# Estimation

SC Agarkar

VNBRIMSThane

## Preamble

- Estimation is an everyday phenomenon as we have to estimate many things in day to day activities.
- Managers need to estimate. The outcomes of the estimates could be serious. For example, sell of houses, passenger in the trip, loan takers from the bank, etc.
- Statistical inference is the branch in statistics concerned with using probability concepts to deal withuncertainityin decision making.

# Types of Estimates

- We can make two types of estimates: Point estimate and interval estimate.
- Point estimate is a single number that is used to estimate an unknown population parameter. For example, department head estimates that there would be 120 students next year for a course.
- An interval estimate is a range of values used to estimate a population parameter. For example, the department head estimates that the enrolment of students next year would be between 100 to 140.

## Estimator and estimates

- An estimator is a sample statistic used to estimate a population parameter. For example, a sample mean (x) can be an estimator of the population mean (μ).
- An estimate is a specific observed value of a statistic. For example, suppose we calculate the mean odometer reading of used taxis as 98,000kilometers. Using this value we can estimate that mileage of all the fleet of used taxis would be 98,000kilometers.

# Tabular summary

| Population | Parameter | Sample Statistic | Estimate |
|---|---|---|---|
| Employees in  a furniture factory | Mean turnover per year | Mean turnover for period of one month | 8.9% turnover per year |
| Applicants for town manager | Mean formal education years | Mean formal education of every fifth applicant | 17.9 years of formal education |
| Teenagers in a given community | Proportions whohave criminalrecords | Proportion of a sample of 50 teenagers who have criminal record | 2 % havecriminalrecords |

# Point Estimates

- The proportion of units that have a particular characteristics in a given population is symbolized as p. if we know the proportion of units in a sample that has same characteristics then we can use it as an estimator of p.
- Suppose the management wishes to estimate the number of carton of syringes that will arrive damaged. They can check a sample of 50 cartons and find that proportions damaged is 0.8. This value can be used to estimate the proportion of damaged cartons in the entire population (0.8).

# Interval Estimate

- An interval estimate describes a range of values within which a population parameter is likely to lie.
- A marketing research director wants to know the average life of a car battery made by his factory. He selects a random sample of 200 batteries and finds that a mean life is 36 months. Then he concludes that this must be true for the entire population.
- SE in this case would be $\sigma \div \sqrt{n}$= 0.707 months assuming that the SD of the population is 10. Thus actual mean life of batteries may lie between 35.292 to 36.77 months (i.e. 35 to 37 months).

