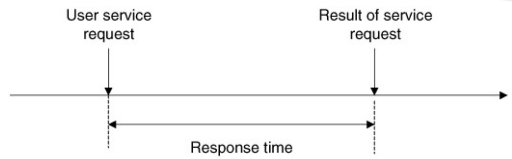
Bab 4 Performance Metrix, dan Teori Probabilitas

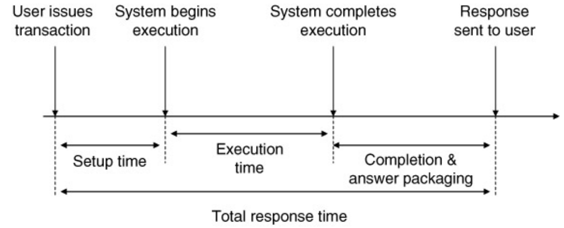
Dr. Ir. Yeffry Handoko Putra, M.T

Response time

- Typical response time measurement

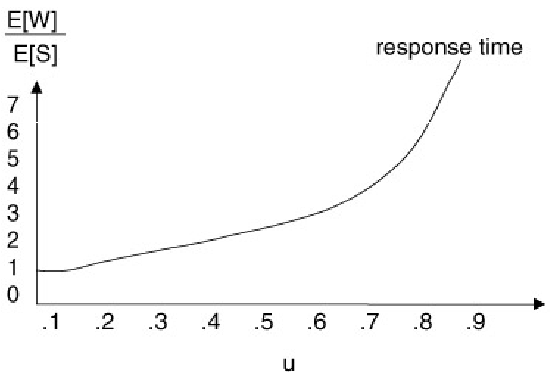


- Transaction processing response



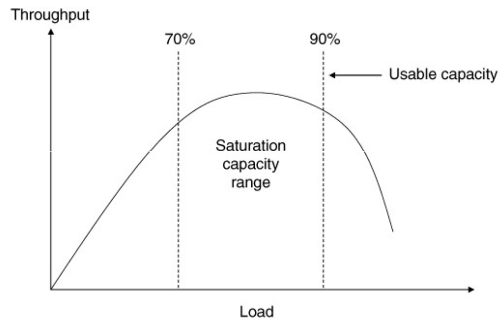
Response Time (cont)

$$\text{Stretch factor } = E[W]/E[S]$$



Throughput

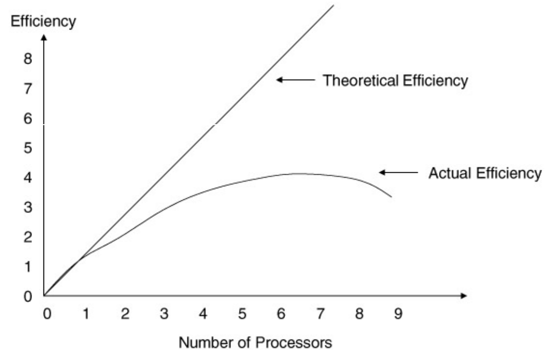
- Throughput curves versus response curves



Efficiency

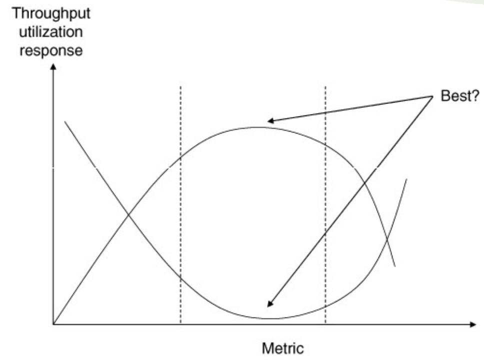
Efficiency = real throughput/theoretical throughput

- Multiprocessor efficiency curve.



Evaluation

- Metrics versus usefulness.



