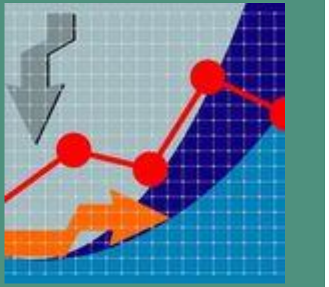


Management Science

Queuing

Dan Brandon, Ph.D., PMP

Session Objectives - Queues

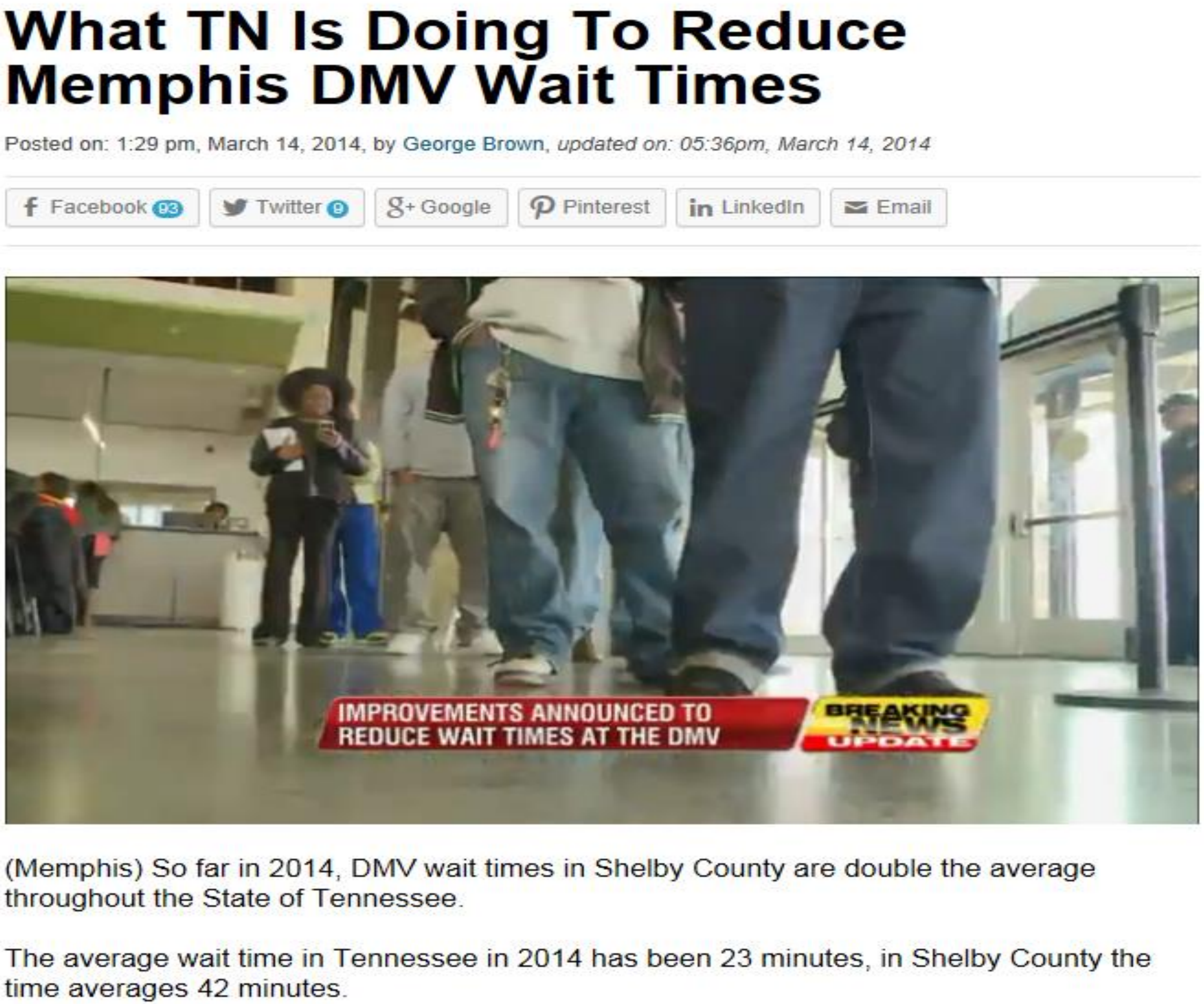
- Describe the trade-off curves for cost-of-waiting time and cost-of-service
- Understand the three parts of a queuing system: the calling population, the queue itself, and the service facility
- Describe the basic queuing system configurations
- Understand the assumptions of the common models dealt with in this chapter
- Analyze a variety of operating characteristics of waiting lines

Introduction

- *Queuing theory* is the study of *waiting lines*
- It is one of the oldest and most widely used quantitative analysis techniques
- Waiting lines are an everyday occurrence for most people
- Queues form in business process as well
- The three basic components of a queuing process are arrivals, service facilities, and the actual waiting line(s)
- Analytical models of waiting lines can help managers evaluate the cost and effectiveness of service systems

Waiting in Line

- Anybody do any waiting in your lives ?
- In what ways, both personally and for business, do you wait ?



What TN Is Doing To Reduce Memphis DMV Wait Times

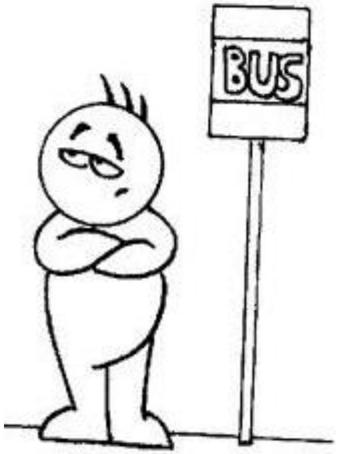
Posted on: 1:29 pm, March 14, 2014, by George Brown, *updated on: 05:36pm, March 14, 2014*

(Memphis) So far in 2014, DMV wait times in Shelby County are double the average throughout the State of Tennessee.

The average wait time in Tennessee in 2014 has been 23 minutes, in Shelby County the time averages 42 minutes.

Waiting in Line (con't)

- Anybody like to wait for service ?
  - Standing in line
  - Idling in traffic
  - Phone call on hold
  - Sitting in the doctor's office
  - Internet service delays (Affordable Health Care Act, IRS, etc.)
- How can organizations improve service ?
- What are the costs of service for an organizations ?
- What are the business implications of waiting customers wait ?



Waiting Line Costs

- Most waiting line problems are focused on finding the ideal level of service a firm should provide
- In most cases, this service level is something management can control
- A large amount of service resources generally results in high service levels of service, <u>but have high costs</u>
- Having few service resources keeps service cost down, <u>but may result in dissatisfied customers</u>

Waiting Line Costs (con't)

- There is generally a trade-off between the cost of providing service and the cost of having customers wait
- Service facilities are evaluated on their *total expected cost* which is the sum of *service costs* and *waiting costs*
- Organizations typically want to find the service level that minimizes the total expected cost

Cost Of Service Resources

- Organizations can also reduce the cost of service resources instead of reducing the amount of service resources
- How do organizations typically do this ?

- Reduce labor costs (or other resource costs)
  - Lower wages and/or benefits
  - Eliminate union costs
  - Part time workers
  - Offshore outsourcing
  - IT
    - Automated call handling
    - Web based service
    - Artificial intelligence (i.e. Expert Systems)
    - Robots

Web Economics

[10+ million shipments per day]

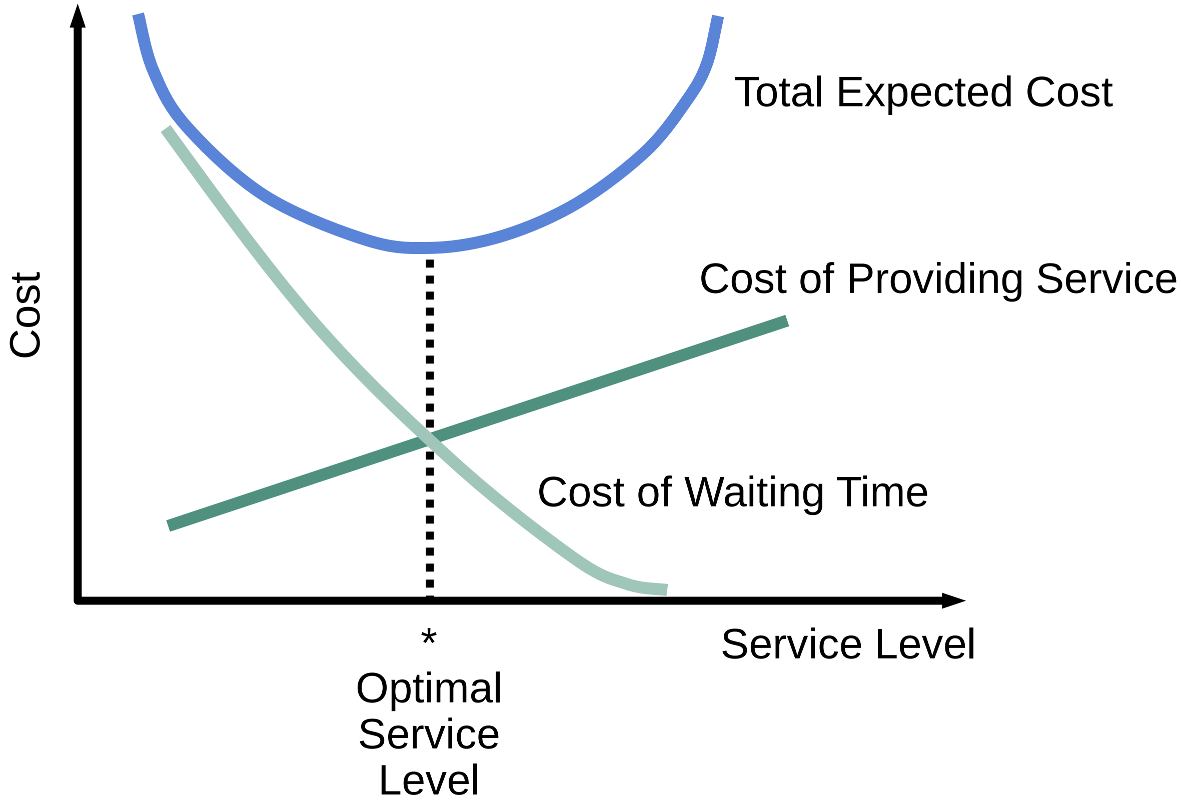
- FEDEX customer service – customers have a choice:
  - 800 Number: 60,000 calls at @ $2 cost per call
  - Website: 2 million hits @ 4 cents cost per hit

Cost Of Waiting

- Organizations can also reduce the cost for customer waiting
- How do organizations typically do this ?

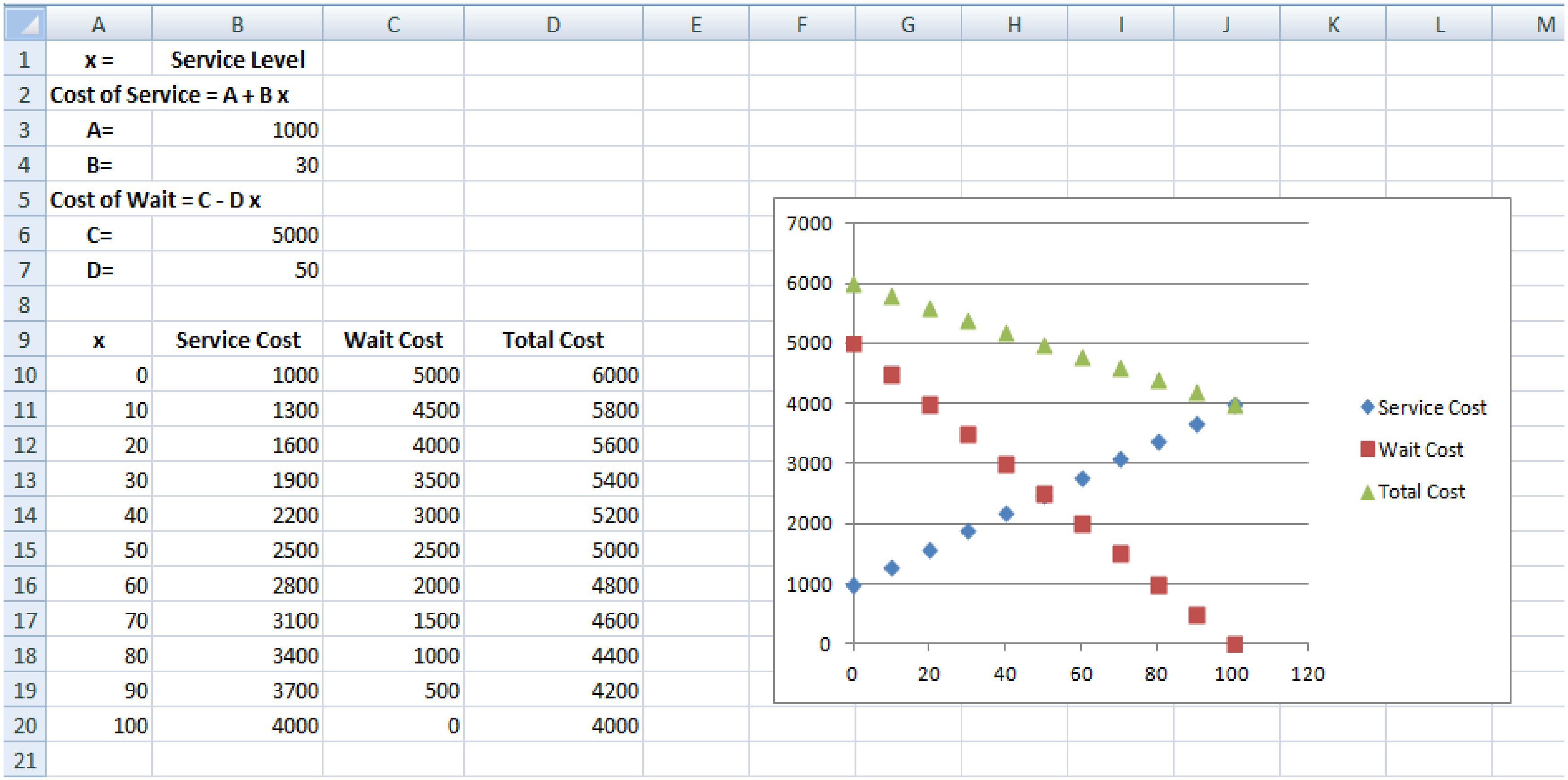
- Distract customers:
  - Mirrors
  - TV
  - Music
  - Serve food/drinks
  - Entertain customers

