

การวิเคราะห์ข้อมูลเชิงสถิติจากข้อมูลงานวิจัยที่ตีพิมพ์

กันตพงศ์ พุ่มอยู่^{1*}, วีรยุทธ เจริญเรืองกิจ²

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเป็นแนวทางในการส่งเสริมอาจารย์ในการทำงานวิจัย ด้วยการศึกษาค้นคว้าข้อมูลเกี่ยวกับการอ้างอิงผลงานวิจัยของอาจารย์และบุคลากรในมหาวิทยาลัยศรีนครินทรวิโรฒ จากชุดข้อมูลการอ้างอิงผลงานวิจัยในช่วง ค.ศ. 2011 ถึง 2020 ที่ถูกเผยแพร่ไว้บนฐานข้อมูล Scopus โดยผู้วิจัยได้แบ่งกลุ่มข้อมูลตามกลุ่มสาขาวิชาเป็น 3 กลุ่ม ได้แก่ กลุ่มสาขาวิชาวิทยาศาสตร์และเทคโนโลยี (Science Technology), กลุ่มสาขาวิชาวิทยาศาสตร์สุขภาพ (Health Science), กลุ่มสาขาวิชามนุษยศาสตร์และสังคมศาสตร์ (Humanities) และแบ่งกลุ่มตามสถานที่ตั้งตามกลุ่มสาขาวิชาเป็น 2 กลุ่ม ได้แก่ ประสานมิตรและองครักษ์ ผู้วิจัยสนใจที่จะศึกษาวิเคราะห์ลักษณะการแจกแจงของข้อมูลจากการทำการทดสอบการกระจายแบบปกติ (Normality Tests) ด้วยวิธีการทดสอบ Anderson-Darling Normality Test และ Lilliefors Normality Test โดยผลลัพธ์จากการทดสอบการกระจายแบบปกติ (Normality Tests) ทำให้ทราบว่าลักษณะการกระจายตัวของข้อมูลกลุ่มสาขาวิชาและกลุ่มสถานที่ตั้งของสาขาวิชานั้นมีรูปแบบการกระจายตัวที่ไม่ปกติ (Non-normal distribution) ผู้วิจัยจึงดำเนินการทดสอบทางสถิติ (Statistical Tests) ด้วยวิธีการทดสอบสถิติไม่อิงพารามิเตอร์ (Nonparametric Statistical Significance Tests) ซึ่งประกอบด้วยวิธีการทดสอบแบบ Mann-Whitney U Test และ Kruskal-Wallis H Test จากการทดสอบทำให้ทราบว่าระหว่างกลุ่มข้อมูลที่แบ่งตามสาขาวิชา 2 กลุ่มเพื่อทดสอบความต่างอย่างมีนัยสำคัญทางสถิติ โดยผู้วิจัยได้นำผลวิเคราะห์จากการทดสอบทางสถิติเพื่อนำเสนอเป็นแนวทางในการส่งเสริมการทำงานวิจัยต่อไป

คำสำคัญ : อ้างอิงผลงานวิจัย, วิเคราะห์ข้อมูลเชิงสถิติ, ทดสอบการกระจายแบบปกติ, ทดสอบสถิติไม่อิงพารามิเตอร์

¹ หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพฯ 10110

² คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพฯ 10110

* Corresponding author: Tel.: 090-9155222 E-mail address: kantapong.phu@g.swu.ac.th

Statistical data analysis from research publications

Kantapong Phumyoo^{1*}, Werayuth Charoenruengkit²

Abstract

The purpose of this research is to be a guideline for promoting teachers in research work by studying information about research citations of professors and personnel in Srinakharinwirot University based on research citation data from 2011 to 2020 published on the Scopus database. The researcher has divided the data according to the field of study from department group into 3 groups as follows: Science Technology, Health Science, Humanities and divided the data from location of department into 2 groups as follows: Prasarnmit and Onkharak. The researcher is interested in analyzing the distribution characteristics of the data from the normality tests using the Anderson-Darling Normality Test and Lilliefors Normality Test. The results from the normality test revealed that the distribution characteristics of the department group data and the location group have a non-normal distribution. Therefore, the researcher conducted statistical tests with non-parametric statistical testing methods, which consisted of Mann-Whitney U Test and Kruskal-Wallis H Test, which revealed that there were statistically significant differences between the data groups divided by two disciplines: the first group Humanities, Health Science and the second group Humanities, Science Technology. The researcher has used the analysis results from the statistical testing to present as a guideline to promote research work further.

Keywords : Citation, Statistical data analysis, Normality Tests, Nonparametric Statistical Significance Tests

¹ Data Science, Faculty of Science, Srinakharinwirot University, Bangkok, 10110, Thailand

² Faculty of Science, Srinakharinwirot University, Bangkok, 10110, Thailand

* Corresponding author: Tel.: 090-9155222 E-mail address: kantapong.phu@g.swu.ac.th

บทนำ

ปัจจุบันการทำงานวิจัยเป็นเครื่องมือหรือวิธีที่ดีที่สุดในการแสวงหาความรู้เพื่อแก้ไขปัญหาต่างๆ และต้องการแสวงหาคำตอบซึ่งส่งผลต่อความเจริญก้าวหน้าพัฒนาองค์ความรู้ในทุกสาขาวิชา ฐานข้อมูล Scopus เป็นฐานข้อมูลของบริษัท Elsevier ซึ่งเป็นหนึ่งในฐานข้อมูลที่ได้รับการยอมรับในระดับนานาชาติที่ครอบคลุมเนื้อหาทุกสาขาวิชา และเป็นหนึ่งในฐานข้อมูลที่บุคลากรและอาจารย์ภายในมหาวิทยาลัยศรีนครินทรวิโรฒเผยแพร่ผลงานวิจัย และสิ่งที่จะช่วยบ่งบอกว่างานวิจัยนั้นมีคุณภาพหรือไม่ คือ การอ้างอิง (Citation) ซึ่งเป็นหนึ่งในตัวชี้วัดที่สำคัญที่ถูกนำไปใช้ในการวัดคุณภาพของงานวิจัย ในกรณีที่งานวิจัยถูกอ้างอิงผลงานเป็นจำนวนมาก เปรียบเสมือนการบ่งบอกว่างานวิจัยนั้นเป็นที่ยอมรับและมีคุณค่ามากในสายต่านักวิจัย

