

A Reference Manual of Road Transport Management

January 9, 2022

Contents

1	Transportation System	1
1.1	Introduction	1
1.2	History of Transportation	1
1.3	Transportation Sector and Modes of Transportation	4
1.3.1	Transportation Sector	4
1.3.2	Modes of Transportation	4
1.4	Factors influencing Transport Operations	6
1.4.1	Driver Characteristics	6
1.4.2	Vehicle Characteristics	10
1.4.3	Pedestrian Characteristics	20
1.4.4	Road Characteristics	20
1.5	Review Problems	21
2	Motor Vehicle Act	23
3	Transportation Economics	25
3.1	Introduction	25
3.2	Scope of Transportation Economics	27
3.3	Transportation Demand	27
3.3.1	Causes of Shift in Transportation Demand Curve	29
3.3.2	Transportation Demand Functions	29
3.4	Transportation Supply	31
3.4.1	Causes of Shift in Transportation Supply Curve	32
3.5	Equilibrium of Transportation Demand and Supply	33
3.6	Sensitivity of Travel Demand	34
3.6.1	Elasticity and Revenue Along a Linear Demand Curve	35
3.7	Kraft Demand Function	36
3.8	Factors Affecting Elasticity	38
3.8.1	Income Elasticity	38
3.8.2	Price Elasticity	39
3.9	Elasticity and Total Revenue	39
3.10	Application of Price Elasticity: Consumer Surplus and Latent Demand	40

3.11	Cost Analysis in the Evaluation of Transportation Systems . . .	43
3.11.1	Economic Laws Related to Cost	43
3.11.2	Average Cost	44
3.11.3	Marginal Costs	45
3.11.4	Costs and Production	47
3.11.5	Cost Elasticity	47
3.12	Transportation Projects Evaluation	48
3.13	Cost-Benefit Analysis	49
3.14	<u>Numerical Examples:</u>	52
3.15	Review Problems	55
4	Insurance	57
4.1	Introduction	57
4.2	Basic Characteristics of Insurance	57
4.2.1	Pooling of Losses	57
4.2.2	Payment of Fortuitous Losses	59
4.2.3	Risk Transfer	59
4.2.4	Indemnification	60
4.3	Characteristics of an Ideally Insurable Risk	60
4.3.1	Large Number of Exposure Units	60
4.3.2	Accidental and Unintentional Loss	60
4.3.3	Determinable and Measurable Loss	61
4.3.4	No Catastrophic Loss	62
4.3.5	Calculable Chance of Loss	62
4.3.6	Economically Feasible Premium	63
4.4	Types of Insurance	63
4.4.1	Private Insurance	64
4.4.2	Government Insurance	64
4.5	Readings from Insurance Regulation, 2049 (1993)	65
4.5.1	Definitions	65
4.5.2	Categories of Insurance Business	65
4.5.3	Provisions Relating to a Surveyor	67
4.5.4	Provisions relating to payment against Insurance Claim	69
4.6	Review Problems	70
5	Passenger Transport Operation	71
5.1	Public Transportation	71
5.1.1	Transit Modes	71
5.1.2	Transit Capacity and Level of Service	71
5.1.3	The Role and Future of Public Transportation	73
5.2	Number of Passengers Versus Number of Served Vehicles	74
5.2.1	A Hypothetical Case Study	74
5.3	Passenger Flows in Public Transportation	76
5.4	Passenger Flows Along a Transit Line	77

5.5	Service Frequency and Headways	79
5.5.1	The Maximum Service Frequency	80
5.5.2	Passenger Waiting Time	81
5.5.3	Headway Determination by "Square Root Formula"	82
5.5.4	Headway Determination by Maximum Load Method	84
5.6	Time Table	85
5.7	Transit Line Capacity	87
5.7.1	Transit Line Capacity Utilization	89
5.8	Route Development	91
5.8.1	Properties of a Good Route Set	91
5.9	Stop Location and Stopping Policy	92
5.9.1	Stopping Policy	92
5.9.2	Stop Location	94
5.9.3	Properties of a Good Schedule	95
5.10	Optimal Paths in Transportation Networks	96
5.10.1	Dijkstra's Shortest Path Algorithm	96
5.11	Review Problems	100
6	Freight Transport Operation	103
6.1	Main Actors of Freight Transport and Distribution	103
6.1.1	Shippers	103
6.1.2	Carriers	104
6.1.3	Intermediaries	104
6.2	Modes of Transportation	104
6.2.1	Road	105
6.2.2	Rail	106
6.2.3	Air	108
6.2.4	Sea	108
6.2.5	Intermodal Transportation	109
6.3	Choice of Carrier and Transportation Mode	112
6.4	Shipment Options	112
6.4.1	Direct or Customized Shipments	112
6.4.2	Consolidated Shipments	113
6.5	Distribution Structures	113
6.5.1	Single-echelon	114
6.5.2	Multi-echelon	115
6.6	Transportation Model	116
6.6.1	Definition of the Transportation Model	116
6.6.2	The Transportation Algorithm	118
6.6.3	Determining the Basic Feasible Solution	119
6.7	Assignment Model	130
6.7.1	Definition of the Assignment Problem	130
6.7.2	Hungarian Method	131
6.8	Travelling Salesman Problem	137

6.9	Review Problems	145
6.9.1	Answers	146
7	Fuel Consumption and Emission Models	149
8	Planning for New Transport Organization	151

Chapter 1

Transportation System

1.1 Introduction

Transportation is a crucial aspect of a nation's development and prosperity. However, with the advent of modern means of transportation, managing its operation has become more complex, so there is ample opportunity for engineers to devise new operational and management techniques for the smooth control of various modes of transportation of a country.

1.2 History of Transportation

Early Boats

The first mode of transportation was created in the effort to traverse water: boats. Those who colonized Australia roughly 60,000–40,000 years ago have been credited as the first people to cross the sea, though there is some evidence that seafaring trips were carried out as far back as 900,000 years ago.

The earliest known boats were simple logboats, also referred to as dugouts, which were made by hollowing out a tree trunk. Evidence for these floating vehicles comes from artifacts that date back to around 10,000–7,000 years ago. The Pesse canoe—a logboat—is the oldest boat unearthed and dates as far back as 7600 BCE. Rafts have been around nearly as long, with artifacts showing them in use for at least 8,000 years.

Horses and Wheeled Vehicles

Next, came horses. While it's difficult to pinpoint exactly when humans first began domesticating them as a means of getting around and transporting goods, experts generally go by the emergence of certain human biological and cultural markers that indicate when such practices started to take place.

Based on changes in teeth records, butchering activities, shifts in settlement patterns, and historic depictions, experts believe that domestication took place around 4000 BCE. Genetic evidence from horses, including changes in musculature and cognitive function, support this.

It was also roughly around this period that the wheel was invented. Archaeological records show that the first wheeled vehicles were in use around 3500 BCE, with evidence of the existence of such contraptions found in Mesopotamia, the Northern Caucasus, and Central Europe. The earliest well-dated artifact from that time period is the "Bronocice pot," a ceramic vase that depicts a four-wheeled wagon that featured two axles. It was unearthed in southern Poland.

Steam Engines

In 1769, the Watt steam engine changed everything. Boats were among the first to take advantage of steam-generated power; in 1783, a French inventor by the name of Claude de Jouffroy built the "Pyroscaphe," the world's first steamship. But despite successfully making trips up and down the river and carrying passengers as part of a demonstration, there wasn't enough interest to fund further development.

While other inventors tried to make steamships that were practical enough for mass transport, it was American Robert Fulton who furthered the technology to where it was commercially viable. In 1807, the Clermont completed a 150-mile trip from New York City to Albany that took 32 hours, with the average speed clocking in at about five miles per hour. Within a few years, Fulton and company would offer regular passenger and freight service between New Orleans, Louisiana, and Natchez, Mississippi.

Back in 1769, another Frenchman named Nicolas Joseph Cugnot attempted to adapt steam engine technology to a road vehicle—the result was the invention of the first automobile. However, the heavy engine added so much weight to the vehicle that it wasn't practical. It had a top speed of 2.5 miles per hour.

Another effort to repurpose the steam engine for a different means of personal transport resulted in the "Roper Steam Velocipede." Developed in 1867, the two-wheeled steam-powered bicycle is considered by many historians to be the world's first motorcycle.

Locomotives

One mode of land transport powered by a steam engine that did go mainstream was the locomotive. In 1801, British inventor Richard Trevithick unveiled the world's first road locomotive—called the "Puffing Devil"—and

used it to give six passengers a ride to a nearby village. It was three years later that Trevithick first demonstrated a locomotive that ran on rails, and another one that hauled 10 tons of iron to the community of Penydarren, Wales, to a small village called Abercynon.

It took a fellow Brit—a civil and mechanical engineer named George Stephenson—to turn locomotives into a form of mass transport. In 1812, Matthew Murray of Holbeck designed and built the first commercially successful steam locomotive, “The Salamanca,” and Stephenson wanted to take the technology a step further. So in 1814, Stephenson designed the “Blücher,” an eight-wagon locomotive capable of hauling 30 tons of coal uphill at a speed of four miles per hour.

By 1824, Stephenson improved the efficiency of his locomotive designs to where he was commissioned by the Stockton and Darlington Railway to build the first steam locomotive to carry passengers on a public rail line, the aptly named “Locomotion No. 1.” Six years later, he opened the Liverpool and Manchester Railway, the first public inter-city railway line serviced by steam locomotives. His notable accomplishments also include establishing the standard for rail spacing for most of the railways in use today. No wonder he’s been hailed as “Father of Railways.”

Submarines

Technically speaking, the first navigable submarine was invented in 1620 by Dutchman Cornelis Drebbel. Built for the English Royal Navy, Drebbel’s submarine could stay submerged for up to three hours and was propelled by oars. However, the submarine was never used in combat, and it wasn’t until the turn of the 20th century that designs leading to practical and widely used submersible vehicles were realized.

Along the way, there were important milestones such as the launch of the hand-powered, egg-shaped “Turtle” in 1776, the first military submarine used in combat. There was also the French Navy submarine “Plongeur,” the first mechanically powered submarine.

Finally, in 1888, the Spanish Navy launched the “Peral,” the first electric, battery-powered submarine, which also so happened to be the first fully capable military submarine. Built by a Spanish engineer and sailor named Isaac Peral, it was equipped with a torpedo tube, two torpedoes, an air regeneration system, and the first fully reliable underwater navigation system, and it posted an underwater speed of 3.5 miles per hour.

Aircraft

The start of the twentieth century was truly the dawn of a new era in the history of transportation as two American brothers, Orville and Wilbur Wright, pulled off the first official powered flight in 1903. In essence, they invented the world's first airplane. Transport via aircraft took off from there with airplanes being put into service within a few short years during World War I. In 1919, British aviators John Alcock and Arthur Brown completed the first transatlantic flight, crossing from Canada to Ireland. The same year, passengers were able to fly internationally for the first time.

Around the same time that the Wright brothers were taking flight, French inventor Paul Cornu started developing a rotorcraft. And on November 13, 1907, his "Cornu" helicopter, made of little more than some tubing, an engine, and rotary wings, achieved a lift height of about one foot while staying airborne for about 20 seconds. With that, Cornu would lay claim to having piloted the first helicopter flight.

1.3 Transportation Sector and Modes of Transportation

1.3.1 Transportation Sector

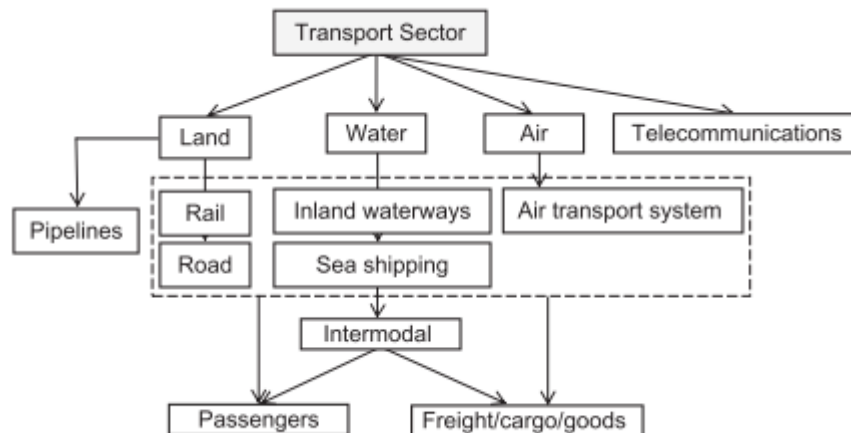
Transportation sector performs transportation services in order to satisfy demand for mobility of people and transport of freight shipments. A range of socioeconomic activities in a society, as well as land uses induce transportation demand. The demand is represented by the number of passengers and/or volumes of cargo to be transported between given origins and destinations during a given (specified) period of time. The supply component in the transportation sector consists of transport services provided by different transport modes and their particular systems. Airports, highways, streets, and ports should be able to meet transportation demand and offer acceptable level-of-service to the users. The transport supply component contributes to the economy of a region, country, and continent it serves. The demand and supply component are in permanent interaction.

1.3.2 Modes of Transportation

The transportation modes constituting the supply component are generally classified according to the way of performing their operations of transporting people and freight shipments. In general, the basic land-based transport modes include road, rail, and pipeline. The water-based mode includes inland waterways and sea shipping. The air transport is the air-based mode.

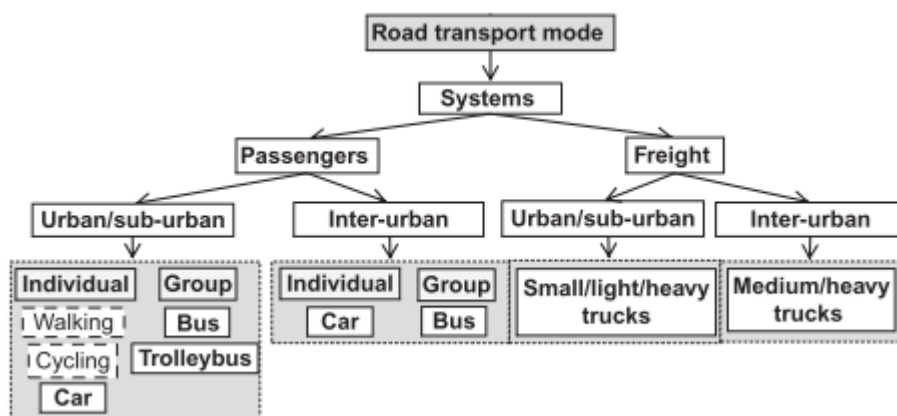
1.3. TRANSPORTATION SECTOR AND MODES OF TRANSPORTATION⁵

The specific mode is intermodal transport consisting of combinations of particular basic modes and their systems. In addition, the mode not carrying out physical entities but just information is telecommunications. Fig. shows a simplified scheme of the structure of transport sector, its modes and their systems.



Transport modes generally consist of two types of systems: the first is intended for serving passenger and the other for serving freight/cargo/goods demand. Each of these consists of subsystems, which will be called systems. The classification of systems within each mode is carried out at three levels: (i) type of the system (passengers, freight), spatial scale of operation (urban/suburban/regional, interurban), and carrier type (individual, group).

The Fig. below shows classification of systems of the road transport mode.



1.4 Factors influencing Transport Operations

There are mainly four components of road transportation: the driver, the pedestrian, the vehicle, and the road, which influence road transport management's design and operational considerations. To provide efficient and safe road transportation, a knowledge of the characteristics and limitations of each of these components is essential. It is also crucial to be aware of the interrelationships among these components to determine the effects they have on each other. Their characteristics are also of primary importance when traffic engineering measures such as traffic control devices are used in the highway mode.

1.4.1 Driver Characteristics

One problem that faces traffic and transportation engineers when they consider driver characteristics in the course of design is the varying skills and perceptual abilities of drivers on the highway, demonstrated by a wide range of capabilities to hear, see, evaluate, and react to information. Studies support that these abilities vary under different conditions, such as the influence of alcohol, fatigue, and time of the day. Therefore, it is essential that criteria used for design purposes be compatible with the capabilities and limitations of most drivers on the highway. The use of an average value, such as mean reaction time, may not be adequate for a large number of drivers. The 85th percentile and the 95th percentile have been used to select design criteria; in general, the higher the chosen percentile, the wider the range covered.

The Human Response Process

Actions taken by the drivers while driving result from their evaluation and reaction to the information they obtain from their certain stimuli. However, assessment and response must be carried out quickly, as the information received while driving the highways is dynamic. Furthermore, it has been suggested that most of the information received by a driver is visual, implying that the ability to see is of fundamental importance in the driving task. Therefore, highway and traffic engineers must have some rudimentary knowledge of visual perception and hearing perception.

Visual Reception : The principal characteristics of the eye are visual acuity, peripheral vision, color vision, glare vision and recovery, and depth perception.

1. Visual Acuity : It is the ability to see fine details of an object. It can be represented by the visual angle, which is the reciprocal of the smallest pattern detail in minutes of arc that can be resolved and given as:

$$\phi = 2 \arctan \left(\frac{L}{2D} \right) \quad (1.1)$$

Where,

L = diameter of the target (letter or symbol)

D = distance from the eye to target in the same units as L

Mathematical Derivation,



Using trigonometry,

$$\tan \left(\frac{X}{2} \right) = \frac{\frac{L}{2}}{D}$$

$$\tan \left(\frac{X}{2} \right) = \frac{L}{2D}$$

$$X = 2 \arctan \left(\frac{L}{2D} \right)$$

The ability to resolve a pattern detail with a visual acuity of one minute of arc (1/60 of a degree) is considered as normal vision of acuity (20/20 vision).

**20/20 vision is a term used to express normal visual acuity measured at a distance of 20 feet. If you have 20/20 vision, you can see clearly at 20 ft what should normally be seen at that distance. If you have 20/100 vision, it means you must be as close as 20ft to see what a person with normal vision can see at 100ft.*

2. Peripheral Vision : It is the ability of people to see objects beyond the cone of clearest vision. Although objects can be seen within this zone, details and color are not clear. The cone for peripheral vision could be one subtending up to 160 degrees; this value is affected by the speed of the vehicle. Age also influences peripheral vision. For instance, at about

age 60, a significant change occurs in a person's peripheral vision.

3. Color Vision : It is the ability to differentiate one color from another, but deficiency in this ability, usually referred to as color blindness, is not of great significance in highway driving because other ways of recognizing traffic information devices (e.g., shape) can compensate for it. Combinations of black and white and black and yellow have been shown to be those to which the eye is most sensitive.

4. Glare Vision and Recovery : There are two types of glare vision: direct and specular. Rowland and others have indicated that direct glare occurs when relatively bright light appears in the individual's field of vision and specular glare occurs when the image reflected by the relatively bright light appears in the field of vision. Both types of glare result in a decrease of visibility and cause discomfort to the eyes. It is also known that age has a significant effect on the sensitivity to glare, and that at about age 40, a significant change occurs in a person's sensitivity to glare.

The time required by a person to recover from the effects of glare after passing the light source is known as glare recovery. Studies have shown that this time is about 3 seconds when moving from dark to light and can be 6 seconds or more when moving from light to dark. Glare vision is of great importance during night driving; it contributes to the problem of serving older people, who see much more poorly at night. This phenomenon should be taken into account in the design and location of street lighting so that glare effects are reduced to a minimum.

Glare effects can be minimized by reducing luminaire brightness and by increasing the background brightness in a driver's field of view. Specific actions taken to achieve this in lighting design include using higher mounting heights, positioning lighting supports farther away from the highway, and restricting the light from the luminaire to obtain minimum interference with the visibility of the driver.

5. Depth Perception : It affects the ability of a person to estimate speed and distance. It is particularly important on two-lane highways during passing maneuvers, when head-on crashes may result from a lack of proper judgment of speed and distance.

The ability of the human eye to differentiate between objects is fundamental to this phenomenon. It should be noted, however, that the human eye is not very good at estimating absolute values of speed, distance, size, and acceleration. This is why traffic control devices are standard in size,

shape, and color. Standardization not only aids in distance estimation but also helps the color-blind driver to identify signs.

Perception-Reaction Process

The process through which a driver, cyclist, or pedestrian evaluates and reacts to a stimulus can be divided into four subprocesses:

1. Perception: the driver sees a control device, warning sign, or object on the road
2. Identification: the driver identifies the object or control device and thus understands the stimulus
3. Emotion: the driver decides what action to take in response to the stimulus; for example, to step on the brake pedal, to pass, to swerve, or to change lanes
4. Reaction or Volition: the driver actually executes the action decided on during the emotion sub-process

Time elapses during each of these subprocesses. The time that elapses from the start of perception to the end of reaction is the total time required for perception, identification, emotion, and volition, sometimes referred to as PIEV time or (more commonly) as perception-reaction time.

Perception-reaction time is an important factor in the determination of braking distances, which in turn dictates the minimum sight distance required on a highway and the length of the yellow phase at a signalized intersection. Perception-reaction time varies among individuals and may, in fact, vary for the same person as the occasion changes. These changes in perception-reaction time depend on how complicated the situation is, the existing environmental conditions, age, whether the person is tired or under the influence of drugs and/or alcohol, and whether the stimulus is expected or unexpected.

Triggs and Harris described this phenomenon in detail. They noted that the 85th-percentile time to brake, obtained from several situations, varied from 1.26 to over 3 seconds. The reaction time selected for design purposes should, however, be large enough to include reaction times for most drivers using the highways. Recommendations made by the American Association of State Highway and Transportation Officials (AASHTO) stipulate 2.5 seconds for stopping-sight distances. This encompasses the decision times for about 90 percent of drivers under most highway conditions. Note, however,

that a reaction time of 2.5 second may not be adequate for unexpected conditions or for some very complex conditions, such as those at multiphase at-grade intersections and ramp terminals. For example, when signals are unexpected, reaction times can increase by 35 percent.

Numerical Example:

A driver with a perception-reaction time of 2.5 sec is driving at 65 mi/h when she observes that an accident has blocked the road ahead. Determine the distance the vehicle would move before the driver could activate the brakes. The vehicle will continue to move at 65 mi/h during the perception-reaction time of 2.5 sec.

Solution:

- Convert mi/h to ft/sec:

$$65 \text{ mi/h} = \left(65 \times \frac{5280}{3600}\right) \text{ ft/sec} = 65 \times 1.47 = 95.55 \text{ ft/sec}$$

- Find the distance traveled:

$$\begin{aligned} D &= vt \\ &= 95.55 \times 2.5 \\ &= 238.9 \text{ ft} \end{aligned}$$

where v = velocity and t = time.

1.4.2 Vehicle Characteristics

Criteria for the geometric design of highways are partly based on the static, kinematic, and dynamic characteristics of vehicles. Static characteristics include the weight and size of the vehicle, while kinematic characteristics involve the motion of the vehicle without considering the forces that cause the motion. Dynamic characteristics involve the forces that cause the motion of the vehicle. Since nearly all highways carry both passenger-automobile and truck traffic, it is essential that design criteria take into account the characteristics of different types of vehicles. A thorough knowledge of these characteristics will aid the highway and/or traffic engineer in designing highways and traffic-control systems that allow the safe and smooth operation of a moving vehicle, particularly during the basic maneuvers of passing, stopping, and turning. Therefore, designing a highway involves the selection of a design vehicle, whose characteristics will encompass those of nearly all vehicles expected to use the highway. The characteristics of the design vehicle are then used to determine criteria for geometric design, intersection design, and sight-distance requirements.

Static characteristics

The size of the design vehicle for a highway is an important factor in the determination of design standards for several physical components of the highway. These include lane width, shoulder width, length and width of parking bays, and lengths of vertical curves. The axle weights of the vehicles expected on the highway are important when pavement depths and maximum grades are being determined.

For many years, each state prescribed by law the size and weight limits for trucks using its highways, and in some cases local authorities also imposed more severe restrictions on some roads. Table 3.1 shows some features of static characteristics for which limits were prescribed. A range of maximum allowable values is given for each feature.

<i>Type</i>	<i>Allowable Lengths (ft)</i>
Bus	35–60
Single truck	35–60
Trailer, semi/full	35–48
Semitrailer	55–85
Truck trailer	55–85
Tractor semitrailer trailer	55–85
Truck trailer trailer	65–80
Tractor semitrailer, trailer, trailer	60–105
<i>Type</i>	<i>Allowable Weights (lb)</i>
Single-axle	18,000–24,000
Tandem-axle	32,000–40,000
State maximum gross vehicle weight	73,280–164,000
Interstate maximum gross vehicle weight	73,280–164,000

SOURCE: Adapted from *State Maximum Sizes and Weights for Motor Vehicles*, Motor Vehicle Manufacturers' Association of the United States, Detroit, MI, May 1982.

Fig: Range of State Limits on Vehicle Lengths by Type and Maximum Weight of Vehicle

Since the passage of the Surface Transportation Assistance Act of 1982, the maximum allowable truck sizes and weights on Interstate and other qualifying federal-aided highways are at most:

- 80,000 lb gross weight, with axle loads of up to 20,000 lb for single axles and 34,000 lb for tandem (double) axles
- 102 in. width for all trucks
- 48 ft length for semitrailers and trailers

- 28 ft length for each twin trailer

(Note: Those states that had higher weight limits before this law was enacted are allowed to retain them for intrastate travel.)

The federal regulations also stipulate that the overall maximum gross weight for a group of two or more consecutive axles should be determined from the equation:

$$W = 500 \left[\frac{LN}{N-1} + 12N + 36 \right] \quad (1.2)$$

Where,

W = overall gross weight (calculated to the nearest 500 lb)

L = the extreme of any group of two or more consecutive axles (ft)

N = number of axles in the group under consideration

The regulations also stipulate that a gross load of 34,000 lb may be carried by two consecutive sets of tandem axles if the overall distance between the first and last axles of the consecutive sets is 36 ft or more.

Numerical Example:

A 5-axle truck traveling on an interstate highway has the following axle characteristics:

Distance between the front single axle and the first set of tandem axles = 20 ft

Distance between the first set of tandem axle and the back set of tandem axles = 48 ft

If the overall gross weight of the truck is 79,500 lb, determine whether this truck satisfies federal weight regulations.

Solution: Although the overall gross weight is less than the maximum allowable of 80,000 lb, the allowable gross weight based on the axle configuration should be checked.

Use Eq. 3.2.

$$\begin{aligned} W &= 500 \left[\frac{LN}{N-1} + 12N + 36 \right] \\ W &= 500 \left[\frac{48 \times 4}{4-1} + 12 \times 4 + 36 \right] \\ &= 74,000 \text{ lb} \end{aligned}$$

which is less than the allowable of 80,000 lb. The truck therefore satisfies the federal truck weight regulations.

Kinematic Characteristics

The primary element among kinematic characteristics is the acceleration capability of the vehicle. Acceleration capability is important in several traffic operations, such as passing maneuvers and gap acceptance. Also, the dimensioning of highway features such as freeway ramps and passing lanes is often governed by acceleration rates. Acceleration is also important in determining the forces that cause motion. Therefore, a study of the kinematic characteristics of the vehicle primarily involves a study of how acceleration rates influence the elements of motion, such as velocity and distance. We therefore review in this section the mathematical relationships among acceleration, velocity, distance, and time.

Let us consider a vehicle moving along a straight line from point o to point m, a distance x in a reference plane T. The position vector of the vehicle after time t may be expressed as:

$$r_{om} = x\hat{i} \quad (1.3)$$

Where,

r_{om} = position vector for m in T
 \hat{i} = a unit vector parallel to line om
 x = distance along the straight line

The velocity and acceleration for m may be simply expressed as

$$u_m = \dot{r}_{om} = \dot{x}\hat{i} \quad (1.4)$$

$$a_m = \ddot{r}_{om} = \ddot{x}\hat{i} \quad (1.5)$$

Where,

u_m = velocity of the vehicle at point m
 a_m = acceleration of the vehicle at point m
 $\dot{x} = \frac{dy}{dx}$
 $\ddot{x} = \frac{d^2y}{dx^2}$

Two cases are of interest: (1) acceleration is assumed constant; (2) acceleration is a function of velocity.

Case 1: Acceleration is constant When the acceleration of the vehicle is assumed to be constant,

$$\ddot{x}_i = a \quad (1.6)$$

$$\frac{d\dot{x}}{dt} = a \quad (1.7)$$

$$\dot{x} = at + C_1 \quad (1.8)$$

$$x = \frac{1}{2}at^2 + C_1t + C_2 \quad (1.9)$$

The constants C_1 and C_2 are determined either by the initial conditions on velocity and position or by using known positions of the vehicle at two different times.

Case 2: Acceleration is the function of velocity The assumption of constant acceleration has some limitations, because the accelerating capability of a vehicle at any time t is related to the speed of the vehicle at that time (u_t). The lower the speed, the higher the acceleration rate that can be obtained. One model that is used commonly in this case is

$$\frac{du_t}{dt} = \alpha - \beta u_t \quad (1.10)$$

Where α and β are constants.

In this model, the maximum acceleration rate that can be achieved is theoretically α , which means that α has units of acceleration as its unit. The term βu_t also should have units of acceleration as its unit, which means that β has the inverse of time (for example, sec^{-1}) as its unit.

Integrating Equation 1.10 gives,

$$-\frac{1}{\beta} \ln(\alpha - \beta u_t) = t + C$$

If the velocity is u_o at $t = 0$,

$$\begin{aligned} C &= -\frac{1}{\beta} \ln(\alpha - \beta u_o) \\ -\frac{1}{\beta} \ln(\alpha - \beta u_t) &= t - \frac{1}{\beta} \ln(\alpha - \beta u_o) \\ \ln \frac{(\alpha - \beta u_o)}{(\alpha - \beta u_t)} &= \beta t \end{aligned}$$

$$\therefore t = \frac{1}{\beta} \ln \frac{(\alpha - \beta u_o)}{(\alpha - \beta u_t)} \quad (1.11)$$

$$\alpha - \beta u_t = (\alpha - \beta u_o)e^{-\beta t}$$

$$u_t = \frac{\alpha}{\beta}(1 - e^{-\beta t}) + u_o e^{-\beta t} \quad (1.12)$$

The distance $x(t)$ traveled at any time t may be determined by integrating equation 1.12

$$x = \int_0^t u_t dt = \int_0^t \left[\frac{\alpha}{\beta}(1 - e^{-\beta t}) + u_o e^{-\beta t} \right] dt = \left(\frac{\alpha}{\beta} \right) t - \frac{\alpha}{\beta^2}(1 - e^{-\beta t}) + \frac{u_o}{\beta}(1 - e^{-\beta t}) \quad (1.13)$$

Numerical Example:

The acceleration of a vehicle can be represented by the following equation.

$$\frac{du_t}{dt} = 33 - 0.04u$$

where u is the vehicle speed in ft/sec. If the vehicle is traveling at 45 mi/h, determine its velocity after 5 sec of acceleration and the distance traveled during that time.

Solution:

- Convert 45 mi/h to ft/sec:
- $45 \times 1.47 = 66.15$ ft/sec
- Use Eq. 3.11 to determine velocity u_t after time t .

$$u = \frac{\alpha}{\beta}(1 - e^{-\beta t}) + u_o e^{-\beta t}$$

$$\alpha = 33$$

$$\beta = 0.04$$

$$\begin{aligned} u_t &= \frac{33}{0.04}(1 - e^{-(0.04 \times 5)}) + 66.15e^{-(0.04 \times 5)} \\ &= 82.5(1 - 0.82) + 66.15 \times 0.82 \\ &= 14.85 + 54.24 \\ &= 69.09 \text{ ft/sec} \end{aligned}$$

- Convert ft/sec to mi/h.

$$\begin{aligned} u_t &= 69.09/1.47 \\ &= 47.00 \text{ mi/h} \end{aligned}$$

- Use Eq. 3.12 to determine distance traveled.

$$\begin{aligned} x &= \left(\frac{\alpha}{\beta} \right) t - \frac{\alpha}{\beta^2}(1 - e^{-\beta t}) + \frac{u_o}{\beta}(1 - e^{-\beta t}) \\ &= \left(\frac{33}{0.04} \right) 5 - \frac{33}{(0.04)^2}(1 - e^{-0.04 \times 5}) + \frac{66.15}{0.04}(1 - e^{-0.04 \times 5}) \\ &= 412.5 - 2062.5(1 - 0.82) + 1653.75(1 - 0.82) \\ &= 412.5 - 371.25 + 297.68 \\ &= 338.93 \text{ ft} \end{aligned}$$

Dynamic Characteristics

Several forces act on a vehicle while it is in motion: air resistance, grade resistance, rolling resistance, and curve resistance. The extents to which these forces affect the operation of the vehicle are discussed in this section.

