1973

통계교육교재

국 민 교 육 헌 장

우리는 민족 중흥의 역사적 사명을 띠고 이 땅에 태어났다. 조상의 빛난 얼을 오늘에 되살려, 안으로 자주 독립의 자세를 확립하고, 밖으로 인류 공영에 이바지할 때다. 이에, 우리의 나아갈 바를 밝혀 교육의 지표로 삼는다.

성실한 마음과 튼튼한 몸으로, 학문과 기술을 배우고 익히며 타고난 저마다의 소질을 계발하고, 우리의 처지를 약진의 발판으로 삼아, 창조의 힘과 개척의 정신을 기른다. 공익과 질서를 앞세우며 능률과 실질을 숭상하고, 경애와 신의에 뿌리박은 상부 상조의 전통을 이어받아, 명랑하고 따뜻한 협동정신을 북돋운다. 우리의 창의와 협력을 바탕으로 나라가 발전하며, 나라의 융성이 나의 발전의 근본임을 깨달아, 자유와 권리에 따르는 책임과 의무를 다하며 스스로 국가 건설에 참여하고 봉사하는 국민 정신을 드높인다.

반공 민주 정신에 투철한 애국 애족이 우리의 삶의 길이며, 자유 세계의 이상을 실현하는 기반이다. 길이 후손에 물려줄 영광된 통일 조국의 앞날을 내다보며, 신념과 긍지를 지닌 근면한 국민으로서, 민족의 슬기를 모아 줄기찬 노력으로, 새 역사를 창조하자.

1968. 12. 5

목 차

Ⅰ. 통 계 개 론

목 차

1. 통계와 통계방법

가. 통계의 의의

첫째 「통계는 기술될 대상을 갖는 수자이다」 그런데 기술될 대상은 시간과 공간의 제약을 받으므로 이러한 의미에서 통계는 구체적인 수자라 할 수 있다. 여기서 기술대상은 보통 복수로 취급되는데 대상의 복수는 곧 집단을 의미하는 것이므로 통계는 집단을 기술하는 수자라고 할 수 있다.

둘째 「통계는 작성되는 수자이다」 통계는 그냥 주어진 수자가 아니고 우리의 노력에 의해서 얻어지는 수자이다.

세째 「통계는 이용되는 수자이다」 통계는 결코 작성을 위해서 작성되는 수자가 아니고 어디까지나 이용하기 위해서 작성되는 수자이다. 보통 통계를 이용하는 주체로서는 개인, 기업, 단체, 정부 국제기구등을 들 수 있다. 그리고 통계가 이용되는 근거는 그것이 현실을 반영하는 수자라는 데에 있다. 즉 어떤 행동지침 내지 정책을 결정하기 위해서는 무엇보다도 현실을 제대로 잘 파악할 필요가 있는데 현실을 반영하는 수자인 통계는 그것을 가능하게 해준다.

나. 통계방법

통계방법은 통계를 작성하며 이용하는 방법을 말한다. 보통 통계의 작성을 통계조사라 하며 통계의 이용을 통계해석이라고 말한다. 따라서 통계방법은 통계조사와 통계해석의 방법으로 구성된다 할 수 있다.

통계방법에서는 개개의 사례를 여러개 모아서 집단을 관찰한다. 그 목적은 개개의 사례에 작용하고 있는 우연적인 요인이 우연성

의 논리인 확률론의 법칙에 따르기 때문이다. 따라서 통계방법의 이론적 기초 확률론에서 찾아지는 것도 바로 이에 기인하는 것이다.

확률론에 의하면 이러한 우연적인 요인은 서로 상쇄 중화작용을 하기 때문에 관찰의 수가 크면 클수록 안전하게 되는 것이다.

다시말해서 우연적 요인의 영향을 받는 사상의 본질적인 작용은 대수의 관찰에 의할 수록 안정적으로 나타나게 되므로 본질적 필연적 요인의 작용을 정확하게 파악하기 위해서는 대수의 사례를 관찰하지 않고서는 안되는 것이다. 이것을 통계에서 대수법칙이라 한다.

통계방법을 대상으로 하는 학문이 통계학이다. 따라서 통계학은 통계조사와 통계해석의 방법을 다루는 학문이라 할 수 있다.

2. 도수분포표 (Table of Frequency Distribution)

도수분포표란 통계조사에서 얻은 지식을 수량적으로 정리하여 통계계열로 만들어 놓은 표를 말한다.

여기서 통계계열이란 통계집단을 장소적, 시간적 및 속성적으로 서로 성질이 같은 것 끼리 분류하여 나열한 수자를 말한다. 이때 분류의 기준이 되는 성질을 분류의 표지라 하며 그 분류의 표지가 장소적인 것이면 장소적계열, 연도별 또는 월별과 같이 시간적인 것일 때에는 시간적계열, 그리고 연령이라든지 신장을 기준으로 하여 분류하였을 경우에는 이를 속성적 계열이라 한다.

그런데 속성적계열 가운데는 연령, 신장과 같이 수량으로 표시할 수 있는 표지가 있고 성별, 직업과 같이 수량으로 표시할 수 없

는 것이 있다. 통계학에서는 이러한 수량으로 표시할 수 있는 표지를 변량이라 하여 수학에서 말하는 변수와 같은 의미를 가지고 있다.

변량중에는 가구인원수와 같이 정수로 표시되는 것이 있는가 하면 신장이라든가 수입액과 같이 단위이하의 소수를 포함하는 것이 있다. 전자를 ~~불연속변량~~이라 하고 후자를 연속변량이라 한다.

또한 이 변량의 전역을 몇개로 구분한 것을 급이라 하고 이 급에 속하는 요소의 수를 도수라고 한다. 따라서 도수분포는 각 급에 도수를 대응시킨 것을 말하며 도수분포표는 도수분포를 표를 형식으로 표현한 것을 말한다.

이 도수분포표에서 한 급의 최대치를 급상한, 최소치를 급하한, 양자를 합쳐서 급한계라 하고 서로 인접하는 두 급의 급하한의 차를 급간격 또는 급폭이라 하고 한급의 급상한과 급하한의 중앙값을 급중심 또는 급대표치라 한다.

다음의 표는 20명의 직공의 임금(단위;원), 21,000, 17,000, 23,000, 33,000, 29,000, 20,000, 26,000, 24,000, 22,000, 17,000, 18,000, 13,000, 23,000, 29,000, 24,000, 21,000, 19,000, 16,000, 23,000, 12,000, 을 급간격을 5,000원으로 하여 도수분포표를 작성한 것이다.

<표 2-1> 도 수 분 포 표

임 금 (급)	직 공 (도수)
(단위 1,000원)	(단위;명)
10 ~ 14	2
15 ~ 19	5
20 ~ 24	9
25 ~ 29	3
30 ~ 35	1

도수분포표에서 급의 상한의 미만 또는 이상의 도수를 합계하여 만든 표를 누적도수분포표라 하며 다음의 표와 같이 작성한다.

<표2-2>

누적도수분포표(미만)

급	누적도수
10 미만	0
15 〃	2
20 〃	7
25 〃	16
30 〃	19
35 〃	20

<표2-3>

누적도수분포표(이상)

급	누적도수
10 이상	20
15 〃	18
20 〃	13
25 〃	4
30 〃	1
35 〃	0

이상에서 설명한 것 이외에 자급의 도수를 총도수에 대한 백분비로 표시한 백분비 도수분포표와 이 표에서 백분비 도수를 누적하여 만든 백분비 도수분포표가 있다.

3. 대표치와 산포도

통계집단의 특징을 일면적으로 규정할 수 있는 일정한 수치로 대표치(평균), 산포도, 비대칭도(외도), 첨도가 있는데 여기서는 이중에서 대표치와 산포도에 대해서만 취급하기로 한다.

가. 대표치

일반적으로 대표치란 한 통계집단을 구성하고 있는 변량의 분포의 위치를 나타내는 것으로 그 결정방법을 기준으로 해서 보

통 계산적 대표치와 위치적 대표치로 나눈다. 계산적 대표치는 집단의 변량을 전부 사용하여 산출결정하는 것을 말하며 이에 속하는 것으로는 산술평균, 기하평균, 조화평균등이 있다.

위치적 대표치는 도수분포의 특정한 위치를 차지하는 한 변량값에 의해서 결정되는 것으로 중위수, 최빈수등이 이에 속한다.

이제 이중에서 가장 널리 사용되고 있는 산술평균과 중위수에 대해서 알아 보기로 하자.

(1) 산술평균

산술평균이란 변량의 총합계를 총도수로 나눈 값을 말하며 이것을 일반식으로 나타내면 다음과 같다.

$$\mu = \frac{\sum X_i}{n} \quad \cdots\cdots\cdots\cdots\cdots\cdots \langle 3-1 \rangle$$

여기서 $\sum$(시그마)는 총 합계를 말하며 X_i는 각 변량값을 n는 변량의 총도수를 말한다. 그런데 이 식은 통계자료를 정리하지 않은 단순계열에서의 계산식을 말하며 통계자료가 정리되어 도수분포표로 작성되었을 때에는 다음의 계산식을 사용한다.

$$\mu = \frac{\sum f_i X_i}{\sum f_i}$$

$$= \frac{\sum f_i X_i}{n} \quad \cdots\cdots\cdots\cdots\cdots\cdots \langle 3-2 \rangle$$

여기서 fi와 Xi는 각급의 도수와 변량값을 나타낸 것이며 $\sum f_i$는 도수의 총 합계 즉 총도수 n를 뜻하는 것이다. 이제 <표2-1>에서 산술평균을 구하여 보면 다음과 같다.

<표3 - 1>

급	급중심 (X_i)	도 수 (f_i)	$f_i \times X_i$
10 ~ 14	12	2	24
15 ~ 19	17	5	85
20 ~ 24	22	[illegible]	198
25 ~ 29	27	3	81
30 ~ 34	32	1	32
합 계		$\sum f_i = 20$	$\sum f_i \cdot X_i = 420$

$$\mu = \frac{\sum f_i \cdot X_i}{\sum f_i} = \frac{420}{20} = 21$$

$\therefore$ 21,000원

그런데 계산을 간편하게 하기 위하여 다음과 같은 공식을 사용하면 아주 편리하다.

$$\mu = X_a + \frac{\sum f_i \times U_i}{n} \times w \cdots\cdots\cdots\cdots \text{<3 - 3>}$$

여기서 X_a는 임의의 변량(보통 도수가 가장 큰 급중심값을 택한다), fi는 자급의 도수, W는 급간격, 그리고 Ui는 급중심값 Xi와 임의의 변량 Xa와의 편차를 다시 급간격 W로 나눈 값 즉 $U_i = \frac{X_i - X_a}{W}$를 말한다. 이제 이 공식을 사용하여 <표2 - 1>에서 산술평균을 계산하면 다음과 같다.

<표 3 - 2>

X_i	f_i	d_i	$f_i \cdot d_i$
12	2	-2	-4
17	5	-1	-5
22	9	0	0
27	3	1	3
32	1	2	2
Σ	20		-4

$$\mu = 22 + \frac{(-4)}{20} \times 5$$

$$= 22 + (-1)$$

$$= 21 \quad (\because \begin{matrix} X_a = 22 \\ w = 5 \end{matrix})$$

$\therefore$ 21,000 원

(2) 중위수

중위수란 변량을 크기의 순으로 나열 하였을 때 중앙에 오는 값을 말한다. 그런데 변량의 총수가 홀수일 때에는 중앙에 오는 값은 하나이나 짝수일 때에는 중앙에 오는 값은 둘이 있게 된다. 따라서 이때에는 중앙에 오는 두 값을 산술평균하여 이를 중위수로 한다.

예컨데 2, 4, 8, 11, 13, 15의 6개 변량값의 중위수는 즉 $\frac{1}{2}(8+11) = 9.5$로 한다. 이것은 단순계열(정리되지 않은 자료)에서의 중위수 산출방법인 데 도수분포표에서 중위수를 산출고자 할 때에는 다음의 계산식을 사용하여야 한다.

$$Me = X_L + \frac{\frac{n}{2} - n_L}{fo} \times W \cdots\cdots\cdots\cdots \langle 3-4 \rangle$$

여기서 X_L은 중위수가 들어있는 급의 급하한, fo는 중위수가 들어있는 급의 도수, n_e는 중위수가 들어있는 급의 바로 앞의

급까지의 누적도수, W는 급간계, n는 총도수를 각각 말한다.

이제 <표2 - 1>의 도수분포표에서 중위수의 값을 계산해 보면 다음과 같다.

우선 중위수가 들어있는 급을 알아야 하는데 $\frac{n}{2}$는 10이므로 「20 ~ 24」의 급에 중위수가 들어 있음을 알 수 있다. 따라서 X_1는 25, f_0는 9, m_0는 7이다. 그리고 W는 5, n는 20이므로 이를 <3 - 4>식에 대입하면 중위수값 21.6을 얻는다.

따라서 임금의 중위는 21,600원이 된다.

<표3 - 3>

급	누 적 도 수
10 ~ 14	2
15 ~ 19	7
20 ~ 24	16
25 ~ 29	19
30 ~ 34	20

$$M_e = 20 + \frac{10-7}{9} \times 5$$

$$= 21.6$$

$$\therefore \underline{21,600 \text{원}}$$

나. 산포도

산포도란 통계집단을 구성하고 있는 각 변량이 평균값을 중심으로 어느 정도 떨어져 있는가 하는 산포의 정도를 나타내는 수치이다.

이제 어느 회사에서 타자수를 채용하기 위하여 2명의 지원자 A와 B에게 실기시험을 치루었다고 하자. 다섯번의 실기 결과

산포도 < 절대적 상대적 <

A는 7, 10, 8, 9, 6자의 오자가, B는 5, 14, 2, 2, 17자의 오자가 각각 나왔다. 여기서 두사람 모두 평균 오자는 8자가 나왔으나 B는 A보다 오자의 범위가 크므로 오자를 더 많이 낼 가능성을 가지고 있다. 따라서 B보다는 A를 채용하는 것이 나을 것이다. 이와같이 산포도는 대표치를 보완하여 집단의 특징을 좀더 명확하게 기술하는 데 필요하다.

산포도에는 이것을 원단위와 동일한 절대수로 표시하는가 또는 그것의 평균에 대한 비율로 표시 하는가에 따라 절대적 산포도와 상대적 산포도로 나눈다. 절대적 산포도에는 레인즈(Range), 4분위편차, 평균편차, 표준편차등이 있으며 상대적 산포도에는 4분위편차계수, 평균편차계수, 변이계수(Coefficient of Variation) 등이 있다.

절대적 산포도중 레인즈나 4분위편차는 어느 특정한 두변량, 또는 집단의 여러 변량중 그 일부를 취하여 계산하는 방식이므로 집단의 분포상태를 완전하게 나타낸다고는 할 수 없다.

평균편차와 표준편차는 이러한 점을 고려하여 집단의 모든 변량을 취하여 계산하는 방식이다. 그런데 평균편차는 「평균에서의 각 변량의 편차에 절대값을 취하여 이를 산술평균」한 것이므로 편차에 절대값을 취하는 불편이 있다.

표준편차는 이러한 불편을 제거한 방식으로 수리적인 처리가 용이하기 때문에 가장 널리 이용되고 있다.

(1) 표준편차(Standard Deviation)

평균에서의 각변량의 편차의 자승값을 산술평균 한 것을 분산(Variance)이라 하고 이를 다시 자승근 한 것을 표준편차(Standard Deviation)라 한다.

표준편차는 보통 σ 로 표시하며 이를 일반식으로 나타내면 다음과 같다.

$$\sigma = \sqrt{\frac{\sum(X_i - \mu)^2}{n}} \quad \cdots\cdots\cdots\cdots \langle 3-5 \rangle$$

이것의 자승값 분산은 다음과 같다.

$$\sigma^2 = \frac{\sum(X_i - \mu)^2}{n} \quad \cdots\cdots\cdots\cdots \langle 3-6 \rangle$$

이제 위의 예에서 두 지원자 A와 B의 타자 실기결과인 오자의 표준편차를 구하여 보기로 하자.

<표 3 - 4>

A	B
7	5
10	14
8	2
9	2
6	17
$\sum X_i = 40$	$\sum X_i = 40$
$\mu = \frac{40}{5} = 8$	$\mu = \frac{40}{5} = 8$

A의 표준편차

$$\sigma_A^2 = \frac{(7-8)^2 + (10-8)^2 + (8-8)^2 +}{5}$$

$$\frac{(9-8)^2 + (6-8)^2}{5}$$

$$= \frac{(-1)^2 + (2)^2 + (0)^2 + (1)^2 + (2)^2}{5}$$

$$= \frac{1+4+0+1+4}{5} = \frac{10}{5} = 2$$

$$\therefore \sigma_A = \sqrt{2} = \underline{\underline{1.41}}$$

B의 표준편차

$$\sigma_B^2 = \frac{(5-8)^2 + (14-8)^2 + (2-8)^2 + (2-8)^2 + (17-8)^2}{5}$$

$$= \frac{(-3)^2 + (6)^2 + (-6)^2 + (-6)^2 + (9)^2}{5}$$

$$= \frac{9+36+36+36+81}{5} = \frac{198}{5} = 39.6$$

$$\therefore \sigma_B = \sqrt{39.6} = \underline{\underline{6.29}}$$

위에서 A , B 두사람의 표준편차를 비교하여 보면 B는 A보다 훨씬 크다. 그러므로 A는 B보다 비교적 안정된 실력을 가졌다고 평가할 수 있을 것이다.

이상에서의 표준편차 계산은 단순계열에서의 계산방식으로 만약 통계자료가 도수분포표의 형태를 취하고 있을 때에는 다음의 계산식을 적용하여야 한다.

$$\sigma = \sqrt{\frac{\sum f_i (X_i - \mu)^2}{\sum f_i}} \cdots\cdots\cdots\cdots \quad <3-7>$$

여기서 X_i는 급중심값을, fi는 도수를 말한다. <3 - 7 >식을 사용하여 <표2 - 1 >의 도수분포표에서 표준편차를 계산해 보면 다음과 같다.

<표3 - 5 >

X_i	f_i	$X_i - \mu$	$(X_i - \mu)^2$	$f_i (X_i - \mu)^2$
12	2	-9	81	162
17	5	-4	16	80
22	9	1	1	9
27	3	6	36	108
32	1	11	121	121
Σ	20			480

$$\sigma = \sqrt{\frac{\sum f_i (X_i - \mu)^2}{n}}$$

$$= \sqrt{\frac{480}{20}}$$

$$= \sqrt{24}$$

$$= 4.9$$

$\therefore$ 4,900원

그런데 ⑤식과 ⑦식을 간편하게 하기 위하여 이것을 풀면 각각 ⑧식과 ⑨식이 된다.

⑤식 : $\sigma = \sqrt{\frac{\sum(X_i - \mu)^2}{n}}$

$= \sqrt{\frac{\sum X_i^2}{n} - \mu^2}$

$= \sqrt{\frac{\sum X_i^2}{n} - (\frac{\sum X_i}{n})^2}$ ………………………… <3 - 8>

⑦식 : $\sigma = \sqrt{\frac{\sum f_i (X_i - \mu)^2}{\sum f_i}}$

$= \sqrt{\frac{\sum f_i X_i^2}{\sum f_i} - \mu^2}$

$= \sqrt{\frac{\sum f_i X_i^2}{\sum f_i} - (\frac{\sum f_i \cdot X_i}{\sum f_i})^2}$

$= \sqrt{\frac{\sum f_i X_i^2}{n} - (\frac{\sum f_i X_i}{n})^2}$ ……………… <3 - 9>

이제 <3 - 9>식을 사용하여 <표2 - 1>에서 표준편차를 계산해 보면 다음과 같다.

<표3 - 6>

X_i	f_i	X_i^2	$f_i X_i$	$f_i X_i^2$
12	2	144	24	288
17	5	289	85	1,445
22	9	484	198	4,356
27	3	729	81	2,187
32	1	1024	32	1,024
$\sum$	20		420	9,300

$\sigma = \sqrt{\frac{\sum f_i X_i^2}{n} - (\frac{\sum f_i X_i}{n})^2}$

$= \sqrt{\frac{9,300}{20} - (\frac{420}{20})^2}$

$= \sqrt{465 - 441}$

$= \sqrt{24}$

$= \underline{4.9}$

이상의 방식은 그 취급하는 수치가 크면 계산이 힘들다. 그러므로 이를 쉽게 계산하기 위하여 <3 - 3>식의 산술평균 계산의 간편식과 같은 개념을 도입하여 이를 계산하면 쉽게 풀수 있다.

$$\sigma = W \times \sqrt{\frac{\sum f_i U_i^2}{n} - \left(\frac{\sum f_i X_i}{n}\right)^2} \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad <3-10>$$

여기서 W는 급간격, U_i 는 급중심값 X_i 와 임의의 변량 X_a (보통 도수가 가장 큰 급중심 값을 택한다)와의 편차를 급간격 W로 나눈값, 즉 $U_i = \frac{X_i - X_a}{W}$ 를 말한다.

이제 <3 - 10>식을 사용하여 표준편차를 계산하여 보면 다음과 같다.

<표3 - 7>

X_i	f_i	U_i	U_i^2	f_iU_i	$f_iU_i^2$
12	2	-2	4	-4	8
17	5	-1	1	-5	5
22	9	0	0	0	0
27	3	1	1	3	3
32	1	2	4	2	4
$\sum$	20			-4	20

$$\sigma = W\sqrt{\frac{\sum f_i U_i^2}{n} - \left(\frac{\sum f_i U_i}{n}\right)^2}$$

$$= 5 \times \sqrt{\frac{20}{20} \quad \left(\frac{-4}{20}\right)^2}$$

$$= 5 \times \sqrt{1 - \left(-\frac{1}{5}\right)^2}$$

$$= 5 \times \sqrt{1 - \frac{1}{25}}$$

$$5 \times \sqrt{\frac{24}{25}}$$

$$= 5 \times \frac{\sqrt{24}}{5} = \sqrt{24}$$

$$= \underline{4.9}$$

(2) 변이계수 (Coefficient of Variation)

절대적 산포도는 서로 다른 단위를 가지는 이질적인 두 집단의 분포상태를 비교할 수 없다. 따라서 이를 서로 비교할 수 있도록 그 집단의 평균값에 대한 비율로 표시한 것이 상대적 산포도이다.

상대적 산포도중에서 가장 대표적인 것은 변이계수이며 이것은 표준편차를 산술평균으로 나누어 여기에 100을 곱하여 보통 평균에 대한 백분비로 표시한다.

$$V = \frac{\sigma}{\mu} \times 100 \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad \langle 3-11 \rangle$$

<3 - 11>식을 사용하여 <표3 - 1>에서 변이계수를 구하면 다음과 같다.

$$V = \frac{4.9}{21} \times 100$$
$$= 19.6\,\%$$

이것은 표준편차가 평균값의 19.6% 만큼 크기를 가졌다는 뜻이다.

변이계수의 성질과 그 표준편차와의 관계는 다음의 표에 나타낸 간단한 예에 의해서 충분히 이해할 수 있을 것이다.

<표3 - 8>

	X_i	μ	σ	V
A	1 , 2 , 3 , 4 , 5	3	1.41	47.0
B	10, 20, 30, 40, 50	30	14.1	47.0

4. 확률과 확률분포

가. 확 률

확률은 우연적 사상을 주제로 하는 것이며 이것은 미리 알 수도 없으며 또 지배할 수도 없는 상태에서 발생한다. 그러나 이미 알고 있는 어떤 공통적인 원인에 의하여 발생하는 경우에는 몇가지의 사상이 일어날 수 있는가 하는 것은 알 수 있다. 또한 어떤 사상이 포함되어 있다면 그 원인에 의하여 그 사상이 나타날지도 모른다고 생각할 수 있다. 이와같이 어떤 우연적 사상이 주어진 조건하에서 일어나는 출현가능성의 정도를 수량적으로 규정한 것이 즉 확률이다.

확률은 실험 또는 경험에 의하여 정의된 경험적확률과 선험적으로 정의된 선험적확률이 있다.

경험적확률의 정의는 다음과 같다.

N가 대단히 큰 수일 적에 N회의 실험 또는 경험에 있어서 사상 E가 일어난 회수가 r일 적에 $\frac{r}{N}$를 일회의 실험 또는 경험에 있어서 E가 일어날 확률이라고 한다. 또는 같은 확실성을 가지고 일어날 수 있는 경우의 수가 전부 n, 그중 사상E가 일어날 수 있는 경우의 수가 r이라 하면 E가 일어나는 확률은 $\frac{r}{n}$ 인 것이다.

이와같이 확률은 비율로 계산되며 확률을 알기 위해서는 순열과 조합의 지식이 필요하다. 따라서 먼저 순열과 조합에 대하여 알아 보자.

(1) 순 열

1, 2, 3의 3개의 숫자가 있다. 이것을 여러가지로 배

임하여 몇가지의 세자리수를 만들 수 있는가를 생각해 보면 먼저 첫자리에 놓을 수 있는 수는 3가지, 두째자리에 놓을 수 있는 수는 첫자리에 놓을 수 하나를 뺀 나머지 2개가 있을 것이다.

이와같이 헤나가면 총 방법의 수는 $3 \times 2 \times 1 = 6$ 이다.

즉 123, 132, 213, 231, 312, 321 이다. 또 3가지의 수 중에서 2개만을 뽑아 이를 나열하는 방법은 첫자리에 올 수 있는 수는 3가지 두째자리에 올 수 있는 수는, 첫자리에 갈 수 하나를 뺀 2가지이므로 즉 $3 \times 2 = 6$이다.

이와같이 어떤 수에서 몇개를 뽑아 나열하는 방법의 수를 순열이라 하며 이를 일반식으로 나타내면 다음과 같다.

$$N_{pr} = N(N-1) \cdots\cdots (N-r+1) \cdots\cdots \quad <4-1>$$

또 어떤 수 전부를 나열하는 방법은

$$N(N-1)(N-2) \cdots\cdots (N-r+1) \cdots\cdots X_2 \; X_1$$

이며 이것을 N의 계승이라 하여 N!로 표시한다.

(2) 조 합

이번에는 A B C D 4분자를 분자의 순서에는 관계없이 2분자씩 취하여 조합(Combination)시키는 방법을 생각해 보면 즉 AB, AC, AD, BC, BD, CD인 6가지가 있다. 즉 AB와 BA와의 순서는 관계없으므로 이를 같은 것으로 생각하는 것이다. 일반적으로 N개의 서로 다른 것에서 r개의 것을 취하여 조합시키는 방법의 수는

$$NC_r = \binom{N}{r} = \frac{N(N-1) \cdots\cdots (N-r+1)}{r!}$$

$$= \frac{N!}{r!(N-r)!} \cdots\cdots \quad <4-2>$$

$_5P_5 = 5\times4\times3\times2\times1 = 120$

20명 학생중 3명을 뽑는법

$_{20}C_3 = \frac{20\times19\times18}{3\times2\times1} = 1140$

으로 표시한다. 따라서 위의 예를 이 계산식에 적용하여 풀면

$$_4C_2 = \frac{4\times3}{2\times1} = \frac{12}{2} = 6$$

※ 알파벳 5자를 3자씩 암기를 만드는 방법수

$_5P_3 = 5\times4\times3 = 60$개

순열

이 된다.

(3) 확률의 성질

확률의 정의는 이미 위에서 하였다. 따라서 여기에서는 확률의 정의에서 유도되는 간단한 성질에 대하여 알아보기로 한다.

(가) 사상E의 확률은 0과 1 사이에 있는 수치이다. 그리고 사상E가 결코 나타나지 않는 확률은 0이며 반드시 나타나는 확률은 1이다.

$0 \le Prob \le 1$

예를 들면 주사위를 던졌을 때에는 1, 2, 3, 4, 5, 6의 눈 이외의 눈은 결코 나타나지 않으므로 주사위를 던져서 1내지 6 이외의 눈이 나올 확률은 0이며 3의눈이 나올 확률은 $\frac{1}{6}$ 이다. 또 1내지 6의 눈이 나올 확률은 1이다. 경험적인 확률에 대해서도 같은 말을 할 수 있다.

(나) 사상E가 나타나지 않는 사상을 사상E의 여사상 (Complementary event)이라고 한다. 여사상의 확률은 사상E의 확률을 1에서 뺀 값이다.