Service Level vs Costs



Excel Scatter Plot – Linear Functions

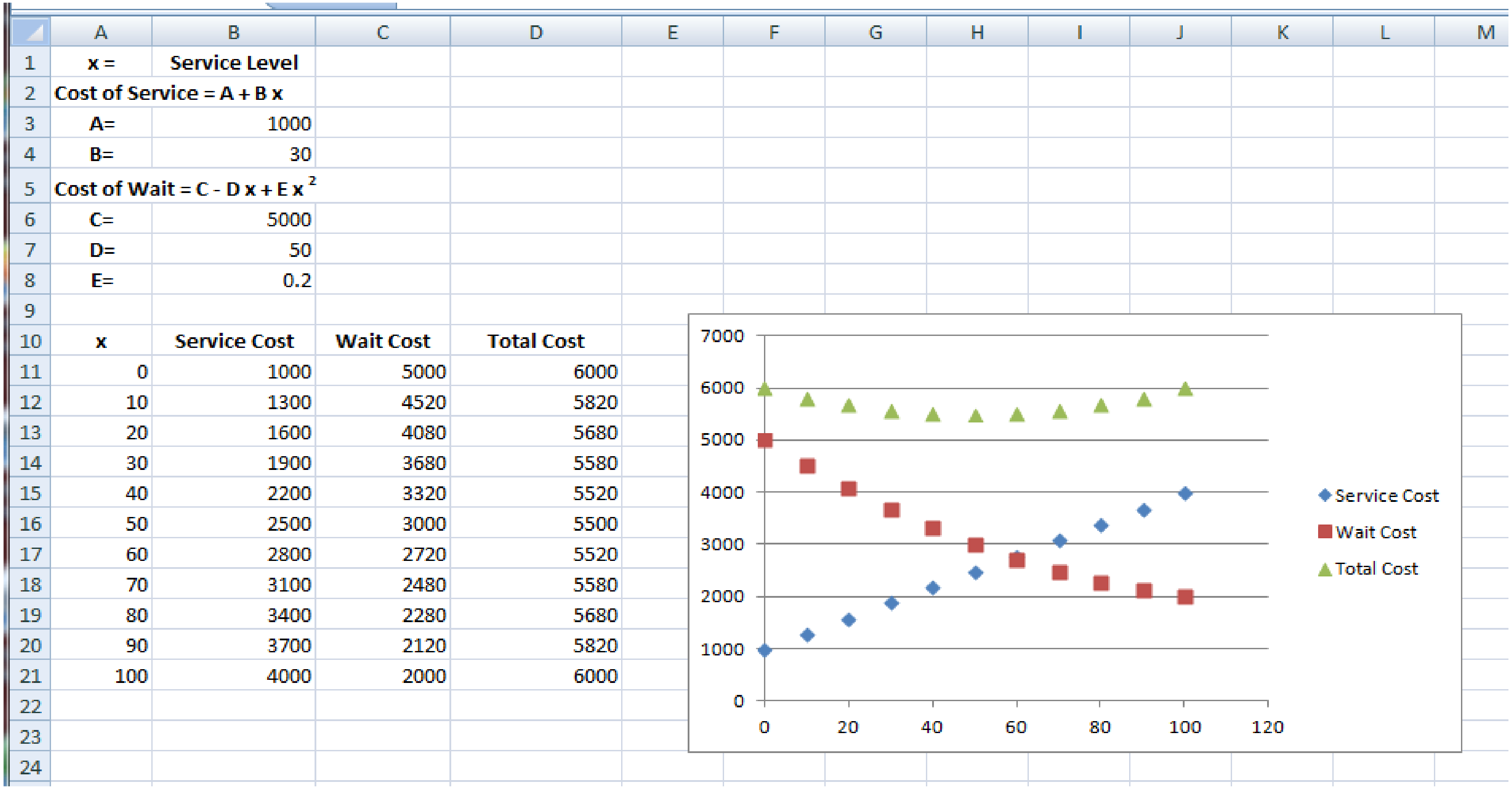
[if cost of wait is linear, min total cost will be at zero wait cost]



	A	B	C	D
1	x =	Service Level		
2	Cost of Service = A + B x			
3	A=	1000		
4	B=	30		
5	Cost of Wait = C - D x			
6	C=	5000		
7	D=	50		
8				
9	x	Service Cost	Wait Cost	Total Cost
10	0	1000	5000	6000
11	10	1300	4500	5800
12	20	1600	4000	5600
13	30	1900	3500	5400
14	40	2200	3000	5200
15	50	2500	2500	5000
16	60	2800	2000	4800
17	70	3100	1500	4600
18	80	3400	1000	4400
19	90	3700	500	4200
20	100	4000	0	4000
21				

Copyright – Dan Brandon

Excel Scatter Plot – Nonlinear Wait Cost

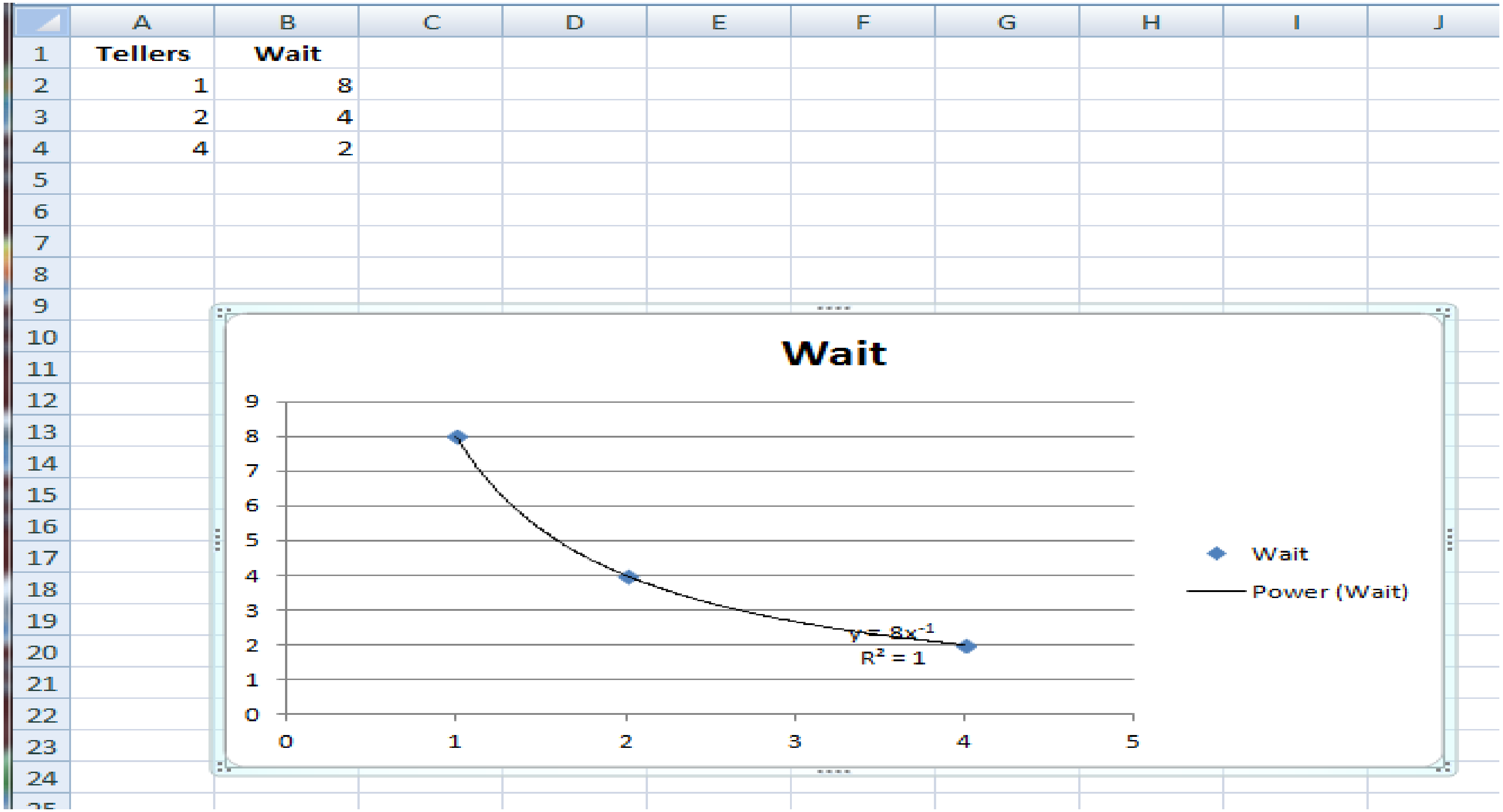


	A	B	C	D
1	x =	Service Level		
2	Cost of Service = A + B x			
3	A=	1000		
4	B=	30		
5	Cost of Wait = C - D x + E x^2			
6	C=	5000		
7	D=	50		
8	E=	0.2		
9				
10	x	Service Cost	Wait Cost	Total Cost
11	0	1000	5000	6000
12	10	1300	4520	5820
13	20	1600	4080	5680
14	30	1900	3680	5580
15	40	2200	3320	5520
16	50	2500	3000	5500
17	60	2800	2720	5520
18	70	3100	2480	5580
19	80	3400	2280	5680
20	90	3700	2120	5820
21	100	4000	2000	6000
22				
23				
24				

Wait Time is Typically Non- Linear

- Consider a bank lobby, and say one teller can serve 2 people per minute
- If there is an average of 16 people in line, a new arrival would wait 8 minutes
- With two tellers, then that new arrival would only wait 4 minutes
- And with 4 tellers, then that new arrival would only wait two minutes
- There is a non-linear relationship between wait time and the number of tellers

“Power” Relationship [wait = 8 / tellers]



	A	B	C	D	E	F	G	H	I	J
1	Tellers	Wait								
2	1	8								
3	2	4								
4	4	2								

Note use of Excel “trendline”, to get equation and R^2

Shipping Company Example

- A shipping company operates a docking facility on a river
- An average of 5 ships arrive to unload their cargos each shift
- Idle ships are expensive
- More staff can be hired to unload the ships, but that is expensive as well
- The shipping company wants to determine the optimal number of teams of stevedores to employ each shift to obtain the minimum total expected cost

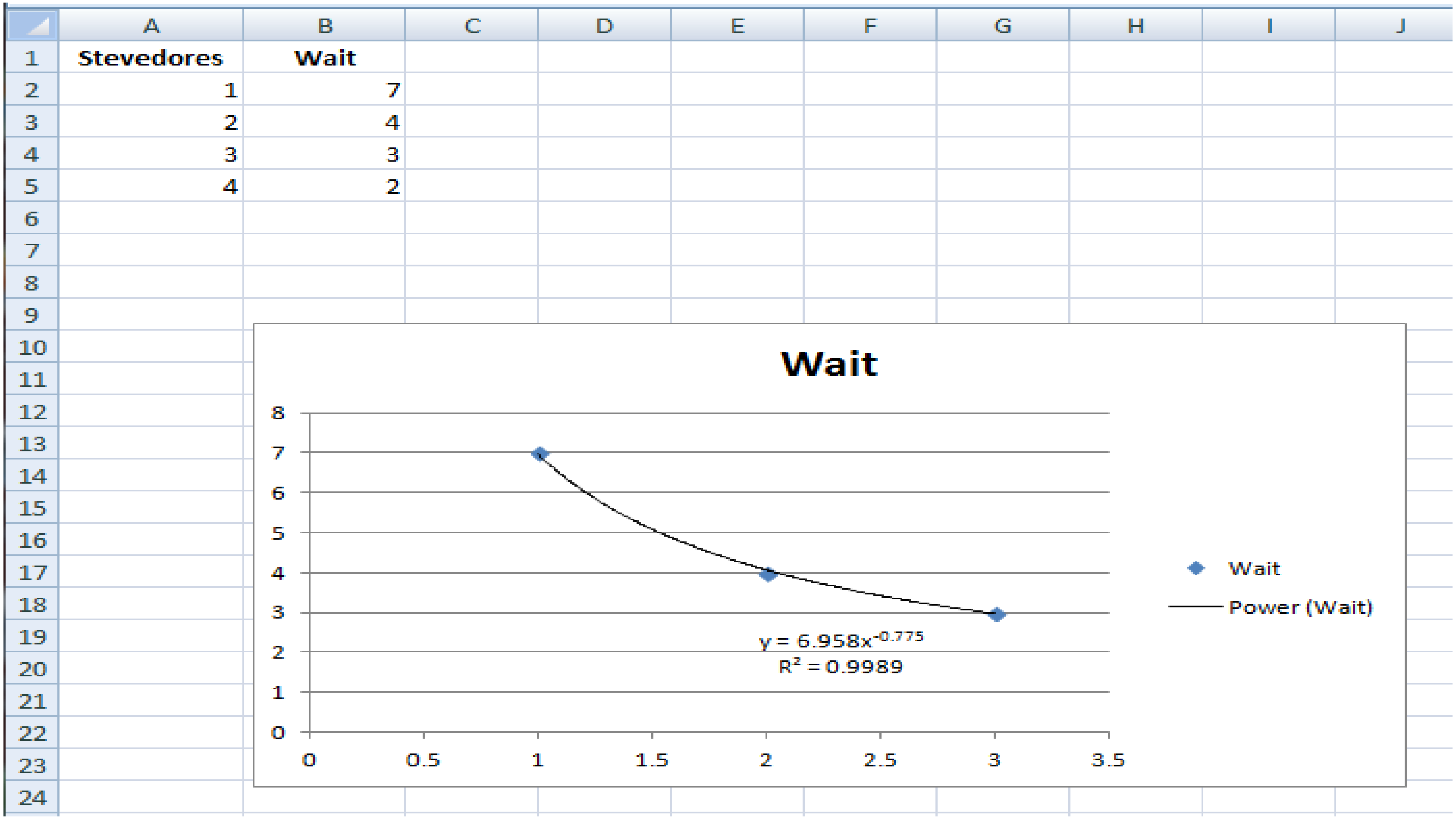
Shipping Company Example (con’t)

- Waiting line cost analysis

		NUMBER OF TEAMS OF STEVEDORES WORKING			
		1	2	3	4
(a)	Average number of ships arriving per shift	5	5	5	5
(b)	Average time each ship waits to be unloaded (hours) - nonlinear	7	4	3	2
(c)	Total ship hours lost per shift (a x b)	35	20	15	10
(d)	Estimated cost per hour of idle ship time	$1,000	$1,000	$1,000	$1,000
(e)	Value of ship’s lost time or waiting cost (c x d)	$35,000	$20,000	$15,000	$10,000
(f)	Stevedore team salary or service cost	$6,000	$12,000	$18,000	$24,000
(g)	Total expected cost (e + f)	$41,000	$32,000	$33,000	$34,000

Optimal cost

Wait Curve & Equation in Excel



	A	B
1	Stevedores	Wait
2	1	7
3	2	4
4	3	3
5	4	2

Copyright – Dan Brandon

Characteristics of a Queuing System

- There are three parts to a queuing system
  1. The arrivals or inputs to the system (sometimes referred to as the *calling population*)
  2. The queue or waiting line itself
  3. The service facility
- These components have their own characteristics that must be examined before mathematical models can be developed

Arrival Characteristics

- Even though the overall arrivals of objects to a service area may be known in total
  - For example the customers seeking teller service at the peak time between 11am and 1pm is X
- Do those customers typically arrive at a constant rate ?
  - Such that the tellers have a constant load ?

