試料의 크기의 合理的 決定方法

金星紡織株式會社

企劃調査室長

徐　　丙　　三

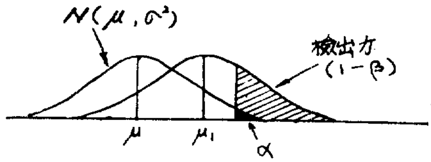
試料의 크기(Sample size)를 얼마로 하면 모집단의 특성을 가장 잘 대변하여 줄수 있으며 또 훌륭한 檢出力을 발휘할 수 있을까?

檢査나 試驗 或은 다른 目的으로 샘플링을 해본 사람이라면 누구나 이러한 의문을, 한 두번은 반드시 가졌으리라 믿는다. 왜냐하면 우리는 母集團에 關한 지식을 試料를 通하여 얻고자 하기 때문에 試料의 크기를 마음대로 定할수 없는 것이며 統計的으로 가장 妥當한 數를 샘플링 하여야 하기 때문이다.

試驗에 의하여 얻어진 데이터로 檢定을 하고자 할 때 檢出力은 생각지도 않고 제멋대로 試料의 크기를 定하여 그저 機械的으로 平均値의 檢定을 한다면 그 結論은 순수한 統計學的 의미에서 볼때에는 뜻이 있을런지 모르나 實際的으로는 아무런 價値가 없는 무의미한 것이며, 이것을 보고 判斷을 내린다는것은 아주 위험한 일이다. 自己가 가지고 있는 그릇된 생각에 의하여 試料의 크기를 늘였다, 줄였다, 한다는 것은 있을수 없는 일이며 어디까지나 통계학적 근거를 가진 試料의크기를 定해야 試料로서 참다운 價値가있는 것이다.

試料의 크기를 너무 크게 하면 실제로는 아주 무시해도 좋은 조그만한 差까지 檢出되어 마치 큰 變化나 일어난 것 같이 判斷을 내리는 경우가 있고 또 너무 작게하면 실제는 당연히 문제가 되는 큰 差가 있는데도 檢出되지 않고 넘어가는 수가 있다. 따라서 통계적 검정(엄격한 의미로 볼때 모든 검사나 실험은 一種의 검정이라고 생각할 수 있다)을 해야 할 때 어느 정도의 檢出力을 가진 검정을 해야 할것인가를 결정해야 하고, 실제로 差가 있으면 그 差를 90% 或은 95% 검출할수있도록 試料의 크기 n을 決定해야 한다.

$N(\mu, \sigma^2)$되는 正規分布를 하는 母集團에서 試料를 取했을때 이試料에 依한 母平均의 推定値가 μ_1이라고 하면 μ와 μ_1사이에는 $\mu_1-\mu$란 差가 있게 되고이 差로 因하여 $N(\mu, \sigma^2)$의 信賴限界 밖으로 나가는 檢出力$(1-\beta)$은 커지게 된다. 即 제1도에서 보는 바와 같이 一定한 有意水準 α일때 얻어지는 試料 分布의 검출력이 커짐에 따라 母集團의 母平均과 試料集團의 평균과의 差가 커진다. 正規分布 $N(\mu, \sigma^2)$의 유의수준 α밖으로 나가는 확률이 試料集團의 검출력이 된다.



第1圖 檢出力圖

α를 一定하게 두고 試料平均이 變하면, 即 $\mu-\mu_1$의 差가 변하면 이에 따라 검출력도 달라진다.

모집단의 標準편차 σ를 알때 使用하는 규준측도

$$u=\frac{\mu-\mu_1}{\sigma/\sqrt{n}}$$

를 利用하여 $\mu-\mu_1$의 變化에 따라 取해야 될 試料의 크기를 求하는 式을 유도할 수 있다.