예를 들면 주사위를 던져서 4이하의 눈이 나타나는 확률은 $\frac{4}{6}$ 이며, 이것과 여사상인 5, 6의 눈이 나올 확률은

$1 - \frac{4}{6} = \frac{2}{6} = \frac{1}{3}$ 이다.

(다) 사상 E_1 과 E_2 가 배반사상 (exclusive event)일 때, 즉 사상 E_1 과 E_2 가 동시에 나타날 수 없는 사상일 때에는 E_1 과 E_2 중의 어떤 하나가 나타나는 확률은 E_1 의 확률과 E_2 의 확률의 합계에 의하여 부여된다.

여사상의
확률 = 1 — 事象의 확률
2 배반사상 Prob(E1) + Prob(E2) = 가법정리
3 독립사상. 확률은 Prob(E1) × Prob(E2) 승법정리

이것을 확률의 가법정리 (addition Theorem)라고 한다.

예를 들면 주사위를 던져서 1 , 2 , 3이 나오는 것은 서로 배반사상이며 각눈이 나올 확률은 $\frac{1}{6}$ 이므로 확률은 $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$ 이다.

(라) 사상 E_1 과 E_2 가 동시에 나올 수 있는 확률은 E_1 이 나올 확률과 E_1 이 나올 조건하에서 E_2 가 나올 확률을 곱하여 얻는다. 이것을 확률의 승법정리 (Multiplication Theorem)이라고 한다. 그리고 E_1 이 나올 조건하에서 E_2 가 나올 확률을 조건부 확률이라 하며 만약 E_1 나올 조건에 관계없이 E_2 가 나올 수 있다면 E_1 과 E_2 가 나올 확률은 E_1 이 나올 확률에 E_2 가 나올 확률을 곱하면 된다. 이때 E_1 과 E_2 와의 관계를 독립사상 (independant event)이라 한다.

예를 들면 백구가 3개 , 흑구가 2개 들어있는 주머니에서 2개의 구슬을 뽑는다면 나올 수 있는 모든 사상은 다음과 같다.

이제 백구를 W로 , 흑구를 B로 표시한다면

	첫번째	두번째
E_1	W	W
E_2	W	B
E_3	B	W
E_4	B	B

여기서 만약에 첫번째 구슬을 뽑은 후 주머니 속에 다시 넣지 않고 두번째 구슬을 뽑는다면 두번째 구슬이 나올 확률은 첫번째 구슬이 백구이냐 , 흑구이냐에 따라 다를 것이다. 이때 이것을 비복원 추출 (Without

Replacement)라 하며 이때 각 사상이 나올 확률은 다음과 같이 계산한다.

$$P(E_1) = P(W) \times P(W/W) = \frac{3}{5} \times \frac{2}{4} = \frac{6}{20}$$

$$P(E_2) = P(W) \times P(B/W) = \frac{3}{5} \times \frac{2}{4} = \frac{6}{20}$$

$$P(E_3) = P(B) \times P(W/B) = \frac{2}{5} \times \frac{3}{4} = \frac{6}{20}$$

$$P(E_4) = P(B) \times P(B/B) = \frac{2}{5} \times \frac{1}{4} = \frac{2}{20}$$

그리고 위의 모든 사상이 나올 확률은 반드시 1이된다.

즉 $P(E_1 E_2, E_3, E_4) = P(E_1) + P(E_2) + P(E_3) + P(E_4)$

$$= \frac{6}{20} + \frac{6}{20} + \frac{6}{20} + \frac{2}{20}$$

$$= \frac{20}{20} = 1$$

이때 사상 E_1, E_2, E_3, E_4 는 모두 서로 배반사상이다.

다음은 주머니에서 첫번째 구술을 뽑아서 본다음 다시 주머니에 넣고 두번째 구술은 뽑는다면 첫번째 구술이 어떤것이 뽑히던지 상관없다. 이때 이것을 복원 추출법 (With Replacment)이라 하며 이때 각 사상이 나올 확률은 다음과 같이 계산한다.

$$P(E_1) = P(W) \times P(W) = \frac{3}{5} \times \frac{3}{5} = \frac{9}{25}$$

$$P(E_2) = P(W) \times P(B) = \frac{3}{5} \times \frac{2}{5} = \frac{6}{25}$$

$$P(E_3) = P(B) \times P(W) = \frac{2}{5} \times \frac{3}{5} = \frac{6}{25}$$

$$P(E_4) = P(B) \times P(B) = \frac{2}{5} \times \frac{2}{5} = \frac{4}{25}$$

여기서도 모든 사상이 나올 확률은 반드시 1이되야 한다.

즉, $P(E_1, E_2, E_3, E_4) = \frac{9}{25} + \frac{6}{25} + \frac{6}{25} + \frac{4}{25}$

$$= \frac{25}{25} = 1$$

나 . 확률분포

확률변수가 취할 수 있는 값과 이들 값의 확률을 같이 배열한 것을 확률분포 (Probability distribution) 라 하며 확률분포의 가장 대표적이며 널리 이용되고 있는 것은 이항분포 (Binomial distribution) , 포아슨분포 (Poisson distribution) 및 정규분포 (Normal distribution) 등 세가지가 있다. 그런데 확률분포를 알기 위해서는 확률변수와 기대치에 대하여 먼저 알아둘 필요가 있다.

(1) 확률변수

확률변수 (Random Voriable) 란 확률을 가지고 있는 많은 값을 취할 수 있는 변수를 말한다. 확률변수는 우연적변수 (Statistic or Chance Voriable) 이라고도 하는 데 예를 들면 주사위를 던졌을 때 나타나는 눈을 확률변수라 한다. 왜냐 하면 그것은 $\frac{1}{1}$ 의 확률값을 가지고 값 1 , 2 , 3 , 4 , 5 , 6 을 취할 수 있기 때문이다. 또한 어떤 것은 이와 같은 정수의 값을 취하지 않고 소수점이하의 값을 갖는 변수도 있다.

그러므로 이것에 의하여 확률변수를 비연속적인 것과 연속적인 것으로 나눈다.

(2) 기대치

확률분포의 가장 중요한 특성치의 하나는 확률변수의 기대치 (Expectation) 인 데 이것은 확률변수의 산술평균과 같은 의미를 갖고 있다.

기대치는 모든 가능한 확률변수와 그 각각의 확률을 곱하여 합산 한것을 말하며 이것은 즉 분포의 중심적 경향을 특징지어 준다.

기대치
변수에 대해서
편차치
3.1
$\frac{1}{6}$ (1×2+3 ... +6)

우선 확률변수가 비연속적인 것일 경우에는 확률변수(X)의 기대치 E(X)는 다음식에 의하여 계산한다.

$$E(X) = \mu = \sum X_i P_i \cdots\cdots\cdots\cdots\cdots\cdots \langle 4-3 \rangle$$

다음에는 확률변수가 연속적인 것인 경우에는 확률변수의 구간 (Sample Space)을 ($X_1 \sim X_n$), 확률변수에 대한 확률을 P(X)로 표시하면 확률변수의 기대치 E(X)는 다음식에 의하여 계산된다.

$$E(X) = \mu = \int_{X_1}^{X_n} X \cdot P(X)\, dX \cdots\cdots\cdots\cdots \langle 4-4 \rangle$$

(3) 2항분포 (Binomial distribution)

4개의 동전을 동시에 던져 나올 수 있는 방법의 가지수는 다음과 같다. 여기서 편의상 표면 H, 이면을 T로 표시하면

<표 4 - 1>

표면이 나올 수(X)		도 수	상대도수
0	TTTT,	1	0.0625
1	HTTT, THTT, TTHT, TTTH,	4	0.2500
2	HHTT, THHT, TTHH, HTHT, THTH, HTTH	6	0.3750
3	HHHT, HHTH, HTHH, THHH,	4	0.2500
4	HHHH,	1	0.0625
계		16	1.0000

0.0
16) 1.000

동전 표면과 이면이 나온 확률 4개 $\left(\frac{1}{2}+\frac{1}{2}\right)^4$

위의 도수는 n을 4로 한 경우의 2항정리 $(P+q)^n$의 전개식

(각각 동전 4개)

$$(P+q)^4 = {}_4C_0P^4 + {}_4C_1P^3q + {}_4C_2P^2q^2 + {}_4C_3Pq^3 + {}_4C_4q^4$$

$$= 1P^4 + 4P^3q + 6P^2q^2 + 4Pq^3 + 1q^4$$

의 계수와 똑 같으며 상대도수는 표면이 나온 확률 0.5원 이면이 나올 확률 $1-0.5=0.5$를 전개식에 대입하여 계산한

$$(0.5+0.5)^4 = 0.0625 + 0.2500 + 0.3750 + 0.2500 + 0.0625$$

와 똑 같다. 상대도수는 확률의 의미를 가지는 것으로 이 계산에서 얻어진 각항의 값은 동전의 표면이 X회 나올 확률을 표시한 것이라 할 수 있다.

이제 이상에서 설명한 것을 요약하면 일반적으로 배반사상의 경우에 1회의 시행에 있어서 사상E가 나올 확률을 P, 나오지 않을 확률을 $q\ (1-P)$로 하면, 독립적인 시행을 n회 반복했을 때 E가 X회 나올 확률을 다음식에 의해서 표시한다.

$$f(X;n,P) = {}_nC_XP^nq^{n-X} \quad (X=0,1,2 \cdots n;\ P+q=1) \cdots\cdots <4-5>$$

이 (4-5)식을 2항분포함수 또는 2항분포라 하며 불연속분포함수의 가장 대표적인 것이다.

이제 확률변수 X_i가 2항분포를 할 때 X_i 기대치 즉 산술평균은

$$E(X) = \mu = \sum_{2}^{n} X_i\, Xf(X)$$

$$= nP \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots <4-6>$$

가 된다. 이것을 위의 예에서 계산해 보면

$$E(X) = \mu = 0\times0.0625 + 1\times0.2500 + 2\times0.3750 + 3\times0.2500 + 4\times0.0625$$

$$= 2$$

${}_nC_0p^0q^{n-0} + {}_nC_1p^1q^{n-1} + {}_nC_2p^2q^{n-2} \cdots + {}_nC_np^nq^0 (n-n)$

이렇게

實例: $(\frac{1}{2}+\frac{1}{2})^4 = \underbrace{{}_4C_0(\frac{1}{2})^0(\frac{1}{2})^{4-0}}_{\frac{1}{16}} + {}_4C_1(\frac{1}{2})^1(\frac{1}{2})^3 + {}_4C_2(\frac{1}{2})^2(\frac{1}{2})^2 + {}_4C_3(\frac{1}{2})^3(\frac{1}{2})^1 + {}_4C_4(\frac{1}{2})^4(\frac{1}{2})^0$

즉 이것은 4×(0.5)와 같다. 따라서 n=4와 P=0.5를 곱한 수 nP와 같다. $(\frac{1}{2})^0=1$, $(\frac{1}{2})^4=\frac{1}{16}$

이와 같은 방법으로서 X_i의 분산을 구하여 보면

$$Var(X) = E(X_i^2) - (E(X_i))^2$$
$$= \sum_{0}^{n} X_i^2 \times f(X) - (np)^2$$
$$= npq \quad \cdots\cdots\cdots\cdots \quad \langle 4-7 \rangle$$

가 된다. 이것을 역시 위의 예에서 계산해 보면

$$E(X_i^2) = 0^2\times 0.0625 + 1^2\times 0.2500 + 2^2\times 0.3750 + 3^2\times 0.2500 + 4^2\times 0.0625$$
$$= 5$$
$$[E(X)]^2 = (3p)^2$$
$$= 4$$
$$E(X_i^2) - [E(X)]^2$$
$$= 5 - 4$$
$$= 1$$
$$Var(X) = npq$$
$$= 4\times 0.5\times 0.5$$
$$= 1$$

0.5 + 0.5
0.1 + 0.9) 넣는 1이는

즉 이항분포 2항분포에 있어서의 X의 분산값은 npq이며 따라서 표준편차 값은 $\sqrt{npq}$ 가 된다.

(4) 포아슨분포 (Poisson distribution) 이항분포의 P가 아주 작은데

포아슨분포는 이항분포의 극한의 경우를 말하는 것으로 이항분포에서 n의 값이 무한이 크며 P의 값이 아주 작을 때 (즉 0에 가까울 때) 이항분포는 포아슨분포로 바뀐다. 이를

0.001 + 0.999
나올 확률이 극히 적은 때

일반식으로 나타 내면

$$f(X;m) = \lim_{\substack{n \to \alpha \\ p \to e}} {}_nC_X p^X q^{n-X}$$

$$= \frac{m^X e^{-m}}{X!} \quad (\therefore m = np) \cdots\cdots\cdots\cdots \langle 4-8 \rangle$$

된다.

포아손 분포에서 평균값과 분산값은 같은값 m이며 표준편차 값은 $\sqrt{m}$ 이다.

연속분포

(15) 정규분포 (Normal distribution)

2항분포와 포아손분포는 비연속적인 것이지만 여기서 기술하고자 하는 정규분포는 연속분포이다.

이것은 그 응용범위가 넓을 뿐아니라 특히 표본이론에서 아주 중요한 역할을 하고 있다.

(가) 정규분포의 정리

2항분포함수 $f(X;n,P) = {}_nC_X P^n q^{n-X}$

에서 P를 주어진 것으로 하고 n을 무한대로 하면 다음과 같은 식이 증명된다.

$$\lim_{n \to \alpha} {}_nC_X P^X q^{n-X} = \frac{1}{\sqrt{2\pi npq}} e - \frac{(X-np)^2}{2npq} \cdots\cdots \langle 4-9 \rangle$$

그런데 2항분포의 평균치와 분산은 각각 $\mu = np$, $\sigma^2 = npq$ 이므로 이를 위의 식에 대입하면

$$f(X;\mu\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e - \frac{(X-\mu)}{2\sigma^2} \cdots\cdots\cdots\cdots \langle 4-10 \rangle$$

가 된다. 이 식에서 정해지는 이론적 분포를 정규분포라 하며 또는 가우스분포(Gauss distribution)라고도 한다. 그리고 평

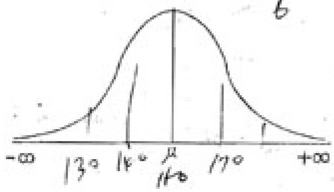
균이 μ이고 표준편차가 σ인 확률변수 X가 정규분포를 한다면 $X \sim N(\mu, \sigma)$로 표시한다.

(나) 정규분포의 성질

(1) 정규분포는 평균치를 중심으로 종모양 (Bell-Shaped) 을 한 <도4 - 1>와 같은 완전 대칭 분포이다.

<도4 - 1>

$\frac{x_i - \bar{x}}{\sigma}$

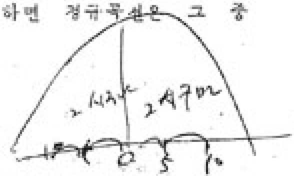


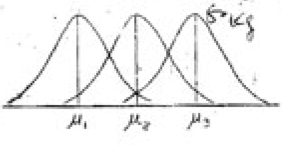
(2) 정규분포의 곡선을 간단히 정규곡선 (Normal Curve) 이라 하는데 정규곡선의 모양은 평균치 μ와 표준편차 σ의 값에 의해서 변한다.

먼저 평균치 μ의 값이 변하고 σ의 값이 일정하다고 하면 정규곡선은 그 중심만이 변하고 모양은 변하지 않는다.

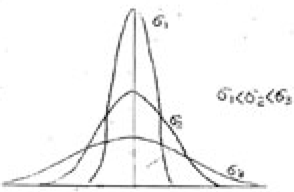
또한 μ는 일정하고 표준편차 σ가 변하면 곡선의 모양은 변한다. <도4 - 3>

<도4 - 2>





<도4 - 3>



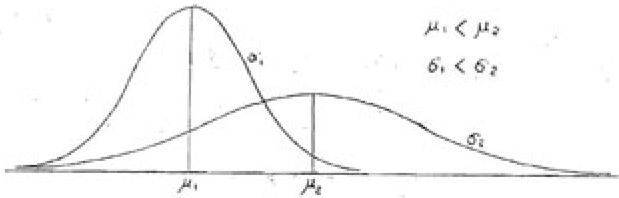
$\bar{x} = 50kg$

$\sigma = 4\ kg$

$z = \frac{x_i - \bar{x}}{\sigma}$

표준정규변량

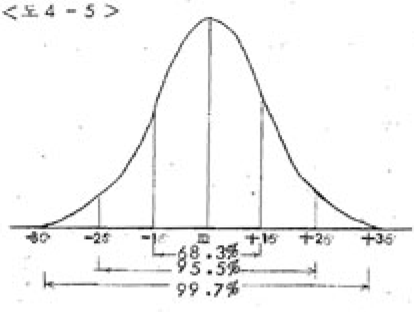
그러고 μ와 σ가 모두 변하면 중심도 이동되며 분포의 모양도 변한다. <도4 - 4>



즉 μ값이 작아지면 정규곡선의 중심은 좌측으로, 커지면 우측으로 이동하며 σ값이 작아지면 곡선의 모양은 좁고 높아지며, 커지면 정규곡선은 평평하게 된다.

(3) 정규곡선에서 산술평균 μ를 중심으로 1표준편차 구간 즉 1σ의 거리를 취하면 그 구간내에 집단의 총개체중 68.3%가 포함되며, 2σ의 거리를 취하면 그 구간내에 95.5%, 3σ의 거리를 취하면 99.7%, 즉 거의 모든 개체가 이 범위내에 들어가게 된다. 이것을 그림으로 그려보면 <도4 - 5>와 같다.

<도4 - 5>



(다) 표준정규분포

정규분포함수의 확률변수 X를

$$z = \frac{X - \mu}{\sigma} \quad \cdots\cdots\cdots\cdots\cdots\cdots \ \langle 4-11 \rangle$$

으로 변환하면 이것을 표준정규변량 (Standard Variable) 이라 하며 이것은 평균이 μ, 표준편차가 σ인 정규분포를 하는 변량 X를 평균 0, 표준편차가 1인 정규분포를 하는 변량으로 변환한 것을 말한다. 어떤 변량 Z가 표준정규분포를 한다면 $Z \sim N(0, 1)$로 표시하며 이것은 서로 다른 μ와 σ 값을 가진 정규분포를 하나의 통계수표를 사용할 수 있도록 하기 위하여 만들어진 것이다.

따라서 어떤 모수 (μ와 σ) 를 가지고 있는 정규분포라도 이를 표준정규변량으로 변환하여 부록의 표준정규분포 수표에서 그 특성을 쉽게 알아 낼 수 있기 때문이다.

이상에서 우리는 확률의 이론과 확률분포에 관해서 보았는데 이는 앞으로 표본이론에서 언급될 표본분포의 기초 이론이 되는 것으로 아주 중요하다.

5. 상관관계 (Corelation)

$-1 \leq r \geq +1$

가. 상관관계의 의의와 종류

상관관계는 변량간의 평균적 관계를 말한다. 즉 상관관계는 남편과 연령과 아내의 연령과의 관계 처럼 한 변량의 개개의 수치와 딴 변량의 다수의 수치에서 유도된 평균치와의 관계를 말한다. 이 점에서 한 변량의 개개의 수치와 딴 변량의 개개의 수치와의 관계를 의미하는 함수관계와 차이점을 갖는다. 그러나 다른 한편에서는 상호 의존적인 관계라는 점에서 서로 공통점을 갖

相關 1. 平均치의 관계
2. 개개 수치의 관계
함수 관계

고 있다고 하겠다.

상관관계는 인과관계와도 차이점을 갖는다. 상관관계는 상호 의존적인데 반해 인과관계는 원인에 의한 결과의 일방적 관계이다. 따라서 상관관계와 인과관계와는 그 의미가 전혀 다르다고 할 수 있다.

상관관계는 단순상관과 중상관으로 크게 나누어 볼 수 있다. 단순상관은 2개의 변량간의 관계를 말하는데 중상관은 3개이상의 변량간의 관계를 말한다. 다시 이들 관계가 직선으로 나타나는가 또는 곡선으로 나타나는가에 따라 직선상관과 곡선상관으로 나뉜다.

상관관계의 정도는 상관도표를 통하여 쉽게 알 수 있지만 일반적으로 수치에 의해서 파악된다. 따라서 다음에 상관관계의 종류와 상관관계의 정도를 표시하는 측도를 대응시켜 보면 아래표와 같다.

상관관계의 종류		상관 관계의 측도
단순상관	직선상관	(단순)상관계수
	비직선상관	(단순)상관비, (단순)상관지수
중상관	직선상관	중상관계수
	비직선상관	중상관비, 중상관지수

그러나 이중에서 가장 중요하고 기본적인 것은 단순직선상관이다. 따라서 여기서는 이것만을 취급하기로 한다.

r=0 무상관
1= 완全상관

$r = \frac{6xy}{6x6y}$ 共分散
x,y 표준편차

나. 회귀방정식과 회귀선

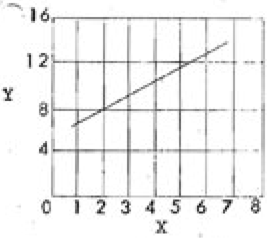
상관관계를 표시하는 식을 회귀방정식이라 하여 이 식을 그라프로 도시한 것을 회귀선이라 하고 이 직선의 기울기를 회귀계수라 한다.

다음의 <표5-1>을 사용해서 그림 <도5-1>에서 보면 도표상의 직선이 회귀선이 되고 이 직선의 식이 회귀방정식이 되는 셈이다. 그러고 회귀방정식은 최소자승법에 의하여 구한다.

<표5-1>

觀察回數	X	Y
1	1	7
2	4	4
3	5	13
4	3	11
5	7	13

<도5-1>



최소자승법은 실제치와 계산치와의 편차의 자승의 합계를 최소가 되도록 방정식의 상수를 정하는 방법을 말한다.

이 방법에 의해서 실제로 계산할 때에는 선형방정식에서 극소의 조건을 만족하도록 유도된 정규방정식을 연립방정식을 푸는 방법에 의하여 상수의 값을 정한다.

회귀방정식이 직선일 때에는 선형방정식은 $Y = a + bX$가 되며 정규방정식은

$$\left.\begin{aligned} \sum y_i &= \sum a + b\sum X_i \\ \sum X_i \cdot y_i &= a\sum X_i + b\sum X_i^2 \end{aligned}\right\} \cdots\cdots\cdots \langle 5-2 \rangle$$

이 된다. 여기서 $\sum y_i$, $\sum X_i$, $\sum X_i \cdot y_i$ 및 $\sum X_i^2$ 은 자료에서 구할 수 있고 다만 a와 b만이 미지수이다. 따라서 a와 b를 변수로 한 2원 1차방정식을 연립방정식을 푸는 방법에 의하여 계산하면 다음과 같다.

<표 5 - 2>

X_i	y_i	$X_i \cdot y_i$	X_i^2
1	7	7	1
4	4	16	16
5	15	75	25
3	11	33	9
7	13	91	49
$\sum X_i = 20$	50	222	100

위의 <표 5 - 2>에서 얻은 각 수치를 <5 - 2>식의 정규방정식에 대입하면 된다. (이 경우 n는 5이다)

$$50 = 5a + 20b \cdots\cdots\cdots (1)$$

$$222 = 20a + 100b \cdots\cdots\cdots (2)$$

$(2) - (1) \times 4$

$$\begin{aligned} 222 &= \cancel{20a} + 100b \\ -)\ 200 &= \cancel{20a} + 80b \\ \hline 22 &= \qquad 20b \end{aligned}$$

$$r = \frac{22}{20}$$

$$= 1.1$$

b 값을 (1)식에 대입하면

$50 = 5a + 20 \times 1.1$

$50 = 5a + 22$

$5a = 28$

$$a = \frac{28}{5}$$

$$= 5.6$$

a 와 b 의 값을 $Y_c = a + bX$에 대입하면 $Y_c = 5.6 + 1.1X$가 된다. 이것이 곧 구하고자 하는 회귀방정식이다.

다. 상관계수의 계산

여기서 말하는 상관계수는 단순상관계수를 말하며 일반적으로 상관계수라고 할 때에는 이 단순상관계수를 의미한다.

상관계수란 두 변량간의 상관관계의 강도를 측정하는 것으로 두 변량 X와 y의 각각의 산술평균에서의 편차의 적을 산술평균한 값 즉 승분산의 개념으로서 나타낼 수 있을 때 이 값은 단위를 가지고 있으므로 이를 다시 각각의 표준편차로 나눈 무명수로 만들어 이 값을 상관계수로 한 것이다.

이제 이것을 일반식으로 나타내면

$$r = \frac{\sigma_{X \cdot y}}{\sigma_X \cdot \sigma_y} = \frac{\frac{1}{N}\sum(X_i - \mu_X)(y_i - \mu_y)}{\sqrt{\frac{1}{N}\sum(X_i - \mu_X)^2} \cdot \sqrt{\frac{1}{N}\sum(y_i - \mu_y)^2}}$$

$$= \frac{N\sum X_i y_i - \sum X_i \cdot \sum y_i}{\sqrt{N\sum X_i - (\sum X_i)^2} \cdot \sqrt{N\sum y_i^2 - (\sum y_i)^2}} \quad \cdots \langle 5-3 \rangle$$

된다. <표5 - 1>에서 상관계수를 구하여 보면 다음과 같다.

<표5 - 3>

관찰회수	X_i	Y_i	X_i^2	Y_i^2	$X_i Y_i$
1	1	7	1	49	7
2	4	4	16	16	16
3	5	15	25	225	75
4	3	11	9	121	33
5	7	13	49	169	91
합 계	20	50	100	580	222

<표5 - 3>에서 얻은 값을 식<5 - 3>에 대입하면 상관계수는

$$r = \frac{5 \times 222 - 20 \times 50}{\sqrt{5\times100-(20)^2}\cdot\sqrt{5\times580-(50)^2}}$$

$$= \frac{+110}{\sqrt{100 \times 400}}$$

$$= \frac{+110}{200}$$

$= +0.55$ 이 된다.

라. 상관계수의 성질

상관계수를 해석하기 위해서는 상관계수의 성질에 관한 지식이 필요하다. 그 성질을 요약하면 다음과 같다.

(1) Y와 X의 상관관계가 직선적인 경우에 $r = 0$이면 Y와 X는 아무런 관계가 없다. 이 경우의 상관관계를 무상관이라고 한다.

버스가격
자전거가격 -0.3

(2) r의 절대치는 항상 1보다 작다.

즉 $-1 \leq r \leq 1$ 이다. 그러고 그 절대치가 클 수록 두 변량간의 상관관계는 크다.

(3) r이 플러스이면 두 변량간의 관계는 병행적이며 r이 마이너스이면 두 변량간의 관계는 역행적이다. 전자의 경우를 정의 상관관계, 후자의 경우를 부의 상관관계라 한다.

(4) $r = \pm 1$일 때에는 두 변량간의 상관관계는 완전하다고 할 수 있다. 이 경우의 상관관계를 완전상관이라고 한다.

6. 시계열 (Time Series)

가. 시계열의 구성요소

시계열이란 동종의 통계집단 또는 부분집단을 기술하는 통계숫자를 시간적으로 배열한 것이다.

시계열의 분석에 있어서 시계열에 나타나는 변동을 보면 대체로 단순하지 않고 여러가지 성질이 서로 관련되어 몇가지 변동의 중첩적 결과로 나타나고 있다.

보통 시계열에 나타나는 변동의 구성요소로서 다음의 네가지를 들 수 있다.

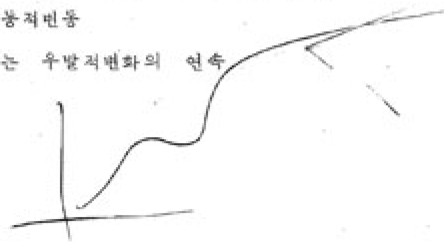
T Trend 추세변동 : 장기간에 걸친 비반복적인 변동

S Seasonal variation 계절변동 : 기후같은 물리적계절, 습관같은 사회적계절에 기인해서 매년 대략 규칙적으로 반복하는 파동적변동

C Cyclical Fluctuation 순환변동 : 1년이상의 기간으로 해서 확장과 축소의 순환적 형태를 표시하는 파동적변동

I Irregular 불규칙변동 : 아무런 규칙성이 없는 우발적변화의 연속

시계열 = T × C × S × I



이상의 여러가지 시계열 변동에는 각기 그 측정방법이 있으나 여기서는 추세변동의 측정방법에 대해서만 취급하기로 한다.

