สำนักพิมพ์มหาวิทยาลัยธรรมศาสตร์

เศรษฐมิติเบื้องต้น
Introductory Econometrics

พนิต วัฒนกูล

เศรษฐมิติเบื้องต้น

Introductory Econometrics

เศรษฐมิติเบื้องต้น
Introductory Econometrics

ดร.พนิต วัฒนกูล
ผู้ช่วยศาสตราจารย์
คณะเศรษฐศาสตร์ มหาวิทยาลัยธรรมศาสตร์

สำนักพิมพ์มหาวิทยาลัยธรรมศาสตร์
2565

พนิต วัฒนกูล.
เศรษฐมิติเบื้องต้น = Introductory Econometrics.
1. เศรษฐมิติ.

HB139
ISBN (ebook) 978-616-314-946-6

ฉบับอิเล็กทรอนิกส์ (ebook) ธันวาคม 2565

จัดพิมพ์และจัดจำหน่ายโดย**สำนักพิมพ์มหาวิทยาลัยธรรมศาสตร์**
ท่าพระจันทร์: อาคารธรรมศาสตร์ 60 ปี ชั้น U1 มหาวิทยาลัยธรรมศาสตร์
ถนนพระจันทร์ กรุงเทพฯ 10200 โทร. 0-2223-9232
ศูนย์รังสิต: อาคารโดมบริหาร ชั้น 3 ห้อง 317 มหาวิทยาลัยธรรมศาสตร์
ตำบลคลองหนึ่ง อำเภอคลองหลวง จังหวัดปทุมธานี 12120
โทร. 0-2564-2859-60 โทรสาร 0-2564-2860
http://thammasatpress.tu.ac.th, e-mail: unipress@tu.ac.th

ราคาเล่มละ 230.- บาท

สารบัญ

คำนำ

ตำราวิชาเศรษฐมิติเบื้องต้นเล่มนี้จัดทำขึ้นไม่เพียงเพื่อให้นักศึกษาคณะเศรษฐศาสตร์ใช้ประกอบการเรียนในชั้นเรียนวิชาเศรษฐมิติเบื้องต้นเท่านั้น แต่ยังเพื่อเผยแพร่เครื่องมือการวิเคราะห์ทางเศรษฐมิติที่เป็นประโยชน์ต่อการวิเคราะห์ทฤษฎีเศรษฐศาสตร์กับข้อมูลเชิงประจักษ์ทั้งในทางวงการวิชาการและภาคธุรกิจอีกด้วย ทั้งนี้เพื่อความสะดวกและประโยชน์ในการศึกษา แต่ละบทจะมีจุดประสงค์และคำสำคัญเพื่อให้ผู้อ่านได้ใช้เป็นหลักยึดในการอ่าน อย่างน้อยที่สุดเมื่อศึกษาจนจบบทแล้ว ผู้อ่านควรที่เข้าใจเนื้อหาตามจุดประสงค์และคำสำคัญต่างๆที่ระบุไว้ในเบื้องต้น

ตำราเล่มนี้แบ่งออกเป็นสองภาค โดยภาคแรกเริ่มต้นด้วยการแนะนำ การประยุกต์ใช้ พร้อมทั้งนำเสนอประโยชน์ที่ได้จากการใช้แบบจำลองถดถอยประกอบด้วยบทที่ 1 ถึงบทที่ 5 ซึ่งในบทแรกนั้นจะแนะนำวิชาเศรษฐมิติที่เป็นการผสมผสานศาสตร์สาขาต่างๆเข้าด้วยกันและวิธีการประยุกต์ใช้ ในบทที่ 2 ได้นำเสนอวิชาสถิติพื้นฐานที่จำเป็นสำหรับศึกษาเศรษฐมิติในบทถัดๆไป จากนั้นจะศึกษาเฉพาะแบบจำลองถดถอยสองตัวแปรในบทที่ 3 และผู้สนใจเพิ่มเติมสามารถศึกษาการพิสูจน์สูตรได้ในท้ายบท ลำดับถัดไป จะมีการเพิ่มตัวแปรอิสระเข้าไปในแบบจำลองและศึกษาเครื่องมือทางสถิติในการทดสอบสมมติฐานหลากหลายประเภทมาประยุกต์ใช้ให้สอดคล้องกับทฤษฎีทางเศรษฐศาสตร์ในบทที่ 4 และบทสุดท้ายของภาคแรกจะนำตัวแปรเชิงคุณภาพเข้ามาในแบบจำลองเพื่ออธิบายตัวแปรต้น

ในภาคที่สองตั้งแต่บทที่ 6 ถึง 9 นั้น จะเป็นการศึกษาเกี่ยวกับการละเมิดข้อสมมติที่สำคัญในการสร้างแบบจำลองทางเศรษฐมิติ ประกอบด้วยปัญหาความสัมพันธ์กันของตัวแปรอิสระ ปัญหาตัวรบกวนเชิงสุ่มมีความแปรปรวนไม่คงที่ ปัญหาตัวรบกวนเชิงสุ่มไม่เป็นอิสระต่อกันและปัญหาการระบุแบบจำลองผิดพลาด ทั้งในแง่ของลักษณะและผลที่เกิดขึ้น วิธีการตรวจสอบ ตลอดจนวิธีการแก้ไขปัญหาดังกล่าว ในภาคที่สองนี้ ผู้เขียนมิได้เพิ่มเติมส่วนของการพิสูจน์ทางคณิตศาสตร์ของปัญหาแต่ละประเภทเนื่องจากจำต้องอาศัยทักษะทางคณิตศาสตร์ในระดับที่สูงขึ้นไปเกินกว่าในระดับเบื้องต้น ซึ่งผู้สนใจสามารถศึกษาเพิ่มเติมได้ในตำราเศรษฐมิติระดับสูงขึ้นไป

ผู้เขียนขอขอบคุณ อ.ดร.วานิสสา เสือนิล สำหรับการช่วยอ่านและตรวจทาน ตลอดจนคุณเพ็ญพิมล บุญบันดาลฤทธิ์ คุณธนพล กองพาลี คุณภัทริศ สิทธิกูล คุณชนิกานต์ สุขมั่นคงเสมอ คุณศวิตา ชาลีวงษ์ และนักศึกษาทุกคนสำหรับข้อคิดเห็นในการจัดทำต้นฉบับ ท้ายที่สุด ขอขอบคุณคณะเศรษฐศาสตร์ มหาวิทยาลัยธรรมศาสตร์ที่ให้การสนับสนุนในการจัดทำตำราเล่มนี้ ผู้เขียนหวังเป็นอย่างยิ่งว่าตำราเล่มนี้จะเป็นประโยชน์ต่อนักศึกษา นักวิชาการ และผู้สนใจทั่วไปที่ต้องการศึกษาการใช้เครื่องมือทางเศรษฐมิติเพื่อตอบปัญหาทางเศรษฐศาสตร์ของสังคมไทยต่อไปในอนาคต

พนิต วัฒนกูล
คณะเศรษฐศาสตร์ มหาวิทยาลัยธรรมศาสตร์

บทที่ 1

อารัมภบท

จุดประสงค์

- เพื่อแนะนำวิชาเศรษฐมิติแก่ผู้อ่าน ตลอดจนความหมายและขั้นตอนกระบวนการใช้เศรษฐมิติกับปัญหาทางเศรษฐศาสตร์
- เพื่อบ่งชี้ถึงความแตกต่างระหว่างการวิเคราะห์ถดถอย (regression analysis) ความสัมพันธ์ระหว่างเหตุและผล (causation) และสหสัมพันธ์ (correlation)
- เพื่อแสดงลักษณะของข้อมูลที่สามารถนำมาประยุกต์ใช้ในวิชาเศรษฐมิติ

คำสำคัญ

- เศรษฐมิติ
- คณิตเศรษฐศาสตร์
- แบบจำลองทางเศรษฐศาสตร์
- แบบจำลองพหุสมการ
- แบบจำลองสมการเดี่ยว
- ตัวแปรอิสระ
- ตัวแปรต้น
- ตัวแปรตาม
- ตัวรบกวนเชิงสุ่ม
- แบบจำลองถดถอยเชิงเส้น
- การวิเคราะห์ถดถอย
- ความสัมพันธ์ระหว่างเหตุและผล
- สหสัมพันธ์
- ข้อมูลอนุกรมเวลา
- ข้อมูลภาคตัดขวาง
- ข้อมูลภาคตัดขวางตามเวลา
- ข้อมูลแผง

1.1 ความสำคัญและความจำเป็นในการศึกษาวิชาเศรษฐมิติ

เศรษฐศาสตร์เป็นศาสตร์เพื่อศึกษาและค้นคว้าวิธีการจัดสรรทรัพยากรที่มีอยู่อย่างจำกัด เพื่อตอบสนองความต้องการที่ไม่สิ้นสุดของมนุษย์อย่างมีประสิทธิภาพสูงสุด นักเศรษฐศาสตร์จึงพยายามสร้างทฤษฎีเศรษฐศาสตร์ขึ้นมาอธิบายปรากฏการณ์ทางเศรษฐกิจและเพื่อเสนอแนวทางการแก้ไขปัญหาที่อาจจะเกิดขึ้นในระบบอีกด้วย

อย่างไรก็ดี ทฤษฎีทางเศรษฐศาสตร์ที่สามารถอธิบายการเปลี่ยนแปลงของพฤติกรรมทางเศรษฐกิจเพียงอย่างเดียวอาจไม่เพียงพอสำหรับการเสนอแนวนโยบายสำหรับการจัดการบางสิ่งที่เฉพาะเจาะจง เช่น ไม่สามารถระบุชัดได้ว่ารัฐบาลควรกระตุ้นเศรษฐกิจผ่านการใช้จ่ายภาครัฐเท่าใดเพื่อให้เศรษฐกิจมีเสถียรภาพ กล่าวคือ ทำให้เศรษฐกิจขยายตัวในระดับที่เหมาะสมโดยไม่สร้างภาวะฟองสบู่ในระบบเศรษฐกิจและภาระทางการคลังให้กับตัวรัฐบาลเอง ดังนั้น นักเศรษฐศาสตร์จึงได้เริ่มนำเครื่องมือสถิติและคณิตศาสตร์มาประยุกต์ใช้กับทฤษฎีเศรษฐศาสตร์ โดยนิยามการประยุกต์ใช้ดังกล่าวว่า **เศรษฐมิติ (Econometrics)**

