**การเลือกใช้สถิติเบื้องต้นและการวิเคราะห์ข้อมูลโดยใช้โปรแกรม Epi Info**

นพ.ยงเจือ เหล่าศิริถาวร

การวิจัยหรือการศึกษาทางระบาดวิทยาที่มีคุณภาพนั้น จำเป็นต้องเลือกรูปแบบการศึกษาที่เหมาะสมกับวัตถุประสงค์ การค้นคว้าวรรณกรรมที่เกี่ยวข้องอย่างมีระบบแบบแผน การดำเนินการวิจัยที่รัดกุมรอบคอบ ไม่มีอคติในวิธีการศึกษา นอกจากนี้แล้วยังต้องการการวิเคราะห์ข้อมูลที่ถูกต้องตามระเบียบวิธีทางสถิติ รวมทั้งการแปลผลการวิเคราะห์ด้วย

การวิเคราะห์ข้อมูลทางสถิติที่เป็นพื้นฐานทั่วไปที่ควรรู้ ได้แก่ การวิเคราะห์สัดส่วน การวิเคราะห์ค่าเฉลี่ย และการวิเคราะห์ข้อมูลสำหรับการศึกษาทางระบาดวิทยาแต่ละรูปแบบ เช่น Cross-sectional study, Cohort study, Case-control study ซึ่งสามารถนำไปใช้เป็นพื้นฐานในการวิเคราะห์การศึกษาทางระบาดวิทยาประยุกต์ เช่น การเฝ้าระวังโรค การสอบสวนโรค เป็นต้น

**สิ่งที่ต้องพิจารณาในการวิเคราะห์ข้อมูล** คือ

1. ชนิดของตัวแปร ได้แก่ ตัวแปรเชิงคุณภาพ ตัวแปรเชิงปริมาณ

**แผนภูมีที่ 1** ชนิดตัวแปร

ตัวแปร (Variables)

เชิงคุณภาพ (Qualitative) เชิงปริมาณ (Quantitative)

ไม่มีอันดับ มีอันดับ ไม่ต่อเนื่อง ต่อเนื่อง

(Nominal) (Ordinal) (Discrete) (Continuous)

2 กลุ่ม > 2 กลุ่ม

(Diclotomous) (Polychotomous)

เช่น เช่น เช่น เช่น เช่น

เพศ สถานภาพสมรส ระดับการศึกษา จำนวนบุตร ส่วนสูง

การป่วย อาชีพ ความรุนแรงของโรค จำนวนการตั้งครรภ์ น้ำหนัก

**Categorical data Numerical data**

การวิเคราะห์ การวิเคราะห์

ค่าเฉลี่ย (Mean), มัฐยฐาน (Median), ฐานนิยม (Mode), พิสัย (Range), ส่วนเบี่ยงเบนมาตรฐาน (SD)

อัตรา (Rate), อัตราส่วน (Ratio), สัดส่วน (Proportion)

1. ชนิดของสถิติที่จะใช้ ได้แก่

* สถิติเชิงพรรณนา (Descriptive statistics) เป็นการสำรวจและอธิบายข้อมูลที่รวบรวมมาแบบสรุป เพื่อให้เห็นภาพรวม เช่น อัตรา อัตราส่วน สัดส่วน ค่าเฉลี่ย ค่ามัธยฐาน ฐานนิยม
* สถิติเชิงอนุมาน (Inferential statistics) การใช้ค่าประมาณที่ได้จากกลุ่มตัวอย่างเพื่ออธิบายลักษณะของประชากร เช่น การทดสอบความแตกต่างของสัดส่วน การทดสอบความแตกต่างของค่าเฉลี่ย การทดสอบความสัมพันธ์

1. แนวทางการวิเคราะห์ทางสถิติ

* การทดสอบสมมติฐาน (hypothesis testing)
* การคาดประมาณ (estimation) ประกอบด้วย คาดประมาณเฉพาะค่า (point estimation) และช่วงเชื่อมั่น (confidence interval)

**การทดสอบสมมติฐาน (Hypothesis testing)**

* สมมติฐานทางเดียว เช่น ผู้สูบบุหรี่ป่วยเป็นมะเร็งปอดมากกว่าหรือเท่ากับผู้ไม่สูบ , ผู้ได้รับวัคซีนหัดป่วยเป็นโรคหัดน้อยกว่าผู้ที่ไม่ได้รับ
* สมมติฐานสองทาง เช่น ผู้รับประทานยาป้องกันโรคมาลาเรียป่วยเป็นโรคมาลาเรียไม่ต่างจากผู้ที่ไม่ได้รับประทาน
* การตั้งสมมติฐาน Null hypothesis และ alternative hypothesis เช่น
  + Ho : P1 > P2 ; Ha : P1 < P2
  + Ho : P1 < P2 ; Ha : P1 > P2
  + Ho : P1 = P2 ; Ha : P1 ≠ P2
* ความคลาดเคลื่อนทางระบาดวิทยา (Error)

ความเป็นจริง

|  |  |  |
| --- | --- | --- |
|  | A เท่ากับ B | A ไม่เท่ากับ B |
| A เท่ากับ B | Confidence | β error |
| A ไม่เท่ากับ B | α error | Power |

การสรุปผลการศึกษา

* สถิติที่ใช้ในการทดสอบสมมติฐาน เช่น Z test, t test, χ2 test
* ค่า α นิยมใช้ที่ระดับ 0.05
* ใช้ p - value เป็นตัววัดความสอดคล้องระหว่างสมมติฐานกับข้อมูล ซึ่ง เป็นความน่าจะเป็น มีค่าระหว่าง 0 ถึง 1
  + p - value ต่ำหมายถึงข้อมูลมีความสอดคล้องกับ Null hypothesis ต่ำ โดยทั่วไปใช้ค่าต่ำกว่า 0.05 ( ค่า α )
  + p - value สูงหมายถึงข้อมูลมีความสอดคล้องกับ Null hypothesisสูง โดยทั่วไปใช้ค่ามากกว่าหรือเท่ากับ 0.05
* ความหมายของ p - value กรณีใช้ระดับ α = 0.05
  + p - value > 0.05 : ยอมรับ null hypothesis
  + p - value < 0.05 : ปฏิเสธ null hypothesis 🡪 เกิด α error ไม่เกิน 5%
* ตัวอย่างการแปรผลการทดสอบสมมติฐาน (α = 0.05) : สมมติฐาน คือ อัตราป่วยโรคมะเร็งปอดในกลุ่มผู้สูบบุหรี่ไม่แตกต่างกับผู้ไม่สูบ ได้ p - value เท่ากับ 0.02
  + อัตราป่วยแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ???
  + อัตราป่วยแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ที่ระดับ α = 0.05 ??
  + อัตราป่วยแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ที่ระดับ α = 0.05 (p – value = 0.02) ?
  + ถ้าความจริงอัตราป่วยไม่แตกต่างกัน การใช้ข้อมูลชุดนี้สรุปว่าแตกต่างกันมีโอกาสผิดพลาด 2%
* ข้อด้อยของการใช้ p - value
  + บอกแค่ความแตกต่างทางสถิติ (โดยเฉพาะการทดสอบสมมติฐานทางเดียว) ในขณะที่ระบาดวิทยาต้องการทราบทั้งทิศทางและขนาดของความแตกต่าง
  + อาจชักนำให้เกิดการแปลผลผิดพลาดได้ เช่น หากค่าสถิติบ่งชี้ว่าน่าจะไม่แตกต่าง ในกรณีขนาดตัวอย่างน้อยๆ , หรือหากค่าสถิติบ่งชี้ว่าน่าจะแตกต่าง ในกรณีขนาดตัวอย่างมากๆ

**การคาดประมาณ (Estimation)**

* การใช้ค่าสถิติจากการศึกษา อนุมานไปถึงค่าประชากร
  + point estimation : การคาดประมาณเป็นค่าเดียว ใช้น้อย เนื่องจากโอกาสที่ค่าสถิติที่ได้จากการศึกษาจะเท่ากับค่าของประชากรจริงมีน้อย
  + interval estimation : การคาดประมาณค่าของประชากร (parameter) เป็นช่วง โดยคาดประมาณช่วงที่จะมีค่า parameter อยู่ เช่น การคำนวณช่วงเชื่อมั่น (confidence interval) โดยทั่วไป คำนวณที่ 95% confidence interval (95% CI) โดยความกว้างของช่วงจะบ่งชี้ความเที่ยงของการศึกษา
* Null value : ค่าที่แสดงไม่มีความแตกต่าง ระหว่างกลุ่มที่เปรียบเทียบในการวิเคราะห์ข้อมูล
  + ในกรณี relative risk เป็น ratio : Null value เท่ากับ 1
  + ในกรณี relative risk เป็น difference : Null value เท่ากับ 0
* ความหมายกรณี 95% confidence interval
  + หากการทำศึกษาชนิดนี้หลายๆครั้ง แต่ละครั้งของการศึกษาก็จะได้ช่วงเชื่อมั่นหนึ่งช่วง ช่วงเชื่อมั่นทั้งหมดจะครอบคลุมค่า parameter อยู่ 95% (หากมี 100 ช่วงเชื่อมั่น จะครอบคลุมค่าจริงของประชากรอยู่ 95 ช่วงเชื่อมั่น ) แต่ในความเป็นจริงไม่มีใครทำการศึกษาซ้ำกันเป็นร้อยครั้ง จึงไม่สามารถยืนยันว่าช่วงเชื่อมั่นในครั้งที่ศึกษานี้จะครอบคลุมค่าจริงของประชากร ดังนั้นจึงอาจแปรผลอย่างง่ายๆ ได้คือ มีความเชื่อมั่น 95% ว่าช่วงเชื่อมั่นในการศึกษานี้จะครอบคลุมค่าจริงของประชากร ซึ่งถ้าหากคลุมค่าจริงแล้ว ค่าจริงนั้นจะอยู่ระหว่าง …. ถึง ……
  + ถ้า 95% CI คร่อม null value ( 1 ในกรณี ratio ) จะพบว่าหากทดสอบสมมติฐานที่ α = 0.05 โดยใช้ statistics เดียวกันทั้งการคำนวณค่าช่วงเชื่อมั่นและการทดสอบสมมติฐาน จะพบว่า p - value > 0.05 ด้วย
* ตัวอย่างการแปรผล 95% confidence interval : การศึกษาความสัมพันธ์ระหว่างการสูบบุหรี่กับโรคมะเร็งปอด : ได้ค่า risk ratio = 9 , 95% CI = 5.2, 15.4
  + ผู้ที่สูบบุหรี่เสี่ยงต่อการป่วยเป็นโรคมะเร็งปอด 9 เท่าเทียบกับผู้ที่ไม่สูบบุหรี่
  + มีความเชื่อมั่นว่าหากทำการศึกษาแบบนี้ซ้ำๆกันหลายๆครั้ง จะมีจำนวนการศึกษา ร้อยละ 95 ที่ช่วงเชื่อมั่นจะครอบคลุมค่าเสี่ยงจริง โดยหากการศึกษาครั้งนี้มีช่วงเชื่อมั่นที่คลุมค่าเสี่ยงสัมพัทธ์จริงของประชากรแล้ว ค่าเสี่ยงสัมพัทธ์จริงจะมีค่าอยู่ระหว่าง 5.2 เท่า ถึง 15.4 เท่า หรือ แปลอย่างง่ายว่า มีความเชื่อมั่น 95% ว่าช่วงเชื่อมั่นในการศึกษาครั้งนี้จะครอบคลุมค่าเสี่ยงสัมพัทธ์จริงของประชากร โดย ค่าความเสี่ยงสัมพัทธ์จริงของประชากรจะอยู่ระหว่าง 5.2 เท่า ถึง 15.4 เท่า

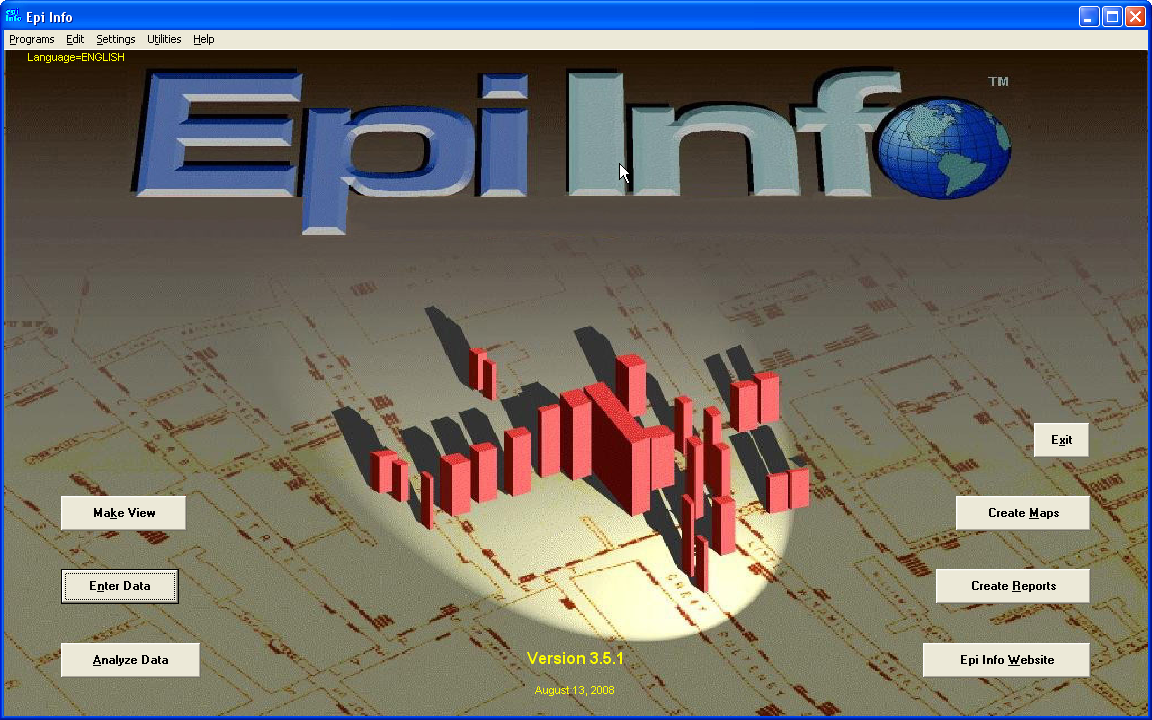
**การเปิดโปรแกรม Analysis และแฟ้มข้อมูล**

* 1. Start 🡪All program 🡪 Epi Info 🡪 Epi Info



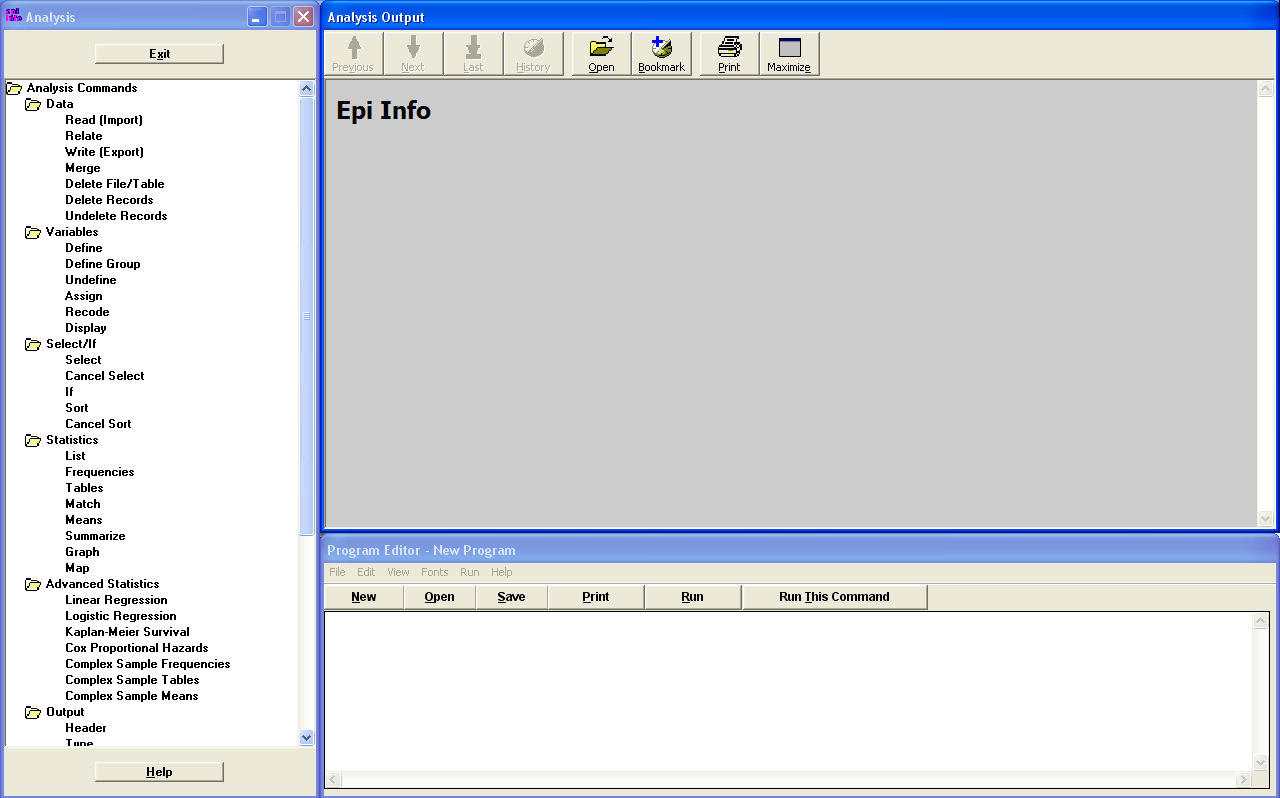
เลือก

* 1. เลือก Analysis



เลือก

* 1. หน้าต่างโปรแกรม Analysis:



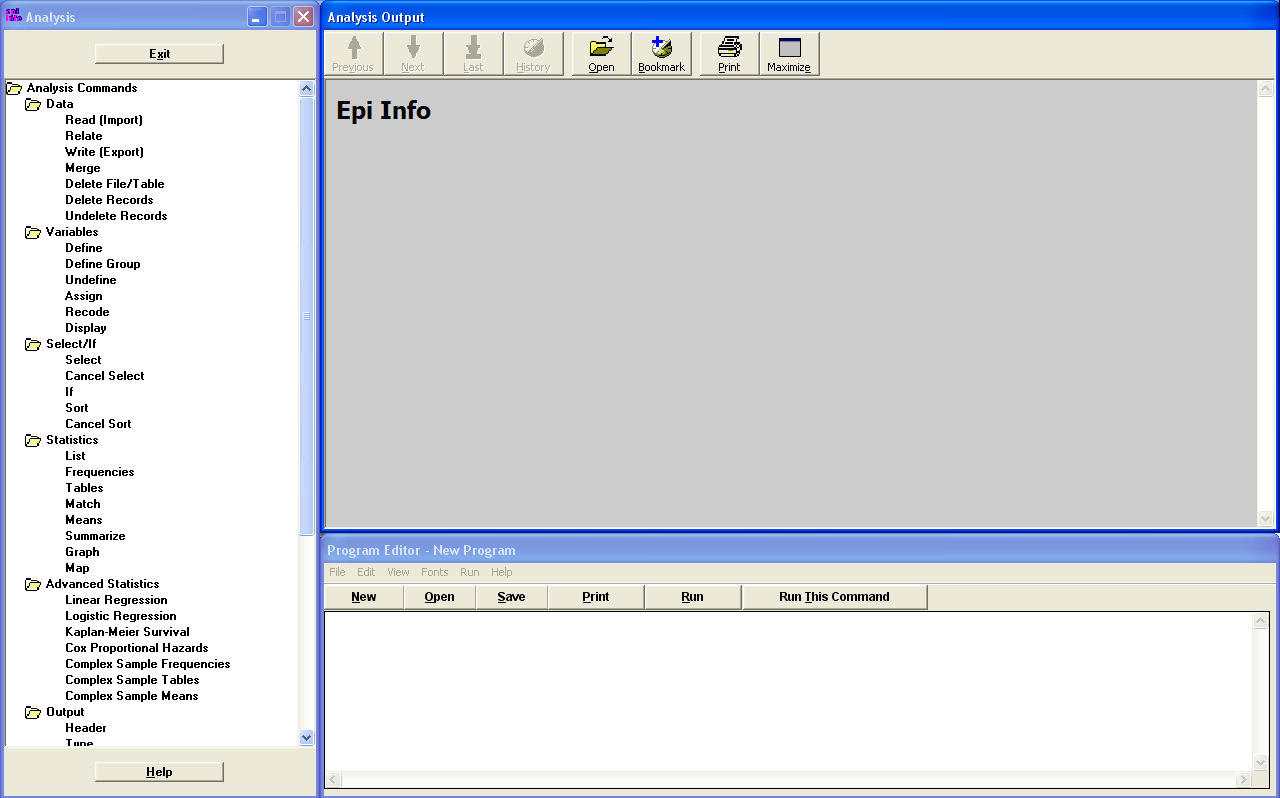
Program Editor

Output window

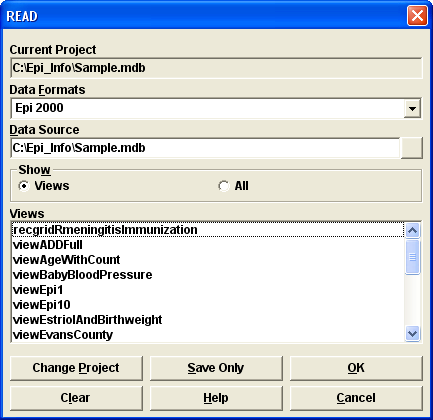
Command window

**การเปิดแฟ้มข้อมูลชนิด .mdb**

1. ในเอกสารแบบฝึกหัดนี้ ให้สร้างโฟลเดอร์ชื่อ training ใน Drive C แล้ว copy แฟ้มข้อมูลตัวอย่างชื่อ Anginal.mdb ไปไว้ในโฟลเดอร์ดังกล่าว
2. ใช้คำสั่ง Read ใน Command window จะมีหน้าต่าง Read ขึ้นมา ให้ click ที่ปุ่ม Chang Project แล้ว browse ไปหาโฟลเดอร์ c:\training เลือก Angina.mdb (ถ้าข้อมูลที่จะวิเคราะห์ ไม่ใช่นามสกุล .mdb จะใช้ Change Project ไม่ได้ ให้เปลี่ยน Data fomat เป็นชนิดของข้อมูล แล้ว browse Data Source)



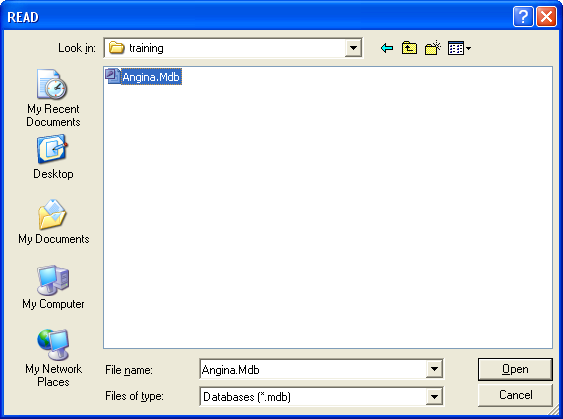
Read แฟ้มข้อมูล



Click

ใช้กรณีข้อมูลเป็น format อื่นที่ไม่ใช่ .mdb

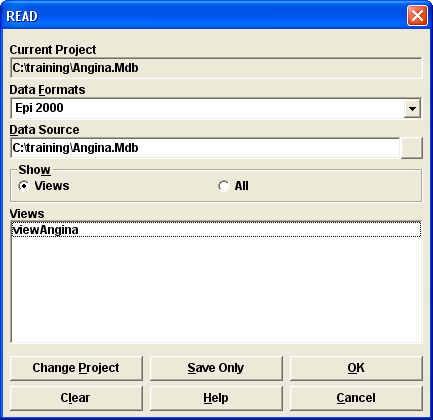
ไม่ใช่ข้อมูลที่จะวิเคราห์



2) Double click ที่แฟ้มข้อมูลที่ต้องการ

1) Browse

1. เลือก ViewAngina (หากข้อมูลเป็น .mdb ที่ไม่ได้สร้างฐานข้อมูลจาก EpiInfo อื่น ให้เลือก Show All แล้วเลือกตารางข้อมูลที่จะวิเคราะห์)



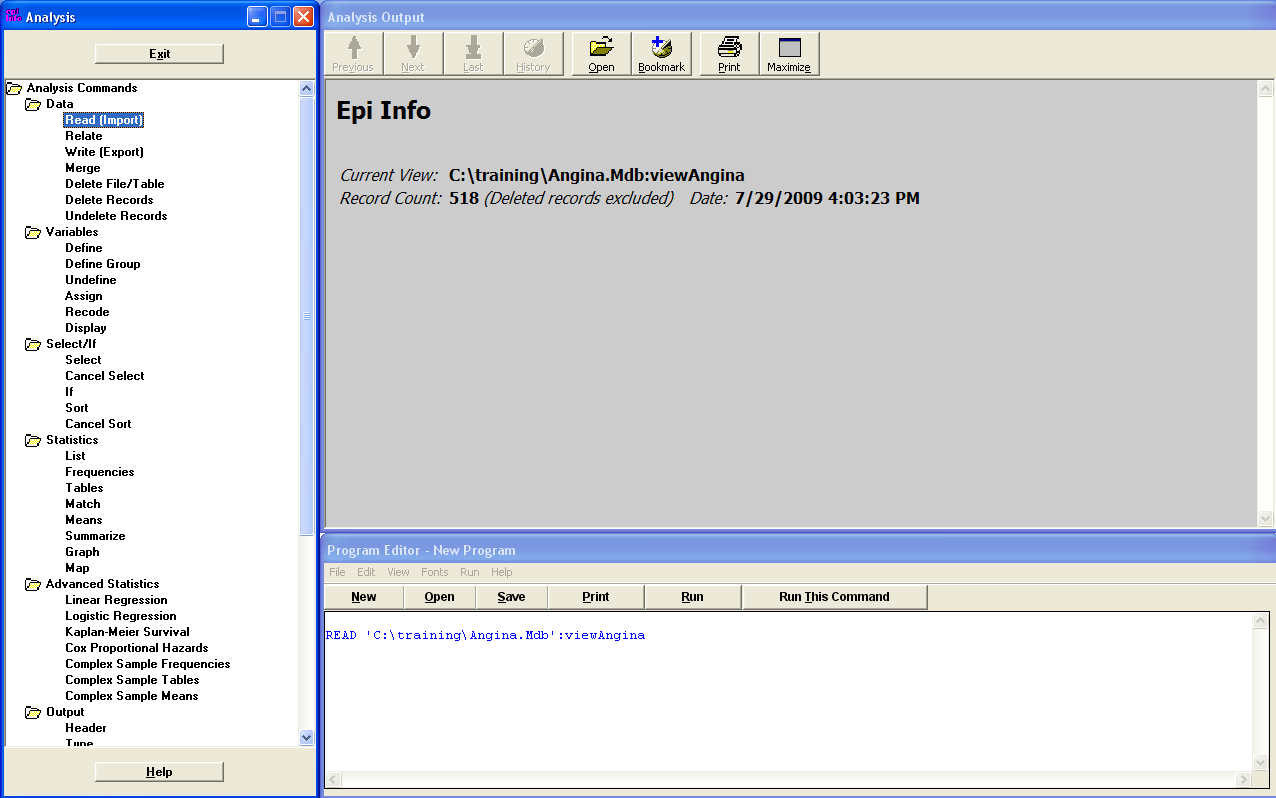
2) Click

1) Click

Click All แล้วเลือก ตารางข้อมูลที่ต้องการ (หากเป็นฐานข้อมูลที่ไม่ได้สร้างจาก Epi Info

ข้อมูลที่จะวิเคราะห์

1. ชื่อแฟ้มข้อมูล และจำนวน record จะปรากฏใน Output window, command ที่ใช้จะปรากฏใน Program Editor window



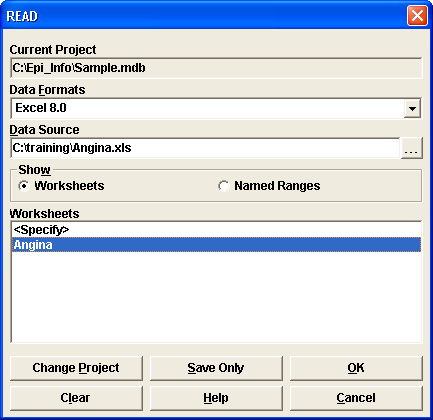
แฟ้มข้อมูล Angina.mdb

Command

**การเปิดแฟ้มข้อมูลชนิดอื่นที่ไม่ใช่ .mdb**

*ตัวอย่างการเปิดแฟ้มข้อมูลชนิด .xls (Microsoft Excel version 97-2003)*

1. เปลี่ยน Data format เป็น Excel 8.0 แล้ว browse ไปหาโฟลเดอร์ c:\training เลือก Angina.xls แล้ว เลือก Angina ที่ Worksheets (ในตัวอย่างนี้มีเพียง worksheet เดียวคือ Angina)



4) Click

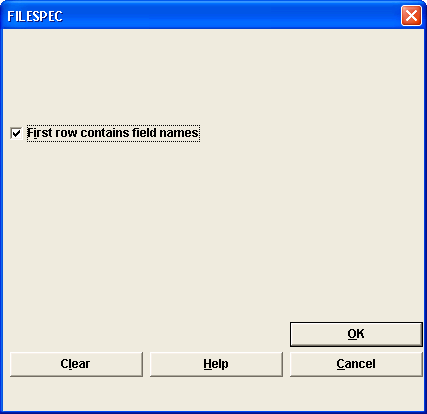
3) เลือก Worksheets ที่ต้องการ

2) Browse

1) Click เลือก Format เป็น Excel 8.0

ไม่ใช่ข้อมูลที่จะวิเคราะห์

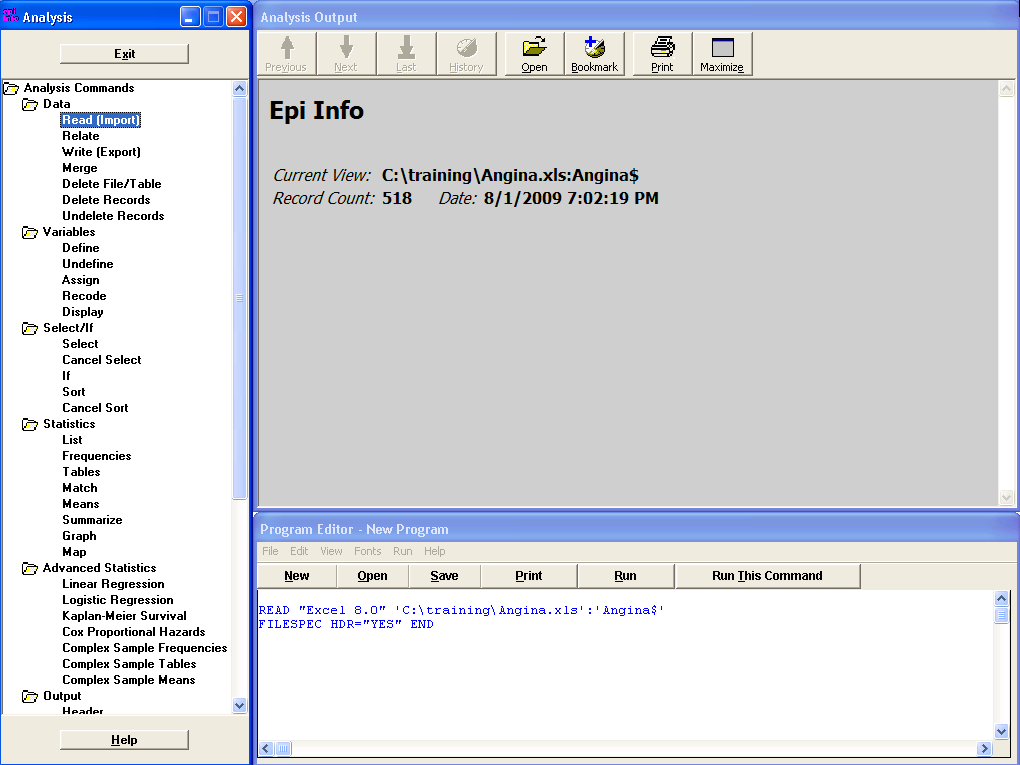
1. จะมีหน้าต่างใหม่ขึ้นมาถามว่า แถวแรกใน Worksheet Angina เป็นชื่อตัวแปรในฐานข้อมูลหรือไม่



2) Click

1) Check ถ้าแถวแรกเป็นชื่อตัวแปร, Uncheck ถ้าแถวแรกไม่ใช่ชื่อตัวแปร แต่เป็นข้อมูลของ record ที่ 1

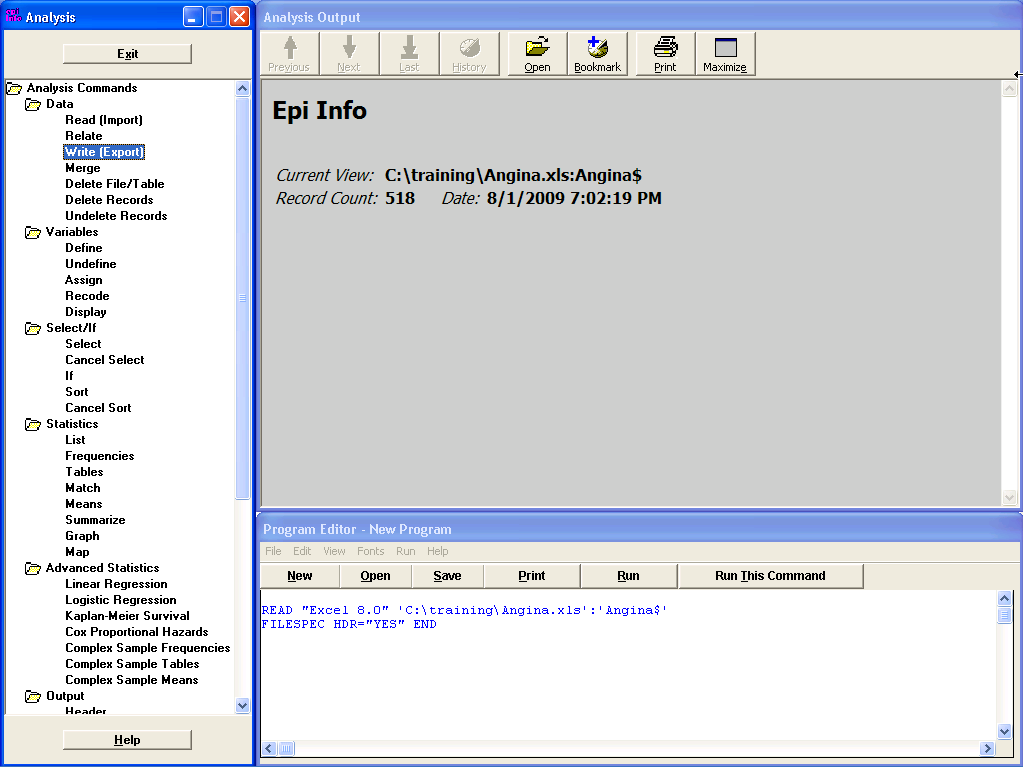
1. ชื่อแฟ้มข้อมูล และจำนวน record จะปรากฏใน Output window, command ที่ใช้จะปรากฏใน Program Editor window



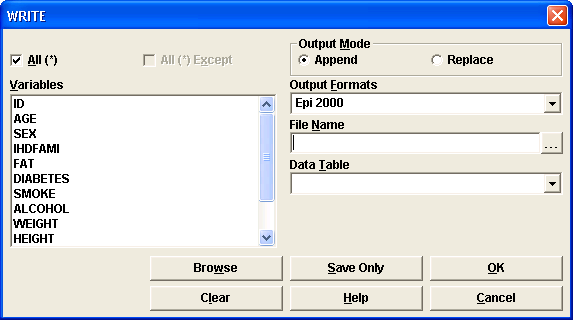
Command

แฟ้มข้อมูล Angina.mdb

1. สร้างฐานข้อมูล .mdb ให้กับแฟ้ม Anginal.xls ที่เปิดมาเพื่อ save งานที่จะวิเคราะห์ในขั้นตอนต่อไปให้สามารถจัดเก็บในแฟ้มเดียวกันได้ ไม่เช่นนั้นจะวิเคราะห์ได้ แต่ไม่สามารถ save งานที่จะทำต่อมาได้



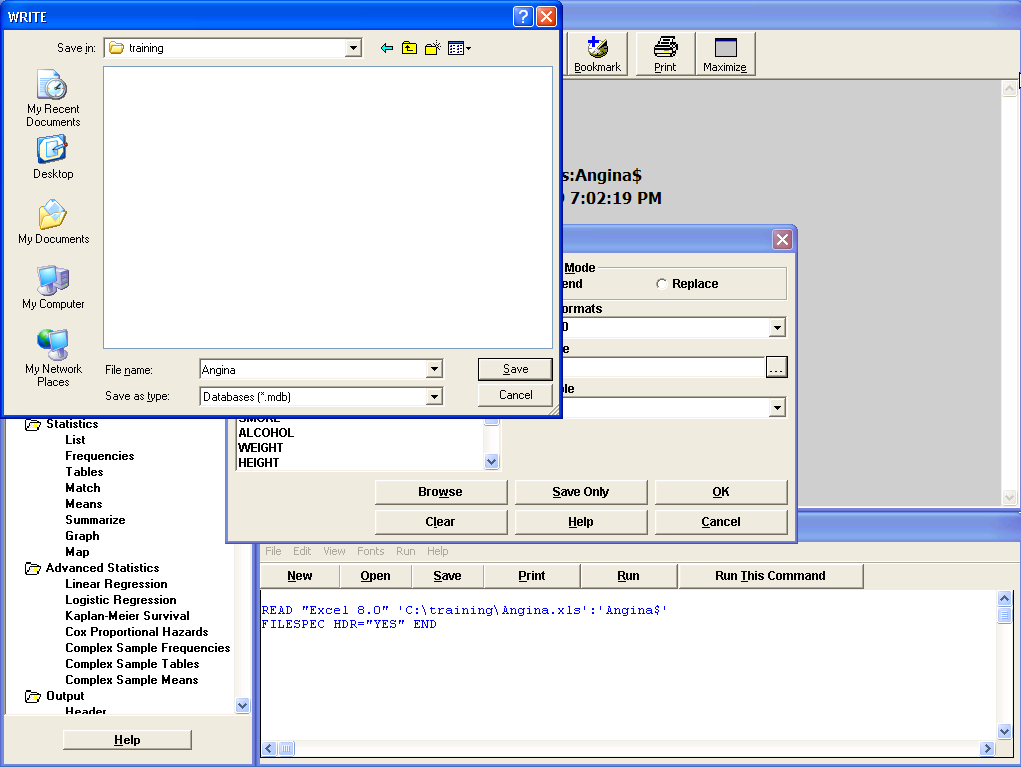
Click



3) Browse ไป Folder ที่ต้องการ

2) เลือก Epi 2000

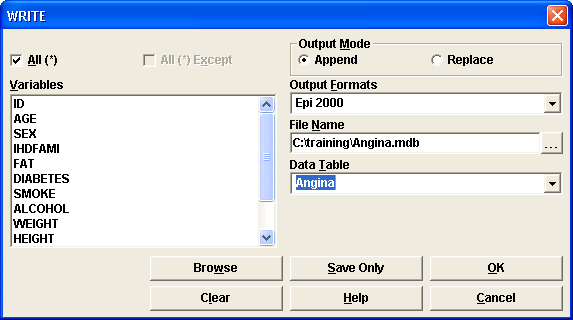
1) เลือก Append



2) ตั้งชื่อ File เช่น Angina

3) Click Save

1) Browse



2) Click

1) พิมพ์ชื่อData table ที่ต้องการจัดเก็บใน Anginal.mdb เช่น Angina

ชื่อ File ที่ตั้งจากขั้นตอนที่แล้ว

จากนั้นก็ใช้คำสั่ง Read ไปยัง Angina.mdb ที่เพิ่งสร้าง ดังในตัวอย่าง **การเปิดแฟ้มข้อมูลชนิด .mdb**

**การวิเคราะห์สัดส่วน**

การวิเคราะห์สัดส่วนใช้เมื่อข้อมูลเป็น Categorical data ซึ่งสรุปข้อมูลโดยใช้ค่าสัดส่วน (proportion) การแจกแจงให้เป็นสัดส่วนทำด้วยการสร้างตารางแจกแจงความถี่

**แผนภูมีที่ 2** การวิเคราะห์สัดส่วน

การวิเคราะห์สัดส่วน

กลุ่มเดียว

สองกลุ่ม

สามกลุ่มขึ้นไป

อิสระต่อกัน

ไม่อิสระต่อกัน

อิสระต่อกัน

ไม่อิสระต่อกัน

* p , 95%CI
* Generalized estimating equations (GEEs)
* Binomial exact probability test
* Freman Halton conditional exact test
* Fisher’s exact test

**กรณีตัวอย่างมีขนาดเล็ก**

* p1 – p2 , 95%CI
* Z test , χ2 test
* p1 – p2 , 95%CI
* Z test , Macnemar test
* p1 , 95%CI

p2 , 95%CI

p3 , 95%CI

* χ2 test

*ที่มา: บัณฑิต ถิ่นคำรพ.   คู่มือปฏิบัติการชีวสถิติ: ตำราสำหรับเรียนรู้ชีวสถิติด้วยตนเอง*

ใช้ File ข้อมูล “angina.mdb” ซึ่งเป็น file การศึกษาภาวะเจ็บหน้าอกชนิดโรคหัวใจขาดเลือดโดยสุ่มตัวอย่างจากประชากรผู้ใหญ่ในหมู่บ้านสนธยา

โครงสร้างแฟ้มข้อมูล

**ตัวแปร คำอธิบาย**

ID หมายเลขลำดับของ record

AGE อายุ (ปี)

SEX เพศ

0 = หญิง 1 = ชาย

IHDFAMI ประวัติการเป็นโรคหัวใจขาดเลือดในครอบครัว

0 = ไม่มี 1 = มี

FAT การรับประทานอาหารไขมัน

0 = ไม่รับประทาน 1 = รับประทานปานกลาง 2 = รับประทานมาก

DIABETES ประวัติการเป็นโรคเบาหวาน

0 = ไม่เป็น 1 = เป็น

SMOKE การสูบบุหรี่

0 = ไม่เคยสูบ 1 = เคยสูบแต่เลิกแล้ว 2 = ยังสูบอยู่

ALCOHOL การดื่มสุรา

0 = ไม่เคยดื่ม 1 = เคยดื่มแต่เลิกแล้ว 2 = ยังดื่มอยู่

WEIGHT น้ำหนัก (กิโลกรัม)

HEIGHT ส่วนสูง (เซนติเมตร)

SBP ความดัน systolic (มิลลิเมตรปรอท)

DBP ความดัน diastolic (มิลลิเมตรปรอท)

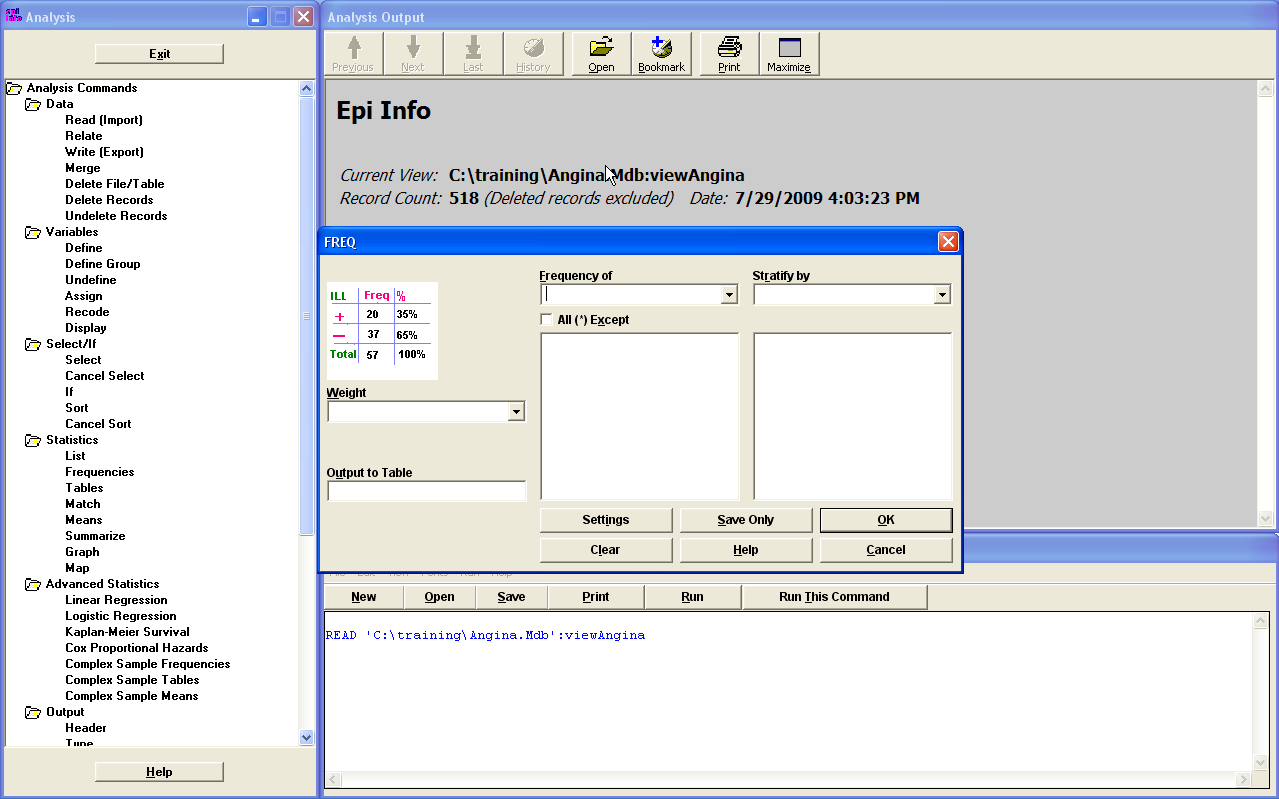
ANGINA อาการเจ็บหน้าอก ตามเกณฑ์การวินิจฉัยภาวะหัวใจขาดเลือด

0 = ไม่เป็น 1 = เป็น

1. **การวิเคราะห์สัดส่วน 1 กลุ่ม**

*ตัวอย่าง* : การแจกแจงสัดส่วนของการมีอาการเจ็บหน้าอก

การสร้างตารางทางเดียวแสดงสัดส่วนการมีอาการเจ็บหน้าอก



3) Click OK

2) กด drop down menuเลือกตัวแปร Angina

1) เลือก Frequencies

command ที่ใช้จะปรากฏใน Program Editor window อาจพิมพ์เองโดยตรงใน Program Editor หากจำคำสั่งและชื่อตัวแปรได้ แล้วกด Run This Command ที่แถบเมนู

**FREQ ANGINA** ↵

หมายเหตุ : เครื่องหมาย ↵ หมายถึง กดปุ่มคำสั่ง “Run This Command”

**FREQ ANGINA**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ANGINA** | **Frequency** | **Percent** | **Cum Percent** |  |
| **0** | 424 | 81.9% | 81.9% | YELLOW |
| **1** | 94 | 18.1% | 100.0% | YELLOW |
| **Total** | 518 | 100.0% | 100.0% | ORANGE |

|  |  |  |
| --- | --- | --- |
| **95% Conf Limits** | | |
| 0 | 78.2% | 85.0% |
| 1 | 15.0% | 21.8% |

* + มีประชากรที่มีอาการเจ็บหน้าอกจำนวน 94 คน จาก 518 คน คิดเป็นความชุกเท่ากับ 18.1%
  + 95% CI มีให้สำหรับสัดส่วนในแต่ละกลุ่ม หมายความว่า ถ้าทำการสุ่มตัวอย่างโดยปราศจากอคติจากประชากร แปลผลได้ว่า มีความเชื่อมั่น 95% ว่าช่วงเชื่อมั่นในการศึกษาครั้งนี้จะครอบคลุมค่าความชุกจริง ซึ่งถ้าหากคลุมค่าจริงแล้ว ค่าความชุกของ อาการเจ็บหน้าอก ในประชากรจะอยู่ระหว่างร้อยละ 15 ถึง ร้อยละ 21.8

**หมายเหตุ :** การรายงานช่วงเชื่อมั่นจะมีความเหมาะสม หากการศึกษามีวิธีการศึกษาและสุ่มตัวอย่างที่เหมาะสมสำหรับการตอบคำถามการวิจัยในข้อนี้ ไม่เช่นนั้น ถึง Epi Info จะแสดงผลออกมา ก็ไม่เหมาะสมจะนำไปใช้

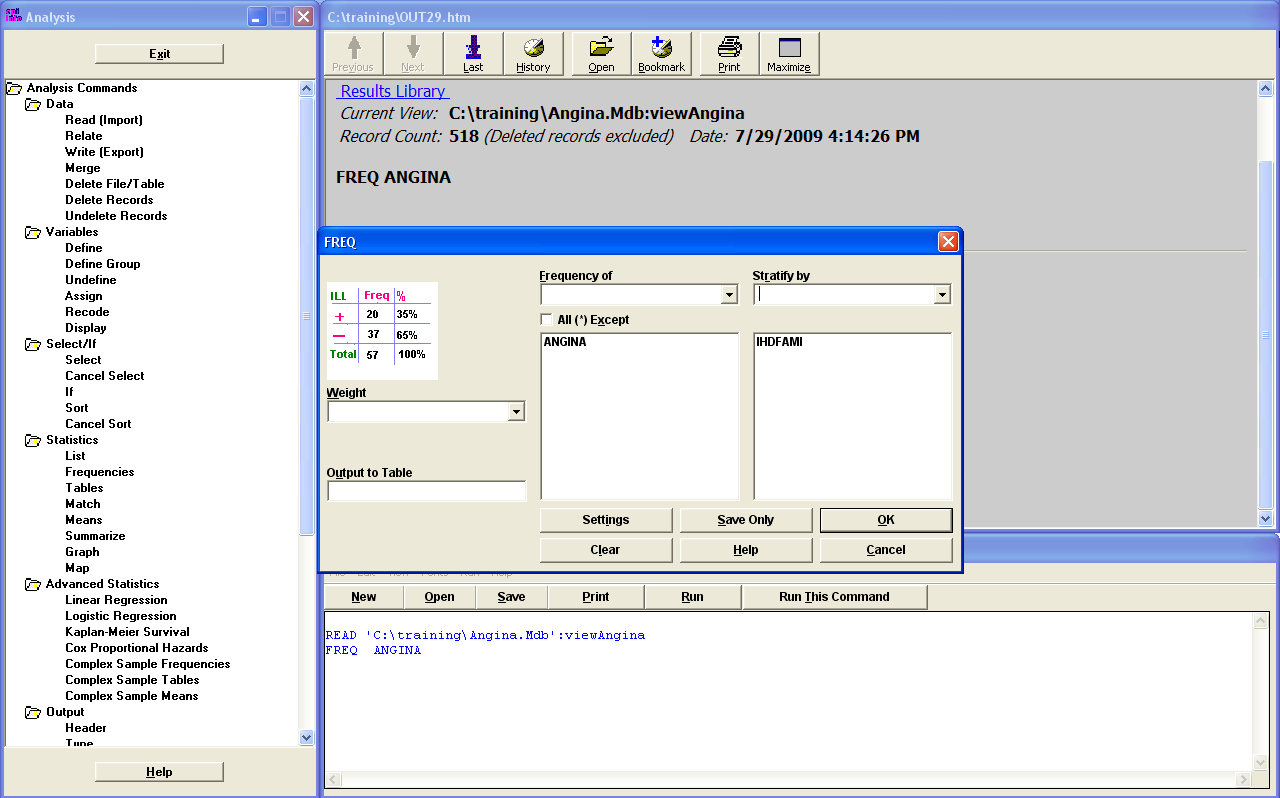
*แบบฝึกหัด*

* + แจกแจงสัดส่วนของเพศและช่วงเชื่อมั่น
  + แจกแจงสัดส่วนของการมีประวัติโรคหัวใจในครอบครัวและช่วงเชื่อมั่น
  + แจกแจงสัดส่วของการรับประทานอาหารไขมันและช่วงเชื่อมั่น
  + แจกแจงสัดส่วนของประวัติการเป็นโรคเบาหวานและช่วงเชื่อมั่น
  + แจกแจงสัดส่วนของการสูบบุหรี่และช่วงเชื่อมั่น
  + แจกแจงสัดส่วนของการดื่มสุราและช่วงเชื่อมั่น

1. **การวิเคราะห์เปรียบเทียบสัดส่วน 2 กลุ่ม**
   1. **การแจกแจงความถี่ของข้อมูล**

*ตัวอย่าง* : การแจกแจงสัดส่วนของการมีอาการเจ็บหน้าอก

* + การสร้างตารางแจกแจงความถี่แสดงสัดส่วนการมีอาการเจ็บหน้าอก จำแนกตามการมีประวัติโรคหัวใจขาดเลือดในครอบครัว และแสดงค่าสัดส่วนเป็นร้อยละตามการมีประวัติโรคหัวใจขาดเลือดในครอบครัว



3) กด drop down menuเลือกตัวแปร IHDFAMI

2) กด drop down menuเลือกตัวแปร Angina

1) เลือก Frequencies

4) Click OK

command ที่ใช้จะปรากฏใน Program Editor window อาจพิมพ์เองโดยตรงใน Program Editor หากจำคำสั่งและชื่อตัวแปรได้ แล้วกด Run This Command ที่แถบเมนู

**FREQ ANGINA STRATAVAR = IHDFAMI** ↵

**FREQ ANGINA STRATAVAR = IHDFAMI**

**ANGINA, IHDFAMI=0**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ANGINA** | **Frequency** | **Percent** | **Cum Percent** |  |
| **0** | 412 | 83.4% | 83.4% | YELLOW |
| **1** | 82 | 16.6% | 100.0% | YELLOW |
| **Total** | 494 | 100.0% | 100.0% | ORANGE |

|  |  |  |
| --- | --- | --- |
| **95% Conf Limits** | | |
| 0 | 79.8% | 86.5% |
| 1 | 13.5% | 20.2% |

**ANGINA, IHDFAMI=1**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ANGINA** | **Frequency** | **Percent** | **Cum Percent** |  |
| **0** | 12 | 50.0% | 50.0% | YELLOW |
| **1** | 12 | 50.0% | 100.0% | YELLOW |
| **Total** | 24 | 100.0% | 100.0% | ORANGE |

|  |  |  |
| --- | --- | --- |
| **95% Conf Limits** | | |
| 0 | 29.1% | 70.9% |
| 1 | 29.1% | 70.9% |

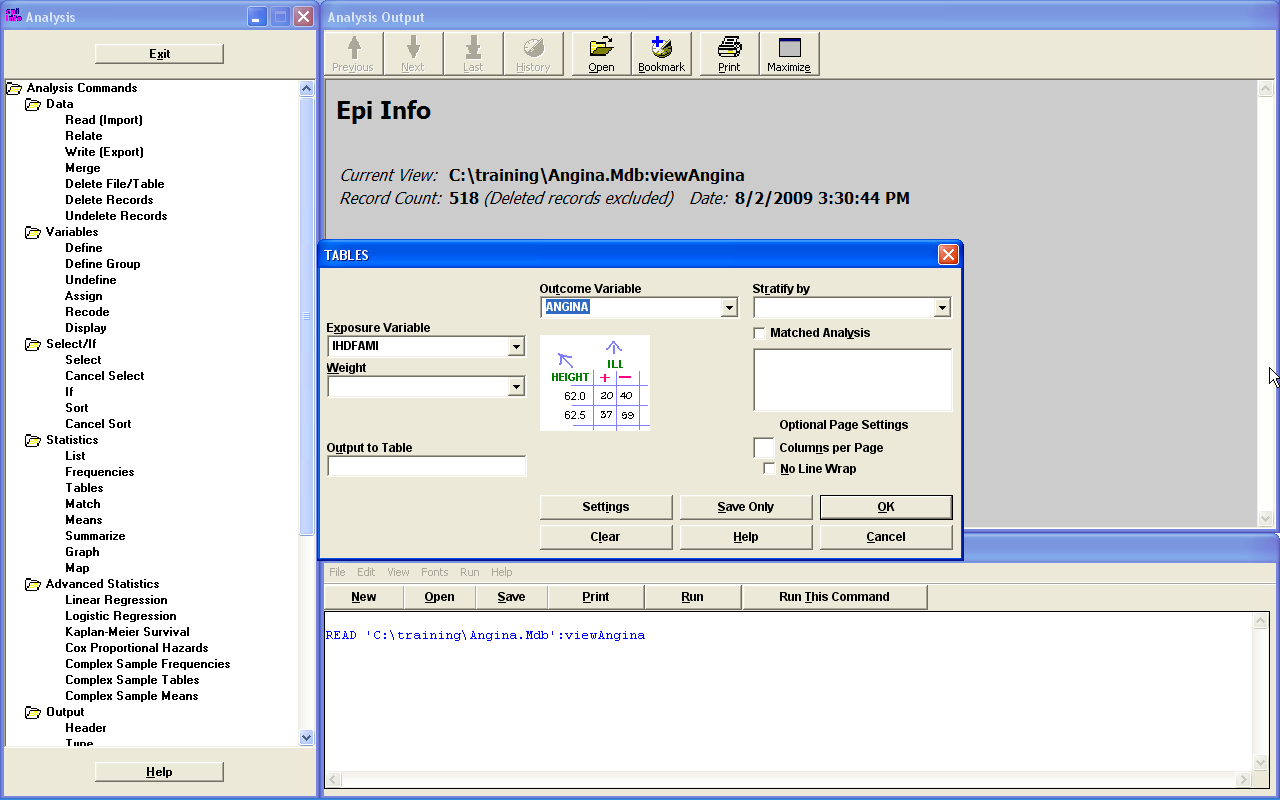
* + สัดส่วนการมีอาการเจ็บหน้าอกในกลุ่มที่ไม่มีประวัติโรคหัวใจขาดเลือดครอบครัว และในกลุ่มที่มีประวัติโรคหัวใจขาดเลือดในครอบครัว เท่ากับ 16.6% และ 50.0% ตามลำดับ
  + มีความเชื่อมั่น 95% ว่าช่วงเชื่อมั่นในการศึกษาครั้งนี้จะครอบคลุมค่าความชุกจริง ซึ่งถ้าหากคลุมค่าจริงแล้ว ความชุกของอาการเจ็บหน้าอกในประชากรที่ไม่มีประวัติโรคหัวใจขาดเลือดในครอบครัวอยู่ระหว่างร้อยละ 13.5 ถึง 20.2 และ ความชุกของอาการเจ็บหน้าอกในประชากรที่มีประวัติโรคหัวใจขาดเลือดในครอบครัวอยู่ระหว่างร้อยละ 29.1 ถึง 70.9

**หมายเหตุ :** การรายงานช่วงเชื่อมั่นจะมีความเหมาะสม หากการศึกษามีวิธีการศึกษาและสุ่มตัวอย่างที่เหมาะสมสำหรับการตอบคำถามการวิจัยในข้อนี้ ไม่เช่นนั้น ถึง Epi Info จะแสดงผลออกมา ก็ไม่เหมาะสมจะนำไปใช้

* 1. **การทดสอบความแตกต่างของสัดส่วน 2 กลุ่ม**

*ตัวอย่าง* : การวิเคราะห์ความแตกต่างของสัดส่วนการมีอาการเจ็บหน้าอก จำแนกตามการมีประวัติโรคหัวใจขาดเลือดในครอบครัว และทดสอบความแตกต่างของสัดส่วนดังกล่าวทางสถิติ

* + การสร้างตารางแจกแจงความถี่แสดงสัดส่วนการมีอาการเจ็บหน้าอก จำแนกตามการมีประวัติโรคหัวใจขาดเลือดในครอบครัว แสดงค่าสัดส่วนเป็นร้อยละตามการมีประวัติ และทดสอบความแตกต่างของสัดส่วน



2) กด drop down menuเลือกตัวแปร IHDFAMI

4) Click OK

3) กด drop down menuเลือกตัวแปร Angina

1) เลือก Tables

command ที่ใช้จะปรากฏใน Program Editor window อาจพิมพ์เองโดยตรงใน Program Editor หากจำคำสั่งและชื่อตัวแปรได้ แล้วกด Run This Command ที่แถบเมนู

**TABLES IHDFAMI ANGINA** ↵

|  |  |  |  |
| --- | --- | --- | --- |
| **ANGINA** | | | |
| **IHDFAMI** | **0** | **1** | **TOTAL** |
| **0** Row % Col % | 412 83.4 97.2 | 82 16.6 87.2 | 494 100.0 95.4 |
| **1** Row % Col % | 12 50.0 2.8 | 12 50.0 12.8 | 24 100.0 4.6 |
| **TOTAL** Row % Col % | 424 81.9 100.0 | 94 18.1 100.0 | 518 100.0 100.0 |

|  |  |
| --- | --- |
| RED | ORANGE |
| YELLOW | GREEN |

**Single Table Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Point | 95% Confidence Interval | |
|  | Estimate | Lower | Upper |
| PARAMETERS: Odds-based |  |  |  |
| Odds Ratio (cross product) | 5.0244 | 2.1810 | 11.5748 (T) |
| Odds Ratio (MLE) | 5.0019 | 2.1324 | 11.7406 (M) |
|  |  | 1.9793 | 12.6508 (F) |
| PARAMETERS: Risk-based |  |  |  |
| Risk Ratio (RR) | 1.6680 | 1.1159 | 2.4934 (T) |
| Risk Difference (RD%) | 33.4008 | 13.1293 | 53.6723 (T) |
|  |  |  |  |
| (T=Taylor series; C=Cornfield; M=Mid-P; F=Fisher Exact) | | | |
|  |  |  |  |
| STATISTICAL TESTS | Chi-square | 1-tailed p | 2-tailed p |
| Chi-square - uncorrected | 17.1905 |  | 0.0000349817 |
| Chi-square - Mantel-Haenszel | 17.1573 |  | 0.0000355776 |
| Chi-square - corrected (Yates) | 15.0154 |  | 0.0001078082 |
| Mid-p exact |  | 0.0001574785 |  |
| Fisher exact |  | 0.0002675274 |  |

Warning: The expected value of a cell is <5. Fisher Exact Test should be used.

* + โปรแกรม Epi Info จะให้ผลลัพธ์ออกมาทั้งหมดหลายชนิด สัดส่วนจำแนกตามแนวตั้ง สัดส่วนจำแนกตามแนวนอน (การเลือกนำเสนอตัวใด ขึ้นอยู่กับรูปแบบการศึกษา) ค่า Odds ratio (และ 95%CI) ค่า Risk ratio (และ 95%CI) ค่า p value (χ2 ทั้ง 1 tailed และ 2 tailed , Mid – p exact, fisher exact) ในที่นี้ขอให้ดูเพียงส่วนท้ายตารางเรื่อง Statistical tests จะเห็นได้ว่ามี ค่า p value จากการทดสอบ 3 ชนิด โดยทั่วๆ ไปแล้ว หากขนาดตัวอย่างมากพอ และจำนวน expected value ภายใต้ Null hypothesis ที่มีค่าน้อยกว่า 5 มีไม่เกิน 20% ของจำนวน expected frequency ทั้งหมด (ในกรณี 2 x 2 table คือ ต้องไม่มีเลยซักเซลล์เดียว) มักจะนิยมใช้ uncorrected χ2 (แต่ที่เหมาะสม อาจจะใช้ Yates-corrected Chi-square) แต่ถ้าหากขนาดตัวอย่างน้อยหรือ จำนวน expected frequency ภายใต้ Null hypothesis ที่มีค่าน้อยกว่า 5 มีเกิน 20% ของจำนวน expected frequency ทั้งหมด (ในกรณี 2 x 2 table คือ มีอย่างน้อยหนึ่งเซลล์) ควรจะใช้ Fisher exact p-value ซึ่งโปรแกรม Epi Info จะเตือนหากพบกรณีดังกล่าว ดังแสดงอยู่ที่ท้ายตาราง ในกรณีนี้จึงต้องใช้ Fisher exact p-value แต่ใน output จะเป็น 1-tailed p หากจะทำให้เป็น 2-tailed p ต้องเอา p ไปคูณสอง ดังนั้น Fisher exact p-value ชนิด 2-tailed = 0.0002675274\*2 = 0.0005350548 หรือ <0.001
  + แปลผลได้ว่า หากในความเป็นจริง กลุ่มที่ไม่มีประวัติโรคหัวใจขาดเลือดในครอบครัวมีสัดส่วนการมีอาการเจ็บหน้าอกไม่แตกต่างไปจากกลุ่มที่มีประวัติโรคหัวใจขาดเลือดในครอบครัว การใช้ข้อมูลนี้สรุปว่าสัดส่วนดังกล่าวมีความแตกต่างกัน จะมีโอกาสสรุปผิดไปจากความเป็นจริง น้อยกว่า ร้อยละ 0.1
  + แปลผลอย่างง่ายได้ว่า กลุ่มที่ไม่มีประวัติโรคหัวใจขาดเลือดในครอบครัวมีสัดส่วนการมีอาการเจ็บหน้าอกแตกต่างจากกลุ่มที่มีประวัติโรคหัวใจขาดเลือดในครอบครัว อย่างมีนัยสำคัญทางสถิติที่ α 0.05 (p value < 0.001)

*แบบฝึกหัด*

* + แจกแจงสัดส่วนของการมีอาการเจ็บหน้าอกจำแนกตามการเป็นเบาหวานและทดสอบความแตกต่างของสัดส่วน
  + แจกแจงสัดส่วนของการมีอาการเจ็บหน้าอกจำแนกตามการเพศและทดสอบความแตกต่างของสัดส่วน
  + แจกแจงสัดส่วนของการเป็นเบาหวาน จำแนกตามเพศ และทดสอบความแตกต่างของสัดส่วน
  + แจกแจงสัดส่วนของการมีประวัติโรคหัวใจขาดเลือดในครอบครัวจำแนกตามเพศและทดสอบความแตกต่างของสัดส่วน

**การวิเคราะห์ค่าเฉลี่ย**

การวิเคราะห์สัดส่วนใช้เมื่อข้อมูลเป็น Numerical data ซึ่งสรุปข้อมูลโดยใช้ค่าเฉลี่ย (mean)

**แผนภูมีที่ 2** การวิเคราะห์ค่าเฉลี่ย

การวิเคราะห์ค่าเฉลี่ย

กลุ่มเดียว

สองกลุ่ม

สามกลุ่มขึ้นไป

อิสระต่อกัน

ไม่อิสระต่อกัน

อิสระต่อกัน

ไม่อิสระต่อกัน

* , 95%CI
* Repeated measure ANOVA
* Wilcoxon Matched pairs Signed Ranks Test
* Kruskall – Wallis Test
* Mann – Whitney Test

**Nonparametric statistics**

* 1– 2 , 95%CI
* 2 sample t test
* 1– 2 , 95%CI
* pair t test
* 1 - 2 , 95%CI

1 - 2 , 95%CI

1 - 2 , 95%CI

* one way ANOVA

********

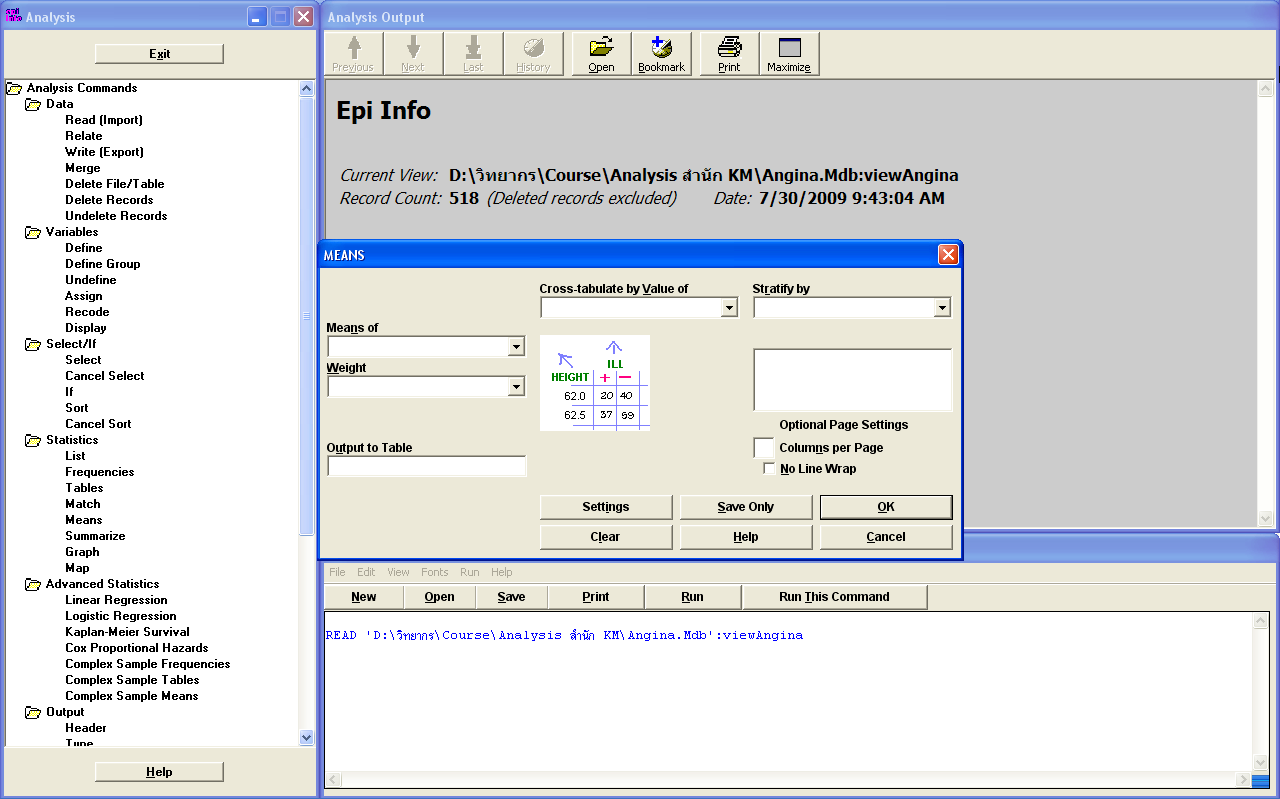
****

****

*ที่มา: บัณฑิต ถิ่นคำรพ.   คู่มือปฏิบัติการชีวสถิติ: ตำราสำหรับเรียนรู้ชีวสถิติด้วยตนเอง*

1. **การแสดงค่าเฉลี่ย 1 กลุ่ม**

*ตัวอย่าง* : การแสดงค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของอายุ



3) Click OK

2) กด drop down menuเลือกตัวแปร Angina

1) เลือก Means

command ที่ใช้จะปรากฏใน Program Editor window อาจพิมพ์เองโดยตรงใน Program Editor หากจำคำสั่งและชื่อตัวแปรได้ แล้วกด Run This Command ที่แถบเมนู

**MEANS AGE** ↵

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **AGE** | **Frequency** | **Percent** | **Cum Percent** |  |
| **25** | 1 | 0.2% | 0.2% |  |
| **26** | 2 | 0.4% | 0.6% |  |
| **27** | 2 | 0.4% | 1.0% |  |
| **28** | 2 | 0.4% | 1.4% |  |
| **29** | 2 | 0.4% | 1.7% |  |
| **30** | 3 | 0.6% | 2.3% | YELLOW |
| **31** | 2 | 0.4% | 2.7% |  |
| **32** | 2 | 0.4% | 3.1% |  |
| **33** | 3 | 0.6% | 3.7% | YELLOW |
| **34** | 2 | 0.4% | 4.1% |  |
| **35** | 4 | 0.8% | 4.8% | YELLOW |
| **36** | 2 | 0.4% | 5.2% |  |
| **37** | 3 | 0.6% | 5.8% | YELLOW |
| **38** | 4 | 0.8% | 6.6% | YELLOW |
| **39** | 4 | 0.8% | 7.3% | YELLOW |
| **40** | 6 | 1.2% | 8.5% | YELLOW |
| **41** | 12 | 2.3% | 10.8% | YELLOW |
| **42** | 21 | 4.1% | 14.9% | YELLOW |
| **43** | 28 | 5.4% | 20.3% | YELLOW |
| **44** | 46 | 8.9% | 29.2% | YELLOW |
| **45** | 39 | 7.5% | 36.7% | YELLOW |
| **46** | 37 | 7.1% | 43.8% | YELLOW |
| **47** | 30 | 5.8% | 49.6% | YELLOW |
| **48** | 18 | 3.5% | 53.1% | YELLOW |
| **49** | 28 | 5.4% | 58.5% | YELLOW |
| **50** | 29 | 5.6% | 64.1% | YELLOW |
| **51** | 18 | 3.5% | 67.6% | YELLOW |
| **52** | 25 | 4.8% | 72.4% | YELLOW |
| **53** | 21 | 4.1% | 76.4% | YELLOW |
| **54** | 19 | 3.7% | 80.1% | YELLOW |
| **55** | 19 | 3.7% | 83.8% | YELLOW |
| **56** | 13 | 2.5% | 86.3% | YELLOW |
| **57** | 11 | 2.1% | 88.4% | YELLOW |
| **58** | 10 | 1.9% | 90.3% | YELLOW |
| **59** | 11 | 2.1% | 92.5% | YELLOW |
| **60** | 8 | 1.5% | 94.0% | YELLOW |
| **61** | 13 | 2.5% | 96.5% | YELLOW |
| **62** | 3 | 0.6% | 97.1% | YELLOW |
| **63** | 4 | 0.8% | 97.9% | YELLOW |
| **64** | 2 | 0.4% | 98.3% |  |
| **65** | 2 | 0.4% | 98.6% |  |
| **66** | 2 | 0.4% | 99.0% |  |
| **67** | 1 | 0.2% | 99.2% |  |
| **68** | 3 | 0.6% | 99.8% | YELLOW |
| **69** | 1 | 0.2% | 100.0% |  |
| **Total** | 518 | 100.0% | 100.0% | ORANGE |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Obs | Total | Mean | Variance | Std Dev |
|  | 518 | 25081.0000 | 48.4189 | 54.4683 | 7.3803 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Minimum | 25% | Median | 75% | Maximum | Mode |
|  | 25.0000 | 44.0000 | 48.0000 | 53.0000 | 69.0000 | 44.0000 |

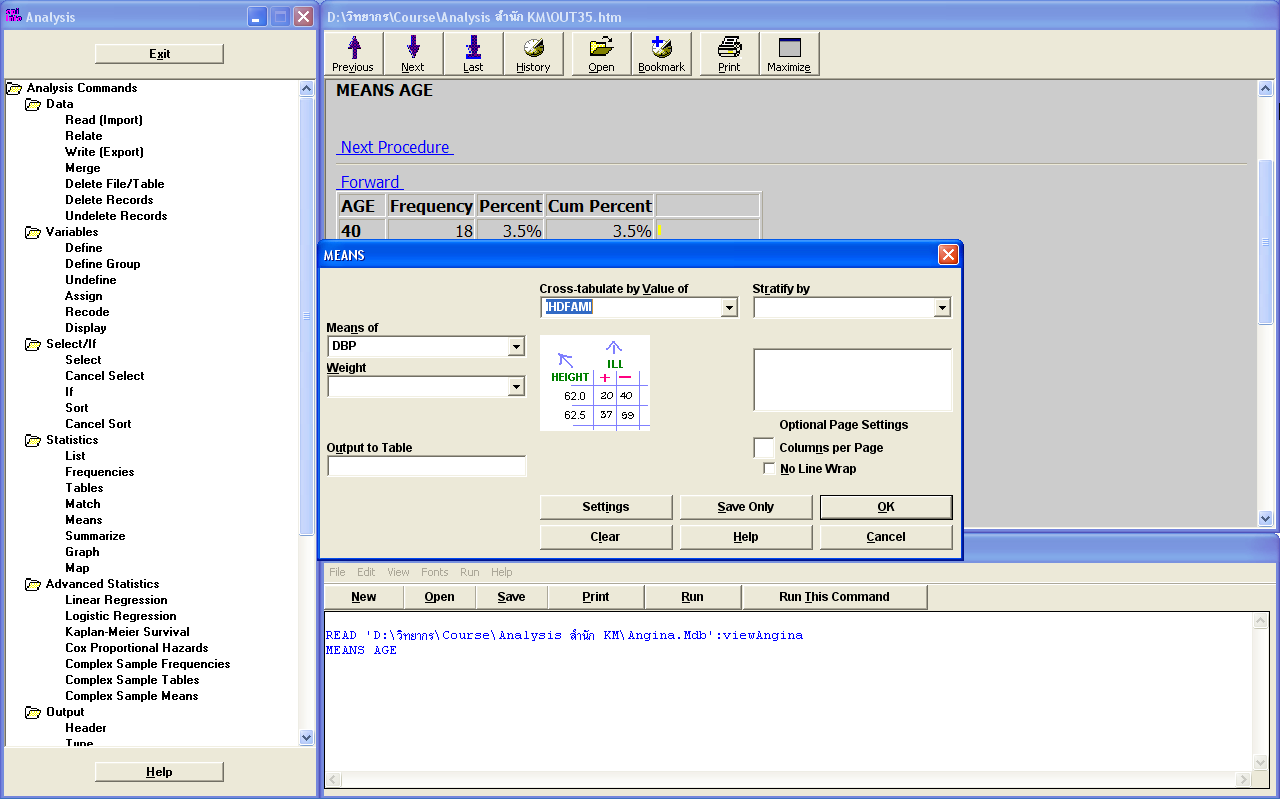
* + ค่าเฉลี่ยเลขคณิตของอายุของประชากรตัวอย่างในการศึกษานี้เท่ากับ 48.4 ปี ส่วนเบี่ยงเบนมาตรฐาน 7.4 ปี ค่ามัธยฐานเท่ากับ 48 ปี ค่าต่ำสุด 25 ปี และสูงสุด 69 ปี ตามลำดับ

*แบบฝึกหัด*

* + ค่าเฉลี่ยของความดัน SBP
  + ค่าเฉลี่ยของความดัน DBP

1. **การเปรียบเทียบค่าเฉลี่ย 2 กลุ่ม**

*ตัวอย่าง* : การเปรียบเทียบค่าเฉลี่ยของความดันโลหิต DBP ระหว่างกลุ่มที่มีกับไม่มีประวัติโรคหัวใจขาดเลือดในครอบครัว และทดสอบความแตกต่างของค่าเฉลี่ยดังกล่าวทางสถิติ



3) กด drop down menuเลือกตัวแปร IHDFAMI

2) กด drop down menuเลือกตัวแปร DBP

4) Click OK

1) เลือก Means

command ที่ใช้จะปรากฏใน Program Editor window อาจพิมพ์เองโดยตรงใน Program Editor หากจำคำสั่งและชื่อตัวแปรได้ แล้วกด Run This Command ที่แถบเมนู

**MEANS DBP IHDFAMI** ↵

#### Descriptive Statistics for Each Value of Crosstab Variable

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Obs | Total | Mean | Variance | Std Dev |
| 0 | 494 | 35808.0000 | 72.4858 | 110.0394 | 10.4900 |
| 1 | 24 | 1901.0000 | 79.2083 | 133.3895 | 11.5494 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Minimum | 25% | Median | 75% | Maximum | Mode |
| 0 | 50.0000 | 65.0000 | 75.0000 | 80.0000 | 120.0000 | 80.0000 |
| 1 | 60.0000 | 75.0000 | 80.0000 | 83.5000 | 110.0000 | 80.0000 |

#### ANOVA, a Parametric Test for Inequality of Population Means

(For normally distributed data only)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variation | SS | df | MS | F statistic |
| Between | 1034.3571 | 1 | 1034.3571 | 9.3118 |
| Within | 57317.3591 | 516 | 111.0802 |  |
| Total | 58351.7162 | 517 |  |  |

T Statistic = 3.0515

P-value = 0.0024

#### Bartlett's Test for Inequality of Population Variances

|  |  |  |  |
| --- | --- | --- | --- |
| Bartlett's chi square= | 0.4254 | df=1 | P value=0.5142 |

A small p-value (e.g., less than 0.05 suggests that the variances are not homogeneous and that the ANOVA may not be appropriate.

#### Mann-Whitney/Wilcoxon Two-Sample Test (Kruskal-Wallis test for two groups)

|  |  |
| --- | --- |
| Kruskal-Wallis H (equivalent to Chi square) = | 9.5455 |
| Degrees of freedom = | 1 |
| P value = | 0.0020 |

*ผลที่แสดงออกมาจะประกอบด้วย 5 ส่วนหลักๆ คือ*

* ตารางที่มีตัวแปร continuous อยู่แนวนอน และตัวแปรแบ่งกลุ่มอยู่ในแนวสดมภ์ (ในที่นี้ไม่ได้นำมาแสดงให้ดู)
* **Descriptive Statistics for Each Value of Crosstab Variable** : ผลวิเคราะห์ระบาดวิทยาเชิงพรรณนาของตัวแปร continuous ในแต่ละกลุ่ม ซึ่งได้แก่ จำนวนข้อมูล, ค่าเฉลี่ย(mean), ค่าความแปรปรวน(variance), ค่า standard deviation, ค่าต่ำสุดสูงสุด, ค่าของข้อมูลอันดับที่ 25th กับ 75th , ค่ามัธยฐาน และค่าฐานนิยม
* **ANOVA, a Parametric Test for Inequality of Population Means** : ตาราง Analysis of variance (ANOVA) และ ค่า p value ที่บ่งบอกว่าค่าเฉลี่ยของทั้งสองกลุ่มแตกต่างกันหรือไม่
* **Bartlett's Test for Inequality of Population Variances** : การทดสอบทางสถิติว่าค่าความแปรปรวนของสองกลุ่มแตกต่างกันหรือไม่ เนื่องจากมีผลต่อการเลือกใช้สถิติมาทดสอบว่า จะใช้สถิติกลุ่ม parametric หรือ non – parametric
* **Mann-Whitney/Wilcoxon Two-Sample Test (Kruskal-Wallis test for two groups)** : การทดสอบทางสถิติโดยวิธี non – parametric ซึ่งมีหลักการแปลผลเหมือน independent t – test กับ one – way ANOVA
  + ก่อนอื่น ต้องพิจารณาว่าการกระจายของข้อมูลของแต่ละกลุ่มมีการกระจายแบบ normal distribution หรือไม่
    1. Test of normality เช่น Shapiro-Wilk test แต่ Epiinfo ไม่มี test of normality
    2. ใช้ plot ดูการกระจายของข้อมูล เช่น histogram หรือ scatter plot

หากพิจารณาแล้วมีการกระจายแบบ normal distribution ก็พิจารณาต่อเรื่องความแปรปรวนของแต่ละกลุ่มแตกต่างกันหรือไม่ หากไม่เป็น normal distribution อย่างน้อยหนึ่งกลุ่ม ก็ให้ไปใช้ non-parametric test (Kruskal-Wallis test) เลย

* + จากนั้นมาดูหนึ่งในเงื่อนไขของการใช้สถิติ ANOVA ว่าความแปรปรวนของแต่ละกลุ่มควรมีค่าใกล้เคียงกัน วิธีการหนึ่งที่จะทดสอบเงื่อนไขนี้เป็นจริงหรือไม่คือการใช้ Bartlett’s test for homogeneity of variance ซึ่ง Epi Info แสดงผล Bartlett’s chi square กับ degree of freedom และ p-value พร้อมข้อแนะนำในการเลือกใช้ค่าสถิติที่เหมาะสมว่าควรจะเป็น ANOVA หรือ non – parametric กันแน่ จากตัวอย่าง p value สูงมากกว่า 0.05 (p value = 0.514) นั่นคือ ค่าความแปรปรวนน่าจะไม่แตกต่างกัน จึงเลือกใช้ ANOVA test (ในกรณีที่ความแปรปรวนแตกต่างกันอย่างมีนัยสำคัญทางสถิติ หรือ Bartlett’s chi square p-value < 0.05 ควรเลือกใช้ผลของ non – parametric แทน คือ Kruskal-Wallis test p-value = 0.002)
  + แปลผลได้ว่า หากในความเป็นจริง กลุ่มที่ไม่มีประวัติโรคหัวใจขาดเลือดในครอบครัวมีค่าเฉลี่ยความดันโลหิต DBP ไม่แตกต่างไปจากกลุ่มที่มีประวัติโรคหัวใจขาดเลือดในครอบครัว การใช้ข้อมูลนี้สรุปว่าค่าเฉลี่ยความดันโลหิต DBP ดังกล่าวมีความแตกต่างกัน จะมีโอกาสสรุปผิดไปจากความเป็นจริง ร้อยละ 0.2
  + แปลผลอย่างง่ายได้ว่า ค่าเฉลี่ยความดันโลหิต DBP ของกลุ่มที่ไม่มีประวัติโรคหัวใจขาดเลือดในครอบครัวต่างจากกลุ่มที่มีประวัติโรคหัวใจขาดเลือดในครอบครัว อย่างมีนัยสำคัญทางสถิติที่ α 0.05 (p value = 0.002)

*แบบฝึกหัด*

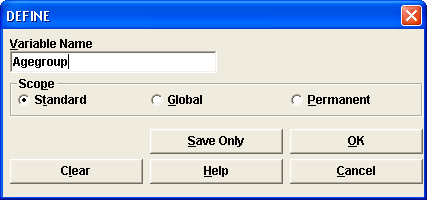
* + ทดสอบความแตกต่างของค่าเฉลี่ยความดัน SBP ระหว่างกลุ่มที่ไม่มีและมีประวัติโรคหัวใจขาดเลือดในครอบครัว
  + ทดสอบความแตกต่างของค่าเฉลี่ยของอายุ ระหว่างเพศชายและหญิง
  + ทดสอบความแตกต่างของค่าเฉลี่ยของน้ำหนักตัว ระหว่างเพศชายและหญิง
  + ทดสอบความแตกต่างของค่าเฉลี่ยของส่วนสูง ระหว่างเพศชายและหญิง
  + ทดสอบความแตกต่างของค่าเฉลี่ยความดัน SBP ระหว่างเพศชายและหญิง
  + ทดสอบความแตกต่างของค่าเฉลี่ยความดัน DBP ระหว่างเพศชายและหญิง
  + ทดสอบความแตกต่างของค่าเฉลี่ยความดัน SBP ระหว่างผู้ที่ไม่มีและมีอาการเจ็บหน้าอก
  + ทดสอบความแตกต่างของค่าเฉลี่ยความดัน DBP ระหว่างผู้ที่ไม่มีและมีอาการเจ็บหน้าอก
  + ทดสอบความแตกต่างของค่าเฉลี่ยของน้ำหนักตัว ระหว่างผู้ที่ไม่มีและมีอาการเจ็บหน้าอก

**การจัดการข้อมูล**

**Define**

คำสั่งนี้ใช้เพื่อ สร้างตัวแปรใหม่ เพื่อเก็บค่าที่จะสร้างขึ้นมาใหม่ หรือจากค่าตัวแปรอื่นๆ

*ตัวอย่างการสร้างตัวแปรใหม่ชื่อ Agegroup เพื่อจัดกลุ่มอายุจากตัวแปร Age*



2) Click

1) พิมพ์ชื่อตัวแปรใหม่ที่ต้องการ

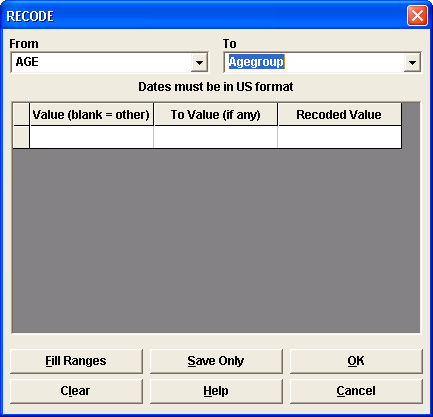
**Undefine**

ใช้เพื่อลบตัวแปรที่สร้างมาจากคำสั่ง Define

**Recode**

ใช้เพื่อเปลี่ยนค่าจากตัวแปรเดิมหนึ่งตัว มาเป็นค่าใหม่ในตัวแปรที่สร้างจากคำสั่ง Define

1) Browse ตัวแปรเดิม



5) Click

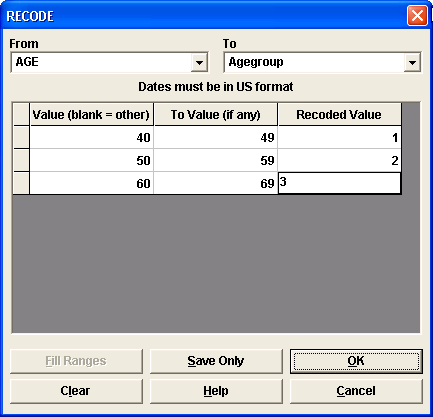
4) พิมพ์ค่าของ ตัวแปรใหม่แล้วกดปุ่ม Enter เพื่อเปลี่ยนค่าอื่นจากตัวแปรเดิมอีก

3.2) พิมพ์ค่าของ ตัวแปรตัวแปรเดิม (ใช้ในกรณีที่ต้องการเปลี่ยนค่าตัวแปรเดิมเป็นช่วงข้อมูล)

3.1) พิมพ์ค่าของ ตัวแปรเดิม

2) Browse ตัวแปรใหม่

*ตัวอย่าง : การจัดกลุ่มอายุจากตัวแปร Age มาใส่ในตัวแปร Agegroup*



Age 40 ถึง 49 ให้เป็น Agegroup 1

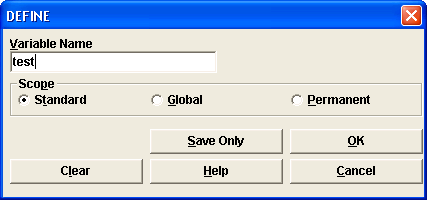
Age 50 ถึง 59 ให้เป็น Agegroup 2

Age 60 ถึง 69 ให้เป็น Agegroup 3

**Assign**

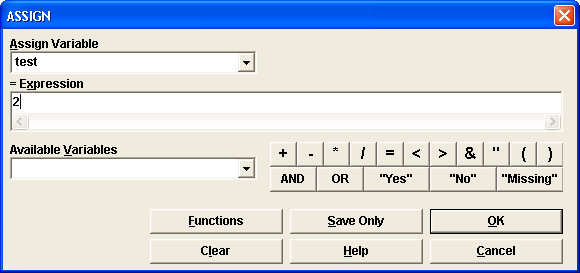
ใช้ตามหลังคำสั่ง Define เช่นเดียวกับ คำสั่ง Recode แต่เป็นการให้ค่าตัวแปรที่สร้างใหม่หรือ เปลี่ยนค่าจากตัวแปรเดิมโดยมีเงื่อนไขที่ซับซ้อนได้มากกว่าคำสั่ง Recode

*ตัวอย่างการสร้างตัวแปรใหม่ชื่อ Test แล้วให้ค่าเป็น 2*



2) Click

1) พิมพ์ชื่อตัวแปรใหม่ที่ต้องการ



3) Click

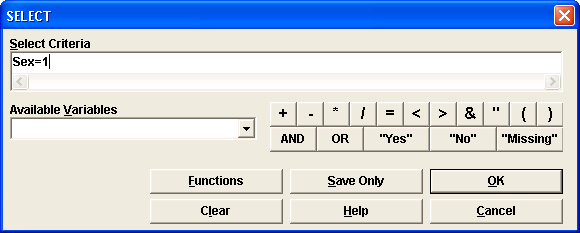
2) พิมพ์ค่าที่ต้องการให้ค่า

1) Browse ตัวแปรใหม่

**Select**

เป็นคำสั่งที่เลือกบาง records มาวิเคราะห์โดยการให้เงื่อนไขกับตัวแปรที่ต้องการ

*ตัวอย่างการเลือกเฉพาะเพศชาย*

****

2) Click

1) พิมพ์ชื่อตัวแปรและเงื่อนไข

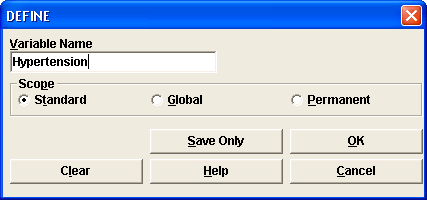
**Cancel Select**

ใช้เพื่อยกเลิกคำสั่ง Select

**If**

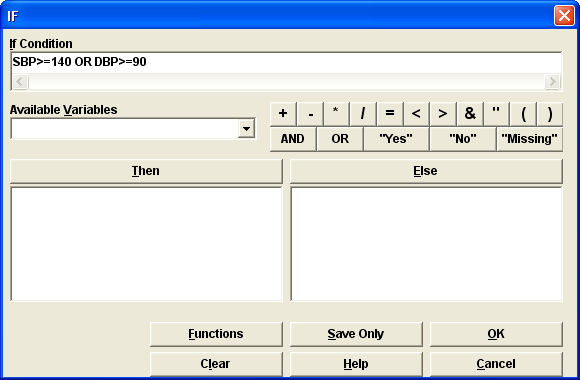
เป็นคำสั่งที่สร้างเงื่อนไข มักใช้ร่วมกับคำสั่ง Assign เพื่อสร้างเงื่อนไขจากตัวแปรเดิมมากกว่าหนึ่งตัวมาเก็บเป็นค่าใหม่ในตัวแปรใหม่ที่สร้างจากคำสั่ง Define

*ตัวอย่าง : การสร้างตัวแปรใหม่ชื่อ Hypertension แล้วให้ค่าเป็น 1 หากค่า SPB>=140* ***หรือ*** *DBP>=90 ให้ค่าเป็น 2 หาก* *SPB<140* ***และ*** *DBP<90*



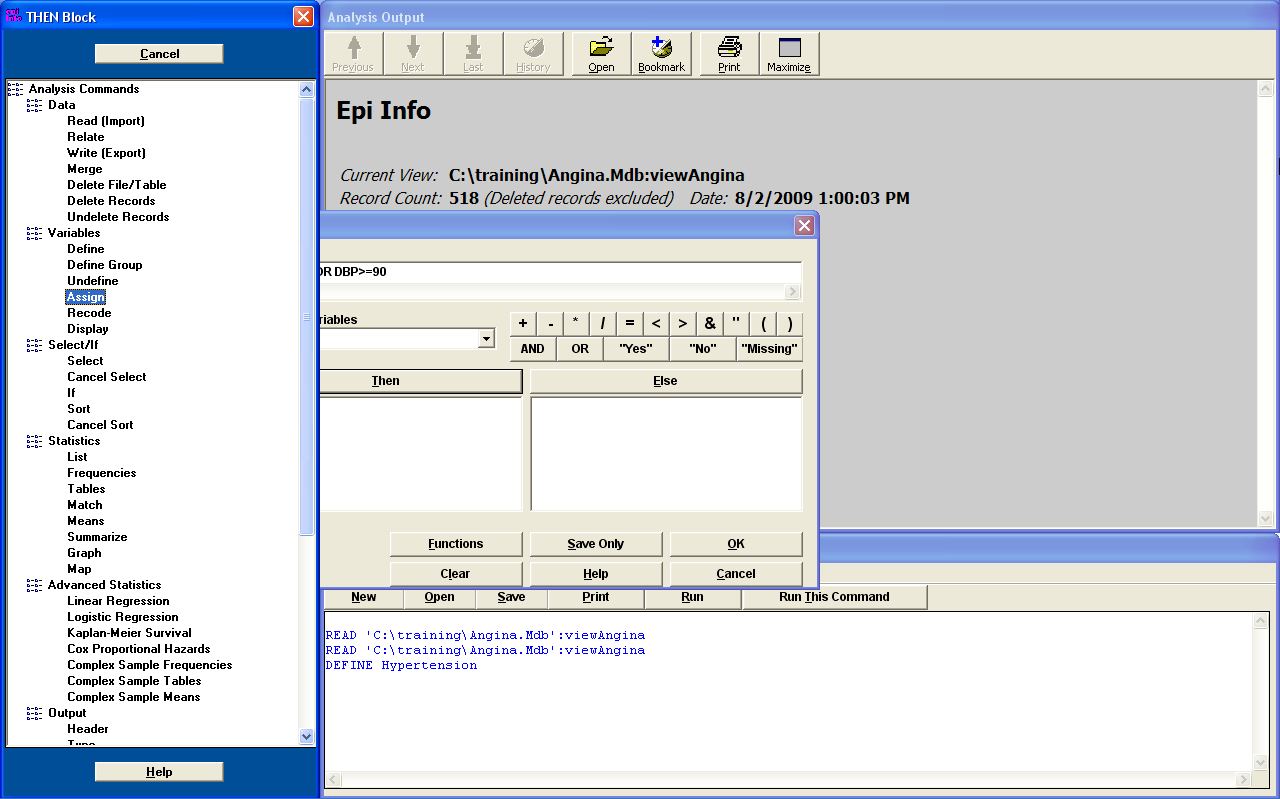
2) Click

1) พิมพ์ชื่อตัวแปรใหม่ที่ต้องการ

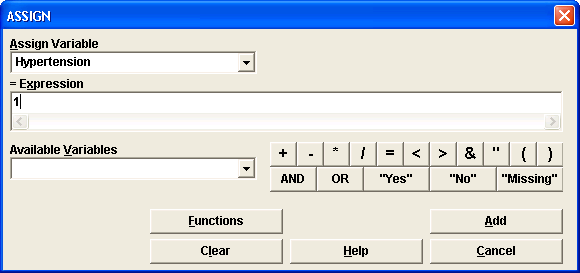


2) Click เพื่อไปใช้คำสั่ง Assign ให้ค่าตัวแปร Hypertension ตามเงื่อนไข

1) พิมพ์เงื่อนไขที่ต้องการ



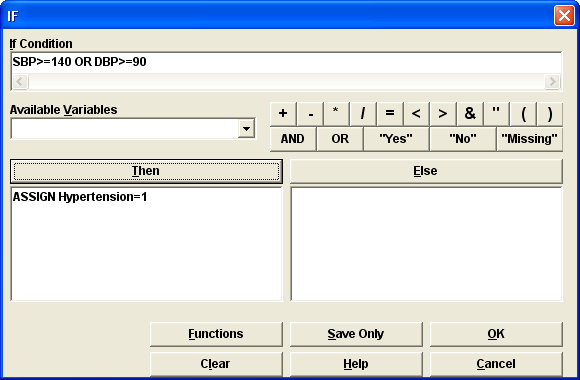
Click เลือก Assign



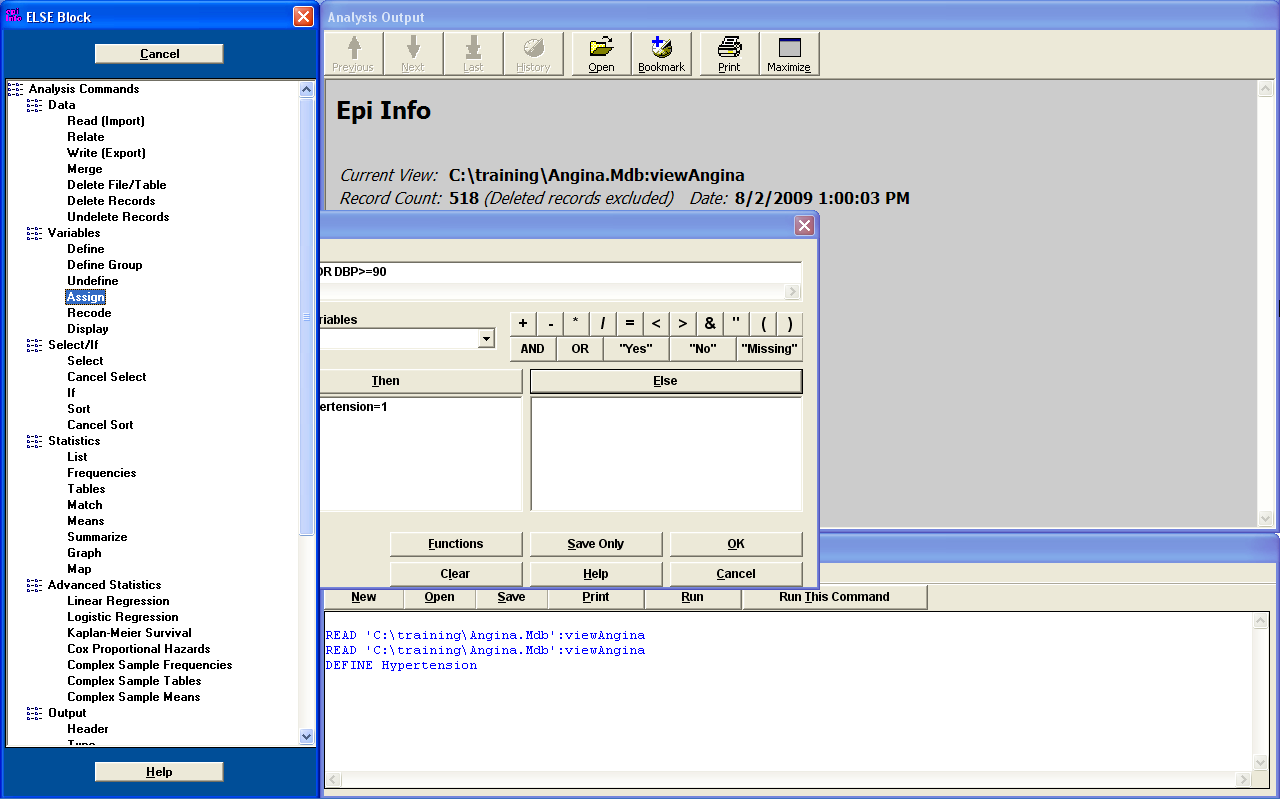
3) Click

2) พิมพ์ค่าที่ต้องการ

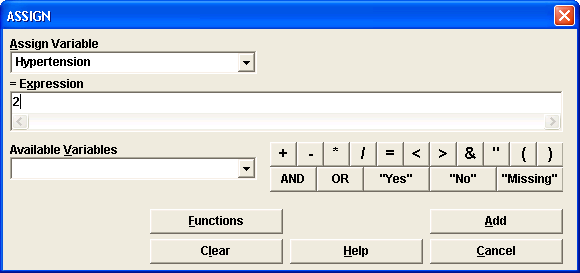
1) Click เลือกตัวแปร



Click เพื่อไปใช้คำสั่ง Assign ให้ค่าตัวแปร Hypertension ตามเงื่อนไขที่เหลือ



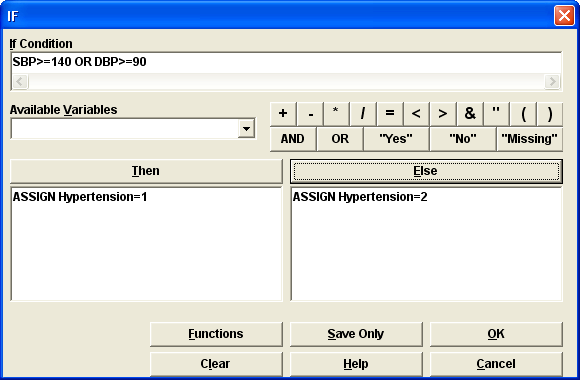
Click เลือก Assign



1) Click เลือกตัวแปร

3) Click

2) พิมพ์ค่าที่ต้องการ



Click

หากพิมพ์คำสั่งเองที่ Program Editor จะเป็นดังนี้

IF SBP>=140 OR DBP>=95 THEN

ASSIGN Hypertension=1

ELSE

ASSIGN Hypertension=2

END

*หมายเหตุ :* ควรทำการตรวจสอบตัวแปรที่สร้างใหม่โดยใช้คำสั่ง List หรือ Tables กับตัวแปรเก่าและตัวแปรใหม่ เพื่อดูว่าให้ค่าได้ถูกต้องหรือไม่

**การวิเคราะห์ขนาดของความสัมพันธ์**

การวัดขนาดของความสัมพันธ์ระหว่างปัจจัย (Exposure) กับผล (Outcome) ศึกษาได้จากรูปแบบการศึกษาระบาดวิทยาเชิงวิเคราะห์

* Cohort study : การวัดผลใช้ risk ratio (RR)

*ตัวอย่าง การวิเคราะห์*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Outcome (+ve) | Outcome (-ve) |  |
| Exposed | A | B | A + B |
| Unexposed | C | D | C + D |

RR = Risk ในกลุ่ม exposed = A / (A + B)

Risk ในกลุ่ม unexposed C / (C + D)

* Case-control study : การวัดผลใช้ odds ratio (OR)

*ตัวอย่าง การวิเคราะห์*

|  |  |  |
| --- | --- | --- |
|  | Case | Control |
| Exposed | A | B |
| Unexposed | C | D |
|  | A + C | B + D |

OR = Odds ของการ exposed ในกลุ่ม Case = A / C = A\*D

Odds ของการ exposed ในกลุ่ม Control B / D B\*C

* Cross-sectional study : การวัดผลใช้ prevalent ratio (PR)

*ตัวอย่าง การวิเคราะห์*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Outcome (+ve) | Outcome (-ve) |  |
| Exposed | A | B | A + B |
| Unexposed | C | D | C + D |

PR = Prevalence ในกลุ่ม exposed = A / (A + B)

Prevalence ในกลุ่ม unexposed C / (C + D)

หมายเหตุ : การใช้สูตรทางสัญลักษณ์ที่ประกอบด้วย A, B, C, D จำเป็นต้องตั้งตาราง Exposure และ Outcome ดังแบบที่แสดงด้านบนด้วย

**การวิเคราะห์ความสัมพันธ์โดยใช้โปรแกรม Epi Info**

โปรแกรม Epi Info จะตั้งตาราง 2 x 2 โดยมีหัวตาราง ตาม Code ค่าของตัวแปรดังนี้

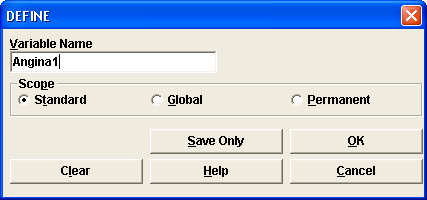
|  |  |  |
| --- | --- | --- |
|  | Code เลขน้อย | Code เลขมาก |
| Code เลขน้อย | A | B |
| Code เลขมาก | C | D |

การคำนวณ RR, OR, PR จะใช้ สูตรทางสัญลักษณ์ที่ประกอบด้วย A, B, C, D ดังที่กล่าวมา กับการตั้งตารางโดยใช้ Code ค่าของตัวแปร เลขน้อยแทนการมีปัจจัยหรือการมี outcome ส่วนเลขมากแทนการไม่มีปัจจัยหรือการไม่มี outcome ดังนั้น หากใช้ Epi Info วิเคราะห์หาความสัมพันธ์ ตัวแปรปัจจัย และ Outcome ต้องเป็น Dichotomous และตั้ง Code ค่าของตัวแปร เลขน้อยแทนการมีปัจจัยหรือการมี outcome ส่วนเลขมากแทนการไม่มีปัจจัยหรือการไม่มี outcome

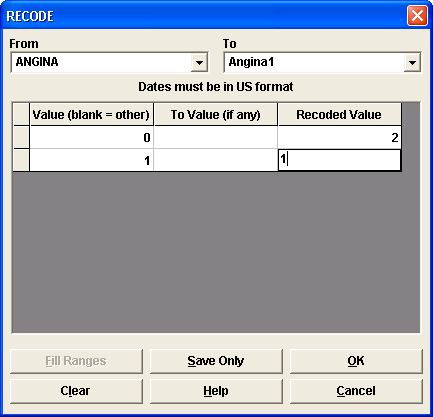
*ตัวอย่าง : การวิเคราะห์ความสัมพันธ์ระหว่างการเป็นเบาหวานกับการมีอาการเจ็บหน้าอก*

ในตัวอย่างนี้ Diabetes และ Angina มี code 0 (เลขน้อย) แทนการไม่มี ส่วน code 1 (เลขมาก) แทนการมี ดังนั้นจึงต้องสร้างตัวแปรใหม่แล้ว Recode ตัวแปรเดิมให้อยู่ในรูปแบบที่เหมาะสมกับการวิเคราะห์

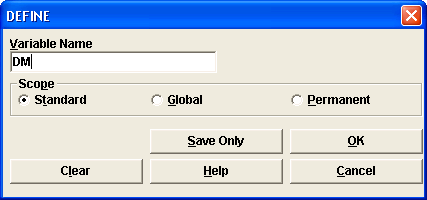
การสร้างตัวแปรใหม่ Angina1



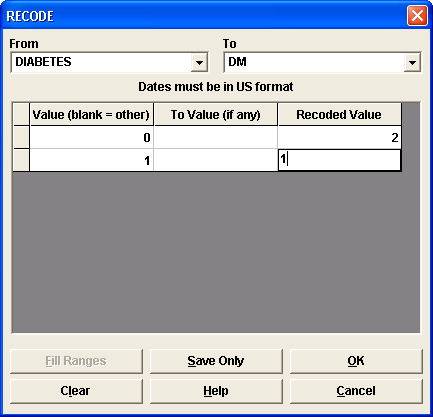
การ Recode ตัวแปรจาก Angina เป็น Angina1 โดยค่าเดิม 0 เปลี่ยนเป็น 2 ส่วนค่า 1 เป็น 1 ดังเดิม



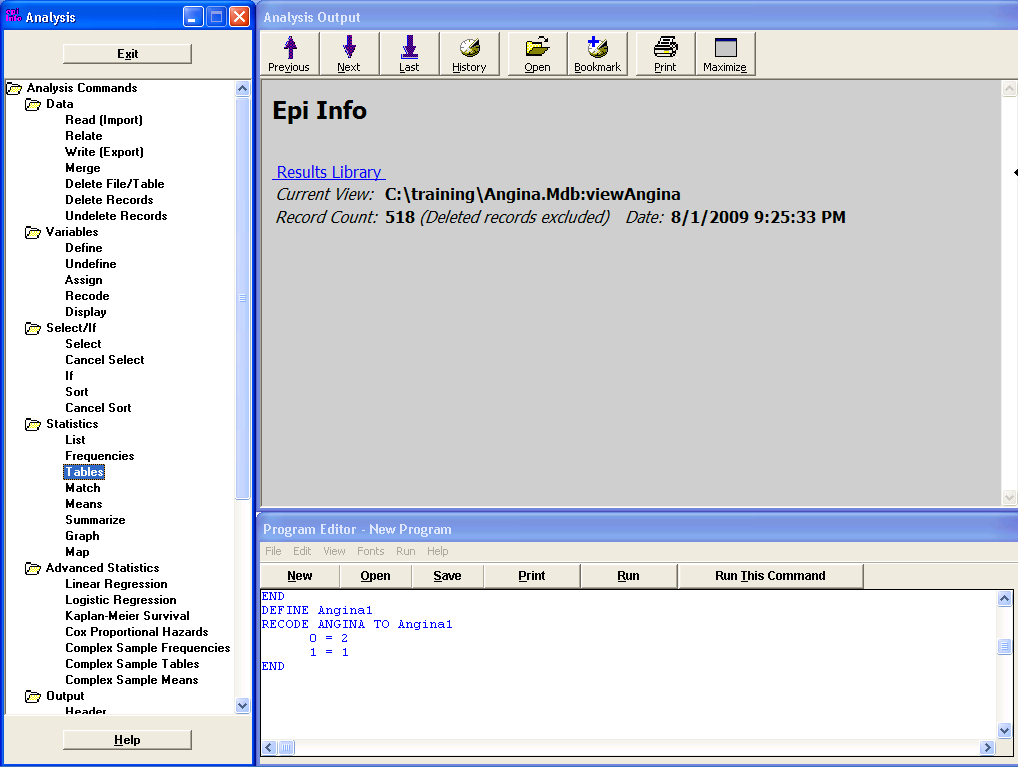
การสร้างตัวแปรใหม่ DM



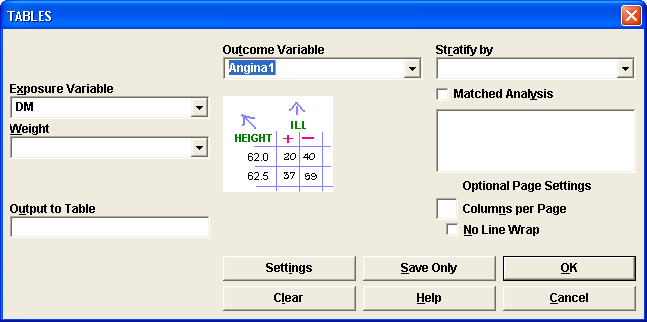
การ Recode ตัวแปรจาก Diabetes เป็น DM โดยค่าเดิม 0 เปลี่ยนเป็น 2 ส่วนค่า 1 เป็น 1 ดังเดิม



การใช้คำสั่ง Tables



Click เลือก Tables



3) Click

2) Click เลือกตัวแปร

1) Click เลือกตัวแปร

|  |  |  |  |
| --- | --- | --- | --- |
| **ANGINA1** | | | |
| **DM** | **1** | **2** | **TOTAL** |
| **1** Row % Col % | 10 38.5 10.6 | 16 61.5 3.8 | 26 100.0 5.0 |
| **2** Row % Col % | 84 17.1 89.4 | 408 82.9 96.2 | 492 100.0 95.0 |
| **TOTAL** Row % Col % | 94 18.1 100.0 | 424 81.9 100.0 | 518 100.0 100.0 |

|  |  |
| --- | --- |
| RED | ORANGE |
| YELLOW | GREEN |

**Single Table Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Point | 95% Confidence Interval | |
|  | Estimate | Lower | Upper |
| PARAMETERS: Odds-based |  |  |  |
| Odds Ratio (cross product) | 3.0357 | 1.3313 | 6.9220 (T) |
| Odds Ratio (MLE) | 3.0272 | 1.2813 | 6.9042 (M) |
|  |  | 1.1843 | 7.3857 (F) |
| PARAMETERS: Risk-based |  |  |  |
| Risk Ratio (RR) | 2.2527 | 1.3343 | 3.8035 (T) |
| Risk Difference (RD%) | 21.3884 | 2.3945 | 40.3823 (T) |
|  |  |  |  |
| (T=Taylor series; C=Cornfield; M=Mid-P; F=Fisher Exact) | | | |
|  |  |  |  |
| STATISTICAL TESTS | Chi-square | 1-tailed p | 2-tailed p |
| Chi square - uncorrected | 7.6055 |  | 0.0058201012 |
| Chi square - Mantel-Haenszel | 7.5909 |  | 0.0058676817 |
| Chi square - corrected (Yates) | 6.2338 |  | 0.0125349345 |
| Mid-p exact |  | 0.0064640245 |  |
| Fisher exact |  | 0.0100790020 |  |

Warning: The expected value of a cell is <5. Fisher Exact Test should be used.

การแปรผลตาราง

* + ความชุกของอาการเจ็บหน้าอกในกลุ่มที่มีประวัติโรคเบาหวาน และในกลุ่มที่ไม่มีประวัติเบาหวาน เท่ากับ 38.5% และ 17.1% ตามลำดับ

การแปรผล PR

* + ผู้ที่มีประวัติเป็นโรคเบาหวานมีความชุกของการเจ็บหน้าอกเป็น 2.26 เท่าเทียบกับผู้ที่ไม่เป็นโรคเบาหวาน

(หมายเหตุ : ใช้ค่าของ RR เนื่องจาก PR และ RR ใช้สูตรคำนวณเดียวกัน)

การแปรผล 95%CI

* + มีความเชื่อมั่น 95% ว่าช่วงเชื่อมั่นในการศึกษาครั้งนี้จะครอบคลุมค่าความชุกสัมพัทธ์จริง ซึ่งถ้าหากคลุมค่าจริงแล้ว ค่าความชุกสัมพัทธ์ ในประชากรจะอยู่ระหว่าง 1.33 เท่า ถึง 3.80 เท่า จะเห็นได้ว่า 95%CI ไม่คร่อมค่า 1 ดังนั้นประวัติโรคเบาหวานกับอาการเจ็บหน้าอกน่าจะมีความสัมพันธ์กันจริง

หมายเหตุ: การศึกษานี้เป็น Cross-sectional study ความสัมพันธ์ที่พบไม่สามารถแปลผลไปถึงว่า Exposure ที่ศึกษาจะเป็นสาเหตุของ Outcome หรือไม่ ความสัมพันธ์แปลผลได้เพียงว่า Exposure พบร่วมไปกับ Outcome

การแปรผล p-value

* + จะเห็นได้ว่าใต้ตารางมี warning เรื่อง expected value จึงควรใช้ Fisher exact test ซึ่งได้ค่า p value = 0.01\*2 = 0.02
  + แปลผลได้ว่า หากในความเป็นจริง กลุ่มที่ไม่มีประวัติเบาหวานมีความชุกของอาการเจ็บหน้าอก ไม่แตกต่างไปจากกลุ่มที่มีประวัติเบาหวาน การใช้ข้อมูลนี้สรุปว่าความชุก ดังกล่าวมีความแตกต่างกัน จะมีโอกาสสรุปผิดไปจากความเป็นจริง ร้อยละ 1
  + แปลผลอย่างง่ายได้ว่า ความชุกของอาการเจ็บหน้าอก ของกลุ่มที่ไม่มีประวัติโรคเบาหวานต่างจากกลุ่มที่มีประวัติโรคเบาหวาน อย่างมีนัยสำคัญทางสถิติที่ α 0.05 (p value = 0.02)

**การ Save การวิเคราะห์ข้อมูล**

การ Save งานมีอยู่ 3 ประเภท คือ แฟ้มข้อมูล, คำสั่งที่ใช้ในการวิเคราะห์, และ ผลการวิเคราะห์

1. **การ Save แฟ้มข้อมูล**

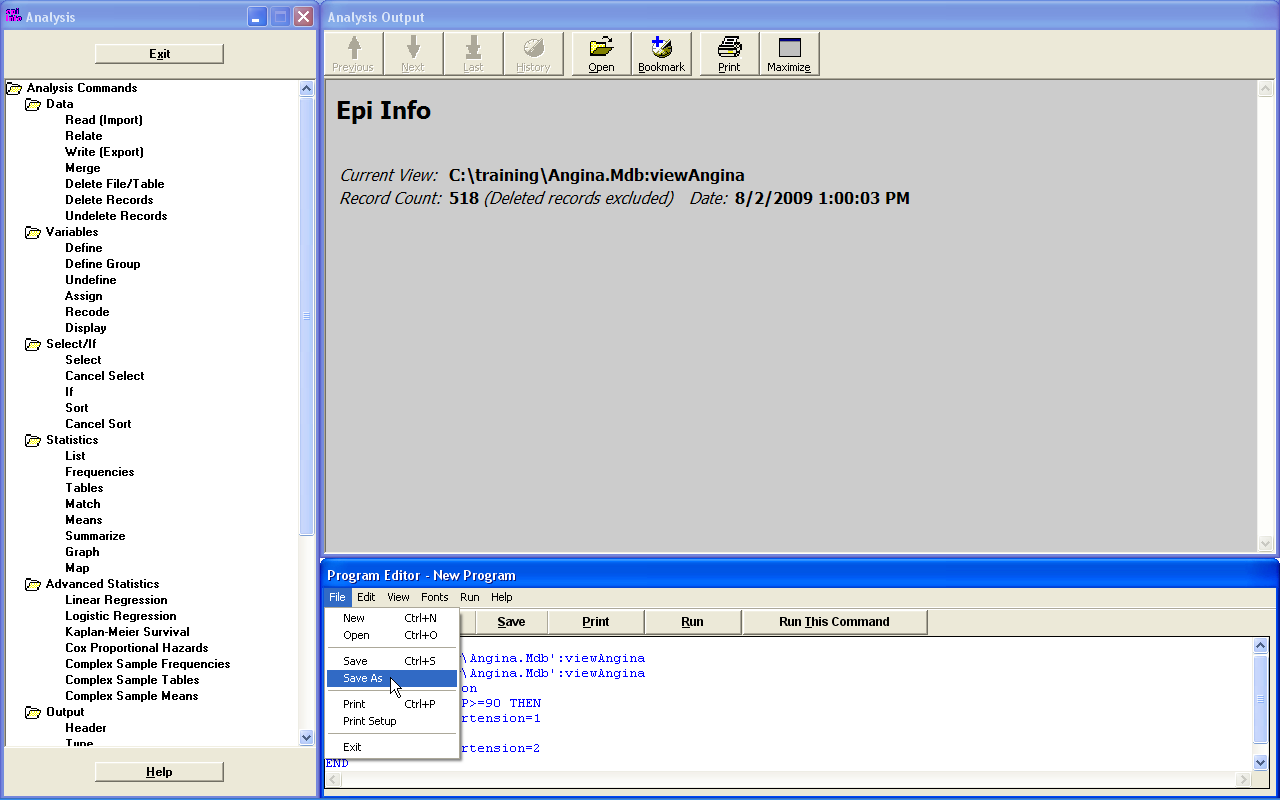
เป็นการ Save ข้อมูลและตัวแปรที่สร้างมาใหม่เป็นการถาวร (ตัวแปรที่สร้างมาใหม่ด้วยคำสั่ง Define จะหายไปหากไม่ได้ทำการ save ด้วยวิธีนี้แล้วออกจากโปรแกรมไป ดังนั้นหากจะวิเคราะห์อีกครั้งต้องมาทำการ Define แล้ว Recode หรือ Assign ใหม่) วิธีการทำก็ใช้ค่ำสั่ง Write (Export) เช่นเดียวกับ การสร้างฐานข้อมูล .mdb ให้กับแฟ้ม Angina.xls ที่เปิดมา ดังตัวอย่างในหัวข้อ การเปิดแฟ้มข้อมูลชนิดอื่นที่ไม่ใช่ **.mdb**

การ Save แฟ้มข้อมูลมีข้อควรระวังว่าขณะจะ save ผู้วิเคราะห์ข้อมูลได้ทำการ Select records อะไรไว้หรือไม่ เพราะการ save แฟ้มข้อมูลจะสร้างแฟ้มที่มีข้อมูลเฉพาะ records ที่เลือกไว้ ซึ่งหากไม่ได้ Select ไว้ ก็จะเป็นการ save ข้อมูลทุก records

1. **การ Save คำสั่งที่ใช้ในการวิเคราะห์ข้อมูล**

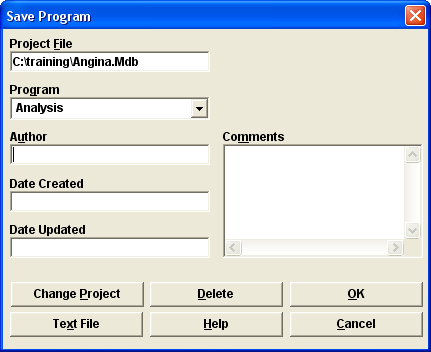
เป็นการ Save คำสั่งที่ปรากฏใน Program Editor ทั้งคำสั่งการวิเคราะห์ข้อมูล และคำสั่งการจัดการข้อมูล (ซึ่งอาจมีทั้งที่ทำได้ถูกต้องและผิดพลาด) เพื่อนำมาใช้ใหม่ในภายหลังหากจะทำการวิเคราะห์ข้อมูลอีก โดยเปิดมา Run โดยไม่ต้องสั่งใหม่อีก (จะได้ทั้งการสร้างตัวแปรใหม่ และผลการวิเคราะห์ข้อมูล) ดังนั้นก่อน save คำสั่ง ควรจะตรวจดูคำสั่งต่างๆว่าไม่มีข้อผิดพลาดอะไร หรือแก้ไขให้เรียบร้อย ก่อนทำการ Save

หมายเหตุ : ควร save คำสั่งให้อยู่ในแฟ้มข้อมูลเดียวกับฐานข้อมูล ดังตัวอย่างต่อไปนี้



2) Click ที่ Save as

1) Click ที่ File

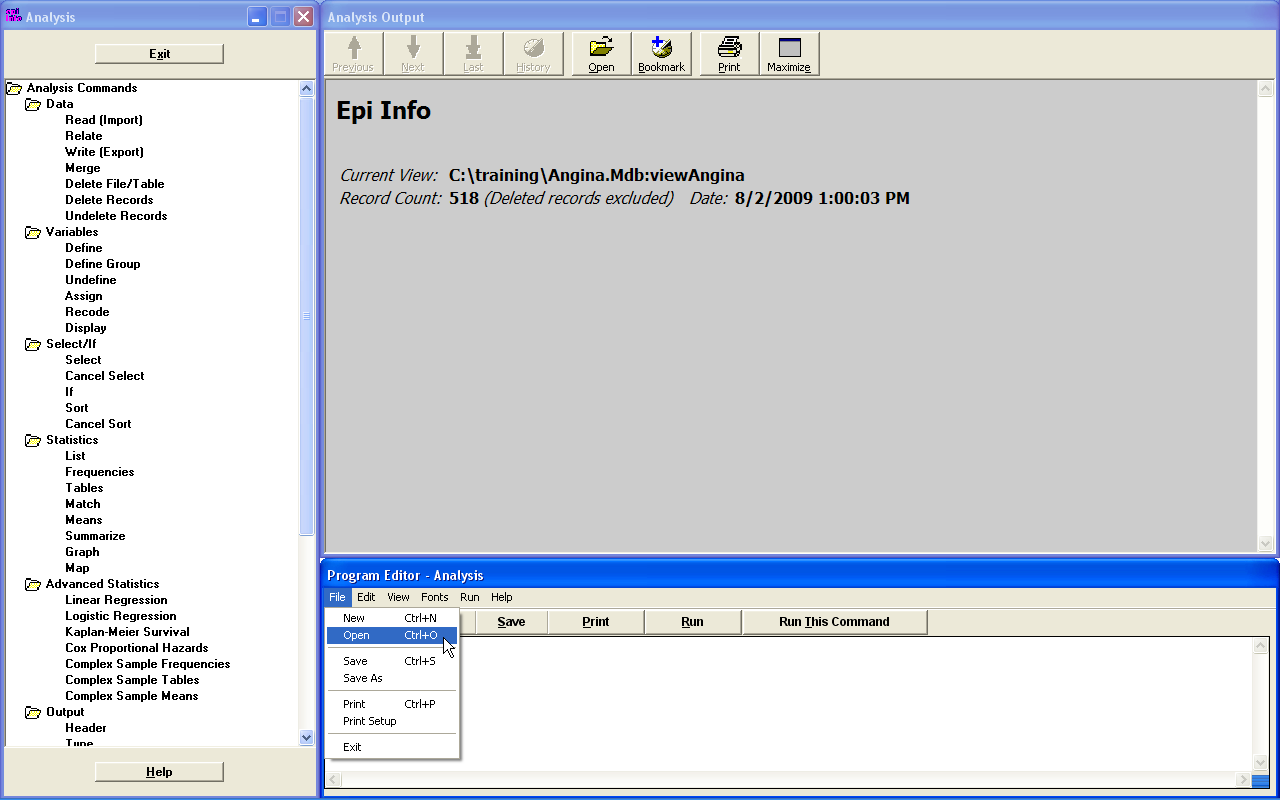


เลือก Change Project แล้ว Browse File ในกรณีที่ข้อ 1) ไม่ใช่แฟ้มข้อมูลที่กำลังวิเคราะห์อยู่

2) พิมพ์ตั้งชื่อโปรแกรม หากกดที่ drop down menu แล้วเลือกชื่อโปรแกรมเก่า จะเป็นการ save ทับ

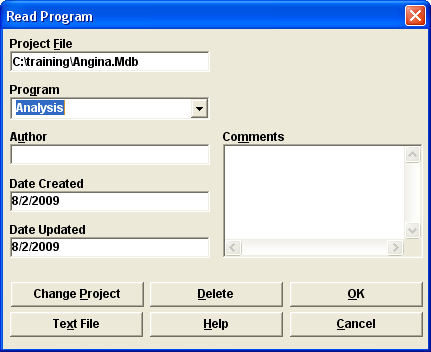
1) ตรวจดูว่าใช่แฟ้มข้อมูลที่วิเคราะห์อยู่หรือไม่ หากไม่ใช่ ให้กดที่ปุ่ม Change Project แล้ว browse ไปแฟ้มที่ถูกต้อง

**การเปิดโปรแกรมที่เคยบันทึกไว้มาใช้ซ้ำ**



2) Click ที่ Open

1) Click ที่ File

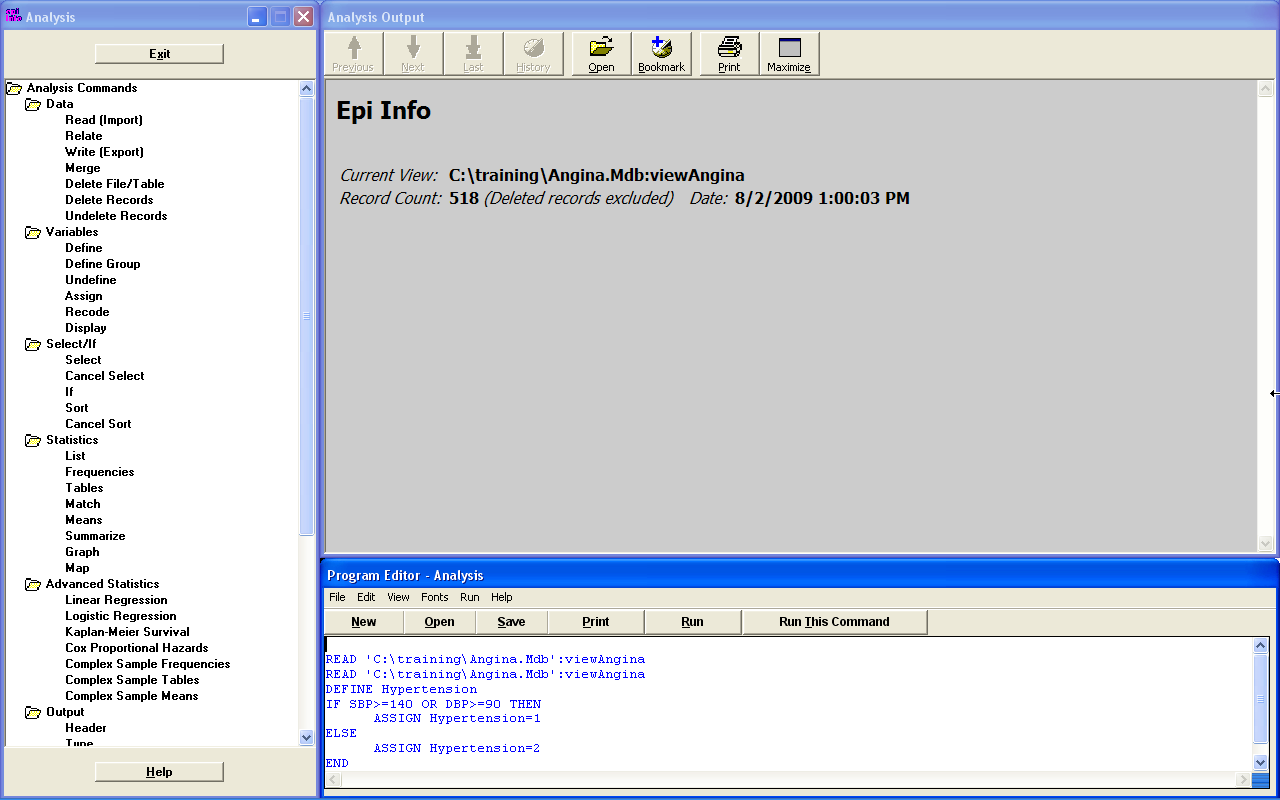


เลือก Change Project แล้ว Browse File ในกรณีที่ข้อ 1) ไม่ใช่แฟ้มข้อมูลที่ต้องการ

2) Click ที่ drop down menu แล้วเลือกชื่อโปรแกรมที่ต้องการ

1) ตรวจดูว่าใช่แฟ้มข้อมูลที่เก็บ program ไว้ หากไม่ใช่ ให้กดที่ปุ่ม Change Project แล้ว browse ไปแฟ้มที่ถูกต้อง

คำสั่งที่เคย Save ไว้จะปรากฏขึ้นที่ Program Editor แต่ยังไม่ได้ทำงาน จากนั้น ให้เลือก Run (เพื่อสั่งให้ Epi Info ทำงานคำสั่งทั้งหมดที่มีในหน้าต่าง Program Editor) หรือ Run This Command (เลื่อน cursor ไปยังบรรทัดที่มีคำสั่งที่ต้องการแล้วกดปุ่ม Run This Command เพื่อสั่งให้ Epi Info ทำงานคำสั่งนั้นคำสั่งเดียว)



Run ทุกคำสั่ง

Run คำสั่งเดียว

1. **การ Save ผลการวิเคราะห์ข้อมูล**

โปรแกรม Epi Info จะทำการ Save ผลการวิเคราะห์ให้โดยอัตโนมัติที่โฟลเดอร์ที่เก็บแฟ้มข้อมูลนั้นๆอยู่ ในรูปแบบ .htm โดยตั้งชื่อเป็น OUT1.htm, OUT2.htm, … (หากไม่ได้ set ชื่อให้เป็นการเฉพาะ) ไปเรื่อยๆ การเปิดสามารถใช้โปรแกรม Web browser ได้ทุกชนิด ในโฟลเดอร์นั้นจะมี File ที่เป็นดรรชนีอยู่ ชื่อว่า IResult.htm ดังนั้นอาจเปิด File นี้ก่อนเพื่อทำการ link ไปหาผลใน files อื่นๆ