สถิติเบื้องต้นโดยใช้ MS Excel
สำหรับ R2R

ดร. ดำรง ไม้เรียง

ห้องปฏิบัติการเทคโนโลยีชีวภาพทางการแพทย์ (ศิริราช)

BIOTEC

สวทช.

BIOTEC
a member of NSTDA

มหาวิทยาลัยมหิดล
คณะแพทยศาสตร์ศิริราชพยาบาล

หัวข้อบรรยาย

- การแสดงข้อมูล (Data visualization)
- สถิติเชิงพรรณนา (Descriptive statistics)
- การคำนวณขนาดตัวอย่าง (Sample size calculation)
- การทดสอบสมมติฐาน (Hypothesis testing)

~~หัวข้อบรรยาย~~ วันนี้เราจะมาคุยเรื่อง

- ~~การแสดงข้อมูล (Data visualization)~~ วาดกราฟ ตาราง และอื่นๆ
- ~~สถิติเชิงพรรณนา (Descriptive statistics)~~ สรุปข้อมูล
- ~~การคำนวณขนาดตัวอย่าง (Sample size calculation)~~ เก็บตัวอย่างเท่าไหร่ดี
- ~~การทดสอบสมมติฐาน (Hypothesis testing)~~ สิ่งที่เห็นว่าต่าง แน่ใจว่าต่างกันจริงไหม?

วาดกราฟ ตาราง และอื่นๆ

วาดกราฟ ตาราง และอื่นๆ

- ตัวอย่างข้อมูลผู้ป่วยไข้เลือดออก (สมมติ)

Patient_ID	Enrol_date	Site	Serotype	Age
00001	01/01/10	Khon Kaen	Dengue 3	12
00002	01/01/10	Loei	Dengue 1	2
00003	01/02/10	Loei	Dengue 3	4
00004	01/04/10	Loei	Dengue 1	13
00005	01/05/10	Songkhla	Dengue 3	12
00006	01/05/10	Khon Kaen	Dengue 4	5
00007	01/07/10	Songkhla	Dengue 3	3
00008	01/09/10	Loei	Dengue 2	7
00009	01/10/10	Loei	Dengue 3	13
00010	01/11/10	Loei	Dengue 2	6

วาดกราฟ ตาราง และอื่นๆ

- เข้าใจข้อมูล โดย Filter
  - รพ. ศิริราช
  - รพ. ขอนแก่น
  - รพ. เลย
  - รพ. สงขลา

To Excel!!!

วาดกราฟ ตาราง และอื่นๆ

- เข้าใจข้อมูล โดย Filter
  - วันที่ผู้ป่วยรับเข้า รพ. (ค.ศ. 2010 – 2014 หรือ พ.ศ. 2553 - 2557)
  - ผลตรวจชนิดของไวรัส (ชนิดที่ 1, 2, 3 และ 4)
  - อายุผู้ป่วย (0 – 15 ปี)

To Excel!!!

วาดกราฟ ตาราง และอื่นๆ

- อยากรู้ง่าย ๆ
  - ผู้ป่วยกี่ราย (นับแถว)
  - อายุเฉลี่ย (ใช้สูตร)
  - นับชนิดไวรัส (ใช้สูตร หรือ filter)

To Excel!!!

วาดกราฟ ตาราง และอื่นๆ

- อยากรู้ยาก ๆ
  - ผู้ป่วยกี่ราย แบ่งตามปี และโรงพยาบาล (กราฟแท่ง)
  - ผู้ป่วยกี่ราย แบ่งตามเดือนแบบอนุกรมเวลา แบ่งตามโรงพยาบาล (กราฟเส้น)
  - ผู้ป่วยกี่ราย แบ่งตามอายุ และโรงพยาบาล (กราฟแท่ง)

วาดกราฟ ตาราง และอื่นๆ

- Pivot table/chart เครื่องมือที่ถูกลืม
  - “ตารางหมุน”
  - ค่อนข้างใช้เวลาเรียนรู้
  - เมื่อเข้าใจจะมีประโยชน์สูงมาก

To Excel!!!

สรุปข้อมูล

สรุปข้อมูล

- เขียน Function
  - Count นับ
  - Mean, Median ค่าเฉลี่ย (ค่ากลางแบบคำนวณ), ค่ามัธยฐาน (ค่ากลางแบบนับ)
  - Min, Max, Quantile/Percentile มากสุด, น้อยสุด, ค่าเปอร์เซ็นไทล์
  - Standard deviation ค่าเบี่ยงเบนมาตรฐาน (ข้อมูลกระจายตัวอย่างไร)

To Excel!!!

