

2021/03/30

---

# 靜電量測異常機率分析

RYAN.CHOU

---



---

# 問題與目的

---

- ❖ 說明：目前公司每人每日於作業前，皆會執行靜電量測，若量測結果高於1000M與低於1M則判定為FAIL，需進行重測，直到量測結果落於區間內才可執行生產作業。
- ❖ 單位：阻值(歐姆)
- ❖ 目的：找出量測機率較高的人員，並給於協助，例：更換靜電鞋，檢視人員穿著的服裝。



---

# 資料清理

---

- ❖ 資料處理:

- ❖ 資料字串處理：

- ❖ 1K=1,000

- ❖ 1M=1,000,000

- ❖ 1G=1,000,000,000

- ❖ 後續皆以1M作為一個單位

- ❖ 定義左右腳

- ❖ 清理後df:

- ❖  $216488 \times 13 > 136045 \times 14$

- ❖ 列入分析人員:

- ❖ 848人>350人

```
# 日期處理
```

```
df['DATE']=df['DATE'].apply(lambda x:datetime.strptime(x, "%Y-%m-%d"))
```

```
# 刪除測試值大於2000的資料（極端值）
```

```
df=df[df['DATA']<=2000]
```

```
# 刪除測試次數小於30的人
```

```
df_valuesCounts=df['ID'].value_counts().reset_index()
```

```
df_valuesCounts.columns=['ID','TEST_QTY']
```

```
df=pd.merge(df,df_valuesCounts,how='left',on='ID')
```



