for multi-brand calculations. There are two versions of each program, to allow the user to choose whether to tabulate the raw data and automatically derive the NBD or Dirichlet predictions, or merely to input some summary statistics to obtain the theoretical predictions. Test data are given on the disc, along with some examples of the output.

In the US, consumer panel data are available on a commercial basis from MRCA Information Services, from IRI (Information Resources Inc.), and from NPD (National Purchase Diary), other services are also being developed. In Britain the main source of such data is AGB Ltd. Consumer panels are also operated in a number of other countries in Western Europe,

*The disc is being distributed by Dr M. D. Uncles at the Centre for Marketing and Communication, London Business School, London NW14SA, England.

Japan, etc. Most firms are prepared to release small amounts of data for academic research. It is important when asking for data to have a sample of continuous reporters over the whole period concerned (e.g. a year), including non-buyers of the particular product, or at least to know approximately how many there are (see **A.14**).

APPENDIX A

CALCULATIONS FOR AN INDIVIDUAL BRAND

A. 1. Tabulations and Calculations

In this Appendix we mainly give a worked numerical example of the mathematical calculations involved in using the NBD/LSD repeat-buying theory. Application of the theory also involves two other things, namely tabulation of the observed data, and a background of past experience of similar work to help in interpreting the results. In § A.12 to A.14 we give guidance on the tabulations.

The tabulations which are required in repeat-buying studies are in principle quite simple. This is primarily because in studying the observed repeat-buying of any particular brand or other type of item, only purchases of that item are involved and purchasing of other brands can be ignored. For the newcomer to this kind of work it is best to "get his hands dirty" by carrying out some initial tabulations himself, so as to get the feel of the data. Only after that should the work be delegated to a computer. The latter is of value for extensive work on a more or less repetitive basis, but it can take time and effort to get a sufficiently wide and flexible range of programs working; time and money can be saved if you use the programs on our disc to gain initial experience and then adapt them for your own software needs.

The theoretical formulae which are then needed may be tackled in one of three ways. Firstly, using special numerical tables from which the required answers can simply be read off (or estimated by interpolation). A number of such tables are set out in Appendix B, although they are mainly intended to provide a "feel" for the results, and are not always detailed enough for routine use. Secondly, the NBD calculations can be computerised, if and when their routine use for large data files is required, as on our disc.
Thirdly, the theoretical formulae can be worked out by direct hand-calculation in each particular case. This is important to do for the newcomer who wishes to understand how the formulae work (or for those engaged in further research into buyer behaviour). A worked example of such calculations is now given.

A.2. A Worked Example

To familiarize the calculations that are required, a worked numerical example is given of all the theoretical calculations required in Chapter 3.

The most general theoretical formulae are the NBD ones, but for hand-calculation the LSD formulae or the simplifying approximations to them which were set out in Chapters 4 and 8 are easier to use in those ranges of the parameters b and w where they give virtually the same results as the NBD. In what follows, we therefore give for each case the NBD, the LSD and the simplified "approximate" forms of calculation.

Brand E has been chosen for the example because it is in the parameter range where both the LSD and the further simplifications apply, in that they give almost the same results as the NBD. (In working through a further numerical exercise, it may be best to choose Brand A in Chapter 3 to learn to what extent the LSD results *differ* from the NBD ones for high penetration levels.) In what follows some deliberate variation has been included in the time-period chosen to illustrate the *detailed* calculations, but the numerical results for all lengths of time-periods covered in Chapter 3 are given.

A.3. The Basic NBD and LSD Parameters

In most of the calculations, the NBD parameter k (or $a = m/k$) and the LSD parameter q (or $a = q/(1-q)$) are required for the given time-period. These values are obtained now for Brand E in the 4-, 12-, 24- and 48-week periods covered in Chapter 3 *. They are set out in Tables Al and A2 for ease of subsequent reference.

For the NBD parameter k, a knowledge of the penetration b and the average buying frequency per buyer w is required as input. The observed values for Brand E are given for various time-periods in the "observed" columns of Tables 3.1 a and 3.2a respectively. (The values are generally averages of several periods of the stated length, but this does not affect the nature of the calculations here. Results for each separate quarter are for example given in Tables 3.1 and 3.2.)

* In a 1-week period, the observed w is only 1.01 (see Table 3.2a) and no NBD or LSD can usefully be fitted.

Table A1. The Calculation of the NBD Parameters k or a from b and w

Brand E	Period of Length (in weeks)			
	4	12	24	48
Given Data:				
b (Table 3.1a)	.04	.07	.09	.12
w (Table 3.2a)	1.6	3.0	4.9	6.8
$p_0 = 1-b$	.96	.93	.91	.88
$m = bw$ *	.064	.21	.441	.816
Derived Statistics				
$c = -m/\ln(p_0)$	1.57	2.89	4.68	6.38
a (Table B3)*	1.32	5.34	12.00	19.18
$k = m/a$ *	.0485	.0393	.0367	.0425

* For footnote see below.

Table A2. The Calculation of the LSD Parameters q or a from w

Brand E	Period of Length (in weeks)			
	4	12	24	48
Given Data:				
w (Table 3.2a)	1.6	3.0	4.9	6.8
Derived Statistics:				
a (Table B3)*	1.40	5.71	12.89	21.0
$q = a/(1+a)$	.584	.851	.928	.955

* Note that the situation is not *quite* stationary (i.e. the values of m and a *are* not quite proportional to the lengths of the periods, and k is not quite constant).

One case, for the 4-weekly data, will be given in some detail; the calculations for the other time-periods will be shown in summary form only. (This form of presentation will also be followed in other sections, where appropriate.)

The NBD Parameter k. From Tables 3.1a and 3.2a, b (the penetration expressed as a proportion) and w (the buying frequency per buyer) for Brand E in 4 weeks are

$$b = .04,\ w = 1.6\ .$$

The proportion of non-buyers, p_0, is therefore given by

$$p_0 = 1 - b = .96\ ,$$

and the mean number of purchases per sample-member, m, by

$$m = bw = .064\ ^*.$$

To obtain the value of the NBD parameters k or a we have to solve the equation $p_0 = (1+m/k)^{-k}$. This is usually best done (as mentioned in §4.2) by calculating a certain quantity $c = -m/\ln(p_0)$ from m and p_0, where "ln" stands for the Natural Logarithm **. We then use Table B2 in Appendix B to read off the corresponding value of a for the given value of c. Thus

$$c = \frac{-m}{\ln(p_0)} = \frac{-.064}{\ln(.96)} = \frac{.064}{.0408}$$

$$= 1.569.$$

From Table B3 (by interpolation)

$$a = 1.32$$

and hence

$$k = m/a = 0.0485\ .$$

Table Al summarises this calculation for the 4-week period and the corresponding ones for the remaining time-periods, of 12, 24 and 48 weeks.

