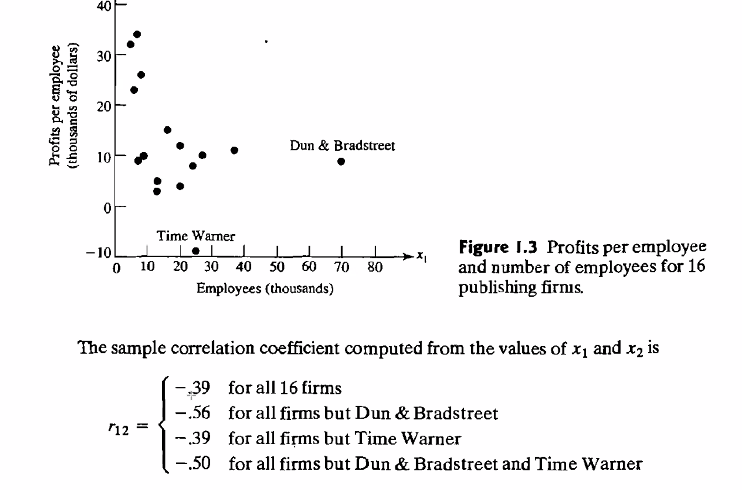
Tuần 2

# **Bài 2**

1. **Nhìn vào đồ thị họ hiểu về 16 công ty như thế nào**

****

Số lượng công nhân không tỉ lệ thuận với năng suất lao động trên đầu người (scc không đổi so với mức trung bình -0.39 khi thiếu công ty Time Warner).

1. **Đọc ví dụ 1.9 rồi nêu ra kết luận:**

Xem biểu đồ 3D Scatter plot xoay theo các chiều hướng khác nhau giúp dễ dàng nhìn ra những trường hợp ngoại lệ (ngoại lai).

Điểm ngoại lai luôn bị phân tán ra xa so với các cá thể còn lại của dữ liệu và có thể ảnh hưởng nặng nề đến các suy luận dựa trên các mô hình tạo dữ liệu chuẩn.

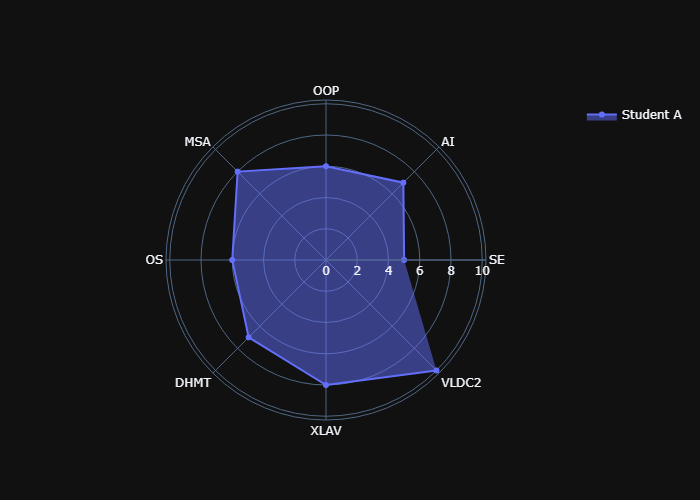
1. **Vẽ trục hình sao điểm 8 môn học :**

| SE | AI | OOP | MSA | OS | DHMT | XLAV | VLDC2 |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 5 | 7 | 6 | 8 | 6 | 7 | 8 | 10 |

1. **Viết python để vẽ đồ thị hình sao:**

| import pandas as pd  import numpy as np  import plotly.express as px  import plotly.graph\_objects as go  # read dataset from csv and perform preprocessing  data = pd.read\_csv('data.csv')  categories = ['SE','AI','OOP','MSA','OS','DHMT','XLAV','VLDC2']  # plot unfilled scatter plot  fig = go.Figure()  fig.add\_trace(go.Scatterpolar(  r=pd.Series(data.loc[0, categories].values),  theta=categories,  fill='toself',  name='Student A'  ))  fig.update\_layout(  polar=dict(  radialaxis=dict(  visible=True,  ),  ),  template='plotly\_dark',  showlegend=True,  )  fig.show()  fig.write\_image("raderChart.png") |
| --- |