# Characteristics of normal curve

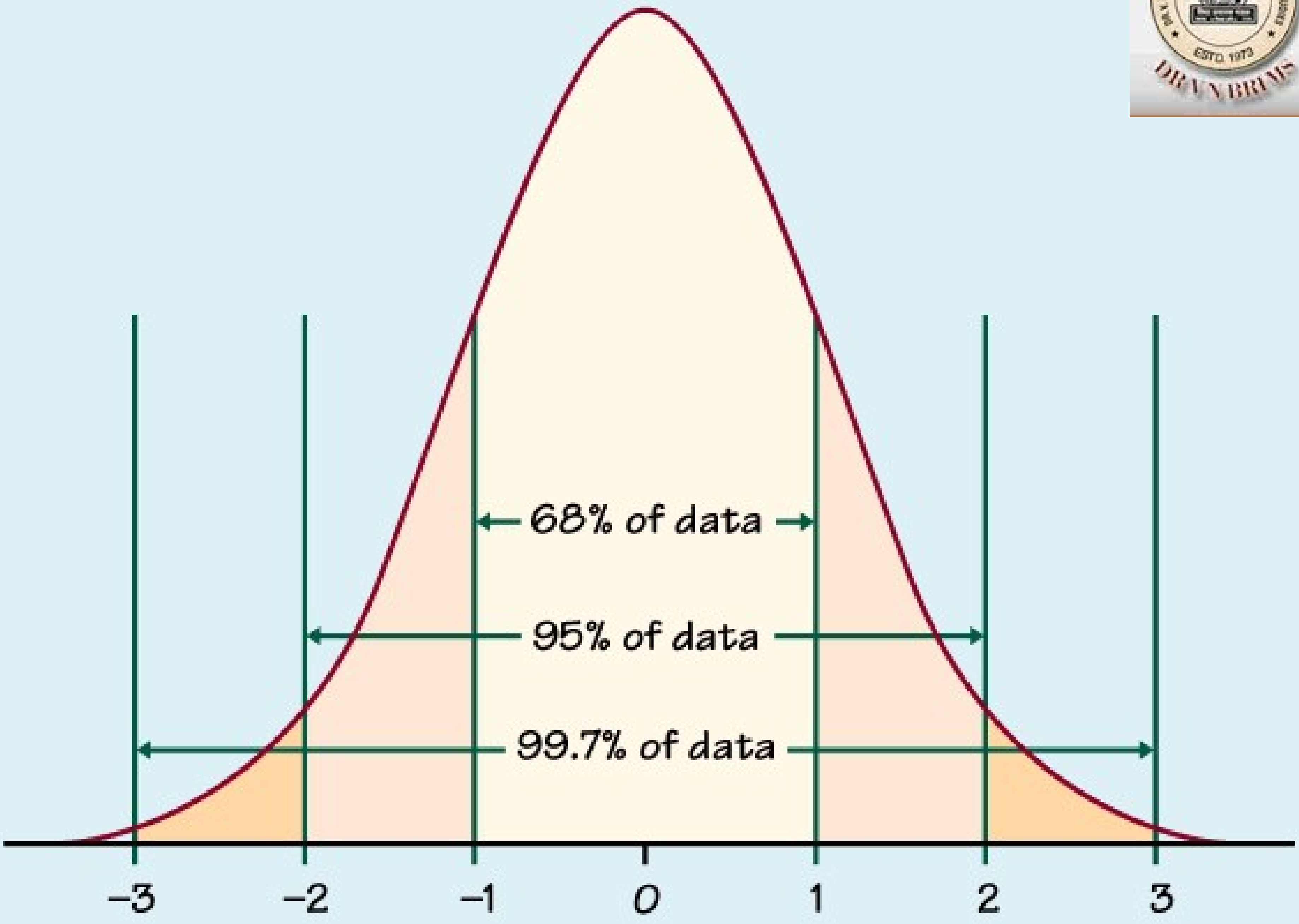
- The curve has a single peak, it is unimodal.
- The mean of the population lies at the centre.
- The mean, median and mode all coincide.
- Two tails extends indefinitely and never touch the horizontal axis.
- To define a particular normal probability distribution we need only two parameters, the mean and the standard deviation.

# Characteristics of normal curve

- The curve has a single peak, itis unimodal.
- The mean of the population lies at the centre.
- The mean, median and mode all coincide.
- Two tails extends indefinitely and never touch the horizontal axis.
- To define a particular normal probability distribution we need only two parameters, the mean and the standard deviation.

# Areas under normal curve

- Approximately 68 percent of all values in a normally distributed population lie within 1 SD (plus or minus) from the mean.
- Approximately 95.5 percent of all the values in a normally distributed population lie within 2 SD (plus or minus) from the mean.
- Approximately 99.7 percent of all the values in a normally distributed population lie within 3 SD (plus or minus) from the mean.
- Every time it is not necessary to look at the area under the normal curve. We can refer to statistical tables constructed showing the portions of the area covered.


68% of data
95% of data
99.7% of data
−3
−2
−1
0
1
2
3

# Formula for measuring distances under the normal curve

- The value of z is derived from the formula $z = (x-\mu)/\sigma$
- $x$ = value of random variable with which we are concerned
- $\mu$= mean of the distribution of this random variable
- $\sigma$= standard deviation of the distribution
- $z$= number of standard deviations from x to the mean of this distribution.
- Normally distributed random variable take on many different units (Rs, meter, degrees, etc). To avoid these units we talk in terms ofunitlessentity called z.

# Interval estimates and confidence intervals

- According to the z value table plus and minus 1.64 SE includes about 90 percent of the area under the curve (it includes .4495 of the area on either side of the mean in a normal distribution). Similarly, plus and minus 2.58 SE includes about 99 percent of the area or 49.5 percent on each side of the mean.
- In statistics the probability that we associate with interval level estimate is called the confidence level. The confidence interval is the range of estimate we are making.
- Upper limit= x+1.64SE and lower limit= x-1.64SE for confidence level of 90 percent.

