Multivariate Analysis Lecture 3: Random Vectors and A Random Sample

Zhaoxia Yu
Professor, Department of Statistics

2023-04-12

Section 1

Review: Random Variables (Univariate) and A Random Sample

Subsection 1

Random Variables

What Is a Random Variable?

- A random variable is a numerical quantity that takes on different values with certain probabilities.
- e.g., a normal distributed random variable takes values between $-\infty$ to ∞.
- It represents the outcome of a random event or experiment.
- e.g., the BMI of a randomly chosen adult living in Canada
- Random variables can be discrete or continuous.

The Mean of a Random Variable

- The mean of a random variable X measures its central tendency, often denoted by μ or $E(X)$.
- It is the expected value of the random variable, weighted by the probabilities of each possible outcome:
  - Continuous: $\mu = E(X) \overset{def}{=} \int_{-\infty}^{\infty} xf(x)dx$
  - Discrete: $\mu = E(X) \overset{def}{=} \sum_{i=1} x_i p_i$
- $E(aX + b) = aE(X) + b$, where X is random and a and b are fixed.

Variance of a Random Variable

- The variance of a random variable is a measure of how spread out its values are around the mean.
- It represents the expected value of the squared deviation of the random variable from its mean. $\sigma^2 \stackrel{def}{=} E[(X-\mu)^2]$, specifically,
  - Continuous: $\sigma^2 \stackrel{def}{=} \int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx$
  - Discrete: $\sigma^2 \stackrel{def}{=} \sum_{i=1}(x_i-\mu)^2 p_i$
- σ, the square root of the variance, is called the standard deviation (SD) of X.

Properties of Variance

- The variance is a non-negative quantity.
- The variance of a constant is 0: Var(c) $= 0$, where c is a constant.
- The variance is affected by changes in the scale of the random variable but not by a shift in locations:
  $Var(aX + b) = a^2 Var(X)$, where a is a constant.
- The variance of a sum of independent random variables is the sum of their individual variances:
  $Var(X + Y) = Var(X) + Var(Y)$, provided that X and Y are independent.More general, if $X_1, \cdots, X_n$ are mutually independent, then $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$.

Subsection 2

A Random Sample of Random Variables

Random Samples (from Simple Random Sampling)

- In a simple random sample, each member of the population is selected independently and with equal probability.
- Obtaining a truly random sample can often be challenging. Reasons:
  - it may be difficult or impossible to obtain a complete list of all members of the population of interest.
  - it may be costly or time-consuming to sample from the entire population.
  - there may be practical constraints on the sampling process, such as geographic distance, language barriers, or legal restrictions.
  - certain subgroups of the population may be underrepresented or difficult to reach, leading to potential biases in the sample.
- Nevertheless, we assume the samples are simple random samples for theoretical derivations

Sample Mean and Variance from a Simple Random Samples

- Let $(X_1, \cdots, X_n)$ be a simple random sample from a distribution with mean μ and variance σ^2. The notation we will use is

$$X_1, \cdots, X_n \overset{iid}{\sim} (\mu, \sigma^2)$$

- Summary Statistics and their Expectations:
    - The sample mean $\bar{X}$ is defined as $\bar{X} \overset{def}{=} \frac{1}{n}\sum_{i=1}^{n} X_i$.
    - $\bar{X}$ is unbiased for μ, i.e., $E(\bar{X}) = \mu$. $Var(\bar{X}) = \sigma^2/n$.
    - The sample variance $S^2 \overset{def}{=} \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.
    - S^2 is unbiased for σ^2, i.e., $E(S^2) = \sigma^2$.

Sample Mean is Unbiased

- The proof of unbiasedness follows from the linearity of the expected value operator:

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$$

- The unbiasedness of the sample mean is a fundamental property of statistical estimation.

The Variance of the Sample Mean

$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right)$$
$$= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n}$$

- The variability of the sample means decreases as the sample size increases.
- The result is important for the design of experiments and surveys. E.g., what is a minimum sample size to achieve a desired level of precision?

