(一)(般)(論)(叢)

시멘트 經營과 統計的方法(1)

郭 昌 權

目 次

統計的 方法 序論
1. 平均——통계로 본 우리나나 시멘트 産業
가. 1人當 시멘트 生産量
나. 變動率 및 기타의 平均計算
2. 統計資料의 정리——통계표
3. 統計分析技術
가. 工記法과 添數
나. 誤差와 偏倚
다. 作業表(Technical paper)

1. 平均—統計로 본 우리나라 시멘트産業

가. 1人當 시멘트生産量

統計라하면 먼저 平均値를 생각하게 된다. 가령 1人當平均消費量·平均賃金水準·月平均事務用品費 등이 그것이다. 물론 이러한 平均値를 얻기 위하여서는 基礎統計資料로서 總量規模와 單位가 미리 갖추어져 있어야 할 것인바 調査統計나 業務統計資料에 의해서 각 統計표지의 크기와 統計單位의 個數를 먼저 파악함으로써 통계의 특성을 나타내는 代表値로서 「平均」을 구하는 것이다. 평균에는 數理的 平均과 位置的 平均이 있다.

① **數理時 平均** : 個個의 變量 전부를 사용하여 계산한다. 이에는 i) 算術平均 ii) 幾何平均 iii) 調和平均 및 iv) 平方平均이 있다.

② **位置的 平均** : 變量分布上의 特定位置를 점하는 1個 또는 數個의 變量을 선택하여 平均値로 정한다. 이에는 i) 最頻數, ii) 中位數가 있다.

이들 중 統計分析에서 주로 쓰이는 것은 數理

〈表—1〉 各國의 시멘트生産力比較 1969년도

第1구룹	총생산 1,000M/T	1人當生産 kg	第2구룹	총 생 산	1人當生産	第3구룹	총 생 산	1人當生産
美 國	68,316	336.17	멕시코	6,972	142.48	브라질	7,824	86.13
프랑스	27,696	550.29	터 키	5,796	186.61	泰 國	2,403	69.20
獨 逸	35,004	596.25	中 國	4,092	296.52	필립핀	2,304	61.97
日 本	50,964	496.07	韓 國	4,872	156.50	印 度	13,620	25.36

資料 : UN통계월보

的 平均의 算術平均이며 그밖에 幾何平均과 位置的 平均의 中位數가 각각의 용도와 分析目的에 따라 비교적 많이 쓰인다.

먼저 **算術平均**부터 보면,

1969년에 주요국의 시멘트 生産量과 그것의 各國別 1人當生産量은 〈表-1〉과 같다.

이 表는 대개 1人當시멘트生産量이 300kg 이상인 제1구룹과 300~100kg의 제2구룹, 그리고 100kg 미만인 제3구룹으로 구별하여 볼 수 있도록 각국을 선정한 것이다. 시멘트 生産力이 그 나라의 産業化 수준 및 經濟發展段階를 나타내는 주요지표의 하나로 본다면 이 점에서 한국은 156.5kg으로 제2구룹 즉 中進級에 속하고 있음을 볼 수 있다. 이들 나라 중에서 브라질·인도와 같은 나라는 全國總生産量規模가 餘他國보다 훨씬 큰데 비해 1人當으로는 각 구룹의 하위층에 속하고 있다. 이때 1人當이란 뜻은 그 나라의 「人口 1人當 平均値로서의 시멘트生産量」인바 즉 각국의 시멘트總生産量을 그 나라의 人口數로 나눈 것이다.

平均値=總合計/人口數

한편 1970년에 한국에서 7개의 시멘트회사가 생산한 總量은 5,811,625톤(M/T)이며 이때 最大規模는 双龍의 2,210,806톤, 最低規模는 현대의 316,002톤, 그리고 平均規模는 830,261톤이었다.

〈表-2〉 **各社別시멘트生産量**(1970)

(M/T)

會 社	生 産 量	備 考
大 韓 洋 灰	450,199	
東 洋	855,596	合計
双 龍	2,210,806	=5,811,625
星 信	747,574	平均(算術)
忠 北	422,197	=830,261
韓 一	801,220	
現 代	316,002	

資料 : 韓國洋灰工業協會(가, 나, 다順)

算術平均은 이와같이 먼저 전체의 合計를 계산하고 계산에 포함된 項數(n)로 나누어 구한다. 즉

$$M(x_1+x_2+x_3\cdots\cdots\cdots+x_n)/n$$

시멘트회사의 平均生産規模計算時에 項數(n)를 7個회사 대신 1970년의 우리나라 人口數(31,793,000人)로 나누면 1970년도 한국의 1人當시멘트生産 수준은 182kg으로 산출된다.

나. 變動率 및 기타의 平均計算

한편 **幾何平均**은 비율 특히 變動比率의 平均計算에 쓰인다. 이는 變量의 전부를 서로 곱하고 그 곱한 項數로 「展開한 累乘根」을 산출하는 것이나 실제의 계산에서는 對數(Log)値에 의한 算術平均方法을 사용한다. 즉

$$G=\sqrt[n]{x_1x_2x_3\cdots\cdots\times x_n}$$

$$\log G=\frac{1}{n}(\log x_1+\log x_2+\log x_3\cdots\cdots+\log x_n)$$

꼭 같은 數字에 의해 平均을 계산한다면 幾何平均은 算術平均보다 적다(表-3 參照).

