PCBs sampling desgin summary

Xuelong Wang

2020-01-13

Contents

1   PCBs data summary

There are 3 surveys were conducted during the 6 years from 1999-2014. The number of PCBs which are measured are varied from each survey. Besides, the measurement sensitivity were improved along the time. There are some details of the PCBs measurements can be found in **this one** at page 3.

Another thing is the NHANES adopted a sub-sampling method after 2005 for PCBs measurements. It seems that they collect all the blood samples from all the subjects but they only chose a sub-sample of them to measure the PCBs value. The details of the pool sampling method can be found in this **this one**. The basic idea is following:

1. Divide the whole subjects into 32 demographic groups

2. For all the subjects in each demographic group, split them into pools with sample size around 8. Note that the number of pools are proportion to the total number of subjects in the demographic group.

3. Draw a random sample from each pools, so that those sub-samples keeps same pattern in term of demographic groups ratio.

The following is a summary table of sub-samples of 2005-2006 subjects' for getting PCBs measurements.

Table 2. Number of subjects per demographic group in the NHANES 2005–2006 one-third subsample, number of individual serum samples available, number of usable samples, and number of pools formed from usable samples.

Race/Ethnicity	Gender	Number of subjects in the one-third subsample/ Number of Samples Available/Number of Usable Samples (Number of Pools)			
		12–19 years (RIDAGGRP=1)[1]	20–39 years (RIDAGGRP=2)	40–59 years (RIDAGGRP=3)	60+ years (RIDAGGRP=4)
Non-Hispanic White (ETHNICTY=1)[1]	Male (RIAGENDR=1)[1]	81/76/72 (9)	110/108/96 (12)	114/110/96 (12)	141/137/120 (15)
	Female (RIAGENDR=2)	94/85/80 (10)	143/136/128 (16)	116/111/104 (13)	149/146/136 (17)
Non-Hispanics Black (ETHNICTY=2)	Male (RIAGENDR=1)	129/114/104 (13)	60/57/48 (6)	56/51/40 (5)	55/52/40 (5)
	Female (RIAGENDR=2)	132/117/112 (14)	74/66/56 (7)	65/62/56 (7)	55/48/40 (5)
Mexican American (ETHNICTY=3)	Male (RIAGENDR=1)	106/96/88 (11)	87/84/72 (9)	44/43/32 (4)	38/38/32 (4)
	Female (RIAGENDR=2)	143/133/128 (16)	88/84/72 (9)	50/50/48 (6)	38/37/24 (3)
Other (ETHNICTY=4)	Male (RIAGENDR=1)	20/19/16 (2)	27/26/24 (3)	24/23/23[2] (3)	9/8/8 (1)
	Female (RIAGENDR=2)	31/26/24 (3)	38/34/32 (4)	18/18/16 (2)	10/6/6[3] (1)

1 Value of this categorical variable in the data set.
2 With only 23 usable samples, two 8 sample pools and one 7 sample pool were created.
3 With only 6 usable samples, one 6 sample pool was created.

Followings are some brief summary of the PCBs data.

1.1 1999-2004

The types of PCBs measured for each survey is

```
$`1999`
 [1] "PCB028" "PCB052" "PCB066" "PCB074" "PCB099" "PCB101" "PCB105" "PCB118" "PCB128" "PCB138"
[11] "PCB146" "PCB153" "PCB156" "PCB157" "PCB167" "PCB170" "PCB172" "PCB177" "PCB178" "PCB180"
[21] "PCB183" "PCB187"

$`2001`
 [1] "PCB052" "PCB066" "PCB074" "PCB087" "PCB099" "PCB101" "PCB105" "PCB110" "PCB118" "PCB128"
[11] "PCB138" "PCB146" "PCB149" "PCB151" "PCB153" "PCB156" "PCB157" "PCB167" "PCB170" "PCB172"
[21] "PCB177" "PCB178" "PCB180" "PCB183" "PCB187" "PCB189" "PCB194" "PCB195" "PCB196" "PCB199"
[31] "PCB206"

$`2003`
 [1] "PCB028" "PCB066" "PCB074" "PCB105" "PCB118" "PCB156" "PCB157" "PCB167" "PCB189" "PCB044"
[11] "PCB049" "PCB052" "PCB087" "PCB099" "PCB101" "PCB110" "PCB128" "PCB138" "PCB146" "PCB149"
[21] "PCB151" "PCB153" "PCB170" "PCB172" "PCB177" "PCB178" "PCB180" "PCB183" "PCB187" "PCB194"
[31] "PCB195" "PCB196" "PCB199" "PCB206" "PCB209"
```

