

ชีวสถิติ และวิทยาการระบาด

BIostatISTICS AND EPIDEMIOLOGY

พิมพ์ครั้งที่ 2

ภักรินี ไตรสถิตย

คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

ประวัติและผลงาน

รองศาสตราจารย์ ดร. กัทรีณี ไตรสภิตย์



ประวัติการศึกษา

- พ.ศ. 2540 วิทยาศาสตร์บัณฑิต สาขาวิชาสถิติ มหาวิทยาลัยเชียงใหม่
- พ.ศ. 2541 Diplôme d'études approfondies (D.E.A) สาขา Biomathematics, DENIS DIDEROT– Paris VII UNIVERSITY
- พ.ศ. 2543 วุฒิบัตร สาขา Statistics in Biologie, CESAM/D.I.U, PIERRE ET MARIE CURIE – Paris VI UNIVERSITY
- พ.ศ. 2548 ปริญญาเอก สาขา Biomathematics, DENIS DIDEROT–Paris VII UNIVERSITY
- พ.ศ.2551 Postdoctoral, Program for HIV Prevention and Treatment

ตำแหน่งปัจจุบัน

ผู้รับผิดชอบหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติประยุกต์
ผู้รับผิดชอบหลักสูตรวิทยาศาสตรดุษฎีบัณฑิต สาขาวิชาสถิติประยุกต์
Editorial Board, Thailand Statistician: Journal of Thai Statistical Association

ผลงานวิชาการ

1. โครงการวิจัยเรื่อง "Infectious Disease Challenges and Public Health in South East Asia
2. โครงการวิจัยเรื่อง “การพัฒนาาระบบปัญญาประดิษฐ์เพื่อการตรวจวินิจฉัยและรักษาโรคมะเร็งในตับได้รวดเร็วและแม่นยำด้วยภาพถ่ายเอกซเรย์คอมพิวเตอร์
3. โครงการวิจัยเรื่อง “เมื่อโลกไซเบอร์รังแกฉัน: ผลกระทบและการป้องกันในเยาวชนผ่านสื่อสร้างสรรค์”
4. โครงการเรื่อง “การเฝ้าระวังการแพร่ระบาดของสารเสพติดจากงานนิติเวชศาสตร์ตามยุทธศาสตร์ชาติด้านความมั่นคง”
5. โครงการวิจัยเรื่อง ปัจจัยที่เกี่ยวข้องกับการพัฒนาการส่งสัยล่าช้าในเด็กปฐมวัยกลุ่มเสี่ยงของประเทศไทย

รางวัล

- พ.ศ. 2558 รางวัล “บทความวิจัยดีเด่น” ผู้มีบทความวิจัยได้รับการตีพิมพ์ในวารสารทางวิชาการระดับนานาชาติที่มีค่าการวัดความถี่ของการอ้างอิงบทความวารสาร ปี พ.ศ. 2557 สูงสุดอันดับ 2 โดยเป็นผู้เขียนชื่อแรก (First Author) หรือผู้เขียนหลัก (Corresponding Author) คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

ชีวสถิติและวิทยาการระบาด

Biostatistics and Epidemiology

พิมพ์ครั้งที่ 2

ภากรินี ไตรสถิตย

ชีวสถิติและวิทยาการระบาด

(Biostatistics and Epidemiology)

พิมพ์ครั้งที่ 2

บรรณาธิการ: กรกฎ งานวงศ์พาณิชย์
ผู้แต่ง: ภัทรินี ไตรสถิตย์
ISBN (e-Book) : 978-616-398-566-8
เจ้าของและผู้จัดพิมพ์: สำนักพิมพ์มหาวิทยาลัยเชียงใหม่
สำนักงานบริหารงานวิจัย สำนักงานมหาวิทยาลัย
มหาวิทยาลัยเชียงใหม่ 50200
โทรศัพท์: 0 5394 3603-4
โทรสาร: 0 5394 3600
<http://cmupress.cmu.ac.th>
E-mail: cmupress.th@gmail.com

พิมพ์ครั้งที่ 1: เมษายน 2561
พิมพ์ครั้งที่ 2: เมษายน 2564
ราคา: 175 บาท

ออกแบบปก: สมสืบ คราฟท์

©สงวนลิขสิทธิ์ พ.ศ. 2537 ห้ามลอกเลียนแบบ ทำซ้ำ ดัดแปลง ส่วนหนึ่งส่วนใดของหนังสือเล่มนี้ รวมทั้งการจัดเก็บถ่ายทอไม่ว่ารูปแบบหรือวิธีการใดๆ ด้วยกระบวนการทางอิเล็กทรอนิกส์ การถ่ายภาพ การบันทึกหรือวิธีการอื่นใดโดยไม่ได้รับอนุญาต

คำนิยม

สถิติเป็นเรื่องที่อยู่ใกล้ตัวมากเข้าไปเกี่ยวข้องกับหลายต่อหลายวงการ วงการหนึ่งที่สถิติเข้าไปมีบทบาทมากคือวงการแพทย์ ทำให้มีหนังสือสถิติทางการแพทย์เขียนออกมาจำนวนมาก และส่วนใหญ่จะเขียนโดยบุคลากรทางการแพทย์ที่ศึกษาความรู้ทางด้านสถิติเพิ่มเติม แต่สำหรับหนังสือ “ชีวสถิติและวิทยาการระบาด” เล่มนี้ กลับเขียนโดยนักสถิติ ที่ไปทำงานเกี่ยวข้องกับทางการแพทย์ ถือว่าเป็นหนังสือที่มีเนื้อหาครอบคลุมครบถ้วนในสาระของชีวสถิติรวมถึงสถิติที่เกี่ยวข้องกับการระบาด ครอบคลุมทั้งการเลือกใช้สถิติที่ถูกต้องกับงานด้านนี้ การคำนวณกลุ่มตัวอย่าง การออกแบบการศึกษาด้านระบาดวิทยา การจัดและวิเคราะห์ข้อมูลทางคลินิก ซึ่งเป็นหนังสือที่เหมาะสมสำหรับผู้ที่ทำงานด้านระบาดวิทยา รวมถึงผู้ทำงานงานวิจัยด้านนี้ด้วย จะสามารถทำให้ออกแบบการศึกษาได้อย่างถูกต้อง โดยผู้เขียนได้อธิบายแต่ละหัวข้ออย่างละเอียดด้วยภาษาที่ถูกต้อง ง่ายแก่การอ่านและทำความเข้าใจ มีการอ้างอิงที่ถูกต้อง ตามหลักสากล นอกจากนี้ในส่วนของเนื้อหาทางวิชาการของหนังสือเล่มนี้ก็ได้ถูกตรวจสอบความถูกต้องแล้วโดยผู้ทรงคุณวุฒิในสาขาวิชาไม่น้อยกว่าสามคน ให้ผู้อ่านสามารถเชื่อมั่นในความถูกต้องทางวิชาการของหนังสือนี้ได้ จัดว่าเป็นหนังสือเล่มหนึ่งที่เหมาะสมสำหรับใช้เป็นแบบอย่างในการเขียนหนังสือวิชาการที่ดีได้อีกด้วย

รองศาสตราจารย์ นายสัตวแพทย์ ดร.กรกฎ งานวงศ์พาณิชย์
บรรณาธิการบริหาร สำนักพิมพ์มหาวิทยาลัยเชียงใหม่