6.25動亂이후 經濟復興作業이 본궤도에 들어선 1956년만 하여도 국내의 시멘트 生産量은 불과 46,564톤이었다. 그러나 14년후인 1970년에는 그것이 5,811,625톤으로 무려 125倍(124.81倍)나 늘어났다. 이때의 年平均 增加率은 얼마인가 하는 것은 幾何平均으로 계산하여야 한다. 이에는 다음의 세가지 方法을 생각할 수 있다. 즉

i) 複利計算法
ii) 最少自乘法
iii) 年率平均計算法

複利計算法에 의하면 初年度의 生産量을 X_0라 하고 1年에 r%만큼씩 늘어나서 n年을 지난 末年度에 X_1만큼 되었다면 末年度의 생산량은

$$X_1=X_0(1+r)^n$$

이 될 것이다. 여기에 初年度(56年) 生産量 X_0에 46,564톤, n=14年이 지난 末年度(70年)의 生産量 5,811,625톤을 代入하여 年平均增加率 r를 구한다. 즉

$5,811,625=46,564(1+r)^{14}$에서 兩邊에 Log를 취하면

$$\text{Log}\,5,811,625=\log 46,564+\log(1+r)\times 14$$

$$6.7643=4.6681+14\times\log(1+r)$$

$$\frac{6.7643-4.6681}{14}=\log(1+r)$$

$$0.1497=\log(1+r)$$

의 逆對數를 구하면

$$1.412=1+r$$
$$r=0.412$$

즉 年平均 增加率(r)는 41.2%이었음을 알게 된다.

한가지 더 예를 들자. 1,천萬원짜리 설비를 定率法으로 10년간 償却하고 殘額評價가 500萬원으로 되게 하려면 年平均 몇%씩 상각할 것인가

期初評價額	1000萬원
期末殘額	500萬원
平均償却率	r(%)
償却年數(n)	10년

$$500=1000(1-r)^{10}$$
$$0.5=(1-r)^{10}$$
$$\log 0.5=10\times\log(1-r)$$
$$\frac{\bar{1}.6990}{10}=\log(1-r)$$
$$\bar{1}.9699=\log(1-r)$$
$$\log 0.9330=\log(1-r)$$
$$0.9330=1-r$$
$$r=1-0.9330$$
$$y=0.067$$

즉 年平均은 償却率 6.7%이다.

그러나 이렇게하여 구한 平均(幾何平均)은 전적으로 初期와 末期의 두 數値에 의해 결정되며 그 중간에 낀 해의 變動率은 전혀 반영되어 있지 않다. 만약 1970年의 數値가 580萬톤이 아니고 경기작용 등에 의해서 '69년이나 별차이 없는 500萬톤이었다면 이때의 평균증가율은 39.7%로 1년간에 1.5%나 낮아진다.

最少自乘法에 의한 年平均增加率 계산은 時系列方程式에 의한 限界値—彈力性係數가 곧 그것이다. 만약에 時系列方程式이 직선(例 $y=b+at$)이면 그때의 時間變數에 대한 係數(a)가 一定期間(1年 또는 $\frac{1}{2}$年)에 이룩한 시멘트생산량의 平均增分(實數 : ㎏)이 되며 時系列方程式이 指數曲線($y=b\cdot a^t$)이라면 그때의 a値는 생산량의 시간탄력성으로서 바로 年平均增加率(100을 乘하여 %)이 된다. 직선이거나 指數曲線이거나 常數(b)項은 바로 생산량 자체의 산술 또는 幾何平均値이다. 따라서 最少自乘法에 의한 時系列方程式에서 年平均增加率을 구한다면 이는 바로 계산기간('56~'70)이 달라짐에 따라 변동할 가능성있는 「生產量의 평균치에서의 平均增加率」이 된다.

1956~1970년간의 時系列方程式과 그에 의한 「平均値에서의 年平均增加率」은 다음과 같다.

直線인 경우 $y=1,653,062+359,115t$
(증가율 21.7%)

指數曲線인 경우 $y=833,900\times 1,353^t$
(증가율 35.3%)

(y : 시멘트생산량. 단위 : ㎏
基準時點 $t=0$: 1963年. t의 單位 1年)

年率平均計算法에 의하면 複利計算法에서와 동일한 數値의 결과를 얻는다. 複利計算의 설명에서는 초년도와 말년도 수치로서 直接平均變動率을 계산하는 년순을 밝힌 것이지만 요컨데 變動率의 계산은 幾何平均이 타당한 것을 증명한 것이다.

즉 1957년은 1956년보다 104.9% 늘어나고 1958년은 57년보다 170.6……1970년은 1969년보다 19.5%씩 시멘트생산이 증가하였으며 이 증가율의 평균은 변동률(%)+100 즉 각년의 前年을 100으로한 指數(57年 204.9, 58年 270.6…, 70年 119.5)를 서로 곱하고 그 곱한 項數(14)로 전개한 累乘根을 산출하여 다시 100을 빼도록 계산하여야만 타당한 14년간의 平均變動率이 된다는 것이다. 이때 실제계산에서는 對數値에 의한 算術平均計算方式이 채용됨을 앞서 언급하였다. 〈表-3〉은 韓國시멘트生產의 1956년 이래 年度別 推移와 變動率 및 그때의 平均變動率計算平順을 표시한 것이다. 이 表에서 「平均」欄에서 ①②는 經濟學的인 의미는 별로 없고 다만 數理式에의 적용을 위한 것이며 ③과 ④는 變動率의 算術平均計算을, ⑤는 그 幾何平均(複利計算과 一致) 計算을 보인 것이다.

