



INSTITUTE FOR
HEALTHCARE
IMPROVEMENT

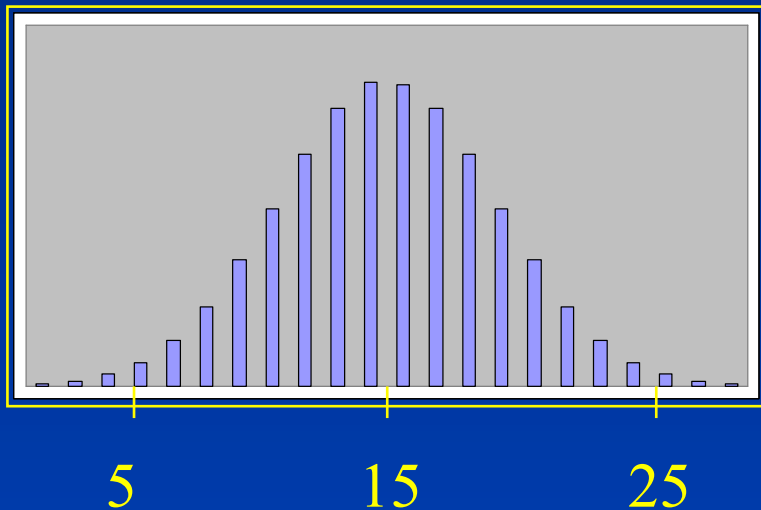
Flow, Random Arrivals, and Queuing

Jody Crane, MD, MBA
jcranmd@gmail.com

Objectives

- Understand the effects of variation on throughput
- Understand the effects of scheduled and non-scheduled arrival patterns on flow
- Understand basic queuing theory and calculations

