CRE : Certified Reliability Engineer

Ch.2 Basic Statistical Concepts.

中国可靠性
KekaoXing.com

中国最专业、最有影响力的可靠性行业网站

Industrial Engineering & Management Systems Research Center

https://m.kekaoxing.com/topic/cre

Key Data Definitions

■ Data Collection, Analysis and Reporting : Some guidelines

1. Formulate a clear statements of the problem.
2. Define precisely what is to be measured.
3. Carefully select the right measurement techniques.
4. List all the important characteristics to be measured.
5. Construct an uncomplicated data form.
6. Decide who will collect the data.
7. Arrange for an appropriate sampling method.
8. Decide who will analyze and interpret the results.
9. Decide who will report the results.

Key Data Definitions.

[CRE Primer Ⅲ 2]

Key Data Definitions.

- Statistic : A numerical data measurement taken from a sample that may be used to make an inference about a population. (e.g. $\bar{x}$, V, s, R, . . .)

- Parameter : The true population value, often unknown, estimated by statistic. (e.g. μ, σ^2, . . .)

- A Continuous Distribution : Contains infinite (variable) data points that may be displayed on a continuous measurement scale.
  (e.g. normal, exponential, Weibull distributions, . . .)

- A Discrete Distribution : Results from countable (attribute) data that has a finite number of possible values.
  (e.g. binominal, Poisson and hypergeometric distributions.)

Basic Statistics.

[CRE Primer Ⅲ 3-6]

■ Measures of Central Tendency.

- The Mean ($\bar{x}$) : $\bar{x} = \frac{\sum_{i}^{n} x_i}{n}$

  (where, x : each number, n : the sample size)

- The Mode ($Mode$, M_o) : The mode is the most frequently occurring number in a data set.

- The Median (M_e, $\tilde{x}$) : The median is the middle value when the data is arranged in ascending or descending order.

Basic Statistics.

[CRE Primer Ⅲ 3-6]

■ Measures of Central Tendency.

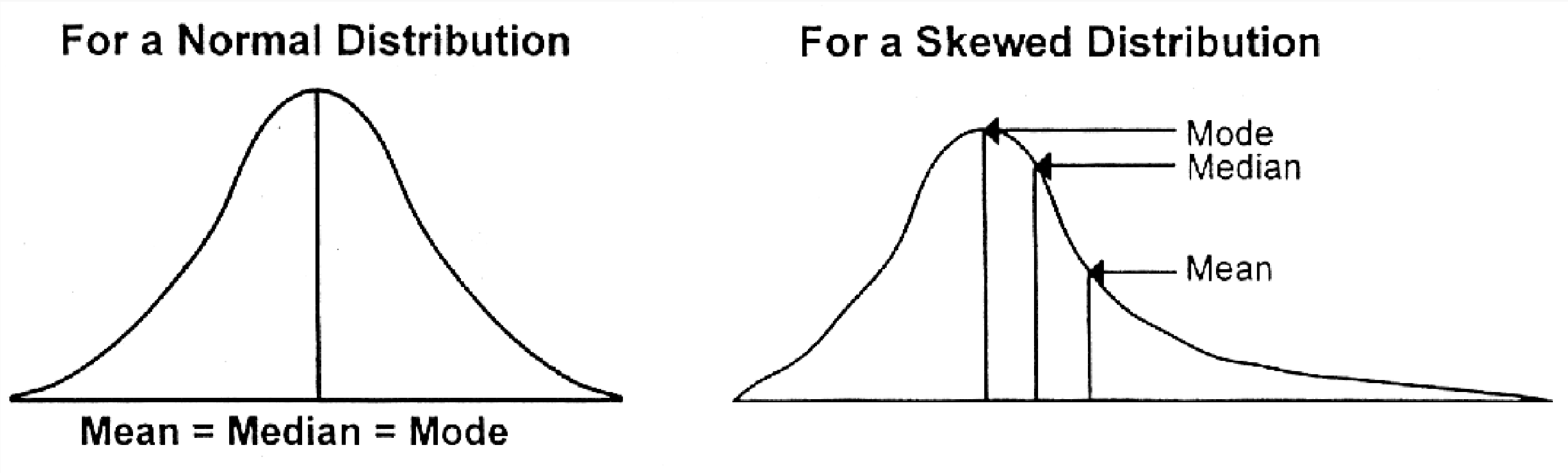


Figure 1. Mean, Mode and Median

Basic Statistics.

[CRE Primer Ⅲ 3-6]

■ The Central Limit Theorem.

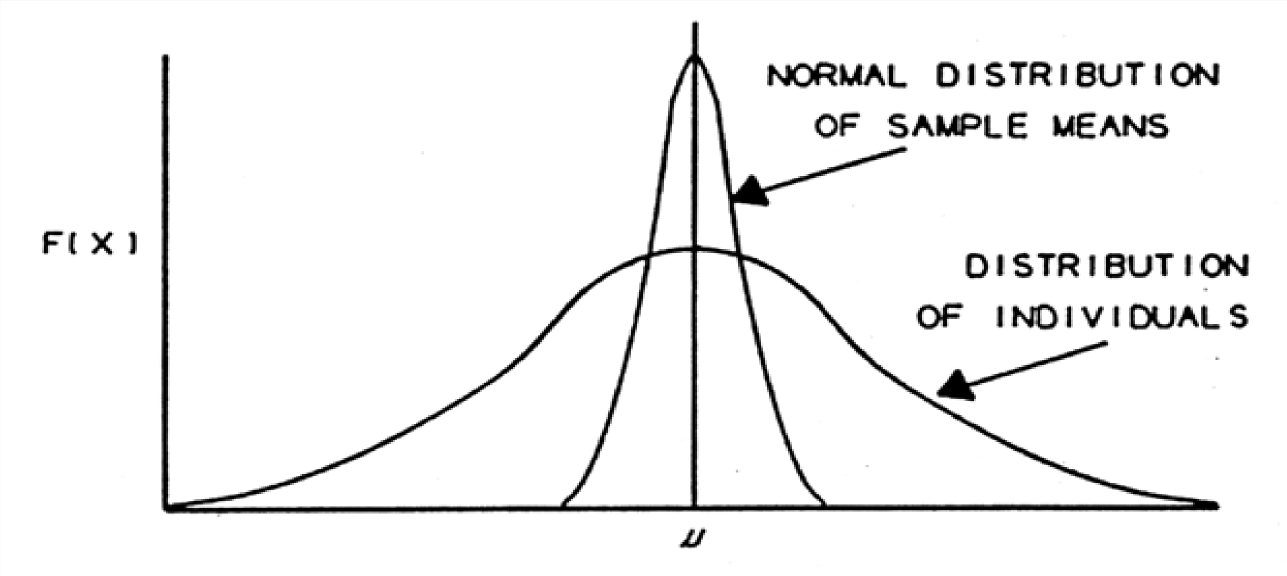


Figure 2. Distributions of Individuals vs. Means

- The sample means $\overline{x}_s$ will be more normally distributed around μ than individual readings x_s.

