# โครงการ

# การเรียนการสอนเพื่อเสริมประสบการณ์

ชื่อโครงการ การประมาณค่าของพารามิเตอร์ที่บ่งชี้ระยะห่างจากกฎของเบนฟอร์ด
Estimation of parameters reflecting distance from Benford's law


ชื่อนิสิต นางสาวกมลรัตน์ ผลคิมหันต์ 5833501723

ภาควิชา คณิตศาสตร์และวิทยาการคอมพิวเตอร์
สาขาวิชา คณิตศาสตร์


ปีการศึกษา 2561


คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย



บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของโครงงานทางวิชาการที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของโครงงานทางวิชาการที่ส่งผ่านทางคณะที่สังกัด
The abstract and full text of senior projects in Chulalongkorn University Intellectual Repository(CUIR)
are the senior project authors' files submitted through the faculty.

# การประมาณค่าของพารามิเตอร์ที่บ่งชี้ระยะห่างจากกฎของเบนฟอร์ด

นางสาวกมลรัตน์ ผลคิมหันต์

โครงงานนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต
สาขาวิชาคณิตศาสตร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2561
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

# Estimation of parameters reflecting distance from Benford's law

Miss Kamonrat Phonkimhan

A Project Submitted in Partial Fulfillment of the Requirements

for the Degree of Bachelor of Science Program in Mathematics

Department of Mathematics and Computer Science

Faculty of Science

Chulalongkorn University

Academic Year 2018

Copyright of Chulalongkorn University

| | |
|---|---|
| หัวข้อโครงงาน | การประมาณค่าของพารามิเตอร์ที่บ่งชี้ระยะห่างจากกฎของเบนฟอร์ด |
| โดย | นางสาวกมลรัตน์ ผลคิมหันต์ เลขประจำตัว 5833501723 |
| สาขาวิชา | คณิตศาสตร์ |
| อาจารย์ที่ปรึกษาโครงงานหลัก | รองศาสตราจารย์ ทิพวัลย์ สันติวิภานนท์ |
| อาจารย์ที่ปรึกษาโครงงานร่วม | รองศาสตราจารย์ ดร.ทรงเกียรติ สุเมธกิจการ |

---

ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับโครงงานฉบับนี้เป็นส่วนหนึ่ง ของการศึกษาตามหลักสูตรปริญญาบัณฑิต ในรายวิชา 2301499 โครงงานวิทยาศาสตร์ (Senior Project)

..... หัวหน้าภาควิชาคณิตศาสตร์
และวิทยาการคอมพิวเตอร์

(ศาสตราจารย์ ดร.กฤษณะ เนียมมณี)

คณะกรรมการสอบโครงงาน

..... อาจารย์ที่ปรึกษาโครงงานหลัก

(รองศาสตราจารย์ ทิพวัลย์ สันติวิภานนท์ )

..... อาจารย์ที่ปรึกษาโครงงานร่วม

(รองศาสตราจารย์ ดร. ทรงเกียรติ สุเมธกิจการ )

..... กรรมการ

(รองศาสตราจารย์ยุวรีย์ พันธ์กล้า)

..... กรรมการ

(รองศาสตราจารย์ ดร. พิเชฐ ชาวหา)

กมลรัตน์ ผลคิมหันต์: การประมาณค่าของพารามิเตอร์ที่บ่งชี้ระยะห่างจากกฎของเบนฟอร์ด (ESTIMATION OF PARAMETERS REFLECTING DISTANCE FROM BENFORD'S LAW) อ.ที่ปรึกษาโครงงานหลัก: รองศาสตราจารย์ ทิพวัลย์ สันติวิภานนท์, อ.ที่ปรึกษาโครงงานร่วม : รองศาสตราจารย์ ดร. ทรงเกียรติ สุเมธกิจการ, 64 หน้า

โครงงานนี้เป็นการทดสอบความเป็นเบนฟอร์ดของข้อมูล โดยการสังเกตระยะห่างระหว่าง 1 กับตัวประมาณของ $\alpha$ ที่ได้จากการประมาณด้วยวิธีความควรจะเป็นสูงสุด พร้อมทั้งพิจารณาช่วงของจำนวนนับ $q$ ที่เหมาะสมในการนำไปทดสอบความเป็นเบนฟอร์ดของข้อมูล ซึ่งทั้งสองพารามิเตอร์นี้ เป็นพารามิเตอร์ของฟังก์ชันการแจกแจงของการแจกแจงที่เป็นลิมิตของกลุ่มการแจกแจงตัดปลาย ที่ศึกษาในงานของวิริยะไกรกุล, สุเมธกิจการ และสันติวิภานนท์

จากนั้นจะนำตัวประมาณของพารามิเตอร์ $\alpha$ และช่วงของจำนวนนับ $q$ ที่เหมาะสมไปใช้ในการทดสอบความเป็นเบนฟอร์ดของข้อมูลค่าใช้จ่ายไฟฟ้ารายเดือน เราพบว่าข้อมูลนี้มีการแจกแจงเลขหลักแรกค่อนข้างสอดคล้องกับกฎของเบนฟอร์ด เราหวังว่าโครงงานนี้จะนำเสนออีกทางเลือกหนึ่งของการทดสอบความเป็นเบนฟอร์ดของชุดข้อมูล

| | | |
|---|---|---|
| ภาควิชา | คณิตศาสตร์และวิทยาการคอมพิวเตอร์ | ลายมือชื่อนิสิต ..... |
| สาขาวิชา | คณิตศาสตร์ | ลายมือชื่อ อ.ที่ปรึกษาโครงงานหลัก ..... |
| ปีการศึกษา | 2561 | ลายมือชื่อ อ.ที่ปรึกษาโครงงานร่วม ..... |

## 5833501723 : MAJOR MATHEMATICS.
KEYWORDS: BENFORD'S LAW, ESTIMATION OF PARAMETER

KAMONRAT PHONKIMHAN: ESTIMATION OF PARAMETERS REFLECTING DISTANCE FROM BENFORD'S LAW. ADVISOR: ASSOC. PROF. TIPPAWAN SANTIWIPANONT , CO-ADVISOR : ASSOC. PROF. SONGKIAT SUMETKIJAKAN, 64 PP.

We test Benfordness of data by observing distance between 1 and the estimator of parameter $\alpha$ which is estimated by *maximum likelihood method*. Moreover, we also find a range of parameter $q \in N$ suitable for testing Benfordness of data. These two parameters are parameters of the limiting distributions in the study of truncated distributions by Wiriyakraikul, Sumetkijakan and Santiwipanont.

Next, the estimator of $\alpha$ and the suitable range of $q$ is used to test Benfordness of electrical monthly payment data. We find that the first digit distributions of these data seem to follow Benford's law. We hope that this project offers an alternative way to test the Benfordness of a set of data.

| | | |
|---|---|---|
| Department | Mathematics and Computer Science | Student's Signature Kamonrat |
| Field of Study | Mathematics | Advisor's Signature Tippawan S. |
| Academic | 2561 | Co-Advisor's Signature |

# กิตติกรรมประกาศ

ในการดำเนินงานครั้งนี้ได้รับความอนุเคราะห์ และความช่วยเหลือจากบุคคลหลายท่านด้วยกัน จึงขอขอบคุณไว้ ณ โอกาสนี้

ขอขอบพระคุณรองศาสตราจารย์ ทิพวัลย์ สันติวิภานนท์ ที่กรุณารับเป็นอาจารย์ที่ปรึกษาโครงงาน และรองศาสตราจารย์ ดร. ทรงเกียรติ สุเมธกิจการ ที่กรุณารับเป็นอาจารย์ที่ปรึกษาโครงงานร่วม พร้อมทั้งได้ให้ความรู้ คำปรึกษาแนะนำ สละเวลาเวลาคอยติดตามความก้าวหน้า ให้ความช่วยเหลือตลอดมาตั้งแต่เริ่มจนกระทั่งโครงงานสำเร็จตามจุดประสงค์ที่ตั้งไว้

ตลอดจนขอบพระคุณ รองศาสตราจารย์ยุวรีย์ พันธ์กล้า และรองศาสตราจารย์ ดร. พิเชฐ ชาวหา ที่กรุณาเป็นคณะกรรมการในการสอบโครงงาน และได้ให้คำปรึกษาและแก้ไขปรับปรุงโครงงานนี้ให้มีความถูกต้องและสมบูรณ์มากยิ่งขึ้น

นอกจากนี้ขอขอบพระคุณคณาจารย์ทุกท่านที่ได้ให้ความรู้และคำแนะนำ รวมทั้งอบรมข้าพเจ้าตลอดระยะเวลาที่เข้ามาศึกษาที่จุฬาลงกรณ์มหาวิทยาลัย

สุดท้ายนี้ขอขอบพระคุณครอบครัว รวมถึงเพื่อนๆทุกคน ที่คอยรับฟังและให้กำลังใจ รวมทั้งนายพีรพงษ์ ปัญญาสกุลวงศ์ ที่คอยช่วยเหลือและให้คำแนะนำต่างๆ มาตลอดการทำโครงงานครั้งนี้

# สารบัญ

หน้า

**บทคัดย่อ (ภาษาไทย)** **ง**

**บทคัดย่อ (ภาษาอังกฤษ)** **จ**

**กิตติกรรมประกาศ** **ฉ**

**สารบัญ** **ช**

**สารบัญรูป** **ญ**

**สารบัญตาราง** **ฎ**

**1 บทนำและความรู้พื้นฐาน** **1**

1.1 ที่มาและความสำคัญ . . . . . 1

1.2 ความรู้พื้นฐานทางคณิตศาสตร์ . . . . . 3

1.2.1 ตัวแปรสุ่ม . . . . . 3

1.2.2 วิธีของความควรจะเป็นสูงสุด . . . . . 5

1.2.3 แมนทิสซาและกฎของเบนฟอร์ด . . . . . 7

**2 ฟังก์ชันการแจกแจงของเลขหลักแรกและ การหาตัวประมาณของพารามิเตอร์** **9**

**3 การเลือกช่วงของ $q$ ที่เหมาะสม** **14**

3.1 ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}$ สำหรับแต่ละ $\alpha$ และ $q$ . . . . . 14

3.1.1 กรณี $q = 1$ . . . . . 15

3.1.2 กรณี $q = 2$ . . . . . 15

3.1.3 กรณี $q = 3$ . . . . . 16
3.1.4 กรณี $q = 4$ . . . . . 16
3.1.5 กรณี $q = 5$ . . . . . 17
3.1.6 กรณี $q = 6$ . . . . . 17
3.1.7 กรณี $q = 7$ . . . . . 18
3.1.8 กรณี $q = 8$ . . . . . 18
3.1.9 กรณี $q = 9$ . . . . . 19
3.1.10 กรณี $q = 10$ . . . . . 19
3.2 พิจารณาค่าลอกอริทึมของฟังก์ชันความควรจะเป็นของข้อมูลที่มีการแจกแจงสอดคล้องกับกฎของเบนฟอร์ด . . . . . 20
3.2.1 ข้อมูลตามกฎของเบนฟอร์ด . . . . . 20
3.2.2 ข้อมูลที่อยู่อาศัย . . . . . 21
3.2.3 ข้อมูลประชากร . . . . . 23
3.2.4 ข้อมูลพื้นที่ของแม่น้ำแต่ละแห่งทั่วโลก . . . . . 23
3.2.5 ข้อมูลค่าพลังงานของการแผ่รังสีของวัตถุดำ . . . . . 24

**4 ทดสอบความเป็นเบนฟอร์ดของข้อมูลค่าไฟฟ้า 27**
4.1 การหาค่าประมาณของพารามิเตอร์ $\alpha$ สำหรับข้อมูลค่าไฟฟ้าแต่ละเดือน . . . . . 28
4.1.1 เดือนกุมภาพันธ์ . . . . . 29
4.1.2 เดือนมีนาคม . . . . . 30
4.1.3 เดือนเมษายน . . . . . 31
4.1.4 เดือนพฤษภาคม . . . . . 32
4.1.5 เดือนมิถุนายน . . . . . 33
4.1.6 เดือนกรกฎาคม . . . . . 34
4.1.7 เดือนสิงหาคม . . . . . 35
4.1.8 เดือนกันยายน . . . . . 36
4.1.9 เดือนตุลาคม . . . . . 37
4.1.10 เดือนพฤศจิกายน . . . . . 38
4.1.11 เดือนธันวาคม . . . . . 39
4.2 สรุปผลการทดสอบความเป็นเบนฟอร์ดของข้อมูลค่าไฟฟ้ารายเดือน . . . . . 40

4.2.1 ราคาค่าไฟฟ้า . . . . . 41
4.2.2 ราคาค่าไฟฟ้าหน่วยละ 5 บาท . . . . . 42
4.3 ปัญหาและข้อเสนอแนะ . . . . . 43

**ภาคผนวก ก** **45**

**ภาคผนวก ข** **49**

**ประวัติผู้เขียน** **51**

# สารบัญรูป

3.1 ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 1$ . . . . 15
3.2 ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 2$ . . . . 15
3.3 ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 3$ . . . . 16
3.4 ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 4$ . . . . 16
3.5 ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 5$ . . . . 17
3.6 ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 6$ . . . . 17
3.7 ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 7$ . . . . 18
3.8 ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 8$ . . . . 18
3.9 ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 9$ . . . . 19
3.10 ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 10$ . . . . 19

# สารบัญตาราง

1.1 ตารางการแจกแจงของกฎของเบนฟอร์ด . . . . . 7

3.1 จำนวนข้อมูลที่มีเลขหลักแรกเป็น $d$ ของข้อมูลตามกฎของเบนฟอร์ด . . . . . 21
3.2 ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลตามกฎของเบนฟอร์ด . . . . . 21
3.3 จำนวนข้อมูลที่มีเลขหลักแรกเป็น $d$ ของข้อมูลที่อยู่อาศัย . . . . . 22
3.4 ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลที่อยู่อาศัย . . . . . 22
3.5 จำนวนข้อมูลที่มีเลขหลักแรกเป็น $d$ ของข้อมูลที่อยู่อาศัย . . . . . 23
3.6 ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลประชากร . . . . . 23
3.7 จำนวนข้อมูลที่มีเลขหลักแรกเป็น $d$ ของข้อมูลพื้นที่แม่น้ำแต่ละแห่งทั่วโลก . . . . 24
3.8 ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลพื้นที่ของแม่น้ำแต่ละแห่งทั่วโลก . . . . . 24
3.9 จำนวนข้อมูลที่มีเลขหลักแรกเป็น $d$ ของข้อมูลค่าพลังงานของการแผ่รังสีของวัตถุดำ . . . . . 25
3.10 ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลค่าพลังงานของการแผ่รังสีของวัตถุดำ . . . . . 25

4.1 ความถี่ที่ได้จากการทดลองของเลขโดดหลักแรกในเดือนกุมภาพันธ์ ปี พ.ศ. 2557 จากการสุ่มจำนวน 1,000 ตัวอย่าง . . . . . 29
4.2 ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาท ในเดือนกุมภาพันธ์ พ.ศ. 2557 . . . . . 29

4.3 ความถี่ที่ได้จากการทดลองของเลขโดดหลักแรกในเดือนมีนาคม ปี พ.ศ. 2557 จากการสุ่มจำนวน 1,000 ตัวอย่าง . . . . . 30
4.4 ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาท ในเดือนมีนาคม พ.ศ. 2557 30
4.5 ความถี่ที่ได้จากการทดลองของเลขโดดหลักแรกในเดือนเมษายน ปี พ.ศ. 2557 จากการสุ่มจำนวน 1,000 ตัวอย่าง . . . . . 31
4.6 ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาท ในเดือนเมษายน พ.ศ. 2557 31
4.7 ความถี่ที่ได้จากการทดลองของเลขโดดหลักแรกในเดือนพฤษภาคม ปี พ.ศ. 2557 จากการสุ่มจำนวน 1,000 ตัวอย่าง . . . . . 32
4.8 ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาท ในเดือนพฤษภาคม พ.ศ. 2557 . . . . . 32
4.9 ความถี่ที่ได้จากการทดลองของเลขโดดหลักแรกในเดือนมิถุนายน ปี พ.ศ. 2557 จากการสุ่มจำนวน 1,000 ตัวอย่าง . . . . . 33
4.10 ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาท ในเดือนมิถุนายน พ.ศ. 2557 33
4.11 ความถี่ที่ได้จากการทดลองของเลขโดดหลักแรกในเดือนกรกฎาคม ปี พ.ศ. 2557 จากการสุ่มจำนวน 1,000 ตัวอย่าง . . . . . 34
4.12 ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาท ในเดือนกรกฎาคม พ.ศ. 2557 . . . . . 34
4.13 ความถี่ที่ได้จากการทดลองของเลขโดดหลักแรกในเดือนสิงหาคม ปี พ.ศ. 2557 จากการสุ่มจำนวน 1,000 ตัวอย่าง . . . . . 35
4.14 ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาท ในเดือนสิงหาคม พ.ศ. 2557 35
4.15 ความถี่ที่ได้จากการทดลองของเลขโดดหลักแรกในเดือนกันยายน ปี พ.ศ. 2557 จากการสุ่มจำนวน 1,000 ตัวอย่าง . . . . . 36
4.16 ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาท ในเดือนกันยายน พ.ศ. 2557 36

4.17 ความถี่ที่ได้จากการทดลองของเลขโดดหลักแรกในเดือนตุลาคม ปี พ.ศ. 2557 จากการสุ่มจำนวน 1,000 ตัวอย่าง . . . . . 37
4.18 ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาท ในเดือนตุลาคม พ.ศ. 2557 37
4.19 ความถี่ที่ได้จากการทดลองของเลขโดดหลักแรกในเดือนพฤศจิกายน ปี พ.ศ. 2557 จากการสุ่มจำนวน 1,000 ตัวอย่าง . . . . . 38
4.20 ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาท ในเดือนพฤศจิกายน พ.ศ. 2557 . . . . . 38
4.21 ความถี่ที่ได้จากการทดลองของเลขโดดหลักแรกในเดือนธันวาคม ปี พ.ศ. 2557 จากการสุ่มจำนวน 1,000 ตัวอย่าง . . . . . 39
4.22 ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาท ในเดือนธันวาคม พ.ศ. 2557 39
4.23 ผลการทดสอบความเป็นเบนฟอร์ดของข้อมูลราคาค่าไฟฟ้าของแต่ละเดือนในปี พ.ศ. 2557 ด้วยการทดสอบไคสแควร์ และสังเกตระยะห่างของค่าประมาณ $\alpha$ กับ 1 . . . . . 41
4.24 ผลการทดสอบความเป็นเบนฟอร์ดของข้อมูลราคาค่าไฟฟ้าหน่วยละ 5 บาท ของแต่ละเดือนในปี พ.ศ. 2557 ด้วยการทดสอบไคสแควร์ และสังเกตระยะห่างของค่าประมาณ $\alpha$ กับ 1 . . . . . 42

# บทที่ 1
# บทนำและความรู้พื้นฐาน

## 1.1 ที่มาและความสำคัญ

ด้วยวิถีชีวิตที่รวดเร็วและสะดวกสบายของคนในสังคมปัจจุบันส่งผลให้ธุรกิจทุกแขนงมีการแข่งขันที่สูงขึ้น ทั้งธุรกิจออนไลน์และออฟไลน์ จึงต้องมีการพัฒนาให้ธุรกิจของตนเอื้ออำนวยความสะดวกต่อกลุ่มเป้าหมายให้มากและเร็วที่สุดเท่าที่จะทำได้ แต่การที่แต่ละธุรกิจจะทำเช่นนั้นได้สิ่งที่ทำให้เขาทราบความต้องการของกลุ่มเป้าหมายได้ ก็คือข้อมูลต่างๆ ที่เกี่ยวข้องกับกลุ่มเป้าหมาย อาทิเช่น อายุ เพศ อัตราเงินเดือน รายจ่ายในแต่ละเดือน เป็นต้น แน่นอนว่าข้อมูลเหล่านี้เป็นทั้งข้อมูลเชิงคุณภาพและเชิงปริมาณที่มีจำนวนมาก อาจเป็นการยากที่จะทำการจัดเก็บ จัดการ หรือตรวจสอบการปลอมแปลงของข้อมูลนั้นๆ เราจึงสนใจการตรวจจับความผิดปกติในเบื้องต้นของข้อมูลเชิงปริมาณที่มีจำนวนข้อมูลมาก ด้วยเครื่องมือทางคณิตศาสตร์ที่เรียกว่า กฎของเบนฟอร์ด ที่นำเลขหลักแรกของข้อมูลมาพิจารณา

