

สถิติสำหรับงานวิจัย



Statistics for Research

การวิเคราะห์คุณภาพเครื่องมือ

มีจุดประสงค์ที่สำคัญ 2 ประการ คือ

- การวิเคราะห์เพื่อหาคุณภาพรายข้อ
- การวิเคราะห์เพื่อหาคุณภาพทั้งฉบับ

สถิติสำหรับงานวิจัย



Statistics for Research

การวิเคราะห์คุณภาพเครื่องมือ
สามารถแบ่งได้ **2** แนวทาง คือ

- การวิเคราะห์ตามทฤษฎีทดสอบดั้งเดิม
(Classical test theory : CTT)
- การวิเคราะห์ตามทฤษฎีตอบสนองข้อสอบ
(Item response theory : IRT)



การวิเคราะห์ตามทฤษฎีทดสอบดั้งเดิม (Classical test theory : CTT)

- การวิเคราะห์คุณภาพรายข้อ
 - ความมีอำนาจจำแนก(discrimination) **
 - ความเชื่อมั่นรายข้อ(item-reliability)
 - ความเที่ยงตรงรายข้อ((item-validity)
 - ค่าความยากรายข้อ(item-difficulty) **
 - ตัวลวง(distractor)

สถิติสำหรับงานวิจัย



Statistics for Research

การวิเคราะห์ตามทฤษฎีทดสอบดั้งเดิม (Classical test theory : CTT)

- การวิเคราะห์คุณภาพทั้งฉบับ
 - ความเชื่อมั่น (reliability)
 - ความเที่ยงตรง (validity)



การวิเคราะห์คุณภาพรายข้อ ความมีอำนาจจำแนก (Discrimination)

รายข้อนั้นต้องสามารถจำแนกสิ่งที่มีลักษณะต่างกันอย่างชัดเจน

- ความแตกต่าง ของ สัดส่วนจำนวนคนที่ตอบถูกของกลุ่มสูงกับต่ำ
- ความสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวม
 - สหสัมพันธ์แบบพอยท์ ไบซีเรียล (point biserial correlation)
 - สหสัมพันธ์แบบเพียร์สัน (pearson correlation coefficient)
- การทดสอบค่าที (t - test)



Statistics for Research

การวิเคราะห์คุณภาพรายข้อ

ความมีอำนาจจำแนก (Discrimination) (ต่อ)

ความแตกต่างของสัดส่วนจำนวนคนที่ตอบถูกของกลุ่มสูงกับต่ำ
วิธีการ

1. นำเครื่องมือไปทดสอบ \longrightarrow เรียงคะแนนจากมาก \longrightarrow น้อย
2. แบ่งกลุ่มผู้สอบออกเป็น 2 กลุ่มเท่า ๆ กัน คือ กลุ่มที่ได้คะแนนสูง (T)
และกลุ่มที่ได้คะแนนต่ำ (L)
3. นับจำนวนคนที่ตอบถูกในกลุ่มสูง และ ต่ำ



การวิเคราะห์คุณภาพรายข้อ ความมีอำนาจจำแนก (Discrimination) (ต่อ)

วิธีการ

4. คำนวณหาค่าดัชนีอำนาจจำแนกรายข้อ จากสูตร $D = P_T - P_L$

เมื่อ D แทนดัชนีอำนาจจำแนกรายข้อของแบบทดสอบ

P_T แทนจำนวนคนที่ตอบถูกในกลุ่มสูง/จำนวนคนในกลุ่มสูง

P_L แทนจำนวนคนที่ตอบถูกในกลุ่มต่ำ/จำนวนคนในกลุ่มต่ำ



การวิเคราะห์คุณภาพรายข้อ

ความมีอำนาจจำแนก (Discrimination) (ต่อ)

การแปลความหมาย

- ดัชนีอำนาจจำแนกมีค่าอยู่ระหว่าง -1 กับ $+1$ นั่นคือ $-1 < D < 1$
- ถ้า D มีค่าบวก และเข้าใกล้ 1 แสดงว่ารายข้อนั้น **มีค่าอำนาจจำแนกดี**
- ถ้า D มีค่า 0 และเข้าใกล้ -1 แสดงว่ารายข้อนั้น **มีค่าอำนาจจำแนกไม่ดี**

หมายเหตุถ้า ค่า D มีค่ามากกว่า 0.6 แสดงว่ารายข้อนั้นสามารถจำแนกได้ดี

$0.3 - 0.6$

“

พอสมควร

ต่ำกว่า 0.3

“

น้อย



การวิเคราะห์คุณภาพรายข้อ

ค่าความยากรายข้อ (Item - Difficulty)

