

ความรู้เบื้องต้นเกี่ยวกับประเภทของ สถิติในการวิจัย

ดร.นำชัย ศุภฤกษ์ชัยสกุล
สถาบันวิจัยพฤतिकกรรมศาสตร์
มหาวิทยาลัยศรีนครินทรวิโรฒ



•BSRI



ตัวแปร

- Variable comes from Vary + Able
- So variable is anything that its values can be varied.
- Variables are the phenomena or the topics that researchers want to investigate.

ปรากฏการณ์
Variable

การศึกษาปรากฏการณ์

- ถ้าปรากฏการณ์เป็นเรื่องใหม่ ที่ยังไม่มีผู้ใดศึกษา หรือยังขาดข้อมูลพื้นฐานเกี่ยวกับปรากฏการณ์

Describe



- จะวัดตัวแปรอย่างไร?
 - ตัวเลขที่ใช้แทนค่าตัวแปรมีความหมายอย่างไร
- จะเก็บข้อมูลอย่างไร?
 - เก็บจากประชากรได้ทั้งหมดหรือไม่
- จะบรรยายตัวแปรอย่างไร?
 - สรุปภาพรวม หรือบรรยายความแตกต่าง

การศึกษาปรากฏการณ์

- ถ้าต้องการอธิบายปรากฏการณ์ที่ศึกษา ด้วยการนำเอาปรากฏการณ์อื่นมาช่วยอธิบาย



- จะศึกษาตัวแปรกี่ตัว?
- จะวัดตัวแปรอย่างไร?
- จะเก็บข้อมูลอย่างไร?
- จะบรรยายตัวแปรอย่างไร?
- จะหาความสัมพันธ์ระหว่างตัวแปรอย่างไร?



การเลือกใช้สถิติในงานวิจัย

- ระดับการวัดตัวแปร
 - Difference Testing or Measure of Association
- ประชากรหรือกลุ่มตัวอย่าง
 - Descriptive Statistics or Inferential Statistics
- จำนวนตัวแปรที่ศึกษา
 - Univariate, Bivariate, Multivariate



การวัดคืออะไร?

- การวัด คือ การให้ค่าสัญลักษณ์ ต่อ คุณสมบัติของ สิ่งที่ต้องการศึกษา
- การวัด ก็คือ การให้ค่าตัวเลข ต่อ คุณสมบัติ ของ สิ่งที่ต้องการศึกษา เพื่อที่จะสามารถจำแนกหน่วยตัวอย่างตามลำดับมากหรือน้อยของคุณสมบัติของสิ่งที่ศึกษาได้

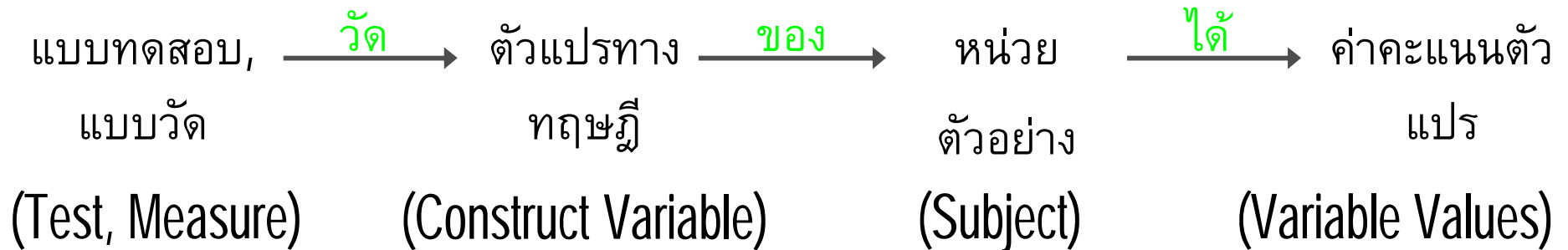
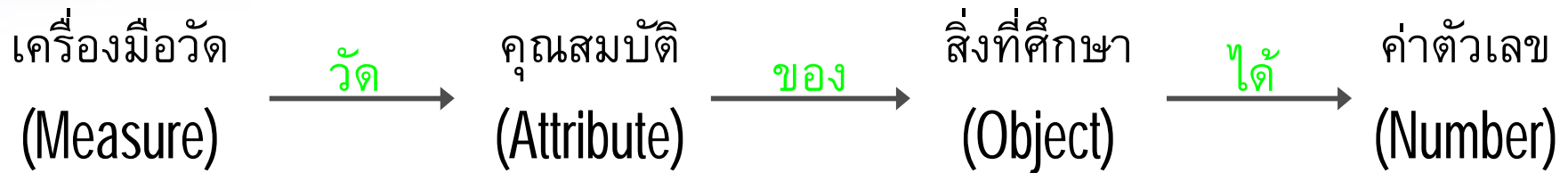
การวัดตัวแปรจริง ประกอบด้วย

เครื่องมือวัด (Measure) $\xrightarrow{\text{วัด}}$ คุณสมบัติ (Attribute) $\xrightarrow{\text{ของ}}$ สิ่งที่ศึกษา (Object) $\xrightarrow{\text{ได้}}$ ค่าตัวเลข (Number)

ไม้บรรทัด $\xrightarrow{\text{วัด}}$ ความยาว $\xrightarrow{\text{ของ}}$ หนังสือ $\xrightarrow{\text{ได้}}$ นิ้ว, ซม.

ตาชั่ง $\xrightarrow{\text{ชั่ง}}$ น้ำหนัก $\xrightarrow{\text{ของ}}$ นักเรียน $\xrightarrow{\text{ได้}}$ กิโลกรัม

ตัวอย่างการวัดในการวิจัยทางสังคมศาสตร์



ตัวอย่างการวัดในการวิจัยทางสังคมศาสตร์

TOEFLS $\xrightarrow{\text{วัด}}$ ความสามารถ $\xrightarrow{\text{ของ}}$ นักศึกษา $\xrightarrow{\text{ได้}}$ คะแนน
ภาษาอังกฤษ

JDI $\xrightarrow{\text{วัด}}$ ความพึงพอใจใน $\xrightarrow{\text{ของ}}$ พนักงาน $\xrightarrow{\text{ได้}}$ ค่าคะแนน
งาน

(Job Description Index) (Job Satisfaction)



ระดับการวัดของตัวแปร

- Nominal Scale
- Ordinal Scale
- Interval Scale
- Ratio Scale

ระดับการวัดของตัวแปร

ระดับ	ตัวเลขใช้เพื่อ	ค่าตัวเลขใช้ระบุถึง	ตัวอย่าง
Nominal	<ul style="list-style-type: none"> จัดกลุ่ม, จัดประเภท 	<ul style="list-style-type: none"> ความแตกต่าง 	ชาย = 1, หญิง = 2
Ordinal	<ul style="list-style-type: none"> จัดกลุ่ม, จัดประเภท เรียงลำดับ 	<ul style="list-style-type: none"> ความแตกต่าง มากกว่าหรือน้อยกว่า 	ประถมศึกษา = 1 มัธยมศึกษา = 2 ปริญญาตรี = 3
Interval	<ul style="list-style-type: none"> จัดกลุ่ม, จัดประเภท เรียงลำดับ มีหน่วยการวัดที่มีช่วงเท่ากัน ไม่มี 0 สมบูรณ์ 	<ul style="list-style-type: none"> ความแตกต่าง มากกว่าหรือน้อยกว่า แต่ละค่าตัวเลขมีระยะห่างเท่ากัน 0 ไม่ได้หมายถึง ไม่มีค่าของตัวแปรนั้นเลย 	คะแนนเจตคติต่อสื่อมวลชน คนที่ 1 = 2.84 คะแนน คนที่ 2 = 3.65 คะแนน คนที่ 3 = 4.12 คะแนน เป็นต้น
Ratio	<ul style="list-style-type: none"> จัดกลุ่ม, จัดประเภท เรียงลำดับ มีหน่วยการวัดที่มีช่วงเท่ากัน มี 0 สมบูรณ์ 	<ul style="list-style-type: none"> ความแตกต่าง มากกว่าหรือน้อยกว่า แต่ละค่าตัวเลขมีระยะห่างเท่ากัน 0 หมายถึง ไม่มีค่าของตัวแปรนั้นเลย 	รายได้ต่อเดือน คนที่ 1 = 0 บาท คนที่ 2 = 7,650 บาท คนที่ 3 = 9,000 บาท เป็นต้น