# 資料分佈

❖ 左圖顯示全公司的直方圖(上)與核密度函數曲線圖(下)，並將資料依冬天、夏天、左腳、右腳進行區分。

❖ 直方圖：

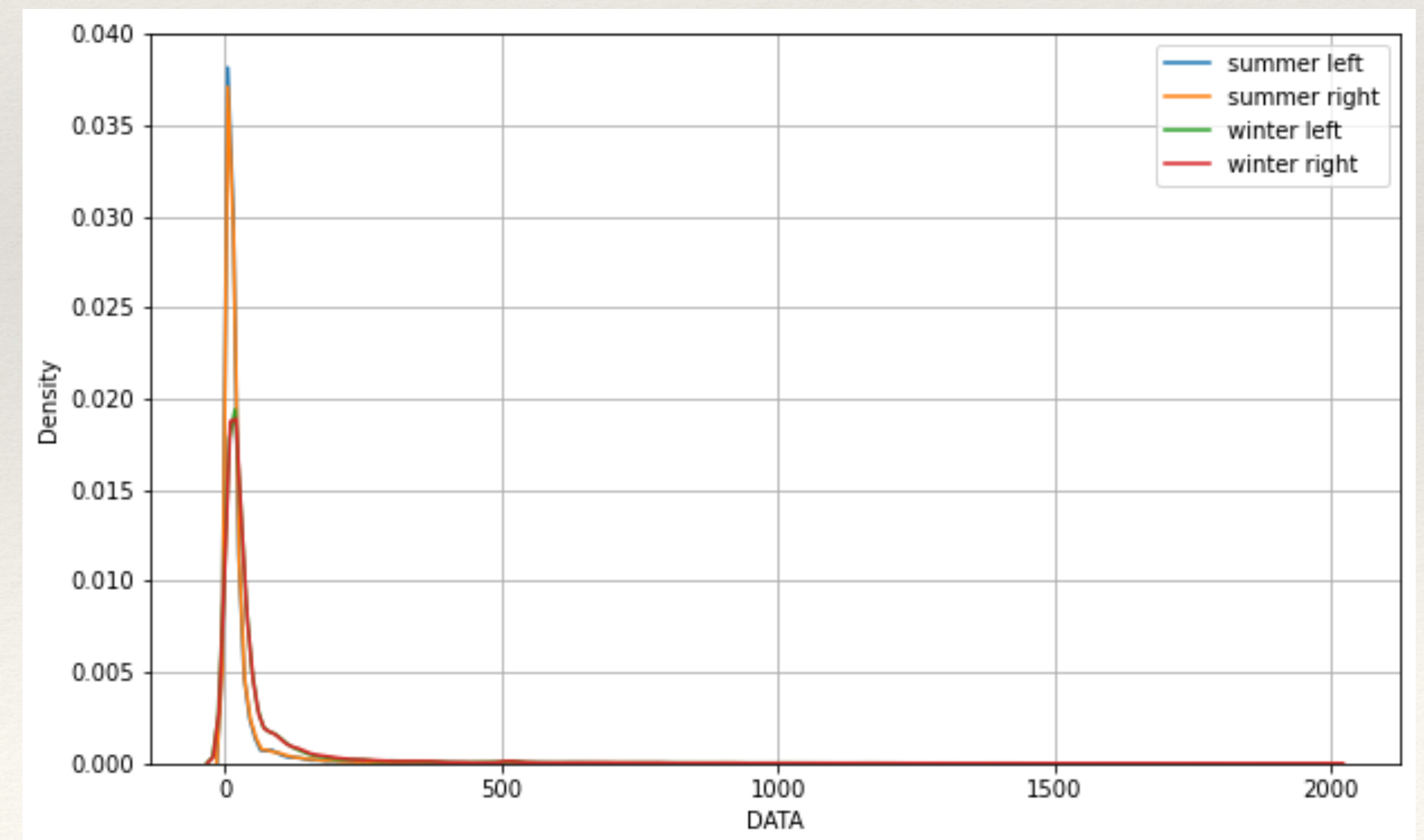
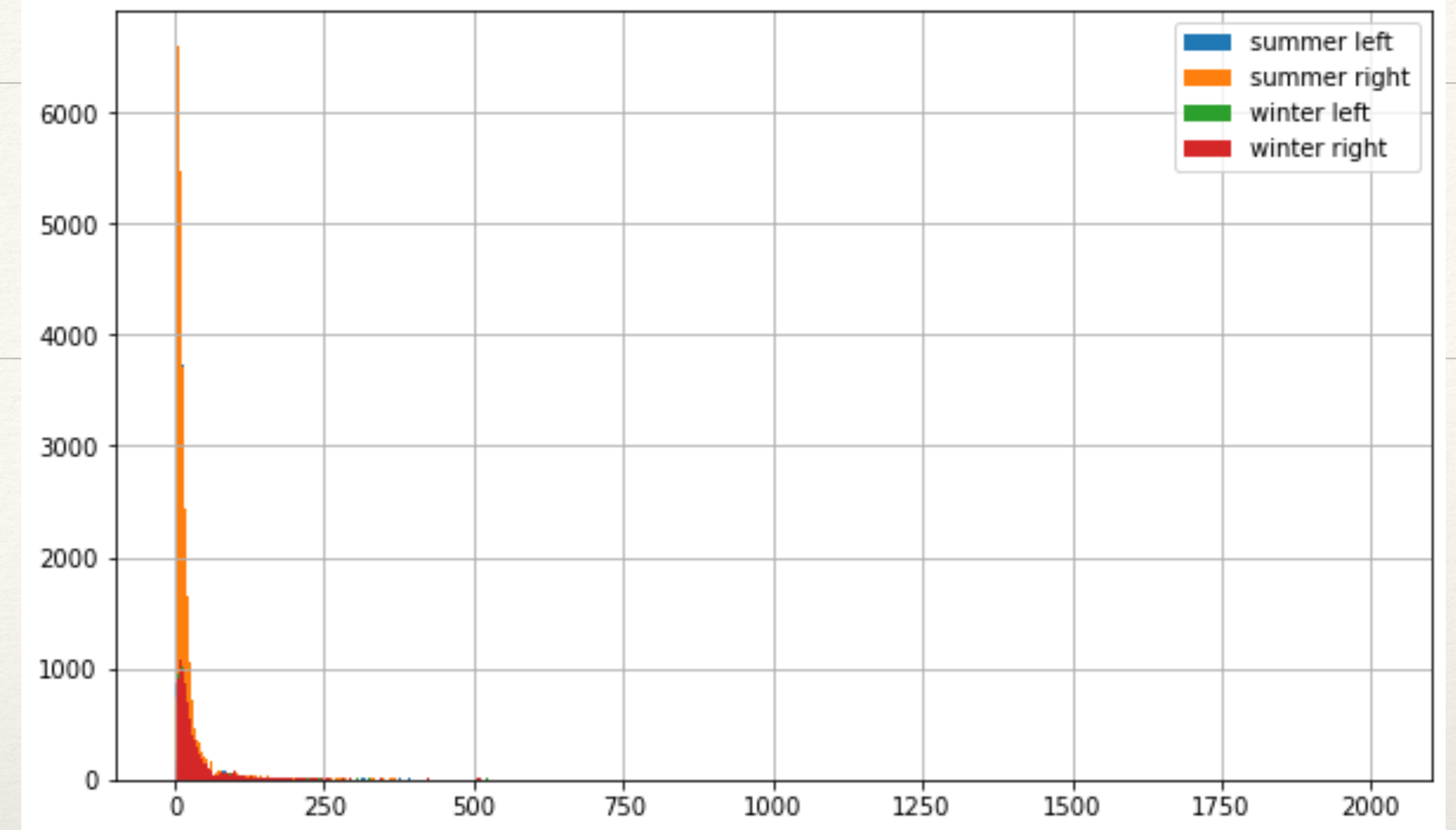
❖ `import matplotlib.pyplot as plt`

