

LAB 7.5 CPE213 Data Models

62070501064 อรวิภา คุณเจริญไพศาล

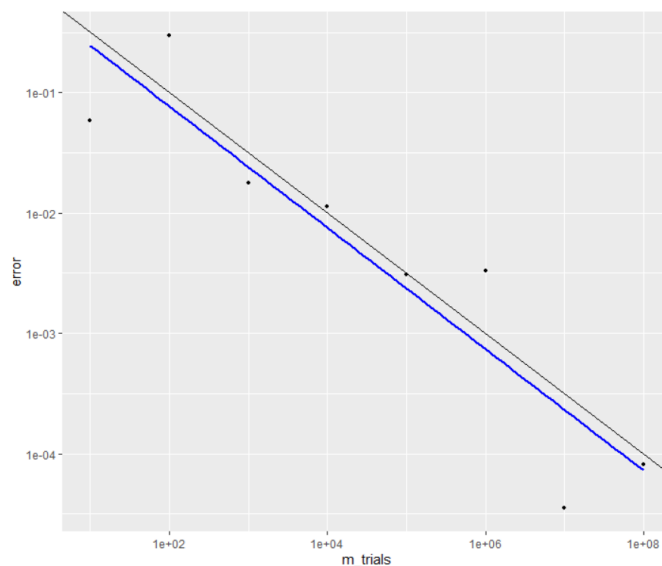
Task 1

จากการประมาณค่าทั้งหมด 10,100,1000,10000,100000,1000000,10000000,100000000 รอบ จากโค้ด

```
library(tidyverse)
# Simple Pi
n_hits <- 0
m_trials <- 100000000
counter <- 0
while(counter < m_trials){
  x <- runif(n = 1, min = 0, max = 1)
  y <- runif(n = 1, min = 0, max = 1)
  if(x**2 + y**2 < 1){
    n_hits <- n_hits + 1
  }
  counter <- counter + 1
}
pie <- 4*n_hits/m_trials
pie

#error <-c()
error <-append(error, abs(pi - pie))
error
ggplot() + geom_point(aes(x = c(10,100,1000,10000,100000,1000000,10000000,100000000), y = error), size = 1) +
  scale_x_log10() + scale_y_log10() + geom_abline(slope = -0.5) + xlab("m_trials") +
  geom_smooth(aes(x = c(10,100,1000,10000,100000,1000000,10000000,100000000), y = error),method = 'lm', color = 'blue', se = FALSE)
```

เมื่อเอามาพลอตกราฟและใช้ geom_smooth ในการพลอตแนวโน้มของค่า pi ที่ได้จากการเปลี่ยนจำนวนรอบในการประมาณค่า จะเห็นได้ดังรูป



จะได้ว่า ยิ่งเพิ่มจำนวนรอบในการประมาณค่า จะทำให้มี error น้อยลง ซึ่งจะสังเกตได้จากแนวโน้มของเส้นสีน้ำเงิน (geom_smooth) และเส้นสีดำ (เส้นที่มีความชันเป็น -0.5)

Task 2

Jimmy prints a neighborhood newspaper with 10 subscribers. He also sells it to whoever comes by from his front lawn on Friday afternoons. His mother has kept track of his demand (including requests made after he had sold out) for the past 100 weeks, and observed the pattern shown below.

Papers Demanded	Number of weeks	Probability	Cumulative Probability
13	1	0.01	0.01
14	2	0.02	0.03
15	4	0.04	0.07
16	9	0.09	0.16
17	10	0.10	0.26
18	15	0.15	0.41
19	16	0.16	0.57
20	15	0.15	0.72
21	12	0.12	0.84
22	9	0.09	0.93
23	4	0.04	0.97
24	2	0.02	0.99
25	1	0.01	1.00
26	0	0.00	1.00
Total	100	1	

The papers cost 30 cents to print and Jimmy sells them for 50 cents. Assume that he prints 20 copies a week. Mom makes him throw away unsold copies. Simulate his sales for the next 12 weeks and determine his earnings.

Week	Random number	Number Demanded	Number Sold	Number Thrown	Revenue	Cost	Profit
1	0.4175	19	19	1	950	600	350
2	0.8434	22	20	0	1000	600	400
3	0.5227	19	19	1	950	600	350
4	0.1624	17	17	3	850	600	250
5	0.0149	14	14	6	700	600	100
6	0.0067	13	13	7	650	600	50
7	0.0957	16	16	4	800	600	200
8	0.6233	20	20	0	1000	600	400
9	0.9990	25	20	0	1000	600	400
10	0.0391	15	15	5	750	600	150
11	0.2901	18	18	2	900	600	300
12	0.0779	16	16	4	800	600	200
Total							3150

Should Jimmy increase his paper copies to 30? Would it increase the earning? Why?

