

1.1.3 Introduction to the use of computers in survey analysis[1]

[Updated 23 August 2010 from 2006, 1992 and 1973 versions, but still with (historically interesting) output from pre-Windows versions of SPSS.]

Survey researchers daily need to solve problems caused by human error. Interviewers may ask the wrong questions, or mark the wrong answers: respondents either mishear or misinterpret questions they are asked. Hence a great deal of attention should be paid, not only to interviewer training and fieldwork checks, but also to questionnaire wording and design, in order to minimize error from these sources.

When completed schedules are returned, coders may assign wrong codes to answers, and keypunchers may punch cards[2] with wrong numbers in the right columns or right numbers in the wrong columns: hence large proportions of coding should be double-checked, and all card punching verified, on the theory that the chances of making the same mistake twice are quite low.

However, even with quality control operating at its most efficient, errors may still appear in the raw data, and these need to be detected and corrected before analysis can begin. It is at this stage that the computer becomes useful as a research tool.

It is sound practice in all survey research to allocate to each survey a unique name or number which is included in the data for each interview. Each interview should also be given a unique number to distinguish it from other interviews in the same survey, and if the data from the interviews are punched on more than one card then each card should also be numbered. If cards are in free format, or if paper tape is used, then a "rogue value" should be used to mark the end of each interview. These simple rules should enable the researcher to check by computer that each interview has the correct number of codes, and that, if cards are used, they are all present (ie no missing cards) and in the correct sequence, and that there no duplicate cases or cards within cases.

Wherever the necessary equipment and programs are available, it is both easier and safer to transfer all the punched data from cards or paper tape direct to magnetic tape or disk and to use the computer for all subsequent processing. The computer can also be used to provide the researcher with a copy of the raw data, and a copy of the latest edition of magnetic tapes - just in case. It is strongly recommended to keep at least one copy, and preferably two, of each edition of a data set. Sometimes data are entered into the computer in batches as questionnaires are coded, in which case they may be out of serial number sequence: at other times data are entered after all questionnaires have been coded and in serial number order.

---

1 **Author's note:** This short paper was first written in 1973 for trainee researchers and for clients of the then SSRC Survey Unit, to help explain data processing and tabulation to people with little or no experience of survey research or computing. The original text was later retyped into the Vax computer at PNL and the output tables are from various versions of SPSS current on the CDC2000 at ULCC or the Dec-10/Dec 20/Vax cluster at PNL at the time. The whole document was later then downloaded into WordStar4. It has now been converted and edited for MSWord. There have been many developments in the 37 years since it was written, particularly in personal computers, computer assisted interviews, on-line surveys and the Windows version of SPSS, but I have left the original text rather than update the whole document.

The data used in figs 2 to 10 inclusive are all from the **Quality of Life in Britain** surveys conducted at the SSRC Survey Unit by the late Dr Mark Abrams and myself between 1971 and 1975. Data from the surveys are deposited with the UK Data Archive at Essex University and are available as SPSS portable files for secondary analysis, together with user documentation in pdf format. For a full description of survey content and of technical information on material deposited at Essex click here (some is also available on this site).

2 In 1973 data prep and analysis specs were done using 80-column Hollerith cards, but at least results came off a line-printer. When the author started doing survey research in 1965 it was all done using paper tape, including the results!

An example of data out of sequence is given in Figure 1a below.

