

방직공장의 QC. Engineer를 위한 품질관리(4-1)

3. 통계적 방법(1)

3.1 통계학이란?

자연계에 있는 여러 가지 현상가운데에 무질서하게 흩어져 있는 것을 질서있게 체계적으로 관찰하여 그 결과를 수량화시켜 연구하고자 하는 현상을 추측하는 것이 통계학이다.

즉 모집단의 전체현상을 대상으로 소량의 샘플(시료)을 추하고 그 샘플이 제시하는 수량화된 수치를 기초로 하여 모집단전체의 상태를 추측하는 것이다.

우리가 분석할때나 측정할 때 얻어지는 데이터는 시료에 대한 데이터이지 모집단전체에 대한 데이터가 아니다. 따라서 모집단(공정이나 로트)에 대한 성질을 정확히 알 수 있도록 데이터를 분석할 필요가 있고 모집단과 시료사이의 관계를 계량적으로 파악하여야 된다.

이와 같이 모집단과 시료사이의 계량적 역은 계수적 관계를 연구하는 학문이 통계학이다.

통계학은 추측과 계획의 과학이며 추측통계학이라고도 한다. 이것을 추계학이라고도 한다.

19세기 이전에는 주로 생물학, 국세조사, 의학등에 이용되었으나 20세기에 접어들면서 공업분야에 응용되기 시작하여 오늘날에는 통신, 전기, 화학, 섬유, 기계, 토목 등 거의 모든 공업분야에 유효하게 응용되고 있으며 특히 통계적 품질관리의 강력한 시기로 등장하였다.

3.2 모수와 통계량

모집단은 그 중심되는 위치와 흩어짐을 나타내는 어떤 수치로 표시된다. 이처럼 모집단의 흩어진 성질을 계량적이나 계수적인 수로 표시한 것을 모수라고 하며 모수는 그 중심의 위치를 나타내는 모평균과 그 흩어짐을 나타내는 모분산으로 표시된다.

이와 반대로 모집단으로부터 얻어진 시료의 중심위치를 표시하는 시료평균치와 그 흩어짐을 표시하는 시료분산을 통계량이라고 하여 모수와 구별한다. 일반적으로 이들을 사용하기 편하도록 기호로 표시했는데 그 기호와 읽는 법은 제1표와 같다.

<표 1>

구 분 표시수치	모 수			통 계 량		
	기 호	읽 는 법		기 호	읽 는 법	
중심을 나타내는 값	M	뮤	모평균	$\bar{x}$	엑스바	평균치
흩어짐을 나타내는 값	δ	시그마	모표준편차	$\hat{\delta}$	시그마하아트	표준편차
	δ^2	시그마제곱	모분산	$\hat{\delta}^2$	시그마제곱하아트	분산

주[4] 시료의 표준편차 $\hat{\delta}$ 는 “불편분산의 제곱근”이라고도 하며 시료의 분산 $\hat{\delta}^2$ 은 불편분산이라고도 하며 V로 표시하기도 한다.

3.2.1 평균치의 계산법

평균치는 산술평균치의 준말이며 모집단 역은 시료의 흩어짐의 중심을 표시하는 값이다.

이 평균치를 계산하는데는 개개의 측정치의 개수로 나누는 직접법과 임의

의 값을 미리 정하여 이 값과 각 측정치와의 차를 구하여 그 차의 평균을 임의의 값에 더하여 계산하는 변환법이 있다.

(1) 직접법

순 서	방 법	계 산
1	측정치를 기록한다.	62.5 60.6 59.8 57.6 60.5
2	측정치를 전부 더한다.	$62.5+60.6+59.8+57.6+60.5=301$
3	측정치의 합계를 측정개수로 나누어 평균치를 구한다.	$301 \div 5=60.20$

위의 방법과 같이 계산하는 것이 직접법이다. 그러나 이것을 공식으로 표현하려면 숫자로서는 부적당하므로 일반화된 공식으로 나타내기 위하여서는 숫자 대신 문자로 표시해야 한다.

그러면 숫자와 문자를 연관시켜 평균치를 구하는 공식을 만들어 보자.

