



รายงาน

Mini-project

การวิเคราะห์ปัจจัยที่มีผลต่อผลการเรียนของนักศึกษา Performance Index โดยใช้การถดถอยเชิงเส้น

เสนอ

รศ. ฤสา แม็คแนล

จัดทำโดย

6520310001 นางสาว เขมิกา อติอระ

6520310003 นางสาว ซอพิยะ เจะเลาะ

6520310024 นางสาว สามียะห์ ลาเตะ

6520310137 นางสาว อามีหน๊ะ อนันต์เดช

6520310202 MISS LIZA THEA

สาขาคณิตศาสตร์และวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์และเทคโนโลยี

รายงานฉบับนี้เป็นส่วนหนึ่งของ

รายวิชา 747-336 ตัวแบบเชิงสถิติ STATISTICAL MODELING

ภาคเรียนที่ 2 ปีการศึกษา 2567

มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตปัตตานี

## Mini-project

1. หัวข้อ: การวิเคราะห์ปัจจัยที่มีผลการเรียนของนักศึกษา Performance Index โดยใช้การถดถอยเชิงเส้น

### 2. คำถามในการวิจัย

คำถามวิจัย : ปัจจัยใดที่มีผลต่อผลการเรียนของนักศึกษา Performance Index

### 3. วัตถุประสงค์ของการศึกษา

เพื่อสำรวจความสำคัญสัมพันธ์ของตัวแปรอิสระต่างๆ ที่ส่งผลต่อผลการเรียนของนักศึกษา

### 4. ที่มาและความสำคัญ

หัวข้อเรื่อง **Student Performance** ถูกเลือกเนื่องจากผลสัมฤทธิ์ทางการเรียนเป็นส่วนสำคัญของการศึกษา และการทำความเข้าใจปัจจัยที่ส่งผลต่อผลการเรียนสามารถนำไปสู่การพัฒนาผลลัพธ์ทางการเรียนที่ดีขึ้น ชุดข้อมูลนี้เปิดโอกาสให้วิเคราะห์ตัวแปรสำคัญ เช่น ชั่วโมงการเรียน รูปแบบการนอน และการเข้าร่วมกิจกรรมนอกหลักสูตร ซึ่งส่งผลโดยตรงต่อผลการเรียน ผลการศึกษานี้สามารถช่วยให้ครู นักเรียน และผู้กำหนดนโยบายพัฒนากลยุทธ์เพื่อเพิ่มความสำเร็จทางการศึกษาและแก้ไขปัญหาที่เกี่ยวข้องกับการศึกษาได้อย่างมีประสิทธิภาพ

### 5. รายละเอียดของตัวแปรที่เลือกใช้

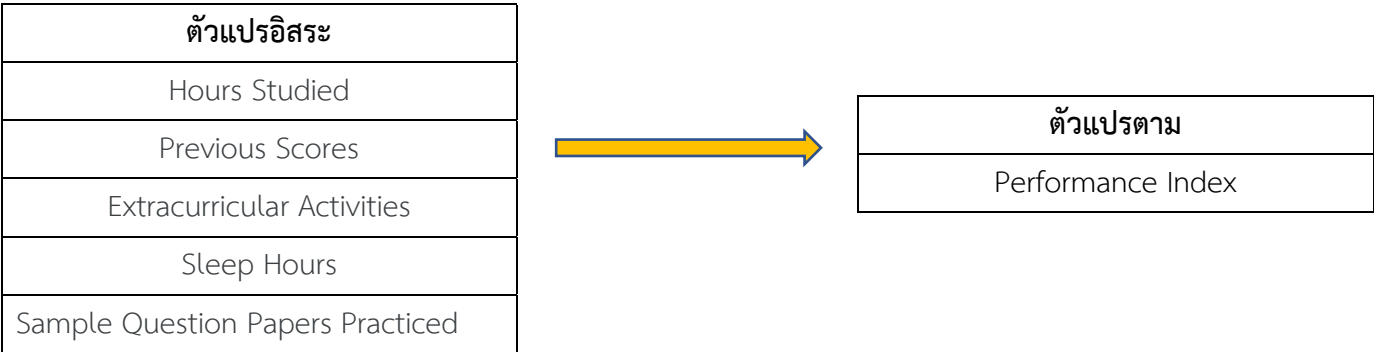
ข้อมูลที่ใช้ในการวิเคราะห์ในรายงานนี้มาจากชุดข้อมูล 'Student Performance (Multiple Linear Regression)' ซึ่งเผยแพร่บนเว็บไซต์ Kaggle (<https://www.kaggle.com/datasets/nikhil7280/student-performance-multiple-linear-regression/data>) ข้อมูลทั้งหมดมี 10,000 รายและ 6 ตัวแปรโดยมีตัวแปรตามตารางดังนี้ :

