



# 기초통계의 개념과 이해

이 혜 선

포항공대 산업공학과 책임연구원

e-mail: [stat@postech.ac.kr](mailto:stat@postech.ac.kr)

<http://www.promoim.co.kr>



# Review

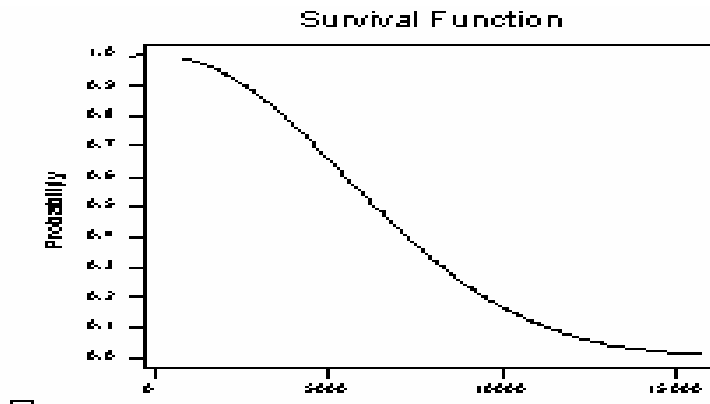
## Reliability & Confidence Level

# Review : 신뢰성(Reliability) - 정의와 척도

신뢰성 - 어떤 제품이 주어진 기능을 수행하는 능력(정성적 개념)

1. 신뢰도(reliability) - 특정시간까지 고장없이 기능을 수행할 확률

$$R(t) = P(X \geq t) = 1 - F(t)$$



2. 고장률(Failure rate or hazard rate) : 시간  $t$ 까지 고장이 없다가 그다음 1개월, 혹은 1년 후, 10년후 에 고장(사망)을 일으킬 확률을 의미한다.

$$\lambda(t) = \frac{f(t)}{P(X \geq t)} = \frac{f(t)}{1 - F(t)}$$

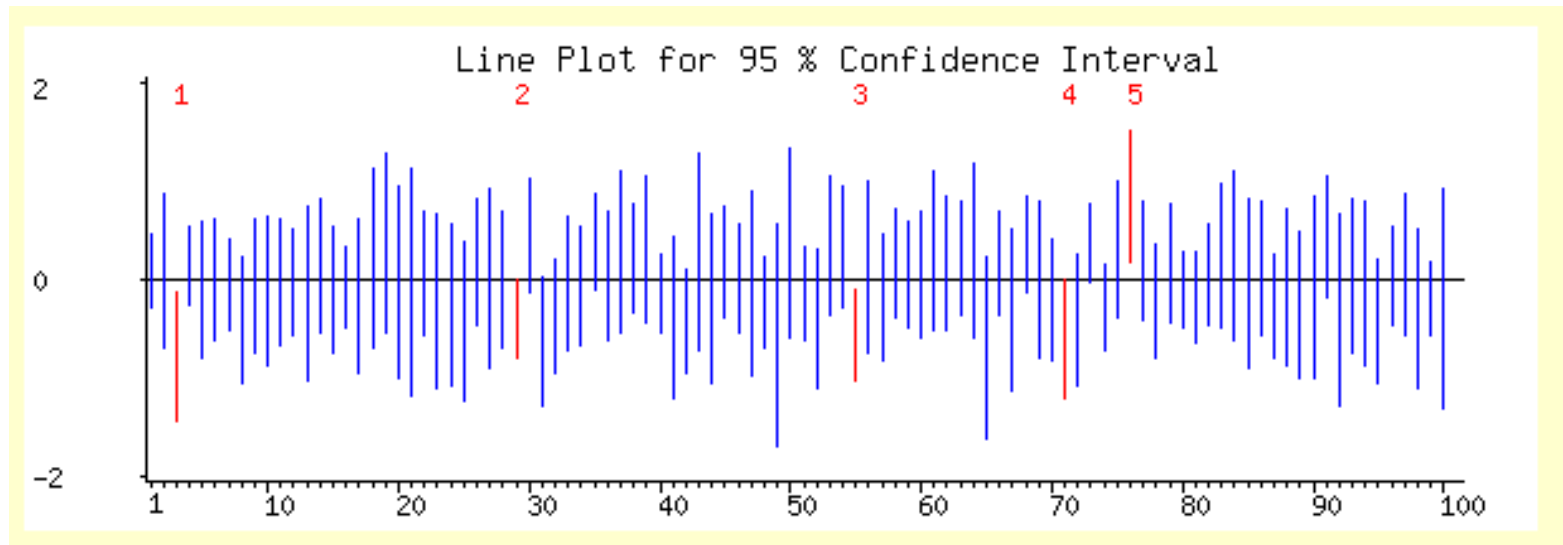
$\lambda(t)$  는  $(t, t + \Delta t)$  사이에 고장을 일으킬 확률이다.

