品	質
管	理

第二講

品質管理의 統計的 基礎 (Ⅰ)

白光藥品株式會社 化學工場
品質管理委員會

第一講에서 品質管理 槪要라는 題目 아래 基礎的 定義 및 組織, 그리고 一般的 順序로서 標準, 試料, 檢査와 分析에 對해 일별하였다. 그러나, 品質管理의 中樞는 亦是 管理圖技法으로, 그 知識과 方法을 習得하는 것도 重要하지만, 그 基礎的 原理를 理解하는 것이 더욱 重要한 過程이라고 생각된다. 그래서, 以下 品質管

Table Ⅰ. 統計 Tool-kit

解決해야 할 問題	이에 쓰일 統計 Tool-kit
1. 表, 그림 및 數學的 方法을 써서 많은 量의 數値를 槪括하는 問題	Probability Distribution Statistical Parameter.
2. 두 Set의 Data 또는 한 Set의 Data와 標準値와의 意味의 差異를 決定하는 問題	Statistical Inference.
3. 特定한 危險水準에서의 假說을 評價하는데 必要한 Sample의 크기를 定하는 問題	Sample Size Determination.
4. 工程에서의 偏差가 어떤 Assignable Cause에 基因하는가를 早期 發見하는 問題 a. 品質이 合, 不合을 基準으로 測定될 때. b. 品質이 各 Item마다 測定値를 取할 때.	Control Charts.
5. 한 品質特性의 眞價(True Value)를 推算함에 있어 Sample에 依한 結果의 信憑度를 決定하는 問題	Confidence Limits.
6. 定해진 許用可能値에 對하여 製品의 品質을 評價하는 問題 a. 品質이 合, 不合을 基準으로(go-not-go-basis) 測定될 때 (Attributes). b. 品質이 各 Item마다 測定値를 取할 때(Variables).	Acceptance Sampling Plan.
7. 모든 問題가 最小의 試驗으로 解決되도록 試驗計劃을 세우는 問題	Design of Experiment.
8. 둘 혹은 그 以上의 標本平均値의 差異의 意味를 定하는 問題	Analysis of Variance.
9. 둘 以上의 變數 사이의 相關關係를 求하는 問題	Regression Analysis.
10. 中古裝置의 信憑를 推測하는 問題	Basic Reliability Formula.
11. 定해진 標準에 對하여 製品의 信憑度를 求하는 問題	Reliability Sampling Plans.

Table Ⅱ. Data 蒐集의 要點

問　　題　　點	考慮해야 할 Factor
1. Data 의 Type(類型)	計量値(Variables)로 測定할 것인가, 合, 不合의 判定(Attributes)으로 할 것인가. (計量値일 경우는 費用이 많이 들지만 많은 Information 을 준다)
2. Data 의 量	Sample size 는 結果値의 所望 正確度, 統計的 危險率, Data 의 變化度 및 測定의 精密度에 딸린다.
3. 測定	測定의 精密度는 所要 Sample size 와 結果値의 正確度에 影響을 준다.
4. 測定 順序	時間이 Key parameter 일 때나 特定한 結果의 原因이 어떤 計量値 때문인가를 찾는 것이 critical 한 問題일 때 測定順序는 매우 重要한 Factor 가 된다.
5. 合當한 Subgroup 의 選定	Subgroup 內의 各單位値들을 一定期間 동안 一團의 機械에서 無作爲로 選別할 것인가. 또는 한 機械에서 그 産出速度(output rate)에 따라 一定間隔으로 取할 것인가.
6. Data 의 分析方法	必要한 Type 의 Data 가 充分히 蒐集되야겠지만 그에 앞서 그 Data 를 分析하는데 쓰여질 formula 를 選定해 두어야 한다.

