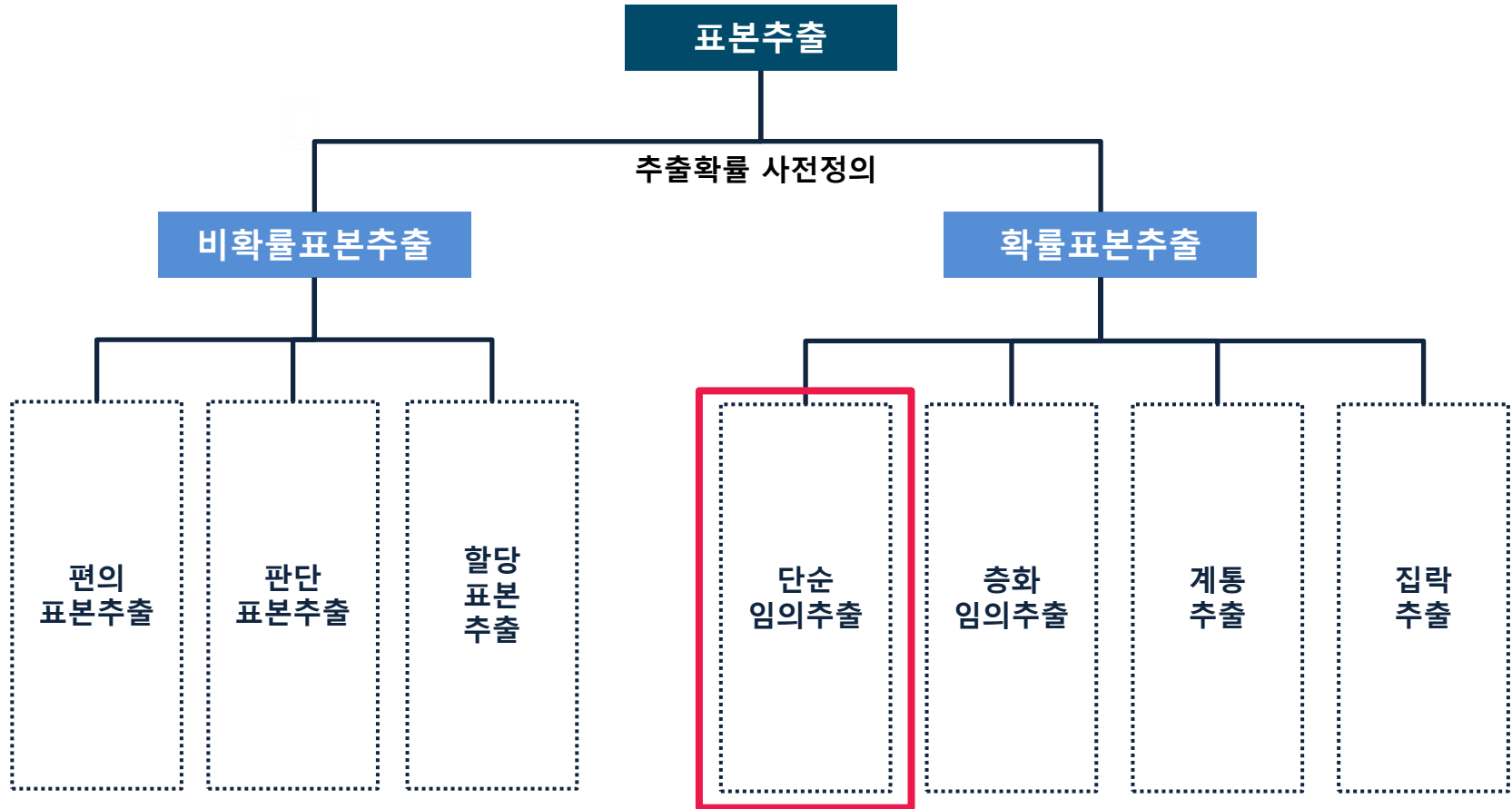

엑셀을 활용한 랜덤 샘플링



Simple random sampling

1. 새로운 열을 하나 만들어서 열 이름(random)을 지정한 뒤, 2번째 행에 “=rand()” 입력 후 enter를 누른다.

