

Giant with Feet of Clay: Leo Strauss's Criticism
And the Insecurity of Quantitative Research's Body of Knowledge?**

Ekkapollanut Nuttapattanun*

Abstract

Humans are ordinarily afraid of uncertainty in life. They therefore hope to be able to predict future happenings so as to be prepared to deal with the problems which might arise. It is such that they could ensure a good future. Accordingly, "science" is called into service as it has a strict methodology and, it is believed, is able to foresee and predict natural happenings of the future. The scientific method thus became popular. However, when it came to be applied in a variety of disciplines in the form of quantitative research, with the use of statistics in testing hypotheses, the much praised principles of "science", such as neutrality and objectivity, became blemished. This was when it became impossible for researchers, who after all are ordinary human beings, to completely avoid being biased in line with their own preferences. The scientific method and its practice thus became out of sync in the testing of hypotheses and consequently in arriving at research results. The most that could be done was to arrive at "statistical probability", in fact prediction on the basis of uncertainty. However, present day technological advances have lessened the problem. It is therefore time to reconsider Leo Strauss's defamatory criticism that modern body of knowledge is like a giant with feet of clay.

Keywords: **Leo Strauss, scientism, objectivity**

*Lecturer at Department of Social Sciences, Faculty of Arts, Silpakorn University, Sanam Chandra Palace Campus, 6 Rachamakka Nai Rd., Phra Pathom Chedi, Muang Nakhon Pathom, Nakhon Pathom 73000.

Email: Nuttapattanun_E@su.ac.th and ekkapollanut@gmail.com

**Received September 3, 2017; Accepted December 4, 2017

วารสารสังคมศาสตร์

คณะรัฐศาสตร์

ปีที่ 48 ฉบับที่ 1 (2561):....

จุฬาลงกรณ์มหาวิทยาลัย

<http://www.library.polsci.chula.ac.th/journal2>

ยักรัษที่เ้าทำด้วยดินเหนียว: บทวิจารณ์ของ Leo Strauss
กับควมม่งอน่งนขององค้ควมมรู้ในการวิจัยเชิงปริมาณ?*

เอกพลณัฐ ญัฐพัทธนันท์*

บทคัดย่อ

มนุษย์โดยทั่วไปต่างหวาดกลัวความไม่แน่นอนในชีวิต พวกเขาจึงคาดหวังที่จะสามารถคาดการณ์สิ่งที่เกิดขึ้นในอนาคตได้ และสามารถเตรียมพร้อมรับมือปัญหาที่เกิดขึ้นล่วงหน้าได้ทันทั่วๆ เพื่อจะได้ทำให้อนาคตของพวกเขากลายเป็นสิ่งที่ดี ซึ่งแนวทางดังกล่าวทำให้ “วิทยาศาสตร์” สามารถตอบโจทย์ความต้องการดังกล่าวได้ เพราะเป็นแนวทางการศึกษาที่มีกฎระเบียบเคร่งครัด สามารถคาดการณ์และทำนายปรากฏการณ์ที่จะเกิดขึ้นตามธรรมชาติในอนาคตได้ แต่ถึงแม้ว่าการทดลองทางวิทยาศาสตร์จำนวนมากจะสร้างความศรัทธาให้กับมนุษยชาติ จนกระทั่งมีการนำแนวทางการศึกษาแบบวิทยาศาสตร์เข้าไปประยุกต์ใช้ในการศึกษาสาขาวิชาต่าง ๆ อย่างกว้างขวาง ในลักษณะของการทำวิจัยเชิงปริมาณ โดยมีการใช้เทคนิคทางสถิติเป็นเครื่องมือในการทดสอบสมมติฐานของการวิจัย แต่แล้วหลักการแบบวิทยาศาสตร์ที่สวดยหรู ไม่ว่าจะเป็นเรื่อง “ความเป็นกลาง” (neutrality) และเรื่อง “ความเป็นภววิสัย” (objectivity) ก็ต้องถูกทำให้มัวหมองไป เมื่อการศึกษาดังกล่าวหลีกเลี่ยงไม่พ้นอคติ หรืออารมณ์ความรู้สึกของผู้วิจัยที่เป็นปฏุชนผู้สร้างกฎเกณฑ์ต่าง ๆ ขึ้นมา ก่อให้เกิดความไม่ลงรอยกันทั้งในทางหลักการและในทางปฏิบัติเกี่ยวกับการทดสอบสมมติฐานและการสร้างข้อสรุปผลการวิจัย เพื่อสร้างองค์ความรู้ที่ทำได้อย่างมากเพียงแค่การอาศัย “ความน่าจะเป็นทางสถิติ” ซึ่งก็คือ การคาดการณ์ว่าปรากฏการณ์ที่ศึกษาจะเกิดขึ้นจริงบนฐานของความไม่แน่นอนนั่นเอง แต่ด้วยความก้าวหน้าของเทคโนโลยีในปัจจุบันได้ทำให้ปัญหาดังกล่าวลดลง ค่าปราคมาศของ ลีโอ สเตราส์ ที่ว่าองค์ควมมรู้สมัยใหม่เปรียบเสมือนยักรัษที่เ้าเป็นดินเหนียวจึงต้องถูกนำกลับมาพิจารณาใหม่อีกครั้ง

คำสำคัญ: ลีโอ สเตราส์, วิทยาศาสตร์นิยม, ความเป็นกลาง

*อาจารย์ ดร. ประจักษ์วิชาสังคมศาสตร์ คณะอักษรศาสตร์ มหาวิทยาลัยศิลปากร วิทยาเขตพระราชวังสนามจันทร์ เลขที่ 6 ถนนราชมรรคาใน ตำบลพระปฐมเจดีย์ อำเภอเมืองฯ จังหวัดนครปฐม 73000 อีเมล Nuttapattanun_E@su.ac.th และ ekkapollanut@gmail.com

**ได้รับบทความ 3 กันยายน 2560; อนุมัติให้จัดพิมพ์ 4 ธันวาคม 2560

1. บทนำ: การใช้เทคนิคทางสถิติกับหายนะของงานวิจัยเชิงปริมาณ

เมื่อมนุษย์ทุกคนต่างมีจุดหมายของชีวิตอยู่ที่การมีชีวิตที่ดีและมีความสุข หนทางหนึ่งที่จะทำให้ความคาดหวังดังกล่าวเป็นจริงได้ คือ จะต้องการสร้างองค์ความรู้ให้เพิ่มมากขึ้นเรื่อย ๆ เพื่อไขความลับที่ดำรงอยู่ในธรรมชาติ ด้วยความเชื่อว่ามีมนุษย์มีศักยภาพเพียงพอที่จะสร้างความก้าวหน้าขึ้นอย่างไม่มีที่สิ้นสุด เพียงแค่สถานการณ์แวดล้อมของเขานั้นพร้อมเอื้ออำนวย (ถูกพื้นที่และเวลา) ให้มนุษยชาติได้สะสมความรู้ นั่น ดังเช่น การมีเครื่องมือเก็บรวบรวมและวิเคราะห์ข้อมูลที่ทันสมัย การมีทรัพยากรการศึกษาที่มากเพียงพอ ฯลฯ สิ่งเหล่านี้จะช่วยให้มนุษย์สามารถดึงศักยภาพที่มีออกมาใช้ได้อย่างเต็มที่มากขึ้น

การศึกษาโลกทางสังคมและโลกทางวัตถุเป็นพื้นที่สำคัญของการผลิตองค์ความรู้ใหม่ ๆ ที่ได้รับความสนใจหลากหลายแง่มุม ดังเช่น พฤติกรรมที่มนุษย์แสดงออก การผลิตยารักษาโรค การสร้างยานพาหนะที่ปลอดภัยและเดินทางได้อย่างรวดเร็ว การก่อสร้างอาคารที่คงทนต่อการเปลี่ยนแปลงของสภาพดิน ฟ้า และอากาศ ฯลฯ แต่การศึกษาพื้นที่ดังกล่าวโดยตรงย่อมเป็นไปได้ยาก โดยเฉพาะในอดีตที่เครื่องมือและทรัพยากรที่ใช้ในการศึกษายังไม่ทันสมัย การศึกษาที่จำเป็นต้องรวบรวมข้อมูลจากประชากร (population) ทั้งหมดอย่างครบถ้วนมีความเป็นไปได้น้อยมาก หรือถ้าหากจะทำเช่นนั้นจริง ก็จะต้องใช้ทรัพยากรจำนวนมหาศาลทำให้สิ้นเปลืองทั้งเวลาและเงินทอง

การวิจัยเชิงปริมาณที่อาศัยการสุ่มกลุ่มตัวอย่าง (sampling) จากประชากรของสิ่งที่สนใจศึกษา ด้วยเทคนิคการอนุมานทางสถิติ (inferential statistics) จึงถูกนำมาใช้เพื่อลดต้นทุนและทำให้เกิดความเป็นไปได้ที่จะได้ภาพรวมของสิ่งที่สนใจศึกษา โดยไม่ต้องเสียเวลาเก็บรวบรวมข้อมูลจากประชากรของสิ่งที่สนใจศึกษาทั้งหมด แต่หนทางลัดดังกล่าวก็ต้องแลกด้วยความเสี่ยงที่จะได้ข้อมูลที่ไม่ตรงกับความเป็นจริงเช่นกัน เพราะอาจจะเกิดปัญหาว่ากลุ่มตัวอย่างที่ศึกษาไม่ได้มีความเป็นตัวแทนของประชากรที่ผู้ศึกษาสนใจอย่างแท้จริง

ความต่างของจำนวนทั้งสองกลุ่มกลายเป็นช่องว่างที่นำมาสู่ข้อผิดพลาดขององค์ความรู้ที่สร้างขึ้นใหม่ ดังที่ผู้เขียนจะแสดงให้เห็นในหัวข้อต่าง ๆ ของบทความชิ้นนี้ ไฉนเลยไปตั้งแต่ความก้าวหน้าอย่างรวดเร็วขององค์ความรู้แบบวิทยาศาสตร์จนนำมาสู่การผูกขาดวิธีการสร้างองค์ความรู้ในสาขาวิชาต่าง ๆ จากการทำวิจัยเชิงปริมาณโดยอาศัยเทคนิคทางสถิติ ก่อนที่หนทางดังกล่าวจะเปิดเผยให้เห็นความผิดพลาดและคลาดเคลื่อนของข้อค้นพบที่ได้ จากการแอบอ้างการเข้าถึงความรู้ที่แท้จริงบนฐานของความไม่แน่นอน ทำให้องค์ความรู้สาขาวิชาต่าง ๆ ที่อิงแอบการศึกษาแบบวิทยาศาสตร์ต้องสั่นคลอนและเสื่อมถอยลง

ในบทความนี้ ผู้เขียนจะพยายามขยายให้เห็นว่าปัญหาขององค์ความรู้ในการวิจัยเชิงปริมาณเกิดขึ้นเพราะอะไร เริ่มจากการกล่าวถึงข้อวิพากษ์วิจารณ์ของ Leo Strauss ที่มีต่อองค์ความรู้แบบวิทยาศาสตร์สมัยใหม่ว่าจะนำมาสู่วิกฤตของอารยธรรมตะวันตก ก่อนที่จะนำเสนอพัฒนาการของการนำวิธีทดสอบนัยสำคัญทางสถิติมาใช้ในการพิสูจน์สมมติฐานการวิจัยของนักวิจัยเชิงปริมาณ หลังจากนั้นจึง

กล่าวถึงสภาพปัญหาของการวิจัยเชิงปริมาณ ทั้งที่เกิดขึ้นในหลักการและในทางปฏิบัติ แล้วจึงกล่าวถึง บทสรุปถึงความเป็นไปได้ของคำวิพากษ์วิจารณ์ของ Strauss ที่มีต่อองค์ความรู้ในการวิจัยเชิงปริมาณ

II. ยักษ์ที่เท้าทำด้วยดินเหนียวกับความหลงระเห็จในข้อค้นพบของนักวิทยาศาสตร์

นับตั้งแต่เกิดการฟื้นฟูศิลปวิทยาการในช่วงคริสต์ศตวรรษที่ 14 ถึง 16 ที่นำมาสู่การปฏิวัติ วิทยาศาสตร์ ได้มีข้อค้นพบจากงานวิจัยทางวิทยาศาสตร์จำนวนมากที่น่าทึ่งและสั่นสะเทือนวงการใน สาขาวิชาต่าง ๆ ดังเช่น สำนักข่าวออนไลน์ของไทยแห่งหนึ่งซึ่งชื่อว่า The Matter ได้พาดหัวข่าวว่า “20 การ ค้นพบทางวิทยาศาสตร์ที่น่าจดจำที่สุดแห่งปี 2016” (Thanes Rattanukul 2016) โดยจัดอันดับงานวิจัยที่ ขึ้นขอบและสร้างความหวังให้มนุษยชาติ จำนวน 20 ชิ้น ดังตัวอย่างเช่น ‘ปลาคลิลีฟิช’ กลายพันธุ์ตัวเอง เพื่อให้ทนสารเคมี 8,000 เท่า, พบทางไดโนเสาร์ขนาด ๕ ในอำพัน เปิดพื้นที่ความรู้ใหม่ระหว่างไดโนเสาร์ และนก, พบ ‘แบคทีเรีย’ เก่าแก่ที่สุดในโลก 5 พันล้านปี ก่อนที่โลกจะมีออกซิเจนซะอีก, แผ่นดินไหวใน นิวซีแลนด์ ยกกันทะเลขึ้นถึง 2 เมตร ภายในไม่กี่วินาที, ทีมวิจัยจีนใช้เทคนิคสร้าง ‘แกะลายจุด’ ด้วยการ ดัดแปลงพันธุกรรม, วิทยาศาสตร์เปลี่ยน CO₂ ให้กลายเป็น ‘หิน’ ใช้เวลาเพียงไม่กี่เดือน ฯลฯ