3. MTTF(Mean time to Failure) : 기대수명 혹은 평균수명



# Review

1. 신뢰도(reliability) – 특정시간까지 고장없이 기능을 수행할 확률
2. 신뢰수준(Confidence level) - 신뢰수준은 실제 모수(parameter=모평균,  $\alpha$ ,  $\lambda$ )를 추정하는데 몇 퍼센트의 확률로 그 신뢰구간이 실제모수를 포함하게 될것이나 하는 것이다. 예를 들어 모평균( $\mu$ )의 추정을 위해 100번의 sampling을 통해 표본평균과 표본분산을 구하여 100개의 신뢰구간을 얻었을때, 그 100개의 신뢰구간중 95개에 모평균( $\mu$ )이 포함되게 설정된 신뢰구간을 95% 신뢰구간이라고 한다.



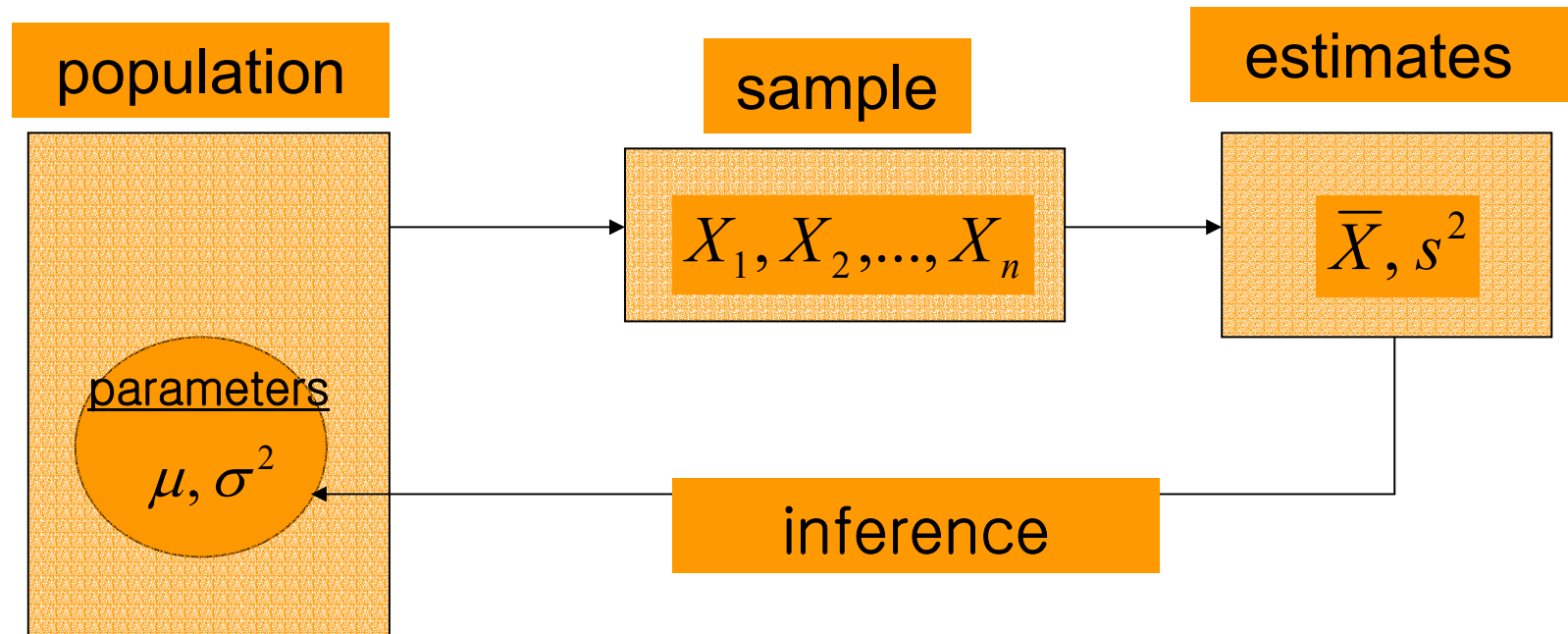


# 목 차

1. Descriptive Statistics - Point Estimation(점추정)
2. Probability Distribution ( 확률분포 )
3. Interval Estimation (구간 추정)
4. 시료개수별 시험거리 산출방법



# 1. Descriptive Statistics



- Parameter :  $\mu, \sigma^2$
- Estimates
  - point estimates :  $\bar{X}, s^2$
  - interval estimates :  $\left| \longleftrightarrow \right|$



# Point Estimation (점 추정)

## • Measure for Location

(1) Mean (평균)  $\hat{\mu} = \bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i$

(2) Median (중앙값)

(3) Mode (최빈값)