ประเภทตัวแปร

- ตัวแปรจัดประเภทหรือมีค่าไม่ต่อเนื่อง (Categorical or Discrete Variable)
 - Nominal Scale
 - Ordinal Scale
- ตัวแปรเชิงปริมาณหรือมีค่าต่อเนื่อง (Numerical or Continuous Variable)
 - Interval Scale
 - Ratio



ประเภทของสถิติ

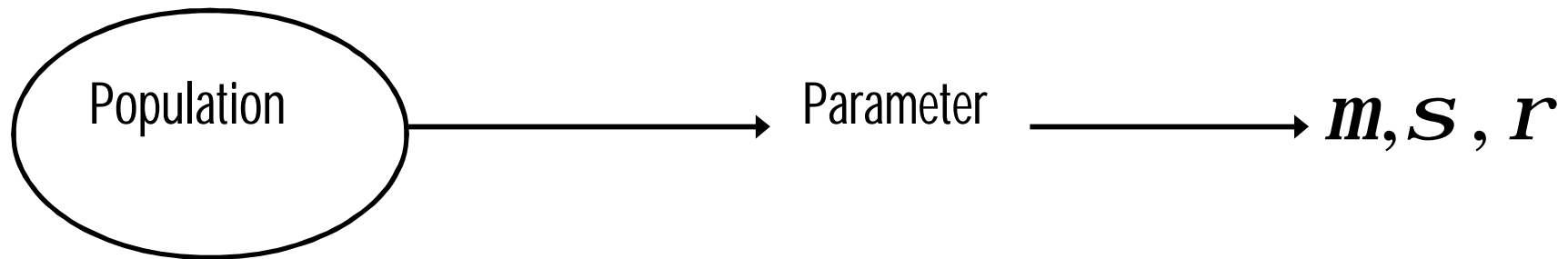
- สถิติเชิงพรรณนา (Descriptive Statistics)
- สถิติเชิงอนุมาน (Inferential Statistics)



2 กรณีในการวิเคราะห์ข้อมูลงานวิจัย

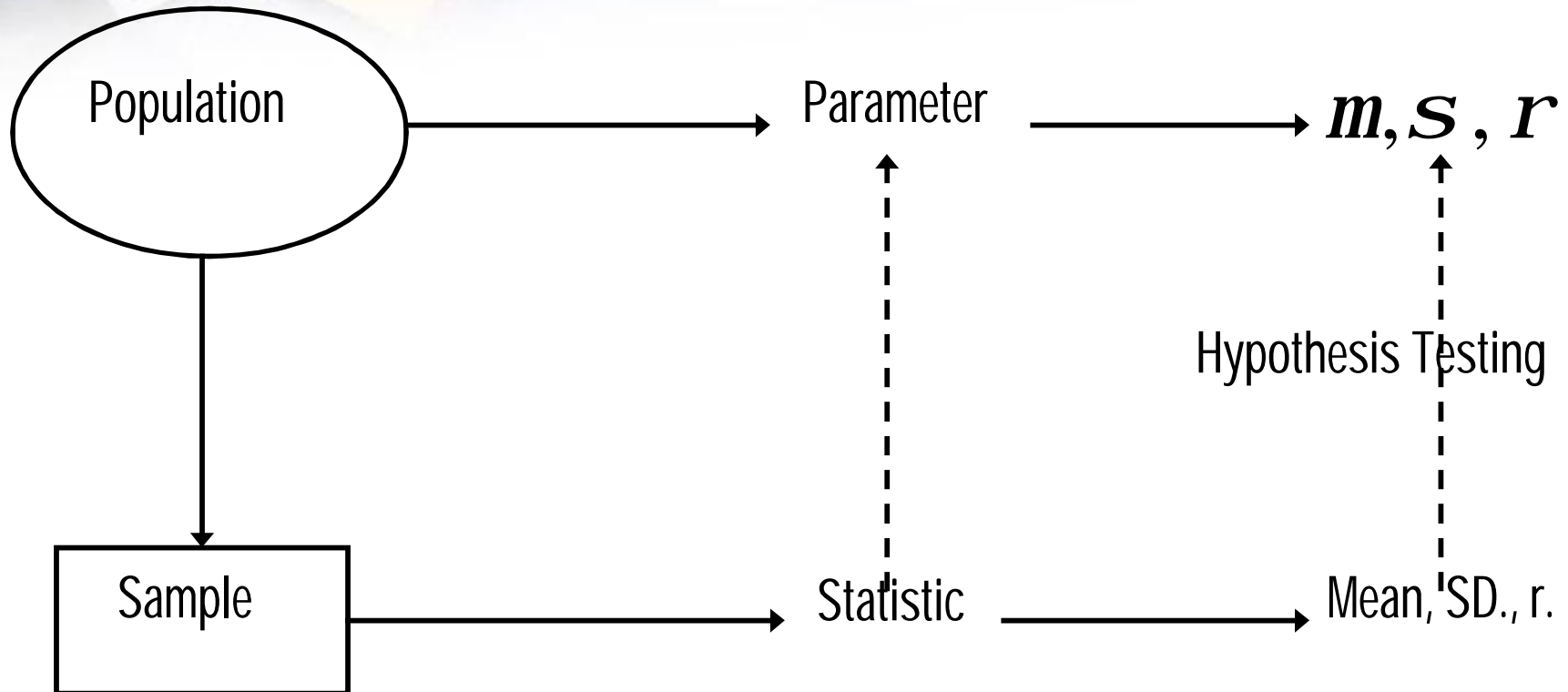
- ไม่มีการทดสอบสมมติฐาน
 - ศึกษาจากประชากรทั้งหมด
- มีการทดสอบสมมติฐาน
 - ศึกษาจากกลุ่มตัวอย่าง แล้วทดสอบสมมติฐาน เพื่ออ้างอิงกลับไปยังประชากร

ศึกษาจากประชากรทั้งหมด



- ใช้สถิติพรรณนา (Descriptive Statistics)

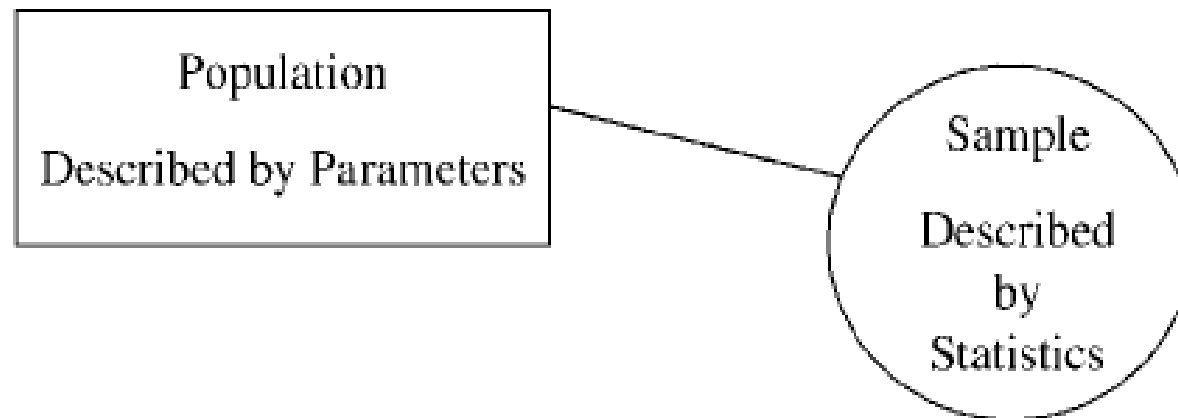
มีการทดสอบสมมติฐาน



- ใช้สถิติพรรณนา (Descriptive Statistics)
- ใช้สถิติอนุมาน (Inferential Statistics)

