รายงานการศึกษา ฝึกอบรม ประชุม ดูงาน สัมมนา ปฏิบัติการวิจัย ในประเทศ และต่างประเทศ
(ระยะสั้นไม่เกิน ๙๐ วัน และ ระยะยาวตั้งแต่ ๙๐ วันขึ้นไป)

ส่วนที่ ๑ ข้อมูลทั่วไป

๑.๑ ชื่อ – นามสกุล นางสาวอภิชญา ตั้งจุฑารัตน์
อายุ ๒๘ ปี การศึกษา วิทยาศาสตร์บัณฑิต
ความเชี่ยวชาญเฉพาะด้าน สาขาเวชระเบียน

๑.๒ ตำแหน่ง นักวิชาการเวชสถิติปฏิบัติการ
หน้าที่ความรับผิดชอบ

๑. บริหารจัดการงานเวชระเบียนและสถิติ

๒. ให้รหัสทางการแพทย์ในผู้ป่วยนอกและผู้ป่วยในเพื่อเป็นฐานข้อมูลสุขภาพของโรงพยาบาลและการเบิกจ่ายเงินตามกองทุนต่าง ๆ

๓. ตรวจสอบความสมบูรณ์ของเวชระเบียน

๔. จัดทำสถิติและวิเคราะห์ข้อมูลของโรงพยาบาล

๕. ให้คำปรึกษาและวิเคราะห์ผลสถิติในงานวิจัยของโรงพยาบาล

๖. เลขาคณะกรรมการเวชระเบียน

๗. คณะกรรมการส่งเสริมการวิจัย และคณะกรรมการพัฒนาโครงร่างงานวิจัยของโรงพยาบาลราชพิพัฒน์

๑.๓ ชื่อเรื่อง / หลักสูตร อบรมเชิงปฏิบัติการ Data Science for Healthcare (DASH) วิทยาศาสตร์ข้อมูลทางการแพทย์และสาธารณสุข ครั้งที่ ๑

เพื่อ ☐ ศึกษา ☑ ฝึกอบรม ☐ ประชุม ☐ ดูงาน ☐ สัมมนา ☐ ปฏิบัติการวิจัย

งบประมาณ ☐ เงินงบประมาณกรุงเทพมหานคร ☑ เงินบำรุงโรงพยาบาล

☐ ทุนส่วนตัว

จำนวนเงิน ๒๑,๙๐๐ บาท ระหว่างวันที่ ๑ - ๕ กรกฎาคม ๒๕๖๗

สถานที่ ณ ห้องประชุม Function Six – A ชั้น ๖ โรงแรม Siam@Siam Design Hotel Bangkok

คุณวุฒิ / วุฒิบัตรที่ได้รับ

การเผยแพร่รายงานผลการศึกษา / ฝึกอบรม / ประชุม สัมมนา ผ่านเว็บไซต์สำนักการแพทย์ และกรุงเทพมหานคร

☑ ยินยอม ☐ ไม่ยินยอม

ส่วนที่ ๒ ข้อมูลที่ได้รับจากการศึกษา ฝึกอบรม ประชุม ดูงาน สัมมนา ปฏิบัติการวิจัย (โปรดให้ข้อมูลในเชิงวิชาการ)

๒.๑ วัตถุประสงค์

เพื่อเสริมสร้างความรู้และทักษะด้านการวิเคราะห์ข้อมูลสุขภาพ สามารถตีความและใช้ประโยชน์จากข้อมูลทางการแพทย์ได้อย่างมีความหมายและมีประสิทธิผล เช่น การวิเคราะห์แนวโน้มของโรค การจัดการทรัพยากรสุขภาพอย่างมีประสิทธิภาพ โดยใช้ภาษาอาร์ (Programming R) ภาษาไพธอน (Python) และภาษาเอสคิวแอล (SQL)

๒.๒ เนื้อหา

หัวข้อ ชีวสถิติขั้นพื้นฐานสำหรับวิทยาศาสตร์ข้อมูลสาธารณสุข (Basic Biostatistics for Health Data Science)

๑. ในการพัฒนางานวิจัยทางด้านสาธารณสุขเราจำเป็นที่จะต้องนำความรู้ด้านชีวสถิติ (Biostatistics) หลักการระบาดวิทยา (Epidemiology) และวิทยาศาสตร์ข้อมูล (Data science) มาผนวกเข้าด้วย

๒. หลักการระบาดวิทยา (Epidemiology) ใช้ในการออกแบบการศึกษา เพื่อดูผลลัพธ์สุขภาพปัจจัยเสี่ยงของการเกิดโรคและการแก้ไขปัญหารวมถึงการกำหนดกลุ่มประชากรศึกษาขนาดตัวอย่างและวิธีการเก็บข้อมูลที่เหมาะสม

๓. ชีวสถิติ (Biostatistics) ใช้สำหรับเลือกโมเดลทางสถิติที่เหมาะสมสำหรับงานวิจัยและวิเคราะห์ผลข้อมูล

๔. วิทยาศาสตร์ข้อมูล (Data science) ใช้สำหรับการประมวลและแสดงผลข้อมูลในรูปแบบกราฟหรือแผนภูมิต่าง ๆ เพื่อใช้ในการอธิบายผลลัพธ์

สำหรับการวิเคราะห์ข้อมูลเราจะใช้หลักการ DDPP ซึ่งมีความหมายดังนี้

๑. การวิเคราะห์พรรณนา (Descriptive) เป็นการวิเคราะห์ข้อมูลแบบพื้นฐานเพื่ออธิบายปัญหาของงานวิจัยที่เราศึกษา