ตัวอย่างข้อค้นพบทางวิทยาศาสตร์ดังกล่าวแสดงให้เห็นถึงความสามารถของมนุษย์ที่พัฒนา ศักยภาพได้มากขึ้นเรื่อย ๆ จนกระทั่งกระทำสิ่งต่าง ๆ ที่อยู่นอกเหนือสิ่งที่ธรรมชาติได้สร้างขึ้นมา หรือ สามารถค้นพบสิ่งลึกลับและแปลกใหม่ในธรรมชาติที่มนุษย์อาจไม่เคยพบเห็นมาก่อน ซึ่งช่วยเพิ่มความ น่าเชื่อถือและศรัทธาในแนวทางการศึกษาแบบวิทยาศาสตร์ให้เพิ่มสูงขึ้นด้วย แต่ขณะเดียวกันก็มีตัวอย่าง ของข้อค้นพบจากงานวิจัยทางวิทยาศาสตร์ที่แปลกประหลาด หรือแม้กระทั่งถูกมองว่าเป็นงานวิจัยที่ไม่มี ประโยชน์สำหรับบางคนเช่นกัน ดังเช่น สำนักข่าวไทยรัฐออนไลน์ ที่ได้พาดหัวข่าวว่า “10 ผลวิจัย วิทยาศาสตร์สุดแปลก ประจำปี 2013” (“10 Phon Wichai Witthayasat ...” 2013) โดยได้อ้างถึง ผลงานวิจัยทางวิทยาศาสตร์ เช่น กลิ่นมีทั้งหมด 10 กลิ่น และกลิ่นข้าวโพดคั่วเป็นหนึ่งในนั้น, การปรบมือ ติดต่อกันได้, ทากทะเลออกอวัยวะเพศใหม่ได้ 3 ครั้ง, กินพิริกไทยเสฉวนจะรู้สึกชาเหมือนเหมือนถูกตะ แฉกหนักรั่ว ๆ สุนัขแกว่งหางไปทางซ้ายหรือขวา มีความหมายต่างกัน เป็นต้น

เหตุผลที่ผู้เขียนนำตัวอย่างผลงานทางวิทยาศาสตร์ดังกล่าวมานำเสนอก็เพื่อแสดงให้เห็นว่า แท้ที่ จริงแล้ววิทยาศาสตร์ก็เป็นแนวทางการศึกษาอย่างหนึ่งที่ท่ามกลางการศึกษาที่หลากหลาย ถึงแม้จะสามารถ สร้างความน่าทึ่งและความแปลกประหลาดใจได้ แต่ก็ไม่ใช่ข้อค้นพบทุกอย่างของวิทยาศาสตร์ที่เป็นเช่นนั้น ยิ่งกว่านั้นข้อค้นพบที่ได้จากการศึกษาด้วยแนวทางแบบวิทยาศาสตร์ ยังสามารถนำมาสู่ข้อผิดพลาดได้ไม่ ต่างไปจากการศึกษาด้วยแนวทางอื่น ๆ ดังจะเห็นได้ว่ามีหนังสือและบทความวิชาการจำนวนมาก ได้ ออกมาได้แย่งความเข้าใจผิด ๆ ของคนที่เกิดขึ้นจากความเชื่อที่มีต่อข้อค้นพบทางวิทยาศาสตร์ โดยเฉพาะ อย่างยิ่งที่เกิดขึ้นจากการนำ “สถิติ” มาใช้เป็นเครื่องมืออย่างหนึ่งในการวิเคราะห์และสร้างข้อค้นพบ

ตัวอย่างของงานที่ออกมาแสดงให้เห็นถึงความผิดพลาดในการนำ “สถิติ” มาใช้ ซึ่งผู้เขียนมองว่า น่าสนใจมีดังนี้ เช่น ใน ค.ศ. 1954 Darrell Huff (1954) ได้ตีพิมพ์หนังสือชื่อ *How to Lie with Statistics* โดยเปิดเผยให้เห็นถึงการนำสถิติมาใช้ปั่นหัวคนในชีวิตประจำวัน ด้วยการสร้างความเชื่อผิด ๆ ที่ “ดูเหมือนจะมี เหตุมีผล” ดังที่ในบทนำของหนังสือเล่มนี้ Huff ได้ยกตัวอย่างครอบครัวหนึ่งในสหรัฐอเมริกาที่มีพ่อเลี้ยงที่เพิ่ง ย้ายจาก Iowa มายัง California ได้พูดกับลูกเลี้ยงขึ้นมาระหว่างอ่านหนังสือพิมพ์ว่า “มีอาชญากรรมเกิดขึ้น **แถวนี้เยอะมาก**” ข้อความดังกล่าวถือว่าเป็นข้อสรุปทางสถิติที่ถูกนำมาใช้อย่างไม่เป็นทางการ (Huff 1954, 7 – 8)

Huff ได้แสดงให้เห็นว่าข้อสรุปทางสถิติดังกล่าวมีปัญหาเพราะวางอยู่บนกลุ่มตัวอย่าง (sample) ที่มีอคติ กล่าวคือ พ่อเลี้ยงได้ทักท้วงเขาว่า “รายงานข่าวอาชญากรรมที่ปรากฏบนพื้นที่ข่าวในหน้า หนังสือพิมพ์เป็นมาตรวัดอัตราการก่ออาชญากรรม” (Huff 1954, 8) ทั้งที่จริงรายงานข่าวอาชญากรรม ดังกล่าว อาจเป็นสิ่งที่ถูกนักข่าวเลือกมานำเสนอ เพราะมองว่าสามารถดึงดูดความสนใจของผู้อ่านก็เป็นไปได้

นอกจากนี้ Huff ยังได้ยกตัวอย่างว่า ในช่วงฤดูหนาว 2-3 ปีที่ผ่านมาได้มีกลุ่มผู้สังเกตการณ์นับ 10 รายออกมารายงานเรื่องยาแก้แพ้ (antihistamine pills) แต่ละคนได้แสดงให้เห็นถึงร้อยละของการลดลง ของอาการหวัดเมื่อได้รับการรักษาด้วยยาดังกล่าว เหตุการณ์ดังกล่าวได้นำมาสู่การเฟื่องฟูของโฆษณา ประชาสัมพันธ์และการผลิตยาดังกล่าว ทั้งที่จริงไม่ได้มีการสำรวจข้อมูลทางสถิติระยะยาวเพียงพอ ก็สร้าง ข้อสรุปขึ้นมาบนความหวังของผู้คนที่ไม่ต้องการเป็นหวัดหรืออยากหายจากอาการดังกล่าวโดยเร็ว จนกระทั่งแม้แต่ Henry G. Felsen ดาราตลกซึ่งไม่ได้มีความรู้ที่น่าเชื่อถือด้านการแพทย์ ได้ออกมาล้อเลียน ว่า “การรักษาที่ถูกวิธีจะทำให้ผู้ป่วยหายจากอาการหวัดได้ภายใน 7 วัน แต่ถ้าหากปล่อยไข้หวัดที่เป็นอยู่ ทั้งไว้จะหายภายใน 1 สัปดาห์” (ขีดเส้นใต้เน้นคำโดยผู้เขียน) (Huff 1954, 8) เพราะอาการหวัดจริง ๆ แล้ว ไม่ได้รักษาให้หายขาดได้ สิ่งที่ได้ทำคือการรักษาตามอาการ ดังเช่น น้ำมูกไหล คัดจมูก ไอ เป็นไข้ เป็นต้น (Pornlapas Boonsorn n.d.)

นอกจากการนำข้อความทางสถิติมาใช้อย่างผิด ๆ ในชีวิตประจำวันแล้ว ยังมีปัญหาความ น่าเชื่อถือของงานวิจัยทางการแพทย์ที่นำข้อมูลทางสถิติมาใช้เพื่อพัฒนาความรู้ ดังเช่น บทความเรื่อง “การศึกษาความผิดพลาดของการใช้สถิติสำหรับการวิจัยทางวิสัญญีวิทยา” ของ วราภรณ์ ไวกกุล, ไรจ นรินทร์ โกมลหิรัญ และภัทราภรณ์ สอนคำมี ที่ได้แสดงให้เห็นว่าในวงการแพทย์มีงานวิจัยที่มีความ ผิดพลาดจากการใช้สถิติในการศึกษาด้วยหลากหลายสาเหตุ (Waraporn Waikakul, Rojnarin Komonhirun and Patthrapon Sonkhammee 2014) ตัวอย่างเช่น ความผิดพลาดจากการเจตนาละเลย การรายงานค่าสถิติที่จำเป็น ไม่ว่าจะเป็นการรายงานเฉพาะค่า p-value แต่ไม่รายงานค่าเฉลี่ยและค่า ความแม่นยำ การไม่คำนวณขนาดตัวอย่างกลุ่มตัวอย่างในการวิจัย ไม่มีการสุ่ม การไม่ทดสอบความเป็น ปกติของโค้งแจกแจงข้อมูลแบบต่อเนื่อง (Waraporn Waikakul, Rojnarin Komonhirun and Patthrapon

Sonkhammee 2014, 65-69) ความผิดพลาดที่เกิดจากการใช้งาน เช่น การนำเสนอแผนภาพหรือกราฟที่สื่อความหมายไม่ตรงกับข้อมูลที่จะนำเสนอ ใช้ค่าสถิติพรรณนาไม่ถูกต้อง มีความสับสนจากระหว่าง “ค่านัยสำคัญทางสถิติ (statistical significance) กับความสำคัญทางคลินิก (clinical importance) เป็นต้น (Waraporn Waikakul, Rojnarin Komonhirun and Patthrapon Sonkhammee 2014, 69-73)

สภาพปัญหาของการนำสถิติมาใช้เป็นเครื่องมือสร้างองค์ความรู้แบบวิทยาศาสตร์ในลักษณะดังกล่าว ได้ถูกวิพากษ์วิจารณ์เอาไว้อย่างน่าสนใจโดย Leo Strauss¹ นักปรัชญาการเมืองเชื้อสายยิวที่เกิดในปรัสเซีย (เยอรมนี) แต่ภายหลังได้ย้ายมาอาศัยอยู่ในสหรัฐอเมริกาผ่านบทความที่ชื่อว่า Progress or Return?: The Contemporary Crisis in Western Civilization ซึ่ง Strauss มองว่า ความคิดเรื่องความก้าวหน้านั้นมีจุดกำเนิดมาจากการมองว่ามนุษย์นั้นไม่ได้มีชีวิตที่ดีและสมบูรณ์แบบ แต่มนุษย์

¹ ผู้เขียนหยิบยกประเด็นวิพากษ์วิจารณ์ของ Leo Strauss ที่มีต่อองค์ความรู้สมัยใหม่ขึ้นมาในที่นี้ เนื่องจาก Strauss เป็นนักปรัชญาการเมืองคนสำคัญที่ต้องการรื้อฟื้นภูมิปัญญาแบบโบราณให้กลับมาได้รับความสำคัญอีกครั้งในยุคสมัยใหม่ และได้ผลิตผลงานทางวิชาการเพื่อตอบโต้แนวทางการศึกษาแบบวิทยาศาสตร์นิยม (scientism) การศึกษาแบบปฏิฐานนิยมเชิงตรรกะ (Logical Positivism) และการศึกษาในเชิงพฤติกรรมศาสตร์ (behavioralism) ที่ถูกนำมาใช้อย่างกว้างขวางในการศึกษาสมัยใหม่ของสาขาวิชาต่าง ๆ ในด้านสังคมศาสตร์ รวมถึงรัฐศาสตร์ ซึ่งทำให้การศึกษาการเมืองในยุคสมัยใหม่ มุ่งเน้นไปที่ระเบียบวิธีการศึกษา (methodology) ที่เป็นวิทยาศาสตร์ โดยใช้คณิตศาสตร์และสถิติเป็นเครื่องมือสำคัญในการวิเคราะห์ข้อมูลอย่างมาก ด้วยความหวาดระแวงว่าจะเกิด “อคติ” ในการตัดสินเชิงคุณค่า/ค่านิยม (value judgement) ของผู้ศึกษาที่มีต่อสิ่งที่ศึกษาซึ่ง Strauss มองว่าบุคคลเหล่านั้นไม่รู้ไม่ว่าเรื่องดังกล่าวแท้จริงแล้วก็เป็นอคติอย่างหนึ่งเช่นกันและท้ายที่สุดความเชื่อที่ว่าวิธีการศึกษาดังกล่าวคือ ความก้าวหน้าในการใช้เหตุผลของมนุษย์แบบสมัยใหม่ (modern rationalism) ที่สามารถเอาชนะธรรมชาติได้ นับตั้งแต่คริสต์ศตวรรษที่ 17 เป็นต้นมา จะกลายเป็น “วิกฤตทางจิตวิญญาณของการใช้เหตุผลแบบสมัยใหม่” (the Spiritual Crisis of Modern Rationalism) เนื่องจากองค์ความรู้ดังกล่าวไม่ต่างจาก “ยักษ์ตาบอด” ที่ไม่สามารถแยกแยะความดี-ความชั่วได้ อันเป็นผลมาจากการจงใจละเลยภูมิปัญญาแบบโบราณของทั้ง Athens (ปราชญ์กรีก) และ Jerusalem (คัมภีร์ไบเบิล) ของโลกยุคใหม่ (Strauss 1989a, 249 – 310) ซึ่งข้อโต้แย้งของ Strauss ดังกล่าว สามารถพบเห็นได้ทั้งจากบทความที่เขาเขียนขึ้นใน ค.ศ. 1956 เรื่อง Social Science and Humanism และบทความที่เขียนขึ้นใน ค.ศ. 1961 เรื่อง Relativism ซึ่งงานทั้งสองชิ้นดังกล่าว Thomas L. Pangle ได้นำมารวมเล่มตีพิมพ์ซ้ำในบทที่ 1 และ 2 ของหนังสือเรื่อง The Rebirth of Classical Political Rationalism: An Introduction to the Thought of Leo Strauss (Strauss 1989b, 3 – 26) และยังมีบทความที่ Strauss เขียนขึ้นอีกชิ้นใน ค.ศ. 1971 เรื่อง Philosophy as Rigorous Science and Political Philosophy ซึ่งถูกนำมาตีพิมพ์ซ้ำไว้ในบทที่ 1 ของหนังสือเรื่อง Studies in Platonic Political Philosophy โดยมี Thomas L. Pangle เป็นผู้เขียนบทนำของหนังสือเล่มนี้ (Strauss 1983, 29 – 37) นอกจากนี้ อิทธิพลทางความคิดของ Strauss ที่มีต่อการศึกษการเมืองอเมริกัน ได้ขยายออกไปอย่างกว้างขวาง ผ่านบรรดาสถาบันศิษย์คนสำคัญรุ่นต่าง ๆ ที่ได้ออกไปทำงานสอนหนังสือในมหาวิทยาลัยชั้นนำต่าง ๆ ของสหรัฐอเมริกา ซึ่งถูกเรียกว่าเป็นสมาชิกของสำนักความคิดแบบ Strauss หรือ Straussian School ดูเพิ่มเติมได้จากบทที่ 6 และ 7 ของหนังสือเรื่อง The Truth about Leo Strauss: Political Philosophy and American Democracy (Zuckert and Zuckert 2006, 197 – 259)