❖ `plt.hist(df['DATA'],bins=1000)`

❖ 核密度函數曲線圖：

❖ `import seaborn as sns`

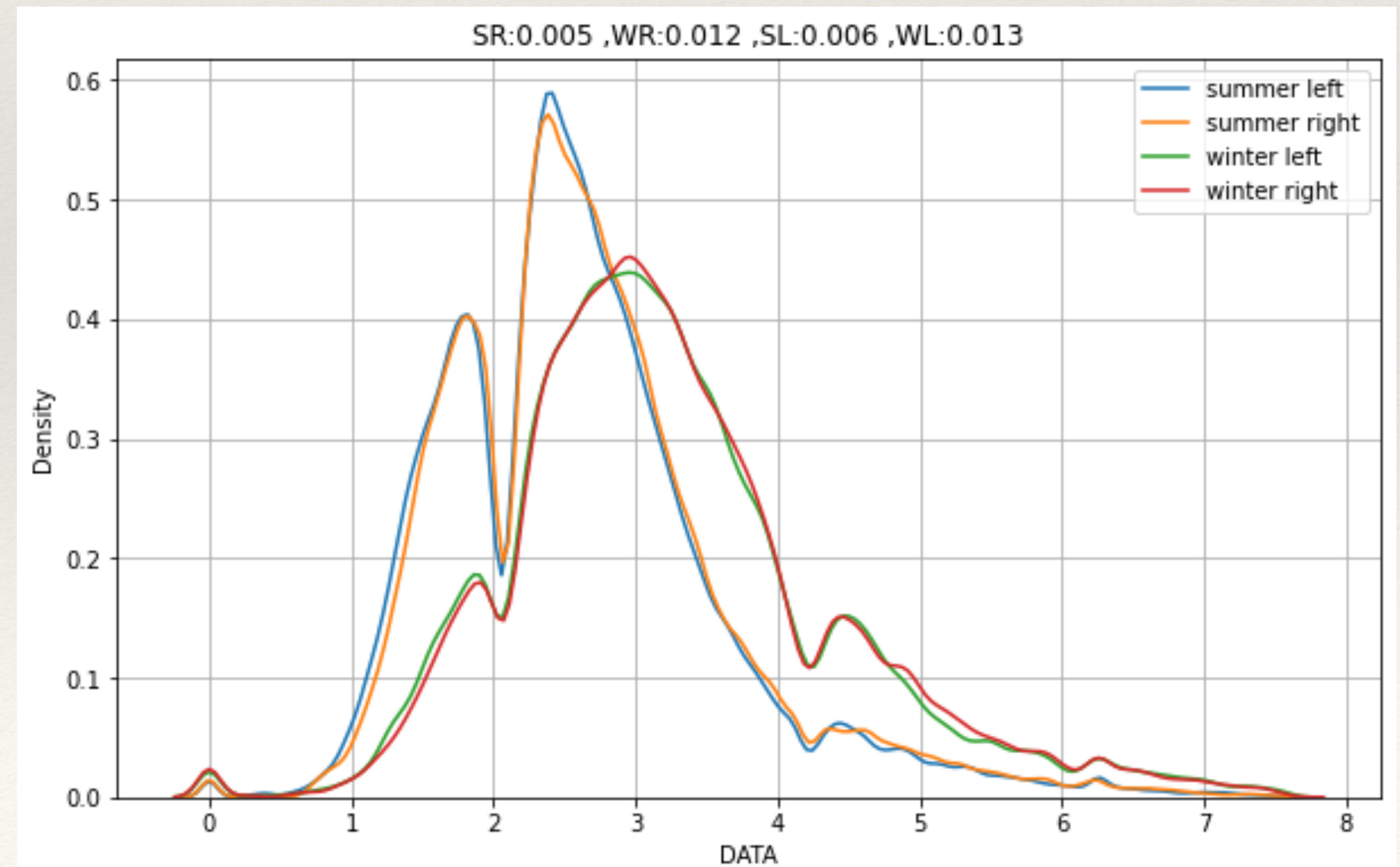