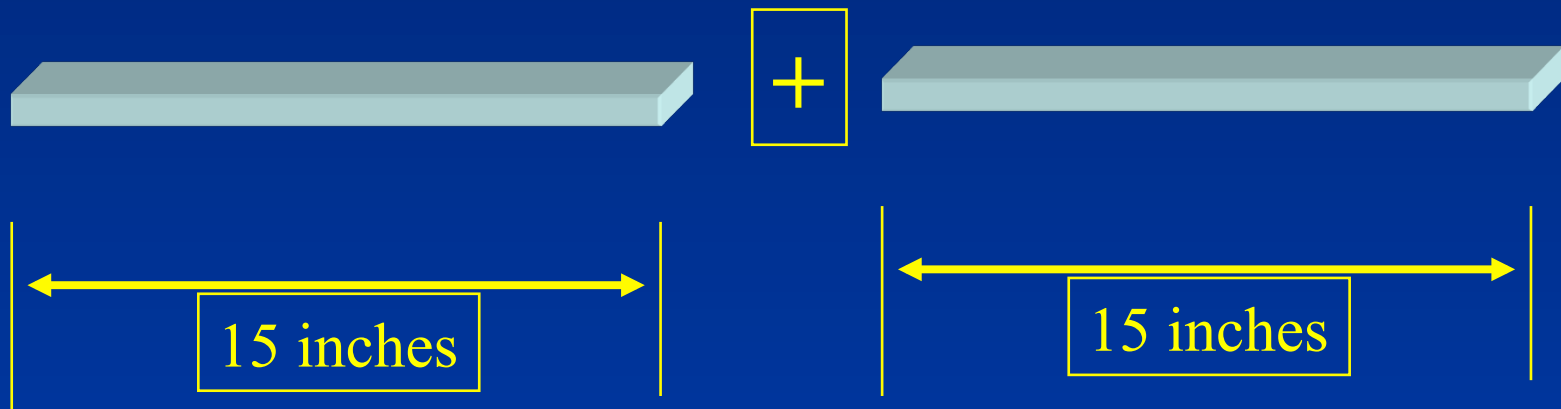
Variation



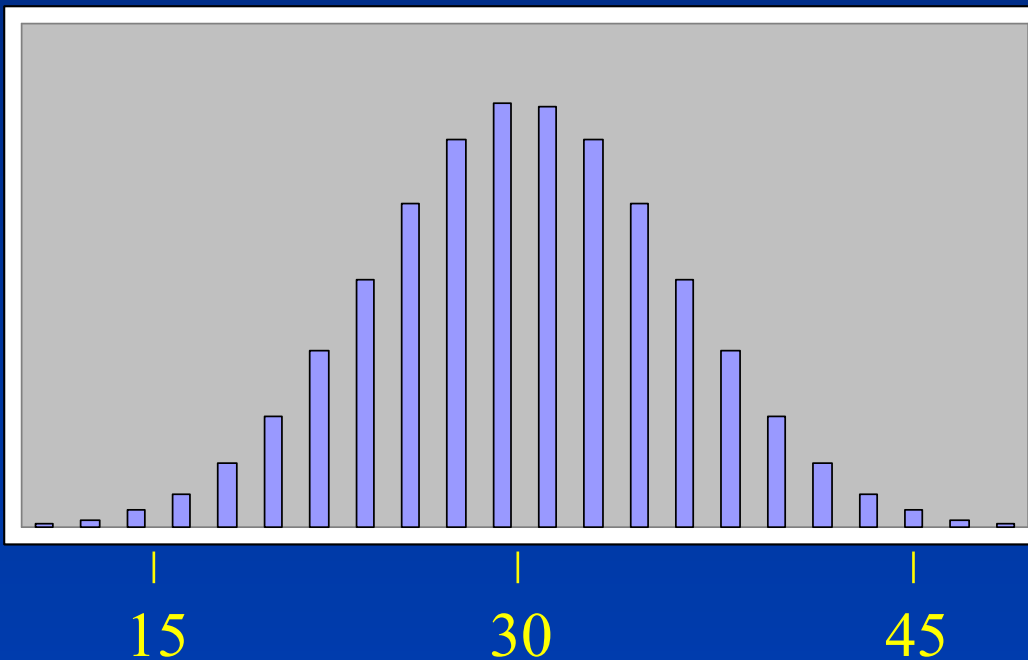
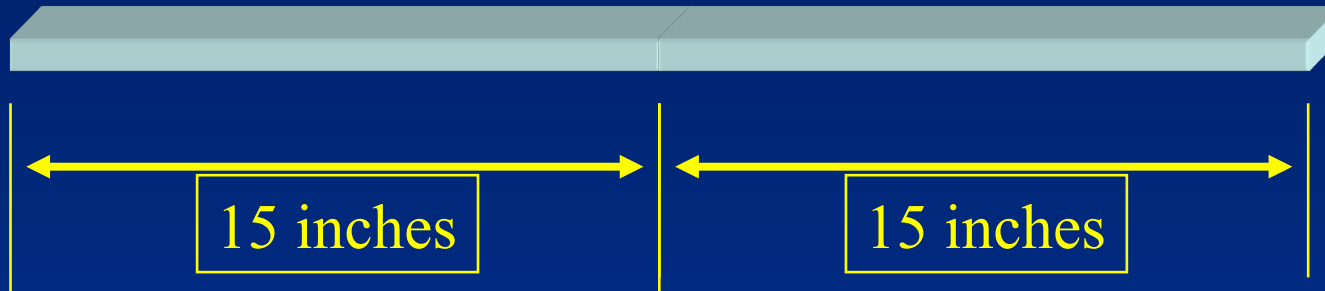
Typical expression of variation for a very bad carpenter might be as follows:

Lengths are *normally* distributed with a *mean* of 15 inches and a *standard deviation* of 5 inches.

What happens when you try to add the 2 pieces?



The Variation is Additive



Mean = 30

Standard Deviation = 7.071

Note: Variances are additive, not S.D.

$\text{SQRT}(25 + 25) = 7.071$

Clinic Example

- Day Clinic - General medical care by appointment only during the day (8am - 4pm).
- Evening Clinic - Walk-in clinic operated during the evening (4pm-midnight). No appointments accepted
- Both are staffed by a single physician.

Σ Sigma Medical

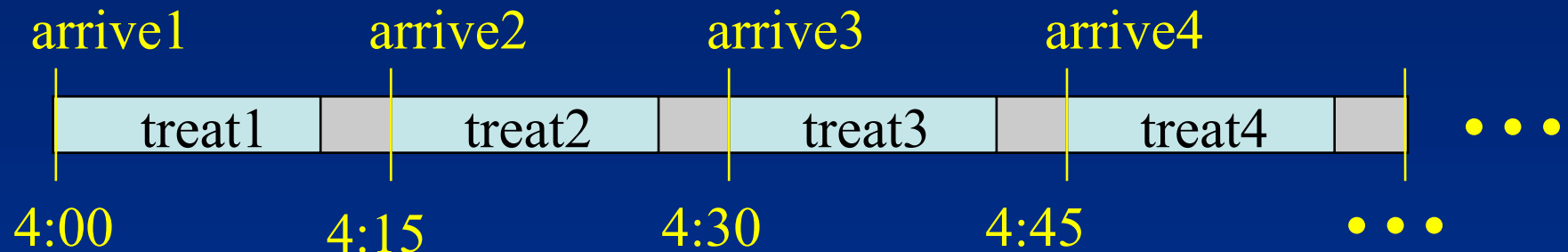


- Day Clinic - general medical care by appointment only during the day (8am - 4pm).
- Evening Clinic - immediate care walk-in clinic operated during the evening (4pm-midnight).
- Both are staffed by a single physician.

Evening Clinic - Assumptions

1. The physician is always *on time* for the start of his shift.
2. On average, 4 patients arrive per hour.
Assume 1 patient arrives *every* 15 minutes.
3. The time it takes the physician to treat a patient averages 12 minutes (can treat 5 per hour).
Assume *exactly* 12 minutes per patient.

Evening Clinic - Capacity & Performance



- Daily number of patients? 32
- Average patient wait time? 0
- Percent of patients who experience wait? 0
- Percent of time the physician is idle? 20