Arrival Characteristics

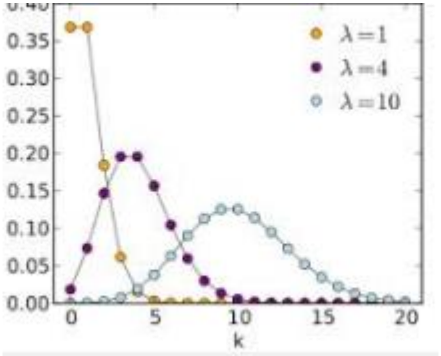
- Arrival Characteristics have three major characteristics, *size*, *pattern*, and *behavior*
  - Size of the calling population
    - Can be either unlimited (essentially *infinite*) or limited (*finite*)
  - Pattern of arrivals
    - Can arrive according to a known pattern or can arrive *randomly*
    - Random arrivals generally follow a *Poisson distribution*

Arrival Characteristics (con’t)

- The Poisson distribution is

$$P(X) = \frac{e^{-\lambda}\lambda^{X}}{X!} \text{ for } X = 0, 1, 2, 3, 4, \ldots$$

where



$P(X)$ = probability of X arrivals

X = number of arrivals per unit of time

λ = average arrival rate

e = 2.7183

The standard deviation is **equal to the square-root of the mean**

Arrival Characteristics (con't)

- We can use textbook App C to find the values of $e^{-\lambda}$
- If λ = 2, we can find the values for X = 0, 1, and 2

$$P(X) = \frac{e^{-\lambda}\lambda^X}{X!}$$

$$P(0) = \frac{e^{-2}2^0}{0!} = \frac{0.1353(1)}{1} = 0.1353 = 14\%$$

$$P(1) = \frac{e^{-2}2^1}{1!} = \frac{e^{-2}2}{1} = \frac{0.1353(2)}{1} = 0.2706 = 27\%$$

$$P(2) = \frac{e^{-2}2^2}{2!} = \frac{e^{-2}4}{2(1)} = \frac{0.1353(4)}{2} = 0.2706 = 27\%$$

Arrival Characteristics (con’t)

$$P(0) = \frac{e^{-2}2^{0}}{0!} = \frac{0.1353(1)}{1} = 0.1353 = 14\%$$

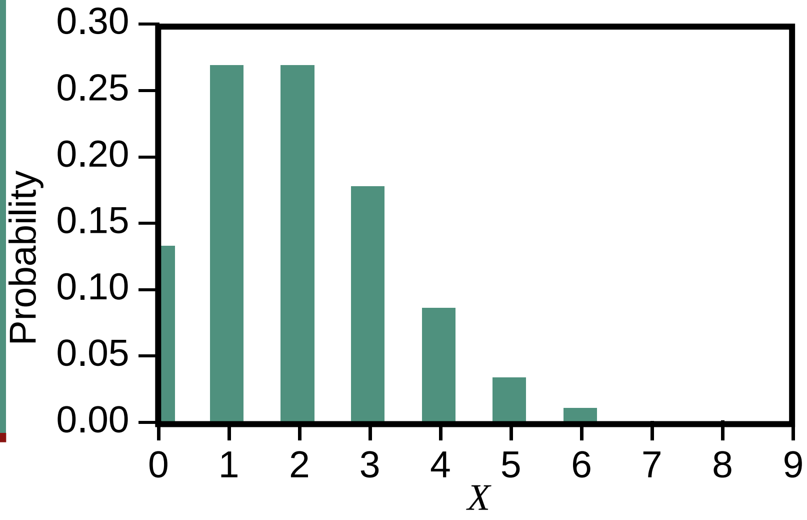
$$P(1) = \frac{e^{-2}2^{1}}{1!} = \frac{e^{-2}2}{1} = \frac{0.1353(2)}{1} = 0.2706 = 27\%$$

$$P(2) = \frac{e^{-2}2^{2}}{2!} = \frac{e^{-2}4}{2(1)} = \frac{0.1353(4)}{2} = 0.2706 = 27\%$$

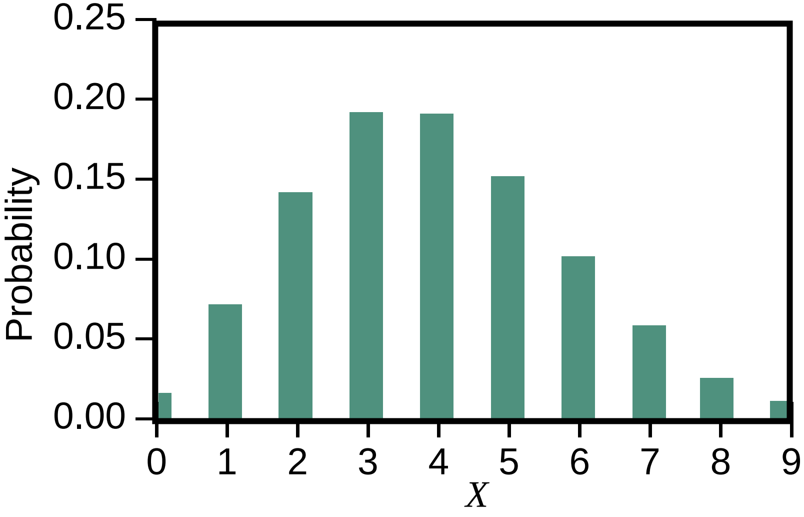
- If the average arrival rate is 2 per hour, then:
  - The probability that no one arrived this hour is 14%
  - The probability that one person arrived this hour is 27%
  - The probability that two people arrived this hour is 27%

Arrival Characteristics (con’t)

- Two examples of the Poisson distribution for arrival rates



λ = 2 Distribution

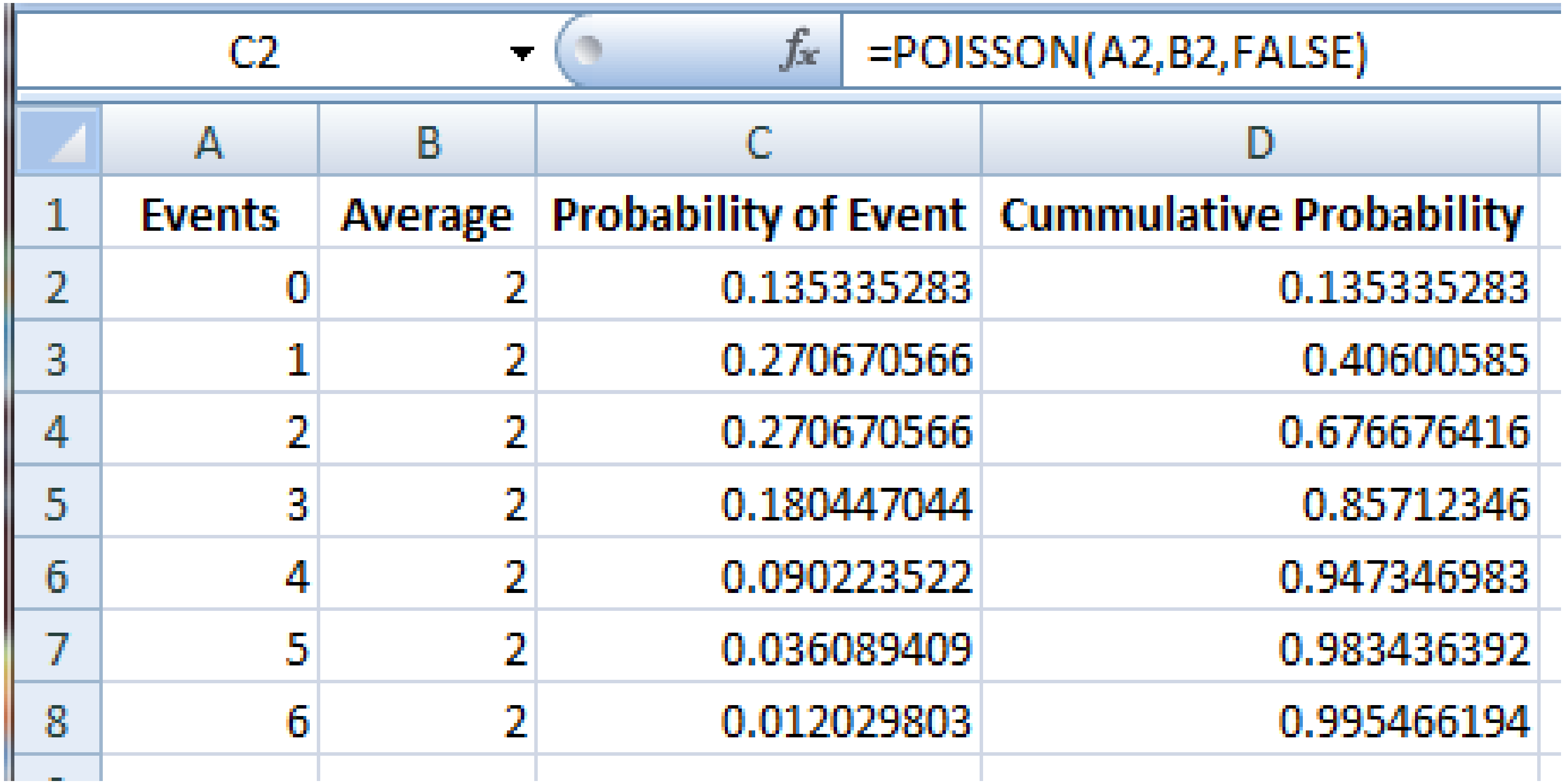


λ = 4 Distribution

Excel Poisson Function

- **POISSON**(**x**,**mean**,**cumulative**)
  - **X** is the number of events
  - **Mean** is the expected numeric value
  - **Cumulative** is “true” for a cumulative value (area) under the probability curve, or “false” for a point value

Excel Poisson Function (con't)



C2 =POISSON(A2,B2,FALSE)

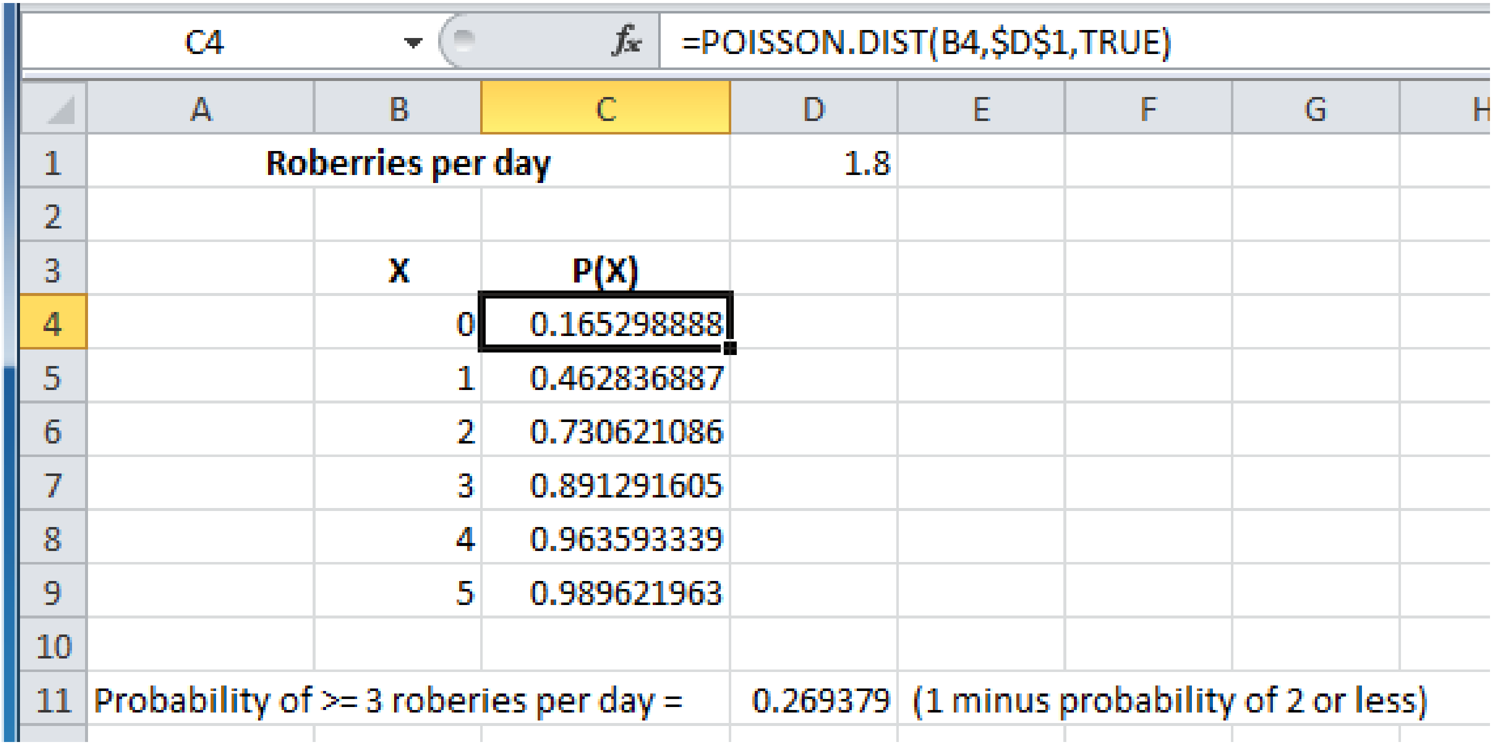
	A	B	C	D
1	Events	Average	Probability of Event	Cummulative Probability
2	0	2	0.135335283	0.135335283
3	1	2	0.270670566	0.40600585
4	2	2	0.270670566	0.676676416
5	3	2	0.180447044	0.85712346
6	4	2	0.090223522	0.947346983
7	5	2	0.036089409	0.983436392
8	6	2	0.012029803	0.995466194

Poisson Exercise

- In our city for the last year there have averaged 1.8 bank robberies per day
- The police can handle 2 robberies per day
- What is the probability there will be 3 or more robberies on any given day ?

- Do not look ahead !

Poisson Example (con’t)

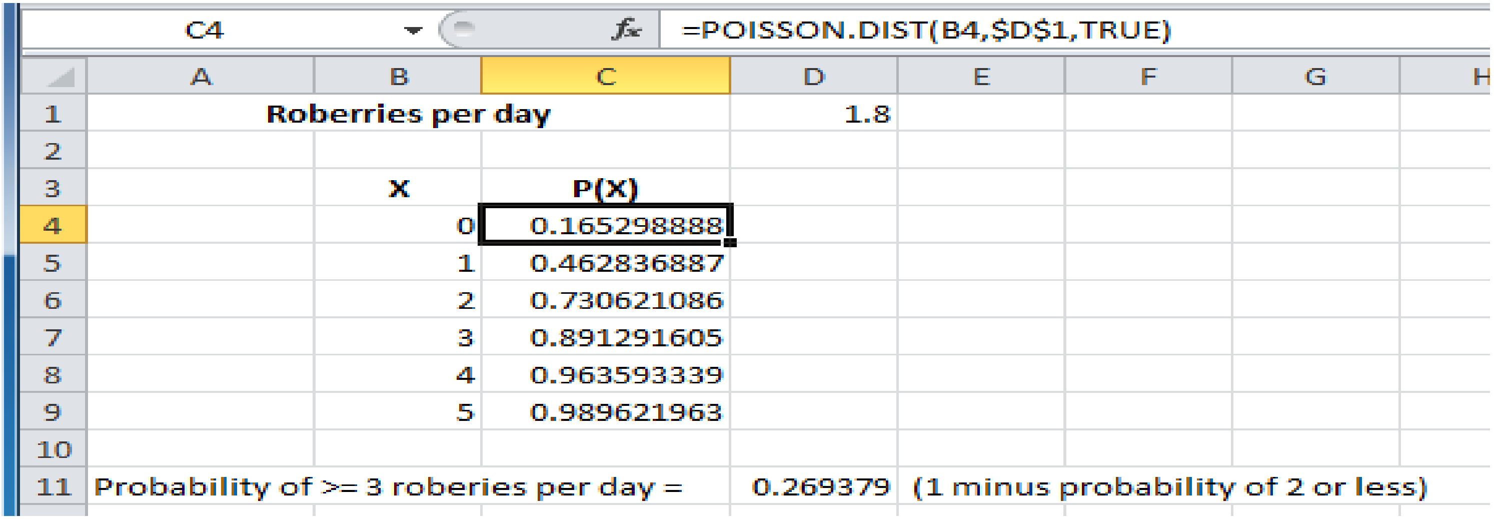


C4 =POISSON.DIST(B4,$D$1,TRUE)

	A	B	C	D	E	F	G	H
1	Roberries per day			1.8				
2								
3		X	P(X)					
4		0	0.165298888					
5		1	0.462836887					
6		2	0.730621086					
7		3	0.891291605					
8		4	0.963593339					
9		5	0.989621963					
10								
11	Probability of >= 3 roberies per day =			0.269379	(1 minus probability of 2 or less)			

Copyright – Dan Brandon

- So even though the police have set enough capacity (2) to handle over the average of 1.8 bank robberies per day, 27% of the time there will be more than they can handle



C4 =POISSON.DIST(B4,$D$1,TRUE)

	A	B	C	D	E	F	G	H
1	Roberries per day			1.8				
2								
3		X	P(X)					
4		0	0.165298888					
5		1	0.462836887					
6		2	0.730621086					
7		3	0.891291605					
8		4	0.963593339					
9		5	0.989621963					
10								
11	Probability of >= 3 roberies per day =			0.269379	(1 minus probability of 2 or less)			

Loss Productivity

- If a bank provides enough tellers to handle their average load, there will still be:
  - Times that more than average number of people will arrived resulting in wait time
  - Times that less than average people will arrive resulting in idle tellers (lost productivity)
  - That lost productivity can never be regained