คำนำ

หนังสือชีวสถิติและวิทยาการระบาด ผู้เขียนจัดทำขึ้นเพื่อให้ผู้ที่มีความสนใจศึกษาเกี่ยวกับชีวสถิติและวิทยาการระบาดได้เข้าใจเกี่ยวกับแนวคิดทางชีวสถิติและทางวิทยาการระบาด เข้าใจถึงความจำเป็นในการมีความรู้ความเข้าใจพื้นฐานบางประการ เช่น ความรู้เกี่ยวกับความไม่แน่นอนและความน่าจะเป็น นิยามของประชากร ตัวอย่างและตัวอย่างสุ่ม เป็นต้น รวมถึงหัวข้อความรู้พื้นฐานต่าง ๆ ที่เหมาะสมสำหรับผู้สนใจศึกษาชีวสถิติและวิทยาการระบาด โดยเนื้อหาหนังสือประกอบด้วย แนวคิดทางชีวสถิติ แนวคิดทางวิทยาการระบาด การวัดความถี่ของการเกิดโรค การออกแบบการศึกษาสำหรับงานวิจัยทางวิทยาการระบาด การรายงานผู้ป่วย การศึกษาแบบภาคตัดขวาง การศึกษาแบบตามรุ่น การศึกษาแบบกลุ่มควบคุม การทดลองทางคลินิก การสุ่มและการปกปิด การจัดการข้อมูลทางคลินิก การวิเคราะห์ผลลัพธ์แบบทวิภาค การวิเคราะห์การศึกษาตามยาว และการพิจารณาขนาดตัวอย่าง ผู้เขียนหวังว่าหนังสือเล่มนี้จะเป็นประโยชน์แก่ผู้ที่สนใจศึกษาทางชีวสถิติและวิทยาการระบาด ทั้งนี้ผู้เขียนได้เล็งเห็นถึงความสำคัญในนำผลงานวิจัยจริงมาประกอบการอธิบายยกเป็นตัวอย่างเพื่อช่วยให้ผู้อ่านสามารถเข้าใจประเด็นต่าง ๆ ได้ง่ายขึ้นและสามารถนำองค์ความรู้ที่ได้ไปใช้ในการดำเนินงานวิจัยหรือการศึกษาต่อไป

ผู้เขียนขอกราบขอบพระคุณ ครู-อาจารย์ทุกท่านที่ได้ให้ความรู้แก่ผู้เขียนตลอดมา ขอขอบคุณครอบครัว ญาติพี่น้อง ผู้ใกล้ชิด เพื่อนและเพื่อนร่วมงานทุกท่านที่สนับสนุน ให้กำลังใจและส่งเสริมการผลิตหนังสือของผู้เขียนมาโดยตลอด และขอขอบคุณภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ ที่ให้การสนับสนุนและเอื้อประโยชน์ให้ผู้เขียนได้ผลิตหนังสือเล่มนี้

ผู้เขียนต้องขอบคุณผู้ประเมินหนังสือเล่มนี้ที่ได้ให้ข้อเสนอแนะที่เป็นประโยชน์ต่อการแก้ไขหนังสือเล่มนี้ให้สมบูรณ์มากยิ่งขึ้น อีกทั้งยังขอขอบคุณสำนักพิมพ์โอเดียนสโตร์ ที่ได้ให้ความกรุณาในการจัดพิมพ์หนังสือเล่มนี้ หากมีข้อผิดพลาดประการใดในหนังสือเล่มนี้ ผู้เขียนขออภัยมา ณ ที่นี้ และจะรู้สึกขอบพระคุณอย่างสูง หากท่านกรุณาช่วยแก้ไขและแจ้งกลับมาที่ผู้เขียนได้ที่ patrinee.t@cmu.ac.th เพื่อจะได้นำมาปรับแก้ไขให้ถูกต้องในโอกาสต่อไป

ภัทรีณี ไตรสถิตย์

ตุลาคม 2560

คำนำ (พิมพ์ครั้งที่ 2)

หนังสือชีวสถิติและวิทยาการระบาด ผู้เขียนจัดทำขึ้นเพื่อให้ผู้ที่มีความสนใจศึกษาเกี่ยวกับชีวสถิติและวิทยาการระบาดได้เข้าใจเกี่ยวกับแนวคิดทางชีวสถิติและทางวิทยาการระบาด เข้าใจถึงความจำเป็นในการมีความรู้ความเข้าใจพื้นฐานบางประการ เช่น ความรู้เกี่ยวกับความไม่แน่นอนและความน่าจะเป็น นิยามของประชากร ตัวอย่างและตัวอย่างสุ่ม เป็นต้น รวมถึงหัวข้อความรู้พื้นฐานต่าง ๆ ที่เหมาะสมสำหรับผู้สนใจศึกษาชีวสถิติและวิทยาการระบาด โดยเนื้อหาหนังสือ ประกอบด้วย แนวคิดทางชีวสถิติ แนวคิดทางวิทยาการระบาด การวัดความถี่ของการเกิดโรค การออกแบบการศึกษาสำหรับงานวิจัยทางวิทยาการระบาด การรายงานผู้ป่วย การศึกษาแบบภาคตัดขวาง การศึกษาแบบตามรุ่น การศึกษาแบบกลุ่มควบคุม การทดลองทางคลินิก การสุ่มและการปกปิด การจัดการข้อมูลทางคลินิก การวิเคราะห์ผลลัพธ์แบบทวิภาค การวิเคราะห์การศึกษาตามยาว การพิจารณาขนาดตัวอย่าง และได้มีการปรับปรุงเพิ่มจากการพิมพ์ครั้งที่ 1 ในบทสุดท้ายคือ กรณีสึกษาและการประยุกต์ใช้โปรแกรมสำเร็จรูปทางสถิติ ผู้เขียนหวังว่าหนังสือเล่มนี้จะเป็นประโยชน์แก่ผู้ที่สนใจศึกษาทางชีวสถิติและวิทยาการระบาด ทั้งนี้ผู้เขียนได้เล็งเห็นถึงความสำคัญในนำผลงานวิจัยจริงมาประกอบการอธิบาย ยกเป็นตัวอย่างเพื่อช่วยให้ผู้อ่านสามารถเข้าใจประเด็นต่าง ๆ ได้ง่ายขึ้นและสามารถนำองค์ความรู้ที่ได้ไปใช้ในการดำเนินงานวิจัยหรือการศึกษาต่อไป

ผู้เขียนขอกราบขอบพระคุณ ครู-อาจารย์ทุกท่านที่ได้ให้ความรู้แก่ผู้เขียนตลอดมา ขอขอบคุณครอบครัว ญาติพี่น้อง ผู้ใกล้ชิด เพื่อนและเพื่อนร่วมงานทุกท่านที่สนับสนุน ให้กำลังใจและส่งเสริมการผลิตหนังสือของผู้เขียนมาโดยตลอด และขอขอบคุณภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ ที่ให้การสนับสนุนและเอื้อประโยชน์ให้ผู้เขียนได้ผลิตหนังสือเล่มนี้

ผู้เขียนต้องขอบคุณผู้ประเมินหนังสือเล่มนี้ที่ได้ให้ข้อเสนอแนะที่เป็นประโยชน์ต่อการแก้ไขหนังสือเล่มนี้ให้สมบูรณ์มากยิ่งขึ้น อีกทั้งยังขอบพระคุณสำนักพิมพ์มหาวิทยาลัยเชียงใหม่และทีมงานของสำนักพิมพ์มหาวิทยาลัยเชียงใหม่ทุกท่านที่ได้ให้ความกรุณาในการจัดพิมพ์หนังสือเล่มนี้ หากมีข้อผิดพลาดประการใดในหนังสือเล่มนี้ ผู้เขียนขออภัยมา ณ ที่นี้ และจะรู้สึกขอบพระคุณอย่างสูง หากท่านกรุณาช่วยแก้ไขและแจ้งกลับมาที่ผู้เขียนได้ที่ patrinee.t@cmu.ac.th เพื่อจะได้นำมาปรับปรุงแก้ไขให้ถูกต้องในโอกาสต่อไป

ภทรีณี ไตรสถิตย์

เมษายน 2564

The background is a light gray with a subtle pattern of hexagons. Overlaid on this are several large, soft, white, organic shapes that resemble cells or blobs. Three DNA double helix structures are also present: one on the left side, one on the right side, and one at the bottom left corner. A white rectangular frame is positioned in the center-left of the image, containing the title text.

