



เรื่อง การวิเคราะห์ข้อมูลรายรับรายจ่ายและข้อมูลการเดินทาง

จัดทำโดย

นายชิษณุพงศ์ วรวิจิตรชัยกุล รหัสประจำตัว 6010502543

เสนอ

ผศ.ดร.สุภาพร เอื้อจงมานี

รายงานนี้เป็นส่วนหนึ่งของการเรียนวิชา

01204314 สถิติสำหรับการประยุกต์ทางวิศวกรรมคอมพิวเตอร์

มหาวิทยาลัยเกษตรศาสตร์

คำนำ

รายงานฉบับนี้เป็นส่วนหนึ่งของของวิชาสถิติสำหรับการประยุกต์ทางวิศวกรรมคอมพิวเตอร์ โดยมีจุดประสงค์เพื่อให้ผู้จัดทำได้ฝึกการรวบรวมข้อมูล และนำข้อมูลที่ได้มาสร้างเป็นกราฟเพื่อวิเคราะห์ข้อมูล รวมทั้งการตั้งสมมติฐานทางสถิติ การทดสอบสมมติฐาน และการสรุปผลการทดสอบสมมติฐาน ซึ่งเป็นการประยุกต์ใช้หลักวิชาสถิติที่สำคัญ

ทั้งนี้ทางผู้จัดทำหวังเป็นอย่างยิ่งว่ารายงานฉบับนี้จะสามารถเป็นประโยชน์ต่อผู้ที่เข้ามาศึกษาไม่มากนักน้อย และหากมีข้อผิดพลาดประการใด ทางผู้จัดทำต้องขออภัยมา ณ ที่นี้ด้วย

นาย ชิณพงศ์ วรวิจิตรชัยกุล

ผู้จัดทำ

สารบัญ

| หัวข้อ | หน้า |
|---|-------|
| การเก็บข้อมูล | 3 |
| Visualization | |
| การเตรียมข้อมูลสำหรับทำ visualization | 4 |
| Time-series | 5-6 |
| Part-to-whole | 7 |
| Deviation | 8-10 |
| Hypothesis test | |
| การเตรียมข้อมูลสำหรับทำ Hypothesis test | 11-12 |
| Hypothesis test | 13-14 |
| Anova on single factor | |
| การเตรียมข้อมูลสำหรับทำ Anova on single factor | 15 |
| Anova on single factor | 16 |
| Anova on two factor | |
| การเตรียมข้อมูลสำหรับทำ Anova on two factor | 17 |
| Anova on two factor | 18 |
| Categorical data analysis | |
| การเตรียมข้อมูลสำหรับทำ Categorical data analysis | 19 |
| Categorical data analysis | 20 |
| สรุป | 21 |

การเก็บข้อมูล

ข้อมูลที่นำมาใช้ในการวิเคราะห์ คือข้อมูลค่าใช้จ่ายทั้งหมดของนายชิษณุพงศ์
วรวิจิตรชัยกุล ตั้งแต่วันพฤหัสบดีที่ 9 มกราคม 2563 ถึงวันพุธที่ 22 มกราคม 2563
รวมเป็นระยะเวลา 14 วัน

| Date | รายการ | รายรับ (บาท) | รายจ่าย (บาท) | ระยะทาง (กิโลเมตร) | จุดเริ่มต้น | จุดหมาย | หมายเหตุ |
|-----------------------|------------------------------------|--------------|---------------|--------------------|-------------|---------|--------------------------|
| 01/01/2020 | เงินเดือน | 16000 | | | | | |
| 01/1/2020 - 08/1/2020 | ค่าใช้จ่ายก่อนวันที่ 9 (โดยประมาณ) | | 9400 | | | | *รวมค่าหอ, ค่าน้ำ, ค่าไฟ |
| 09/01/2020 | รถเมล์ | | 13 | 2.1 | คอนโด | ม.เกษตร | |
| | ข้าวหมูกรอบ | | 52 | | | | |
| | น้ำเปล่า | | 7 | | | | |
| | ค่าสนามบอล | | 121 | | | | |
| | รถเมล์ | | 15 | 2.2 | ม.เกษตร | คอนโด | |
| | ข้าวขาหมู | | 35 | | | | |
| 10/01/2020 | น้ำแข็ง | | 18 | | | | |
| | ยานพาหนะ | | 60 | | | | |
| | ข้าวขาหมู | | 35 | | | | |
| 11/01/2020 | เกสเซอร์ | | 40 | | | | |
| | ข้าวกะเพราหมูกรอบไข่ดาว | | 80 | | | | |
| | ปูอัดผัดกะหล่ำราดข้าว | | 85 | | | | |
| | ไอติม | | 113 | | | | |
| | มาน้ำ | | 24 | | | | |
| | ไข่ออนเซ็น | | 17 | | | | |
| 12/01/2020 | ยานพาหนะ | | 60 | | | | |
| | หมูย่าง | | 40 | | | | |
| 13/01/2020 | รถเมล์ | | 13 | 2.1 | คอนโด | ม.เกษตร | |
| | รถเมล์ | | 15 | 2.2 | ม.เกษตร | คอนโด | |
| | ข้าวกะเพราหมูกรอบไข่ดาว | | 80 | | | | |