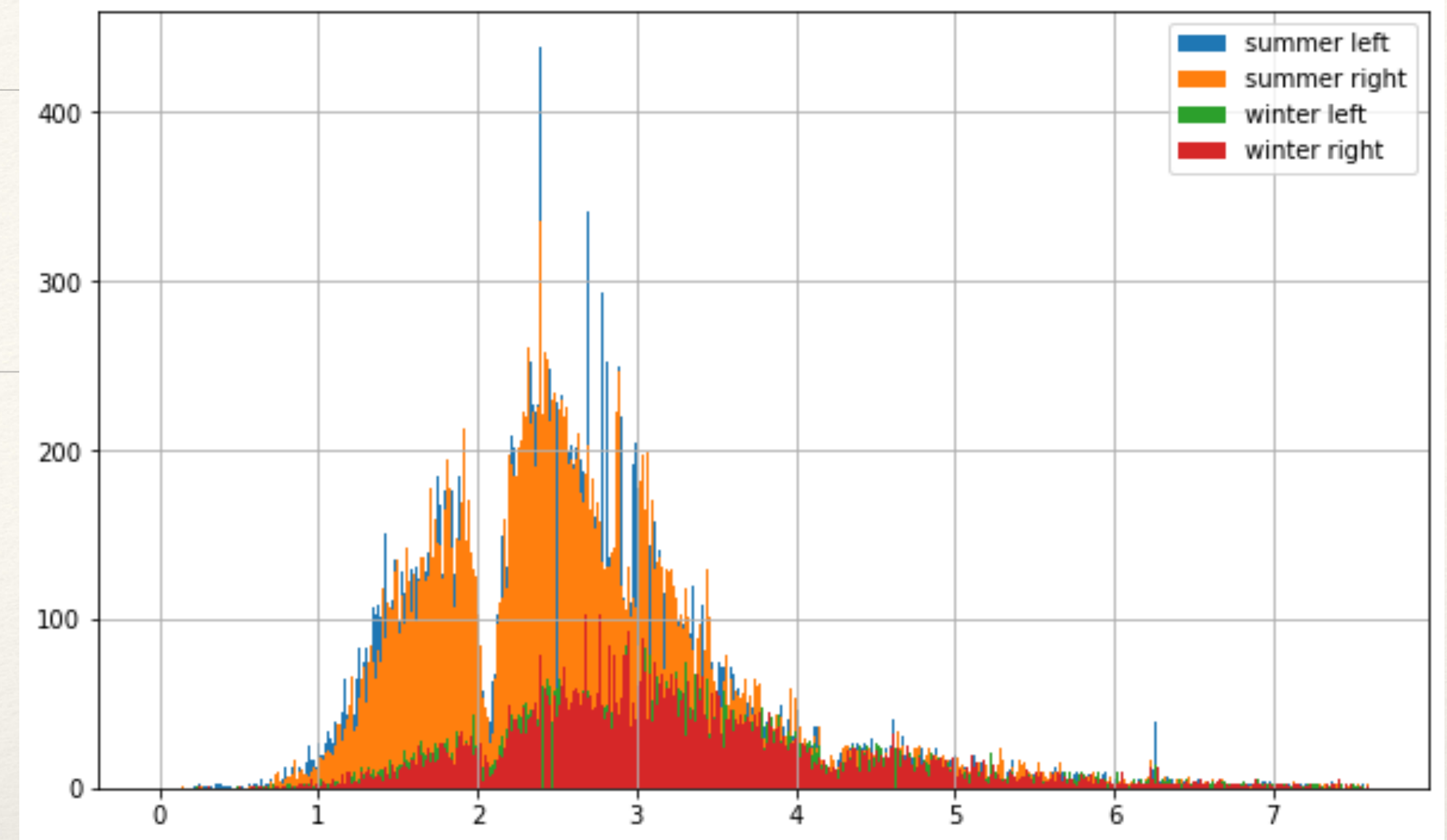
❖ `sns.kdeplot(df['DATA'],kernel='gau',bw_a  
djust=0.5)`