Basic Statistics.

[CRE Primer Ⅲ 7-9]

Measure of Dispersions.

- Range (R) : The range of a set of data is the difference between the largest and smallest values.

- Variance (σ^2, s^2) : The sum of squared deviation from the mean divided by the sample size.

Population : $\sigma^2 = \dfrac{\sum_{i=1}^{N}(X-\mu)^2}{N}$ Sample : $s^2 = \dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$

- Standard Deviation (σ, s) : The square root of the variance.

Basic Statistics.

[CRE Primer III 7-9]

Tchebysheff's Theorem

- $P(\,|\,X-\mu\,|\ \langle\ k\sigma\,)\ \geq\ 1-\frac{1}{k^2}$ where $k\ \rangle\ 1$
- This means that when we use $\overline{x}$ to estimate μ, at least 100(1 - (1 / k^2))% of the observed values will fall within $\mu + k\sigma$ regardless of what distribution it is.

Probability

[CRE Primer III 10-17]

Conditions for Probability.

- Event(E) : The probability of any event lies between 0 and 1.

- Sample space(S) : The sum of the probabilities of all possible events in a sample space = 1

- The Probability of E : $P(E) = \frac{n_E}{N}$

  (where N = the number of sample space, n_E = the number of event)

Probability.

[CRE Primer Ⅲ 10-17]

■ Compound Events.

1. Union of A and B (The Additive Law)

   - $P(A \cup B) = P(A) + P(B)$ (mutual exclusive events)
   - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (not mutually exclusive)

2. Intersection of A and B (The Multiplicative Law)

   - $P(A \cap B) = P(A)P(B)$ (events A and B are independent)
   - $P(A \cap B) = P(A)P(B \mid A)$ (events A and B are dependent)

Probability.

[CRE Primer Ⅲ 10-17]

■ Permutations & Combinations.

· Permutations : An ordered arrangement of n distinct objects

$$ {}_nP_r = n(n-1)(n-2)\cdot\cdot\cdot(n-r+1) = \frac{n!}{(n-r)!} $$

· Combinations : The number of distinct combinations of n distinct objects taken r at a time.

$$ {}_nC_r = \frac{n!}{r!(n-r)!} $$

Probability Density Functions.

[CRE Primer Ⅲ 18-29]

Probability Density Function.

- Continuous distributions

: $f(x) \geq 0, \quad x$ is real.

: $P(a \leq x \leq b) = \int_a^b f(x)dx$

: $\int_{-\infty}^{\infty} f(x)dx = 1$

- Discrete distributions

: $p(x) \geq 0, \quad x$ is real

: $P(X = x) = p(x)$

: $\sum_{i=0}^{k} f(x_i) = 1, \quad n(S) = k$

Probability Density Functions.

中国可靠性 KekaoXing.com

[CRE Primer Ⅲ 18-29]

■ Reliability & Hazard Functions.

- $R(x) = \int_x^{\infty} f(\tau)\ d\tau$

- $\lambda(x) = \dfrac{f(x)}{R(x)}$

  $\rightarrow R(x) = e^{-\int_0^x \lambda(x)\ dx}$

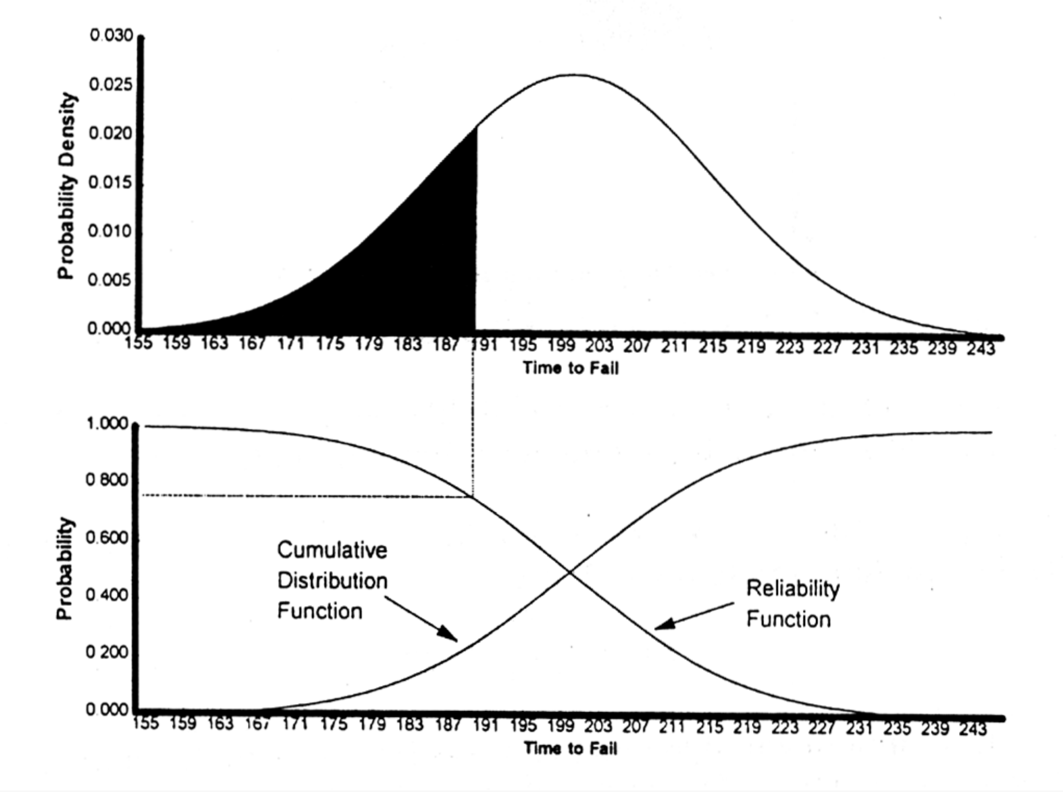
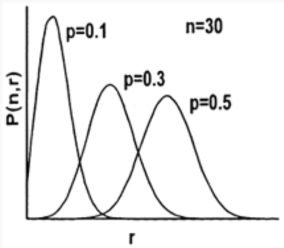
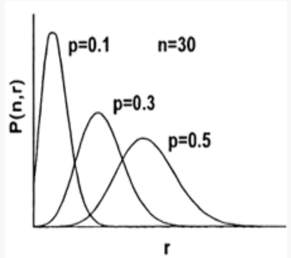
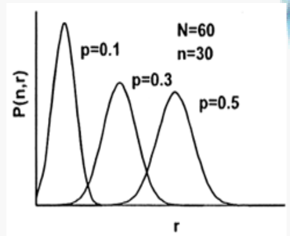


