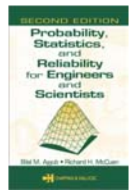
CHAPTER

Probability, Statistics, and Reliability for Engineers and Scientists

Second Edition

FUNDAMENTALS OF STATISTICAL ANALYSIS

• A. J. Clark School of Engineering •Department of Civil and Environmental Engineering



8a

Probability and Statistics for Civil Engineers

Department of Civil and Environmental Engineering

University of Maryland, College Park

CHAPMAN HALL/CRC

Introduction

- Definition
  - **Statistics:** is the estimation of certain parameters (i.e., mean, *COV*, distribution type) needed to quantify uncertainty and to describe the probability functions.

Introduction

- Estimation of Parameters of a Distribution
  - Once the distribution type of a random variable is assumed, it is necessary to define it uniquely by evaluating its parameters.
  - Some distributions have only one parameter, while others have multiple.

Introduction

- Examples: Distribution Parameters
  - Normal Distribution:
    - Two parameters, μ_X and σ_X
  - Lognormal
    - Two parameters, μ_Y and σ_Y (or λ and ζ)
  - Poisson
    - One parameter, λ
  - Beta
    - Four parameters, *a*, *b*, *q*, and *r*

Introduction

- Sampling
  - Values of random variables obtained from sample measurements are used to estimate the parameters of a distribution.
  - They are also used to in making important engineering decision.
  - For example, in determining the maximum wind speed for the design of a tall building, past records of measured wind velocities near the building site are important.

Introduction

- Sampling
  - Samples of river water are collected to estimate the average level of a pollutant in the entire river at that location.
  - The average of sample measurements of the compressive strength of concrete collected during the pouring of a large concrete slab, such as a deck of of a parking garage, is used to help decide whether or not the deck strength met the specifications.

Introduction

- Sampling
  - The estimated mean for a random variable is considered by itself to be a random variable, because different samples about the random variable can produce different estimated mean values.
  - Hence, ***randomness*** in the estimated mean.

Introduction

- Sampling
  - When a sample of *n* measurements of a random variable is collected, the *n* values are not necessarily identical.
  - The sample is characterized by variation.
  - If two different samples of 5 measurements of concrete strength are collected, their mean values would not necessarily be identical.

Introduction

- <u>Example:</u> Strength of Concrete (psi) in a Parking Garage Deck

	Sample 1	Sample 2
	3250	3650
	3610	3360
	3460	3328
	3380	3420
	3510	3260
Mean	**3442**	**3404**
StDev	**135.9**	**149.3**

Introduction

- <u>Example:</u> Strength of Concrete (psi) in a Parking Garage Deck
  - Assume that the building code requires a mean compressive strength of **3500 psi**.
  - Since the mean of 3442 psi of Sample1 is less than 3500 psi
  - **<u>Should we conclude that the garage deck dose not meet the specifications?</u>**

Introduction

- Example: Strength of Concrete (psi) in a Parking Garage Deck
  - Unfortunately, decision making is not that simple.
  - If a third sample of 5 measurements had been randomly collected from other locations on the garage deck, the following values as just likely to have been obtained:
    3720, 3440, 3590, 3270, and 3610 psi.
  - This sample would have different mean and different standard deviation as shown next.

Introduction

- Example: Strength of Concrete (psi) in a Parking Garage Deck

	Sample 1	Sample 2	Sample 3
	3250	3650	3720
	3610	3360	3440
	3460	3328	3590
	3380	3420	3270
	3510	3260	2610
Mean	**3442**	**3404**	**3526**
StDev	**135.9**	**149.3**	**174.4**

Introduction

- Example: Strength of Concrete (psi) in a Parking Garage Deck

  The third sample (Sample 3) produces a mean of 3526 psi and standard deviation of 174.4 psi.

  In this case, the mean value is greater than the specified value. The question now arises:

  Can we conclude that the concrete is of an adequate strength?

  Unfortunately, we cannot conclude with certainty that the strength is adequate.

Introduction

- Example: Strength of Concrete (psi) in a Parking Garage Deck

  "The fact that different samples lead to different means is an indication that we cannot conclude that the design specification is not met just because the sample mean is less than the design standard."

Introduction

- Example: Strength of Concrete (psi) in a Parking Garage Deck
  - A systematic decision process is needed to take into account the variation that can be expected from one sample to another.
  - The decision process must also capable to reflect the *risk* of making incorrect decision.
  - This decision making can be made using hypothesis testing.

CHAPTER 8a. FUNDAMENTALS OF STATISTICAL ANALYSIS

Sample and Population

- Sample Parameters

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}$$

- Population Parameters

For normal distribution :

μ_X

σ_X

Sample and Population

- The data that are collected represent sample information but it is not complete by itself.
- Predictions are not made directly from the sample.
- The intermediate step between sampling and prediction is the identification of the underlying ***population***.

Sample and Population

- The ***sample*** is used to identify the ***population***.
- Then, the ***population*** is used to make prediction or decision.
- The *sample-to-population-to-predictions* sequence is true for both univariate, bivariate, and multivariate methods.

Sample and Population

- Population Models
  - A known model or function is often used to represent the population.
  - The normal and logmormal distributions are commonly used to model the population for a univariate problem
  - For bivariate and multivariate prediction, linear and power model are assumed.

Sample and Population

- Example: Population Models
  - Linear Model

$$\hat{Y} = a + bX$$

  - Power Model

$$\hat{Y} = aX^b$$

Sample and Population

– When using a probability function to represent the population, it is necessary to estimate the parameters

- Normal Distribution

  Estimate : μ and σ

- Lognormal

  Estimate : λ and ζ

- Exponential

  Estimate : λ