調和平均은 각 「變量値의 逆數의 算術平均의 逆數」로서 다음과 같은 算式을 가진다. 즉

$$H=1\Big/\left[\left(\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}\cdots\cdots\frac{1}{x_n}\right)\Big/n\right]$$

中位數는 各變量을 큰것부터 차례로 적은 것에 이르기까지 N個를 정리 나열하여 놓고 순서

〈表-3〉 韓國시멘트生産推移와 平均計算

	年間生産量	生産量의 對數値	對前年變動率 (%)	對前年變動率의 平均計算	
				前年을 100으로 한 指數	指數의 對數値
	M/T				
1956年	46,564	4.6681	—	—	—
57	95,427	4.9796	104.9	204.9	2.3115
58	258,178	5.4120	170.6	270.6	2.4323
59	419,821	5.6230	62.6	162.6	2.2111
60	464,265	5.6668	10.6	110.6	2.0437
61	511.371	5.7087	10.1	110.1	2.0418
62	789,744	5.8975	54.4	154.4	2.1886
63	778,298	5.8912	—1.4	98.6	1.9939
64	1,242,784	6.0944	59.7	159.7	2.2034
65	1,614,141	6.2079	29.9	129.9	2.1137
66	1,884,353	6.2751	16.7	116.7	2.0661
67	2,441,026	6.3876	29.5	129.5	2.1123
68	3,573,538	6.5532	46.4	146.4	2.1656
69	4,864,797	6.6871	36.1	136.1	2.1338
70	5,811,625	6.7643	19.5	119.5	2.0774
合計	24,795,932	88.8165	649.6	2049.6	3.00952
平均	①1,653,062	5.9211 (② 833,900)	③46.4	④146.4	2.1497 (⑤ 141.2)

資料：韓國洋灰工業協會
註：①③④는 算術平均
②⑤는 幾何平均

로 보아 $\frac{n}{2}+1$ 번째(또는 $\frac{n}{2}$번째와 $\frac{n}{2}+1$ 번째의 평균) 즉 變量이 15개라면 8번째를 변량이 크든 적든 代表値로 보는 것이다(〈表-3〉에서는 62年의 789.744M/T).

最頻數는 變量이 가장 여러번 되풀이되는 數値이다.

調和平均. 平方平均. 中位數 및 最頻數의 계산에 대한 例示說明은 생략한다.

2. 統計資料의 정리—통계표

統計資料는 統計表 또는 圖表로 表示된다. 따라서 統計的 방법이라 할 때 그 대상인 統計表에 대한 인식부터 바로 잡아야 할 것이며 統計資料의 選擇·整理 및 그 分析的 處理方法은 統計資料—통계표의 생산과 이용방법의 이해로부터 出發한다.

統計的 방법은 크게 統計調査와 統計分析으로도 구분한다.

統計集團 → 統計調査 → 統計數字 → 統計解析 → 統計法則
(統計調査) (統計分析)

통계는 본래 구체적이고도 경험적인 特定標識를 공통적으로 가지고 있는 개체를 모아 집단화하고 그를 대상으로 관찰한 결과로 얻어진 數値를 의미한다. 따라서 統計的 방법의 第1次的 내용은 「統計集團」「統計標識」「觀察體系」 및 「數字作業」으로서 구성된다. 즉

첫째 : 「統計集團」이 있어야 할 것

둘째 : 특정한 「統計標識」가 정립되어 있어야 할 것

세째 : 特定「基準에 입각한 觀察體系」가 확립되어 있어야 할 것

네째 : 관찰 결과는 일정 양식의 「表」 또는 「圖」로 「數字化」되어야 할 것

등을 요구하고 있다.

「統計集團」의 構成個體는 「統計單位」를 이루며 「統計標識」는 概念定義 등에 관한 基準이 된다.

이 基準에 의한 「觀察體系」는 調査方法을 의미하며 그것의 「數字化」가 즉 「統計表」 또는 「統計圖表」인 것이다.

「統計集團」 및 「統計單位」는 시간적으로 장소적으로 범위가 한정될 뿐 아니라 속성면에서 質 또는 量的으로 統計集團의 특성을 한정시킨다.

統計的 방법의 第一次的 접근은 이처럼 統計集團에 관한 資料의 蒐集·整理를 요한다. 따라서 統計分析에 앞서 分析資料의 선택에서도 調査方法 일반에 관한 單位·基準·觀察方法 등을 먼저 파악하여야 한다. 統計資料는 物價統計·家計調査·市場調査와 같이 애당초 統計를 목적으로 실시된 조사에 의한 것과 犯罪件數·通貨量·豫算執行·生產品出荷·入出庫傳票·賃金支拂台帳같이 該當業務處理過程에서 얻어지는 것이다. 전자를 調査統計資料, 후자를 業務統計資料라고 한다. 어느 경우에나 觀察·蒐集·集計한 統計表에는 單位·基準·槪念·觀察·蒐集方法 등 이른바 統計의 범위와 특성을 식별할 수 있도록 디자인되어야 한다.

統計表는 統計作成者와 統計利用者의 입장에 따라 다음과 같이 구분할 수 있다. 즉

統計表

① 原表 } 統計作成者의 立場
② 集計表 }
③ 簡易表(狹意의 統計表) } 統計利用者의 立場
④ 分析表(圖表를 包含함) }

통계표는 통계작성에 있어서 最終生產物이며 또한 統計分析을 위한 이용에 있어서 기초가 되는 것이다. 위의 구분에서 ①原表는 調査原票類와 傳票·領收證·申告書 및 그 일차적인 綜合集計表·試算表 같은 것이며 ②의 集計表는 原表에서의 綜合表·試算表 등을 포함하여 대개 「○○調査報告」 「第×期決算報告書」 또는 特定目的·事項·內容 등에 관한 「○○統計表」라는 형식으로 공표된 것과 未公表된 것(作業表—Technical Paper) 등 일체의 상세한 통계표이며 통계수자의 原典으로서 작성된 것이다. ③의 簡易表는 대개 狹意의 「統計表」로서 統計月報·統計年鑑 등의 綜合刊行物이나 速報·要約表·總括表 등의 형식으로 된 것이다.

