

รหัสนักศึกษา.....ชื่อ-สกุล.....

ปฏิบัติการ 5-6 การวิเคราะห์ข้อมูล 1 ตัวแปร โดยใช้สถิติเบื้องต้น การหาความสัมพันธ์ระหว่าง 2 ตัวแปรโดยใช้สถิติอ้างอิง และการสร้างตัวแบบการถดถอยเชิงเส้นพหุคูณ

ปฏิบัติการนี้มีวัตถุประสงค์เพื่อ ฝึกการวิเคราะห์ข้อมูลสำหรับ 1 ตัวแปร และ การหาความสัมพันธ์ของข้อมูลสำหรับ 2 ตัวแปร โดยใช้ two-sample t-test, anova, และ simple linear regression

สรุปเนื้อหา

กราฟ box plot แสดงให้เห็นข้อมูลเชิงประจักษ์ วิธีการทางสถิติให้ค่า p-value และทำให้สร้างช่วงความเชื่อมั่นได้ ช่วยให้การเปรียบเทียบความแตกต่างระหว่างกลุ่มชัดเจนยิ่งขึ้น การเปรียบเทียบความแตกต่างของค่าเฉลี่ยน้ำหนักรกคลอดระหว่างการสูบบุหรี่ของแม่ ใช้สถิติ two sample t- test หรือ การวิเคราะห์การถดถอยอย่างง่าย (simple linear regression) ส่วนการเปรียบเทียบความแตกต่างของค่าเฉลี่ยของน้ำหนักรกคลอดระหว่างกลุ่มสีผิวของแม่ ใช้สถิติ one-way ANOVA หรือ การวิเคราะห์การถดถอยอย่างง่าย

กรณีตัวแปรตามแบบปริมาณตัวแปรอิสระแบบกลุ่ม 2 ระดับ two sample t- test และ การถดถอยอย่างง่ายให้ค่า p-value เท่ากัน และกรณีตัวแปรตามแบบปริมาณตัวแปรอิสระแบบกลุ่มหลายระดับ (> 2 ระดับ) one-way ANOVA และ การถดถอยอย่างง่ายให้ค่า p-value เท่ากัน แต่การนำเสนอผลการวิเคราะห์อยู่ในรูปแบบที่แตกต่างกัน

การวิเคราะห์การถดถอยอย่างง่ายและการวิเคราะห์การถดถอยพหุคูณ การตรวจสอบตัวแบบ การแปลผล

แบบฝึกปฏิบัติการ

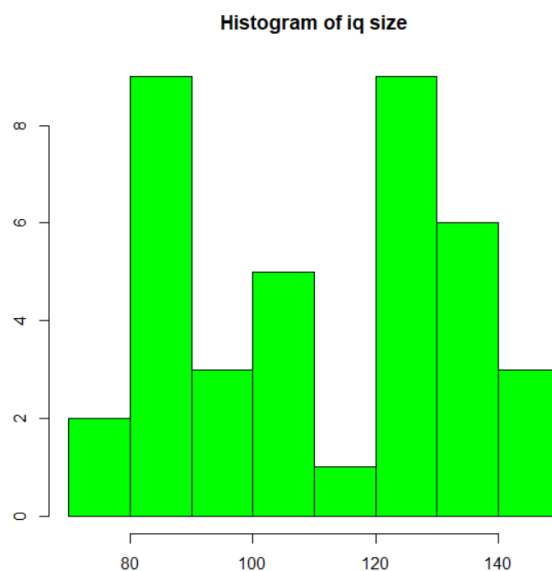
ใช้ข้อมูลตัวอย่าง ของนักวิจัยท่านหนึ่งต้องการศึกษาคุณลักษณะทางกายภาพของมนุษย์มีความสัมพันธ์กับความฉลาดหรือ IQ ของหรือไม่รายละเอียดข้อมูลดังตารางที่ 1

ตารางที่ 1 รายละเอียดของข้อมูล "iqsize.csv"

Variables	Descriptions
Response (y): PIQ	Performance IQ scores (PIQ) Wechsler Adult Intelligence Scale.
Brain	Brain size from MRI scans (given as count/10,000).
Height	Height in inches.
Weight	Weight in pounds. 1="106-140", 2=">140-160", 3=">160-190"
Gender	gender (1=male, 2=female)

จากข้อมูลข้างต้น ให้นักศึกษาวิเคราะห์ข้อมูลและตอบคำถามต่อไปนี้

- ข้อมูล IQ มีค่าสังเกต.....ค่า และตัวแปร.....ตัว
- ตัวแปรตาม คือ.....เป็นประเภท.....
- ตัวแปรอิสระประเภทต่อเนื่อง คือ
ตัวแปรอิสระประเภทกลุ่ม คือ
- การวิเคราะห์ข้อมูลของตัวแปรตาม โดยใช้กราฟฮิสโตแกรม เพื่อแสดงการกระจายของข้อมูล



รูปที่ 1 แสดงการแจกแจงของตัวแปร ระดับ iq

อธิบายกราฟ

.....

.....

.....

