

สถิติความสัมพันธ์: เลือกใช้อย่างไร Correlation in Statistics: How to Use

สมถวิล วิจิตรวรรณ
อาจารย์ประจำหลักสูตรประกาศนียบัตรบัณฑิตวิชาชีพครู มหาวิทยาลัยราชพฤกษ์
E-mail: sowiji@rpu.ac.th
Somtawin Wjitwanna
Lecturer, Graduate Diploma in Teaching Profession, Rajaprak University

รับเข้า: 6 พฤศจิกายน 2563 แก้ไข: 12 ธันวาคม 2563 ตอรับ: 16 ธันวาคม 2563

บทคัดย่อ

สถิติวิเคราะห์ความสัมพันธ์ได้นำไปใช้ในการอธิบายความเกี่ยวข้องของตัวแปรตั้งแต่ 2 ตัวแปรขึ้นไป บทความนี้มีจุดมุ่งหมายในการเลือกใช้และการแปลผลของสถิติการวิเคราะห์ความสัมพันธ์ หลักเกณฑ์การเลือกใช้ที่สำคัญ คือ ระดับการวัดของตัวแปรและเงื่อนไขเฉพาะของสถิติ โดยที่สัมประสิทธิ์ Phi และสถิติ Chi Square ใช้วิเคราะห์ความสัมพันธ์ตัวแปรระดับนามบัญญัติ สัมประสิทธิ์ Spearman ใช้วิเคราะห์ความสัมพันธ์ของตัวแปรระดับเรียงอันดับ สัมประสิทธิ์ Pearson ใช้วิเคราะห์ความสัมพันธ์ของตัวแปรระดับอันดับและอัตราส่วน การแปลผลของสถิติที่ใช้วิเคราะห์ความสัมพันธ์ไม่เหมือนกัน ผู้วิจัยจึงต้องทำความเข้าใจสถิติความสัมพันธ์เหล่านี้ นอกจากนี้ยังกล่าวถึงข้อควรระวังในการใช้ผลการวิเคราะห์ความสัมพันธ์

คำสำคัญ: การวิเคราะห์ความสัมพันธ์ Phi, Chi square, Spearman, Pearson

Abstract

Correlation in statistics is used to describe the relevance of two or more variables. The aim of this article is how to select and interpret the statistics used in correlation analysis. The selection of correlation statistics takes into account measurement scale of the variables and the specific conditions of those statistics. In this paper, Phi coefficient and Chi Square statistic are used to analyze the relationship of nominal scale variables. Spearman coefficient is used to analyze the relationship of the ordinal-scale variables. The Pearson coefficient is used to analyze the relationship of

the interval and the ratio scale variables. The interpretations of correlation statistics are not the similar meaning. Researchers need to understand these correlation statistics. In addition, caution is required to use of correlation analysis results.

Keywords: correlation analysis, Phi, Chi square, Spearman, Pearson

บทนำ

การศึกษาความสัมพันธ์เริ่มจากผู้วิจัยเห็นการเปลี่ยนแปลงของปรากฏการณ์ระหว่าง 2 ตัวแปร เช่น รีวรอยของกางเกงยีนยิ่งมาก ประสบการณ์ของการเดินทางก็มากด้วย อากาศร้อนมาก ค่าไฟฟ้ามากขึ้น ด้วย คือ ความสัมพันธ์ระหว่างจำนวนรีวรอยของกางเกงยีนส์ กับจำนวนครั้งของการเดินทาง ความสัมพันธ์ระหว่างอุณหภูมิห้องกับค่าไฟฟ้า ความสัมพันธ์ลักษณะนี้มีแนวโน้มไปในทางเดียวกันเป็น ความสัมพันธ์ทางบวก ในทางกลับกัน คนที่อายุยิ่งมากความสามารถในการฟังน้อยลง ความสัมพันธ์ ลักษณะนี้มีแนวโน้มในทิศทางผกผันกันเป็นความสัมพันธ์ทางลบ ในการศึกษาความสัมพันธ์ผู้วิจัยมี ข้อมูลเชิงปริมาณ 2 ชุดขึ้นไปที่มีกรอบแนวคิดที่น่าจะมีความเกี่ยวข้องหรือมีความสัมพันธ์กัน สถิติ ความสัมพันธ์มีความสำคัญนำไปใช้ในหลายด้าน เช่น การวิเคราะห์คุณภาพเครื่องมือวิจัย รวมถึงเป็น สถิติที่สำคัญที่มีการขยายต่อ ยอดไปเป็นสถิติขั้นสูงที่ตอบคำถามการวิจัยที่ซับซ้อนขึ้น บทความนี้มี จุดมุ่งหมายในการเลือกใช้และการแปลผลของสถิติการวิเคราะห์ความสัมพันธ์ หลักเกณฑ์การเลือกใช้ ที่สำคัญ คือ ระดับการวัดของตัวแปรและเงื่อนไขเฉพาะของสถิติ จึงขอกล่าวถึงแนวคิดเกี่ยวกับตัวแปร และข้อมูลก่อน ดังนี้

1. แนวคิดเกี่ยวกับตัวแปรและข้อมูล

สถิติที่ใช้วิเคราะห์ความสัมพันธ์มีหลายประเภท บางสถิติให้ผลวิเคราะห์ทั้งทิศทางและขนาด ความสัมพันธ์ บอกได้ว่า 2 ตัวแปร มีความสัมพันธ์กันหรือไม่มีความสัมพันธ์กัน ไม่บอกทิศทางและ ขนาดความสัมพันธ์การเลือกใช้สถิติประเภทความสัมพันธ์ขึ้นอยู่กับระดับการวัดข้อมูล จึงขอกล่าวถึง แนวคิดเกี่ยวกับตัวแปรและข้อมูลซึ่งเป็นพื้นฐานสำคัญในการเลือกใช้สถิติ

ตารางที่ 1 ตัวอย่างตัวแปรและข้อมูลที่ได้จากการวัดตัวแปร

ตัวแปร	ข้อมูล (data)
เพศ	หญิง ชาย เพศทางเลือก
ระดับการศึกษา	ประถม มัธยม ปริญญาตรี ปริญญาโท ปริญญาเอก
ประสบการณ์ทำงาน	น้อยกว่า 5 ปี 5-10 ปี 11-15 ปี 15 ปีขึ้นไป
คุณภาพสินค้า	ระดับ 1 (ดี) 2 (ปานกลาง) 3 (ต้องปรับปรุง)
ความพึงพอใจ	5 (มากที่สุด) 4 (มาก) 3 (ปานกลาง) 2 (น้อย) 1 (น้อยที่สุด)