สรุปข้อมูล

- ใช้ Status bar
  - เลือก แล้วดูผลเร็วๆ
  - Average
  - Count
  - Numerical count (นับตัวเลขเท่านั้น)
  - Min
  - Max
  - Sum

To Excel!!!

สรุปข้อมูล

- Pivot table
  - สรุปข้อมูลยากๆ แบบเร็ว
  - ค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐานของอายุ ของแต่ละ รพ.

To Excel!!!

เก็บตัวอย่างเท่าไหร่ดี

เก็บตัวอย่างเท่าไหร่ดี

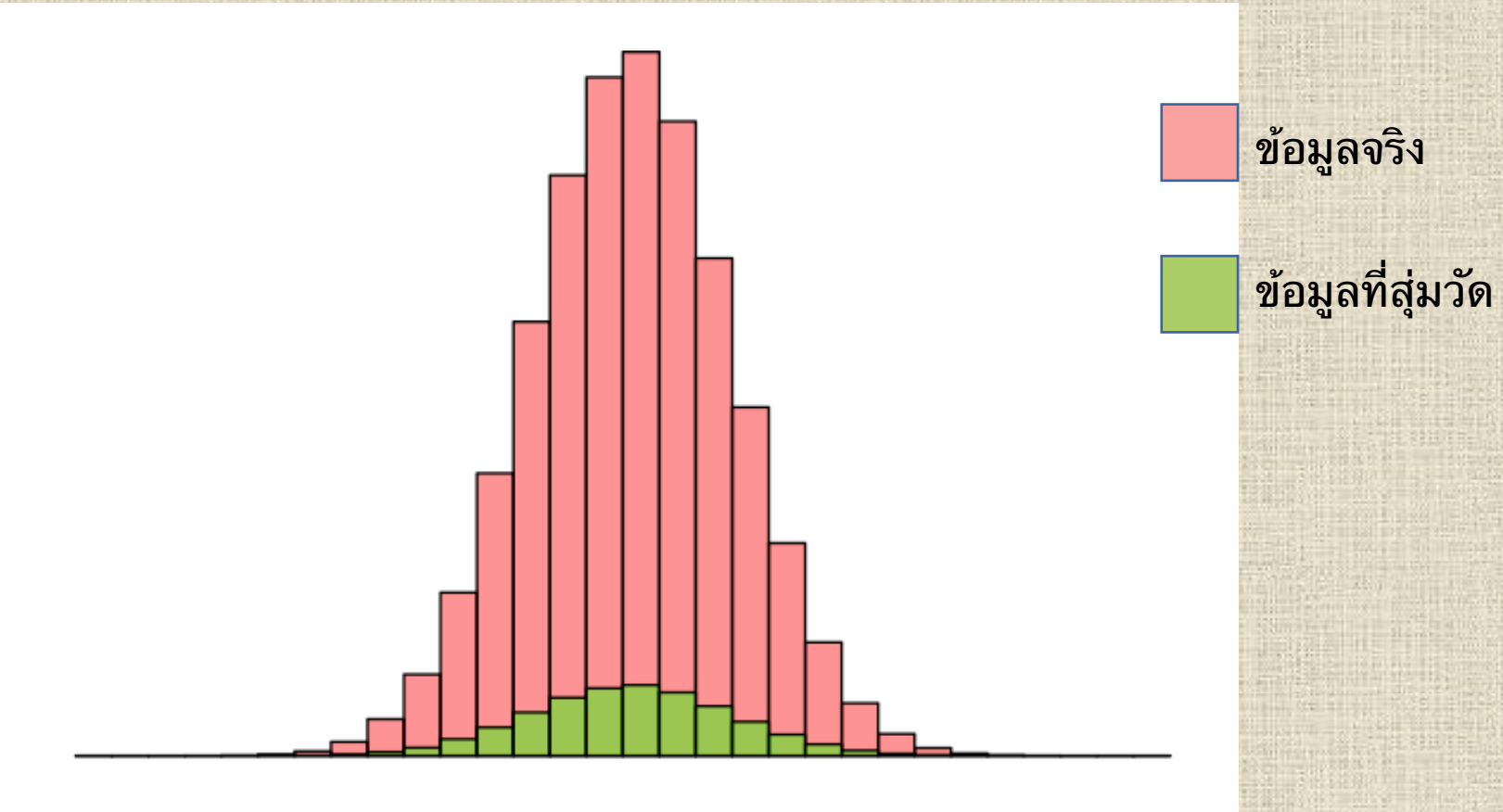
- ผลจากการทำวิจัย → เอาไปใช้ “จริง”
- แล้วมั่นใจได้อย่างไรว่าผล เป็น “จริง”

เก็บตัวอย่างเท่าไหร่ดี

- ส่วนสูงของเด็กไทย (อายุ < 15 ปี)
- ปี 2559 (CIA world fact book)
  - ชาย 6,000,434
  - หญิง 5,714,464
- วัดส่วนสูง 12 ล้านคน???