나. 추세변동의 측정

추세변동의 측정방법에는 목축법, 이동평균법 및 최소자승법이 있는데 이중 가장 대표적인 방식이 역시 최소자승법인 것이다.

앞에서도 이미 언급된 바와 마찬가지로 최소자승법에서는 우선 선형방정식을 찾아내야 하는데 일반적으로 다음의 선형식을 주로 사용하고 있다.

(1) 직선 : $y = a + bt$

(2) 2차곡선 : $y = a + bt + ct^2$

(3) 지수곡선 : $y = ab^t$ 또는 $\log y = \log a + t \log b$

이상의 여러가지 선형중에서 어느 하나가 정해지면 여기에서 정규방정식이 유도되고 이 정규방정식과 실제치를 사용하여 선형방정식의 상수를 정하면 된다. 정해진 상수를 선형방정식에 대입한 것을 추세방정식이라고 한다.

지금 선형방정식을 직선식으로 가정하고 $y = a + bt_i$로 하면 정규방정식은

$$\begin{cases} \sum y_i = \sum a + b\sum_t \\ \sum y_i \cdot t_i = a\sum t + b\sum_t{}^2 \end{cases} \quad \cdots\cdots\cdots\cdots \quad \langle 6-1 \rangle$$

이 된다.

이제 <표 6 - 1>에서 추세방정식을 구하여 보기로 하자. 그런데 원자료를 그대로 사용하면 $\sum^t$, $\sum^{t2}$ 및 $\sum t \cdot y_i$의 계산은 번잡하게 된다. 따라서 여기에 계산을 간편하게 하기 위하여 t의 원점을 이동하는 방법을 생각하게 된다.

전원에
전남도원에
전남도기비

추세를 알아내는 方法
1) 目測法
2) 移動 평균법 (양쪽 끝년도 없기때문 국민경제가 상대적 으로 ...)
3) 최소자승법
4) 반분평균법

<표 6 − 1>

연 차	Y_i
1950	3
1951	2
1952	8
1953	6
1954	11

이제 주어진 연수(자료의 수) N이 기수일 때와 우수일 때와를 나누어 생각해 보자.

(가) N가 기수일 경우

먼저 N가 기수(홀수)일 때, 즉 $N = 2n + 1$에는 중앙년도를 0으로 하고 위로 −1, −2, −3, ……………… 으로 하고 아래로 1, 2, 3……… 으로 기입하면 정규방정식에서 $\sum t$는 0이 된다. 따라서 <6 − 1>의 정규방정식은 <6 − 2>의 방정식으로 변환할 수 있다.

$$\begin{cases} \sum Y_i = \sum a \\ \sum Y_i t_i = b \sum t^2 \end{cases} \quad \cdots\cdots\cdots\cdots\cdots <6-2>$$

이제 <표 6 − 2>에서 추세방정식을 계산하여 보면 다음과 같다.

<표 6 − 2>

연 차	t	Y_i	tY_i	t^2
1950	−2	3	−6	4
1951	−1	2	−2	1
1952	0	8	0	0
1953	1	6	6	1
1954	2	11	22	4
합계 ($\sum_{i=1}^{n}$)	0	30	20	10

$30 = 5a$

$20 = 10b$

그러므로

$$\begin{cases} a = \dfrac{30}{5} = 6 \\ b = \dfrac{20}{10} = 2 \end{cases}$$

계절변동을 제거하는 方法
1) 月別 평균법
2) 12개월 이동평균법

$Y = T \times C \times S \times I$

$\dfrac{Y}{S \times I} = T \times C$

따라서 추세변동의 직선방정식은

$y = 6 + 2t$

이다. 이와같은 방법에 의하여 방정식을 구하였으면 기준으로 한 원점의 연도를 반드시 명시하여 두어야 한다.

즉 $y = 6 + 2t$ (원점 : 1952 년 , 단위 : 1 년)

(나) N가 우수일 경우

계열의 항수가 기수인 경우에는 중앙을 원점으로한 계열로 변환할 수 있었으나 우수일 경우에는 중앙에 오는 수가 2항이므로 중앙에 오는 2항중 위의 항에 − 1 , 아래항에 + 1 , 을 주어 각항간의 시간단위의 간격을 2로하여 배열하면 된다.

즉 , ······· −5 , − 3 , − 1 , + 1 , + 3 , + 5 ·····································

로 시간단위를 부여하면 $\sum t$ 는 0이 되므로 간편하게 계산할 수 있다.

7 . 지수 (Index Number)

가 . 지수의 의의

지수라는 것은 일반적으로 동일한 통계 계열에 있어서 그 통계수자의 크기를 비교하기 위하여 일정한 수치를 기준으로 하고 이에 대응하는 다른 각 수치를 이에 대한 비율로 표시한 것을 말한다.

박보미

예컨데 1960 년에 비하여 1970 년에는 우리나라의 인구가 얼마나 증가되었는 가를 알고자 할 때 또는 65년을 기준으로 하여 볼 때 70 년에는 어느정도 물가가 올랐는가 , 지난 제2차 경제계획기간 동안 경제성장은 얼마나 이루어졌는가 등 어느 두 시점간

의 변화를 측정하는 방법으로서 흔히 사용되고 있는 것이다.

이와같이 통계적 계열의 시간적(또는 장소적) 변화를 비교하기 위하여 백분비(Percentage)로 나타낸 것을 지수라 하며 이것은 가장 기본적인 통계해석 방법의 하나이다. 왜냐하면 아주 복잡하고 복합적인 사상을 가지고 있는 생계비, 산업생산 및 경기순환등의 변화를 나타내는 것도 이 지수에 의하기 때문이다.

이것은 높은 가격이 다른 물량등의 요소를 가지고 전체적인 변화의 정도를 하나의 수로서 표현되는 것이다.

이와같이 지수는 경제분야 뿐아니라 사회 각 분야에 널리 보급되어 중요한 지표로 이용되고 있는 것이다.

나. 지수의 종류와 작성방법

지수에는 단순지수(Unweighted Index Number)와 가중지수(Weighted Index Number)의 구별이 있으며 이를 계산방법에 따라 다시 개별지수와 종합지수로 나누어 볼 수 있고, 또 표지의 종류에 따라 가격지수, 물량지수, 그러고 금액지수로 나누어 볼 수 있다.

지수를 작성할 때에는 반드시 다음과 같은 문제점에 유의하여야 한다.

첫째, 자료의 유용성 및 비교가능성의 검토

둘째, 지수품목선정

세째, 기준시점의 선택

네째, 지수품목의 중요도(가중치)산출

다섯째, 적합한 지수산식의 적용

(1) 단순지수 (Unweighted Index Number)

이제 지수의 계산방법을 알아보기 위하여 1970년 1971년의 곡물가격을 비교하여 보면 다음과 같다.

<표 7 - 1>

	1970년	1971년
쌀	692	904
보리쌀	422	520
콩	652	748
팥	902	1,122
밀가루	773	959

지금 위의 표에서 1971년의 곡물가격 합계를 1970년의 곡물가격 합계로 나누면,

$$\frac{904+520+748+1,122+959}{692+422+652+902+773}=\frac{4253}{3441}=1.246$$

의 값을 얻는다. 이것은 1971년의 곡물가격이 1970년의 곡물가격에 비하여 24.6%가 등귀하였다는 뜻이다.

이때 여기서 사용한 산출방식을 단순총화법이라 하여 일반적으로 다음과 같은 수식으로 표시한다.

$$I=\frac{\sum P_n}{\sum P_o}\cdot 100 \quad\cdots\cdots\cdots\cdots \quad <7-1>$$

여기서 $\sum P_n$은 비교시 가격을 나타내는 것이고 $\sum P_o$는 기준시 가격의 총화를 뜻하는 것이다. 그리고 이 두값의 비율에 100을 곱한 것은 지수를 백분비로 나타내기 위함이다.

단순총화지수에 있어서의 가장 큰 결점은 지수에 사용된 여러품

① 개별지수. 종합지수

쌀. $\frac{73\text{가격}}{72\text{가격}}\times 100$

연결 등가

③ { 단순지수 / 가중지수.   { 가격지수 - 도매물가, 소비자물가 / 물량지수 → 생산지수, 출하지수 / 금액지수 → 수입액

목의 가계의 물량단위가 달라짐에 따라 다른 값을 나타내기 때문에 서로 다른 단위를 가지는 품목을 복합하여 지수를 작성할 때에는 곤란한 점이 많다. 따라서 실제에 있어서는 단순총화법에 의한 지수작성은 거의 하지않고 있다.

또 다른 방법으로서는 개별품목의 지수를 산출하여 얻은 개별지수를 평균하여 얻는 방법으로 평균법이 있다.

이제 평균법에 의한 곡물가격지수를 계산하여 보면,

먼저 각 품목별 개별지수를 다음과 같이 산출한다.

<u>개 별 가 격 지 수</u>

쌀	$\frac{904}{692} \times 100 = 130.6$
보 리	$\frac{520}{422} \times 100 = 123.2$
콩	$\frac{748}{652} \times 100 = 114.7$
팥	$\frac{1,122}{902} \times 100 = 124.4$
밀가루	$\frac{959}{773} \times 100 = 124.1$

이것을 다시 종합한 종합지수를 계산하기 위해서 산술평균, 중위수, 최빈수 또는 다른 평균법을 사용하여 계산할 수 있다.

이제 이것을 산술평균방법을 적용하여 계산하면

$$\frac{130.6 + 123.2 + 114.7 + 124.4 + 124.1}{5} = \frac{617.0}{5} = 123.4$$

의 값을 얻게 되는데 이를 일반식으로 나타내면

$$I = \frac{\sum \frac{P_n}{P_o} \times 100}{k} \cdots\cdots\cdots\cdots \langle 7-2 \rangle$$

된다. 여기서 k는 지수작성에 사용된 품목의 수를 의미한다.

(2) 가중지수 (Weighted Index Numbers)

지수를 구성하고 있는 개개의 품목들은 상품거래 또는 소비생활상 그 중요도가 다르다. 따라서 이와같은 중요도의 차이를 지수작성에 고려하여야 하는데 다음은 중요도를 고려한 지수작성 방법에 대하여 생각해 보기로 한다.

<표7 - 2>

	가 격		능 량 (생산량)
	1970 년	1971 년	1970 년
	(원)	(원)	(1,000톤)
쌀	692	904	3,939
보리	422	520	819
콩	652	748	232
팥	902	1,122	24
밀가루	773	959	917

위의 표에서 기준년도 (1970년) 의 곡농생산량을 가중치로한 기준년도의 곡농가격과 비교년도 (1971 년) 의 곡농가격의 가중산술평균값을 구하여 지수로 환산하면

$$\frac{904(3,939) + 520(819) + 748(232) + 1,122(24) + 959(917)}{692(3,939) + 422(819) + 652(232) + 902\ (\ 24) + 773(917)}$$

$$= \frac{5,066,603}{3,953,159} = 1,282 \text{ (또는 } 128.2\% \text{)}$$

의 값을 얻는다. 이것은 위에서 계산한 단순지수의 값 124.6에 비하여 3.6 포인트나 큰데 이는 쌀의 개별지수가 타품목에 비하여

훨씬 클 뿐아니라 가중치로 사용된 생산량 역시 타품목에 비하여 상당히 크므로 결국 쌀의 기여도가 크게 작용한데 원인이 있음을 알 수 있다.

여기서 사용한 계산식을 일반식으로 나타내면

$$I = \frac{\sum P_n q_0}{\sum P_0 q_0} \times 100 \cdots\cdots\cdots\cdots\cdots\cdots \langle 7-3 \rangle$$

이 된다.

그런데 기준년도의 가격에 기준년도수량을 가중치로 사용하고 비교년도의 가격에는 비교년도의 수량을 가중치로 사용하는 것이 더 좋을 것이라 생각할 수도 있지만 이것은 전혀 다른 의미의 금액지수라 하는 것이다.

V5

I. 상 용 대 수 표

N	0	1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010

N	0	1	2	3	4	5	6	7	8	9
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445

N	0	1	2	3	4	5	6	7	8	9
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996

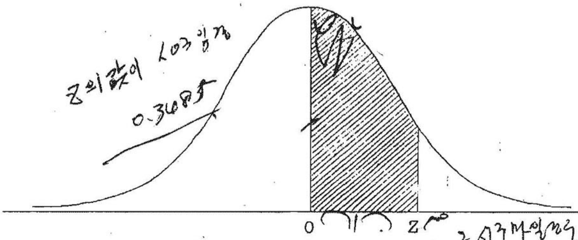
Ⅱ. 난 수 표

04433	80674	24520	18222	10610	05794	37515
60298	47829	72648	37414	75755	04717	29899
67884	59651	67533	68123	17730	95862	08034
89512	32155	51906	61662	64130	16688	37275
32653	01895	12506	88535	36553	23757	34209
95913	15405	13772	76638	48423	25018	99041
55864	21694	13122	44115	01601	50541	00147
35334	49810	91601	40617	72876	33967	73830
57729	32196	76487	11622	96297	24160	09903
86648	13697	63677	70119	94739	25875	38829
30574	47609	07967	32422	76791	39725	53711
81307	43694	83580	79974	45929	85113	72268
02410	54905	79007	54939	21410	86980	91772
18969	75274	52233	62319	08598	09066	95288
87863	82384	66860	62297	80198	19347	73234
68397	71708	15438	62311	72844	60203	46412
28529	54447	58729	10854	99058	18260	38765
44285	06372	15867	70418	57012	72122	36634
86299	83430	33571	23309	57040	29285	67870
84842	68668	90894	61658	15001	94055	36308
56970	83609	52098	04184	54967	72938	56834
83125	71257	60490	44369	66130	72936	69848
55503	52423	02464	26141	68779	66388	75242
47019	76273	33203	29608	54553	25971	69573
84828	32592	79526	29554	84580	37859	28504

68921	08141	79227	05748	51276	57143	31926
36458	96045	30424	98420	72925	40729	22337
95752	59445	36847	87729	81679	59126	59437
26768	47323	58454	56958	20575	76746	49878
42613	37056	43636	58085	06766	60227	96414
95457	30566	65482	25596	02678	54592	63607
95276	17894	63564	95958	39750	64379	46059
66954	52324	64776	92345	95110	59448	77249
17457	18481	14113	62462	02798	54977	48349
03704	36872	83214	59337	01695	60666	97410
21538	86497	33210	60337	27976	70661	08250
57178	67619	98310	70348	11317	71623	55510
31048	97558	94953	55866	96283	46620	52087
69799	55380	16498	80733	96422	58078	99643
90595	61867	59231	17772	67831	33317	00520
33570	04981	98939	78784	09977	29398	93896
15340	93460	57477	13898	48431	72936	78160
64079	42483	36512	56186	99098	48850	72527
63491	05546	67118	62063	74958	20946	28147
92003	63868	41034	28260	79708	00770	88643
52360	46658	66511	04172	73085	11795	52594
74622	12142	68355	65635	21828	39539	18988
04157	50079	61343	64315	70836	82857	35335
86003	60070	66241	32836	27573	11479	94114
41268	80187	20351	09636	84668	42486	71303

※ Based on parts of Table of 105,000 Random Decimal Digits, Interstate Commerce Commission, Bureau of Transport Economics and Statistics, Washington, D.C.

Ⅲ. 정 규 분 포 표



Z의값이 1.03일경우 0.3485

2시그마일경우 0.4772

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1654	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.480[illegible]	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4951
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Ⅳ. 2 항 계 수 표

n	$\binom{n}{0}$	$\binom{n}{1}$	$\binom{n}{2}$	$\binom{n}{3}$	$\binom{n}{4}$	$\binom{n}{5}$	$\binom{n}{6}$	$\binom{n}{7}$	$\binom{n}{8}$	$\binom{n}{9}$	$\binom{n}{10}$
0	1										
1	1	1									
2	1	2	1								
3	1	3	3	1							
4	1	4	6	4	1						
5	1	5	10	10	5	1					
6	1	6	15	20	15	6	1				
7	1	7	21	35	35	21	7	1			
8	1	8	28	56	70	56	28	8	1		
9	1	9	36	84	126	126	84	36	9	1	
10	1	10	45	120	210	252	210	120	45	10	1
11	1	11	55	165	330	462	462	330	165	55	11
12	1	12	66	220	495	792	924	792	495	220	66
13	1	13	78	286	715	1287	1716	1716	1287	715	286
14	1	14	91	364	1001	2002	3003	3432	3003	2002	1001
15	1	15	105	455	1365	3003	5005	6435	6435	5005	3003
16	1	16	120	560	1820	4368	8008	11440	12870	11440	8008
17	1	17	136	680	2380	6188	12376	19448	24310	24310	19448
18	1	18	153	816	3060	8568	18564	31824	43758	48620	43758
19	1	19	171	969	3876	11628	27132	50388	75582	92378	92378
20	1	20	190	1140	4845	15504	38760	77520	125970	167960	184756

For $K > 10$ it may be necessary to make use of the identity $\binom{n}{K} = \binom{n}{n-K}$.

2. 표 본 조 사

목 차

1. 표 본 조 사

우리나라의 인구는 얼마나 되며 그중에서 실업자는 얼마나 될 것인가? 또 금년의 미곡수확량은 얼마인가? 이를 알기 위해서는 인구나 농가의 미곡수확량을 전부 조사해 보지 않으면 정확히 알 수 없다.

이와같이 조사하고자 하는 대상을 하나도 빠짐없이 모두 조사하는 것을 전수조사라고 한다. 그러나 우리나라의 미곡수확량총계는 이와같은 전수조사에 의해서 계산된 것은 아니다. 실제로는 지역별로 일정한 방법에 의하여 선정된 일부분의 면적에 대해서만 그 지역의 평당수확량을 조사하고 여기에 그 지역의 전경지면적을 곱하여 그 지역의 총생산량으로 보고 다시 각 지역의 총량을 합계하여 전국의 총생산량이라고 보는 것이다. 이와같이 대상의 일부분만을 조사하고 이것을 기초로 하여 전체를 추정하는 조사방법이 바로 표본조사인 것이다.

그런데 표본조사는 우리가 조사하여 결론을 얻고자 하는 대상전체(이를 모집단이라 한다)를 조사하지 않고 그 일부분만을 선정하여(이 선정된 부분을 표본〈Sample〉이라 한다) 조사하게 되므로 전체를 조사한 결과와 완전히 일치할 수는 없고 어느정도의 차이가 날것은 당연한 일이다.

따라서 표본조사에서는 그 결과가 전수조사의 결과와 가능한 한 근사하게 접근시키는 것이 중요하다. 그러기 위해서는 무엇보다 먼저 선정된 표본이 모집단의 특성이나 구조를 그대로 잘 반영하고 있어야만 할 것이다.

그러면 표본을 얻는 여러가지 방법에 대해서 알아보기로 하자.

가. 유의표본추출법 (Purposive Sampling)

예컨데 농가의 소득을 알고자 할 때에 전국에서 가장 전형적이라고 생각되는 농촌을 몇군데 적당히 표본으로 선정하는 방법 (전형법 Typical Method)이 있고 또 여론조사를 하고자 할때에 시민의 성별, 년령, 학력, 직업, 수입등의 특성을 고려하여 몇가지의 부류로 분류하고 각 부류에 적당한 수의 표본을 할당하여 조사원으로 하여금 추출하게 하는 방법 (할당법 Quota Method) 이 있다.

그런데 전형적인 표본을 얻기 위하여는 조사담당자의 광범한 지식, 풍부한 경험, 적절한 판단이 필요하며 또한 좋은 표본을 얻었다 하더라도 여기에서 얻어지는 결과가 전수조사하였을 때의 결과와 얼마나 상이할 것인가를 판단할 수 없다. 그 이유는 전형적인 표본을 선정하는 과정이 주로 조사설계자의 주관에 의존하고 있으므로 그 표본이 모집단을 대표하는 정도를 객관적으로 아는 방법이 없는 까닭이다. 따라서 전형적인 표본은 설계자의 능력에 따라 좋은 결과를 얻을 수도 있고 그렇지 못할 수도 있다

할당법은 전형법에 비하면 여러가지 특성에 대하여 모집단과 같은 구조를 갖게끔 표본을 뽑는 것이므로 보다 합리적이라 할 수 있다. 그러나 이 방법에서도 얻고자 하는 정보와 관련이 있는 몇가지 요인에 의해서 모집단을 관리할 수는 있어도 모든 요인을 철저히 관리할 수는 없는 결점이 있다.

위와같은 방법들은 조사설계자가 자기의 주관에 의해서 표본을 선정하게 되므로 이를 유의추출법이라고 한다.

이외에도 백화점등에서 투표함을 설치하여 두고 특정상품이나 써어비스에 대하여 손님의 의견을 구하는 방법도 있으나 이 방법은 매우 만족한 사람, 또는 매우 불만인 사람들의 투표율이 높아

양극단의 의견이 반영될 경향이 있으므로 모든 손님의 의견을 공평하게 대표하였다고 생각할 수 없다.

이러한 의미에 있어서 이 방법을 불완전 일부조사라고도 한다.

이상과 같은 종류의 방법은 모집단에서 좋은 표본을 얻기 위한 완전한 방법이라고는 말할 수 없으나 사용방법에 따라서는 유용한 정보를 얻을 수 있다는 것을 잊어서는 안된다.

나. 확률표본추출법 (Probability Sampling)

지금 우리나라의 도시근로자가구의 가계수지를 알기 위하여 표본조사를 하려고 한다 하자. 이때에 모집단(전도시 근로자가구)에서 제비뽑는 식으로 표본을 뽑는 방법을 생각할 수 있다. 즉 어느가구나 표본으로 뽑힐 기회를 가질 수 있도록 하는 추출방법이다.

이러한 방법은 우리가 일상생활에서도 흔히 사용하고 있다. 그런데 이 방법에 대해서는 특히 주목하여야 할 사항이 두가지 있다.

첫째로 이 방법에서는 표본은 전연 우연적으로 선정되고 따라서 모집단의 모든 개체가 모두 표본으로 뽑힐 가능성, 즉 확률을 가진다는 점이며,

둘째로는 이와같이 하여 표본이 선정되는 것이기 때문에 만약 다시 표본을 추출하여 보면 딴 표본이 추출되고, 이에 의하여 밝혀진 결과는 처음의 표본에 의하여 밝혀진 결과와는 틀리게 된다

즉 예를들어 근로자가구의 평균수입을 조사하였다면 모집단(전도시 근로자가구)의 평균치는 확정치(하나밖에 없는)인데 표본의 평균치는 확률변수인 점이다.

그러므로 표본조사에 의하여 모수(모집단의 특성치)를 추정할

때에는 어느정도의 폭을 가지고 추정하게 된다.

따라서 이와같은 방법은 표본을 확률적으로 추출하는 방법이므로 이를 확률추출법 (Probability Sampling) 또는 임의추출법 (Random Sampling) 이라고 한다.

오늘날 표본조사라고 하면 곧 이 확률표본에 의한 조사를 말하는 것이고 또 이것만이 진정한 의미에 있어서의 표본조사라고 할 수 있다. 그 이유는 이 방법에 의하면 조사결과의 정도를 객관적으로 판단할 수 있으나 유의표본 조사에서는 그렇지 못하기 때문이다.

2. 표본조사의 장점

표본조사의 장점을 요약하면 다음의 세가지를 들 수 있다.

가. 비용의 절약

나. 집계의 신속 (시간의 절약)

다. 정도 (精度) 의 향상

처음의 두 이점에 대해서는 이해하기 어렵지 않을 것이다. 전수조사와는 달라서 표본조사는 조사대상의 수가 적기 때문에 비용이 적게 들것은 당연하다.

또 조사대상의 수가 몇분의 일로 축소되기 때문에 그만큼 집계 소요시간도 줄게 되어 결과를 빨리 이용할 수 있다.

통계조사의 결과가 신속하게 이용될 수 있다는 것은 대단히 중요한 일로서 아무리 내용이 풍부하고 정확한 통계라 할지라도 3년이나 5년이 지난 후에야 결과가 발표된다면 별로 쓸모가 없게

된다. 물가나 임금통계따위는 그때 그때의 최신자료를 필요로 하는 것이기 때문에 불가불 표본조사에 의할 수 밖에 없다.

세번째의 이점인 정도의 향상에 대해서는 좀 더 설명이 필요할 것이다. 일반적으로 표본조사는 일부분만을 조사하여 전체를 추정하는 방법이기 때문에 전수를 조사하는 전수조사에는 미치지 못하는 것으로 생각되기 쉽다. 그러나 이는 반드시 그렇지 않다.

즉 조잡하게 실시된 전수조사보다는 정밀하게 기획되고 운용된 표 표본조사의 결과가 보다 신뢰성이 높기 때문이다. 왜냐하면 통계조사에 있어서는 여러가지 사정으로 정확한 자료를 얻기가 곤란하고 또 조사관계자 특히 피조사자와 직접 대하는 조사원이 언제나 정확한 조사활동을 하리라고는 보장할 수가 없기 때문이다. 인간인 이상에는 여러가지 실수를 범하게 된다. 그런데 인구센서스와 같이 수만명의 조사담당자가 동원되는 대규모조사가 될 것 같으면 이들에 대한 조사방법의 전달지도가 획일적으로 정확하게 행해지기 어렵다. 상당히 상세한 조사지시서를 준비하고 세심한 주의를 한다 하더라도 전달의 단계가 중앙에서 시.도로 시.도에서 다시 읍.면으로 단계가 겹칠수록 말단에서의 지시는 잘못 전달되는 수가 많다. 따라서 조사원의 조사활동도 획일적으로 행해지리라 고는 말할 수 없고 그러므로써 실사에서 여러가지 불통일과 과오를 저지르게 된다. 이와같이 통계조사에서는 조사상의 오차가 발생하는 것인데 그 크기는 조사의 규모가 커짐에 따라 증대하는 것으로 인구센서스와 같은 대규모의 조사에 이르면 일반인이 생각하는 것보다 훨씬 크게 나타나는 것이다.

그런데 표본조사에서는 이 조사오차의 애로를 크게 해소하게 된다.

첫째로 조사규모가 작기 때문에 그만큼 담당조사자간의 지도연락이 간편하고 통일을 기할 수 있다.

둘째로 소수의 조사원으로 조사할 수 있으므로 유능한 조사원을 채용하여 충분한 훈련을 실시함으로써 조사상의 실수를 면하게 한다

세째로 조사를 정밀하게 행할 수 있다.

조사원은 한가구 또는 한사람의 피조사자에게 충분한 시간을 주어 많은 질문을 하여 정확한 판단을 할 수 있게 된다. 예컨데 가계조사에서는 가구에서 매일 구입하는 모든 품목에 대한 수량과 금액에 대하여 조사하고 있는데 이때 한가구의 1개월간의 구입회수는 수십회에 이르는 것이기 때문에 전국의 모든 가구에 대하여 실시한다는 것은 도저히 불가능한 일이고 비록 그렇게 하였다 하더라도 정확을 기하지는 못할 것이다. 그러나 이를 미리 음미하고 소수의 추출된 가구에 대하여 실시하는 것은 가능하며 또 이들 소수의 가구에 대하여 충분한 지도를 함으로써 정확한 조사를 하게 된다.

이상에서 조사오차의 견지로 보아 표본조사가 전수조사보다 좋은 조사결과를 얻을 수 있음을 알았는데 다음에 표본오차의 견지에서 볼때에도 조사결과의 정도와 관련하여 표본조사의 우수성은 한층 명료하게 된다.

즉 표본조사는 부분적으로 표본만을 조사하기 때문에 일어나는 부정확성을 가지는데 이를 표본오차(Sampling Error)라 하고 앞에서 본 조사오차와 구별하고 있다. 그런데 확률표본에 의한 표본조사는 확률론의 지식을 이용하여 그 결과가 가지는 표본오차의 크기를 추정할 수 있고 또 이를 미리 고려에 넣어 필요한 정도(精度)의 결과를 얻을 수 있도록 표본을 설계할 수 있다.