	A	B	C	D	E	F	G	H	I
1	id	bweight	lowbw	gestwks	preterm	matage	hyp	sex	random
2	1	2974	0	38.52	0	34	0	2	=rand()
3	2	3270	0	NA	NA	30	0	1	
4	3	2620	0	38.15	0	35	0	2	

	A	B	C	D	E	F	G	H	I
1	id	bweight	lowbw	gestwks	preterm	matage	hyp	sex	random
2	1	2974	0	38.52	0	34	0	2	0.201423
3	2	3270	0	NA	NA	30	0	1	
4	3	2620	0	38.15	0	35	0	2	

2. 자동 채우기를 통해 전체 데이터에 대해 난수를 생성한다.

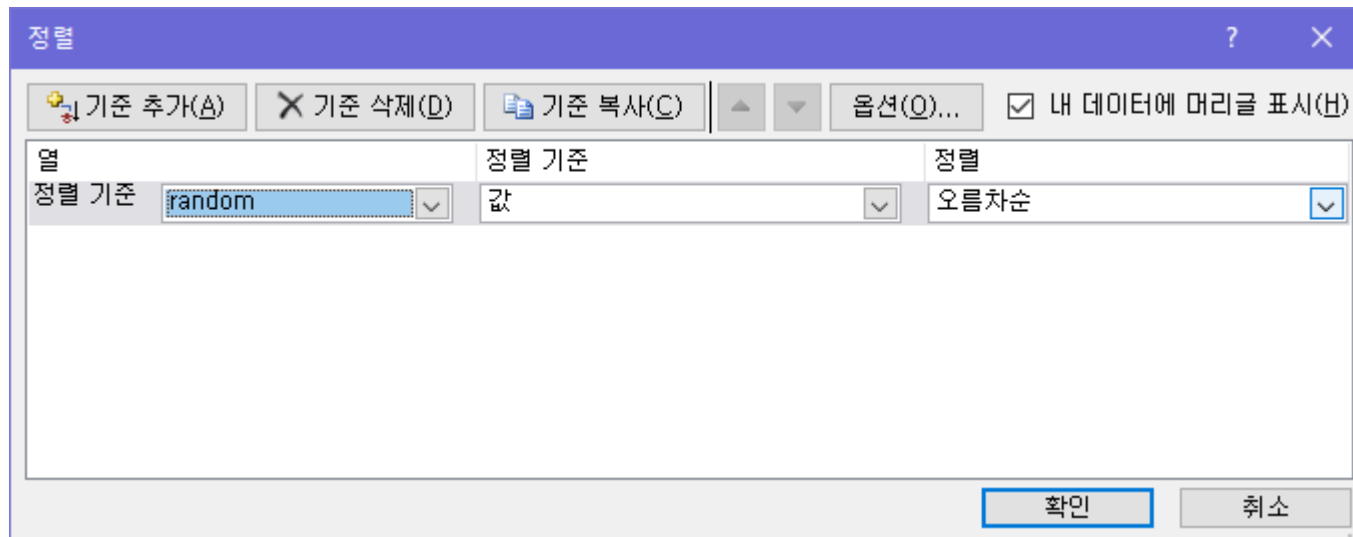
	A	B	C	D	E	F	G	H	I
1	id	bweight	lowbw	gestwks	preterm	matage	hyp	sex	random
2	1	2974	0	38.52	0	34	0	2	0.219685
3	2	3270	0	NA	NA	30	0	1	0.938391
4	3	2620	0	38.15	0	35	0	2	0.205863
5	4	3751	0	39.8	0	31	0	1	0.162122
6	5	3200	0	38.89	0	33	1	1	0.592787

Simple random sampling

3. rand() 함수값이 지속적으로 변경되는 것을 막기 위해, 값을 복사하여 옆에 붙여 넣는다.

