

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

หลักการและแนวคิดของการทำงานวิจัยเรื่องการประยุกต์ใช้การวิเคราะห์การรอดชีพในการศึกษาอัตราการออกกลางคันของนักศึกษามหาวิทยาลัยราชภัฏเชียงใหม่ที่ได้จากการศึกษาค้นคว้าเอกสาร และงานวิจัยที่ เกี่ยวข้อง แบ่งออกเป็น 3 ตอนได้แก่ ตอนแรก การออกกลางคัน ตอนที่ 2 หลักการ และ แนวคิดของ การวิเคราะห์การอยู่รอด ตอนที่ 3 การออกกลางคันและการวิเคราะห์การอยู่รอด ในตอนที่ 3 นี้ผู้วิจัยนำเสนอการเปรียบเทียบการวิเคราะห์การออกกลางคันโดยเทคนิคการวิเคราะห์การอยู่รอดกับการวิเคราะห์การออกกลางคันในรูปแบบอื่น ๆ ในอดีต รายละเอียดเนื้อหาแต่ละตอนมีดังต่อไปนี้

ตอนที่ 1 การออกกลางคัน (Dropout)

จากระเบียบ ข้อบังคับ มหาวิทยาลัยราชภัฏเชียงใหม่ว่าด้วยการศึกษาในระดับปริญญาตรี (2557) หมวดที่ 8 ข้อ 41 การพ้นสถานภาพนักศึกษา เป็นไปตามกรณีต่อไปนี้

- (1) สำเร็จการศึกษาตามหลักสูตรและได้รับอนุมัติให้ปริญญา
- (2) ผลการประเมินได้ระดับคะแนนสะสมเฉลี่ยต่ำกว่า 1.50 เมื่อสิ้นภาคการศึกษาปกติที่ 2 นับตั้งแต่เริ่มเข้าเรียน โดยนับจำนวนภาคการศึกษารวมทั้งภาคการศึกษาปกติที่มีการลาพักการศึกษาด้วย
- (3) ผลการประเมินได้ระดับคะแนนสะสมเฉลี่ยต่ำกว่า 1.70 เมื่อสิ้นภาคการศึกษาปกติที่ 4 นับตั้งแต่เริ่มเข้าเรียน โดยนับจำนวนภาคการศึกษารวมทั้งภาคการศึกษาปกติที่มีการลาพักการศึกษาด้วย
- (4) ผลการประเมินได้ระดับคะแนนสะสมเฉลี่ยต่ำกว่า 1.80 เมื่อสิ้นภาคการศึกษาปกติที่ 6 ที่ 7 ที่ 10 ที่ 12 หรือที่ 14 และเมื่อสิ้นภาคการศึกษาปกติที่ 16 หรือที่ 18 สำหรับหลักสูตรปริญญาตรี 5 ปีโดยนับจำนวนภาคการศึกษารวมทั้งภาคการศึกษาปกติที่มีการลาพักการศึกษาด้วย
- (5) นักศึกษาลงทะเบียนเรียนครบตามหลักสูตร แต่ได้ระดับค่าระดับคะแนนสะสมเฉลี่ยต่ำกว่า 1.80
- (6) ใช้เวลาศึกษาเกินกวาระยะเวลาการศึกษาที่กำหนดไว้ตามเกณฑ์มาตรฐานหลักสูตรระดับปริญญาตรี

(7) มหาวิทยาลัยมีคำสั่งให้พ้นสภาพการเป็นนักศึกษา หรือเป็นไปตามระเบียบและข้อบังคับของมหาวิทยาลัยที่กำหนดไว้

(8) ลาออก

(9) ตาย

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการออกกลางคันของนักเรียน นิสิตนักศึกษา พบว่างานวิจัยแต่ละครั้งมีวัตถุประสงค์ของการวิจัยหลายประเด็น โดยมีรายละเอียดดังนี้

วัชรินทร์ แสงมา และคณะ (2551) ได้ศึกษา อัตราเสี่ยงของการออกกลางคัน และปัจจัยที่มีอิทธิพลต่ออัตราเสี่ยงของการออกกลางคันของนักศึกษาระดับปวสมหาวิทยาลัยราชภัฏวชิรเวศน์นครราชสีมา ผลการวิจัยพบว่า ปัจจัยที่มีอิทธิพลต่ออัตราเสี่ยงการออกกลางคันคือสถานที่ตั้งของสถานศึกษาเดิม ภูมิภาคของสถานศึกษาเดิม เกรดเฉลี่ยที่จบ ปวช. โดยนักศึกษาที่จบมาจากสถานศึกษาในภาคกลาง มีอัตราเสี่ยงออกกลางคั้นน้อยกว่าของนักศึกษาที่จบมาจากสถานศึกษาในกรุงเทพ และปริมณฑล หรือมีอัตราเสี่ยงน้อยกว่าอยู่ร้อยละ 43.8 นักศึกษาที่จบ ปวช. ที่ได้เกรดเฉลี่ยในช่วง 3.10 ขึ้นไป มีอัตราการเสี่ยงในการออกกลางคั้นน้อยกว่าอยู่ร้อยละ 43.1เมื่อเทียบกับกลุ่มที่ได้เกรดเฉลี่ยในช่วง 2.00-2.50

