

자료분석노트

과제: 통계청 순환표본

2017.01.13. ver 8

수정내용

1. 2.5절, 3.2절, 3.3.1절에서 보조변수들이 추가됨
2. SAS 예시 3-1, 3-2의 옵션 구성이 변경됨
2. 수정내용은 파란색으로 표기됨

1. 자료이해

1.1 자료개요

1.1.1 프레임 자료

□ 프레임 파일 I (인충 조사구 프레임) 1)

- 파일 1
 - frame_0730.sas7bdat
 - 2015 표본설계 당시 사용한 전국 조사구 파일

□ 프레임 파일 II (등록-인충 프레임 연계 파일) 2)

- 파일 1
 - FRAME1016_all.csv
 - 등록 전수 프레임 (대전, 제주)
 - 대전/제주 전체 조사구+등록센서스 (가구, 인구, 주택) 연계 자료
- 파일 2
 - FRAME1016_ADM.csv
 - 등록 전수 프레임 (대전, 제주) 중 조사구 정보 없는 등록자료를 포함한 읍면동별 집계
 - 2015 인충 제주/대전 읍면동별 집계결과
- 파일 3
 - FRAME1016_SAM.csv
 - 표본 20% 인충 프레임 (전국)
 - 전국 전체 조사구 중 20% 표본조사구에 등록센서스 및 20% 표본조사 자료 연계 자료
- 파일 4
 - 등록센서스(제공항목).xlsx
 - 등록인충 연계자료 제공항목 코드북
 - 인구/가구/주택 별 설명

1) 16.10.7 김영원 교수님 이메일 확인; 10.06 박법민 주무관 송부

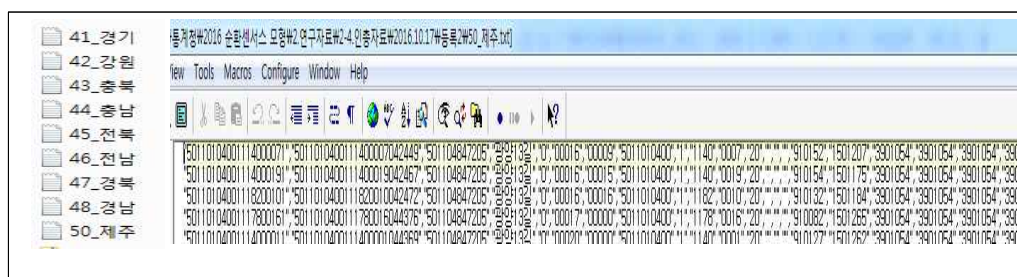
2) 2016.10.17.

2.연구자료 ▶ 2-4.인종자료 ▶ 2016.10.17 ▶ 1.등록인종20프레임1016 ▶	
이름	수정한 날짜
1.등록인종20프레임1016	2016-10-17 오후...
FRAME1016_ADM	2016-10-29 오후...
FRAME1016_all	2016-10-29 오후...
FRAME1016_SAM	2016-10-29 오후...
등록센서스(제공항목)	2016-10-29 오후...
등록예조사구 연결과정	2016-10-29 오후...
표본프레임분포및추가변수	2016-10-29 오후...

□ 프레임 파일 III (등록센서스 파일) 3)

● 파일 1

— 등록 text 파일



3) 2016.10.17.

1.1.2 표본조사 및 표본가중치 파일

□ 제주도 표본조사 파일 I 4)

● 파일 1

- 2015인총 인구 조사자료(제주도)_제공용.xlsx
- 2015인총 가구 조사자료(제주도)_제공용.xlsx
- 2015인총 주택 조사자료(제주도)_제공용.xlsx
- 2015년 제주도 인총 20% 표본조사 자료

● 파일 2

- 인구_가구_ 및_주택_읍면동_20161006133040.xlsx
- KOSIS 읍면동, 특광역시도별 추계자료

● 파일 3

- 일반조사표_컬럼명표기pdf
- 설문지 변수 코드 기재

현재 살고 있는 집의 주소를 적어 주십시오.						
시·도	시·군·구	읍·면·동	도로명 주소	공동주택명	동	호
		TS_NM	ROD_NM	AH_NM	DONG_CNT	DR_CNT

□ 제주도 표본가중치 파일 II

● 파일 15)

- 2015 인구주택총조사 제주도 가중치.xlsx
- 2015 인총 20% 표본설계 제주도 가중치
- 인구/가구/주택 별 탭 제공되고 ‘읍면동 가중치’ (EMD_WGT2) 사용

4) 16.10.7 김영원 교수님 이메일 확인; 10.06 박법민 주무관 송부

5) 2016.10.13. 이메일

	A	B	C	D	E	F	G	H	I	J
1		시도	시군구	읍면동	조사구	거처번호	가구번호	가구원수	시군구가중치	읍면동가중치
2		CP_CD	CDW_CD	TTB_CD	ED_NUM1	HU_NUM1	HHD_NUM	HM_SN	SGG_WGT2	EMD_WGT2
3	1	39	010	66	0095	0014	001	4	5	5
4	2	39	010	66	0095	0013	001	2	5	5
5	3	39	010	66	0095	0018	001	3	5	6
6	4	39	010	66	0102	0002	001	2	5	5

□ 경기·대전 인총 20% 표본조사 파일 II

● 파일 16)

- 1대전자료및레이아웃.zip
- 2hhd31_가구.zip
- 3HM_31_2_인구.zip
- 인총20%대전,경기 자료

1.1.3 표본설계 정보자료

□ 표본설계 정보자료

● 파일 17)

- 2.2015 인구주택총조사 표본규모_0806(최종).xlsx
- 표본규모(추출률 포함) 및 예상 CV 기술 파일
- 시군구/읍면동 별로 따로 제공

● 파일 28)

- 2015 인구주택총조사 표본설계 결과.hwp
- 2015 인총 20% 표본설계안

6) 2016.11.12. 이메일/최종 내검 안 끝난 자료이고 제주 보다 제공 칼럼이 줄고 경기도 용량이 너무 커서 줄였다고 밝힘

7) 2016.10.17. 이메일

8) 2016.09.29. 이메일

1.1.4 제주도 시험조사 자료

□ 순환센서스 제2차 시험조사 표본설계

- 표본설계안⁹⁾

- 순환센서스_제2차_시험조사_표본추출_결과.hwp

□ 제주도 조사구 병합 구축 자료

- 파일 1

- 제주_병합조사구리스트.xlsx
 - 2016년도 시험조사를 위한 조사구 병합 자료 결과물

현황(제주)		개별조사구리스트_제주					비고분류코드				테이블정의서			
시도	시군구	읍면동	읍면동명	조사구 수량				합계				성명코드		
				계	보통(1)	성(2)	아파트(A)	계	1개	2개	3개	계	1개	2개
1	2	43	43	8,027	5,733	64	2,230	4,215	902	2,814	499	2,852	345	2,133
		3901011	한림읍	246	227	2	17	124	20	86	18	112	15	79
SIDO_CD	SIGUNGU_CD	ADM_CD	ADM_NM	IND_CD	IND_TXT									
시도코드	시군구코드	읍면동코드	읍면동명	개별조사구대표	개별조사구전체									
39	39010	3901011	한림읍	001	001-1									
39	39010	3901011	한림읍	002	002-1									





□ 제주도 시험조사¹⁰⁾

- 제주도 2016년 7월분(1607월분)
- 제주도 2차 시험조사 1(15.10~16.8)
- 제주도 2차 시험조사 2(15.10~16.8)
- 2016년 11월 08일 최종 (16.9월분)




9) 2016.11.08. 이메일

10) 2016.9.13. 이메일


연구자료 ▶ 2-3.시험조사 ▶ 시험조사_1607월분_제공용 ▶

이름	수정한 날짜
 1. 주택(2016년 7월분)	2016-09-13
 2. 가구(2016년 7월분)	2016-09-13
 3. 가구원(2016년 7월분)	2016-09-13
 시험조사_1607월분_제공용	2016-09-13




구자료 ▶ 2-3.시험조사 ▶ 제주도 2차 시험조사2 ▶

이름	수정일
 1. 주택(15.10~16.8)	2016-
 2. 가구(15.10~16.8)	2016-
 3. 가구원(15.10~16.8)	2016-

연구자료 ▶ 2-3.시험조사 ▶ 2016.11.08 ▶ 최종 ▶

이름	수정한 날짜
 1. 주택(16.9)	2016-11-(
 2. 가구(16.9)	2016-11-(
 3. 가구원(16.9)	2016-11-(

구자료 ▶ 2-3.시험조사 ▶ 제주도 2차 시험조사1 ▶

이름
 1. 주택(TN_SC_POPL_HOUS)_입력자료(15.10~16.8)
 2. 가구(TN_SC_POPL_HHD)_입력자료(15.10~16.8)
 3. 가구원(TN_SC_POPL_HM)_입력자료(15.10~16.8)

1.2 세부자료 및 관련 기술

1.2.1 인충 조사구 추출틀 (frame_0730.sas7bdat)

□ 변수 및 자료

- 변수 코드: 확인 필요
- 13개 변수에 694937 cases

변수와 속성 리스트(오름차순)

#	변수	유형	길이	출력형식	입력형식	레이블
11	apt	숫자	8			apt
8	char	문자	2	\$2.	\$2.	char
10	dandok	숫자	8			dandok
6	emd	문자	22	\$22.	\$22.	emd
5	emd_cd	문자	14	\$14.	\$14.	emd_cd
13	etc	숫자	8			etc
9	gagu	숫자	8			gagu
7	jsg	문자	6	\$6.	\$6.	jsg
4	sgg	문자	8	\$8.	\$8.	sgg
3	sgg_cd	문자	10	\$10.	\$10.	sgg_cd
2	sido	문자	14	\$14.	\$14.	sido
1	sido_cd	문자	4	\$4.	\$4.	sido_cd
12	yun	숫자	8			yun

SAS 시스템

OBS	sido_cd	sido	sgg_cd	sgg	emd_cd	emd	jsg	char	gagu	dandok	apt	yun	etc
1	11	서울특별시	11010	종로구	1101053	사직동	001	1	29	7	0	22	0
2	11	서울특별시	11010	종로구	1101053	사직동	002	1	28	7	0	21	0
3	11	서울특별시	11010	종로구	1101053	사직동	003	1	37	15	0	22	0
4	11	서울특별시	11010	종로구	1101053	사직동	004	1	27	12	0	15	0
5	11	서울특별시	11010	종로구	1101053	사직동	005	4	1	0	0	0	1
6	11	서울특별시	11010	종로구	1101053	사직동	006	1	26	17	0	8	1
7	11	서울특별시	11010	종로구	1101053	사직동	007	1	22	16	0	0	6
8	11	서울특별시	11010	종로구	1101053	사직동	008	1	40	26	0	12	2
9	11	서울특별시	11010	종로구	1101053	사직동	009	1	30	30	0	0	0
10	11	서울특별시	11010	종로구	1101053	사직동	010	1	21	16	0	0	5

1.2.2 등록 인종 추출틀 연계자료 (FRAME1016_all.csv)

□ 변수 및 자료

data명: 등록인종20프레임1016-FRAME1016_all	
대전/제주 전체조사구+등록센서스(가구, 인구, 주택) 연계 자료	
sido_cd	시도코드
sido	시도
sgg_cd	시군구코드
sgg	시군구
emd_cd	읍면동코드
emd	읍면동
jsg	조사구
char	특성번호
gagu07	가구수
dandok07	단독
apt07	아파트
yun07	연립,다세대
etc07	기타주택
key	key