BIostatistics
AND EPIDEMIOLOGY

สารบัญ

คำนำ	I
สารบัญ	II
สารบัญภาพ	IX
สารบัญตาราง	

บทที่ 1 แนวคิดทางชีวสถิติ	1
1.1 ความไม่แน่นอนและความน่าจะเป็น	2
1.2 นิยามของประชากร ตัวอย่างและตัวอย่างสุ่ม	4
1.3 ความเอนเอียงและความผันแปร	5
1.4 ข้อมูลและชนิดของตัวแปร	6
1.5 ปัจจัยกวนและอันตรายกิริยา	7
1.6 สถิติเชิงพรรณนาและเชิงอนุมาน	13
1.7 การทดสอบสมมติฐานและค่าพี	14
แบบฝึกหัดท้ายบท	21
เอกสารอ้างอิง	22
เฉลยแบบฝึกหัดท้ายบท	23

บทที่ 2 แนวคิดทางวิทยาการระบาด	25
2.1 บทนำ	25
2.2 วิธีการทางวิทยาศาสตร์	26
2.3 นิยามของวิทยาการระบาด	27
2.4 สมมติฐานงานวิจัย	34
2.5 การวัด	36
2.6 ความเที่ยงของการวัด	38
2.7 ความตรงของการวัด	43
2.8 จริยธรรมการวิจัยในคน	45
2.9 บทสรุป	48
แบบฝึกหัดท้ายบท	49
เอกสารอ้างอิง	50
เฉลยแบบฝึกหัดท้ายบท	51

สารบัญ

บทที่ 3 การวัดความถี่ของการเกิดโรค	51
3.1 ความสำคัญของการวัดความถี่ของการเกิดโรค	55
3.2 ดัชนีที่ใช้สรุปข้อมูล	57
3.3 ความชุก	59
3.4 อุบัติการณ์	60
3.5 ความสัมพันธ์ระหว่างค่าความชุกและอุบัติการณ์	63
3.6 บทสรุป	64
แบบฝึกหัดท้ายบท	65
เอกสารอ้างอิง	66
เฉลยแบบฝึกหัดท้ายบท	67
 บทที่ 4 การออกแบบการศึกษาสำหรับงานวิจัยทางวิทยาการระบาด	 69
4.1 รูปแบบของการศึกษา	70
4.2 ข้อพิจารณาในการออกแบบการศึกษา	75
4.3 ประชากรที่ใช้ศึกษา	78
4.4 ปัจจัยที่สัมพันธ์และผลลัพธ์	82
4.5 การเปรียบเทียบรูปแบบการศึกษาเชิงทดลองและการศึกษาเชิงสังเกต	84
4.6 การระบุความสัมพันธ์ที่เป็นสาเหตุจากความสัมพันธ์ที่พบในการศึกษา	85
แบบฝึกหัดท้ายบท	89
เอกสารอ้างอิง	90
เฉลยแบบฝึกหัดท้ายบท	92
 บทที่ 5 การรายงานผู้ป่วย	 95
5.1 ประเภทของการรายงานผู้ป่วยจำแนกตามวัตถุประสงค์ของการวิจัย	96
5.2 องค์ประกอบของการรายงานผู้ป่วย	97
5.3 ขั้นตอนการดำเนินการวิจัยในการรายงานผู้ป่วย	99
5.4 ข้อดีและข้อจำกัดของการรายงานผู้ป่วย	99
5.5 คุณภาพของการรายงานผู้ป่วย	99
แบบฝึกหัดท้ายบท	101

สารบัญ

เอกสารอ้างอิง	102
เฉลยแบบฝึกหัดท้ายบท	103
บทที่ 6 การศึกษาแบบภาคตัดขวาง	105
6.1 ลักษณะของการศึกษาแบบภาคตัดขวาง	106
6.2 องค์ประกอบของการศึกษาแบบภาคตัดขวาง	108
6.3 ขั้นตอนการดำเนินการวิจัยในการศึกษาแบบภาคตัดขวาง	109
6.4 ข้อดีและข้อจำกัดของการศึกษาแบบภาคตัดขวาง	112
แบบฝึกหัดท้ายบท	114
เอกสารอ้างอิง	115
เฉลยแบบฝึกหัดท้ายบท	116
บทที่ 7 การศึกษาแบบตามรุ่น	119
7.1 ลักษณะของการศึกษาแบบตามรุ่น	120
7.2 องค์ประกอบของการศึกษาแบบตามรุ่น	123
7.3 ขั้นตอนการดำเนินการวิจัยในการศึกษาแบบตามรุ่น	123
7.4 ข้อดีและข้อจำกัดของการศึกษาแบบตามรุ่น	132
แบบฝึกหัดท้ายบท	135
เอกสารอ้างอิง	135
เฉลยแบบฝึกหัดท้ายบท	136
บทที่ 8 การศึกษาแบบกลุ่มควบคุม	139
8.1 ลักษณะของการศึกษาแบบกลุ่มควบคุม	139
8.2 องค์ประกอบของการศึกษาแบบกลุ่มควบคุม	141
8.3 ขั้นตอนการดำเนินการวิจัยในการศึกษาแบบกลุ่มควบคุม	147
8.4 ข้อดีและข้อจำกัดของการศึกษาแบบกลุ่มควบคุม	149
แบบฝึกหัดท้ายบท	151
เอกสารอ้างอิง	151
เฉลยแบบฝึกหัดท้ายบท	152

สารบัญ

บทที่ 9 การทดลองทางคลินิก	155
9.1 หลักการและเหตุผลของการทดลองทางคลินิก	156
9.2 ระยะของการทดลองทางคลินิก	156
9.3 การดำเนินการทดลองทางคลินิก	158
9.4 ข้อจำกัดของการวิจัยเชิงทดลองแบบสุ่มโดยมีการควบคุม	164
แบบฝึกหัดท้ายบท	165
เอกสารอ้างอิง	166
เฉลยแบบฝึกหัดท้ายบท	167
 บทที่ 10 การสุ่มและการปกปิด	 169
10.1 ตัวแบบการสุ่ม	169
10.2 วิธีการของการสุ่ม	171
10.3 การจัดสรรเข้ากลุ่ม	173
10.4 การปกปิด	174
10.5 บทสรุป	175
แบบฝึกหัดท้ายบท	176
เอกสารอ้างอิง	177
เฉลยแบบฝึกหัดท้ายบท	178
 บทที่ 11 การจัดการข้อมูลทางคลินิก	 179
11.1 แนวคิดพื้นฐานในการจัดการข้อมูลทางคลินิก	179
11.2 ความสำคัญของการปฏิบัติการวิจัยทางคลินิกที่ดี	180
11.3 กระบวนการจัดการข้อมูลที่ดีในงานวิจัยด้านวิทยาศาสตร์สุขภาพ	181
11.4 กระบวนการบริหารจัดการข้อมูล	182
11.5 บทบาทของเจ้าหน้าที่จัดการข้อมูล	187
แบบฝึกหัดท้ายบท	189
เอกสารอ้างอิง	190
เฉลยแบบฝึกหัดท้ายบท	191

