Journal of Research Methodology, Volume 11, Number 2 (July - December 1998)

วารสารวิธีวิทยาการวิจัย ปีที่ 11 ฉบับที่ 2 (กรกฎาคม-ธันวาคม 2541)

ทำไม α ต้อง .05 ?

สุชาดา บวรกิติวงศ์

บทคัดย่อ

เป็นที่เชื่อกันว่า R.A. Fisher เป็นบุคคลแรกที่ใช้คำว่า ที่ระดับนัยสำคัญ 5% ในภาษาเขียน (Cowles, M. and Davis, C., 1982) ปัจจุบันระดับนัยสำคัญ 5% ได้ถูกใช้โดยคนทั่วโลก จนกลายเป็นระดับมาตรฐาน บทความนี้มุ่งเน้นที่จะอธิบายว่าทำไมทุกคนจึงใช้ระดับนัยสำคัญ 5% และใช้อย่างไร ผู้เขียนได้เสนอแนะการกำหนดขนาดของ α ที่เหมาะสมกับงานวิจัย การวางแผนงานวิจัยที่ดี และการเขียนรายงานผลทางสถิติไว้ด้วย

Why α is .05 ?

Suchada Bowarnkittiwong

ABSTRACT

According to Cowles, M. and Davis,C. (1982), R.A. Fisher is the first person using the word "at a 5% level of significance" in writing. Currently, the .05 level of significance has become "standard"used by people around the world. This article aims to describe how and why everyone uses the .05 significance level. Ways to determine the size of alpha, plan for good research, and report statistical results are recommended.

บทนำ

ในการทดสอบสมมติฐานทางสถิติ ความคิดของนักวิจัยส่วนใหญ่จะมีอยู่ใน 2 ค่าย ค่ายแรกเป็นกลุ่มที่กำหนด α (significance level) ไว้ล่วงหน้า ค่ายที่สองไม่ได้กำหนด α ไว้ล่วงหน้า แต่จะรายงานค่า p (p-value) หรือค่าสถิติ (statistic value) ที่คำนวณได้แทน

นักวิจัยในค่ายที่หนึ่งจะกำหนด α ไว้ก่อน ค่า α ที่นิยมใช้กันมากคือ .05 .01 และ .10 ตามลำดับแล้วนำค่า p ที่คำนวณได้ไปเปรียบเทียบกับค่า α ที่กำหนด หรือนำค่าสถิติที่คำนวณได้ไปเปรียบเทียบกับค่าวิกฤติที่เปิดจากตาราง ที่ระดับ α กำหนดไว้ ถ้าค่า p ที่คำนวณได้มีค่าน้อยกว่า α หรือค่าสถิติที่คำนวณได้มีค่าตกอยู่ในบริเวณวิกฤติ ผลการทดสอบจะปฏิเสธสมมติฐาน H_0 ที่ตั้งไว้ มิฉะนั้นแล้วจะยอมรับสมมติฐาน H_0 ตัวอย่างเช่น ในการทดสอบความแปรปรวนครั้งหนึ่ง สมมติมีค่า p ที่คำนวณได้เป็น .04 นำไปเปรียบเทียบกับ α ที่ระดับ .05 (top-hit value) เนื่องจาก $p < \alpha$ ดังนั้นจึงปฏิเสธ H_0 สรุปว่า ข้อมูลสนับสนุนสมมติฐานรอง H_a หรือ H_1 ในทางกลับกัน ถ้า p ที่คำนวณได้มีค่าเป็น .06 นำไปเทียบกับ α ที่ระดับ .05 เนื่องจาก $p > \alpha$ ก็จะตัดสินว่ายอมรับสมมติฐาน H_0 สรุปได้ว่าข้อมูลไม่ได้สนับสนุนสมมติฐานรอง H_a หรือ H_1 ที่ระดับนัยสำคัญ 5% เป็นที่น่าสังเกตว่านักวิจัยในค่ายนี้ให้ความสำคัญกับงานวิจัยที่มีค่า p เป็น .04 มากกว่างานวิจัยที่มีค่า p เป็น .06 มาก เนื่องจากงานวิจัยที่มีค่า p เป็น .04 ได้คำตอบว่าค่าพารามิเตอร์เหล่านั้นมีความแตกต่างกันอย่างมีนัยสำคัญ (significant result) แต่งานวิจัยที่มีค่า p เป็น .06 ได้คำตอบว่าค่าพารามิเตอร์เหล่านั้นมีความแตกต่างกันอย่างไม่มีนัยสำคัญ (non-significant result) นักวิจัยกลุ่มนี้จะให้ความสำคัญกับ α ในตำแหน่งไม่ต่อเนื่อง (discrete point) คือ .05 .01 หรือ .10 มากเกินไป จึงมีนักสถิติหลายคนตั้งคำถามว่าทำไม α ที่ระดับ .05 จึงมีความสำคัญมากกว่า α ที่ระดับ .06 หรือ .07 มากมายนัก?

