บทที่ 5
การทดสอบสมมุติฐาน

5.1 บทนำ

งานหลักของวิชาการทางสถิติคือการสร้างโมเดลทางสถิติขึ้นมาโดยอาศัยข้อมูลและข้อสนเทศจากกลุ่มตัวอย่าง[1] แล้วนำโมเดลนั้นไปใช้งานในด้านการคาดหมายหรือพยากรณ์ผลลัพธ์ที่จะพึงบังเกิดขึ้นในอนาคตสำหรับสถานการณ์นั้น ๆ กระบวนการในการศึกษาวิชาสถิติในลักษณะนี้เรียกว่า "สถิติอนุมาน" (Statistical Inference)

ในงานสถิติอนุมานนั้น งานหลักที่สำคัญสำหรับสายนี้คือการกะประมาณค่าพารามิเตอร์ (Estimation of Parameter) และการทดสอบข้อสมมุติฐาน (Test of Hypothesis หรือ Hypothesis Testing) การกะประมาณค่าพารามิเตอร์และการทดสอบข้อสมมุติฐาน มีส่วนผูกพันกันอยู่ในลักษณะของความต่อเนื่องหรือแม้แต่ใช้แทนกันได้ในบางสถานการณ์ เช่นการกะประมาณด้วยช่วงเชื่อมั่น (Interval Estimate) สามารถใช้แทนการทดสอบสมมุติฐาน เป็นต้น แต่อย่างไรก็ตามในหลักทางทฤษฎีวิธีการทั้งสองมีวิธีการพัฒนาขึ้นมาในลักษณะที่ต่างกัน แต่สิ่งหนึ่งที่ใช้ร่วมกันอย่างจะขาดเสียมิได้ก็คือ การแจกแจงของตัวสถิติหรือตัวแทน (Sampling Distribution)[2] และด้วยเหตุที่สถิติอนุมานเป็นเรื่องที่ต้องอาศัยทฤษฎีการแจกแจงของตัวสถิติ (Statistics) โดยตลอด เราจึงมักจัดจำแนกสถิติอนุมานนี้ไว้ในส่วนที่ว่าด้วยการศึกษาทฤษฎีของตัวสถิติ (Theory of Statistics)

---

[1] ข้อมูลสามารถเก็บได้ใน 2 ลักษณะคือข้อมูลสำรวจหรือข้อมูลที่ได้จากการสังเกต (Observed Data) กับข้อมูลที่ได้จากการทดลอง (Experimental Data) นักวิจัยสามารถเลือกใช้ข้อมูลประเภทใดก็ได้ ทั้งนี้ขึ้นอยู่กับลักษณะของงานวิจัยเป็นสำคัญ เทคนิคการเก็บรวบรวมข้อมูลทั้งสองประเภทตลอดจนเทคนิคการวิเคราะห์ นักศึกษาสามารถศึกษาได้จากวิชาเทคนิคการสำรวจด้วยตัวอย่างและวิชาเทคนิคการวางแผนการทดลองและวิจัย

[2] ได้กล่าวถึงการแจกแจงของตัวสถิติไว้ค่อนข้างจะละเอียดในหนังสือทฤษฎีสถิติ 2 โดยผู้เขียนคนเดียวกัน

ในงานทางปฏิบัติด้านสถิติอนุมานนั้น บางครั้งเป็นเรื่องของการประมาณค่าพารามิเตอร์เพียงอย่างเดียว แต่ขณะเดียวกัน นักวิจัยอาจมองลึกไปถึงการเปรียบเทียบด้วยว่า ค่าที่ประมาณได้นั้นมีลักษณะแตกต่างไปจากค่าประมาณของกลุ่มประชากรย่อย (Subpopulation) อื่น ๆ หรือกลุ่มประชากรเดียวกันแต่ในต่างกาลต่างวาระกันหรือไม่ เมื่อความสนใจของนักวิจัยหันเหมาทางด้านการศึกษาเปรียบเทียบ ความจำเป็นที่นักวิจัยพึงศึกษาต่อไปก็คือการทดสอบข้อสมมุติฐาน เช่น นักวิจัยทำการศึกษา IQ ของนักศึกษาวิชาเอกคณิตศาสตร์ พบว่าโดยถัวเฉลี่ยแล้วนักศึกษาวิชาเอกนี้มี IQ ประมาณ 110[1] การกระทำเช่นนี้เรียกว่า การประมาณค่าพารามิเตอร์โดยที่พารามิเตอร์คือ IQ โดยถัวเฉลี่ยของนักศึกษาวิชาเอกคณิตศาสตร์ ค่า 110 คือค่ากะประมาณ (Estimates) แต่ถ้านักวิจัยพิจารณาแล้วเห็นว่าการได้คำตอบเพียงเท่านี้ดูจะน้อยเกินไปไม่คุ้มกับการลงทุนลงแรงวิจัย และเห็นว่า น่าจะได้ศึกษาเปรียบเทียบว่า IQ ของนักศึกษาวิชาเอกคณิตศาสตร์ในระหว่างกลุ่มอายุต่าง ๆ ในระหว่างเพศ ในระหว่างพื้นฐาน การศึกษาระดับมัธยมศึกษา (เช่น มศ.5 อาชีวศึกษา ปกศ.) ในระหว่างภูมิภาคและอื่น ๆ จะต่างกันหรือไม่ กลุ่มอายุใด IQ สูงกว่าเพศไหน IQ สูงกว่าหรือว่านักศึกษาจากภูมิภาคใด IQ สูงกว่า ดังนี้ก็หมายความว่า นักวิจัยกำลังก้าวล้ำเข้ามาในวิชาสถิติอนุมาน สาขาการทดสอบสมมุติฐานแล้ว ปัญหาก็มีอยู่ว่าจะตอบคำถามเช่นนั้นได้อย่างไร มีวิธีการเสาะหาคำตอบอย่างไร มีข้อจำกัดหรือข้อตกลงเพิ่มเติมหรือไม่ ควรวางแผนการเสาะหาคำตอบอย่างไร ถ้า IQ โดยถัวเฉลี่ยของนักศึกษาชายเท่ากับ 115 และ IQ โดยถัวเฉลี่ยของนักศึกษาหญิงเท่ากับ 100 จะสรุปได้หรือไม่ว่านักศึกษาชายมี IQ สูงกว่า จะใช้กฎเกณฑ์ใดเป็นหลักในการตัดสินใจ ตัดสินใจแล้วจะทราบได้อย่างไรว่าตัดสินใจผิดหรือถูก มีอะไรเป็นหลักประกันความ "เชย" หรือ "เฉียบแหลม" ของงานและประการที่สำคัญที่สุดก็คือเพราะเหตุใดจึงใช้กติกานั้น ๆ เป็นเครื่องช่วยในการตัดสินใจ

ที่กล่าวมาทั้งหมดนี้ นักศึกษาคงพอจะมองเห็นเค้าโครงและปัญหาของงานวิจัยหรือสถิติอนุมานได้พอลาง ๆ แล้ว และคงจะเข้าใจความผูกพันระหว่างการประมาณค่าพารามิเตอร์และการทดสอบสมมุติฐานได้พอสมควร ปัญหาต่อไปจึงเป็นเรื่องของการศึกษาและพัฒนาตัวทดสอบ (Test) สำหรับสมมุติฐานชนิดต่าง ๆ ตลอดจนเงื่อนไขและหลักประกันแห่งความถูกต้องของการตัดสินใจ

---

[1] เทคนิคการสร้างตัวประมาณค่า (Estimator) และการประมาณค่าพารามิเตอร์นักศึกษาสามารถศึกษาได้จากวิชาทฤษฎีสถิติ 1 (Estimation) และเทคนิคการสำรวจด้วยตัวอย่าง

5.2 ความสำคัญของข้อสมมุติฐาน

ข้อสมมุติฐานคือข้อสงสัยหรือข้อคิดคำนึงที่เกี่ยวกับธรรมชาติของกลุ่มประชากรที่เรากำลังสนใจศึกษา เมื่อเป็นเช่นนี้ ข้อสมมุติฐานจึงมีอิทธิพลต่อกระบวนการวิจัยเป็นอย่างยิ่ง เพราะข้อสงสัยหรือข้อคำนึงดังกล่าวจะเป็นตัวจุดความคิดคำนึงของนักวิจัยให้เสาะหาหนทางตอบคำถามในความสงสัยหรือความคิดคำนึงดังกล่าว หรือกล่าวได้ว่า ข้อสมมุติฐานจะเป็นตัวการที่ช่วยให้นักวิจัยวางแผน (Planning) และกำหนดแนวทาง (Guiding) ในการหาคำตอบ หมายความว่าข้อสมมุติฐานที่ตั้งขึ้นจะช่วยให้นักวิจัยวางแผนเก็บรวบรวมข้อมูล วางแผนออกแบบสำรวจที่เหมาะสม วางแผนวิเคราะห์ข้อมูลหรือจัดจำแนกข้อมูลวางแผนจำกัดขอบเขตของการวิจัยและวางแผนลดหรือตัดทอนข้อมูลที่ไม่จำเป็นออกจากแบบสอบถามหรือแผนวิเคราะห์ข้อมูล กล่าวเช่นนี้นักศึกษาอาจมองไม่เห็นภาพ และความสำคัญของข้อสมมุติฐานได้ชัดเจนพอ จึงขอยกตัวอย่างประกอบการพิจารณาดังนี้ สมมติว่า นักวิจัยมีความสงสัยว่า นักศึกษาในมหาวิทยาลัยเปิดในระดับชั้นปีต่างกันจะมีทัศนคติต่อตัวอาจารย์ผู้สอนต่างกันหรือไม่? และจากการเฝ้าติดตามสังเกตหรือจำเจอยู่กับสถานการณ์ในมหาวิทยาลัยเปิดมานานพอควร นักวิจัยรู้สึกว่า นักศึกษาในมหาวิทยาลัยเปิด ยิ่งอยู่ในมหาวิทยาลัยหลายปีเพียงใดจะยิ่งห่างเหินและขาดความสัมพันธ์กับอาจารย์มากเพียงนั้น เราจึงตั้งข้อสงสัยหรือข้อคิดคำนึงว่า "นักศึกษาในมหาวิทยาลัยเปิดในระดับชั้นปีที่ต่างกันจะมีทัศนคติต่อตัวอาจารย์ผู้สอนแตกต่างกัน" หรือ "นักศึกษาในมหาวิทยาลัยเปิดในชั้นปีสูงกว่าจะมีทัศนคติที่ไม่ดีต่อตัวอาจารย์ผู้สอนมากกว่านักศึกษาในชั้นปีที่ต่ำกว่า" ความสงสัยหรือความคิดคำนึงเช่นนี้เรียกว่า ข้อสมมุติฐานเพื่อการวิจัย ความสงสัยหรือข้อคิดคำนึงนี้ไม่จำเป็นต้องถูกต้อง และจะยังไม่มีการตัดสินใจว่าถูกหรือผิด จนกว่าจะมีการทดสอบเสียก่อน จากข้อสมมุติฐานที่ตั้งขึ้นนี้ นักวิจัยก็จะมีขอบเขตของการวิจัยเพียงเฉพาะนักศึกษาในมหาวิทยาลัยเปิดเท่านั้น นักศึกษาในมหาวิทยาลัยปิดอื่น ๆ จะถูกตัดออกจากการพิจารณาหรือกรอบตัวอย่าง (Sampling Frame) ปัญหาต่อไปจึงเป็นเรื่องของการวางแผนวิจัย เริ่มต้นตั้งแต่การสร้างแบบสอบถาม กำหนดประเภทของข้อถาม จำนวนข้อถาม จำนวนตัวอย่าง แผนสำรวจ เรื่อยไปจนถึงแผนวิเคราะห์ข้อมูล การสร้างแบบสอบถามนั้นมักเป็นปัญหาสำหรับนักวิจัยเสมอ ปัญหาที่กล่าวถึงก็คือ ไม่ทราบว่าจะถามอะไร ถามทำไม ถามแล้วจะเอาไปใช้อะไร เมื่อไร ตอนไหน เท่าที่พบเห็นมา นักวิจัยประเภทนี้มักถามอะไรต่อมิอะไรเปะปะไปหมด บางคำถามก็เกินความจำเป็น ถามแล้วไม่ได้ใช้หรือใช้ไม่ได้ บางเรื่องน่าจะถามก็ไม่ถาม ดังนี้เป็นต้น ต่อข้อสมมุติฐานนี้จะเห็นได้ว่า เป็นข้อสมมุติฐานที่มุ่งเปรียบเทียบทัศนคติ แบบสอบถามจึงต้องถามในประเด็นต่าง ๆ ที่พึงวัดทัศนคติต่อตัวอาจารย์ได้ อาจเป็นการแต่งกายการสอน ท่าทีในการสอนหน้าชั้นความเป็นกันเองกับนักศึกษา ฯลฯ เมื่อเป็นเรื่องเกี่ยวกับทัศนคติ ข้อถามก็น่าจะเป็นข้อถามประเภท

สเกลประเมินค่าหรือกราฟฟิค (Rating Scale หรือ Graphic Scale) จำนวนข้อถามมีกี่ข้อขึ้นอยู่ที่ว่าอะไรบ้างที่ผู้วิจัยจะใช้เป็นตัววัดทัศนคติและเพราะเหตุที่เป็นการศึกษาที่มุ่งเปรียบเทียบแผนสำรวจจึงควรเป็น Stratified Random Sampling หรือ Stratified Systematic Sampling โดยถือว่าชั้นปีเป็นชั้นภูมิ (Stratum) เมื่อวางแผนถึงขั้นนี้ ต่อไปจึงเป็นเรื่องของการกำหนดขนาดตัวอย่าง วิธีกำหนดตัวอย่างก็ต้องให้สอดคล้องกับแผนสำรวจด้วย โดยทั่วไปจะต้องทำ Pretest หรือทดลองแบบสอบถามครั้งหนึ่งก่อน เพื่อตรวจสอบคุณภาพของแบบทดสอบ และนำข้อมูลมาวิเคราะห์เพื่อกำหนดขนาดตัวอย่าง การกำหนดขนาดตัวอย่างอาจกำหนด เป็นร้อยละก็ได้ถ้ามีความชำนาญพอ แผนการต่อไปก็คือแผนวิเคราะห์ข้อมูล ข้อมูลสำหรับสมมุติฐานนี้ควรจัดเป็นตารางจำแนก (Cross Classification Table) โดยจำแนกข้อมูลตามชั้นปีและประเภทหรือระดับทัศนคติตลอดจนประเภทของทัศนคติและอื่น ๆ แผนวิเคราะห์นี้ต้องรวมถึงตัวทดสอบ (Test) หรือตัวสถิติที่จะใช้วัดหรือตัดสินด้วย สำหรับสถานการณ์การเปรียบเทียบนี้ โดยทั่วไปจะใช้ Nonparametric Statistics เช่น Kruskal Wallis Test หรือ Kolmogorov-Smirnov Test หรืออื่น ๆ อาจใช้ t-test หรือ F-test ได้ถ้าแน่ใจหรือสามารถยืนยันได้ว่ากลุ่มประชากรเป็นกลุ่มที่มีการกระจายแบบปกติ

เห็นได้ว่าข้อสมมุติฐานมีบทบาทมากในงานวิจัยทุกประเภท เพราะเมื่อกำหนดสมมุติฐานขึ้น แนวคิดและแผนการของนักวิจัยจะเริ่มขึ้นตั้งแต่ชั้นวางแผนเตรียมการและไหลไปจนถึงแผนวิเคราะห์ แต่ก็เป็นที่น่าประหลาดที่พบว่าหลายคนยังมองไม่เห็นความสำคัญของข้อสมมุติฐาน คิดจะทำงานวิจัยเรื่องอะไร เพียงแต่มีข้อสงสัยก็ลงมือเก็บรวบรวมข้อมูลโดยไม่มีการวางแผนตามที่กล่าวมาแล้ว เมื่อได้ข้อมูลมาแล้วจึงคิดได้ว่ายังไม่ทราบว่าจะวิเคราะห์อย่างไร จัดจำแนกข้อมูลอย่างไร ใช้สถิติตัวไหนเป็นตัวทดสอบ และบ่อยครั้งที่ข้อมูลที่มีอยู่ไม่เพียงพอที่จะใช้ตอบคำถามหรือทดสอบสมมุติฐาน เช่นอยากทราบว่าแนวคิดทางการเมืองของประชาชนในต่างภูมิภาค ระดับอายุ การศึกษา เพศและอาชีพจะแตกต่างกันหรือไม่ ปรากฏว่าข้อถามมุ่งเน้นที่แนวคิด ขาดข้อมูลส่วนตัวที่จะใช้เป็นหลักในการจำแนก เมื่อเป็นเช่นนี้ นักสถิติก็ไม่อาจให้ความช่วยเหลือได้ ที่น่าตลกที่สุดก็คือ มีแต่ข้อมูลคือไปเก็บรวบรวมข้อมูลมาเรียบร้อยแล้ว อยากรู้อะไรก็ถามกันเรื่อยไป ได้ข้อมูลแล้วไม่รู้จะเอาข้อมูลมาประมวลผลอย่างไร ถามว่ามีสมมุติฐานว่าอย่างไรก็ได้รับคำตอบว่ายังไม่ได้ตั้งขึ้นเลยกะว่ามาตั้งทีหลัง อย่างนี้ก็มีและพบบ่อย ๆ ด้วย น่าประหลาดว่าเก็บข้อมูลมาได้อย่างไร การทำงานวิจัยถ้าไม่มีข้อสมมุติฐานก็เปรียบเสมือนเดินเรือโดยไม่มีเข็มทิศและแผนที่ก็ได้ไม่ใช่ไปไม่ได้ แต่จะถึงฝั่งหรือไม่? ระหว่างทางจะแวะพักที่ไหน? เรือจะวิ่งในเส้นทางหรือไม่ อันนี้ตอบไม่ได้