รูปที่ 1 ตัวอย่างข้อมูลจาก Google sheet : 6010502543_Assignment_1 (Collecting_data)

Visualization

จากข้อมูลในรูปที่ 1 นำมาจัดเรียงใหม่ให้เหมาะสำหรับนำไปใช้ในการวิเคราะห์ โดยแบ่งรายจ่ายออกเป็น 3 ส่วน คือ 1. ค่าอาหาร 2. ค่าเดินทาง และ 3. ค่าใช้จ่ายอื่นๆ เช่น ค่าอุปกรณ์ทำความสะอาด ค่าตัดผม ค่าสนามฟุตบอล ฯลฯ

| | Day | Food expenses | Travel expenses | Other expenses |
|----|-----|---------------|-----------------|----------------|
| 0 | 9 | 94 | 28 | 121 |
| 1 | 10 | 113 | 0 | 0 |
| 2 | 11 | 359 | 0 | 0 |
| 3 | 12 | 100 | 0 | 0 |
| 4 | 13 | 80 | 28 | 0 |
| 5 | 14 | 372 | 28 | 0 |
| 6 | 15 | 130 | 29 | 0 |
| 7 | 16 | 176 | 28 | 69 |
| 8 | 17 | 590 | 31 | 0 |
| 9 | 18 | 460 | 16 | 657 |
| 10 | 19 | 137 | 15 | 0 |
| 11 | 20 | 127 | 27 | 0 |
| 12 | 21 | 210 | 27 | 1000 |
| 13 | 22 | 157 | 15 | 0 |

รูปที่ 2 ข้อมูลจาก Google sheet : 6010502543_Assignment_1 (Data_visualization)



รูปที่ 3 กราฟ Time_series_FoodExpenses

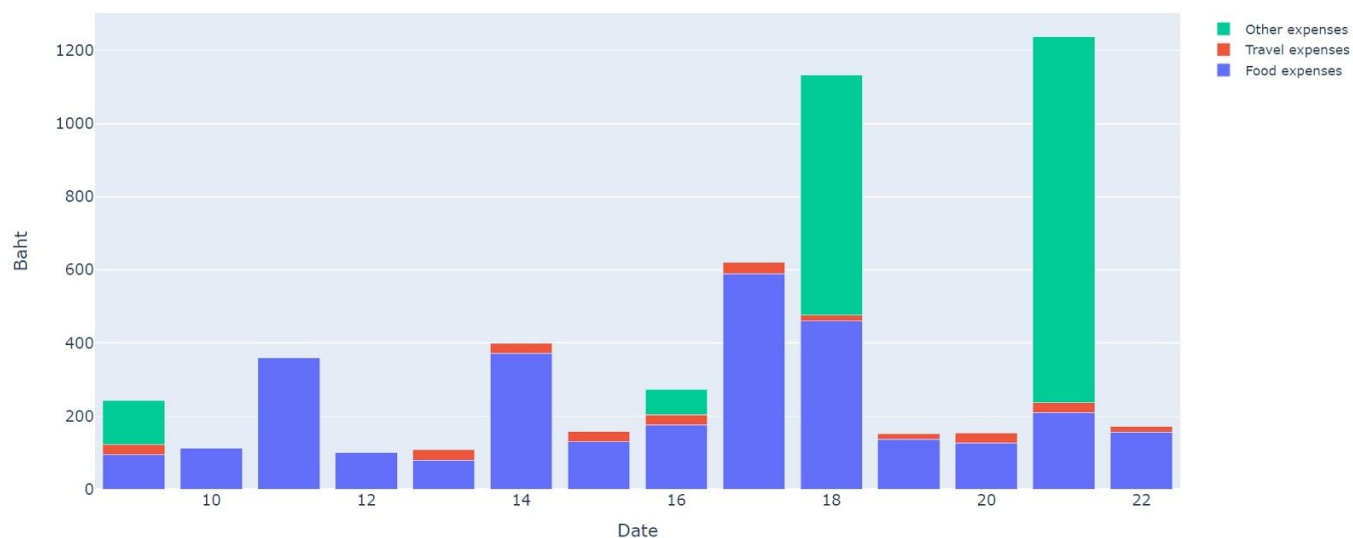
- กราฟแสดงค่าอาหารตั้งแต่วันที่พฤหัสบดีที่ 9 ถึงวันพุธที่ 22 มกราคม 2563
- กราฟ time-series แสดงให้เห็นถึงแนวโน้มของค่าใช้จ่ายในเรื่องอาหารที่เปลี่ยนแปลงไปในแต่ละวันเป็นระยะเวลา 14 วัน
- จากกราฟแสดงให้เห็นถึงค่าใช้จ่ายในเรื่องอาหารที่ไม่คงที่ในแต่ละวัน บางวันมีค่าใช้จ่ายเพียง 100 บาท ในขณะที่บางวันค่าใช้จ่ายกลับสูงถึง 600 บาท แต่จะสังเกตเห็นว่าในวันที่ 17 และ 18 มกราคม มีค่าใช้จ่ายสูงกว่าในวันอื่นๆ เนื่องจากเป็นวันศุกร์และเสาร์ ซึ่งเป็นวันที่นิสิตไม่มีเรียนและกลับไปอยู่ที่บ้านทำให้มีค่าใช้จ่ายสูงกว่าปกติ