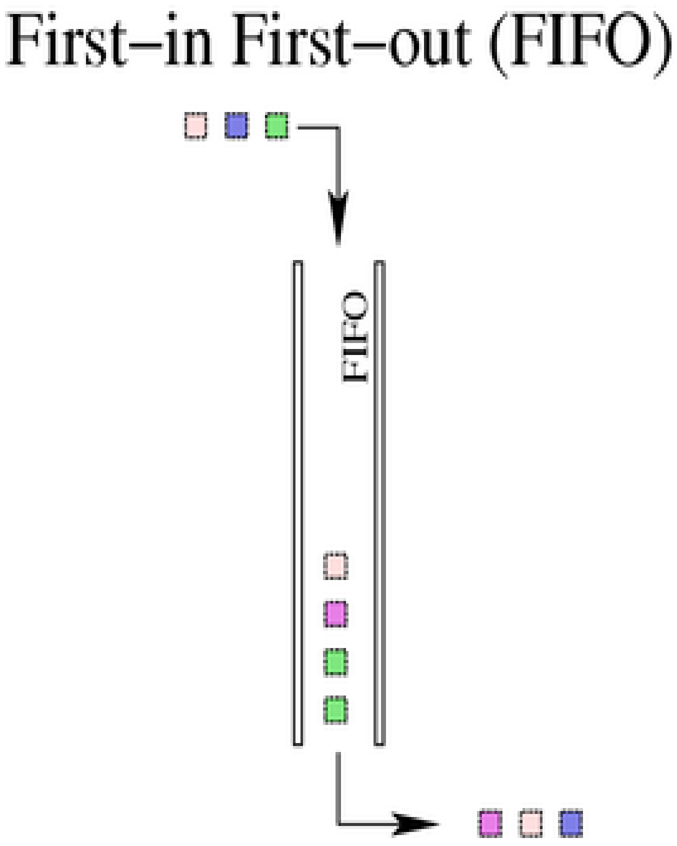
Behavior of Arrivals

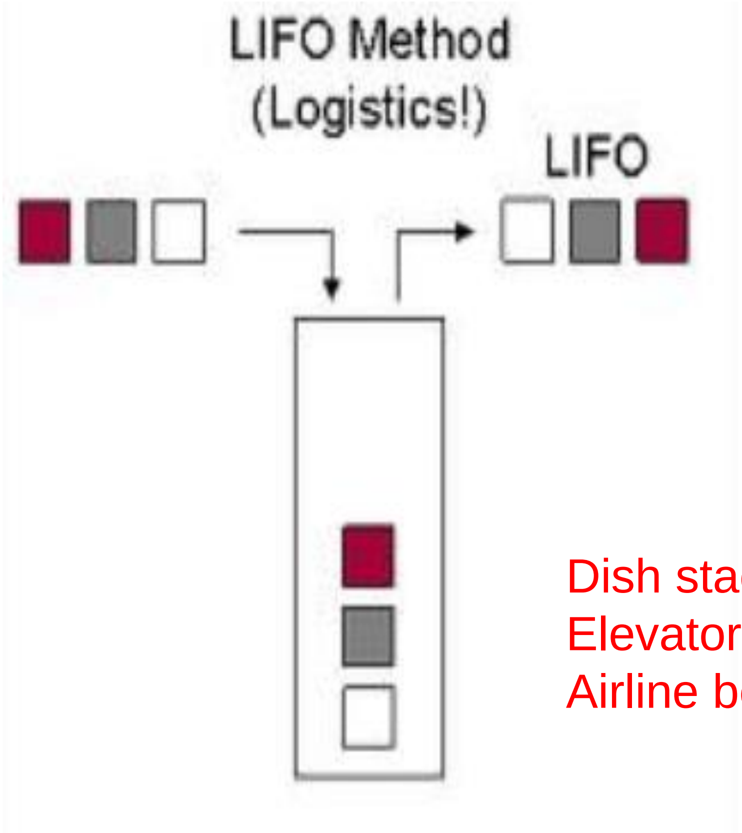
- Most queuing models assume customers are patient and will wait in the queue until they are served and do not switch lines
- *Balking* refers to customers who refuse to join the queue
- *Reneging* customers enter the queue but become impatient and leave without receiving their service
- That these behaviors can exist is an argument for the use of more complex queuing theory to managing waiting lines in some situations

Waiting Line Characteristics

- Waiting lines can be either *limited* or *unlimited*
- Queue discipline refers to the rule by which customers in the line receive service
- The most common rule is *first-in, first-out* (*FIFO*)
- Other rules are possible and may be based on other important characteristics
- Other rules can be applied to select which customers enter which queue, but may apply FIFO once they are in the queue
- Multiple queue types (i.e. express checkout)

FIFO (queue) & LIFO (stack)





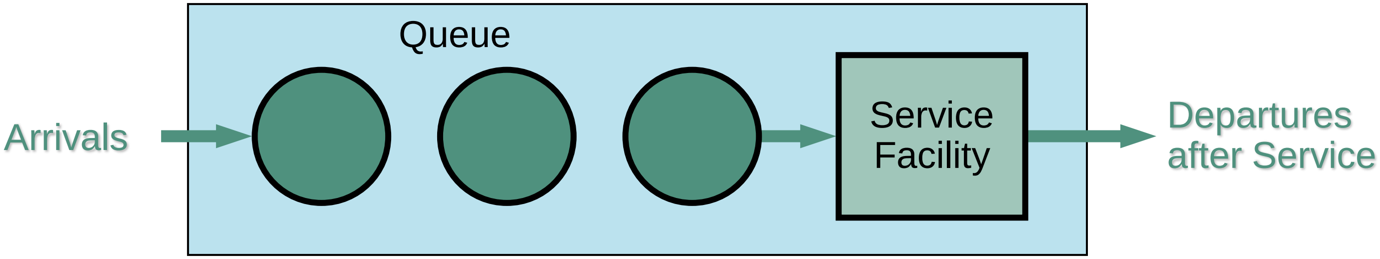
Dish stack
Elevator
Airline boarding

Service Facility Characteristics

- Basic queuing system configurations
  - Service systems are classified in terms of the number of channels (or servers), and the number of phases (service stops)
    - A *single-channel system* with one server is quite common
    - *Multichannel systems* exist when multiple servers are fed by one common waiting line
    - In a *single-phase system* the customer receives service from just one server
    - If a customer has to go through more than one server, it is a *multiphase system*

Service Facility Characteristics (con’t)

- Four basic queuing system configurations



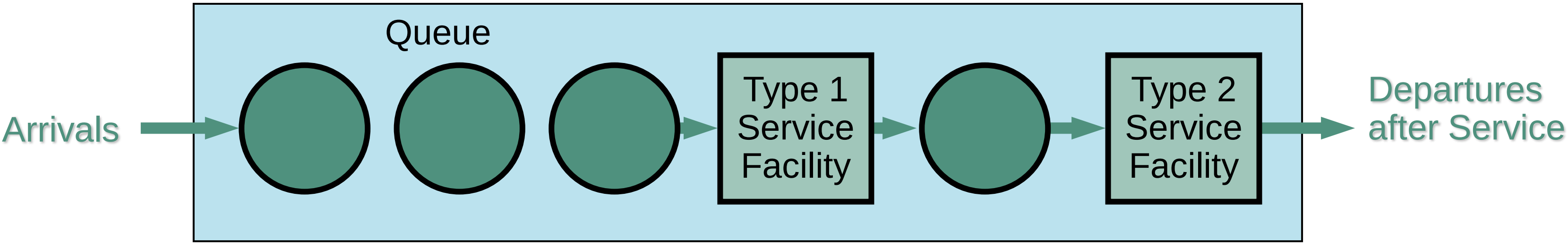
Queue

Arrivals

Service Facility

Departures after Service

Single-Channel, Single-Phase System



Queue

Arrivals

Type 1 Service Facility

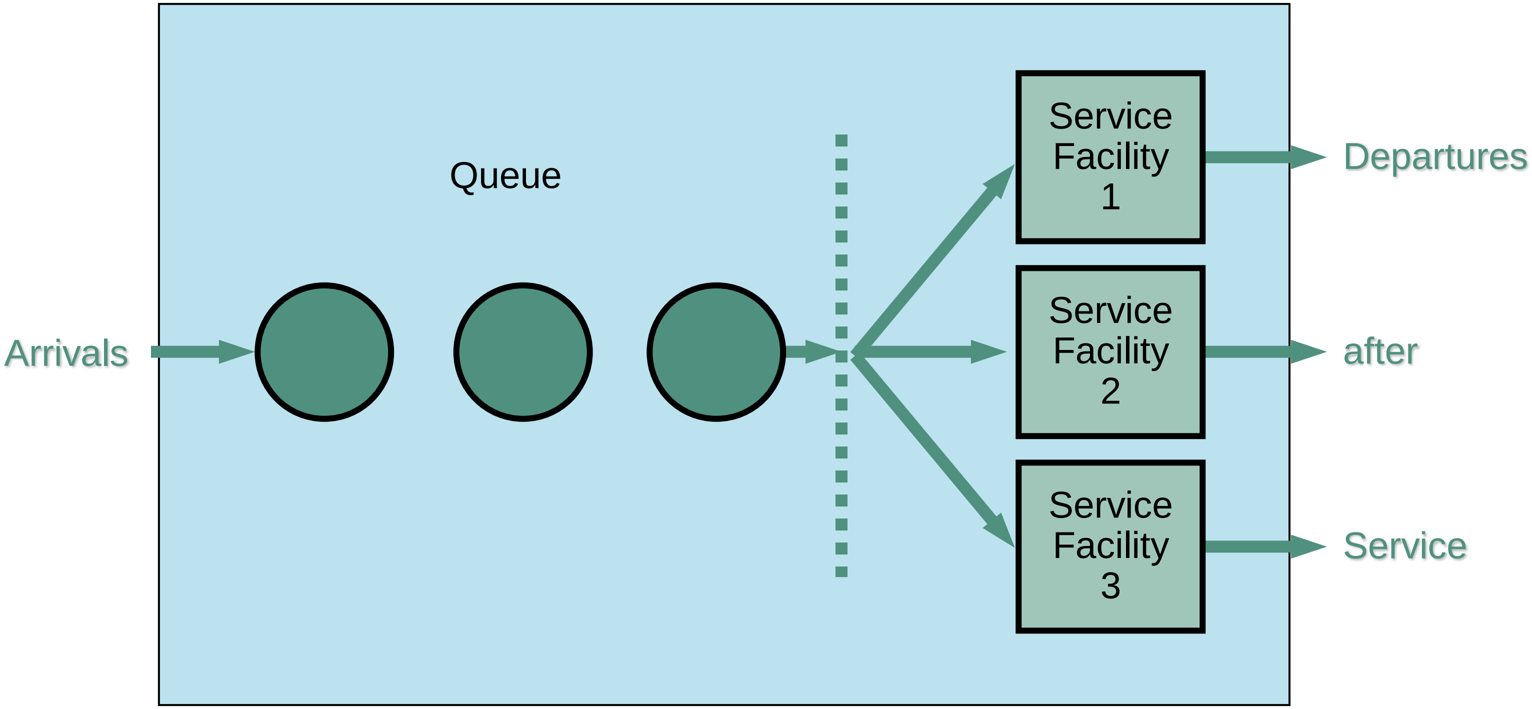
Type 2 Service Facility

Departures after Service

Single-Channel, Multiphase System

Service Facility Characteristics (con't)

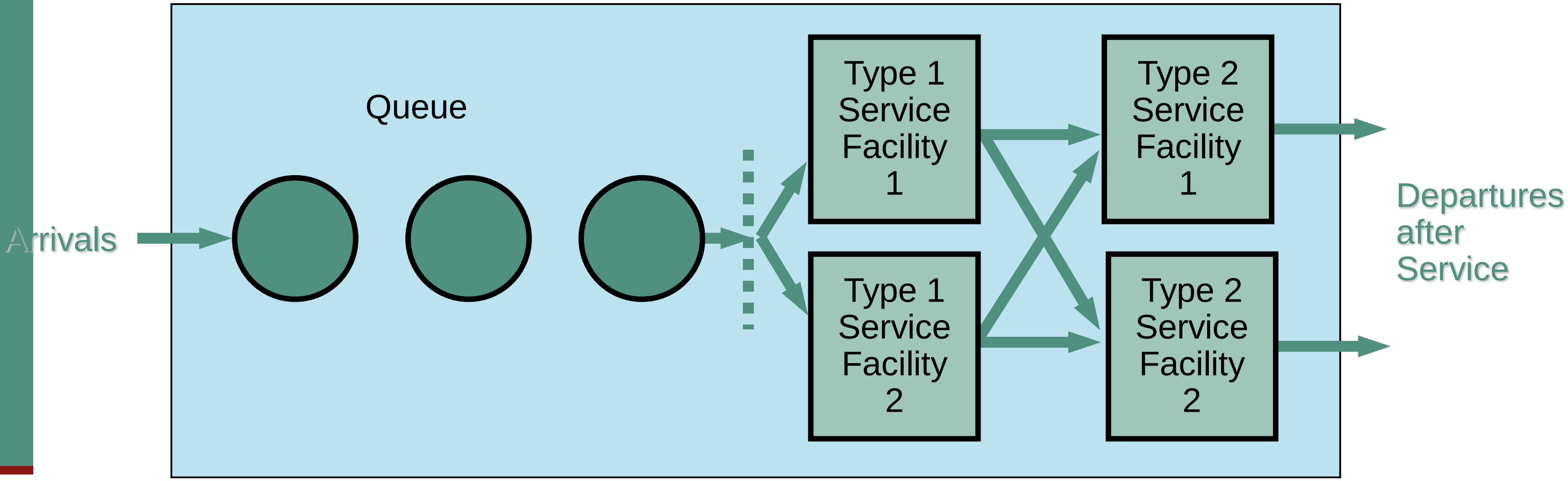
- Four basic queuing system configurations



Multichannel (multiple servers), Single-Phase System

Service Facility Characteristics (con't)

- Four basic queuing system configurations


Queue
Arrivals
Type 1 Service Facility 1
Type 2 Service Facility 1
Type 1 Service Facility 2
Type 2 Service Facility 2
Departures after Service

Multichannel, Multiphase System

Service Time Distribution

- Service patterns can be either constant or random
- Constant service times are often machine controlled
- More often, service times are randomly distributed according to a *negative exponential probability distribution*
- Models are based on the assumption of particular probability distributions
- One should take care to ensure observations fit the assumed distributions when applying these models

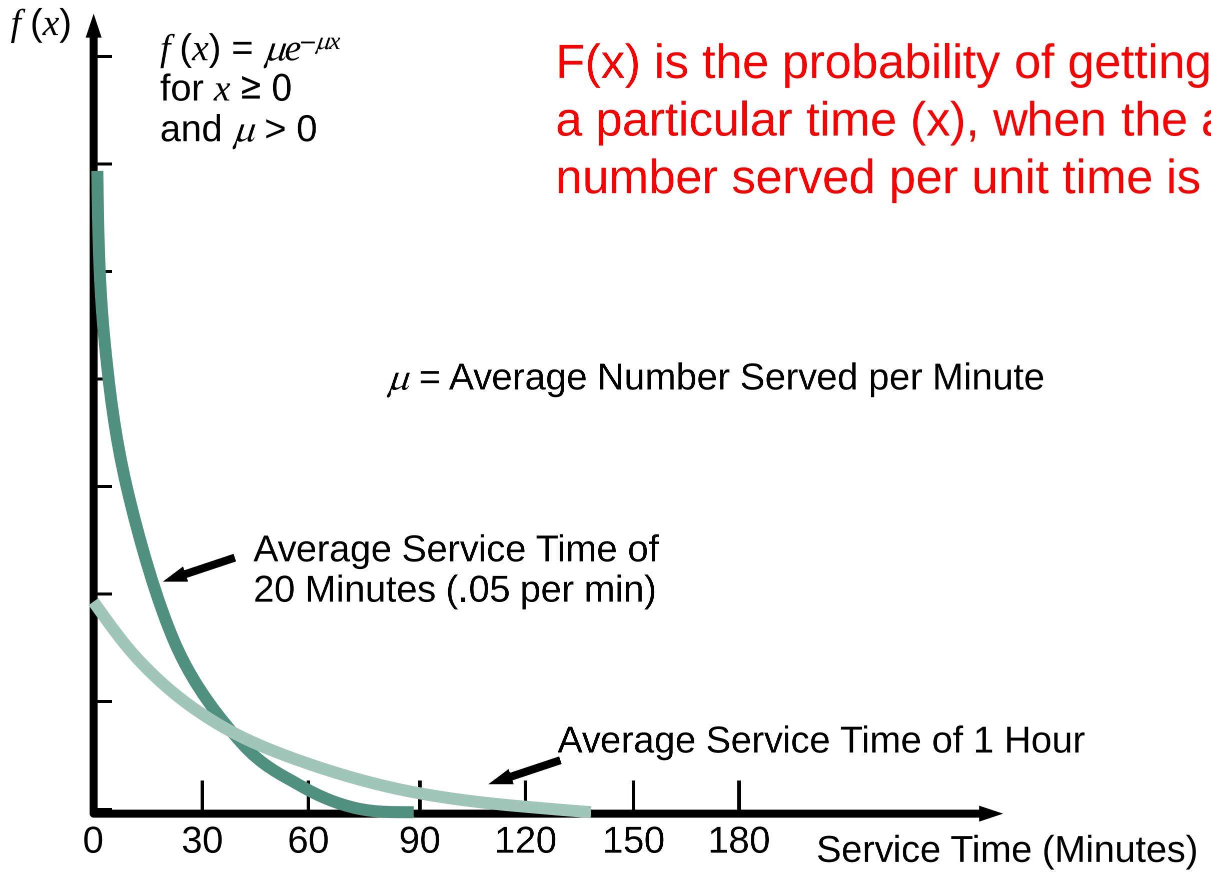
Service Time Distribution (con't)

- Take a bank for example
- Does it take the same amount of time to service each customer ?

