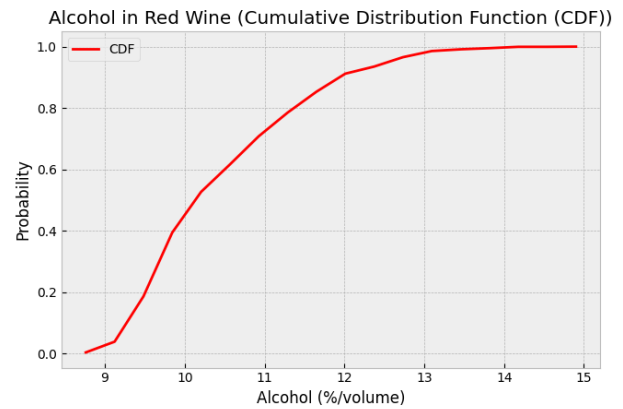
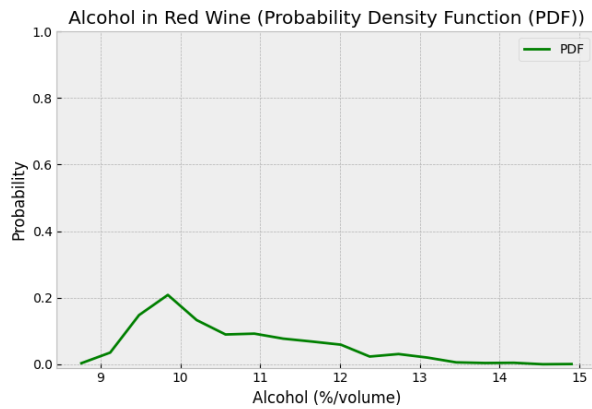


Homework #3

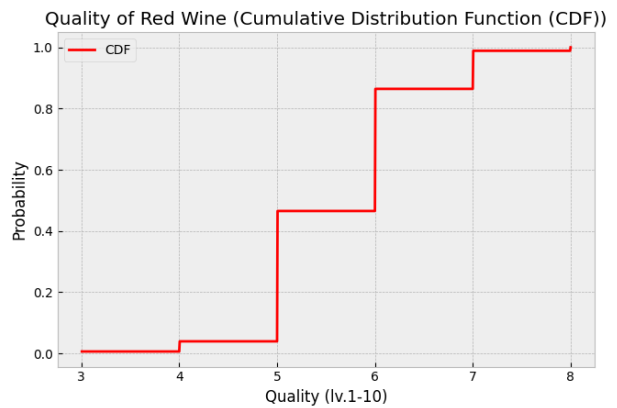
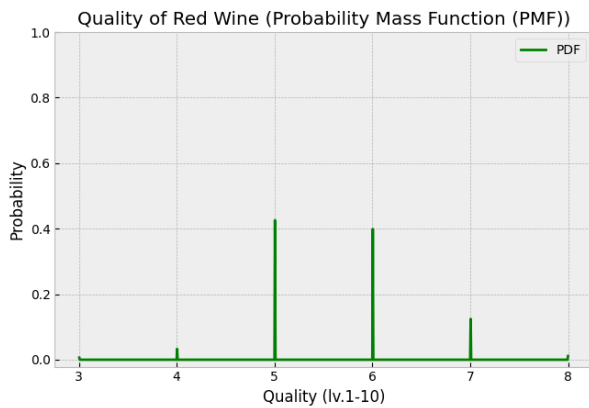
กราฟ Probability Density Function (PDF)/ Cumulative Distribution Function (CDF)

ปริมาณแอลกอฮอล์ในไวน์แดง



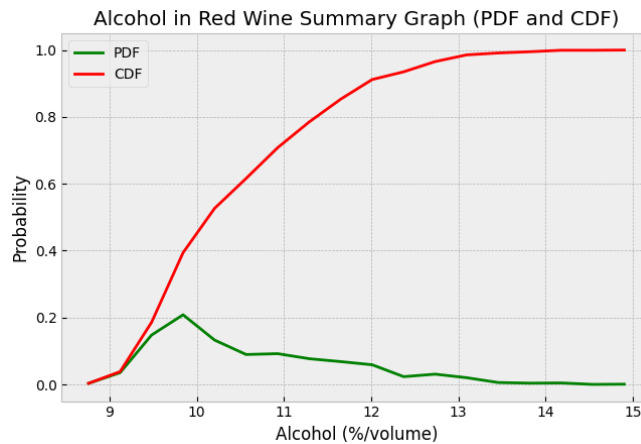
กราฟ Probability Mass Function (PMF)/ Cumulative Distribution Function (CDF)