พงศ์พันธุ์ คำพรรณ และ จีรเกียรติ อภิบุญโยภาส (2562) ศึกษาโมเดลเชิงสาเหตุของการออกกลางคันของนักศึกษามหาวิทยาลัยวลัยลักษณ์ เก็บข้อมูลโดยใช้แบบสอบถามเก็บข้อมูลจากกลุ่มตัวอย่างจำนวน 380 ตัวอย่างวิเคราะห์ข้อมูลโดยใช้โมเดลเชิงสาเหตุด้วยโปรแกรมคอมพิวเตอร์ผลการศึกษาจากโมเดลเชิงสาเหตุ พบว่า ตัวแปรแฝงด้านสถานะทางเศรษฐกิจและสังคมการบูรณาการทางวิชาการการบูรณาการทางสังคม ความผูกพันกับเป้าหมายประสงค์และความผูกพันกับสถาบันมีอิทธิพลทางตรงต่อในทิศทางตรงข้ามกับการออกกลางคันของนักศึกษามหาวิทยาลัยวลัยลักษณ์อย่างมีนัยสำคัญทางสถิติที่ .05 ข้อเสนอแนะเพื่อแก้ปัญหาการออกกลางคันของนักศึกษา ได้แก่ (1) พัฒนาสภาพแวดล้อมและสิ่งอำนวยความสะดวกในมหาวิทยาลัย (2) สนับสนุนทุนการศึกษาและการสร้างรายได้เสริมให้กับนักศึกษา (3) การรับนักศึกษาต้องเน้นที่คุณภาพและมีการสอนเสริมในกลุ่มที่มีความต้องการเป็นพิเศษ (4) สร้างแรงจูงใจและสร้างเป้าหมายในชีวิตให้นักศึกษา (5) มีการฝึกอบรมอาจารย์ให้เป็นที่ปรึกษามืออาชีพ

สุกัญญา ทารส และ ทรงศักดิ์ ภูสีอ่อน (2562) ศึกษาปัจจัยที่มีความสัมพันธ์กับการออกกลางคัน ของนิสิตมหาวิทยาลัยมหาสารคาม ระดับปริญญาตรี ชั้นปีที่ 1 โดยกลุ่มตัวอย่างที่ใช้ในการวิจัยคือ นิสิตมหาวิทยาลัยมหาสารคาม ระดับปริญญาตรี ชั้นปีที่ 1 ที่เข้าศึกษาอย่างน้อย 1 ภาคการศึกษา ในปีการศึกษา 2558 จำนวน 239 คน สถิติที่ใช้ในการวิเคราะห์ข้อมูลได้แก่การวิเคราะห์จำแนกประเภท(discriminant analysis) แบบขั้นตอน (stepwise method) ผลการวิจัยปรากฏพบว่า ปัจจัยที่สามารถจำแนกการออกกลางคันของนิสิตมหาวิทยาลัยมหาสารคาม ระดับปริญญาตรี ชั้นปีที่ 1 มี 2 ตัวแปร ได้แก่ ปัญหาด้านนิสิต และปัญหาด้านครอบครัวนิสิต ซึ่งตัวแปรในสมการจำแนกประเภททั้ง 2 ตัวแปร มีส่วนในการจำแนก กลุ่มนิสิตที่กำลังศึกษาอยู่ และกลุ่มนิสิตที่ออกกลางคัน ได้อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และร่วมกันทำนายการเป็นสมาชิกของกลุ่มนิสิตที่กำลังศึกษาอยู่ และกลุ่มนิสิตที่ออกกลางคันโดยทำนายการออกกลางคันถูกต้องร้อยละ 76.60

แคทรียา ทาเวอร์มย์ (2543) ได้ประยุกต์ใช้การวิเคราะห์ความอยู่รอด ในการศึกษาฟังก์ชันการอยู่รอดมัธยฐานระยะเวลาการอยู่รอดอัตราเสี่ยงอันตรายต่อการออกกลางคันสูงสุดและเพื่อศึกษาฟังก์ชันการอยู่รอดและโมเดลฟังก์ชันความเสี่ยงอันตรายของนิสิตที่มีคุณลักษณะแตกต่างกัน ตามตัวแปรทำนายได้แก่เพศอายุประเภทการเรียน เกรดเฉลี่ยสะสมระดับมปลาย จังหวัดและภาควิชา ของนิสิตระดับ ปริญญาบัณฑิตคณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยจำนวนสามรุ่น พบว่าตัวทำนายที่มีผลต่อระยะเวลาการอยู่รอดอย่างมีนัยสำคัญทางสถิติ ได้แก่อายุเกรดเฉลี่ยสะสมระดับมัธยมปลายจังหวัดและภาควิชา