สารบัญ

บทที่ 12 การวิเคราะห์ผลลัพธ์แบบทวิภาค	193
12.1 การเปรียบเทียบสัดส่วนระหว่าง 2 กลุ่ม	194
12.2 การเปรียบเทียบผลลัพธ์แบบทวิภาคระหว่าง 2 กลุ่ม	207
12.3 การถดถอยลอจิสติก	211
12.4 การควบคุมปัจจัยกวน	219
แบบฝึกหัดท้ายบท	220
เอกสารอ้างอิง	221
เฉลยแบบฝึกหัดท้ายบท	222
 บทที่ 13 การวิเคราะห์การศึกษาตามยาว	 225
13.1 การศึกษาตามยาว อัตรา และการแจกแจงปัวส์ซอง	226
13.2 การเปรียบเทียบอัตรา	226
13.3 การปรับเข้ามาตรฐาน	227
13.4 การวิเคราะห์การรอดชีพ	234
แบบฝึกหัดท้ายบท	245
เอกสารอ้างอิง	246
เฉลยแบบฝึกหัดท้ายบท	247
 บทที่ 14 การพิจารณาขนาดตัวอย่าง	 249
14.1 แนวคิดการพิจารณาขนาดตัวอย่าง	250
14.2 ความรู้พื้นฐานที่เกี่ยวข้องกับการคำนวณขนาดตัวอย่าง	252
14.3 การแจกแจงค่าสถิติจากตัวอย่าง	253
14.4 สูตรคำนวณขนาดตัวอย่างของแต่ละรูปแบบการวิจัย	263
14.5 การปรับขนาดตัวอย่างในกรณีมีผู้ตกสำรวจ	
หรือสูญหายจากการติดตาม (Dropout)	274
แบบฝึกหัดท้ายบท	276
เอกสารอ้างอิง	278
เฉลยแบบฝึกหัดท้ายบท	279

สารบัญ

บทที่ 15 การพิจารณาขนาดตัวอย่าง	283
15.1 โปรแกรมสำเร็จรูปทางสถิติ	283
15.2 การเตรียมข้อมูลสำหรับการวิเคราะห์	287
15.3 กรณศึกษาและการแปรผลการวิเคราะห์ด้วยโปรแกรมสำเร็จรูปทางสถิติ	290
แบบฝึกหัดท้ายบท	311
เอกสารอ้างอิง	315
เฉลยแบบฝึกหัดท้ายบท	316
 ภาคผนวก	 321
ดัชนี	343
Index	361
ประวัติผู้เขียน	379

สารบัญภาพ

ภาพที่ 1.1	ความสัมพันธ์เชิงทฤษฎีระหว่างการดื่มเครื่องดื่มผสมแอลกอฮอล์ โรคมะเร็งในปากและการสูบบุหรี่	8
ภาพที่ 1.2	หลักการอนุมานทางสถิติ	14
ภาพที่ 1.3	เขตวิกฤตและเขตการยอมรับสมมติฐาน	17
ภาพที่ 1.4	เขตวิกฤตแบบสองทาง	17
ภาพที่ 1.5	เขตวิกฤตแบบทางเดียวด้านขวา	18
ภาพที่ 1.6	เขตวิกฤตแบบทางเดียวด้านซ้าย	18
ภาพที่ 1.7	การคำนวณค่า p-value กรณีการทดสอบแบบ 2 ทาง	19
ภาพที่ 1.8	การคำนวณค่า p-value กรณีการทดสอบทางเดียวทางด้านซ้าย	20
ภาพที่ 1.9	การคำนวณค่า p-value กรณีการทดสอบทางเดียวทางด้านขวา	20
ภาพที่ 2.1	แผนภาพจำลองผลลัพธ์ในการปาเป้าเพื่ออธิบายความเที่ยงและความตรง	37
ภาพที่ 3.1	สถานการณ์การขาดสารไอโอดีนในประเทศไทย พ.ศ. 2558	56
ภาพที่ 3.2	ระยะเวลาของการเสี่ยงต่อการเกิดโรคเบาหวานชนิดที่ 2 ในผู้ใหญ่ที่ติดเชื้อเอชไอวี และสถานะโรคเบาหวาน	62
ภาพที่ 4.1	รูปแบบของการศึกษาทางระบาดวิทยา	77
ภาพที่ 4.2	ขั้นตอนการเลือกประชากรและตัวอย่าง	78
ภาพที่ 6.1	แนวคิดของการศึกษาแบบภาคตัดขวาง	105
ภาพที่ 6.2	องค์ประกอบของการศึกษาแบบภาคตัดขวาง	108
ภาพที่ 7.1	การศึกษาแบบตามรุ่นไปข้างหน้า (Prospective cohort study) เริ่มต้นศึกษาในปี พ.ศ. 2559	120
ภาพที่ 7.2	การศึกษาแบบตามรุ่นย้อนหลัง (Retrospective cohort study) เริ่มต้นศึกษาในปี พ.ศ. 2550	122
ภาพที่ 7.3	อัตราอุบัติการณ์โรคเบาหวานในผู้ติดเชื้อเอชไอวี (ต่อ 1,000 คน-ปี) จำแนกตามข้อมูลพื้นฐานและข้อมูล ทางเศรษฐกิจ ณ วันที่เริ่มต้นรับยาต้านไวรัส	128

สารบัญภาพ

ภาพที่ 8.1	โครงสร้างของการศึกษาแบบกลุ่มควบคุม	140
ภาพที่ 9.1	การทดลองทางคลินิก (Clinical trials)	155
ภาพที่ 11.1	กระบวนการบริหารจัดการข้อมูล	188
ภาพที่ 12.1	ตัวอย่าง 2 กลุ่มที่เป็นอิสระกัน	199
ภาพที่ 13.1	การตรวจสอบข้อสมมติเบื้องต้นของ Cox proportional hazard ด้วยแผนภาพ Kaplan-Meier	238
ภาพที่ 13.2	การตรวจสอบข้อสมมติเบื้องต้นของ Cox proportional hazard ด้วยแผนภาพการกระจายระหว่างค่าตกค้างของตัวแปรอิสระกับเวลา	253
ภาพที่ 14.1	การอนุมานทางสถิติ	253
ภาพที่ 14.2	เปรียบเทียบการแจกแจงแบบปกติและการแจกแจง t	258
ภาพที่ 14.3	การแจกแจงไคสแควร์	262
ภาพที่ 15.1	หน้าต่าง Data editor ของโปรแกรม SPSS	285
ภาพที่ 15.2	หน้าต่าง Output viewer ของโปรแกรม SPSS	286
ภาพที่ 15.3	หน้าต่าง Syntax editor ของโปรแกรม SPSS	286
ภาพที่ 15.4	ตัวอย่างแบบสอบถาม	289
ภาพที่ 15.5	ตัวอย่างการตอบแบบสอบถาม	290
ภาพที่ 15.6	ตัวอย่างการป้อนข้อมูลในโปรแกรม SPSS	290
ภาพที่ 15.7	ตัวอย่างข้อมูลด้วยโปรแกรม SPSS	291
ภาพที่ 15.8	ตัวอย่างคำสั่งสำหรับการวิเคราะห์ทางสถิติในโปรแกรม SPSS	291