๒. การวิเคราะห์หาสาเหตุของปัญหา (Diagnosis) ใช้สำหรับหาเหตุผลที่เกิดขึ้น โดยการทดสอบสมมติฐาน (hypothesis testing)

๓. การทำนายผลลัพธ์ (Predictive) เป็นการทำนายแนวโน้มของปัญหาที่จะเกิดขึ้นในอนาคต โดยใช้โมเดลทางสถิติต่าง ๆ

๔. การวิเคราะห์แบบให้คำแนะนำ (Prescriptive) เป็นการกำหนดถึงสิ่งที่ต้องดำเนินการเพื่อจัดการกับปัญหาในอนาคตหรือใช้เพื่อระบุแนวโน้มเทรนด์ต่าง ๆ ที่จะเกิดขึ้นได้

การวิเคราะห์พรรณนา (Descriptive) ประกอบไปด้วย

ชนิดของข้อมูล

๑. ข้อมูลในรูปแบบตัวเลข ข้อมูลจัดกลุ่ม

๒. มาตรวัดตัวแปร ๔ ระดับ

๒.๑ นามบัญญัติ (Nominal scale)

๒.๒ เรียงลำดับ (Ordinal scale)

๒.๓ อันตราภาค (Interval scale)

๒.๔ อันตราส่วน (Ratio scale)

การแจกแจงของข้อมูล

๑. การแจกแจงแบบปกติ (Normal distribution) เป็นข้อมูลที่มีการกระจายอยู่รอบค่าเฉลี่ยของข้อมูลทางด้านซ้ายและขวาในปริมาณที่ใกล้เคียงกัน

๒. การแจกแจงแบบทวินาม (Binomial distribution) เป็นการแจกแจงข้อมูลที่ไม่ต่อเนื่องจะเกิดผลลัพธ์ได้เพียง ๒ อย่าง เช่น การทดลองโยนเหรียญ หัว ก้อย เป็นต้น

๓. การแจกแจงแบบปัวซง (Poisson distribution) เป็นการทำนายข้อมูลความน่าจะเป็นที่ความสำเร็จจะเกิดขึ้น ณ ช่วงเวลาหนึ่ง เช่น ทำนายโอกาสการเกิดฝนในช่วงหน้าฝน

ค่าวัดตำแหน่งของข้อมูล

๑. ค่าเฉลี่ย (Mean)

๒. ค่ามัธยฐาน (Median)

๓. ค่านิยม (Mode)

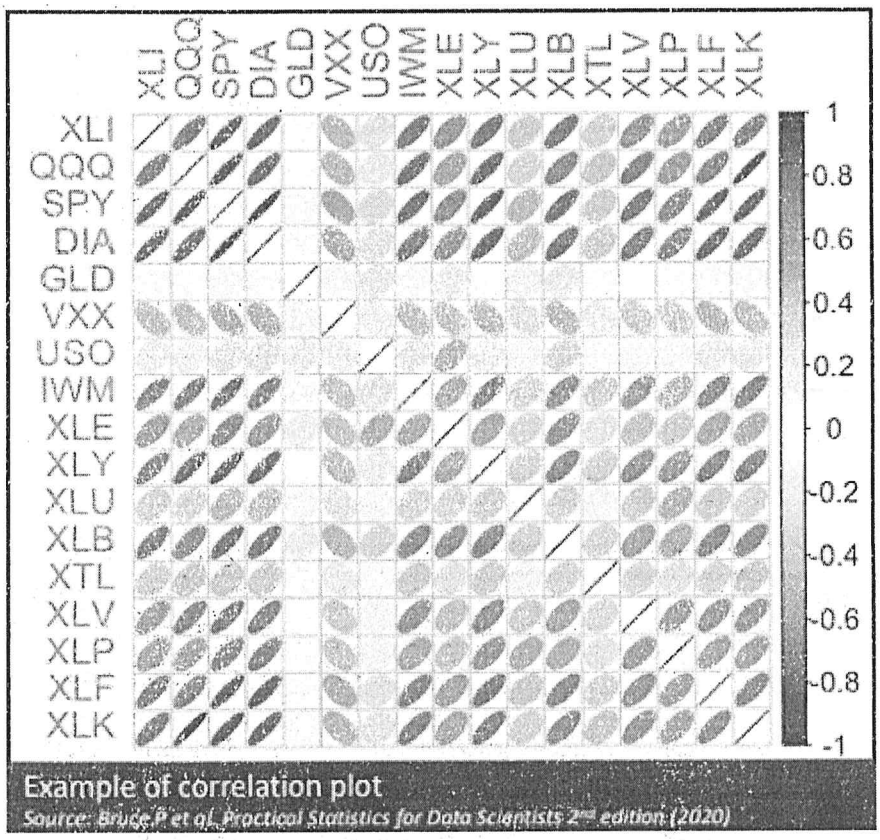
ค่าวัดการกระจายของข้อมูล

๑. ค่าความเบี่ยงเบนมาตรฐาน (Standard deviation)

๒. ค่าควอร์ไทล์ (Quantiles)