이들 統計表는 통계의 生產者 또는 작성자가 이용자의 입장에서 또는 統計利用의 효율을 높이기 위하여 갖가지 배려를 기우려 작성하여야 하므로 集計項目의 設定 및 그 數, 槪念定義는 물론 表의 내용이 누구에게나 쉽게 이해할 수 있도록 用語를 음미하고 정리된 체제를 갖추어야 한다. 그중에서도 무엇보다 統計表上의 형식요건으로 필수적으로 갖추어야 하는 것은 「單位」와 「基準」이 어딘가에 표시되도록 「디자인」하는 것이다.

統計表의 구성은 대개 表題·頭註·表體·脚註의 네개의 큰 부분으로 이루어지고 있으며 表體는 表頭와 欄, 表側과 行으로 이루어진다. 그중에 表題는 해당통계의 내용, 편제 또는 表의 이용 목적을 집약하여 표현할 수 있도록 用語를 선정한다.

「統計分析」은 이처럼 갖추어진, 原資料로서의 統計表를 이용하여 소정의 절차와 방법에 의하여 解析하고 그에 따라 「統計法則」을 정립하는 것이다. 統計利用者는 統計分析者이며 이는 統計作成者와 동일인일 수도 있으나 대개는 구분된다.

統計分析者가 統計資料의 이용시에 유의하여야 할 일반적 사항은 다음과 같이 요약할 수 있다.

① 統計利用者는 먼저 表體의 計數보다 먼저 表題·頭註·脚註나 表頭·表側부터 파악할 줄 알아야 한다.

② 表體는 그 자체가 이미 統計圖表로 보이겠금 되어야 한다. 특별한 경우가 아니고는 대개의 統計表가 백지에 黑字로써 구성되어 있어 구체적으로 數字를 들여다 보지 않아도 이미 統計圖表와 같은 視覺的인 효과를 나타내고 있는 것이다.

③ 表體의 表側과 表頭에 사용한 用語 또는 記號는 일반 상식적인 개념과는 반드시 일치하지 않고 또 심지어는 전혀 별개의 개념일 수도 있다. 이것은 이른바 「統計的 用語」와 「統計的 槪念」의 특성(抽象的이 아니라 具體的이고 理論的인 것보다 技術的인 것 등)에 基因한 것이다.

④ 統計의 各數字는 어느 경우에나 調査上의 「誤差」를 포함하고 있으며 統計作成時의 人的-

時間的·場所的 또는 政治·社會的인 要因 등이른바 非統計的인 偏倚(Bias)가 있다는 것을 항상 기억하여야 한다. 따라서 利用目的에 따라서는 「統計숫자의 조정」이 불가피하게 된다.

끝으로 統計表의 내용상 분류를 요약하면 다음과 같다.

① 屬性的 統計表(構造系列)
  (ㄱ. 量的屬性統計表(變量系列)
  (ㄴ. 質的屬性統計表(構成比系列)
② 時間的 統計表(時系列)
③ 場所的 統計表(場所系列)
(〈表-4〉~〈表-7〉 참조)

〈表-4〉
<量的構造系列統計表> **토석 및 유리제품제조업의 생산비(1968年)** 단위 : 천원

	원재료비	연료비	구입전력비	구입용수비	위탁생산비	합계
건설용점토제품	666,155	421,243	76,707	1,590	1,076	1,353,439
유리및유리제품	2,981,423	234,215	48,903	2,738	6,700	3,474,850
도자기및토기제품	406,375	311,839	40,504	1,358	570	884,915
시멘트	4,402,026	2,045,552	2,157,563	9,174	3,043	9,476,864
콘크리이트제품	4,245,186	112,111	72,778	11,583	1,198	4,703,714
기타비금속광물제품	520,238	62,080	27,106	5,821	4,809	664,931
합계	13,220,403	3,187,040	2,423,561	32,264	17,396	20,558,713

資料 : 韓國産業銀行·經濟企劃院「鑛工業센서스報告書」1968年. 시리즈 1. 기본통계편 PP. 168

〈表-5〉 <質的構造系列統計表>

1969年度 시멘트企業의 資産·資本構成 단위 : %

	資産構成				負債·資本構成				
							自己資本		
	流動資産	固定資産	移延資産	計	流動負債	固定負債	小計	당기순이익	計
製造業平均	50.4	46.4	3.1	100.0	41.9	31.1	27.0	3.7	100.0
中小企業	53.8	45.0	1.2	100.0	51.6	12.0	48.4	6.6	100.0
大企業	50.4	46.4	3.2	100.0	41.9	31.3	26.8	3.6	100.0
시멘트産業平均	28.9	69.8	1.4	100.0	34.7	38.4	26.8	−1.0	100.0
中小企業	37.7	62.3	—	100.0	35.7	59.2	5.1	−37.1	100.0
大企業	28.8	69.8	1.4	100.0	34.7	38.4	26.9	−0.9	100.0

資 料 : 1969年度「기업경영분석」한국은행

〈表-6〉
<時系列統計表> **시멘트産業總括表**

項目 / 年 / 單位	①生産實績	②輸出實績	③輸入實績	④主要都市建築許可	⑤시멘트都賣價格	⑥全國都賣物價值數
	1,000㌧	1,000㌧	1,000㌧	1,000m³	원	1965=100.0
1965	1,614.1	50.4	—	2,964.3	202	100.0
1966	1,884.4	85.1	190.3	3,683.6	222	108.8
1967	2.441.0	53.9	533.7	5,087.3	227	115.8
1968	3,573.5	77.2	106.3	6,729.3	253	125.2
1969	4,864.8	227.1	15.6	7,467.0	262	133.7
1970	5,811.6	397.9	—	9,210.0	289	145.9