กฎของเบนฟอร์ด (Benford's law) หรือกฎของเลขโดดหลักแรก เป็นกฎที่ได้จากการสังเกตจำนวนต่างๆ จากข้อมูลในชีวิตจริง อาทิเช่น ค่าไฟฟ้า บ้านเลขที่ ราคาหุ้น จำนวนประชากร เป็นต้น และพบว่าจำนวนจากข้อมูลเหล่านี้นำหน้าหรือขึ้นด้วยตัวเลข 1 มากถึงร้อยละ 30 ของจำนวนทั้งหมด นอกจากนี้ยังได้อีกว่า เมื่อจำนวนเหล่านี้มีค่ามากขึ้น โอกาสที่จะนำหน้าหรือขึ้นต้นด้วยตัวเลขที่มีค่ามากก็จะลดลง จากการสังเกต จึงได้ว่าจำนวนที่นำหน้าหรือขึ้นต้นด้วยตัวเลข 9 น้อยกว่าจำนวนที่นำหน้าหรือขึ้นต้นด้วยตัวเลข 1 ถึง 7 เท่า ชื่อของกฎเบนฟอร์ดมาจากนักฟิสิกส์และวิศวกรไฟฟ้าชาวอเมริกันนามว่า แฟรงก์ เบนฟอร์ด (Frank Benford) ซึ่งกล่าวถึงปรากฏการณ์นี้ในปี ค.ศ. 1938 อย่างไรก็ตามนักดาราศาสตร์และนักคณิตศาสตร์ชาวแคนาดา-อเมริกัน ไซมอน นิวคอมบ์

(Simon Newcomb) เคยอ้างถึงปรากฏการณ์นี้ไว้เมื่อปี ค.ศ. 1881 อีกทั้งกฎนี้ยังได้รับการพิสูจน์ทางคณิตศาสตร์อย่างเคร่งครัดเป็นครั้งแรกโดย ธี โอดอร์ ฮิลล์ (Theodore P. Hill) ในปี 1988 อีกด้วย ซึ่งกฎของเบนฟอร์ดได้อธิบายความถี่ของการกระจายเลขโดดหลักแรกแต่ละตัว กล่าวคือ ตัวแปรสุ่ม $X$ เป็นไปตามกฎของเบนฟอร์ดเมื่อ ความน่าจะเป็นที่เลขโดดหลักแรกของ $X$ เป็น $d$ มีค่าเท่ากับ $\log(1+\frac{1}{d})$ เมื่อ $d = 1, 2, 3, ..., 9$ นั่นคือ

$$P(D(X) = d) = \log(1 + \frac{1}{d})$$

เมื่อ $D(x)$ คือเลขโดดหลักแรกของ $x$

การที่เราจะทราบว่าการแจกแจงของข้อมูลนั้นๆ เป็นไปตามกฎของเบนฟอร์ดหรือไม่ สามารถทำได้โดยการทดสอบทางสถิติที่ชื่อว่าการทดสอบไคแสควร์ (Chi-Square test) แต่ว่าการทดสอบการแจกแจงของข้อมูลด้วยวิธีนี้มีจุดบกพร่องคือ เมื่อจำนวนข้อมูลมีปริมาณมาก ผลการทดสอบมักจะปฏิเสธสมมติฐานที่ว่าการแจกแจงนั้นสอดคล้องกับกฎของเบนฟอร์ด

ในโครงงานนี้เราจึงสนใจการทดสอบความเป็นเบนฟอร์ดของข้อมูลค่าใช้จ่ายไฟฟ้ารายเดือน เนื่องจากเป็นข้อมูลเชิงปริมาณและเป็นข้อมูลในชีวิตจริง โดยการเลือกการแจกแจงที่มีฟังก์ชันการแจกแจง $G_{\alpha,q}(s)$ โดยที่ $\alpha > 0, q$ เป็นจำนวนนับ และ $s \in [0, 10]$ ซึ่ง

$$\lim_{n\to\infty} Man_{F_{\alpha,q}^{(n)}}(s) = G_{\alpha,q}(s)$$

เมื่อ $Man_{X_n}(s) = \sum_{l=-\infty}^{\infty} \mathbb{P}\left(10^{-l} \leq X_n \leq s10^{-l}\right)$
และ $F_{\alpha,q}^{(n)}$ คือ ฟังก์ชันการแจกแจงที่สร้างจากการสุ่มตัวอย่างในช่วงที่มีการตัดปลายแบบลูกโซ่ $n$ ครั้งของการแจกแจงเบต้าคงตัวเป็นช่วง
เนื่องจากเราทราบว่าตัวแปรสุ่ม $X$ สอดคล้องกับกฎของเบนฟอร์ดเมื่อ $Man_X(s) = \log_{10} s$ ซึ่งฟังก์ชันการแจกแจง $G_{\alpha,q}(s)$ ลู่เข้าสู่ค่า $\log_{10} s$ เมื่อ $q \to \infty$ และ $G_{\alpha,q}(s) = \log_{10} s$ เมื่อ $\alpha = 1$ สำหรับทุก $q \in \mathbb{N}$ เราจึงคิดว่าการแจกแจงที่มีฟังก์ชันการแจกแจง $G_{\alpha,q}(s)$ จะเป็นแบบจำลองของเลขหลักแรกในธรรมชาติได้ดี และจะกล่าวได้ว่าการแจกแจงดังกล่าวลู่เข้าสู่กฎของเบนฟอร์ด เมื่อ $q \to \infty$ หรือ $\alpha = 1$ สำหรับทุก $q \in \mathbb{N}$

เราจึงต้องการหาตัวประมาณของพารามิเตอร์ $\alpha$ โดยวิธีความควรจะเป็นสูงสุด (Maximum Likelihood) ซึ่งจะแสดงไว้ในบทที่ 2 โดยเริ่มจากการพิจารณาผลคูณของฟังก์ชันความน่าจะเป็น

$g_{\alpha,q}(s)$ โดยที่

$$g_{\alpha,q}(d) = G_{\alpha,q}(d+1) - G_{\alpha,q}(d)$$

เมื่อ $d = 1, 2, 3, ..., 9$

จากนั้นเราจะทำการหาช่วงของจำนวนนับ $q$ ที่เหมาะสมในการนำไปทดสอบความเป็นเบนฟอร์ดของข้อมูล ซึ่งจะแสดงไว้ในบทที่ 3 โดยการสังเกตว่าค่าสูงสุดของฟังก์ชันความควรจะเป็น (likelihood function) มักเกิดขึ้นเมื่อ $q$ อยู่ในช่วงใด เมื่อได้ตัวประมาณของ $\alpha$ จากบทที่ 2 และช่วงของ $q$ ที่เหมาะสมจากบทที่ 3 ก็จะนำไปทดสอบความเป็นเบนฟอร์ดของข้อมูลค่าใช้จ่ายไฟฟ้ารายเดือนในบทที่ 4 และทำการเลือกค่าประมาณของ $\alpha$ และค่า $q$ ที่ทำให้ฟังก์ชันความควรจะเป็นมีค่าสูงสุด เพื่อดูระยะห่างระหว่างค่าประมาณ $\alpha$ สำหรับแต่ละข้อมูล และ 1

## 1.2 ความรู้พื้นฐานทางคณิตศาสตร์

### 1.2.1 ตัวแปรสุ่ม

**บทนิยาม 1.2.1.1. ตัวแปรสุ่ม (Random Variables)** คือฟังก์ชันที่มีค่าเป็นจำนวนจริง ซึ่งกำหนดโดยแต่ละสมาชิกในปริภูมิตัวอย่าง

นิยมเขียนตัวอักษรตัวพิมพ์ใหญ่ เช่น $X, Y, Z$ แทนตัวแปรสุ่ม และใช้ตัวอักษรพิมพ์เล็ก เช่น $x, y, z$ แทนค่าที่เป็นไปได้ของตัวแปรสุ่ม

ในโครงงานนี้จะทดสอบความเป็นเบนฟอร์ดของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาทของผู้ใช้ไฟฟ้าที่อาศัยอยู่ในกรุงเทพฯ และปริมณฑลที่ได้จากการสุ่ม โดย [4] ฉะนั้นตัวแปรสุ่มของข้อมูลข้างต้น จะเป็นตัวแปรสุ่มประเภทไม่ต่อเนื่อง

**บทนิยาม 1.2.1.2.** เราจะกล่าวว่า ตัวแปรสุ่ม $X$ เป็น **ตัวแปรสุ่มไม่ต่อเนื่อง (discrete random variable)** ถ้าพิสัยของ $X$ เป็นเซตนับได้ และเราเรียกฟังก์ชัน $f : R \to [0, 1]$ ที่กำหนดโดย

$$f(x) = P(X = x)$$

ว่าฟังก์ชันความน่าจะเป็น (probability function) ของ $X$

**บทนิยาม 1.2.1.3.** เราจะเรียกฟังก์ชัน $F : R \to [0, 1]$ ว่า **ฟังก์ชันการแจกแจงสะสม (cumulative distribution function)** หรือเรียกโดยย่อว่า **ฟังก์ชันการแจกแจง (distribution func-**

**tion)** ของตัวแปรสุ่ม $X$ ถ้า สำหรับจำนวนจริง $x$ ใดๆ

$$F(x) = P(X \leq x)$$

**ทฤษฎีบท 1.2.1.4.** *ให้ $F : R \to [0,1]$ จะได้ว่า $F$ เป็นฟังก์ชันการแจกแจง ก็ต่อเมื่อ*

1. *$F$ เป็นฟังก์ชันไม่ลด นั่นคือ ถ้า $x_1 \leq x_2$ แล้ว $F(x_1) \leq F(x_2)$*
2. *$\lim_{x\to\infty} F(x) = 1$ และ $\lim_{x\to-\infty} F(x) = 0$*
3. *$F$ มีความต่อเนื่องทางขวา นั่นคือ สำหรับจำนวนจริง $x_0$ ใดๆ*

$$\lim_{x\to x_0^+} F(x) = F(x_0)$$

**บทนิยาม 1.2.1.5. การแจกแจงที่ได้จากสุ่มตัวอย่างในช่วงที่มีการตัดปลายแบบลูกโซ่ (chains of truncated distribution)** ให้ $D[0,\infty]$ เป็นการแจกแจงบนช่วง $[0,\infty]$ นั่นคือฟังก์ชันการแจกแจง $F$ โดยที่ $F(x) = 0$ สำหรับทุก $x < 0$ เราจะนิยามการแจกแจงที่ได้จากการตัดปลายช่วงของ $D[x,\infty]$ ที่จุด $a \in (0,\infty)$ ว่าคือการแจกแจง $D[0,a]$ บนช่วง $[0,a]$ ที่มีฟังก์ชันการแจกแจง $F^{[0,a]}$ ซึ่งถูกนิยามโดย

$$F^{[0,a]}(x) = \begin{cases} \frac{F(x)}{F(a)} & \text{สำหรับ } x \leq a \\ 1 & \text{สำหรับ } x > a \end{cases}$$

เมื่อเราให้ตัวแปรสุ่ม $X_1 \sim D[0,\infty]$ นั่นคือ $X_1$ มีการแจกแจงแบบ $D[0,\infty]$ และให้ $X_n \sim D[0,X_{n-1}]$ สำหรับ $n \geq 2$ โดยที่ $D[0,X_{n-1}]$ เป็นการแจกแจงที่ได้จากการตัดปลายช่วงของ $D[0,X_{n-2}]$ ที่จุด $X_{n-1}$

เราจะเรียกลำดับของตัวแปรสุ่ม $X_n$ ว่า การแจกแจงที่ได้จากการตัดปลายช่วงแบบลูกโซ่ ซึ่งก่อกำเนิดโดย $D[0,\infty]$ หรือเรียกอีกอย่างได้ว่า การแจกแจงที่ได้จากการตัดปลายช่วงแบบลูกโซ่ ซึ่งก่อกำเนิดโดยฟังก์ชันการแจกแจง $F$

เนื่องจากตัวแปรสุ่มที่มีการแจกแจงแบบเบต้าคงตัวเป็นช่วงเป็นตัวแปรสุ่มต่อเนื่อง และในโครงงานนี้เราจะทำการประมาณค่าพารามิเตอร์ของการแจกแจงเลขหลักแรกที่ได้จากการสุ่มตัวอย่างในช่วงที่มีการตัดปลายแบบลูกโซ่ของการแจกแจงเบต้าคงตัวเป็นช่วง ดังนั้นตัวแปรสุ่มที่มีการแจกแจงดังกล่าว จะเป็นตัวแปรสุ่มต่อเนื่อง

**บทนิยาม 1.2.1.6.** เราจะกล่าวว่า ตัวแปรสุ่ม $X$ เป็น **ตัวแปรสุ่มต่อเนื่อง (continuous random variable)** ถ้ามีฟังก์ชัน $f : R \to [0, \infty)$ ซึ่งสำหรับทุก $x \in R$

$$F_X(x) = \int_{-\infty}^{x} f(t)dt$$

เราเรียก $f$ ว่า **ฟังก์ชันความหนาแน่นความน่าจะเป็น (probability density function)** ของ $X$

*ข้อสังเกต.* $\int_{-\infty}^{\infty} f(t)dt = \lim_{x\to\infty} F(x) = 1$

## 1.2.2 วิธีของความควรจะเป็นสูงสุด

ในขั้นตอนการประมาณค่าพารามิเตอร์ เราจะใช้วิธีความควรจะเป็นสูงสุดในการหาตัวประมาณของพารามิเตอร์ $\alpha$ ซึ่งเป็นพารามิเตอร์ตัวหนึ่งของการแจกแจงเลขหลักแรกที่ได้จากการสุ่มตัวอย่างในช่วงที่มีการตัดปลายแบบลูกโซ่ของการแจกแจงเบต้าคงตัวเป็นช่วง

**บทนิยาม 1.2.2.1.** ให้ $X_1, X_2, ..., X_n$ เป็นตัวอย่างสุ่มที่ให้ค่า $\{x_1, x_2, ..., x_n\}$ จากประชากรที่มีการแจกแจง (ฟังก์ชันความหนาแน่นความน่าจะเป็น ในกรณีต่อเนื่อง หรือ ฟังก์ชันความน่าจะเป็นในกรณีที่ไม่ต่อเนื่อง) เป็น $f_\theta(x)$ ซึ่งมีพารามิเตอร์ $\theta \in \Theta$ แล้ว ฟังก์ชันความควรจะเป็น (Likelihood function) ของตัวอย่างสุ่มนี้คือ

$$L(\theta) = L(\theta; x_1, x_2, ..., x_n) = f_\theta(x_1) \cdots f_\theta(x_n) = \prod_{i=1}^{n} f_\theta(x_i)$$

ตัวประมาณแบบความควรจะเป็นสูงสุด (maximum likelihood estimator or MLE) ของพารามิเตอร์ $\theta$ คือ $\hat{\theta} = \hat{\theta}(X_1, X_2, ...X_n)$ ที่ทำให้ฟังก์ชันควรจะเป็น $L(\theta)$ มีค่าสูงที่สุด สำหรับ $\theta \in \Theta$

**ตัวอย่าง 1.2.2.2.** *สุ่มตัวอย่างขนาด $n$ จากการแจกแจงแบร์นูลลีที่มีพารามิเตอร์ $p$ จงหาตัวประมาณแบบความควรจะเป็นสูงสุด $p$*

*วิธีทำ.* การแจกแจงแบร์นูลลี เป็นการแจกแจงของตัวแปรสุ่มที่มีผลลัพธ์ที่เป็นไปได้ 2 ทาง เช่น ความสำเร็จและความไม่สำเร็จ มีกับไม่มี เป็นสมาชิกกับไม่เป็นสมาชิก ชายกับหญิง เป็นต้น ซึ่งตัวแปรเชิง

สุ่ม $X$ แทนผลลัพธ์จากการทดลอง 2 ทางนี้ โดยกำหนดให้

$$X = \begin{cases} 1 & \text{เมื่อผลการทดลองสำเร็จ} \\ 0 & \text{เมื่อผลการทดลองไม่สำเร็จ} \end{cases}$$

และ $p$ คือ ความน่าจะเป็นของการเกิดเหตุการณ์ความสำเร็จ ฉะนั้น $1-p$ (หรือเขียนแทนด้วย $q$) คือ ความน่าจะเป็นของการเกิดเหตุการณ์ความไม่สำเร็จ โดยมีฟังก์ชันการแจกแจงความน่าจะเป็นดังนี้

$$f(x) = (1-p)^{1-x}p^x$$

เราจึงได้ปริภูมิพารามิเตอร์ $\Theta = \{p : 0 \leq p \leq 1\}$ และฟังก์ชันควรจะเป็นคือ $L(p) = f_p(x_1)f_p(x_2)\cdots f_p(x_n)$ เมื่อ $f_p(x) = (1-p)^{1-x}p^x$ และ $x_1, x_2, ..., x_n$ เป็นค่าที่ได้จากตัวแปรสุ่ม $X_1, X_2, ..., X_n$ ดังนั้นเราต้องหาค่า $p$ ในรูปของ $x_1, x_2, ..., x_n$ ที่ให้ค่าสูงสุดของ

$$L(p) = \prod_{i=1}^{n}(1-p)^{1-x_i}p^{x_i} = (1-p)^{\sum_{i=1}^{n}(1-x_i)}p^{\sum_{i=1}^{n}x_i}$$

เนื่องจากฟังก์ชันลอการิทึมเป็นฟังก์ชันเพิ่ม เราจึงสามารถหา $p$ ซึ่งให้ค่าสูงสุดของ

$$\ln L(p) = \left(\sum_{i=1}^{n}(1-x_i)\right)\ln(1-p) + \left(\sum_{i=1}^{n}x_i\right)\ln p$$

โดยการหาจุดวิกฤตจากสมการ

$$\frac{d}{dp}\ln L(p) = \frac{\sum_{i=1}^{n}x_i}{p} - \frac{n-\sum_{i=1}^{n}x_i}{1-p} = 0$$

คูณตลอดทั้งสมการด้วย $\frac{p(1-p)}{n}$ จะได้

$$\begin{aligned} \tfrac{(1-p)\sum_{i=1}^{n}x_i}{n} - \tfrac{(n-\sum_{i=1}^{n}x_i)p}{n} &= 0 \\ (1-p)\tfrac{\sum_{i=1}^{n}x_i}{n} - p\left(1-\tfrac{\sum_{i=1}^{n}x_i}{n}\right) &= 0 \\ \tfrac{\sum_{i=1}^{n}x_i}{n} - p &= 0 \\ p &= \tfrac{\sum_{i=1}^{n}x_i}{n} \end{aligned}$$

จุดวิกฤตเพียงจุดเดียวซึ่งแสดงได้ว่าเป็นจุดสูงสุดคือ $p = \frac{1}{n}\sum_{i=1}^{n} x_i$ จึงสรุปได้ว่าตัวประมาณแบบความควรจะเป็นสูงสุดของ $p$ คือ $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}_n$ □

### 1.2.3 แมนทิสซาและกฎของเบนฟอร์ด

**บทนิยาม 1.2.3.1.** สำหรับจำนวนจริง $x > 0$ เราสามารถเขียน $x = M \times 10^k$ ได้เพียงวิธีเดียว โดยที่ $k \in Z$ และ $1 \leq M < 10$ เราจะเรียก $M$ ว่า แมนทิสซาของ $x$

**บทนิยาม 1.2.3.2.** ให้ $X$ เป็นตัวแปรสุ่มแทนข้อมูลที่เราสนใจที่ให้ค่า $x$ และ $D$ เป็นฟังก์ชัน $D : x \to \{1, 2, 3, ..., 9\}$ นิยามโดย $D(x)$ คือเลขโดดหลักแรกของ $x$ จะได้ว่า $X$ สอดคล้องกับกฎของเบนฟอร์ด (Benford's law) ถ้า $X$ มีการแจกแจงของเลขโดดหลักแรก เป็นดังนี้

$$P(D(X) = d) = \log_{10}(1 + \frac{1}{d}) = \log_{10}(\frac{d+1}{d})$$

เมื่อ $d = 1, 2, ..., 9$

จากบทนิยามกฎของเบนฟอร์ด จะได้ความน่าจะเป็นที่เลขโดดหลักแรกจะเป็น 1,2,3,...,9 ดังตารางต่อไปนี้

| $d$ | $P(D(X) = d)$ |
|---|---|
| 1 | 0.3010 |
| 2 | 0.1761 |
| 3 | 0.1249 |
| 4 | 0.0969 |
| 5 | 0.0792 |
| 6 | 0.0669 |
| 7 | 0.0580 |
| 8 | 0.0512 |
| 9 | 0.0458 |

ตารางที่ 1.1: ตารางการแจกแจงของกฎของเบนฟอร์ด

จะได้ว่าฟังก์ชันการแจกแจงสะสมของเลขหลักแรก เมื่อเป็นไปตามกฎของเบนฟอร์ด คือ

$$\begin{aligned} F(d) = P(D(X) \leq d) &= \log_{10}(1+\tfrac{1}{1}) + \log_{10}(1+\tfrac{1}{2}) + \cdots + \log_{10}(1+\tfrac{1}{d}) \\ &= \log_{10}(\tfrac{2}{1}) + \log_{10}(\tfrac{3}{2}) + \cdots + \log_{10}(\tfrac{d+1}{d}) \\ &= \log_{10}\left(\tfrac{2}{1} \times \tfrac{3}{2} \times \cdots \times \tfrac{d+1}{d}\right) \\ &= \log_{10}(d+1) \end{aligned}$$