Air Resistance A vehicle in motion has to overcome the resistance of the air in front of it as well as the force due to the frictional action of the air around it. The force required to overcome these is known as the air resistance and is related to the cross-sectional area of the vehicle in a direction perpendicular to the direction of motion and to the square of the speed of the vehicle. Claffey has shown that this force can be estimated from the formula

$$R_a = \frac{1}{2} \frac{(2.15\rho C_d A u^2)}{g} \quad (1.14)$$

Where,

R_a = air resistance force (lb)

ρ = density of air (0.0766 lb/ft³) at sea level; less at higher elevations

C_d = aerodynamic drag coefficient

(current average value for passenger cars is 0.4; for trucks, this value ranges from 0.5 to 0.8, but a typical value is 0.5)

A = frontal cross-sectional area (ft²)

u = vehicle speed (mi/h)

g = acceleration of gravity (32.2 ft /sec²)

Grade Resistance When a vehicle moves up a grade, a component of the weight of the vehicle acts downward, along the plane of the highway. This creates a force acting in a direction opposite that of the motion. This force is the grade resistance. A vehicle traveling up a grade will therefore tend to lose speed unless an accelerating force is applied. The speed achieved at any point along the grade for a given rate of acceleration will depend on the grade. Figure below shows the relationships between speed achieved and distance traveled on different grades by a typical heavy truck of 200 lb/hp during maximum acceleration. Note: grade resistance = weight * grade, in decimal.



Fig: Speed-Distance Curves for Acceleration of a Typical Heavy Truck of 120 kg/kw [200 lb/hp] on Upgrades and Downgrades

Rolling Resistance There are forces within the vehicle itself that offer resistance to motion. These forces are due mainly to frictional effect on moving parts of the vehicle, but they also include the frictional slip between the pavement surface and the tires. The sum effect of these forces on motion is known as rolling resistance. The rolling resistance depends on the speed of the vehicle and the type of pavement. Rolling forces are relatively lower on smooth pavements than on rough pavements.

The rolling resistance force for passenger cars on a smooth pavement can be determined from the relation

$$R_r = (C_{rs} + 2.15C_{rv}u^2)W \quad (1.15)$$

Where,

R_r = rolling resistance force (lb)

C_{rs} = constant (typically 0.012 for passenger cars)

C_{rv} = constant (typically $0.65 * 10^{-6}$ sec²/ft² for passenger cars)

u = vehicle speed (mi/h)

W = gross vehicle weight (lb)

For trucks, the rolling resistance can be obtained from

$$R_r = (C_a + 1.47C_b u)W \quad (1.16)$$

Where,

R_r = rolling resistance force (lb)

C_a = constant (typically 0.02445 for trucks)

C_b = constant (typically 0.00044 sec/ft for trucks)

u = vehicle speed (mi/h)

W = gross vehicle weight (lb)

The surface condition of the pavement has a significant effect on the rolling resistance. For example, at a speed of 50 mi/h on a badly broken and patched asphalt surface, the rolling resistance is 51 lb/ton of weight, whereas at the same speed on a loose sand surface, the rolling resistance is 76 lb/ton of weight.

Curve Resistance When a passenger car is maneuvered to take a curve, external forces act on the front wheels of the vehicle. These forces have components that have a retarding effect on the forward motion of the vehicle. The sum effect of these components constitutes the curve resistance. This resistance depends on the radius of the curve, the gross weight of the vehicle, and the velocity at which the vehicle is moving. It can be determined as

$$R_c = 0.5 \frac{(2.15u^2W)}{gR} \quad (1.17)$$

Where,

R_c = curve resistance (lb)

u = vehicle speed (mi/h)

W = gross vehicle weight (lb)

g = acceleration due to gravity (ft/sec^2)

R = radius of curvature (ft)

Power Requirements Power is the rate at which work is done. It is usually expressed in horsepower (a U.S. unit of measure), where 1 horsepower is 550 lb-ft /sec. The performance capability of a vehicle is measured in terms of the horsepower the engine can produce to overcome air, grade, curve, and friction resistance forces and put the vehicle in motion. Figure below shows how these forces act on the moving vehicle.



The power delivered by the engine is

$$P = \frac{1.47Ru}{550} \quad (1.18)$$

Where,

P = horsepower delivered (hp)

R = sum of resistance to motion (lb)

u = speed of vehicle (mi/h)

Numerical Example:

Determine the horsepower produced by a passenger car traveling at a speed of 65 mi/h on a straight road of 5% grade with a smooth pavement. Assume the weight of the car is 4000 lb and the cross-sectional area of the car is 40 ft².

Solution: The force produced by the car should be at least equal to the sum of the acting resistance forces.

$$R = (\text{air resistance}) + (\text{rolling resistance}) + (\text{upgrade resistance})$$

Note: There is no curve resistance since the road is straight.

- Use Eq. 3.13 to determine air resistance.

$$\begin{aligned} R_a &= 0.5 \left(\frac{2.15pC_D A u^2}{g} \right) \\ &= 0.5 \frac{2.15 \times 0.0766 \times 0.4 \times 40 \times 65 \times 65}{32.2} \\ &= 172.9 \text{ lb} \end{aligned}$$

- Use Eq. 3.14 to determine rolling resistance.

$$\begin{aligned} R_r &= (C_{rr} + 2.15C_{rv}u^2)(4000) \\ &= (0.012 + 2.15 \times 0.65 \times 10^{-6} \times 65 \times 65)(4000) \\ &= (0.012 + 0.006)(4000) \\ &= 0.018 \times 4000 \\ &= 72 \text{ lb} \end{aligned}$$

$$\text{Grade resistance} = 4000 \times \frac{5}{100} = 200 \text{ lb}$$

- Determine total resistance.

$$R = R_a + R_r + \text{grade resistance} = 172.9 + 72 + 200 = 444.9 \text{ lb}$$

- Use Eq. 3.17 to determine horsepower produced.

$$\begin{aligned} P &= \frac{1.47Ru}{550} \\ &= \frac{1.47 \times 444.9 \times 65}{550} \\ &= 77.3 \text{ hp} \end{aligned}$$

1.4.3 Pedestrian Characteristics

Pedestrian characteristics relevant to traffic and highway engineering practice include those of the driver, discussed in the preceding sections. In addition, other pedestrian characteristics may influence the design and location of pedestrian control devices. Such control devices include special pedestrian signals, safety zones and islands at intersections, pedestrian underpasses, elevated walkways, and crosswalks. Apart from visual and hearing characteristics, walking characteristics play a major part in the design of some of these controls. For example, the design of an all-red phase, which permits pedestrians to cross an intersection with heavy traffic, requires knowledge of the walking speeds of pedestrians. Observations of pedestrian movements have indicated that walking speeds vary between 3.0 and 8.0 ft /sec. Significant differences have also been observed between male and female walking speeds. At intersections, the mean male walking speed has been determined to be 4.93 ft /sec, and for females, 4.63 ft /sec. A more conservative value of 4.0 ft /sec is normally used for design purposes. However, Rouphail and others have shown that the average walking speed depends on the population of elderly pedestrians. For example, the average walking speed is 4.0 ft /sec when the percentage of elderly pedestrians is 20 percent or lower, but reduces to 3.0 ft /sec when the percentage of elderly pedestrians is higher than 20 percent. This factor therefore should be taken into consideration for the design of intersection pedestrian signals at locations where a high number of older pedestrians is expected. Consideration also should be given to the characteristics of handicapped pedestrians, such as the blind. Studies have shown that accidents involving blind pedestrians can be reduced by installing special signals. The blind pedestrian can turn the signal to a red phase by using a special key, which also rings a bell, indicating to the pedestrian that it is safe to cross. Ramps are also now being provided at intersection curbs to facilitate the crossing of the intersection by the occupant of a wheelchair. Also, consideration should be given to the relatively lower average walking speed of the handicapped pedestrian, which can vary from a low of 1.97 ft /sec to 3.66 ft /sec.

1.4.4 Road Characteristics

The characteristics of the highway discussed in this section are related to stopping and passing because these have a more direct relationship to the characteristics of the driver and the vehicle discussed earlier.

Sight Distance

Sight distance is the length of the roadway a driver can see ahead at any particular time. The sight distance available at each point of the highway must be such that, when a driver is traveling at the highway's design speed, adequate time is given after an object is observed in the vehicle's path to make the necessary evasive maneuvers without colliding with the object. The two types of sight distance are (1) stopping sight distance and (2) passing sight distance.

1. Stopping Sight Distance The stopping sight distance (SSD), for design purposes, is usually taken as the minimum sight distance required for a driver to stop a vehicle after seeing an object in the vehicle's path without hitting that object. This distance is the sum of the distance traveled during perception-reaction time and the distance traveled during braking. The SSD for a vehicle traveling at u mi/h is

$$SSD = 1.47ut + \frac{u^2}{30 \left(\frac{a}{g} \pm G \right)} \quad (1.19)$$

Where,

- u = speed of the vehicle when brakes are applied
- t = distance travelled between perception-reaction time
- g = acceleration due to gravity
- a = acceleration of the vehicle
- $G = \tan \gamma$ (% Grade/100)

2. Passing Sight Distance The passing sight distance is the minimum sight distance required on a two-lane, two-way highway that will permit a driver to complete a passing maneuver without colliding with an opposing vehicle and without cutting off the passed vehicle. The passing sight distance will also allow the driver to successfully abort the passing maneuver (that is, return to the right lane behind the vehicle being passed) if he or she so desires. In determining minimum passing sight distances for design purposes, only single passes (that is, a single vehicle passing a single vehicle) are considered. Although it is possible for multiple passing maneuvers to occur (that is, more than one vehicle pass or are passed in one maneuver), it is not practical for minimum design criteria to be based on them.

1.5 Review Problems

Chapter 2

Motor Vehicle Act

Please refer the "Motor Vehicles and Transport Management Act, 2049 (1993)" for this chapter.

Chapter 3

Transportation Economics

3.1 Introduction

All decisions related to planning, design, and improvement of transportation infrastructure have economic implications. Transport economics includes the issues such as transport location, movement of people and freight/goods, transport demand, transport planning and forecasting, direct and indirect cost of transport, pricing of transport services, investments in transport infrastructure and services, transport and social-economic development, and transport regulation. In this chapter the transport economics is considered from the micro economic perspective. We consider various aspects of the direct costs and pricing of transport infrastructure and services of different transportation modes.

Considered from this microeconomic perspective, it can be said that the transport sector consists of the demand and supply component. The transport demand is derived demand due to needs of people and freight/goods shipments to change the physical place. For example, many people live at one place but work and/or have a leisure on the others, which requires them to travel forward and backward. The location of companies providing raw materials is different than those of the users of these materials—manufacturers of the semifinal and final products, which requires transportation of these raw materials from the former to the latter. In addition, the manufacturers of the final products are often located far away from the retailers of these products, which again require transportation, this time of the final products. Thus, it can be said that the transport demand is derived demand. In many cases transportation demand is proportional to the volumes of peoples' activities and the quantities of final products they consume during a given period of time.

The transport demand is handled by the transport supply/capacity pro-

vided by transport companies. The transport companies generally provide transport infrastructure with the supportive facilities and equipment, and rolling stock/vehicles carrying out transport services. In order to make them operational, the corresponding material, labor, ie, employed staff/personnel, and energy/fuel, are consumed. In terms of time, the transport infrastructure has particularly the long life-cycle, which is, for example, about 20, 30, 40, or even 60 years. That of rolling stock/vehicles is shorter (20–25 years) mainly due to its/their physical and also technological obsolescence, after they need to be replaced. In this context, two categories of transport companies can be distinguished: that providing transport infrastructure called “the infrastructure providers,” and that providing transport services called “transport operators.” They both constitute the transport systems within particular transport modes.

According to the economic jargon, the above-mentioned components of transport supply/capacity represent the main inputs to transport processes. The outputs of the transport process are the transport services produced in the given quantities and at the specified quality. They are consumed by users—passengers and/or freight shippers/receivers at the same time as they are produced. This implies that they are the short-lived similarly as transport demand without possibility to be stored/warehoused and left to be consumed sometimes latter on.

The most important economic categories of a given transport system are its costs, revenues, and their relationship.

The costs are expenses for maintaining the transport infrastructure and carrying out transport services. These costs are passed to users of these services (passengers and freight/goods shippers and receivers) in the form of prices/charges. These charges bring revenues to transport companies. In general, these prices/charges, ie, revenues, are set up to cover the company’s costs and provide some profits. Nevertheless, the difference between revenues and costs can generally be positive or negative, thus representing the profits or losses. Profits or losses are calculated for the given period of time (usually for a quarter of year, year, or few years).

The prices charged to users of transport services represent for them direct costs. In addition, users are imposed the indirect costs, which are usually considered to be the cost of time during trip/travel/ transportation. In such case, the sum of direct and indirect costs is called the users’ generalized trip/travel/transportation costs for users.

We describe in this chapter the main elements of economics of transportation systems considered from the engineering perspective. The chapter

analyzes direct costs and revenues of the providers of transport infrastructure and transport operators. Costs and revenues could be analyzed at the level of the individual component (infrastructure, rolling stock) and/or of the entire company.

3.2 Scope of Transportation Economics

The study of economics is divided into macro-economics and micro-economics. Macro-economics is associated with the wealth of society on a regional scale, and deals with the behavior of aggregate concepts. On the other hand, micro-economics involves the behavior of relatively smaller entities such as firms and individuals. Transportation economics, while considered a branch of applied micro-economics, is associated with certain unique issues such as:

1. the demand for transportation is not direct, but is derived
2. the consumption of each transportation facility (i.e., each trip) is unique in time and space
3. technological differences among different modes and economies of scale
4. governmental interventionist policies and regulations in transportation

Transportation economics specifically addresses demand of transportation services, supply of transportation facilities, elasticity of demand and supply, price mechanisms, and transportation cost analysis

3.3 Transportation Demand

Like all other goods and services, the demand for any specific transportation facility depends on factors pertaining to the consumer such as income, and characteristics of the facility such as the cost associated with its use (in terms of time and price) relative to rival facilities. A typical example of such demand is that for auto travel: lower incomes of consumers, coupled with lower costs and travel times associated with transit are expected to lead to reduced demand for auto travel. Transportation demand analysis involves demand functions (which represents the willingness of consumers to purchase the transportation “product” at various alternative prices, i.e., the demand-price curve, and demand models (which estimate the probability that an individual (or fraction of a set of individuals) will choose a particular product over the other.



A hypothetical example of an aggregate transportation demand function is provided as shown in the above figure. This represents the amount of travel people are willing to make by transit at various transit fare (price) levels. Transportation demand functions, either in the form of a graph or an equation, are useful in transportation planning because they enable the determination of expected demand at any price. A specific demand curve represents the demand-price relationship given a set of conditions specific to the transportation product in question (referred to as alternative-specific attributes, such as travel time, comfort, convenience), and also specific to the users (income levels and other socioeconomic characteristics). Changes in such conditions often result in changes in the levels of transportation demand, even at fixed price of that product. For example, increased unemployment would likely lead to reduced demand for travel. Also, an increase in costs associated with auto use is likely to result in increased transit demand, even if transit fares remain the same. When such changes in conditions (other than price) occur, they are represented as a shift in the demand curve shown in the figure below (upward shift for increased demand, $D_1 \rightarrow D_2$; and downward shift for decreased demand, $D_1 \rightarrow D_3$).



3.3.1 Causes of Shift in Transportation Demand Curve

There is a possibility of a change in demand of a transportation facility even when its price remains same, and this is reflected as a shift in the demand curve for that transportation facility. Factors that influence such demand shifts are discussed below:

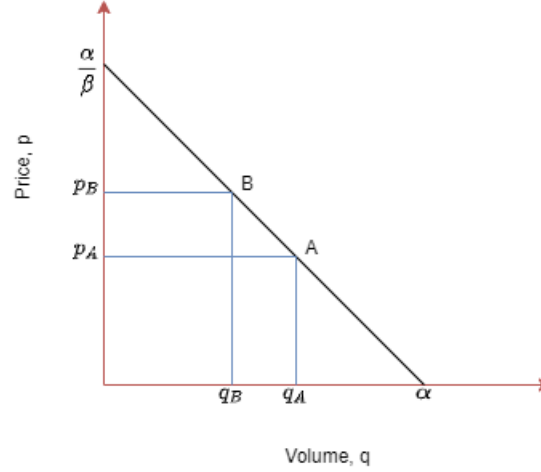
- ***Sudden change in customer preference:*** (Season, life, style, e.t.c)
For example, more people seem to ride transit in the winter season.
- ***Change in the level of attribute of interest (e.g., increase in price) of related goods:*** For complementary products, a decrease in price of product increases the demand for other product, shifting the latter's demand curve to the right (e.g., parking spaces, automobile use). For rival products, an increase in price of a product increases the demand for its rival product, shifting the latter's demand curve to the right (e.g., transit, auto).
- ***Change in regional income:*** An increase in income shifts the demand curve to the right. A normal good is one whose demand increases as a person's income increases.
- ***Change in number of potential customers:*** An increase in market size shifts the demand curve towards right.
- ***Expectation of impending change in the level of the attribute of interest:*** e.g., A news report predicting higher prices in the future can cause a shift in the demand curve at the current price as customers purchase increases in the anticipation of the price change.

3.3.2 Transportation Demand Functions

Transportation demand models are used to determine the volume of travel demanded, at various levels of service and have been described as “a representation of human behavior which can be used to predict how individuals or firms will change transportation choices in response to changes in future conditions”. Within the context of transportation economics, a trip maker is defined as a consumer, in the economics meaning of the word, as the trip maker, by planning a trip, seeks to consume the service offered by transportation facilities. There are two types of demand functions:

1. ***Disaggregate demand functions:*** These predict the behavior of a single consumer in response to changes in future conditions.
2. ***Aggregate demand functions:*** These predict the behavior of a group of consumers such as a household, in response to changes in future conditions.

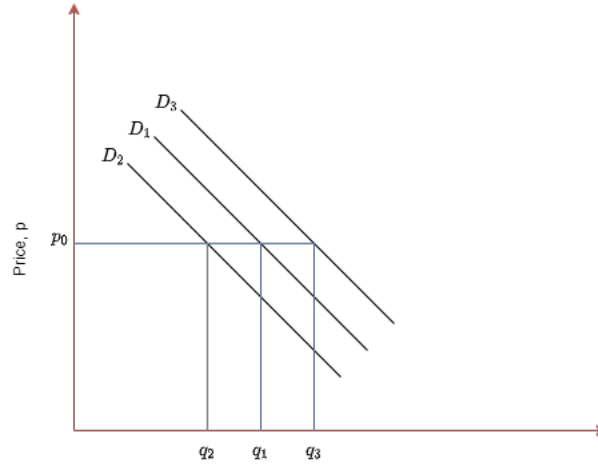
Linear Demand Function



A linear demand function for travel as shown in the above figure for a given pair of origin and destination points, at a specific time of day and for a particular purpose. Such a demand function is useful for predicting travel over a wide range of conditions. This demand function is useful for predicting travel over a wide range of conditions. This demand function assumes a particular level and distribution of income, population, and socioeconomic characteristics. Note that it is an aggregate demand curve, representing the volume of trips demanded at different prices by a group of travelers. Functionally,

$$q = \alpha - \beta p \quad (3.1)$$

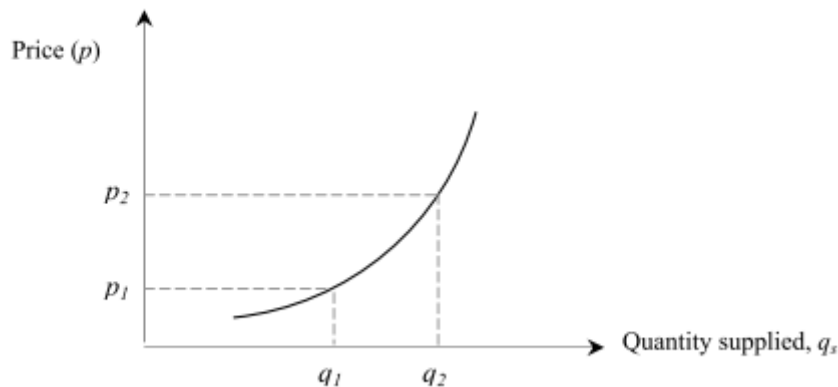
Where, q is the quantity of trips demanded, p is the price, and α and β are constant demand parameters. The demand function is drawn with a negative slope expressing a similar situation where a decrease in perceived price usually results in an increase in travel, although it is not always true.



The above figure shows a series of shifted demand curves, representing changes in the quantity of travel due to variables other than the perceived price. Naturally, at a price p_0 , one could expect different quantities q_1 , q_2 , and q_3 as the demand curve changes from D_1 , D_2 , and D_3 . If the curve moves upward (from D_1 to D_3), it indicates an increase in trips.

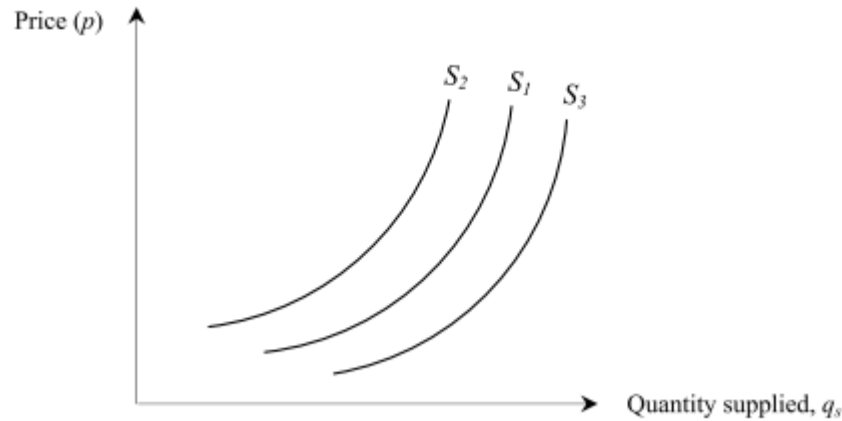
It is important to distinguish short-run changes in the quantity of travel due to price changes represented by movement along a demand curve from long-run changes due to activity or behavioral variables, represented by shifts in demand functions.

3.4 Transportation Supply



The supply of a transportation product represents the quantity of that product a producer is willing to offer at a given price. However, transportation supply may also be associated with the quality of the product. At a given time, transportation facility supply depends on price: higher prices are an incentive for producers to make more profits who therefore increase supply levels. A hypothetical example of a transportation supply function is provided as shown in the figure above. This represents the amount of transportation products that suppliers are willing to make available at various prices. Transportation supply functions are useful in transportation planning because they enable the determination of expected supply at any future price. A specific supply curve represents the supply-price relationship given a set of conditions specific to the transportation product in question (referred to as alternative-specific attributes, such as travel time, comfort, convenience), and also specific to the producers (such as technology, policy and governmental intervention through policies and regulation. Changes in such conditions often result in changes in the levels of transportation supply, even at fixed price of that product. For example, improved increased

unemployment would likely lead to reduced demand for travel. Also, an increase in costs associated with auto use is likely to result in increased transit demand, even if transit fares remain the same. When such changes in conditions (other than price) occur, they are represented as a shift in the supply curve shown in the figure below (upward shift for increased demand, $S_1 \rightarrow S_2$; and downward shift for decreased demand, $S_1 \rightarrow S_3$).



Increases in transportation supply may be traditionally thought of in terms of increasing the fleet size of a transit company or building new roads or increasing the number of lanes for existing roads. However, it is possible to increase supply without such physical capital-intensive investments. For instance, the use of intelligent transportation systems could lead to increased supply without any physical enlargements of the road network.

3.4.1 Causes of Shift in Transportation Supply Curve

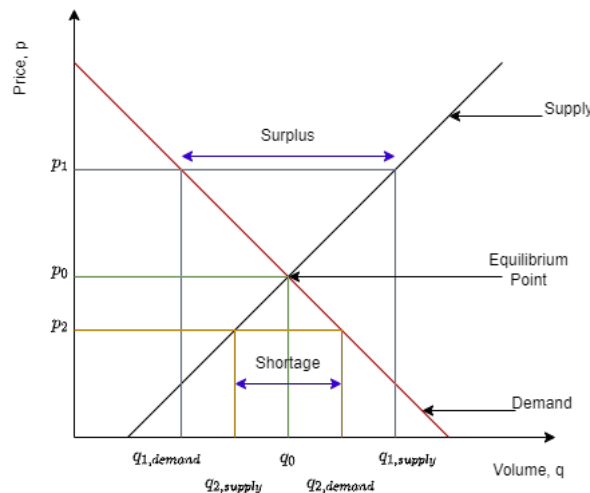
The supply of transportation service may change even if the price remains constant, for reasons such as:

- **Price of rival transportation services:** The supply of a service may decrease if there is a decrease in the price of competing transportation service. This may apply to toll roads, where profit is the primary intent, and to a lesser extent, non-toll roads.
- **Number of transportation modes:** An increased number of modes, such as construction of a subway in a city that already has buses and light rail transit and facilities for autos, indicates an increase in supply, shifting the supply curve towards right.
- **Price of relevant input:** If the cost of resources used to produce a transportation service increases, the transportation agency would be less capable of supplying the same quantity at a given price, and the supply curve will shift towards left.

3.5. EQUILIBRIUM OF TRANSPORTATION DEMAND AND SUPPLY³³

- **Technology:** Technological advances that increase facility capacity or efficiency causes the supply curve to shift towards right.

3.5 Equilibrium of Transportation Demand and Supply



We have seen that the demand function is a relation between quantity demanded of a good and its price. Similarly, the supply function (Service function) represents the quantity of goods a producer is willing to offer at a given price. If the demand and supply function for a transportation facility are known, then it is possible to deal with the concept of equilibrium. Equilibrium is said to be attained when factors that affect the quantity demanded and those that determine the quantity supplied result in being statically equal (or converging toward equilibrium).

Numerical example:

The travel time on a stretch of highway lane connecting two activity centers has been observed to follow the equation representing the service function:

$$t = 15 + 0.02v$$

Where t and v are measured in minutes and vehicles per hour respectively. The demand function for travel connecting the two activity centers is $v = 4000 - 120t$.

- Sketch these two equations and determine the equilibrium time and speed of travel.
- If the length of the highway lane is 20 miles, what is the average speed of vehicles traversing this length?

Solution:

Demand function: $v = 4000 - 120t$

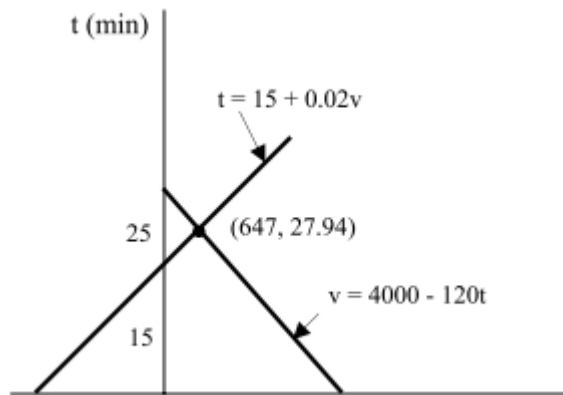
Supply function: $v = 4000 - 120t$

Solving these two equations simultaneously yields:

$v = 647$ vehicles/hour

$t = 27.94$ minutes

$$\therefore \text{speed} = \frac{20 \times 60}{27.94} = 42.95 \text{ mph}$$



3.6 Sensitivity of Travel Demand

In the planning and evaluation of transportation systems and associated investments, it is often useful to have knowledge of the changes in transportation demand caused by changes in attributes of the transportation system or its environment. A particular instance is the change in demand for a given mode in response to changes in price of that mode. Given the functional form of the travel demand function, it is possible to derive a marginal effects model that estimates any one of the following:

- Change in demand in response to unit change in attribute
- Change in demand in response to unit percent change in attribute
- Percent change in demand in response to unit percent change in attribute

The law of demand states that a fall in the price of a good raises the quantity demanded. The price elasticity of demand measures how much the quantity demanded responds to a change in price. Demand for a good is said to be elastic if the quantity demanded responds substantially to changes in the

price. Demand is said to be inelastic if the quantity demanded responds only slightly to changes in the price.

Transportation demand elasticity may be defined as the degree of responsiveness of transportation demand in response to a unit change in demand-related or attributes such as price or income. This is typically expressed as follows:

$$e_p = \frac{\text{percentage change in quantity demanded}}{\text{percentage change in price}}$$

$$e_p = \frac{\frac{\partial q}{q}}{\frac{\partial p}{p}} = \frac{\partial q}{\partial p} \times \frac{p}{q} \quad (3.2)$$

Arc elasticity can be calculated as:

$$\text{Arc elasticity} = \frac{\frac{\partial q}{\partial p}}{\frac{p}{q}} = \frac{\partial q}{\partial p} \times \frac{p}{q} = \frac{\frac{Q_1 - Q_0}{\frac{Q_1 + Q_0}{2}}}{\frac{P_1 - P_0}{\frac{P_1 + P_0}{2}}} = \frac{Q_1 - Q_0 \times (P_1 + P_0)/2}{P_1 - P_0 \times (Q_1 + Q_0)/2} \quad (3.3)$$

Where, Q_0 and Q_1 represent quantity of travel demanded corresponding to prices P_0 and P_1 respectively.

3.6.1 Elasticity and Revenue Along a Linear Demand Curve

For a linear demand function, we can determine the elasticity of demand by taking derivative of equation 3.1:

$$\frac{dq}{dp} = -\beta$$

Also from equation 3.1,

$$p = \frac{\alpha - q}{\beta}$$

We have,

$$e_p = \frac{\partial q}{\partial p} \times \frac{p}{q}$$

Substituting the values of $\frac{dq}{dp}$ and p we get,

$$\therefore e_p = -\beta \times \frac{\alpha - q}{\beta q} = 1 - \frac{\alpha}{q} \quad (3.4)$$

When the elasticity is less than -1 (i.e., more negative than -1), the demand is described as being elastic, meaning that the resulting percentage change in quantity will be larger than the percentage change in price. In this case, demand is relatively sensitive to price change. However, when elasticity is between 0 and -1 , the demand is described as being inelastic or relatively insensitive. These ranges are shown in the figure below:



A linear demand curve has several interesting properties. Notice, as one moves down the the demand curve, the price elasticity of demand becomes smaller (i.e., more inelastic). In fact, the elasticity at a given point equals the length of the demand line segment below the point divided by the length of the line segment above it. Another point to is that the slope of the line is constant, but the elasticity changes from ∞ at the top, where the demand line intersects the vertical axis, ti 0, where the demand line intersects the horizontal axis. Because elasticity changes along the demand curve, it is essential to specify what range of prices or quantity the elasticity was measured.

<i>value</i>	<i>Nature of Demand</i>	<i>Relation between Price and Revenue</i>	<i>Impact of increase in price</i>
$e > 1$	elastic	negatively related	reduces revenue
$e < 1$	inelastic	positively related	increases revenue
$e = 1$	unit elasticity	none	remains the same

3.7 Kraft Demand Function

We occasionally come across a demand function where the elasticity of demand for travel with respect to its price is essentially constant. The demand function for such a situation corresponds to the equation.

$$Q = \alpha(p)^\beta \quad (3.5)$$

Where α and β are constant parameters of the demand function. To prove that this function is has a constant elasticity, we differentiate this function

with respect to price, we get:

$$\frac{dQ}{dp} = \alpha \beta p^{\beta-1}$$

Now substituting this value in standard elasticity equation 3.2,

$$\begin{aligned} e_p &= \frac{dQ}{dp} \times \frac{p}{Q} \\ &= \alpha \beta p^{\beta-1} \times \frac{p}{Q} \\ &= \alpha \beta p^{\beta-1} \times \frac{p}{\alpha(p)^\beta} \text{ [from equation 3.5]} \\ &\therefore e_p = \beta \end{aligned} \tag{3.6}$$

Thus, β , the exponent of price (a constant parameter), is the price elasticity.