Parameter and Statistic

- สถิติที่ได้จากประชากรทั้งหมดเรียกว่า Parameter
- สถิติที่ได้จากกลุ่มตัวอย่างเรียกว่า Statistic

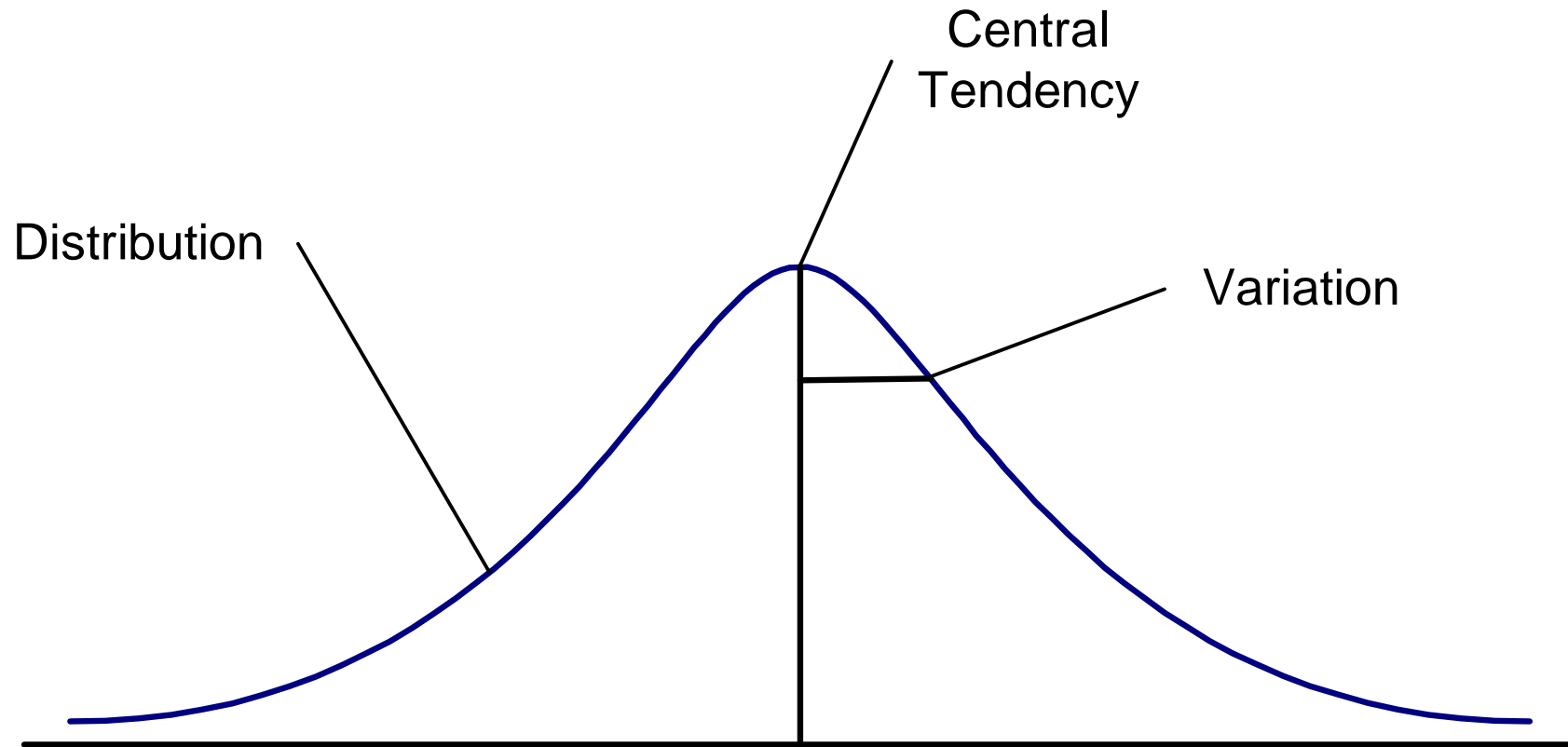




การใช้สถิติบรรยายข้อมูล

- สถิติที่บ่งบอกถึงค่ากลาง (Central Tendency)
- สถิติที่บ่งบอกถึงการกระจาย (Variation)
- สถิติที่บ่งบอกถึงรูปร่างหรือลักษณะของการแจกแจงของข้อมูล (Distribution)

การใช้สถิติบรรยายข้อมูล





สถิติที่บ่งบอกถึงค่ากลาง

Numbers

Nominal

Ordinal

Interval

Ratio

Permissible Measures

Mode

Mode, median

Mode, median, mean

Mode, median, mean



สถิติที่บ่งบอกถึงการกระจาย

- Ordinal Scale
 - Interquartile Range
- Interval and Ratio Scale
 - Range (Max. - Min.)
 - Variance
 - Standard Deviation