เศรษฐมิติ (Econometrics) เป็นการนำเครื่องมือทางสถิติและคณิตศาสตร์มาคำนวณหรือประมาณค่าตัวแปรต่าง ๆ ตาม*แบบจำลองทางเศรษฐศาสตร์ (Economic Model)* ที่ได้คิดค้นขึ้นจากการใช้*คณิตเศรษฐศาสตร์ (Mathematical Economics)* ดังนั้น เศรษฐมิติเป็นวิชาที่มีประโยชน์ในแง่ของการนำทฤษฎีเศรษฐศาสตร์มาเชื่อมโยงกับเหตุการณ์ทางเศรษฐกิจที่เกิดขึ้นจริง

เพื่อให้เห็นภาพของการใช้เครื่องมือทางเศรษฐมิติ สมมติให้บริษัท ก. ซึ่งขายสินค้า X ต้องการทราบถึงกลยุทธ์ที่บริษัทควรจะใช้เพื่อให้ได้กำไรสูงสุด สิ่งที่บริษัทจำต้องศึกษาคืออุปสงค์ของสินค้า X ที่บริษัท ก. ขายได้เอง ทฤษฎีเศรษฐศาสตร์กล่าวว่า เมื่อราคาสินค้า (P_X) ปรับตัวสูงขึ้น ความต้องการในการบริโภค (Q_X^d) จะลดลง ซึ่งคณิตเศรษฐศาสตร์จะสร้างแบบจำลองเพื่ออธิบายเหตุการณ์ดังกล่าวเป็นสมการทางคณิตศาสตร์ได้ดังนี้

$$Q_X^d = a + bP_X$$

จากนั้น นักเศรษฐศาสตร์จะรวบรวมข้อมูลราคาและปริมาณของสินค้า X แล้วใช้เศรษฐมิติในการประมาณค่าพารามิเตอร์ a และ b เพื่อให้บริษัท ก. ทราบถึงอุปสงค์ของสินค้า X และสามารถนำข้อมูลดังกล่าวไปคำนวณหาความยืดหยุ่นของอุปสงค์ต่อราคาของสินค้า X พร้อมทั้งกำหนดยุทธศาสตร์ในการปรับราคาสินค้าต่อไป

1.2 ขั้นตอนการวิเคราะห์แบบจำลองเศรษฐมิติ

สำหรับขั้นตอนการวิเคราะห์แบบจำลองเศรษฐมิตินั้น[1] โดยทั่วไปมีทั้งสิ้น 7 ขั้นตอน ในบทนี้จะทดลองนำทฤษฎีการบริโภคของเคนส์ในวิชาหลักเศรษฐศาสตร์มหภาคมาประยุกต์ใช้เศรษฐมิติเพื่อสร้างสมการถดถอยหาความสัมพันธ์ระหว่างรายได้และรายจ่ายเพื่อการบริโภค

1. เลือกทฤษฎีหรือสมมติฐานที่สนใจ

ในทางเศรษฐศาสตร์ มีทฤษฎีที่สามารถอธิบายพฤติกรรมการใช้จ่ายเพื่อการบริโภคมากมาย หนึ่งในนั้นคือทฤษฎีการบริโภคของ จอห์น เมย์นาร์ด เคนส์ (John Maynard Keynes) นักเศรษฐศาสตร์ชาวอังกฤษที่ได้เสนอแนวคิดไว้ว่า มนุษย์จะบริโภคสินค้าและบริการเพิ่มขึ้น เมื่อระดับรายได้ของพวกเขาเพิ่มขึ้น แต่อัตราส่วนการบริโภคและรายได้ที่เพิ่มขึ้นนั้นจะไม่เพิ่มในอัตราส่วนที่เท่ากัน หรือกล่าวอีกในหนึ่ง แนวโน้มการบริโภค (Marginal Propensity to Consume: MPC) หรืออัตราการเปลี่ยนแปลงของการบริโภคต่อการเปลี่ยนแปลงรายได้จะมีค่ามากกว่า 0 แต่น้อยกว่า 1

2. กำหนดรูปแบบแบบจำลองทางคณิตศาสตร์ของทฤษฎีดังกล่าว

ทฤษฎีดังกล่าวข้างต้นระบุถึงการเปลี่ยนแปลงของการบริโภคและรายได้ แต่ไม่ได้กำหนดชัดถึงอัตราการเปลี่ยนแปลงหรือแม้แต่รูปแบบของแบบจำลองทางคณิตศาสตร์ที่จะอธิบายทฤษฎีดังกล่าว นักคณิตเศรษฐศาสตร์จึงอาจกำหนดแบบจำลองได้ดังต่อไปนี้

$$C = a + bY \text{ และ } 0 < b < 1 \tag{1.1}$$

โดยกำหนดให้

C คือ รายจ่ายเพื่อการบริโภค
Y คือ รายได้ประชาชาติ และ
a กับ b คือ พารามิเตอร์ (parameters) ของแบบจำลองที่แสดงค่าคงที่และค่าความชัน ตามลำดับ

แบบจำลองทฤษฎีการบริโภคดังกล่าวเป็น **แบบจำลองสมการเดี่ยว (single-equation model)** แต่หากในแบบจำลองนั้นมีหลายสมการที่จะอธิบายความสัมพันธ์ของตัวแปรต่างๆ เราจะเรียกว่า **แบบจำลองพหุสมการ (multiple-equation model)** จากแบบจำลองข้างต้น แนวโน้มการบริโภค (MPC) จะมีค่าเท่ากับพารามิเตอร์ b สมการดังกล่าวอธิบายถึงทฤษฎีการบริโภคของเคนส์ เมื่อมีการเปลี่ยนแปลงรายได้ (Y) ซึ่งเป็น**ตัวแปรอิสระหรือตัวแปรต้น (independent variable)** จะทำให้การใช้จ่ายเพื่อการบริโภคซึ่งเป็น**ตัวแปรตาม (dependent variable)** เปลี่ยนแปลงไปด้วย

[1]Gujarati, Damodar N. and Dawn C. Porter, *Basic Econometrics*, fifth edition, McGraw-Hill Education, 2009

3. กำหนดรูปแบบแบบจำลองทางสถิติหรือทางเศรษฐมิติ

ลำดับถัดมาคือการกำหนดแบบจำลองทางเศรษฐมิติเพื่ออธิบายการใช้จ่ายของครัวเรือนเพื่อการบริโภค ดังนี้

$$C_t = a + bY_t + u_t \tag{1.2}$$

u_t คือ **ตัวรบกวนเชิงสุ่มหรือตัวแปรสุ่มคลาดเคลื่อน (disturbance หรือ random term)** ซึ่งเป็นตัวแปรสุ่ม (random variable)[2]

ทั้งนี้วิชาเศรษฐมิติเบื้องต้นจะประมาณค่าพารามิเตอร์ a และ b จากการใช้**แบบจำลองถดถอยเชิงเส้น (Linear Regression Model)** เพื่อพิสูจน์ว่าทฤษฎีการบริโภคดังกล่าวนั้นสามารถอธิบายสิ่งที่เกิดขึ้นจริงในระบบเศรษฐกิจได้มากน้อยเพียงใด โดยมีตัวรบกวนเชิงสุ่มทำหน้าที่เป็นปัจจัยอื่นๆ ไม่ได้อยู่ในแบบจำลองที่สามารถอธิบายตัวแปรด้านซ้ายได้ เราจะศึกษารายละเอียดเรื่องนี้ในบทที่ 3

4. รวบรวมข้อมูล

สำหรับข้อมูลการบริโภคภาคเอกชนและผลิตภัณฑ์มวลรวมภายในประเทศของประเทศไทยระหว่าง พ.ศ. 2533 - 2564 นั้น สภาพัฒนาการเศรษฐกิจและสังคมแห่งชาติ (สศช.) เป็นผู้รวบรวม ดังที่ได้แสดงในตารางที่ 1-1

ตารางที่ 1-1: ข้อมูลการใช้จ่ายเพื่อการบริโภคภาคเอกชน (C) และผลิตภัณฑ์มวลรวมภายในประเทศ (Y) ณ ราคาคงที่แบบดัชนีลูกโซ่ หรือ CVM (ล้านล้านบาท)

ปี	C	Y	ปี	C	Y	ปี	C	Y
2533	1.89	3.37	2544	3.02	5.44	2555	4.71	8.90
2534	1.99	3.66	2545	3.21	5.77	2556	4.76	9.14
2535	2.18	3.99	2546	3.45	6.18	2557	4.79	9.23
2536	2.36	4.34	2547	3.70	6.57	2558	4.91	9.52
2537	2.55	4.69	2548	3.86	6.85	2559	5.05	9.85
2538	2.76	5.07	2549	3.97	7.19	2560	5.21	10.26
2539	2.89	5.36	2550	4.01	7.58	2561	5.45	10.69
2540	2.85	5.21	2551	4.13	7.71	2562	5.66	10.92
2541	2.56	4.81	2552	4.09	7.66	2563	5.61	10.25
2542	2.67	5.03	2553	4.31	8.23	2564	5.63	10.40
2543	2.85	5.25	2554	4.39	8.30			