ลำเพย สนิธิ (2539) ได้ศึกษาสาเหตุของปัญหา วิธีการแก้ไขและแนวทางการปฏิบัติของการออกกลางคันในด้านตัวนักเรียน หลักสูตรและการเรียนการสอน บริการและสภาพแวดล้อมของโรงเรียน ครอบครัว และเศรษฐกิจและสังคมของนักเรียนโครงการขยายโอกาสทางการศึกษาขั้นพื้นฐานสังกัด สำ นักงานคณะกรรมการการประถมศึกษาแห่งชาติ ภาคกลาง พบว่า สาเหตุของปัญหาการออกกลางคัน ด้านตัวนักเรียน คือ ผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์และภาษาอังกฤษต่ำ ขาดเรียนบ่อย มีปัญหากับครูบางครั้ง ถูกครูลงโทษเพราะไม่ค่อยทำ การบ้าน มีปัญหากับเพื่อนบ้าง เจ็บป่วยค่อนข้างบ่อย ด้านหลักสูตรและการเรียนการสอน คือ ใช้เวลามากในการทำ ความเข้าใจบทเรียนไม่ค่อยสนใจเข้ารับการสอนซ่อมเสริม ด้านบริการและสภาพแวดล้อมของโรงเรียน คือ การติดต่อบริการระหว่างบ้านกับโรงเรียนไม่สม่ำเสมอ ด้านครอบครัว คือ ขาดการเอาใจใส่และดูแลจาก

ผู้ปกครองเนื่องจากผู้ปกครองต้องไปประกอบอาชีพ ด้านเศรษฐกิจและสังคม คือ รายได้ภายในครอบครัวไม่เพียงพอ และคนในชุมชนขาดความกระตือรือร้นในการศึกษาหาความรู้

พูนศักดิ์ อดิชาติ (2536) ได้ศึกษาสถานภาพทั่วไปของนักศึกษาที่ออกกลางคัน สาเหตุของการออกกลางคัน และวิเคราะห์ระหว่างสถานภาพของนักศึกษาที่ออกกลางคันกับสาเหตุของการออกกลางคันของนักศึกษาหลักสูตรวิชาชีพพระยาศน์ ในวิทยาลัยสารพัดช่างสมุทรปราการ พบว่า ผลการศึกษาทั้ง 6 ด้าน ได้แก่ ด้านภูมิหลัง ด้านหลักสูตร ด้านครูและการสอนของครู ด้านการบริการ และสภาพแวดล้อม ด้านวัสดุ อุปกรณ์และด้านการประเมินผลนั้น สาเหตุของการออกกลางคันที่สำคัญที่สุดคือ ด้านภูมิหลังของนักศึกษา โดยเฉพาะการไม่มีเวลาเรียนเพราะต้องประกอบอาชีพ

กัลยา อนุพัฒน์ (2533) ได้ศึกษาเปรียบเทียบอัตราการรายงานตัว อัตราการออกกลางคัน ภูมิหลังทางการศึกษา เศรษฐกิจและสังคม ผลสัมฤทธิ์ทางการเรียนและเจตคติต่อมหาวิทยาลัยสงขลานครินทร์ของนักศึกษาที่ผ่านการสอบคัดเลือกโดยวิธีรับตรงและวิธีรับรวม มหาวิทยาลัยสงขลานครินทร์ รุ่นปีการศึกษา 2527-2531 ผลการวิจัย พบว่า นักศึกษารับตรงมีอัตราการออกกลางคันต่ำกว่านักศึกษารับรวมอย่างมีนัยสำคัญทางสถิติ นักศึกษารับตรงมีผลการเรียนชั้นมัธยมศึกษาตอนปลายสูงกว่านักศึกษารับรวมในเกือบทุกคณะ ยกเว้นคณะวิทยาศาสตร์และเทคโนโลยีอาชีพของบิตามารดาของนักศึกษาแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ระดับการศึกษาของบิตามารดาของนักศึกษาทั้งสองกลุ่มส่วนใหญ่สำเร็จการศึกษาชั้นประถมศึกษา สำหรับผลสัมฤทธิ์ทางการเรียนในมหาวิทยาลัยของนักศึกษารับตรงสูงกว่านักศึกษารับรวมในเกือบทุกชั้นปี คณะและรุ่นปีการศึกษาก่อนหน้านั้นปี 1 คณะศึกษาศาสตร์ รุ่นปีการศึกษา 2529 และนักศึกษารับตรงมีเจตคติต่อมหาวิทยาลัยสงขลานครินทร์ในระดับที่สูงกว่านักศึกษารับรวม

Han and Gages (1995) ได้ศึกษาระยะเวลาของการเสี่ยงที่จะเกิดการออกกลางคันของนักศึกษาที่รับเข้าพิเศษและปีที่นักศึกษาออกไปโดยส่วนใหญ่ และความเสี่ยงของการออกจากมหาวิทยาลัยเมื่อนักศึกษามีคุณลักษณะส่วนตัวและคุณลักษณะพื้นฐานการศึกษาที่แตกต่างกันโดยใช้การวิเคราะห์ความอยู่รอด ซึ่งกลุ่มตัวอย่างประชากรเป็นนักศึกษาที่รับเข้ามาในปี 1986 -1989 มหาวิทยาลัย Northern Illinois จำนวน 1,639 คนข้อมูลที่ใช้วิเคราะห์ได้แก่เพศเชื้อชาติคะแนนสอบย่อยและคะแนนรวม อันดับของโรงเรียนมัธยมปลายจำนวนหน่วยกิตสะสม สถานะทางการศึกษาในแต่ละเทอมและวันที่อนุมัติให้สำเร็จการศึกษา พบว่าตัวแปรที่มีผลต่อความเสี่ยงของการออกกลางคัน

