

รายงานการศึกษา ฝึกอบรม ประชุม ดูงาน สัมมนา ปฏิบัติการวิจัย ในประเทศ ระยะสั้นไม่เกิน ๙๐ วัน  
 รายงานการอบรมเชิงปฏิบัติการ โครงการอบรมเข้มข้นระยะสั้นสถิติศาสตร์คลินิก ประจำปี ๒๕๖๗  
 “Short Intensive Course Diploma in Clinical Statistics 2024”

.....

ส่วนที่ ๑

ข้อมูลทั่วไป

๑.๑ ชื่อ - นามสกุล นางสาวศรสวรรค์ คิ้ววิจิตร

อายุ ๓๒ ปี การศึกษาปริญญาตรี สาธารณสุขศาสตรบัณฑิต

ความเชี่ยวชาญเฉพาะด้าน สาธารณสุข

ตำแหน่ง นักวิชาการสาธารณสุขปฏิบัติการ

หน้าที่ความรับผิดชอบ จัดทำโครงการอบรมด้านการวิจัยให้กับบุคลากรทางการแพทย์และสาธารณสุข จัดทำโครงการทุนส่งเสริมการวิจัยทางการแพทย์และสาธารณสุข ส่งเสริม สนับสนุน และให้คำแนะนำด้านระบาดวิทยา ชีวสถิติ แก่ผู้สนใจทำวิจัยในโรงพยาบาล

ชื่อเรื่อง / หลักสูตร การอบรม โครงการอบรมเข้มข้นระยะสั้นสถิติศาสตร์คลินิก ประจำปี ๒๕๖๗

“Short Intensive Course Diploma in Clinical Statistics 2024”

เพื่อ  ศึกษา  ฝึกอบรม  ประชุม  ดูงาน  สัมมนา  ปฏิบัติการวิจัย

งบประมาณ  เงินงบประมาณกรุงเทพมหานคร  เงินบำรุงโรงพยาบาล

ทุนส่วนตัว

จำนวนเงิน ๒๕,๐๐๐.- บาท (สองหมื่นห้าพันบาทถ้วน)

ระหว่างวันที่ ๑ - ๒๙ มีนาคม ๒๕๖๗

สถานที่ ณ อาคารคุณากร คณะแพทยศาสตร์ มหาวิทยาลัยธรรมศาสตร์

คุณวุฒิ / วุฒิบัตรที่ได้รับ -

การเผยแพร่รายงานผลการศึกษา/ฝึกอบรม/ ประชุม สัมมนา ผ่านเว็บไซต์สำนักงานการแพทย์และกรุงเทพมหานคร

ยินยอม

ไม่ยินยอม

ส่วนที่ ๒

ข้อมูลที่ได้รับจากการศึกษา ฝึกอบรม ประชุม ดูงาน สัมมนา ปฏิบัติการวิจัย

๒.๑ วัตถุประสงค์

๒.๑.๑ เพื่อเพิ่มพูนความรู้ความเข้าใจในการวิเคราะห์ข้อมูลทางคลินิก และการเลือกใช้สถิติให้สอดคล้องกับข้อมูลและระเบียบวิธีวิจัยที่ศึกษาได้อย่างถูกต้อง

๒.๑.๒ เพื่อเพิ่มพูนความรู้ความสามารถในการใช้โปรแกรม STATA ซึ่งเป็นโปรแกรมที่ใช้ในการวิเคราะห์สถิติ โดยเฉพาะด้านการวิจัยทางการแพทย์และสาธารณสุขได้อย่างถูกต้อง

๒.๒ เนื้อหา

๒.๒.๑ สถิติพื้นฐาน

การวิเคราะห์ทางสถิติที่เกี่ยวข้องกับตัวทำนายหลายตัวเป็นการสรุปของเทคนิคให้เรียบง่ายขึ้น โดยถูกพัฒนาขึ้นเพื่อการสำรวจความสัมพันธ์ระหว่างผลลัพธ์กับตัวทำนายเดียว บทนี้จะศึกษาวิธีพื้นฐาน

สำหรับผลลัพธ์ที่เป็นตัวแปรต่อเนื่อง ซึ่งรวมถึง t-test และการวิเคราะห์ความแปรปรวนแบบหนึ่งทาง (one-way analysis of variance) ค่าสหสัมพันธ์ (correlation coefficient)

(๑) การทดสอบ t-test เบื้องต้นใช้ในการเปรียบเทียบสองตัวอย่างที่เป็นอิสระกัน ค่าสถิติ t ที่ใช้ในการทดสอบคือความแตกต่างระหว่างค่าเฉลี่ยของสองตัวอย่าง หาดด้วยค่าคลาดเคลื่อนมาตรฐานของความแตกต่างนั้น การทดสอบ t-test ถูกออกแบบให้ทำงานได้ในตัวอย่างที่มีขนาดเล็ก ในขณะที่การทดสอบ Z-test ทำไม่ได้