เก็บตัวอย่างเท่าไหร่ดี

- สุ่มตัวอย่าง = วัดส่วนน้อย สะท้อนส่วนใหญ่



เก็บตัวอย่างเท่าไหร่ดี

- สุ่มตัวอย่าง = วัดส่วนน้อย สะท้อนส่วนใหญ่
- จำนวนตัวอย่าง

sample size	mean	stdev
12000000	100.02792	19.9797
1200000	100.04108	19.98406
120000	99.97564	20.04038
12000	99.87991	20.07676
1200	99.51937	20.32255
120	103.61993	21.13232
12	98.07442	19.27547
6	97.72924	12.53156

ค่าจริง

เก็บตัวอย่างเท่าไหร่ดี สำหรับค่าเฉลี่ย

$$n = \frac{z^2_{1-\frac{\propto}{2}} \sigma^2}{d^2}$$

- ต้องการความเชื่อมั่นเท่าไหร่ (ทั่วไป 95%)
  - คำนวณ ค่า z
- ค่าเบี่ยงเบนมาตรฐาน
  - คิดว่าเป็นเท่าไหร่ (ดูจากการศึกษาก่อนหน้า)
  - ยิ่งมาก ตัวอย่างยิ่งเยอะ
- อนุญาตให้ค่าเฉลี่ยที่หาได้มีความคลาดเคลื่อนจาก “ของจริง” เท่าไหร่
  - ยิ่งน้อย ตัวอย่างยิ่งเยอะ

To Excel!!!

เก็บตัวอย่างเท่าไหร่ดี สำหรับ %/สัดส่วน

- ต้องการความเชื่อมั่นเท่าไหร่ (ทั่วไป 95%)
  - คำนวณ ค่า z
- สัดส่วน / %
  - คิดว่าเป็นเท่าไหร่ (ดูจากการศึกษาก่อนหน้า)
  - ยิ่งใกล้ 50% หรือ 1:1 ตัวอย่างยิ่งเยอะ
- อนุญาตให้ค่าสัดส่วนที่หาได้มีความคลาดเคลื่อนจาก “ของจริง” เท่าไหร่
  - ยิ่งน้อย ตัวอย่างยิ่งเยอะ

$$n = \frac{z^2_{1-\frac{\propto}{2}}\, p(1-p)}{d^2}$$

To Excel!!!

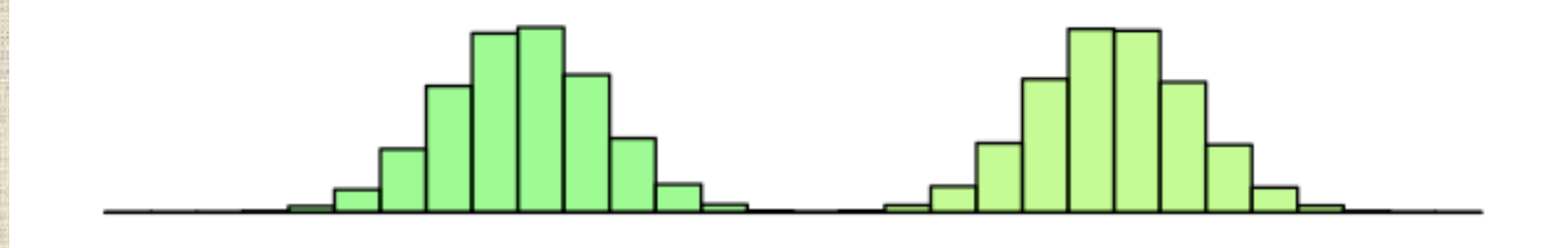
สิ่งที่เห็นว่าต่าง แน่ใจว่าต่างกันจริงไหม?

(ทดสอบ 2 กลุ่ม)

สิ่งที่เห็นว่าต่าง แน่ใจว่าต่างกันจริงไหม?

