

สถิติกับงานวิทยาเอ็นโดดอนต์ ตอนที่ 4: การเปรียบเทียบค่าเฉลี่ย การทดสอบสถิติพื้นฐานทางเอ็นโดดอนติกส์ (Statistics in Endodontics Part 4: Comparing Means—Basic Statistical Tests in Endodontics)

สิทธิโชค โอศิริ

สาขาวิชาวิทยาเอ็นโดดอนต์ ภาควิชาทันตกรรมหัตถการและวิทยาเอ็นโดดอนต์ คณะทันตแพทยศาสตร์ มหาวิทยาลัยมหิดล

บทคัดย่อ

การเลือกใช้การทดสอบทางสถิติที่เหมาะสมเป็นสิ่งจำเป็นสำหรับการสรุปผลที่ถูกต้องในงานวิจัยทางเอ็นโดดอนติกส์และการปฏิบัติทางคลินิกที่อิงหลักฐานเชิงประจักษ์ บทความนี้ครอบคลุมเกี่ยวกับการทดสอบแบบพาราเมตริกและนอนพาราเมตริกสำหรับการเปรียบเทียบค่าเฉลี่ย โดยเน้นย้ำความสำคัญอย่างยิ่งของการตรวจสอบข้อสมมติทางสถิติก่อนการวิเคราะห์ รวมข้อสมมติพื้นฐาน ได้แก่ ความต่อเนื่องของตัวแปรตาม ความเป็นอิสระของการสังเกต การกระจายตัวแบบปกติ ความเท่าเทียมของความแปรปรวน และการไม่มีค่าผิดปกติ พร้อมวิธีการประเมินโดยใช้การตรวจสอบด้วยสายตาและการทดสอบเชิงตัวเลข การทดสอบแบบพาราเมตริกที่นำเสนอประกอบด้วย One-Sample t-test, Independent Samples t-test, Paired Samples t-test, One-Way ANOVA และ One-Way Repeated Measures ANOVA พร้อมคำแนะนำเกี่ยวกับการทดสอบหลังการวิเคราะห์สำหรับการเปรียบเทียบหลายกลุ่ม เมื่อข้อสมมติแบบพาราเมตริกไม่เป็นไปตามที่กำหนด บทความจะนำเสนอทางเลือกแบบนอนพาราเมตริกพร้อมการทดสอบหลังการวิเคราะห์ที่สอดคล้องกัน กรอบการตัดสินใจจะแนะนำการเลือกการทดสอบโดยพิจารณาจากการออกแบบการวิจัย ลักษณะของข้อมูล และการตรวจสอบข้อสมมติ การทดสอบแต่ละชนิดประกอบด้วยข้อสมมติเฉพาะ สมมติฐาน และตัวอย่างการประยุกต์ใช้ในงานวิจัยทางเอ็นโดดอนติกส์ การเข้าใจและประยุกต์ใช้หลักการเหล่านี้อย่างถูกต้องจะเพิ่มความน่าเชื่อถือของผลการวิจัย ลดความเสี่ยงของข้อผิดพลาดทางสถิติ และส่งเสริมการตัดสินใจทางคลินิกที่อิงหลักฐานในทางเอ็นโดดอนติกส์

คำสำคัญ: การเปรียบเทียบค่าเฉลี่ย, การทดสอบแบบนอนพาราเมตริก, การทดสอบแบบพาราเมตริก, ระเบียบวิธีวิจัย, สถิติ

Abstract

Appropriate statistical test selection is essential for valid conclusions in endodontic research and evidence-based clinical practice. This article provides a comprehensive guide to parametric and non-parametric tests for comparing means, emphasizing the critical importance of verifying statistical assumptions before analysis. The review covers fundamental assumptions—continuous dependent variables, independence of observations, normal distribution, homogeneity of variances, and absence of outliers—with practical methods for assessment using visual inspection and numerical tests. Parametric tests discussed include One-Sample t-test, Independent Samples t-test, Paired Samples t-test, One-Way ANOVA, and One-Way Repeated Measures ANOVA, with detailed guidance on post-hoc testing procedures for multiple comparisons. When parametric assumptions are violated, non-parametric alternatives are presented with corresponding post-hoc procedures. A decision framework guides test selection based on research design, data characteristics, and assumption verification. Each test includes specific assumptions, hypotheses, and practical examples from endodontic research. Understanding and correctly applying these principles enhances research reliability, reduces statistical errors, and promotes evidence-based clinical decision-making in endodontics.

Keywords: Comparing means, Non-parametric tests, Parametric tests, Research methodology, Statistics

Correspondence: สิทธิโชค โอศิริ

เลขที่ 6 ถนนโยธี แขวงทุ่งพญาไท เขตราชเทวี กรุงเทพฯ 10410

โทรศัพท์ 0-2200-7825-6

Email address: sittichoke.osi@mahidol.ac.th

Received: 14 October 2025

Revised: 5 December 2025

Accepted: 12 December 2025

บทนำ (Introduction)

การปฏิบัติงานทางวิทยาเอ็นโดดอนต์ในยุคปัจจุบัน จำเป็นต้องอาศัยหลักฐานเชิงประจักษ์ (evidence-based practice) เป็นรากฐานสำคัญในการตัดสินใจรักษา เพื่อให้ผู้ป่วยได้รับการดูแลที่มีประสิทธิภาพสูงสุดและปลอดภัยที่สุดในขณะที่วงการทันตกรรมมีการพัฒนาวัสดุ เทคนิคการรักษา และแนวทางการดูแลผู้ป่วยใหม่ ๆ อย่างต่อเนื่อง การประเมินผลอย่างเป็นระบบและมีหลักวิชาการจึงมีความจำเป็น การเปรียบเทียบทางสถิติจึงเป็นเครื่องมือที่มีบทบาทสำคัญในกระบวนการนี้

การเปรียบเทียบค่าเฉลี่ยเป็นงานพื้นฐานในการวิจัยทางคลินิกและชีววิทยา โดยเป็นรากฐานสำหรับการทดสอบสมมติฐานและการตัดสินใจที่มีข้อมูลรองรับ วิธีการทางสถิติช่วยให้ทันตแพทย์สามารถหาข้อสรุปที่ชัดเจน เชื่อถือได้ และทำซ้ำได้เกี่ยวกับประสิทธิผลของการรักษา คุณสมบัติของวัสดุทันตกรรม และการตอบสนองของผู้ป่วยต่อการรักษา การเปรียบเทียบค่าเฉลี่ย (means) ของตัวแปรต่าง ๆ ทางสถิติ เช่น แรงยึดติดเฉลี่ย (mean bond strength) หรือจำนวนโคโลนีแบคทีเรียเฉลี่ย (mean bacterial colony counts) ระหว่างกลุ่มหรือสถานะต่างๆ ช่วยให้ทันตแพทย์สามารถกำหนดได้ว่าความแตกต่างที่สังเกตได้ในข้อมูลตัวอย่างสะท้อนถึงความแตกต่างที่แท้จริงในประชากรต้นแบบหรือเป็นเพียงผลจากความบังเอิญ (random chance) เท่านั้น [1]

การเปรียบเทียบค่าเฉลี่ยมีบทบาทสำคัญในการขับเคลื่อนนวัตกรรมทางเอ็นโดดอนติกส์ ทั้งการพัฒนาวัสดุใหม่และการประเมินผลการรักษาทางห้องปฏิบัติการและทางคลินิก อย่างไรก็ตาม การวิเคราะห์กลายเป็นเรื่องที่ต้องพิจารณาเมื่อต้องจัดการกับมากกว่าสองกลุ่ม เนื่องจากการเปรียบเทียบแบบคู่หลายครั้งเพิ่มความเสี่ยงของข้อผิดพลาดประเภทที่ 1 (type I error) จึงจำเป็นต้องมีการจัดการทางสถิติอย่างระมัดระวัง [2] หากขาดความเข้าใจที่ถูกต้อง การตัดสินใจทางคลินิกอาจอาศัยหลักฐานที่ไม่น่าเชื่อถือหรือตีความผิด ซึ่งอาจส่งผลกระทบต่อคุณภาพการดูแลผู้ป่วยและความถูกต้องในการเลือกวิธีการรักษา

การตรวจสอบข้อสมมติอย่างเพียงพอเป็นสิ่งสำคัญยิ่งก่อนดำเนินการเปรียบเทียบค่าเฉลี่ย ข้อสมมติหลัก ได้แก่ ความเป็นปกติของข้อมูล ความเท่าเทียมของความแปรปรวน

ระหว่างกลุ่ม และความเป็นอิสระของการสังเกต [3] การละเมิดข้อสมมติเหล่านี้สามารถนำไปสู่อัตราความผิดพลาดที่สูงขึ้นหรืออำนาจ (power) ทางสถิติที่ลดลง ทำให้ความถูกต้องของการอนุมาน (inferential) ลดลง

บทความฉบับนี้จัดทำขึ้นโดยมีจุดมุ่งหมายเพื่อให้ทันตแพทย์และทันตแพทย์สามารถเลือกใช้วิธีการทางสถิติที่เหมาะสมกับลักษณะของข้อมูลและคำถามการวิจัย ตีความผลได้อย่างถูกต้องและมีวิจารณ์โดยเข้าใจทั้งความหมายทางสถิติและความสำคัญทางคลินิก มีส่วนร่วมในการสร้างองค์ความรู้ที่มีคุณภาพ และนำไปสู่การพัฒนาวงการวิทยาเอ็นโดดอนต์ต่อไป บทความนี้มุ่งเน้นการนำเสนอการทดสอบทางสถิติพื้นฐานที่ใช้กันอย่างแพร่หลายในงานวิจัยทางเอ็นโดดอนติกส์ โดยจะไม่กล่าวถึงการทดสอบทางสถิติขั้นสูง (advanced statistics) เช่น การวิเคราะห์ regression หลายตัวแปร (multiple regression) การวิเคราะห์การรอดชีพ (survival analysis) หรือ factorial ANOVA ซึ่งต้องการความเข้าใจเชิงลึกและการประยุกต์ใช้ที่ซับซ้อน

1. ข้อสมมติหลักของการทดสอบแบบพารามेटริก (Key Assumptions of Parametric Tests)

