Multivariate Normal Distribution – I

- We will almost always assume that the joint distribution of the $p \times 1$ vectors of measurements on each sample unit is the p-dimensional *multivariate normal distribution*.

- The MVN assumption is often appropriate:
  - Variables can sometimes be assumed to be multivariate normal (perhaps after transformation)
  - Central limit theorem tells us that distribution of many multivariate sample statistics is approximately normal, regardless of the form of the population distribution.

- As a bonus, the MVN assumption leads to tractable results (but mathematical convenience should not be the reason for choosing the MVN as the probability model).

Multivariate Normal Distribution

- The MVN is a generalization of the univariate normal distribution for the case $p \geq 2$.

- Recall that if X is normal with mean μ and variance σ^2, and its density function is given by

$$f(x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp[-\frac{1}{2\sigma^2}(x-\mu)^2], \quad -\infty < x < \infty.$$

- We can write the *kernel* of the density as:

$$\exp[-\frac{1}{2\sigma^2}(x-\mu)^2] = \exp[-\frac{1}{2}(x-\mu)'(\sigma^2)^{-1}(x-\mu)].$$

Multivariate Normal Distribution

Now consider a $p \times 1$ random vector $X = [X_1, X_2, \ldots, X_p]'$. The kernel shown above generalizes to

$$\exp[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})],$$

where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} \quad \text{and} \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

is the $p \times p$ population covariance matrix. This is the *kernel* of the MVN distribution.

Multivariate Normal Distribution

- The quadratic form $(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})$ in the kernel is a statistical distance measure, of the type we described earlier. For any value of $\mathbf{x}$, the quadratic form gives the squared statistical distance of $\mathbf{x}$ from $\boldsymbol{\mu}$ accounting for the fact that the variances of the p variables may be different and that the variables may be correlated.

- This quadratic form is often referred to as *Mahalanobis distance*.

- The density function of the MVN distribution is:

$$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})],$$

  where the normalizing constant $(2\pi)^{p/2}|\Sigma|^{1/2}$ makes the volume under the MVN density equal to 1.