ตัวแปรที่ใช้พิจารณาในการเลือกสถิติประเภทความสัมพันธ์แบ่งเป็น 4 ประเภทตามระดับการวัดข้อมูล คือ

1. ระดับการวัดนามบัญญัติ (Nominal Scale) เป็นการวัดที่จำแนกกลุ่มข้อมูล เช่น เพศ อาชีพ ประสบการณ์ทำงานที่แบ่งเป็นกลุ่ม เป็นต้น

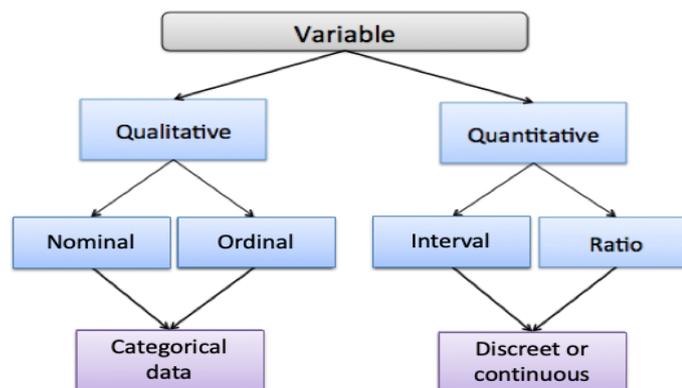
2. ระดับการวัดเรียงอันดับ (Ordinal Scale) เป็นการวัดที่จำแนกกลุ่มได้ เรียงลำดับกลุ่มข้อมูลได้โดยใช้เกณฑ์ที่ชัดเจน เช่น ระดับการศึกษา คุณภาพสินค้า เป็นต้น

ข้อมูลระดับที่ 1 และ 2 เป็นข้อมูลเชิงคุณภาพ (Qualitative Data) ไม่ใช่ตัวเลข หากเป็นตัวเลขเป็นตัวเลขสมมติที่ใช้เป็นรหัสเรียก การวิเคราะห์ข้อมูลทั้ง 2 ระดับ แต่ไม่สามารถนำมาคำนวณได้ เพียงแต่ใช้นับจำนวน (ความถี่) ของแต่ละกลุ่ม เรียกข้อมูลการวัดทั้งสองระดับว่าข้อมูลจัดกลุ่ม (categorical data) และเรียกตัวแปรทั้ง 2 ระดับนี้ว่า ตัวแปรจัดกลุ่ม (categorical variable)

3. ระดับการวัดอันตรภาค (Interval Scale) เป็นการวัดที่จำแนกได้เป็นกลุ่มเรียงลำดับได้ และช่วงห่างของแต่ละมาตรวัดเท่ากัน จึงสามารถนำมาบวกลบคูณหารได้ แต่ปริมาณ 0 ไม่ใช่ค่าศูนย์แท้หรือไม่ใช่จุดเริ่มต้นที่ ได้แก่ คะแนน อุณหภูมิ ความดันโลหิต ความพึงพอใจ เป็นต้น

4. ระดับการวัดอัตราส่วน (Ratio Scale) เป็นการวัดที่เทียบสิ่งที่วัดกับมาตรฐานที่มีจุดเริ่มต้นที่ 0 ที่มีความหมายว่าไม่มีค่าใด ๆ ดังนั้นจึงนำมาบวก ลบ คูณ หารได้ ได้แก่ ระยะทาง เงินเดือน ประสบการณ์ทำงานที่ให้เติมข้อมูล เป็นต้น

ข้อมูลระดับที่ 3 และ 4 เป็นข้อมูลเชิงปริมาณ (Quantitative Data) ข้อมูลทั้งสองระดับนี้ส่วนใหญ่มีความต่อเนื่อง มักจะเรียกตัวแปรทั้งสองระดับว่า ตัวแปรต่อเนื่อง (continuous variable) ในโปรแกรม SPSS เรียกข้อมูล 2 ระดับนี้ว่า scale สำหรับตัวแปรเชิงปริมาณ เช่น อายุ ประสบการณ์ทำงาน ที่ให้ผู้ตอบเติมข้อมูลตามความจริงจัดเป็นตัวแปรต่อเนื่องสามารถหาค่าเฉลี่ยได้ แต่ถ้าผู้วิจัยจำแนกเป็นกลุ่ม ๆ เช่น อายุ น้อยกว่า 20 ปี 21-40 ปี 41-60 ปี และ 60 ปีขึ้นไป ตัวแปรอายุในลักษณะนี้เป็นตัวแปรจัดกลุ่ม



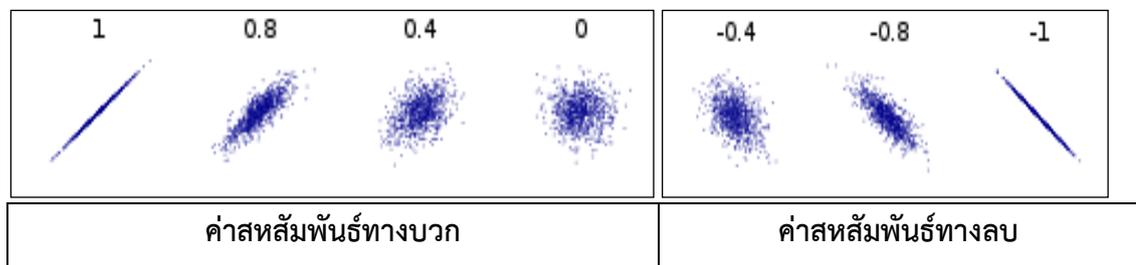
ภาพที่ 1 ตัวแปรจำแนกตามระดับการวัดข้อมูล

ที่มา: C. Muller (1973)

2. การเลือกใช้สถิติประเภทความสัมพันธ์

การวิเคราะห์ความสัมพันธ์ของข้อมูลที่เก็บรวบรวมจากประชากร สามารถแปลผลความสัมพันธ์โดยใช้เกณฑ์ที่กล่าวมา แต่หากเก็บข้อมูลจากกลุ่มตัวอย่างจะต้องมีการทดสอบสมมติฐาน หากไม่มีนัยสำคัญทางสถิติ แสดงว่าตัวแปรไม่มีความสัมพันธ์กันจะไม่สามารถใช้เกณฑ์นี้อ้างอิงได้ แต่ถ้าพบว่ามีนัยสำคัญทางสถิติแสดงว่า ตัวแปรมีความสัมพันธ์กันจะสามารถใช้เกณฑ์นี้

