

방직공장의 QC. Engineer를 위한 품질관리(6)

3.4.2 이항분포

어떤 일이 일어나는 확률이 p , 일어나지 않는 확률이 $q=1-p$ 일 경우, 아무렇게 추한 n 회중 x 회만이 일어나는 확률(이것을 이항정리라 한다)은

$$p_x = \binom{n}{x} p^x q^{n-x} \quad (x = 0, 1, 2, \dots, n)$$

즉 $0 < p < 1$ 로 표시되고 x 는 $0, 1, 2, \dots, n$ 가 각각 확률 $p_0, p_1, \dots, p_n$ 를 가지고 나타난다.

$\binom{n}{x}$ 는 다음 식

$$\binom{n}{x} = \frac{n(n-1) \cdots (n-x+1)}{x!}$$

로 계산할 수 있고 분자는 n 으로부터 하나씩 줄여서 x 개를 취하여 그 곱셈으로 구한 것이고 분모는 x 의 계승(Factorial)으로서

$x! = x(x-1)(x-2) \cdots 3 \cdot 2 \cdot 1$ 이란 뜻이다.

즉 1로부터 x 까지의 정수의 적이다.

일반적으로 불량률 p 인 Lot로부터 독립적으로 1개씩 n 회 발취하여 검사할 때 불량품이 검출되는 회수는 이항분포에 의해서 지배된다.

지금 불량률 20%의 Lot로부터 한번에 1개씩 10회 발취했을 때 불량품이 검출되는 확률분포는

$$n=10 \quad p=0.2 \quad q=0.8$$

로 하여 표 10과 같이 된다.

표 10을 Graph로 그리면 그림 18과 같다.

이항분포 곡선에 지배받는 확률변수 x 의 기대치 $E(x)$ 와 표준편차 $D(x)$ 는

$$E(x) = np$$

$$D(x) = \sqrt{npq} \quad \text{이다.}$$

(예 1) 표10의 $n=10$, $p=0.2$ 일 때의 기대치 $E(x) = \sum x p_x$ 는 $E(x)=np$ 와 일치하고 분산 $V(x) = \sum (x - E(x))^2 p_x$ 는 $D(x)^2 = npq$ 와 일치함을 살펴보자.

표11에서 $E(x) = \sum x p_x = 2,000$ 이 되어 $np=10 \times 0.2=2$ 와 잘 일치하며 $V(x) = \sum (x - 2)^2 p_x = 1.600$ 이 되어 $npq=10 \times 0.2 \times 0.8=1.6$ 과 잘 일치한다.

(예 2) 표12는 매일 검사하는 Cheese 2,500개중에 있는 불량 Cheese수를 나타낸다.

이 표에서 평균 불량률과 표준편차를 구하라.

$$\text{불량률 } p = \frac{1,800}{2,500 \times 30} = 0.024$$

$$\bar{x} = np = 2,500 \times 0.024 = 60$$

$$\sigma = \sqrt{npq}$$

$$= \sqrt{2,500 \times 0.024 \times \sqrt{(1 - 0.024)}} = 7.71$$

위의 예에서 보는 바와 같이 이항분포는 n 이 작고 p 가 클 때는 성립되나 n 이 아주 크고 p 가 작게 되면 3.4.3에서 말하는 Poisson의 분포를 따르게 된다.

또 불량률 p 가 커지면(즉 np 가 커진다) 좌우 맞선꼴이 되어 정규분포에 가까운 분포를 하게 된다. 물론 정규분포는 연속분포이고 이항분포는 개개 값의 이산분포이지만 $np \geq 3$ 일 때(실용면에서는 1이하가 되어도 된다) 이항분포

는 대체로 정규분포라고 생각해도 좋다.

불량률 p 가 커지고 n 이 일정할 때 분포곡선의 모양을 보여 주는 것이 그림 19이다.

3.4.3 포아송분포(Poisson Distribution)

직물의 흠, 소면 Web의 Nep과 같이 결점수가 모집단내에 Random하게 분포되어 있을 때 이러한 모집단으로부터 Random하게 Sampling하여 그 결점수를 세어보면 포아송분포를 하게 된다.

즉 $1m^2$ 당 m 개의 흠을 갖는 직물로부터 $1m^2$ 의 직물을 Random Sampling하였을 때

x 개의 흠이 나타나는 확률 p_x 는

$$p_x = e^{-m} \frac{m^x}{x!} \quad (x=0, 1, 2, \dots, m>0)$$

라는 확률분포로 나타난다. 이것을 Poisson분포라고 한다.