ข้อสมมุติฐานสำหรับงานวิจัยมี 3 ลักษณะคือ สมมุติฐานหลัก (H_0, Null Hypothesis หรือ Zero Hypothesis) กับสมมุติฐานรอง (H_1, Aternative Hypothesis) สมมุติฐานหลักโดยปกติ

เป็นข้อสมมุติฐานที่ตั้งไว้กลาง ๆ ไม่ชี้ทิศทางแต่อย่างไร แต่ก็มีเหมือนกันที่สมมุติฐานหลักตั้งไว้ในลักษณะที่ชี้ทิศทางด้วย แต่ไม่ค่อยนิยม สมมุติฐานหลักทำหน้าที่เสมือน "ผู้ต้องหา" โดยปกติผู้ต้องหามิใช่เป็นผู้ผิดแต่ถูกกล่าวหาหรือสงสัย ผู้ต้องหาจึงอยู่ในฐานะกลางคืออาจผิดก็ได้ ถูกก็ได้ งานหลักทางการวิจัยก็มุ่งยอมรับ (Accept หรือ Not Reject) หรือปฏิเสธ (Reject) สมมุติฐานหลักเป็นสำคัญ ตัวอย่างของสมมุติฐาน เช่น ชายและหญิงมีความไวต่อความรู้สึกในอารมณ์โกรธพอ ๆ กัน หรือ บุหรี่ของโรงงานยาสูบทุกยี่ห้อได้รับความนิยมเท่ากัน หรือ ข้าวโพดสายพันธุ์ใหม่ให้ผลผลิต (ถังต่อไร่) ไม่ต่างจากพันธุ์พื้นเมือง หรือชายและหญิงมีความสามารถในการเรียนภาษาศาสตร์ได้ไม่ต่างกัน หรือ รายได้มิได้มีอิทธิพลต่อการบริโภค หรือรสนิยม รายได้และพื้นฐานการศึกษามิได้มีอิทธิพลต่อความต้องการซื้อ (อุปสงค์) เสื้อแอร์โรว์ เป็นต้น จะเห็นได้ว่าข้อสมมุติฐานหลักเป็นข้อความที่กล่าวได้อย่างกลาง ๆ มิได้ชี้ทิศทางว่าชายหรือหญิงไวต่อความรู้สึกในอารมณ์โกรธมากกว่ากัน บุหรี่ยี่ห้อใดได้รับความนิยมสูงกว่า ฯลฯ แต่ประการใดเหตุที่สมมุติฐานหลักกล่าวไว้กลาง ๆ เช่นนี้ทำให้จำเป็นต้องมีสมมุติฐานรอง (Alternative Hypothesis) ไว้ มิเช่นนั้นจะกลายเป็นว่าทำการวิจัยทั้งที่มิได้ผลลัพธ์อะไรที่เด่นชัดเลย สมมุติฐานรองเป็นสมุติฐานที่มีลักษณะตรงข้ามกับสมมุติฐานหลัก กล่าวคือสมมุติฐานรองจะเป็นตัวระบุหรือชี้ทิศทางที่แน่ชัด เช่นระบุอย่างชัดเจนว่าชายหรือหญิงไวต่ออารมณ์โกรธมากกว่า ข้าวโพดสายพันธุ์ใดให้ผลผลิตสูงกว่า เป็นต้น ขอให้ระลึกไว้ ณ ที่นี้ว่า ไม่ว่าจะเป็นสมมุติฐานหลักหรือสมมุติฐานรองก็ตาม ทั้งคู่ก็ยังคงเป็นเพียงข้อสมมุติมิได้ตัดสินถูกผิด เพียงแต่สมมุติฐานหลักวางตัวเป็นกลางไม่กล่าวโทษโจษความผู้ใด สมมุติฐานรองวางตัวในฐานะผู้กล่าวโทษโจษความเมื่อเป็นเช่นนี้ ความยุ่งยากในการตั้งข้อสมมุติฐานจึงอยู่ที่การตั้งข้อสมมุติฐานรองและในงานวิจัยทั่วไปจะกล่าวถึงเฉพาะสมมุติฐานรอง เรียกว่าสมมุติฐานเพื่อการวิจัย โดยละสมมุติฐานหลักไว้ในฐานะที่เข้าใจ เมื่อมีการทดสอบทางสถิติก็ปฏิเสธหรือยอมรับสมมุติฐานหลักที่ละไว้นั้น ผลลัพธ์จากการปฏิเสธหรือยอมรับสมมุติฐานหลักก็จะปรากฏแก่สมมุติฐานรองได้เองโดยอัตโนมัติ ทั้งนี้เพราะการยอมรับสมมุติฐานหลักก็หมายถึงปฏิเสธสมมุติฐานรอง การปฏิเสธสมมุติฐานหลักก็หมายถึงการยอมรับสมมุติฐานรอง และเหตุที่สมมุติฐานรองเป็นตัวบ่งชี้ที่ชัดเจน การตั้งสมมุติฐานรองจึงต้องรัดกุมพอควรมิใช่ตั้งได้ตามอารมณ์หรือความพอใจ การตั้งสมมุติฐานรองโดยทั่วไปจะต้องอาศัยข้อมูลหรือประสบการณ์ เช่นอาศัยประสบการณ์ที่พบเห็นหรือสถานการณ์เช่นนั้นอยู่บ่อยครั้ง หรือทดลองสุ่มตัวอย่างแล้วอาศัยข้อมูลจากตัวอย่างช่วยบ่งชี้ ตัวอย่างเช่นอยากทราบว่าจะตั้งสมมุติฐานรองว่าอย่างไร สำหรับสมมุติฐานหลักว่า "ชายและหญิงมีความสามารถในการเรียนภาษาศาสตร์ไม่ต่างกัน" ผู้วิจัยอาจอาศัยประสบกาณณ์ที่พบเห็นอยู่บ่อย ๆ คือ ทุกครั้งที่มีการสอบนักศึกษาหญิงมักจะทำคะแนนเฉลี่ยสูงกว่านักศึกษาชายเสมอ หรือลองสุ่ม

ตัวอย่างนักศึกษาทั้งชายและหญิงมาเข้ารับการทดสอบด้วยแบบทดสอบมาตรฐานและปรากฏผลว่านักศึกษาหญิงทำคะแนนเฉลี่ยได้สูงกว่า ดังนี้ควรตั้งสมมุติฐานรองว่า "นักศึกษาหญิงมีความสามารถในทางภาษาศาสตร์สูงกว่านักศึกษาชาย" ดังนี้เป็นต้น หรือแม้แต่จะอาศัยจากประสบการณ์ที่ผู้อื่นทำการศึกษาค้นคว้ามาแล้วก็ได้

5.3 ข้อผิดพลาดในการตัดสินใจ

ในการตัดสินใจยอมรับหรือปฏิเสธสมมุติฐานหลักนั้น สิ่งที่พึงระมัดระวังก็คือข้อผิดพลาดที่อาจเกิดขึ้น เหตุที่สมมุติฐานมี 2 ลักษณะคือสมมุติฐานหลักและสมมุติฐานรอง ข้อผิดพลาดที่พึงบังเกิดจึงมี 2 ประเภท คือข้อผิดพลาดอันเกิดขึ้นจากการปฏิเสธสมมุติฐานหลักทั้ง ๆ ที่สมมุติหลักถูกต้อง (หรือยอมรับสมมุติฐานรองทั้ง ๆ ที่สมมุติฐานรองไม่ถูกต้อง หรือยอมรับสมมุติฐานรองทั้ง ๆ ที่สมมุติฐานหลักถูกต้อง) เรียกว่าความผิดพลาดประเภทที่ 1 (Type I Error) หรือ α-Error เขียนเป็นฟังก์ชั่นความน่าจะเป็นดังนี้

$$\alpha = \Pr(\text{Reject } H_o \mid H_o)^{1}$$

หรือ $$\alpha = \Pr(\text{Accept } H_1 \mid H_o)$$

หรือ $$\alpha = \Pr(\text{Accept } H_1 \mid H_1 \text{ ไม่จริง})$$

และความผิดพลาดอีกประเภทหนึ่งคือ ยอมรับสมมุติฐานหลักทั้ง ๆ ที่สมมุติฐานรองถูกต้อง (หรือยอมรับสมมุติฐานหลักทั้ง ๆ ที่สมมุติฐานหลักไม่ถูกต้อง หรือปฏิเสธสมมุติฐานรองทั้ง ๆ ที่สมมุติฐานรองถูกต้อง) เรียกว่า ความผิดพลาดประเภทที่ 2 (Type II Error) หรือ β-Error สามารถเขียนเป็นรูปฟังก์ชั่นความน่าจะเป็นดังนี้

$$\beta = \Pr(\text{Accept } H_o \mid H_1)$$

หรือ $$\beta = \Pr(\text{Accept } H_o \mid H_o \text{ ไม่จริง})^{2}$$

หรือ $$\beta = \Pr(\text{Reject } H_1 \mid H_1)$$

---

[1] อ่านว่าความน่าจะเป็นที่จะปฏิเสธสมมุติฐานหลักทั้ง ๆ ที่สมมุติฐานหลักถูกต้องมีค่าเท่ากับ α

[2] ดังที่กล่าวไว้ในตอน 5.2 แล้วว่าการตัดสินใจนั้นเรามุ่งตัดสินใจว่าจะยอมรับหรือปฏิเสธสมมุติฐานหลักเป็นสำคัญ ผลที่ติดตามมาก็จะเป็นปฏิเสธหรือยอมรับสมมุติฐานรองได้เองโดยอัตโนมัติ ดังนั้นเราจึงนิยาม α-error และ β-error เป็น $\alpha = \Pr(\text{Reject } H_o \mid H_o)$ และ $\beta = \Pr(\text{Accept } H_o \mid H_1)$ เท่านั้น ไม่นิยมนิยามเป็นรูปอื่น ๆ ดังที่ผู้เขียนเสนอเพิ่มเติม เหตุที่เสนอเพิ่มเติมไว้ในลักษณะอื่นด้วยก็มีเจตนาเพียงขยายความให้นักศึกษาเข้าใจดีขึ้นและมองภาพเดียวกันได้ในหลาย ๆ มุมเท่านั้น

ความผิดพลาดทั้งสองประเภทมีผลเสียต่องาน วิจัยได้ทัดเทียมกัน จะขอยกตัวอย่างเปรียบเทียบให้ดูเพื่อให้สามารถมองเห็นภาพของความผิดพลาดทั้งสองประเภทได้ชัดเจนขึ้นดังนี้ สมมุติว่านายดำถูกกล่าวหาว่าฆ่าคนตายและถูกตำรวจจับไว้เป็นผู้ต้องหา เมื่อเป็นเช่นนี้ นายดำจึงอาจผิดจริงหรือว่าเป็นผู้บริสุทธิ์ก็ได้ การตัดสินของศาลยังไม่เกิดขึ้นเพียงใด การตัดสินว่านายดำเป็นผู้ผิดหรือบริสุทธิ์ก็ยังไม่เกิดขึ้นเพียงนั้น สมมุติว่านายดำฆ่าคนตายจริง แต่ตำรวจหาหลักฐานไม่ได้ศาลยกฟ้อง ก็แปลว่าศาลปฏิเสธว่านายดำผิดทั้ง ๆ ที่นายดำผิดจริง กรณีนี้เรียกว่า α-error โดยนัยกลับกัน สมมุติว่านายดำเป็นประชาชนผู้บริสุทธิ์ แต่เผอิญไปอยู่ในเหตุการณ์ขณะมีการฆ่ากันตาย ผู้ร้ายทำหลักฐานต่าง ๆ ไว้ทำให้ดูเหมือนว่านายดำเป็นผู้ร้ายเสียเอง เมื่อมีหลักฐานมัดตัวแน่นหนาศาลจึงตัดสินจำคุกหรือประหารนายดำ แปลว่าศาลยอมรับว่านายดำผิดจริงทั้ง ๆ ที่ นายดำไม่ผิด กรณีเช่นนี้เรียกว่า β-error อีกตัวอย่างหนึ่ง กองทัพบกทำการรับมอบกระสุนปืนจำนวนหนึ่งจากพ่อค้าผู้ชนะการประกวดราคา การส่งมอบสินค้าจะสมบูรณ์เมื่อมีคณะกรรมการตรวจรับได้ทำการตรวจสอบกระสุนปืนนั้นแล้วว่าตรงตามข้อกำหนด (Specification) เช่น กระสุนไม่ด้าน อายุไม่เกิน 6 เดือนนับแต่วันผลิตความหนาของปลอกไม่เกิน 0.25 ม.ม. และกระสุนต้องผิดข้อกำหนดต่างไม่เกิน 1% เป็นต้น สมมุติว่าตามความเป็นจริงแล้ว กระสุนปืนที่ส่งมอบมีสัดส่วน (Proportion) ที่ผิดข้อกำหนดเพียง 0.5% แต่กรรมการตรวจรับไม่ยอมรับมอบ กรณีเช่นนี้เรียกว่า α-error ขณะเดียวกันสมมุติว่าตามความเป็นจริงแล้วกระสุนดังกล่าวมีสัดส่วนที่ผิดข้อกำหนดถึง 5% แต่กรรมการตรวจรับยอมรับมอบ ดังนี้เรียกว่าเกิด β-error

จากตัวอย่างทั้งสอง นักศึกษาคงจะมองเห็นแล้วว่าทั้ง α-error และ β-error ล้วนมีอิทธิพลต่องานได้ทัดเทียมกันคือก่อให้เกิดผลเสียหายได้ร้ายแรงพอกัน ไม่ว่าจะตัดสินยกฟ้องผู้ตัดสินจำคุกผู้บริสุทธิ์ ไม่รับมอบกระสุนคุณภาพดี ยอมรับมอบกระสุนคุณภาพเลว ล้วนเป็นผลเสียทั้งสิ้น และเมื่อพิจารณากันให้ลึกซึ้ง β-error จะเป็นความเสี่ยงที่มีความสำคัญมากกว่า น่าหวาดกลัวมากกว่า

ในการตัดสินใจทางสถิติ β-error จะมีอิทธิพลมากในการเลือกกติกาการตัดสินใจ เทคนิคการเลือกกติกาที่ใช้กันอยู่คือเลือกกติกาการตัดสินใจที่ให้ β-error ต่ำกว่าหรือต่ำที่สุดในระหว่างกติกาทั้งหลายให้ α-error ระดับเดียวกัน ซึ่งกติกาที่พึงเลือกใช้จะต้องเป็นกติกาที่ให้ β ต่ำที่สุดในบรรดากติกาทั้งหลายเหล่านั้น แต่ β-error แม้จะมีความสำคัญมากกว่า แต่กลับถูกมองข้ามเสมือนปิดทองหลังพระ คนทั่วไปรู้จักแต่ α-error เช่น $\alpha = .05$, $\alpha = .01$ พอเอ่ยถึง β พาลไม่รู้จักและบอกว่าไม่ใช่สิ่งสำคัญ ที่เป็นเช่นนี้เพราะ β-error เป็นสิ่งที่เข้าไปมีอิทธิพลในตอนเริ่มต้นเมื่อเริ่มสร้างตัวทดสอบ (Test) ตัวทดสอบที่ได้ในขั้นสุดท้ายคือขั้นที่จะนำมาใช้ทดสอบสมมุติฐานเป็นตัวทดสอบที่ผ่านการกลั่นกรองคัดเลือกโดยเปรียบเทียบ β-error กันดีแล้วว่าเป็นตัวทดสอบ

