통계조사론입문 HW5

2019150445 신백록

텍스트이(가) 표시된 사진

자동 생성된 설명

2.

/\* 2 ratio\_est(srs)\*/

**DATA** cluster;

INFILE 'C:\cs1\_srs\_sample.dat';

INPUT cluster element y1 y2 y3;

**RUN**;

**DATA** cluster; SET cluster;

SamplingWeight\_1=**250**/**25**;

**RUN**;

**PROC** **SURVEYSELECT** DATA=cluster OUT=cluster\_2 METHOD=SRS SAMPSIZE=(**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**)

SEED=**2019150435**;

STRATA cluster;

**RUN**;

**PROC** **PRINT** DATA=cluster\_2;

**RUN**;

**DATA** cluster\_2; SET cluster\_2;

SamplingWeight\_2=SamplingWeight\_1\*SamplingWeight;

**RUN**;

**PROC** **SURVEYMEANS** DATA=cluster\_2 TOTAL=**250**;

CLUSTER cluster;

VAR y1 y2 y3;

WEIGHT SamplingWeight\_2;

**RUN**;

/\*2 ratio\_est,unb\_est(pps) (since K\_hat\_wr=K, y\_bar\_ratio=y\_bar\_unb) \*/

**DATA** cluster\_pps;

INFILE 'C:\cs1\_pps\_sample.dat';

INPUT cluster element y1 y2 y3 x qi selection\_prop1 sampling\_weight1;

**RUN**;

**proc** **sql**;

create table cluster\_pps1 as

select \*

from cluster\_pps

where qi=**2**;

**quit**;

**DATA** cluster\_pps2;

set cluster\_pps;

if cluster=**54** then cluster=**53**;

if cluster=**136** then cluster=**135**;

**run**;

**data** cluster\_pps3;

set cluster\_pps2 cluster\_pps1;

**run**;

**proc** **print** data=cluster\_pps3;

**run**;

**proc** **sort** data=cluster\_pps3;

by cluster;

**run**;

**PROC** **SURVEYSELECT** DATA=cluster\_pps3 OUT=cluster\_pps4 SAMPSIZE=(**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**)

SEED=**2019150445**;

STRATA cluster;

**RUN**;

**DATA** cluster\_pps4; SET cluster\_pps4;

SamplingWeight\_2=sampling\_weight1\*SamplingWeight;

**RUN**;

**PROC** **SURVEYMEANS** DATA=cluster\_pps4 TOTAL=**250**;

CLUSTER cluster;

CLASS y3;

VAR y1 y2 y3;

WEIGHT SamplingWeight\_2;

**RUN**;

/\*#2 unb\_est(srs) \*/

**Data** cluster\_unb;

INFILE 'C:\cs1\_srs\_sample.dat';

INPUT cluster element y1 y2 y3;

IF cluster=**1** THEN Mi=**47**;

IF cluster=**19** THEN Mi=**58**;

IF cluster=**21** THEN Mi=**64**;

IF cluster=**29** THEN Mi=**58**;

IF cluster=**50** THEN Mi=**75**;

IF cluster=**64** THEN Mi=**81**;

IF cluster=**72** THEN Mi=**92**;

IF cluster=**76** THEN Mi=**101**;

IF cluster=**78** THEN Mi=**109**;

IF cluster=**85** THEN Mi=**84**;

IF cluster=**92** THEN Mi=**114**;

IF cluster=**95** THEN Mi=**95**;

IF cluster=**99** THEN Mi=**94**;

IF cluster=**132** THEN Mi=**117**;

IF cluster=**134** THEN Mi=**129**;

IF cluster=**154** THEN Mi=**134**;

IF cluster=**176** THEN Mi=**151**;

IF cluster=**183** THEN Mi=**147**;

IF cluster=**190** THEN Mi=**147**;

IF cluster=**198** THEN Mi=**152**;

IF cluster=**215** THEN Mi=**141**;

IF cluster=**219** THEN Mi=**135**;

IF cluster=**225** THEN Mi=**169**;

IF cluster=**241** THEN Mi=**181**;

IF cluster=**243** THEN Mi=**165**;

K=**28462**;

N=**250**;

Mi\_y1=Mi\*y1;

Mi\_y2=Mi\*y2;

Mi\_y3=Mi\*y3;

weight=(**250**/**25**)\*(Mi/**20**);

**RUN**;

**PROC** **SURVEYSELECT** DATA=cluster\_unb OUT=cluster\_unb2 METHOD=SRS SAMPSIZE=(**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**,**20**)

