

คณะสถิติประยุกต์
สถาบันบัณฑิตพัฒนบริหารศาสตร์

Final Examination 2020/03
LM7204 Business Forecasting

2021-07-31
09:00-12:00

คำชี้แจงและคำสั่ง

1. การทุจริตถือเป็นความผิดทางวินัยอย่างร้ายแรง
 - (a) ห้ามปรึกษาหรือติดต่อสื่อสารกับบุคคลอื่น
 - (b) ห้าม search internet
2. ข้อสอบเป็น open book: อนุญาตให้เปิดเอกสาร และ/หรือตำราได้ สามารถใช้คอมพิวเตอร์หรือเครื่องคิดเลขได้
3. การทำข้อสอบ
 - a. สามารถเขียนใส่กระดาษแล้วถ่ายรูป หรือเขียนใน computer/tablet ก็ได้
 - b. หนึ่งไฟล์ต่อหนึ่งข้อเท่านั้น ส่งในรูปแบบ pdf (หากข้อนั้นมีหลายหน้าให้ save เป็น pdf ไฟล์เดียว)
 - c. ให้แสดงวิธีทำโดยละเอียด. คำตอบโดยไม่มีที่มาหรือวิธีทำอาจไม่ได้รับคะแนนเต็มสำหรับส่วนนั้น หากคำตอบ มาจากการคำนวณโดยใช้โปรแกรม ระบุ function ที่ใช้ด้วย.
 - d. ให้วงกลมหรือใส่กรอบที่คำตอบสุดท้ายด้วย เพื่อให้ทราบว่าคำตอบอยู่ตรงไหนใน output ของ program
4. การส่งข้อสอบ
 - a. ส่งใน assignment แต่ละข้อตามช่วงเวลาส่งในตาราง. (ใน MS Teams หลังจาก Turned in, วันที่และเวลาส่งจะถูกระบุไว้) หากในข้อใด, ส่ง late หรือส่งไม่ตรงตาม assignment, จะไม่ตรวจข้อนั้น และถือว่าคะแนนข้อนั้นเป็น 0 (ศูนย์)
 - b. ตั้งชื่อตาม Filename สำหรับแต่ละข้อดังในตาราง

Problem	ช่วงเวลาส่ง	Filename
FN Problem1	09:00-09:35	Problem1-SID6XXXXXXXXXX
FN Problem2	09:30-10:05	Problem2-SID6XXXXXXXXXX
FN Problem3	10:00-10:35	Problem3-SID6XXXXXXXXXX
FN Problem4	10:30-11:05	Problem4-SID6XXXXXXXXXX
FN Problem5	11:00-11:35	Problem5-SID6XXXXXXXXXX
FN Problem6	11:30-12:05	Problem6-SID6XXXXXXXXXX

5. ข้อสอบมีทั้งหมด 7 แผ่น ไม่รวมคำชี้แจง
6. ข้อสอบคะแนนเต็ม 100 คะแนน

ใช้ data set ต่อไปนี้สำหรับปัญหาข้อ 1-5 (ทุกข้อยกเว้นข้อสุดท้าย) data file: **LM7204Final2020.xlsx**

Sheet: dataP1

FuelConsumption	CarAge	FemaleDriver	CarModel	Distance
22.3	3.5	0	A	129407
22.0	4	1	P	147563
23.7	2.3	1	P	84260
24.2	2.1	0	A	77621
25.5	1.2	1	A	44673
21.1	5.5	0	A	203401
20.6	4.5	0	A	165231
24.0	1.5	0	A	55745
26.0	1.5	1	A	55017
23.1	2.7	0	A	100102
24.8	2.5	1	A	93612
20.2	5	0	P	183615

บริษัทผู้รถ taxi สนใจศึกษาความสัมพันธ์ระหว่างอัตราการใช้น้ำมัน (Fuel Consumption หน่วยเป็น kilometer/liter) และตัวแปรต่างๆ เช่น อายุรถ (Car Age หน่วยเป็นปี), เพศคนขับ (Female Driver = 1 ผู้หญิง, 0 ผู้ชาย), รุ่นรถ (Car Model มีสองรุ่นคือ A และ P) และระยะทางที่วิ่งได้ (Distance หน่วยเป็น kilometer)

รถทั้งหมดของบริษัทผู้รถนี้มาจาก manufacturer เดียวกัน. ทั้งรุ่น A และ P มีขนาดความจุเครื่องยนต์ (cc) เท่ากัน.

Problem 1 (10 คะแนน)

จงสร้าง scatter plot ให้ y = Fuel Consumption และ x = Car Age. ทำด้วย Excel หรือ R หรือจะ plot ด้วยมือเองบนกระดาษกราฟก็ได้. (ข้อนี้เป็น “warm up” เพื่อให้คุ้นเคยกับวิธีการสอบ เช่น ส่งแต่ละข้อใน pdf file เดียว และกด turn in ภายในเวลาที่กำหนด. หากทำข้อนี้เสร็จแล้ว, ทำข้อต่อไปได้เลย.)

Problem 2 (25 คะแนน)

ให้ใช้ independent variable เดียวคือ Car Age.