สัดส่วน หรือร้อยละของคนที่ตอบถูกในข้อนั้น

วิธีการ หาค่าความยากรายข้อ จาก
$$p = \frac{\text{จำนวนคนที่ตอบรายข้อนั้นถูก}}{\text{จำนวนคนทั้งหมด}}$$

การแปลความหมาย

- $0 < p < 1$
- ถ้า ค่า p สูง แสดงว่ารายข้อนั้นเป็นข้อที่ค่อนข้างง่าย
- ถ้า ค่า p ต่ำ แสดงว่ารายข้อนั้นเป็นข้อที่ค่อนข้างยาก

การวิเคราะห์คุณภาพรายข้อ ตัวลวง (Distractor)

สถิติสำหรับงานวิจัย



Statistics for Research

ตัวเลือกอื่น ๆ ที่ไม่ใช่ตัวเลือกที่ถูกต้องในแบบทดสอบ

วิธีการ หาค่า $\text{อำนาจตัวลวงที่คาดหวัง} = \frac{\text{จำนวนคนที่ ตอบผิดในข้อ นั้น}}{\text{จำนวนตัวลวง}}$

การแปลความหมาย **ตัวลวงที่ดี** จะต้องมีจำนวนคนเลือกตอบในแต่ละตัวลวงเท่า ๆ กัน นั่นคือ ใกล้เคียงกับอำนาจตัวลวงที่คาดหวัง ดังนั้น ถ้าตัวลวงใดมีคนเลือกตอบแตกต่างจาก อำนาจตัวลวงที่คาดหวังมาก ๆ แสดงว่าตัวลวงนั้นเป็น **ตัวลวงที่ไม่ดี** ต้องปรับปรุง

การวิเคราะห์คุณภาพทั้งฉบับ ความเชื่อมั่น (Reliability)

สถิติสำหรับงานวิจัย



Statistics for Research

สามารถให้ผลการวัดที่คงเส้นคงวา คงที่แน่นอน ในการวัดแต่ละครั้ง

- วิธีการสอบซ้ำ (test retest method)
- วิธีการใช้เครื่องมือวัดที่คู่ขนานกัน (parallel form method)
- วิธีการหาความเชื่อมั่นแบบความคงที่ภายใน (internal consistency) **



การวิเคราะห์คุณภาพทั้งฉบับ

วิธีการหาความเชื่อมั่นแบบความคงที่ภายใน (internal consistency)

- วิธีการแบ่งครึ่ง (split - half method)
- วิธีการของคูเดอร์และริชาร์ดสัน (Kuder and Richardson)
- วิธีการหาสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's coefficient alpha)
- วิธีการวิเคราะห์ความแปรปรวนของฮอยท์ (Hoyt's analysis of variance)



การวิเคราะห์คุณภาพทั้งฉบับ

วิธีการของคูเดอร์และริชาร์ดสัน (Kuder and Richardson)

ใช้กับกรณีที่เครื่องมือมีระบบการให้คะแนนเป็น 0 กับ 1

- สูตร KR 20 โดยหาค่าความเชื่อมั่นจาก $r_{tt} = \frac{k}{k-1} \left[1 - \frac{\sum pq}{s_t^2} \right]$
- เมื่อ k แทน จำนวนข้อของเครื่องมือ
- p แทน สัดส่วนของคนที่ทำถูกในแต่ละข้อ
- q แทน สัดส่วนของคนที่ทำผิดในแต่ละข้อ $1 - p$
- s_t^2 แทน ความแปรปรวนของคะแนนสอบของเครื่องมือทั้งฉบับ



การวิเคราะห์คุณภาพทั้งฉบับ

วิธีการของคูเดอร์และริชาร์ดสัน (Kuder and Richardson)

• สูตร KR 21 โดยหาค่าความเชื่อมั่นจาก $r_{tt} = \frac{k}{k-1} \left[1 - \frac{\bar{x}(k-\bar{x})}{ks_t^2} \right]$

เมื่อ k แทน จำนวนข้อของเครื่องมือ

\bar{x} แทน คะแนนเฉลี่ยของเครื่องมือ

s_t^2 แทน ความแปรปรวนของคะแนนสอบของเครื่องมือทั้งฉบับ



การวิเคราะห์คุณภาพทั้งฉบับ

วิธีการสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's coefficient alpha)

ใช้กับกรณีที่เครื่องมือมีระบบการให้คะแนนเป็น 0 กับ 1 หรือไม่ใช้

• โดยหาสัมประสิทธิ์ความเชื่อมั่นจาก $\alpha(\text{Alpha}) = \frac{k}{k-1} \left[1 - \frac{\sum s_i^2}{s_t^2} \right]$