Figure 1a: Unsorted data file (with card layout template)

```
0        1         2         3         4         5         6         7         8
12345678901234567890123456789012345678901234567890123456789012345678901234567890

0119 11190580 8341D           048001112814S         1 13231                   HQ
0451 11190580 8411C           048001222816Q         1 22031                   HQ
0148 15160580 8153H           107907122844S         1 12191                   HQ
1234 12200580 8428BFI       20058000                  22431                   HQ
0149 13210580 8411C         19058001111811E           12813                   HQ
0122 12200580 8427C         18067701  1842B           12512                   HQ
1236 13210580 8215H           0879011338 0NP          22713 13815             HQ
1235 13210580 8481CJK       06058001111843CEI         22113                   HQ
1237 13210580 8145H         17077901111814H           23013                   HQ
1205 13220580 8147H           0474 311294XN           22315                   HQ
1204 13220580 8451CJK       07058001111811EJ          13114                   HQ
1203 13210580 811YEG              00111             1 14132                   HQ
1202 13210580 8164KS          0479 1122815NS          22513 1281421           HQ
0123 15230580 8118BE              00111               21812                   HQ
1127 15230580 8327D           06780111191XTW          22212112491             HQ
1128 15300580 8413CJK         057805111841CE          22331                   HQ
1179 15300580 8 Y1M         260580 4111841CEJ         12331 21912             HQ
0123 15230580 8418BE              00111               21812                   HQ
1176 15230580 8414BJ          0980 1111816K           21913                   HQ
0124 15230580 8423CJ          127804111842CJ          12014                   HQ
1238 12270580 8158H               00111               16021                   HQ
1241 12270580 8144B         29127901444840N           118133217  21           HQ
1177 12270580 8326DES               112               13314 2238              HQ
0125 12270580 8214H             6525122810N           14413124122             HQ
0050 12270580 8321DJK         057903111811H         1 23031112513             HQ
1242 12270580 8161K         20058001111812F           1273                    HQ
1240 12270580 8211H         20048001111832D         1 22031                   HQ
1239 12270580 8171K         25058001111814H           1535                    HQ
1243 12270580 8423B           127901111815E           22314                   HQ
1178 13280580 8314DJK         107903111834H           22131 12612             HQ
1001 13280580 8235H           0279  113815N           22413112381             HQ
1126 13280580 8166Q         22097503211841AF          21812                   HQ
1152 13280580 8162K         09048001111840C           11721                   HQ
0536 15160380 8124H         15017902111814H           13415 2272211           HQ
0223 11190580 8146H           087801111814H           14015                   HQ
0537 11190580 8247B                                   21612                   HQ
```

In the above example the first four columns contain the serial number from the original questionnaire and the last two columns contain a code for which survey they came from. The blank columns are deliberate as they help to give a visual clue as to whether data have been entered in the correct columns.

The computer can also be used to sort the data so that all the cases and all the cards (records) within cases are in the right order. An example of sorted data is given in Figure 1b below.

Figure 1b: Sorted data file (same data as Figure 1a)

```
0        1         2         3         4         5         6         7         8
12345678901234567890123456789012345678901234567890123456789012345678901234567890

0050 12270580 8321DJK         057903111811H         1 23031112513             HQ
0119 11190580 8341D           048001112814S         1 13231                   HQ
0122 12200580 8427C         18067701  1842B           12512                   HQ
0123 15230580 8118BE              00111               21812                   HQ
0123 15230580 8418BE              00111               21812                   HQ
0124 15230580 8423CJ          127804111842CJ          12014                   HQ
0125 12270580 8214H             6525122810N           14413124122             HQ
0148 15160580 8153H           107907122844S         1 12191                   HQ
0149 13210580 8411C         19058001111811E           12813                   HQ
0223 11190580 8146H           087801111814H           14015                   HQ
0451 11190580 8411C           048001222816Q         1 22031                   HQ
0536 15160380 8124H         15017902111814H           13415 2272211           HQ
0537 11190580 8247B                                   21612                   HQ
1001 13280580 8235H           0279  113815N           22413112381             HQ
1126 13280580 8166Q         22097503211841AF          21812                   HQ
1127 15230580 8327D           06780111191XTW          22212112491             HQ
1128 15300580 8413CJK         057805111841CE          22331                   HQ
1152 13280580 8162K         09048001111840C           11721                   HQ
1176 15230580 8414BJ          0980 1111816K           21913                   HQ
1177 12270580 8326DES               112               13314 2238              HQ
1178 13280580 8314DJK         107903111834H           22131 12612             HQ
1179 15300580 8 Y1M         260580 4111841CEJ         12331 21912             HQ
1202 13210580 8164KS          0479 1122815NS          22513 1281421           HQ
1203 13210580 811YEG              00111             1 14132                   HQ
1204 13220580 8451CJK       07058001111811EJ          13114                   HQ
1205 13220580 8147H           0474 311294XN           22315                   HQ
1234 12200580 8428BFI       20058000                  22431                   HQ
1235 13210580 8481CJK       06058001111843CEI         22113                   HQ
1236 13210580 8215H           0879011338 0NP          22713 13815             HQ
1237 13210580 8145H         17077901111814H           23013                   HQ
1238 12270580 8158H               00111               16021                   HQ
1239 12270580 8171K         25058001111814H           1535                    HQ
1240 12270580 8211H         20048001111832D         1 22031                   HQ
1241 12270580 8144B         29127901444840N           118133217  21           HQ
1242 12270580 8161K         20058001111812F           1273                    HQ
1243 12270580 8423B           127901111815E           22314                   HQ
```

