

01076253 PROBABILITY AND STATISTICS HW 4

ในการหา Confidence Interval (CI) ทางผู้จัดทำได้เลือก คอลัมน์ Average ratings on parks

```
> sample.mean <- mean(Google_Review_Ratings_.$'Average ratings on parks')
> print(sample.mean)
[1] 2.958941
```

```
> sample.n <- length(Google_Review_Ratings_.$'Average ratings on parks')
> sample.sd <- sd(Google_Review_Ratings_.$'Average ratings on parks')
> sample.se <- sample.sd/ sqrt(sample.n)
> print(sample.se)
[1] 0.01812849
```

sample.mean	2.95894061583578
sample.n	5456L
sample.sd	1.33905646061777
sample.se	0.0181284932154164

จากรูปจะเห็นค่าของ Mean, Sample Size, Standard Deviation และ Standard error

Formula >

$$CI = \bar{x} \pm z \frac{s}{\sqrt{n}}$$

CI = confidence interval

\bar{x} = sample mean

z = confidence level value

s = sample standard deviation

n = sample size

นี่คือสูตรในการใช้คำนวณหาค่า Confidence Interval โดยในที่นี้จะใช้จากเว็บไซต์

[Easy Confidence Interval Calculator \(socscistatistics.com\)](https://www.socscistatistics.com/easy-confidence-interval-calculator/)