การตรวจสอบข้อสมมติทางสถิติควรดำเนินการตั้งแต่ขั้นตอนการออกแบบการวิจัย ไม่ใช่เพียงในขั้นตอนการวิเคราะห์ข้อมูล การละเมิดข้อสมมติอาจนำไปสู่ข้อผิดพลาดประเภทที่ 1 คือการสรุปว่ามีความแตกต่างเมื่อความจริงแล้วไม่มี หรือข้อผิดพลาดประเภทที่ 2 (type II error) คือการไม่พบความแตกต่างเมื่อความจริงแล้วมี ซึ่งอาจส่งผลกระทบต่อความถูกต้องของผลการวิจัยและการดูแลผู้ป่วย [4-6]

ข้อสมมติพื้นฐานสำหรับการทดสอบแบบพารามेटริกส่วนใหญ่ประกอบด้วย

1.1 ตัวแปรตามแบบต่อเนื่อง: ระดับช่วงหรืออัตราส่วน (Continuous Dependent Variable: Interval/Ratio Data)

ตัวแปรตาม (dependent variable) ต้องถูกวัดในระดับช่วง (interval scale) หรืออัตราส่วน (ratio scale) ซึ่งหมายความว่าข้อมูลมีค่าตัวเลขที่มีความหมาย ความแตกต่างระหว่างค่าต่างๆ มีความสม่ำเสมอ และอัตราส่วนสามารถ

ตีความได้ การทดสอบแบบพาราเมตริกใช้การคำนวณค่าเฉลี่ย (mean) และความแปรปรวน (variance) ซึ่งต้องการข้อมูลที่มีช่วงระยะห่างเท่ากันระหว่างค่าต่างๆ การใช้ข้อมูลระดับลำดับ (ordinal scale) หรือนามบัญญัติ (nominal scale) กับ การทดสอบแบบพาราเมตริกอาจให้ผลลัพธ์ที่ไม่มีความหมายหรือเข้าใจผิด

1.2 ความเป็นอิสระของการสังเกต (Independence of Observations)

ข้อมูลภายในกลุ่มและระหว่างกลุ่มต้องเป็นอิสระต่อกัน หมายความว่า การวัดผู้เข้าร่วมหรือการสังเกตรายหนึ่งไม่ส่งผลต่อ หรือ ไม่มีความสัมพันธ์กับรายอื่น ตัวอย่างการออกแบบที่ไม่เป็นอิสระ เช่น การใช้ฟันหลายซี่จากผู้ป่วยคนเดียวกันโดยไม่คำนึงถึง clustering effect อาจทำให้ค่าความคลาดเคลื่อนมาตรฐาน (standard error) เล็กเกินจริง เพิ่มโอกาสของข้อผิดพลาดประเภทที่ 1

1.3 การกระจายตัวแบบปกติของข้อมูล (Normal Distribution of Data)

ตัวแปรตามควรมีการกระจายตัวแบบปกติภายในแต่ละกลุ่มที่เปรียบเทียบ การกระจายตัวแบบปกติมักเรียกว่า “เส้นโค้งรูประฆัง” (bell-shaped curve) ซึ่งเป็นพื้นฐานสำคัญของการทดสอบแบบพาราเมตริก ในการกระจายตัวแบบปกติ

ที่สมบูรณ์แบบ ข้อมูลจะกระจายตัวอย่างสมมาตรรอบค่าเฉลี่ย โดยค่าเฉลี่ย ค่ามัธยฐาน (median) และฐานนิยม (mode) จะมีค่าใกล้เคียงกัน ความเบ้ (skewness) และความโด่ง (kurtosis) มีค่าเท่ากับศูนย์ นอกจากนี้ ประมาณ 68% ของข้อมูลจะอยู่ภายใน ± 1 ส่วนเบี่ยงเบนมาตรฐานจากค่าเฉลี่ย และประมาณ 95% ของข้อมูลจะอยู่ภายใน ± 2 ส่วนเบี่ยงเบนมาตรฐานจากค่าเฉลี่ย

วิธีการตรวจสอบความเป็นปกติ (รูปที่ 1)

การตรวจสอบความเป็นปกติของข้อมูลสามารถทำได้หลายวิธี ซึ่งแบ่งออกเป็น การตรวจสอบด้วยสายตา (visual inspection) และการทดสอบเชิงตัวเลข (numerical tests)

1. การตรวจสอบด้วยสายตา (Visual Inspection)

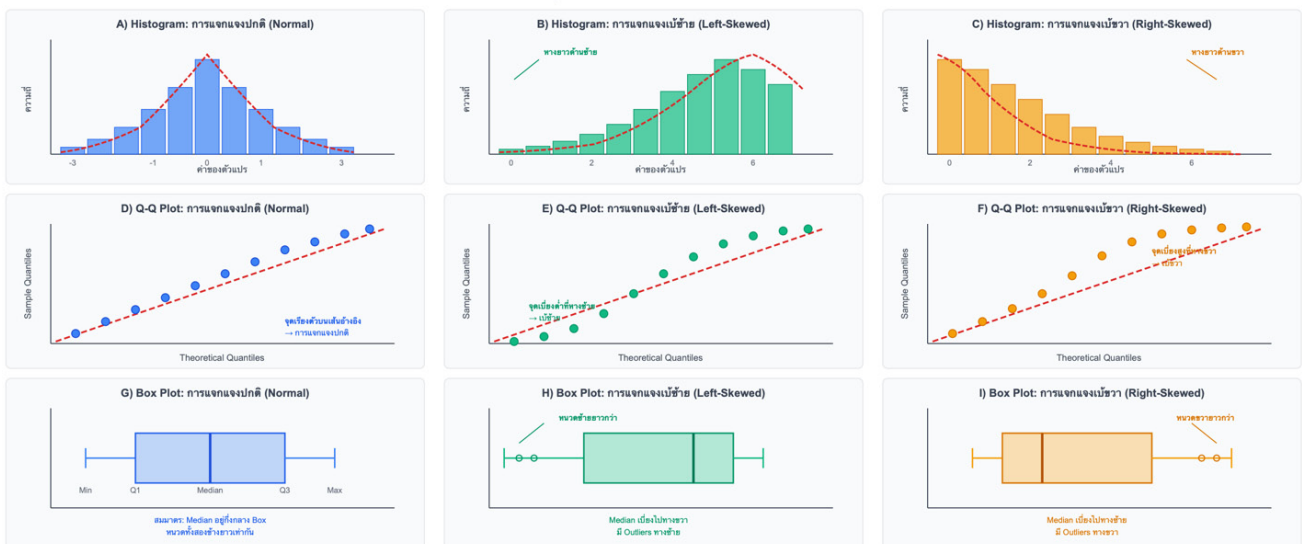
1.1 ฮิสโตแกรม (Histogram) เป็นเครื่องมือพื้นฐานที่แสดงการกระจายตัวของข้อมูลในรูปแบบแท่งกราฟ ซึ่งควรมีรูปร่างคล้ายระฆังและสมมาตร โดยเหมาะสมสำหรับข้อมูลขนาดใหญ่ ($n > 50$)

1.2 กราฟ Q-Q (Quantile-Quantile Plot)

เป็นวิธีที่มีประสิทธิภาพที่สุดในการตรวจสอบด้วยสายตา โดยเปรียบเทียบควอนไทล์ของข้อมูลจริงกับควอนไทล์ของการกระจายตัวปกติทฤษฎี หากข้อมูลเป็นไปตามการกระจายตัวแบบปกติ จุดต่าง ๆ จะอยู่บนเส้นทแยงมุมโดยประมาณ การเบี่ยงเบนจากเส้นทแยงมุมบ่งบอกถึงการไม่เป็นปกติ

การเปรียบเทียบการกระจายข้อมูลแบบปกติและไม่ปกติ

Comparison of Normal and Non-Normal Distributions



รูปที่ 1 แผนภาพแสดงการกระจายตัวของข้อมูล Histogram: การแจกแจงปกติ (A) การแจกแจงเบ้ซ้าย (B) การแจกแจงเบ้ขวา (C); Q-Q Plot: การแจกแจงปกติ (D) การแจกแจงเบ้ซ้าย (E) การแจกแจงเบ้ขวา (F); Box Plot: การแจกแจงปกติ (G) การแจกแจงเบ้ซ้าย (H) การแจกแจงเบ้ขวา (I)

1.3 แผนภาพกล่อง (Box Plot) เป็นอีกเครื่องมือที่มีประโยชน์ในการประเมินการกระจายตัวของข้อมูล โดยแสดงค่าสถิติสำคัญ 5 ค่า ได้แก่ ค่าต่ำสุด ควอร์ไทล์ที่ 1 (Q1) มัชฌิมฐาน ควอร์ไทล์ที่ 3 (Q3) และค่าสูงสุด สำหรับข้อมูลที่มีการกระจายตัวแบบปกติ กล่องจะมีลักษณะสมมาตร โดยเส้นมัธยฐานอยู่ตรงกลางกล่อง และเส้นหนวด (whiskers) ทั้งสองข้างมีความยาวใกล้เคียงกัน ในทางตรงกันข้าม หากข้อมูลมีการกระจายตัวแบบเบ้ขวา (positive skewness) เส้นมัธยฐานจะเบี่ยงไปทางซ้ายของกล่อง เส้นหนวดด้านขวาจะยาวกว่าด้านซ้าย และมักพบค่าผิดปกติ (outliers) ทางด้านขวา ส่วนการกระจายตัวแบบเบ้ซ้าย (negative skewness) จะมีลักษณะตรงกันข้าม คือ เส้นมัธยฐานเบี่ยงไปทางขวา เส้นหนวดด้านซ้ายยาวกว่า และมักพบค่าผิดปกติทางด้านซ้าย นอกจากนี้ยังมีข้อดีในการแสดงค่าผิดปกติอย่างชัดเจน ซึ่งเป็นข้อมูลสำคัญที่อาจส่งผลกระทบต่อผลการตัดสินใจเลือกใช้สถิติพารามетริกหรือนอนพารามетริก

2. การทดสอบเชิงตัวเลข (Numerical Tests)

2.1 การคำนวณความเบ้และความโด่ง (Skewness and Kurtosis)

สำหรับความเบ้ ค่าเท่ากับศูนย์แสดงถึงการกระจายตัวที่สมมาตร ค่ามากกว่าศูนย์แสดงการเบ้ขวา (positive skew) ที่มีหางยาวทางขวา ค่าน้อยกว่าศูนย์แสดงการเบ้ซ้าย (negative skew) ที่มีหางยาวทางซ้าย สำหรับความโด่ง ค่าเท่ากับศูนย์แสดงถึงความโด่งปกติ (mesokurtic) ค่ามากกว่าศูนย์แสดงการกระจายตัวที่แหลมกว่าปกติ (leptokurtic) ซึ่งมีหางหนาและค่าสุดขั้วมากกว่า ค่าน้อยกว่าศูนย์แสดงการกระจายตัวที่แบนกว่าปกติ (platykurtic)