สามารถเปลี่ยนไปสู่สิ่งที่ดีกว่าได้ในอนาคต ซึ่งความคิดแบบนี้ก็มีอยู่ในปรัชญายุคกรีกโบราณ (Strauss 1989a, 259 – 260) แต่จะแตกต่างออกไปจากแนวคิดเรื่องความก้าวหน้าในคริสต์ศตวรรษที่ 17 ซึ่งเป็นความคิดสมัยใหม่ (modern thought) ด้วยเงื่อนไขพื้นฐานจากความก้าวหน้าทางภูมิปัญญาและสังคมที่เปลี่ยนแปลงคู่ขนานกันไป ได้นำมาสู่ความคิดว่าภูมิปัญญาและสังคมของมนุษย์สามารถก้าวหน้าไปได้อย่างไม่สิ้นสุด (infinite intellectual and social progress) สามารถเกิดขึ้นได้อย่างแท้จริง (Strauss 1989a, 263-264)

บริบททางภูมิปัญญาและสังคมดังกล่าวได้ทำให้วิทยาศาสตร์และเทคโนโลยีสมัยใหม่กลายเป็นที่ยอมรับของสังคมโดยทั่วไปได้อย่างรวดเร็ว มนุษย์ต่างศรัทธาความคิดเรื่องความก้าวหน้าและโดยทั่วไปต่างเชื่อว่ามนุษย์สามารถเอาชนะธรรมชาติได้ มนุษย์จึงกลายเป็นผู้ควบคุมและเป็นเจ้าของธรรมชาติ (Strauss 1989a, 264) ภูมิปัญญาแบบโบราณ ดังเช่น ปรัชญาที่ใช้วิธีการศึกษาแตกต่างออกไปจากวิทยาศาสตร์ถูกดูแคลนลงไป เพราะไม่มีความเป็นกลาง (neutrality) และไม่มีความเป็นภววิสัย (objectivity)

อย่างไรก็ตาม Strauss กลับมองว่าถึงแม้ภูมิปัญญาของมนุษย์สมัยใหม่ (modern man) อย่างวิทยาศาสตร์ จะแสดงให้เห็นถึงความก้าวหน้าขององค์ความรู้อย่างไม่หยุดยั้งจนเปรียบเสมือนยักษ์ใหญ่เมื่อเปรียบเทียบกับภูมิปัญญาของมนุษย์สมัยโบราณ (ancient man) แต่สิ่งที่เกิดขึ้นกลับไม่ได้แสดงให้เห็นถึงการเพิ่มขึ้นของความรู้ความเข้าใจเรื่องความดีงามมากกว่ามนุษย์สมัยโบราณ มนุษย์สมัยใหม่จึงเปรียบเสมือนยักษ์ที่ไม่รู้ว่าอะไรดีหรืออะไรเลวมากกว่าที่มนุษย์ยุคก่อนหน้านั้นรู้ เพราะภูมิปัญญาสมัยใหม่แบบวิทยาศาสตร์ไม่ได้ให้ความสำคัญกับการตัดสินเชิงคุณค่า จึงไม่สามารถแยกความดีกับความชั่วร้ายได้ ดังนั้น ถ้าหากมนุษย์สมัยใหม่เป็นยักษ์ก็เป็นได้เพียงแค่ “ยักษ์ตาบอด” ที่จะนำมาสู่วิกฤตของอารยธรรมตะวันตกทั้งหมด และในคริสต์ศตวรรษที่ 19 ก็ปรากฏให้เห็นแล้วว่าการแยกแยะระหว่างสิ่งที่ดีกับสิ่งที่เลว ความดีงามกับความชั่วร้าย ได้ถูกแทนที่ด้วยการแยกออกจากกันระหว่างความก้าวหน้า (progress) กับสิ่งที่ปฏิกิริยาต่อความก้าวหน้า (reactionary) (Strauss 1989a, 264)

ยิ่งกว่านั้น Strauss ยังวิพากษ์วิจารณ์ว่าการค้นพบทฤษฎีความรู้ของวิทยาศาสตร์ที่กล่าวอ้างว่าเป็นวิถีทางที่จะทำให้มนุษย์มีความเข้าใจโลกตามธรรมชาติได้ ก็ไม่ใช่ความรู้ความเข้าใจอย่างสมบูรณ์แบบด้วยเหตุที่ความรู้ของวิทยาศาสตร์นับแต่คริสต์ศตวรรษที่ 19 มีพื้นฐานอยู่บนข้อสมมติฐาน (hypotheses) นั่นคือ ถ้าหากในอนาคตมีข้อสมมติฐานใหม่ที่พิสูจน์ได้มาหักล้างข้อสมมติฐานเก่าที่มีอยู่ ความรู้เดิมที่เคยรับรู้ก็จะไม่ใช่ความรู้ที่สมบูรณ์อีกต่อไป การศึกษาแบบที่วิทยาศาสตร์ใช้จึงเป็นเพียงแค่นั้นหนึ่งในหลากหลายวิธีที่ใช้ในการทำความเข้าใจโลกแค่นั้น (Strauss 1989a, 266) ความรู้ของวิทยาศาสตร์สมัยใหม่จึงไม่มีทางเหนือกว่าความรู้สมัยกรีกโบราณได้ กล่าวคือ แม้ว่าวิทยาศาสตร์จะมีความหยั่งทะนงในศักดิ์ศรีอย่างมากมาว่าเป็นองค์ความรู้ที่มีอำนาจเหนือกว่าอำนาจใด ๆ ในโลกสมัยใหม่ แต่ถ้าหากพิจารณาจากรากฐานขององค์ความรู้แล้วเปรียบเทียบกับวิทยาศาสตร์กับยักษ์ จะเห็นได้ว่าวิทยาศาสตร์เป็นได้เพียงแค่ว่า

“ยักษ์ที่มีเท้าเป็นดินเหนียว” ที่ไม่สามารถยืนได้อย่างมั่นคง (Strauss 1989a, 267) ดังที่ผู้เขียนจะขยายความคำวิพากษ์วิจารณ์ดังกล่าวของ Strauss ให้กระจ่างชัดในหัวข้อต่อไป

III. P-Value: กำเนิดปีศาจร้ายของการวิจัยเชิงปริมาณ²

ความไม่ลงรอยกันในแนวทางการศึกษาด้านสถิติ โดยเฉพาะที่ใช้ในการตรวจสอบสมมติฐานการวิจัย ปรากฏให้เห็นผ่านพัฒนาการทางประวัติศาสตร์ของการทดสอบค่านัยสำคัญ (significance test) ที่แบ่งออกได้เป็น 2 แนวทางการศึกษาหลัก คือสำนักการศึกษาแบบ Sir Ronald A. Fisher หรือ เรียกว่า Fisherian School และสำนักการศึกษาแบบ Neyman-Pearson School (Haig 2013, 12) ซึ่งการถกเถียงทางวิชาการดังกล่าว เกิดขึ้นมาโดยตลอด นับตั้งแต่กลางคริสต์ทศวรรษที่ 1930 จนกระทั่ง Fisher เสียชีวิตลงในปี 1962 ที่เมือง Adelaide ในออสเตรเลีย อย่างไรก็ตาม ในที่นี้ผู้เขียนจะกล่าวถึงยุคต่าง ๆ โดยใช้เกณฑ์กาลเวลาเริ่มต้นนำเสนอวิธีการทดสอบค่านัยสำคัญแต่ละวิธี จึงแบ่งออกเป็น 3 ยุค

เริ่มต้นจากยุคการสร้างวิธีการตรวจสอบสมมติฐานโดยอาศัยค่า p-value ของ Sir Ronald A. Fisher มาสู่ยุคการตรวจสอบสมมติฐานโดยอาศัยการสร้างช่วงวิกฤต (critical region) ของ Jerzy Neyman และ Egon Pearson และยุคสุดท้ายที่จะนำมากล่าวถึง คือ การตรวจสอบสมมติฐานโดยอาศัยช่วงความเชื่อมั่น (confidence interval) ของ Jerzy Neyman โดยมีรายละเอียดตามลำดับดังนี้

- กำเนิด P-Value กับค่านัยสำคัญ (significance)

งานของ Fisher (1950) *Statistical Methods for Research Workers* ซึ่งถูกตีพิมพ์ครั้งแรกในปี 1925 เป็นงานวิชาการด้านสถิติที่มีอิทธิพลอย่างมากต่อการทำงานวิจัยเชิงปริมาณยุคต่อ ๆ มา เพราะเขาได้คิดหาวิธีการทดสอบสมมติฐานการวิจัย โดยใช้สมมติฐานหลักเพียงอย่างเดียวในการศึกษา เรียกว่า “hypothesis” แต่ในปัจจุบันจะเรียกว่า “null hypothesis” (ใช้สัญลักษณ์ H_0 ในปัจจุบัน) หมายถึง ปรากฏการณ์ที่ผู้ศึกษาสนใจ/สังเกตซึ่งจะต้องถูกทำให้ใช้ไม่ได้ (เป็นโมฆะ) แต่สมมติฐานดังกล่าวไม่จำเป็นต้องมีความแตกต่างเท่ากับศูนย์ก็ได้

ด้วยเหตุผลดังกล่าว Fisher จึงไม่กำหนดสมมติฐานทางเลือกและไม่มีการคิดค่า Type II error (ซึ่งจะเกิดขึ้นก็ต่อเมื่อสมมติฐานการวิจัยผิดแต่ผู้วิจัยกลับยอมรับสมมติฐานนั้น) เพราะ Fisher มองว่าไม่มีใครที่จะ “ยอมรับ” สมมติฐานอยู่แล้ว (มีเพียงแค่จะปฏิเสธสมมติฐานได้หรือยังไม่ได้ในเวลานั้น) (Thompson

² ในการเขียนหัวข้อนี้ ผู้เขียนได้ประโยชน์อย่างมากจากการติดตามข้อมูลและเกร็ดความรู้ต่าง ๆ จากเพชฌัญญู DataRockie ดูได้ใน <https://goo.gl/Q1Wwnh>

³ เมื่อ Karl Pearson (ตอนเกิดใช้ชื่อ Carl) เกษียณจาก University College มหาวิทยาลัยลอนดอน ประเทศอังกฤษในปี 1933 ภาคิชาของเขาได้แตกออกเป็น 2 ภาคิชา คือ ภาคิชาสถิติ มีบุตรชายของเขาชื่อ Egon S. Pearson เป็นหัวหน้าภาคิชา และภาคิชาสุพันธุศาสตร์ (Eugenics) ที่มี Sir Ronald A. Fisher เป็นหัวหน้าภาคิชา และในปี 1934 ภาคิชาของ Egon S. Pearson ได้มีสมาชิกใหม่บรรจุเข้ามาทำงาน นั่นคือ Jerzy Neyman (see Thompson 2013, 8)

2013, 12) เว้นเสียแต่ว่าสมมติฐานดังกล่าวอยู่บนฐานของความเชื่อซึ่งอาจจะเกิดขึ้นจากการสังเกตเห็นปรากฏการณ์ที่สนใจเกิดขึ้นอย่างซ้ำ ๆ จนกระทั่งทำให้เชื่อว่าสิ่งเหล่านั้นเป็นความจริงแท้ที่จะต้องเกิดขึ้น

