抽樣調查作業二

（分層抽樣）

指導教授：沈葆聖 老師

學生：王奎貿

```
dm'log;clear;output;clear;';
options nodate nocenter ls=78;
/* two-stage cluster sampling: the first stage:srs; the second stage:srs */
data one;
    do clusterid=1 to 10;
       do i=1 to 1+int(50*ranuni(12346));
          if i=1 then personid=1;
         else personid+1;
          output;
       end;
    end;
    drop i;
    proc print;
    proc sort;by clusterid;
proc means noprint;var personid;by clusterid;
output out=test n=n;
proc print;var clusterid n;
/*These statements select a simple random sample of three clusters without
replacement:*/
proc surveyselect data=one out=sample method=srs sampsize=4 seed=377183 noprint;
   samplingunit clusterid;
*proc print;
proc surveyselect data=sample out=sample2    method=srs n=6;
strata clusterid;
proc print;
/* stratified pps (wor)*/
 data travel;
        input id$    amount @@;
        datalines;
   110   237.18    002   567.89    234   118.50
   743    74.38    411 1287.23    782   258.10
   216   325.36    174   218.38    568 1670.80
   302   134.71    285 2020.70    314    47.80
   139 1183.45    775   330.54    425   780.10
   506   895.80    239   620.10    011   420.18
   672   979.66    142   810.25    738   670.85
   192   314.58    243    87.50    263 1893.40
   496   753.30    332   540.65    486 2580.35
   614   230.56    654   185.60    308   688.43
   784   505.14    017   205.48    162   650.42
   289 1348.34    691    30.50    545 2214.80
   517   940.35    382   217.85    024   142.90
   478   806.90    107   560.72
   ;
proc means noprint;var amount;
output out=test mean=truemean var = truevar;
```

U(1,51)

做10群每群從均勻分配隨機抽樣整數個

根據群數內的人,算人數並輸出報表

抽四群

抽群內6人

算母體平均數、變異數

```
proc print;var truemean truevar;
data gender1;seedg = 1234;
do i = 1 to 41;
u = ranuni(seedg);
if u > 0.5 then gender = 'female';
else gender = 'male';
output;
end;
proc print; var gender;
data travel2;set travel;
seed=2233;
y=0.03+0.02*ranuni(seed);
salary=(amount/y);
data travel2;set travel2;merge gender1;
drop amount i u seedg;
proc sort;by gender;
proc print;
 title1 'Travel Expense Audit';
    title2 'Stratified PPS (Dollar-Unit) Sampling';
    proc surveyselect data=travel2
            method=pps n=(9 6) MAXSIZE= 41
            seed=47268 out=sample;
        size salary;
        strata gender;
   title1 'Travel Expense Audit';
   title2 'Sample Selected by Stratified PPS Design';
    proc sort;by gender;
    proc print data=sample;
   proc means noprint data=travel2;by gender;var salary;
output out =total sum=tx mean=mx n=nh;
proc print;var gender tx mx nh;
data travel2;merge travel2 total;by gender;
if gender='male' then sn=6;
if gender='female' then sn=9;
prob=sn*salary/tx;
proc print;var gender id salary tx nh sn;
proc means noprint data=travel2;var salary;
output out=total2 n=tn;
data travel2;set travel2;
if _n_=1 then set total2;
proc print; var gender id salary tx tn    nh sn prob;
proc sort data=travel;by    id;
proc sort data=sample;by id;
data final(keep=id gender amount);merge travel sample;by id;
if gender=' ' then delete;
proc print;
```

產生資料檔gender 1 ,do迴圈產生gender變數，u為均勻分布。If u ＞0.5，gender = female, o.w.