รูปที่ 4 กราฟ Time_series_TravelExpenses

- กราฟแสดงค่าเดินทางตั้งแต่วันที่พฤหัสบดีที่ 9 ถึงวันพุธที่ 22 มกราคม 2563
- กราฟ time-series แสดงให้เห็นถึงแนวโน้มของค่าใช้จ่ายในการเดินทางที่เปลี่ยนแปลงไปในแต่ละวันเป็นระยะเวลา 14 วัน
- จากกราฟแสดงให้เห็นถึงค่าใช้จ่ายในการเดินทางที่แตกต่างกันไปในแต่ละวัน ซึ่งแบ่งออกเป็น 3 แบบ คือ 1. ประมาณ 30 บาท 2. ประมาณ 15 บาท และ 3. ไม่มีค่าใช้จ่ายในการเดินทาง โดยแบบที่ 1 เกิดจากค่าใช้จ่ายทั้งขาไปและขากลับจากมหาวิทยาลัย แบบที่ 2 เกิดจากมีค่าใช้จ่ายแค่ขาไปเพียงอย่างเดียว ขากลับติดรถเพื่อนกลับมา และแบบที่ 3 ไม่มีค่าใช้จ่ายเนื่องจากไม่มีเรียน

Expenses between 09-22 Jan 2022



รูปที่ 5 กราฟ Part_to_Whole_Expenses

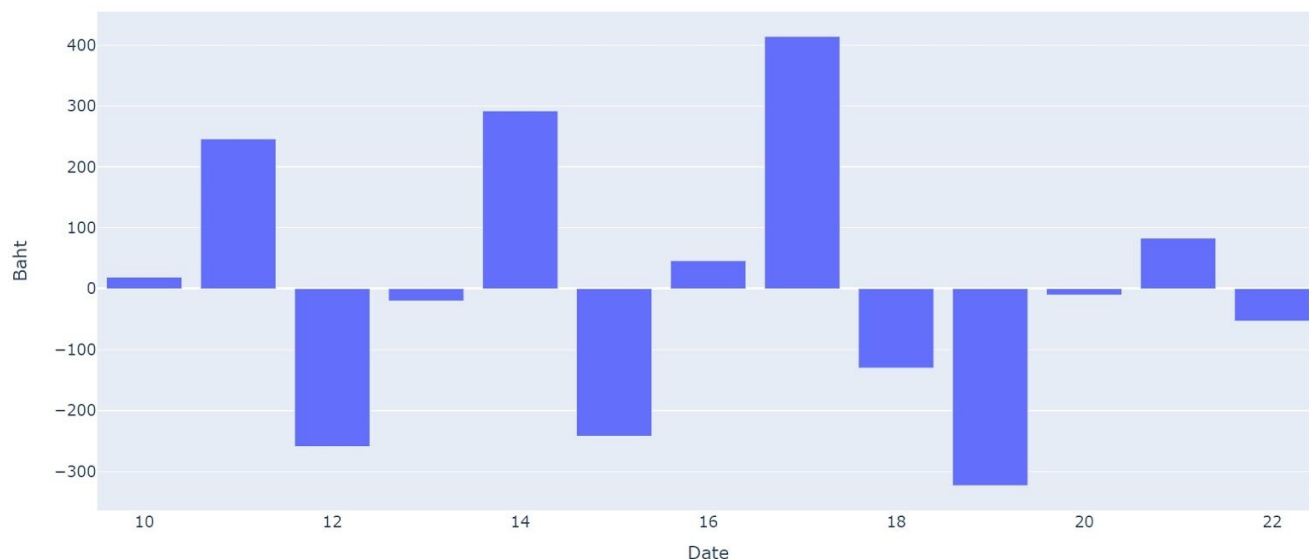
- กราฟแสดงค่าใช้จ่ายทั้งหมดตั้งแต่วันที่ 9 ถึงวันที่ 22 มกราคม 2563 โดยแบ่งรายจ่ายออกเป็น 3 ส่วน คือ 1. ค่าอาหาร 2. ค่าเดินทาง และ 3. ค่าใช้จ่ายอื่นๆ
- Stacked Bar Graph แสดงให้เห็นสัดส่วนของค่าใช้จ่ายทั้งหมดที่แตกต่างกันไปในแต่ละวันเป็นระยะเวลา 14 วัน
- จากกราฟแสดงให้เห็นถึงค่าใช้จ่ายในเรื่องอาหารที่มีทุกวันแต่อาจแตกต่างกันในเรื่องปริมาณ เนื่องจากรับประทานอาหารทุกวัน แต่อาหารไม่เหมือนกันในแต่ละวัน ค่าใช้จ่ายในการเดินทางไม่ได้มีทุกวันเนื่องจากบางวันไม่มีการเดินทาง และค่าใช้จ่ายอื่นๆที่มีแค่บางวันเนื่องจากเป็นค่าใช้จ่ายที่เพิ่มมาจากปกติ เช่น ค่าตัดผม ค่าสนามฟุตบอล