In the above example the data have now been sorted into order by serial number. This makes it much easier to check for missing or duplicate serial numbers, to correct data errors and also to check against the original questionnaires (which should also be kept in serial number order).

Next the computer is used to check that all codes punched fall within the range allocated for each item. For instance the replies to a question "How do you normally travel to work?" may be coded "Walk" = 1, "Bus" = 2, "Train" = 3, "Car" = 4, "Other replies" = 5. The computer is used to make sure that all codes for that item are within the range 1 - 5 and to print out the serial numbers of questionnaires for which data fall outside the range. A typical error message[3] might read:

CASE 329 CARD 1 COLUMN 24 NOT IN RANGE. SHOULD BE 1 TO 5 IS 7.

The computer will not detect coding errors such as people coded as "Walk", but whose actual reply was "Bus", which should have been detected by the interviewer or by the coder. Sometimes these errors can be detected by the second type of data check.

The computer can now be used to carry out logical checks[4] on the coded responses. Typical examples of this are checks carried out to see that single 16 year-old girls are not coded as having been married for 20 years, or that a head of household coded as AB is not classed somewhere else as DE, or that a woman coded as full-time housewife has a personal income from a paid job in excess of £5000 p.a. In the journey to work example above, a person coded as "Walk", but with a travel-to-work cost per week greater than zero, would require further inspection of the original completed schedule.

If the data set is not too large, it is very easy to spot certain kinds of errors simply by printing out the contents of the data file on the line-printer or listing them on a computer screen. Listing them in case-order card-within-case will help to check that all cards are present for each case. A line editor with a facility for jumping data lines in multiples of the number of cards per case is particularly helpful in this respect as the card numbers should always be the same. Listing each card separately (e.g. all card one's followed by all card two's etc.) helps to spot entries in columns which are supposed to be blank, or blanks where there should be entries. It is particularly useful to leave blank columns deliberately in fixed locations as these will show up as vertical straight lines. (See figs 1a and 1b)

[3] This example was produced by Survey Data Tabulation (SDTAB), a program written by Peter Wakeford, then Director of Computer Services at LSE

[4] Modern computer-assisted personal interview software (CAPI) can be programmed to pick such errors up during the interview itself.

Once the data have been edited and cleaned, the researcher needs preliminary results in convenient form. One of the first things that is usually done is a simple holecount for each column, to reveal the distributions of responses.