Figure 3. The Reliability Function.

Discrete Modeling Distribution.

[CRE Primer Ⅲ 60-71]

■ Common Discrete Distribution.

	Binomial	Poisson	Hypergeometric
Shape			
Formulas	$P(X) = {}_nC_r \ p^X q^{n-X}$ n = Sample size. X = Number of occurrences. p = Probability. $q = 1-p$	$P(X) = \frac{(n\bar{p})^X e^{-n\bar{p}}}{X!}$ n = Sample size. X = Number of occurrences. p = Probability. $n\bar{p}$ = μ = average.	$P(X)=\frac{\binom{d}{X}\binom{N-d}{n-X}}{\binom{N}{n}}$ n = Sample size. X = Number of occurrences. d = Occurrences in population N = Population size
Application	The binomial is an approximation to the hypergeometric. Sampling is with replacement. The Sample size is less than 10 % of N. The normal distribution approximates the binomial when np $\geq$ 5.	The Poisson is used as a distribution for defect counts and can be used as an approximation to the binomial. For np < 5 the binomial is better approximated by the Poission than the Normal.	Used when the number of defects (d) is known. Sampling is without replacement. The population size(N) is frequently small. Applied when the sample(n) is a relatively large proportion of the population.

Discrete Modeling Distribution.

[CRE Primer Ⅲ 60-71]

■ Binomial Distribution

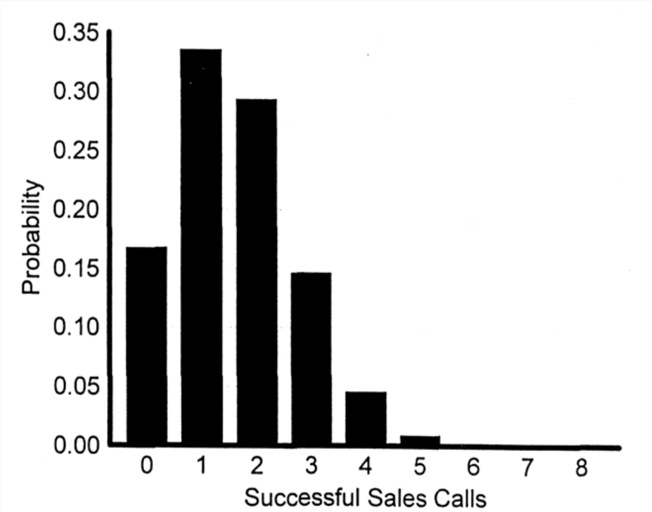
- It is used to model situation having only 2 possible outcomes, usually labeled as success or failure.

- The binomial probability density function.

$$p(x : n, p) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

- Mean and variance.

$$E(x) = np, \qquad V(x) = np(1-p)$$



Discrete Modeling Distribution.

[CRE Primer Ⅲ 60-71]

Poisson Distribution

- It is used to rates, such as rabbits per acre, defects per unit, or arrivals per hour.
- The Poisson probability density function.

$$p(x : \mu) = \frac{e^{-\mu}\mu^{x}}{x!}$$

- The term $p(x, \mu)$ represents the probability of exactly x occurrence in an interval having an average of μ occurrences.

- Mean and variance.

$E(x) = \mu,$ $V(x) = \mu$

Discrete Modeling Distribution.

[CRE Primer Ⅲ 60-71]

Hypergeometric Distribution

- It is similar to the binomial distribution. Both are used to model the number of successes given :
  1. A fixed number of trials, and
  2. Two possible outcome on each trial.

- The difference is that the binomial distribution requires the probability of successes to be the same for all trials, while the hypergeometric distribution does not.

$$p(x : N, n, m) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}} \qquad x = 0,\ 1,\ 2,\ \cdot\ \cdot\ \cdot\ \min(n, m)$$

$$E(x) = \frac{nm}{N} = np\ , \qquad V(x) = np(1-p)\left(\frac{N-n}{N-1}\right)$$

Discrete Modeling Distribution.

[CRE Primer Ⅲ 60-71]

Geometric Distribution

- The geometric distribution requires exactly 1 success, and the random variable is the number of trials required to obtain the first success.

- The geometric distribution.

$$p(x : p) = p(1-p)^{x-1}$$

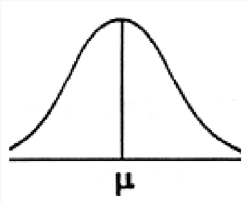
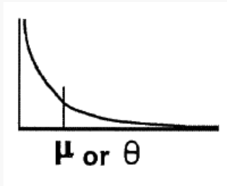
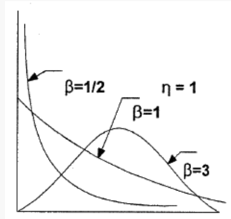
- Mean and variance.

$$E(x) = \frac{1}{p} \quad , \quad V(x) = \frac{1-p}{p^2}$$

Continuous Modeling Distribution.