* In practice the value of m would usually be given by direct tabulation of the data.

** Appropriate values of Naperian or Natural Logarithms to base e are given in Table B3 in Appendix B.

The LSD Parameter q. Next, we describe the estimation of the LSD parameter q (or $a = q/(1-q)$). Here only w is required as input and we need to solve the equation $w = -q/(1-q)\ln(1-q)$. The best way is to use Table B2 which was used in the NBD calculation above but which also gives the value of the LSD parameter a for a given w, and hence the LSD parameter q (as $q = a/(1+a)$). In the 4-week period for Brand E, we have (from Tables 3.2a or A2) that

$$w = 1.6.$$

Using Table B2 gives

$$a = 1.402\,,$$

and q therefore is

$$q = \frac{a}{1+a} = \frac{1.402}{2.402} = .584\,.$$

Table A2 summarises this calculation and those for the other time-periods.

A table from which the value of the LSD q can be read off *directly* for a given w is given as Table B 1 in Appendix B (see also Table 2.2). Accurate interpolation for relatively high values of q is however relatively difficult, and it is usually better to calculate a first, using Table B2.

More generally, we may note that the values of b and w in the various time-periods here are mostly in the range where the LSD theory tends to give very similar results to the NBD (as will be illustrated by the following sections). The NBD and the LSD versions of the parameter a *are* however not identical, although Tables Al and A2 show them to be similar.

A.4. Penetration Growth (Table 3.1a)

The theoretical norms for the change of penetration of Brand E in various length time-periods in Table 3.1a were derived from the observed results in the average *12-week* period. For the NBD, we need to solve the equation $b_T = 1-(1+Tm/k)^{-k}$ as given in §4.4 (Table 4.10) and in §7.5. The lengths of the other time-periods for which the predictions are made are expressed as multiples of the 12-week period, i.e.

by calling this the base-period with $T = 1$. For the other periods, we therefore have

$$\begin{aligned} &\text{1 week;} && T = 1/12 = .083, \\ &\text{4 weeks;} && T = 4/12 = .333, \\ &\text{24 weeks;} && T = 24/12 = 2, \\ &\text{48 weeks;} && T = 48/12 = 4. \end{aligned}$$

The full NBD calculation is illustrated now for the one-week period. Using the NBD formula $b_T = 1-(1+Tm/k)^{-k} = 1-(1+Ta)^{-k}$, where $m = .2$, $a = 5.34$, $k = .0393$ are the parameters obtained from the 12-week period (Table Al), and $T = .083$, we have

$$\begin{aligned} b_{.083} &= 1-(1+.083 \times 5.34)^{-.0393} \\ &= 1-(1.44)^{-.0393} \\ &= 1-.986 = .014, \end{aligned}$$

or just over 1% (the "Theoretical" estimate shown in the l-week column of Table 3. la). In other words, given that 7% of the sample bought Brand E on average 3.0 times in the 12-week period (Table Al), the NBD estimate is that about 1.4% of the sample should buy Brand E in the typical *l-week* period (and this compares with the *observed* 1-week penetration of roughly 2%, as was shown in Table 3.1 a). This result and these for the longer time-periods are summarised in Table A3.

The corresponding LSD formula (Table 4.10 and § 8.5) is expressed in terms of the ratio of b_T to b, the penetration in the 12-week period ($T = 1$),

$$\frac{b_T}{b} = 1 - \frac{\ln(1+(T-1)q)}{\ln(1-q)},$$

where $b = .07$ and $q = .85$ 1, the LSD parameters in the 12-week period (Table A2).

The quantity $\ln(1-q)$ in this formula enters into the calculation for periods of *any* length and may first be calculated:

$$\ln(1-q) = \ln(.149) = -1.904.$$

Table A3. The NBD Estimates of the Penetration b_T from the **12-Weekly** Data ($T=1$)

Brand E	Period of length (in weeks)			
	1	4	24	48
T	.083	.333	2	4
$1+Ta$	*1.44*	*2.78*	*11.68*	*22.36*
$(1+Ta)^{-k}$	.986	.961	.908	.885
b_T	.014	.039	.093	.115
$100\,b_T$	1%	4%	9%	11%
Observed 100 b_T	2%	4%	9%	12%

Table **A4.** The LSD Estimates of b_T

Brand E	Period of length (in weeks)			
	1	4	24	48
T	.083	.333	2	4
$(T-1)q$	–.780	–.568	.851	2.553
$\ln(1+(T-1)q)$	-1.514	–.839	,615	1.268
b_T/b	.205	.559	1.323	1.666
b_T	.014	.039	.093	.117
100 b_T	1%	4%	9%	12%

Table **A5.** Estimates of b_T from the "Approximate" Formula

Brand E	Period of length (in weeks)			
	1	4	24	48
T	.083	.333	2	4
$T^{.82}$	.13	.41	1.76	3.10
b_T/b	.198	.552	1.33	1.67
b_T	.014	.039	.093	.117
100 b_T	1%	4%	9%	12%

For a single week ($T = .083$), we therefore have

$$\frac{b_{.083}}{b} = 1 - \frac{\ln\{1+(.083-1).851\}}{-1.904}$$

$$= 1 - \frac{\ln(1-.917 \times .851)}{-1.904}$$

$$= 1 - \frac{1.514}{1.904} = .205, \qquad \text{so that}$$

$$b_{.083} = .07 \times .205 = .014 \text{ or } 1\%,$$

as for the NBD. The calculations are summarised in Table A4. Note that for the 48-week period the NBD and LSD estimates differ slightly (.115 and .117).

Finally, we can use the approximation to the LSD formulae

$$\frac{b_T}{b} = \frac{Tw}{1+(w-1)T^{.82}},$$

where w is the buying frequency in the unit period (Tables 4.10 and §8.5). For Brand E in 12 weeks, w = 3.0, so that

$$\frac{b_T}{b} = \frac{3T}{1+2T^{.82}},$$

The estimate of the penetration in a single week ($T = .083$) is therefore

$$\frac{b_{.083}}{b} = \frac{.249}{1+2(.083)^{.82}}$$

$$= \frac{.249}{1+2 \times .13} = \frac{.249}{1.26}$$

$$= .198, \qquad \text{so that}$$

$$b_{.083} = .198 \times .07 = .014,$$

the same value as given by the LSD calculation itself. Table A5 summarises these calculations for all the time-periods.