Figure 2: Typical commercial holecount[5] (including multipunches)

	1	2	3	4	5	6	7	8	9	0	X	Y	REJ	SUM	CRDS
1	67	0	0	0	0	0	0	0	0	88	0	0	0	155	155
	43.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	56.8	0.0	0.0	0.0	100.0	
2	20	20	20	20	20	16	10	9	10	10	0	0	0	155	155
	12.9	12.9	12.9	12.9	12.9	10.3	6.5	5.8	6.5	6.5	0.0	0.0	0.0	100.0	
3	30	0	0	0	0	0	0	0	0	125	0	0	0	155	155
	19.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	80.6	0.0	0.0	0.0	100.0	
4	16	16	16	16	15	16	15	15	14	16	0	0	0	155	155
	10.3	10.3	10.3	10.3	9.7	10.3	9.7	9.7	9.0	10.3	0.0	0.0	0.0	100.0	
5	5	4	10	0	0	0	0	0	0	16	0	4	121	39	155
	3.2	2.6	6.5	0.0	0.0	0.0	0.0	0.0	0.0	10.3	0.0	2.6	78.1	25.2	
6	5	2	2	12	2	2	2	5	2	2	0	0	122	36	155
	3.2	1.3	1.3	7.7	1.3	1.3	1.3	3.2	1.3	1.3	0.0	0.0	78.7	23.2	
7	0	0	155	0	0	0	0	0	0	0	0	0	0	155	155
	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	
8	155	0	0	0	0	0	0	0	0	0	0	0	0	155	155
	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	
9	0	0	0	0	0	0	0	0	155	0	0	0	0	155	155
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	
10	155	0	0	0	0	0	0	0	0	0	0	0	0	155	155
	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	
11	155	0	0	0	0	0	0	0	0	0	0	0	0	155	155
	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	
12	133	0	0	0	0	0	0	0	0	17	0	0	5	150	155
	85.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.0	0.0	0.0	3.2	96.8	
13	34	9	20	27	15	5	0	0	15	29	0	0	1	154	155
	21.9	5.8	12.9	17.4	9.7	3.2	0.0	0.0	9.7	18.7	0.0	0.0	0.6	99.4	
14	29	19	31	20	18	0	0	0	0	37	0	0	1	154	155
	18.7	12.3	20.0	12.9	11.6	0.0	0.0	0.0	0.0	23.9	0.0	0.0	0.6	99.4	
15	0	1	3	0	57	0	1	0	0	92	0	0	1	154	155
	0.0	0.6	1.9	0.0	36.8	0.0	0.6	0.0	0.0	59.4	0.0	0.0	0.6	99.4	
16	146	0	0	0	0	0	0	0	0	6	0	0	3	152	155
	94.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.9	0.0	0.0	1.9	98.1	

The headings are the possible hole-sites on each card-column, REJ means the column has no hole punched, SUM is the sum of all holes punched in that column for the whole sample and CRDS is the number of cases.

In the example above, for each column of the card, the upper row of figures gives the number of cases with each hole-site punched and the lower row the percentage of cases to one decimal place.

This analysis can only be done one column at a time by some computer programs, but it provides a quick visual check on distributions and can help to spot rogue data.

[5] This example is of output from Donovan Data Systems as used by Research Services Ltd

The next step is to run (unlabelled) frequency counts on some or all of the variables from the survey (source data: SSRC Survey Unit Quality of Life in Britain survey,1973)

FIG.3: FREQUENCY COUNT[6] WITHOUT LABELS

AGEGRP	Code	Absolute freq	Relative freq (%)	Adjusted freq (%)	Cumulative freq (%)
	1.	206	22.1	22.1	22.1
	2.	214	23.0	23.0	45.1
	3.	242	26.0	26.0	71.0
	4.	256	27.5	27.5	98.5
	99.	14	1.5	1.5	100.0
		------	------	-------	
	Total	932	100.0	100.0	

Valid cases 932 Missing cases 0

A more advanced presentation of frequency distributions for simple variables is to use titles and captions. Data can be presented as numbers or percentages in categories of the variable (marginal distribution) or pictorially in a bar-chart or histogram. For variables measured on ordinal, interval or ratio scales some statistical measures of location, dispersion and shape may be wanted, but all these measures need to take account of codes given for missing answers (Refused, Don't Know, Not applicable, etc).