สถิติประเภทความสัมพันธ์ระหว่าง 2 ตัวแปร มีค่าระหว่าง -1.00 ถึง +1.00 ขนาดและทิศทางของความสัมพันธ์พิจารณาจากค่าดัชนีสัมประสิทธิ์สหสัมพันธ์ ข้อตกลงเบื้องต้นคือ ตัวแปรระดับการวัดแบบอันตรภาคและแบบอัตราส่วน (Interval Scale และ Ratio Scale) กลุ่มตัวอย่างได้จากการสุ่มจากประชากรที่มีการแจกแจงแบบโค้งปกติ ความหมายของขนาดของความสัมพันธ์ เพื่อให้เข้าใจระดับความสัมพันธ์ระหว่างตัวแปรได้ง่ายขึ้น สามารถแสดงเป็นภาพได้ดังนี้



ภาพที่ 2 ระดับความสัมพันธ์ระหว่างตัวแปร

จากภาพแสดงทิศทางและขนาดของความสัมพันธ์ พิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ มีค่าอยู่ระหว่าง -1.0 ถึง +1.0 สำหรับทิศทางของความสัมพันธ์พิจารณาจากเครื่องหมาย ค่าสหสัมพันธ์เป็นบวกแสดงว่าตัวแปรทั้งสองเปลี่ยนแปลงไปในทิศทางเดียวกัน ส่วนค่าความสัมพันธ์เป็นลบ แสดงว่า ตัวแปรทั้งสองเปลี่ยนแปลงไปในทิศทางกลับตรงกันข้าม ขนาดของความสัมพันธ์ ตัวเลขเป็นการบอกค่าความสัมพันธ์ที่ค่ายิ่งมากใกล้ -1.00 หรือ +1.00 เช่น 0.9 หรือ -0.9 แสดงว่ามีความสัมพันธ์สูงมาก ค่าเป็น 0 หมายถึง ทั้งสองตัวแปรไม่มีความสัมพันธ์ต่อกัน

การวิเคราะห์ความสัมพันธ์ของข้อมูลที่เก็บรวบรวมจากประชากร สามารถแปลผลความสัมพันธ์โดยใช้เกณฑ์ของที่กล่าวมา แต่หากเก็บข้อมูลจากกลุ่มตัวอย่างจะต้องมีการทดสอบสมมติฐาน หากไม่มีนัยสำคัญทางสถิติ แสดงว่าตัวแปรไม่มีความสัมพันธ์กันจะไม่สามารถใช้เกณฑ์นี้อ้างอิงได้ แต่ถ้าพบว่ามีนัยสำคัญทางสถิติแสดงว่า ตัวแปรมีความสัมพันธ์กันจะสามารถใช้เกณฑ์นี้อ้างอิงได้ ตัวแปรทุกระดับการวัดสามารถหาความสัมพันธ์ได้ สถิติแต่ละประเภทจะมีเงื่อนไขแตกต่างกันไปตามระดับการวัดของตัวแปร ดังนั้นในการเลือกใช้สถิติประเภทนี้ ผู้วิจัยจึงต้องศึกษาเงื่อนไขแต่ละสถิติเพื่อแปลผลการวิเคราะห์ที่ได้ถูกต้อง ในที่นี้ผู้เขียนได้สังเคราะห์และสรุปแนวทางการเลือกใช้สถิติสำหรับใช้วิเคราะห์ความสัมพันธ์ของข้อมูลวิจัยดังตารางที่ 2

ตารางที่ 2 แนวทางในการเลือกสถิติวิเคราะห์ความสัมพันธ์

ระดับการวัด		สถิติที่ใช้วิเคราะห์ ความสัมพันธ์	ตัวอย่าง
ตัวแปรตัวที่ 1	ตัวแปรตัวที่ 2		
Nominal scale แบ่งสองกลุ่ม	Nominal scale แบ่งสองกลุ่ม	Phi Coefficient	เพศของนักศึกษากับการ รับประทานอาหารเช้า
Nominal scale	Nominal scale Ordinal scale	Chi-Square test	การสอบผ่าน-ไม่ผ่านกับเพศ ของนักเรียน
Ordinal scale	Ordinal scale	Spearman Rank Correlation	การจัดอันดับผลงานของ นักเรียน กับเกรดของวิชา
Interval scale หรือ Ratio scale)	Interval scale หรือ Ratio scale)	Pearson Correlation	คะแนนสอบกับเกรดดีต่อการศึกษา เรียน

สำหรับการเก็บข้อมูลจากกลุ่มตัวอย่างที่สุ่มมาจากประชากรเป้าหมาย การสรุปผลจากกลุ่มตัวอย่างไปยังที่ประชากร จำเป็นต้องมีการใช้สถิติอ้างอิง เพื่อทดสอบนัยสำคัญทางสถิติว่าผลสรุปนี้สามารถอ้างอิงไปยังประชากรได้หรือไม่ จึงนำสมมติฐานการวิจัยมากำหนดเป็นสมมติฐานทางสถิติ คือ (นงลักษณ์ วิรัชชัย, 2548; สุชาติดา บวรกิติวงศ์, 2548)

H_0 : ตัวแปรทั้งสองเป็นอิสระกันหรือไม่มีความสัมพันธ์กัน

H_1 : ตัวแปรทั้งสองไม่เป็นอิสระกันหรือมีความสัมพันธ์กัน

การตัดสินใจว่าจะปฏิเสธหรือยอมรับ H_0 พิจารณาจากค่า Sig. หรือ p-value ดังนี้

1. หากค่า p-value ในตารางมีค่ามากกว่า 0.05 ผลการตัดสินใจคือ ยอมรับ H_0 นั่นคือ **ตัวแปรทั้งสองไม่มีความสัมพันธ์กัน** ผลวิเคราะห์นี้มีความเชื่อมั่นในการที่จะตัดสินใจถูก ร้อยละ 95
2. หาก p-value ในตารางมีค่าน้อยกว่า 0.05 แต่มากกว่า 0.01 ผลการตัดสินใจคือ **ปฏิเสธ H_0** (ยอมรับสมมติฐานทางเลือก H_1) นั่นคือ **ตัวแปรทั้งสองมีความสัมพันธ์กัน** ผลวิเคราะห์นี้มีความเชื่อมั่นได้ ร้อยละ 95 หรือภาษาสถิติใช้คำว่า มีนัยสำคัญทางสถิติที่ระดับ 0.05
3. หาก p-value (หรือค่า **Sig.(2-tailed)**) ในตารางมีค่าน้อยกว่า 0.01 ผลการตัดสินใจคือ **ปฏิเสธ H_0** (ยอมรับสมมติฐานทางเลือก H_1) นั่นคือ **ตัวแปรทั้งสองมีความสัมพันธ์กัน** ผลวิเคราะห์นี้มีความเชื่อมั่นได้ ร้อยละ 99 หรือภาษาสถิติใช้คำว่า มีนัยสำคัญทางสถิติที่ระดับ 0.01