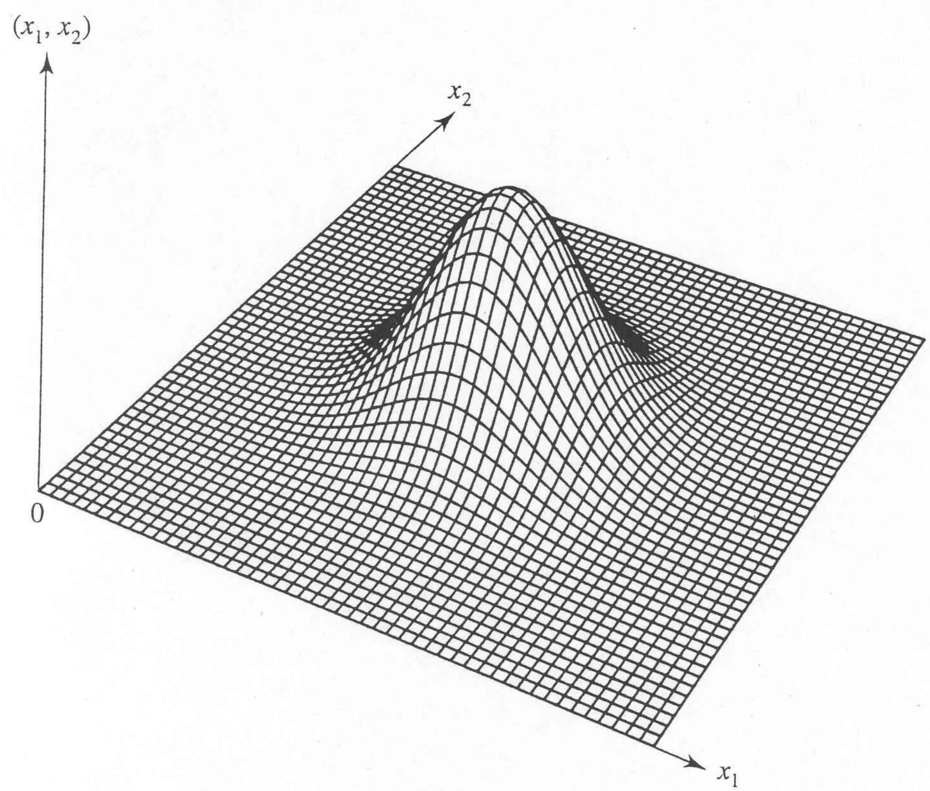
Multivariate Normal Distribution

- We use the notation

$$X \sim \mathsf{N}_p(\boldsymbol{\mu}, \Sigma) \text{ or } X \sim \mathsf{MVN}(\boldsymbol{\mu}, \Sigma)$$

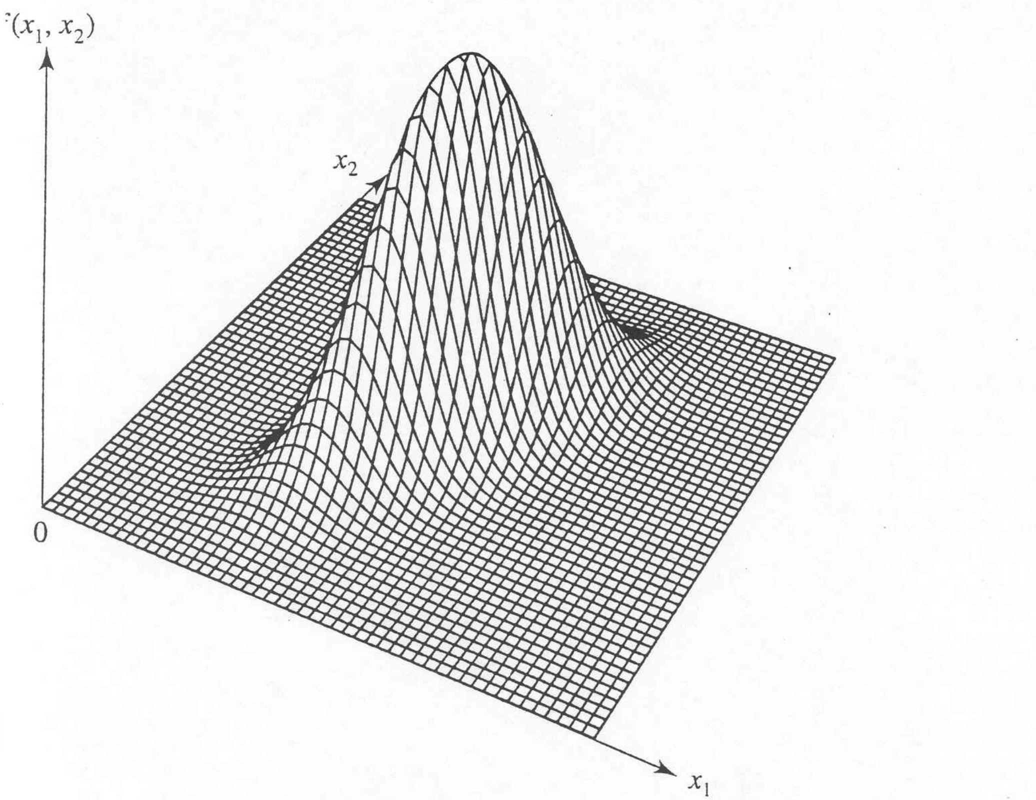
  to indicate that the p-dimensional random vector X has a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ.

Example: Bivariate Normal Density $\sigma_{11} = \sigma_{22}$, and $\rho_{12} = 0$



(a)

Example: Bivariate Normal Density $\sigma_{11} = \sigma_{22}$, and $\rho_{12} > 0$



(b)

Density Contours

- The x values that yield a constant height for the density form ellipsoids centered at $\boldsymbol{\mu}$.

- The MVN density is constant on surfaces or *contours* where

$$(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) = c^2.$$

- Definition of a constant probability density contour is all x's that satisfy the expression above.