FIG.4: FREQUENCY COUNT WITH LABELS

AGEGRP: Age group	Code	Absolute freq	Relative freq (%)	Adjusted freq (%)	Cum freq (%)
17-29	1.	206	22.1	22.4	22.4
30-44	2.	214	23.0	23.3	45.8
45-59	3.	242	26.0	26.4	72.1
60+	4.	256	27.5	27.9	100.0
	99.	14	1.5	Missing	100.0
		------	------	-------	
	Total	932	100.0	100.0	

Valid cases 918 Missing cases 14

These initial frequency distributions are useful to the researcher when deciding upon groupings or checking on the representativeness of the sample.

[6] This and all subsequent examples of tables are from earlier mainframe versions of SPSS at ULCC and PNL

Some computer programs can also provide graphical output which is often easier to understand than simply looking at sets of numbers. This is particularly useful with interval scaled data as the shape of the distribution can be seen at a glance.

FIG.5: HISTOGRAM[7] PLOT (WITH OPTIONAL STATISTICS)

```
VAR147    Q147 SATISFACTION WITH WHOLE LIFE


    Code
         I
       1 ** (     1)
         I
       2 ** (     1)
         I
       3 ** (     2)
         I
       4 ****** (     9)
         I
       5 ********** (    18)
         I
       6 ********** (    17)
         I
       7 ******************* (    35)
         I
       8 ******************************* (    60)
         I
       9 ****************** (    34)
         I
      10 ****************** (    33)
         I.........I.........I.........I.........I.........I
         0        20        40        60        80       100
         Frequency

Mean      7.610     Median     7.867     Std dev    1.801

Valid cases    210     Missing cases    0
```

[7] This was done many years before the graphics facilities were added to SPSS!

Some variables, e.g. age last birthday, may have been coded in single years across two card-columns, but the computer will produce tables of these quite easily.

Figure 6: Condensed[8] format frequency count with full range of statistics

AGE AGE OF R IN COMPLETE YEARS

Code	Freq	Adj %	Cum %	Code	Freq	Adj %	Cum %	Code	Freq	Adj %	Cum %
18	15	2	2	42	14	2	42	66	14	2	83
19	16	2	3	43	14	2	44	67	20	2	85
20	19	2	5	44	19	2	46	68	12	1	87
21	17	2	7	45	11	1	47	69	18	2	89
22	19	2	9	46	15	2	49	70	13	1	90
23	16	2	11	47	14	2	50	71	8	1	91
24	16	2	13	48	17	2	52	72	8	1	92
25	14	2	14	49	15	2	54	73	12	1	93
26	19	2	16	50	24	3	56	74	9	1	94
27	25	3	19	51	16	2	58	75	8	1	95
28	13	1	21	52	15	2	60	76	7	1	96
29	16	2	22	53	19	2	62	77	7	1	97
30	13	1	24	54	14	2	63	78	6	1	97
31	13	1	25	55	15	2	65	79	4	0	98
32	24	3	28	56	13	1	66	80	5	1	98
33	7	1	29	57	19	2	68	81	3	0	98
34	19	2	31	58	16	2	70	82	6	1	99
35	13	1	32	59	19	2	72	83	2	0	99
36	7	1	33	60	10	1	73	85	1	0	99
37	12	1	34	61	15	2	75	86	1	0	100
38	14	2	36	62	17	2	77	87	1	0	100
39	13	1	37	63	14	2	78	88	2	0	100
40	15	2	39	64	17	2	80	90	1	0	100
41	17	2	41	65	15	2	82				

M i s s i n g d a t a

Code	Freq	Code	Freq	Code	Freq
Wild	15				

Mean	47.033	Std err	0.586	Median	47.464
Mode	27.000	Std dev	17.733	Variance	314.466
Kurtosis	-1.054	Skewness	0.100	Range	72.000
Minimum	18.000	Maximum	90.000		

8 This table was produced with the command:

FREQUENCIES AGE /FORMAT CONDENSE /STATISTICS ALL

but the keyword **CONDENSE** is no longer available (Shame on you, SPSS!).