2.2 การทดสอบ Shapiro-Wilk เป็นการทดสอบทางสถิติที่นิยมใช้มากที่สุดสำหรับตรวจสอบความเป็นปกติ เหมาะสำหรับกลุ่มตัวอย่างขนาดเล็กถึงปานกลาง ($n < 50$) และมีอำนาจทางสถิติสูงในการตรวจจับการเบี่ยงเบนจากความเป็นปกติ

- H_0 : ข้อมูลมีการกระจายตัวแบบปกติ
- H_1 : ข้อมูลไม่มีการกระจายตัวแบบปกติ
- หาก $p\text{-value} > 0.05$ ไม่ปฏิเสธ H_0 แสดงว่าข้อมูลเป็นไปตามการกระจายตัวแบบปกติ
- หาก $p\text{-value} \leq 0.05$ ปฏิเสธ H_0 แสดงว่าข้อมูลไม่เป็นไปตามการกระจายตัวแบบปกติ

2.3 การทดสอบ Kolmogorov-Smirnov

เหมาะสำหรับกลุ่มตัวอย่างขนาดใหญ่ ($n \geq 50$) แต่มีอำนาจทางสถิติต่ำกว่า Shapiro-Wilk สำหรับกลุ่มตัวอย่างขนาดเล็ก ข้อดีคือสามารถใช้ทดสอบการกระจายตัวอื่น ๆ ได้ ไม่เฉพาะการกระจายตัวปกติ

1.4 ความเท่าเทียมของความแปรปรวน (Homogeneity of Variances)

ข้อสมมตินี้เรียกอีกอย่างว่า homoscedasticity กำหนดให้ความแปรปรวน (variance) ของตัวแปรตามต้องมีค่าใกล้เคียงกันในทุกกลุ่มที่เปรียบเทียบ เมื่อข้อสมมตินี้ถูกละเมิดจะเรียกว่า heteroscedasticity หรือ heterogeneity of variance การทดสอบแบบพารามетริก เช่น t-test และ ANOVA สันนิษฐานว่าความแปรปรวนของข้อมูลในแต่ละกลุ่มเท่ากัน หากความแปรปรวนไม่เท่ากัน การทดสอบอาจให้ผลลัพธ์ที่ไม่ถูกต้อง โดยเฉพาะเมื่อขนาดกลุ่มตัวอย่างไม่เท่ากัน เนื่องจากกลุ่มที่มีความแปรปรวนมากกว่าจะมีอิทธิพลมากเกินไปต่อผลการทดสอบ

วิธีการทดสอบทางสถิติ:

1. การทดสอบ Levene's Test เป็นการทดสอบที่นิยมใช้มากที่สุด มีความทนทานต่อการเบี่ยงเบนจากความไม่เป็นปกติ (robust) และเหมาะสำหรับการเปรียบเทียบมากกว่าหรือเท่ากับ 2 กลุ่ม

- H_0 : ความแปรปรวนของทุกกลุ่มเท่ากัน
- H_1 : ความแปรปรวนของกลุ่มอย่างน้อย 1 กลุ่มแตกต่างจากกลุ่มอื่น
- หาก $p\text{-value} > 0.05$ ไม่ปฏิเสธ H_0 แสดงว่าความแปรปรวนเท่ากัน
- หาก $p\text{-value} \leq 0.05$ ปฏิเสธ H_0 แสดงว่าความแปรปรวนไม่เท่ากัน

2. การทดสอบ F-test of Equality of Variances

ใช้สำหรับเปรียบเทียบ 2 กลุ่ม โดยคำนวณอัตราส่วนของความแปรปรวนทั้งสองตามสูตร $F = s_1^2/s_2^2$ (ใช้ความแปรปรวนที่ใหญ่กว่าเป็นตัวเศษ) อย่างไรก็ตาม การทดสอบนี้ไม่ทนทานต่อการเบี่ยงเบนจากความเป็นปกติ (not robust) หากข้อมูลไม่เป็นปกติ F-test อาจให้ผลที่ไม่ถูกต้อง ดังนั้น Levene's test จึงเป็นทางเลือกที่ดีกว่า

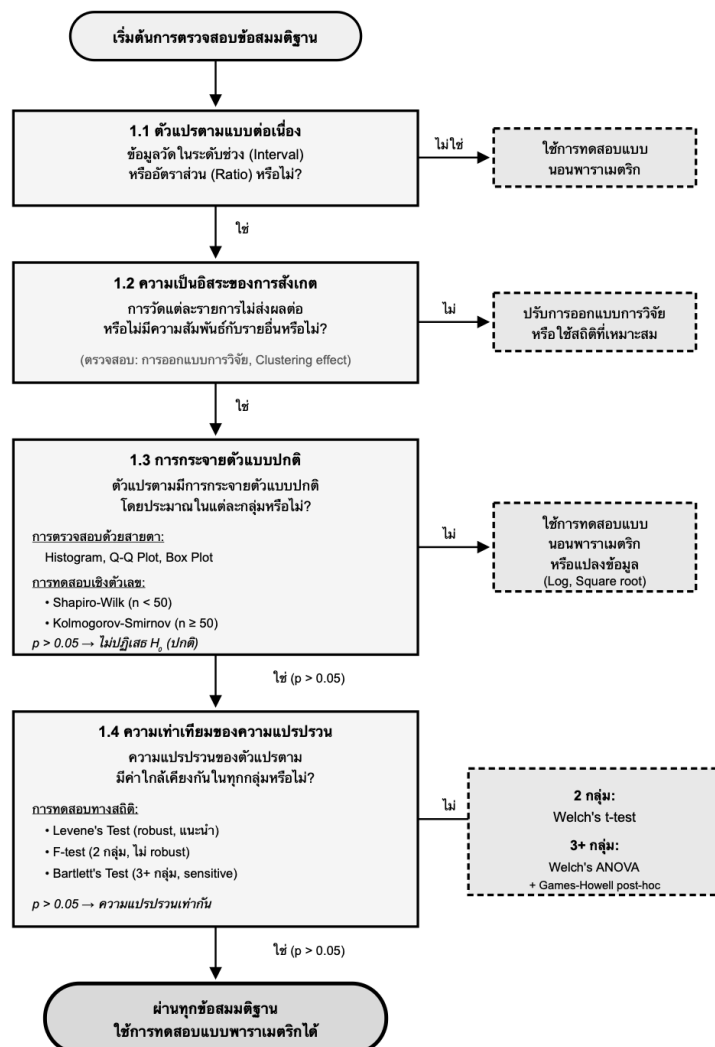
3. การทดสอบ Bartlett's Test ใช้สำหรับเปรียบเทียบ 3 กลุ่มขึ้นไป มีอำนาจทางสถิติสูงกว่า Levene's test แต่มีข้อจำกัดคือไวต่อการเบี่ยงเบนจากความเป็นปกติมาก (very sensitive) จึงไม่แนะนำให้ใช้หากข้อมูลไม่เป็นปกติ

วิธีการแก้ไขกรณีมีความเท่าเทียมของการแปรปรวนที่ผิดปกติ:

สำหรับ 2 กลุ่ม (Independent t-test): ใช้ Welch's t-test (Welch's Unequal Variances t-test) ซึ่งไม่สันนิษฐานว่าความแปรปรวนเท่ากัน มีการปรับค่า degrees of freedom ให้เหมาะสม และให้ผลลัพธ์ที่เชื่อถือได้แม้ความแปรปรวนไม่เท่ากัน แนะนำให้ใช้เป็นค่าเริ่มต้น (default) สำหรับการเปรียบเทียบ 2 กลุ่มอิสระ

สำหรับ 3 กลุ่มขึ้นไป (ANOVA): ใช้ Welch's ANOVA ซึ่งเป็นรูปแบบของ ANOVA ที่ไม่สันนิษฐานความเท่าเทียมของความแปรปรวน เหมาะเมื่อ Levene's test แสดงว่าความแปรปรวนไม่เท่ากัน นอกจากนี้ยังมีการแปลงข้อมูล (data transformation) เช่น log transformation สำหรับข้อมูลที่มีความแปรปรวนเพิ่มตามค่าเฉลี่ย หรือ square root transformation สำหรับข้อมูล count

ข้อควรพิจารณา: หากขนาดกลุ่มตัวอย่างเท่ากันในทุกกลุ่ม การทดสอบแบบพารามेटริกค่อนข้างทนทานต่อการละเมิดข้อสมมตินี้ แต่หากขนาดกลุ่มตัวอย่างต่างกันมาก และความแปรปรวนก็ต่างกันมาก ควรใช้ความระมัดระวังสูงและพิจารณาใช้ Welch's test หรือการทดสอบแบบนอนพารามेटริก จากบทความนี้สามารถสรุปได้ รูปที่ 2



รูปที่ 2 แผนภูมิการตรวจสอบข้อสมมติหลักของการทดสอบแบบพารามेटริก (Key assumptions of parametric test)

2. การทดสอบแบบพาราเมตริกทั่วไป

การทดสอบแบบพาราเมตริกที่ใช้สำหรับการเปรียบเทียบค่าเฉลี่ยมีหลายชนิด โดยการใช้ขึ้นอยู่กับจำนวนกลุ่ม ลักษณะความสัมพันธ์ของกลุ่มตัวอย่าง และวัตถุประสงค์ของการวิจัย ตารางที่ 1 สรุปการทดสอบค่าเฉลี่ยทางสถิติแบบพาราเมตริกและทางเลือกแบบนอนพาราเมตริกที่สอดคล้องกัน ส่วนรูปที่ 3 แสดงแผนภูมิการตัดสินใจเลือกการทดสอบที่เหมาะสม

2.1 การทดสอบที่ แบบกลุ่มเดียว (One-Sample t-test): เปรียบเทียบค่าเฉลี่ยของกลุ่มตัวอย่างกับค่าที่ทราบ

การทดสอบที่ แบบกลุ่มเดียวเป็นวิธีการทางสถิติที่ใช้เพื่อพิจารณาว่ากลุ่มตัวอย่างของการสังเกตที่กำหนดสามารถถูกสร้างขึ้นโดยกระบวนการที่มีลักษณะเฉพาะด้วยค่าเฉลี่ยประชากรที่ตั้งสมมติฐานไว้เฉพาะเจาะจงหรือไม่ [1]