ที่มา: สศช. ณ กันยายน พ.ศ. 2565

[2]รายละเอียดของตัวแปรสุ่มสามารถทบทวนได้ในบทที่ 2

5. ประมาณค่าตัวแปรของแบบจำลองทางเศรษฐมิติ

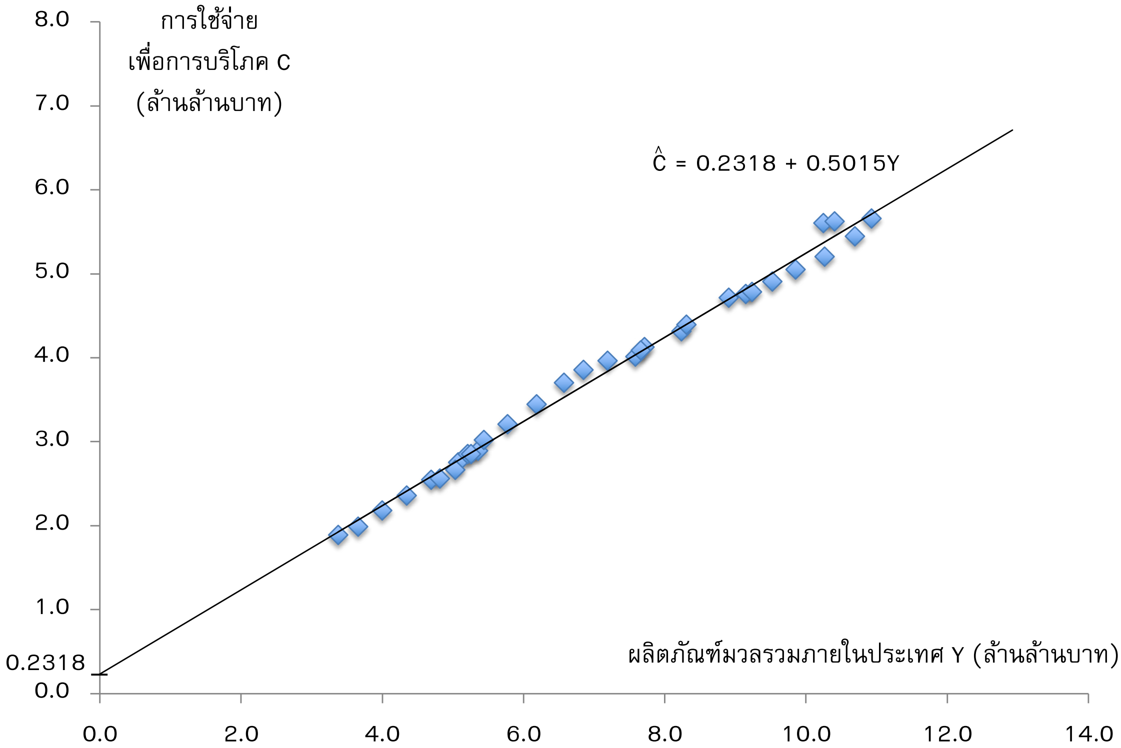
เมื่อได้ข้อมูลจากการใช้จ่ายเพื่อการบริโภคและผลิตภัณฑ์มวลรวมภายในประเทศของไทยแล้ว เราสามารถใช้เครื่องมือทางเศรษฐมิติเพื่อสร้างสมการถดถอย (regression equation) วิเคราะห์ลักษณะของข้อมูลทั้งสองชนิดและสามารถประมาณค่าพารามิเตอร์ a และ b ออกมาได้ดังนี้

$$\hat{C}_t = 0.1021 + 0.525Y_t \tag{1.3}$$

$\hat{C}_t$ คือ รายจ่ายเพื่อการบริโภคที่ถูกประมาณค่าจากแบบจำลอง และ
Y_t คือ ผลิตภัณฑ์มวลรวมภายในประเทศ

ภาพที่ 1-1 แสดงฟังก์ชันการบริโภคตามสมการ 1.3 เราได้สร้างแบบจำลองข้างต้นโดยใช้ข้อมูลระหว่างปี พ.ศ. 2533 ถึง 2564 พบว่า เมื่อรายได้ที่แท้จริงเปลี่ยนแปลงไป 1 ล้านล้านบาท โดยเฉลี่ยแล้ว ระบบเศรษฐกิจจะมีค่าใช้จ่ายเพื่อการบริโภคที่แท้จริงเพิ่มขึ้น 0.5015 ล้านล้านบาท และเมื่อระดับผลิตภัณฑ์มวลรวมภายในประเทศมีค่าเท่ากับ 0 ประชาชนยังคงต้องบริโภคเพื่อความอยู่รอดที่ 0.2318 ล้านล้านบาท

ภาพที่ 1-1: การใช้จ่ายเพื่อการบริโภคและผลิตภัณฑ์มวลรวมภายในประเทศ ณ ราคาคงที่แบบดัชนีลูกโซ่ ตั้งแต่ปี พ.ศ. 2533 - 2564 (ล้านล้านบาท)



6. ทดสอบสมมติฐาน

หลังจากที่ได้สมการ 1.3 จากวิธีการประมาณค่าตัวแปรทางเศรษฐมิติแล้ว สิ่งที่เราจะดำเนินการต่อไปคือทดสอบสมมติฐานว่าตัวแปรทั้ง a และ b ที่มีค่าเท่ากับ 0.2318 และ 0.5015 ตามลำดับนั้นมีนัยสำคัญทางสถิติหรือไม่[3]

โดยเริ่มต้นพิจารณาเครื่องหมายและขนาดของตัวแปรดังกล่าวว่า*สอดคล้องกับทฤษฎีเศรษฐศาสตร์หรือไม่* ดังที่จะเห็นได้จากเครื่องหมายที่บ่งบอกทิศทางการเปลี่ยนและขนาดของตัวแปร b ที่เป็นบวกและอยู่ระหว่าง 0 ถึง 1 ซึ่งสอดคล้องกับทฤษฎีของเคนส์ที่กล่าวไว้ข้างต้น สำหรับพารามิเตอร์ a นั้น มีขนาดเป็นบวกซึ่งเป็นไปตามทฤษฎีที่ว่าแม้คนจะไม่มีรายได้ก็ตาม แต่เพื่อความอยู่รอด ผู้คนจำต้องใช้จ่ายเพื่อการบริโภคอยู่ดี ทั้งนี้ค่าใช้จ่ายดังกล่าวอาจมาจากการขายทรัพย์สินที่มีอยู่เดิมหรือจากการกู้ยืม

จากนั้นจึงวินิจฉัยต่อไปว่า ตัวแปรอิสระ (ผลิตภัณฑ์มวลรวมภายในประเทศ) มีความสามารถในการอธิบายการเปลี่ยนแปลงของตัวแปรตาม (การใช้จ่ายเพื่อการบริโภค) ได้มากน้อยเพียงใด โดยเราจะศึกษาเพิ่มเติมต่อไปในบทที่ 3 และบทที่ 4 นอกจากนั้น ยังต้องทดสอบถึงความถูกต้องและความน่าเชื่อถือของแบบจำลองอีกด้วยว่าการประมาณค่าดังกล่าวนี้ประสบปัญหาต่างๆที่เราจะกล่าวถึงในบทที่ 6 ถึงบทที่ 9 หรือไม่ ถ้าหากประสบปัญหาดังกล่าว เราก็อาจจะเสนอแนวทางการแก้ไขซึ่งจะได้ศึกษากันต่อไป

7. ประยุกต์ใช้แบบจำลองเศรษฐมิติเพื่อพยากรณ์หรือกำหนดนโยบาย

สมมติว่าแบบจำลองเศรษฐมิติที่เรายกตัวอย่างมาข้างต้นสามารถใช้ได้จริงในทางสถิติ เราอาจจะนำตัวแปรในแบบจำลองมาคำนวณหาตัวทวีคูณ (multiplier) สำหรับกรณีเศรษฐกิจมหภาคระบบปิดและรัฐบาลไม่เก็บภาษีเงินได้ ตัวทวีคูณจะมีค่าดังต่อไปนี้

$$\text{Multiplier} = \frac{1}{1 - \text{MPC}} = \frac{1}{1 - b} \tag{1.4}$$

เช่นนี้แล้ว เราสามารถคำนวณหาค่าตัวทวีคูณได้เนื่องจากค่าแนวโน้มการบริโภค หรือ Marginal Propensity to Consume (MPC) ถูกประมาณค่าจากแบบจำลองเศรษฐมิติแล้ว ซึ่งมีค่าเท่ากับ 0.5015 พบว่าตัวทวีคูณมีค่าเท่ากับ 2.006 ความหมายคือ หากภาคเอกชนหรือรัฐบาลใช้จ่ายเพิ่มขึ้น 1 บาท ตัวทวีคูณจะช่วยกระตุ้นเศรษฐกิจทำให้ผลิตภัณฑ์มวลรวมภายในประเทศเพิ่มขึ้น 2.006 บาท ผู้ออกแบบนโยบายเศรษฐกิจมหภาคจึงอาจนำข้อมูลดังกล่าวไปใช้เพื่อวางแผนดำเนินนโยบายต่อไปได้

นอกจากนี้ สมมติว่าผู้กำหนดนโยบายคาดการณ์ว่าหากประชาชนจับจ่ายใช้สอยเป็นมูลค่า 4.5 ล้านล้านบาทแล้วนั้น จะทำให้อัตราการว่างงานในประเทศลดลงร้อยละ 0.5 หากการคาดคะเนดังกล่าวเป็นจริง รัฐบาลต้องวางแผนเพื่อกระตุ้นเศรษฐกิจทำให้ผลิตภัณฑ์มวลรวมภายในประเทศปรับตัวอยู่ในระดับที่ทำให้มูลค่าการใช้จ่ายเพื่อการบริโภคปรับตัวตามเป้าหมายให้ได้ โดยสามารถคำนวณได้จากสมการดังต่อไปนี้

[3]รายละเอียดของการทดสอบสมมติฐานสามารถศึกษาเพิ่มเติมได้ในบทที่ 3

$$\begin{aligned}
\hat{C}_t &= 0.2318 + 0.5015Y_t \\
4.5 &= 0.2318 + 0.5015Y_t \\
Y_t &= \tfrac{4.5-0.2318}{0.5015} \\
\therefore Y_t &= 8.5108
\end{aligned}$$

จากการคำนวณพบว่ามูลค่าของผลิตภัณฑ์มวลรวมภายในประเทศจำต้องปรับตัวสูงขึ้นถึงระดับ 8.5108 ล้านล้านบาท จึงจะทำให้ค่าใช้จ่ายเพื่อการบริโภคเป็นไปตามเป้าหมายและลดระดับการว่างงานลงได้ตามกำหนด