Copyright – Dan Brandon

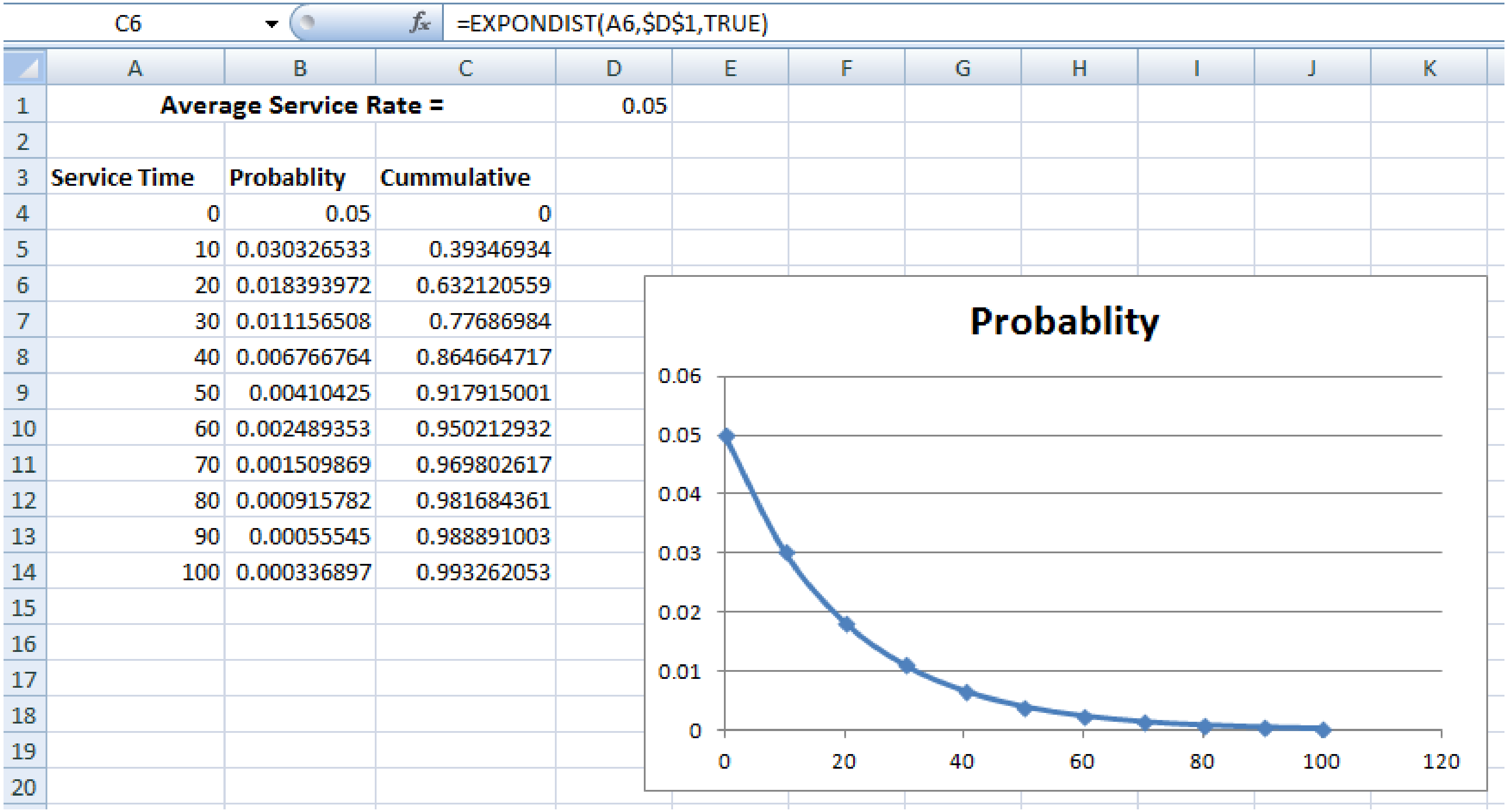
Service Time Distribution (con’t)

- Two examples of exponential distribution for service times



F(x) is the probability of getting served in a particular time (x), when the average number served per unit time is u

Excel Exponential Function



C6 =EXPONDIST(A6,$D$1,TRUE)

	A	B	C	D
1	Average Service Rate =			0.05
2				
3	Service Time	Probablity	Cummulative	
4	0	0.05	0	
5	10	0.030326533	0.39346934	
6	20	0.018393972	0.632120559	
7	30	0.011156508	0.77686984	
8	40	0.006766764	0.864664717	
9	50	0.00410425	0.917915001	
10	60	0.002489353	0.950212932	
11	70	0.001509869	0.969802617	
12	80	0.000915782	0.981684361	
13	90	0.00055545	0.988891003	
14	100	0.000336897	0.993262053	

Copyright – Dan Brandon

Identifying Models Using Kendall Notation

- D. G. Kendall developed a notation for queuing models that specifies the pattern of arrival, the service time distribution, and the number of channels
- It is of the form

Arrival distribution / Service time distribution / Number of service channels open

- Specific letters are used to represent probability distributions

  M = Poisson distribution for arrival, or exponential distribution for service (usual cases)

  D = constant (deterministic) rate

  G = normal distribution with known mean and variance

Kendall Notation (con't)

- So a single channel model with Poisson arrivals and exponential service times would be represented by

$$M/M/1$$

- If a second channel (server) is added we would have

$$M/M/2$$

- A three channel (server) system with Poisson arrivals and constant service time would be

$$M/D/3$$

- A four channel (server) system with Poisson arrivals and normally distributed service times would be

$$M/G/4$$

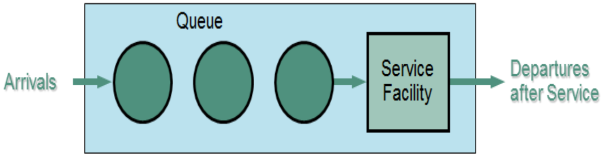
Single-Channel Model, Poisson Arrivals, Exponential Service Times (M/M/1)

- Assumptions of the model
  - Arrivals are served on a FIFO basis
  - No balking or reneging
  - Arrivals are independent of each other but rate is constant over time
  - Arrivals follow a Poisson distribution
  - Service times are variable and independent but the average is known
  - Service times follow a negative exponential distribution
  - Average service rate is greater than the average arrival rate

M/M/1 (con't)

- When these assumptions are met, we can develop a series of equations that define the queue's *operating characteristics*

- Queuing Equations



  - We let

λ = mean number of arrivals per time period

μ = mean number of people or items served per time period

M/M/1 (con't)

1. The average number of customers or units in the system, L

$$L = \frac{\lambda}{\mu - \lambda}$$

2. The average time a customer spends in the system, W

$$W = \frac{1}{\mu - \lambda}$$

3. The average number of customers in the queue, L_q

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)}$$

M/M/1 (con’t)

4. The average time a customer spends waiting in the queue, W_q

$$W_q = \frac{\lambda}{\mu(\mu-\lambda)}$$

5. The *utilization factor* for the system, ρ, the probability the service facility is being used

$$\rho = \frac{\lambda}{\mu}$$

M/M/1 (con’t)

6. The percent idle time, P_0, the probability no one is in the system

$$P_0 = 1 - \frac{\lambda}{\mu}$$

7. The probability that the number of customers in the system is greater than k, $P_{n>k}$

$$P_{n>k} = \left(\frac{\lambda}{\mu}\right)^{k+1}$$

Muffler Shop Case

- A mechanic can install mufflers at a rate of 3 per hour
- Customers arrive at a rate of 2 per hour

λ = 2 cars arriving per hour

μ = 3 cars serviced per hour

$$L = \frac{\lambda}{\mu - \lambda} = \frac{2}{3-2} = \frac{2}{1}$$ = 2 cars in the system on the average

$$W = \frac{1}{\mu - \lambda} = \frac{1}{3-2}$$ = 1 hour that an average car spends in the system

Muffler Shop Case (con't)

$$L_q = \frac{\lambda^2}{\mu(\mu-\lambda)} = \frac{2^2}{3(3-2)} = \frac{4}{3(1)}$$ = 1.33 cars waiting in line on the average

$$W_q = \frac{\lambda}{\mu(\mu-\lambda)} = \frac{2}{3}\text{ hour}$$ = 40 minutes average waiting time per car (time in queue)

$$\rho = \frac{\lambda}{\mu} = \frac{2}{3} = 0.67$$ = percentage of time mechanic is busy

$$P_0 = 1 - \frac{\lambda}{\mu} = 1 - \frac{2}{3} = 0.33$$ = probability that there are 0 cars in the system

Muffler Shop Case (con’t)

- Probability of more than *k* cars in the system

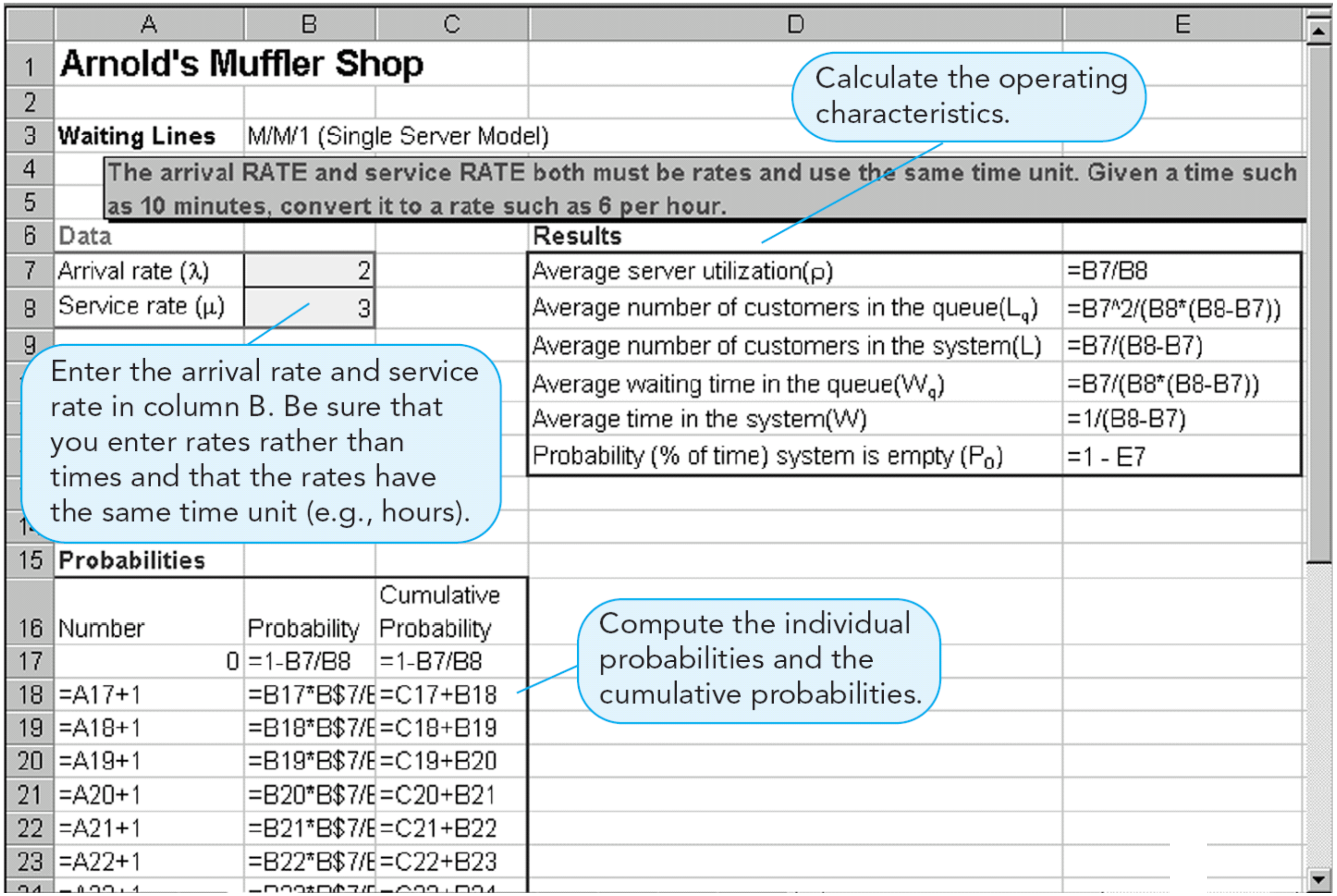
k	$P_{n>k} = (^2/_3)^{k+1}$
0	0.667
1	0.444
2	0.296
3	0.198
4	0.132
5	0.088
6	0.058
7	0.039

Note that this is equal to 1 – P_0 = 1 – 0.33 = 0.667

Implies that there is a 19.8% chance that more than 3 cars are in the system

Muffler Shop Case (con’t)

- Excel/QM model:



	A	B	C	D	E
1	**Arnold's Muffler Shop**				
2					
3	**Waiting Lines**	M/M/1 (Single Server Model)			
4		**The arrival RATE and service RATE both must be rates and use the same time unit. Given a time such**			
5		**as 10 minutes, convert it to a rate such as 6 per hour.**			
6	**Data**			**Results**	
7	Arrival rate (λ)	2		Average server utilization(ρ)	=B7/B8
8	Service rate (μ)	3		Average number of customers in the queue(L_q)	=B7^2/(B8*(B8-B7))
9				Average number of customers in the system(L)	=B7/(B8-B7)
				Average waiting time in the queue(W_q)	=B7/(B8*(B8-B7))
				Average time in the system(W)	=1/(B8-B7)
				Probability (% of time) system is empty (P_0)	=1 - E7
15	**Probabilities**				
16	Number	Probability	Cumulative Probability		
17	0	=1-B7/B8	=1-B7/B8		
18	=A17+1	=B17*B$7/E	=C17+B18		
19	=A18+1	=B18*B$7/E	=C18+B19		
20	=A19+1	=B19*B$7/E	=C19+B20		
21	=A20+1	=B20*B$7/E	=C20+B21		
22	=A21+1	=B21*B$7/E	=C21+B22		
23	=A22+1	=B22*B$7/E	=C22+B23		
24	=A23+1	=B23*B$7/E	=C23+B24		

