

การวัดแบบก่อเป็นรูป และอรรถประโยชน์ของแบบทดสอบ

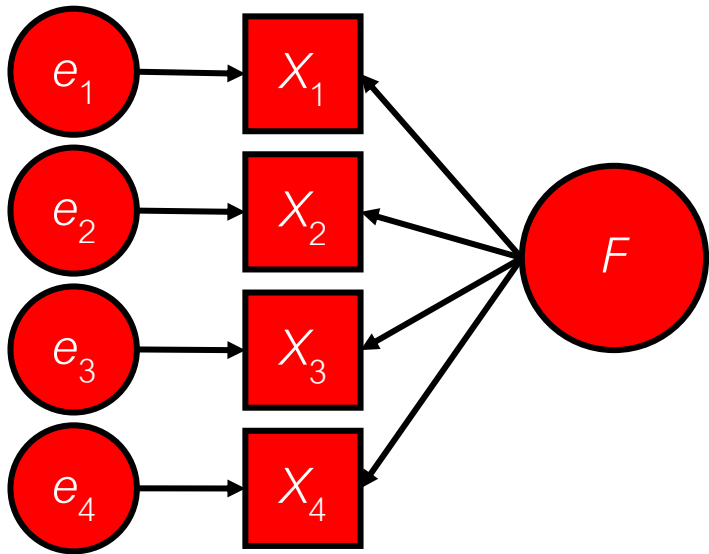
การประเมินลักษณะมนุษย์
สันทัด พรประเสริฐมานิต

โครงร่างการนำเสนอ

- การวัดแบบก่อเป็นรูป
- อรรถประโยชน์
- การสร้างจุดตัด

การวัดแบบก่อเป็นรูป (Formative Measurement)

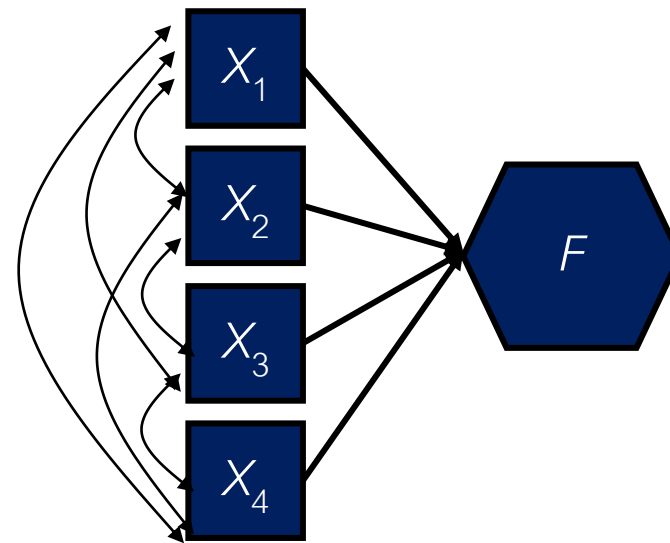
- ทำไมมาตรวัดการตอบสนองต่อสถานการณ์ (Situational judgment test), ศูนย์ประเมิน (Assessment center), และการทดลองงาน (Work sample) ถึงมีความสอดคล้องภายในระหว่างข้อคำถามต่ำ แต่สามารถทำนายผลการปฏิบัติงานได้ดี
- ทำไมเรายังคงใช้มาตรวัดที่ความ “เที่ยง” ต่ำอยู่ แล้วแน่ใจเหลือ ว่าคะแนนที่ได้จากผู้ถูกประเมิน จะสะท้อนคะแนนที่แท้จริง (True score) ของเขา
- มาตรวัดเหล่านี้สะท้อน “ภาวะสันนิษฐาน” อะไรกันแน่? หากข้อคำถามไม่สอดคล้องกัน



การวัดแบบสะท้อน (Reflective Measurement Model)

โมเดลองค์ประกอบร่วม (Common Factor Model)

ข้อคำถาม หรือตัวบ่งชี้แต่ละตัว จะสอดคล้องกัน เป็น “ผล” ของสาเหตุเดียวกัน แม้ว่าบางครั้งข้อคำถามจะจับเป็นกลุ่มย่อย ที่เรียกว่าองค์ประกอบ ข้อคำถามในกลุ่มย่อยก็จะเป็น “ผล” ขององค์ประกอบย่อยเดียวกัน



การวัดแบบก่อเป็นรูป (Formative Measurement Model)

โมเดลส่วนประกอบ (Composite Model)

ข้อคำถาม หรือตัวบ่งชี้แต่ละตัว จะเป็นอิสระจากกัน การนำมารวมกัน เหมือนเป็นการสร้างสมการทำนายเชิงเส้น (Linear Regression) ที่ค่าทำนายได้ (Predicted Scores) นี้ จะเกิดจากการหาความสัมพันธ์กับตัวแปรอื่นให้มากที่สุด (ทำนายให้ดีที่สุด) ตัวบ่งชี้ไม่จำเป็นต้องสัมพันธ์กันสูง

การวัดแบบก่อเป็นรูป (Formative Measurement)

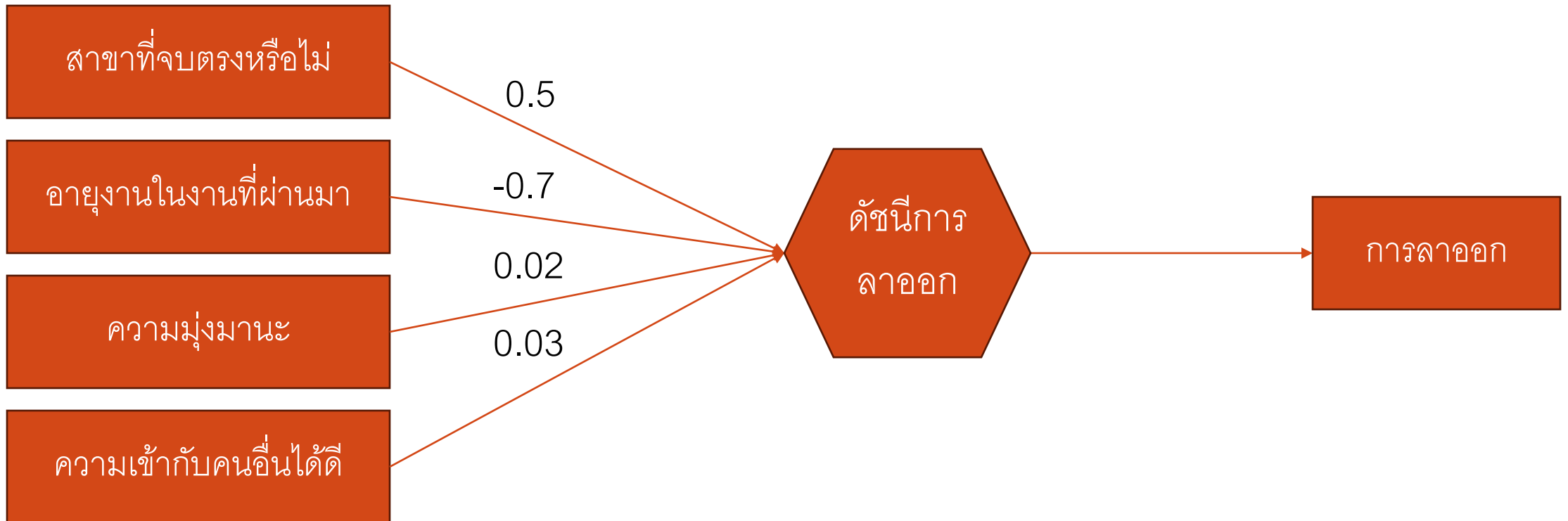
- มาตรการวัดใหม่ๆ ที่สาขาจิตวิทยามักใช้ คือ โมเดลการวัดแบบสะท้อน
- สมมติฐานของโมเดลนี้ คือ มีสิ่งใดสิ่งหนึ่ง ที่เราไม่สามารถวัดได้โดยตรง ที่เรียกว่าภาวะสันนิษฐาน (Construct) ข้อคำถาม สิ่งเร้า ที่เราใช้สร้างคะแนนเป็นตัวสะท้อนภาวะสันนิษฐานออกมา
 - เช่น ความเครียด อาจสะท้อนจาก ความคิดหมกมุ่น อาการนอนไม่หลับ ไม่มีสมาธิ ปวดหัว ฯลฯ อาการเหล่านี้ล้วนมาจากสาเหตุเดียวกัน คือ ความเครียด
 - เราไม่สามารถวัดความเครียดได้โดยตรง แต่ต้องวัดผ่านอาการต่างๆ แล้วนำอาการไปอ้างอิงถึงระดับความเครียด
- โมเดลการวัดแบบสะท้อนจะมุ่งเน้นหาค่าของภาวะสันนิษฐานให้ถูกต้องที่สุด
 - ข้อคำถาม หรือตัวบ่งชี้แต่ละตัว จะสอดคล้องกัน เป็น “ผล” ของสาเหตุเดียวกัน แม้ว่าบางครั้งข้อคำถามจะจับเป็นกลุ่มย่อย ที่เรียกว่าองค์ประกอบ ข้อคำถามในกลุ่มย่อยก็จะเป็น “ผล” ขององค์ประกอบย่อยเดียวกัน
 - โมเดลการวัดแบบสะท้อน หาคำตรงของคะแนนที่ได้ จากการหาหลักฐานมาสนับสนุนว่าคะแนนที่ได้ สะท้อนภาวะสันนิษฐานนั้นจริงๆ

การวัดแบบก่อเป็นรูป (Formative Measurement)

- โมเดลการวัดแบบสะท้อน ภาวะสันนิษฐานเป็นสาเหตุ และข้อคำถามหรือตัวบ่งชี้เป็นผล
- โมเดลการวัดแบบก่อเป็นรูป อยู่ในทิศทางตรงข้าม ข้อคำถามหรือตัวบ่งชี้เป็นสาเหตุ ที่ทำให้เกิดผลเป็น “ดัชนี” (จะเรียกว่าภาวะสันนิษฐานก็พอได้ เพราะผู้วิจัยสันนิษฐานว่ามันมีอยู่)
- เช่น แนวโน้มการลาออก อาจมีตัวบ่งชี้เป็น (ก) สาขาที่เรียนตรงกับตำแหน่งงานหรือไม่ (ข) อายุงานในงานที่ผ่านมา (ค) ความมุ่งมั่น และ (ง) ความเข้ากับคนอื่นได้ดี คะแนนเหล่านี้จะถูกนำมารวมกัน ด้วยการถ่วงน้ำหนักที่แตกต่างกัน เช่น

$$\text{Quitting Index} = 0.5 * \text{Area} - 0.7 * \text{Tenure} + 0.02 * \text{Grit} + 0.03 * \text{Agreeableness}$$

การวัดแบบก่อเป็นรูป (Formative Measurement)



การวัดแบบก่อเป็นรูป (Formative Measurement)

- การวัดแบบก่อเป็นรูป “คล้าย” กับการสร้างสมการทำนายในการวิเคราะห์ถดถอย (Regression Analysis) ที่นักวิจัยเลือกตัวแปรอิสระมาทำนายตัวแปรเป้าหมาย
- การวัดแบบก่อเป็นรูป สามารถแบ่งออกเป็น 2 รูปแบบ คือ
 - การวัดแบบก่อเป็นรูปเชิงสาเหตุ (Causal Formative Measurement) ตัวบ่งชี้ทั้งหมดมาจากทฤษฎี แนวคิด หลักการพื้นฐาน ว่าทำไมตัวแปรทั้งหมดถึงได้เป็นสาเหตุของตัวแปรเป้าหมาย เช่น ดัชนีความตั้งใจซื้อบริการ ที่ใช้ทฤษฎีพฤติกรรมที่วางแผนไว้ (Theory of Planned Behavior) ในการสร้างดัชนี
 - การวัดแบบก่อเป็นรูปจากการประกอบ (Composite Formative Measurement) ตัวบ่งชี้เกิดจากการกวาดตัวแปรที่เป็นไปได้ทั้งหมดมาทำนายตัวแปรตาม โดยไม่สนใจว่าจะมีทฤษฎีสนับสนุนหรือไม่ เช่น การทำนายผลการปฏิบัติงานผ่านข้อมูลอัตชีวประวัติ (Biodata)
- แม้การแบ่งทั้งสองรูปแบบจะไม่ชัดเจน เพราะตัวแปรที่นำมาทำนาย มักจะนำมาจากทฤษฎีด้วย