A.5. *The* Average Frequency of Buying (Table 3.2a)

We now turn to estimating w_T, the average frequency of purchase per buyer, in time-periods of relative length T, from the b and w results in the "unit" time-period, here 12 weeks (as given in Table 3.2a).

The NBD formula for w_T is

$$w_T = \frac{Tm}{\{1-(1+Tm/k)^{-k}\}},$$

where m and k are the parameters of the 12-week period for Brand E with

$$m = .21, \quad k = .0396.$$

We also have $m/k = a = 5.34$.

To calculate the 4-week value of w_T (with $T = 4/12 = .333$), *we* therefore have

$$w_{.333} = \frac{.333 \times .21}{\{1-(1+.333 \times 5.34)^{-.0396}\}}$$

$$= \frac{.07}{1-(2.778)^{-.0396}}$$

$$= \frac{.07}{.039} = 1.8.$$

Note that the denominator in the last expression is the estimated value of $b_{.333}$ which can be read off from Table A3 and need not really be recalculated. Table A6 summarises the calculations for the various periods of time and makes use of b_T from Table A3 in this way.

The LSD formula expressed as the ratio w_T/w is

$$\frac{w_T}{w} = \frac{T\ln(1-q)}{\ln(1-q)-\ln(1+(T-1)q)},$$

in terms of q, where $q = .85\,1$ and w = 3.0 for Brand E in 12 weeks ($T = 1$).

Table A6. The NBD Estimates of the Average Purchase Frequency w_T

Brand E	1
T	.083
$T^{.82}$	.13
w_T	1.3

For 4 weeks (with $T = .333$ again),

$$\frac{w_T}{w} = \frac{,333 \ln(.149)}{\ln(.149) - \ln(1-.568)}$$

$$= .333 \left(\frac{-1.904}{-1.904+.839}\right)$$

$$= \frac{.333}{.559}$$

$$= .596 .$$

We therefore have the estimates $w_T = 3.0 \times .596 = 1.8$ in 4 weeks. The numerical value of the denominator in the expression for w_T/w (i.e. .559) need not be directly calculated if the LSD calculations for b_T have been carried out, since it is the value of b_T/b in Table A4. This is the procedure used in the summary Table A.7.

Finally we have the approximation to the LSD formula (Table 4.12)

$$(w_T - 1) = (w-1) T^{.82} .$$

Substituting w = 3 and rearranging gives

$$w_T = 1 + 2T^{.82} .$$

In 4 weeks this gives

$$w_{.333} = 1 + 2(.333)^{.82}$$

$$= 1 + 2 \times .406$$

$$= 1.8.$$

The values of $T^{.82}$ are already available in Table AS, for the various values of T, and Table A8 shows the approximation for all the time-periods.

It should perhaps be stressed that the reason why the calculations for b_T (Table 3.1 a) and w_T (Table 3.2a) have so much in common is that if b_T is known then w_T can easily be calculated from the two simple relationships

$$m_T = Tm \quad and \quad w_T = m_T/b_T.$$

The calculations set out here are slightly more lengthy to illustrate the various formulae for w_T explicitly.

A.6. Light and Heavy Buyers (Table 3.4)

We now describe the calculations for the theoretical frequency distributions as shown in Table 3.4. Given b and w (and hence k or a) in a given period, we can estimate how many people bought 0, 1, 2, 3 etc. times in this period.

In the NBD, p_r is the theoretical proportion of the population buying the brand r times in the chosen time-period (see §§4.2 and 7.3). The numerical values of p_r *are* best calculated using the recursive formula,

$$p_r = \left(\frac{a}{1+a}\right)\left(1 - \frac{a-m}{ar}\right)p_{r-1} .$$

For Brand E in 48 weeks, we have (Table Al) $a = 19.18$, and $m = .816$.

Therefore,

$$p_r = \left(\frac{19.18}{20.18}\right)\left(1 - \frac{19.18-.816}{19.18r}\right)p_{r-1}$$

$$= 0.9504\left(1 - \frac{.9575}{r}\right)p_{r-1} .$$

The observed penetration of Brand E in 48 weeks was $b = .12$ (Table 3.1 a or Table Al), and so $p_0 = 1-b = .88$. Using this value we can now estimate p_1, and proceed to higher values:

$$p_1 = .9504(1-.9575)\,.88 = .0355,$$

$$p_2 = .9504(1-.4787)\,.0355 = .0176,$$

$$p_3 = .9504(1-.3192)\,.0176 = .0114,$$

$$p_4 = .9504(1-.2394)\,.0114 = .0083,$$

$$p_5 = .9504(1-.1915)\,.0083 = .0064,$$

and so on. To estimate the cumulative "tail" for 6 or more purchases, we have

$$p_{6+} = 1 - p_0 - p_1 - p_2 - p_3 - p_4 - p_5 = .0408.$$

These proportions are of the population as a whole. Expressed as p'_r, the proportion of ***buyers*** (i.e. the 12% of the population who bought E in the year) who bought r times, they are

$$p'_r = 100\,p_r/12,$$

which is the form given in Table 3.4.

The LSD formula is of course expressed directly in p'_r, i.e.

$$p'_r = \frac{-1}{\ln(1-q)}\,\frac{q^r}{r}\ .$$

The proportion buying once is therefore

$$p'_1 = \frac{-1}{\ln(1-q)}\,q = \frac{-.955}{\ln(.045)}$$

$$= .3080 \text{ or } 31\%.$$

Values for p'_r for $r > 1$ can best be calculated by the recursive formula

$$p'_r = (r-1)\,q p'_{r-1}/r\,.$$

This gives

$$p'_2 = .3080 \times .955/2 = .2941/2 = .15,$$

$$p'_3 = .2941 \times .955/3 = .2809/3 = .09,$$

$$p'_4 = .2809 \times .955/4 = .2683/4 = .07,$$

$$p'_5 = .2683 \times .955/5 = .2562/5 = .05,$$

$$p'_{6+} = 1 - .31 - .15 - .09 - .07 - .05 = .33.$$

Table A9 compares the observed percentages of light and heavy buyers with these NBD and LSD norms in the form $100p'_r$.

A.7. The Sales Importance of Light and Heavy Buyers (Table 3.5)

Table 3.5 shows in percentage terms what proportion of total sales are accounted for by those consumers who make precisely r purchases each in the analysis period. This amounts to

$$100\, rp_r/m$$

(or $100Nrp_r/Nm$ for a sample of N consumers). Since r and m are given, the theoretical values of this ratio can be simply estimated from the NBD values of p_r in § A.5, and the computational problem is almost trivial. For example, for Brand E in the 48-week period and for $r = 2$, p_2 was .176, so that the share of sales accounted for out of an m-value of .816 (Table Al), is

$$\frac{100 \times 2 \times .0176}{.816} = 4.4\%.$$

The other NBD values follow similarly, with the "tail" (e.g. purchases made by buyers buying 6+ times, say) being obtained by subtraction. The results are set out in Table A 10.

