

# What is Statistic

- ↳ ចំណាំថ្មី pol តារាងអនុសញ្ញា ឬដឹងទៅ ពាណិជ្ជកម្មនូវការបង្ហាញទិន្នន័យ និងរាយការណ៍
  - ↳ statistics is by and large the process of collecting information and presenting that in lots of different ways
  - ↳ ក្រឡាតេសample ឱ្យខ្សោយលើភេទជាន់ដែល រួម និងកំណត់ស្ថានឱ្យតែង និង ការបង្ហាញគ្រប់គ្រងទេសទៀត ឬវិវាទការបង្ហាញក្នុងសាលា → ពាណិជ្ជកម្មនូវការបង្ហាញទិន្នន័យបានបង្ហាញឡើង
  - ↳ ឲ្យវាតែងទៅការsample និងលើកទីនិងកំណត់ផ្សេងៗ និងកំណត់ទិន្នន័យនិងការបង្ហាញទិន្នន័យ នៅពេលវិនិច្ឆ័យ គ្រប់គ្រងទេសទៀត
  - ↳ កំណត់ទេសទៀតទៅការsample ជាមួយការបង្ហាញទិន្នន័យ ឬប៉ារ៉ា
  - ↳ so how can we draw conclusion about a bigger sample population
  - ↳ គឺនូវរាយការណ៍នៃក្រឡាតេសទៀត
  - ↳ មេដូចជាក្រឡាតេសទៀត 250 បុរាណតែម្ដោះ → ពាណិជ្ជកម្មនិងការបង្ហាញទិន្នន័យណាតែង ?
  - ↳ ក្នុងសាលាអាស៊ា 70% ជាលោកស្រីក្នុងបុរាណ 30% ត្រូវបានបង្ហាញទិន្នន័យ (សុបន្ទោះ)
  - ↳ នៅសាលាអាស៊ា ត្រូវបានបង្ហាញទិន្នន័យនៅពេលវិនិច្ឆ័យ ?
  - ↳ ក្នុងបុរាណ ជាបុរាណ
  - ↳ ① ដោយ 250 បុរាណ និងការបង្ហាញទិន្នន័យ 70% និង 30% និងទិន្នន័យ ?
  - ↳ ② ការបង្ហាញទិន្នន័យ → បានបានឯណាមីនីមូល មិនមែនបានបង្ហាញទិន្នន័យឡើង ការបង្ហាញទិន្នន័យបានបង្ហាញទិន្នន័យ
  - ↳ ឱ្យ ①, ② បានបង្ហាញទិន្នន័យ និងបង្ហាញទិន្នន័យ និងបង្ហាញទិន្នន័យ និងបង្ហាញទិន្នន័យ
  - ↳ “បានបង្ហាញទិន្នន័យ និងបង្ហាញទិន្នន័យ”
  - ↳ we are interesting in having information from a small sample of data but we're much interesting in using that information to make broad conclusions about a large set of data because it's impractical to ask everybody in the world a question

# What is Population in Statistic

↳ definition

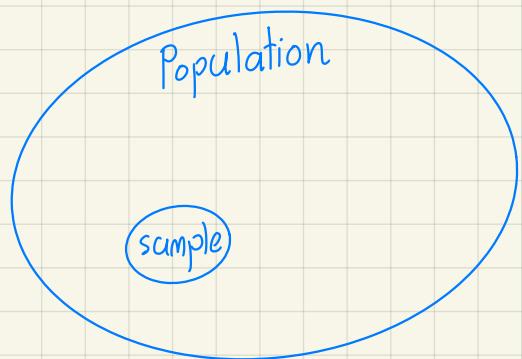
↳ statistic : gathering, describing and analyzing data

: actual numeric description of sample data → find max, min, ... this is statistics

↳ data : information gathered

- come together {
- ↳ population : a particular group of interesting
  - ↳ parameter : numerical description of a population characteristic.
    - ↳ might be mean height of all males in world, 75% of kid between 6-9 play vide game
  - come together {
    - ↳ sample : subset of population from which data is collect
    - ↳ sample statistics : numeric description of particular sample characteristic
      - ↳ of 100 female asked, 47% like chocolate

