★ 히스토그램의 작성법 ★

이수화학 TPM 컨설팅 교육자료
KSA (한국표준협회) TPM 컨설턴트 權五運

■ 제1편 히스토그램의 작성법

(1) 히스토그램 (도수분포도)

(가) 히스토그램의 정의

히스토그램(histogram)이란 데이터가 존재하는 범위를 몇 개의 구간으로 나누어 각 구간에 들어가는 데이터의 출현도수를 세어서 도수표(도수분포표)를 만든 다음 이것을 도형화한 것으로 데이터의 분포파악 및 평균치, 산포크기가 파악가능한 그림을 말한다.

(나) 도수분포표에 의한 히스토그램의 작성

(a) 도수분포의 의의

우리는 좋은 품질의 물건을 만들기 위해 매일 노력하고 있다. 그러나 똑같은 공정에서 작업자가 동일한 작업표준에 의해서 똑같은 재료를 사용, 동일한 설비에서 물건을 만들어도 완성된 물건의 품질(품질특성)에는 반드시 어느 정도의 산포가 있게 마련이다.

품질특성(예를 들어 길이, 무게, 시간, 온도 등)을 측정해서 얻어지는 데이터의 도수는, 어떤 값을 중심으로 해서 가장 많고 그 중심에서 멀어짐에 따라 그 비율이 점차 적어지는 경우가 많다. 이것을 가리켜 공정 혹은 물건의 집단이 분포를 가진다고 말한다. 즉, 집단으로서의 물건치 성질을 알기 위해서는 도수분포가 도움이 된다.

도수분포(frequency distribution)란『데이터가 존재하는 범위를 몇 개의 구간으로 나누어, 각 구간에 들어가는 것이 몇 개 있는가를 도수를 세어서 그림이나 표로 한 것』으로서 전체의 분포를 파악하기 쉽다.

도수분포를 나타내는 것으로 도수표(도수분포표), 히스토그램, 누적도수 등의 3가지 방법이 있다. 품질관리의 기본은 품질에 산포를 가져오는 원인을 파악해서 그것을 관리하는 일로, 우선 품질을 나타내는 특성치에 대해서 그 변동상태를 올바르게 파악할 필요가 있는데, 이와 같은 특성치의 변동을 눈에 보이는 모양으로 만드는 기술이 도수분포이다.

(b) 도수분포의 작성목적

도수분포의 작성목적으로서는 ①데이터의 분포의 형태를 파악, ②평균치를 구함, ③표준편차를 구함, ④규격과 대조 등의 목적으로 작성한다.

(다) 도수분포표에 의한 히스토그램의 작성순서

다음 <표 7-2>의 데이터는 실의 인장강도의 측정결과를 나타낸 것이다. 회사규격은 $5.6 \pm 0.6(g)$ 이라 한다. 주어진 실의 인장강도를 예로 들어 도수분포표에 의한 히스토그램의 작성순서를 설명하여 보기로 한다.

<표 7-2> 실의 인장강도의 데이터 (단위 : g)

4.7	4.9	5.1	5.7	5.9	4.7	5.4	5.5	5.8	5.1
5.2	5.3	5.4	5.0	5.1	5.2	6.0	5.4	5.1	5.3
5.5	5.7	4.7	5.3	5.4	5.5	5.3	5.6	5.8	5.6
4.9	5.1	5.2	5.7	5.9	4.8	5.3	5.1	5.3	5.3
5.3	5.4	5.5	5.0	5.2	5.3	5.6	5.8	5.0	5.6

▶순서1 : data를 수집함. 이 경우 $n=50$ 이 된다.

▶순서2 : 최대치(L : Largest value) 및 최소치(S : Smallest value)를 구한다. 이 경우 최대치 $L=6.0$, 최소치 $S=4.7$ 이 얻어진다.