$$\text{即}\quad u=\frac{\mu-\mu_1}{\sigma/\sqrt{n}}=\frac{\sqrt{n}(\mu-\mu_1)}{\sigma}$$

$$\sqrt{n}=\frac{u\sigma}{\mu-\mu_1}$$

$$\therefore\ n=\left(\frac{u\sigma}{\mu-\mu_1}\right)^2$$

이 式에서 u는 試料의 크기가 n이고 표준편차가σ일때 시료평균이 母平均으로 부터 σ의 몇倍 되는 位置에 떨어져 있느냐를 意味하는 것이므로

$$u=K(\alpha)-K(\beta)$$

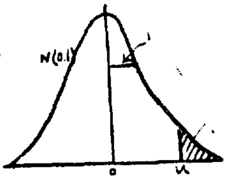
란 一般式으로 놓을수 있다. 양쪽 검정일때는

$$u=K(\alpha)-K(\beta_1)$$
$$u=K(\alpha)-K(\beta_2)$$

라고 표시할수 있고 여기서 K란 第1表 正規分布表에서의 K의 값을 나타낸다.

第1表 正規分布表

$$P=\frac{1}{\sqrt{2\pi}}\int_u^\infty e^{-\frac{u^2}{2}}du$$



第 1 表 正規分布表

$K \to P$

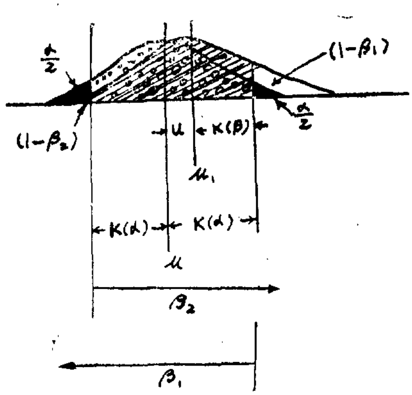
K	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	5000	4960	4920	4880	4840	4801	4761	4721	4681	4641
0.1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
0.2	4207	4168	4129	4090	4052	4013	3974	3936	3897	3859
0.3	382	3783	3745	3707	3669	3632	3594	3557	3526	3483
0.4	3446	3409	3372	3336	3300	3264	3228	3192	3156	3121
0.5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
0.6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
0.7	2420	2389	2358	2327	2296	2266	2236	2206	2177	2148
0.8	2119	2090	2061	2033	2005	1977	1949	1922	1891	1867
0.9	1841	1814	1788	1762	1736	1711	1685	1660	1635	1611
1.0	1587	1562	1539	1515	1492	1469	1446	1423	1401	1379
1.1	1357	1335	1314	1292	1271	1251	1230	1210	1190	1170
1.2	1151	1131	1112	1093	1075	1056	1038	1020	1003	0985
1.3	0968	0951	0934	0918	0901	0885	0869	0853	0838	0823
1.4	0808	0793	0778	0764	0749	0735	0721	0708	0694	0681
1.5	0668	0655	0643	0630	0618	0606	0594	0582	0571	0559
1.6	0548	0537	0526	0516	0505	0495	0485	0475	0465	0455
1.7	0446	0436	0427	0418	0409	0401	0392	0384	0375	0367
1.8	0359	0351	0344	0336	0329	0322	0314	0307	0301	0294
1.9	0287	0281	0274	0268	0262	0256	0250	0244	0239	0233
2.0	0228	0222	0217	0212	0207	0202	0197	0192	0188	0183
2.1	0179	0174	0170	0166	0162	0158	0154	0150	0146	0143
2.2	0139	0136	0132	0129	0125	0122	0119	0116	0113	0110
2.3	0107	0104	0102	0099	0096	0094	0091	0089	0087	0084
2.4	0082	0080	0078	0075	0073	0071	0069	0068	0066	0064
2.5	0062	0060	0059	0057	0055	0054	0052	0051	0049	0048
2.6	0047	0015	0044	0013	0041	0040	0039	0038	0037	0036
2.7	0035	0034	0033	0032	0031	0030	0029	0028	0027	0026
2.8	0026	0025	0024	0023	0023	0022	0021	0021	0020	0019
2.9	0019	0018	0018	0017	0016	0016	0015	0015	0014	0014
3.0	0013	0013	0013	0012	0012	0011	0011	0011	0010	0010

양쪽 검정일때의 검출력은$(1-\beta_1)+(1-\beta_2)$이므로 $2-(\beta_1+\beta_2)$가 된다.