Poisson분포에 따르는 확률변수의 기대치는 변수 m 과 같고 표준편차는 $\sqrt{m}$ 이다.

$$\text{즉 } E(x) = m$$

$$D(x) = \sqrt{m}$$

또 m 가 커지면 그림 20과 같이 정규분포에 가까운 분포를 하게 된다.

Poisson분포

$$p_x = e^{-m} \frac{m^x}{x!}$$

에서 e 는 자연대수의 밑으로 $2.71828 \dots$ 이란 값을 가졌고 e^{-m} 은 “ e 의 $(-m)$ 제곱”으로 $1/e^m$ 와 같다.

e^m 은 다음과 같은 급수로 전개된다.

$$e^m = 1 + \frac{m}{1!} + \frac{m^2}{2!} + \frac{m^3}{3!} + \dots$$

따라서

$$1 = e^{-m} + e^{-m} \frac{m}{1!} + e^{-m} \frac{m^2}{2!} + e^{-m} \frac{m^3}{3!} + \dots$$

m이 정수라면 이 값은 정이고 합계는 1이 된다.

이제 m=2일 때 Poisson분포를 표 12로 계산해 보자.

합계가 1보다 0.00003 많은 것은 반올림으로 인한 오차때문이다.

표 12의 p_x 의 값을 그림으로 그리면 그림 21과 같다.

직물의 흠이나 딱지, 소면 Web의 Nep, 실의 사반등 결점분포는 공정이 안정되어 있어야 Poisson분포를 따르게 된다. 즉 모집단의 변동이 심할 때에는 이 분포가 적용되지 않기 때문이다.

(예 1) 과거의 Data에 의하면 소면 Web의 Nep가 A원면조합으로 방출하였을 때 100평방inch당 3이었다고 한다.

최근에 갑자기 5개/100"구의 측정치가 발생했다면 이 모집단에서 5개/100"구가 발생할 확률은 얼마나?

$$(풀이) p_x = e^{-m} \frac{m^x}{x!}$$

에서 m=3, x=5 이므로

$$\begin{aligned} p_x &= e^{-3} \frac{3^5}{5!} \\ &= \frac{1}{(2.71828)^3} \times \frac{3^5}{5!} \\ &= \frac{1}{20.0855} \times \frac{243}{120} = 0.1008 \end{aligned}$$

즉, 약 10%의 출현확률이므로 이 결과를 유의차 검정하여 보면 이상 유무를 판단할 수 있다.

(설명)

직물의 결점, 실의 사반, 딱지, Nep 등을 이 Poisson분포에 의해서 계산하려면 계산이 복잡해지기 때문에 Thorndike의 누적 확률곡선도에 의해서 계산하면 아주 쉽고 간단하게 구할 수 있기 때문에 이 방법을 권하고 싶다.

Thorndike곡선도에서 p_n 은 m 에 해당하고 C 는 x 에 해당하며 이 문제에서 $x=5$ 가 나타나는 누적확률에서 $x=4$ 가 나타나는 누적확률을 빼면 $x=5$ 만이 나타나는 확률이 된다.

$$x_1=4\text{일 때 } p_{x_1}=0.915$$

$$x_2=5\text{일 때 } p_{x_2}=0.810$$

$$p_{x_1}-p_{x_2}=0.915-0.810=0.105$$

가 되어 앞의 계산과 비슷한 결과가 된다(Thorndike 곡선도는 발취 검사법을 설명한 책에는 거의 전부 실려 있으므로 참고할 것).

3.4.2의 (예2)에서 검사한 Cheese 2,500개의 평균치와 표준편차를 Poisson분포와 비교하여 보자.

Poisson분포에서는

$$\bar{x} = m = np$$

$$\sigma = \sqrt{m}$$

니까

$$x=2,500 \times 0.024=60$$

$$\sigma = \sqrt{60} = 7.75$$

로서 이항분포때의 $\sigma=7.71$ 과 거의 비슷하다. 만약 검사개수가 많고 불량률이 더 적으면 σ 는 더욱 비슷한 값이 될 것이다.

이제 실제문제를 들어 n 이 커지고 p 가 작아짐에 따라 Poisson분포의 값이 이항분포의 값과 일치함을 살펴보자.