จากข้อมูลในรูปที่ 2 นำมาคำนวณหาค่าใช้จ่ายที่เปลี่ยนแปลงไปเมื่อเทียบกับวันก่อนหน้าเพื่อใช้ในการสร้างกราฟ Deviation โดยข้อมูลเริ่มต้นที่วันที่ 10 (นำวันที่ 10 เทียบกับวันที่ 9)

| | Day | Food expenses | Travel expenses |
|----|-----|---------------|-----------------|
| 0 | 10 | 19 | -28 |
| 1 | 11 | 246 | 0 |
| 2 | 12 | -259 | 0 |
| 3 | 13 | -20 | 28 |
| 4 | 14 | 292 | 0 |
| 5 | 15 | -242 | 1 |
| 6 | 16 | 46 | -1 |
| 7 | 17 | 414 | 3 |
| 8 | 18 | -130 | -15 |
| 9 | 19 | -323 | -1 |
| 10 | 20 | -10 | 12 |
| 11 | 21 | 83 | 0 |
| 12 | 22 | -53 | -12 |

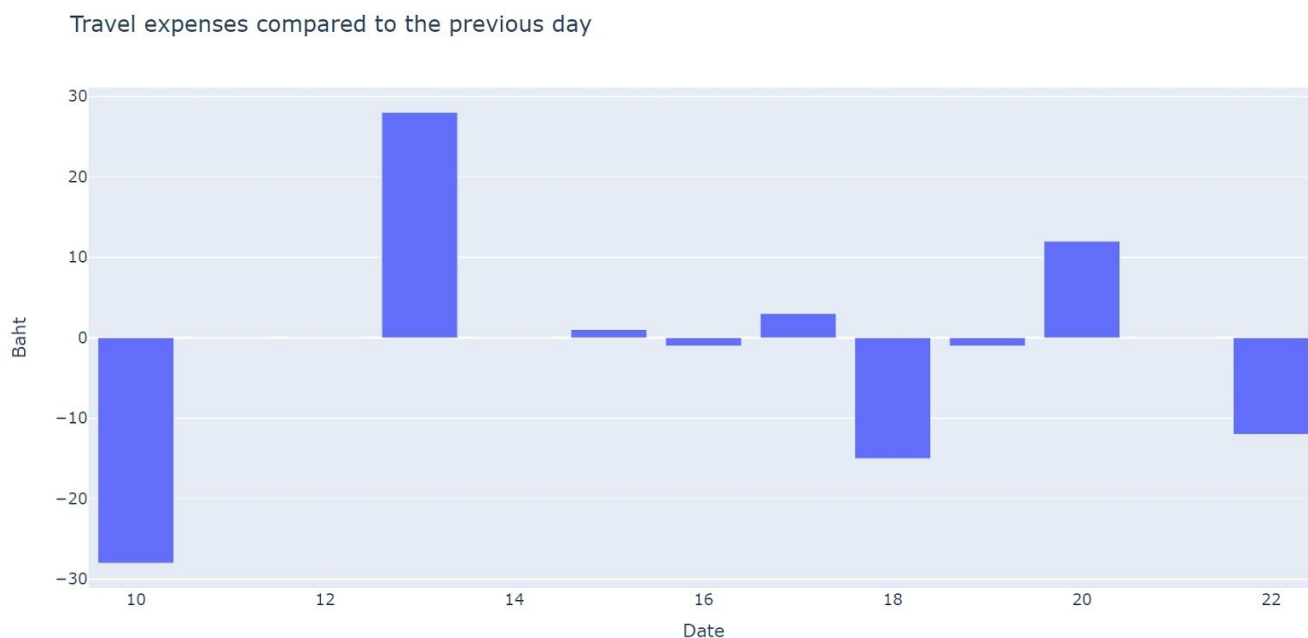
รูปที่ 6 ข้อมูลจาก Google sheet : 6010502543_Assignment_1 (Data_deviation)