Numerical Example:

The elasticity of transit demand with respect to price has been found to be equal to -2.75, which means that a 1% increase in transit fare will result in 2.75 decrease in the number of passengers using the system. A transit line on this system carries 12,500 passengers per day, charging 50 cents per ride. The management wants to raise the fare to 70 cents per ride. What advice would you offer to the management?

Solution:

Since the elasticity of the demand is a constant, the demand function is a Kraft Demand Model,

$$\begin{aligned} Q_1 &= 12500 \\ p_1 &= 50 \\ p_2 &= 70 \\ \beta &= -2.75 \end{aligned}$$

We have,

$$\begin{aligned} Q_1 &= \alpha(p_1)^\beta \\ 12500 &= \alpha \times 50^{-2.75} \\ \therefore \alpha &= 5.876 \times 10^8 \end{aligned}$$

An increase in fare from 50 to 70 cents will attract a demand of,

$$\begin{aligned} Q_2 &= \alpha(p_2)^\beta \\ \therefore Q_2 &= 5.876 \times 10^8 \times 70^{-2.75} = 4955 \text{ passengers} \end{aligned}$$

Therefore, the increase in fare from 50 to 70 cents (40% increase) is likely to reduce the patronage on this line from 12500 passengers per day to 4955

(60 % decrease). In terms of revenue, the results are as follows:

Revenue before increase in fare (R_1) = $50 \times 12500 = \$6250$
 Revenue after increase in fare (R_2) = $70 \times 4955 = \$3468.50$
 \therefore Loss in revenue = $R_1 - R_2 = \$6250 - \$3468.50 = \$3406$

Hence, the management is not advised to increase the fare.

3.8 Factors Affecting Elasticity

As stated earlier, elasticity is the change in demand in response to a unit change in levels of attributes of the transportation system or its environment. Such attributes include characteristics of the transportation system such as price and level of service associated with a given mode, price and level of service of competing modes, and characteristics of the socio-economic system such as income, level of employment, household size, car ownership, etc. Of these factors, of particular interest are price and income. The elasticity of demand in response to price and income are known as price elasticity and income elasticity, respectively.

3.8.1 Income Elasticity

The elasticity of demand with respect to income, or income elasticity, is the change in demand for a good in response to a unit change in income of the consumer of that good. Income elasticity has a special significance in travel demand modeling. Often, the transportation planner seeks to evaluate the impact of changing socioeconomic trends on the demand for or share of various modes of transportation. A major indicator of economic trends is income. In disaggregate demand modeling, it is sought to determine the sensitivity of changing income on the demand for a particular mode. Income elasticity is generally defined as the change in demand in response to a unit change in income. In transportation economics, a good service is considered normal if there is a direct relationship between the demand for that commodity and the income of the consumer. Besides increased demand, if the share of the demand for that good (in the consumer's total income) also increases, then the good is described as a superior good. If the demand of a good decreases with increasing income then the good is described as inferior. In developed countries, automobile travel is considered superior while mass transit is considered an inferior good. Mathematically,

$$e_i = \frac{\% \text{ change in quantity of goods/service demanded}}{\% \text{ change in income}} \quad (3.7)$$

3.8.2 Price Elasticity

The elasticity of demand with respect to price, or price elasticity, is the change in demand for a good in response to a unit change in the price of the good. A study of price elasticity is important because it is often sought to assess the impacts of changing prices of a good or rival goods (due to past supply and demand conditions) on the demand of that good. The level of price elasticity depends on factors such as price of rival goods, income-share of the good, the scope of definition of the good, and whether the good is considered a luxury or a necessity. Some factors that affect price elasticity are as follows:

- *Price of rival goods:* A consumer who spends a substantial percentage of income on a particular good is more likely to seek a substitute good when the price of the good increases.
- *Scope of definition:* Goods that have narrow definitions are more likely to have more substitutes, and are therefore expected to have a more elastic demand.
- *Price and availability of rival goods:* The lower the price and greater availability of substitutes, the greater the elasticity of demand of the good with respect to price.
- *Luxury vs. substitute goods:* Goods that are considered necessities typically have inelastic price, while luxury goods have relatively elastic price.

3.9 Elasticity and Total Revenue

It is possible to tell what the revenue (price \times output) of a firm is likely to be if the price of a unit changes. Some general relationship between elasticity and revenue are given below:

- When demand is inelastic ($e < 1$), price and total revenue move in the same direction: If the price increases, total revenue also increases.
- When demand is elastic ($e > 1$), price and total revenue move in opposite directions: If the price increases, total revenue decreases.
- If demand is unit elastic ($e = 1$), total revenue remains constant when the price changes.

3.10 Application of Price Elasticity: Consumer Surplus and Latent Demand

Consumer surplus is a measure of the monetary value made available to consumers by the existence of a facility. It is defined as the difference between what the consumers might be willing to pay for a service and what they actually pay. A patron of a bus service pays a fare of, say, 50 cents per trip but would be willing to pay up to as much as 75 cents per trip. In this case, her consumer surplus is 25 cents.



Consumer surplus may be determined from a survey of customers to determine how much they are willing to pay for a good or service. Recent studies have evaluated the willingness of road users to pay for improvements that would enhance safety on their highways. Such information can be combined with data on actual payments to determine the amount of consumer surplus per person or for a group of persons. Alternatively, consumer surplus can also be determined from the demand-supply curve, as the demand curve can be considered as an indicator of the utility of the good or service in terms of price, as illustrated in the figure above. The area ABC represents the total consumer surplus. Maximization of consumer surplus is the maximization of the economic utility of the consumer. The use of the consumer surplus concept is common in the area of the evaluation of transit systems.



Generally, the economic impacts of improvements in a transportation system can be evaluated in terms of consumer surplus, and can be represented as the area under a demand curve and a shift in the supply curve. The figure above indicates the case of an urban arterial street with a traffic supply S_1 , intersecting a demand curve at E_1 . An additional lane is added, thus shifting the supply curve to S_2 and therefore intersecting the demand curve at E_2 . The change in consumer surplus can be quantified as the trapezoidal area P_1, P_2, E_1, E_2 . In other words,

$$E = (P_1 - P_2) \times \frac{(Q_1 + Q_2)}{2}$$

The consumer surplus can be measured as the difference between the maximum amount which consumers are willing to pay for a specified quantity of a good or service rather than going without it. In general, the area $AOQB$ in Figure 5-9 represents the total community benefit, $BCOQ$ is equal to the market value, and ACB is equal to the consumer surplus or net community benefit. For a constant supply curve, a higher elasticity is associated with a smaller consumer surplus, and vice versa.

The foremost figure also illustrates an additional concept that is useful to transportation engineers: latent demand. It can be observed that travelers between Q and D do not make trips at the current time, but would do so if the price per trip were lower than the equilibrium price. The number of such potential travelers is referred to as latent demand. An application of this concept is the investigation of latent demand associated with introduction of incentives for non-peak travel, such as reducing transit fares during non-peak hours. From Figure 5-8, it can be seen that if transit travel were free (zero trip price), the travel demand (quantity of trips demanded) would be $Q_L - Q$.

Numerical Example:

A bus company with an existing fleet of one hundred 40-seater buses increases its fleet size by 20% and reduces its \$1.00 fare to 90 cents per ride. Determine the impact of the price change on the company's profitability, change in consumer surplus, and the price elasticity of demand. It is assumed that the current load factor of 90% would increase to 95% after the price change. Assume that all buses in the fleet are being used during the peak hours.

Solution:

Existing situation:

$$\text{Demand} = 100 \text{ buses} \times 40 \text{ seats} \times 0.90 = 3600 \text{ persons per hour}$$

$$\text{Revenue} = \text{Demand} \times \text{Price} = 3600 \times 1.00 = \$3600 \text{ per hour}$$

After the price drop and supply increase:

$$\text{Demand} = 120 \text{ buses} \times 40 \text{ seats} \times 0.95 = 4560 \text{ persons/hour}$$

$$\text{Revenue} = 4560 \times 0.90 = \$4104 \text{ per hour}$$

After the price change, the company gains $\$4104 - \$3600 = \$504$ per hour.

$$\text{Change in consumer surplus} = (1.00 - 0.90) \times (3600 + 4560)/2 = \$408/\text{hr}$$

$$\text{Also, Price elasticity of demand, } e_p = \frac{Q_1 - Q_0 \times (P_1 + P_0)/2}{P_1 - P_0 \times (Q_1 + Q_0)/2} = \frac{960 \times 0.95}{0.1 \times 4080} = -2.235$$

Numerical Example:

The Demand for a transit service between a city and its largest suburb during an off-peak hour V, is given by $2500 - 350t$ where t is the travel time in minutes. At the current time, the transit trip takes an average of 5 minutes. Determine the time elasticity of demand and the latent demand at this travel.

Solution:

$$V = 2500 - 350t$$

Diff the above equation with respect to t we get,

$$\frac{dV}{dt} = -350$$

Time elasticity of demand,

$$e_t = \frac{dV/V}{dT/t}$$

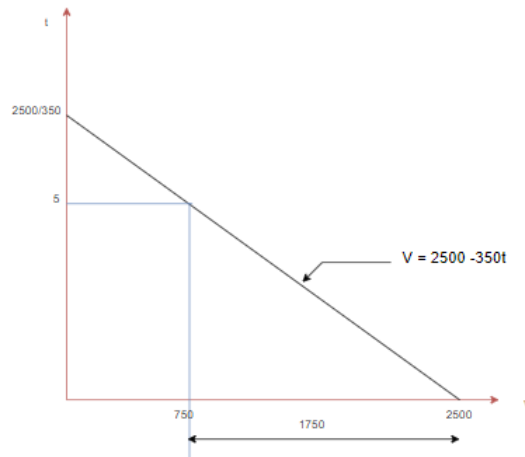
$$\text{or, } e_t = -350 \times \frac{5}{(2500 - 350 \times 5)}$$

$$\therefore e_t = -2.3$$

Also, at e = 5 minutes,

$$V = 2500 - 350 \times 5 = 750$$

$$\therefore \text{Latent demand} = 2500 - 750 = 1750$$



3.11 Cost Analysis in the Evaluation of Transportation Systems

A complete and balanced evaluation of alternative investment options is possible only by giving due consideration to both benefits and costs associated with each alternative. For this reason, it is essential to have knowledge of the costs of each aspect of the provision of a transportation good or service. This way, the future costs of such aspects can be determined using average cost values or better still, cost models that estimate cost as a function of investment and facility attributes.

In economic theory, three types of costs are encountered: fixed costs (which are independent of the volume of goods produced), variable costs (which depend on the volume of good produced), and total costs (the sum of fixed and variable costs). For each of these costs, it is possible to find the average cost (dividing the total production cost by the number of goods produced) and marginal cost (the incremental cost of producing an additional unit). It is therefore possible to determine average fixed costs, marginal fixed costs, average total costs, etc.

3.11.1 Economic Laws Related to Cost

- *Law of Diminishing Returns*: States that an increase in input of one unit of a factor of production generally causes an increase in output, but only up to a point, after which increasing inputs of that factor will result in progressively less increase in output.

- *Law of Increasing Returns to Scale*: States that in practice, the production of units is often likely to increase at a faster rate than the increase in the factors of production. This may be due to technological features, specialization.

3.11.2 Average Cost

The Average Cost is the per unit cost of production obtained by dividing the total cost by the total output. By per unit cost of production, we mean that all the fixed and variable cost is taken into the consideration for calculating the average cost. Thus, it is also called as Per Unit Total Cost. Mathematically,

$$C = cq = \alpha + \beta(q) \quad (3.8)$$

where,

C = total cost of a product

c = unit cost

q = magnitude of output

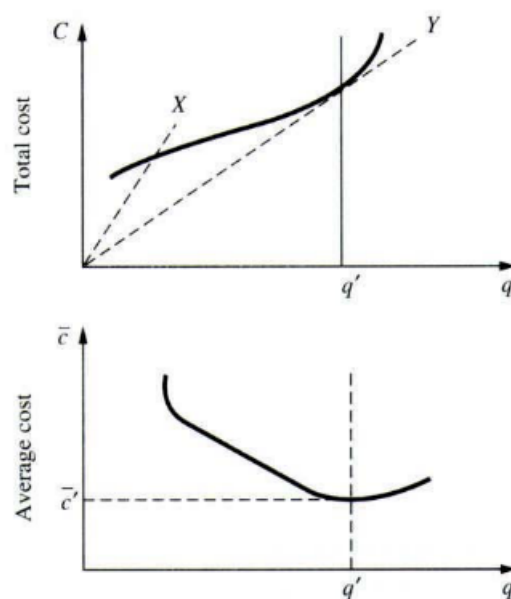
α = fixed cost of production

$\beta(q)$ = variable cost of production

Therefore, Average cost:

$$\bar{c} = \frac{C}{q} = \frac{\alpha + \beta(q)}{q} = \frac{\alpha}{q} + \frac{\beta(q)}{q} \quad (3.9)$$

The relationships of the total and average cost functions are shown the figure below. It can be seen that as output q increases, the average cost of production decreases and then increase at higher levels of production. When the production level reaches q' , the average cost is a minimum \bar{c}' . The decrease in average cost with increasing output is referred to as economies of scale. In the figure, there is obviously there is economy of scale for production levels between 0 and q' . However, there is no economy of scale beyond q' because the average cost increases. This concept is useful to engineers in deciding whether additional capacity or growth would yield higher profits, and is important in the economic evaluation of transportation system improvements.

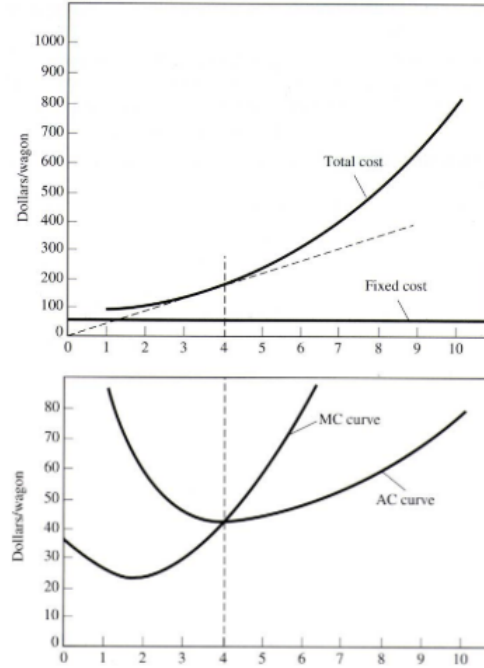


3.11.3 Marginal Costs

The marginal costs of a transportation good or service is the additional cost associated with the production of an additional unit of output. The following example illustrates the concepts of average and marginal costs.

Case Study

(1)	(2)	(3)	(4)	(5)	(6)
Number of wagon/train	Fixed cost, FC	Variable cost, VC	Total cost, TC	Average cost, AC	Marginal cost/unit, MC
1	55	30	85	85.0	
2	55	55	110	55.0	25
3	55	75	130	43.3	20
4	55	105	160	40.0	30
5	55	155	210	42.0	50
6	55	225	280	46.7	70
7	55	315	370	52.9	90
8	55	425	480	60.0	110
9	55	555	610	67.8	130
10	55	705	760	76.0	150



The table above presents the cost of running a train system with variable number of wagons. For each system size, the fixed and variable costs are provided in the first three columns. The total, average and marginal costs are then computed and presented in the next three columns.

The various costs are computed as follows:

$$\text{Total cost} = TC(x) = FC + VC(x)$$

$$\text{Average Total Cost} = AC(x) = \frac{TC(x)}{x} = \frac{FC}{x} + \frac{VC(x)}{x}$$

$$\text{Marginal Cost} = MC(x) = TX(x) - TC(x-1)$$

The second graph in the figure above presents the curves corresponding to average cost, marginal cost and total costs for the train system. It can be seen that the point of minimum cost (\$40) occurs at the intersection of the average cost (AC) and marginal cost (MC) curves. It is also observed that the projection of this point to the first graph in the figure above corresponds to the point where the gradient of the tangent drawn from the origin has the minimum slope.

When the output is a continuous function, the differential form of the marginal cost is used, in which the marginal cost is the rate of change of total cost with respect to a change in output. In this form the equation is:

$$MC(x) = \frac{dTC(x)}{dx} = \frac{dVC(x)}{dx} \quad (3.10)$$

From the geometry of the AC and MC curves, it is also seen that the average costs is proportional to the slope of a line connecting the origin of the total cost curve with a point on that curve corresponding to the total output. In our example, the slope of such a line begins at infinity at zero output and then decreases to its lowest point, when $x = 4$. Beyond this point, the slope increases again. On the other hand, the marginal cost curve is the slope of the tangent drawn at any point on the total cost curve.

3.11.4 Costs and Production

In general, a private company will continue to produce and market a good or service as long as it is returning a profit. The net profit (P) is equal to the total revenue (R) minus the total costs (C).

$$P = R - C = pq - cq \quad (3.11)$$

Where p is the selling price of one unit of product q , and c is the production cost of one unit.

The necessary condition for profit maximization is:

$$\begin{aligned} \frac{dP}{dq} &= \frac{dR}{dq} - \frac{dC}{dq} = 0 \\ \text{or, } \frac{dP}{dq} &= \frac{d(pq)}{dq} - \frac{d(cq)}{dq} = 0 \\ \therefore \frac{d(pq)}{dq} &= \frac{d(cq)}{dq} \end{aligned}$$

Let,

$$\text{Marginal Revenue (MR)} = \frac{dR}{dq} = \frac{d(pq)}{dq}$$

$$\text{Marginal Cost (MC)} = \frac{dC}{dq} = \frac{d(cq)}{dq}$$

$$\therefore MR = MC$$

This equation implies that the firm seeking to maximize profits should produce at the point where marginal revenue equals marginal costs.

3.11.5 Cost Elasticity

The cost elasticity of a good or service is defined as the ratio of percentage change in cost C to a unit percentage change in supply q . The difference between cost elasticity and price elasticity is obvious.

$$e_c = \frac{\% \text{change cost}}{\% \text{change in supply}} = \frac{\Delta C/C}{\Delta q/q} = \frac{q}{C} \times \frac{\Delta C}{\Delta q} \quad (3.12)$$

In the limit when $\Delta q = 0$, $\frac{\Delta C}{\Delta q} = \frac{dC}{dq}$, then,

$$e_c = \frac{dC/dq}{C/q} = \frac{MC}{AC} \quad (3.13)$$

Numerical Example:

A transport company hauling goods by truck has cost function $C = 1.5q^{1.25}$, where C is the total costs of supply, q.

- Determine the average cost and the marginal cost of production
- Prove that the cost elasticity is 1.25
- Is there an economy of scale?

Solution:

$$\bar{C} = \frac{C}{q} = \frac{1.5q^{1.25}}{q} = 1.5q^{0.25}$$

$$MC = \frac{dC}{dq} = 1.5 \times 1.25 \times q^{0.25} = 1.875q^{0.25}$$

$$e = \frac{MC}{AC} = \frac{1.875q^{0.25}}{1.5q^{0.25}} = 1.25$$

\therefore Economy of scale does not exist because the average cost increases with increases q.

3.12 Transportation Projects Evaluation

Construction of highways along with the other associated infrastructures, construction of a new freight center, acquisition of new vehicles for the transit e.t.c. represent transportation projects faced by transportation experts. In the initial stage of these projects, it is necessary to perform their evaluation, that is, as precise as possible, analyze and review economic, environmental, equity, as well as other project impacts. The planners and engineers must properly answer the following questions: (a) Are the transportation project's benefits greater than the projects' costs?; (b) What is the best project's alternative in the case when project has few mutually exclusive alternatives?; (c) How to allocate available funds among competitive transportation projects?; (d) When to start the considered project?

Frequently, financial resources are scarce, and appropriate engineering economic analysis can significantly help planners and decision-makers to allocate available resources properly. In the first step of any project evaluation, it is necessary to analyze the project's socio-economic context, and to clearly define the project's objectives. In the next steps, the analysts should clearly

recognize the type of costs and benefits, compare them and make recommendations to the decision makers.

Transportation projects usually extend over many years. On the other hand, the purchasing power of money decreases over time. The main cause of this phenomenon is the inflation that exists in every society. A discount rate regulates the value of money for time. This rate is used to represent future monetary quantities in terms of their today's value. Compounding and discounting are techniques that enable us to compare money values at different points in time. Let us briefly explain compounding technique.

Let us assume that we want to invest \$100 these days, at an annual interest rate (r) of 5%. It will be $\$100 + \$5 = \$105$ in 1 year. After two years it will be $105\$ + \$0.05 * 105 = \$110.25$. After 3 years we will have $\$110.25 + \$0.05 * \$110.25 = \115.7625 . Discounting represents reverse operation of compounding. The compounding technique helps us to find the answer to the following question: what is the present value (PV) of a known future amount of money?

In our example, the PV of \$105 next year, when $r = 5\%$, is \$100. The PV equals

$$PV = V_t / (1 + r)^t \quad t = 0, 1, 2, \dots, n \quad (3.14)$$

Where n is the project duration (in years), r is the rate of discount, and V_t is the value in year t .

There are two interest rates that are used in transportation projects evaluation. The first one is the real interest rate that is exclusive of inflation, while the second one is the nominal interest rate that is inclusive of inflation. Project value is usually expressed as a NPV. This value represents a project's value or cost for its whole life cycle in today's dollars.

When evaluating transportation projects many governments and funding agencies in the world (OECD, World Bank, etc.) require a cost-benefit analysis (CBA) to be performed. In this way, the CBA represents common evaluation language between the governments, funding agencies and the transportation project supporters.

3.13 Cost-Benefit Analysis

The CBA is a method that calculate and compare project's costs and benefits to society over period of time. The CBA monetizes all project's inputs and outputs. In other words, the CBA converts the inputs and the outputs

into a monetary values. The CBA helps decision-makers to rank and prioritize various project's alternatives including also alternative "no action" ("no action," or "do nothing" case assumes continued operation of the existing facility, exclusive of any major investments). The specific transportation project should start only when the CBA clearly shows to the decision makers that the total benefits to society outweigh the total costs. When performing CBA, the analysts enumerate all project's costs and benefits to society. In the next step, they assign monetary values to costs and benefits, and discount them to a NPV. All costs, as well as all benefits are added into a single number. The transportation project is evaluated by using total costs, and the total benefits values.

In the first step of the CBA, it is necessary to identify transportation project's alternatives to be evaluated. Alternatives may represent "do nothing" case, rehabilitation of existing facility, construction of a new facility, etc. Transportation projects have consequences over time. The analysts should also define the time period over which the life cycle costs and benefits of all of the alternatives will be calculated.

The project's economic performances are measured by the following indicators:

NPV: Net Present Value

IRR: Internal Rate of Return

B/C: The Benefit-Cost ratio

The majority of experts consider the NPV as the most important CBA indicator. The NPV is defined in the following way:

$$NPV = \sum_{t=0}^{\infty} \frac{B_t - C_t}{(1 + r)^t} \quad (3.15)$$

Where,

B_t : Benefits in year t

C_t : Costs in year t

n : Project duration (in years)

r : interest rate

t : year index

Project's benefits and costs are forecast over the project duration. For example, benefits from road investment could be shorter traveling distance, shorter travel time, reduced number of traffic accidents, etc. The road improvement costs could be project design costs, labor costs equipment costs, material costs, etc.

The analysts use a NPV to express a project's worth for its complete life cycle in today's money value. We see that the NPV decreases in r (interest rate) increases. In case when $NPV > 0$, the project may be accepted. However, when $NPV < 0$, the considered project should be rejected. Finally, when $NPV = 0$, we conclude that the considered project adds no monetary value. The final decision about such transportation project should be based on some additional criteria.

The internal rate of return (IRR) is the indicator that also measures the project's performances. The IRR is the discount rate/interest rate at which the $NPV = 0$. We calculate the IRR by solving the following equation:

$$\sum_{t=0}^{\infty} \frac{B_t - C_t}{(1 + IRR)^t} = 0 \quad (3.16)$$

The average values of the observed IRR 's in a sample of investment projects sponsored by the European Union (EU) at the end of the 20th century are approximately equal to 15% in the cases of roads and highways, 10% in the cases of railways and underground, and 25% in the cases of ports and airports.

After calculation net benefits B and net costs C , the benefit/cost ratio ($\frac{B}{C}$) should be also calculated. (Frequently, it is not easy to estimate future costs, and, especially project's benefits.) The benefit/cost ratio ($\frac{B}{C}$) informs us about the improvement in traffic operations (expressed in dollars) per dollar invested.

Analysts and engineers usually perform a sensitivity analysis to conclude how sensitive final results are to changes in hypothesis about the costs, benefits, and discount rate.

The main weakness of the CBA is that all transportation project benefits are evaluated only in monetary terms. It is very complicated to value all the costs and benefits of transportation projects in monetary terms. In other words, in many situations, social and environmental aspects of the considered transportation projects are not treated adequately. For example, many traffic safety programs, actions and projects involve the prevention of loss of life. The logical and ethical question is how should we value a life saved? A pure economic approach would suggest to us that the value of life is equal to the PV of lifetime earnings. Obviously, there are numerous opponents to such an oversimplified and ethically questioned approach.

3.14 Numerical Examples:

Example #1: Selecting a transportation mode

An individual is planning to take a trip between the downtown area of two cities, A and B, which are 400 miles apart. There are three options available:

1. **Travel by air:** This trip will involve driving to the airport near city A, parking, waiting at the terminal, flying to airport B, walking to a taxi stand, and taking a taxi to the final destination.
2. **Travel by auto:** This trip will involve driving 400 miles through several congested areas, parking in the downtown area, and walking to the final destination.
3. **Travel by rail:** This trip will involve taking a cab to the railroad station in city A, a direct rail connection to the downtown area in city B, and a short walk to the final destination.

Since this is a business trip, the person making the trip is willing to pay up to \$25 for each hour of travel time reduced by a competing mode. (For example, if one mode is two hours faster than another, the traveler is willing to pay \$50 more to use the faster mode.) After examining all direct costs involved in making the trip by air, auto, or rail (including parking, fuel, fares, tips, and taxi charges) the traveler concludes that the trip by air will cost \$250 with a total travel time of five hours, the trip by auto will cost \$200 with a total travel time of eight hours and the trip by rail will cost \$150 with a total travel time of 12 hours.

- Which mode is selected based on travel time and cost factors alone?
- What other factors might be considered by the traveler in making a final selection?

Solution: Since travel time is valued at \$25/hr, the following costs would be incurred:

$$\text{Air : } 250 + 25 * 5 = \$375$$

$$\text{Auto : } 200 + 25 * 8 = \$400$$

$$\text{Rail : } 150 + 25 * 12 = \$450$$

In this instance, the air alternate reflects the lowest cost and is the selected mode. However, the traveler may have other reasons to select another alternative. These may include the following considerations.

Safety: While each of these modes is safe, the traveler may feel “safer” in one mode than another. For example, rail may be preferred because of concerns regarding air safety issues.

Reliability: If it is very important to attend the meeting, the traveler may select the mode that will provide the highest probability of an on-time arrival. If the drive involves travel through work zones and heavily congested areas, rail or air would be preferred. If potential air delays are likely due to congestion, flight cancellations, or inclement weather, another mode may be preferred.

Convenience : The number of departures and arrivals provided by each mode could be a factor. For example, if the railroad provides only two trains/day and the airline has six flights/day, the traveler may prefer to go by air.

Example #2: Computing the Toll to Maximize Revenue Using a Supply—Demand Curve

A toll bridge carries 5000 veh/day. The current toll is 150 cents. When the toll is increased by 25 cents, traffic volume decreases by 500 veh/day. Determine the amount of toll that should be charged such that revenue is maximized. How much additional revenue will be received?

Solution: Let x the toll increase in cents.

Assuming a linear relation between traffic volume and cost, the expression for V is

$$V = 5000 - x/25(500)$$

$$\text{The toll is, } T = 150 + x$$

Revenue is the product of toll and volume,

$$\begin{aligned} R &= V * T \\ &= (5000 - x/25 * 500)(150 + x) \\ &= (5000 - 20x)(150 + x) \\ &= 750,000 + 2000 - 20x^2 \end{aligned}$$

For maximum value of x ,

$$\begin{aligned} \frac{dR}{dx} &= 0 \\ 20,000 - 40x &= 0 \\ \therefore x &= 50 \text{cents} \end{aligned}$$

The new toll is the current toll plus the toll increase.

Toll for maximum revenue = $150 + 50 = 200 \text{cents}$

The additional revenue, AR

$$\begin{aligned}
 &= V_{max} * T_{max} - V_{current} * T_{current} \\
 &= 5000 - 50/25 * 500 * 2 - 5000 * 1.5 \\
 &= 4000 * 2 - 7500 \\
 &= 8000 - 7500 \\
 &\therefore AR == \$500
 \end{aligned}$$

Example #3:

The demand function for transportation from the suburbs to downtown in a large city is as follows:

$$Q = T^{-0.3} C^{-0.2} A^{0.1} I^{-0.25}$$

Where,

Q = number of transit trips

T = travel time on transit(hours)

C = fare on transit (dollars)

A = cost of automobile trip (dollars)

I = average income (dollars)

1. There are currently 10,000 persons per hour riding the transit system, at a flat fare of \$1 per ride. What would be the change in ridership with a 90 cent fare? What would be company gain/lose per hour?
2. By auto, the trip costs \$ (including parking). If the parking charge were raised by 30 cents, how would it affect the transit ridership?
3. The average income of auto riders is \$15,000 per year. What raise in salary will rider require to cover their costs in view of the change in parking charge noted in the second question?

Solution:

Part 1:

This is essentially a Kraft model. The price elasticity of demand for transit trips,

$$e_p = -0.2$$

$$\% \text{ Reduction in fare } \left(\frac{dC}{C} \right) = \frac{100-90}{100} = 10\%$$

$$\% \text{ Increase of Patronage } \left(\frac{dQ}{Q} \right) = e_p \times \frac{dC}{C} = 0.2 \times 0.1 = 0.02 = 2\%$$

Now,

Initial Patronage (Q_1) = 10000

Final Patronage (Q_2) = $10000 + 0.02 \times 10000 = 10200$

Also,

Revenue before increment of patronage (R_1) = $Q_1 \times 1 = 10000$

Revenue after increment of patronage (R_2) = $Q_2 \times 0.9 = 9180$

\therefore The company will lose \$820 per hour.

Part 2:

Similarly, the automobile price cross-elasticity of demand, $e_{cp} = 0.1$

% rise in auto costs ($\frac{dA}{A}$) = $\frac{3.3-3}{3} = 10\%$

% increase of patronage ($\frac{dQ}{Q}$) = $e_{cp} \times \frac{dA}{A} = 0.1 \times 0.1 = 1\%$

Now,

Final patronage = $10000 + 0.01 \text{ times } 10000 = 10100$

\therefore The transit patronage increases from 10000 riders to 10100 riders.

Part 3:

Similarly, the income elasticity of demand, $e_i = -0.25$

$\frac{dQ}{Q} = 1\% = 0.01$

We have,

$$\frac{dI}{I} = \frac{\frac{dQ}{Q}}{e_i} = \frac{0.01}{-0.25} = 0.04 = 4\%$$

So, a 4% increase in income would cover a 30 cent increase in auto cost.

If a the average income is % 15000, \$600 raise in salary would change the minds of those auto drivers who were planning to ride the transit system.