คุณภาพของไวน์แดง



บทวิเคราะห์ข้อมูลจากกราฟทั้งหมด

จะเห็นว่าเมื่อนำข้อมูลกราฟทั้งสองมารวมกัน จะเห็นความสัมพันธ์ระหว่างกราฟ PDF - CDF และ PMF - CDF ดังนี้



ปริมาณแอลกอฮอล์ในไวน์แดง

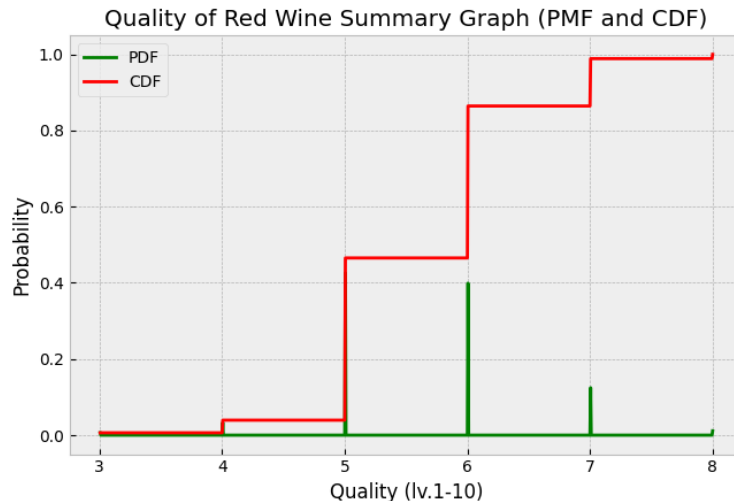
ในกราฟปริมาณแอลกอฮอล์ **เส้นสีเขียว (PDF)** สังเกตได้ ดังนี้

1. จะมีลักษณะเป็น **ภูเขาสูง** ระหว่างช่วงปริมาณแอลกอฮอล์ตั้งแต่ 9.2 – 10.5 %/volume
2. จะมีลักษณะเป็น **ที่ราบแบน** ระหว่างช่วงปริมาณแอลกอฮอล์ตั้งแต่ 8.5 – 9.1 %/volume และ 10.6 – 15.0 %/volume

ในกราฟปริมาณแอลกอฮอล์ **เส้นสีแดง (CDF)** สังเกตได้ ดังนี้

1. เป็นการรวมค่าความน่าจะเป็น(ความถี่สัมพัทธ์) จากเส้นกราฟสีเขียว เพิ่มขึ้นเรื่อย ๆ สะสมจนครบ 1.0 เมื่อสิ้นสุดกราฟนั้น ๆ ซึ่งเป็นความถี่สะสมสัมพัทธ์
2. จะมีค่าความน่าจะเป็น = 0.5 ที่ประมาณ 10.2 %/volume ซึ่งเป็นจุดกึ่งกลางของกราฟพอดี (Median) ทำให้ทางซ้ายและทางขวาของพื้นที่ใต้กราฟสีเขียวแบ่งออกเป็น 50% เท่า ๆ กัน
3. ความชันของกราฟช่วงประมาณ 9.2 - 9.7 %/volume มีความชันโดยรวมมาก และความชันจะค่อย ๆ ลดหลั่นลงไปเรื่อย ๆ เพื่อมีปริมาณแอลกอฮอล์มากขึ้น