產生資料檔，將travel覆蓋上去

製造1機率P

使得P x Salary = amount

產生資料檔，覆蓋並合併，把變數amount, I, u, seedy丟棄。照gender排序

最大樣本數

對男女兩層抽男6人、女9人的salary

輸出男女的sum of salary, mean of salary男個數、女個數（母體）

照gender合併資料男抽6人女抽9人

定義變數prob = 樣本數 x Salary / Total salary

因為要合併，所以產生資料，且覆蓋（多併一）

Travel 2 再多一次為總母體數

Travel 沒 gender 所以併 gender

將salary 母體，照id排並合併final，再照gender, Id排。假如gender為空白則delete資料。

```
proc sort data=travel2;by id;
data final2;merge travel2 final;by id;
if amount='.' then delete;   Travel 2 沒amount 所以
proc sort;by gender id;      併amount
proc print; var gender id salary tx tn nh sn prob amount;
-
data final2;set final2;
estimate=(1/nh)*amount/prob;
proc means noprint;var estimate;by gender;
output out=three mean=hmean sum = esum;
proc print;run;

data four;merge three final2;by gender;
w = nh/tn;
proc print;by gender;run;

data five;merge three final2;by gender;
westimate=(nh/tn)*hmean;
proc means noprint;var westimate;
output out=six sum=ht;
proc print; var ht;
run;
```

將母體資料與final依id做合併。假如amount為NA則delete。最後照id排。

$\frac{1}{N_h} \times \frac{y_i}{p_i}$

$mean = \frac{1}{n_h} \sum_{i=1}^{n_h} \frac{1}{N_h} \times \frac{y_i}{p_i}$

$sum = \sum_{i=1}^{n_h} \frac{1}{N_h} \times \frac{y_i}{p_i}$

∵有 mean

∵有 sum

計算男女兩層的 W_h

$\bar{y}_{HT,h}$ 報表出現

最後：男女兩層各有 $n_h^{(1)}$ $n_h^{(2)}$

$\therefore \underbrace{\frac{1}{n_h} \sum_{i=1}^{n_h} \frac{1}{N_h} \frac{y_i}{p_i}}_{mean} \cdot \underbrace{n_h}_{sum} = \sum_{i=1}^{n_h} \frac{1}{N_h} \frac{y_i}{p_i}$

goal

$$\bar{y}_{HT,h} = \frac{1}{N_h} \sum_{i=1}^{n_h} \frac{y_{hi}}{p_{hi}}$$

Travel Expense Audit
Sample Selected by Stratified PPS Design

Obs	clusterid	n
1	1	28
2	2	36
3	3	12
4	4	27
5	5	6
6	6	35
7	7	32
8	8	14
9	9	38
10	10	42

do迴圈生成的群組及每群下的人數

Travel Expense Audit
Sample Selected by Stratified PPS Design

Obs	clusterid	personid	SelectionProb	SamplingWeight
1	1	6	0.21429	4.66667
2	1	7	0.21429	4.66667
3	1	17	0.21429	4.66667
4	1	18	0.21429	4.66667
5	1	21	0.21429	4.66667
6	1	26	0.21429	4.66667
7	3	1	0.50000	2.00000
8	3	3	0.50000	2.00000
9	3	6	0.50000	2.00000
10	3	7	0.50000	2.00000
11	3	9	0.50000	2.00000
12	3	11	0.50000	2.00000
13	4	5	0.22222	4.50000
14	4	6	0.22222	4.50000
15	4	9	0.22222	4.50000
16	4	20	0.22222	4.50000
17	4	21	0.22222	4.50000
18	4	24	0.22222	4.50000
19	8	2	0.42857	2.33333
20	8	3	0.42857	2.33333
21	8	5	0.42857	2.33333
22	8	10	0.42857	2.33333
23	8	11	0.42857	2.33333
24	8	14	0.42857	2.33333

