

## Solution to Exercise 1: A basic program in EpiData Analysis

### Key Point(s):

- It is important to make a comment first in the program. This preferably is what you want to do in that program. Comments are preceded by an asterisk and are bypassed by the analysis program.
- The F9 key runs the whole program whilst the F8 key runs only the selected part of the program.

### Tasks:

- o Determine the year of birth (new variable created from age and date of registration), then make groups of examinees (another variable) born respectively between 1900 and 1929, 1930 and 1949, 1950 and 1999, and those without known year of birth.*
- o Use two approaches, one with text field coding and the other with numeric coding and value labels.*

### Solution:

The output solution is as follows:

```
Examinee's sex
Exercise birth years Female Male Total
text coding
1900-1929                3    2    5
1930-1949                4    5    9
1950-1999               25   34   59
Unknown                  1    1    2
Total                   33   42   75
. tables sex birthgrp2
Examinee's sex
Exercise birth years Female Male Total
numeric coding
Born 1900 to 1929        3    2    5
Born 1930 to 1949        4    5    9
Born 1950 to 1999       25   34   59
Unknown birth year       1    1    2
Total                   33   42   75
```

The program B\_EX01.PGM might look as follows (exercises and tasks):

```
* This is the program b_ex01.pgm, my first program

*****
* Preparatory steps 1-3

* 1: Start with a clean slate
cls
logclose
close

* 2: Tell EpiData where the work is done
```

```

cd c:\epidata_course

* 3: Read the file you wish to analyze
read "a.rec"

*****
* Examples with the two most frequently used commands:

*   FREQ
*   TABLES

*   FREQ
freq sex
freq sex reason
freq *

*   TABLES
tables sex reason
tables reason res1

*****
* Options in EpiData Analysis:
*   Without options, you get just the numbers
*   Options allow obtaining more information
*   Options start with a forward slash at the end of a line: /
*   After the forward slash you enter the option(s) you need

* Options to get different percentages:

tables sex reason /c
tables sex reason /r
tables sex reason /c /r
set table percent format col=P1[]
tables sex reason /c /r
tables sex reason /c /r /pct
tables sex reason /c /d0

* Options to see values and value labels

freq sex
freq sex /vl
freq sex /v
freq sex /vn
freq sex /vnl

*****
* Selecting records meeting some criteria:
*   SELECT
*   followed by the field name,
*   followed by an operator,
*   followed by a value of that field
*   gives you the desired selection
*   SELECT
*   without anything restores the full dataset

cls
freq reason /vl
select reason=0

```

```
tables sex res1
select
tables sex res1
```

\* For operators see Help (F1)

\*\*\*\*\*

\* Define new variables in memory  
\* DEFINE  
\* followed by a new variable name  
\* followed by the variable definition and variable length  
\* gives a new variable

\* Define a text variable of length 5  
\* Then give values to the new variable based  
\* on an existing variable, e.g., WHO age groups:

```
* define WHO age groups
      define agegrp1 _____
      agegrp1="other"
if age>=00 and age<15 then agegrp1="00-14"
if age>=15 and age<25 then agegrp1="15-24"
if age>=25 and age<35 then agegrp1="25-34"
if age>=35 and age<45 then agegrp1="35-44"
if age>=45 and age<55 then agegrp1="45-54"
if age>=55 and age<65 then agegrp1="55-64"
if age>=65 and age<99 then agegrp1="65 + "
if age>=99          then agegrp1="unkn"
```

```
cls
tables sex agegrp1
select age<>99
tables sex agegrp1
select
```

\* Define a numeric variable of length 1  
\* Then give values to the new variable based  
\* on an existing variable, e.g., WHO age groups  
\* Then give a field LABEL to the new variable  
\* and a LABELVALUE (field values):

```
* define WHO age groups
      define agegrp2 #
      agegrp2=8
if age>=00 and age<15 then agegrp2=1
if age>=15 and age<25 then agegrp2=2
if age>=25 and age<35 then agegrp2=3
if age>=35 and age<45 then agegrp2=4
if age>=45 and age<55 then agegrp2=5
if age>=55 and age<65 then agegrp2=6
if age>=65 and age<99 then agegrp2=7
if age>=99          then agegrp2=9
label agegrp2 "WHO age groups"
labelvalue agegrp2 /1="0 to 14 years"
labelvalue agegrp2 /2="15 to 24 years"
labelvalue agegrp2 /3="25 to 34 years"
labelvalue agegrp2 /4="35 to 44 years"
labelvalue agegrp2 /5="45 to 54 years"
labelvalue agegrp2 /6="55 to 64 years"
labelvalue agegrp2 /7="65 years and older"
labelvalue agegrp2 /8="Other" //Note this should not exist - control
```

```

labelvalue agegrp2 /9="Unknown age"

cls
tables sex agegrp2
select age<>99
tables sex agegrp2
select

*****
*****
* Solution of task

* define the year of birth
define birthyr ####
birthyr=year(regdate)-age

* Using text variables
* define groupings for birth years
define birthgrp1 _____
                                let birthgrp1="other"
if birthyr>1899 and birthyr<1930 then birthgrp1="1900-1929"
if birthyr>1929 and birthyr<1950 then birthgrp1="1930-1949"
if birthyr>1949 and birthyr<2000 then birthgrp1="1950-1999"
if age=99                                then birthgrp1="Unknown"
label birthgrp1 "Exercise birth years text coding"

* Using numeric variables
* define groupings for birth years
define birthgrp2 #
                                let birthgrp2=8
if birthyr>1899 and birthyr<1930 then birthgrp2=1
if birthyr>1929 and birthyr<1950 then birthgrp2=2
if birthyr>1949 and birthyr<2000 then birthgrp2=3
if age=99                                then birthgrp2=9
label birthgrp2 "Exercise birth years numeric coding"
labelvalue birthgrp2 /1="Born 1900 to 1929"
labelvalue birthgrp2 /2="Born 1930 to 1949"
labelvalue birthgrp2 /3="Born 1950 to 1999"
labelvalue birthgrp2 /8="Unaccounted for"
labelvalue birthgrp2 /9="Unknown birth year"

cls
tables sex birthyr
tables sex birthgrp1
tables sex birthgrp2

```