

นาย ภัทรพัทธ์ ชัยอมรเวทย์ รหัสนักศึกษา : 62010684

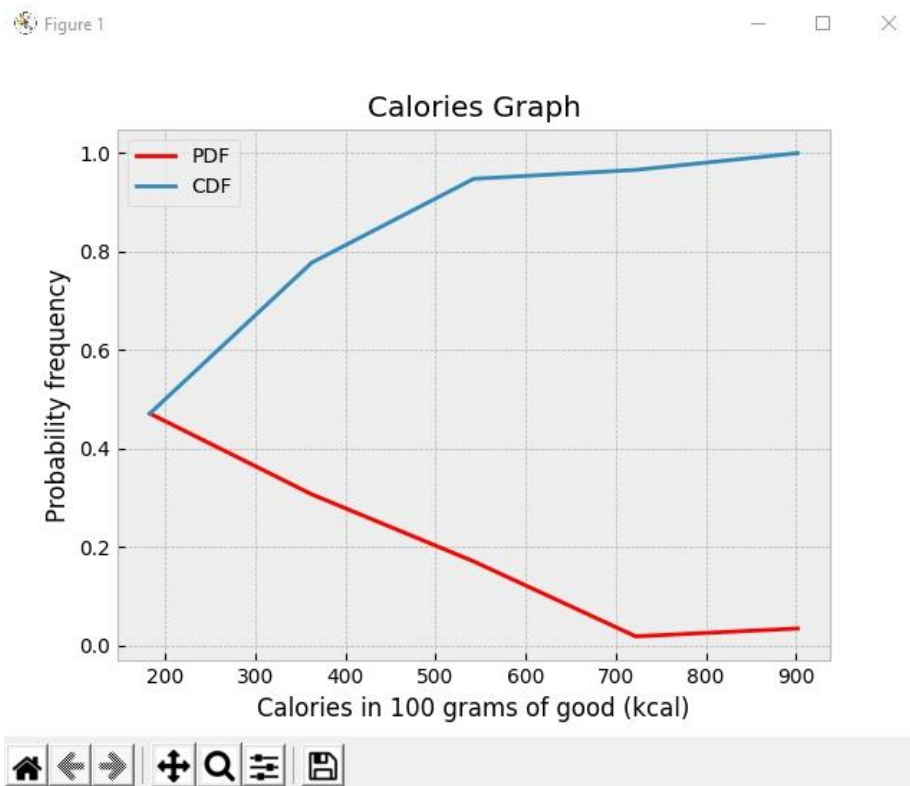
คณะวิศวกรรมศาสตร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

Topic : Probability Density Function / Cumulative Prob Function

ภาษาที่ใช้ : Python

Column ที่เลือกใช้ : Calories / Total_fat

1. Calaires with Probability Density Function / Cumulative Prob Function



ในแกน x ของกราฟ -> จะบอกถึงจำนวน Calories ในอาหารทุกๆ 100 กรัม หน่วย kcal

หน่วยของพลังงานในอาหาร 500 ชนิด มีตั้งแต่ 0-900 kcal ในแกน x จึงมีเลขตั้งแต่ 0-900

ในแกน y ของกราฟ -> จะบอกถึงโอกาสที่จะมี Calories จำนวน ... kcal ในอาหารทุกๆ 100 กรัม ไม่มีหน่วย (prob)

จะคิดจาก จำนวนอาหารที่มี Calories เท่านั้น แล้วหารด้วยจำนวน Calories เช่น

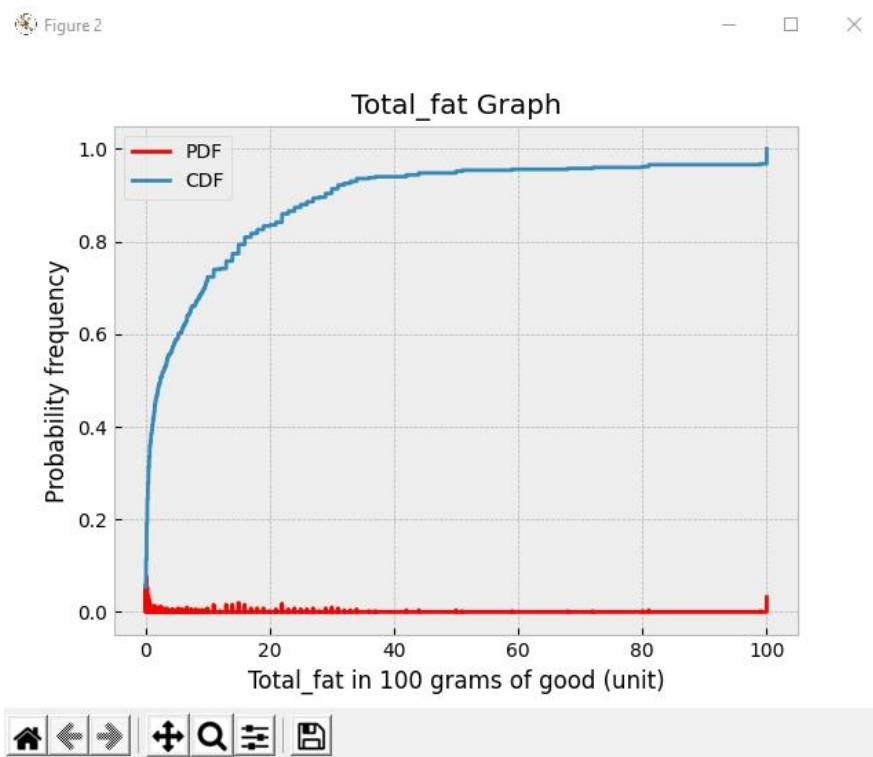
$0.48 = \text{prob}(\text{อาหารที่มี } 200 \text{ kcal ต่อ } 100 \text{ g})$

