I. Introduction

II. Statistical and Notational Preliminaries

A. Introduction

B. Elements of Statistical Theory

1. Prerequisites:

❖ Random Variables

❖ Distributions: Population vs. Sample

❖ Descriptive Measures

Today {

❖ Estimation

❖ Sampling distributions.

❖ Inference: CIs and Hypothesis Testing.

}

❖ ***Estimation***

- You want a "best guess" about a ***population parameter***, say, μ.
- How?

More details/definitions:

- ***Population Parameters***:
- ***Estimator***:
- ***Estimate***:
- ***Sample statistic:***

❖ Sampling Distributions

- An ***estimator*** is a ***random variable***. Explain.

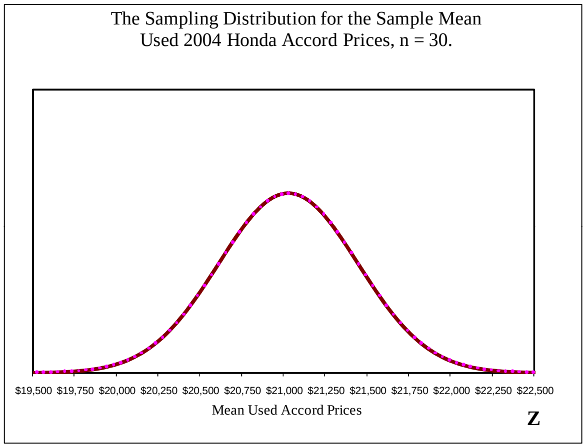
Review: What are the formulas for descriptive measures of the following distributions?

- Population
- Sample
- Sampling Distribution

(Assume the sampling distribution is ***normally distributed***.)

Experiment: Suppose we repeatedly sample from our population of used 2004 Honda Accords. What does the sampling distribution for the sample mean look like?

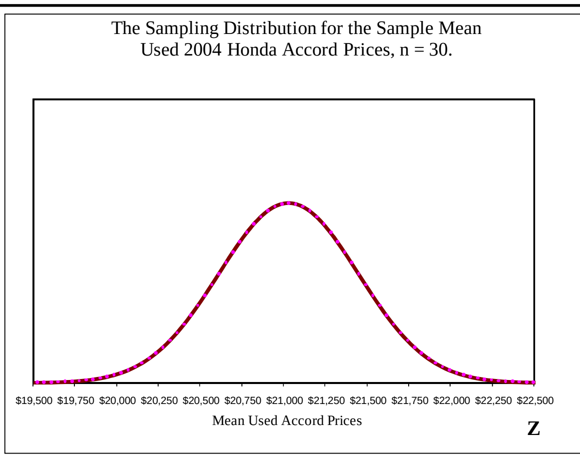
- The shape of the population distribution is not known to be normal. What should we do?
- Characterize the sampling distribution:
  1. Shape:
  2. Center:
  3. Variation:



Example: Sampling Distribution Probability interval.

- Determine the lower and upper limits for a 90% interval for the sample mean.
- Symmetric interval around the true mean.
- Contains 90% of the area under the sampling distribution between these values.

$$P(\mu_Y - z_{\alpha/2} \cdot \sigma_{\bar{Y}} < \bar{Y} < \mu_Y + z_{\alpha/2} \cdot \sigma_{\bar{Y}}) = 0.90$$



❖ Inference: CIs and Hypothesis Tests

- Confidence Intervals – Before drawing your sample of size 30:

$$P(\bar{Y} - \underbrace{z_{\alpha/2} \cdot \sigma_{\bar{Y}}}_{\text{Margin of Error}} < \mu_Y < \bar{Y} + z_{\alpha/2} \cdot \sigma_{\bar{Y}}) = 1 - \alpha$$

- Before sampling, there is a 0.90 probability of choosing a sample that will result in a 90% CI that contains the true population mean.

Example: Confidence Interval Estimation.

- Draw a sample.
- Calculate a point estimate:
- Which CI formula should you use?

❖ ***Inference***

♦ ***CI*** when σ_Y is known.

Procedure:

- Choose the level of confidence:
- Determine the correct $z_{\alpha/2}$ value.
- Select a sample – estimate
- CI estimator:
- We can be $(1 - \alpha)$x100% confident that the sample we draw will result in an interval that contains the true value, μ_Y .

Why are we 90% confident that the true population parameter is contained in our interval?

- The sampling distribution shows that 90% of our sample means will fall within 1.645 standard errors of μ_X.
- All of these sample means will result in CIs that contain μ_X.
- Thus, we are 90% confident that our mean is one of those!

♦ ***CI*** when σ_Y is NOT known.

Procedure:

- Choose the level of confidence:
- Determine the correct ___ **values**
- Select a sample – estimate:
- CI estimator:
- We can be $(1 - \alpha)$x100% confident that the sample we draw will result in an interval that contains the true value, μ_Y .

t-distribution:

- The ***t-distribution*** - "fatter tails" than the ***z***.
- ***More conservative –*** wider intervals given $(1–\alpha)$
- ***More uncertainty*** - we also estimated $\sigma_{\mathbf{Y}}$.
- Shape depends on ***degrees of freedom***:

  df = (n – 1).
- ***Center is zero –*** standardized distribution.
- As ***df*** $\rightarrow \infty$, *t-distribution* converges to *z-distribution*.