Sample Variance is Unbiased

- The proof of unbiasedness follows from the properties of the variance operator and the linearity of the expected value operator:

$$
\begin{aligned}
E(S^2) &= \frac{1}{n-1}\sum_{i=1}^{n} E[(X_i - \bar{X})^2] = \frac{1}{n-1}\sum_{i=1}^{n} E[(X_i - \mu + \mu - \bar{X})^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n} E[(X_i - \mu)^2 + 2(X_i - \mu)(\mu - \bar{X}) + (\mu - \bar{X})^2] \\
&= \frac{1}{n-1}[n\sigma^2 - 2nE[(\mu - \bar{X})^2] + nE[(\mu - \bar{X})^2]] \\
&= \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2
\end{aligned}
$$

Section 2

Random Vectors (Multivariate) and A Random Sample

Subsection 1

Random Vectors

Notations for Random Vectors

- A random vector is a vector whose elements are random variables. e.g.,

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$$

where each X_i is a random variable

The Expectation of A Random Vector

- Let $E(\mathbf{X})$ denote the mean vector of $\mathbf{X}_{p\times 1}$. We have

$$\boldsymbol{\mu} = E(\mathbf{X}) \stackrel{def}{=} \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix},$$

where $\mu_i = E(X_i), i = 1, \cdots, p$.

- Suppose $X_1, \cdots, X_n \stackrel{iid}{\sim} (\mu, \sigma^2)$, and $\mathbf{X} = (X_1, \cdots, X_n)^T$. What is $E(\mathbf{X})$?

$$E(\mathbf{X}) = E[\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}] = \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix} = \mu\mathbf{1}$$

The Variance-Covariance of A Random Vector

- The variance-covariance matrix of a random vector X is a square matrix that summarizes the variability and dependence among its components.
- It is denoted by the symbol $Var(\mathbf{X})$, $Cov(\mathbf{X})$, or Σ and is given by:

$$
\begin{aligned}
\Sigma &\stackrel{def}{=} E[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T] \\
&= \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) & \cdots & Cov(X_1, X_p) \\ Cov(X_2, X_1) & Var(X_2) & \cdots & Cov(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_p, X_1) & Cov(X_n, X_2) & \cdots & Var(X_p) \end{bmatrix}
\end{aligned}
$$

The Variance-Covariance of A Random Vector

- Alternative notations

$$Var(\mathbf{X}) = \Sigma = (\sigma_{ij}) \stackrel{def}{=} \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{bmatrix}$$

- Remarks
  - The covariance between two components measures how much they vary together, and it can be positive, negative, or zero.
  - Σ is a symmetric matrix because $\sigma_{ij} = Cov(X_i, X_j) = \sigma_{ji}$.
  - The diagonal elements of Σ represent the variances of the components of the random vector: $\sigma_i^2 = Var(X_i) = Cov(X_i, X_i)$.

Correlation Matrix

- A correlation matrix is a table showing correlation coefficients between different variables.
- The correlation coefficient measures the strength and direction of the linear relationship between two variables.

$$\mathrm{Corr}(X_i, X_j) = \frac{\mathrm{Cov}(X_i, X_j)}{\sqrt{Var(X_i)}\sqrt{Var(X_j)}} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

- The correlation coefficient ranges from -1 to 1, with values close to -1 indicating a strong negative linear relationship, values close to 1 indicating a strong positive linear relationship, and values close to 0 indicating no linear relationship.

Correlation Matrix

$$\mathbf{R} = \begin{bmatrix} 1 & \mathrm{Corr}(X_1, X_2) & \cdots & \mathrm{Corr}(X_1, X_p) \\ \mathrm{Corr}(X_2, X_1) & 1 & \cdots & \mathrm{Corr}(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Corr}(X_p, X_1) & \mathrm{Corr}(X_p, X_2) & \cdots & 1 \end{bmatrix}$$

- $\rho_{ij} \overset{def}{=} Corr(X_i, X_j)$ - The diagonal ρ_{ii} of the correlation matrix shows the correlation of each variable with itself, which is always equal to 1. - The matrix is symmetric since the correlation between X and Y is the same as the correlation between Y and X: $\rho_{ij} = \rho_{ji}$.
- Correlation matrix can help identify variables that are correlated.

Covariance Matrix of Two Random Vectors

- The covariance matrix of two random vectors $\mathbf{X} = (X_1, \ldots, X_p)^T$ and $\mathbf{Y} = (Y_1, \ldots, Y_q)^T$ is a $p \times q$ matrix defined as

$$\begin{aligned}\mathbf{Cov}(\mathbf{X}, \mathbf{Y}) &\overset{def}{=} E[(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{Y} - \boldsymbol{\mu}_Y)^T] \\ &= \begin{bmatrix} \mathrm{Cov}(X_1, Y_1) & \cdots & \mathrm{Cov}(X_1, Y_q) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_p, Y_1) & \cdots & \mathrm{Cov}(X_p, Y_q) \end{bmatrix}\end{aligned}$$

- Each element of the matrix is the covariance between two corresponding elements of the vectors.