(예) 2, 3, 3, 4, 38      평균=10, 중앙값=3, 최빈값=3

## • Measure for Variation

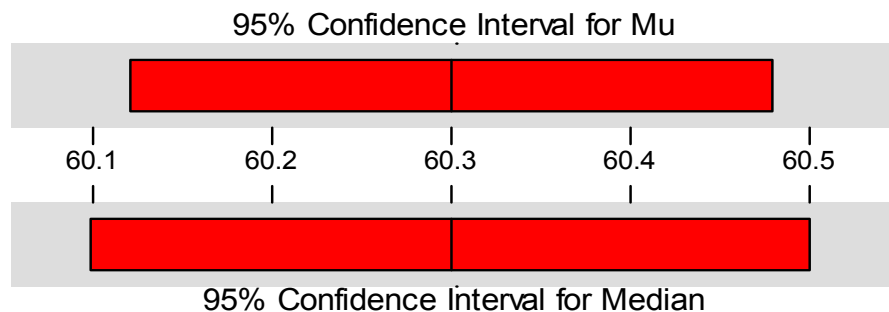
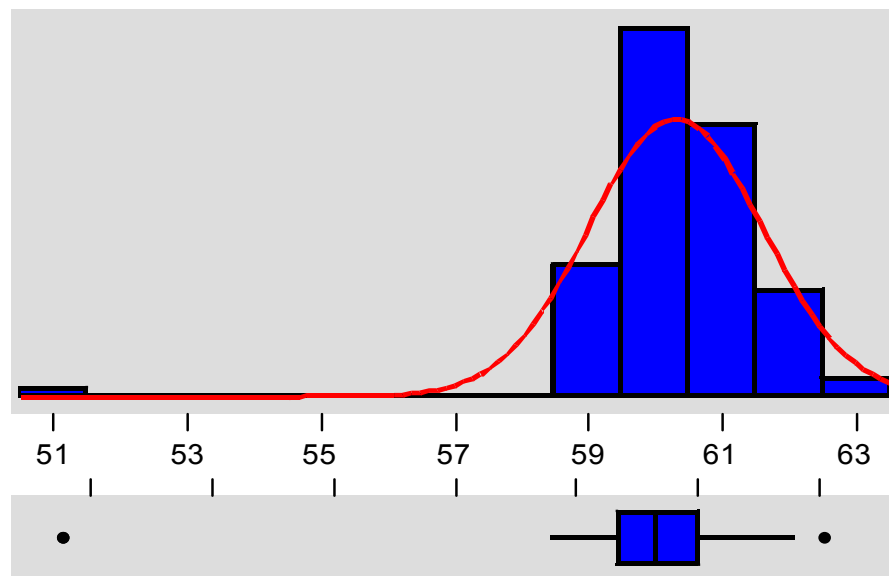
(1) Variance (분산)  $\hat{\sigma}^2 = s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

(2) Standard Deviation(표준편차)      s



# 기술통계량 by Minitab

## Descriptive Statistics



Variable: BP

Anderson-Darling Normality Test

A-Squared: 6.609  
P-Value: 0.000

Mean 60.2995  
StDev 1.3021  
Variance 1.69556  
Skewness -3.84685  
Kurtosis 28.5218  
N 206

Minimum 50.6000  
1st Quartile 59.7000  
Median 60.3000  
3rd Quartile 61.0000  
Maximum 63.1000