ของสามารคือเพศเชื้อชาติคะแนนสอบและอันดับที่ของโรงเรียนมัธยมปลาย ช่วงระยะเวลาที่เกิดความเสี่ยงของการออกสูงสุดคือภาคการศึกษาที่สองสามและสี่ ทางเพศชายและหญิง กลุ่มเชื้อชาติทั้งสี่กรุง เน้นทั้งกลุ่มที่มีความแตกต่างของคะแนนสอบเข้ามหาวิทยาลัยหรืออันดับที่โรงเรียนมปลาย นอกจากนั้นช่วงเวลาหลังจากมัธยมต้นระยะเวลาความอยู่รอดนั้นลักษณะการกระจายของความอยู่รอดคงที่แต่จะมีแนวโน้มลดลงเล็กน้อยประมาณสี่ภาคการศึกษาแล้วอัตราเสี่ยงจะเพิ่มขึ้นอีกครั้งในภาคการศึกษาต่อมา

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการออกกลางคันของนักเรียน นิสิตนักศึกษาที่กล่าวมาทั้งหมด พบว่า เป็นการศึกษาข้อมูลพื้นฐานหรือสภาพทั่วไปของการออกกลางคัน ศึกษาสาเหตุหรือเหตุผลของการออกกลางคัน ศึกษาความสัมพันธ์ระหว่างสภาพกับสาเหตุหรือเหตุผลของการออกกลางคัน ศึกษาองค์ประกอบที่สัมพันธ์กับการออกกลางคันและ ศึกษาในลักษณะของการเปรียบเทียบ ซึ่งงานวิจัยเรื่องการออกกลางคันนั้น นอกจากจะศึกษาในประเด็นที่กล่าวมาแล้วพบว่า งานวิจัยในต่างประเทศได้ศึกษาในประเด็นอื่นอีก คือ การพยากรณ์การออกกลางคันและการประมาณค่าความน่าจะเป็นของนักเรียนที่จะออกกลางคัน (Willett & Singer, 1991) ซึ่งผลจากการวิจัยศึกษาในประเด็นนี้ทำให้พยากรณ์ได้ว่าผู้เรียนจะออกกลางคันเมื่อไร (when) ได้ล่วงหน้า สำหรับการวิเคราะห์ในการศึกษาวิจัยประเด็นดังกล่าว คือ การวิเคราะห์การอยู่รอด(survival analysis)

ตอนที่ 2 การวิเคราะห์การอยู่รอด (Survival Analysis)

1) คำศัพท์และความหมายที่เกี่ยวข้องกับการวิเคราะห์ความอยู่รอด

1.1 ระยะเวลาเหตุการณ์ (Time to event หรือ Survival time)

คือ ระยะเวลาระหว่างเมื่อเริ่มต้นจนถึงเมื่อเกิดเหตุการณ์ คำที่ใช้บ่อยเกี่ยวกับระยะเวลาเหตุการณ์คือ Survival time ซึ่งหมายถึงตั้งแต่เริ่มต้นการสังเกตจนถึงตาย แต่อาจหมายถึงอย่างอื่นได้เช่น การไม่กลับมาเกิดอาการอีก (Relapse-free period) ระยะเวลาโรคสงบ (Remission duration) หรือ ช่วงเวลาที่อาการไม่กำเริบ (Progression-free survival)

1.2 เวลาเริ่มต้น (Begin Date)

เป็นจุดเริ่มต้นจริงของกระบวนการที่นำไปสู่การเกิดเหตุการณ์ที่สนใจศึกษา ตัวอย่างเช่น วันเริ่มต้น ในการรับวัคซีน คือวันเริ่มต้น ในการเข้าตรวจ หรือวันเริ่มต้นในการเข้าศึกษา เป็นต้น

1.3 เหตุการณ์ (Event)

คือสิ่งที่สนใจศึกษา ตัวอย่างเช่น การตาย การออกกลางคัน การกลับเข้าเรือนจำ หลังจากได้รับการปล่อยตัวไปแล้ว หรือการสิ้นสุดอายุการใช้งานของหลอดไฟ เป็นต้น

1.4 การวัดการเกิดเหตุการณ์

เมื่อเกิดเหตุการณ์ที่สนใจขึ้นจะเรียกว่า Failure โดยที่หากเหตุการณ์ยังไม่เกิดเมื่อสิ้นสุดระยะเวลาที่ศึกษา เรียกว่า Censored เช่น ขาดการติดต่อก่อนจะทราบว่าหายป่วยจากโรคที่กำลังศึกษาอยู่ หรือ เลิกกินยาต่อเนื่อง โดยที่เหตุการณ์ยังไม่เกิด หรือ เสียชีวิตด้วยสาเหตุอื่นก่อนเกิดเหตุการณ์ที่สนใจเป็นต้น โดยกำหนดตัวแปรคือ Event เป็นตัวแปรสุ่ม ดังนี้