- The axes of the ellipses are in the directions of the eigenvectors of Σ and the length of the $j-th$ longest axis is proportional to $(\sqrt{\lambda_j})$, where λ_j is the eigenvalue associated with the $j-th$ eigenvector of Σ.

Density Contours

- Recall that if (λ_j, e_j) is an eigenvalue-eigenvector pair for Σ and Σ is positive definite, then (λ_j^{-1}, e_j) is an eigenvalue-eigenvector pair of Σ^{-1}.

- The jth axis is $\pm c\sqrt{\lambda_j} e_j$, for $j = 1, ..., p$.

- We show later that if

$$c^2 = \chi_p^2(\alpha),$$

where $\chi_p^2(\alpha)$ is the upper (100α)th percentile of a χ^2 distribution with p degrees of freedom, then the probability is $1-\alpha$ that the value of a random vector will be inside the ellipsoid defined by

$$(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) \leq \chi_p^2(\alpha)$$

Bivariate Normal Density Contours

- Eigenvalues and eigenvectors of Σ are obtained from $|\Sigma - \lambda I| = 0$. Using $\sigma_{12} = \sigma_{21} = \rho\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}$, we have

$$
\begin{aligned}
0 &= \left| \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| \\
&= \left| \begin{array}{cc} \sigma_{11} - \lambda & \rho\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} \\ \rho\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} & \sigma_{22} - \lambda \end{array} \right| \\
&= (\sigma_{11} - \lambda)(\sigma_{22} - \lambda) - \rho^2 \sigma_{11}\sigma_{22} \\
&= \lambda^2 - (\sigma_{11} + \sigma_{22})\lambda + \sigma_{11}\sigma_{22}(1 - \rho^2).
\end{aligned}
$$

Bivariate Normal Density Contours

- Solutions to this quadratic equation are:

$$\lambda_1 = \frac{1}{2}\left[\sigma_{11} + \sigma_{22} + \sqrt{(\sigma_{11} + \sigma_{22})^2 - 4\sigma_{11}\sigma_{22}(1-\rho^2)}\right]$$

$$\lambda_2 = \frac{1}{2}\left[\sigma_{11} + \sigma_{22} - \sqrt{(\sigma_{11} + \sigma_{22})^2 - 4\sigma_{11}\sigma_{22}(1-\rho^2)}\right]$$

Solving quadratic equations:

$$ax^2 + bx + c = 0$$

The solutions are

$$x = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad \text{and} \quad x = \frac{-b - \sqrt{b^2 - 4ac}}{2a}$$

Bivariate Normal Density Contours

- An eigenvector associated with λ_1 satisfies

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{11} \end{bmatrix} \begin{bmatrix} e_{11} \\ e_{12} \end{bmatrix} = \lambda_1 \begin{bmatrix} e_{11} \\ e_{12} \end{bmatrix},$$

where e_{ij} denotes the jth element in the ith eigenvector. We must solve

$$\begin{aligned} (\sigma_{11} - \lambda_1)e_{11} + \rho\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}e_{12} &= 0 \\ (\sigma_{22} - \lambda_1)e_{12} + \rho\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}e_{11} &= 0 \end{aligned}$$

Bivariate Normal Density Contours

Solutions:

$$\mathbf{e}_1 = \begin{bmatrix} e_{11} \\ e_{12} \end{bmatrix} = \begin{bmatrix} d \\ db \end{bmatrix}$$

where d is any scalar and

$$b = \frac{\sigma_{22} - \sigma_{11} + \sqrt{(\sigma_{11} + \sigma_{22})^2 - 4\sigma_{11}\sigma_{22}(1 - \rho^2)}}{2\rho\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}}$$

Bivariate Normal Density Contours

Choose d to satisfy $1 = \mathbf{e}_1'\mathbf{e}_1 = d^2(1+b^2)$. Then

$$d = \frac{1}{\sqrt{1+b^2}} \quad \text{or} \quad d = \frac{-1}{\sqrt{1+b^2}}$$

and

$$\mathbf{e}_1 = \begin{bmatrix} \frac{1}{\sqrt{1+b^2}} \\ \frac{b}{\sqrt{1+b^2}} \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \frac{-1}{\sqrt{1+b^2}} \\ \frac{-b}{\sqrt{1+b^2}} \end{bmatrix}$$

Bivariate Normal Density Contours

The eigenvector corresponding to λ_2 is

$$\mathbf{e}_2 = \begin{bmatrix} \frac{1}{\sqrt{1+c^2}} \\ \frac{c}{\sqrt{1+c^2}} \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \frac{-1}{\sqrt{1+c^2}} \\ \frac{-c}{\sqrt{1+c^2}} \end{bmatrix}$$

where

$$c = \frac{\sigma_{22} - \sigma_{11} - \sqrt{(\sigma_{11} + \sigma_{22})^2 - 4\sigma_{11}\sigma_{22}(1 - \rho^2)}}{2\rho\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}}$$

These eigenvectors have the following properties:

- $||\mathbf{e}_i|| = \sqrt{\mathbf{e}_i'\mathbf{e}_i} = 1$
- $\mathbf{e}_i'\mathbf{e}_j = 0$ for $i \neq j$

Bivariate Normal Contours ($\sigma_{11} = \sigma_{22}$)

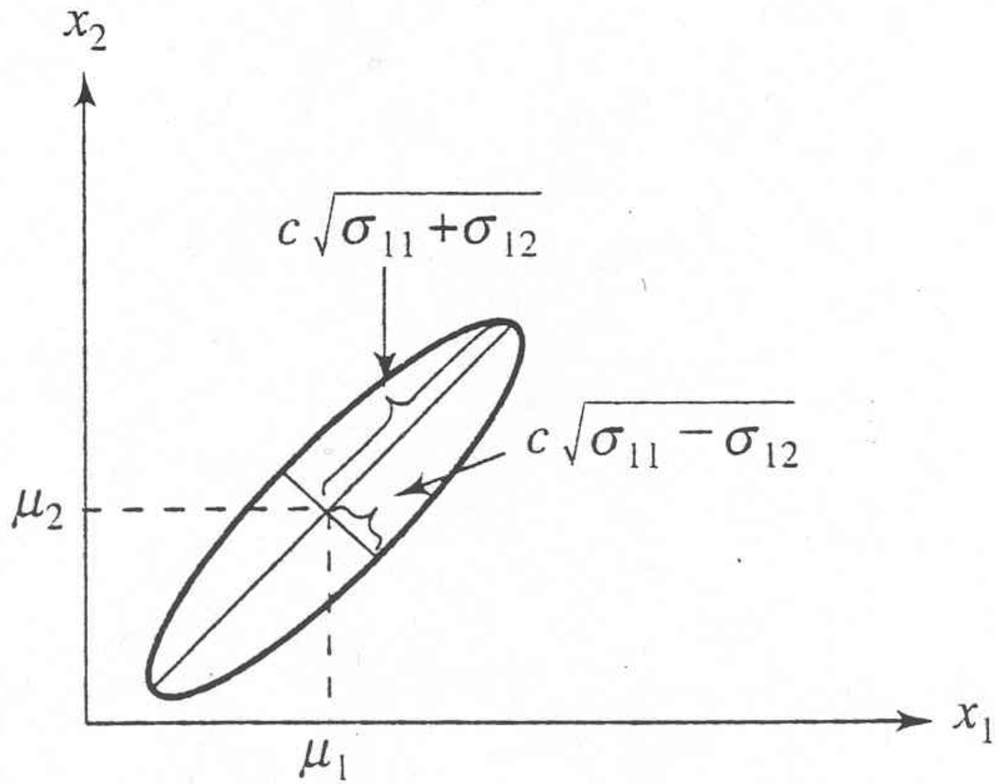
- When $\sigma_{11} = \sigma_{22}$, the formulas simplify to

  $\lambda_1 = \sigma_{11}(1+\rho) = \sigma_{11}+\sigma_{12}$ and $\lambda_2 = \sigma_{11}(1-\rho) = \sigma_{11}-\sigma_{12}$

$$e_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \qquad e_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

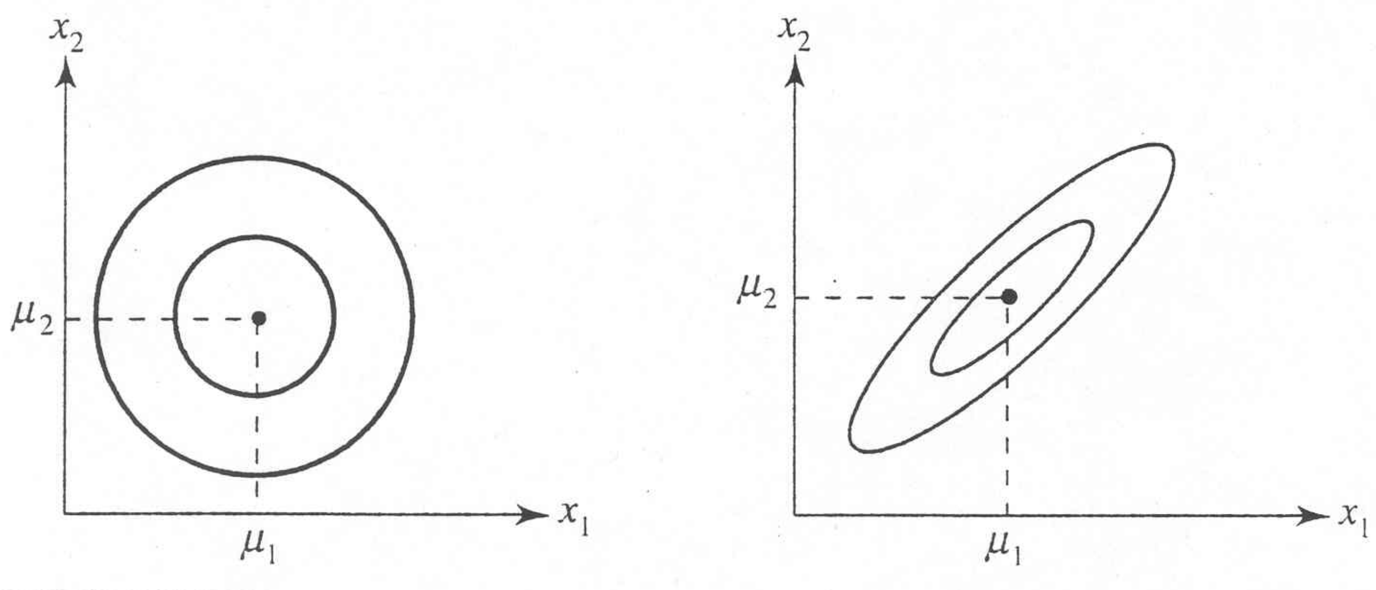
- If $\sigma_{12} > 0$, the major axis of the ellipse will be in the direction of the 45^o line. The actual value of σ_{12} does not matter. If $\sigma_{12} < 0$, then the major axis will be perpendicular to the 45^o line.

Bivariate Normal Contours ($\sigma_{11} = \sigma_{22}$)



More Bivariate Normal Examples

- 50% and 90% contours of two bivariate normal densities. Density is the highest when $\mathbf{x} = \boldsymbol{\mu}$.



Central $(1-\alpha)\times 100\%$ Region of a Bivariate Normal Distribution

- The ratio of the lengths of the major and minor axes is

$$\frac{\text{Length of major axis}}{\text{Length of minor axis}} = \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}}$$

- If $1-\alpha$ is the probability that a randomly selected member of the population is observed inside the ellipse, then the half-length of the axes are given by

$$\sqrt{\chi^2_2(\alpha)}\sqrt{\lambda_i}$$

- This is the smallest region that has probability $1-\alpha$ of containing a randomly selected member of the population

Central $(1-\alpha)\times 100\%$ Region of a Bivariate Normal Distribution

- The area of the ellipse containing the central $(1-\alpha)\times 100\%$ of a bivariate nornal population is

$$area = \pi\chi_2^2(\alpha)\sqrt{\lambda_1}\sqrt{\lambda_2} = \pi\chi_2^2(\alpha)|\Sigma|^{1/2}$$

- Note that

$$det(\Sigma) = det\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = det\begin{bmatrix} \sigma_{11} & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_{22} \end{bmatrix} = \sigma_{11}\sigma_{22}(1-\rho^2)$$

Central $(1-\alpha)\times 100\%$ Region of a Bivariate Normal Distribution

- For fixed variances, σ_{11} and σ_{22}, area of the ellipse is largest when $\rho = 0$
- The area becomes smaller as ρ approaches 1 or as ρ approches -1.
- For $\sigma_{11} = \sigma_{22}$ and $\rho = 0$ the contours of constant density are concentric circles and $\lambda_1 = \lambda_2$
- For $\sigma_{11} > \sigma_{22}$ the axes of the ellipse are parallel to the coordinate axes, with the major axis parallel to the horizontal axis.
- For $\sigma_{22} > \sigma_{11}$ the axes of the ellipse are parallel to the coordinate axes, with the major axis parallel to the vertical axis.

Central $(1 - \alpha) \times 100\%$ Region of a Multivariate Normal Distribution

- For a p-dimensional normal distribution, the smallest region such that there is probability $1 - \alpha$ that a randomly selected observation will fall in the region is
  - a p-dimensional ellipsoid
  - with hypervolume

$$\frac{2\pi^{p/2}}{p\Gamma(p/2)} \left[\chi_p^2(\alpha)\right]^{p/2} |\Sigma|^{1/2}$$

where $\Gamma(\cdot)$ is the gamma function

Gamma Function

$$\Gamma\left(\frac{p}{2}\right) = \left(\frac{p}{2} - 1\right)\left(\frac{p}{2} - 2\right)\cdots(2)(1)$$

when p is an even integer, and

$$\Gamma(\frac{p}{2}) = \frac{(p-2)(p-4)\cdots(3)(1)}{2^{(p-1)/2}}\sqrt{\pi}$$

when p is an odd integer

Overall Measures of Variability

- Generalized variance:

$$|\Sigma| = \lambda_1 \lambda_2 \cdots \lambda_p$$

- Generalized standard deviation

$$|\Sigma|^{1/2} = \sqrt{\lambda_1 \lambda_2 \cdots \lambda_p}$$

- Total variance

$$\begin{aligned} trace(\Sigma) &= \sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp} \\ &= \lambda_1 + \lambda_2 + \cdots + \lambda_p \end{aligned}$$

Sample Estimates: Air Samples

$$X_i = \begin{bmatrix} X_{i1} \leftarrow \ CO \text{ concentration} \\ X_{i2} \leftarrow \ N_2O \text{ concentration} \end{bmatrix}$$

$$X_1 = \begin{bmatrix} 7 \\ 12 \end{bmatrix} \quad X_2 = \begin{bmatrix} 4 \\ 9 \end{bmatrix} \quad X_3 = \begin{bmatrix} 4 \\ 5 \end{bmatrix} \quad X_4 = \begin{bmatrix} 5 \\ 8 \end{bmatrix} \quad X_5 = \begin{bmatrix} 4 \\ 8 \end{bmatrix}$$

Sample Estimates: Air Samples

- Sample mean vector:

$$\bar{X} = \begin{bmatrix} 4.8 \\ 8.4 \end{bmatrix}$$

- Sample covariance matrix:

$$S = \begin{bmatrix} 1.7 & 2.6 \\ 2.6 & 6.3 \end{bmatrix}$$

- Sample correlation: r_{12}=0.7945
- Generalized variance: $|S| = 3.95$
- Total variance; $trace(S) = 1.7 + 6.3 = 8.0$

Example 3.8

$$S = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} \qquad S = \begin{bmatrix} 5 & -4 \\ -4 & 5 \end{bmatrix} \qquad S = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$$

$$r_{12} = 0.8 \qquad r_{12} = -0.8 \qquad r_{12} = 0$$

$$|S| = 9 \qquad |S| = 9 \qquad |S| = 9$$

$$tr(S) = 10 \qquad tr(S) = 10 \qquad tr(S) = 6$$