資 料 : ①, ②, ③, ④ 韓國洋灰工業協會「시멘트」1971年 2月
⑤, ⑥ 韓銀調査部
註 : ① 白시멘트除外

② 輸出은船積基準. 國內軍納은 不包含하되 越南軍納包含
③ 輸入은 通關基準
④ 建築許可面積은 70年 6月以前은 韓銀調査部 그 以後는 建設部資料임. 서울 및 韓銀支店所在地의 各市廳이 許可한 新築, 改築, 再築, 增築만을 包含함
⑤ 시멘트 國內都賣價格은 水硬性袋當(42kg)이며 서울地方의 都賣商販賣價格임

〈表-7〉

<場所系列統計表> **地域別 시멘트 消費 實積**(1970年) (單位 : M/T)

	서 울	忠 北	忠 南	慶 南	慶 南	全 北	全 南	濟 州	計
東 洋	306,714	1,550	9,536	95,269	298,743	5,352	32,989	26,389	776,572
大 韓	261,034	11,040	13,939	45,024	27,426	23,393	8,634	—	390,490
雙 龍	856,862	27,148	126,800	104,717	353,514	73,995	163,512	26,669	1,720,301
韓 一	458,730	20,622	47,295	112,060	8,979	31,956	9,872	—	689,514
現 代	172,329	20,721	8,137	22,542	31,760	—	1,544	—	257,003
忠 北	260,144	13,322	42,248	24,057	—	14,126	35,006	—	388,903
星 信	522,307	19,192	54,796	36,321	57,200	16,653	16,092	—	722,561
其 他	221,290	27,745	8,995	38,283	30,987	18,920	33,439	—	379,659
計	3,059,410	141,340	311,746	478,273	808,609	184,395	301,088	53,058	5,337,919

費 料 : 韓國洋灰工業協會
註 : 1. 出荷基準 2. 其他는 工場自家消費 및 寄贈包含

끝으로 度數分布表에 대하여 간단히 說明한다.

度數分布는 觀察値의 크기 즉 統計標識를 가진 값의 크기에 대하여 그 값을 가지는 單位의 갯수를 度數라고 하며 표식에 따라 度數를 하나의 표로 만든 것을 度數分布表라 한다. 앞서 통계는 집단을 전제로 한다고 하였지만 度數分布表에 기록되는 개체의 범위는 일정한 統計條件(統計內容)을 충족하고 있는 개체의 집합인 것이다. 즉 度數分布는 統計集團의 단위를 變量의 크기에 따라 분류한 表로서 그 때 變量을 적당한 폭을 두고 구분한 것이다. 구분된 각 부분을「級」, 各 級變量의 幅을「級間隔」, 級과 이웃級과의 경계를「級限界」, 變量을「級標識」그리고 각급에 속하는 단위의 수를「級度數」또는 단순히 度數라고 한다.

통계표는 대개 이런 度數分布表의 형식으로 정리하여 공표되었거나 그렇지 않으면 統計分析者에 의해서 度數分布表化하여 이용된다. 따라서 度數分布表는 제 1 차적인 통계분석표가 되는 것이다.

3. 統計分析技術

가. 工記法과 添數

지금 國內의 7 個시멘트회사에 대하여 會社名 대신 記號 A를 사용하여 A_1, A_2, A_3……A_7로 표시하고 이들 각사의 시멘트생산량을 각기 x_1, x_2, x_3……x_7으로 할때 一定期間內에 각사의 시멘트生產量合計(X)는

$$X=x_1+x_2+x_3\cdots\cdots+x_7 \quad \text{——(1)}$$

이다.

그와 마찬가지로 전국의 시멘트實需要者 예컨데 시멘트브럭이나 벽돌을 만드는 工場, 建設業會社 등의 수가 가령 500개라고 할 때 이들 實需要者의 記號를 B로 하고 각각을 B_1, B_2, B_3,… B_{500}, 그들의 시멘트消費量(Y)을 역시 y_1, y_2, y_3……, y_{500}으로 한다면 다음과 같은 等式을 생각할 수 있다(輸出入 및 期初在庫는 없다고 假定한다). 즉

$$X=Y+S \quad \text{——————(2) 또는}$$

$$x_1+x_2+x_3\cdots\cdots+x_7=y_1+y_2+y_3\cdots\cdots+y_{500}+S \quad \text{——(3)}$$

(단 S는 期末在庫量)

(2)式은 總量數式(定義式)이나, 실제로는 (3)式에 의하여 사후적으로 파악되는 추상적인 算式이다. 그러나 (3)式은 그 허다한 生產, 消費量에 관한 숫자들을 x와 y에 1, 2, 3,……, 7 또는 1, 2, 3,……, 500과 같은 數字를 添加하여

代數式에 응용할 수 있게 한 것이다. 이처럼 代數文字의 右下에 숫자를 기입하여 區別하는 것을 添數記法이라 한다. 그러나 添數記法을 사용하여도 수의 갯수가 많으면 불편하므로 이를 다시 다음과 같은 방법으로 표시한다.

$A_i(i=1,\ 2,\ 3,\cdots\cdots,\ 7)$

$x_i(i=1,\ 2,\ 3,\cdots\cdots,\ 7)$

$B_i(i=1,\ 2,\ 3,\cdots\cdots,\ 500)$

$y_i(i=1,\ 2,\ 3,\cdots\cdots,\ 500)$

이것은 A_i, x_i, B_i 또는 y_i의 i에 1, 2, 3과 같은 添數를 대입한 전체라는 것이다. 이때 i가 붙은 A_i, x_i, B_i, y_i 등을 一般項이라고 하는데 이러한 一般項의 표시는 i이외에도 j를 사용하며 특정한 一般項에는 대개 k나 l, m, 또는 n(특히 末項의 경우)를 사용한다. 요컨데 一般項에 표시되는 添字 i나 j 代身 1, 2, 3을 대입하면 이는 變量의 크기가 아니라 통계표나 設明事項의 序列順으로 이해하면 되는 것이다. 그러면 이 添數記號의 代數算式은 어떻게 表現하는가, 바로 Σ 記法이 그것이다.