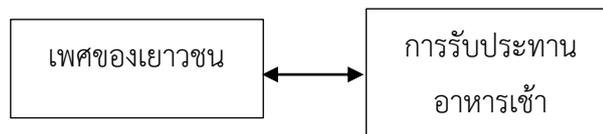
3. สถิติประเภทความสัมพันธ์ Phi

สถิติที่วิเคราะห์ความสัมพันธ์ ประเภทสัมประสิทธิ์ Phi (Phi Coefficient) สัมประสิทธิ์ครามเมอร์วี (Cramer's V) เป็นการทดสอบความเป็นอิสระต่อกันของสองตัวแปร การเลือกใช้สถิตินี้ใช้เมื่อข้อมูลมีระดับการวัดนามบัญญัติ แต่ละตัวแปรจำแนกเป็น 2 กลุ่ม เช่น

เพศ ข้อมูลได้แก่ ชาย หญิง การซื้อโฆษณา ข้อมูลได้แก่ ซื้อ ไม่ซื้อ
ระดับการศึกษา ข้อมูลได้แก่ ปริญญาตรี สูงกว่าปริญญาตรี

การพิสูจน์ว่าตัวแปรแรกมีความอิสระกับตัวแปรที่สองหรือไม่ ข้อมูลอยู่ในรูปของความถี่ หรือร้อยละ นำเสนอเป็นตาราง Contingency Table สำหรับสมมติฐานทางสถิติคือ

ตัวอย่างที่ 1 ความสัมพันธ์ที่วิเคราะห์ด้วย สัมประสิทธิ์ Phi (Phi Coefficient) การศึกษาความสัมพันธ์ระหว่างเพศของเยาวชนกับการรับประทานอาหารเช้า โดยสำรวจกับกลุ่มตัวอย่างนิสิต 275 คน **สมมติฐานการวิจัย** เพศของเยาวชนกับการรับประทานอาหารเช้ามีความสัมพันธ์กัน



ภาพที่ 3 กรอบแนวคิดในการวิจัยตัวอย่างที่ 1

จากสมมติฐานการวิจัย มากำหนดสมมติฐานทางสถิติได้ ดังนี้

H_0 : เพศของเยาวชนกับการรับประทานอาหารเช้า ไม่มีความสัมพันธ์กัน

H_1 : เพศของเยาวชนกับการรับประทานอาหารเช้า มีความสัมพันธ์กัน

การวิเคราะห์ด้วยโปรแกรมสำเร็จรูป SPSS คำสั่งที่สำคัญ คือ

Analyze → Descriptive Statistics → Crosstabs ปุ่ม statistic เลือก Phi Cramer's V

คำสั่งที่เลือก Phi จะได้ค่า Cramer's V ด้วยผลการวิเคราะห์จากโปรแกรมสำเร็จรูปจะได้ผลลัพธ์การวิเคราะห์ความสัมพันธ์ด้วย สัมประสิทธิ์ Phi ที่ได้จากโปรแกรมสำเร็จรูป SPSS มี 2 ตาราง ทั้งค่าสถิติพรรณนาเป็นความถี่ และร้อยละในตารางแรกและสถิติ Phi Coefficient ตัวอย่างการนำเสนอผลการวิเคราะห์ข้อมูล

ตารางที่ 3 ความสัมพันธ์ระหว่างเพศของเยาวชนกับการรับประทานอาหารเช้า

ตัวแปร		ความคิดเห็น		รวม	Phi	p-value
		รับประทาน อาหารเช้า	ไม่รับประทาน อาหารเช้า			
ชาย	จำนวน	68	57	125	.307	.000
	%	54.4%	45.6%	100.0%		
หญิง	จำนวน	124	26	150		
	%	82.7%	17.3%	100.0%		
	จำนวน	192	83	275		
	%	69.8%	30.2%	100.0%		

จากตารางที่ 3 นิสิตชายส่วนใหญ่ร้อยละ 54.4 รับประทานอาหารเช้า ส่วนนิสิตหญิงส่วนใหญ่ร้อยละ 82.7 รับประทานอาหารเช้า เมื่อทดสอบค่า $\Phi=0.307$ มีนัยสำคัญทางสถิติ ($\text{Sig}=.000$) เป็นการปฏิเสธ H_0 แปลผลได้ว่า ตัวแปรทั้งสองมีความสัมพันธ์กัน นั่นคือเพศของเยาวชนกับการรับประทานอาหารเช้ามีความสัมพันธ์เกี่ยวข้องกัน ที่มีขนาดความสัมพันธ์เท่ากับ 0.307 โดยนิสิตหญิงมีร้อยละของการรับประทานอาหารเช้ามากกว่านิสิตชาย

4. สถิติประเภทความสัมพันธ์ไคสแควร์

สถิติที่วิเคราะห์ความสัมพันธ์ ประเภทไคสแควร์ (Chi Square: χ^2) เป็นการทดสอบความเป็นอิสระต่อกันของสองตัวแปร การเลือกใช้สถิตินี้ใช้เมื่อข้อมูลลักษณะเป็น Category data หรือตัวแปรตัวหนึ่งมีระดับการวัดนามบัญญัติ แต่ละตัวแปรจำแนกตั้งแต่ 2 กลุ่มขึ้นไป เช่น

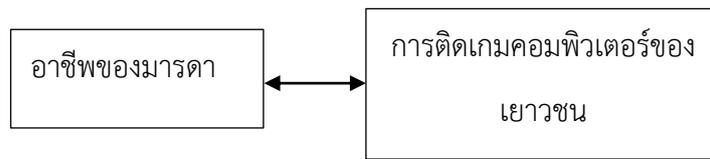
การซื้อโฆษณา ข้อมูลได้แก่ ซื้อ ไม่ซื้อ

อาชีพ ข้อมูลได้แก่ ราชการ พนักงานบริษัท รับจ้าง ค้าขาย ฯลฯ

ประเภทของสื่อมวลชน ข้อมูลได้แก่ หนังสือพิมพ์ โทรทัศน์ วิทยุ ฯลฯ

การพิสูจน์ว่าตัวแปรแรกมีความสัมพันธ์เกี่ยวข้องกันกับตัวแปรที่สอง การทดสอบข้อมูลด้วยการทำเป็นตาราง Contingency Table ตัวแปรหนึ่งเป็นแถว อีก ตัวแปรเป็นคอลัมน์