Event = 1 เมื่อเกิดเหตุการณ์ที่สนใจ(Failure)

Event = 0 เมื่อยังไม่เกิดเหตุการณ์ที่สนใจเมื่อสิ้นสุดระยะเวลาศึกษา (Censored)

1.5 ประเภทของ Censoring

1.5.1 Left censoring

คือระยะปลอดเหตุการณ์ (Survival time) ไม่สมบูรณ์ทางด้านซ้ายของ ระยะเวลาในการติดตาม (Follow-up period) กล่าวคือไม่สามารถบอกได้ว่าเหตุการณ์เกิดเมื่อไร รู้เพียงว่าเกิดก่อนระยะเวลาที่เริ่มสังเกต เช่น ศึกษาระยะเวลาปลอดโรคเอดส์ในกลุ่มผู้ป่วยโรคเอดส์ เหตุการณ์ที่ศึกษาคือการติดเชื้อเอดส์ วันที่เริ่มสังเกตคือวันที่ได้รับการวินิจฉัยว่าป่วยคือแสดงอาการและผลตรวจเลือดเป็นบวก แต่ไม่ทราบวันที่ได้รับเชื้อ ข้อสังเกตคือระยะเวลาที่สังเกตจะยาวนานกว่าระยะปลอดเหตุการณ์

1.5.2 Right censoring

คือระยะเวลาปลอดเหตุการณ์ (Survival time) ไม่สมบูรณ์ทางด้านขวาของ Follow-up period เป็นประเภทของ Censoring ที่พบได้บ่อย ข้อสังเกตคือจะตรงข้ามกับ Left-censoring และระยะเวลาที่สังเกตจะสั้นกว่าระยะเวลาปลอดเหตุการณ์

1.5.3 Interval Censoring

ช่วงเวลาในการสังเกตไม่ได้เป็นระยะเวลาแบบต่อเนื่อง แต่แบ่งเป็นช่วง ๆ เช่นมาพบตรวจทุก ๆ 3 เดือน เป็นต้น ดังนั้นหากเหตุการณ์ที่ศึกษาเกิดขึ้น ณ เวลาใดเวลาหนึ่งของ ช่วงเวลาที่ดังกล่าว จะเรียกว่า interval censored เนื่องจากไม่ทราบเวลาที่แท้จริงของการเกิด เหตุการณ์ แต่ทราบเวลาครั้งล่าสุดที่มาสังเกตและยังไม่เกิดเหตุการณ์ที่สนใจนั้น

1.6 ฟังก์ชันที่เกี่ยวข้องกับการวิเคราะห์การอยู่รอด (Survival analysis)

1.6.1 ฟังก์ชันความอยู่รอด Survival function

เป็นพื้นฐานของ Survival analysis บอกความน่าจะเป็นที่คนคนหนึ่งยังคง ปลอดเหตุการณ์มากกว่าเวลา t โดยที่ปลอดเหตุการณ์มาโดยตลอดก่อนหน้านี้ เขียนแทนด้วย

$$S(t) = P(T > t)$$

โดย T = ระยะเวลาการอยู่รอด ($T \geq 0$)

เช่น $S(5) = P(T > 5)$ หมายถึงความน่าจะเป็นที่จะเกิดเหตุการณ์ที่สนใจ หลังจากผ่านไป 5 ปี เป็นต้น $S(t)$ เป็น ฟังก์ชันขั้นบันได (Step Functions) โดยหากว่า $t=0$ จะได้ว่า $S(t) = S(0) = 1$ เพราะยังไม่มีเหตุการณ์ที่สนใจศึกษาเกิดขึ้น และในทางตรงข้าม เมื่อ $t=\infty$ จะได้ว่า $S(t) = S(\infty) = 0$ เพราะไม่มีใครรอด หรือกล่าวได้ว่าความน่าจะเป็นของการอยู่รอด หรือการยังไม่เกิด เหตุการณ์ที่สนใจเมื่อระยะเวลาผ่านไปเป็นอนันต์ จะมีค่าเท่ากับ 0

1.6.2 ฟังก์ชันพิบัติ (Hazard function)

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t}$$

โดย $h(t)$ คืออัตราการเกิดเหตุการณ์ ณ จุดเวลา t โดยที่มีการปลอดเหตุการณ์จนกระทั่งถึงเวลานั้น อัตรานี้เรียกว่า Hazard rate มีค่า 0 ถึง ∞ และ $h(t)$ ให้ค่าเกี่ยวกับการเกิดเหตุการณ์ แต่ $S(t)$ ให้ค่าเกี่ยวกับการปลอดเหตุการณ์