ประเด็น	สะท้อน	ก่อเป็นรูป
คะแนน	มาตร (Scale)	ดัชนี (Index)
มุ่งหมาย	สะท้อนระดับของภาวะสันนิษฐาน	มัดกลุ่มตัวแปรไว้ด้วยกัน เพื่อความสะดวกในการอธิบายความสัมพันธ์กับตัวแปรอื่นๆ เช่น ดัชนีประสิทธิภาพทีม ที่มีความสัมพันธ์กับรูปแบบภาวะผู้นำ
ความสัมพันธ์กับตัวบ่งชี้	ตัวบ่งชี้เป็นผลที่เกิดขึ้นจากภาวะสันนิษฐาน	ตัวบ่งชี้เป็นสาเหตุที่ทำให้เกิด (ประกอบกันเป็น) ดัชนี
ความสัมพันธ์ระหว่างตัวบ่งชี้	ตัวบ่งชี้ต้องมีความสัมพันธ์กันเองสูง เพราะมีสาเหตุมาจากภาวะสันนิษฐานเดียวกัน	ตัวบ่งชี้ไม่จำเป็นต้องสัมพันธ์กัน ตัวบ่งชี้ที่ดีต้องไม่สัมพันธ์กับตัวบ่งชี้อื่นภายในดัชนี และสัมพันธ์กับตัวแปรนอกดัชนี (เช่น ตัวแปรตาม) สูง
การนำตัวบ่งชี้ออก	ความหมายของภาวะสันนิษฐานไม่เปลี่ยนแปลง	ความหมายของดัชนีเปลี่ยน ความสัมพันธ์กับตัวแปรนอกดัชนีเปลี่ยน
ลักษณะของตัวบ่งชี้	สะท้อนสิ่งที่ซ่อนอยู่ ไม่สามารถวัดได้โดยตรง มีบางสิ่งบางอย่างบงการตัวบ่งชี้ ที่ทำให้ตัวบ่งชี้มีค่าสูงหรือต่ำ	จะผสมไปด้วยตัวบ่งชี้ ที่ดูแล้วไม่ได้เกี่ยวข้องกับตัวบ่งชี้อื่นเลย อาจเป็นลักษณะของสิ่งแวดล้อมภายนอกด้วยซ้ำ เช่น นำตัวแปรข้อมูลพื้นฐาน (เช่น เพศ อายุ เรียนจบเมืองนอกหรือไม่) มาร่วมกันสร้างดัชนีภาวะผู้นำ
ถ้านำไปวิเคราะห์องค์ประกอบ	ได้การจับกลุ่มของตัวบ่งชี้ชัดเจน	ได้ผลออกมาจับกลุ่มกันได้ไม่หมด ไม่ชัดเจน ตัวแปรสำคัญอาจไม่เข้าพวกกับผู้อื่น
การพัฒนา	ไม่จำเป็นต้องมีตัวแปรอื่นนอกองค์ประกอบมาอ้างอิง เน้นความสัมพันธ์ภายในตัวบ่งชี้	“ควร” ตัวแปรอื่นมาอ้างอิง ความหมายดัชนีขึ้นอยู่กับความสัมพันธ์กับตัวแปรอื่นข้างนอก เช่น นำภาวะผู้นำปฏิรูป แบบแลกเปลี่ยน แบบปล่อยอิสระ มาเป็นตัวแปรตาม เพื่อหาดัชนีภาวะผู้นำจากตัวแปรข้อมูลพื้นฐาน

การวัดแบบก่อเป็นรูป (Formative Measurement)

- การวัดสิ่งเหล่านี้ เป็น “มาตร” หรือ “ดัชนี”
 - ผลการปฏิบัติงาน
 - ภาวะผู้นำ
 - การทำงานเป็นทีม
- ถ้าวัด “คุณลักษณะ” จะเป็นมาตร แต่ถ้าวัด “ส่วนประกอบ” ที่ทำให้เกิดสิ่งเหล่านี้ จะเป็นดัชนี
- ถ้าถามว่า หัวหน้าชั้นเป็นผู้นำที่ดี หัวหน้าชั้นสามารถทำให้ทีมประสบความสำเร็จ หัวหน้าชั้นเป็นศูนย์รวมจิตใจของคนในทีม สามตัวบ่งชี้นี้เป็นลักษณะของมาตร
- แต่ถ้าถามว่า หัวหน้านำความคิดคนได้ หัวหน้าใส่ใจคนในงาน หัวหน้ากระตุ้นให้เกิดความคิดใหม่ หัวหน้าติดตามงานคนในทีม หัวหน้าแบ่งปันผลประโยชน์ของคนในทีมให้ลงตัว หัวหน้าเจรจาต่อรองได้ดี สิ่งเหล่านี้ อาจมองว่าเป็นส่วนประกอบที่ทำให้เป็นหัวหน้าที่ดี จะเป็นลักษณะของดัชนี

การวัดแบบก่อเป็นรูป (Formative Measurement)

- ลองนึกภาพว่าตัวบ่งชี้ในแต่ละมาตรมีสาเหตุร่วมกันหรือไม่ ถ้าดึงตัวบ่งชี้หนึ่งออกไป ทำให้ความหมายเปลี่ยนหรือไม่ ถ้าใช่ก็เป็นดัชนี ถ้าไม่ใช่ก็เป็นมาตร
- มาตรที่ดูแล้วอาจเหมาะสมกับโมเดลแบบก่อเป็นรูป (เปลี่ยนเป็นดัชนี) เช่น
 - ต้นทุนทางจิตวิทยา (Psychological Capital) คือ ภาวะจิตวิทยาทางบวกที่ส่งผลให้เกิดผลงานที่ดีและประสบความสำเร็จ (แบ่งเป็น Hope, Self-Efficacy, Resilience, Optimism)
 - ผลการปฏิบัติงาน (Job Performance) งานอาจแบ่งออกเป็นส่วนย่อยๆ ที่คะแนนไม่ได้สอดคล้องกัน แต่นักวิจัยหรือองค์กรจำเป็นต้องรวมเพื่อเปรียบเทียบผลงานระหว่างบุคคล
 - คุณภาพของชีวิต (Quality of Life)
 - ความเชี่ยวชาญด้านเทคโนโลยี

การวัดแบบก่อเป็นรูป (Formative Measurement)

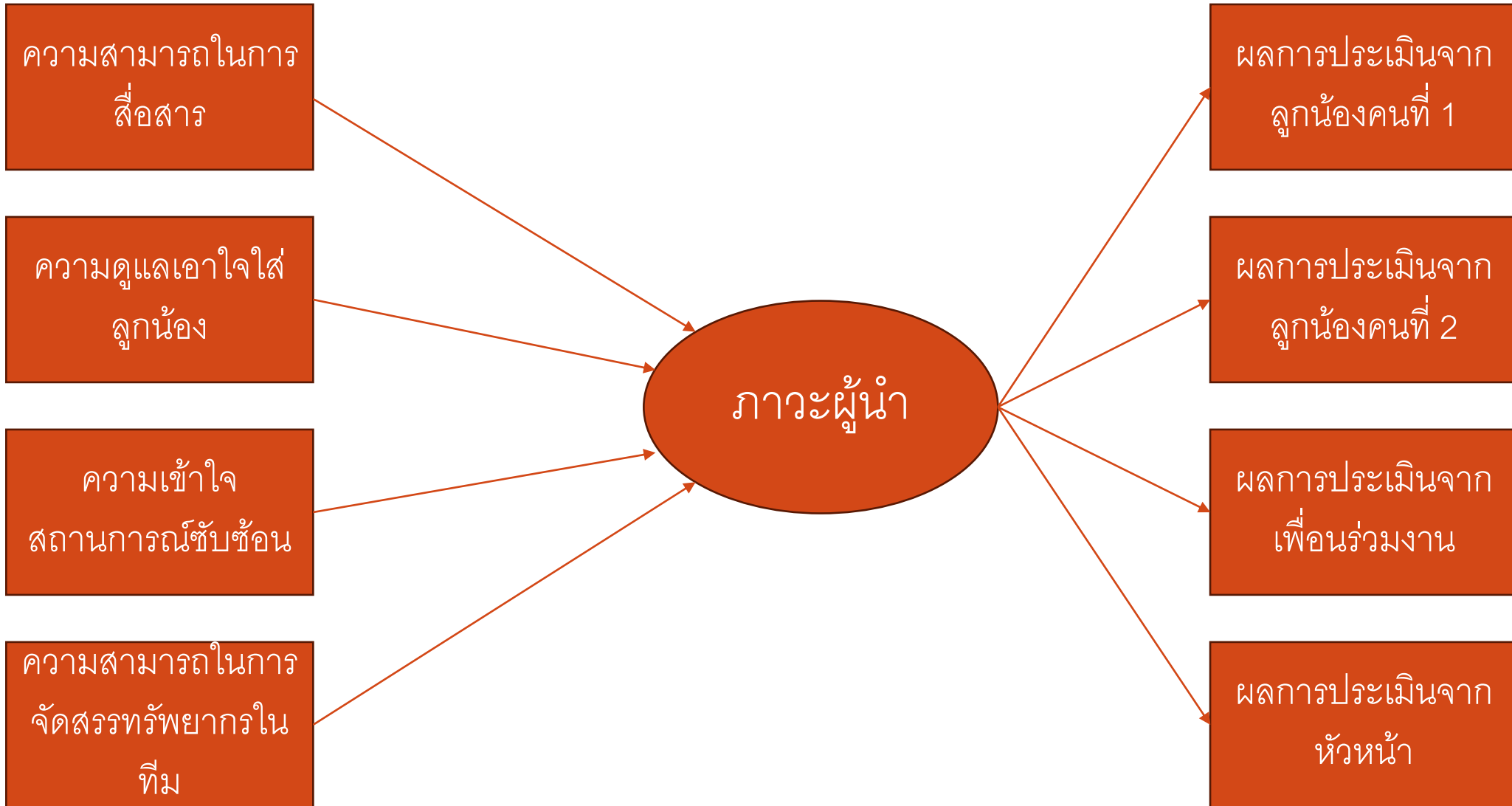
- ดังนั้น การวัดทางจิตวิทยาที่ดี ควรตั้งเจตนาว่าเป็นมาตรฐานหรือดัชนี กระบวนการสร้างจะต่างกันเล็กน้อย
- กระบวนการสร้างมาตรวัดแบบสะท้อน จะเริ่มจาก
 - เขียนข้อคำถามที่สะท้อนภาวะสันนิษฐาน ให้ตรง และครอบคลุม ข้อคำถามต้องเป็น “ผล” ของภาวะสันนิษฐาน
 - วิเคราะห์ข้อคำถาม ดูการจับกลุ่มเป็นองค์ประกอบย่อย ตัดข้อคำถามที่มีน้ำหนักองค์ประกอบต่ำออก
 - วัดความเที่ยงรูปแบบต่างๆ รวมถึงความสอดคล้องภายใน
 - ตรวจสอบความตรงตามความสัมพันธ์กับตัวแปรภายนอก
 - สร้างกลุ่มอ้างอิง เพื่อสามารถนำคะแนนเปรียบเทียบกับกลุ่มอ้างอิง หรือเทียบกับจุดตัดได้

การวัดแบบก่อเป็นรูป (Formative Measurement)

- กระบวนการสร้างมาตรวัดแบบก่อเป็นรูป
 - ทบทวนทฤษฎี ว่าตัวแปรอะไรเป็น “สาเหตุ” ที่ทำให้เกิดตัวแปรตาม หาตัวแปรสาเหตุให้ครอบคลุมมากที่สุด ตามทฤษฎี ตามกรอบการทำงาน ตามโดเมนที่ตัวแปรสาเหตุครอบคลุม เช่น คุณสมบัติที่ทำให้เป็นหัวหน้าที่ดี, รูปแบบปัญหาที่หัวหน้าจะพบเจอ เป็นต้น
 - อาจเลือกเฉพาะกรอบแนวคิดที่สนใจได้ เช่น วิธีการคิด การรับรู้ ที่ทำให้คนมีความสุข ซึ่งจะคัดเลือกเฉพาะคุณลักษณะที่เป็นแนวคิด การรับรู้เท่านั้น ไม่สนใจประเด็นอื่น เช่น ความมั่นคงทางอารมณ์ ฐานะการเงิน ฯลฯ
 - “อาจ” นำไปทำนายตัวแปรเป้าหมายที่ต้องการ ตัดตัวแปรที่ไม่สามารถทำนายตัวแปรเป้าหมายได้ดีออก
 - วัดความเที่ยงในรูปแบบต่างๆ เช่น แบบวัดซ้ำ แบบผู้ประเมิน โดยไม่สนใจเรื่องความสอดคล้องภายใน
 - ตรวจสอบความตรงตามความสัมพันธ์กับตัวแปรภายนอก “อื่น” ที่ไม่ใช่ตัวแปรเป้าหมายที่นำไปคำนวณน้ำหนัก
- ดังนั้น ทั้งมาตรและดัชนีต้องการความเที่ยงและความตรงทั้งคู่ แต่แนวคิดว่ามาตรที่ดี หรือดัชนีที่ดีเป็นอย่างไรนั้นแตกต่างกัน

การวัดแบบก่อเป็นรูป (Formative Measurement)

- ตัวแปรเป้าหมายที่เหมาะสมควรเป็นอย่างไร
 - ตัวแปรเป้าหมายบางครั้ง ชัดเจนอยู่แล้ว จับต้องได้ เช่น การลาออก การเกิดภาวะเสียชีวิตรหว่างผ่าตัด
 - ตัวแปรเป้าหมายบางครั้งไม่ชัดเจน วัดโดยตรงไม่ได้ เช่น ผลการปฏิบัติงาน ความสุข ภาวะผู้นำ
- ถ้าตัวแปรเป้าหมายไม่ชัดเจน ตัวแปรเป้าหมายอาจมีลักษณะของภาวะสันนิษฐานได้เช่นกัน ก็ต้องใช้หลักการวัดแบบสะท้อน มาหาคะแนนของตัวแปรเป้าหมายอยู่ดี
 - เช่น ภาวะผู้นำ จะใช้เป็นการประเมินตนเอง หรือการประเมินจากผู้ใต้บังคับบัญชา?
 - จะประเมินด้วยข้อคำถามข้อเดียว หรือประเมินด้วยหลากหลายข้อคำถาม หรือประเมินจากผู้ใต้บังคับบัญชาหลายคน
 - กล่าวคือ ใช้ “มาตร” เป็นตัวแปรเป้าหมายในการสร้าง “ดัชนี”



Formative

Reflective



Formative

Reflective??