# Descriptive Verses Inferential Statistic



Branch of Statistics

↳ Descriptive Statistics

↳ gather, sort, summarize data from sample

↳ ឈរកសាងពេញ

↳ ងារចំណាំរបស់រាយការណ៍

↳ Inferential Statistics

↳ uses descriptive statistics (data) to estimate population parameters

↳ ស្ថិតិយកសាងពេញមួយទៅក្នុងក្រុងក្រប់ក្រង់ និង sample នៃកំសើសដែលមិនមែនជាប្រភេទនៃ population ទេ

↳ នូវរាយការណ៍ដូចជា ឱសាងនគរបាល និងការស្នើសុំរាយការណ៍

↳ ex បានកំសាន់ទីតាំងរាយការណ៍ 22% រាយការណ៍នេះមិនមែនជាប្រភេទនៃ inferential ឬនៃ descriptive

↳ ទេពេល ទៀតែជា descriptive statistics

↳ ជាមុននឹងពេល inferential មែនជា បានកំសាន់

ស្ថិតិយកសាងពេញ

estimate population parameter

↳ ex 65% of senior at a local high school applying to college plan to major in business.

↳ ទេជា descriptive statistics

↳ ជាមុន inference នឹងនឹងទាំង data នៅក្នុងរាយការណ៍ និងកំសាន់ទីតាំងរាយការណ៍ និងកំសាន់ទីតាំងរាយការណ៍

# Applying Definition in Statistics

- ↳ ចាប់ពីរបៀបដែលមិនមែនការស្នើសុំសាងសង់ទេ  
        → មិនមែនសាងសង់ទេ
  - ↳ ex. the height of every fourth bottle on an assembly line. Population or Sample
    - ↑
  - ↳ ex. ages of all USA presidents. P or S → សរុបអត្ថលេខាដែលជាប្រធានារដ្ឋាភិបាល
  - ↳ ពីរបៀបគឺជាព័ត៌មាននៃបណ្តុះបណ្តុះជាប្រជាធិបតេយ្យ population និងជាប្រជាធិបតេយ្យ sample
    - population
    - sample
  - ↳ ex. a researcher stop 100 people in a store to ask a survey of household income.
  - ↳ ex. the average number of hours per week a sample of 10 year olds spends watching TV is 20 hr
  - ↳ ex. 87% of all patients in a hospital report having an alcohol problem
- parameter  
statistics  
for sample

# The Frequency Distribution

↳ frequency distribution is table that divides data into groups (classes) and shows how many data value occur in each group.

classes	frequency	jumlah data dalam setiap kelas
a - b	n	
c - d	m	
e - f	o	

jumlah data dalam setiap kelas = class

$n + m + o = \text{sample size}$

↳ class width =  $b - a$

↳ class limit = a, c, e → จำกัดค่าสุดต่ำและสุดสูงของชั้น

↳ class boundaries =  $\frac{b+c}{2}$  → กำหนดจุดกลางระหว่าง 2 class ที่ → แบ่งช่วงของคลาสออกเป็นครึ่งๆ

# Relative and Cumulative Frequency Distribution

↳ relative frequency is the % of the data set that fall in a class (ຕາງກີ່ສິນທີ່ອະນຸຍາ)

$$\hookrightarrow \text{relative frequency} = \frac{\text{class frequency}}{\sum \text{frequency}}$$

↳ ex

classes	frequency	relative frequency	cumulative frequency
a - b	n	$\frac{n}{\sum f}$	n
c - d	m	$\frac{m}{\sum f}$	n + m
e - f	o	$\frac{o}{\sum f}$	n + m + o
$\sum f = n + m + o$			ດອກຢູ່ % ກໍ x 100 (ວຽກ)

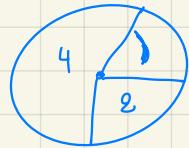
↳ cumulative frequency is the sum of a frequency for a class and all previous classes (ຕາງກີ່ສະສະ)