1.6.3 ความสัมพันธ์ระหว่าง Survivor function กับ Hazard function

$h(t)$ เป็น Mathematical model สำหรับ Survival analysis และ $h(t)$ เป็นปฏิภาคผกผันกับ $S(t)$ ตัวหนึ่งเพิ่มอีกตัวจะลด และเมื่อรู้ตัวหนึ่งก็สามารถรู้อีกตัวได้ แสดงได้ดังนี้

$$H(t) = \int_0^t h(u) du = -\ln[S(t)] \text{ หรือ } S(t) = e^{-H(t)}$$

2) เป้าหมาย Survival analysis

เป้าหมายของ การวิเคราะห์ความอยู่รอด มี 3 ประการหลักคือ

- 2.1 เพื่อประมาณค่าและพรรณนาลักษณะ $h(t)$ หรือ $S(t)$ หรือทั้ง 2 ฟังก์ชัน
- 2.2 เพื่อเปรียบเทียบค่า $h(t)$ หรือ $S(t)$ หรือทั้ง 2 ฟังก์ชัน ระหว่างกลุ่มที่ศึกษา
- 2.3 เพื่อหาความสัมพันธ์ระหว่างปัจจัยต่างๆ กับระยะปลอดเหตุการณ์

3) วิธีการวิเคราะห์การอยู่รอด

3.1 วิธีของ Kaplan-Meier

ขั้นแรกจัดเรียงลำดับข้อมูลเวลาก่อน จากนั้นคำนวณค่าความน่าจะเป็น หรือโอกาสในการรอดชีพ ณ จุดเวลาที่ข้อมูลมี ค่าเหล่านี้สามารถนำเสนอเป็นกราฟต่อไป เป็นลำดับดังนี้สมมติการศึกษาหนึ่ง ศึกษาการรอดชีพผู้ป่วยมะเร็งปอดหลังรับการรักษา เหตุการณ์ที่ศึกษาคือ "เสียชีวิต (จากมะเร็งปอด)" เพื่อทราบระยะเวลาการรอดชีพ ได้ทำการศึกษาโดยมีระยะเวลาทั้งหมด 4 ปี จากต้นปี พ.ศ.2560 ถึงปลายปี พ.ศ. 2564 สรุปข้อมูลผู้ป่วย 6 รายได้ดังนี้

คนที่	ระยะปลอด เหตุการณ์ (เดือน)	ผลเมื่อวันสุดท้ายในการศึกษา (48เดือน)	เหตุการณ์
1	48	ยังมีชีวิตอยู่เมื่อสิ้นสุดระยะเวลาการศึกษา	Censored
2	22	เสียชีวิตเนื่องจากสาเหตุอื่นที่ไม่ใช่มะเร็ง ปอด	Censored
3	14	เสียชีวิตเนื่องจากสาเหตุมะเร็งปอด	Failure
4	40	เสียชีวิตเนื่องจากสาเหตุมะเร็งปอด	Failure
5	26	ยังมีชีวิตอยู่เมื่อสิ้นสุดระยะเวลาการศึกษา	Censored
6	13	ขาดการติดต่อระหว่างระยะเวลาที่ศึกษา โดยทราบว่าเมื่อตอนเริ่มการศึกษายังมีชีวิต อยู่	Censored

3.1.1 เตรียมข้อมูลสำหรับการคำนวณ

ข้อมูลเดิม			เรียงลำดับข้อมูลตามเวลา		
ID	TIME	DEAD	ID	TIME	DEAD
1	48	0	6	23	0
2	22	0	3	14	0
3	14	1	2	22	1
4	40	1	5	26	1
5	26	0	4	40	0
6	13	0	1	48	0