3.15 Review Problems

Chapter 4

Insurance

4.1 Introduction

There is no single definition of insurance. Insurance can be defined from the viewpoint of several disciplines, including law, economics, history, actuarial science, risk theory, and sociology. But each possible definition will not be examined at this point. Instead, we will examine the common elements that are typically present in any insurance plan. However, before proceeding, a working definition of insurance—one that captures the essential characteristics of a true insurance plan—must be established.

After careful study, the Commission on Insurance Terminology of the American Risk and Insurance Association has defined insurance as follows. *Insurance is the pooling of fortuitous losses by transfer of such risks to insurers, who agree to indemnify insureds for such losses, to provide other pecuniary benefits on their occurrence, or to render services connected with the risk.* Although this lengthy definition may not be acceptable to all risk managers and insurance scholars, it is useful for analyzing the common elements of a true insurance plan.

4.2 Basic Characteristics of Insurance

Based on the preceding definition, an insurance plan or arrangement typically includes the following characteristics:

4.2.1 Pooling of Losses

Pooling or the sharing of losses is the essence of insurance. *Pooling is the spreading of losses incurred by the few over the entire group, so that in the process, average loss is substituted for actual loss.* In addition, pooling involves the grouping of a large number of exposure units so that the law of

large numbers can operate to provide a substantially accurate prediction of future losses. Ideally, there should be a large number of similar, but not necessarily identical, exposure units that are subject to the same perils. Thus, pooling implies (1) the sharing of losses by the entire group and (2) the prediction of future losses with some accuracy based on the law of large numbers.

The primary purpose of pooling, or the sharing of losses, is to reduce the variation in possible outcomes as measured by the standard deviation or some other measure of dispersion, which reduces risk. For example, assume that two freight business owners each own an identical storage building valued at \$50,000. Assume there is a 10% chance in any year that each building will be destroyed by a peril, and that a loss to either building is an independent event. The expected annual loss for each owner is \$5,000 as shown below:

$$\text{Expected Loss} = 0.90 \times \$0 + 0.10 \times \$50000 = \$5000$$

A common measure of risk is the standard deviation, which is the square root of the variance. The standard deviation (SD) for the expected value of the loss is \$15,000, as shown below:

$$\text{SD} = \sqrt{0.9 \times (0 - 5000)^2 + 0.1 \times (50000 - 5000)^2} = \$15000$$

Suppose instead of bearing the risk of loss individually, the two owners decide to pool (combine) their loss exposures, and each agrees to pay an equal share of any loss that might occur. Under this scenario, there are four possible outcomes:

Possible Outcomes	Probability
Neither Building is destroyed	$0.9 \times 0.9 = 0.81$
First building is destroyed and second building is intact	$0.1 \times 0.9 = 0.09$
First building is intact and second building is destroyed	$0.9 \times 0.1 = 0.09$
Both buildings are destroyed	$0.1 \times 0.1 = 0.01$

If neither building is destroyed, the loss for each owner is \$0. If one building is destroyed, each owner pays \$25,000. If both buildings are destroyed, each owner must pay \$50,000. The expected loss for each owner remains \$5,000 as shown below:

$$\text{Expected loss} = 0.81 \times \$0 + 0.09 \times \$25000 + 0.09 \times \$25000 + 0.01 \times \$50000 = \$5000$$

Note that while the expected loss remains the same, the probability of the extreme values, \$0 and \$50,000, have declined. The reduced probability of the extreme values is reflected in a lower standard deviation as shown below:

$$\text{SD} = \sqrt{0.81(0 - 5000)^2 + 0.09(25000 - 5000)^2 + 0.09(25000 - 5000)^2 + 0.01(50000 - 5000)^2}$$

$$\text{SD} = \$10607$$

Thus, as additional individuals are added to the pooling arrangement, the standard deviation continues to decline while the expected value of the loss remains unchanged. For example, with a pool of 100 insureds, the standard deviation is \$1,500; with a pool of 1,000 insureds, the standard deviation is \$474; and with a pool of 10,000, the standard deviation is \$150.

In addition, by pooling or combining the loss experience of a large number of exposure units, an insurer may be able to predict future losses with greater accuracy. From the viewpoint of the insurer, if future losses can be predicted, objective risk is reduced. Thus, another characteristic often found in many lines of insurance is risk reduction based on the law of large numbers.

The law of large numbers states that the greater the number of exposures, the more closely will the actual results approach the probable results that are expected from an infinite number of exposures. For example, if you flip a balanced coin into the air, the a priori probability of getting “heads” is 0.5. If you flip the coin only 10 times, you may get heads eight times. Although the observed probability of getting heads is 0.8, the true probability is still 0.5. If the coin were flipped 1 million times, however, the actual number of heads would be approximately 500,000. Thus, as the number of random tosses increases, the actual results approach the expected results.

4.2.2 Payment of Fortuitous Losses

A second characteristic of private insurance is the payment of fortuitous losses. Most insurance policies exclude intentional losses. *A fortuitous loss is one that is unforeseen and unexpected by the insured and occurs as a result of chance.* In other words, the loss must be accidental. The law of large numbers is based on the assumption that losses are accidental and occur randomly. For example, a person may slip on an icy sidewalk and break a leg. The loss would be fortuitous.

4.2.3 Risk Transfer

Risk transfer is another essential element of insurance. With the exception of self-insurance, a true insurance plan always involves risk transfer. *Risk*

transfer means that a pure risk is transferred from the insured to the insurer, who typically is in a stronger financial position to pay the loss than the insured. From the viewpoint of the individual, pure risks that are typically transferred to insurers include the risk of premature death, excessive longevity, poor health, disability, destruction and theft of property, and personal liability lawsuits.

4.2.4 Indemnification

A final characteristic of insurance is indemnification for losses. *Indemnification means that the insured is restored to his or her approximate financial position prior to the occurrence of the loss.* Thus, if your home burns in a fire, a homeowners policy will indemnify you or restore you to your previous position. If you are sued because of the negligent operation of an automobile, your auto liability insurance policy will pay those sums that you are legally obligated to pay. Similarly, if you become seriously disabled, a disability-income insurance policy will restore at least part of the lost wages.

4.3 Characteristics of an Ideally Insurable Risk

Private insurers generally insure only pure risks. However, some pure risks are not privately insurable. From the viewpoint of a private insurer, an insurable risk ideally should have certain characteristics. There are ideally six characteristics of an insurable risk:

4.3.1 Large Number of Exposure Units

The first requirement of an insurable risk is a large number of exposure units. Ideally, there should be a large group of roughly similar, but not necessarily identical, exposure units that are subject to the same peril or group of perils. For example, a large number of wood frame dwellings in a city can be grouped together for purposes of providing property insurance on the dwellings.

The purpose of this first requirement is to enable the insurer to predict losses based on the law of large numbers. Loss data can be compiled over time, and losses for the group as a whole can be predicted with some accuracy. The loss costs can then be spread over all insureds in the underwriting class.

4.3.2 Accidental and Unintentional Loss

A second requirement is that the loss should be accidental and unintentional; ideally, the loss should be unforeseen and unexpected by the insured

and outside of the insured's control. Thus, if an individual deliberately causes a loss, he or she should not be indemnified for the loss.

There are several reasons for this requirement. First, the loss should be accidental because the law of large numbers is based on the random occurrence of events. A deliberately caused loss is not a random event because the insured knows when the loss will occur. Thus, prediction of future experience may be highly inaccurate if a large number of intentional or nonrandom losses occur. Second, moral hazard is increased if the insured deliberately intends to cause a loss. Finally, it is poor public policy to allow insureds to collect for intentional losses.

4.3.3 Determinable and Measurable Loss

A third requirement is that the loss should be both determinable and measurable. This means the loss should be definite as to cause, time, place, and amount. Life insurance, in most cases, meets this requirement easily. The cause and time of death can usually be readily determined, and if the person is insured, the face amount of the life insurance policy is the amount paid.

Some losses, however, are difficult to determine and measure. For example, under a disability-income policy, the insurer promises to pay a monthly benefit to the disabled person if the definition of disability stated in the policy is satisfied. Some dishonest claimants may deliberately fake sickness or injury to collect from the insurer. Even if the claim is legitimate, the insurer must still determine whether the insured satisfies the definition of disability stated in the policy. Sickness and disability are highly subjective, and the same event can affect two persons quite differently. For example, two accountants who are insured under separate disability-income contracts may be injured in an auto accident, and both may be classified as totally disabled. One accountant, however, may be more determined to return to work. If that accountant undergoes rehabilitation and returns to work, the disability-income benefits will terminate. Meanwhile, the other accountant would still continue to receive disability-income benefits according to the terms of the policy. In short, it is often difficult to determine when a person is actually disabled. However, all losses ideally should be both determinable and measurable.

The basic purpose of this requirement is to enable an insurer to determine if the loss is covered under the policy, and if it is covered, how much should be paid. For example, assume that Shannon has an expensive fur coat that is insured under a homeowners policy. It makes a great deal of difference to the insurer if a thief breaks into her home and steals the coat, or the coat is missing because her husband stored it in a drycleaning establishment but

forgot to tell her. The loss is covered in the first example but not in the second.

4.3.4 No Catastrophic Loss

The fourth requirement is that ideally the loss should not be catastrophic. This means that a large proportion of exposure units should not incur losses at the same time. As we stated earlier, pooling is the essence of insurance. If most or all of the exposure units in a certain class simultaneously incur a loss, then the pooling technique breaks down and becomes unworkable. Premiums must be increased to prohibitive levels, and the insurance technique is no longer a viable arrangement by which losses of the few are spread over the entire group.

Insurers ideally wish to avoid all catastrophic losses. In reality, however, that is impossible, because catastrophic losses periodically result from floods, hurricanes, tornadoes, earthquakes, forest fires, and other natural disasters. Catastrophic losses can also result from acts of terrorism.

Several approaches are available for meeting the problem of a catastrophic loss. First, reinsurance can be used by which insurance companies are indemnified by reinsurers for catastrophic losses. Reinsurance is an arrangement by which the primary insurer that initially writes the insurance transfers to another insurer (called the reinsurer) part or all of the potential losses associated with such insurance. The reinsurer is then responsible for the payment of its share of the loss.

Second, insurers can avoid the concentration of risk by dispersing their coverage over a large geographical area. The concentration of loss exposures in a geographical area exposed to frequent floods, earthquakes, hurricanes, or other natural disasters can result in periodic catastrophic losses. If the loss exposures are geographically dispersed, the possibility of a catastrophic loss is reduced.

Finally, financial instruments are now available for dealing with catastrophic losses. These instruments include catastrophe bonds, which are designed to help fund catastrophic losses.

4.3.5 Calculable Chance of Loss

Another requirement is that the chance of loss should be calculable. The insurer must be able to calculate both the average frequency and the average severity of future losses with some accuracy. This requirement is necessary so that a proper premium can be charged that is sufficient to pay all claims

and expenses and yields a profit during the policy period.

Certain losses, however, are difficult to insure because the chance of loss cannot be accurately estimated, and the potential for a catastrophic loss is present. For example, floods, wars, and cyclical unemployment occur on an irregular basis, and prediction of the average frequency and severity of losses is difficult. Thus, without government assistance, these losses are difficult for private carriers to insure.

4.3.6 Economically Feasible Premium

A final requirement is that the premium should be economically feasible. The insured must be able to afford the premium. In addition, for the insurance to be an attractive purchase, the premiums paid must be substantially less than the face value, or amount, of the policy.

To have an economically feasible premium, the chance of loss must be relatively low. One view is that if the chance of loss exceeds 40 percent, the cost of the policy will exceed the amount that the insurer must pay under the contract. For example, an insurer could issue a \$1,000 life insurance policy on a man who is age 99, but the pure premium would be close to that amount, and an additional amount for expenses would also have to be added. The total premium would exceed the face amount of insurance.

Based on the preceding requirements, most personal risks, property risks, and liability risks can be privately insured because the ideal characteristics of an insurable risk generally can be met. In contrast, most market risks, financial risks, production risks, and political risks are difficult to insure by private insurers. These risks are speculative, and calculation of a correct premium may be difficult because the chance of loss cannot be accurately estimated. For instance, insurance that protects a retailer against loss because of a change in consumer tastes, such as a style change, generally is not available. Accurate loss data are not available. Thus, it would be difficult to calculate an accurate premium. The premium charged may or may not be adequate to pay all losses and expenses. Since private insurers are in business to make a profit, certain risks are difficult to insure because of the possibility of substantial losses.

4.4 Types of Insurance

Insurance can be classified as either private or government insurance. Private insurance includes life and health insurance as well as property and liability insurance. Government insurance includes social insurance programs and other government insurance plans.

4.4.1 Private Insurance

Life Insurance

Life insurance pays death benefits to designated beneficiaries when the insured dies. The benefits pay for funeral expenses, uninsured medical bills, estate taxes, and other expenses. The death proceeds can also provide periodic income payments to the deceased's beneficiary. Life insurers also sell annuities, individual retirement account (IRA) plans, 401(k) plans, and individual and group retirement plans. Some life insurers also sell (1) individual and group health insurance plans that cover medical expenses because of sickness or injury, (2) disability income plans that replace income lost during a period of disability, and (3) long-term care policies that cover care in nursing facilities.

Health Insurance

Although many life insurers we described also sell some type of individual or group health insurance plan, the health insurance industry overall is highly specialized and controlled by a relatively small number of insurers. These companies include Blue Cross Blue Shield Association, AETNA, United Health Group, and Well Point. Medical expense plans pay for hospital and surgical expenses, physician fees, prescription drugs, and a wide variety of additional medical costs.

Property and Liability Insurance

Property insurance indemnifies property owners against the loss or damage of real or personal property caused by various perils, such as fire, lightning, windstorm, or tornado. Liability insurance covers the insured's legal liability arising out of property damage or bodily injury to others; legal defense costs are also paid.

Property and liability insurance is also called property and casualty insurance. In practice, nonlife insurers typically use the term property and casualty insurance (rather than property and liability insurance) to describe the various coverages and operating results. Casualty insurance is a broad field of insurance that covers whatever is not covered by fire, marine, and life insurance; casualty lines include auto, liability, burglary and theft, workers compensation, and health insurance.

4.4.2 Government Insurance

Numerous government insurance programs are in operation at the present time. Social insurance is the one discussed in this chapter, whereas other

government insurance also exist but differ as per the national and state policy of a country.

Social Insurance

Social insurance programs are government insurance programs with certain characteristics that distinguish them from other government insurance plans. These programs are financed entirely or in large part by mandatory contributions from employers, employees, or both, and not primarily by the general revenues of government. The contributions are usually earmarked for special trust funds; the benefits, in turn, are paid from these funds. In addition, the right to receive benefits is ordinarily derived from or linked to the recipient's past contributions or coverage under the program; the benefits and contributions generally vary among the beneficiaries according to their prior earnings, but the benefits are heavily weighted in favor of low-income groups. Moreover, most social insurance programs are compulsory. Covered workers and employers are required by law to pay contributions and participate in the programs. Finally, eligibility requirements and benefit rights are usually prescribed exactly by statute, leaving little room for administrative discretion in the award of benefits. %

4.5 Readings from Insurance Regulation, 2049 (1993)

The following are the readings extracted from "*Insurance Regulation, 2049*".

4.5.1 Definitions

Unless the subject or context otherwise requires, in this Regulation :

- "Act" means the Insurance Act, 2049
- "Certificate" means an Insurer Registration Certificate provided to an Insurer for operating the Insurance Business by registering him as an Insurer pursuant to Rule 8
- "Premium" means the Insurance Premium to be collected by the Insurer from the insured in consideration of the Insurance Business
- "Advisory Committee" means the Insurance Tariff Advisory Committee constituted pursuant to section 41 of the Act

4.5.2 Categories of Insurance Business

Subject to the provisions made in the Act and this Regulation, the Insurance Business to be operated by an Insurer shall be divided into the following categories :

- Life Insurance Business
- Non-Life Insurance Business
- Re-Insurance Business

Notwithstanding anything contained in sub-rule (1), Nepal Government may prescribe other categories of Insurance Business as required on the advise of the Board.

Life Insurance Business

1. The Insurer may operate the following Insurance Business under the Life Insurance Business:
 - (a) Whole Life Insurance
 - (b) Endowment Life Insurance
 - (c) Term Life Insurance
2. Notwithstanding anything contained in sub-rule (1), the Board may prescribe other categories of the Life Insurance Business as required.
3. The conditions and privileges of the Life Insurance Policy to be executed pursuant to this Rule shall be as specified by the Board.

Non-Life Insurance Business

1. The Insurer may operate the following Insurance Business under the Non-Life Insurance Business :
 - (a) Fire Insurance
 - (b) Motor Insurance
 - (c) Marine Insurance
 - (d) Engineering and Contractor's Risk Insurance
 - (e) Aviation Insurance
 - (f) Miscellaneous Insurance
2. Notwithstanding anything contained in sub-rule (1), the Board may prescribe other categories of Non-Life Insurance Business as required.
3. The conditions and privileges of the Non-Life Insurance Policy to be executed pursuant to this rule shall be as specified by the Board.

Re-Insurance Business

1. The Insurer may re-insure the risks which are in excess from the risks assumed by it.
2. The Categories of Re-insurance Business to be made pursuant to sub-rule (1) and other arrangement shall be as specified by the Board

4.5.3 Provisions Relating to a Surveyor

Application to be submitted for surveyor's license

Any person or any corporate body having a qualification, as mentioned in regulation, and desirous to work as a Surveyor pursuant to sub-section (1) of Section # 30A of the Act, shall submit an application to the office of the Board in the format of Schedule -12.

Surveyor's License to be provided

1. After receiving an application for the Surveyor's license pursuant to rule-26, the Board shall make an inquiry whether the applicant is qualified or not pursuant to Rule 28, and if it deems appropriate to provide the Surveyor's license to him it shall register his name as a Surveyor in the registration-book pursuant to Schedule - 13.
2. After making the registration of the name of the applicant on the registration-book pursuant to sub-rule (1), the Board shall provide a Surveyor's license to the applicant in the format of Schedule - 14 by collecting from the applicant a fee of twelve thousand Rupees for the class A license, that of nine thousand Rupees for the class B license, that of seven thousand rupees for the class C license and that of five thousand Rupees for the class D license.

Qualifications of a Surveyor

1. An applicant desirous of making an application for the Surveyor's License pursuant to Rule 26 shall have possessed any one of the following qualifications :
 - (a) Having gained at least ten years of work experiences on the Insurance Business, holding an officer level post at the office of any Insurer
 - (b) Having possessed at least a Bachelor Degree in Engineering subject
 - (c) Having possessed at least a Bachelor Degree in Insurance subject from a Chartered Insurance Institute of international standard or from an organization recognized by such institute, or

- (d) Having passed the Chartered Accountancy Examination
- 2. An applicant having possessed the qualification as referred to in clause (a) or (b) or (c) or (d) of sub-rule (1) shall, prior to obtaining the Surveyor's License, have obtained a certificate indicating his participation in, and completion of, the surveyor training conducted by the Board

Classification of the Surveyor

The Surveyors who have been working as Surveyors after having obtained the Surveyor's license and completed the following period shall be classified as follows and provided with the Surveyor's License, after the commencement of this Regulation:

- A Surveyor who has regularly worked as a Surveyor for a period more than fifteen years, Class "A"
- A Surveyor who has regularly worked as a Surveyor for a period from ten to fifteen years, Class "B"
- A Surveyor who has regularly worked as a Surveyor for a period From five to ten years, Class "C"
- A Surveyor who has regularly worked as a Surveyor for a period of five years, Class "D"

Provisions relating to the renewal of Surveyor's License

1. The Surveyor shall submit an application to the office of the Board in the format of Schedule-15 along with the renewal fee of →twelve thousand Rupees for "A" class license, nine thousand Rupees for "B" class license, seven thousand Rupees for "C" class license and five thousand Rupees for "D" class license within the time-limit pursuant to subsection (1) of Section 31 of the Act for the renewal of the License. On receipt of such application, the Board shall renew the Surveyor's license
2. If a Surveyor has submitted an application to the office of the Board stating the reasons for inability to submit an application for the renewal of his license within the time-limit pursuant to sub-rule (1), and if the reasons are found to be appropriate, the Board may renew the license by receiving an additional fee of five hundred Rupees for first two months from the date of expiry of the renewal time-limit and after that fifty Rupees per day for up to four months

Limitation of Survey

1. The limitation relating to survey which a surveyor of each category pursuant to Rule 28A can make shall be as prescribed by the Board
2. The Code of Conduct of the Surveyor shall be as prescribed by the Board

In the case of cancellation of Surveyor's License

If the Surveyor's license is cancelled pursuant to Section 33 of the Act, no other Surveyor's license shall be provided to him to work as a Surveyor up to a period of five years from the date of such cancellation.

4.5.4 Provisions relating to payment against Insurance Claim

Process against the payment of Life Insurance Claim

1. The Insurer shall issue a discharge voucher in the name of the Insured who has already paid the last installment of the Life Insurance Premium requesting him to come to collect payment against the claim along with the Insurance Policy and other documents required for making payment against such Life Insurance claim within fifteen days from the date of payment of such installment
2. In case an Insured submits the Insurance Policy and other documents including the discharge voucher to the Insurer for the payment of claim against the Life Insurance claim pursuant to sub-rule (1), the Insurer shall conduct an inquiry as required and make a payment against the Life Insurance claim within seven days from the date of expiry of the period of the Life Insurance Policy
3. In case any person who has taken up an Insurance Policy dies before the expiry of the period of the Insurance Policy, the person designated by him, if any, and in case no person has been designated, the nearest heir from among the persons mentioned in sub-section (1) of Section 38 of the Act shall submit an application for the payment against the claim to the Insurer to receive the amount of the Life Insurance stating the details as follows :
 - (a) The details relating to the claim
 - (b) A Certificate of death of the insured
 - (c) In case the insured has died in an accident and if such risk is covered by the Life Insurance, the postmortem report of the government physician relating to the cause of death, and if there is no such report, a report of the police

- (d) A certificate of relationship with the insured
 - (e) The documents regarding the certification of the age in case the age has not been certified
 - (f) Other details specified by the Board
4. After the receipt of the application pursuant to sub-rule (3), the Insurer shall make an inquiry into the details including the documents submitted regarding to the claim of Life Insurance, and shall examine other matters also if necessary, and shall determine the liability within fifteen days from the date of receipt of such documents by it and shall issue the discharge voucher in the name of the applicant requesting him to come to collect the payment against the claim. The Insurer shall make the payment against the Insurance claim within fifteen days from the date of receipt of the discharge voucher from the applicant
 5. If it is found, while making an inquiry into the details pursuant to sub-rule (4) that the Insurance claim need not to be paid by determining the liability, the Insurer shall provide a written information to the applicant clearly stating the reasons thereof.

Process of payment against Non-Life Insurance Claim

1. If any claim has to be made under the Insurance Policy by an Insured who has taken up a Non-Life Insurance Policy, the Insured shall submit an application to the Insurer stating all the details relating to it
2. On receipt of an application of the Insured for the payment against the Insurance claim of the Non-Life Insurance Pursuant to sub-rule (1), the Insurer shall immediately designate a Surveyor to make necessary inquiry, if necessary
3. The Surveyor deputed pursuant to sub-rule (2) shall make necessary inquiry and shall determine the liability of the Insurer within fifteen days and shall submit a report to the Insurer including the comprehensive details and also inform the Insured relating to it mentioning the amount to be received by the Insured subject to the terms and conditions and facilities of the Insurance Policy
4. The Insurer shall determine the liability and shall provide the payment against the claim of the Non-Life Insurance to the Insured generally within thirty-five days from the submission of the report by the surveyor pursuant to sub-rule (3).

4.6 Review Problems

Chapter 5

Passenger Transport Operation

5.1 Public Transportation

Public transportation is a generic term used to describe the family of transit services available to urban and rural residents. Thus, it is not a single mode but a variety of traditional and innovative services, which should complement each other to provide system-wide mobility.

5.1.1 Transit Modes

The modes included within the realm of public transportation are:

- **Mass transit**, characterized by fixed routes, published schedules, designated networks, and specified stops. Mass-transit vehicles include buses, light rail (trolleys) or rapid transit that either share space in mixed traffic or operate on grade-separated rights of way.
- **Paratransit** is characterized by flexible and personalized service intended to replace conventional fixed-route, fixed-schedule mass-transit lines. Paratransit is available to the public on demand, by subscription, or on a shared-ride basis. Examples include taxi, car rental, dial-a-ride, and specialized services for elderly, medical, and other designated users.
- **Ridesharing** (as the name implies) is characterized by two or more persons traveling together by prearrangement, such as carpool, vanpool, or shared-ride taxi.

5.1.2 Transit Capacity and Level of Service

A basic attribute of any transit mode is its carrying capacity, defined as the number of vehicles or persons that pass a given point in a specified time

(usually an hour). The numerical value of carrying capacity (usually referred simply as capacity), is dependent on two variables: (1) the number of vehicles that pass a point at a given time and (2) the number of passengers within each vehicle. For example, if for a given lane along a section of highway there are 60 buses that pass by in an hour (or one per minute), and each bus carries 50 seated passengers, then the carrying capacity of this highway lane is 60 buses/ln/hr or $50 \times 60 = 3000$ passengers/ln/hr.

Carrying capacity is influenced by (1) the “spacing” in seconds between each vehicle (called the headway) and (2) the “comfort factor” experienced by passengers (called the level of service). Thus, carrying capacity can be increased in two ways: (1) reduce the headway or (2) increase the number of passengers per vehicle. In the bus capacity example, the headway was 60 seconds and the level of service was that all passengers had a seat. Time spacing between buses could possibly be reduced, but there are limits to lowering headway values dictated by safe distance requirements between vehicles and/or the time spent at transit stops and terminals (called dwell time). Similarly, passenger loading could be increased by allowing standees, but this would decrease the comfort level for passengers. Were the bus equipped with computer tables and a refreshment area (thus offering a higher level of service), fewer passengers could be accommodated resulting in a lower carrying capacity but a higher level of service.

Accordingly, when reporting transit capacity, it is important to specify the units as either vehicles or passengers/hour and the corresponding level of service in terms of passengers/vehicle. Public transit is often compared with the automobile when issues of carrying capacity are involved, as it is commonly believed that transit capacity is superior to auto capacity. As will be discussed in greater detail in Chapters 9 and 10, the capacity of a single lane of passenger vehicles is approximately 2,000 vehicles/hour which represents a headway of 1.8 seconds. Since most cars have at least five seats, the person capacity of a highway lane could be as great as $5 \times 2000 = 10000$ per/hour. Capacities of this magnitude never have been achieved, since most cars carry only one person with an average car occupancy of about 1.5. Why is this so? Have you ever driven in a car carrying five people? Not a pleasant experience, and a reason why car pooling is not very popular. Given the opportunity, most people choose to drive alone or with just one other person.

Travelers usually consider many more factors than simply the in-vehicle level of service, and they don't really consider how they can contribute to increasing “carrying capacity.” In fact, if drivers were to optimize the carrying capacity of a highway, they would all drive at 35 miles per hour! Other major considerations in selecting the travel mode include: reliability, punctuality, cost, travel time, and safety. Transit systems that receive “high

marks” for the out-of-vehicle level-of-service factors are typically the ones that use exclusive lanes or tracks with no interference from other vehicles or pedestrians and have adequate capacity at station stops and terminals. Thus, rapid transit services (whether bus or fixed guideway), are the superior mode but are more costly to build and maintain and require high volumes of demand to be feasible.

5.1.3 The Role and Future of Public Transportation

Public transportation is an important element of the total transportation services provided within large and small metropolitan areas. A major advantage of public transportation is that it can provide high-capacity, energy-efficient movement in densely traveled corridors. It also serves medium- and low-density areas by offering an option for auto owners who do not wish to drive and an essential service to those without access to automobiles, such as school children, senior citizens, single-auto families, and others who may be economically or physically disadvantaged.

For most of this century, public transportation was provided by the private sector. However, increases in auto ownership, shifts in living patterns to low-density suburbs, and the relocation of industry and commerce away from the central city, along with changes in lifestyle (which have been occurring since the end of World War II) have resulted in a steady decline in transit ridership. Since the early 1960s, most transit services have been provided by the public sector. Income from fares no longer represent the principal source of revenue, and over a 25- to 30-year period, the proportion of funds for transit provided by federal, state, and local governments has increased steadily. While it generally is believed that highways and motor transport will play a dominant role in providing personal transportation in the beginning decades of the twenty-first century, there are many unforeseen changes that could alter the balance between public and private transportation. Some could contribute to a decline in transit ridership while others might cause transit to become stronger, and for the remainder, there would be little or no effect. The potential changes that could influence transit usage are categorized here from the book *Urban Mass Transportation Planning*.

Factors Bad for Transit

- Growth of suburbs
- Industry and employment moving from the central city
- Increased suburb-to-suburb commuting
- Growth in private vehicle ownership

- Increased diversity in vehicle types such as SUVs, pickup trucks, and RVs
- High labor costs

Factors Good for Transit

- Emphasis by the government on air quality
- Higher prices of gasoline
- Depletion of energy resources
- Trends toward higher-density living
- Legislation to encourage “livable cities” and “smart growth”
- Increased number of people who cannot or choose not to drive

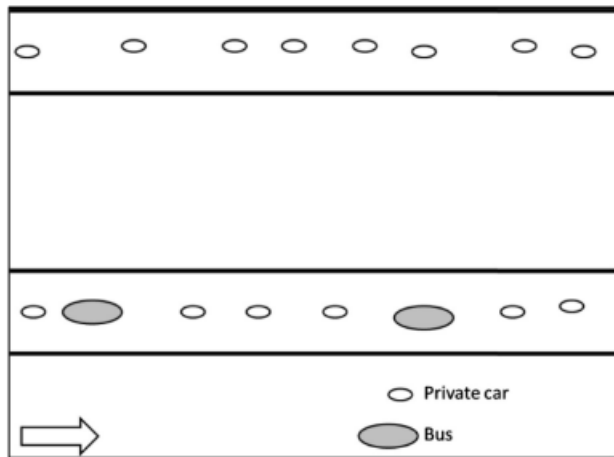
5.2 Number of Passengers Versus Number of Served Vehicles

Traffic engineers try by various techniques and measures to enable as many vehicles as possible to pass, during a specific period of time, through a traffic intersection or through a road section. In other words, traffic experts try to maximize the number of vehicles that are served during a certain period of time. On the other hand, when it comes to public transit, we are trying to maximize the number of transported passengers. The number of passengers that can be served during the observed time interval represents transit line capacity. Similarly, the number of vehicles that can be served during the observed time interval represents the vehicle line capacity.

5.2.1 A Hypothetical Case Study

Let us assume that the freeway lane can serve 2200 vehicles per hour (the freeway lane capacity equals 2200 vehicles per hour). The upper part of the figure shows the situation when only private cars used freeway lane. The lower part of the figure refers to the situation when some buses also participate in freeway lane traffic. We assume that average number of passengers in bus and private car (average occupancies) are, respectively, equal to 50 and 1.4. We also assume that instead on any two private cars we can allow one bus to participate in freeway lane traffic.

5.2. NUMBER OF PASSENGERS VERSUS NUMBER OF SERVED VEHICLES75

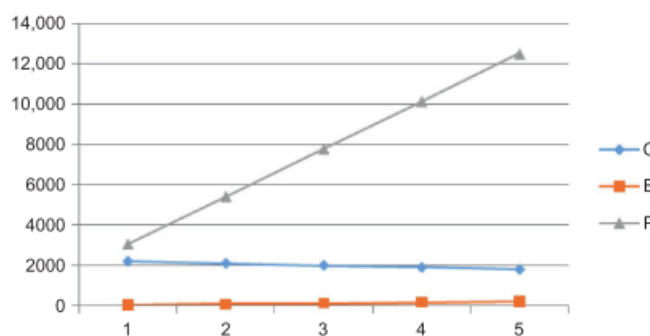


We denote by C , B , and P , respectively, the total number of cars, total number of buses, and total number of transported passengers. If we allow, for example, 50 buses to enter the freeway lane, we need to reduce the total number of private cars for $50 \times 2 = 100$ private cars. In this case, the total number of transported passengers P by private cars and buses equals:

$$P = C \times 1.4 + B \times 50 = 2100 \times 1.4 + 50 \times 50 = 5440 \text{ Passengers}$$

The table and the graph below show increase in the total number of transported passengers with the increase of the number of buses engaged.