1.3 ความแตกต่างระหว่างการถดถอย (Regression) ความสัมพันธ์ระหว่างเหตุและผล (Causation) และสหสัมพันธ์ (Correlation)

การวิเคราะห์ถดถอย (Regression analysis) เป็นการหาความสัมพันธ์ทางสถิติระหว่างตัวแปรตามและตัวแปรต้น โดยเชื่อว่าตัวแปรต้นจะมีความสามารถในการอธิบายตัวแปรตามได้ ซึ่งการวิเคราะห์ดังกล่าวไม่จำเป็นต้องหมายความว่าตัวแปรต้นนั้นคือสาเหตุทำให้ตัวแปรตามเปลี่ยนแปลง (Causation) เสมอไป การวิเคราะห์สาเหตุและผลนั้นเพียงแค่ข้อมูลและวิธีการทางสถิติไม่เพียงพอสำหรับการสรุปความสัมพันธ์ นักเศรษฐศาสตร์จำต้องอาศัยข้อมูลอื่นๆเพิ่มเติม เช่น ทฤษฎีเศรษฐศาสตร์หรือสามัญสำนึกทั่วไปอีกด้วย

ตัวอย่างเช่น โดยทั่วไป เราเชื่อว่าผลผลิตข้าวต่อไร่จะขึ้นอยู่กับปริมาณน้ำฝนที่ตกลงในช่วงเวลาเพาะปลูกข้าว (causation) และเราสามารถใช้การวิเคราะห์ถดถอยเพื่อหาความสัมพันธ์ของตัวแปรทั้งสองได้ (regression) ในทางกลับกัน ปริมาณน้ำฝนไม่มีทางจะขึ้นอยู่กับผลผลิตข้าวต่อไร่ได้โดยสามัญสำนึก (no causation) แต่ในทางสถิติแล้ว เราสามารถให้ผลผลิตข้าวต่อไร่เป็นตัวแปรต้นและปริมาณน้ำฝนเป็นตัวแปรตามเพื่อวิเคราะห์ถดถอยได้ (still regression) แม้จะมีความหมายทางสถิติแต่ไม่มีเหตุผลหรือทฤษฎีใดๆรองรับการวิเคราะห์ดังกล่าวได้ ดังนั้น โดยหลักแล้ว การวิเคราะห์ถดถอยที่ดีควรคำนึงถึงสาเหตุและผลที่มีทฤษฎีรองรับ มิฉะนั้นการวิเคราะห์ดังกล่าวจะไม่นำมาซึ่งผลสรุปที่สมเหตุสมผล

สำหรับสหสัมพันธ์ (correlation) กับการวิเคราะห์ถดถอยนั้น (regression) เครื่องมือทางสถิติทั้งสองถูกนำมาใช้หาความสัมพันธ์ของตัวแปรต่างๆได้เหมือนกัน แต่การตีความจากการวิเคราะห์ถดถอยและสหสัมพันธ์จะเป็นไปในทางที่ต่างกัน

โดยสหสัมพันธ์มีเป้าหมายเพื่อวัดระดับความสัมพันธ์ของตัวแปรทั้งสองตัวทั้งสองทาง เช่น นักสถิติอาจต้องการหาความสัมพันธ์ระหว่างน้ำหนักและส่วนสูงของบุคคล คะแนนระหว่างวิชาฟิสิกส์และวิชาคณิตศาสตร์ เป็นต้น ในขณะที่การวิเคราะห์ถดถอยจะเป็นการประมาณหรือทำนายค่าเฉลี่ยของตัวแปรหนึ่งจากค่าคงที่ของอีกตัวแปรหนึ่ง กล่าวคือ เราต้องการทำนายคะแนนวิชาฟิสิกส์โดยใช้คะแนนวิชาคณิตศาสตร์ หรือทำนายน้ำหนักโดยใช้ส่วนสูงของบุคคล เป็นต้น

ข้อแตกต่างอีกประการที่น่าสนใจ คือ การวิเคราะห์สหสัมพันธ์นั้น ตัวแปรทั้งสองตัวที่เราต้องการหาความสัมพันธ์จะถือว่ามีสถานะเท่ากัน คือสุ่ม (random หรือ stochastic) ซึ่งตรงข้ามกับการวิเคราะห์

ถดถอย เนื่องจากตัวแปรอิสระหรือตัวแปรต้นจะถูกสมมติว่าเป็นค่าคงที่ (fixed value and nonstochastic) ที่มาจากการสุ่มตัวอย่างซ้ำ ๆกัน (repeated sampling) และตัวแปรตามจะสุ่ม (random) ตามแต่ละค่าของตัวแปรต้น เช่น การศึกษาระดับรายได้และรายจ่ายเพื่อการบริโภคต่อเดือนของหมู่บ้านแห่งหนึ่งกำหนดให้ตัวแปรต้นและตัวแปรตาม คือ ระดับรายได้ครัวเรือนต่อเดือน และระดับรายจ่ายเพื่อการบริโภคของครัวเรือนต่อเดือน ตามลำดับ ผู้ศึกษาจะศึกษาครัวเรือนที่มีระดับรายได้ที่ 12,000 บาทต่อเดือน (fixed value and nonstochastic) ว่ามีรายจ่ายเพื่อการบริโภคเท่าใด โดยการสุ่มทุกครัวเรือนที่มีระดับรายได้ที่ 12,000 บาทต่อเดือน (repeated sampling) จากนั้นจึงพิจารณาครัวเรือนที่มีระดับรายได้ที่ 13,000 บาทต่อเดือน และระดับรายได้อื่นๆต่อไป ในขณะที่การวิเคราะห์สหสัมพันธ์จะนำเพียงข้อมูลที่มีทั้งหมดมาคำนวณหาค่าสถิติโดยไม่มีการแบ่งแยกสถานะของแต่ละตัวแปรแต่อย่างใด

1.4 ประเภทของข้อมูลในการวิเคราะห์ทางเศรษฐมิติ

ข้อมูลในทางเศรษฐศาสตร์มีหลากหลายประเภท โดยทั่วไป นักเศรษฐศาสตร์จะนำทฤษฎีประกอบกับข้อมูลต่างๆมาประมวลผลและวิเคราะห์เพื่ออธิบายเหตุการณ์ต่างๆ เช่น อัตราการขยายตัวทางเศรษฐกิจมีผลอย่างไรต่อการเปลี่ยนแปลงของดัชนีตลาดหลักทรัพย์ของไทย ราคาน้ำมันในต่างประเทศส่งผลอย่างไรต่ออัตราเงินเฟ้อภายในประเทศ เป็นต้น อีกทั้ง ข้อมูลบางชนิดอาจถูกบันทึกในคาบเวลาที่ต่างกัน เช่น สศช. จะประกาศมูลค่าผลผลิตมวลรวมภายในประเทศรายไตรมาสและรายปี ในขณะที่ดัชนีตลาดหลักทรัพย์จะถูกเก็บข้อมูลทุกวันทำการ เป็นต้น

ด้วยลักษณะข้อมูลที่หลากหลาย จึงได้มีการแบ่งประเภทของข้อมูลออกเป็น 4 ประเภท เพื่อให้เราทราบถึงคุณสมบัติเฉพาะของข้อมูลแต่ละประเภทนำไปใช้ในการวิเคราะห์เศรษฐมิติ โดยข้อมูลทั้ง 4 ประเภท มีดังต่อไปนี้

1.ข้อมูลอนุกรมเวลา (Time series data) คือ ข้อมูลที่เก็บในแต่ละช่วงเวลา เช่น ต่อปี ต่อเดือน ต่อสัปดาห์ เป็นต้น ตารางที่ 1-2 เป็นตัวอย่างของข้อมูลอนุกรมเวลา เราจะศึกษากันต่อไปในบทที่ 8 ว่าจากลักษณะของการเก็บข้อมูลของตัวอย่างเดิมในหลายช่วงเวลา ข้อมูลอนุกรมเวลาจึงมีแนวโน้มที่จะก่อให้เกิดปัญหาตัวรบกวนเชิงสุ่มไม่เป็นอิสระต่อกัน (Autocorrelation)

ตารางที่ 1-2: ตัวอย่างข้อมูลอนุกรมเวลา

ปี	GDP	C	I	G	X - M
2533	4.2	2.0	3.0	1.0	-2.2
2534	4.6	2.1	2.1	1.4	-1.0
⋮	⋮	⋮	⋮	⋮	⋮
2565	8.4	2.3	4.0	2.0	0.1

2.ข้อมูลภาคตัดขวาง (Cross-sectional data) คือ ข้อมูลที่เก็บจากกลุ่มตัวอย่างในช่วงระยะเวลาหนึ่ง เช่น ข้อมูลสำมะโนประชากรที่สอบถามข้อมูลหลาย ๆอย่างจากกลุ่มตัวอย่าง (ตารางที่ 1-3) จากลักษณะของข้อมูลที่มีความหลากหลายนี้เอง บทที่ 7 จะกล่าวว่าเมื่อเราใช้ข้อมูลภาคตัดขวางกับแบบจำลองที่

สนใจ แบบจำลองนั้นมักจะประสบปัญหาตัวรบกวนเชิงสุ่มมีความแปรปรวนไม่คงที่ (Heteroscedasticity)

ตารางที่ 1-3: ตัวอย่างข้อมูลภาคตัดขวาง

ตัวอย่างปี 2565	รายได้	การศึกษา	เพศ	สถานะสมรส	รายจ่าย
นาย ก. (1)	15,000	ปริญญาตรี	ชาย	สมรส	10,000
นาง ข. (2)	35,000	ปริญญาโท	หญิง	โสด	30,000
นาย ค. (3)	10,000	ปวส.	ชาย	โสด	15,000
⋮	⋮	⋮	⋮	⋮	⋮
นาง Z (500)	100,000	ปริญญาเอก	หญิง	สมรส	20,000