#### ตารางที่ 1 : คำอธิบายของตัวแปร

ตัวแปร	คำอธิบายของตัวแปร	ประเภทของตัวแปร	รายละเอียดของตัวแปร (คำอธิบายของตัวแปร)
Performance Index (ตัวแปรตาม)	ดัชนีผลการเรียน (คะแนนที่ได้)	ตัวแปรต่อเนื่อง	ดัชนีมีช่วงค่าตั้งแต่ 10 ถึง 100 โดยค่าที่สูงกว่าบ่งชี้ถึงผลการเรียนที่ดียิ่งขึ้น
Hours Studied	จำนวนชั่วโมงที่ใช้ในการเรียน	ตัวแปรต่อเนื่อง	
Previous Scores	คะแนนก่อนหน้า	ตัวแปรต่อเนื่อง	

Extracurricular Activities	การเข้าร่วมกิจกรรมนอกหลักสูตร	ตัวแปรกลุ่ม	“Yes” = 1, “No” = 0
Sleep Hours	จำนวนชั่วโมงการนอน	ตัวแปรต่อเนื่อง	
Sample Question Papers Practiced	จำนวนข้อสอบตัวอย่างที่ฝึกทำ	ตัวแปรต่อเนื่อง	

ตารางที่ 2 : ตัวแปรอิสระ 5 ตัวแปรและตัวแปรตาม 1 ตัวแปร



### 6. การจัดการข้อมูลและทำความสะอาดของข้อมูล

ข้อมูล Student Performance เก็บไว้ในตัวแปร d โดยมีการเลือกใช้ library(epiDisplay), library(dplyr) ในการวิเคราะห์ข้อมูลโดยใช้ R

```
> str(d)
'data.frame':  10000 obs. of  6 variables:
 $ hours_studied   : int  7 4 8 5 7 3 7 8 5 4 ...
 $ prev_scores     : int  99 82 51 52 75 78 73 45 77 89 ...
 $ extra_activities: chr   "Yes" "No" "Yes" "Yes" ...
 $ sleep_hours     : int  9 4 7 5 8 9 5 4 8 4 ...
 $ sample_practiced: int  1 2 2 2 5 6 6 6 2 0 ...
 $ performance     : num  91 65 45 36 66 61 63 42 61 69 ...
```

## 6.1 การตรวจสอบข้อมูลที่ซ้ำกัน

```
> # Check for duplicate rows
> dup <- duplicated(d)
> num_duplicates <- sum(dup)
> num_duplicates
[1] 127
```

จากผลการวิเคราะห์โดยใช้ฟังก์ชัน duplicated() พบว่ามีชุดข้อมูลที่ซ้ำกันจำนวน 127 ชุด ดังนั้นเราจะลบข้อมูลที่ซ้ำกันออกจากชุดข้อมูล

```
> str(d)
'data.frame':  9873 obs. of  6 variables:
 $ hours_studied   : int  7 4 8 5 7 3 7 8 5 4 ...
 $ prev_scores    : int  99 82 51 52 75 78 73 45 77 89 ...
 $ extra_activities: chr  "Yes" "No" "Yes" "Yes" ...
 $ sleep_hours    : int  9 4 7 5 8 9 5 4 8 4 ...
 $ sample_practiced: int  1 2 2 2 5 6 6 6 2 0 ...
 $ performance    : num  91 65 45 36 66 61 63 42 61 69 ...
```