A	B	
1	data명: 등록인종20프레임1016-FRAME1016_all	
2	대전/제주 전체조사구+등록센서스(가구, 인구, 주택) 연계 자료	
18	gagu	가구수
19	gaguwon1-gaguwon5	가구원수
20	gaguwon1	1인가구
21	gaguwon2	2인가구
22	gaguwon3	3인가구
23	gaguwon4	4인가구
24	gaguwon5	5인가구

A	B	
1	data명: 등록인종20프레임1016-FRAME1016_all	
2	대전/제주 전체조사구+등록센서스(가구, 인구, 주택) 연계 자료	
66	HM_CNT	가구원수
67	FMLY_CNT	가족수
68	ingu	인구수
69	male	남
70	female	녀
71		10세간격인구
72	a0	0~14세
73	a15	15~19
74	a20	20~29
75	a30	30대
76	a40	40대
77	a50	50대
78	a60	60세이상
79	a65	65세이상
80	ma0	0~14세 남
81	ma15	15~19 남
82	ma20	20~29 남
83	ma30	30대 남
84	ma40	40대 남
85	ma50	50대 남
86	ma60	60세이상 남
87	ma65	65세이상 남
88	fa0	0~14세 여
89	fa15	15~19 여
90	fa20	20~29 여
91	fa30	30대 여
92	fa40	40대 여
93	fa50	50대 여
94	fa60	60세이상 여
95	fa65	65세이상 여

A	B	
1	data명: 등록인종20프레임1016-FRAME1016_all	
2	대전/제주 전체조사구+등록센서스(가구, 인구, 주택) 연계 자료	
108		형제자매의자녀,그배우자
109		부모의형제자매,그배우자
110		기타친인척
111	hrel8	그외같이사는사람
112	hrel9	
113	meanage	평균연령
114	house	주택
115	dwelling	거처
116	dandok	단독
117	dandok1	일반단독
118	dandok2	다가구단독
119	dandok3	영업겸용단독
120	apt	아파트
121	miniapt	소형아파트
122	midapt	중형아파트
123	largeapt1	대형아파트
124	largeapt2	대대형아파트
125	yun	연립,다세대
126	etc_h	기타주택
127	officetel	오피스텔
128	nohouse	주택외거처

1.2.3 2015 인총20%표본조사 추계결과 (표본추정 결과) (FRAME1016_ADM.csv)

□ 변수 및 자료

FRAME1016_ADM.csv에서

○ 가구기준-가구원수

data명: 등록인총20프레임1016-FRAME1016_ADM	
2015인총 제주/대전 읍면동별 집계결과	
변수명	변수설명
ADM_CD	
FREQ	
gagu	가구수
gaguwon1-gaguwon5	가구원수
gaguwon1	1인가구
gaguwon2	2인가구
gaguwon3	3인가구
gaguwon4	4인가구
gaguwon5	5인가구

○ 가구기준-거처수: Q. 주택수 인지 가구수 인지 확인 필요 (to 이주희 사무관)

1	data명: 등록인총20프레임1016-FRAME1016_ADM	
2	2015인총 제주/대전 읍면동별 집계결과	
3	변수명	변수설명
104	house	주택
105	dwelling	거처
106	dandok	단독
107	dandok1	일반단독
108	dandok2	다가구단독
109	dandok3	영업겸용단독
110	apt	아파트
111	miniapt	소형아파트
112	midapt	중형아파트
113	largeapt1	대형아파트
114	largeapt2	대대형아파트
115	yun	연립,다세대
116	etc_h	기타주택
117	officetel	오피스텔
118	nohouse	주택외거처

FRAME1016_ADM.csv에서

○ 가구원 기준-성x연령층별 가구원수

1	data명: 등록인종20프레임1016-FRAME1016_ADM	
2	2015인종 제주/대전 읍면동별 집계결과	
3	변수명	변수설명
56	ingu	인구수
57	male	남
58	female	녀
59	10세간격인구	
60	a0	0~14세
61	a15	15~19
62	a20	20~29
63	a30	30대
64	a40	40대
65	a50	50대
66	a60	60세 이상
67	a65	65세 이상
68	ma0	0~14세 남
69	ma15	15~19 남
70	ma20	20~29 남
71	ma30	30대 남
72	ma40	40대 남
73	ma50	50대 남
74	ma60	60세 이상 남
75	ma65	65세 이상 남
76	fa0	0~14세 여
77	fa15	15~19 여
78	fa20	20~29 여
79	fa30	30대 여
80	fa40	40대 여
81	fa50	50대 여
82	fa60	60세 이상 여
83	fa65	65세 이상 여

2. 기본점검 및 기본 가중치 산출

2.1 제주도 시험조사 표본설계

2.1.1 2차 시험조사 제주도 지역 표본설계

□ 조사개요

- 등록센서스로 매년 모수가 작성되므로, 표본부분도 매월 조사하는 순환표본조사로 변경하여 자료의 시의성 확보를 위함 (표본과, 2015.9.23)
- 일정지역을 여러 시점에 걸쳐 중복되지 않게 반복조사하며 조사된 결과를 일정 기간 (또는 규모) 누적함
- 제2차 시험조사는 제주 지역을 대상으로 하며 2015.10~2017.9까지 총 24개월간 진행
- 제주도를 선정한 이유는 타 통계와 비교 가능한 최소 지역으로 고려
- 6개월을 1년으로 간주하여 4년을 실시하는 (가상적) 개념
- 2015년 11월은 인총조사와 중복되는 표본조사구로 대체

□ 모집단 분석

- 제주도의 8,207개 조사구 내 모든 거처, 가구 및 가구원
- 2015 인총을 위해 설정한 조사구 중 3가지 특성인 A/1/2 형태의 조사구 및 조사구 내 주택/가구/인구로 구성

〈표 2-1〉 제주도 조사구 모집단 현황

구분	조사구특성							
	계	일반조사구				시설조사구		
		소계	A	1	2	소계	3	4
조사구수	8,207	8,027	2,230	5,733	64	178	77	101
가구수	229,761	229,583	60,089	167,658	1,836	178	77	101

출처: 표본과(2015.9.23.) 제2차 시험조사 표본추출 결과

□ 표본설계

- 24개월간 동일한 규모의 표본이 되도록 약 20%에 가까운 1,584개 조사구
- 읍면동, 조사구 특성 (A, 기타(1,2))로 층화. 단, 추자면은 섬(2) 조사구로 조사여건 상 대상에서 제외시킴
- 층별로 조사구 특성, 단독 비율, 조사구 번호 순으로 정렬 후 계통추출 (2015 인총과 동일)

— 월별 배분 결과(표본과, 2015.9.23)

<표 2-2> 제주도 2차 시험조사 그룹 및 차수별 표본조사구 현황¹¹⁾

그룹 (6월단위)	차수 (월단위)	연월	전체	특성별		시군구별	
				아파트	기타	제주시	서귀포시
합계			1,584	386	1,198	1,060	524
A	1	'15.10	66	16	50	45	21
	2	11	66	17	49	44	22
	3	12	66	16	50	44	22
	4	'16. 1	66	16	50	44	22
	5	2	6566	16	4950	44	2122
	6	3	66	16	50	44	22
B	7	4	66	16	50	44	22
	8	5	66	16	50	44	22
	9	6	66	16	50	44	22
	10	7	66	16	50	44	22
	11	8	66	16	50	44	22
	12	9	66	16	50	45	21
C	13	10	66	16	50	45	21
	14	11	66	17	49	44	22
	15	12	66	16	50	44	22
	16	'17. 1	66	16	50	44	22
	17	2	66	16	50	44	22
	18	3	66	16	50	44	22
D	19	4	66	16	50	44	22
	20	5	66	16	50	44	22
	21	6	66	16	50	44	22
	22	7	66	16	50	44	22
	23	8	66	16	50	44	22
	24	9	66	16	50	45	21

☞ 시험조사 표본설계 내역 <표 2-3>에 대한 검증 필요 !

11) 서귀포시 기타조사구 수는 21개로 확인됨. 따라서 제주도 전체 기타조사구 수도 49개이고 해당 월 조사구 수도 65개임 (수진 메일, 2016.12.27.)

〈표 2-3〉 제주도 2차 시험조사 읍면동별 표본조사구 현황

구 분	합 계	A						B						C						D					
		1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
제주	1,584	66	66	66	66	66	66	66	66	66	66	66	66	66	66	66	66	66	66	66	66	66	66	66	66
제주시	1,060	45	44	44	44	44	44	44	44	44	44	44	45	45	44	44	44	44	44	44	44	44	44	44	45
한림읍	40	2	1	2	1	2	2	2	1	2	1	2	2	2	1	2	1	2	2	2	1	2	1	2	2
애월읍	52	2	3	2	2	2	2	2	3	2	2	2	2	2	3	2	2	2	2	2	3	2	2	2	2
구좌읍	34	2	1	2	1	1	2	1	1	2	1	1	2	2	1	2	1	1	2	1	1	2	1	1	2
조천읍	40	1	2	2	2	1	2	1	2	2	2	1	2	1	2	2	2	1	2	1	2	2	2	1	2
한경면	30	1	1	2	2	1	1	1	1	2	1	1	1	1	1	2	2	1	1	1	1	2	1	1	1
우도면	18		1	1		1	1		1	1	1	1	1		1	1		1	1		1	1	1	1	1
일도1동	26	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1
일도2동	64	2	3	2	3	3	3	2	3	2	3	3	3	2	3	2	3	3	3	2	3	2	3	3	3
이도1동	30	2	1	1	1	2	1	1	1	1	1	2	1	2	1	1	1	2	1	1	1	1	1	2	1
이도2동	92	4	4	4	4	3	4	4	4	4	4	3	4	4	4	4	4	3	4	4	4	4	4	3	4
삼도1동	36	2	1	1	2	2	1	2	1	1	2	2	1	2	1	1	2	2	1	2	1	1	2	2	1
삼도2동	32	1	1	1	2	2	1	1	2	1	1	2	1	1	1	1	2	2	1	1	2	1	2	1	1
용담1동	30	2	1	1	1	1	1	2	1	2	1	1	1	2	1	1	1	1	1	2	1	2	1	1	1
용담2동	36	2	2	1	1	2	1	2	2	1	1	2	1	2	2	1	1	2	1	2	2	1	1	2	1
건입동	32	1	2	2	1	1	2	1	1	1	1	2	1	1	2	2	1	1	2	1	1	1	1	2	1
화북동	48	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
삼양동	36	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1
봉개동	24	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
아라동	44	2	2	2	2	1	2	2	2	2	2	1	2	2	2	2	2	1	2	2	2	2	2	1	2
오라동	30	1	2	1	1	1	1	1	2	1	1	1	2	1	2	1	1	1	1	1	2	1	1	1	2
연동	96	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
노형동	100	4	4	4	5	4	4	4	4	4	5	4	4	4	4	4	5	4	4	4	4	5	4	4	4
외도동	40	2	1	1	2	2	2	2	1	1	2	2	2	2	1	1	2	2	2	2	1	1	2	2	2
이호동	26	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	2
도두동	24	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
서귀포시	524	21	22	22	22	22	22	22	22	22	22	22	21	21	22	22	22	22	22	22	22	22	22	22	21
대정읍	36	2	1	2	2	1	1	2	1	2	2	1	1	2	1	2	2	1	1	2	1	2	2	1	1
남원읍	36	1	2	1	1	2	2	1	2	1	1	2	2	1	2	1	1	2	2	1	2	1	1	2	2
성산읍	34	1	1	2	1	1	2	2	1	2	1	1	2	1	1	2	1	1	2	2	1	2	1	1	2
안덕면	30	1	2	1	1	1	1	1	2	1	1	2	1	1	2	1	1	1	1	1	2	1	1	2	1
표선면	32	2	1	1	1	1	2	2	1	1	2	1	1	2	1	1	1	1	2	2	1	1	2	1	1
송산동	28	1	1	1	2	1	1	1	1	1	2	1	1	1	1	1	2	1	1	1	1	1	2	1	1
정방동	26	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	2	1
중앙동	28	1	1	1	1	2	1	1	1	1	1	2	1	1	1	1	1	2	1	1	1	1	1	2	1
천지동	26	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1
효돈동	26	1	2	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1
영천동	26	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1
동홍동	48	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
서홍동	30	1	2	2	1	1	1	1	2	1	1	1	1	1	2	2	1	1	1	1	2	1	1	1	1
대륜동	30	1	1	1	2	2	1	2	1	1	1	1	1	1	1	1	2	2	1	2	1	1	1	1	1
대천동	30	1	1	1	1	1	2	1	2	1	1	1	2	1	1	1	1	1	2	1	2	1	1	1	2
중문동	32	2	1	1	1	2	1	1	1	2	2	1	1	2	1	1	1	1	2	1	1	2	2	1	1
예래동	26	1	1	2	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1