또 한쪽 檢定일때의 u는 $u=K(\alpha)-K(\beta)$로 주어지고檢出力은 $1-\beta$가 된다.

第2圖와 第3圖는 各各 양쪽 檢定일때와 한쪽檢定일때의 檢出力을 그림으로 표시한것이다. 이 그림을 보면 正規分布에 있어서 檢出力이라는 것이 어떤의 미를 지니고 있는지 알수 있다.

실제로 試料의크기를 決定할때 표준편차(σ)나 平均值의 差 $\mu-\mu_1$은 쉽게 구할수 있으나 문제는 u의



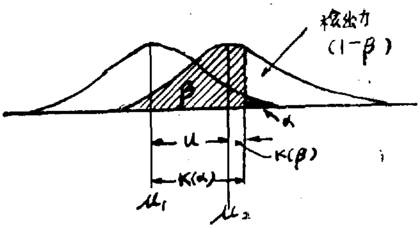
第2圖 양쪽檢定일때의 檢出力圖

값이다.

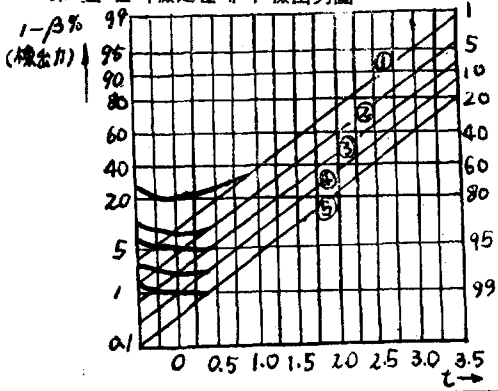
이 u의 값만 쉽게 求한다면 언제든지 必要할때 即席에서 試料의 크기를 求할수 있을 것이다. 그러나 n는 正規分布의 確率에 依하여 구하여야 되므로 統計學에 아주 숙달하지 못한 사람은 그 計算이 어려울뿐만 아니라 어설프게 잘못 계산하면 오히려 엉터리 값을 구하게 될 위험성마저 內包하고 있다. 따라서 이러한 難點을 해결하기 위하여 α와 β만을 주고 u를 구할수 있도록 圖表化한 것이 있다. 이러한 그림을 平均値의 差의 검정 (σ를 알때)에 있어서의 OC曲線圖라고 하며 第4圖는 이것을 보여준다.

第2表는 第4圖를 第5圖(A)(B)와 같이 檢出力 $(1-\beta)$가 有意水準 α가 5%에서 95%까지 變하는 동안 의 여러가지의 값을 구한 것이다.

第4圖는 $\beta, 1-\beta$일때의 u의 값을 그린 그림이며 이 그림에서 u를 구하고 앞의 公式을 利用하여 合理的 試料의 크기 n을 구할수 있다.