In the LSD model, the results can either be obtained by the corresponding calculations in terms of p'_r or through a special formula which gives the answer directly (see § $2.3, 4.4 and 8.4). Thus the proportion of sales accounted for by buyers who make more than r purchases is

$$q^r\,.$$

The proportion of sales accounted for by those buyers who each make exactly r purchases is therefore

$$q^{r-1} - q^r\,.$$

Only a tabulation of q^{r-1} for various values of r is therefore required, from which the required estimates can be obtained by subtracting successive terms. This is illustrated for Brand E in Table A10, the value of the 6+ tail being either obtained by subtraction or simply as q^5.

Table A9. The Percentage of Buyers Making **1, 2, 3, etc. Purchases in the Year**

Brand E	No. of Purchases in a Year					
	1	2	3	4	5	6+
Obs.	36	12	6	5	6	36
NBD	30	15	10	7	5	34
LSD	31	15	9	7	5	33

Table A10. LSD and NBD Estimates of the Percentage of the Annual Sales of Brand E Accounted for by Buyers Making *r* Purchases of the Brand E

r	1	2	3	4	5	6+
LSD Estimate						
4^{r-i}	.955	.912	.871	.832	.795	.759
$100(q^{r-1}-q^r)\%$	4.5	4.3	4.1	3.9	3.7	79.5*
NBD Estimate						
$100rp_r/m\%$	4.4	4.4	4.2	4.1	4.0	78.9*

* The 6+ tail is obtained by subtracting previous terms from 1 (and equals q^5 for the LSD).

A. 8. The Incidence of Repeat-Buyers (Table 3.6)

The commonest way of calculating the norms for repeat-buying from one period to another is from the values of *b* and w (or $m = bw$) in the first of the two periods. Sometimes it is also relevant to calculate the results backwards, i.e. given the pattern of buying in the *second* period, how many of the buyers should also have bought in the preceding one? The theoretical calculations which were given in Tables 3.6-3.9 are slightly different still, in that they were based on the *b* and w values averaged for both periods (and averaged across all the pairs of periods analysed). If we are dealing with a strictly stationary situation, the *b* and w values are the same in all the periods and none of these variations matters. In practice, some minor fluctuations in the *b* and w values occur from period to period (as documented in Tables 3.1 and 3.2), and in any specific problem-solving work this needs to be allowed for.

To illustrate the nature of the calculations here, we give the calculation based on the *average* values of *b* and w for all the periods of each length, which are given in Tables 3.1a and 3.2a. (Some of the resultant theoretical estimates therefore differ slightly from the theoretical values given in the tables in Chapter 3, which were averages of the separate theoretical estimates based on the values of b and w in each individual period.) The data for the average 4-week period are used to illustrate the detailed calculations, for the NBD, the LSD and the "approximate" formulae.

The NBD formula for b_R, the proportion of the population who are repeat-buyers, is (Table 4.6 and §7.6)

$$b_R = 1 - 2(1+a)^{-k} + (1+2a)^{-k}.$$

For Brand E in 4 weeks we have $a = 1.32$ and $k = .0485$, so that

$$\begin{aligned} b_R &= 1 - 2(1+1.32)^{-.0485} + (1+2 \times 1.32)^{-.0485} \\ &= 1 - 2 \times .9599 + .9391 \\ &= .0193\,. \end{aligned}$$

The ratio of those repeat-buying to those who bought in the first period, b_R/b, is the theoretical quantity given in Table 3.6. Thus with $b = .04$ for Brand E in 4 weeks,

$$b_R/b = .0193/.04 = .48, \text{ or } 48\,\%.$$

The NBD calculations are summarised in Table Al 1.

The LSD formula is expressed directly as the ratio of repeat-buyers to buyers in the first period,

$$b_R/b = 1 + \frac{\ln(1+q)}{\ln(1-q)}\,.$$

In 4 weeks for Brand E we have $q = .584$ (Table A2) and so

$$\begin{aligned} b_R/b &= 1 + \frac{\ln(1+.584)}{\ln(.416)} \\ &= 1 - \frac{.4600}{.8771} = .48 \text{ or } 48\%. \end{aligned}$$

The LSD calculations for all the time-periods are summarised in Table A12.

Table A11. The NBD Estimates of the Incidence of Repeat-Buyers

(The values of a, k and b for each period are from Table A1)

Brand E	Periods of length (in weeks) 4	12	24
$1+a$	2.32	6.34	13.00
$1+2a$	3.64	11.68	25.00
$(1+a)^{-k}$	.9599	.9299	.9101
$(1+2a)^{-k}$	.9391	.9080	.8886
b_R	.0193	.048	.068
b_R/b	.48	.69	.75

Table A12. The LSD Estimates of the Proportion of Repeat-Buyers

Brand E	Periods of length (in weeks) 4	12	24
q	.584	.851	.928
$\ln(1+q)$	.460	.616	.656
$\ln(1-q)$	−.877	-1.904	-2.631
b_R/b	.48	.68	.75

Table A13. The "Approximate" Estimates of the Proportion of Repeat-Buyers

Brand E	Periods of length (in weeks) 4	12	24
w	1.6	3.0	4.9
$2(w-1)$	1.2	4.0	7.8
$2.3w-1$	2.7	5.9	10.3
b_R/b	.45	.68	.76

The approximate formula for the proportions of repeat-buyers in Table 4.6 and §8.6 was

$$\frac{b_R}{b} = \frac{2(w-1)}{2.3w-1},$$

where w is the only input required. For Brand E in the average 4-week period,

$$w = 1.6$$

$$\frac{b_R}{b} = \frac{2 \times .6}{2.3 \times 1.6-1} \frac{1.2}{2.68}.$$

$$= .45 \text{ or } 45\%.$$

This and the approximate calculations for the other periods are set out in Table A13.

The observed values of b_R are necessarily the same in both periods which are being analysed (they are the same people), but the observed b_R/b ratios can vary if b varies from the 1 st to the 2nd period if there is some non-stationarity. similarly, the theoretical values of b_R or b_R/b depend on whether the input is that observed in the first or in the second period (or an average of the two). This only matters slightly when dealing with cases of slight non-stationarity, but becomes crucial when using the theoretical norm to interpret major changes from one period to the other (as for example in §§ 2.4 and 6.2).