	A	B	C	D	E	F	G	H	I	J
1	id	bweight	lowbw	gestwks	preterm	matage	hyp	sex	random	random
2	1	2974	0	38.52	0	34	0	2	0.777894	0.219685
3	2	3270	0	NA	NA	30	0	1	0.743014	0.938391
4	3	2620	0	38.15	0	35	0	2	0.102717	0.205863
5	4	3751	0	39.8	0	31	0	1	0.965773	0.162122
6	5	3200	0	38.89	0	33	1	1	0.689452	0.592787

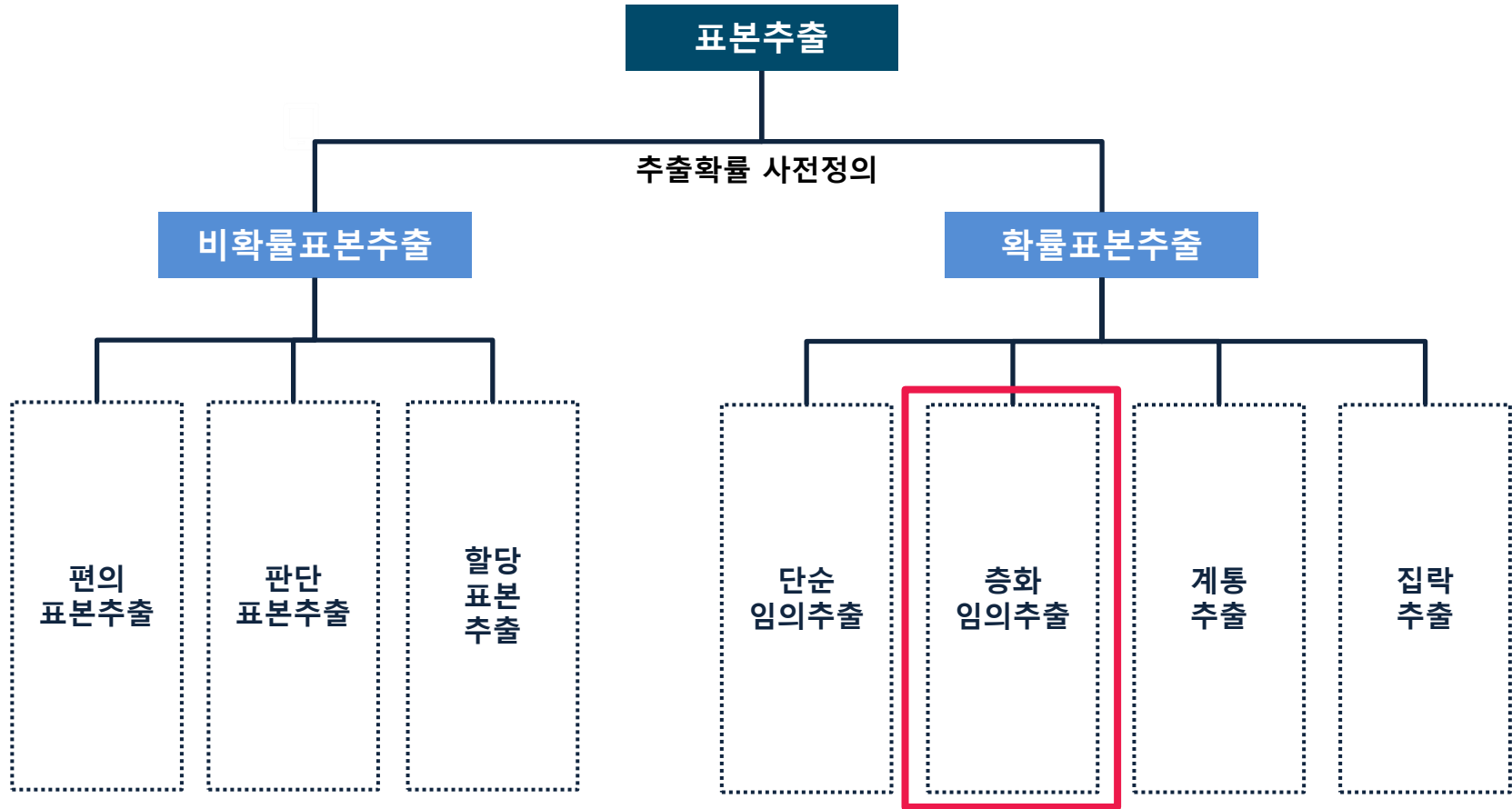
4. 정렬 기능을 통해 복사한 값을 오름차순으로 정렬한다.