Once all groupings and recodings are complete, the computer can be used to produce initial cross-tabulations. Usually these tabulate the response to the substantive part of the questionnaire against standard demographic information (Sex, age group, social class, marital status, house-type etc). The simplest output of this kind gives a table title and a table with no headings other than the card and column numbers of the variables being tabulated, and no captions other than the code numbers. The table may contain raw counts, percentages, or both, depending on the options selected. The base for percentages may be row totals, column totals, or the global total for the table.

Figure 7: Contingency table without labels

Count Row %	1	2	3	Row Total
1	24 6.2	230 59.7	131 34.0	385 41.6
2	33 6.1	286 52.9	222 41.0	541 58.4
Column Total	57 6.2	516 55.7	353 38.1	926 100.0

Number of missing observations = 6

The researcher is left to add row and column captions from the coding list. A more advanced presentation to include captions is normally used only by experienced researchers, especially in market research, who have little time for playing with data and require an output in a form which can be (photocopied and) included directly in a research report. Whilst such presentation is the most convenient to read, it is not necessarily recommended for beginners because of the complex preparations required. Moreover all text and related programming occupies useful core storage in the computer, and the processing may add considerably to the time, and therefore cost, of the job.

Figure 8 gives an example of output with row and column captions.

Figure 8: Contingency table with labels (see figure 7 above)

```
  SEX       SEX OF RESPONDENT   by  HAPPY     HOW HAPPY IS R?


                        HAPPY
           Count  :
          Row  %  :NOT TOO  PRETTY   VERY      Row
                  :HAPPY    HAPPY    HAPPY     Total
                  :     1  :     2  :     3  :
 SEX      --------:--------:--------:--------:
                1  :    24  :   230  :   131  :   385
   MEN             :   6.2  :  59.7  :  34.0  :  41.6
                 -:--------:--------:--------:
                2  :    33  :   286  :   222  :   541
   WOMEN           :   6.1  :  52.9  :  41.0  :  58.4
                 -:--------:--------:--------:
          Column      57       516      353      926
           Total      6.2      55.7     38.1    100.0

 Number of missing observations =     6
```

A good researcher will not be content with analysing two variables at a time, but will want to test apparent relationships between two variables by controlling for a third variable. Any good survey analysis program should allow this, up to three, four or even more levels of control.

Figure 9: Conditional contingency tables - illustrating elaboration

```
   SEX        SEX OF RESPONDENT   by  HAPPY     HOW HAPPY IS R?

 Controlling for:

   AGEGROUP  GROUPED AGE OF R           Value:      1   17-29

                HAPPY
           Count  :
          Row  %  :NOT TOO  PRETTY   VERY       Row
                  :HAPPY    HAPPY    HAPPY     Total
                  :     1  :     2  :     3  :
 SEX      --------:--------:--------:--------:
               1  :     6  :    56  :    25  :    87
   MEN            :   6.9  :  64.4  :  28.7  :  42.4
                 -:--------:--------:--------:
               2  :     4  :    66  :    48  :   118
   WOMEN          :   3.4  :  55.9  :  40.7  :  57.6
                 -:--------:--------:--------:
          Column       10      122       73      205
           Total      4.9     59.5     35.6    100.0

  SEX        SEX OF RESPONDENT   by  HAPPY     HOW HAPPY IS R?

 Controlling for:

  AGEGROUP  GROUPED AGE OF R            Value:      2   30-44

                HAPPY
           Count  :
          Row  %  :NOT TOO  PRETTY   VERY       Row
                  :HAPPY    HAPPY    HAPPY     Total
                  :     1  :     2  :     3  :
 SEX      --------:--------:--------:--------:
               1  :     8  :    55  :    27  :    90
   MEN            :   8.9  :  61.1  :  30.0  :  42.1
                 -:--------:--------:--------:
               2  :     4  :    63  :    57  :   124
   WOMEN          :   3.2  :  50.8  :  46.0  :  57.9
                 -:--------:--------:--------:
          Column       12      118       84      214
           Total      5.6     55.1     39.3    100.0
```