이와같이 조사결과의 정도를 객관적으로 평가할 수 있고 또 반대

로 필요한 정도(精度)의 결과를 얻으려면 어떠한 표본을 사용하면 좋은가 하는 점에 대하여 따질 수 있는것이 표본조사의 가장 뚜렷한 특징이고 장점이다.

그렇다면 표본조사를 하기 위해서는 반드시 표본의 추출이 필요한데 어떠한 방법들이 있는가 다음 절에서 그 추출법에 대하여 알아 보기로 하자.

3. 단순임의 추출법

가장 기본적인 표본추출방법으로서 어떤 모집단으로부터 그 일부분을 뽑아서 표본으로 할때 그 집단에 속하는 모든 개체 즉 추출단위가 표본으로 선출될 기회를 동등하게 가질 수 있도록 하는 추출방법을 단순임의추출법(Simple Random Sampling)이라 하는데 우선 이의 구체적인 추출방법을 기술하기에 앞서 기초적인 이론을 먼저 살펴보기로 하자.

내

2

가. 가능한 표본의 조합수

확률표본의 기구를 이해하기 위하여 우선 가장 간단한 경우를 들어 모집단과 표본의 관계를 밝혀 보자. 일반적으로 조사항목은 다수이고 그 종류도 속성, 변량의 두가지가 있다.

그러나 지금 간단하게 하기 위하여 조사항목이 하나일 때를 생각해 보자.

모집단은 N개(이것을 모집단의 크기라고 한다)의 단위로 구성되고 이 중에서 n개(이것을 표본의 크기라고 한다)의 단위를 표본으로 추출한다고 하자. 이때 변량 값을 X라고 한다면 모집단은 N개의 X값으로 되어 있다고 생각할 수 있다.

즉. 모집단 : X_1 , X_2 , X_3 ………………………… , X_N

표 본 : x_1 , x_2 , x_3 ………………………… , x_n

로 구성되며 이때 x_i는 X_1 , X_2 ……………… , X_N중 어느 한 값이 될 것이며 모집단과 표본의 평균치와 분산은 각각 다음과 같이 정의된다.

모 평 균 : $\mu = \frac{1}{N}\sum X_i$

모 분 산 : $\sigma^2 = \frac{1}{N}\sum(X_i - \mu)^2$ 또는 $S^2 = \frac{1}{N-1}\sum(X_i - \mu)^2$

표본평균 : $\bar{x} = \frac{1}{n}\sum x_i$

표본분산 : $s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$

이제 크기 N인 모집단에서 크기 n인 표본을 단순임의 추출하면 몇조의 표본이 있을 수 있을 것인가? 숫자의 조합론에 의하면 NCn 조 있게 된다.

여기서 한가지 주의하여야 할 것은 이 NCn 조라고 하는 것은 N개 중에서 어느 하나를 추출하고 이 추출된 것은 제하고 나머지 N－1개 중에서 다시 하나를 추출하는 것과 같이 할때의 가능한 방법의 수이다. 이와는 다른 방법으로는 N개의 단위 중에서 하나를 추출하고 다음에 이를 제거하지 않고 다시 원상으로 되돌려 놓고 재차 N개의 단위중에서 하나를 추출하는 것과 같은 방법을 n번 반복하여 크기 n의 표본을 얻는 방법이 있다. 이와같은 방법을 무별한단순임의추출법 또는 반복을 허용하는 단순임의추출법이라고 하는데 이때에는 가능한 표본의 조합수는 N^n 조이다. 그러므로 전자의 방법을 비복원추출법이라 하고 후자의 방법을 복원추출법이라 한다.

나. 표본평균과 표본분산의 기대치

크기 N인 모집단에서 크기 n인 표본을 단순임의추출한다면 NCn 조 (복원추출의 경우 N^n 조) 의 가능한 표본을 얻을 수 있는데 이들 각조의 표본평균과 분산을 구하여 다시 평균하면 각각 모평균과 모분산에 일치한다. 이때 이 평균치를 기대치라 하며,

$$E(\bar{X}) = \frac{1}{NCn}\sum_{i=1}^{NCn}\bar{X}_i = \mu \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (1)$$

$$E(S^2) = \frac{1}{NCn}\sum_{i=1}^{NCn}S_i^2 = S^2 = \sigma^2 \quad \cdots\cdots\cdots\cdots \quad (2)$$

로 표시할 수 있다. 또한 이들 표본평균에서부터 표본평균의 기대치 즉, 모평균까지의 편차의 제곱의 기대치 (이것은 표본평균의 기대치 즉, 모평균까지의 편차의 제곱의 기대치 (이것은 표본평균의 분산을 뜻한다) 를 구하면 다음과 같이 된다.

$$\sigma_{\bar{X}}^2 = E\{\bar{X} - E(\bar{X})\} = \frac{1}{NCn}\sum_{i=1}^{NCn}(\bar{X} - \mu^2)$$

$$= \frac{N-n}{N-1} \cdot \frac{\sigma^2}{N}$$

$$\left(\text{또는 } \frac{N-n}{N} \cdot \frac{S^2}{n}\right) \quad \cdots\cdots\cdots \quad (3)$$

이제 위의 사실들을 예를 들어 증명하여 보기로 하자.

〔예〕 크기가 4이고 변량값이 각각 1, 2, 3, 4, 5인 모집단에서 크기 2인 표본을 추출한다고 하면 모두 $5C_2 = \frac{5 \times 4}{2} =$ 10개조의 표본이 가능하다.

이들을 열거하여 직접 비교하여 보면,

$$\mu = \frac{\sum Xi}{N} = \frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$$

$$S^2 = \frac{1}{N-1}\sum(X_i - \mu)^2$$

$$= \frac{1}{4}\{(1-3)^2+(2-3)^2+(3-3)^2+(4-3)^2+(5-3)^2\}$$

$$= \frac{1}{4}(4+1+0+1+4) = 2.5$$

$$\sigma_{\bar{X}}^2 = \frac{N-n}{N} \cdot \frac{S^2}{N} = \frac{3}{5} \cdot \frac{2.5}{2} = 0.75$$

標本番号	X_i	$\sum X_i$	標本平均 $\bar{X}_i$	標本分散 S_i^2	$(\bar{X}_i - \mu)$	$(\bar{X}_i - \mu)^2$
1	1.2	3	1.5	0.5	− 1.5	2.25
2	1.3	4	2	2	− 1.0	1.00
3	1.4	5	2.5	4.5	− 0.5	0.25
4	1.5	6	3	8	0	0
5	2.3	5	2.5	0.5	− 0.5	0.25
6	2.4	6	3	2	0	0
7	2.5	7	3.5	4.5	0.5	0.25
8	3.4	7	3.5	0.5	0.5	0.25
9	3.5	8	4	2	1.0	1.00
10	4.5	9	4.5	0.5	1.5	2.25
計			30	25		7.5
(平均 (期待値)			3	2.5		0.75

위의 표에서 보는 바와같이 표본평균 및 분산의 기대치가 모평균과 모분산에 일치함을 알았다.

이상에서는 우리는 몇가지 사실을 증명하여 보았다. 이들은 후에 아주 유용하게 사용될 것이다.

표본을 추출하는 것은 표본에서 얻은 값을 가지고 모집단의 값을 추정하고자 하는 것이었다. 즉, 표본평균 $\bar{X}$는 모평균 μ의 추정치이다. 이 추정치 $\bar{X}$의 기대치는 즉 가능한 모든 표본 NCn 조에 대하여 만약 자기의 표본평균을 계산하고 이를 평균한다면 (1)에서 보는 바와같이 모평균 μ와 같아진다. 이와같이 표본추정치의 기대치가 모수와 같아지는 경우에는 이 추정치를 불편추정치 (Unbiased estimate) 라고 하고 이 추정치는 편의 (Bias) 가 없다고 한다.

(1) 및 (2)에 의하여 단순임의추출에서는 표본평균 $\bar{X}$ 및 표본분산 S^2 가 각각 모평균 μ 및 모분산 σ^2 의 불편추정치임을 알았다.

그런데 추정치는 항상 불편(不偏)하다고는 할 수 없다.

특히 표본표준편차 S의 기대치는 모표준편차 σ 와 일치하지 않는다. 그러나 근사하므로 모표준편차의 추정치로 사용함에 모순이 없다고 할 수 있다. 따라서 추정치는 일반적으로 불편추정인 것을 바라나 불편추정치이어야만 사용할 수 있는것이 아니고 때에 따라서는 편의가 있는 추정치를 사용하는 것이 더 좋을때도 있다.

다. 표본평균의 분포와 표본오차

앞에서도 본바와 같이 크기 N의 모집단에서 n의 표본의 모든 가능한 방법의 수는 NCn 개이었다. 지금 이들 모든 가능한 표본에 대해서 모두 조사하여 그 평균치의 분포상태를 보기로 한다. 말할것도 없이 이 분포는 모집단의 분포와는 상이할 것이다. 그러나 확률론에 의하면 표본평균의 분포에 대하여 다음의 두정리가 증명되고 있다.

내

3

즉, 단순임의표본의 경우에는

(1) 모 집단의 분포가 정규분포이면 표본평균도 정규분포가 된다.

(2) 모 집단의 분포가 어떠하든 n을 크게 하면 표본평균의 분포는 극한에 있어서는 정규분포가 된다.

위의 정리에 의하면 모집단의 분포여하에 불구하고 표본수 n을 충분히 크게 하면 표본평균의 분포는 근사적으로 정규분포로 보아도 좋다.

그런데 실제문제로서 n은 얼마던지 크게 할 수는 없는 것이므

로 n가 어느 정도이면 좋은가가 문제가 되는데 다행이도 변량조사의 경우에는 n가 30이상이면 표본평균의 분포는 정규분포로 볼 수 있다. 이 성질은 대단히 편리한 것으로 표본조사에서는 모집단의 분포형을 일단 고려할 필요가 없게 한다. 정규분포에 대해서는 이미 통계개론에서 언급된 바 있으나 여기서 다시 복습하여 보기로 하자.

이제 어느 통계집단이 평균을 μ이고 표준편차는 σ인 정규분포를 한다고 하면

(1) 이 집단의 분포곡선은 평균치 μ를 중심으로 좌우 대칭이고,

(2) 이때 표준편차의 값은 곡선의 모양을 정하고 σ의 값이 클수록 곡선은 편평하게 되고 σ의 값이 적을수록 좁고 높게 된다.

(3) 평균치 μ를 중심으로 σ만큼의 거리를 취하면 그 구간내에 집단의 총개체중 68.3%가 포함되고, 2σ의 거리를 취하면. 95.5%, 3σ의 거리를 취하면 99.7% 즉, 집단의 모든 개체가 거의 포함되게 된다. 이것을 도시하면 다음과 같다.

이제 표본평균의 분포에 위에서 살핀 정규분포의 성질을 적용하면 즉, 모든 가능한 표본에 얻은 표본평균들은 모평균을 중심으로 2배의 표본평균의 표준편차구간을 취할때 그 구간내에 이들 표본평균중 95% 이상이 포함된다. 따라서 단순임의추출에 의하여 얻은 어느 하나의 표본평균값 $\overline{X}$와 모평균값 μ와의 차이가 2배의 표본평균의 표준편차값보다 적을 확률은 95% 이상이라고 귀납할 수 있겠다.

위에서 표본평균의 표준편차를 표준오차(Standard Error) 또는

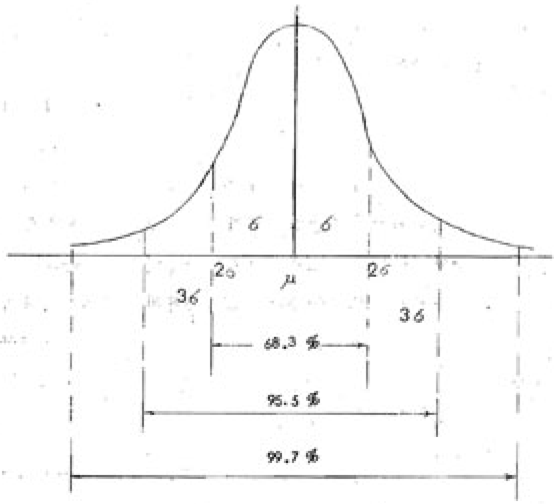


그림 4

표본오차(Sampling Error)라고 부르며 식(3)을 제곱근을 취하여 얻게 된다.

$$\sigma_{\bar{X}} = \sqrt{\frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}} \quad (\text{또는 } \sqrt{\frac{N-n}{N} \cdot \frac{S^2}{n}}) \cdots\cdots\cdots \quad (4)$$

그런데 모분수 σ^2는 실제에 있어 모르고 있으므로 표본추정치 S^2로 대치하여 쓰게되며 식(4)는 다시 다음과 같이 된다.

$$S_{\bar{X}} = \sqrt{\frac{N-n}{N} \cdot \frac{S^2}{n}} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (5)$$

이때 $S_{\bar{X}}$는 표본오차(또는 표준오차)의 추정치가 되는 것이다 따라서 표본에서 계산할 수 있는 것은 $\sigma_{\bar{X}}$가 아니고 $S_{\bar{X}}$인

것이다. 그렇다면 이들에 대해서도 앞의 이론이 성립하는가 하는 것에 의문을 갖는것이 당연한데 실수에 있어서는 일단 성립하는 것으로 보아도 무방하므로 증명없이 받아 들이기로 한다.

이상에서 증명한 내용을 다시 요약하여 본다면 즉 평균이 μ. 표준편차가 σ인 모집단에서 적당한 크기의 표본을 추출하면 그 표본평균 $\bar{x}$는 평균이 μ이고 표준편차가 $\sigma_{\bar{x}}$인 정규분포를 한다고 볼 수 있다. 따라서 모평균추정치로 사용된 어느 하나의 표본평균에서 2배의 표본오차구간을 취하면 그 구간내에 모평균의 진가가 위치할 확률이 95% 이상이 된다고 할 수 있다.

이제 이것을 일반식으로 나타내면.

$$Pr\{\bar{x} - K\sigma_{\bar{x}} < \mu < \bar{x} + K\sigma_{\bar{x}}\} = 0.954 \cdots\cdots\cdots\cdots(6)$$

(K=2일때 즉. 2배의 표본오차구간을 취하면)

되고. 여기서 K를 신뢰계수라 하며 K=1일때에는 68.3%, K=3일때에는 99.7%의 신뢰구간을 각각 얻게 된다.

라. 필요한 표본수의 결정

표본수가 크면 클수록 그만큼 조사비용은 많이 들지만 표본오차는 적어지게 된다. 그런데 실제에 있어서 비용은 제한을 받는 것이기 때문에 이 제한된 예산내에서 여하히 조사결과의 정도를 높일수 있는가가 문제된다.

그러나 비용이란 여러가지 조건으로 조사종류에 따라 달라지기 때문에 측정이 곤란하므로 여기서는 비용을 생각하지 않고 조사결과의 정도(精度)에 의해서만 표본수를 결정하는 방법에 대하여 알아 보기로 하자.

예컨대 어떤 지역의 전가구 평균 소비지출액을 알고자 할때 이

에 사용될 표본의 크기를 결정키 위해서는 표본조사결과에 의한 추정치가 얼마만큼의 정확성을 가지면 충분한지 또는 얼마만큼의 정확성을 필요로 하는가를 미리 정하여야 한다. 이제 신뢰수준 α에서 추정된 평균 소비지출액의 표본오차의 허용한계를 E이내로 할려고 한다면

$$K\sigma_{\bar{X}} = K\sqrt{\frac{N-n}{n}\cdot\frac{\sigma^2}{n}} \leqq E \quad \cdots\cdots\cdots\cdots(7)$$

이란 부등식을 만들고 양변에 제곱을 취하여 n에 대하여 풀면,

$$N \geqq \frac{K^2 N \sigma^2}{K^2 \sigma^2 + NE^2} \quad \cdots\cdots\cdots\cdots(8)$$

얻게 된다. 그런데 n을 얻기 위해서는 반드시 모분수 σ^2를 알고 있어야 하는데 보통 정확히 알지 못한다.

그러므로 과거조사의 결과라던지 또는 타조사의 결과에 의한 추정치를 사용하게 된다.

〔예〕 인천시의 가구당 평균 소비지출액을 조사하기 위하여 표본을 얻을려고 한다. 그런데 인천시의 총가구수는 10만이고, 가구당 평균 소비지출액의 표준편차는 1,000원일 것이라 한다.

이때 95%의 신뢰한계에서 허용오차의 한계를 200원 이내로 하는 표본을 얻을려고 한다면 표본을 어느정도 크기로 하면 좋은가?

(답) $N = 100{,}000$

$\sigma = 1{,}000$

$E = 200$

$K = 2$ (신뢰수준이 95%이기 때문)

이므로 식(8)에 의해서

$$n \geqq \frac{4 \times 100{,}000 \times 1{,}000{,}000}{4 \times 1{,}000{,}000 + 100{,}000 \times 40{,}000} \fallingdotseq 99.01$$

따라서 표본수는 적어도 100이 되어야 한다.

마. 구체적인 추출방법

위에서도 설명한바와 같이 단순임의 추출법이란 모집단의 모든 개체가 표본으로 선출될 확률을 동일하게 갖게 하는 방법으로서 말하자면 제비 뽑는 식의 추출법이다.

그런데 모집단의 개체의 수가 많을때에는 제비를 만드는 일만도 상당한 작업량이 되고 또 그것들이 균일하게 만들어지기 어렵다.

그래서 이와같은 추첨의 도구로서 통계에서는 「난수표」라는 것을 사용한다. 난수표라는 것은 0에서 9까지의 숫자를 임의로 배열해 놓은 것으로서 말하자면 0~9의 숫자를 추첨으로 몇번이고 반복하여 뽑아 그 결과를 기록한 표라고 생각하면 좋다.

부록에 있는 난수표는 다섯자리의 수로서 배열해 놓았는데 이는 편의상 그렇게 한것뿐이지 별다른 뜻이 있는 것은 아니다.

지금 모집단이 30개의 요소로 구성되어 있다 하자. 이때에 먼저 이들 30개의 개체에 어떠한 순으로도 좋으니 01, 02, 03, ………… 28, 29, 30과 같이 일련번호를 붙인다. 그리고서 난수표에서 임의의 자리를 골라 거기서부터 시작하여 30이하의 수가 나오면 그 수에 해당하는 번호의 개체를 표본으로 하면 된다.
가령 부록의 난수표를 좌상의 처음 두줄부터 시작하여 아래로 내려 읽기로 한다면 최초로 나오는 수는 02이므로 2번의 개체를 표본으로 한다. 만약 또 하나의 표본을 추출코자 한다면 그 밑은 85인데 이는 30보다 큰 수이므로 그냥 건너뛰고 그 다음을 보면 26이니 26번의 개체를 추출하면 된다.

이와같은 방법을 계속하여 나가면 얼마던지 필요한 수의 표본을 뽑을 수 있을 것이다.

다시 모집단의 개체의 수가 300개일때를 생각해 보자.

이때에도 앞에서와 마찬가지로 우선 난수표의 어느 열, 어느 행부터 시작할 것인가를 정한다. 이의 결정도 전적으로 임의로 결정하는 것이므로 예를 들자면 눈을 감고 연필을 떨어뜨려 그 끝이 맞는 점에서 가장 가까운 두 수를 읽어 처음 수는 열을, 다음 수는 행의 번호를 정하고 그 점부터 시작하여 세자리씩 읽어 나가는 식으로 하면 된다. 지금 연필 끝이 제시한 점의 수가 95라 한다면 9열5행을 기점으로 하여 세줄의 수를 아래로 읽어 나가면서 추출하게 된다. 이때 열으로 세숫자씩 읽어나가도 좋다. 부록의 난수표에서 보면 9열5행을 기점으로 하면 (394), 221, (321), 005, (742), (945), (452), (615), (948), (806) (750), (838), (697), 197, (766), (257), (488), 019 ……… 와 같이 된다. 여기에서 ( )내의 수는 모두 300을 넘으므로 버리고 221, 005, 197, 019, ………… 를 취하게 된다. 그런데 위에서 본 바와 같이 세자리의 수를 읽어가자면 300을 넘는수가 많이 나온다. 이때에 이들을 전부 버리고 있으면 약 50개의 표본을 추출하기에도 상당한 시간이 걸린다. 그래서 이를 절약하기 위하여 다음과 같은 방법을 쓰면 좋다. 즉 지금 난수표에서 나온 숫자를 N이라 하면 이를 300으로 나누어 나머지 수를 이용하면 된다. 즉,

$$\frac{N}{300} = a + R$$

와 같이 되는데 이때 이 R은 300보다 작은 수일 것이다.

이렇게 한 결과를 보면 94, 221, 21, 005, 142, 45, 15, 48,

206, 150 ………………… 과 같이 되어 이들 수에 해당하는 번호의 개체를 표본으로 뽑게 된다.

난수표의 사용에서 주의해야 할 일은 언제나 같은 표를 사용하지 않도록 하는 일이다. 시작하는 거점이나 읽는 순서 또는 방향을 임의로 자주 바꾸지 않으면 안된다.

4. 계통추출법

난수표를 사용하면 임의표본을 간단히 선정할 수 있다고 했다 그러나 실제에 있어서는 표본수가 많아지면 여간 시간이 걸리지 않는다. 일례를 들어 어느 학교 학생수 2,000명 중에서 50명의 표본학생을 뽑는 일을 생각해 보자. 지금 부록의 난수표의 우상에서부터 시작하여 아래로 읽어가면서 결정하기로 한다면 모집단이 네자리의 수이므로 0294, 8569, 2673, 4782, 7660, 4752, 7010, 8681, ………과 같이 읽어가게 되는데 이렇게 하여 이 세줄의 제일 끝까지 내려가서 당선되는 표본수는 0294, 1793, 0290, 1419, 1997, 1523의 6명에 불과하게 된다. 이와같이 하여 간다면 50명의 표본을 뽑기 위하여 이 부록 전부를 읽어 보아야 할 것이다. 더우기 선정된 번호는 일정한 순서로 나오는 것이 아니므로 중복하여 뽑히지 않았나를 확인한다던지 번호순으로 정리하는데 상당한 시간과 노력이 요하게 된다. 겨우 2,000명중에서 50명을 뽑는데도 이러한데 만약 모집단의 개체의 수나 표본수가 다수가 되면 그 표본추출작업은 얼마나 힘드는 일인가를 짐작하기에 충분하다.

그러면 실제에 있어서는 어떻게 하는 것일까? 위의 예를 가지고 보면 우선 표본으로 뽑힐 확률을 계산한다. 표본수는 50이므

로 이 확률은 1/40 (= 50/2,000) 이기 때문에 어쨌든 40명에 1명꼴로 표본을 뽑으면 기계적으로 50명을 선정할 수가 있다.

그러기 위해서는 우선 1 ~ 40에서 임의로 한 수를 뽑고 단순임의추출법으로 다음부터는 이 수에서 40번째의 번호를 표본으로 선정하면 된다.

이와같이 모집단의 각개체에 일련번호를 붙여놓고 처음의 하나만을 임의추출하고 나머지의 표본은 이 표본으로 부터 일정간격으로 선정하는 방법을 계통추출법 또는 등간격추출법이라고 한다.

이때에 처음의 임의로 선정된 수를 임의출발점 (Random Start) 이라 하고 건너뛰는 간격을 추출간격이라고 한다.

그러면 이제 이 계통추출법은 과연 임의추출법이라고 말할 수 있는 것인가 검토해 보지 않으면 안된다. 지금 2,000명의 학생을 다음과 같이 배열해 보자.

↑	0001, 0002, 003, ……,	i, ……………………		0040,
	0041, 0042, ……………,	$i + 40 \times 1$. ………		0080,
(50行)	………………………………			
	1921, 1922, ……………,	$i + 40 \times 49$, ……		1960,
↓	1961, 1962, …………….	$i + 40 \times 50$, ……		2000,
	…………………… (40列) ……………………			

1번에서 40번까지의 숫자 중에서 임의로 하나를 정하여 이하 일정한 간격마다 (40번마다) 표본으로 선정한다는 것은 위와 같은 50행 40열의 수열의 표에서 어느 일열을 임의로 선정하는 것이 된다.

즉, 단순임의추출법에서는 모집단의 각개체에서 하나하나 임의로 선정하게 되는 것인데 대하여 계통추출법에서는 모집단의 개체를

번호순으로 몇개의 소집단에 등분하여 두고 그중에서 임의로 한 소집단을 추출하여 표본으로 선정하는 것이 된다. 따라서 만약 소집단으로 분할하였을 때에 그 소집단마다 일정한 경향을 갖게 된다면 표본은 한편으로 편의되는 결과가 된다. 그러나 모집단의 크기가 크고 복잡한 내용을 갖는 개체로 되어 있으면 이와같은 경향은 우선 면하게 되는것이 보통이고 경험상으로도 단순임의 추출법과 같이 생각하여도 좋은 것으로 되어 있다. 그런데 계통추출법은 단순임의추출법보다 정도(精度)가 못하나 추출에 사용되는 List의 작성방법에 따라 보다 정도(精度)를 높일 수 있다.

그렇다면 계통추출이 List의 순서에 따라 여하히 그 유효성이 달라지느냐 하는 점과 그 문제점에 대하여 알아보기로 하자.

첫째. List형태의 검토

(1) List가 크기의 순으로 나열되어 있을때는 단순임의추출보다 훨씬 정도(精度)가 좋으나 층화추출보다는 못하다.

(2) List에 주기성이 있을 때에는 어느 한편에 편의될 위험이 있으므로 이를 검토해 보아야 한다.

(3) 자연적인 지역순으로 나열되어 있을 때에는 경험상 단순임의추출법보다 약간 정도(精度)가 좋다.

(4) 생산공장의 제품등 시간적으로 계속성 있는 제품검사등에 활용하면 편리하다.

둘째. 문제점

(1) 계통추출에서는 가능한 표본수가 K개(추출간격)에 한정되기 때문에 표본평균의 분포가 정규분포에 접근한다는 전제를 적용할 수 없고 따라서 신뢰구간을 계산한다 하더라도 그 의미가 불명료하다.

(2) 표본평균의 분산을 계산할 실제적 방법이 없다.

(3) 모집단의 크기 N가 표본의 크기 n로 꼭 나누어 떨어지지 않는 때가 더 많다. 이 단수(端數)의 영향을 어떻게 처리할 것인가가 문제다.

5. 층화추출법

精度 (Precision)
正確度 (Accuracy)

모집단에 대한 어떤 지식을 가지고 있을때 이를 조사의 설계에 이용하면 그만큼 이득이 될 것이라는 것은 누구나 짐작할 수 있다 이 예비지식의 가장 간단한 이용방법으로서 층화추출법(Stratified Sampling)이 있다.

층화추출에 있어서는 먼저 어떠한 특성에 의하여 모집단을 분할한다. 예컨데 사업체조사에서 각 사업체를 규모나 업종에 의하여 분할하는 것과 같다.

이와같이 모집단에 속하는 모든 개체를 어느 지표에 의하여 몇개의 집단으로 분할하되 그 집단의 구성요소인 각 개체는 동질적인 것으로 하는 것을 층화(Stratification)라고 하고 그 집단을 층(Stratum)이라고 한다.

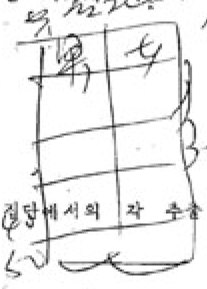
따라서 층화추출이란 모집단을 여러개의 층으로 분할하여 각층에서 독립적으로 일정수의 표본을 임의추출하는 방법을 말한다.

그렇다면 층화는 무엇때문에 하는가 하는 의문을 가지게 되는데 층화의 목적을 기술하면 다음과 같다.

가. 추정치의 정도(精度)를 높이기 위한 것.

나. 조사목적에 따른 것.

다. 조사실시상의 편의에 의한



상기한 목적중 가의 것이 주목적인데 이는 모집단에서의 각 추출

단위가 변이성이 클때 이것이 표본의 정도(精度)에 미치는 영향을 가능한 한 감소시키기 위한 수단이다. 따라서 층화는 추출의 수단이 아니고 추출을 하기 위한 집단화의 수단으로서 그 방법은 추출과 같이 일정한 기계적인 수속에 의하지 않아도 이론의 적용에는 하등의 지장이 없으므로 조사계획자의 유의판단에 의하여도 관계는 없겠으나 그 판단여하에 따라 효과의 차이는 클 것이다.