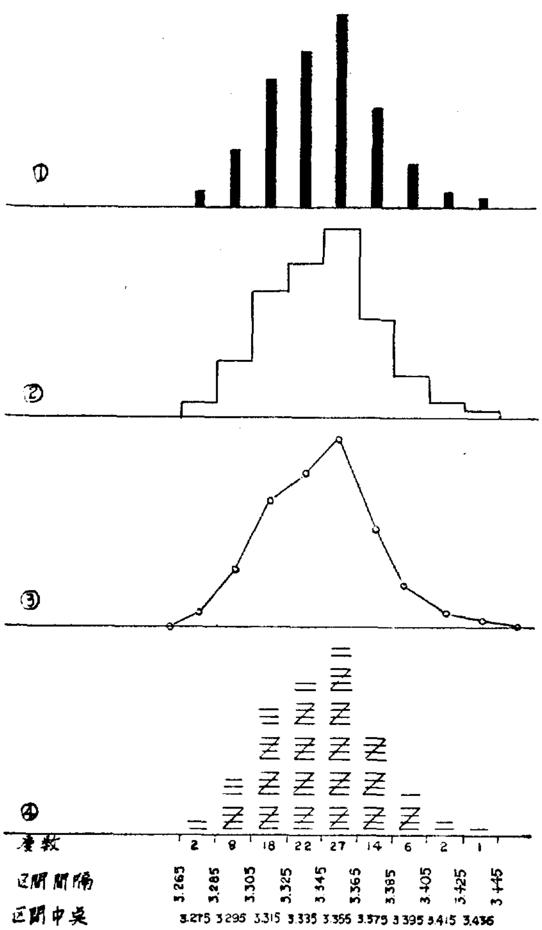

Fig. I. 度數分布의 回式的 表現에 例

理의 基本이 되는 統計의 諸問題를, 應用面을 念頭에 두면서 簡單히, 그러나 可能한 限 廣範圍하게 說明하려 한다. 普通 統計(Statistics)란 語彙는 두 가지 意味로 使用되는데, 그 하나는 單純한 數値資料(Numerical Data)란 意味요, 다른 하나는 이들 數値資料의 Collection, Organization, Analysis 및 Interpretation 을 取扱하는 學問(Science)이란 意味이다. 그러나 應用하는 사람의 立場에서 볼 때, 統計란 여러가지 問題를 解決하는 道具(tool)로 생각할 수 있다. 그런 뜻에서 Table Ⅰ에 品質管理에서 提起되는 諸問題와 그 問題에 應用되는 統計의 Tool-kit 를 실었다.

1. Data 의 蒐集

어떤 調査를 爲해 Data 의 蒐集을 計劃할 때는, 그 Data 를 分析하는데 어떤 統計方法을 使用할 것인가를 먼저 考慮하여야 한다. 그래서 Table Ⅱ에, Data 를 蒐集하기 前에 考慮하여야 할 여섯가지 Keypoints 를 要約하여 실었다. 實際로 Data 의 蒐集은 技術者와 統計 專門家가 서로 協同하여 實施하여야 한다.

2. 度數 分布表(Frequncy Tables)

Table Ⅲ과 같은 100 개의 Coil 의 電氣低抗(Ohms)을 測定한 Data 는 Table Ⅶ과 같이 度數로 再配列 表示하며, Table Ⅴ와 같이 小區間(Cell)으로 나누어 그 小區間의 度數로 表示할 수 있다. 이 度數 分布表는 모든 統計資料 整理의 基本이 되는 것으로, Fig. Ⅰ과 같이 여러가지 모양의 그림表로 表現된다.

3. 代表値(Measures of Central Tendency)

모든 測定値를 가장 잘 代表하여 주는 하나의 값은

Table Ⅲ. 100 코일의 抵抗

3.37	3.34	3.38	3.32	3.33	3.28	3.34	3.31	3.33	3.34
3.29	3.36	3.30	3.31	3.33	3.34	3.34	3.36	3.39	3.34
3.35	3.36	3.30	3.32	3.33	3.35	3.35	3.34	3.32	3.38
3.32	3.37	3.34	3.38	3.36	3.37	3.36	3.31	3.33	3.30
3.35	3.33	3.38	3.37	3.44	3.31	3.36	3.32	3.29	3.35
3.38	3.39	3.34	3.32	3.30	3.39	3.36	3.40	3.32	3.33
3.29	3.41	3.27	3.36	3.41	3.37	3.36	3.37	3.33	3.36
3.31	3.33	3.35	3.34	3.35	3.34	3.31	3.36	3.37	3.35
3.40	3.35	3.37	3.35	3.35	3.36	3.38	3.35	3.31	3.34
3.35	3.36	3.39	3.31	3.31	3.30	3.35	3.33	3.35	3.31

무엇일가를 나타내는 것으로, 算術平均, 幾何平均, 調和平均, 代數平均 等의 平均値(Averages or Means)와 中位數(Median), 그리고 最頻數(Mode) 等이 있는데, 一般으로 算術平均이 가장 많이 쓰이는 것으로, 아래와 같이 數式化할 수 있다.

$$X=\frac{\Sigma f_i x_i}{\Sigma f_i} \quad \text{or} \quad X=\frac{\Sigma x_i}{n} \qquad (1)$$

4. 散布度(Measures of Dispersion)

代表値를 中心으로 Data가 흩어져 있는 範圍(程度)를 散布度로 表示하는데, 平均偏差, 標準偏差, 分散(Variance), 範圍(Range) 等이 있다. Data가 적을 때는, 最大値에서 最小値를 뺀 差로 表示하는 Range를 使用하기도 하나, 普通은 標準偏差를 쓰며,

$$S=\sqrt{\frac{\Sigma(x-x_i)^2}{n}}=\sqrt{\frac{\Sigma x^2}{n}} \qquad (2)$$

으로 計算한다. Variance는 後述되는 Variance Analysis에서 母集團에 對한 몇 個의 Data群 (例컨데, 같은 工程에서 나온 몇 個의 月間 Data 들)을 比較評價할 때 利用되는 것으로, 標準偏差의 自乘으로 計算한다.

要컨데, 어떤 Data가 度數로 整理되어 그 分布狀態를 檢討할 때, 그 分布形態를 適切히 表現해 주는 값 ──統計値는 代表値와 散布度인데, Data가 적을 때 結果値의 正確度에 比한 計算의 煩雜을 避하기 위해 平均値와 範圍(Range)로 나타내기도 하지만, 一般으로 平均値 X와 標準偏差 S로 表示한다. 그러므로, 이 두 값의 槪念은 統計──品質管理에 있어 매우 重要한 意味를 갖는다.