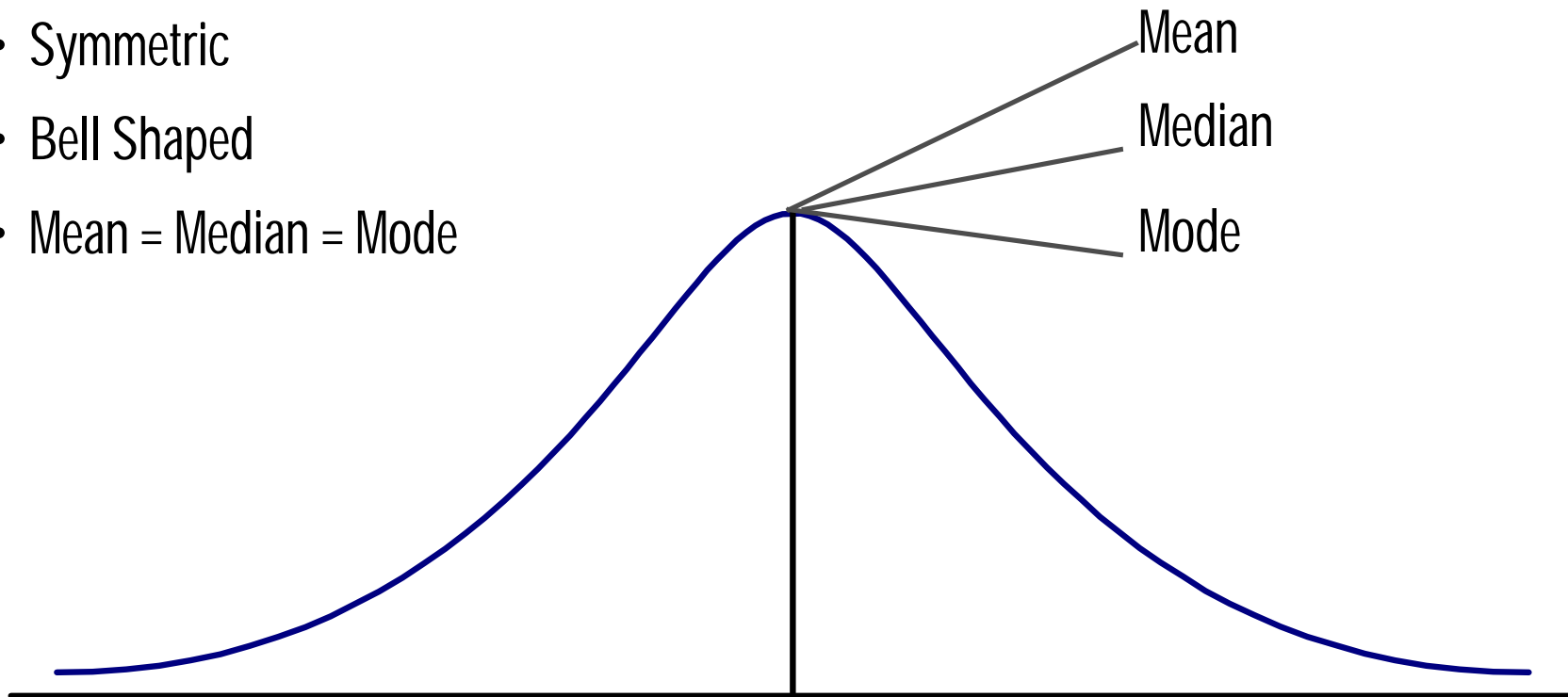


ลักษณะการกระจาย

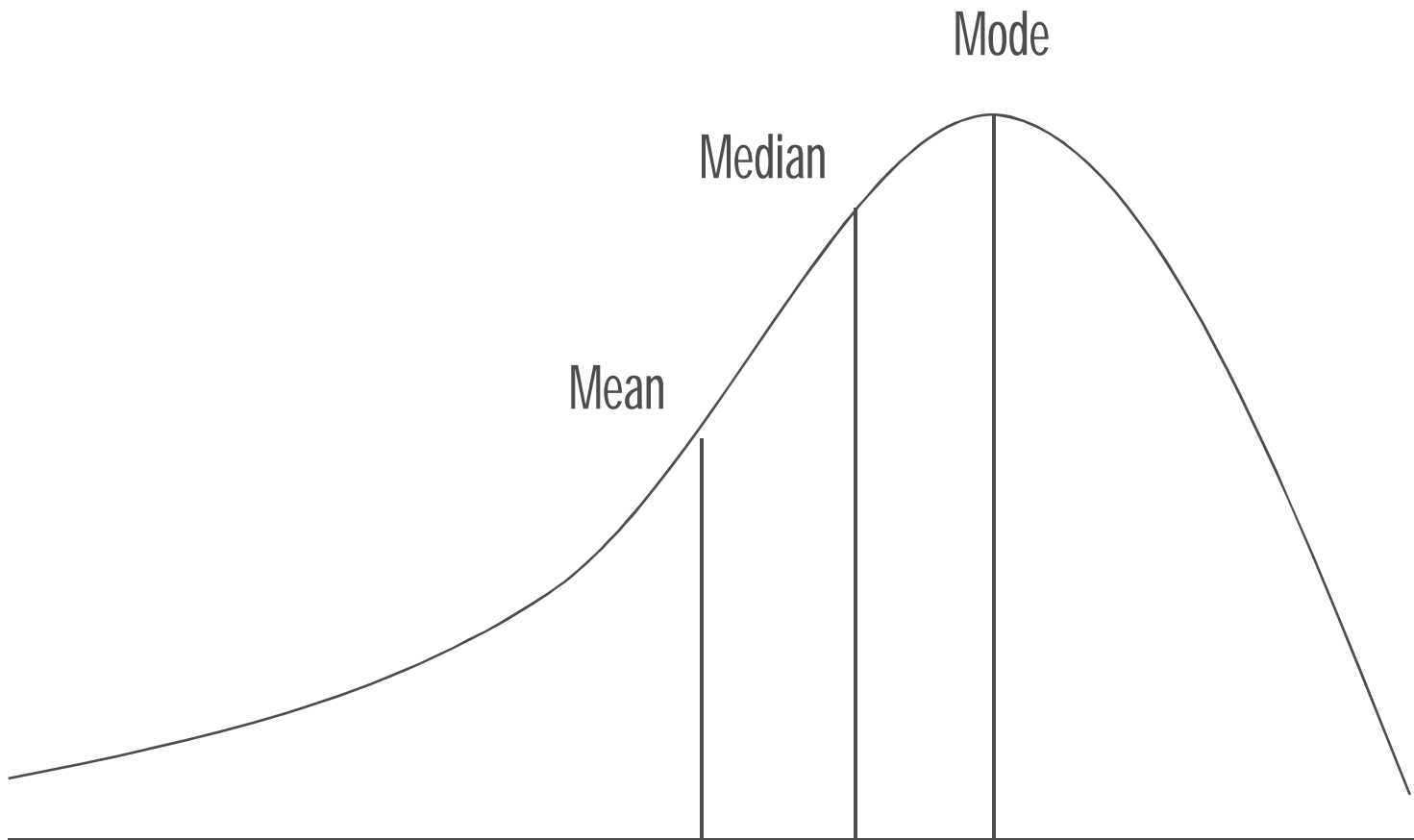
- ความเบ้ (Skewness)
- ความโด่ง (Kurtosis)

การกระจายแบบโค้งปกติ (Normal Distribution)

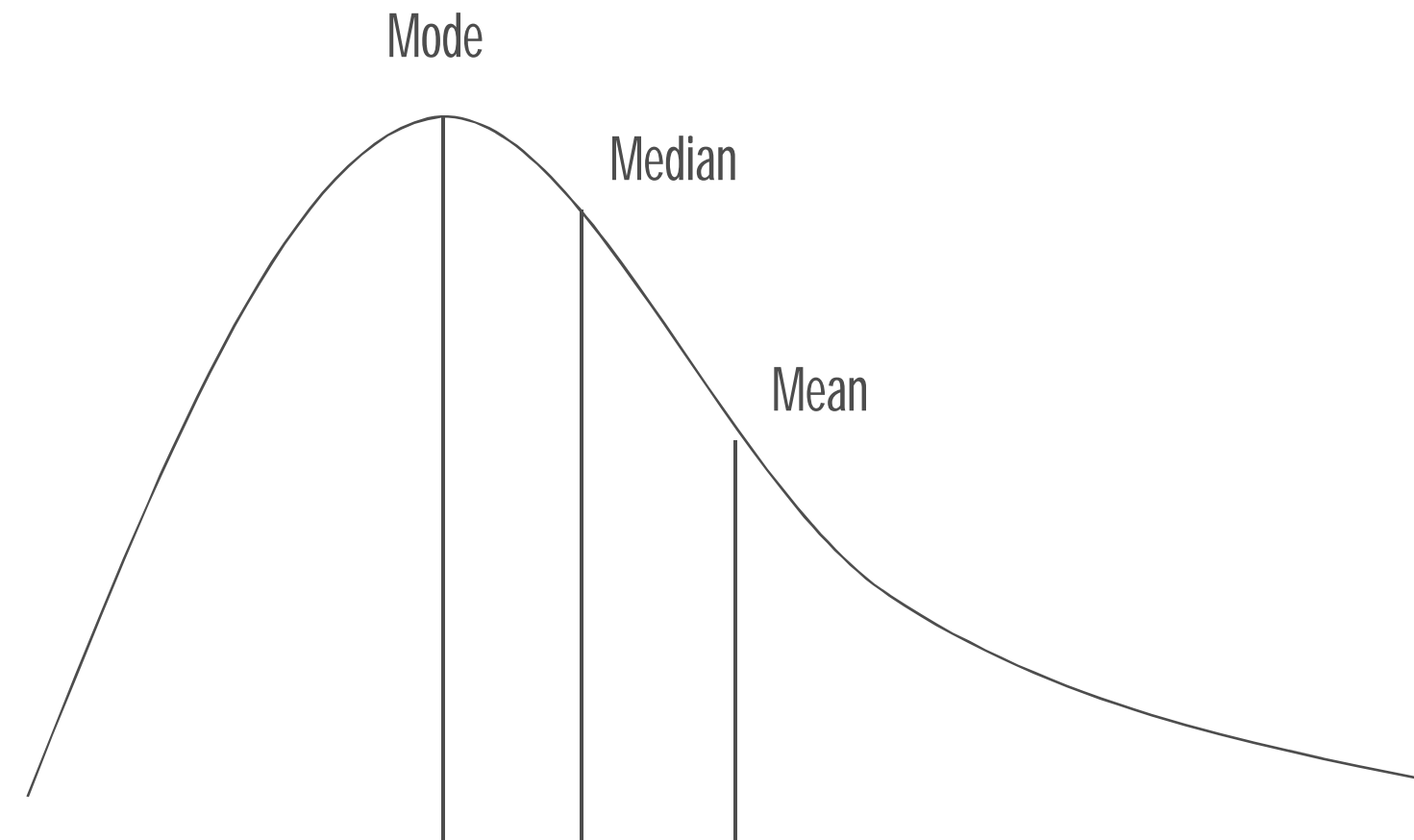
- Symmetric
- Bell Shaped
- Mean = Median = Mode