# 資料分佈

- ❖ 因資料嚴重向左偏斜，因此我們透過 `np.log1p()` 去除偏態，結果如右圖。
- ❖  $\text{np.log1p}(x) = \ln(x+1)$
- ❖





---

# 機率

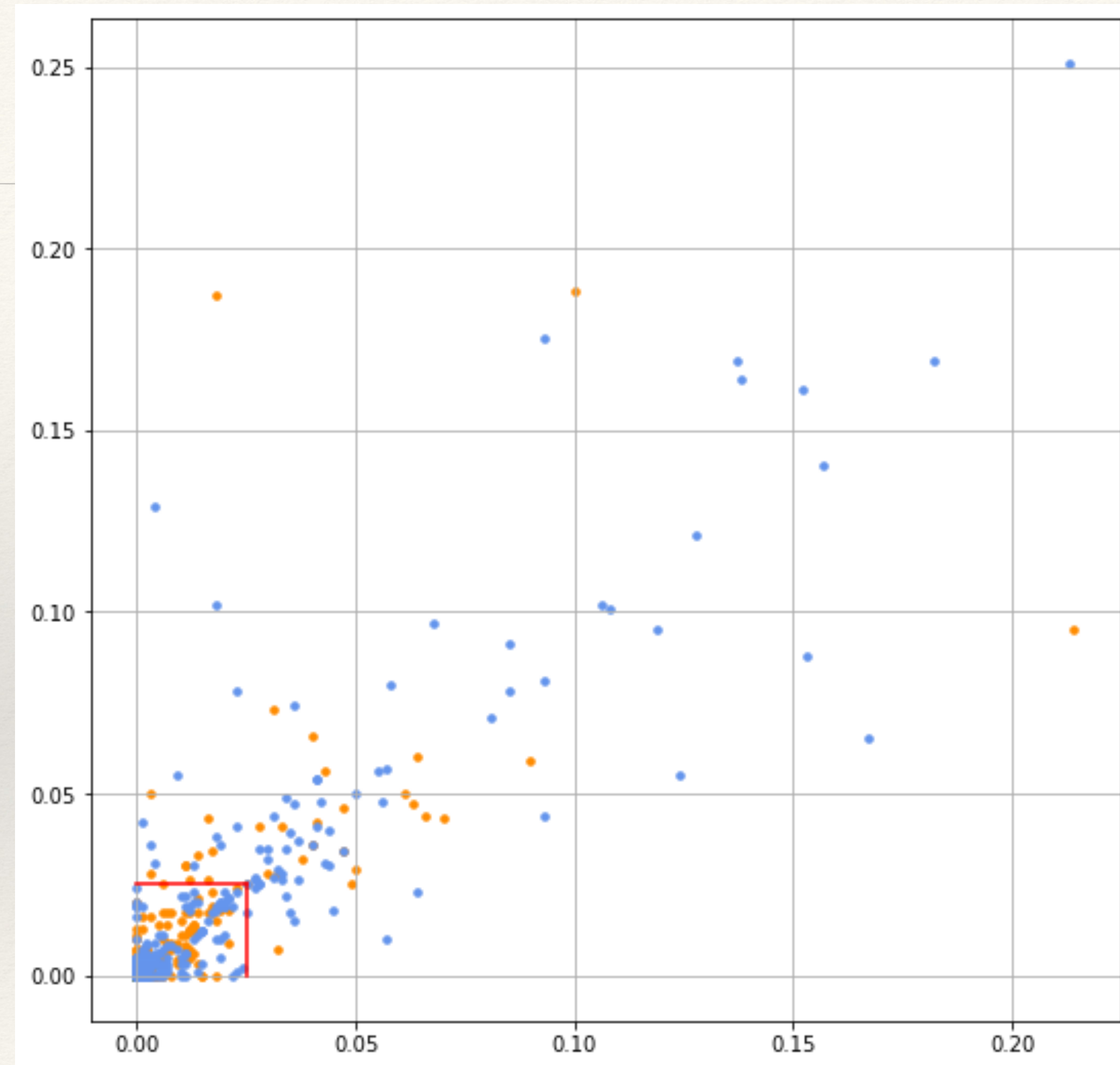
---

- ❖ 透過以下式子，可得每個人的於KDE落於區間內的機率
- ❖ `scipy.stats.gaussian_kde(SR).integrate_box_1d(low=0.69314,high=6.908)`
- ❖ 合格區間：1~1000
  - ❖  $\ln(1+1)=0.69314$
  - ❖  $\ln(1000+1)=6.908$



# 機率散布圖

- ❖ X 軸：右腳
- ❖ Y 軸：左腳
- ❖ 落於紅線以外皆判定為異常機率過高人員
- ❖ `plt.scatter(df_prob['P_SR'],df_prob['P_SL'],color='darkorange',s=12)`
- ❖ `plt.scatter(df_prob['P_WR'],df_prob['P_WL'],color='cornflower blue',s=12)`





---

# 未合格人員

---

- ❖ 冬天右腳、冬天左腳、夏天右腳、夏天左腳  
任一異常機率大於0.025即判定為不合格。
- ❖  $\text{df\_probFail} = \text{df\_prob}[(\text{df\_prob}['P\_SR'] > 0.025) | (\text{df\_prob}['P\_WR'] > 0.025) | (\text{df\_prob}['P\_SL'] > 0.025) | (\text{df\_prob}['P\_WL'] > 0.025)]$
- ❖ 不合格人員為86人，比率為：24.57%