Food expenses compared to the previous day



รูปที่ 7 กราฟ Deviation_FoodExpense

- กราฟแสดงค่าอาหารตั้งแต่วันศุกร์ที่ 10 ถึงวันพุธที่ 22 มกราคม 2563 ที่เปลี่ยนแปลงไปเมื่อเทียบกับวันก่อนหน้า
- Bar Graph แสดงให้เห็นค่าใช้จ่ายในเรื่องอาหารที่เพิ่มขึ้นและลดลงในแต่ละวัน
- จากกราฟแสดงให้เห็นถึงค่าใช้จ่ายในเรื่องอาหารที่มีการเพิ่มขึ้นและลดลงค่อนข้างมาก ซึ่งเกิดจากการรับประทานอาหารที่ไม่เหมือนกันในแต่ละวัน



รูปที่ 8 กราฟ Deviation_TravelExpense

- กราฟแสดงค่าเดินทางตั้งแต่วันที่ 10 ถึงวันพุธที่ 22 มกราคม 2563 ที่เปลี่ยนแปลงไปเมื่อเทียบกับวันก่อนหน้า
- Bar Graph แสดงให้เห็นค่าใช้จ่ายในการเดินทางที่เพิ่มขึ้นและลดลงในแต่ละวัน
- จากกราฟแสดงให้เห็นถึงค่าใช้จ่ายในการเดินทางในวันที่ 10 ที่ลดลงมามาก เนื่องจากในวันที่ 9 มีการเดินทางไปมหาชัยแต่วันที่ 10 ไม่มีการเดินทางไปไหน วันที่ 11 และ 12 ไม่มีการเดินทางเช่นกันทำให้ไม่เกิดการเปลี่ยนแปลงของค่าใช้จ่ายและกราฟมีค่าเป็น 0 แต่ในวันที่ 13 มีเรียนทำให้มีค่าใช้จ่ายในการเดินทาง และกราฟกลายเป็นบวก ในวันต่อๆมามีเรียนเป็นปกติจึงทำให้ค่าใช้จ่ายไม่ค่อยมีการเปลี่ยนแปลง

Hypothesis test on two data sets

จากข้อมูลในรูปที่ 2 มีการเปลี่ยนแปลงตารางข้อมูลใหม่โดยทำการเพิ่มคอลัมน์ Weekend_or_Weekday เข้ามาแยกระหว่างวันธรรมดาและวันหยุด เพื่อใช้ในการทำ Hypothesis test

| | Day | Food expenses | Travel expenses | Other expenses | Weekend_or_Weekday |
|----|-----|---------------|-----------------|----------------|--------------------|
| 0 | 9 | 94 | 28 | 121 | Weekday |
| 1 | 10 | 113 | 0 | 0 | Weekday |
| 2 | 11 | 359 | 0 | 0 | Weekend |
| 3 | 12 | 100 | 0 | 0 | Weekend |
| 4 | 13 | 80 | 28 | 0 | Weekday |
| 5 | 14 | 372 | 28 | 0 | Weekday |
| 6 | 15 | 130 | 29 | 0 | Weekday |
| 7 | 16 | 176 | 28 | 69 | Weekday |
| 8 | 17 | 590 | 31 | 0 | Weekday |
| 9 | 18 | 460 | 16 | 657 | Weekend |
| 10 | 19 | 137 | 15 | 0 | Weekend |
| 11 | 20 | 127 | 27 | 0 | Weekday |
| 12 | 21 | 210 | 27 | 1000 | Weekday |
| 13 | 22 | 157 | 15 | 0 | Weekday |

รูปที่ 9 ข้อมูลจาก Google sheet : 6010502543_Assignment_1 (Data_Weekday-Weekend)

จุดประสงค์ : เพื่อหาค่าเฉลี่ยของค่าอาหารในช่วงวันธรรมดา และวันเสาร์-อาทิตย์ เท่ากันหรือไม่

```
df_weekday = df3[df3.Weekend_or_Weekday=='Weekday']
df_weekday
```

| | Day | Food expenses | Travel expenses | Other expenses | Weekend_or_Weekday |
|----|-----|---------------|-----------------|----------------|--------------------|
| 0 | 9 | 94 | 28 | 121 | Weekday |
| 1 | 10 | 113 | 0 | 0 | Weekday |
| 4 | 13 | 80 | 28 | 0 | Weekday |
| 5 | 14 | 372 | 28 | 0 | Weekday |
| 6 | 15 | 130 | 29 | 0 | Weekday |
| 7 | 16 | 176 | 28 | 69 | Weekday |
| 8 | 17 | 590 | 31 | 0 | Weekday |
| 11 | 20 | 127 | 27 | 0 | Weekday |
| 12 | 21 | 210 | 27 | 1000 | Weekday |
| 13 | 22 | 157 | 15 | 0 | Weekday |

```
df_weekend = df3[df3.Weekend_or_Weekday=='weekend']
df_weekend
```

| | Day | Food expenses | Travel expenses | Other expenses | Weekend_or_Weekday |
|----|-----|---------------|-----------------|----------------|--------------------|
| 2 | 11 | 359 | 0 | 0 | Weekend |
| 3 | 12 | 100 | 0 | 0 | Weekend |
| 9 | 18 | 460 | 16 | 657 | Weekend |
| 10 | 19 | 137 | 15 | 0 | Weekend |

```
df_weekday.describe()
```