ตัวอย่างที่ 2 การศึกษาความสัมพันธ์ระหว่างอาชีพของมารดากับการติดเกมคอมพิวเตอร์ของเยาวชนโดยสำรวจกับกลุ่มตัวอย่างเยาวชนจากชุมชนแออัดต่าง ๆ จำนวน 110 คน การทดสอบสมมติฐานความสัมพันธ์และสรุปผล ณ ระดับนัยสำคัญทางสถิติ .05 สมมติฐานการวิจัย อาชีพของมารดากับการติดเกมคอมพิวเตอร์ของเยาวชนมีความสัมพันธ์กัน



ภาพที่ 4 กรอบแนวคิดในการวิจัยตัวอย่างที่ 2

จากสมมติฐานการวิจัย มากำหนดสมมติฐานทางสถิติได้ ดังนี้

H₀: อาชีพของมารดากับการติดเกมคอมพิวเตอร์ของเยาวชนไม่มีความสัมพันธ์กัน

H₁: อาชีพของมารดากับการติดเกมคอมพิวเตอร์ของเยาวชนมีความสัมพันธ์กัน

การวิเคราะห์ด้วยโปรแกรมสำเร็จรูป SPSS คำสั่งที่สำคัญ คือ

Analyze → Descriptive Statistics → Crosstabs ปุ่ม statistic เลือก Chi-square

ผลการวิเคราะห์จากโปรแกรมสำเร็จรูปจะได้ทั้งค่าสถิติพรรณนาและสถิติการทดสอบ χ^2

ผลการวิเคราะห์ด้วยโปรแกรมสำเร็จรูป เป็น 2 ตาราง ตารางแรก Crosstabulation เป็นผลการวิเคราะห์สถิติพรรณนาที่เป็นความถี่และร้อยละ และตารางที่สองคือ Chi-Square Tests ตัวอย่างการนำผลการวิเคราะห์มานำเสนอในรายงานวิจัย

ตารางที่ 4 ความสัมพันธ์ระหว่างอาชีพของมารดากับการติดเกมคอมพิวเตอร์ของเยาวชน

อาชีพของมารดา		ความคิดเห็น		รวม	χ^2	p-value
		ไม่การติดเกม	การติดเกม			
รับราชการ/ รัฐวิสาหกิจ	จำนวน	8	22	30	1.12	0.325
	ร้อยละ	26.7%	73.3%	100.0%		
พนักงานบริษัท	จำนวน	12	20	32		
	ร้อยละ	37.5%	62.5%	100.0%		
รับจ้าง	จำนวน	18	30	48		
	ร้อยละ	37.5%	62.5%	100.0%		
รวม	จำนวน	38	72	110		
	ร้อยละ	34.5%	65.5%	100.0%		

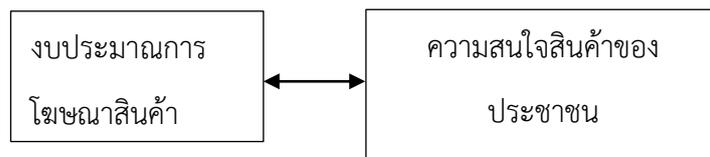
จากตารางพบว่า มารดาที่มีอาชีพรับราชการ/วิสาหกิจ อาชีพพนักงานบริษัทและรับจ้างมีความคิดเห็นว่า บุตรหลานติดเกมคอมพิวเตอร์จำนวนร้อยละ 73.3, 62.5 และ 62.5 ตามลำดับ เมื่อหาความสัมพันธ์ระหว่างอาชีพของมารดากับการติดเกมคอมพิวเตอร์ของเยาวชนที่ พบว่า ค่าไคสแควร์

เท่ากับ 1.12 หรือไม่มีนัยสำคัญทางสถิติ Sig. (2-tailed) เท่ากับ .325 นั่นคือ อาชีพของมารดากับการติดเกมคอมพิวเตอร์ของเยาวชนไม่มีความสัมพันธ์กัน แผลผลตามข้อมูลได้ว่า อาชีพของมารดาทุกอาชีพมีความคิดเห็นเกี่ยวกับการติดเกมคอมพิวเตอร์ของเยาวชนไม่ต่างกัน โดยส่วนใหญ่บอกว่าเยาวชนติดเกมคอมพิวเตอร์

5. สถิติประเภทความสัมพันธ์ Spearman Correlation

การเลือกใช้สถิติสัมประสิทธิ์สเปียร์แมน (Spearman's Rank Coefficient: ρ) เมื่อตัวแปรทั้งสองที่นำมาหาความสัมพันธ์มี ระดับการวัดแบบเรียงอันดับ (Ordinal Scale) เช่น การจัดอันดับความสามารถของนักเรียนโดยครู 2 คน ผลการสอบกลางภาคกับปลายภาคของบุคคลเดียวกัน เป็นต้น

ตัวอย่างที่ 3 งานวิจัยเรื่อง ความสัมพันธ์ระหว่างงบประมาณการโฆษณาสินค้ากับความสนใจสินค้าของประชาชน ผู้วิจัยจัดกลุ่มงบประมาณการโฆษณาสินค้าเป็น 4 กลุ่ม คือ งบประมาณน้อยกว่า 1 ล้านบาท 1-10 ล้านบาท และมากกว่า 10 ล้านบาท และสำรวจถามระดับความสนใจสินค้ากับกลุ่มตัวอย่างผู้บริโภคคือ 89 คน



ภาพที่ 5 กรอบแนวคิดในการวิจัยตัวอย่างที่ 3

สมมติฐานการทางสถิติ

H_0 : งบประมาณการโฆษณาสินค้ากับความสนใจสินค้าของประชาชนไม่มีสัมพันธ์กัน

H_1 : งบประมาณการโฆษณาสินค้ากับความสนใจสินค้าของประชาชนสัมพันธ์กัน

การวิเคราะห์ด้วยโปรแกรมสำเร็จรูป SPSS มีให้เลือกได้ 2 คำสั่ง

คำสั่งที่ 1 กรณีมีตัวแปรตั้งแต่ 2 ตัวแปรขึ้นไป จะได้ผลวิเคราะห์เฉพาะค่าความสัมพันธ์คือ

