Appendix 1

Formulas Derivation for Simmons Model under Three-stage Sampling

Let the overall survey cost C be

$$C = C_0 + C_1 n_1 + C_2 n_1 \bar{n}_2 + C_3 n_1 \bar{n}_2 \bar{n}_3. \quad (10)$$

The variance of p can also be written in the following alternative form:

$$V(p) = -\frac{\sigma_1^2}{N_1} + \frac{1}{n_1}\left(\sigma_1^2 - \sigma_2^2/\bar{N}_2\right) + \frac{1}{n_1\bar{n}_2}\left(\sigma_2^2 - \sigma_3^2/\bar{N}_3\right) + \frac{\sigma_3^2}{n_1\bar{n}_2\bar{n}_3}. \quad (11)$$

To minimize the sampling cost C under a given variance ($V(p) = V$), the optimum sampling size can be considered as the minimal values of function (10) subject to the constraint (11). The Lagrange function L is defined as

$$L(n_1, n_1\bar{n}_2, n_1\bar{n}_2\bar{n}_3, \lambda) = C + \lambda\left[V - V(P)\right],$$

that is,

$$L(n_1, n_1\bar{n}_2, n_1\bar{n}_2\bar{n}_3, \lambda) = C_0 + C_1 n_1 + C_2 n_1\bar{n}_2 + C_3 n_1\bar{n}_2\bar{n}_3 + \lambda\left[V - V(P)\right],$$

where λ is a Lagrange multiplier.

The necessary conditions for the solution of the problem are

$$\begin{cases} \dfrac{\partial L}{\partial n_1} = C_1 + \dfrac{1}{n_1^2}\left(\sigma_1^2 - \sigma_2^2/\bar{N}_2\right) \times \lambda = 0 \\ \dfrac{\partial L}{\partial (n_1\bar{n}_2)} = C_2 + \dfrac{1}{(n_1\bar{n}_2)^2}\left(\sigma_2^2 - \sigma_3^2/\bar{N}_3\right) \times \lambda = 0 \\ \dfrac{\partial L}{\partial (n_1\bar{n}_2\bar{n}_3)} = C_3 + \dfrac{1}{(n_1\bar{n}_2\bar{n}_3)^2}\sigma_3^2 \times \lambda = 0 \end{cases}.$$

Equations above-mentioned gives

$$\bar{n}_2 = \sqrt{\frac{\sigma_2^2 - \sigma_3^2/\bar{N}_3}{\sigma_1^2 - \sigma_2^2/\bar{N}_2} \bullet \frac{C_1}{C_2}}, \quad (12)$$

$$\bar{n}_3 = \sqrt{\frac{\sigma_3^2}{\sigma_2^2 - \sigma_3^2/\bar{N}_3} \bullet \frac{C_2}{C_3}}, \quad (13)$$

$$n_1 = \frac{\sigma_1^2 - \sigma_2^2/\bar{N}_2 + \left(\sigma_2^2 - \sigma_3^2/\bar{N}_3\right)/\bar{n}_2 + \sigma_3^2/\bar{n}_2\bar{n}_3}{V + \sigma_1^2/N_1}. \quad (14)$$

http://dx.doi.org/10.21037/apm-20-2262

The minimum value of $V(p)$ under a cost function (fixed survey cost C), the optimum sampling size is obtained as the minimum values of function (11) subject to the constraint (10). Consider the following Lagrange function :

$$L\left(n_1, n_1\bar{n}_2, n_1\bar{n}_2\bar{n}_3, \lambda\right)=V\left(P\right)+\lambda\left[C-C_0-C_1n_1-C_2n_1\bar{n}_2-C_3n_1\bar{n}_2\bar{n}_3\right],$$

and the detailed form is as follows,

$$\begin{aligned}&L\left(n_1, n_1\bar{n}_2, n_1\bar{n}_2\bar{n}_3, \lambda\right)=\\&-\frac{\sigma_1^2}{N_1}+\frac{1}{n_1}\left(\sigma_1^2-\sigma_2^2/\bar{N}_2\right)+\frac{1}{n_1\bar{n}_2}\left(\sigma_2^2-\sigma_3^2/\bar{N}_3\right)+\frac{\sigma_3^2}{n_1\bar{n}_2\bar{n}_3}\\&+\lambda\left[C-C_0-C_1n_1-C_2n_1\bar{n}_2-C_3n_1\bar{n}_2\bar{n}_3\right],\end{aligned}$$

where λ is a Lagrange multiplier.

The optimum n_1, $\bar{n}_2$ and $\bar{n}_3$ are the solutions of the following numerical problem:

$$\begin{cases}\dfrac{\partial L}{\partial n_1}=-\dfrac{1}{n_1^2}\left(\sigma_1^2-\sigma_2^2/\bar{N}_2\right)-\lambda C_1=0\\[2ex]\dfrac{\partial L}{\partial\left(n_1\bar{n}_2\right)}=-\dfrac{1}{\left(n_1\bar{n}_2\right)^2}\left(\sigma_2^2-\sigma_3^2/\bar{N}_3\right)-\lambda C_2=0.\\[2ex]\dfrac{\partial L}{\partial\left(n_1\bar{n}_2\bar{n}_3\right)}=-\dfrac{1}{\left(n_1\bar{n}_2\bar{n}_3\right)^2}\sigma_3^2-\lambda C_3=0\end{cases}$$

Results are presented as below:

$$\bar{n}_2=\sqrt{\frac{\sigma_2^2-\sigma_3^2/\bar{N}_3}{\sigma_1^2-\sigma_2^2/\bar{N}_2}\bullet\frac{C_1}{C_2}},\left(12\right)$$

$$\bar{n}_3=\sqrt{\frac{\sigma_3^2}{\sigma_2^2-\sigma_3^2/\bar{N}_3}\bullet\frac{C_2}{C_3}},\left(13\right)$$

$$n_1=\frac{C-C_0}{C_1+C_2\bar{n}_2+C_3\bar{n}_2\bar{n}_3}.\left(15\right)$$

http://dx.doi.org/10.21037/apm-20-2262

Appendix 2

Questionnaire of Random Response Technique Method

Hello! This survey was conducted on the sensitive issues concerning personal sexual behaviors, which was jointly carried out by Chinese Center for Disease Control and Prevention and Xichang Dermatology and Venereal Disease Prevention and Control Station.