平均値는 水平座標에서 度數分布度의 位置를 指示하여 준다. 또 平均値를 中心으로 集中하려는 傾向, 또는 物理學에서의 무게 中心(重心)과 같은 槪念이라고도 볼 수 있을 것이다.

標準偏差의 槪念은 한층 어렵다. 한 方法은 標準偏差를 各個 偏差의 2次 平均(Quatratic Mean)으로 생각하는 것이다. Table Ⅳ의 8個의 偏差를 各各 8個의

Table Ⅳ. 100 coil의 抵抗値

Resistance ohms	Tabulation	Frequency
3.45		
3.44	|	1
3.43		
3.42		
3.41	||	2
3.40	||	2
3.39	||||	4
3.38	卌 |	6
3.37	卌 |||	8
3.36	卌 卌 |||	13
3.35	卌 卌 ||||	14
3.34	卌 卌 ||	12
3.33	卌 卌	10
3.32	卌 ||||	9
3.31	卌 ||||	9
3.30	卌	5
3.29	|||	3
3.28	|	1
3.27	|	1
3.26		
Total.		100

Table Ⅴ. 抵抗値의 度數分布表

Resistance	Frequency
3.425—3.445	1
3.405—3.425	2
3.385—3.405	6
3.365—3.385	14
3.345—3.365	27
3.325—3.345	22
3.305—3.325	18
3.285—3.305	8
3.265—3.285	2
	100

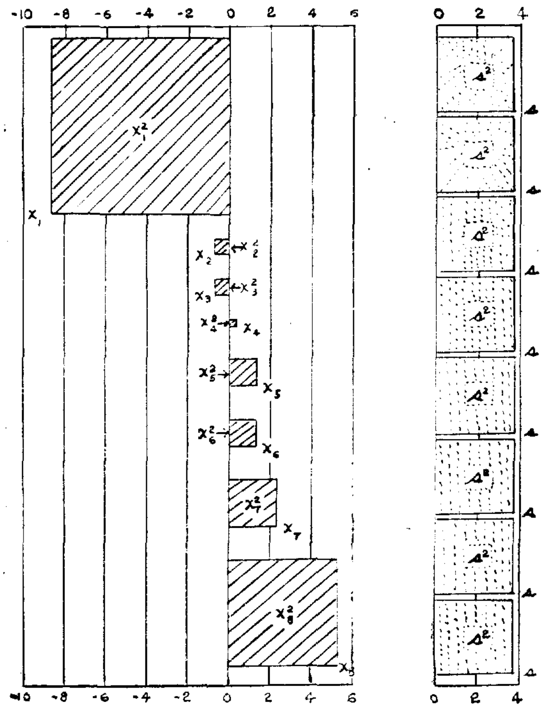

Fig. Ⅱ. 標準偏差의 幾何學的 表現
(Table Ⅵ 에서)

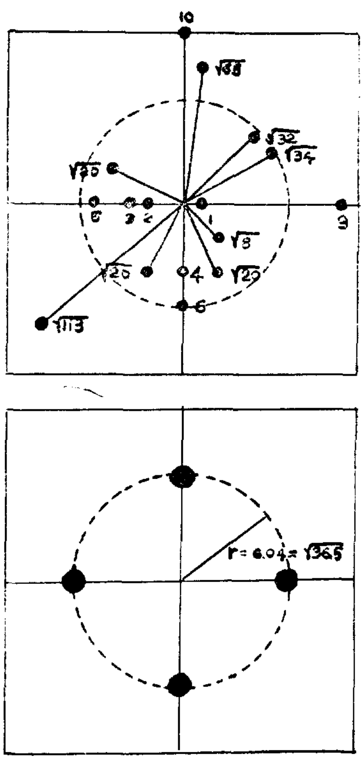

Fig. Ⅲ. 무게中心과 Gyration 半徑

Table Ⅵ.

Specimen	Breaking strength (pounds) X	Deviations from mean $x=X-\overline{X}$	x^2
X_1	57	−8.75	76.5625
X_2	65	−0.75	0.5625
X_3	65	−0.75	0.5625
X_4	66	0.25	0.0625
X_5	67	1.25	1.5625
X_6	67	1.25	1.5625
X_7	68	2.25	5.0625
X_8	71	5.25	27.5625
	526 $\overline{X}=65.75$	0.00* 0.00	113. 500. $S=14.1875$

*$\Sigma\,|x|=20.50$

正方形의 一邊의 길이라고 하면, 標準偏差는 이들 正方形의 平均面積을 가지는 正方形의 一邊의 길이와 같다(Fig. Ⅱ 參照). 한편 標準偏差를 物理學에서의 Gyration 半徑과 같은 槪念으로 생각할 수도 있다. 說明은 Fig. Ⅲ으로 代한다. 또 다른 가장 效果的인 생각은 어떤 度量의 尺度로 보는 것으로, 定規曲線에서는 平均値에서 이 定規曲線의 變曲點까지의 水平距離를 나타낸다. 定規曲線은 Fig. Ⅳ에 그려 놓았다.