Week	Random number	Number Demanded	Number Sold	Number Thrown	Revenue	Cost	Profit
1	0.4175	19	19	11	950	900	50
2	0.8434	22	22	8	1100	900	200
3	0.5227	19	19	11	950	900	50
4	0.1624	17	17	13	850	900	-50
5	0.0149	14	14	16	700	900	-200
6	0.0067	13	13	17	650	900	-250
7	0.0957	16	16	14	800	900	-100
8	0.6233	20	20	10	1000	900	100
9	0.9990	25	25	5	1250	900	350
10	0.0391	15	15	15	750	900	-150
11	0.2901	18	18	12	900	900	0
12	0.0779	16	16	14	800	900	-100
Total							-100

Answer: จากตารางด้านบนจะเห็นได้ว่าเมื่อเพิ่มจำนวนเป็น 30 ฉบับ และคำนวณกำไรที่ได้รวมทั้ง 12 สัปดาห์แล้ว ปริมาณกำไรรวมที่จิมมี่จะได้ จะกลายเป็นติดลบหรือก็คือขาดทุนนั่นเอง ดังนั้นจิมมี่ไม่ควรเพิ่มเป็น 30 ฉบับ

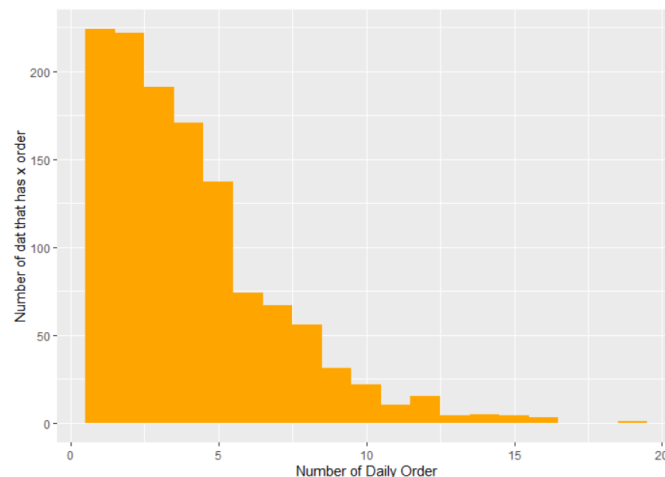
Task 3

1. Analyze the distribution of daily demand of each order.

Code

```
# 1.
data %>%
  select(Order.ID, Order.Date) %>%
  distinct() %>%
  group_by(Order.Date) %>%
  summarise(n = n()) %>%
  ggplot() + geom_histogram(aes(x = n), binwidth = 1, fill = 'orange') +
  xlab('Number of Daily Order') + ylab('Number of dat that has x order')
```

Graph



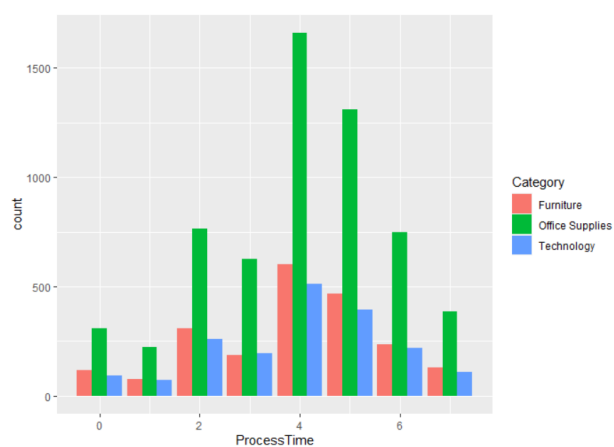
จากกราฟ จะเห็นได้ว่ากราฟเป็นลักษณะปัวส์ซอง หรือก็คือมีค่าที่สูงในช่วงแรกนั่นเอง

2. Analyze the time between order of different product category.

Code

```
# 2.
data %>%
  mutate(ProcessTime = as.Date(Ship.Date) - as.Date(Order.Date)) %>%
  group_by(Category) %>%
  ggplot(aes(x = ProcessTime, fill = Category)) + geom_bar(position = 'dodge')
```

