

# Joint Distributions & Hypothesis Testing

อ.ปรัชญ์ ปิยะวงศิศาล

Probability and Statistics for Engineering @ RMUTL อ.ปรัชญ์

## Topics

- Joint Distributions – Two RVs
  - Joint PMF/PDF
  - Marginal PMF/PDF
  - Independent RVs
  - Covariance
- Hypothesis Testing

Probability and Statistics for Engineering @ RMUTL อ.ปรัชญ์

## Joint Distributions – Two RVs

- ในชีวิตจริง เรามักสนใจความสัมพันธ์ระหว่างตัวแปรสุ่ม 2 ตัว  $X, Y$  เช่น
  - ปริมาณน้ำที่รด กับ ความสูงของต้นไม้
  - จำนวนห้องนอน กับ ราคาบ้าน
  - ความดันโลหิต กับ ปริมาณคอเลสเตอรอลในเลือด
  - จำนวนลูกค้าในร้าน กับ เวลาที่ลูกค้าต้องคอยในคิว
  - ฯลฯ
- เราสามารถหา **joint distribution (PMF/PDF)** ของ  $X, Y$  ได้
- ในอนาคตสามารถต่อยอดไปเป็น 3+ ตัวแปรได้ไม่ยาก

Probability and Statistics for Engineering @ RMUTL อ.ปริญญ์

## Joint Distributions – Two RVs

- ในกรณีที่  $X, Y$  เป็น **discrete RV** จะได้ว่า:
  - Joint PMF คือ:  $P_{XY}(x, y) = P(X = x, Y = y)$
  - โดยที่  $\sum_{(x,y) \in S_{XY}} P_{XY}(x, y) = 1$
- ในกรณีที่  $X, Y$  เป็น **continuous RV** จะได้ว่า:
  - Joint PDF คือ:  $f_{XY}(x, y)$
  - โดยที่  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy = 1$
  - และ  $P((X, Y) \in A) = \iint_A f_{XY}(x, y) dx dy$



สำคัญ

Probability and Statistics for Engineering @ RMUTL อ.ปริญญ์


## Exercise

	$Y = 0$	$Y = 1$	$Y = 2$
$X = 0$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{8}$
$X = 1$	$\frac{1}{8}$	$\frac{1}{6}$	$\frac{1}{6}$

- กำหนดให้ตัวแปรสุ่ม  $X$  และ  $Y$  มี joint PMF ดังตาราง
- จงหา
  - $P_{XY}(0, 2)$
  - $P(X = 1, Y \leq 1)$
  - $P(Y = 1)$
  - $P(Y = 1 | X = 0)$
  - $X$  และ  $Y$  เป็นอิสระจากกันหรือไม่

Probability and Statistics for Engineering @ RMUTL อ.ปริญญ์

## Marginal Distributions

- เราสามารถหา PMF ของตัวแปรเดียว  $X$  หรือ  $Y$  จาก joint PMF ได้ โดย
  - $P_X(x) = \sum_{y \in S_Y} P_{XY}(x, y)$
  - $P_Y(y) = \sum_{x \in S_X} P_{XY}(x, y)$

สำคัญ
- เราเรียก  $P_X(x), P_Y(y)$  ที่สกัดมาจาก joint PMF ว่า marginal PMF ของ  $X, Y$
- ในทำนองเดียวกัน สำหรับ  $X, Y$  ที่เป็น continuous เราสามารถหา marginal PDF ของตัวแปรเดียว  $X$  หรือ  $Y$  จาก joint PDF ได้ โดย
  - $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy$
  - $f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx$

Probability and Statistics for Engineering @ RMUTL อ.ปริญญ์

## Exercise

	$Y = 0$	$Y = 1$	$Y = 2$
$X = 0$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{8}$
$X = 1$	$\frac{1}{8}$	$\frac{1}{6}$	$\frac{1}{6}$