เพื่อตรวจสอบความเชื่อมโยงและศึกษาข้อมูลข้อมูลเกี่ยวกับการอ้างอิงผลงานวิจัยของบุคลากรและอาจารย์ในมหาวิทยาลัยศรีนครินทรวิโรฒ โดยได้แบ่งแยกกลุ่มผลงานวิจัยตามกลุ่มสาขาวิชาตามคณะเป็น 3 กลุ่ม ได้แก่ กลุ่มสาขาวิชาวิทยาศาสตร์และเทคโนโลยี (Science Technology), กลุ่มสาขาวิชาวิทยาศาสตร์สุขภาพ (Health Science) และกลุ่มสาขาวิชามนุษยศาสตร์และสังคมศาสตร์ (Humanities) และแบ่งกลุ่มตามสถานที่ตั้งของกลุ่มสาขาวิชา ได้แก่ ประสานมิตรและองครักษ์ โดยอ้างอิงจากข้อมูลงานวิจัยของอาจารย์และบุคลากรของมหาวิทยาลัยศรีนครินทรวิโรฒที่ได้เผยแพร่ผลงานบนฐานข้อมูล Scopus ในช่วงปี ค.ศ. 2011 ถึง 2020 ในงานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาวิเคราะห์ลักษณะการแจกแจงของข้อมูลจากการทำการทดสอบการกระจายแบบปกติ (Normality Tests) และการทดสอบสถิติไม่อิงพารามิเตอร์ (Nonparametric Statistical Significance Tests) สำหรับค่ามัธยฐาน (median) ของกลุ่มประชากร รวมถึงการตรวจสอบและวิเคราะห์ข้อมูลค่าเฉลี่ยในแต่ละกลุ่มสาขาวิชาด้วยวิธีการทางสถิติเพื่อนำเสนอเป็นแนวทางในการส่งเสริมให้อาจารย์และบุคลากรในมหาวิทยาลัยศรีนครินทรวิโรฒทำผลงานวิจัย เช่น กรณีในกลุ่มสาขาวิชาที่ได้รับการอ้างอิงผลงานการวิจัยน้อย ควรส่งเสริมให้อาจารย์และบุคลากรในกลุ่มสาขาดังกล่าวทำผลงานวิจัยมากขึ้น เพื่อให้สอดคล้องกับพันธกิจของมหาวิทยาลัยศรีนครินทรวิโรฒที่ต้องการสร้างสรรค์งานวิจัยและนวัตกรรมที่มีคุณภาพ มีประโยชน์อย่างยั่งยืนต่อสังคมทั้งในระดับชาติและนานาชาติ

งานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้ทำการศึกษาค้นคว้าบทความงานวิจัยที่เกี่ยวข้องดังนี้

1. บทความวิจัยเรื่อง Citation analytics: Data exploration and comparative analyses of CiteScores of Open Access and Subscription-Based publications indexed in Scopus (2014–2016) (Atayero, Popoola, Egeonu, & Oludayo, 2018) [1] ได้บรรยายถึงผลลัพธ์และประสบการณ์การเรียนรู้จากการศึกษาทำความเข้าใจงานวิจัยข้อมูลของรูปแบบการตีพิมพ์งานวิจัยแบบ Open Access (OA) sources หรือ Subscription-Based (SB) ที่ส่งผลต่อ citation impact และในระยะยาวข้อมูลในบทความนี้จะช่วยให้ผู้กำหนดนโยบายและนักวิจัยในระดับสถาบันอุดมศึกษา (HEI) ระบุประเภทแหล่งที่มาของสิ่งพิมพ์และหมวดหมู่ที่เหมาะสมสำหรับการเผยแพร่ผลการวิจัยทางวิชาการที่มีผลกระทบต่ออ้างอิงสูงสุด โดยใช้ข้อมูล citation scores (CiteScores) จาก Scopus index ของปี ค.ศ. 2014 ถึง 2016

2. บทความวิจัยเรื่อง Status of India in science and technology as reflected in its publication output in the Scopus international database, 1996–2006 (Gupta & Dhawan, 2009) [2] ศึกษาหาตัวบ่งชี้ความก้าวหน้าทางด้านงานวิจัยของประเทศ วิเคราะห์การเติบโตของการตีพิมพ์ผลงานของอินเดียว่ามีอัตราส่วนการตีพิมพ์เท่าไร เมื่อเปรียบเทียบกับประเทศชั้นนำที่เลือกเพื่อหาตัวชี้วัดเหล่านี้ที่มีความสำคัญเพื่อวัดความก้าวหน้าใน S&T และใช้ในการปรับนโยบายในการพัฒนาด้าน S&T

(science and technology) โดยใช้ข้อมูลงานวิจัยที่ถูกตีพิมพ์ของประเทศอินเดียและอีก 20 ประเทศชั้นนำจาก Scopus database ในช่วงปี ค.ศ. 1996 ถึง 2006

3. บทความวิจัยเรื่อง Measuring research performance of Iraqi universities using Scopus data Ghassan (Abdul-Majeed, Mahmood, & Namer, 2021) [3] ได้บรรยายถึงผลลัพธ์ได้บรรยายถึงผลลัพธ์ตรวจสอบผลการวิจัยของมหาวิทยาลัยของรัฐและเอกชนในอิรักโดยใช้ฐานข้อมูลการอ้างอิงจากฐานข้อมูล Scopus โดยวัตถุประสงค์ของงานวิจัยนี้คือ วิเคราะห์ข้อมูลตัวบ่งชี้การวิจัย (Research Indicator) ของมหาวิทยาลัยอิรักโดยใช้ฐานข้อมูลจากฐานข้อมูล Scopus

4. บทความวิจัยเรื่อง An Examination of Citation Counts in a New Scholarly Communication Environment (Bauer, 2005) [4] ในงานวิจัยนี้ได้นำเสนอกรณีศึกษาที่เปรียบเทียบจำนวนการอ้างอิงจาก Web of Science, Scopus และ Google Scholar โดยใช้ข้อมูลจาก Journal of the American Society for Information Science and Technology (JASIST) ช่วงปี ค.ศ. 1985 และปี ค.ศ. 2000 เพื่อจับคู่ทดสอบ t-test เพื่อกำหนดนัยสำคัญทางสถิติ