A.9. The Buying-Frequency per Repeat-Buyer (Table 3.7)

To illustrate the theoretical calculations of the average purchase frequency for repeat-buyers, we take the average 12-week period in Table 3.7.

The NBD formula for m_R, the number of purchases made by repeat-buyers but expressed on a *per informant* basis, gives

$$m_R = m\,\{1-(1+m/k)^{-k-1}\}$$

$$= m\,\{1-(1+a)^{-k-1}\}.$$

Table A14. The NBD Estimates of m_R or w_R

Brand E	Periods of length (in weeks) 4	12	24
m	.064	.21	.44
a	1.32	5.34	12.00
k	.0485	.0393	.0367
$(1+a)^{-k-1}$	.4143	.1466	.0700
m_R	.0375	.1790	.4092
b_R(T.A11)	.0193	.048	.068
w_R	1.9	3.7	6.0

Table **A15.** The **LSD** and "Approximate" Estimates of w_R

Brand E	Periods of length (in weeks) 4	12	24
q	.584	.851	.928
$\ln(1-q^2)$	-.417.341	-1.287.724	-1.973.861
$(1-q)\ln(1-q^2)$	−.173	−.192	−.142
w_R	2.0	3.8	6.1
w	1.6	3.0	4.9
1.23w	2.0	3.7	6.0

If w_R, the rate of repeat-buying *per repeat-buyer,* is wanted, this is easily obtainable from the relationship

$$w_R = m_R/b_R ,$$

the theoretical value of b_R being available from Table A. 11.

In 12 weeks we have $m = .21$, $a = 5.34$, $k = .0393$ from Table Al, so that

$$m_R = .21 \{(1 - (6.34)^{-1.0393})\} ,$$

$$= .21\ (1 - .1466) = 0.1790 ,$$

and

$$w_R = \frac{.179}{.048} = 3.7.$$

Table A 14 summarises these calculations for all the time-periods.

The corresponding LSD formula for w_R is

$$w_R = \frac{-q^2}{(1-q)\ln(1-q^2)},$$

so that in 12 weeks, where $q = .851$,

$$w_R = \frac{-.724}{.149(-1.287)} = \frac{.724}{.192}$$

$$= 3.8$$

Table A15 gives the calculations for 4-, 12- and 24-week repeat-buying. The approximate formula for w_R which can be derived from the LSD is

$$w_R \doteqdot 1.23w,$$

and the results are also set out in Table A15.

Unlike for b_R, the observed values of w_R can differ from the first to the second time-periods because of any non-stationarity (even if only slight), and the theoretical values similarly depend on which period's *observed* values are used as input.

A. 10. The Buying-Frequency per "New" Buyer (Table 3.8)

To illustrate the calculations for the theoretical frequency of purchase of "new" or of "lapsed" buyers, we take the two 24-week periods.

The new buyers' purchasing frequency is given on a per informant basis in the NBD theory, as

$$m_N = \frac{m}{(1+a)^{k+1}}.$$

In 24 weeks, $m = .44$, $a = 12.00$, $k = .0367$, so that

$$m_N = \frac{.44}{(13.00)^{1.0367}} = \frac{.44}{14.280} = .0308.$$

The buying-frequency per "new" buyer is given by

$$w_N = m_N / b_N \ ,$$

where the proportion of "new" buyers b_N is given by

$$b_N = b - b_R \ .$$

The total penetration b of Brand E in the 24-week periods being .09 (Table 3.2a) and the theoretical NBD value for b_R at .068 being available from Table Al 1, we have

$$b_N = .09 - .068 = .022 \ ,$$

so that

$$w_N = \frac{.0308}{.022} = 1.4 \ .$$

Table A16. The NBD Estimates for m_N and w_N

Brand E	Periods of length (in weeks) 4	12	24
a	.064	.21	.44
k	.0485 1.32	.0393 5.34	.0367 12.00
$(1+a)^{1+k}$	2.417	6.818	14.280
m_N	.0265	.0308	.0308
$b_N = b - b_R$	.021	.022	.022
w_N	1.3	1.4	1.4

Table A17. The LSD and Approximate Estimates for w_N

Brand E	Periods of length (in weeks) 4	12	24
q	.584	.851	.928
$\ln(1+q)$	.460	.616	.656
w_N	1.3	1.4	1.4
Approximate Formula	1.4	1.4	1.4

This and the calculations for the average 4- and 12-week periods are given in Table Al 6.

The LSD formula for w_N is

$$w_N = \frac{q}{\ln(1+q)},$$

and in 24 weeks, $q = .928$, so that

$$w_N = \frac{,928}{\ln(1.928)} = \frac{.928}{.656}$$

$$\doteqdot 1.4.$$

Table Al 7 shows the calculation of w_N for the average 4- and 12-week periods as well as for the 24-week period, together with the values of the "approximate" formulae, which is simply

$$w_N \doteqdot 1.4.$$

The latter generally holds for $w > 2$ and values of b not too high (see §8.6), and here gives a fractionally different result from the LSD and NBD in the 4-week period, where $w < 2$.

The corresponding buying rates for the "lapsed" buyers (who buy in the first but not the second period) are equal to the above, subject to the effects of any non-stationarity, as already discussed in the previous section.

A. 11. Repeat-Buying by Light and Heavy Buyers (Table 3.10)*

In the case of the "Conditional Trend" analysis in Tables 3.10 and 3.1 Oa, only the NBD calculations (§7.6) need to be considered, since the LSD theory does not lead to any effective simplification here.

Tables 3.10 and 3.1 Oa referred specifically to the repeat-buying in Quarters II and III, so that the NBD parameters for Quarter II are required. Their method of determination is that illustrated in §A3 at the

* The theoretical calculations required for Table A9 are not discussed here as they are like those already given in §A7 for Table 3.6a, but as usual, care has to be taken over using the appropriate base-period.

beginning of the Appendix, and here only the numerical results are quoted. For Quarter II, $b = .062$ and w = 3.2 *, and this leads to $m = .198$, $a = 6.037$, and $k = .0361$.