2.2 제주도 표본추출틀

2.2.1 등록인총20프레임 파일 (FRAME1016_all.csv)

□ 등록인총20프레임 파일

- 파일명: FRAME1016_all.csv
- 조사구 ID: 두 변수 emd_cd 와 jsg 결합한 KEY 변수 사용
- 조사구 종류: char로 A=아파트, 1/2=기타
- 표본층:

emd_cd
ed_type = 1 기타 (char =1 or 2)
 = 2 아파트

- PSU: 조사구 jsg
- PSU 규모: gagu (Q. 추후 확인 필요)
- PSU 특성: 여러 변수*

* 등록센서스(제공항목).xlsx 참조 (§ 1.1.1 절 설명 참고)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	sido_cd	sido	sgg_cd	sgg	emd_cd	emd	jsg	char	gagu07	dandok07	apt07	yun07	etc07	key	FREQ	gagu
20888	39	제주특별자치도	39010	제주시	3901011	한림읍	1	1	25	21	0	0	4	3901011001	38	
20889	39	제주특별자치도	39010	제주시	3901011	한림읍	2	1	26	23	0	0	3	3901011002	19	
20890	39	제주특별자치도	39010	제주시	3901011	한림읍	3	1	21	20	0	0	1	3901011003	20	
20891	39	제주특별자치도	39010	제주시	3901011	한림읍	4	1	38	35	0	0	3	3901011004	35	
20892	39	제주특별자치도	39010	제주시	3901011	한림읍	5	1	32	28	0	0	4	3901011005	32	
20893	39	제주특별자치도	39010	제주시	3901011	한림읍	6	3	1	0	0	0	1	3901011006	1	
20894	39	제주특별자치도	39010	제주시	3901011	한림읍	7	1	27	26	0	0	1	3901011007	29	
20895	39	제주특별자치도	39010	제주시	3901011	한림읍	8	1	31	31	0	0	0	3901011008	30	
20896	39	제주특별자치도	39010	제주시	3901011	한림읍	9	1	38	36	0	0	2	3901011009	36	
20897	39	제주특별자치도	39010	제주시	3901011	한림읍	10	A	32	0	32	0	0	3901011010	33	
20898	39	제주특별자치도	39010	제주시	3901011	한림읍	11	A	32	0	32	0	0	3901011011	32	

□ 가구기준 자료 점검

- SAS로 읽고 자료개수, 변수명 등 점검
- 조사구가 일차추출단위(PSU)가 됨
- (표본층*조사구)
 - (1) 조사구는 시도코드+시군구코드+읍면동코드+조사구번호로 묶어 확인
 - (2) 표본층은 읍면동코드와 조사구특성(아파트(A), 기타(1,2))으로 묶어 확인
- 표본층(h) 별 표본층 내 모집단 조사구 수 N_h 를 확인할 것. 표본추출틀은 조사구 단위로 구성되어 있는지 확인할 것. 또한 다음의 통계표의 결과와 일치

하는지 확인할 것

□ 읍면동별 (모집단) 가구 · 가구원 규모 확인

- 읍면동별 모집단 내 가구 및 가구원 규모를 확인하여 표로 만들 것
- 추가적으로 분할표를 구성하되 행은 조사구 수, 열은 가구 수
이때 조사구 범위: [1~5], [6~20], [21,50], [51,100], [101,200], [201,∞]
가구수 범위: (0,20000), [20000,65000], {65000,∞}

2.2.2 등록인총20프레임 파일 (FRAME1016_sam.csv)

□ 필요시 “_all” 파일과 비교

<표 2-4> 제주도 2015 인종표본조사 표본추출 규모

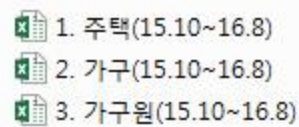
시군구코드	읍면동	읍면동코드	전체 조사구			표본 조사구			추출률		
			합계	아파트	기타	합계	아파트	기타	합계	아파트	기타
제주시	3901011	한림읍	246	17	229	45	4	41	18.29	23.53	17.90
제주시	3901012	애월읍	370	33	337	75	7	68			
제주시	3901013	구좌읍	178	0	178	37	0	37	20.79	-	20.79
제주시	3901014	조천읍	248	11	237	45	2	43	18.15	18.18	18.14
제주시	3901031	한경면	108	0	108	31	0	31	28.70	-	28.70
제주시	3901032	추자면	34	0	34	26	0	26	26.00	-	76.47
제주시	3901033	우도면	22	0	22	17	0	17	17.00	-	77.27
제주시	3901051	일도1동	53	6	47	29	4	25	29.00	66.67	53.19
제주시	3901052	일도2동	450	174	276	89	35	54	19.78	20.11	19.57
제주시	3901053	이도1동	113	41	72	33	12	21	29.20	29.27	29.17
제주시	3901054	이도2동	686	209	477	131	40	91	19.10	19.14	19.08
제주시	3901055	삼도1동	204	44	160	41	9	32	20.10	20.45	20.00
제주시	3901056	삼도2동	135	23	112	34	6	28	25.19	26.09	25.00
제주시	3901057	용담1동	115	10	105	33	3	30	28.70	30.00	28.57
제주시	3901058	용담2동	207	40	167	41	8	33	19.81	20.00	19.76
제주시	3901059	건입동	144	41	103	35	10	25	24.31	24.39	24.27
제주시	3901060	화북동	312	185	127	52	30	22	16.67	16.22	17.32
제주시	3901061	삼양동	201	118	83	40	23	17	19.90	19.49	20.48
제주시	3901062	봉개동	32	1	31	24	1	23	24.00	100.00	74.19
제주시	3901063	아라동	287	93	194	50	17	33	17.42	18.28	17.01
제주시	3901064	오라동	117	7	110	33	2	31	28.21	28.57	28.18
제주시	3901065	연동	711	252	459	135	48	87	18.99	19.05	18.95
제주시	3901066	노형동	755	374	381	142	71	71	18.81	18.98	18.64
제주시	3901067	외도동	235	145	90	43	26	17	18.30	17.93	18.89
제주시	3901068	이호동	49	2	47	29	2	27	29.00	100.00	57.45
제주시	3901069	도두동	32	0	32	24	0	24	24.00	-	75.00
서귀포시	3902011	대정읍	208	19	189	37	4	33	17.79	21.05	17.46
서귀포시	3902012	남원읍	204	9	195	37	2	35	18.14	22.22	17.95
서귀포시	3902013	성산읍	176	10	166	34	2	32	19.32	20.00	19.28
서귀포시	3902031	안덕면	116	0	116	30	0	30	25.86	-	25.86
서귀포시	3902032	표선면	135	5	130	31	2	29	22.96	40.00	22.31
서귀포시	3902051	송산동	60	12	48	27	6	21	27.00	50.00	43.75
서귀포시	3902052	정방동	43	11	32	27	7	20	27.00	63.64	62.50
서귀포시	3902053	중양동	65	4	61	27	2	25	27.00	50.00	40.98
서귀포시	3902054	천지동	55	10	45	27	5	22	27.00	50.00	48.89
서귀포시	3902055	효돈동	53	1	52	27	1	26	27.00	100.00	50.00
서귀포시	3902056	영천동	55	4	51	27	2	25	27.00	50.00	49.02
서귀포시	3902057	동홍동	318	212	106	49	32	17	15.41	15.09	16.04
서귀포시	3902058	서홍동	114	41	73	30	11	19	26.32	26.83	26.03
서귀포시	3902059	대륜동	119	19	100	30	5	25	25.21	26.32	25.00
서귀포시	3902060	대천동	90	12	78	28	4	24	28.00	33.33	30.77
서귀포시	3902061	중문동	131	35	96	31	9	22	23.66	25.71	22.92
서귀포시	3902062	예래동	41	0	41	27	0	27	27.00	-	65.85
제주시			6044	1826	4218	1314	360	954	21.74	19.72	22.62
서귀포시			1983	404	1579	526	94	432	26.53	23.27	27.36
제주도			8027	2230	5797	1840	454	1386	22.92	20.36	23.91

2.3 제주도 시험조사 자료

2.3.1 개요

□ 2차 시험조사 2 파일 사용

- 파일: “제주도 2차 시험조사2 엑셀” 사용.
참고로 김영원 교수님의 추가 설명을 고려할 것 ¹²⁾
- 3종류의 파일을 사용할 것. 1) 주택, 2) 가구, 3) 가구원



2.3.2 자료점검

□ 가구기준자료 기본점검

- SAS로 읽고 자료개수, 변수명 등 점검
- 조사구가 일차추출단위(PSU)가 됨
- (표본층*조사구)가 조사 월별로 서로 중복되지 않는지 확인할 것.¹³⁾ 이때
 - (1) 조사구는 시도코드+시군구코드+읍면동코드+조사구번호로 묶어 확인
 - (2) 표본층은 읍면동코드와 조사구특성(아파트(A), 기타(1,2))으로 묶어 확인
(h:1=일반조사구, 2=아파트)
- 표본층(h) 별 조사월 t별 표본조사구 수 n_{th} 를 확인할 것. 이때 표본조사구 수를 셀 때는 조사구 내 가구수와 관계없이 유일한 조사구 수로 확인할 것. 예로, proc sort로 확인한다면 nodupkey 옵션을 사용할 수 있음. § 2.1.1의 표들로 통계가 맞는지 확인할 것¹⁴⁾

12) 김영원 교수 이메일 (2016.10.17.) “제주도 시험조사 2015.10~2016.08 자료 참고하십시오. 압축파일이 두개인데 파일1과 파일2의 차이는... (담당 주무관이 차이를 몰라서 제가 비교해 본 결과...) 파일1이 원자료 파일인데 일부 읍면동의 경우 2016년 9월 자료도 붙어 있는게 있어서 이 부분 제외하고, 개인정보보호를 위해 거쳐번호를 다시 부여(랜덤조사구번호1자리+기존거쳐번호2자리) 해서 만든 것이 파일2로 보입니다. (** 아마도 파일2만 보내면 될 것을 박주무관이 잘 몰라서 두개 다 보낸 것을 추측됨) 따라서 작업은 파일2를 갖고 하시면 되고, 혹시 거쳐번호가 헷갈리면 파일1의 해당 조사구내 거쳐번호 참고하시면 됩니다.”

13) 15년10월~16년8월의 시험조사자료에서 중복되어 조사된 조사구는 없는 것으로 확인됨 (수진메일, 2016.12.27.)

14) 2016년2월 시험조사(A/5)에 차이가 존재. 서귀포 기타 조사구 수는 22가 아닌 21개 (수진메일, 2016.12.27.)

