

Solution to Exercise 1: A relational database and “Aggregating” vs from “long-to-wide”

Key points:

- A relational database is the solution to a varying number of observations per individual
- The child file is merged with the parent file to give a dataset of all observations
- To obtain means for an individual from continuous variables, aggregating the data is the strategy of choice
- To reduce the dataset to individuals with information on each examination, one must copy the information from observations in the vertical to newly created fields in the first observation of each individual before selecting that record from the individual (“long-to-wide”)

Task

- *Prepare a data documentation sheet*

The documentation sheet is shown on the next page.

Task

- *Prepare the EpiData Manager form for the relational database*

The data entry forms may be made as follows:

D_EX01_PATIENT.EPX	D_EX01_VISIT.EPX
<p>Entry form for the patient</p> <p>Unique patient identifier <input type="text"/></p> <p>Patient's sex <input type="text"/> label_sex</p> <p>Marital status <input type="text"/> label_marital</p>	<p>Entry form for the visit</p> <p>Unique patient identifier <input type="text"/></p> <p>Unique visit identifier <input type="text"/></p> <p>Date of visit <input type="text"/></p> <p>Plasma glucose in mMol/L <input type="text"/></p> <p>Quality aspect of sputum <input type="text"/> label_sputum</p> <p>Microscopy result <input type="text"/> label_micres</p>

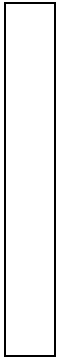
The data documentation sheet:

Data documentation sheet

	Field name	Field label	Field type	Field length	Field values	Value label	Field comment
Patient file	idpat	Unique patient identifier	U	1	A,...,Z		Any given unique ID
	sex	Patient's sex	I	1	1 Female 2 Male 3 Unknown		
	marital	Patient's marital status	I	1	1 Single 2 Married 3 Cohabiting 4 Annulled 5 Divorced 6 Widowed 7 Separated 8 Engaged 9 Not recorded		

NOTE: If SEX is given during an earlier examination and left empty in a subsequent examination, update the information to known
 If SEX is different in different examinations, record as UNKNOWN
 If MARITAL is given during an earlier examination and left empty in a subsequent examination, keep the initial information from earlier
 If MARITAL is different in different examinations, update to most recent information

Examination file	idvisit	Unique examination identifier	S	12	A-2007-01-31,...		Automatically calculated
	visitdate	Date of visit	dd/mm/yyyy	10	01/01/2007,...,31/12/2007 01/01/1800		Legal visit date recordings Enter if visit date is missing
	bs	Fasting plasma blood glucose (mMol)	F	4	2.5,...,19.9 99.9		Legal valid value Enter if blood sugar is missing
	sputum	Macroscopic sputum aspect	I	1	1 Mucoid 2 Purulent		



micres

Microscopy result

|

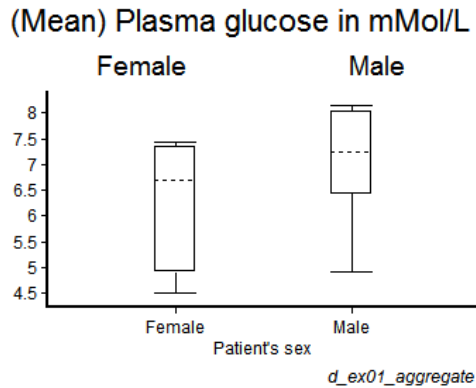
1

- 3 Muco-purulent
- 4 Blood-tinged
- 5 Salivary
- 9 Not recorded
- 0 Negative
- 1 "1+ positive"
- 2 "2+ positive"
- 3 "3+ positive"
- 4 "Scanty positive"
- 9 "Not recorded"

Task:

- Write a program *D_EX01.PGM* that merges the two files, then prepare sets for the aggregated data and for the “long-to-wide” transformation to produce the following output respectively:

From aggregating the data:



From transformation “long-to-wide”:

"Table 1. Pattern of serial smear results"

