CENTRAL LIMIT THEOREM

- Specifies a theoretical distribution
- Formulated by the selection of all possible random samples of a fixed size n
- A sample mean is calculated for each sample and the distribution of sample means is considered

DATA FOLKZ
CATAPULT DATA LEADERS

SAMPLING DISTRIBUTION OF THE MEAN

- The mean of the sample means is equal to the mean of the population from which the samples were drawn.
- The variance of the distribution is σ divided by the square root of n. (the standard error.)

STANDARD ERROR

Standard Deviation of the Sampling Distribution of Means

$$\sigma_x = \sigma / \sqrt{n}$$

HOW LARGE IS LARGE?

- If the ~~sample is **normal**, then the sampling distribution of $\bar{x}$ will al~~so be normal, no matter what the sample size.
- When the sample population is approximately **symmetric**, the distribution becomes approximately normal for relatively small values of *n*.
- When the sample population is **skewed**, the sample size must be **at least 30** before the sampling distribution of $\bar{x}$ becomes approximately normal.

EXAMPLE

- A certain brand of tires has a mean life of 25,000 miles with a standard deviation of 1600 miles.

- What is the probability that the mean life of 64 tires is less than 24,600 miles?

EXAMPLE CONTD

- The sampling distribution of the means has a mean of 25,000 miles (the population mean)

μ = 25000 mi.

- And a standard deviation (i.e.. standard error) of:

1600/8 = 200

EXAMPLE CONTD

- Convert 24,600 mi. to a z-score and use the normal table to determine the required probability.

  $z = (24600-25000)/200 = -2$

  $P(z< -2) = 0.0228$

or 2.28% of the sample means will be less than 24,600 mi.

ESTIMATION OF POPULATION VALUES

- Point Estimates
  - A single value given as an estimate of a parameter of a population
- Interval Estimates
  - An interval within which the value of a parameter of a population has a stated probability of occurring.

CONFIDENCE INTERVAL ESTIMATES FOR LARGE SAMPLES

- The sample has been randomly selected
- The population standard deviation is known or the sample size is at least 30.

CONFIDENCE INTERVAL ESTIMATE OF THE POPULATION MEAN

$$\bar{X} - z\frac{s}{\sqrt{n}} \le \mu \le \bar{X} + z\frac{s}{\sqrt{n}}$$

$\bar{X}$: sample mean

s: sample standard deviation

n: sample size

DATA FOLKZ
CATAPULT DATA LEADERS

clt

EXAMPLE

- Estimate, with 95% confidence, the lifetime of nine volt batteries using a randomly selected sample where:

$X = 49$ hours

$s = 4$ hours

$n = 36$

EXAMPLE CONTINUED

- Lower Limit: 49 - (1.96)(4/6)
  49 - (1.3) = 47.7 hrs

- Upper Limit: 49 + (1.96)(4/6)
  49 + (1.3) = 50.3 hrs

We are 95% confident that the mean lifetime of the population of batteries is between 47.7 and 50.3 hours.

DATA FOLKZ
CATAPULT DATA LEADERS

CONFIDENCE BOUNDS

- Provides a upper or lower bound for the population mean.
- To find a 90% confidence bound, use the z value for a 80% CI estimate.

DATA FOLKZ
CATAPULT DATA LEADERS

EXAMPLE

- The specifications for a certain kind of ribbon call for a mean breaking strength of 180 lbs. If five pieces of the ribbon have a mean breaking strength of 169.5 lbs with a standard deviation of 5.7 lbs, test to see if the ribbon meets specifications.
- Find a 95% confidence interval estimate for the mean breaking strength.

clt

15

- You sample 36 apples from your farms harvest of 200,000 apples. The mean weight of sample is 112g with 40g samples standard deviation. What is the probability that the mean weight of all 200,000 apples is between 100 and 124g?

DATA FOLKZ
CATAPULT DATA LEADERS

- I don’t know how population is distributed, mean and sd is not known., and I know sample is coming from normal distribution. How should we solve?
- What should we find out? (is population mean is between 100 or 124?)
- We have to find if mean = x+/- 12, similar is x= mea+/-12