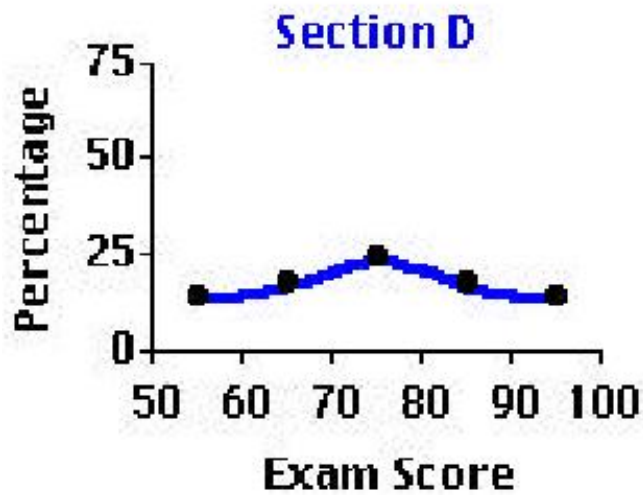
เบ้ซ้ายหรือเบ้ทางลบ (Negatively Skewed)



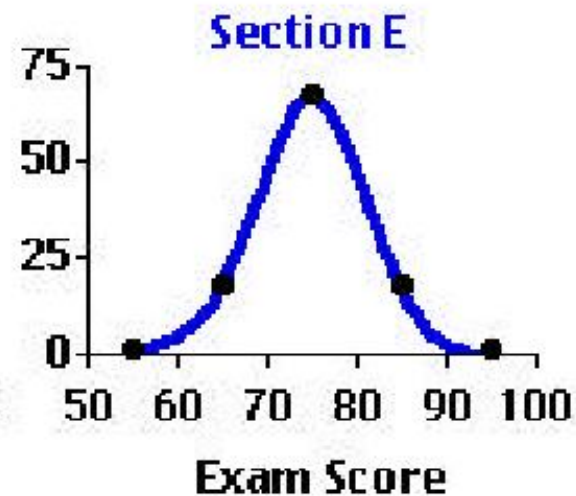
เบ้ขวาหรือเบ้ทางบวก (Positively Skewed)



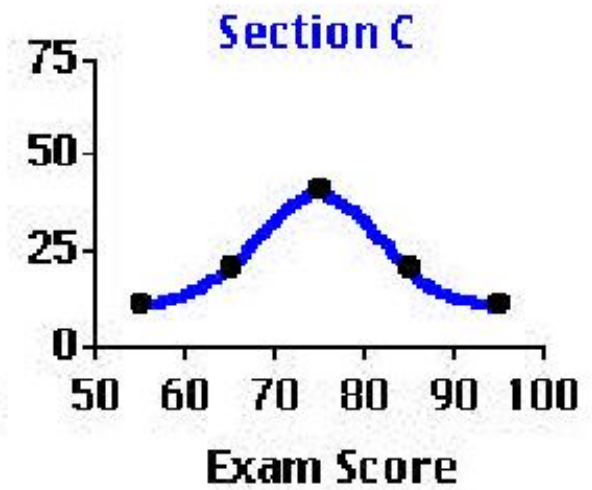
ความโด่ง (Kurtosis)



Platykurtic



Leptokurtic



Mesokurtic



What is random variable?

- ตัวแปรสุ่ม (Random Variable) เป็นตัวแปรที่โอกาสในการเกิดค่าของตัวแปรนั้นเป็นไปได้ในลักษณะของการสุ่ม
- $\text{Random Variable} = \text{Random Sample} + \text{Variable}$
- ตัวอย่างเช่น เหรียญ ค่าของตัวแปรคือ หัว และ ก้อย
- ตัวอย่างเช่น เพศ ค่าของตัวแปร คือ ชาย และ หญิง
- ตัวอย่างเช่น ส่วนสูง น้ำหนัก



Probability

- ความน่าจะเป็น (Probability)
- ตัวอย่างเช่น มีลูกแก้ว 100 ลูก เป็นลูกแก้วสีแดง 20 ลูก สีน้ำเงิน 80 ลูก
- เลือกลูกแก้วขึ้นมา 1 ลูกจากประชากร 100 ลูก จะเป็นสีอะไร?
- ตัวแปรคือสีของลูกแก้ว ค่าตัวแปรคือสีแดงและสีน้ำเงิน กลุ่มตัวอย่างคือลูกแก้ว 1 ลูก



Probability

- โอกาสที่จะเลือกได้ลูกแก้วสีแดง = $20/100 = .20$
- โอกาสที่จะเลือกได้ลูกแก้วสีน้ำเงิน = $80/100 = .80$
- ดังนั้นโอกาสที่ลูกแก้ว 1 ลูกที่เลือกจากประชากรจะเป็นสีน้ำเงินมากกว่าสีแดง
- ความน่าจะเป็นจึงเป็นโอกาสของการเกิดปรากฏการณ์หรือโอกาสที่จะเกิดค่าของตัวแปรของกลุ่มตัวอย่าง



Probability

- ถ้ามีขวดโหลอยู่ 2 ขวด
 - ขวดที่ 1 มีลูกแก้วสีแดง 20 ลูก สีน้ำเงิน 80 ลูก
 - ขวดที่ 2 มีลูกแก้วสีแดง 80 ลูก สีน้ำเงิน 20 ลูก
- ถ้าเลือกลูกแก้วมาหนึ่งกลุ่มตัวอย่าง จำนวน 10 ลูก เป็นลูกแก้วสีน้ำเงิน 3 ลูก สีแดง 7 ลูก ถ้ามว่า ลูกแก้วกลุ่มตัวอย่างนี้มาจากประชากรขวดโหลใด?
- ประชากรน่าจะเป็นขวดที่ 2



Probability

- แม้ว่าจะมีโอกาสที่ลูกแก้วกลุ่มตัวอย่าง 10 ลูกนี้จะมาจากประชากรขวดที่ 2 ก็ตาม
- ดังนั้นถึงแม้ว่าเราจะไม่รู้ว่าประชากรลูกแก้วในขวดไหนคืออะไร แต่เราสามารถใช้ความรู้จากกลุ่มตัวอย่างในการอนุมานกลับไปยังประชากรได้โดยอาศัยความน่าจะเป็น
- ความน่าจะเป็นจึงเป็นสะพานเชื่อมระหว่างกลุ่มตัวอย่างกับประชากร

