

4. 표본크기와 표본배분

4.1 표본크기 결정

- 표본크기를 목표허용오차를 기준으로 계산하는 방안과 목표상대표준오차를 이용하여 계산할 수 있는데 이때 사용되는 되는 산식은 아래와 같음
 - 목표허용오차와 표준편차가 각각 d_0 와 S 로 주어졌으면 아래와 같이 계산함

$$n = \frac{\left(\frac{z_{\alpha/2} S}{d_0}\right)^2}{1 + \frac{1}{N} \left(\frac{z_{\alpha/2} S}{d_0}\right)^2}$$

단, $z_{\alpha/2}$ 는 신뢰수준 $(1-\alpha)100\%$ 에서 신뢰계수이고 N 은 모집단크기임

- 기존의 표본조사 데이터에서 상대표준오차를 알 수 있고 여기에 목표상대표준오차를 설정할 경우 필요한 표본크기는 아래 식으로 계산할 수 있음

$$n_{2018} = n_{2017} \cdot \left(\frac{rse_{2017}}{rse_{2018}}\right)^2$$

단, 여기서 rse_{2017} 는 2017년 조사데이터에서 산출한 상대표준오차이고, rse_{2018} 은 2018년 조사의 목표상대표준오차를 나타내며 n_{2017} 은 2017년도 표본크기를 나타냄

- 2018년도 표본크기는 예산규모를 고려하여 2017년 조사와 동일한 규모인 4,500개로 정하고, 이때 집체훈련시간과 집체훈련비용의 조사항목에 대한 기대 상대표준오차를 계산하면 각각 7.41%와 7.67%로 산출되어 표본규모는 적합한 것으로 판단됨. 더불어, 표본크기가 4,500개인 경우에 모비율 추정에서 95% 신뢰수준에서 기대표본오차는 $\pm 1.5\%p$ 이므로 역시 적절한 것으로 판단됨

4.2 표본 배분

- 2017년 조사데이터 중에서 ‘집체훈련 시간’에 대한 분산 분석한 결과는 앞서와 같으며, 결론적으로 업종 대분류와 종사자 규모를 증화변수로 이용하는데 있어서 적합성을 보였으므로 업종 대분류와 종사자 규모를 명시적 증화변수로 사용하여 표본배분을 산출할 것임.

먼저 대분류 업종별 배분법으로 기업체수와 종사자수에 대한 표준편차를 근거로 한 네이만배분법, 기업체수의 제곱근비례배분법, 업종별 20개 우선할당 및 기업체수*표준편차의 0.5승근 비례배분법, 30개 우선할당 후 비례배분법 등을 적용하여 시뮬레이션 하여, 산출된 표본규모에 대한 기대표본오차의 크기를 계산하여 최적의 배분법을 찾아 내고자 함. 각 배분법에 대한 계산식은 아래와 같음

- 네이만 배분법

$$n_h = 4500 \cdot \frac{N_h \cdot S_h}{\sum_{h=1} N_h \cdot S_h} \text{ --- (1)}$$

- 제곱근 비례배분법

$$n_h = 4500 \cdot \frac{\sqrt{N_h}}{\sum_{h=1} \sqrt{N_h}} \text{ --- (2)}$$

- 20+ 0.5승근 비례배분법

$$n_h = 20 + 4160 \cdot \frac{(N_h \cdot S_h)^{0.5}}{\sum_{h=1} (N_h \cdot S_h)^{0.5}} \text{ --- (3)}$$

- 30+ 비례배분법

$$n_h = 30 + 3990 \cdot \frac{N_h}{\sum_{h=1} N_h} \text{ --- (4)}$$

단, 위의 식에서 N_h 와 n_h 는 각각 h 업종 대분류의 모집단 크기와 표본크기를 나타내고, S_h 는 h 업종 대분류에서 상용 종사자수에 대한 표준편차를 의미함

○ 위의 식 (1) ~ 식(4)를 이용하여 업종대분류별로 표본할당을 실시하여 각 배분방법별로 표본 크기를 산출하였으며, 이를 기준으로 기대표본오차를 계산하여 아래 <표 3>에 정리함

<표 3 > 업종 대분류별 표본할당결과와 기대표본오차

업종 대분류	기업체수	분산	할당결과				기대표본오차			
			네이만 배분	제공근 비례	20+0.5 승근	30+비례	네이만 배분	제공근 비례	20+0.5 승근	30+비례
B	266	503.508	2	55	35	37	70.44	12.01	15.75	15.25
C	56144	45947.057	1188	783	691	1452	2.87	3.55	3.78	2.59
D	140	150110.998	5	39	65	34	43.92	13.6	9.08	14.92
E	1732	837.380	5	137	63	74	44.66	8.2	12.37	11.37
F	14657	22888.932	219	400	308	400	6.71	4.93	5.64	4.93
G	21307	57336.176	504	482	457	570	4.4	4.5	4.63	4.13
H	6750	74604.283	182	271	283	201	7.31	5.95	5.82	6.95
I	4881	15158.532	59	231	170	154	12.94	6.42	7.53	7.93
J	7022	143540.269	263	277	336	208	6.05	5.89	5.32	6.83
K	1100	1251564.261	121	110	235	58	8.58	9.05	5.78	12.78
L	2415	28201.695	40	162	143	91	15.68	7.59	8.11	10.28
M	10006	2000663.895	1397	330	748	283	2.48	5.41	3.52	5.86
N	6595	154087.001	256	268	331	197	6.13	5.98	5.36	7.02
P	4111	192583.557	178	212	280	134	7.33	6.69	5.77	8.5
Q	13954	1139.344	46	390	153	383	14.72	4.99	8.04	5.04
R	1129	10585.948	11	111	86	59	30	9.01	10.36	12.67
S	5348	2148.886	24	242	116	165	20.37	6.28	9.18	7.66
합계	157557	185431.856	4500	4500	4500	4500				

- 표본 배분법의 적합성을 비교하기 위해서 기대표본오차의 평균, 표준편차와 범위를 계산하여 <표 4>에 요약하였고, 이들을 기준으로 비교 분석 결과에서 제공근비례배분법이 전체 평균, 표준편차에서 가장 작게 나타남. 따라서 업종대분류의 배분은 제공근비례배분법으로 할당한 결과를 적용하고 이를 기준으로 각 업종 내에서 종사자 규모별 배분을 실시할 것임

<표 4 > 업종대분류의 4가지 배분법의 할당결과에 따른 기대표본오차 통계량 비교

구분	네이만배분	제공근비례	20+0.5승근 비례	30+비례배분
평균	17.92	7.06	7.41	8.51
표준편차	18.89	2.65	3.21	3.78
범위	67.96	10.05	12.23	12.66