# Sampling and Sampling Theorem

## A Short Introduction

by Brad Morantz
brad@cognitive-decisions.com

# Example

- Makebucks Coffee Company is developing an automatic coffee vending machine
- Need to test how full it fills the cup
  - Variation because of other water usage and system variables
  - Too full and it spills out
    - Waste of coffee
    - Spills on floor and can cause law suit
  - Not full enough
    - People kick machine causing damage
    - People get irritated, feeling cheated, buy elsewhere

# Solution

- Put one of these machines in the hall
- Randomly through the day send someone to go get 4 cups of coffee from the machine.
- Measure the sample of the 4 cups of coffee
- Compile *statistics* of all of the samples over several days or weeks
- Calculate a confidence interval for average amount of coffee in a cup

# Other Sampling Uses

- Surveys
- Quality control
- Product design
- Warranty
- Detective work
- Basically, take a sample that is representative of the population, and use that to describe the population or to perform some test

# Definitions

- Sampling frame – definition of the population
- Law of large numbers: if add up sufficiently large number of random variables, the sum will be normally distributed – Laplace

# Central Limit Theorem

- If there is a population, of any distribution
- And you take a number of samples from it
  - The mean of the samples are approximately normally distributed
  - The mean of the samples is an estimator of the population mean
  - The sample variance of the samples is an estimator of the population variance
- A larger number of samples gets these estimators closer to the true values
- Laplace 1776

# Sampling

- Have population of unknown distribution
- Take many samples from population
- If take many samples, distribution of the sample means will be normally distributed
- Use this to infer upon the population

# Types of Sampling

- Simple random
  - All have equal probability of being sampled
- Systematic
  - Method for taking sample, like every tenth one
- Cluster
  - Organizing into clusters, then randomly select cluster to sample
- Stratified
  - Population split into strata, and then those are sampled

# Non-Probability Sampling

- Judgment sample
  - Selected by expert
- Convenience sample
  - Done in a manner that is easier to do
  - Often biased

# Another Example

- Badpoor tire company wants to know how many miles to warrant a certain tire
- If they test the tire, they will have nothing to sell
- If they claim too much, then they will get too many warranty claims and bad rap.
- If they warrant too little, they could have either raised the price or put on less rubber

# Solution to Tire Example

- Sell tires to employees at good price
  - They put on either 2 or 4 at a time
- Go out to employee parking lot and measure tires and read odometers
  - Do this on regular basis
- From this data, calculate confidence interval
- Adjust warranty & price of tire

# Sample Statistics

- $\bar{X}$ or Xbar is the sample mean

  And is an unbiased estimator of μ

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

- $S^2$ is the sample variance

  And is an unbiased estimator of $\sigma^2$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

Notice that we use (n-1)

- $\sigma_x$ is the standard error of the mean

$$\sigma_x = S^2 / \sqrt{n}$$

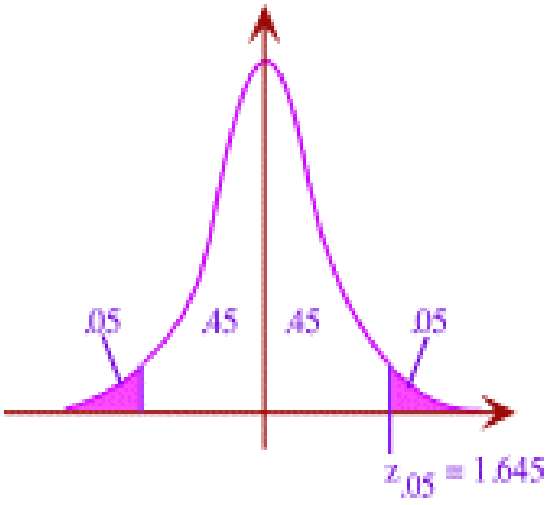
# Confidence Interval

- In previous examples, suppose you now have sample data, and have calculated the sample statistics.
- How do you calculate a range that you expect the tires to wear, or how much coffee in a cup?
- Answer:  Calculate the confidence interval or CI
  - They are typically 90%, 95%, or 99% interval

# How to Calculate CI

- Lower limit is Xbar – Z * $\sigma_x$
- Upper limit is Xbar + Z * $\sigma_x$

$$\bar{Y} - Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{Z}_{\alpha 1}\frac{\sigma}{\sqrt{n}}$$




The lower equation is the same as the two upper equations
Decide on what percent confidence interval you need
Look up the Z values (or t values) in the table or computer
Plug in the values from your samples
Calculate the upper and lower limit

# Hypothesis Testing

- Well, do the tires have a mean wear life of 40,000 miles?
- Do the samples that we took show this with any confidence.
- Most companies use 95% confidence
- We need to test that hypothesis
- Calculate the 95% CI for the sample tires on employees cars
- Does that include the desired 40,000 miles?