Definition of case by microscopy			
of 4 serial smears	Negative	Positive	Total
N---	2	0	2
NN--	1	0	1
NP9P	0	1	1
NPF-	0	3	3
PNF-	0	1	1
PP--	0	2	2
Total	3	7	10

"Table 2. Incremental yield among positive results"

Incremental yield of Female first 3 smears	Patient's sex		
	% Female	% Male	% Total
NPx	2 {66.7}	2 {50.0}	4 {57.1}
Px	1 {33.3}	2 {50.0}	3 {42.9}
Total	3 {100.0}	4 {100.0}	7

Percents: {Col}

The *D_EX01.PGM* reads:

```

* Part D, Exercise 1
* Merging files and aggregating files
* Copying and transposing data from "Long-to-wide"

* Written by:    Hans L Rieder
* First version: 17 Jan 2010
* Last revision: 11 Nov 2016

cls
close
logclose

*****
* Merge child and parent files

cls
close

read "d_ex01_visit.epx"
merge idpat /file="d_ex01_patient.epx" /table

sort idpat visitdate
gen i visit=1
if idpat=idpat[_n-1] then visit=visit[_n-1]+1
label visit "Visit number"

savedata "temp_01.rec" /replace

*****
* Create an aggregate data set
* to determine means

cls
close
read "temp_01.rec"

```

```

aggregate idpat sex /mean="bs" /close /save="d_ex01_aggregate.rec" /replace

cls
close
read "d_ex01_aggregate.rec"

* set display databrowser=on
* browse
* tables sex meabs //Testing here only, will be done in Analysis
* boxplot meabs /by=sex //Testing here only, will be done in Analysis

*****
* Transpose , copy "long-to-wide"

cls
close
read "temp_01.rec"

* freq exam
* => Maximum is 4 visits

cls
gen d visitdate1
gen d visitdate2
gen d visitdate3
gen d visitdate4

visitdate1=visitdate
if (idpat[_n])=(idpat[_n+1]) then visitdate2=visitdate[_n+1]
if (idpat[_n])=(idpat[_n+2]) then visitdate3=visitdate[_n+2]
if (idpat[_n])=(idpat[_n+3]) then visitdate4=visitdate[_n+3]

cls
define bs1 ##.#
define bs2 ##.#
define bs3 ##.#
define bs4 ##.#

bs1=bs
if (idpat[_n])=(idpat[_n+1]) then bs2=bs[_n+1]
if (idpat[_n])=(idpat[_n+2]) then bs3=bs[_n+2]
if (idpat[_n])=(idpat[_n+3]) then bs4=bs[_n+3]

cls
gen i sputum1
gen i sputum2
gen i sputum3
gen i sputum4

sputum1=sputum
if (idpat[_n])=(idpat[_n+1]) then sputum2=sputum[_n+1]
if (idpat[_n])=(idpat[_n+2]) then sputum3=sputum[_n+2]
if (idpat[_n])=(idpat[_n+3]) then sputum4=sputum[_n+3]

cls
gen i micres1
gen i micres2
gen i micres3
gen i micres4

micres1=micres
if (idpat[_n])=(idpat[_n+1]) then micres2=micres[_n+1]
if (idpat[_n])=(idpat[_n+2]) then micres3=micres[_n+2]
if (idpat[_n])=(idpat[_n+3]) then micres4=micres[_n+3]

cls
label visitdate1 "Date of 1st visit"
label visitdate2 "Date of 2nd visit"
label visitdate3 "Date of 3rd visit"

