Multivariate Analysis (Slides 2)

- In today’s class we will look at some important preliminary material that we need to cover before we look at many multivariate analysis methods.

- This material will include topics that you are likely to have seen in courses in probability and linear algebra.
Notation: observational unit

- A vector of observed values of each of $m$ variables for observational unit $i$ can be denoted as a column vector:

$$
\mathbf{x}_i = \begin{pmatrix}
  x_{i1} \\
  x_{i2} \\
  \vdots \\
  x_{im}
\end{pmatrix}
$$

i.e., $x_{ij}$ represents the value of variable $j$ for observational unit $i$.

- Observations will generally be denoted by lower case letters.
- For univariate data $m = 1$. 
Notation: multivariate data

- Multivariate data sets are generally denoted by a data matrix \( \mathbf{X} \) with \( n \) rows and \( m \) columns (\( n = \) total number of observational units, \( m = \) number of variables).

\[
\mathbf{X} = \begin{pmatrix}
    x_{11} & x_{12} & \cdots & x_{1m} \\
    x_{21} & x_{22} & \cdots & x_{2m} \\
    \vdots & \vdots & \ddots & \vdots \\
    x_{n1} & x_{n2} & \cdots & x_{nm}
\end{pmatrix}
\]

- The \( i \)th row of \( \mathbf{X} \) is therefore \( \mathbf{x}_i^T \).
Random variables

- A random variable (r.v.) is a mapping which assigns a real number to each outcome of a variable.

- Say our variable of interest is ‘gender’. The outcomes of this variable are therefore ‘male’ and ‘female’. The random variable $X$ assigns a number to these outcomes, *i.e.*,

$$
X = \begin{cases} 
1 & \text{if female} \\
0 & \text{if male.}
\end{cases}
$$

- A random variable whose value is unknown will generally be denoted by upper case letters.
Expectation of a Random Variable

- If $X$ is a discrete random variable with probability mass function $P(X)$, then the expected value of $X$ (also known as its mean) is defined as
  
  $$\mu = \mathbb{E}[X] = \sum_x xP(X = x).$$

- If $X$ is a continuous random variable with probability density function $f(x)$ defined on the space $\mathbb{R}$, then
  
  $$\mu = \mathbb{E}[X] = \int_{-\infty}^{\infty} xf(x)dx.$$

- The expectation of a function $u(X)$ is given by:
  a) $\mu = \mathbb{E}[u(X)] = \sum_x u(x)P(X = x)$ for a discrete r.v.
  b) $\mu = \mathbb{E}[u(X)] = \int_{-\infty}^{\infty} u(x)f(x)dx$ for a continuous r.v.
Variance, Covariance and Correlation

- Let’s consider random variables $X_1, X_2, \ldots, X_m$.

- The variance of $X_i$ is defined to be 
  $$\text{Var}[X_i] = \mathbb{E}[(X_i - \mathbb{E}[X_i])^2] = \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2.$$ 

- The covariance of $X_i$ and $X_j$ is defined to be 
  $$\text{Cov}[X_i, X_j] = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])].$$

- The correlation of $X_i$ and $X_j$ is defined to be 
  $$\text{Cor}[X_i, X_j] = \frac{\text{Cov}[X_i, X_j]}{\sqrt{\text{Var}[X_i]\text{Var}[X_j]}}.$$ 

- Correlation is hence a ‘normalized’ form of covariance, with equality between the two existing if the random variables have unit variance.
Covariance Matrix

- Frequently, it is convenient to record the variance and covariance of a set of random variables $\mathbf{X} = (X_1, X_2, \ldots, X_m)$ using a matrix

$$
\Sigma = \begin{pmatrix}
  s_{11} & s_{12} & \cdots & s_{1m} \\
  s_{21} & s_{22} & \cdots & s_{2m} \\
  \vdots & \vdots & \ddots & \vdots \\
  s_{m1} & s_{m2} & \cdots & s_{mm}
\end{pmatrix},
$$

where $s_{ij} = \text{Cov}[X_i, X_j]$ and $s_{ii} = \text{Var}[X_i]$.

- We call this matrix a covariance matrix.

- $\Sigma = \text{Cov}[\mathbf{X}] = \text{Var}[\mathbf{X}]$ (usually referred to by former equality).
Independence

• Two random variables $X_1$ and $X_2$ are said to be independent if and only if:

$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2)$$

• For two independent random variables $X_1$ and $X_2$:

$$\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1] \mathbb{E}[X_2]$$

• Proof: Exercise (refer to the definition).
Linear Combinations

• Suppose that $a$ and $b$ are constants and the random variable $X$ has expected value $\mu$ and variance $\sigma^2$, then
  a) $E[aX + b] = aE[X] + b = a\mu + b$
  b) $\text{Var}[aX + b] = a^2\text{Var}[X] = a^2\sigma^2$.