การวัดแบบก่อเป็นรูป (Formative Measurement)

- กรณีที่ตัวแปรตามหลายๆ ตัว ไม่ได้สะท้อนมาจากภาวะสันนิษฐานเดียวกัน อาจเป็นไปได้ที่ดัชนีเดียว ไม่เพียงพอในการอธิบายชุดของตัวแปรตามทั้งหมด เช่น
 - ความพึงพอใจในงาน และภาวะหมดไฟ ถูกทำนายด้วยดัชนี $0.5H + 0.2E + 0.1R + 0.8O$ ได้ดี
 - ผลการปฏิบัติงาน ถูกทำนายด้วยน้ำหนัก $0.1H + 0.9E + 0.2R + 0.1O$ ได้ดี
 - สุขภาพกาย ถูกทำนายด้วยน้ำหนัก $0.1H + 0.2E + 0.8R + 0.3O$ ได้ดี
- กรณีเลือกดัชนีเดียว ดัชนีนี้จะพยายามทำนายตัวแปรตามให้มากที่สุด โดยใช้สมการทำนายเดียว เช่น ดัชนี $0.5H + 0.25E + 0.4R + 0.6O$ ทำนายตัวแปรตามความพึงพอใจในงานได้ 20% ภาวะหมดไฟได้ 15% ผลการปฏิบัติงานได้ 10% และสุขภาพกายได้ 7% ซึ่งเป็นค่าสูงสุด
- ตามหลักต้องทดสอบว่า หากเพิ่มดัชนีที่สองขึ้นมา ทำให้อธิบายตัวแปรตามได้เพิ่มขึ้นอย่างมีนัยสำคัญหรือไม่

Table 2.1: Regressions to Explain Average Happiness across Countries (Pooled OLS)

Independent Variable	Dependent Variable			
	Cantril Ladder (0-10)	Positive Affect (0-1)	Negative Affect (0-1)	Cantril Ladder (0-10)
Log GDP per capita	0.36 (0.066)***	-.013 (0.009)	0.0001 (0.007)	0.388 (0.065)***
Social support	2.420 (0.368)***	0.316 (0.055)***	-.328 (0.049)***	1.778 (0.361)***
Healthy life expectancy at birth	0.029 (0.01)***	-.0007 (0.001)	0.003 (0.001)***	0.03 (0.01)***
Freedom to make life choices	1.305 (0.298)***	0.368 (0.041)***	-.090 (0.04)**	0.509 (0.284)*
Generosity	0.583 (0.265)**	0.09 (0.032)***	0.024 (0.027)	0.378 (0.254)
Perceptions of corruption	-.704 (0.271)***	-.006 (0.027)	0.094 (0.022)***	-.704 (0.259)***
Positive affect				2.222 (0.333)***
Negative affect				0.173 (0.395)
Year fixed effects	Included	Included	Included	Included
Number of countries	156	156	156	156
Number of obs.	1853	1848	1852	1847
Adjusted R-squared	0.753	0.439	0.322	0.777

Notes: This is a pooled OLS regression for a tapered panel explaining annual national average Cantril ladder responses from all available surveys from 2005 through 2021. See **Technical Box 2** for detailed information about each of the predictors. Coefficients are reported with robust standard errors clustered by country in parentheses. ***, **, and * indicate significance at the 1, 5 and 10 percent levels respectively.



การวัดแบบก่อเป็นรูป (Formative Measurement)

- แล้วทำไมต้องทำดัชนีการวัด ทำไมเราไม่วัดตัวแปรเป้าหมายเลย
 - ตัวแปรที่ต้องการวัดยังไม่เกิดขึ้น เป็นสิ่งที่เกิดขึ้นในอนาคต ต้องการทำนาย
 - ตัวแปรที่นำมาสร้างดัชนี มีส่วนประกอบ ที่ทำให้เข้าใจกระบวนการเกิดตัวแปรเป้าหมาย

การวัดแบบก่อเป็นรูป (Formative Measurement)

- บางครั้ง การสร้างดัชนี ผู้พัฒนาอาจยังไม่ได้วัดตัวแปรเป้าหมาย สร้างดัชนีจากการรวมคะแนนจากตัวบ่งชี้ แล้วนำดัชนีดังกล่าวไปใช้เลย สามารถทำได้หรือไม่
 - สังเกตว่า ดัชนีจะสร้างจากการรวมคะแนนตัวบ่งชี้เข้าด้วยกัน ไม่มีน้ำหนักแบบละเอียด เป็นการรวมคะแนนแบบง่าย ๆ เช่น เอาค่ามาบวกกันตรง ๆ
 - ตัวบ่งชี้ที่นำมาประกอบเป็นดัชนี มักจะสร้างจากทฤษฎี หรือสุ่มจากสถานการณ์ในชีวิตประจำวัน เช่น นำสถานการณ์ในการทำงานด้านต่างๆ มาใช้ (หรือการหาเหตุการณ์วิกฤต [Critical Incidents])
 - สมรรถนะ (Competency) มีลักษณะเหมือนดัชนี ที่อาจไม่มีตัวแปรเป้าหมาย เช่น การคัดเลือกบุคลากรจากสมรรถนะ อาจเกิดจากแนวคิดของผู้บริหารที่ต้องการให้พนักงานของตนเป็น เพื่อผลักดันองค์กรสู่เป้าหมาย ซึ่งเป้าหมายดังกล่าว ไม่ได้มีการวัดแล้วเอาสมรรถนะไปถ่วงน้ำหนักจริงๆ
 - การมีตัวแปรเป้าหมายมาอ้างอิง ย่อมทำให้ดัชนีที่สร้างขึ้นมีหลักฐานความตรงมากขึ้น การสร้างจากทฤษฎีอย่างเดียว อาจสร้างความสงสัยให้แก่ผู้ใช้ ว่าเป็นตัวแปรที่ทำนายได้จริงหรือไม่ น้ำหนักที่ให้เท่ากันนั้นเหมาะสมหรือไม่

Herde & Lievens, 2022

- บทบาทสมมติ 18 รูปแบบ มีนักแสดงที่แสดงบทบาทเป็นผู้บริหารโปรเจ็ค มีปัญหาต่างๆ มาปรึกษา กระจายออกเป็นปัญหาต่างๆ ด้านต่างๆ
- ผลการแสดงผลบทบาทสมมติ จะถูกประเมินจากผู้ประเมิน 3 คน จาก 1 (ควรพัฒนา) จนถึง 9 (ตอบสนองดีจนเป็นแบบอย่าง)
- คะแนนถูกนำมาเฉลี่ยในแต่ละบทบาท แล้วนำมารวมเป็นคะแนนรวม
- เป็นดัชนี ที่ไม่ได้อิงกับตัวแปรเป้าหมาย
- มีความสัมพันธ์กับผลงาน .54

This conceptual underpinning leads to several implications. Due to the heterogeneity of the domain, the diversity of the situational demands of the simulations, and the variability in people's performances across situations, we do not assume a unidimensional latent variable as the "single behavior-generating" mechanism across all simulations (see Bledow & Frese, 2009, p. 241). Hence, we expect that simulation scores will not correlate highly (but at best moderately or even lowly) with each other. Relatedly, the overall score that people obtain based on multiple, speeded simulations will depend on the kind and variety of the simulations sampled. As a critical implication, this means that removing specific simulations might change what is being measured and predicted by this overall simulation score. In other words, removing particular simulations (or overlooking to include simulations) might reduce domain coverage and be detrimental for what is being measured, thereby underscoring the importance of stimulus domain sampling.

- สร้างดัชนี โดยยังไม่มีตัวแปรเป้าหมาย แต่มาทดสอบกับตัวแปรเป้าหมายในอนาคต
- ถ้าการถ่วงน้ำหนัก ไม่ได้มาจากตัวแปรเป้าหมาย ความสัมพันธ์กับผลงาน สามารถใช้เป็นหลักฐานความตรงได้

การวัดแบบก่อเป็นรูป (Formative Measurement)

- ถ้านำดัชนี ไปวิเคราะห์แบบมาตร จะทำให้ข้อคำถามสำคัญ ถูกตัดออกไป เช่น แหล่งความเครียด เป็นเรื่องการเสียชีวิตของลูกตนเอง ตัวแปรนี้เป็นแหล่งความเครียดที่สำคัญมาก แต่จำนวนคนที่เจอความเครียดนี้ไม่เยอะเลยอาจไม่เจอความสัมพันธ์กับตัวบ่งชี้อื่น ถ้ายึดแนวการสร้างมาตร อาจทำให้ข้อนี้ถูกตัดออก ทั้งที่เป็นแหล่งความเครียดที่ทำนายความเครียดได้สูง
- ถ้านำมาตร ไปวิเคราะห์แบบดัชนี อาจทำให้คะแนนไม่สะท้อนภาวะสันนิษฐานที่ต้องการ เพราะข้อคำถามบางข้อถูกตัดออก เนื่องจากไม่สามารถทำนายตัวแปรตามได้เพิ่มเติมได้ ทั้งที่เป็นตัวสะท้อนภาวะสันนิษฐานที่ดี
- อย่างไรก็ตาม บางอย่างอาจเลือกยาก ว่าเป็นมาตรหรือดัชนี บางครั้งอาจต้องสร้างข้อคำถามมาก่อน แล้วตรวจสอบความสัมพันธ์ระหว่างตัวบ่งชี้ หรือวิเคราะห์องค์ประกอบ จนเข้าใจว่าเป็นมาตร หรือดัชนี ถ้าวัดรู้ว่า เป็นดัชนี ค่อยไปอิงกับตัวแปรเป้าหมาย
- ดู Rhemtulla, van Bork, & Borsboom (2020) ถึงผลกระทบจากการวิเคราะห์แบบมาตรวัดสะท้อน ไปใช้กับมาตรวัดแบบก่อรูป

การวัดแบบก่อเป็นรูป (Formative Measurement)

- การวิเคราะห์การวัดแบบก่อเป็นรูปจากการประกอบ
 - หากตัวแปรตามมีเพียงแค่ตัวเดียว สามารถใช้การวิเคราะห์ถดถอยได้เลย
 - หากตัวแปรตามมีหลายตัวแปร ที่ไม่ได้มาจากภาวะสันนิษฐานเดียวกัน ให้ใช้การวิเคราะห์ความซ้ำซ้อน (Redundancy Analysis)
 - หากตัวแปรตามมีหลายตัวแปร และคาดว่ามาจากภาวะสันนิษฐานเดียวกัน อาจใช้โมเดล Multiple Causes Multiple Indicators Model (MIMIC Model) แทน แต่การวิเคราะห์ความซ้ำซ้อนจะดีกว่า
 - สามารถสร้างดัชนี ผสมกับการวิเคราะห์สมการเชิงโครงสร้าง (Structural Equation Modeling: SEM) เช่น ต้องการวิเคราะห์ตัวแปรส่งผ่าน ระหว่างตัวแปรผลลัพธ์ ดูรายละเอียดที่ Hair et al. (2022), Yu, Schubert, & Henseler (2023), หรือ Gu, Yung, & Cheung (2019)

การวัดแบบก่อเป็นรูป (Formative Measurement)