Evening Clinic - Assumptions

1. The physician is always *on time* for the start of his shift.

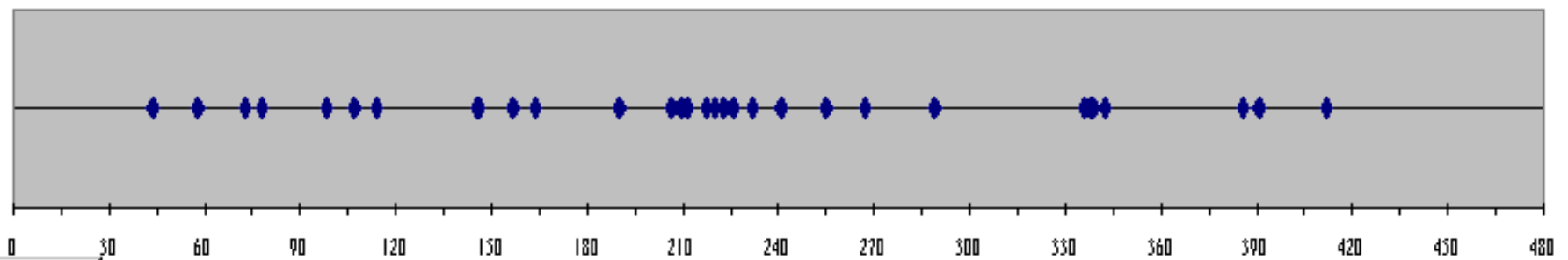
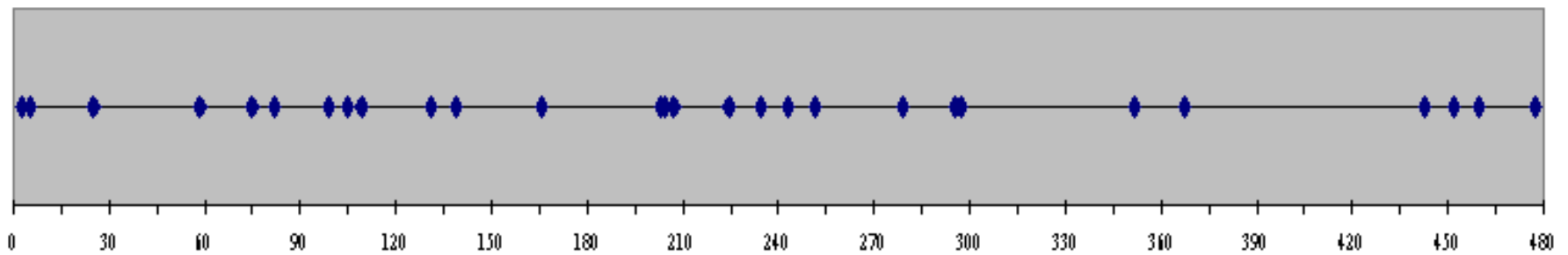
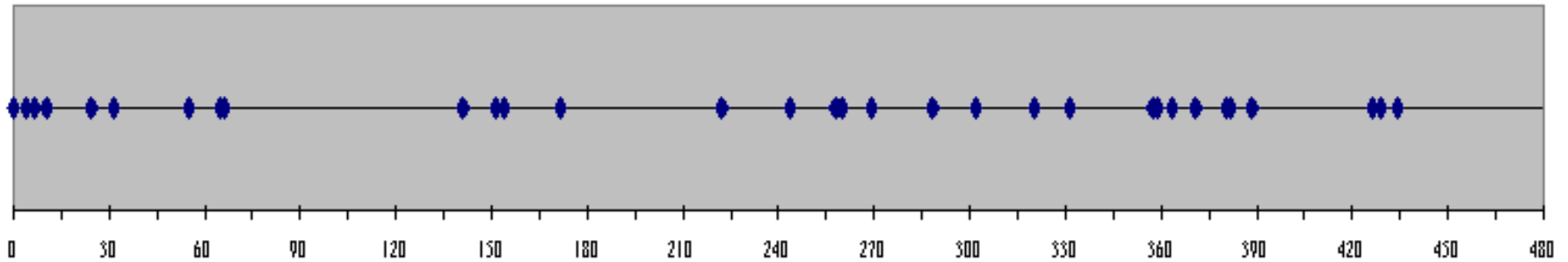
2. On average, 4 patients arrive per hour.

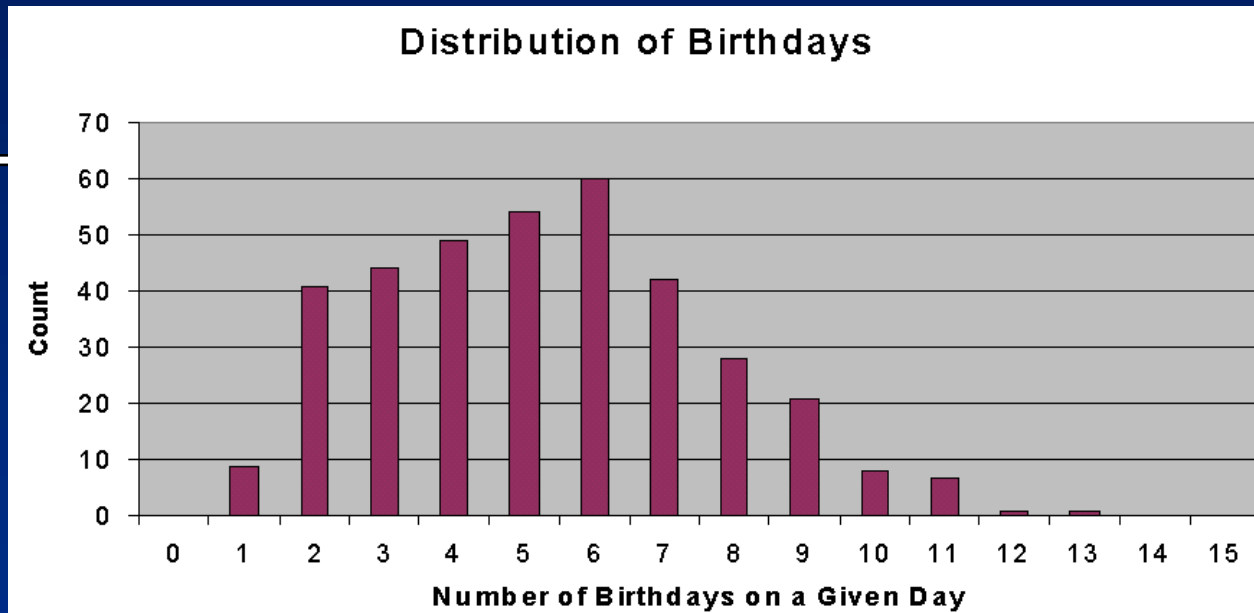
Assume ~~1 patient arrives every 15~~
minutes. **Poisson arrival process.**

