

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.
- Never use 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 *Take your A_EX04.REC file, go to “Tools” “Copy structure” and copy the A_EX04.REC including its A_EX04.CHK file to:
A_EX05 A.REC and A_EX05A.CHK files
A_EX05 B.REC and A_EX05B.CHK files*
- o *Enter the 15 records using the A_EX05A.REC file.*

After completing data entry, enter the same data again into to the A_EX05B.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 any discordance. Save the *.NOT file as A_EX05AB.NOT*
- o *Use “6. Export Data” “Epidata” to export either one of the two files to a new A_EX05F.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 21. Jun 2008 15:45
```

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

```
-----
File name:      C:\epidata_course_Mongolia\course_exercise_files\a_ex05a.rec
File label:     Exercise 5: Data entry and validation
File date:      21. Jun 2008 15:44
Records total: 15
```

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

```
-----
File name:      C:\epidata_course_Mongolia\course_exercise_files\a_ex05b.rec
File label:     Exercise 5: Data entry and validation
File date:      21. Jun 2008 15:44
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:
SERNO,REGDATE,SEX,AGE,REASON,RES1,RES2,RES3
```

```
Fields excluded from data file 1:
None
```

```
Fields excluded from data file 2:
None
```

```
Fields used as index keys:
SERNO
```

```
-----
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:  8
Total number of fields checked:      120
```

```
2 out of 15 records had errors (13.33 pct.)
2 out of 120 fields had errors ( 1.67 pct.)
```

```
-----
DATA FILE 1 | DATA FILE 2
-----
Record key field(s): (Rec. # 11) | Record # 11
serno      = 3307 |
res3 = 0.0 | res3 = 9.0
-----
Record key field(s): (Rec. # 12) | Record # 12
serno      = 3308 |
age = 29 | age = 39
-----
```

While your 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 records), you will not be able to identify the error. For uniformity, you should overwrite your existing file with the A_EX05F.REC file that is provided with the solution.