

Solution to Exercise 4: Data entry and validation

Key Point(s):

- It should be routine that two persons work on data entry, and never one.
- The only and acceptable way to minimize data entry errors is to enter the data twice into two different files, and then compare the two files for discordances.
- Avoid using the mouse to move around fields during data entry, because the Check file cannot be applied to fields you skip by moving the mouse from one field to another.

Tasks:

- o **Download the solution of Exercise 3 and overwrite your A_EX03 triplet files. Go to “Tools” “Copy structure” and copy the A_EX03.REC including its A_EX03.CHK file to:**

A_EX04 A.REC **and** A_EX04A.CHK **files**

A_EX04 B.REC **and** A_EX04B.CHK **files**

- o **Enter the 15 records using the A_EX04A.REC file.**

After completing data entry, enter the same data again into to the A_EX04B.REC file.

- o **After you have completed the two files, go to “5. Document” “Validate Duplicate Files” and produce a *.NOT file giving you a list of discordances, if any. Save the *.NOT file as A_EX04AB_validation.NOT**
- o **Use “6. Export Data” “Epidata” to export either one of the two files [recommended: the A_EX04A.REC] to a new A_EX04F.REC file, and then make all corrections in this file. This is your final dataset.**

Solution:

Depending on the errors you made, you will get an output like the following:

```
VALIDATE DUPLICATE DATA FILES REPORT
=====
```

```
Report generated 19. Jun 2012 12:32
```

```
-----
Data file 1:
```

```
-----
File name:      C:\temp\a_ex04a.rec
File label:     Exercise 4, 1st entry, Part A
File date:      19. Jun 2012 12:23
Records total: 15
```

```
-----
Data file 2
-----
```

File name: C:\temp\a_ex04b.rec
File label: Exercise 4, 2nd entry, Part A
File date: 19. Jun 2012 12:32
Records total:15

Options for validation:
Ignore deleted records: Yes
Ignore text fields: No
Ignore letter-case in text fields: No
Report differences in field types: No
Ignore missing records in data file 2 No

Fields in both data files that were used in the validation:
LABCODE, IDCODE, LABNAME, SERNO, REGDATE, SEX, AGE, REASON, RES1, RES1SC, RES2,
RES2SC, RES3, RES3SC

Fields excluded from data file 1:
None

Fields excluded from data file 2:
None

Fields used as index keys:
IDCODE

RESULTS OF VALIDATION:

Records missing in data file 1: 0
Records missing in data file 2: 0

Number of common records found: 15
Number of fields checked per record: 14
Total number of fields checked: 210

3 out of 15 records had errors (20.00 pct.)
3 out of 210 fields had errors (1.43 pct.)

DATA FILE 1	DATA FILE 2
Record key field(s): (Rec. # 3) idcode = ML_J-2003-3300 res3 = 0	Record # 3 res3 = 9
Record key field(s): (Rec. # 11) idcode = ML_J-2003-3307 res3 = 0	Record # 11 res3 = 9
Record key field(s): (Rec. # 12) idcode = ML_J-2003-3308 age = 29	Record # 12 age = 39

After making correction in the “F” file, your data should be correct, or not? While your final data file should be correct, there is still a slim chance that it has errors. How is this possible? If by chance the same error was entered in both files (which can happen particularly if the same person enters the data in both files), you will not be able to identify the error. For uniformity, you should overwrite your existing file with the A_EX04F.REC file that is provided with the solution.