Fisher ได้สร้างสิ่งที่เรียกว่าค่า p -value ขึ้นมา (Fisher 1950, 120) แล้วอธิบายว่าค่าดังกล่าวหมายถึง ค่าความน่าจะเป็นของสิ่งที่สนใจศึกษาว่าจะเกิดขึ้น (หรือเกิดขึ้นมากกว่านั้น) เมื่อสมมติฐานหลักเป็นจริง แต่การพิสูจน์ว่าผลการศึกษาที่ได้นั้นน่าเชื่อถือหรือไม่ จำเป็นต้องมีค่าเปรียบเทียบซึ่งในประเด็นนี้เองที่ทำให้ Fisher ได้สมมติค่าเปรียบเทียบสำหรับตรวจสอบสมมติฐานตามความสะดวกของตนเองขึ้นมา มีค่าเท่ากับ 0.05 โดยไม่ได้มีการทดลองอย่างเป็นวิชาการ เพื่อเป็นพื้นฐานอ้างอิงการตัดสินใจกำหนดค่าดังกล่าวแต่อย่างใด Fisher ใช้เพียงแค่ความรู้สึกของตนเองประเมินว่า 0.05 มีค่าน้อยเพียงพอแล้วที่จะยอมให้เกิดความผิดพลาดในการทดสอบสมมติฐานได้⁴ โดยยกตัวอย่างเปรียบเทียบกับ การปลูกพืชในแต่ละฤดูกาล

- กำหนดสมมติฐานทางเลือก (alternative hypothesis – H_a) กับช่วงวิกฤต (critical region)

เมื่อถึงปี 1933 ได้มีนักสถิติจำนวน 2 คน คือ Jerzy Neyman กับ Egon S. Pearson พิมพ์บทความเผยแพร่ เรื่อง “On the Problem of the Most Efficient Tests of Statistical Hypothesis” (Neyman and Pearson 1933, 289 – 337) ซึ่งเสนอให้มีสมมติฐานทางเลือก (alternative hypothesis – H_a) และได้ตั้งชื่อให้กับค่า “0.05” ของ Fisher เรียกว่า “ ϵ ”⁵ ซึ่งก็คือค่าความผิดพลาดแบบที่ 1 (type I error หรือเรียกอีกอย่างว่า false positive) หมายถึง การสรุปผลผิดในการทดสอบสมมติฐานของการวิจัย กล่าวคือ ผู้วิจัยปฏิเสธสมมติฐานหลัก ทั้งที่สมมติฐานดังกล่าวเป็นความจริง (Fisher 1926, 336)

นอกจากนี้ Neyman กับ Pearson ยังได้เสนอให้ใช้ “พื้นที่วิกฤต” (Critical Region: CR) แทนการใช้ p -value ของ Fisher ในการทดสอบสมมติฐาน (Fisher 1926, 336 – 337) ซึ่งพื้นที่วิกฤตดังกล่าว คือ พื้นที่ปลายเส้นความแตกต่างจากค่าเฉลี่ยที่ถูกกำหนดขึ้นมามีขนาดเท่ากับ 0.05 ดังนั้น เมื่อพิจารณาข้อเสนอต่าง ๆ ของ Neyman และ Pearson จะทำให้เห็นแนวทางที่แตกต่างไปจาก Fisher กล่าวคือ ฝ่ายแรกเป็นพวกที่เน้นการทดสอบสมมติฐาน (hypothesis testing) ขณะที่อีกฝ่ายเน้นทดสอบนัยสำคัญ (significance testing) (Haig 2013, 13)

- ช่วงความเชื่อมั่น (confidence interval) กับความแม่นยำในการคาดการณ์

เมื่อถึงปี 1937 Jerzy Neyman ได้เผยแพร่บทความ เรื่อง Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability เพื่อเสนอวิธีทดสอบสมมติฐานขึ้นมาใหม่ เรียกว่า “ช่วงความเชื่อมั่น” (Confidence Interval: CI) (Neyman 1937, 378) ที่เกิดจากการได้ค่าทางสถิติที่เกิดขึ้นเมื่อมีการทดลองแบบซ้ำ ๆ ในลักษณะที่เป็นการสุ่ม (random) ซึ่งโดยทั่วไปจะไม่ได้มีค่าเท่ากันทุก

⁴ ดังที่ Fisher กล่าวว่า “Personally, the writer prefers to set a low standard of significance at the 5 per cent. Point, and ignore entirely all results which fail to reach this level...” (see Fisher 1926, 504)

⁵ ในปัจจุบันจะใช้สัญลักษณ์ α (ค่า alpha)

ครั้งแต่ก็มักจะเป็นการกระจายตัวของค่าทางสถิติจากกลุ่มตัวอย่าง (sampling distribution) ที่มีความใกล้เคียงกัน

กล่าวคือ ค่าทางสถิติที่ได้จะมีค่าลู่เข้าสู่ค่ากลางค่าใดค่าหนึ่งเสมอ เป็นไปตามทฤษฎีค่ากลาง (Central Limit Theorem: CLT)⁶ ของ Pierre Simon Laplace (Stigler 1986, 136-137) ที่อาศัยการสร้างกราฟการกระจายตัวของค่าทางสถิติจากกลุ่มตัวอย่างมีลักษณะเป็นการกระจายตัวแบบปกติ (normal distribution) ลักษณะเหมือนระฆังคว่ำ โดยมีค่าความคลาดเคลื่อนมาตรฐาน (standard error)⁷ ที่แตกต่างไปจากค่าเฉลี่ย ซึ่งทำให้ผู้วิจัยสามารถสร้างเป็นค่าของช่วงความเชื่อมั่นได้ว่าปรากฏการณ์ที่สนใจศึกษามีความน่าจะเป็นว่าจะเกิดขึ้นอย่างแน่นอน เพราะเมื่อมีการนำ CI ไปใช้ในการทดสอบสมมติฐานการวิจัยซ้ำ ๆ ในรอบ 100 ครั้ง (หรือมากกว่านั้น) แล้วเปรียบเทียบจะพบว่าปรากฏการณ์ที่สังเกตมีความแตกต่างไปจากเดิมเพียงแค่ว่าค่าเท่านั้น⁸

อย่างไรก็ตาม ถึงแม้ว่าจะมีการพัฒนาแนวทางการทดสอบนัยสำคัญและสมมติฐานทางสถิติมากขึ้นเรื่อย ๆ แต่ก็ไม่ได้หมายความว่าองค์ความรู้ดังกล่าวจะกลายเป็นองค์ความรู้ที่สมบูรณ์ เพราะแนวทางทั้ง 3 ดังกล่าว ไม่ว่าจะเป็นการใช้ค่า p-value การใช้พื้นที่วิกฤต (CR) หรือการใช้ช่วงความเชื่อมั่น (CI) ก็ยังคงเกิดปัญหาอยู่ เพราะนักวิจัยที่ใช้สถิติในการศึกษาได้นำแนวทางการศึกษาแต่ละสำนักมาผสมรวมกัน และใช้กันอย่างสับสน (เรียกว่า hybrid logic) (Haig 2013, 14)

ต่อมาในปี 1969 จึงมีนักสถิติชื่อว่า Jacob Cohen เขียนหนังสือชื่อ Statistical Power Analysis for the Behavioral Sciences เพื่อแสดงให้เห็นว่าแนวทางการศึกษาแบบดั้งเดิมทั้ง 3 ดังกล่าวนั้นมีปัญหาตั้งแต่ต้น กล่าวคือ การตั้งสมมติฐานหลัก (H_0) ว่าค่าเฉลี่ยที่ได้จากการทดลองไม่มีความแตกต่างกันนั้นเป็นการตั้งข้อสมมติฐานที่เป็นไปไม่ได้ตั้งแต่แรก (Cohen 1977, 1 – 2) เพราะไม่มีทางที่จะมีการทดลองที่ได้ค่าเฉลี่ยที่เท่ากันทุกประการจนกระทั่งถึงระดับของทศนิยมดังที่ Tukey (1991, 100) กล่าวว่า “พวกเราทุก

⁶ ทฤษฎีค่ากลาง (CLT) จำเป็นต้องอาศัยเงื่อนไขจากการสุ่มกลุ่มตัวอย่างแบบอิสระ (random sampling) อย่างซ้ำ ๆ หลายครั้งเพื่อลดอคติในการศึกษา และมีขนาดของกลุ่มตัวอย่างที่สุ่มมาจำนวนไม่ต่ำกว่า 30 คน จึงจะสามารถใช้ทฤษฎีค่ากลางได้

⁷ ค่าความคลาดเคลื่อนมาตรฐาน (Standard Error: S.E.) คือ ค่าส่วนเบี่ยงเบนมาตรฐานที่แตกต่างออกไปจากค่าเฉลี่ยตามการกระจายตัวของกลุ่มตัวอย่าง (sampling distribution) ซึ่งเป็นสิ่งที่เกิดขึ้นในจินตนาการทางสถิติ แต่ในกรณีของค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) จะหมายถึง ค่าทางสถิติที่ได้จากการทดลองแต่ละครั้งซึ่งแตกต่างไปจากค่าเฉลี่ยของค่าทางสถิติจากการทดลองจริงแบบซ้ำ ๆ หลายครั้ง

⁸ ถึงแม้ CI จะทำให้การทดสอบสมมติฐานมีประสิทธิภาพมากขึ้น แต่ก็มีข้อเสียคือ เมื่อ CI มีช่วงที่กว้างขึ้น จะยังทำให้ความแม่นยำในการทำนายลดลง

คนต่างรู้เกี่ยวกับโลกที่สอนเราว่าผลกระทบจาก A และ B มักแตกต่างกันเสมอ – ในระดับทศนิยม – สำหรับ A และ B ใด ๆ ก็ตาม ดังนั้น การถามว่าผลกระทบที่เกิดขึ้นนี้แตกต่างกันหรือไม่ จึงเป็นสิ่งที่โง่"⁹

และในปี 1994 Cohen ได้เขียนบทความชื่อว่า The Earth Is Round ($p < .05$) (Cohen 1994) เพื่อแสดงให้เห็นว่าโลกนี้ไม่ได้กลม จากการปฏิเสธสมมติฐานหลักที่ว่า “โลกกลม” ที่ระดับนัยสำคัญน้อยกว่า 0.05 แต่การตั้งสมมติฐานหลักดังกล่าวนี้เป็นสิ่งที่ผิดตั้งแต่ต้น เพราะปรากฏการณ์ที่สังเกตนั้นแตกต่างกันอยู่แล้วโดยธรรมชาติ (เพราะตามความเป็นจริงโลกไม่ได้กลมแต่เป็นวงรี) การถามถึงความแตกต่างจึงเป็นการตั้งคำถามผิดในการทดสอบสมมติฐานการวิจัย เพราะสิ่งที่สมควรถามควรเป็นขนาดของความแตกต่างในปรากฏการณ์ที่สนใจศึกษาแต่ละครั้งนั้นมีมากขนาดไหนเสียมากกว่า การเชื่อและรับเอาแนวคิดเรื่องการทดสอบนัยสำคัญและการทดสอบสมมติฐานที่ผิดตั้งแต่ต้นมาใช้ต่อ ๆ กันไปของนักวิจัยจึงกลายเป็นปัญหาที่ตกทอดมาจนถึงปัจจุบัน

IV. อคติของนักวิจัยกับการบิดเบือนเจตนารมณ์ของ Karl Popper

การศึกษาปรากฏการณ์ต่าง ๆ ที่สามารถพิสูจน์ว่าผิดได้ (falsification) ถือเป็นหลักการสำคัญของการสร้างองค์ความรู้แบบวิทยาศาสตร์ที่ Karl Popper ได้คิดขึ้นมา เพื่อให้ความรู้ก้าวหน้าไปได้เรื่อย ๆ โดยไม่ต้องเสียเวลาในการพิสูจน์ด้วยวิธีการศึกษาแบบอุปนัย (induction) ที่ต้องเก็บข้อมูลรายละเอียดปลีกย่อยต่าง ๆ ให้ครบถ้วนสมบูรณ์เสียก่อนแล้วจึงสร้างข้อสรุปผลการศึกษา หากแต่สามารถเก็บข้อมูลเพียงจำนวนหนึ่งที่แสดงให้เห็นถึงความสอดคล้องไปในทิศทางเดียวกันแล้วอนุมานมาสู่ข้อสรุปผลการศึกษาที่เป็นไปตามสมมติฐานด้วยวิธีการศึกษาแบบนิรนัย (deduction) (Ekkapollanun Nuttappattanun 2017, 19 – 20; Keuth 2005, 2 and 11; O’Hear 1980, 57 – 58; Williams 2000, 20 – 21)

ดังนั้น Popper จึงมองว่าไม่ใช่ความรู้ทุกอย่างของมนุษย์จะเป็นความรู้ที่แท้จริงแบบวิทยาศาสตร์ได้ ถ้าหากไม่สามารถพิสูจน์ได้ว่าเป็นสิ่งที่ถูกหรือผิด นั่นคือ ถ้าหากเป็นความรู้ที่เหนือธรรมชาติหรือไม่มีทางที่จะพิสูจน์ได้ แม้ว่าจะมีความน่าเชื่อถือแต่ก็จะเป็นองค์ความรู้ ดังเช่น ความรู้เกี่ยวกับเทพเจ้า ชีวิตหลังความตาย กำเนิดจักรวาล ฯลฯ ที่ไม่สามารถหาข้อมูลพิสูจน์หักล้างหรือทำลายชุดความรู้นั้นโดยใช้ชุดความรู้ใหม่ ๆ ที่สร้างขึ้นมา กล่าวคือ Popper ได้เสนอให้มีการสร้างองค์ความรู้ในลักษณะที่เป็นสมมติฐานของปรากฏการณ์ที่สนใจศึกษาขึ้นมา ตรวจสอบดูที่ไม่มีชุดข้อมูลใดที่ถูกนำมาทำลายแล้วหักล้างสมมติฐานของงานวิจัยได้ ก็จะทำให้สมมติฐานดังกล่าวยกสถานะขึ้นไปเป็นทฤษฎีที่ใช้อธิบายต่อไปได้เรื่อย ๆ เป็นพื้นฐานขององค์ความรู้ที่จะใช้ต่อยอดศึกษาเรื่องอื่น ๆ ที่เกี่ยวข้องได้ต่อไป

