

Solution to Exercise 5: 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 4 and save your a_ex04.epx file as a_ex05_a.epx and a_ex05_b.epx.*
- o *Enter the 15 records using the a_ex05_a.epx file. After completing data entry, enter the same data again into the a_ex05_b.epx file.*
- o *After you have completed the two files, proceed to validation as explained here.*
- o *After ensuring that no record is missing in either file, export the a_ex05_a.epx file to a a_ex05_f.epx file, check out the discordances if any and correct them. This is your final dataset.*

Solution

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

```
=====
Report: Double Entry Validation Report.
Created: 29-04-2015 22:03:30
=====

-----
File 1: C:\EpiData_course\a_ex05_a.epx
File 2: C:\EpiData_course\a_ex05_b.epx
-----

File 1: C:\EpiData_course\a_ex05_a.epx
-----
Title      EpiData course
Created    28-04-2015 09:50:32
Last Edited 29-04-2015 20:33:53
Version    1
Cycle      29
-----
Backup on shutdown: yes
Encrypted data: no

Dataforms:
-----
Caption      Created      Structure Edited  Data Edited  Sections Fields
Records Deleted
```


Caption Fields in key

Microscopy lab (serno:Laboratory serial number) + (regyy:Registration year) + (lab:Laboratory)

=====
DataForm: Microscopy lab
=====

Selections for validation:

Options:

Option Selected

Ignore deleted records No
Ignore missing records No
Add result to field No
Case sensitive text No

Key Fields:

lab serno regyy

Compared Fields:

id: Unique identifier
regdd: Registration day
regmm: Registration month
sex: Examinee's sex
age: Examinee's age in years
reason: Reason for examination
res1: Result of specimen 1
res1sc: Result of specimen 1 scanty
res2: Result of specimen 2
res2sc: Result of specimen 2 scanty
res3: Result of specimen 3
res3c: Result of specimen 3 scanty

Result of Validation:

Overview

Test Result

Records missing in main file 0
Records missing in duplicate file 0
Non-unique records in main file 0
Non-unique records in duplicate file 0
Number of fields checked 12
Common records 15
Records with errors 1
Field entries with errors 1
Error percentage (#records) 6.67
Error percentage (#fields) 0.56

Datasets comparison:

Main Dataset: Duplicate dataset:

Record no: 5 Record no: 5

Key Fields:
lab = ML_J
serno = 3302

```
regyy = 2003
Compared Fields:
sex = 2          sex = 1
-----
```

After making correction in the “F” file, your data should be correct, or are they 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_ex05_f.epx` file that is provided with the solution.