[CRE Primer Ⅲ 30-59]

■ Common Continuous Distributions.

	Normal	Exponential	Weibull
Shape			
Formulas	$P(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$ μ = Mean σ = Standard deviation.	$P(X) = \frac{1}{\mu} e^{-\frac{X}{\mu}}$ μ = Mean	$P(X) = \frac{\beta}{\theta}\left(\frac{X-\delta}{\theta}\right)^{\beta-1} e^{-\left(\frac{X-\delta}{\theta}\right)^{\beta}}$ θ = Scale parameter. β = Shape parameter. δ = Location parameter.
Application	Numerous applications. Useful when it is equally likely that readings will fall above or below the average.	Describes constant failure rate conditions. Applies for the useful life cycle of many products. Often, time (t) is used for X.	Used for many reliability applications. Can test for the end of the infant mortality period. Can also describe the normal and exponential distribution.

Continuous Modeling Distribution.

[CRE Primer Ⅲ 30-59]

■ Weibull Distribution.

- The Weibull distribution is one of the most widely used distributions in reliability. It is commonly used to model time to fail, time to repair and material strength. There are two common versions of the Weibull distribution used in reliability. The two parameter Weibull and three parameter Weibull. The difference is the three parameter Weibull distribution has a location parameter when there is some non zero time to first failure.

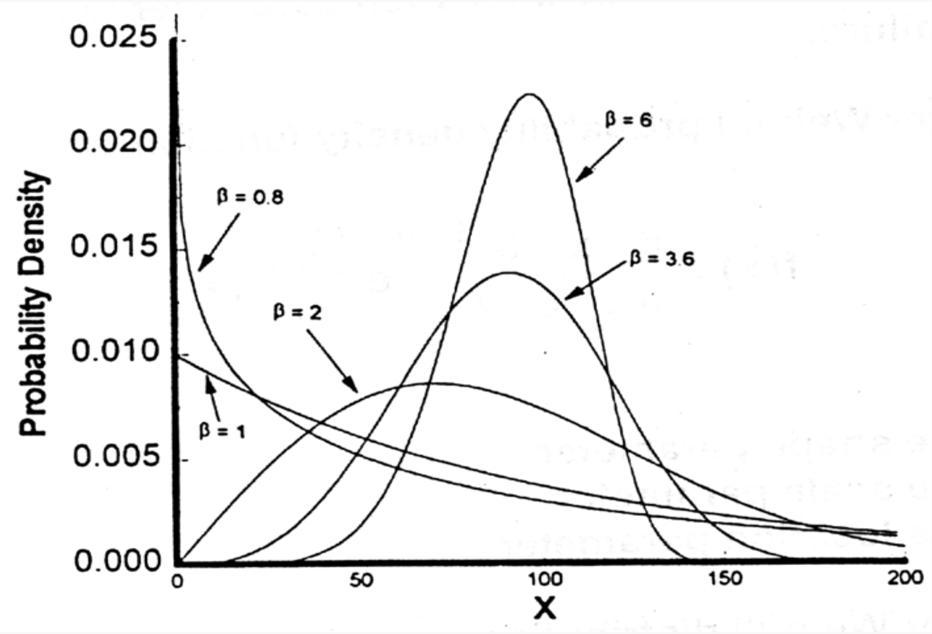
- $f(x) = \frac{\beta}{\theta}\left(\frac{x-\delta}{\theta}\right)^{(\beta-1)} e^{-\left(\frac{x-\delta}{\theta}\right)^{\beta}}$, $x \geq \delta$

  where β : the shape parameter, θ : the scale parameter.
  δ : the location parameter.

Continuous Modeling Distribution.

[CRE Primer Ⅲ 30-59]

■ Weibull Distribution : Shape Parameter.

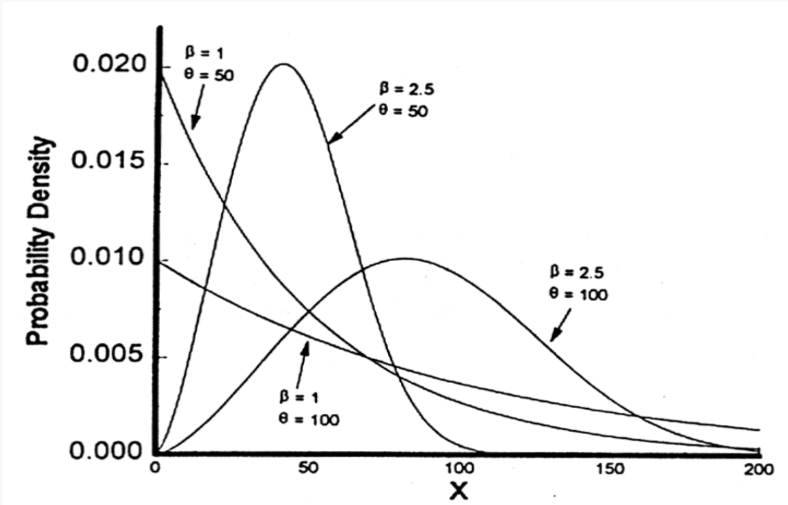


- If β = 1, the Weibull distribution is identical to the exponential distribution.
- If β = 2, the Weibull distribution is identical to the Rayleigh distribution.
- If β is between 3 and 4, the Weibull distribution approximates the normal distribution.

Continuous Modeling Distribution.

[CRE Primer Ⅲ 30-59]

■ Weibull Distribution : Scale Parameter.