นักวิจัยในค่ายที่สองจะให้ความสำคัญกับ α ทุกค่าเท่ากันหมด คือมอง α เป็นค่าต่อเนื่อง นักวิจัยกลุ่มนี้จะรายงานค่า p ที่คำนวณได้เพื่อให้ผู้อ่านตัดสินใจเองว่า ที่ α ระดับใดจะปฏิเสธ H_0 แต่โดยส่วนใหญ่ผู้วิจัยจะสรุปผลการทดสอบว่าจะปฏิเสธ H_0 ในช่วงใด ตัวอย่างเช่น ถ้า p ที่คำนวณได้มีค่าเป็น .006 จะสรุปว่าปฏิเสธสมมติฐาน H_0 ที่ $p < .01$ ถ้า p ที่คำนวณได้ มีค่าเป็น .015 ก็จะปฏิเสธสมมติฐาน H_0 ที่ $p<.02$ ถ้า p ที่คำนวณได้มีค่าเป็น .045 ก็จะปฏิเสธสมมติฐาน H_0 ที่ $p<.05$ เป็นต้น

ทำไม α ที่ระดับ 5% จึงเป็นที่นิยมกันทั่วโลก?

มีนักวิจัยและนักสถิติจำนวนไม่น้อยที่เชื่อว่า R.A. Fisher เป็นคนแรกที่ใช้ α เท่ากับ 5% หรือ .05 ในความเป็นจริงแล้วมีการใช้ระดับนัยสำคัญในการสรุปผลการทดสอบทางสถิติมาก่อนหน้านั้น (Cowles and Davis,1982) Cowles และ Davis ให้ความเห็นว่า R.A.Fisher เป็นคนแรกที่เขียนบันทึกคำว่าระดับนัยสำคัญ (significance level) เป็นลายลักษณ์อักษร ในการสรุปผลการทดสอบ

สมมติฐานทางสถิติในการวิเคราะห์ความแปรปรวน (Analysis of Variance) ในหนังสือของเขา ชื่อ Statistical Methods for Research Workers ในปี ค.ศ. 1925 Fisher ทำงานในสถานีทดลอง (experiment station) และให้เหตุผลว่า ถ้ามีการทดลอง 5 ครั้ง ที่ให้ผลคลาดเคลื่อน จากการทดลองทั้งหมด 100 ครั้ง ผลการทดลองนั้นก็น่าจะยอมรับได้ Fisher จะสรุปว่าสมมติฐานที่ถูกยอมรับนั้นเชื่อถือได้

เป็นที่ถกเถียงกันในหมู่นักสถิติว่า R.A. Fisher ผู้ให้กำเนิดคำอาถรรพ์ (magic word) คือ α ที่ .05 ว่าเขาใช้ α ในความหมายของค่าต่อเนื่องหรือค่าไม่ต่อเนื่อง Cochran (1976) (อ้างถึงใน Cowles and Davis, 1982) กล่าวว่า Fisher ใช้ α ในความหมายของค่าต่อเนื่องมากกว่า Cochran สรุปจากข้อความตอนหนึ่งในหนังสือของ Fisher ที่กล่าวไว้ว่า