5. บทความเรื่อง Normality Tests for Statistical Analysis: A Guide for Non-Statisticians (Ghasemi & Zahediasl, 2012) [5] การทดสอบการกระจายแบบปกติ Normality Tests (Test for goodness of fit) คือการประเมินสมมติฐานว่า ข้อมูลมีการกระจายตัวแบบปกติหรือไม่ หากการทดสอบมีนัยสำคัญทางสถิติ การแจกแจงในรูปแบบนั้นไม่ปกติ นั่นสำหรับตัวอย่างของข้อมูลที่มีขนาดเล็กแล้ว การทดสอบการกระจายตัวแบบปกติมีผลเพียงเล็กน้อยในการปฏิเสธสมมติฐาน ดังนั้นตัวอย่างข้อมูลที่มีขนาดเล็กมักจะผ่านการทดสอบการกระจายตัวแบบปกติ สำหรับตัวอย่างขนาดใหญ่ ผลลัพธ์ที่มีนัยสำคัญจะได้รับแม้ในกรณีที่มีการเบี่ยงเบนเล็กน้อย จะไม่ส่งผลต่อผลลัพธ์ของการทดสอบพารามेटริก โดยในงานวิจัยนี้ได้เลือกวิธีการทดสอบความเป็นอิสระต่อกันของกลุ่มประชากร

6. บทความเรื่อง สถิตินอนพารามेटริกและการประยุกต์ใช้ในงานวิจัยทางการแพทย์ (ศุภามณ จันทร์สกุล และสุชาดา บวรกิตติวงศ์, 2017) [6] ได้กล่าวถึงสถิติอ้างอิง (inferential statistics) ในการวิเคราะห์ ข้อมูลแบ่งเป็นสถิติพารามेटริก (parametric statistics) และสถิตินอนพารามेटริก (nonparametric statistics) บางครั้งเรียกสถิติที่ไม่ใช่พารามิเตอร์โดยพารามิเตอร์ หมายถึง ค่าที่แสดงคุณลักษณะของประชากร พารามิเตอร์ ที่สำคัญได้แก่ค่าเฉลี่ย มัชยฐาน ความแปรปรวน เป็นต้น การประมาณค่าพารามิเตอร์ทำได้โดยคำนวณค่าสถิติที่ได้ จากตัวอย่างไปทดสอบสมมติฐาน การตัดสินใจเลือกใช้ สถิติแบบมีพารามิเตอร์หรือแบบไม่มีพารามิเตอร์ขึ้นอยู่กับ ลักษณะข้อมูลเป็นไปตามข้อตกลงเบื้องต้นทางสถิติหรือไม่ เช่น ตัวอย่างสุ่มมาจากประชากรที่มีการแจกแจงปกติ ความแปรปรวนของประชากรแต่ละกลุ่มมีค่าเท่ากัน การวัด ตัวแปรตามอยู่ในระดับช่วงเป็นอย่างน้อย เป็นต้น

วิธีดำเนินการ

ขั้นตอนที่ 1 : การศึกษาชุดข้อมูล

ในขั้นตอนการศึกษาชุดข้อมูล ผู้วิจัยได้ดำเนินการแบ่งข้อมูลที่ใช้ในการศึกษาจะแบ่งออกเป็น 3 ชุดข้อมูล ได้แก่ ชุดข้อมูลการอ้างอิงผลงานวิจัยของอาจารย์และบุคลากรมหาวิทยาลัยศรีนครินทรวิโรฒ ผู้วิจัยได้รับชุดข้อมูลนี้จากเว็บไซต์ของฐานข้อมูล Scopus ซึ่งเป็นข้อมูลเกี่ยวกับข้อมูลงานวิจัยที่ถูกเผยแพร่บนฐานข้อมูล Scopus ในช่วงปี ค.ศ. 2011 ถึง 2020, ชุดข้อมูลบุคลากรของมหาวิทยาลัยศรีนครินทรวิโรฒ ผู้วิจัยได้รับชุดข้อมูลจากสำนักคอมพิวเตอร์ มหาวิทยาลัยศรีนครินทรวิโรฒ, ชุดข้อมูลคณะโดยแบ่งตามสาขาวิชา ผู้วิจัยได้รับข้อมูลจากสำนักคอมพิวเตอร์ มหาวิทยาลัยศรีนครินทรวิโรฒ โดยอ้างอิงการแบ่งกลุ่ม

ตามกลุ่มสาขาวิชาของมหาวิทยาลัยเชียงใหม่และปริกษารูปแบบการแบ่งกลุ่มข้อมูลกับทางบรรณารักษ์ของสำนักหอสมุดกลาง มหาวิทยาลัยศรีนครินทรวิโรฒ โดยแบ่งกลุ่มข้อมูลตามสาขาวิชาและสถานที่ตั้งของสาขาวิชา