- สุ่มตัวอย่าง = วัดส่วนน้อย สะท้อนส่วนใหญ่
- การเลือกตัวอย่างจากประชากร

A B



ข้อมูลจริงกลุ่ม A

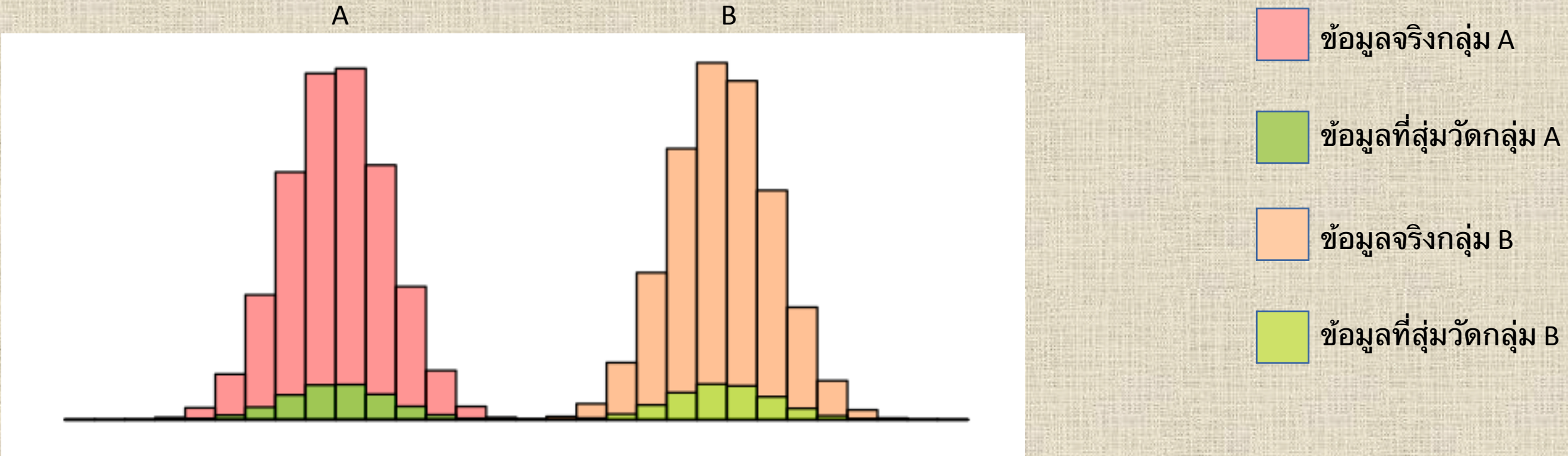
ข้อมูลที่สุ่มวัดกลุ่ม A

ข้อมูลจริงกลุ่ม B

ข้อมูลที่สุ่มวัดกลุ่ม B

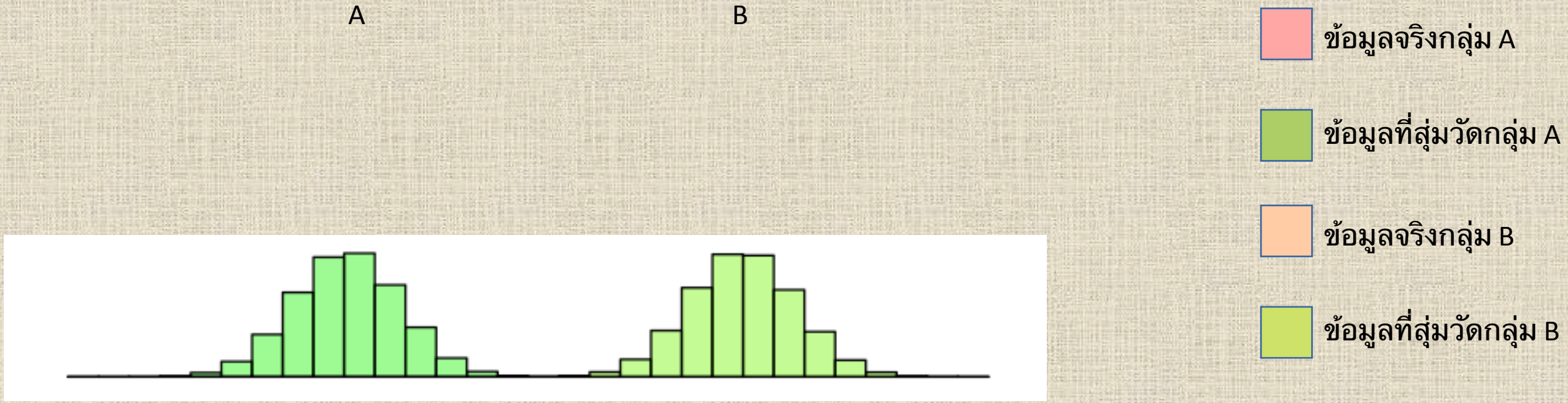
สิ่งที่เห็นว่าต่าง แน่ใจว่าต่างกันจริงไหม?