2.3.3 월별 기준 점검 및 기본 가중치 산출

□ 월별 기준 개요

- 조사 기준 월별 그룹 및 차수 구분 정보는 <표 2-2>에 요약됨. 시험조사 그룹 및 차수별 표본조사구 계획과 실제 현황이 조금 다른 것으로 확인됨 (§ 2.3.2 참고)

□ 가중치 산출을 위한 사전작업, 연구대상 한정, 결측 현황 파악

- 시험조사에 대한 개체수준별 정보파일, 즉 주택, 가구, 인구의 3종류 파일간의 비교를 통해 다음의 사실을 확인함¹⁵⁾¹⁶⁾

- 인구파일에는 있고 가구/주택 파일에 없는 경우 - 총 10 명¹⁷⁾
- 인구파일에는 없고 가구/주택 파일에 있는 경우 - 총 861가구
- 두 파일에 동시에 존재하는 가구원은 총 40965명

- 향후 가중치 연구를 위한 작업 단순화를 위해 다음의 조건을 만족하는 가구/가구원 만을 대상으로 제한

- 조건 1: 인구/가구/주택 파일 모두에 존재하는 개인 (명)
- 조건 2: 가구 내 가구원 모두 성별, 연령을 응답한 가구에 속한 개인 (명)
- 조건 3: 가구 주택유형 문항을 응답한 가구 (개)

- 시험조사를 위한 기본가중치 산출은 조건 1과 조건 2를 만족하는 자료에 대해서만 수행할 것

15) 수진메일, 엑셀파일로 정리 (2017.01.02.).

16) 수진메일 (2017.01.06.) “저번에 가구&주택파일에 조사시점이 2월,3월로 나누어져있었던 조사구가 4개라서 그것만 인구파일의 조사시점을 따르게 설정하겠다고 했었는데, 그 외에 추가로 3개의 조사구도 인구파일에서는 조사월이 16년 3월로 가구&주택파일에서는 16년 2월로 되어있어 그냥 인구파일의 조사시점을 따르게 merge했습니다.” 즉, 가구&주택 파일내에 2월과 3월 조사로 분리되어 있던 조사구에 대해 인구파일 기준으로 수정하였고, 추가적으로 인구와 가구&주택 파일간 조사시점의 상이한 경우가 발견되어 인구파일 기준으로 수정함

17) 수진메일 (2017.01.06.) “16년 5월, 시군구010 읍면동14 거쳐번호 714 조사구번호 186 가구번호 (3,4,5,6)에 해당하는 데이터인데 가구/주택 파일에는 시군구010 읍면동14 거쳐번호 714 조사구번호 186 가구번호 1에대한 정보만” 있다고 함

□ 조사구 표본 가중치 산출

- 조사월 t 에 대한 표본틀 내 조사구 수와 시험조사 파일 상의 표본 조사구 수로 다음의 조사구 수준 표본가중치(wEDmd)를 산출할 것¹⁸⁾

$$w_{thi}^d = \frac{N_{th}}{n_{th}}$$

여기서 N_{th} 와 n_{th} 는 각각 조사월 t 의 표본층 h 의 전체 조사구 수와 표본 조사구 수 (시험조사 파일에 존재하고 위의 조건 1과 조건 2를 만족하는 자료 기준)를 나타냄

□ 조사구 내 가구 표본 가중치 산출

- 조사월 t 에 대한 조사구 내 가구 표본 가중치(wHHmc)를 산출할 것

$$w_{tjhi}^d = \max\left\{\frac{M_{thi}}{m_{thi}}, 1\right\}$$

여기서 M_{thi} 와 m_{thi} 는 각각 조사월 t 의 표본층 h 의 조사구 i 내 총 가구 수와 표본 (응답)가구 수를 나타내고, j 는 가구 식별번호임

- 조사현황파악을 위해 조사구 내 총 가구수에 비해 조사된 가구수가 많은 조사구(즉, $M_{thi} < m_{thi}$)들을 (읍면동, 조사구 특성, 조사연월 포함하여) 보고하고, 이들에 대한 두 특성값 M_{thi} , m_{thi} 도 포함할 것

□ 가구 표본 가중치 산출

- 조사월 t 에 대한 가구 표본 가중치(wHHmd)를 산출할 것

$$w_{thij}^d = w_{thi}^d w_{tjhi}^d$$

18) 한국인구학회(2015,p.26)의 조사구추출 가중치 조정을 고려할 수는 있지만 정보부재로 생략하고자 함

□ 조사월 기준 통계표 작성

- 읍면동 별 그룹 및 차수에 따른 표본 조사구 수 분포를 산출하고 <표 2-3>과 같은지 확인할 것
- 읍면동 별 가중치 분포를 다음과 같이 작성할 것

<표 2-5> 제주도 2차 시험조사 읍면동별 가구 표본 가중치 분포

시 군 구	읍 면 동	표본수			전체							월평균기준 가중치분포						
		조사구	가구	가구원	\bar{w}	최소	중위수	최대	s_w	cv_w	L_w	\bar{w}	최소	중위수	최대값	s_w	cv_w	

- 전체는 9개월 전체 표본자료에 대해 가구 표본 가중치 분포

예, \bar{w} 는 가구 표본 가중치 평균; 최소는 최소값

s_w 는 가구 표본 가중치 표본오차

cv_w 는 가구 표본 가중치 상대표본오차 ($=s_w/\sqrt{\bar{w}}$)

$$L_w = 1 + cv_w^2$$

- 월평균기준 가중치 통계는 9개월 평균 가구 표본 가중치의 분포

예, \bar{w} 는 9개월 평균 가구 표본 가중치의 평균

최소는 9개월 평균 가구 표본 가중치의 최소값

s_w 는 9개월 평균 가구 표본 가중치의 표본오차

cv_w 는 9개월 평균 가구 표본 가중치의 상대표본오차 ($=s_w/\sqrt{\bar{w}}$)

- 읍면동별 총 9개월 전 기간 동안의 조사구 표본 가중치와 가구 표본 가중치에 대해 상자그림으로 분포를 비교할 것. 예로, 다음의 SAS 명령어 사용 가능
proc sgplot data=가중치파일; vbox 가중치 /category=읍면동; run;

2.3.4 분기·반기 가중치 산출

2.3.4.1 개요

□ 분기·반기 가중치 개요

- 제주도 2차 시험조사를 위한 표본설계 당시에는 24개월간 조사하고 6개월을 1년으로 간주하여 총 4년을 실시하는 가상적 상황을 상정함. 하지만 현재 통계청은 순환표본조사에서 월별 조사는 고려하고 있지 않고 다음의 3가지 방안을 고려함
 - (i) 매년 4% 씩 5년조사,
 - (ii) 매년 2% 씩 조사하고, 등록센서스 연도에 10%를 추가
 - (iii) 매년 2.5%씩 조사하고, 등록센서스 연도에 7.5%를 추가
- <표 2-2> 기준으로 분기와 반기 표본을 각각 통합하면 <표 2-6>과 같이 제주도 기준으로 2.5%와 4.9%의 추출률 규모의 표본임을 확인할 수 있음. 따라서 분기와 반기는 각각 통계청에서 현재 고려하고 있는 옵션 (ii)/(iii)과 옵션 (i)와 견줄 수 있는 규모에 해당함

<표 2-6> 2차 시험조사 분기 및 반기 통합표본 추출률 산정

시험조사 (균등분할순환조사 대비)

시	2015		분기					반기				
	조사구	2015 가구		조사구	가구	e가구	추출률		ef	조사구	가구	e가구
제주시	6,044	181,320	132	3,960	2,475	2.18%	1.36%	264	7,920	4,950	4.37%	2.73%
서귀포시	1,983	59,490	66	1,980	1,238	3.33%	2.08%	132	3,960	2,475	6.66%	4.16%
제주도	8,027	240,810	198	5,940	3,713	2.47%	1.54%	396	11,880	7,425	4.93%	3.08%

주:* e가구와 ef는 각각 일종의 유효표본수와 유효추출률에 해당하며 표본수와 추출률을 설계효과 $1.6=1+0.01*30$ 값으로 나누어 환산해본 값

- 분기와 반기 표본은 시산용으로 일단 첫 3달과 첫 6달을 포함하고자 함. 2015년11월은 인종과 중복되는 조사구로 대체됨에 따라 A-2는 가중치 산정에서 제외함

분기: A-1 (15/10), A-3 (15/12), A-4 (16/01)

반기: A-1 (15/10), A-3 (15/12), A-4 (16/01),
A-5 (16/02), A-6 (16/03), B-7 (16/04)

□ 가중치 산출을 위한 사전작업, 연구대상 한정, 결측 현황 파악

- 본 작업은 § 2.3.3절의 월별 기준 점검 및 가중치 산출에서 확인한 것으로 대신함. 따라서 월별 점검 자료로부터 작업을 수행하면 됨

2.3.4.2 분기 가중치 산출

□ 조사구 표본 가중치 산출

- 분기자료 (A-1,A-3,A-4)만을 추출 것
- 표본틀 내 조사구 수와 시험조사 파일 상의 표본 조사구 수로 다음의 조사구 수준 표본가중치(w_{qhi}^d)를 산출할 것

$$w_{qhi}^d = \frac{N_h}{n_{qh}}$$

여기서 N_h 는 표본층내 조사구 총수와 n_{qh} 는 분기자료내 표본조사구 총수

□ 조사구 내 가구 표본 가중치 산출

- 분기자료에 대한 조사구 내 가구 표본 가중치(w_{qjhi}^d)를 산출할 것

$$w_{qjhi}^d = \max\left\{\frac{M_{hi}}{m_{qhi}}, 1\right\}$$

여기서 M_{hi} 는 조사구내 총 가구수, m_{qhi} 는 분기 표본의 표본층 h 의 조사구 i 내 표본(응답)가구 수를 나타내고, j 는 가구 식별번호

□ 가구 표본 가중치 산출

- 분기자료의 가구 표본 가중치(w_{HHqd})를 산출할 것

$$w_{qhij}^d = w_{qhi}^d w_{qjhi}^d$$

2.3.4.3 반기 가중치 산출

□ 조사구 표본 가중치 산출

- 반기자료 (A-1,A-3,A-4,A-5,A-6,B-7)만을 추출 것
- 표본틀 내 조사구 수와 시험조사 파일 상의 표본 조사구 수로 다음의 조사구 수준 표본가중치(wEDbd)를 산출할 것

$$w_{bhi}^d = \frac{N_h}{n_{bh}}$$

여기서 N_h 는 표본층내 조사구 총수와 n_{bh} 는 반기자료내 표본조사구 총수

□ 조사구 내 가구 표본 가중치 산출

- 반기자료에 대한 조사구 내 가구 표본 가중치(wHHbc)를 산출할 것

$$w_{bjhi}^d = \max\left\{\frac{M_{hi}}{m_{bhi}}, 1\right\}$$

여기서 M_{hi} 는 조사구내 총 가구수, m_{bhi} 는 반기 표본의 표본층 h 의 조사구 i 내 표본(응답)가구 수를 나타내고, j 는 가구 식별번호