สารบัญตาราง

ตารางที่ 1.1	คุณลักษณะของผู้ป่วยจำแนกตามการดื่มเครื่องดื่มผสมแอลกอฮอล์	7
ตารางที่ 1.2	ผลการศึกษาเปรียบเทียบความเสี่ยงของการเกิดโรคมะเร็งในปาก	8
ตารางที่ 1.3	สัดส่วนของผู้ที่สูบบุหรี่จำแนกตามการดื่มเครื่องดื่มผสมแอลกอฮอล์	9
ตารางที่ 1.4	ผลการศึกษาเปรียบเทียบความเสี่ยงของการเกิดโรคมะเร็งในปากในผู้ที่สูบบุหรี่ (N=114)	10
ตารางที่ 1.5	ผลการศึกษาเปรียบเทียบความเสี่ยงของการเกิดโรคมะเร็งในปากในผู้ที่ไม่สูบบุหรี่ (N=886)	10
ตารางที่ 1.6	ความสัมพันธ์ของการสูบบุหรี่และการดื่มเครื่องดื่มแอลกอฮอล์ที่มีต่อการเกิดมะเร็งกล่องเสียง	12
ตารางที่ 4.1	แสดงเกณฑ์การวินิจฉัยโรคเบาหวาน	84
ตารางที่ 6.1	จำนวนผู้มีภาวะหลอดเลือดแดงอุดตันจำแนกตามภาวะความดันโลหิตสูง	110
ตารางที่ 6.2	การคำนวณอัตราส่วนความชุกและอัตราส่วนออดส์ตามความชุก	111
ตารางที่ 7.1	ตัวอย่างการบันทึกข้อมูลการติดตามในการศึกษาแบบตามรุ่น	125
ตารางที่ 7.2	อัตราอุบัติการณ์โรคเบาหวานในผู้ติดเชื้อเอชไอวีที่รับยาต้านไวรัสในโครงการ PHPT Cohort ระหว่าง พ.ศ. 2543–2554	128
ตารางที่ 7.3	อัตราอุบัติการณ์โรคเบาหวานในผู้ติดเชื้อเอชไอวี (ต่อ 1,000 คน-ปี) จำแนกตามข้อมูลพื้นฐานและข้อมูลทางเศรษฐกิจ ณ วันที่เริ่มต้นรับยาต้านไวรัส	129
ตารางที่ 7.4	การเก็บรวบรวมข้อมูลในการวิจัยแบบไปข้างหน้า	130
ตารางที่ 8.1	ข้อมูลตัวอย่างการบริโภคอาหารที่มีไขมันสูงและการเกิดภาวะความดันโลหิตและภาวะคอเลสเตอรอลสูง	143
ตารางที่ 8.2	ข้อมูลตัวอย่างการบริโภคอาหารที่มีไขมันสูงในผู้ที่มีภาวะความดันโลหิตสูงเทียบกับคนทั่วไป	144
ตารางที่ 8.3	การนำเสนอข้อมูลจากผลการศึกษาแบบกลุ่มควบคุม โดยใช้ตาราง 2x2	147
ตารางที่ 12.1	ตารางการถ่วงของการทดสอบไคสแควร์สำหรับข้อมูล 2 กลุ่มที่เป็นอิสระกัน	201
ตารางที่ 12.2	ตารางบันทึกข้อมูลของกลุ่มตัวอย่าง 2 กลุ่มที่เกี่ยวข้องกัน	203

สารบัญตาราง

ตารางที่ 12.3	ตารางการถ่วงน้ำหนักของการทดสอบของแมคนีมาร์สำหรับข้อมูล 2 กลุ่มที่เกี่ยวข้องกัน	204
ตารางที่ 12.4	การแจกแจงข้อมูลสำหรับการทดสอบความน่าจะเป็นของฟิชเชอร์	206
ตารางที่ 12.5	การเก็บรวบรวมข้อมูลในการวิจัยแบบไปข้างหน้า	208
ตารางที่ 12.6	ตารางแจกแจงความถี่ของการทดสอบความเหมาะสมของตัวแบบ ด้วยวิธีตารางจัดจำพวก	218
ตารางที่ 13.1	เปรียบเทียบความแตกต่างระหว่างอัตราส่วน อัตรา และสัดส่วน	228
ตารางที่ 13.2	ข้อดีและข้อด้อยของการคำนวณหาอัตราแต่ละรูปแบบ	234
ตารางที่ 14.1	ตารางการสุ่มตัวอย่างของ Krejcie และ Morgan	274
ตารางที่ 15.1	ตัวอย่างคู่มือลงรหัส	289

แนวคิดทางชีวสถิติ (Concepts of biostatistics)

ชีวสถิติ (Biostatistics) เป็นการศึกษาทางด้านสถิติ (Statistics) ที่ประยุกต์เน้นไปทางวิทยาศาสตร์ชีวภาพ ประกอบด้วยการศึกษาทางชีววิทยา การศึกษาวิจัยทางการแพทย์และทางด้านบริการสุขภาพ โดยคำว่า “สถิติ” มีรากศัพท์มาจากภาษาเยอรมันว่า Statistik^[1] หมายถึงข้อมูลหรือสารสนเทศ คำว่า “ข้อมูลทางสถิติ (Statistical data)” จึงหมายถึงตัวเลขหรือข้อมูลที่ได้จากการเก็บรวบรวม เช่น อุณหภูมิร่างกาย จำนวนเด็กทารกแรกเกิด น้ำหนักและส่วนสูงของนักเรียน จำนวนผู้ป่วยที่เข้ารับการรักษาในโรงพยาบาล เป็นต้น

สถิติเป็นเครื่องมือหนึ่งที่สำคัญในกระบวนการเพื่อให้ได้มาซึ่งองค์ความรู้ ผ่านการแสวงหา หรือสร้างองค์ความรู้ที่เรียกว่าการทำวิจัย สถิติเกี่ยวข้องกับกระบวนการเก็บรวบรวมข้อมูล การนำเสนอข้อมูล การวิเคราะห์และการตีความหมายข้อมูล ซึ่งองค์ความรู้ที่ได้จะนำไปใช้ประโยชน์ในการอธิบายปรากฏการณ์ที่ศึกษาในกลุ่มที่มีขอบเขตกว้างกว่ากลุ่มที่ถูกนำมาศึกษา ในทางสถิติเรียกกลุ่มที่เราต้องการนำผลจากการศึกษานี้ไปใช้อธิบายว่าประชากร (Population) โดยเราสามารถได้องค์ความรู้เกี่ยวกับประชากรได้โดยอาศัยข้อมูลที่เราสุ่มมาเพียงบางส่วนที่เรียกว่าตัวอย่าง (Sample) ซึ่งค่าที่ใช้ในการสรุปลักษณะของกลุ่มตัวอย่างเรียกว่าค่าสถิติ (Statistic) และเมื่อมีการสุ่ม (Sampling) จึงเกี่ยวข้องกับโอกาสที่ถูกสุ่ม ค่าสถิติที่ได้นั้นจึงแปรเปลี่ยนไปตามตัวอย่างที่สุ่มมาได้ โดยค่าที่ได้จากการศึกษาหลายครั้งที่ทำการศึกษาในวัตถุประสงค์เดียวกันจึงอาจจะไม่เท่ากันเพราะโอกาส (Chance) หรือความน่าจะเป็น (Probability) ของผลของการทดลองที่ออกมาแตกต่างกัน จึงสะท้อนให้เห็นถึงความหมายของคำว่าความไม่แน่นอน (Uncertainty)

ดังนั้นในการศึกษาความน่าจะเป็นของการเกิดผลเป็นแบบใดแบบหนึ่งว่ามีค่าเท่าใด โดยอาศัยข้อมูลที่ได้จากตัวอย่างไปอธิบายลักษณะของประชากรหรือค่าพารามิเตอร์ (Parameter) ที่เรียกว่าการอนุมานทางสถิติ (Inferential statistics) จึงต้องประยุกต์ใช้ทฤษฎีความน่าจะเป็น (Probability theory) เข้าช่วย

1.1 ความไม่แน่นอนและความน่าจะเป็น (Uncertainty and probability)

ความไม่แน่นอน (Uncertainty) หมายถึง สถานภาพความลังเล สงสัย หรือความไม่แน่ใจ ส่งผลให้ผลลัพธ์เกิดความไม่แน่นอน ซึ่งความไม่แน่นอนอาจทำให้เกิดความเบี่ยงเบนของข้อมูลซึ่งเกิดมาจากปัจจัยต่าง ๆ เช่น ข้อมูลไม่ครบถ้วน แหล่งที่มาของข้อมูลขาดความน่าเชื่อถือ เป็นต้น วิธีการแก้ปัญหาความไม่แน่นอนโดยใช้ทฤษฎีสถิติที่สำคัญ ได้แก่ ทฤษฎีความน่าจะเป็น ซึ่งแนวคิดเกี่ยวกับความน่าจะเป็นเกิดขึ้นในปี ค.ศ. 1654 จากนักคณิตศาสตร์ชาวฝรั่งเศส คือ แบลส ปาสกาล (Blaise Pascal) (ค.ศ. 1623-1662) และปิแยร์ เดอ แฟร์มาต์ (Pierre de Fermat) (ค.ศ. 1601-1665) เพื่อพยายามอธิบายโอกาสที่จะชนะในการพนัน ซึ่งต่อมาแนวคิดดังกล่าวได้เป็นพื้นฐานของสถิติเชิงอนุมานและการประยุกต์ในงานวิจัยด้านต่าง ๆ^[2]