3.ข้อมูลภาคตัดขวางตามเวลา (Pooled data) คือ ข้อมูลที่เก็บจากกลุ่มตัวอย่างที่แตกต่างกันในแต่ละช่วงเวลา ตัวอย่างเช่นในตารางที่ 1-4 ที่นักสถิติได้ทำการสอบถามข้อมูลด้านต่าง ๆจากประชาชน 500 คน ในปี 2533 เรื่อยไปจนถึงปี 2565 อย่างไรก็ดี กลุ่มตัวอย่างทั้ง 500 คนของแต่ละปีจะแตกต่างกันไป ขึ้นอยู่กับว่าในปีนั้น ๆจะสุ่มสัมภาษณ์ประชาชนคนใด

ตารางที่ 1-4: ตัวอย่างข้อมูลภาคตัดขวางตามเวลา

ตัวอย่าง	รายได้	การศึกษา	เพศ	สถานะสมรส	รายจ่าย
ปี 2533					
นาย ก. (1)	10,000	มัธยมศึกษา	ชาย	โสด	2,000
นาง ข. (2)	20,000	ปริญญาตรี	หญิง	โสด	18,000
⋮	⋮	⋮	⋮	⋮	⋮
นาย ง. (500)	350,000	ปริญญาเอก	ชาย	สมรส	400,000
ปี 2534					
นาง ฮ. (1)	20,000	ปริญญาโท	หญิง	โสด	1,000
นาง ค. (2)	22,000	ปริญญาตรี	หญิง	โสด	20,000
⋮	⋮	⋮	⋮	⋮	⋮
นาย จ. (500)	350,000	ปริญญาเอก	ชาย	สมรส	500,000
⋮	⋮	⋮	⋮	⋮	⋮
ปี 2565					
นาย A (1)	40,000	ปริญญาตรี	ชาย	สมรส	35,000
นาง B (2)	70,000	ปริญญาตรี	หญิง	สมรส	40,000
⋮	⋮	⋮	⋮	⋮	⋮
นาง Z (500)	80,000	ปริญญาตรี	หญิง	สมรส	100,000

4.ข้อมูลแผง (Panel data) คือ ข้อมูลที่เก็บจากกลุ่มตัวอย่างที่เหมือนกันในแต่ละช่วงเวลา จากตารางที่ 1-5 แสดงอย่างข้อมูลแผงที่ผู้วิจัยได้ทำการสัมภาษณ์ประชาชนทั้งสิ้น 500 คน ตั้งแต่ปี 2533 ถึง ปี 2565 โดยในแต่ละปี จะมีการเจาะจงถามคำถามเดิมกับประชาชนคนเดิม (ทั้งหมด 500 คน) เพื่อดูแนวโน้มการเปลี่ยนแปลงเมื่อเวลาผ่านไป

ตารางที่ 1-5: ตัวอย่างข้อมูลแผง

ตัวอย่าง	รายได้	การศึกษา	เพศ	สถานะสมรส	รายจ่าย
ปี 2533					
นาย ก. (1)	10,000	มัธยมศึกษา	ชาย	โสด	2,000
นาง ข. (2)	20,000	ปริญญาตรี	หญิง	โสด	18,000
⋮	⋮	⋮	⋮	⋮	⋮
นาย ง. (500)	350,000	ปริญญาเอก	ชาย	สมรส	400,000
ปี 2534					
นาย ก. (1)	12,000	มัธยมศึกษา	ชาย	โสด	4,000
นาง ข. (2)	22,000	ปริญญาตรี	หญิง	โสด	20,000
⋮	⋮	⋮	⋮	⋮	⋮
นาย ง. (500)	360,000	ปริญญาเอก	ชาย	สมรส	420,000
⋮	⋮	⋮	⋮	⋮	⋮
ปี 2565					
นาย ก. (1)	15,000	ปริญญาตรี	ชาย	โสด	10,000
นาง ข. (2)	35,000	ปริญญโท	หญิง	โสด	30,000
⋮	⋮	⋮	⋮	⋮	⋮
นาย ง. (500)	800,000	ปริญญาเอก	ชาย	สมรส	780,000

สรุป

1.เศรษฐมิติคือวิชาที่ผสมผสานระหว่างศาสตร์วิชาสามสาขาด้วยกัน ได้แก่ คณิตศาสตร์ สถิติ และเศรษฐศาสตร์ เพื่อพิสูจน์ทฤษฎีเศรษฐศาสตร์ด้วยหลักฐานเชิงประจักษ์ โดยสามารถนำเศรษฐมิติไปประยุกต์ใช้ในทางปฏิบัติได้อย่างหลากหลาย

2.ขั้นตอนการวิเคราะห์ทางเศรษฐมิติมี 7 ขั้นตอน คือ เลือกทฤษฎีหรือสมมติฐานที่ต้องการทดสอบ จากนั้นจึงกำหนดรูปแบบจำลองทางคณิตศาสตร์เพื่อนำมาปรับใช้กับแบบจำลองเศรษฐมิติ แล้วจึงรวบรวมข้อมูล พร้อมทั้งประมาณค่าตัวแปรของแบบจำลองและทดสอบสมมติฐานว่า ตัวประมาณค่าดังกล่าวมีนัยสำคัญทางสถิติหรือไม่ เมื่อแบบจำลองเศรษฐมิติถูกต้องจึงนำมาประยุกต์ใช้เพื่อค้นหาคำตอบต่อคำถามทางเศรษฐศาสตร์ต่อไป

3.ในทางปฏิบัติ การวิเคราะห์ถดถอยจะต้องคำนึงถึงความเหมาะสมของเหตุและผลก่อนเสมอ ส่วนสหสัมพันธ์ก็มีความแตกต่างจากการวิเคราะห์ถดถอยตรงที่การแปลความหมายทางสถิติ สหสัมพันธ์พิจารณาความสัมพันธ์ระหว่างสองตัวแปรทั้งสองทาง แต่การวิเคราะห์ถดถอยเป็นการพยายามอธิบายตัวแปรตัวหนึ่งด้วยตัวแปรอีกตัวหนึ่ง

4.ข้อมูลแบ่งออกเป็น 4 ชนิด คือ ข้อมูลอนุกรมเวลา (Time series data) ข้อมูลภาคตัดขวาง (Cross-sectional data) ข้อมูลภาคตัดขวางตามเวลา (Pooled data) และข้อมูลแผง (Panel data)

บทที่ 2
สถิติเบื้องต้น

จุดประสงค์

- เพื่อทบทวนความรู้ทางสถิติสำหรับประยุกต์ใช้กับวิชาเศรษฐมิติ
- เพื่อพิจารณาตัวประมาณค่าและคุณสมบัติที่ดีของตัวประมาณค่าซึ่งเป็นส่วนสำคัญในการวิเคราะห์และปรับใช้วิชาเศรษฐมิติกับข้อมูลจริง

คำสำคัญ

- ผลรวม
- ความน่าจะเป็น
- ตัวแปรสุ่ม
- ตัวแปรสุ่มที่มีค่าไม่ต่อเนื่อง
- ตัวแปรสุ่มที่มีค่าต่อเนื่อง
- รูปแบบการแจกแจงหลายตัวแปร
- รูปแบบการแจกแจงความน่าจะเป็นร่วม
- รูป แบบ การ แจกแจง ความ น่า จะ เป็น ตาม ขอบ
- รูปแบบการแจกแจงความน่าจะเป็นแบบมีเงื่อนไข
- ความเป็นอิสระทางสถิติ
- ค่าคาดหวัง
- ความแปรปรวน
- ความแปรปรวนร่วม
- สหสัมพันธ์
- ทฤษฎีขีดจำกัดกลาง
- รูปแบบการแจกแจงแบบปกติ
- รูปแบบการแจกแจงแบบไคกำลังสอง
- รูปแบบการแจกแจงแบบที
- รูปแบบการแจกแจงแบบเอฟ
- ความคลาดเคลื่อนที่เกิดจากการสุ่ม
- ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนที่เกิดจากการสุ่ม
- ความโน้มเอียง
- ความต่อเนื่อง

เพื่อเป็นการเตรียมความพร้อมก่อนเริ่มศึกษาวิชาเศรษฐมิติ เราจะเริ่มทบทวนความรู้ทางสถิติที่จำเป็นต่อการวิเคราะห์ทางเศรษฐมิติ โดยเราจะทบทวนนิยามและคุณสมบัติของการรวม (summation) ที่จำเป็นสำหรับการรวมตัวแปรต่างๆ มาเน้นย้ำ จากนั้นจะกล่าวถึงความน่าจะเป็นและตัวแปรสุ่ม ลำดับถัดมาคือการอธิบายรูปแบบการแจกแจงที่สำคัญสำหรับการศึกษา พร้อมทั้งอภิปรายเกี่ยวกับตัวประมาณค่าและคุณสมบัติของตัวประมาณค่าที่ดีอีกด้วย[1]

2.1 นิยามและคุณสมบัติของผลรวม (Summation)

เครื่องหมาย $\sum$ (ซิกมา) ในทางคณิตศาสตร์หมายถึงการรวมตัวแปรจากตัวแปร X ที่ i (สัญลักษณ์ด้านล่างเครื่องหมาย $\sum$) จนถึงตัวแปร X ที่ n (สัญลักษณ์ด้านบนเครื่องหมาย $\sum$) โดยสามารถเขียนเป็นสัญลักษณ์ทางคณิตศาสตร์ได้ดังนี้

$$\sum_{i=1}^{n} X_i = X_1 + X_2 + \dots + X_n \tag{2.1}$$

เช่น หากเรามีข้อมูลสามชุดได้แก่ $X_1 = 1$ $X_2 = 2$ และ $X_3 = 3$ เมื่อเราต้องการหาผลรวมของชุดข้อมูลนี้จะได้ว่า

$$\sum_{i=1}^{3} X_i = X_1 + X_2 + X_3 = 1 + 2 + 3 = 6$$

คุณสมบัติของเครื่องหมายผลรวม (Summation) มีดังต่อไปนี้

1. $\sum_{i=1}^{n} k = nk$

2. $\sum_{i=1}^{n} kX_i = k\sum_{i=1}^{n} X_i$

3. $\sum_{i=1}^{n}(a + bX_i) = na + b\sum_{i=1}^{n} X_i$

4. $\sum_{i=1}^{n}(X_i + Y_i) = \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} Y_i$