เมื่อ $d = 1, 2, ..., 9$ ดังนั้น

$$\begin{aligned} f(d) &= P(D(X) = d) \\ &= \log_{10}(d+1) - \log_{10}(d) \\ &= F(d+1) - F(d) \end{aligned}$$

เมื่อ $f$ เป็นฟังก์ชันความน่าจะเป็น และ $F$ เป็นฟังก์ชันความน่าจะเป็นสะสมหรือฟังก์ชันการแจกแจง

**บทนิยาม 1.2.3.3.** ให้ $D$ เป็นการแจกแจงบนช่วง $[0, k]$ สำหรับ $k > 0$ จะกล่าวว่าตัวแปรสุ่ม $X \sim D[0, k]$ สอดคล้องกับกฎของเบนฟอร์ดเมื่อ ความน่าจะเป็นที่แมนทิสซาของตัวแปรสุ่ม $X$ จะมีค่าอยู่ในช่วง $[1, s]$ เท่ากับ $\log_{10} s$ สำหรับ $s \in [1, 10]$ และจะกล่าวได้ว่าลำดับของตัวแปรสุ่ม $X_n$ โดยที่ $n \in N$ สอดคล้องกับกฎของเบนฟอร์ดในลิมิตเมื่อ ความน่าจะเป็นที่แมนทิสซาของลำดับ $\{X_n\}_{n \in N}$ จะมีค่าอยู่ในช่วง $[1, s]$ ลู่เข้าสู่ค่า $\log_{10} s$ เมื่อ $n \to \infty$ นั่นคือ

$$\lim_{n \to \infty} Man_{X_n}(s) = \log_{10} s$$

เมื่อ $Man_{X_n}(s) = \sum_{l=-\infty}^{\infty} \mathbb{P}\left(10^{-l} \leq X_n \leq s10^{-l}\right)$

# บทที่ 2
# ฟังก์ชันการแจกแจงของเลขหลักแรกและการหาตัวประมาณของพารามิเตอร์

จากการศึกษาบทความ [1] และจากข้อคาดการณ์ที่ว่า เมื่อให้ $F$ เป็นฟังก์ชันการแจกแจง $D$ บนช่วง $[0,k]$ เมื่อ $k \in R$ และให้ $\{F^{(n)}\}_{n\in N}$ เป็นฟังก์ชันการแจกแจงของการแจกแจงที่สร้างจากการสุ่มตัวอย่างในช่วงที่มีการตัดปลายแบบลูกโซ่ $n$ ครั้ง ซึ่งก่อกำเนิดโดย $F$ จะได้ว่าความน่าจะเป็นที่แมนทิสซาของ $F^{(n)}$ จะมีค่าอยู่ในช่วง $[1,s]$ เมื่อ $s \in [1,10]$ ลู่เข้าสู่กฎของเบนฟอร์ด เมื่อ $n \to \infty$ แต่จากการพิสูจน์ของนายธีรพจน์ วิริยะไกรกุล ในบทความ [1] ที่ว่า เมื่อให้ $\{F_{\alpha,q}^{(n)}\}_{n\in N}$ โดยที่ $\alpha > 0$ และ $q \in N$ เป็นฟังก์ชันการแจกแจงของการแจกแจงที่สร้างจากสุ่มตัวอย่างในช่วงที่มีการตัดปลายแบบลูกโซ่ $n$ ครั้ง ซึ่งก่อกำเนิดโดยการแจกแจงเบต้าคงตัวเป็นช่วง จะได้ว่าความน่าจะเป็นที่แมนทิสซาของ $F_{\alpha,q}^{(n)}$ จะมีค่าอยู่ในช่วง $[1,s]$ เมื่อ $s \in [1,10]$ ไม่ได้ลู่เข้าสู่กฎของเบนฟอร์ดเมื่อ $n \to \infty$ แต่ลู่เข้าสู่ฟังก์ชันการแจกแจง $G_{\alpha,q}(s)$ นั่นคือ

$$\lim_{n\to\infty} Man_{F_{\alpha,q}{}^{(n)}}(s) = G_{\alpha,q}(s)$$

โดยที่

$$G_{\alpha,q}(s) = \begin{cases} 0 & , s < 1 \\ \frac{1}{\alpha}\log_{10}\left[\left(\frac{10^{\frac{\alpha}{q}}-1}{10^{\frac{1}{q}}-1}\right)10^{(\alpha-1)(b_{s,q}/q)}(s-10^{b_{s,q}/q})+10^{(\alpha b_{s,q})/q}\right] & , s \in [1,10] \\ 1 & , s > 10 \end{cases}$$

เมื่อ $b_{s,q} = \lfloor q \log_{10} s \rfloor$
และได้ด้วยว่า ถ้า $q \to \infty$ แล้ว $\lim_{n\to\infty} Man_{F^{(n)}{}_{\alpha,q}}(s) = G_{\alpha,q}(s)$ ลู่เข้าสู่ค่า $\log_{10} s$ อย่างสม่ำเสมอ บนช่วง [1,10] อีกทั้งยังเห็นได้ชัดว่า $G_{\alpha,q}(s) = \log_{10} s$ เมื่อ $\alpha = 1$ สำหรับทุก $q \in N$

โดยบทนิยาม 1.2.3.3 เราจึงคิดว่าการแจกแจงที่มี $G_{\alpha,q}(s)$ เป็นฟังก์ชันการแจกแจง จะเป็นแบบจำลองของเลขหลักแรกของข้อมูลในธรรมชาติได้ดี และจะเห็นได้ว่าการแจกแจงเลขหลักแรกของข้อมูลเหล่านั้น สอดคล้องกับกฎของเบนฟอร์ดเมื่อ $q \to \infty$ หรือ $\alpha = 1$ เราจึงสนใจการทดสอบความเป็นเบนฟอร์ดของข้อมูลต่างๆ ด้วยการหาตัวประมาณของพารามิเตอร์ $\alpha$ โดยวิธีความควรจะเป็นสูงสุด เพื่อเปรียบเทียบค่าประมาณของพารามิเตอร์ $\alpha$ กับ 1 และหาช่วง $q$ ที่เหมาะสม ซึ่งจะกล่าวในบทถัดไป

เราจะทำการหาตัวประมาณของพารามิเตอร์ $\alpha$ ของฟังก์ชันการแจกแจง $G_{\alpha,q}(s)$ ด้วยวิธีความควรจะเป็นสูงสุด โดยเริ่มจากพิจารณาฟังก์ชันความน่าจะเป็นของการแจกแจง จากกฎของเบนฟอร์ดที่ว่า

$$f(d) = F(d+1) - F(d)$$

เมื่อ $f$ เป็นฟังก์ชันความน่าจะเป็น, $F$ เป็นฟังก์ชันความน่าจะเป็นสะสมหรือฟังก์ชันการแจกแจง และ $d = 1, 2, 3, ..., 9$

ดังนั้น เมื่อเรากำหนดให้ $g_{\alpha,q}(s)$ เป็นฟังก์ชันความน่าจะเป็นของการแจกแจงที่มีฟังก์ชันการแจกแจง $G_{\alpha,q}(s)$ เราจะได้ว่าฟังก์ชันความน่าจะเป็น

$$g_{\alpha,q}(d) = G_{\alpha,q}(d+1) - G_{\alpha,q}(d)$$

เมื่อ $d = 1, 2, 3, ..., 9$

$$\begin{aligned}
\text{เนื่องจาก } G_{\alpha,q}(s) &= \tfrac{1}{\alpha}\log_{10}\left[\left(\tfrac{10^{\frac{\alpha}{q}}-1}{10^{\frac{1}{q}}-1}\right)10^{(\alpha-1)(b_{s,q}/q)}(s-10^{b_{s,q}/q})+10^{(\alpha b_{s,q})/q}\right] \\
&= \tfrac{1}{\alpha}\log_{10}\left[10^{\frac{\alpha b_{s,q}}{q}}\left(\left(\tfrac{10^{\frac{\alpha}{q}}-1}{10^{\frac{1}{q}}-1}\right)10^{(-b_{s,q}/q)}(s-10^{b_{s,q}/q})+1\right)\right] \\
&= \tfrac{b_{s,q}}{q}+\tfrac{1}{\alpha}\log_{10}\left(\left(\tfrac{10^{\frac{\alpha}{q}}-1}{10^{\frac{1}{q}}-1}\right)10^{(-b_{s,q}/q)}(s-10^{b_{s,q}/q})+1\right) \\
\text{ดังนั้น } G_{\alpha,q}(d+1) &= \tfrac{b_{d+1,q}}{q}+\tfrac{1}{\alpha}\log_{10}\left(\left(\tfrac{10^{\frac{\alpha}{q}}-1}{10^{\frac{1}{q}}-1}\right)10^{(-b_{d+1,q}/q)}(d+1-10^{b_{d+1,q}/q})+1\right)
\end{aligned}$$

จึงได้ว่า $g_{\alpha,q}(d) = \frac{b_{d+1,q}-b_{d,q}}{q} + \frac{1}{\alpha}\log_{10}\left(\frac{\left(\frac{10^{\frac{\alpha}{q}}-1}{10^{\frac{1}{q}}-1}\right)10^{(-b_{d+1,q}/q)}\left(d+1-10^{(b_{d+1,q}/q)}\right)+1}{\left(\frac{10^{\frac{\alpha}{q}}-1}{10^{\frac{1}{q}}-1}\right)10^{(-b_{d,q}/q)}\left(d-10^{(b_{d,q}/q)}\right)+1}\right)$

$= \frac{b_{d+1,q}-b_{d,q}}{q} + \frac{1}{\alpha \ln(10)}[\ \ln((10^{\frac{\alpha}{q}}-1)a_{d+1}+1) - \ln((10^{\frac{\alpha}{q}}-1)a_d+1)\ ]$

เมื่อ $a_d = \frac{10^{(-b_{d,q}/q)}(d-10^{b_{d,q}/q})}{10^{\frac{1}{q}}-1} = \frac{d10^{(-b_{d,q}/q)}-1}{10^{\frac{1}{q}}-1}$

ต่อมาเราจะพิจารณาผลคูณของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d_i)$ โดยที่ $d_i = 1,2,3,...,9$ และ $i \in \{1,2,3,...,n\}$ จะได้ว่า

$$\begin{aligned}\prod_{i=1}^{n} g_{\alpha,q}(d_i) &= g_{\alpha,q}(d_1)g(d_2)\cdots g_{\alpha,q}(d_n)\\ &= g_{\alpha,q}(1)^{N_1}g_{\alpha,q}(2)^{N_2}\cdots g_{\alpha,q}(9)^{N_9}\\ &= \prod_{d=1}^{9} g_{\alpha,q}(d)^{N_d}\end{aligned}$$

เมื่อ $N_d$ คือความถี่ข้อมูลที่มีเลขหลักแรกเป็น $d$
ดังนั้น เมื่อให้ $L(\alpha)$ เป็นฟังก์ชันความควรจะเป็น จะได้ว่า

$$\begin{aligned}L(\alpha) &= \prod_{i=1}^{n} g_{\alpha,q}(d_i) = \prod_{d=1}^{9} g_{\alpha,q}(d)^{N_d}\\ \text{ฉะนั้น } \ln L(\alpha) &= \sum_{d=1}^{9} N_d \ln\left(g_{\alpha,q}(d)\right)\\ \text{และ } \frac{d\ln L(\alpha)}{d\alpha} &= \sum_{d=1}^{9} N_d \frac{1}{g_{\alpha,q}(d)}\frac{d(g_{\alpha,q}(d))}{d\alpha}\end{aligned}$$

พิจารณา $\frac{dg_{\alpha,q}(d))}{d\alpha}$ จะได้ว่า

$$\begin{aligned}\frac{dg_{\alpha,q}(d)}{d\alpha} &= 0+\frac{1}{\alpha\ln(10)}\left(\frac{a_{d+1}\ln(10)10^{\frac{\alpha}{q}}}{\left((10^{\frac{\alpha}{q}}-1)a_{d+1}+1\right)q}-\frac{a_d\ln(10)10^{\frac{\alpha}{q}}}{\left((10^{\frac{\alpha}{q}}-1)a_d+1\right)q}\right)-\frac{1}{\alpha^2}\log_{10}\left(\frac{\left(10^{\frac{\alpha}{q}}-1\right)a_{d+1}+1}{\left(10^{\frac{\alpha}{q}}-1\right)a_d+1}\right)\\ &= \frac{10^{\frac{\alpha}{q}}}{\alpha q}\left(\frac{a_{d+1}}{(10^{\frac{\alpha}{q}}-1)a_{d+1}+1}-\frac{a_d}{(10^{\frac{\alpha}{q}}-1)a_d+1}\right)-\frac{1}{\alpha^2}\log_{10}\left(\frac{\left(10^{\frac{\alpha}{q}}-1\right)a_{d+1}+1}{\left(10^{\frac{\alpha}{q}}-1\right)a_d+1}\right)\end{aligned}$$

และจัดรูป $g_{\alpha,q}(d)$

$$\begin{aligned}g_{\alpha,q}(d) &= \frac{b_{d+1,q}-b_{d,q}}{q}+\frac{1}{\alpha\ln(10)}[\,\ln((10^{\frac{\alpha}{q}}-1)a_{d+1}+1)-\ln((10^{\frac{\alpha}{q}}-1)a_d+1)\,]\\ &= \frac{b_{d+1,q}-b_{d,q}}{q}+\frac{1}{\alpha}\log_{10}\left(\frac{\left(10^{\frac{\alpha}{q}}-1\right)a_{d+1}+1}{\left(10^{\frac{\alpha}{q}}-1\right)a_d+1}\right)\end{aligned}$$

จึงได้ว่า

$$\frac{d\ln L(\alpha)}{d\alpha}=\sum_{d=1}^{9}N_d\frac{\frac{10^{\frac{\alpha}{q}}}{\alpha q}\left(\frac{a_{d+1}}{(10^{\frac{\alpha}{q}}-1)a_{d+1}+1}-\frac{a_d}{(10^{\frac{\alpha}{q}}-1)a_d+1}\right)-\frac{1}{\alpha^2}\log_{10}\left(\frac{\left(10^{\frac{\alpha}{q}}-1\right)a_{d+1}+1}{\left(10^{\frac{\alpha}{q}}-1\right)a_d+1}\right)}{\frac{b_{d+1,q}-b_{d,q}}{q}+\frac{1}{\alpha}\log_{10}\left(\frac{\left(10^{\frac{\alpha}{q}}-1\right)a_{d+1}+1}{\left(10^{\frac{\alpha}{q}}-1\right)a_d+1}\right)}$$

โดยวิธีความควรจะเป็นสูงสุด เราต้องการหาตัวประมาณของพารามิเตอร์ $\alpha$ ที่ทำให้ฟังก์ชันความควรจะเป็นนั่นคือ $L(\alpha)$ มีค่าสูงสุด กล่าวคือเราต้องหาตัวประมาณของพารามิเตอร์ $\alpha$ ที่ทำให้ $\frac{dL(\alpha)}{d\alpha}=0$ แต่เนื่องจากฟังก์ชันลอการิทึมเป็นฟังก์ชันเพิ่มและเพื่อความสะดวกในการคำนวณเราจึงพิจารณาตัวประมาณของพารามิเตอร์ $\alpha$ ที่ทำให้ $\frac{d\ln L(\alpha)}{d\alpha}=0$ แทน ดังนั้นเราจึงจะหาตัวประมาณของพารามิเตอร์ $\alpha$ จากสมการ

$$\sum_{d=1}^{9}N_d\frac{\frac{10^{\frac{\alpha}{q}}}{\alpha q}\left(\frac{a_{d+1}}{(10^{\frac{\alpha}{q}}-1)a_{d+1}+1}-\frac{a_d}{(10^{\frac{\alpha}{q}}-1)a_d+1}\right)-\frac{1}{\alpha^2}\log_{10}\left(\frac{\left(10^{\frac{\alpha}{q}}-1\right)a_{d+1}+1}{\left(10^{\frac{\alpha}{q}}-1\right)a_d+1}\right)}{\frac{b_{d+1,q}-b_{d,q}}{q}+\frac{1}{\alpha}\log_{10}\left(\frac{\left(10^{\frac{\alpha}{q}}-1\right)a_{d+1}+1}{\left(10^{\frac{\alpha}{q}}-1\right)a_d+1}\right)}=0 \qquad (2.1)$$

เราจะใช้คำสั่ง $uniroot$ ของโปรแกรม $R$ ที่หารากของสมการโดยใช้วิธีแบ่งครึ่งช่วง (bisection method) ในการคำนวณหาตัวประมาณของพารามิเตอร์ $\alpha$ ซึ่งถือเป็นรากของสมการ (2.1) โดยจะแสดงการคำนวณไว้ในภาคผนวก ข

เราสามารถหาตัวประมาณของพารามิเตอร์ $\alpha$ สำหรับแต่ละข้อมูลได้โดยการแทนค่า $N_d$ ของ

แต่ละข้อมูล และค่า $q$ ที่อยู่ในช่วงที่เราจะทำการเลือกอย่างเหมาะสมในบทถัดไป ลงในสมการ (2.1) และใช้คำสั่ง $uniroot$ ของโปรแกรม $R$ ในการคำนวณหาตัวประมาณของพารามิเตอร์ $\alpha$

# บทที่ 3
# การเลือกช่วงของ $q$ ที่เหมาะสม

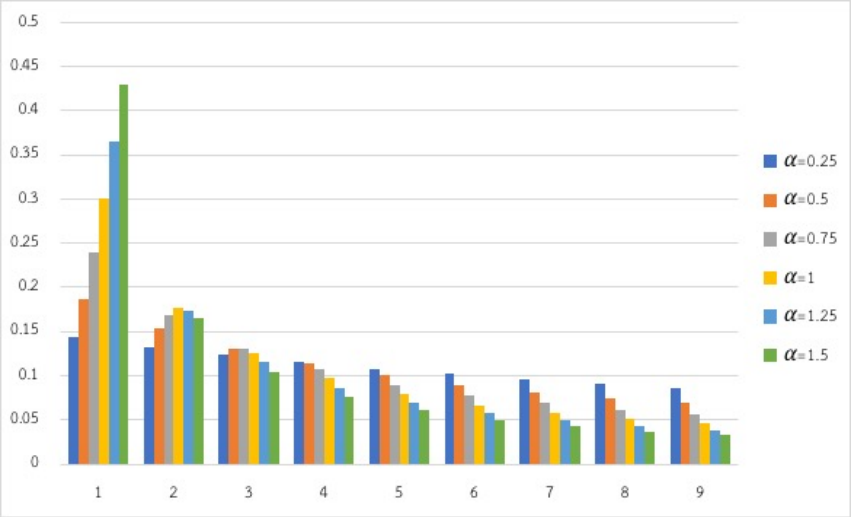
ในหัวข้อนี้เราจะทำการเลือกพิจารณา $q$ ในช่วง ๆ หนึ่ง เพื่อความสะดวกในการนำค่าประมาณ $\alpha$ และ $q$ ในช่วงจำนวนนับนี้ไปทดสอบความเป็นเบนฟอร์ดของข้อมูลที่มีปริมาณมาก

เนื่องจากเราทราบอยู่แล้วว่า เมื่อ $q \to \infty$ การแจกแจงของเลขหลักแรกที่มีฟังก์ชันการแจกแจง $G_{\alpha,q}(s)$ จะลู่เข้าสู่กฎของเบนฟอร์ด เราจึงจะทำการเลือกช่วงของ $q$ โดยการพิจารณาค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $q$ และ $\alpha$ ที่มีค่าใกล้ ๆ 1 เพื่อสังเกตว่าค่าของ $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ มีการเปลี่ยนแปลงอย่างไรเมื่อ $q$ มีค่ามากขึ้น มากไปกว่านั้นเราได้พิจารณาค่าลอการิทึมของฟังก์ชันความจะเป็นของข้อมูลจริงที่มีการแจกแจงสอดคล้องกับกฎของเบนฟอร์ด เพื่อสังเกตดูว่าหากข้อมูลมีการแจกแจงสอดคล้องกับกฎของเบนฟอร์ดแล้ว ค่าสูงสุดของฟังก์ชันความน่าจะเป็นมักจะเกิดขึ้นที่ $q$ ในช่วงใด