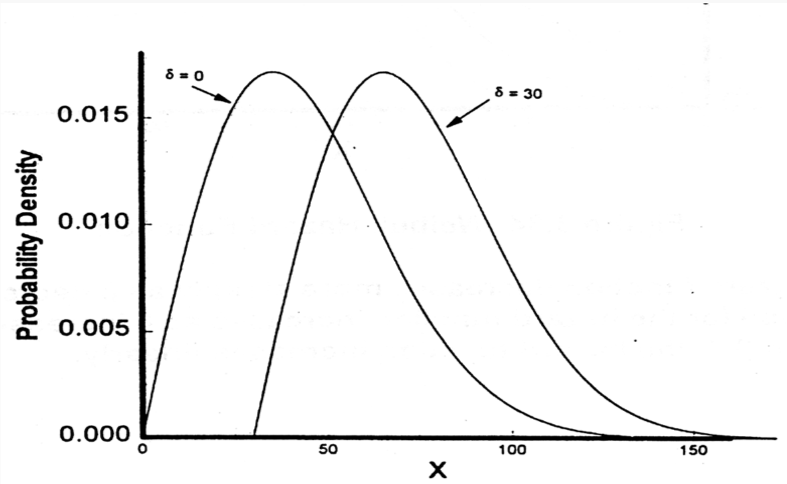


- The scale parameter is also known as the *characteristic life* if the location parameter is equal to zero.
- If δ does not equal zero, the characteristic life is equal to $\theta + \delta$; 63.2% of all value fall below the characteristic life regardless of the value of the shape parameter.

Continuous Modeling Distribution.

[CRE Primer Ⅲ 30-59]

■ Weibull Distribution : Location Parameter.



- The probability of failure when x is less than δ is zero.
- When $\delta > 0$, there is a period when no failures can occur.
- When $\delta < 0$, failures have occurred before time equal 0.
- A negative location parameter is caused by shipping failed units, failures during transportation, and shelf life failures.

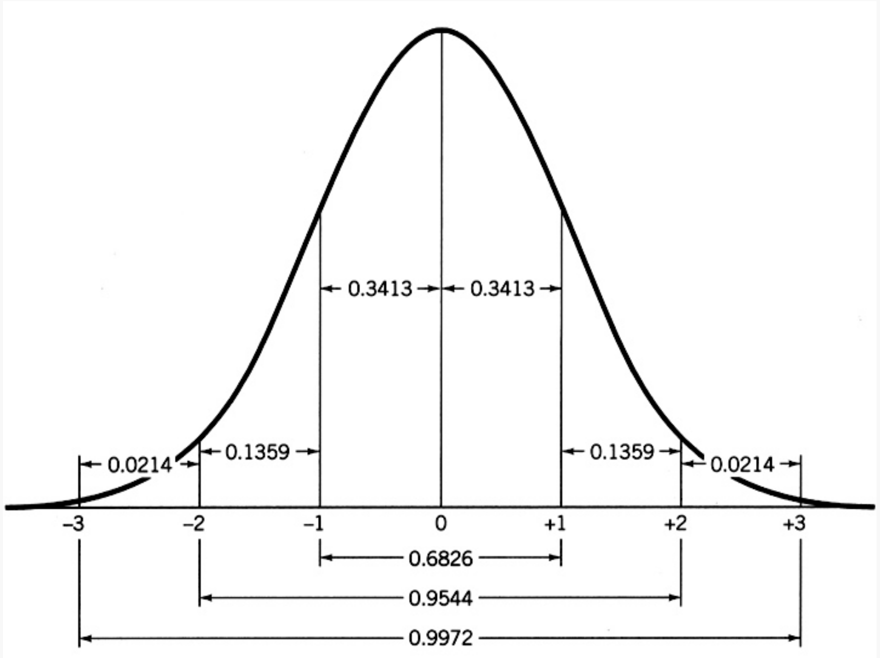
Continuous Modeling Distribution.

[CRE Primer Ⅲ 30-59]

■ Normal Distribution

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

where, μ : the mean, σ : the standard deviation.



Continuous Modeling Distribution.

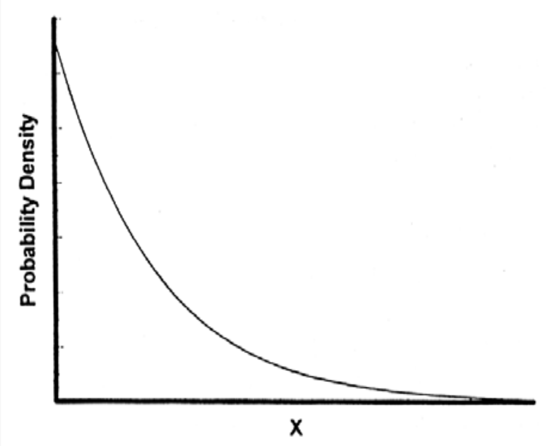
[CRE Primer Ⅲ 30-59]

Exponential Distribution

- The exponential distribution is used to model items with a constant failure rate, usually electronics. The exponential distribution is closely related to the Poisson distribution. If random variable, x is exponentially distributed, then the reciprocal of x $y=1/x$ follows a Poisson distribution. https://m.kekaoxing.com

- $$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}} = \lambda e^{-\lambda x}, \quad x \geq 0$$

  where, λ = the failure rate. θ = the mean



- Mean = $\frac{1}{\lambda}$
- Variance = $\sigma^2 = \frac{1}{\lambda^2}$

Continuous Modeling Distribution.

[CRE Primer Ⅲ 30-59]

※ Estimation of the mean

- $\theta = \frac{T}{r}$ where, T = the total test time for all items.

  r = the number failed.

- A (1 - α)% confidence limit for θ

$$\frac{2T}{\chi^2_{\frac{\alpha}{2},\,2r+2}} \le \theta \le \frac{2T}{\chi^2_{1-\frac{\alpha}{2},\,2r}} \qquad \text{(Time censoring)}$$

$$\frac{2T}{\chi^2_{\alpha,\,2r}} \le \theta \le \frac{2T}{\chi^2_{1-\alpha,\,2r}} \qquad \text{(Failure censoring)}$$

Continuous Modeling Distribution.

[CRE Primer Ⅲ 30-59]

■ Lognormal Distribution.

- Transforming the data by taking a logarithm yields a data set that is approximately normally distributed.

Lognormal data	Normal data	IN
12	ln(12)	2.48
16	ln(16)	2.77
28	ln(28)	3.33
48	ln(48)	3.87
87	ln(87)	4.47
143	ln(143)	4.96

$$y = x_1 x_2 x_3 \quad \rightarrow \quad \ln y = \ln x_1 + \ln x_2 + \ln x_3$$

Continuous Modeling Distribution.