Utilization, reliability, and availability

Utilization = time busy/(time busy + time idle)

- Kegagalan :
  - Failure rate : kegagalan sistem yang tidak berhubungan dengan sumber daya kritis
  - Hazardous rate : kegagalan sistem yang berhubungan dengan sumber daya kritis (nyawa, alat yang penting)

- Perbaikan
- Pergantian
- Pemeliharaan

Availabilitas dan servisability

Parameter Reliabilitas

- MTBF : Mean Time Between Failure
  Nilai waktu rata-rata terjadinya kegagalan pertama yang bisa diperbaiki
  Dalam sistem komputer dikenal MTBI (Mean Time Between Interrupt)
- MTTF: Mean Time True Failure
  Nilai waktu rata-rata terjadinya kegagalan yang harus diganti
- MTTR : Mean Time to Repair
  Nilai waktu rata-rata untuk memperbaiki

Tool menyelesaikan reliabilitas

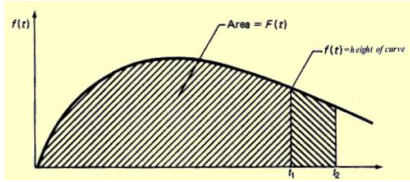
- Probabilitas
  - Probability density function :

$$f(t)=\frac{dF(t)}{dt}$$

  - Cumulative density function :

$$F(t)=\int_{-\infty}^{t} f(x)\,dx$$

Model dasar reliabilitas



$$F(t)=1-e^{-\left(\frac{t}{\alpha}\right)^{\gamma}}\quad,\quad f(t)=\frac{\gamma}{t}\left(\frac{t}{\alpha}\right)^{\gamma}e^{-\left(\frac{t}{\alpha}\right)^{\gamma}}$$

CDF: $F(t)=1-e^{-\left(\frac{t}{\alpha}\right)^{\gamma}}$

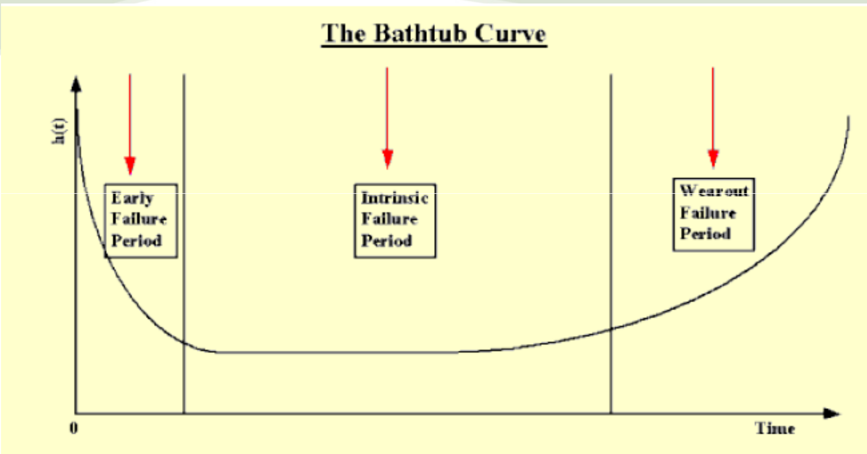
RELIABILITY: $e^{-\left(\frac{t}{\alpha}\right)^{\gamma}}$

PDF: $f(t)=\frac{\gamma}{t}\left(\frac{t}{\alpha}\right)^{\gamma}e^{-\left(\frac{t}{\alpha}\right)^{\gamma}}$

FAILURE RATE: $\frac{\gamma}{\alpha}\left(\frac{t}{\alpha}\right)^{\gamma-1}$

MEAN: $\alpha\Gamma\left(1+\frac{1}{\gamma}\right)$

Model Kurva Bathup untuk proses wearing



Probability Theory and Statistic

Dr. Ir. Yeffry Handoko Putra, M.T

After Follow this lecture

you will be able to answer questions such as the ones that follow.

1. How should you report the performance as a single number? Is specifying the mean the correct way to summarize a sequence of measurements?
2. How should you report the variability of measured quantities? What are the alternatives to variance and when are they appropriate?
3. How should you interpret the variability? How much confidence can you put on data with a large variability?
4. How many measurements are required to get a desired level of statistical confidence?
5. How should you summarize the results of several different workloads on a single computer system?
6. How should you compare two or more computer systems using several different workloads? Is comparing the mean performance sufficient?
7. What model best describes the relationship between two variables? Also, how good is the model?