ที่ให้ β-error ต่ำสุด เมื่อถึงขั้นนำมาใช้ก็เป็นอันรู้กันว่า ตัวทดสอบนั้นดีที่สุดแล้ว หน้าที่ของผู้ใช้จึงเหลือเพียงการควบคุมมิให้ α-error มีค่าสูงเกินไป โดยทั่วไปเรานิยมควบคุมให้มีได้เพียง 5% หรือ 1% จะต่ำไปกว่านั้นก็ได้ถ้าเป็นงานที่อาจเกิดอันตรายได้ เช่น การวิจัยทางยาซึ่งถ้าให้มี α-error สูงถึง 1% ดูจะสูงเกินไป หมายความว่า ยาชนิดนั้นคน 100 คน รับประทานยาแล้วจะต้องแพ้ยาหรือตายเพราะยา 1 คน เช่นนี้ก็จำเป็นต้องคุม α-error ให้ต่ำมาก ๆ อาจเป็น $\alpha = 0.00001\%$ หรือต่ำกว่า จะเห็นได้ว่า α-error ถูกนำมาใช้หรือกล่าวถึงอย่างออกหน้าออกตามากกว่า ทั้ง ๆ ที่ β-error มีบทบาทมากกว่าและน่าจะหวาดกลัวกว่า อย่างไรก็ตามขอเตือนไว้ ณ ที่นี้ด้วยว่า ถ้าไม่จำเป็นจริง ๆ แล้วอย่าคุม α-error ให้ต่ำมากนักเพราะมันจะมีผลให้ β-error มีค่าสูงขึ้นทั้งนี้เพราะ α และ β สัมพันธ์กันอยู่ในรูปสัดส่วนผกผัน (Inverse Relationship) หลักทั่วไปสำหรับการเลือกระดับของ α และ β คือการวางแผนควบคุมให้ $\alpha = \beta$ ในขั้นวางแผนกำหนดขนาดตัวอย่าง แต่ก็มิใช่หลักสากลที่จะต้องให้ $\alpha = \beta$ เสมอไป อาจอยู่ในระดับใดก็ได้ตามที่เห็นสมควรที่พอจะยอมรับได้

ข้อคิดสำหรับเรื่องนี้ก็คือ ไม่ว่าจะวางแผนการวิจัยอย่างรัดกุมเพียงใด ควบคุมงานดีเพียงใด ความผิดพลาดจะต้องเกิดขึ้นได้เสมอ จะต่างกันก็แต่เพียงเกิดขึ้นในดีกรีที่สูงต่ำเพียงใดเท่านั้น โดยทั่วไปเราจะควบคุมให้เกิดขึ้นในดีกรีต่ำหรือค่อนข้างต่ำ สำหรับ α-error นิยมใช้กันเป็นสากลให้อยู่ในระดับ 5% และ 1% แต่จะสูงต่ำกว่านั้นก็ได้ทั้งนี้ขึ้นอยู่กับประเภทของงานว่าเสี่ยงภัยเพียงใดสำหรับผู้ที่จะใช้ผลงานวิจัยส่วน β-error เรานิยมคำนวณค่าเป็นตัวเลขในภายหลัง โดยทั่วไปก็ควรควบคุมให้อยู่ในระดับต่ำ ๆ เช่นกัน ซึ่งเราสามารถทำได้ด้วยการกำหนดขนาดตัวอย่างที่เหมาะสม หรือกำหนดค่า α และ β ลงไปได้เลย เช่นในงานตรวจสอบคุณภาพ (Sampling Inspection หรือ Acceptance Sampling) ที่อาศัย Sequential Analysis เรื่องนี้นักศึกษาจะได้ศึกษาโดยละเอียดในโอกาสต่อไป

ในทางทฤษฎีเราจะเสนอ α-error และ β-error ในรูปของฟังก์ชั่นความน่าจะเป็นที่สอดคล้องกับเขตวิกฤติ (Critical Region หรือ Rejection Region) ดังนี้

ให้ R คือเขตวิกฤติ และ $X_1, X_2, \ldots, X_n$ เป็น Sampled Random Variables ที่สุ่มมาจากกลุ่มประชากรที่มีพารามิเตอร์ θ โดยมี $f_X(x; \theta)$ และ มีสมมุติฐานดังนี้คือ $H_0 : \theta = \theta_0$ vs $H_1: \theta \neq \theta_0$ ดังนั้น

$$\begin{aligned} \alpha &= \Pr(\text{Reject } H_0 | H_0) \\ &= \sum_{i=1}^{n} \Pr(x_i \in R | \theta = \theta_0) \end{aligned}$$

และ

$$\beta = \Pr(\text{Accept } H_0|H_1)^{1}$$

$$= \sum_{i=1}^{n} \Pr(x_i \in \bar{R}|\theta \neq \theta_0) \text{ ; } \bar{R} \text{ คือ Acceptance Region}$$

ในขณะเดียวกัน

$$1 - \alpha = 1 - \Pr(\text{Reject } H_0|H_0)$$

$$= \Pr(\text{Accept } H_0|H_0)$$

$$= \sum_{i=1}^{n} \Pr(x_i \in \bar{R}|\theta = \theta_0)$$

ค่า $1 - \alpha$ เรียกว่า Sensitivity
ขณะเดียวกัน

$$1 - \beta = 1 - \Pr(\text{Accept } H_0|H_1)$$

$$1 - \beta = \Pr(\text{Reject } H_0|H_1)$$

$$= \Pr(\text{Reject } H_0|H_0 \text{ ไม่จริง})$$

$$= \sum_{i=1}^{n} \Pr(x_i \in R|\theta \neq \theta_0)$$

ค่า $1 - \beta$ เรียกว่าอำนาจของการตรวจสอบ (Power of Test) โดยปกติในการทดสอบข้อสมมุติฐานเดียวกันจะมีตัวทดสอบ (กติกาการตัดสินใจ) ได้หลายตัว และตัวทดสอบทั้งหลายเหล่านั้นจะมีอำนาจในการตรวจสอบไม่เท่ากัน ตัวทดสอบใดมีอำนาจตรวจสอบสูงกว่าจะเป็นตัวที่เหมาะสมกว่า และดังที่เคยกล่าวแล้วว่าโดยปกติแล้ว เราจะใช้ β-error หรือ $1 - \beta$ อย่างใดอย่างหนึ่งเป็นหลักในการเปรียบเทียบเลือกตัวทดสอบ (กติกาการตัดสินใจ) จึงเห็นได้ว่า ถ้า β มีค่าต่ำ $1 - \beta$ จะมีค่าสูง β มีค่าสูง $1 - \beta$ จะมีค่าต่ำ เมื่อเป็นเช่นนี้ โดยทั่วไปเราจึงเลือกตัวทดสอบ

---

[1] $\beta = \Pr(\text{Accept } H_0|H_1)$ ถ้า $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1 > \theta_0$

$\Rightarrow \beta(\theta_1) = \Pr(\text{Accept } H_0|\theta = \theta_1)$

เมื่อแปรค่า θ ไปในทุกค่าที่เป็นไปได้ของ $\theta = \theta_1 > \theta_0$ ค่าความน่าจะเป็นที่ได้คือ Probability of Acceptance ณ ค่า $\theta = \theta_1$ ถ้านำคู่ลำดับ $(\theta_1, \beta(\theta_1))$ ในทุกค่าของ θ ไปพล็อตจะได้โค้งที่เรียกว่า Probability of Acceptance Curve หรือ Operating Characteristic Curve เรียกย่อ ๆ ว่า OC-Curve โค้ง OC ใช้สำหรับเปรียบเทียบแผนการ หรือ test เพื่อดูว่า แผนการ (Sampling Plan) หรือ test ใดดีกว่ากัน test ที่ให้โค้ง OC ต่ำกว่าย่อมดีกว่า

อนึ่งโค้ง OC จะกลับกันกับโค้ง Power เพราะ $\pi = 1 - \beta$ โดยนัยนี้ test ใดให้ Power Curve สูงกว่าย่อมเป็น test ที่ดีกว่า

ที่ให้อำนาจทดสอบสูงกว่าในบรรดาตัวทดสอบทั้งหลายที่ให้ α-error ในระดับเดียวกัน ดังนี้ โดยหลักทางทฤษฎีแล้ว เราจึงนิยมใช้อำนาจการทดสอบ $(1 - \beta)$ หรือ β อย่างใดอย่างหนึ่งเป็นเครื่องมือคัดเลือกกติกาการตัดสินใจ

5.4 ขั้นตอนในการดำเนินการทดสอบสมมุติฐาน

ในการดำเนินการทดสอบสมมุติฐาน นักวิจัยพึงยึดถือปฏิบัติตามขั้นตอนทั่วไปดังนี้

ก. ศึกษาและกำหนดรูปการแจกแจงของตัวแปรสุ่ม

ข. ตั้งข้อสมมุติฐาน

ค. คำนวณและสร้างฟังก์ชั่นความน่าจะเป็นของตัวสถิติ (Sampling Distribution)

ง. เลือกระดับนัยสำคัญและเขตวิกฤติที่สอดคล้องกับระดับนัยสำคัญ

จ. คำนวณค่าของตัวทดสอบ (Test Statistics) ที่ได้จากกลุ่มตัวอย่างตามลักษณะที่ได้รับในข้อ ค. และสอดคล้องกับข้อ ง.

ฉ. ตัดสินใจ

การตัดสินใจในการดำเนินงานในแต่ละขั้นตอนจะได้กล่าวถึงโดยละเอียดดังต่อไปนี้

ก. การศึกษาและกำหนดรูปการแจกแจงของตัวแปรสุ่ม

งานในขั้นต้นที่สุดที่นักวิจัยจะต้องทำในการทดสอบสมมุติฐานก็คือ ตรวจสอบดูว่ากลุ่มประชากรที่มุ่งศึกษามีลักษณะใด แบบปกติ ทวินาม แกมม่า เบต้า หรือ เอกโพเนนเชียล ฯลฯ กล่าวเช่นนี้หลายคนยังอาจมองไม่เห็น โดยเฉพาะผู้ที่ศึกษาวิชาสถิติแต่เพียงเล็กน้อยพอเพียงแก่การใช้งานบางระดับ โดยปกติสมมุติฐานที่มุ่งทดสอบจะเสนอออกมาในรูปของข้อความซึ่งเมื่อถึงขั้นทดสอบจริงจะต้องแปลงรูปออกมาเป็นสัญลักษณ์ในรูปของพารามิเตอร์ของกลุ่มประชากรอีกทีหนึ่ง เช่น มีสมมุติฐานเพื่อการวิจัย (Research Hypothesis) ว่า "ผลผลิตมันสำปะหลังของประเทศไทยในปี 2523 สูงกว่าปีกลายในช่วงเดียวกัน" เมื่อจะทำการทดสอบจำเป็นต้องแปลงเป็นสมมุติฐานทางสถิติ (Statistical Hypothesis) เสียก่อน คือ $H_0 : \mu = \mu_0$ vs $H_1 : \mu = \mu_1 > \mu_0$ โดยที่ μ_0 คือผลผลิตโดยถัวเฉลี่ย (ต่อหน่วย) ในปี พ.ศ. 2522 ในช่วงเดือนเดียวกันกับปี 2523 μ คือ ผลผลิตโดยถัวเฉลี่ยสำหรับปีพ.ศ. 2523 μ_1 คือค่าเฉพาะค่าหนึ่งของผลผลิตสำหรับปีพ.ศ. 2523 ดังนี้เป็นต้น จะเห็นได้ว่าเมื่อแปลงรูปสมมุติฐานเพื่อการวิจัยมาสู่รูปสมมุติฐานทางสถิติแล้ว รูปร่างหน้าตาของพารามิเตอร์ก็ปรากฏให้เห็น ซึ่งจะได้ช่วยชี้ให้เห็นแนวทางสำหรับการเลือกตัวทดสอบหรือกระบวนการทดสอบที่เหมาะสมต่อไป ปัญหาเฉพาะหน้าที่นักวิจัยจะพบต่อไปก็คือจะใช้ตัวทดสอบใดจึงจะเหมาะหรือมองลึกลงไปถึงกระบวนการของการสร้างตัวทดสอบก็คือการแจกแจงของตัวสถิติ (Sampling Distribution) มีฟังก์ชั่นความน่าจะเป็นแบบใด เรื่องการทดสอบสมมุติฐานนี้ผู้เขียนขอย้ำอีกครั้งหนึ่งว่า หัวใจของงานอยู่ที่การสร้างหรือพัฒนาฟังก์ชั่นความน่า-

จะเป็นของตัวสถิติถ้าไม่มีความรู้ความเข้าใจในเรื่องนี้มาก่อนก็ยากที่จะตอบคำถามว่าเพราะเหตุใด จึงใช้ตัวสถิตินั้น ๆ มาทดสอบได้และมีผลต่อความถูกต้องแม่นยำของงานวิจัยอย่างมากอีกด้วย ปัญหาขั้นต้นที่จะต้องแก้ให้ได้ก็คือตัวแปรสุ่มที่มีพารามิเตอร์ตามสมมุติฐานหลักมีการแจกแจงแบบใด? เพราะถ้าแก้ปัญหานี้ได้ ปัญหาเรื่องการแจกแจงของตัวสถิติก็ไม่ยากเกินไปทั้งนี้เพราะการแจกแจงของตัวสถิติพัฒนาขึ้นมาจากฟังก์ชั่นความน่าจะเป็นของตัวแปรสุ่มหรือประชากร เช่น $X \sim N(\mu, \sigma^2)$ $Y = \sum_{i=1}^{n} X_i$ จะมีการแจกแจงแบบ $N(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2)$ หรือ $X \sim EX(\lambda)$ $Y = \sum_{i=1}^{n} X_i$ จะมีการแจกแจงแบบ $G(n, \lambda)$ หรือ $Y = 2\lambda \sum_{i}^{n} x_i$ มีการแจกแจงแบบ $x^2_{(2n)}$ เป็นต้น

การจะทราบได้ว่าประชากรที่มุ่งศึกษามีการแจกแจงแบบใดนั้น เรามีวิธีอยู่มากมายหลายวิธี วิธีแรกที่ต้องรู้อยู่แล้วเป็นทุนเดิมก็คือ "สถานการณ์ของตัวแปรสุ่มแต่ละชนิด" เช่นสถานการณ์ใดเป็นสถานการณ์ของตัวแปรสุ่มปกติ สถานการณ์ใดเป็นสถานการณ์ของตัวแปรสุ่มพัวซอง สถานการณ์ใดเป็นสถานการณ์ของตัวแปรสุ่มแกมม่า ฯลฯ นอกจากวิธีนี้แล้วอีกวิธีหนึ่งที่นิยมใช้คือการใช้ Probability Paper Probability Paper เป็นแผ่นกระดาษคล้ายกระดาษกราฟ มีสเกลเฉพาะตัวสำหรับตัวแปรสุ่มแต่ละชนิด และมีรูปร่างของโค้งที่ Plot ได้จากกลุ่มข้อมูลเฉพาะตัว ส่วนใหญ่จะเป็นเส้นตรง เช่น ถ้านำข้อมูลที่มีอยู่ไป Plot ใน Normal Probability Paper แล้ว โค้งที่ได้เป็นรูปเส้นตรงก็ถือได้ว่า กลุ่มประชากรของกลุ่มตัวอย่างที่สุ่มมานั้นมีการแจกแจงแบบปกติ ถ้าเป็นเส้นโค้งลักษณะอื่นก็แสดงว่ากลุ่มประชากรนั้นมิได้มีการแจกแจงแบบปกติ ให้ทดลองใช้ Probability Paper แบบอื่นต่อไป เช่น Exponential Probability Paper หรือ Log Normal Probability Paper หรืออื่น ๆ นอกจากวิธีนี้แล้ว อีกวิธีหนึ่งที่นิยมใช้ก็คือ การสร้าง Probability Distribution ของประชากรดูโดยอาศัย Monte Carlo Techniques การใช้วิธี Simulation และการทดสอบด้วย Goodness of fit test วิธีอื่น ๆ นอกจากนี้ก็ยังมีอีก เช่น การวิเคราะห์ฮิสโตแกรม วิธีที่ไม่ดีที่สุดแต่นิยมใช้ที่สุดก็คือวิธีกำหนดลงไปในข้อตกลงเบื้องต้น เช่น "ถือว่ากลุ่มประชากรนี้มีการแจกแจงแบบปกติ" ซึ่งนอกจากจะเป็นการบังคับข้อมูลมากเกินไป วิธีนี้เหมาะสมสำหรับผู้ที่ชำนาญงานหรือมีประสบการณ์มาก ๆ แต่ค่อนข้างเสี่ยง แม้จะมีทฤษฎีการโน้มเข้าสู่เกณฑ์กลาง (CLT) ค้ำประกันไว้ก็ตาม โดยทั่วไปและที่พบเห็นบ่อยครั้งจะเป็นการตกลงให้กลุ่มประชากรมีการแจกแจงแบบปกติมากกว่าการแจกแจงแบบอื่น ๆ ทั้งนี้เพราะ ถ้าตกลงเช่นนี้ก็หมายถึงความสะดวกสบายในการใช้ตัวทดสอบซึ่งก็คงไม่พ้น t-test, F-test หรืออื่น ๆ ที่อิงการแจกแจงแบบปกติ การตกลงเช่นนี้ดูจะเหมาะสมสำหรับงานที่ใช้กลุ่มตัวอย่างขนาดใหญ่เพราะมีทฤษฎีการโน้มเข้าสู่เกณฑ์กลางคอยค้ำจุนก็จริงอยู่ แต่ควรระลึกไว้ด้วยว่า CLT มิใช่จะใช้ได้ในทุกสถานการณ์[1]