- ตัวอย่างวิเคราะห์ความซ้ำซ้อน (Redundancy Analysis)
- ตัวแปรอิสระ นำมาจากมาตรวัดคุณลักษณะความเป็นผู้นำ มีทั้งหมด 9 ด้าน โดยลูกน้องประเมินหัวหน้า คือ
 - ความรู้สึกมีเสน่ห์ของผู้นำ (Attributed Idealized Influence)
 - พฤติกรรมทำตามความคิดความเชื่อของผู้นำ (Behavioral Idealized Influence)
 - ความใส่ใจรายบุคคล (Individual Consideration)
 - ความสามารถในการดลใจ จูงใจ (Inspirational Motivation)
 - การกระตุ้นให้คิด (Intellectual Stimulation)
 - การให้รางวัลตอบแทน (Contingent Reward)
 - การจัดการเชิงป้องกัน ช่มชู้ (Active Management by Exception)
 - การจัดการเมื่อเกิดปัญหาเท่านั้น (Passive Management by Exception)
 - การปล่อยลูกน้องอิสระ (Laissez-faire)

การวัดแบบก่อเป็นรูป (Formative Measurement)

- ต้องการสร้างดัชนีจากคุณสมบัติทั้ง 9 ด้าน ในการทำนายผลลัพธ์ 3 ด้าน โดยลูกน้องประเมินตนเอง คือ
 - ลูกน้องมีประสิทธิภาพในการทำงาน (Effectiveness)
 - ลูกน้องมีความพยายามเพิ่มเติมในการทำงาน (Extra Effort)
 - ลูกน้องมีความพึงพอใจในการทำงาน (Satisfaction)
- เก็บข้อมูลจากลูกน้อง 201 คน นำตัวแปรทำนาย 9 ตัวทำนายตัวแปรผลลัพธ์ 3 ตัว ได้ผลดังนี้ (ดูวิธีการวิเคราะห์ข้อมูลด้วย R ที่เว็บ)

รายการ	ผลรวมที่ 1	ผลรวมที่ 2	ผลรวมที่ 3
ตัวแปรทำนาย			
ความรู้สึกมีเสน่ห์ของผู้นำ (Attributed Idealized Influence)	-0.795	0.089	0.040
พฤติกรรมทำตามความคิดความเชื่อของผู้นำ (Behavioral Idealized Influence)	-0.734	-0.389	-0.425
ความใส่ใจรายบุคคล (Individual Consideration)	-0.560	-0.084	0.388
ความสามารถในการดลใจ จูงใจ (Inspirational Motivation)	-0.860	0.071	-0.270
การกระตุ้นให้คิด (Intellectual Stimulation)	-0.750	0.314	-0.112
การให้รางวัลตอบแทน (Contingent Reward)	-0.830	-0.199	0.261
การจัดการเชิงป้องกัน ช่มชู้ (Active Management by Exception)	-0.450	0.262	-0.199
การจัดการเมื่อเกิดปัญหาเท่านั้น (Passive Management by Exception)	0.130	-0.149	0.005
การปล่อยลูกน้องอิสระ (Laissez-faire)	0.337	-0.480	-0.117
ตัวแปรตาม			
ลูกน้องมีประสิทธิภาพในการทำงาน (Effectiveness)	-1.998	0.346	0.090
ลูกน้องมีความพยายามเพิ่มเติมในการทำงาน (Extra Effort)	-2.088	-0.116	-0.306
ลูกน้องมีความพึงพอใจในการทำงาน (Satisfaction)	-1.764	-0.255	0.261
ค่าสถิติ			
ความแปรปรวน (Eigenvalues)	1.40**	0.02	0.02
สัดส่วนการทำนาย	46.8%	0.8%	0.7%

มีเฉพาะผลรวมแรกที่ถึงระดับนัยสำคัญ และสัดส่วนการทำนายเกิน 1% จึงสนใจแค่ผลรวมแรก

ผลรวมแรกได้รับอิทธิพลจากตัวแปรทำนาย 6 ตัวแรกเยอะ ซึ่งรวมๆ เรียกว่า ผู้นำแบบปฏิรูปและแลกเปลี่ยน ซึ่งหากค่าผู้นำนี้สูง จะทำให้ผลรวมมีค่าต่ำ (เครื่องหมายลบ) แล้วหากผลรวมมีค่าต่ำ จะทำให้ลูกน้องมีประสิทธิภาพในการทำงาน มีความพยายาม และพอใจ (ผลลบในตัวแปรตาม)

ผลรวมแรกทำนายตัวแปรตามทั้ง 3 ตัวได้ 46.8%

การวัดแบบก่อเป็นรูป (Formative Measurement)

- นักวิจัยสามารถใช้วิธีการคัดเลือกตัวแปรจากการวิเคราะห์ถดถอย เช่น วิธี Stepwise เพื่อลดจำนวนตัวแปรทำนายให้น้อยลง
- สามารถนำคะแนนของตัวแปรทำนายแต่ละตัว ไปใช้หาค่าผลรวมที่ 1 ซึ่งอาจตั้งชื่อว่า “ดัชนีหัวหน้าที่ประสบความสำเร็จในการบริหารลูกน้อง”

อรรถประโยชน์

- แม้ว่าแบบทดสอบจะมีความเที่ยงสูง มีหลักฐานความตรงมาก แต่อาจไม่มีประโยชน์ในการใช้
 - เช่น ตำแหน่งงานปัจจุบันขาดแคลนอยู่แล้ว และรับผู้สมัครทุกคน
 - การทำแบบทดสอบจึงไม่ได้มีประโยชน์ในการช่วยคัดเลือกบุคลากร
- บางครั้งอาจไม่คุ้มค่าในการใช้
 - เช่น ค่าใช้จ่ายในการทดสอบผู้สมัครคนละ 1,000 บาท
 - ตำแหน่งหนึ่งมีผู้สมัคร 1,000 คน รับ 10 คน
 - การใช้แบบทดสอบนี้กับผู้สมัครทุกคนอาจไม่คุ้มค่า

อรรถประโยชน์

- อรรถประโยชน์ของแบบทดสอบขึ้นอยู่กับ
 - ราคาของแบบทดสอบ ทั้งในเชิงตัวเลขและไม่ใช่ตัวเลข ยิ่งสูงยิ่งทำให้ความคุ้มค่าต่ำลง
 - ประโยชน์จากการใช้แบบวัด ทั้งในเชิงตัวเลขและไม่ใช่ตัวเลข ยิ่งประโยชน์มาก ความคุ้มค่าก็จะสูงขึ้น
 - คุณสมบัติของแบบทดสอบ ความตรง ความเที่ยง ความสามารถในการทำนายสูง ทำให้ประโยชน์จากการใช้แบบวัดสูง

$$\text{อรรถประโยชน์ (Utility)} = \text{ประโยชน์ (Benefit)} - \text{ราคา (Cost)}$$

การวิเคราะห์หรือรถประโยชน์

- ตัวทำนาย (Predictor)
 - สอบผ่าน / ไม่ผ่าน
- เกณฑ์ (Criterion)
 - ต้องการ / ไม่ต้องการ

ผลการปฏิบัติงาน	สำเร็จ	ไม่ผ่าน	ผ่าน
	ล้มเหลว	ไม่ผ่าน	ผ่าน
		คะแนนสอบ	

การวิเคราะห์หรือรถประโยชน์

แบ่งพื้นที่เป็น 4 ส่วน

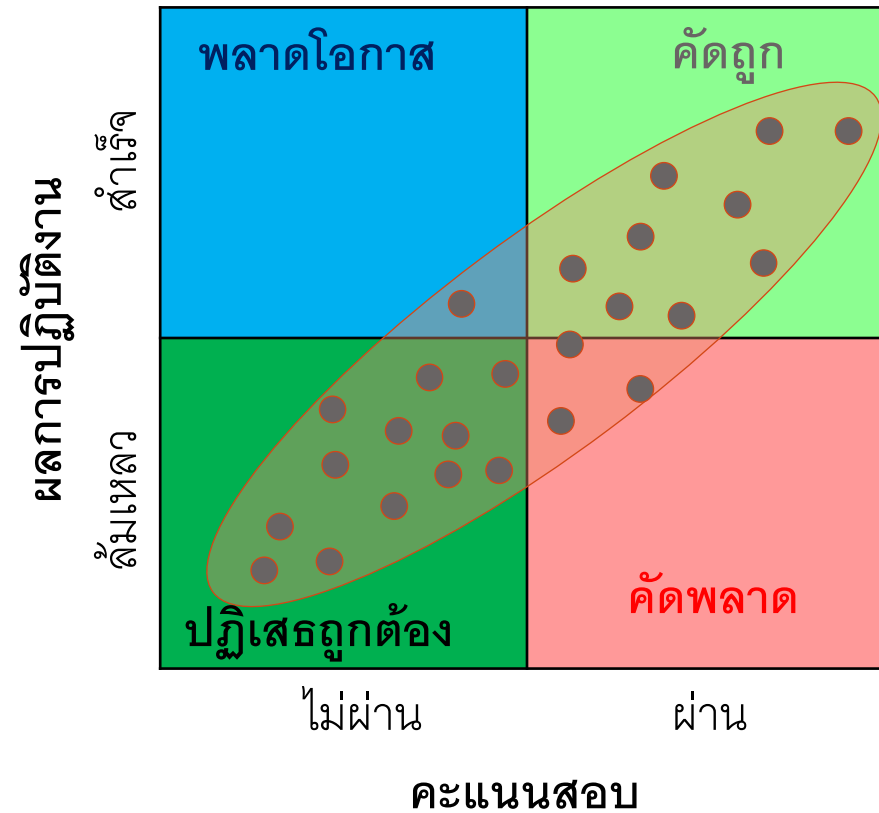
- คัดถูก (Hit; Correct Acceptance)
- ปฏิเสธถูกต้อง (Correct Rejection)
- พลาดโอกาส (Miss; False Rejection)
- คัดพลาด (False Acceptance)

ผลการปฏิบัติงาน	สำเร็จ	พลาดโอกาส	คัดถูก
	ล้มเหลว	ปฏิเสธถูกต้อง	คัดพลาด
		ไม่ผ่าน	ผ่าน
		คะแนนสอบ	