Covariance Matrix of Two Random Vectors

- E.g.,

$$\mathbf{X}_{2\times 1} = \begin{pmatrix} X_1 & X_2 \end{pmatrix}^T, \mathbf{Y}_{3\times 1} = \begin{pmatrix} Y_1 & Y_2 & Y_3 \end{pmatrix}$$

$$\begin{aligned}
& Cov(\mathbf{X}, \mathbf{Y}) \\
=& E[\begin{pmatrix} X_1 - \mu_{x1} \\ X_2 - \mu_{x2} \end{pmatrix} (Y_1 - \mu_{y1}, Y_2 - \mu_{y2}, Y_3 - \mu_{y3})^T] \\
=& \begin{bmatrix} E[(X_1 - \mu_{x1})(Y_1 - \mu_{y1})] & E[(X_1 - \mu_{x1})(Y_2 - \mu_{y2})] & E[(X_1 - \mu \\ E[(X_2 - \mu_{x2})(Y_1 - \mu_{y1})] & E[(X_2 - \mu_{x2})(Y_2 - \mu_{y2})] & E[(X_3 - \mu \end{bmatrix} \\
=& \begin{bmatrix} Cov(X_1, Y_1) & Cov(X_1, Y_2) & Cov(X_1, Y_3) \\ Cov(X_2, Y_1) & Cov(X_2, Y_2) & Cov(X_2, Y_3) \end{bmatrix}
\end{aligned}$$

- Note: $\mathbf{Cov}(\mathbf{X}, \mathbf{Y}) = [\mathbf{Cov}(\mathbf{Y}, \mathbf{X})]^T$

Subsection 2

A Random Sample of Random Vectors

Notations about A Random Sample of Random Vectors

- Consider a random sample from a multivariate distribution with mean vector $\boldsymbol{\mu}_{p\times 1}$ and covariance $\boldsymbol{\Sigma}_{p\times p}$
- A random sample of random vectors is a collection of n independent and identically distributed random vectors, denoted as $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$.
- The random sample of random vectors is denoted by

$$\mathbf{X}_{n\times p} = \begin{pmatrix} \mathbf{X}_1^T \\ \vdots \\ \mathbf{X}_n^T \end{pmatrix}$$

- Each random vector $\mathbf{X}_i$ is of dimension p and can be represented as:

$$\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{ip})^T$$

Sample Mean Vector $\bar{\mathbf{X}}_{p\times 1}$

- The sample mean vector, denoted as $\bar{\mathbf{X}}$, is a random vector of dimension p, defined as:

$$\bar{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i$$

- It is unbiased for the population mean vector $\boldsymbol{\mu}$ because

$$E[\bar{\mathbf{X}}] = E[\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i] = \frac{1}{n}\sum_{i=1}^{n}E[\mathbf{X}_i] = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\mu} = \boldsymbol{\mu}$$

- The sample mean vector $\bar{\mathbf{X}}$ is often used to estimate the population mean vector $\boldsymbol{\mu}$.

The Covariance of the Sample Mean Vector

- The sample mean vector, denoted as $\bar{\mathbf{X}}$, is a random vector of dimension p. We can also compute its covariance matrix
- Because $(\mathbf{X}_1, \cdots, \mathbf{X}_n)$ are iid ,

$$Cov(\bar{\mathbf{X}}) = Cov(\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i) = \frac{1}{n^2}\sum_{i=1}^{n} Cov(\mathbf{X}_i) = \frac{1}{n}\boldsymbol{\Sigma}$$

- Similar to the population mean vector, the population covariance $\boldsymbol{\Sigma}$ is typically unknown. If we have a random sample, we can estimate it - the sample covariance matrix.S