The common PCBs that were measured by each survey is

```
 [1] "PCB052" "PCB066" "PCB074" "PCB099" "PCB101" "PCB105" "PCB118" "PCB128" "PCB138" "PCB146"
[11] "PCB153" "PCB156" "PCB157" "PCB167" "PCB170" "PCB172" "PCB177" "PCB178" "PCB180" "PCB183"
[21] "PCB187"
```

I will use those 21 PCBs to calculate their covariance matrix.

The total number of PCBs data from 1999-2004 is 7106. After I remove all the missing data, what I get is total 4873 observations.

1.2 2005-2014

The types of PCBs measured for each survey is

```
$`2005`
 [1] "PCB028" "PCB044" "PCB049" "PCB052" "PCB066" "PCB074" "PCB087" "PCB099" "PCB101" "PCB105"
[11] "PCB110" "PCB114" "PCB118" "PCB123" "PCB128" "PCB138" "PCB146" "PCB149" "PCB151" "PCB153"
[21] "PCB156" "PCB157" "PCB167" "PCB170" "PCB172" "PCB177" "PCB178" "PCB180" "PCB183" "PCB187"
[31] "PCB189" "PCB194" "PCB195" "PCB196" "PCB199" "PCB206" "PCB209"

$`2007`
 [1] "PCB028" "PCB044" "PCB049" "PCB052" "PCB066" "PCB074" "PCB087" "PCB099" "PCB101" "PCB105"
[11] "PCB110" "PCB114" "PCB118" "PCB123" "PCB128" "PCB138" "PCB146" "PCB149" "PCB151" "PCB153"
[21] "PCB156" "PCB157" "PCB167" "PCB170" "PCB172" "PCB177" "PCB178" "PCB180" "PCB183" "PCB187"
[31] "PCB189" "PCB194" "PCB195" "PCB196" "PCB199" "PCB206" "PCB209"

$`2009`
 [1] "PCB028" "PCB066" "PCB074" "PCB099" "PCB105" "PCB114" "PCB118" "PCB138" "PCB146" "PCB153"
[11] "PCB156" "PCB157" "PCB167" "PCB170" "PCB178" "PCB180" "PCB183" "PCB187" "PCB189" "PCB194"
[21] "PCB196" "PCB199" "PCB206" "PCB209"
```

```
$`2011`
 [1] "PCB028" "PCB066" "PCB074" "PCB099" "PCB105" "PCB114" "PCB118" "PCB138" "PCB146" "PCB153"
[11] "PCB156" "PCB157" "PCB167" "PCB170" "PCB178" "PCB180" "PCB183" "PCB187" "PCB189" "PCB194"
[21] "PCB196" "PCB199" "PCB206" "PCB209"

$`2013`
 [1] "PCB028" "PCB066" "PCB074" "PCB099" "PCB105" "PCB114" "PCB118" "PCB138" "PCB146" "PCB153"
[11] "PCB156" "PCB157" "PCB167" "PCB170" "PCB178" "PCB180" "PCB183" "PCB187" "PCB189" "PCB194"
[21] "PCB196" "PCB199" "PCB206" "PCB209"
```

The common PCBs that were measured by each survey is

```
 [1] "PCB028" "PCB066" "PCB074" "PCB099" "PCB105" "PCB114" "PCB118" "PCB138" "PCB146" "PCB153"
[11] "PCB156" "PCB157" "PCB167" "PCB170" "PCB178" "PCB180" "PCB183" "PCB187" "PCB189" "PCB194"
[21] "PCB196" "PCB199" "PCB206" "PCB209"
```

I will use those 24 PCBs to calculate their covariance matrix.

The total number of PCBs data from 2005-2014 is 1347. After I remove all the missing data, what I get is total 1228 observations.

Note that I only work on the PCBs measurement without adjustment, but I used the under the limit adjustment.

2 Some resource

All the information of NHANES data could be found in their **webseit**, all the information of PCBs are in the Laboratory data section, for example **webseit**.