กำหนดให้ a, b และ k เป็นค่าคงที่

[1] Gujarati, Damodar N. *Econometrics By Example*. Palgrave Macmillian, 2011

สำหรับการรวมหลายครั้ง (Multiple Summation) นั้น คือการรวมตัวแปรที่อยู่ในรูปเมตริก ซึ่งเป็นไปตามสมการด้านล่าง

$$\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij} = X_{11} + X_{12} + \dots + X_{1m} + X_{21} + X_{22} + \dots + X_{2m} + \dots + X_{nm} \qquad (2.2)$$

โดย

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1m} \\ X_{21} & X_{22} & \dots & X_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{nm} \end{bmatrix}_{nXm}$$

ยกตัวอย่างเช่น หากเรามีเมตริก **X** ดังแสดงด้านล่าง และต้องการผลรวมของสมาชิกในเมตริก **X** เราสามารถเขียนสัญลักษณ์ทางคณิตศาสตร์ได้ดังนี้

$$\mathbf{X} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}_{2X2}$$

$$\sum_{i=1}^{2}\sum_{j=1}^{2} X_{ij} = X_{11} + X_{12} + X_{21} + X_{22} = 1 + 2 + 3 + 4 = 8$$

คุณสมบัติของ Multiple summation มีดังต่อไปนี้

1. $\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij} = \sum_{j=1}^{m}\sum_{i=1}^{n} X_{ij}$

2. $\sum_{i=1}^{n}\sum_{j=1}^{m} X_i Y_j = \sum_{i=1}^{n} X_i \times \sum_{j=1}^{m} Y_j$

3. $\sum_{i=1}^{n}\sum_{j=1}^{m}(X_{ij} + Y_{ij}) = \sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{m} Y_{ij}$

4. $(\sum_{i=1}^{n} X_i)^2 = \sum_{i=1}^{n} X_i^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} X_i X_j$

2.2 Sample Space, Sample Point and Events

Sample Space (SS) คือ เซตที่เป็นไปได้จากการทดลอง เช่น การโยนเหรียญสองครั้งนั้น ผลลัพธ์ที่เป็นไปได้ทั้งหมด คือ หัวทั้งสองครั้ง หัวครั้งที่หนึ่งและก้อยครั้งที่สอง ก้อยครั้งที่หนึ่งและหัวครั้งที่สอง และก้อยทั้งสองครั้ง (โดยกำหนดให้ H แทนเหตุการณ์ที่เหรียญออกหน้าหัว และ T แทนเหตุการณ์ที่เหรียญออกหน้าก้อย) ผลที่เป็นไปได้จากการทดลองทั้งหมดเรียกว่า Sample Space ดังนี้

$$SS = \{HH, HT, TH, TT\}$$

Sample Point (SP) คือ สมาชิกแต่ละตัวของ Sample Space เช่น การได้หัวทั้งสองครั้ง หัวครั้งที่หนึ่งและก้อยครั้งที่สอง ก้อยครั้งที่หนึ่งและหัวครั้งที่สอง หรือก้อยทั้งสองครั้ง จากการโยนเหรียญสองครั้ง

$$SP = HH \text{ หรือ } HT \text{ หรือ } TH \text{ หรือ } TT$$

Events คือ เหตุการณ์จากการทดลองบางส่วนจากผลของการทดลองทั้งหมด เช่น เหตุการณ์ที่จะได้หัวจากการโยนเหรียญทั้งสองครั้ง

$$A = \text{เหตุการณ์ที่จะได้หัวทั้งสองครั้งจากการโยนเหรียญ} = \{HH\}$$

ลักษณะของเหตุการณ์ที่น่าสนใจคือ *เหตุการณ์ที่ไม่อาจเกิดร่วมกัน (mutually exclusive)* คือ การเกิดขึ้นของเหตุการณ์หนึ่งไปแล้วทำให้ไม่อาจเกิดอีกเหตุการณ์หนึ่งได้อีก เช่น ในการโยนเหรียญหนึ่งครั้ง สองเหรียญพร้อมกันนั้น กำหนดให้เหตุการณ์ C คือเหตุการณ์ที่ได้ผลลัพธ์ออกหน้าหัวทั้ง 2 เหรียญ และเหตุการณ์ D ได้หน้าก้อยทั้ง 2 เหรียญ จะเห็นได้ว่า เหตุการณ์ C และ D ไม่อาจเกิดร่วมหรือพร้อมกันได้

สำหรับกรณีที่เหตุการณ์อาจเกิดร่วมกันได้นั้น เช่น สมมติให้เหตุการณ์ E คือ การจั่วไพ่ 52 ใบให้ได้คิง และเหตุการณ์ F ได้ดอกจิก ซึ่งมีความเป็นไปได้ว่าเหตุการณ์ E และ F จะเกิดขึ้นพร้อมกัน กล่าวคือ จั่วได้ไพ่คิงดอกจิก เช่นนี้ เหตุการณ์ E และ F จึงเกิดร่วมกันได้ คือ ไม่ mutually exclusive

เหตุการณ์อีกลักษณะนั้นเรียกว่า *collectively exhaustive* คือ ลักษณะของเหตุการณ์ที่ประกอบกันแล้วครอบคลุมสิ่งที่อาจจะเกิดขึ้นทั้งหมดจากการทดลอง (Sample Space) เช่น เหตุการณ์ A คือการโยนเหรียญแล้วได้หัวทั้งสองครั้ง เหตุการณ์ B คือการโยนเหรียญแล้วได้ก้อยทั้งสองครั้ง และเหตุการณ์ C คือการโยนเหรียญแล้วได้หัวหนึ่งครั้งก้อยหนึ่งครั้ง เหตุการณ์ทั้ง A, B และ C นั้นเป็น collectively exhaustive เนื่องจากทั้งสามเหตุการณ์ครอบคลุมสิ่งที่อาจจะเกิดขึ้นทั้งหมดจากการทดลองสุ่ม (Sample Space) ทั้งหมด

2.3 ความน่าจะเป็น (Probability) และตัวแปรสุ่ม (Random Variable)

ความน่าจะเป็น (Probability) คือโอกาสที่เหตุการณ์หนึ่งๆจะเกิดขึ้น กำหนดให้ A เป็นเหตุการณ์ที่เกิดขึ้นใน Sample Space และ $P(A)$ เป็นโอกาสที่เหตุการณ์ A จะเกิดขึ้น จะได้ว่า

$$P(A) = \frac{\text{จำนวนเหตุการณ์ A}}{\text{จำนวนเหตุการณ์ที่อาจจะเกิดขึ้นทั้งหมด (sample space)}} \tag{2.3}$$

ตัวอย่างเช่น กำหนดให้เหตุการณ์ A คือเหตุการณ์ที่จั่วไพ่หนึ่งใบให้ได้เลข 2 ความน่าจะเป็นของเหตุการณ์ A คำนวณได้จากการนับจำนวนเลข 2 ที่มีในสำรับ นั่นคือ 4 (จาก 2 ดอกจิก 2 โพดำ 2 โพแดง และ 2 ข้าวหลามตัด) และจำนวนผลลัพธ์ที่อาจจะเกิดขึ้นทั้งหมดจากการจั่วไพ่ คือ 52 (จากทั้งหมด 52 ใบในไพ่ 1 สำรับ) ดังนั้นจึงกล่าวได้ว่า ความน่าจะเป็นของเหตุการณ์ A เท่ากับ $\frac{4}{52}$ หรือ $\frac{1}{13}$

โดยคุณสมบัติของความน่าจะเป็น (Probability) มีดังต่อไปนี้:

1. $0 \leq P(A) \leq 1$
2. ถ้า A, B และ C ประกอบกันเป็น exhaustive set แสดงว่า
$$P(A) + P(B) + P(C) = 1$$
3. ถ้า A, B และ C เป็นเหตุการณ์ mutually exclusive แสดงว่า
$$P(A + B + C) = P(A) + P(B) + P(C)$$

คำถามชวนคิด

ปัญหา Let's make a deal หรือ Monty Hall Problem

รายการโทรทัศน์รายการหนึ่งในประเทศสหรัฐอเมริกาได้ให้ผู้เข้าแข่งขันเลือกเปิดป้ายใดป้ายหนึ่งจาก 3 ป้าย โดยที่หนึ่งในแผ่นป้ายเหล่านั้นจะทำให้ผู้เข้าแข่งขันได้รับรถยนต์เป็นรางวัล แต่อีกสองแผ่นป้ายนั้นจะมีแพะ หมายความว่าผู้เข้าแข่งขันจะกลับบ้านมือเปล่า

เมื่อผู้เข้าแข่งขันรายการได้เลือกแผ่นป้ายแล้ว พิธีกรที่รู้ว่ารถยนต์อยู่ที่แผ่นป้ายใด จะเลือกเปิดแผ่นป้ายอีกใบหนึ่งที่ไม่มีรางวัลให้ และจะถามผู้เข้าแข่งขันว่าต้องการเปลี่ยนแผ่นป้ายหรือไม่

คำถามคือการเปลี่ยนแผ่นป้ายจะเพิ่มโอกาสในการได้รับรางวัลหรือไม่?