BASIC PROBABILITY AND STATISTICS CONCEPTS

1. **Independent Events:** Two events are called independent if the occurrence of one event does not in any way affect the probability of the other event. Thus, knowing that one event has occurred does not in any way change our estimate of the probability of the other event.
2. **Random Variable:** A variable is called a random variable if it takes one of a specified set of values with a specified probability.
3. **Cumulative Distribution Function:** The Cumulative Distribution Function (**CDF**) of a random variable maps a given value *a* to the probability of the variable taking a value less than or equal to *a*:
   $F_x(a) = P(x \le a)$
4. **Probability Density Function:** The derivative

$$f(x) = \frac{dF(x)}{dx}$$

of the CDF $F(x)$ is called the probability density function (**pdf**) of x. Given a pdf $f(x)$, the probability of x being in the interval (x_1, x_2) can also be computed by integration:

$$P(x_1 < x \le x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)\,dx$$

5. **Probability Mass Function:** For discrete random variable, the CDF is not continuous and, therefore, not differentiable. In such cases, the probability mass function (**pmf**) is used in place of pdf. Consider a discrete random variable x that can take n distinct values $\{x_1, x_2, \ldots, x_n\}$ with probabilities $\{p_1, p_2, \ldots, p_n\}$ such that the probability of the ith value x_i is p_i. The pmf maps x_i to p_i:

$$f(x_i)=p_i$$

The probability of x being in the interval (x_1, x_2) can also be computed by summation:

$$P(x_1 < x \le x_2) = F(x_2) - F(x_1) = \sum_{x_1 < x_i \le x_2} p_i$$

6. **Mean or Expected Value**

$$\text{Mean } \mu = E(x) = \sum_{i=1}^{n} p_i x_i = \int_{-\infty}^{+\infty} x f(x)\,dx$$

Summation is used for discrete and integration for continuous variables, respectively.

7. **Variance:** The quantity $(x - \mu)^2$ represents the square of distance between x and its mean. The expected value of this quantity is called the variance x:

The variance is traditionally denoted by σ^2. The square root of the variance is called the **standard deviation** and is denoted by σ.

$$\text{Var}(x) = E[(x-\mu)^2] = \sum_{i=1}^{n} p_i (x_i - \mu)^2 = \int_{-\infty}^{+\infty} (x_i - \mu)^2 f(x)\,dx$$

8. **Coefficient of Variation:** The ratio of the standard deviation to the mean is called the Coefficient of Variation (**C.O.V.**):

$$\text{C.O.V.} = \frac{\text{standard deviation}}{\text{mean}} = \frac{\sigma}{\mu}$$

9. **Covariance:** Given two random variables x and y with means μ_x and μ_y, their covariance is

$$\text{Cov}(x,y) = \sigma_{xy}^2 = E[(x-\mu_x)(y-\mu_y)] = E(xy) - E(x)E(y)$$

For independent variables, the covariance is zero since $E(xy) = E(x)E(y)$
Although independence always implies zero covariance, the reverse is not true. It is possible for two variables to be dependent and still have zero covariance.

10. **Correlation Coefficient:** The normalized value of covariance is called the correlation coefficient or simply the **correlation**

$$\text{Correlation}(x,y) = \rho_{xy} = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y}$$

The correlation always lies between -1 and +1.

11. **Mean and Variance of Sums:** If $x_1, x_2, \ldots, x_k$ are k random variables and if $a_1, a_2, \ldots, a_k$ are k arbitrary constants (called weights), then $E(a_1x_1 + a_2x_2 + \ldots + a_kx_k) = a_1E(x_1) + a_2E(x_2) + \ldots + a_kE(x_k)$
For independent variables, $\text{Var}(a_1x_1 + a_2x_2 + \ldots + a_kx_k) = a^2_1\text{Var}(x_1) + a^2_2\text{Var}(x_2) + \ldots + a^2_k\text{Var}(x_k)$

12. **Quantile:** The x value at which the CDF takes a value α is called the α-quantile or 100α-percentile. It is denoted by x_α and is such that the probability of x being less than or equal to x_α is α:

$$P(x \le x_\alpha) = F(x_\alpha) = \alpha$$

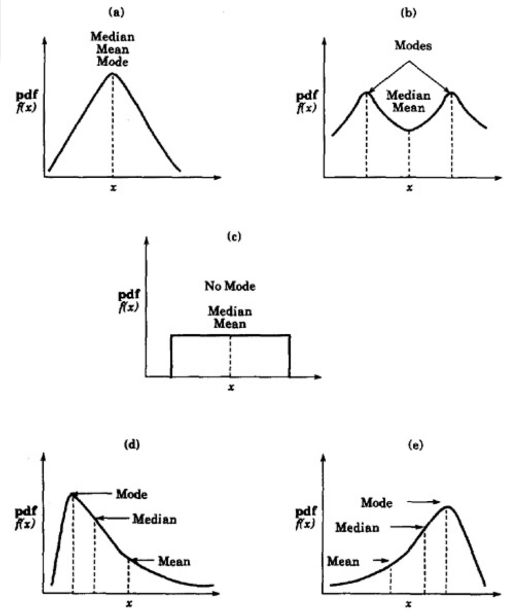
13. **Median:** The 50-percentile (or 0.5-quantile) of a random variable is called its median.
14. **Mode:** The most likely value, that is, x_i, that has the highest probability p_i, or the x at which pdf is maximum, is called the mode of x.
15. **Normal Distribution:** This is the most commonly used distribution in data analysis. The sum of a large number of independent observations from any distribution has a normal distribution. Also known as Gaussian distribution, its pdf is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty \le x \le +\infty$$

There are two parameters μ and σ, which are also the mean and standard deviations of x. A normal variate is denoted by $N(\mu, \sigma)$. A normal distribution with zero mean and unit variance is called a **unit normal** or **standard normal distribution** and is denoted as $N(O, 1)$.

SUMMARIZING DATA BY A SINGLE NUMBER

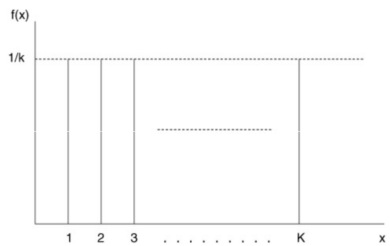
- Three popular alternatives to summarize a sample are to specify its **mean, median, or mode**
- These measures are what statisticians call **indices of central tendencies**. The name is based on the fact that these measures specify the center of location of the distribution of the observations in the sample.



Some Continue Probability Distribution
Ref. [Fortier chap 5.6)

- **Uniform distribution**

$f(x) = 1/k \qquad x = x_1, x_2, \ldots, x_k$



Mean: $E[X] = \sum_{i=1}^{k} x_i (1/k)$

$$\sigma^2 = E\left[(X-\mu)^2\right] = \sum_{i=1}^{k} (x_i - \mu)^2 f(x_i)$$

$$= \sum_{i=1}^{k} \frac{(x_i-\mu)^2}{k}$$

- **Binominal distribution**

$P(x \text{ successes in } n \text{ trials}) = p^x q^{(n-x)}$

$f(x) = C(n,x)\, p^x q^{n-x} \qquad x = 1, 2, 3, \ldots, n$

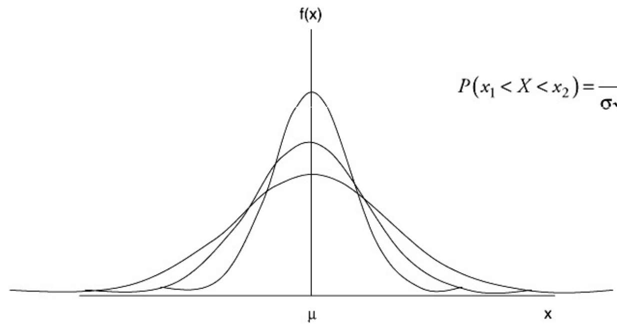
Mean: E[x]=np
Var[x]=npq

- **Poisson distribution**

$f(x) = (e^{-\mu}\mu^x)/x! \qquad x = 0, 1, 2, \ldots$

- **Gaussian distribution**

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



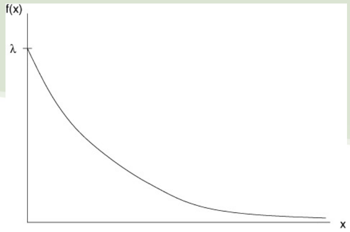
$$P(x_1 < X < x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)} dx$$

- **Exponential distribution**

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & x>0\\ 0 & \text{otherwise}\end{cases}$$

$$F(x)=1-e^{-\lambda kx}\left[1+\sum_{i=0}^{k-1}\frac{(k\lambda x)^i}{i!}\right]$$