The calculations for Tables 3.10 and 3.1 Oa use the conditional formula referred to in §7.6 of Chapter 7. This gives $p_{s/r}$, i.e. the proportion of those buying r times in the first period who buy s times in the second period,

$$p_{s/r} = (1+a')^{-k'} \frac{\Gamma(k'+s)}{\Gamma(k')\Gamma(s+1)} \left(\frac{a'}{1+a'}\right)^s ,$$

where

$$k' = k+r \text{ and } a' = a/(1+a) .$$

For Table 3.10, we only require the proportion $p_{./r}$ who buy *at* ***all*** in the second period of those who made r purchases in the first period. This is 1 minus $p_{0/r}$, the proportion *not* buying in the second period, i.e.

$$\begin{aligned} p_{./r} &= 1 - p_{0/r} \\ &= 1 - (1+a')^{-k'} \\ &= 1 - (1+a')^{-(k+r)} , \end{aligned}$$

from the above expression for $p_{s/r}$, putting $s = 0$. This is easily evaluated numerically once $(1+a')^{-1}$ and $(1+a')^{-k}$ have been determined. Thus

$$a = 6.037,$$

$$a' = \frac{6.037}{1+6.037} = .8579,$$

$$(1+a')^{-1} = .5382,$$

$$\begin{aligned} (1+a')^{-k} &= (1.8579)^{-.0361} \\ &= .9781, \end{aligned}$$

* **These** values are given to ***two*** significant figures (compared with one in Tables 3.1 and **3.2**), as the detailed calculations here are susceptible to rounding errors.

Table **A18.** NBD Estimates of Repeat-Buying in the Second Period

Brand E	Number of Purchases in 1st Period $r=0$	$r=1$	$r=2+$	Total
Buyers in 2nd Period				
As % of 1st period buyers $p_{./r}$	2.2	47	86	-
As % of total sample *	1.9	1.3	3.0 ***	6.2
Av. Purchases in 2nd Period				
Per buyer of *r in* 1st period, $m_{./r}$	.031	.889	4.04	-
Per repeat-buyer, $w_{./r}$	1.41	1.88	4.70	-
Per informant	.0268 **	.0237 **	.1475 **	.198

* Prop. of 1st Period Buyers **times** $p_{./r}$.
** Numerically the same as the proportion of sample buying $(r+1)$ times in the 1 **st** period (Table **A19).**
*** By subtraction.

and from this we get $p_{./r}$ either directly as $1-(1+a')^{-k}\{(1+a')^{-1}\}^r$, or or by the recurrence formulae $(1-p_{./r})=(1-p_{./r-1})(1+a')^{-1}$.

For Table 3.1 Oa we need $m_{./r}$, the mean of the theoretical distribution of the number of purchases in the second period conditional upon having made r purchases in the first. This is given by

$$m_{./r}=a'(k+r).$$

With these general expressions for $p_{./r}$ and $m_{./r}$ for any r, we can now calculate the specific values for $r=0$, 1, 2, etc.

Starting with non-buyers in the first period, i.e. $r=0$, we have

$$\begin{aligned} p_{./0} &= 1-(1+a')^{-k} \\ &= 1-.9781 \\ &= .0219 \text{ or } 2.2\%, \end{aligned}$$

and

$$\begin{aligned} m_{./0} &= a'k \\ &= .8579\times .0361 = .0310. \end{aligned}$$

Hence the buying frequency,

$$w_{./0} = \frac{m_{./0}}{b_{./0}} = \frac{.0310}{.0219} = 1.41.$$

(This is simply w_N, the buying frequency per "new buyer", as in Table 3.8.) These values are set out in Table A 18 for $r = 0$ *.

Next, for once-only buyers in the first period, i.e. $r = 1$, we have

$$\begin{aligned} p_{./1} &= 1 - (1+a')^{-k-1} \\ &= 1 - (1+a')^{-k}(1+a')^{-1} \\ &= 1 - .9781 \times .5382 \\ &= 1 - .5264 \\ &= .474 \text{ or } 47\%, \end{aligned}$$

and

$$\begin{aligned} m_{./1} &= a'(k+1) \\ &= .8579 \times 1.0361 = 0.889. \end{aligned}$$

Hence

$$w_{./1} = \frac{m_{./1}}{b_{./1}} = \frac{.889}{.474} = 1.88.$$

These values are set out in the $r = 1$ column in Table A 18.

And so one can proceed for $r = 2$, $r = 3$, etc. In practice one will stop at some "maximum" value of r. Beyond this we then require the repeat-buying estimates for all buyers who bought more often than this maximum in the first period. These theoretical values of the number of repeat-buyers amongst the "more than r" buyers, and their average purchase frequency in the second period, have to be found by subtracting the appropriate values for $r = 0$, for $r = 1$, for $r = 2$, etc. from the *total* number of buyers and the *total* number of purchases predicted for the second period (expressing all values on a "per informant" basis so as to avoid having explicitly to introduce the sample size). This can

* The derivation of the number of repeat-buyers as a percentage of the total sample (i.e. 1.9% for $r = 0$ in Table A18) is given later in this section.

be illustrated here in terms of the "more than once only" buyers in the first period.

We therefore need to express the repeat-buying values $p_{./r}$ on a "per informant in the total sample" basis. The proportion of informants buying in the first period was .062 (see above) and so the proportion *not* buying is .938. Under the NBD, we have worked out above that 2.2% of these non-buyers (r = 0) in the first period buy in the second period (see $p_{./0}$ above). Therefore the theoretical value for the percentage of *all* informants of those who, not having bought in the first period, buy in the second is

$$.02 \times .938 = .019 \text{ or } 1.9\%.$$

Further, they buy at a rate of 1.41 ($w_{./0}$ above), giving a theoretical total of

$$(1.41 \times 1.9)/100 = .0268$$

purchases per informant, as is shown in Table Al 8.

For r = 1, the percentage of households who bought once only in the first period is first of all required. This can of course be calculated directly from the p_r formula in § A6 with r = 1, but the calculation is simplified for the present purpose by a certain feature of the NBD, namely that the number of purchase occasions in the second period made by those households who bought r times in the first period is numerically the same as the sheer *number* of households who bought (r+ 1) times in the first period! It follows that the proportion of households who bought only once in the first period is .0268 (i.e. the numerical value calculated – for a quite different purpose – in the preceding paragraph). This is set out in Table A 19.

Table A19. The NBD Estimates of p_r in the First Period

Brand E	r, the Number of Purchases			Total
	0	1	2+	
NBD Proportions p_r	.9380	.0268	.0352 *	1.000

* By subtraction from 1.

Next, of this proportion .0268 of once-only buyers in the first period, 47% were expected to buy in the second period (the $p_{./1}$ value calculated above). The number of these repeat-buyers on a per informant basis is therefore

$$.0268 \times 47 = .0126 \text{ or } 1.3\%.$$

These particular repeat-buyers are expected to buy at a rate of 1.88 ($w_{./1}$ above) and therefore account for

$$1.88 \times .0126 = .0237 \text{ purchases on a per informant basis.}$$

We can now calculate the expected proportion of buyers in the second period who bought *more than once* in the first period, namely by subtraction,

$$6.2 - 1.9 - 1.3 = 3.0\%,$$

where 6.2% is of course the total number of buyers in the second period (equal to the number of buyers in the first period). The number of purchases they are expected to make is, similarly,

$$.198 - .0268 - .0237 = .1475 \text{ per informant,}$$

where .198 is the total number of purchases in each period per informant (as given at the beginning of this section).