[1] อ่านตอน 4.2.4

บางสถานการณ์ก็ใช้ไม่ได้ หรือใช้ได้แต่มีความผิดพลาดสูงและที่สำคัญก็คือ CLT ให้ค่าได้เพียงค่าประมาณมิใช่ค่าที่ถูกต้องจริง (Exact Value) ดังนั้น ถ้าจะมิให้งานวิจัยยืนอยู่บนความเสี่ยงเพราะการประมาณและทำให้เกิดความเสี่ยงเพิ่มขึ้นจากการมี α-error และ β-error ซึ่งต้องปรากฏอยู่ตามปกติแล้ว นักวิจัยพึงหาทางตรวจสอบการกระจายของประชากรไว้ด้วย

ข. ตั้งข้อสมมุติฐาน

สมมุติฐานที่จะทดสอบจะต้องเสนอไว้ในรูปของพารามิเตอร์ที่ต้องการ พารามิเตอร์ของใคร? พารามิเตอร์ที่ต้องการก็คือพารามิเตอร์ของกลุ่มประชากรที่กำหนดหรือตรวจสอบไว้ได้แล้วในตอน ก. กลุ่มประชากรอาจมีพารามิเตอร์เดียว เช่นการแจกแจงแบบทวินาม การแจกแจงแบบอนุกรมเรขาคณิต การแจกแจงแบบพัวซอง การแจกแจงแบบเอกโพเนนเชียล หรือกลุ่มประชากรที่มีพารามิเตอร์ตั้งแต่ 2 ตัวขึ้นไปเช่น การแจกแจงแบบปกติ การแจกแจงแบบแกมม่า การแจกแจงแบบเบต้ารวมตลอดถึงการแจกแจงแบบพหุนามและอื่น ๆ กรณีของกลุ่มประชากรที่มีพารามิเตอร์เดียวไม่ค่อยมีปัญหาของข้อจำกัดและข้อกำหนดมากนักแต่ถ้าเป็นกรณีของกลุ่มประชากรที่มีพารามิเตอร์ตั้งแต่ 2 ตัวขึ้นไปจำเป็นต้องศึกษาหรือมีข้อกำหนดเกี่ยวกับพารามิเตอร์ตัวอื่น ๆ ด้วย เช่น กรณีการแจกแจงแบบปกติ ถ้าจะทำการทดสอบค่าเฉลี่ย μ ก็จำเป็นต้องศึกษากำหนดหรือควบคุม σ^2 หรือกรณีการแจกแจงแบบแกมม่า ถ้าจะทดสอบพารามิเตอร์ r จำเป็นต้องกำหนดหรือควบคุมพารามิเตอร์ λ ดังนี้เป็นต้น

สมมุติฐานมี 2 ชนิดคือ Simple Hypothesis หรือ Exact Hypothesis กับ Composite Hypothesis หรือ Inexact Hypothesis, Simple Hypothesis เป็นสมมุติฐานที่ระบุค่าพารามิเตอร์ไว้อย่างชัดเจน เช่น $H_0: \mu = 10$ หรือ $H_0: \lambda = 2$ หรือ $H_0: p = .20$ หรือ $H_0: r = 2.77$ ส่วน Composite Hypothesis หรือ Inexact Hypothesis คือสมมุติฐานที่ระบุค่าของพารามิเตอร์ไว้กว้าง ๆ คลุมไว้หลวม ๆ ไม่แจ้งชัด ส่วนใหญ่จะระบุไว้เป็นเซท เช่น $H_0: \mu > 10$ หรือ $H_0: \mu \geqslant 10$ หรือ $H_0: \mu < 10$ หรือ $H_0: \leqslant 10$ หรือ $H_0: \mu \neq 10$ หรืออื่น ๆ ในทำนองนี้ จะเห็นว่าเสนอไว้ในรูปของเซท การทดสอบสมมุติฐานนั้นสามารถเสนอสมมุติฐานเป็น Simple Hypothesis หรือ Composite Hypothesis ก็ได้ขึ้นอยู่กับข้อมูลข้อสนเทศที่มีอยู่ว่าสนับสนุนให้จัดไว้ในรูปใด โดยทั่วไปจะเสนอสมมุติฐานหลักเป็น Simple Hypothesis ส่วนสมมุติฐานรองจะเสนอเป็น Simple Hypothesis หรือ Composite Hypothesis ก็ได้แล้วแต่ข้อมูลข้อสนเทศดังกล่าวแล้ว ถ้ามีข้อมูลยืนยันได้ชัดเจน ก็สามารถกำหนดสมมุติฐานรองเป็น Simple Hypothesis ได้เช่น $H_0: \mu = 10$ vs $H_1: \mu = 12$ ถ้ามีข้อมูลยืนยันพอควรไม่มากมายชัดเจนนักก็เสนอเป็น Composite Hypothesis เช่น $H_0: \mu = 10$ vs $H_1: \mu < 10$ หรือ $H_0: \mu = 10$ vs $H_1: \mu > 10$ กรณี $H_0: \mu = 10$ vs $H_1: \mu < 10$ เป็นกรณีที่มีข้อมูลยืนยันว่าค่าเฉลี่ย

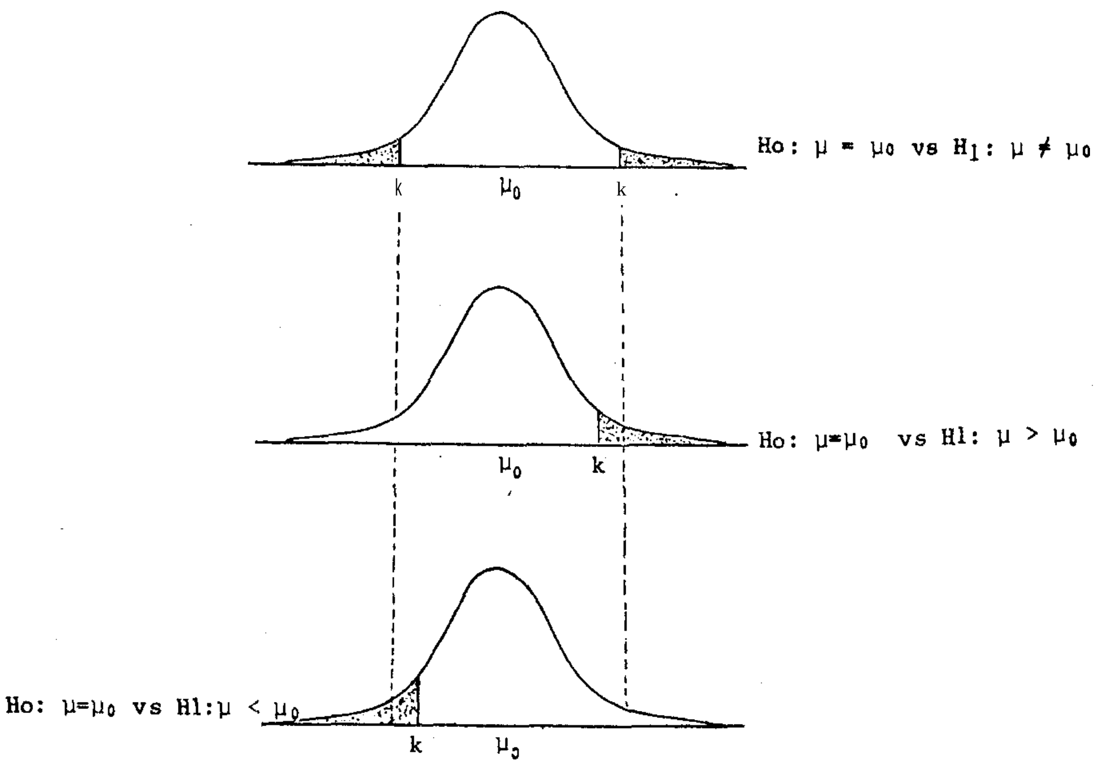
จะมีค่าต่ำกว่า 10 หรืออยู่ด้านซ้ายมือของเลข 10 ในสเกลของจำนวน (Real Line) การทดสอบสมมุติฐานดังกล่าวลักษณะนี้เรียกว่า การทดสอบสมมุติฐานหางซ้าย (Left Tail Hypothesis Testing) เพราะเขตวิกฤติจะปรากฏที่ปลายซ้ายของโค้งตัวสถิติ ถ้าสมมุติฐานเป็น $H_0 = \mu = 10$ vs $H_1 : \mu > 10$ แสดงว่าข้อมูลข้อสนเทศที่มีอยู่ (อาจเป็นผลที่ได้มาจากการสังเกตหรือประสบการณ์) ยืนยันว่าค่าเฉลี่ยจะมีค่าสูงกว่า 10 การทดสอบสมมุติฐานลักษณะนี้เรียกว่า การทดสอบสมมุติฐานหางขวา ทั้งนี้เพราะเขตวิกฤติปรากฏที่ปลายด้านขวาของโค้งของตัวสถิติ ขอให้สังเกตว่าการทดสอบทั้งสองชนิดนี้มิได้ระบุค่าที่ชัดเจนของ μ ว่าเป็นเท่าไรแน่ บอกไว้เพียงกว้าง ๆ ว่ามากกว่าหรือน้อยกว่าเท่านั้น กรณีสุดท้ายคือกรณี $H_0 : \mu = 10$ vs $H_1 : \mu \neq 10$ กรณีนี้เป็นกรณีที่ผู้ทดสอบไม่มีข้อสนเทศไว้ในมือเลย ทำให้ต้องตั้งสมมุติฐานไว้กว้างมาก ๆ ไม่ระบุทิศทางใดที่แน่ชัด เขตวิกฤติที่ใช้จะปรากฏที่ปลายโค้งทั้งสองของตัวสถิติ การทดสอบสมมุติฐานลักษณะนี้เรียกว่า การทดสอบสมมุติฐานสองหาง (Two Tails Hypothesis Testing)

โดยหลักการของการทดสอบสมมุติฐานนั้นเรามุ่งปฏิเสธสมมุติฐานหลักเป็นสำคัญ แต่จะปฏิเสธได้หรือไม่ขึ้นอยู่กับข้อมูลที่ได้รับว่าสนับสนุนให้ปฏิเสธได้หรือไม่ ถ้าปฏิเสธไม่ได้เราก็ไม่ปฏิเสธ คำนี้ฟังดูเหมือนกำปั้นทุบดิน แต่โดยหลักการเขาเรียกกันอย่างนั้นคือไม่ปฏิเสธสมมุติฐานหลัก (Not Reject H_0) ไม่นิยมคำยอมรับสมมุติฐานหลัก แต่โดยหลักทางปฏิบัติ การไม่ปฏิเสธสมมุติฐานหลักกับการยอมรับสมมุติฐานหลักก็คือสิ่งเดียวกัน เหตุที่ต้องเรียกว่าไม่ปฏิเสธสมมุติฐานหลักหรือปฏิเสธสมมุติฐานหลักไม่ได้ก็เพราะเหตุว่างานทดสอบสมมุติฐานเป็นงานที่มุ่งปฏิเสธสมมุติฐานหลักเป็นสำคัญดังกล่าวแล้ว[1] เมื่อนักศึกษาทราบเจตนาของงานทดสอบสมมุติฐานตามนัยนี้แล้วลองย้อนไปพิจารณาสมมุติฐานอีกที่ว่า การทดสอบหางเดียวหรือสองหางอย่างไหนดีกว่ากันที่ต้องพูดเรื่องนี้ไว้ก็เพราะผู้เขียนเคยได้รับคำถามลักษณะนี้บ่อย ๆ ถ้ามองเผิน ๆ จะพบว่าการทดสอบสองหางดีกว่าเพราะมีเขตวิกฤติถึง 2 เขตและไม่ยุ่งยากในการกำหนดสมมุติฐาน นี่ก็พอนับว่าเป็นข้อดีของการทดสอบสองหางได้ แต่ไม่ใช่ข้อดีตามหลักที่ถือว่าทางสถิติ เพราะตัวทดสอบใดจะถือว่าดีโดยนัยสถิติได้จะต้องเป็นตัวทดสอบที่ให้ β-error ต่ำที่สุดหรือมีอำนาจทดสอบ (Power of Test) สูงที่สุด และจากการเปรียบเทียบอำนาจตรวจสอบซึ่งนักศึกษาจะได้พบในบทต่อไปจะพบว่า การทดสอบหางเดียวมีอำนาจตรวจสอบสูงกว่าการตรวจสอบสองหาง

---

[1] คำว่า "ไม่อาจปฏิเสธได้" กับคำว่า "ยอมรับ" มีน้ำหนักต่างกัน โดยปกติเราไม่นิยมที่จะให้ใช้คำว่า "ยอมรับ" เพราะในงานทดสอบสมมุติฐานนั้นมี error ปะปนอยู่หลายแหล่งที่สำคัญคือ α-error β-error และ Sampling Variation ในส่วนที่ Sampling Variation นั้นหมายถึงผลอันเนื่องมาจากการสุ่มตัวอย่าง โดยปกติกลุ่มตัวอย่างจะเป็นไปได้ถึง $\binom{N}{n}$ ชุดแตกต่างกัน ตัวอย่างหนึ่งให้ค่าสถิติตกอยู่นอกเขตวิกฤติแต่ตัวอย่างอื่นอาจให้ค่าสถิติตกอยู่ในเขตวิกฤติก็ได้ การผลีผลามใช้คำว่า "ยอมรับ" จึงดูเป็นการเชื่อง่ายหรือหูเบามากสักหน่อย

และเมื่อพิจารณาพื้นที่ของเขตวิกฤติของการทดสอบทั้งสอง (หางเดียวกับ 2 หาง) เมื่อใช้ระดับนัยสำคัญเดียวกัน การทดสอบหางเดียวมีพื้นที่เขตวิกฤติในด้านเดียวกันสูงกว่า ทำให้มีโอกาสปฏิเสธสมมุติฐานหลักได้มากกว่า การทดสอบหางเดียวจึงมีความไวต่อการปฏิเสธสมมุติฐานหลักสูงกว่า ดังภาพ



ค. คำนวณและสร้างฟังก์ชั่นความน่าจะเป็นของตัวสถิติ

งานในขั้นนี้ถ้าจะแยกพิจารณาก็สามารถแยกเป็นได้ 2 ตอนคือ การคำนวณหาตัวสถิติที่เกี่ยวข้องกับการทดสอบตอนหนึ่งและการหาฟังก์ชั่นความน่าจะเป็นของตัวสถิติที่ได้อีกตอนหนึ่ง แต่ถ้านักศึกษามีความรู้พื้นฐานเกี่ยวกับการกระจายของตัวแปรสุ่มและตัวสถิติดีพอ งานขั้นนี้จะต่อเนื่องเป็นตอนเดียวกัน

งานทดสอบสมมุติฐานนั้นเปรียบเสมือนกระบวนการยุติธรรมหรือวิธีการทางศาล สมมุติฐานหลักทำหน้าที่เป็นผู้ต้องหา (จำเลย) สมมุติฐานรองทำหน้าที่เป็นโจทก์ ผู้ทดสอบทำหน้าที่เป็นศาลซึ่งจะต้องรวบรวมพยานหลักฐานและโดยประมวลกฎหมายอาญาผู้ต้องหาจะได้รับการพิจารณาในฐานะผู้บริสุทธิ์ไว้ก่อนเสมอ ถ้าหลักฐานอ่อนแม้ผู้ต้องหาจะรับสารภาพศาลก็ยกฟ้อง