| | Day | Food expenses | Travel expenses | Other expenses |
|-------|-----------|---------------|-----------------|----------------|
| count | 10.000000 | 10.000000 | 10.0000 | 10.000000 |
| mean | 15.700000 | 204.900000 | 24.1000 | 119.000000 |
| std | 4.423423 | 158.955934 | 9.5038 | 312.301848 |
| min | 9.000000 | 80.000000 | 0.0000 | 0.000000 |
| 25% | 13.250000 | 116.500000 | 27.0000 | 0.000000 |
| 50% | 15.500000 | 143.500000 | 28.0000 | 0.000000 |
| 75% | 19.250000 | 201.500000 | 28.0000 | 51.750000 |
| max | 22.000000 | 590.000000 | 31.0000 | 1000.000000 |

```
df_weekend.describe()
```

| | Day | Food expenses | Travel expenses | Other expenses |
|-------|-----------|---------------|-----------------|----------------|
| count | 4.000000 | 4.000000 | 4.000000 | 4.00 |
| mean | 15.000000 | 264.000000 | 7.750000 | 164.25 |
| std | 4.082483 | 173.652911 | 8.958236 | 328.50 |
| min | 11.000000 | 100.000000 | 0.000000 | 0.00 |
| 25% | 11.750000 | 127.750000 | 0.000000 | 0.00 |
| 50% | 15.000000 | 248.000000 | 7.500000 | 0.00 |
| 75% | 18.250000 | 384.250000 | 15.250000 | 164.25 |
| max | 19.000000 | 460.000000 | 16.000000 | 657.00 |

Claim : Average food expenses on weekdays is the same as the average food expenses on weekends.

Let

μ_1 = average food expenses on weekdays

μ_2 = average food expenses on weekends.

$\mu_1 - \mu_2$ = difference average food expenses.

$\Delta_0 = 0$

$H_0 : \mu_1 - \mu_2 = 0$

$H_a : \mu_1 - \mu_2 \neq 0$

```
meanx = df_weekday["Food expenses"].mean()
meanx
```

204.9

```
meany = df_weekend["Food expenses"].mean()
meany
```

264.0

```
stdx = df_weekday["Food expenses"].std()
stdx
```

158.95593379578156

```
stdy = df_weekend["Food expenses"].std()
stdy
```

173.6529105236458

```
countx = df_weekday["Food expenses"].count()
countx
```

10

```
county = df_weekend["Food expenses"].count()
county
```

4

```
mean = meanx - meany
mean
```

-59.099999999999994

```
sd = np.sqrt((pow(stdx,2)/countx)+(pow(stdy,2)/county))
sd
```

100.32712605383563

```
test_statistics = (mean-0)/sd
test_statistics
```

-0.5890729887776002

```
x = (pow(stdx,2)/countx)
y = (pow(stdy,2)/county)
```

```
degree_of_freedom = math.floor(pow(x+y,2)/((pow(x,2)/countx-1)+(pow(y,2)/county-1)))
degree_of_freedom
```

6

```
alpha = 0.05/2
t.ppf(1-alpha, degree_of_freedom)
```

2.4469118487916806

- Test statistics = -0.5891
- Given $\alpha = 0.05$
- Degree of freedom = 6
- $t_{0.05/2,6} = 2.4469$
- Rejection region : $t \geq 2.4469$ or $t \leq -2.4469$
- Test statistics fall outside the rejection region.
- Null hypothesis is not rejected.
- **Average food expenses on weekdays is the same as the average food expenses on weekends.**

Anova on single factor

เนื่องจากข้อมูลเดิมไม่สามารถนำมาทำ Anova on single factor ได้ ดังนั้นจึงนำข้อมูลใหม่มาใช้ โดยข้อมูลที่นำมาใช้คือ ข้อมูลการค่าเดินทางโดย มอเตอร์ไซด์ รถเมล์ และ รถไฟฟ้า ของนิสิตทุกคนที่เรียนวิชาสถิติสำหรับการประยุกต์ทางวิศวกรรมคอมพิวเตอร์

| | Motocycle Expense (Baht) | Bus Expense (Baht) | BTS Expense (Baht) |
|----|--------------------------|--------------------|--------------------|
| 0 | 15.0 | 13 | 43.0 |
| 1 | 10.0 | 15 | 43.0 |
| 2 | 10.0 | 13 | 15.0 |
| 3 | 10.0 | 15 | 26.0 |
| 4 | 10.0 | 14 | 10.0 |
| 5 | 40.0 | 14 | 10.0 |
| 6 | 13.0 | 14 | 10.0 |
| 7 | 10.0 | 15 | 10.0 |
| 8 | 20.0 | 13 | 10.0 |
| 9 | 15.0 | 15 | 44.0 |
| 10 | 10.0 | 8 | 53.0 |
| 11 | 20.0 | 8 | 44.0 |
| 12 | 40.0 | 8 | 44.0 |
| 13 | 40.0 | 8 | 44.0 |
| 14 | 20.0 | 14 | 44.0 |
| 15 | 10.0 | 1 | 44.0 |
| 16 | 12.0 | 13 | NaN |