หลังจากลบข้อมูลที่ซ้ำกัน ข้อมูลที่เหลือมี 9873 ชุดและ 6 ตัวแปร

## 6.2 การตรวจสอบข้อมูลสูญหาย

ตรวจสอบข้อมูลสูญหายของแต่ละตัวแปร

```
> d[d == ""] <- NA
> missing_value <- is.na(d)
> missing_count <- colSums(missing_value)
> missing_count
  hours_studied    prev_scores extra_activities    sleep_hours
            0             0             0             0
sample_practiced    performance
            0             0
```

จากผลการวิเคราะห์โดยใช้ฟังก์ชัน is.na() และ sum() พบว่าไม่มีตัวแปรใดที่มีข้อมูลสูญหายในชุดข้อมูล

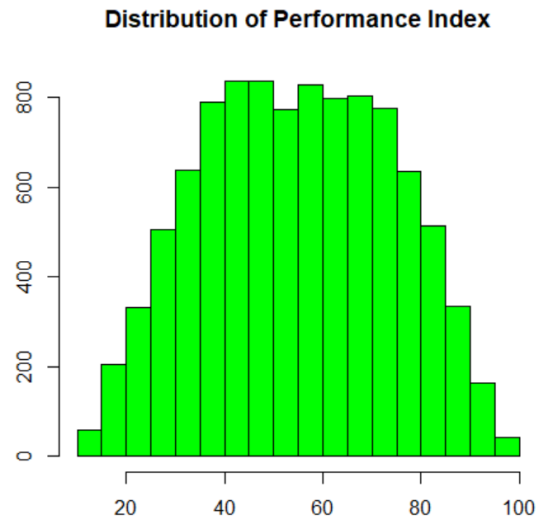
## 7. ผลการวิเคราะห์ข้อมูลเบื้องต้น

### 7.1. การวิเคราะห์ข้อมูล 1 ตัวแปร

สรุปสถิติเบื้องต้นของแต่ละตัวแปร

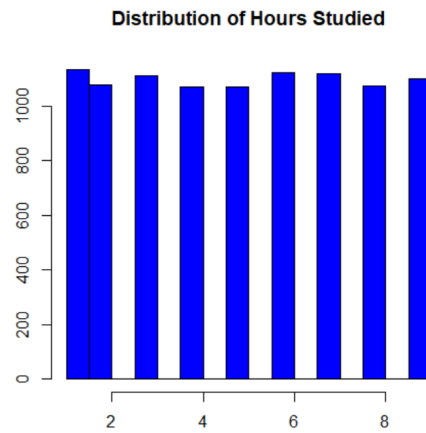
ชื่อตัวแปร	ขนาด	ค่าเฉลี่ย	ค่ามัธยฐาน	ค่าส่วนเบี่ยงเบนมาตรฐาน	ค่าต่ำสุด	ค่าสูงสุด
จำนวนชั่วโมงที่ใช้ในการเรียน	9873	4.99	5	2.59	1	9
คะแนนก่อนหน้า	9873	69.44	69	17.33	40	99
การเข้าร่วมกิจกรรมนอกหลักสูตร	9873	0.49	0	0.5	0	1
จำนวนชั่วโมงการนอน	9873	6.53	7	1.7	4	9
จำนวนข้อสอบตัวอย่างที่ฝึกทำ	9873	4.58	5	2.87	0	9
ดัชนีผลการเรียน (คะแนนที่ได้)	9873	55.22	55	19.21	10	100