ดังนั้น อาหารที่มี 200 kcal ต่อ 100 g จะมีค่าเท่ากับ $0.48 * 500 = 240$ อย่าง

เมื่อเราได้ความน่าจะเป็นนี้มาแล้ว เราสามารถหาจำนวนของอาหารให้ได้

หมายเหตุ : ในการวิเคราะห์ครั้งนี้ nutrition.scv มีข้อมูลอยู่ 500 rows หมายความว่า มีอาหารอยู่ 500 อย่าง

2.Total_fat with Probability Density Function / Cumulative Prob Function



ในแกน x ของกราฟ -> จะบอกถึงจำนวนไขมัน (total_fat) ในอาหารทุกๆ 100 กรัม ไม่มีหน่วย

หน่วยไขมันของในอาหาร 500 ชนิด มีตั้งแต่ 0-100 หน่วย ในแกน x จึงมีเลขตั้งแต่ 0-100

ในแกน y ของกราฟ -> จะบอกถึงโอกาสที่จะมี Total fat จำนวน... หน่วย ในอาหารทุกๆ 100 กรัม ไม่มีหน่วย (prob)

โดยคิดเหมือน calories แต่ข้อมูลจะต่างกันเนื่องจากช่วงของหน่วยมีค่าต่างกัน 0-100 และ 0-900

บทวิเคราะห์

จากกราฟที่ผ่านมา เราได้ทำการวิจัยข้อมูล 2 คอลัมน์ โดยแต่ละคอลัมน์จะมีกราฟอยู่ 2 ประเภท ได้แก่

Probability Density Function (PDF) และ Cumulative Prob Function (CPF) ซึ่งแต่ละกราฟจะบ่งบอกข้อมูลต่างกันไป

PDF graph จะบ่งบอกว่าข้อมูลนี้แบ่งได้เป็นส่วนๆเท่าไร และเมื่อข้อมูลขนาดเท่านี้จะมีแนวโน้มจะเป็นเท่าไร ส่วน CDF graph จะบ่งบอกว่าข้อมูลนี้มีการสะสมความน่าจะเป็นมากน้อยแค่ไหน การสะสมความน่าจะเป็นของข้อมูลนั้น จะขึ้นอยู่กับความชันของกราฟ ถ้าความชันมาก ข้อมูลในช่วงนั้น ก็จะมีมาก ถ้าความชันน้อย ข้อมูลในช่วงนั้น ก็จะมีน้อย

จากกราฟข้างต้น อาหารที่นำมาวิจัย 500 อย่าง และแต่ละอาหารจะมีขนาด 100 กรัม ความน่าจะเป็นที่อาหารมีพลังงาน (Calories) ที่มีค่าสูงที่สุดก็คือ 0.48 ซึ่ง ค่าพลังงานก็ประมาณ 170 kcal ทางฝั่งไขมันความน่าจะเป็นสูงที่สุดก็คือ 0.9 กว่าๆ โดยจะมีไขมันอยู่ที่ 90 กว่า หน่วย จะสรุปผลได้ว่า อาหารที่นำมาทำการวิจัย 500 อย่างนั้น เราได้พบความน่าจะเป็นของไขมัน มีค่ามากกว่าความน่าจะเป็นของค่าพลังงาน หมายความว่า เราสามารถค้นหาข้อมูลของคอลัมน์ไขมันได้ง่าย

กว่าเมื่อเปรียบเทียบกับค่าพลังงานในอาหาร **Prob max (ไขมัน) 0.9 , Prob max (ค่าพลังงาน)** อย่างไรก็ตาม อาหารประกอบไปด้วยสารอาหารมากมาย ตัวความน่าจะเป็นอาจจะขึ้นอยู่กับส่วนผสมของอาหารอย่างอื่นด้วย เช่น ค่าพลังงานอาจจะมาจากไขมัน โปรตีน และ คาร์โบไฮเดรต

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3 import pandas as pd
4
5 plt.style.use('bmh')
6 pd.options.mode.chained_assignment = None
7 df = pd.read_csv('nutrition.csv')
8 df.dropna(inplace=True)
9
10 Calories_rate, Calories_rating = np.histogram(df["calories"], bins=5)
11 totalfat_review, totalfat_review = np.histogram(df["total_fat"], bins=1000000)
12 pdf_rating = Calories_rate / sum(Calories_rate)
13 pdf_review = Calories_rate / sum(Calories_rate)
14 cdf_rating = np.cumsum(pdf_rating)
15 cdf_review = np.cumsum(pdf_review)
16 plt.figure()
17 plt.plot(Calories_rating[1:], pdf_rating, color="red", label="PDF")
18 plt.plot(Calories_rating[1:], cdf_rating, label="CDF")
19 plt.title("Calories Graph")
20 plt.xlabel('Calories in 100 grams of good (kcal)')
21 plt.ylabel('Probability frequency')
22 plt.figure()
23 plt.plot(totalfat_review[1:], pdf_review, color="red", label="PDF")
24 plt.plot(totalfat_review[1:], cdf_review, label="CDF")
25 plt.title("Total_fat Graph")
26 plt.xlabel('Total_fat in 100 grams of good (unit)')
27 plt.ylabel('Probability frequency')
28 plt.show()
```

Source Code Python

Matplotlib.pyplot -> plot graph และแสดง label ต่างๆ

Pandas -> อ่านข้อมูลและคอลัมน์จากไฟล์ csv