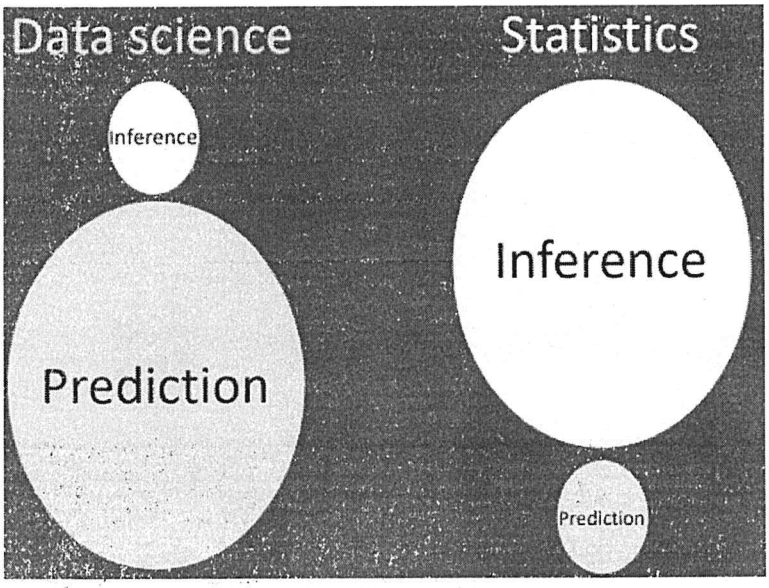
๓. ค่าพิสัย (Range) ค่าสูงสุด ต่ำสุด

กราฟ เช่น กราฟแผนภูมิแท่ง (Bar graph) ฮิสโตแกรม (Histogram) กราฟกล่อง (Box plot) กราฟแสดงค่าสหสัมพันธ์ (Correlation plot) เป็นต้น



Example of correlation plot

Source: Bruce P et al. Practical Statistics for Data Scientists 2nd edition (2020)



ความแตกต่างระหว่างนักวิทยาศาสตร์ข้อมูล (Data science) กับนักสถิติ (Statistics)
(Data science) จะทำหน้าที่สร้างโมเดลที่ใช้ในการพยากรณ์ปัญหาที่จะเกิดขึ้นในอนาคตมากกว่าในขณะที่นักสถิติจะหาข้อสรุปสาเหตุของปัญหามากกว่า ซึ่งในทางสถิติจะเรียกว่าการอนุมาน

สถิติอนุมาน (Inference Statistics)
เป็นการอนุมานลักษณะกลุ่มตัวอย่างของประชากร เพื่อหาสาเหตุของปัญหา โดยการอนุมานจะขึ้นอยู่กับสมมติฐานที่เราตั้งไว้ ซึ่งจะมีการใช้โมเดลทางสถิติ และความน่าจะเป็นมาใช้ในการอนุมาน เช่น การทดสอบพารามิเตอร์โดยใช้สถิติ (Chi-square) หรือการทดสอบแบบไม่ใช่พารามิเตอร์ เช่น (Wilcoxon signed-rank) เป็นต้น

โมเดลพยากรณ์ด้วยสถิติถดถอย (Regression methods for predictive models) ยกตัวอย่าง เช่น

๑. โมเดลถดถอยเชิงเส้นง่าย (Simple linear regression)

๒. โมเดลถดถอยเชิงทั่วไป (Generalized linear model)

๓. โมเดลการถดถอยโลจิสติก (Logistic regression)

๔. โมเดลการถดถอยปัวซอง (Poisson regression)

๕. โมเดลการถดถอยลาโซ (Lasso regression) การปรับแก้แบบถดถอยเชิงเส้น

๖. โมเดลเพิ่มเติมแบบทั่วไป (Generalized additive models)

การวัดความอคติของข้อมูล (Bias-Variance Trade off)

ในการใช้โมเดลทางสถิติเพื่อทำนายผลลัพธ์ที่เกิดขึ้นนั้น จำเป็นจะต้องดูความอคติของข้อมูล เพื่อให้โมเดลมีการทำนายที่แม่นยำ

๑. ถ้าโมเดลที่เราสร้างมีความเรียบง่าย ไม่ซับซ้อน จะทำให้เกิดอคติของข้อมูลสูง แต่ความแปรปรวนของผลลัพธ์ต่ำ

๒. ถ้าโมเดลที่เราสร้างมีความซับซ้อนมาก อคติของข้อมูลน้อย แต่ความแปรปรวนมาก

หัวข้อวิทยาศาสตร์ข้อมูลสำหรับสุขภาพ (introduction to Data Science for Healthcare)

วิทยาศาสตร์ข้อมูล คือ การนำข้อมูลมาใช้ประโยชน์ โดยครอบคลุมตั้งแต่ขั้นตอนการเก็บข้อมูล (Collect) > การจัดการข้อมูล (Manage) > การวิเคราะห์ข้อมูล (Analyze) > ไปจนถึงขั้นตอนการนำข้อมูลมาช่วยตัดสินใจ (Decision) ซึ่งในสายสุขภาพวิยาศาสตร์ข้อมูล (Data science) จำเป็นต้องมีความรู้ทางด้านการแพทย์สถิติและการเขียนโปรแกรมเพื่อที่จะสามารถสร้างโมเดลทำนายข้อมูลได้

ความสำคัญของวิทยาศาสตร์ข้อมูล (Data science) กับสายสุขภาพมีหลายอย่าง เช่น เป็นข้อมูลสำหรับใช้พัฒนาการป้องกันดูแลรักษาโรคต่าง ๆ ใช้ในการทำนายการเกิดความเสี่ยงของการเกิดโรค การระบาดทดสอบประสิทธิภาพของกระบวนการรักษาการพัฒนาเครื่องมือการดูแลรักษาโรคหรือการคิดค้นยาใหม่ ๆ เป็นต้น

การเก็บข้อมูลทางสุขภาพและการจัดการข้อมูล มีวิธีการดังนี้

๑. ข้อมูลเวชระเบียนซึ่งอาจอยู่ในรูปแบบอิเล็กทรอนิกส์หรือเป็นกระดาษ

๒. ข้อมูลพันธุกรรมใช้สำหรับดูความเสี่ยงของการเกิดโรคซึ่งอาจเกิดจากพันธุกรรม การตอบสนองต่อยาบางชนิด เป็นต้น