조사항목이 단순한 조사에서는 조사항목의 상관도가 높은 과거의 자료 또는 기존지식을 규준으로 하여 서로 유사한 것끼리 모으면 되나 조사항목이 복잡하고 다기(多岐)할 때에는 여하한 점에 초점을 맞추어야 하는가는 어려운 문제이므로 층화기준으로 다음과 같은 조건을 고려하여 결정하면 효과적이다.

가. 조사항목의 가장 중심이 되는 항목에 관련이 깊은 특성을 기준으로 한다.

나. 양적인 특성에 대해서는 모집단의 분포가 매우 편의(偏倚)된 것 또는 표준편차가 큰 것을 기준으로 한다.

다. 정성적(定性的)인 특성에 대해서는 모집단에 있어서 비율이 적은 것을 기준으로 한다.

라. 시간적인 안정성이 없는 특성은 기준으로 하지 않는다.

6. 집 락 추 출

모집단을 여러개의 부분집단으로 분할하고 그중 몇개의 집단을 임의추출하여 그 추출된 집단을 전수조사(Complete Survey)하던지 또는 추출된 집단에서 그 일부를 다시 추출하여 표본조사하는 방법을 집락추출(Cluster Sampling)이라 한다. 그 대표적인

장연

소지역 정확한 자료를 구할 때
표본조사로 예산 증가 있음

것으로는 지역추출 (Area Sampling) 이란 것이 있는데 이는 우리가 조사하고자 하는 각 개체가 실재하여 있는 전지역을 모집단으로 하고 이를 여러개의 소지역으로 분할하여 그중 몇개의 소지역을 추출하고 추출된 소지역내에 산재되어 있는 각 개체를 전수조사하던지 또는 추출된 소지역내에 산재되어 있는 전개체에 대하여 추출단위명부를 작성하고 단순임의추출법 또는 계통추출법에 의하여 그 일부를 다시 추출하여 추출된 개체에 대해서만 조사하는 방법을 말한다. 이때 각 조사단위는 소지역내로 한정되게 되므로

가. 조사를 위한 여비가 절약되고

나. 조사의 지도, 감독에 편리하고

다. 추출된 소지역에 대해서만 추출단위명부를 작성하면 되니까 모집단에 대한 충분한 정보가 없다고 하더라도 추출작업이 가능하다.

내 6

이상과 같이 집락추출의 경우 조사의 경제성과 신속성의 조건은 가장 잘 충족하나 정확성의 조건에서 보면 다른 추출법에 비하여 표본의 정도 (精度) 가 낮다. 그러므로 이의 보완책으로 층화의 방법을 사용하게 되는데 이때 이방법을 층화집락추출법 (Stratified Cluster Sampling) 이라 하며 실제 표본설계에 있어서는 위에서 설명한 여러 추출법을 따로따로 독립하여 사용하는 것이 아니라 동시에 적용하게 되는 것이다.

Slonim, M.J
Sampling in a Nutshell
출판사 Simon and Schuster, Inc
New york 1960

층화집락추출법

상대 ...

계통추출법

附錄 亂 數 表

(이表는 H.Burke Horton의 Random Decimal Digits의 一部를 転載한것이다)

90 개

추출간격

Sampling interval

계통추출법

02946	96520	81881	56247	17623	47441	27821
85697	62000	87957	07258	45054	58410	92081
26734	68426	52067	23123	73700	58730	06111
47829	32363	95941	72169	58374	03905	06865
76603	99339	40571	41186	04981	17531	97372
47526	26522	11045	835 5	45639	02485	43905
70100	85732	19741	92951	98832	38188	24080
86819	50200	50889	06493	66638	03619	90906
41614	30074	23403	03656	77580	87772	86877
17930	26194	53836	63692	67125	98175	00912
24649	31845	25736	75231	83808	98997	71829
79899	34061	54308	59358	56462	58116	97302
76801	49594	81002	30397	52728	15108	72070
62567	08480	61873	63162	44873	35302	04511
49723	15275	09399	11211	67352	41526	23497
42658	70183	89417	57656	35370	14915	16569
65080	35569	79392	14937	06081	74957	87787
02906	38119	72407	71427	58478	99297	43519
75153	86376	63852	60557	21211	77299	74967
14192	49528	78844	13664	98964	64425	33636

32059	11548	86264	74406	81496	23996	56872
81716	80301	96704	57214	71361	41989	92589
43315	50483	02950	09611	36341	20326	37489
27510	10769	09921	46721	34183	22856	18724
81782	04769	36716	82519	98272	13969	12429
19975	48346	91029	78902	75689	70722	88553
98356	76855	18769	52843	64204	95212	31320
29708	17814	31556	68610	16574	42305	56300
88014	27583	78167	25057	93552	74363	30951
94491	19238	17396	10592	48907	79840	34607
56957	05072	53948	07850	42569	82391	20435
50915	31924	80621	17495	81618	15125	48087
49631	93771	80200	84622	31413	33756	15218
99683	58162	45516	39761	77600	15175	67415
86017	20264	94618	85979	42009	78616	45210
77339	64605	82583	85011	02955	84348	46436
61714	57933	37342	26000	93611	93346	71212
15232	48027	15832	62924	11509	95853	02747

3. 統計調査方法論

目　　次

Ⅰ. 統計調査一般

가. 統計調査의 定義와 種類

統計라고 하면 우리는 곧 우리나라의 總人口, 通貨量, 物価指數, 鑛工業事業体數 等과 같이 社会現象을 數字로서 나타낸 것을 연상하게 된다. 統計란 이와같이 人口, 出生, 死亡, 物価, 通貨量, 生産, 輸出, 貯蓄, 失業等과 같은 社会現象에 관한 것인만큼 그것은 어떤 集団에 関한 事実을 説明하는 것이며, 따라서 集団이 아닌 個体(例 金某人, A企業体等)에 関한 個別的인 事実을 説明하는것은 統計가 아니다. 이와같이 集団에 関한 事実을 説明하는 統計를 作成하기 위하여 우리는 統計調査를 実施한다.

統計調査라함은 合目的的인 根拠에 依하여 統計集団 또는 部分集団(試料)을 形成함을 말하는 것이나 簡単히 말하면 統計資料를 蒐集整理하여 統計를 作成하는 節次이다.

이러한 統計調査의 種類를 살펴보면 다음과 같다.

(1) 全數調査와 標本調査

統計調査는 統計集団에 関한 事実을 알려고하는 것이므로 統計集団을 構成하고 있는 單位 全部를 調査하는 것이 所望스러운 일이나 時間, 費用, 人員등 많은 制約이 있고, 또 正確한 調査라는 觀点에서도 問題되는 点이 있어 오히려 調査単位의 一部만을 調査하는 경우가 더 많다. 이와같이 調査単位를 全部 調査하는 것을 全數調査라 하고, 調査単位의 一部만을 調査하는 것을 標本調査라고 한다

全數調査와 標本調査는 統計調査의 代表的인 것이며 兩調査가 密接한 関係가 있으므로 다음 節에서 詳述하기로 한다.

제1의 통계조사 → 통계의 目的을 실시
제2의 통계조사 → 이미의 目的 조작된 고

(2) 第一義統計調査와 第二義統計調査

第一의統計調査는 統計를 作成하는 그 自体를 目的으로 하여 実施하는 調査이며 第二義統計調査는 統計以外의 目的으로 作成한 記錄, 書類를 利用하여 統計를 만드는 調査이다. 普通 統計調査의 大部分은 前者의 경우이다. 第二義統計調査의 例를 들면 人口動態調査 建築許可統計, 貿易統計와 같은 것이 있다. 第二義統計調査는 報告 또는 記錄의 制度만 있으면 調査의 困難은 容易性, 經費, 被調査者의 負担等을 考慮할 때 第一義統計調査를 하는것 보다 쉽게 統計를 作成할 수 있는 것은 明白한 일이다.

反面에 第二義統計調査의 欠点도 많다. 戶籍上으로 申告되는 出生死亡, 離婚等의 統計에 있어서 相当한 期間이 経過後 発生事項이 申告되며 이를 다시 具体化하여 統計를 作成할때 까지는 많은 時間을 消費하여 利用에 不便하다. 뿐만아니라 申告를 하지 않는 일도 있어 精度를 低下시킨다.

(3) 静態統計調査와 動態統計調査

統計는 静態統計가 아니면 動態統計에 属한다. 一定한 時点에 있어서의 調査対象의 状態를 調査함으로써 얻는 統計를 静態統計라 하며 一定한 期間内에 継続 発生하는 事象을 調査함으로써 얻는 統計를 動態統計라고 한다.

5年 또는 10年에 1回씩 実施하는 人口센서스와 季別調査인 経済活動人口調査 또는 毎月調査인 人口動態調査, 그리고 鉱工業센서스와 鉱工業動態調査, 都小売業센서스와 都小売業動態調査등은 서로 補完하는 調査이며 静態와 動態를 把握하는 面에서 若干 差異가 있다. 統計를 利用할 때에는 両者를 関聯시켜 利用하면 그 効用을 높일 수 있다.

나. 統計調査의 必要性

必要한 統計를 既存資料에 의하여 作成할 수만 있다면 資料蒐集을 위한 時間, 費用, 努力이 節約될 수 있으므로 그 위에 좋은 일이 없을 것이다. 그러나 社会現象은 不断히 変化発展하고 있으므로 항상 必要한 既存資料를 갖고 있기는 어려운 일이다. 그 뿐아니라 既存資料란 그것이 統計를 만들기 위하여서 蒐集된 것이 아니고 다른 目的에 쓰기 위하여서 蒐集된 것이므로 이를 統計化하는데는 여러가지 制約이 따르는 것이다.

따라서 基本이 되는 資料의 蒐集이 먼저 遂行되지 않으면 안된다.

그런데 統計의 対象이 되는 것은 統計集団이지만 統計調査의 対象이 되는 것은 統計集団 그 自体가 아니고 統計集団을 構成하고 있는 個体(統計単位)이다. 即 統計는 集団 그 自体에 関한 数量的 知識을 얻는 것이지만 이 統計의 基本資料를 얻는 統計調査는 現実的으로 個個 単位를 対象으로 하여 実施되는 것이다. 이때 統計調査의 現実的인 対象이 되는 個個 単位를 調査単位라고 한다.

그러나 統計調査가 비록 個個単位를 対象으로 한다고 하더라도 궁극적으로는 集団에 関한 知識(統計)을 얻는데 目的이 있는 것이므로 당초부터 個体 그 自体에 関한 知識을 얻고자 하는 調査(個別調査)와는 厳密하게 区別된다. 統計調査가 個別調査와 区別되는 点으로 첫째, 統計調査는 集団에 関한 知識을 얻기 위한 実践的인 過程으로서 그 結果는 반듯이 統計로 俱現되는 것이며, 둘째, 集団의 構成単位(調査単位)를 大量으로 観察하는 것이며, 세째, 統計化가 可能하도록 調査単位에 대하여 統一된 標識을 부여하여야 하는

것이다.

統計調査는 繁多한 社会現象을 統計作成者의 意図(目的)에 맞추어 統計集団으로 만들어야 하고 統計調査의 目的을 이룰 수 있도록 各調査単位에 대한 正確한 実査가 이루어져야 하며 이에 뒤따라 当初의 目的에 맞는 統計表를 作成 発表하여야 한다.

統計調査의 全過程을 図示하면 다음과 같다.

$\boxed{\text{企 劃}} \rightarrow \boxed{\text{準 備}} \rightarrow \boxed{\text{実 査}} \rightarrow \boxed{\text{審査, 集計}} \rightarrow \boxed{\text{統計表作成}}$

外

1

統計調査의 方法을 統計作成者는 勿論 統計를 利用하기만 하는 사람도 統計를 올바로 보고 利用하기 위하여서는 充分히 理解하여야만 한다. 왜냐하면 統計는 그 目的과 作成者의 立場 및 統計調査方法의 如何에 따라서 얼마든지 다른 結果가 나올 수 있기 때문이다.

統計調査方法은 다음에 調査過程에 따라 詳述하기로 한다.

2. 全数調査와 標本調査

統計調査는 앞에서 말한바와 같이 統計集団에 関한 事実을 알려고 하는 것이므로 統計集団을 編成하고 있는 単位 全部를 調査하는 것이 가장 所望스러운 것이다. 그러나 実際로 調査単位를 全部調査한다는 것은 時間, 費用, 人員등 많은 制約이 있고 또 正確한 調査라는 点에서도 問題되는 点이 있어 오히려 調査単位의 一部만을 調査하는 경우가 더 많을 뿐아니라 一部만을 調査하고서도 統計集団에 関하여서 正確한 知識을 얻을 수만 있다면 오히려 一部만을 調査하는 것이 所望스러울 것이다.

위에서와 같이 調査単位를 全部調査하는 것을 全数調査라고 하며 調査単位의 一部만을 調査하는 것을 標本調査라고 한다.

가. 全數調査

全數調査는 調査対象이라고 생각되는 모든 部分을 全部調査하는 것을 말한다. 센서스같은 것이 그 좋은 例이다. 그러나 또한 調査対象의 範囲가 한個의 事務室의 境遇라도 또한 한 洞里라든가 한 学校内의 한 学級의 境遇라도 그 調査対象의 最終単位를 全部調査하는 限 이것은 全數調査인 것이다.

外

그러한 意味에서 어떤 種類의 全數調査는 事象의 量的記述을 目的으로 하는 統計調査에 対하여 어떤 特定된 事象을 깊이 파고 내려가서 綿密한 観察을 行하여 이것을 記述的으로 表現하는 事例調査와 같다고 할 수도 있을 것이다.

2

그러나 統計調査의 境遇 全數調査는 一般的으로 그 規模가 크다 여기서 한가지 注意할 것은 어떤 調査対象의 範囲内에 있는 모든 要素를 調査하는 것만이 全數調査가 아니다. 例를 들면 우리가 어떤 地域内의 農場을 調査하려고 할 때, 그 地域内의 農場만을 全部 調査하면 全數調査가 되는 것이며 農場 以外의 家屋이라든가 学校等 그 地域의 모든 것을 調査하여야 全數調査가 되는 것은 아니다. 그리하여 農場의 数, 農場과 農場 以外의 土地와의 比例 年度別 農場数 또는 農土面積의 増減, 農場의 種類別 比較, 収穫高 農事從業員人口 이들의 年令別, 性別, 比率等을 計量的으로 調査 表示하는 것이다.

나. 標本調査

標本調査를 部分調査라고도 한다. 即 이것은 前述한 諸標本抽出法에 따라서 調査対象 全体의 一部分을 選出하여 그 全体를 推定하는 調査를 말한다. 이때 調査対象 全体를 母集団이라고 하며

하며, 選出된 部分을 標本이라고 한다. 따라서 여기서 問題가 되는 것은 全体를 代表하는 部分을 如何히 抽出하는가 하는 点이다. 많은 調査가 이와같은 標本調査法에 依하여 이루어지고 있다.

그러나 原則的으로 時間이라든가 費用 또는 人員이 許諾한다면 全数調査가 좋다는 것은 말할 必要가 없다. 이것은 勿論 精密性이라는 点에 基準을둔 말이나, 調査目的에 비추어 어느 程度의 誤差는 問題視 안되는 境遇라든가 또는 調査目的이 많은 사람들 中에 大多数의 어떤 傾向만을 알고자 할 때에는 구태여 많은 人力과 時間을 浪費할 必要가 없을 것이다. 勿論 어느 程度 精密한 知識이 必要한가에 따라서 決定할 問題이며, 이 点에 関하여는 앞의 標本調査論에서 說明하였다.

다. 全数調査와 標本調査의 比較

全数調査와 標本調査는 어느것이 더 낫다든가 못하다고 한 말로 말할 수가 없다. 両者는 各各 長短点이 있으므로 調査의 性格에 비추어서 両者中 하나를 択하여야 한다. 全数調査와 標本調査의 長点을 가려내어 両者의 選択基準을 추려보면 다음과 같다.

(1) 全数調査에는 標本誤差가 없다. 따라서 센서스와 같이 誤差가 전혀 없거나 或은 그것이 最小限에서 그쳐야 할 경우에는 不得已 全数調査를 할 수 밖에 없다.

(2) 事後의 分類가 세분된 各層마다 이루어져야 하고 그것이 또한 正確한 것이 要求될 때에는 標本調査를 하더라도 많은 標本을 必要로 하므로 이 경우 차라리 全数調査를 하는 것이 좋다.

(3) 母集団(統計集団)이 比較的 작은 경우에는 標本調査를 하더라도 推定의 精度를 높이기 위하여서는 全数調査와 거의 対等한 程

度의 많은 標本을 推出하여야 한다. 따라서 이때에도 全数調査를 하는 것이 낫다.

(4) 또 어떤 目的을 위하여서는 精度가 높은 標本調査도 다른 目的을 위하여서는 精度가 떨어지는 일이 있으므로 多面的으로 利用하기 위하여서는 標本調査는 많은 制約이 있으므로 이 경우에도 全数調査를 하여야 한다.

(5) 標本調査를 하기 위하여서는 標本抽出, 推定, 誤差計算等 보다 專門的인 統計知識이 必要하므로 이러한 知識을 가진 사람을 求하기 힘들 때에는 全数調査를 하는 것이 安全하다.

(6) 標本抽出을 위하여서는 母集団에 関한 基礎資料가 있어야 하는데 이러한 것이 없는 경우에는 不得已 全数調査를 할 수 밖에 없다.

(7) 全数調査는 調査의 規模가 크기 때문에 巨額의 費用을 必要로 하는데 이러한 費用의 制約이 있는 경우에는 標本調査를 하면 費用을 節約할 수 있다.

(8) 全数調査는 実査와 集計에 많은 時間이 所要되는데 対하여 標本調査는 이를 短縮할 수 있으므로 調査結果를 迅速히 公表하여야 할 경우에는 標本調査를 하는 것이 좋다.

(9) 全数調査에서는 많은 調査人員을 必要로 하며, 따라서 未熟한 調査員을 使用하지 않을 수 없음에 対하여 標本調査에서는 훨씬 小数의 調査員으로서도 調査가 可能하기 때문에 熟練된 調査員을 使用하지 않으면 안될 경우에도 訓練하기가 容易하다.

(10) 全数調査는 標本誤差가 없는 대신 調査의 大規模性으로 因하여 非標本誤差가 標本調査에서 보다 크다. 따라서 非標本誤差를 줄이는 것이 重要한 경우에는 標本調査를 하는 것이 좋다.

(11) 그리고 現實的으로 全數調査를 實施할 수 없거나 調査의 性質上 全數調査가 不適当한 경우도 많은데 이러한 경우에는 不得已 標本調査를 하여야 한다.

以上 列擧한 것 以外에 標本調査는 標本抽出의 基礎資料(List등)를 全數調査結果에 依存하여야 하므로 標本調査를 위하여서도 數年에 一回씩 全數調査를 할 必要가 있고 또 標本調査에서는 推定値보다 平均이라든가 比率에 더 重点을 두는데 이는 全數調査에서 얻어지는 構造等과 結付되므로써 더욱 利用價値가 높아지는 것이므로 全數調査와 標本調査는 서로 補完關係에 있다고도 할 것이다.

3. 統計調査의 方法

全數調査와 標本調査의 槪念과 兩調査의 性格을 앞에서 略述하였다. 그러나 이러한 調査를 實施하고자할 때에는 調査主体와 客体, 調査事項, 調査時期 그의 地域 및 方法등에 여러가지 制約을 받게 된다.

다음에 調査上 一般的으로 要請되는 몇가지 注意事項을 든다.

(1) 合目的性: 調査対象이 그의 目的과 合理的으로 適合되어야 한다. 例를 들면 貨幣의 一般的 購買力을 調査하려면 都売物価를 調査하여야 하고, 勞働者의 生活程度를 觀察하려면 名目賃金보다 實質賃金의 調査가 더욱 重要하다.

같은 対象의 調査에 關하여서도 目的에 따라서 調査의 時期, 場所 方法等에 差異가 要求될 것은 勿論이다.

(2) 合理性: 統計対象에 따라서 調査時期와 場所 및 그의 方法이 科学的인 同時에 經済的이어야 한다. 이를테면 不必要하게 細密한

數字的 調査는 時間과 勞力 및 費用의 經濟上 반드시 合理的인 方途가 아니다.

(3) 資料의 同質性 : 統計資料의 時間的 및 場所的인 比較利用度를 可及的으로 向上시키기 위하여 調査項目, 調査時期, 其他 調査方法의 同質性으로 維持함이 必要하다. (例 國際的 各種統計基準의 定立과 必要性)

(4) 少聞多知의 原則 : 標識의 選定에 있어서 簡單한 調査로서 充分한 實質的인 目的을 達할 수 있도록 最善의 方法을 講求함이 必要하다. 이를 少聞多知의 原則이라 한다. 이를테면 道德統計에 있어서 國內의 善行을 일일이 調査하는 것보다 犯罪件數를 調査함이 便利하고, 國民所得의 調査에 있어서 收入部門을 일일이 調査 合計하는 것보다 消費額, 貯蓄額등의 支出面에서 이를 把握함이 簡便할 수가 있다.

(5) 外部的制約 : 觀察對象이 合目的的이라 할지라도 이를 實際調査함이 事實上 不可能한 境遇, 法律 또는 社會慣習上 不容認된 境遇 또는 經濟上 不合理한 境遇가 있으니, 이러한 外部的 制約은 調査에 앞서 미리 留念하여야 한다.

以上에서 列擧한 調査上 留意点을 念頭에 두고 統計調査를 企劃하여야 하며 調査準備 實際調査, 調査票의 審査와 集計, 統計表의 作成과 分析等 統計調査의 企劃부터 統計表의 作成까지 段階的으로 다음에 略述한다.

가. 統計調査의 企劃과 準備

(1) 調査目的

統計調査의 企劃中에서 가장 먼저 하여야 될일은 調査目的

ex. Why

U.N. recommendation

의 設定이다. 調査의 目的이 있으므로서 調査가 實施되는 것이며 모든 調査過程은 調査目的에 依하여서 規制되어야 하는 것이다. 그러므로 調査目的의 設定은 매우 重要하다.

그러나 간혹 調査目的이 不明確하고 抽象的이어서 具体的으로 어떠한 것인지 모호한 것이 있다. 그뿐아니라 調査過程에 있어서 調査目的이 忘却되어 調査目的과는 엉뚱한 다른 方向으로 調査가 行하여지기도 한다. 이와같이 調査目的에서 離脫되거나 調査目的이 変質되지 않기 위하여서는 첫째 調査目的을 明確하게 그리고 具体的으로 設定하여야 한다. 調査目的에는 一般的으로 무엇에 関한 知識을 必要로 하며 그 知識은 어디에 利用할 것인가를 明白히 하여야 한다

둘째, 調査目的과 아울러 생각하여야 할것은 調査目的을 實現하기 위하여 統計的方法이 採択될 수 있는가 하는 統計調査의 適合性問題이다. 統計調査는 社会現象을 把握하는데 있어 매우 有力한 手段이며 漸次 그 利用度도 높아가고 있지만 모든 分野에서 完全한 手段은 되지 못하고 있어 그 自体 커다란 限界를 가지고 있는 것이다 統計調査가 그 調査目的을 實現함에 있어 可能하고 適合한 것인가를 判断하기 위하여서는 다음 事項을 検討하면 된다.

1 調査의 目的으로 하는 知識이 集団에 関한 知識인가

2 調査対象이 現實的인 集団이거나 集団化시킬 수 있는 것인가.

3 그 集団에 統一된 標識을 付与할 수 있는가

4 그 調査結果가 数量으로 集団을 正確히 表現할 수 있는 것인가.

5 그리고 그 結果가 客観性과 普遍性이 있는가

6 統計調査가 可能하다고 하더라도 現實的으로 隨伴되는 困乱

Tax

National wealth

조사日程 ~~(Time Schedule)~~

조사목표
조사모집
조사범위 (Scope Coverage)

即 大量觀察에 따르는 費用, 時間, 人員을 克服할 수 있는가

乙. 大量觀察로 因한 얇고 平面的인 知識이 統計目的을 充足시킬 수 있는가

(2) 調査範圍

調査目的이 決定되면 다음에는 調査範圍를 定하는 것이 重要한 일이다. 調査範圍를 定하는 일은 곧 統計集団의 範圍를 定하는 것이며, 바꾸어 말하면 現實의 社会現象을 여러가지 要素에 의하여서 統計集団化시키는 것이다. 따라서 調査範圍는 곧 統計集団인 것이다.

調査範圍는 이를 概念的 範囲, 時間的範圍 및 場所的 範圍로 나누어 볼 수 있다. 概念的 範圍라함은 알려고 하는 社会現象의 属性(標識)을 말하는 것으로서 例컨데 鉱工業, 卸売物価등 社会現象의 種類를 限定하고 그 概念을 明確히 規定하여야 하는 것이다. 따라서 鉱工業이라 하더라도 그것이 어떤 規模의 것을 말하는 것인지 明確히 定하여야 하는 것이다.

時間的 範圍는 언제 現在의 社会現象을 調査할 것인가 하는 問題이다. 따라서 静態統計調査에 있어서는 一定한 時点을 그리고 動態統計調査에 있어서는 一定한 期間을 定하는 것을 말한다.

그리고 場所的 範圍라 함은 全国에 対한 社会現象을 調査할 것인가 또는 一部地域에 限한 社会現象을 調査할 것인가 하는 問題이다 調査範圍는 調査目的에 의하여서 規定되는 것이지만 人員, 費用, 時間등 与件에 맞추어서 이를 定하지 않으면 안될 것이다.

(3) 調査単位

統計調査에 있어서 実際로 調査의 対象이 되는 것이 調査単位이다. 即 必要한 情報를 提供하여 주는 것이 調査単位이다.

조사단위 (survey unit)
통계단위 (Statistical unit)
보고단위 (Reporting unit)

(1) Establishment unit
② Enterprise unit
③ Economic activity
④ Local.

조사의 단위

따라서 調査를 實施하기 以前에 調査單位를 무엇으로 할 것인가를 定하여야 한다.

그런데 調査單位는 統計集團을 構成하는 單位와 반드시 一致하는 것이 아님을 注意하여야 한다. 例를 들면 人口센서스에 있어서 人口라는 統計集團의 單位는 한사람 한사람의 個人이지만 調査는 家口를 調査單位로 하여 實施되는 것이다. 이와 마찬가지로 調査單位는 또한 集計單位, 分類適用單位, 抽出單位등과도 다르다. 集計單位는 集計에 있어서의 單位이며 分類適用單位는 各種 分類를 適用시키는데 있어서의 單位로서 모두 統計集團의 構成單位와 一致하는 것이다. 그리고 抽出單位는 標本調査에 있어서의 標本을 抽出하는 單位를 말하는 것으로서 全然 別個의 概念인 것이다.

調査單位를 定하는데 있어서 問題가 되는 것은 첫째로 무엇을 調査單位로 할 것인가 하는 것과, 둘째로 調査單位를 全部調査할 것인가 또는 一部만을 調査할 것인가 하는 問題이다. 後者는 곧 全數調査와 標本調査의 問題로서 言及되었으므로 여기서는 調査單位의 決定에 關하여서 說明코자 한다.

調査單位는 統計集團의 構成單位와 一致하는 것이 普通이지만 때에 따라서는 調査單位를 순전히 調査上의 便宜를 위하여 別途로 定하기도 한다. 調査單位는 가장 正確한 情報를 얻을 수 있고 또한 빠짐없이 情報를 모을 수 있는 最少限의 單位로 하는 것이 가장 좋다. 人口調査에 있어서는 家口가 調査單位로 되는 것이 普通이지만 鑛工業調査와 같은 産業調査에 있어서는 一般的으로 事業体를 調査單位로 한다.

(4) 調査項目 (Items to be covered)

調査目的이 確定되면 그 다음에 重要한 것은 調査項目을 決

기본항목 — 1. 身分的 항목
2. 家事的 항목

單記式. Individual
連記式 Collective.
한사람 하나씩 연기로 기재하는것

1. 만기수
2. 여기서
3. 결심이음

조사표 問題票 Form

Sheet or Booklet
· 問答形式 생산량 · 用語는 응답자가 충분히 이해할 수 있도록

質問의 배열 (결사시기) 회답하기 쉽도록 (충분한 space 해요)

定하는 일이다. 調査項目은 調査目的이 되어있는 커다란 主題를 여러 次元으로 細分한 것으로 調査票上에 印刷되어 實際로 調査되는 具体的인 事項이다.