ค่า Confidence Interval ที่ Confidence Level 90%

$$M = 2.958941$$

$$Z = 1.64$$

$$S_M = \sqrt{(1.3390562/5456)} = 0.02$$

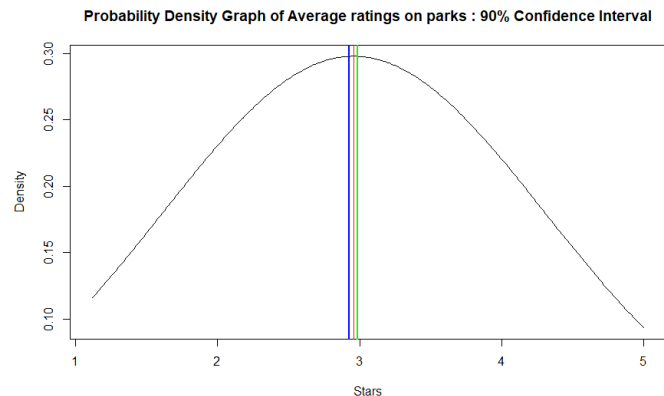
$$\mu = M \pm Z(S_M)$$

$$\mu = 2.958941 \pm 1.64 \cdot 0.02$$

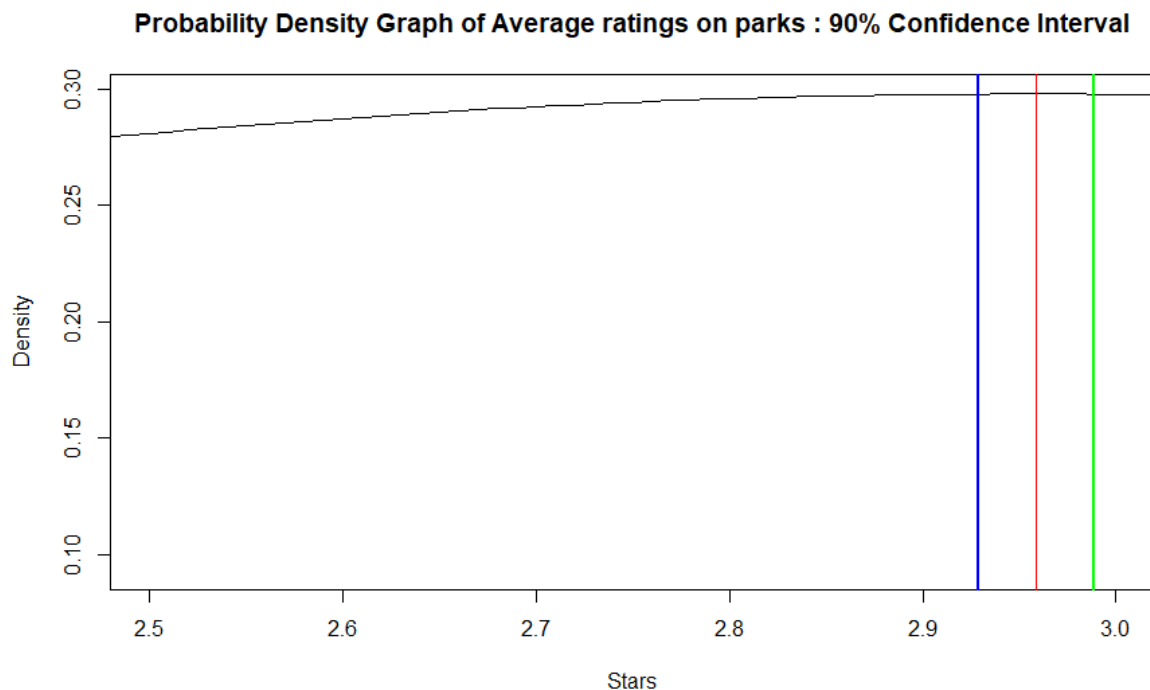
$$\mu = 2.958941 \pm 0.02981871$$

ค่า Confidence Interval ของคะแนนการรีวิว คอลัมน์ Average ratings on parks

ที่ Confidence Level 90% คือระหว่างช่วง [2.92912229, 2.98875971] คะแนน



จากกราฟจะเห็นได้ว่าช่วง Confidence Level 90% คือระหว่างช่วง
[2.92912229, 2.98875971] คะแนน



สูตรที่ใช้หา Confidence Level 90%

```
> plot(sort(Google_Review_Ratings_.$'Average ratings on parks'),dnorm(sort(Google_Review_Ratings_.$'Average ratings on parks'),sample.mean,sample.sd),type = 'l',main = "Probability Density Graph of Average ratings on parks : 90% Confidence Interval",xlab="Stars",ylab="Density", xlim = c(2.5,3))
> abline(v=sample.mean,col="red")
> abline(v=2.92912229,col="blue",lwd=2)
> abline(v=2.98875971,col="green",lwd=2)
```