> *If one in twenty does not seem high enough odds, we may, if we prefer it, draw the line at one in fifty (the 2 percent point), or one in a hundred (the 1 percent point). Personally, the writer* ***prefers*** *to set a low standard of significance at the 5 percent point, and ignore entirely all results which fail to reach this level. A scientific fact should be regarded as experimentally established only if a properly designed experiment* ***rarely fails*** *to give this level of significance.* (p. 504)

Cochran กล่าวว่า การที่ Fisher ใช้คำว่า prefers นั้น หมายถึง Fisher เองก็ให้ความสำคัญกับระดับนัยสำคัญที่ระดับต่างๆ มากพอสมควร และงานเขียนของ Fisher โดยส่วนใหญ่ก็สรุปผลที่ระดับนัยสำคัญ ณ ระดับหนึ่งเสมอ แต่ไม่ได้หมายความว่าต้องเป็น α ที่ 5% เสมอไป

จากประสบการณ์การสอนวิชาสถิติทั้งในประเทศไทยและในสหรัฐอเมริกา ผู้เขียนสังเกตว่านักวิจัยในเมืองไทยส่วนใหญ่จะอยู่ในค่ายที่หนึ่ง คือนำค่า p ที่คำนวณได้ไปเทียบกับค่า α ที่กำหนดไว้แล้ว หรือนำค่าสถิติที่คำนวณได้ไปเทียบกับค่าวิกฤติที่เปิดจากตาราง ตามระดับ α ที่กำหนดไว้ ในขณะที่นักวิจัยในสหรัฐอเมริกามีแนวโน้มอยู่ในค่ายที่สอง คือ รายงานค่า p ที่คำนวณได้ (obtained p-value) พร้อมทั้งสรุปว่าจะปฏิเสธ H_0 ในช่วงใด (ไม่ได้กำหนด α ไว้ล่วงหน้า)

ความจริงแล้วค่าที่สำคัญในการเขียนสรุปผลการทดสอบสมมติฐานทางสถิติไม่ได้มีเฉพาะค่า α หรือค่า p เท่านั้น ยังมีคำถามต่างๆ อีกมากมายที่ถูกมองข้าม เช่น งานวิจัยลักษณะใดควรจะกำหนด α ขนาดใหญ่ ลักษณะใดควรใช้ α ขนาดเล็ก ปัจจัยอะไรบ้างที่มีผลต่ออำนาจการทดสอบ (power of the test) จะออกแบบงานวิจัยอย่างไรให้มีอำนาจการทดสอบสูง และมีค่าความคลาดเคลื่อนชนิดที่ 1 ต่ำ การทดสอบสมมติฐานทางสถิติในลักษณะใดควรเน้นความคลาดเคลื่อนชนิดที่ 1 ในลักษณะใดควรเน้นความคลาดเคลื่อนชนิดที่ 2 เป็นต้น ผลงานวิจัย

ที่รายงานเฉพาะยอมรับหรือปฏิเสธสมมติฐาน H_0 เท่านั้นจะให้คุณค่าแก่ literature น้อยมาก

