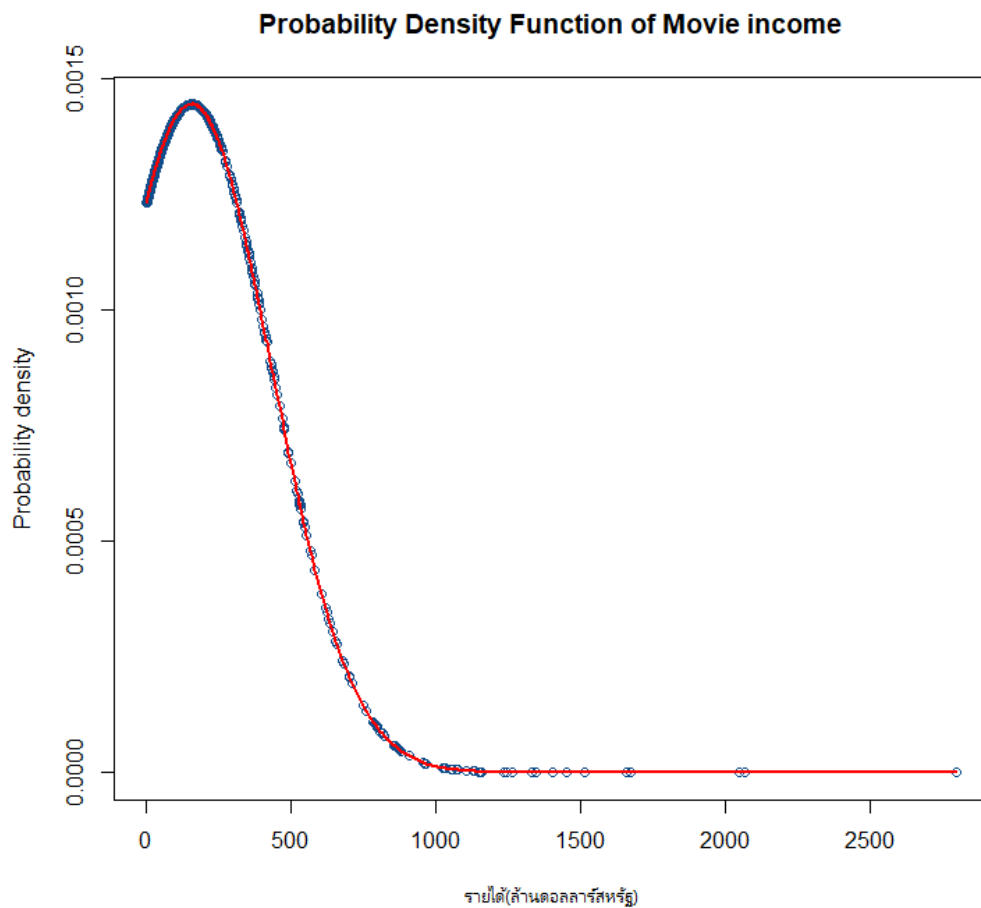


## HW3 : Probability Density Function/Cumulative Prob Function

### IMDb Movies

#### Probability Density Function

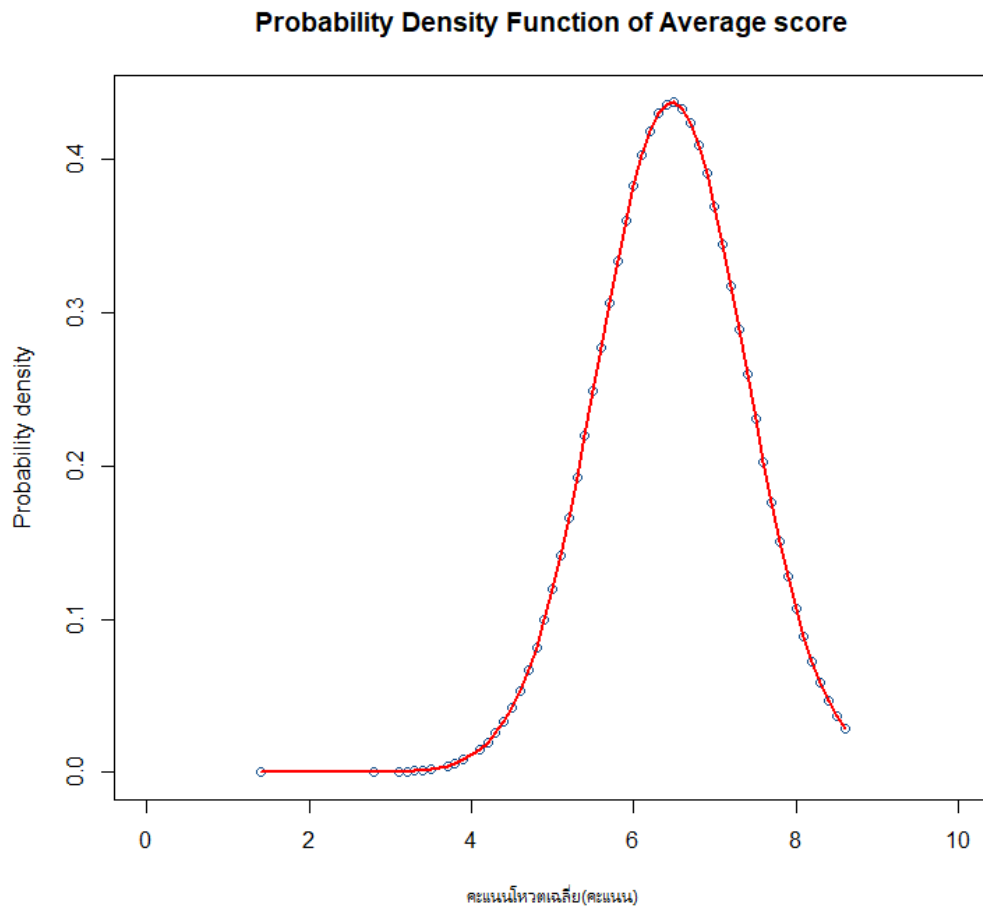
-รายได้ภาพยนตร์ทั่วโลก



```
pdfIncome = dnorm(x=income,mean=mean(income),sd = sd(income))
plot(income,pdfIncome,col = "dodgerblue4",main = "Probability Density Function of Movie income",ylab="Probability density",
      xlab = "รายได้(ล้านดอลลาร์สหรัฐ)")
lines(smooth.spline(income,pdfIncome), col='red',lwd=2)
```

แกน x เป็นรายได้(ล้านดอลลาร์สหรัฐ) แกน y เป็น ค่าความหนาแน่นที่สอดคล้องกับ mean และ sd ของรายได้(ล้านดอลลาร์สหรัฐ)