ค่า Confidence Interval ที่ Confidence Level 95%

$$M = 2.958941$$

$$Z = 1.96$$

$$S_M = \sqrt{(1.339056^2/5456)} = 0.02$$

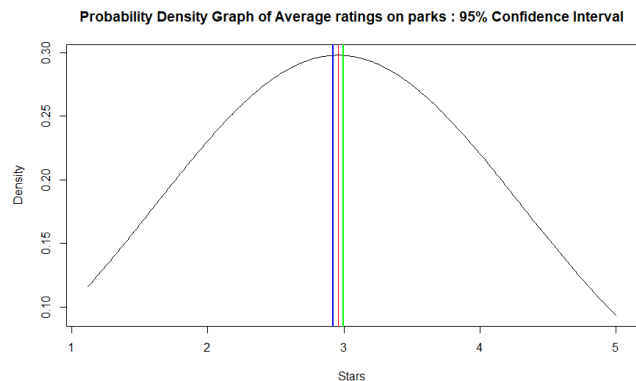
$$\mu = M \pm Z(S_M)$$

$$\mu = 2.958941 \pm 1.96 * 0.02$$

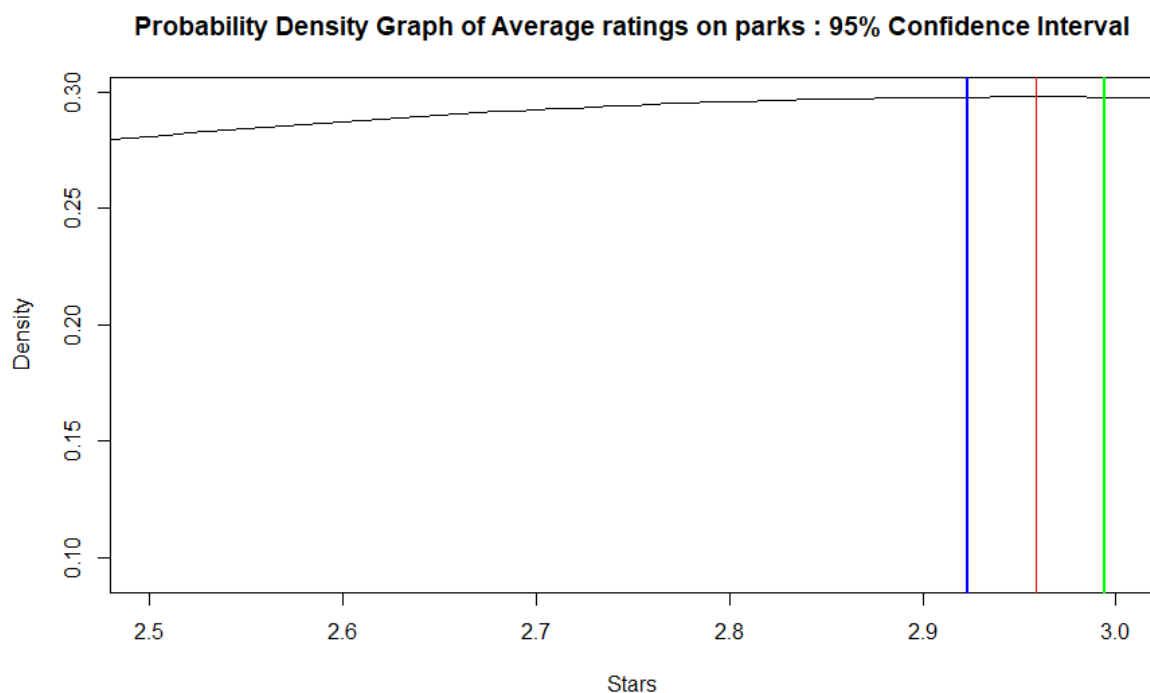
$$\mu = 2.958941 \pm 0.03553118$$

ค่า Confidence Interval ของคะแนนการรีวิว คอลัมน์ Average ratings on parks

ที่ Confidence Level 95% คือระหว่างช่วง [2.92340982, 2.99447218] คะแนน



จากกราฟที่ขยายจะเห็นได้ว่าช่วง *Confidence Level 95%* คือระหว่างช่วง
[2.92340982, 2.99447218] คะแนน



สูตรที่ใช้หา *Confidence Level 95%*

```
> plot(sort(Google_Review_Ratings_.$'Average ratings on parks'),dnorm(sort(Google_Review_Ratings_.$'Average ratings on park
s'),sample.mean,sample.sd),type = 'l',main = "Probability Density Graph of Average ratings on parks : 95% Confidence Inte
rval",xlab="stars",ylab="Density", xlim = c(2.5,3))
> abline(v=sample.mean,col="red")
> abline(v=2.92340982,col="blue",lwd=2)
> abline(v=2.99447218,col="green",lwd=2)
```