do迴圈生成資料抽樣結果

Travel Expense Audit

Sample Selected by Stratified PPS Design

Obs	truemean	truevar
1	696.335	399303.47

針对 amount 的平均、变方

Travel Expense Audit

Sample Selected by Stratified PPS Design

SURVEYSELECT 程序

選取方法	PPS (沒有取代)
大小量值	salary
最大大小量值	41
分層變數	gender

輸入資料集	TRAVEL2
亂數種子	47268
分層數目	2
總樣本大小	15
輸出資料集	SAMPLE

男女兩層

Travel Expense Audit

Sample Selected by Stratified PPS Design

Obs	gender	tx	mx	nh
1	female	227136.66	16224.05	14
2	male	481036.10	17816.15	27

Obs	gender	id	seed	y	salary	AdjustedSize	SelectionProb	SamplingWeight
1	female	691	2233	0.044231	689.57	41	0.64286	1.55556
2	female	024	2233	0.037620	3798.52	41	0.64286	1.55556
3	female	614	2233	0.035686	6460.77	41	0.64286	1.55556
4	female	775	2233	0.030463	10850.47	41	0.64286	1.55556
5	female	332	2233	0.047118	11474.29	41	0.64286	1.55556
6	female	784	2233	0.038627	13077.29	41	0.64286	1.55556
7	female	142	2233	0.045826	17681.10	41	0.64286	1.55556
8	female	289	2233	0.044501	30298.81	41	0.64286	1.55556
9	female	545	2233	0.043972	50368.05	41	0.64286	1.55556
10	male	782	2233	0.037926	6805.28	41	0.22222	4.50000
11	male	110	2233	0.030428	7794.70	41	0.22222	4.50000
12	male	216	2233	0.039450	8247.43	41	0.22222	4.50000
13	male	107	2233	0.047056	11916.08	41	0.22222	4.50000
14	male	162	2233	0.037999	17116.72	41	0.22222	4.50000
15	male	263	2233	0.037496	50496.36	41	0.22222	4.50000

Obs	gender	id	salary	tx	tn	nh	sn	prob
1	female	775	10850.47	227136.66	41	14	9	0.42994
2	female	425	25422.45	227136.66	41	14	9	1.00733
3	female	506	20987.55	227136.66	41	14	9	0.83161
4	female	142	17681.10	227136.66	41	14	9	0.70059
5	female	332	11474.29	227136.66	41	14	9	0.45465
6	female	614	6460.77	227136.66	41	14	9	0.25600
7	female	654	3859.35	227136.66	41	14	9	0.15292
8	female	784	13077.29	227136.66	41	14	9	0.51817
9	female	289	30298.81	227136.66	41	14	9	1.20055
10	female	691	689.57	227136.66	41	14	9	0.02732
11	female	545	50368.05	227136.66	41	14	9	1.99577
12	female	517	27693.95	227136.66	41	14	9	1.09734
13	female	382	4474.50	227136.66	41	14	9	0.17730
14	female	024	3798.52	227136.66	41	14	9	0.15051
15	male	110	7794.70	481036.10	41	27	6	0.09722
16	male	002	16623.42	481036.10	41	27	6	0.20735
17	male	234	2567.52	481036.10	41	27	6	0.03202
18	male	743	2140.83	481036.10	41	27	6	0.02670
19	male	411	26393.28	481036.10	41	27	6	0.32921
20	male	782	6805.28	481036.10	41	27	6	0.08488
21	male	216	8247.43	481036.10	41	27	6	0.10287
22	male	174	6177.24	481036.10	41	27	6	0.07705
23	male	568	49430.18	481036.10	41	27	6	0.61655
24	male	302	3142.86	481036.10	41	27	6	0.03920
25	male	285	49566.61	481036.10	41	27	6	0.61825
26	male	314	1037.83	481036.10	41	27	6	0.01294
27	male	139	27787.09	481036.10	41	27	6	0.34659
28	male	239	13828.40	481036.10	41	27	6	0.17248
29	male	011	9943.37	481036.10	41	27	6	0.12402
30	male	672	30312.43	481036.10	41	27	6	0.37809
31	male	738	14382.24	481036.10	41	27	6	0.17939
32	male	192	7143.98	481036.10	41	27	6	0.08911
33	male	243	2498.10	481036.10	41	27	6	0.03116
34	male	263	50496.36	481036.10	41	27	6	0.62984
35	male	496	19439.82	481036.10	41	27	6	0.24247
36	male	486	53929.04	481036.10	41	27	6	0.67266
37	male	308	14518.61	481036.10	41	27	6	0.18109
38	male	017	5885.38	481036.10	41	27	6	0.07341
39	male	162	17116.72	481036.10	41	27	6	0.21350
40	male	478	21911.30	481036.10	41	27	6	0.27330
41	male	107	11916.08	481036.10	41	27	6	0.14863

Travel Expense Audit

Sample Selected by Stratified PPS Design

Obs	gender	_TYPE_	_FREQ_	hmean	esum
1	female	0	9	73.718	663.465
2	male	0	6	114.001	684.008