๓. ข้อมูลในแอพลิเคชันสุขภาพต่าง ๆ เช่น แอปเปิ้ลวอช (Apple watch) ที่บันทึกอัตราการเต้นของหัวใจ

๔. ข้อมูลรูปภาพ เช่น ผล X-ray MRI CT scan

๕. ข้อมูลโครงสร้างทางสังคม เช่น ที่อยู่อาศัย อาชีพ อายุ เป็นต้น

หลักธรรมาภิบาลและจริยธรรมข้อมูล (Data Givernance and Ethical Consideration)

การเก็บข้อมูลทางสุขภาพต้องคำนึงถึงความปลอดภัยของข้อมูลและจริยธรรมในการศึกษาวิจัย ซึ่งประกอบไปด้วย

๑. ความเป็นส่วนตัวและความปลอดภัย (Privacy and Security) ต้องมีวิธีการเก็บรักษาข้อมูลที่ปลอดภัย เช่น ในอเมริการจะมีการรักษาความปลอดภัยของข้อมูลตามข้อกำหนดของ (HIPAA)

๒. การกำหนดนโยบายการใช้ข้อมูล (Data stewardship) ว่ามีวิธีการเก็บ ดูแลรักษา และใช้ข้อมูลอย่างไรให้มีความปลอดภัย และไม่ส่งผลเสียต่อเจ้าของข้อมูลซึ่งก็คือผู้ป่วย

๓. จริยธรรมการใช้ข้อมูล (Ethical Use) การนำข้อมูลสุขภาพมาศึกษาวิจัยต้องคำนึงถึงประโยชน์ที่ผู้ป่วยจะได้รับเป็นสำคัญ

หัวข้อ การเขียนโปรแกรมภาษา R เบื้องต้น (Introduction to R)

ภาษา R เป็นภาษาทางคอมพิวเตอร์ที่ใช้สำหรับการสร้างกราฟเพื่อใช้อธิบายข้อมูลทางสถิติ ซึ่งถูกพัฒนามาจากภาษา S สามารถติดตั้งผ่านเว็บไซต์ของ http://cran.r-project.org/ หลังจากที่ติดตั้งและเราสามารถเขียนโปรแกรมภาษา R ผ่านโปรแกรมอาร์สตูดิโอ (RStudio) ได้แต่ในการอบรมครั้งนี้จะเขียนโปรแกรมผ่านจูปีเตอร์โน๊ตบุ๊ค (Jupyter notebook)

สำหรับภาษา R จะมีชุดคำสั่งต่าง ๆ ที่สนับสนุนในการสร้างกราฟและแผนภูมิในรูปแบบต่าง ๆ เช่น

๑. (ggplot2) ใช้สำหรับสร้างกราฟโดยอาศัยแนวคิดของชั้นข้อมูล (data layering) และกฎการกำหนดลักษณะกราฟ (aesthetic mapping)

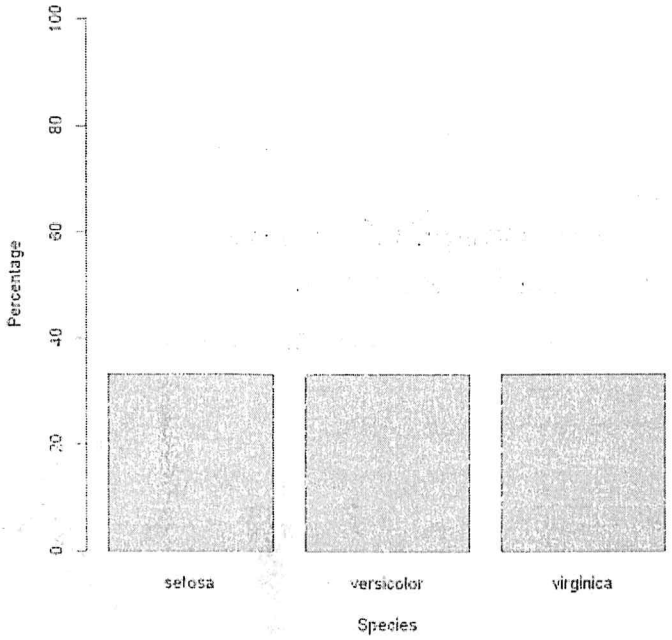
๒. (tidyr package) ใช้สำหรับจัดระเบียบข้อมูลให้เป็นระเบียบมากขึ้น

๓. (rmarkdown) ใช้สำหรับรายงานผลข้อมูล

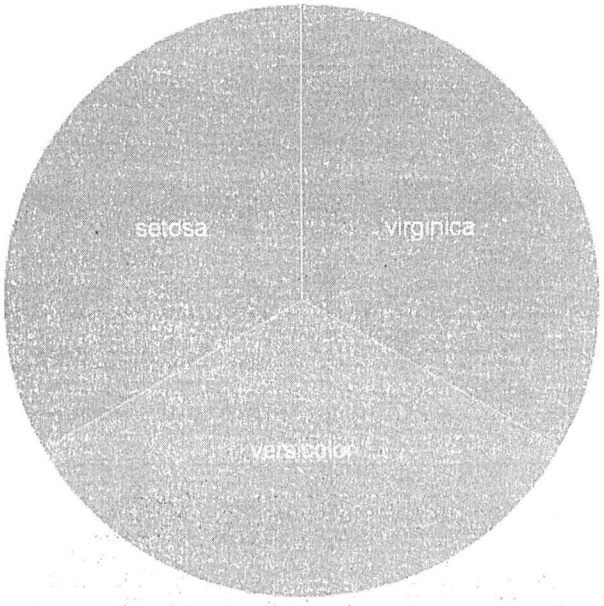
๔. (shiny) ใช้สำหรับการพัฒนาเว็บและแอพพลิเคชัน