95% Confidence Interval for Mu

60.1206 60.4784

95% Confidence Interval for Sigma

1.1874 1.4417

95% Confidence Interval for Median

60.1000 60.5000



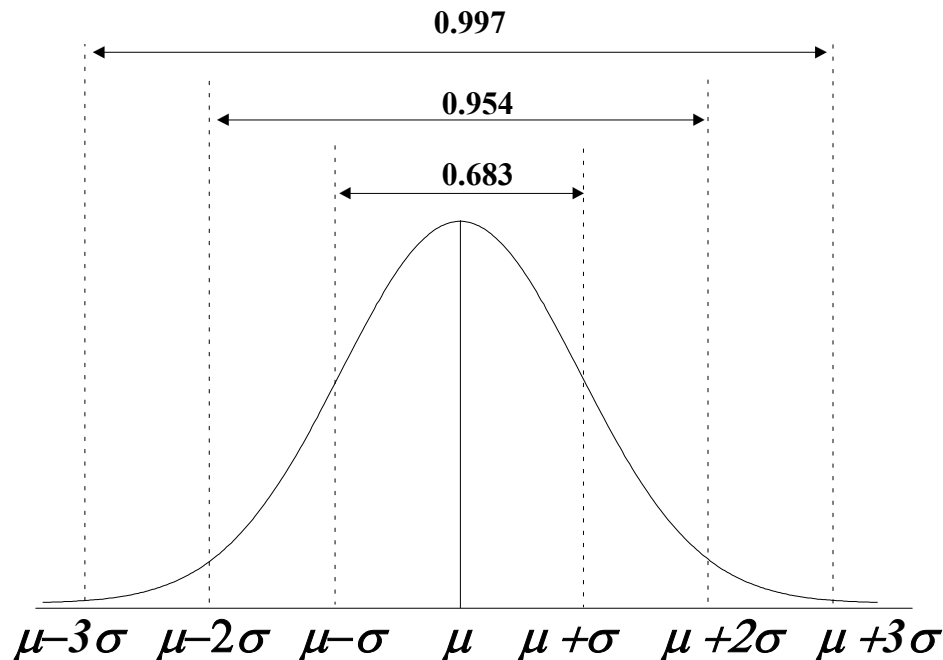


## 2. Probability Distribution ( 확률분포 )

### 😊 정규분포 (Normal Distribution)

정규분포는 평균을 중심으로 대칭을 이루는 종모양의 연속확률분포이다

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





## 2. Probability Distribution (cont'd)

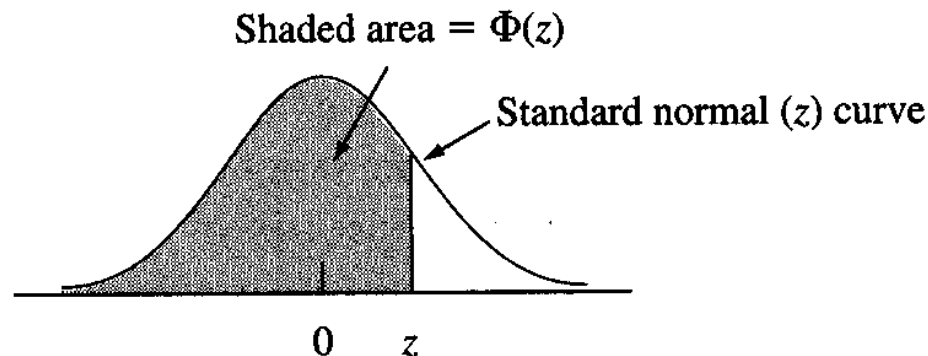
### ☺ 표준정규분포 (Standard Normal Distribution)

The normal distribution with parameter values  $\mu = 0$  and  $\sigma=1$  is called a **standard normal distribution**. A random variable that has a standard normal distribution is called a **standard normal random variable** and will be denoted by  $Z$ . The pdf of  $Z$  is

$$f(z;0,1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty$$

the cdf of  $Z$  is

$$P(Z \leq z) = \int_{-\infty}^z f(y;0,1)dy \quad \text{is denoted by } \Phi(z).$$



**Figure 4.14** Standard normal cumulative areas tabulated in Appendix Table A.3



## 2. Probability Distribution (cont'd)

☺ 지수분포 (Exponential Distribution) -  $\lambda$  ( $\alpha=1$ 일때)

☺ 대수정규분포 (lognormal Distribution) -  $\mu, \sigma^2$

☺ 와이블분포 (Weibull Distribution) -

$\alpha$ : (shape parameter)     $\lambda$ : (scale parameter)



모든 분포함수에 있어서는 **모수 추정**이 필요하다.

고장 데이터로부터 적합한 분포를 찾고 그에 대한 모수를 추정한다.



### 3. Interval Estimation (구간 추정)

An Interval Estimates states the range within which a population parameter probably lies.

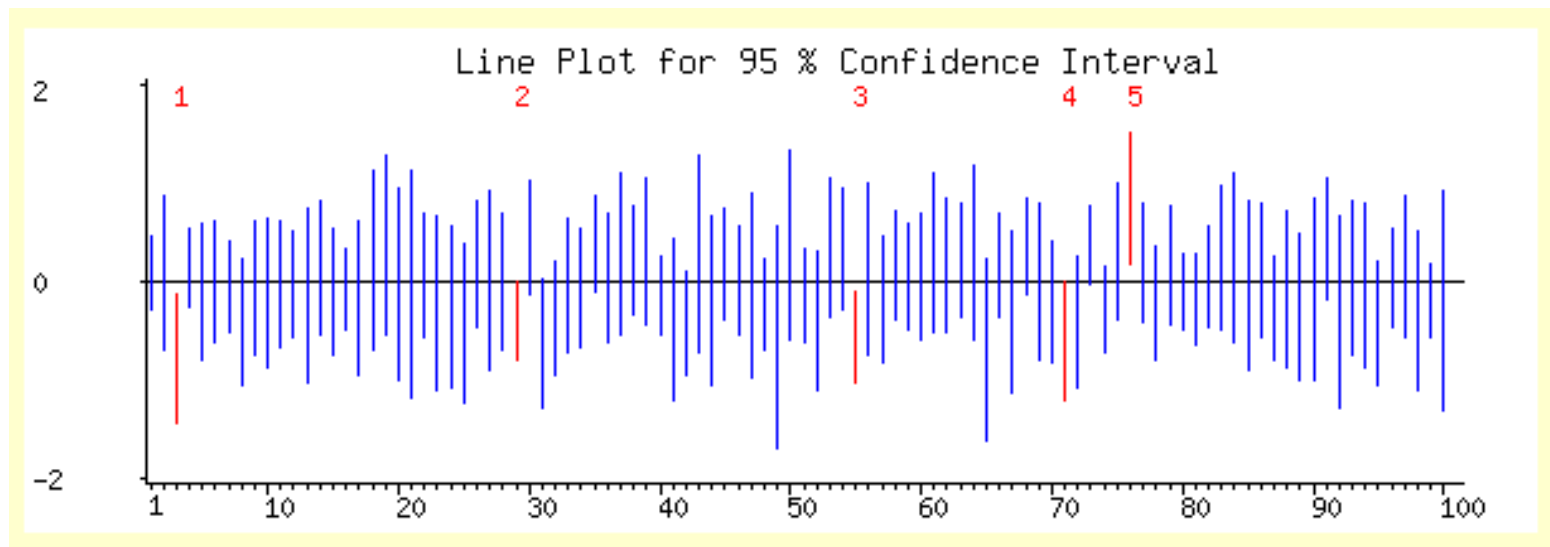
$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (\text{혹은} \quad \bar{X} \pm t_{\alpha/2(n-1)} \frac{s}{\sqrt{n}})$$