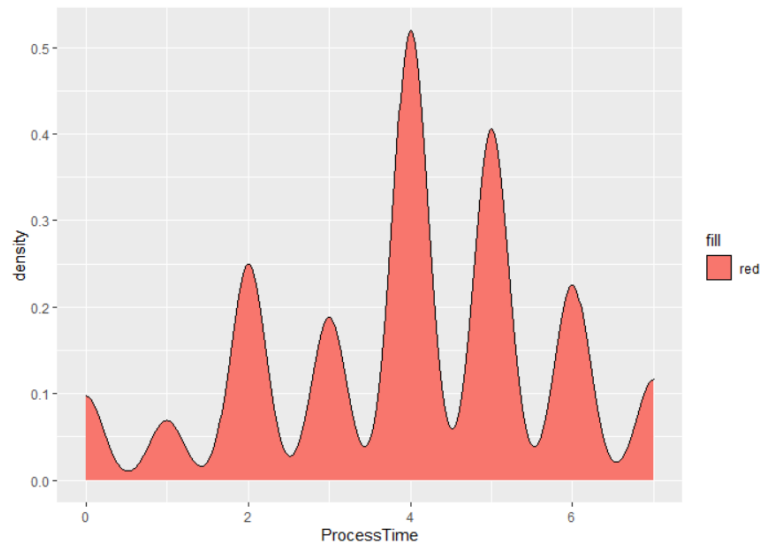
Graph



จากกราฟแสดงให้เห็นว่า ในภาพรวมแล้วจำนวนวันในการจัดการ (ระยะเวลาตั้งแต่สั่งสินค้าจนถึงวันส่งสินค้า) ของสินค้าแต่ละประเภทจะเท่า ๆ กัน โดยระยะเวลาสูงสุดคือ 4-5 วัน ต่างกันเพียงจำนวนสินค้าที่ขายได้ในแต่ละประเภทเท่านั้น

- Analyze the distribution of order processing time of each order.

```
# 3.
data %>%
  mutate(ProcessTime = as.Date(Ship.Date) - as.Date(Order.Date)) %>%
  ggplot(aes(x = ProcessTime, fill = 'red')) + geom_density()
```



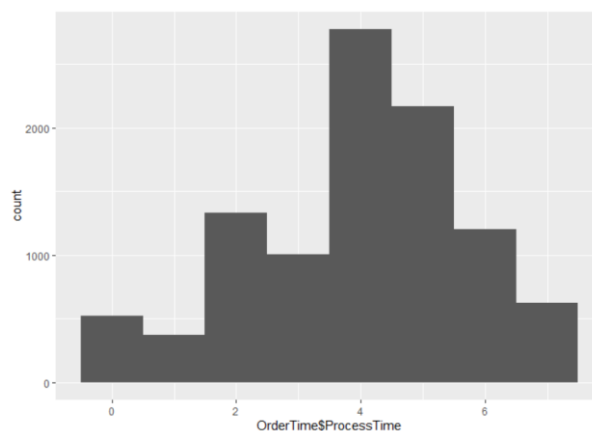
จากกราฟ จะแสดงให้เห็นว่าการกระจายตัวของจำนวนวันในการจัดการของสินค้า (ระยะเวลาตั้งแต่สั่งสินค้าจนถึงวันส่งสินค้า) ส่วนใหญ่จะกระจุกตัวอยู่ที่ประมาณ 4 – 5 วัน

- Simulate the superstore processing time.

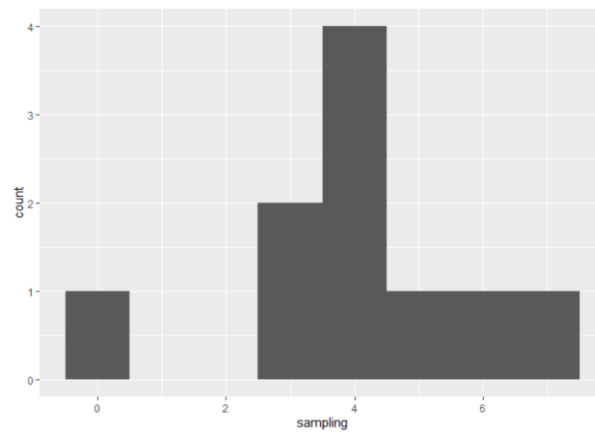
Code

```
# 4.
data %>%
  mutate(ProcessTime = as.Date(Ship.Date) - as.Date(Order.Date)) %>%
  select(ProcessTime) -> OrderTime
# Sampling
quantile(OrderTime$ProcessTime, runif(1000)) -> y
sampling <- c()
sampling <- append(sampling, as.numeric(y))
ggplot() + geom_histogram(aes(x = sampling), binwidth = 1)
# raw data
ggplot() + geom_histogram(aes(x = OrderTime$ProcessTime), binwidth = 1)
```