# ดัชนีผลการเรียน (ตัวแปรตาม)



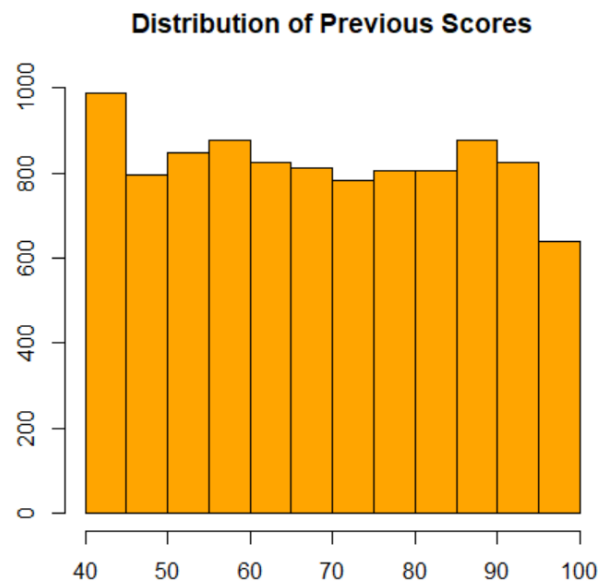
จากกราฟ Histogram พบว่าข้อมูลคะแนนที่ได้จากการเรียนมีการแจกแจงแบบปกติและโดยเฉลี่ยมีค่าเท่ากับ 55.22

# จำนวนชั่วโมงที่ใช้ในการเรียน



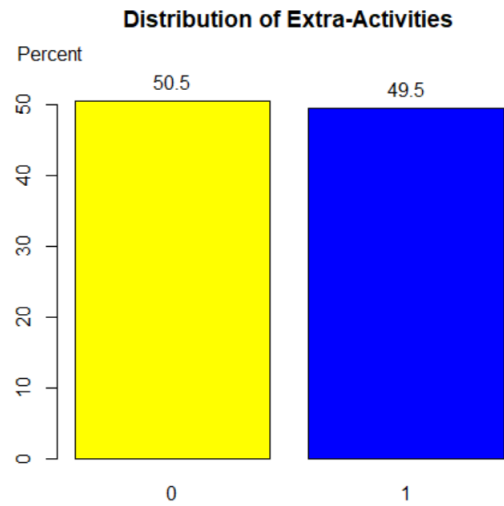
จากกราฟHistogram พบว่าข้อมูลจำนวนชั่วโมงที่ใช้ในการเรียนมีการแจกแจงแบบปกติและโดยเฉลี่ยมีค่าเท่ากับ 4.992

# คะแนนก่อนหน้า



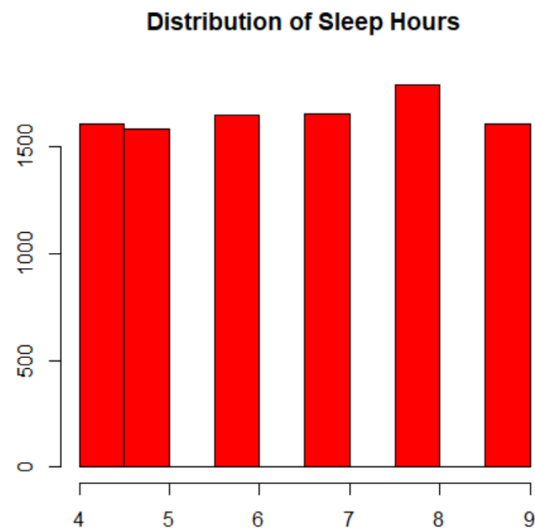
จากกราฟ Histogram พบว่าข้อมูลคะแนนก่อนหน้ามีการแจกแจงแบบปกติและโดยเฉลี่ยมีค่าเท่ากับ 69.44

