Covariation, Correlations
Covariance - A number to measure dependence between random variables

\[ \text{Cov}(X,Y) \text{ or } \sigma_{xy} \]

\[ \sigma_{xy} = \mathbb{E} \left[ (X - \mu_x) \cdot (Y - \mu_y) \right] = \]

\[ = \mathbb{E}(X,Y) - \mu_x - \mu_y \]

- \( \text{Var}(X) = \text{Cov}(X,X) \)
- If \( X \& Y \) are independent
  \[ \text{Cov}(X, Y) = \mathbb{E}[X - \mu_x] \cdot \mathbb{E}[Y - \mu_y] = 0 \]
- \( -\infty < \text{Cov}(X, Y) < +\infty \) (can be negative!)


Covariance Defined

Covariance is a number quantifying average dependence between two random variables.

The covariance between the random variables $X$ and $Y$, denoted as $\text{cov}(X, Y)$ or $\sigma_{XY}$, is

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X \mu_Y \quad (5-14)$$

The units of $\sigma_{XY}$ are units of $X$ times units of $Y$.

Unlike the range of variance, $-\infty < \sigma_{XY} < \infty$. 

Sec 5-2 Covariance & Correlation
Covariance and PMF tables

The probability distribution of Example 5-1 is shown.

By inspection, note that the larger probabilities occur as $X$ and $Y$ move in opposite directions. This indicates a negative covariance.

<table>
<thead>
<tr>
<th>$y$ = number of times city name is stated</th>
<th>$x$ = number of bars of signal strength</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.01</td>
<td>0.02</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.02</td>
<td>0.03</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.02</td>
<td>0.10</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.15</td>
<td>0.10</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>
Covariance and Scatter Patterns

Figure 5-13  Joint probability distributions and the sign of $\text{cov}(X, Y)$. Note that covariance is a measure of linear relationship. Variables with non-zero covariance are correlated.
Independence Implies $\sigma = \rho = 0$ but **not vice versa**

- If $X$ and $Y$ are independent random variables,
  \[ \sigma_{XY} = \rho_{XY} = 0 \]  \hspace{1cm} (5-17)

- $\rho_{XY} = 0$ is necessary, but **not a sufficient condition** for independence.

\[ \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \]

*Independent\n
\[ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \]

*NOT independent*
Correlation is “normalized covariance”

- Also called: Pearson correlation coefficient

\[ \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \]

is the covariance normalized to be \(-1 \leq \rho_{xy} \leq 1\)

Karl Pearson (1852–1936)
English mathematician and biostatistician
Prove that $\rho_{xy}$ is in $[-1, 1]$.

$Z_x = \frac{x - \mu_x}{\sigma_x}$; $Z_y = \frac{y - \mu_y}{\sigma_y}$

$0 \leq E((Z_x - Z_y)^2) = E(Z_x^2) + E(Z_y^2) - 2E(Z_x \cdot Z_y) = 2 - 2 \frac{1}{\sigma_x \sigma_y} E((x - \mu_x)(y - \mu_y)) = 2 - 2 \rho_{xy} \implies \rho_{xy} \leq 1$

$0 \leq E((Z_x + Z_y)^2) = E(Z_x^2) + E(Z_y^2) + 2E(Z_x \cdot Z_y) = 2 + 2 \rho_{xy} \implies \rho_{xy} \geq -1$
Spearman rank correlation

- **Pearson correlation** tests for **linear relationship** between X and Y
- **Unlikely** for variables with **broad distributions** → non-linear effects dominate
- **Spearman correlation** tests for any **monotonic relationship** between X and Y
- Calculate ranks (1 to n), \( r_X(i) \) and \( r_Y(i) \) of variables in both samples. Calculate Pearson correlation between ranks: \( \text{Spearman}(X,Y) = \text{Pearson}(r_X, r_Y) \)
- **Ties**: convert to fractions, e.g. tie for 6s and 7s place both get 6.5. This can lead to artefacts.
- If lots of ties: use **Kendall rank correlation** (Kendall tau)
Matlab exercise: Correlation/Covariation

• Generate a sample with Stats=100,000 of two Gaussian random variables \( r_1 \) and \( r_2 \) which have mean 0 and standard deviation 2 and are:
  
  – Uncorrelated
  – Correlated with correlation coefficient 0.9
  – Correlated with correlation coefficient -0.5
  – Trick: first make uncorrelated \( r_1 \) and \( r_2 \). Then change \( r_1 \) to: \( r_{1\text{mix}} = \text{mix} \cdot r_2 + (1-\text{mix}^2)^{0.5} \cdot r_1 \); where \( \text{mix} = \text{corr. coeff} \).