# Confidence leveland confidence interval

- A high confidence level signify a high degree of accuracy in estimate. In practice, however, high confidence levels produce large confidence intervals and such large intervals are not precise.
- Confidence level of delivering an item in a short time is low. Instead it is high if the time interval is large.
- Confidence level of finding a medical shop within a radius of 10kilometersis very high.

# Sampling and confidence interval estimation

- In a battery example discussed earlier if we say "we are 95 percent confident that the mean battery life of the population lies between 30 and 42 months.
- This statement does not mean that the chance is .95 that the mean life falls between the limits. Instead, it means that if we select many random samples of this sample size and if we calculate a confidence interval for each of these samples, then in about 95 percent of these cases, the population mean would lie within that interval.

# Calculating interval estimates of themean from large samples

- In this case Standard error is given by SE=σ/√n
- The wholesaler of wiper blades have selected a sample of 100 blades and calculated sample mean as 21 months and SD of the population as 6 months. Hence SE=6/√100 = 0.6 months
- At the 95 percent confidence level will include 47.5 percent of the area on either side of the mean of the sampling distribution. We find that .475 of the area under the normal curve is contained between the mean and a point 1.96 SE.
- Upper limit= x+1.96 SE= 22.18 months, Lower limit= x-1.96 SE= 19.82 months
- We can report that we estimate the mean life of the population of wiper blades to be between 19.82 and 22.18 months with 95 percent confidence.

# When the population SD is unknown

- A social service agency is interested in estimating the mean annual income of 700 families in a locality. By taking a sample of 50 we get sample mean $4800 and SD $950.
- SE =σ/√n *√N-n/N-1

- Substituting the values we get SE= $129.57
- At 90 percent confidence level
- Upper confidence limit: x+1.64SE= $5,012.50
- Lower confidence limit: x-1.64SE= $4,587.50
- We report that with 90 percent confidence we estimate the average annual income of all 700 families living in the locality falls between $4,587.50 and $5,012.50.

# Calculating interval estimates of the proportion from large sample

- SE of proportion=$\sqrt{pq/n}$
- n= number of trials, p= probability of success and q = probability of a failure (1-p).
- An industry wants to estimate what proportion of employees prefer to provide their own retirement benefit in lieu of company sponsored plan.
- Taking a random sample of 75 employees they have found out that .4 of them are interested in providing their own benefit. Thus n=75, p= .4 and q= 1-.4=.6.

# Continued

- SE= =√pq/n = .057
- A 99 percent confidence level would include 49.5 percent of the area on either side of the mean in the area under the normal curve. This refers to 2.58 standard error from the mean.
- Upper confidence limit = p+2.58 SE= .547
- Lower confidence limit = p-2.58 SE = .253
- Thus, we estimate from the sample of 75 employees that with 99 percent confidence we believe that the proportion of the total population of employees who wish to establish their own retirement plans lies between .253 and .547.

# Summary Table

| | When the populationis finite | When the population is |
|---|---|---|
| Estimatingµ(the population mean)whenσ(population SD) is known | Upperlimit:x+zσ÷√n* √N-n/N-1<br>Lower limit: x-zσ÷√n* √N-n/N-1 | Upper Limit=x+zσ÷√n<br>Lower Limit=x-zσ÷√n |
| Whenσ(the population SD) is not known<br>When n(sample size) is larger than 30 | Upperlimit:x+zs÷√n* √N-n/N-1<br>Lower limit: x-zs÷√n* √N-n/N-1 | Upper Limit=x+zσ÷√n<br>Upper Limit=x-zσ÷√n |
| When n (sample sizeis less than 30 | Upperlimit:x+ts÷√n* √N-n/N-1<br>Lower limit: x-ts÷√n* √N-n/N-1 | Upper Limit=x+ts÷√n<br>Upper Limit=x-ts÷√n |
| Estimating p (the population proportion)when n (sample size) is larger than 30 | Upperlimit:x+zσ÷√n* √N-n/N-1<br>Lower limit: x-zσ÷√n* √N-n/N-1 | Upper Limit=p+zSE<br>Upper Limit=p-zSE |


20

# Thank you.Nowestimatethe number of birds next morning in the garden?