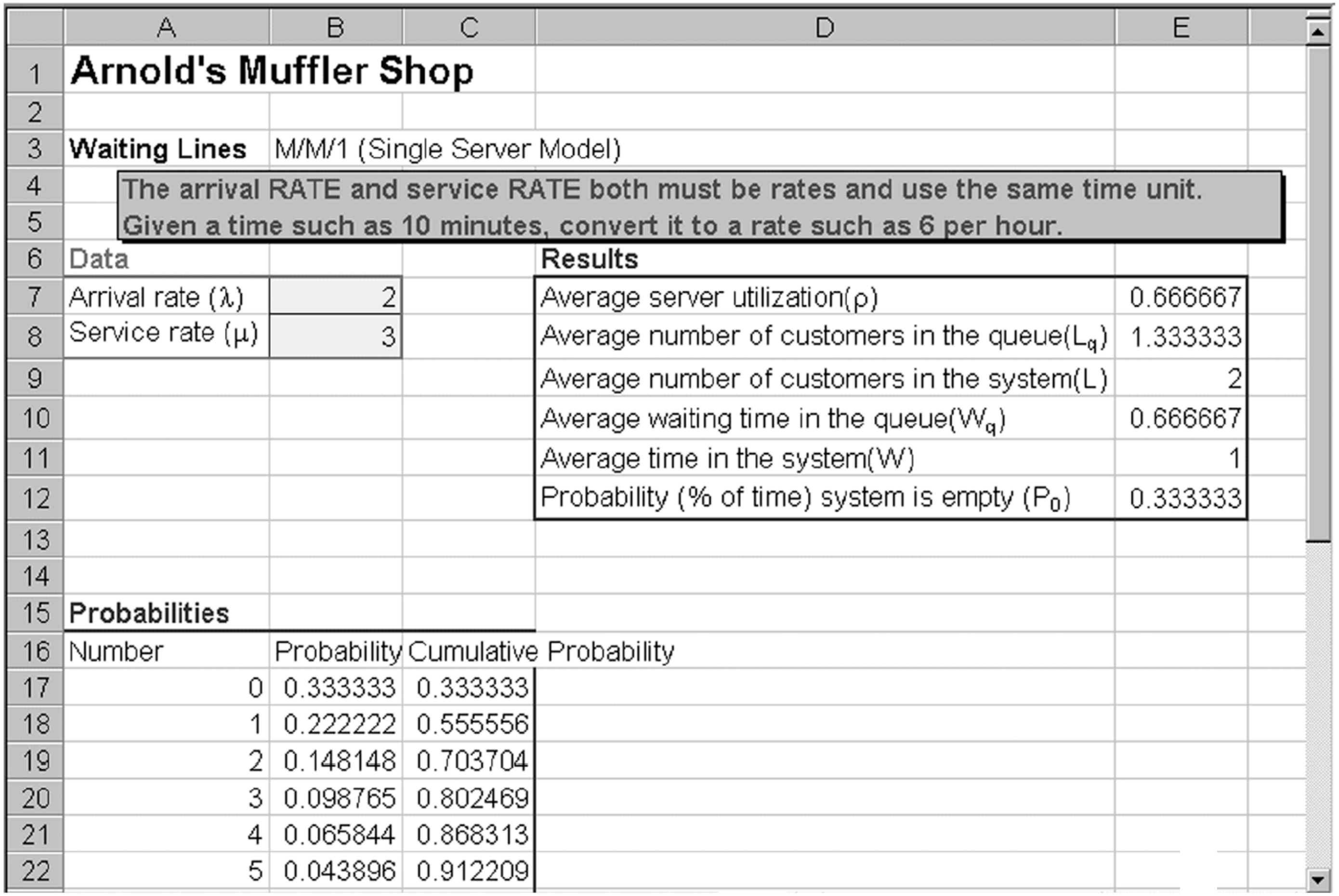
Calculate the operating characteristics.

Enter the arrival rate and service rate in column B. Be sure that you enter rates rather than times and that the rates have the same time unit (e.g., hours).

Compute the individual probabilities and the cumulative probabilities.

Muffler Shop Case (con’t)

- Output from Excel QM analysis



	A	B	C	D	E
1	**Arnold's Muffler Shop**				
2					
3	**Waiting Lines**	M/M/1 (Single Server Model)			
4		The arrival RATE and service RATE both must be rates and use the same time unit.			
5		Given a time such as 10 minutes, convert it to a rate such as 6 per hour.			
6	Data			**Results**	
7	Arrival rate (λ)	2		Average server utilization(ρ)	0.666667
8	Service rate (μ)	3		Average number of customers in the queue(L_q)	1.333333
9				Average number of customers in the system(L)	2
10				Average waiting time in the queue(W_q)	0.666667
11				Average time in the system(W)	1
12				Probability (% of time) system is empty (P_0)	0.333333
13					
14					
15	**Probabilities**				
16	Number	Probability	Cumulative Probability		
17	0	0.333333	0.333333		
18	1	0.222222	0.555556		
19	2	0.148148	0.703704		
20	3	0.098765	0.802469		
21	4	0.065844	0.868313		
22	5	0.043896	0.912209		

Muffler Shop Case (con’t)

- Introducing costs into the model
  - We want to do an economic analysis of the queuing system and determine the waiting cost and service cost
  - The total service cost is

Total service cost = (Number of channels) x (Cost per channel)

Total service cost = mC_s

where

m = number of channels
C_s = service cost of each channel

Muffler Shop Case (con’t)

- Waiting cost when the cost is based on time in the system

Total waiting cost = (Total time spent waiting by all arrivals) x (Cost of waiting)

= (Number of arrivals) x (Average wait per arrival)C_w

Total waiting cost = $(\lambda W)C_w$

- If waiting time cost is based on time in the queue

Total waiting cost = $(\lambda W_q)C_w$

Muffler Shop Case (con’t)

- So the total cost of the queuing system when based on time in the system is

Total cost = Total service cost + Total waiting cost

Total cost = $mC_s + \lambda W C_w$

- And when based on time in the queue

Total cost = $mC_s + \lambda W_q C_w$

Muffler Shop Case (con't)

- We estimate the cost of customer *waiting* time in line is $10 per hour

Total daily waiting cost = (8 hours per day)$\lambda W_q C_w$

= (8)(2)($^2/_3$)($10) = $106.67

- The mechanics wage is $7 per hour as the *service* cost

Total daily service cost = (8 hours per day)mC_s

= (8)(1)($7) = $56

- So the total cost of the system is

Total daily cost of the queuing system = $106.67 + $56 = $162.67

Muffler Shop Case (con’t)

- Let’s think about hiring a more skilled mechanic who can install mufflers at a faster rate
- The new operating characteristics would be

λ = 2 cars arriving per hour

μ = 4 cars serviced per hour (instead of 3)

$$L = \frac{\lambda}{\mu - \lambda} = \frac{2}{4-2} = \frac{2}{2}$$ = 1 car in the system on the average

$$W = \frac{1}{\mu - \lambda} = \frac{1}{4-2}$$ = 1/2 hour that an average car spends in the system

Muffler Shop Case (con't)

$$L_q = \frac{\lambda^2}{\mu(\mu-\lambda)} = \frac{2^2}{4(4-2)} = \frac{4}{8(1)}$$ = 1/2 cars waiting in line on the average

$$W_q = \frac{\lambda}{\mu(\mu-\lambda)} = \frac{1}{4}\text{ hour}$$ = 15 minutes average waiting time per car (in queue)

$$\rho = \frac{\lambda}{\mu} = \frac{2}{4} = 0.5$$ = percentage of time mechanic is busy

$$P_0 = 1 - \frac{\lambda}{\mu} = 1 - \frac{2}{4} = 0.5$$ = probability that there are 0 cars in the system

Muffler Shop Case (con’t)

- Probability of more than *k* cars in the system

k	$P_{n>k} = ({}^{2}/_{4})^{k+1}$
0	0.500
1	0.250
2	0.125
3	0.062
4	0.031
5	0.016
6	0.008
7	0.004

Muffler Shop Case (con't)

- The customer *waiting* cost is the same $10 per hour

Total daily waiting cost $= (8 \text{ hours per day})\lambda W_q C_w$

$= (8)(2)(^1/_4)(\$10) = \$40.00$

- The more skilled mechanic is more expensive at $9 per hour

Total daily service cost $= (8 \text{ hours per day})mC_s$

$= (8)(1)(\$9) = \$72$

- So the total cost of the system is

Total daily cost of the queuing system $= \$40 + \$72 = \$112$

Muffler Shop Case (con’t)

- The total time spent waiting for the 16 customers per day was formerly

  (16 cars per day) x ($^2/_3$ hour per car) = 10.67 hours

- It is now

  (16 cars per day) x ($^1/_4$ hour per car) = 4 hours

- The total system costs are less with the more skilled mechanic resulting in a $50 per day savings

  $162 – $112 = $50

- Are there other ways we can reduce cost ?

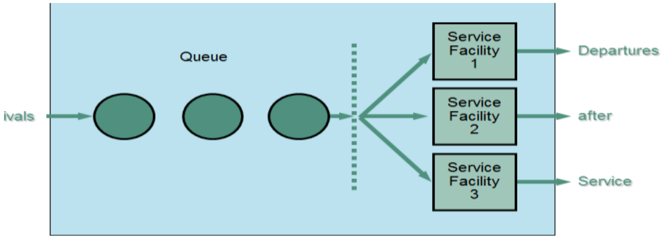
- Do not look ahead !

Reducing cost:

- Have more service areas (channels)
- Hire a cheaper mechanic
- Reduce customer wait cost
  - Find some way to entertain customers while they wait
  - Drive customer back home (or to work) instead of then waiting at the shop
  - Pick up customer’s car from his home/work

Multichannel Model, Poisson Arrivals, Exponential Service Times (M/M/m)

- Assumptions of the model
  - Arrivals are served on a FIFO basis
  - No balking or reneging
  - Arrivals are independent of each other but rate is constant over time
  - Arrivals follow a Poisson distribution
  - Service times are variable and independent but the average is known
  - Service times follow a negative exponential distribution
  - Average service rate is greater than the average arrival rate



M/M/m (con’t)

- Equations for the multichannel queuing model
- We let

m = number of channels open

λ = average arrival rate

μ = average service rate at each channel

1. The probability that there are zero customers in the system

$$P_0 = \frac{1}{\left[\sum_{n=0}^{n=m-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n\right] + \frac{1}{m!}\left(\frac{\lambda}{\mu}\right)^m \frac{m\mu}{m\mu - \lambda}} \text{ for } m\mu > \lambda$$

M/M/m (con’t)

2. The average number of customer in the system

$$L = \frac{\lambda\mu(\lambda/\mu)^m}{(m-1)!(m\mu-\lambda)^2}P_0 + \frac{\lambda}{\mu}$$

3. The average time a unit spends in the waiting line or being served, in the system

$$W = \frac{\mu(\lambda/\mu)^m}{(m-1)!(m\mu-\lambda)^2}P_0 + \frac{1}{\mu} = \frac{L}{\lambda}$$

M/M/m (con't)

4. The average number of customers or units in line waiting for service

$$L_q = L - \frac{\lambda}{\mu}$$

5. The average time a customer or unit spends in the queue waiting for service

$$W_q = W - \frac{1}{\mu} = \frac{L_q}{\lambda}$$

6. Utilization rate

$$\rho = \frac{\lambda}{m\mu}$$

Muffler Shop Revisited

- We want to investigate building a second garage bay
- We would hire a second worker who works at the same rate as the first worker (3/hr)
- The customer arrival rate remains the same (2/hr)

$$P_0 = \frac{1}{\left[\sum_{n=0}^{n=m-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n\right] + \frac{1}{m!}\left(\frac{\lambda}{\mu}\right)^m \frac{m\mu}{m\mu-\lambda}} \text{ for } m\mu > \lambda$$

$$P_0 = 0.5$$

$=$ probability of 0 cars in the system

Muffler Shop Revisited (con't)

- Average number of cars in the system

$$L = \frac{\lambda\mu(\lambda/\mu)^m}{(m-1)!(m\mu-\lambda)^2}P_0 + \frac{\lambda}{\mu} = 0.75$$

- Average time a car spends in the system

$$W = \frac{L}{\lambda} = \frac{3}{8}\text{ hours} = 22\,{}^1\!/_2\text{ minutes}$$

Muffler Shop Revisited (con’t)

- Average number of cars in the queue

$$L_q = L - \frac{\lambda}{\mu} = \frac{3}{4} - \frac{2}{3} = \frac{1}{12} = 0.083$$

- Average time a car spends in the queue

$$W_q = W - \frac{1}{\mu} = \frac{L_q}{\lambda} = \frac{0.083}{2} = 0.0415 \text{ hour} = 2\,{}^{1}\!/_{2} \text{ minutes}$$

Muffler Shop Revisited (con’t)

- Effect of service level on operating characteristics

	LEVEL OF SERVICE		
OPERATING CHARACTERISTIC	ONE MECHANIC $\mu = 3$	TWO MECHANICS $\mu = 3$ FOR BOTH	ONE FAST MECHANIC $\mu = 4$
Probability that the system is empty (P_0)	0.33	0.50	0.50
Average number of cars in the system (L)	2 cars	0.75 cars	1 car
Average time spent in the system (W)	60 minutes	22.5 minutes	30 minutes
Average number of cars in the queue (L_q)	1.33 cars	0.083 car	0.50 car
Average time spent in the queue (W_q)	40 minutes	2.5 minutes	15 minutes

Muffler Shop Revisited (con’t)

- Adding the second service bay reduces the waiting time in line but will increase the service cost as a second mechanic needs to be hired

  Total daily waiting cost = (8 hours per day)$\lambda W_q C_w$
  = (8)(2)(0.0415)($10) = $6.64

  Total daily service cost = (8 hours per day)mC_s
  = (8)(2)($7) = $112

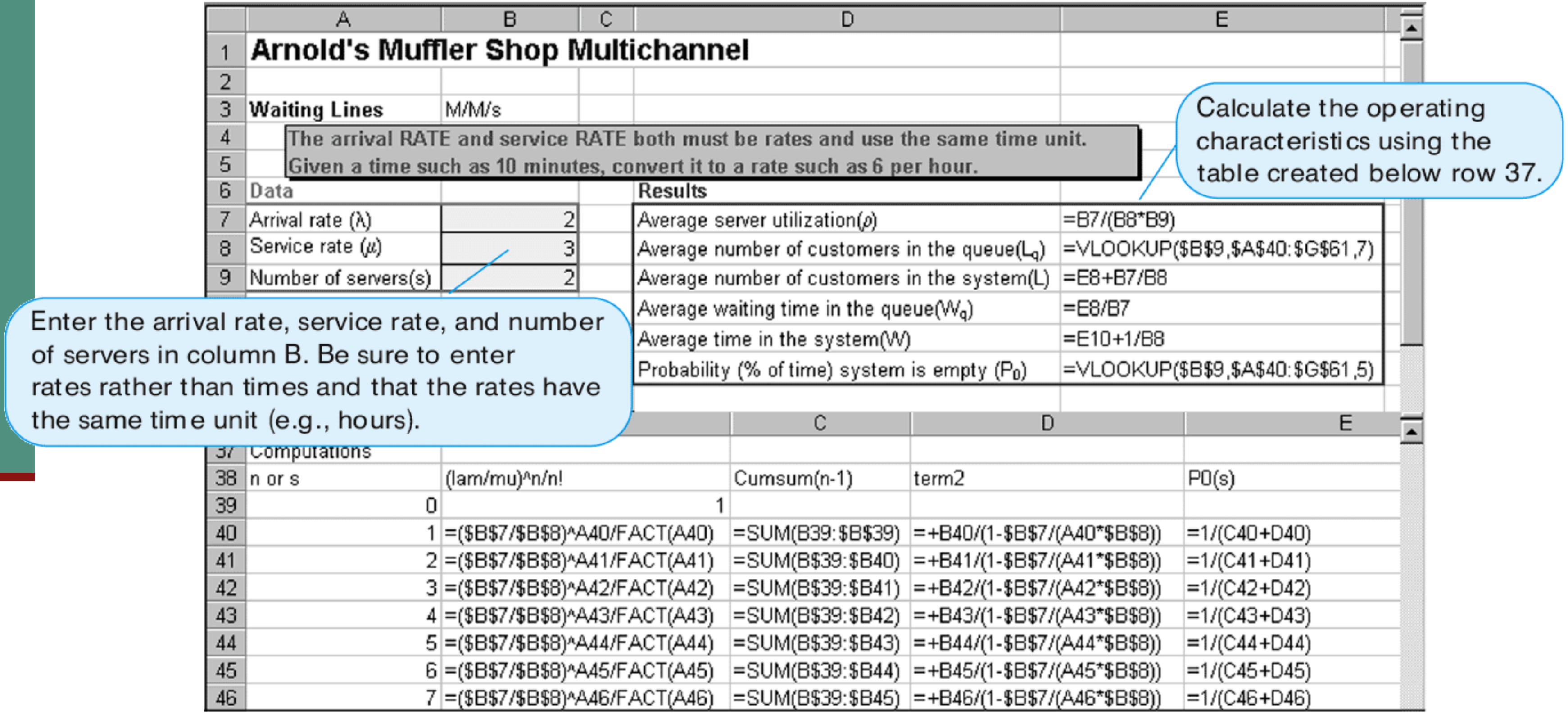
- So the total cost of the system is

  Total system cost = $6.64 + $112 = $118.64

- The more skilled (faster) mechanic is the cheapest option

Muffler Shop Revisited (con't)