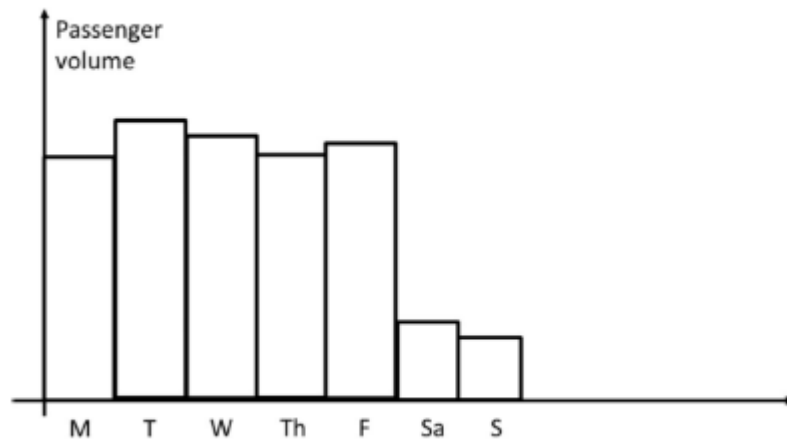
Number of Cars per Hour (C)	Number of Buses per Hour (B)	Number of Passengers per Hour (P)
2200	0	3080
2100	50	5440
2000	100	7800
1900	150	10,160
1800	200	12,520



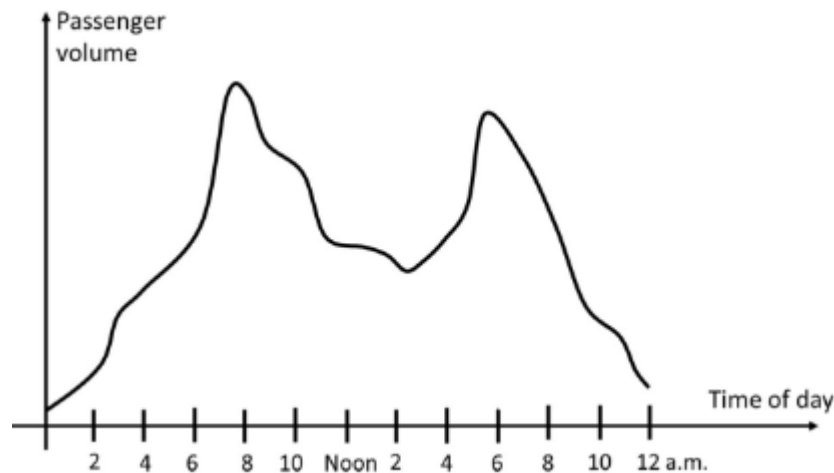
The graph above illustrates the effects that public transit can achieve. With a very small percentage share of public transport vehicles in the total number of vehicles, it is possible to increase dramatically the total number of passengers carried. To achieve such effects in reality, it is necessary constantly to increase the public transit attractiveness.

5.3 Passenger Flows in Public Transportation

Passenger flows vary considerably in public transportation. During the working day, in most cities in the world, the daily numbers of passengers on most transit lines are relatively uniform. This is due to the fact that many passengers, each working day, use the same transportation mode and the same route when going to work. During a weekend the number of passengers in public transit is significantly smaller. The figure below shows daily variations in passenger volume.

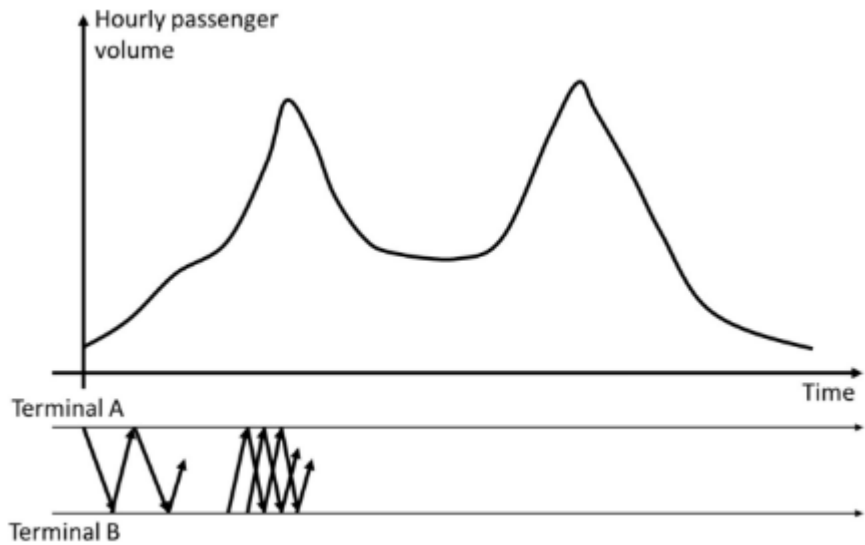


The fig below shows hourly variations in passenger volume. The variations shown in figure are typical for many cities in the world. There are morning and evening peaks when people go, and when people come back from work. The differences in hourly passenger volumes could be very high by hours of a day.



Hourly variations in passenger volume have, as a consequence, different number of vehicle departures from the terminals during certain time intervals as shown in the figure below. In this way, hourly variations in passenger

volume require the engagement of different number of vehicles during certain time periods. The number of engaged vehicles is much higher during the rush-hours. Outside the peak periods the transit operator has a surplus of vehicles and drivers. The transit operator, therefore, meets with a range of organizational problems that have to be solved (“empty” vehicle trips to garage and from garage, drivers working hours divided in two shifts, vehicle maintenance planning, etc.).

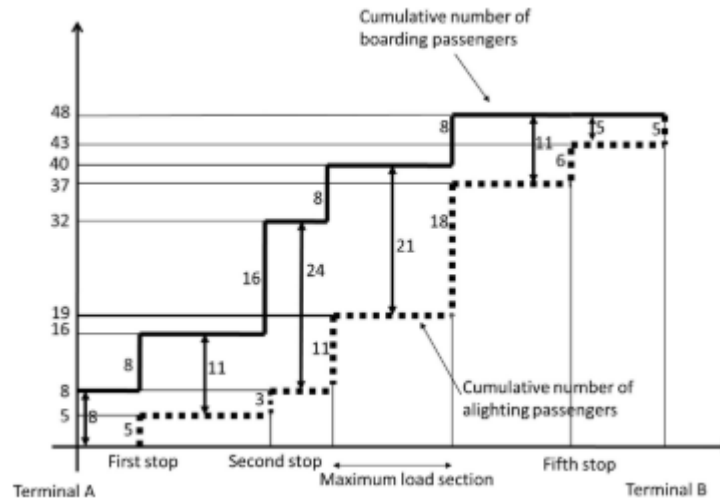


5.4 Passenger Flows Along a Transit Line

Headway in public transportation operations represents the time interval between vehicles past a specific point as shown in the fig below. Headways are expressed in minutes. It is essential to study passenger flow along the transit line, in order to determine the appropriate transit line headway.

Let us assume that we have data on the number of boarding passengers and number of alighting passengers on individual line stops and terminals (end stations on a transit line) as shown in the table below. The transit line has 5 stations. The terminals are denoted respectively by A and B.

The last column of the table shows the numbers of passengers in the vehicle, after departing from bus stops. Thus, for example, after leaving the station 2, there are 21 passengers in the vehicle that travels along line section between stop 3 and stop 4. The number of boarding passengers, number of alighting passengers and the number of passengers in the vehicle for any line section are shown in the figure below.

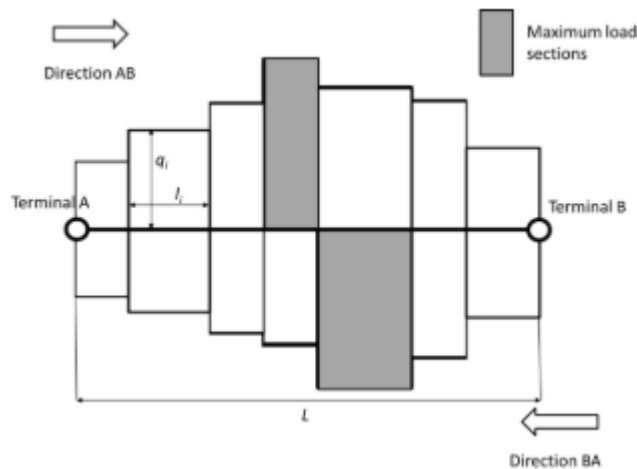


Bus Stop	Number of boarding Passengers	Number of Alighting Passengers	Number of Passengers in the Vehicle
Terminal A	8	0	8
1	8	5	11
2	16	3	24
3	8	11	21
4	8	18	11
5	0	6	5
Terminal B	0	5	0

The numbers of passengers on certain line sections respectively equal 8, 11, 24, 21, 11, and 5. The maximum passenger volume equals $\max 8, 11, 24, 21, 11, 5 = 24$ and corresponds to the section between stop 2 and stop 3. We call section between stop 2 and stop 3 the maximum load section (MLS).

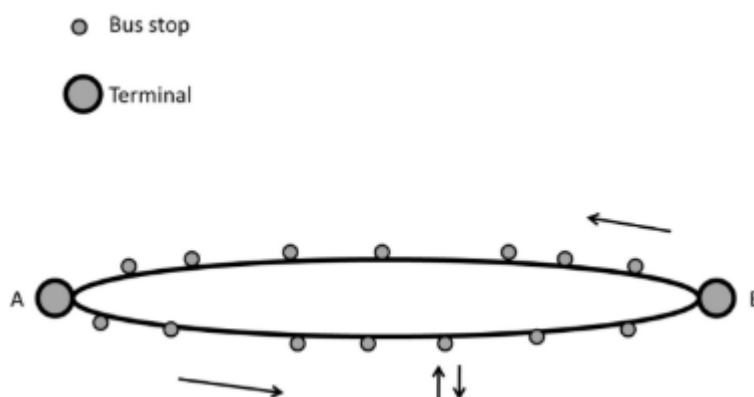
The data shown in the table and figure above are related to one vehicle trip. In real-life applications, 1 h is the basic time unit that is used when describing cumulative number of alighting and boarding passengers, as well as the maximum passenger volume on MLS. In other words, the maximum passenger volume is expressed in passengers per hour. The passenger volume profile is calculated for both directions of the transit line.

In the figure shown below, the line length is denoted by L . This length represents the one-way distance between the line terminals along the line alignment. The line length is measured in miles or kilometers. We denote by q_i passenger volume on the i^{th} section. The length of the i^{th} transit section is denoted by l_i . The MLS in one direction is usually different from the MLS in another direction. The maximum passenger volumes in both directions should be taken into account when determining the number of vehicles to be engaged on the transit line.



5.5 Service Frequency and Headways

One of the most important problems encountered by public transit operators is how to match transportation supply and passenger demand on individual transit line. The matching problem is far more complex over the entire route network than on individual routes. Service frequencies and vehicle departure times on transit lines in the network reflect the manner in which transportation supply and passenger demand are matched. Service frequencies and vehicle departure times depend on passenger volume profile and on the number and type of vehicles in the fleet. The number of passengers that decide in the end to use public transit on a particular transit line depends, to the highest degree, on service frequency and vehicle departure times. For example, if service frequency is low or if vehicle departure times during the day are not convenient, a number of potential passengers will instead choose other modes of transportation.



Let us note the bus line shown in the figure above. Vehicles move from

Terminal A to Terminal B. On the way to the Terminal B, vehicles stop at pre-defined bus-stops, where passengers enter and exit the vehicle. On arrival at the Terminal B, the driver rests for a while, and then the vehicle travels to Terminal A. On the way to Terminal A, car stops at bus-stops where passengers enter and exit the vehicle.

We denote by T turnaround time. This time is the time that elapses from the moment when vehicle leaves Terminal A to the moment when vehicle returns to Terminal A. Let us assume that we have on our disposal N vehicles that we can assign to the bus line. The service frequency represents the number of vehicles per time unit past a specific point in the same direction. The frequency equals:

$$f = \frac{N}{T} \quad (5.1)$$

The frequency is expressed in the number of vehicle per hour. Headway h in public transportation operations represents the time interval between vehicles past a specific point. Headways are expressed in minutes. Since the frequency represents the number of vehicles per time unit past a specific point in the same direction, we conclude that the frequency is the inverse of the headway, ie:

$$h = \frac{1}{f} \quad (5.2)$$

When calculating and rounding headways, it is desirable to obtain a so called clock headway. Clock headways have a feature that enables the generation of timetable that is repeated every hour, starting on the hour. Thus, for example, in the case when the headway is equal to 15 min, it is possible to have the vehicle departures from the terminal in 8:00, 8:15, 8:30, 8:45, 9:00, 9:15, 9:30, 9:45, 10:00, etc.

5.5.1 The Maximum Service Frequency

The maximum service frequency is defined by the maximum number of transit vehicles passing through a given point of a line/route i in one direction during a given period of time (usually 1 h) under prevailing operating conditions. This can be estimated as follows:

$$f_{max/i}(\tau) = \frac{\tau}{h_{min/i}} \quad (5.3)$$

Where,

τ is the given period of time

$h_{min/i}$ is the minimum headway, ie, the time interval between the successive transit vehicles passing through a given point of a line/route i in the same direction (min).

The minimum headway $h_{min/i}$ in Eq. 5.3 can be determined according to different criteria, but in many cases the prevailing factors are characteristics of the system such as technology and way of operations along the line and at the stations. These factors influence the minimum headway for a given line/ route and for the stops/stations along it. Most frequently the stop/station headway is as much greater than that of the line(s)/route(s). In addition, the headway $h_{s/i}$ at the stop/station on a given line/route i , should be greater than the vehicle stop time $t_{s/i}$ at that stop/station. Consequently, the “ultimate” capacity of this stop/station will be:

$$C_{ss/i}(\tau) = \frac{\tau}{\max[h_{s/i}, t_{s/i}]} \quad (5.4)$$

In addition, the following must be satisfied for all stops/stations on the line/route i :

$$f_{max/i}(\tau) \leq C_{ss/i} \quad (5.5)$$

In other words, the maximum number of transit vehicles passing through a given stops/stations cannot exceed the “ultimate” capacity of any of these stops/stations.

5.5.2 Passenger Waiting Time

Passengers’ walk to a stop and passenger waiting time are the basic attributes of the public transit level of service. Walk to stop in the range of 400–800 m is considered acceptable for public transit users.

In order to estimate the average passenger waiting time, let us first consider the situation when the bus arrives at the bus stop regularly, according to the published timetable. We also assume that all passengers at the bus stop can enter the vehicle, and that the passengers appear at the bus stop in random moments of time. It has been shown that, in this case, the average waiting time per passenger at the station w is equal to the one half of the vehicle headway h , ie:

$$w = \frac{h}{2} \quad (5.6)$$

The average waiting time per passenger could be longer in the case of irregular bus arrivals. In the case of irregular bus arrivals, vehicle headway is not any more deterministic quantity. In this case, the vehicle headway is a random variable. If the planned headway equals, 10 min, in the case of irregular arrivals headway values could be, for example, 8, 9, 12, 15,... minutes. It has been shown, that in the case of irregular vehicle arrivals at the stop, the average passenger waiting time equals:

$$E(W) = \frac{E(H)}{2} + \frac{\text{var}(H)}{2 \times E(H)} \quad (5.7)$$

Where, $E(H)$ is the expected value of the random variable H ; and $var(H)$ is the variance of the random variable H (variance represents the square of standard deviation).

5.5.3 Headway Determination by "Square Root Formula"

Transit operator cost and passenger cost depend on chosen headway. Passenger cost, in the case when vehicles arrive regularly, has linear increase with headway. The greater the headway, the greater the passenger waiting time and passenger cost. On the other hand, greater headway means for transit operator smaller number of departures and lower costs.

Z is the total cost per hour;

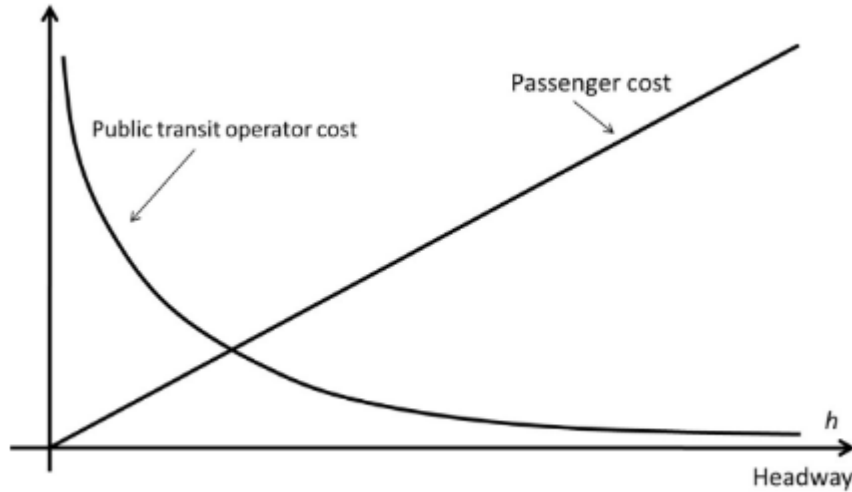
c is the transit operator cost per bus hour;

ν is the value of passenger waiting time per hour;

r is the total number of passengers on line per hour (ridership per hour);

N is the number of vehicles assigned to the bus line; and

h is the headway.



The transit operator cost per hour is equal to $N \times c$. We assume regular vehicle arrivals. In this case, the average waiting time per passenger at the station w is equal to the one half of the vehicle headway h , ie:

$$w = \frac{h}{2} \quad (5.8)$$

The waiting cost of all passengers is equal to the $\nu \times r \times \frac{h}{2}$. The total cost is:

$$Z = N \times c + \nu \times r \times \frac{h}{2} \quad (5.9)$$

Since, $N = \frac{T}{h}$, we can write,

$$Z = \frac{T}{h} \times c + \nu \times r \times \frac{h}{2} \quad (5.10)$$

The optimal headway is found by setting the derivative of Z with respect to h equal to zero:

$$\frac{dZ}{dh} = -c \times \frac{T}{h^2} + \frac{\nu \times r}{2} = 0 \quad (5.11)$$

Therefore the Optimum Headway,

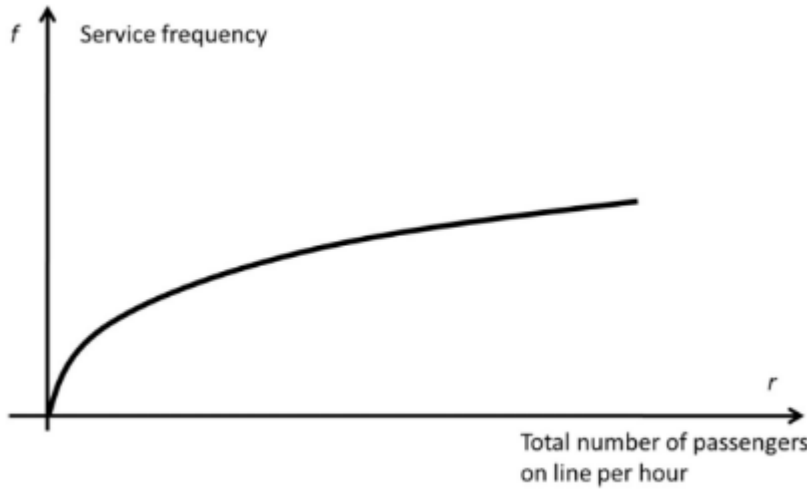
$$h = \sqrt{\frac{2 \times c \times T}{\nu \times r}} \quad (5.12)$$

Eq. 5.12 represents the “square root formula” for optimizing headway and service frequency. Minimal headway values in real-life are usually between 2 and 3 min. Maximal headways values are between 15 and 30 min. Outside of peak periods, on some transit lines, maximum headway values reach 60 min.

The service frequency f is the inverse of the headway, ie:

$$f = \frac{1}{h} = \sqrt{\frac{\nu \times r}{2 \times c \times T}} \quad (5.13)$$

Service frequency dependence of the total number of passengers on line per hour is shown in the figure below.



Numerical Example:

Let us assume the following line parameters: $c = 120\$$ per hour, $\nu = 10\$$ per passenger hour, $r = 1200$ passengers per hour, and $T = 1.5$ hr. Calculate the optimal headway.

Solution:

$$h = \sqrt{\frac{2 \times c \times T}{\nu \times r}} = \sqrt{\frac{2 \times 120 \times 1.5}{10 \times 1200}} = 0.173 \text{ hr}$$

$$\therefore h \approx 10 \text{ min}$$

5.5.4 Headway Determination by Maximum Load Method

When determining headways, transit operators try to provide enough space (especially during peak hours) to meet passenger demand. Majority of transit operators also define maximum headways on transit lines. For example, operator could define that maximum headway on a specific route, is equal to 30 min. Maximum headways guarantee a minimum service frequency offered to the passengers. The prescribed maximum headway is usually called *policy headway* and denoted by h_p .

The Maximum load method (that can have few variations) is based on counting passengers on the transit stop that is at the beginning of the MLS. Depending on the time of a day, the location of the MLS could change. For example, the transit stop, that is at the beginning of the MLS, could be Stop #5 between 10:00 am and 11:00 am, while Stop #7 could be at the beginning of the MLS between 5:00 pm and 6:00 pm. Frequently, passenger counting is performed at the bus stop that has the highest daily passenger volume. The location of this transit stop is usually well known to the transit operator. The counting interval (whole day, between 7:00 am and 10:00 am, between 4:00 pm and 7:00 pm, etc.) is different for different transit operators and different cities.

We denote by P_{max} the average value of the maximum daily passenger volume. For example, transit operator monitored during seven days period the daily number of passengers that departed from the station # 6. The following 7 values were recorded: 1262, 1348, 1439, 1285, 1290, 1391, and 1287. The P_{max} , in this case is equal to:

$$P_{max} = \frac{1262 + 1348 + 1439 + 1285 + 1290 + 1391 + 1287}{7} = 1329$$

The service frequency, f , that should be offered in order to satisfy maximum passenger volume and desired vehicle occupancy is equal to:

$$f = \frac{P_{max}}{\alpha \times C_{car}} \quad (5.14)$$

Where, C_{car} is maximum number of passengers per car; and $0 \leq \alpha \leq 1$ is load factor.

The Load factor α is related to the concept of desired vehicle occupancy.

The product $\alpha \times C_{car}$ defines the desired vehicle occupancy during the observed time period.

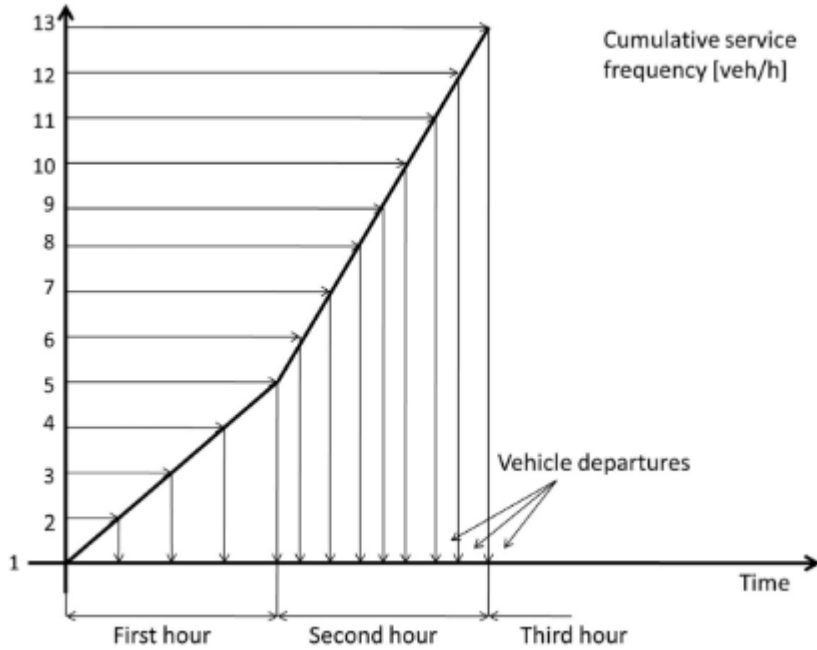
The corresponding headway is equal to:

$$h = \frac{1}{f} = \frac{\alpha \times C_{car}}{P_{max}} \quad (5.15)$$

5.6 Time Table

The transit line timetable is generated at one point (usually terminal). In the next step, by using information about average travel times between transit stops, timetable is generated for all transit stops. The transit operator's timetable contains information about vehicle departure times at all transit stops.

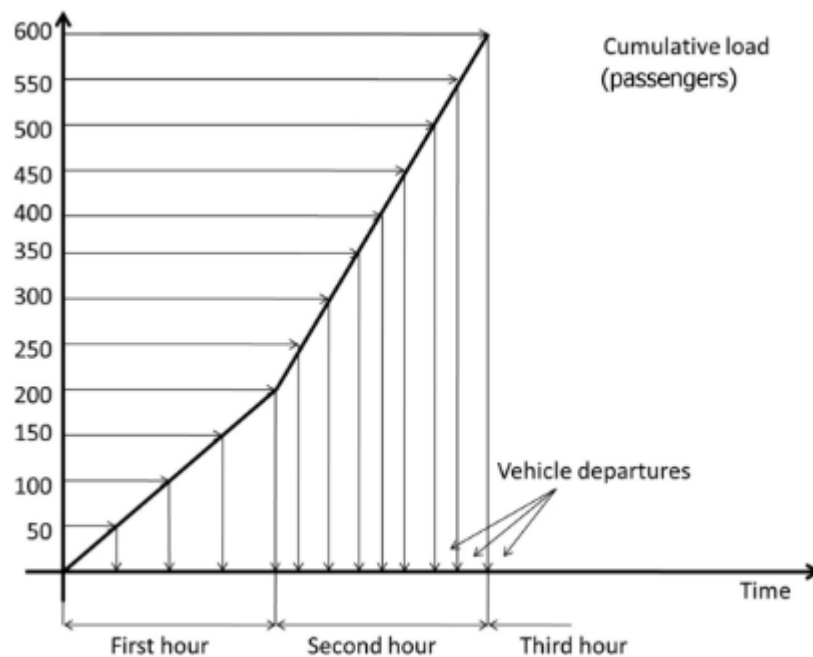
ansit operator's timetable contains information about vehicle departure times at all transit stops. The basic input data for the timetable construction represent service frequency values in certain intervals of time during the day. Timetable can easily be generated in the following way. Let the abscissa represents time. Based on the known values of service frequency in certain intervals of time, we draw a cumulative frequency.



The cumulative frequency shown in the figure above is related to the case when service frequency within first hour is equal to 4, and service frequency within second hour is equal to 8. Let us first vehicle departure happen at the

beginning of the first hour. We can go horizontally for every vehicle departure, until intersecting the cumulative curve. From the point of intersection we go vertically. The intersection of this vertical line and abscissa represents vehicle departure time. In this way, we generate timetable at specific point. In this way, we generate constant headways during specific time intervals.

Because of demand variability, constant headways frequently do not generate equal number of passengers at every vehicle departure. In order to achieve even-load at every vehicle departure, we draw cumulative loads on the transit stop that is at the beginning of the MLS.



Unequal number of passengers at specific vehicle departures can cause overcrowding in some vehicles, long boarding time at some transit stops, “bunching” of vehicles, and a decrease in the level of service offered to the passengers. On the other hand, equal load is related to unequal headways which could be inconvenient for passengers.

Trips between nodes in public transit networks may be made with or without making transfers. Transfers generally cause inconvenience to passengers. Given that inadequately coordinated transfers can increase waiting times considerably, it is particularly important (when constructing timetables) to synchronize schedules cautiously in cases of larger headways. Unsuccessfully coordinated transfers can also reduce the number of passengers using public transit as a result of switching to competitor modes. When designing synchronized schedules, it is essential to try to minimize the total waiting times

of all passengers at transfer nodes in a transit network.

5.7 Transit Line Capacity

The transit line is the basic element of public transit system. The transit route lengths in one direction are usually between 40 and 90 min, while stop spacing in urban areas are in the range from 120 to 400 m. It is desirable that transit route intersects few other transit routes. In this way, transfer points are generated that enable passengers to create various itineraries when making a trip. The waiting time at transfer points, which is less than 8 min, is considered as an acceptable waiting time. The transit line operating hours by weekdays are usually between 5:00 am and midnight. The spacing between transit lines, in the majority of cities, is in the range of 700–1000 m.

We denote by C_{car} the maximum number of passengers per car. Bus capacity represents the sum of number of seated passengers and legal standees. Public transit agencies and operators usually assume six passengers per square meter, as legal standees. In some countries, this figure could be higher.

The line/route (offered) “ultimate” capacity $C_{car}(\tau)$ is expressed by the maximum number of spaces, which can be transported in one direction during a given period of time τ (usually 1 hr) under prevailing operating conditions. In other words, the “ultimate” capacity of a public transportation lines represents the product of the maximum frequency of the service $f_{max}(\tau)$ and the maximum number of passengers in the transit vehicle.

$$C_{car}(\tau) = f_{max/i}(\tau) \times N_{car} \times C_{car}$$

The practical capacity C_{line} of a public transit line represents the product of the offered service frequency f and the maximum number of passengers in the transit vehicle.

$$C_{line} = f \times N_{car} \times C_{car} \quad (5.16)$$

Where, N_{car} is the number of cars/vehicles, and C_{car} s the maximum number of passengers per car.

The number of vehicles N_{car} per service frequency f can be variable. In the case of bus operations, $N_{car} = 1$, and the capacity of a public transportation line C_{line} equals:

$$C_{line} = f \times C_{car}(\text{spaces/hr}) \quad (5.17)$$

The turnaround time equals:

$$T = \frac{2 \times L}{u} \quad (5.18)$$

Where, L is the distance between Terminals A and B (for simplicity we assume that distance from A to B is equal to the distance from B to A); and u is the average vehicle speed.

Then,

$$C_{line} = \frac{N}{T} \times C_{car} \quad (5.19)$$

$$C_{line} = \frac{N}{\frac{2 \times L}{u}} \times C_{car} \quad (5.20)$$

$$\therefore C_{line} = \frac{N \times u \times C_{car}}{2 \times L} \quad (5.21)$$

As we can see, the line capacity C_{line} depends on the number of engaged vehicles N , the average speed u , vehicle capacity C_{car} , and length of the line L . By changing some of these quantities, it is possible to change line capacity.

Numerical Example:

The public transit line length equals 10 km in one direction. The average bus speed on a city heavy traffic equals 20 km/h. The total of 12 buses is assigned to the line. The capacity of every vehicle equals 50. Calculate the line capacity, frequency, and headway.