By subtraction we also find that the proportion of informants who bought more than once in the first period is

$$1.000 - .938 - .0268 = .0352 \text{ or } 3.5\%,$$

as shown in Table Al 9.

Therefore the 3.0% who are expected to buy again in the second period represent

$$\frac{3.0}{3.5} = 86\%$$

of the initial more-than-once buyers. The rate at which they are expected to buy in the second period is

$$\frac{.1475 \times 100}{3.0} = 4.9 .$$

These results are shown in the 1+ column of Table A 18.

A. 12. Tabulating Sales, Penetration, and Purchase Frequency

Condensing the raw data from a consumer panel for a single brand into the summary statistics like m, b, and w discussed so far is relatively simple because we need only consider the data for that single brand. It involves identifying the panel member, the brand bought on each purchase occasion and the point in time.

As discussed in § 1.4, all the analyses are in terms of purchase occasions, not amounts bought or paid. (The definition of a purchase occasion may raise problems in some data, where two separate purchases of the same brand made at the same retail outlet in a week may not be distinguishable from two packs of the brand being bought on the same occasion.)

The only step needed beyond simple counts concerns the penetration b of a brand in a given time period; for this, we only count the first time a panel-member buys the brand in the period. In the tabulations we therefore need to allow for whether the buyer has bought that brand before in that time-period.

Data records can be provided in one of several different ways, e.g.

(a) Ordered by panel member. Here all purchases for an individual consumer or household are shown consecutively for a time-period. This time-period (e.g. a year) may be longer than the period we intend to study (e.g. a particular 4 weeks).

(b) Ordered by week. All purchases by different consumers in Week 1 are arranged in some order, then all purchases in Week 2, and so on.

(c) No apparent order.

Especially for hand tabulations it is usually easiest to aim at layout (a) and then order by (i) brand and by (ii) time sequence within brand for each individual. This simplifies the tabulation of penetrations.

To illustrate with a very small-scale hypothetical example, Table A20 sets out the purchases of three brands *X*, Y and *Z* over 12 weeks. The total sample is 200 households, of whom 180 did not buy the product-category in that 12-week quarter.* Suppose we want to tabulate summary statistics

*The example will also be used in the new Appendix C. It covers multi-brand buying and therefore is more elaborate than the numerical example used earlier in this Appendix.

for Brand X. We consider statistics mainly for the first 4 week period (similar tabulations will be needed later for the other two months and for the whole quarter), namely:

- Sales, or average purchase occasions for Brand X in the period per individual, denoted *m* as in the algebraic notation of the main text.
- Penetration, or % buying X at least once in the period, *b*.
- The frequency distribution of the number of purchases, p_r.
- The average number of purchases per buyer, w = *m*/*b*.

In doing the counts by hand, we can use five-barred "gates" (e.g. ~~||||~~), as shown on the right of Table A20. To count "buying at least once" we only note the first purchase by that individual in the time-period.

The tabulation shows that in Weeks 1-4, 6 buyers bought on 8 purchase occasions (with the tally of gate counts and "totals" both summing to 8). Hence with a sample of 200,

$$\begin{aligned} m &= 8/200 = .04, \\ b &= 6/200 = .03 \text{ or } 3\%, \\ w &= 8/6 = 1.33. \end{aligned}$$

The frequency distribution of purchases, counting the number of Os, ls, 2s, etc out of 200 in the total column, gives

Number of Purchase Occasions in Weeks 1-4

0	1	2	3	4+
194	5	–	1	–

The figures add up to 200 and give 5 x 1 + 1 x 3 = 8 purchase occasions, calculations which are essential checks on the arithmetic.

The answers for all three 4-week periods and the total 12 weeks from Table A20 should come out to be

For Brand X	1st 4 weeks	2nd 4 weeks	3rd 4 weeks	12 weeks
Total Purchase Occasions	8	9	8	25
Total Buying at least once	6	7	7	13
m	,040	.045	.040	.125
b	.030	.035	.035	.065
w	1.33	1.28	1.14	1.92

Table A20. A Simplified Buying Pattern: Three Brands over 12 Weeks
(Tabulating X in weeks 1 to 4)

Panel-Member	Purchases in Week: 1	2	3	4	5	6	7	8	9	10	11	12	Brand X in Weeks 1–4: Gate Count	Total	Buying at least once
1	X	X	Y	X	X	X	Y	X	.	Y	X	Y	///	3	1
2	.	.	X	.	Z	X	.	.	.	Z	X	X	/	1	1
3		.	Z	.	.	Z	.	X	X	.	Z	.		-	-
4	.	.	.	.	Y	.	Y	Y	.	X	.	.		-	-
5	.	X	.	.	X	.	.	.	.	.	.	Y	/	1	1
6	Y	.	.	.	.	.	.	Z	Z.	Y	.	.	/	1	1
7	X	Z	X	.	.	.	.	.	.	.	.	X	/	1	1
8	.	Y	.	Y	.	.	.	.	X	.	.	.		-	-
9	.	.	.	.	.	.	X	.	.	.	Y	.		-	-
10	Z	.	.	X	.	Y	.	.	.	.	.	.	/	1	1
11	.	.	.	Y	.	.	X	.	.	.	.	.		-	-
12	.	.	.	.	.	.	.	.	.	X	.	.		-	-
13	Y	.	.	.	.	.	.	.	.	.	.	.		-	-
14	.	.	.	.	.	.	.	Y	.	.	.	.		-	-
15	.	.	Z	.	.	.	.	.	.	.	.	.		-	-
16	.	.	.	.	.	.	.	.	Y	.	.	.		-	-
17	.	.	.	.	Z	.	.	.	.	.	.	.		-	-
18	.	.	.	.	.	.	.	.	.	.	Z	.		-	-
19	.	.	.	.	.	.	.	.	.	Z	.	.		-	-
20	.	.	.	.	.	X	.	.	.	.	.	.		-	-
21–200	.	.	.	.	.	.	.	.	.	.	.	.		-	-
Totals	2	2	2	2									8	8	6

It is easy to make mistakes and so it is once more essential to do the following checks:

(a) The three 4-week totals should add up to the 12-week total of 25,
(b) The three 4-week *ms* should add up to the 12-week *m* of .125,
(c) The sum of the three 4-week *bs* (.1OO) should be greater (or possibly equal) to the 12-week *b* (.065).