ความน่าจะเป็นกับการอนุมานทางสถิติ

$H_0 : m_1 = m_2$ → ขวดโหลที่ 1 สีแดง 20 สีน้ำเงิน 80

$H_1 : m_1 \neq m_2$ → ขวดโหลที่ 2 สีแดง 80 สีน้ำเงิน 20

↓
กลุ่มตัวอย่าง

ผลต่างค่าเฉลี่ย = 10
($p=.03$)

↓
ประชากรน่าจะเป็น H_1

↓
กลุ่มตัวอย่าง

สีแดง 7 ลูก
สีน้ำเงิน 3 ลูก ($p=.03$)

↓
ประชากรน่าจะเป็น ขวดโหลที่ 2



ตรรกะการทดสอบสมมติฐาน

- จงทดสอบว่า ถ้าไฟเป็นตัวอักษรสระ ด้านหลังจะเป็นเลขคู่ โดยพลิกไฟเพื่อทดสอบจำนวนน้อยครั้งที่สุด

E

K

4

7



ตรรกะการทดสอบสมมติฐาน

- H_0 : อักษรสระ = เลขคู่
- H_1 : อักษรสระ \neq เลขคู่

E

K

4

7



ตรรกะการทดสอบสมมติฐาน

- Verification – การทดสอบโดยพิสูจน์หรือยืนยันว่าทฤษฎีนั้นถูก
- Karl Popper's Falsification – การทดสอบโดยหาหลักฐานหัก
ค้านว่าทฤษฎีนั้นผิด
- Logic of Disconfirming Null Hypothesis
- Null hypothesis can never be proved to be true, only
falsifying or disconfirming



Type I and Type II Error

		Actual situation	
		H_0 true	H_0 not true
Decision	reject H_0	Type 1 error	correct decision
	fail to reject H_0	correct decision	Type 2 error

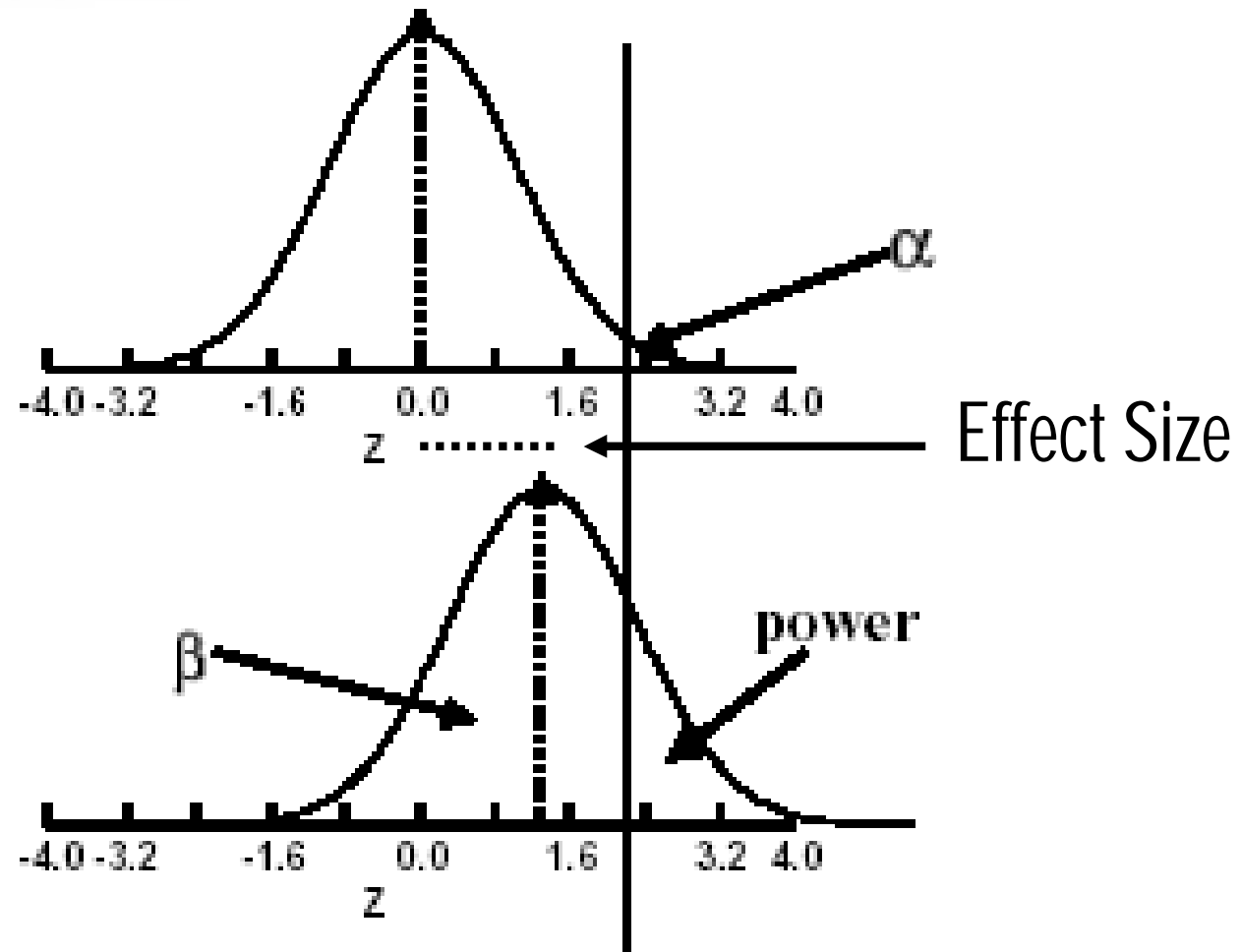
- Type I Error คือ การปฏิเสธ H_0 เมื่อ H_0 เป็นจริง
- Type II Error คือ การไม่ปฏิเสธ H_0 เมื่อ H_0 เป็นเท็จ



Type I and Type II Error

- Type I Error คือ False Alarm ใช้สัญลักษณ์ α (Alpha) แทนความคลาดเคลื่อนประเภทนี้
- Type II Error คือ Failed Alarm ใช้สัญลักษณ์ β (Beta) แทนความคลาดเคลื่อนประเภทนี้

Type I and Type II Error





Power

- อำนาจทดสอบทางสถิติ คือ ความน่าจะเป็นในการปฏิเสธ Null Hypothesis เมื่อเป็นเท็จ
- ขึ้นอยู่กับ
 - จำนวนกลุ่มตัวอย่าง
 - การกระจายของข้อมูล
 - ขนาดของอิทธิพลหรือความแตกต่าง (Effect Size)
 - ระดับนัยสำคัญ



Type I and Type II Error

- H_0 : นาย ก ไม่ได้ฆ่า นาย ข
- H_1 : นาย ก ฆ่า นาย ข
- Type I Error = จริงๆ แล้ว นาย ก ไม่ได้ฆ่า นาย ข แต่ศาลตัดสินว่า นาย ก เป็นคนฆ่า
- Type II Error = จริงๆ แล้ว นาย ก ฆ่า นาย ข แต่ศาลตัดสินยกฟ้อง



Type I and Type II Error

- H_0 : ยายี่ห้อ A ไม่สามารถลดไข้ได้ภายใน 1 ชม.
- H_1 : ยายี่ห้อ A สามารถลดไข้ได้ภายใน 1 ชม.
- Type I Error = จริงๆ แล้ว ยายี่ห้อ A ไม่สามารถลดไข้ได้ภายใน 1 ชม. แต่บริษัทสรุปว่า ยาลดไข้ได้
- Type II Error = จริงๆ แล้ว ยายี่ห้อ A สามารถลดไข้ได้ภายใน 1 ชม. แต่บริษัทสรุปว่า ยาลดไข้ไม่ได้