Sample Covariance Matrix $\mathbf{S}_{p\times p}$

- The sample covariance matrix, denoted as $\mathbf{S}$, is a $p \times p$ symmetric matrix, defined as:

$$\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T$$

- Next, we show that the sample covariance matrix $\mathbf{S}$ is an unbiased estimator of $\boldsymbol{\Sigma}$:

$$\mathbb{E}[\mathbf{S}] = \boldsymbol{\Sigma}$$

The Sample Covariance Matrix is Unbiased: Lemmas

- Lemma 1: $E(\mathbf{X}_i\mathbf{X}_i^T) = \boldsymbol{\mu}\boldsymbol{\mu}^T + Cov(\mathbf{X}_i) = \boldsymbol{\mu}\boldsymbol{\mu}^T + \boldsymbol{\Sigma}$. -Proof. By the definition of Cov, we have

$$\begin{aligned}
\boldsymbol{\Sigma} &= E[(\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})^T] \\
&= E[\mathbf{X}_i\mathbf{X}_i^T - \boldsymbol{\mu}\mathbf{X}_i^T - \mathbf{X}_i\boldsymbol{\mu}^T + \boldsymbol{\mu}\boldsymbol{\mu}^T] \\
&= E[\mathbf{X}_i\mathbf{X}_i^T] - \boldsymbol{\mu}E[\mathbf{X}_i^T] - E[\mathbf{X}_i]\boldsymbol{\mu}^T + \boldsymbol{\mu}\boldsymbol{\mu}^T \\
&= E[\mathbf{X}_i\mathbf{X}_i^T] - \boldsymbol{\mu}\boldsymbol{\mu}^T - \boldsymbol{\mu}\boldsymbol{\mu}^T + \boldsymbol{\mu}\boldsymbol{\mu}^T \\
&= E[\mathbf{X}_i\mathbf{X}_i^T] - \boldsymbol{\mu}\boldsymbol{\mu}^T
\end{aligned}$$

  As a result, $E[\mathbf{X}_i\mathbf{X}_i^T] = \boldsymbol{\mu}\boldsymbol{\mu}^T + \boldsymbol{\Sigma}$.
- Similarly, we have Lemma 2:

$$E(\bar{\mathbf{X}}\bar{\mathbf{X}}^T) = \boldsymbol{\mu}\boldsymbol{\mu}^T + Cov(\bar{\mathbf{X}}) = \boldsymbol{\mu}\boldsymbol{\mu}^T + \frac{1}{n}\boldsymbol{\Sigma}$$

The Sample Covariance Matrix is Unbiased: Proof

- Proof:Expand the product:

$$
\begin{aligned}
\mathbf{S} &= \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i\mathbf{X}_i^T - \mathbf{X}_i\bar{\mathbf{X}}^T - \bar{\mathbf{X}}\mathbf{X}_i^T + \bar{\mathbf{X}}\bar{\mathbf{X}}^T) \\
&= \frac{1}{n-1}[\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^T - n\bar{\mathbf{X}}\bar{\mathbf{X}}^T - n\bar{\mathbf{X}}\bar{\mathbf{X}}^T + n\bar{\mathbf{X}}\bar{\mathbf{X}}^T] \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^T - \frac{n}{n-1}\bar{\mathbf{X}}\bar{\mathbf{X}}^T
\end{aligned}
$$

The Sample Covariance Matrix is Unbiased: Proof (continued)

- Taking the expected value:

$$
\begin{aligned}
\mathbb{E}[\mathbf{S}] &= \mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^T - \frac{n}{n-1}\bar{\mathbf{X}}\bar{\mathbf{X}}^T\right] \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}[\mathbf{X}_i\mathbf{X}_i^T] - \frac{n}{n-1}\mathbb{E}[\bar{\mathbf{X}}\bar{\mathbf{X}}^T] \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^T) - \frac{n}{n-1}(\frac{\mathbf{1}}{\boldsymbol{n}}\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^T) \\
&= \frac{n}{n-1}\boldsymbol{\Sigma} + \frac{n}{n-1}\boldsymbol{\mu}\boldsymbol{\mu}^T - \frac{1}{n-1}\boldsymbol{\Sigma} - \frac{n}{n-1}\boldsymbol{\mu}\boldsymbol{\mu}^T \\
&= \boldsymbol{\Sigma}
\end{aligned}
$$

- Therefore, the sample covariance matrix is unbiased

Examples: The Iris Setosa Data

- The iris data consists of three random samples, one for each species. Consider the setosa sample.
- It is a random sample (let's assume it) of size 50.
- The data matrix has $n = 50$ rows and $p = 4$ columns

The Data Matrix of Iris Setosa

```r
setosa=as.matrix(iris[iris$Species=="setosa", 1:4])
dim(setosa)
```

```
## [1] 50  4
```

```r
head(setosa)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width
## 1          5.1         3.5          1.4         0.2
## 2          4.9         3.0          1.4         0.2
## 3          4.7         3.2          1.3         0.2
## 4          4.6         3.1          1.5         0.2
## 5          5.0         3.6          1.4         0.2
## 6          5.4         3.9          1.7         0.4
```