- สุ่มตัวอย่าง = วัดส่วนน้อย สะท้อนส่วนใหญ่
- แล้วประชากรจริง?



สิ่งที่เห็นว่าต่าง แน่ใจว่าต่างกันจริงไหม?

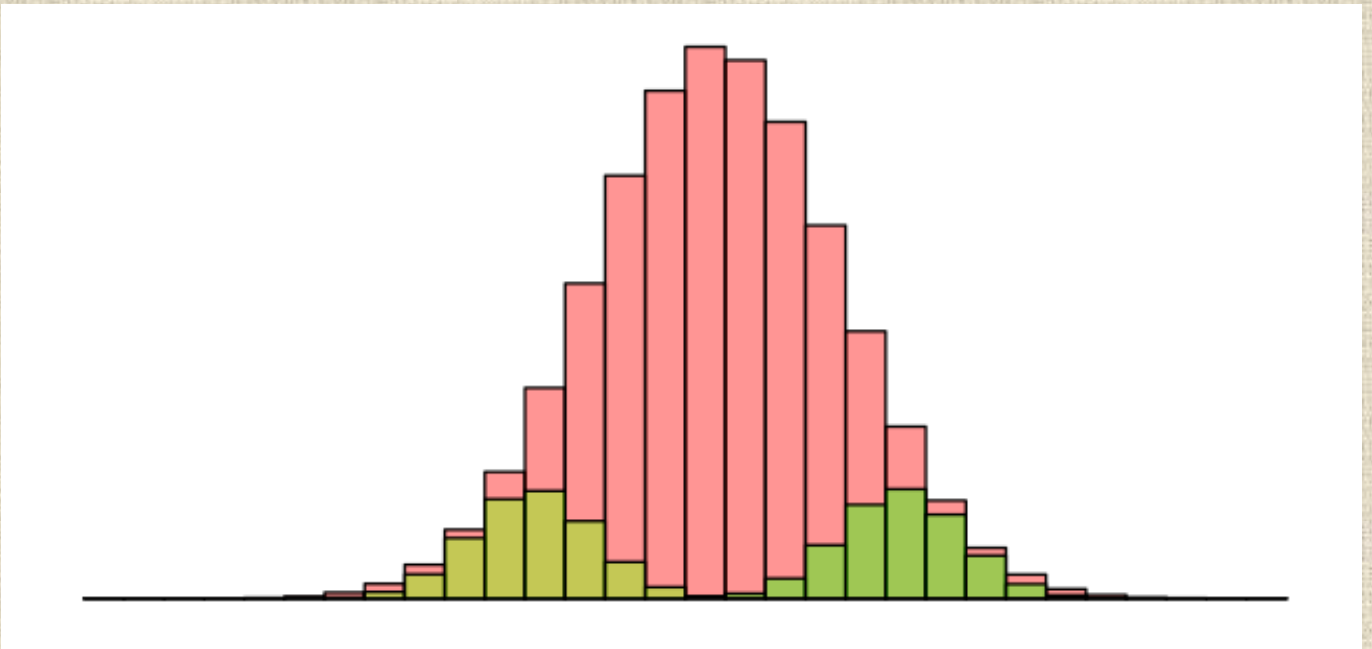
- สุ่มตัวอย่าง = วัดส่วนน้อย สะท้อนส่วนใหญ่



สิ่งที่เห็นว่าต่าง แน่ใจว่าต่างกันจริงไหม?