Travel Expense Audit

Sample Selected by Stratified PPS Design

Obs	gender	id	salary	tx	tn	nh	sn	prob	amount
1	female	024	3798.52	227136.66	41	14	9	0.15051	142.90
2	female	142	17681.10	227136.66	41	14	9	0.70059	810.25
3	female	289	30298.81	227136.66	41	14	9	1.20055	1348.34
4	female	332	11474.29	227136.66	41	14	9	0.45465	540.65
5	female	545	50368.05	227136.66	41	14	9	1.99577	2214.80
6	female	614	6460.77	227136.66	41	14	9	0.25600	230.56
7	female	691	689.57	227136.66	41	14	9	0.02732	30.50
8	female	775	10850.47	227136.66	41	14	9	0.42994	330.54
9	female	784	13077.29	227136.66	41	14	9	0.51817	505.14
10	male	107	11916.08	481036.10	41	27	6	0.14863	560.72
11	male	110	7794.70	481036.10	41	27	6	0.09722	237.18
12	male	162	17116.72	481036.10	41	27	6	0.21350	650.42
13	male	216	8247.43	481036.10	41	27	6	0.10287	325.36
14	male	263	50496.36	481036.10	41	27	6	0.62984	1893.40
15	male	782	6805.28	481036.10	41	27	6	0.08488	258.10

gender=female

Obs	_TYPE_	_FREQ_	hmean	esum	id	seed	y	salary	tx	mx	nh	sn	prob	tn	amount	estimate	w
1	0	14	73.7183	663.465	024	2233	0.037620	3798.52	227136.66	16224.05	14	9	0.15051	41	142.90	67.8164	0.34146
2	0	14	73.7183	663.465	142	2233	0.045826	17681.10	227136.66	16224.05	14	9	0.70059	41	810.25	82.6088	0.34146
3	0	14	73.7183	663.465	289	2233	0.044501	30298.81	227136.66	16224.05	14	9	1.20055	41	1348.34	80.2215	0.34146
4	0	14	73.7183	663.465	332	2233	0.047118	11474.29	227136.66	16224.05	14	9	0.45465	41	540.65	84.9390	0.34146
5	0	14	73.7183	663.465	545	2233	0.043972	50368.05	227136.66	16224.05	14	9	1.99577	41	2214.80	79.2677	0.34146
6	0	14	73.7183	663.465	614	2233	0.035686	6460.77	227136.66	16224.05	14	9	0.25600	41	230.56	64.3304	0.34146
7	0	14	73.7183	663.465	691	2233	0.044231	689.57	227136.66	16224.05	14	9	0.02732	41	30.50	79.7334	0.34146
8	0	41	73.7183	663.465	775	2233	0.030463	10850.47	227136.66	16224.05	14	9	0.42994	41	330.54	54.9152	0.34146
9	0	14	73.7183	663.465	784	2233	0.038627	13077.29	227136.66	16224.05	14	9	0.51817	41	505.14	69.6323	0.34146

gender=male

Obs	_TYPE_	_FREQ_	hmean	esum	id	seed	y	salary	tx	mx	nh	sn	prob	tn	amount	estimate	w
10	0	27	114.001	684.008	107	2233	0.047056	11916.08	481036.10	17816.15	27	6	0.14863	41	560.72	139.725	0.65854
11	0	27	114.001	684.008	110	2233	0.030428	7794.70	481036.10	17816.15	27	6	0.09722	41	237.18	90.353	0.65854
12	0	27	114.001	684.008	162	2233	0.037999	17116.72	481036.10	17816.15	27	6	0.21350	41	650.42	112.833	0.65854
13	0	27	114.001	684.008	216	2233	0.039450	8247.43	481036.10	17816.15	27	6	0.10287	41	325.36	117.141	0.65854
14	0	27	114.001	684.008	263	2233	0.037496	50496.36	481036.10	17816.15	27	6	0.62984	41	1893.40	111.338	0.65854
15	0	27	114.001	684.008	782	2233	0.037926	6805.28	481036.10	17816.15	27	6	0.08488	41	258.10	112.617	0.65854

Travel Expense Audit

Sample Selected by Stratified PPS Design

Obs	ht
1	676.993