# Pie Charts, Bar graphs, and Pareto Charts

## ↳ pie charts

↳ សំណើលើក្នុង software មួយចន្ល័យ

↳ ការងារកាត់ទីតាមតម្លៃ 10%  $\Rightarrow 0.1 \times 360 = 36^\circ$



## ↳ Bar graph



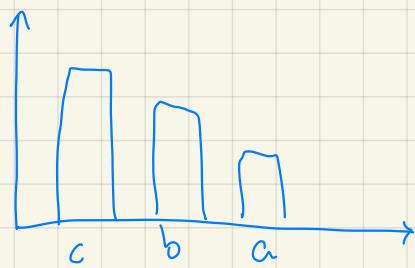
ជូនត្រួតពេលវេលា: bar chart

បានរាយការណ៍

ស្ថិតិមានដែលមានចំណោមខ្ពស់

## ↳ Pareto graph

↳ bar graph in descending order



ពេលវេលាភិនិត្យបាន

# Histogram

↳ basically is a bar charts that represent frequency distribution.

↳ ພາຍຕົ້ນ

frequency

10

8

6

4

2

boundaries \*

ບົດແຜງຂອງ ລົງ bar ຕົກນ → ໂພະວານໃຫ້ histogram

↳ ລົງ bar ທີ່ຕົກນ ດີວ່າເປັນ class ສິ້ນທີ່ຕົກນ

ກົດໝູນໃຫ້ class boundaries

60

65

70

75

80

85

↳ relative frequency histogram

↳ the height show relative frequency.

# Stem and Leaf Plots

↳ example is easier than explanation , it just shows ways to represent data

↳ `datas`: 18 23 24 31 19 27 26 99 32 18

stem | leaf

1	8 8 9
2	2 3 4 6 7
3	1 2

represent data above

၃၁။ ထိခိုက် histogram နှင့် များစွာ အသေခြင်း ပေါ်လိမ့်မှုပါ။

ၷ မြန်မာစီရင် အောက် မြန်မာစုပ္ပန္လ

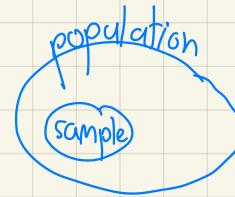
# Population Mean and Sample Mean

↳ center of that data set

↳ means average value of the data

sample mean: mean of sample values collected  $(\bar{x} = \frac{\sum x}{n})$

population means: concept same as sample mean  $(\mu = \frac{\sum x}{N})$



↳  $\bar{x}$  and  $\mu$  are measures of location

# Calculating Mean

↳ weighted mean : When each value is not equally important

$$\text{weighted mean} = \frac{\sum (w \cdot x)}{\sum w}$$

↳ សម្រាប់ការគិតផលនៃទូទាត់

ទូទាត់	រូបតាមរយៈរឹង	ស្នើសុំ
1	4	2
2	3	4

$$\begin{aligned}\text{ស្នើសុំ} &= \frac{4(2) + 3(4)}{7} \\ &= \frac{20}{7} \\ &\approx 2.8\end{aligned}$$

↳ median and mode

↳ កំណត់ទីទូទាត់នៃពេលវេលាដែលមិនមែនចំណាំរាយទៅជាការគិតផលនៃទូទាត់ និងជាការគិតផលនៃទូទាត់ដែលមិនមែនចំណាំរាយទៅជាការគិតផលនៃទូទាត់

↳ ជាផ្លូវការងារប្រចាំថ្ងៃនៃទូទាត់

↳ មួយចំណាំនៃទូទាត់

↳ median : the middle value in an ordered list of data.

↳ ការគិតផលនៃទូទាត់ (ទូទាត់ទាំងអស់) ដែលមិនមែនចំណាំរាយទៅជាការគិតផលនៃទូទាត់

↳ mode : value in data set that occurs most frequently

↳ ទូទាត់ដែលមិនមែនចំណាំរាយទៅជាការគិតផលនៃទូទាត់

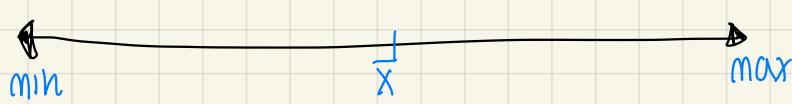
↳ Range of Data

↳ how the data is spread out

↳ range of data = max - min

# Population and Sample Variance

↳ Variance: how spread apart our data is



; brand of dataset

↳ for each data point

b deviation from mean =  $x_i - \bar{x}$  ; និងបោរកការកំណត់សម្រាប់ទីផ្សារ គឺជាដែនលក់ដែលមិនស្ម័គ្រប់ពីលក្ខណៈ