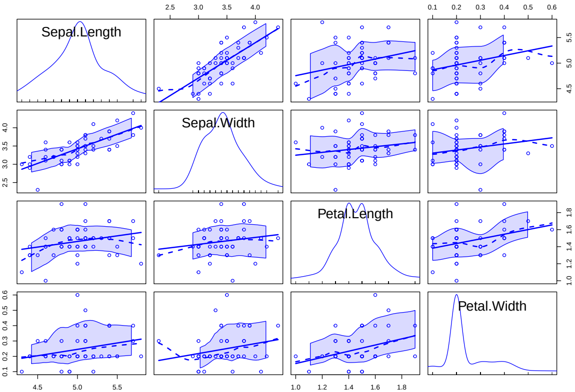
The Sample Mean of Iris Setosa

```
sample.meanvec=matrix(colMeans(setosa), 4, 1)
rownames(sample.meanvec)=colnames(setosa)
colnames(sample.meanvec)="mean"
sample.meanvec
```

```
##               mean
## Sepal.Length 5.006
## Sepal.Width  3.428
## Petal.Length 1.462
## Petal.Width  0.246
```

Pairwise Scatter Plot of the Features of Iris Setosa

```
scatterplotMatrix(setosa)
```



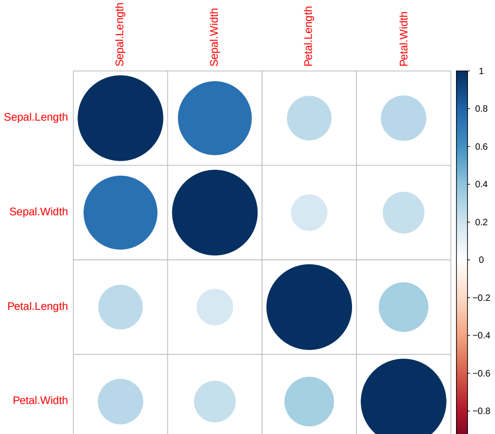
The Sample Covariane Matrix of Iris Setosa

```
sample.cov=cov(setosa)
round(sample.cov,2)
```

```
##              Sepal.Length Sepal.Width Petal.Length Petal
## Sepal.Length         0.12        0.10         0.02
## Sepal.Width          0.10        0.14         0.01
## Petal.Length         0.02        0.01         0.03
## Petal.Width          0.01        0.01         0.01
```

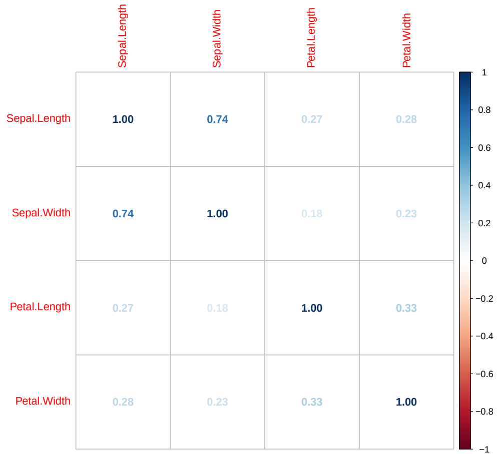
The Sample Correlation Matrix of Iris Setosa

```
sample.corr=cor(setosa)
corrplot(sample.corr)
```



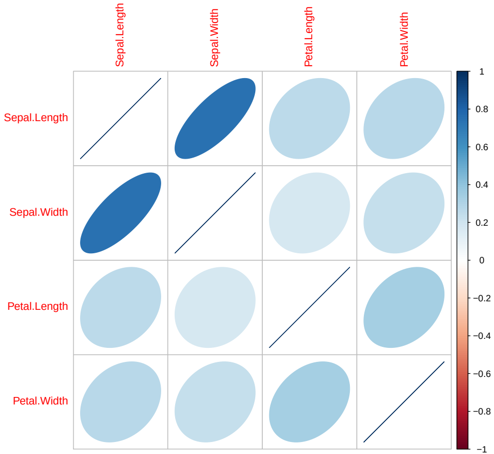
The Sample Correlation Matrix of Iris Setosa

```r
corrplot(sample.corr, method="number")
```