กล่าวโดยสรุป จากข้อมูลที่ได้ ไวน์จะมีปริมาณแอลกอฮอล์ประมาณ 9.2 – 10.5 %/volume เป็นส่วนมาก เนื่องจากมีพื้นที่ใต้กราฟเส้นสีเขียวเยอะ (Density หนาแน่น) และไวน์จะมีปริมาณแอลกอฮอล์ประมาณ 8.5 – 9.1 %/volume และ 10.6 – 13.0 %/volume เป็นส่วนที่น้อย เนื่องจากพื้นที่ใต้กราฟเส้นสีเขียวน้อย (Density เบาบาง) และถ้ายังมีปริมาณแอลกอฮอล์ที่มากขึ้นไปกว่า 13.0 %/volume ก็จะมีจำนวนไวน์ที่จะลดลงตามลำดับลงไป (Density เบาบางมาก ๆ) ซึ่งข้อมูลข้างต้นสามารถนำมาดูเปรียบเทียบกับได้กับเส้นสีแดงควบคู่กัน ถ้าเส้นสีแดงมีความชันมาก ก็จะมีการเพิ่มจำนวนไวน์ในปริมาณแอลกอฮอล์ช่วงนั้นที่มากขึ้น ส่วนถ้ามีความชันน้อย ก็จะมีการเพิ่มจำนวนไวน์ในปริมาณแอลกอฮอล์ช่วงนั้นที่น้อยลง



คุณภาพของไวน์แดง

ในกราฟคุณภาพไวน์แดง **เส้นสีเขียว (PMF)** สังเกตได้ ดังนี้

1. คุณภาพระดับ 5 และ 6 จะมีปริมาณมาก เมื่อเทียบกับคุณภาพไวน์อื่นๆ เนื่องจากมีแท่งสีเขียวที่ค่อนข้างสูงโดดเด่นเป็นพิเศษ และมีค่าความน่าจะเป็น(ความถี่สัมพัทธ์) โดยประมาณ = 0.4
2. คุณภาพระดับ 7 จะมี ความน่าจะเป็น(ความถี่สัมพัทธ์) โดยประมาณ = 0.1
3. คุณภาพระดับ 3 4 และ 8 จะมี ความน่าจะเป็น(ความถี่สัมพัทธ์) โดยประมาณน้อยกว่า 0.05 ซึ่งเป็นค่าที่น้อยมากๆ

ในกราฟคุณภาพไวน์แดง **เส้นสีแดง (CDF)** สังเกตได้ ดังนี้

1. เป็นการรวมค่าความน่าจะเป็น(ความถี่สัมพัทธ์) จากเส้นกราฟสีเขียว เพิ่มขึ้นเรื่อย ๆ สะสมจนครบ 1.0 เมื่อสิ้นสุดกราฟนั้น ๆ ซึ่งเป็นความถี่สะสมสัมพัทธ์
2. จะมีค่าความน่าจะเป็น = 0.5 ที่คุณภาพระดับ 6 ซึ่งเป็นจุดกึ่งกลางของกราฟพอดี (Median)
3. ความชันของกราฟคุณภาพระดับที่ 5 และ 6 จะลากยาว และมีความสูงเป็นพิเศษกว่าระดับคุณภาพอื่นๆ เนื่องจากมีจำนวนคุณภาพระดับ 5 และ 6 ที่มีเยอะมากกว่าระดับอื่นๆ

กล่าวโดยสรุป จากข้อมูลที่ได้จากเส้นสีเขียว ไวน์จะมีระดับคุณภาพอยู่ที่ประมาณ 5 และ 6 เป็นส่วนใหญ่ (คิดเป็นประมาณ 80% ของไวน์ทั้งหมด) รองลงมาไวน์จะมีคุณภาพอยู่ในระดับ 7 มีปริมาณปานกลาง(คิดเป็นประมาณ 12% ของไวน์ทั้งหมด) และ คุณภาพระดับ 3 4 และ 8 มีปริมาณที่ค่อนข้างน้อย (คิดเป็นประมาณ 8% ของไวน์ทั้งหมด) ซึ่งสามารถนำข้อมูลข้างต้นมาดูเปรียบเทียบกับเส้นสีแดงควบคู่กัน ถ้าเส้นสีแดงมีความชันมาก ก็จะมีการจำนวนคุณภาพของไวน์ในระดับนั้นมาก ถ้ามีความชันน้อย ก็จะมีจำนวนคุณภาพของไวน์ในระดับนั้นน้อย