- กำหนดให้ตัวแปรสุ่ม  $X$  และ  $Y$  มี joint PMF ดังตาราง
- จงหา marginal PMF  $P_X(x)$  และ  $P_Y(y)$

## Independent RVs

- ความอิสระจากกันระหว่างตัวแปรสุ่ม  $X, Y$  มีหลักการคล้ายกับความอิสระจากกันระหว่างเหตุการณ์  $A, B$  นั่นคือ

- สำหรับ discrete RV:  $X, Y$  จะเป็นอิสระจากกัน (independent) ก็ต่อเมื่อ

$$P_{XY}(x, y) = P_X(x)P_Y(y) \quad ; \text{สำหรับทุกๆ } x, y$$



สำคัญ

- หากมีมากกว่า 2 ตัวแปรที่เป็นอิสระจากกัน เราสามารถแตก joint PMF เป็นผลคูณของ marginal PMF ได้ในทำนองเดียวกัน

## Covariance

- หากพิจารณา 2 ตัวแปรสุ่ม  $X, Y$  ร่วมกัน
- เราสามารถหา  $Cov(X, Y)$  คือ covariance ของ  $X, Y$  โดย  

$$Cov(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$
- ค่าของ covariance จะบ่งชี้ว่า  $X, Y$  มีค่าไปในทิศทางเดียวกันหรือไม่
  - หาก  $Cov(X, Y)$  เป็น + แสดงว่า  $X, Y$  นั้นมีค่าไปในทิศทางเดียวกัน (positively correlated)
  - หาก  $Cov(X, Y)$  เป็น - แสดงว่า  $X, Y$  นั้นมีค่าที่สวนทางกัน (negatively correlated)
- นอกจากนี้เราสามารถนำ covariance มา normalize โดยหารด้วยค่า S.D. ของ  $X, Y$  เพื่อให้ค่าอยู่ในช่วง -1 ถึง 1 เรียกว่า correlation coefficient

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

Probability and Statistics for Engineering @ RMUTL อ.ปริญญ์

## Hypothesis Testing



- ในการทำวิจัย เรามักจำเป็นต้องพิสูจน์ว่า สมมติฐาน (hypothesis) ที่เราตั้ง เป็นจริงอย่างที่คิดหรือไม่
  - เช่น ยา M ที่บริษัท N คิดค้นขึ้นมา ช่วยลดความดันโลหิตได้จริงหรือไม่
- ซึ่งสมมติฐานจะมี 2 ได้กรณี คือ
  - สมมติฐาน  $H_0$ : ยา M ไม่มีผลต่อความดันโลหิต (Null Hypothesis)
  - สมมติฐาน  $H_1$ : ยา M มีผลต่อความดันโลหิต (Alternate Hypothesis)
- ในการจะพิสูจน์ว่า  $H_0$  หรือ  $H_1$  ที่เป็นจริง:
  - เราจำเป็นต้องทดลองกับกลุ่มตัวอย่าง (sample) จำนวนไม่มาก เช่น
    - กลุ่ม A (ได้รับยา) และกลุ่ม B (ไม่ได้รับยา) กลุ่มละ 30 คน
  - วัดค่าความดันโลหิตของทั้งสองกลุ่ม ถ้าพบว่าค่าเฉลี่ย 2 กลุ่มห่างกันมากพอ เราสามารถปฏิเสธ (reject)  $H_0$  ได้
  - เราเรียกรีวิวทดสอบในลักษณะนี้ว่า Hypothesis Testing