모수의 추정치를 하나의 값으로 나타내기 보다는 실제 모수가 존재할 가능성이 높은 구간을 제시하는 방식이 보다 바람직할 수 있는데, 이를 구간 추정이라 한다.

➡ 신뢰수준 (Confidence Level) : 90%, 95%, 99%

## • Confidence Interval

신뢰구간은 실제 모수(parameter=모평균, 모분산등)를 추정하는데 몇 퍼센트의 확률로 그 신뢰구간이 실제모수를 포함하게 될것이나 하는 것이다. 예를 들어 모평균( $\mu$ )의 추정을 위해 100번의 sampling을 통해 표본평균과 표본분산을 구하여 100개의 신뢰구간을 얻었을때, 그 100개의 신뢰구간중 95개에 모평균( $\mu$ )이 포함되게 설정된 신뢰구간을 95% 신뢰구간이라고 한다.





신뢰구간은 신뢰수준이 높고 구간의 크기는 좁도록 정하는 것이 바람직하나, 일반적으로 신뢰수준이 높아지면 신뢰구간도 넓어진다. 신뢰수준 90%, 95%, 99% 등이 어느 정도 절충된 형태라 할 수 있겠다. 신뢰수준을 95%에서 99%로 4% 올리기 위해서는 신뢰구간을 약 1.3배로 하여야 함을 알 수 있다. 한편 동일한 신뢰수준에서 신뢰구간의 크기는 표본평균의 표준편차와 관련되므로 신뢰구간을 좁게 하기 위해서는 표본 수를 늘려야 한다.

예 2) 스테인레스 냉연코일의 평균 경도를 알아보기 위하여 7 개의 표본을 추출하여 경도를 측정한 결과 다음과 같은 값을 얻었다.

84.9, 83.4, 81.0, 82.7, 83.4, 85.4, 82.9

표준편차는 1.5. 95% 신뢰수준에서 이 코일의 평균 경도에 대한 신뢰구간을 구하라.

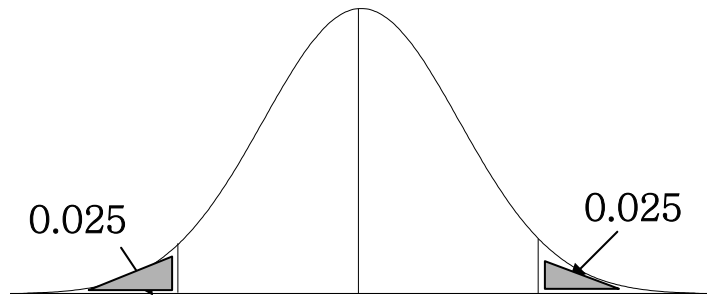
풀이)  $\bar{X}=83.386, z_{\alpha/2} = z_{0.025} = 1.96$  이 되므로

$$\begin{aligned}\bar{X} - z_{0.025} \frac{\sigma}{\sqrt{n}} &\leq \mu \leq \bar{X} + z_{0.025} \frac{\sigma}{\sqrt{n}} \\ \Rightarrow 83.386 - 1.96(1.5/\sqrt{7}) &\leq \mu \leq 83.386 + 1.96(1.5/\sqrt{7}) \\ \Rightarrow 83.386 - 1.111 &\leq \mu \leq 83.386 + 1.111\end{aligned}$$

이고, 평균 경도에 대한 신뢰구간은 (82.275, 84.497)이다. 즉 평균경도가 (82.275, 84.497) 사이에 있을 것이라는 주장을 95% 신뢰할 수 있다.