統計計算에서는 個體單位의 표지에 따른 값을 합계하여 統計集團의 값을 구하는 일이 많다. 이때 일정한 범위에 속하는 값을 「모두 合算한다」는 數學的記號로서 Σ(씨그마 英語로는 Summation의 뜻을 簡略한 것임)의 활용법을 익혀 둘 必要가 있다.

(1)式에서 $x_1+x_2+x_3\cdots\cdots+x_7$를 x_i로 표시하면 「x_i에 대하여 i를 1에서 7까지 合計」하여 X를 求한다는 것을

$$X=\sum_{i=1}^{7} x_i \text{——(4)}\quad (i=1,\ 2,\ 3,\cdots\cdots,7)$$

와 같이 한다. 같은 방법으로 消費量 Y는 「y_i에 대하여 i가 1에서 500까지 變하는 것을 모두 합한다」는 것을

$$Y=\sum_{i=1}^{500} y_i \text{—(5)}\quad (i=1,\ 2,\ 3,\cdots\cdots,\ 500)$$

로 한다. 만약에 生產工場이나 實需要者가 7 또는 500이 아니고 n개나 k개가 있다면

$$\left.\begin{aligned} X&=\sum_{i=1}^{n} x_i \\ Y&=\sum_{j=1}^{k} y_j \end{aligned}\right\} \text{——————(6)} \quad \begin{pmatrix} i=1,\ 2,\ 3,\cdots\cdots n \\ j=1,\ 2,\ 3,\cdots\cdots k \\ i\neq j \end{pmatrix}$$

로 표시한다. 이때는 X와 Y를 구분하기 위하여 添字의 一般項을 i와 j로 구분한 것 뿐이다.

$\sum_{i=1}^{n}$나 $\sum_{j=1}^{k}$는 이처럼 하나의 計算手續을 나타내는 記號이지만 만약 그 計算手續의 내용이 명백할 때는 $\sum_i$, $\sum_j$ 또는 단순히 Σ로만 생략하여 쓰인다.

Σ와 비슷한 것으로는 $\Pi(\overset{n}{\Pi}_{i=1})$가 있다. 이는 添字 「$i$가 1에서 n까지 이르는 모든 變量을 서로 곱한다」는 計算手續을 意味한다. 즉

$$\prod_{i=1}^{n} x_i=x_1\times x_2\times x_3\cdots\cdots\times x_n \text{——(7)}$$

統計學에서 公式이나 계산에는 工記法이 자주 사용되므로 여기에 그 중요한 性質을 설명하여 둔다.

① $\sum_{i=1}^{n} x_i+\sum_{i=1}^{n} y_i=\sum_{i=1}^{n}(x_i+y_i)$

② $\sum_{i=1}^{n} c\cdot x_i=c\cdot\sum_{i=1}^{n} x_i$(C는 定數)

③ $\sum_{i=1}^{n} c=n\cdot c$

④ $\sum_{i=1}^{n}=\sum_{i=1}^{n} 1=n$

⑤ $(\sum_{i=1}^{n} x\cdot\sum_{i=1}^{n} y)\cdot\sum_{i=1}^{n} y=\sum_{i=1}^{n} x(\sum_{i=1}^{n} y\cdot\sum_{i=1}^{n} y),\ \sum_{i=1}^{n}\frac{x^2y}{xy^2}$

$=\sum_{i=1}^{n}\frac{x}{y}$,

⑥ $\Sigma(xy)\neq\Sigma x\cdot\Sigma y,\ \frac{\Sigma x^2y}{\Sigma xy^2}\neq\frac{\Sigma x}{\Sigma y}\neq\Sigma\frac{x}{y}$,

$\Sigma x\cdot\Sigma(y+z)=\Sigma(x+y)\cdot\Sigma z$

($\sum_{i=1}^{n}$의 n과 $i=1$를 省略하였음)

나. 誤差와 偏倚

統計分析—統計計算에서 나타나는 수자의 誤差는 調査誤差와 計算誤差가 있다. 조사오차는 통계조사시나, 자료수집 및 集計時에 발생한 것이며 특히 標本調査인 경우의 標本誤差, 추출오차가 있고 申告者나 公票記載者의 주관적 차이, 부주의로 말미암은 調査記載, 集計過程 등에서 나타나는 非標本誤差같은 것이 있다.

계산오차는 e, $\sqrt{2}$, π의 값, 確率積分값과 같이 순전히 숫자상의 특성에서 오는 誤差이다. 이 計算誤差도 近似値誤差, 循環誤差, 計算不注意

誤差 등이 있다.

실제의 통계분석에서는 이상의 여러 誤差가 몇 개씩 동시적으로 혼합되어 있다. 그리하여 分析作業過程을 거쳐 구한 計數에는 誤差가 극히 복잡하게 얽혀 있는 것이 보통이다. 이런 경우 誤差를 정확히 가려낸다는 것은 불가능하기도 하지만 별로 중요한 것도 아니다. 다만 誤差의 대체적인 범위를 파악하고 따라서 이 誤差의 범위를 가능한한 축소시키도록 배려한다. 이른바 誤差法側이라는 것이 있다. 이는 이상의 誤差들이 단순히 우연적으로 발생한 것이며 그것은 대개 正規分布(이를 Gauss 分布라고 함)를 하고 있다고 보아 통계분석으로 처리할 수 있는 각종 연구가 이미 되어 있다. 요컨데 誤差없는 통계가 없으며 또한 통계 숫자는 誤差가 포함되어 있다는 것이다.