ขั้นตอนที่ 2 : การนำเข้าข้อมูล การทำความสะอาดข้อมูล และการรวมชุดข้อมูล

ผู้วิจัยได้ทำการดาวน์โหลดชุดข้อมูลการอ้างอิงผลงานวิจัยของบุคลากรในมหาวิทยาลัยศรีนครินทรวิโรฒ ในช่วงปี ค.ศ. 2011 ถึง 2020 จากฐานข้อมูล Scopus ในรูปแบบไฟล์ CSV โดยแยกเป็นปีละ 1 ไฟล์ ต่อมาผู้วิจัยได้นำเข้าไฟล์และดำเนินการสร้าง Data frame ของชุดข้อมูลการอ้างอิงผลงานวิจัยของบุคลากรในมหาวิทยาลัยศรีนครินทรวิโรฒ โดยแยกข้อมูลของผู้ทำวิจัยเป็นชื่อผู้แต่ง (Authors) และหน่วยงานในสังกัดของผู้แต่งงานวิจัย (Authors Organization) โดยอิงจาก ID ผู้แต่งเพื่อทำความสะอาดข้อมูล (Cleansing Data) ด้วยวิธีการตัดคำนำหน้าชื่อของผู้แต่ง (Authors) และตรวจสอบรูปแบบหน่วยงานในสังกัดของผู้ทำงานวิจัย (Authors Organization) ที่พบว่ามีรูปแบบที่ผิด พร้อมตรวจสอบความถูกต้องของข้อมูลโดยการเปรียบเทียบปริมาณชุดข้อมูลก่อนและหลังการแยกข้อมูล ต่อมาผู้วิจัยได้ดำเนินการ รวมชุดข้อมูลบุคลากรของมหาวิทยาลัยศรีนครินทรวิโรฒและชุดข้อมูลคณะโดยแบ่งตามสาขาวิชาโดยอิงจากชื่อสาขาวิชา (Department Name) โดยได้ดำเนินการปรับรูปแบบชื่อผู้แต่งของชุดข้อมูลบุคลากรของมหาวิทยาลัยศรีนครินทรวิโรฒให้เหมือนกับชื่อผู้แต่งของชุดข้อมูลการอ้างอิงผลงานวิจัยจากฐานข้อมูล Scopus และดำเนินการรวมกับชุดข้อมูลการอ้างอิงผลงานวิจัยและข้อมูลบุคลากรของมหาวิทยาลัยศรีนครินทรวิโรฒที่ได้ทำความสะอาดชุดข้อมูลเรียบร้อยแล้วเข้า พร้อมตรวจสอบความถูกต้องโดยการตรวจสอบจากการเปรียบเทียบจำนวนการอ้างอิงผลงานวิจัย (Citation) และจำนวนผลงานวิจัย (Title) ก่อนและหลังการรวมชุดข้อมูลว่ามีข้อมูลที่ใกล้เคียงกันหรือไม่ ชุดข้อมูลก่อนและหลังมีความแตกต่างกันเกิดจากกรณีในผลงานวิจัยบางเรื่องผู้ทำวิจัยที่จากต่างกลุ่มสาขาวิชา หรือไม่มีชื่อในชุดข้อมูลบุคลากรมหาวิทยาลัยศรีนครินทรวิโรฒซึ่งตีความได้ว่าถูกเขียนโดยนิสิตและมีอาจารย์ที่ปรึกษาเป็นบุคลากรในมหาวิทยาลัย

ขั้นตอนที่ 3 : การตรวจสอบความถูกต้องของการทำความสะอาดชุดข้อมูล

หลังจากผู้วิจัยทำความสะอาดชุดข้อมูล (Cleaning Data) เรียบร้อยแล้ว ผู้วิจัยได้ทำการตรวจสอบความถูกต้องของการทำความสะอาดชุดข้อมูล โดยมีรายละเอียดการตรวจสอบดังนี้

1. หลังจากผู้วิจัยดำเนินการ Cleansing Data โดยการตัดคำนำหน้าชื่อของผู้แต่ง (Authors) และรูปแบบหน่วยงานในสังกัดของผู้ทำงานวิจัย (Authors Organization) ที่พบว่ามีรูปแบบที่ผิด ผู้วิจัยได้ตรวจสอบโดยวิธีการเปรียบเทียบข้อมูล ชื่อผู้แต่ง, ID ผู้แต่ง และในสังกัดของผู้ทำงานวิจัยว่ามีจำนวนที่เท่ากันหรือไม่

2. หลังจากผู้วิจัยได้รวมชุดข้อมูลการอ้างอิงผลงานวิจัย (Citation) และชุดข้อมูลบุคลากรมหาวิทยาลัยศรีนครินทรวิโรฒ ผู้วิจัยได้ทำการเปรียบเทียบจำนวนการอ้างอิงผลงานวิจัย (Citation) และจำนวนผลงานวิจัย (Title) ก่อนและหลังการรวมชุดข้อมูลว่ามีข้อมูลที่ใกล้เคียงกันหรือไม่ ชุดข้อมูลก่อนและหลังมีความแตกต่างกันเพราะในผลงานวิจัยบางเรื่องถูกเขียนโดยผู้ทำวิจัยที่จากต่างกลุ่มภาควิชา สาขาวิชา หรือไม่มีชื่อในฐานข้อมูลมหาวิทยาลัยศรีนครินทรวิโรฒ

ขั้นตอนที่ 4 : การทดสอบการกระจายแบบปกติ

ผู้วิจัยได้ดำเนินการทดสอบการแจกแจงของข้อมูลจากการทำการทดสอบการกระจายแบบปกติ (Normality Tests) โดยมีรายละเอียดการทดสอบดังนี้

1. การทดสอบการกระจายแบบปกติด้วยวิธี Anderson-Darling Normality Tests มีรายละเอียดการคำนวณดังนี้

Statistic หมายถึง ค่าทดสอบทางสถิติจากการทดสอบ Anderson-Darling

Critical values หมายถึง ค่าวิกฤตที่อยู่ในแต่ละช่วงตามระดับนัยสำคัญ

Significance level หมายถึง ช่วงระดับนัยสำคัญโดยจะแบ่งเป็นเปอร์เซ็นต์สอดคล้องกับค่าวิกฤต

ในการทดสอบ Anderson-Darling Normality Tests หากค่าสถิติที่ได้จากการคำนวณ (Statistic) มีค่ามากกว่าค่าวิกฤต (Critical values) ในระดับนัยสำคัญ (Significance level) ที่สอดคล้อง โดยในงานวิจัยนี้ผู้วิจัยจะใช้ช่วงระดับนัยสำคัญที่ 5 เปอร์เซ็นต์ หรือ $\alpha = 0.05$ ส่งผลให้ค่าสมมติฐานว่างของข้อมูลที่มาจากการแจกแจงนั้นถูกปฏิเสธ (Reject H0) หรือมีรูปแบบการแจกแจงไม่ปกติ (Non-normal distribution)

2. การทดสอบการกระจายแบบปกติด้วยวิธี Lilliefors Normality Tests มีรายละเอียดการคำนวณดังนี้