↳ add up all deviation  $\Sigma(x - \bar{x})$

Deviation =  $\frac{\sum(X - \bar{X})}{N}$  ; averaging deviation

↳ զարգացնելու համար առաջարկություններ - , + Առաջարկավորչական Mean  
գործադրություն

↳  $\omega_0 + \text{ເອົາກອນພວກົບ} - (\text{ເກີດອະນຸຍັງຫຼືຈິງ}) \rightarrow \text{ກອບມືກລົງ} = 0$

$$S^2 = \frac{\sum (x-u)^2}{N}; \text{ always positive deviation}$$

↳ eisann data ຖະແຫຼງນະບາງ → population variance

## b sample variance

$$S^2 = \frac{\sum (X - \bar{X})^2}{n-1}$$

“ S ” sample variance

# Population and Sample Standard Deviation

↳ sample variance :  $s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$

population variance :  $\sigma^2 = \frac{\sum(x-\mu)^2}{N}$

} ចិត្តកំណត់ទៅកាន់ square នាយកស្រី មួយតុលាបែងការ (បញ្ចប់ទិន្នន័យ)  
នឹង data ម៉ោងនៅពេល

↳ មិនដឹងមុនថា ធមធានភាពណានឹង variance = 15 មុនតែ 15 គឺរបៀបណា?

↳ ឱ្យដឹងមុនថា គឺជាអ្នកបានដោះស្រាយ មួយតុលាបែងការ

↳ សេវាដំឡើងបែងប្រឈមបែងប្រឈម

↳ sample standard deviation

$$s = \sqrt{s^2} = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

population standard deviation

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(x-\mu)^2}{N}}$$

↳ ចាប់ពី μ ដល់ s ទៅជាបន្ទាន់ បានការពិនិត្យថាបានចូលរួមនៅក្នុងបែងប្រឈមការបែងប្រឈម ± s

↳  $\mu=90$ ,  $s=5 \rightarrow$  ពីចូលរួមនៅក្នុង  $(85, 95)$  បានចូលរួមនៅក្នុង

# Coefficient of Variation

- ↳ standard deviation and variance → កំណត់ភារិសាងការងាររបស់ខ្លួនដោយលក្ខណៈការកំណត់
- ↳ តាម (S O V C) និងកំណត់នៅលាយមិនអាចគាំទ្រ
- ↳ សម្រាប់ដឹងដឹងអ្នកនីមួយៗ

$$\begin{array}{l} \text{ចំណាំ 1} \quad \bar{x} = 120,000 \\ \quad s = 2,000 \end{array}$$

$$\begin{array}{l} \text{ចំណាំ 2} \quad \bar{x} = 900,000 \\ \quad s = 10,000 \checkmark \end{array}$$

មានភាពការកំណត់នៅលាយមិនអាចគាំទ្រ:  $s_1 > s_2$

↳ ព័ត៌មាននៃលាយមិនអាចគាំទ្រនៅ mean នៅ

↳ វិនិច្ឆ័យនៃសំណើរបស់វា ការកំណត់នៅលាយមិនអាចគាំទ្រនៅ

↳ "co-efficient of variation"

- ↳ basically, it's a way to let you compare 2 datasets to tell you which one has more spread.