표 13은 $m=2$ 되는 Poisson분포의 확률과 $np=2$ 가 되는 여러 가지 이항분포

의 확률을 비교한 것이다.

이 표에서 $n=100$ $p=0.02$ 일 때와 $m=2$ 일 때의 두 가지 분포를 비교하면 그 값이 거의 일치함을 알 수 있다.

즉 p 가 적어지면 적어질수록 $q=1-p$ 는 1에 가까워지므로 $np=m$ 가 되는 것이다.

3.5 검정, 추정

우리가 공장에서 일할 때 측정치의 “흠어짐”으로 인하여 여러 가지 문제에 부딪칠 때가 많다. 특히 단위 섬유의 결속으로 이루어지는 집합체인 섬유제품은 그것이 완제품이던 반제품이던 간에 부동이 필연적으로 따르게 마련이고 또한 이 분산으로 인한 평균치의 변동이 생기게 된다.

이러한 변동이 생기는 원인은 여러 가지가 있겠지만 크게 나누어 우연적인 변동과 보아넘길 수 없는 변동, 두 가지로 나눌 수 있다.

매일 똑 같은 기계로 같은 사람이 작업을 했는데도 제품의 차가 심하다든가, 기계부속의 변경으로 인한 품질의 변동, 이유없이 실이나 직물의 인장강력이 떨어진다는가 하는 공정상의 문제로부터 A사의 호료를 사용하다가 B사의 호료로 바꾸어 호부를 하고자 한다든가, 원면조합의 변경으로 인한 수율의 경제성을 검토한다든가, 직물의 조성변경으로 인한 여러 가지 특성치의 비교라든가, 기료품의 양부를 판단한다든가, 하는 2 가지 이상의 문제를 놓고 그 우열을 판단해야 할 경우가 우리 주변에는 무수히 많다.

이런 경우에 우리는 대개 모집단의 값으로 판정을 내릴 수는 없고 Sample에서 나온 Data에 의할 때가 많다. 따라서 이 Data로 모집단을 판단해야 되는데 언제나 경험하는 일이지만 아무리 여러 번 모집단에서 Sample을 취하더라도 그 평균치는 전부 똑 같을 수 없기 때문에 Sample의 Data로 모집단의

상태를 그대로 믿어버린다는 것은 통계학적으로 상당히 위험한 일이다. 즉 통계량은 반드시 분포를 갖게 되며 모집단과 차가 있나 없나 하느것은 통계적으로 판단해야 한다.

대상이 되는 모집단의 평균이나 분산의 차이가 있는가, 없는가?를 객관적으로 확인하는 것이 검정이고 이 결과를 될 수 있는대로 정확히 추측하는 것을 추정이라고 한다.

검정, 추정의 대상은 모집단이지만 Sample로부터 얻어지는 Data에 의하여 결론을 얻어야 된다는 관점에서 볼 때 될 수 있는대로 오차가 적고 신뢰성이 있는 결론을 얻기 위해서 이미 알려진 확률을 충분히 염두에 두고 판단 기준이 되는 한계($u\sigma$)에 의하여 경제적이며 기술적으로 평균치의 차($\mu_x - \mu_0$)를 검출하는 것이 검정이다. 또 이것을 추측하는 것이 추정이다.

3.5.1 검정, 추정에 필요한 여러 가지 분포

(가) 시료 평균치의 분포

평균치가 μ , 표준 편차가 σ 인 모집단으로부터 Random Sampling한 Sample size n 인 시료의 평균치($\bar{x}$)분포는 $N(\mu, \frac{\sigma^2}{n})$ 인 정규분포를 한다.

즉 통계량의 중심치를 표시하는 방법으로 시료평균 $\bar{x}$ 가 있다. 시료를 Random으로 Sampling하여 시료의 평균 $\bar{x}$ 를 여러 번 반복하여 구하면 $\bar{x}$ 는 모평균 μ 에 일치하게 되나 $\bar{x}$ 의 분포는 모집단분포의 약 반으로 줄어드는 사실을 알게 된다. 즉

평균은 μ

분산은 $\frac{\sigma^2}{n}$

이 된다. 따라서 표준편차는 $\frac{\sigma}{\sqrt{n}}$ 가 된다. 이것을 그림으로 표시하면 그림 22와 같다.