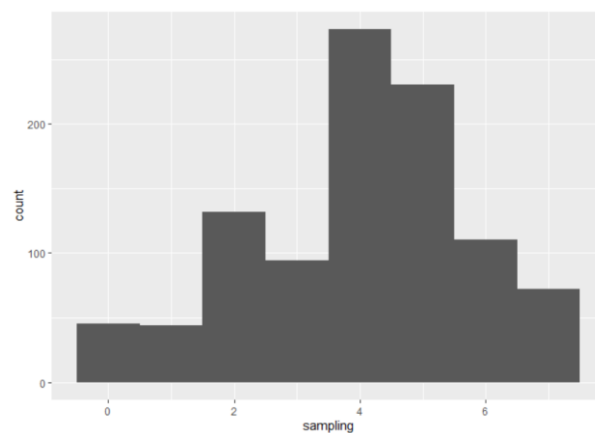
ข้อมูลเดิม



Number of Run = 10



Number of Run = 1000



จากกราฟทั้ง 3 ภาพ จะเห็นว่าเมื่อเทียบกับข้อมูลเดิม หากเพิ่ม Number of run หรือจำนวนการสุ่มข้อมูลให้มากขึ้น กราฟจะมีลักษณะใกล้เคียงกับกราฟข้อมูลเดิมมากขึ้น ซึ่งก็จะมีระยะเวลาดังแต่ส่งสินค้าจนถึงวันที่ส่งสินค้าที่กระจายตัวตั้งแต่ 0-8 วัน โดยข้อมูลส่วนใหญ่จะกระจุกตัวอยู่ที่ประมาณ 4-6 วันนั่นเอง

5. Is there relationship between order size and average processing time?

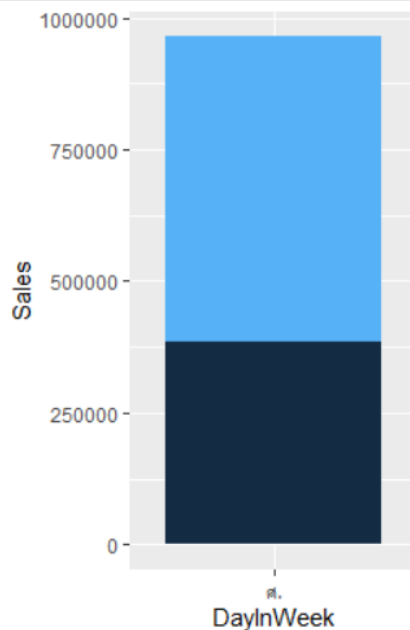
ปริมาณสินค้าในแต่ละออเดอร์จะมีความสัมพันธ์กับระยะเวลาตั้งแต่วันที่ส่งออเดอร์จนถึงวันที่ส่งสินค้า โดยที่หากปริมาณสินค้ามากขึ้นจะส่งผลให้มีระยะเวลาในการจัดการที่เพิ่มขึ้น

6. What would happen, if every Friday, the number of orders is boosted by 50%?

```
# 6.
data %>%
  mutate(DayInWeek = format(as.Date(Order.Date), format= "%a")) %>%
  group_by(DayInWeek) %>%
  summarise(n = n(), Sales = sum(Sales)) %>%
  filter(DayInWeek == 'ศ.') -> oldFri

data %>%
  mutate(DayInWeek = format(as.Date(Order.Date), format= "%a")) %>%
  group_by(DayInWeek) %>%
  summarise(n = n(), Sales = sum(Sales)) %>%
  mutate(Sales = ifelse(DayInWeek == 'ศ.', Sales * 1.5, Sales)) %>%
  filter(DayInWeek == 'ศ.') -> newFri

rbind(oldFri, newFri) %>%
  ggplot() + geom_col(aes(x = DayInWeek, y = Sales, fill = Sales))
```



สีฟ้า = วันศุกร์ที่เพิ่มยอดขาย 50%, สีน้ำเงิน = วันศุกร์เดิม

จากกราฟ จะแสดงให้เห็นว่าถ้าหากมีการเพิ่มออเดอร์ในวันศุกร์เป็น 50 % จะมียอดขายเพิ่มขึ้นกว่าเดิมเกือบ 2 เท่า