

ใบงาน รายวิชา 747-341 Data analytics and data visualization module

week 5-6: inferential statistics

1. ข้อมูลและรายละเอียดข้อมูล

ข้อมูลน้ำหนักของทารกแรกคลอด ที่เก็บรวบรวม จากศูนย์การแพทย์ Baystate (Baystate Medical Center) ในเมือง Springfield รัฐแมสซาชูเซตส์ จากหญิงตั้งครรภ์จำนวน 189 ราย ข้อมูลนี้ถูกเก็บในชื่อ “Bweight.csv” รายละเอียดข้อมูลแสดงดังตารางที่ 1 การวิเคราะห์ข้อมูล 1 ตัวแปร โดยใช้สถิติเบื้องต้น

ตารางที่ 1 รายละเอียดข้อมูล

ตัวแปร	คำอธิบายตัวแปร	คำอธิบายค่าตัวเลข
low	น้ำหนักแรกคลอดต่ำกว่า 2,500 กรัม	0 คือ $\geq 2,500$, 1 คือ $< 2,500$
age	อายุของมารดาเป็นปี	
lwt	น้ำหนักของมารดาที่มีประจำเดือนครั้งสุดท้าย (ปอนด์)	
race	เชื้อชาติของมารดา	1=ผิวขาว, 2=ผิวดำ, 3=อื่นๆ
smoke	การสูบบุหรี่ขณะตั้งครรภ์	0= ไม่สูบ, 1= สูบ
ptl	จำนวนของการคลอดก่อนกำหนด	
ht	ประวัติการเป็นโรคความดันโลหิตสูง	0=ไม่มี, 1=มี
ui	การกลั่นปัสสาวะไม่อยู่	0=ไม่มี, 1=มี
ftv	จำนวนครั้งในการพบแพทย์ของไตรมาสแรก	
bwt	น้ำหนักแรกคลอดเป็นกรัม (ตัวแปรตาม)	

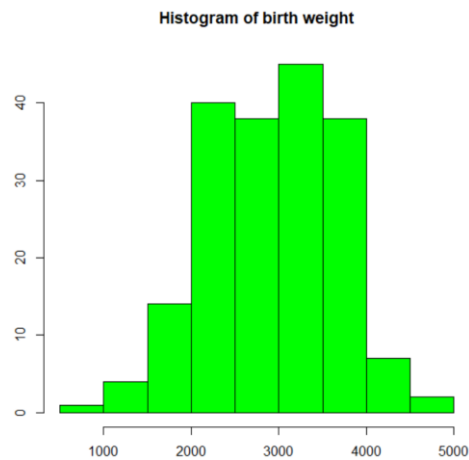
จากข้อมูลข้างต้นให้ นศ. ตอบคำถามต่อไปนี้

- จากข้อมูล “Bweight.csv” ตัวแปรทั้งหมดมีกี่ตัว.....
- จงระบุชื่อตัวแปรต่อเนื่อง คือ

ลำดับที่รหัส นศ.ชื่อ-สกุล.....

3. จงระบุชื่อตัวแปรกลุ่ม คือ

4. กราฟฮิสโตแกรมแสดงการแจกแจงของน้ำหนักของทารกแรกเกิด



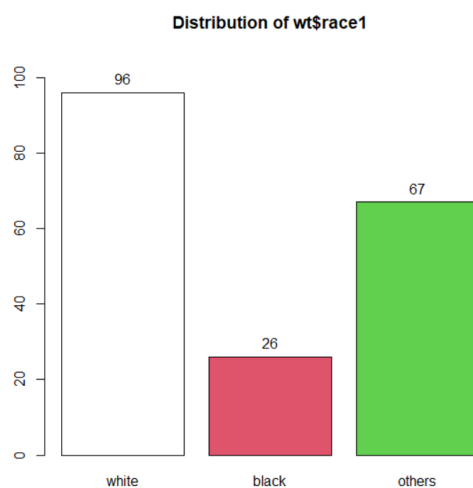
```
> summ(wt$bwt)
obs. mean    median  s.d.   min.   max.
189  2944.587  2977    729.214  709   4990
```

จงอธิบายกราฟพร้อมทั้งแสดงค่าสถิติเบื้องต้น

.....

.....

5. กราฟแท่งแสดงการแจกแจงของตัวแปรสีผิวของแม่



จงอธิบายกราฟโดยใช้สถิติเบื้องต้นร้อยละ

.....

.....

การวิเคราะห์ข้อมูลรายคู่ 2 ตัวแปร

6. ผลการวิเคราะห์ข้อมูลแสดงความสัมพันธ์ระหว่างอัตราการเต้นของชีพจรของ ก่อนวิ่งและหลังวิ่งแตกต่างกันหรือไม่

```
> t.test(pul$PULSE2_NR, pul$PULSE1_NR, paired=TRUE, var.equal=TRUE)

Paired t-test

data:  pul$PULSE2_NR and pul$PULSE1_NR
t = -0.70662, df = 30, p-value = 0.4853
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -1.6313753  0.7926656
sample estimates:
mean difference
 -0.4193548
```

จากผลการวิเคราะห์ข้อมูลข้างต้น จงแสดงลำดับการทดสอบสมมติฐาน

.....

.....

.....

.....

.....

.....

.....

7. ผลการวิเคราะห์ข้อมูลแสดงความสัมพันธ์ระหว่างการสูบบุหรี่ระหว่างตั้งครรภ์กับน้ำหนักของทารกแรกคลอด

```
> t.test(wt$bwt~wt$smoke1, var.equal=T)

Two Sample t-test

data:  wt$bwt by wt$smoke1
t = 2.6529, df = 187, p-value = 0.008667
alternative hypothesis: true difference
95 percent confidence interval:
  72.75612 494.79735
sample estimates:
mean in group no mean in group yes
 3055.696          2771.919
```

จากผลการวิเคราะห์ข้อมูลข้างต้น จงแสดงลำดับการทดสอบสมมติฐาน

.....

.....

.....

.....

.....

.....

.....

.....

.....

8. ผลการวิเคราะห์ข้อมูลแสดงความสัมพันธ์ระหว่างชนิดผิวของมารดากับน้ำหนักของทารกแรกคลอด

```
> oneway.test(wt$bwt~wt$race1, var.equal=T)

One-way analysis of means

data: wt$bwt and wt$race1
F = 4.9125, num df = 2, denom df = 186, p-value = 0.008336
```

จากผลการวิเคราะห์ข้อมูลข้างต้น จงแสดงลำดับการทดสอบสมมติฐาน

.....

.....

.....

.....

.....

.....

.....

.....

.....

ตัวแบบการถอดถอยเชิงเส้น

9. ผลการวิเคราะห์ข้อมูลแสดงความสัมพันธ์ระหว่างจำนวนการคลอดก่อนกำหนดของมารดากับน้ำหนักของทารกแรกคลอด

```
> # simple linear regression
> m1 <- lm(bwt~ptl1, data=wt)
> summary(m1)

Call:
lm(formula = bwt ~ ptl1, data = wt)

Residuals:
    Min       1Q   Median       3Q      Max
-1992.49  -495.40    -8.49   586.51  1976.51

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3013.49      56.59  53.252  < 2e-16 ***
ptl11-3      -434.09     142.04   -3.056  0.00257 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 713.6 on 187 degrees of freedom
Multiple R-squared:  0.04757,    Adjusted R-squared:  0.04248
F-statistic:  9.34 on 1 and 187 DF,  p-value: 0.00257
```

จากผลการวิเคราะห์ข้อมูลข้างต้น จึงแสดงลำดับการทดสอบสมมติฐาน

[illegible]