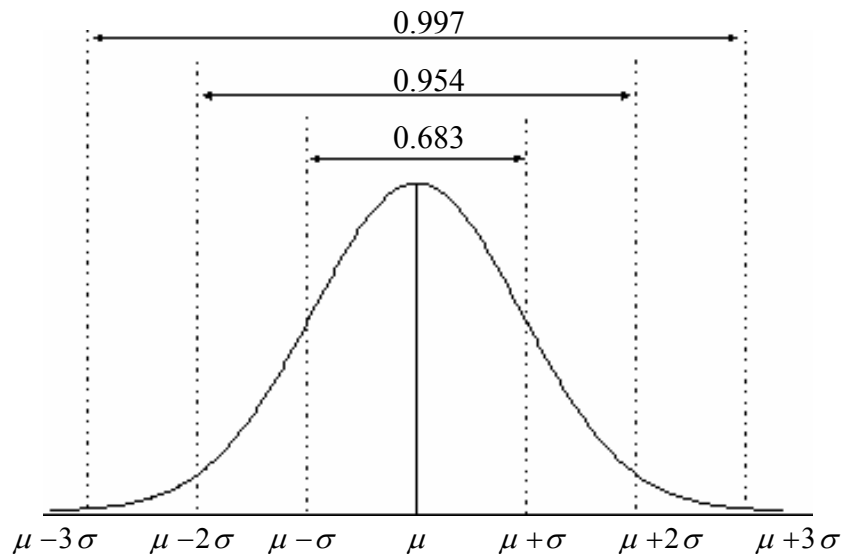
- 데이터가 정규분포일때

## 95% 신뢰구간



$$P\{\bar{X} - d \leq \mu \leq \bar{X} + d\} = 0.95$$

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$





## (참조) 표준정규분포 확률표

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0986	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143



# 4. 시료개수별 시험거리 산출방법

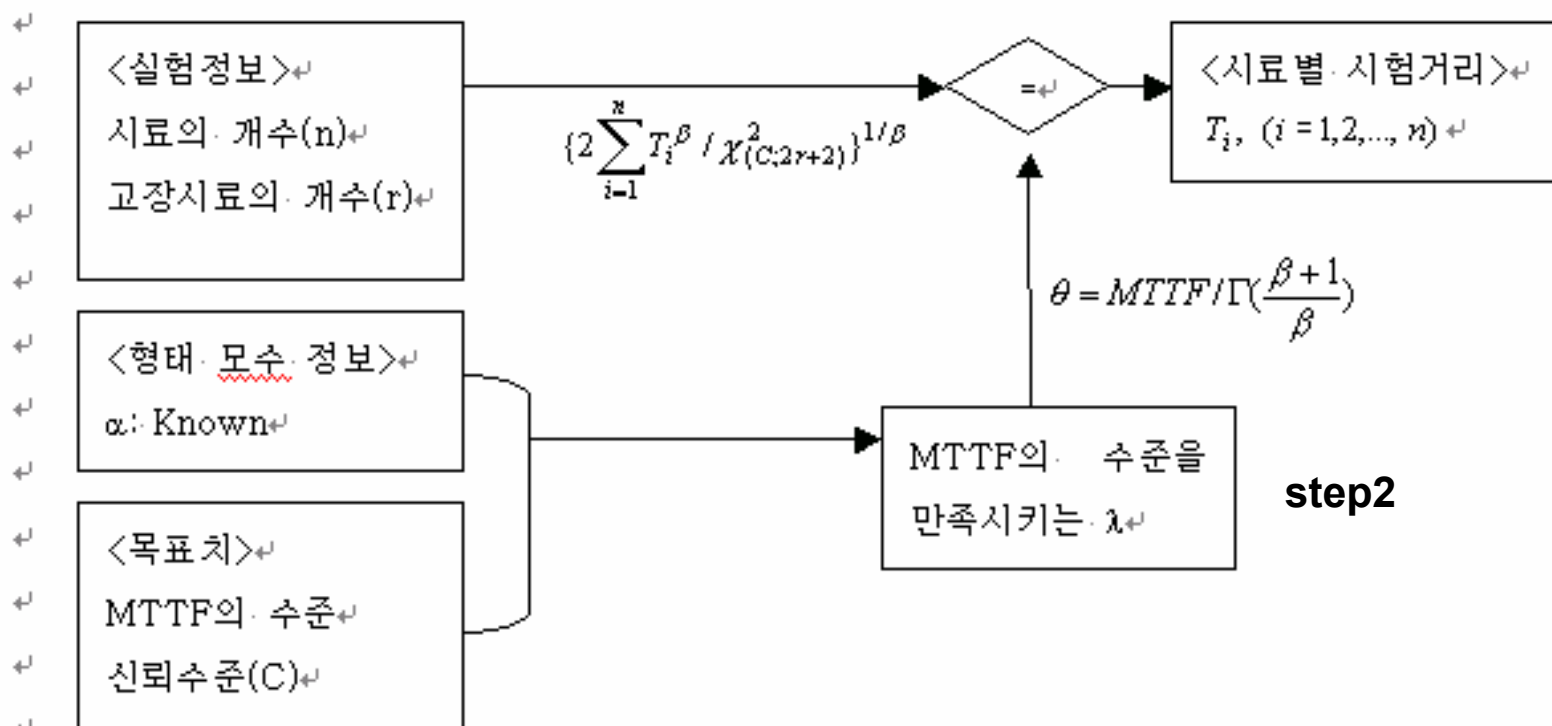
● → The lower C% confidence interval for weibull scale parameter  $\theta$  is

$$\theta = \left\{ 2 \sum_{i=1}^n T_i^\beta / \chi^2_{(C; 2r+2)} \right\}^{1/\beta} \dots \text{for } r \geq 0$$