□ 가구 표본 가중치 산출

- 반기자료의 가구 표본 가중치(wHHbd)를 산출할 것

$$w_{bhij}^d = w_{bhi}^d w_{bjhi}^d$$

2.4 제주도 인총 20% 표본조사 자료

2.4.1 자료점검

□ 가중치 점검

- <표 2-4>의 제주도 2015 인총표본조사 표본추출 규모를 기준으로 조사구 및 가구 설계가중치 산출할 것
- 조사구 가중치: 읍면동 내 층별 (즉, 일반조사구 (기타), 아파트조사구)로 다음과 같이 조사구 설계가중치 (wEDcd) 산출

$$w_{hi}^d = N_h / n_h$$

여기서 N_h 와 n_h 는 각각 표본층 h 의 전체 조사구 수와 표본 조사구 수

- 가구 (가구원) 가중치: 조사구 내 가구가중치 산출 ?? 인총20%표본파일 읍면동 가중치 값으로 확인

$$w_{hij}^d = \text{읍면동 가중치 (wHHcd)}^{19)}$$

- 가구내 가구원 조건부 가중치를 다음과 같이 산출할 것

$$w_{j|hi}^d = w_{hij}^d / w_{hi}^d \quad (\text{wHHcc})$$

- 제주도 20% 인총 표본 내 가구수 및 가구원 수를 확인하고 시험조사와 함께 분석하기에 비교 가능한 표본수인지 확인 필요

19) wHHcd = design weight for household's 20% census

2.4.2 가구 가중치 분포 비교

□ 기간단위 및 종류별 가중치 분포 비교

- 월 가중치는 편의상 A-1만을 취할 것
- 제주도내 시별로 가중치 분포를 다음과 같이 작성할 것
- 추가로 boxplot도 생성할 것

〈표 2-7〉 제주도 2차 시험조사 시 및 기간별 가구 표본 가중치 분포

분류		표본수			전체						
시	기간	조사 구	가구	가구 원	\bar{w}	최 소	중 위 수	최 대	s_w	cv_w	L_w
제주시	인총										
	월										
	분기										
	반기										
서귀포	인총										
	월										
	분기										
	반기										
제주도	인총										
	월										
	분기										
	반기										

2.5 GREG 관련 변수 점검

2.5.1 보조변수 x 점검: 표본추출틀 자료

□ 보조변수 관련 파일 추가설명

- FRAME1016_all.csv: 인총 표본추출을 위해 등록센서스 자료로부터 변환된 표본추출틀로 볼 수 있고 조사구 수준의 정보를 갖고 있음
- FRAME1016_sam.csv: 위의 파일에 표본조사 추계결과를 붙이고 있음
- FRAME1016_ADM.csv: 나열단위는 읍면동으로 읍면동 수준의 추계결과를 갖고 있음²⁰⁾

□ 보조변수 선택 기준 및 명부

- 프레임 상의 보조변수는 3가지 수준별로 존재하는데 이는 인구·가구·주택임. 가구는 표본추출단위이고 인구는 가구 내 모든 가구원을 조사하는 관계로 가구·인구 통합가중치를 고려하고 있으므로 가구와 가구원 정보만 보조변수로 선택하는 것이 바람직 할 수 있음. 조사구 내 모든 가구, 즉, 모든 주택이 포함됨으로 사실상 주택, 가구, 인구(가구원)는 모두 동일한 (조건부 및 무조건부) 가중치를 갖게 됨. 보조정보의 속성에 따라 단위수준을 주의하여 취급하여야 함
- 가중치 작업을 위한 보조정보이므로 지나치게 많은 정보를 사용하는 것은 바람직하지 않을 수 있고 단순한 몇 가지 변수만 선택하고자 함
- 인구부분: 성별 × 연령층(5세 구분*)²¹⁾
- 가구부분: 거처 종류**, 가구원수***

* 연령구분: 0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-

** 거처구분: 단독주택, 아파트, 연립·다세대, 기타

*** 가구원수: 1인, 2인, 3인, 4인, 5인 이상 가구

20) ADM은 아마도 읍면동에 대한 표기인 듯함

21) FRAME1016_all.csv 내 10세 간격 등 collapsing 된 나이구분 확인 필요. 한국인구학회(2016, 28쪽)은 5세 간격의 연령별 인구를 고려하고 있지만, 이주희 사무관이 제공한 파일에는 거의 10살 묶음으로만 제공됨

□ 프레임 정보(FRAME1016_all.csv)를 활용한 총합값 (t_x)

— 예시 변수를 사용하여 총합값을 유도²²⁾할 것. 여기서 총합은 추정단위 기준으로 고려할 것이므로 읍면동 수준이 됨

— 인구 총합변수·카테고리

■ 성·연령별 인구총합 : 제주도 내 읍면동별로 표를 만들 것

(남, 여) x (0-14 / 15-19 / 20-29 / 30대 / 40대 / 50대 / 60대 이상)

■ 조합별 변수명은 다음과 같이 할 것

TSexAge11 = 남자 0-14세 총합
 TSexAge12 = 남자 15-19세 총합
 TSexAge13 = 남자 20-29세 총합
 TSexAge14 = 남자 30-39세 총합
 TSexAge15 = 남자 40-49세 총합
 TSexAge16 = 남자 50-59세 총합
 TSexAge17 = 남자 60세이상 총합
 TSexAge21 = 여자 0-14세 총합
 TSexAge22 = 여자 15-19세 총합
 TSexAge23 = 여자 20-29세 총합
 TSexAge24 = 여자 30-39세 총합
 TSexAge25 = 여자 40-49세 총합
 TSexAge26 = 여자 50-59세 총합
 TSexAge27 = 여자 60세이상 총합

■ 표의 layout은 다음 형식을 갖출 것

시군구	읍면동	TSexAge11	TSexAge12	...	TSexAge26	TSexAge27
제주시	한림읍					

■ SAS 파일로 저장하되 TotSexAge라 명명할 것

22) Q. 연령구분이 0-14, 15-19, 등 rough한 구분

- TotSexAge2 SAS 파일을 앞서와 동일한 방법으로 생성하되 성·연령 조합을 다음과 같이 조정할 것

TSexAge211 = 남자 0-19세
 TSexAge212 = 남자 20-39세
 TSexAge213 = 남자 40-59세
 TSexAge214 = 남자 60세 이상
 TSexAge221 = 여자 0-19세
 TSexAge222 = 여자 20-39세
 TSexAge223 = 여자 40-59세
 TSexAge224 = 여자 60세 이상

— 가구 총합변수·카테고리

- 가구원 수별 가구총합 : 제주도 내 읍면동별로 표를 만들 것

1인 / 2인 / 3인 / 4인 / 5인 이상 가구

- 조합별 변수명은 다음과 같이 할 것

THHSize1 = 1인 가구
 THHSize2 = 2인 가구
 THHSize3 = 3인 가구
 THHSize4 = 4인 가구
 THHSize5 = 5인 이상 가구

- 거처 종류별 가구총합 : 제주도 내 읍면동별로 표를 만들 것

단독주택²³⁾ / 아파트 / 연립·다세대 / 기타주택

- 조합별 변수명은 다음과 같이 할 것

TDuType1 = 단독주택 거주 가구수
 TDuType2 = 아파트 거주 가구수

23) “호”로 구분되는 것으로 보아 주택수준 정보인 듯. 추출단위는 가구이므로 주택(거처 dwelling unit)은 달리 처리해야 할 것

TDuType3 = 연립·다세대 거주 가구수

TDuType4 = 기타주택 거주 가구수

- 표의 layout은 다음 형식을 갖출 것

시군구	읍면동	THHSize1	THHSize2	...	TDuType3	TDuType4
제주시	한림읍					

- SAS 파일로 저장하되 TotHHSizeDuType 이라 명명할 것

- TotHHSizeDuType2 SAS 파일을 앞서와 동일한 방법으로 생성하되 그룹 구성을 다음과 같이 조정할 것

THHSize21 = 1인 가구

THHSize22 = 2인 가구

THHSize23 = 3인 이상 가구

TDuType21 = 단독주택 거주 가구수

TDuType22 = 비단독주택 거주 가구수

- 2015 인총 20%표본조사 추계결과 (표본추정 결과)로 비교 시 참고 자료로 활용

— FRAME1016_ADM.csv에서 15년 인총 조사구별 추정값 (\hat{t}_x) 확인가능

— 앞서의 프레임정보와 동일한 구조로 성·연령 및 가구원수·거처종류에 대한 인총 20% 표본조사 총합 추정치를 취합하여 SAS 파일을 만들 것

- 성·연령 추정치 파일은 TotSexAgeCS20 라 명명하고 변수명도 끝(postfix)에 “CS20” 를 붙일 것. 예로, TSexAge11CS20
- 가구원수·거처종류 추정치 파일은 TotHHSizeDuTypeCS20 이라 명명하고 변수명도 앞서와 동일하게 “CS20” 을 끝에 붙일 것

— TotSexAge2CS20과 TotHHSizeDuType2CS20 파일을 생성하되 앞서 프레임파일에 기술한 동일한 방식으로 변수를 생성할 것

□ FRAME1016_ADM

	A	B	C	D	E	F	G	H	I	J	K
1	ADM_CD	_FREQ_	gagu	gaguwon1	gaguwon2	gaguwon3	gaguwon4	gaguwon5	hmale	hfemale	htype1
2	2501053	9281	9717	1739	2420	2317	2401	840	6995	2722	96
3	2501055	4005	3955	1068	1145	808	699	235	2642	1313	39
4	2501056	5400	5188	2323	1497	781	440	147	2904	2284	51
5	2501057	6688	11379	5712	1954	1725	1400	588	7754	3625	112

O	P	Q
REQ_	gagu	gagu
21	20	9717
36	33	

all 자료 gagu 더하면 위의 9717이 됨

□ FRAME1016_ALL

C	D	E	F	G	H	I	J	K	L	M	N	O	P
sgg_cd	sgg	emd_cd	emd	jsg	char	gagu07	dandok07	apt07	yun07	etc07	key	_FREQ_	gagu
25010	동구	2501053	효동		1 A	21	0	21	0	0	2501053001	21	
25010	동구	2501053	효동		2 A	36	0	36	0	0	2501053002	36	
25010	동구	2501053	효동		3 A	36	0	36	0	0	2501053003	36	

□ FRAME1016_SAM

- gagu07은 프레임 상 정보이고 gagu 조사된 가구 아닌지 점검해 볼 것 (수진)
- 만약 그렇다면 조사구별 프레임 상 정보와 실제 조사결과의 비교도 가능할 것임

E	F	G	H	I	Q
emd_cd	emd	jsg	char	gagu07	gagu
1101053	사직동	1	1	29	
1101053	사직동	2	1	28	26
1101053	사직동	3	1	37	
1101053	사직동	4	1	27	26
1101053	사직동	5	4	1	1
1101053	사직동	6	1	26	
1101053	사직동	7	1	22	
1101053	사직동	8	1	40	
1101053	사직동	9	1	30	
1101053	사직동	10	1	21	19
1101053	사직동	11	1	39	
1101053	사직동	12	1	32	
1101053	사직동	13 A		18	
1101053	사직동	14	1	33	
1101053	사직동	15	1	37	
1101053	사직동	16	1	25	
1101053	사직동	17	1	25	19
1101053	사직동	18	1	20	