순 서	방 법	계 산	문 자 표 시
1	측정치를 기록한다.	62.5 60.6 59.8 57.6 60.5	$X_1 X_2 X_3 X_4 X_5$
2	측정치를 전부 더한다.	$62.5+60.6+59.8+57.6+60.5=301$	$X_1+X_2+X_3+X_4+X_5=\sum_{i=1}^5 x_i (5)$
3	측정치의 합계를 측정개수로 나누어 평균치를 구한다.	$301 \div 5=60.20$	$\frac{\sum_{i=1}^n x_i}{n} \bar{x}$

표[5]는 $\sum$ 는 총합계를 의미하는 것으로 그리스 문자이며 시그마라 읽는다. 그러나 시그마는 표준편차에도 있으므로 혼동되기 쉬워서 썸(sum)이라고 읽기도 한다.

$\sum_{i=1}^n x_i$ 는 x라는 측정치를 첫번째부터 n번째까지 전부 더했다는 것을 의미한

다.

N은 측정개수를 의미하므로 평균치 $\bar{x}$ 는 $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ 라는 일반식으로 표시하

게 된다.

(2) 변환법

順序	方 法	計 算	文 字 表 示
1	測定値를 記錄한다	62.5 60.6 59.8 57.6 60.5	x_1, x_2, x_3, x_4, x_5
2	假平均을 定한다 假平均은 될수있는데로 測定値의 中心이 되는 數로서 整数이어야 한다	假平均을 60으로 定한다	$x_0 = 60$ x_0 = 假平均의 記号이다
3	各 測定値에서 假平均 을 빼고 그 값에 10 을 곱한다 <u>10을 곱하는 理由</u> 測定値에서 假平均을 빼면 그 答이 小數點 以下 한 자리까지 나 오므로 計算上 煩雜하 여 10곱하여 整数로 만들어 計算하면 簡便 하다	$(62.5 - 60) \times 10 = 25$ $(60.6 - 60) \times 10 = 6$ $(59.8 - 60) \times 10 = -2$ $(57.6 - 60) \times 10 = -24$ $(60.5 - 60) \times 10 = 5$	$(x_1 - x_0) \times h = x_1$ $(x_2 - x_0) \times h = x_2$ $(x_3 - x_0) \times h = x_3$ $(x_4 - x_0) \times h = x_4$ $(x_5 - x_0) \times h = x_5$
4	(3)에서 구한 答을 全 部 더하여 平均을 求 한다	$\frac{25+6+(-2)+(-24)+5}{5}$ $= \frac{10}{5} = 2$	$\frac{x_1+x_2+x_3+x_4+x_5}{n}$ $= \frac{\sum_{i=1}^n x_i}{n} = \bar{x}$
5	(4)에서 求한 平均値는 變換値이므로 0.8을 10으로 나누어 假平 均 50에 더해주어야 眞平均이 나온다	$60 + \frac{2}{10} = 60.2$	$\bar{x} = x_0 + \frac{1}{h} \times \bar{x}$ $= x_0 + \frac{\bar{x}}{h}$

변환법으로 구한 값도 직접법으로 구한 것과 똑 같은 것을 위의 본보기에서 알 수 있다. $\bar{x}$ 는 변환치라고 부르며 측정치 x 와 구별된다.

변환치를 사용하여 평균치를 쉽게 구하는 방법을 변화법 역은 변수 변환이라고 한다.

3.2.2 분산의 계산법

흠어짐의 정도를 표시하는 방법으로 표준편차(Standard Deviation)와 분산(Variance)이 있다.

표준편차나 분산은 한 개 한 개의 측정치가 평균치나 중심치에서 얼마나 떨어져 있는가? 그 거리를 집약된 숫자로 표시한 것으로 그 값의 크기로 흠어짐의 정도를 알 수 있게 한 것이다.

지금까지 우리는 모든 것을 생각할 때 흠어짐보다는 시료의 중심적 경향을 나타내는 평균치를 더 많이 사용하였으며 또 평균치를 비교하여 크고 작은 것을 판단하여 왔다.

그러면 왜 표준편차나 분산을 즐겨 쓰지 않았을까? 거기에는 그만한 이유가 있다. 표준편차나 분산의 계산방법이 조금 복잡하기 때문에 어려운 표준편차는 계산하기 싫어하고 그 반면에 간단한 평균치만 즐겨 썼던 것이다.

그러나 중심적인 움직임만을 알고서는 도저히 균일한 품질을 만들기 위한 관리에는 효과를 얻을 수 없으며 반드시 흠어짐을 나타내는 표준편차와 함께 표시되어야만 비로소 훌륭한 판단을 내릴 수 있을 것이다.

특히 20세기초부터 통계학이 발달되어 표준편차의 계산방법중의 한가지인 아주 쉬운 범위법이 발견되어 모든 관리에 많은 도움을 주고 있다.