⁹ ข้อความในต้นฉบับภาษาอังกฤษ คือ All we know about the world teaches us that the effects of A and B are always different – in some decimal place- for any A and B. Thus asking ‘Are the effects different?’ is foolish

อย่างไรก็ตาม สิ่งที่น่าสังเกตในการตั้งสมมติฐานตามความคิดของ Popper ซึ่งแตกต่างอย่างมากจากการพิสูจน์สมมติฐานงานวิจัยในปัจจุบัน ที่มีการตั้งสมมติฐานหลัก (Null Hypothesis: H_0) และสมมติฐานทางเลือก (alternative hypothesis: H_a) โดยจะใช้สมมติฐานของงานวิจัยซึ่งเป็นสิ่งที่นักวิจัยต้องการพิสูจน์ว่าเป็นองค์ความรู้ใหม่กำหนดให้เป็น H_a การกระทำดังกล่าวถือเป็นเรื่องกลับตาลปัตรไปจากความคิดของ Popper¹⁰ ที่ต้องการให้มีการพิสูจน์สิ่งที่เป็นองค์ความรู้ที่นักวิจัยสร้างขึ้นใหม่หรือเสนอขึ้นมาหาใช่การตรวจสอบด้วยการพิสูจน์ผิดต่อสมมติฐานหลัก (H_0) (ที่ไม่มีทางเป็นจริง) แล้วพิจารณาว่าข้อมูลทางสถิติที่ได้สามารถปฏิเสธ H_0 ได้สำเร็จหรือไม่ เพื่อที่จะสรุปว่าสมมติฐานการวิจัย ซึ่งเป็นสมมติฐานทางเลือก หรือ H_a นั้นเป็นจริง เพราะการกระทำดังกล่าวเท่ากับว่า ข้อเสนอที่ว่า H_a เป็นจริงนั้นไม่ได้รับการตรวจสอบว่าเป็นองค์ความรู้ที่แท้จริงหรือไม่

ผู้เขียนขอยกตัวอย่างสมมติ เช่น เมื่อนักรัฐศาสตร์มีความเชื่อว่า “เมื่อเศรษฐกิจเติบโตขึ้นจนทำให้ประชาชนทั่วไปมีค่าเฉลี่ยรายได้ขยับสูงขึ้นและยกสถานะเป็นชนชั้นกลาง จะทำให้พวกเขากระตือรือร้นที่จะเข้ามีส่วนร่วมทางการเมืองสูงขึ้น (ดังเช่นพิจารณาจากค่าเฉลี่ยการออกมาลงคะแนนเสียงเลือกตั้ง) อันเป็นพื้นฐานสำคัญของการเมืองแบบประชาธิปไตย” แต่กระนั้นก็ตาม ด้วยความเชื่อในสังคมไทยที่ว่าการเมืองไทยมีลักษณะเป็นระบบอำมาตยาธิปไตย หรือระบบการเมืองที่ข้าราชการมีอำนาจมากในการตัดสินใจ ขณะที่ประชาชนมีความเฉื่อยชาทางการเมือง เพราะมองว่าการเมืองไม่ใช่เรื่องของตนเองหากแต่เป็นเรื่องของชนชั้นนำผู้มีอำนาจ แม้ว่าตนเองจะมีสถานะทางเศรษฐกิจดีขึ้นหรือไม่ก็ตาม

ถ้าหากนักรัฐศาสตร์ต้องการพิสูจน์ว่า ความเชื่อของตนเรื่องความสัมพันธ์ระหว่างการพัฒนาทางเศรษฐกิจ (การยกระดับรายได้เป็นชนชั้นกลาง) กับการพัฒนาประชาธิปไตย (การมีส่วนร่วมทางการเมือง) เป็นความจริงและความเชื่อที่มีอยู่เดิมที่ว่า ไทยมีลักษณะของการเป็นระบบอำมาตยาธิปไตยและประชาชนมีความเฉื่อยชาทางการเมืองนั้นใช้ไม่ได้อีกต่อไป โดยตั้งสมมติฐานหลัก (H_0) และสมมติฐานทางเลือกซึ่งเป็นสมมติฐานการวิจัย (H_a) ดังนี้

H_0 : การพัฒนาเศรษฐกิจไม่มีความสัมพันธ์กับการพัฒนาประชาธิปไตย

H_a : การพัฒนาเศรษฐกิจมีความสัมพันธ์กับการพัฒนาประชาธิปไตย

ถ้าหากการเก็บข้อมูลโดยสุ่มกลุ่มตัวอย่างจากประชาชนไทยจำนวนหนึ่ง เพื่อค่าเฉลี่ยของรายได้ของประชาชนและค่าเฉลี่ยของการออกมาลงคะแนนเสียงเลือกตั้งของคนไทย 2 กลุ่มเปรียบเทียบค่า

¹⁰ ตามความคิดเรื่องการพิสูจน์ผิด (falsification) ของ Popper ถ้าหากนักวิจัยต้องการจะพิสูจน์ว่าทฤษฎีใดใช้ได้หรือไม่ ก็ต้องพยายามหาหลักฐานมาโต้แย้งเพื่อจะล้มทฤษฎีนั้น ถ้าหากไม่สามารถโต้แย้งได้จึงจะทำให้ทฤษฎีนั้นใช้ได้ต่อไปจนกว่าจะถูกหักล้าง ไม่ใช่ความพยายามหาหลักฐานมายืนยันว่าสอดคล้องกับทฤษฎีที่นักวิจัยต้องการพิสูจน์ (Hawkesworth 2004, 11; Keuth 2005, 13; Smith 2000, 9)

สหสัมพันธ์ (correlation) โดยแยกตามระดับรายได้ แล้วพบว่าข้อมูลที่ได้มีนัยสำคัญ กล่าวคือ ค่า p -value มีค่าต่ำกว่า 0.05 ($p < 0.05$) จึงปฏิเสธ H_0 และยอมรับ H_a พร้อมกับสรุปว่า H_a เป็นจริงที่ระดับนัยสำคัญที่ 0.05 กล่าวคือ การพัฒนาเศรษฐกิจมีความสัมพันธ์กับการพัฒนาประชาธิปไตยเป็นจริง เพราะค่าสหสัมพันธ์ของกลุ่มตัวอย่างประชาชนที่มีรายได้เฉลี่ยสูงกว่าที่สุ่มมา มีอัตราการเข้ามีส่วนร่วมทางการเมือง ออกมาลงคะแนนเสียงเลือกตั้ง (voter turnout) สูงกว่ากลุ่มตัวอย่างประชาชนอีกกลุ่มหนึ่ง

ข้อสรุปดังกล่าวหาใช่การพิสูจน์ความถูกต้องของสมมติฐานการวิจัยแบบที่ Popper ต้องการ เพราะชุดความรู้ที่สร้างขึ้นใหม่ไม่ได้ถูกพิสูจน์จริง ๆ เพียงแต่มีการยืนยันแนวคิดใหม่ที่ถูกตั้งเป็นข้อสมมติฐานทางเลือกเท่านั้น ด้วยการพิสูจน์ว่าความสัมพันธ์ระหว่างการพัฒนาเศรษฐกิจกับการพัฒนาประชาธิปไตยมีอยู่จริงด้วยเงื่อนไขของข้อมูลจากกลุ่มตัวอย่างที่เก็บมาพิจารณา เพราะค่าสหสัมพันธ์ของรายได้เฉลี่ยของประชาชนที่ถูกสุ่มกลุ่มตัวอย่างมา มีความสัมพันธ์กับอัตราการมีส่วนร่วมทางการเมืองในรูปแบบของการออกมลงคะแนนเสียงเลือกตั้งที่สูงขึ้น การที่หักเอาว่าระบบการเมืองไทยแบบอำมาตยาธิปไตยได้เปลี่ยนแปลงไปเป็นประชาธิปไตยแล้ว โดยไม่ได้มีการหักล้างทฤษฎีเดิมย่อมไม่ถูกต้อง

เพราะถ้าหากจะสรุปเช่นนั้นขึ้นแรกจะต้องทดสอบด้วยว่าตัวแปรแต่ละคู่ ได้แก่ ระบบอำมาตยาธิปไตย – การพัฒนาเศรษฐกิจ – การมีส่วนร่วมทางการเมือง มีความสัมพันธ์กันจริงหรือไม่ แล้วจึงพยายามหักล้างแนวคิดเรื่องระบบอำมาตยาธิปไตยว่าไม่มีอยู่จริงให้สำเร็จก่อน แล้วจึงหาข้อมูลมาหักล้างความเชื่อเรื่องความสัมพันธ์ระหว่างการพัฒนาเศรษฐกิจกับการพัฒนาประชาธิปไตย ถ้าหากพบว่าตัวแปรบางคู่มีความสัมพันธ์กันจริง ก็ควรจะต้องพิสูจน์ให้ได้ว่าความสัมพันธ์ดังกล่าวมีมากเพียงใดมากกว่าจะบอกว่ามีความสัมพันธ์อยู่จริงสถานการณ์สมมติดังกล่าวเป็นตัวอย่างของอคติของนักวิจัยซึ่งผู้เขียนจะกล่าวถึงอีกครั้งในหัวข้อ อคติของผู้ศึกษากับการตีความผลการวิเคราะห์ทางสถิติ

V. Null Hypothesis (H_0): ต้นไม้พิษของการทดสอบสมมติฐาน

จุดเริ่มต้นของการสร้างข้อสมมติฐานหลักนับตั้งแต่ยุคของ Fisher ซึ่งเปรียบเสมือนการปฏิวัติการศึกษาปรากฏการณ์ต่าง ๆ ด้วยวิธีการทางสถิติ นับเป็นอย่างก้าวสำคัญในการทดสอบสมมติฐานการวิจัย แม้ว่าในช่วงเริ่มต้นจะมีเพียงแค่การตั้งสมมติฐานหลักเพียงประการเดียวเท่านั้น แต่ก็ทำให้สามารถนำข้อค้นพบที่ได้จากปรากฏการณ์ที่ศึกษาไปพิสูจน์ผิด อันนำมาสู่การสร้างข้อค้นพบใหม่ ๆ จนกระทั่งเกิดเป็นทฤษฎีขององค์ความรู้ด้านนั้น ๆ ได้ แต่ทว่าจุดเริ่มต้นดังกล่าวแท้ที่จริงแล้ว คือ “ต้นไม้พิษ” ที่จะนำมาสู่การเกิด “ผลไม้พิษ” (ข้อสรุปจากการทดสอบสมมติฐานงานวิจัย) ที่จะถูกนำไปแพร่พันธุ์ต่อไปในรูปของการเป็นฐานต่อยอดความรู้ในการศึกษาเรื่องอื่น ๆ

สาเหตุที่ผู้เขียนเปรียบเทียบจุดเริ่มต้นของการสร้างสมมติฐานหลัก (H_0) ว่าเป็นต้นไม้พิษ ก็เพราะว่าสมมติฐานหลักที่ตั้งต้นขึ้นนั้นอยู่บนฐานของความเป็นไปไม่ได้ตั้งแต่แรกแล้ว¹¹ กล่าวคือ Fisher และนักสถิติคนอื่น ๆ ที่ใช้แนวทางการตรวจสอบสมมติฐานการวิจัยในลักษณะเดียวกันนี้ในเวลาต่อมา หรือที่เรียกว่าเป็นพวกสำนัก Fisherian (Fisherian School) ได้ทำให้เกิดการตั้ง H_0 ที่พร้อมจะถูกหักล้างได้อยู่ตลอดเวลา ไม่ใช่การสร้าง H_0 เพื่อจะพิสูจน์ว่าเป็นทฤษฎีที่ใช้ได้หรือไม่เพราะโดยแท้จริงแล้วสมมติฐานทางเลือก (H_a) ต่างหากที่เป็นสมมติฐานการวิจัยที่ผู้วิจัยต้องการพิสูจน์ว่าใช้ได้จริง

นอกจากนี้เกณฑ์ในการตรวจสอบสมมติฐานของ Fisher (ระดับนัยสำคัญ 0.05) ก็เป็นเกณฑ์ที่ถูกสร้างขึ้นตามอำเภอใจไม่ได้มีหลักการรองรับ และแทนที่จะเป็นเกณฑ์ที่ช่วยสร้างความชัดเจนในการทำการทดลองให้กับนักวิจัยกลับเป็นเกณฑ์ที่คลุมเครือ ทำให้มีการนำเกณฑ์ดังกล่าวไปใช้อย่างผิด ๆ ไม่ได้ก่อประโยชน์ นำมาสู่การเริ่มสั่งห้ามตีพิมพ์ผลงานวิจัยที่ใช้ค่า p-value ในการสร้างข้อสรุปผลการศึกษาเผยแพร่ในวารสารทางวิชาการบางฉบับในที่สุด (Siegfried 2015)