(๒) การทดสอบ t-test แบบผู้ทดลองผูกพัน (paired t-test) ใช้ในสถานการณ์ที่บุคคลหรือการสังเกตถูกเชื่อมโยงกันระหว่างสองกลุ่มตัวอย่าง ตัวอย่างเช่นการวัดที่ทำไว้ในสองจุดเวลาต่างกันของบุคคลเดียวกัน หรือในคู่ที่เชื่อมโยงโดยธรรมชาติอื่น ๆ เช่นในการทดลองทางคลินิกที่ตาหนึ่งถูกรักษาและตาอีกตาหนึ่งเป็นตัวควบคุม ในกรณีนี้ สองตัวอย่างไม่เป็นอิสระกันและการละเว้นความสัมพันธ์แบบคู่อาจทำให้ข้อมูลสูญเสียและเป็นไปในทางที่ผิดพลาดได้

ขั้นตอนการทดสอบ t-test แบบผู้ทดลองผูกพันจะเริ่มต้นด้วยการคำนวณความแตกต่างแบบคู่สำหรับแต่ละบุคคลหรือคู่ที่เชื่อมโยง ในตัวอย่างแรกนี้ นั่นคือการเปลี่ยนแปลงของผลลัพธ์จากจุดเวลาที่หนึ่งไปยังอีกจุดเวลาหนึ่ง และในตัวอย่างที่สองคือความแตกต่างระหว่างผลลัพธ์ของตาที่ได้รับการรักษาและตาที่เป็นตัวควบคุม จากนั้นจึงใช้การทดสอบ t-test เพื่อประเมินว่าค่าเฉลี่ยของประชากรของความแตกต่างแบบคู่เหล่านี้แตกต่างจากศูนย์หรือไม่ การเพิ่มพลังขึ้นเนื่องจากความแปรปรวนระหว่างบุคคลถูกกำจัดในขั้นตอนแรก การทดสอบ t-test แบบผู้ทดลองผูกพันยังสามารถทำได้โดยใช้คำสั่ง ttest ใน STATA

(๓) การวิเคราะห์ความแปรปรวนแบบหนึ่งทาง (One-Way Analysis of Variance, ANOVA) สมมติว่าเราต้องการเปรียบเทียบค่าเฉลี่ยตัวอย่างระหว่างกลุ่มการทดลองทางคลินิกที่มีการรักษาหลายประเภทหรือทั้งหมดต่อกลุ่มตัวอย่างที่อิสระกันมากกว่าสองกลุ่ม สำหรับวัตถุประสงค์นี้ การวิเคราะห์ความแปรปรวนแบบหนึ่งทาง (ANOVA) และการทดสอบ F-test จะมาแทนที่ของการทดสอบ t-test

### ๒.๒.๒ การถดถอยเชิงเส้น (Linear Regression)

(๑) การถดถอยเชิงเส้นแบบง่าย (Simple Linear Regression) มีผลลัพธ์เป็นตัวแปรต่อเนื่องและตัวทำนายเป็นตัวแปรต่อเนื่องเดียว ในการถดถอยเชิงเส้น เราจำแนกโมเดลค่าเฉลี่ยของตัวแปรผลลัพธ์ที่เป็นตัวเลข  $y$  เป็นฟังก์ชันของตัวทำนายตัวเลข  $x$  เดียวโดยใช้ความสัมพันธ์เชิงเส้นตามรูปแบบนี้

$$E[y|x] = \beta_0 + \beta_1 x.$$

(๒) การวิเคราะห์ถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) การถดถอยเชิงเส้นพร้อมกับตัวทำนายหลายตัวแปร เป็นเครื่องมือที่ยืดหยุ่นและใช้กันอย่างแพร่หลายในการประเมินความสัมพันธ์ร่วมกันของตัวทำนายหลายตัวแทนกับตัวแปรผลลัพธ์ที่เป็นตัวแปรต่อเนื่อง

$$E[y|\mathbf{x}] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p,$$

### ๒.๒.๓ การถดถอยโลจิสติก (Logistic Regression)

(๑) การถดถอยโลจิสติกแบบง่าย (Simple Logistic Regression) มีผลลัพธ์เป็นตัวแปรเจนนับและตัวทำนายมี ๑ ตัวแปร