第3圖 한쪽檢定일때의 檢出力圖

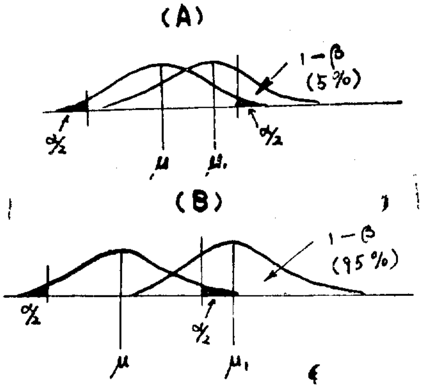


α %	①	②	③	④	⑤	
한 쪽	10	5	2.5	1.0	0.5	직선 만 사용
양 쪽	20	10	5.0	2.0	1.0	곡선 부분사용

第4圖 CC曲線圖

第2表 OC 曲線 算出表

β	1-β	α=5%일때		α=10%일때		α=20%일때	
第2種의 差課	(檢出力)	$K(\beta)$	u	$K(\beta)$	u	$K(\beta)$	u
0.0505	0.9495	0.32	3.60	0.00	3.28	—	—
0.1003	0.8997	0.68	3.24	0.36	2.92	0.00	2.56
0.1515	0.8485	0.93	2.99	0.61	2.67	0.25	2.31
0.2005	0.7995	1.22	2.80	0.80	2.48	0.44	2.12
0.2514	0.7486	1.29	2.63	0.97	2.31	0.61	1.95
0.2015	0.6985	1.44	2.48	1.12	2.16	0.76	1.80
0.3520	0.6480	1.58	2.34	1.26	2.02	0.90	1.66
0.4013	0.5987	1.71	2.21	1.39	1.89	1.03	1.53
0.4522	0.5478	1.84	2.08	1.52	1.76	1.16	1.40
0.5000	0.5000	1.96	1.96	1.64	1.64	1.28	1.28
0.5478	0.4522	2.08	1.84	1.76	1.52	1.40	1.16
0.5987	0.4013	2.21	1.71	1.89	1.39	1.53	1.03
0.6480	0.3520	2.34	1.58	2.02	1.26	1.66	0.90
0.6985	0.3015	2.47	1.45	2.16	1.12	1.80	0.76
0.7486	0.2514	2.54	1.38	2.31	0.97	1.95	0.61
0.7995	2.2005	2.80	1.12	2.48	0.80	2.12	0.44
0.8485	0.1515	2.99	0.93	2.67	0.61	2.31	0.25
0.8991	0.1003	3.24	0.68	2.92	0.36	2.56	0.00
0.9495	0.0505	3.60	0.32	3.28	0.00	—	—



第5圖 檢出力 比較表

第4圖를 利用하여 試料의 크기를 決定하는 方法을 說明하기로 하자.

어떤 規格品의 실(絲)은 그 强度의 모평균이 68kg 모표준편차가 2.5kg 이라는 것을 알고 있다. 原料混合方法을 變更하여 실을 紡出하였을때 그 母平均이 번했나? 或은 그대로 있는가?를 알고 싶다. 이때 試料의 크기를 몇個로 하면 올바른 判斷을 할수 있을까? 이 실에 있어서 萬若 强度가 2kg以上 差가 생긴다면 當然히 문제시해야 한다고 한다.

위에 주어진 값을 文字로 표시하면 $\mu=68.$ $\sigma=2.5$ $\mu-\mu_1=2$이고 $\alpha=5\%$와 $\beta=10\%$의 수준에서 判斷을 내리고 싶다면 第 4圖에서 $\alpha=5\%$와 $\beta=10\%$일때의 값을 구하면 $u=3.24$가 된다

$$n=\left(\frac{u\sigma}{\mu-\mu_1}\right)^2=\left(\frac{3.24\times2.5}{2}\right)^2\fallingdotseq16$$

即 새로운 原料의 混合方法으로 紡出한 실의 試料를 16個 取하여 强度를 測定하면 된다.

위에 說明한 方法은 試料를 채취할 모집단의 σ를 알고 있을 때에만 利用할수 있기 때문에 標準偏差를 모르는 試料集團에는 利用할 수 없다. 따라서 이러한 境遇에는 시료를 채취하여 직접 標準偏差를 계산하여야 되는데 여러로트에서 採取한 시료의 크기가 같을때의 平均標準偏差 $\bar{\sigma}$는

$$\bar{\sigma}=\frac{\sigma_1+\sigma_2+\sigma_3+\cdots\cdots+\sigma n_{n-1}+\sigma_m}{m}$$

라는 식으로 계산할 수 있고 시료의 크기가틀릴때는

$$\bar{\sigma}=\sqrt{\frac{n_1\sigma_1^2+n_2\sigma_2^2+n_3\sigma_3^2+\cdots\cdots+n_{m-1}\sigma^2_{m-1}+n_m\sigma^2}{n_1+n_2+n_3+\cdots\cdots+n_{m-1}+n_m}}$$

라는 식으로 계산하게 된다.