หัวข้อ การเขียนโปรแกรมภาษา R การสร้างกราฟอธิบายข้อมูล (Data Visualization with R)

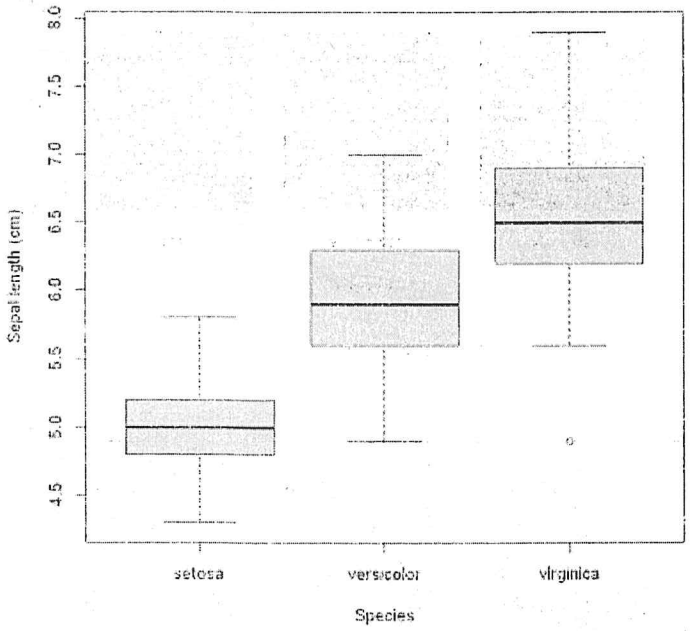
เราสามารถใช้ภาษา R ในการสร้างกราฟในรูปแบบต่าง ๆ ได้ โดยใช้ชุดคำสั่ง ggplot2 ตัวอย่างกราฟที่สามารถสร้างได้



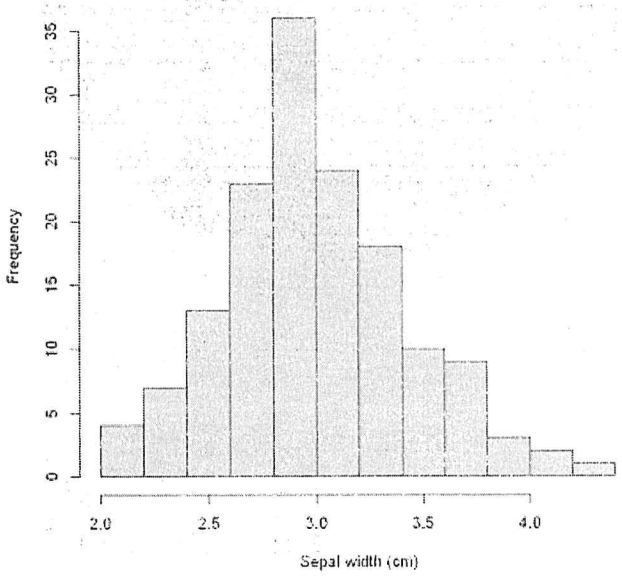
แผนภูมิแท่ง (Bar graph)



แผนภูมิวงกลม (Pie chart)



กราฟกล่อง (Box plot)



ฮิสโตแกรม (Histogram)

หัวข้อการเขียนโปรแกรม R สำหรับการสร้างการเรียนรู้ของเครื่อง (Machine learning)

การเรียนรู้ของเครื่อง (Machine Learning) คือ การทำให้ระบบคอมพิวเตอร์เรียนรู้ได้ด้วยตนเอง โดยใช้ข้อมูลที่เราใส่เข้าไป เพื่อให้โปรแกรมสามารถตอบคำถามในสิ่งที่เราใส่ข้อมูลได้ในอนาคตว่าผลลัพธ์ที่เกิดขึ้นเป็นอย่างไร เช่น การแยกแยะฟิล์มเอ็กซ์เรย์ปอดว่าปกติหรือเกิดปอดบวม เป็นต้น

ขั้นตอนการสร้างการเรียนรู้ของเครื่อง (Machine Learning) ด้วยโปรแกรม R มีดังนี้

๑. การจัดการข้อมูล (Data management) ประกอบไปด้วย
การนำเข้าข้อมูล (Import Data) โดยใช้คำสั่ง setwd และ read.csv

๑.๑ การสำรวจข้อมูลทั้งหมด (Explore Data) โดยใช้คำสั่ง str() เพื่อดูว่าข้อมูลในแต่ละแถวหน้าตาเป็นอย่างไรและเป็นข้อมูลชนิดไหน เช่น ข้อมูลตัวเลข ตัวอักษร ตัวเลขทศนิยม เป็นต้น

๑.๒ การแปลงข้อมูล (Transform Data) โดยใช้คำสั่งไทดี้แพคเกจ (tidy package) เพื่อแปลงข้อมูลให้พร้อมใช้งานตามวัตถุประสงค์ที่เราต้องการ เช่น เพศหญิง = ๑ เพศชาย = ๒ เลขดังกล่าวเป็นตัวอักษรไว้แบ่งกลุ่มไม่ใช้เป็นตัวเลขที่ไว้สำหรับคำนวณเราต้องใช้คำสั่งเพื่อบอกโปรแกรมตามที่เราต้องการหรือการจัดการ ข้อมูลที่มีค่าผิดปกติ ซึ่งอาจเกิดจากการคีย์ข้อมูลผิดพลาด เช่น ผู้ป่วยอายุ ๓๐๐ ปี โดยการใช้คำสั่งปรับ ข้อมูลให้อยู่ในค่าปกติอาจจะเปลี่ยนเป็นข้อมูลค่าเฉลี่ยหรือวิธีการอื่น ๆ ขึ้นอยู่กับข้อมูลและหลักการทางชีวสถิติ