## 3.1 ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}$ สำหรับแต่ละ $\alpha$ และ $q$

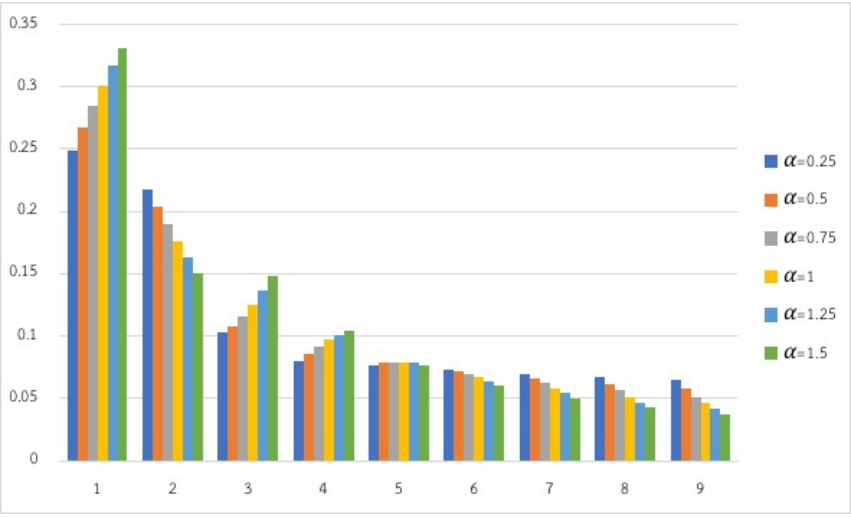
ในหัวข้อนี้เราจะใช้โปรแกรม R ในการคำนวณหาค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $q$ โดยที่ $q$ เป็นจำนวนนับในช่วง 1 ถึง 10 และ $\alpha$ ที่มีค่าใกล้ๆ 1 ได้แก่ $\alpha = 0.25, 0.5, 0.75, 1, 1.25$ และ $1.5$ ตามลำดับ เพื่อทำการสังเกตว่าค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ มีการเปลี่ยนแปลงอย่างไรเมื่อค่า $q$ มีค่าสูงขึ้นตามลำดับ โดยการสร้างกราฟเปรียบเทียบค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $q$ และ $\alpha$ ที่มีเลขโดดหลักแรก $d$ ซึ่ง $d = 1, 2, 3, ..., 9$ เป็นแกนนอน และค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $q$ และ $\alpha$ เป็นแกนตั้ง จะได้ผลลัพธ์ออกมาดังนี้

### 3.1.1 กรณี $q = 1$




รูปที่ 3.1: ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 1$

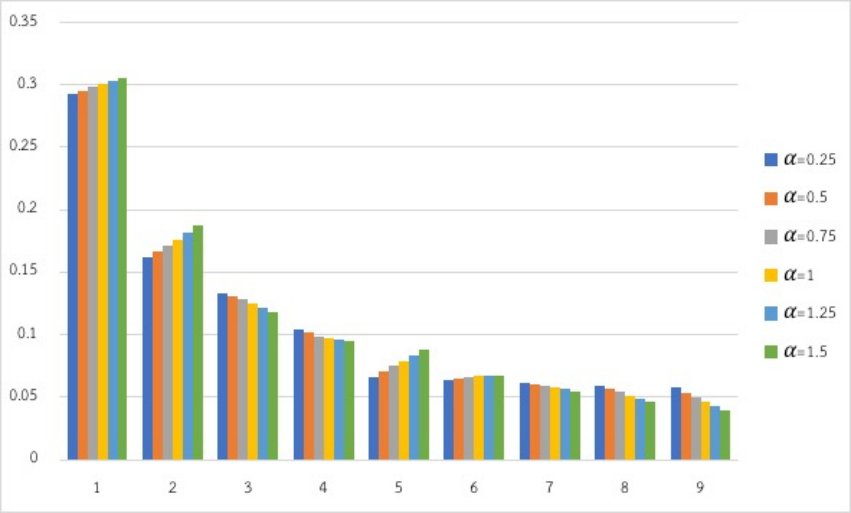
### 3.1.2 กรณี $q = 2$




รูปที่ 3.2: ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 2$

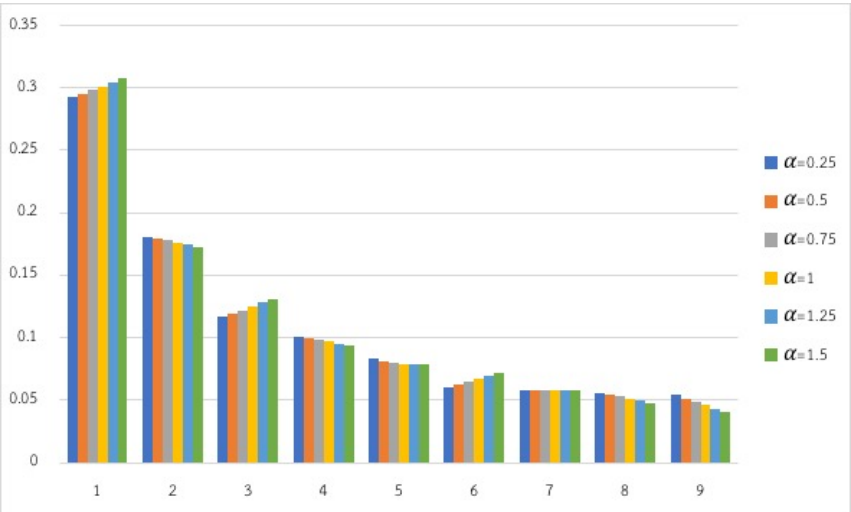
จะเห็นได้ว่าค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ ในกรณีที่ $q = 1$ และ 2 สำหรับแต่ละ $\alpha$ มีค่าแตกต่างกันมาก

### 3.1.3 กรณี $q = 3$




รูปที่ 3.3: ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 3$

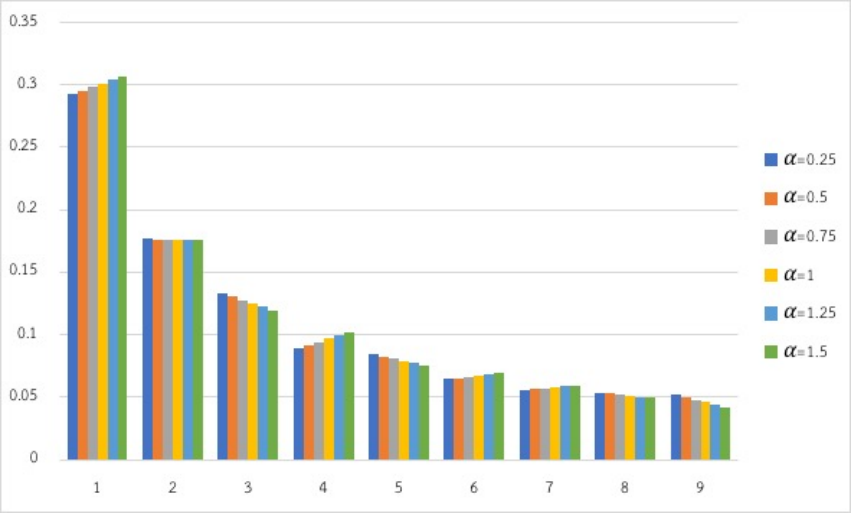
### 3.1.4 กรณี $q = 4$




รูปที่ 3.4: ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 4$

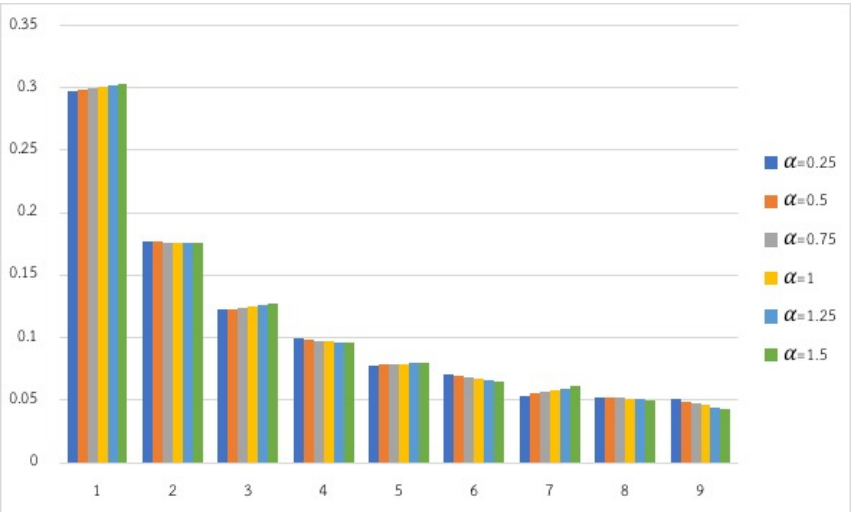
จะเห็นได้ว่าค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ ในกรณีที่ $q = 3$ ยังคงมีค่าแตกต่างกันมาก ยกเว้นในกรณีที่เลขโดดหลักแรก $d = 6$ และในกรณีที่ $q = 4$ ที่เลขโดดหลักแรก $d = 5$ และ 7 ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ มีค่าใกล้เคียงกันมาก

### 3.1.5 กรณี $q = 5$




รูปที่ 3.5: ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 5$

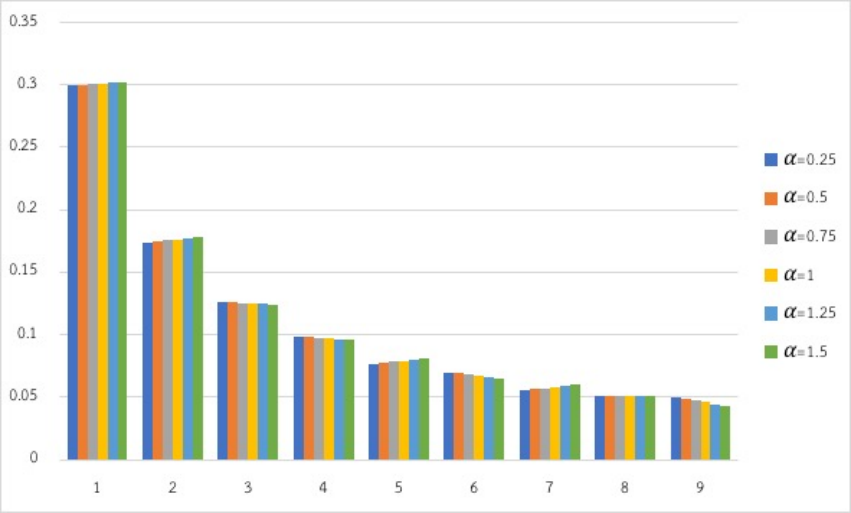
### 3.1.6 กรณี $q = 6$




รูปที่ 3.6: ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 6$

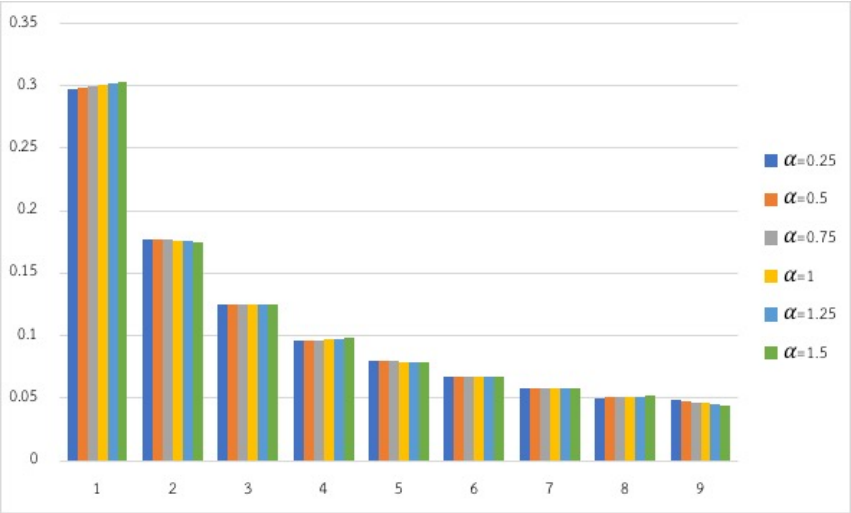
จะเห็นได้ว่าค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 5$ และเลขโดดหลักแรก $d = 3, 4, 5$ ยังคงมีค่าแตกต่างกันมาก แต่ในกรณีที่ $q = 6$ ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ มีค่าใกล้เคียงกันมาก

### 3.1.7 กรณี $q = 7$




รูปที่ 3.7: ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 7$

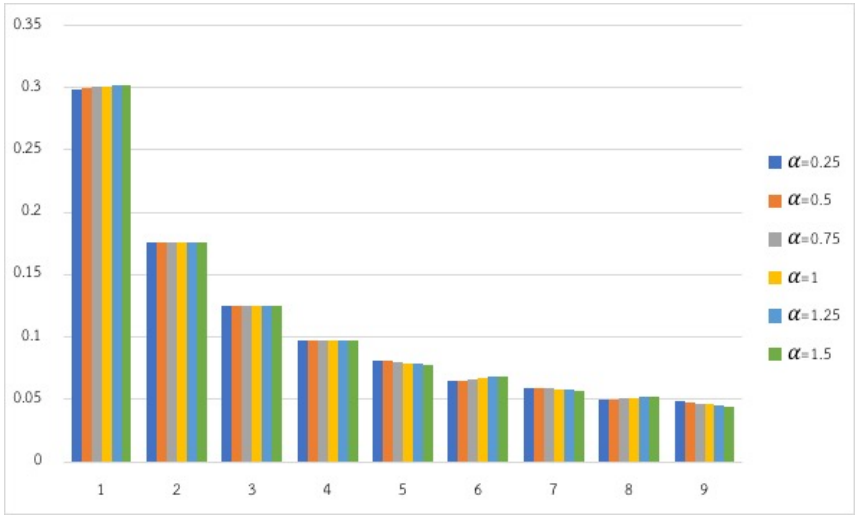
### 3.1.8 กรณี $q = 8$




รูปที่ 3.8: ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 8$

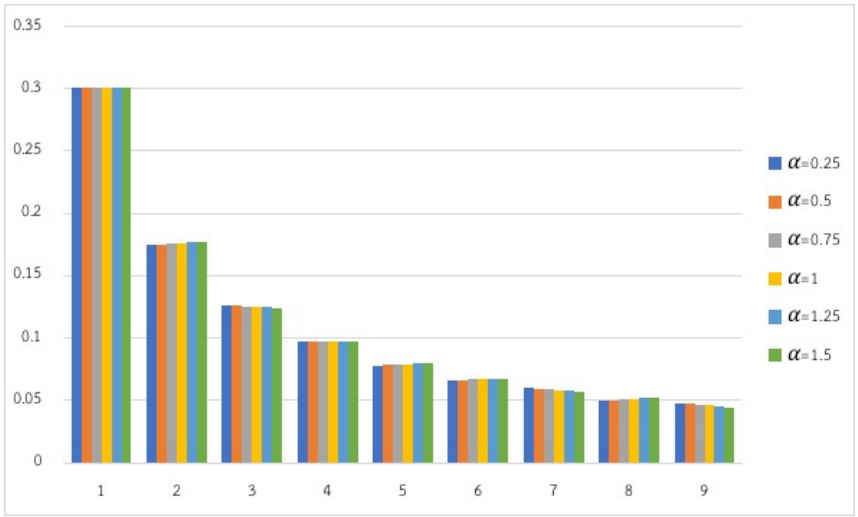
จะเห็นได้ว่าค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 7$ มีความใกล้เคียงกันมาก ส่วนในกรณีที่ $q = 8$ ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ แทบจะไม่แตกต่างกันเลย

### 3.1.9 กรณี $q = 9$




รูปที่ 3.9: ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 9$

### 3.1.10 กรณี $q = 10$




รูปที่ 3.10: ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 10$

จะเห็นได้ว่าค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับแต่ละ $\alpha$ ในกรณีที่ $q = 9$ และ 10 มีค่าใกล้เคียงกันมาก แทบจะเท่ากันเลย

จากการพิจาณาค่าของ $g_{\alpha,q}(d)$ โดยที่ $\alpha$ มีค่าใกล้ ๆ 1 นั่นคือ $\alpha = 0.25, 0.5, 0.75, 1.00, 1.25$ และ $1.50$ ตามลำดับ สำหรับแต่ละ $q$ ที่เป็นจำนวนนับในช่วง 1 ถึง 10 จะสังเกตเห็นได้ว่าใน

กรณีที่ $q$ มีค่าตั้งแต่ 6 ขึ้นไป ค่า $g_{\alpha,q}(d)$ สำหรับ $\alpha$ แต่ละตัวที่เราพิจารณามีค่าแตกต่างกันน้อยมาก แทบจะเท่ากันเลย

ต่อไปเราจึงจะพิจารณาค่าลอกอริทึมของฟังก์ชันความควรจะเป็นของข้อมูลที่มีการแจกแจงสอดคล้องกับกฎของเบนฟอร์ด เพื่อสังเกตว่าค่าของฟังก์ชันความควรจะเป็นสูงสุดจะเกิดขึ้นในกรณีที่ $q$ มีค่ามากกว่า 5 หรือไม่

## 3.2 พิจารณาค่าลอกอริทึมของฟังก์ชันความควรจะเป็นของข้อมูลที่มีการแจกแจงสอดคล้องกับกฎของเบนฟอร์ด

เนื่องจากลอการิทึมเป็นฟังก์ชันเพิ่ม เราจึงทำการคำนวณหาค่าลอการิทึมของฟังก์ชันความควรจะเป็นสูงสุดของข้อมูลที่มีการแจกแจงสอดคล้องกับกฎของเบนฟอร์ด เพื่อพิจารณาว่าค่าของฟังก์ชันความควรจะเป็นสูงสุดสำหรับแต่ละ $\alpha$ และ $q$ จะเกิดค่าสูงสุดในกรณีที่ $\alpha$ และ $q$ เป็นเท่าใด

โดยการแทนค่าจำนวนของข้อมูลที่มีเลขหลักแรกเป็น $d$ เมื่อ $d = 1, 2, ..., 9$ ของข้อมูลที่เรานำมาพิจารณาได้แก่ ข้อมูลตามกฎของเบนฟอร์ด ข้อมูลที่อยู่อาศัย ข้อมูลประชากร ข้อมูลพื้นที่ของแม่น้ำแต่ละแห่งทั่วโลก และข้อมูลค่าพลังงานของการแผ่รังสีของวัตถุดำ ลงในฟังก์ชันลอการิทึมของฟังก์ชันความควรจะเป็นสูงสุด

กล่าวคือ เมื่อให้ $N_d$ เป็นจำนวนข้อมูลที่มีเลขหลักแรกเป็น $d$ โดยที่ $d = 1, 2, ..., 9$ เราจะทำการแทนค่า $N_d$ สำหรับแต่ละ $d$ ของแต่ละข้อมูล ลงใน

$$\ln L(\alpha) = \sum_{d=1}^{9} N_d \ln(g(d))$$

ทำให้แต่ละข้อมูลได้ผลลัพธ์ออกมาดังนี้

### 3.2.1 ข้อมูลตามกฎของเบนฟอร์ด

จากกฎของเบนฟอร์ด จะได้จำนวนของข้อมูลที่มีเลขหลักแรกเป็น $d$ นั่นคือ $N_d$ สำหรับแต่ละ $d$ ดังตาราง

| $d$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| $N_d$ | 30.103 | 17.6091 | 12.4939 | 9.691 | 7.91812 | 6.69468 | 5.79919 | 5.11525 | 4.57575 |

ตารางที่ 3.1: จำนวนข้อมูลที่มีเลขหลักแรกเป็น $d$ ของข้อมูลตามกฎของเบนฟอร์ด

เมื่อแทนค่า $q$ ตั้งแต่ 1 ถึง 10, $N_d$ ของข้อมูลตามกฎของเบนฟอร์ด สำหรับแต่ละ $d$ และค่าประมาณ $\alpha$ ที่คำนวณได้จากสมการ (2.1) โดยโปรแกรม R ลงในฟังก์ชัน $\ln L(\alpha)$ จะได้ผลลัพธ์ดังตาราง

| $q$ | $\alpha$ | $\ln L(\alpha)$ |
|---|---|---|
| 1 | 1 | -199.3433 |
| 2 | 1.000003 | -199.3433 |
| 3 | 0.9999961 | -199.3433 |
| 4 | 1.000006 | -199.3433 |
| 5 | 0.9999961 | -199.3433 |
| 6 | 1.000002 | -199.3433 |
| 7 | 0.9999897 | -199.3433 |
| 8 | 1.000004 | -199.3433 |
| 9 | 1.000005 | -199.3433 |
| 10 | 0.9999862 | -199.3433 |

ตารางที่ 3.2: ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลตามกฎของเบนฟอร์ด

จากตารางเราจะได้ว่าค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับทุก $q$ มีค่าเท่ากันที่ทศนิยม 4 ตำแหน่ง นั่นคือเท่ากับ -199.3433 และค่าประมาณ $\alpha$ สำหรับทุก $q$ มีค่าเท่ากันที่ทศนิยม 6 ตำแหน่ง นั่นคือเท่ากับ 1

ต่อไปเราจะทำการคำนวณหาค่าประมาณของ $\alpha$ และค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับ $q$ แต่ละกรณี ของแต่ละตัวอย่างข้อมูลที่มีการแจกแจงเลขหลักแรกสอดคล้องกับกฎของเบนฟอร์ด