(가) 편차의 기하학적 뜻

지금 다음과 같은 2개의 측정치가 있다고 하자.

A	62.5	60.6	59.8	57.6	60.5
B	63.2	61.3	59.1	57.3	60.1

(A)와 (B)의 평균치를 구하여 보면

$$(A) \frac{62.5 + 60.6 + 59.8 + 57.6 + 60.5}{5} = \frac{301}{5} = 60.2$$

$$(B) \frac{63.2 + 61.3 + 59.1 + 57.3 + 60.1}{5} = \frac{301}{5} = 60.2$$

로써 (A)와 (B)는 똑같다.

평균치만 본다면 (A)나 (B)가 똑같으니까 우열을 가리기 어렵다.

그러므로 흠어짐을 살펴보면 차이를 알 수 있으므로 측정치와 평균치와의 차를 구하여 보자.

$$(A) 62.5 - 60.2 = 2.3$$

$$60.6 - 60.2 = 0.4$$

$$59.8 - 60.2 = -0.4$$

$$57.6 - 60.2 = -2.8$$

$$60.5 - 60.2 = 0.3$$

$$(B) 63.2 - 60.2 = 3.0$$

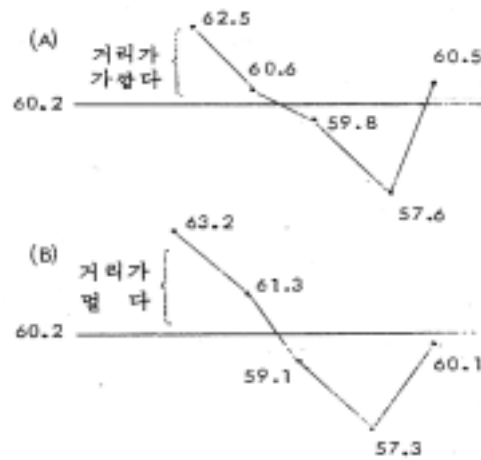
$$61.3 - 60.2 = -1.1$$

$$59.1 - 60.2 = -2.9$$

$$57.3 - 60.2 = -2.9$$

$$60.1 - 60.2 = -0.1$$

위의 (A)(B) 각 측정치와 평균치의 차가 평균치와 각 측정치와의 거리가 되며 이것을 편차라고 한다. 이 편차를 평균치에 대한 위치로 표시하여 보면 그림 4와 같다.



(A) 평균치는 같으나 흩어짐이 작은 경우

(B) 평균치는 같으나 흩어짐이 큰 경우

<그림 4>

여기서 우리는 비록 평균치가 같더라도 흩어짐이 틀리면 그 평가방법도 달라야 된다는 것을 알 수 있다.

흩어짐의 비교는 편차의 절대치의 평균을 비교하면 된다.

$$(A) \frac{2.3+0.4+0.4+2.8+0.3}{5} = 1.24$$

$$(B) \frac{3.0+1.1+1.1+2.9+0.1}{5} = 1.64$$

(A) (B) 두 조의 데이터는 평균치가 같으나 각 측정치의 흩어짐은 차가 크다.

즉 평균치와의 떨어진 거리가 다르다는 것을 알 수 있다.

(나) 분산과 표준편차의 계산방법 흩어짐의 척도로서 통계학에서 가장 많이 사용하는 것이 표준편차이다.

분산은 표준편차를 제곱한 것이므로 표준편차나 같은 뜻을 가지며 많이 사용된다.

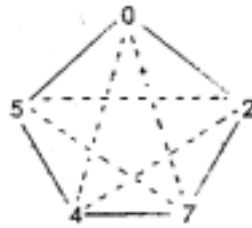
순 서	방 법	계 산	문 자 표 시
1	측정치를 기록하고 데이터의 수를 확인한다. N=5	101 102 100 99 98	$X_1=101$ $X_4=99$ $X_2=102$ $X_5=98$ $X_3=100$
2	평균치를 구한다.	$\frac{101+102+100+99+98}{5}$	$\frac{X_1+X_2+X_3+X_4+X_5}{5}$ $= \frac{\sum_{i=1}^5 x_i}{5}$
3	각 측정치에서 평균치를 빼다. 이 값이 편차이다.	$101-100=1$ $102-100=2$ $100-100=0$ $99-100=-1$ $98-100=-2$	$(X_1 - \bar{X}) = X_1$ $(X_2 - \bar{X}) = X_2$ $(X_3 - \bar{X}) = X_3$ $(X_4 - \bar{X}) = X_4$ $(X_5 - \bar{X}) = X_5$
4	(3)에서 구한 편차의 제곱을 구하여 합계한 후 자유도(6)로 나눈다. 이것이 분산이다.	$\frac{(101-100)^2 + (102-100)^2 + (100-100)^2}{5-1} + \frac{(99-100)^2 + (98-100)^2}{5-1} = 2.5$	$\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2}{n-1} + \frac{(X_3 - \bar{X})^2 + (X_4 - \bar{X})^2}{n-1} + \frac{(X_5 - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \bar{V}$
5	(4)의 값의 제곱근을 구한다. 이것은 표준편차이다.	표준편차 = $\sqrt{\text{분산}}$ $= \sqrt{2.5} = 1.8$	$\delta = \sqrt{\bar{V}}$