การวิเคราะห์หรือรถประโยชน์

จากผู้สมัคร 25 คน

- คัดถูก 9 คน
- ปฏิเสธถูกต้อง 12 คน
- พลาดโอกาส 1 คน
- คัดพลาด 3 คน

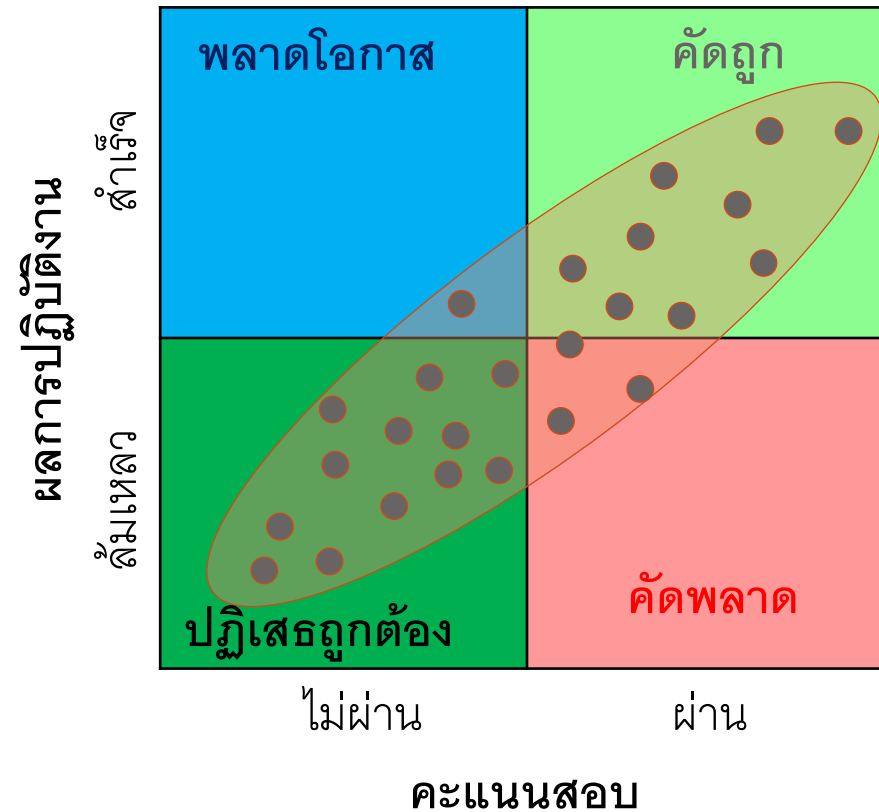


การวิเคราะห์หรือรถประโยชน์

จากผู้สมัคร 25 คน

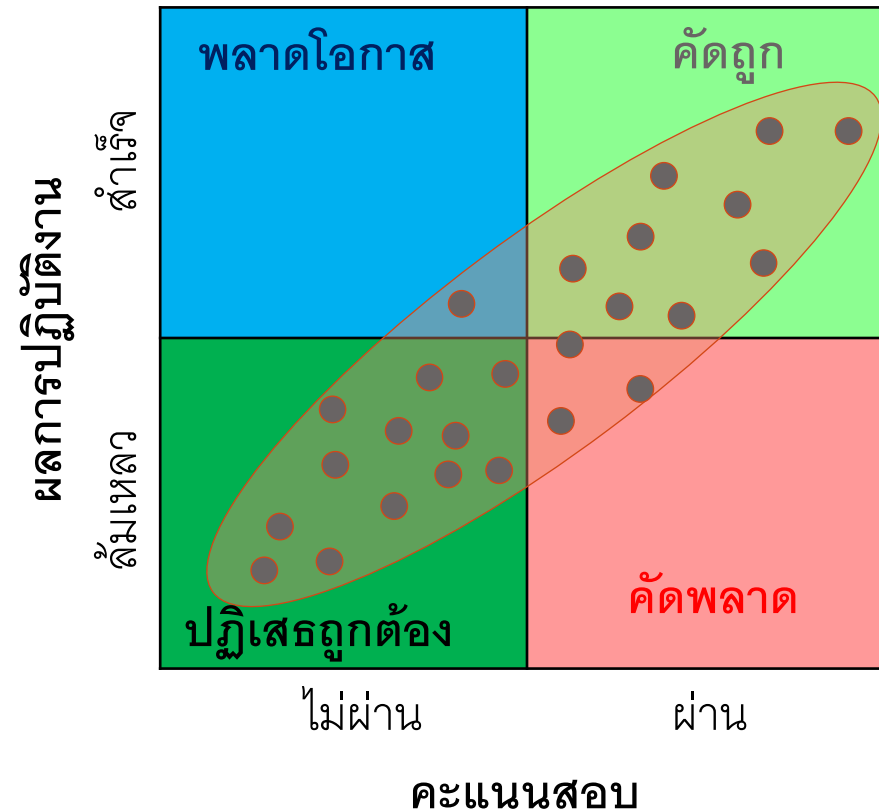
- สัดส่วนคัดสำเร็จ (Hit ratio) คือ สัดส่วนของคนที่ถูกคัดเลือกไปตามที่พึงประสงค์

$$\text{Hit Ratio} = \frac{\text{คัดถูก}}{\text{คัดมาทั้งหมด}} = 9 / (9 + 3) = .75$$



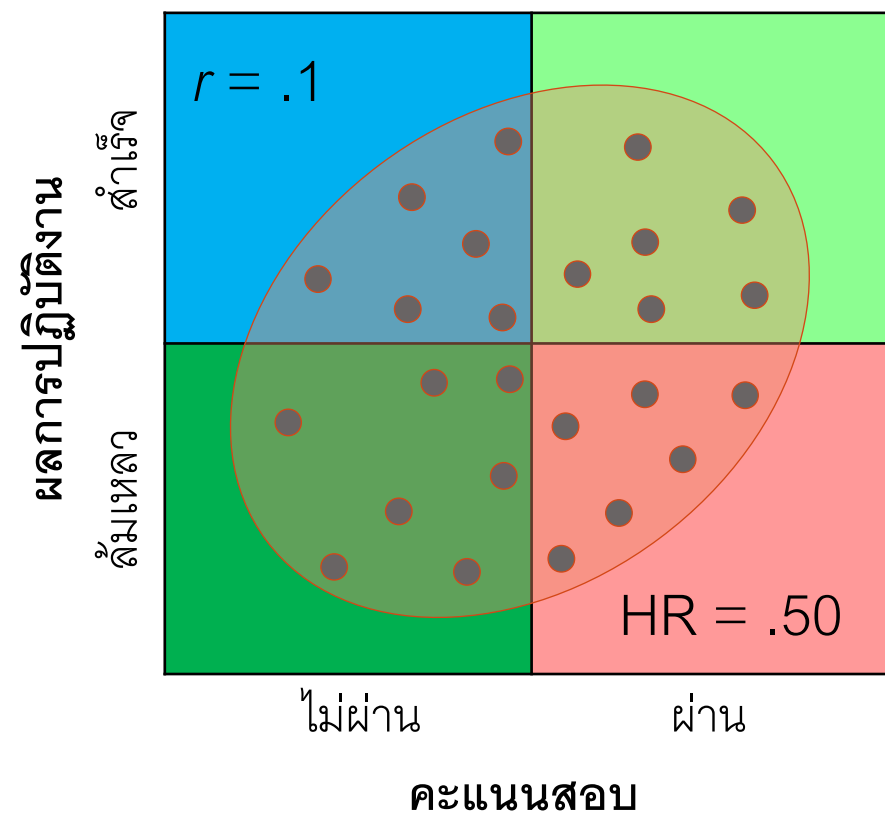
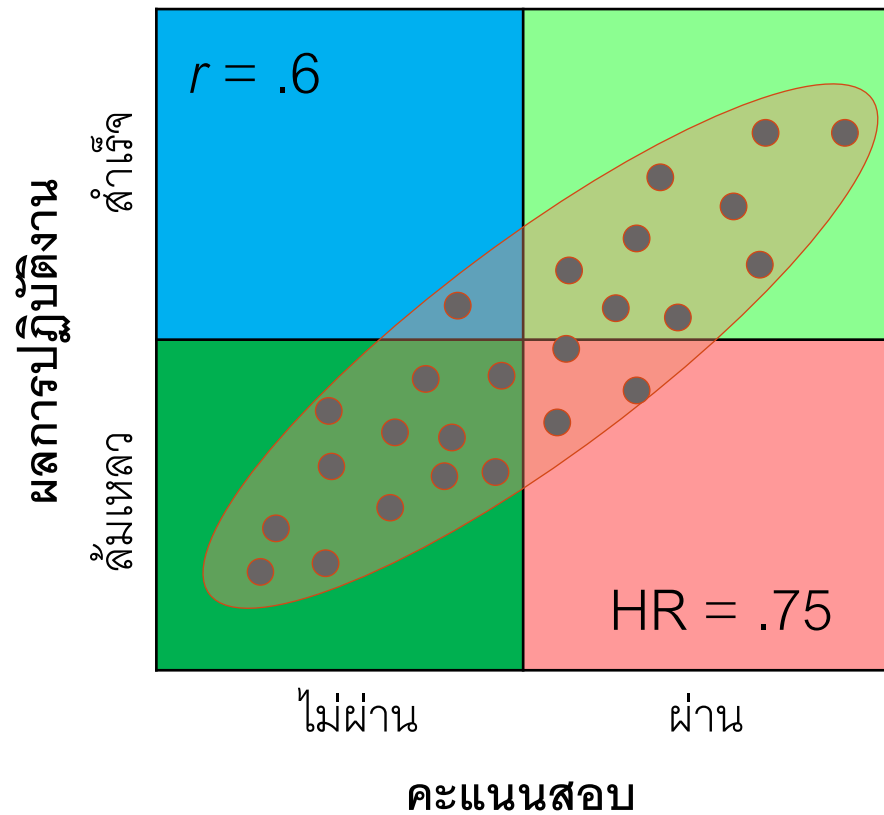
การวิเคราะห์หรือรรถประโยชน์

- สัดส่วนคัดสำเร็จ (Hit ratio) คือ อรรถประโยชน์ของแบบทดสอบ
- สัดส่วนคัดสำเร็จขึ้นอยู่กับ 3 ปัจจัย
 - ความสัมพันธ์ระหว่างตัวทำนายและเกณฑ์
 - จุดตัดของคะแนนที่ใช้คัดเลือก
 - ระดับผลการปฏิบัติงานที่แสดงว่าสำเร็จหรือล้มเหลว



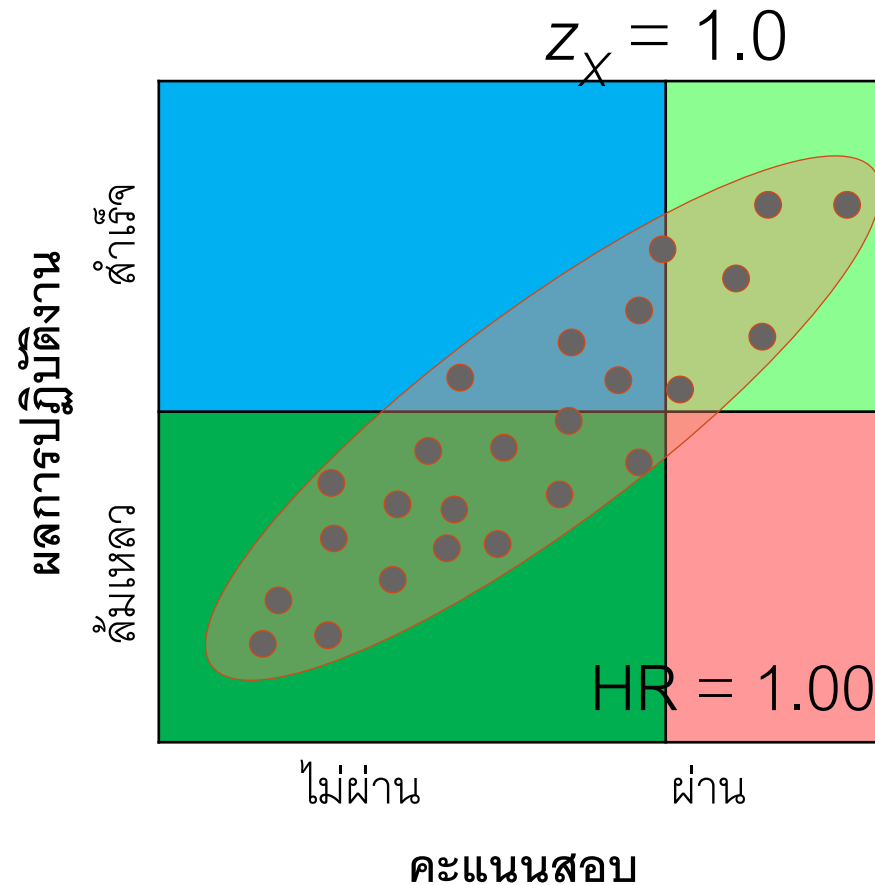
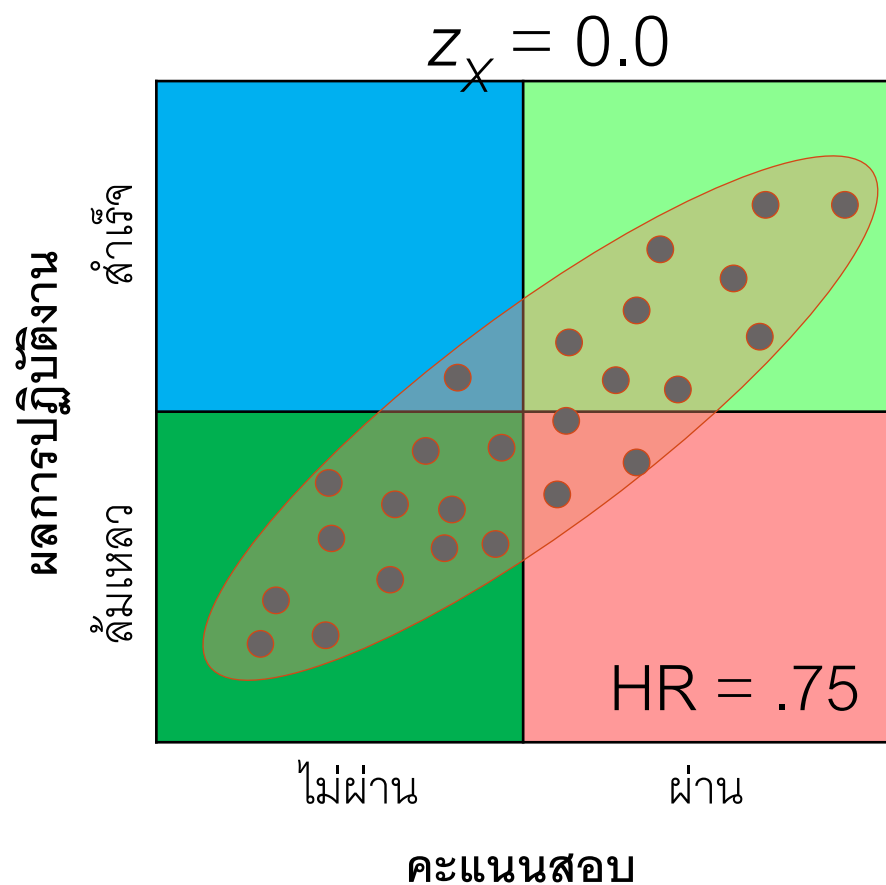
การวิเคราะห์หรือรรถประโยชน์

- ความสัมพันธ์ระหว่างตัวทำนายและเกณฑ์ (Predictive Validity)