สิ่งหนึ่งที่ควรระลึกถึงเสมอคือในการทดสอบสมมติฐานทางสถิติทุกครั้ง จะเกิดความคลาดเคลื่อนได้ 2 ชนิด คือ ความคลาดเคลื่อนชนิดที่ 1 หรือ Type I error ซึ่งจะเกิดขึ้นเมื่อ นักวิจัยปฏิเสธสมมติฐานที่เป็นจริง (reject a true null) และความคลาดเคลื่อนชนิดที่ 2 หรือ Type II error จะเกิดขึ้นเมื่อนักวิจัยยอมรับสมมติฐานที่เป็นเท็จ (fail to reject a false null) ในทางปฏิบัตินิยมกำหนดสัญลักษณ์ α แทนโอกาสที่จะเกิดความคลาดเคลื่อนชนิดที่ 1 (probability of rejecting a true null) กำหนด β แทนโอกาสที่จะเกิดความคลาดเคลื่อนชนิดที่ 2 (probability of accepting a false null) และกำหนด $1-\beta$ แทนอำนาจการทดสอบ หรือ power of the test (probability of correctly rejecting a false null) โดยทฤษฎีค่า α และ β จะแปรผกผันซึ่งกันและกัน นั่นคือ ถ้า α มีค่าเล็ก β จะมีค่าใหญ่ และถ้า β มีค่าเล็ก α จะมีค่าใหญ่ ดังนั้น ถ้าผู้วิจัยต้องการควบคุมให้ทั้ง α และ β มีค่าเล็ก จะทำได้โดยการเพิ่มขนาดตัวอย่าง

เป็นที่น่าประหลาดใจว่า ในการทดสอบสมมติฐานทางสถิติแต่ละครั้ง ผู้วิจัยมักจะคำนึงถึงแต่ α และระดับของ α ที่นิยมใช้จะเป็นแบบอัตโนมัติ คือ 5% หรือ 1% ข้อเท็จจริงอันนี้ทำให้ผู้เขียนคิดว่าหลังจากคำนวณค่าสถิติทดสอบและได้ค่า p แล้ว นักวิจัยจะนำค่า p ที่คำนวณได้ไปเทียบกับ α ที่ 5% หรือ 1% แล้วสรุปผลทันทีว่าจะยอมรับหรือปฏิเสธสมมติฐาน H_0 โดยที่นักวิจัยไม่ได้หยุดคิดเลยสักนิดว่า เกณฑ์เหล่านี้มีความเหมาะสมกับงานวิจัยของตนแค่ไหน เป็นเรื่องที่น่าเป็นห่วงอย่างยิ่งว่าทำไมต้อง 5% และทำไม α ที่ระดับ 5% จึงกลายเป็นระดับมาตรฐานของโลกไปแล้ว

ในฐานะที่สอนวิชาสถิติและมีประสบการณ์ในงานวิจัยมาบ้าง ผู้เขียนอยากจะฝากข้อเตือนใจไว้ว่าไม่ใช่ทุกงานวิจัยที่ α = 5% หรือ .05 จะใช้ได้ดีเสมอไป และเป็นการแน่นอนว่าแนวความคิดในการรายงานค่า p ไม่ใช่ของใหม่ ผู้เขียนกลับมีความเห็นว่า การที่รายงานค่า p เพียงอย่างเดียว โดยที่ผู้วิจัยไม่ได้ระบุว่าควรใช้ α ขนาดเท่าไรในงานวิจัยนั้น ผู้อ่านหรือผู้ที่จะนำผลงานวิจัยนั้นไปใช้จะไม่สามารถวิเคราะห์ได้เลยว่างานวิจัยนั้นควรจะ accept หรือ reject H_0 ที่ α ระดับใด กลับจะเป็นการซ้ำเติมผู้อ่านมากขึ้นไปอีก เพราะการรายงานเฉพาะค่า p เพียงอย่างเดียวจะทำให้ผู้อ่านได้ information น้อยลงไปอีก เพราะผู้อ่านไม่แน่ใจเกี่ยวกับการออกแบบงานวิจัย (research design) ไม่แน่ใจเกี่ยวกับระดับของการควบคุมตัวแปรแทรกซ้อน ฉะนั้น ทางออกที่ดีทางหนึ่งคือ ควรจะรายงานทั้ง 2 ค่า คือค่า p ที่เกิดขึ้นจริง และค่า α ที่ควรใช้ในงานวิจัยนั้น อันจะทำให้ภาพของงานวิจัยสมบูรณ์ขึ้น ชัดเจนมากขึ้น

α ควรมีขนาดเท่าไร?