- Excel/QM model for multiple servers:



	A	B	C	D	E
1	**Arnold's Muffler Shop Multichannel**				
2					
3	**Waiting Lines**	M/M/s			
4	The arrival RATE and service RATE both must be rates and use the same time unit.				
5	Given a time such as 10 minutes, convert it to a rate such as 6 per hour.				
6	Data			**Results**	
7	Arrival rate (λ)	2		Average server utilization(ρ)	=B7/(B8*B9)
8	Service rate (μ)	3		Average number of customers in the queue(L_q)	=VLOOKUP($B$9,$A$40:$G$61,7)
9	Number of servers(s)	2		Average number of customers in the system(L)	=E8+B7/B8
				Average waiting time in the queue(W_q)	=E8/B7
				Average time in the system(W)	=E10+1/B8
				Probability (% of time) system is empty (P_0)	=VLOOKUP($B$9,$A$40:$G$61,5)

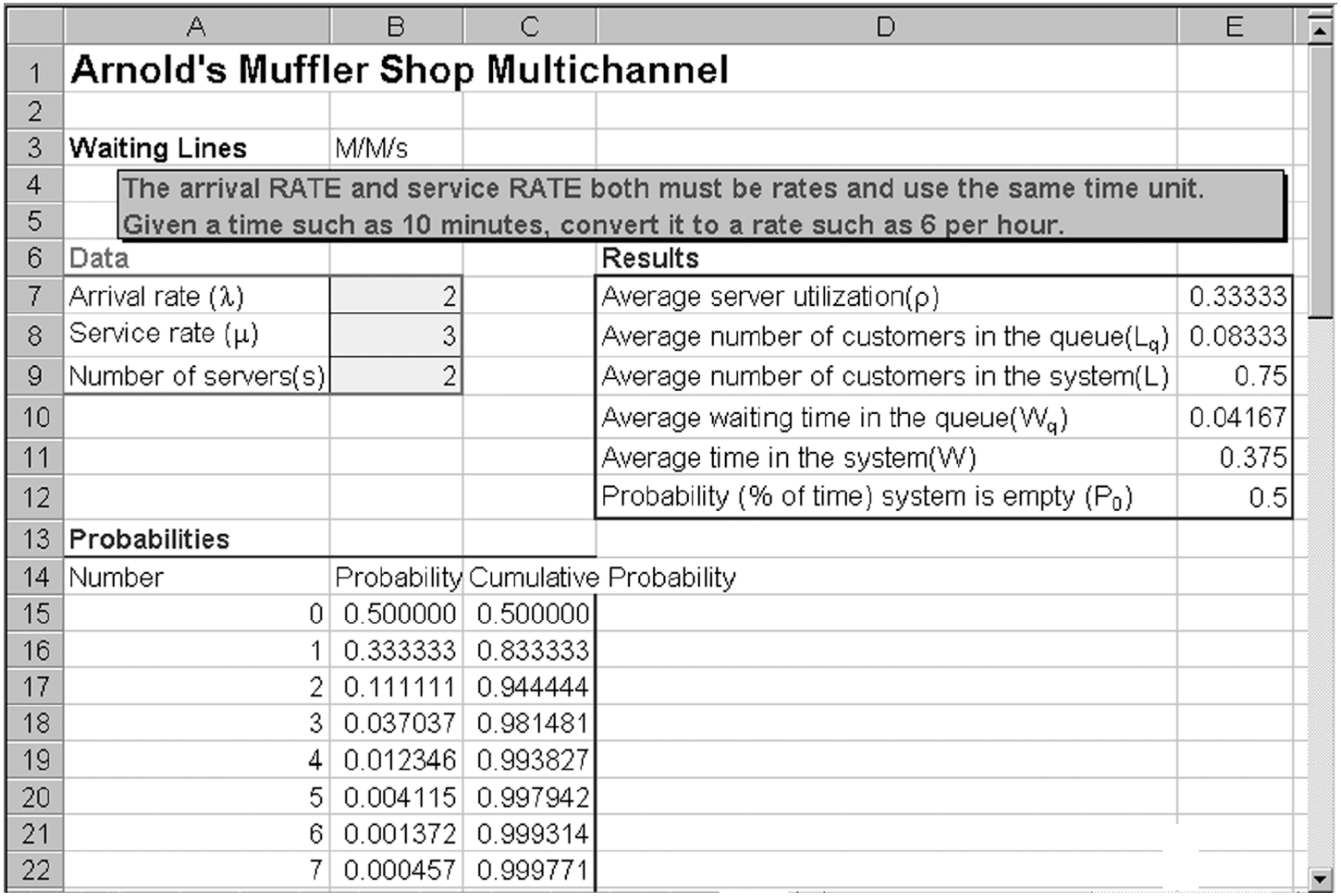
	A	B	C	D	E
37	Computations				
38	n or s	(lam/mu)^n/n!	Cumsum(n-1)	term2	P0(s)
39	0	1			
40	1	=($B$7/$B$8)^A40/FACT(A40)	=SUM(B39:$B$39)	=+B40/(1-$B$7/(A40*$B$8))	=1/(C40+D40)
41	2	=($B$7/$B$8)^A41/FACT(A41)	=SUM(B$39:$B40)	=+B41/(1-$B$7/(A41*$B$8))	=1/(C41+D41)
42	3	=($B$7/$B$8)^A42/FACT(A42)	=SUM(B$39:$B41)	=+B42/(1-$B$7/(A42*$B$8))	=1/(C42+D42)
43	4	=($B$7/$B$8)^A43/FACT(A43)	=SUM(B$39:$B42)	=+B43/(1-$B$7/(A43*$B$8))	=1/(C43+D43)
44	5	=($B$7/$B$8)^A44/FACT(A44)	=SUM(B$39:$B43)	=+B44/(1-$B$7/(A44*$B$8))	=1/(C44+D44)
45	6	=($B$7/$B$8)^A45/FACT(A45)	=SUM(B$39:$B44)	=+B45/(1-$B$7/(A45*$B$8))	=1/(C45+D45)
46	7	=($B$7/$B$8)^A46/FACT(A46)	=SUM(B$39:$B45)	=+B46/(1-$B$7/(A46*$B$8))	=1/(C46+D46)

Calculate the operating characteristics using the table created below row 37.

Enter the arrival rate, service rate, and number of servers in column B. Be sure to enter rates rather than times and that the rates have the same time unit (e.g., hours).

Muffler Shop Revisited (con’t)

- Output from Excel QM analysis



	A	B	C	D	E
1	**Arnold's Muffler Shop Multichannel**				
2					
3	**Waiting Lines**	M/M/s			
4		The arrival RATE and service RATE both must be rates and use the same time unit.			
5		Given a time such as 10 minutes, convert it to a rate such as 6 per hour.			
6	Data			**Results**	
7	Arrival rate (λ)	2		Average server utilization(ρ)	0.33333
8	Service rate (μ)	3		Average number of customers in the queue(L_q)	0.08333
9	Number of servers(s)	2		Average number of customers in the system(L)	0.75
10				Average waiting time in the queue(W_q)	0.04167
11				Average time in the system(W)	0.375
12				Probability (% of time) system is empty (P_0)	0.5
13	**Probabilities**				
14	Number	Probability	Cumulative Probability		
15	0	0.500000	0.500000		
16	1	0.333333	0.833333		
17	2	0.111111	0.944444		
18	3	0.037037	0.981481		
19	4	0.012346	0.993827		
20	5	0.004115	0.997942		
21	6	0.001372	0.999314		
22	7	0.000457	0.999771		

Constant Service Time Model (M/D/1)

- Constant service times are used when customers or units are processed according to a fixed cycle
- The values for L_q, W_q, L, and W are always less than they would be for models with variable service time
- In fact both average queue length and average waiting time are ***halved*** in constant service rate models

M/D/1 (con’t)

1. Average length of the queue

$$L_q = \frac{\lambda^2}{2\mu(\mu-\lambda)}$$

2. Average waiting time in the queue

$$W_q = \frac{\lambda}{2\mu(\mu-\lambda)}$$

M/D/1 (con't)

3. Average number of customers in the system

$$L = L_q + \frac{\lambda}{\mu}$$

4. Average time in the system

$$W = W_q + \frac{1}{\mu}$$

Recycling Company

- A company collects and compacts aluminum cans and glass bottles
- Trucks arrive at an average rate of 8 per hour (Poisson distribution)
- Truck drivers currently wait about 15 minutes before they empty their load
- Drivers and trucks cost $60 per hour
- A new automated machine can process truckloads at a constant rate of 12 per hour
- New compactor will be amortized at $3 per trip

M/D/1 (con't)

- Analysis of cost versus benefit of the purchase

Current waiting cost/trip = ($^1/_4$ hour waiting time)($60/hour cost)
= $15/trip

New system: λ = 8 trucks/hour arriving
μ = 12 trucks/hour served

Average waiting
time in queue = $W_q = {}^1/_{12}$ hour

Waiting cost/trip
with new compactor = ($^1/_{12}$ hour wait)($60/hour cost) = $5/trip

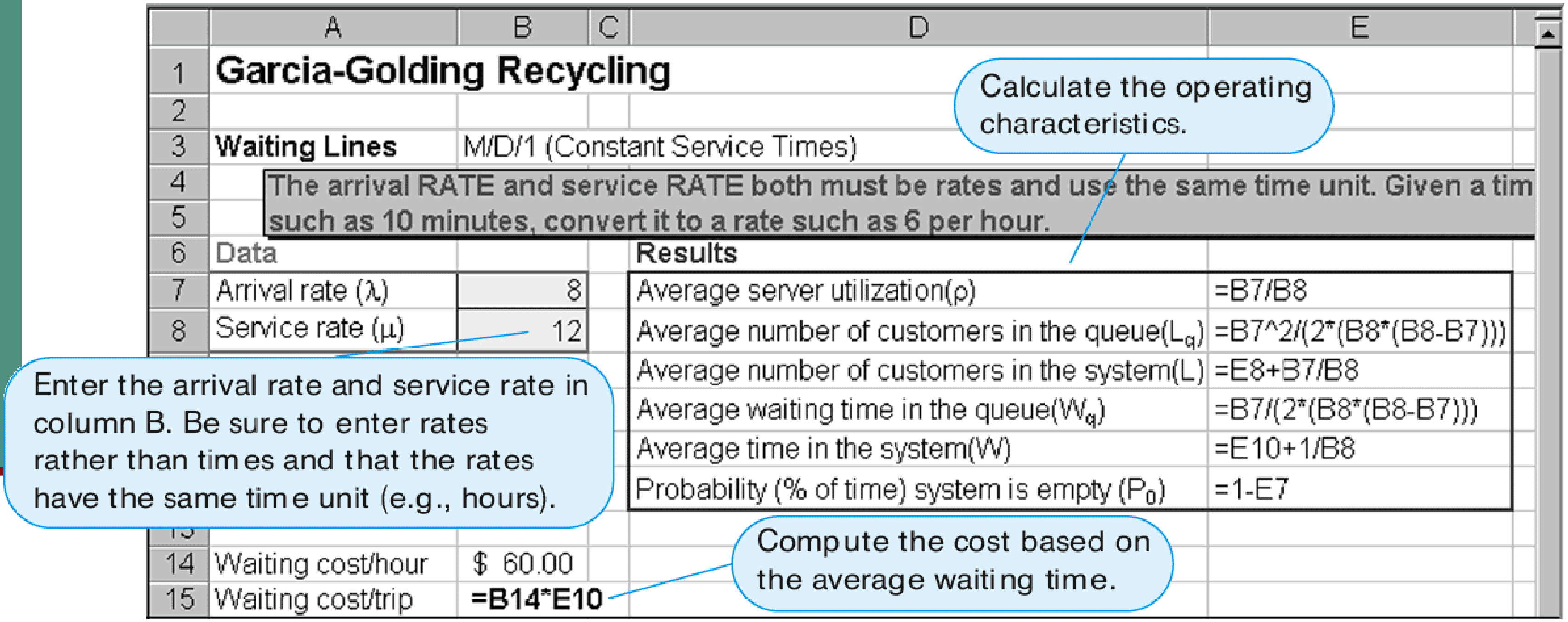
Savings with
new equipment = $15 (current system) – $5 (new system)
= $10 per trip

Cost of new equipment
amortized = $3/trip

Net savings = $7/trip

M/D/1 (con’t)

- Excel Model:



	A	B	C	D	E
1	**Garcia-Golding Recycling**				
2					
3	**Waiting Lines**	M/D/1 (Constant Service Times)			
4	**The arrival RATE and service RATE both must be rates and use the same time unit. Given a tim**				
5	**such as 10 minutes, convert it to a rate such as 6 per hour.**				
6	Data			**Results**	
7	Arrival rate (λ)	8		Average server utilization(ρ)	=B7/B8
8	Service rate (μ)	12		Average number of customers in the queue(L_q)	=B7^2/(2*(B8*(B8-B7)))
				Average number of customers in the system(L)	=E8+B7/B8
				Average waiting time in the queue(W_q)	=B7/(2*(B8*(B8-B7)))
				Average time in the system(W)	=E10+1/B8
				Probability (% of time) system is empty (P_0)	=1-E7
14	Waiting cost/hour	$ 60.00			
15	Waiting cost/trip	**=B14*E10**			

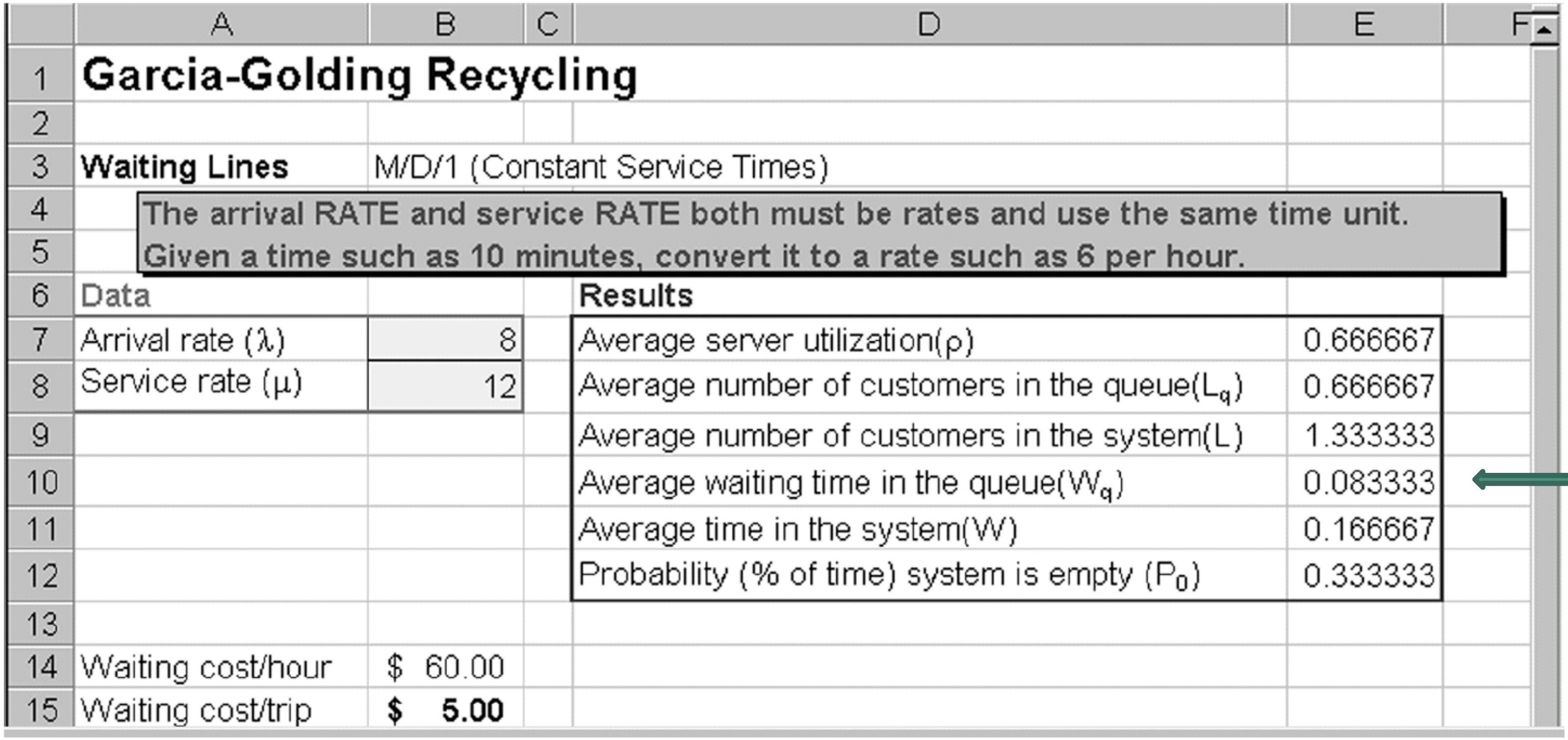
Calculate the operating characteristics.