SEED=**2019150435**;

STRATA cluster;

**RUN**;

**proc** **means** mean var data=cluster\_unb2;

ID Mi;

CLASS cluster;

VAR y1 Mi\_y1 y2 Mi\_y2 y3 Mi\_y3;

OUTPUT OUT=cluster\_unb3 MEAN= y1\_meani y1\_thati y2\_meani y2\_thati y3\_meani y3\_thati VAR=y1\_vari junk y2\_vari junk2 y3\_vari junk3;

**RUN**;

**data** cluster\_unb3; set cluster\_unb3;

IF cluster=**.** THEN DELETE;

small\_mi=**20**;

DROP \_type\_ \_freq\_ junk junk2 junk3;

**run**;

**proc** **means** sum var data=cluster\_unb3;

var y1\_thati y2\_thati y3\_thati;

output out=that\_sum\_var sum=y1\_thati\_sum y2\_thati\_sum y3\_thati\_sum var=y1\_that\_var y2\_that\_var y3\_that\_var;

**run**;

**DATA** mean\_unb; set that\_sum\_var;

y1\_mean\_unb=**1**/**28462**\***250**/**25**\*y1\_thati\_sum;

y2\_mean\_unb=**1**/**28462**\***250**/**25**\*y2\_thati\_sum;

y3\_mean\_unb=**1**/**28462**\***250**/**25**\*y3\_thati\_sum;

y1\_comp1=**250**\*\***2**\*(**1**-**25**/**250**)\*y1\_that\_var/**25**;

y2\_comp1=**250**\*\***2**\*(**1**-**25**/**250**)\*y2\_that\_var/**25**;

y3\_comp1=**250**\*\***2**\*(**1**-**25**/**250**)\*y3\_that\_var/**25**;

KEEP y1\_mean\_unb y2\_mean\_unb y3\_mean\_unb y1\_comp1 y2\_comp1 y3\_comp1;

**run**;

**data** cluster\_unb3; set cluster\_unb3;

comp2\_y1=(**1**-**20**/Mi)\*(Mi\*\***2**)\*y1\_vari/**20**;

comp2\_y2=(**1**-**20**/Mi)\*(Mi\*\***2**)\*y2\_vari/**20**;

comp2\_y3=(**1**-**20**/Mi)\*(Mi\*\***2**)\*y3\_vari/**20**;

**RUN**;

**proc** **print** data=cluster\_unb3;

**run**;

**proc** **means** mean data=cluster\_unb3;

VAR comp2\_y1 comp2\_y2 comp2\_y3;

output out=mean\_unb\_comp2 mean=y1\_comp2 y2\_comp2 y3\_comp2;

**RUN**;

**data** mean\_unb\_comp2; set mean\_unb\_comp2;

comp2\_y1=y1\_comp2\***250**;

comp2\_y2=y2\_comp2\***250**;

comp2\_y3=y3\_comp2\***250**;

DROP \_type\_ \_freq\_;

**Run**;

**data** mean\_unb\_srs;

merge mean\_unb mean\_unb\_comp2;

y1\_se\_unb=sqrt((y1\_comp1+comp2\_y1)/**28462**\*\***2**);

y2\_se\_unb=sqrt((y2\_comp1+comp2\_y2)/**28462**\*\***2**);

y3\_se\_unb=sqrt((y3\_comp1+comp2\_y3)/**28462**\*\***2**);

keep y1\_mean\_unb y2\_mean\_unb y3\_mean\_unb y1\_se\_unb y2\_se\_unb y3\_se\_unb;

**run**;

**proc** **print** data=mean\_unb\_srs;

**run**;

|  |  |  |
| --- | --- | --- |
|  | srs | pps |
| unb | Y1;100.479, 6.4614 | 99.1598, 0.43001 |
| Y2:6.7013, 0.45123 | 6.96238,0.307049 |
| Y3:0.57189,0.039587 | 0.618,0.024777 |
| ratio | Y1:100.6984,0.5508 | 99.1598, 0.43001 |
| Y2:6.715932,0.212810 | 6.96238,0.307049 |
| Y3:0.573134,0.019705 | 0.618,0.024777 |

3.

일반적으로 이단계 확률비례 집락추출법이 이단계 단순임의집락추출법보다 효율적이다. 각 집락의 모집단 총합인 ti가 집락의 크기인 Mi에 비례하는 성질을 이용하는 추출법이기 때문이다. 따라서 unbiased estimation, ratio estimation 모두 평균적으로 pps의 se가 srs의 se보다 작음을 볼 수 있다.