จากการวิเคราะห์ทั้งหมด จะมีข้อสรุปได้ว่า ไวน์ส่วนใหญ่ที่ได้จากการผลิตในความสัมพันธ์ระหว่างปริมาณแอลกอฮอล์ในไวน์และคุณภาพของไวน์ ปริมาณแอลกอฮอล์ในไวน์ช่วงตั้งแต่ประมาณ 9.2 – 10.5 %/volume จะมีไวน์ปริมาณมากที่อยู่ในช่วงนี้ ซึ่งจะมีคุณภาพของไวน์เฉลี่ยอยู่ในระดับ 5 และ 6 เป็นส่วนใหญ่ด้วยเช่นกัน ถือว่าไวน์ส่วนใหญ่ที่มีระดับแอลกอฮอล์ 9.2 – 10.5 %/volume จะมีระดับคุณภาพที่พอใช้-ดี ในขณะที่ปริมาณแอลกอฮอล์ในไวน์ในช่วงตั้งแต่ประมาณ 8.5-9.1 %/volume และ 10.6 – 13.0 %/volume เป็นส่วนที่มีจำนวนไวน์ที่หาได้ยากหรือมีจำนวนที่ลดลง จะมีคุณภาพของไวน์เฉลี่ยอยู่ในระดับ 7 จึงอาจตีความได้ว่า เมื่อปริมาณแอลกอฮอล์ในไวน์ยิ่งสูงขึ้น จำนวนไวน์จะยิ่งหาได้ยากขึ้นหรือมีจำนวนที่ลดลง คุณภาพของไวน์อาจจะเพิ่มขึ้นไปอยู่ที่ระดับ 7 ซึ่งเป็นคุณภาพในระดับที่ดี-ดีมาก ส่วนคุณภาพในระดับ 3 4 และ 8 จะไม่นำมาคิด เพราะมีค่าความน่าจะเป็นที่น้อยมากๆ จากค่า outlier ของข้อมูล รวมถึงไม่คิดในส่วนปริมาณแอลกอฮอล์ที่มากกว่า 13.0 %/volume ด้วย

ดังนั้น ปริมาณแอลกอฮอล์ในไวน์แดงที่มากขึ้น อาจมีแนวโน้มที่จะทำให้คุณภาพของไวน์แดงเพิ่มขึ้นตามไปด้วย โดยความสัมพันธ์ระหว่างปริมาณแอลกอฮอล์ในไวน์แดง และคุณภาพของไวน์แดงเป็นความสัมพันธ์แบบแปรผันตรงหรือคล้ายตามกัน

อย่างไรก็ตาม ทั้งหมดนี้ยังไม่สามารถกล่าวได้อย่างชัดเจน 100% เป็นเพียงแค่แนวโน้มเท่านั้น เนื่องจากการผลิตไวน์จริง จะมีส่วนผสมอื่นๆ และมีอีกหลายปัจจัยในการกำหนดคุณภาพของไวน์แดง เช่น ค่าความเป็นกรด, น้ำตาลคงค้างที่เหลือในไวน์แดง, ระยะเวลาการผลิตไวน์แดง, คุณภาพขององุ่นที่นำมาใช้ในการผลิต และ เกณฑ์การวัดคุณภาพของไวน์แดง เป็นต้น ซึ่งเกณฑ์การวัดคุณภาพของไวน์แดง (ระดับ 1-10) ในครั้งนี้ อ้างอิงมาจากโรงงานผลิตไวน์ ในจังหวัด Minho ทางตอนเหนือของประเทศโปรตุเกสเท่านั้น