ในการทดสอบ Lilliefors Normality Tests ใช้ในการคำนวณค่า P (P-value) ของการทดสอบทางสถิติ ในกรณีที่ค่า P มีค่าน้อยกว่าค่าระดับนัยสำคัญที่กำหนด ในงานวิจัยนี้ผู้วิจัยจะใช้ช่วงระดับนัยสำคัญที่ 5 เปอร์เซ็นต์ หรือ $\alpha = 0.05$ ส่งผลให้สมมติฐานว่าง (null hypothesis) ของข้อมูลที่มาจากการแจกแจงนั้นถูกปฏิเสธ (Reject H0) หรือมีรูปแบบการแจกแจงไม่ปกติ (Non-normal distribution)

ขั้นตอนที่ 5 : ดำเนินการทดสอบทางสถิติ

เนื่องด้วยจากการทดสอบการแจกแจงของข้อมูลผู้วิจัยพบว่าชุดข้อมูลทั้งหมดมีรูปแบบการแจกแจงไม่ปกติ (Non-normal distribution) ผู้วิจัยได้ดำเนินการทดสอบทางสถิติ (Statistical Tests) ด้วยวิธีการทดสอบสถิติไม่อิงพารามิเตอร์ (Nonparametric Statistical Significance Tests) โดยมีรายละเอียดการทดสอบดังนี้

1. การทดสอบสถิติไม่อิงพารามิเตอร์ด้วยวิธี Mann-Whitney U Test มีรายละเอียดการคำนวณดังนี้

ในการทดสอบ Mann-Whitney U เป็นการทดสอบแบบไม่อิงพารามิเตอร์ (Nonparametric Statistical Significance Tests) ของสมมติฐานว่างว่าการแจกแจงต้นแบบ x นั้นเหมือนกับการแจกแจงที่อยู่ภายใต้ตัวอย่าง y มักใช้เป็นการทดสอบความแตกต่างของชุดข้อมูล โดยในกรณีที่ค่า P มีค่าน้อยกว่าค่าระดับนัยสำคัญที่กำหนด ในงานวิจัยนี้ผู้วิจัยจะใช้ช่วงระดับนัยสำคัญที่ 5

เปอร์เซ็นต์ หรือ $\alpha = 0.05$ ส่งผลให้สมมติฐานว่าง (null hypothesis) ของข้อมูลนั้นถูกปฏิเสธ (Reject H0) หรือกลุ่มตัวอย่างมีความต่างกันอย่างมีนัยสำคัญทางสถิติ (Sample distributions are not equal)

2. การทดสอบสถิติไม่อิงพารามิเตอร์ด้วยวิธี Kruskal-Wallis H Test มีรายละเอียดการคำนวณดังนี้

ในการทดสอบ Kruskal-Wallis H Test เปรียบเสมือนการทดสอบ แบบไม่อิงพารามิเตอร์ (Nonparametric Statistical Significance Tests) ของรูปแบบการทดสอบ ANOVA ซึ่งเป็นหนึ่งในตัวเลือกที่มีประสิทธิภาพสำหรับการวิเคราะห์ความแปรปรวนทางเดียว (One-way analysis of variance) ของข้อมูล โดยทำการทดสอบค่ามัธยฐาน (median) ของกลุ่มประชากรว่ามีค่าเท่ากันหรือไม่

ขั้นตอนที่ 6 : ตรวจสอบกลุ่มสาขาวิชาที่ได้รับค่าเฉลี่ยการอ้างอิงผลงานมากที่สุดกับกลุ่มที่มีค่าเฉลี่ยการอ้างอิงต่ำสุด

หลังการทดสอบทางสถิติผู้วิจัยได้ตรวจสอบกลุ่มสาขาวิชาที่ได้รับค่าเฉลี่ยการอ้างอิงผลงานมากที่สุดกับกลุ่มที่มีค่าเฉลี่ยการอ้างอิงต่ำสุด เพื่อวัดคุณภาพของงานวิจัยที่แต่ละสาขาวิชาได้เผยแพร่ โดยถ้าสาขาวิชานั้นตีพิมพ์ผลงานเป็นจำนวนมากแต่ค่าเฉลี่ยของการถูกอ้างอิงน้อยแสดงให้เห็นว่าคุณภาพงานวิจัยที่ตีพิมพ์ผลงานมีคุณภาพน้อยกว่าสาขาวิชาตีพิมพ์ผลงานน้อยแต่ได้รับการอ้างอิงมาก เพื่อนำผลที่ได้ไปต่อยอดด้านการจัดกลุ่มเพื่อทำงานวิจัยในแต่ละกลุ่มสาขาวิชา โดยควรใช้วิธีการให้กลุ่มสาขาวิชาที่ได้รับค่าเฉลี่ยการอ้างอิงผลงานมากที่สุดกับกลุ่มที่มีค่าเฉลี่ยการอ้างอิงต่ำสุดมาทำงานรวมเพื่อที่จะได้ช่วยกันผลิตงานวิจัยที่มีคุณภาพมากขึ้น

ผลการวิจัยและอภิปรายผลการวิจัย

1. การทดสอบการกระจายแบบปกติ (Normality Tests)

สมมติฐานที่ 1 ข้อมูลมีรูปแบบการแจกแจงปกติ

Fail to Reject H0 หมายถึง ข้อมูลมีรูปแบบการแจกแจงปกติ

Reject H0 หมายถึง ข้อมูลมีรูปแบบการแจกแจงไม่ปกติ

ตารางที่ 1 การทดสอบการกระจายแบบปกติด้วยวิธี Anderson-Darling Normality Tests แยกตามกลุ่มสาขาวิชา

Dept Group	Anderson-Darling Normality Test ($\alpha = 0.05$)		
	statistic	critical values	Null hypothesis
Health Science	111.837	0.783	Reject H0

Humanities	50.0222	0.779	Reject H0
Science Tech	132.987	0.784	Reject H0

ตารางที่ 2 การทดสอบการกระจายแบบปกติด้วยวิธี Anderson-Darling Normality Tests แยกตามสถานที่ตั้งสาขาวิชา

Dept Location	Anderson-Darling Normality Test ($\alpha = 0.05$)		
	statistic	critical values	Null hypothesis
Prasarnmit	159.137	0.784	Reject H0
Onkharak	145.661	0.784	Reject H0

