56:295-001 Multivariate Statistics and Advanced Quality Control

Instructor: Yong Chen

M/W 5:00 - 6:15

1

## About Me

- Ph. D., 2003, Industrial and Operations Engineering, University of Michigan
- Assistant professor, Dept. of Mechanical and Industrial Engineering
- Research area: Quality & reliability engineering; Complex sensor systems

2

## Availability

- Office hours: W 1:30 - 2:30
- Or by appointment

3

## Textbook & Lecture Notes

- Applied Multivariate Statistical Analysis, 5e, by R. A. Johnson and D. W. Wichern (on reserve in Math/Sci library)
  - Reference text: Introduction to Statistical Quality Control, 5e, by D. C. Montgomery, not required, but highly recommended (4e on reserve in Eng. library)
- Lecture notes: To be posted on course website weekly

4

## Course Website and Email List

- Website
  - http://www.engineering.uiowa.edu/~yongchen
- Email List
  - Will send a test email to the class. Let me know if the test email is not received.

5

## Computing

- Matlab: computation
- Minitab: statistics/quality control

6

## Attendance Policy

- Attendance is required
- Attendance is a factor in final grading

7

## Prerequisites

- Basic knowledge on statistics/probability, calculus and linear algebra
- E.g., I expect you to be familiar with the concept of
  - Mean, variance, covariance/correlation, joint/marginal density functions, conditional distribution/expectation
  - Sample mean, variance, covariance/correlation
  - Normal distribution, $t$-distribution, $F$-distribution, $\chi^2$ distribution
  - Hypothesis testing, confidence interval
  - Eigenvalues and eigenvectors
  - Positive definite matrix, quadratic form
- Quality control background, will help, but are not required

8

## Grading Policy

- Homework 25%
- Group presentation&report on a research paper 15%
- Exam 1 (October 17) 30%
- Exam 2 (December 7) 30%

  - Homework should be handed in by the end of class on the due date.
  - Only a randomly selected subset of problems (~50%) will be graded. Solutions to all problems will be posted.
  - A pool of research papers will be provided about mid-term. Each group of two students selects a paper and give a presentation on it in the last two lectures before Exam 2.



## Difference from Statistics Course

- Multivariate analysis has broad applications in areas such as biology, social science, and imaging. The course is not intended to cover all important topics in multivariate statistics
- Focus on topics exhibiting potentials in **quality control** applications
- Usage of generic statistical software packages for multivariate analysis will not be covered
- Most multivariate statistics topics are followed by the corresponding quality control techniques/applications

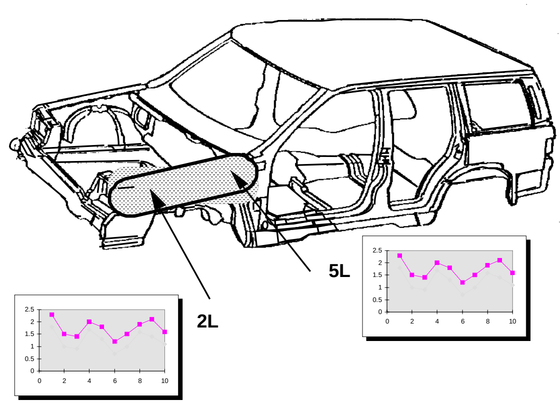
# Random Vectors and Random Sampling



## Examples of Multivariate Data

- Bookstore example
- Quality control examples

Cowl side reinforcement panel I/O variation



## Purpose of Multivariate Analysis

- In general, multivariate analysis can be used for
  - Reduction of data dimension
  - Grouping data
  - Prediction
  - Hypothesis testing
  - Investigation of dependence among variables
- In quality control applications, multivariate analysis is used for
  - Monitoring process changes
  - Diagnosis of root causes of quality problems

13

## Descriptive Statistics

- Much of the information contained in the data can be assessed by calculating certain summary numbers.
- Sample mean
- Sample variance and standard deviation
- Sample covariance/correlation: measures of ***linear*** association

14

## Data Displays

- It is generally impossible to simultaneously plot all the measurements made on several variables and study the configurations
- Plots of individual variables and plots of pairs of variables can still be very informative
  - Histogram
  - Scatter plot
  - Boxplot
- Others (Stars, Chernoff faces, Control Charts, ect.)

15

## Random Vectors and Matrices

- A **random vector** is a vector whose elements are random variables.
- A **random matrix** is a matrix whose elements are random variables.
- Expectation (population mean): $E[\mathbf{X}+\mathbf{Y}]=E[\mathbf{X}]+E[\mathbf{Y}]$, $E[\mathbf{AXB}]=\mathbf{A}E[\mathbf{X}]\mathbf{B}$, where **X** and **Y** are random vectors/matrices of the same dimension, and **A** and **B** are matrices of constants.
- (Population) variance-covariance matrix and population correlation matrix
- Statistically independent random variables

16

## Linear Combinations of Random Variables

- $E[\mathbf{CX}]=\mathbf{C}E[\mathbf{X}]$
- $\mathrm{Cov}[\mathbf{CX}]=\mathbf{C}\mathrm{Cov}[\mathbf{X}]\mathbf{C}^T$
- Example 2.15 of J&W

17

## Random Samples

- A single multivariate observation is the collection of measurements on $p$ different variables taken on the same item. If $n$ observations have been obtained, the entire data set can be placed in an $n \times p$ matrix $\mathbf{X}$
- The row vectors of $\mathbf{X}$, denoted by $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, are said to form a **random sample** from a common density function $f(\mathbf{x})=f(x_1, x_2, \ldots, x_p)$. We say the data are from a sample of size $n$ from a $p$-variate "population". The sample consists of $n$ observations, each of which has $p$ components.
- The measurements of the $p$ variables in a *single* observation, such as $\mathbf{X}_j=[\mathbf{x}_{j1}, \ldots, \mathbf{x}_{jp}]$, will usually be correlated.
- The measurements from *different* observations must, however, be independent.

18