2.5.2 보조변수 x 점검: 표본조사 자료

2.5.2.1 표본가중치 파일

□ 월별 기준 가중치 산출자료

— § 2.3.3절의 월별 가중치 파일에서 매칭변수, 설계변수, 가중치 변수를 취합

- 매칭변수: 시도, 시군구, 조사구 식별번호(예, KEY)
- 설계변수: 읍면동, 조사구 종류, 조사월(t)
- 가중치 변수:

$$wEDmd (=w_{thi}^d)$$

$$wHHmc (=w_{tj|hi}^d)$$

$$wHHmd (=w_{thij}^d)$$

□ 분기 가중치 산출자료

— § 2.3.4절의 분기 가중치 파일에서 매칭변수, 설계변수, 가중치 변수를 취합

- 매칭변수: 시도, 시군구, 조사구 식별번호(예, KEY)
- 설계변수: 읍면동, 조사구 종류
- 가중치 변수:

$$wEDqd (=w_{qhi}^d)$$

$$wHHqc (=w_{qj|hi}^d)$$

$$wHHqd (=w_{qhij}^d)$$

□ 반기 기준 가중치 산출자료

— § 2.3.4절의 반기 가중치 파일에서 매칭변수, 설계변수, 가중치 변수를 취합

- 매칭변수: 시도, 시군구, 조사구 식별번호(예, KEY)
- 설계변수: 읍면동, 조사구 종류
- 가중치 변수:

$$wEDbd (=w_{bhi}^d)$$

$$wHHbc (=w_{bj|hi}^d)$$

$$wHHbd (=w_{bhij}^d)$$

□ 인총 20% 표본조사 가중치 산출자료

— § 2.4.1절의 인총 파일에서 매칭변수, 설계변수, 가중치 변수를 취합

- 매칭변수: 시도, 시군구, 조사구 식별번호(예, KEY)
- 설계변수: 읍면동, 조사구 종류
- 가중치 변수:

$$wEDbd (=w_{bhi}^d)$$

$$wHHbc (=w_{bj|hi}^d)$$

$$wHHbd (=w_{bhij}^d)$$

2.5.2.2 표본조사자료 파일

- 표본조사는 가구별로 진행. 한 설문지에 주택/가구/(가구원별) 가구원 문항 포함
- 단, 조사단위, 즉 주택, 가구, 가구원별로 자료파일이 구성됨

□ 제주도 2차 시험조사 2 가구 기준 주택 관련 자료

— § 2.3.1절에 언급한 3종류 파일 중 ‘1.주택(15.10~16.8)’ 사용

- 매칭변수: 시도, 시군구, 조사구 식별번호(예, KEY), 거처번호, 가구번호
- 설계변수: 조사월(t)
- 조사변수: C1 (거처종류)

□ 제주도 2차 시험조사 2 가구 기준 가구 관련 자료

— § 2.3.1절에 언급한 3종류 파일 중 ‘2.가구(15.10~16.8)’ 사용

- 매칭변수: 시도, 시군구, 조사구 식별번호(예, KEY), 거처번호, 가구번호
- 설계변수: 조사월(t)

- 조사변수:

C1=HHC1 (가구구분)

(☞ 주택 자료에 동일한 명칭의 변수가 존재하므로 rename 할 것)

□ 제주도 2차 시험조사 2 가구 기준 개인 관련 자료

— § 2.3.1절에 언급한 3종류 파일 중 ‘3.주택(15.10~16.8)’ 사용

- 매칭변수: 시도, 시군구, 조사구 식별번호(예, KEY), 거처번호, 가구번호, 가구원번호
- 설계변수: 조사월(t)
- 조사변수:
 - C2 (성별)
 - C3 (나이)
 - C9 (교육정도)

2.5.2.3 인총20% 표본조사 파일

- 인총 20% 표본조사 자료도 제주도 2차 시험조사와 같거나 매우 유사한 형식을 주지 할 것임. 한 설문지에 주택/가구/(가구원별) 가구원 문항 포함
- 조사단위도 주택, 가구, 가구원별로 자료파일이 구성됨. 따라서 2.5.2.2절의 시험조사와 같은 기준으로 파악해 볼 것

2.5.2.4 표본분석 파일

□ 시험조사 표본자료 구축

- 앞서 생성한 가구가중치 파일(a)에 표본조사 파일 (주택 b, 가구 c, 개인 d)들을 매칭변수에 의해 통합(merge)할 것
- 2.5.2.2절에서 생성된 가구가중치는 월, 분기, 반기의 세 종류 가중치 세트를 포함하게 되지만 분기와 반기 가중치는 포함 기간이 일부이므로 미포함 기간

자료들에 대해서는 각각 결측값(.)을 갖게 됨

- 가구수준과 개인수준별로 표본자료를 구축할 것
- 가구수준으로는 조사구 내 다수 가구
- 개인수준으로는 가구 내 하나 이상의 가구원 자료들로 구성
- § 2.3.3절에 기술된 자료제한 조건에 따라 수준별 파일 병합 절차는 단순화될 수 있음

- 가구자료: JJT2PSS1ALL.sasb7dat 으로 생성
- 개인자료: JJT2HHS1ALL.sasb7dat 으로 생성

주택		TABLE : TN_SC_POPL_HOUS
41 거처의 종류 이 집은 어떤 종류입니까? • 주거용 부분과 영업용 부분이 함께 있는 건물의 경우 주거용 부분이 같거나 더 넓으면 ① 단독주택 중 ② 영업 겸용 단독주택, 영업용 부분이 더 넓으면 ⑤ 비거주용 건물 내 주택에 해당합니다.		
C1 ① 단독주택 ② 아파트 ③ 연립주택 ④ 다세대주택 ⑤ 비거주용 건물(상가, 공장, 여관 등) 내 주택 ⑥ 오피스텔 ⑦ 호텔·여관 등 숙박업소의 객실 ⑧ 기숙사 및 사회시설 ⑨ 판잣집, 비닐하우스 ⑩ 기타	C2 ① 일반 단독주택 ② 다가구 단독주택 ③ 영업 겸용 단독주택	C3
42 층 방 수 이 주택에는 방, 거실, 식사용 방이 각각 몇 개입니까? • 세 든 가구를 포함하여 주택 전체의 방 수를 기입하며, 사용하지 않는 방도 포함합니다.		
• 방 C6 개 • 거실(서면이 벽 또는 문으로 막혀 있는 대청마루 포함) C7 개 • 식사용 방(식탁 등이 놓여 있어 식사를 할 수 있는 별도 공간) ... C8 개		
43 주거시설 수 이 주택에는 독립된 출입구, 부엌, 화장실이 각각 몇 개입니까? • 독립된 출입구는 다른 가구의 주거 부분을 차지 않고 외부와 자유로이 출입할 수 있는 현관문도 포함합니다.		
• 독립된 출입구 C12 개 • 부엌 C10 개 • 화장실 C11 개		

컬럼명	컬럼 내용
SORT_DT	조사월
HHD_LTNM_ID	가구명부ID
C1	주택형태(550041)
C2	단독주택유형(550042)
C6	방수
C7	거실수
C8	식사용방수
C10	부엌수
C11	화장실수
C12	독립된출입구수
CP_CD	시도코드
CDW_CD	시군구코드
TTB_CD	읍면동코드
HU_NUM	거처번호
HHD_NUM	가구번호
ED_NUM	조사구번호
ED_PRPT_CD	조사구특성코드

□ 인총 20% 표본조사 표본자료 구축

- 가구자료: JJCSPSS1.sasb7dat 으로 생성
- 개인자료: JJCSHHS1.sasb7dat 으로 생성

3. 가구·가구원 통합가중치 시산

3.1 가중치 개요

3.1.1 표본설계 및 설계가중치 구성

□ 표본설계 및 기호

- 개인특성은 $(y_{jk}, \mathbf{x}_{jk})$ 으로 표기. 단, 가구 j 내 k 번째 가구원 조사/보조변수 값
- 가구특성은 편의상 $(y_{jc}, \mathbf{x}_{jc})$ 로 표기

— 가구모집단 $U_I = \{j : j = 1, \dots, N\}$

— 가구원모집단 $U_j = \{k : k = 1, \dots, N_j\}$

— 모집단 $U = \cup_{j=1}^N U_j = \cup_{j \in U_I} U_j$

— 가구표본 $s_I = \{j : j = 1, \dots, n\}$

— 전체표본 $s = \cup_{j=1}^n U_j = \cup_{j \in s_I} U_j = \{(jk) : k = 1, \dots, n_j, j = 1, \dots, n\}$

- 가구 설계가중치: 가구(집락)는 조사구 내 전수조사 되므로 (층 내) 조사구 추출률 $\pi_{hi} (= \pi_{Ij})$ 의 역수로 정의됨

$$d_{Ij} = 1/\pi_{Ij} (= d_i)$$

- 가구 내 가구원도 전수조사 되므로 (조건부) 설계 가중치도 1로 정의

$$d_{kj} = 1$$

- 가구원 설계 가중치는 가구 설계가중치에 가구 내 가구원 설계가중치 곱으로 정의됨. 가구 내 모든 가구원이 조사되므로 역시 전수조사

$$d_{jk} = d_{Ij} d_{kj} = d_{Ij} \quad (3-1)$$

3.1.2 칼리브레이션 가중치²⁴⁾

□ 칼리브레이션 등식

— 집락수준-칼리브레이션 등식

$$\sum_{j \in s_I} w_{Ij} \mathbf{x}_{jc} = \sum_{j \in U_I} \mathbf{x}_{jc} \quad (3-2)$$

— 가구수준-칼리브레이션 등식

$$\sum_{(jk) \in s} w_{jk} \mathbf{x}_{jk} = \sum_{(jk) \in U} \mathbf{x}_{jk} \quad (3-3)$$

□ 비통합 칼리브레이션 (non-integrated calibration, NC)

- 단계 1: 가구 설계가중치 d_{Ij} 에 대해 가구수준 보조정보 \mathbf{x}_{jc} 를 사용하여 (3-2) 식을 만족할 수 있도록 칼리브레이션 가중치를 산출
- 단계 2: 단계 1과는 독립적으로 개인 설계가중치 $d_{jk} (= d_{Ij} d_{k|j})$ 에 대해 개인수준 보조정보 \mathbf{x}_{jk} 를 사용하여 (3-3)을 만족할 수 있도록 칼리브레이션 가중치를 산출

□ 통합 칼리브레이션 (integrated calibration, IC)

- 단계 1: 식 (3-3)에서 $w_{jk} = d_{k|j} w_{Ij}$ 으로 정의하고 개인수준-칼리브레이션 등식을 가구 칼리브레이션 가중치의 함수로 표현함
- 단계 2: 가구 및 가구원 보조변수벡터를 다음과 같이 새롭게 정의함

$$\mathbf{x}_{jc}^* = \begin{pmatrix} \mathbf{x}_{jc} \\ \tilde{\mathbf{x}}_{j\cdot} \end{pmatrix} \quad (3-4)$$

$$\text{여기서 } \hat{\mathbf{x}}_{j\cdot} \text{ 은 가구총합추정치, 즉 } \tilde{\mathbf{x}}_{j\cdot} = \begin{cases} \sum_{k \in s_j} d_{k|j} \mathbf{x}_{jk} & \text{표본} \\ \sum_{k \in U_j} \mathbf{x}_{jk} & \text{전수} \end{cases}$$

24) Sarndal, C-E. (2007), Estevao, V. M. and Sarndal, C-E. (2006) 참조

- 단계 3: 가구 설계가중치 d_{Ij} 에 대해 단계 2에서 새롭게 정의된 가구수준 보조 정보 \mathbf{x}_{jc}^* 를 사용하여 다음 식 (3-5)을 만족하는 칼리브레이션 가중치를 산출

$$\sum_{j \in s_I} w_{Ij} \mathbf{x}_{jc}^* = \sum_{j \in U_I} \mathbf{x}_{jc}^* \quad (3-5)$$

※ 식 (3-5)는 다시 표기하면 다음과 같음

$$\begin{pmatrix} \sum_{s_I} w_{Ij} \mathbf{x}_{jc} \\ \sum_{s_I} w_{Ij} \sum_{s_j} d_{k|j} \mathbf{x}_{jk} \end{pmatrix} = \begin{pmatrix} \sum_{U_I} \mathbf{x}_{jc} \\ \sum_U \mathbf{x}_{jk} \end{pmatrix} \quad (3-5a)$$

- 단계 4: 단계 3에서 산출된 가구 칼리브레이션 가중치로 개인 칼리브레이션 가중치를 다음과 같이 산출

$$w_{jk} = d_{k|j} w_{Ij} \quad (3-6)$$

□ Lemaître & Dufour (1987) 통합 가중 방법론 (L-D calibration, LC)

