

การบ้านที่ 2 (1/2564)

รายวิชา คพ.348 แบบจำลองสำหรับวิทยาการข้อมูล

กำหนดส่ง วันที่ 10 พฤศจิกายน 2564 (ก่อนเวลา 23:55 น.)

คำชี้แจง/ข้อกำหนด

1. การบ้านชิ้นนี้เป็นการบ้านรายบุคคล นักศึกษาจะต้องทำโจทย์ปัญหาด้วยตนเองทุกข้อ หากมีการสงสัยว่านักศึกษามีการทุจริตไม่ว่าด้วยวิธีการใด ผู้สอนรายวิชานี้จะเรียกสอบสวน หากพบว่ามีการทุจริตจริง นักศึกษาจะไม่สามารถรับคะแนนของการบ้านชิ้นนี้ และผู้สอนขอสงวนสิทธิ์ในการดำเนินการตามระเบียบมหาวิทยาลัยจนถึงที่สุด
2. การบ้านมีทั้งหมด 2 ข้อ คะแนนเต็ม 100 คะแนน ดังนี้

ข้อที่	1	2
คะแนนเต็ม	60	40

3. การส่งการบ้านล่าช้าเกิน 5 วัน หลังจากกำหนดส่ง ผู้สอนขอสงวนสิทธิ์ในการไม่ตรวจให้คะแนนการบ้านไม่ว่ากรณีใดทั้งสิ้น แต่หากหากนักศึกษาส่งการบ้านล่าช้าภายใน 5 วัน คะแนนของการบ้านชิ้นนี้จะถูกหักร้อยละ 20 ของคะแนนเต็มต่อ 1 วันที่ส่งล่าช้า (โดยนักศึกษามีสิทธิ์ในการขอละเว้นการตัดคะแนนในกรณีส่งล่าช้านี้ 1 ครั้งจากจำนวนการบ้านทั้งหมด)
4. ให้นักศึกษาส่งการบ้านในส่วนที่เป็นข้อเขียน (ถ้ามี) ในรูปแบบไฟล์ PDF โดยมีรูปแบบการตั้งชื่อไฟล์ดังนี้ CS348_XXXXXXXXXX_1-2564.pdf โดยที่ XXXXXXXXXXXX คือรหัสนักศึกษา ให้นักศึกษาระบุข้อที่ตอบให้ชัดเจนสำหรับคำตอบของปัญหาแต่ละข้อ ลงในไฟล์เดียว
5. ให้นักศึกษาส่งการบ้านในส่วนที่เป็นการบ้านเขียนโปรแกรม (ถ้ามี) ในรูปแบบไฟล์ของภาษาโปรแกรมนั้น ๆ โดยมีรูปแบบการตั้งชื่อไฟล์ดังนี้ CS348_XXXXXXXXXX_YY_1-2564.ZZZ โดยที่ XXXXXXXXXXXX คือรหัสนักศึกษา 10 หลัก YY คือข้อที่นักศึกษาเขียนโปรแกรมเพื่อแก้ปัญหา และ ZZZ คือ file extension ที่สอดคล้องกับภาษาโปรแกรมที่นักศึกษาใช้ ตัวอย่าง ถ้านักศึกษาใช้ภาษา C++ เพื่อแก้ปัญหาข้อ 3 ให้นักศึกษาตั้งชื่อไฟล์ดังนี้ CS348_6209610123_03_1-2564.cpp
6. **นักศึกษาไม่ต้อง Hard copy ของการบ้านฉบับนี้**
7. หากการส่งการบ้านไม่ตรงตามข้อกำหนดในข้อ 4., 5. และ 6. การบ้านชิ้นนี้อาจไม่ได้รับการตรวจให้คะแนน
8. นักศึกษามีสิทธิ์ในการโต้แย้งคะแนนของนักศึกษาสำหรับการบ้านชิ้นนี้ภายใน 7 วันหลังจากการประกาศคะแนน หลังจากนั้นจะถือว่านักศึกษายอมรับคะแนนการบ้านชิ้นนี้โดยปราศจากข้อโต้แย้งใด ๆ

ข้อที่ 1 (ข้อเขียน, 60 คะแนน)

- 1.1 ให้นักศึกษยกตัวอย่างสถานการณ์ที่มีลักษณะที่เป็นการทดลองสุ่มที่เหมือนกันและเป็นอิสระต่อกัน (Independent and Identical) n ครั้ง มา 3 สถานการณ์ พร้อมอธิบายตามความเข้าใจของนักศึกษาว่าการที่เหตุการณ์แต่ละครั้งเหมือนกันและเป็นอิสระต่อกันหมายความว่าอย่างไร (15 คะแนน)

