ข้อสอบครั้งที่ 3 (ข้อสอบ Take-home)

วิชา: 2756886 เรื่องคัดเฉพาะทางการวิจัยทางการศึกษา (การใช้โปรแกรม R)

อาจารย์ผู้สอน: อ. ดร. สันทัด พรประเสริฐมานิต

วัน-เวลาทดสอบ: 2-5 พฤษภาคม พ.ศ. 2559 (60 คะแนน)

# คำสั่ง

1. จงเปิดไฟล์ R ขึ้นมาไฟล์หนึ่ง ตั้งชื่อโดยนำด้วยรหัสประจำตัวนิสิต แล้วตามด้วยคำว่า Test3 เช่น 5947426527Test3.R ไม่ต้องใส่ชื่อนามสกุลของคุณลงในไฟล์คำตอบ
2. จงตอบคำถามทุกข้อ โดยก่อนตอบแต่ละคำข้อคำถามให้ใส่เลขข้อลงใน comment เช่น #### Item 1 หากคำตอบไม่ตรงตำแหน่งของข้อคำถาม ผมขอไม่คิดคะแนนให้
3. เมื่อตอบเสร็จแล้ว ให้ส่งคำตอบของคุณเข้ามาที่ psunthud@gmail.com โดยให้แนบไฟล์คำตอบ พร้อมทั้งใส่ชื่อ นามสกุล และเลขประจำตัวนิสิตภายในเนื้อหาของอีเมลล์ และตรวจสอบกับอาจารย์ว่าได้รับไฟล์แล้วก่อนออกจากห้อง
4. ให้ทำ comment ในทุกๆ 4-5 บรรทัด อธิบายว่าคุณกำลังทำอะไรอยู่ หากผมไม่เข้าใจและต้องใช้ความพยายามอย่างที่ไม่สมควรจะเป็น ในการอ่าน code ของคุณ ผมขอหักคะแนนส่วนดังกล่าว
5. การสอบครั้งนี้เป็นการสอบจากการทำที่บ้าน คือ คุณทำของคุณคนเดียวจากที่ใดก็ได้ โดยไม่ต้องมีเพื่อนมานั่งทำด้วยกัน ผมอนุญาตให้คุณใช้ Internet และเปิดหนังสือและเอกสารทุกอย่าง
6. การสอบครั้งนี้ ไม่อนุญาตให้คุณสื่อสารกับผู้อื่น (รวมถึงเพื่อนของคุณ) เกี่ยวกับการสอบนี้ ไม่ว่าวิธีทางใดทั้งสิ้น ทั้งการคุย ส่งไฟล์ แชท โพสเวปบอร์ด ถาม pantip และพวกคุณไม่สามารถแชร์เอกสารเกี่ยวกับการสอบได้ครับ หากผมพบว่าพวกคุณมีลักษณะการเขียน code คล้ายกัน ผมจะปรับ F ทั้งสองคนครับ ดังนั้น คุณต้องระวังไม่ให้เพื่อนเห็น code ของคุณอย่างเด็ดขาดครับ
7. ผมมีสิทธิ์ในการถามซ้ำกับคุณอีกครั้งหนึ่ง เพื่ออธิบาย code ของคุณ หากคุณตอบไม่ได้หรือตอบได้ไม่ชัดเจน ผมถือว่าเป็นตัวบ่งชี้ว่าคุณไม่ได้ทำเอง ผมขออนุญาตสอบคุณใหม่อีกครั้งหนึ่ง และหากคะแนนต่ำลงอย่างเห็นได้ชัด (เช่นต่ำลงมากกว่า 30%) ผมถือว่าคุณอาจจะทุจริตในการสอบ ผมจะใช้คะแนนล่าสุด ผมจะส่งเรื่องให้คณะพิจารณาต่อไป
8. การละเมิดคำสั่งในข้อ 5-7 ถือเป็นการทุจริตการสอบ ต้องได้รับโทษตามระเบียบของจุฬาลงกรณ์มหาวิทยาลัย ซึ่งการลงโทษต่ำสุด คือ ได้ F ในวิชาที่สอบและพิจารณาพักการศึกษาไม่ต่ำกว่า 1 ภาคการศึกษา
9. คุณต้องตรวจสอบคำตอบของคุณก่อนส่งทุกครั้ง ว่าสามารถทำคำสั่งได้ โดยไม่มี error ใดๆ เกิดขึ้น
10. ผมขอไม่รับผิดชอบกับอุบัติเหตุทุกชนิดที่อาจเกิดขึ้นระหว่างสอบ เช่น เครื่องคอมพิวเตอร์เสีย ไฟล์ค้าง และอื่นๆ หากคุณไม่สามารถส่งไฟล์คำตอบได้ ถือว่าคุณสละสิทธิ์การสอบครั้งนี้ ดังนั้นผมขอแนะนำให้คุณบันทึกข้อมูลระหว่างทำบ่อยๆ และ backup ไฟล์ตลอด
11. เมื่อคุณอ่านเอกสารนี้แล้ว คุณไม่โต้แย้งผมภายใน 23:59 วันจันทร์ ไม่ว่าทางใดทางหนึ่ง ผมถือว่าคุณยอมรับกติกาในข้อที่ผ่านมาโดยอัตโนมัติ