รูปที่ 10 ข้อมูลจาก Google sheet : 6010502543_Assignment_1 (Data_TravelExpenses_shared)

จุดประสงค์ : เพื่อหาค่าเฉลี่ยของค่าใช้จ่ายในการเดินทางของยานพาหนะแต่ละประเภทเท่ากันหรือไม่

- Analyze 3 types of travel expenses.
- Show that travel expenses from 3 travel type are the same at $\alpha = 0.05$
- Let

μ_1 = mean of Motorcycle expense.

μ_2 = mean of Bus expense.

μ_3 = mean of BTS expense.

$I = 3$, $J_1 = 51$, $J_2 = 57$, $J_3 = 16$, $N = 124$

- Hypothesis

$H_0 : \mu_1 = \mu_2 = \mu_3$

H_a : Not all μ_i 's are equal. (at least two of the μ_i 's are different.)

* The calculation method is in Google sheet : [6010502543_Assignment_1_Calculation_method](#)
(Anova_one_factor)

| | df | SS | MS | f |
|-----------|-----|------------|-----------|--------|
| Treatment | 2 | 2714.0353 | 1357.0176 | 9.0679 |
| Error | 121 | 18107.7067 | 149.6505 | |
| Total | 123 | 20821.7419 | | |

- p-value is 0.000214

`1-f.cdf(9.0679,2,121)`

0.00021402381039281604

- $\alpha = 0.05$
- At $\alpha = 0.05 > p\text{-value} = 0.000214$
- Test statistics fall inside the rejection region.
- Null hypothesis is rejected.
- **Travel expenses from 3 travel types are not the same.**

Anova on two factors (additive)

ข้อมูลที่จะนำมาใช้ทำ Anova on two factors คือ ข้อมูลค่าอาหาร โดยมี 2 ตัวแปร คือ

1. ช่วงเวลาในการรับประทานอาหาร (มื้ออาหาร) แบ่งเป็น 2 เวลา ได้แก่
 1. อาหารกลางวัน
 2. อาหารเย็น
2. วันในสัปดาห์ แบ่งได้ 7 วัน

| | Day | Food expenses (lunch) | Food expenses (dinner) |
|---|-----------|-----------------------|------------------------|
| 0 | thursday | 106 | 164 |
| 1 | friday | 292 | 160 |
| 2 | saturday | 295 | 330 |
| 3 | sunday | 117 | 120 |
| 4 | monday | 67 | 140 |
| 5 | tuesday | 414 | 168 |
| 6 | wednesday | 117 | 170 |

รูปที่ 11 ข้อมูลจาก Google sheet : 6010502543_Assignment_1 (Data_FoodExpenses_by_day_of_week)

จุดประสงค์ : เพื่อหาว่าวันในสัปดาห์และมื้ออาหารส่งผลต่อค่าใช้จ่ายในเรื่องอาหารหรือไม่

- Test 2 meals on 7 days of week at significance level = 0.05
- A = Days of week, I = 7
- B = Meals, J = 2
- Hypothesis

$H_{0A} : \alpha_1 = \alpha_2 = \alpha_3$ (Factor A has no effect on Food expenses.)

$H_{aA} : \text{Not all } \alpha_i \text{'s are equal. (Factor A has an effect on Food expenses.)}$

$H_{0B} : \beta_1 = \beta_2 = \beta_3$ (Factor B has no effect on Food expenses.)

$H_{aB} : \text{Not all } \beta_i \text{'s are equal. (Factor B has an effect on Food expenses.)}$

- SST = 133908.00, df = 13
- SSA = 88570.00, df = 6
- SSB = 1738.29, df = 1
- SSE = 43599.71, df = 6

f.ppf(0.95,6,6)

4.283865713822639

f.ppf(0.95,1,6)

5.987377607273699

* The calculation method is in Google sheet : [6010502543_Assignment_1_Calculation_method](#)
(Anova_two_factor)

| | df | SS | MS | f | Rejection region |
|-------|----|-----------|----------|------|------------------|
| A | 6 | 88570.00 | 14761.67 | 2.03 | 4.28 |
| B | 1 | 1738.29 | 1738.29 | 0.24 | 5.99 |
| Error | 6 | 43599.71 | 7266.62 | | |
| Total | 13 | 133908.00 | | | |

- Both two test statistics fall outside the rejection region.
- H_{0A} is not rejected. Factor A (Days of week) has no effect on Food expenses.
- H_{0B} is not rejected. Factor B (Meals) has no effect on Food expenses.
- **Days of week and meals have no effect on food expenses.**

Categorical data analysis (homogeneity)