$$P(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}$$

(๒) การวิเคราะห์ถดถอยโลจิสติกพหุคูณ (Multiple Logistic Regression) การศึกษาทางคลินิกและระบาดวิทยาเกี่ยวกับผลลัพธ์ที่เป็นไบนารีมักเน้นไปที่ผลกระทบที่เป็นไปได้ของตัวทำนายหลายตัวแทน ในกรณีที่เหล่านี้เป็นข้อมูลที่เป็นหมวดหมู่และมีจำนวนน้อย วิธีการตารางการตรวจสอบเพียงพอสำหรับการวิเคราะห์ข้อมูล อย่างไรก็ตาม สำหรับจำนวนตัวทำนายที่มากกว่าและเมื่อบางส่วนเป็นการวัดต่อเนื่อง วิธีการถดถอยมีข้อดีหลายประการ ตัวอย่างเช่น การศึกษา WCGS ได้ทำการวัดตัวทำนายที่เป็นไปได้ของโรคหัวใจพบว่ามีความเสี่ยงต่อโรคหัวใจในเลือดทั้งหมด ความดันโลหิตช่วงต่ำและสูง การสูบบุหรี่ อายุ ขนาดร่างกาย และแบบแพทเทิร์น นักวิจัยรู้ถึงว่าตัวแปรเหล่านี้มีส่วนร่วมในความเสี่ยงของผลลัพธ์นอกจากการเชื่อมโยงกัน และในการประเมินผลกระทบของตัวทำนายที่เลือก อาจมีความสำคัญที่จะควบคุมผลกระทบที่สับสนของตัวแปรอื่น ๆ เนื่องจากมีตัวทำนายที่เป็นตัวเลือกหลายตัวและบางส่วนอาจมองเห็นได้เป็นการวัดต่อเนื่อง เทคนิคการถดถอยหลายตัวแทนมีความน่าสนใจมากในการวิเคราะห์ข้อมูลเช่นนี้ โมเดลการถดถอยโลจิสติกสำหรับตัวแปรตัวทำนายหลายตัวแทนเป็นการแยกประเภทโดยตรงจากเวกเตอร์สำหรับตัวทำนายเดี่ยวสำหรับผลลัพธ์ที่เป็นไบนารี  $y$  และตัวทำนาย  $p$  ตัว  $x_1, x_2, \dots, x_p$  ส่วนระบบของโมเดลถูกกำหนดไว้ดังนี้:

$$P(x_1, x_2, \dots, x_p) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}$$

### ๒.๒.๔ การถดถอยการรอดชีพ (Cox Regression)

โมเดลการถดถอยการรอดชีพที่สัมพันธ์กัน การวิเคราะห์การรอดชีพ (Cox Regression) เป็นเครื่องมือที่ยืดหยุ่นสำหรับการประเมินความสัมพันธ์ของตัวทำนายหลายตัวแทนกับผลลัพธ์ที่เป็นเหตุการณ์ซึ่งมีเวลาถึงการเกิด และมีหน่วยเวลาที่ถูกต้องชัดเจน และมีอยู่เหมือนกันกับโมเดลเชิงเส้นและโลจิสติก ในการเข้าใจวิธีการทำงานของการถดถอยการรอดชีพเราจะพิจารณาว่าโมเดลที่สัมพันธ์กันอย่างสมสัดส่วนมีความแตกต่างหรือมีข้อไหนดีกว่ากัน ในโมเดลการถดถอยการรอดชีพที่สัมพันธ์กันอย่างสมสัดส่วน หากอัตราการรอดชีพเป็นไปตามสมมติฐานของอันตรายอันตรายที่สัมพันธ์กันอย่างสมสัดส่วน และไม่ขึ้นอยู่กับเวลา เราสามารถเขียนได้ว่า

$$\log[h(t|\mathbf{x})] = \log[h_0(t)] + \beta_1 x_1 + \dots + \beta_p x_p$$

### ๒.๒.๕ การถดถอยโลจิสติกเชิงลำดับ และ แบบมัลติโนเมียลโลจิสติก (Ordinal และ Multinomial Logistic Regression)

(๑) การถดถอยโลจิสติกเชิงลำดับ (Ordinal Logistic Regression) โมเดลสัมพัทธ์ความสัมพันธ์ที่ใช้กันอย่างแพร่หลายของโมเดลโลจิสติกซึ่งเหมาะสำหรับการตอบสนองทางมีขั้นขั้นมากเท่าไรก็ได้ ต่างจากการจำแนกโอกาสของการตอบสนองในหมวดหมู่ที่เฉพาะเจาะจงหรือแค่ ๒ ตัวเลือก การวิเคราะห์นี้ขึ้นอยู่กับความน่าจะเป็นสะสมที่การตอบสนองไม่มากกว่าหมวดหมู่ที่ถูกเลือก ความขึ้นต่อกันนี้ขึ้นอยู่กับตัวทำนายในรูปแบบเดียวกับการวิเคราะห์โลจิสติก พารามิเตอร์ความหมายของ  $\beta$  ถูกตีความว่าเป็นอัตราส่วนความเสี่ยงลอกล็อกที่เกี่ยวข้องกับการเพิ่มขึ้นของ  $x$  ซึ่งถูกสมมติว่าคงที่ในทุกๆ ระดับของการตอบสนอง (กล่าวคือ ระดับการตอบสนองเป็นขนาดเดียวกัน แต่ระดับมีความชันเป็น  $\beta$ ) สมมติฐานนี้เทียบเท่ากับข้อจำกัดที่แข็งแกร่งในการผลิตผลของตัวทำนายต่อการตอบสนอง และต้องได้รับการตรวจสอบความถูกต้อง โปรดทราบว่า มีทางเลือกมากมายสำหรับโมเดลสัมพัทธ์อัตราต่อเนื่อง รวมถึงโมเดลอัตราต่อเนื่อง สำหรับตัวอย่างของการปวดหลัง (โดยสมมติว่ามี การตอบสนองระดับ ๑๐ และตัวทำนายเพียงตัวเดียว  $x$ ) รูปแบบของโมเดลนี้สำหรับความน่าจะเป็นของการตอบสนองที่มีความรุนแรงไม่เกินระดับ ๕ มีดังนี้