ผู้เขียนขอเสนอแนวทางกว้างๆ ในการกำหนดขนาดของ α ดังนี้

1. การนำผลการทดสอบไปใช้หลังการทดลองหรือผลสืบเนื่องในทางปฏิบัติที่จะตามมา เช่น

ถ้าผู้วิจัยต้องการทดสอบประสิทธิภาพของยาชนิดหนึ่งว่าสามารถรักษาโรคได้จริงหรือไม่ ดังนั้น

Type I error rate = α = โอกาสที่จะปฏิเสธ H_o ที่เป็นจริง
Type II error rate = β = โอกาสที่จะยอมรับ H_o ที่เป็นเท็จ
Power of the test = $1-\beta$ = โอกาสที่จะปฏิเสธ H_o ที่เป็นเท็จ

ผู้วิจัยอาจเริ่มต้นด้วยการถามตัวเองว่า ความคลาดเคลื่อนประเภทไหนที่ควรให้ความสำคัญ หรือควรควบคุมมากกว่า α ในที่นี้จะหมายถึง โอกาสที่จะปฏิเสธยาที่รักษาโรคได้ β หมายถึง โอกาสที่จะยอมรับยาที่ไม่สามารถรักษาโรคได้ ถ้าผู้วิจัยเห็นว่างานวิจัยนี้ควรเน้น β เพราะการยอมรับยาที่ไม่มีประสิทธิภาพในการรักษาโรค น่าจะมีความเสี่ยงมากกว่าการปฏิเสธยาที่มีประสิทธิภาพ ถ้าเป็นเช่นนี้ก็ควรให้ β มีค่าเล็ก (เช่น $\beta = .001$) และ α มีค่าใหญ่ (เช่น $\alpha = .10$) เป็นต้น

คราวนี้สมมติว่ายานั้นเป็นยาวิตามินชนิดหนึ่งที่ถูกโฆษณาว่าสามารถทำให้ผู้ที่รับประทานอย่างสม่ำเสมอ แลดูอ่อนกว่าวัย

α = โอกาสที่จะปฏิเสธยา โดยที่ในความเป็นจริงแล้ว ยาวิตามินนี้ช่วยให้ผู้รับประทานอย่างสม่ำเสมอดูอ่อนกว่าวัยจริง

β = โอกาสที่จะยอมรับยา เมื่อในความจริงแล้ว ยาวิตามินนี้ไม่ช่วยให้ผู้รับประทานอย่างสม่ำเสมอดูอ่อนกว่าวัย

$1-\beta$ = โอกาสที่จะปฏิเสธยา เมื่อในความเป็นจริงแล้ว ยาวิตามินนี้ไม่ช่วยให้ผู้รับประทานอย่างสม่ำเสมอดูอ่อนกว่าวัย

หัวใจของการทดสอบสมมติฐานทางสถิติคือ นักวิจัยต้องการผลการทดสอบที่มีค่า α ต่ำ และ $1-\beta$ สูง

เมื่ออ่านมาถึงตรงนี้ ท่านคิดว่างานวิจัยนี้ควรเน้นการควบคุม α หรือ β เรื่องนี้อาจจะไม่ใช่เรื่องรุนแรงหรือสำคัญมากนักกับคนทั่วไปเกี่ยวกับการกำหนด α และ β ถ้าหากเปลี่ยนยาวิตามินนี้เป็นยาลดความอ้วน ก็อาจจะยังไม่สำคัญเช่นกัน แต่ถ้าเป็นยารักษาโรค AIDS หรือ HIV หรือมะเร็ง ที่มีความเสี่ยงต่อชีวิตสูงมาก และเสียค่าใช้จ่ายแพงมาก ท่านก็จะพอมองออกมากขึ้นว่าความคลาดเคลื่อนชนิดที่ 1 หรือ 2 ที่จะสำคัญมากกว่า การที่ท่านคิดว่าท่านควรจะให้ความสำคัญกับความคลาดเคลื่อนชนิดใดมากกว่า ก็คือ ท่านจะยอมรับความเสี่ยงแบบไหนน้อยกว่านั่นเอง