-คะแนนโหวตเฉลี่ย

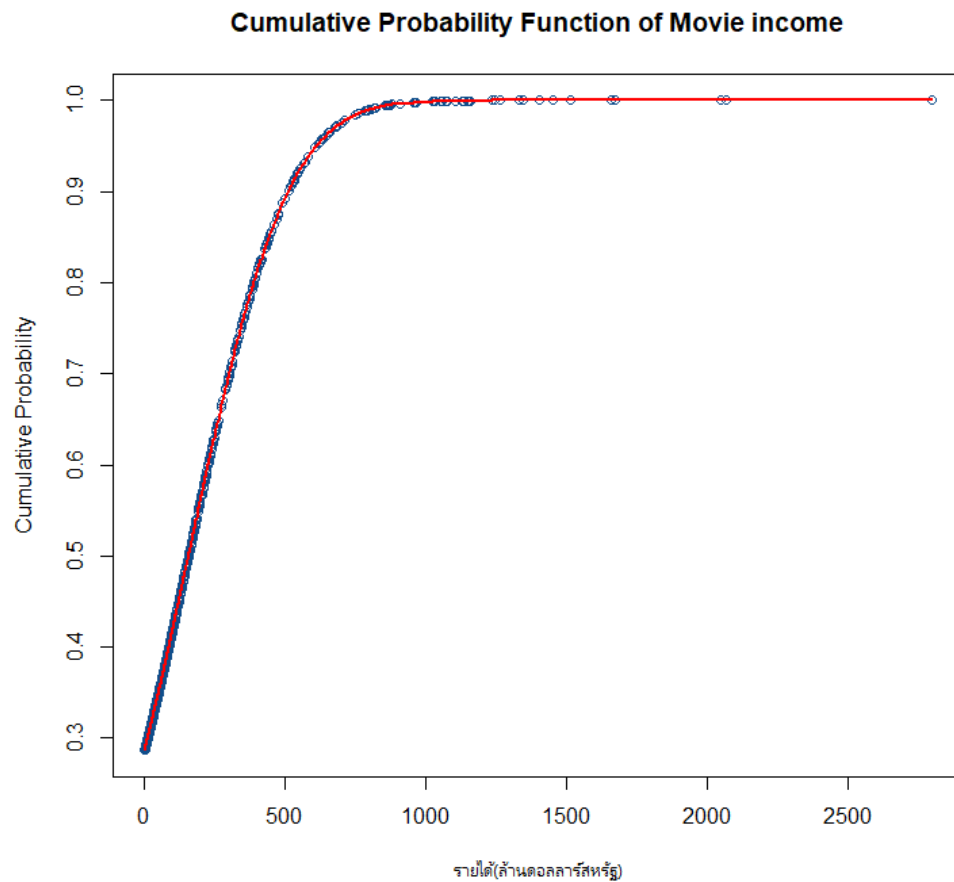


```
pdfAvgvote = dnorm(x=avg_vote,mean = mean(avg_vote),sd=sd(avg_vote))
plot(avg_vote,pdfAvgvote,xlim = c(0,10),col = "dodgerblue4",main = "Probability Density Function of Average score",ylab="Probability density",
      xlab = "คะแนนโหวตเฉลี่ย(คะแนน)")
lines(smooth.spline(avg_vote,pdfAvgvote), col='red',lwd=2)
```

แกน x เป็นคะแนนโหวตเฉลี่ย(คะแนน) แกน y เป็น ค่าความหนาแน่นที่สอดคล้องกับ mean และ sd ของคะแนนโหวตเฉลี่ย(คะแนน)

