



## แนวทางการเลือกใช้โปรแกรมประยุกต์ทางสถิติ

เรียงเรียงใจ... วิชชา อัครพงษ์พงศ์

การวิเคราะห์ข้อมูลทางสถิตินั้น เป็นการประมวลผลข้อมูลที่มีปริมาณมาก ดังนั้นการนำคอมพิวเตอร์มาช่วยในการวิเคราะห์ข้อมูล จะช่วยประหยัดเวลาและค่าใช้จ่าย มีความสะดวกมากกว่าที่จะคำนวณด้วยเครื่องคิดเลขทั่วๆ ไป และในการใช้คอมพิวเตอร์ ส่วนใหญ่นักวิจัยนิยมที่จะใช้โปรแกรมสำเร็จรูป (Package Program) มากกว่าที่จะเขียนโปรแกรมขึ้นมาเอง โปรแกรมสำเร็จรูปที่นำมาใช้วิเคราะห์ข้อมูลจะเป็นโปรแกรมสำเร็จรูปทางสถิติโดยตรง แต่ในปัจจุบันมีโปรแกรมสำเร็จรูปที่ทำงานด้านอื่นๆ สามารถนำมาวิเคราะห์ข้อมูลทางสถิติได้ เช่น โปรแกรม EXCEL ซึ่งเป็นโปรแกรมประเภทกระดาษทศ (Spread Sheet) แต่เนื่องจากโปรแกรมประเภทนี้ถูกสร้างขึ้นมาสำหรับงานบางอย่างที่ไม่ใช่สถิติโดยตรง ดังนั้นการวิเคราะห์ข้อมูลด้วยวิธีการทางสถิติบางอย่าง อาจทำโปรแกรมประเภทนี้ไม่ได้ หรือทำได้แต่อาจจะให้รายละเอียดน้อยกว่า ดังนั้นผู้วิจัยส่วนใหญ่จึงนิยมใช้โปรแกรมสำเร็จรูปทางสถิติมากกว่า แต่อย่างไรก็ตาม โปรแกรม EXCEL จะมีความสามารถในด้านการนำเสนอข้อมูลด้วยกราฟหรือแผนภูมิ

### ประเภทของโปรแกรม

การวิเคราะห์ข้อมูลสถิติที่มีจำนวนมาก ควรเลือกคอมพิวเตอร์ช่วยในการวิเคราะห์ ซึ่งต้องอาศัยโปรแกรมหรือซอฟต์แวร์มาช่วย

โปรแกรมจำแนกได้เป็น 2 ประเภท คือ

**1. โปรแกรมที่ผู้ใช้เขียนขึ้นเอง (User's Written Program)** เป็นโปรแกรมที่ผู้ใช้เขียนสั่งให้คอมพิวเตอร์ทำงานได้ตามความต้องการ หรือ ตรงตามวัตถุประสงค์

และเหมาะสมด้วยภาษาคอมพิวเตอร์ โดยใช้เทคนิคและความชำนาญของผู้เขียนโปรแกรมด้วยภาษาคอมพิวเตอร์ที่นิยมใช้ เช่น ภาษาเบสิก ภาษาซี เป็นต้น

**2. โปรแกรมสำเร็จรูป (Package Program)** เป็นโปรแกรมที่มีผู้เขียนได้เขียนไว้เรียบร้อยแล้ว โปรแกรมสำเร็จรูปจะให้ความสะดวกในการใช้งานมาก โดยที่ผู้ใช้ไม่จำเป็นต้องมีความรู้ทางด้านคอมพิวเตอร์มากนัก เพียงแต่เรียนรู้วิธีการใช้งาน ซึ่งส่วนมากจะมีคำอธิบายการใช้โปรแกรมมาให้ และในขณะที่ทำงานก็สามารถขอรายละเอียดเพิ่มเติมได้ตลอดเวลาในการใช้โปรแกรมสำเร็จรูป