# คำถาม

1. ท่านต้องการทดสอบว่า trimmed mean กับ mean แบบปกติ วิธีการใดที่ประมาณค่าพารามิเตอร์ได้ดีกว่ากัน เมื่อประชากรมีการกระจายแบบเบ้ จงดูตัวอย่าง code ด้านล่าง แล้วใช้ for-loop เพื่อทดสอบว่า trimmed mean หรือ mean แบบปกติ ดีกว่ากันในการประมาณค่าเฉลี่ย (เปรียบเทียบผ่านอคติในการประมาณค่าเฉลี่ย) (5 คะแนน)

**x <- rchisq(20, df = 3) # Population mean = 3**

**m <- mean(x)**

**tm <- psych::winsor.mean(x, trim = 0.2)**

2. จากคำถามในข้อที่ 1 จงใช้คำสั่ง sapply หรือ lapply ในทดสอบ trimmed mean หรือ mean แบบปกติ แทนที่จะใช้ for-loop (5 คะแนน)

3. จงใช้ for-loop เพื่อแสดงว่า series ดังต่อไปนี้ สามารถประมาณค่า e ได้จริงหรือไม่

$$e=\sum\_{k=0}^{\infty }\frac{1}{k!}$$

แน่นอนว่าคุณไม่สามารถสร้าง for-loop ได้ถึง infinity ได้ ให้คุณแสดงผลเมื่อกระจายผลการบวกไปจนกระทั่ง k = 3, 5, 10, 20, 50, และ 100 (5 คะแนน)

4. จงใช้ข้อมูล HBAT.csv ซึ่งเป็นไฟล์ข้อมูลจาก Hair et al. (2010) โดยข้อมูลนี้เป็นข้อมูลที่ได้จากบริษัทที่ปรึกษาทางการตลาด ที่ทำให้กับบริษัทจำหน่ายกระดาษแห่งหนึ่ง มีตัวแปรดังต่อไปนี้