▪ Simple random sampling

5. 정렬된 데이터의 위에서부터 추출하려는 개수만큼 행 기준으로 데이터를 선택한다.

	A	B	C	D	E	F	G	H	I
1	id	bweight	lowbw	gestwks	preterm	matage	hyp	sex	random
2	49	3305	0	39.19	0	36	0	1	0.002145
3	369	2784	0	38.23	0	35	0	2	0.0041
4	59	3886	0	40.92	0	40	0	1	0.005705
5	176	3575	0	38.6	0	40	0	1	0.006171
6	341	2855	0	38.07	0	35	0	2	0.007819
7	60	2849	0	37.62	0	33	1	1	0.008108
8	86	4141	0	39.01	0	33	0	1	0.01129
9	337	1325	1	34.67	1	36	0	2	0.011363
10	483	3455	0	38.85	0	35	0	1	0.015328



Stratified random sampling

1. 새로운 열을 하나 만들어서 열 이름(random)을 지정한 뒤, 2번째 행에 “=rand()” 입력 후 enter를 누른다.

	A	B	C	D	E	F	G	H	I
1	id	bweight	lowbw	gestwks	preterm	matage	hyp	sex	random
2	1	2974	0	38.52	0	34	0	2	=rand()
3	2	3270	0	NA	NA	30	0	1	
4	3	2620	0	38.15	0	35	0	2	

	A	B	C	D	E	F	G	H	I
1	id	bweight	lowbw	gestwks	preterm	matage	hyp	sex	random
2	1	2974	0	38.52	0	34	0	2	0.201423
3	2	3270	0	NA	NA	30	0	1	
4	3	2620	0	38.15	0	35	0	2	

2. 자동 채우기를 통해 전체 데이터에 대해 난수를 생성한다.

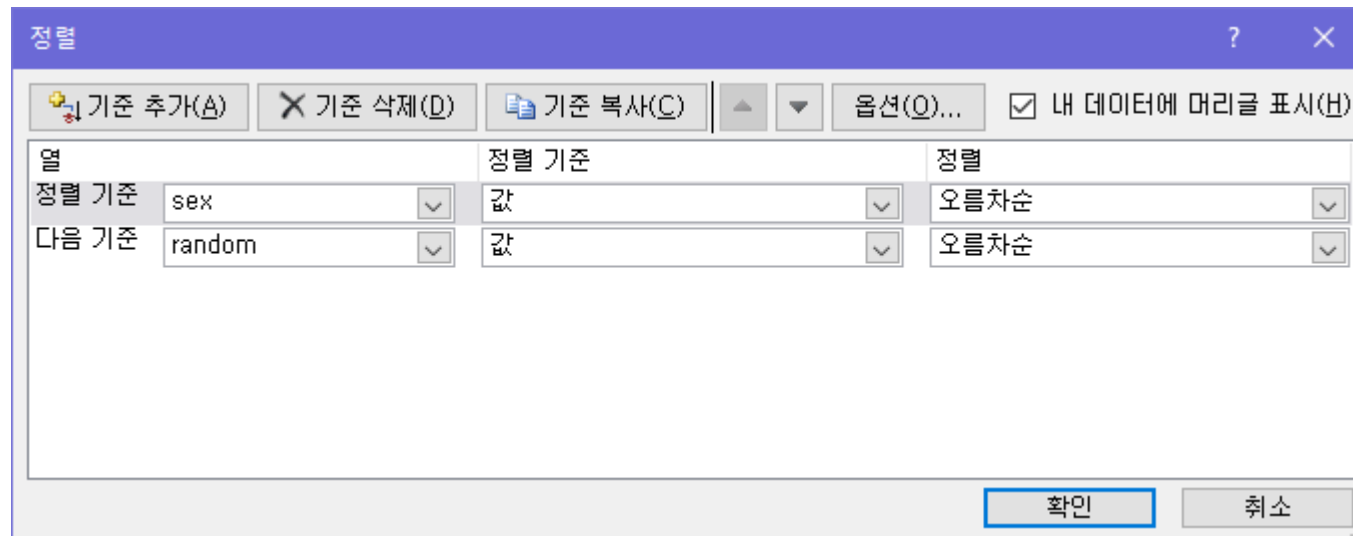
	A	B	C	D	E	F	G	H	I
1	id	bweight	lowbw	gestwks	preterm	matage	hyp	sex	random
2	1	2974	0	38.52	0	34	0	2	0.219685
3	2	3270	0	NA	NA	30	0	1	0.938391
4	3	2620	0	38.15	0	35	0	2	0.205863
5	4	3751	0	39.8	0	31	0	1	0.162122
6	5	3200	0	38.89	0	33	1	1	0.592787