ข้อสมมติ:

1. ตัวแปรตามเป็นข้อมูลต่อเนื่องที่วัดในระดับช่วงหรืออัตราส่วน
2. การสังเกตเป็นอิสระจากกัน (independent observations)
3. ข้อมูลมีการกระจายตัวแบบปกติ
4. ไม่มีค่าผิดปกติที่มีนัยสำคัญ (no significant outliers)

ตารางที่ 1 การทดสอบค่าเฉลี่ยทางสถิติ

ชื่อการทดสอบ	วัตถุประสงค์	จำนวนกลุ่ม	ลักษณะกลุ่ม	ประเภทข้อมูล	ทางเลือกแบบ non-parametric
One-Sample t-test	เปรียบเทียบค่าเฉลี่ยกลุ่มตัวอย่างกับค่าที่ตั้งสมมติฐาน	1	-	ต่อเนื่อง	Wilcoxon Signed-Rank (One-sample)
Independent Samples t-test	เปรียบเทียบค่าเฉลี่ยของ 2 กลุ่มที่ไม่เกี่ยวข้อง	2	อิสระ	ต่อเนื่อง	Mann-Whitney U Test
Paired Samples t-test	เปรียบเทียบค่าเฉลี่ยของ 2 กลุ่มที่เกี่ยวข้อง	2	เกี่ยวข้อง (จับคู่)	ต่อเนื่อง	Wilcoxon Signed-Rank Test
One-Way ANOVA	เปรียบเทียบค่าเฉลี่ยของ 3 กลุ่มอิสระขึ้นไป	3+	อิสระ	ต่อเนื่อง	Kruskal-Wallis H Test
One-Way Repeated Measures ANOVA	เปรียบเทียบค่าเฉลี่ยของ 3 จุดเวลาขึ้นไปในกลุ่มเดียวกัน หรือแต่ละกลุ่มมีความเกี่ยวข้องกัน	3+	เกี่ยวข้อง (จับคู่)	ต่อเนื่อง	Friedman Test

สมมติฐาน:

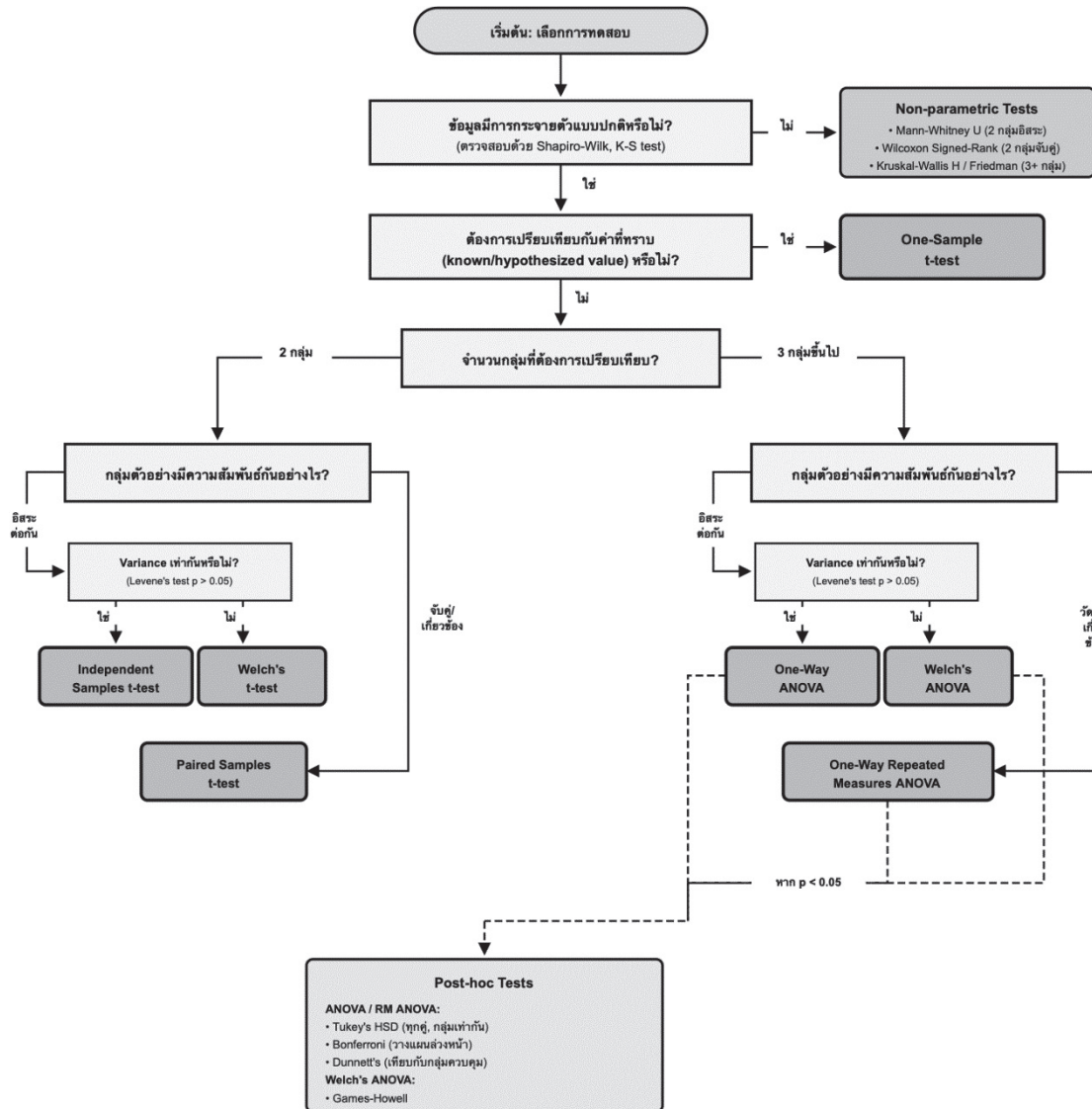
- H_0 (สมมติฐานว่าง): ค่าเฉลี่ยประชากรที่แท้จริง (μ) เท่ากับค่าที่ตั้งไว้ (μ_0)
- H_1 (สมมติฐานทางเลือก): ค่าเฉลี่ยประชากรที่แท้จริง แตกต่างจากค่าที่ตั้งไว้

ตัวอย่างการใช้ในงานเอ็นโดดอนติกส์:

- เปรียบเทียบค่าเฉลี่ยของความยาวการทำงาน (working length) ที่วัดได้จากเครื่อง apex locator กับค่ามาตรฐานจากภาพรังสี
- ทดสอบว่าค่าเฉลี่ยของแรงบิตที่ทำให้ไฟล์หักแตกต่างจากค่าที่ผู้ผลิตระบุหรือไม่
- ตรวจสอบว่าค่าเฉลี่ยของคะแนนความเจ็บปวดหลังการรักษาแตกต่างจากระดับพื้นฐานที่กำหนดไว้หรือไม่

2.2 การทดสอบที่ แบบอิสระ (Independent Samples t-test หรือ student t-test): เปรียบเทียบค่าเฉลี่ยของสองกลุ่มที่ไม่เกี่ยวข้อง

การทดสอบที่ แบบอิสระเป็นการทดสอบทางสถิติเชิงอนุมานที่ออกแบบมาเพื่อพิจารณาว่ามีความแตกต่างอย่างมีนัยสำคัญทางสถิติระหว่างค่าเฉลี่ยของสองกลุ่มที่ไม่เกี่ยวข้องกันหรือไม่ ซึ่งหมายความว่าบุคคลหรือหน่วยที่แตกต่างกันมีส่วนร่วมในการให้ข้อมูลแก่แต่ละกลุ่ม [2]



รูปที่ 3 แผนภูมิการเลือกการทดสอบทางสถิติสำหรับการเปรียบเทียบค่าเฉลี่ย

ข้อสมมติ:

1. ตัวแปรตามเป็นข้อมูลต่อเนื่องที่วัดในระดับช่วงหรืออัตราส่วน
2. ตัวแปรอิสระประกอบด้วยสองกลุ่มที่เป็นอิสระต่อ
3. ข้อมูลมีการกระจายตัวแบบปกติในแต่ละกลุ่ม
4. ความเท่าเทียมของความแปรปรวน (homogeneity of variances / homoscedasticity) ตรวจสอบด้วย: Levene's test หรือ F-test
 - หาก Levene's test $p > 0.05$: ความแปรปรวนเท่ากัน (ใช้ standard t-test)

- หาก Levene's test $p \leq 0.05$: ความแปรปรวนไม่เท่ากัน (ใช้ Welch's t-test)
- 5. ไม่มีค่าผิดปกติที่มีนัยสำคัญในแต่ละกลุ่ม (no significant outliers)

สมมติฐาน:

- H_0 : ค่าเฉลี่ยของกลุ่ม 1 ไม่แตกต่างจากค่าเฉลี่ยของกลุ่ม 2
- H_1 : ค่าเฉลี่ยของกลุ่ม 1 แตกต่างจากค่าเฉลี่ยของกลุ่ม 2

ตัวอย่างการใช้ในงานเอ็นโดดอนติกส์:

- เปรียบเทียบแรงยึดติดของซีลเลอร์ชนิด A กับชนิด B
- ทดสอบความแตกต่างของเวลาในการทำงานระหว่างการใช้ไฟล์มือกับไฟล์หมุน

2.3 การทดสอบที แบบจับคู่ (Paired Samples t-test): เปรียบเทียบค่าเฉลี่ยของสองกลุ่มที่เกี่ยวข้อง

การทดสอบที แบบจับคู่ หรือที่เรียกว่า dependent t-test หรือ paired-samples t-test ใช้เพื่อพิจารณาว่าค่าเฉลี่ยของตัวแปรตามคงที่หรือไม่ในสองกลุ่มที่เกี่ยวข้องกัน การทดสอบนี้เหมาะสมอย่างยิ่งสำหรับการออกแบบที่เกี่ยวกับกลุ่มตัวอย่างที่จับคู่กันหรือการวัดซ้ำที่ดำเนินการจากอาสาสมัครคนเดียวกัน [1]

ข้อสมมติ:

1. ตัวแปรตามเป็นข้อมูลต่อเนื่อง ที่วัดในระดับช่วงหรืออัตราส่วน
2. ตัวแปรอิสระประกอบด้วยสองกลุ่มที่เกี่ยวข้องกัน (two related groups / matched pairs)
3. ความแตกต่างระหว่างคู่ (differences) มีการกระจายตัวแบบปกติ
4. การสังเกตแต่ละคู่เป็นอิสระจากคู่อื่น ๆ (independence of pairs)
5. ไม่มีค่าผิดปกติที่มีนัยสำคัญในความแตกต่าง (no significant outliers in differences)

สมมติฐาน:

- H_0 : ค่าเฉลี่ยของความแตกต่างระหว่างสองกลุ่ม เท่ากับศูนย์
- H_1 : ค่าเฉลี่ยของความแตกต่างแตกต่างจากศูนย์

ตัวอย่างการใช้ในงานเอ็นโดดอนติกส์:

- ศึกษาการเปลี่ยนแปลงของปริมาณแบคทีเรียในรากฟันก่อนและหลังการใช้ยาฆ่าเชื้อ
- ประเมินการเปลี่ยนแปลงของความหนาแน่นของกระดูกรอบรอบปลายรากฟันหลังการรักษาในช่วงเวลาต่าง ๆ
- วัดการเปลี่ยนแปลงของสีฟันหลังการฟอกสีภายในรากฟัน

จุดเด่นของการทดสอบนี้คือความสามารถในการจัดความแปรปรวนระหว่างอาสาสมัครโดยการวิเคราะห์ผลรวมของความแตกต่างภายในคู่ ทำให้เพิ่มอำนาจทางสถิติในการตรวจจับการเปลี่ยนแปลงภายในอาสาสมัคร

2.4 การวิเคราะห์ความแปรปรวนทางเดียว (One-Way ANOVA): เปรียบเทียบค่าเฉลี่ยของสามกลุ่มอิสระขึ้นไป

การวิเคราะห์ความแปรปรวนทางเดียว (One-Way ANOVA) เป็นการทดสอบแบบพาราเมตริกที่มีประสิทธิภาพซึ่งออกแบบมาเพื่อพิจารณาว่ามีความแตกต่างอย่างมีนัยสำคัญทางสถิติระหว่างค่าเฉลี่ยของสามกลุ่มอิสระ (ที่ไม่เกี่ยวข้อง) ขึ้นไปหรือไม่ [7]

ข้อสมมติ:

1. ตัวแปรตามเป็นข้อมูลต่อเนื่องที่วัดในระดับช่วงหรืออัตราส่วน
2. ตัวแปรอิสระประกอบด้วยสามกลุ่มอิสระขึ้นไป
3. การสังเกตเป็นอิสระทั้งภายในกลุ่มและระหว่าง
4. ข้อมูลมีการกระจายตัวแบบปกติในแต่ละกลุ่ม
5. ความเท่าเทียมของความแปรปรวนในทุกกลุ่ม ตรวจสอบด้วย Levene's test หรือ Bartlett's test
 - หาก $p > 0.05$: ความแปรปรวนเท่ากัน
 - หาก $p < 0.05$: ความแปรปรวนไม่เท่ากัน → ใช้ Welch's ANOVA [7]
6. ไม่มีค่าผิดปกติที่มีนัยสำคัญในแต่ละกลุ่ม (no significant outliers)

การทดสอบหลังการวิเคราะห์ (Post-Hoc Tests) หรือการเปรียบเทียบหลายกลุ่ม (Multiple comparisons) สำหรับ One-Way ANOVA

หาก One-Way ANOVA แสดงว่ามีความแตกต่างอย่างมีนัยสำคัญ ($p < 0.05$) จำเป็นต้องทำการทดสอบ post-hoc เพื่อระบุว่ากลุ่มใดแตกต่างจากกัน การเลือก post-hoc test ขึ้นอยู่กับหลายปัจจัย ได้แก่ จำนวนการเปรียบเทียบที่ต้องการ การวางแผนการเปรียบเทียบล่วงหน้า (planned) หรือไม่ (unplanned) และความเข้มงวดที่ต้องการในการควบคุมข้อผิดพลาดประเภทที่ 1 การทดสอบแต่ละชนิดมีข้อดีข้อเสียและความเหมาะสมในการใช้งานที่ต่างกัน [1, 7, 8]

เมื่อทำการเปรียบเทียบหลายคู่พร้อมกัน ความเสี่ยงที่จะเกิดข้อผิดพลาดประเภทที่ 1 จะเพิ่มขึ้นตามจำนวนการเปรียบเทียบ ตัวอย่างเช่น หากกำหนดระดับนัยสำคัญ $\alpha = 0.05$ สำหรับการเปรียบเทียบแต่ละครั้ง หมายความว่ามีโอกาส 5% ที่จะสรุปผิดพลาดว่ามีความแตกต่างเมื่อความจริงแล้วไม่มี เมื่อทำการเปรียบเทียบหลายคู่ โอกาสที่จะเกิดข้อผิดพลาดอย่างน้อยหนึ่งครั้งจะสะสมเพิ่มขึ้น โดยคำนวณได้จากสูตร $1 - (1 - \alpha)^k$ เมื่อ k คือจำนวนการเปรียบเทียบ ตัวอย่างเช่น การศึกษาที่เปรียบเทียบซีลเลอร์ 4 ชนิด จะมีการเปรียบเทียบทั้งหมด 6 คู่ (คำนวณจาก $[4 \times 3] / 2 = 6$) หากใช้ $\alpha = 0.05$ สำหรับแต่ละคู่โดยไม่มีการปรับค่า โอกาสที่จะเกิดข้อผิดพลาดประเภทที่ 1 อย่างน้อยหนึ่งครั้งจะเพิ่มขึ้นเป็น $1 - (1 - 0.05)^6 = 0.265$ หรือประมาณ 26.5% ซึ่งสูงกว่าระดับที่ยอมรับได้มาก “Familywise error rate” คืออัตราความผิดพลาดโดยรวมของการเปรียบเทียบทั้งหมดที่ทำในการวิเคราะห์เดียวกัน การควบคุม familywise error rate ให้อยู่ในระดับที่กำหนด (เช่น 0.05) จึงเป็นหลักการสำคัญของการทดสอบหลังการวิเคราะห์ ซึ่งแต่ละวิธีมีกลไกการปรับค่าที่แตกต่างกันเพื่อรักษาความน่าเชื่อถือของผลการวิจัยโดยรวม

ในบริบททางสถิติ คำว่า “อนุรักษ์นิยม” (conservative) หมายถึงการทดสอบที่มีความระมัดระวังสูงในการสรุปว่ามีความแตกต่างอย่างมีนัยสำคัญ กล่าวคือ การทดสอบแบบอนุรักษ์นิยมจะควบคุมข้อผิดพลาดประเภทที่ 1 (false positive) อย่างเข้มงวด ทำให้มีโอกาสน้อยที่จะสรุปผิดพลาดว่ามีความแตกต่างเมื่อความจริงแล้วไม่มี อย่างไรก็ตาม ความอนุรักษ์นิยมมีข้อแลกเปลี่ยน (trade-off) ที่สำคัญ การทดสอบที่อนุรักษ์นิยมเกินไปจะมีอำนาจทางสถิติต่ำ ส่งผลให้โอกาสสูงที่จะไม่พบความแตกต่างแม้ว่าความแตกต่างนั้นมีอยู่จริง ซึ่งเป็นข้อผิดพลาดประเภทที่ 2 (false negative) ตัวอย่างเช่น Bonferroni correction เป็นวิธีที่อนุรักษ์นิยมมาก เนื่องจากใช้การหารระดับนัยสำคัญด้วยจำนวนการเปรียบเทียบทั้งหมด เมื่อมีการเปรียบเทียบจำนวนมาก ค่า α ที่ปรับแล้วจะเล็กมากจนยากที่จะพบความแตกต่างที่มีนัยสำคัญ ดังนั้น ในกรณีที่มีการเปรียบเทียบหลายคู่ การใช้วิธีอื่นที่มีความสมดุลระหว่างการควบคุมข้อผิดพลาดประเภทที่ 1 และการรักษาอำนาจทางสถิติ เช่น Tukey’s HSD อาจเป็นทางเลือกที่เหมาะสมกว่า

Tukey’s Honestly Significant Difference (Tukey’s HSD) เป็นการทดสอบที่นิยมใช้มากที่สุดสำหรับการเปรียบเทียบทุกคู่ (all pairwise comparisons) การทดสอบนี้ควบคุม familywise error rate ได้ดีและมีอำนาจทางสถิติปานกลาง โดยเหมาะสมมากเมื่อขนาดกลุ่มตัวอย่างของแต่ละกลุ่มเท่ากัน ข้อดีของ Tukey’s HSD คือสมดุลระหว่างการควบคุมข้อผิดพลาดประเภทที่ 1 และการรักษาอำนาจทางสถิติ มีอำนาจการทดสอบมากกว่า Bonferroni correction เมื่อมีการเปรียบเทียบหลายคู่ การทดสอบนี้เหมาะสมสำหรับสถานการณ์ที่ต้องการเปรียบเทียบทุกกลุ่มกับกลุ่มอื่น ๆ และไม่มีแผนการเปรียบเทียบล่วงหน้าว่าต้องการเปรียบเทียบคู่ใดโดยเฉพาะ

Bonferroni correction เป็นวิธีการที่เข้าใจและใช้งานง่าย เหมาะสำหรับการเปรียบเทียบที่วางแผนไว้ล่วงหน้าจำนวนน้อย วิธีการทำงานคือการแบ่งระดับนัยสำคัญ ($\alpha = 0.05$) ด้วยจำนวนการเปรียบเทียบทั้งหมด ตัวอย่างเช่น หากมีการเปรียบเทียบ 5 คู่ ค่า α ใหม่จะเท่ากับ $0.05/5 = 0.01$ ข้อดีหลักของวิธีนี้คือควบคุมข้อผิดพลาดประเภทที่ 1 ได้ดีมาก และเป็นวิธีที่อนุรักษ์นิยมสูง อย่างไรก็ตาม ข้อเสียคือเมื่อมีการเปรียบเทียบจำนวนมาก (มากกว่า 5 คู่) วิธีนี้จะอนุรักษ์เกินไปและเพิ่มโอกาสของข้อผิดพลาดประเภทที่ 2 ทำให้อาจไม่ตรวจพบความแตกต่างที่แท้จริงได้ ดังนั้น Bonferroni จึงเหมาะสมเมื่อมีการเปรียบเทียบน้อย และต้องการควบคุมข้อผิดพลาดประเภทที่ 1 อย่างเข้มงวด