งานวิจัยในสาขาสังคมศาสตร์ที่พบมากที่สุดเรื่องหนึ่ง คือ การเปรียบเทียบการสอน 2 วิธี ว่าให้ผลต่างกันหรือไม่

α = prob. ที่จะปฏิเสธว่าการสอนทั้ง 2 วิธี ให้ผลไม่ต่างกัน เมื่อในความเป็นจริงแล้ว การสอนทั้ง 2 วิธี ให้ผลไม่ต่างกัน

β = prob. ที่จะยอมรับว่าการสอนทั้ง 2 วิธี ให้ผลไม่ต่างกัน เมื่อในความเป็นจริงแล้ว การสอนทั้ง 2 วิธี ให้ผลต่างกัน

ในฐานะเป็นผู้ทดลอง ท่านอยากจะควบคุมความเสี่ยงตัวไหนมากกว่า ถ้าท่านต้องการ β เล็ก ท่านก็ต้องกำหนด α ขนาดใหญ่ เป็นต้น นักวิจัยที่ดีจะระมัดระวังในการออกแบบงานวิจัยมาก นักวิจัยจะทราบว่าตนกำลังต้องการควบคุมความคลาดเคลื่อนชนิดใดมากกว่า จะเห็นว่า α = .05 ไม่ควรจะเป็นค่าอาถรรพ์ (magic value) หรือค่าอัตโนมัติ (automatic value) อีกต่อไป บางงานวิจัย α อาจจะมีค่ามากถึง .2 ก็ได้ แต่บางงานวิจัย α ต้องมีค่าน้อย อาจเป็น .001 ก็ได้

2. วัตถุประสงค์ของงานวิจัย

บางครั้งผู้วิจัยต้องการตรวจสอบว่า ผลงานวิจัยนั้นยืนยัน (confirm) ทฤษฎี (theory) หรือตัวแบบ (model) ที่มีคนคิดไว้แล้วหรือไม่ การทำงานวิจัยเพื่อสนับสนุนทฤษฎี หรือ model ที่มีอยู่แล้ว อาจใช้ α ขนาดใหญ่ได้ แต่ถ้าต้องการสร้างทฤษฎีใหม่ หรือ model ใหม่ ควรกำหนด α ขนาดเล็ก

3. อำนาจการทดสอบ (power of the test)

ดังที่ได้กล่าวมาก่อนหน้านี้ว่าอำนาจการทดสอบ คือ โอกาสที่จะปฏิเสธสมมติฐานที่ผิด ที่ผู้วิจัยต้องการให้มีค่าสูง ๆ ในทางทฤษฎี อำนาจการทดสอบจะแปรผันตรงกับขนาดตัวอย่าง ดังนั้น ถ้าขนาดตัวอย่างใหญ่ขึ้น อำนาจการทดสอบจะสูงขึ้นด้วย ในขณะเดียวกันจะมีผลทำให้ α สูงขึ้นด้วย อันเป็นผลเนื่องมาจากค่า standard error ซึ่งจะแปรผกผันกับขนาดตัวอย่าง อันจะทำให้ α มีค่ามากเมื่อขนาดตัวอย่างใหญ่ ถ้าขนาดตัวอย่างเล็ก โอกาสที่จะยอมรับสมมติฐาน H_0 จะมาก ดังนั้นในทางปฏิบัติ

α ขนาดเล็ก (.01, .001) ควรใช้ควบคู่กับขนาดตัวอย่างใหญ่

α ขนาดใหญ่ (.10, .15) ควรใช้ควบคู่กับขนาดตัวอย่างเล็ก

ผู้อ่านคงจำได้ว่า α และ β จะแปรผกผันกัน ถ้าลด α จะเป็นการเพิ่ม β ดังนั้น α ที่ .05 จะให้ β ต่ำกว่า α ที่ .01

