

การวิเคราะห์ทางสถิติสำหรับข้อมูลที่ศึกษา ทั้งประชากร

อรุณ จิรวัดนกุล

ภาควิชาชีวสถิติและประชากรศาสตร์ คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น

การใช้สถิติในงานวิจัยเกือบทั้งหมดจะเป็นการใช้สถิติพรรณนาบรรยายลักษณะตัวอย่าง และใช้สถิติอนุมานสรุปลักษณะประชากรโดยการประมาณค่าพารามิเตอร์ สรุปผลการเปรียบเทียบ หรือหาความสัมพันธ์ของตัวแปร แต่บางกรณีที่นักวิจัยสามารถเก็บข้อมูลจากทุกหน่วยศึกษาของประชากร จะใช้สถิติอะไรในการสรุปผลการศึกษา^(๑)

สถิติ (ศาสตร์) ที่ใช้ในการสรุปลักษณะข้อมูล เช่น ค่าเฉลี่ย ถ้าคำนวณจากข้อมูลของ 'ตัวอย่าง' เรียกว่า **ค่าสถิติ** แต่ถ้าคำนวณจากข้อมูลทุกหน่วยศึกษาของ **'ประชากร'** เรียกว่า **ค่าพารามิเตอร์**

งานวิจัยที่ศึกษาด้วยตัวอย่าง จะนำข้อมูลจากตัวอย่างมาคำนวณค่าสถิติ เช่นค่าเฉลี่ย ของตัวอย่าง (\bar{x}) แล้วนำค่าเฉลี่ยของตัวอย่างที่ได้ไปอนุมานค่าเฉลี่ยของประชากร (μ) วิธีการทางสถิติที่ใช้อนุมานค่าพารามิเตอร์ของประชากรจะทำได้สองวิธีคือ ประมาณด้วยช่วงเชื่อมั่น เช่น นำ \bar{x} มาประมาณค่า 95% ช่วงเชื่อมั่นของ μ หรือการทดสอบสมมติฐาน ($H_0: \mu = 60, H_1: \mu < 60$) เช่นจาก \bar{x} ที่ได้นำไปทดสอบสมมติฐานได้ค่า p-value < 0.001 สรุปว่าค่าเฉลี่ยของประชากรมีค่าน้อยกว่า 60 อย่างมีนัยสำคัญที่ระดับความเชื่อมั่น 95% งานวิจัยที่เก็บข้อมูลจากทุกหน่วยศึกษาในประชากร เมื่อใช้สถิติสรุปลักษณะข้อมูลที่ศึกษาทั้งหมดค่าที่คำนวณได้จะเป็นค่าพารามิเตอร์ เช่น ค่าเฉลี่ย (μ) ค่าสัดส่วน (π) ค่าสหสัมพันธ์ (ρ) เป็นต้น

ดังนั้นการสรุปผล นักวิจัยไม่ต้องประมาณค่า หรือทดสอบสมมติฐานกับข้อมูลดังกล่าว เช่น ในการศึกษาเรื่องการสวมใส่หมวกนิรภัยในวิทยาลัยแห่งหนึ่ง ผู้วิจัยเก็บรวบรวมข้อมูลจากนักศึกษาทั้งวิทยาลัย การวิเคราะห์เพื่อตอบคำถามว่าสัดส่วนของนักศึกษาหญิงสวมใส่หมวกนิรภัยในขณะที่ขับขี่จักรยานยนต์ต่างจากนักศึกษาชายหรือไม่ โดยการนำค่าสัดส่วนของสวมใส่หมวกนิรภัยในขณะที่ขับขี่จักรยานยนต์ของนักศึกษาหญิงลบกับของนักศึกษาชายโดยตรง ผลต่างไม่เท่ากับศูนย์(0) แสดงว่าต่างกันจริง ถ้าผลการวิเคราะห์พบว่าสัดส่วนของสวมใส่หมวกนิรภัยในขณะที่ขับขี่จักรยานยนต์ของนักศึกษาหญิงเท่ากับ 0.28 ของนักศึกษาชายเท่ากับ 0.26 ผลต่างของสัดส่วน ($\pi_{หญิง} - \pi_{ชาย}$) เท่ากับ 0.02 จะสรุปว่าสัดส่วนของนักศึกษาหญิงสวมใส่หมวกนิรภัยมากกว่านักศึกษาชาย 0.02 หรือร้อยละ 2 โดยไม่ต้องทดสอบสมมติฐาน (p-value) หรือหาค่า 95% ช่วงเชื่อมั่น

การแปลผลนอกจากจะระบุว่าต่างกันแล้วควรพิจารณาขนาดความต่างที่พบด้วยว่าขนาดความต่างมีความสำคัญหรือไม่ ในกรณีนี้เป็นปัญหาสาธารณสุข จะพิจารณาว่าความต่างที่พบร้อยละ 2 มีความสำคัญหรือไม่ (public health meaningful) โดยพิจารณาจากการนำข้อมูลนี้ไปใช้งาน เช่นถ้าใช้ข้อมูลนี้ไปพิจารณาว่าการรณรงค์สวมใส่หมวกนิรภัยในขณะที่ขับขี่จักรยานยนต์ให้กับนักศึกษาหญิงและชายควรใช้วิธีต่างกันหรือไม่ ถ้าผู้จัดรณรงค์เห็นว่าอัตราการสวมใส่ที่ต่างกันร้อยละ 2