## Cumulative Probability Function

-รายได้ภาพยนตร์รวมทั่วโลก



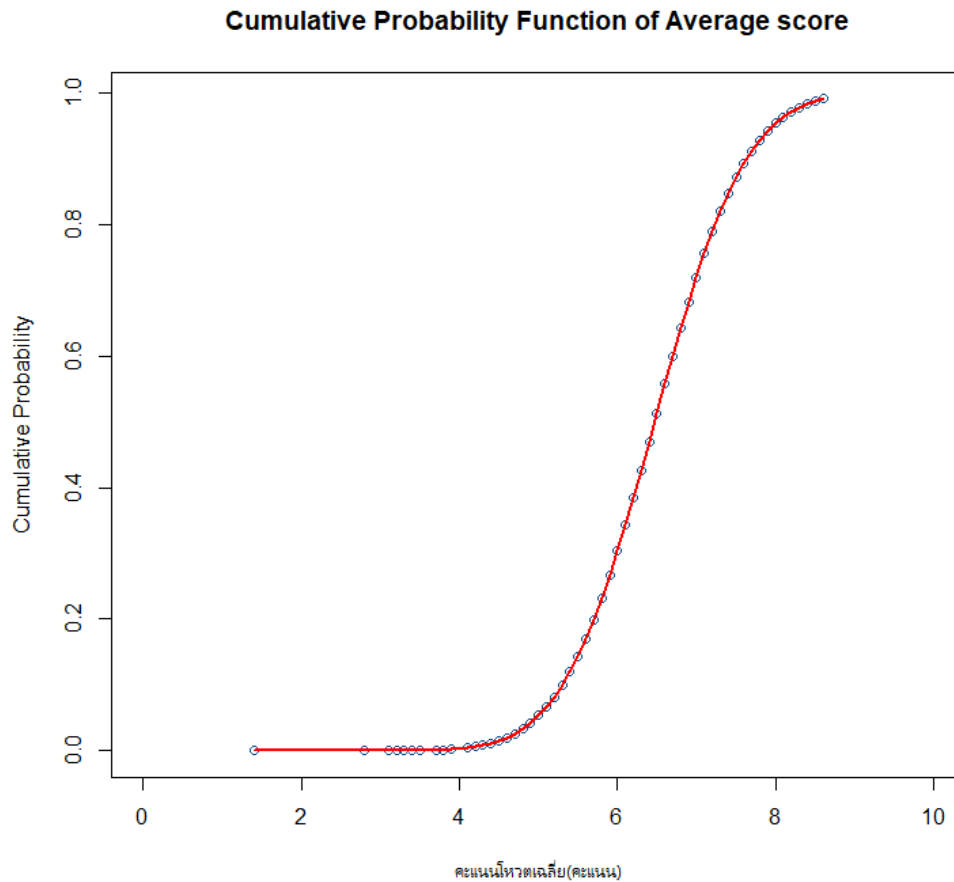
```

cpfIncome = pnorm(q=income,mean = mean(income),sd=sd(income))
plot(income,cpfIncome,col = "dodgerblue4",main = "Cumulative Probability Function of Movie income",ylab="Cumulative Probability",
      xlab = "รายได้(ล้านดอลลาร์สหรัฐ)")
lines(smooth.spline(income,cpfIncome), col='red',lwd=2)

```

แกน x เป็นรายได้(ล้านดอลลาร์สหรัฐ) แกน y เป็น ค่าสะสมที่สอดคล้องกับ mean และ sd ของรายได้(ล้านดอลลาร์สหรัฐ)

-คะแนนโหวตเฉลี่ย



```

cpfAvgVote = pnorm(q=avg_vote,mean = mean(avg_vote),sd=sd(avg_vote))
plot(avg_vote,cpfAvgVote,xlim = c(0,10),col = "dodgerblue4",main = "Cumulative Probability Function of Average score",ylab="Cumulative Probability",
      xlab = "คะแนนโหวตเฉลี่ย(คะแนน)")
lines(smooth.spline(avg_vote,cpfAvgVote), col='red',lwd=2)

```

แกน x เป็นคะแนนโหวตเฉลี่ย(คะแนน) แกน y เป็น ค่าสะสมที่สอดคล้องกับ mean และ sd ของคะแนนโหวตเฉลี่ย (คะแนน)

## บทวิเคราะห์ข้อมูลจากกราฟ

### Probability Density Function

- จากกราฟรายได้ของภาพยนตร์ จะเห็นได้ว่าในช่วงรายได้ประมาณ 100-400 ล้านดอลลาร์สหรัฐจะมีค่าความหนาแน่นเยอะที่สุด และในช่วงรายได้ประมาณ 400 ล้านดอลลาร์สหรัฐเป็นต้นไปก็จะค่อยๆลดลงเรื่อยๆ วิเคราะห์ได้ว่า รายได้ของภาพยนตร์ส่วนใหญ่จะอยู่ในช่วงประมาณ 100-400 ล้านดอลลาร์สหรัฐ