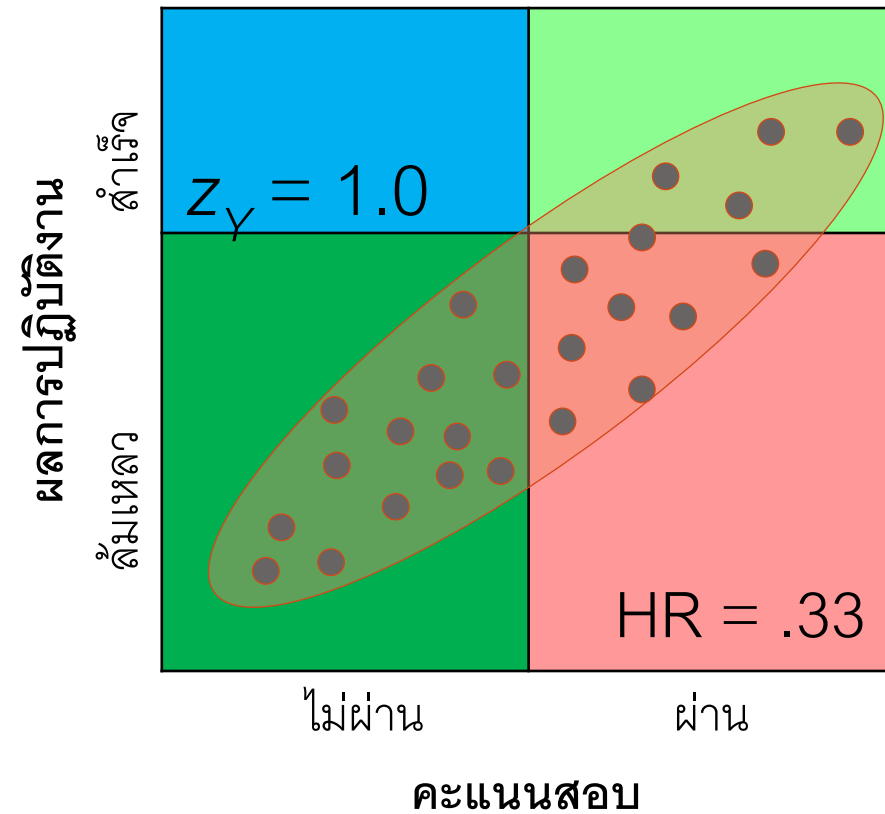
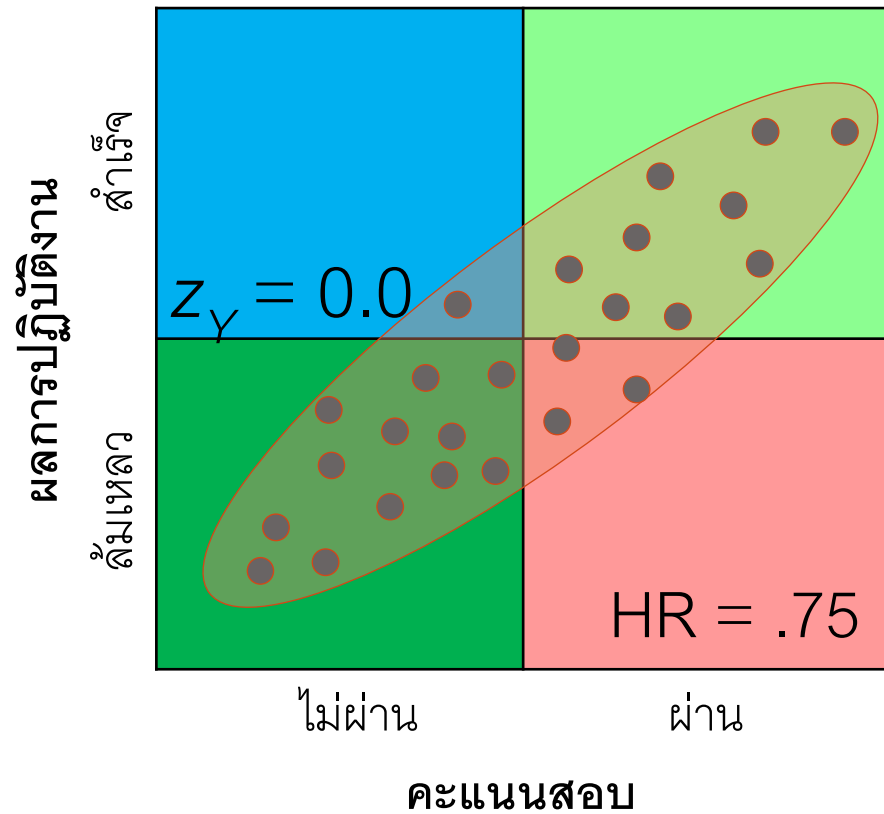
การวิเคราะห์ห้หรือรณประโยชน์

- จุดตัดของคะแนนที่ใช้คัดเลือก



การวิเคราะห์หรือรณประโยชน์

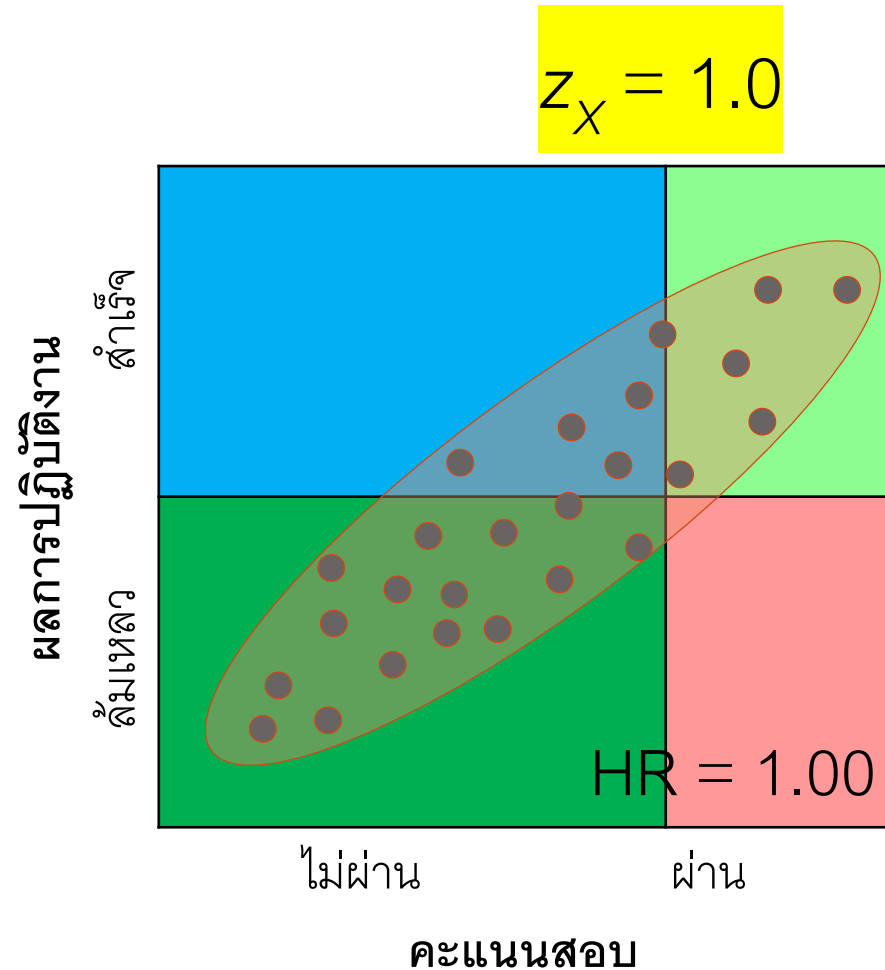
- ระดับผลการปฏิบัติงานที่แสดงว่าสำเร็จหรือล้มเหลว



การวิเคราะห์หรือรรถประโยชน์

- ระดับคะแนนที่ใช้คัดเลือก อาจรายงาน
ในรูปของสัดส่วนการคัดเลือก
(Selection Ratio) คือ สัดส่วน
ของผู้สมัครทั้งหมดที่ถูกคัดเลือกไป

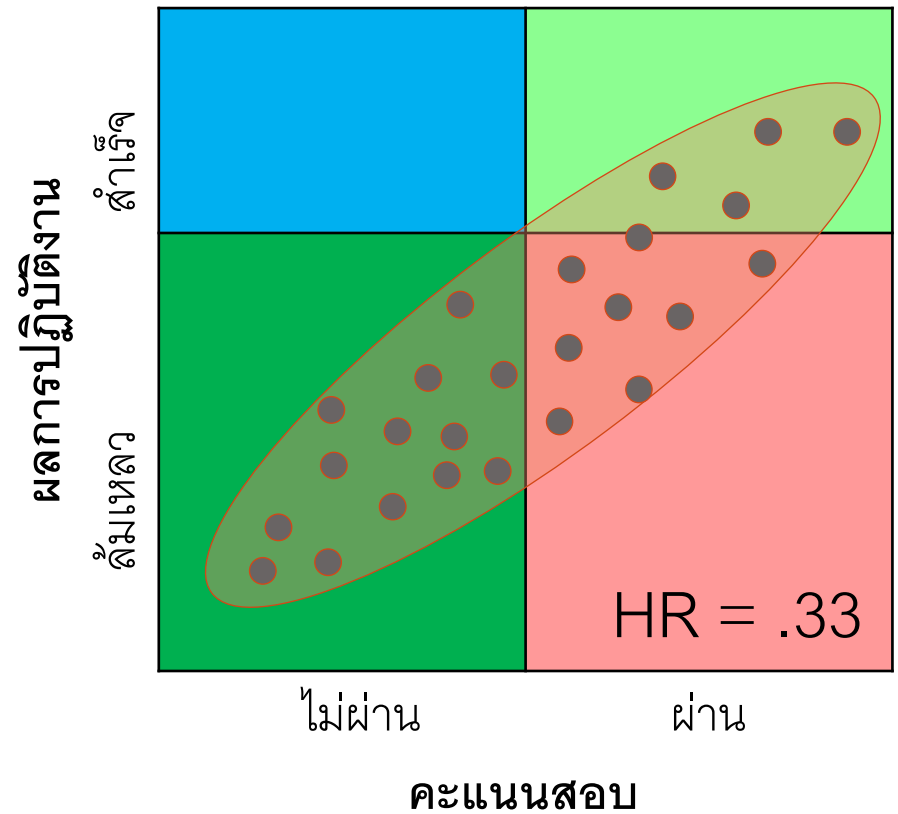
$$\text{Selection Ratio} = \frac{\text{ผ่าน}}{\text{ทั้งหมด}} \\ = 5 / 25 = .20$$



การวิเคราะห์หรือรรถประโยชน์

- ระดับผลการปฏิบัติงานที่แสดงว่าสำเร็จหรือล้มเหลว อาจรายงานในรูปของสัดส่วนพื้นฐาน (Base Rate) คือ สัดส่วนของคนทั้งหมดที่ประสบความสำเร็จ โดยที่ยังไม่ใช้คะแนนสอบเลย

$$\begin{aligned} \text{Base Rate} &= \frac{\text{สำเร็จ}}{\text{ทั้งหมด}} \\ &= 4 / 25 = .16 \end{aligned}$$



$$z_Y = 1.0$$

การวิเคราะห์หรือสรุปประโยชน์

- หากมีข้อมูลดิบ นักวิเคราะห์สามารถหาสัดส่วนคัดเลือกถูกต้อง (Hit Ratio) ได้เลย โดย
 - แปลงตัวแปรทำนาย (X) ให้เป็นตัวแปรใหม่ (D_X) มีสองกลุ่ม คือ ผ่านหรือไม่ผ่าน
 - แปลงตัวแปรเกณฑ์ (Y) ให้เป็นตัวแปรใหม่ (D_Y) มีสองกลุ่ม คือ สำเร็จหรือไม่สำเร็จ
 - สร้างตารางไขว้ (Crosstab) ระหว่าง D_X และ D_Y
 - คำนวณสัดส่วนคัดเลือกสำเร็จ

$$HR = P(D_Y = 1 | D_X = 1) = \frac{N(D_X = 1 \cap D_Y = 1)}{N(D_X = 1)}$$

การวิเคราะห์หรือรรถประโยชน์

- ในบางครั้ง ผู้วิจัยยังไม่มีข้อมูลดิบ แต่พอทราบข้อมูลของ
 - ความสัมพันธ์ระหว่างตัวแปรทำนายและเกณฑ์ (Predictive validity)
 - สัดส่วนการคัดเลือก (Selection ratio)
 - สัดส่วนพื้นฐาน (Base rate)
- หากข้อมูลการกระจายของตัวแปรทำนายและเกณฑ์เป็นโค้งปกติร่วม (Bivariate normality) แล้วสามารถใช้ตารางของ Taylor & Russell (1939) ในการหาสัดส่วนค่านวณถูกต้อง (Hit ratio) ได้