Dunnett’s test ถูกออกแบบมาโดยเฉพาะสำหรับการเปรียบเทียบกลุ่มทดลองหลายกลุ่มกับกลุ่มควบคุมกลุ่มเดียว ซึ่งเป็นสถานการณ์ที่พบบ่อยในงานวิจัยทางการแพทย์และทันตกรรม การทดสอบนี้มีอำนาจทางสถิติสูงกว่า Tukey’s HSD เนื่องจากไม่เปรียบเทียบกลุ่มทดลองระหว่างกัน จึงลดจำนวนการเปรียบเทียบทั้งหมด ตัวอย่างเช่น การศึกษาที่มี 3 กลุ่มยาทดลองและ 1 กลุ่ม placebo Dunnett’s test จะทำการเปรียบเทียบเพียง 3 ครั้ง (แต่ละยา vs placebo) แทนที่จะเป็น 6 ครั้งเหมือน Tukey’s HSD การทดสอบนี้เหมาะสมเมื่อคำถามการวิจัยมุ่งเน้นที่ว่า “กลุ่มทดลองแต่ละกลุ่มแตกต่างจากกลุ่มควบคุมหรือไม่” มากกว่า “กลุ่มทดลองแตกต่างกันเองหรือไม่”

Fisher's Least Significant Difference (LSD) เป็นการทดสอบที่มีอำนาจทางสถิติสูงที่สุด แต่ไม่ควบคุม familywise error rate ซึ่งหมายความว่าเมื่อมีการเปรียบเทียบจำนวนมาก โอกาสข้อผิดพลาดประเภทที่ 1 จะเพิ่มขึ้นอย่างมาก การทดสอบนี้เหมาะสมเฉพาะเมื่อมีการวางแผนการเปรียบเทียบไว้ล่วงหน้าเพียง 2-3 คู่เท่านั้น และไม่แนะนำให้ใช้สำหรับการเปรียบเทียบหลายคู่หรือการเปรียบเทียบที่ไม่ได้วางแผนไว้ในงานวิจัยสมัยใหม่ การใช้ Fisher's LSD ได้รับการวิพากษ์วิจารณ์และไม่ค่อยได้รับการยอมรับเว้นแต่จะมีเหตุผลที่มากและได้วางแผนการเปรียบเทียบไว้อย่างชัดเจนตั้งแต่ก่อนเก็บข้อมูลในทางปฏิบัติ

การเลือก post-hoc test ควรพิจารณาจากวัตถุประสงค์ของการวิจัย หากต้องการเปรียบเทียบทุกกลุ่มและขนาดตัวอย่างค่อนข้างเท่ากัน Tukey's HSD เป็นทางเลือกแรก หากมีกลุ่มควบคุมที่ชัดเจน Dunnett's test จะให้อำนาจทางสถิติที่ดีกว่า หากมีการเปรียบเทียบจำนวนน้อยที่วางแผนไว้และต้องการควบคุมข้อผิดพลาดประเภทที่ 1 อย่างเข้มงวด Bonferroni เป็นทางเลือกที่เหมาะสม

การทดสอบหลังการวิเคราะห์สำหรับ Welch's ANOVA

เมื่อใช้ Welch's ANOVA เนื่องจากความแปรปรวนของข้อมูลในแต่ละกลุ่มไม่เท่ากัน การใช้ post-hoc tests มาตรฐานเช่น Tukey's HSD จะไม่เหมาะสมเพราะ tests เหล่านั้นสันนิษฐานว่าความแปรปรวนเท่ากัน ในกรณีนี้ Games-Howell test เป็นการทดสอบที่ออกแบบมาโดยเฉพาะสำหรับสถานการณ์ที่ความแปรปรวนไม่เท่ากัน และปรับค่า degrees of freedom ให้เหมาะสมในแต่ละการเปรียบเทียบคู่ การทดสอบนี้มีข้อดีหลายประการ คือทนทานต่อการละเมิดข้อสมมติความเท่าเทียมของความแปรปรวน ควบคุมข้อผิดพลาดประเภทที่ 1 ได้ดีแม้ขนาดกลุ่มตัวอย่างไม่เท่ากัน และมีอำนาจทางสถิติสูงกว่า Bonferroni correction เมื่อความแปรปรวนไม่เท่ากัน

2.5 การวิเคราะห์ความแปรปรวนทางเดียวแบบวัดซ้ำ (One-Way Repeated Measures ANOVA): เปรียบเทียบค่าเฉลี่ยของสามจุดเวลาขึ้นไปในกลุ่มเดียวกัน

One-Way Repeated Measures ANOVA เป็นการขยายของ Paired Samples t-test ที่ใช้สำหรับเปรียบเทียบค่าเฉลี่ย

ของตัวแปรตามที่วัดในสามจุดเวลาหรือมากกว่าในกลุ่มตัวอย่างเดียวกัน หรือในสภาวะที่แตกต่างกันในผู้เข้าร่วมคนเดียวกัน เหมาะสำหรับการศึกษาแบบ longitudinal หรือการศึกษาที่ต้องการติดตามการเปลี่ยนแปลงเมื่อเวลาผ่านไป ความสามารถในการควบคุมความแปรปรวนระหว่างบุคคล (between-subject variability) โดยแต่ละคนทำหน้าที่เป็น control ของตนเอง ทำให้มีอำนาจทางสถิติสูงกว่า One-Way ANOVA แบบปกติเมื่อใช้ขนาดตัวอย่างเท่ากัน [1]

ข้อสมมติ:

1. ตัวแปรตามเป็นข้อมูลต่อเนื่องที่วัดในระดับช่วงหรืออัตราส่วน
2. ตัวแปรอิสระประกอบด้วยสามจุดเวลาหรือมากกว่าที่เกี่ยวข้องกัน (related time points)
3. ไม่มีค่าผิดปกติที่มีนัยสำคัญ
4. ข้อมูลมีการกระจายตัวแบบปกติสำหรับแต่ละจุดเวลา
5. ความเป็น Sphericity: ความแปรปรวนของความแตกต่างระหว่างระดับทั้งหมดของตัวแปรอิสระควรเท่ากัน ตรวจสอบด้วย Mauchly's Test of Sphericity หากปฏิเสธสมมติฐาน sphericity ต้องปรับแก้โดยใช้ Greenhouse-Geisser correction เมื่อ $\epsilon < 0.75$ หรือ Huynh-Feldt correction เมื่อ $\epsilon > 0.75$ มีความเข้มงวดน้อยกว่า

สมมติฐาน:

- H_0 : ค่าเฉลี่ยของตัวแปรตามเท่ากันในทุกจุดเวลา ($\mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$)
- H_1 : ค่าเฉลี่ยของตัวแปรตามแตกต่างกันอย่างน้อยหนึ่งจุดเวลา

ตัวอย่างการใช้ในงานเอ็นโดดอนติกส์:

- ประเมินความหนาแน่นของกระดูกรอบรอบปลายรากฟันในผู้ป่วยคนเดียวกันที่ baseline, 3 เดือน, 6 เดือน และ 12 เดือนหลังการรักษา
- วัดการเปลี่ยนแปลงของ pH ในรากฟันหลังการใช้ calcium hydroxide ในช่วงเวลาต่าง ๆ (1 สัปดาห์, 2 สัปดาห์ และ 4 สัปดาห์)

การทดสอบหลังการวิเคราะห์สำหรับ One-Way Repeated Measures ANOVA

หาก One-Way Repeated Measures ANOVA แสดงว่ามีความแตกต่างอย่างมีนัยสำคัญ ($p < 0.05$) จำเป็นต้องทำ

post-hoc test เพื่อระบุว่าจุดเวลาหรือสภาวะใดแตกต่างจากกัน การเลือก post-hoc test สำหรับ Repeated Measures ANOVA มีความแตกต่างจาก One-Way ANOVA แบบปกติ เนื่องจากข้อมูลมีความสัมพันธ์กันภายในผู้เข้าร่วมคนเดียวกัน ทำให้ไม่สามารถใช้ Tukey's HSD แบบมาตรฐานได้ แต่ต้องใช้วิธีการที่ปรับให้เหมาะสมกับโครงสร้างของข้อมูลแบบวัดซ้ำ

Bonferroni correction เป็นวิธีที่นิยมใช้มากที่สุดสำหรับ post-hoc test หลัง Repeated Measures ANOVA และมักเป็นค่าเริ่มต้น (default) ในโปรแกรมสถิติส่วนใหญ่ Holm-Bonferroni จึงเป็นทางเลือกที่แนะนำมากกว่า Bonferroni แบบธรรมดาในกรณีส่วนใหญ่

Fisher's Least Significant Difference (LSD) บางครั้งยังคงถูกใช้สำหรับ post-hoc test หลัง Repeated Measures ANOVA แต่ไม่แนะนำอย่างยิ่ง เนื่องจากไม่มีการปรับค่า p เลย และไม่ควบคุม familywise error rate การใช้ LSD จะทำให้โอกาสของข้อผิดพลาดประเภทที่ 1 เพิ่มขึ้นอย่างมากเมื่อมีการเปรียบเทียบหลายคู่ ควรใช้เฉพาะเมื่อมีการวางแผนการเปรียบเทียบไว้ล่วงหน้าเพียง 2-3 คู่เท่านั้น และต้องมีเหตุผลทางทฤษฎีที่ชัดเจน

3. การทดสอบแบบนอนพาราเมตริกสำหรับการเปรียบเทียบค่าเฉลี่ย (Non-parametric tests for comparing means)

3.1 นิยามและเมื่อใดควรใช้

การทดสอบแบบนอนพาราเมตริกเป็นกลุ่มของวิธีการทางสถิติที่ไม่จำเป็นต้องมีข้อสมมติว่าข้อมูลสอดคล้องกับการกระจายตัวของความน่าจะเป็นที่เฉพาะเจาะจง เช่น การกระจายตัวแบบปกติ ด้วยเหตุนี้ จึงมักเรียกว่า “การทดสอบที่ปราศจากการกระจาย” (distribution-free tests) [1, 4, 9]

สถานการณ์ที่ควรใช้การทดสอบแบบนอนพาราเมตริก:

1. **ข้อมูลที่ไม่มีการกระจายตัวแบบปกติ:** เมื่อข้อมูลเบี่ยงเบนมีนัยสำคัญหรือเบี่ยงเบนจากการกระจายตัวแบบปกติในลักษณะอื่น