Analyze → Correlation → Bivariate เลือก Spearman

คำสั่งที่ 2 กรณีมีตัวแปร 2 ตัวแปร จะได้ผลวิเคราะห์ความสัมพันธ์และร้อยละแต่ละระดับคือ

Analyze → Descriptive → Crosstabs ที่ปุ่ม statistic เลือก Spearman

ผลการวิเคราะห์ด้วยโปรแกรมสำเร็จรูป เป็น 2 ตาราง ตารางแรก Crosstabulation เป็นผลการวิเคราะห์สถิติพรรณนาที่เป็นความถี่และร้อยละ และตารางที่สอง คือ Spearman's Rho เป็นสถิติอ้างอิงเพื่อใช้อ้างอิงไปยังประชากร ตัวอย่างการนำผลการวิเคราะห์มานำเสนอในรายงานวิจัย

ตารางที่ 5 ความสัมพันธ์ระหว่างงบประมาณการโฆษณาสินค้ากับความสนใจสินค้าของประชาชน

ระดับความสนใจ		งบประมาณ			รวม	Spearman's Rho	p-value
		น้อยกว่า 1 ล้านบาท	1-10 ล้านบาท	มากกว่า 10 ล้านบาท			
น้อย	จำนวน	18	3	1	22	.600**	.000
	ร้อยละ	81.8%	13.6%	4.5%	100.0%		
ปานกลาง	จำนวน	23	110	30	163		
	ร้อยละ	14.1%	67.5%	18.4%	100.0%		
มาก	จำนวน	1	25	64	90		
	ร้อยละ	1.1%	27.8%	71.1%	100.0%		
รวม	จำนวน	42	138	95	275		
	ร้อยละ	15.3%	50.2%	34.5%	100.0%		

** มีนัยสำคัญทางสถิติที่ระดับ .01

จากตารางที่ 5 ผลทดสอบความสัมพันธ์พบว่างบประมาณการโฆษณาสินค้ามีความสัมพันธ์กับระดับความสนใจสินค้าของประชาชนในทางบวก ค่าความสัมพันธ์สเปียร์แมนเท่ากับ .600 มีนัยสำคัญทางสถิติที่ระดับ .01 หากพิจารณาแต่ละกลุ่มงบประมาณการโฆษณาสินค้า พบว่า กลุ่มงบประมาณน้อยกว่า 1 ล้านบาท ประชาชนส่วนใหญ่ร้อยละ 81.8 มีความสนใจสินค้าระดับน้อย กลุ่มงบประมาณ 11-20 ล้านบาท ประชาชนส่วนใหญ่ร้อยละ 67.5 มีความสนใจสินค้าระดับปานกลาง และกลุ่มงบประมาณการโฆษณาสินค้ามากกว่า 10 ล้านบาท ประชาชนส่วนใหญ่ ร้อยละ 71.1 มีความสนใจสินค้าระดับมาก

6. สถิติประเภทความสัมพันธ์ Pearson Correlation

การเลือกใช้สถิติ สหสัมพันธ์เพียร์สัน (Pearson's Product Moment Correlation: r) เมื่อตัวแปรทั้งสองที่มาหาความสัมพันธ์เป็นตัวแปรต่อเนื่อง ที่มีระดับการวัดแบบอันตรภาคหรืออัตราส่วน (Interval or Ratio Scale) และที่ต่างจากสถิติความสัมพันธ์ประเภทอื่น ๆ คือ การแจกแจงของ

ข้อมูลของทั้ง 2 ตัวแปรต้องมีการแจกแจงปกติเช่น แรงจูงใจกับคะแนนการสอบ ค่าสัมประสิทธิ์ r มีค่าอยู่ระหว่าง -1.00 ถึง $+1.00$

ข้อตกลงเบื้องต้นของการวิเคราะห์สหสัมพันธ์แบบเพียร์สัน ดังนี้

1. ข้อมูลที่ได้จากตัวแปร X และ Y อยู่ในระดับการวัดแบบอันดับหรืออัตราส่วน
2. การแจกแจงของข้อมูลที่ได้จากการวัดตัวแปร X และ Y มีการแจกแจงปกติ

การแปลผลการวิเคราะห์ความสัมพันธ์

การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรว่า มีความเกี่ยวข้องกันหรือไม่ เกี่ยวข้องกันมากน้อยเพียงใด ไปในทิศทางเดียวกันหรือไม่ ผลการวิเคราะห์ความสัมพันธ์จะให้คำตอบเป็น 3 ลักษณะ คือ

1. ตัวแปรมีความสัมพันธ์กันหรือไม่ การที่ตัวแปรมีความสัมพันธ์กันแสดงว่ามีความเกี่ยวข้องกัน ตัวแปรหนึ่งเท่านั้น (แต่ไม่ได้แปลความหมายว่าเป็นสาเหตุของอีกตัวแปรหนึ่ง) เมื่อทดสอบสมมติฐานพบว่ามีความสำคัญทางสถิติ แสดงว่า ตัวแปรทั้งสองมีความสัมพันธ์กัน

2. สัมพันธ์กันในทิศทางใด พิจารณาจากเครื่องหมายของค่าสัมประสิทธิ์สหสัมพันธ์ ถ้าค่าเป็นบวกแสดงว่าตัวแปรทั้งสองเปลี่ยนแปลงไปในทิศทางเดียวกัน คือ เมื่อตัวแปรหนึ่งมีค่าสูง อีกตัวแปรก็มีค่าสูง หากมีค่าต่ำ อีกตัวแปรก็มีค่าต่ำด้วย ส่วนค่าความสัมพันธ์เป็นลบ แสดงว่า ตัวแปรทั้งสองมีการเปลี่ยนแปลงไปในทิศทางกลับตรงกันข้าม คือ เมื่อตัวแปรหนึ่งมีค่าสูง อีกตัวแปรก็มีค่าต่ำกลับทิศทางกัน