Mean = $1/\lambda$
Var[x] = $1/\lambda^2$

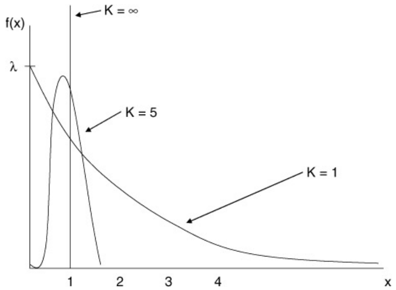


- **Erlang distribution**

$$f(x)=\begin{cases}\dfrac{\lambda k(\lambda kx)^{k-1}e^{-\lambda kx}}{(k-1)!} & x>0\end{cases}$$

$$F(x)=1-e^{-\lambda kx}\left[1+\sum_{i=0}^{k-1}\frac{(k\lambda x)^i}{i!}\right]$$

Mean = $1/\lambda$
Var[x] = $1/k\lambda^2$



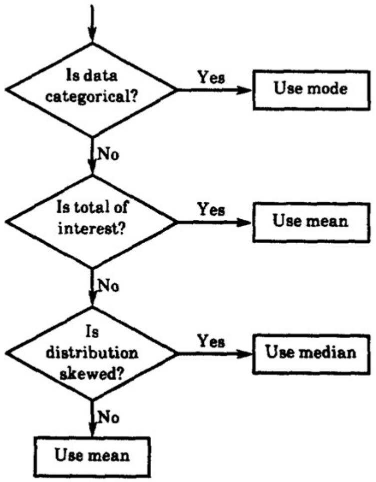
There are two main reasons for the popularity of the normal distribution

- *The sum of n independent normal variates is a normal variate.* If $x_i \sim N(\mu_i, \sigma_i)$, then has a normal distribution with mean $\mu = \sum_{i=1}^{n} a_i\mu_i$ and variance $x = \sum_{i=1}^{n} a_i x_i$ $\sigma^2 = \sum_{i=1}^{n} a_i^2\sigma_i^2$.

  As a result of this linearity property, normal processes remain normal after passing through linear systems, which are popular in electrical engineering.

- *The sum of a large number of independent observations from any distribution tends to have a normal distribution.* This result, which is called the **central limit theorem**, is true for observations from all distributions. As a result of this property, experimental errors, which are contributed by many factors, are modeled with a normal distribution

Guidelines to select a proper index of central tendency (Mean, Mode, Median)



The following are examples of selections of indices of central tendencies:

- *Most Used Resource in a System*: Resources are categorical and hence the mode must be used.
- *Interarrival Time*: Total time is of interest and so the mean is the proper choice.
- *Load on a Computer*: The median is preferable due to a highly skewed distribution.
- *Average Configuration*: Medians of number devices, memory sizes, and number of processors are generally used to specify the configuration due to the skewness of the distribution.

COMMON MISUSES OF MEANS

- *Using Mean of Significantly Different Values*:
  For example, it is not very useful to say that the mean CPU time per query is 505 milliseconds when the two measurements come out to be 10 and 1000 milliseconds. An analysis based on 505 milliseconds would lead nowhere close to the two possibilities. In this particular example, the mean is the correct index but is useless.
- *Using Mean without Regard to the Skewness of Distribution*:
- *Multiplying Means To Get the Mean of a Product*:

$$E(xy) \neq E(x)E(y)$$

- *Taking a Mean of a Ratio with Different Bases*:

Example 12.1 On a timesharing system, the total number of users and the number of subprocesses for each user are monitored. The average number of users is 23. The average number of subprocesses per user is 2. What is the average number of subprocesses?

TABLE 12.1 System Response Times for 5 Days

	System A	System B
	10	5
	9	5
	11	5
	10	4
	10	31
Sum	50	50
Mean	10	10
Typical	10	5

Is it 46? No! The number of subprocesses a user spawns depends upon how much load there is on the system. On an overloaded system (large number of users), users try to keep the number of subprocesses low, and on an underloaded system, users try to keep the number of subprocesses high. The two variables are correlated, and therefore, the mean of the product cannot be obtained by multiplying the means. The total number of subprocesses on the system should be continuously monitored and then averaged.