Solution:

The Line Capacity,

$$C_{line} = \frac{N \times u \times C_{car}}{2 \times L}$$

$$C_{line} = \frac{12 \times 20 \times 50}{2 \times 10}$$

$$\therefore C_{line} = 600(spaces/hr)$$

Then,

$$T = \frac{2 \times L}{u} = \frac{2 \times 10}{20} = 1hr$$

We have,

$$f = \frac{N}{T} = \frac{12}{1}$$

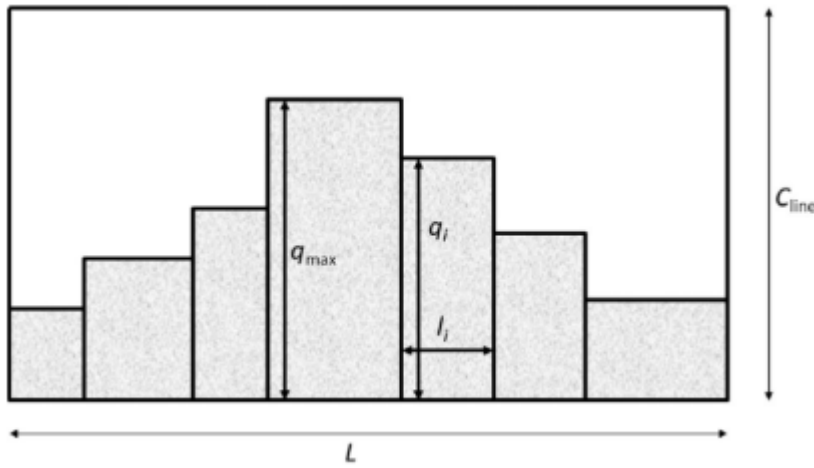
$$\therefore f = 12vehicles/hr$$

$$\text{Also, } h = \frac{1}{f} = \frac{1}{12} = \frac{1}{\frac{12}{60}}$$

$\therefore h = 5 \text{ minutes}$

5.7.1 Transit Line Capacity Utilization

The transit line capacity represents the number of spaces offered to passengers that pass a specific point in one direction during 1 h. Transit operator offers specific number of spaces to the passengers during specific period of time. It could happen that the offered capacity is insufficient in certain situations. On the other hand, is it possible that the offered capacity is underutilized. Therefore, there is a need to measure the utilization of the offered capacity. The figure below shows transit line capacity and passenger volume profile on the transit line between terminal A and terminal B.



The transportation work w_i , made by the transit operator when carrying q_i passengers along the section that has length equal to l_i , equals:

$$w_i = q_i \times l_i \quad (5.22)$$

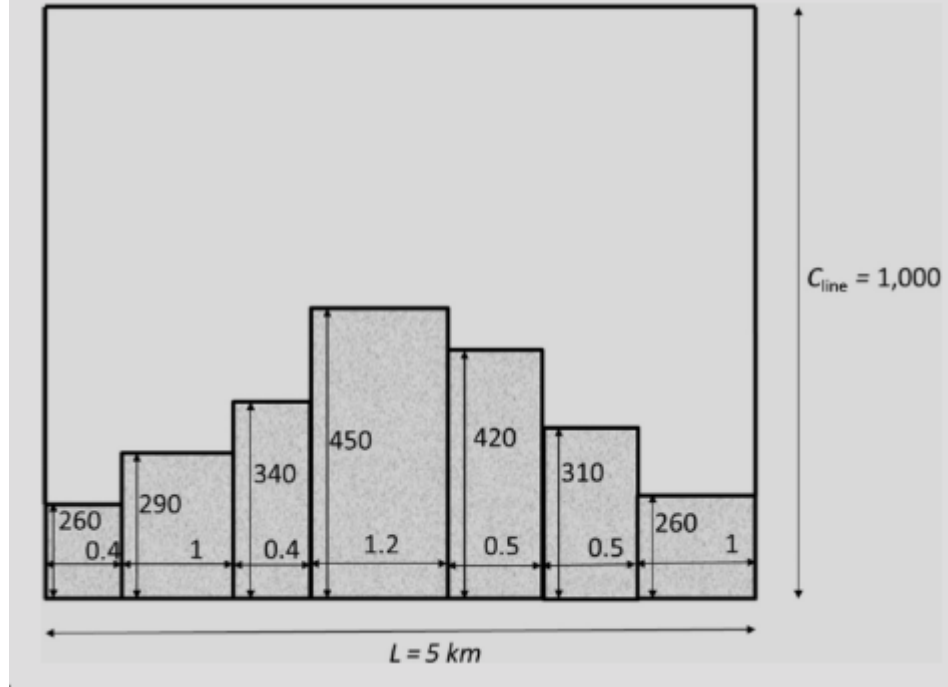
Transit operator offers to the passengers the capacity that is equal to C_{line} . The transportation work that is possible to make is equal to the area of a rectangle with sides L and C_{line} . The realized transportation work is equal to the sum of the areas of shaded rectangles. The average transit line capacity utilization α is given by:

$$\alpha = \frac{\sum_{i=1}^n q_i \times l_i}{C_{line} \times L} \quad (5.23)$$

Where n is the number of line sections.

Numerical Example: The public transit line length equals 5 km in one direction. The average bus speed on a city heavy traffic equals 20 km/h. The

total of 10 buses is assigned to the line. The capacity of every vehicle equals 50. The passenger volume profile of the line is given in the figure below.



Calculate the turnaround time, service frequency, headway, line capacity, and the average transit line capacity utilization α .

Solution:

Turnaround time,

$$T = \frac{2 \times L}{u} = \frac{2 \times 5}{20} = 0.5 \text{ hr}$$

$$\therefore T = 30 \text{ minutes}$$

Then, service frequency:

$$\therefore f = \frac{N}{T} = \frac{10}{0.5} = 20 \text{ vehicles/hr}$$

Also, headway:

$$\therefore h = \frac{1}{f} = \frac{1}{20} = \frac{1}{\frac{20}{60}} = 3 \text{ minutes}$$

Now, Line Capacity,

$$\therefore C_{\text{line}} = \frac{N \times u \times C_{\text{car}}}{2 \times L} = \frac{10 \times 20 \times 50}{2 \times 5} = 1000 \text{ spaces/hr}$$

Also, the average transit line capacity utilization:

$$\alpha = \frac{\sum_{i=1}^n q_i \times l_i}{C_{line} \times L}$$

$$\alpha = \frac{260 \times 0.4 + 290 \times 1 + 340 \times 0.4 + 450 \times 1.2 + 420 \times 0.5 + 310 \times 0.5 + 260 \times 1}{1000 \times 5}$$

$$\therefore \alpha = 0.339$$

5.8 Route Development

A transit system typically consists of many routes on each of which different number of transit units (for example, buses, trams, street, cars, and the like) ply. A route is a path that a transit unit follows during its journey from the origin terminus to the destination and back. A route is a path that a transit unit follows during its journey from the origin terminus to the destination terminus and back. Along the route there can be many stops at which the transit units halt to let passengers alight and board. The subject of this section is to discuss what constitutes a good set of routes and how we can determine such a set. The next section how we can determine the location of stops on a route.

5.8.1 Properties of a Good Route Set

A transit system's aim or for that matter any public transportation system's aim is to provide transport to a large portion of the population efficiently. In order to achieve this goal, the following properties of a route set are desirable.

Ridership: The percentage of the potential transit users who are served by the designed route set should be as high as possible. Given a route set and a statement of the origins and destinations of the potential transit users, we can determine the above quantity easily. An example is shown later to illustrate how this can be done.

Riding Time: The time spent by any passenger on the transit unit should not be more than the time the passenger would have to spend going from his/her origin to his/her destination directly.

Transfer: Often in a transit network, we may have to change routes at some intermediate stops to go from the origin to the destination. This changing of routes is referred to as a transfer. The route set should be designed such that passengers are not required to make too many transfers while travelling from their origins to their destinations.

5.9 Stop Location and Stopping Policy

In this section, some issues related to then number of stops (and their locations) to be provided on a route and the stopping policy to be followed by the transit units plying on a route are discussed. The first subsection is on stopping policy and the next on stop locations.

5.9.1 Stopping Policy

Stopping policy relates to the policy followed by the transit system as regards to stopping on the route. There are different stopping policies of which any one may be used by a transit system. These are: (i) All stop, (ii) On-call stop, and (iii) Demand stop.

In the all stop case, the transit stops at all the stops at all the stops on the route irrespective of whether there is any demand for that stop. For a given transit unit, the demand for a stop exists if someone wants to board the transit unit from the stop or someone on the transit unit wants to alight at that stop. In the on-call stop case, the transit unit stops at a stop if and only if there is a demand for that stop. In the demand stop case, the transit unit stops anywhere along the route where it needs to stop to pick up or drop off passengers. In this case, although the route is fixed, the stop locations are not fixed or alternatively all points on the route are viable stop locations (although the transit unit does not stop at most of them).

In order to decide which is the best policy for a transit system (or a route in the transit system) a simple analysis, based on the expected number of places at which a transit unit a given route may have to stop, is done. The analysis procedure is shown below.

Assume that the transit units are operating on a route with n stops (i.e. stop locations). Note that n could be infinitely many. Further, assume that a transit unit has to stop at a stop only if someone is waiting at the stop or someone on the transit unit has to stop at a stop only. Also assume that the passenger demand for boarding alighting at any stop follows a Poisson distribution with a rate of λ . Now if h is the time headway at which transit units operate on the route being analyzed, then the probability that k persons demand to use a stop is given by,

$$\text{Prob. (} k \text{ persons demand to use a stop)} = P(k) = \frac{(\lambda h)^k \exp(-\lambda h)}{k!} \quad (5.24)$$

Hence,

$$\text{Prob. (transit unit stops at a STOP)} = 1 - P(0) = 1 - \exp(-\lambda h) \quad (5.25)$$

Therefore, on the entire route of n stops, the probability the transit unit stops s times, $\Pi(s)$, is given by:

$$\Pi(s) = \binom{n}{s} [1 - \exp(-\lambda h)]^s [\exp(-\lambda h)]^{n-s} \quad (5.26)$$

Hence, the expected value of s or the average number of times a transit unit stops, $E(s)$, is given by (consult any introductory book on probability theory to see how the following relation can be obtained from the earlier one).

$$E(s) = n[1 - \exp(-\lambda h)] \quad (5.27)$$

The rate λ , at which passenger demand for boarding/alighting arises at any stop, can be written in terms of p , the total passenger demand per unit time for the entire route. If p is the total passenger demand per unit time, then the total demand for boarding or alighting per unit time is $2p$. If n is the total number of stops, then the total passenger demand per unit time for boarding/alighting at a particular stop is $2p/n$ (assuming all stops have, on an average, equal demand). Hence, $\lambda = (2p/n)$. Therefore,

$$E(s) = n[1 - \exp(-2ph/n)] \quad (5.28)$$

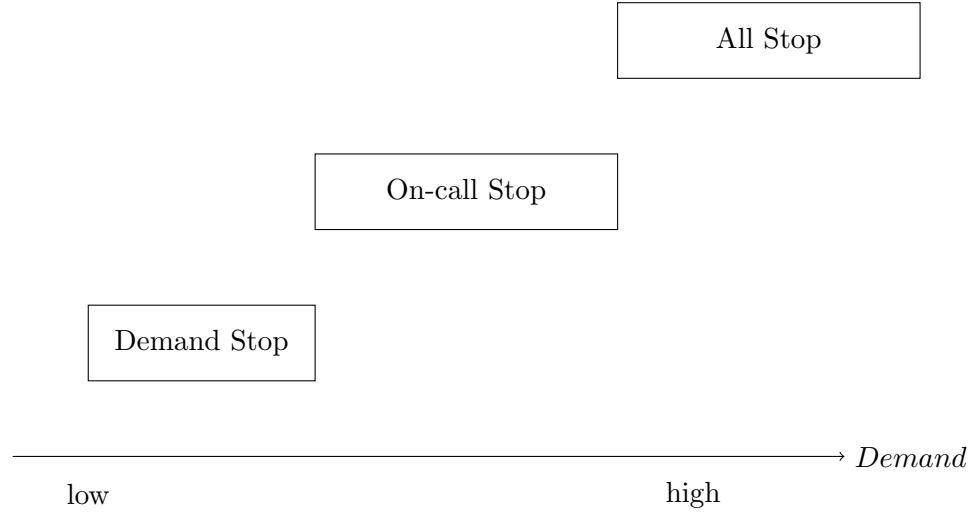
If the demand is very large, that is, $ph \rightarrow \infty$, then from the above relation,

$$\lim_{ph \rightarrow \infty} E(s) = n \quad (5.29)$$

The above relation implies that when the demand is large then on an average, a transit unit will have to stop at all the stops. Hence, in this case it is better to use an all-stop stopping policy.

If, on the other hand, demand ph is small compared to the total number of stops, n , then by expanding the exponential term and ignoring all higher order terms of (ph/n) in the expansion, we can obtain $E(s) = 2ph$. This implies that the average number of places where the transit unit will have to stop is twice the number of passengers who use the transit unit that period h . That is, no special benefit is obtained by providing a fixed number of stops and stop locations. Hence, in the low demand case it is better to follow a demand stop stopping policy with no pre-defined stop locations on the route. For the intermediate values of demand, $E(s) < n$ and therefore following the on-call stop stopping policy makes sense.

The above analysis, in effect, brings out two aspects of stopping policy: (i) it relates demands to stopping policy and (ii) it shows that for the entire range of demand for a route, the expression $n[1 - \exp(-2ph/n)]$ describe the average number of times a transit unit will have to stop.



5.9.2 Stop Location

Decisions on the number of stops to be on a route and where to put those stops effect the efficiency of operation on the route the route as well as the ridership on the route. The number of stops affect the efficiency of operation because every stop adds to the travel time on the route and hence to the round trip time. The number and location of stops also affect the ridership as too few and poorly located stops increase (i) the access time of passengers from their origins to a stop and (ii) the egress time of passengers from a stop to their destination; and too many stops increase the riding time of passengers as the operating speed falls with the increasing number of stops.

Hence in deciding the number of stops and stop locations, we have to strike a compromise between the operating efficiency and passenger convenience. Now, one of the simple models for deciding the number of stops is described.

A good way of determining the number of stops on a route (or the stop spacing) would be to determine the number of stops (or stop spacing) which minimizes the sum of the passenger travel time cost C_p and the operator cost C_o . The passenger travel time should include the access/egress time plus the in-vehicle riding time of the passengers. The operator cost is related to the number of transit units the operator has to operate in order to maintain the required frequency of service (or the required headway between the successive transit units). Mathematically, the total cost per unit time C_T is given by:

$$C_T = C_p + C_o = pT_u k_p + N k_o \quad (5.30)$$

Where,

p is the number of passengers who use the route per unit time.

T_u is the average passenger travel time.

k_p is the cost of unit interval of time to a passenger.

N is the fleet of transit units running in order to maintain the time headway of h .

k_o is the cost of operating a single transit unit per unit time.

However, before attempting to minimize the total cost we need to determine T_u and N as these are functions of the number of stops n .

Determination of T_u

Let l be the average distance a passenger (or user) travels on a route and let L be the total one way length of the route. Let v be the operating speed of the transit unit on the route and let T_v be the total travel time of the transit unit over the length L .

Every time a transit unit stops at a stop location, it needs to decelerate from v to zero speed, remain stopped till all boarding and alighting is over, and then accelerate to v again. In this process, the transit unit loses time due to acceleration and deceleration and information can always be derived once the schedule of arrivals and departures at all stops is known.

5.9.3 Properties of a Good Schedule

A good schedule is one which distributes the available number of transit units in such a way that the following properties are satisfied.

Initial waiting time: The time between the arrival of a passenger at its origin stop for a particular route and the arrival of the next transit unit on that route is referred to as the initial waiting time. This should be as small as possible.

Transfer time: The time spent by passengers at an intermediate stop in order to transfer to another route (from the route on which the passenger came to the intermediate stop) is referred to as transfer time. This time should be small. Further, the transfer time of any passenger should not be greater than some allowable maximum value.

Policy headway: The time headway between the transit units of given route should not be greater than pre-specified maximum value, referred to as policy headway.

5.10 Optimal Paths in Transportation Networks

When traveling through the network we are faced with the problem of finding the paths that are “optimal.” In other words, paths that we are looking for must possess optimal properties. The problems of finding optimal paths in transportation networks are known as shortest path problems. Depending on the context of the problem considered, the “shortest path” could be the shortest path, the fastest path, the most reliable path, the path with the greatest capacity, etc. network links are characterized by length. “Link length” could represent distance, travel time, travel cost, link reliability, etc. Link lengths are mostly treated as deterministic quantities.

Link lengths are treated in some problems as random variables. Most frequently, these link lengths represent travel times. There are random variations in travel times caused by weather conditions, randomness in traffic flows, traffic accidents, and other factors. In these cases, the shortest paths in probabilistic network should be determined. In some cases, when searching for the optimal path, we simultaneously try to optimize two or more objectives. For example, when searching for the best path, we could try simultaneously to take care of travel time as well as travel costs. In such cases, we are dealing with multicriteria shortest path problems.

5.10.1 Dijkstra’s Shortest Path Algorithm

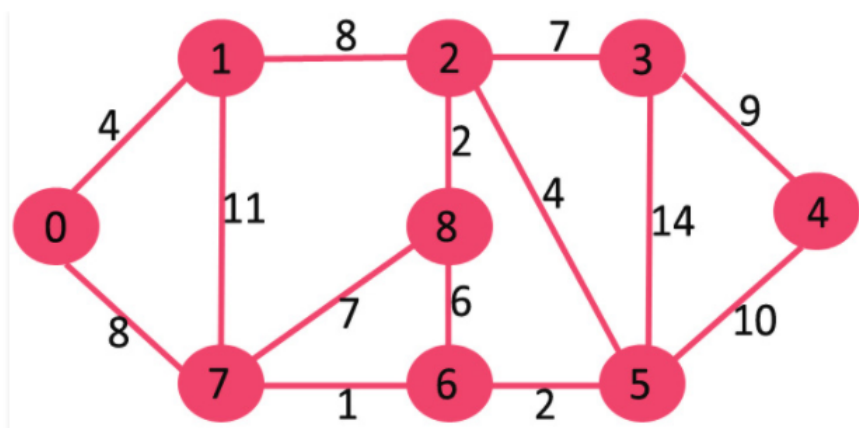
Dijkstra’s algorithm is an algorithm we can use to find shortest distances or minimum costs depending on what is represented in a graph. You’re basically working backwards from the end to the beginning, finding the shortest leg each time. The steps to this algorithm are as follows:

1. Start at the ending vertex by marking it with a distance of 0, because it’s 0 units from the end. Call this vertex your current vertex, and put a circle around it indicating as such.
2. Identify all of the vertices that are connected to the current vertex with an edge. Calculate their distance to the end by adding the weight of the edge to the mark on the current vertex. Mark each of the vertices with their corresponding distance, but only change a vertex’s mark if it’s less than a previous mark. Each time you mark the starting vertex with a mark, keep track of the path that resulted in that mark.
3. Label the current vertex as visited by putting an X over it. Once a vertex is visited, we won’t look at it again.
4. Of the vertices you just marked, find the one with the smallest mark, and make it your current vertex. Now, you can start again from step 2.

5. Once you've labeled the beginning vertex as visited - stop. The distance of the shortest path is the mark of the starting vertex, and the shortest path is the path that resulted in that mark.

Numerical Example:

From the figure below, find the shortest path from node "0" to all other nodes, and also find the shortest route for node "4".



Solution:

Assign cost value of ∞ for all nodes which are not directly connected to node "0", and compute the distance for the nodes "1" and "7".

from/To	1	2	3	4	5	6	7	8
0	4*	∞	∞	∞	∞	∞	8	∞

Since $4 < 8$, we will assign node "7" as our next vertex, then:

from/To	1	2	3	4	5	6	7	8
0	4*	∞	∞	∞	∞	∞	8	∞
1	-	12	∞	∞	∞	∞	8* ($\because 8 < 15$)	∞

In the above step, we assigned the value 8 for node "7", because $8 < 15$ (computed value from node "1"). Also, in this step, since $8 < 12$, we will assign node "7" as our next vertex, then:

from/To	1	2	3	4	5	6	7	8
0	4*	∞	∞	∞	∞	∞	8	∞
1	-	12	∞	∞	∞	∞	8*	∞
7	-	12	∞	∞	∞	9*	-	15

In the above step, node "7" is not directly connected to node "2" and the node is not solved yet, so the value remains as it is. Since, $9 < 12 < 15$, we assign node "6" as our next vertex, then:

from/To	1	2	3	4	5	6	7	8
0	4*	∞	∞	∞	∞	∞	8	∞
1	-	12	∞	∞	∞	∞	8*	∞
7	-	12	∞	∞	∞	9*	-	15
6	-	12	∞	∞	11*	-	-	15

In the above step, the computed value for node "8" from node "6" was 15, since the value didn't change we set the value as it is, that is, 15. Also, $11 < 12 < 15$, so we assign node "5" as our next vertex, then:

from/To	1	2	3	4	5	6	7	8
0	4*	∞	∞	∞	∞	∞	8	∞
1	-	12	∞	∞	∞	∞	8*	∞
7	-	12	∞	∞	∞	9*	-	15
6	-	12	∞	∞	11*	-	-	15
5	-	12* ($\because 12 < 15$)	25	21	-	-	-	15

In the above step, the computed value for node "2" from node "5" was 15, and since $12 < 15$, we retain the earlier value, that is, 12. Also, $12 < 15 < 21 < 25$, we assign node "2" as our next vertex, then:

from/To	1	2	3	4	5	6	7	8
0	4*	∞	∞	∞	∞	∞	8	∞
1	-	12	∞	∞	∞	∞	8*	∞
7	-	12	∞	∞	∞	9*	-	15
6	-	12	∞	∞	11*	-	-	15
5	-	12*	25	21	-	-	-	15
2	-	-	19 ($\because 19 < 25$)	21	-	-	-	14* ($\because 14 < 15$)

in the above step, since $14 < 19 < 21$, we assign node "8" as our next vertex, then:

from/To	1	2	3	4	5	6	7	8
0	4*	∞	∞	∞	∞	∞	8	∞
1	-	12	∞	∞	∞	∞	8*	∞
7	-	12	∞	∞	∞	9*	-	15
6	-	12	∞	∞	11*	-	-	15
5	-	12*	25	21	-	-	-	15
2	-	-	19	21	-	-	-	14*
8	-	-	19*	21	-	-	-	-

In the above step, node "8" had no direct connection with either of the unsolved nodes, "3" and "4", so we resolve to assign node "3" as our next vertex as $19 < 21$, then:

from/To	1	2	3	4	5	6	7	8
0	4*	∞	∞	∞	∞	∞	8	∞
1	-	12	∞	∞	∞	∞	8*	∞
7	-	12	∞	∞	∞	9*	-	15
6	-	12	∞	∞	11*	-	-	15
5	-	12*	25	21	-	-	-	15
2	-	-	19	21	-	-	-	14*
8	-	-	19*	21	-	-	-	-
3	-	-	-	21* ($\because 21 < 28$)	-	-	-	-

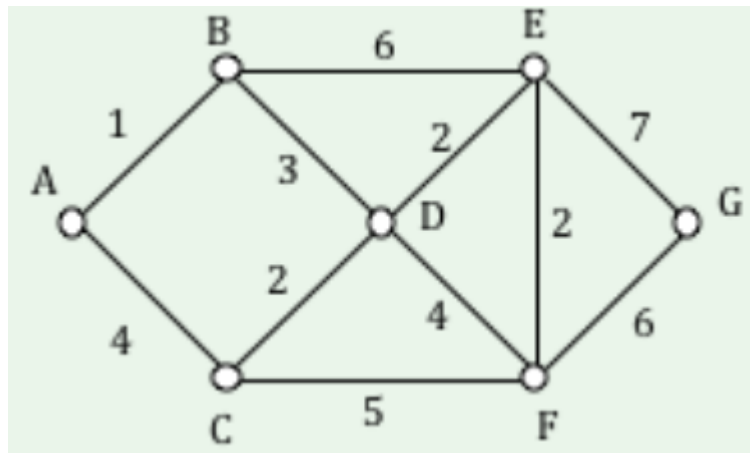
Hence, the shortest distance from node "0" to all other nodes are given below:

node	shortest distance from node "0"/source
1	4
2	12
3	19
4	21
5	11
6	9
7	8
8	14

Hence the shortest route to reach node "4" from node "0" is 0-7-6-5-4.

5.11 Review Problems

1. Define the term public transportation. Briefly describe the different types of transit modes in public transportation.
2. List out the factors which are good and bad for transit.
3. Derive the expression for headway determination by "Square Root Formula".
4. Derive the expression for Transit Line Capacity, and use the expression to determine the average transit line utilization(α).
5. Let us assume that transit line parameters are respectively equal: $c = 120\$$ per hour, $\nu = 12\$$ per passenger hour, $r = 1400$ passengers per hour, and $T = 120$ min = 2hr. Calculate the optimal headway by using the "square root formula."
6. Derive an expression for the the average number of times a transit unit stops, $E(s)$, and use the relation to illustrate the relationship between stopping policy and demand.
7. Describe the properties of a good route set.
8. Discuss the properties of a good schedule.
9. The transit operator monitored during 10 days period the daily number of passengers that departed from the station #5. The following 10 values were recorded: 1160, 1245, 1440, 1280, 1180, 1380, 1220, 1358, 1178 and 1382. The maximum number of passengers per car is equal to 100. The transit operator would like to achieve the load factor that is equal to 0.75. Calculate the average value of the maximum daily passenger volumes. Determine the service frequency f that should be offered in order to satisfy maximum passenger volume and the desired vehicle occupancy.
10. The public transit line length equals 12 km in one direction. The average bus speed on a city heavy traffic equals 25 km/h. The total of 14 buses is assigned to the line. The capacity of every vehicle equals 60. Calculate the line capacity, frequency, and headway.
11. Calculate the shortest path from node "A" to all other nodes, and also determine the shortest route from node "A" to node "G".



12. The number of boarding passengers, number of alighting passengers and the number of passengers in the vehicle are given in the following table:

Bus stop	# of boarding passengers	# of alighting passengers
Terminal A	9	0
1	10	6
2	14	4
3	9	11
4	7	17
5	0	7
Terminal B	0	4

Graphically illustrate the number of boarding passengers, number of alighting passengers and the number of passengers in the vehicle for all line sections. Also, graphically show the passenger load profile and identify the Maximum Load Section (MLS).

13. A public transit operator wants to provide service frequency between 6:00 and 7:00 that is equal to 4. Between 7:00 am and 8:00 am the operator wants to offer service frequency equals to 6. The operator provides service frequency that is equal to 8 in the time interval between 8:00 am and 9:00 am. Generate vehicles timetable that will provide constant vehicle headways within time interval from 6:00 am to 9:00 am. Show the solution graphically.

Chapter 6

Freight Transport Operation

Some important terminologies:

1. **Transport** is the actual movement of goods from one location to another using a means or a vehicle of transport (e.g. trains, trucks, boats) and a transport infrastructure (e.g. roads, railways, canals).
2. **Distribution** often denotes all activities relevant to physical movement of goods, including transportation, but also transshipment and warehousing.
3. **Logistics** is generally used as an overarching term that includes all activities related to the movement and coordination of goods from their source of origin to the final point of delivery, and includes production and distribution. Here, movement does not just correspond to physical movement of goods but also the flow of information.

6.1 Main Actors of Freight Transport and Distribution

While freight transport and distribution include many elements, including various organizations and people, the three fundamental actors who actively take part in the domain are described here.

6.1.1 Shippers

The demand for freight transportation is generated by shippers. Each shipper will have their own logistics strategy, which includes whether to operate their own fleet or to use an external party to which it will outsource their logistics and distribution activities, as well as choosing the mode(s) of transport. The process through which the shippers will define their logistics activities generally takes a three-level decision structure:

1. The long-term decisions in the first level involve defining strategies in line with their customer network and production activities.
2. The second-level medium-term decisions include levels of inventories at production, warehousing, and distribution facilities, frequency and amount of shipping and flexibility of service.
3. At the short-term level, shippers decide on the attributes of the services required for its shipments, such as maximum rates, transport time, reliability and safety.

In making these decisions, they will consider the availability and the characteristics of the services offered on the market by carriers and intermediaries, such as freight brokers and third-party logistics providers.

6.1.2 Carriers

Carriers are people, businesses or organisations that operate and offer transportation services for shippers. They may either provide a customised service, where a vehicle or a fleet will be dedicated exclusively to a particular customer, or operate on the basis of consolidation, where each vehicle contains several pieces of freight for different customers with possibly different origins and destinations. In the latter case, carriers generally operate their services according to a published timetable, which prescribes routes, schedules and rates they offer.

6.1.3 Intermediaries

In some cases, the shipper operates their own fleets of vehicles and does not require an external carrier to ship goods on their behalf. In this case, the management of the relevant transportation and distribution is done inhouse. If a shipper does not own a fleet, then it may choose to work directly with one or several carriers.

Shippers may alternatively use a freight forwarder, an intermediary person or organization that acts as a third party and manages the shipments on behalf of the shipper by contracting one or several carriers. They also help to identify a suitable mode or a combination of modes for the shipper. Freight forwarders work closely with shippers and carriers, as well as other entities in the transportation network, such as ports or terminals, particularly if they additionally undertake ancillary services such as customs clearance.

6.2 Modes of Transportation

There exist different means of transporting freight over the network, each of which is referred to as a mode of transport. Transportation modes can be

differentiated with respect to the type and specification of the vehicle used, the underlying technology, the relevant infrastructure and the nature of the associated operations. The three main modes of transport are air (e.g., cargo planes), land (including road, rail and off-road) and water (e.g., ships in oceans, barges in rivers). Other modes of transport, such as pipelines (e.g., to transport gas) and cable transport (e.g., elevators and cable-cars), also exist.

The term mode can also be used to denote different types of vehicles within a given domain of transport. For example, trucks, vans and bicycles can be seen as three separate modes operating within road transportation due to their distinct features, such as different capacities, capabilities and restrictions. In this example, trucks have the largest capacity but often have restrictions in traveling in urban areas, whereas bicycles are much smaller in capacity but do not suffer from the same type of restrictions as trucks or vans. Brief descriptions of the three main modes of transportation are explained here.

6.2.1 Road

Road has been, and continues to be, the most widely used mode of freight transport, both nationally and globally. One of the main reasons behind its popularity is the ability of road transport to offer a very quick service and often be available on demand.

A wide variety of vehicles are used for road freight transportation, which can be differentiated on the basis of size, capacity, weight and the type of energy used. Vehicle classification on road transportation is generally based on the Gross Vehicle Weight Rating (GVWR), which refers to the maximum allowable total weight of a vehicle including its empty mass, fuel and any load carried. The empty mass of the vehicle, but with fuel and fluids such as engine oil, is named as the curb weight. Vehicle classifications vary from one country to another. In the United States, eight classes exist, with vehicles in the lightest class having a GVWR up to around 3 tonnes, and those in the heaviest class with a GVWR higher than 15 tonnes. In the United Kingdom, more classes exist, with those of at most 3.5 tonnes gross weight described as light goods vehicles (LGVs) and those between 3.5 and 44 tonnes gross weight named as lorries or heavy goods vehicles (HGVs).

Most vehicles used in road freight transportation run on gasoline or diesel engines. Vehicles using alternative sources of fuel or energy have also been developed, such as those running on batteries, biofuels (such as bioalcohol or ethanol), biodiesel, compressed natural gas, hydrogen, and liquefied petroleum gas (LPG), for use in freight distribution. Within urban areas,

human-powered vehicles, such as bicycles and tricycles, can also be used for goods deliveries. To overcome the sole dependency on human power, some of the freight bicycles have power assist motors to aid the cyclist.

The road network is composed of motorways (or highways), urban roads, rural roads, lanes or graded roads and includes bridges and tunnels. Traffic on the road network is controlled by means of traffic signals, signs or markings on the pavement. Various legal requirements are imposed on freight vehicles traveling on the road network, which include limitations on vehicle weight, dimensions, mandatory equipment, licenses and insurances. As for truck drivers, there also exist regulations on driving and working hours, which restrict the duration of driving time and require break and rest periods in long-haul journeys. These regulations aim at reducing driver fatigue, which is known to have adverse affects on road and driver safety. The regulations usually differentiate between on-duty time, which is the time spent working, including driving, waiting, loading and unloading and doing paperwork, and off-duty time, where the driver has no obligation to work. In the United States, for example, these regulations are known as Hours of Service, which limit the maximum consecutive driving time between two rest periods to 11 hours, at which point the driver must be off-duty for at least 10 consecutive hours. Furthermore, a truck driver cannot drive if 8 hours or more have elapsed since the end of the last off-duty period of at least 30 minutes. Similar regulations prevail in other countries, albeit with differences.