3. The time it takes the physician to treat a patient averages 12 minutes (can treat 5 per hour).

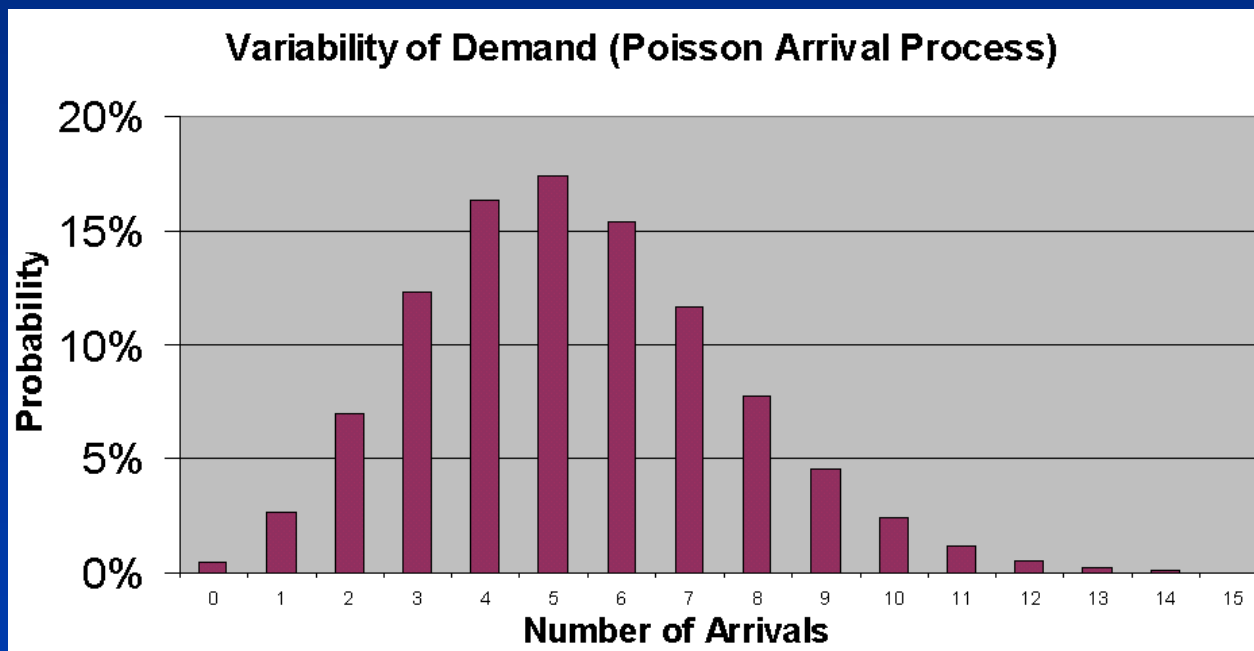
Assume *exactly* 12 minutes per patient.

Evening Clinic - Poisson Arrivals





From birthdate information for 1,938 patients



From Poisson expression using average arrival rate of 5

Evening Clinic - Assumptions

1. The physician is always *on time* for the start of his shift.

2. On average, 4 patients arrive per hour.

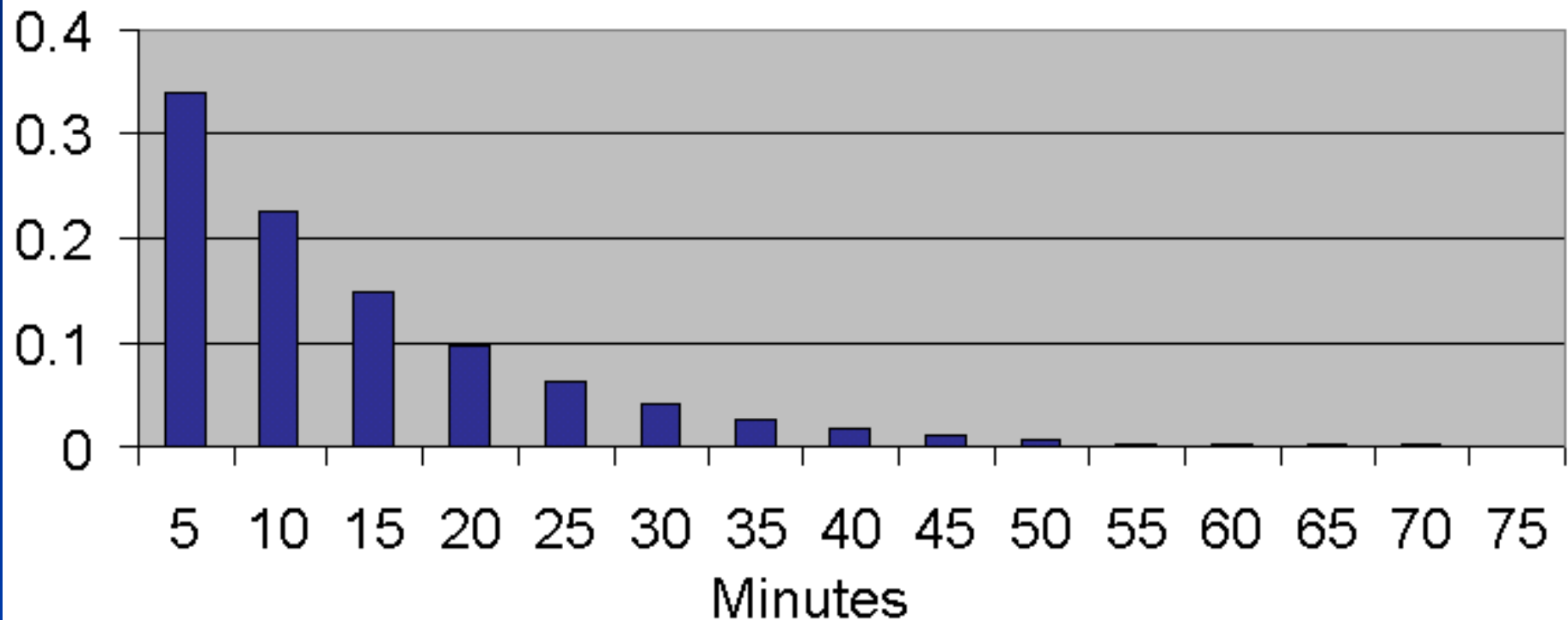
Assume ~~1 patient arrives every 15~~
minutes. Poisson arrival process.