Figure 9 Conditional contingency tables (contd)

```
SEX       SEX OF RESPONDENT   by  HAPPY     HOW HAPPY IS R?

Controlling for:

 AGEGROUP  GROUPED AGE OF R            Value:      3   45-59

              HAPPY
          Count  :
         Row  %  :NOT TOO  PRETTY   VERY       Row
                 :HAPPY    HAPPY    HAPPY     Total
                 :     1  :     2  :     3  :
SEX      --------:--------:--------:--------:
              1  :     3  :    73  :    34  :   110
  MEN            :   2.7  :  66.4  :  30.9  :  45.5
                -:--------:--------:--------:
              2  :    11  :    75  :    46  :   132
  WOMEN          :   8.3  :  56.8  :  34.8  :  54.5
                -:--------:--------:--------:
         Column       14      148       80      242
          Total      5.8     61.2     33.1    100.0
```

```
 SEX       SEX OF RESPONDENT   by  HAPPY     HOW HAPPY IS R?

Controlling for:

 AGEGROUP  GROUPED AGE OF R              Value:      4   60+

              HAPPY
          Count  :
         Row  %  :NOT TOO  PRETTY   VERY       Row
                 :HAPPY    HAPPY    HAPPY     Total
                 :     1  :     2  :     3  :
SEX      --------:--------:--------:--------:
              1  :     7  :    40  :    42  :    89
  MEN            :   7.9  :  44.9  :  47.2  :  35.5
                -:--------:--------:--------:
              2  :    14  :    78  :    70  :   162
  WOMEN          :   8.6  :  48.1  :  43.2  :  64.5
                -:--------:--------:--------:
         Column       21      118      112      251
          Total      8.4     47.0     44.6    100.0

Number of missing observations =    20
```

Finally, percentaging in tables may be needed not only in rows, but also in columns or even on the whole table.

Figure 10: Contingency table with all percentages

```
SEX        SEX OF RESPONDENT    by  AGEGROUP  GROUPED OF R

                     AGEGROUP
             Count  :
           Row   %  :17-29    30-44    45-59    60+        Row
           Col   %  :                                      Total
           Tot   %  :      1  :      2  :      3  :      4  :
 SEX               --------:--------:--------:--------:--------:
               1    :     88  :     90  :    110  :     92  :    380
   MEN              :   23.2  :   23.7  :   28.9  :   24.2  :   41.4
                    :   42.7  :   42.1  :   45.5  :   35.9  :
                    :    9.6  :    9.8  :   12.0  :   10.0  :
                   -:--------:--------:--------:--------:
               2    :    118  :    124  :    132  :    164  :    538
   WOMEN            :   21.9  :   23.0  :   24.5  :   30.5  :   58.6
                    :   57.3  :   57.9  :   54.5  :   64.1  :
                    :   12.9  :   13.5  :   14.4  :   17.9  :
                   -:--------:--------:--------:--------:
            Column       206      214      242      256      918
             Total      22.4     23.3     26.4     27.9    100.0

 Number of missing observations =    14
```

[**NB**: Such tables may occasionally be needed for checking samples against known population parameters, but their extensive use in analysis is usually a sign of inexperience and anxiety in researchers who are either too proud to ask for advice and assistance or who are possibly even completely incompetent. It is also a waste of paper, time and money!]

Now move on to **1.2: Coding data from questionnaires**

[Back to Block 1 menu]