調査項目은 調査目的에 関聯되는 必要하고도 充分한 것이 아니면 안되므로 調査項目을 設定하는데는 細心한 注意와 検討가 必要하다. 調査項目을 設定할 때에는 누구나 興昧가 있고 価値가 있을듯 하다고 생각되는 것은 무엇이나 다 包含시키고 싶은 誘惑을 느끼게 되는데 이는 絶対로 피하지 않으면 안된다. 調査項目은 必要하고도 充分한 最少限의 것이 가장 좋은 것이다.

調査項目 設定에 있어서 検討하여야 할 点은 첫째로 그 項目이 実際로 正確한 資料를 얻을 수 있는 것인가하는 것과, 둘째로 그 項目이 集計할 수 있는 것인가, 세째로 그 項目이 重点的으로 選定되고 調査項目에 寄与하는 것이 明白한가, 네째 調査時間, 調査員 被調査者 集計費등 現実的인 条件이 具備되어 있는가 等이다.

調査項目에는 그것이 集計되고 分析되는 基本項目과 附随項目으로서 基本項目을 보다 容易하게 調査할 수 있도록 誘導하는 誘導項目, 調査上의 오류를 予防하고 内容을 照会할 수 있는 対照項目 (Check Item)이 있다.

(5) 調査票 즉 問題형성제

調査票는 調査項目을 一定한 様式으로 配列한 紙面이다. 調査票는 統計調査에 있어서 매우 重要한 用具로서 그것은 첫째로 調査한 項目을 明確하게 함으로써 調査事項의 脱落을 防止할 수 있고, 둘째로 各異한 調査員의 観察을 標準化하고 統一시키며, 세째로 観察을 強化하고 正確히 할 수 있는 것이다.

調査票作成에 있어서는 調査票体裁, 用語形式 項目의 配列이 重要

7. 面接 (직접 記入할 것인가)
(우편으로 할 것인가)
8. 비고 . 란
9. 조사표색깔

한 問題이다.

1 体制: 調査票의 크기는 그것이 包含할 項目數와 調査票의 使用方法에 따라서 左右되겠으나 携帶하기 쉽고 取扱保管에 便利한 크기로 하여야 하며 比較的 簡單한 調査票에 의하여서 目的을 이룰 수 있는 경우에는 一般的으로 접을 必要가 없는 程度의 크기가 좋다. 調査票의 紙質은 記入, 分類, 集計, 保管등에 便利한 堅固한 것으로서 카드가 좋을 것이다. 그리고 事前 分類가 可能한 것은 調査票의 色을 달리하여 만드는 것도 分類 및 集計를 위하여 便利하다.

調査票의 体制에 있어서 그 크기와 紙質外에 重要한 것은, 첫째로 實査의 方法을 考慮하여야 하는 点이다. 實査의 方法에는 面接調査法, 留置調査法, 集合調査法, 郵便調査法, 電話調査法이 있으며 調査票의 記入方法에는 被調査者가 스스로 記入하는 自計法과 調査者가 記入하는 他計法이 있다. 調査票를 設計할 때에는 위의 어느 方法을 採用할 것인가가 이미 決定되어 있지 않으면 안된다.

그것은 어느 實査方法을 採用할 것인가에 따라서 調査票의 体制가 달라지기 때문이다. 그리고, 둘째로 이와 아울러 생각하여야 할 것은 調査員의 質도 考慮하여야 한다는 点이다. 有能하지 못한 調査員에게 複雜한 調査票는 極히 避하여야 한다. 調査票는 無能한 調査員도 容易하게 調査할 수 있도록 쉽게 꾸며지는 것이 좋은 것이다. 세째로 考慮하여야 할 것은 集計의 方法이다. 機械集計에 依하고자 할 때에는 機械集計의 專門家와 協議하여 集計에 便易하도록 調査票를 設計하여야 한다.

2. 項目配列: 調査員은 調査票上의 項目配列 順序에 따라 質問하게 되는 것이므로 質問의 順序가 論理的으로 矛盾이 없고 応

答의 効果를 높이기 위하여 調査項目의 配列에 특히 神経을 쓰지 않으면 안된다. 項目配列은 大体로 다음과 같은 順序에 의하는 것이 좋다.

가 被調査者가 応答하기 쉬운것 부터 始作할것.

나 可能한 限 論理的인 順序에 의하되 難問題는 中間 또는 끝가까이에 配列할 것.

다 一般的인 項目에서 細部的인 項目으로 展開할 것.

라 可及的 서로 関聯된 項目이나 集計할 때 함께 必要한 項目은 調査票의 같은 部分에 位置하도록 할 것.

以上에 依하여 項目을 配列하면 다음에 各項目에 一連番号 또는 符号를 부치는 것을 잊어서는 안된다. 그리고 調査票에는 반듯이 符号欄을 設置하여야 한다. 符号欄에는 各種 分類를 表示하는 番号 또는 符号를 記入하는 것이다. 符号欄은 調査票上의 適切한 位置에 設置하되 可能한 限 集計員이 보기 좋은 곳에 모으거나 当該回答의 가까이에 設置하는 것이 좋을 것이다.

3 用語形式: 調査票上의 用語나 文句는 그 意味가 完全하게 表現되어야 하며 普通의 知識을 가진 사람이면 充分히 明確하게 理解할 수 있는 것이 아니면 안된다. 調査説明書가 別途로 있다고 하더라도 可能한 限 調査票만 가지고도 調査項目의 意味를 正確히 알 수 있도록 하는것이 좋다. 調査票의 用語形式에 있어서, 특히 注意하여야 할 것은 調査項目이 主観的인 対答을 要求하는 것이어서는 안된다는 것이다. 調査票는 어디까지나 客観的이며 明確하고 簡単한 答을 얻을 수 있는 形式으로 꾸며져야 한다는 것이다.

(6) 分　類

統計調査의 結果 蒐集된 調査票는 調査対象에 関한 가장 詳細한 情報이긴 하지만 이를 一定한 方式에 의하여서 整理하고 集計하지 않으면 集団에 関한 知識을 얻을 수는 없다. 統計에 있어서는 一定한 方式에 의한 整理, 即 分類가 絶対的으로 必要한 것이며 이는 반듯이 全調過程에서 그 基準이 設定되지 않으면 안되는 것이다. 分類라 함은 調査対象을 몇個의 구룹으로 나누는 것으로서 同一한 구룹에 나누어진 것은 同質的인 것으로 取扱되고 統計数字는 이 구룹에 関하여서만 나타나게 되는 것이다. 따라서 구룹内에 있어서의 異質性은 無視되고 個別的인 特殊한 情報는 捨象되어 버리고 마는 것이다.

이와같이 分類란 調査結果의 複雑한 情報의 集計中에서 必要한 것을 가려내어 簡単한 形態로 整理하는 것을 말하는 것이다. 分類에는 質的分類와 量的分類가 있다. 質的分類란 事業(産業)의 種類나 職業의 種類 或은 性別과 같은 質的인 内容에 関한 分類이며 量的分類란 사람의 年令, 事業体의 従業員数나 資本金과 같은 量的인 内容에 관한 分類이다. 위의 어느것이든 事前에 한 個. 한 個의 구룹을 어떤 基準에 의하여서 어느 範囲까지로 하느냐 하는 것을 定하지 않으면 안되는 것이다.

各種分類를 作成하는데 있어서는 다음과 같은 点을 考慮하여야 한다

1 調査의 目的에 따라서 分類를 作成할것

2 分되된 各 구룹内에 있어서는 모든 単位가 同質的이 되도록 할것.

3 구룹과 구룹間은 異質性이 뚜렷이 나타나도록 할것

4 分類된 単位가 하나도 남김없이 그리고 重複됨이 없이 어

어떤 구룹엔가 歸屬되도록 할것.

5 다른 同種의 調査結果와 比較될 수 있도록 分類를 作成할 것. 따라서 標準分類가 있는 것은 되도록 이를 採用할것.

6 量的分類에 있어서는 合理的인 区間을 設定하되 統計의 一般原則에 따를것.

7 質的分類에 있어서 그 内容이 複雜한 것은 十進分類法에 의하여서 小分類, 中分類, 大分類와 같이 体系的으로 分類하는 것이 便利하다.

8 分類가 細分되어 구룹이 過多할 때에는 集団의 内容은 자세히 볼 수 있으나 全体的 把握에 不便하다. 그 反面 分類가 大分되어 구룹이 過少할 때에는 1個의 구룹속에 異質的인 것이 無理하게 区分되기 쉽다.

9 2個 以上의 구룹에 重複되어 屬하게 되는것 또는 2個의 구룹 中間에 位置하는 것은 그 歸屬을 明白히 規定하여 두어야 한다.

(7) 調査方法 — 全數조사로 할것인가 標본조사에 의할것인가, 回收하나 面接하나:

여기서 말하는 調査方法이란 實査(現地調査)의 方法을 말하는 것인데 이는 調査票設計나 調査対象의 決定과 同時에 定하여져야 하는 것이다. 實査의 方法에는 大体로 다음 다섯가지가 있다.

1 面接調査法:

調査者가 被調査者를 直接 面接하여 質問과 應答을 通하여 調査하는 方法이다. 이 方法에 依하면 被調査者의 應答을 確實히 얻을 수 있고 調査票의 回收率을 높이며 比較的 길고 複雜한 調査票도 使用할 수 있을뿐 아니라 正確한 調査를 期할 수 있고 補充的인 情報도 얻을 수 있다는 長点이 있는 反面 많은 調査員

調査時期
가. 基準時点 (Base Period) 70.10.1 10.31 31
나. 対象기간 (Reference period) 61 — 12.31 71—6.30
다. 実査기간 (Field Survey 기) 10.1 —10 6.1

가 費用이 所要되고 調査員에 따라서 応答이 달라지기 쉬우며 調査員의 不正行為가 発生할 念慮가 있는点 및 特殊한 階層의 사람은 面接할 수 없다는 点等의 短点이 있다.

2 配票調査法:

調査者가 被調査者에게 調査票를 配付하고 一定期間 後에 이를 回収하는 方法이다. 이 方法은 面接調査法에 比하여 追及訪問이 덜 必要하고 調査票의 回収率도 比較的 좋으며 費用도 적게 들지만 被調査者 本人이 記入하지 않거나 本人이 記入한다고 하여라도 調査項目을 誤解하여 記入이 不正確하게 되기 쉬우며 또 内容이 造作되거나 記入漏落이 생길 可能性이 많은 것이다.

3 集合調査法:

被調査者를 一定한 場所에 集合시켜 同時的으로 調査票에 記入케 하는 方法이다. 이 方法에 依하면 調査의 説明이나 条件을 被調査者 全員에게 対하여 統一시킬 수 있는 利点이 있으며 被調査者를 容易하게 集合시킬 수만 있다면 費用이 적게 들고 簡便하게 調査를 行할 수 있으며 調査員도 小数로 足할 것이다. 그러나 集合調査法은 特別한 경우를 除하고는 使用하기 困難할 뿐 아니라 一般的으로 出席率이 나쁜 点과 自計式의 一般的 欠陥을 免할 수 없는 点, 被調査者의 差를 無視하게 되는 点, 出席者의 日当이나 交通費를 支給하게 된다면 오히려 費用이 많이 든다는 点等의 欠陥이 있다.

4 郵便調査法:

調査票를 被調査者에게 郵送하고 이를 다시 郵便에 依하여 回収하는 方法이다. 이 方法은 費用이 적게 들며 넓은 地域에 調査票를 配布할 수 있으며, 面接調査가 어려운 特殊한 階層의

사람에게도 調査를 할 수 있는 長点이 있다. 이 反面 調査票의 回収率이 낮고 回収가 어려움에 따라 督促状을 여러번 보내야 되므로 回収에 時間을 要한다. 그리고 配票調査에 있어서와 같이 自計式 一般의 欠点이 있으며 調査忌避를 防止하기 為하여서 調査票가 必然的으로 簡単해지는 等의 短点이 있다.

5 電話調査法:

이는 調査者가 被調査者에게 電話로 質問하여 調査하는 것이다. 이 方法에 依하면 簡単하고 迅速하게 調査를 할 수 있으며 費用도 적게 들 수 있다. 그러나 電話를 가진 사람만을 対象으로 하여야 하는 点, 調査時間이 짧아야 되므로 簡単한 調査 以外는 할 수 없다는 点等이 決定的인 短点이다.

위의 여러 方法中 集合調査法 等은 制限된 特殊한 調査에만 使用되는 方法이며 一般的으로는 面接調査法, 配票調査法, 郵便調査法의 順으로 많이 使用되는데 特히 面接調査法은 回収率이 높고 正確하고 詳細한 調査를 期할 수 있다는 長点으로 因하여 가장 널리 使用되고 있다.

따라서, 위의 方法 以外에 調査票 記入方法으로서 自計式과 他計式이 있는 데 自計式이란 被調査者가 스스로 記入하는 것이고, 他計式은 調査員이 記入하는 것이다. 따라서 面接調査法과 電話調査法은 他計式에 依하는 것이며, 配票調査法, 集合調査法, 郵便調査法은 自計式에 依하는 것이다. 이 中 他計式은 費用의 問題와 調査員 偏倚의 問題가 있긴 하여도 自計式에 比하여 훨씬 優越한 것으로 생각되고 있다.

(8) 調査員

調査員은 広義에 있어서는 調査管理者, 現地調査員 및 集計員

을 包含한 모든 調査從事者를 意味하나 狹義로는 現地調査員만을 가르킨다. 여기서는 現地 調査員中에서도 特히 面接調査法이 採用되는 경우의 調査員(面接調査員)을 中心으로 하여 說明하고자 한다

社会現象을 다루는 모든 調査에 있어서는 被調査者가 얼마나 眞実한 情報를 提供하여 주는가 하는 被調査者의 協力的 態度와 調査員의 正確하고 熱誠的인 觀察態度가 가장 重要하다 할 것이다. 被調査者는 各己 性別, 年令, 学層, 階層, 思想, 性格 等이 다른 各様의 属性과 個性을 가지고 있어서 그들에게서 한결같은 協力을 얻고 그들을 正確히 觀察한다는 것은 容易한 일이 아니다. 調査員은 이와같이 어려운 일을 担当하는 것이며 調査員의 能力과 努力 如何에 따라서 그 調査의 成功 与否가 左右된다고 하여도 過言이 아니다.

따라서 優秀한 調査員을 確保하고 充分히 訓練시키는 일이 調査를 成功케하는 하나의 關鍵이 라고 할 수 있는 것이다.

1. 調査員의 資質과 性格:

調査員에 適合한 性格으로서는 責任感, 誠実, 明朗, 順応性 探究心, 忍耐力, 公正性 等이 要求되며 正確한 理解力과 判断力 그리고 調査에 対한 興味 等이 要望된다. 그外의 条件으로서는 健康하고 普通水準 以上의 教養과 知識이 必要하며 調査에 専念 할 수 있는 時間的 餘裕가 있을것이 要求된다.

調査員의 年令은 20歳 ~ 30歳가 適合하며 性別은 一般的으로 男性이 適合하나 경우에 따라서는 女性이 適合 할 때도 있다. (例 家計調査) 또 経験이 없는 調査員보다는 経験이 있는 調査員이 一般的으로 낫다고 할 수 있겠으나 正規의 訓練을 받지 않고 漠然하게 調査員으로서의 経験만을 가지고 있는 것은 반듯이 낫다고

할 수 없으며, 오히려 調査에 無誠意하고 內容을 造作할 可能性이 있다.

2 調査員의 偏倚 :

調査員은 그의 性格과 資質에 따라서 一定한 偏倚를 갖기 쉬운데 이는 調査員의 偏見, 先入觀, 思想, 性格 等에 基因하는 것으로 同一한 対象의 回答을 듣고도 各異한 두사람의 調査員에 있어서는 各異한 結果가 나타나는 것이다.

調査員에게는 自己의 立場이나 主觀, 信念에 따라서 事物을 理解하려는 傾向이 있기 때문이다. 調査員의 이러한 偏倚는 調査結果에 커다란 誤差를 가져오기 쉬우므로 調査員을 採用 할 때나 調査員을 指導訓練 할 때 特히 이를 避하고 除去하도록 努力하여야 한다.

3 調査員의 訓練 :

調査員의 適合한 資質은 반드시 生來的인 것은 아니며 各種의 訓練에 依하여서 鍊磨되고 完成 될 수 있는 것이다.
勿論 生來的인 素質도 重要한 것이므로 좋은 調査員을 얻기 위하여서는 知能, 性格 等에 対한 「테스트」에 依하여서 미리 選拔하여야 하겠으나 實際問題로서 調査員採用은 그다지 慎重히 다루어지지 않고 있으므로 特히 調査員의 指導와 訓練을 徹底하게 하지 않으면 안되는 것이다.

調査員訓練의 目的은 調査員에게 適合한 資質을 키워 주고 調査에 関한 直接的인 知識을 付与하며 面接技術을 鍊磨시키는 데 있는 것이다.

따라서 實際로 調査員訓練을 為한 講習을 實査 直前의 調査準備過程에서 實施하게 되는데 이때에 調査員에게 調査의 目的과 意義

를 充分히 認識시키고 아울러 具体的으로 調査項目의 定義, 記入方法 其他 実査中에 일어날 것으로 予想되는 여러가지 事項에 関한 処理方法 等을 詳細히 指示하여야 한다. 그리고 이러한 具体的인 指示는 반듯이 調査要領書나 調査指示書로서 作成되어 調査員에게 交付되어야 한다.

調査員의 訓練方法으로는 実地로 練習調査를 実施하여 그 結果를 1枚씩 精査하게 検討하여 調査員이 誤解하고 있는 点 調査票記入方法의 不充分한 点을 指摘하고 注意를 喚起하는 方法이 많이 使用되고 있다.

그런데 調査員에 対한 調査要領의 指示는 調査員을 한 場所에 集合시켜 一時에 하는 것이 좋다. 그것은 調査指示者와 時間 및 場所에 따라 指示内容이 달라질 念慮가 있으며 그로 因하여 調査員의 統一된 知識과 観点을 期할 수 없게 되기 때문이다.

(9) 調査区 Enumeration District Sealing

調査의 範囲가 狭小한 1個地域에 局限되어 있거나 標本調査인 경우에는 特別히 調査区를 設定할 必要가 없으나 「센서스」와 같이 全国 또는 道와 같은 広大한 地域을 調査의 範囲로 하는 경우에는 한사람의 調査員으로서는 全地域을 担当할 수 없고 여러사람의 調査員이 地域을 分担하여야 하므로 이들 各 調査員의 分担地域을 設定할 必要가 있다.

이와같이 各調査員의 分担区域을 調査区라고 하는 것이다. 調査区를 設定하는 目的은 調査地域内에 있는 調査対象을 하나도 脱落하지 않고 또한 重複됨이 없이 포착하려는 데 있다. 따라서 調査区를 設定할 때에는 脱落된 地域이 없는가 調査区와 調査区間의 境界가 明確한가를 特히 念頭에 두어야 한다. 한 調査区에는 한 調査員을

調査区 없이는 소재지

配置하는 것이 原則이므로 調査区의 數는 動員 可能한 調査員의 數에 依하여 定하는 것이 普通이다. 그리고 調査区의 設定方法으로는 行政区域에 依하여 市·郡単位 또는 里·洞単位로 하는 方法과 地勢및 交通에 따라 設定하는 方法 및 各 地域内에 있는 予想調査対象數에 依하여 決定하는 方法 等이 있다. 調査区를 設定할 때는 위의 方法中 어느 方法을 択할 것인가를 決定하고 特히 全国的인 「센서스」에서와 같이 많은 調査区를 設定하여야 할 때에는 可及的 크게 拡大된 地圖를 準備하여 이에 依하여 区域을 明確히 測定하는 것이 좋다. 그리고 調査区가 많을 때에는 数個의 調査区를 指導 및 管理할 指導区를 設定하는 것이 調査管理上 効果가 있다.

(10) 調査時期

調査의 企図이 完了되고 이에 따른 準備가 갖추어지면 調査를 実施하게 되는 것이므로 調査時期를 特別히 考慮할 必要가 없을 듯 하나 事実은 調査의 時期가 企劃過程에서 먼저 定하여지고 이에 맞추어 企劃 및 準備日程이 짜여지는 것이 原則이다. 그것은 調査時期가 調査의 進行과 結果에 적지 않은 影響을 주는 때문이다 万一 調査의 時期가 調査의 時間的 基準과 너무 떨어진 1年 또는 數年前의 事実을 調査한다면 그 調査結果는 内容이 不正確할 뿐 아니라 結果의 利用価値도 낮아지게 될것이며 또 調査時期가 調査対象을 찾아가기 어려운 季節이라면 그 調査는 中途에서 일단 保留를 하지 않을 수 없게 될 것이다.

따라서 調査時期는 調査가 可能하고 容易하게 되도록 調査基準時点으로 부터 너무 오랜 時日이 지나지 않은 때를 択하도록 하여야 한다. 具体的으로 말하면 調査가 可能하고 容易하기 위하여서는 먼저 그 以前에 企劃과 準備가 完了될 可能性이 있는가 하는 것과

그때에 必要한 調査員을 動員할 수 있는가 하는 点, 그리고 調査對象이 正確히 把握될 수 있고 被調査者가 充分히 情報를 提供할 수 있는 準備가 되어 있는가 等을 考慮하여야 한다는 것이다. 調査時期를 調査基準時点에서 너무 멀리 택하는 것도 좋지않지만 調査基準時点에 너무 近接시켜도 좋지 않은 경우가 있는 것이다. 例를 들면 企業体의 生産 또는 財務活動調査에 있어서는 적어도 企業体가 帳簿整理를 끝내거나 決算을 完了한 後를 調査時期로 택하여야 할 것이다.

그리고 調査를 年1回 實施하는 경우에는 季節的인 條件을 勘案하여 炎署酷寒을 避하도록 하여야 하며 또 被調査者가 大部分 바쁜 時期는 택하지 않는 것이 좋다.

實査期間은 위의 調査時期에 一定한 期間을 設定한 것인데 그 期間의 長短은 調査對象數와 調査內容 및 調査員數에 依하여서 決定된다. 實査期間은 짧을수록 좋은 것이므로 有能한 調査員을 많이 動員할 수만 있다면 費用의 問題는 있으나 可及的 實査期間을 短縮시키도록 하는 것이 좋다.

(11) 實査管理, 集計等 計劃

統計調査의 企劃은 調査準備로 부터 結果公表의 最終段階까지의 全過程에 걸친 事項을 全部 網羅하여 이루어져야 하므로 實査段階 以後에 이루어질 實査管理, 調査票審査, 集計 等에 關하여서도 企劃過程에서 미리 計劃되지 않으면 안된다. 實査管理에 關하여서는 正確하고 迅速한 實査의 運營方法을 細密한 部分까지 計劃하여야 하며 調査票審査에 關하여서는 調査票의 內容檢討 및 分類를 効果的으로 할 수 있는 方法을 研究하여야 한다.

그리고 集計에 있어서는 集計方法, 集計期間, 集計場所, 集計員, 集計表樣式, 公表樣式 等에 関하여서 具体的인 計劃이 이루어져야 한다.

特히 集計計劃은 調査項目의 設定과 調査票設計時에 있어 同時에 考慮되어야 한다. 그러므로써 集計가 不可能한 것이나 利用上에 意味가 없는 項目은 除外될 수 있는 것이다.

以上의 実査管理, 調査票審査 및 集計에 関한 具体的 事項은 後述하고자 한다.

(12) 被調査者의 協調

実査는 被調査者를 相対로 하는 것이며 被調査者가 積極的으로 調査에 協調하지 않고서는 調査의 目的을 達成할 수 없는 것이므로 事前에 被調査者의 協調를 얻는 方案을 講究하여 두는 것이 主要한 일이다.

実査는 그것이 어떤 種類의 것이든 他人의 時間과 情報를 侵害하는 것이므로 被調査者의 協調를 얻기 위하여서는 무엇보다 먼저 이와 같은 侵害를 正当化시키고 相対方에 認識시키는 것이 緊要하다. 이를 위하여서는 이調査가 被調査者와 社会를 위하여 実施되는 것이며 有益한 結果를 가져올 것이라는 点, 그리고 調査実施機関이 權威있는 機関이거나 公益機関인 点等을 強調하여야 한것이다.

또한 被調査者가 調査에 非協調的인 理由의 하나가 秘密漏洩에 対한 疑慮에 있으므로 이러한 疑惑를 払拭시키도록 努力하는 것도 重要하다.

따라서 調査員이 実査에 着手하기 前에 被調査者에게 調査의 目的과 趣旨, 調査協調에 関한 懇曲한 付託, 調査内容은 絶対로 秘密로 한다는 内容의 調査 協調依頼文을 보내는 것이 좋다. 그리고 予算이 許容하면 이러한 個別的인 協調依頼文 外에 新聞, 放送, 포스타

等을 通하여 調査의 重要性을 널리 周知시키는 것도 좋다. 이밖에 調査實施機關이 權限이 없거나 公的機關이 아닌 때에는 다른 有力機關의 支援을 받는 것이 매우 効果가 있다.

被調査者의 協調를 얻는 方法으로서 謝礼品을 주는 것도 좋은 方法이다. 謝礼品을 주기 위하여서는 莫大한 費用이 所要되지만 被調査者의 協調와 手苦에 報答하는 뜻으로 實査 後에 謝礼品을 贈呈하는 것이 効果的이다.

또한 實査를 担当한 調査員에게는 調査員의 身分을 証明하고 調査機関을 明示한 調査員身分証을 発給하여 實査時에 被調査者에 提示케 할 뿐 아니라 服裝이나 言語를 端正하게 하여 좋은 印象을 주도록 하는 것도 被調査者의 協調를 얻는데 크게 도움이 될 것이다.

(13) 予備調査 (試驗調查)

實査에 들어가기 前에 企劃된 内容들이 現實的으로 妥当한가 하는 것을 確認하기 위하여 予備調査(準備調査 또는 試驗調査라고도 함)를 하는 것이 普通이다. 予備調査에서는 主로 調査単位 分類方法, 調査票様式, 調査項目의 定義, 調査方法, 調査員의 業務量, 被調査者의 協力程度 等과 같은 調査企劃上에 나타난 諸事項에 關하여 그 適否를 検討, 確認하고자 實施하는 것이지만 때에 따라서는 「센서스」의 予備名簿를 作成하거나 標本名簿를 作成하기 위한 것과 같은 實査의 前提的인 作業으로서 實施되는 경우도 있다. 따라서 企劃上의 欠陷을 是正하기 위하여서 實施되는 予備調査를 特히 試驗調査라고 하며 이것은 企劃이 全部 完了된 後에 實施되기보다 企劃途中에 必要에 따라 數次에 걸쳐 實施되는 경우가 더 많다.

나. 実験調査(現地調査)

(1) 実査의 重要性

統計調査의 企劃과 準備가 끝나면 実査의 段階로 들어가게 되는데 이 段階에서는 調査員이 被調査者로 부터 情報를 얻고 이를 調査票에 記入하며, 調査管理者는 実査를 管理하고 記入이 完了된 調査票를 蒐集하게 된다. 実査의 方法에는 面接調査法, 配票調査法, 集合調査法, 郵便調査法, 電話調査法등이 있음은 既述한 바 있으나 여기서는 面接調査의 경우를 中心으로 하여 説明하기로 한다.

統計調査가 物理現象을 다루는 경우에 있어서는 대개의 경우 調査対象이 下等動物이건 無生物이건 간에 対象과 論議하는 일이란 있을 수 없는 것이다. 그러나 統計調査가 社会現象을 다루는 경우에는 調査対象인 個人 또는 社会集団이 積極的으로 이에 參与하지 않고서는 不可能한 때가 많다.

다시말하면, 社会現象에 関한 調査가 成功하느냐 失敗하느냐의 重要한 関鍵은 被調査者가 얼마나 積極的으로 協力하여 真実한 情報를 提供하여 주느냐에 달려있다.