If the data are almost stationary, then all the 4-week figures will be very similar. (The rather "low" w of 1.14 in the 3rd 4-week period is, given the sample size, in effect due to 1 purchase or 1 buyer. Thus 8/7 = 1.14, whereas 9/7 = 1.29 or 8/6 = 1.33, close to the figures for the previous two 4-week periods).

As a further example of the tabulations, the frequency distribution of purchases of X in the whole 12-week period, using gate-counts as an intermediate step, is

Number of Purchase Occasions in Weeks 1-12

0	1	2	3	4	5	6	7	8+
	𝍸 |||	|||		|			|	
187	8	3	–	1	–	--	1	

which correctly adds to 200 individuals again and to 25 purchases

(i.e. 8 x 1 + 3 x 2 + 1 *x* 4 + 1 x 7 = 25).

A. 13. **Tabulating Period-by-Period Repeat-Buying**

To illustrate the tabulation of the repeat-buying statistics from one period to another equal-length period, we consider weeks 1-4 and 5-8 from Table A20. Table A21 shows one possible way of arranging the hand-tabulations. (More elaborate versions are used later.) The table gives the counts that are required for each period, care being needed again in tabulating the penetrations (i.e. counting those who buy at least once in *both* periods, or in only one or the other).

Table **A21.** Repeat Buying of Brand X, Weeks 1-4 and 5-8

	Purchase Occasions			
	Counts		Totals	
Panel- member	1-4	5-8	**1-4**	5-8
1	///	///	(3	3)
2		I	(1	1)
3			–	1
4			–	–
5			(1	1)
6	/		1	–
7	/		1	–
8			–	–
9		I		1
10	1		1	–
11		/	–	1
12			–	–
13				
14				
15			–	–
16				
17				
18			–	–
19			–	–
20		/	–	1
Number of Purchases				
All			8	9
by Repeat Buyers			5	5
by Lapsed Buyers			3	–
by New Buyers				4
Number of Buyers				
All			6	7
Repeat Buyers			3	3
Lapsed Buyers			3	–
New Buyers				4

The first two columns give the five-barred-gate counts done by hand (which are more helpful than here when one is dealing with a larger number of purchases in longer time-periods). They are to be filled in one row and column at a time (i.e. period-by-period for each panel-member) from Table A20. We then form the totals for each period, as shown in type-set form in the last two columns, and visually differentiate repeat-buyers who bought in each period, by encircling them by hand. We can now form various totals for the total panel as follows, first for the number of purchases and then for the number of buyers:

Purchases

(a) Period totals, i.e. total purchases in Weeks 1-4 (8) and Weeks 5-8 (9).

(b) By repeat-buyers, i.e. purchases in each period by the (encircled) buyers who bought in both (the two sets of numbers happen to be equal, giving a total of 5 in each period).

(c) By lapsed buyers, i.e. (non-encircled) purchases in the first period (which is 3 purchases by those who bought in that period only).

(d) By new buyers, i.e. (non-encircled) purchases in the second period.

Buyers

In the same way we can count the number of buyers (i.e. their sheer occurrence, without taking account of how often they buy), for

(a) Each period (6 and 7),
(b) Repeat-buyers (3 and 3),
(c) Lapsed buyers (3),
(d) New buyers (4).

It is unavoidable to run down each column separately six times to perform these counts. (The first time round this is usually quicker than getting a computer program to work!)

Various checks on the arithmetic can be done. Thus

— The total number of purchases in each period should agree with those recorded in Table A20 (i.e. 8 and 9).
— The two entries in (a) for Weeks 1-4 and 5-8 must each equal the sums of (b) and (c) or of (b) and (d).

— The numbers of repeat-buyers in each of the two periods must be equal, whereas how often they bought need not be equal.

The repeat-buying summary statistics can now be calculated with the sample-size $n = 200$.

	Purchase Occasions			Buying at Least Once	
Weeks	1-4	5-8	Weeks	1-4	5-8
m	.040	.045	b	.030	0.35
w	1.33	1.29	–	–	–
m_R	.025	.025	b_R	.015	,015
w_R	1.67	1.67	b_R/b	50%	43%
m_L	.015	–	b_L	.015	–
w_L	1.00	–	b_N	–	.020
m_N	–	.020	b_L/b	50%	–
w_N	–	1.00	b_N/b	–	57%

To check the arithmetic, we note that

(i) $m = m_R + m_L$ in Weeks 1-4, i.e. $.040 = .025 + .015$, and $m = m_R + m_N$ in Weeks 5-8.

(ii) $b = b_R + b_L$ in Weeks 1-4, i.e. $.030 = .015 + .015$, and $= b_R + b_N$ in Weeks 5-8.

(iii) We also have $b_R/b + b_L/b = 100$ and $b_R/b + b_N/b = 100$ in each period.

Extensions of these tabulations, such as separate "conditional" ones for Light, Medium or Heavy buyers (defined as desired) are straightforward in principle. In practice, it is best to mark the panel-members first (e.g. by *L, M* and *H*) using their purchases in period 1, say, as the criterion, and then run three separate sets of counts for the summary statistics at the bottom of the table. (As a check, the three sets of figures should add up to the relevant totals of the foot of Table **A21**). (It may reduce counting errors to make several Xerox copies of Table A20 and block out Light, Medium or Heavy buyers separately on each sheet.)

A. 14. **Continuous Reporters**

The tabulations in A.12 and A.13 help to emphasise the need for basing the analysis on "continuous reporters", sometimes called a "static sample'', over the period analysed. In a diary panel, for example, it is misleading if a panel-member is recorded as a non-buyer of brand **X** in some week or weeks when in fact he or she did not return a diary that week.

Continuous reporting can be difficult to ascertain because in some panel operations no count is kept of whether a usable diary is returned by each panel-member every week, or at least such data are not readily accessible. A partial (but time-consuming) solution is to form a panel of continuous reporters by checking whether each individual recorded any purchases in ***any*** of the product-categories measured by the panel.

A definition of a "continuous reporter" is however somewhat fuzzy. Absence on holiday should presumably be counted as "non-purchasing but reporting'', if the aim is to cover "purchases at home". But absence (or illness) of the normal diary keeper only (with other household members still buying and consuming at home) necessarily introduces "errors" in the data, hopefully minor ones.

In general, for the type of studies discussed in this book it seems better to have a relatively small sample of "more rather than less" continuous reporters, even if this is potentially biased in other statistical respects, e.g. by age of or even by their amount of purchasing. Repeat-buying patterns for a slightly biased sample are likely to be more valid than ones that are wrong because of discontinuous reporting, even though the sample itself may be more directly representative of the population at large.