$$\log \left[ \frac{\Pr(y \leq 5)}{\Pr(y > 5)} \right] = \alpha_5 - \beta x.$$

(๒) การถดถอยโลจิสติกแบบมัลติโนเมียลโลจิสติก (Multinomial Logistic Regression) เมื่อไม่มีลำดับธรรมชาติที่ถาวรในการตอบสนองทางประเภทหรือเมื่อสมมติฐานที่ถาวรในโมเดลด้านบนไม่สามารถใช้กับผลลัพธ์ที่เป็นตัวแปรอันดับได้ โมเดลโพลีโทมัสโลจิสติกสามารถนำมาใช้ในการวิเคราะห์การถดถอยได้ สำหรับตัวทำนายเดี่ยว  $x$  โมเดลระบุว่าแต่ละระดับการตอบสนองเป็นตัวแสดงของโมเดลโลจิสติกสำหรับ  $x$  โดยระดับที่เลือกไว้จะระบุว่าเป็นหมวดอ้างอิง พารามิเตอร์การถดถอยสำหรับแต่ละระดับจะไม่ซ้ำกัน ดังนั้นสำหรับตัวอย่างการปวดหลัง โมเดลจะรวมถึงพารามิเตอร์ตัดเชิงเดียวกับสองพารามิเตอร์เส้นทแยงมุมสำหรับระดับ ๕ และระดับแรกเป็นหมวดอ้างอิง รูปแบบของโมเดลจะมีรูปแบบดังนี้

$$\log \left[ \frac{\Pr(y = 5)}{\Pr(y = 1)} \right] = \alpha_5 + \beta_5 x.$$

### ๒.๒.๖ การถดถอยการรอดชีพแบบพาราเมตริก (Parametric Survival Regression)

เรามีตัวเลือกสองอย่างในการจัดการกับความเสี่ยงในระดับพื้นฐาน หนึ่งคือการโมเดลด้วยฟังก์ชันพารามิเตอร์เชิงพาราเมตริก เช่นในโมเดลการอยู่รอดแบบไฟบูลหรือแบบเอ็กโพเนนเชียล ในกรณีนี้ ความเสี่ยงในระดับพื้นฐาน ถูกกำหนดโดยจำนวนของพารามิเตอร์เพิ่มเติมที่น้อยมาก ซึ่งจะถูกระบุพร้อมกับ หากความเสี่ยงในระดับพื้นฐานถูกกำหนดไว้อย่างถูกต้อง วิธีการนี้จะมีประสิทธิภาพ จัดการกับการตัดไปทางขวา รวมถึงการวางแผนระบบการตัดที่ซับซ้อนมากขึ้นได้โดยง่าย และทำให้ง่าย (แม้ว่ายังมีความเสี่ยง) ในการยับยั้งที่เกินไปนอกเหนือจากข้อมูล แน่แน่นอนว่าความเหมาะสมของโมเดลสำหรับความเสี่ยงในระดับพื้นฐานจะต้องถูกตรวจสอบ

### ๒.๒.๗ โมเดลเชิงเส้นผสม (Linear mixed model)

ในการวิเคราะห์ข้อมูลระยะยาว (longitudinal data) โดยทั่วไปเราสนใจในการเปลี่ยนแปลงของค่าของตัวแปรภายใน "รายการ" และเราเก็บข้อมูลอย่างสม่ำเสมอตลอดเวลา ตัวอย่างเช่น การศึกษาผลกระทบของแอลกอฮอล์อาจบันทึกการวัดระดับความง่วงก่อนและหลังการให้แอลกอฮอล์หรือยาเสพติดเสมือน ความสนใจอยู่ที่การประเมินผลของแอลกอฮอล์ต่อการเปลี่ยนแปลงในความง่วงนั่นเอง นี่เป็นกลยุทธ์การออกแบบที่ดีเนื่องจากแต่ละรายการมีการใช้ตัวควบคุมตนเอง ทำให้สามารถลดความแปรปรวนในการวัดความง่วงจากบุคคลไปยังบุคคลหรือแม้กระทั่งตามโอกาสภายในบุคคลเดียวกันได้ ในการที่กลยุทธ์นี้จะมีประสิทธิภาพ การวัดก่อนและหลังจะต้องมีความสัมพันธ์กันอย่างเชื่อถือได้อย่างน้อย (มิฉะนั้น การรับค่าความแตกต่างจะทำให้ความแปรปรวนเพิ่มขึ้นแทนที่จะลดลง)

### ๒.๒.๘ สมการการประมาณค่าทำนายทั่วไป (Generalized Estimating Equation: GEE)

เราจะพิจารณาเทคนิคที่เรียกว่า สมการการประมาณค่าทำนายทั่วไป generalized estimating equations (GEE) ซึ่งมีก้อยย่อว่า GEE ลักษณะสำคัญของวิธีนี้คือตัวเลือกในการประมาณค่าโครงสร้างความสัมพันธ์จากข้อมูลโดยไม่ต้องสมมติว่ามันเป็นโครงสร้างที่ระบุไว้ล่วงหน้า โดยจะต้องพิจารณา ๕ ด้านของข้อมูล ได้แก่