Probability and Statistics for Engineering @ RMUTL อ.ปริญญ์

# Hypothesis Testing

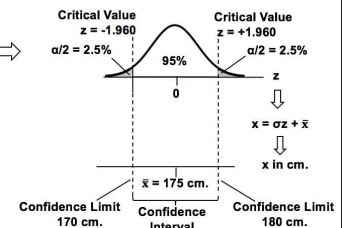
- วิธีหนึ่งที่ใช้กับตัวอย่างในข้างต้นได้ คือ **Student's t-test** มีขั้นตอนคร่าวๆ ดังนี้
  - คำนวณหาค่า **t-value**
    - $t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
    - โดยที่ค่า **t-value** วัดความห่างระหว่าง **mean** ของ 2 กลุ่มตัวอย่าง
  - เช็คค่า **t-value** มีค่ามากกว่า **critical value @ 0.05 significance level** หรือไม่ (ต้องเปิดตาราง)
    - ถ้า **t-value > critical value** เราสามารถ **reject  $H_0$**  ได้ (=mean ของ 2 กลุ่มต่างกันอย่างมีนัยสำคัญทางสถิติ)
    - ถ้า **t-value  $\leq$  critical value** เราไม่สามารถ **reject  $H_0$**  ได้ (=ไม่มีข้อสรุปว่า  $H_1$  จะเป็นจริงหรือไม่)
- **Remark:** ปัจจุบันการทดสอบประสิทธิภาพของยาจะต้องใช้กระบวนการที่ซับซ้อนกว่านี้ ได้แก่ **randomized controlled trial (RCT)**

Probability and Statistics for Engineering @ RMUTL อ.ปริญญ์

## รายละเอียดเพิ่มเติมที่ควรรู้เกี่ยวกับ t-test

- **significance level  $\alpha$**  (เช่น 0.05, 0.01, ...)
  - ความน่าจะเป็นที่จะสังเกตเจอค่าที่ **extreme** มากๆ หาก  $H_0$  เป็นจริง
  - = ความน่าจะเป็นที่เราจะ **reject  $H_0$**  อย่างผิดพลาด หาก  $H_0$  เป็นจริง
- **confidence interval**
  - ช่วงของค่าของ **t-value** ระหว่าง [-critical value, +critical value]
  - สำหรับ 0.05 sig level, ค่า **t-value** มีโอกาสจะตกในช่วงนี้ = 0.95 (95%)
- **degree of freedom**
  - สำหรับ **t-test** ในตัวอย่าง ค่า  $df = n_1 + n_2 - 2$
  - ยิ่งเยอะยิ่งมั่นใจเวลา **reject  $H_0$**
- ชนิดของ **t-test**
  - **one-sample vs two-sample vs paired sample**
  - **one-tailed vs two-tailed**
  - (ตัวอย่างที่เราทำคือ **two-sample, two-tailed t-test**)

I select  $\alpha = 5\%$ .



Reproduced by permission of John Wiley and Sons, Inc.  
from the book Statistics from A to Z – Confusing Concepts Clarified

<https://www.statisticfromator.com/blog/statistics-tip-of-the-week-from-alpha-to-critical-value-to-confidence-interval>

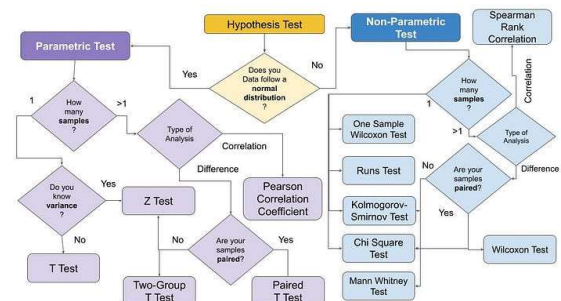
Probability and Statistics for Engineering @ RMUTL อ.ปริญญ์

## Other hypothesis testing methods

- z-test
- F-test, ANOVA
- Chi-square test
- Mann-Whitney
- Wilcoxon

• ...และอื่นๆ

- [https://en.wikipedia.org/wiki/Test\\_statistic](https://en.wikipedia.org/wiki/Test_statistic)
- [https://en.wikipedia.org/wiki/Category:Statistical\\_tests](https://en.wikipedia.org/wiki/Category:Statistical_tests)
- YouTube อธิบายวิธีเลือกค่าต่างๆ: <https://www.youtube.com/watch?v=UaptUhOushw>



<https://towardsdatascience.com/hypothesis-tests-explained-8a070636bd28>