6.2.2 Rail

Rail freight transportation is known for its ability to offer cost-effective long-haul transportation services, primarily, but not exclusively, for bulk cargo. There are two major components of a rail system, namely the rail network infrastructure and freight trains.

The rail network is a large and complex structure composed of nodes and tracks (or track segments) as links between the nodes. The former include yards or terminals where classification or marshalling operations are performed, stations where cargo is picked up from or delivered to and junctions that are signal-controlled points in the rail network to allow trains to switch from one route to another.

Freight trains are composed of one or more locomotives, and several rail wagons (or cars). Locomotives move the train along the tracks by either pulling it from the front or push from the rear and range from the earlier types powered by steam to contemporary ones using electricity, magnetic force or diesel engine. Rail wagons carry the freight and come in a variety of forms, including specialized wagons for carrying particular types of cargo

(e.g., autoracks for carrying automobiles or refrigerator cars for temperature-sensitive goods). A train is characterized by its route, origin, destination, intermediate stops, the physical path it travels on and the schedule information that includes departure and arrival times at each station where it stops. Each wagon also has an itinerary that specifies an origin and a destination station and need not correspond to the origin and destination of the train on which it is carried. Wagons may travel on several trains during their journey, usually in groups called blocks. Each block is assigned an origin and a destination, although individual cars in a single block may have different origins and destinations. A block is treated as a single unit for handling purposes. Once formed at its origin yard, a block will not be classified again until it arrives at its destination yard.

Classification or marshalling refers to a set of operations carried out at yards or terminals, where incoming trains are disassembled by decoupling the rail cars and new trains are formed using individual cars or blocks. Bektaş et al. (2009) provide a detailed description of the operations at a classification yard, according to which a train arriving at a yard first enters a receiving area, where the engines are taken away for inspection and maintenance, blocks are separated and cars are inspected. The classification operation begins from this point on and can be performed in two ways, depending on the type of the rail yard. In flat yards, a switching engine is used to push a group of cars out of the receiving tracks onto one of the classification tracks. In hump yards, classification is performed by using an artificially built hill, called the hump, where an engine pushes a group of cars out of the receiving area and up the ramp until it reaches the top of the hill. Due to the pull of the gravitational force, the cars roll down the incline on the other side of the hump, usually one car at a time, and are directed onto one of the classification tracks. Following this operation, each classification track becomes occupied by a group of cars that form the block. Each block then waits until the departure time of its outbound train. When the train is due to leave, they are pulled out of the classification tracks onto the departure tracks and are attached to the train. Following one last inspection of the whole train, the train and blocks leave the yard. The figure below shows the general structure of a classification hump yard.



A schematic representation of a rail classification (hump) yard showing the receiving area (RA), classification tracks (CT) and departure tracks (DT).

Classification is not the only operation that is performed at a rail yard. Other types of operations include inspection, crew change, refueling the

trains and dropping and picking up blocks of cars. Among these, however, classification is known to be the major time-consuming operation.

6.2.3 Air

Air transportation is mainly used for goods that are time sensitive or have high value-to-weight ratios, including perishables due to the need for speed, fashion goods, emergency supplies and spare parts. Speed is a major advantage of air transportation but comes with the significant drawback of having a very high cost. This is one of the reasons as to why air transport has always had a marginal share in the amount of freight shipped worldwide, in comparison to other modes of transport.

Air freight transportation uses a system of air hubs, where outgoing cargo to be shipped is received and consolidated for shipping to a different hub in the network or incoming cargo from another hub is separated and shipped to its final destination. The network tends to have a hub-and-spoke structure to make the best use of economies of scale and help to achieve the lowest possible cost of transport. Specialized pallets or containers named unit load devices are used to stow cargo in aircraft, in a safe and efficient way. Air cargo can be carried in passenger planes along with passenger baggage. Alternatively, it can be transported in cargo aircraft designed specifically for freight, or by helicopters for access to areas that are difficult to reach.

6.2.4 Sea

Sea or maritime transport is the dominant mode with which a large majority of international trade is moved. It is suitable for transporting bulk goods, large packages or containers and similar types of cargo that are not time sensitive.

Vessels are the main vehicles of maritime transport, ranging from small boats to larger ships. Some of the common ship types include break-bulk vessels that are used to transport any type of loose cargo excluding liquid or loose bulk commodities (e.g. boxes, barrels, casks and bagged cargo), Roll-on/Roll-off (RoRo) vessels that carry wheeled cargo (e.g. cars, trucks, trailers), oil tankers that carry crude oil in liquid form, dry bulk vessels that carry bulk goods (e.g. coal, ore, grain) and reefer ships that carry perishable goods that require refrigeration (e.g. fish, meat, fresh produce).

The maritime transportation network consists of ports, which are hubs where ships depart from, stop at, or return to, used for loading, unloading and handling cargo. Ports come in different types based on their location and the nature of the cargo they handle. Seaports are the most common. Some

ports only work with containerized cargo and include specialized equipment such as a gantry crane used to transfer containers between a vessel and other vehicles of land-based transport, such as trucks and rail wagons, and a stacker that is used to lift and swap containers, as well as perform loading and unloading operations on trucks and wagons.

In maritime transportation, there are larger variations in travel as compared with other modes of transport, caused by, for example, weather conditions or congestion at ports. In addition, travel and loading/unloading times are longer, particularly in the case of intercontinental trade.

There are three main modes of operation in maritime transportation:

1. **Liner shipping** is where vessels operate on regular routes with fixed schedules. The routes and schedules are published in a timetable, similar to bus services for passenger transportation. The routes are fixed, each of which is called a service route. There may be several vessels operating on a given service route, and this depends on the desired frequency over a given period of time (e.g., weekly).
2. **Industrial shipping** is where the cargo owner operates their own vessel fleet.
3. **Tramp shipping** is where a carrier operates a fleet of vessels to transport mandatory cargo between specified ports and within a certain time frame for contracted shippers but can also optionally carry spot cargo for other shippers, in order to increase their revenues. Tramp operators will need to decide on whether to accept or reject any spot cargos, which will depend on capacity and timing considerations in the light of the already scheduled routes for mandatory cargo.

6.2.5 Intermodal Transportation

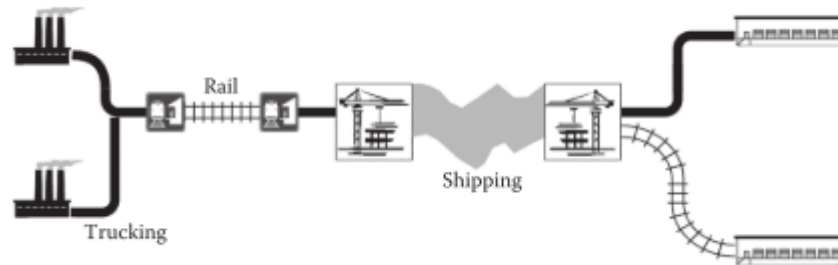
A transportation network is said to be of unimodal type if it only operates a single mode of transport or of multimodal type if it operates at least two different modes of transport. Uni-modal transport cannot take advantage of the use of several modes of transportation, and it is either impractical or impossible to assume a single mode of transport for performing door-to-door delivery. If freight originates from an inland warehouse, for example, and is destined to another inland warehouse on another continent, then part of the journey will have to involve air or water transport.

However, even if there are no physical or geographical limitations, carrying small amounts of cargo on a relatively large vehicle from origin to destination is inefficient. One way to improve efficiency is consolidation. The aim

of intermodal transportation is exactly this, that is, to consolidate loads for efficient long-haul transportation (e.g., by rail or large ocean vessels), while taking advantage of the efficiency of local pick-up and delivery operations by truck. For any shipment to be intermodal, freight must be transported from origin to destination by a sequence of at least two modes of transport, where transfers from one mode to another take place at intermodal terminals (such as sea ports or rail yards). An illustrative example is provided here.

Example:

A multimodal transportation network is illustrated in the figure below, on which three modes of transportation, namely road (trucks), rail and water (shipping), operate. Cargo originates from the two facilities seen on the far left of the figure and is destined to the warehouses appearing on the far right of the figure. The figure also shows how the cargo is shipped using an intermodal chain of modes. In particular, loaded trucks leave the origin facilities to a rail yard, where they are consolidated into a train and sent to another rail yard. Trucks are again used to transport the containers from this rail yard to a sea container terminal. This last operation may not be necessary if the sea container terminal has an interface to the rail network, in which case freight is transferred directly from the former onto the latter. Containers are then transported to a port on another continent by ocean shipping, from where they leave by either trucking or rail (or both) to their destinations.



One of the advantages of intermodal freight transport is the possibility of modal shift, which is defined as partially or fully transferring freight from one mode to the other in the network when it is being shipped to its final destination.

Intermodal Terminals

In intermodal transportation, transfers from one mode to another, and all associated operations related to, for example, loading and unloading, temporary storage or intermediate buffer, and even pre-delivery inspection or enhancement work on the goods are carried out at intermodal terminals. Here are some examples of intermodal terminals:

- **Rail yards** where transfers between the two modes of road (e.g. trucks and lorries) and rail take place.
- **Sea ports** that act as the interface between sea transport (vessels, ferries) and land transport (road or rail). A barge terminal is a smaller port used in inland river transportation and is linked to larger sea terminals.
- **Inland distribution** centers that are used only for road transportation, and act as transfer points between different modes of road transport, for example lorries and vans. Warehouses and cross-docking centres are examples to inland distribution centers.

Containerized Transport

Cargo flow in an intermodal transportation network can be carried by a variety of means, for example crates, pallets or boxes for fresh produce, specialized vessels or tanks for liquid bulk cargo. Irregularity of general cargo, in shape or size, has led to the development of standardized units of shipment, which came to be known as containers. A container is a steel or Aluminium box, with a standard size of 20 ft in length, known as the Twenty Foot Equivalent Unit (TEU), is widely used around the globe, but 40 and 45 ft containers are more commonly used in North America. Containers offer a number of advantages over non-containerized cargo, which are summarized as follows:

1. Containers can be locked, which means that their contents cannot easily be modified except at origin or destination, which implies increased safety and reduced loss and damage.
2. Due to its standard structure, transfer operations at terminals can also be tailored to process containerized cargo fast and with minimal amount of effort. Container ports are excellent examples where such operations can be seen, using container-specific equipment such as container gantry cranes, implying reduced cargo handling.
3. Containers can be stacked, which implies a more efficient use of storage space within the port.
4. Cargo of various sizes, shapes, dimensions or weight can be carried in a container, making it a flexible enough means of transport to accommodate for a variety of cargo types.

Due to these advantages, intermodal transportation, as it is implemented and used today, heavily relies on containerization, to the point that intermodal transportation is often equated to moving containers over long distances, on multimodal networks. Intermodal transportation is not restricted, however, to containers and intercontinental exchanges.

6.3 Choice of Carrier and Transportation Mode

There are a number of factors that shippers use to decide which particular mode or modes of transportation to use for shipping their freight. Earlier studies on this topic have identified six main factors that influence a shipper's decision:

- Freight rates (including cost and charges)
- Reliability (delivery time)
- Transit times (time-in-transit, speed, delivery time)
- Over, short and damaged shipments (loss, damage, claims processing, tracing)
- Shipper market considerations (customer service, user satisfaction, market competitiveness, market influences)
- Carrier considerations (availability, capability, reputation, special equipment)

Later research identified frequency and flexibility as additional factors in particular industries, and that factors such as international dimension, economies of scale, security, environmental concerns and energy use, integration with the supply chain and information technologies, and particularly the use of the Internet are all relevant to the transportation mode choice but remain as under-researched areas.

Other studies have recognized that the perception of a mode of transport is likely to have a significant effect on the choice of mode. Of the more prominent ones, earlier research has identified communication, quality of service, consistency (in delivery), transit times and rates as being the main factors, with the first two being more significant than the others.

6.4 Shipment Options

There are at least two ways in which shipments can be made, which are described in more detail here.

6.4.1 Direct or Customized Shipments

Direct or customized shipments arise when goods are shipped from the source of supply to the demand point without the use of any intermediate facilities or demand points. Full-load trucking is an example of customized transportation, where, upon the call of a customer wishing to have goods

shipped from one (source) point to another (destination) point, a dispatcher assigns a truck to this task. The truck travels to the customer location, is loaded and then moves to the destination, where it is unloaded. Following this, the driver is assigned a new task by the dispatcher, kept waiting until a new demand appears in the near future, or repositioned to a location where a load exists or is expected to be available about the arrival time. The advantages of full-load trucking come from its flexibility in adapting to a highly dynamic environment and uncertain future demands, offering reliability in service and low tariffs compared to other modes of transportation. The fill efficiency of full-load trucking is achieved through the implementation of resource management and allocation strategies that seek to make the best use of the available resources, while maximizing the volume of demand satisfied and the associated profits. Customized services are also offered, for example, by chartered sea or river vessels and planes.

6.4.2 Consolidated Shipments

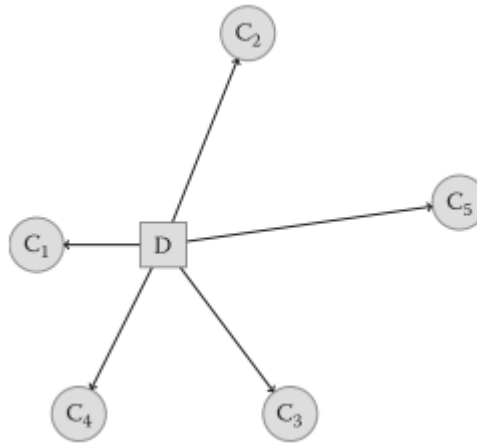
In many cases, trade-offs between volume and frequency of shipping, along with the cost of transportation, render customized services impractical. In such cases, consolidation is used for serving a number of demand points using a particular service. Freight consolidation transportation is performed by Less-Than-Truckload (LTL) motor carriers, railways, ocean shipping lines, regular and express postal services, etc. One example of LTL transportation is when the loads of a given set of customers are consolidated at a given source (e.g. a depot), and the deliveries are multiple destination points. In this case, a truck or a lorry is loaded with all the goods at the depot and is dispatched to visit the customers in a particular order to perform the deliveries. A consolidation-based transportation system can also be structured as a hub-and-spoke network, where shipments for a number of origin-destination points may be transferred via intermediate consolidation facilities, or hubs, such as airports, seaport container terminals, rail yards, truck break-bulk terminals and intermodal platforms.

6.5 Distribution Structures

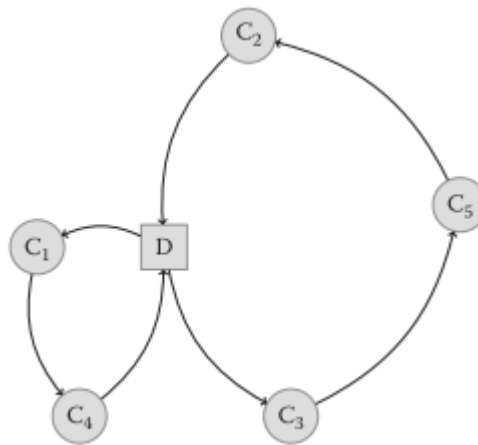
The structure of the way in which goods are distributed is generally in the form of single-echelon (or single-tier) or multi-echelon (or multi-tier) distribution network. The type of structure used depends on a number of factors, such as the type of area (e.g. urban, suburban, rural) or the size of the area (e.g. countries, continents), the type of product(s) being shipped, the types of vehicles used and the demand requirements in terms of volume and time.

6.5.1 Single-echelon

A single-echelon distribution structure does not involve the use of any intermediate facilities between the source(s) of supply and the source(s) of demand and operates on the basis that deliveries are made from the former to the latter. This might be in the form of direct shipments from the supply point to the customer, or, as in the case of consolidation, from one supply point to many customers.

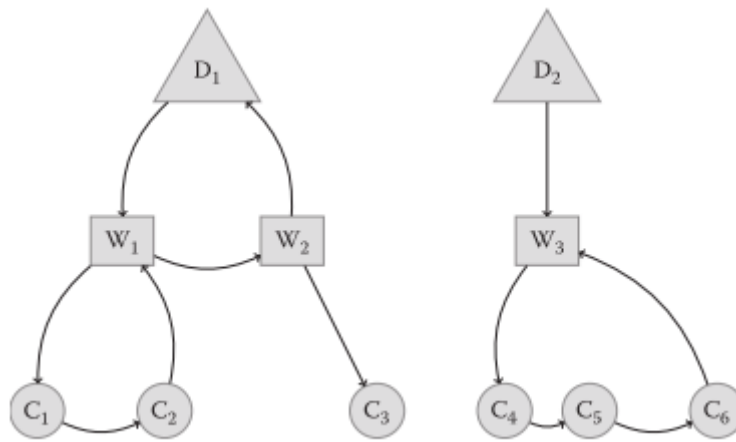


The figure above shows a single echelon distribution structure where a central depot, shown by D , serves five customers labeled C_1 – C_5 , all using direct shipments. In this example, it is assumed that each link carries at least one full-truckload shipment. The figure below shows the same network in which it is assumed that the deliveries can be consolidated. In particular, one LTL vehicle serves customers C_1 and C_4 in the given order, and the other serves customers C_3, C_5 and C_2 in the given order.



6.5.2 Multi-echelon

If the distribution network includes several layers of intermediate facilities, such as warehouses, consolidation facilities, distribution centres and cross-docks, to which the goods are sent as part of the distribution from their origin(s) to their destination(s), then the network is said to have a multi-echelon structure. Supply chain networks often have such a structure, for which the planning problems entail the management and operation of the facilities in the network as well as the assignment of (aggregated) flow between the facilities. In the context of freight distribution, however, the planning problems are more specific to the management of transport modes and the vehicle fleets used in between the different layers.



A two-echelon network distribution structure is shown in the figure above, where the first echelon includes all movements between the first layer of depots shown by D_1 and D_2 and the second layer of intermediate points (e.g. warehouses) shown by W_1 , W_2 and W_3 . In this example, consolidated shipments depart from D_1 to serve the two warehouses W_1 and W_2 . In contrast, a direct shipment using a full-truckload is made between D_2 and W_3 . The second echelon consists of all the distribution activities between the second layer and the third layer, where the latter includes six demand points shown by C_1 – C_6 . It is clear from the figure that this network uses a mix of consolidated and direct shipments at this echelon.

Structures such as those shown in the figure above have been suggested for use within urban areas, where the first layer of depots consists of those that are located outside the boundaries of the urban zone and the second layer consists of intermediate points located on the boundaries. The expectation is that bulk deliveries are made from the first to the second layer using heavy goods vehicles, which would stop here. The distribution activities

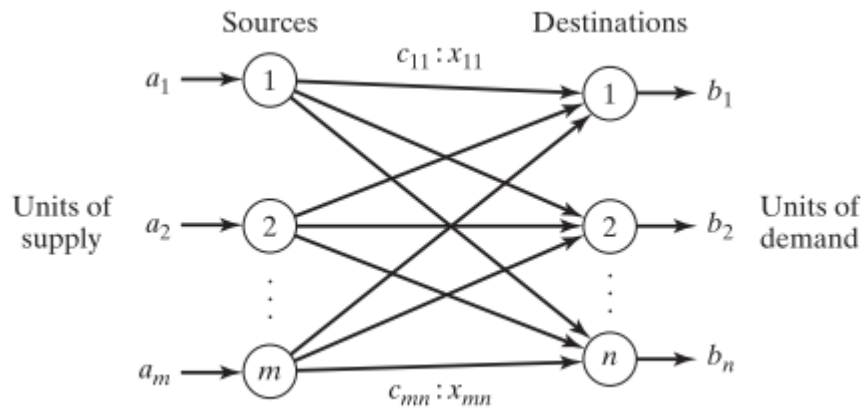
within the urban zone would then be carried out using smaller vehicles that would operate between the second layer of intermediate points and the third layer of demand points. Such a strategy prevents large and heavy vehicles from entering urban zones.

6.6 Transportation Model

The transportation model deals with a special class of linear programming problem in which the objective is to transport a homogeneous commodity from various origins or factories to different destinations or markets at a total minimum cost.

6.6.1 Definition of the Transportation Model

The problem is represented by the network in the figure below. There are m sources and n destinations, each represented by a node. The arcs represent the routes linking the sources and the destinations. Arc (i, j) joining source i to destination j carries two pieces of information: the transportation cost per unit, c_{ij} , and the amount shipped, x_{ij} . The amount of supply at source i is a_i , and the amount of demand at destination j is b_j . The objective of the model is to minimize the total transportation cost while satisfying all the supply and demand restrictions.



Example #1:

MG Auto has three plants in Los Angeles, Detroit, and New Orleans and two major distribution centers in Denver and Miami. The quarterly capacities of the three plants are 1000, 1500, and 1200 cars, and the demands at the two distribution centers for the same period are 2300 and 1400 cars. The transportation costs per car on the different routes, rounded to the closest

dollar is given in the table below. Determine the amount of cars to distributed to each distribution center such that the cost transportation cost is minimum, satisfying all demand and supply constraints.

Transportation cost per car		
From/To	Denver(1)	Miami(2)
LA (1)	\$80	\$215
Detroit (2)	\$100	\$108
New Orleans (1)	\$102	\$68

Solution:

The linear Programming model of the problem is:

Minimize, $z = 80x_{11} + 215x_{12} + 100x_{21} + 108x_{22} + 102x_{31} + 68x_{32}$

Subject to:

$$\begin{aligned}
 x_{11} + x_{12} + 0 + 0 + 0 + 0 &= 1000 \text{ (LA)} \\
 0 + 0 + x_{21} + x_{22} + 0 + 0 &= 1500 \text{ (Detroit)} \\
 0 + 0 + 0 + 0 + x_{31} + x_{32} &= 1200 \text{ (New Orleans)} \\
 x_{11} + 0 + x_{21} + 0 + x_{31} + 0 &= 2300 \text{ (Denver)} \\
 0 + x_{12} + 0 + x_{22} + 0 + x_{32} &= 1400 \text{ (Miami)} \\
 x_{ij} &\geq 0, i = 1, 2, 3 \text{ and } j = 1, 2
 \end{aligned}$$

All the constraints are equations because the total supply (= 1000+ 1500 + 1200 = 3700) equals the total demand (= 2300 + 1400 = 3700).

From / To	Denver	Miami	Supply
	x_{11}	x_{12}	
LA	80	215	1000
	x_{21}	x_{22}	
Detroit	100	108	1500
	x_{31}	x_{32}	
New Orleans	102	68	1200
Demand	2300	1400	

The special structure of the transportation problem allows a compact representation of the problem using the transportation tableau format in shown above.

The optimal solution of the given transportation problem ships 1000 cars from LA to Denver ($x_{11} = 1000$), 1300 from Detroit to Denver ($x_{21} = 1300$), 200 from Detroit to Miami ($x_{22} = 200$), and 1200 from New Orleans to Miami ($x_{31} = 1200$). The associated minimum transportation cost is computed as $100 \times \$80 + 1300 \times \$100 + 200 \times \$108 + 1200 \times \$68 = \$313,200$.

A transportation problem is said to be balanced if the total supply equals to the total demand. However, if the problem is not balanced then it can be balanced by introducing a dummy row or column having cost value of zero, i.e:

- **If Total Supply < Total Demand:** A dummy row, whose each cell have zero cost value, is added to the table with a corresponding supply value = Total Demand - Total Supply.
- **If Total Supply > Total Demand:** A dummy column, whose each cell have zero cost value, is added to the table with a corresponding demand value = Total Supply - Total Demand.

6.6.2 The Transportation Algorithm

The special transportation algorithm that will be presented in this section was developed early on when hand computations were the norm and shortcuts were warranted. Today, powerful computer codes can solve transportation models of any size as a regular LP. But there is more to the transportation model than the hand computations. First, its historical significance in the evolution of OR techniques is important and must be preserved. Second, the special transportation tableau format can facilitate modeling a number of situations that do not deal directly with transporting goods. Third, the algorithmic hand computations boast such (almost intuitive) simplicity that a beginner can get a “feel” of what optimization is about. Lastly, the transportation algorithm does provide insight into the use of the theoretical primal–dual relationships to achieve a practical end result—that of developing simple rules for hand computations. The exercise is theoretically intriguing.

The basic steps of the Transportation Algorithm are as follows:

- **Step 1:** Balance the given transportation problem.
- **Step 2:** Compute the starting basic feasible solution. A basic feasible solution can be computed using either of these methods:
 1. North - west corner method.
 2. Least - cost cell method. (Or Inspection method Or Matrix minimum - row minimum - column minimum method).

3. Vogel's Approximation Method, generally known as VAM.

- **Step 3:** After getting the basic feasible solution conduct a optimality test to check whether the solution is optimal or not. There are two methods of computing optimality test, they are:

1. Stepping Stone Method.
2. Modified Distribution Method, generally known as MODI method.

For this subject, we will be only computing up to the basic feasible solution of the transportation problem since solving for optimality test is a cumbersome task on its own and doing so causes to deviate away from the course domain of the subject. These topics are only introduced for the sake of familiarizing students with the scope of transportation engineering. For an in-depth study about these topics, students are highly encouraged to peruse books related to Operation Research.

6.6.3 Determining the Basic Feasible Solution

a. North-West Corner Method

The following are the steps used for determining the basic feasible solution of a transportation problem using the North-West Corner Method:

- Balance the problem, that is, check whether the total demand is equal to the total demand or not. If not, open a dummy column or dummy row as the case may be and balance the problem.
- Start from the left hand side top corner or cell and make allocations depending on the availability and requirement constraint. If the availability constraint is less than the requirement constraint, then for that cell make allocation in units which is equal to the availability constraint. In general, verify which is the smallest among the availability and requirement and allocate the smallest one to the cell under question. Then proceed allocating either side ways or downward to satisfy the rim requirement. Continue this until all the allocations are over.
- Once all the allocations are over, i.e., both rim requirement (column and row i.e., availability and requirement constraints) are satisfied, write allocations and calculate the cost of transportation.

Numerical Example:(Sunray Transport Problem)

SunRay Transport Company ships truckloads of grain from three silos to four mills. The supply (in truckloads) and the demand (also in truckloads) together with the unit transportation costs per truckload on the different routes are summarized in the table below. The unit transportation costs are in hundreds of dollars. The model seeks the minimum cost shipping schedule between the silos and the mills.

Transportation Cost					
Silo / Mill	1	2	3	4	Supply
1	\$10	\$2	\$20	\$11	15
2	\$12	\$7	\$9	\$20	25
3	\$4	\$14	\$16	\$18	10
Demand	5	15	15	15	

Solution:

$$\text{Total Supply} = 15 + 25 + 10 = 50$$

$$\text{Total Demand} = 5 + 15 + 15 + 15 = 50$$

\therefore Total Supply = Total Demand, the transportation model is balanced.

At first, we choose the cell that is in the extreme north-west position, that is, cell corresponding to transport operation from silo 1 to mill 1. For that cell, we have supply of 15 and demand of 5. $\therefore 5 < 15$, we allocate the value of 5 this cell. Allocating 5 truckloads from silo 1 to mill 1 leaves silo 1 with a remainder of 10 truckloads of supply. Also demand for mill 1 has been satisfied so we gray out the column corresponding to mill 1 indicating it requires no further action. This is shown in the transportation tableau below:

Silo / Mill	1	2	3	4	Supply
	5				
1	\$10	\$2	\$20	\$11	15 10
2	\$12	\$7	\$9	\$20	25
3	\$4	\$14	\$16	\$18	10
Demand	5 0	15	15	15	

Similarly, cell corresponding to transport operation from silo 1 to mill 2 is in the north-west position considering the remaining table (Table excludes the grayed out region). For this cell, supply is 10 and demand is 15. $\therefore 10 < 15$, we allocate value of 10 for this cell. Allocating 10 truckloads from silo 1 leaves the supply capacity of silo 1 to 0 and demand of mill 2 is reduced to 5 truckloads. Since silo 1 can no longer supply any truckloads, we gray out the corresponding row.

Silo / Mill	1	2	3	4	Supply
	5	10			
1	\$10	\$2	\$20	\$11	15 0
2	\$12	\$7	\$9	\$20	25
3	\$4	\$14	\$16	\$18	10
Demand	5 0	15 5	15	15	

Similarly, cell corresponding to transport operation from silo 2 to mill 2 is in the north-west position considering the remaining table. For this cell, supply is 25 and demand is 5. $\therefore 5 < 25$, we allocate value of 5 for this cell. Allocating 5 truckloads from silo 2 reduces the supply capacity of silo 2 to 20 and demand of mill 2 is reduced to 0 truckloads. Since mill 2 can no longer receive truckloads, we gray out the corresponding column.

Silo / Mill	1	2	3	4	Supply
	5	10			
1	\$10	\$2	\$20	\$11	15 0
2	\$12	\$7	\$9	\$20	25 20
3	\$4	\$14	\$16	\$18	10
Demand	5 0	15 5 0	15	15	

Similarly, cell corresponding to transport operation from silo 2 to mill 3 is in the north-west position considering the remaining table. For this cell, supply is 20 and demand is 15. $\therefore 15 < 20$, we allocate value of 15 for this cell. Allocating 15 truckloads from silo 2 reduces the supply capacity of silo 2 to 5 and demand of mill 3 is reduced to 0 truckloads. Since mill 3 can no longer receive truckloads, we gray out the corresponding column.

Silo / Mill	1	2	3	4	Supply
1	<div>\$10</div> <div>5</div>	<div>\$2</div> <div>10</div>	<div>\$20</div>	<div>\$11</div>	15 10 0
2	<div>\$12</div>	<div>\$7</div> <div>5</div>	<div>\$9</div> <div>15</div>	<div>\$20</div>	25 20 5
3	<div>\$4</div>	<div>\$14</div>	<div>\$16</div>	<div>\$18</div>	10
Demand	5 0	15 5 0	15 0	15	

Similarly, cell corresponding to transport operation from silo 2 to mill 4 is in the north-west position considering the remaining table. For this cell, supply is 5 and demand is 15. $\therefore 5 < 15$, we allocate value of 5 for this cell. Allocating 5 truckloads from silo 2 reduces the supply capacity of silo 2 to 0 and demand of mill 3 is reduced to 10 truckloads. Since, silo 2 can no longer supply any truckloads, we gray out the corresponding row.

Silo / Mill	1	2	3	4	Supply
1	<div>\$10</div> <div>5</div>	<div>\$2</div> <div>10</div>	<div>\$20</div>	<div>\$11</div>	15 10 0
2	<div>\$12</div>	<div>\$7</div> <div>5</div>	<div>\$9</div> <div>15</div>	<div>\$20</div> <div>5</div>	25 20 5 0
3	<div>\$4</div>	<div>\$14</div>	<div>\$16</div>	<div>\$18</div>	10
Demand	5 0	15 5 0	15 0	15 10	

Finally, there is only one cell left, cell corresponding to transport operation from silo 3 to mill 4, so we assign the remaining 10 truckloads to this cell and the transportation supply and demand constraints are also met.

From the basic feasible solution obtained using the North-West Corner method, following assignments are obtained: $x_{11} = 5$, $x_{12} = 10$, $x_{22} = 5$, $x_{23} = 15$, $x_{24} = 5$, and $x_{34} = 10$.

Also, the associated cost of the schedule is,

$$z = 5 \times \$10 + 10 \times \$2 + 5 \times \$7 + 15 \times \$9 + 5 \times \$20 + 10 \times \$18 = \$520$$

b. Least-Cost Cell Method

The least-cost method finds a better starting solution by targeting the cheapest routes. It assigns as much as possible to the cell with the smallest unit cost (ties are broken arbitrarily). Next, the satisfied row or column is crossed out and the amounts of supply and demand are adjusted accordingly. If both a row and a column are satisfied simultaneously, only one is crossed out, the same as in the northwest-corner method. Next, select the uncrossed-out cell with the smallest unit cost and repeat the process until exactly one row or column is left uncrossed out.