The Sample Correlation Matrix of Iris Setosa

```
corrplot(sample.corr, method="ellipse")
```



Sample Covariate Matrix as a Quadratic Form

$$(n-1)S = \sum (X_i - \bar{X})(X_i - \bar{X})^T = \begin{pmatrix} X_1 - \bar{X} & \cdots X_n - \bar{X} \end{pmatrix} \begin{pmatrix} (X_1 - \bar{X})^T \\ \vdots \\ (X_n - \bar{X})^T \end{pmatrix}$$

Note that

$$\begin{pmatrix} (X_1 - \bar{X})^T \\ \vdots \\ (X_n - \bar{X})^T \end{pmatrix} = \begin{pmatrix} X_1 - \bar{X} \\ \vdots \\ X_n - \bar{X} \end{pmatrix}^T = \mathbf{CX}$$

where $\mathbf{C}$ is the centering matrix defined in assignment 1, i.e., $\mathbf{C} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T$. In addition, it can be verified that $\mathbf{C}^T\mathbf{C} = \mathbf{C}$.

Therefore,

$$(n-1)S = (C\mathbf{X})^T C\mathbf{X} = \mathbf{X}^T C\mathbf{X}$$

Section 3

Linear Combination of a Random Vector: $Y = a^T X$

Definition of a Linear Combination of a Random Vector

- Let $\mathbf{X}$ be a p-dimensional random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$.
- Consider a linear combination of the form:

$$Y = \mathbf{a}^T \mathbf{X}$$

  where $\mathbf{a}$ is a p-dimensional constant vector.
- E.g., $\mathbf{X} = (X_1, X_2, X_3)^T$, $a = (1/3, 1/3, 1/3)^T$. Then

$$Y = a^T X = \frac{1}{3}(X_1 + X_2 + X_3)$$

Mean of $Y = a^T X$

- The mean of Y can be expressed as:

$$\begin{aligned} E(Y) &= E(\mathbf{a}^T \mathbf{X}) \\ &= \mathbf{a}^T E(\mathbf{X}) \\ &= \mathbf{a}^T \boldsymbol{\mu} \end{aligned}$$

- Intuitively, the mean of Y is a weighted average of the components of $\mathbf{X}$, with weights given by the corresponding components of $\mathbf{a}$.

Variance of Y

- The variance of Y can be expressed as:

$$\begin{aligned} \mathrm{Var}(Y) &= \mathrm{Var}(\mathbf{a}^T\mathbf{X}) \\ &= \mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a} \end{aligned}$$

- The variance of Y depends on the covariance structure of $\mathbf{X}$, as well as the weights given by $\mathbf{a}$. If the components of $\mathbf{a}$ are uncorrelated or orthogonal, then the variance of Y is simply a weighted sum of the variances of the components of $\mathbf{X}$. However, if the components of $\mathbf{a}$ are correlated, then the covariance structure of $\mathbf{X}$ affects the variance of Y as well.

Linear Combinations of Iris Setosa Features

- Recall that for the iris setosa, $\mathbf{X}$ is 50×4.
- Consider a linear combination of the features $Y = Xb$, where

$$b = \begin{pmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{pmatrix}$$

- Yb is a 50×1 vector, with the ith row be the average of the four features of the ith iris setosa flower. To see this

Linear Combinations of Iris Setosa Features

$$Y = Xb = \begin{pmatrix} X_1^T \\ \vdots \\ X_n^T \end{pmatrix} b = \begin{pmatrix} X_1^T b \\ \vdots \\ X_n^T b \end{pmatrix} = \begin{pmatrix} \frac{x_{11}+x_{12}+x_{13}+x_{14}}{4} \\ \vdots \\ \frac{x_{n1}+x_{n2}+x_{n3}+x_{n4}}{4} \end{pmatrix}$$

Linear Combinations of Iris Setosa Features: sample mean

```
b=matrix(1/4, 4, 1)
Y=setosa%*%b
#sample mean of Y: the following two results are the same
mean(Y)
```

```
## [1] 2.5355
```

```
t(b)%*%sample.meanvec
```

```
##          mean
```

Linear Combinations of Iris Setosa Features: sample variance

```
#sample variance of Y: the following two results are the s
var(Y)
```

```
##             [,1]
## [1,] 0.03844617
```

```
t(b)%*%cov(setosa)%*%b
```

```
##             [,1]
## [1,] 0.03844617
```