[CRE Primer III 30-59]

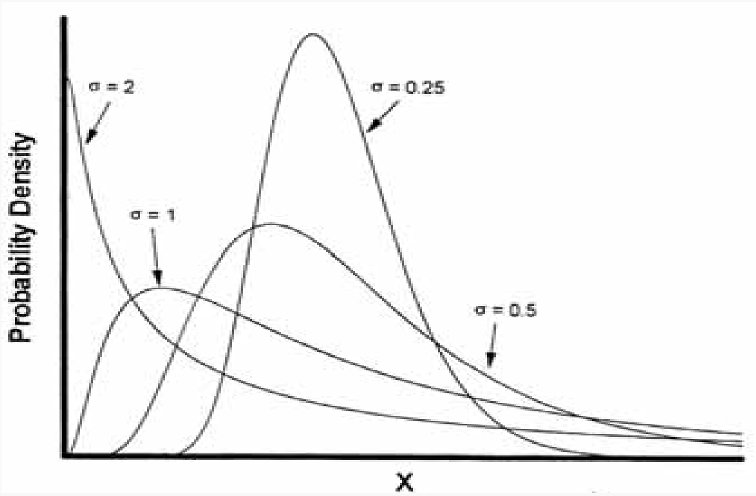
Lognormal Distribution.

- The lognormal probability density function :

$$f(x) = \frac{1}{x\ \sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2} \quad , x > 0$$

where, μ : the location parameter or log mean.

σ : the scale parameter or log standard deviation.



Continuous Modeling Distribution.

[CRE Primer Ⅲ 30-59]

Gamma Distribution.

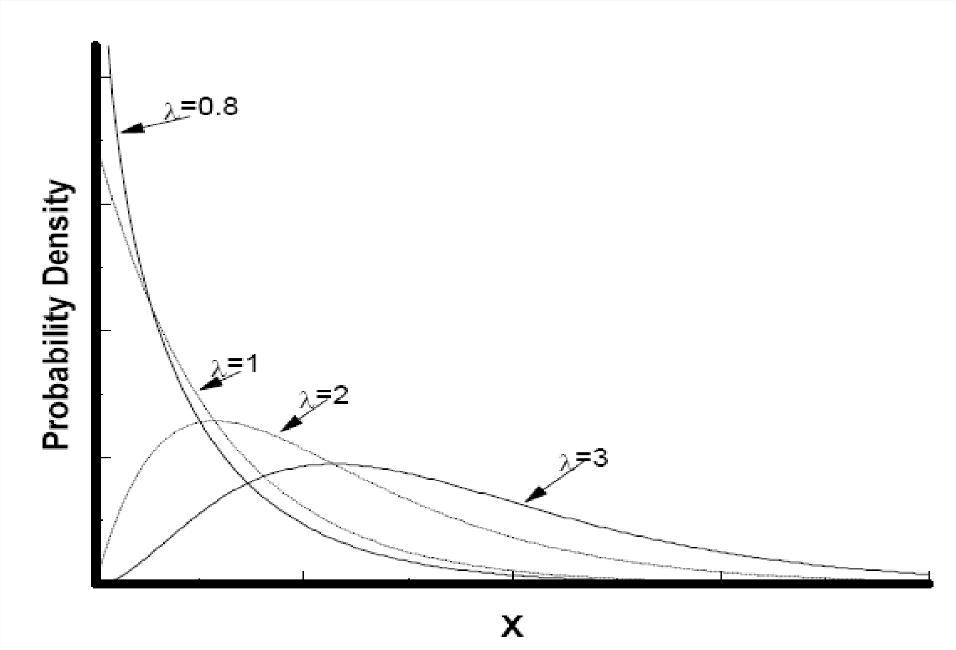
- Probability density function.

$$f(x) = \frac{\lambda^{n}}{\Gamma(n)} x^{n-1} e^{-\lambda x}$$

where $x \geq 0$, $n > 0$ and $\lambda > 0$

n is the shape parameter.

λ is the scale parameter.



Continuous Modeling Distribution.

[CRE Primer Ⅲ 30-59]

Gamma Distribution.

- Where the failure distribution is know to be exponential and the concern is to model the time to the nth failure rather than the first failure, the gamma distribution is the appropriate model.

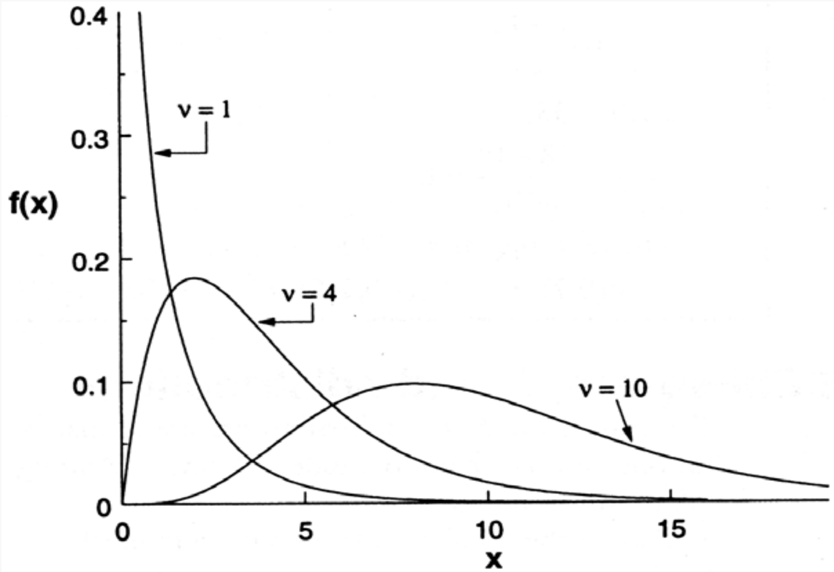
- Reliability function.

$$R(x) = \sum_{n-1}^{k=0} \frac{(\lambda x)^{k} e^{-\lambda x}}{k!}$$

Sampling Distribution.