The following are the steps used for determining the basic feasible solution of a transportation problem using the Least-Cost Cell Method:

- *Step 1:* Identify the lowest cost cell in the given matrix. When more than one cell has the same cost, then both the cells are competing for allocation. This situation in transportation problem is known as tie. To break the tie, select any one cell of your choice for allocation. Make allocations to this cell either to satisfy availability constraint or requirement constraint. Once one of these is satisfied, then mark crosses (\times) in all the cells in the row or column which ever has completely allocated.
- *Step 2:* Are all allocations completed? If yes the basic solution has been obtained, else go to the *Step 1*.

Numerical Example: Determine the basic feasible solution for the Sunray Transport Problem using the Least-Cost Cell method.

Solution:

$$\text{Total Supply} = 15 + 25 + 10 = 50$$

$$\text{Total Demand} = 5 + 15 + 15 + 15 = 50$$

\therefore Total Supply = Total Demand, the transportation model is balanced.

First, we have to choose the cell having the least cost, in this problem the cell corresponding to transport operation from silo 1 to mill 2 has the least cost (\$2). For this cell, the supply is 15 and demand is also 15. So we assign the value of 15 for this cell, which reduces the supply of silo 1 to 0 and demand of mill 2 to 0. Since silo 1 can no longer supply and mill 2 can no longer receive truckloads, we gray out the corresponding row and column.

Silo / Mill	1	2	3	4	Supply
		15			
1	\$10	\$2	\$20	\$11	15 0
2	\$12	\$7	\$9	\$20	25
3	\$4	\$14	\$16	\$18	10
Demand	5	15 0	15	15	

Similarly for the remaining active cells, the cell corresponding to transport operation from silo 3 to mill 1 has the least cost (\$4) among other active cells. For this cell, the supply is 10 and demand is 5. Since $5 < 10$ we assign the value of 5 for this cell, which reduces the supply of silo 3 to 5 and demand of mill 1 to 0. Since mill 1 can no longer receive truckloads, we gray out the corresponding column.

Silo / Mill	1	2	3	4	Supply
		15			
1	\$10	\$2	\$20	\$11	15 0
2	\$12	\$7	\$9	\$20	25
3	\$4	\$14	\$16	\$18	10 5
Demand	5 0	15 0	15	15	

Similarly for the remaining active cells, the cell corresponding to transport operation from silo 2 to mill 3 has the least cost (\$9) among other active cells. For this cell, the supply is 25 and demand is 15. Since $15 < 25$ we assign the value of 15 for this cell, which reduces the supply of silo 2 to 10 and demand of mill 3 to 0. Since mill 3 can no longer receive truckloads, we gray out the corresponding column.

Silo / Mill	1	2	3	4	Supply
1	\$10	\$2 15	\$20	\$11	15 0
2	\$12	\$7	\$9 15	\$20	25 10
3	\$4 5	\$14	\$16	\$18	10 5
Demand	5 0	15 0	15 0	15	

Similarly for the remaining active cells, the cell corresponding to transport operation from silo 3 to mill 4 has the least cost (\$18) among other active cells. For this cell, the supply is 5 and demand is 15. Since $5 < 15$ we assign the value of 5 for this cell, which reduces the supply of silo 3 to 0 and demand of mill 4 to 10. Since silo 3 can no longer supply truckloads, we gray out the corresponding row.

Silo / Mill	1	2	3	4	Supply
1	\$10	\$2 15	\$20	\$11	15 0
2	\$12	\$7	\$9 15	\$20	25 10
3	\$4 5	\$14	\$16	\$18 5	10 5 0
Demand	5 0	15 0	15 0	15 10	

Finally the cell corresponding to transport operation from silo 2 to mill 4 is only left which has a both demand and supply value of 10. Hence we assign the value of 10 to this cell.

From the basic feasible solution obtained using the North-West Corner method, following assignments are obtained: $x_{13} = 5, x_{12} = 15, x_{23} = 15, x_{24} = 10$, and $x_{34} = 5$.

Also, the associated cost of the schedule is,

$$z = 15 \times \$2 + 5 \times \$4 + 15 \times \$9 + 10 \times \$20 + 5 \times \$18 = \$475.$$

c. Vogel's Approximation Method (Opportunity Cost Method)

VAM is an improved version of the least-cost method that generally, but not always, produces better starting solutions. In this method, we use concept of opportunity cost. Opportunity cost is the penalty for not taking correct decision. This is the penalty for not taking correct decision and hence the opportunity cost. The following are the steps used for determining the basic feasible solution of a transportation problem using the Least-Cost Cell Method:

- Step 1: For each row (column), determine a penalty measure by subtracting the smallest unit cost in the row (column) from the next smallest unit cost in the same row (column). This penalty is actually a measure of lost opportunity one forgoes if the smallest unit cost cell is not chosen.
- Identify the row or column with the largest penalty, breaking ties arbitrarily. Allocate as much as possible to the variable with the least unit cost in the selected row or column. Adjust the supply and demand, and cross out the satisfied row or column. If a row and a column are satisfied simultaneously, only one of the two is crossed out, and the remaining row (column) is assigned zero supply (demand).
- 1. If exactly one row or column with zero supply or demand remains uncrossed out, stop.
 2. If one row (column) with positive supply (demand) remains uncrossed out, determine the basic variables in the row (column) by the least-cost method. Stop.
 3. If all the uncrossed-out rows and columns have (remaining) zero supply and demand, determine the zero basic variables by the least-cost method. Stop.
 4. Otherwise, go to step 1.

Numerical Example: *Determine the basic feasible solution for the Sunray Transport Problem using the Vogel's Approximation method.*

Solution:

$$\text{Total Supply} = 15 + 25 + 10 = 50$$

$$\text{Total Demand} = 5 + 15 + 15 + 15 = 50$$

\therefore Total Supply = Total Demand, the transportation model is balanced.

At first, we compute the values of penalties for each row and column of the transportation model. For first row (supply row of silo 1), 10 and 2 are the two minimum values. Thus, the value of penalty for the corresponding row is 8. For first column (demand column of mill 1), 10 and 4 are the two minimum value. Thus, the value of penalty for the corresponding column is 6. Similarly the value of penalty for all rows and columns are computed as shown in the table below. Comparing the values of penalty of all rows and column, it is found that the value of penalty corresponding to the third row, that is 10, is the maximum value. Since 10 is the max value, we determine the cell to be assigned for the corresponding row using least-cost cell method.

S/M	1	2	3	4	Supply	P_1
1	\$10	\$2	\$20	\$11	15	$10 - 2 = 8$
2	\$12	\$7	\$9	\$20	25	$9 - 7 = 2$
3	\$4	\$14	\$16	\$18	10 5	$14 - 4 = 10 \leftarrow$
Demand	5 0	15	15	15		
P_1	$10 - 4 = 6$	5	7	7		

Similarly, we compute the value of penalty for all active rows and columns. Since 9, which is the penalty value for row 1, is the maximum value among all values of penalty, we determine the cell to be assigned for the corresponding row using least-cost cell method.

S/M	1	2	3	4	Sup	P_1	P_2
1	\$10	\$2	\$20	\$11	15 0	8	$9 \leftarrow$
2	\$12	\$7	\$9	\$20	25	2	2
3	\$4	\$14	\$16	\$18	10 5	$10 \leftarrow$	2
Dem	5 0	15 0	15	15			
P_1	6	5	7	7			
P_2	-	5	7	7			

Similarly,

S/M	1	2	3	4	Sup	P_1	P_2	P_3
1	\$10	\$2	\$20	\$11	15 0	8	9 ←	-
2	\$12	\$7	\$9	\$20	25 10	2	2	11 ←
3	\$4	\$14	\$16	\$18	10 5	10 ←	2	2
Dem	5 0	15 0	15 0	15				
P_1	6	5	7	7				
P_2	-	5	7	7				
P_3	-	-	7	2				

Similarly,

S/M	1	2	3	4	Sup	P_1	P_2	P_3	P_4
1	\$10	\$2	\$20	\$11	15 0	8	9 ←	-	-
2	\$12	\$7	\$9	\$20	25 10	2	2	11 ←	20 ←
3	\$4	\$14	\$16	\$18	10 5	10 ←	2	2	18
Dem	5 0	15 0	15 0	15 5					
P_1	6	5	7	7					
P_2	-	5	7	7					
P_3	-	-	7	2					
P_4	-	-	-	2					

The basic feasible solution is shown below:

Silo / Mill	1	2	3	4	Supply
1	10	15 \$2	20	11	15
2	12	7	15 \$9	10 \$20	25
3	5 \$4	14	16	5 \$18	10
Demand	5	15	15	15	

The associated cost of the schedule is,

$$z = 15 \times \$2 + 5 \times \$4 + 15 \times \$9 + 10 \times \$20 + 5 \times \$18 = \$475.$$

North-West Corner Method	Least-Cost Cell Method	Vogel's Approximation Method
1. The allocation is made from the left hand side top corner irrespective of the cost of the cell.	The allocations are made depending on the cost of the cell. Lowest cost is first selected and then next highest etc.	The allocations are made depending on the opportunity cost of the cell.
2. As no consideration is given to the cost of the cell, naturally the total transportation cost will be higher than the other methods.	As the cost of the considered while cell is making allocations, the total cost of transportation will be comparatively less.	As the allocations are made depending on the opportunity cost of the cell, the basic feasible solution obtained will be very nearer to optimal solution.
3. It takes less time. This method is suitable to get basic feasible solution quickly.	The basic feasible solution, we get will be very nearer to optimal solution. It takes more time than northwest corner method.	It takes more time for getting basic Feasible solution. But the solution we get will be very nearer to Optimal solution.
4. When basic feasible solution alone is asked, it is better to go for northwest corner method.	When optimal asked, better solution to go is for least-cost cell method for basic feasible solution and MODI for optimal solution.	VAM and MODI is the best option to get optimal solution.

6.7 Assignment Model

An assignment problem is a particular case of transportation problem where the objective is to assign a number of resources to an equal number of activities so as to minimize total cost or maximize total profit of allocation.

The problem of assignment arises because available resources such as men, machines etc. have varying degrees of efficiency for performing different activities, therefore, cost, profit or loss of performing the different activities is different.

Thus, the problem is “How should the assignments be made so as to optimize the given objective”. Some of the problem where the assignment technique may be useful are assignment of workers to machines, salesman to different sales areas.

6.7.1 Definition of the Assignment Problem

Suppose there are n jobs to be performed and n persons are available for doing these jobs. Assume that each person can do each job at a term, though with varying degree of efficiency, let c_{ij} be the cost if the i -th person is assigned to the j -th job. The problem is to find an assignment (which job should be assigned to which person one on-one basis) So that the total cost of performing all jobs is minimum, problem of this kind are known as assignment problem.

For solving the assignment problem we use Assignment technique or Hungarian method or Flood's technique. All are one and the same. Above, it is mentioned that one origin is to be assigned to one destination. This feature implies the existence of two specific characteristics in linear programming problems, which when present, give rise to an assignment problem. The first one being the pay of matrix for a given problem is a square matrix and the second is the optimum solution (or any solution with given constraints) for the problem is such that there can be one and only one assignment in a given row or column of the given payoff matrix. The transportation model is a special case of linear programming model (Resource allocation model) and assignment problem is a special case of transportation model, therefore it is also a special case of linear programming model. Hence it must have all the properties of linear programming model. That is it must have: (i) an objective function, (ii) it must have structural constraints, (iii) It must have non-negativity constraint and (iv) The relationship between variables and constraints must have linear relationship. In our future discussion, we will see that the assignment problem has all the above properties.

There are different types of assignment problems. They are:

1. Assigning the jobs to machines when the problem has square matrix to minimize the time required to complete the jobs. Here the number of rows i.e. jobs are equals to the number of machines i.e. columns.
2. The second type is maximization type of assignment problem. Here we have to assign certain jobs to certain facilities to maximize the returns or maximize the effectiveness.
3. Assignment problem having non-square matrix. Here by adding a dummy row or dummy columns as the case may be, we can convert a non-square matrix into a square matrix and proceed further to solve the problem.
4. Assignment problem with restrictions. Here restrictions such as a job cannot be done on a certain machine or a job cannot be allocated to a certain facility may be specified. In such cases, we should neglect such cell or give a high penalty to that cell to avoid that cell to enter into the program.
5. Traveling sales man problem (cyclic type). Here a salesman must tour certain cities starting from his hometown and come back to his hometown after visiting all cities.

6.7.2 Hungarian Method

The Hungarian method (developed by Hungarian mathematician D. Konig) is an efficient method of finding the optimal solution of an assignment problem without making a direct comparison of every solution. The method works on the principle of reducing the given cost matrix to a matrix of opportunity costs. Opportunity costs show the relative penalties associated with assigning a resource to an activity. Hungarian method reduces the cost matrix to the extent of having at least one zero in each row and column so as to make optimal assignments.

The Hungarian method (minimization case) can be summarized in the following steps:

Step A: Develop the cost matrix from the given problem If the number of rows are not equal to the number of columns, then add required number of dummy rows or columns. The cost element in dummy rows/columns are always zero.

Step B: Compute the opportunity cost matrix 1. Identify the smallest element in each row of cost matrix and then subtract it from each element of that row.

2. In the reduced matrix obtained from 2(a), identify the smallest element in each column and then subtract it from each element of that column. Each row and column now have at least one zero element.

Step C: Make assignments in the opportunity cost matrix The procedure of making assignments is as follows:

1. First round for making assignments:
 - (a) Identify rows successively from top to bottom until a row with exactly one zero element is found. Make an assignment to this single zero by enclosing it with a square bracket $([])$, and then cross off, (\times) , all other zeros in the corresponding column.
 - (b) Identify columns successively from left to right hand with exactly one zero element that has not been assigned. Make an assignment to this single zero by enclosing it with a square bracket $([])$, and then cross off, (\times) , all other zeros in the corresponding row.
2. Second round for making assignments:
 - (a) If a row and/or column has two or more unmarked zeros and one cannot be chosen by inspection, then choose zero element arbitrarily for assignment.
 - (b) Repeat sub-steps (1) and (2) successively until one of the following situations arise

Step D: Optimality Critereon The following are the steps involved in determining the optimality of the solution:

1. If all zero elements in the cost matrix are either marked with square bracket $([])$ or are crossed off (\times) and there is exactly one assignment in each row and column, then it is an optimal solution. The total cost associated with this solution is obtained by adding the original cost elements in the occupied cells.
2. If a zero element in a row or column was chosen arbitrarily for assignment in Step D(1), there exists an alternative optimal solution.
3. If there is no assignment in a row (or column), then this implies that the total number of assignments are less than the number of rows/columns in the square matrix. In such a situation proceed to Step E.

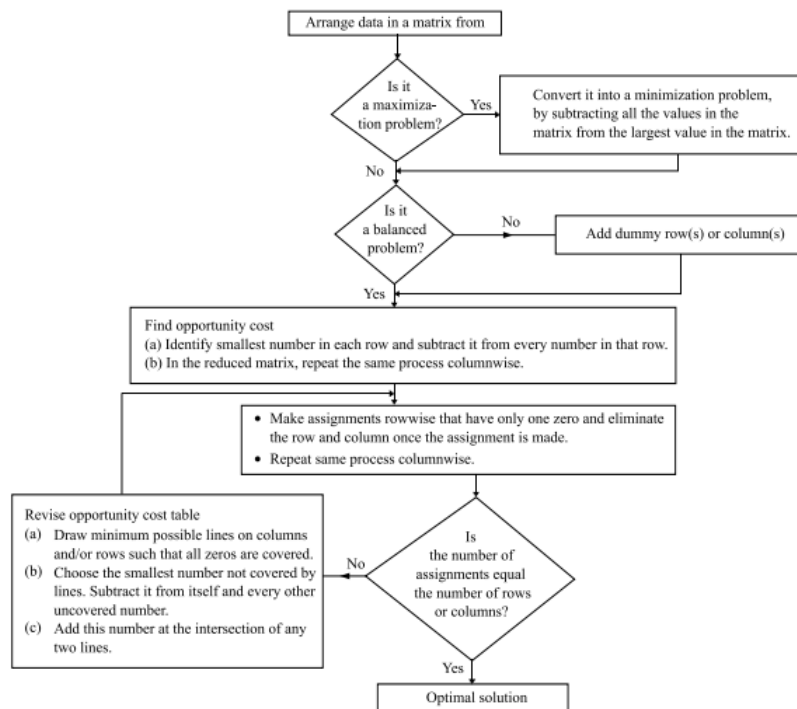
Step E: Revise the opportunity cost matrix Draw a set of horizontal and vertical lines to cover all the zeros in the revised cost matrix obtained from Step C, by using the following procedure:

1. For each row in which no assignment was made, mark a tick (✓)
2. Examine the marked rows. If any zero element is present in these rows, mark a tick (✓) to the respective columns containing zeros.
3. Examine marked columns. If any assigned zero element is present in these columns, tick (✓) the respective rows containing assigned zeros.
4. Repeat this process until no more rows or columns can be marked.
5. Draw a straight line through each marked column and each unmarked row.

Step F: Develop the new revised opportunity cost matrix The steps are as follows:

1. Among the elements in the matrix not covered by any line, choose the smallest element. Call this value k .
2. Subtract k from every element in the matrix that is not covered by a line.
3. Add k to every element in the matrix covered by the two lines, i.e. intersection of two lines.
4. Elements in the matrix covered by one line remain unchanged.

Step G: Repeat Repeat steps C to F until an optimal solution is obtained.



Numerical Example: Time taken by trucks A, B, C, D, and E to complete jobs 1, 2, 3, 4, and 5 are given in the table below and each truck is to be assigned a job. Determine the job assignment for each truck such that the time taken to complete all the jobs is minimum.

Jobs / Truck	A	B	C	D	E
1	13	8	16	18	19
2	9	15	24	9	12
3	12	9	4	4	4
4	6	12	10	8	13
5	15	17	18	12	20

Solution:

Rows 1, 2, 3, 4, and 5 have minimum value of 8, 9, 4, 6, and 12 respectively. Using these values we find out the row reduction matrix.

	A	B	C	D	E	
1	5	0	8	10	11	
2	0	6	15	0	3	
3	8	5	0	0	0	
4	0	6	4	2	7	
5	3	5	6	0	8	

Similarly, all the columns have minimum value of 0. Using these values we find out the column reduction matrix.

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	
1	5	0	8	10	11	
2	0	6	15	0	3	
3	8	5	0	0	0	
4	0	6	4	2	7	
5	3	5	6	0	8	

Then, rows 1, 4, and 5 have only one 0, so we assign these cells with square bracket and cross out any other value of 0 in the respective columns which have been assigned. Similarly column wise, column C has got only one 0, so we assign that cell and cross out the remaining 0 values in the row corresponding to this assigned cell.

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	
1	5	[0]	8	10	11	
2	0×	6	15	0×	3	
3	8	5	[0]	0×	0×	
4	[0]	6	4	2	7	
5	3	5	6	[0]	8	

Since there are cost matrix is of size 5×5 , we require 5 assignments for the solution to be optimal. However, we have only got 4 assignments, so the solution is not optimal. Now, we should draw minimum number of horizontal and vertical lines such that it covers all zeros to revise the cost matrix for optimal solution.

Out of all five rows row 2 is not assigned, so we tick mark this row. In the marked row we have two cells having zero which belongs to column A and D, so we mark these columns. In the marked column A and D we have a cell which consists an assigned 0 for each column, so we mark the rows corresponding to these cells (row 4 and 5). Now we draw horizontal lines on

the rows which has not been marked, and draw vertical lines on the columns which has not been marked.

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	
1	5	[0]	8	10	11	
2	0×	6	15	0×	3	✓
3	8	5	[0]	0×	0×	
4	[0]	6	4	2	7	✓
5	3	5	6	[0]	8	✓
	✓			✓		

For the revised cost matrix we require the value of k , which is the minimum value among the cells which remain uncovered by the lines. In this case, the value of k is 3. Then, we add the value of k to the cell values which lies in the intersection of the drawn lines and subtract the value of k from the cell values which remained uncovered by the lines. Now, the cost matrix has been revised.

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	
1	8	0	8	13	11	
2	0	3	12	0	0	
3	11	5	0	0	0	
4	0	3	1	2	4	
5	3	2	3	0	5	

Again rows 1, 4, and 5 have only one cell having the value of 0 (1-B, 4-A, and 5-D), so we assign these cells with square brackets and cross out the cells having the value of 0 in their respective columns. Similarly column wise column C has only one cell having the value of 0 (3-C), so we assign this cell with square bracket and cross out the cell having the value of 0 in its respective row.

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	
1	8	[0]	8	13	11	
2	0	3	12	0×	[0]	
3	11	5	[0]	0×	0×	
4	[0]	3	1	2	4	
5	3	2	3	[0]	5	

\therefore the size of the matrix = the number of assignments = 5, the solution obtained is the optimal solution.

The optimum assignment is given below:

Job	Truck
1	B
2	E
3	C
4	A
5	D

Also, minimum time = $8 + 12 + 4 + 6 + 12 = 42$.

6.8 Travelling Salesman Problem

The travelling salesman problem asks the following question: "Given a list of cities and the distances between each pair of cities, what is the shortest possible route that visits each city exactly once and returns to the origin city?".

Just consider how a postman delivers the post to the addressee. He arranges all the letters in an order and starts from the post office and goes from addressee to addressee and finally back to his post office. If he does not arrange the posts in an order he may have to travel a long distance to clear all the posts. Similarly, a traveling sales man has to plan his visits. Let us say, he starts from his head office and go round the branch offices and come back to his head office. While traveling he will not visit the branch already visited and he will not come back until he visits all the branches.

There are different types of traveling salesman's problems. One is **cyclic** problem. In this problem, he starts from his head quarters and after visiting all the branches, he will be back to his head quarters. The second one is **Acyclic** problem. In this case, the traveling salesman leaves his head quarters and after visiting the intermediate branches, finally reaches the last branch and stays there. The first type of the problem is solved by Hungarian method or Assignment technique. The second one is solved by Dynamic programming method.

Note: The traveling salesman problem, where we sequence the cities or branches he has to visit is a SEQUENCING PROBLEM. But the solution is solved using the Assignment technique. Hence basically, the traveling salesman problem is a SEQUENCING PROBLEM; the objective is to minimize the total distance traveled.

Numerical Example:

A salesman stationed at city A has to decide his tour plan to visit cities B, C, D, E and back to city A I the order of his choice so that total distance traveled is minimum. No sub touring is permitted. He cannot travel from city A to city A itself. The distance between cities in Kilometers is given below:

Cities	A	B	C	D	E
A	∞	16	18	13	20
B	21	∞	16	27	14
C	12	14	∞	15	21
D	11	18	19	∞	21
E	16	14	17	12	∞

Solution:

Performing row reduction we get the row opportunity cost matrix (ROCM).

	A	B	C	D	E	
A	∞	3	5	0	7	
B	7	∞	2	13	0	
C	0	2	∞	3	9	
D	0	7	8	∞	10	
E	4	2	5	0	∞	

Performing column reduction on the ROCM we get the total opportunity cost matrix (TOCM).

	A	B	C	D	E	
A	∞	1	3	0	7	
B	7	∞	0	13	0	
C	0	0	∞	3	9	
D	0	5	6	∞	10	
E	4	0	3	0	∞	

Now we assign using the same technique used in the Assignment Problem.

	A	B	C	D	E	
A	∞	1	3	[0]	7	
B	7	∞	[0]	13	0×	
C	0×	0×	∞	3	9	
D	[0]	5	6	∞	10	
E	4	[0]	3	0×	∞	

Since we there are only 4 assignments but the matrix is of size 5×5 , we need to modify the opportunity cost matrix for optimum solution. Now, we will draw the minimum number of vertical and horizontal lines to cover all the 0's.

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	
<i>A</i>	∞	1	3	[0]	7	✓
<i>B</i>	7	∞	[0]	13	0×	
<i>C</i>	0×	0×	∞	3	9	✓
<i>D</i>	[0]	5	6	∞	10	✓
<i>E</i>	4	[0]	3	0×	∞	✓
	✓	✓		✓		

For new opportunity cost matrix, $k = \min(3, 6, 7, 9, \infty) = 3$. The new opportunity cost matrix is:

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	
<i>A</i>	∞	1	0	0	4	
<i>B</i>	10	∞	0	16	0	
<i>C</i>	0	0	∞	3	6	
<i>D</i>	0	5	3	∞	7	
<i>E</i>	4	0	0	0×	∞	

Again we assign the cells for the solution.

	A	B	C	D	E	
A	∞	1	0	0	4	
B	10	∞	$0 \times$	16	$[0] \times$	
C	$0 \times$	$[0]$	∞	3	6	
D	$[0]$	5	3	∞	7	
E	4	$0 \times$	0	0	∞	

Here we have a tie situation, that is, more than one solution can be obtained. At first we assign the cell A-D, we get:

	A	B	C	D	E	
A	∞	1	$0 \times$	$[0]$	4	
B	10	∞	$0 \times$	16	$[0] \times$	
C	$0 \times$	$[0]$	∞	3	6	
D	$[0]$	5	3	∞	7	
E	4	$0 \times$	$[0]$	$0 \times$	∞	

We have got 5 assignments, however, the solution is not valid for TSP since there are two cycles, that is, $(A \rightarrow D, D \rightarrow A)$ and $(B \rightarrow E, E \rightarrow C, C \rightarrow B)$.

Now, we assign the cell A-C, we get:

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	
<i>A</i>	∞	1	[0]	$0\times$	4	
<i>B</i>	10	∞	$0\times$	16	[0] \times	
<i>C</i>	$0\times$	[0]	∞	3	6	
<i>D</i>	[0]	5	3	∞	7	
<i>E</i>	4	$0\times$	$0\times$	[0]	∞	

We have 5 assignments, and the solution is valid for TSP since it makes a one complete cycle travelling all cities starting from city A ($A \rightarrow C$, $C \rightarrow B$, $B \rightarrow E$, $E \rightarrow D$, $D \rightarrow A$).

The minimum distance = $18 + 14 + 14 + 12 + 11 = 69$ Km.

Numerical Example:

Given the travel costs below, show how to sequence the travel between the cities so as to minimize the total setup cost per cycle.

Cities	A	B	C	D	E
A	∞	2	5	7	1
B	6	∞	3	8	2
C	8	7	∞	4	7
D	12	4	6	∞	5
E	1	3	2	8	∞

Solution:

ROCM:

	A	B	C	D	E	
A	∞	1	4	6	0	
B	4	∞	1	6	0	
C	4	3	∞	0	3	
D	8	0	2	∞	1	
E	0	2	1	7	∞	

TOCM:

	A	B	C	D	E	
A	∞	1	3	6	0	
B	4	∞	0	6	0	
C	4	3	∞	0	3	
D	8	0	1	∞	1	
E	0	2	0	7	∞	

Now, we assign the cells for the solution.

	A	B	C	D	E	
A	∞	1	3	6	[0]	
B	4	∞	[0]	6	0×	
C	4	3	∞	[0]	3	
D	8	[0]	1	∞	1	
E	[0]	2	0×	7	∞	

We have got 5 assignments, however, the solution is not valid for TSP since there are two cycles, that is, $(A \rightarrow E, E \rightarrow A)$ and $(B \rightarrow C, C \rightarrow D, D \rightarrow B)$.

The next best solution is obtained by assigning minimum non-zero cell in the matrix (In this case 1). At first, we have to assign all 1's and then the 0's. Also we have a tie between the cells D-C and D-E. At first, we assign the cell D-C.

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	
<i>A</i>	∞	[1]	3	6	$0\times$	
<i>B</i>	4	∞	$0\times$	6	[0]	
<i>C</i>	4	3	∞	[0]	3	
<i>D</i>	8	$0\times$	[1]	∞	$1\times$	
<i>E</i>	[0]	2	$0\times$	7	∞	

We have got 5 assignments, however, the solution is not valid for TSP since there are two cycles, that is, $(C \rightarrow D, D \rightarrow C)$ and $(A \rightarrow B, B \rightarrow E, E \rightarrow A)$.

Now, we assign the cell D-E.

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	
<i>A</i>	∞	[1]	3	6	$0\times$	
<i>B</i>	4	∞	[0]	6	$0\times$	
<i>C</i>	4	3	∞	[0]	3	
<i>D</i>	8	$0\times$	$1\times$	∞	[1]	
<i>E</i>	[0]	2	$0\times$	7	∞	

We have 5 assignments, and the solution is valid for TSP since it makes a one complete cycle travelling all cities starting from city A ($A \rightarrow B, B \rightarrow C, C \rightarrow D, D \rightarrow E, E \rightarrow A$).

minimum distance = $2 + 3 + 4 + 5 + 1 = 15$

However, since we violated the assignment rule by assigning two non-zero values, we have to add those penalties in our minimum distance, that is, minimum distance = $15 + 1 + 1 = 17$.

6.9 Review Problems

1. Describe the three main actors involved in freight transportation and distribution.
2. List out the factors that influences the decision of shippers regarding the choice of carrier and transportation mode.
3. Describe the two ways in which shipments can be made.
4. Describe the three distribution structures used in freight transport and distribution.
5. Write the differences among the North-West Corner Method, Least-Cost Cell Method, and Vogel's Approximation Method.
6. A dairy firm has three plants located in a state. The daily milk production at each plant is as follows: Plant 1 : 6 million litres, Plant 2 : 1 million litres, and Plant 3 : 10 million litres.

Each day, the firm must fulfil the needs of its four distribution centres. The minimum requirement of each centre is as follows: Distribution centre 1 : 7 million litres, Distribution centre 2 : 5 million litres, Distribution centre 3 : 3 million litres, and Distribution centre 4 : 2 million litres.

Cost (in hundreds of rupees) of shipping one million litre from each plant to each distribution centre is given in the following table:

Plant / Distribution Center	D_1	D_2	D_3	D_4
P_1	2	3	11	7
P_2	1	0	6	1
P_3	5	8	15	9

Find the initial basic feasible solution for given problem by using following methods:

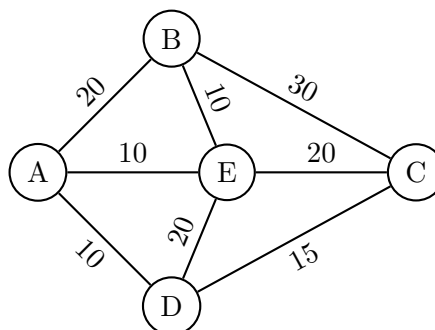
- (a) North-West Corner Method
- (b) Least-Cost Cell Method
- (c) Vogel's Approximation Method

7. A logistics department of a freight company has five trucks with five jobs to be performed. The time (in hours) that each truck takes to perform each job is given in the effectiveness matrix.

Job / Truck	1	2	3	4	5
A	10	5	13	15	16
B	3	9	18	13	6
C	10	7	2	2	2
D	7	11	9	7	12
E	7	9	10	4	12

How should the jobs be allocated, one per truck, so as to minimize the total truck-hours?

8. An Online Delivery organization is planning to deliver packages to 5 cities. Starting from the head office at city A then going round B, C, D and E and then come back to A. Their objective is to minimize the total distance covered. Help them in sequencing the cities. A, B, C, D and E as the shown in the figure. The numbers on the arrows show the distances in Km.



6.9.1 Answers

6 .

- a $x_{11} = 6, x_{21} = 1, x_{32} = 5, x_{33} = 3, x_{34} = 2$, and Total Cost = Rs. 11600
- b $x_{11} = 6, x_{22} = 1, x_{31} = 1, x_{32} = 4, x_{33} = 3, x_{34} = 2$, and Total Cost = Rs. 11200
- c $x_{11} = 1, x_{12} = 5, x_{24} = 1, x_{31} = 6, x_{33} = 3, x_{34} = 1$, and Total Cost = Rs. 10200

7 .

Job	Truck	Time
A	2	5
B	1	3
C	5	2
D	3	9
E	4	4
		Total = 23

- 8 . The shortest route is $(A \rightarrow E, E \rightarrow D, D \rightarrow C, C \rightarrow B, B \rightarrow A)$ and the distance travelled = 95 Km.

Chapter 7

Fuel Consumption and Emission Models

Chapter 8

Planning for New Transport Organization