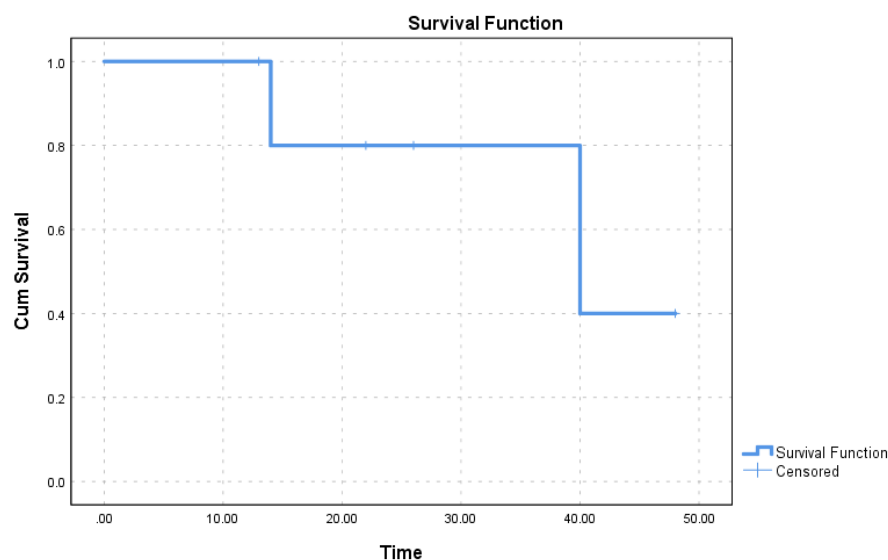
3.1.2 คำนวณ Survival Probability

Time (t_j)	Number at risk (n_j)	Dead (d_j)	Censored (q_j)	Survival Probability $S(t_j)$
13	6	0	1	$6/6 = 1.0$
14	$6-1=5$	1	0	$[(5-1)/5] \times 1.0 = 0.8$
22	$5-1=4$	0	1	$[(4-0)/4] \times 0.8 = 0.8$
26	$4-1=3$	0	1	$[(3-0)/3] \times 0.8 = 0.8$
40	$3-1=2$	1	0	$[(2-1)/2] \times 0.8 = 0.4$
48	$2-1=1$	0	1	$[(1-0)/1] \times 0.4 = 0.4$

เมื่อเวลา 13 เดือน ทั้ง 6 คนยังมีชีวิต ความน่าจะเป็นในการรอดชีพจึงเท่ากับ $6/6$ เท่ากับ 1 พอมาถึงเดือนที่ 14 เหลือคนที่อยู่ในการศึกษา 5 คนเพราะคนที่ ID = 6 ที่ขาดการติดตามเมื่อเดือนที่ 13 (ดูข้อมูลสรุปผู้ป่วย 6 ราย) และในเดือนที่ 14 นี้มี 1 คนที่เสียชีวิต จากทั้งหมด 5 คน ตาย 1 คน เหลือรอด 4 คน ดังนั้นความน่าจะเป็นของการรอดชีพจึงเท่ากับ $4/5$ เท่ากับ 0.8 แต่ด้วยโอกาสรอดชีพดังกล่าวมีเงื่อนไขว่าต้องรอดชีพมาก่อน จึงต้องคูณด้วยค่าความน่าจะเป็นของการรอดชีพเมื่อ 13 เดือน ซึ่งเท่ากับ 1 ได้เท่ากับ 0.8 นี่คือ Conditional probability ทำเช่นนี้เรื่อยไปจนครบจากตารางข้างต้น ทำให้เราทราบได้ว่า โอกาสรอดชีพของผู้ป่วยมะเร็งปอดเมื่อ 48 เดือนหลังการรักษาเท่ากับ 0.4 กล่าวอีกนัยหนึ่ง มีเพียงร้อยละ 40 ของผู้ป่วยที่ยังคงมีชีวิตอยู่ถึงปีที่ 4 หลังรักษา เราสามารถทราบค่าเหล่านี้ได้ทุกจุดเวลาที่มีข้อมูล เช่น 80% ของผู้ป่วยที่ได้รับการผ่าตัดมีโอกาสมีชีวิตอยู่ได้ถึงเดือนที่ 26 เป็นต้น

3.1.3 นำเสนอเป็น Kaplan-Meier survival curve

การสร้าง Kaplan-Meier Survival Curve ทำได้โดยนำค่าความน่าจะเป็นจากตาราง 2.1 นำเสนอเป็นกราฟที่มีลักษณะขั้นบันได (Survival function เป็น Step function) ดังภาพที่ 2.1



ภาพที่ 2.1 Kaplan-Meier Survival Curve

Kaplan-Meier Survival Probability ไม่มี Assumption เกี่ยวกับการแจกแจงข้อมูล จึงเป็น Nonparametric Assumption for Kaplan-Meier Survival Curve คือสาเหตุของ Censoring ต้องไม่เกี่ยวข้องกับการเกิดเหตุการณ์ ถ้า Censored เพราะผลการรักษาไม่ดีขึ้นจึงเลิกการติดต่อกับผู้วิจัย Kaplan-Meier Survival Curve ที่ได้ถือว่าไม่เหมาะสม

3.2 วิธีการของ Cox Model

Cox Model หรือ Cox Regression สามารถวิเคราะห์ข้อมูลได้คล้ายกันกับ Multiple Regression เพียงแต่ตัวแปรตามใน Cox Model เป็นความเสี่ยงต่อการเกิดเหตุการณ์ ณ เวลา t

ถ้ามีตัวแปรอิสระ หรือตัวแปรต้น p ตัว แทนด้วย $X_1, X_2, X_3, \dots, X_p$ สามารถเขียน Model ได้ดังนี้

$$h(t) = h_0(t) \times \text{EXP}(b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots + b_p X_p)$$

เมื่อ

$h_0(t)$ เป็น Baseline หรือ Underlying hazard function ได้จากเมื่อให้ค่า X ทุกตัวเป็น 0

$\text{EXP}(\theta)$ คือ Exponential ของค่า θ อาจเขียนใหม่เป็น e^θ ซึ่งเท่ากับ 2.71828^θ

สิ่งควรทราบคือ $\text{Log}_e[\text{EXP}(\theta)] = \text{Ln}[\text{EXP}(\theta)] = \theta$

bp คือค่าสัมประสิทธิ์ (Coefficient) ที่ได้จากการ Fit model

จากสมการข้างต้น จะเห็นว่า Hazard ของ X คือ Exponential ของ bX ดังนั้น อาจกล่าวอีกนัยหนึ่งได้ว่า bX ก็คือ Log Hazard ของ X นั่นเอง