• Let $X_1$ and $X_2$ denote two independent random variables with respective means $\mu_1$ and $\mu_2$ and variances $\sigma_1^2$ and $\sigma_2^2$. If $a_1$ and $a_2$ are constants, then
  c) $E[a_1X_1 + a_2X_2] = a_1E[X_1] + a_2E[X_2] = a_1\mu_1 + a_2\mu_2$.
  d) $\text{Var}[a_1X_1 + a_2X_2] = a_1^2\text{Var}[X_1] + a_2^2\text{Var}[X_2] = a_1^2\sigma_1^2 + a_2^2\sigma_2^2$.

• If $X_1$ and $X_2$ are not independent then c) above still holds but d) is replaced by
  e) $\text{Var}[a_1X_1 + a_2X_2] = a_1^2\text{Var}[X_1] + a_2^2\text{Var}[X_2] + 2a_1a_2\text{Cov}[X_1, X_2]$.

• Proofs: Exercise (return to the definitions).
Linear Combinations for Covariance

Suppose that $a$, $b$, $c$ and $d$ are constants and $X$, $Y$, $W$ and $Z$ are random variables with non-zero variance, then

a) $\text{Cov}[aX + b, cY + d] = ac\text{Cov}[X, Y]$

b) $\text{Cov}[aX + bY, cW + dZ] =$


- Proofs: Exercise (as with the rest, manipulation of algebra from the definitions).
Matrix representation: Expected Value

- Similar ideas follow in the more general \( a_1X_1 + \cdots + a_mX_m \) case.
- Remember \( \mathbb{E}[a_1X_1 + a_2X_2 + \cdots + a_mX_m] = a_1\mu_1 + a_2\mu_2 + \cdots a_m\mu_m \).
- Let \( \mathbf{a} = (a_1, a_2, \ldots, a_m)^T \) be a vector of constants and \( \mathbf{X} = (X_1, X_2, \ldots, X_m)^T \) be a vector of random variables.
- Then we can write

\[
 a_1X_1 + a_2X_2 + \cdots a_mX_m = (a_1, a_2, \cdots, a_m) \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix} = \mathbf{a}^T \mathbf{X}
\]

- Hence we can write \( \mathbb{E}[\mathbf{a}^T \mathbf{X}] = \mathbf{a}^T \mu \), where \( \mu = (\mu_1, \mu_2, \ldots, \mu_m)^T \).
Matrix representation: Variance

- Similarly $\text{Var}[a_1X_1 + a_2X_2 + \cdots + a_mX_m] = \text{Var}[a^T X]$, and

\[
\text{Var}[a^T X] = a_1^2 \text{Var}[X_1] + a_2^2 \text{Var}[X_2] + \cdots + a_m^2 \text{Var}[X_m]
\]

\[
+ a_1 a_2 \text{Cov}[X_1, X_2] + \cdots + a_1 a_m \text{Cov}[X_1, X_m]
\]

\[
+ \cdots + a_{m-1} a_m \text{Cov}[X_{m-1}, X_m]
\]

\[
= \sum_{i=1}^{m} a_i^2 \text{Var}[X_i] + \sum_{i=1}^{m} \sum_{j \neq i} a_i a_j \text{Cov}[X_i, X_j]
\]

\[
= \sum_{i=1}^{m} a_i^2 s_{ii} + \sum_{i=1}^{m} \sum_{j \neq i} a_i a_j s_{ij}
\]

\[
= a^T \Sigma a
\]
Matrix representation: Covariance

- Suppose that $U = a^T X$ and $V = b^T X$.

  \[
  \text{Cov}[U, V] = \sum_{i=1}^{m} a_i b_i s_{ii} + \sum_{i=1}^{m} \sum_{j \neq i} a_i b_j s_{ij}.
  \]

- In matrix notation,

  \[
  \text{Cov}[U, V] = a^T \Sigma b = b^T \Sigma a.
  \]

- Proof: Exercise
Example

• Let $E[X_1] = 2$, $\text{Var}[X_1] = 4$, $E[X_2] = 0$, $\text{Var}[X_2] = 1$ and $\text{Cor}[X_1, X_2] = 1/3$.

• Exercise:
  – What is the expected value and variance of $X_1 + X_2$?
  – What is the expected value and variance of $X_1 - X_2$?
Eigenvalues and Eigenvectors

• Suppose that we have a $m \times m$ matrix $A$.