๒. การแบ่งข้อมูล (Data splitting) ในการสร้างโมเดลเพื่อทำนายข้อมูลจำเป็นจะต้องแบ่งข้อมูลออกเป็น ๒ กลุ่ม ใช้สำหรับฝึก (Train) เพื่อให้โมเดลจดจำข้อมูลที่เราป้อนเข้าไปกับอีกกลุ่มใช้สำหรับ Test เพื่อทดสอบหาความแม่นยำของโมเดลที่เราสร้าง

๒.๒ การแบ่งข้อมูลเท่าๆกัน (Cross validation) ในการแบ่งข้อมูลสำหรับฝึก (Train) โมเดลเราจำเป็นต้องมีเครื่องมือที่ช่วยตัดสินใจว่าควรแบ่งข้อมูลส่วนใดไปทำการฝึกฝน (Trainging) โดยการทำ cross validation คือ การแบ่งข้อมูลจำนวนหนึ่งที่เท่า ๆ กัน ซึ่งมาจากการสุ่มเพื่อให้ข้อมูลการกระจายที่เท่า ๆ กันแล้วไปสร้างกราฟถอยเชิงเส้นเพื่อดูว่าข้อมูลชุดไหนจะทำนายผลได้แม่นที่สุดจึงเลือกข้อมูลชุดนั้นมาฝึกฝน (Training)

๓. การวัดประสิทธิภาพของโมเดล (Evaluate model performance with test data) โดยใช้เส้นโค้งประสิทธิภาพของโมเดล (ROC-AUC Curve) และเมทริกซ์แห่งความสับสน (Confusion Matrix) ซึ่งใช้วัดประสิทธิภาพความแม่นยำของการทำนายผลของโมเดล

หัวข้อแนะนำการใช้การเรียนรู้เชิงลึก Deep Learning (Introduction to Deep Learning)

การเรียนรู้เชิงลึก (Deep Learning) เป็นเทคโนโลยีที่พัฒนาต่อยอดมาจากการเรียนรู้ของเครื่อง (Machine Learning) โดยการเรียนรู้เชิงลึก (Deep Learning) จะเรียนรู้ประมวลผลและตัดสินใจจากข้อมูลที่ได้รับแต่ไม่ต้องอาศัยการแทรกแซงจากมนุษย์เพราะการเรียนรู้เชิงลึก (Deep Learning) สามารถปรับปรุงผลลัพธ์การตัดสินใจได้ด้วยตัวมันเอง ซึ่งอาศัยสถาปัตยกรรมคอมพิวเตอร์ที่เรียกว่าโครงข่ายประสาทเทียม (Neural Network)

โดยโครงข่ายประสาทเทียม (Neural Network) ที่ได้รับแรงบันดาลใจมาจากโครงสร้างระบบประสาทของมนุษย์ ซึ่งอัลกอริทึมของโครงข่ายประสาทเทียม (Neural Network) จะมีลักษณะเป็นปุ่ม (Node) เรียงกัน ๔ ชั้น คือ ชั้นขาเข้า (Input Layer), ชั้นที่ซ่อนไว้ (Hidden Layer), ชั้นขาออก (Output Layer) และ การทำนายผล (Prediction)

๑. ชั้นขาเข้า (Input Layer) คือ ชั้นข้อมูลขาเข้าอาจเป็นข้อมูลที่มีโครงสร้างหรือไม่มีโครงสร้างก็ได้ โดยข้อมูลเหล่านี้จะถูกเปลี่ยนให้อยู่ในรูปแบบตัวแปรที่ต่างกัน

๒. ชั้นที่ซ่อนไว้ (Hidden Layer) คือ ชั้นประมวลผล โดยในชั้นที่ซ่อนไว้ (Hidden Layer) จะมีชั้นของหน่วยประมวลผลอยู่มากกว่า ๑ ชั้น ขึ้นอยู่กับโครงสร้างของสถาปัตยกรรมเครือข่าย (Network Architecture)

๓. ชั้นขาออก (Output Layer) คือ ชั้นที่นำผลลัพธ์จาก ชั้นที่ซ่อนไว้ (Hidden Layer) มาประมวลในฟังก์ชันประมวลผล (Activation Function) อีกรอบหนึ่ง

๔. การทำนายผล (Prediction) คือ ชั้นที่นำเอาผลลัพธ์จาก ชั้นขาออก (Output Layer) มาตัดสินใจในฟังก์ชันการตัดสินใจ (Decision Function)

ประโยชน์ของการเรียนรู้เชิงลึก (Deep Learning) ที่ใช้ในสาขาต่างๆ เช่น

๑. ด้านสุขภาพใช้สำหรับวินิจัยโรคจากภาพทางรังสีหรือผลพยาธิวิทยา การหาวิธีการผลิตสูตรยาใหม่ ๆ โดยใช้เทคโนโลยีแทนการใช้วิธีการแบบดั้งเดิม

๒. ด้านวิศวกรรมยานยนต์ ใช้ในการสร้างเซนเซอร์ กล้อง หรือเรดาห์ที่ช่วยอำนวยความสะดวกในการขับรถหรือสร้างเทคโนโลยีป้องกันการเกิดอุบัติเหตุ