1.1.1 คำศัพท์ที่เกี่ยวข้องกับความน่าจะเป็นและตัวแปรสุ่ม

ก่อนจะกล่าวถึงการนำทฤษฎีความน่าจะเป็นไปใช้ ควรทราบถึงคำศัพท์ที่เกี่ยวข้องกับความน่าจะเป็นและตัวแปรสุ่ม^[3] ดังนี้

(1) การทดลองสุ่ม (Random experiment)

การทดลองสุ่ม หมายถึง การทดลองที่ไม่ทราบผลลัพธ์ที่จะเกิดขึ้นอย่างแน่นอนแต่สามารถบอกผลลัพธ์ที่เป็นไปได้ทั้งหมดของการทดลองสุ่มนั้น

(2) ปริภูมิตัวอย่าง (Sample space)

ปริภูมิตัวอย่าง หมายถึง เซตของผลลัพธ์ที่เป็นไปได้ทั้งหมดที่เกิดขึ้นจากการทดลองสุ่ม ใช้สัญลักษณ์แทนด้วย S เช่น ในการโยนลูกเต๋า 1 ครั้ง ปริภูมิตัวอย่างจะเท่ากับ $\{1, 2, 3, 4, 5, 6\}$

(3) เหตุการณ์ (Event)

เหตุการณ์ หมายถึง เซตย่อย (Subset) ของปริภูมิตัวอย่าง ใช้สัญลักษณ์ แทนด้วย E หรือกล่าวอีกนัยหนึ่ง เหตุการณ์คือส่วนหนึ่งของผลการทดลองสุ่มที่มีลักษณะที่เราสนใจ

(4) ความสัมพันธ์ระหว่างเหตุการณ์

จากความหมายที่กล่าวถึงเหตุการณ์ว่าเป็นเซตย่อยของปริภูมิตัวอย่าง S ซึ่งเหตุการณ์ที่เราสนใจศึกษาบางเหตุการณ์อาจเกิดขึ้นร่วมกันได้หรือมีความเกี่ยวข้องกันและอาจจะต้องใช้ความรู้ทางพีชคณิตของเซตมาประยุกต์ใช้ในการกำหนดเหตุการณ์ที่สนใจนั้นเพื่อทำให้การหาความน่าจะเป็นของเหตุการณ์ที่สนใจกระทำให้สะดวกขึ้น จึงนิยามความสัมพันธ์ระหว่างเหตุการณ์ต่าง ๆ ด้วยสัญลักษณ์การดำเนินการเกี่ยวกับเซตดังนี้

- อินเตอร์เซกชัน (Intersection) ของเหตุการณ์ใช้สัญลักษณ์ $A \cap B$ หมายถึง เหตุการณ์ที่มีสมาชิกอยู่ในเหตุการณ์ A และเหตุการณ์ B

- ยูเนียน (Union) ของเหตุการณ์ใช้สัญลักษณ์ $A \cup B$ หมายถึงเหตุการณ์ที่มีสมาชิกอยู่ในเหตุการณ์ A หรือเหตุการณ์ B หรืออยู่ทั้ง A และ B
- คอมพลีเมนต์ (Complement) ของเหตุการณ์ A ใช้สัญลักษณ์ A' หมายถึงเหตุการณ์ที่มีสมาชิกอยู่ใน S แต่ไม่อยู่ใน A

(5) ตัวแปรสุ่ม (Random variable)

ตัวแปรสุ่ม หมายถึง ฟังก์ชันที่ทำการเปลี่ยนเหตุการณ์แต่ละเหตุการณ์ที่สามารถเกิดขึ้นได้จากการทดลองให้เป็นตัวเลข ตัวแปรสุ่มนี้โดยมากเราจะนิยมแทนด้วยอักษรตัวใหญ่ภาษาอังกฤษ เช่น X , Y หรือ Z และค่าของตัวแปรสุ่มเขียนแทนด้วยอักษรตัวเล็กภาษาอังกฤษ เช่น x , y หรือ z

ตัวแปรสุ่มแบ่งเป็น 2 ชนิด คือ ตัวแปรสุ่มชนิดไม่ต่อเนื่อง และตัวแปรสุ่มชนิดต่อเนื่อง

1) ตัวแปรสุ่มชนิดไม่ต่อเนื่อง

ถ้า X เป็นตัวแปรสุ่มชนิดไม่ต่อเนื่อง (Discrete random variable) ค่าของ X จะมีได้เพียงบางค่าและมักจะเป็นจำนวนนับ จำนวนค่าของตัวแปรสุ่มชนิดไม่ต่อเนื่องอาจมีจำกัดหรือเป็นค่าอนันต์ที่นับได้ เช่น ถ้าโยนเหรียญ 1 อัน 3 ครั้ง ถ้าให้ X = จำนวนเหรียญที่ขึ้นหัว ดังนั้น X จะเป็นตัวแปรสุ่มชนิดไม่ต่อเนื่อง โดยที่ $x = 0, 1, 2, 3$ หรือจำนวนคนงานที่ลาป่วยในโรงงานแห่งหนึ่ง $x = 0, 1, 2, 3, 4, \dots$ เป็นต้น

2) ตัวแปรสุ่มชนิดต่อเนื่อง

ถ้า X เป็นตัวแปรสุ่มชนิดต่อเนื่อง (Continuous random variable) X จะมีค่าจริงในช่วงที่ต่อเนื่อง เช่น ความสูง น้ำหนัก ระยะเวลา เป็นต้น ตัวอย่างเช่น ถ้า X เป็นอายุของนักศึกษาชั้นปีที่ 1 ค่าของ X อาจจะอยู่ในช่วง 19-21 ปี เขียนได้ว่า $19 \leq x \leq 21$

1.1.2 แนวคิดการคำนวณความน่าจะเป็น (Probability)

(1) Classical approach

ในการทดลองสุ่มซึ่งแต่ละผลลัพธ์มีโอกาสเกิดขึ้นเท่ากัน ถ้า $N(S)$ เป็นจำนวนผลลัพธ์ที่เป็นไปได้ทั้งหมดของการทดลองสุ่ม และ $N(A)$ เป็นจำนวนผลลัพธ์ของเหตุการณ์ A ความน่าจะเป็นแบบ Classical approach คืออัตราส่วนระหว่าง $N(A)$ ต่อ $N(S)$ ตัวอย่างเช่น ความน่าจะเป็นที่เลือกตัวเลขมา 1 ตัวจากชุดตัวเลข 1-99 แล้วตรงกับเลข 27 มีค่าเท่ากับ 0.000102

(2) Relative frequency approach

การคำนวณความน่าจะเป็นแบบนี้เกิดจากการนับจำนวนครั้งที่เกิดขึ้นของเหตุการณ์ที่สนใจจากการทดลองสุ่ม ถ้า N เป็นจำนวนครั้งที่ทั้งหมดของการทดลองสุ่มโดยที่ $N \rightarrow \infty$ และ n เป็นจำนวนครั้งที่เกิดขึ้นของเหตุการณ์ A ความน่าจะเป็นแบบ Relative frequency approach คืออัตราส่วนระหว่าง n ต่อ N โดยความน่าจะเป็นแบบ Relative frequency approach นี้ ใช้คำนวณภายหลังการทดลองจริงโดย

อาศัยผลลัพธ์ที่เกิดขึ้น บางครั้งจึงเรียกความน่าจะเป็นแบบนี้ว่า Empirical approach ตัวอย่างเช่น จากการรวบรวมข้อมูลที่ผ่านมา สามารถคำนวณค่าความน่าจะเป็นที่ผู้ป่วยมะเร็งร้ายภายหลังการรักษา จะกลับเป็นซ้ำอีกภายใน 5 ปี เท่ากับ 0.10

