

หนังสืออ่าน (Book Reviews)

How to Lie With Statistics: Book Review

น.ท.พงษ์เพชร คงพ่วง

ข้อมูลหนังสือ

ชื่อ: How to lie with statistics
ผู้แต่ง: Darrell Huff
ผู้วาดภาพประกอบ: Irving Geis
สำนักพิมพ์: W. W. Norton & Company
รูปแบบ: ป็อกเก็ตบุ๊กปกอ่อน
ความหนา: 142 หน้า
ปีที่พิมพ์ครั้งแรก: ค.ศ. 1954

ตำนานของข้อมูลเกี่ยวข้องกับค่าเฉลี่ย ระดับความสัมพันธ์ แนวโน้ม และการนำเสนอข้อมูลด้วยกราฟ ... คุณควรอ่านหนังสือเล่มนี้ ถ้าคุณอยากรู้ทันผู้ใช้ตัวเลขและวิธีการทางสถิติเป็นเครื่องมือสำหรับหลอกลวงผู้อ่าน ... คุณต้องอ่านหนังสือเล่มนี้

ดาร์เรลล์ ฮัฟ (Darrell Huff : ชาวอเมริกัน มีชีวิตระหว่าง ค.ศ. 1913 ถึง 2001) เป็นนักสื่อสารมวลชนที่แต่งหนังสือประเภท "How to ..." จำนวนมาก เขาไม่ได้เป็นนักสถิติศาสตร์ แต่หนังสือ How to Lie with Statistics ของเขาซึ่งได้รับการตีพิมพ์ครั้งแรกใน ค.ศ. 1954 มีการตีพิมพ์กว่า 30 ครั้ง ยอดพิมพ์ในภาคภาษาอังกฤษสูงถึง 1.5 ล้านฉบับ และได้รับการแปลในอีกหลาย ๆ ภาษา มีผู้กำลังแปลหนังสือเล่มนี้เป็นไทยอยู่ ฉบับล่าสุดตีพิมพ์ใน ค.ศ. 1991 โดยสำนักพิมพ์ Penguin ประเทศไทย

J.M. Steele เขียนบทความเรื่อง "Darrell Huff and Fifty Years of How to Lie with Statistics" ในวารสาร Statistical Science ฉบับที่ 20(3) ค.ศ. 2005 ส่วนหนึ่งของบทความกล่าวว่า "ในช่วงห้าสิบปีที่ผ่านมา มีผู้ซื้อหนังสือเล่มนี้เป็นจำนวนมากกว่าตำราทางสถิติใดๆ" นักวิจารณ์หนังสือบางท่านกระเชาะว่าชื่อของหนังสือเล่มนี้ ที่ถูกต้องควรจะมีชื่อเป็น "How not to get lied to with statistics" มากกว่า

คำปรารภในหนังสือ How to Lie with Statistics ฉบับพิมพ์ครั้งที่ 31 เขียนว่า หนังสือเล่มนี้เป็นการแนะนำวิชาสถิติสำหรับผู้อ่านทั่วไป เป็นงานเขียนแนวเบาๆ ไม่ใช่ศัพท์ทางคณิตศาสตร์ เพื่อให้เข้าใจสาระของสถิติ และเพื่อให้ทราบความไร้สาระของสถิติเมื่อได้พบเห็น มันเป็นหนังสือทางสถิติที่ขายดีที่สุดเท่าที่มีมา

H.G. Wells นักเขียนนิยายวิทยาศาสตร์ผู้มีชื่อเสียงได้คาดการณ์ว่า "สักวันหนึ่งวิธีคิดทางสถิติจะกลายเป็นสิ่งจำเป็นสำหรับประชาชนที่มีคุณภาพ เช่นเดียวกับความสามารถในการอ่านและเขียน" ดังนั้นผู้ที่ต้องพิมพ์ตีพิมพ์เอกสารหรือข้อมูลตัวเลข เพื่อนำมาหาข้อสรุป นำมาทำนายแนวโน้ม และนำเสนอในรูปแบบกราฟ รวมทั้งผู้ที่ต้อง "อ่าน" เอกสารวิจัย บทความ หรือแม้แต่โฆษณา ที่อ้างข้อมูลทางสถิติ จึงควรทราบความคลาดเคลื่อนของการใช้สถิติในการแปลผลข้อมูล ไม่ว่าจะโดยตั้งใจหรือไม่ก็ตาม ซึ่งหนังสือเล่มนี้จะช่วยให้ผู้อ่านทำความเข้าใจได้อย่าง