ตัวแปรสุ่ม (Random Variable) คือ ตัวแปรที่ใช้แทนผลที่เกิดขึ้นจากการทดลอง ประกอบด้วยตัวเลขที่มีค่าจำกัดแบบไม่ต่อเนื่อง (Discrete value) และมีค่าต่อเนื่อง (Continuous value)

สำหรับตัวอย่างตัวแปรสุ่มที่ให้ผลเป็นตัวเลขที่มีค่าจำกัดแบบไม่ต่อเนื่อง คือความน่าจะเป็นจากการได้ผลรวมของตัวเลขที่ได้จากการทอยลูกเต๋า 2 ลูก กล่าวคือ ผลรวมของตัวเลขจากการทอยลูกเต๋า 2 ลูกจะอยู่ระหว่าง 2 ถึง 12 จะไม่มีโอกาสได้ผลลัพธ์เป็น 2.5 หรือ 3.5 อย่างแน่นอน

แต่หากเป็นกรณีตัวแปรสุ่มที่มีค่าต่อเนื่องนั้น พิจารณาความสูงของจำนวนนักเรียนชั้นมัธยมระหว่าง 160 ถึง 180 เซนติเมตร จะเห็นได้ว่าค่าของความสูงเป็นเซนติเมตรดังกล่าวจะไม่จำกัดเพียงเฉพาะจำนวนเต็มเท่านั้น แต่อาจรวมไปถึงความสูงที่ระดับ 160.50 หรือ 160.52 เซนติเมตร

ลักษณะของตัวแปรสุ่มทั้งสองแบบที่ได้กล่าวมาแล้วข้างต้นมีความแตกต่างกันออกไปย่อมส่งผลให้เราสามารถแจกแจงความน่าจะเป็นออกเป็นฟังก์ชันของตัวแปรสุ่มได้ในรูปแบบที่แตกต่างกันออกดังไปตามที่ปรากฏในหัวข้อที่ 2.4 รูปแบบการแจกแจงหลายตัวแปร หรือ Probability Density Function

2.4 รูปแบบการแจกแจงหลายตัวแปร (Probability Density Function)

เนื่องจากตัวแปรสุ่มเกิดจากการทดลอง การแจกแจงค่าความน่าจะเป็นในแต่ละเหตุการณ์จะช่วยให้พิจารณาภาพรวมของเหตุการณ์สุ่มได้ชัดเจนขึ้น โดยสามารถแยกรูปแบบการแจกแจงดังกล่าวให้ขึ้นอยู่กับลักษณะของตัวแปรสุ่มนั้น เช่น การแจกแจงตัวแปรสุ่มของข้อมูลไม่ต่อเนื่อง ข้อมูลต่อเนื่อง และอื่นๆ ดังนี้

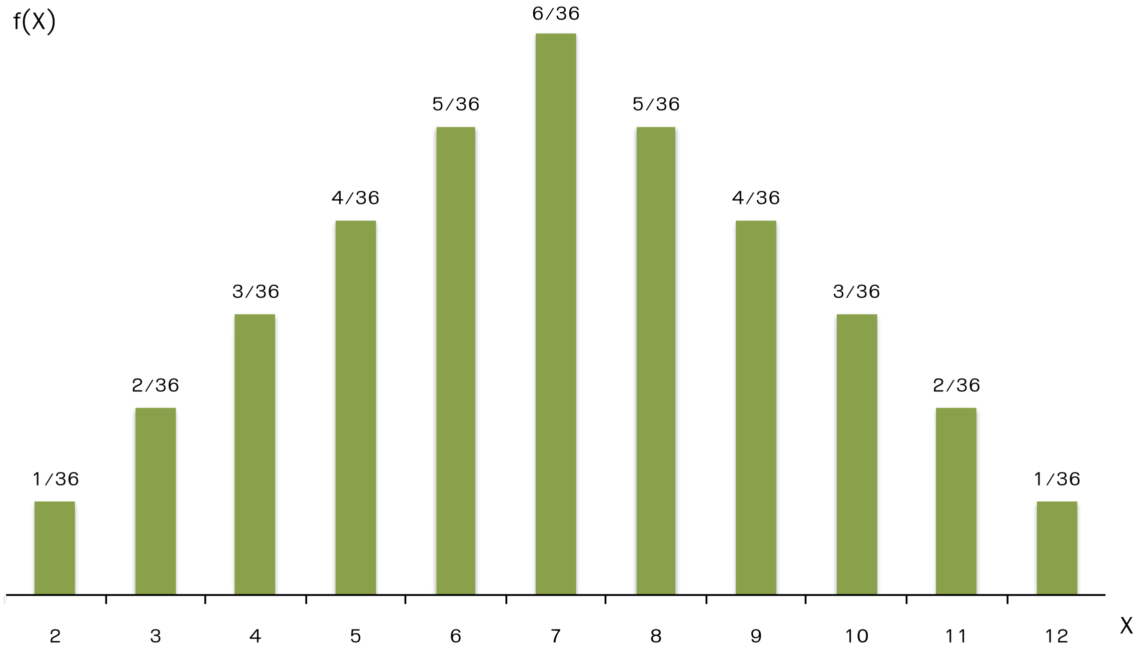
2.4.1 รูปแบบการแจกแจงความน่าจะเป็นของตัวแปรสุ่มที่มีค่าไม่ต่อเนื่อง (Probability Density Function for Discrete Random Variable)

กำหนดให้ X เป็นตัวแปรสุ่มที่มีค่าไม่ต่อเนื่อง (Discrete Random Variable) ซึ่งมีค่า X_1 X_2 จนถึง X_n จะได้ว่า

$$f(X) = \begin{cases} P(X = X_i) & i = 1, 2, ..., n \\ 0 & X \neq X_i \end{cases}$$

ตัวอย่างที่ 2-1 กำหนดให้ตัวแปรสุ่ม X คือผลรวมของตัวเลขที่ปรากฏขึ้นจากการโยนลูกเต๋าสองครั้ง ซึ่งผลรวมที่เป็นไปได้อาจจะเป็น 2 (จากการที่ลูกเต๋าทั้งสองลูกออกแต้ม 1) ถึง 12 (จากการที่ลูกเต๋าทั้งสองลูกออกแต้ม 6) แสดงรายละเอียดได้ตามภาพที่ 2-1 ซึ่งระบุถึงความน่าจะเป็นของแต่ละเหตุการณ์ เช่น การที่ผลรวมของหน้าลูกเต๋าจะมีค่าเท่ากับ 3 ก็ต่อเมื่อลูกเต๋าออกหน้า (1,2) หรือ (2,1) นับเป็น 2 เหตุการณ์จากเหตุการณ์ที่เป็นไปได้ทั้งหมด 36 เหตุการณ์ ความน่าจะเป็นที่ผลรวมจะเท่ากับ 3 คือ $\frac{2}{36}$#

ภาพที่ 2-1: แผนภูมิแจกแจงความน่าจะเป็นของผลรวมของตัวเลขที่ปรากฏขึ้นจากการโยนลูกเต๋าสองครั้ง



2.4.2 รูปแบบการแจกแจงความน่าจะเป็นของตัวแปรสุ่มที่มีค่าต่อเนื่อง (Probability Density Function for Continuous Random Variable)

กำหนดให้ X เป็นตัวแปรสุ่มที่มีค่าต่อเนื่อง (Continuous Random Variable) จะได้ว่า $f(X)$ คือรูปแบบการแจกแจงความน่าจะเป็นของตัวแปรสุ่มที่มีค่าต่อเนื่อง ก็ต่อเมื่อเงื่อนไขทั้งสามด้านล่างนี้สมบูรณ์

1.โอกาสที่ผลลัพธ์ของตัวแปรสุ่มจะเกิดค่าต่อเนื่องใด ๆที่มีค่ามากกว่าหรือเท่ากับ 0

$$f(X) \geq 0$$

2.ผลรวมของโอกาสที่เหตุการณ์ของตัวแปรสุ่มที่มีค่าต่อเนื่องจะเกิดขึ้นจะต้องมีค่าเท่ากับ 1

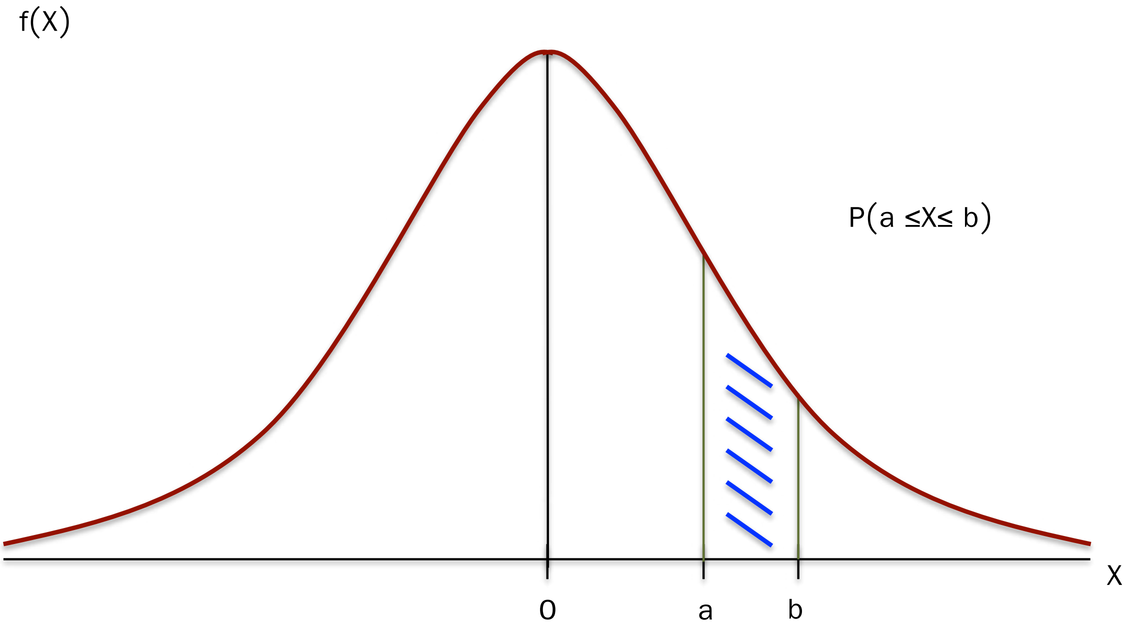
$$\int_{-\infty}^{\infty} f(X)dX = 1$$

3.ภายใต้ช่วงตัวแปรสุ่มที่สนใจ (ระหว่าง a กับ b) พื้นที่ใต้กราฟจะแสดงความน่าจะเป็นที่ตัวแปรสุ่มที่มีค่าต่อเนื่องนั้นอาจเกิดขึ้น

$$\int_{a}^{b} f(X)dX = P(a \leq X \leq b)$$

ภาพที่ 2-2 แสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่มที่มีค่าต่อเนื่อง พื้นที่ใต้เส้นแสดงถึงความน่าจะเป็นในแต่ละช่วงของค่าตัวแปรสุ่ม โดยที่ $P(a \leq X \leq b)$ หมายถึงความน่าจะเป็นที่จะเกิดขึ้นเมื่อตัวแปรสุ่ม X มีค่าอยู่ระหว่างช่วงใดๆ (ในกรณีนี้คือ a และ b)