그런데 誤差와는 달리 偏倚(편이)라는 것이 있다. Bias(바이아스)라고 英譯하지만 통계숫자의 偏差에는 위의 誤差(Error)와 偏倚(Bias)가 있는 중 위에서 언급한대로 誤差의 처리는 통계적 방법으로 어느 정도 가능하나 Bias의 처리는 애당초 非統計的인 요소로 보고 있는 것이다. 현재 分散分析, χ(카이) test 등으로 통계편차의 크기가 적정한 危險水準을 넘으면 위에 말한 誤差法則의 한계를 넘은 Bias로 간주하고 「統計的으로 기각」시키고 있다. 따라서 일반적으로 통계의 信憑性을 논할 때는 제1차적으로 이 Bias를 다룬다. 特定調査에서 조사대상을 無作爲抽出(Random Sampling)로 할 때에는 Error는 크지만 Bias는 적게 된다. 그러나 앙케이트調査와 같은 有意抽出時에는 Error의 크기는 적게 할 수 있지만 이때에는 選擇者의 주관에 의하기 때문에 Bias가 커지게 마련이다.

統計分析者의 임무는 통계자료의 해석에 의한 統計法則을 발견하는 것이지만 그때의 통계법칙의 발견이란 다름 아닌 誤差範圍, 誤差法則의 발견과 Bias의 抽出로 集約할 수 있다. 특히 Bias의 抽出에 의한 統計資料의 信賴度 區間 評價는 분석자료의 「調整」作業을 필요로 하게 되는 소이이다. 統計調査者가 정확하게 조사하였다고 그대로 正直한 統計숫자로 간주할 수는 없는 것이다.

다. 作業表——Technical paper

모든 企劃業務 또는 管理業務에서와 같이 統計分析作業에는 그때그때에 적합한 作業表를 작성하게 된다. 이 작업표는 作業者의 능력·습관·작업환경 또는 保有하고 있는 分析機器에 따라 한갓되게 설명할 수는 없지만 일반적으로 다음과 같은 요건과 특성을 갖는다.

① 作業表는 作業者 자신의 分析作業을 위하여 작성한다. 따라서 作業表上의 각종 用語나 符號 또는 계수의 취급은 철저하게 작업자 本位이다.

② 作業表作成時의 筆器具는 가장 손쉬운 것(예컨데 연필이나 볼펜)으로 또 그밖에 目表나 표의 양식을 그릴때 필요한 器具도 단순한 것으로 한다. 그러기 위하여 作業표는 애당초 유동적이나마 특정한 form(가령 從橫의 눈금, 細線을 갖춘)을 갖춘다.

③ 作業表는 統計原典과 분석표를 링크시키는 역할을 하는 Draft이다. 따라서 Draft는 원칙적으로 對外秘이거나 최소한도 공표할 의무는 없는 것이다.

④ 統計分析作業은 작업단계가 아무리 많고 또 복잡하여도 특히 算式展開나 計量模型의 설정에 있어서는 每段階마다 그 直前段階의 작업을 檢證(檢定, 評價)할 수 있도록 작업표상에 배려되어야 한다.

作業表는 一般事業計劃書·資料目錄·參考文獻索引 등 가장 일반적인데서 基礎原資料整理·調整·算式展開·分析原稿 같은 복잡한 것에 이르기까지 요컨데 最終作業이 完成되어 그 結果가 公表될 때까지의 모든 作業實績이다.

統計分析의 성과에 대한 평가는 최종作業結果에서 나타나겠지만 성과를 기대할 수 있는가의 與否는 바로 작업표를 얼만큼 효과적으로 활용하는가에 달렸다고 할 것이다.

가장 적은 人員과 時間 그리고 勞力으로서 애당초 예기하였던 성과를 얻기 위하여는 작업자의 試行錯誤를 가장 적게 하는데 있으므로 말하자면 작업표는 作業者의 試行錯誤가 집약되어 있는 것이다. 근래 統計分野에서의 品質管理論이 중시되는 이유도 바로 이 試行錯誤를 축소시키기 위

한作業表의 관리를 얼만큼 효율적으로 할 것인가에 있는 것이다.

<作業表例示 1>

X. M. G

枚中 枚

$$\frac{\triangle x}{\triangle M}=\frac{\eta_m \cdot g_w \cdot X}{m \cdot g \cdot Y}\cdot =1$$

$$=\frac{7.4932\times 0.052}{3.7762}\cdot\frac{X}{Y}\cdot\frac{1}{g}=1$$

$X/Y=x$ 라 하면

$$\frac{0.3896}{3.7762}\cdot\frac{x}{g}=1$$

$g=0.1032x$

$x=0.25\sim0.30$ 일 때

$g=0.026\sim0.031$

즉 2.6%~3.1%

그러나 이는 한계 輸入性向(m)이 3.7762 나 되는 크기에 연유하므로 만약에 이를 1, 즉 수입의 증가율이 소득의 증가율과 같이 조정만된다면

(즉 $\triangle x=m(\triangle Y)$ m의 m=1 이라면)

$g=0.3896x$ 에서 $x=0.25\sim0.30$ 일 때

$g=0.097\sim0.117$

즉 9.7%~11.7%를 期待할 수 있다

① $x=\frac{1}{0.1032}\cdot g$

$=9.6899g$

$g=0.09\sim0.11$

$x=0.8721\sim1.0659$

② $x=\frac{1}{0.3896}\cdot g$

$=2.5667g$

$g=0.09\sim0.11$

$x=0.231\sim0.282$

(第1分科)