[CRE Primer Ⅲ 72-77]

■ The Chi-Square Distribution $\left(\chi^2 = \frac{s^2(n-1)}{\sigma^2}\right)$.



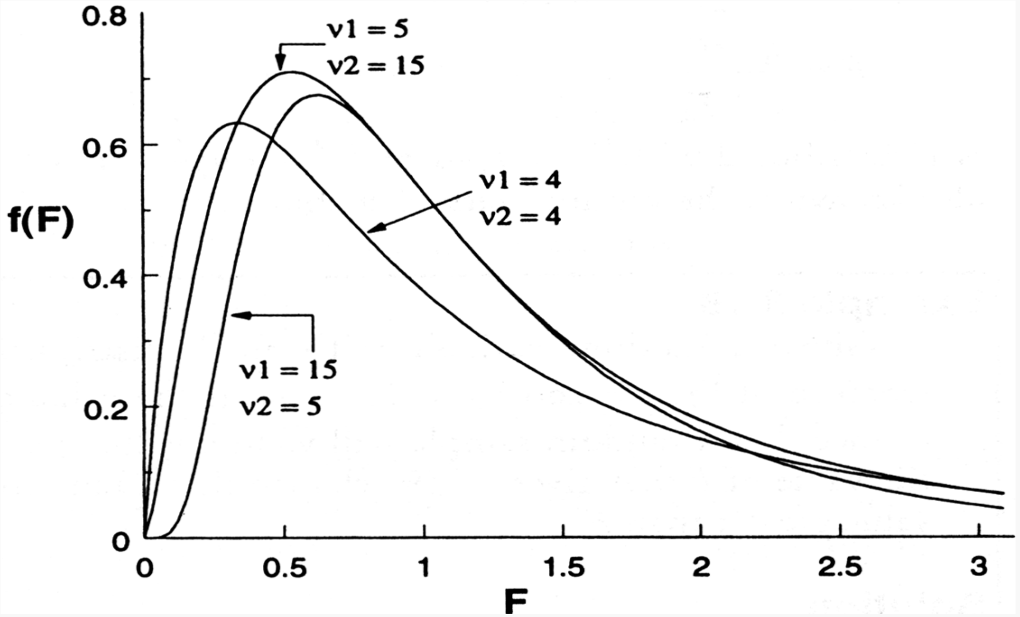
$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu/2-1)} e^{-x/2}, \quad x > 0$$

Sampling Distribution.

[CRE Primer III 72-77]

■ The F Distribution $\left(F = \frac{s_1^2}{s_2^2}\right)$.

- The *F* distribution is used to make inferences about variances and to construct confidence limits.

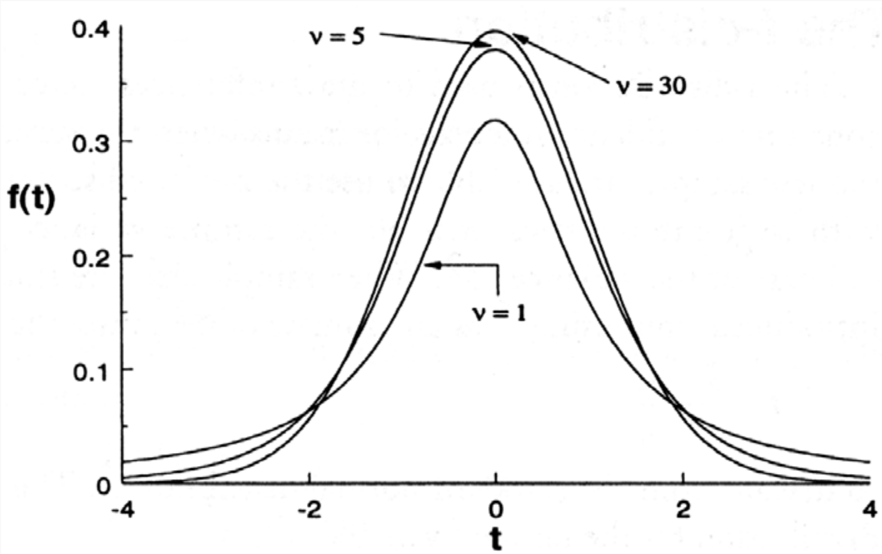


$$F_{1-\alpha}(\nu_1, \nu_2) = \frac{1}{F_\alpha(\nu_2, \nu_1)}$$

Sampling Distribution (cont.)

[CRE Primer Ⅲ 72-77]

■ The t Distribution.



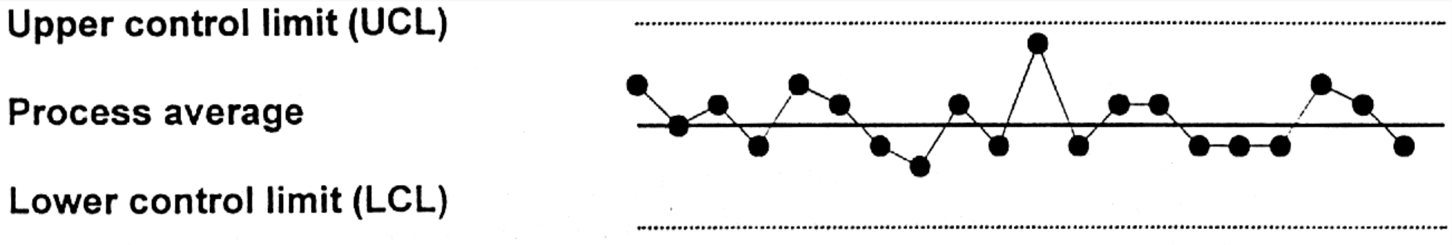
$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

- The t distribution is used to draw inferences concerning means and to construct confidence intervals for means when the variance is unknown, or too few samples are available to use the normal distribution.

Statistical Process Control

[CRE Primer Ⅲ 78-88]

■ Control Chart.



- Line graphs that display a dynamic picture of process behavior.
- Require approximately 100 data values to calculate upper and lower control limits, but
- Require only periodic small subgroups or x_s to continue to monitor the process.

Statistical Process Control.

[CRE Primer Ⅲ 78-88]

■ Control Chart.