이러한 평균치 분포의 성질을 요약하면

i) $\bar{x}$ 분포의 평균치는 $\mu_{\bar{x}} = \mu$ 가 된다.

ii) $\bar{x}$ 분포의 표준편차는 $\sigma/\sqrt{n}$ 이 된다.

iii) $\bar{x}$ 분포는 정규분포를 한다.

Sampling에 의한 평균치의 값으로 검정이나 추정을 하고자 할 때 평균치 분포의 표준편차는 반드시 $\sigma/\sqrt{n}$ 을 써야 한다.

(나) t분포

모든 측정이나 검정, 추정을 할 때 모평균은 쉽게 구할 수 있지만 모표준편차는 모를때가 많다.

$$\text{규준측정 } u = \frac{\mu - \mu_0}{\sigma}$$

는 모표준편차를 알때만 사용할 수 있고 부지일 때는 시료의 불편분산의 평방근($\sqrt{V}$)을 이용하게 된다.

$$\text{즉 } t = \frac{\bar{x} - \mu}{\sqrt{V}/\sqrt{n}}$$

로 변형하여 t분포를 사용하는 것이다.

$$\text{그런데 } \hat{\sigma} = \sqrt{V} = \sqrt{\frac{S}{\phi}}$$

$$S = \sum (x_i - \bar{x})^2$$

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

$$\phi = n - 1$$

이와 같이 t분포가 표준정규분포와 틀리는 점은 모표준편차 σ 대신에 추정치 $\hat{\sigma}$ 을 사용하는 점이다.

또 t분포는 그림 23과 같이 평균0의 좌우 맞선꼴인 것은 정규분포와 같으나 출정이 좀 둔한 것이 다르다.

t분포는 Sample의 크기에서 1을 뺀 자유도 $\Phi(n-1)$ 에 따르는 분포이며 Φ 가 커짐에 따라 정규분포에 가까워지고 $n > 25$ 이면 정규분포와 거의 일치하게 된다.

(다) F분포

제조공정에서 기종을 바꾼다든가, 부품을 바꾸었을 때 제품의 분산을 비교하여 그 정도를 평가하고 싶을 때가 있다.

이때 F분포를 사용하여 검정, 추정을 하면 된다.

하나의 정규 모집단으로부터 Sample Size n_1, n_2 개의 시료를 Random으로 취하여 각각의 불편분산 V_1, V_2 를 구하고 V_1 과 V_2 의 비를 F로 표시한다.

$$F = \frac{V_1}{V_2}$$

F는 시료에서 계산한 값이기 때문에 실제로 1이 되기 어렵고 여러 번 반복하여 F의 값을 구하면 1을 중심으로 한 일정한 모양의 분포를 하게 된다.

이것을 F분포라고 하고 분포의 흩어짐은 시료의 크기에 따라 결정된다.

F분포표에서 그 값을 찾을 때는 V의 값이 큰쪽을 분자로 하여 계산한다.

즉 $V_1 > V_2$

F분포표를 사용할때는 Φ_1, Φ_2 란 2개의 자유도가 필요하며 $F(\Phi_1, \Phi_2)$ 로 표시한다. 또 F분포의 모양은 자유도 Φ_1, Φ_2 에 따라서 여러 가지로 변하며 1이상의 F에 대해서 확률 p가 되는 확률 한계치를 나타낸 거시 표 15이다.

그러면 F분포의 값을 찾는 법을 알아 보기로 하자.

$F(4, 15, 0.05)$ 의 값을 $\Phi_1=4, \Phi_2=15, p=0.05$ 일 때이므로 표 15의 Φ_1 에서 4를

찾고 Φ_2 에서 15를 찾는다.

Φ_1 , Φ_2 의 가로, 세로 줄에서 교차하는 곳에 상단에는 3.06, 하단에는 4.89가 있다. 상단은 5%의 값을 표시하고 하단의 것은 1%의 F분포값을 나타낸다.

즉 $F(4, 15, 0.05)=3.06$ 이다.

보통 수치표에는 95%나 99%의 경우는 없을 때가 많다.

이런때에는 다음 식이 성립하는 것을 이용하여 5%와 1%의 수식표에서 계산하여 구할 수가 있다.

$$F(\Phi_1, \Phi_2, p) = \frac{1}{F(\Phi_2, \Phi_1, 1-p)}$$

가령 자유도 4, 5인 F분포의 상측 95%점(하측 5%)과 99%의 점(하측 1%)의 값이 각각 얼마인가, 알아보자.