3. The time it takes the physician to treat a patient averages 12 minutes (can treat 5 per hour)
service times exponentially distributed.

Assume ~~exactly 12 minutes per patient.~~

Exponential Distribution of Treatment Times

Distribution of Treatment Times



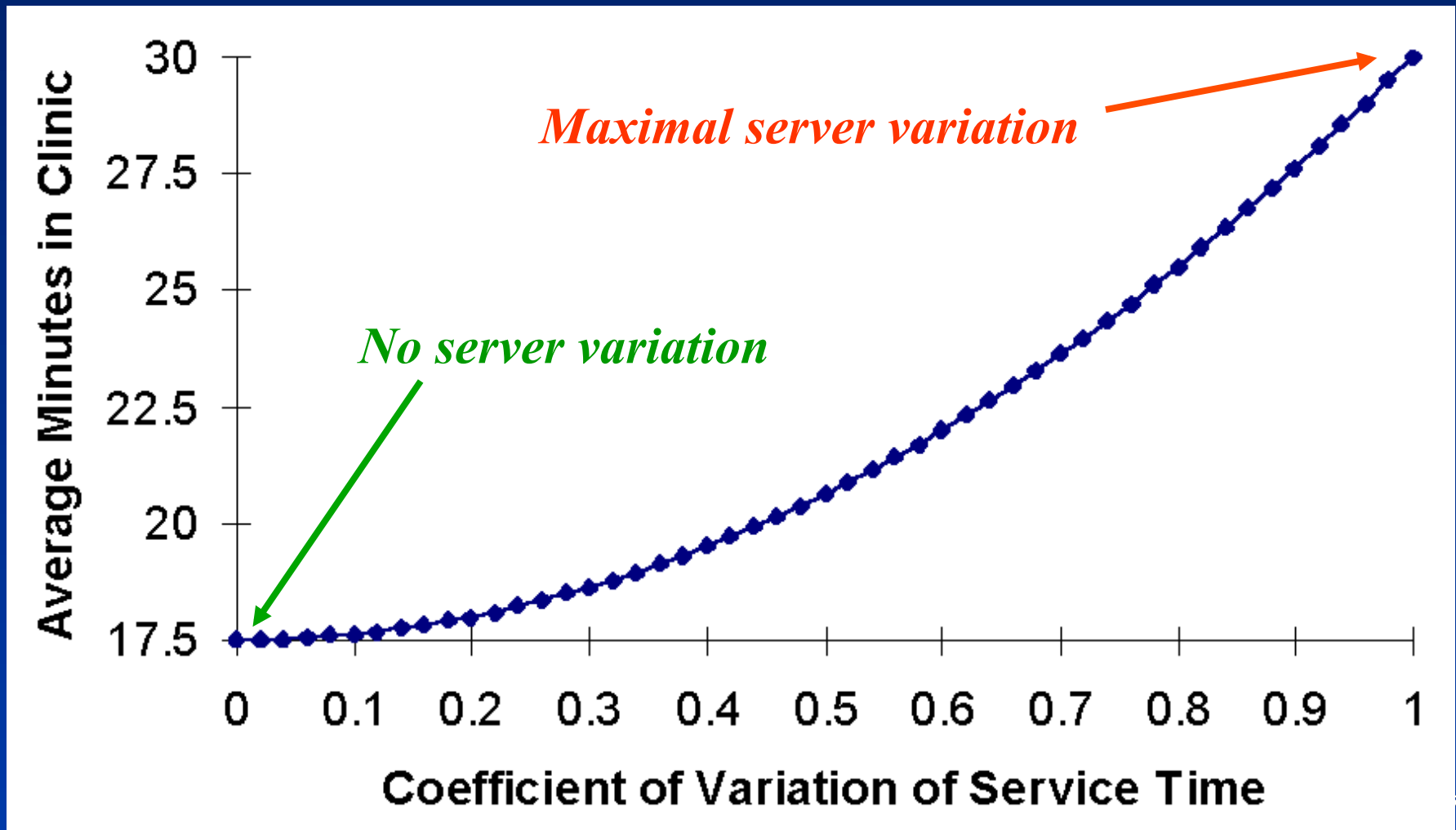
Evening Clinic - Waiting Times

- Average patient wait time? 30
- Percent of patients who experience ...
 - any wait? 70
 - more than 10 minutes of wait? 55
 - more than 30 minutes of wait? 30
 - more than 60 minutes of wait? 12
- Percent of time the physician is idle? 20
- Average number of patients waiting? 3

Evening Clinic – Take Aways

- Unscheduled arrivals tend to follow a Poisson process and impose considerable load variation on a system.