定規曲線은 뒤의 分布型 項에서 詳述하겠지만, 大部分의 管理工程에서의 標本의 分布는 定規曲線과 같거나, 이에 近似하게 나타난다. 우선 쉽게 定規曲線이란 管理의 理想狀態에서의 分布曲線이라고 생각하자.

전혀 다른 두 개의 對象을 比較함에 있어서, 이들 Data 를 다음과 같은 標準化된 값으로 環元하여 比較하면 좋다.

$$Z=\frac{X-\overline{X}}{S} \tag{3}$$

但 $\overline{X}$: 平均値
　S: 標準偏差

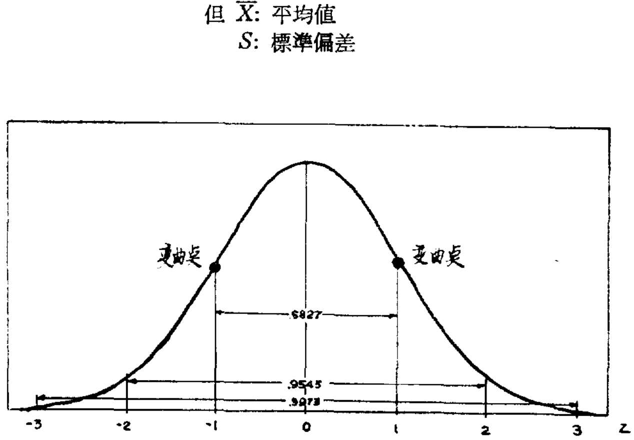

Fig. Ⅳ. 定規曲線

Table Ⅶ.

Z	$\mu \pm Z\sigma$ 範圍 內에 들어갈 確率
0.6745	0.5
1	.6827
1.96	.95
2	.9545
2.576	.99
3	.9973
3.291	.999

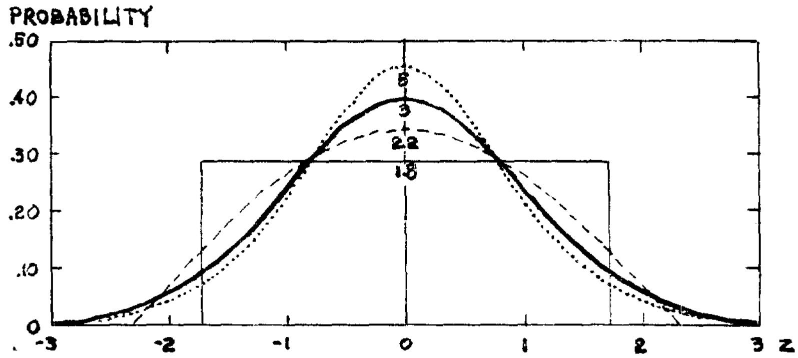

Fig. Ⅵ. α_4와 對應하는 分布曲線

定規曲線은 이 Z와 X가 그 값(Z)을 나타낼 確率을 橫 및 縱座標로 하여 構成한 것이다. 그러므로 定規曲線은 여러 分布를 比較檢討하는 標準으로 使用할 수 있음을 알 수 있다. 定規曲線에서 $\mu \pm Z\sigma$(μ와 σ는 母集團의 平均値와 標準偏差) 範圍 內에 들어갈 確率을 몇 個 Table Ⅶ에 싣는다. 例를 들어 表에서 Z가 1이란 말은, 分布 中에서의 어떤 값에서 平均値($Z=0$인 點)까지의 거리가 標準偏差와 같은 點은 平均値를 中心으로 두 點이 있는데, 이 點에 測定値가 들어 갈, 即 어떤 測定値에서 平均値를 뺀 값이 標準偏差보다 작거나 같을 確率은 0.6827이란 말이다.

5. 分布形을 說明하는 統計値

普通 分布의 形(Form)은 Skewness와 Kurtosis로 表示한다. 이들 값은 品質管理에서 매우 興味있는 意味를 갖는데, 그 理由는 代表値나 散布度는 分布의 모양에 크게 左右되며 測定値가 標準範圍 內에 들어 갈 確率도 一部는 分布의 모양에 依持하기 때문이다.

Skewness는 對稱에서의 偏歪의 程度를, Kurtosis는 曲線의 屈曲의 程度――最頻點에서의 Curvature를 뜻한다.

曲線은 leptokurtic(屈曲이 甚한 것), platykurtic(平偏한 것), mesokurtic(中間程度의 것) 等으로 屈曲의 程度를 表現한다.

Relative Skewness는 α_3, Relative Kurtosis는 α_4로 表示하며, 다음 式으로 計算한다.

$$\alpha_3 = \frac{\Sigma Z^3}{n} \tag{4}$$

$$\alpha_4 = \frac{\Sigma Z^4}{n} \tag{5}$$

但, $n = \Sigma f_i$, $Z = \frac{X - X_i}{S}$

Fig. Ⅴ, Ⅵ을 보면 α_3, α_4의 各各의 값과 曲線의 모양과의 關係를 알 수 있을 것이다. 要컨데 α_3가 零이면 對稱이고, 零보다 크면 왼 쪽으로, 작으면 오른 쪽으로 기우러 졌음을 알 수 있다. α_4는 크면 클수록 leptokurtic, 작으면 작을수록 platykurtic이 된다. 또 한 가지 重要한 것은 定規曲線의 α_3는 0, α_4는 3이란 事實로서, 普通의 Chemical Process에서는 그들 分布가 正常的으로 管理될 때 定規曲線을 나타내므로, 管理의 狀態를 이들 α_3와 α_4가 定規曲線의 그 것, 即 $\alpha_3=0$, $\alpha_4=3$에 어느 程度 接近하는 가를 봄으로서 大略 알 수 있다는 것이다.