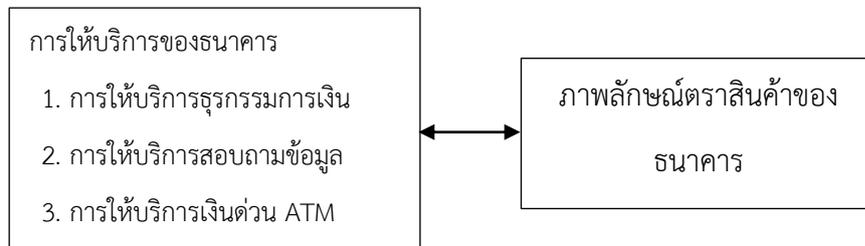
3. สัมพันธ์กันขนาดเท่าใด พิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน ตัวเลขเป็นการบอกค่าความสัมพันธ์ที่ค่ายิ่งมากใกล้ 1.0 แสดงว่ายิ่งมีความสัมพันธ์กันสูงค่าที่เข้าใกล้ -1.00 หรือ $+1.00$ เช่น 0.9 หรือ -0.9 แสดงว่ามีความสัมพันธ์สูงมาก ค่าเป็น 0 หมายถึงทั้งสองตัวแปรไม่มีความสัมพันธ์ต่อกัน

การหาความสัมพันธ์ระหว่างตัวแปรที่เก็บข้อมูลจากกลุ่มตัวอย่าง จะแปลความหมายขนาดของความสัมพันธ์ได้ต่อเมื่อพบว่าทดสอบด้วยสถิติแล้วพบว่ามีความสำคัญทางสถิติเท่านั้น

ความหมายของขนาดของความสัมพันธ์ มีผู้ให้ความหมายต่าง ๆ กัน ในที่นี้เสนอเกณฑ์การแปลผล ขนาดของความสัมพันธ์บาร์ทซ์ (Bartz, 1999 อ้างถึงใน สุขมาส อังศุโชติ, 2556) ที่นิยมใช้ดังนี้

ค่าสัมประสิทธิ์สหสัมพันธ์		ความหมาย
ค่าสหสัมพันธ์ทางบวก	ค่าสหสัมพันธ์ทางลบ	
0.81 ถึง 1.0	-0.81 ถึง -1.0	สูงมาก
0.61 ถึง 0.80	-0.61 ถึง -0.80	สูง
0.41 ถึง 0.60	-0.41 ถึง -0.60	ปานกลาง
0.21 ถึง 0.40	-0.21 ถึง -0.40	ต่ำ
0.0 ถึง 0.20	0.0 ถึง -0.20	ต่ำมาก

ตัวอย่างที่ 4 การศึกษาความสัมพันธ์ระหว่างการให้บริการกับภาพลักษณ์ตราสินค้าของธนาคาร การหาความสัมพันธ์ตัวแปรการเข้ารับอบรมฯ จึงจำแนกเป็น 3 ด้านการให้บริการธุรกรรม การเงิน การให้บริการสอบถามข้อมูล การให้บริการเงินด่วน ATM ดังตัวอย่างในตาราง



ภาพที่ 6 กรอบแนวคิดในการวิจัยตัวอย่างที่ 4

สมมติฐานการทางสถิติ

H_0 : การให้บริการกับภาพลักษณ์ตราสินค้าของธนาคาร ไม่มีความสัมพันธ์กัน

H_1 : การให้บริการกับภาพลักษณ์ตราสินค้าของธนาคาร มีความสัมพันธ์กัน

การวิเคราะห์ด้วยโปรแกรมสำเร็จรูป SPSS คำสั่งที่สำคัญ คือ

```
Analyze → Correlation → Bivariate เลือก Pearson
```

ผลการวิเคราะห์จากโปรแกรมสำเร็จรูป จะได้ค่าความสัมพันธ์แบบ Pearson ตัวอย่างการนำผลการวิเคราะห์มานำเสนอในรายงานวิจัย

ตารางที่ 6 ความสัมพันธ์ระหว่างการให้บริการกับภาพลักษณ์ตราสินค้าของธนาคาร

ความสัมพันธ์	r	P-value	ระดับความสัมพันธ์
การให้บริการธุรกรรมการเงินกับ ภาพลักษณ์ตราสินค้า	0.53**	0.00	ปานกลาง
การให้บริการสอบถามข้อมูลกับ ภาพลักษณ์ตราสินค้า	0.55**	0.00	ปานกลาง
การให้บริการเงินด่วน ATM กับ ภาพลักษณ์ตราสินค้า	0.72**	0.00	สูง

** มีนัยสำคัญทางสถิติที่ระดับ .01

การแปลผล การให้บริการทั้งสามด้าน พบว่า ทุกด้านมีความสัมพันธ์ทางบวกกับภาพลักษณ์ตราสินค้าของธนาคารอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 เรียงลำดับความสัมพันธ์จากมากไปน้อย ดังนี้ การให้บริการเงินด่วน ATM มีความสัมพันธ์กับภาพลักษณ์ตราสินค้าของธนาคารในระดับสูง ($r=0.72$) รองลงมาคือ การให้บริการธุรกรรมการเงินและการให้บริการสอบถามข้อมูลมีความสัมพันธ์กับภาพลักษณ์ตราสินค้าของธนาคารในระดับปานกลาง ($r=0.53$ และ 0.55)

7. ข้อผิดพลาดที่ควรระวังในการใช้ค่าสหสัมพันธ์

การอธิบายผลของค่าสหสัมพันธ์สหสัมพันธ์บางครั้งพบว่าไม่ถูกต้องหรือเข้าใจคลาดเคลื่อนที่ต้องแก้ไข จึงขอแนะนำข้อผิดพลาดที่พบเพื่อเป็นแนวทางในการใช้ค่าสหสัมพันธ์ ดังนี้

1. ค่าความสัมพันธ์ไม่ใช่บอกความเป็นเหตุเป็นผล การที่ตัวแปรมีความสัมพันธ์กันแสดงว่า มีความเกี่ยวข้องกัน ไม่สามารถกล่าวได้ว่าตัวแปรหนึ่งเป็นสาเหตุของอีกตัวแปรหนึ่ง อาทิเช่น ยอดขายคอนโดกับโทรศัพท์มือถือ มีแนวโน้มไปในทิศทางเดียวกัน ไม่ได้หมายความว่า ยอดขายโทรศัพท์มือถือเกิดจากยอดขายคอนโด ในทางกลับกัน ยอดขายคอนโดก็ไม่ได้เกิดจากยอดขายมือถือโทรศัพท์ แต่ทั้งสองเป็นผลมาจาก Spurious Correlation เช่น ปัจจัยเศรษฐกิจ เป็นต้น