Stratified random sampling

3. rand() 함수값이 지속적으로 변경되는 것을 막기 위해, 값을 복사하여 옆에 붙여 넣는다.

	A	B	C	D	E	F	G	H	I	J
1	id	bweight	lowbw	gestwks	preterm	matage	hyp	sex	random	random
2	1	2974	0	38.52	0	34	0	2	0.777894	0.219685
3	2	3270	0	NA	NA	30	0	1	0.743014	0.938391
4	3	2620	0	38.15	0	35	0	2	0.102717	0.205863
5	4	3751	0	39.8	0	31	0	1	0.965773	0.162122
6	5	3200	0	38.89	0	33	1	1	0.689452	0.592787

4. 정렬 기능을 통해 층화변수 > 복사한 값을 순차적으로 오름차순으로 정렬한다.

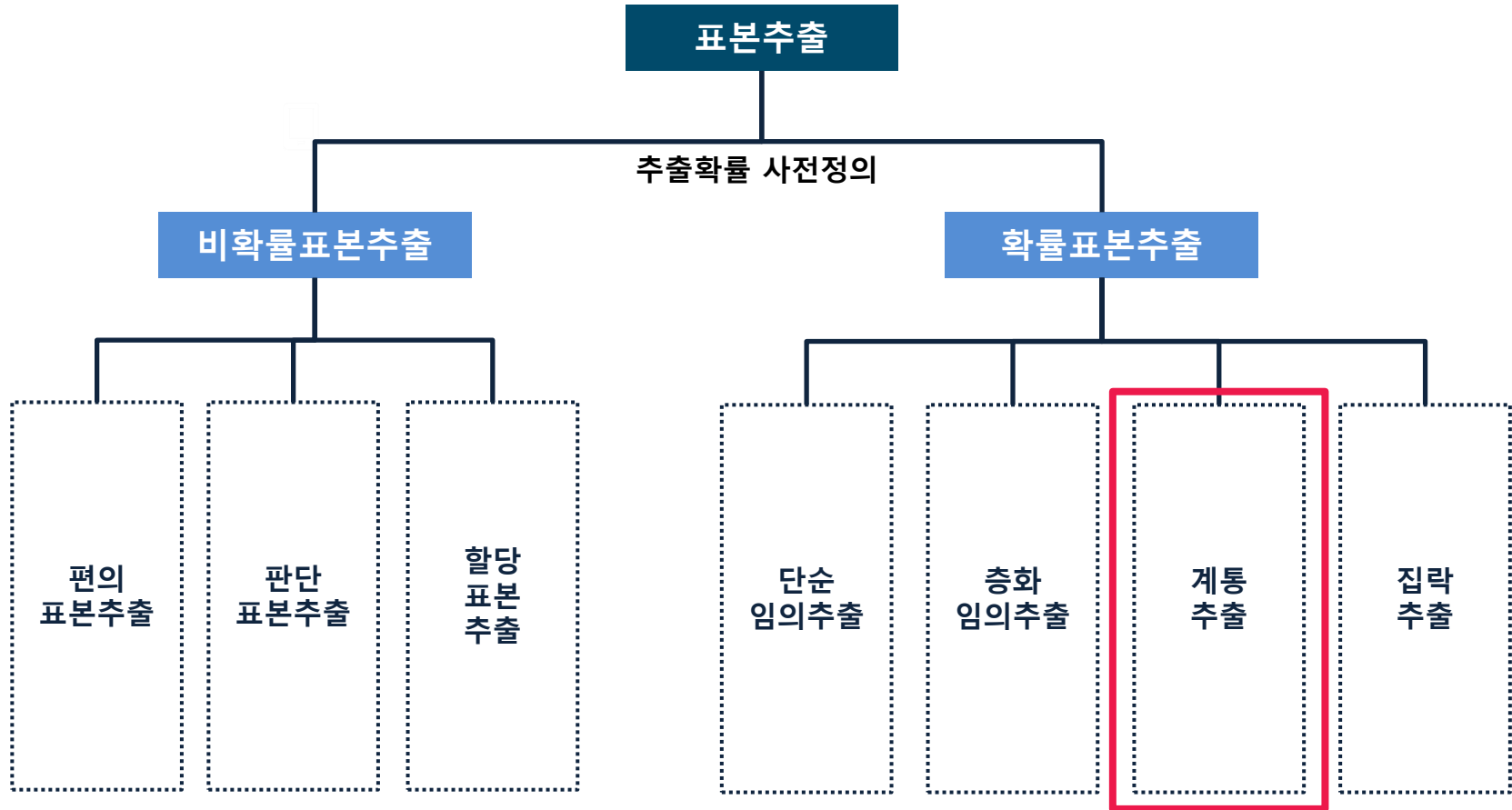


Stratified random sampling

- 정렬된 데이터에서 그룹별로 위에서부터 추출하려는 개수만큼 행 기준으로 데이터를 선택한다.

	A	B	C	D	E	F	G	H	I
1	id	bweight	lowbw	gestwks	preterm	matage	hyp	sex	random
2	78	4512	0	38.68	0	36	0	1	0.004284
3	266	3505	0	38.66	0	39	0	1	0.00767
4	291	3590	0	37.83	0	32	0	1	0.008093