ภาพที่ 2-2: แผนภูมิแจกแจงความน่าจะเป็นของตัวแปรสุ่มที่มีค่าต่อเนื่อง



ตัวอย่างที่ 2-2 กำหนดให้ X เป็นตัวแปรสุ่มแบบต่อเนื่อง โดยมีฟังก์ชันความน่าจะเป็นดังนี้

$$f(X) = -6X + 4, \text{ โดยที่ } 0 \leq X \leq 1$$

จงตอบคำถามต่อไปนี้

1.จงหาพื้นที่ใต้กราฟ f(X) ในช่วง $0 \leq X \leq 1$
2.จงหา $f(X = 0.3)$
3.จงหา $P(X \geq 0.5)$
4.จงหา $P(0.25 \leq X \leq 0.50)$

ตอบ

พิจารณาคุณสมบัติของรูปแบบการแจกแจงความน่าจะเป็นของตัวแปรสุ่มที่มีค่าต่อเนื่องจะพบว่า $f(X) \geq 0$ ก็ต่อเมื่อ $0 \leq X \leq \frac{2}{3}$

1.$F(X) = \int_0^{\frac{2}{3}}(-6X+4)dX + \int_{\frac{2}{3}}^{1}(-6X+4)dX = \frac{4}{3} + 1 - \frac{4}{3} = 1$

2.$F(X = 0.3) = \int_{0.3}^{0.3}(-6X+4)dX = 0$

3.$P(X \geq 0.5) = \int_{\frac{1}{2}}^{\frac{2}{3}} (-6X+4)dX = \frac{1}{12}$

4.$P(0.25 \leq X \leq 0.50) = \int_{\frac{1}{4}}^{\frac{1}{2}} (-6X+4)dX = \frac{7}{16}$#

2.4.3 รูปแบบการแจกแจงความน่าจะเป็นร่วม (Joint Probability Density Function)

รูปแบบการแจกแจงลำดับถัดมาคือ รูปแบบการแจกแจงความน่าจะเป็นร่วม (Joint Probability Density Function) โดยในที่นี้ จะขอยกตัวอย่างถึงรูปแบบการแจกแจงลักษณะดังกล่าวของข้อมูลที่มีค่าไม่ต่อเนื่อง

กำหนดให้ X และ Y เป็นตัวแปรสุ่มที่มีค่าไม่ต่อเนื่อง ฟังก์ชันความน่าจะเป็นร่วมจะระบุความน่าจะเป็นสำหรับเหตุการณ์ที่ตัวแปรสุ่ม X และ Y เกิดขึ้นพร้อมกัน แสดงได้ดังต่อไปนี้

$$f(X,Y) = P(X = x \text{ และ } Y = y)$$

ตัวอย่างที่ 2-3 ตารางต่อไปนี้แสดงถึงความน่าจะเป็นร่วมของตัวแปรสุ่ม X และ Y

ตารางที่ 2-1: ตัวอย่างความน่าจะเป็นร่วมของตัวแปรสุ่ม X และ Y

		X		
		-1	0	1
	1	0.11	0.08	0.05
Y	2	0.09	0.05	0.03
	3	0.35	0.07	0.17

จากตารางข้างต้นพบว่า ความน่าจะเป็นที่ตัวแปรสุ่ม X จะมีค่าเท่ากับ 0 พร้อมๆกับที่ตัวแปรสุ่ม Y มีค่าเท่ากับ 3 อยู่ที่ร้อยละ 7 หรือ 0.07 โดยแสดงเป็นสัญลักษณ์ทางคณิตศาสตร์ได้ว่า

$$f(X=0, Y=3) = 0.07\#$$

2.4.4 รูปแบบการแจกแจงความน่าจะเป็นตามขอบ (Marginal Probability Density Function)

การแสดงรูปแบบการแจกแจงความน่าจะเป็นร่วมข้างต้น $f(X,Y)$ นั้นเป็นการแสดงความน่าจะเป็นร่วมของตัวแปรสุ่มสองตัวแปร อย่างไรก็ดี การแจกแจงความน่าจะเป็นตามขอบ (Marginal Probability Density Function) จะแสดงความน่าจะเป็นเดี่ยวของตัวแปรสุ่มได้ เช่น $f(X)$ หรือ $f(Y)$ โดยสามารถแปลงค่าจากความน่าจะเป็นร่วมได้ด้วยวิธีการต่อไปนี้

สำหรับกรณีข้อมูลแบบไม่ต่อเนื่อง (Discrete Value)

$f(X) = \sum_Y f(X,Y)$ เรียกว่า marginal PDF ของ X

$f(Y) = \sum_X f(X,Y)$ เรียกว่า marginal PDF ของ Y

สำหรับกรณีข้อมูลแบบต่อเนื่อง (Continuous Value)

$f(X) = \int_a^b f(X,Y)dY$ เรียกว่า marginal PDF ของ X

$f(Y) = \int_c^d f(X,Y)dX$ เรียกว่า marginal PDF ของ Y

โดย

$\sum_Y$ หรือ $\sum_X$ หมายถึง การรวมค่าความน่าจะเป็นของตัวแปร Y หรือ X ตามลำดับ

a กับ b คือค่าคงที่ตัวแปรสุ่มต่อเนื่อง Y เป็นได้น้อยที่สุดและมากที่สุด ตามลำดับ และ

c กับ d คือค่าคงที่ตัวแปรสุ่มต่อเนื่อง X เป็นได้น้อยที่สุดและมากที่สุด ตามลำดับ

ตัวอย่างที่ 2–3 จากตารางที่ 2-1 ของตัวอย่างที่ 2-2 จงหา marginal PDF ของ X และ marginal PDF ของ Y

ตอบ

หา marginal PDF ของ X ได้ดังนี้

$$\begin{aligned}
f(X=-1) &= \textstyle\sum_Y f(X=-1,Y) \\
&= f(X=-1,Y=1)+f(X=-1,Y=2)+f(X=-1,Y=3) \\
&= 0.11+0.09+0.35 \\
&= 0.55 \\
f(X=0) &= \textstyle\sum_Y f(X=0,Y) \\
&= f(X=0,Y=1)+f(X=0,Y=2)+f(X=0,Y=3) \\
&= 0.08+0.05+0.07 \\
&= 0.20 \\
f(X=1) &= \textstyle\sum_Y f(X=1,Y) \\
&= f(X=1,Y=1)+f(X=1,Y=2)+f(X=1,Y=3) \\
&= 0.05+0.03+0.17 \\
&= 0.25
\end{aligned}$$

และ marginal PDF ของ Y ได้ดังนี้

$$\begin{aligned}
f(Y=1) &= \textstyle\sum_X f(X,Y=1) \\
&= f(X=-1,Y=1)+f(X=0,Y=1)+f(X=1,Y=1) \\
&= 0.11+0.08+0.05 \\
&= 0.24 \\
f(Y=2) &= \textstyle\sum_X f(X,Y=2) \\
&= f(X=-1,Y=2)+f(X=0,Y=2)+f(X=1,Y=2) \\
&= 0.09+0.05+0.03 \\
&= 0.17 \\
f(Y=3) &= \textstyle\sum_X f(X,Y=3) \\
&= f(X=-1,Y=3)+f(X=0,Y=3)+f(X=1,Y=3) \\
&= 0.35+0.07+0.17 \\
&= 0.59
\end{aligned}$$

จากการคำนวณข้างต้น สามารถสรุปได้ดังตารางที่ 2-2 แสดง marginal probability #

ตารางที่ 2-2: ตารางแสดงตัวอย่างความน่าจะเป็นร่วมของตัวแปรสุ่ม X และ Y

			X		
		-1	0	1	
	1	0.11	0.08	0.05	$f(Y=1)$ = 0.24
Y	2	0.09	0.05	0.03	$f(Y=2)$ = 0.17
	3	0.35	0.07	0.17	$f(Y=3)$ = 0.59
		$f(X=-1)$ = 0.55	$f(X=0)$ = 0.20	$f(X=1)$ = 0.25	$f(X)=1$ $f(Y)=1$

2.4.5 รูปแบบการแจกแจงความน่าจะเป็นแบบมีเงื่อนไข (Conditional PDF)

ความน่าจะเป็นแบบมีเงื่อนไข คือการหาความน่าจะเป็นของเหตุการณ์หนึ่งๆโดยมีเงื่อนไขว่าอีกเหตุการณ์หนึ่งเกิดขึ้นแล้ว โดยมีฟังก์ชันดังต่อไปนี้

$$f(X|Y) = P(X = x|Y = y)$$

โดยสามารถคำนวณผ่านความน่าจะเป็นร่วม (Joint PDF) ได้ดังนี้

$$f(X|Y) = \frac{f(X,Y)}{f(Y)}$$

ตัวอย่างที่ 2-4 จากข้อมูลในตารางที่ 2-1 จงหา $f(X=0|Y=1)$ และ $f(Y=2|X=0)$

<u>ตอบ</u>

$$\begin{aligned} f(X=0|Y=1) &= \tfrac{f(X=0,Y=1)}{f(Y=1)} \\ &= \tfrac{0.08}{0.24} \\ &= 0.33 \end{aligned}$$

$$\begin{aligned} f(Y=2|X=0) &= \tfrac{f(Y=2,X=0)}{f(X=0)} \\ &= \tfrac{0.05}{0.20} \\ &= 0.25\# \end{aligned}$$