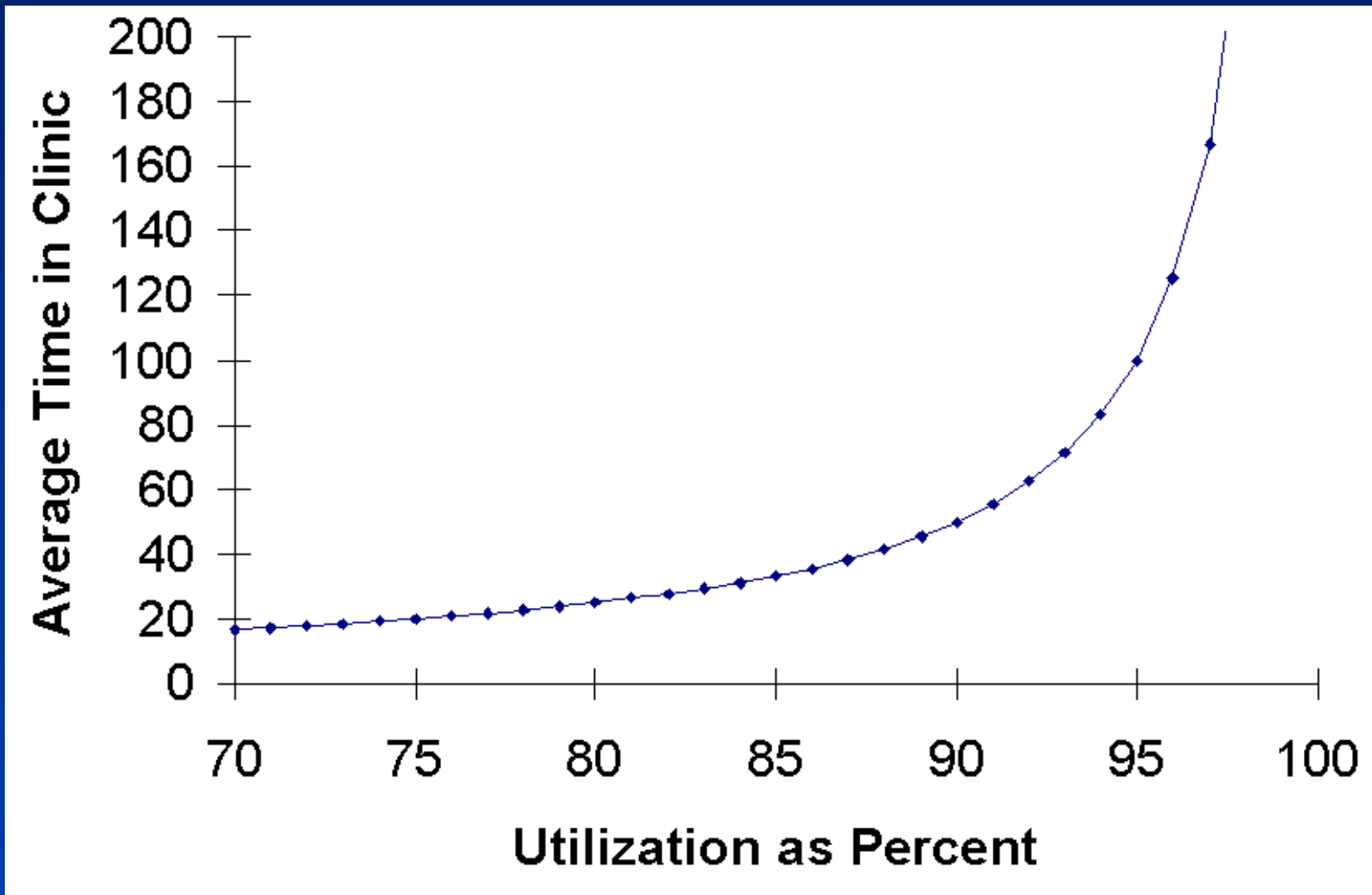
Arrival Rate of 10/hour, Service Rate of 12/hour, Server Utilization of 83.33%



Evening Clinic – Take Aways

- Unscheduled arrivals tend to follow a Poisson process and impose considerable load variation on a system.
- Even systems with built-in “capacity cushions” can exhibit extremely long waits.

Queue Behavior as a Function of Server Utilization



Evening Clinic – Take Aways

- Unscheduled arrivals tend to follow a Poisson process and impose considerable load variation on a system.
- Even systems with built-in “capacity cushions” can exhibit extremely long waits.
- Customer experiences in a walk-in system are considerably more varied than in a scheduled system.
- For either of the two service systems, a calculation of “capacity” is not easy.

Take Aways

- Identify sources of variation and try to reduce the ones that you can.
- Use information to reduce variation.
- Adjust utilization expectations according to demand variation and tolerance for waiting.
- Look for pooling opportunities but keep in mind that random demands result in random amounts of idle time.