Darrell Huff เขียนเริ่มต้นในบทนำของหนังสือเล่มนี้ว่า "พ่อตาของผมบ่นว่า แถบนี้ก็มีอาชญากรรมเกิดขึ้นเยอะจัง โดยดูจากข่าวที่ลงในหนังสือพิมพ์ แสดงว่าพ่อตาของผมหาข้อสรุปแบบสถิติอย่างไม่เป็นทางการ ข้อสรุปนี้มาจากตัวอย่างที่มีอคติ ซึ่งขัดแย้งการเกิดอาชญากรรมจากพื้นที่ในหนังสือพิมพ์ที่ลงข่าวอาชญากรรม" นับเป็นความแสบคายยิ่งที่นำเสนอปัญหาอคติของการสุ่มตัวอย่างทางสถิติจากเหตุการณ์จริงที่เรพบเห็นกัน เป็นจุดเริ่มต้นให้ผู้อ่านอยากติดตาม และขมวดท้ายในบทนำว่า "หนังสือเล่มนี้ จะเป็นส่วนนำโด่งเด่นแรกที่บอกวิธีการใช้สถิติเพื่อหลอกลวง ทำให้เราดูเหมือนกับเป็นคู่มือของคนที่ฉ้อโกง แต่ขอให้คิดเสมอว่าผมเป็นโจรที่วางมือแล้ว หันมาเปิดหลักสูตรสอนวิธีฉ้อโกงเขา และขอยกเขา ซึ่งพวกโจรรู้เรื่องนี้อยู่แล้ว แต่สำหรับคนดีนั้น จำเป็นต้องเรียนรู้ไว้เพื่อปกป้องตนเอง"

หนังสือเล่มนี้แบ่งออกเป็น 10 บท มุ่งเน้นถึงวิธีใช้สถิติในการหลอกลวง เช่น การเลือกสุ่มข้อมูลอย่างใจ การบิดเบือนโดยการนำเสนอด้วยกราฟ เป็นต้น เนื้อหาของทั้ง 10 บท มีดังนี้