2. **ขนาดกลุ่มตัวอย่างเล็ก:** ในการศึกษาที่มีการสังเกตจำกัด อาจเป็นเรื่องท้าทายในการสันนิษฐานหรือทดสอบ

ความเป็นปกติอย่างน่าเชื่อถือ ทำให้การทดสอบแบบนอนพาราเมตริกเป็นทางเลือกที่เหมาะสมกว่า

3. **มีค่าผิดปกติ:** การทดสอบแบบนอนพาราเมตริกโดยทั่วไปมีความอ่อนไหวน้อยต่อค่าสุดขั้ว (outliers) ในข้อมูล ซึ่งสามารถส่งผลกระทบต่อทดสอบแบบพาราเมตริกอย่างไม่สมส่วน

4. **ข้อมูลระดับลำดับหรือนามบัญญัติ:** เมื่อตัวแปรตามถูกวัดในระดับลำดับ (ประเภทที่มีลำดับ) หรือระดับนามบัญญัติ (ประเภทที่ไม่มีลำดับโดยธรรมชาติ) การกำหนดคะแนนตัวเลขสำหรับการวิเคราะห์แบบพาราเมตริกอาจไม่เหมาะสมหรือทำให้เข้าใจผิด

ข้อจำกัดของการทดสอบแบบนอนพาราเมตริก:

- **อำนาจน้อยกว่า:** ข้อเสียที่สำคัญคือการทดสอบแบบนอนพาราเมตริกอาจมีอำนาจทางสถิติต่ำกว่าเมื่อเทียบกับคู่มือแบบพาราเมตริก โดยเฉพาะเมื่อข้อสมมติสำหรับวิธีการแบบพาราเมตริกที่สอดคล้องกันเป็นจริง ซึ่งแปลว่ามีโอกาสลดลงในการตรวจจับผลกระทบที่แท้จริงหากมีอยู่จริงในประชากร
- **ใช้ข้อมูลน้อยกว่า:** โดยการแปลงข้อมูลดิบเป็นอันดับ การทดสอบแบบนอนพาราเมตริกจะละทิ้งข้อมูลรายละเอียดบางส่วนที่มีอยู่ในข้อมูลต้นฉบับ ซึ่งอาจนำไปสู่ผลลัพธ์ที่มีนัยสำคัญทางสถิติที่น้อยลงแม้ว่าจะมีผลกระทบที่แท้จริงอยู่

3.2 การทดสอบ Mann-Whitney U หรือ Wilcoxon Rank-Sum Test (ทางเลือกแทน Independent Samples t-test)

การทดสอบ Mann-Whitney U เป็นการทดสอบทางสถิติแบบนอนพาราเมตริกที่ใช้เพื่อประเมินความแตกต่างระหว่างสองกลุ่มอิสระ โดยเฉพาะเมื่อตัวแปรตามถูกวัดในระดับลำดับหรือเป็นข้อมูลต่อเนื่องแต่ไม่มีการกระจายตัวแบบปกติ

ข้อสมมติ

- ตัวแปรตามวัดในระดับลำดับหรือต่อเนื่อง
- ตัวแปรอิสระประกอบด้วยสองกลุ่มที่เป็นอิสระต่อกัน
- การสังเกตเป็นอิสระ

สมมติฐาน:

- H_0 : การกระจายตัวของสองกลุ่มเหมือนกัน
- H_1 : การกระจายตัวของสองกลุ่มแตกต่างกัน

ตัวอย่างการใช้ในงานเอ็นโดดอนติกส์:

- เปรียบเทียบคะแนนความเจ็บปวดระหว่างสองกลุ่มเมื่อข้อมูลไม่มีการกระจายตัวแบบปกติ
- ศึกษาความแตกต่างของระดับความพึงพอใจของผู้ป่วยระหว่างเทคนิคการรักษา 2 วิธี
- เปรียบเทียบลำดับของความยากง่ายในการใช้งานระหว่างไฟล์ชนิดต่าง ๆ ที่ประเมินโดยทันตแพทย์
- ทดสอบความแตกต่างของจำนวนโคไลน์แบคทีเรียที่มีค่าผิดปกติสูง

3.3 การทดสอบ Wilcoxon Signed-Rank (ทางเลือกแทน Paired Samples t-test)

การทดสอบ Wilcoxon Signed-Rank Test เป็นวิธีการทางสถิติแบบนอนพาราเมตริกที่ออกแบบมาสำหรับการเปรียบเทียบสองกลุ่มตัวอย่างที่เกี่ยวข้องกัน เช่น คู่ที่จับคู่กันหรือการวัดซ้ำจากกลุ่มตัวอย่างเดียว การทดสอบนี้มักใช้เป็นทางเลือกแทนการทดสอบที แบบจับคู่เมื่อไม่สามารถสันนิษฐานว่าข้อมูลมีการกระจายตัวแบบปกติได้ การทดสอบทำงานโดยการประเมินความแตกต่างระหว่างคู่ของการสังเกต จัดอันดับขนาดสัมบูรณ์ของความแตกต่างเหล่านี้ จากนั้นพิจารณาเครื่องหมาย (บวกหรือลบ) ของแต่ละความแตกต่าง

ข้อสมมติ

- ตัวแปรตามเป็นข้อมูลต่อเนื่องหรือลำดับ
- กลุ่มตัวอย่างเป็นแบบจับคู่หรือเกี่ยวข้องกัน
- ความแตกต่างมีการกระจายตัวแบบสมมาตรรอบศูนย์

สมมติฐาน:

- H_0 : ค่ามัธยฐานของความแตกต่างระหว่างการสังเกตแบบจับคู่เท่ากับศูนย์
- H_1 : ค่ามัธยฐานของความแตกต่างแตกต่างจากศูนย์อย่างมีนัยสำคัญ

ตัวอย่างการใช้ในงานเอ็นโดดอนติกส์:

- เปรียบเทียบคะแนนความเจ็บปวดก่อนและหลังการรักษาเมื่อข้อมูลเบ้
- ศึกษาการเปลี่ยนแปลงของระดับความวิตกกังวลของผู้ป่วยก่อนและหลังได้รับข้อมูลเกี่ยวกับการรักษา
- ประเมินการเปลี่ยนแปลงของคุณภาพชีวิตที่วัดด้วยแบบสอบถามก่อนและหลังการรักษารากฟัน

- ทดสอบความแตกต่างของคะแนนการรับรู้การเจ็บปวดในช่วงเวลาต่างๆ ในผู้ป่วยคนเดียว

3.4 การทดสอบ Kruskal-Wallis H (ทางเลือกแทน One-Way ANOVA)

การทดสอบ Kruskal-Wallis H เป็นการทดสอบแบบนอนพาราเมตริกที่อิงจากอันดับซึ่งใช้เพื่อพิจารณาว่ามีความแตกต่างอย่างมีนัยสำคัญทางสถิติระหว่างสองกลุ่มหรือมากกว่าของตัวแปรอิสระ

ลักษณะสำคัญ:

- คล้ายกับ ANOVA คือ การทดสอบ Kruskal-Wallis H ให้ความหมายว่ามีความน้อยสองกลุ่มที่แตกต่างกันอย่างมีนัยสำคัญหรือไม่ แต่ไม่ระบุว่ากลุ่มเฉพาะใดแตกต่างจากกัน
- ดังนั้น หากการทดสอบ Kruskal-Wallis H ให้ผลลัพธ์ที่มีนัยสำคัญ จำเป็นต้องมีการทดสอบหลังการวิเคราะห์เพิ่มเติม เพื่อทำการเปรียบเทียบแบบคู่เฉพาะและระบุตำแหน่งของความแตกต่าง

ข้อสมมติ:

- สองตัวแปรวัดในระดับลำดับหรือต่อเนื่อง
- ตัวแปรอิสระประกอบด้วยสองกลุ่มอิสระหรือมากกว่า
- การสังเกตเป็นอิสระ

สมมติฐาน

- H_0 : การกระจายตัวของทุกกลุ่มเหมือนกัน
- H_1 : การกระจายตัวของอย่างน้อยหนึ่งกลุ่มแตกต่างจากกลุ่มอื่น

ตัวอย่างการใช้ในงานเอ็นโดดอนติกส์

- เปรียบเทียบคะแนนความเจ็บปวดระหว่างกลุ่มยา 4 ชนิดเมื่อข้อมูลไม่ปกติ
- ศึกษาความพึงพอใจของผู้ป่วยต่อเทคนิคชลประทานหลายวิธี
- ประเมินระดับความวิตกกังวลระหว่างกลุ่มที่ได้รับข้อมูลในรูปแบบต่าง ๆ

การทดสอบหลังการวิเคราะห์สำหรับ Kruskal-Wallis H Test

หาก Kruskal-Wallis H test แสดงว่ามีความแตกต่างอย่างมีนัยสำคัญ ($p < 0.05$) จำเป็นต้องทำ post-hoc test

เพื่อระบุว่าคุณได้แตกต่างจากกัน การเลือก post-hoc test สำหรับ Kruskal-Wallis มีหลายทางเลือก โดยแต่ละวิธีมีข้อดีข้อเสียและความเหมาะสมแตกต่างกัน

Dunn's test เป็น post-hoc test มาตรฐานที่ยอมรับกันทั่วไปสำหรับ Kruskal-Wallis H test วิธีการทำงานคือการเปรียบเทียบอันดับเฉลี่ย (mean rank) ระหว่างกลุ่มแต่ละคู่ โดยใช้ค่าความแปรปรวนรวม (pooled variance) จากข้อมูลทุกกลุ่มในการคำนวณ พร้อมปรับค่า p เพื่อควบคุม familywise error rate ด้วย Bonferroni adjustment ซึ่งมีความอนุรักษ์นิยมน้อยกว่า การทดสอบนี้ทำงานได้ดีแม้ขนาดกลุ่มตัวอย่างไม่เท่ากัน

Mann-Whitney U test แบบ pairwise พร้อมปรับค่า p เป็นอีกทางเลือกหนึ่งที่สามารถใช้ได้ วิธีการคือทำ Mann-Whitney U test สำหรับทุกคู่ พร้อมปรับค่า p ด้วย Bonferroni correction ข้อดีของวิธีนี้คือเข้าใจง่ายและสามารถทำได้ด้วยโปรแกรมสถิติทั่วไป อย่างไรก็ตาม วิธีนี้ไม่ใช้วิธีมาตรฐานและมักมีความอนุรักษ์นิยมมากกว่า Dunn's test เนื่องจากการคำนวณแต่ละคู่เป็นอิสระต่อกัน ไม่ได้ใช้ข้อมูลจากกลุ่มอื่นมาช่วยประมาณค่าความแปรปรวน ทำให้มีอำนาจทางสถิติต่ำกว่า ดังนั้นจึงควรใช้เฉพาะเมื่อไม่มี Dunn's test ในโปรแกรมสถิติที่ใช้