Enter the arrival rate and service rate in column B. Be sure to enter rates rather than times and that the rates have the same time unit (e.g., hours).

Compute the cost based on the average waiting time.

M/D/1 (con’t)

- Output from Excel QM constant service time model



	A	B	C	D	E	F
1	**Garcia-Golding Recycling**					
2						
3	**Waiting Lines**	M/D/1 (Constant Service Times)				
4	**The arrival RATE and service RATE both must be rates and use the same time unit.**					
5	**Given a time such as 10 minutes, convert it to a rate such as 6 per hour.**					
6	Data			**Results**		
7	Arrival rate (λ)	8		Average server utilization(ρ)	0.666667	
8	Service rate (μ)	12		Average number of customers in the queue(L_q)	0.666667	
9				Average number of customers in the system(L)	1.333333	
10				Average waiting time in the queue(W_q)	0.083333	
11				Average time in the system(W)	0.166667	
12				Probability (% of time) system is empty (P_0)	0.333333	
13						
14	Waiting cost/hour	$ 60.00				
15	Waiting cost/trip	**$ 5.00**				

Finite Population Model (M/M/1 with Finite Source)

- When the population of potential customers is limited, the models are different
- There is now a dependent relationship between the length of the queue and the arrival rate
- The model has the following assumptions
  1. There is only one server
  2. The population of units seeking service is finite
  3. Arrivals follow a Poisson distribution and service times are exponentially distributed
  4. Customers are served on a first-come, first-served basis

M/M/1 with Finite Source (con't)

- Equations for the finite population model
  - Using

    λ = mean arrival rate, μ = mean service rate, *N* = size of the population
  - The operating characteristics are

1. Probability that the system is empty

$$P_0 = \frac{1}{\sum_{n=0}^{N} \frac{N!}{(N-n)!}\left(\frac{\lambda}{\mu}\right)^n}$$

M/M/1 with Finite Source (con't)

2. Average length of the queue

$$L_q = N - \left(\frac{\lambda + \mu}{\lambda}\right)(1 - P_0)$$

3. Average number of customers (units) in the system

$$L = L_q + (1 - P_0)$$

4. Average waiting time in the queue

$$W_q = \frac{L_q}{(N - L)\lambda}$$

M/M/1 with Finite Source (con't)

5. Average time in the system

$$W = W_q + \frac{1}{\mu}$$

6. Probability of n units in the system

$$P_n = \frac{N!}{(N-n)!}\left(\frac{\lambda}{\mu}\right)^n P_0 \text{ for } n = 0,1,\ldots,N$$

Department of Commerce Example

- The Department of Commerce has five printers that each need repair after about 20 hours of work
- Breakdowns follow a Poisson distribution
- The technician can service a printer in an average of about 2 hours, following an exponential distribution

$$\lambda = {}^{1}/_{20} = 0.05 \text{ printer/hour}$$

$$\mu = {}^{1}/_{2} = 0.50 \text{ printer/hour}$$

Department of Commerce Example (con’t)

1.

$$P_0 = \frac{1}{\sum_{n=0}^{5} \frac{5!}{(5-n)!}\left(\frac{0.05}{0.5}\right)^n} = 0.564$$

2.

$$L_q = 5 - \left(\frac{0.05+0.5}{0.05}\right)\left(1 - P_0\right) = 0.2 \text{ printer}$$

3.

$$L = 0.2 + \left(1 - 0.564\right) = 0.64 \text{ printer}$$

Average # of printers in repair

Department of Commerce Example (con't)

4.
$$W_q = \frac{0.2}{(5-0.64)(0.05)} = \frac{0.2}{0.22} = 0.91\text{hour}$$

5.
$$W = 0.91 + \frac{1}{0.50} = 2.91\text{hours}$$

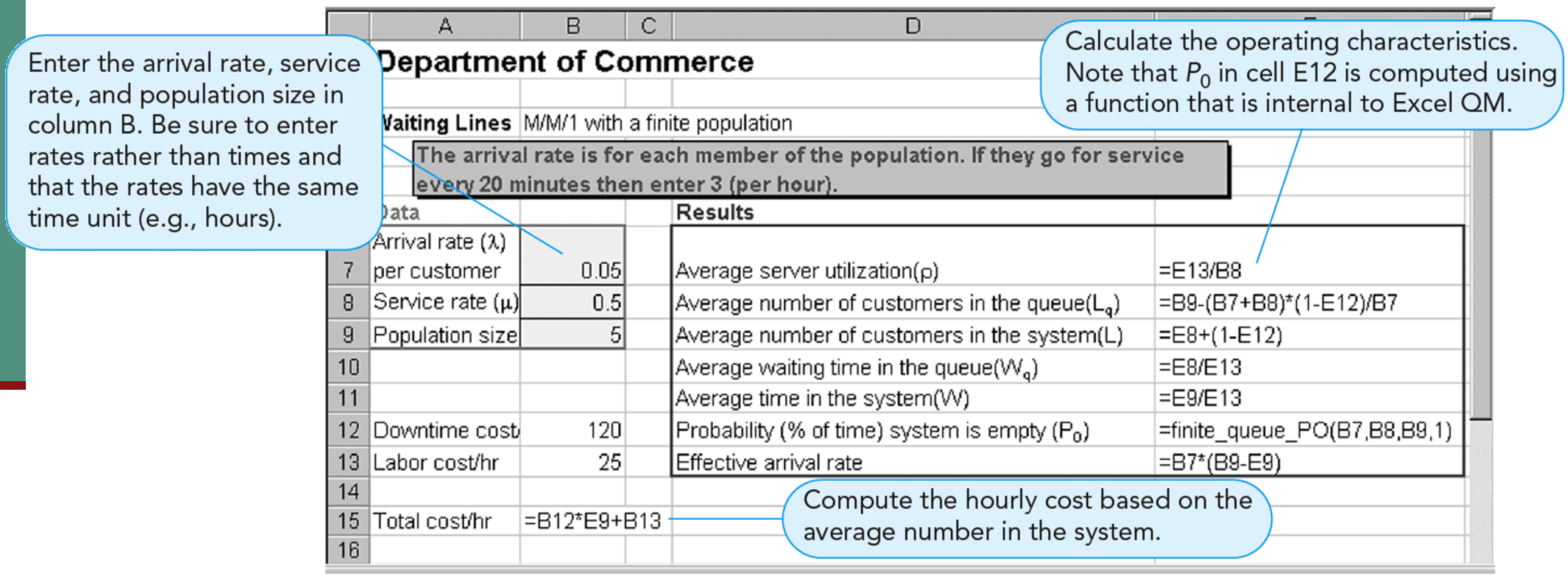
- If printer downtime costs $120 per hour and the technician is paid $25 per hour, the total cost is

Total hourly cost	=	(Average number of printers down) (Cost per downtime hour) + Cost per technician hour

= (0.64)($120) + $25 = $101.80

Department of Commerce Example (con’t)

- Excel model:



Enter the arrival rate, service rate, and population size in column B. Be sure to enter rates rather than times and that the rates have the same time unit (e.g., hours).

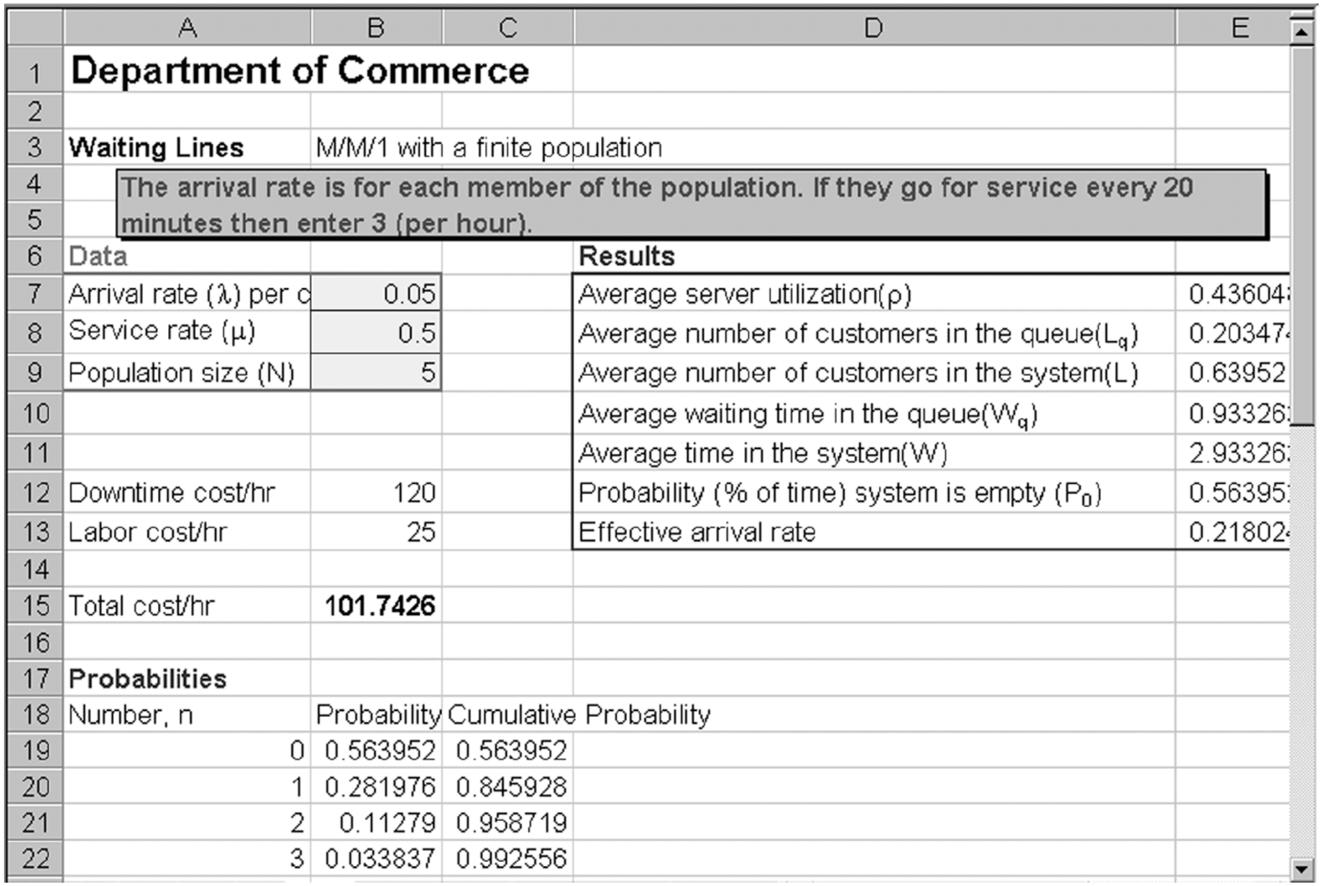
Calculate the operating characteristics. Note that P_0 in cell E12 is computed using a function that is internal to Excel QM.

	A	B	C	D	E
	Department of Commerce				
	Waiting Lines	M/M/1 with a finite population			
	The arrival rate is for each member of the population. If they go for service every 20 minutes then enter 3 (per hour).				
	Data			**Results**	
	Arrival rate (λ)				
7	per customer	0.05		Average server utilization(ρ)	=E13/B8
8	Service rate (μ)	0.5		Average number of customers in the queue(L_q)	=B9-(B7+B8)*(1-E12)/B7
9	Population size	5		Average number of customers in the system(L)	=E8+(1-E12)
10				Average waiting time in the queue(W_q)	=E8/E13
11				Average time in the system(W)	=E9/E13
12	Downtime cost	120		Probability (% of time) system is empty (P_0)	=finite_queue_PO(B7,B8,B9,1)
13	Labor cost/hr	25		Effective arrival rate	=B7*(B9-E9)
14					
15	Total cost/hr	=B12*E9+B13			
16					

Compute the hourly cost based on the average number in the system.

Department of Commerce Example (con’t)

- Output from Excel QM finite population queuing model



	A	B	C	D	E
1	**Department of Commerce**				
2					
3	**Waiting Lines**	M/M/1 with a finite population			
4		The arrival rate is for each member of the population. If they go for service every 20			
5		minutes then enter 3 (per hour).			
6	Data			**Results**	
7	Arrival rate (λ) per c	0.05		Average server utilization(ρ)	0.43604
8	Service rate (μ)	0.5		Average number of customers in the queue(L_q)	0.20347
9	Population size (N)	5		Average number of customers in the system(L)	0.63952
10				Average waiting time in the queue(W_q)	0.93326
11				Average time in the system(W)	2.93326
12	Downtime cost/hr	120		Probability (% of time) system is empty (P_0)	0.56395
13	Labor cost/hr	25		Effective arrival rate	0.21802
14					
15	Total cost/hr	**101.7426**			
16					
17	**Probabilities**				
18	Number, n	Probability	Cumulative Probability		
19	0	0.563952	0.563952		
20	1	0.281976	0.845928		
21	2	0.11279	0.958719		
22	3	0.033837	0.992556		

Periods with Different Characteristics

- For some types of business, the key characteristics in terms of average arrival rates are constant during the period in question, such as the business day
- And for other types of businesses the characteristics change during the business period
- For example in a bank, generally the periods from 9am to 11am, 11am to 1pm, 1pm to 4pm, and 4pm to 6pm all have different average arrival rates and must be optimized separately by varying the number of tellers on duty in each period has one average arrival time

More Complex Queuing Models and the Use of Simulation

- In the real world there are often *variations* from basic queuing models
- *Computer simulation* can be used to solve these more complex problems
- Simulation allows the analysis of controllable factors
- Simulation should be used when standard queuing models provide only a poor approximation of the actual service system

References

- Applied Probability and Queues by Søren Asmussen (2010)
- Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling by William J. Stewart (Jul 6, 2009)
- Queues, Inventories and Maintenance: The Analysis of Operational Systems with Variable Demand and Supply (Dover Phoenix Editions) by Philip McCord Morse
- Stochastic Modeling and Optimization: With Applications in Queues, Finance, and Supply Chains (Springer series in operations research) by David D. Yao, Hanqin Zhang, and Xun Yu Zhou

Homework

- Textbook Chapter 12
- Quiz on these slides and Chapters 12 next session
- Discussion Questions to be answered: 1, 3, 4 from Chapter 12
- Project 11 →

Project 11

- USA Internet Bank operates a banking service to US customers over the Internet
- During their peak operating hours they have 20,000 customers per hour needing to access the system for transactions (customer access follows a Poisson distribution)
- Their computer servers can each handle 6,000 transactions per hour, and the service follows an exponential distribution
- They do not want to exceed an average customer wait time of one second
- How many servers should they use ?