แสดงว่าศาลยอมให้ยกฟ้องผู้ผิดได้มากกว่าเอาผิดผู้บริสุทธิ์หรือนัยหนึ่งศาลยอมเสี่ยงต่อการเกิด α-error มากกว่า β-error ในงานขั้นรวบรวมหลักฐานนั้นอุปมาเสมือนการสุ่มตัวอย่างเพราะในการทดสอบสมมุติฐานนั้น งานสำคัญที่สุดที่จะขาดเสียมิได้ก็คือการสุ่มตัวอย่างเพื่อเก็บรวบรวมข้อมูลมายืนยันสนับสนุนหรือคัดค้านสมมุติฐานหลัก และเนื่องจากกลุ่มตัวอย่าง (Random Sample) ก็คือกลุ่มของตัวแปรสุ่มที่สุ่มมาจากประชากรที่กำลังศึกษา (Sampled Random Variables) ดังนั้นกลุ่มของตัวแปรสุ่มเหล่านี้จึงมีฟังก์ชั่นความน่าจะเป็นร่วมกันอยู่รูปหนึ่งเรียก joint probability Density Function หรือ Likelihood Function ปัญหาก็คือฟังก์ชั่นนี้สนับสนุนสมมุติฐานหลัก (จริงตามสมมุติฐานหลัก) หรือสนับสนุนสมมุติฐานรอง (จริงตามสมมุติฐานรอง) วิธีตัดสินว่าฟังก์ชั่นดังกล่าวสนับสนุนสมมุติฐานใด เราอาศัยทฤษฎีที่ใช้เป็นหลักสำคัญอยู่ 2 ทฤษฎีคือ ทฤษฎีของเนย์แมน-เบียร์สัน (Nayman-Pearson Lemma) และ Maximum Likelihood Ratio อย่างใดอย่างหนึ่ง ถ้าเป็นการทดสอบ Simple Hypothesis หรือ Composite Hypothesis ที่พอจัดเป็น Simple Hypothesis ได้ ก็ใช้ทฤษฎีของเนย์แมน-เบียร์สัน ถ้าเป็นการทดสอบ Composite Hypothesis หรือกลุ่มประชากรที่กำลังศึกษามีพารามิเตอร์เกินกว่า 1 ตัว จำเป็นต้องควบคุมพารามิเตอร์ตัวอื่นนอกเหนือไปจากตัวที่ระบุไว้ในสมมุติฐาน ก็ใช้ Maximum Likelohood Ratio Test นักศึกษาจะได้ศึกษาเรื่องเหล่านี้โดยการดูรายละเอียดในบทต่อไป ในที่นี้นำมากล่าวไว้ให้เห็นเป็นนิทรรศนะเท่านั้น และโดยอาศัยทฤษฎีของเนย์แมน-เบียร์สัน หรือ Maximum Likelihood Ratio Test สิ่งที่จะปรากฏออกมาก็คือ เขตวิกฤติในรูปของฟังก์ชั่นของตัวสถิติ เช่นปฏิเสธสมมุติฐานหลักเมื่อ $\sum_{i=1}^{n} X_i > k$, $|\sum_{i=1}^{n} X_i| > k$, $\sum_{i=1}^{n} X_i < k$, $\prod_{i=1}^{n} X_i > k$ หรือ $\text{Sup}\, X < k$, $\text{Inf}\, X > k$ เป็นต้น เมื่อถึงขั้นนี้ก็แปลว่านักวิจัยทำงานมาเกือบครึ่งทางแล้ว เรียกว่าพอเห็นค้าโครงหรือแนวทางการตัดสินใจได้ลาง ๆ แล้ว ถ้าเปรียบเทียบกับกระบวนการยุติธรรมก็เหมือนได้พยานหลักฐานไว้เกือบครบขาดแต่พยานปากเอก หน้าที่ของนักวิจัยหรือนักสถิติในลำดับต่อไปก็คือการตรวจสอบหรือคำนวณดูว่าตัวสถิติ (ฟังก์ชั่นของตัวแปรสุ่ม) ที่ได้คือ $\sum_{i=1}^{n} X_i$, $\prod_{i=1}^{n} X_i$ $\text{Sup}\, X$, $\text{Inf}\, X$ หรืออื่น ๆ มีฟังก์ชั่นความน่าจะเป็นรูปใด เท่าที่เคยพบมานักศึกษามักมีปัญหาตรงนี้ เรียกว่ามีพื้นฐานความรู้เรื่องการกระจายของตัวสถิติ (Sampling Distribution) ไม่ดีพอ ถ้าผ่านด่านนี้ไปได้ก็เหมือนได้พยานปากเอก ตัดสินคดีได้ ถ้าผ่านไม่ได้ก็ตัดสินไม่ได้หรือตัดสินได้ก็พลาดพลั้งได้มากกว่าปกติ ที่กล่าวเช่นนี้ก็เพราะไม่ต้องการยืนยันอย่างแข็งขันนักว่าตัดสินไม่ได้ ความจริงนั้นเราพอมีลู่ทางอยู่แต่เป็นเรื่องของการประมาณการ (Approximation) ซึ่งไม่แม่นยำคมชัดเท่าของจริง กล่าวคือ เมื่อสร้างตัวสถิติได้แล้วและพบปัญหาว่าไม่อาจทราบลักษณะฟังก์ชั่นการกระจายของตัวสถิติเหล่านั้น นักศึกษาสามารถใช้ทฤษฎีการโน้มเข้าสู่เกณฑ์กลาง (Central Limit Theorem, CLT)

เข้าช่วยเหลือ ซึ่งมีผลให้ได้กติกาการตัดสินใจตามต้องการและดังที่ได้กล่าวแล้วว่า กติกานี้ใช้ได้พอเป็นแต่เพียงค่าประมาณยังรวมความผิดพลาดไว้ด้วย ดังนั้นถ้าไม่จำเป็นจริง ๆ แล้วผู้เขียนก็ไม่อยากแนะนำให้ใช้เพราะนอกจากจะให้ผลที่ไม่ถูกต้องสมบูรณ์แล้วยังชี้ให้เห็นถึงความอ่อนภูมิรู้เชิงวิชาการสถิติของผู้ใช้อีกด้วย

เทคนิคสำคัญที่ใช้คำนวณหาฟังก์ชั่นความน่าจะเป็นของตัวสถิติที่นิยมใช้กันอย่างกว้างขวางในปัจจุบันมีอยู่ 3 ประการคือ

1. Moment Generating Function Technique (mgf) วิธีนี้เหมาะสำหรับคำนวณหาฟังก์ชั่นของตัวแปรสุ่มที่อยู่ในรูปของการประกอบกันเชิงเส้น (Linear Combination of Random Variables) คือ $Y = a_1X_1 + a_2X_2 + \dots + a_nX_n$ เช่น $Y = X_1 + X_2 + \dots + X_n$ หรือ $Y = 2\lambda X_1 + 2\lambda X_2 + \dots + 2\lambda X_n$ หรือ $Y = (\frac{1}{2}\lambda)X_1 + (\frac{1}{2}\lambda)X_2 + \dots + (\frac{1}{2}\lambda)X_n$ หรือ $Y = \frac{1}{n}(X_1 + X_2 + \dots + X_n) = \overline{X}$ หรือ $Y = \overline{X}_1 - \overline{X}_2$ ฯลฯ

2. Comulative Density Function Technique (cdf) วิธีนี้เหมาะสำหรับคำนวณหาฟังก์ชั่นของตัวแปรสุ่มที่อยู่ในรูปยกกำลัง คือ $Y = aX^r$; r เป็นเลขจำนวนจริง เช่น $Y = X^{1/2}$, $Y = X^2$, $Y = X^3$ เป็นต้น นอกจากนี้เทคนิคของวิธี cdf ยังใช้ผสมกับวิธี mgf ได้เป็นอย่างดีโดยเฉพาะในกรณีของการประกอบกันที่มิใช่เชิงเส้นคือ $Y = a_1X_1^{r_1} + a_2X_2^{r_2} + \dots + a_nX_n^{r_n}$ เมื่อ a_i และ r_i เป็นเลขจำนวนจริง เช่น $Y = X_1^2 + X_2^2 + \dots + X_n^2$ มีการกระจายแบบ $\chi^2_{(n)}$ เมื่อ $X_i \sim N(0, 1)$ หรือ $Y = \sum_{i}^{n}(\frac{X_i - \mu_i}{\sigma_i})^2$ มีการกระจายแบบ $\chi^2_{(n)}$ เมื่อ $X_i \sim N(\mu_i, \sigma_i^2)$ เป็นต้น

3. Transformation of Random Variables Technique วิธีนี้สามารถใช้คำนวณหาการแจกแจงของฟังก์ชั่นของตัวแปรสุ่มได้ทุกลักษณะ โดยเฉพาะอย่างยิ่งในกรณีที่ไม่อาจคำนวณหาได้ด้วยเทคนิคทั้งสองประการที่ผ่านมา โดยทั่วไปแล้วฟังก์ชั่นของตัวแปรสุ่มที่จำเป็นต้องใช้วิธีนี้คือฟังก์ชั่นที่เสนออยู่ในรูปของผลคูณและผลหารของตัวแปรสุ่ม ส่วนรูปอื่น ๆ ให้พิจารณาเลือกใช้เอาตามความเหมาะสม

เทคนิคสองวิธีแรกคือ mgf และ cdf ได้เสนอไว้แล้วในตอนต้นโดยเฉพาะอย่างยิ่ง mgf ได้กล่าวไว้ค่อนข้างละเอียดเพราะเห็นว่าใช้ง่าย สะดวกรวดเร็ว และโดยปกติตัวสถิติตลอดจนฟังก์ชั่นของตัวแปรสุ่มเท่าที่พบเห็นอยู่เสมอ ๆ นั้นมักปรากฏอยู่ในรูปที่เอื้ออำนวยให้สามารถนำ mgf เข้าช่วยวิเคราะห์ได้ ส่วนกรณีของการแปลงรูปตัวแปรสุ่ม (Transformation of Random Variables Technique) ยังมิได้กล่าวถึงโดยละเอียดนักแต่ก็เคยกล่าวถึงไว้บ้างแล้วในตอนที่ว่าด้วยการกระจายของตัวแปรสุ่มที่เนื่องถึงการกระจายแบบปกติ (t, F, χ, χ^2) นักศึกษาควรศึกษาหาความรู้เพิ่มเติมจากตำราเล่มอื่นที่อธิบายเรื่องนี้ไว้ด้วย

ง. เลือกระดับนัยสำคัญและเขตวิกฤติที่สอดคล้องกับระดับนัยสำคัญ

การเลือกระดับนัยสำคัญและเขตวิกฤติเป็นกระบวนการที่ต่อเนื่องกัน เพราะเมื่อกำหนดระดับนัยสำคัญลงไปแล้วผลที่ติดตามมาก็คือเขตวิกฤติที่สอดคล้องกับระดับนัยสำคัญนั้น ระดับนัยสำคัญ (Significant Level) ก็คือขนาดของ α-error หรือความเสี่ยงที่พึงบังเกิดขึ้นจากการปฏิเสธสมมุติฐานหลักที่เป็นจริง เขตวิกฤติ (Critical Region-Rejection Region) คือกลุ่มหรือเซทของค่าของตัวสถิติที่พึงเป็นไปได้ทั้งหมดที่ทำให้ปฏิเสธสมมุติฐานหลัก โดยหลักทางทฤษฎีนั้นเมื่อกำหนดระดับนัยสำคัญให้แล้วหน้าที่ของนักวิจัยขั้นต่อไปก็คือการปรับค่าคงที่ที่ช่วยชี้ขีดจำกัดล่างหรือขีดจำกัดบนของเขตวิกฤติที่ทำให้พื้นที่ใต้โค้งของตัวสถิติในเขตวิกฤติมีค่าเท่ากับ (หรือใกล้เคียง) ระดับนัยสำคัญ ตัวอย่างเช่น ต้องการทดสอบสมมุติฐานเกี่ยวกับค่าเฉลี่ย μ ของกลุ่มประชากรปกติเมื่อไม่ทราบค่าของ σ^2 และมีสมมุติฐานดังนี้ $H_0: \mu = \mu_0$ vs $H_1: \mu > \mu_0$ ปรากฏว่าเขตวิกฤติคือ $\frac{\overline{X} - \mu_0}{S/\sqrt{n}} > k$ ในกรณีนี้ k คือค่าคงที่ที่แสดงขีดจำกัดล่างของเขตวิกฤติ เมื่อกำหนดระดับนัยสำคัญให้มีค่าเท่ากับ 5% หน้าที่ของนักวิจัยก็คือปรับค่าคงที่ K ไปเรื่อย ๆ จนกระทั่งได้ค่า k ที่ทำให้พื้นที่ของเขตวิกฤติใต้โค้งของตัวสถิติ $\frac{\overline{X} - \mu_0}{S/\sqrt{n}}$ มีค่าเท่ากับ 5% พอดี ดังนี้เป็นต้น

งานของนักวิจัยในด้านการกำหนดระดับนัยสำคัญมิใช่งานหนัก เพราะนักวิจัยสามารถเลือกใช้ระดับนัยสำคัญใดก็ได้ที่พิจารณาเห็นว่าสมควร โดยทั่วไปนิยมใช้ระดับ 5%, 1% และ .001% แต่จะใช้ระดับอื่นนอกเหนือไปจากนี้ก็ได้ดังที่ได้กล่าวไว้แล้วในตอน 5.3 งานที่หนักก็คือการปรับค่า k เพราะต้องอาศัย ความรู้เกี่ยวกับ การแจกแจงของตัวสถิติ (Sampling Distribution) ค่า k คือค่า ควอนไทล์[1] หรือค่าตัวเลขบนแกนนอนใต้โค้งของตัวสถิติ จะเป็นควอนไทล์ลำดับ (order) ใดขึ้นอยู่กับระดับนัยสำคัญ ถ้ากำหนดระดับนัยสำคัญไว้เท่ากับ 5% ค่า k ก็คือควอนไทล์ลำดับที่ 0.05 หรือถ้ากำหนดระดับนัยสำคัญเป็น 1% ค่า k ก็คือควอนไทล์ลำดับที่ 0.01 ลองย้อนพิจารณาตามตัวอย่างที่ผ่านมาพบว่าเขตวิกฤติคือ $\frac{\overline{X} - \mu_0}{S/\sqrt{n}} > k$ และกำหนดให้ $\alpha = .05$ กรณีนี้ k ที่คำนวณได้จะเป็นควอนไทล์อันดับที่ .05 และพบว่า $k = t_{n-1,.95}$ = ควอนไทล์อันดับ .05 ดังนั้นเขตวิกฤติที่ต้องการคือ $\frac{\overline{X} - \mu_0}{S/\sqrt{n}} > t_{n-1,.95}$ หรือ $\overline{X} > \mu_0 + st_{n-1,.95}/\sqrt{n}$ ขอให้สังเกตว่าเขตวิกฤติจะอยู่ในรูปของเซทเสมอในที่นี้เขตวิกฤติคือเซทของ $\overline{X}$ ที่มีค่าเกินกว่า $\mu_0 + S.t_{n-1,.95}/\sqrt{n}$ หมายความว่าเมื่อใดก็ตามที่ค่าเฉลี่ย $\overline{X}$ ที่ได้จากกลุ่มตัวอย่างขนาด n มีค่าเท่ากับค่าใดค่าหนึ่งในเซทนี้ <u>เราจะปฏิเสธสมมุติฐาน</u>หลักทันที

[1] ควอนไทล์ (Quantile) นิยามไว้ดังนี้ ถ้า X เป็นตัวแปรสุ่มที่มี pdf เป็น $f_X(x)$ และ α คือ พื้นที่ใต้โค้งของ $f_X(x)$ ที่สอดคล้องกับสมการ $\alpha = \int_{-\infty}^{q} f_X(x)dx$ แล้วค่า q ที่ได้เรียกว่า ควอนไทล์อันดับ α

คำว่า "มีนัยสำคัญ" หรือ "แตกต่างกันอย่างมีนัยสำคัญ" (Significant หรือ Significance Difference) หมายความว่า ความแตกต่างระหว่างค่าตามสมมุติฐาน (Hypothetical Value) กับค่าที่ปรากฏจากกลุ่มตัวอย่าง (Sample Value) แตกต่างกันมากจริง ๆ มิใช่ต่างกันโดยเหตุบังเอิญ (Chance) บางครั้งเรียกความแตกต่างนี้ว่า "นัยสำคัญทางสถิติ" (Statistical Significant) ตัวอย่างเช่น กองตำรวจจราจรทำสถิติจำนวนอุบัติเหตุบริเวณสี่แยกไฟแดงไว้พบว่า โดยถัวเฉลี่ยแล้วจะมีอุบัติเหตุตรงบริเวณสี่แยกไฟแดงวันละ 10 ราย ภายหลังเมื่อกรมตำรวจปรับปรุงระบบจราจรบริเวณสี่แยกทุกแห่ง เปลี่ยนระบบสัญญาณไฟจราจรเป็นสัญญาณอัตโนมัติและบังคับช่องทางวิ่ง พบว่ามีอุบัติเหตุบริเวณสี่แยกถัวเฉลี่ยวันละ 8 ราย จากอุทาหรณ์นี้เห็นได้ว่า $\mu = 10$ และ $\bar{X} = 8$, 10 และ 8 ต่างกันอย่างมีนัยสำคัญหรือไม่ ถ้าต่างกันอย่างมีนัยสำคัญ ก็แปลว่าระบบสัญญาณไฟอัตโนมัติ และการบังคับช่องทางวิ่งมีผลต่อการลดจำนวนอุบัติเหตุ ถ้าต่างกันแต่ไม่มีนัยสำคัญก็แสดงว่าระบบสัญญาณไฟอัตโนมัติและการบังคับช่องทางวิ่งไม่มีผลในการลดจำนวนอุบัติเหตุ[1]