(3) Subjective approach

เป็นการกำหนดความน่าจะเป็นโดยอาศัยประสบการณ์และเหตุผลส่วนตัวโดยไม่มีสูตรคำนวณที่แน่นอน บางครั้งจึงเรียกความน่าจะเป็นแบบนี้ว่า Personal approach ตัวอย่างเช่น โอกาสที่คนรักแมวจะไม่รักสุนัขมีมาก

1.2 นิยามของประชากร ตัวอย่างและตัวอย่างสุ่ม (Definition of a population, sample and random sample)

1.2.1 ประชากร (Population)

ประชากร หมายถึง หน่วยต่าง ๆ ทุกหน่วยของสิ่งที่เราสนใจจะทำการศึกษา โดยลักษณะของประชากรที่ศึกษาอาจมีจำนวนจำกัด สามารถนับจำนวนได้ เรียกว่า ประชากรจำกัด (Finite population) เช่น การศึกษานำหนักของนักศึกษาสาขาสถิติชั้นปีที่ 1 หรืออาจมีจำนวนไม่จำกัด ไม่สามารถนับจำนวนที่แน่นอนได้ เรียกว่า ประชากรอนันต์ (Infinite population) เช่น การศึกษาเกี่ยวกับประสิทธิภาพของยาชนิดหนึ่ง ประชากรจะเป็นผลการทดสอบประสิทธิภาพของยาในผู้ป่วยที่ใช้ยานี้ ซึ่งไม่สามารถบอกถึงจำนวนทั้งหมดได้ อย่างไรก็ตามผู้วิจัยจำเป็นต้องให้นิยามของประชากรที่ตนเองศึกษาไว้ให้ชัดเจนมากที่สุด เช่น ในการศึกษาเกี่ยวกับค่าเฉลี่ยของความดันโลหิตของประชากรไทย แพทย์อาจจะระบุขอบเขตในการศึกษาเฉพาะในช่วงอายุ 40-59 นอกจากนี้ยังจำเป็นต้องพิจารณาในรายละเอียดให้ชัดเจน คำว่าประชากรไทยว่าให้หมายถึงใคร เป็นต้น

1.2.2 ตัวอย่าง (Sample)

ตัวอย่าง หมายถึง ส่วนหนึ่งของประชากรที่นำมาศึกษาซึ่งเป็นตัวแทนของประชากร การที่กลุ่มตัวอย่างจะเป็นตัวแทนที่ดีของประชากรเพื่อการอ้างอิงไปยังประชากรอย่างน่าเชื่อถือได้นั้น จะต้องมีการสุ่มตัวอย่างที่สอดคล้องตามกรอบประชากรที่กำหนดไว้และจะต้องอาศัยสถิติเข้ามาช่วยในการสุ่มตัวอย่างและการกำหนดขนาดของกลุ่มตัวอย่าง ตัวอย่างที่ดีจะให้ข้อมูลของประชากรและทำให้ความเชื่อมั่นทางสถิติมีค่าสูง และยังลดค่าใช้จ่ายในการดำเนินการวิจัยด้วย

1.2.3 ตัวอย่างสุ่ม (Random sample)

ในทางสถิติเมื่อพูดถึงตัวอย่าง เรามักจะหมายถึงตัวอย่างสุ่ม ซึ่งได้แก่ตัวอย่างที่เลือกโดยอาศัยหลักเกณฑ์เกี่ยวกับความน่าจะเป็นที่จะถูกเลือกมาเป็นตัวอย่างตามที่ได้กำหนดไว้ล่วงหน้า เช่น ให้แต่ละหน่วยประชากรมีความน่าจะเป็นที่จะถูกเลือกเท่า ๆ กัน เป็นต้น

1.3 ความเอนเอียงและความผันแปร (Bias and variability)

ในการเก็บรวบรวมข้อมูล สามารถทำได้ 2 วิธี คือ 1) การทำสำมะโน (Census) ซึ่งเก็บรวบรวมข้อมูลจากทุกหน่วยประชากร อย่างไรก็ตามการทำสำมะโนเป็นไปได้ยากสำหรับการศึกษาในประชากรที่มีขนาดใหญ่ หรือในสถานการณ์ที่มีงบประมาณหรือเวลาจำกัด และ 2) การสำรวจด้วยตัวอย่าง (Sample survey) เป็นการเลือกรวบรวมข้อมูลเพียงบางส่วนของประชากร เพื่อที่จะสามารถนำข้อมูลที่ได้จากชุดข้อมูลตัวอย่างไปอ้างอิงประชากรที่ศึกษาได้ โดยข้อมูลนั้นควรได้มาจากตัวอย่างที่เป็นตัวแทนที่ดีของประชากร กล่าวคือควรเป็นตัวอย่างสุ่ม (Random sample) เลือกมาอย่างสุ่มโดยอาศัยความน่าจะเป็นในการเลือกตัวอย่าง จึงจะไม่มี ความเอนเอียงจากการเลือกหน่วยตัวอย่าง (Selection bias)

ความเอนเอียง (Bias) หรืออคติ มักเกิดขึ้นได้เสมอทั้งโดยตั้งใจและไม่ตั้งใจ ความเอนเอียงอาจเกิดได้ในขณะจัดกลุ่มทดลอง ขณะที่ทำการรวบรวมข้อมูล ขณะที่ทำการวิเคราะห์หรือการสรุปผลการศึกษา เช่น หากผู้วิจัยมีความเอนเอียงว่าวิธีใหม่รักษาได้ผลดีกว่า ก็อาจเลือกผู้ป่วยที่มีอาการไม่รุนแรงให้ได้รับรักษาด้วยวิธีใหม่ หรือรวบรวมแต่ผลที่ดีไว้กับการรักษาโดยวิธีใหม่ เป็นต้น

ในการเชื่อมโยงแนวคิดของการทำวิจัยเพื่อแสวงหาความรู้จากข้อมูลที่สนใจ โดยใช้การวัด (Measurement) ผู้วิจัยจะทำการบันทึกค่าที่ได้จากการสังเกตและกำหนดเป็นสัญลักษณ์หรือตัวเลขให้กับสิ่งที่ต้องการวัดตามกฎเกณฑ์อย่างใดอย่างหนึ่ง โดยค่าความคลาดเคลื่อนจากการวัด (Measurement error : e) คือ ค่าความแตกต่างระหว่างค่าที่วัดได้ (Observed score : X) กับค่าจริง (True ability : T) ทั้งนี้ในการวัดที่ดีควรต้องมีค่า e เกิดขึ้นน้อย ๆ การเกิดขึ้นของค่าความคลาดเคลื่อนจากการวัด สามารถจำแนกได้เป็น 2 ประเภท ดังนี้

1) ความคลาดเคลื่อนอย่างสุ่ม (Random error) อันเนื่องมาจากความแปรปรวนของค่าที่วัดได้ ซึ่งมีลักษณะและทิศทางของความคลาดเคลื่อนไม่แน่นอน มีได้ทั้งค่าบวกและลบ ความคลาดเคลื่อนประเภทนี้ไม่ร้ายแรงมากนัก สามารถจัดการกับความคลาดเคลื่อนนี้ได้ด้วยวิธีการทางสถิติ เช่น ระดับความเครียดของผู้เข้ารับการทดสอบสมรรถภาพร่างกายที่แตกต่างกัน อาจมีส่วนที่ส่งผลให้คะแนนการทดสอบสมรรถภาพร่างกายแตกต่างกัน โดยหากผู้เข้ารับการทดสอบมีความเครียดน้อย การปฏิบัติของเขาก็อาจจะได้คะแนนสูงกว่าเมื่อผู้ที่มีความเครียดมาก