▶순서3 : 구간의 수(k)를 식(7.1)에 의하여 구한다.

$$k=\sqrt{n} \qquad (7.1)$$

식(7.1)에 의한 구간의 수는 다음과 같이 구해진다.

$$k=\sqrt{n}=\sqrt{50}=7.1\approx 7$$

그리고 구간의 수(k)를 스터지스(H.A. Sturges)가 제시한 식(7.2)에 의하여 구할 수도 있다.

$$k=1+3.3\log n \qquad (2.2)$$

식(7.2)에 의하여 구간의 수(k)를 구하면 식(7.1)과 동일한 값이 나온다.

$$k=1+3.3\log n=1+3.3\log 50=1+5.6=6.6\approx 7$$

▶순서4 : 구간의 폭(h)를 정한다. 그리고 측정단위의 정수배로 수치맺음한다.

$$h=\frac{L-S}{k}=\frac{6.0-4.7}{7}=0.19\approx 0.2$$

(측정단위 0.1이므로 정수배는 0.1×2=0.2)

▶순서5 : 각 계급의 경계치를 구한다.

$$1\text{급하한}=S-\frac{\text{측정단위}}{2}=4.7-\frac{0.1}{2}=4.65$$

$$1\text{급상한}=1\text{급하한}+h=4.65+0.2=4.85$$

$$2\text{급상한}=1\text{급상한}+h=4.85+0.2=5.05$$

⋮

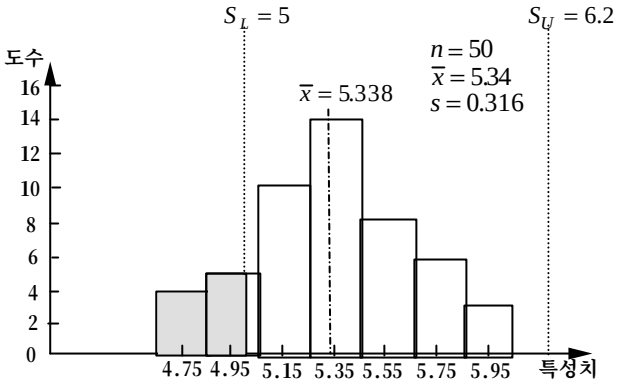
▶순서6 : 도수표를 작성(평균치, 표준편차 계산시는 아래의 도수표와 같이 작성하면 편리함)

<표 7-3> 도수표

No.	급의 구간	중앙치($\tilde{x}_i$)	도수체크	f	u	fu	fu^2
1	4.65 4.85	4.75	////	4	-3	-12	36
2	4.85 5.05	4.95	////	5	-2	-10	20
3	5.05 5.25	5.15	//// ////	10	-1	-10	10
4	5.25 5.45	5.35	//// //// ////	14	0	0	0
5	5.45 5.65	5.55	//// ///	8	1	8	8
6	5.65 5.85	5.75	//// /	6	2	12	24
7	5.85 6.05	5.95	///	3	3	9	27
계	-	-		50		-3	125

여기서 u 는 도수가 제일 큰 것을 0으로 한다.

▶순서7 : 히스토그램을 작성한다.



〔그림 7-3〕 실의 인장강도의 히스토그램

▶순서8 : 평균치 $\bar{x}$, 표준편차 s 를 계산한다.

평균치 $\bar{x}$는 다음 식(7.3)에 의하여 구한다.

$$\bar{x} = x_0 + \frac{\sum fu}{\sum f} \times h \qquad (7.3)$$

상기의 식(7.3)에 의하여 산술평균을 구하면,

$$\bar{x} = x_0 + \frac{\sum fu}{\sum f} \times h = 5.35 + \frac{-3}{50} \times 0.2 = 5.338 \approx 5.34$$

그리고 표준편차 s 는 식(7.4)에 의하여 구한다.

$$s = \sqrt{\frac{S}{\sum f}} = \sqrt{\frac{\sum fu^2 - (\sum fu)^2 / \sum f}{\sum f}}$$

$$= h\sqrt{\frac{\sum fu^2}{\sum f} - \left(\frac{\sum fu}{\sum f}\right)^2} \qquad (7.4)$$

주어진 식(7.4)에 의하여 표준편차를 구하면,

$$\therefore s = 0.2\sqrt{\frac{125}{50} - \left(\frac{-3}{50}\right)^2} = 0.316$$