4. ระดับการควบคุมตัวแปรต่าง ๆ ใน design

การที่ R.A. Fisher ตั้ง α ไว้สูงที่ .05 แทนที่จะเป็น .01 หรือ .001 นั้น เป็นเพราะเขาทำการทดลองในห้องทดลอง เขาสามารถควบคุมตัวแปรแทรกซ้อนและตัวแปรที่กำลังศึกษาได้อย่างค่อนข้างเต็มที่ ดังนั้น α ที่มีขนาดใหญ่ก็น่าจะยอมรับได้ สำหรับ design ที่ไม่สามารถควบคุมตัวแปรและปัจจัยได้เต็มที่เหมือนการทดลองในห้อง lab เพื่อให้ผู้วิจัยมั่นใจในผลการวิจัย ควรใช้ α ขนาดเล็ก

5. ความถูกต้องของผลการทดสอบ (robustness of the test)

ถ้าต้องการให้ผลการทดสอบมีความถูกต้อง (robust) ควรใช้ α ขนาดเล็ก พูดอีกนัยหนึ่งคือ ถ้าลักษณะของข้อมูลมีความสอดคล้องกับข้อตกลงเบื้องต้น (assumptions) ของตัวสถิติทดสอบ α ขนาดใหญ่ก็น่าจะยอมรับได้ ในทางกลับกัน ถ้าลักษณะข้อมูลมีความสอดคล้องน้อยหรือไม่สอดคล้องกับ assumptions ก็ควรใช้ α ขนาดเล็ก

6. การทดสอบแบบ 1 หาง หรือ 2 หาง (one-tailed or two-tailed test)

ผู้เขียนได้รับคำถามทั้งจากนิสิตและจากเพื่อนนักวิจัยบ่อยครั้งว่า การทดสอบนี้ควรเป็น 1 หาง หรือ 2 หางดี คำตอบสำหรับคำถามนี้คือ คงขึ้นอยู่กับคำถามวิจัย (research question) ของงานวิจัยนั้น ๆ ถ้าผู้วิจัยสนใจจะทดสอบในทิศทางน้อยกว่าอย่างเดียว (one-tailed) หรือมากกว่าอย่างเดียว (one-tailed) หรือไม่สนใจทิศทาง (two-tailed) เช่น บทความหนึ่งกล่าวสรุปไว้ว่า นิสิตที่สามารถเรียนจบระดับปริญญาตรีได้ จะมี I.Q. เฉลี่ยประมาณ 110 สมมติว่าท่านสุ่มนิสิตที่กำลังเรียนระดับปริญญาเอก ชั้นปีที่ 2 มา 50 คน ท่านคิดว่าถ้าตั้ง H_0 : I.Q. เฉลี่ยเป็น 110 H_a : ควรมีทิศทางหรือไม่ และถ้ามีควรเป็นทิศทางมากกว่าหรือน้อยกว่า เป็นต้น สิ่งสำคัญประการหนึ่งคือ การทดสอบแบบ one-tailed จะมีโอกาสที่จะ reject H_0 ได้ง่ายกว่าการทดสอบแบบ two-tailed

นักวิจัยบางท่าน รวมทั้งนักวิจัยในอเมริกาเอง มีความเข้าใจผิด ๆ ว่า ถ้าผลงานวิจัยออกมาว่าปฏิเสธ H_0 หรือที่เรียกว่า got a significant result หรือที่คนไทยเรียกสั้น ๆ ว่า ได้ผล significant แล้ว งานวิจัยนั้นจะทำประโยชน์ต่อวรรณคดี (literature) มากกว่าผลงานวิจัยที่ยอมรับ H_0 หรือที่เรียกว่าได้ผลไม่ significant และโอกาสที่งานวิจัยที่ให้ผลต่างกันอย่างมีนัยสำคัญ (significant result) จะถูกยอมรับตีพิมพ์ลงใน Journal มากกว่างานวิจัยที่ให้ผลไม่ต่างกันอย่างมีนัยสำคัญ (non-significant result) ในอดีตความเชื่อเหล่านี้มีแนวโน้มค่อนข้างสูงว่าเป็นความจริง แต่ปัจจุบันความเชื่อเหล่านี้กำลังถูกเปลี่ยนไป โดยส่วนตัวผู้เขียนมีความเห็นว่างานวิจัยที่ให้ผลทั้ง significant และไม่ significant จะส่งผลหรือเป็นประโยชน์ต่อวรรณคดี (literature) หรือผู้สนใจในศาสตร์นั้น ๆ ไม่ต่างกัน เพราะทั้ง 2 ผลงานวิจัยได้ให้ information แก่ผู้สนใจในเรื่องนั้นได้พอสมควร สิ่งหนึ่งที่ผู้วิจัยควรคิดให้มากขึ้น คือ จะทำอย่างไรให้ได้ผลงานวิจัยที่สะท้อนความเป็นจริงมากที่สุด