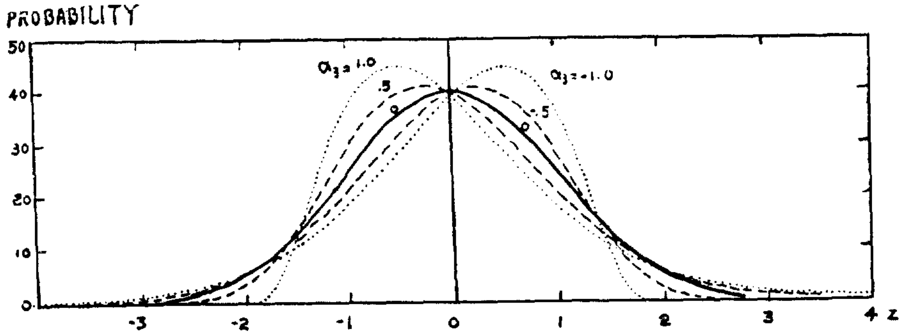

Fig. Ⅴ. α_3와 對應하는 分布曲線

여기서 文字 使用上의 混沌을 避하기 위해, 測定値에서 統計値(statistics)는 Alphabet, 母集團의 統計値(parameter)는 對應하는 希臘 文字를 使用하기로 하고 지금 까지의 것들을 Table Ⅷ에 整理한다. 그리고 參考로 本 會社에서 이들을 計算하는데, 使用하는 Work Sheet를 Fig Ⅷ에 실어둔다.

6. 確率 初步

品質管理에 使用되는 統計方法은 確率에 基礎를 두고 있다는 意味에서, 初步이나마 그 原則을 整理해 두기로 한다.

完全한 주사위를 無作爲로 던졌을 때, 1이 나올 確率은 $\frac{1}{6}$이다. 이때, ① 나오는 數字는 반드시 1, 2, 3, 4, 5, 6 中 어느 하나는 반드시 나오며, ② 만일 1이 나왔으면 다른 數字는 絶對로 나올 수 없고, ③ 1에서 6까지의 數字中 어느 것이든 等可能으로 나온다. 이런 것을 前題로, 確率을 다음과 같이 定義할 수 있다.

定義 I. 한 번의 試行 中에서, n個의 事象 中 어느 하나는 반드시 일어나고, 또 어느 하나가 일어나면 다른 것은 絶對 일어나지 않으며, 等可能으로 期待될 때, 하나의 事象 E가 n個의 等可能한 경우 중 r個의 경우에 對해서 일어나면 事象 E가 일어날 確率은 $\frac{r}{n}$이다.