### 3.2.2 ข้อมูลที่อยู่อาศัย

จากตารางข้อมูลใน [6] ทำให้สามารถคำนวณหาจำนวนของข้อมูลที่มีเลขหลักแรกเป็น $d$ นั่นคือ $N_d$ สำหรับแต่ละ $d$ ดังตาราง

| $d$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| $N_d$ | 988.38 | 656.64 | 465.12 | 300.96 | 2684.7 | 218.88 | 191.52 | 171 | 171 |

ตารางที่ 3.3: จำนวนข้อมูลที่มีเลขหลักแรกเป็น $d$ ของข้อมูลที่อยู่อาศัย

เมื่อแทนค่า $q$ ตั้งแต่ 1 ถึง 10, $N_d$ ของข้อมูลที่เป็นตัวอย่างสุ่มของข้อมูลที่อยู่อาศัย สำหรับแต่ละ $d$ และค่าประมาณ $\alpha$ ที่คำนวณได้จากสมการ (2.1) โดยโปรแกรม R ลงในฟังก์ชัน $\ln L(\alpha)$ จะได้ผลลัพธ์ดังตาราง

| $q$ | $\alpha$ | $\ln L(\alpha)$ |
|---|---|---|
| 1 | 0.4265987 | -199.8719 |
| 2 | 0.8842688 | -200.1218 |
| 3 | 3.081027 | -200.3792 |
| 4 | 1.809717 | -201.2523 |
| 5 | ไม่แสดงผลลัพธ์ในโปรแกรม R | |
| 6 | 2.000511 | -14666.37 |
| 7 | 1.175658 | -200.5607 |
| 8 | 3.446191 | -16893.9 |
| 9 | ไม่แสดงผลลัพธ์ในโปรแกรม R | |
| 10 | ไม่แสดงผลลัพธ์ในโปรแกรม R | |

ตารางที่ 3.4: ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลที่อยู่อาศัย

จากตารางจะได้ว่าโปรแกรม R ไม่แสดงผลลัพธ์ของค่าประมาณ $\alpha$ และค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ ในกรณีที่ $q = 5, 9$ และ 10 เนื่องจากเราพิจารณาค่าประมาณของ $\alpha$ ที่มีค่ามากกว่าศูนย์ และสามารถเลือกค่าขอบเขตบนของตัวประมาณ $\alpha$ ที่เราต้องการได้ โดยการระบุไว้ในส่วนของการคำนวณโดยโปรแกรม R ฉะนั้นอาจเป็นไปได้ว่าค่าประมาณของ $\alpha$ สำหรับ $q$ แต่ละกรณีดังกล่าวไม่ได้อยู่ในช่วงที่เรากำหนด เราจึงจะไม่พิจารณาค่าประมาณของ $\alpha$ และค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ ในกรณีที่ $q = 5, 9$ และ 10

จากตารางจึงได้ว่าค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ มีค่าสูงสุดเท่ากับ -199.8719 ที่กรณี $q = 1$ และ $\alpha = 0.4265987$

### 3.2.3 ข้อมูลประชากร

จากตารางข้อมูลใน [6] ทำให้สามารถคำนวณหาจำนวนของข้อมูลที่มีเลขหลักแรกเป็น $d$ นั่นคือ $N_d$ สำหรับแต่ละ $d$ ดังตาราง

| $d$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| $N_d$ | 1104.801 | 664.836 | 462.778 | 263.979 | 234.648 | 202.058 | 133.619 | 120.583 | 71.698 |

ตารางที่ 3.5: จำนวนข้อมูลที่มีเลขหลักแรกเป็น $d$ ของข้อมูลที่อยู่อาศัย

เมื่อแทนค่า $q$ ตั้งแต่ 1 ถึง 10, $N_d$ ของข้อมูลที่เป็นตัวอย่างสุ่มของข้อมูลประชากร สำหรับแต่ละ $d$ และค่าประมาณ $\alpha$ ที่คำนวณได้จากสมการ (2.1) โดยโปรแกรม R ลงในฟังก์ชัน $\ln L(\alpha)$ จะได้ผลลัพธ์ดังตาราง

| $q$ | $\alpha$ | $\ln L(\alpha)$ |
|---|---|---|
| 1 | 1.209473 | -6134.403 |
| 2 | 1.355776 | -6145.846 |
| 3 | 1.88017 | -6145.846 |
| 4 | 2.500862 | -6651.571 |
| 5 | 1.605275 | -6209.364 |
| 6 | 4.571258 | -7661.623 |
| 7 | 3.179336 | -7661.623 |
| 8 | 4.445957 | -7610.056 |
| 9 | 5.994532 | -8178.239 |
| 10 | 6.221491 | -8178.239 |

ตารางที่ 3.6: ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลประชากร

จากตารางจะได้ว่าค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ มีค่าสูงสุดเท่ากับ -6134.403 ที่กรณี $q = 1$ และ $\alpha = 1.209473$

### 3.2.4 ข้อมูลพื้นที่ของแม่น้ำแต่ละแห่งทั่วโลก

จากตารางข้อมูลใน [6] ทำให้สามารถคำนวณหาจำนวนของข้อมูลที่มีเลขหลักแรกเป็น $d$ นั่นคือ $N_d$ สำหรับแต่ละ $d$ ดังตาราง

เมื่อแทนค่า $q$ ตั้งแต่ 1 ถึง 10, $N_d$ ของข้อมูลที่เป็นตัวอย่างสุ่มของข้อมูลพื้นที่ของแม่น้ำแต่ละแห่งทั่วโลก สำหรับแต่ละ $d$ และค่าประมาณ $\alpha$ ที่คำนวณได้จากสมการ (2.1) โดยโปรแกรม

| d | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| $N_{(d)}$ | 103.85 | 54.94 | 35.845 | 37.855 | 24.12 | 28.81 | 18.425 | 14.07 | 17.085 |

ตารางที่ 3.7: จำนวนข้อมูลที่มีเลขหลักแรกเป็น $d$ ของข้อมูลพื้นที่แม่น้ำแต่ละแห่งทั่วโลก

R ลงในฟังก์ชัน ln $L(\alpha)$ จะได้ผลลัพธ์ดังตาราง

| $q$ | $\alpha$ | ln $L(\alpha)$ |
|---|---|---|
| 1 | 1.004459 | -669.0577 |
| 2 | 1.021995 | -669.0757 |
| 3 | 0.9589126 | -669.1812 |
| 4 | 1.131923 | -669.9771 |
| 5 | 2.117064 | -715.9303 |
| 6 | ไม่แสดงผลลัพธ์ในโปรแกรม R | |
| 7 | ไม่แสดงผลลัพธ์ในโปรแกรม R | |
| 8 | 0.9803607 | -669.0921 |
| 9 | 1.779086 | -695.4806 |
| 10 | ไม่แสดงผลลัพธ์ในโปรแกรม R | |

ตารางที่ 3.8: ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลพื้นที่ของแม่น้ำแต่ละแห่งทั่วโลก

จากตารางจะได้ว่าโปรแกรม R ไม่แสดงผลลัพธ์ของค่าประมาณ $\alpha$ และค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ ในกรณีที่ $q = 6, 7$ และ $10$ เนื่องจากเราพิจารณาค่าประมาณของ $\alpha$ ที่มีค่ามากกว่าศูนย์ และสามารถเลือกค่าขอบเขตบนของตัวประมาณ $\alpha$ ที่เราต้องการได้ โดยการระบุไว้ในส่วนของการคำนวณโดยโปรแกรม R ฉะนั้นอาจเป็นไปได้ว่าค่าประมาณของ $\alpha$ สำหรับ $q$ แต่ละกรณีดังกล่าวไม่ได้อยู่ในช่วงที่เรากำหนด เราจึงจะไม่พิจารณาค่าประมาณของ $\alpha$ และค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ ในกรณีที่ $q = 6, 7$ และ $10$

จากตารางจะได้ว่าค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ มีค่าสูงสุดเท่ากับ -669.0577 ที่กรณี $q = 1$ และ $\alpha = 1.004459$

### 3.2.5 ข้อมูลค่าพลังงานของการแผ่รังสีของวัตถุดำ

จากตารางข้อมูลใน [6] ทำให้สามารถคำนวณหาจำนวนของข้อมูลที่มีเลขหลักแรกเป็น $d$ นั่นคือ $N_d$ สำหรับแต่ละ $d$ ดังตาราง

| $d$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| $N_d$ | 361.15 | 201.545 | 164.265 | 101.355 | 76.89 | 81.55 | 60.58 | 54.755 | 62.91 |

ตารางที่ 3.9: จำนวนข้อมูลที่มีเลขหลักแรกเป็น $d$ ของข้อมูลค่าพลังงานของการแผ่รังสีของวัตถุดำ

เมื่อแทนค่า $q$ ตั้งแต่ 1 ถึง 10, $N_d$ ของข้อมูลที่เป็นตัวอย่างสุ่มของค่าพลังงานของการแผ่รังสีของวัตถุดำ สำหรับแต่ละ $d$ และค่าประมาณ $\alpha$ ที่คำนวณได้จากสมการ (2.1) โดยโปรแกรม R ลงในฟังก์ชัน $\ln L(\alpha)$ จะได้ผลลัพธ์ดังตาราง

| $q$ | $\alpha$ | $\ln L(\alpha)$ |
|---|---|---|
| 1 | 1.037141 | -2306.356 |
| 2 | 1.08893 | -2306.905 |
| 3 | 0.7165661 | -2329.478 |
| 4 | 1.362301 | -2326.051 |
| 5 | 0.5269492 | -2366.65 |
| 6 | 0.3952591 | -2403.023 |
| 7 | ไม่แสดงผลลัพธ์ในโปรแกรม R | |
| 8 | 0.05399579 | -2534.587 |
| 9 | 1.368742 | -2326.789 |
| 10 | ไม่แสดงผลลัพธ์ในโปรแกรม R | |

ตารางที่ 3.10: ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลค่าพลังงานของการแผ่รังสีของวัตถุดำ

จากตารางจะได้ว่าโปรแกรม R ไม่แสดงผลลัพธ์ของค่าประมาณ $\alpha$ และค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ ในกรณีที่ $q = 7$ และ 10 เนื่องจากเราพิจารณาค่าประมาณของ $\alpha$ ที่มีค่ามากกว่าศูนย์ และสามารถเลือกค่าขอบเขตบนของตัวประมาณ $\alpha$ ที่เราต้องการได้ โดยการระบุไว้ในส่วนของการคำนวณโดยโปรแกรม R ฉะนั้นอาจเป็นไปได้ว่าค่าประมาณของ $\alpha$ สำหรับ $q$ แต่ละกรณีดังกล่าวไม่ได้อยู่ในช่วงที่เรากำหนด เราจึงจะไม่พิจารณาค่าประมาณของ $\alpha$ และค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ ในกรณีที่ $q = 7$ และ 10

จากตารางจะได้ว่าค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ มีค่าสูงสุดเท่ากับ -2306.356 ที่กรณี $q = 1$ และ $\alpha = 1.037141$

จากการพิจารณาค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ โดยที่ $\alpha$ ที่มีค่าใกล้ ๆ 1 นั่นคือ $\alpha = 0.25, 0.5, 0.75, 1.00, 1.25$ และ $1.50$ ตามลำดับ สำหรับแต่ละ $q$ ที่เป็นจำนวนนับในช่วง 1 ถึง 10 จะเห็นได้ว่าในกรณีที่ $q$ มีค่าตั้งแต่ 6 ขึ้นไป ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(d)$ สำหรับ $\alpha$ แต่ละตัวที่เราพิจารณามีค่าแตกต่างกันน้อยมาก ประกอบกับการที่เรานำกลุ่มข้อมูลตัวอย่างที่มีการแจกแจงเลขหลักแรกสอดคล้องกับกฎของเบนฟอร์ด จะเห็นได้ว่าค่าสูงสุดของ

ฟังก์ชันความควรจะเป็น ไม่ได้เกิดขึ้นในกรณีที่ $q$ มีค่ามากกว่า 5

ดังนั้นเราจึงเลือกพิจารณาค่า $q$ ในช่วงจำนวนนับ 1 ถึง 5 เท่านั้น

# บทที่ 4
# ทดสอบความเป็นเบนฟอร์ดของข้อมูลค่าไฟฟ้า

ในโครงงานนี้สนใจทดสอบความเป็นเบนฟอร์ดของข้อมูลค่าไฟฟ้ารายเดือน จากการวิเคราะห์ข้อมูลที่ได้จากฝ่ายเศรษฐกิจและพลังงานไฟฟ้า การไฟฟ้านครหลวง โดยนางสาวธัญญาดา คงพริ้ว ซึ่งแสดงการวิเคราะห์ไว้ใน [4] ทำให้เราทราบว่ามีจำนวนผู้ใช้ไฟฟ้าในบริเวณกรุงเทพมหานคร และปริมณฑล ตั้งแต่เดือนกุมภาพันธ์ถึงเดือนธันวาคม ในปีพ.ศ. 2557 มีประมาณ 340,000 ราย ซึ่งสามารถแบ่งตามการใช้งานได้หลายประเภท ประกอบไปด้วย ที่อยู่อาศัย กิจการขนาดเล็ก กิจการขนาดกลาง กิจการขนาดใหญ่ กิจการเฉพาะอย่าง ส่วนราชการ และส่วนที่มีการใช้ไฟฟ้าชั่วคราว โดยแต่ละประเภทของผู้ใช้จะมีปริมาณการใช้ไฟฟ้าที่แตกต่างกัน หากมีปริมาณการใช้ไฟฟ้าที่น้อย อย่างเช่น การใช้ไฟฟ้าในบ้านซึ่งเป็นที่อยู่อาศัยแต่ละครัวเรือน ก็จะส่งผลให้ค่าใช้จ่ายไฟฟ้ารายเดือนต่ำ เมื่อเทียบกับค่าใช้จ่ายไฟฟ้ารายเดือนของกิจการขนาดใหญ่ที่มีค่าใช้จ่ายไฟฟ้ารายเดือนสูง เนื่องจากมีปริมาณการใช้ไฟฟ้ามากในแต่ละเดือน ซึ่งหากการแจกแจงเลขหลักแรกของข้อมูลเหล่านี้ไม่เป็นไปตามกฎของเบนฟอร์ด ก็อาจจะเป็นข้อสังเกตหนึ่งที่แสดงให้เห็นถึงความผิดปกติของข้อมูลในเบื้องต้น และเป็นข้อสงสัยที่จะนำไปสู่การตรวจสอบในเชิงลึกต่อๆไป เนื่องจากการที่ข้อมูลมีปริมาณที่มาก ส่งผลให้เกิดความผิดพลาดในการจัดเก็บและการจัดการข้อมูลได้ง่าย อาจเกิดขึ้นโดยเจตนาหรือไม่เจตนา เช่น ความผิดพลาดจากการจดบันทึกข้อมูล ความผิดพลาดจากการคำนวณค่าไฟฟ้า หรือการดัดแปลงเครื่องสำหรับวัดปริมาณการใช้ไฟฟ้าซึ่งส่งผลให้ค่าไฟฟ้ารายเดือนต่ำแม้ว่าจะใช้ไฟฟ้ามาก ซึ่งถือเป็นการทุจริต เป็นต้น

ในการนำข้อมูลไฟฟ้ารายเดือน ในปีพ.ศ. 2557 ที่จะนำข้อมูลตั้งแต่เดือนกุมภาพันธ์ถึงเดือนธันวาคม มาใช้ในการทดสอบการแจกแจง เนื่องจากเกิดความผิดพลาดในการสุ่มตัวอย่างของข้อมูล

ในเดือนมกราคม ได้มีการแบ่งการศึกษาในแต่ละเดือนออกเป็น 2 ส่วนคือ ราคาค่าไฟฟ้าและราคาค่าไฟฟ้าที่เราทำการคูณค่าคงที่ 5 ที่หน่วยการใช้ไฟฟ้า และเรียกว่าราคาค่าไฟฟ้าหน่วยละ 5 บาท ซึ่งจะนำข้อมูลทั้ง 2 ส่วนนี้มาทดสอบว่ามีการแจกแจงของเลขหลักแรกเป็นไปตามกฎของเบนฟอร์ดหรือไม่ โดยการแทนค่า $N_d$ สำหรับแต่ละ $d$ ของแต่ละข้อมูล, $q$ ที่เป็นจำนวนนับตั้งแต่ 1 ถึง 5 และค่า $\alpha$ ที่ได้จากการประมาณค่า โดยการแทนค่า $N_d$ สำหรับแต่ละ $d$ ลงในฟังก์ชันความควรจะเป็น $L(\alpha)$ ลงในฟังก์ชันลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ นั่นคือ

$$\ln L(\alpha) = \sum_{d=1}^{9} N_d \ln(g_{\alpha,q}(d))$$

เมื่อ $N_d$ เป็นจำนวนข้อมูลที่มีเลขหลักแรกเป็น $d$ โดยที่ $d = 1, 2, ..., 9$ จากนั้นทำการสังเกตว่าค่า $q$ และค่าประมาณ $\alpha$ คู่ใดที่ทำให้ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ มีค่าสูงสุด

## 4.1 การหาค่าประมาณของพารามิเตอร์ $\alpha$ สำหรับข้อมูลค่าไฟฟ้าแต่ละเดือน

เราจะนำข้อมูลที่ได้จากการทดลองสุ่มข้อมูลราคาค่าไฟฟ้า และราคาค่าไฟฟ้าหน่วยละ 5 บาท ของแต่ละเดือนมาวิเคราะห์จำนวนของข้อมูลที่มีเลขหลักแรกเป็น 1 ถึง 9 ซึ่งคือ $N_d$ สำหรับแต่ละเลขโดดหลักแรก $d$ จากนั้นนำ $N_d$ ของข้อมูลแต่ละเดือนแทนค่าลงในสมการ (2.1) และใช้โปรแกรม R ในการคำนวณหาค่าประมาณของพารามิเตอร์ $\alpha$ แล้วพิจารณาค่า $\alpha$ ที่ทำให้ค่าของฟังก์ชันลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ มีค่าสูงสุด จะได้ค่าประมาณของพารามิเตอร์ $\alpha$ สำหรับข้อมูลค่าไฟฟ้าแต่ละเดือน ดังนี้

### 4.1.1 เดือนกุมภาพันธ์

| เลขโดดหลักแรก | ความถี่ที่ได้จากการทดลองเดือนกุมภาพันธ์ | | ความถี่ที่ได้จากกฎของเบนฟอร์ด |
|---|---|---|---|
| | ราคาค่าไฟฟ้า | ราคาค่าไฟฟ้าหน่วยละ 5 บาท | |
| 1 | 331 | 304 | 301 |
| 2 | 161 | 168 | 176 |
| 3 | 113 | 113 | 125 |
| 4 | 115 | 88 | 97 |
| 5 | 83 | 84 | 79 |
| 6 | 64 | 68 | 67 |
| 7 | 51 | 75 | 58 |
| 8 | 47 | 49 | 51 |
| 9 | 35 | 51 | 46 |
| รวม | 1,000 | 1,000 | 1,000 |

ตารางที่ 4.1: ความถี่ที่ได้จากการทดลองของเลขโดดหลักแรกในเดือนกุมภาพันธ์ ปี พ.ศ. 2557 จากการสุ่มจำนวน 1,000 ตัวอย่าง

จะเห็นว่าการแจกแจงความถี่ของเลขโดดหลักแรกของข้อมูลทั้ง 2 ชนิด มีความใกล้เคียงกับกฎของเบนฟอร์ดอย่างมาก

เมื่อนำความถี่ของข้อมูลที่มีเลขหลักแรกเป็น $d$ ซึ่งก็คือ $N_d$ ของแต่ละ $d$ โดยที่ $d = 1, 2, 3..., 9$ ของทั้งสองข้อมูล, ค่า $q$ ตั้งแต่ 1 ถึง 5 และค่าประมาณ $\alpha$ ที่คำนวณได้จากสมการ (2.1) โดยโปรแกรม R ลงในฟังก์ชัน $\ln L(\alpha)$ จะได้ผลลัพธ์ดังตาราง

| $q$ | ราคาค่าไฟฟ้า | | ราคาค่าไฟฟ้าหน่วยละ 5 บาท | |
|---|---|---|---|---|
| | $\alpha$ | $\ln L(\alpha)$ | $\alpha$ | $\ln L(\alpha)$ |
| 1 | 1.098468 | -195.5175 | 0.977861 | -201.0501 |
| 2 | 1.25294 | -195.9022 | 0.8732962 | -201.2475 |
| 3 | 1.392251 | -196.8362 | 0.9433783 | -201.0712 |
| 4 | 1.282207 | -196.0558 | 0.8241461 | -201.4814 |
| 5 | 1.988022 | -204.8185 | 0.9851678 | -201.0511 |

ตารางที่ 4.2: ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาท ในเดือนกุมภาพันธ์ พ.ศ. 2557