ขั้นตอนต่าง ๆ ที่ควรคิดพิจารณาก่อนทำวิจัย

1. ผู้วิจัยต้องมีความเข้าใจในงานวิจัยนั้น ๆ อย่างแท้จริง เข้าใจถึงปัจจัยเชิงสาเหตุ ปัจจัยแทรกซ้อน รวมถึงตัวแปรต่าง ๆ ที่เกี่ยวข้อง หรือที่อาจเกี่ยวข้องได้เป็นอย่างดี

2. ผู้วิจัยควรมี concepts ของตัวสถิติทดสอบที่เลือกไว้ เข้าใจถึงข้อตกลงเบื้องต้นของตัวสถิตินั้นว่ามีอะไรบ้าง ข้อมูลที่มีอยู่สอดคล้องกับข้อตกลงเบื้องต้นที่มีอยู่หรือไม่ และจะทำการทดสอบอย่างไร การอ่านผลลัพธ์ที่ได้จากคอมพิวเตอร์อย่างถูกต้องก็เป็นสิ่งสำคัญมาก

3. ผู้วิจัยสามารถเสนอแนะได้ว่า งานวิจัยนั้นควรใช้ α ที่ระดับใด เพราะเหตุไร

4. ผู้วิจัยสามารถออกแบบงานวิจัยโดยคำนึงถึงอำนาจการทดสอบ โดยดูจากปัจจัยต่าง ๆ ที่จะมีผลต่ออำนาจการทดสอบ เช่น การกำหนดขนาดตัวอย่าง การกำหนดเทคนิคการสุ่มตัวอย่าง การทำซ้ำและผลที่จะได้รับจากการทำซ้ำ การควบคุมตัวแปรแทรกซ้อน คุณภาพของเครื่องมือที่จะวัด สถิติที่จะใช้ทดสอบ เป็นต้น

ควรรายงานอะไรบ้างในการสรุปผลทางสถิติ?

เพื่อให้ผลงานวิจัยมีคุณค่ามากขึ้นผู้เขียนขอถือโอกาสนี้เสนอแนะการเขียนสรุปผลงานวิจัยว่าผู้วิจัยควรจะรายงานสถิติภาคบรรยาย (descriptive statistics) ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ขนาดตัวอย่าง องศาแห่งความเป็นอิสระ (degrees of freedom) ค่าสหสัมพันธ์ ค่าสถิติทดสอบที่ใช้ ค่า p ที่คำนวณได้ (ถ้าเป็นไปได้รวมถึงช่วงความเชื่อมั่นของ p) ขนาดของ α ที่เหมาะสมในงานวิจัยนั้น ค่าขนาดอิทธิพล (effect size) พร้อมทั้งความคิดเห็น (judgement) และข้อเสนอแนะของผู้วิจัย เพื่อประโยชน์สูงสุดต่อผู้สนใจ และผู้ที่จะนำผลวิจัยไปใช้ต่อไป

เอกสารอ้างอิง

Cowles, M., & Davis, C. (1982). On the origins of the .05 level of statistical significance. ***Journal of American Psychologist, 37(5),*** 553-558.