(๑) การกระจายของข้อมูล (สำหรับค่าคงที่ของตัวแปรความสนใจ) ที่เหมาะสมที่จะใช้สำหรับตัวแปรผลลัพธ์คืออะไร? ตัวอย่างเช่น การกระจายปกติ (normal), ตัวแปรแบบไบนารี (binary), และและตัวแปรแบบโบนิเมียว

(๒) ตัวทำนายที่เราจะรวมอยู่ในโมเดลคืออะไร

(๓) ในวิธีใดเราจะเชื่อมโยงตัวทำนายกับข้อมูล (ผ่านค่าเฉลี่ย ผ่านโลจิสของความเสี่ยง วิธีอื่น ๆ)

(๔) โครงสร้างความสัมพันธ์ที่จะใช้หรือสมมติชั่วคราวเพื่อสร้างการประมาณค่าคืออะไร

(๕) ตัวแปรใดบ้างที่ข้อมูลถูกจัดกลุ่มอย่างไร

### ๒.๒.๙ การวิเคราะห์รูปแบบเชิงคาดการณ์ของข้อมูลทางคลินิก (Clinical Prediction Models)

การแพทย์ระดับบุคคล การพยากรณ์โรคที่บ่งบอกหรือโรคที่เกิดขึ้น (การวินิจฉัย) หรือผลลัพธ์สำหรับเส้นทางของโรคในอนาคต (การพยากรณ์) เป็นสิ่งสำคัญมากขึ้นสำหรับการจัดการผู้ป่วยโดยบุคคลในด้านสุขภาพ โมเดลการพยากรณ์ทางคลินิกถูกใช้ในการสำรวจความสัมพันธ์ระหว่างผลลัพธ์ที่เกิดขึ้นในอนาคตหรือที่รู้ล่วงหน้า (จุดสิ้นสุด) และสถานะสุขภาพเริ่มต้น (จุดเริ่มต้น) ในกลุ่มคนที่มีเงื่อนไขเฉพาะพวกเขาทั่วไปรวมกันพารามิเตอร์หลายรายการเพื่อให้เห็นภาพของผลกระทบที่สำคัญของตัวทำนายแต่ละตัวในโมเดล การแพทย์ขึ้นอยู่กับข้อมูลที่เป็นหลักฐานที่มีความเข้มงวดที่สุด ซึ่งรวมถึงผลจากการทดลองแบบสุ่มควบคุม การวิเคราะห์รวมข้อมูลจากหลายศูนย์ และการรวบรวมข้อมูลจากการทบทวนทางระบบ แม้ว่าโมเดลการพยากรณ์ทางคลินิกจะพึ่งบางส่วนบนการแพทย์ที่มีหลักฐานที่มีการสนับสนุน ผู้ใช้ต้องยอมรับความประสมพันธ์และการใช้วิธีที่เหมาะสมในการสร้างโมเดลที่เกี่ยวข้องกับผู้ใช้เป้าหมายอย่างเข้าถึงได้และความหมายทางคลินิก

โมเดลความน่าจะเป็นควรทำนายเหตุการณ์ที่เฉพาะเจาะจงอย่างแม่นยำและมีความง่ายและใช้งานง่ายในระดับหนึ่ง หากโมเดลการพยากรณ์ให้การประมาณค่าไม่แม่นยำของเหตุการณ์ที่เกิดขึ้นในอนาคตมันจะทำให้ผู้ให้บริการด้านสุขภาพให้การจัดการกับผู้ป่วยหรือทรัพยากรอย่างไม่เพียงพอ อย่างไรก็ตาม

หากโมเดลมีพลังในการพยากรณ์สูง แต่ยากในการใช้งาน (เช่น มีการคำนวณที่ซับซ้อนหรือมีค่าถาม/รายการหรือหน่วยที่ไม่เคยเจอมาก่อน) เสียเวลา มีค่าใช้จ่าย หรือไม่เกี่ยวข้องมากนัก (เช่น โมเดลยุโรปสำหรับชาวเกาหลี หรือเหตุการณ์ที่อยู่ไกลเกินไป) มันจะไม่ได้รับการใช้งานอย่างแพร่หลาย ตัวอย่างเช่น โมเดลการพยากรณ์เบาหวานที่พัฒนาโดย Lim et al. มีพื้นที่ใต้เส้นโค้งการทำงานของผู้รับรู้อัจฉริยะ (AUC, ๐.๗๗) ในขณะที่ตรวจสุขภาพที่วัดฮีโมโกลบิน A๑c, ไฮโดรเจนไลโปโปรตีนโคเลสเตอรอลความหนาและไตรกลีเซอไรด์ถูกรวมอยู่ในคะแนนความเสี่ยง ซึ่งมักจะต้องมีความเกี่ยวข้องของแพทย์ผู้ประเมินจึงเป็นอุปสรรคสำคัญในการใช้ในสถานการณ์ชุมชน เมื่อโมเดลการพยากรณ์ประกอบด้วยสมการทางคณิตศาสตร์ที่ซับซ้อน, การใช้งานในรูปแบบเว็บสามารถเสริมประสิทธิภาพในการใช้งานได้ (เช่น การคำนวณความเสี่ยงในระยะเวลา ๑๐ ปีและทั่วชีวิตสำหรับโรคหัวใจและหลอดเลือด [CVD] ดังนั้น การประสบความสำเร็จในการพยากรณ์และความง่ายตายเป็นจุดสำคัญในการสร้างโมเดลการพยากรณ์ทางคลินิกที่ดี