• **Definition:** $\lambda$ is an *eigenvalue* of $A$ if there exists a non-zero vector $v$ such that

\[ A v = \lambda v. \]

• The vector $v$ is said to be an *eigenvector* of $A$ corresponding to the eigenvalue $\lambda$.

• We can find eigenvalues by solving the equation,

\[ \det(A - \lambda I) = 0. \]
Finding Eigenvalues

• If \( \mathbf{v} \) is an eigenvector of \( \mathbf{A} \) with eigenvalue \( \lambda \), then \( \mathbf{A}\mathbf{v} - \lambda \mathbf{I}\mathbf{v} = \mathbf{0} \).

• Hence \((\mathbf{A} - \lambda \mathbf{I})\mathbf{v} = \mathbf{0}\).

• If there exists an inverse \((\mathbf{A} - \lambda \mathbf{I})^{-1}\) then the trivial solution \( \mathbf{v} = \mathbf{0} \) is obtained.

• When there does not exist a trivial solution there is no inverse and hence \( \det(\mathbf{A} - \lambda \mathbf{I}) = 0 \).
• The eigenvectors are only scaled by the matrix $A$. 

\[ \begin{align*}
\text{x} \\
\text{Ax}
\end{align*} \]
Other vectors are rotated and scaled by the matrix $A$. 

\begin{align*}
Ax
\end{align*}
Unit Eigenvectors

- **Property:** If $v$ is an eigenvector corresponding to eigenvalue $\lambda$, then $u = \alpha v$ will also be an eigenvector corresponding to $\lambda$.

- **Definition:** The eigenvector $v$ is a *unit eigenvector* if $\sum_{i=1}^{m} v_i^2 = v^T v = 1$.

- We can turn any eigenvector $v$ into a unit eigenvector by multiplying it by the value $\alpha = \frac{1}{\sqrt{v^T v}}$. 
Orthogonal and Orthonormal Vectors

- **Definition:** Two vectors \( \mathbf{u} \) and \( \mathbf{v} \) are *orthogonal* if \( \mathbf{u}^T \mathbf{v} = 0 = \mathbf{v}^T \mathbf{u} \).

- **Definition:** Two vectors \( \mathbf{u} \) and \( \mathbf{v} \) are *orthonormal* if they are orthogonal and \( \mathbf{u}^T \mathbf{u} = 1 \) and \( \mathbf{v}^T \mathbf{v} = 1 \).
Eigenvalues of Covariance Matrices

- **Fact:** The eigenvalues of a covariance matrix $\Sigma$ are non-negative.
- If $\lambda$ is an eigenvalue of $\Sigma$, then
  \[ \Sigma v = \lambda v, \]
  where $v$ is an eigenvector corresponding to $\lambda$.
- Hence,
  \[ v^T \Sigma v = v^T \lambda v = \lambda v^T v \]
  \[ \Rightarrow \lambda = \frac{v^T \Sigma v}{v^T v}. \]
- The numerator and denominator are both non-negative (why?), so $\lambda$ must be non-negative.
Eigenvectors of Covariance Matrices

- **Fact:** An $m \times m$ covariance matrix $\Sigma$ has $m$ orthonormal eigenvectors.
- Proof Omitted, but uses Spectral Decomposition (or similar theorem) of a symmetric matrix.
- This result will be used when developing principal components analysis.
Example

• Suppose that we have the matrix

\[ A = \begin{pmatrix} 1 & -1 \\ 3 & 5 \end{pmatrix}. \]

• **Question:** What are the eigenvalues and corresponding eigenvectors of \( A \)?
Example

- **Answer:** The eigenvalues are 4 and 2 and the corresponding eigenvectors are

\[
\begin{pmatrix}
\frac{1}{\sqrt{10}} \\
-\frac{3}{\sqrt{10}}
\end{pmatrix} = \begin{pmatrix}
0.32 \\
-0.95
\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}
\frac{1}{\sqrt{2}} \\
-\frac{1}{\sqrt{2}}
\end{pmatrix} = \begin{pmatrix}
0.71 \\
-0.71
\end{pmatrix}.
\]
Example: Frog Cranial Measurements

- **Example:** The cranial length and cranial breath of 35 female frogs are believed to have expected value \((23, 24)^T\) and covariance matrix

\[
\begin{pmatrix}
17.7 & 20.3 \\
20.3 & 24.4
\end{pmatrix}
\]

- **Question:** What are the eigenvalues and eigenvectors of the covariance matrix?