A Basic Queue

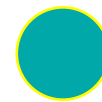
Server

A Basic Queue



A Basic Queue

Server



A Basic Queue

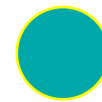


A Basic Queue



Queuing Analysis

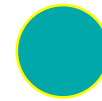
Service
Rate (μ)



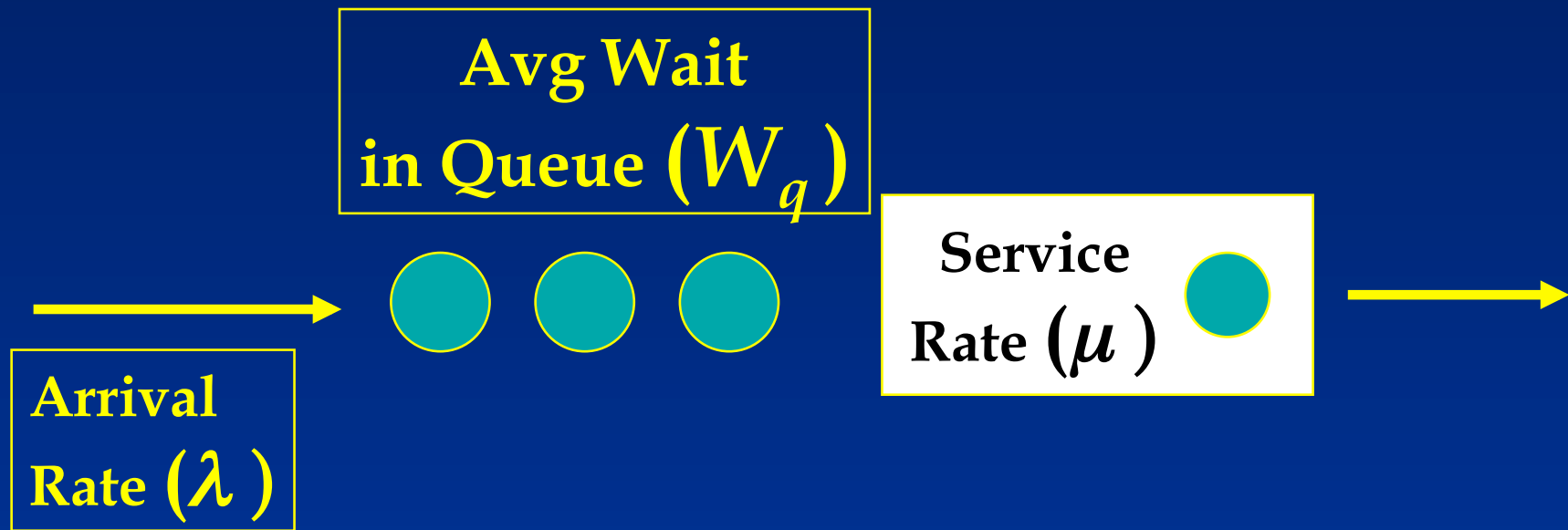
Queuing Analysis


**Arrival
Rate (λ)**

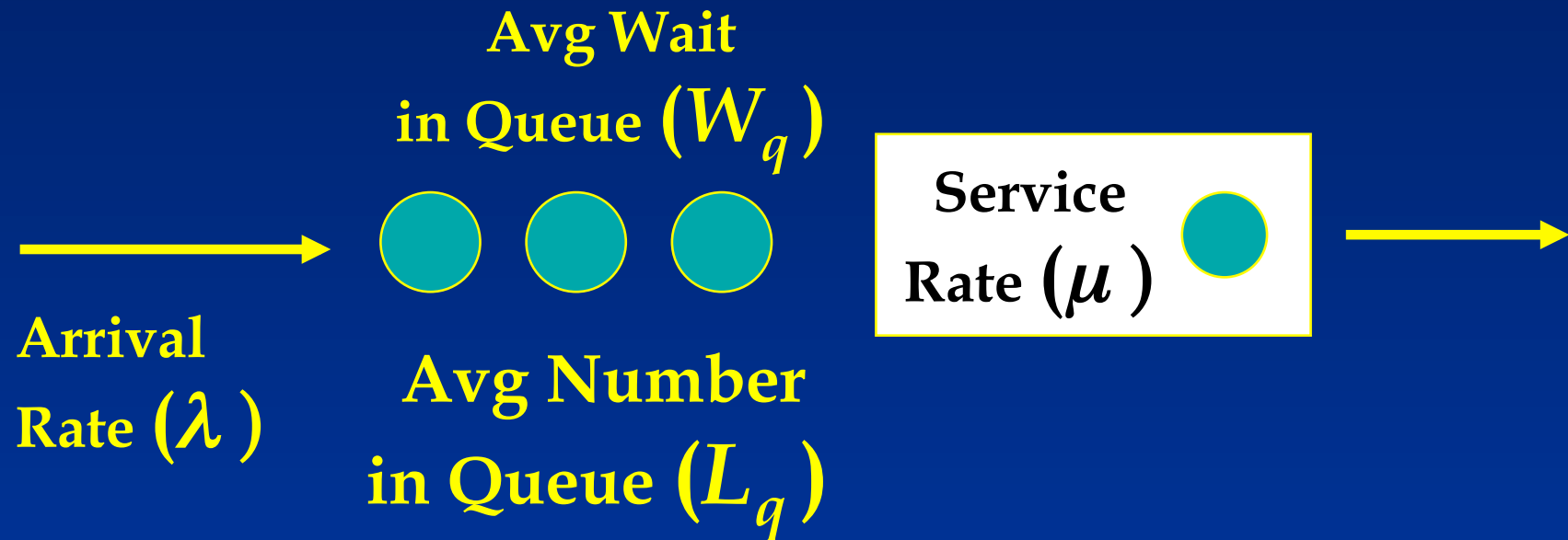
**Service
Rate (μ)**



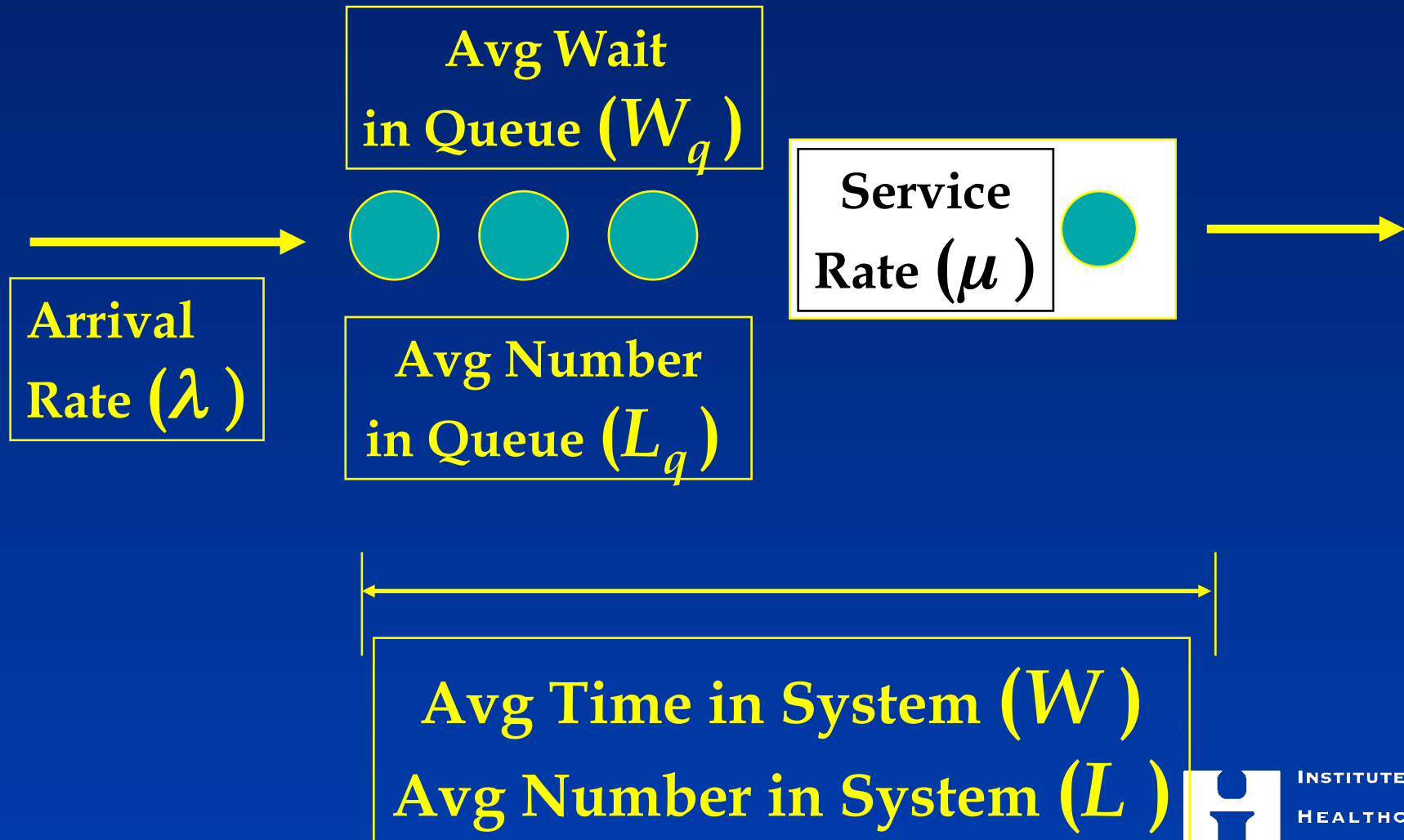
Queuing Analysis



Queuing Analysis



Queuing Analysis



Queuing Basics

For a single server system with Poisson arrivals (of rate λ) and Exponential distribution of service times (of rate μ), we can calculate steady-state estimates of:

1. Utilization (ρ)
2. Average Time in the System (W)
3. Average Number of People in the System (L)
4. Average Wait Time (Wq)
5. Average Line Length (Lq)

1. Utilization (ρ)

$$\text{Utilization } \rho = \lambda / \mu$$

For the example with $\lambda=4/\text{hour}$ and $\mu=5/\text{hour}$,

$\rho = 4/5 = .80$ or 80% utilization of physician.

2. Average Time-in-System (W)

$$W = 1 / (\mu - \lambda)$$

For the example with $\lambda=4/\text{hour}$ and $\mu=5/\text{hour}$,

$$W = 1/(5-4) = 1 \text{ hour.}$$

3. Average Number of People-in-System (L)

$$L = \lambda W$$

For the example with $\lambda=4/\text{hour}$, $\mu=5/\text{hour}$ and $W = 1/(5-4) = 1 \text{ hour}$,

$L = 4 W = 4$ people in the system.

4. Average Wait Time (W_q)

$$W_q = W - 1/\mu$$

For the example with $\lambda=4/\text{hour}$, $\mu=5/\text{hour}$ and $W = 1/(5-4) = 1 \text{ hour}$,

$$W_q = W - 1/5 = .80 \text{ hour wait.}$$

5. Average Line Length (L_q)

$$L_q = \lambda W_q$$

For the example with $\lambda=4/\text{hour}$, $\mu=5/\text{hour}$ and $W_q = 1 - 1/5 = .80$ hour,

$L_q = 4 W_q = 3.2$ people waiting in line.