ตัวอย่างปัญหาของการใช้ค่า p-value ในการทดสอบสมมติฐานการวิจัยมีดังนี้ เมื่อวิเคราะห์ข้อมูลทางสถิติแล้วพบว่าได้ค่า p-value ที่ต่ำมาก (ต่ำกว่า 0.05) ไม่ได้หมายความว่าปรากฏการณ์ที่สังเกตนั้นมีความแตกต่างจากสมมติฐานหลัก (H_0) แล้วจะสามารถสรุปได้ว่าข้อค้นพบในการวิจัยนั้นเป็นสิ่งที่มีความจริง (true) โดยอ้างว่าค่าที่ได้ของผู้วิจัยมีนัยสำคัญทางสถิติ เพราะสิ่งที่เกิดขึ้นบอกผู้วิจัยได้แค่ว่าโอกาสที่จะเกิดปรากฏการณ์ที่ศึกษาเป็นไปได้ยากมากเมื่อ H_0 เป็นจริง และที่กล่าวว่าไม่สามารถสรุปได้ว่าสมมติฐานทางเลือก (H_a) เป็นจริงก็เพราะในการทดลองครั้งต่อไป ก็ยังคงมีโอกาสที่จะไม่สามารถปฏิเสธ H_0 ได้อีกเช่นกัน เนื่องจากการสุ่มกลุ่มตัวอย่างที่นักวิจัยทำขึ้นมาไม่ได้ครอบคลุมถึงข้อมูลทั้งหมดของประชากร ซึ่งอาจจะมีข้อมูลที่ทำให้ได้ข้อสรุปในการศึกษาที่แตกต่างออกไปได้

ดังนั้น กรณีศึกษาการเลือกตั้งประธานาธิบดีสหรัฐอเมริกาในปี 2016 ระหว่าง Hillary Clinton กับ Donald Trump ว่าเพศของผู้ลงคะแนนเสียงเลือกตั้งส่งผลต่อการตัดสินใจเลือกผู้สมัครรับเลือกตั้งตำแหน่งประธานาธิบดีสหรัฐอเมริกาที่แตกต่างกันหรือไม่ โดยตั้งสมมติฐานดังนี้

H_0 : เพศของผู้ลงคะแนนเสียงเลือกตั้งที่ต่างกันไม่มีผลต่อการตัดสินใจเลือกผู้ลงสมัครรับเลือกตั้งประธานาธิบดีสหรัฐอเมริกา

¹¹ ไม่ว่าสมมติฐานหลัก (H_0) จะถูกกำหนดให้มีค่าเท่ากับ 0 หรือถูกให้กำหนดให้เท่ากับค่าใดค่าหนึ่งก็ตาม ต่างก็มีโอกาสที่จะถูกปฏิเสธได้โดยง่าย ดังเช่น ถ้าหากสมมติให้ H_0 คือ อัตราเฉลี่ยของการออกมาลงคะแนนเสียงเลือกตั้งของผู้มีสิทธิลงคะแนนเสียงเลือกตั้งเพศชาย (μ_a) และอัตราเฉลี่ยของการออกมาลงคะแนนเสียงเลือกตั้งของผู้หญิง (μ_b) ไม่ต่างกัน ($H_0: \mu_a = \mu_b$) หรือถ้าสมมติให้ H_0 คือ อัตราเฉลี่ยของการออกมาลงคะแนนเสียงเลือกตั้งของผู้หญิงต่างกันเท่ากับ 4 ($H_0: \mu_a - \mu_b = 4$) เมื่อจะทดสอบสมมติฐานก็เป็นไปได้สูงที่จะหาความแตกต่างค่าเฉลี่ยของกลุ่มตัวอย่างได้ค่าที่ต่างออกไปจากที่กำหนดไว้ใน H_0 ซึ่งจะทำให้ H_0 ถูกหักล้างไป

H_a : เพศของผู้ลงคะแนนเสียงเลือกตั้งที่ต่างกันมีผลต่อการตัดสินใจเลือกผู้ลงสมัครรับเลือกตั้งประธานาธิบดีสหรัฐอเมริกา

ผลการวิเคราะห์ข้อมูลทางสถิติโดยเปรียบเทียบค่าเฉลี่ยคะแนนเสียงเลือกตั้งที่ผู้สมัครรับเลือกตั้งตำแหน่งประธานาธิบดีสหรัฐอเมริกาได้รับจากผู้ลงคะแนนเสียงเลือกตั้งแต่ละเพศ ปรากฏว่ามีค่า $p\text{-value} = 0.001$ ($p < 0.05$) จะทำให้ผู้วิจัยต้องปฏิเสธ H_0 และยอมรับ H_a นำมาสู่ข้อสรุปว่า เพศของผู้ลงคะแนนเสียงเลือกตั้งที่ต่างกันมีผลต่อการตัดสินใจเลือกผู้ลงสมัครรับเลือกตั้งประธานาธิบดีสหรัฐอเมริกา แต่ก็ได้ไม่ได้หมายความว่าข้อสรุปนี้จะเป็นความจริงแท้แน่นอน เพราะอาจจะพบว่าโดยแท้จริงแล้วปัจจัยที่ส่งผลต่อการตัดสินใจเลือกผู้สมัครรับเลือกตั้งมาจากการยึดถือชุดอุดมการณ์ทางการเมืองที่แตกต่างกันระหว่างอนุรักษนิยมกับเสรีนิยมมากกว่าปัจจัยความแตกต่างทางเพศของผู้ลงคะแนนเสียงเลือกตั้ง นั่นคือ เพศที่แตกต่างกันของผู้ลงคะแนนเสียงเลือกตั้งไม่ได้มีผลต่อการตัดสินใจเลือกผู้สมัครรับเลือกตั้ง ซึ่งถือว่าการสรุปผลการศึกษานี้ขึ้นต้นได้เกิดความผิดพลาดในการตัดสินใจของผู้วิจัยแบบ Type I error (ตัดสินใจปฏิเสธ H_0 ทั้งที่ H_0 มีอยู่จริง) อย่างไรก็ตาม ในการทดลองครั้งถัดไป ผู้วิจัยอาจจะได้ข้อสรุปที่ไม่สามารถหักล้าง H_0 ก็ได้

ในทางกลับกัน สมมติว่าผู้วิจัยวิเคราะห์ข้อมูลทางสถิติแล้วพบว่าได้ค่า $p\text{-value}$ ที่สูงกว่า 0.05 ก็ได้ไม่ได้หมายความว่า ผู้วิจัยจะต้องยอมรับ H_0 เพราะในความเป็นจริง H_0 ดังกล่าวไม่ได้เป็นจริงตั้งแต่แรกเริ่ม หากเพียงแต่ผู้วิจัยยังไม่สามารถปฏิเสธได้เท่านั้น เพราะยังไม่มีข้อมูลหลักฐานเพียงพอที่จะนำไปหักล้าง ผู้วิจัยจึงไม่มีทางที่จะสรุปว่ายอมรับ H_0 นั้นได้ด้วย

ดังเช่นถ้าหากผู้เขียนยกตัวอย่างต่อเนื่องจากเดิมพบว่า ผลการวิเคราะห์ข้อมูลทางสถิติโดยเปรียบเทียบค่าเฉลี่ยคะแนนเสียงเลือกตั้งที่ผู้สมัครรับเลือกตั้งตำแหน่งประธานาธิบดีสหรัฐอเมริกาได้รับจากผู้ลงคะแนนเสียงเลือกตั้งแต่ละเพศ ปรากฏว่ามีค่า $p\text{-value} = 0.1$ ($p > 0.05$) ซึ่งจะทำให้ผู้วิจัยล้มเหลวในการปฏิเสธ H_0 นำมาสู่ข้อสรุปว่า เพศของผู้ลงคะแนนเสียงเลือกตั้งที่ต่างกันไม่มีผลต่อการตัดสินใจเลือกผู้ลงสมัครรับเลือกตั้งประธานาธิบดีสหรัฐอเมริกา แต่ก็ได้ไม่ได้หมายความว่าข้อสรุปนี้จะเป็นความจริงแท้แน่นอน เพราะอาจจะพบว่าข้อมูลที่เก็บรวบรวมมาในครั้งนี้มาจากพื้นที่ฐานคะแนนเสียงของผู้ลงสมัครรับเลือกตั้งคนใดคนหนึ่งเป็นจำนวนมาก ทำให้ผู้ลงคะแนนเสียงไม่ว่าจะเพศใดก็เลือกผู้สมัครรับเลือกตั้งคนนี้แต่ในการทดลองครั้งถัดไปอาจจะไม่ได้ผลสรุปเป็นเช่นนี้ ถ้าหากการสุ่มกลุ่มตัวอย่างมีการกระจายตัวครอบคลุมทุกพื้นที่มากขึ้น เป็นต้น

ดังนั้น การใช้วิธีการทดสอบนัยสำคัญโดยตั้งสมมติฐานหลักแบบ Fisherian School ที่ให้ความสำคัญกับการใช้เกณฑ์การตัดสินใจด้วยค่า $p\text{-value}$ ที่มีความคลุมเครือ ประกอบกับการตั้งสมมติฐานหลัก (H_0) ที่ไม่มีทางเป็นจริงตั้งแต่เริ่มแรก จึงเป็นปัญหาของการวิจัยเชิงปริมาณที่จะผลิตแนวคิด/ทฤษฎีที่ผิด ๆ ซึ่งได้จากการสรุปผลการวิจัยจากข้อสมมติฐานที่ไม่ถูกตรวจสอบ กลายเป็นฐาน

ให้กับงานวิจัยชิ้นอื่น ๆ นำไปต่อยอดเรื่อย ๆ จึงเปรียบเสมือนต้นไม้พืชที่ออกดอกออกผลเป็นผลไม้พืชแพร่ขยายพันธุ์ต่อไป

VI. บทบาทนักวิจัยกับการสร้างความคลาดเคลื่อนในการวิจัยเชิงปริมาณ

ปัญหาของการทำงานวิจัยเชิงปริมาณไม่ได้อยู่เพียงแคในระดับของหลักการ อันสืบเนื่องมาจากการประยุกต์ใช้เทคนิคทางสถิติในการตรวจสอบและวิเคราะห์ข้อมูลที่อยู่บนพื้นฐานของความไม่แน่นอนเท่านั้น หากแต่ยังเกิดขึ้นในระดับของการนำไปปฏิบัติ อันเป็นปัญหาที่เกิดขึ้นจากอคติของตัวผู้ศึกษาเองที่ไม่มีความเข้าใจเจตนารมณ์ของผู้สร้างหลักการและแนวคิดที่ตนนำมาใช้ ครอบคลุมไปถึงความทะเยอทะยานของตัวผู้ศึกษาที่ต้องการสร้างข้อค้นพบที่น่าตื่นเต้นและดึงดูดความสนใจจากผู้อื่นในแวดวงวิชาการด้านนั้น ๆ ซึ่งนับได้ว่าเป็นการแสดงถึงความก้าวหน้าขององค์ความรู้ด้วยเช่นกัน ปัญหาของผู้วิจัยที่จะถูกหยิบยกมาเป็นตัวอย่างในบทความนี้มีอยู่ 2 ประเด็น ได้แก่ ความสับสนของนักวิจัยกับการนำเทคนิคการสุ่มกลุ่มตัวอย่างของ Taro Yamane มาใช้ และอคติของผู้ศึกษากับการตีความผลการวิเคราะห์ทางสถิติโดยมีรายละเอียดดังนี้

- ความสับสนของนักวิจัยกับการเลือกใช้เทคนิคการกำหนดขนาดกลุ่มตัวอย่างของ Taro Yamane

เทคนิคการกำหนดขนาดกลุ่มตัวอย่างของ Taro Yamane (1967) เป็นที่รู้จักกันอย่างกว้างขวางในหมู่นักวิจัย โดยเฉพาะผู้ที่ศึกษาในระดับบัณฑิตศึกษาของไทยที่มีการนำเทคนิคนี้มาใช้ เพราะใช้ง่ายและสะดวกสบาย เพียงแค่รู้ขนาดของประชากรที่ศึกษา และกำหนดความคลาดเคลื่อนของการสุ่มตัวอย่างที่ต้องการก็สามารถคำนวณได้ ดังเห็นได้จากสูตร (formula) (Yamane 1967, 886) ที่ว่า

$$n = \frac{N}{1 + Ne^2}$$

เมื่อ n = ขนาดของตัวอย่าง

N = ขนาดของประชากร

e = ความคลาดเคลื่อนของการสุ่มตัวอย่าง

นอกจากนี้ ผู้วิจัยยังสามารถอาศัยจำนวนกลุ่มตัวอย่างตามตารางกำหนดขนาดกลุ่มตัวอย่างสำเร็จรูปที่สร้างขึ้นจากสูตรของ Yamane ได้ทันทีจากตารางที่ 6 (Yamane 1967, 886-887) แต่ทว่าความสะดวกสบายดังกล่าวได้นำมาสู่ปัญหาดังเห็นได้จากบทความ อันเนื่องมาแต่สูตรของยามานะ ของ รศ.ดร. อัจฉรวรรณ งามบุญ (Acharawan Ngamyuan 2011) ซึ่งมีประสบการณ์ควบคุมการทำวิทยานิพนธ์ระดับปริญญาโทที่คณะพาณิชยศาสตร์และการบัญชี มหาวิทยาลัยธรรมศาสตร์ พบว่า นักศึกษาระดับปริญญาโทโดยทั่วไปเลือกใช้สูตรของ Yamane เพื่อหลีกเลี่ยงการคำนวณค่าความแปรปรวน (variance) ซึ่งไม่ทราบว่าจะต้องไปหาจากที่ไหน เนื่องจากยังไม่ได้ลงมือเก็บข้อมูลจริง ๆ (Acharawan Ngamyuan 2011, 48)