****** ใหยกตัวอย่างสถานการณ์นอกเหนือจาก เครื่องวัดอุณหภูมิ เครื่องวัดความเร็วแสง การโยนเหรียญ การโยนลูกเต๋า

- 1.2 กำหนดให้ X_i เป็นตัวแปรสุ่มตัวที่ i โดยตัวแปรสุ่มมีทั้งหมด n ตัว ให้แสดงวิธีทำหาค่า $\bar{X}_n, E[\bar{X}_n], Var(\bar{X}_n)$ (10 คะแนน)

- 1.3 ให้อธิบายสมการต่อไปนี้อย่างละเอียด (20 คะแนน)

1.3.1 $\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| < a) = 1$ (LoLN)

1.3.2 $\lim_{n \rightarrow \infty} F_{Z_n}(a) = \Phi(a)$ (CLT)

1.4 ให้นักศึกษาทำการทดลองที่ลิงก์ http://onlinestatbook.com/stat_sim/sampling_dist/ และถ่ายภาพการทดลองของนักศึกษาและอภิปราย

1.4.1 เกี่ยวกับ Central Limit Theorem ของค่า Means ที่พบในการทดลอง (5 คะแนน)

1.4.2 เกี่ยวกับความสัมพันธ์ระหว่าง Population distribution กับ Sampling distribution ให้ชัดเจนที่สุดและอธิบายว่าพารามิเตอร์แต่ละค่าที่เกี่ยวข้องกับ Population distribution และ Sampling distribution สำหรับค่า Sample means และ Sample proportions มีอะไรบ้าง รายละเอียดเป็นอย่างไร มีที่มาอย่างไร อย่างละเอียดที่สุด ถ้ามีข้อมูลนอกเหนือจากการทดลองในข้อ 1.3 ต้องอธิบายข้อมูลอื่นนั้นด้วย (10 คะแนน)

ข้อที่ 2 (ข้อเขียน, 40 คะแนน)

2.1 อธิบายการเปิดตาราง Z และตาราง T (อาจจะวาดหรือหาตัวอย่างเพิ่มเติมจากอินเทอร์เน็ตประกอบคำอธิบาย) ว่าค่าที่อยู่ในตารางดังกล่าวหมายถึงค่าอะไร และค่าที่สอดคล้องกับค่าที่อยู่ในตารางดังกล่าวมีค่าอะไรบ้าง (5 คะแนน)

2.2 อธิบายความหมายของช่วงความเชื่อมั่น (Confidence Interval) ว่ามีความหมายว่าอย่างไร พร้อมเขียนสูตรการหาช่วงความเชื่อมั่นสำหรับ Population Mean และ Population Proportion (5 คะแนน)

2.3 อธิบายคุณสมบัติ Random, Normal, Independence สำหรับการทำให้ Inference สำหรับค่า Mean และ Proportion ของกลุ่มประชากร (5 คะแนน)

2.4 ในการทดลองที่มีชื่อเสียงหนึ่งของ Mendel เกี่ยวกับถั่ว คือ เขาทำนายว่าร้อยละ 25 ถั่วรุ่นลูกจะมีสีเหลือง แต่ในสิ่งที่ได้จากการทดลองคือเขาได้ถั่วสีเหลือง 152 และได้สีเขียว 428 ต้น ให้เราช่วยกันหาช่วงความเชื่อมั่นที่ 95% สำหรับอัตราส่วนของถั่วสีเหลืองในรุ่นลูก ให้แสดงวิธีหาช่วงความเชื่อมั่นดังกล่าวอย่างละเอียด (10 คะแนน)

2.5 ผู้จัดการค่ายธรรมะแห่งหนึ่งซึ่งมีผู้เข้าร่วมนับพันคนสนใจว่าค่าเฉลี่ยของจำนวนจดหมายที่ผู้เข้าร่วมเขียนถึงคนข้างนอกเป็นเท่าไร สมมติว่าเรารู้ว่า standard deviation ของคนในค่ายทั้งหมดคือ 2.5 จากการสอบถามจากผู้ที่อยู่ในค่าย 25 คนได้ค่าอยู่ที่ 7.9 ฉบับต่อหนึ่งคน และ sample standard deviation คือ 2.8 (15 คะแนน)

2.5.1 ให้ระบุค่า \bar{X}, σ, n

2.5.2 ให้ระบุว่าในกรณีนี้ตัวแปรสุ่ม X และ \bar{X} หมายถึงอะไร

2.5.3 ให้สร้างช่วงความเชื่อมั่นสำหรับค่า Population Mean ของจำนวนจดหมายที่ถูกส่งออกไปนอกค่าย