Type I and Type II Error

- H_0 : ยาที่เพิ่งคิดค้นใหม่ไม่สามารถฆ่าเชื้อ HIV ได้
- H_1 : ยาที่เพิ่งคิดค้นใหม่สามารถฆ่าเชื้อ HIV ได้
- Type I Error = จริง ๆ แล้ว ยาที่เพิ่งคิดค้นใหม่ไม่สามารถฆ่าเชื้อ HIV ได้ แต่นักวิจัยสรุปว่าสามารถฆ่าเชื้อได้
- Type II Error = จริง ๆ แล้ว ยาที่เพิ่งคิดค้นใหม่สามารถฆ่าเชื้อ HIV ได้ แต่นักวิจัยสรุปว่าไม่สามารถฆ่าเชื้อได้



Type I and Type II Error

- H_0 : ไฟฟ้าไม่ได้ไหม้อาคาร
- H_1 : ไฟฟ้าไหม้อาคาร
- Type I Error = จริงๆ แล้ว ไฟฟ้าไม่ได้ไหม้ แต่สัญญาณเตือนภัยไฟไหม้ดัง
- Type II Error = จริงๆ แล้ว ไฟฟ้าไหม้อาคาร แต่สัญญาณเตือนภัยไฟไหม้ไม่ดัง



Type I and Type II Error

- Type I และ Type II Error จะสวนทางกัน ถ้าความผิดพลาดอันใดอันหนึ่งมาก ความผิดพลาดอีกอันหนึ่งจะน้อย
- งานวิจัยส่วนใหญ่จะกำหนด Type I ไว้ เนื่องจากการกำหนด Type II จะเป็นเรื่องยุ่งยากซับซ้อนกว่า และพยายามควบคุม Type II โดยการศึกษากลุ่มตัวอย่างจำนวนมาก ๆ



การทดสอบสมมติฐานทางสถิติ

- Significance Testing
 - ทดสอบนัยสำคัญโดยใช้สถิติทดสอบสมมติฐาน
- Confidence Interval
 - ประมาณค่าช่วงความเชื่อมั่นของประชากร



Confidence Interval

- การประมาณค่าแบ่งออกเป็น

- Point Estimate
- Interval Estimate

$$\bar{X} \pm Z \times \frac{S}{\sqrt{n}}$$

- Confidence Interval คือ การประมาณค่าแบบช่วงของค่าประชากรว่ามีค่าอยู่ระหว่างเท่าไร
- Confidence Interval = Statistics + Sampling Error (Margin Error)



Confidence Interval

- ดูว่าช่วงของค่าสถิติกว้างมากน้อยเพียงใด ถ้าค่าประมาณยิ่งแคบแสดงว่ายิ่งมีความแม่นยำ (Precision)
- ดูช่วงของค่าประมาณว่าครอบคลุมค่าสถิติที่ต้องการทดสอบใน H_0 หรือไม่



Confidence Interval

$$\bar{X} \pm Z \times \frac{S}{\sqrt{n}}$$

$$10,200 \pm 1.96 \times \frac{2000}{\sqrt{400}}$$

$$10,200 \pm 196 = 10,004 - 10,396$$



Significance

- **Statistical Significance** มีนัยสำคัญทางสถิติ
- **Practical Significance** มีนัยสำคัญทางปฏิบัติ



Types of Statistics

- Difference Testing Statistics
 - เปรียบเทียบค่าเฉลี่ย ร้อยละ (Descriptive Statistics)
 - t-test, ANOVA, ANCOVA, Chi-Square
- Measures of Association Statistics
 - Pearson's r , Regression, Canonical Correlation
- Interdependence Statistics
 - Factor Analysis, Cluster Analysis, Multidimensional Scaling



Types of Statistics

- Parametric Statistics
 - t-test, ANOVA, Pearson's r , Regression
- Nonparametric Statistics
 - Mann-Whitney U, Wilcoxon, Kruskal Wallis
- Robust Statistics
 - Trimmed Mean, Winsorized Mean, Winsorized Variance, Resampling Statistics



สถิติสองตัวแปร (Bivariate Statistics)

- t-test
 - $Y = X$
 - โดยที่ Y เป็นตัวแปรเชิงปริมาณ และ X เป็นตัวแปรจัดกลุ่มแบ่งได้ 2 กลุ่ม
 - ถ้ากลุ่มเป็นอิสระกัน ใช้ Independent Sample
 - ถ้าละเมิด Assumptions ใช้ Mann-Whitney U
 - ถ้ากลุ่มไม่เป็นอิสระกัน ใช้ Dependent Sample
 - ถ้าละเมิด Assumptions ใช้ Wilcoxon Signed Rank



สถิติสองตัวแปร (Bivariate Statistics)

- ANOVA
 - $Y = X$
 - โดยที่ Y เป็นตัวแปรเชิงปริมาณ และ X เป็นตัวแปรจัดกลุ่มแบ่งได้มากกว่า 2 กลุ่ม
 - ถ้ากลุ่มเป็นอิสระกัน ใช้ ANOVA ตามปกติ
 - ถ้าละเมิด Assumptions ใช้ Kruskal Wallis
 - ถ้ากลุ่มไม่เป็นอิสระกัน ใช้ ANOVA for Repeated Measures
 - ถ้าละเมิด Assumptions ใช้ Friedman Test



สถิติสองตัวแปร (Bivariate Statistics)

- Pearson's r
 - $Y = X$
 - โดยที่ Y เป็นตัวแปรเชิงปริมาณ และ X เป็นตัวแปรเชิงปริมาณ
- Chi-square
 - $Y = X$
 - โดยที่ Y เป็นตัวแปรจัดกลุ่ม และ X เป็นตัวแปรจัดกลุ่ม



สถิติหลายตัวแปร (Multivariate Statistics)

- Multiple Regression
- Manova (Multivariate Analysis of Variance)
- Discriminant Analysis
- Canonical Correlation
- Factor Analysis



สถิติหลายตัวแปร (Multivariate Statistics)

- Multiple Regression

- $Y = X_1 + X_2 + X_3 + \dots + X_n$

- โดยที่ Y เป็นตัวแปรเชิงปริมาณ และ X เป็นตัวแปรเชิงปริมาณ (อาจมีตัวแปรจัดกลุ่มได้ แต่ต้องแปลงเป็นตัวแปร Dummy)

- Manova (Multivariate Analysis of Variance)

- $Y_1 + Y_2 + Y_3 + \dots + Y_n = X_1 + X_2 + X_3 + \dots + X_n$

- โดยที่ Y เป็นตัวแปรเชิงปริมาณ และ X เป็นตัวแปรจัดกลุ่ม



สถิติหลายตัวแปร (Multivariate Statistics)

- Discriminant Analysis

- $Y = X_1 + X_2 + X_3 + \dots + X_n$

- โดยที่ Y เป็นตัวแปรจัดกลุ่ม และ X เป็นตัวแปรเชิงปริมาณ
(อาจมีตัวแปรจัดกลุ่มได้ แต่ต้องแปลงเป็นตัวแปร Dummy)

- Canonical Correlation

- $Y_1 + Y_2 + Y_3 + \dots + Y_n = X_1 + X_2 + X_3 + \dots + X_n$

- โดยที่ Y เป็นตัวแปรเชิงปริมาณ และ X เป็นตัวแปรเชิงปริมาณ



สถิติหลายตัวแปร (Multivariate Statistics)

