Solution to Exercise 3: Incremental yield from serial smears

Key Learning Points

When you have a hypothesis to test, remember that it may be logical to:

a. Create and use a subset of the working dataset
b. Create new variable(s)
c. Make use of other software applications e.g. a spreadsheet to make calculations.

Tasks:

Exercise hypothesis:

H_0: Not more than 125 third smear examinations have to be made to find one additional case of tuberculosis in each of the four study countries

- ***Determine with a program C_EX03.PGM the number of suspects with the patterns listed above***
- ***Create a table in spreadsheet by country***
- ***Interpret the findings***

Solution:

The following output was created in EpiData Analysis:

Pattern of serial smears

Study country	N99	NN9	NNN	NNP	NPX	PX	Total
Moldova	1381	1579	8424	34	84	1013	12515
Mongolia	414	708	12264	12	42	1663	15103
Uganda	10713	3325	14736	107	487	6686	36054
Zimbabwe	1795	2706	17740	155	325	2969	25690
Total	14303	8318	53164	308	938	12331	89362

using the following program C_EX03.PGM:

```
This is b_ex03 EpiData Analysis program
* to determine the incremental yield from serial smears

cls
close
logclose

cd c:\epidata_course

read  "c_ex01.rec"

* Definition positive: any AFB in any of three results
* Values: "P" (positive) or "N" (negative)
* or "9" (unknown)

define restxt1 _
```

```
if result1=0               then restxt1="N"
if result1=9               then restxt1="9"
if result1>0 and result1<9 then restxt1="P"

define restxt2 _
if result2=0               then restxt2="N"
if result2=9               then restxt2="9"
if result2>0 and result2<9 then restxt2="P"

define restxt3 _
if result3=0               then restxt3="N"
if result3=9               then restxt3="9"
if result3>0 and result3<9 then restxt3="P"

define pattern ___
pattern=restxt1+restxt2+restxt3
label pattern "All observed patterns"

define case #
case=0
if result1>0 and result1<9 then case=1
if result2>0 and result2<9 then case=1
if result3>0 and result3<9 then case=1
label case "Case definition"
labelvalue case /0="Non-case"
labelvalue case /1="Case"

cls
* Define essential patterns from
* all possible patterns

define esspatt #
                                esspatt=9
if substr(pattern,1,3)="NNN" then esspatt=4
if substr(pattern,1,3)="NN9" then esspatt=5
if substr(pattern,1,3)="N99" then esspatt=6
if substr(pattern,3,1)="P"   then esspatt=1
if substr(pattern,2,1)="P"   then esspatt=2
if substr(pattern,1,1)="P"   then esspatt=3
label esspatt "Essential patterns"
labelvalue esspatt /1="NNP"
labelvalue esspatt /2="NPx"
labelvalue esspatt /3="Px"
labelvalue esspatt /4="NNN"
labelvalue esspatt /5="NN9"
labelvalue esspatt /6="N99"
labelvalue esspatt /9="Not allocated"

select reason=0

drop restxt1 restxt2 restxt3
savedata "temp.rec" /replace

*********************************
* Produce requested output
*********************************
* First approach: output table for
* into spreadsheet

cls
close
logclose

read "temp.rec"

set echo=off
logopen "b_ex03.txt" /replace
freq pattern
tables esspatt country /r
select esspatt<4
tables esspatt country /r
logclose
select
set echo=on
```

This table was pasted into the spreadsheet `C_EX03.XLS` to calculate the proportion of cases:

	A	B	C	D	E	F	G	H	I	J
1	Exercise 3. Patterns of serial sputum smear examinations									
2										
3										
4		Pattern of serial smear examination results								Proportion
5	**Country**	**N99**	**NN9**	**NNN**	**NNP**	**NPX**	**PX**	**Total**		**Cases**
6	**Total**	14,303	8,318	53,164	308	938	12,331	89,362		0.152
7										
8	**Moldova**	1,381	1,579	8,424	34	84	1,013	12,515		0.090
9	**Mongolia**	414	708	12,264	12	42	1,663	15,103		0.114
10	**Uganda**	10,713	3,325	14,736	107	487	6,686	36,054		0.202
11	**Zimbabwe**	1,795	2,706	17,740	155	325	2,969	25,690		0.134

Using “Copy”, “Paste special” and “Transpose” the information was carried into a second sheet of the same spreadsheet and the calculations completed:

Exercise 3. Number of additional smears that have to be examined to find one additional case on a third serial sputum smear examination following Two negative results, by country

		Moldova	Mongolia	Uganda	Zimbabwe	Total
Total		12,515	15,103	36,054	25,690	89,362
Pattern						
	N99	1,381	414	10,713	1,795	14,303
	NN9	1,579	708	3,325	2,706	8,318
	NNN	8,424	12,264	14,736	17,740	53,164
	NNP	34	12	107	155	308
	Npx	84	42	487	325	938
	Px	1,013	1,663	6,686	2,969	12,331
Yield						
	First	0.896	0.969	0.918	0.861	
	Second	0.074	0.024	0.067	0.094	
	Third	0.030	0.007	0.015	0.045	
	X	125	125	125	125	NA
	P	0.00402	0.00098	0.00721	0.00866	
	SE(P)	0.00069	0.00028	0.00069	0.00069	
	95% low	0.00267	0.00042	0.00585	0.00730	
	95% high	0.00537	0.00153	0.00857	0.01002	
Smears		248.8	1,023.0	138.7	115.5	
	95% low	186.3	653.5	116.7	99.8	
	95% high	374.3	2,354.6	171.0	136.9	
Hypothesis:		Refute	Refute	Accept	Accept	

Interpretation: The research hypothesis is refuted for Moldova and Mongolia, but accepted for Uganda and Zimbabwe.

Moldova has a higher incremental yield from the third smear than Uganda for example, but the prevalence of cases is much smaller. Zimbabwe has both a high yield from the third serial smear and a high prevalence of cases. As a result, the relative efficiency of sputum smear examination is the best among all countries (still a very large number of smears has to be examined). Mongolia has a very poor yield of the third smear, and together with the finding in the previous exercise that there is very little variation among those with at least one positive result, it is suggestive that this low yield is attributable to the failure to examine thoroughly a third smear after two had already been negative. In other words, there is probably little point in requiring three examinations if the third is not done properly to begin with.

Alternative solution circumventing the need for the spreadsheet

It is possible to get EpiData Analysis to provide all your essential final output:

country	sm95low	smpoint	sm95high	hypothesis
Moldova	186.3	248.8	374.3	Refute
Mongolia	653.5	1023.0	2354.0	Refute
Uganda	116.7	138.7	171.0	Accept
Zimbabwe	99.8	115.5	136.9	Accept

The portion of the program that follows the main program above and accomplishes this would be something like:

```
*******************************
* Second approach: handle everything in
* EpiDat Analysis

cls
close
logclose

read "temp.rec"

aggregate esspatt country /save="yield.rec" /replace /close

cls
close
read "yield.rec"
select esspatt=1
define nnp #####
nnp=n
savedata "nnp.rec" /replace

cls
close
read "yield.rec"
select esspatt=2
define npx #####
npx=n
savedata "npx.rec" /replace

cls
close
read "yield.rec"
select esspatt=3
define px #####
px=n
savedata "px.rec" /replace

cls
close
read "yield.rec"
select esspatt=4
define nnn #####
nnn=n
savedata "nnn.rec" /replace

cls
close
read "yield.rec"
select esspatt=5
define nn9 #####
nn9=n
savedata "nn9.rec" /replace

cls
close
read "yield.rec"
select esspatt=6
define n99 #####
```

```
n99=n
savedata "n99.rec" /replace

cls
close
read "nnp.rec"
merge country /file="npx.rec"
merge country /file="px.rec"
merge country /file="nnn.rec"
merge country /file="nn9.rec"
merge country /file="n99.rec"

define tot #####
tot=nnp+npx++px+nnn+nn9+n99

define totpos #####
totpos=nnp+npx++px

drop n esspatt mergevar
savedata "esspatt.rec" /replace

cls
close
read "esspatt.rec"

define p #.#######
p=nnp/(nnp+nnn)

define sep #.#######
sep=sqrt(p*(1-P)/(nnp+nnn))

define cilow #.#######
cilow=p-1.96*sep

define cihigh #.#######
cihigh=p+1.96*sep

define smpoint ###.#
smpoint=1/p

define sm95low ###.#
sm95low=1/cihigh

define sm95high ###.#
sm95high=1/cilow

             define hypothesis  ______
                    hypothesis="Accept"
if sm95low>125 then hypothesis="Refute"

cls
logopen "c_ex03.txt" /replace
list country sm95low smpoint sm95high hypothesis
logclose
```