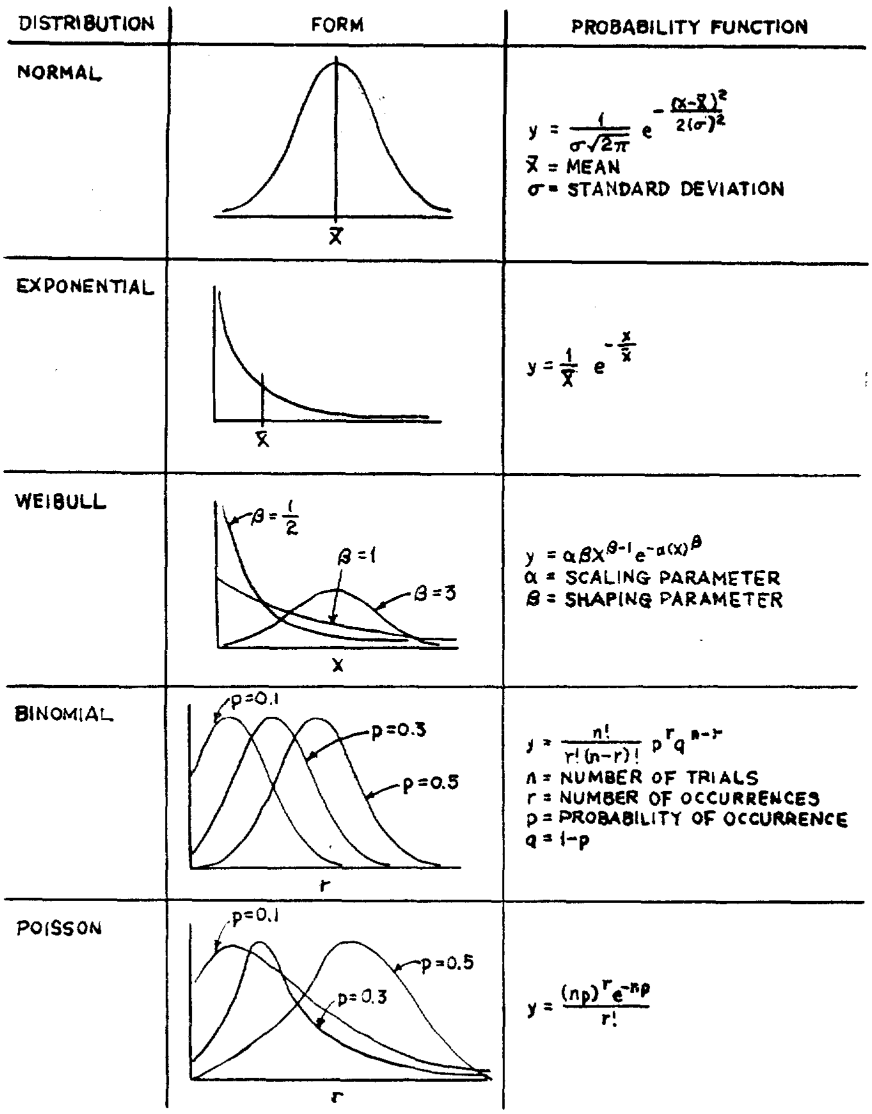

Fig. Ⅶ. 分 布 型

定義 Ⅱ. 어떤 種類의 偶然的 事象 E가 몇 번이고 反復하여 試行할 수 있고, 各 回의 試行 結果는 相互 影響을 미치지 않는다고 하자. 이때 n_1回의 試行 中 E가 r_1回 일어나고, 또 다른 n_2回의 試行 中 E가 r_2回 일어나고, 이것을 多數回 反復할 때

$$\frac{r_1}{n_1},\ \frac{r_2}{n_2}\cdots\cdots$$

의 數列이 一定數 P에 接近하면, 事象 E가 일어날 確率은 $P_1(E)=P$로 定義된다.

이렇게 定義해 놓고 보면, 이들 定義는 確率의 意味를 미리 알아야만 理解될 槪念을 써서 이루어진 것이라는 事實을 알게 된다. 그러나 確率의 存在를 假定치 않았을 때, 統計方法은 無益한 것이 되고, 品質管理도 無意味하게 된다. 아마도 確率이란 語彙를 定義하지 않은 채 그 意味를 알아차려 두고 確率의 몇 가지 定理를 理解하여 두는 것이 나을 것이다.

몇 개의 確率 定理를 들어 보면,

a) 餘事象의 確率

주사위를 던질 때 1의 눈이 나오는 事象을 E라고 하고, 1의 눈이 나오지 않는 事象을 餘事象이라고 하며, E가 일어나지 않을 確率은,

$$P_r(E\text{가 일어나지 않을})=1-P_r(E) \tag{6}$$

이다.

b) 排反事象의 確率의 加法

2個의 事象 E_1과 E_2가 그 中 어느 하나가 일어나면 다른 한 事象은 일어나지 않을 때, 事象 E_1과 E_2는 서로 排反的(Exclusive)이라고 한다. 지금 두 개의 事象 E_1과 E_2가 서로 排反的이라고 하면, 事象 E_1이든가 E_2든가 어느 쪽이든 일어나는 確率은,

$$P_r(E_1 \text{ or } E_2)=P_r(E_1)+P_r(E_2) \tag{7}$$

이다.

c) 獨立事象의 確率의 乘法

두 개의 事象 E_1과 E_2가 相互 獨立的이라면(即 E_1이 E_2에, 또는 E_2가 E_1에 아무 影響을 주지 않는다면), E_1과 E_2가 同時에 일어날 確率은,

$$P_r(E_1 \text{ and } E_2 \text{ both})=P_r(E_1)\cdot P_r(E_2) \tag{8}$$

이다.

7. 分布型(Types of Distribution)

經驗에 依하면 大部分의 工業 Data는 몇 개의 標準確率分布 中 어느 하나에 거의 一致한다. 이와같은 標本分布는 全的으로 理論的 觀點에서 探求된 것이며, 그들의 性質은 實際的인 應用을 爲해서 Table로 表現되어 있다. Fig. Ⅶ에 一般으로 자주 應用되는 몇 개의 理論分布를 要約하였다.

Table Ⅷ.

	測定値에서의	母集團의
平　　均	$\bar{X}$	μ
標 準 偏 差	S	σ
Relative Skewness	a_3	α_3
Relative Kurtosis	a_4	α_4

a) 正規分布(Normal Distribution)

連續分布 中에서, 가장 重要하고 널리 通用되는 分布型이다. 理論的으로 생각할 때, 正規分布는 問題의 測定値가 여러가지 原因의 結果일 때(例를 들어 濃度, 溫度, 壓力 等 여러 要因에 依한 反應으로 製造된 製品의 品質) 많이 볼 수 있다. 品質管理에서 나타나는 많은 Data는 Dimension 혹은 다른 測定 可能의 特性으로 이루어져 있으며, Control되는 狀態에서는 거의 近似的으로 正規分布를 나타낸다.

正規分布의 確率密度는,

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{9}$$

로 나타나며, 平均値 μ, 標準偏差 σ의 正規分布를 $N(\mu,\sigma^2)$로 略記할 때도 있다. 正規分布의 몇 가지 性質을 적어보면,

① 對稱이고 鍾形이며, 兩端은 橫座標軸에 漸近한다. (Fig. Ⅳ 參照)

② 正規分布는 $\alpha_3=0$, $\alpha_4=3$이다.