๓. ด้านเศรษฐศาสตร์ใช้ในดูแนวโน้มของราคาสินค้า การป้องกันการทุจริตทางการเงิน

๔. ด้านธุรกิจค้าปลีกใช้ศึกษาพฤติกรรมการซื้อสินค้าของผู้บริโภค แนวโน้มของยอดขายการประมาณการสินค้าในคลัง

๕. การประมวลผลภาษาธรรมชาติ (Natural language processing) เช่น การพัฒนาแชทบอท Chatbot เพื่อโต้ตอบระหว่างมนุษย์กับเครื่องจักรหรือการสร้างแอพลิเคชันสำหรับการแปลภาษาต่าง ๆ

๖. ด้านธุรกิจบันเทิงการสร้างเนื้อหาแนะนำสิ่งที่ผู้บริโภคสนใจซึ่งพบได้ในแพลตฟอร์มของเน็ตฟลิกซ์ (Netflix) หรือยูทูป (Youtube) การสร้างคอมพิวเตอร์กราฟิกส์ในภาพยนต์หรือเกม

๗. ด้านอุตสาหกรรมการสร้างเทคโนโลยีควบคุมคุณภาพการผลิต การสร้างระบบบำรุงรักษาเชิงพยากรณ์ สำหรับคาดการณ์ความเสียหายของเครื่องจักรและอุปกรณ์ต่าง ๆ

๘. ด้านพลังงานการคาดการณ์ความต้องการในการใช้พลังงาน การปรับสมดุลทรัพยากร

๙. ด้านความปลอดภัยทางไซเบอร์ การสร้างระบบตรวจจับการบุกรุก การป้องกันการรั่วไหลของข้อมูล

๑๐. ด้านเกษตรกรรมการทำนายการเกิดโรคของเกษตรกร การใช้เทคโนโลยีช่วยในการเก็บเกี่ยวผลผลิต

๒.๓ ประโยชน์ที่ได้รับ

๒.๓.๑ ☑ ต่อตนเอง ได้รับความรู้และเทคนิคการจัดการข้อมูล การวิเคราะห์ข้อมูลแขนงใหม่ ๆ ที่ช่วยทำให้การทำงานด้านการจัดการข้อมูลของโรงพยาบาลได้อย่างรวดเร็ว มีประสิทธิภาพ และมีความถูกต้องตามหลักสถิติมากขึ้น

๒.๓.๒ ☑ ต่อหน่วยงาน สามารถนำความรู้ที่ได้มาปรับใช้กับงานประจำ เพื่อลดระยะเวลาการทำงาน วิเคราะห์ข้อมูลในเชิงลึกได้อย่างมากขึ้น เพราะปัจจุบันโรงพยาบาลเป็นโรงเรียนแพทย์ ซึ่งในอนาคตต้องมีการใช้เทคนิคต่าง ๆ สำหรับวิเคราะห์ข้อมูลทางการแพทย์ที่มีความซับซ้อน

๒.๓.๓ ☑ อื่น ๆ (ระบุ) สามารถนำมาปรับใช้กับการวิเคราะห์ข้อมูลของโครงการตรวจสุขภาพล้านคน เพื่อศึกษาข้อมูลสุขภาพของประชาชนที่อาศัยในกรุงเทพมหานครได้ และนำผลการวิเคราะห์ข้อมูลมาใช้ในการวางแผนส่งเสริมสุขภาพของประชาชนในกรุงเทพมหานคร

ส่วนที่ ๓ ปัญหาและอุปสรรค

๓.๑ ☑ **การปรับปรุง** ปรับเนื้อหาให้เหมาะสมกับกลุ่มเป้าหมาย เนื่องจากมีเนื้อหาค่อนข้างเฉพาะทางค่อนข้างมากและไม่มีการปูเนื้อหาพื้นฐานก่อนทำให้เข้าใจได้ยาก

๓.๒ ☑ **การพัฒนา** นำความรู้มาใช้ในการวิเคราะห์ข้อมูลทางการแพทย์ในโรงพยาบาลที่มีความเฉพาะเจาะจงและมีประสิทธิภาพมากขึ้น

ส่วนที่ ๔ ข้อคิดเห็นและข้อเสนอแนะ เนื่องจากเป็นการจัดอบรมครั้งแรกของหน่วยงาน การอบรมจึงมีความติดขัดและมีความไม่เข้าใจเนื้อหาในบางส่วน เนื่องจากมีการใช้ศัพท์เทคนิคค่อนข้างมาก รวมถึงการใช้ความรู้ทางด้านชีวสถิติ ซึ่งผู้อบรมจำเป็นต้องมีความรู้ในระดับหนึ่งจึงจะมีความเข้าใจอย่างชัดเจน อาจต้องปรับเนื้อหาและวิธีการอบรมให้เหมาะสมกับกลุ่มเป้าหมาย เพื่อให้ผู้อบรมได้รับความรู้และประโยชน์ครบถ้วน

ลงชื่อ.....ผู้รายงาน

(นางสาวอภิชญา ตั้งจุฑารัตน์)
นักวิชาการเวชสถิติปฏิบัติการ

ส่วนที่ ๕ ความคิดเห็นของผู้บังคับบัญชา .....

เห็นควรนำความรู้ที่ได้ไปปฏิบัติงาน

.....

.....