5	360	3646	0	40.03	0	36	0	1	0.012735
6	11	4020	0	40.92	0	37	0	1	0.014092
7	49	3305	0	39.19	0	36	0	1	0.017947
8	95	2671	0	34.87	1	31	0	1	0.021164

	A	B	C	D	E	F	G	H	I
1	id	bweigh	lowbw	gestwks	preterm	matage	hyp	sex	random
266	212	3134	0	38.88	0	35	0	2	0.002073
267	271	3558	0	37.49	0	37	0	2	0.003082
268	134	2959	0	37.26	0	36	1	2	0.003897
269	108	2830	0	37.08	0	41	0	2	0.00522
270	262	1570	1	34.02	1	28	1	2	0.006017
271	260	3695	0	42.86	0	38	0	2	0.01645
272	101	3784	0	39.01	0	32	0	2	0.016711



Systematic sampling

1. 500명 중 50명을 계통추출 하고자 한다면, $k = \text{모집단크기} / \text{표본크기} = 500 / 50 = 10$ 을 구한다. 1~10 중 임의의 숫자를 골라 그 뒤로 k번씩 건너뛰며 추출한다.
2. 새로운 열을 하나 만들어서 열 이름(random)을 지정한 뒤, 2번째 행에 "=rand()" 입력 후 enter를 누른다.

