

## Solution to Exercise 4: Aggregating data and saving the summary data

### Key Point(s):

- If aggregate data are collected, there is no way to get back to the individual records. Conversely, it is easy to aggregate data from individual records.

### Task:

- The B\_EX04\_WORKLOAD.REC has been edited to contain only three laboratories (out of the original 30) and only the year 2002. Nonsensical results (e.g., first examination not recorded, followed by a valid result) have been excluded. Create a program B\_EX04.PGM to provide you with the necessary information to determine the number of smears performed on average on each day on which at least one examinee was examined for each of the three laboratories.*

### Solution

There are two ways to solve the problem, both done here in the same program:

- \* Program B\_EX04.PGM aggregating the data to
- \* determine the workload in three laboratories
- \* Data courtesy: Biggie Mabaera, University of Zimbabwe

- \* The following is the first method

```
cls
close
logclose

cd c:\epidata_course
read "b_ex04_workload.rec"

cls
                                define smears #
                                let smears=9
if result1<9 and result2<9 and result3<9 then smears=3
if result1<9 and result2<9 and result3=9 then smears=2
if result1<9 and result2=9 and result3=9 then smears=1
if result1=9 and result2=9 and result3=9 then smears=0

cls
erase "b_ex04a.txt"
logopen "b_ex04a.txt"
tables smears laboratory
select laboratory=401
freq regdate
select

cls
select laboratory=411
freq regdate
select
```

```

cls
select laboratory=416
freq regdate
select

logclose
close

* The following is the second method
* Alternative to the above program
* using the aggregate function

cls
close
logclose

cd c:\epidata_course

read "b_ex04_workload.rec"

                                define smears ###
                                let smears=9
if result1<9 and result2<9 and result3<9 then smears=3
if result1<9 and result2<9 and result3=9 then smears=2
if result1<9 and result2=9 and result3=9 then smears=1
if result1=9 and result2=9 and result3=9 then smears=0

erase "b_ex04b.rec"

aggregate regdate laboratory /sum=smears /save="b_ex04b.rec" /replace

cls
close
read "b_ex04.rec"
sort laboratory regdate

keep regdate laboratory n sumsmears
savedata "b_ex04.rec" /replace

close
erase "b_ex04b.rec"
logclose

```

The first method uses what you learned in an earlier exercise using simple frequencies to aggregate the data. You may have noted that it is not possible to use the TABLES command as there are too many categories. Note that it is possible to write the data directly into an Excel™ \*.XLS file (in addition to the earlier shown possibilities of \*.TXT and \*.HTML files).

The second approach is to use the AGGREGATE command of EpiData Analysis. Any table a user produces in EpiData Analysis is built on this AGGREGATE command, but the user never gets to see it. It is, however, possible to use it as a function and to write the aggregated data directly into a file. This file (named here B\_EX04.REC) contains only the aggregated 618 records produced from the 12,087 records of the original B\_EX04\_WORKLOAD.REC file.

**Task:**

- o Use your spreadsheet program to do the necessary final calculations and save it as B\_EX04.XLS. Summarize the result in a simple table.*

If you used the first approach, you created an output text file (named here B\_EX04A.TXT) which could be imported into a spreadsheet, and with some cutting and pasting getting the table as required.

If you used the second approach, EpiData Analysis produced the summary aggregate file (named here B\_EX04.REC). In EpiData Entry this file could then be exported to a text file, exporting only the necessary variables LABORATORY to identify the laboratory, NSMEARS, the number of examinees on each day, and SUMSMEARS, the number of smears done on each day.

Both approaches must provide the same result:

**Method 1: Workload by laboratory**

Laboratory	Days	Smears	Smears/day
401	242	23,044	95.2
411	135	1,211	9.0
416	241	6,328	26.3

**Method 2: Workload by laboratory**

Laboratory	Days	Smears	Smears/day
401	242	23,044	95.2
411	135	1,211	9.0
416	241	6,328	26.3