3.5 การทดสอบ Friedman (Friedman Test): ทางเลือกแทน One-Way Repeated Measures ANOVA

การทดสอบ Friedman เป็นการทดสอบทางสถิติแบบนอนพารามेटริก ที่ใช้เพื่อตรวจหาความแตกต่างในการรักษาทั่วทั้งการวัดซ้ำหลายครั้ง ถือเป็นทางเลือกแบบนอนพารามेटริกของ One-Way Repeated Measures ANOVA และเป็นทางเลือกของ Wilcoxon Signed-Rank Test สำหรับมากกว่าสองจุดเวลาหรือสภาวะ

การทดสอบนี้ทำงานโดยการจัดอันดับข้อมูลภายในแต่ละบุคคล (within-subject ranking) จากนั้นวิเคราะห์ว่าอันดับเฉลี่ยแตกต่างกันระหว่างจุดเวลาหรือสภาวะต่าง ๆ หรือไม่ การทดสอบนี้ไม่ต้องการข้อสมมติเรื่องการกระจายตัวแบบปกติหรือความเป็น sphericity ทำให้เหมาะสำหรับข้อมูลที่ละเมิดข้อสมมติของ Repeated Measures ANOVA

ข้อสมมติ

- ตัวแปรตามอย่างน้อยต้องเป็นข้อมูลลำดับหรือข้อมูลต่อเนื่อง
- ตัวแปรอิสระประกอบด้วยสามกลุ่มที่เกี่ยวข้องกันหรือมากกว่า
- การสังเกตแต่ละชุดเป็นอิสระจากชุดอื่น ๆ แต่การวัดภายในบุคคลเดียวกันมีความเกี่ยวข้องกัน

สมมติฐาน:

- H_0 : การกระจายตัวของตัวแปรตามเหมือนกันในทุกจุดเวลาหรือสภาวะ (เท่ากับการทดสอบว่าค่ามัธยฐานเท่ากัน)
- H_1 : การกระจายตัวของตัวแปรตามแตกต่างกันในอย่างน้อยหนึ่งจุดเวลาหรือสภาวะ

ตัวอย่างการใช้ในงานเอ็นโดดอนติกส์:

- ประเมินระดับความเจ็บปวดของผู้ป่วยคนเดียวกันที่ 24 ชั่วโมง 48 ชั่วโมง และ 72 ชั่วโมง หลังการรักษา เมื่อข้อมูลไม่มีการกระจายตัวแบบปกติ
- เปรียบเทียบคะแนนความพึงพอใจของผู้ป่วยต่อขั้นตอนการรักษา 4 ขั้นตอน (การวินิจฉัย การเตรียมคลองรากฟัน การอุดคลองรากฟัน และการติดตามผล)
- วัดการเปลี่ยนแปลงของระดับความวิตกกังวล (anxiety score) ก่อนการรักษา, ระหว่างการรักษา และหลังการรักษาในผู้ป่วยคนเดียวกัน
- ศึกษาการเปลี่ยนแปลงของคะแนนคุณภาพชีวิตที่ baseline 1 สัปดาห์ 1 เดือน และ 3 เดือน หลังการรักษารากฟัน เมื่อข้อมูลมีลักษณะเบ้

การทดสอบหลังการวิเคราะห์สำหรับ Friedman Test

หาก Friedman test แสดงว่ามีความแตกต่างอย่างมีนัยสำคัญ ($p < 0.05$) จำเป็นต้องทำ post-hoc test เพื่อระบุว่าจุดเวลาหรือสภาวะใดแตกต่างจากกัน เนื่องจาก Friedman test ใช้สำหรับข้อมูลแบบวัดซ้ำ (repeated measures) ที่ไม่เป็นไปตามข้อสมมติของ parametric tests การเลือก post-hoc test จึงต้องคำนึงถึงลักษณะพิเศษของข้อมูลประเภทนี้

Dunn's test สำหรับ Friedman (บางครั้งเรียกว่า Durbin-Conover test) เป็น post-hoc test มาตรฐานที่ปรับให้เหมาะสมกับข้อมูลแบบวัดซ้ำ วิธีการทำงานคล้ายกับ Dunn's test สำหรับ Kruskal-Wallis แต่มีการปรับปรุงให้เหมาะสมกับโครงสร้างของข้อมูลที่มีความสัมพันธ์กันภายในผู้เข้าร่วมคนเดียว

Wilcoxon Signed-Rank test แบบ pairwise พร้อมปรับค่า p เป็นทางเลือกที่ใช้งานได้ง่าย วิธีการคือทำ Wilcoxon Signed-Rank test สำหรับทุกคู่ของจุดเวลาหรือสถานะ พร้อมปรับค่า p เพื่อควบคุมข้อผิดพลาดประเภทที่ 1 การปรับค่า p แบบ Bonferroni ทำโดยการหารระดับนัยสำคัญ ($\alpha = 0.05$) ด้วยจำนวนการเปรียบเทียบทั้งหมด ตัวอย่างเช่น หากมี 4 จุดเวลา จะมีการเปรียบเทียบทั้งหมด 6 คู่ (คำนวณจาก $4 \times 3 / 2 = 6$) ดังนั้น α ที่ปรับแล้วจะเท่ากับ $0.05 / 6 = 0.0083$ ค่า p ที่น้อยกว่า 0.0083 จึงจะถือว่ามีความสำคัญ วิธีนี้เข้าใจง่ายแต่อนุรักษ์นิยมมาก

4. การเลือกการทดสอบที่เหมาะสม: กรอบการตัดสินใจ (Choosing the Right Test: A Decision Framework)

การเลือกการทดสอบทางสถิติที่เหมาะสมที่สุดสำหรับการเปรียบเทียบค่าเฉลี่ยต้องใช้แนวทางที่เป็นระบบและพิจารณาอย่างรอบคอบ โดยคำนึงถึงปัจจัยสำคัญหลายประการที่เกี่ยวข้องกับคำถามการวิจัยและลักษณะของข้อมูล (รูปที่ 2-3 และตารางที่ 1)

ปัจจัยหลักที่ต้องพิจารณาได้แก่

1. คำถามการวิจัย (Research Question)
2. จำนวนกลุ่ม (Number of Groups)
3. ประเภทข้อมูล (Type of Data)
4. ความเป็นอิสระของการสังเกต (Independence of Observations)
5. การกระจายตัวของข้อมูล (Data Distribution - Normality)
6. ความเท่าเทียมของความแปรปรวน (Homogeneity of Variances)
7. ค่าผิดปกติ (Outliers)

5. สรุปและขอเสนอแนะ (Conclusion and Recommendations)

การเปรียบเทียบค่าเฉลี่ยทางสถิติเป็นเครื่องมือที่ทรงพลังในการวิจัยทางเอ็นโดดอนติกส์ ช่วยให้เราสามารถตัดสินใจบนพื้นฐานของหลักฐานเชิงประจักษ์แทนการพึ่งพาเพียงประสบการณ์หรือความคิดเห็นส่วนตัว อย่างไรก็ตาม เครื่องมือเหล่านี้จะมีประโยชน์ก็ต่อเมื่อใช้อย่างถูกต้องและเหมาะสม

จุดสำคัญที่ควรจดจำ:

1. การเลือกการทดสอบที่เหมาะสมเป็นสิ่งสำคัญ - ขึ้นอยู่กับลักษณะของข้อมูล คำถามการวิจัย และความเป็นไปตามข้อสมมติ
2. การตรวจสอบข้อสมมติ - เป็นความรับผิดชอบทางจริยธรรมที่มีผลต่อความถูกต้องของผลการวิจัยและการดูแลผู้ป่วย
3. การวางแผนล่วงหน้าเป็นสิ่งจำเป็น - การปรึกษาผู้เชี่ยวชาญด้านสถิติตั้งแต่ขั้นตอนการออกแบบการศึกษาช่วยป้องกันปัญหาในภายหลัง
4. การรายงานผลอย่างโปร่งใสและครบถ้วน - ช่วยให้ผู้อ่านสามารถประเมินคุณภาพและความน่าเชื่อถือของผลการวิจัย

สุดท้ายนี้ เป้าหมายของการใช้สถิติในงานวิจัยทางเอ็นโดดอนติกส์คือการสร้างองค์ความรู้ที่เชื่อถือได้และนำไปใช้ประโยชน์ได้จริงในการปรับปรุงคุณภาพการดูแลรักษาผู้ป่วย การเข้าใจและประยุกต์ใช้หลักการเปรียบเทียบค่าเฉลี่ยอย่างถูกต้องจะช่วยให้ทันตแพทย์และนักวิจัยสามารถมีส่วนร่วมอย่างมีความหมายในการพัฒนาสาขาวิทยาเอ็นโดดอนติกส์ต่อไป

เอกสารอ้างอิง

1. Hazra A, Gogtay N. Biostatistics Series Module 3: Comparing Groups: Numerical Variables. *Indian J Dermatol.* 2016;61(3):251-60.
2. Kim TK. T test as a parametric statistic. *Korean J Anesthesiol.* 2015;68(6):540-6.
3. Sainani KL. Dealing With Non-normal Data. *PM&R.* 2012;4(12):1001-5.
4. Field A. Discovering Statistics Using IBM SPSS Statistics: SAGE Publications; 2018.

5. Herberich E, Sikorski J, Hothorn T. A Robust Procedure for Comparing Multiple Means under Heteroscedasticity in Unbalanced Designs. **PLoS ONE**. 2010;5.
6. Saculinggan M, Balase EA. Empirical Power Comparison Of Goodness of Fit Tests for Normality In The Presence of Outliers. **J Phys Conf Ser**. 2013;435:012041.
7. Kim HY. Analysis of variance (ANOVA) comparing means of more than two groups. **Restor Dent Endod**. 2014;39(1):74-7.
8. Kim H-Y. Statistical notes for clinical researchers: post-hoc multiple comparisons. **Restor Dent Endod**. 2015;40(2):172-6.
9. Schober P, Vetter TR. Nonparametric Statistical Methods in Medical Research. **Anesthesia & Analgesia**. 2020;131(6):1862-3.