	A	B	C	D	E	F	G	H	I
1	id	bweight	lowbw	gestwks	preterm	matage	hyp	sex	random
2	1	2974	0	38.52	0	34	0	2	=rand()
3	2	3270	0	NA	NA	30	0	1	
4	3	2620	0	38.15	0	35	0	2	

3. 자동 채우기를 통해 1~9번째 개체에 대해 난수를 생성한다.

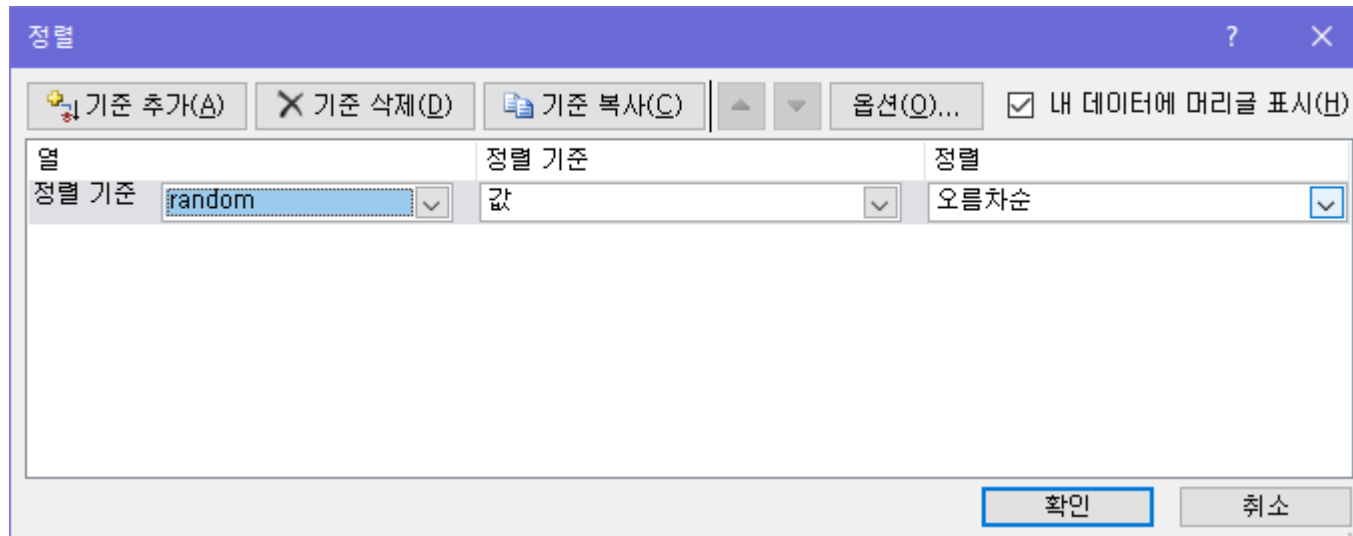
	A	B	C	D	E	F	G	H	I
1	id	bweight	lowbw	gestwks	preterm	matage	hyp	sex	random
2	1	2974	0	38.52	0	34	0	2	0.219685
3	2	3270	0	NA	NA	30	0	1	0.938391
4	3	2620	0	38.15	0	35	0	2	0.205863
5	4	3751	0	39.8	0	31	0	1	0.162122
6	5	3200	0	38.89	0	33	1	1	0.592787

Systematic sampling

3. rand() 함수값이 지속적으로 변경되는 것을 막기 위해, 값을 복사하여 옆에 붙여 넣는다.

	A	B	C	D	E	F	G	H	I	J
1	id	bweight	lowbw	gestwks	preterm	matage	hyp	sex	random	random
2	1	2974	0	38.52	0	34	0	2	0.777894	0.219685
3	2	3270	0	NA	NA	30	0	1	0.743014	0.938391
4	3	2620	0	38.15	0	35	0	2	0.102717	0.205863
5	4	3751	0	39.8	0	31	0	1	0.965773	0.162122
6	5	3200	0	38.89	0	33	1	1	0.689452	0.592787

4. 정렬 기능을 통해 복사한 값을 오름차순으로 정렬한다.