ปัญหาก็เกิดขึ้นตรงตีความหมายของคำว่า "มีนัยสำคัญ" นี่เองเพราะไม่ว่าจะมองดูอย่างไร 10 กับ 8 ก็ต่างกันเสมอ ไม่ใช่เลขจำนวนเดียวกันแน่ ๆ หลายคนสงสัยว่าถ้าผลปรากฏว่าไม่ต่างกัน (ไม่มีนัยสำคัญทางสถิติ) มันจะเป็นไปได้อย่างไรในเมื่อผลที่ปรากฏให้เห็นนี้ 10 กับ 8 มันต่างกันชัด ๆ ? ถ้าสงสัยอย่างนี้ก็แปลว่ายังไม่เข้าใจคำว่า "มีนัยสำคัญ" และ "ไม่มีนัยสำคัญ" ความจริง 10 กับ 8 แตกต่างกันแน่ ๆ แต่แตกต่างกันด้วยเหตุบังเอิญหรือต่างกันในลักษณะที่ยืนยันได้ ถ้าผลสรุปว่าไม่มีนัยสำคัญก็แปลว่าต่างกันด้วยเหตุบังเอิญ อย่างไรก็ตาม ผลสรุปจะเป็นรูปใดขึ้นอยู่กับระดับนัยสำคัญ α ที่กำหนดขึ้น จะต่างกันแน่หรือต่างกันโดยบังเอิญก็ต้องพูดกันโดยพื้นฐานของระดับนัยสำคัญ สมมุติว่ากำหนดระดับนัยสำคัญเป็น $\alpha = .05$ และสมมุติว่าผลสรุปที่ได้คือ "มีนัยสำคัญทางสถิติ" ก็แปลว่า 10 กับ 8 แตกต่างกันแน่ คือ ถ้าทำการทดลองหรือวิจัยเรื่องนี้ 100 ครั้ง ผลสรุปจะยืนยันว่าค่าจากตัวอย่างกับค่าตามสมมุติฐานหลักจะต่างกัน 95 ครั้ง ไม่ต่างกัน 5 ครั้ง แต่ถ้าปรากฏผลสรุปว่า "ไม่มีนัยสำคัญทางสถิติ" ก็แปลว่า ถ้าทำการทดลองหรือวิจัยลักษณะนี้ 100 ครั้ง ค่าจากตัวอย่างกับค่าตามสมมุติฐานหลักจะต่างกันไม่ถึง 95 ครั้ง ไม่ต่างกันเกินกว่า 5 ครั้ง ซึ่งแสดงว่ามีเหตุบังเอิญที่มาทำให้ค่าทั้งสองเกิดความแตกต่างกันอยู่

[1] โดยปกตินัยสำคัญจะผูกพันอยู่กับระดับความเสี่ยง α เสมอ การที่เราใช้ระดับความเสี่ยง $\alpha = 1\%$ แล้วผลการทดสอบชี้ว่าไม่มีนัยสำคัญ มิได้หมายความว่าไม่มีความแตกต่างกันแต่หมายถึงเฉพาะไม่มีความแตกต่าง ณ. ระดับ $\alpha = 1\%$ เท่านั้น อาจแตกต่างกันในระดับอื่น ๆ เช่น $\alpha = 5\%, 6\%, 10\%, 20\%$, (หรือไม่มีนัยสำคัญ) ณ. ระดับนัยสำคัญ 5% ผลสรุปนี้จะเป็นเพียงการยืนยันที่ระดับความเสี่ยง $\alpha = 5\%$ นักวิจัยไม่อาจยืนยันได้ในลักษณะทั่วไป เพราะถ้าใช้ = 61, 7% หรือ 10%, 20%- อาจมีนัยสำคัญก็ได้ นอกจากนี้การทดลองยังเกี่ยวข้องกับขนาดตัวอย่าง n อีกด้วย การใช้ขนาดตัวอย่างไม่เหมาะสม ย่อมส่งผลให้ขาดข้อมูลข้อสนเทศอย่างเพียงพอในการยังผลสรุป

นักศึกษาจึงมองเห็นได้ว่า ตัวการที่ทำให้สามารถยังผลสรุปว่ามีนัยสำคัญหรือไม่นั้นคือระดับนัยสำคัญ α ถ้าค่า α มีมากขึ้นเพียงใด ความมีนัยสำคัญก็ปรากฏง่ายขึ้นเพียงนั้นแต่ความเสี่ยงต่อการตัดสินใจผิดก็เพิ่มขึ้นเป็นเงาตามตัว ดังนั้นในทางปฏิบัติ (แม้จะไม่เสมอไป) นักวิจัยควรคาดหมายหรือหวังผลลัพธ์ลักษณะใดลักษณะหนึ่งไว้ตั้งแต่เริ่มงาน คือจะปฏิเสธหรือยอมรับสมมุติฐานหลัก หลายคนอาจแคลงใจว่าทำเช่นนั้นได้หรือ? คำถามนี้จะเป็นคำถามสำหรับนักวิจัยหน้าใหม่หรือผู้ที่ยังไม่เคยทำการวิจัยเท่านั้น โดยปกติการวิจัยเรื่องใดนักวิจัยจะต้องมีความรู้ความเข้าใจในเรื่องที่จะทำนั้นมาเป็นอย่างดีหรือมีประสบการณ์ในเรื่องนั้นมาดีแล้ว แม้แต่ผู้ที่ไม่จำเจหรือคุ้นเคยกับงานนั้นก็จำเป็นต้องศึกษาในลักษณะวรรณคดีที่เกี่ยวข้อง (Related Literature) เพื่อสอบย้ำทิศทางของผลการวิจัย การมีประสบการณ์ การรู้ทิศทางของงานย่อมหมายถึงรู้ทิศทางว่างานทดสอบสมมุติฐานนั้นจะมีนัยสำคัญหรือไม่มีนัยสำคัญนั่นเอง ในทางปฏิบัติถ้านักวิจัยต้องการปฏิเสธสมมุติฐานก็ควรกำหนดค่า α ให้มีขนาดใหญ่ ถ้าไม่ปรารถนาจะปฏิเสธสมมุติฐานก็กำหนดค่า α ให้มีขนาดเล็ก แต่ก็ต้องระวังค่า β ไว้ด้วยเพราะค่าของ β มีจำนวนเป็นสัดส่วนผกผันกับ α และไม่ควรกำหนด α ให้มีขนาดต่างไปจากระดับที่นิยมใช้กันเป็นสากลคือ .05 และ .01 มากนัก นอกจากปัจจัยที่ทำให้เกิด "เหตุบังเอิญ" อีกประการหนึ่งก็คือความผันแปรจากกลุ่มตัวอย่าง (Sampling Variation) ทั้งนี้เพราะในการดำเนินการวิจัยนั้นจำเป็นต้องอาศัยข้อมูลจากกลุ่มตัวอย่าง ค่าของตัวสถิติจากกลุ่มตัวอย่างที่พึงเป็นไปได้ทั้งหมด (All Possible Samples) ย่อมผันแปรไปได้ แม้ว่าตัวประมาณค่าจะเป็นตัวแทนที่ไม่มีอคติ (Unbiased Estimator) ของพารามิเตอร์ก็ตามและในทางปฏิบัติเราก็สุ่มตัวอย่างเพียงตัวอย่างเดียว (Single Sample) มิใช่ใช้จากทุกตัวอย่าง ดังนั้นความผันแปรหรือผันผวนไปได้ในค่าของตัวประมาณค่าจึงย่อมมีส่วนที่ก่อให้เกิด "เหตุบังเอิญ" ขึ้นได้ดังกล่าวแล้ว

ในประการสุดท้ายก่อนจะผ่านตอนนี้ไป ขอให้ข้อสังเกตไว้ว่า ค่า α ในงานทดสอบสมมุติฐานกับค่า α ในงานประมาณค่าด้วยช่วงเชื่อมั่น (Interval Estimation) นั้นมีความหมายเชิงสถิติ แตกต่างกัน α ในงานทดสอบสมมุติฐานคือ ค่าขนาดของความน่าจะเป็นที่จะปฏิเสธสมมุติฐานหลักทั้ง ๆ ที่สมมุติฐานหลักเป็นจริง หรือพื้นที่ทั้งหมดใต้โค้งของตัวสถิติซึ่งใช้เป็นเขตวิกฤติ แต่ α ในงานประมาณค่าด้วยช่วงเชื่อมั่นที่สร้างขึ้นนั้นจะไม่นับรวมค่าพารามิเตอร์จริง (True Parameter)

จ. คำนวณค่าของตัวทดสอบ (Computing the Test Statistics)

ตัวทดสอบ (Test Statistics) คือตัวสถิติที่ใช้เป็นเครื่องมือในการตัดสินใจว่าจะปฏิเสธหรือไม่ปฏิเสธสมมุติฐานหลัก ตัวทดสอบนี้จะมีลักษณะใดขึ้นอยู่กับปัจจัยต่าง ๆ ดังที่เสนอมาแล้วในตอน ก. ถึง ง. อย่างไรก็ตามขอสรุปไว้ในที่นี้อีกครั้ง

ปัจจัยที่กำหนดลักษณะของตัวทดสอบคือ

1. การกระจายของกลุ่มประชากร
2. สมมุติฐานที่มุ่งทดสอบ
3. การกระจายของตัวสถิติ
4. ระดับนัยสำคัญ

การทดสอบสมมุติฐานของพารามิเตอร์ของประชากรแต่ละชนิดย่อมให้ตัวทดสอบคนละลักษณะแม้ในกลุ่มประชากรเดียวกัน (ในกรณีที่มีพารามิเตอร์มากกว่า 1 ตัว) ก็ให้ตัวทดสอบแตกต่างกันสำหรับพารามิเตอร์ต่างกัน ที่เป็นเช่นนี้เพราะแต่ละสมมุติฐานของต่างกลุ่มประชากรจะให้การแจกแจงของตัวสถิติคนละแบบ นอกจากนี้ระดับนัยสำคัญย่อมให้ผลต่อรูปร่างหน้าตาของตัวทดสอบด้วย รายละเอียดเหล่านี้นักศึกษาจะได้พบในบทต่อไป

การคำนวณค่าของตัวทดสอบ คำนวณตามลักษณะของตัวทดสอบโดยอาศัยข้อมูลจากกลุ่มตัวอย่างเป็นหลัก ขั้นตอนนี้มิได้ยุ่งยากแต่ประการใดเพียงแต่คำนวณตามแบบและอาศัยความรู้เรื่องการใช้ตารางสถิติของประชากรแต่ละแบบเท่านั้น เช่น ในการทดสอบค่าความแปรปรวน σ^2 ของกลุ่มประชากรปกติ โดยมีสมมุติฐานดังนี้คือ $H_0: \sigma^2 = 2$ vs $H_1: \sigma^2 > 2$ ปรากฏว่าตัวทดสอบคือปฏิเสธสมมุติฐานหลักเมื่อ $(n-1)s^2 > \sigma_0^2\chi^2_{(n-1),1-\alpha}$ หรือ $(n-1)s^2 > 2\chi^2_{(n-1),1-\alpha}$ หน้าที่ของนักวิจัยที่จะต้องทำก็เพียงแต่สุ่มตัวอย่างมา n ชุด แล้วอาศัยข้อมูลจากตัวอย่างคำนวณหาค่า $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ และเปิดตาราง χ^2 ณ. df = n − 1 พื้นที่ $1-\alpha$ การตัดสินใจก็สามารถกระทำได้ตามเกณฑ์ของเขตวิกฤติ มิได้ยุ่งยากแต่ประการใด ดังนี้เป็นต้น

5.5 นิยามและสัญลักษณ์

เพื่อประโยชน์ต่อการศึกษาและความเข้าใจงานทดสอบสมมุติฐานไปในแนวทางร่วมกัน จำเป็นต้องเข้าใจในนิยามดังต่อไปนี้ [1]

1. สมมุติฐานทางสถิติ (Statistical Hypothesis)

สมมุติฐานทางสถิติคือ ข้อกำหนดหรือบ่งชี้เกี่ยวกับการแจกแจงของตัวแปรสุ่ม ถ้าสมมติฐานใดบ่งชี้ลักษณะการแจกแจงของตัวแปรสุ่มไว้อย่างชัดเรียกสมมุติฐานนั้นว่า Simple Statistical Hypothesis มิเช่นนั้นจะเรียกว่า Composite Statistical Hypothesis

---

[1] เพื่อความเข้าใจขอให้ย้อนกลับไปอ่านตอน 5.1 ถึง 5.4

2. ตัวทดสอบ (Test)

ตัวทดสอบก็คือกฎเกณฑ์หรือกติกาที่นำไปสู่การยอมรับหรือปฏิเสธสมมุติฐานทางสถิติ ตัวทดสอบโดยปกติจะอยู่ในรูปของตัวสถิติ[1]

3. เขตวิกฤติ (Crital Region)

เขตวิกฤติคือ ส่วนหนึ่งหรืออนุเซทของ Sample Space ที่สอดคล้องกับตัวทดสอบ อันมีผลให้นำไปสู่การตัดสินใจยอมรับหรือปฏิเสธสมมุติฐาน

4, Power Function *ของตัวทดสอบ*

Power Function ของตัวทดสอบคือฟังก์ชั่นที่ใช้สำหรับคำนวณหาค่าความน่าจะเป็นที่จะปฏิเสธสมมุติฐานหลัก ค่าของ Power Function ณ ค่าพารามิเตอร์ที่กำหนดให้เรียกว่า Power of Test ณ ค่าพารามิเตอร์นั้น

หรือ $\Pi(\theta) = \Pr(\text{Reject } H_1|H_1)$[2]

เช่น $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1 > \theta_0$

$=> \ \Pi(\theta_1) = \Pr(\text{Reject } H_0|\theta = \theta_1)$

และเมื่อนำค่า $\Pi(\theta)$ ไปพล๊อตจะได้โค้งของ Power of Test

5. ระดับนัยสำคัญ (Significant Level) *หรือขนาดของพื้นที่ในเขตวิกฤติ*[3]

ระดับนัยสำคัญคือค่าสูงสุด (ที่พอจะยอมรับได้) ของความน่าจะเป็นที่จะต้องปฏิเสธสมมุติฐานหลักทั้ง ๆ ที่สมมุติฐานหลักถูกต้อง

หรือนัยหนึ่ง $\alpha = \Pr(\text{Reject } H_0|H_0)$

6. เขตวิกฤติที่ดีที่สุด (Best Critical Region, BCR) *ของ* Test *ขนาดเดียวกัน*

เขตวิกฤติที่ดีที่สุดคือเขตวิกฤติที่ให้ Power สูงที่สุดในบรรดา test ขนาดเดียวกัน[4]

---

[1] ตัวสถิติ (Statistics) คือฟังก์ชั่นของตัวแปรสุ่มและต้องเป็นฟังก์ชั่นที่ไม่มีพารามิเตอร์ปะปนอยู่หรือแม้จะปะปนอยู่ก็จะต้องเป็นพารามิเตอร์ที่ทราบค่าหรือกำหนดค่าเอาไว้

[2] ใช้สัญลักษณ์ Π แทน Power Function บางครั้งใช้อักษร K เช่น $K(\theta_1)$, $\Pr(\text{Reject } H_0|H_1)$ อ่านว่าความน่าจะเป็นที่จะปฏิเสธสมมุติฐานหลักเมื่อสมมุติฐานรองเป็นจริง (หรือสมมุติฐานหลักไม่จริง) ดังนี้ ถ้า test ใหทำให้มีเปอร์เซนต์ที่ปฏิเสธสมมุติฐานที่ไม่ถูกต้องได้สูงเพียงใดแสดงว่า test นั้นมีคุณภาพหรืออำนาจ (Power) ตรวจจับความผิดได้สูงเพียงนั้น