③ 많은 分布는(例, 二項 分布) 標本의 크기가 커짐에 따라 正規分布에 接近한다. 實際로 管理圖(Control Charts)는 正規分布에 根據를 두고 있는 것으로, 平均値 管理圖는, 비록 各個 觀測値의 分布는 正規가 아니라도 平均値의 分布, 例를 들어 日間 Data의 平均値로 이루어진 月間 Data는, 正規 혹은 이에 매우 近似하다는 것을 前題로 하여 만들어 지는 것이다. 적은 標本으로 分析할 때는, 標本結果値는 普通 數學的으로 t-分布(後述)로서 說明되는 한 Index로 轉換된다. t-分布는 모양에 있어 正規分布와 비슷하며 標本의 크기가 無限大일 때 正規分布와 一致한다.

b) 指數分布(Exponential Distribution)

이는 電子器機의 部分品에 對한 Data를 分析하는데, 그리고 複雜한 裝置의 故障模型을 說明하는데 有用하다. 例를 들어, 大部分의 電子部品의 Life Characteristics(壽命特性)와 여러 裝置의 無故障運轉期間의 分布(Distribution of "Time Between Failures")는 指數分布를 따른다.

指數分布의 實際的인 應用을 爲해 Reliability 槪念을 생각해 보자. AGREE(Advisory Group on Reliability of Electronic Equipment)의 定義에 따르면, Reliability 는 한 裝置가 特定한 期間 동안 주어진 條件下에서 자기의 任務를 事故 없이 遂行할 수 있는 確率을 말한다. 例를 들면, 彈頭誘導彈이 地上의 通常 氣候條件下에서 한 時間 동안 故障 없이 運轉되어 目標地點 1,000 ft 以內에 適中할 Reliability 는 0.8 이다.

經驗上 한 System 의 事故模型(Failure Patterns)은 相互 類似하다는 것을 우리는 안다.

勿論 實際的인 面에서 볼 때, 한 裝置가 故障 없이 運轉되는 期間이 主觀心事일 것이다. 一般으로 故障率(Failure Rate)이 一定할 때, TBF(Time Between Failures)의 分布는 指數分布를 따른다. 이때 Reliability 를 說明해 주는 分布函數는 다음과 같다.

$$P_s = R = e^{-\frac{t}{m}} = e^{-\lambda t} \tag{10}$$

但, $P_s = R$: Reliability

$e = 2.718$

m: MTBF(Mean Time Between Failures)

$\lambda = \dfrac{1}{m}$: 故障率(Failure Rate)

例를 들어서, 平均 50 日 만에 故障이 나는 眞空 pump 를 30 日間 無事故로 運轉할 수 있는 確率은 $e^{-\frac{30}{50}} \fallingdotseq 0.541$ 이다.

以上에서 본 바와 같이, 어떤 裝置의 平均補修間隔(MTBF)을, 다시 말해서 平均故障率을 알고 있으면, 그 裝置의 無事故運轉期間을 說明해 주는 Reliability 를 計算할 수 있는데, 그 分布는 指數分布를 이루는 것이다.

c) Weiball 分布

確率密度가 $y = \alpha\beta x \times {}^{\beta-1} e^{\alpha} x^{\beta}$ 로 表示되는 數學式을 基礎로 하는 것으로, Curve 의 모양은 Scaling Factor α 와 Shaping Factor β 에 따라 定해진다. 이 分布는 電子管의 壽命特性(Life Characteristics), Ball Bearing 의 疲勞壽命(Fatigue Life), 또는 Milk 瓶의 壽命等의 分布에 適用되나 一般的인 것이 아니므로 자세한 說明은 省略한다.

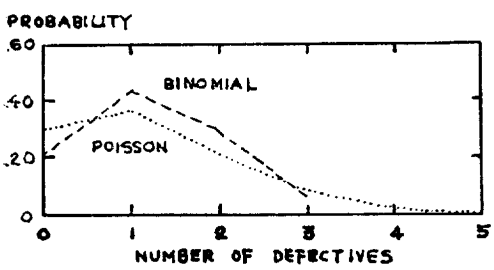

Fig. Ⅷ. 二項(n=3, p=1.4) 및 poisson(np=1.2) 分布의 比較

d) 二項分布(Binomial Distribution)

지금 하나의 주머니에 검은 구슬 A 개와 흰 구슬 B 개, 합계 N 개의 구슬이 들어 있다고 하자. 이 주머니에서 無作爲로 하나의 구슬을 꺼내어, 검은 구슬이 나올 確率은 $\frac{A}{N} = P$, 흰 구슬이 나올 確率은 $\frac{B}{N} = q = 1 - P$ 이다. 이 주머니에서 한 개를 無作爲로 꺼내서 검은가 흰가를 確認한 후, 도로 집어 넣고 잘 흔들어 섞어논 다음, 다시 한 개를 꺼내 黑白을 確認하고, 다시 집어 넣고, 하여 n 번 反復, r 回 黑이 나올 確率은

$$P_r(r) = \binom{n}{r} p^r q^{n-r} \tag{11}$$

$$但, \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

이다.