อย่างไรก็ตาม การนำสูตรของ Yamane มาใช้จะไม่เป็นปัญหาเลย ถ้าหากนักศึกษาที่นำสูตรดังกล่าวมาใช้ตระหนักถึงที่มาของสูตรและทราบเงื่อนไขที่สอดคล้องกับการใช้ ตามที่ Yamane ได้ชี้แจงไว้ กล่าวคือ สูตรดังกล่าวเหมาะสมกับการวิจัยแบบสำรวจ (survey research) ที่พิจารณาค่าความคลาดเคลื่อนในการสุ่มกลุ่มตัวอย่างแบบสุ่มจากผลลัพธ์การสำรวจ (margin of error) ที่เหมาะสม โดยอาศัยทฤษฎีการสำรวจตัวอย่าง (theory of sample survey) ที่ใช้การสุ่มกลุ่มตัวอย่างประเภทการสุ่มอย่างง่าย (simple random sampling) เป็นพื้นฐาน (Yamane 1967, 133-134) นอกจากนี้ คำถามที่ใช้จะต้องอยู่ในลักษณะคำถามที่เลือกตอบได้อย่างใดอย่างหนึ่งจาก 2 ตัวเลือก (เช่น เห็นด้วย/ไม่เห็นด้วย, ชอบ/ไม่ชอบ, ชาย/หญิง ฯลฯ) ซึ่งจะเป็นการแจกแจงแบบ Bernoulli (Bernoulli distribution) และงานวิจัยที่ใช้สูตรนี้จะต้องสนใจค่าสัดส่วนของประชากร (proportion) อีกด้วย

แต่ในความเป็นจริงกลับพบว่า นักวิจัยโดยเฉพาะที่เป็นนักศึกษาระดับปริญญาโทไม่ได้สนใจเงื่อนไขต่าง ๆ ดังกล่าว เพราะงานวิจัยของตนเองใช้แบบสอบถามที่มีระดับมาตรวัดแตกต่างออกไป กล่าวคือ เป็นคำถามที่มีหลายตัวเลือกแบบ 5 ระดับ (เช่น เรียงลำดับจาก ไม่เห็นด้วยอย่างยิ่งไปจนถึงเห็นด้วยอย่างยิ่ง) หรือที่เรียกว่า Likert scale ซึ่งจะทำให้ตัวแปรสุ่มที่ได้มีการแจกแจงแบบ multinomial (multinomial distribution) และไม่สามารถหาค่าสัดส่วนประชากร (π) ซึ่งตามสมมติฐานของ Yamane กำหนดไว้ว่าจะต้องมีค่าเท่ากับ 0.5 (Yamane 1967, 886) อันเป็นค่าที่ทำให้เกิดความแปรปรวนของสัดส่วนตัวอย่าง (σ_p^2) สูงสุดที่ 0.25 เมื่อขนาดของตัวอย่าง (n) มีค่าสูงสุด (Acharawan Ngamyang 2011, 56)

ความผิดพลาดดังกล่าวในการเลือกใช้เทคนิคการสุ่มกลุ่มตัวอย่างและสูตรการกำหนดขนาดตัวอย่าง จึงทำให้ผลการศึกษาที่ได้มีความคลาดเคลื่อนจากความเป็นจริงสูงมาก ข้อค้นพบที่ได้จึงขาดความน่าเชื่อถือเพราะผู้ศึกษาขาดความรู้และความเข้าใจอย่างแท้จริงในสิ่งที่ตนกำลังทำอยู่ อันนับเป็นปัญหาที่เกิดขึ้นจากตัวนักวิจัยเอง

- อคติของผู้ศึกษากับการตีความผลการวิเคราะห์ทางสถิติ

โดยทั่วไปผู้วิจัยทุกคนต่างมีความต้องการสร้างข้อค้นพบใหม่ ๆ ในประเด็นที่ศึกษา เพื่อเพิ่มความน่าสนใจให้กับงานของตน และเชื่อว่าการกระทำดังกล่าวเป็นการตอ่ยอดองค์ความรู้ที่มีอยู่ให้ขยายออกไปเป็นความก้าวหน้าที่ไม่มีจุดสิ้นสุด ความคิดดังกล่าวของนักวิจัยถือได้ว่าเป็นอคติอย่างหนึ่ง ดังที่ผู้เขียนได้กล่าวถึงไว้ส่วนหนึ่งก่อนหน้านี้แล้วในประเด็นการกำหนดสมมติฐานการวิจัย ซึ่งโดยทั่วไปผู้วิจัยจะนำมาใช้เป็นสมมติฐานทางเลือก (H_a) เพราะสมมติฐานหลัก (H_0) มีโอกาสสูงที่จะถูกปฏิเสธอยู่แล้ว การกระทำดังกล่าวจะทำให้ได้ผลการศึกษามี H_a เป็นข้อสรุปกลายเป็นข้อค้นพบใหม่ ทั้งที่ข้อค้นพบดังกล่าวไม่ได้ถูกตรวจสอบอย่างแท้จริง ตามความคิดการพิสูจน์ผิด (falsification) ในทางวิทยาศาสตร์

อคติของนักวิจัยดังกล่าวถือเป็นแค่เศษเสี้ยวหนึ่งที่เกิดขึ้น เพราะไม่ว่าโดยตั้งใจหรือไม่ตั้งใจก็ตาม ได้เกิดความพยายามที่จะบิดเบือนผลการวิจัย เพื่อให้ได้ผลการศึกษาน่าสนใจและได้รับการยอมรับโดย

อาศัยการโกงค่า p-value ที่ระดับนัยสำคัญ 0.05 ดังที่ อ.ดร. อานนท์ ศักดิ์วรวิชญ์ จากคณะสถิติประยุกต์ ผู้อำนวยการศูนย์คลังปัญญาและสารสนเทศ สถาบันบัณฑิตพัฒนบริหารศาสตร์ ได้นำเสนอไว้ 7 ประการ ตัวอย่างเช่น การเพิ่มขนาดกลุ่มตัวอย่างภายหลังจากวิเคราะห์แล้วไม่พบค่านัยสำคัญ การจงใจเลือกตัดกลุ่มตัวอย่างบางกลุ่มออกเพื่อให้พบค่านัยสำคัญ การกวาด/ค้นหาวิธีการทางสถิติหรือแบบจำลองทางสถิติที่จะทำให้มีค่านัยสำคัญ เป็นต้น (Anon Sakworawitch 2017)

ปัญหาของการยึดติดกับค่านัยสำคัญที่ระดับ 0.05 โดยมองว่าเป็นเป้าหมายของการวิจัยที่จะต้องไปให้ถึง เพื่อให้ได้รับผลการศึกษาที่พึงประสงค์ดังกล่าว ได้กลายเป็นอคติของนักวิจัยที่ทำเพื่อให้ตนได้ลงตีพิมพ์ผลงานในวารสารทางวิชาการ ซึ่งเกิดขึ้นมาเป็นระยะเวลานานนับร้อยปีแล้วนับตั้งแต่ Fisher ได้เสนอเกณฑ์ดังกล่าวในการทดสอบสมมติฐานการวิจัย แต่ในปัจจุบันได้มีนักวิชาการด้านต่าง ๆ ที่ใช้สถิติในการศึกษาเริ่มออกมาตั้งข้อสังเกตและประณามการกระทำ รวมถึงเปิดโปงให้เห็นถึงสภาพปัญหาดังกล่าว (Briggs 2015) ตัวอย่างเช่นคำกล่าวที่ว่า “การโกงค่า P เกิดขึ้นเมื่อนักวิจัยทั้งที่โดยสำนึกและไม่สำนึกได้วิเคราะห์ข้อมูลหลายครั้งหลายครา หรือด้วยหลากหลายหนทาง จนกระทั่งพวกเขาได้รับผลลัพธ์ที่พึงประสงค์ ถ้าการโกงค่า P เกิดขึ้นเหมือน ๆ กัน ผลลัพธ์การศึกษาต่าง ๆ ที่เกินจริงสามารถนำมาสู่ข้อสรุปที่หลอกลวงตาคน แม้ว่าหลักฐานที่ได้จะมาจากการศึกษาหลากหลายแนวทางก็ตาม” (Head 2015)

การตีความผลการตรวจสอบสมมติฐานการวิจัยและการสร้างข้อสรุปในงานวิจัยโดยมีอคติจากการมุ่งเป้าไปที่การได้ค่า p-value ที่ระดับนัยสำคัญ 0.05 ดังกล่าวถือเป็นอคติทางวิชาการที่นำมาสู่การเกิดข้อค้นพบในการศึกษาที่ไม่ใช่ความก้าวหน้าขององค์ความรู้อย่างแท้จริง เพราะเป็นการบิดเบือนหลักการโดยพื้นฐานของการศึกษาแบบวิทยาศาสตร์ที่ไม่เพียงแต่จะต้องมีการพิสูจน์ผิดได้ แต่ยังต้องสามารถทดสอบได้อย่างซ้ำ ๆ ด้วย เพื่อให้มั่นใจว่าผลการศึกษาที่ได้มาไม่ใช่เรื่องบังเอิญ แต่ทว่าผลการศึกษาที่เกิดจากการโกงค่า p-value นั้นกลับไม่สามารถทำการทดสอบซ้ำ ๆ แล้วได้ผลที่เป็นดังเดิมเพื่อยืนยันว่าองค์ความรู้ใหม่นั้นใช้ได้จริง (Halsey et al. 2015, 179 – 185) นี่จึงเป็นอีกปัญหาหนึ่งที่เกิดขึ้นจากพฤติกรรมของนักวิจัยที่นำเทคนิคทางสถิติไปใช้

VII. บทสรุป: ฤกษ์จะล้มลง?

บทความชิ้นนี้ได้แสดงให้เห็นถึงปัญหาที่เกิดขึ้นทั้งในทางหลักการและในทางปฏิบัติของการทำวิจัยเชิงปริมาณ โดยเฉพาะที่เกิดขึ้นจากการใช้ค่า p-value เป็นเกณฑ์การตัดสินใจในการทดสอบสมมติฐานการวิจัย ซึ่งทำให้นักวิจัยไม่ได้มองค่า p-value เป็นแค่เครื่องมือในการวัดค่า หากแต่กลายเป็นเป้าหมายของการศึกษาที่จะต้องทำให้เกิดผลการวิจัยที่มีค่า p-value อยู่ในระดับนัยสำคัญ 0.05 ให้จงได้ เพื่อให้เกิดข้อสรุปผลการศึกษาเป็นข้อค้นพบใหม่ สภาพปัญหาดังกล่าวได้ทำให้คำวิจารณ์ของ Leo Strauss เรื่อง

วิกฤตขององค์ความรู้สมัยใหม่ คู่มือให้นักและใกล้เคียงจะเป็นจริงอย่างยิ่ง เมื่อ “ยักษ์ที่เท้าทำด้วยดินเหนียว” คนนี้มีที่ท่าจะล้มลงเพราะยืนอยู่บนหลักการที่ไม่มั่นคง

ปัญหาดังกล่าวได้ทำให้เกิดการเคลื่อนไหวในแวดวงวิชาการที่อาศัยเทคนิคทางสถิติเป็นเครื่องมือในการสร้างองค์ความรู้โดยเฉพาะในสาขาที่เกี่ยวข้องกับวิทยาศาสตร์โดยทั่วไป จนถึงขั้นที่มีวารสารทางวิชาการบางฉบับสั่งห้ามส่งบทความวิจัยที่ใช้ค่า p -value เป็นเกณฑ์ในการตรวจสอบสมมติฐานการวิจัยเพียงอย่างเดียวลงตีพิมพ์ ดังเช่น วารสารจิตวิทยาสังคมพื้นฐานและประยุกต์ (Basic and Applied Social Psychology)¹² เป็นต้น ขณะที่สมาคมสถิติอเมริกัน (American Statistical Association: ASA) ได้ออกแถลงการณ์เรื่อง นโยบายทางสถิติและค่า p -value ในวันที่ 7 มีนาคม 2016 โดยมีข้อความเผยแพร่เป็นหลักการ จำนวน 6 ข้อ (Wasserstein 2016) ที่มีใจความสำคัญอยู่ที่ความต้องการให้นักวิจัยที่ใช้สถิติในการศึกษา จะต้องมิใช่สถิติอย่างมีคุณภาพและมีความเป็นวิทยาศาสตร์ ที่สามารถทดสอบผลการทดลองซ้ำได้ เพื่อให้ได้ข้อสรุปผลการศึกษาที่ถูกต้อง¹³