```

```

label visitdate4 "Date of 4th visit"
cls
label bs1 "Plasma glucose at 1st visit"
label bs2 "Plasma glucose at 2nd visit"
label bs3 "Plasma glucose at 3rd visit"
label bs4 "Plasma glucose at 4th visit"
cls
label sputum1 "Macroscopic sputum of 1st visit"
label sputum2 "Macroscopic sputum of 2nd visit"
label sputum3 "Macroscopic sputum of 3rd visit"
label sputum4 "Macroscopic sputum of 4th visit"
labelvalue sputum1-sputum4 /1="Mucoid"
labelvalue sputum1-sputum4 /2="Purulent"
labelvalue sputum1-sputum4 /3="Muco-purulent"
labelvalue sputum1-sputum4 /4="Blood-tinged"
labelvalue sputum1-sputum4 /5="Salivary"
labelvalue sputum1-sputum4 /9="Not recorded"
cls
label micres1 "Microscopy result of 1st visit"
label micres2 "Microscopy result of 2nd visit"
label micres3 "Microscopy result of 3rd visit"
label micres4 "Microscopy result of 4th visit"
labelvalue micres1-micres4 /0="Negative"
labelvalue micres1-micres4 /1="1+ positive"
labelvalue micres1-micres4 /2="1+ positive"
labelvalue micres1-micres4 /3="1+ positive"
labelvalue micres1-micres4 /4="Scanty positive"
labelvalue micres1-micres4 /9="Not recorded"
cls
label marital "Civil status"

select visit=1
drop visitid mergevar visit

savedata "temp_02.rec" /replace

cls
close
read "temp_02.rec"

define restxt1 _
define restxt2 _
define restxt3 _
define restxt4 _

cls
restxt1="-"
if micres1>0 and micres1<9 then restxt1="P"
if micres1=0 then restxt1="N"
if micres1=9 then restxt1="9"
cls
restxt2="-"
if micres2>0 and micres2<9 then restxt2="P"
if micres2=0 then restxt2="N"
if micres2=9 then restxt2="9"
cls
restxt3="-"
if micres3>0 and micres3<9 then restxt3="P"
if micres3=0 then restxt3="N"
if micres3=9 then restxt3="9"
cls
restxt4="-"
if micres4>0 and micres4<9 then restxt4="P"
if micres4=0 then restxt4="N"
if micres4=9 then restxt4="9"

```

```

cls
define pattern ____
pattern=restxt1+restxt2+restxt3+restxt4
label pattern "Pattern of 4 serial smears"

* freq pattern

cls
define case #
                                case=0
if substr(pattern,1,1)="P" then case=1
if substr(pattern,2,1)="P" then case=1
if substr(pattern,3,1)="P" then case=1
if substr(pattern,4,1)="P" then case=1

label case "Definition of case by microscopy"
labelvalue case /0="Negative"
labelvalue case /1="Positive"

define yield ____
if substr(pattern,1,3)="N--" then yield="N99"
if substr(pattern,1,3)="NN-" then yield="NN9"
if substr(pattern,1,3)="NP9" then yield="NPx"
if substr(pattern,1,3)="NPP" then yield="NPx"
if substr(pattern,1,3)="PNP" then yield="Px "
if substr(pattern,1,3)="PP-" then yield="Px"
label yield "Incremental yield of first 3 smears"

keep idpat sex marital \
    visitdate1 visitdate2 visitdate3 visitdate4 \
    sputum1 sputum2 sputum3 sputum4 \
    micres1 micres2 micres3 micres4 \
    case pattern yield
savedata "d_ex01.rec" /replace

*****
* Produce tables on smear pattern and incremental yield

cls
close
read "d_ex01_aggregate.rec"

set option graph /sizex=400
set graph footnote="d_ex01_aggregate"

set echo=off
cls
boxplot meabs /by=sex /bw /sub="    Female                                Male"

cls
close
read "d_ex01.rec"

title "Table 1.  Pattern of serial smear results"
tables case pattern
select case=1
title "Table 2.  Incremental yield among positive results"
tables sex yield /c /PCT
set echo=on

*****
* Clean up

set echo=off
define yesno # global
cls

```

```
yesno=?Delete temporary files: 1=yes 0=no?
imif yesno=1 then
  erasepng /all /noconfirm
  erase "temp_01.rec"
  erase "temp_01.chk"
  erase "temp_02.rec"
  erase "temp_02.chk"
  erase "d_ex01_aggregate.chk"
  erase "d_ex01_aggregate.rec"
  select
  cls
  type "All temporary files erased" /h2
else
  select
  type "File D_EX01_EXAMINEE.REC remains open" /h2
endif
set echo=on
```