GEOMETRIC MEAN

- The geometric mean of n values $x_1, x_2, \ldots, x_n$ is obtained by multiplying the values together and taking the nth root of the product:

$$\dot{x} = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$$

The mean discussed in other sections is what should be termed the arithmetic mean. The arithmetic mean is used if the sum of the observations is a quantity that is of interest. Similarly, the geometric mean is used if the product of the observations is a quantity of interest.

- **Example 12.2** The performance improvements in the latest version of seven layers of a new networking protocol was measured separately for each layer. The observations are listed in Table 12.2. What is the average improvement per layer? The improvements in the seven layers work in a "multiplicative" manner, that is, doubling the performance of layer 1 and layer 2 shows up as four times the improvement in performance.

TABLE 12.2 Improvement in Each Layer of Network Protocol

Protocol Layer	Performance Improvement (%)
7	18
6	13
5	11
4	8
3	10
2	28
1	5

Average improvement per layer

$= \{(1.18)(1.13)(1.11)(1.08)(1.10)(1.28)(1.05)\}^{1/7} - 1$

$= 0.13$

Thus, the average improvement per layer is 13%.

- Other examples of metrics that work in a multiplicative manner are as follows:
  - Cache hit ratios over several levels of caches
  - `Cache miss ratios
  - Percentage performance improvement between successive versions
  - Average error rate per hop on a multihop path in a network
- The geometric mean can be considered as a function gm( ), which maps a set of responses $\{x_1, x_2, \ldots, x_n\}$ to a single number . It has the following multiplicativity property:

$$\mathrm{gm}\left(\frac{x_1}{y_1}, \frac{x_2}{y_2}, \ldots, \frac{x_n}{y_n}\right) = \frac{\mathrm{gm}(x_1, x_2, \ldots, x_n)}{\mathrm{gm}(y_1, y_2, \ldots, y_n)} = \frac{1}{\mathrm{gm}(y_1/x_1, y_2/x_2, \ldots, y_n/x_n)}$$

- That is, the geometric mean of a ratio is the ratio of the geometric means of the numerator and denominator. Thus, the choice of the base does not change the conclusion. It is because of this property that sometimes the geometric mean is recommended for ratios. However, if the geometric mean of the numerator or denominator does not have any physical meaning, the geometric mean of their ratio is meaningless as well. Means of ratios are discussed further in Section 12.7.
-

HARMONIC MEAN

- The harmonic mean $\dot{x}$ of a sample $\{x_1, x_2, \ldots, x_n\}$ is defined as follows:

$$\dot{x} = \frac{n}{1/x_1 + 1/x_2 + \cdots + 1/x_n}$$

- A harmonic mean should be used whenever an arithmetic mean can be justified for $1/x_i$. For example, suppose repeated measurements are made for the elasped time of a benchmark on a processor. In the *i*th repetition, the benchmark takes t_i seconds. Now suppose the benchmark has m million instructions, the MIPS x_i computed from the *i*th repetition is
- $$x_i = \frac{m}{t_i}$$
- For the same reasons, x_i's should be summarized using the harmonic mean since the sum of $1/x_i$'s has a physical meaning. The average MIPS rate for the processor is

$$\dot{x} = \frac{n}{\frac{1}{m/t_1} + \frac{1}{m/t_2} + \cdots + \frac{1}{m/t_n}} = \frac{m}{(1/n)(t_1 + t_2 + \cdots + t_n)}$$

MEAN OF A RATIO

- Given a set of n ratios, a common problem is to summarize them in a single number

$$\text{Average}\left(\frac{a_1}{b_1}, \frac{a_2}{b_2}, \ldots, \frac{a_n}{b_n}\right) = \frac{a_1 + a_2 + \cdots + a_n}{b_1 + b_2 + \cdots + b_n} = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} = \frac{(1/n)\sum_{i=1}^{n} a_i}{(1/n)\sum_{i=1}^{n} b_i} = \frac{\bar{a}}{\bar{b}}$$

TABLE 12.3 CPU Utilization Measured over Five Intervals

Measurement Duration	CPU Duration Busy (%)
1	45
1	45
1	45
1	45
100	20
sum	200%
Mean	≠ 200/5 or 40%

$$\text{Mean CPU utilization} = \frac{\text{sum of CPU busy times}}{\text{sum of measurement durations}} = \frac{0.45 + 0.45 + 0.45 + 0.45 + 20}{1 + 1 + 1 + 1 + 100} = 21\%$$