ตารางที่ 1 สถิติที่ใช้ในการวิเคราะห์ของข้อมูลวิจัย

การวิเคราะห์	ศึกษาจากตัวอย่าง		ศึกษาทั้งประชากร ค่าพารามิเตอร์
	ค่าสถิติ	ประมาณค่า พารามิเตอร์	
ค่าเฉลี่ย	\bar{x}	95% CI of μ	μ
ค่าสัดส่วน	p	95% CI of π	π
เปรียบเทียบค่าเฉลี่ย	$\bar{x}^1 - \bar{x}^2$	Test hypothesis (p value) or 95% CI of $\mu_1 - \mu_2$	$\mu_1 - \mu_2$
เปรียบเทียบค่าสัดส่วน	$p_1 - p_2$	Test hypothesis (p value) or 95% CI of $\pi_1 - \pi_2$	$\pi_1 - \pi_2$
ความสัมพันธ์ตัวแปรต่อเนื่องสองตัว	r	Test hypothesis (p value) or 95% CI of ρ	ρ
	$y = a + bx$	Confidence bands	$Y = \alpha + \beta x$
	RR	95% CI of RR	RR
ความสัมพันธ์ ตัวแปรกลุ่มสองตัว	Logistic regression	Test hypothesis (p value) or 95% CI of RR	Logistic regression

ไม่มีผลต้องกำหนดให้มีวิธีการรณรงค์ที่ต่างกัน จะสรุปผลการศึกษาว่าอัตราการสวมใส่หมวกนิรภัยของนักศึกษาหญิงมากกว่านักศึกษาชายร้อยละ 2 แต่ขนาดความต่างไม่มีความสำคัญที่จะทำให้ต้องเปลี่ยนแปลงวิธีการทำงานในการแก้ไขปัญหาด้านสาธารณสุข

ในการวิเคราะห์หาความสัมพันธ์ระหว่างตัวแปรสามารถใช้สถิติที่วิเคราะห์หาความสัมพันธ์ที่ใช้กับการศึกษาจากตัวอย่าง เช่น สัมประสิทธิ์สหสัมพันธ์ สมการถดถอย Relative Risk (RR) สมการถดถอยลอจิสติก ผลการวิเคราะห์ที่ได้จะสามารถใช้ระบุความสัมพันธ์ในประชากรได้โดยตรงโดยไม่ต้องทดสอบสมมติฐาน เช่น คำนวณค่าสหสัมพันธ์ $\rho = 0.8$ สามารถสรุปได้เลยว่าตัวแปรทั้งสองมีความสัมพันธ์ในระดับสูง

ตารางที่ 1 แสดงสถิติที่ใช้ในการวิเคราะห์ของข้อมูลวิจัยที่ศึกษาจากตัวอย่าง และศึกษาทั้งประชากร

ในการหาความสัมพันธ์ของตัวแปรกลุ่มสองตัวจะใช้ Odds Ratio (OR) ไม่ได้ เพราะ OR ใช้ในกรณีที่ไม่ทราบค่าประชากร เมื่อศึกษาทั้งประชากรจะต้องใช้ RR ในการแสดงขนาดความสัมพันธ์

โดยปกติในการศึกษาทั้งประชากร คำถามวิจัยจะเป็นการหาขนาดของปัญหา และมุ่งที่จะหากลุ่มย่อยที่

เกิดปัญหา เพื่อจะได้หาวิธีการแก้ไขปัญหานั้น โดยจะไม่ต้องการดูความสัมพันธ์ เช่น การศึกษาสมรรถนะการทำงานของข้าราชการสังกัดกระทรวงสาธารณสุขภายในจังหวัด ผู้วิจัยส่งแบบสอบถามไปเก็บข้อมูลจากเจ้าหน้าที่ทุกคนที่ทำงานอยู่ในจังหวัด (ทั้งประชากร) การวิเคราะห์ข้อมูลจะทำให้ทราบว่า มีผู้ที่มีสมรรถนะการทำงานต่ำกว่าที่กำหนดมีมากน้อยเท่าไร และเป็นผู้ที่สังกัดหน่วยงานใด หรือมีลักษณะอย่างไร เพื่อจะได้จัดกิจกรรมในการแก้ไขปัญหาดังกล่าวโดยตรง มากกว่าจะตั้งคำถามว่าสมรรถนะการทำงานมีความสัมพันธ์กับเพศหรือไม่

สรุป ในการใช้สถิติสรุปลักษณะของข้อมูลที่ได้จากการศึกษาทั้งประชากร นักวิจัยสามารถคำนวณค่าพารามิเตอร์ต่าง ๆ ได้โดยตรง ไม่ต้องใช้การประมาณค่าหรือทดสอบสมมติฐาน ในการสรุปผลการศึกษาของประชากร

เอกสารอ้างอิง

1. อรุณ จิรวัดน์กุล. สถิติทางวิทยาศาสตร์สุขภาพเพื่อการวิจัยที่ใช้ในงานวิจัยทางวิทยาศาสตร์สุขภาพ. พิมพ์ครั้งที่ 1. กรุงเทพมหานคร: วิทยพัฒน์; 2552.