단, 치우침이 없는 불편분산 V 의 제곱근 $\sqrt{V}$ 에 의거 산정할 경우($s \approx \sqrt{V}$) 다음과 같이 계산된다.

$$\sqrt{V} = h\sqrt{\frac{\sum fu^2 - (\sum fu)^2 / \sum f}{\sum f - 1}} = 0.2\sqrt{\frac{125 - (-3)^2/50}{50 - 1}} = 0.319$$

▶순서9 : 히스토그램에 필요사항을 기재

데이터의 수 n , 평균치 $\bar{x}$, 표준편차 s 등을 히스토그램에 기재한다.

예제4-4 다음 데이터는 A공장에서 생산된 어느 기계부품 중에서 랜덤하게 64개를 취하여 크기를 측정한 것이다. (단위 : cm)

53	52	46	49	51	57	60	45
64	45	49	58	46	62	53	50
51	50	52	54	63	49	45	58
50	65	53	58	49	50	54	44
53	61	48	54	50	52	58	51
39	51	58	52	50	54	48	61
66	52	46	54	50	53	48	54
50	53	54	60	46	64	51	57

(1) 위의 데이터로 제일 낮은 구간의 하한이 38.5가 되도록 하고 k(급의 수)를 H.A. Sturges법으로 구하라.

1) k(급의 수)를 구하라.

2) h(급의 폭)을 구하라.

(2)위의 데이터로 1)도수표를 작성하고 도수표를 사용하여 2)평균 $\bar{x}$, 3)불편분산 V , 4)표준편차 s 를 구하라.

〔풀이〕 (1) 1) $k = 1 + 3.3\log n = 1 + 3.3\log 64 = 6.96 \approx 7$

$$2) h = \frac{x_{\max} - x_{\min}}{k} = \frac{L - S}{k} = \frac{66 - 39}{7} = 3.9 \approx 4$$

(2) 1)도수표의 작성

급의 번호	경계치	중앙치	f	u	fu	fu^2
1	38.5 42.5	40.5	1	-3	-3	9
2	42.5 46.5	44.5	8	-2	-16	32
3	46.5 50.5	48.5	15	-1	-15	15
4	50.5 54.5	52.5	23	0	0	0
5	54.5 58.5	56.5	7	1	7	7
6	58.5 62.5	60.5	5	2	10	20
7	62.5 66.5	64.5	5	3	15	45
계			64		-2	128

$$2)\ \bar{x} = x_0 + \frac{\sum fu}{\sum f} \times h = 52.5 + \frac{(-2)}{64} \times 4 = 52.375$$

$$3)\ V = \frac{S}{n-1} = \frac{\left\{\sum fu^2 - (\sum fu)^2 / \sum f\right\} h^2}{\sum f - 1} = \frac{\{128 - (-2)^2/64\} \times (4)^2}{64 - 1} = 32.492$$

$$4)\ s = \sqrt{\frac{S}{\sum f}} = h\sqrt{\frac{\sum fu^2}{\sum f} - \left(\frac{\sum fu}{\sum f}\right)^2} = 4 \times \sqrt{\frac{128}{64} - \left(\frac{-2}{64}\right)^2} = 5.655$$

유 제 다음 <표 7-2>의 데이터는 실의 인장강도의 측정결과를 나타낸 것이다.

회사규격은 $5.4 \pm 0.6(g)$이라 한다. 주어진 실의 인장강도를 예로 들어 도수분포표에 의한 히스토그램의 작성순서를 보기이고 히스토그램을 작성한 후 필요사항인 평균치($\bar{X}$) 및 표준편차(s)를 기재하고 규격을 그린 후 규격과 대비하여 보아라.

<표 7-2> 실의 인장강도의 데이터 (단위 : g)

4.7	4.9	5.1	5.7	5.9	4.7	5.4	5.5	5.8	5.1	5.7
5.2	5.3	5.4	5.0	5.1	5.2	6.0	5.4	5.1	5.3	5.0
5.5	5.7	4.7	5.3	5.4	5.5	5.3	5.6	5.8	5.6	5.3
4.9	5.1	5.2	5.7	5.9	4.8	5.3	5.1	5.3	5.3	5.7
5.3	5.4	5.5	5.0	5.2	5.3	5.6	5.8	5.0	5.6	5.0

단, $\sqrt{55} = 7.4$, $5.35 + \frac{-3}{55} \times 0.2 = 5.339$, $0.2\sqrt{\frac{141}{55} - \left(\frac{-3}{55}\right)^2} = 0.320$임.