주[6] 자유도라는 것은 초보자에게는 매우 이해하기 어려운 낱말이다.

자유도는 실험으로부터 얻어지는 정보량의 하나치로서 시료의 크기 n에서 1을 뺀 숫자가 자유도이며 Φ 로 표시하고 Φ 는 “파이”라 읽는다.

$$\Phi = n-1$$

그러면 Φ 는 도대체 어떤 의미를 가지는 것일까? 속자들은 궁금할 것이다.
 시료의 크기 $n=5$ 일때 실험결과가 각각 4, 5, 0, 7, 2로 알려졌다면 이 다섯가지 실험결과에서 우리는 $\Phi = 5-1 = 4$
 즉 4가지의 정보량을 얻을 수 있다.



<그림 5>

그림 5와 같이 5가지 실험결과로서 얻어지는 정보량은 4가지다.
 다시 말하면 7과 2의 차, 2와 0의 차, 5와 0과의 차, 5와 4와의 차의 4가지만 비교하여 알게되면 나머지 정보량은 자연스럽게 알게 된다.

그림 5에서 실선(-)은 자유도를 표시하고 점선($\cdot \cdot \cdot$)은 자유도의 정보량만 알면 자연스럽게 알 수 있는 것을 표시한다.

이것을 풀이 해 보자.

7과 2의 차 $7-2=5$

2와 0의 차 $2-0=2$

0과 5의 차 $0-5 = -5$

4와 5의 차 $4-5 = -1$

는 5가지 실험에서 얻어진 정보이다. 이 정보만 알면 나머지 점선으로($\cdot \cdot \cdot$) 연결된 실험차의 정보도 알 수 있다.

(7-0)의 차는

$$(7-2)+(2-0)=(7-0)$$

$$5+2=7$$

로서 (7-2)와 (2-0) 두 가지 정보를 이용하여 구할 수 있다.

(7-5)의 차는

$$(7-2)+(2-0)+(0-5)=(7-5)$$

$$5+2-5=2$$

로서 (7-2), (2-0)과 (0-5)의 세 가지 정보를 이용하여 알 수 있다.

(7-4)의 차는

$$(7-2)+(2-0)+(0-5)+(5-4)=(7-4)$$

$$5+2+(-5)+1=+3$$

로서 (7-2), (2-0), (0-5), (5-4)의 네 가지 정보를 이용하여 구할 수 있다.

(4-2)의 차는

$$(0-2)+(5-0)+(4-5)=(4-2)$$

$$(-2)+5+(-1)=2$$

로서(0-2), (5-0), (4-5)의 세 가지 정보로서 구할 수 있다.

(4-0)의 차는

$$(5-0)+(4-5)=(4-0)$$

$$5+(-1)=4$$

로서 (5-0), (4-5)의 두 가지 정보로서 구할 수 있다.

(5-2)의 차는

$$(0-2)+(5-0)=(5-2)$$

$$(-2)+5=3$$

으로서 (0-2), (5-0)의 두 가지 정보로서 구할 수 있다.