### # การเข้าร่วมกิจกรรมนอกหลักสูตร



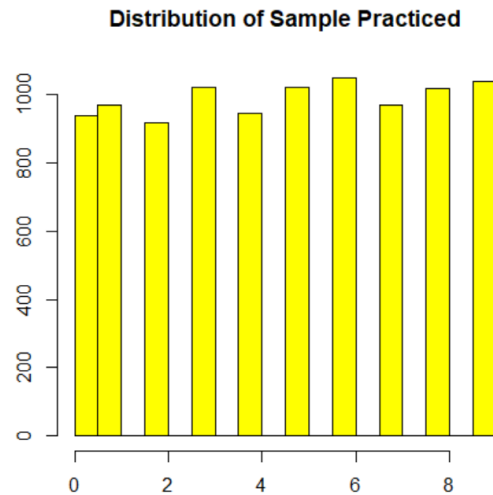
จากกราฟพบว่านักศึกษาส่วนใหญ่ไม่เข้าร่วมกิจกรรมนอกหลักสูตรคิดเป็นร้อยละ 50.5

### # จำนวนชั่วโมงการนอน



จากกราฟ Histogram พบว่าจำนวนชั่วโมงการนอนมีการแจกแจงแบบปกติและโดยเฉลี่ยมีค่าเท่ากับ 6.532

# จำนวนข้อสอบตัวอย่างที่ฝึกทำ



จากกราฟ Histogram พบว่าจำนวนข้อสอบตัวอย่างที่นักศึกษาฝึกทำมีการแจกแจงแบบปกติและโดยเฉลี่ยมีค่าเท่ากับ 4.583

## 8. ขั้นตอนการวิเคราะห์ข้อมูล

### 8.1 การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตาม

ข้อตกลง (Assumption):

- ค่าความคลาดเคลื่อน (error) มีการแจกแจงแบบปกติและมีค่าเฉลี่ยเท่ากับศูนย์
- ค่าความคลาดเคลื่อน (error) มีความแปรปรวนคงที่ (homoscedasticity)
- ค่าความคลาดเคลื่อน (error) เป็นอิสระต่อกัน
- ไม่มี multicollinearity ระหว่างตัวแปรอิสระ

สมมติฐานของการทดสอบ

$$H_0: \beta_1 = 0$$

(ไม่มีอิทธิพลระหว่างจำนวนชั่วโมงที่ใช้ในการเรียนกับดัชนีผลการเรียน)

$$H_1: \beta_1 \neq 0$$

สถิติที่ใช้ทดสอบ  $H_0$  คือสถิติทดสอบ  $t$

$$t = \frac{b_1 - \beta_1}{se(b_1)} = \frac{2.8527 - 0}{0.0079} = 358.94$$



$$\text{มี } v = n - p - 1 = 9867$$

ค่า p-value ประมาณศูนย์ ซึ่งมีค่าน้อยกว่า 0.05 จึงปฏิเสธ  $H_0$  แสดงว่า ตัวแปรจำนวนชั่วโมงที่ใช้ในการเรียน มีอิทธิพลต่อดัชนีผลการเรียนที่ระดับนัยนัยสำคัญ 0.05

$$H \quad H_0: \beta_2 = 0$$

(ไม่มีอิทธิพลระหว่างคะแนนก่อนหน้ากับดัชนีผลการเรียน)

$$H_1: \beta_2 \neq 0$$

สถิติที่ใช้ทดสอบ  $H_0$  คือสถิติทดสอบ  $t$

$$t = \frac{b_2 - \beta_2}{se(b_2)} = \frac{1.0183 - 0}{0.0012} = 857.43$$

$$\text{มี } v = n - p - 1 = 9867$$

ค่า p-value ประมาณศูนย์ ซึ่งมีค่าน้อยกว่า 0.05 จึงปฏิเสธ  $H_0$  แสดงว่า ตัวแปรคะแนนก่อนหน้ามีอิทธิพลต่อดัชนีผลการเรียนที่ระดับนัยนัยสำคัญ 0.05