척도모수:  $(\lambda) \theta$

형태모수:  $(\alpha) \beta$

## 2. → Procedure

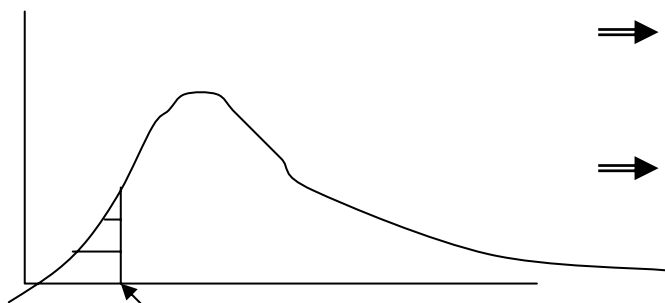


(예제) : 수문개폐 작동 이동거리로 243km이상을 설정, 와이블분포의 형태모수( $\beta$ )는 2로 알려짐.  
 시료 2개를 실험할때 **confidence level**(신뢰수준) 90% 신뢰하한은 얼마인가?

Step1 :  $MTTF = \theta \Gamma[(\beta + 1)/\beta] = \theta \Gamma(3/2)$

Step2 :  $MTTF = 243$ 을 만족시키는  $\theta$ 를 구한다.  $\theta = MTTF / \Gamma(3/2) = 243 / 0.8862 = 274.2$

Step3 :  $\theta = \{2 \sum_{i=1}^n T_i^\beta / \chi^2_{(C; 2r+2)}\}^{1/\beta}$  에 따라 계산하면 다음과 같다.



$\chi^2_{(0.1,2)} = 4.61$

$\Rightarrow 2 \sum_{i=1}^n T_i^\beta = \theta^\beta \chi^2_{(0.1;2)}$

$\Rightarrow$  여기서  $\chi^2_{(0.1,2)} = 4.61$ 이고 고장데이터의 수는 0이므로  $n=2, \beta=2$ 이다. ( $t_1=t_2$ )

$\Rightarrow 2(t_1^2 + t_2^2) = \theta^\beta \chi^2_{(0.1;2)} \Rightarrow 4t^2 = \theta^2 4.61$



$$t = \left( \frac{\theta^2 4.61}{4} \right)^{\frac{1}{2}} = \left( \frac{274.2^2 4.61}{4} \right)^{\frac{1}{2}} = 294.4$$

⇒ 신뢰수준(Confidence level) 90%에서 243km의 MTTF를 보장하기 위해서는 시료 2개에 대해서 평균 294.4km수명이 보장되어야 한다.





# 예제 : 온간단조의 다이수명 예측 및 합격기준

예제(온간단조 현대의 데이터로 추정, 여기서 H-C는 제외)

$\alpha$ : 1.792 (형태모수)  $\lambda$ : 0.0003 (척도모수), MTTF= 24782

H-A	KCW1	균열	16,585	25,868	11,786	13,217	25,172	19,966	21,067	12,601	41,555	16,969
H-B	KCW1	균열	21,549	3,534	11,262	18,937	69,713					
H-C	KCW1	균열	2,993	6,642	55,985							
H-D	KCW1	균열	6,401	37,424	43,029	23,986	28,805	50,264	27,762	25,383	19,836	

분석목적 : 목표MTTF에 부합하기 위한 합격기준 산정

1. Weibull, Lognormal분포에 적합시켜 모수를 추정
2. 목표 MTTF를 정함- 여기서는 20000으로 설정
3. 신뢰수준(confidence level)을 정함 – 여기서는 90%, 95%, 99%
4. 시료개수에 따른 평균고장시간을 제시 –예를 들어 3개의 시료를 실험할 경우 어느시점에서 1개가 고장나면 나머지 두개에 대한 합격기준이 제시된다. 즉, 실험개수와 실험도중 고장난 시기에 따라 생존한 부품의 수명에 대한 합격기준은 높아진다.