|  |  |
| --- | --- |
| ตัวแปร | คำอธิบาย |
| custype | ระยะเวลาที่ลูกค้าซื้อจาก HBAT |
| industype | กลุ่มธุรกิจของลูกค้าที่ซื้อกระดาษจาก HBAT |
| firmsize | จำนวนลูกจ้างของลูกค้า (น้อยกว่า 500 คน หรือมากกว่า 500 คน) |
| region | ที่อยู่ของลูกค้า |
| dissys | วิธีการสั่งซื้อของลูกค้า (ผ่านเอเยนต์หรือซื้อขายโดยตรง) |
| product | การรับรู้ของลูกค้าว่า HBAT เด่นทางด้านคุณภาพของกระดาษหรือไม่ (1 คือแย่ 10 คือเยี่ยม) |
| ecommerce | การรับรู้ของลูกค้าว่า HBAT เด่นทางด้านเวปไซต์ (ใช้ได้ง่าย) หรือไม่ (1 คือแย่ 10 คือเยี่ยม) |
| techsup | การรับรู้ของลูกค้าว่า HBAT เด่นทางด้านการสนับสนุนทางด้านเทคนิค (ช่วยแก้ปัญหาได้) หรือไม่ (1 คือแย่ 10 คือเยี่ยม) |
| compliant | การรับรู้ของลูกค้าว่า HBAT เด่นทางด้านการจัดการกับคำร้องเรียนได้หรือไม่ (1 คือแย่ 10 คือเยี่ยม) |
| advertise | การรับรู้ของลูกค้าว่า HBAT เด่นทางด้านการโฆษณาทั้งหมดหรือไม่ (1 คือแย่ 10 คือเยี่ยม) |
| prodline | การรับรู้ของลูกค้าว่า HBAT เด่นทางด้านความลึกและกว้างของประเภทของสินค้า (เพื่อตอบสนองความต้องการลูกค้า) หรือไม่ (1 คือแย่ 10 คือเยี่ยม) |
| saleimage | การรับรู้ของลูกค้าว่า HBAT เด่นทางด้านภาพลักษณ์ของเซลล์หรือไม่ (1 คือแย่ 10 คือเยี่ยม) |
| pricing | การรับรู้ของลูกค้าว่า HBAT เด่นทางด้านราคาที่แข่งขันกับคู่แข่งหรือไม่ (1 คือแย่ 10 คือเยี่ยม) |
| warranty | การรับรู้ของลูกค้าว่า HBAT เด่นทางด้านการรับประกันและการจัดการการคืนสินค้าหรือไม่ (1 คือแย่ 10 คือเยี่ยม) |
| newprod | การรับรู้ของลูกค้าว่า HBAT เด่นทางด้านการพัฒนาและขายสินค้าใหม่หรือไม่ (1 คือแย่ 10 คือเยี่ยม) |
| billing | การรับรู้ของลูกค้าว่า HBAT เด่นทางด้านการรับคำสั่งซื้อและการเรียกเก็บเงินว่าทำได้มีประสิทธิภาพหรือไม่ (1 คือแย่ 10 คือเยี่ยม) |
| priceflex | การรับรู้ของลูกค้าว่า HBAT เด่นทางด้านความสามารถในการเจรจาต่อรองทางด้านราคาหรือไม่ (1 คือแย่ 10 คือเยี่ยม) |
| speed | การรับรู้ของลูกค้าว่า HBAT เด่นทางด้านเวลาที่ใช้ในการส่งสินค้าหลังจากมีการสั่งซื้อหรือไม่ (1 คือแย่ 10 คือเยี่ยม) |
| cussat | ความพึงพอใจของลูกค้า (1 คือแย่ 10 คือเยี่ยม) |
| recommend | แนวโน้มที่จะแนะนำ HBAT ต่อบริษัทอื่นที่ต้องการสินค้ากระดาษ (1 คือแย่ 10 คือเยี่ยม) |
| futurebuy | แนวโน้มที่จะสั่งซื้อสินค้าจาก HBAT ในอนาคต |
| percentbuy | ร้อยละของกระดาษที่สั่งซื้อจาก HBAT |
| partnership | แนวโน้มที่บริษัทจะเป็นหุ้นส่วนทางด้านกลยุทธ์กับ HBAT ในอนาคต |

 1) จงทำนาย partnership ด้วย cussat ด้วย logistic regression แปลความหมาย พร้อมทั้งสร้างกราฟแสดงความสัมพันธ์ระหว่างตัวแปรทั้งสอง โดยให้ความพึงพอใจของลูกค้าอยู่แกน *X* และสัดส่วนที่จะเป็นหุ้นส่วนอยู่แกน *Y* (5 คะแนน)