### ๒.๒.๑๐ การจัดการข้อมูลที่สูญหาย (Missing data)

การหลีกเลี่ยงข้อมูลที่ขาดหายเป็นสิ่งสำคัญที่นักวิจัยพยายามทำเพราะพวกเขามีอยู่ในทุก ๆ การศึกษาเกือบทุกเรื่อง การละเว้นข้อมูลที่ขาดหายในการวิเคราะห์ทางสถิติสามารถทำให้ผลการศึกษารับอิทธิพลที่ผิดหวังอย่างรุนแรง Rubin เป็นคนแรกที่พัฒนารอบการของประเภทของข้อมูลที่ขาดหายต่าง ๆ (กลไกของข้อมูลที่ขาดหาย) ที่สำคัญที่จะกำหนดขั้นตอนต่อไปในการจัดการข้อมูลที่ขาดหาย มีทั้งหมด ๓ ประเภทคือ ข้อมูลที่ขาดหายแบบสมบูรณ์แบบสุ่ม (MCAR), ข้อมูลที่ขาดหายแบบสุ่ม (MAR), และข้อมูลที่ขาดหายไม่สุ่ม (MNAR) MCAR หมายความว่าค่าที่ขาดหายแบบสุ่มกระจายตัวอย่างข้อมูล เหตุผลที่ข้อมูลขาดหายไม่ได้เกี่ยวข้องกับตัวแปรหรือผลลัพธ์ของการศึกษาที่เกี่ยวข้อง ตัวอย่างเช่น สมมติว่ามีการศึกษาซึ่งเชิญชวนคนที่มีความดันเลือดในครอบครัวมายังศูนย์วิจัยเพื่อวัดความดันเลือดและความสัมพันธ์ของตัวแปรที่เกี่ยวข้องกับความดันเลือดในกลุ่มประชากรนี้ เมื่อข้อมูลเกี่ยวกับความดันเลือดขาดหาย เพราะบางคนไม่สามารถมาเยี่ยมศูนย์วิจัยได้เนื่องจาก เช่น การสัตรีคในการขนส่งสาธารณะ ข้อมูลที่ขาดหายเหล่านี้จะเป็น MCAR MAR หมายความว่าความน่าจะเป็นของข้อมูลที่ขาดหายเกี่ยวข้องกับตัวแปรอื่น ๆ เช่น เมื่อข้อมูลที่เกี่ยวกับความดันเลือดขาดหายมากขึ้นในผู้ที่มีดัชนีมวลกายสูง ข้อมูลที่ขาดหายเหล่านี้จะเป็น MAR ข้อมูลที่ขาดหายไม่สุ่ม (MNAR) คือเมื่อความน่าจะเป็นของข้อมูลที่ขาดหายเกี่ยวข้องกับค่าของตัวเอง เช่น เมื่อคนที่มีความดันเลือดสูงที่สุดไม่ไปเยี่ยมศูนย์วิจัย สถานการณ์นี้ที่สุดนี้มีปัญหา เพราะคุณไม่รู้ว่าเป็นเช่นนั้นหรือไม่ ในกรณีที่ข้อมูลที่ขาดหายเป็น MNAR ไม่มีวิธีการที่ง่ายในการสร้างผลลัพธ์ที่ถูกต้อง หนึ่งในทางเลือกคือการดำเนินการทำการวิเคราะห์ความไว้วางใจเพื่อศึกษาผลกระทบของข้อมูลที่ขาดหายต่อผลลัพธ์ของการศึกษา ควรรู้ว่ากลไกของข้อมูลที่ขาดหายขึ้นอยู่กับตัวแปรซึ่งในศึกษาหนึ่งๆ ข้อมูลที่ขาดหายบางส่วนอาจจะเป็น MCAR ในขณะที่สำหรับตัวแปรอื่น ๆ ข้อมูลที่ขาดหายอาจเป็น MAR หรือ MNAR เกี่ยวกับกลไกของข้อมูลที่ขาดหายไม่ว่าตัวแปรเฉพาะที่ขาดหายจะเป็นตัวแปรผลลัพธ์ของการศึกษาหรือหนึ่งในตัวแปรที่เกี่ยวข้อง