ค่า Confidence Interval ที่ Confidence Level 99%

$$M = 2.958941$$

$$Z = 2.58$$

$$S_M = \sqrt{(1.3390562/5456)} = 0.02$$

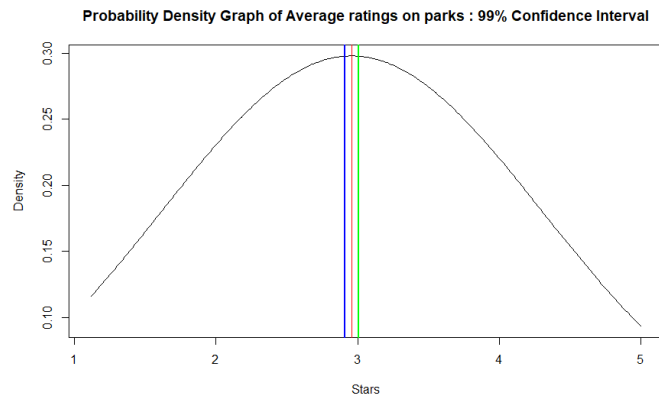
$$\mu = M \pm Z(S_M)$$

$$\mu = 2.958941 \pm 2.58 \cdot 0.02$$

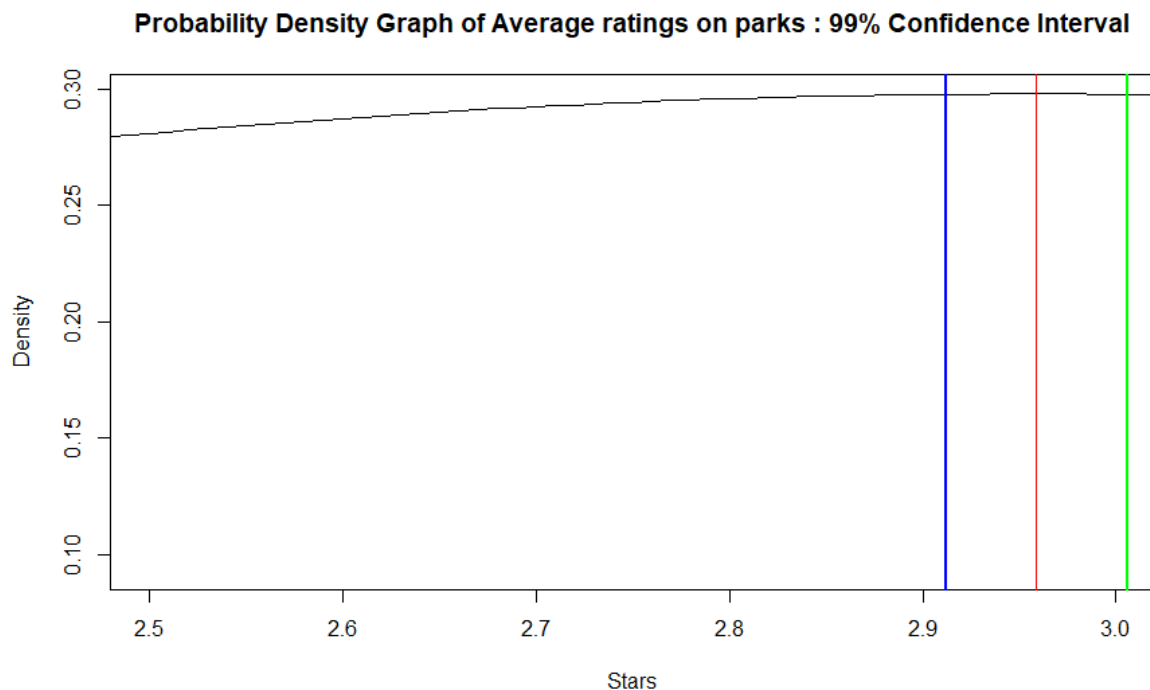
$$\mu = 2.958941 \pm 0.04669589$$

ค่า Confidence Interval ของคะแนนการรีวิว คอลัมน์ Average ratings on parks

ที่ Confidence Level 99% คือระหว่างช่วง [2.91224511, 3.00563689] คะแนน



จากกราฟจะเห็นได้ว่าช่วง Confidence Level 99% คือระหว่างช่วง
[2.91224511, 3.00563689] คะแนน



สูตรที่ใช้หา Confidence Level 99%

```
> plot(sort(Google_Review_Ratings_.$'Average ratings on parks'),dnorm(sort(Google_Review_Ratings_.$'Average ratings on park
s'),sample.mean,sample.sd),type = 'l',main = "Probability Density Graph of Average ratings on parks : 99% Confidence Inte
rval",xlab="Stars",ylab="Density", xlim = c(2.5,3))
> abline(v=sample.mean,col="red")
> abline(v=2.91224511,col="blue",lwd=2)
> abline(v=3.00563689,col="green",lwd=2)
```