จากตารางที่ 1 และ 2 การทดสอบการกระจายแบบปกติด้วยวิธี Anderson-Darling Normality Tests ตามสมมติฐานที่ 1 จะเห็นได้ว่าข้อมูลทั้งหมดมีรูปแบบการแจกแจงไม่ปกติ (Non-normal distribution) ซึ่งสามารถตรวจสอบได้จากค่าสถิติที่ได้จากการคำนวณ (Statistic) มีค่ามากกว่าค่าวิกฤต (Critical values) ในระดับนัยสำคัญ (Significance level) ที่ช่วงระดับนัยสำคัญที่ 5 เปอร์เซ็นต์ หรือ $\alpha = 0.05$ ตามที่ผู้วิจัยกำหนด

ตารางที่ 3 การทดสอบการกระจายแบบปกติด้วยวิธี Lilliefors Normality Test แยกตามกลุ่มสาขาวิชา

Dept Group	Lilliefors Normality Test ($\alpha = 0.05$)	
	P-value	Null hypothesis
Health Science	1.248e-222	Reject H0
Humanities	3.572e-95	Reject H0
Science Tech	1.047e-237	Reject H0

ตารางที่ 4 การทดสอบการกระจายแบบปกติด้วยวิธี Lilliefors Normality Test แยกตามสถานที่ตั้งสาขาวิชา

Dept Location	Lilliefors Normality Test ($\alpha = 0.05$)	
	P value	Null hypothesis
Prasarnmit	2.015e-288	Reject H0
Onkharak	2.831e-279	Reject H0

จากตารางที่ 3 และ 4 การทดสอบการกระจายแบบปกติด้วยวิธี Lilliefors Normality Test ตามสมมติฐานที่ 1 จะเห็นได้ว่าข้อมูลทั้งหมดมีรูปแบบการแจกแจงไม่ปกติ (Non-normal distribution) ซึ่งสามารถตรวจสอบได้จากค่า P (P-value) มีค่าน้อยกว่าค่าระดับนัยสำคัญที่กำหนด ในงานวิจัยนี้ผู้วิจัยจะใช้ช่วงระดับนัยสำคัญที่ 5 เปอร์เซ็นต์ หรือ $\alpha = 0.05$

สรุปจากการทดสอบการกระจายแบบปกติด้วยวิธี Anderson-Darling Normality Tests และ Lilliefors Normality Test ตามตารางที่ 1, 2, 3, 4 มีผลลัพธ์ที่สอดคล้องกันคือข้อมูลทั้งหมดมีรูปแบบการแจกแจงไม่ปกติ (Non-normal distribution)

2. การทดสอบสถิติไม่อิงพารามิเตอร์ (Nonparametric Statistical Significance Tests)

สมมติฐานที่ 2 กลุ่มสาขาวิชาที่ต่างกัน จะมีผลต่อปัจจัยที่ส่งผลต่อการถูกอ้างอิงผลงานที่ไม่แตกต่างกัน

Fail to Reject H0 หมายถึง กลุ่มสาขาวิชาที่แตกต่างกัน จะมีผลต่อปัจจัยที่ส่งผลต่อการถูกอ้างอิงผลงานไม่แตกต่างกัน

Reject H0 หมายถึง กลุ่มสาขาวิชาที่แตกต่างกัน จะมีผลต่อปัจจัยที่ส่งผลต่อการถูกอ้างอิงผลงานแตกต่างกัน

ตารางที่ 5 การทดสอบสถิติด้วยวิธี Mann-Whitney U Test และ Kruskal-Wallis H Test แยกตามกลุ่มสาขาวิชา

Dept Group	Mann-Whitney U Test (P-value)	Kruskal-Wallis H Test (P-value)	Null hypothesis
Health Science, Humanities	4.989e-06	9.974e-06	Reject H0
Health Science, Science Tech	0.104	0.209	Fail to reject H0
Humanities, Science Tech	6.106e-08	1.220e-07	Reject H0

จากตารางที่ 5 การทดสอบสถิติไม่อิงพารามิเตอร์ด้วยวิธี Mann-Whitney U Test และ Kruskal-Wallis H Test แยกตามกลุ่มสาขาวิชา ตามสมมติฐานที่ 2 จะเห็นได้ว่ากลุ่มสาขาวิชาที่แตกต่างกัน จะมีผลต่อปัจจัยที่ส่งผลการถูกอ้างอิงผลงานแตกต่างกันอย่างมีนัยสำคัญทางสถิติยกเว้นกลุ่มสาขาวิชาวิทยาศาสตร์สุขภาพ (Health Science) และกลุ่มสาขาวิชาวิทยาศาสตร์และเทคโนโลยี (Science Technology) ซึ่งมีค่า P-value ของ Mann-Whitney U Test อยู่ที่ 0.104 และมีค่า P-value ของ Kruskal-Wallis H Test 0.209 ซึ่งมากกว่าค่าระดับนัยสำคัญที่กำหนด ในงานวิจัยนี้ผู้วิจัยจะใช้ช่วงระดับนัยสำคัญที่ 5 เปอร์เซ็นต์ หรือ $\alpha = 0.05$

สมมติฐานที่ 3 สถานที่ตั้งสาขาวิชาที่ต่างกัน จะมีผลต่อปัจจัยที่ส่งผลการถูกอ้างอิงผลงานที่ไม่แตกต่างกัน

Fail to Reject H0 หมายถึง สถานที่ตั้งสาขาวิชาที่แตกต่างกัน จะมีผลต่อปัจจัยที่ส่งผลการถูกอ้างอิงผลงานไม่แตกต่างกัน

Reject H0 หมายถึง สถานที่ตั้งสาขาวิชาที่แตกต่างกัน จะมีผลต่อปัจจัยที่ส่งผลการถูกอ้างอิงผลงานแตกต่างกัน

ตารางที่ 6 การทดสอบด้วยวิธี Mann-Whitney U Test และ Kruskal-Wallis H Test แยกตามสถานที่ตั้งสาขาวิชา

Dept Group	Mann-Whitney U Test (P-value)	Kruskal-Wallis H Test (P-value)	Null hypothesis
Prasanmit, Ongkharak	0.054	0.108	Fail to Reject H0