ข้อมูลที่จะนำมาใช้ทำ Categorical data analysis คือ ข้อมูลค่าอาหาร โดยมี 2

ตัวแปร คือ 1. ประเภทของค่าใช้จ่าย แบ่งเป็น 3 ประเภท ได้แก่

1. ค่าอาหาร
2. ค่าเดินทาง
3. ค่าใช้จ่ายอื่นๆ

2. วันในสัปดาห์ แบ่งได้ 7 วัน

| | Day | Food expenses | Travel expenses | Other expenses |
|---|-----------|---------------|-----------------|----------------|
| 0 | thursday | 270 | 56 | 190 |
| 1 | friday | 703 | 31 | 0 |
| 2 | saturday | 819 | 16 | 657 |
| 3 | sunday | 237 | 15 | 0 |
| 4 | monday | 207 | 55 | 0 |
| 5 | tuesday | 582 | 55 | 1000 |
| 6 | wednesday | 287 | 44 | 0 |

รูปที่ 12 ข้อมูลจาก Google sheet : 6010502543_Assignment_1 (Data_Expenses_by_day_of_week)

จุดประสงค์ : เพื่อหาค่าใช้จ่ายสำหรับแต่ละวันในสัปดาห์มีอัตราส่วนแบ่งเป็นแต่ละประเภท (ค่าอาหาร, ค่าเดินทาง, ค่าใช้จ่ายอื่นๆ) เท่ากันหรือไม่

- In one week has 7 days.
- Test in expenses types at significance level = 0.05
(Food expenses, Travel expenses and other expenses)
- Hypothesis
 H_0 : All days of week are homogeneous in term of expenses types
 (Food expenses, Travel expenses, other expenses)
 I = Days of week = 7
 J = expenses types = 3
 That is we test whether $p_{1j} = p_{2j} = \dots = p_{7j}$ for $j = 1, 2, 3$
 H_a : All days of week are not homogeneous

* The calculation method is in Google sheet : [6010502543_Assignment_1_Calculation_method](#)
 (Categorical_data_analysis)

- Test statistics = 1595.66
- Degree of freedom = 12
- p-value is 0.0
- $\alpha = 0.05$
- At $\alpha = 0.05 > p\text{-value} = 0.0$
- Test statistics fall inside the rejection region.
- Null hypothesis is rejected.
- **All days of week are not homogeneous in terms of expenses types.**

```
1-chi2.cdf(1595.66,12)
```

```
0.0
```

สรุป

จากการวิเคราะห์ข้อมูลค่าใช้จ่ายทั้งหมดของนายชิษณุพงศ์ วรวิจิตรชัยกุล ตั้งแต่วันพฤหัสบดีที่ 9 มกราคม 2563 ถึงวันพุธที่ 22 มกราคม 2563 ทำให้ทราบว่าค่าใช้จ่ายในเรื่องอาหารในแต่ละวันค่อนข้างแตกต่างกัน ส่วนค่าใช้จ่ายในการเดินทางจะแตกต่างกันไปขึ้นอยู่กับว่าวันนั้นมีเรียน และมีเพื่อนกลับพร้อมกันหรือไม่ ในส่วนของค่าใช้จ่ายอื่นๆจะมีแค่บางวันเนื่องจากเป็นค่าใช้จ่ายที่เพิ่มมาจากค่าใช้จ่ายปกติในชีวิตประจำวัน เช่น ค่าตัดผม ค่าสนามฟุตบอล

จากการทดสอบถึงแม้ว่าค่าใช้จ่ายในเรื่องอาหารของแต่ละวันจะแตกต่างกันแต่ค่าเฉลี่ยของค่าอาหารในช่วงวันธรรมดา และวันเสาร์-อาทิตย์นั้นเท่ากัน

ค่าเฉลี่ยของค่าใช้จ่ายในการเดินทางของยานพาหนะแต่ละประเภทยังไม่เท่ากันเนื่องจากค่าใช้จ่ายสำหรับการโดยสารรถไฟฟ้า BTS นั้นสูงกว่าค่าใช้จ่ายในการเดินทางประเภทอื่น

วันในสัปดาห์และมื้ออาหารไม่ส่งผลต่อค่าใช้จ่ายในเรื่องอาหาร ทำให้ทราบว่าไม่ว่าจะทานอาหารในวันไหนและในช่วงเวลาใดค่าใช้จ่ายในการรับประทานอาหารก็ไม่แตกต่างกัน

ค่าใช้จ่ายสำหรับแต่ละวันในสัปดาห์มีอัตราส่วนแบ่งเป็นแต่ละประเภท (ค่าอาหาร, ค่าเดินทาง, ค่าใช้จ่ายอื่นๆ) ไม่เท่ากัน เนื่องจากค่าใช้จ่ายในเรื่องอาหารมากกว่าค่าใช้จ่ายในการเดินทางอยู่พอสมควร และค่าใช้จ่ายอื่นๆจะมีแค่ในบางวัน