- A process which is under statistical control : have plot points that do not exceed the upper or the lower control limits.

- Types of Charts.

  1. Control charts for Variables :

     $\bar{x}-R$ charts, Run charts,

     $M\overline{X}-MR$ charts, $\bar{x}-s$ charts, $Median$ charts

  2. Control charts for Attributes :

     p charts, np charts, c charts, u charts

Statistical Process Control.

[CRE Primer Ⅲ 78-88]

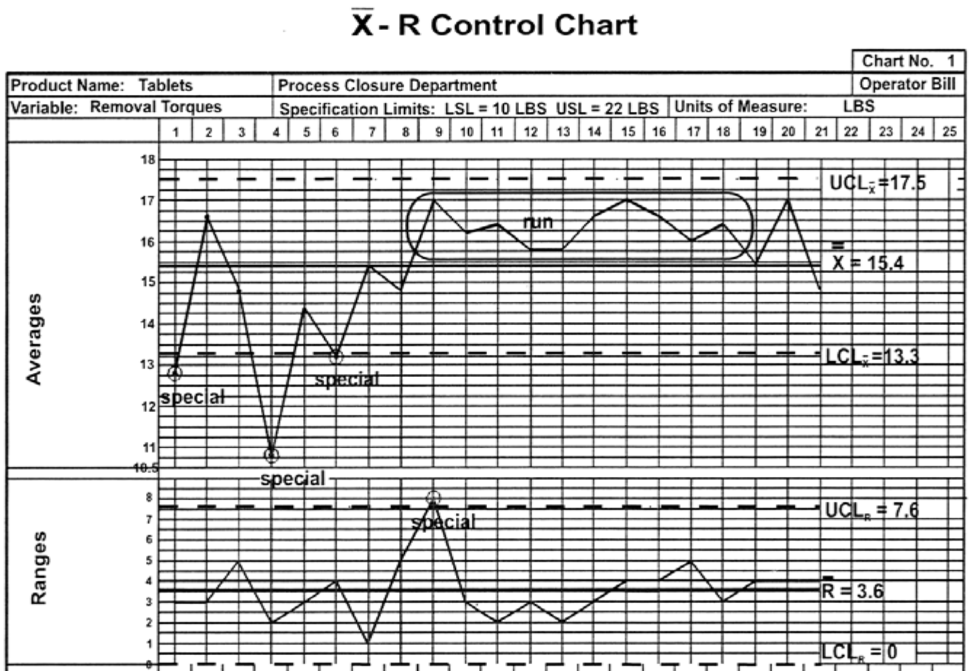
■ Steps for Constructing $\overline{x}-R$ Charts

1. Determine the sample size ($n=3,\ 4,\ 5$) and the frequency of sampling.
2. Collect 20 to 25 sets of time-sequenced samples.
3. Calculate the average for each set of samples, equals $\overline{X}$.
4. Calculate the range for each set of samples, equals R
5. Calculate $\overline{\overline{X}}$. This is the center line of the $\overline{X}$ chart.
6. Calculate $\overline{R}$. This is the center line of the R chart.
7. Calculate the control limits :
8. Plot the data and interpret the chart for special or assignable causes.

Statistical Process Control.

[CRE Primer Ⅲ 78-88]

■ Example of $\bar{x} - R$ Charts



		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Date Start 1/12/98		1/12																1/13								
Time		7:05	7:10	7:35	8:10	8:15	9:10	9:12	9:33	11:40	11:43	12:05	13:05	13:45	13:55	14:20	14:55	7:00	7:55	9:00	9:12	9:32				
Sample Measurement	1	12	15	13	10	13	15	15	15	22	16	16	15	17	16	17	19	16	16	17	19	14				
	2	12	17	18	12	16	12	16	17	17	15	18	16	15	15	19	17	19	15	13	18	17				
	3	13	16	14	11	14	13	15	16	15	17	16	17	16	18	17	15	16	17	17	17	16				
	4	15	17	14	10	15	15	16	14	17	15	16	17	15	18	15	15	15	16	15	15	14				
	5	12	18	15	11	14	11	15	12	14	18	16	14	16	16	17	17	14	18	14	16	13				
Total		64	83	74	54	72	66	77	74	85	81	82	79	79	83	85	83	80	82	76	85	74				
Average $\bar{x}$		12.8	16.6	14.8	10.8	14.4	13.2	15.4	14.8	17	16.2	16.4	15.8	15.8	16.6	17	16.6	16	16.4	15.2	17	14.8				
Range R		3	3	5	2	3	4	1	5	8	3	2	3	2	3	4	4	5	3	4	4	4				

Statistical Process Control.

[CRE Primer Ⅲ 78-88]

■ Basic Control Chart Interpretation Rules.

1. Special causes are any points above the UCL or below the LCL.

2. A **Run** is 7 or more consecutive points above or below the center lines.

3. A **1-in-20 violation** is more than one point in twenty consecutive points close to the control limits.

4. A **trend violation** is any upward or downward movement of 5 or more consecutive points or drifts of 7 or more points.

Statistical Process Control.

[CRE Primer Ⅲ 78-88]

Attribute Charts.

Chart	Records	Subgroup size
p	Fraction Defective.	Varies
np	Number of Defectives.	Constant
c	Number of Defects.	Constant
u	Number of Defects per Unit.	Varies
100 p	Percent Defectives.	Varies

Statistical Process Control.

[CRE Primer Ⅲ 78-88]

■ p Chart (% defectives, Binomial Distribution).

- UCL_p /LCL_p = $\bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{\bar{n}}}$

■ np Chart (Defectives, Binomial Distribution).

- UCL_{np} /LCL_{np} = $n\bar{p} \pm 3\sqrt{n\bar{p}(1-\bar{p})}$

Statistical Process Control.

[CRE Primer Ⅲ 78-88]

■ c Chart (Number of Defects, Poisson Distribution).

- $UCL_c / LCL_c = \bar{c} \pm 3\sqrt{\bar{c}}$

■ u Chart (Average Number of Defects, Poisson Distribution).

- $UCL_u / LCL_u = \bar{u} \pm 3\frac{\sqrt{\bar{u}}}{\sqrt{n}}$