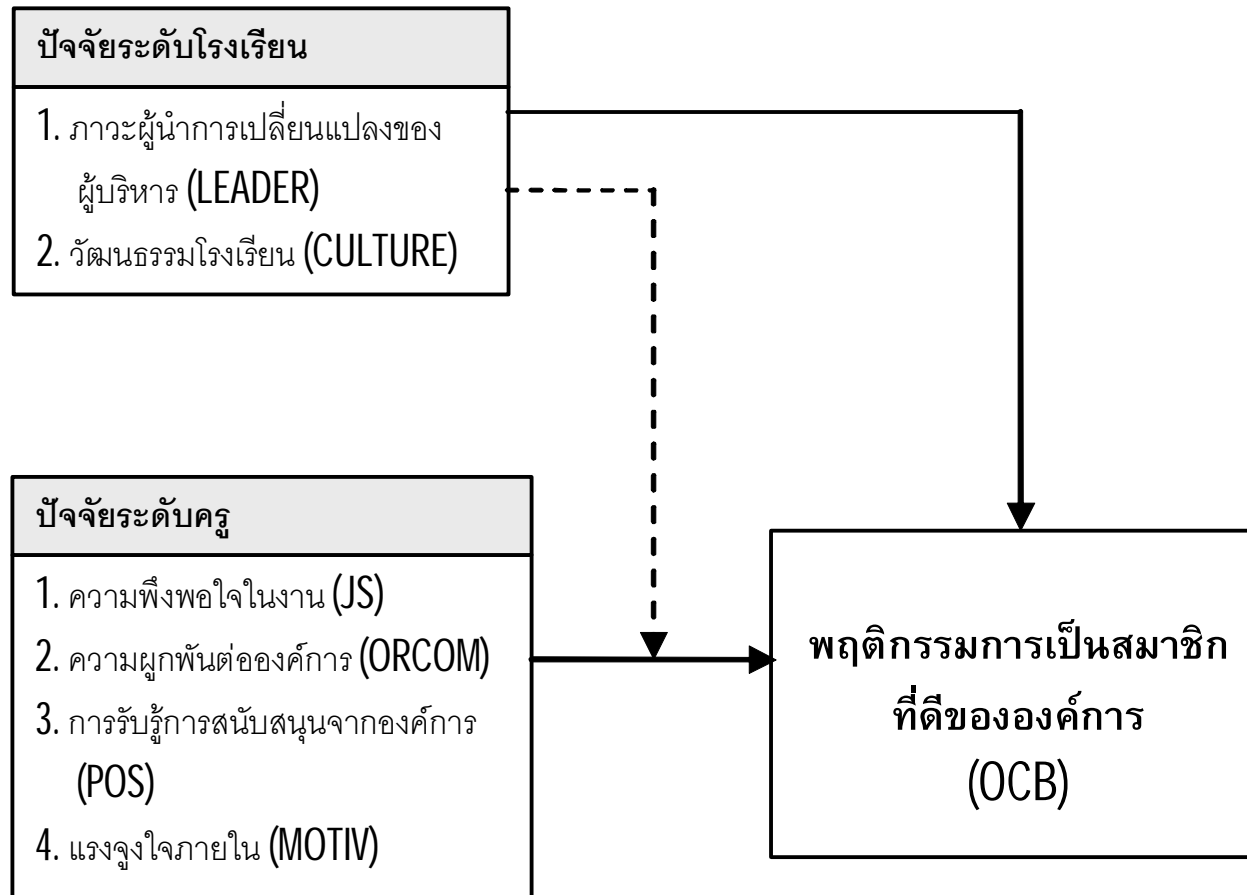
- Factor Analysis

- $X_1 + X_2 + X_3 + \dots + X_n$
- โดยที่ X เป็นตัวแปรเชิงปริมาณ

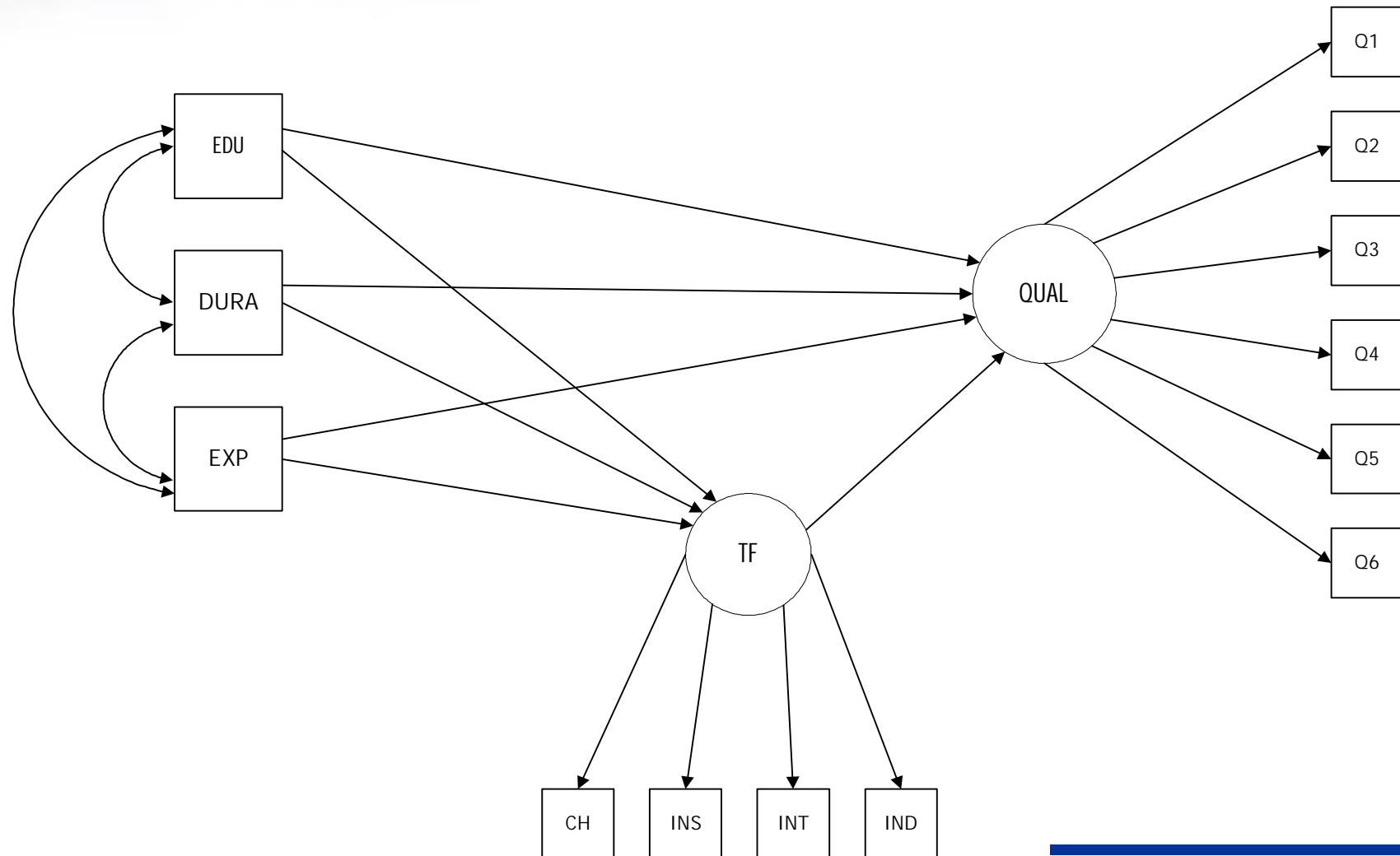
- Multilevel Analysis

- $Y = X_1 + X_2 + X_3 + \dots + X_n$
- โดยที่ Y เป็นตัวแปรเชิงปริมาณ และ X เป็นตัวแปรเชิงปริมาณ (อาจมีตัวแปรจัดกลุ่มได้ แต่ต้องแปลงเป็นตัวแปร Dummy) แต่มี X บางตัวที่อยู่คนละระดับการวิเคราะห์

Multilevel Analysis



Structural Equation Modeling





จบการบรรยาย

ดร.นำชัย ศุภฤกษ์ชัยสกุล