<作業表例示 2>

Korea

	Y	P	M	X	Y_w
60	−29.0	−19.4	−51.5	−158.1	−19.9
61	−25.2	−14.6	−56.5	−148.9	−16.5
62	−22.2	−9.7	−37.5	−132.8	−11.4
63	−13.8	−2.3	−12.9	−96.0	−7.0
64	−5.4	−1.3	−40.8	59.2	−0.9
65	2.5	0.6	−30.2	5.1	4.9
66	17.5	12.4	15.0	91.4	11.8
67	28.1	14.0	65.0	171.8	16.2
68	47.2	20.2	149.2	327.0	23.1
M	113.8	102.3	112.9	196.0	107.0

$y^2=5518.43$ $My=ay^2+byP$ $xy_w=ay^2_w+by_wP$

$p^2=1,448.75$ $Mp=ayp+bp^2$ $xp=ay_wp+bp^2$

$M^2=36,704.48$

$X^2=222,347.51$

$y^2_w=1,807.33$ $M\begin{cases}5,518.43a+2,749.96b=13,203.88\\2,749.96a+1,448.75b=6,362.18\end{cases}$

$y_p=2,749.96$ $X\begin{cases}1,807.33a+1,597.49b=19,176.05\\1,597.49a\ 1,448.75b=16,820.10\end{cases}$

$y_m=13,203.88$

$mp=6,362.18$

$xp=16,820.10$ $5,518.43a+2,749\cdot96b=13,203.88$

$y_wp=1,597.49$ $a+0.498323b=2.392.688$

$y_wx=19,176.05$

$$-\begin{vmatrix}2,749.96a+1,448.75b=6,362.18\\2,749.96a+1,370.368317b=6,579,797093\end{vmatrix}$$

0 $78.381683b=-217.617093$

$b=-2776377$

$a=3776220$

$R=0.958$ $c=-31810469$

私有機關車8臺導入運營

祝

1971年 2月 12日

韓國洋灰工業協會

品目名 : 시 멘 트 No. 31 포오트란트시멘트, 1,000MJ BOK 시멘트 : 포래 42kg 도매 統

	수급동향				대표가격	월별가격추이											
	국내생산	수입	수출	국내소비	의 추이	1	2	3	4	5	6	7	8	9	10	11	12
1965	1,614				202	176	176	176	183	198	208	207	213	221	221	221	221
66	1,884				222	221	221	221	223	223	223	223	223	223	223	223	223
67	2,441				227	230	223	223	223	223	223	223	223	223	232	234	243
68	3,572				253	249	253	253	253	253	253	253	253	253	253	253	253
69	4,865				262	253	253	253	253	253	268	268	268	268	268	268	270
70	504				288	270	286	294	294	294	280	284	294	294	294		
연평균 증가율																	

	A회사	B회사	C회사		회사	회사		회사	회사		회사
유동자산	32.95 3,117,093	13.92 3,975,642		인건비	87,610	2.26 231,497	재료비	2,341,010	4,728,602	고정자산비율	83.57
고정자산	66.66 6,305,978	83.57 23,867,956		교통비	8,938	0.20 20,379	노무비	208,353	701,970	자기자본비율	9.18
이연재정	0.39 37,095	2.51 717,366		통신비	8,825	0.24 24,404	경비	959,001	3,780,631	유동비율	56.97
유동부채	33.58 3,176,352	24.43 6,978,753		수도광열비		0.08 8,404	외주가공비			고정비율	911.42
고정부채	35.71 3,378,117	66.39 18,960,574		세금·공과	260,102	0.01 104,048	전력비	418,734	1,349,253	부채비율	989.43
자본금	35.20 3,330,000	11.97 3,418,000		지대·집세	21,745	0.08 7,984	수도광열비		22,013	고자회전율	0.43
잉여금	△4.49 △424,303	△3.14 △679,363		감가상각비		0.16 16,848	운임	4,178	451,378	총자본회전율	0.36
				보험료	9,779	0.13 13,154	감각상각비	419,821	1,697,475	총자본이익률	−18.90
총자산	100.00 9,460,166	100.00 28,560,964		교제비	162,092	0.24 24,322	수선비	14,624	23,480	금융비용률	19.85
순매출액	6,265,910	100.00 10,264,381		광고선전비	10,481	0.32 32,473	세금공과	3,748	7,350	영업비용률	13.77
매출원가	4,093,603	77.89 7,994,501		보관비	510,780	3.21 329,962	보험료	3,717	3,450	광고비율	0.32
매출이익	2,172,307	22.11 2,269,880		기타	401,700	5.85 600,154	여비교통비	3,450	25,544	교제비율	0.24
판관비	1,483,052	13.77 1,413,629					통신비	1,105	7,588	감가비율	23.71
영업이익	689,255	8.34 856,251		판관비합계	1,483,052	13.77 1,413,629	기타잡비	89,624	196,100	조세부담률	1.08
영업의수익	200,909	6.25 641,020					당기제조비용	3,508,364	9,211,203	부가가치액	4,264,443
영업의비용	1,105,909	△19.85 △2,036,972					조정항목	△741,312	△1,154,357	부가가치율	41.54
당기순손익	△215,745	△5.26 △539,701					당기제품원가	2,767,052	8,056,846	투자효율	14.93

규격	1.	2.	3.	4.	총합평균
생산량					
금액					
단가					

구분	내용				
협회또는조합명					
회원수	년	년	년	년	년
설립년월					
주요기능업무	1.	2.	3.		
가격조정기능					
가격형태					
대정부건의사항	1.	2.	3.		

<韓國産業開發研究所>