จะเข้าใจว่าตามนิยามแล้วไม่สามารถประเมินได้ว่าข้อมูลที่ขาดหายนั้นเป็น MAR หรือ MNAR ความแตกต่างระหว่างสองนั้นคือเมื่อข้อมูลที่ขาดหายเป็น MNAR ข้อมูลที่ขาดหายนั้นเกี่ยวข้องกับข้อมูลที่ไม่เคยสังเกตเห็นและเนื่องจากข้อมูลเหล่านั้นไม่เคยสังเกตเห็นและเป็นที่ยอมรับ จึงเป็นไปได้ที่จะประเมินว่าข้อมูลที่ไม่มีเคยสังเกตเห็นเกี่ยวข้องกับข้อมูลที่ขาดหายหรือไม่ อย่างไรก็ตาม ยังมีวิธีการหลายวิธีที่สามารถใช้สำรวจว่าข้อมูลเป็น MCAR หรือไม่ การทดสอบแบบ t-test และการวิเคราะห์โลจิสติกสามารถใช้สำหรับสำรวจว่ามีความสัมพันธ์ระหว่างตัวแปรที่มีข้อมูลที่ขาดหายและไม่มีข้อมูลที่ขาดหายหรือไม่ ตัวแปรที่มีข้อมูลที่ขาดหายสามารถเข้ารหัสเป็น ๐ สำหรับข้อมูลที่สังเกตเห็นและ ๑ สำหรับข้อมูลที่ขาดหาย เมื่อใช้ตัวแปรนี้ (เช่น ตัว

แปรไอondiเคเตอร์ของข้อมูลที่ขาดหาย) เป็นตัวแบ่งกลุ่มในการทดสอบ t-test หรือเป็นผลลัพธ์ในการวิเคราะห์โลจิสติก ความสัมพันธ์กับตัวแปรอื่น ๆ สามารถสำรวจได้ วิธีการอื่น ๆ ที่สามารถใช้งานได้คือการทดสอบ MCAR ของ Little

### ๒.๓ ประโยชน์ที่ได้รับ

๒.๓.๑  ต่อตนเอง เพิ่มศักยภาพ ความรู้ ความเข้าใจด้านหลักการสถิติคลินิก สามารถวิเคราะห์ได้ตั้งแต่ระดับพื้นฐานจนถึงระดับสูง พัฒนาทักษะด้านการวิเคราะห์ และสร้างองค์ความรู้ใหม่ทางวิชาการ โดยวิธีการทางวิทยาศาสตร์ที่เป็นระบบ สามารถวิเคราะห์ข้อมูลทางสุขภาพได้ทุกรูปแบบ นอกจากนี้ยังได้เรียนรู้เกี่ยวกับการใช้โปรแกรม STATA ซึ่งเป็นโปรแกรมทางสถิติที่นิยมใช้กันในงานวิจัยทางคลินิกอีกด้วย

๒.๓.๒  ต่อหน่วยงาน สามารถนำความรู้ที่ได้รับมาส่งเสริม สนับสนุน และให้คำแนะนำด้านระบาดวิทยา ชีวสถิติ และโปรแกรม STATA ในการหาขนาดตัวอย่างหรือการวิเคราะห์ข้อมูลสถิติให้แก่บุคลากรโรงพยาบาลตากสินที่สนใจทำวิจัยทางการแพทย์และสาธารณสุข

๒.๓.๓  อื่น ๆ (ระบุ) สามารถนำเสนอผลงานวิชาการและวิจัยตีพิมพ์ในวารสารวิชาการระดับชาติหรือนานาชาติที่มีมาตรฐาน เชื่อมโยงการทำงานระหว่างผู้ทำวิจัยกับกลุ่มงานส่งเสริมการวิจัยในโรงพยาบาล

### ส่วนที่ ๓ ปัญหาและอุปสรรค

### ส่วนที่ ๔ ข้อคิดเห็นและข้อเสนอแนะ

๔.๑  ข้อคิดเห็น เนื่องจากเป็นการอบรมในหลักสูตรเฉพาะการวิเคราะห์สถิติเท่านั้น ทำให้เนื้อหายังไม่ครอบคลุมด้านระบาดวิทยาคลินิก เลยอาจจะยังไม่สามารถเข้าใจเนื้อหาทั้งหมดของหลักสูตรได้

๔.๒  ข้อเสนอแนะ ควรจะได้ลงทะเบียนอบรมในหลักสูตรต่อไป ซึ่งเป็นหลักสูตรโครงการอบรมเข้มข้นระยะสั้น ระบาดวิทยาคลินิก Short Intensive Course Diploma in Clinical Epidemiology ซึ่งเป็นการต่อเนื่องกับการอบรมครั้งนี้ ที่มีเนื้อหาครอบคลุมในด้านระบาดวิทยาคลินิกที่จะเข้าใจการใช้สถิติต่าง ๆ มากขึ้น

ลงชื่อ..... dl .....ผู้รายงาน

(นางสาวศรสวรรค์ คิววิจิตร)

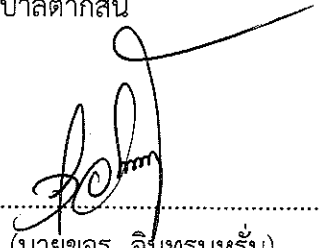
นักวิชาการสาธารณสุขปฏิบัติการ

ส่วนที่ ๕

ความคิดเห็นของผู้บังคับบัญชา

ขอให้นำความรู้ที่ได้ มาพัฒนาหน่วยงาน และโรงพยาบาลตากสิน

ลงชื่อ.....