การเกิดฐานข้อมูลที่มีขนาดใหญ่ มีความหลากหลายของข้อมูล และมักจะเป็นข้อมูลที่เกิดขึ้นแบบทันทีทันใดเวลานั้น ๆ เช่นข้อมูลผ่านสื่อสังคมออนไลน์ (social media) ซึ่งข้อมูลลักษณะดังกล่าว เรียกว่า “Big data” (Molek 2015) ทำให้ผู้วิจัยสามารถนำข้อมูลไปบริหารจัดการ และคัดกรองเพื่อเลือกข้อมูลไปใช้ประโยชน์ตามต้องการ ดังเช่น การนำข้อมูลพื้นฐานส่วนบุคคลไปวิเคราะห์อย่างลึกซึ้งในการจัดทำนโยบายหาเสียงเลือกตั้งแบบรายบุคคลของ Barack Obama ในการเลือกตั้งประธานาธิบดีสหรัฐอเมริกา ปี 2012 (Atip Assawanun 2014) ซึ่งทำให้ Obama และคณะทำงานรณรงค์หาเสียงเลือกตั้งสามารถออกนโยบาย

¹² ในเวลานั้นบรรณาธิการวารสารฉบับนี้ คือ David Trafimow และ Michael Marks จาก New Mexico State University ได้เผยแพร่บรรณาธิการวารสารฉบับปีที่ 37 (ค.ศ. 2015) เพื่อชี้แจงถึงการสั่งห้ามการใช้ p - value ตามกระบวนการทดสอบนัยสำคัญทางสถิติของสมมติฐานหลัก (the null hypothesis significance testing procedure: NHSTP) แต่ทั้งนี้ได้อนุญาตให้ผู้ส่งบทความลงตีพิมพ์วารสาร โดยใช้ NHSTP ในการนำเสนอผลการศึกษา ซึ่งได้ส่งบทความมาให้กองบรรณาธิการวารสารเผยแพร่ก่อนหน้านี้ สามารถเผยแพร่ผลงานกับทางวารสารได้ ถ้าหากปรับปรุงบทความตามข้อกำหนด ดังเช่น กรณีที่ต้นฉบับของบทความดังกล่าวได้ผ่านการพิจารณาตรวจสอบเบื้องต้นจากกองบรรณาธิการวารสารไปแล้วก่อนหน้านี้ ทางกองบรรณาธิการจะส่งให้ผู้ทรงคุณวุฒิประเมินบทความต่อไปแต่ผู้เขียนบทความต้องปรับแก้บทความด้วยการถอดเอาข้อความทั้งหมดที่เกี่ยวข้องกับเรื่อง NHSTP ออกจากเนื้อหาก่อนที่จะตีพิมพ์เผยแพร่เป็นต้น (Trafimow and Marks 2015, 1-2)

¹³ แถลงการณ์ดังกล่าวเริ่มจากการตั้งคำถามเกี่ยวกับปัญหาการใช้ p - value ที่ระดับนัยสำคัญ 0.05 ไปยัง ASA โดย George Cobb ศาสตราจารย์กิตติคุณสาขาคณิตศาสตร์และสถิติ จาก Mount Holyoke College ในเดือนกุมภาพันธ์ 2014 ซึ่งเป็นผลมาจากการกระตุ้นผ่านบทความทางวิชาการที่ออกมาวิพากษ์วิจารณ์การใช้ค่า p - value ในการทดสอบสมมติฐานกันอย่างกว้างขวางในวารสารวิชาการและ website วิชาการต่าง ๆ ตั้งแต่ พ.ศ. 2010 เป็นต้นมา (Wasserstein and Lazar 2016, 129 – 133)

ที่ตรงใจผู้ลงคะแนนเสียงเลือกตั้งได้มากขึ้น ดังนั้น การเกิดขึ้นของ big data จึงช่วยลดปัญหาการอนุมานจากกลุ่มตัวอย่างไปยังประชากรที่นักสถิติในอดีตเคยประสบ เมื่อต้องทำวิจัยที่มีกลุ่มประชากรขนาดใหญ่

โดยสรุป ถึงแม้ว่าองค์ความรู้แบบวิทยาศาสตร์ที่ใช้เทคนิคทางสถิติเป็นเครื่องมือทดสอบสมมติฐานในการสร้างความก้าวหน้า จะต้องประสบปัญหาจากความพยายามทำนายปรากฏการณ์ที่สนใจศึกษาโดยอยู่บนฐานของความไม่แน่นอน ด้วยการสุ่มกลุ่มตัวอย่างเพียงบางส่วนจากประชากรทั้งหมดที่ศึกษา ซึ่งทำให้มีโอกาสเกิดความคลาดเคลื่อนของข้อมูลที่ไม่มีความเป็นตัวแทนของประชากรทั้งหมด องค์ความรู้ที่ได้จึงไม่ใช่ความจริงที่แน่นอนและการเข้าถึงธรรมชาติอย่างแท้จริง ดังที่ Strauss ได้วิจารณ์ว่าจะเป็นวิกฤตขององค์ความรู้สมัยใหม่ที่เกิดขึ้น โดยเปรียบเทียบกับ “ยักษ์ที่มีเท้าเป็นดินเหนียว” ดังที่ผู้เขียนได้กล่าวไว้ตั้งแต่ต้น

แต่กระนั้นก็ตาม สิ่งที่คุณเขียนจะทิ้งท้ายไว้ให้ได้คิด คือ ในปัจจุบันเทคโนโลยีการติดต่อสื่อสารที่ทันสมัยและเทคนิคการวิเคราะห์ข้อมูลทางสถิติได้ก้าวหน้าขึ้นไปไกลเกินกว่าวิธีที่ Strauss รู้จักแล้ว ทำให้นักสถิติมีหนทางในการเข้าถึงข้อมูลที่มีกลุ่มประชากรขนาดใหญ่ได้ (เกือบ) ทั้งหมด (หรืออาจจะทั้งหมดในกรณีที่สามารติดต่อถึงกันได้ในเครือข่ายอินเทอร์เน็ต) โดยไม่ต้องสิ้นเปลืองทรัพยากรมูลค่ามหาศาลเป็นข้อจำกัด ดังนั้น การสุ่มกลุ่มตัวอย่างจากประชากรจึงไม่มีความจำเป็นอีกต่อไป ปรากฏการณ์ดังกล่าวได้ทำให้ข้อค้นพบของนักสถิติมีความน่าเชื่อถือมากขึ้น เพราะความคลาดเคลื่อนที่เคยเป็นปัญหาในอดีตลดต่ำลงอย่างมาก ถ้าหากทิศทางของการศึกษาแบบวิทยาศาสตร์ด้วยเทคนิคทางสถิติยังคงก้าวหน้าต่อไปได้เรื่อย ๆ เช่นนี้ “ยักษ์” ตอนนี้ยอมไม่ใช่ “ยักษ์ที่มีเท้าทำด้วยดินเหนียว” เหมือนกับที่ Strauss ได้ปรามาสไว้อย่างแน่แท้!

- Acharawan Ngamyang. 2011. "An Nueang Ma Tae Sut Yamane." [Because of Yamane's Formula]. *Journal of Business Administration* 34(131): 46 – 60. (in Thai)
- Anon Sakworawitch. 2017. "Hacking P-value Withikan Kong Hai Phon Kan Totsop Thang Sathiti Mi Nai Samkhan Thang Sathiti 7 Prakan." [Hacking P-value; the 7 Methods of Cheating the Significance of Statistical Results]. *Business Analytics and Intelligence @NIDA*. April 1. Accessed August 29, 2017. <https://goo.gl/nJuyax> (in Thai)
- Atip Assawanun. 2014. "Khwaam Khaochai Thi Phit Phit Kiaokap Big Data Lae Analytics (1)." [Misunderstandings about Big Data and Analytics (1)]. *Bangkok Biz News*, September 30. Accessed August 29, 2017. <https://goo.gl/NrLrbk>. (in Thai)
- Briggs, William M. 2015. *P-Value Hacking Is Finally Being Noticed*. Accessed August 29, 2017. <https://goo.gl/rjt2Jc>.
- Cohen, Jacob. 1994. "The Earth Is Round ($p < .05$)." *American Psychologist* 49(12): 997 – 1003.
- . 1977. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Revised. 1969. New York: Academic Press.
- Ekkapollanut Nuttappattanun. 2017. *Kanwikhro Kanmueang: Khwamru Bueangton Thang Ratthasat*. [Analyzing Politics: An Introduction to Political Science]. Nakhon Pathom: Silpakorn University Printing House. (in Thai)
- Fisher, Ronald A. 1926. "The Arrangement of Field Experiment." *Journal of Ministry of Agriculture of Great Britain* 33: 503 - 513.
- . 1950. *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.
- Haig, Brian D. 2013. "The Philosophy of Quantitative Methods." In *The Oxford Handbook of Quantitative Methods*, edited by Todd D. Little, 7 – 31. Oxford: Oxford University Press.
- Halsey, Lewis G., Douglas Curran-Everett, Sarah L. Vowler, and Gordon B. Drummond. 2015. "The Fickle P Value Generates Irreproducible Results." *Nature Method* 12(3): 179 – 185.
- Hawkesworth, Mary. 2004. "Political Science in a New Millennium: Issues of Knowledge and Power." In *Encyclopedia of Government and Politics*, edited by Mary Hawkesworth and Maurice Kogan, 3 – 34. London: Routledge.
- Head, Megan. 2015. *Scientists Unknowingly Tweak Experiments*. Accessed August 29, 2017. <https://goo.gl/a4jye7>.
- Huff, Darrell. 1954. *How to Lie with Statistics*. New York: W. W. Norton & Company, Inc.
- Keuth, Herbert. 2005. *The Philosophy of Karl Popper*. Cambridge: Cambridge University Press.

- Molek [pseud.]. 2015. "Big Datas Khomun Mahasan Thi Thamhai Kantham Kantalat Nan Atchariya Khuen." [Big Data and Better Marketing]. **MarketingOops!**, January 24. Accessed August 29, 2017. <https://goo.gl/uJpJts>. (in Thai)
- Neyman, Jerzy. 1937. "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability." **Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences** 236(77): 333 – 380.
- Neyman, Jerzy, and Ergon S. Pearson. 1933. "On the Problem of the Most Efficient Tests of Statistical Hypotheses." **Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character** 231: 289 – 337.
- O'Hear, Anthony. 1980. **Karl Popper**. London: Routledge and Kegan Paul.
- Pornlapas Boonsorn. n.d. "Ya Kae Wat." [Medication for Colds]. **Haamor.COM**. Accessed August 28, 2017. <https://goo.gl/QM9cFy>. (in Thai)
- "10 Phon Wichai Witthayasat Sut Plaek Pracham Pi 2013." [10 Strange Scientific Research Findings in 2013]. 2013. **Thairath Online News**, December 27. Accessed August 28, 2017. <https://goo.gl/Cx8mvK>. (in Thai)
- Siegfried, Tom. 2015. "P Value Ban: Small Step for a Journal, Giant Leap for Science." **Science News**, March 17. Accessed August 28, 2017. <https://goo.gl/PQdikD>.
- Smith, Peter K. 2000. "Philosophy of Science and Its Relevance for the Social Sciences," In **Research Training for Social Scientists**, edited by Dawn Burton, 5 – 20. London: SAGE.
- Stigler, Stephen M. 1986. **The History of Statistics: The Measurement of Uncertainty before 1900**. Cambridge, MA: The Belknap Press of Harvard University Press.
- Strauss, Leo. 1983. **Studies in Platonic Political Philosophy**. Chicago: The University of Chicago Press.
- Strauss, Leo. 1989a. **An Introduction to Political Philosophy: Ten Essays**. Detroit: Wayne State University Press.
- Strauss, Leo. 1989b. **The Rebirth of Classical Political Rationalism: An Introduction to the Thought of Leo Strauss**. Chicago: The University of Chicago Press.
- Thanes Rattanakul. 2016. "20 Kan Khonphop Thang Witthayasat Thi Na Chotcham Thisut Haeng Pi 2016." [The 20 Most Noteworthy Scientific Findings in 2016]. **The Matter**, December 22. Accessed August 28, 2017. <https://goo.gl/TfCaGg>. (in Thai)

- Trafimow, David and Michael Marks. 2015. "Editorial." **Basic and Applied Social Psychology** 37(1): 1-2.
- Thompson, Bruce. 2013. "Overview of Traditional/Classical Statistical Approaches." In **The Oxford Handbook of Quantitative Method**, edited by Todd D. Little, 7-25. Oxford: Oxford University Press.
- Tukey, John W. 1991. "The Philosophy of Multiple Comparisons." **Statistical Science** 6(1): 100 – 116.
- Waraporn Waikakul, Rojnarin Komonhirun, and Patthrapon Sonkhammee. 2014. "Kansueksa Khwam Phitphlat Khong Kan Chai Sathiti Samrap Kanwichai Thang Wisan-yi-witthaya." [Statistics in Anesthesia Research: A Review of Most Common Pitfalls]. **Thai Journal of Anesthesiology** 40(1): 64 – 77. (in Thai)
- Wasserstein, Ronald L. 2016. "American Statistical Association Release Statement on Statistical Significance and P-values." **ASA News**, March 7. Accessed August 29, 2017. <https://goo.gl/QNhzdQ>.
- Wasserstein, Ronald L., and Nicole A. Lazar. 2016. "The ASA's Statement on p-Values: Context, Process, and Purpose." **The American Statistician** 70(2): 129 – 133. Accessed August 29, 2017. <https://goo.gl/hf4OMA>.
- Williams, Malcom. 2000. **Science and Social Science: An Introduction**. London: Routledge.
- Yamane, Taro. 1967. **Statistics: An Introductory Analysis**. 2nd ed. New York: Harper & Row.
- Zuckert, Catherine and Michael Zuckert. 2006. **The Truth about Leo Strauss: Political Philosophy and American Democracy**. Chicago: The University of Chicago Press.