$$H \quad H_0: \beta_3 = 0$$

(ไม่มีอิทธิพลระหว่างการเข้าร่วมกิจกรรมนอกหลักสูตรกับดัชนีผลการเรียน)

$$H_1: \beta_3 \neq 0$$

สถิติที่ใช้ทดสอบ  $H_0$  คือสถิติทดสอบ  $t$

$$t = \frac{b_3 - \beta_3}{se(b_3)} = \frac{0.6167 - 0}{0.0412} = 14.98$$

$$\text{มี } v = n - p - 1 = 9867$$

ค่า p-value ประมาณศูนย์ ซึ่งมีค่าน้อยกว่า 0.05 จึงปฏิเสธ  $H_0$  แสดงว่า ตัวแปรการเข้าร่วมกิจกรรมนอกหลักสูตรมีอิทธิพลต่อดัชนีผลการเรียนที่ระดับนัยนัยสำคัญ 0.05

$$H \quad H_0: \beta_4 = 0$$

(ไม่มีอิทธิพลระหว่างจำนวนชั่วโมงการนอนกับดัชนีผลการเรียน)

$$H_1: \beta_4 \neq 0$$

สถิติที่ใช้ทดสอบ  $H_0$  คือสถิติทดสอบ  $t$

$$t = \frac{b_4 - \beta_4}{se(b_4)} = \frac{0.4803 - 0}{0.0121} = 39.62$$

$$\text{มี } v = n - p - 1 = 9867$$

ค่า p-value ประมาณศูนย์ ซึ่งมีค่าน้อยกว่า 0.05 จึงปฏิเสธ  $H_0$  แสดงว่า ตัวแปรจำนวนชั่วโมงการนอนมีอิทธิพลต่อดัชนีผลการเรียนที่ระดับนัยนัยสำคัญ 0.05

$$H_0: \beta_5 = 0$$

(ไม่มีอิทธิพลระหว่างจำนวนข้อสอบตัวอย่างที่ฝึกทำกับดัชนีผลการเรียน)

$$H_1: \beta_5 \neq 0$$

สถิติที่ใช้ทดสอบ  $H_0$  คือสถิติทดสอบ  $t$

$$t = \frac{b_5 - \beta_5}{se(b_5)} = \frac{0.1939 - 0}{0.0071} = 27.02$$

$$\text{มี } v = n - p - 1 = 9867$$

ค่า p-value ประมาณศูนย์ ซึ่งมีค่าน้อยกว่า 0.05 จึงปฏิเสธ  $H_0$  แสดงว่า ตัวแปรจำนวนข้อสอบตัวอย่างที่ฝึกทำมีอิทธิพลต่อดัชนีผลการเรียนที่ระดับนัยนัยสำคัญ 0.05

ตารางที่1: ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation)

ชื่อตัวแปร	จำนวนชั่วโมงที่ใช้ในการเรียน	คะแนนก่อนหน้า	การเข้าร่วมกิจกรรมนอกหลักสูตร	จำนวนชั่วโมงการนอน	จำนวนข้อสอบตัวอย่างที่ฝึกทำ	ดัชนีผลการเรียน
จำนวนชั่วโมงที่ใช้ในการเรียน	1.0000	-0.107	0.0049	0.0021	0.0157	0.3753
คะแนนก่อนหน้า	-0.011	1.0000	0.0095	0.0080	0.0087	0.9151
การเข้าร่วมกิจกรรมนอกหลักสูตร	0.0049	0.0095	1.0000	-0.0240	0.0138	0.0261
จำนวนชั่วโมงการนอน	0.0021	0.0080	-0.0240	1.0000	0.0049	0.0504
จำนวนข้อสอบตัวอย่างที่ฝึกทำ	0.0157	0.0087	0.0139	0.0049	1.0000	0.0434
ดัชนีผลการเรียน	0.3753	0.9151	0.0261	0.0504	0.0434	1.0000