ค่า b คือค่าสัมประสิทธิ์ คำนวณด้วยวิธีการที่สลับซับซ้อนโดยอาศัยคอมพิวเตอร์ เรียกว่า Fit Model วิธีการที่ใช้ประมาณค่า b เรียกว่า Partial maximum likelihood

ค่า Likelihood (L) คือค่าความน่าจะเป็นที่ค่าสังเกตจะถูกอธิบายโดย Model ค่า ยิ่งสูง ยิ่งดีปกติผลจากคอมพิวเตอร์รายงานค่านี้เป็นสเกล Log เรียก Log-likelihood เพราะง่ายที่จะนำไป คำนวณขั้นต่อไป ด้วยค่าความน่าจะเป็นอยู่ระหว่าง 0 - 1 ค่าที่สูงคือค่าที่เข้าใกล้ 1 เมื่อเป็นค่า Log แล้ว ค่าที่เข้าใกล้ 1 จะได้ค่า Log เข้าใกล้ 0 และ Log ของค่าที่น้อยกว่า 1 มีค่าติดลบเสมอ ดังนั้น ค่า Log-likelihood จึงเป็นค่าที่ติดลบเสมอ และค่ายิ่งใกล้ 0 ยิ่งดี

การได้มาซึ่งค่า b นั้นเป็นการเปรียบเทียบ Likelihood ของ Model ที่กำหนดให้ ตัวแปรต้นทุกตัวไม่มีผลต่อ Model เขียนแทนด้วย $L(0)$ กับ Likelihood ของ Model ที่ตัวแปรต้นทุกตัวมีผลต่อ Model เขียนแทนด้วย $L(b)$ แล้วคำนวณหาค่า $L(b)$ ที่มากที่สุดเท่าที่จะเป็นไปได้ที่ยังผลให้ได้ค่าสัดส่วนระหว่าง $L(0)/L(b)$ หรือ Likelihood ratio นี้มีค่าต่ำสุด จึงเรียกว่า Maximum likelihood ค่าของ Linear combination คือ $b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_pX_p$ ของคนคนหนึ่ง หาได้จากการแทนค่า X ต่างๆ ลงในสมการที่ได้จากการ Fit model ค่าที่ได้เรียกว่า Prognostic Index หรือ Risk score มีค่าเท่ากับ Log Hazard

ภายใต้ Model นี้มี Assumption เกี่ยวกับ Proportional hazard เป็นหลัก กล่าวคือความแตกต่างสัมพัทธ์หรือสัดส่วนของ Log hazard ระหว่างกลุ่มนั้น เท่ากันตลอดเวลา ที่ศึกษา ไม่ว่าจะมียุคใดก็ตาม จะเพิ่มหรือลดเมื่อเวลาเปลี่ยนไป หรือเพิ่มแล้วลด หรือลดแล้วเพิ่ม ขอเพียงแต่ต้องเป็นสัดส่วนคงที่ระหว่างกลุ่มหรือหน่วยศึกษาทุกหน่วยนั้นคือ Cox proportional hazard ไม่มีการระบุลักษณะของ Baseline hazard แต่ให้คงที่ (Baseline hazard

is left unparameterized but fixed) แต่ยังคงมีการประมาณค่าพารามิเตอร์ของสัมประสิทธิ์ จึงเรียกว่าเป็น Semiparametric model แบบหนึ่ง

Hazard บอกความเสี่ยงต่อการเกิดเหตุการณ์ ณ เวลา t ถ้าต้องการทราบความเสี่ยงต่อการเกิดเหตุการณ์ระหว่างเวลา 0 ถึง เวลา t ก็เพียงแต่รวมค่า hazard ทุกจุดเวลาจนถึงเวลา t เรียกว่า Cumulative hazard เขียนแทนด้วย $H(t)$ ดังนั้น Model สามารถเขียนใหม่เป็น

$$H(t) = H_0(t) \times \text{EXP}(b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots + b_p X_p)$$

จากที่กล่าวข้างต้น เราทราบว่า โอกาสที่อยู่อย่างปลอดภัยจนถึงเวลา t ซึ่งเขียนแทนด้วย $S(t)$ นั้น สามารถหาได้จาก $\text{EXP}[-H(t)]$ ดังนั้นเราจึงสามารถหาโอกาสที่คนคนหนึ่งจะอยู่อย่างปลอดภัยเมื่อเราทราบค่าตัวแปรต้นต่างๆ ที่เป็นข้อมูลของเขา การแปลความหมายจาก

$$h(t) = h_0(t) \times \text{EXP}(b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots + b_p X_p)$$

สามารถดูที่ขนาดของค่า b ค่ายิ่งมาก บ่งชี้ว่ายิ่งมีโอกาสเกิดเหตุการณ์ได้มาก Exponentiate ค่า b เรียกว่า Hazard แปลความหมายคล้ายอัตรา (Rate)

กรอบแนวความคิดของโครงการวิจัย