### โปรแกรมสำเร็จรูปทางสถิติ (Statistical Package)

โปรแกรมสำเร็จรูปทางสถิติ เป็นโปรแกรมสำหรับการวิเคราะห์ข้อมูลทางสถิติโดยตรง ในระยะแรกโปรแกรมสำเร็จรูปทางสถิติที่ใช้กับคอมพิวเตอร์ มีไม่มากนักและมีใช้เฉพาะเครื่องคอมพิวเตอร์ขนาดใหญ่ ได้แก่ ซุปเปอร์คอมพิวเตอร์ (Super Computer) หรือ เมนเฟรม (Mainframe computer) โปรแกรมที่มีชื่อเสียงและเป็นที่ยอมรับกันดี คือโปรแกรมสำเร็จรูป SPSS (Statistical Package for the Social Sciences) สร้างขึ้นในประเทศสหรัฐอเมริกา โปรแกรมสำเร็จรูป SAS (Statistical Analysis System) ซึ่งพัฒนาโดยบริษัท SAS Institute INC. และโปรแกรมสำเร็จรูป BMDP (Biomedical Computer Program) เป็นต้น

ในปัจจุบันนี้ได้มีความก้าวหน้าทางเทคโนโลยีด้านไมโครคอมพิวเตอร์ ทำให้มีขีดความสามารถในการบันทึกข้อมูลจำนวนมากขึ้น มีหน่วยความจำสูงขึ้น ทำให้การ



ประมวลผลข้อมูลมีประสิทธิภาพมากขึ้น ดังนั้นนักวิจัยจึงหันมาใช้ไมโครคอมพิวเตอร์มากขึ้น เพราะมีความสะดวกกว่า จึงทำให้มีผู้ผลิตโปรแกรมสำเร็จรูปทางสถิติ สำหรับเครื่องไมโครคอมพิวเตอร์มากขึ้น ซึ่งโปรแกรม SPSS SAS และ BMDP ก็ได้ถูกพัฒนามาใช้กับเครื่องไมโครคอมพิวเตอร์เช่นกัน

### โปรแกรมสำเร็จรูปทางสถิติ

จำแนกตามองค์ประกอบของโปรแกรมได้เป็น 2 ประเภทใหญ่ ๆ คือ

1. โปรแกรมเบ็ดเสร็จ เป็นโปรแกรมที่เขียนขึ้นสำหรับวิเคราะห์ข้อมูลทางสถิติแบบอเนกประสงค์ มีความสามารถในการวิเคราะห์ข้อมูลเชิงสถิติทุกประเภท โดยทั่วไปแล้วจะเป็นโปรแกรมที่มีขนาดใหญ่ เช่น โปรแกรม SPSS, R, SAS, BMDP, GENSTAT, MINITAB และ STATA เป็นต้น

2. โปรแกรมที่เน้นเฉพาะด้าน เป็นโปรแกรมที่เขียนขึ้นเพื่อเน้นการวิเคราะห์เชิงสถิติประเภทใดประเภทหนึ่งโดยเฉพาะ เช่น

- โปรแกรมสำหรับกำหนดการเชิงเส้น (Linear Programming) เช่น โปรแกรม LINDO โปรแกรม GINO โปรแกรม LINGO

- โปรแกรมที่เน้นการพยากรณ์ของข้อมูลอนุกรมเวลา เช่น โปรแกรม TSP (Time Series Program)

- โปรแกรมที่เน้นทางด้าน Statistical Modeling เช่น โปรแกรม GLIM

- โปรแกรมทางด้านระบาดวิทยา เช่น โปรแกรม Epidata

ความสามารถในการวิเคราะห์ข้อมูลหรือวิเคราะห์สถิติที่เป็นพื้นฐานทั่วไป (Basic Statistics) ทั้งสถิติเชิงพรรณนาและสถิติอ้างอิง ทุกโปรแกรมจะมีเหมือนๆ กัน แต่ในส่วนที่เป็น advance statistics แต่ละโปรแกรมจะมีความแตกต่างกันไป ตามลักษณะเฉพาะ แม้แต่การ

วิเคราะห์สถิติแบบเดียวกัน ผลลัพธ์ที่ได้จากโปรแกรมเหล่านั้นอาจมีความสมบูรณ์ตามหลักสถิติแตกต่างกันไป

ลักษณะการใช้งานของโปรแกรมทางสถิติอาจแบ่งออกได้เป็น 2 ประเภทใหญ่ ๆ คือ

1. โปรแกรมชนิดสั่งให้ทำงานด้วยคำสั่ง (Command Driven) ก่อนที่จะใช้โปรแกรมชนิดนี้ ผู้ใช้จะต้องศึกษาคำสั่ง (Commands) ต่างๆ ให้เข้าใจเสียก่อน จากนั้นจึงเขียนชุดคำสั่งเพื่อทำการวิเคราะห์ข้อมูล ตัวอย่างของโปรแกรมประเภทนี้ เช่น SPSS/PC, R, MINITAB, SAS และ STATA เป็นต้น

2. โปรแกรมชนิดสั่งให้ทำงานโดยอาศัยเมนู (Menu Driven) โปรแกรมประเภทนี้จะใช้งานง่ายกว่าโปรแกรมชนิดคำสั่ง จึงเหมาะสำหรับผู้เริ่มใช้โปรแกรมสำเร็จรูปทางสถิติ ทั้งนี้เนื่องจาก โปรแกรมประเภทนี้จะแสดงรายการต่างๆ ออกทางหน้าจอ แล้วให้ผู้ใช้เลือกว่าต้องการใช้รายการใด เช่น โปรแกรม STATPACK, SPSS for Windows

โปรแกรมทางสถิติ เมื่อพิจารณาจากที่มาอาจแบ่งออกได้เป็น 2 ประเภทคือ

1. Free software ได้แก่ R, Epi Info

2. โปรแกรมลิขสิทธิ์ ได้แก่ SAS, SPSS, STATA

### โปรแกรมสำเร็จรูป R

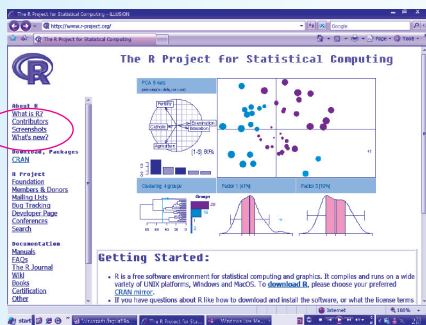
สร้างโดยนักสถิติชาวนิวซีแลนด์ชื่อ Ross Ihaka และ Robert Gentleman โดยเห็นว่าโปรแกรมสถิติที่ใช้กันอยู่นั้นส่วนใหญ่จะมีราคาแพง จึงลองเขียนโปรแกรมขึ้นมาใช้เอง และเผยแพร่โค้ดโปรแกรมออกมาและเนื่องจากเป็นโปรแกรมที่สามารถดาวน์โหลดมาใช้ได้ฟรีไม่ต้องเสียค่าลิขสิทธิ์ ก็อาจจะไม่มีการรับประกันใดๆ อย่างไรก็ตามโปรแกรมนี้เป็นโปรแกรมที่ได้รับการยอมรับจากนักสถิติทั่วโลกเลยทีเดียว (Paper หลายฉบับใช้โปรแกรมนี้ในการคำนวณ) มักใช้ในงานสถิติและงานด้านคณิตศาสตร์ มีลักษณะเป็น Object Oriented Language ทำงานแบบ Interactive ผู้ใช้สามารถเขียนฟังก์ชันที่ต้องการสำหรับใช้



งานเฉพาะด้านของตนได้ และสามารถสร้างกราฟฟิคต่างๆ รวมทั้งกราฟฟิคใน 3 มิติ ได้ดี ข้อมูลที่จัดเก็บด้วยโปรแกรมสำเร็จรูปอื่น เช่น EXCEL, SPSS, STATA, EPI DATA, หรือ ASCII file สามารถนำมาใช้กับซอฟต์แวร์ R ได้ด้วย

### ขั้นตอนการ download โปรแกรม R

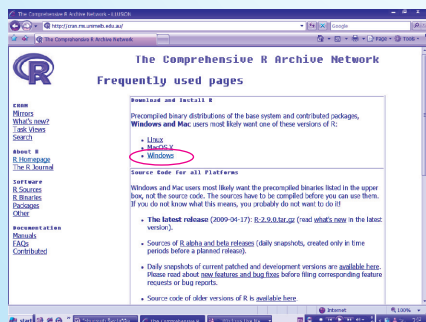
1. เข้า <http://www.r-project.org/> เลือก Download ที่หัวข้อ CRAN



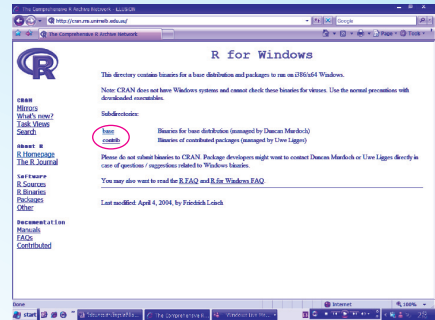
2. เลือกสถานที่ที่ใกล้ประเทศไทยที่สุดเพื่อจะได้โหลดเร็วๆ (เมืองไทยก็มี) แต่แนะนำ Australia เพราะเมืองไทยมักไม่อัปเดต



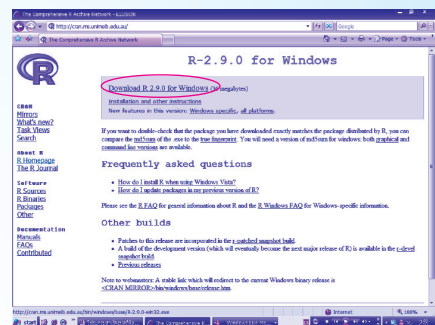
3. เลือกวินโดวส์ ถ้าคุณใช้วินโดวส์



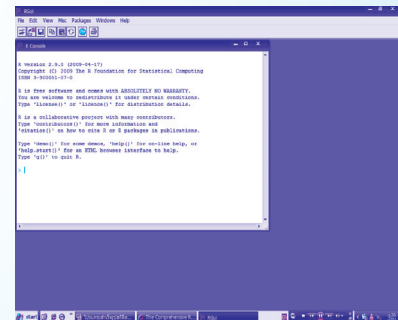
4. เลือก base



5. ขณะนี้เวอร์ชัน 2.9.0



6. Downloads และติดตั้งโปรแกรม R แล้วจะพบหน้าต่างโปรแกรม



เปิด RGui แล้วพิมพ์ `help()` <enter> หรือถ้าต้องการหาที่ `help.search("...")` เช่น ต้องการเรียนรู้คำสั่งที่เกี่ยวกับการหาค่าเฉลี่ย (mean) ก็ทำดังนี้

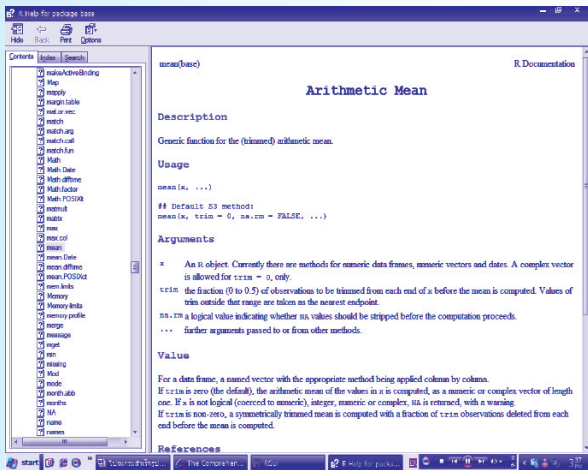
> `help.search("mean")` <enter> จะได้คำตอบเป็น Help files with alias or concept or title matching 'mean' using regular expression matching:

<code>base::Date</code>	Date Class
<code>base::Date</code>	Date Class
<code>base::colSums</code>	Form Row and Column Sums and Means
<code>base::difftime</code>	Time Intervals
<code>base::mean</code>	Arithmetic Mean
<code>boot::sunspot</code>	Annual Mean Sunspot Numbers
<code>cluster::meanabsdev</code>	Internal cluster functions



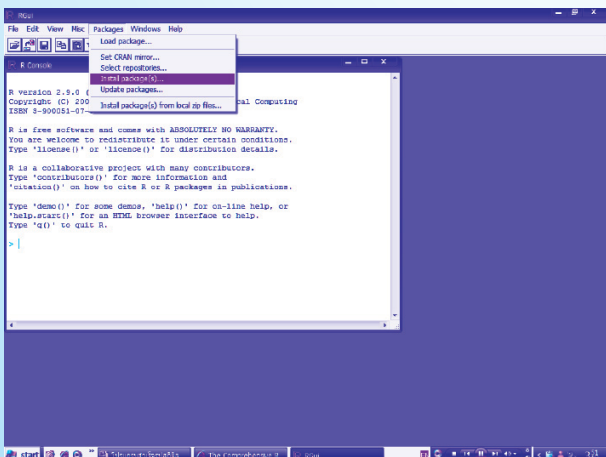
lattice::tmd	Tukey Mean-Difference Plot
Matrix::Matrix-class	Virtual Class "Matrix" Class of Matrices
Matrix::colSums	Form Row and Column Sums and Means
Matrix::dgeMatrix-class	Class "dgeMatrix" of Dense Numeric (S4 Class) Matrices
rpart::meanvar	Mean-Variance Plot for an Rpart Object
stats::kmeans	K-Means Clustering
stats::oneway.test	Test for Equal Means in a One-Way Layout
stats::weighted.mean	Weighted Arithmetic Mean

ดูตรงบรรทัดที่เขียนว่า "base::mean Arithmetic Mean" ให้พิมพ์ `> help(mean) <enter>` ก็จะได้คำตอบของการใช้คำสั่ง mean ที่อยู่ในส่วนของ package base ซึ่งมีตัวอย่างให้ดูด้วยดังนี้

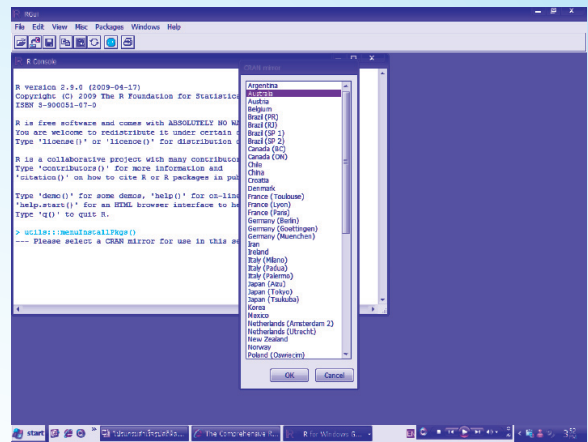


ข้อดีของโปรแกรม R นั้นนอกจากจะฟรีแล้วยังสามารถใส่ความสามารถอื่นๆ เข้าไปได้ เช่นต้องการใส่โปรแกรมเสริม EpiCalc

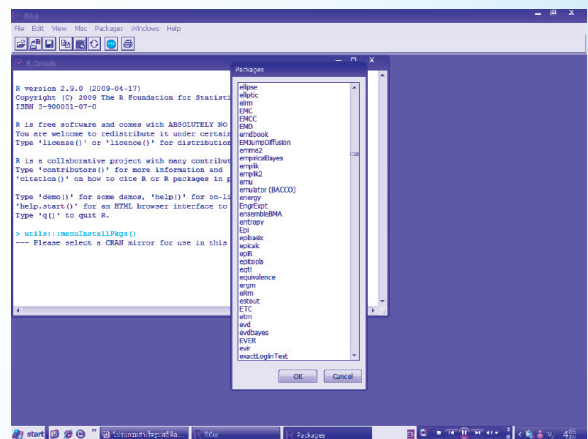
1. เลือก Packages -> Install Package(s)...



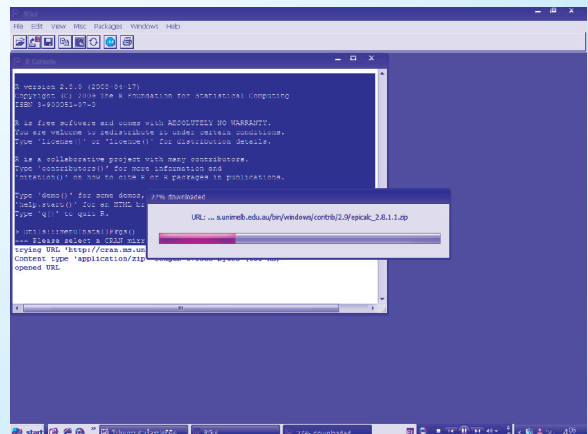
2. เลือกประเทศไหนก็ได้ (แนะนำ Australia)



3. จะเจอโปรแกรมเสริม (เรียกว่า packages) มากมาย เลือก epicalc ที่ต้องการ



4. จะทำการดาวน์โหลดโปรแกรมเสริมมาให้เอง





5. ดาวนโหลดเสร็จแล้วยังใช้ไม่ได้ ต้องเลือก Load package... เพื่อเอามาใช้ก่อน

6. แล้วเลือก epicalc เหมือนเดิม

สำหรับขั้นที่ 5-6 นั้น ต้องทำทุกครั้งหลังเปิดโปรแกรมใหม่

ปัจจุบันมีโปรแกรมที่ใช้ในการวิเคราะห์ข้อมูลทางสถิติจำนวนมาก การจัดหาโปรแกรมสำเร็จรูปทางสถิติ มาใช้งานนั้น ควรต้องคำนึงถึงองค์ประกอบหลายประการ เช่น

คำนึงถึงขนาดของข้อมูลที่จะนำมาประมวลผล, ความสามารถของเครื่องคอมพิวเตอร์ที่มีอยู่, ลิขสิทธิ์ของโปรแกรมฯ, งบประมาณที่มีอยู่, ความสามารถของโปรแกรมฯ ที่ต้องการใช้งาน ในปัจจัยหลายประการดังกล่าวนี้ ผู้เขียนไม่สามารถกำหนดได้ว่าปัจจัยใดเป็นส่วนที่สำคัญที่สุด แต่อาจต้องประกอบด้วยหลายปัจจัย และที่สำคัญอีกประการหนึ่งนอกเหนือจากประการทั้งปวงคือ “โปรแกรมทางสถิติเหล่านี้ สร้างขึ้นบนพื้นฐานที่ว่าผู้ใช้จะต้องมีความรู้ทางสถิติและเลือกสถิติในการคำนวณ วิเคราะห์ และทดสอบสมมุติฐานได้อย่างถูกต้อง” เพราะไม่มีโปรแกรมใดสามารถบอกถึงความผิดพลาดในจุดนี้ได้เลย

## เอกสารอ้างอิง

1. โปรแกรมสำเร็จรูปทางสถิติ [ออนไลน์] [อ้างถึงวันที่ 3 มิถุนายน 2552] เข้าถึงได้จาก <http://tsl.tsu.ac.th/file.php/1/courseware/math2/lesson1/index.htm>
2. การพัฒนาโปรแกรม [ออนไลน์] [อ้างถึงวันที่ 8 มิถุนายน 2552] เข้าถึงได้จาก <http://www.2poto.com/vbwebboard/00468.html>
3. ข้อดี ข้อเสียของโปรแกรม [ออนไลน์] [อ้างถึงวันที่ 10 มิถุนายน 2552] เข้าถึงได้จาก <http://www.dmbn.net/board/?task=main&order=message&cat=1&board=1&topic=328>
4. การใช้โปรแกรม R [ออนไลน์] [อ้างถึงวันที่ 10 มิถุนายน 2552] เข้าถึงได้จาก <http://www.clinicalepi.com/2009/02/r-odds-ratio.html>

สำนักบริหารและรับรองห้องปฏิบัติการ กรมวิทยาศาสตร์บริการ  
โทร. 0-2201-7194  
e-mail : [kanittaa@dss.go.th](mailto:kanittaa@dss.go.th)