먼저 $F(15, 4, 0.05)=5.86$

$F(15, 4, 0.01)=14.2$ 임을 표 15로 알 수 있다. 따라서,

$$F(4, 15, 0.95) = \frac{1}{F(15, 4, 0.05)} = \frac{1}{5.86} = 0.171$$

$$F(4, 15, 0.99) = \frac{1}{F(15, 4, 0.01)} = \frac{1}{14.2} = 0.070$$

으로 계산하여 95%, 99%때의 값을 쉽게 구할 수 있음을 알 수 있다.

(라) χ^2 분포(Chi Distribution)

Q, C를 실시해 나가면 Data를 해석할 때 여러 가지로 곤란을 느낄 때가 많다.

연조나 조방 같은 공정에서 Sliver나 Roving의 중량분석이 심할 때 이 원인이 과연 어디에 있느냐? 즉 원면이나 섬유원료에 있는 것인지, 아니면 공정 자체에서 일어나는 것인지?를 쉽게 알 수 있는 방법이 없다. 설혹 안다고 하더라도 원면에 어느 정도의 분산이 있었고 이로 인하여 공정에서 어느 정도의 분산이 일어나는 것인지를 잘 모르기 때문에 그 원인을 명확히 판별해

보고 싶을 때가 많다.

이와 같이 Data로서 모분산을 검정하거나 추정하고 싶을 때에는 χ^2 분포를 쓰게 된다.

정규 모집단으로부터 n 개의 Sample을 취하여 평방화 S 를 구하고 이런 방법을 여러 번 되풀이하면 매번 나오는 S 의 값은 각 각 틀리게 되며 어떤 값을 중심으로 흠어짐이 일어나고 분포곡선을 이루게 된다. 평방화 S 를 모분산 σ^2 으로 나누어 규준화한 통계량을 χ^2 (가이제곱)이라고 한다.

즉 정규모집단 $N(\mu, \sigma^2)$ 으로부터 크기 n 인 Sample을 취하여 평방화 S 를 구한다. S 를 모분산 σ^2 으로 나눈 것은 자유도 $\Phi(n-1)$ 의 χ^2 분포를 한다.

$$\chi^2 = \frac{S}{\sigma^2} \quad \phi = n - 1$$

또 χ^2 분포의 기대치는 자유도 Φ 와 같고 표준편차는 $\sqrt{2\phi}$ 와 같다.

$$E(\chi^2) = \phi$$

$$D(\chi^2) = \sqrt{2\phi}$$

그림 25는 χ^2 분포 곡선을 나타내는 그림이다.

χ^2 분포가 자유도 Φ 에 따라 변화하는 모양을 좀 더 자세히 그리면 그림 26과 같다. χ^2 분포의 값은 표 16에 주어져 있다.

(마) t 분포, F 분포, χ^2 분포의 관계 χ^2 분포표로부터 얻을 수 있는 값 $\chi^2(\Phi, p)$ 과 t 분포표로부터 얻을 수 있는 값 $t(\Phi, p)$ 는 다음과 같은 관계식으로 F 분포표에서도 얻을 수 있다.

(예 1) 자유도 5인 χ^2 분포의 상측 5%점의 값을 F 분포표로 구하라

$$\chi^2(\Phi, p) = \Phi F(\Phi, \infty, p)$$

에서 $\chi^2(5, 0.05) = 5 \times F(5, \infty, 0.05) = 5 \times 2.21 = 11.05$

이 값을 실제 χ^2 분포표에서 찾아보면 11.07이 되어 거의 같은 값이 된다.

(예 2) 자유도 5인 χ^2 분포의 하측 5%점(상측 95%)의 값을 F분포표로부터 구하라.

$$\chi^2(\phi, p) = \frac{\phi}{F(\infty, \phi, 1-p)}$$

$$\text{예시 } \chi^2(5, 0.95) = \frac{5}{F(\infty, 5, 0.05)} = \frac{5}{4.36} = 1.146$$

이 값을 χ^2 분포표에서 찾으면 1.145로서 거의 같은 값이다.

(예 3) 자유도 7인 t분포의 양측 5%의 값을 F분포표로부터 구하라

$$t(\Phi, p) = \sqrt{F(1, 7, 0.05)} = \sqrt{5.59} = 2.368$$

t분포표의 $t(7, 0.05)$ 의 값 2.365와 거의 같은 값이다.