เมื่อ k แทน จำนวนข้อของเครื่องมือ

s_i^2 แทน ความแปรปรวนของคะแนนสอบในแต่ละข้อ

s_t^2 แทน ความแปรปรวนของคะแนนสอบของเครื่องมือทั้งฉบับ



สถิติสำหรับงานวิจัย

Statistics for Research

การแปลความหมาย

- ค่าความเชื่อมั่นจะมีค่าระหว่าง **0** กับ **1**
- เครื่องมือที่ดีจะมีค่าความเชื่อมั่นสูง หรือใกล้เคียงกับ **1**

การวิเคราะห์คุณภาพทั้งฉบับ

ความเที่ยงตรง (Validity)

สถิติสำหรับงานวิจัย



Statistics for Research

มีคุณสมบัติวัดในสิ่งที่ต้องการวัดได้

วัดได้ตรงวัตถุประสงค์ เนื้อหาที่ต้องการวัด

ความเที่ยงตรง แบ่งออกเป็น **3** ประเภท คือ

- ความเที่ยงตรงตามเนื้อหา (content validity)
- ความเที่ยงตรงตามโครงสร้าง (construct validity)
- ความเที่ยงตรงตามเกณฑ์ที่เกี่ยวข้อง (criterion-related validity)



การวิเคราะห์คุณภาพทั้งฉบับ ความเที่ยงตรงตามเนื้อหา (Content validity)

เครื่องมือที่วัดได้ตรงเนื้อหาสาระที่ต้องการวัด

และวัดได้ตรงจุดที่ต้องการวัดมากที่สุด

มีวิธีการดังนี้

1. พิจารณาความสอดคล้อง หรือเป็นตัวแทนของข้อความ และได้
ตัดส่วนของจำนวนข้อความกับ นำหนัก เนื้อหารายวิชา
2. หาคความเที่ยงตรงตามเนื้อหา โดยผู้เชี่ยวชาญ

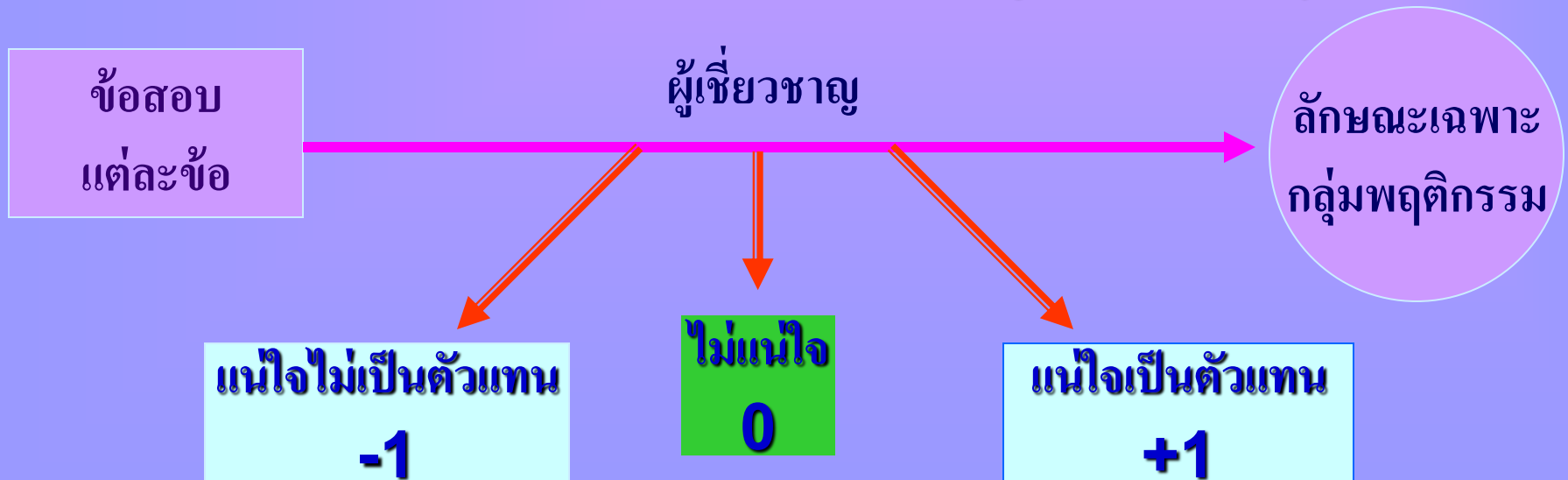


การวิเคราะห์คุณภาพทั้งฉบับ

ความเที่ยงตรงตามเนื้อหา (Content validity)