จากตารางจะได้ว่าค่า $\ln L(\alpha)$ ที่ $N_d$ สอดคล้องกับข้อมูลราคาค่าไฟและราคาค่าไฟฟ้าหน่วยละ 5 บาทมีค่าสูงสุดเท่ากับ -195.5175 และ -201.0501 ตามลำดับ โดยที่ $\alpha = 1.098468$ และ 0.977861 ตามลำดับ ที่กรณี $q = 1$ เช่นเดียวกัน

### 4.1.2 เดือนมีนาคม

| เลขโดดหลักแรก | ความถี่ที่ได้จากการทดลองเดือนมีนาคม | | ความถี่ที่ได้จากกฎของเบนฟอร์ด |
|---|---|---|---|
| | ราคาค่าไฟฟ้า | ราคาค่าไฟฟ้าหน่วยละ 5 บาท | |
| 1 | 323 | 304 | 301 |
| 2 | 184 | 185 | 176 |
| 3 | 116 | 130 | 125 |
| 4 | 103 | 78 | 97 |
| 5 | 80 | 72 | 79 |
| 6 | 74 | 71 | 67 |
| 7 | 46 | 72 | 58 |
| 8 | 35 | 51 | 51 |
| 9 | 39 | 37 | 46 |
| รวม | 1,000 | 1,000 | 1,000 |

ตารางที่ 4.3: ความถี่ที่ได้จากการทดลองของเลขโดดหลักแรกในเดือนมีนาคม ปี พ.ศ. 2557 จากการสุ่มจำนวน 1,000 ตัวอย่าง

จะเห็นว่าการแจกแจงความถี่ของเลขโดดหลักแรกของข้อมูลทั้ง 2 ชนิด มีความใกล้เคียงกับกฎของเบนฟอร์ด มีเพียงความถี่ของราคาค่าไฟฟ้าหน่วยละ 5 บาท ที่จะสังเกตเห็นความแตกต่างที่เลขโดด 4 และ 7 โดยแตกต่างจากกฎของเบนฟอร์ดเล็กน้อย

เมื่อนำความถี่ของข้อมูลที่มีเลขหลักแรกเป็น $d$ ซึ่งก็คือ $N_d$ ของแต่ละ $d$ โดยที่ $d = 1, 2, 3..., 9$ ของทั้งสองข้อมูล, ค่า $q$ ตั้งแต่ 1 ถึง 5 และค่าประมาณ $\alpha$ ที่คำนวณได้จากสมการ (2.1) โดยโปรแกรม R ลงในฟังก์ชัน $\ln L(\alpha)$ จะได้ผลลัพธ์ดังตาราง

| $q$ | ราคาค่าไฟฟ้า | | ราคาค่าไฟฟ้าหน่วยละ 5 บาท | |
|---|---|---|---|---|
| | $\alpha$ | $\ln L(\alpha)$ | $\alpha$ | $\ln L(\alpha)$ |
| 1 | 1.10321 | -194.5368 | 1.022958 | -198.3937 |
| 2 | 1.156217 | -194.5854 | 0.9294044 | -198.5561 |
| 3 | 1.692733 | -199.3691 | 1.077997 | -198.4476 |
| 4 | 1.593893 | -198.0349 | 1.534077 | -202.3344 |
| 5 | 1.961313 | -203.5834 | 1.130904 | -198.5978 |

ตารางที่ 4.4: ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาท ในเดือนมีนาคม พ.ศ. 2557

จากตารางจะได้ว่าค่า $\ln L(\alpha)$ ที่ $N_d$ สอดคล้องกับข้อมูลราคาค่าไฟและราคาค่าไฟฟ้าหน่วยละ 5 บาทมีค่าสูงสุดเท่ากับ -194.5368 และ -198.3937 ตามลำดับ โดยที่ $\alpha = 1.10321$ และ 1.022958 ตามลำดับ ที่กรณี $q = 1$ เช่นเดียวกัน

### 4.1.3 เดือนเมษายน

| เลขโดดหลักแรก | ความถี่ที่ได้จากการทดลองเดือนเมษายน | | ความถี่ที่ได้จากกฎของเบนฟอร์ด |
|---|---|---|---|
| | ราคาค่าไฟฟ้า | ราคาค่าไฟฟ้าหน่วยละ 5 บาท | |
| 1 | 312 | 199 | 301 |
| 2 | 184 | 193 | 176 |
| 3 | 119 | 128 | 125 |
| 4 | 85 | 79 | 97 |
| 5 | 77 | 72 | 79 |
| 6 | 63 | 62 | 67 |
| 7 | 62 | 67 | 58 |
| 8 | 40 | 48 | 51 |
| 9 | 39 | 52 | 46 |
| รวม | 1,000 | 1,000 | 1,000 |

ตารางที่ 4.5: ความถี่ที่ได้จากการทดลองของเลขโดดหลักแรกในเดือนเมษายน ปี พ.ศ. 2557 จากการสุ่มจำนวน 1,000 ตัวอย่าง

จะเห็นว่าการแจกแจงความถี่ของเลขโดดหลักแรกของข้อมูลทั้ง 2 ชนิด มีความใกล้เคียงกับกฎของเบนฟอร์ด มีเพียงความถี่ของราคาค่าไฟฟ้าหน่วยละ 5 บาท ที่จะสังเกตเห็นความแตกต่างที่เลขโดด 1 และ 2 ที่ต่างจากกฎของเบนฟอร์ดพอสมควร

เมื่อนำความถี่ของข้อมูลที่มีเลขหลักแรกเป็น $d$ ซึ่งก็คือ $N_d$ ของแต่ละ $d$ โดยที่ $d = 1, 2, 3..., 9$ ของทั้งสองข้อมูล, ค่า $q$ ตั้งแต่ 1 ถึง 5 และค่าประมาณ $\alpha$ ที่คำนวณได้จากสมการ (2.1) โดยโปรแกรม R ลงในฟังก์ชัน $\ln L(\alpha)$ จะได้ผลลัพธ์ดังตาราง

| $q$ | ราคาค่าไฟฟ้า | | ราคาค่าไฟฟ้าหน่วยละ 5 บาท | |
|---|---|---|---|---|
| | $\alpha$ | $\ln L(\alpha)$ | $\alpha$ | $\ln L(\alpha)$ |
| 1 | 1.083091 | -192.0171 | 1.012359 | -198.8845 |
| 2 | 1.034603 | -192.0585 | 0.8295372 | -199.5239 |
| 3 | 1.383186 | -193.4215 | 0.9373069 | -198.9895 |
| 4 | 1.415552 | -193.7186 | 0.9030622 | -199.1092 |
| 5 | 1.336582 | -193.0375 | 0.5218161 | -203.7233 |

ตารางที่ 4.6: ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาท ในเดือนเมษายน พ.ศ. 2557

จากตารางจะได้ว่าค่า $\ln L(\alpha)$ ที่ $N_d$ สอดคล้องกับข้อมูลราคาค่าไฟและราคาค่าไฟฟ้าหน่วยละ 5 บาทมีค่าสูงสุดเท่ากับ -192.0171 และ -198.8845 ตามลำดับ โดยที่ $\alpha = 1.083091$ และ 1.012359 ตามลำดับ ที่กรณี $q = 1$ เช่นเดียวกัน

### 4.1.4 เดือนพฤษภาคม

| เลขโดดหลักแรก | ความถี่ที่ได้จากการทดลองเดือนพฤษภาคม | | ความถี่ที่ได้จากกฎของเบนฟอร์ด |
|---|---|---|---|
| | ราคาค่าไฟฟ้า | ราคาค่าไฟฟ้าหน่วยละ 5 บาท | |
| 1 | 323 | 302 | 301 |
| 2 | 186 | 179 | 176 |
| 3 | 121 | 118 | 125 |
| 4 | 104 | 96 | 97 |
| 5 | 77 | 91 | 79 |
| 6 | 66 | 54 | 67 |
| 7 | 42 | 77 | 58 |
| 8 | 51 | 45 | 51 |
| 9 | 30 | 38 | 46 |
| รวม | 1,000 | 1,000 | 1,000 |

ตารางที่ 4.7: ความถี่ที่ได้จากการทดลองของเลขโดดหลักแรกในเดือนพฤษภาคม ปี พ.ศ. 2557 จากการสุ่มจำนวน 1,000 ตัวอย่าง

จะเห็นว่าการแจกแจงความถี่ของเลขโดดหลักแรกของข้อมูลทั้ง 2 ชนิด ใกล้เคียงกับกฎของเบนฟอร์ดอย่างมาก แต่ความถี่ของเลขโดดหลักแรกเป็น 7 ของทั้งสองข้อมูลจะแตกต่างจากกฎของเบนฟอร์ดเล็กน้อย

เมื่อนำความถี่ของข้อมูลที่มีเลขหลักแรกเป็น $d$ ซึ่งก็คือ $N_d$ ของแต่ละ $d$ โดยที่ $d = 1, 2, 3..., 9$ ของทั้งสองข้อมูล, ค่า $q$ ตั้งแต่ 1 ถึง 5 และค่าประมาณ $\alpha$ ที่คำนวณได้จากสมการ (2.1) โดยโปรแกรม R ลงในฟังก์ชัน $\ln L(\alpha)$ จะได้ผลลัพธ์ดังตาราง

| $q$ | ราคาค่าไฟฟ้า | | ราคาค่าไฟฟ้าหน่วยละ 5 บาท | |
|---|---|---|---|---|
| | $\alpha$ | $\ln L(\alpha)$ | $\alpha$ | $\ln L(\alpha)$ |
| 1 | 1.110056 | -194.0474 | 1.00593 | -198.996 |
| 2 | 1.185373 | -194.1451 | 0.9874919 | -199.0021 |
| 3 | 1.599363 | -197.5359 | 1.310316 | -200.5176 |
| 4 | 1.476695 | -196.1126 | 0.8910385 | -199.1092 |
| 5 | 1.798868 | -200.3637 | 1.223305 | -199.7965 |

ตารางที่ 4.8: ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาท ในเดือนพฤษภาคม พ.ศ. 2557

จากตารางจะได้ว่าค่า $\ln L(\alpha)$ ที่ $N_d$ สอดคล้องกับข้อมูลราคาค่าไฟและราคาค่าไฟฟ้าหน่วยละ 5 บาทมีค่าสูงสุดเท่ากับ -194.0474 และ -198.996 ตามลำดับ โดยที่ $\alpha = 1.110056$ และ 1.00593 ตามลำดับ ที่กรณี $q = 1$ เช่นเดียวกัน

### 4.1.5 เดือนมิถุนายน

| เลขโดดหลักแรก | ความถี่ที่ได้จากการทดลองเดือนมิถุนายน | | ความถี่ที่ได้จากกฎของเบนฟอร์ด |
|---|---|---|---|
| | ราคาค่าไฟฟ้า | ราคาค่าไฟฟ้าหน่วยละ 5 บาท | |
| 1 | 339 | 317 | 301 |
| 2 | 183 | 156 | 176 |
| 3 | 124 | 120 | 125 |
| 4 | 80 | 89 | 97 |
| 5 | 79 | 80 | 79 |
| 6 | 47 | 57 | 67 |
| 7 | 59 | 72 | 58 |
| 8 | 46 | 54 | 51 |
| 9 | 43 | 55 | 46 |
| รวม | 1,000 | 1,000 | 1,000 |

ตารางที่ 4.9: ความถี่ที่ได้จากการทดลองของเลขโดดหลักแรกในเดือนมิถุนายน ปี พ.ศ. 2557 จากการสุ่มจำนวน 1,000 ตัวอย่าง

จะเห็นว่าการแจกแจงความถี่ของเลขโดดหลักแรกของข้อมูลทั้ง 2 ชนิด มีความแตกต่างจากกฎของเบนฟอร์ดเล็กน้อย

เมื่อนำความถี่ของข้อมูลที่มีเลขหลักแรกเป็น $d$ ซึ่งก็คือ $N_d$ ของแต่ละ $d$ โดยที่ $d = 1, 2, 3..., 9$ ของทั้งสองข้อมูล, ค่า $q$ ตั้งแต่ 1 ถึง 5 และค่าประมาณ $\alpha$ ที่คำนวณได้จากสมการ (2.1) โดยโปรแกรม R ลงในฟังก์ชัน $\ln L(\alpha)$ จะได้ผลลัพธ์ดังตาราง

| $q$ | ราคาค่าไฟฟ้า | | ราคาค่าไฟฟ้าหน่วยละ 5 บาท | |
|---|---|---|---|---|
| | $\alpha$ | $\ln L(\alpha)$ | $\alpha$ | $\ln L(\alpha)$ |
| 1 | 1.157422 | -192.9864 | 1.01386 | -200.0987 |
| 2 | 1.130825 | -192.9984 | 0.9865929 | -200.1115 |
| 3 | 1.285812 | -193.2512 | 0.655558 | -202.4634 |
| 4 | 1.141823 | -192.9905 | 0.6025886 | -203.239 |
| 5 | 0.8664813 | -194.5404 | 0.7410264 | -201.4498 |

ตารางที่ 4.10: ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาท ในเดือนมิถุนายน พ.ศ. 2557

จากตารางจะได้ว่าค่า $\ln L(\alpha)$ ที่ $N_d$ สอดคล้องกับข้อมูลราคาค่าไฟและราคาค่าไฟฟ้าหน่วยละ 5 บาทมีค่าสูงสุดเท่ากับ -192.9864 และ -200.0987 ตามลำดับ โดยที่ $\alpha = 1.157422$ และ 1.01386 ตามลำดับ ที่กรณี $q = 1$ เช่นเดียวกัน

### 4.1.6 เดือนกรกฎาคม

| เลขโดดหลักแรก | ความถี่ที่ได้จากการทดลองเดือนกรกฎาคม | | ความถี่ที่ได้จากกฎของเบนฟอร์ด |
|---|---|---|---|
| | ราคาค่าไฟฟ้า | ราคาค่าไฟฟ้าหน่วยละ 5 บาท | |
| 1 | 328 | 319 | 301 |
| 2 | 186 | 175 | 176 |
| 3 | 122 | 133 | 125 |
| 4 | 88 | 86 | 97 |
| 5 | 63 | 76 | 79 |
| 6 | 62 | 52 | 67 |
| 7 | 57 | 59 | 58 |
| 8 | 49 | 50 | 51 |
| 9 | 45 | 50 | 46 |
| รวม | 1,000 | 1,000 | 1,000 |

ตารางที่ 4.11: ความถี่ที่ได้จากการทดลองของเลขโดดหลักแรกในเดือนกรกฎาคม ปี พ.ศ. 2557 จากการสุ่มจำนวน 1,000 ตัวอย่าง

จะเห็นว่าการแจกแจงความถี่ของเลขโดดหลักแรกของข้อมูลทั้ง 2 ชนิด มีความใกล้เคียงกับกฎของเบนฟอร์ดอย่างมาก แต่จะเห็นว่าที่เลขโดด 5 และ 6 ของราคาค่าไฟฟ้า และราคาค่าไฟฟ้าที่ราคาหน่วยละ 5 บาท จะต่างจากกฎของเบนฟอร์ดเล็กน้อย

เมื่อนำความถี่ของข้อมูลที่มีเลขหลักแรกเป็น $d$ ซึ่งก็คือ $N_d$ ของแต่ละ $d$ โดยที่ $d = 1, 2, 3..., 9$ ของทั้งสองข้อมูล, ค่า $q$ ตั้งแต่ 1 ถึง 5 และค่าประมาณ $\alpha$ ที่คำนวณได้จากสมการ (2.1) โดยโปรแกรม R ลงในฟังก์ชัน $\ln L(\alpha)$ จะได้ผลลัพธ์ดังตาราง

| $q$ | ราคาค่าไฟฟ้า | | ราคาค่าไฟฟ้าหน่วยละ 5 บาท | |
|---|---|---|---|---|
| | $\alpha$ | $\ln L(\alpha)$ | $\alpha$ | $\ln L(\alpha)$ |
| 1 | 1.116149 | -194.7732 | 1.070552 | -196.7573 |
| 2 | 1.041219 | -194.8722 | 1.087666 | -196.7624 |
| 3 | 1.010208 | -194.973 | 0.8798518 | -197.4286 |
| 4 | 1.196721 | -194.8821 | 0.9151443 | -197.1992 |
| 5 | 1.337812 | -195.5556 | 0.5584847 | -201.9009 |

ตารางที่ 4.12: ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาท ในเดือนกรกฎาคม พ.ศ. 2557

จากตารางจะได้ว่าค่า $\ln L(\alpha)$ ที่ $N_d$ สอดคล้องกับข้อมูลราคาค่าไฟและราคาค่าไฟฟ้าหน่วยละ 5 บาทมีค่าสูงสุดเท่ากับ -194.7732 และ -196.7573 ตามลำดับ โดยที่ $\alpha = 1.116149$ และ 1.070552 ตามลำดับ ที่กรณี $q = 1$ เช่นเดียวกัน

### 4.1.7 เดือนสิงหาคม

| เลขโดดหลักแรก | ความถี่ที่ได้จากการทดลองเดือนสิงหาคม | | ความถี่ที่ได้จากกฎของเบนฟอร์ด |
|---|---|---|---|
| | ราคาค่าไฟฟ้า | ราคาค่าไฟฟ้าหน่วยละ 5 บาท | |
| 1 | 319 | 307 | 301 |
| 2 | 189 | 166 | 176 |
| 3 | 113 | 121 | 125 |
| 4 | 92 | 85 | 97 |
| 5 | 78 | 86 | 79 |
| 6 | 57 | 58 | 67 |
| 7 | 60 | 78 | 58 |
| 8 | 56 | 43 | 51 |
| 9 | 36 | 56 | 46 |
| รวม | 1,000 | 1,000 | 1,000 |

ตารางที่ 4.13: ความถี่ที่ได้จากการทดลองของเลขโดดหลักแรกในเดือนสิงหาคม ปี พ.ศ. 2557 จากการสุ่มจำนวน 1,000 ตัวอย่าง

จะเห็นว่าการแจกแจงความถี่ของเลขโดดหลักแรกของข้อมูลทั้ง 2 ชนิด ใกล้เคียงกับกฎของเบนฟอร์ดอย่างมาก แต่จะมีที่เลขโดด 7 ของราคาค่าไฟฟ้าทีราคาหน่วยละ 5 บาท ที่สูงกว่ากฎของเบนฟอร์ดเล็กน้อย

เมื่อนำความถี่ของข้อมูลที่มีเลขหลักแรกเป็น $d$ ซึ่งก็คือ $N_d$ ของแต่ละ $d$ โดยที่ $d = 1, 2, 3..., 9$ ของทั้งสองข้อมูล, ค่า $q$ ตั้งแต่ 1 ถึง 5 และค่าประมาณ $\alpha$ ที่คำนวณได้จากสมการ (2.1) โดยโปรแกรม R ลงในฟังก์ชัน $\ln L(\alpha)$ จะได้ผลลัพธ์ดังตาราง

| $q$ | ราคาค่าไฟฟ้า | | ราคาค่าไฟฟ้าหน่วยละ 5 บาท | |
|---|---|---|---|---|
| | $\alpha$ | $\ln L(\alpha)$ | $\alpha$ | $\ln L(\alpha)$ |
| 1 | 1.082572 | -196.0049 | 0.9927722 | -200.4504 |
| 2 | 0.9825018 | -196.1862 | 0.9283604 | -200.5244 |
| 3 | 1.331141 | -197.0012 | 0.8767914 | -200.6935 |
| 4 | 0.9369705 | -196.3936 | 0.6863251 | -202.2137 |
| 5 | 1.364534 | -197.2705 | 0.6493069 | -202.6788 |

ตารางที่ 4.14: ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาท ในเดือนสิงหาคม พ.ศ. 2557

จากตารางจะได้ว่าค่า $\ln L(\alpha)$ ที่ $N_d$ สอดคล้องกับข้อมูลราคาค่าไฟและราคาค่าไฟฟ้าหน่วยละ 5 บาทมีค่าสูงสุดเท่ากับ -196.0049 และ -200.4504 ตามลำดับ โดยที่ $\alpha = 1.082572$ และ 0.9927722 ตามลำดับ ที่กรณี $q = 1$ เช่นเดียวกัน

### 4.1.8 เดือนกันยายน

| เลขโดดหลักแรก | ความถี่ที่ได้จากการทดลองเดือนกันยายน | | ความถี่ที่ได้จากกฎของเบนฟอร์ด |
|---|---|---|---|
| | ราคาค่าไฟฟ้า | ราคาค่าไฟฟ้าหน่วยละ 5 บาท | |
| 1 | 334 | 307 | 301 |
| 2 | 177 | 193 | 176 |
| 3 | 123 | 107 | 125 |
| 4 | 96 | 87 | 97 |
| 5 | 73 | 75 | 79 |
| 6 | 61 | 68 | 67 |
| 7 | 47 | 70 | 58 |
| 8 | 45 | 51 | 51 |
| 9 | 44 | 42 | 46 |
| รวม | 1,000 | 1,000 | 1,000 |

ตารางที่ 4.15: ความถี่ที่ได้จากการทดลองของเลขโดดหลักแรกในเดือนกันยายน ปี พ.ศ. 2557 จากการสุ่มจำนวน 1,000 ตัวอย่าง