In order to better protect your privacy, we have adopted random response technique (RRT) method for some more sensitive issues, that is, fill in the corresponding survey according to the figure on the ball.

Your participation is very important for us to carry out AIDS prevention and research work in the future. We hope to get your understanding and cooperation. We will keep the results strictly confidential. The survey will take about 10 minutes. Thank you! Please write down the corresponding characters or values on_____ and the relevant values or order numbers of answers in ☐.

I agree to participate and truly answer（fill in ☐ with √）:
1、☐ Yes　　2、☐ No （The investigation is over.）

B、General conditions

B01 Your age: _____ one full year of life

B02 Your marital status: ☐

1、Married cohabitation　2、Married separation　3、Unmarried cohabitation
4、Unmarried single　5、Divorced / widowed cohabitation　6、Divorced / widowed single

B03 Your place of domicile: ☐

1、Xichang　2、Liangshan Prefecture barring Xichang
3、Sichuan Province barring Liangshan Prefecture（Please specify: _____city/ Prefecture）
4、other provinces / Municipality（Please specify: _____province/ Municipality）

B04 Your living time in Xichang: ☐

1、≤3 months　2、3～6 months　3、6～12 months　4、1～2 years　5、≥2 years

B05 Your nationality: ☐

1、Han nationality　2、Yi nationality　3、other（Please specify: _____ nationality）

B06 Your degree of education: ☐

1、illiteracy　2、Primary school　3、middle school
4、High school or special secondary school　5、Junior College
6、Undergraduate and above

http://dx.doi.org/10.21037/apm-20-2262

B07 Are you still engaged in other work: ☐

1、No 2、Yes（Please specify:_____）

B08 Where is the main place of your sexual services in Xichang now: _____ Street / Road

B09 Your current place of providing sexual services:_____

（Please answer the specific type of venue. The following categories are for reference only.）

1、Separate sauna/bath center/massage center

2、Separate nightclub/ Karaoke Hall / discotheque /bar

3、Star-rated hotel（sauna/bath center/ hair salon / nightclub / Karaoke Hall / discotheque /bar etc.）

4、Non-star hotel（sauna/bath center/ hair salon / nightclub / Karaoke Hall / discotheque /bar etc.）

5、Small hotel or rest house 6、Own home / private home

7、Roadside/streetside: hair salon/beauty salon/hairdressing salon/shampoo room/footbath room

8、Station pile 9、Other（Please specify:_____）

C、Dichotomous sensitive questions

C01：

- **Draw No. 1 ball：**Please answer the question: "Whether there is a spouse or any regular sex partner other than the sexual serving object that charges?"
- **Draw No. 2 ball：** Please answer the question: "Is your date of birth an odd number?"

☐1、Yes 2、No

（Note：No. 1 ball：No. 2 ball= 6:4）

C02：

- **Draw No. 1 ball：** Please answer the question: "When you were diagnosed with a sexually transmitted disease by a doctor, did you stop sexual services?"
- **Draw No. 2 ball：** Please answer the question: "Is your date of birth an odd number?"

☐1、Yes 2、No

（Note：No. 1 ball：No. 2 ball= 6:4）

http://dx.doi.org/10.21037/apm-20-2262

C03：

- **Draw No. 1 ball：** Please answer the question: “Someone suggested that the state should allow prostitution to be legalized. Do you agree?”
- **Draw No. 2 ball：** Please answer the question: “Is your date of birth an odd number?”

☐1、Yes　　2、No

（Note：No. 1 ball：No. 2 ball= 6:4）

The investigation is over. Thank you for your participation!

Questionnaire quality controller： _____ **Date：** _____

http://dx.doi.org/10.21037/apm-20-2262

Appendix 3

```
%macro simulated_population(N=,
                            N1=,
                            N2=,
                            seed=);
data xc1;
do id=1 to &N;
    seed=&seed;
    h=rantbl(&seed,0.5525,0.4475);
    birthdate=rantbl(&seed,0.5,0.5);
    primary_unit=ceil(0+(&N1-0)*ranuni(&seed));
    second_unit=ceil(0+(&N2-0)*ranuni(&seed));
output;
end;
run;
proc sort data=xc1 out=xc2;
    by h primary_unit second_unit;
run;
data xc_population;
    set xc2;
    district=put(primary_unit,2.);
    site=catx('-',district,put(second_unit,2.));
    if h=1 then do;
        c3_real=233.3122+256.23*rannor(&seed);
        d1_real=rantbl(&seed,0.5188,0.4812);
        e4_real=rantbl(&seed,0.1153,0.1504,0.7343);
    end;
    else do;
        c3_real=189.4191+201.3559*rannor(&seed);
        d1_real=rantbl(&seed,0.6095,0.3950);
        e4_real=rantbl(&seed,0.1123,0.1318,0.7559);
```


http://dx.doi.org/10.21037/apm-20-2262

```
        end;
        drop primary_unit second_unit;
run;
proc sort data=xc_population;
        by h district site;
run;
%mend simulated_population;
%simulated_population (N=60000,
                                N1=10,
                                N2=12,
                                seed=1);
run;
```


http://dx.doi.org/10.21037/apm-20-2262