5. จงแสดงผลการวิเคราะห์ข้อมูลที่ถูกต้องในตารางที่ 2

ตารางที่ 2 ผลการวิเคราะห์ข้อมูลเบื้องต้น 1 ตัวแปร

Variables	Descriptive statistics	บอกชื่อ graph
ตัวแปรต่อเนื่อง: mean±sd		
piq (ตัวแปรตาม)		
mri		
height		
ตัวแปรกลุ่ม: n(%)		
Weight (105,140] (140,160] (160,192]		
Gender Male Female		

อธิบายผลจากตารางที่ 2

.....

.....

.....

.....

.....

6. จงทดสอบสมมติฐานต่อไปนี้ กรณี univariate analysis

6.1 ทดสอบสมมติฐานระหว่างความฉลาด (piq) และ ขนาดของสมอง (brain)

ผลการวิเคราะห์	
	<pre> > mod0 <- lm(piq~brain, data=iq) > summary(mod0) Call: lm(formula = piq ~ brain, data = iq) Residuals: Min 1Q Median 3Q Max -40.077 -17.508 -2.095 17.097 41.571 Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 4.6519 43.7118 0.106 0.9158 brain 1.1766 0.4806 2.448 0.0194 * --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 21.21 on 36 degrees of freedom Multiple R-squared: 0.1427, Adjusted R-squared: 0.1189 F-statistic: 5.994 on 1 and 36 DF, p-value: 0.01935 </pre>
Hypothesis testing	
H	
A	
T	
P	
C	

6.2 ทดสอบสมมติฐานระหว่าง ความฉลาด (piq) และ ความสูง (height)

ผลการวิเคราะห์	
	<pre> > plot(iq\$piq,iq\$height) > mod1 <- lm(piq~height, data=iq) > summary(mod1) Call: lm(formula = piq ~ height, data = iq) Residuals: Min 1Q Median 3Q Max -42.20 -21.82 5.94 17.33 41.23 Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 147.4067 64.3498 2.291 0.0279 * height -0.5271 0.9389 -0.561 0.5780 --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 22.81 on 36 degrees of freedom Multiple R-squared: 0.008678, Adjusted R-squared: -0.01886 F-statistic: 0.3151 on 1 and 36 DF, p-value: 0.578 </pre>
Hypothesis testing	
H	
A	
T	
P	
C	

6.3 ทดสอบสมมติฐานระหว่าง ความฉลาด (piq) และ น้ำหนัก (weight)

ผลการวิเคราะห์	
<pre>> oneway.test(iq\$piq~iq\$weight1, var.equal=T)</pre> <p>One-way analysis of means</p> <p>data: iq\$piq and iq\$weight1</p> <p>F = 0.72188, num df = 2, denom df = 35, p-value = 0.4929</p>	
Hypothesis testing	
H	
A	
T	
P	
C	

7. ผลการวิเคราะห์ข้อมูลรายคู่

ตารางที่ 3 ผลการวิเคราะห์ข้อมูลรายคู่ 2 ตัวแปร

ตัวแปรอิสระ	ตัวแปรตาม ค่า mean±sd และค่า r	ค่าสถิติ t-test/F-test	p-value
mri			
height			
Weight (105,140] (140,160] (160,192]			

ตัวแปรอิสระ	ตัวแปรตาม ค่า mean±sd และค่า r	ค่าสถิติ t-test/F-test	p-value
Gender			
Male			
Female			

อธิบายผลจากตารางที่ 3

.....

.....

.....

.....

.....

8. ผลการวิเคราะห์จากตัวแบบ linear regression

8.1 ผลการวิเคราะห์ full model

ตารางที่ 4 ผลการวิเคราะห์ full models (multiple linear regression)

Determinants	Estimate	Std. Error	t value	Pr(> t)
(Intercept)		76.99	2.04	
brain		0.58	3.41	
height		1.27	-2.72	
weight1(140,160]				
weight1(160,192]				
gender1female				

จงระบุตัวแปรที่มีความสัมพันธ์กับตัวแปรตาม

.....

.....

จงระบุตัวแปรที่ไม่มีความสัมพันธ์กับตัวแปรตาม

.....

.....

8.2 ผลการคัดเลือกตัวแบบโดยวิธีของ “Backward”

ตารางที่ 5 ผลการคัดเลือกตัวแบบ โดยใช้วิธี back ward selection

```

Start:  AIC=232.21
piq ~ brain + height + weight1 + gender1

      Df Sum of Sq  RSS    AIC
- weight1  2      447.0 12937 229.55
- gender1  1      666.6 13157 232.19
<none>                 12490 232.21
- height   1     2877.1 15367 238.09
- brain    1     4547.4 17038 242.01

Step:  AIC=229.55
piq ~ brain + height + gender1

      Df Sum of Sq  RSS    AIC
- gender1  1      384.6 13322 228.66
<none>                 12937 229.55
- height   1     3189.3 16127 235.92
- brain    1     5697.1 18634 241.42

Step:  AIC=228.66
piq ~ brain + height

      Df Sum of Sq  RSS    AIC
<none>                 13322 228.66
- height  1     2875.6 16198 234.09
- brain   1     5408.8 18731 239.61

Call:
lm(formula = piq ~ brain + height, data = iq)

Coefficients:
(Intercept)      brain      height
    111.276      2.061     -2.730

```

จากตารางที่ 4 ค่า AIC ต่ำสุด คือ

จงระบุตัวแปรที่ถูกคัดออกจากตัวแบบมีกี่ตัวอะไรบ้าง พร้อมทั้งแสดงเหตุผล.....