จากตารางที่ 6 การทดสอบสถิติไม่อิงพารามิเตอร์ด้วยวิธี Mann-Whitney U Test และ Kruskal-Wallis H Test แยกตามสถานที่ตั้งสาขาวิชาตามสมมติฐานที่ 3 จะเห็นได้ว่าสถานที่ตั้งสาขาวิชาที่แตกต่างกัน จะมีผลต่อปัจจัยที่ส่งผลการถูกอ้างอิงผลงานไม่แตกต่างกัน โดยมีค่า P-value ของ Mann-Whitney U Test อยู่ที่ 0.054 และมีค่า P-value ของ Kruskal-Wallis H Test 0.108 ซึ่งมากกว่าค่าระดับนัยสำคัญที่กำหนด ในงานวิจัยนี้ผู้วิจัยจะใช้ช่วงระดับนัยสำคัญที่ 5 เปอร์เซ็นต์ หรือ $\alpha = 0.05$

3. กลุ่มสาขาวิชาที่ได้รับค่าเฉลี่ยการอ้างอิงผลงานมากที่สุดกับกลุ่มที่มีค่าเฉลี่ยการอ้างอิงต่ำสุด

ตารางที่ 7 ตารางผลจำนวนผลงานที่ถูกตีพิมพ์และค่าเฉลี่ยการถูกอ้างอิงของกลุ่มสาขาวิชามนุษยศาสตร์และสังคมศาสตร์ (Humanities)

2021 1st Proceeding of the Data Science Conference

DEPT NAME	Paper	Mean Cited
คณะวัฒนธรรมสิ่งแวดล้อมและการท่องเที่ยวเชิงนิเวศ	21	5.714
สำนักนวัตกรรมการเรียนรู้	20	6.428
วิทยาลัยอุตสาหกรรมสร้างสรรค์	10	4.823
คณะมนุษยศาสตร์	9	7.695
สถาบันวิจัยพฤติกรรมศาสตร์	9	2.566
วิทยาลัยนานาชาติเพื่อศึกษาความยั่งยืน	8	3.833
วิทยาลัยโพธิวิชชาลัย	6	3.461
คณะบริหารธุรกิจเพื่อสังคม	6	3.956
คณะสังคมศาสตร์	5	1.176
คณะศึกษาศาสตร์	4	0.578
วิทยาลัยนวัตกรรมการสื่อสารสังคม	3	1.142
คณะเศรษฐศาสตร์	3	2.562
สำนักงานอธิการบดี	3	5
สำนักทดสอบทางการศึกษาและจิตวิทยา	3	5
สถาบันวิจัย พัฒนา และสาธิตการศึกษา	2	0.5
คณะศิลปกรรมศาสตร์	1	0
สำนักหอสมุดกลาง	1	2
สถาบันพัฒนาการท่องเที่ยวเพื่ออนุรักษ์สิ่งแวดล้อม	0	0
สถาบันยุทธศาสตร์ทางปัญญาและวิจัย	0	0
สถาบันวัฒนธรรมและศิลปะ	0	0
สำนักงานสภามหาวิทยาลัย	0	0
สำนักสื่อและเทคโนโลยีการศึกษา	0	0
สถาบันสิ่งแวดล้อมและทรัพยากร	0	0
ศูนย์วิชาศึกษาทั่วไป	0	0
ศูนย์นวัตกรรมการเรียนรู้ตลอดชีวิต	0	0
สำนักวิชาเศรษฐศาสตร์และนโยบายสาธารณะ	0	0

จากตารางที่ 7 แสดงให้เห็นว่าคณะมนุษยศาสตร์เป็นคณะที่ผลิตผลงานวิจัยที่มีคุณภาพมากที่สุดในกลุ่มสาขาวิชา เนื่องจากมีการตีพิมพ์ผลงาน 9 ครั้งแต่ค่าเฉลี่ยการถูกอ้างอิงสูงถึง 7.695 ครั้งต่อผลงาน ต่างกับคณะวัฒนธรรมสิ่งแวดล้อมและการท่องเที่ยวเชิงนิเวศที่จำนวนการเผยแพร่ผลงานเป็นอันดับแรกด้วยจำนวน 21 ครั้ง แต่ถูกอ้างอิงต่อผลงานโดยมีค่าเฉลี่ย 5.714 เป็นอันดับสามของผลเฉลี่ยการถูกอ้างอิงต่อผลงานวิจัย โดยในกลุ่มสาขาวิชานี้มีหน่วยงานที่ไม่มีผลงานวิจัยเลย

จำนวน 9 หน่วยงาน ได้แก่ สถาบันพัฒนาการท่องเที่ยวเพื่ออนุรักษ์สิ่งแวดล้อม, สถาบันยุทธศาสตร์ทางปัญญาและวิจัย, สถาบันวัฒนธรรมและศิลปะ, สำนักงานสภามหาวิทยาลัย, สำนักสื่อและเทคโนโลยีการศึกษา, สถาบันสิ่งแวดล้อมและทรัพยากร, ศูนย์วิชาศึกษาทั่วไป, ศูนย์นวัตกรรมการเรียนรู้ตลอดชีวิต, สำนักวิชาเศรษฐศาสตร์และนโยบายสาธารณะ

ตารางที่ 8 ตารางผลจำนวนผลงานที่ถูกตีพิมพ์และค่าเฉลี่ยการถูกอ้างอิงของกลุ่มสาขาวิชาวิทยาศาสตร์สุขภาพ (Health Science)

DEPT NAME	Paper	Mean Cited
คณะแพทยศาสตร์	53	7.770
คณะเภสัชศาสตร์	32	8.146
คณะทันตแพทยศาสตร์	18	6.672
ศูนย์การแพทย์ปัญญานันทภิกขุ ชลประทาน	18	8.585
มหาวิทยาลัยศรีนครินทรวิโรฒ	14	8.136
คณะกายภาพบำบัด	13	7.312
คณะพลศึกษา	12	5.458
ศูนย์การแพทย์สมเด็จพระเทพรัตนราชสุดาฯ	7	4
คณะพยาบาลศาสตร์	6	26.75
บัณฑิตวิทยาลัย	5	6.545
คณะสหเวชศาสตร์	0	0