รายละเอียด Source Code ของโปรแกรม WineGraph2.py

```

wineGraph2.py X
wineGraph2.py > ...
    You, 14 minutes ago | 1 author (You)
1  # defining the libraries
2
3  import numpy as np
4  import matplotlib.pyplot as plt
5  import pandas          # collection for data
6
7
8  # Init Style of Graph and Insert table of data in form of columns
9  plt.style.use('bmh')
10 columns = pandas.read_csv('testgraphredwine.csv')
11
12 # All columns
13 x = columns['alcohol']      # x (independent variable) = alcohol
14 y = columns['quality']      # y (dependent variable) = quality
15
16
17 # getting data of the histogram
18 al_count, al_bins_count = np.histogram(x, bins=18)
19 qu_count, qu_bins_count = np.histogram(y, bins=1000) #6
20
21 # finding the PDF of the histogram using count values
22 al_pdf = al_count / sum(al_count)
23 al_cdf = np.cumsum(al_pdf)
24
25 qu_pdf = qu_count / sum(qu_count)
26 qu_cdf = np.cumsum(qu_pdf)
27
28
29 figure, al_func = plt.subplots(1, 2, figsize=(14, 5))
30 plt.tight_layout(pad=4, w_pad=3, h_pad=1.0)
31
32 al_func[0].set_title('Alcohol in Red Wine (Probability Density Function (PDF))')
33 al_func[0].set_xlabel("Alcohol (%/volume)")
34 al_func[0].set_ylabel("Probability")
35 al_func[0].plot(al_bins_count[1:], al_pdf, color="green", Label="PDF", )
36 al_func[0].legend()
37 al_func[0].axis(ymax=1)
38
39 al_func[1].set_title('Alcohol in Red Wine (Cumulative Distribution Function (CDF))')
40 al_func[1].set_xlabel("Alcohol (%/volume)")
41 al_func[1].set_ylabel("Probability")
42 al_func[1].plot(al_bins_count[1:], al_cdf, color="red", Label="CDF")
43 al_func[1].legend()
44
45
46 figure, qu_func = plt.subplots(1, 2, figsize=(14, 5))
47 plt.tight_layout(pad=4, w_pad=3, h_pad=1.0)
48
49 qu_func[0].set_title('Quality of Red Wine (Probability Mass Function (PMF))')
50 qu_func[0].set_xlabel("Quality (lv.1-10)")
51 qu_func[0].set_ylabel("Probability")
52 qu_func[0].plot(qu_bins_count[1:], qu_pdf, color="green", Label="PDF", )
53 qu_func[0].legend()
54 qu_func[0].axis(ymax=1)
55
56 qu_func[1].set_title('Quality of Red Wine (Cumulative Distribution Function (CDF))')
57 qu_func[1].set_xlabel("Quality (lv.1-10)")
58 qu_func[1].set_ylabel("Probability")
59 qu_func[1].plot(qu_bins_count[1:], qu_cdf, color="red", Label="CDF")
60 qu_func[1].legend()

```

```
63 figure, total = plt.subplots(1, 2, figsize=(14, 5))
64 plt.tight_layout(pad=4, w_pad=3, h_pad=1.0)
65
66 total[0].set_title('Alcohol in Red Wine Summary Graph (PDF and CDF)')
67 total[0].set_xlabel("Alcohol (%/volume)")
68 total[0].set_ylabel("Probability")
69 total[0].plot(al_bins_count[1:], al_pdf, color="green", label="PDF")
70 total[0].plot(al_bins_count[1:], al_cdf, color="red", label="CDF")
71 total[0].legend()
72
73 total[1].set_title('Quality of Red Wine Summary Graph (PMF and CDF)')
74 total[1].set_xlabel("Quality (lv.1-10)")
75 total[1].set_ylabel("Probability")
76 total[1].plot(qu_bins_count[1:], qu_pdf, color="green", label="PDF")
77 total[1].plot(qu_bins_count[1:], qu_cdf, color="red", label="CDF")
78 total[1].legend()
79
80
81
82 plt.show()
83
```

แหล่งที่มาของชุดข้อมูล (Reference/URL) :

- ที่มาของชุดข้อมูล Winequality-red.csv

<https://www.kaggle.com/uciml/red-wine-quality-cortez-et-al-2009>

- ที่มาคำอธิบายแต่ละส่วนประกอบของไวน์

<https://archive.ics.uci.edu/ml/datasets/wine+quality>

https://rstudio-pubs-static.s3.amazonaws.com/57835_c4ace81da9dc45438ad0c286bcbb4224.html

<https://waterlibrary.com/th-รู้ไหมว่า-ระดับปริมาณแ:/#:~:text=ปัจจุบันมีแอลกอฮอล์อยู่ใน,สูงขึ้นด้วยเช่นกัน>

- วิธีการทำไวน์

https://www.youtube.com/watch?v=7gguYRxLMFI&ab_channel=Insider

- ประเภทของไวน์

<https://www.unlockmen.com/terrazas-unlock-wine-101-1/>

<https://thewinelist.shop/blog/news/wine-101>

- รายละเอียดอื่นๆ เกี่ยวกับคุณภาพและวิธีรับรสที่ดีของไวน์

<https://www.blockdit.com/posts/5e5f68d77b00780ed6462939>

<https://www.dummies.com/food-drink/drinks/wine/the-special-technique-for-tasting-wine/>

<https://www.quickanddirtytips.com/house-home/entertaining/wine/4-ways-to-know-if-your-wine-is-good>