# Statistical Preliminaries and Mathematical Notation

## Introduction

In this lecture we want to introduce the standard mathematical and statistical notation used in most text books for the following topics.

- Vector and Matrix Notation
- Random Variable Notation
- The Term “Distribution”
- Mathematical Expectation
- Expectation, Variance, and Covariance as Mathematical Operators

## Vector and Matrix Notation

- A *scalar* is a number. Scalars are represented by lower case letters from the beginning of the alphabet such as $a$, $b$, $c$, etc.

- A *vector* is a $n \times 1$ array defined with the mathematical operations of addition and multiplication. The standard convention is for all vectors to be *column vectors*, i.e. they are "long" with $n$ rows and 1 column. Vectors are represented by bold face lower case letters, frequently from the end of the alphabet, such as $\mathbf{x}$, $\mathbf{u}$, and $\mathbf{v}$, but $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ could also represent vectors. The $i$th entry of a vector $\mathbf{u}$ is denoted by $\mathbf{u}[i] = u_i$.

- A *matrix* is a $n \times m$ array defined with the mathematical operations of addition and multiplication. Matrices are represented by upper case letters such as $\mathbf{A}$, $\mathbf{W}$, $\mathbf{X}$, etc. The $(i, j)$th entry of a matrix $\mathbf{A}$ is denoted by $\mathbf{A}[i, j] = a_{ij}$

# The Vector and Matrix Transpose

- The transpose of a $(n \times 1)$ column vector $\mathbf{a}$ is the $(1 \times n)$ row vector $\mathbf{a}^T = [a_1 \cdots a_n]$. Sometimes the transpose $\mathbf{a}^T$ is denoted by $\mathbf{a}'$.

- The transpose of a $(n \times m)$ matrix $\mathbf{A}$ is the $(m \times n)$ matrix $\mathbf{A}^T$ where $\mathbf{A}[i, j] = \mathbf{A}^T[j, i]$. When a matrix is transposed, the rows become the columns and the columns become the rows.

- It is preferred to use the $T$ notation $\mathbf{a}^T$ instead of the "prime notation" $\mathbf{a}'$.

## Random Variable Notation

- Random variables are denoted by capital letters from the end of the alphabet such as $U$, $V$, $X$, $Y$, or $Z$.

- The *observed value* of a random variable is denoted by the lower case counterpart $u$, $v$, $x$, $y$, or $z$.

- When we have a *random sample* of independent and identically distributed (iid) random variables, we will index the variables in a set such as $\{X_1, X_2, \ldots, X_n\}$ for the random variables and $\{x_1, x_2, \ldots, x_n\}$ for the observed values.

- Random variable notation can become convoluded when we move to multivariate random variables. We will not establish any notational pretexts here. Instead, we should simply follow the conventions established in the text.

## What does the term "distribution" mean?

The term *distribution* is used throughout all statistical applications and discussions. Loosely speaking, the term distribution is meant to describe how a group of values are related either to each other or to the range of values on which they are defined (their *support*).

There are many mathematical notations for characterizing a statistical distribution. The choice of the characterization will depend on the context and the existence of the characterization. A random variable can be characterized by any of the following functions.

- The *cumulative distribution function* (cdf), denoted by $F(x) = \Pr(X \leq x)$. The cdf will exist for all random variables, and in general is why we use the term "distribution" so loosely throughout statistics.

## What does the term "distribution" mean? - Continued

- The *probability density function* (pdf) for continuous random variables, denoted by $f(x)$, or the *probability mass function* (pmf) for discrete random variables, denoted by $p(x)$. Note that neither of these functions are guaranteed to exist. A random variable that can be described with a cdf will not always possess a pdf or a pmf.

- Transformation functions such as the *moment generating function* $m(t) = \mathbb{E}[\exp(tX)]$ and the *characteristic function* $\phi(t) = \mathbb{E}[\exp(itX)]$.

- Specialized representations for particular applications such as the *hazard function* $h(t) = f(t)/S(t)$ and the *survival function* $S(t) = 1 - F(t)$ used in Survival Analysis.

- In data analysis distributions can be analyzed using the empirical cdf, the histogram, the Quantile-Quantile plot, and the Kolmogorov-Smirnov test.

Mathematical Expectation is the theoretical averaging of a random variable with respect to its distribution function. In this sense the pdf or pmf act as a weight function that allow you to find the "center" of the distribution.

- For a continuous random variable $X$ with pdf function $f(x)$, the mathematical expectation of $X$ can be computed by

$$\mathbb{E}[X] = \int x f(x) dx. \tag{1}$$

- For a discrete random variable $X$ with pmf function $p(x) = \Pr(X = x)$, the mathematical expectation of $X$ can be computed by

$$\mathbb{E}[X] = \sum_x x p(x). \tag{2}$$

- $\mathbb{E}[X]$ is also referred to as the *first moment* of $X$.

## Expectation, Variance, and Covariance as Mathematical Operators

Let $X$ denote a random variable. Consider the affine transformation $aX + b$.

- $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$
- $\mathsf{Var}[aX + b] = a^2\mathsf{Var}[X]$

Let $X$ and $Y$ be random variables with a joint distribution function. (In the continuous case we would denote this joint distribution function by the joint density function $f(x, y)$.) Consider the linear transformations $aX$ and $bY$.

- $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$
- $\mathsf{Var}[aX + bY] = a^2\mathsf{Var}[X] + b^2\mathsf{Var}[Y] + ab\mathsf{Cov}[X, Y]$

Here the reader should note that in general $\mathsf{Cov}[aX+b, cY+d] = ac\mathsf{Cov}[X, Y]$. If $X$ and $Y$ are independent random variables, then $\mathsf{Cov}[X, Y] = 0$. The converse of this statement is not true except when both $X$ and $Y$ are normally distributed. In general $\mathsf{Cov}[X, Y] = 0$ does not imply that $X$ and $Y$ are independent random variables.

- Use the sample mean $\bar{X} = \sum_{i=1}^{n} X_i$ to show that the sample mean of $aX_i + b$ is $a\bar{X} + b$.

- Use the sample variance

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 \tag{3}$$

  to show that the sample variance of $X_i + b$ is the same as $X_i$ and that the sample variance of $aX_i$ is $a^2$Var$[X]$. (Hint: This does not need to be difficult.)

## Suggested Reading

---

- An introductory overview of the concepts of matrix algebra related to linear regression and multivariate analysis can be found in Chapter 2 *Matrix Algebra* from *Methods of Multivariate Analysis* (Third Edition) by Rencher and Christensen.

- The concepts related to random variables can be found in any 'engineering' introduction to statistics book, or higher level mathematical statistics book.