ความเที่ยงตรงตามเนื้อหา โดยผู้เชี่ยวชาญ

- ความเที่ยงตรงที่มีความเป็นอัตนัย (subjective validity)
- ความเที่ยงตรงที่มีความเป็นปรนัย (objective validity)





การวิเคราะห์คุณภาพทั้งฉบับ

ความเที่ยงตรงตามเนื้อหา (Content validity)

คำนวณหาดัชนีสอดคล้องระหว่างข้อคำถามกับลักษณะพฤติกรรม (IC) จาก

$$IC = \frac{\sum R}{N}$$

ผลรวมของคะแนนความคิดเห็นในแต่ละข้อ

จำนวนผู้เชี่ยวชาญ

การแปลผล

$$IC \geq 0.5$$

ข้อคำถามนั้นเป็นตัวแทนของลักษณะเฉพาะกลุ่มพฤติกรรมนั้น

$$IC < 0.5$$

ข้อคำถามนั้นควรตัดทิ้งไป หรือปรับปรุงใหม่

ตัวอย่าง จากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ซึ่งประกอบด้วย 5 คำถาม ๆ ละ 5 ตัวเลือก จำนวน 7 คน

ข้อมูลดิบ

ข้อมูลที่ใช้คำนวณ

คนที่	ข้อที่				
	1	2	3	4	5
1	2	3	1	5	1
2	2	3	2	5	4
3	2	5	4	4	5
4	2	3	1	4	3
5	2	3	1	4	3
6	3	4	3	4	3
7	2	3	1	4	3

คนที่	ข้อที่				
	1	2	3	4	5
1	1	1	1	0	0
2	1	1	0	0	0
3	1	0	0	1	0
4	1	1	1	1	1
5	1	1	1	1	1
6	0	0	0	1	1
7	1	1	1	1	1

ข้อมูลที่ใช้คำนวณ

คนที่	ข้อที่				
	1	2	3	4	5
1	1	1	1	0	0
2	1	1	0	0	0
3	1	0	0	1	0
4	1	1	1	1	1
5	1	1	1	1	1
6	0	0	0	1	1
7	1	1	1	1	1

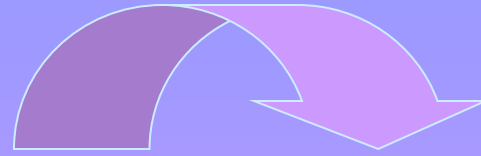
รวม

3
2
2
5
5
2
5

รวม

6 5 4 5 4

จำนวนคนที่ตอบถูกในแต่ละข้อ



- เรียงลำดับจากมากไปน้อย
- แบ่งกลุ่มสูง ต่ำได้
- หาอำนาจจำแนกได้
- \bar{x} s_t^2 s_i^2
- KR 21 , Alpha

ค่าความยากรายข้อ

- p และ q
- KR 20

$$\text{อำนาจตัวลงที่คาดหวัง} = \frac{\text{จำนวนคนที่ตอบผิดในข้อนั้น}}{\text{จำนวนตัวลง}}$$

- ก (9)
- ข (17)
- ค (19)
- ง (37)
- จ (18)

คนที่	ข้อที่				
	1	2	3	4	5
1	2	3	1	5	1
2	2	3	2	5	4
3	2	5	4	4	5
4	2	3	1	4	3
5	2	3	1	4	3
6	3	4	3	4	3
...
100	2	3	1	4	3

ถ้าข้อ ง ถูก
 จะมีคนตอบข้อ 3 ผิด 63 คน
 อำนาจตัวลงที่คาดหวัง = $63/4 = 15.75$
 จะเห็นว่าตัวลงที่ต้องปรับปรุงใหม่ คือ
 ข้อ ก

คนที่	ข้อที่				
	1	2	3	4	5
1	2	3	1	5	1
2	2	3	2	5	4
3	2	5	4	4	5
4	2	3	1	4	3
5	2	3	1	4	3
6	3	4	3	4	3
...
100	2	3	1	4	3

- ก (21)
- ข (48)
- ค (4)
- ง (4)
- จ (6)

ถ้าข้อ ง ถูก
 จะมีคนตอบข้อ 4 ผิด 96 คน
 จำนวนตัวลวงที่คาดหวัง = $96/4 = 24$
 จะเห็นว่าตัวลวงที่ต้องปรับปรุงใหม่ คือ
 ข้อ ข ค และ จ