[3] หรือเรียกว่า Size of Test

[4] หรือให้ β ต่ำที่สุดในบรรดา test ขนาดเดียวกัน เมื่อ

$$\beta(\theta) = \Pr(\text{Accept } H_0|H_1) = 1 - \Pi(\theta)$$

ถ้าให้ R′ คือเขตวิกฤติขนาด α ใด ๆ และ R คือขนาดของเขตวิกฤติที่ดีที่สุดขนาด α ดังนั้น R จะเป็นเขตวิกฤติที่ดีที่สุดเมื่อ

(1) $Pr\{X_1, X_2, \ldots X.) \in R|H_0\} = a = Pr\{(X_1, X_2, \ldots X_n) \in R'|H_0\}$

(2) $Pr\{(X_1, X_2, \ldots X,) \in R/H,\} \geqslant Pr\{(H_1, X_2, \ldots X_n) \in R'|H_1\}$[4]

7. Uniform Most Powerful Critical Region (UMPC)

คือเขตวิกฤติที่ยังคงให้ Power สูงกว่าทุกเขตในขนาดเดียวกันในทุกค่าของพารามิเตอร์ตามสมมุติฐานรอง (หรือ H_1 จริง)

Test ที่สอดคล้องกับ UMPC เรียกว่า Uniform Most Powerful Test (UMPT) เป็น Test ที่ให้ Power สูงกว่าทุก Test ขนาดเดียวกันในทุกค่าของพารามิเตอร์ θ_1 ตาม H_1 หรือนัยหนึ่งไม่ว่าจะแปรค่าพารามิเตอร์ตาม H_1 ไปอย่างไร UMPT ก็จะให้ Power สูงกว่า Test อื่น ๆ ตลอดเวลา

5.6 การทดสอบสมมุติฐาน

โดยปกติการทดสอบสมมุติฐานจะมีทั้งสมมุติฐานหลัก (H_0) และสมมุติฐานรอง (H_1) ซึ่งมีสิทธิถูกต้องได้ทั้งคู่ ปัญหาก็คือจะทราบได้อย่างไรว่า H_0 ถูกต้องหรือ H_1 ถูกต้อง

อย่างไรก็ตามการที่จะลงข้อยุติว่าสมมุติฐานใดควรจะได้รับการยอมรับนั้นนักวิจัยจำเป็นจะต้องรวบรวมข้อมูลข้อสนเทศต่าง ๆ เพื่อสนับสนุนหรือร่วมพิจารณา โดยนัยทางสถิติ วิธีการรวบรวมข้อมูลข้อสนเทศดังกล่าวก็คือการสุ่มตัวอย่าง แล้วนำข้อมูลจากทุกหน่วยในตัวอย่างมาร่วมสร้างกฎเกณฑ์การตัดสินใจ

ให้ $(X_1, X_2, \ldots, X_n)$ เป็นกลุ่มตัวอย่างที่สุ่ม (Sampled Random Variable) มาขนาด $n; n \geqslant 2$ ที่สุ่มมาจากประชากรที่มีพารามิเตอร์ θ มี pdf เป็น $f_X(x; \theta)$ โดยมีสมมุติฐานที่ต้องตัดสินใจดังนี้คือ

$$H_0 : \theta = \theta' \text{ vs } H_1 : \theta = \theta''$$

ดังนั้นย่อมหมายความว่ากลุ่มประชากรที่สนใจอาจมี pdf เป็น $f_X(x; \theta'')$ หรือ $f_X(x; \theta'')$ ก็ได้

---

[4] หรือ $Pr\{(X_1, X_2, \ldots X_n) \in \bar{R}|H_1\} \leqslant Pr\{(X_1, X_2, \ldots X_n) \in \bar{R}'|H_1\}$

สำหรับ test ที่สอดคล้องกับ BCR เรียกว่า Best Test, Best Test จึงเป็น Test หนึ่งในบรรดา Test ขนาดเดียวกัน (α) สำหรับทดสอบสมมุติฐานเดียวกันที่ให้ β ต่ำที่สุด หรือ power สูงที่สุด

และด้วยเหตุที่เราจำเป็นต้องอาศัยข้อมูลข้อสนเทศทุกหน่วยมาร่วมพิจารณาโดยนัยนี้ย่อมหมายความว่าเมื่อถึงวาระจะต้องตัดสินใจ เราจำเป็นต้องตัดสินใจเลือกระหว่าง Likelihood Function (หรือ joint pdf) ที่เป็นจริงตาม H_0 หรือ H_1 อย่างใดอย่างหนึ่งเท่านั้น

เมื่อ $L = \prod_{i=1}^{n} f_X(x_i;\ 6)$ คือ Likelihood Function'

และ $L_0 = \prod_{i=1}^{n} f_X(x_i;\ \theta')$ คือ Likelihood Function ที่สอดคล้องกับ $H_0 : \theta = \theta'$

และ $L_1 = \sum_{i=1}^{n} f_X(x_i;\ \theta'')$ คือ Likelihood Function ที่สอดคล้องกับ H, : $\theta = \theta''$

ปัญหาในปัจจุบันก็คือในระหว่าง L_0 และ L_1 นั้นเราควรเชื่อถือใคร? จะมีวิธีการใดคัดเลือกหรือช่วยชี้ให้เห็นว่า L_0 ถูกต้อง (ซึ่งสะท้อนให้เห็นว่า $\theta = \theta'$) หรือว่า L_1 ถูกต้อง (ซึ่งสะท้อนให้เห็นว่า $\theta = \theta''$)

สำหรับปัญหานี้เรามีทฤษฎีที่สามารถให้คำตอบได้โดยจำแนกเป็น 2 กรณีคือ กรณีของ Simple Hypothesis และ Composite Hypothesis คือ

ก. ในกรณี Simple Hypothesis ทฤษฎีที่ช่วยชี้ตัดสินใจว่าควรเลือกใช้ L_0 หรือ L_1 คือ Neyman-Pearson Lemma (NPL) กรณีนี้ทั้ง H_0 และ H_1 จะต้องเป็น Simple Hypothesis ทั้งคู่

ข. ในกรณี Composite Hypothesis ทฤษฎีที่ช่วยตัดสินใจว่าควรเลือกใช้ L_0 หรือ L_1 คือ Maximum Likelihood Ratio Test (MLRT) กรณีนี้ H_0 และ H_1 อาจเป็น Composite Hypothesis ด้วยกันทั้งคู่ หรือเฉพาะ H_1 เท่านั้น ที่เป็น Composite Hypothesis

อย่างไรก็ตาม ในกรณีของ Composite Hypothesis เรายังคงใช้ NPL เป็นเครื่องช่วยในการตัดสินใจได้โดยการจัดค่าของพารามิเตอร์ตาม Composite Hypothesis ออกเป็นกลุ่มของ Simple Hypothesis ในรูปของ Set of Points เช่น $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$ ซึ่ง H_1 เป็น Composite Hypothesis ซึ่งสามารถจัดเป็น Simple Hypothesis ได้โดยใช้ $H_1 : \theta = \theta_1$ โดยถือว่า $\theta_1 > \theta_0$ นั่นคือเราจัด H_1 เป็น $H_1 : \theta = \theta_1 > \theta_0$ ดังนี้เป็นต้น

---

[1] โดยทั่วไปจะเขียน Likelihood Function $\prod_{i=1}^{n} f_{X_i}(x_i;\ \theta)$ เป็น $L(x_1, x_2, \ldots, x_n)$ หรือ $L(x_1, x_2, \ldots, x_n;\ \theta)$ ในที่นี้ใช้ย่อ ๆ เป็น L แต่อาจเขียนได้ทั้ง 3 แบบ ซึ่งก็ขอให้นักศึกษาได้เข้าใจว่าคือสิ่งเดียวกัน

5.6.1 การทดสอบ Simple Hypothesis

คำว่าการทดสอบ Simple Hypothesis ในที่นี้หมายความว่าทั้งสมมุติฐานหลักและสมมุติฐานรองเป็น Simple Hypothesis ด้วยกันทั้งคู่หรือมิเช่นนั้นก็เป็นสมมุติฐานที่สามารถจัดให้เป็น Simple Hypothesis ได้

ทฤษฎีที่ช่วยให้สามารถตัดสินใจเลือกระหว่าง L_0 หรือ L_1 หรือนัยหนึ่งตัดสินใจว่าจะเชื่อถือสมมุติฐานใดในระหว่าง H_0 และ H_1 ก็คือ NPL ซึ่งปรากฏดังนี้

ทฤษฎี 6.1 Neyman Pearson Lemma (NPL)

ให้ $(X_1, x_2, \ldots X_n)$ เป็น Sampled Random Variables ที่สุ่มมาจากกลุ่มประชากรที่มี pdf เป็น $f_X(x; \theta)$ และมีสมมุติฐานที่ต้องการทดสอบดังนี้คือ

$$H_0 : \theta = \theta' \text{ vs } H_1 : \theta = \theta''$$

ให้ $L_0 = \prod_{i=1}^{n} f_X(x; \theta')$ เป็น Likelihood Function ที่สอดคล้องกับพารามิเตอร์ตาม H_0

ให้ $L_1 = \prod_{i=1}^{n} f_X(x; \theta'')$ เป็น Likelihood Function ที่สอดคล้องกับพารามิเตอร์ตาม H_1

ให้ k เป็นค่าคงที่ใด ๆ และ R เป็นอนุเซทของ Sample Space ที่มีผลให้

1. $\frac{L_0}{L_1} \leqslant k$ เมื่อ $(x_1, x_2, \ldots x_n) \in R'$

2. $\frac{L_0}{L_1} \geqslant k$ เมื่อ $(x_1, x_2, \ldots, x_n) \in \overline{R}$

3. $Pr.\{(X_1, X_2, \ldots X_n) \in R | H_0\} = a$

แล้ว R จะเป็น BCR ขนาด α สำหรับทดสอบสมมุติฐาน $H_0 : \theta = \theta'$ vs $H_1 : \theta = \theta''$

หมายเหตุ ถ้าดูตามทฤษฎีแล้วนักศึกษาอาจไม่เข้าใจและมองไม่เห็นประโยชน์ตลอดจนความเกี่ยวเนื่องกันในระหว่างนิยามและสัญลักษณ์ที่เกี่ยวข้อง ก่อนที่จะพิสูจน์ผู้เขียนขอขยายความเกี่ยวกับ NPL ในเชิงปฏิบัติเพื่อความเข้าใจที่ดีขึ้นดังนี้

---

[1] อสมการในที่นี้เราอาจใช้ $\leqslant$ หรือ $<$ และ $\geqslant$ หรือ $>$ ก็ได้เพราะไม่มีความหมายในทางปฏิบัติมากนักเพราะค่าของตัวสถิติที่เท่ากับค่าวิกฤติมักก่อให้เกิดปัญหาในการตัดสินใจ โดยปกติแล้วเราจะไม่กล้าตัดสินใจเมื่อพบกรณีเช่นนี้

เมื่อต้องการทดสอบสมมุติฐาน $H_0 : \theta = \theta'$ vs $H_1 : \theta = \theta''$ และสุ่มตัวอย่าง $(X_1, X_2, ..., X_n)$ มาจากกลุ่มประชากรที่มี pdf เป็น $f_X(x; \theta)$ โดยมีพารามิเตอร์ θ เป็นตัวแสดงคุณลักษณะทางประชากรของกลุ่มประชากรดังกล่าว

ดังนั้น เราจะปฏิเสธสมมุติฐานหลัก (หรือยอมรับสมมุติฐานรอง) เมื่อ $\frac{L_0}{L_1} \leqslant k$ และยอมรับสมมุติฐานหลัก (หรือเชื่อว่า L_0 ถูกต้อง) เมื่อ $\frac{L_0}{L_1} \geqslant k$ ทั้งนี้ k คือค่าคงที่ (Quantile) บนแกนนอนใต้โค้งของตัวสถิติ $u(x_1, x_2, ..., x_n; \theta)$ ที่เกิดขึ้นจาก Likelihood Ratio ซึ่งสามารถปรับค่าไปมาได้ จนกระทั่งทำให้พื้นที่ในเขตวิกฤติมีค่าเท่ากับระดับความเสี่ยงประเภทที่ 1 (α-error) ตามคุณสมบัติข้อ 3 ของ NPL

โดยนัยนี้ หลักปฏิบัติเพื่อพัฒนาตัวทดสอบ (Test Statistics) สำหรับสมมุติฐานที่สนใจจึงปรากฏดังนี้

1. ศึกษาดูว่าตัวแปรสุ่มที่สนใจนั้นมีการแจกแจงแบบใด
2. กำหนด Parameter Space เพื่อจะได้ทราบว่าพารามิเตอร์ของตัวแปรสุ่มประกอบด้วยอะไรบ้าง มีตัวใดที่ทราบค่าแล้วหรือว่าไม่ทราบค่าทั้งหมด
3. กำหนดข้อสมมุติฐาน
4. สร้าง Likelihood Function L_0 และ L_1 แล้วสร้าง Likelihood Ratio ในรูปอสมการ $\frac{L_0}{L_1} \leqslant k$ ซึ่งมีความหมายว่า เราจะปฏิเสธ H_0 (หรือไม่เชื่อว่า L_0 ถูกต้อง) ถ้า $\frac{L_0}{L_1} \leqslant k$
5. คำนวณหาตัวสถิติจาก Likelihood Ratio, $\frac{L_0}{L_1}$ ซึ่งก็คือฟังก์ชั่นของตัวแปรสุ่มคือ $u(x_1, x_2, ..., x_n; \theta)$ ดังที่ได้กล่าวมาแล้ว
6. พิสูจน์โดยอาศัยเทคนิคต่าง ๆ ที่เห็นว่าเหมาะสมเพื่อหาฟังก์ชั่นการแจกแจง (Sampling Distribution) ของ $u(x_1, x_2, ..., x_n; \theta)$ เทคนิคดังกล่าวอาจเป็น cdf, mgf หรือ Transformation of Random Variable ก็ได้
7. คำนวณหาค่า k โดยอาศัย Sampling Distribution ในข้อ 6 ในลักษณะที่ k เป็นค่าคงที่ที่สามารถปรับค่าไปมาได้จนกระทั่งทำให้เกิดความเสี่ยงที่จะปฏิเสธสมมุติฐานหลักทั้ง ๆ ที่สมมุติฐานหลักเป็นจริงมีค่าเท่ากับค่าระดับความเสี่ยงที่นักวิจัยพอจะยอมรับได้ (คำนวณโดยอาศัยคุณสมบัติข้อ 3 ของ NPL)
8. ตัวทดสอบที่ได้มาโดยวิธีนี้จะเป็นตัวทดสอบที่ดีที่สุดเสมอคือเป็น Best Test ซึ่งนักศึกษาจะตรวจสอบความจริงข้อนี้ได้จากการพิสูจน์ทฤษฎีซึ่งจะได้กล่าวถึงต่อไป

อนึ่ง ขอให้นักศึกษาตัดความกังวลเกี่ยวกับลักษณะของอสมการทิ้งไปเสียเพราะ NPL อาจเสนอไว้ในรูปของ $\leqslant$ หรือ $<$ และ $\geqslant$ หรือ $>$ ก็ได้ เรื่องนี้ไม่มีความหมายในทางปฏิบัติมากนัก โดยปกติแล้วค่าของตัวสถิติที่เท่ากับค่าวิกฤติมักก่อให้เกิดความรวนเรในการตัดสินใจเสมอ เช่นในการทดสอบสมมุติฐานเกี่ยวกับค่าเฉลี่ยของกลุ่มประชากรปกติเมื่อทราบค่าของความแปรปรวนและพบว่า $Z_c = Z_{tab}$ เราจะมิกล้าที่จะตัดสินใจว่าควรยอมรับหรือปฏิเสธสมมุติฐานหลักในทันที ดังนั้นเพื่อตัดความกังวลในเรื่องอสมการและขจัดปัญหาที่ก่อให้เกิดความรวนเรไม่แน่ใจที่จะตัดสินใจเราจึงนิยมปฏิบัติเป็น 2 วิธีคือ

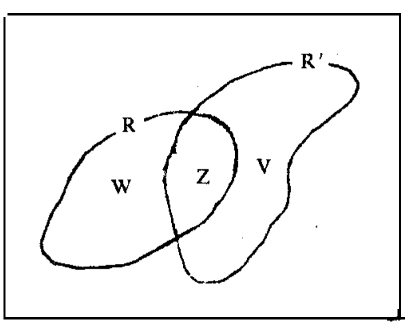
ก. กำหนดให้แน่ชัดเสียแต่ต้นว่าจะให้เขตวิกฤตินับรวมเครื่องหมายเท่ากับไว้ด้วยหรือไม่ ถ้านับรวมด้วยก็ใช้กติกาการตัดสินใจตามวิธีของเนย์แมน-เปียสัน ว่า "เราจะปฏิเสธสมมุติฐานหลักเมื่อ $\frac{L_0}{L_1} \leqslant k$" ถ้าไม่นับรวมก็ให้ใช้กติกาว่า "เราจะปฏิเสธสมมุติฐานหลักเมื่อ $\frac{L_0}{L_1} < k$" แล้วคำนวณหาค่า k ตามวิธีข้างต้น