【풀이】

▶순서1 : data를 수집함. 이 경우 $n =$ 가 된다.

▶순서2 : 최대치(L : Largest value) 및 최소치(S : Smallest value)를 구한다.

이 경우 최대치 $L =$, 최소치 $S =$ 이 얻어진다.

▶순서3 : 구간의 수(k)를 식(7.1)에 의하여 구한다.

$$k = \sqrt{n} \qquad (7.1)$$

식(7.1)에 의한 구간의 수는 다음과 같이 구해진다.

$$k = \sqrt{n} = \sqrt{\quad} =$$

▶순서4 : 구간의 폭(h)를 정한다. 그리고 측정단위의 정수배로 수치맺음한다.

$$h = \frac{L - S}{k} =$$

(측정단위 0.1이므로 정수배는 0.1×2=0.2)

▶순서5 : 각 계급의 경계치를 구한다.

$$1\text{급하한} = S - \frac{\text{측정단위}}{2} =$$

$1\text{급상한} = 1\text{급하한} + h =$

$2\text{급상한} = 1\text{급상한} + h =$

$3\text{급상한} = 2\text{급상한} + h =$

$4\text{급상한} = 3\text{급상한} + h =$

$5\text{급상한} = 4\text{급상한} + h =$

$6\text{급상한} = 5\text{급상한} + h =$

$7\text{급상한} = 6\text{급상한} + h =$

▶순서6 : 도수표를 작성(평균치, 표준편차 계산시는 아래의 도수표와 같이 작성하면 편리함)

<표 7-3> 도수표

No.	급의 구간	중앙치($\tilde{x}_i$)	도수체크	f	u	fu	fu^2
1							
2							
3							
4							
5							
6							
7							
계	-	-					

여기서 u 는 도수가 제일 큰 것을 0으로 한다.

▶순서7 : 히스토그램을 작성한다.(아래 그림 7.3 참조)

▶순서8 : 평균치 $\bar{x}$, 표준편차 s 를 계산한다.

평균치 $\bar{x}$는 다음 식(7.3)에 의하여 구한다.

$$\bar{x} = x_0 + \frac{\sum fu}{\sum f} \times h \qquad (7.3)$$

상기의 식(7.3)에 의하여 산술평균을 구하면,

$$\bar{x} = x_0 + \frac{\sum fu}{\sum f} \times h =$$

그리고 표준편차 s 는 식(7.4)에 의하여 구한다.

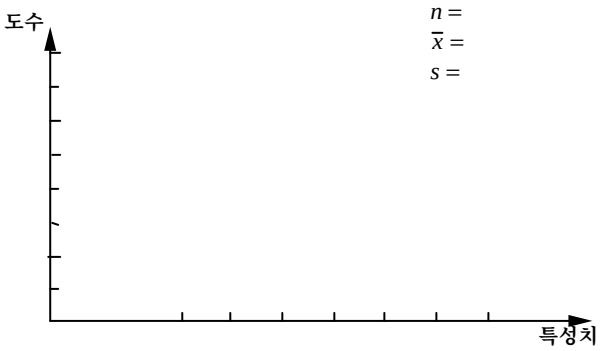
$$s = h\sqrt{\frac{\sum fu^2}{\sum f} - \left(\frac{\sum fu}{\sum f}\right)^2} \qquad (7.4)$$

주어진 식(7.4)에 의하여 표준편차를 구하면,

$\therefore\ s =$

▶순서9 : 히스토그램에 필요사항을 기재

데이터의 수 n , 평균치 $\bar{x}$, 표준편차 s 등을 히스토그램에 기재한다.



〔그림 7-3〕 실의 인장강도의 히스토그램