Proportion of Employees Considered Satisfactory $\pm .20$
 Selection Ratio

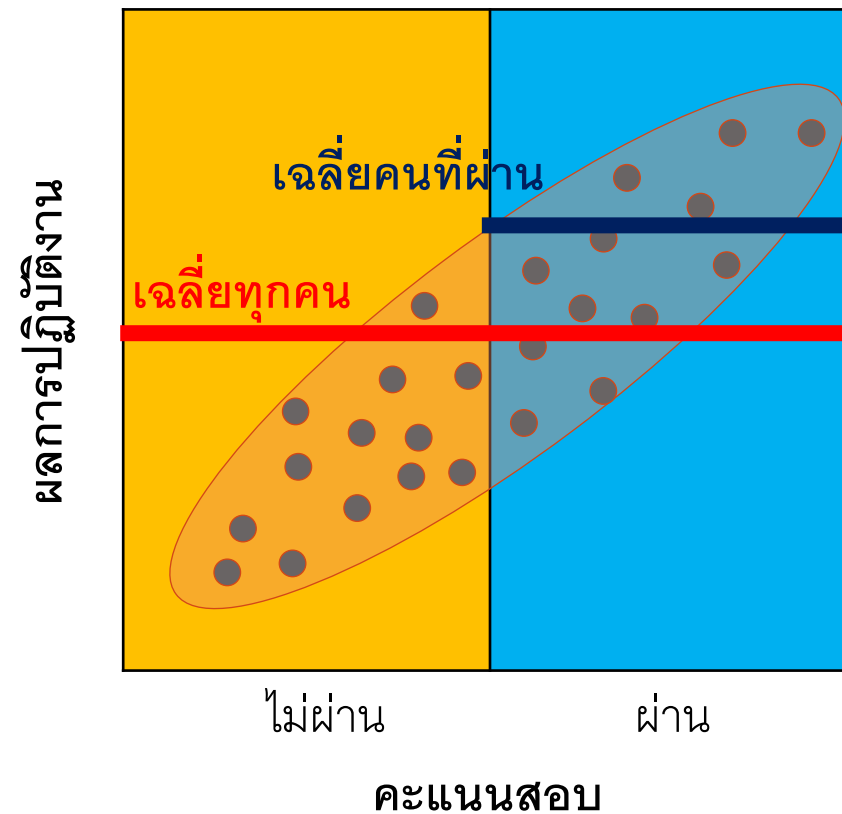
r	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
.00	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20
.05	.23	.23	.22	.22	.21	.21	.21	.21	.20	.20	.20
.10	.26	.25	.24	.23	.23	.22	.22	.21	.21	.21	.20
.15	.30	.28	.26	.25	.24	.23	.23	.22	.21	.21	.20
.20	.33	.31	.28	.27	.26	.25	.24	.23	.22	.21	.21
.25	.37	.34	.31	.29	.27	.26	.24	.23	.22	.21	.21
.30	.41	.37	.33	.30	.28	.27	.25	.24	.23	.21	.21
.35	.45	.41	.36	.32	.30	.28	.26	.24	.23	.22	.21
.40	.49	.44	.38	.34	.31	.29	.27	.25	.23	.22	.21
.45	.54	.48	.41	.36	.33	.30	.28	.26	.24	.22	.21
.50	.59	.52	.44	.38	.35	.31	.29	.26	.24	.22	.21
.55	.63	.56	.47	.41	.36	.32	.29	.27	.24	.22	.21
.60	.68	.60	.50	.43	.38	.34	.30	.27	.24	.22	.21
.65	.73	.64	.53	.45	.39	.35	.31	.27	.25	.22	.21
.70	.79	.69	.56	.48	.41	.36	.31	.28	.25	.22	.21
.75	.84	.74	.60	.50	.43	.37	.32	.28	.25	.22	.21
.80	.89	.79	.64	.53	.45	.38	.33	.28	.25	.22	.21
.85	.94	.85	.69	.56	.47	.39	.33	.28	.25	.22	.21

แบบทดสอบจะมีอัตราประโยชน์มากที่สุด เมื่อความสัมพันธ์ระหว่างเกณฑ์และตัวแปรทำนายสูง และสัดส่วนในการคัดเลือกต่ำ

```
> ### Specify three values
> sr <- 0.05 # Selection Ratio
> br <- 0.2 # Base Rate
> r <- 0.7 # Predictive Validity
>
> ##### calculation #####
> library(mvtnorm)
> zx <- qnorm(1 - sr)
> zy <- qnorm(1 - br)
> r <- matrix(c(1, r, r, 1), 2, 2)
> q1 <- pmvnorm(lower = c(zx, zy), mean=c(0,0), corr=r)
> q1q4 <- pmvnorm(lower = c(zx, -Inf), mean=c(0,0), corr=r)
> q1/q1q4
[1] 0.7856052
attr(,"error")
[1] 1e-15
attr(,"msg")
[1] "Normal Completion"
```

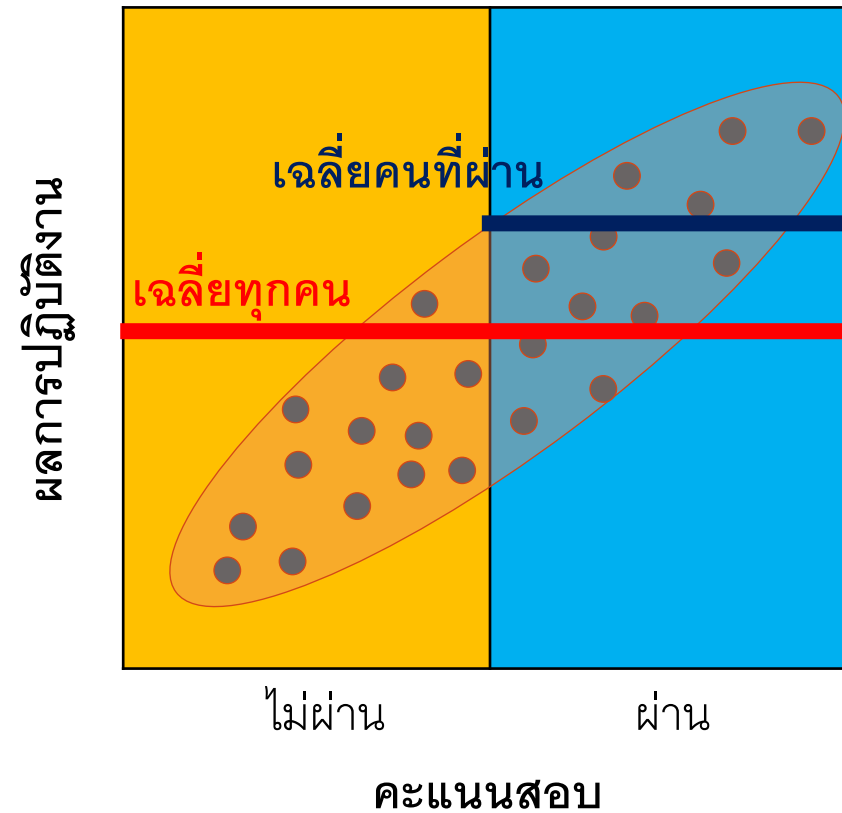
การวิเคราะห์หรือสรุปประโยชน์

- บางครั้งผู้ใช้แบบทดสอบไม่ได้แบ่งเกณฑ์ออกเป็นสำเร็จและล้มเหลว
- แต่ผู้ใช้แบบทดสอบต้องการทราบว่า เมื่อคัดเลือกแล้ว คะแนนเฉลี่ยของเกณฑ์จะเป็นอย่างไร



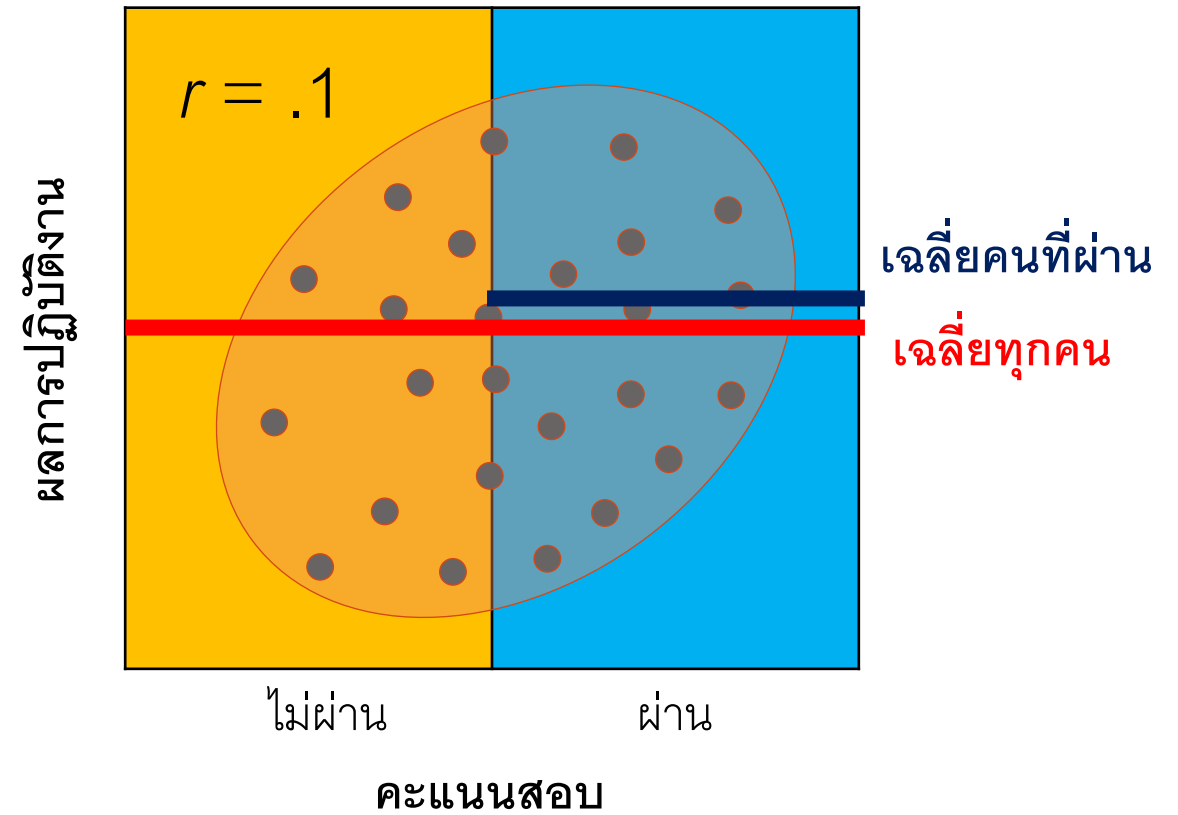
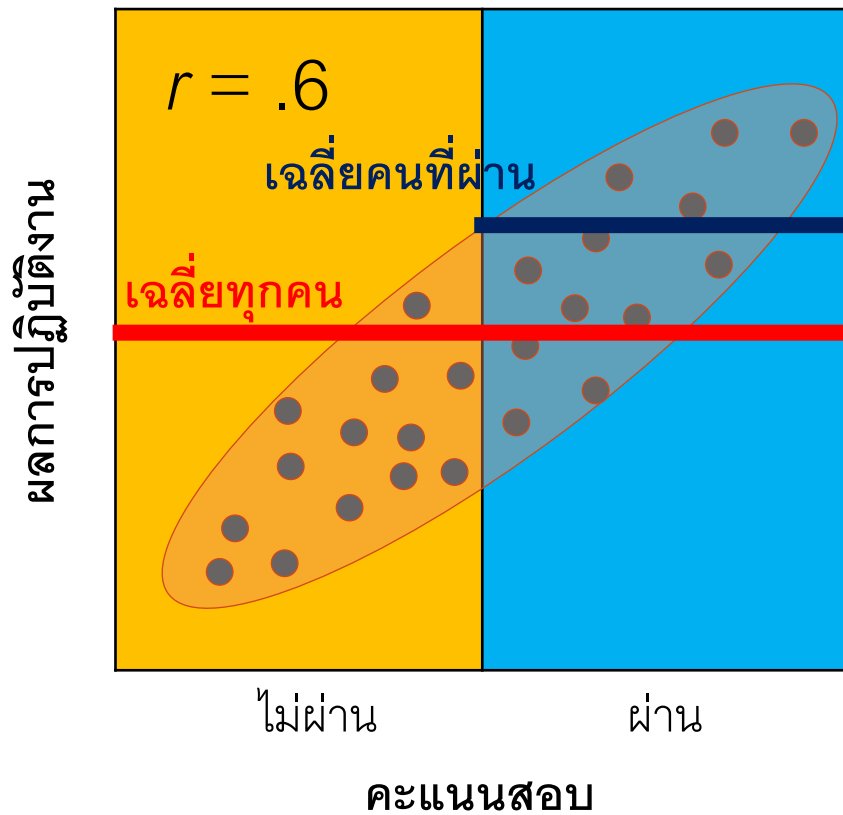
การวิเคราะห์หรือสรุปประโยชน์