 2) จงทำนาย partnership ด้วย cussat และ industype ด้วย logistic regression แปลความหมาย พร้อมทั้งสร้างกราฟแสดงความสัมพันธ์ระหว่างตัวแปรทั้งสอง โดยให้ความพึงพอใจของลูกค้าอยู่แกน *X* และสัดส่วนที่จะเป็นหุ้นส่วนอยู่แกน *Y* และเส้นที่แตกต่างกันแสดงถึงประเภทธุรกิจที่แตกต่างกัน (5 คะแนน)

 3) จงทำนาย custype ด้วย region ด้วย Ordinal logistic regression แปลความหมาย พร้อมทั้งสร้างกราฟแท่งเพื่อแสดงสัดส่วนของลูกค้าแต่ละประเภท ในแต่ละที่อยู่ของลูกค้า ให้เหมือนหรือใกล้เคียงกับกราฟข้างล่าง (5 คะแนน)



 4) จงทำนาย custype ด้วย region และ cussat ด้วย Ordinal logistic regression แปลความหมาย พร้อมทั้งสร้างกราฟเส้นเพื่อแสดงสัดส่วนของลูกค้าแต่ละประเภท ในแต่ละที่อยู่ของลูกค้าและในแต่ละระดับของความพึงพอใจของลูกค้า โดยให้กราฟเหมือนหรือใกล้เคียงกับกราฟข้างล่าง (5 คะแนน)



 5) จงใช้ lavaan เพื่อตรวจสอบโมเดลดังต่อไปนี้ และใช้วิธี Bootstrap เพื่อตรวจสอบอิทธิพลทางอ้อมทั้ง 6 รูปแบบ (5 คะแนน)

product

complaint

speed

cussat

recommend

futurebuy

 6) จงใช้ lavaan เพื่อใช้ confirmatory factor analysis ตรวจสอบโมเดลข้างล่างว่าเหมาะสมกับข้อมูลดีหรือไม่ และตัวแปรอะไรที่มี standardized loadings สูงที่สุด (5 คะแนน)

cussat

recommend

futurebuy

percentbuy

 7) จากโมเดลในข้อที่ 6 จงตรวจสอบว่ามี measurement invariance across region หรือไม่ หากพบว่ามี scalar invariance จงหาค่าเฉลี่ยของ attitude ในแต่ละกลุ่ม พร้อมทั้งทดสอบว่าค่าเฉลี่ยของทั้งสองกลุ่มแตกต่างกันอย่างมีนัยสำคัญหรือไม่ (5 คะแนน)

 8) จงใช้ exploratory factor analysis หาองค์ประกอบภายใต้ตัวแปรต่อไปนี้ ให้เลือกจำนวนองค์ประกอบที่เหมาะสมจากวิธีการคัดเลือกองค์ประกอบด้วย scree plot, parallel test, chi-square goodness-of-fit test, RMSEA, TLI, Likelihood ratio test comparing between solutions with k and k + 1 factors, AIC, BIC (4 คะแนน)

product, ecommerce, techsup, complaint, advertise, prodline, saleimage, pricing, warranty, newprod, billing, priceflex, speed

 9) จากข้อมูลในข้อที่ 8 จงหาจำนวนองค์ประกอบที่เป็นไปได้ แล้วหมุนแกนด้วยวิธี Quartimin และคัดเลือกผลการวิเคราะห์ที่ดีที่สุด แล้วหาคะแนนองค์ประกอบด้วยวิธีของ ten Berge et al. (1999) (3 คะแนน)

10) จากข้อมูลในข้อที่ 8 จงหาจำนวนองค์ประกอบที่เป็นไปได้ แล้วหมุนแกนด้วยวิธี Bifactor และคัดเลือกผลการวิเคราะห์ที่ดีที่สุด แล้วหาคะแนนองค์ประกอบด้วยวิธีของ ten Berge et al. (1999) (3 คะแนน)