- ↳ in relation to compare to its mean

$$C o V_o = \frac{s}{\bar{x}} \times 100$$

↳ សម្រាប់ការរាយលទ្ធផលបានប្រើប្រាស់

$$C o V_{o1} = \frac{2,000}{120,000} \times 100 = 1.66\% \checkmark$$

→ % នៃលាយមិនអាចគាំទ្រនៅ mean

$$C o V_{o2} = \frac{10,000}{900,000} \times 100 = 1.11\%$$

→ % នៃលាយមិនអាចគាំទ្រនៅ mean

↳ ធ្វើតាមរាយការ 2 កំណត់នៅលាយមិនអាចគាំទ្រនៅ mean នៃសំណើរបស់វា

↳ ស្ថិតិសាស្ត្រការណ៍នៃលទ្ធផលបានប្រើប្រាស់

# Standard Deviation of Frequency Data

↳ recall  $s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$

↳ recall frequency distribution

classes	frequency
A - B	n

↳ ප්‍රතිච්චීයා මෙහේ සඳහා නොමැත්තු නොමැත්තු නොමැත්තු නොමැත්තු නොමැත්තු

↳ it can be proven that

$$s = \sqrt{\frac{n(\sum(F \cdot x^2)) - (\sum(F \cdot x))^2}{n-1}}$$

$n$  = sample size

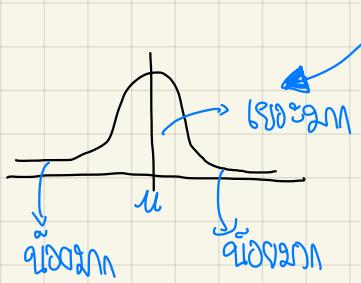
$x$  = mid point of class  $\frac{A+B}{2}$

$F$  = frequency

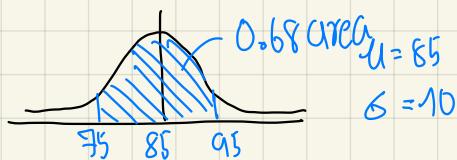
↳ Variance = standard deviation<sup>2</sup>

# Empirical Rule of Standard Deviation

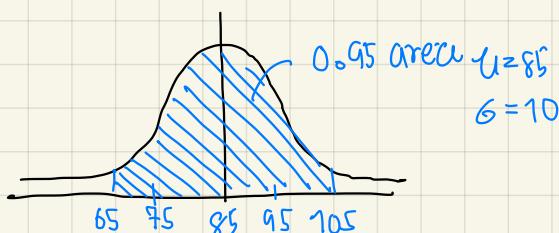
- ↳ how we can start to draw some basic conclusion from data using standard deviation.
- ↳ සිංහලයෙන් standard deviation නෑතුවන්දීමූදාන්, මැන්සා → ගණකාධිකීම්?
- ↳ සේවන් නෑතුවන්දීමූදාන් පරිශ්‍යාපිත ප්‍රමාණය අනුව ප්‍රමාණය පිළිබඳ ප්‍රතිශතය පිළිබඳ යුතු ඇති අතර මෙයින් ප්‍රතිශතය පිළිබඳ යුතු නොවායි.
- ↳ only applies to Bell Shaped data



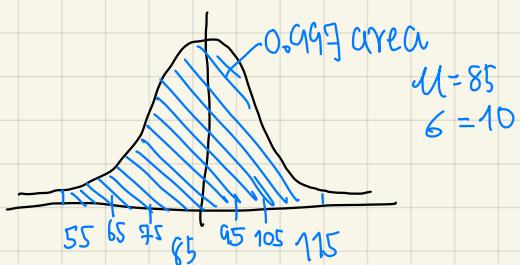
- ↳ the empirical rule for standard deviation
- ↳ about 68% of data lie within 1 standard deviation of the mean



- ↳ about 95% of the data lie within 2 standard deviation of the mean



- ↳ about 99.7% of the data lie within 3 standard deviation of the mean



# Chebychev's Theorem

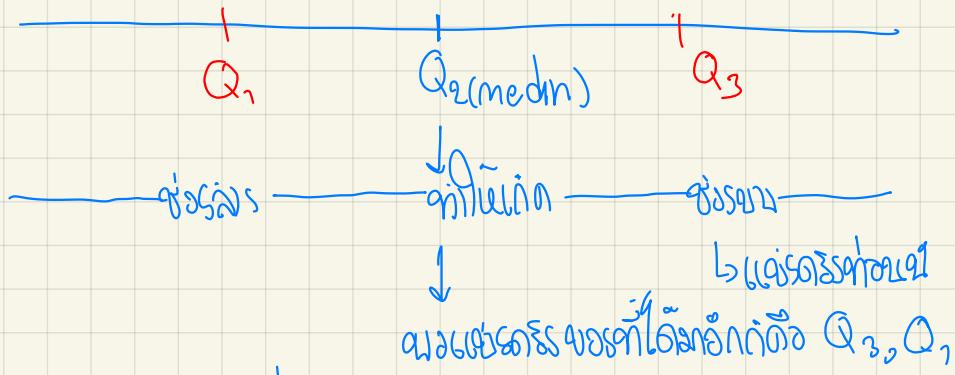
- ↳ cheb - by - shev theorem
  - ↳ សម្រាប់ការបង្កើតលទ្ធផលរបស់ពួកគេ តើតុលាតិត្យនៃ data នឹងជារូប bell shape
  - ↳ និងការបង្កើតលទ្ធផលរបស់ពួកគេ តើតុលាតិត្យនៃ data នឹងជារូប bell shape
- ↳ It doesn't tell us that 95% is gonna fall between this, it give us a minimum that applies.
- ↳ at least x percent is gonna fall between one standard deviation
- ↳ Chebbychev's Theorem
  - ↳ the % of the data that lies within "k" standard deviation is at least.
$$= 1 - \frac{1}{k^2} ; (k > 1)$$
- ↳ ex for ( $k=2$  at least)  $1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4} = 75\%$
- ↳ at least 75% of the data lie within 2G mean