그런데 被調査者는 各異한 個性과 特徵을 가지고 있으므로 이러한 被調査者로 부터 한결같이 積極的인 協調를 얻고 正確한 調査를 期한다는 것은 결코 容易한 일이 아니다. 社会現象을 다루는 統計調査에 있어서 実査의 技術的인 方法이 크게 重要視되는 것은 이 때문이다.

아무리 統計調査의 企劃이 잘 되었다 할지라도 実査過程에서 無能하고 無誠意한 調査員의 熟練되지 못한 調査는 内容이 不正確한 結果를 가져올 뿐만 아니라 被調査者의 非協調的인 態度만 助長시킴으로써 将来의 調査마저도 困難하게 만들 憂慮가 있는 것이다.

大体로 被調査者는 調査를 忌避하거나 真実한 対答을 拒否하려는 傾向이 있다.

이것은 自己의 事業上의 秘密이 外部에 漏泄됨으로써 同業者間의 競争에 不利하여지거나 納税額에 影響을 주지 않을까 하는 疑懼心과 調査에 応하려면 바쁜 때의 貴重한 時間이 浪費된다고 생각하기 때문이다. 그러나 우리가 調査目的을 達成하기 위하여는 이러한 被調査者를 우리가 意図하는 대로 이끌고가서 真実한 対答을 하도록 만들지 않으면 안되는 것이다.

그러므로 이를 위하여서는 무엇보다도 調査員이 훌륭한 資質과 態度 그리고 充分한 知識을 具備하고 誠意있게 調査에 臨하여야 하며 또한 그때 그때의 状況에 適応하여 適切한 技術的인 面接方法을 항상 研究하지 않으면 안되는 것이다.

(2) 実査의 管理

調査員에 의한 実査는 調査가 企図된 대로 원활히 遂行되도록 厳格히 管理되지 않으면 안된다. 먼저 調査員이 実査에 着手하기 前에 調査管理者는 調査員으로 부터 実査計画書 또는 実査日程表를 作成토록하여 이를 検討하고 調節하여야 한다. 그리고 이 実査計画書에 依拠 調査員의 実査를 管理監督하는 것이다. 調査員이 実査를 進行하고 있는 동안에는 調査員이 調査対象을 틀림없이 訪問하여 正確한 調査를 하도록 이를 統制하여야 한다. 調査員中에는 被調査者를 만나지도 않고 卓上에서 調査票를 造作記入하는 者도 있으며, 被調査者를 訪問하지도 않고 또는 但一回 訪問하고서는 被調査者의 不在 或은 調査不能이라고 報告하는 事例가 있으므로 調査管理者는 이러한 일이 発生하는 것을 最大限 防止하여야 하는 것이다.

調査員의 이러한 行為를 統制하는 方法으로는 調査管理者가 不時에

任意의 調査対象을 訪問하거나 調査対象에게 郵便 또는 電話로 調査員의 訪問与否와 態度를 照会하는 等의 方法이 있으며 이 外에 調査票上에 対照項目을 設置하여 調査員을 統制하는 方法도 있다.

卓上造作은 않더라도 他人에게 調査를 依頼하거나 電話 또는 郵便으로 調査하는 것도 이를 防止하여야 한다. 実査 途中 調査不能이 나오는 경우에는 調査管理者는 그 事由를 検討하고 再調査를 指示하거나 標本調査인 경우에는 調査対象인 標本을 代替하여야 하는데 이때 注意할 것은 標本의 代替는 調査員이 任意로 하여서는 결코 안되며 調査管理者에 의하여 厳格히 다루어져야 한다는 것이다. 그리고 이와 別途로 調査管理者는 調査不能의 原因을 詳密히 分析하여야 한다.

調査不能에는 여러가지 原因이 있겠으나 大体로 調査対象이 所在不明인 경우, 調査対象이 不適格한 경우, 被調査者가 不在인 경우, 被調査者가 調査에 不応하는 경우등에 調査不能이 나타나는 것이다.

調査対象의 所在가 不明한 것은 当初 名簿가 잘못 作成되었거나 名簿가 作成된 後에 調査対象에 変動이 생긴 때문인데 이러한 경우에는 可能한 限 名簿를 修正하고 調査를 하여야 한다. 그리고 調査対象이 不適格한 경우, 例컨데 調査対象이 調査範囲 밖에 있는 때와 같은 경우에는 大体로 実査를 할 必要는 없으나 調査管理者는 반드시 그 事実与否를 確認하여야 한다. 被調査者가 出他하여 不在인 경우에는 결코 調査不能으로 処理하여서는 안된다. 이 때에는 반드시 再次 訪問하여 調査를 遂行하여야 한다.

그런데 가장 困難한 것은 被調査者가 調査에 応하기를 拒絶하는 경우이다. 그러나 이 경우에도 調査員은 1次 拒絶当하였다 하여 調査를 포기하여서는 안되며 再次 誠意있는 努力을 하여야 한다.

그래도 調査를 不應하는 경우에는 調査管理者가 直接 被調査者를 訪問하여 実査를 하여야 한다.

이 外에 調査管理者는 調査員이 実査 途中에 実査에 関한 疑問点을 問議하여 왔을 때 이에 対答하여 주어야 하며 調査員의 実査上의 各種 隘路点을 解決하여야 한다. 뿐만아니라 調査員이 不意의 事故로 調査를 遂行할 수 없게 되는 경우에는 期間内에 実査를 完了할 수 있도록 格別한 対策을 講究하여야 한다.

調査員은 実査가 끝난 調査票를 実査管理者에게 提出케 되는 데 이에 앞서 調査員으로 하여금 스스로 調査票를 徹底히 検討하고 完全히 整理하여 提出케 하여야 한다. 即 調査員으로 하여금 毎日 調査가 끝났을 때 그날의 調査된 調査票에 대하여서는 記入漏落, 記入의 不完全, 記入上의 錯誤를 스스로 発見하여 是正케하며 또한 글자를 알기 쉽게 하고 記入法의 統一을 期하며 計算한 것이 있으면 이를 検算하고 경우에 따라서는 計算値의 記入을 行하게 하는 등 調査票를 整理하도록 한다. 그리고 調査員이 割当된 調査対象에 대하여서 調査를 全部 끝냈을 때는 指定된 対象에 대하여 빠짐없이 그리고 錯誤없이 調査가 完了되었는가를 스스로 確認케한 後에 調査票를 提出하게 하여야 한다.

調査員이 記入 完了한 調査票를 提出하였을 때에는 되도록 現地에서 調査員을 面前에 두고 調査票를 審査하여야 한다. 여기서 調査票上의 記入漏落, 記入不完全, 記入의 錯誤, 調査員의 造作등을 発見하여 이를 修正케 하거나 再調査를 指示하여야 한다.

이어서 調査員이 割当된 実査를 모두 終了하였을 때에는 所定期日에 一定한 場所에 調査票를 収集하고 여기서 調査対象別로 調査票를 点検하여 調査漏落이나 重複調査 및 調査対象의 同一性 与否

등을 確認하여야 하는 것이다.

(3) 被調査者와의 面接方法

(가) 面接의 準備: 調査員은 被調査者와의 面接을 實施하기 전에 먼저 面接調査를 가장 効果的으로 할 수 있도록 計劃을 세워야 한다. 即, 첫째로 어떻게 하면 時間과 距離를 短縮하여 被調査者를 訪問할 수 있는가를 硏究하여 路程日程表를 짜야하며 調査에 앞서 調査要領書나 調査指示書를 熟読하고 調査票, 調査要領書, 記入道具, 地図등 持參할 物件을 点検하여야 한다. 다음 누구를 面接対象으로 할 것인가를 定하여야 한다. 面接対象으로 가장 適合한 사람은 말할것도 없이 調査内容에 가장 精通하고 同時에 調査員에게 自由로히 말할 수 있는 位置에 있는 사람이어야 한다. 事業体内容을 調査하기 위하여서는 그 事業体의 幹部를 面接하지 않으면 안될 것이다.

面接対象이 定하여지면 調査員은 可能하면 그를 訪問하기 전에 그에 관한 知識, 即 그의 性格, 過去 및 그가 属하여 있는 部分社会에 있어서의 慣行등 予備知識을 얻도록 하여야 한다.

이와같은 予備知識은 面接을 원활하게 이끌고 被調査者의 積極的인 協調를 얻는데 큰 도움이 된다. 다음 調査員은 被調査者에게서 調査할 問題에 관하여 大略的인 知識을 갖추어야 한다. 그리고 状況이나 条件에 따라 面接進行의 大綱을 事前에 計劃하여 두는 것도 面接의 時間을 節約하고 序顺를 効果있게 만들며 要点을 強調할 수 있다는 点에서 반드시 잊어서는 안될 것이다.

끝으로 面接에 앞서 面接場所와 面接時間을 定하는 것도 重要하다. 面接場所와 時間은 調査員의 便宜만에 의하여서 定할 것이 아니라 被調査者가 가장 安楽하게 이야기할 수 있는 条件을 考慮

하여 定하여야 한다. 一般的으로 事業体의 内容을 調査하는 것이라면 面接場所로는 그 事業体의 事務室이 가장 適当하다. 그것은 必要한 帳簿나 記録을 即時 参考할 수 있기 때문이다. 그리고 面接時間은 被調査者가 그다지 바쁘지 않고 별다른 일이 없으며 피로를 느끼지 않은 때를 択하는 것이 좋다. 可能하면 事前에 場所와 時間을 被調査者와 約束하는 것이 좋을 것이다.

(나) 面接의 順序: 面接은 大概 다음과 같은 順序에 따라 進行된다.

第1段階: 最初로 被調査者와 接触하는 段階로서 먼저 所持한 身分証이나 명함을 提示하고 人事를 交換하며 自己의 所属된 調査機関을 밝힌다. 그리고 분위기를 造成하여 가면서 面接의 趣旨와 目的을 相対가 納得할 수 있도록 説明한다. 이때에 所持한 調査協調依頼文이 있으면 이를 주도록 한다.

第2段階: 여기서는 두사람의 呼吸을 調整하고 親密하고 부드러운 분위기를 만들어 相対方이 親密感을 갖고 積極的으로 協調하도록 努力한다.

第3段階: 最初의 接触이 끝나고 또 調査員과 被調査者間에 親密한 분위기가 造成되면 被調査者를 主役으로 하여 一般的이고 全体的인 이야기를 交換하게된다. 이 段階에 있어서 이야기하는 사람은 主로 被調査者이며 調査員은 主로 相対方의 이야기를 関心깊게 들어야 한다. 例를 들면 相対方으로 하여금 一般的인 景気라든가 業界의 動向을 이야기 시키는 것이다.

第4段階: 이 段階에 있어서는 調査員이 主役이 되어 本格的인 質問을 한다. 調査員은 大体로 주어진 調査票의 順序에 따라 한項目 한項目씩 質問하고 그 対答을 調査票에 記入하는 것이다.

第 5 段階 : 이 段階에서는 相対方이 特히 하고 싶은 이야기나 意見을 듣고 이를 参考로 하여 調査를 補完한다. 그리고 被調査者가 前段階에서 錯誤한 것을 깨닫고 스스로 訂正하여 주는 경우에는 이를 確認하여 調査票의 記入을 訂正한다.

第 6 段階 : 이 段階에서는 調査가 完全히 이루어졌는가를 確認한다. 即 調査員이 調査票上에 記入漏落이나 記入不完全, 記入錯誤가 없는가 調査票를 検討하는 것이다. 그리하여 만약 調査票上의 不備를 発見하면 即席에서 이를 補完하여야 한다. 이 段階에서의 調査票検討는 매우 重要한 일인데도 一般的으로 소홀히 생각되기 쉬운 데 여기서 이를 소홀히 하면 後에 이를 補完하기란 極히 힘든 일이므로 특히 주의를 要한다.

第 7 段階 : 面接이 끝났으므로 感謝의 뜻을 表示하고 面接場所를 물러난다. 感謝는 真心으로 表示하여야 하며 설혹 目的을 達成하지 못한 경우라도 他人의 時間을 多少나마 소비하게 하였으므로 반드시 感謝의 뜻을 表示하여야 한다. 그리고 이때 今後의 継続調査를 確保할 수 있는 素地를 만들도록 親密하고 좋은 인상을 남기도록 하여야 한다.

以上으로 面接의 順序를 段階的으로 区分하여 説明하였으나 이 順序는 柔軟性이 있는 것이므로 경우에 따라 상당히 変形되지 않으면 안되는 것이다. 그리고 週期的으로 調査되는 継続調査에 있어서는 第 1 段階와 第 2 段階는 省略되는 것이 보통이다.

(다) 面接上의 注意

1 端正하고 素朴한 외모를 갖추어 被調査者에게 좋은 印象을 줄것.

2 부드러운 態度와 말씨로 自然스럽고 率直하게 이야기할

수 있는 温和한 분위기를 만들어 겸손과 예절을 잃지 않도록 할 것.

3 自由롭게 이야기하면서도 真摯性과 沈着性을 나타내어 相対方이 軽蔑感을 갖지 않도록 하며 相対方의 信頼를 얻도록 할것.

4 相対方에게 本心에서 우러나오는 따뜻한 理解와 同情을 가지고 接한다.

5 相対方의 気分, 言外의 말, 익숙치못한 表現에서 相対方의 真意를 把握할 수 있도록 細心한 洞察力을 発揮할 것.

6 相対方과 論争을 하거나 相対方을 軽蔑하는 빛을 보이지 말것.

7 自己의 態度 質問 判断이 正当한가를 反省하고 検討할 것. 自己의 判断에 너무 確信을 갖는 것은 훌륭한 調査員이 아니다. 8

8 忍耐와 責任感을 가질것. 指定된 被調査者를 찾아서 만나지 못한 때에는 몇 번이라도 다시 찾아가야 하며 때로는 相対方에 의하여 참을 수 없는 일을 당하더라도 忍耐할 것. 그리고 真実이 아니라고 생각되는 것은 安易하게 処理하지 말고 끝까지 真実을 찾아낼것. 이것은 任務에 忠実한 責任感에서 생기는 것이다

9 끊임없는 関心과 研究로서 充分한 知識을 錬磨할 것.

10 相対方과 面接하지도 않고 調査員이 멋대로 調査票를 造作記入하는 일이 결코 없도록 할것.

다. 調査票의 審査와 集計

(1) 調査表의 審査

調査表의 審査는 実査에 의하여서 蒐集된 調査票의 内容을

審査하여 錯誤나 記入漏落을 発見하고 이를 訂正하여 集計할 수 있도록 調査票의 記入内容을 完全하게 하는 作業이다.

調査票審査作業의 内容을 살펴보면 다음과 같다.

(가) 指定된 調査対象에 관하여 調査票가 確保되었는가를 点検하는 것.

(나) 調査票中에 記入漏落된 項目이 없는가 検討하는것.

(다) 記入이 不完全한 項目을 検出하는것.

(라) 判読하기 어려운 文字나 数字를 고쳐 쓰는것.

(마) 調査票記入法을 統一하는것.

(바) 計算錯誤를 発見하여 訂正하는것.

(사) 調査票記入錯誤를 発見하는것.

(아) 調査員의 不正記入을 発見하는것.

(자) 計算値를 記入하는것.

以上의 것中 (사)의 調査票의 記入錯誤는 調査員 또는 被調査者가 不注意에 基因하는 内容上의 錯誤이므로 이를 調査表審査 過程에서 発見하기는 어려운 것이다. 이는 調査票의 構成이나 調査員의 選択 또는 指導에서 充分히 注意를 하여 미리 予防토록 하여야 하나 調査票審査 過程에 있어서도 이러한 錯誤를 発見하도록 最大의 努力을 기울이지 않으면 안되는 것이다. 이를 위하여서는 다음과 같은 方法을 使用할 수가 있다.

첫째로 個個의 調査票中에서 関聯되는 다른 項目과 対照하여 그 間에 모순이 発見되면 錯誤가 存在하는 可能性이 있다.

둘째로 調査員마다 簡単한 集計를 하여 다른 調査員의 그것과 比較하여 현저한 差異가 있는 경우에는 어느 調査員인가 不注意를 犯했을지도 모른다.

세째는 全体에 대하여서 簡単한 集計를 하여 予想과 동떨어진 結果가 나올 경우에는 일단 그 項目의 内容을 疑心할 수 있을 것이다.

調査票審査에는 実査管理者에 의한 現地審査와 別途審査者에 의한 事後審査의 段階가 있으며, 後者는 다시 調査規模에 따라 地方審査와 中央審査의 段階로 나누어진다. 다만, 郵便調査法을 使用할 때에는 審査者에 의한 審査만이 1段階 行하여 질 뿐이다.

実査管理者에 依한 審査는 대개 実査期間中 調査員이 記入 完了한 調査票를 提出하였을 때 調査員의 面前에서 한다. 여기서는 記入漏落, 記入不完全, 記入의 錯誤, 調査員의 不正等 発見에 注力하고 発見된 경우에는 再調査를 命한다. 同時에 보기 힘든 글자, 記入法의 不統一에 注意한다. 現地審査는 実査期間中 毎日하는 것이 좋으며 적어도 몇번은 하여야 한다. 그리고 調査員이 割当된 実査를 모두 끝마쳤을 때는 調査漏落이나 重複없이 割当対象의 調査가 完了되었는가를 確認하여야 한다.

実査가 모두 完了되고 調査票가 一定한 場所에 収集되어 審査員에 의하여서 一斉히 実施되는 調査票審査에 있어서는 簡単한 集計에 의하여서 記入上의 錯誤나 調査員의 不正을 発見하여 訂正 或은 調査票의 破棄를 하고, 또 計算을 検算하며 計算値를 記入하고 경우에 따라서는 空白欄에 回答을 統計的으로 推定하여 메꾸기도 한다.

以上과 같이 調査票審査는 実査過程에서 発生하는 各種誤差를 最少限으로 줄일 수 있는 最終的인 機会이므로 이를 徹底히 実施하지 않으면 안되며 審査를 担当하는 사람은 調査業務에 経験이 있고 이에 精通한 사람이어야 하는 것이다.

(2) 符号化 (Coding)

調査票審査에 뒤따르는 作業은 符号化作業이다. 이것은 調査票의 各調査項目에 대하여 記入된 回答을 몇個의 구룹으로 分類하고 그 구룹에 対応한 一定한 符号를 各調査票에 부쳐서 分類集計에 便利하게 하는 作業이다. 따라서 符号化에 앞서서 分類基準이 定하여져 있지 않으면 안되는데 이 分類基準은 全調過程에서 미리 決定되는 것이 普通이지만 調査項目에 따라서는 調査結果를 보지 않고는 適当한 分類基準을 定할 수 없는 것도 있으므로 이러한 경우에는 審査가 끝난 調査票를 하나 하나 検討하여 事後에 適当한 分類基準을 定하여야 하는 것이다.

그리고 分類基準이 定하여지면 이에 対応한 符号를 決定하여야 한다. 그런데 符号는 機械集計를 하는 경우와 手集計를 하는 경우에 따라 若干 相異하다. 機械集計에 있어서의 符号는 1.2.3……9,0,X,Y의 10数字 2文字의 範囲에서 選択하지 않으면 안된다. 이때 1,2,3………으로 적은 数字를 順次로 使用하고 9를 「其他」X를 「不明」Y를 「非該当」하는 式으로 곧数字와 文字를 特定한 範疇에 該当시키는 것이 普通이다.

手集計에 있어서는 記入이 簡便하고 判読하기 쉬운 것이면 어떤 数字나 文字를 使用하여도 좋으나 一見하여 그 範疇의 内容이 認知되는 것이면 더욱 좋다. 例컨데 性別에 있어서 男性을 m, 女性을 f로 表示하는 것이다. 또 集計에 있어서 가까이 있는 項目과 錯誤하지 않도록 数字 가,나,다 大文字 小文字의 「알파벳트」를 섞어서 使用하는 것도 좋은 方法이다.

符号가 決定되면 具体的으로 어디에 무슨 符号를 부치는 가를 細密하게 指示한 符号表을 作成하여 符号記入者에게 나누어 주어

審査가 끝난 調査票에 每장마다 符号를 부치도록 한다. 符号記入者는 符号集에 따라 各項目의 符号欄에 色鉛筆(色은 統一하여야 한다)로 符号를 記入한다. 이 作業이 끝나면 다른 符号記入者가 바꾸어서 符号가 제대로 記入되어 있는가 検査하지 않으면 안된다.

符号化作業은 項目에 따라서는 매우 쉬운것도 있으나 때로는 매우 어려운 것도 있어 아무리 符節集이 細密하게 되어 있다고 하더라도 具体的인 경우에 어떤 符号를 부쳐야 할지 困難할 때가 많다. 따라서 符号記入者는 相当한 知識과 熟練을 必要로 하며 注意깊은 作業을 하여야 하는 것이다. 기껏 애써서 收集된 資料라도 符号化를 잘못하면 그 結果는 쓸모없는 것이 되고 마는 것이다.

그런데 以上은 모두 事後符号化의 경우이나 分類基準과 符号가 調査企劃 過程에서 決定되고 그 基準에 따라 分類(구룹化)하기가 매우 쉬운 경우에는 符号記入을 調査員에게 시키는 경우가 많은데 이 경우에는 実査를 始作하기 前에 調査員에 대한 徹底한 訓練이 있어야 하며 또 事後에 検査가 따라야 할 것이다.

(3) 集　計

調査票審査와 符号化作業이 끝나면 모든 調査票를 符号에 따라 分類하여 集計를 하여야 한다. 集計란 調査票의 各項目에 記入된 内容을 計算하는 作業을 말하는 것이다. 集計에는 다음과 같은 種類가 있다.

(가) 単純集計와 相関集計:

이것은 만들어 지는 統計表의 種類에 의한 分類인데 単純集計란 一般的인 度数分布表를 만드는 集計를 말하며, 相関集計는 2個以上의 標識에 関하여 度数를 나타낸 相関表를 만드는 集計를 말한다.

(나) 中央集計와 地方集計 :

中央集計란 調査票를 中央에 送付케하여 中央에서 一括集計하는 것을 말하며 地方集計는 調査票를 中央에 送付하지 않고 地方調査機關에서 集計를 하여 作成된 統計表만을 中央에 送付되어 中央에서 各地로 부터 收集된 統計表를 取合하여 最終的인 것을 만드는 方法이다. 地方集計의 唯一한 利点은 集計가 迅速하게 끝난다는 데 있으나 機械集計가 発達된 現在에는 地方集計를 하여야 할 理由가 없으며 또 地方集計는 正確性을 欠하기 쉬운 致命的인 欠陷이 있는 것이다.

(다) 手集計와 機械集計 :

手集計란 統計機械를 使用하지 않고 사람의 손으로써 集計하는 方法이며 機械集計란 一群의 統計機械를 使用하여 集計하는 方法을 말한다. 手集計에는 다시 여러가지 方法이 있는 데 大体로 調査票를 그대로 集計作業에 使用하는 方法과 調査票의 內容을 다른 集計用 카-드에 転記하여 集計하는 方法으로 나눌 수 있다. 機械集計는 電子計算組織에 의한 集計인데 調査票 한장 (또는 一部) 마다 펀치카아드 (Punch card)를 만들어 이를 펀치에 의하여 分類하고 다시 計算하는 것이다. 手集計와 다른 点은 펀치. 카아드의 穿孔부터 始作하여 카아드의 分類, 카아드의 計算에 걸쳐 一切가 機械에 의하여 이루어지며 따라서 操作이 매우 迅速 正確하게 行하여 진다는 것이다.

機械集計와 手集計와의 優劣은 한마디로 말할 수 없으나 手集計는 1) 集計가 単純할때, 2) 調査対象이 적을 때, 3) 分類가 極히 細分되어 分類된 各 구룹의 調査票 枚数가 적을 때 等에 有利하고 機械集計는 1) 調査対象이 많을때, 2) 集計가 比較的 複雜할 때 3) 網羅的인 集計를 할 때 等에 有利하다.

라. 統計表의 作成과 分析

(1) 統計表의 種類

調査票의 分類 集計가 끝나면 마지막으로 統計表를 作成하여야 한다. 統計表는 一連의 統計調査 業務의 最終 生産物로서 統計의 作成과 利用의 媒介物이 되는 것이므로 統計資料의 觀察 比較 解釋이 容易하도록 調査에서 蒐集된 資料(統計數字)를 体系있게 分類하고 簡潔하게 行과 列(欄)로 配列, 整理하여서 다른 說明을 듣지 않더라도 그 表만 보고 調査結果를 把握할 수 있도록 그 構成을 잘 考慮하여야 한다.

統計表의 構成(樣式)은 統計表의 優劣을 左右하기도 하지만 또한 集計, 製表의 全過程을 規制하게 되는 것이므로 企劃過程에서 가장 먼저 考慮되고 決定되어야 하는 것이다. 그런데 統計表는 그 表示方法, 形式內容(性質) 統計數字의 加工 與否 等에 따라 다음과 같이 分類할 수가 있다.

(가) 文中統計表와 正式統計表:

이것은 統計를 表示하는 方法에 의한 分類인데 文中統計表는 文章中에 表示한 統計表로서 表題가 없고 文章과 直接 關係되어 있어 文章을 읽지 않으면 統計의 意味를 알 수 없는 統計表이다. 即 獨立性이 없는 統計表이다. 이에 대해서 正式統計表는 表頭, 表題, 表側等 表의 体制를 完備하여 다른 說明을 듣지 않더라도 그 表만 보고 內容을 알 수 있고, 여러가지 分析이 可能하도록 作成된 統計表이다. 普通 統計表라고 할 때는 이 正式統計表를 말하는 것이다.

(나) 單純分類表와 相關表:

이것은 形式에 의한 分類로서 단 한가지의 分類에 依하여

作成된 統計表를 單純分類表라 하며 數種의 分類를 結合하여 만든 統計表를 相關表라고 한다.

(나) 構造統計表와 系列統計表 :

이것은 統計表의 內容에 의한 分類로서 構造統計表는 時間과 場所를 一定하게 하고 其他標識 (例 産業分類, 規模, 性別등) 에 의하여서 分類하여 놓은 統計인 데 이를 보통 度數分布表라고도 한다. 例를 들면 業種別, 規模別, 生産額 또는 性別, 職業別, 從業員數등 統計表가 이에 屬한다. 이에 대하여 系列統計表는 同種의 統計數字를 時間的 또는 場所的으로 配列한 統計表로서 例컨데 年度別 生産額 또는 道別 從業數등 統計表가 이에 屬한다. 그리고 系列統計表中 統計數字를 時間的으로 配列한 統計表를 時系列表라고 하며 統計數字를 場所的으로 配列한 統計表를 場所的 (地域的) 分布라고 한다. 이 中 특히 時系列表는 어떤 現象의 時間的인 變遷을 表示하기 위한 것으로서 生産指數, 物價指數등과 같이 指數化하여 表示되는 경우가 많다.

(다) 非加工統計表와 加工統計表 :

이것은 統計數字의 加工與否에 의한 分類로 非加工統計表는 統計調査 結果 蒐集된 資料의 總和値 (合計値) 를 아무런 加工없이 그대로 表示한 統計이며 加工統計表는 非加工統計表에 나타난 統計數字 (總和値) 를 各種 比率, 平均등으로 加工하여 表示한 統計表이다. 普通 統計調査 過程은 非加工統計表를 作成 公表하는 것으로 一段落되지만 利用者를 위하여서는 보다 一目瞭然하게 알 수 있는 加工統計表를 作成하여 表示하여 주는 것이 親切하다. 例를 들면 各種指數나 國民所得등이 加工統計表에 屬한다.

(2) 統計表의 構成

統計表 (正式統計表) 는 그 本質的 要素가 되는 統計數字를 除外하면 다음과 같은 8個의 部分으로 構成된다.

(가) 表題 (heading)

(나) 頭註 (headnote)

(다) 表頭 (box head)

(라) 表側 (stub)

(마) 欄 (column)

(바) 行 (line)

(사) 表体 (field or body)

(아) 脚註 (foodnote)

正式統計表의 構造

表 題 (頭 註)

	表		頭	
表 側	— — — —	— 表 — — —	— 体 — — —	— — — —

脚 註

1 表題는 다시 表番号와 表名으로 成立된다. 表番号는 統計表 表의 番号인데 이것은 한 統計表의 系列中에 있어서의 그 統計表의 關係位置를 表示하는 것이며, 또 統計表의 索引도 되는 것이다.