ข. ใช้ควอนไทล์ลำดับที่ละเอียดมาก ๆ โดยใช้ทศนิยมอย่างน้อย 3 หลัก ซึ่งจะมีผลให้ค่าของตัวสถิติกับค่าวิกฤติคลาดเคลื่อนกันอยู่เสมอ

ต่อไปนี้จะแสดงการพิสูจน์ทฤษฎีของเนย์แมน-เบียร์สัน ขอให้นักศึกษาพยายามใช้ความสังเกตให้มากเพราะอาจสับสนและเข้าใจยากในตอนต้น

พิสูจน์ ให้ R เป็นเขตวิกฤติที่พัฒนาขึ้นมาจากทฤษฎีของเนย์แมน-เบียร์สัน และสมมุติว่าเขตวิกฤตินี้ยังมิใช่เขตที่ดีที่สุดกล่าวคือ ยังเป็นเขตที่ไม่ให้ Power สูงที่สุดหรือนัยหนึ่งยังไม่ให้ความเสี่ยงประเภทที่ 2 ต่ำที่สุด

สมมุติว่า R' คือเขตวิกฤติที่ให้ความเสี่ยงประเภทที่ 2 ต่ำที่สุดและเขตวิกฤติทั้งสองคือ R และ R' มีขนาดเดียวกันเท่ากับ α หรือมีพื้นที่ในเขตวิกฤติเท่ากันเท่ากับ α



ดังนั้น α = P_r (ปฏิเสธสมมุติฐานหลักทั้ง ๆ ที่สมมุติฐานหลักถูกต้อง)

$$\alpha = \int\int_R \cdots \int L_0(x_1, x_2, \ldots x_n)dx_1dx_2\ldots dx_n$$
$$= \int\int_{R'} \cdots \int L_0(x_1, x_1, \ldots, x_n)dx_1dx_2\ldots dx_n$$

ดังนั้น $\int\int_R \cdots \int L_0(x_1, x_2, \ldots, x_n)dx_1dx_2\ldots dx_n = \int\int_{R'} \cdots \int L_0(x_1, x_2, \ldots, x_n)dx_1dx_2\ldots dx_n$

แต่เนื่องจากเขตวิกฤติ R และ R′ มีเขตร่วมกันคือ z เมื่อหักเขตร่วม z ออกจึงพบว่า

$$\int\int_w \cdots \int L_0(x_1,x_2,\ldots,x_n)dx_1dx_2\ldots dx_n = \int\int_v \cdots \int L_0(x_1,x_2,\ldots,x_n)dx_1dx_2\ldots dx_n$$

เมื่อคูณตลอดด้วย k จะพบว่า

$$k\int\int_w \cdots \int L_0(x_1,x_2,\ldots,x_n)dx_1dx_2\ldots dx_n = k\int\int_v \cdots \int L_0(x_1,x_2,\ldots,x_n)dx_1dx_2\ldots dx_n \quad \ldots\ldots\ldots.(1)$$

และพบว่า Power ของเขตวิกฤติ R และ R′ มีค่าเท่ากับ Π และ Π' ตามลำดับโดยที่

$$\Pi = Pr \text{ (ปฏิเสธสมมุติฐานหลักเมื่อสมมุติฐานรองถูกต้อง)}$$
$$= \int\int_R \cdots \int L_1(x_1,x_2,\ldots,x_n)dx_1dx_2\ldots dx_n$$

และในขณะเดียวกันเราจะพบว่า

$$\Pi' = \int\int_{R'} \cdots \int L_1(x_1,x_2,\ldots,x_n)dx_1dx_2\ldots dx_n$$

โดยนัยดังกล่าวข้างต้น เขตวิกฤติ R′ จะทำให้มี Power เพิ่มขึ้นคือเพิ่มขึ้นเป็นปริมาณ $\delta\Pi$ เมื่อ $\delta\Pi = \Pi' - \Pi$

$$\delta\Pi = \int\int_v \cdots \int L_1(x_1,x_2,\ldots,x_n)dx_1dx_2\ldots dx_n \;\; \int\int_w \cdots \int L_1(x_1,x_2,\ldots,x_n dx_1 dx_2dx_3\ldots dx_n \quad \ldots\ldots\ldots.(2)$$

แต่เพราะว่าในขั้นต้นนี้ เรามีเฉพาะเขตวิกฤติ R ดังนั้นเขตนอกเหนือไปจากนี้จะเป็นเขตยอมรับทั้งหมด ด้วยเหตุนี้ w จึงตกอยู่ในเขตวิกฤติเดิม ขณะที่ v จะตกอยู่ในเขตยอมรับเดิม ดังนั้นโดยอาศัยทฤษฎีของเนย์แมน-เบียร์สัน

$$=> \quad \frac{L_0(x_1,x_2,\ldots,x_n|w)}{L_1(x_1,x_2,\ldots,x_n|w)} < k$$

หรือ $\int\int_w \ldots \int L_0(x_1,x_2,\ldots,x_n)dx_1dx_2\ldots dx_n < k \int\int_w \ldots \int L_1(x_1,x_2,\ldots,x_n)$

$dx_1dx_2\ldots dx_n$ ....(3)

และ $$\frac{L_0(x_1,x_2,\ldots,x_n|v)}{L_1(x_1,x_2,\ldots,x_n|v)} > k$$

หรือ $\int\int_v \ldots \int L_0(x_1,x_2,\ldots,x_n)dx_1dx_2\ldots dx_n > k \int\int_v \ldots \int L_1(x_1,x_2,\ldots,x_n)$

$dx_1dx_2\ldots dx_n$ ....(4)

จากสมการ (1) และ (4) จะพบว่า

$\int\int_w L_0(x_1,x_2,\ldots,x_n)dx_1dx_2\ldots dx_n > k \int\int_w \ldots \int L_1(x_1,x_2,\ldots,x_n)$

$dx_1dx_2\ldots dx_n$ ....(5)

จากสมการ (3) และ (5) จะพบว่า

(5) − (3) $= k\int\int_v \ldots \int L_1(x_1,x_2,\ldots,x_n)dx_1dx_2\ldots dx_n - k\int\int_w \ldots \int L_1(x_1,x_2,\ldots,x_n)$

$dx_1dx_2\ldots dx_n < 0$

=> $\int\int_v \ldots \int L_1(x_1,x_2,\ldots,x_n)dx_1dx_2\ldots dx_n - \int\int_w \int L_1(x_1,x_2,\ldots,x_n)$

$dx_1dx_2\ldots dx_n \leqslant \frac{0}{k} = 0$

หรือ $\delta\Pi < 0$

นั่นคือ ถ้าถือว่าเขตวิกฤติ R′ เหมาะสมกว่าเขต R แล้ว จะปรากฏว่า Power เลวลง (ถ้า $\delta\Pi = 0$ แสดงว่ามี Power เท่ากัน)

ดังนั้นเขตวิกฤติ R ที่พัฒนาขึ้นมาจากทฤษฎีของเนย์แมน-เบียร์สัน จึงเป็นเขตวิกฤติที่ให้ Power สูงกว่า เหมาะสมกว่า หรือเป็นเขตวิกฤติที่ดีที่สุด (Best Critical Region)

5.6.2 การทดสอบ Composite Hypothesis

ในกรณีของ Composite Hypothesis ซึ่งก็คือ กรณีเมื่อสมมุติฐานรองจัดหรือเสนอพารา-มิเตอร์ไว้ในรูปของเซท เช่น $H_1 : \theta > \theta_0$ หรือ $H_1 : \theta < \theta_0$ หรือ $H_1 : \theta \neq \theta_0$[1]

---

[1] $H_1 : \theta < \theta_0$ และ $H_1 : \theta > \theta_0$ สามารถใช้ NPL ได้โดยอนุโลมเมื่อถือว่า $H_1 : \theta = \theta < \theta_0$ หรือ $H_1 : \theta = \theta_1 > \theta_0$ ซึ่งหมายความว่าเราจัดพารามิเตอร์ตาม H_1 ออกเป็นเซทของ Simple Hypothesis $H_1 : \theta = \theta_1$ โดยถือว่า σ_j ต้องมีค่าน้อยกว่า θ_0 หรือมากกว่า θ_0 แล้วแต่กรณี

ดังนั้น ถ้า R คือเขตวิกฤติขนาด α และ k คือตัวคงที่ใด ๆ แล้ว $(k < 1)$

$$1.\ \lambda = \frac{L\hat{\omega}}{LB} \leqslant k \quad \text{เมื่อ } (x_1, x_2, \ldots, x_n) \in R\ ;\ k < 1$$ [1]

และ

$$2.\ \lambda = \frac{L\hat{\omega}}{LB} \geqslant k \quad \text{เมื่อ } (x_1, x_2, \ldots, x_n) \in \bar{R}$$

หมายเหตุ MLRT ให้ตัวทดสอบ (test) ที่ดี (good test) แต่มิได้ยืนยันว่าตัวทดสอบที่ได้จะเป็น Best Test เพราะวิธี MLR อาศัยใช้วิธีการของ NPL โดยอนุโลมเท่านั้น อนึ่งในทางปฏิบัติ เราควรจะต้องแจกแจงพารามิเตอร์ตาม ω และ Ω ให้ชัดเจนเพื่อมิให้เกิดความผิดพลาดอันเนื่องมาจากการละเลยไม่ประมาณค่าของพารามิเตอร์บางตัว

เรื่องนี้ขอให้คำอธิบายเพิ่มเติมไว้ดังนี้คือ

จาก Likelihood Ratio คือ $\frac{L_0}{L_1}$ หรือ $\frac{L\hat{\omega}}{L\hat{\Omega}}$ ผลลัพธ์ของการเปรียบเทียบอัตราส่วนที่เราต้องการก็คือ ตัวสถิติ $u(x_1,x_2,\ldots,x_n;\theta)$ สำหรับกรณี Single Parameter หรือ $u(x_1,x_2,\ldots,x_n;\theta_1,\theta_2,\ldots,\theta_k)$ สำหรับกรณี Several Parameter แล้วใช้เทคนิคต่าง ๆ เช่น mgf, cdf หรือ transformation technique หา Sampling Distribution หรือการแจกแจงของตัวสถิติ $u(x_1,x_2,\ldots,x_n;\theta)$ หรือ $u(x_1,x_2,\ldots,x_n;\theta_1,\theta_2,\ldots,\theta_k)$ แล้วแต่กรณีเพื่อคำนวณหาค่า k อีกต่อหนึ่ง ซึ่งกระทำได้โดยอาศัยสมการ Probability of Type I Error ทั้งนี้เพราะ k คือค่าคงที่หรือ quantile บนแกนนอนใต้โค้งของตัวสถิติ $u(x_1, x_2, \ldots, x_n;\theta)$ ซึ่ง k จะสามารถปรับค่าได้จนกระทั่งทำให้พื้นที่ใต้โค้งในเขตวิกฤติมีค่าเท่ากับระดับความเสี่ยง (α) สูงสุดเท่าที่ผู้วิจัยจะยอมให้มีได้ (Max α)

---

[1] หรือ $\frac{\text{Sup } L_\omega}{\text{Sup } L_\Omega} \leqslant k$ เมื่อ $(x_1, x_2, \ldots, x_n) \in R$ และ $\frac{\text{Sup } L_\omega}{\text{Sup } L_\Omega} \geqslant k$ เมื่อ $(x_1, x_2, \ldots, x_n) \in R$ ซึ่งหมายความว่าเราจะปฏิเสธสมมุติฐานหลักเมื่อ $\frac{\text{Sup } L_\omega}{\text{Sup } L_\Omega} \leqslant k$ โดยที่ k คือตัวคงที่ใด ๆ (Quantile) ที่สามารถปรับค่าไปมาได้ จนกระทั่งทำให้พื้นที่ในเขตวิกฤติมีค่าเท่ากับความเสี่ยงประเภทที่ 1 (α-risk)

และด้วยเหตุที่ตัวสถิติคือฟังก์ชั่นของค่าสังเกต (Function of Observation หรือ Function of Random Variable) ที่ไม่มีพารามิเตอร์ปะปนอยู่หรือถ้ามีพารามิเตอร์ปะปนอยู่ พารามิเตอร์นั้นจะต้องเป็นพารามิเตอร์ที่ทราบค่าแล้วหรือประมาณค่าไว้แล้วด้วยตัวประมาณค่าหนึ่งค่าใดที่เหมาะสม ด้วยเหตุนี้ใน MLRT เราจึงมีความจำเป็นต้องประมาณค่าพารามิเตอร์ทุกตัวที่เกี่ยวข้องถ้าเป็นพารามิเตอร์ที่ไม่ทราบค่า ทั้งนี้เพื่อให้ U เป็นตัวสถิติ การเสนอ Parameter Space ω และ Ω อย่างถี่ถ้วน ทำให้เราทราบได้ว่ามีพารามิเตอร์ตัวใดบ้างที่พึงประมาณค่าเสียก่อน ทั้งนี้เพื่อป้องกันความผิดพลาดเนื่องจากละเลยหรือหลงลืมประมาณค่าพารามิเตอร์บางตัวซึ่งมีผลให้พารามิเตอร์ที่ไม่ทราบค่าหรือยังไม่ได้ประมาณค่าปะปนเข้าไปในฟังก์ชั่น U ซึ่งมีผลให้ U ไม่เป็นตัวสถิติ ผลลัพธ์ก็คือเราไม่อาจคำนวณหาฟังก์ชั่น pdf ของ U ได้ ซึ่งส่งผลสะท้อนให้ไม่ทราบค่า quantile[1]

อย่างไรก็ตามปัญหาข้างต้นนี้จะปรากฏขึ้นกับเฉพาะกรณีที่พัฒนา test สำหรับ Composite Hypothesis เท่านั้น และจะไม่ปรากฏกับกรณีของ Simple Hypothesis หรือ One Sided Test เพราะในกรณีเช่นนี้เราทราบค่าที่แจ้งชัดของพารามิเตอร์ตาม H_0 และ H_1 แล้วซึ่งย่อมชี้ให้เห็นว่าจะไม่มีพารามิเตอร์ที่ไม่ทราบค่า หรือยังไม่ได้ประมาณค่าปะปนอยู่ในฟังก์ชั่น U แต่อย่างใด

สิ่งหนึ่งที่ควรทราบก็คือ MLR มิได้ให้ best test และเป็นวิธีการที่ใช้เมื่อไม่อาจสร้าง best test โดยอาศัย NPL ได้ ซึ่งในกรณีเช่นนี้ MLR จะให้ test ที่พอจะใช้แทนกันได้ แต่จะใช้ได้ดีเฉพาะเมื่อกลุ่มตัวอย่างมีขนาดใหญ่มากพอเท่านั้น

---

1 ควอนไทล์เป็นคำกลาง ๆ ใช้แทนความหมายของอันดับต่าง ๆ ที่สอดคล้องกับพื้นที่ใต้โค้ง นักศึกษาเคยพบเห็นแต่เปอร์เซนต์ไทล์ เดไซล์และควอไทล์ เมื่อพบควอนไทล์อาจไม่เข้าใจความจริงควอนไทล์เป็นสถานการณ์ทั่วไปของเรื่องเหล่านั้น ถ้าเป็นควอนไทล์อันดับ .50 ก็คือเปอร์เซนต์ไทล์ที่ 50 หรือเดไซล์ ที่ 5 หรือควอไทล์ที่ 2 ถ้าเป็นควอนไทล์อันดับ .20 ก็คือเปอร์เซนต์ไทล์ที่ 20 หรือเดไซล์ที่ 2 แต่ถ้าเป็นควอนไทล์อันดับ .0015 เราไม่อาจทราบเปอร์เซนต์ไทล์ได้ ขอให้สังเกตว่าควอนไทล์เป็นเรื่องที่เตรียมไว้สำหรับจัดอันดับในสเกลที่ถี่มาก ๆ ที่เปอร์เซนต์ไทล์ที่ไม่อาจรับได้ การจัดอันดับที่หยาบที่สุดคือควอไทล์ จัดได้เพียง 4 อันดับ แก้ด้วยเดไซล์ซึ่งสามารถจัดได้ถึง 10 อันดับ แต่ยังหยาบอยู่จึงแก้ด้วยเปอร์เซนต์ไทล์ซึ่งสามารถจัดได้ถึง 100 อันดับ ถ้าจัดเกินกว่า 100 อันดับ ต้องแก้ด้วยควอนไทล์