บทวิเคราะห์ข้อมูล

จากการทำการหาค่า Confidence Interval ได้ผลลัพธ์ดังนี้

ค่า Confidence Interval ของคะแนนการรีวิว คอลัมน์ Average ratings on parks
ที่ Confidence Level 90% คือระหว่างช่วง [2.92912229, 2.98875971] คะแนน

ค่า Confidence Interval ของคะแนนการรีวิว คอลัมน์ Average ratings on parks
ที่ Confidence Level 95% คือระหว่างช่วง [2.92340982, 2.99447218] คะแนน

ค่า Confidence Interval ของคะแนนการรีวิว คอลัมน์ Average ratings on parks
ที่ Confidence Level 99% คือระหว่างช่วง [2.91224511, 3.00563689] คะแนน

เราสามารถวิเคราะห์ข้อมูลจากค่า Confidence Interval ได้ว่า

คะแนนการรีวิว คอลัมน์ Average ratings on parks ของ User กว่า 90% นั้นมีการให้คะแนนเฉลี่ย
อยู่คือระหว่างช่วง [2.92912229, 2.98875971] คะแนน

คะแนนการรีวิว คอลัมน์ Average ratings on parks ของ User กว่า 90% นั้นมีการให้คะแนนเฉลี่ย
อยู่คือระหว่างช่วง [2.92340982, 2.99447218] คะแนน

คะแนนการรีวิว คอลัมน์ Average ratings on parks ของ User กว่า 90% นั้นมีการให้คะแนนเฉลี่ย
อยู่คือระหว่างช่วง [2.91224511, 3.00563689] คะแนน

และสรุปจากคะแนนทั้งหมดได้ว่า

ค่าคะแนนการรีวิวเกี่ยวกับ Travel Review Rating Dataset (From the Machine Learning Repository of University of California) นั้นมีข้อสังเกต และข้อสงสัยที่ว่าคะแนนที่ได้มานั้น ค่าเฉลี่ยไป
กองรวมกันที่ 2.9 โดยประมาณ และในการหาค่าคะแนนสังเกตว่า มีการให้คะแนนเต็ม (5 ดาว) กันหลาย
User ซึ่งบ่งบอกได้หลายกรณีเช่น อาจมีการรับจ้างรีวิว, อาจจะมีการใช้บอททำหน้าที่ให้คะแนน, อาจจะมี
โดยให้คะแนนสูง, อาจจะมีรีวิวเพื่อแค่แลกของรางวัลที่ทางสถานที่นั้น ๆ จัดขึ้น, หรืออาจจะรีวิวจากการได้รับ
ส่วนลดค่าที่พัก เป็นต้น ซึ่งข้อสันนิษฐานของผู้จัดทำข้างต้นนี้มีความเป็นไปได้สูงที่จะสามารถเกิดขึ้น เพราะ
ด้วยสังคมโลกยุคปัจจุบันการเอาตัวรอดในสถานการณ์ต่าง ๆ เป็นสิ่งสำคัญ สถานที่หรือผู้ประกอบการบริษัท
นั้น ๆ สามารถล่อลวงผู้บริโภคเพื่อประโยชน์ส่วนตนได้ ดังเช่นตัวอย่างจากข่าวนี้ [ร้อง สคบ. เอาผิดเว็บจองที่พัก
ชื่อดัง หลอกลวงผู้บริโภค | ข่าวช่อง 8 \(thaich8.com\)](#) ซึ่งทางผู้จัดทำหวังเป็นอย่างยิ่งว่าผู้ชมที่เข้ามาศึกษางานนี้
จะมีวิธีรับมือ และไม่หลงเชื่ออะไรง่าย ๆ ครับ