- Lemaître & Dufour 방법론 “동일분담(equal share) 보조정보” 방식으로도 불리우며 개인별 보조변수 \mathbf{x}_{jk} 을 다음의 가구수준 평균값으로 대체하고

$$\bar{\mathbf{x}}_{j\cdot} = (1/N_j) \sum_{U_j} \mathbf{x}_{jk} \quad (3-7)$$

- 개인가중치 $d_{jk} = d_{Ij} d_{k|j} = d_{Ij}$ ($d_{k|j} = 1$)에 대해 대체된 가구 평균적 개인수준 보조정보 $\mathbf{x}_{jk}^* := \bar{\mathbf{x}}_{j\cdot}$ 을 사용하여 (3-3)을 만족하도록 개인 칼리브레이션 가중치를 산출함
- Lemaître & Dufour (1987)는 위의 통합가중치를 일반화선형회귀추정(GREG) 방법을 통해 산출함

3.1.3 GREG 가중치

□ GREG 추정량

— GREG 추정량

$$\hat{Y}_{gr} = \begin{cases} \sum_{(jk) \in s} d_{jk} g_{jk} y_{jk} & \text{개체 (예, 개인)} \\ \sum_{j \in s_I} d_{jc} g_{jc} y_{jc} & \text{집락 (예, 가구)} \end{cases} \quad (3-8)$$

— GREG 조정계수 (g-weight)

$$g_{jc} = 1 + (\mathbf{X} - \hat{\mathbf{X}}_{\pi})' \left(\sum_{j \in s_I} c_{Ij}^{-1} d_j \mathbf{x}_{jc} \mathbf{x}_{jc}' \right)^{-1} c_{Ij}^{-1} \mathbf{x}_{jc} \quad (\text{가구}) \quad (3-9)$$

$$g_{jk} = 1 + (\mathbf{X} - \hat{\mathbf{X}}_{\pi})' \left(\sum_{j \in s} a_{jk}^{-1} d_{jk} \mathbf{x}_{jk} \mathbf{x}_{jk}' \right)^{-1} a_{jk}^{-1} \mathbf{x}_{jk} \quad (\text{가구원}) \quad (3-10)$$

여기서 \mathbf{x}_{jc} 는 가구(집락) j 의 보조변수 값, c_{Ij} 와 a_{jk} 는 GREG 모형의 분산구조를 나타냄

□ GREG 모형

— 개인 GREG 모형

$$E_M(y_{jk}) = \mathbf{x}_{jk} \boldsymbol{\beta} \quad , \quad V_M(y_{jk}) = a_{jk} \sigma^2 \quad (3-11)$$

가구조사에서는 y 가 주로 이분변수(dichotomous variable)인데 이때 a_{jk} 는 주로 1의 값으로 놓고, 사업체조사는 사업체 규모 값에 비례한 값을 사용 (Särndal et al, 1992 참조)

— 가구 GREG 모형

$$E_M(y_{jc}) = \mathbf{x}_{jc} \boldsymbol{\beta}_I \quad , \quad V_M(y_{jc}) = c_{Ij} \sigma^2 \quad (3-12)$$

□ Lemaître & Dufour (1987) GREG 방식

— g_{jk} 를 x_{jk} 대신 식 (3-7)의 $x_{jk}^* := \bar{x}_j$ 으로 산출

□ Nieuwenbroek (1993) 방식 고려

— 분산모형

$$c_{Ij} = \sum_{k \in U_j} a_{jk} \quad (3-13)$$

— 가구 GREG 모형 (3-12)의 가정 하에서 식 (3-9)로 g-weight g_{jc} 를 구하여 통합가중치를 산출

— 만약 $a_{jk} = 1$ 이면 $c_{Ij} = N_j$ 이 되며, $c_j = 1$ 을 사용할 수도 있음(Steel and Clark, 2007 참조)²⁵⁾

25)향후 GREG 모형 (3-12)와 (3-13)을 수정하여 고려할 수도 있을 것. 예로, $V_M(y_{jk}, y_{jk'}) = \sigma_{jkk'}$ 을 가정할 수 있을 것

3.2 작업파일 생성

3.2.1 자료정리 I (2차 시험조사 기준)

3.2.1.1 개요

□ 모집단²⁶⁾

- 추정 및 분산추정의 안정성을 검토하기 위해서는 표본 규모와 추출률이 커야 하지만 단위 모집단 또한 적절한 규모를 갖추어야 하므로 제주시와 서귀포시를 각각 모집단으로 간주하고 함. <표 2-4>에 시별 그리고 읍면동별 조사구 규모에 대한 기술이 포함됨

□ 추정항목(y)

— 인구부문

교육정도 (C9*C10)
종교유무 (C146)
출생지 (C23)
거주지1 (1년전, C27)
통근·통학여부 (C43)
평균통근·통학시간 (C58,C58)
경제활동상태 (C74)
종사상 지위 (C77)
혼인상태 (C88)

— 가구부문 (주택부분 포함)

주거 전용·영업 겸용 여부 (C77)
점유형태 (C78)
주인(대표)가구여부(C88)

26) 읍면동: 최하위 모집단으로는 읍면동이 되지만 <표 2-4>에 나타나 있듯이 작게는 우도면 22개, 봉개동과 도두동이 각각 32개, 추자면 34개 등이 있고 많게는 이도2동 686개, 연동 711개, 노형동 755개로 구성됨. 조사구당 평균 30가구로 구성된다고 하더라도 읍면동 규모가 그리 크지 않음. 시: 반면 시단위인 제주시와 서귀포시는 각각 6044개와 1983개 조사구들로 구성되어 있어 읍면동에 비해 규모를 갖추고 있어 가중치 산출에 따른 추정량은 물론 분산추정량의 안정성을 검토하기에 적절할 것으로 판단됨

타지 주택 소유 여부(C87)

□ 표본설계정보 및 설계가중치

— 표본층은 읍면동의 행정영역으로 구분되는 지역층과 읍면동내 조사구 특성(아파트, 기타)에 따라 구분되는 아파트 특성층으로 나뉨. 즉,

- 읍면동 (EMD, TTB_CD)
- 조사구특성코드 (ED_PRPT_CD)

— 일차추출단위(PSU)는 조사구. 즉,

- 조사구 (ED_NUM)

— 개체식별번호(ID)는 시도, 시군구, 조사구, 거처, 가구, 가구원 등에 대해 각각 관리할 것

— 조사시점(월)는 <표 2-2>의 정보, 즉, 그룹, 차수, 연월

- 그룹 (A,B,C,D)
- 차수 (1,2,...,24)²⁷⁾
- 연월 (t)

— 설계가중치는 조사구 및 가구 (설계) 가중치

[월기준]

- 조사월 t 조사구 설계가중치(w_{thi}^d) wEDmd
- 조사월 t 가구무응답조정계수($w_{tj|hi}^d$) wHHmc
- 조사월 t 가구 설계가중치(w_{thij}^d) wHHmd

[분기기준]

- 분기 조사구 설계가중치(w_{ghi}^d) wEDqd
- 분기 가구무응답조정계수($w_{gj|hi}^d$) wHHqc
- 분기 가구 설계가중치(w_{ghij}^d) wHHqd

27) 현재 가용한 월로 2016년7월인 차수 10일 뜻함

[반기기준]

- 반기 설계가중치(w_{bhi}^d) wEDbd
- 반기 가구무응답조정계수($w_{bj|hi}^d$) wHHbc
- 반기 가구 설계가중치(w_{bhi}^d) wHHbd

3.2.1.2 개인 기본파일 생성

□ 개인수준 조사정보와 표본설계 취합

— 개인별 정보만으로 구성된 개인 기본파일을 생성할 것. 즉 다음의 정보 포함

- 표본설계정보 및 설계가중치
- 인구부문 추정항목

□ 보조변수 x_{jk} : 인구부문 (수준)

— SexAge## : 개인별 성·연령 구분자 14개를 다음과 같이 생성할 것

SexAge11 = 1 (남자 & 0-14세), = 0 (아니면)
SexAge12 = 1 (남자 & 15-19세), = 0 (아니면)
SexAge13 = 1 (남자 & 20-29세), = 0 (아니면)
SexAge14 = 1 (남자 & 30-39세), = 0 (아니면)
SexAge15 = 1 (남자 & 40-49세), = 0 (아니면)
SexAge16 = 1 (남자 & 50-59세), = 0 (아니면)
SexAge17 = 1 (남자 & 60세 이상), = 0 (아니면)
SexAge21 = 1 (여자 & 0-14세), = 0 (아니면)
SexAge22 = 1 (여자 & 15-19세), = 0 (아니면)
SexAge23 = 1 (여자 & 20-29세), = 0 (아니면)
SexAge24 = 1 (여자 & 30-39세), = 0 (아니면)
SexAge25 = 1 (여자 & 40-49세), = 0 (아니면)
SexAge26 = 1 (여자 & 50-59세), = 0 (아니면)
SexAge27 = 1 (여자 & 60세 이상), = 0 (아니면)

→ proc freq을 이용하여 읍면동별 14개 구분자 빈도수 파악해 볼 것²⁸⁾

28) 읍면동별 GREG 추정의 안정성을 확인하기 위함.

- SexAge2## : § 2.5 절의 TotSexAge2 파일 생성에서 고려한 조합변수 생성
- 비통합 칼리브레이션 가중치 산출 가능

□ 개인 기본파일 저장

- 생성한 자료를 sas7bdat 형식 파일로 생성할 것
- 이때 기간별로 파일을 생성할 것
- JJT2MPSS1 월자료 (2015년10월 기준)
- JJT2QPSS1 분기자료 (2015년10월 이후 3개월)
- JJT2BPSS1 반기자료 (2015년10월 이후 6개월)
- 파일 기본 내용을 확인할 것 (proc contents)

3.2.1.3 가구파일 생성

□ 가구수준 조사정보와 표본설계 취합

- 가구별 정보로 구성된 가구 기본파일을 생성할 것. 즉 다음의 정보 포함
 - 표본설계정보 및 설계가중치. 단, 가구원 식별번호 제거
 - 가구부문 추정항목

□ 보조변수 x_{jc} : 가구부문 (수준)

- DuType#: 가구별 4가지 거처종류²⁹⁾ 지시자를 생성할 것

DuType1 = 1 (단독주택), =0 (아니면)
 DuType2 = 1 (아파트), =0 (아니면)
 DuType3 = 1 (연립 · 다세대), =0 (아니면)
 DuType4 = 1 (기타, =0 (아니면))

29) 거처종류는 주택정보로 해당 거처종류별 가구수로 확인되어 조사구별로 집계됨

— HM_CNT: 가구원수 (응답가구원수)

— HHSize#: 가구규모 변수를 생성할 것. 조사표 변수 HM_CNT을 이용할 것

HHSize1 = 1 (HM_CNT=1, 1인가구), =0 (아니면)

HHSize2 = 1 (HM_CNT=2, 2인가구), =0 (아니면)

HHSize3 = 1 (HM_CNT=3, 3인가구), =0 (아니면)

HHSize4 = 1 (HM_CNT=4, 4인가구), =0 (아니면)

HHSize5 = 1 (HM_CNT>4, 5인이상), =0 (아니면)

— DuType2# : § 2.5 절의 TotSexAge2 파일에서 고려한 조합변수 생성

— HHSize2# : § 2.5 절의 TotHHSizeDuType2 파일에서 고려한 조합변수 생성

— 비통합 칼리브레이션 가중치 산출 가능

□ 가구 기본파일 저장

— 생성한 자료를 sas7bdat 형식 파일로 생성할 것

— 이때 기간별로 파일을 생성할 것

— JJT2MHHS1 월자료 (2015년10월 기준)