일단 平均標準偏差 $\bar{\sigma}$를구하면 變動率 (Coefficien of Variation) $v=\frac{\sigma}{\bar{x}}\times100$을 구하고

$$n=\frac{u^2\times v^2}{E^2}$$

依하여 시료의 크기 n을 求할 수 있다. 各記號는

u:t分布에서 $n=\infty$일때의 t의값

v : 變動率

E : 샘플링 誤差

를 나타내는 것으로 u의 값은 確率에 따라 그 값이 각각 다르다.

確　率(%)	u의 값
90	1.645
95	1.960
99	2.576

即 위의 表에서 보는 바와 같이 試料에서 얻는 情報의 精度를 높이려면 確率이 커야 하고 따라서 試料의 크기 n도 커야 한다. 또한 試料의 分散狀態가 나쁘면 나쁠 수록 샘플링해야 할 試料의 크기도 커져야 한다. 그러나 大部分의 경우 한번 定한 試料의 크기를 그냥 習慣的으로 長期間 使用하는 例가 많은데 이러한 일은 되도록 삼가야 좋을줄 안다.

變動率v는 試料 平均値에 對한 표준편차 σ의 比率을 百分率로 계산한 것인데 測定値의 分散狀態를 表示하는 좋은 指標가 된다.

샘플링 誤差라는 것은 쉽게 말하면 샘플링을 잘못하여 생긴 誤差를 말한다. 어떤 統計的 結果가 나왔을때 그 결과에는 참값 外에 반드시 誤差라는 것이 따라 다니게 되며 이 誤差는 測定器나 測定하는 사람이 잘못해서 생기는 測定誤差, 계산을 잘못해서 생기는 計算誤差, 周圍환경(온도, 습도, 晝, 夜等)에 따라 發生되는 自然誤差, 試料 採取때에 잘못하여 생기는 샘플링 誤差等 여러가지 種類의 誤差가 內包된다. 이中에서 特히 샘플링을 잘못해서 생기는 誤差를 E 로 表示하여 試料의 크기를 求하는데 利用하는 것이다.

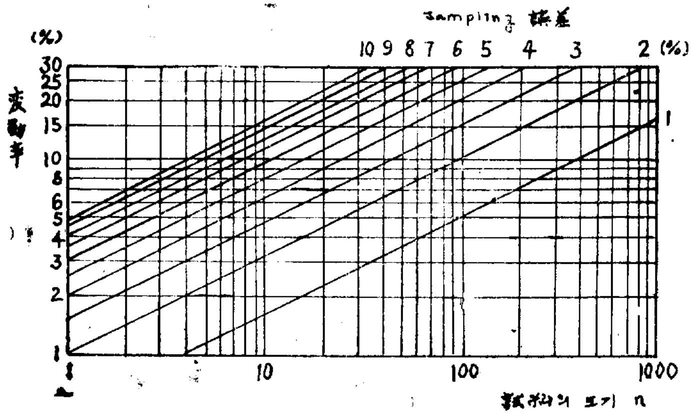
샘플링 誤差 E는 計算上으로 求하는 것보다 주어진 規格이나 經驗的 限界性에 依하여 定하는 것이實用的이다. E는 $0.1\%\sim10\%$까지를 使用하는 것이 一般化된 常識이며 이 範圍內에서 製品의 精度나 信賴

性等을 考慮하여 決定하면 된다.

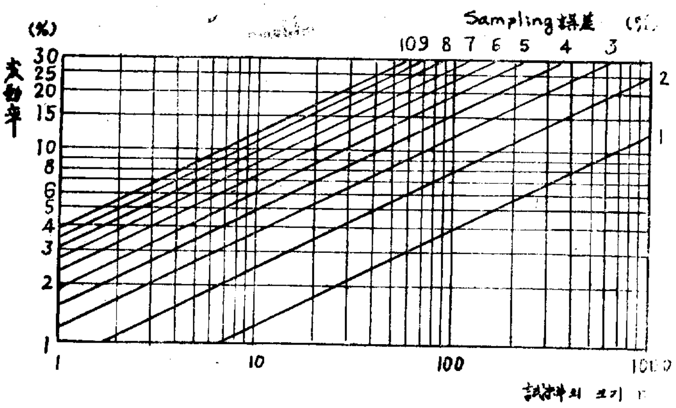
그럼 例를 들어 보자.