2.1 จงเขียน regression equation ที่ได้

2.2 จงทำ diagnostic checks ตรวจสอบ regression model ที่ได้ (ให้ครบทุก assumptions)

2.3 ถ้าอายุของรถเป็น 3.2 ปี จงพยากรณ์อัตราการใช้น้ำมัน พร้อมทั้ง 90% Prediction Interval (PI)

Point Forecast	Lower Bound of 90% PI	Upper Bound of 90% PI

Problem 3 (20 คะแนน)

3.1 จงหา correlation ระหว่าง car age และ distance พร้อมทั้งแปลผล

3.2 พิจารณา สอง models ต่อไปนี้

- Model A ที่มี independent variable เป็น car age
- Model B ที่มี independent variables เป็น car age และ distance

ทำ leave-out cross-validation ใส่ค่า BIC, Adjusted R-squared, R-squared ที่ได้ในตาราง

	BIC	Adjusted R-squared	R-squared
Model A			
Model B			

3.3 ควรเลือก model A หรือ B? ให้เหตุผลในการเลือกว่าพิจารณาอะไรบ้าง? (สามารถตอบอย่างอื่นนอกเหนือจากตารางได้, ถ้าจำเป็น)

Problem 4 (20 คะแนน)

4.1 จงทำ regression ที่มี independent variables ดังต่อไปนี้

(i) car age; (ii) dummy variable สำหรับเพศของคนขับรถ; (iii) dummy variable สำหรับรุ่นของรถ

และเขียน regression equation ที่ได้ ให้ใช้ dummy สำหรับ (ii) และ (iii) คนละตัวแปรกัน

4.2 Forecast อัตราการใช้น้ำมันในกรณีต่างๆต่อไปนี้ สำหรับอายุของรถ 3.2 ปี

อายุรถ	เพศคนขับ	รุ่นรถ	Forecast equation	ค่าพยากรณ์ อัตราการใช้น้ำมัน
3.2	Female	A		
3.2	Female	P		
3.2	Male	A		
3.2	Male	P		

Problem 5 (10 คะแนน)

ให้ y_t แทน observation ที่เวลา t เมื่อ $t = 1, 2, 3, \dots$

จงเขียนรูป NNAR(5,3,1)₁₂

Problem 6 (15 คะแนน)

Sheet: dataP6 (data file: LM7204Final2020.xlsx)

HouseNo	Income	LotSize	MowerOwner
1	60	18.4	Owner
2	85.5	16.8	Owner
3	64.8	21.6	Owner
4	61.5	20.8	Owner
... ในข้อสอบ แสดงไม่ครบทั้งหมด. ให้ใช้ข้อมูลที่ครบถ้วนจาก Sheet dataP6 ...			
23	51	14	Non-Owner
24	63	14.8	Non-Owner

บริษัทขายเครื่องตัดหญ้า (mower) สนใจพยากรณ์ความต้องการเครื่องตัดหญ้า. ตารางข้างต้นเป็นข้อมูลของ 24 ครั้วเรือน รายได้ (Income หน่วยเป็น 1000 USD/year), ขนาดสนามหญ้า (หน่วยเป็น 1000 square feet), และครั้วเรือนนี้เป็นเจ้าของเครื่องตัดหญ้า (Owner) หรือไม่เครื่องตัดหญ้า (Non-Owner). สมมติว่าแต่ละครั้วเรือนที่มีเครื่องตัดหญ้ามียี่ห้อเดียวกัน.

6.1 จงทำ logistic regression เพื่อพยากรณ์โอกาสเป็นเจ้าของเครื่องตัดหญ้า (mower owner) โดยใช้ independent variables ทั้งหมด. เขียนสมการ logistic regression ที่ได้ พร้อมทั้งแปลผล coefficients

6.2 สมมติว่าหมู่บ้านจัดสรรโครงการหนึ่ง มีบ้านขนาดใหญ่จำนวน 10 หลัง มีพื้นที่หลังละ 22.2 (หน่วย 1000 square feet) และโดยเฉลี่ยเจ้าของบ้านมีรายได้ 68.7 (1000USD/ปี). มีบ้านขนาดเล็กจำนวน 25 หลังมีพื้นที่หลังละ 18.0 (หน่วย 1000 square feet) และโดยเฉลี่ยเจ้าของบ้านขนาดใหญ่มีรายได้ 49.5 (1000 USD/ปี).

6.2.1 จงหาความน่าจะเป็นที่บ้านขนาดใหญ่จะมีเครื่องตัดหญ้า

6.2.2 จงหาความน่าจะเป็นที่บ้านขนาดเล็กจะมีเครื่องตัดหญ้า

Extra Credit (10 คะแนน) ให้ใช้โจทย์ต่อจากข้อ 6

ให้ X เป็น random variable แทนจำนวนเครื่องตัดหญ้ารวมของบ้านขนาดใหญ่

ให้ Y เป็น random variable แทนจำนวนเครื่องตัดหญ้ารวมของบ้านขนาดเล็ก

สมมติให้เหตุการณ์ที่บ้านแต่ละหลังมีเครื่องตัดหญ้าเป็นอิสระ (independent) กัน

6.3.1 จงระบุการแจกแจง (distribution) ของ X และ Y (เช่น normal, Poisson, binomial, exponential เป็นต้น) พร้อมทั้ง parameters ของ distribution นั้น

6.3.2 ให้จำนวนเครื่องตัดหญ้ารวมทั้งหลังเล็กและใหญ่แทนด้วย $Z = X + Y$

จงคำนวณค่าเฉลี่ย (expectation) ของจำนวนเครื่องตัดหญ้ารวมทั้งหลังเล็กและใหญ่ พร้อมทั้ง variance และ standard deviation.

Distribution	Parameters	Mean	Variance	Standard deviation
N/A	N/A			