- จากกราฟคะแนนโหวตเฉลี่ย จะเห็นได้ว่าในช่วงคะแนนโหวตเฉลี่ย 0-4 คะแนนค่าความหนาแน่นเพิ่มขึ้นน้อยมาก และเพิ่มขึ้นอย่างรวดเร็วในช่วงคะแนนโหวตเฉลี่ย 4-6 คะแนน และในช่วงคะแนนโหวตเฉลี่ย 6-7 คะแนนจะมีค่าความหนาแน่นจะเยอะที่สุด และในช่วงคะแนนโหวตเฉลี่ย 7 คะแนนเป็นต้นไปค่าความหนาแน่นก็จะลดลงเรื่อยๆ วิเคราะห์ได้ว่า คะแนนโหวตเฉลี่ยของภาพยนตร์ส่วนใหญ่จะอยู่ในช่วง 6-7 คะแนน

### Cumulative Probability Function

- จากกราฟรายได้ของภาพยนตร์ จะเห็นได้ว่าในช่วงรายได้ 0-500 ล้านดอลลาร์สหรัฐค่าสะสมจะเพิ่มขึ้นอย่างรวดเร็ว ในช่วง 500-1000 ล้านดอลลาร์สหรัฐค่าสะสมก็จะเพิ่มขึ้นช้าลง และในช่วง 1000 ล้านดอลลาร์สหรัฐเป็นต้นไปค่าสะสมจะเพิ่มขึ้นน้อยมาก วิเคราะห์ได้ว่า รายได้ของภาพยนตร์ส่วนใหญ่จะอยู่ในช่วง 0-500 ล้านดอลลาร์สหรัฐ

- จากกราฟคะแนนโหวตเฉลี่ยจะเห็นได้ว่าในช่วงคะแนนโหวตเฉลี่ย 0-5 คะแนนค่าสะสมที่เพิ่มขึ้นน้อยมาก แต่ในช่วงคะแนนโหวตเฉลี่ย 5-8 คะแนนค่าสะสมเพิ่มขึ้นอย่างรวดเร็ว และในช่วงคะแนนโหวตเฉลี่ย 8 คะแนนขึ้นไปก็จะค่อยๆเพิ่มขึ้นช้าลง วิเคราะห์ได้คะแนนโหวตเฉลี่ยส่วนใหญ่จะอยู่ในช่วง 5-8 คะแนน

## Source Code

```

setwd("~/CE2D-2/git/Propstat")

df <- read.csv("imdbm.csv")

income <- df$income
avg_vote <- df$avg_vote

pdfIncome = dnorm(x=income,mean=mean(income),sd = sd(income))

plot(income,pdfIncome,col = "dodgerblue4" ,main = "Probability Density Function of Movie income",ylab="Probability density"
,xlab = "รายได้(ล้านดอลลาร์สหรัฐ)")
lines(smooth.spline(income,pdfIncome), col='red',lwd=2)

pdfAvgvote = dnorm(x=avg_vote,mean = mean(avg_vote),sd=sd(avg_vote))

plot(avg_vote,pdfAvgvote,xlim = c(0,10),col = "dodgerblue4" ,main = "Probability Density Function of Average score",ylab="Probability density"
,xlab = "คะแนนโหวตเฉลี่ย(คะแนน)")
lines(smooth.spline(avg_vote,pdfAvgvote), col='red',lwd=2)

cpfIncome = pnorm(q=income,mean = mean(income),sd=sd(income))

plot(income,cpfIncome,col = "dodgerblue4" ,main = "Cumulative Probability Function of Movie income",ylab="Cumulative Probability"
,xlab = "รายได้(ล้านดอลลาร์สหรัฐ)")
lines(smooth.spline(income,cpfIncome), col='red',lwd=2)

cpfAvgvote = pnorm(q=avg_vote,mean = mean(avg_vote),sd=sd(avg_vote))

plot(avg_vote,cpfAvgvote,xlim = c(0,10),col = "dodgerblue4" ,main = "Cumulative Probability Function of Average score",ylab="Cumulative Probability"
,xlab = "คะแนนโหวตเฉลี่ย(คะแนน)")
lines(smooth.spline(avg_vote,cpfAvgvote), col='red',lwd=2)

```