式 (11)은 二項整理에서 r 번째 項을 나타내는 것임을 우리는 쉽게 알 수 있다.

二項分布는, 앞의 例에서처럼, 母集團에서 오직 두 가지의 事象만을 取할 수 있고, 그 確率이 一定할 때 連續抽出에 依한 標本의 平均値의 分布는 二項分布이다.

例컨데 標本의 크기에 比해 Lot 의 크기가 非常히 큰 경우의 拔取檢査에서, Lot 가 合格하는 確率을 計算할 때 쓰인다. 이는 合, 不合의 두 事象만이 있고, Lot 中에서 標本을 拔取해도 Lot 의 不良率 P 는 거의 變하지 않는다고 볼 수 있으므로, 二項分布를 쓸 수 있는 것이다. 二項分布를 다른 말로 表現하면 確率 P 인 事象이 n 번의 試行에서 r 번 發生할 確率을 나타낸다고 할 수 있다.

이 分布의 모양은 n 와 P 의 값에 따라 定해지는데 (Fig. Ⅷ 參照), 不良率管理圖를 運用할 때, 또는 標本抽出計劃을 展開하려고 할 때 쓰인다(後述).

一般으로 離散型分布에서는 平均 μ 는,

$$\mu = X_1 P_r(X_1) + X_2 P_r(X_2) + \cdots\cdots + X_k P_r(X_k) = \Sigma X_i P_r(X_i) \tag{12}$$

分散은

$$\sigma^2 = (X_1-\mu)^2 P_r(X_1) + (X_2-\mu)^2 P_r(X_2) + \cdots\cdots + (X_k-\mu)^2 Pr(X_k) = \Sigma (X_i-\mu)^2 P_r(X_i) \tag{13}$$

으로 주어진다. 이 公式을 써서 二項分布의 平均値, 分散(Variance) 及 標準偏差를 求하면,

平　均: np

$x_1 \sim x_2$	x_i	f_i	$x_i f_i$	$x_i - \bar{x}$	$(x_i-\bar{x})^2$	$(x_i-\bar{x})^2 f_i$	z	z^2	z^3	z^4
TOTAL										
	$\bar{x}=$			$S(x)=$			$t=$			

$z^3 f_i$	$z^4 f_i$	$\frac{f_i}{N}$	$f(z)$	$\frac{f_i}{N}-f(z)$	$\left(\frac{f_i}{N}-f(z)\right)^2$	$\frac{\left(\frac{f_i}{N}-f(z)\right)^2}{f(z)}$	R	Remark
$\alpha_3=$		$\alpha_4=$		$\chi^2=$			$R=$	

Fig. Ⅸ. 品質管理計算表

標準偏差： $\sqrt{npq}$

分　　散： npq

이다.

e) Poisson 分布

標本의 크기 n 가 커지는 反面 母集團에서의 不良率 p 가 적어지고, np 가 變化하지 않을 때, 二項分布는 다음과 같은 limiting form 을 갖는다.

$$P_r(r)=y=\frac{(np)^r}{r!}e^{-np} \qquad (14)$$

이것을 Poisson 分布라고 한다(Fig. Ⅴ參照). Possion 分布는 品質管理에 있어서 두 가지 型의 Data 를 分析하는데 쓰인다. 卽, ① n 가 크고 p 가 작은 不良率管理에, ② Unit output 當의 不良數管理에. ①에 있어서는 二項分布에 對한 近似計算에 쓰이는데, n 가 充分히 크고 p 가 充分히 작다면, 實際的인 計算에 큰 誤差를 주지 않는다. 參考로 $n=3$, $p=0.4$ 일 때의 二項 및 Poisson 分布를 Fig Ⅷ에 表示하였다. ②를 좀 더 분명히 하기 위하여, 不良率 p, 代身에 不良數 c 를 使用하고. 不良率 p 의 平均値를 意味하는 값 m 를 $m=E(p)=np$ 라고 하면, ($E(p)$란 p 의 期待値란 뜻) $m=E(c)$가 되고 式 (14)는,

$$P_r(c)=\frac{m^c}{c!}e^{-m} \quad \cdots\cdots\cdots\cdots\cdots (15)$$

이 된다.

그러면 不良數에 關한 式으로 轉換된다.

Poisson 分布의 경우 平均値, 分散, 標準偏差는,

平　　均： $m(=pq)$

標準偏差： $\sqrt{m}$

分　　散： m

以上에서 몇가지 分布를 들었는데, 大部分의 現場 Data 가 이들 分布中 어느 하나에 거의 一致하지만, 實際로는 二項分布와 正規分布로서 大部分의 Data 를 分析整理할 수 있다. 參考로 本 會社의 鹽酸製品의 HCl 濃度에 對한 確率分布圖와 이와 같은 平均値와 標準偏差를 가지는 正規曲線을 Fig. Ⅸ에 싣는다.

後述되겠지만 大部分의 平均値管理圖는 正規分布, 不良率管理圖는 二項分布를 基礎로 하여 構成되는 것이다.

(次號에 繼續)