2) ความคลาดเคลื่อนอย่างเป็นระบบ (Systematic error) เกิดขึ้นจากรูปแบบ ขั้นตอน หรือวิธีการศึกษาที่ไม่ถูกต้อง ส่งผลให้การวัดคลาดเคลื่อนจากความเป็นจริง ซึ่งทำให้เกิดผลกระทบค่อนข้างรุนแรง โดยความคลาดเคลื่อนอย่างเป็นระบบ อาจแยกตามสาเหตุการเกิดได้ เช่น เกิดจากสภาพแวดล้อมทางธรรมชาติ ในขณะที่ทำการวัดนั้น ไม่ใช่สภาวะเดียวกันกับสภาวะมาตรฐานของการใช้เครื่องมือที่ระบุไว้ เกิดจากความไม่สมบูรณ์ของเครื่องมือ เกิดจากความไม่ระมัดระวัง การขาดความชำนาญของผู้ปฏิบัติงานหรือเกิดจากเครื่องคำนวณอันเนื่องจากการปัดเศษตัวเลข เช่น กรณีที่เครื่องชั่งน้ำหนักคลาดเคลื่อนต่ำกว่าค่ามาตรฐานไป 0.5 กิโลกรัม ส่งผลให้ข้อมูลน้ำหนักของผู้ป่วยทุกคนที่ใช้เครื่องชั่งนี้เกิดความคลาดเคลื่อนอย่างเป็นระบบ

1.4 ข้อมูลและชนิดของตัวแปร (Data and types of variables)

ข้อมูล (Data) คือ ข้อเท็จจริงหรือรายละเอียดของสิ่งต่าง ๆ ที่เราสนใจ เช่น น้ำหนัก ส่วนสูง เพศ อายุ ความดันโลหิต ฯลฯ ซึ่งได้มาจากการเก็บรวบรวมข้อมูลด้วยวิธีการต่าง ๆ ทั้งจากการสังเกต การเก็บแบบสอบถาม ตลอดจนการทดลอง ซึ่งข้อมูลที่ได้จากกระบวนการดังกล่าวเรียกว่า ข้อมูลดิบ (Raw data) เราสามารถแบ่งประเภทของข้อมูลเป็น 2 ประเภท คือ

1) ข้อมูลเชิงคุณภาพ (Qualitative data) หมายถึง ข้อมูลที่วัดค่าเป็นตัวเลขไม่ได้โดยตรง แต่พิจารณาจากคุณสมบัติของข้อมูล เช่น เพศชายหรือหญิง หมู่โลหิตของคนไข้ ซึ่งมีหมู่เลือด A, B, AB และ O เป็นต้น

2) ข้อมูลเชิงปริมาณ (Quantitative data) หมายถึง ข้อมูลที่วัดค่าเป็นตัวเลขได้โดยตรง เช่น น้ำหนัก ส่วนสูง อุณหภูมิ เป็นต้น

นอกจากนี้ในการเก็บรวบรวมข้อมูลจำนวนมาก เราจะทำการกำหนดตัวแปร (Variable) เพื่อใช้แทนข้อมูล ช่วยทำให้การเก็บรวบรวม ตลอดจนการวิเคราะห์ข้อมูลทำได้สะดวกยิ่งขึ้น โดยเราสามารถแบ่งตัวแปรออกเป็น 2 ชนิด คือ

1) ตัวแปรไม่ต่อเนื่อง (Discrete variable) หมายถึง ตัวแปรที่มีค่าไม่ต่อเนื่อง ค่าของตัวแปรเป็นเลขจำนวนเต็ม ซึ่งอาจจะอยู่ในรูปของจำนวนนับหรือความถี่ เช่น จำนวนครั้งของการเกิดอุบัติเหตุ จำนวนผู้ป่วยที่รอดชีพจากการรักษา เป็นต้น

2) ตัวแปรต่อเนื่อง (Continuous variable) หมายถึง ตัวแปรที่มีค่าต่อเนื่อง ค่าของตัวแปรเป็นเลขจำนวนจริง ค่าที่ได้จากการวัด เช่น ส่วนสูง น้ำหนักทารกแรกเกิด อายุการใช้งานอุปกรณ์ในห้องผ่าตัด เป็นต้น

1.5 ปัจจัยกวนและอันตรกิริยา (Confounding factor and interaction)^[4]

1.5.1 ปัจจัยกวน (Confounding factor)

ปัจจัยกวน (Confounding factor) หมายถึง ปัจจัยที่มีอิทธิพลต่อความสัมพันธ์ระหว่างตัวแปรต้นกับตัวแปรตามที่ทำการศึกษาอยู่ ปัจจัยกวนอาจบิดเบือนหรือส่งผลกระทบต่อความสัมพันธ์ระหว่างตัวแปรต้นกับตัวแปรตาม โดยปัจจัยกวนนั้นต้องมีความสัมพันธ์กับทั้งตัวแปรต้นและตัวแปรตามเสมอ ซึ่งถ้าไม่พบความสัมพันธ์ของปัจจัยกวนกับตัวแปรทั้งคู่ ก็จะไม่ถือเป็นปัจจัยกวนและในความสัมพันธ์ที่พบนั้นปัจจัยกวนต้องไม่ใช่สาเหตุของตัวแปรต้นที่ศึกษาอยู่ หากผู้วิจัยต้องการทราบว่าปัจจัยกวนอยู่หรือไม่ ให้ผู้วิจัยเริ่มจากการพิจารณาเบื้องต้นว่ามีปัจจัยใดอีกที่สามารถทำให้เกิดผลลัพธ์ (Outcome) ที่ศึกษานี้ และเมื่อทำการทดสอบทางสถิติ ก็สามารถพิจารณาได้จากค่าสถิติอย่างหยาบ (Crude ratio) หากพบว่ามีความต่างจากค่าสถิติที่มีการปรับค่า (Adjusted ratio) แสดงว่ามีปัจจัยกวน

การทบทวนวรรณกรรมหรือการศึกษาจากรายงานวิจัยเพื่อหาปัจจัยความสัมพันธ์และปัจจัยเสี่ยงจะช่วยให้ผู้วิจัยสามารถกำหนดได้ว่าปัจจัยใดเป็นปัจจัยกวนและการทบทวนวรรณกรรมที่ครอบคลุมจะช่วยในการกำหนดปัจจัยที่มีอิทธิพลได้ครบถ้วน ซึ่งหลักการประเมินปัจจัยกวนโดยทั่วไปประกอบด้วยขั้นตอนดังต่อไปนี้

- 1) ประเมินความสัมพันธ์ระหว่างตัวแปรต้นที่ต้องการศึกษาและตัวแปรที่คาดว่าจะปัจจัยกวน
- 2) ประเมินความสัมพันธ์ระหว่างผลลัพธ์และตัวแปรที่คาดว่าจะปัจจัยกวน
- 3) ประเมินว่าตัวแปรต้นและตัวแปรที่คาดว่าจะปัจจัยกวนมีความสัมพันธ์เชิงสาเหตุหรือไม่

ตัวอย่างที่ 1 นักวิจัยทำการศึกษาเพื่อประเมินอิทธิพลของการสูบบุหรี่ที่อาจเป็นปัจจัยกวนระหว่างความสัมพันธ์ของการดื่มเครื่องดื่มผสมแอลกอฮอล์กับการเกิดโรคมะเร็งในปาก โดยทำการเก็บรวบรวมข้อมูลจากผู้ป่วยจำนวน 1,000 ราย ได้ผลการวิจัยดังนี้

ตารางที่ 1.1 คุณลักษณะของผู้ป่วยจำแนกตามการดื่มเครื่องดื่มผสมแอลกอฮอล์

	ดื่มเครื่องดื่มผสมแอลกอฮอล์ (N=500)	ไม่ดื่มเครื่องดื่มผสมแอลกอฮอล์ (N=500)
เพศชาย	285 (57.0%)	279 (55.8%)
สูบบุหรี่	70 (14.0%)	44 (8.7%)

จากตารางที่ 1.1 จะเห็นว่าสัดส่วนของผู้ที่สูบบุหรี่ในกลุ่มที่ดื่มและไม่ดื่มเครื่องดื่มผสมแอลกอฮอล์แตกต่างกัน แสดงให้เห็นว่าการสูบบุหรี่มีความสัมพันธ์กับการดื่มเครื่องดื่มผสมแอลกอฮอล์ที่เป็นตัวแปรต้น