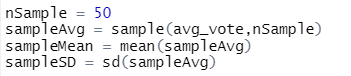
**HW4# - Confidence Interval (CI) of Mean**

**IMDb Movies**

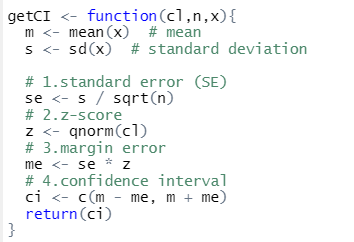
หา Confidence Interval (CI) ของคอลัมน์ คะแนนโหวตเฉลี่ย IMDb Movies

สมมติว่าภาพยนตร์ทั้งหมดมี 1000 เรื่อง มีค่า population mean (mu) ของคะแนนโหวตเฉลี่ย = 6.4693 คะแนน



ทำการสุ่มภาพยนตร์ตัวอย่างมา 50 เรื่อง จะได้ sample mean = 6.552 คะแนน และ sd = 0.8179716 คะแนน

-หา Confidence Interval (CI) ของแต่ละ Confidence Level



Confidence Level = 90%

-90% confidence interval = [6.403752, 6.700248] คะแนน

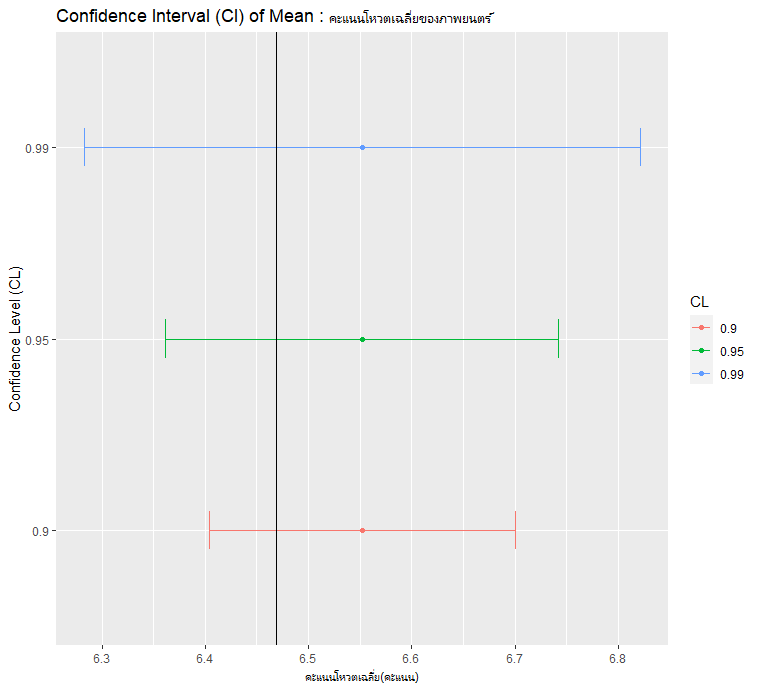
Confidence Level = 95%

-95% confidence interval = [6.361726, 6.742274] คะแนน

Confidence Level = 99%

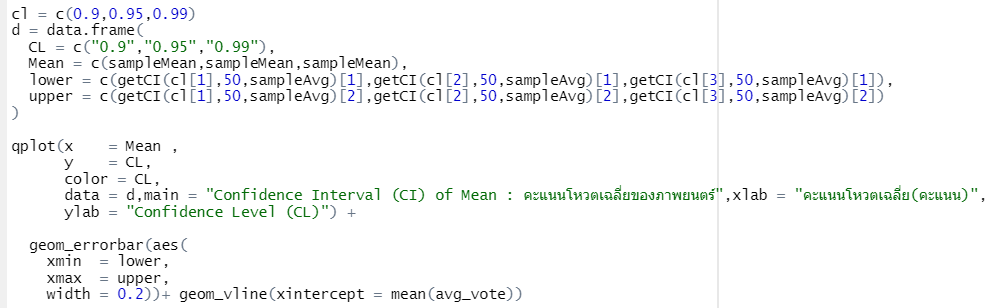
-99% confidence interval = [6.282891, 6.821109] คะแนน

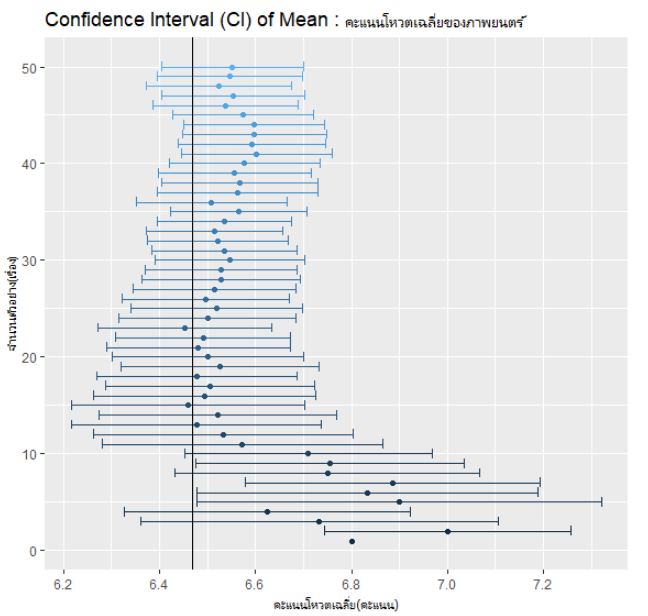
**กราฟ Confidence Interval (CI) of Mean**