# Quartiles of Data

↳ កិច្ចការបង្កើត box, whisker plot

↳ quartiles divide a data set up as follow

↳ this is your data -



↳  $Q_1$  : the number  $\geq 25\%$  of the data

$Q_2$  : the number  $\geq 50\%$  of the data

$Q_3$  : the number  $\geq 75\%$  of the data

$$Q_r =$$

# Box and Whisker Plots

សម្រាប់ five-number summary of the data set → ការរួមចាប់ពីលើខ្លួន ដល់ 5 ចំណាំ

សងគ្គតាមទីរយៈប្រចាំថ្ងៃនៃសម្រាប់ប្រចាំសប្តាហិរញ្ញវត្ថុ

អាណាព្យាប័ណ្ឌ នូវការ

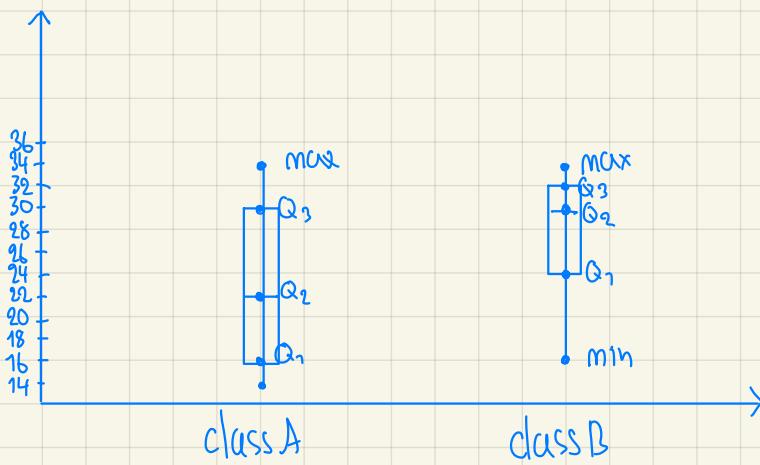
	ចំណាំ 1	ចំណាំ 2
min	14	16
$Q_1$	16	24
$Q_2$	22	30
$Q_3$	30	32
max	34	34

min value,  $Q_1$ ,  $Q_2$ ,  $Q_3$ , max value

ប្រចាំថ្ងៃ 5 ចំណាំ

ប្រចាំសប្តាហិរញ្ញវត្ថុ

ប្រចាំសប្តាហិរញ្ញវត្ថុ



# Standard Score (z-Score)

↳ everything depend on average and standard deviation

↳ សេចក្តីពណ៌នាបានប្រើប្រាស់ស្ថាបន្ទាក់រាយការណ៍ដែលបានតាំងក្នុងស្ថាបន្ទាក់

↳ z-score ជាមួយនាំនូវម៉ោងនៃ mean និងស្ថាបន្ទាក់

↳ easiest way to calculate that is how many standard deviation am I in front of the mean

↳ ហើយទីនាំ (empirical) rule នឹងមាន (9 ពាន់នាទី)

∴ standard score is tell us how many std dev a number is above or below the mean.

$$Z = \frac{X - \mu}{\sigma}$$

↓ population → sample និងសៅរីសៅរី

↳ ឯក z-score នឹងត្រូវបានការពារឡើង

↳ ការគិតតម្លៃនឹងកុណិតុយ និងចំណុចនៃការការពារនឹងបានការពារឡើង