- บทที่ 1 - กลุ่มตัวอย่างที่มีอคติแฝง (The Sample with the Built-in Bias): อธิบายว่าผลสำรวจหรือโพล มักมีอคติออกไปในทางเข้าข้างตัวเอง เช่น ไม่ต้องการเปลี่ยนแปลง มีรายได้สูงขึ้น หรือ การศึกษาสูงขึ้น
- บทที่ 2 - ค่าเฉลี่ยที่คัดเลือกมาอย่างดี (The Well-Chosen Average): อธิบายการใช้คำว่าค่าเฉลี่ย (mean, median, หรือ mode) อย่างผิดๆ และถึงแม้จะใช้ในโอกาสที่เหมาะสม ก็มักสร้างความเข้าใจผิด
- บทที่ 3 - ตัวเลขเล็กๆ ที่ไม่ปรากฏ (The Little Figures That Are Not There): เล่าถึงตัวเลขที่ถูกซ่อนอยู่ เช่น "วิทยาศาสตร์พิสูจน์ได้ว่าการโยนเหรียญจะออก หัว ได้มากถึง 80 เปอร์เซ็นต์" เพราะใช้จำนวนตัวอย่างน้อยเกินไป
- บทที่ 4 - ความวุ่นวายของสิ่งที่ไม่มีความหมายในทางปฏิบัติ (Much Ado about Practically Nothing): ไม่ได้รายงานค่า บวก/ลบ หรือ ความน่าจะเป็นของความคลาดเคลื่อน ทำให้ความแตกต่างที่แสดง ไม่มีความหมาย
- บทที่ 5 - กราฟเตอะตา (The Gee-Whiz Graph): การบิดเบือนกราฟ โดยเลือก scales และ origins ตัวอย่างเช่นกราฟโลกร้อนของนาย Al Gore
- บทที่ 6 - รูปมิติเดียว (The One-Dimensional Picture): การบิดเบือนกราฟรูปภาพ เช่น ความสูงเทียบกับพื้นที่
- บทที่ 7 - แสดงตัวเลขไม่หมด (The Semiattached Figure): ข้อมูลที่รายงานอาจเป็นเรื่องอื่น เช่น ยานวันปากนี้สามารถฆ่าเชื้อโรคได้ (แต่อาจไม่มายังเชื้อโรคในปาก)
- บทที่ 8 - สิ่งที่เกิดขึ้นหลังจากนั้น (Post Hoc Rides Again): ว่าด้วยเหตุและผล หรือสาเหตุร่วม หรือความบังเอิญ หรืออย่างอื่น? เช่น "พบว่ามีความสัมพันธ์อย่างมากระหว่าง รายได้ของสตรีคุณใน Massachusetts กับราคาของเหล้ารัมใน Havana" "มีการแต่งงานมากในเดือนมิถุนายน และมีการฆาตกรรมมากในเดือนมิถุนายน อันไหนเป็นสาเหตุของอันไหน"
- บทที่ 9 - วิธีการนำสถิติมาใช้ (How to Statisticulate): การโกหกด้วยสถิติ เป็นความไม่ซื่อสัตย์หรือไร้ความสามารถ?
- บทที่ 10 - วิธีตอบกลับไปยังสถิติ (How to Talk Back to a Statistic): การจับโกหก (เป็นวัตถุประสงค์หลักของหนังสือเล่มนี้) โดยตั้งคำถามง่าย ๆ 5 ข้อว่า ใครเป็นผู้พูด (มีอคติหรือไม่ ถ้าเป็นผู้เชี่ยวชาญ ให้สอบถามว่าเห็นด้วยกับข้อสรุปนี้หรือไม่), เขาทราบได้อย่างไร (จำนวนข้อมูลมีมากเพียงพอและน่าเชื่อถือหรือไม่), มี (ข้อมูล) อะไรที่ขาดไป, มีใครเปลี่ยนหัวข้อเรื่องหรือไม่ (ดูว่ามีการสลับเปลี่ยนระหว่างข้อมูลดิบกับข้อสรุปหรือไม่), มันดูมีเหตุผลหรือไม่ (มีการใช้ตัวเลขหลอกลวงหรือไม่ ถ้าดูที่จะถูกหลอก ให้หาความเชื่อมโยงระหว่างข้อมูลกับข้อสรุป)

โดยรวม หนังสือเล่มนี้หนาเพียง 140 หน้า เขียนด้วยภาษาง่ายๆ ผู้เขียนตัวอย่างจำนวนมากที่ชาวอเมริกันคุ้นเคย และมีภาพประกอบที่ขบขัน แต่แฝงความหมาย ผู้อ่านจะได้ทราบความลึกซึ้งของเนื้อหาทางสถิติที่เรามักมองข้าม รวมทั้งทราบวิธีนำผลการวิเคราะห์ทางสถิติไปใช้งาน ซึ่งสิ่งเล็กๆ ที่บังคับไว้อาจใช้หลอกลวงผู้ชมจำนวนมากให้เห็นคล้อยตามได้ เนื้อหาในหนังสือทำให้เราเห็นความสำคัญของความถูกต้องและน่าเชื่อถือของข้อมูล การแปลผลอย่างไม่มั่ว และ การนำเสนอที่ไม่บิดเบือน และชี้ให้เห็นว่าปัญหาที่แท้จริงไม่ได้อยู่ที่ตัวสถิติ แต่อยู่ที่การนำมาใช้