(b) 간접법

순서	방법	계산	문자표시
1	데이터를 기록한다.		평균치 x_0
2	수치를 간단히 하기 위하여 변수 변환을 한다. 이때 가 평균 x_0 와 곱수 h 는 변환 후의 수치가 간단히 되도록 정한다.	$\textcircled{1}-\textcircled{2}=\textcircled{3}$ $101-100=1$ $102-100=2$ $100-100=0$ $99-100=-1$ $98-100=-2$	곱수 h 변환치 x_i $X_1=X_2-X_0$ $X_2=X_2-X_0$ $X_3=X_3-X_0$ $X_4=X_4-X_0$ $X_5=X_5-X_0$
3	측정치에서 가평균을 빼어 변환치를 구한다.		
4	변환치의 합계를 구한다.	$x_i \quad x_i^2$	변환수의 합계
5	변환치의 제곱을 구하고 그 합계를 구한다.	$1 \quad 1$ $2 \quad 4$ $0 \quad 0$ $-1 \quad 1$ $-2 \quad 4$	$X_1 + X_2 + X_3 + X_4 + X_5$ $\sum_{i=1}^5 x_i^2$ $i=1$ 변환치 제곱의 합계 $X_1^2 + X_2^2 + X_3^2 + X_4^2 + X_5^2$ $= \sum_{i=1}^5 x_i^2$
6	변환치의 평방차 S 를 구한다.	$S = \frac{1}{1^2} \times (10 - \frac{0^2}{5}) = 10$	$S = \frac{1}{h^2} (\sum x^2 - \frac{(\sum x_i)^2}{n})$ 주(7)
7	분산 V 를 구한다.	$V = \frac{10}{5-1} = 2.5$	$V = \frac{S}{n-1}$
		$\hat{\sigma} = \sqrt{2.5} = 1.58$	$\hat{\sigma} = \sqrt{V}$

주[7] $(x_1 - x_0) = x_i$

$$\sum x_i^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

가 성립되는 것은 다음과 같이 증명된다.

$$\sum (x_i - x_0)^2$$

$$= \sum x_i^2 - 2 \sum x_i x_o + n x_o^2$$

$$(\because \sum_{i=1}^n x_i = n x_o)$$

$$= \sum x_i^2 - 2 \sum x_i \left(\frac{\sum x_i}{n} \right) + \left(\frac{\sum x_i}{n} \right)^2$$

$$= \sum x_i^2 - \frac{2(\sum x_i)^2}{n} + \frac{(\sum x_i)^2}{n}$$

$$= \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$s = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

을 편차 평방지라고 하며 이 식으로 계산하면 쉽게 구할 수 있다.

또 $\frac{(\sum x_i)^2}{n}$ 을 수정항이라 하고 CT역은 CF로 표시한다.

예를 들어 설명하여 보자.

4개의 데이터

120, 125, 119, 124 가 있을 때 이 데이터의 편차평방지를 구하여 보자.

a) 일반적 방법에 의한 계산

데이터의 평균치는

$$\bar{x} = \frac{120+125+119+124}{4} = 122$$

이고 편차평방지 s는

$$s = (120-122)^2 + (125-122)^2 + (119-122)^2 + (124-122)^2$$

$$= 2^2 + 3^2 + (-3)^2 + 2^2$$

$$= 4 + 9 + 9 + 4$$

$$= 26$$

b) $\sum x_i^2 - \frac{(\sum x_i)^2}{n}$ 에 대한 계산

[주8] d_2 란 R관리도의 중앙선(Center Line)의 값을 정하는 계수로서 조내의 시료의 크기에 따라 결정되는 값이며 ASTM Manual on Quality Control of Material(1951)에 발표되었다. 또 d_2 는 조의 표준편차를 구하는 계수도 된다.

시료의 집단을 여러 개의 조로 나누고 각조의 최고와 최저의 차로 표준편차를 구하는 방법이다.

각조에 속하는 시료의 크기(Sample Size) n 은 4~5로 하는 것이 좋다.

N 가 너무 작으면 검출력이 작아지고 정도도 나쁘나 n 이 4~5일 때는 직접 법으로 구하는 표준편차의 값과 거의 비슷하여 가장 정도가 높고 또 계산하기도 쉽다.

<표 2> d_2 의 값

n	d_2	n	d_2
2	1.128	14	3.407
3	1.693	15	3.472
4	2.059	16	3.532
5	2.326	17	3.588
6	2.534	18	3.640
7	2.704	19	3.689
8	2.847	20	3.735
9	2.970	21	3.778
10	3.078	22	3.819
11	3.173	23	3.858
12	3.258	24	3.895
13	3.336	25	3.931

일반적으로 기내변동이나 기간변동을 알기 위해서는 10조 정도의 시료를 택하는 것이 좋다.

가령 n 가 5이고 조가 10이라면 이 시료의 개수는 $5 \times 10 = 50$ 즉 50개가 된다.

n 가 7이상되는 것은 정도가 나빠서 특수한 경우를 제외하고는 잘 이용되지 않는다.