— JJT2QHHS1 분기자료 (2015년10월 이후 3개월)

— JJT2BHHS1 반기자료 (2015년10월 이후 6개월)

— 파일 기본 내용을 확인할 것 (proc contents)

3.2.1.4 가구 확장파일 생성

□ 통합 칼리브레이션 방식 사용을 위한 보조변수 생성

— 가구수준 정보로 구성된 가구 확장파일을 생성

□ § 3.2.1.2에서 생성한 개인파일 JJT2MPSS1, JJT2QPSS1, JJT2BPSS1 으로부터 식 (3-4)의 가구수준 보조정보 \tilde{x}_j 를 생성하여 붙일 것

— HHSexAge## : 가구별로 개인정보 SexAge##를 취합한 총합을 구하고 이를 HHSexAge##로 명명할 것.³⁰⁾

— HHSexAge2## : § 2.5 절의 TotSexAge2 파일에서 고려한 조합변수를 기준으로 한 가구수준 보조정보 생성

— 위 변수정보를 § 3.2.1.3의 해당 기간 가구 기본파일 JJT2MHHS1, JT2QHHS1, JT2BHHS1에 병합하고, 결과자료를 가구 확장 SAS 파일 JJT2MHHS2, JT2QHHS2, JT2BHHS2 로 명명하고 저장할 것

3.2.1.5 개인 확장파일 생성

□ Lemaitre & Dufour 방식 사용을 위한 보조변수 생성

— 개인수준 정보파일에 가구수준 특성값을 equal share 방식 (즉, 평균값 대체)으로 분리하여 개인파일에 병합시킨 개인확장파일을 생성

□ § 3.2.1.2에서 생성한 개인수준 파일 JJT2MPSS1, JJT2QPSS1, JJT2BPSS1 에 § 3.2.1.4에서 생성한 가구수준파일 JJT2MHHS1, JT2QHHS1, JT2BHHS1에 있는 가구수준 보조정보를 식 (3-7)과 같이 평균값을 산출하여 붙일 것.

— 가구파일 JJT2HH1 자료로부터

— AveSexAge## : 가구별 해당 성별/연령별 가구원수 평균값을 산출함

$$\text{AveSexAge##} = \text{HHSexAge##} / \text{HM_CNT}$$

30) §2.3.3의 조건 2를 적용함에 따라 가구내 무응답은 고려하지 않아도 됨

— AveDuType# : 가구별 해당 주택유형 평균값을 산출함

$$\text{AveDuType\#} = \text{DuType\#} / \text{HM_CNT}$$

— AveHHSIZE# : 가구별 해당 주택유형 평균값을 산출함

$$\text{AveHHSIZE\#} = \text{HHSIZE\#} / \text{HM_CNT}$$

— AveSexAge2##, AveDuType2#, AveHHSIZE2# 를 생성

— 앞서 생성한 다음의 변수, 즉

AveSexAge##, AveDuType#, AveHHSIZE#,
AveSexAge2##, AveDuType2#, AveHHSIZE2#

변수들을 개인파일 JJT2MPSS1, JJT2QPSS1, JJT2BPSS1 자료에 가구 KEY변수를 이용하여 병합(merge)하고, 결과자료를 개인 확장 SAS 파일 JJT2MPSS2, JJT2QPSS2, JJT2BPSS2로 명명하고 저장할 것

3.2.2 자료정리 II (인총 20% 표본조사 기준)

□ 추후 고려

3.3.1 비통합가중치

3.3.1.1 개요

□ 비통합가중치 산출 방식

- § 3.1.2 절에 기술된 비통합 칼리브레이션 단계를 거침
- 이는 가구와 가구에 대해 해당 보조변수를 활용하여 GREG 조정계수 식 (3-9)와 (3-10)을 산출함

3.3.1.2 GREG 조정계수 산출 - 가구·월·읍면동³¹⁾

□ 가구기본조사파일과 가구총합파일을 사용³²⁾

- JJT2MHHS1.sas7bdat
- TotHHSIZE.DUTYPE.sas7bdat

□ PSU 모집단 정보 파일 생성

- 시군구, 읍면동, 조사구 종류별 조사구수 파일을 생성할 것. <표 2-4>를 이용하여 다음의 예시가 주는 결과와 같도록 2015년 제주도 인총표본조사 당시 프레임내 조사구 수 정보를 나타내도록 할 것

```
data JJEDFrmInfo;
  input CDW_CD TTB_CD ED_P RTP_CD1 _total_;
  datalines;
3901011 한림읍 1 229
3901011 한림읍 A 17
.....
;
run;
```

31) 수준=(가구,월,칼리브레이션)

32) 보조변수 그룹별 표본 case가 존재하지 않아 불안정한 GREG 발생시 본 작업을 교정해야 할 수도 있음. 예로 해당 읍면동 삭제

□ 보조변수 HT-총합추정치 \hat{X}_π 산출

- 시군구내 개별 읍면동이 단위 모집단임
- 위의 2개 파일 모두를 시군구(코드)와 읍면동(코드)에 대해 sort할 것
- 가구기본조사파일에서 가구설계가중치 wHHmd 를 이용하여 가구별 보조변수 값에 대해 Horvitz-Thompson 가중총계추정치 (HT-estimator)를 산출할 것
- 가구별 보조변수값은 다음과 같음

HHSize1-HHSize5, DuType1-DuType4
 HHSize21-HHSize23, DuType21-DuType22

- 다음의 SAS 프로그램을 통해 $\hat{X}_\pi = \sum_s w_{thij}^d x_{thijc}$ 을 산출할 것

<SAS 예시 3-1>

```
proc sort data=JJT2MHHS1; by CDW_CD TTB_CD; run;
proc surveymeans data=JJT2MHHS1 totol=JJEDFrmInfo
                    sum cvsum sumwgt;
  by CDW_CD TTB_CD; /* 시군구 읍면동 */
  strata ED_P RTP_CD1; /* 조사구 종류 */
  cluster ED_NUM; /* 조사구 ID */
  var HHSize1-HHSize5 DuType1-DuType4
      HHSize21-HHSize23 DuType21-DuType22;
  weight wHHmd;
  ods output StrataInfo = HHMUMDSTE
             Statistics = HHMUMBHTE;
run;
```

- 점검을 위해 총계추정치 \hat{X}_π (HMUMBHTE), 가구총합 X (TotHHSizeDuType, TotHHSizeDuType2)를 비교할 것. 이때 다음의 통계량을 시군구 · 읍면동별로 나열한 엑셀표를 작성할 것

- | | |
|------------|--|
| ■ 총계추정치 | \hat{X}_π |
| ■ 총계추정치 CV | CVSUM |
| ■ 차이: | $\hat{X}_\pi - X$ |
| ■ 상대차이(%): | $100 \times \frac{\hat{X}_\pi - X}{X}$ |

- 읍면동 표본가구수 m
- 읍면동 표본조사구수 n
- 읍면동 크기추정 $\hat{M} = \sum_s w_{hij}^d$

□ 보조변수 $\left(\sum_{j \in s_I} d_{hij} \mathbf{x}_{hijc} \mathbf{x}_{hijc}' \right)^{-1}$ 산출

- 식 (3-9)의 역행렬 부분을 PROC SURVEYREG을 이용하여 구해볼 것
- 예로 다음의 SAS 프로그램을 통해 $\left(\sum_{j \in s_I} d_{hij} \mathbf{x}_{hijc} \mathbf{x}_{hijc}' \right)^{-1}$ 산출할 수 있음
- $\left(\sum_{j \in s_I} d_{hij} \mathbf{x}_{hijc} \mathbf{x}_{hijc}' \right)^{-1}$ 값은 **HHXPXi** 파일에서 확인할 수 있음
- 월자료 경우에 HHSize#와 DuType# cell이 없는 읍면동도 있어 해당 읍면동의 역행렬이 존재하지 않음

<SAS 예시 3-2>

```
proc surveyreg data = JJT2MHHS1 total = JJEDFrmInfo;
  by CDW_CD TTB_CD; /* 시군구 읍면동 */
  strata CDW_CD TTB_CD ED_PRTP_CD1 /list nocollapse;
                                /* 조사구 종류 */
  cluster ED_NUM; /* 조사구 ID */
  model C88 = HHSize21-HHSize25 DuType21-DuType3
          /noint-solution XPX INVERSES DEFF;
                                /* 지시자 수 제한 */
  weight wHHmd; /* 가구(월)설계가중치 */
  ods output StrataInfo = HHMSTR
             ParameterEstimates = HHParaEst
             invXPX = HHXPXi;
run;
```

□ GREG 조정계수 g_{thijc} 를 산출할 것

- GREG 조정계수 g 는 읍면동별로 산출하고 “SET”을 통해 쌓으면 됨. 산출 개요는 다음과 같이 정리할 수 있음
- 단계 1: 읍면동별 보조변수 분리. 아래의 구성요소의 행은 읍면동

- 보조변수 행렬 $X_s = ((x_{hijc}))$ (m x p)
- 설계가중치 행렬 $D_s = diag(w_{hij}^d)$ (m x m)
- 총합추정오차 벡터 $\underline{X} - \hat{\underline{X}}_\pi$ (p x 1)

여기서

n = 읍면동내 사례(가구)수, p = 고려하는 보조변수수

보조변수, 설계가중치 행렬은 JJT2MHHS1 파일로부터 생성

총합추정오차는 TotHHSIZEDUType, [TotHHSIZEDuType2](#) 와 HMUMBHTE 파일로부터 생성

- 단계 2: 읍면동별 GREG 조정계수 g-weight을 다음의 이름으로 생성할 것

- gHHmNI $g_{jc} = 1 + X_s (X_s' D_s X_s)^{-1} (\underline{X} - \hat{\underline{X}})$ (m x 1)

- 단계 3: 읍면동별 g-weight을 쌓은 후(SET), JJT2MHHS1 파일에 merge 시켜고 비통합 가구 GREG 가중치 변수를 생성할 것

- wHHmg = wHHmd x gHHmNI $w_{hijc}^{gr} = w_{hijc}^d g_{hijc}$

- 단계 4: 새로운 파일을 JJT2MHHW1.sas7bdat 이라 칭할 것

□ GREG 계수 및 가중치에 대해 평가할 것

- 세 가지 가중치 및 요소에 대한 산점도 산출

```
proc sgscatter data=JJT2MHHW1;
    matrix wHHmd gHHmNI wHHmg / diagonal=(histogram);
run; quit;
```

- 가중치 및 요소별 산술통계 및 상자그림 산출. 예로, g-weight에 대해 다음 명령어 수행

```
proc means data=JJT2MHHW1
    n sum mean std
    min p1 p5 p10 p25 p50 p75 p90 p95 p99 max
    maxdec = 2;
    var gHHmNI;
    class 읍면동;
run;
proc means data=JJT2MHHW1
    n sum mean std
    min p1 p5 p10 p25 p50 p75 p90 p95 p99 max
    maxdec = 2;
    var gHHmNI;
run;

proc sgplot data=JJT2MHHW1; vbox gHHmNI / category = 읍면동;
run;
```

3.3.1.3 GREG 조정계수 산출 - 가구원

- 개인기본조사파일과 개인총합파일을 사용

- JJT2PSS1.sas7bdat
- TotSexAge.sas7bdat

3.3.2 통합가중치 방안 1

3.3.3 통합가중치 방안 2

3.4 비교논의

3.4.1 가중치 증감

3.4.2 분산 증감

4. 소지역

5. 시점추정

6. 기간추정

부록

A. 파일