어떤 織物에 使用하는 糊液 PH價의 變動率이 5%이고 可能한 限 糊液採取時 發生하는 샘플링 誤差를 3%以內로 하여야 된다고 한다. 이때 $\alpha=5\%$ 水準에서 試料의 크기를 定하고자 한다. 얼마로 하면 될까

먼저 앞의 式 $n=\frac{u^2\times v^2}{E^2}$ 으로 계산하면



第6圖 $p=0.95$일때 샘플링 誤差 1～10%에對한 變動率과 試料의 크기



第7圖 $p=0.99$일때 샘플링 誤差 1～10%에 對한 變動率과 試料의크기

$$n=\frac{(1.96)^2\times(5)^2}{(3)^2}=\frac{3.84\times25}{9}=10.7\fallingdotseq11$$

(但 $\alpha=5\%$일때 $t=1.96$)

가 된다. 그러나 이렇게 일일히 계산하면 번잡하므로 第6圖와 같은 圖表를 使用하면 每檢査時(不確實事象일때)마다 계산하지 않아도 試驗者나 檢査者가 統計的 意味를 모르더라도 機械的으로 試料의 크기를 決定할 수 있으므로 아주 便利하다.

이 圖表는 對數方眼紙에 샘플링 誤差를 直線으로 表示한 것으로 세로에 變動率(%)을, 가로에 試料의 크기(Saple size)를 各各 表示하고 있다

이제 變動率 5%, 샘플링 誤差 3%일때의 試料의 크기를 第6圖에 依하여 求하여 보자.

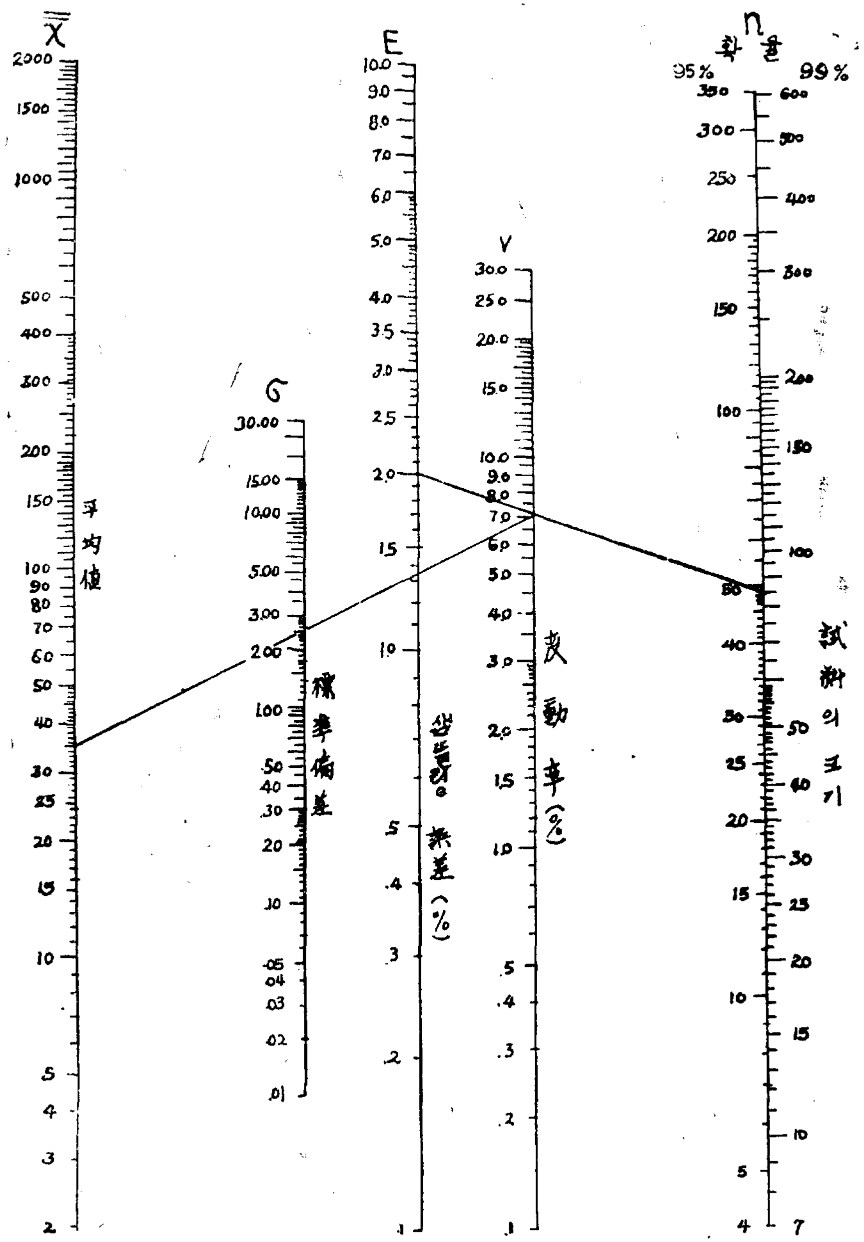
우선 세로의 變動率에서 5%되는 곳을 찾고 이線을 따라 오른쪽으로 옮기면서 샘플링誤差 3%되는 빗금(斜線)과 맞나는 點에서 垂直으로 내려 제일 밑 가로 줄의 試料의 크기가 線上에 닿는 數字를 읽으면 10.7이 나온다. 이 圖表에서는 소숫점 以下까지 자세히 나오지 않았기 때문에 어림수로 대충 읽으면 된다. 이 결과를 보면 앞의 式으로 計算한 것과 같음을 알수있다.