따라서 單独의 統計表에는 表番号를 붙일 必要가 없다. 表名은 表題의 中心을 이루는 것으로서 統計表 內容의 目錄이다. 表名에는 統計의 表示範囲, 分類事項, 地域的範囲, 基準時点등이 表示되어야 하며 이 네 個의 事項이 結合되지 않으면 完全한 目錄의 役割을 할 수 없는 것이다.

2 頭註는 表題와 統計表의 最上位線 사이에 表示되는 注記事項이나 頭註는 表名을 補充하고 統計表 全体를 理解하는데 必要한 事項으로서 表名과 密接한 關係를 갖는 것이다. 따라서 보통 頭註로서 表示되는 것은 그 統計表의 數字 全体에 関한 單位라든가 統計의 基本時点 또는 集計対象範囲와 같은 것이다. 頭註가 統計表 全体에 関한 注記事項인 데 대하여 (8)脚註는 表中의 特定한 統計数字 또는 表頭, 表側의 意味를 明白히 하거나 資料의 出処를 밝히는 데 使用되는 것이다.

脚註는 統計表의 最下位線 아래에 記入한다. 그리고 頭註를 記入할 때는 「註」「備考」등을 表示할 必要가 없지만 脚註는 반드시 「註」라고 表示한 다음에 必要事項을 記載하여야 하며 또한 說明을 必要로 하는 數字의 앞(또는 뒤)에는 合符号를 무쳐서 註와 連結시켜야 한다.

3 表頭와 表側은 形式的으로 統計表의 모양을 만들고 実質的으로는 統計系列을 만들어 數字間의 關係를 明白하게 하는 것이다. 表頭는 統計表의 제일 위쪽에 位置하여 各 分類事項을 表

示하며 表側은 統計表의 左側에 位置하여 이 亦是 分類事項을 表示한다. 따라서 表頭와 表側에 있어서의 問題는 어떤 分類事項을 表頭와 表側의 어느 쪽에 表示하는 것이 数字間의 関係를 보다 明確하게 나타내는가 그리고 紙面을 浪費하지 않고 合理的으로 利用할 수 있는가 하는 것이다.

4 表体는 統計表中 統計数字가 記入되는 場所를 総称하는 것으로서 이것은 다시 欄과 行으로 成立된다. 欄은 統計数字가 縦으로 記列된 것이며 行은 統計数字가 横으로 記列된 것을 말한다.

(3) 統計表 作成上의 注意

統計表는 社会的으로 널리 利用됨으로서 비로소 그 価値가 있는 것이므로 먼저 利用者에게 쉽고 便利하도록 作成되어야 한다. 即 統計表는 統計利用者가 그 統計表를 보고 理解할 수 있거나 誤解 또는 錯覚을 일으키도록 複雑하게 만들어져서는 안되며 되도록 簡潔하고 明確하게 作成되어야 한다. 그리고 보다 利用者에게 親切하기 위하여서는 比率, 指数, 平均등 加工統計를 아울러 表示하여 주거나 統計図表化하여 나타내주는 것도 좋다.

統計表作成에 있어서 또 한가지 注意하여야 할 事項은 統計表에는 어떤 個人의 秘密이 露出되어서는 안된다는 것이다. 따라서 分類가 細分됨으로써 어떤 個体에 関한 事実이 露出될 염려가 있는 것은 技術的으로 適切히 処理하여 個人의 秘密이 나타나지 않도록 하여야 한다. 그리고 特히 誤解하기 쉬운 事項이 있거나 例外的인 事項이 있는 경우에는 반드시 그 内容을 註書하여야 한다.

統計表의 一般的인 記載方法은 다음과 같다.

(가) 該当事項이 없을 때에는 「 − 」로 表示한다.

(나) 未詳인 때는 「 … 」로 表示한다.

外 1 1

(다) 單位未滿인 때 即 表示單位가 千원인 데 集計된 金額이 100원 또는 300원이 되었을 때는 「0」으로 表示한다.

끝으로 調査結果를 發表할 때에는 調査目的, 範圍, 調査單位, 調査項目의 定義, 調査方法, 標本調査일 때는 標本數, 標本抽出方法등 統計表를 利用하는 사람들이 參考하여야 할 事項에 대한 說明書를 統計表의 앞에 붙여야 한다.

外 1

(4) 統計表의 分析

統計表가 作成되면 마지막으로 統計表의 分析이 남는다. 統計表의 分析은 우선 統計表를 알기쉽게 꾸미고 여기에 說明을 加하여 어떤 結論을 끌어내는 것이다. 먼저 統計表를 알기 쉽게 꾸미는 데는 統計表의 數字를 加工하여 百分率이나 比率로 나타내는 方法이 있다. 그러나 百分率이나 比率은 아무 數字에 대하여서나 마구 算出하여서는 안된다.

統計表를 分析할 때에는 다음 事項을 檢討하여야 한다.

(가) 가장 重要한 點은 무엇인가

(나) 그 밖에 무엇이 表示되어 있는가

(다) 平均은 豫期한 것, 또는 다른 資料에 比하여 어떠한가

(라) 各 範疇의 最大數나 最小數는 무엇을 意味하는가 다른 것과 比較하여 그 百分率은 무엇을 말하는가

(마) 一般的인 傾向이 있는가 없는가 그것은 어째서 그러한가 例外的인 것은 무엇을 意味하는가

(바) 因果的인 關係가 나타나는가

(사) 이미 알려져있는 것이나 다른 統計表에 나타나 있는 것과 比較하여 一貫性이 있는가

(아) 이러한 結果가 나온 것은 標本抽出의 方法이나 調査方法

때문이 아닌가

(자) 各部分의 数字를 다시 検討하기 위하여 調査票를 찾아 보아야 하거나 再集計를 하는 편이 낫지는 않을까

(차) 最終的인 結論을 내릴때 統計的 処理를 加할 必要는 없는가

以上과 같이 検討하면서 分析하여 가면 重要한 類似性이나 差異, 連続関係, 因果関係가 発見되어 数字의 限界도 明白하게 될 것이다.

그리고 이와같은 検討와 分析에 이어서 다시 全体的으로 検討하고 分析하지 않으면 안되는 데 이 때 重要한 것은 이러한 分析이 先入見에 이끌리어 不当하게 強調되거나 軽視되는 일이 있어서는 안된다. 一般的인 傾向과 例外的인 경우와를 明確히 区別하고 이에 올바른 評価를 加하여야 하는 것이다. 因果関係를 推論하는 경우에도 単純히 数字上에 相関関係가 있는 것만으로 立論하여서는 안되며 넓은 理論的인 考察과 緻密한 比較分析이 推論의 前提가 되어야만 한다.

4. 統計調査의 実例 (経済活動人口調査)

経済活動人口調査는 1957年부터 1962年 5月까지 地方行政機関을 通하여 労動力調査를 実施하여 毎月 就業者와 失業者의 資料를 蒐集하였다.

그後 経済開発 5個年計劃의 樹立과 그 遂行을 위하여 経済活動에 参与하는 人口의 正確한 資料가 切実히 要請되어 経済企劃院 調査統計局에서 過去에 実施한 労動力調査의 諸欠点을 시정하고 보다 正確한 資料를 作成할 目的으로 統計法 第2条에 依拠하여 経済活動人口調査를 指定統計 第4号로 定하고 1960年 人口센서스에 根拠를 둔

새로운 標本設計와 專門化된 調査員에 依하여 1962年 8月부터 實施하여 왔다.

그러나 時間이 흐름에 따라 調査区内의 特性値가 変하여 1969年 6月부터는 66年 人口센서스의 調査区에 根據를 둔 標本에 의하여 再設計되었고 다시 1970年 人口 및 住宅센서스가 実施됨에 따라 1972年 3月부터는 이에 의한 새로운 調査地域에서 実施하게 되었다.

다음에 経済活動人口調査의 概要, 調査節次, 調査票作成, 調査結果에 対하여 前述한 調査方法에 依據 紹介함으로써 理解를 돕고자 한다.

가. 調査概要

(1) 調査目的

經済, 人口 및 諸 社会条件의 変動에 따르는 国民의 経済活動의 変化를 適期에 正確히 把握하여 雇傭 및 失業의 構造와 変動趨移를 分析하고 이에 대한 政策을 樹立하는 데 基礎資料를 提供하는데 그 目的이 있다.

(2) 調査範囲

調査期間을 基準으로 하여 大韓民国 国籍을 가진 常住人口의 1/300에 該当하는 人口中 満14歳 以上者는 모두 調査対象者로 하나 다음 事項에 該当되는 者는 除外한다.

(가) 現役軍人

(나) 刑이 確定된 교도소 수감자

(다) 外国人

常住者라 함은 調査期間을 基準으로 하여 同一한 場所에서 3個月 以上 居住하였거나 居住하려는 期間이 3個月에 達하는 者를

말한다. 但, 航空, 船舶의 船員은 長期 出他中일지라도 調査한다.

(3) 調査項目

本 調査의 調査票는 18個 調査項目으로 되어 있는 데 姓名등 6個項目의 人的事項과 1週間의 主된 活動状態등 12個項目의 経済活動事項을 調査한다.

(4) 調査方法

(가) 標本調査

1970年 人口 및 住宅센서스 調査区中 普通調査区에서 $\frac{1}{500}$ 에 該当되는 市部 72個 調査区와 郡部 93個 調査区등 総 165個 調査区를 抽出하여 同 調査区内의 全家口 約12,000標本家口를 調査하는 標本調査이며, 標本抽出에서 除外된 調査区는 落島, 閑地調査区, 寄宿舎調査区, 特殊社会施設調査区, 特別調査区등이며 除外된 調査区의 센서스 人口는 約 75万名에 不過하다.

(나) 週間 및 家口単位調査

経済活動人口調査는 1週間의 調査期間中에 일어나는 経済活動状態를 把握하는 方法으로 家口를 標本単位로 調査하고 있다.
調査期間을 1週間으로 定한 理由는 1週間이란 質問時 応答者의 記憶誤差를 적게 하여 正確한 応答을 바랄 수 있고 経常状態를 反映하는 데 適当한 時期이며 公休日, 日曜日로 말미암아 発生되는 変動을 充分히 피할 수 있기 때문이다.

(다) 面接調査方法과 配票調査方法의 併用

経済活動人口調査票는 対象家口를 訪問하여 応答者로 부터 얻어지는 情報를 調査員이 記入하는 他計式方法을 使用하며 就業時間 記入票는 応答者 自身이 記入하는 自計式方法을 쓰고 있다.

(5) 調査時期

季節別로 1年에 4번 實施하며 調査対象期間(調査週間)은 指定된 달(3月, 6月, 9月, 12月)의 指定된 1週間으로 하여 実際 調査期間(実査期間)은 調査週間 다음의 1週間으로 한다.

1972年의 調査時期는 다음과 같이 計劃하고 있다.

	準備調査	調査週間	実地調査
1/4 分期 (3月)	5日 ~ 11日	12日 ~ 18日	19日 ~ 25日
2/4 分期 (6月)	4日 ~ 10日	11日 ~ 17日	18日 ~ 24日
3/4 分期 (9月)	3日 ~ 9日	10日 ~ 16日	17日 ~ 23日
4/4 分期 (12月)	3日 ~ 9日	10日 ~ 16日	17日 ~ 23日

(6) 主要用語定義

(가) 家 口

居住와 家計를 같이 하는 사람의 모임을 家口라 하며 한사람이라도 別途로 居住하고 独立的인 家計를 이룩하고 있는 경우에는 하나의 家口로 간주한다.

1 家庭婦, 其他使用人, 店員, 同居人등은 主家口의 家口員으로 包含시킨다.

2 셋방인, 同居人으로서 主家口와는 家計를 別途로 하는 경우와 食事는 같이 하고 있지마는 主家口에게 房貰라든가 食費를 支払하여 家計를 別途로 하는 사람은 主家口의 家口員에 包含시키

지 않고 別個의 家口로 한다. 그러나 친척이라든가 主人의 子弟를 돌보아 주고 實費程度의 食費, 房貰를 받고 있는 同居人은 主家口의 家口員으로 한다.

3 学校나 工場등의 寄宿舍등에서 살고 있는 独身者는 各各 単一 家口로 하였으나 이 때 全員이 한 家族처럼 살고 있으면 그 家口一員을 하나의 家口로 한다.

(나) 農 家

農家라 함은 農業을 生業(生計와 営利)으로 하는 家口로서 다음 各 号의 1에 該当하는 家口를 말한다.

1 所有如何를 不問하고 논, 밭 水源地등의 総面積이 300坪 以上을 直接 耕作하는 경우

2 大家畜(소 또는 말)을 1마리 以上 飼育하는 경우 (但, 運搬用은 除外)

3 中家畜(돼지, 염소, 양등)을 도합 3마리 以上 飼育하는 경우

4 小家畜(토끼등)을 40마리 또는 닭 30마리 以上을 飼育하는 경우

5 高等園芸, 菜蔬, 特用作物등 100坪 또는 果樹苗圃등을 200坪 以上 耕作하는 경우

6 5郡 以上의 養蜂을 하는 경우

7 蚕種 10g 以上을 소잠하는 경우

農業과 他産業을 함께 経営하는 경우에는 農業收入이 50% 以上인 경우에만 農家에 包含시킨다. (但, 農業收入面으로 区別하기 困難할 때에는 労動力 投入量으로 区別한다.)

(바) 経済活動人口

財貨나 用役을 生産하기 위하여 労動을 提供하는 모든 사람과 提供할 能力이 있는 사람을 말한다.

여기에는 다음과 같은 種類로 分類된다.

- 14歳以上 人口
  - 経済活動 人口
    - 就業者
      - 正常的就業者
      - 非正常的就業者
        - 家事
        - 通学（就業한者)
        - 其他
      - 日時休職者
    - 失業者
      - 求職者
      - 非求職者
        - 日気不順
        - 一時的病
        - 求職断念
        - 待期
  - 非経済活動人口
    - 家事
    - 通学
    - 年老者
    - 不具者
    - 不労所得者
    - 其他

(7) 調査結果

経済活動人口調査의 結果表는 다음 9個表로 되어 있으며 経済活動人口年報에 公表한다.

(가) 経済活動人口年報

(나) 年令階級別 経済活動人口

(다) 非経済活動人口

(라) 年令階級別 就業者

(마) 産業 및 年令階級別 就業者

(바) 産業別 就業者

(사) 職業別 就業者

(아) 従事上의 地位別 就業者

(자) 産業 및 就業時間別 就業者

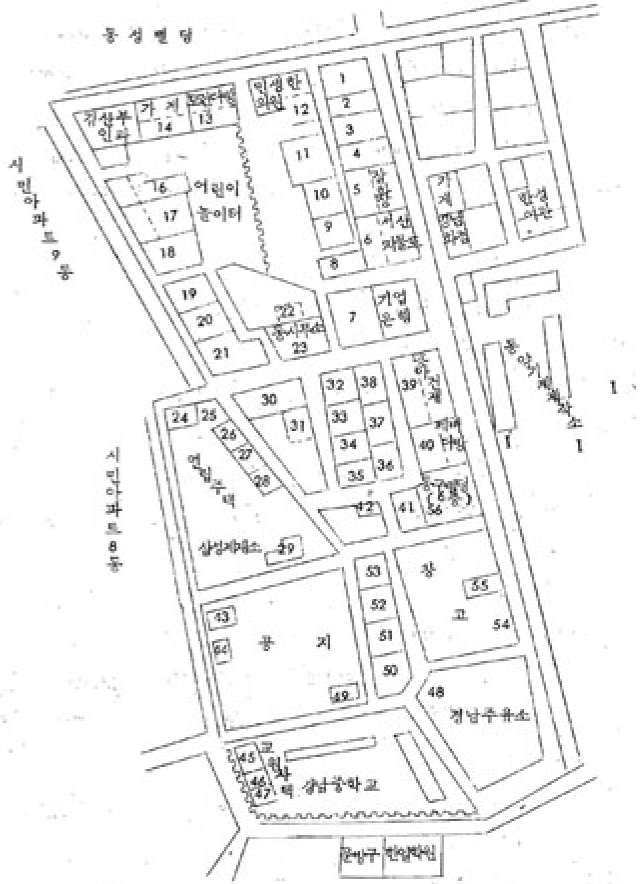
나. 調査節次

(1) 準備調査期間

準備調査期間에는 먼저 標本으로 뽑힌 調査区의 要図를 作成하고 要図에 나타난 居処에 居住하는 家口의 家口員名簿를 作成한다. 調査区要図와 家口員名簿는 本 調査의 基本이 되는 것으로서 要図는 정성을 다하여 깨끗하고 正確하게 作成하여야 하며 名簿는 每 家口마다 빠짐없이 작성하여야 하므로 사람이 살 수 있는 곳은 全部 確認한다.

調査区要図의 作成要領은 먼저 地形, 地物의 表示를 補充記入하면서 境界線을 確認 作成하고, 調査区 境界線과 境界線안쪽에 있는 道路와 建物(또는 居処)을 빠짐없이 □로 그려 넣고 主要建物의 이름이나 地名을 써넣어 位置를 明確히 한 다음 居処의 一連番号를 □안에 調査하기 便利한 順序로 記入한다.

要圖作成例



調査区要図와 家口員名簿가 作成되면 調査員은 調査区를 수시로 순회하며 新築建物(居処)과 建物의 撤去등 地形 또는 建物의 変動을 把握하여 要図를 補完, 修正하여야 하며, 또한 家口員名簿作成後 毎 訪問時마다 家口 및 家口員의 変動事項을 把握하여 要図및 家口員 変動報告書에 記入하여 다음 実査에 漏落이나 重複이 없도록 事前 準備를 하여야 한다.

또한 準備調査期間에 就業時間 記入表를 配付하여 被調査者가 経済活動状況을 事前 記入토록 하여 調査上 便宜와 正確을 期할 수 있도록 한다.

(2) 調査週間

調査週間은 調査対象이 되는 1週間으로서 準備調査期間에 配付한 就業時間記入表를 正確하게 作成하고 있는가의 与否를 確認하고, 만약 記入上의 애로가 있을 경우 이를 指導하여 正確하게 作成되도록 한다.

또한 就業時間記入 確認中 調査区要図와 家口員의 変動이 있을 경우 계속하여 補完하고 이를 変動報告書에 記入하여 中央에 報告토록 한다.

(3) 実査期間

実査期間은 調査週間 다음 1週間으로서 各 家口의 満14歳以上 人口인 各 사람과 直接面談하여 調査票를 記入하되 準備調査時 配付하여 被調査者가 作成한 就業時間記入票를 比較하여 作成토록 한다.

実査期間에도 調査区의 要図와 家口員名簿를 補完하여 調査의 漏落을 방지하고, 次期調査를 便利하게 한다.

調査票의 項目別 作成要領은 다음에 説明토록 한다.

(4) 整理期間

整理期間은 実査가 完了된 後 調査의 正確性을 把握하는 期間으로서 우선 変動報告書의 記入内容을 確認하여 調査의 重複이나 脫落을 検討하고, 各 調査票의 内容을 検査하여 調査票作成이 잘 되었는가 살핀다음 各 調査区別 綜合表를 作成토록 한다.

다. 調査票作成

(1) 調査員의 態度

(가) 調査員의 身分을 確認할 수 있는 身分証을 반드시 提示한다.

(나) 各 家口의 滿 14歲以上 人口인 各 사람과 直接 面談하여 調査하는 것이 原則이다.

(다) 親切하게 自己紹介를 한 後에 経済活動人口調査의 目的을 說明하고 正確한 답변을 하여 줄 것을 부탁한다.

(라) 調査員이 家口를 訪問하였으나 사람이 없어서 調査를 할 수가 없을 경우에는 家口員名簿에 表示를 하였다가 再訪問을 하도록 한다.

(마) 調査票記入을 할 때에는 같은 事項 例를 들면 職業이나 産業이 같은 경우에 「上同」 또는 「〃」으로 記入하지 말고 反複하여 記入하되 該当事項이 없는 欄에는 「＼」을 긋는다.

(2) 調査票의 記入要領

経済活動人口調査는 経済活動人口를 対象으로 하고 있으므로 家口員中 14歲 以上인 者에 対하여서만 調査토록 한다.

調査項目은 이름등 6種의 人的事項과 活動状態등 12種의 経済活動事項이며 第7欄의 「지난 1週日間에 主로 무엇을 하셨읍니까?」

에서는 지난 1週間에 主로 한 行為란 平常時에 했던 活動과는 関係없이 指定된 週間사이에 活動한 行為를 말하는데 9個의 項目中 「1. 일하였음」에 「○」表한 사람은 8欄, 9欄, 10欄은 「＼」을 긋고 11欄부터 質問하며 其外項目의 경우 8～10欄을 作成토록한다

다음에 調査票의 項目을 묻는 方法의 차례를 図表로 나타냈으며 経済活動人口調査票 作成例를 添附하여 参考토록 하였다.

라. 調査方法決定의 理論的 基礎

労動力調査方法은 資料蒐集上의 接近法과 概念上의 接近法이 있다.

(1) 資料蒐集上의 接近法

資料蒐集上의 接近法에는 事業体를 調査対象으로 그 事業体에 雇傭되어 있는 被雇傭者와 그와 関聯된 資料를 蒐集하는 事業体接近法과 家口를 調査対象으로 하여 就業者, 失業者 및 其他 人口学的 特性値를 家口内의 家口員과 関聯하여 資料를 蒐集하는 家口接近法이 있다.

事業体接近法은 被雇傭人의 数를 비롯하여 給与, 賃金, 就業할 計劃時間 및 日数, 実際 就業한 時間등 有用한 資料를 同時에 얻을 수 있고, 얻어진 資料는 雇傭指数, 賃金(日, 週, 月別)指数, 産業間의 移職率을 얻을 수 있으며 産業, 職業, 性別, 就業者의 構成은 調査期間中 취업모형의 変化를 測定할 수 있는 長点이 있으나 調査期間中 한 事業体 以上에 일한 者는 모두 그곳에서 把握되므로 過大로 調査되기 쉬우며 오직 被雇傭者에 限하여 調査되므로 調査의 範囲가 限定되어 있어 雇傭主, 自営業主, 家族従事者, 失業者, 非経済活動人口 등은 把握할 수 없는 短点이 있다.

家口接近法은 就業者, 失業者, 非経済活動人口의 特性을 모두 把握할 수 있으며 이들 特性値와 結付하여 家口의 社会的, 経済的, 人口学的 背景을 同時에 研究할 수 있는 長点이 있으며 正確한 給与額이나 賃金率을 算出할 수 없으며 生産量이나 生産性을 把握할 수 없는 短点이 있다.

(2) 概念上의 接近法

調査의 概念上 接近法에는 平常状態에서 주어진 役割이나 機能에 의하여 労動力人口를 調査를 하는 有業者接近法과 어떤 期間内에 実際로 活動한 状態에 의하여 労動力人口를 調査하는 労動力接近法이 있다.

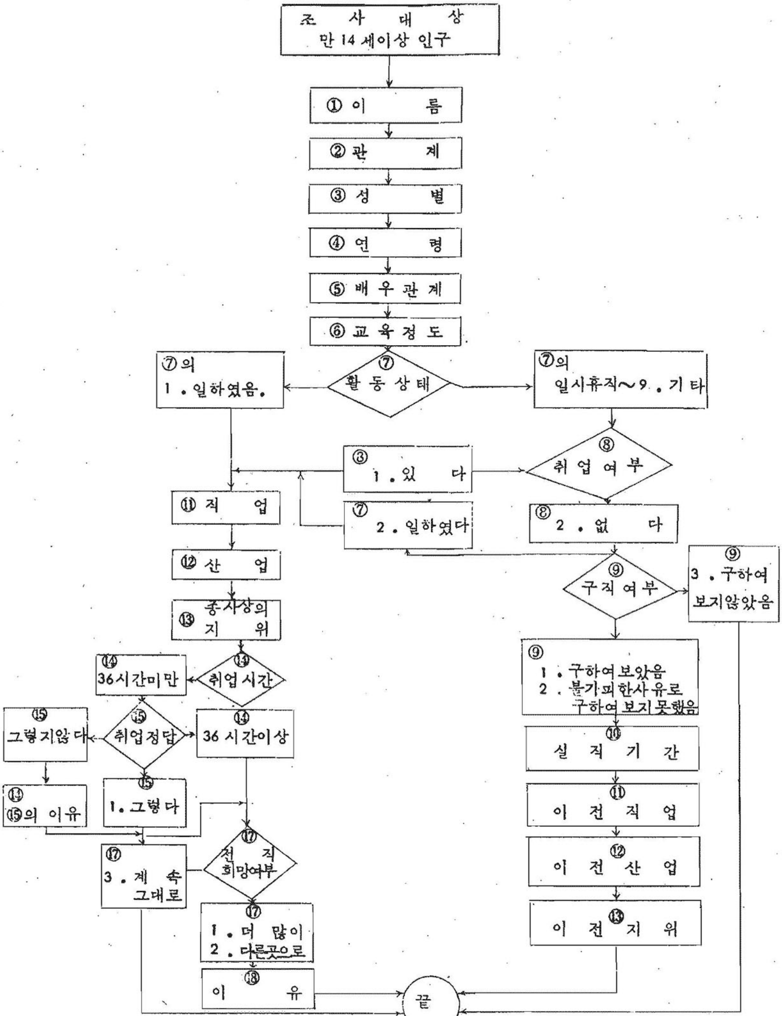
有業者接近法은 職業이 있느냐 없느냐로 就業者와 그 밖에 것으로 区分하기 쉽고 調査員訓練이 容易하며 平常状態에 의하여 얻어진 資料이기 때문에 季節的이나 우발적인 活動에 별로 影響을 입지 않는 長点이 있고, 한가지 職業이나 確固不同한 職業이 없는 者들을 定義하기 어려운 同時에 얻어진 結果는 特定한 期間이 없기 때문에 Bench Mark로 使用하기에 不適当하고 처음 求職活動을 한 者를 失業者로 포착하기 힘들며 또한 経営的 経済動向을 포착할 수 없는 短点이 있다.

労動力接近方法은 짧은 期間동안 일어난 活動을 묻기 때문에 記憶誤謬를 적게하고 Bench Mark로 提供될 수 있으며 経常状態는 平常状態보다 客観的이고 正確性을 期할 수 있고, 失業者, 非経済活動人口를 쉽게 区別할 수 있을 뿐아니라 経常的, 経済動向을 빨리 포착할 수 있는 長点이 있고 反面에 짧은 期間에 일어난 活動状態를 調査하므로 그 期間中에 우발사건이나 気候変動이 일어났을때 그 影響이 크게 미치게 되어 継続的으로 調査하여야 한다는 短点을 가지고 있다.

(3) 以上의 두 方法에 대하여 UN에서도 勧告하고 있으나 어느 方法을 択하여야 할지는 調査의 性格이나 그 나라의 経済的, 社会的 与件에 의하여 左右된다.

우리나라의 経済活動人口調査는 就業者, 失業者, 非経済活動人口의 構成的 特性을 모두 把握할 수 있는 家口接近法과 客観的이고 正確性을 期할 수 있는 労動力接近方法에 依拠 実施되고 있다.

조사표 작성순서


조 사 대 상
만 14 세이상 인구
① 이 름
② 관 계
③ 성 별
④ 연 령
⑤ 배 우 관 계
⑥ 교 육 정 도
⑦ 활 동 상 태
⑦의 1. 일하였음.
⑦의 일시휴직～9. 기 타
⑧ 취 업 여 부
③ 1. 있 다
⑦ 2. 일하였다
⑧ 2. 없 다
⑨ 구 직 여 부
⑨ 3. 구하여 보지않았음
⑨ 1. 구하여 보았음 2. 불가피한사유로 구하여 보지 못했음
⑩ 실 직 기 간
⑪ 이 전 직 업
⑫ 이 전 산 업
⑬ 이 전 지 위
⑪ 직 업
⑫ 산 업
⑬ 종사상의 지 위
⑭ 취업 시간
⑭ 36시간미만
⑭ 36 시간이상
⑮ 취업정답
⑮ 그렇지않다
⑮ 1. 그렇다
⑭ ⑮의 이유
⑰ 전 직 희망여부
⑰ 3. 계 속 그대로
⑰ 1. 더 많이 2. 다른곳으로
⑱ 이 유
끝