Systematic sampling

5. 계통추출을 시작할 시작점이 선택되면 다시 id를 기준으로 정렬한다.

	A	B	C	D	E	F	G	H	I	J
1	id	bweight	lowbw	gestwks	preterm	matage	hyp	sex	random	
2	4	3751	0	39.8	0	31	0	1	0.269896	시작점
3	1	2974	0	38.52	0	34	0	2	0.420982	
4	2	3270	0	NA	NA	30	0	1	0.426897	
5	5	3200	0	38.89	0	33	1	1	0.512129	
6	8	3773	0	40.21	0	37	0	1	0.533472	
7	6	3673	0	40.97	0	33	0	2	0.553856	
8	3	2620	0	38.15	0	35	0	2	0.624498	

정렬

기준 추가(A) 기준 삭제(D) 기준 복사(C) 옵션(O)... 내 데이터에 머리글 표시(H)

정렬 기준: id 정렬 기준: 값 정렬: 오름차순

확인 취소

Systematic sampling

6. 시작점 옆 셀에 =mod(row() -4, 50) 입력 후 enter를 누른다.

	A	B	C	D	E	F	G	H	I	J	K	L
1	id	bweight	lowbw	gestwks	preterm	matage	hyp	sex	random		systematic	
2	1	2974	0	38.52	0	34	0	2	0.420982			
3	2	3270	0	NA	NA	30	0	1	0.426897			
4	3	2620	0	38.15	0	35	0	2	0.624498			
5	4	3751	0	39.8	0	31	0	1	0.269896	시작점	=mod(row() - 5, 50)	
6	5	3200	0	38.89	0	33	1	1	0.512129			
7	6	3673	0	40.97	0	33	0	2	0.553856			

	A	B	C	D	E	F	G	H	I	J	K
1	id	bweight	lowbw	gestwks	preterm	matage	hyp	sex	random		systematic
2	1	2974	0	38.52	0	34	0	2	0.420982		
3	2	3270	0	NA	NA	30	0	1	0.426897		
4	3	2620	0	38.15	0	35	0	2	0.624498		
5	4	3751	0	39.8	0	31	0	1	0.269896	시작점	0
6	5	3200	0	38.89	0	33	1	1	0.512129		1
7	6	3673	0	40.97	0	33	0	2	0.553856		2
8	7	3628	0	42.14	0	29	0	2	0.909653		3
9	8	3773	0	40.21	0	37	0	1	0.533472		4
10	9	3960	0	42.03	0	36	0	2	0.825707		5
11	10	3405	0	39.33	0	39	0	1			6

Systematic sampling

7. 해당 열에서 필터기능을 통해 0에 해당하는 데이터를 선택한다.

	A	B	C	D	E	F	G	H	I	J	K
1	id	bweigh	lowbw	gestwks	preterm	matage	hyp	sex	random		systema
5	4	3751	0	39.8	0	31	0	1	0.269896	시작점	0
55	54	3079	0	38.44	0	34	1	2			0
105	104	3138	0	39.46	0	40	1	2			0
155	154	3621	0	41.88	0	36	0	2			0
205	204	3005	0	39.67	0	35	0	1			0
255	254	2879	0	37.46	0	37	0	1			0
305	304	3502	0	40.64	0	37	0	1			0
355	354	3435	0	40.48	0	34	0	2			0
405	404	3082	0	39.64	0	37	0	2			0
455	454	2885	0	40.03	0	36	1	2			0

Rexsoft

감사합니다

| 문의

 <http://rexsoft.org>

 rexsoft.help@gmail.com

홈페이지 '질문과 답변' 게시판을 통해 Rex 설치, 다운로드, 기능 등 사용문의를 남겨주세요.
신속하고 친절한 상담을 통해 사용자들의 궁금증에 답변해 드립니다.



Make Analysis Fast and Easy