หน้า

คำนำ
สารบัญ
คำศัพท์ทางสถิติที่ใช้ในหนังสือเล่มนี้
คำศัพท์ทางสถิติที่ควรเลี่ยง
คำศัพท์เฉพาะสำหรับโปรแกรม STATA

ส่วนที่ 1 การใช้งานโปรแกรม STATA เบื้องต้น 1

บทที่ 1 ทำความรู้จักกับโปรแกรม STATA 3
หน้าต่างทั้ง 4 บาน เมื่อใช้งานจริง 8
ข้อสังเกต 11

บทที่ 2 การใช้คำสั่ง แสดงผล (display) เพื่อทำการคำนวณอย่างง่าย 15
ข้อสังเกต 18

บทที่ 3 การเขียนคำสั่งด้วย Do-file 23
การเขียนบันทึกข้อความลงใน Do-file 26
5 บรรทัดแรกของ Do-file และ Log file 28
การสั่งให้ชุดคำสั่งใน Do-file ทำงาน 36
ประโยชน์ของ Do-file และ Log file 38
ข้อควรจำ 39

หน้า

บทที่ 4 การ load และ save ชุดข้อมูล 41

การ load ข้อมูลเข้ามาในหน่วยความจำของ STATA 42

วิธีที่ 1 ใช้ command input เพื่อกรอกข้อมูลเข้าโปรแกรมโดยตรง 42

วิธีที่ 2 ใช้ command insheet เพื่อนำข้อมูลจาก ไฟล์ Excel เข้าสู่โปรแกรม STATA 46

วิธีที่ 3 ใช้ command use เพื่อเปิดไฟล์ข้อมูล .dta 53

วิธีที่ 4 ใช้ command set obs เพื่อสร้างชุดข้อมูลจำลอง (simulation dataset) 54

การบันทึกชุดข้อมูล ด้วยคำสั่ง save 56

การบันทึกชุดข้อมูลเป็นไฟล์ .csv ด้วยคำสั่ง outsheet 57

คำแนะนำเกี่ยวกับชุดข้อมูล และโปรแกรมเก็บข้อมูล 59

ตัวอย่าง Do-file ของบทที่ 4 65

ข้อควรจำ 66

บทที่ 5 การตรวจสอบข้อมูลด้วย STATA (ครั้งแรก) 67

list 68

if 72

in 76

sort 79

describe 82

Storage type 84

summarize 88

tabulate 92

ตัวอย่าง Do-file ของบทที่ 5 95

ข้อควรจำ 97

หน้า

บทที่ 6 การบริหารจัดการข้อมูลด้วย STATA (ครึ่งแรก) 99

การบริหารจัดการค่าของตัวแปร 100

generate 100

replace 101

Missing data (ข้อมูลสูญหาย) 108

หลักทั่วไปในการปฏิบัติต่อข้อมูลสูญหาย 109

การบริหารจัดการข้อมูลสูญหายด้วยโปรแกรม STATA 114

Option missing 120

destring 122

duplicates 123

drop/keep 124

ตัวอย่างการบริหารจัดการข้อมูลในสถานการณ์จริง 125

ขั้นที่ 1 ตรวจสอบภาพรวมของชุดข้อมูล 126

ขั้นที่ 2 ตรวจสอบลักษณะข้อมูลของตัวแปรทีละตัว 128

ขั้นที่ 3 จัดการกับข้อมูลสูญหาย 144

ขั้นที่ 4 ตรวจสอบความสัมพันธ์ของตัวแปร 146

ขั้นที่ 5 ตรวจสอบซ้ำทั้งหมดอีกครั้งหนึ่ง 147

ขั้นที่ 6 สร้างตารางสรุปข้อมูล 147

ตัวอย่าง Do-file ของบทที่ 6 149

ข้อควรจำ 153

บทที่ 7 การตรวจสอบข้อมูลใน STATA (ครึ่งหลัง) 155

histogram 156

qnorm 163

graph box 173

twoway scatter 176
twoway lfit 178
Scatter plot และ Pearson's correlation coefficient 180
ตัวอย่าง Do-file ของบทที่ 7 186
ข้อควรจำ 187

บทที่ 8 การ label ข้อมูลใน STATA 189
การระบุชื่อของชุดข้อมูล 190
การระบุชื่อของตัวแปร 192
การระบุค่าของตัวแปร 193
ตัวอย่าง Do-file ของบทที่ 8 197

บทที่ 9 การบริหารจัดการข้อมูลใน STATA (ครึ่งหลัง) 199
การใช้ expression ประกอบคำสั่ง generate 199
ลำดับ (_n) และยอดรวม (_N) 201
bysort 204
egen 206
การสร้างตัวแปร category variable ด้วยคำสั่ง egen 209
ตัวอย่าง Do-file ของบทที่ 9 213

บทที่ 10 การบริหารจัดการข้อมูลตัวอักษร (string function) 215
สำนวนชุดที่ 1 (trim, lower) 216
สำนวนชุดที่ 2 (substr, subinstr, string) 218
การสร้าง dummy variable ด้วยคำสั่ง tabulate 221

บทที่ 11 การบริหารจัดการข้อมูลวันที่ 225
การใส่ข้อมูลวันที่ 226
การแก้ไขข้อมูลวันที่ 230
การประยุกต์ใช้ _n และ _N ในข้อมูล longitudinal data 232

ส่วนที่ 2 ความรู้ทั่วไปเกี่ยวกับสถิติ 243

บทที่ 12 p value คืออะไร 245
Statistical inference หรือการอนุมานทางสถิติ 246
การตั้งสมมุติฐาน และการทดสอบสมมุติฐาน 253
ความหมายของ p value 260
ทำไมต้อง p value < 0.05 หรือ < 0.01 264
ระดับน้ำหนักของหลักฐาน โดย p value 266
Borderline significant (แตกต่างอย่างมีนัยสำคัญระดับก้ำกึ่ง) 267
ความเข้าใจเกี่ยวกับ p value (1) 268
การประเมินค่าคงที่ และช่วงความเชื่อมั่น 95% (95% Confidence Interval) 273
ระดับนัยสำคัญทางสถิติ กับขนาดกลุ่มตัวอย่าง 277
ความเข้าใจเกี่ยวกับค่า p value (2) 281
Size of effect (ขนาดของความแตกต่างหรือขนาดของผลกระทบ) 282
Two-sided test 284
การเขียนสรุปสุดท้ายจากการคำนวณทางสถิติ 285
นัยสำคัญทางสถิติและนัยสำคัญทางปฏิบัติ 286
การวิเคราะห์ข้อมูลกับความลำเอียง (bias) 287

บทที่ 13 การเลือกวิธีทดสอบทางสถิติ (อย่างง่าย) 291
สิ่งสำคัญที่สุดสิ่งแรก : ตัวแปร 292
สิ่งสำคัญอันดับ 2 : ชนิดของตัวแปร 294
สิ่งสำคัญอันดับ 3 : สมมุติฐานหลัก และสมมุติฐานทางเลือก 297
ตารางเลือกวิธีทดสอบสมมุติฐาน 302
ตารางที่ 1 ตารางเลือกวิธีทดสอบสมมุติฐาน ในกรณีที่ตัวแปรผลลัพธ์เป็นตัวแปรชนิดช่วงมาตราและกระจายตัวปกติ 302
ตารางที่ 2 ตารางเลือกวิธีทดสอบสมมุติฐาน ในกรณีที่ตัวแปรผลลัพธ์เป็นตัวแปรชนิดลำดับ หรือตัวแปรผลลัพธ์เป็นตัวแปรชนิดช่วงมาตราแต่กระจายตัวไม่ปกติ 303
ตารางที่ 3 ตารางเลือกวิธีทดสอบสมมุติฐาน ในกรณีที่ตัวแปรผลลัพธ์เป็นตัวแปรชนิดกลุ่ม 304
คำแนะนำทั่วไปสำหรับการเลือกวิธีทดสอบทางสถิติ 305

ส่วนที่ 3 การทดสอบสมมุติฐานทางสถิติ **309**
รายละเอียดชุดข้อมูลตัวอย่าง 311

บทที่ 14 การเปรียบเทียบค่าเฉลี่ย (t-test) **315**
การเปรียบเทียบค่าเฉลี่ยในตัวอย่าง 1 กลุ่มกับค่ามาตรฐาน (One-sample t test) 320
95%CI และ Standard error of the mean (ci) 323
การเปรียบเทียบค่าเฉลี่ยระหว่างกลุ่มตัวอย่าง 2 กลุ่ม (Simple t test) 324
การเปรียบเทียบค่าเบี่ยงเบนมาตรฐานระหว่างกลุ่มตัวอย่าง 2 กลุ่ม (F test) 327
การเปรียบเทียบความแตกต่างของค่าเฉลี่ยของกลุ่มตัวอย่างแบบจับคู่ (Paired t test) 331

บทที่ 15 การคำนวณความแตกต่างของค่าเฉลี่ย โดยการสร้าง model 337
Simple linear regression (1) 338
Pearson's correlation coefficient (pwcorr) 349
Analysis of variance (ANOVA) 352
Simple linear regression (2) 359
Dummy variables 364
xi : 365
Multivariable linear regression model 370
Likelihood ratio test (lrtest) 376
Wald test for multiple combination of variables (testparm) 377
สรุปท้ายบท 383

บทที่ 16 การเปรียบเทียบค่ามัธยฐาน 385
การเปรียบเทียบค่ามัธยฐานในกลุ่มตัวอย่าง 1 กลุ่มกับค่ามาตรฐาน (One-sample sign test) 386
การเปรียบเทียบค่ามัธยฐานในกลุ่มตัวอย่าง 1 กลุ่มกับค่ามาตรฐาน (One-sample Wilcoxon signed-rank test) 392
การเปรียบเทียบค่ามัธยฐานระหว่างกลุ่มตัวอย่าง 2 กลุ่ม (Median test) 397
การเปรียบเทียบค่ามัธยฐานระหว่างกลุ่มตัวอย่าง 2 กลุ่ม (Wilcoxon-Mann-Whitney test) 401
การเปรียบเทียบค่ามัธยฐานระหว่างกลุ่มตัวอย่างแบบจับคู่ (Paired sign test) 406
การเปรียบเทียบค่ามัธยฐานระหว่างกลุ่มตัวอย่างแบบจับคู่ (Paired signed-rank test) 410

บทที่ 17 Kruskal-Wallis test และ
Spearmann correlation coefficient 415
การเปรียบเทียบค่ามัธยฐานระหว่างกลุ่มตัวอย่างตั้งแต่ 2 กลุ่มขึ้นไป
(Kruskal-Wallis test) 415
การหาค่า Spearman's rank correlation coefficient 419

บทที่ 18 การเปรียบเทียบอัตราส่วน 423
การเปรียบเทียบอัตราส่วนในกลุ่มตัวอย่างหนึ่งกลุ่มกับค่ามาตรฐาน
(Binomial test) 423
95% CI และ Standard error of the proportion (ci) 426
การเปรียบเทียบอัตราส่วนระหว่างกลุ่มตัวอย่างตั้งแต่ 2 กลุ่มขึ้นไป
(Chi-square test และ Fisher exact test) 427
95% CI สำหรับค่าความแตกต่างของอัตราส่วน (cs) 430
การเปรียบเทียบอัตราส่วนระหว่างกลุ่มตัวอย่างแบบจับคู่
(McNemar's test) 432
การเปรียบเทียบอัตราส่วนระหว่างกลุ่มตัวอย่างหลายกลุ่มแบบมีทิศทาง
(Chi-square test for trend) 439

บทที่ 19 การเปรียบเทียบอัตราส่วนโดย
Logistic regression model 443
ตัวอย่างที่ 19.1 ตัวแปรปัจจัยเป็น Binary variable 452
ตัวอย่างที่ 19.2 ตัวแปรปัจจัยเป็น Category variable 455
ตัวอย่างที่ 19.3 ตัวแปรปัจจัยเป็นตัวแปรชนิดตัวเลขมาตรา 459
Multivariable logistic regression (ตัวแปรปัจจัยมากกว่า 1 ตัว) 463
ตัวอย่างที่ 19.4 Multivariable logistic regression 463

ส่วนที่ 4 บทเสริมสำหรับผู้ใช้งานระดับกลางขึ้นไป 469

บทที่ 20 การทดสอบสมมุติฐานโดยไม่ต้องมีชุดข้อมูลด้วย Immediate commands 471
การเปรียบเทียบอัตราส่วนระหว่างกลุ่มตัวอย่าง 2 กลุ่ม 473
การเปรียบเทียบสัดส่วน (Odds ratio) ระหว่างกลุ่มตัวอย่าง 2 กลุ่ม 477
การเปรียบเทียบอัตราส่วนระหว่างกลุ่มตัวอย่าง 2 กลุ่มด้วยค่าสถิติ z 482
การเปรียบเทียบอัตราส่วนในกลุ่มตัวอย่าง 1 กลุ่มกับค่ามาตรฐาน 485
การสร้างช่วงความเชื่อมั่นจากกลุ่มตัวอย่าง 1 กลุ่ม 487
การเปรียบเทียบค่าเฉลี่ยระหว่างกลุ่มตัวอย่าง 2 กลุ่ม 490
การเปรียบเทียบค่าเฉลี่ยในกลุ่มตัวอย่าง 1 กลุ่มกับค่ามาตรฐาน 493
การเปรียบเทียบค่าเบี่ยงเบนมาตรฐานระหว่างกลุ่มตัวอย่าง 2 กลุ่ม 496

บทที่ 21 การคำนวณขนาดของกลุ่มตัวอย่าง (sampsi) 501
หลักการคำนวณขนาดของกลุ่มตัวอย่าง 510

บทที่ 22 เทคนิคการใช้ STATA สำหรับผู้ใช้งานระดับกลาง 529
เทคนิคที่ 1 การสร้างตารางสรุป 529
เทคนิคที่ 2 การใช้ผลลัพธ์จากการคำนวณ 535
เทคนิคที่ 3 การใช้ local 538
เทคนิคที่ 4 การใช้ /// 543
เทคนิคที่ 5 การแต่งกราฟ และการบันทึกการแต่งกราฟ 545