  
(นายจร อินทรบูรณ์)

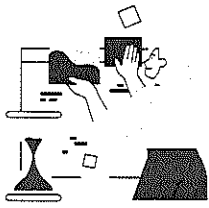
ผู้อำนวยการโรงพยาบาลตากสิน



# โครงการอบรม เข้มข้นระยะสั้น : สถิติศาสตร์คลินิก

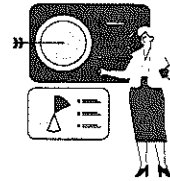


โดยใช้โปรแกรม STATA



## สถิติขั้นพื้นฐาน:

- การเลือกใช้สถิติหลักๆ ขึ้นอยู่กับลักษณะของผลลัพธ์ (outcome) ดังนี้
1. ข้อมูลเป็นตัวแปรต่อเนื่อง หรือ ตัวแปรแจกแจง
  2. กลุ่มตัวอย่างเป็น 1 กลุ่ม 2 กลุ่ม หรือ มากกว่า 2 กลุ่ม
  3. เป็นการแจกแจงปกติหรือไม่



## การวิเคราะห์การถดถอยเชิงเส้น (Linear Regression):

- ตัวแปรตามเป็นข้อมูลต่อเนื่อง
- กราฟเส้นตรง
- เป็นการแจกแจงปกติหรือไม่
- การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตาม
- การพยากรณ์ข้อมูล



## การวิเคราะห์การถดถอยโลจิสติก (Logistic Regression):

- ตัวแปรตามมี 2 ตัวเลือก (Y= 0 หรือ 1)
- กราฟเส้นมีลักษณะตัวเอส (sigmoid)
- การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตาม
- การพยากรณ์ข้อมูล



## การวิเคราะห์การรอดชีพ (Cox Regression):

- ตัวแปรตามมี 2 ตัวเลือก และมีตัวแปรระยะเวลา
- กราฟเส้นมีลักษณะตัวเอส (sigmoid)
- การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตาม
- การพยากรณ์ข้อมูล



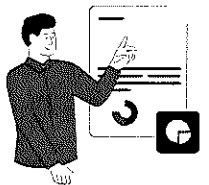
## การวิเคราะห์การถดถอยโลจิสติกเชิงลำดับ และ แบบมัลติโนเมียลโลจิสติก (Ordinal and Multinomial Logistic Regression):

- ตัวแปรตามมีมากกว่า 2 ตัวเลือก
- การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตาม
- การพยากรณ์ข้อมูล



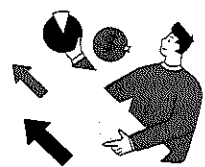
## การคำนวณขนาดตัวอย่าง (Sample size):

- การเลือกใช้สถิติหลักๆ ขึ้นอยู่กับลักษณะของผลลัพธ์ (outcome) ดังนี้
1. ข้อมูลเป็นตัวแปรต่อเนื่อง หรือ ตัวแปรแจกแจง
  2. กลุ่มตัวอย่างเป็น 1 กลุ่ม 2 กลุ่ม หรือมากกว่า 2 กลุ่ม
  3. จำนวนครั้งที่วัดผล
  4. เป็นการศึกษาที่เกี่ยวกับเวลาหรือไม่



## การวิเคราะห์ห่อภิมา (Meta-analysis):

- การค้นหาผลงานวิจัย
- การประเมินคุณภาพเอกสารการวิจัย
- การสังเคราะห์ข้อมูล
- การทดสอบความเป็นเอกพันธ์ของข้อมูล
- อดตีจากการตีพิมพ์



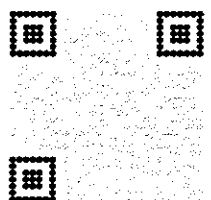
## การวิเคราะห์รูปแบบเชิงคาดการณ์ของข้อมูลทางคลินิก:

- เตรียมการสำหรับการสร้างโมเดล
- การเลือกชุดข้อมูล
- การจัดการตัวแปร
- การสร้างโมเดล
- การประเมินและการตรวจสอบแบบจำลอง (ภายในหรือภายนอก)



## การจัดการข้อมูลที่สูญหาย:

- การป้องกันการสูญหายเสมอคือการแก้ไขข้อมูล
- เมื่อข้อมูลที่สูญหายเป็น MCAR การวิเคราะห์ข้อมูลกรณีสมบูรณ์อาจจะต้อง
- การสังเคราะห์ข้อมูล
- วิธีการเติมค่าเดียนำไปสู่การประมาณค่าของค่าคลาดเคลื่อนที่น้อยกว่า



## การนำมาใช้:

ได้รับความรู้ในด้านสถิติที่ใช้กับงานวิจัยทางคลินิก สามารถเลือกสถิติที่เหมาะสมและสามารถวิเคราะห์ผ่านโปรแกรม STATA ได้อย่างถูกต้อง