• In each case calculate covariance and correlation coefficient

• In each case make scatter plot: \( \text{plot}(r_{1\text{mix}},r_2,'k.'); \)
Matlab exercise: Correlation/Covariation

1. Stats=100000;
2. r1=2.*randn(Stats,1);
3. r2=2.*randn(Stats,1);
4. disp('Covariance matrix='); disp(cov(r1,r2));
5. disp('Correlation='); disp(corr(r1,r2));
6. figure; plot(r1,r2,'k.');
7. mix=0.9; %Mixes r2 to r1 but keeps same variance
8. r1m=mix.*r2+sqrt(1-mix.^2).*r1;
9. disp('Covariance matrix='); disp(cov(r1m,r2));
10. disp('Correlation='); disp(corr(r1m,r2));
11. figure; plot(r1m,r2,'k.');
12. mix=-0.5; %REDO LINES 8-11
Linear Functions of Random Variables

• A function of random variables is itself a random variable.

• A function of random variables can be formed by either linear or nonlinear relationships. We start with linear functions.

• Given random variables $X_1, X_2, ..., X_p$ and constants $c_1, c_2, ..., c_p$
  \[ Y = c_1 X_1 + c_2 X_2 + ... + c_p X_p \]  (5-24)
  is a linear combination of $X_1, X_2, ..., X_p$. 
Mean & Variance of a Linear Function

\[ Y = c_1 X_1 + c_2 X_2 + \ldots + c_p X_p \]

\[ E(Y) = c_1 E(X_1) + c_2 E(X_2) + \ldots + c_p E(X_p) \]  \hspace{1cm} (5-25)

\[ V(Y) = c_1^2 V(X_1) + c_2^2 V(X_2) + \ldots + c_p^2 V(X_p) + 2 \sum_{i<j} c_i c_j \text{cov}(X_i, X_j) \]  \hspace{1cm} (5-26)

If \( X_1, X_2, \ldots, X_p \) are independent, then \( \text{cov}(X_i, X_j) = 0 \),

\[ V(Y) = c_1^2 V(X_1) + c_2^2 V(X_2) + \ldots + c_p^2 V(X_p) \]  \hspace{1cm} (5-27)
Example 5-31: Error Propagation

A semiconductor product consists of three layers. The variances of the thickness of each layer is 25, 40 and 30 nm². What is the variance of the finished product?

Answer:

\[ X = X_1 + X_2 + X_3 \]

\[ V(X) = \sum_{i=1}^{3} V(X_i) = 25 + 40 + 30 = 95 \text{ nm}^2 \]

\[ SD(X) = \sqrt{95} = 9.7 \text{ nm} \]

*If adding SDs one would get \[ \sqrt{25 \text{ nm}} + \sqrt{40 \text{ nm}} + \sqrt{30 \text{ nm}} = 16.08 \text{ nm} \]
IMPORTANT:

$p$ independent identically distributed (i.i.d.) variables

Average \( \bar{X} = \frac{X_1 + X_2 + X_3 + \ldots + X_p}{p} \)

\[
E(\bar{X}) = \frac{p \cdot E(X)}{p} = \frac{p \cdot \mu}{p} = \mu
\]

\[
V(\bar{X}) = \frac{p \cdot V(X)}{p^2} = \frac{p \cdot \sigma^2}{p^2} = \frac{\sigma^2}{p}
\]

Standard deviation \( (\bar{X}) = \sqrt{V(\bar{X})} = \frac{\sigma}{\sqrt{p}} \)
Mean & Variance of an Average

If \( \bar{X} = \frac{X_1 + X_2 + \ldots + X_p}{p} \) and \( E(X_i) = \mu \)

Then \( E(\bar{X}) = \frac{p \cdot \mu}{p} = \mu \) \hspace{1cm} (5-28a)

If the \( X_i \) are independent with \( V(X_i) = \sigma^2 \)