จากตารางที่ 8 แสดงให้เห็นว่าคณะพยาบาลศาสตร์เป็นคณะที่ผลิตผลงานวิจัยที่มีคุณภาพมากที่สุดในกลุ่มสาขาวิชา เนื่องจากมีการตีพิมพ์ผลงาน 6 ครั้งแต่ค่าเฉลี่ยการถูกอ้างอิงสูงถึง 26.75 ครั้งต่อผลงานซึ่งต่างกับคณะแพทยศาสตร์ที่จำนวนการตีพิมพ์เป็นอันดับแรกด้วยจำนวน 53 ครั้งค่าเฉลี่ยการถูกอ้างอิงต่อผลงานโดยมีค่าเฉลี่ย 5.714 ครั้งต่อผลงาน โดยในกลุ่มสาขาวิชานี้มีหน่วยงานที่ไม่มีผลงานวิจัยเลยจำนวน 1 หน่วยงาน ได้แก่ คณะสหเวชศาสตร์

ตารางที่ 9 ตารางผลจำนวนผลงานที่ถูกตีพิมพ์และค่าเฉลี่ยการถูกอ้างอิงของกลุ่มสาขาวิชาวิทยาศาสตร์และเทคโนโลยี (Science Technology)

DEPT NAME	Paper	Mean Cited
คณะวิทยาศาสตร์	70	10.569
คณะวิศวกรรมศาสตร์	51	7.852
คณะเทคโนโลยีและนวัตกรรมผลิตภัณฑ์การเกษตร	26	10.475
ศูนย์วิทยาศาสตร์ศึกษา	0	0
สำนักคอมพิวเตอร์	0	0

จากตารางที่ 9 แสดงให้เห็นว่าคณะเทคโนโลยีและนวัตกรรมผลิตภัณฑการเกษตรเป็นคณะที่ผลิตผลงานวิจัยที่มีคุณภาพมากที่สุดในกลุ่มสาขาวิชา เนื่องจากการตีพิมพ์ผลงาน 26 ครั้งแต่ค่าเฉลี่ยของการถูกอ้างอิงสูงถึง 10.475 ครั้งต่อผลงาน ถึงแม้คณะวิทยาศาสตร์จะมีค่าเฉลี่ยของการถูกอ้างอิง 10.569 ครั้งต่อผลงาน ซึ่งมากกว่าคณะเทคโนโลยีและนวัตกรรมผลิตภัณฑการเกษตรแต่ด้วยคณะวิทยาศาสตร์มีจำนวนผลงานที่ตีพิมพ์มากกว่าถึง 3 เท่าโดยประมาณแสดงให้เห็นว่าคุณภาพของผลงานวิจัยที่ถูกตีพิมพ์มีว่าคุณภาพมากกว่า โดยในกลุ่มสาขาวิชานี้มีหน่วยงานที่ไม่มีผลงานวิจัยเลยจำนวน 2 หน่วยงาน ได้แก่ ศูนย์วิทยาศาสตร์ศึกษา, สำนักคอมพิวเตอร์

สรุปผลการวิจัย

งานวิจัยนี้เป็นการศึกษาการอ้างอิงของงานวิจัยของอาจารย์และบุคลากรของมหาวิทยาลัยศรีนครินทรวิโรฒ โดยแบ่งกลุ่มข้อมูลตามสาขาวิชา และสถานที่ตั้งของสาขาวิชา โดยใช้ข้อมูลจากฐานข้อมูล Scopus ในช่วงปี ค.ศ. 2011 ถึง 2020 ในการทดสอบและการวิเคราะห์ข้อมูลเชิงสถิติในงานวิจัยนี้ ผู้วิจัยได้ข้อสรุปว่าควรส่งเสริมและสนับสนุนให้ในกลุ่มอาจารย์และบุคลากรในกลุ่มหน่วยงานที่ไม่มี การตีพิมพ์ผลงานวิจัยเลยให้ทำงานวิจัยมากขึ้นเป็นอันดับแรก และสนับสนุนสาขาวิชามนุษยศาสตร์และสังคมศาสตร์เป็นอันดับสองจากผลค่าเฉลี่ยการถูกอ้างอิงผลงานวิจัย รวมถึงยังสามารถนำข้อมูลที่แสดงผลจากงานวิจัยนี้ไปต่อยอดเป็นแนวคิดในการจัดกลุ่มเพื่อทำงานวิจัย โดยควรใช้วิธีการให้กลุ่มสาขาวิชาเดียวกันที่ได้รับค่าเฉลี่ยการอ้างอิงผลงานมากที่สุดกับกลุ่มที่มีค่าเฉลี่ยการอ้างอิงต่ำสุดมาทำงานร่วมเพื่อที่จะได้ช่วยกันผลิตงานวิจัยที่มีคุณภาพมากขึ้น

กิตติกรรมประกาศ

การจัดทำวิจัยได้รับการสนับสนุนจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ในการนำเสนอผลงานวิจัย ผู้วิจัยจึงขอขอบคุณมา ณ ที่นี้

เอกสารอ้างอิง

- [1] Atayero, A. A., Popoola, S. I., Egeonu, J., & Oludayo, O. (2018). Citation analytics: Data exploration and comparative analyses of CiteScores of Open Access and Subscription-Based publications indexed in Scopus (2014–2016). *Data in Brief*, 19, 198-213.
- [2] Gupta, B. M., & Dhawan, S. M. (2009). Status of India in science and technology as reflected in its publication output in the Scopus international database, 1996–2006. *Scientometrics*, 80(2), 473-490
- [3] Abdul-Majeed, G., Mahmood, W., & Namer, N. S. M. (2021). Measuring research performance of Iraqi universities using Scopus data. *Scientometrics*, 126(3), 2349-2363.

2021 1st Proceeding of the Data Science Conference

[4] Bauer, K. (2005). An Examination of Citation Counts in a New Scholarly Communication Environment. D-Lib Magazine.

[5] Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. International journal of endocrinology and metabolism, 10(2), 486–489. <https://doi.org/10.5812/ijem.3505>

[6] Chansakul, S., & Bowarnkitiwong, S. (2017). สถิติ นอน พารา เมตริก และ การ ประยุกต์ ใช้ ใน งาน วิจัย ทาง การ พยาบาล. EAU Heritage Journal Science and Technology, 11(1), 38-48.