第7圖는 確率(P)이 99%일때의 試料의 크기를 구하는 圖表이니 第6圖와 比較하면서 參考하여 주기 바란다.

第8圖는 평균치와 표준편차만 알면 變動率 (v)와 試料의 크기 n을 圖表上에서 쉽게 구할 수 있도록 만든 NOMOGRAM이다. 이NOMOGRAM 의特徵은 變動率을 計算하지 않아도 좋다는 點과 確率이 95%일대와 99%일때의 試料의 크기를 한번에 찾을 수 있다는點이다.

그러나 平均値와 標準偏差는 다른 경우와 마찬가지로 계산해야 된다는 點에는 變함이 없다.

지금 평균치가 35kg이고 표준편차가 2.5kg인 銅線의 引張强度 測定値가 있다고 하자. 먼저 $\overline{x}$ 눈금 위의 35와 σ눈금 위의 2.5를 연결하는 直線을 그리고 이 直線이 v눈금과 맞나는 點 7.1을 읽는다. 이 값이 $\overline{x}=35$, $\sigma=2.5$일때의 變動率 v의 값이다. 이 銅線을 샘플링할때 우리가 願하는 샘플링誤差는 2%로 하고 싶다고 하면 $v=7.1$과 E눈금 위의 2.0을 연결하는 直線을 그린다. 이直線이 試料의 크기 위의 눈금과 마주 치는 곳을 보면 確率이 95%일때는 $n=49$이고 99%일때는 $n=85$임을 알 수 있다. 이와같은 方法으로 이미 說明한 여러가지 側의 경우를 적용시켜 보아도 똑 같은 값이 나온



第8圖 試料의 크기를 求하는 NOMOGRAM

다는 것을 확인할 수 있다. 이 방법은 다른 例와 비교할때 시료의 크기가 100以下인 것은 그 값이 精密하게 나타나므로 對數方眼紙를 使用하는것 보다 더욱 實用的인 表이다

이상으로 試料의 크기를 合理的으로 求하는 方法을 몇가지 說明하였는데. 어떠한 경우라도 適用시킬 수 있으니 여러분은 現在 自己가 使用하고 있는 시험이나 검사의 基準, 或은 規定上의 試料의 크기를 한번 檢討하여 보는 것이 좋을 것이다.

參考文獻

1. ASTM *Standard on textile materials*
   by American Society for Testing & materials 1962
2. 品質管理便覽 : 日本規格協會 編 1962.
3. 纖維技術者를 爲한 品質管理:徐丙三著 1968