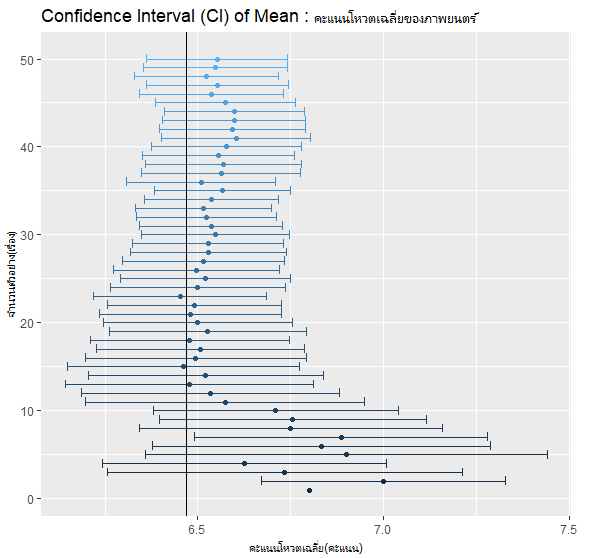
**Population Mean =** 6.4693 คะแนน

**รูปที่ 1**

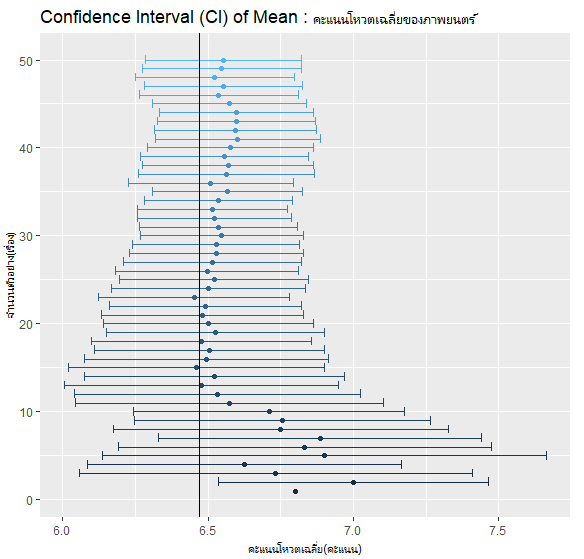


Confidence Interval จาก Confidence Level = 90% ของการสุ่มตัวอย่างตัวที่ 1 – 50

**รูปที่ 2**

Confidence Interval จาก Confidence Level = 95% ของการสุ่มตัวอย่างตัวที่ 1 – 50

**รูปที่ 3**

Confidence Interval จาก Confidence Level = 95% ของการสุ่มตัวอย่างตัวที่ 1 – 50

**รูปที่ 4**

**บทวิเคราะห์ข้อมูลจากกราฟ**

วิเคราะห์ข้อมูลจากกราฟรูปที่ 1

จากกราฟ Confidence Interval (CI) of Mean ของคะแนนโหวตเฉลี่ยของภาพยนตร์ ซึ่งทำการสุ่มภาพยนตร์มาจำนวน 50 เรื่อง มี sample mean = 6.552 คะแนน และ sd = 0.8179716 คะแนน

จาก confidence interval ที่สร้างขึ้นมา ค่า population mean = 6.4693 คะแนน จะอยู่ในช่วง confidence interval ที่สร้างขึ้นมาทั้ง 3 อัน

ถ้าค่า Confidence Level เยอะกว่าจะทำให้ confidence interval กว้างกว่า Confidence Level ที่มีค่าน้อยกว่า

วิเคราะห์ข้อมูลจากกราฟรูปที่ 2-4

ทุกครั้งที่เราสุ่มตัวอย่างใหม่ ค่าสถิติทั้งหมดไม่ว่าจะเป็นค่า mean, sd รวมถึง confidence interval ก็จะเปลี่ยนไปเรื่อย ๆ แต่ถ้าเราสุ่มซ้ำหลายๆครั้ง เช่น ทำซ้ำ 50 ครั้งและทำทุกอย่างเหมือนเดิม

จาก Confidence Level = 90% มี 45 ครั้ง ใน 50 ครั้งที่ ค่า population mean อยู่ในช่วง

confidence interval ที่สร้างขึ้นมา หรือคิดเป็น 90 % และมี 5 % ที่ค่า population mean ไม่ได้อยู่ในช่วง confidence interval

จาก Confidence Level = 95% มี 48 ครั้ง ใน 50 ครั้งที่ ค่า population mean อยู่ในช่วง

confidence interval ที่สร้างขึ้นมา หรือคิดเป็น 96 % และมี 4 % ที่ค่า population mean ไม่ได้อยู่ในช่วง confidence interval

จาก Confidence Level = 99% มี 49 ครั้ง ใน 50 ครั้งที่ ค่า population mean อยู่ในช่วง

confidence interval ที่สร้างขึ้นมา หรือคิดเป็น 98 % และมี 2 % ที่ค่า population mean ไม่ได้อยู่ในช่วง Confidence Level

สามารถวิเคราะห์ได้ว่า ค่า Confidence Level = x % หมายถึง มีโอกาส x % โดยประมาณที่ confidence interval ที่สร้างขึ้นมาจะครอบคลุมค่า population mean

**Source Code**