**Kết quả:**

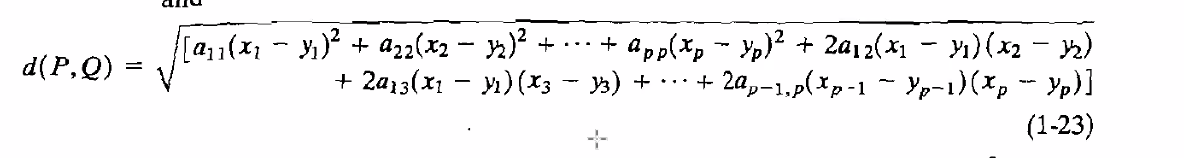
****

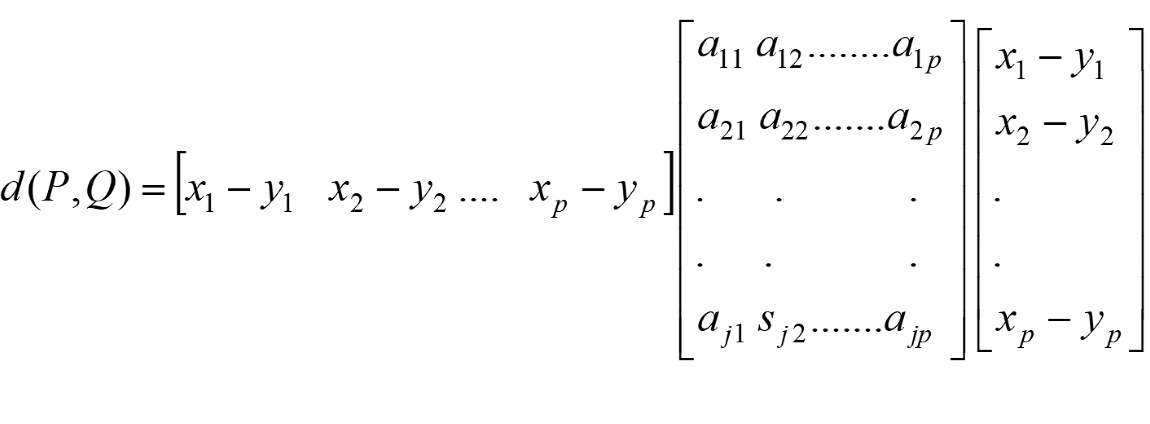
1. **Vẽ biểu tượng khuôn mặt từ năm 1 -> năm 4:**
2. **Ví dụ về tầm quan trọng của statistical distance:**

Nếu tính điểm trung bình học tập mà không sử dụng số tín chỉ để tính thì điểm trung bình đó không thể hiện được tầm quan trọng của những môn quan trọng hơn(có số tín chỉ nhiều hơn)

=> Để đánh giá dữ liệu đúng đắn thì phải xét với khoảng biến thiên dữ liệu (phương sai ).

1. **Viết biểu thức sau dưới dạng ma trận**

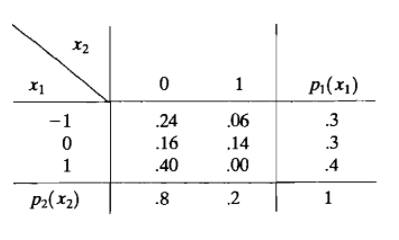
****

****

1. **Giải thích vì sao OP, OQ, OR là như nhau nếu nhìn dưới con mắt thống kê thay vì khoảng cách Euclide.**

N

1. **Tính mean vector/ covariance matrix 2.13**

****

E(X1)=-1\*0.3+0\*0.3+1\*0.4=0.1

E(X2)=0\*0.8+1\*0.2=0.2

Mean vector

[ 0.1

0.2 ]

E(X1-E(X1))(X2-E(X2)) = (-1-0.1)\*(0-0.2)\*0.24 + (-1-0.1)\*(1-0.2)\*0.06 + …+(1-0.1)\*(1-0.2)\*0.0=-0.08

Covariance matrix

[ 0.69 -0.08

-0.08 0.16 ]