# 온간단조의 다이수명 예측 및 합격기준 (예제)

실험개수 <sub>i</sub>	고장개수 <sub>i</sub>	고장시간 <sub>i</sub>	신뢰수준(%) <sub>i</sub>		
$n_i$	$r_i$	$T(i)$	0,1(90%) <sub>i</sub>	0,05(95%) <sub>i</sub>	0,01(99%) <sub>i</sub>
1 <sub>i</sub>	0 <sub>i</sub>		35811,51 <sub>i</sub>	41476,23 <sub>i</sub>	52723,98 <sub>i</sub>
2 <sub>i</sub>	0 <sub>i</sub>		24324,12 <sub>i</sub>	28171,75 <sub>i</sub>	35811,51 <sub>i</sub>
2 <sub>i</sub>	1 <sub>i</sub>	23000 <sub>i</sub>	40324,44 <sub>i</sub>	46679,94 <sub>i</sub>	58786,28 <sub>i</sub>
3 <sub>i</sub>	0 <sub>i</sub>		19398,67 <sub>i</sub>	22467,18 <sub>i</sub>	28559,95 <sub>i</sub>
3 <sub>i</sub>	1 <sub>i</sub>	18000 <sub>i</sub>	29322,24 <sub>i</sub>	33438,08 <sub>i</sub>	41382,14 <sub>i</sub>
3 <sub>i</sub>	2 <sub>i</sub>	28000 <sub>i</sub>	42802,49 <sub>i</sub>	49702,60 <sub>i</sub>	62538,29 <sub>i</sub>
4 <sub>i</sub>	0 <sub>i</sub>		16521,59 <sub>i</sub>	19134,99 <sub>i</sub>	24324,12 <sub>i</sub>
4 <sub>i</sub>	1 <sub>i</sub>	15000 <sub>i</sub>	24133,73 <sub>i</sub>	27343,89 <sub>i</sub>	33576,01 <sub>i</sub>
4 <sub>i</sub>	2 <sub>i</sub>	22000 <sub>i</sub>	32529,60 <sub>i</sub>	36861,67 <sub>i</sub>	45093,93 <sub>i</sub>
4 <sub>i</sub>	3 <sub>i</sub>	32000 <sub>i</sub>	44166,87 <sub>i</sub>	51567,76 <sub>i</sub>	65086,91 <sub>i</sub>
5 <sub>i</sub>	0 <sub>i</sub>		14587,22 <sub>i</sub>	16894,64 <sub>i</sub>	21476,22 <sub>i</sub>
5 <sub>i</sub>	1 <sub>i</sub>	14500 <sub>i</sub>	20650,30 <sub>i</sub>	23375,35 <sub>i</sub>	28670,15 <sub>i</sub>
5 <sub>i</sub>	2 <sub>i</sub>	20000 <sub>i</sub>	26606,33 <sub>i</sub>	29999,82 <sub>i</sub>	36477,47 <sub>i</sub>
5 <sub>i</sub>	3 <sub>i</sub>	25000 <sub>i</sub>	34041,26 <sub>i</sub>	38643,55 <sub>i</sub>	47256,33 <sub>i</sub>
5 <sub>i</sub>	4 <sub>i</sub>	34000 <sub>i</sub>	44657,91 <sub>i</sub>	52562,44 <sub>i</sub>	66766,82 <sub>i</sub>

설명 뒤쪽



## 온간단조의 다이수명 예측 및 합격기준 (예제설명)

목표 MTTF 20000번을 설정했을때 실험 부품의 개수가 3개에 대해서 90% 신뢰수준을 만족할수 있는 기준을 제시해보면 다음과 같다.

- i) 3개의 부품이 모두 19399번까지 작동한다면 이부품은 90% 신뢰수준에서 수명이 20000번 이상을 보장할수 있다고 할 수 있다.
- ii) 그런데, 부품 1개가 19399번 전인 18000번에서 고장이 발생하면 나머지 2개가 모두 29323번(혹은 평균이)까지는 작동해야 90% 신뢰수준에서 부품의 수명이 20000번 이상을 보장할수 있다고 할 수 있다.
- iii) 만일 2개중 1개마저 29323번 전인 28000번에서 고장이 발생했다고 가정하면, 나머지 1개는 최소한 42803번까지 작동을 해야 부품의 수명이 20000번 이상임을 주장할 수 있다.

즉, 신뢰도와 신뢰수준의 차이를 구별하여 설명하면 3개의 부품이 고장없이 모두 19398까지 작동한다면, 그 제품은 **90% 신뢰수준으로** 목표MTTF 20000을 달성한다는 것이다. 여기서 90% 신뢰수준이란 random하게 부품을 10000개 시험하였을때 그 중 10%만 목표MTTF 20000을 달성하지 못할것이라는 것을 의미한다.



(1) 미국의 시의원의 평균연봉을 예측할 때 그 오차수준을 100\$로 하고 신뢰수준은 95%로 한다고 하면 몇 명의 시의원에 대한 조사를 해야하는지 계산할수 있다.

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

여기에서 Error Bound 100으로 정하면  $z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 100$  이라는 의미이다 .  
따라서  $n = [(1.96*s)/100]^2$ 으로 계산된다. 여기서  $s=1000\$$ 로 가정하면  
 $n=385$ 명(384.16)이 된다

(2) A후보자 득표율 24% (허용오차  $\pm 2\%$ , 95% 신뢰수준)이라고 하면, 그 의미는  
 $n$ 명을 표본으로 하여 random하게 반복 조사하였을때 100번중 95번은 22%~26%의  
득표율을 가질것이라는 의미이다.

(Ref : Scheaffer, Mendenhall and Ott, *Elementary Survey Sampling*, 1990)

$n = Npq / ((N-1)D + pq)$   $n$ =sample size,  $N$ =population size

$p, q=1-p$

$D = (\text{bound error})^2 / (1.96)^2$