.....

.....

.....

8.3 ผลการวิเคราะห์ข้อมูลจากตัวแบบที่เหมาะสมที่สุด

ตารางที่ 6 ผลการวิเคราะห์ตัวแบบที่ดีที่สุด

Determinants	Estimate	Std. Error	t-value	p-value
(Intercept)				
brain				
height				

จากตารางที่ 6 จงแสดงสมการเชิงเส้นของ estimated linear regression equation

.....

.....

อธิบายผลจากตารางที่ 6

.....

.....

.....

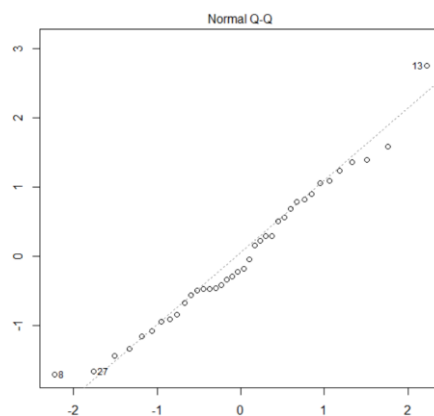
ค่า R-squared ที่ได้จากตัวแบบพร้อมทั้งอธิบายผล

.....

.....

.....

8.4 ผลการตรวจสอบตัวแบบ



รูปที่ 2 กราฟความคลาดเคลื่อนกับ normal score

คำอธิบายรูปที่ 2

.....

.....

```
##### command in R
setwd("D:\\2566-1\\747-341\\data & command")
iq <- read.csv("iqsize.csv", header=T, sep=",")
library(epiDisplay)
str(iq) # structure data
names(iq) <- tolower(names(iq))
des(iq)
summ(iq) # for continuous variables
# data table
#View(iq)
# 1. การแจกแจงของตัวแปรตาม outcome distribution
summ(iq$piq)
hist(iq$piq, col="green", main="Histogram of iq size")

# 2. การแจกแจงของตัวแปรอิสระ determinants distribution and management
#continuous variables based on descriptive results
summ(iq$brain)
hist(iq$brain, col="pink")
shapiro.test(iq$brain)

summ(iq$height)
hist(iq$height, col="pink")
shapiro.test(iq$height)

summ(iq$weight)
hist(iq$weight, col="pink")
shapiro.test(iq$weight)
quantile(iq$weight) # การจัดกลุ่มใหม่ของตัวแปร น้ำหนัก
iq$weight1 <- cut(iq$weight, c(105.00,140.00,160.00,192.00))
tab1(iq$weight1, col=c("green","pink","yellow"))
```

2.1 การแปลงตัวแปร อีสระที่เป็น integer ให้เป็น factor Categorical variable based on descriptive

```
tab1(iq$gender)
iq$gender1 <- factor(iq$gender)
levels(iq$gender1) <- c("male","female")
tab1(iq$gender1, col=c("blue","pink"))
```

3. การหาความสัมพันธ์รายคู่ ระหว่างตัวแปรอิสระแต่ละตัวกับตัวแปรตาม preliminary analysis

relationship between outcome and determinants by t.test and simple linear regression and graph

compare mean, 2 groups

```
summ(iq$piq,by=iq$gender1)
t.test(iq$piq~iq$gender1, var.equal=T)
boxplot(iq$piq~iq$gender1, col=c("yellow","red"),main="iq size between gender")
```

```
summ(iq$piq,by=iq$weight1)
oneway.test(iq$piq~iq$weight1, var.equal=T)
boxplot(iq$piq~iq$weight1, col=c("yellow","red"),main="iq size between weight group")
```

simple linear regression, full model

```
cor(iq$piq,iq$brain)
plot(iq$piq,iq$brain)
mod0 <- lm(piq~brain, data=iq)
summary(mod0)
```

```
cor(iq$piq,iq$height)
plot(iq$piq,iq$height)
```

```
mod1 <- lm(piq~height, data=iq)
summary(mod1)
```

4. การสร้างตัวแบบการถดถอยเชิงเส้นพหุคูณ multiple linear regression, full model

```
mod <- lm(piq~brain+height+weight1+gender1, data=iq)
summary(mod)
```

5. การคัดเลือกตัวแบบที่เหมาะสมที่สุด

```
step(mod,direction = "backward")
```

6. ตัวแบบที่เหมาะสมที่สุด best model

```
mod4 <- lm(piq~brain+height, data=iq)
summary(mod4)
```

```
step(mod4,direction = "backward")
```

```
anova(mod4) # assess whether at least one predictor variable has a non-zero
coefficient
```

```
#Assess model
```

```
plot(mod4)
```