ลงชื่อ.....หัวหน้าส่วนราชการ

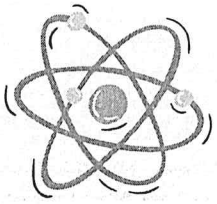
(นางสาววรรณิกา แสงสุริย์)
รองผู้อำนวยการโรงพยาบาล ฝ่ายการแพทย์
รักษาการในตำแหน่งผู้อำนวยการโรงพยาบาลราชพิพัฒน์

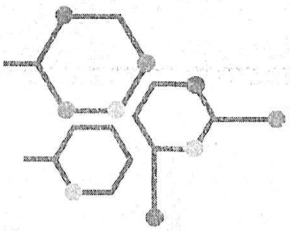
QR Code รายงานอบรม

อบรมเชิงปฏิบัติการวิทยาศาสตร์ข้อมูล ทางการแพทย์และสาธารณสุข ครั้งที่ 1
(DATA SCIENCE FOR HEALTH CARE)

วิทยาศาสตร์ข้อมูล (DATA SCIENCE)

- การนำข้อมูลมาใช้ประโยชน์ โดยครอบคลุมตั้งแต่ขั้นตอนการเก็บข้อมูล (Collect) >การจัดการข้อมูล (Manage) > การวิเคราะห์ข้อมูล (Analyze) > ไปจนถึงขั้นตอนการนำข้อมูลมาช่วยตัดสินใจ (Decision) ซึ่งในสายสุขภาพ นัก Data science จำเป็นต้องมีความรู้ทางด้านการแพทย์ สถิติ และการเขียนโปรแกรม เพื่อที่จะสามารถสร้างโมเดลทำนายข้อมูลได้



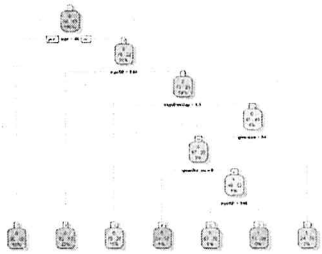


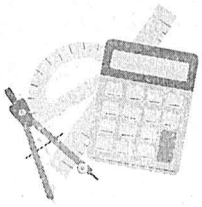
การเรียนรู้ของเครื่องจักร (MECHINE LEARNING)

- เป็นส่วนหนึ่งของ Data Science คือ การทำให้ระบบคอมพิวเตอร์เรียนรู้ได้ด้วยตนเอง โดยใช้ข้อมูลที่เราใส่เข้าไป เพื่อให้โปรแกรมสามารถตอบคำถามในสิ่งที่เราใส่ข้อมูลได้ในอนาคตว่าผลลัพธ์ที่เกิดขึ้นเป็นอย่างไร เช่น การแยกแยะฟิล์มเอ็กซ์เรย์ปอดว่าปกติหรือเกิดปอดบวม เป็นต้น

ขั้นตอนการสร้างการเรียนรู้ของเครื่องจักร (MECHINE LEARNING) ด้วยภาษา R

- การจัดการข้อมูล (Data management) ประกอบไปด้วย
  - การนำเข้าข้อมูล (Import Data) โดยใช้คำสั่ง setwd และ read.csv
  - การสำรวจข้อมูลทั้งหมด (Explore Data) โดยใช้คำสั่ง str() เพื่อดูว่าข้อมูลในแต่ละแถวหน้าตาเป็นอย่างไร และเป็นข้อมูลชนิดไหน
  - การแปลงข้อมูล (Transform Data) โดยใช้คำสั่ง tidy package เพื่อแปลงข้อมูลให้พร้อมใช้
- การแบ่งข้อมูล (Data splitting) ในการสร้างโมเดลเพื่อทำนายข้อมูล จำเป็นจะต้องแบ่งข้อมูลออกเป็น 2 กลุ่ม ใช้สำหรับ ฝึก (Train) เพื่อให้โมเดลจดจำข้อมูลที่เราป้อนเข้าไป กับอีกกลุ่มใช้สำหรับทดสอบ (Test) เพื่อทดสอบหาความแม่นยำของโมเดลที่เราสร้าง
- การวัดประสิทธิภาพของโมเดล (Evaluate model performance with test data) โดยใช้เส้นโค้งลักษณะการทำงานของตัวรับ (ROC-AUC Curve) และ เมทริกซ์ข้อผิดพลาด (Confusion Matrix) ซึ่งใช้วัดประสิทธิภาพความแม่นยำของการทำนายผลของโมเดล





การเรียนรู้เชิงลึก (DEEP LEARNING)

- เป็นเทคโนโลยีที่พัฒนาต่อยอดมาจากการเรียนรู้ของเครื่องจักร (Machine Learning) โดยการเรียนรู้เชิงลึก (Deep Learning) จะเรียนรู้ ประมวลผลและตัดสินใจจากข้อมูลที่ได้รับ แต่ไม่ต้องอาศัยการแทรกแซงจากมนุษย์ เพราะ การเรียนรู้เชิงลึก (Deep Learning) สามารถปรับปรุงผลลัพธ์การตัดสินใจได้ด้วยตัวมันเอง ซึ่งอาศัยสถาปัตยกรรมคอมพิวเตอร์ที่เรียกว่า โครงข่ายประสาทเทียม (Neural Network)

ประโยชน์ที่ได้รับ

- สามารถสร้างโมเดลการพยากรณ์โรค เพื่อใช้พัฒนาการป้องกัน ดูแลรักษาโรคต่างๆ ใช้ในการทำนายการเกิดความเสี่ยงของการเกิดโรค การระบาด ทดสอบประสิทธิภาพของกระบวนการรักษา การพัฒนาเครื่องมือการดูแลรักษาโรค หรือการคิดค้นยาใหม่ๆ เป็นต้น

น.ส.อภิชญา ตั้งจุฑารัตน์
นักวิชาการเวชสถิติปฏิบัติการ