- สุ่มตัวอย่าง = วัดส่วนน้อย สะท้อนส่วนใหญ่
- การเลือกตัวอย่าง

A = B



ข้อมูลจริงกลุ่ม A + B

ข้อมูลที่สุ่มวัดกลุ่ม A

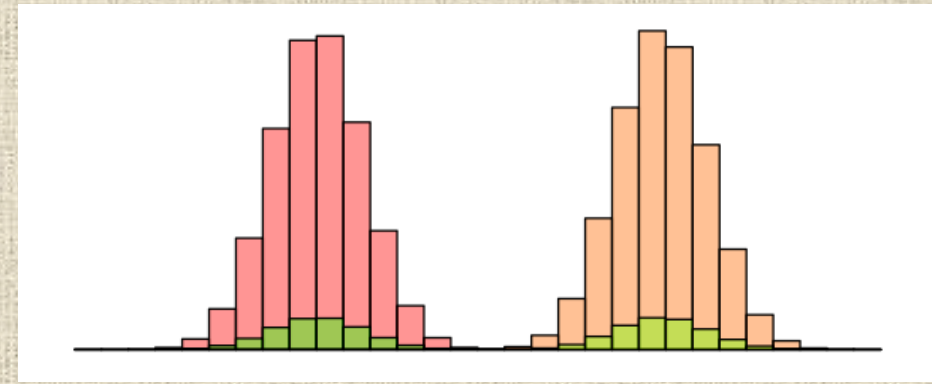
ข้อมูลที่สุ่มวัดกลุ่ม B

สิ่งที่เห็นว่าต่าง แน่ใจว่าต่างกันจริงไหม?

- เราทดสอบว่าโอกาสที่ A = B เกินกว่า 5%, 1% หรือ 0.1% หรือไม่
- เราไม่ได้ทดสอบว่า A ต่างจาก B

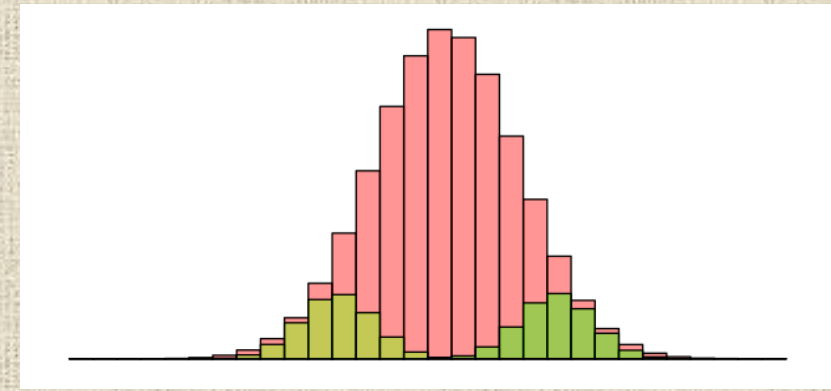
≥ 95%?

A ≠ B



< 5%?

A = B



ข้อมูลจริงกลุ่ม A + B

ข้อมูลที่สุ่มวัดกลุ่ม A

ข้อมูลที่สุ่มวัดกลุ่ม B

แน่ใจว่าต่างกันจริงไหม? สำหรับค่าเฉลี่ย

- Student’s t-test
- ข้อมูลกระจายแบบปกติ เช่น น้ำหนัก ส่วนสูง
  - ไม่ปกติ เช่น เกรด เวลาที่ผู้ป่วยเริ่มกินยาแล้วมีอาการ
- ข้อมูลเทียบเป็นคู่ (Before / After) หรือไม่
- ความแปรปรวน (ดูค่าเบี่ยงเบนมาตรฐาน) เหมือนกันไหม

To Excel!!!

แน่ใจว่าต่างกันจริงไหม? สำหรับ จำนวน/สัดส่วน

- Chi-square test
- ข้อมูลที่นับได้ แบ่งเป็นกลุ่มๆ แล้ว ต้องไม่มีกลุ่มไหนน้อยกว่า 5
- เช่น ผู้ชายสูบบุหรี่ในอัตราส่วนที่ต่างจากผู้หญิงสูบบุหรี่ หรือไม่

To Excel!!!

~~บทสรุป~~ ถ้าลืมหรือไม่เข้าใจควรทบทวนอะไร

- Pivot table/ Pivot chart
- สูตรคำนวณตัวอย่าง
  - ค่าเฉลี่ย $n = z_{1-\alpha/2}^2 \sigma^2 / d^2$
  - สัดส่วน $n = z_{1-\alpha/2}^2 p(1-p)/d^2$

~~บทสรุป~~ ถ้าลืมหรือไม่เข้าใจควรทบทวนอะไร

- chisq.test(ข้อมูลจริง, ข้อมูลที่คาดว่าจะเป็นหากสัดส่วนไม่ต่างกัน)
- t.test(กลุ่ม1, กลุ่ม2, 1 or 2 tailed, คู่ แปรปวนเท่ากัน หรือ แปรปวนไม่เท่ากัน)

ถ้าอยากรู้เพิ่ม

- สถิติวิจัยง่ายนิดเดียว บน Excel 2013