จะเห็นว่าการแจกแจงความถี่ของเลขโดดหลักแรกของข้อมูลทั้ง 2 ชนิด มีความใกล้เคียงกับกฎของเบนฟอร์ด แต่จะมีที่เลขโดด 3 และ 4 ของราคาค่าไฟฟ้าหน่วยละ 5 บาทที่ต่ำกว่ากฎของเบนฟอร์ดเล็กน้อย

เมื่อนำความถี่ของข้อมูลที่มีเลขหลักแรกเป็น $d$ ซึ่งก็คือ $N_d$ ของแต่ละ $d$ โดยที่ $d = 1, 2, 3..., 9$ ของทั้งสองข้อมูล, ค่า $q$ ตั้งแต่ 1 ถึง 5 และค่าประมาณ $\alpha$ ที่คำนวณได้จากสมการ (2.1) โดยโปรแกรม R ลงในฟังก์ชัน $\ln L(\alpha)$ จะได้ผลลัพธ์ดังตาราง

| $q$ | ราคาค่าไฟฟ้า | | ราคาค่าไฟฟ้าหน่วยละ 5 บาท | |
|---|---|---|---|---|
| | $\alpha$ | $\ln L(\alpha)$ | $\alpha$ | $\ln L(\alpha)$ |
| 1 | 1.133816 | -193.8489 | 1.031355 | -198.3742 |
| 2 | 1.192515 | -193.9064 | 0.8157761 | -199.2501 |
| 3 | 1.217825 | -193.9656 | 1.232637 | -199.0541 |
| 4 | 1.30392 | -194.3123 | 0.94673 | -198.505 |
| 5 | 1.34127 | -194.5283 | 1.449082 | -201.0757 |

ตารางที่ 4.16: ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาท ในเดือนกันยายน พ.ศ. 2557

จากตารางจะได้ว่าค่า $\ln L(\alpha)$ ที่ $N_d$ สอดคล้องกับข้อมูลราคาค่าไฟและราคาค่าไฟฟ้าหน่วยละ 5 บาทมีค่าสูงสุดเท่ากับ -193.8489 และ -198.3742 ตามลำดับ โดยที่ $\alpha = 1.133816$ และ 1.031355 ตามลำดับ ที่กรณี $q = 1$ เช่นเดียวกัน

### 4.1.9 เดือนตุลาคม

| เลขโดดหลักแรก | ความถี่ที่ได้จากการทดลองเดือนตุลาคม | | ความถี่ที่ได้จากกฎของเบนฟอร์ด |
|---|---|---|---|
| | ราคาค่าไฟฟ้า | ราคาค่าไฟฟ้าหน่วยละ 5 บาท | |
| 1 | 328 | 325 | 301 |
| 2 | 183 | 161 | 176 |
| 3 | 103 | 118 | 125 |
| 4 | 88 | 79 | 97 |
| 5 | 86 | 71 | 79 |
| 6 | 61 | 67 | 67 |
| 7 | 60 | 65 | 58 |
| 8 | 51 | 57 | 51 |
| 9 | 40 | 57 | 46 |
| รวม | 1,000 | 1,000 | 1,000 |

ตารางที่ 4.17: ความถี่ที่ได้จากการทดลองของเลขโดดหลักแรกในเดือนตุลาคม ปี พ.ศ. 2557 จากการสุ่มจำนวน 1,000 ตัวอย่าง

จะเห็นว่าการแจกแจงความถี่ของเลขโดดหลัก แรกของข้อมูลทั้ง 2 ชนิด มีความใกล้เคียงกับกฎของเบนฟอร์ด แต่จะมีความแตกต่างของข้อมูลทั้ง 2 ชนิดกับกฎของเบนฟอร์ดในช่วงเลขโดด 2 ถึง 5 เล็กน้อย

เมื่อนำความถี่ของข้อมูลที่มีเลขหลักแรกเป็น $d$ ซึ่งก็คือ $N_d$ ของแต่ละ $d$ โดยที่ $d = 1, 2, 3..., 9$ ของทั้งสองข้อมูล, ค่า $q$ ตั้งแต่ 1 ถึง 5 และค่าประมาณ $\alpha$ ที่คำนวณได้จากสมการ (2.1) โดยโปรแกรม R ลงในฟังก์ชัน $\ln L(\alpha)$ จะได้ผลลัพธ์ดังตาราง

| $q$ | ราคาค่าไฟฟ้า | | ราคาค่าไฟฟ้าหน่วยละ 5 บาท | |
|---|---|---|---|---|
| | $\alpha$ | $\ln L(\alpha)$ | $\alpha$ | $\ln L(\alpha)$ |
| 1 | 1.099869 | -195.835 | 1.047747 | -199.0789 |
| 2 | 0.9689762 | -196.1387 | 0.9282223 | -199.3263 |
| 3 | 1.488011 | -198.0632 | 0.6489575 | -201.982 |
| 4 | 0.9138419 | -196.4575 | 0.8735911 | -199.6103 |
| 5 | 1.368227 | -196.9524 | 0.6699908 | -201.6756 |

ตารางที่ 4.18: ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาท ในเดือนตุลาคม พ.ศ. 2557

จากตารางจะได้ว่าค่า $\ln L(\alpha)$ ที่ $N_d$ สอดคล้องกับข้อมูลราคาค่าไฟและราคาค่าไฟฟ้าหน่วยละ 5 บาทมีค่าสูงสุดเท่ากับ -195.835 และ -199.0789 ตามลำดับ โดยที่ $\alpha = 1.099869$ และ 1.047747 ตามลำดับ ที่กรณี $q = 1$ เช่นเดียวกัน

### 4.1.10 เดือนพฤศจิกายน

| เลขโดดหลักแรก | ความถี่ที่ได้จากการทดลองเดือนตุลาคม | | ความถี่ที่ได้จากกฎของเบนฟอร์ด |
|---|---|---|---|
| | ราคาค่าไฟฟ้า | ราคาค่าไฟฟ้าหน่วยละ 5 บาท | |
| 1 | 324 | 339 | 301 |
| 2 | 182 | 157 | 176 |
| 3 | 113 | 115 | 125 |
| 4 | 92 | 88 | 97 |
| 5 | 76 | 78 | 79 |
| 6 | 60 | 57 | 67 |
| 7 | 64 | 71 | 58 |
| 8 | 46 | 51 | 51 |
| 9 | 43 | 44 | 46 |
| รวม | 1,000 | 1,000 | 1,000 |

ตารางที่ 4.19: ความถี่ที่ได้จากการทดลองของเลขโดดหลักแรกในเดือนพฤศจิกายน ปี พ.ศ. 2557 จากการสุ่มจำนวน 1,000 ตัวอย่าง

จะเห็นว่าการแจกแจงความถี่ของเลขโดดหลักแรกของข้อมูลทั้ง 2 ชนิด มีความใกล้เคียงกับกฎของเบนฟอร์ดอย่างมาก

เมื่อนำความถี่ของข้อมูลที่มีเลขหลักแรกเป็น $d$ ซึ่งก็คือ $N_d$ ของแต่ละ $d$ โดยที่ $d = 1, 2, 3..., 9$ ของทั้งสองข้อมูล, ค่า $q$ ตั้งแต่ 1 ถึง 5 และค่าประมาณ $\alpha$ ที่คำนวณได้จากสมการ (2.1) โดยโปรแกรม R ลงในฟังก์ชัน $\ln L(\alpha)$ จะได้ผลลัพธ์ดังตาราง

| $q$ | ราคาค่าไฟฟ้า | | ราคาค่าไฟฟ้าหน่วยละ 5 บาท | |
|---|---|---|---|---|
| | $\alpha$ | $\ln L(\alpha)$ | $\alpha$ | $\ln L(\alpha)$ |
| 1 | 1.090222 | -195.9968 | 1.107742 | -196.3707 |
| 2 | 1.013696 | -196.1004 | 1.111833 | -196.371 |
| 3 | 1.231081 | -196.3241 | 0.9480679 | -196.8077 |
| 4 | 1.03341 | -196.0535 | 1.131428 | -196.3798 |
| 5 | 1.437174 | -197.8328 | 1.307122 | -196.978 |

ตารางที่ 4.20: ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาท ในเดือนพฤศจิกายน พ.ศ. 2557

จากตารางจะได้ว่าค่า $\ln L(\alpha)$ ที่ $N_d$ สอดคล้องกับข้อมูลราคาค่าไฟและราคาค่าไฟฟ้าหน่วยละ 5 บาทมีค่าสูงสุดเท่ากับ -195.9968 และ -196.3707 ตามลำดับ โดยที่ $\alpha = 1.090222$ และ 1.107742 ตามลำดับ ที่กรณี $q = 1$ เช่นเดียวกัน

### 4.1.11 เดือนธันวาคม

| เลขโดดหลักแรก | ความถี่ที่ได้จากการทดลองเดือนตุลาคม | | ความถี่ที่ได้จากกฎของเบนฟอร์ด |
|---|---|---|---|
| | ราคาค่าไฟฟ้า | ราคาค่าไฟฟ้าหน่วยละ 5 บาท | |
| 1 | 329 | 274 | 301 |
| 2 | 186 | 187 | 176 |
| 3 | 118 | 136 | 125 |
| 4 | 92 | 91 | 97 |
| 5 | 71 | 70 | 79 |
| 6 | 66 | 59 | 67 |
| 7 | 46 | 99 | 58 |
| 8 | 44 | 40 | 51 |
| 9 | 48 | 44 | 46 |
| รวม | 1,000 | 1,000 | 1,000 |

ตารางที่ 4.21: ความถี่ที่ได้จากการทดลองของเลขโดดหลักแรกในเดือนธันวาคม ปี พ.ศ. 2557 จากการสุ่มจำนวน 1,000 ตัวอย่าง

จะเห็นว่าการแจกแจงความถี่ของเลขโดดหลักแรกของข้อมูลทั้ง 2 ชนิด มีความใกล้เคียงกับกฎของเบนฟอร์ด ยกเว้นที่เลขโดด 7 ของราคาค่าไฟฟ้าที่ราคาหน่วยละ 5 บาท จะสูงกว่ากฎของเบนฟอร์ดอย่างมาก

เมื่อนำความถี่ของข้อมูลที่มีเลขหลักแรกเป็น $d$ ซึ่งก็คือ $N_d$ ของแต่ละ $d$ โดยที่ $d = 1, 2, 3..., 9$ ของทั้งสองข้อมูล, ค่า $q$ ตั้งแต่ 1 ถึง 5 และค่าประมาณ $\alpha$ ที่คำนวณได้จากสมการ (2.1) โดยโปรแกรม R ลงในฟังก์ชัน $\ln L(\alpha)$ จะได้ผลลัพธ์ดังตาราง

| $q$ | ราคาค่าไฟฟ้า | | ราคาค่าไฟฟ้าหน่วยละ 5 บาท | |
|---|---|---|---|---|
| | $\alpha$ | $\ln L(\alpha)$ | $\alpha$ | $\ln L(\alpha)$ |
| 1 | 1.121388 | -194.3891 | 0.9170115 | -202.1267 |
| 2 | 1.075153 | -194.4265 | 0.8206427 | -202.309 |
| 3 | 1.254865 | -194.6822 | 0.7696889 | -202.5572 |
| 4 | 1.180202 | -194.4475 | 1.061442 | -202.5121 |
| 5 | 1.282388 | -194.8112 | 1.154186 | -203.137 |

ตารางที่ 4.22: ค่าลอการิทึมของฟังก์ชันความควรจะเป็น $L(\alpha)$ สำหรับแต่ละ $\alpha$ และ $q$ ของข้อมูลราคาค่าไฟฟ้าและราคาค่าไฟฟ้าหน่วยละ 5 บาท ในเดือนธันวาคม พ.ศ. 2557

จากตารางจะได้ว่าค่า $\ln L(\alpha)$ ที่ $N_d$ สอดคล้องกับข้อมูลราคาค่าไฟและราคาค่าไฟฟ้าหน่วยละ 5 บาทมีค่าสูงสุดเท่ากับ -194.3891 และ -202.1267 ตามลำดับ โดยที่ $\alpha = 1.121388$ และ 0.9170115 ตามลำดับ ที่กรณี $q = 1$ เช่นเดียวกัน

## 4.2 สรุปผลการทดสอบความเป็นเบนฟอร์ดของข้อมูลค่าไฟฟ้ารายเดือน

เนื่องจากเราทราบว่าการแจกแจงที่มีฟังก์ชันการแจกแจง $G_{\alpha,q}(s)$ ที่เราคิดว่าเป็นแบบจำลองของเลขหลักแรกในธรรมชาติได้ดี จะสอดคล้องกับกฎของเบนฟอร์ดเมื่อพารามิเตอร์ $\alpha$ มีค่าเท่ากับ 1 หรือเมื่อจำนวนนับ $q$ ลู่เข้าสู่ $\infty$ แต่จากการหาค่าประมาณของพารามิเตอร์ $\alpha$ สำหรับข้อมูลราคาไฟฟ้าและราคาไฟฟ้าหน่วยละ 5 บาทของแต่ละเดือน โดยพิจารณาในกรณีที่ $q$ เป็นจำนวนนับในช่วง 1 ถึง 5 จะเห็นได้ว่าค่าประมาณของพารามิเตอร์ $\alpha$ ที่ทำให้ค่าของฟังก์ชันลอการิทึมของฟังก์ชันความควรจะเป็นมีค่าสูงสุดมักเกิดขึ้นในกรณีที่ $q = 1$ เราจึงไม่สนใจการทดสอบความเป็นเบนฟอร์ดของข้อมูลในกรณีที่ $q$ ลู่เข้าสู่ $\infty$ แต่สนใจการทดสอบความเป็นเบนฟอร์ดของข้อมูลโดยการเปรียบเทียบค่าประมาณของพารามิเตอร์ $\alpha$ กับ 1 แทน

การทดสอบความเป็นเบนฟอร์ดของข้อมูลราคาค่าไฟฟ้าโดยวิธีการนี้ เป็นอีกหนึ่งวิธีการทดสอบการแจกแจงของข้อมูล ที่ต้องการแก้ไขข้อจำกัดในการทดสอบการแจกแจงที่ว่า เมื่อข้อมูลมีปริมาณที่มาก การทดสอบไคแสควร์มักให้ผลปฏิเสธสมมติฐานหลัก นั่นคือปฏิเสธว่าการแจกแจงที่เราต้องการทดสอบไม่สอดคล้องกับการแจกแจงตามทฤษฎี เราจึงทำการเปรียบเทียบผลการทดสอบความเป็นเบนฟอร์ดของข้อมูลค่าไฟฟ้ารายเดือนโดยวิธีการหาค่าประมาณของพารามิเตอร์ด้วยวิธีความควรจะเป็นสูงสุดเทียบกับค่าจริงตามทฤษฎีซึ่งคือ 1 และการทดสอบไคแสควร์ ซึ่งได้ทำการทดสอบไว้แล้วใน [4] จึงได้ตารางการเปรียบเทียบการทดสอบสำหรับแต่ละข้อมูล ดังนี้

### 4.2.1 ราคาค่าไฟฟ้า

| เดือน | ค่าไคสแควร์ | ยอมรับ/ปฏิเสธ สมมติฐานหลัก ที่ระดับนัยสำคัญ 0.05 | $\alpha$ | $\alpha$ | $\lvert\alpha - 1\rvert$ | $\lvert\alpha - 1\rvert$ |
|---|---|---|---|---|---|---|
| กุมภาพันธ์ | 5.4129 | ยอมรับ | 1.098468 | | 0.098468 | |
| มีนาคม | 12.3023 | ยอมรับ | 1.10321 | | 0.10321 | |
| เมษายน | 6.2999 | ยอมรับ | 1.083091 | | 0.083091 | |
| พฤษภาคม | 6.2999 | ยอมรับ | 1.110056 | | 0.110056 | |
| มิถุนายน | 14.7364 | ยอมรับ | 1.157422 | | 0.157422 | |
| กรกฎาคม | 7.6282 | ยอมรับ | 1.116149 | | 0.116149 | |
| สิงหาคม | 7.6846 | ยอมรับ | 1.082572 | | 0.082572 | |
| กันยายน | 7.538 | ยอมรับ | 1.133816 | | 0.133816 | |
| ตุลาคม | 9.4165 | ยอมรับ | 1.099869 | | 0.099869 | |
| พฤศจิกายน | 5.5236 | ยอมรับ | 1.090222 | | 0.090222 | |
| ธันวาคม | 8.1781 | ยอมรับ | 1.121388 | | 0.121388 | |

ตารางที่ 4.23: ผลการทดสอบความเป็นเบนฟอร์ดของข้อมูลราคาค่าไฟฟ้าของแต่ละเดือนในปี พ.ศ. 2557 ด้วยการทดสอบไคสแควร์ และสังเกตระยะห่างของค่าประมาณ $\alpha$ กับ 1

จากผลการทดสอบไคสแควร์ของการแจกแจงข้อมูลราคาค่าไฟฟ้าใน [4] ที่ว่า ข้อมูลราคาค่าไฟฟ้าของทุกเดือนในปี พ.ศ. 2557 มีการแจกแจงของข้อมูลสอดคล้องกับกฎของเบนฟอร์ด ประกอบกับการหาระยะห่างหรือความแตกต่างของค่าประมาณของพารามิเตอร์ $\alpha$ กับ 1 สำหรับข้อมูลราคาค่าไฟฟ้าแต่ละเดือน เราได้ว่าข้อมูลราคาค่าไฟฟ้าในเดือนมิถุนายนมีความแตกต่างกันของค่าประมาณพารามิเตอร์ $\alpha$ กับ 1 มากที่สุด แต่แตกต่างกันเพียงแค่ 0.157422 เท่านั้น ซึ่งจะเห็นได้ว่า 1 กับค่าประมาณของพารามิเตอร์ $\alpha$ สำหรับข้อมูลราคาค่าไฟฟ้าในแต่ละเดือนมีความแตกต่างกันน้อยมาก ดังนั้นเราจึงสรุปได้ว่าการแจกแจงของข้อมูลราคาค่าไฟฟ้าแต่ละเดือน ในปี พ.ศ. 2557 มีการแจกแจงของข้อมูลสอดคล้องกับกฎของเบนฟอร์ด

### 4.2.2 ราคาค่าไฟฟ้าหน่วยละ 5 บาท

| เดือน | ค่าไคสแควร์ | ยอมรับ/ปฏิเสธ สมมติฐานหลัก ที่ระดับนัยสำคัญ 0.05 | $\alpha$ | $\alpha$ | $\lvert\alpha - 1\rvert$ | $\lvert\alpha - 1\rvert$ |
|---|---|---|---|---|---|---|
| กุมภาพันธ์ | 3.6353 | ยอมรับ | 0.977861 | | 0.022139 | |
| มีนาคม | 10.411 | ยอมรับ | 1.022958 | | 0.022958 | |
| เมษายน | 8.4166 | ยอมรับ | 1.012359 | | 0.012359 | |
| พฤษภาคม | 8.4166 | ยอมรับ | 1.00593 | | 0.00593 | |
| มิถุนายน | 10.8049 | ยอมรับ | 1.01386 | | 0.01386 | |
| กรกฎาคม | 6.6983 | ยอมรับ | 1.070552 | | 0.070552 | |
| สิงหาคม | 14.4549 | ยอมรับ | 0.9927722 | | 0.0072278 | |
| กันยายน | 8.4326 | ยอมรับ | 1.031355 | | 0.031355 | |
| ตุลาคม | 11.9155 | ยอมรับ | 1.047747 | | 0.07747 | |
| พฤศจิกายน | 12.999 | ยอมรับ | 1.107742 | | 0.107742 | |
| ธันวาคม | 37.874 | ปฏิเสธ | 0.9170115 | | 0.0829885 | |

ตารางที่ 4.24: ผลการทดสอบความเป็นเบนฟอร์ดของข้อมูลราคาค่าไฟฟ้าหน่วยละ 5 บาทของแต่ละเดือนในปี พ.ศ. 2557 ด้วยการทดสอบไคสแควร์ และสังเกตระยะห่างของค่าประมาณ $\alpha$ กับ 1