- คะแนนเฉลี่ยคนที่ผ่านขึ้นอยู่กับ 2 ปัจจัย
 - ความสัมพันธ์ระหว่างเกณฑ์และตัวแปรทำนายสูง
 - ระดับจุดตัดของคะแนนสอบ



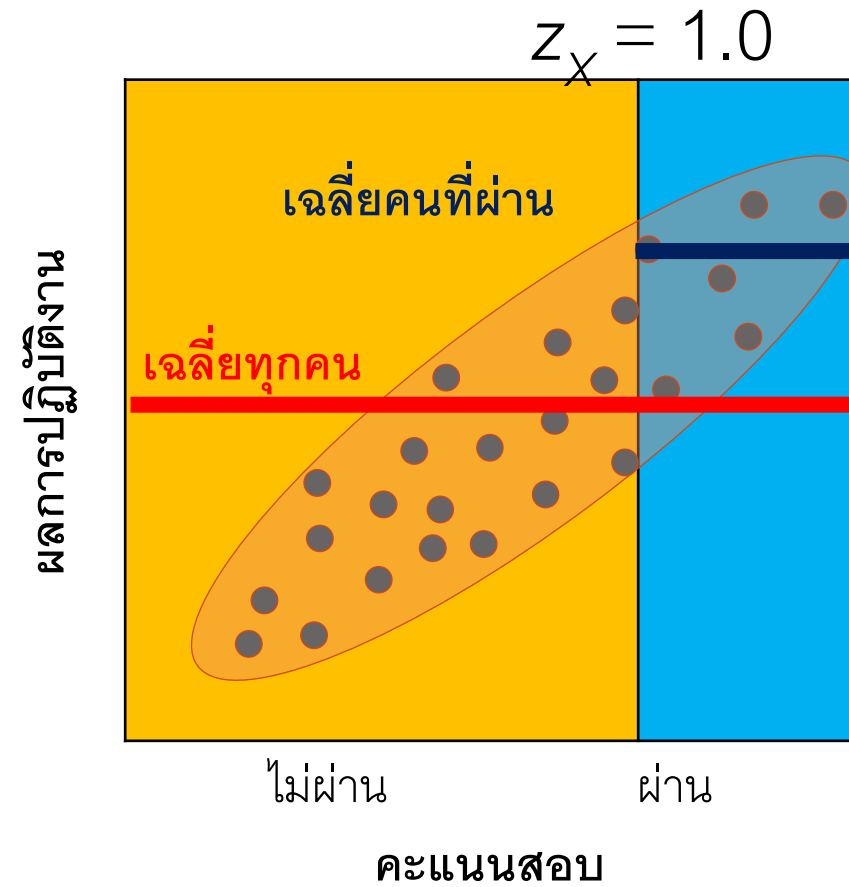
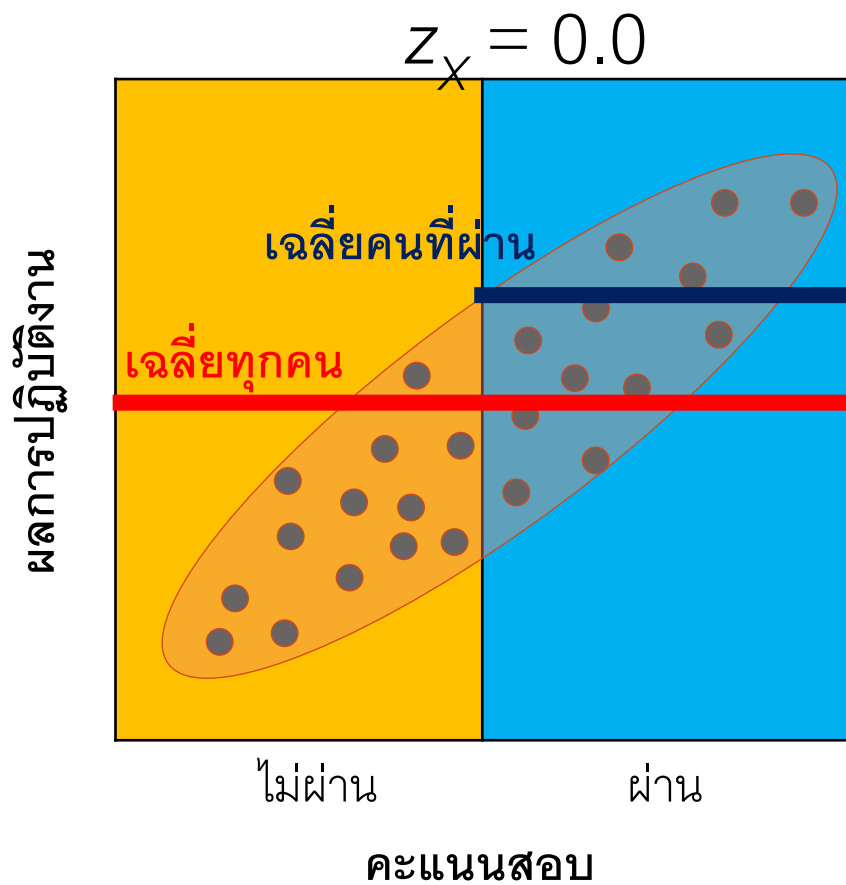
การวิเคราะห์หรือรถประโยชน์

- ความสัมพันธ์ระหว่างเกณฑ์และตัวแปรทำนาย



การวิเคราะห์หรือรูปประโยชน์

- จุดตัดของคะแนนสอบ



การวิเคราะห์อรรถประโยชน์

- หากมีข้อมูลดิบ นักวิเคราะห์สามารถหาค่าเฉลี่ยของผู้ถูกคัดเลือกได้เลย โดย
 - แปลงตัวแปรทำนาย (X) ให้เป็นตัวแปรใหม่ (D_X) มีสองกลุ่ม คือ ผ่านหรือไม่ผ่าน
 - หาค่าเฉลี่ยของตัวแปรเกณฑ์ (Y) ภายในกลุ่มที่ผ่านการคัดเลือก
- ในบางครั้ง ผู้วิจัยยังไม่มีข้อมูลดิบ แต่พอทราบข้อมูลของ
 - ความสัมพันธ์ระหว่างตัวแปรทำนายและเกณฑ์ (Predictive validity)
 - สัดส่วนการคัดเลือก (Selection ratio)
- หากข้อมูลการกระจายของตัวแปรทำนายและเกณฑ์เป็นโค้งปกติร่วม (Bivariate normality) แล้วสามารถใช้ตารางของ Naylor & Shines (1965) ในการหาค่าเฉลี่ยของตัวแปรเกณฑ์ในคนที่ถูกคัดเลือกได้

คำนวณเองโดยใช้ R

```
> ### Specify two values
> sr <- 0.7 # Selection Ratio
> r <- 0.7 # Predictive Validity
>
> ##### Calculation #####
> zx <- qnorm(1 - sr)
> lambda <- dnorm(zx)
> r * lambda / sr
[1] 0.3476926
```

ตามสูตรของ Myers (1993)

การวิเคราะห์อรรถประโยชน์

- นอกจากนี้ สามารถนำมาคำนวณอรรถประโยชน์ของการใช้แบบทดสอบในรูปแบบของเงินได้ หากผลการปฏิบัติงานคำนวณได้ในรูปของเงิน เช่น
- เซลล์โดยเฉลี่ยสามารถทำกำไรเบื้องต้นได้ 50,000 บาทต่อเดือน ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5,000 บาทต่อเดือน
- ตัวแปรทำนายมีค่าเฉลี่ยเท่ากับ 50 คะแนน ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10 คะแนน จะคัดเลือกผู้ที่มีคะแนนสูงกว่า 50 คะแนนเท่านั้น (สัดส่วนการคัดเลือก = 50%)
- ความสัมพันธ์ระหว่างตัวแปรทำนายและยอดขายเท่ากับ .4
- ค่าใช้จ่ายในการใช้แบบทดสอบเท่ากับ 500 บาท คัดเลือกมา 100 คนจาก 200 คน
- เซลล์ที่ถูกคัดเลือก มีอายุงานเฉลี่ย 6 เดือน

การวิเคราะห์หรือรถประโยชน์

1. หาค่าใช้จ่ายรวมทั้งหมด = $200 \times 500 = 100,000$ บาท
2. หาค่าเฉลี่ยของเกณฑ์ในรูปของคะแนนมาตรฐานจากผู้ที่ถูกคัดเลือกเท่ากับ 0.32

```
> ### specify two values
> sr <- 0.5 # selection Ratio
> r <- 0.4 # Predictive Validity
>
> ##### calculation #####
> zx <- qnorm(1 - sr)
> lambda <- dnorm(zx)
> r * lambda / sr
[1] 0.3191538
```

การวิเคราะห์อรรถประโยชน์

3. แปลงคะแนนมาตรฐานของเกณฑ์ให้อยู่ในรูปกำไรสุทธิ

$$0.32 \times 5000 + 50000 = 51600$$

4. กำไรสุทธิที่เพิ่มขึ้นต่อคน จากการใช้แบบทดสอบเท่ากับ 1,600 บาท
5. กำไรสุทธิสะสมตลอดอายุงาน 6 เดือน เท่ากับ 9,600 บาท
6. กำไรสุทธิของผู้ที่ถูกคัดเลือกทั้งหมด 100 คน เท่ากับ 960,000
7. ดังนั้น อรรถประโยชน์ของการใช้แบบทดสอบเท่ากับ $960,000 - 100,000 = 860,000$ บาท

การวิเคราะห์หรือรรถประโยชน์

- การคำนวณที่ผ่านมา คือ การคำนวณตามสูตร Brogden-Cronbach-Gleser

$$\text{Utility Gain} = (N_S)(T)(SD_Y)(r_{XY} \cdot Z_m) - (N)(C)$$

- ค่าสีแดง สามารถแทนได้ด้วยผลจากตารางของ Naylor-Shines
- จากตัวอย่าง

$$\text{Utility Gain} = (100)(6)(5000)(0.32) - (200)(500) = 860000$$

ข้อควรระวัง

- สูตรคำนวณทั้งหมดที่ผ่านมา เกิดจากสมมติฐานที่ว่า ผู้สมัครที่ผ่านเกณฑ์จะตอบรับการทำงานทั้งหมด
- ในความจริงแล้ว พนักงานไม่ได้ตอบรับเราทุกคน ดังนั้นสูตรที่คำนวณได้ อาจจะไม่ถูกต้องทั้งหมด
- อย่างไรก็ตาม จะใช้หรือไม่ใช้แบบทดสอบคัดเลือก ผู้สมัครก็ไม่ได้ตอบรับการทำงานทั้งหมด
- ดังนั้นผลที่ได้จากสูตรที่ผ่านมา น่าจะยังพอเชื่อถือได้

ข้อแนะนำในเชิงปฏิบัติ

- ตรวจสอบว่า ค่าเฉลี่ยเกณฑ์ของผู้ที่ถูกคัดเลือกแล้วหรือสัดส่วนคัดสำเร็จ (Hit ratio) ของการใช้แบบวัด ต้องสูงกว่าการไม่ใช้แบบวัดอย่างมีนัยสำคัญเชิงปฏิบัติ (Practical significance) เสมอ
- การสรรหา (Recruitment) ก็เป็นส่วนที่เอื้อให้เกิดอรรถประโยชน์มากขึ้น
 - เพิ่มจำนวนผู้สมัคร ทำให้ลดสัดส่วนคัดเลือก (Selection ratio)
 - เพิ่มสัดส่วนของผู้ประสบความสำเร็จ (Base rate) หากผู้สมัครถูกคัดเลือกมาจากกลุ่มคนที่พึงประสงค์

กระบวนการคัดเลือก

- การใช้จุดตัดเดียว
 - สร้างจุดตัดอิงกลุ่ม (Norm-referenced cut score)
 - สร้างจุดตัดอิงเกณฑ์ (Criterion-reference cut score)
- การใช้หลายแบบทดสอบ
 - จุดตัดหลายจุดตัด (Multiple cutoff)
 - จุดตัดต่อเนื่อง (Multiple hurdle)
 - การใช้สมการถดถอย (Multiple regression) ซึ่งมาจากแนวคิดที่ว่าจุดเด่นหนึ่งสามารถชดเชยจุดด้อยในอีกส่วนหนึ่งได้ หรือที่เรียกว่าโมเดลชดเชย (Compensatory model)

วิธีการสร้างจุดตัด

- วิธีของ Angoff – ให้ผู้เชี่ยวชาญตัดสินใจว่า
 - ผู้สมัครที่คาบเส้น จะตอบข้อนี้ถูกต้องได้สัดส่วนเท่าไร
- วิธีกลุ่มที่แตกต่าง
 - วิเคราะห์ผ่านฮิสโทแกรม
 - ใช้การวิเคราะห์ด้วย Logistic regression