Then \( V(\bar{X}) = \frac{p \cdot \sigma^2}{p^2} = \frac{\sigma^2}{p} \) \hspace{1cm} (5-28b)
Credit: XKCD comics

why do whales jump
why are witches green
why are there mirrors above beds
why do I say uh
why is sea salt better
why are there trees in the middle of fields
why is there not a pokémon mmo
why is there laughing in tv shows
why are there doors on the freeway
why are there so many schwedexes running
why aren't there any countries in antarctica
why are there scary sounds in minecraft
why is there kokoing in my stomach
why are there two slashes after http
why are there celebrities
why do snakes exist
why do oysters have pearls
why are ducks called ducks
why do they call it the clap
why are kyle and cartman friends
why is there an arrow on aang's head
why are text messages blue
why are there mustaches on clothes
why are there mustaches on cars
why are there mustaches everywhere
why are there so many birds in ohio
why is there so much rain in ohio
why is ohio weather so weird
why are there male and female bikes
why are there tiny spiders in my house
why do spiders come inside
why are there huge spiders in my house
why are there lots of spiders in my house
why are there spiders in my room
why are there so many spiders in my room
why do spiders bite itch
why is dying so scary
why is earth tilted
why is space black
why is outer space so cold
why are there pyramids on the moon
why is nasa shutting down
why is there a line through https
why is there a red line through https on facebook
why is https important
why aren't my arms growing
why are there so many crows in rochester, mn
why are there quadrants of dairy
why is there philos
why are there 47s so expensive
why are there helicopters circling my house
why are there gods
why are there two spoons
why is mt vesuvius there
why do they say t minus
why are there ocelots
why are wrestlers always wet
why are oceans becoming more acidic
why is arwen dying
why aren't my quail laying eggs
why aren't my quail eggs hatching
why aren't there any foreign military bases in america
why aren't there guns in harry potter
why are ultrasounds important
why are underwater pyramids prime
why is stealing wrong
Multivariable statistics and Principal Component Analysis (PCA)

- A table of $n$ observations in which $p$ variables were measured

$p \times p$ symmetric matrix $R$ of corr. coefficients

$$r_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

PCA: Diagonalize matrix $R$
Let’s work with real cancer data!

- Data from Wolberg, Street, and Mangasarian (1994)
- Fine-needle aspirates = biopsy for breast cancer
- Black dots – cell nuclei. Irregular shapes/sizes may mean cancer
- 212 cancer patients and 357 healthy (column 1)
- 30 other properties (see table)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Col 2</th>
<th>Col 12</th>
<th>Col 22</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radius (average distance from the center)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Texture (standard deviation of gray-scale values)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perimeter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoothness (local variation in radius lengths)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compactness (perimeter^2 / area - 1.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concavity (severity of concave portions of the contour)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concave points (number of concave portions of the contour)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symmetry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fractal dimension (&quot;coastline approximation&quot; - 1)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Suppose we have a population measured on $p$ random variables $X_1, \ldots, X_p$. Note that these random variables represent the $p$-axes of the Cartesian coordinate system in which the population resides. Our goal is to develop a new set of $p$ axes (linear combinations of the original $p$ axes) in the directions of greatest variability:

This is accomplished by rotating the axes.

Adapted from slides by Prof. S. Narasimhan, “Computer Vision” course at CMU
Applications of PCA

• Uses:
  – Data Visualization
  – Dimensional Reduction
  – Data Classification

• Examples:
  – How many unique “sub-sets” are in the sample?
  – How are they similar / different?
  – What are the underlying factors that most influence the samples?
  – Which measurements are best to differentiate between samples?
  – How to best present what is “interesting”?
  – Which “sub-set” does this new sample rightfully belong?

Adapted from slides by Prof. S. Narasimhan, “Computer Vision” course at CMU
PCA: General

From $k$ original variables: $x_1, x_2, \ldots, x_k$:

Produce $k$ new variables: $y_1, y_2, \ldots, y_k$:

\[
\begin{align*}
    y_1 &= a_{11} x_1 + a_{12} x_2 + \ldots + a_{1k} x_k \\
    y_2 &= a_{21} x_1 + a_{22} x_2 + \ldots + a_{2k} x_k \\
    &\vdots \\
    y_k &= a_{k1} x_1 + a_{k2} x_2 + \ldots + a_{kk} x_k
\end{align*}
\]

Adapted from slides by Prof. S. Narasimhan, “Computer Vision” course at CMU
PCA: General

From k original variables: $x_1, x_2, ..., x_k$:

Produce k new variables: $y_1, y_2, ..., y_k$:

$y_1 = a_{11}x_1 + a_{12}x_2 + ... + a_{1k}x_k$

$y_2 = a_{21}x_1 + a_{22}x_2 + ... + a_{2k}x_k$

...$

y_k = a_{k1}x_1 + a_{k2}x_2 + ... + a_{kk}x_k$

such that:

$y_k$'s are uncorrelated (orthogonal)
$y_1$ explains as much as possible of original variance in data set
$y_2$ explains as much as possible of remaining variance etc.

Adapted from slides by Prof. S. Narasimhan, “Computer Vision” course at CMU
Adapted from slides by Prof. S. Narasimhan, “Computer Vision” course at CMU
PCA Scores

Adapted from slides by Prof. S. Narasimhan, “Computer Vision” course at CMU
PCA Eigenvalues

Adapted from slides by Prof. S. Narasimhan, “Computer Vision” course at CMU
Choosing the Dimension $K$

- How many eigenvectors to use?
- Look at the decay of the eigenvalues
  - the eigenvalue tells you the amount of variance “in the direction” of that eigenvector
  - ignore eigenvectors with low variance

Adapted from slides by Prof. S. Narasimhan, “Computer Vision” course at CMU
Principle Component Analysis (PCA)

- \( p \times p \) symmetric matrix \( R \) of corr. coefficients \( r_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j} \)
- \( R=n^{-1}Z'Z \) is a “square” of the matrix \( Z \) of standardized r.v.: 
  \[ z_{\alpha k} = \frac{x_{\alpha k} - \mu_k}{\sigma_k} \]  
  all eigenvalues of \( R \) are non-negative
- Diagonal elements=1 \( \Rightarrow \) \( \text{tr}(R)=p \)
- Can be diagonalized: 
  \( R=V*D*V' \) where \( D \) is the diagonal matrix
- \( d(1,1) \) – largest eig. value, \( d(p,p) \) – the smallest one
- The meaning of \( V(i,k) \) – contribution of the data type \( i \) to the \( k \)-th eigenvector
- \( \text{tr}(D)=p \), the largest eigenvalue \( d(1,1) \) absorbs a fraction \( =d(1,1)/p \) of all correlations can be \( \sim 100\% \)
- Scores: \( Y=Z*V \): \( n \times p \) matrix. Meaning of \( Y(\alpha,k) \) – participation of the sample \# \( \alpha \) in the \( k \)-th eigenvector
Let’s work with real cancer data!

- Data from Wolberg, Street, and Mangasarian (1994)
- Fine-needle aspirates = biopsy for breast cancer
- Black dots – cell nuclei. Irregular shapes/sizes may mean cancer
- 212 cancer patients and 357 healthy (column 1)
- 30 other properties (see table)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Col 2</th>
<th>Col 12</th>
<th>Col 22</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radius (average distance from the center)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Texture (standard deviation of gray-scale values)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perimeter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoothness (local variation in radius lengths)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compactness (perimeter² / area - 1.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concavity (severity of concave portions of the contour)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concave points (number of concave portions of the contour)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symmetry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fractal dimension (“coastline approximation” - 1)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Matlab exercise 1

• Download cancer data in cancer_wdbc.mat
• Data in the table X (569x30). First 357 patients are healthy. Next 569-357=212 patients have cancer.
• Calculate the correlation matrix of all-against-all variables: 30*29/2=435 correlations.
  Hint: look at the help page for corr
• Visualize 30x30 table of correlations using pcolor
• Plot the histogram of these 435 correlation coefficients
• Select pairs of variables with strong positive (>0.9) weak positive (~0.5) and the strongest negative (~-0.2) correlation and show the scatter plots of sample data for each pair of variables
Matlab exercise 2

• Carry out PCA of the cancer data
  Use either eigs or pca commands and read the freaking manual

• Which 30 variables give the strongest positive or negative contributions to the 1\textsuperscript{st}, 2\textsuperscript{nd}, and 3\textsuperscript{rd} largest eigenvalues?

• Can PCA be used to separate cancer from normal patients? Plot the scores (S=Z*V) of the 1\textsuperscript{st} vs 2\textsuperscript{nd} eigenvalues for normal and cancer patients separately