จากผลการทดสอบไคสแควร์ของการแจกแจงข้อมูลราคาค่าไฟฟ้าหน่วยละ 5 บาทใน [4] ที่ว่า ข้อมูลราคาค่าไฟฟ้าหน่วยละ 5 บาทของทุกเดือนในปี พ.ศ. 2557 มีการแจกแจงของข้อมูลสอดคล้องกับกฎของเบนฟอร์ด ยกเว้นในเดือนธันวาคมที่ได้ผลการทดสอบว่าข้อมูลราคาค่าไฟฟ้าหน่วยละ 5 บาทไม่สอดคล้องกับกฎของเบนฟอร์ด แต่จากการหาระยะห่างหรือความแตกต่างของค่าประมาณพารามิเตอร์ $\alpha$ กับ 1 สำหรับข้อมูลราคาค่าไฟฟ้าหน่วยละ 5 บาทของแต่ละเดือน เราได้ว่าข้อมูลราคาค่าไฟฟ้าหน่วยละ 5 บาทในเดือนพฤศจิกายนมีความแตกต่างกันของค่าประมาณพารามิเตอร์ $\alpha$ กับ 1 มากที่สุด แต่แตกต่างกันเพียงแค่ 0.107742 เท่านั้น ซึ่งจะเห็นได้ว่า 1 กับค่าประมาณของพารามิเตอร์ $\alpha$ สำหรับข้อมูลราคาค่าไฟฟ้าในแต่ละเดือนมีความแตกต่างกันน้อยมาก ดังนั้นเราจึงสรุปได้ว่าการแจกแจงของข้อมูลราคาค่าไฟฟ้าหน่วยละ 5 บาทของแต่ละเดือน ในปี พ.ศ. 2557 มีการแจกแจงของข้อมูลสอดคล้องกับกฎของเบนฟอร์ด

การสรุปผลดังนี้อาจทำให้เกิดข้อขัดแย้งกับการทดสอบไคสแควร์ของการแจกแจงข้อมูลราคาค่าไฟฟ้าหน่วยละ 5 บาท ที่ว่าข้อมูลราคาค่าไฟฟ้าหน่วยละ 5 บาทในเดือนธันวาคม มีการแจกแจงข้อมูลที่ไม่สอดคล้องกับกฎของเบนฟอร์ด แต่เมื่อเราพิจารณาข้อมูล ประกอบกับสมบัติการไม่เปลี่ยนแปลงมาตราส่วนของกฎของเบนฟอร์ดในบทที่ 2 ของ [4] ที่ว่าข้อมูลที่มีการแจกแจงของเลขหลักแรกสอดคล้องกับกฎของเบนฟอร์ด เมื่อถูกคูณด้วยค่าคงที่ $c > 0$ จะยังคงมีการแจกแจงสะสม

ของเลขโดดหลักแรกเป็นไปตามกฎของเบนฟอร์ด และจากหัวข้อ 4.2.1 เราทราบว่าข้อมูลราคาค่าไฟฟ้าของทุกเดือนสอดคล้องกับกฎของเบนฟอร์ด ฉะนั้นข้อมูลราคาค่าไฟฟ้าหน่วยละ 5 บาทของทุกเดือน ซึ่งเป็นข้อมูลราคาค่าไฟฟ้าของทุกเดือนคูณด้วยค่าคงที่ 5 ควรจะมีการแจกแจงของเลขหลักแรกสอดคล้องกับกฎของเบนฟอร์ดเช่นกัน

จากการเปรียบเทียบผลการทดสอบการแจกแจงเลขหลักแรกของข้อมูลราคาค่าไฟฟ้า โดยการทดสอบทั้งสองวิธี จะเห็นได้ว่าวิธีการพิจารณาระยะห่างของ 1 กับค่าประมาณของพารามิเตอร์ $\alpha$ ที่ได้จากการประมาณด้วยวิธีความควรจะเป็นสูงสุด ให้ผลการทดสอบที่สอดคล้องกับสมบัติของกฎของเบนฟอร์ด ฉะนั้นเราจึงอยากเสนอให้วิธีการในโครงงานนี้ เป็นทางเลือกหนึ่งในการบ่งชี้ระยะห่างจากกฎของเบนฟอร์ด

## 4.3 ปัญหาและข้อเสนอแนะ

เนื่องจากโครงงานนี้เป็นการสังเกตระยะห่างระหว่างค่าประมาณของพารามิเตอร์ $\alpha$ ของข้อมูลต่างๆ และ 1 ซึ่งสามารถบ่งชี้ได้ว่าข้อมูลนั้นๆ มีการแจกแจงเลขหลักแรกใกล้เคียงกับกฎของเบนฟอร์ดมากน้อยเพียงใด แต่ในการสรุปผลว่าข้อมูลเหล่านั้นสอดคล้องกับกฎของเบนฟอร์ดหรือไม่ตามหลักการทางสถิติ ยังคงต้องการหาช่วงความเชื่อมั่นของค่าประมาณพารามิเตอร์ $\alpha$ ที่คาดว่าอาจขึ้นกับค่าของพารามิเตอร์ $q$

# บรรณานุกรม

[1] Wiriyakraikul, T., Sumetkijakan, S., Santiwipanont, T. : Benford's Law for Chains of Truncated Distributions, *Proceeding of the 22nd Annual Meeting in Mathematicas*, PRO-03(2017).

[2] Wiriyakraikul, T., Sumetkijakan, S., Santiwipanont, T. : " Benfordness of Chains of Truncated Beta Distributions via a Piecewise Constant Approximation" in Structure Changes and their Econometric Modeling (International Conference of the Thailand Econometrics Society) , pp. 342-351

[3] Hürlimann, W.: A generalized Benford law and its application. In Advances and Applications in Statistics 3(3), 217-228 (2003).

[4] ธัญญาดา คงพริ้ว, "กฎของเบนฟอร์ดกับข้อมูลการใช้ไฟฟ้า," (วิทยานิพนธ์ปริญญาบัณฑิต ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2557)

[5] รองศาสตราจารย์ ดร. ทรงเกียรติ สุเมธกิจการ, รองศาสตราจารย์ ทิพวัลย์ สันติวิภานนท์ และรองศาสตราจารย์ ดร. ณัฐกาญจน์ ใจดี, *เอกสารประกอบการเรียนวิชาคณิตสถิติศาสตร์*, ปีการศึกษา 2560

[6] J Miller, S. "How a simple observation from the 1800s about patterns in big data sets can fight fraud".2015. [ออนไลน์]. เข้าถึงได้จาก : https://theconversation.com/how-a-simple-observation-from-the-1800s-about-patterns-in-big-data-sets-can-fight-fraud-42683. [สืบค้นเมื่อ 9 พฤศจิกายน 2561]

[7] "uniroot solution in R.".2016. [ออนไลน์]. เข้าถึงได้จาก : https://stackoverflow.com/questions/38961221/uniroot-solution-in-r.[สืบค้นเมื่อ 14 มีนาคม 2562]

# ภาคผนวก ก

แบบเสนอหัวข้อโครงงาน รายวิชา 2301399 Project Proposal

**ปีการศึกษา 2561**

| | |
|---|---|
| **ชื่อโครงงาน (ภาษาไทย)** | การประมาณค่าของพารามิเตอร์ที่บ่งชี้ระยะห่างจากกฎของเบนฟอร์ด |
| **ชื่อโครงงาน (ภาษาอังกฤษ)** | Estimation of parameters reflecting distance from Benford's law |
| **อาจารย์ที่ปรึกษาโครงงาน** | 1. รศ. ทิพวัลย์ สันติวิภานนท์<br>2. รศ.ดร. ทรงเกียรติ สุเมธกิจการ |
| **ผู้ดำเนินการ** | นางสาวกมลรัตน์ ผลคิมหันต์ เลขประจำตัวนิสิต 5833501723<br>สาขาวิชา คณิตศาสตร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์<br>คณะวิทยาศาสตร์ |

---

**หลักการและเหตุผล**

กฎของเบนฟอร์ด (Benford's law) กล่าวว่าความถี่ในการเกิดเลขหลักแรก (first digit) เป็นเลข 1 ถึง 9 ไม่เท่ากัน โดยความน่าจะเป็นที่จะเกิดเลขหลักแรกเป็น 1 มีมากที่สุด และความน่าจะเป็นที่เกิดเลขหลักแรกเป็น 2,3,4,...,9 ก็จะน้อยลงเรื่อย ๆ ตามลำดับ ซึ่งเราสามารถนำหลักการจากกฎของเบนฟอร์ดมาตรวจจับความผิดปกติเบื้องต้นของข้อมูลได้ แต่เราต้องศึกษาก่อนว่าข้อมูลเหล่านั้นมีการแจกแจงแบบเบนฟอร์ดหรือไม่ ด้วยการทดสอบไคสแควร์ (Chi-Square Test)

จากการศึกษาบทความ Benfordness of chains of truncated Beta distributions via a piecewise constant approximation เราได้กลุ่มของการแจกแจงของเลขหลักแรกมาหนึ่งกลุ่มนั่นคือ $G_{\alpha,q}(s)$ โดยที่ $\alpha > 0$, $q$ เป็นจำนวนนับ และ $1 \leq s \leq 10$ ซึ่ง $G_{\alpha,q}(s)$ ลู่เข้าสู่กฎของเบนฟอร์ดนั่นคือ $G_{\alpha,q}(s) \to \log_{10}(s)$ เมื่อ $q \to \infty$ และ $G_{\alpha,q}(s) = \log_{10}(s)$ เมื่อ $\alpha = 1$ สำหรับแต่ละ $q$ เราคิดว่าการแจกแจงกลุ่มนี้น่าจะเป็นแบบจำลองของเลขหลักแรกในธรรมชาติได้ดี จึงหาตัวประมาณค่าของพารามิเตอร์ $\alpha$ โดยวิธีความควรจะเป็นสูงสุด (Maximum Likelihood) เพื่อบ่งชี้ระยะห่างของพารามิเตอร์ $\alpha$ กับ 1 หากทำการศึกษาแล้วพบว่าพารามิเตอร์ $\alpha$ มีค่าใกล้เคียง 1 มาก ๆ จะนำผลการศึกษานี้ไปใช้ในการทดสอบความเป็นเบนฟอร์ดของข้อมูลค่าไฟฟ้ารายเดือน

**วัตถุประสงค์**

ต้องการหาตัวประมาณของพารามิเตอร์ที่บ่งชี้ระยะห่างจากกฎของเบนฟอร์ด และนำผลของ

การประมาณค่าพารามิเตอร์นี้ไปใช้ในการทดสอบความเป็นเบนฟอร์ดของข้อมูลค่าไฟฟ้ารายเดือน

**ขอบเขตของโครงงาน**

1. ประมาณค่าพารามิเตอร์ $\alpha$ โดยวิธีความควรจะเป็นสูงสุด (Maximum Likelihood)
2. พิจารณาพารามิเตอร์ $q$ ในช่วงช่วงหนึ่ง
3. นำผลการประมาณค่าพารามิเตอร์ไปใช้ในการทดสอบความเป็นเบนฟอร์ดของข้อมูลค่าไฟฟ้ารายเดือน

**วิธีการดำเนินงาน**

ก. แผนการศึกษา

1. 1. ศึกษาบทความที่เกี่ยวข้อง ประกอบไปด้วย 3 บทความ ดังนี้
   (a) Benfordness of chains of truncated Beta distributions via a piecewise constant approximation
   (b) Benford's law for chains of truncated distribution
   (c) A generalized Benford's law and its application
2. หาตัวประมาณของพารามิเตอร์ $\alpha$ โดยวิธีความควรจะเป็นสูงสุด (Maximum Likelihood)
3. นำผลการศึกษามาประยุกต์ใช้ โดยการทดสอบความเป็นเบนฟอร์ดของข้อมูลค่าไฟฟ้า รายเดือน

ข. ระยะเวลาที่ศึกษา

| การดำเนินงาน | สิงหาคม 2561 - เมษายน 2562 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | ส.ค. | ก.ย. | ต.ค. | พ.ย. | ธ.ค. | ม.ค. | ก.พ. | มี.ค. | เม.ย. |
| 1.ศึกษาปัญหาและบทความที่เกี่ยวข้อง. | | | | | | | | | |
| 2.คำนวณหารูปแบบทั่วไปของตัวประมาณพารามิเตอร์ $\alpha$ ที่ขึ้นกับค่า $q$ โดยวิธีความควรจะเป็นสูงสุด | | | | | | | | | |
| 3.นำรูปแบบทั่วไปในข้อ 2 มาคำนวณหาค่า $\alpha$ ของแต่ละ $q$ โดยใช้โปรแกรม Wolfram Mathematica | | | | | | | | | |
| 4.นำผลของการประมาณค่าพารามิเตอร์ $\alpha$ มาใช้ในการทดสอบความเป็นเบนฟอร์ดของข้อมูลค่าไฟฟ้ารายเดือน | | | | | | | | | |
| 5.สรุปผลและจัดทำรูปเล่มโครงงาน | | | | | | | | | |
| 6.ตรวจสอบรูปเล่มโครงงาน และจัดเตรียมการนำเสนอโครงงาน | | | | | | | | | |

**ประโยชน์ที่คาดว่าจะได้รับ**

ก. ด้านความรู้และประสบการณ์ต่อผู้จัดทำ

1. นำความรู้คณิตศาสตร์พื้นฐานและความรู้คณิตศาสตร์ทางด้านความน่าจะเป็นและสถิติมา

ประยุกต์ใช้ในการแก้ปัญหา

2. ได้ฝึกทักษะการทำงานวิจัยทางคณิตศาสตร์

3. ฝึกความรับผิดชอบและความตรงต่อเวลา

ข. ความรู้ ความเข้าใจที่นำไปสู่การแก้ไขปัญหาของสังคมหรือสภาพแวดล้อม

นำผลจากการศึกษาโครงงานนี้ไปประยุกต์ใช้ในการตรวจจับความผิดปกติเบื้องต้นของข้อมูลที่มีการทุจริตได้

**อุปกรณ์และเครื่องมือที่ใช้**

ก. วัสดุอุปกรณ์

1. กระดาษ A4 และเครื่องเขียน

2. คอมพิวเตอร์

3. อุปกรณ์จัดเก็บข้อมูล (Flash drive)

ข. โปรแกรม (Soft ware)

1. Wolfram Mathematica

2. Microsoft Word 2013

3. Text editor by www.overleaf.com)

**งบประมาณ**

1. กระดาษขนาด A4 80 แกรม

2. อุปกรณ์จัดเก็บข้อมูล (Flash drive)

3. เมาส์คอมพิวเตอร์

# ภาคผนวก ข

## การคำนวณหาตัวประมาณพารามิเตอร์ $\alpha$ ในโปรแกรม $R$

ในส่วนของการคำนวณหาค่าประมาณของพารามิเตอร์ $\alpha$ โดยโปรแกรม *Wolfram Mathematica* ที่กล่าวไว้ในแบบนำเสนอโครงงาน ในทางปฏิบัตจริง ปรากฏว่าโปรแกรมต้องใช้เวลานานอย่างมากในการคำนวณหาคำตอบ เราจึงหันมาใช้โปรแกรม $R$ ในการคำนวณหาค่าประมาณของพารามิเตอร์ $\alpha$ โดยใช้คำสั่งที่คำนวณหารากของสมการโดยวิธีแบ่งครึ่งช่วง (bisection method) ที่มีชื่อว่า *uniroot* ในคำนวณหารากของสมการ (2.1) ซึ่งผลลัพธ์ที่ได้ก็คือตัวประมาณของพารามิเตอร์ $\alpha$ นั่นเอง เราจะสร้างชุดคำสั่งที่ให้ผลลัพธ์ออกมาเป็นค่าประมาณของพารามิเตอร์ $\alpha$ สำหรับแต่ละ $q$ และ $N_d$ ของแต่ละข้อมูล ที่มีชื่อว่า "Alpha" ดังนี้

```
Alpha = function(N,q,alphamax) {
  b = function(d,q) floor(q*log10(d))
  a = function(d,q) (d*10^(-b(d,q)/q) -1)/(10^(1/q) -1)
  f = function(x,q,d) {
    D1 = a(d+1,q)/((10^(x/q) -1)*a(d+1,q) +1) - a(d,q)/((10^(x/q) -1)*a(d,q) +1)
    D1 = (10^(x/q)/(x*q))*D1
    D2 = (1/x^2)*log10(((10^(x/q) -1)*a(d+1,q) +1)/((10^(x/q) -1)*a(d,q) +1))
    N1 = (b(d+1,q)-b(d,q))/q
    N2 = (1/x)*log10(((10^(x/q) -1)*a(d+1,q) +1)/((10^(x/q) -1)*a(d,q) +1))
    T0 = N[d]*(D1-D2)/(N1+N2)
    return(T0)
  }

  l = function(x,q,d) {
    N1 = (b(d+1,q)-b(d,q))/q
    N2 = (1/x)*log10(((10^(x/q) -1)*a(d+1,q) +1)/((10^(x/q) -1)*a(d,q) +1))
    L = N[d]*log((N1+N2))
  }

  F= function(x) sum(f(x,q,seq(1,9,1)))
  uniroot(F,c(0.0001,alphamax),tol=0.00000001,maxiter=10)$root
}
```

จากภาพ จะเห็นได้ว่า ในช่วงบรรทัดที่ 5 ถึง 12 เป็นการสร้างฟังก์ชัน $f$ นั่นก็คือสมการ (2.1) ที่ยังไม่ได้ถูกคำนวณผลรวมสำหรับเลขหลักแรก $d$ ตั้งแต่ 1 ถึง 9 และถูกแบ่งส่วน เพื่อความสะดวกในการคำนวณ ซึ่งเราจะแทนผลลัพธ์ที่ได้จากการแทนค่า $N_d$, $q$ และขอบเขตบนของค่า $\alpha$ ลงในฟังก์ชัน $f$ ด้วยตัวแปร $x$ นั่นก็คือตัวประมาณของพารามิเตอร์ $\alpha$ แต่ยังไม่ทราบค่า เพราะยังคงอยู่ในรูปตัวแปร $d$ เราจึงให้ฟังก์ชัน $F$ เป็นผลรวมของฟังก์ชัน $f$ ในกรณีที่ $d = 1$ ถึง 9 นั่นคือสมการ (2.1) นั่นเอง

ขั้นตอนสุดท้ายของชุดคำสั่ง "Alpha" จะเป็นการหารากของฟังก์ชัน $F$ โดยใช้คำสั่ง $uniroot$

ที่หารากของสมการโดยใช้วิธีการแบ่งครึ่งช่วง เราจึงต้อง ระบุขอบเขตบนและขอบเขตล่างของราก ระดับนัยสำคัญของราก และจำนวนครั้งที่ต้องการแบ่งครึ่งช่วง ตามลำดับ ดังนั้น เมื่อเราแทนค่า $N_d$ ของแต่ละข้อมูล , ค่า $q$ และค่าขอบเขตบนของ $\alpha$ ลงในชุดคำสั่ง "Alpha" จะได้ผลลัพธ์เป็นค่าประมาณของ $\alpha$ สำหรับแต่ละกรณีที่เราต้องการ

ลำดับต่อไปเราจะนำค่าประมาณของ $\alpha$ สำหรับ $q$ แต่ละกรณี ที่คำนวณได้จากชุดคำสั่ง "Alpha" มาคำนวณหาค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(s)$ โดยการเริ่มจากการนิยามฟังก์ชันความน่าจะเป็น $g_{\alpha,q}$ ด้วยฟังก์ชัน $g$ ในโปรแกรม $R$ ดังนี้

```
18 l = function(x,q,d)
19 {
20   N1 = (b(d+1,q)-b(d,q))/q
21   N2 = (1/x)*log10(((10^(x/q) -1)*a(d+1,q) +1)/((10^(x/q) -1)*a(d,q) +1))
22   L = N[d]*log((N1+N2))
23 }
24
25 g= function(x) sum(l(x,q,seq(1,9,1)))
26 g(Alpha(N,q,alphamax))
```

เมื่อเราแทนค่าประมาณของพารามิเตอร์ $\alpha$ สำหรับแต่ละ $q$ ของแต่ละข้อมูล ที่ได้จากชุดคำสั่ง "Alpha" ลงในฟังก์ชัน $g$ ก็จะได้ค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(s)$ สำหรับแต่ละกรณีดังต้องการ

ตัวอย่างเช่น เราต้องการหาค่าประมาณของพารามิเตอร์ $\alpha$ ของข้อมูลค่าพลังงานของการแผ่รังสีของวัตถุสีดำ ในกรณีที่ $q = 2$ พร้อมทั้งหาค่าของฟังก์ชันความน่าจะเป็น $g_{\alpha,q}(s)$ สามารถทำได้โดยเริ่มจากการสร้างเซตที่มีสมาชิกเป็นจำนวนข้อมูลที่มีเลขหลักแรกเป็น $d$ ของข้อมูลค่าพลังงานของการแผ่รังสีของวัตถุดำ ที่ $d$ เรียงลำดับจาก 1 ถึง 9 ซึ่งเราให้ชื่อเซตนี้ว่า "black" จากนั้นแทนชื่อเซต black ,1 และ 1.5 ซึ่งเป็นขอบเขตบนของค่า $\alpha$ ลงในชุดคำสั่ง "Alpha" และฟังก์ชัน $g$ ตามลำดับ จะได้ผลลัพธ์ดังนี้

```
> black = c(31,17.3,14.1,8.7,6.6,7.0,5.2,4.7,5.4)
> Alpha(black,1,1.5)
[1] 1.037154
> g(Alpha(black,1,1.5))
[1] -188.7793
```

# ประวัติผู้เขียน

นางสาวกมลรัตน์ ผลคิมหันต์
รหัสประจำตัวนิสิต 5833501723
ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย