

[Applied Probability of Statistics Book 2]

# ความน่าจะเป็น และสถิติประยุกต์

เล่ม 2

สำหรับนักศึกษาระดับ ปวส./ปริญญาตรีและผู้ที่สนใจทั่วไป

# ความน่าจะเป็น และ สถิติประยุกต์ เล่ม



APPLIED PROBABILITY AND STATISTICS BOOK 2

ข้าขอประณตน์อมสัการ บุรพคณาจารย์  
ผู้ก่อร้ปเกิดประโยชน์ศึกษา

รองศาสตราจารย์ธีระศักดิ์ อุรัจนานนท์

## “ความน่าจะเป็นและสถิติประยุกต์ เล่ม 2”

พิมพ์ครั้งที่ 1 มกราคม 2546

สงวนลิขสิทธิ์ตามกฎหมาย

ห้ามคัดลอกถ่ายเอกสารหรือพิมพ์

หรือวิธีหนึ่งวิธีใดของหนังสือเล่มนี้ก่อนได้รับอนุญาต

จากบริษัท สกายบุ๊กส์ จำกัด

## ราคา 220 บาท

### ข้อมูลทางบรรณานุกรมของหอสมุดแห่งชาติ

ธีระศักดิ์ อุรัจนานนท์

ความน่าจะเป็นและสถิติประยุกต์ เล่ม 2 -- กรุงเทพฯ : สกายบุ๊กส์, 2546.

564 หน้า

1. ความน่าจะเป็น 2. คณิตศาสตร์สถิติ I. ชื่อเรื่อง

519

ISBN: 974-389-314-8

S7901-30-01-03

### จัดพิมพ์และจำหน่ายโดย



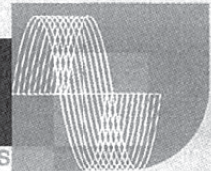
บริษัท สกายบุ๊กส์ จำกัด  
SKYBOOK COMPANY LIMITED  
515/270-8 ถ.รัชดา-ปทุมธานี ต.ประเวศชัย อ.เมือง จ.ปทุมธานี 12130  
โทร. 0-2958-1125-7, 0-2567-5119 โทรสาร. 0-2567-5105  
E-mail: skybook1992@hotmail.com

[www.skybook.co.th](http://www.skybook.co.th)

พิมพ์ที่ บริษัท สยามสปอร์ต ซินดิเคท จำกัด

459 ซอยพิบูลย์อุปถัมภ์ (ลาดพร้าว 48) แขวงสามเสนนอก เขตห้วยขวาง กรุงเทพฯ 10310

โทรศัพท์ : 0-2694-3010



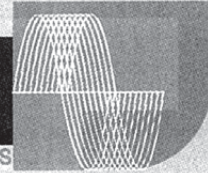
หนังสือความน่าจะเป็นและสถิติประยุกต์ ( APPLIED PROBABILITY AND STATISTICS ) เล่มนี้ ผู้เขียนได้เขียนและเรียบเรียงขึ้นเพื่อให้เป็นตำราประกอบการเรียนการสอนวิชาสถิติในระดับมหาวิทยาลัย ได้จำแนกเป็นความน่าจะเป็นและสถิติประยุกต์ เล่ม 1 และความน่าจะเป็นและสถิติประยุกต์ เล่ม 2 โดยเป็นตำราทางด้านสถิติที่ผู้เขียนได้ตั้งใจเขียนและเรียบเรียงเนื้อหาตั้งแต่ระดับต้นจนถึงระดับสูง ซึ่งมีทั้งทางด้านทฤษฎีซึ่งมีการพิสูจน์สูตรและการนำทฤษฎีไปประยุกต์ใช้ในสถานการณ์ต่าง ๆ และอาศัยความรู้ การค้นคว้าตำรา เอกสาร รวมถึงงานวิจัยต่าง ๆ และประสบการณ์ในการให้คำปรึกษาทางด้านการวิเคราะห์ข้อมูลทางสถิติและการวิจัย

หนังสือความน่าจะเป็นและสถิติประยุกต์ เล่ม 1 เนื้อหาภายในเล่ม ผู้เขียนได้เขียนและเรียบเรียงเนื้อหาวิชาสถิติตั้งแต่พื้นฐานเริ่มต้นซึ่งแบ่งได้ดังนี้ ภาคสถิติพรรณนาหรือสถิติบรรยาย โดยครอบคลุมเนื้อหาตั้งแต่บทที่ 1 ถึง บทที่ 7 ความรู้เบื้องต้นของสถิติที่ใช้บรรยายลักษณะของข้อมูล ภาคความน่าจะเป็น ซึ่งครอบคลุมเนื้อหาตั้งแต่บทที่ 8 ถึง บทที่ 14 เป็นเครื่องมือที่สำคัญที่ช่วยในการวิเคราะห์ข้อมูลทางสถิติ เพราะทฤษฎีต่าง ๆ ทางสถิติส่วนใหญ่ตั้งอยู่บนพื้นฐานของความน่าจะเป็น

สำหรับหนังสือความน่าจะเป็นและสถิติประยุกต์ เล่ม 2 นี้เนื้อหาภายในเล่ม ผู้เขียนได้เขียนและเรียบเรียงเนื้อหาวิชาสถิติซึ่งแบ่งได้ดังนี้ ภาคสถิติอ้างอิงหรือสถิติอนุมาน โดยได้ครอบคลุมเนื้อหาตั้งแต่บทที่ 15 ถึง บทที่ 19 เป็นสถิติที่อยู่บนพื้นฐานของข้อมูลสถิติพรรณนากับความน่าจะเป็นในการตัดสินใจในสถานการณ์ต่าง ๆ ที่ได้นำสถิติไปประยุกต์ใช้ บทที่ 20 ได้กล่าวถึงการวิเคราะห์การถดถอยและสหสัมพันธ์ ซึ่งเป็นเครื่องมือที่ช่วยในการวิเคราะห์ข้อมูลหาความสัมพันธ์ระหว่างตัวแปร และบทที่ 21 กล่าวถึงการทดสอบสถิติพารามเมตริก โดยเป็นวิธีทางสถิติอย่างง่าย ๆ ที่คำนวณไม่ยุ่งยาก และไม่มีเงื่อนไขหรือข้อตกลงเกี่ยวกับการแจกแจงเฉพาะของประชากร

ผู้เขียนหวังเป็นอย่างยิ่งว่าหนังสือเล่มนี้ จะเป็นประโยชน์ในการเรียนการสอน ต่อนักศึกษา นักวิชาการ และผู้ที่ต้องการใช้ประกอบการวิเคราะห์ข้อมูลในการวิจัย และผู้เขียนขอกราบขอบพระคุณ บิดา มารดา คุณครู และอาจารย์ทุก ๆ ท่าน ที่ได้ประสิทธิ์ประสาทความรู้วิชาทางด้านสถิติให้ผู้เขียน หากมีข้อผิดพลาดอันเนื่องมาจากการพิมพ์หรือการเรียบเรียง กรุณาแจ้งให้ผู้เขียนทราบ จักเป็นพระคุณอย่างสูง

รองศาสตราจารย์ธีระศักดิ์ อุรัจนานนท์



|                                                                   | หน้า |
|-------------------------------------------------------------------|------|
| <b>บทที่ 15 การสุ่มตัวอย่าง</b>                                   | 391  |
| ประเภทของสถิติ                                                    | 391  |
| ประชากรและตัวอย่าง                                                | 393  |
| วิธีการสุ่มตัวอย่าง                                               | 395  |
| ทฤษฎีการสุ่มตัวอย่าง                                              | 401  |
| การแจกแจงของตัวอย่างสุ่ม                                          | 407  |
| การแจกแจงของค่าเฉลี่ยเลขคณิตของตัวอย่าง                           | 410  |
| การแจกแจงของผลต่างของค่าเฉลี่ยเลขคณิตของตัวอย่าง                  | 422  |
| การแจกแจงของอัตราส่วนของตัวอย่าง                                  | 426  |
| การแจกแจงของผลต่างของสัดส่วนของตัวอย่าง                           | 429  |
| การแจกแจงที                                                       | 431  |
| การแจกแจงค่าความแปรปรวนของตัวอย่าง                                | 446  |
| การแจกแจงเอฟ                                                      | 451  |
| แบบฝึกหัดบทที่ 15                                                 | 456  |
| คำตอบแบบฝึกหัดบทที่ 15                                            | 462  |
| <b>บทที่ 16 การประมาณค่า</b>                                      | 465  |
| การอนุมานเชิงสถิติ                                                | 466  |
| วิธีการของการประมาณค่า                                            | 466  |
| การประมาณแบบจุด                                                   | 467  |
| สมบัติของตัวประมาณค่าที่ดี                                        | 468  |
| การประมาณแบบช่วง                                                  | 472  |
| การประมาณค่าเฉลี่ยประชากร $\mu$                                   | 473  |
| การประมาณค่าของผลต่างระหว่างค่าเฉลี่ยของประชากรสองกลุ่ม           | 482  |
| การประมาณผลต่างของค่าเฉลี่ยของประชากรสองกลุ่มที่มีความสัมพันธ์กัน | 492  |
| การประมาณค่าสัดส่วน                                               | 494  |
| การประมาณค่าของผลต่างระหว่างสัดส่วนของประชากรสองกลุ่ม             | 498  |
| การประมาณค่าความแปรปรวน                                           | 500  |

|                 |                                                                   |     |
|-----------------|-------------------------------------------------------------------|-----|
|                 | การประมาณค่าอัตราส่วนของความแปรปรวนของประชากรสองกลุ่ม             | 503 |
|                 | สรุปช่วงความเชื่อมั่น $(1 - \alpha)100\%$ สำหรับค่าต่าง ๆ         | 507 |
|                 | แบบฝึกหัดบทที่ 16                                                 | 510 |
|                 | คำตอบแบบฝึกหัดบทที่ 16                                            | 517 |
| <b>บทที่ 17</b> | <b>การทดสอบสมมติฐาน</b>                                           | 519 |
|                 | สมมติฐานสถิติ                                                     | 519 |
|                 | การทดสอบสมมติฐานสถิติ                                             | 521 |
|                 | ความผิดพลาดประเภท I และความผิดพลาดประเภท II                       | 522 |
|                 | การทดสอบทางเดียวและการทดสอบสองทาง                                 | 526 |
|                 | การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของประชากรกลุ่มเดียว            | 530 |
|                 | การทดสอบสมมติฐานเกี่ยวกับผลต่างระหว่างค่าเฉลี่ยของประชากรสองกลุ่ม | 539 |
|                 | การทดสอบสมมติฐานเกี่ยวกับสัดส่วนของประชากร                        | 552 |
|                 | การทดสอบสมมติฐานเกี่ยวกับผลต่างระหว่างสัดส่วนของประชากรสองกลุ่ม   | 556 |
|                 | การทดสอบสมมติฐานเกี่ยวกับความแปรปรวนของประชากรกลุ่มเดียว          | 561 |
|                 | การทดสอบสมมติฐานเกี่ยวกับความแปรปรวนของประชากรสองกลุ่ม            | 562 |
|                 | สรุปสถิติที่ใช้ในการทดสอบสมมติฐาน                                 | 568 |
|                 | แบบฝึกหัดบทที่ 17                                                 | 571 |
|                 | คำตอบแบบฝึกหัดบทที่ 17                                            | 579 |
| <b>บทที่ 18</b> | <b>การทดสอบไคสแควร์</b>                                           | 581 |
|                 | การทดสอบหลายสัดส่วน                                               | 585 |
|                 | การทดสอบความเหมาะสมของการแจกแจง                                   | 587 |
|                 | การทดสอบความเป็นอิสระ                                             | 594 |
|                 | แบบฝึกหัดบทที่ 18                                                 | 606 |
|                 | คำตอบแบบฝึกหัดบทที่ 18                                            | 610 |
| <b>บทที่ 19</b> | <b>การวิเคราะห์ความแปรปรวน</b>                                    | 611 |
|                 | นิยามศัพท์ต่าง ๆ ที่เกี่ยวข้อง                                    | 612 |
|                 | การทดสอบสมมติฐานเกี่ยวกับความแปรปรวนของหลายประชากร                | 616 |
|                 | แผนการทดลองแบบสุ่มตลอด                                            | 622 |
|                 | ข้อมูลจำแนกทางเดียวและวิธีวิเคราะห์ความแปรปรวน                    | 624 |
|                 | แผนการทดลองแบบกลุ่มสุ่ม                                           | 641 |

|          |                                                        |     |
|----------|--------------------------------------------------------|-----|
|          | ข้อมูลจำแนกสองทางและวิธีวิเคราะห์ความแปรปรวน           | 642 |
|          | แผนการทดลองแบบแฟคทอเรียล                               | 656 |
|          | การเปรียบเทียบค่าเฉลี่ยรายคู่                          | 667 |
|          | การวิเคราะห์ความแปรปรวนร่วม                            | 684 |
|          | แบบฝึกหัดบทที่ 19                                      | 699 |
|          | คำตอบแบบฝึกหัดบทที่ 19                                 | 706 |
| บทที่ 20 | การวิเคราะห์การถดถอยและสหสัมพันธ์                      | 707 |
|          | ตัวแบบเชิงคณิตศาสตร์                                   | 708 |
|          | ตัวแบบการถดถอย                                         | 708 |
|          | การถดถอยเชิงเส้น                                       | 710 |
|          | การถดถอยเชิงเส้นอย่างง่าย                              | 713 |
|          | วิธีกำลังสองน้อยที่สุด                                 | 715 |
|          | การวิเคราะห์การถดถอย                                   | 723 |
|          | การประมาณค่าและการทดสอบสมมติฐานพารามิเตอร์ของเส้นถดถอย | 723 |
|          | สหสัมพันธ์                                             | 746 |
|          | สหสัมพันธ์เชิงเส้นอย่างง่าย                            | 746 |
|          | สัมประสิทธิ์สหสัมพันธ์                                 | 748 |
|          | การทดสอบสมมติฐานและการประมาณค่าสัมประสิทธิ์สหสัมพันธ์  | 753 |
|          | สมการถดถอยเชิงเส้นพหุคูณ                               | 761 |
|          | แบบฝึกหัดบทที่ 20                                      | 782 |
|          | คำตอบแบบฝึกหัดบทที่ 20                                 | 788 |
| บทที่ 21 | การทดสอบสถิตินอนพาราเมตริก                             | 789 |
|          | สถิติทดสอบสำหรับกลุ่มตัวอย่าง 1 กลุ่ม                  | 790 |
|          | สถิติทดสอบสำหรับกลุ่มตัวอย่างที่สัมพันธ์กัน 2 กลุ่ม    | 802 |
|          | สถิติทดสอบสำหรับกลุ่มตัวอย่างอิสระ 2 กลุ่ม             | 814 |
|          | สถิติทดสอบสำหรับกลุ่มตัวอย่างหลายกลุ่มที่สัมพันธ์กัน   | 831 |
|          | สถิติทดสอบสำหรับกลุ่มตัวอย่างอิสระหลายกลุ่ม            | 841 |
|          | สถิติทดสอบหลายตัวแปร                                   | 851 |
|          | การวัดความสัมพันธ์และการทดสอบนัยสำคัญ                  | 864 |
|          | การวัดความสัมพันธ์ระหว่างตัวแปรระดับนามบัญญัติ         | 864 |

|                                                                                                                               |      |
|-------------------------------------------------------------------------------------------------------------------------------|------|
| การวัดความสัมพันธ์ระหว่างตัวแปรระดับเรียงอันดับ                                                                               | 871  |
| ตารางเปรียบเทียบการใช้สถิติทดสอบสมมติฐานระหว่างสถิติพาราเมตริกและสถิตินอนพาราเมตริก                                           | 880  |
| ตารางการใช้สถิตินอนพาราเมตริกในการทดสอบสมมติฐาน                                                                               | 881  |
| แบบฝึกหัดบทที่ 21                                                                                                             | 882  |
| คำตอบแบบฝึกหัดบทที่ 21                                                                                                        | 890  |
| <b>ภาคผนวก ก. บรรณานุกรม</b>                                                                                                  | ก-1  |
| <b>ภาคผนวก ข. ตารางสถิติ</b>                                                                                                  | ข-1  |
| ตาราง 1. พื้นที่ใต้โค้งปกติมาตรฐาน                                                                                            | ข-3  |
| ตาราง 2. ค่าวิกฤตของการแจกแจง $t$                                                                                             | ข-4  |
| ตาราง 3. ค่าวิกฤตของการแจกแจง $\chi^2$                                                                                        | ข-5  |
| ตาราง 4. ค่าวิกฤตของการแจกแจง $F$                                                                                             | ข-6  |
| ตาราง 5. เลขสุ่ม                                                                                                              | ข-10 |
| ตาราง 6. ค่าวิกฤตของ $F_{\max}$ สำหรับการทดสอบความเป็นเอกพันธ์ของความแปรปรวนที่ระดับนัยสำคัญ 0.05 และ 0.01 ตามวิธีของ Hartley | ข-11 |
| ตาราง 7. ค่าวิกฤตสำหรับการทดสอบความเป็นเอกพันธ์ของความแปรปรวนตามวิธีของ Cochran                                               | ข-12 |
| ตาราง 8. ค่าวิกฤตของพิสัยสถิติเดนท์ (Studentized Range : $q$ ) สำหรับการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของกลุ่มต่าง ๆ        | ข-13 |
| ตาราง 9. ค่าวิกฤตของ $q$ สำหรับการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของกลุ่มต่าง ๆ ตามวิธีของดันคัน                             | ข-15 |
| ตาราง 10. ค่าความน่าจะเป็นที่ได้จากค่าน้อยสุดของ $X$ ในการทดสอบ Binomial Test                                                 | ข-17 |
| ตาราง 11. ค่าวิกฤตของ D ในการทดสอบ Kolmogorov –Smirnov One- Sample Test                                                       | ข-18 |
| ตาราง 12. ค่าวิกฤตของ $r$ ในการทดสอบ Runs Test                                                                                | ข-19 |
| ตาราง 13. ค่าวิกฤตของ T ในการทดสอบ Sign Test                                                                                  | ข-21 |
| ตาราง 14. ค่าวิกฤตของ T ในการทดสอบ Wilcoxon Signed Ranks Test                                                                 | ข-22 |
| ตาราง 15. ค่าความน่าจะเป็นที่ได้จากค่าน้อยสุดของ $U$ ในการทดสอบ Mann-Whitney U Test                                           | ข-23 |

|                                                                                                                             |      |
|-----------------------------------------------------------------------------------------------------------------------------|------|
| ตาราง 16. ค่าวิกฤตของ U ในการทดสอบ Mann-Whitney U Test ที่ระดับนัย<br>สำคัญต่าง ๆ                                           | ข-26 |
| ตาราง 17. ค่าวิกฤตของ $K_D$ ในการทดสอบ kolmogorov-Smirnov<br>Two-Sample Test (กลุ่มตัวอย่างขนาดเล็ก)                        | ข-28 |
| ตาราง 18. ค่าวิกฤตของ $K_D$ ในการทดสอบ kolmogorov-Smirnov<br>Two-Sample Test (กลุ่มตัวอย่างขนาดใหญ่ในการทดสอบแบบ<br>สองทาง) | ข-29 |
| ตาราง 19. ค่าความน่าจะเป็นที่ได้จากค่าใหญ่สุดของ $\chi^2_r$ ในการทดสอบ<br>Friedman Two-way Analysis of Variance by Ranks    | ข-30 |
| ตาราง 20. ค่าความน่าจะเป็นที่ได้จากค่าใหญ่สุดของ $H$ ในการทดสอบ<br>Kruskal-Wallis One-way Analysis of Variance by Ranks     | ข-32 |
| ตาราง 21. ค่าความน่าจะเป็นที่ได้จากค่าใหญ่สุดของ $S$ ในการทดสอบ<br>Kendall ' s tau ( $\tau$ )                               | ข-34 |
| ตาราง 22. ค่าวิกฤตของ $S$ ในการทดสอบสัมประสิทธิ์ของความสอดคล้อง<br>kendall Coefficient of Concordance ( $W$ )               | ข-35 |
| ตาราง 23. ค่าวิกฤตของสัมประสิทธิ์สหสัมพันธ์เชิงอันดับ $r_s$<br>(Spearman Rank Correlation Coefficient)                      | ข-36 |

## ประเภทของสถิติ

จากประเภทของสถิติที่มีความหมายเกี่ยวกับศาสตร์ที่ว่าด้วยระเบียบวิธีการทางสถิตินั้น สามารถแบ่งขอบเขตของเนื้อหาและวิธีการได้เป็น 2 ประเภทด้วยกัน ดังนี้

### 1. สถิติพรรณนา หรือสถิติบรรยาย (Descriptive Statistics)

เป็นสถิติที่ศึกษาหาคำตอบเชิงตัวเลข ในแง่บรรยายลักษณะข้อมูลหรือการแจกแจงข้อมูลเฉพาะของประชากรหรือกลุ่มตัวอย่างประชากร ตามข้อมูลที่เก็บรวบรวมมา โดยเก็บรวบรวมข้อมูลในกลุ่มเล็กหรือกลุ่มใหญ่ก็ได้ แล้วสรุปและตีความหมายของข้อมูลในกลุ่มที่ทำการศึกษาเท่านั้น โดยไม่นำไปใช้อธิบายหรือสรุปอ้างอิงไปยังประชากรหรือกลุ่มตัวอย่างประชากรกลุ่มอื่น ๆ ซึ่งการบรรยายลักษณะข้อมูลหรือการแจกแจงข้อมูลของประชากร หรือกลุ่มตัวอย่างประชากรกลุ่มใดกลุ่มหนึ่งนั้น ได้ทำการศึกษาในเรื่อง

1. การจัดตำแหน่งเปรียบเทียบ เช่น แสดงความถี่ ร้อยละ สัดส่วน อัตราส่วน เปอร์เซ็นต์ไทล์
2. การวัดแนวโน้มเข้าสู่ส่วนกลาง เช่น ค่าเฉลี่ย มัธยฐาน
3. การวัดการกระจาย เช่น ส่วนเบี่ยงเบนมาตรฐาน
4. การวัดการแจกแจง เช่น ความเบ้ ความโด่ง

การเลือกใช้สูตรสถิติในการบรรยายลักษณะหรือการแจกแจงข้อมูลจะต้องคำนึงถึงมาตรการวัดที่ได้จากข้อมูล หรือคุณลักษณะพื้นฐานของค่าตัวเลขที่ได้จากการวัด เพื่อการใช้สูตรได้ตรงตาม ข้อตกลงและการแปลความหมายที่ถูกต้อง

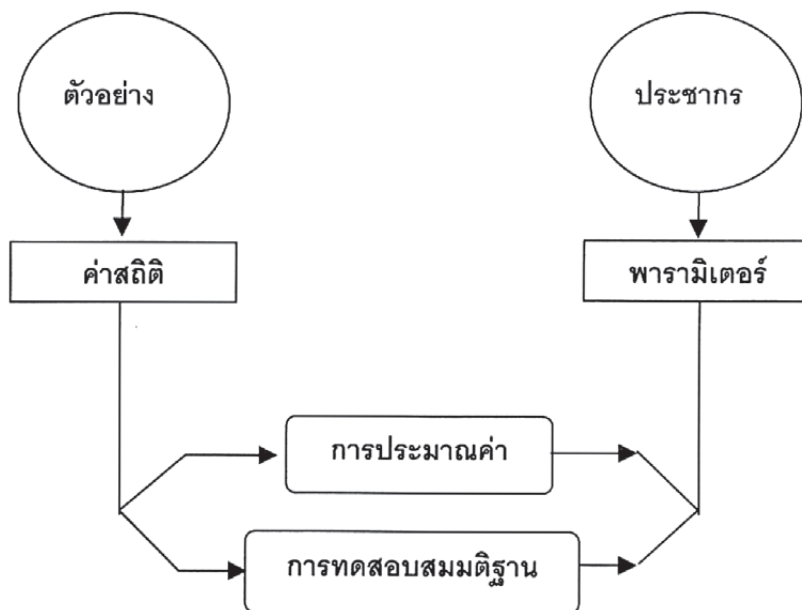
### 2. สถิติอ้างอิงหรือสถิติอนุมาน (Inference Statistics)

เป็นสถิติที่ศึกษาและรวบรวมข้อมูลจากตัวอย่างของประชากร แล้วทำการวิเคราะห์ข้อมูลเพื่อประมาณ คาดคะเน พยากรณ์ และสรุปอ้างอิงหรือนำไปสู่การตัดสินใจไปยังประชากรเป้าหมายในแง่ที่ทำการศึกษา สำหรับการตีความหมายในขั้นสรุปอ้างอิงนั้น ตั้งอยู่บนฐานของข้อมูลจากสถิติบรรยายกับทฤษฎีความน่าจะเป็น ในการเกิดคำตอบที่ต้องการโดยอาศัยหลักการของการอุปมาน (inductive) ซึ่งเป็นการใช้เหตุผลจากกรณีเฉพาะไปสู่กรณีทั่วไป โดยศึกษาในเรื่อง

1. ทฤษฎีความน่าจะเป็น
2. การสุ่มตัวอย่าง
3. การประมาณค่าของประชากร
4. การทดสอบสมมติฐาน

การเลือกใช้สูตรสถิติในการสรุปอ้างอิงไปถึงประชากรเป้าหมายนั้น จะต้องคำนึงถึงคุณสมบัติเบื้องต้นของข้อมูลชุดนั้น ขนาดของกลุ่มตัวอย่าง ประชากร วิธีการสุ่มตัวอย่าง จำนวนของกลุ่มตัวอย่าง เช่น กลุ่มเดียว สองกลุ่ม สามกลุ่มหรือมากกว่า ลักษณะการเป็นอิสระต่อกัน ของข้อมูลหรือขึ้นต่อกันของข้อมูล จำนวนตัวแปรในแต่ละการแจกแจง ลักษณะการแจกแจงของประชากร ตลอดจนความน่าจะเป็นของการเกิดคำตอบที่ต้องการ กล่าวคือ เราต้องการทราบว่า ที่เราประมาณ คาดคะเน พยากรณ์ และสรุปอ้างอิงนั้น มีความน่าจะเป็นที่จะถูกต้องหรือผิดพลาด เพียงใด น่าเชื่อถือหรือไม่

กระบวนการของสถิติอ้างอิงหรือสถิติอนุมาน แสดงได้ด้วยแผนภูมิดังนี้



## ประชากรและตัวอย่าง (Population and Sample)

ในการทดลองทางสถิติ เหตุการณ์ต่าง ๆ หรือผลลัพธ์ต่าง ๆ ที่เกิดขึ้นนั้นอาจมีค่าเป็นตัวเลขหรือไม่เป็นตัวเลขก็ได้ แต่เราสามารถจะกำหนดเป็นค่าตัวเลขได้ เช่น เมื่อโยนเหรียญ 1 อัน เหรียญจะขึ้นหัวหรือก้อย แต่เราอาจกำหนดเป็นตัวเลข เช่น กำหนดหัวเป็น 1 และก้อยเป็น 0 หรือเมื่อโยนเหรียญ 2 อัน ผลที่เกิดขึ้น คือ HH, HT, TH และ TT เราอาจกำหนดโดยนับจำนวนเหรียญที่ขึ้นหัวจะได้เป็น 0, 1, 2 เป็นต้น ซึ่งค่าตัวเลขที่เรากำหนดขึ้นนี้เรียกว่า ตัวแปรสุ่ม (Random Variable) และในการทดลองนั้นค่าตัวเลขที่ได้จากการสังเกตแต่ละครั้งเรียกว่า ค่าสังเกต (Observation) ค่าสังเกตนี้บางครั้งเราเรียกว่า คะแนน ค่าสังเกตหรือคะแนนที่เรารวบรวมได้จากการทดลองหรือจากการสังเกต ก็จะเป็นข้อมูลที่นำมาคำนวณวิเคราะห์ต่อไป

**นิยาม** ประชากร หมายถึง เซตของค่าสังเกตของสิ่งที่ต้องการศึกษาทั้งหมด ที่เกี่ยวข้องในการทดลอง หรือในการสำรวจทางสถิติ หรือหมายถึง กลุ่มของสิ่งต่าง ๆ ทั้งหมดหรือข้อมูลที่เราจะศึกษาทั้งหมด

ประชากรเป็นค่าสังเกตที่ได้มาจากกลุ่มคนหรือกลุ่มของสัตว์ พืช หรือสิ่งอื่น ๆ ที่เราสนใจ เช่น ถ้าต้องการทราบอายุการใช้งานเฉลี่ยของแบตเตอรี่ของโรงงานหนึ่ง ประชากรคืออายุการใช้งานเฉลี่ยของแบตเตอรี่ของโรงงานนั้นทั้งหมด และค่าสังเกตของประชากรก็คือตัวแปรสุ่ม เช่น เมื่อโยนลูกเต๋า 2 ลูก และให้  $X$  เป็นผลบวกของแต้มลูกเต๋านั้น  $X$  จะมีค่าเป็นเลขจำนวนเต็ม คือ 2, 3, 4, ..., 12

จำนวนค่าสังเกตทั้งหมดของประชากร หมายถึง ขนาดของประชากรนั้น แทนด้วยสัญลักษณ์  $N$  เช่น ในการสำรวจความคิดเห็นของนักศึกษามหาวิทยาลัยหนึ่งที่มีจำนวน 2,000 คน ขนาดของประชากรคือ  $N = 2,000$

ประชากรอาจแบ่งตามขนาดได้เป็น 2 ชนิดคือ

1. ประชากรที่มีจำนวนจำกัด (Finite Population) เช่น จำนวนรถยนต์ในจังหวัดเชียงใหม่ จำนวนครัวเรือนในกรุงเทพฯ จำนวนนักศึกษาในมหาวิทยาลัยหนึ่ง เป็นต้น
2. ประชากรที่มีจำนวนอนันต์ (Infinite Population) เช่น จำนวนครั้งที่โยนเหรียญ 1 อัน แล้วได้ก้อย จำนวนข่าวที่เก็บเกี่ยวได้ในปีหนึ่ง ๆ เป็นต้น

ในบางครั้งจำนวนของประชากรที่มีจำนวนจำกัดมีค่ามาก เราสามารถอนุมานให้เป็นประชากรที่มีจำนวนอนันต์ได้ ประกอบกับในการศึกษาทฤษฎีต่าง ๆ เกี่ยวกับประชากรที่มีจำนวน

อนันต์จะง่ายกว่าประชากรที่มีจำนวนจำกัด และการที่เราสมมติให้ประชากรที่มีจำนวนจำกัด เป็นประเภทประชากรที่มีจำนวนอนันต์ ก็ไม่ทำให้เกิดความผิดพลาดมากนัก ดังนั้นจึงอนุโลมให้ประชากรที่มีจำนวนจำกัดแต่มีจำนวนมากและแน่นอนเป็นประชากรที่มีจำนวนอนันต์

**นิยาม** ค่าที่แสดงคุณสมบัติหรือลักษณะของประชากรเรียกว่า **พารามิเตอร์** (Parameter)

จากนิยาม ค่าเฉลี่ยเลขคณิต ส่วนเบี่ยงเบนมาตรฐาน เรียกว่า พารามิเตอร์ของประชากร นั้น นิยมใช้แทนด้วยอักษรกรีก เช่น ค่าเฉลี่ยเลขคณิตแทนด้วย  $\mu$  ส่วนเบี่ยงเบนมาตรฐานแทนด้วย  $\sigma$  ถ้าประชากรที่มีจำนวนจำกัด เราสามารถหาค่าพารามิเตอร์จากประชากรโดยตรง แต่ถ้าประชากรที่มีจำนวนอนันต์ ค่าพารามิเตอร์ก็จะเป็นค่าทางทฤษฎี ซึ่งอาจหาค่าได้จากการแจกแจงความน่าจะเป็นของตัวแปรสุ่มที่สอดคล้องกับประชากร แต่ถ้าเราไม่ทราบการแจกแจงความน่าจะเป็น เราก็ไม่สามารถหาค่าพารามิเตอร์ได้ โดยปกติแล้วประชากรส่วนใหญ่จะมีลักษณะเป็นประเภทประชากรที่มีจำนวนอนันต์ และไม่ทราบการแจกแจงความน่าจะเป็น เราก็ไม่สามารถที่จะแสดงค่าต่าง ๆ ทั้งหมดของประชากรได้ วิธีที่เราหาค่าพารามิเตอร์ได้โดยการประมาณค่าที่คำนวณจาก ตัวอย่าง

**นิยาม** ตัวอย่าง หมายถึง สับเซตของประชากร

เรากล่าวได้ว่าตัวอย่างเป็นข้อมูลบางส่วนที่เราศึกษาเพื่ออ้างอิงถึงประชากร เมื่อเราจะประมาณค่าพารามิเตอร์ของประชากร โดยพิจารณาค่าที่ได้จากตัวอย่าง ซึ่งถ้าหากเราไม่สามารถรวบรวมข้อมูลหรือศึกษาจากประชากรทั้งหมดได้ อาจเนื่องมาจากเปลืองค่าใช้จ่ายและเสียเวลามาก เช่น ถ้าต้องการทราบอายุการใช้งานของหลอดไฟที่ผลิตโดยโรงงานแห่งหนึ่ง เราไม่สามารถจะทดสอบหลอดไฟทุก ๆ หลอดที่ผลิตจากโรงงานนั้นได้ จึงจำเป็นต้องเลือกประชากรมาบางส่วน ซึ่งเป็นตัวแทนของประชากรมาจำนวนหนึ่ง เรียกว่า ตัวอย่าง จึงจะทำให้ประหยัดค่าใช้จ่ายและเวลาเพราะข้อมูลมีจำนวนน้อย แต่ตัวอย่างที่เลือกมานั้น จะต้องเป็นตัวแทนที่ดีของประชากร โดยที่ลักษณะของตัวอย่างที่เลือกมา ควรมีลักษณะเหมือนกับประชากรทั้งหมด เพื่อสามารถประมาณค่าที่ได้จาก ตัวอย่างใกล้เคียงกับประชากร และผลสรุปจะมีความถูกต้องเพียงใดนั้น ก็จะขึ้นอยู่กับองค์ประกอบที่สำคัญคือ

1. วิธีการสุ่มตัวอย่าง
2. วิธีการประมาณค่าต่าง ๆ ทางสถิติ

## วิธีการสุ่มตัวอย่าง (Random Sampling)

วิธีการสุ่มตัวอย่างอาจแบ่งกว้าง ๆ ได้ 2 วิธีคือ

### 1. การสุ่มตัวอย่างโดยไม่ใช้ความน่าจะเป็น (Nonprobability Random Sampling)

วิธีการสุ่มตัวอย่างนี้ ขึ้นอยู่กับความรู้สึกและสามัญสำนึกของผู้เก็บข้อมูลเป็นหลัก ซึ่งสามารถแบ่งออกได้เป็น 3 วิธีคือ

1.1 การสุ่มตัวอย่างโดยความสะดวก ตัวอย่างที่ได้จากบางส่วนของประชากรหาได้สะดวกและง่ายเท่านั้น แต่ข้อมูลที่ได้มาเป็นตัวอย่างจะมีความลำเอียง (bias) มาก

1.2 การสุ่มตัวอย่างโดยเฉพาะเจาะจง ตัวอย่างที่ได้จากบางส่วนของประชากรจะใช้การตัดสินใจ ความรู้สึกส่วนตัวของผู้เก็บข้อมูล ที่คิดว่าจะเป็นตัวแทนที่ดีของประชากร ซึ่งจะต้องอาศัย ผู้ที่มีความชำนาญและเชี่ยวชาญอย่างแท้จริงในการเก็บข้อมูลจากตัวอย่าง

1.3 การสุ่มตัวอย่างตามโควตา ตัวอย่างที่ได้จากบางส่วนของประชากรจะมีการกำหนดจำนวนข้อมูลตามประเภทไว้ หรือควบคุมโควตาตามลักษณะที่ต้องการ เช่น กำหนดเพศ อายุ ระดับการศึกษา เงินเดือน เป็นต้น เมื่อได้ครบจำนวนตัวอย่างที่ต้องการแล้วจะไม่ยอมรับหน่วยตัวอย่างที่พบเพิ่มเติมอีกจะช่วยให้ประหยัดเวลาค่าใช้จ่ายมาก แต่มีความลำเอียงมากเหมือนกัน

### 2. การสุ่มตัวอย่างโดยใช้ความน่าจะเป็น (Probability Random Sampling)

โดยทั่วไปนักสถิตินิยมใช้การสุ่มตัวอย่างโดยใช้ความน่าจะเป็น ซึ่งเป็นการสุ่มอย่างไม่เฉพาะเจาะจงว่า สมาชิกใดของประชากรจะต้องเป็นตัวอย่าง โดยให้แต่ละสมาชิกของประชากรมีความน่าจะเป็นในการรับเลือกเข้าเป็นตัวอย่าง ด้วยความน่าจะเป็นเท่า ๆ กัน โดยไม่มีการใช้ความรู้สึกตัดสินใจ ไม่ใช้ความสะดวกหรือไม่ใช้ความบังเอิญเกี่ยวกับตัวอย่างนั้น ๆ ตัวอย่างที่เลือกมาโดยวิธีการสุ่มนี้ เรียกว่า ตัวอย่างสุ่ม (Random Sample) การได้ตัวอย่างสุ่มนี้จะทำให้เราได้ตัวแทนที่ดีของประชากร ตัวอย่างแต่ละตัวอย่างจะมีความน่าจะเป็นที่จะถูกเลือกมาใช้เท่าใดนั้น ขึ้นอยู่กับแบบแผนของการสุ่มตัวอย่าง ซึ่งแบ่งได้หลายวิธี ซึ่งวิธีการสุ่มตัวอย่างที่นิยมใช้กันมี 5 วิธีคือ

2.1 การสุ่มตัวอย่างแบบง่าย

2.2 การสุ่มตัวอย่างแบบมีระบบ

2.3 การสุ่มตัวอย่างตามระดับชั้น

2.4 การสุ่มตัวอย่างตามกลุ่ม

2.5 การสุ่มตัวอย่างแบบหลายขั้นตอน

การเลือกตัวอย่างชนิดนี้ เป็นการเลือกเพื่อได้มาซึ่งตัวอย่างที่มีความยุติธรรม ถือว่าเป็นการสุ่มตัวอย่างที่ดี เพราะข้อมูลแต่ละค่าหรือแต่ละหน่วยมีโอกาสถูกเลือกได้เท่า ๆ กัน

## 2.1 การสุ่มตัวอย่างแบบง่าย (Simple Random Sampling)

เป็นวิธีการสุ่มตัวอย่าง โดยให้ประชากรมีความน่าจะเป็นถูกเลือกเท่า ๆ กัน และแต่ละประชากรที่ถูกเลือกเป็นอิสระต่อกันไม่เกี่ยวพันกัน ซึ่งประชากรนั้นควรมีลักษณะคล้ายคลึงกัน (Homogeneity) ไม่แตกต่างกันมากนัก การสุ่มตัวอย่างแบบง่ายนี้ บางทีเรียกว่า การสุ่มตัวอย่างแบบธรรมดา ในทางปฏิบัติมี 2 วิธีดังนี้

### 1. แบบจับฉลาก (Lottery)

ในการใช้วิธีแบบจับฉลาก เช่น ถ้าเราต้องการคัดเลือกนักศึกษา 50 คน จากนักศึกษาทั้งหมดซึ่งเป็นประชากร 600 คน ซึ่งวิธีนี้เราเขียนชื่อลงในบัตร 600 แผ่น จับบัตรขึ้นมาทีละ 1 แผ่น จดบันทึกไว้แล้วนำบัตรนั้นคืนไปรวมในกลุ่มบัตรเพื่อรวมให้ได้ 600 แผ่นเสมอ เพื่อให้โอกาสได้รับเลือกเป็น  $\frac{1}{600}$  เสมอ ทำเช่นนี้เรื่อย ๆ จนครบ 50 แผ่น เท่ากับจำนวนตัวอย่างที่ต้องการ หรืออาจจะไม่นำบัตรใส่คืนลงไปก็ได้ แต่จะทำให้ลักษณะการแจกแจงตัวอย่างที่ได้มีความแตกต่างกัน แต่ถ้าจำนวนประชากรที่มีจำนวนอนันต์ ลักษณะการแจกแจงของตัวอย่างที่ได้จากแบบใส่คืนและไม่ใส่คืนจะเป็นแบบเดียวกัน ในทางปฏิบัติเรามักจะใช้วิธีการสุ่มแบบไม่ใส่คืน เพราะโดยทั่วไปแล้วประชากรมักจะมีจำนวนอนันต์

### 2. แบบใช้ตารางเลขสุ่ม (Random Number Table)

ในการใช้ตารางเลขสุ่ม จะต้องกำหนดหมายเลขให้กับทุกค่าของประชากรก่อน แล้วเลือก ตัวเลขตัวแรกอย่างสุ่ม โดยการชี้ไปที่เลขใดเลขหนึ่งโดยไม่เจาะจงในตารางเลขสุ่ม จากนั้นจะอ่าน ตัวเลขจากขวาไปซ้าย จากซ้ายไปขวา จากข้างบนลงข้างล่าง หรืออ่านตามแนวทะแยงก็ได้ การที่จะเลือกใช้เลขที่หลักนั้นพิจารณาจากจำนวนประชากร ถ้าจำนวนประชากรมี 2 หลัก ก็อ่านเลข 2 ตัว ถ้ามี 3 หลักก็อ่าน 3 ตัว ดังตัวอย่างต่อไปนี้

ถ้าประชากรมีขนาด 50 และต้องการตัวอย่างขนาด 5 ก็ให้หมายเลขประชากรตั้งแต่ 00, 01, 02, 03, ..., 49 แล้วอ่านตัวเลขในตารางเลขสุ่มทีละ 2 ตัว โดยการสุ่มตัวเลขตัวหนึ่งเสียก่อนเมื่อเริ่มต้น แล้วอ่านจากซ้ายไปขวาทีละ 2 ตัว ดังนี้