2. ค่าความสัมพันธ์ไม่มีหน่วย ตัวแปรที่นำมาวิเคราะห์ความสัมพันธ์อาจมีหน่วย เช่น งบประมาณการโฆษณา หน่วยเป็นบาท ความนิยมของประชาชน หน่วยเป็นจำนวนคน ผลการสอบ หน่วยเป็นคะแนน เมื่อนำมาหาความสัมพันธ์ เป็นค่ามาตรฐานจึงไม่มีหน่วย

3. ค่าความสัมพันธ์ไม่แปลความหมายเป็นเปอร์เซ็นต์ เช่น คนที่ให้ความหมายความสัมพันธ์ระหว่าง 2 ตัวแปรเท่ากับ .60 คือมีความสัมพันธ์ 60 เปอร์เซ็นต์ซึ่งไม่ถูกต้อง หรือมีความสัมพันธ์เป็น 2 เท่าของความสัมพันธ์ 0.30 ไม่ถูกต้องเช่นกัน แต่หมายความว่า ความสัมพันธ์ระหว่าง 2 ตัวแปรเท่ากับ .60 หมายความว่า สองตัวแปรมีความสัมพันธ์ในระดับปานกลาง

4. การวิเคราะห์ความสัมพันธ์ของข้อมูลที่เก็บรวบรวมจาก **ประชากร** สามารถแปลผลค่าความสัมพันธ์ได้โดยไม่ต้องทดสอบค่านัยสำคัญ สำหรับค่าสหสัมพันธ์สหสัมพันธ์ Spearman และค่าสหสัมพันธ์สหสัมพันธ์ Pearson สามารถแปลผลตามเกณฑ์ของ Bartz (1999) ได้ **โดยไม่ต้องใช้** ค่านัยสำคัญทางสถิติ

แต่หากเก็บข้อมูลจาก **กลุ่มตัวอย่าง** จะต้องมีการทดสอบค่าสหสัมพันธ์สหสัมพันธ์ ถ้าพบว่ามีนัยสำคัญทางสถิติแสดงว่า ตัวแปรมีความสัมพันธ์กันจะสามารถใช้เกณฑ์ของ Bartz (1999) สำหรับการแปลค่าสหสัมพันธ์สหสัมพันธ์ หากไม่มีนัยสำคัญทางสถิติ ไม่สามารถแปลผลค่าสหสัมพันธ์สหสัมพันธ์ได้ เพราะตัวแปรไม่มีความสัมพันธ์กัน (สมถวิล วิจิตรวรรณ และ สุนิสา จุ้ยม่วงศรี, 2559)

5. ค่าสัมประสิทธิ์สหสัมพันธ์ Spearman และค่าสัมประสิทธิ์สหสัมพันธ์ Pearson ที่มีค่าความสัมพันธ์เป็นศูนย์หรือค่าระดับต่ำ บางครั้งไม่ได้หมายความว่าตัวแปรทั้งสองตัวไม่มีความสัมพันธ์กัน เนื่องจากเป็นค่าที่คำนวณจากสูตรลักษณะความสัมพันธ์เชิงเส้น (Linear Correlation) ดังนั้นการที่ตัวแปรที่เรามาหาความสัมพันธ์อาจมีรูปแบบความสัมพันธ์แบบ s curve (หัวและท้ายของการกระจาย ขนานกับแกน X) หรือมีรูปแบบตัว U เราสามารถตรวจสอบด้วยการพล็อตจุดหรือทำ scatter plot และพิจารณาค่าที่อยู่นอก Outlines และพิจารณาให้รอบคอบในการ**ลบค่าผิดปกติออกก่อนคำนวณใหม่** เนื่องจากค่าสัมประสิทธิ์สหสัมพันธ์ มีความอ่อนไหวช่วงที่มีค่าสูงสุดหรือต่ำสุดค่อนข้างมาก

6. การเลือกใช้สถิติความสัมพันธ์ที่มีมาก บางสถิติมีได้นำมากล่าว ได้แก่ Kendall's Tau ใช้กับตัวแปรที่มีระดับเรียงอันดับ ส่วน Biserial ใช้กับตัวแปรนามบัญญัติที่แบ่งเป็น 2 กลุ่ม กับตัวแปรต่อเนื่อง (อันตรภาคหรืออัตราส่วน) เช่น ความสัมพันธ์ระหว่างเพศของผู้สอบกับคะแนนสอบ แต่หากตัวแปรที่แบ่ง 2 กลุ่มมีความต่อเนื่อง เช่น ความสัมพันธ์ระหว่าง IQ ของผู้สอบกับคะแนนสอบ ที่แบ่ง IQ เป็น 2 กลุ่มเป็นกลุ่มคะแนน IQ สูงและต่ำ ลักษณะนี้ต้องใช้ Point Biserial เป็นต้น

สรุปได้ว่า สถิติความสัมพันธ์นำไปใช้ในอธิบายความเกี่ยวข้องของตัวแปร การเลือกใช้สถิติความสัมพันธ์ต้องคำนึงถึงระดับการวัดของตัวแปรเป็นหลักและเงื่อนไขเฉพาะของสถิตินั้น ๆ การแปลผลของสถิติไม่เหมือนกันจึงต้องทำความเข้าใจ สถิติความสัมพันธ์มีความสำคัญนำไปใช้ในหลายด้าน เช่น การวิเคราะห์คุณภาพเครื่องมือวิจัย รวมถึงเป็นสถิติที่สำคัญที่มีการขยายต่อยอดไปเป็นสถิติขั้นสูงที่ตอบคำถามการวิจัยที่ซับซ้อนขึ้น เช่น การทำนายที่ได้การวิเคราะห์ความสัมพันธ์เชิงสาเหตุ การศึกษาพัฒนาการ การวิเคราะห์โครงสร้างของคุณลักษณะ เป็นต้น

เอกสารอ้างอิง

- นงลักษณ์ วิรัชชัย. (2548). **สถิติขั้นสูง**. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- สมถวิล วิจิตรวรรณ และ สุนิสา จัยม่วงศรี. (2559). **การวัดตัวแปรและการเลือกใช้สถิติ**. หน่วยที่ 7 ในเอกสารการสอน ชุดวิชาวิธีการวิจัยและเทคโนโลยีสารสนเทศสำหรับเศรษฐศาสตร์ นนทบุรี: มหาวิทยาลัยสุโขทัยธรรมาธิราช สาขาวิชาเศรษฐศาสตร์.
- สุชาติดา บวรกิตติวงศ์. (2548). **สถิติประยุกต์ทางพฤติกรรมศาสตร์**. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- สุภมาส อังสุโชติ. (2556). “การวิเคราะห์ข้อมูลวิจัยเพื่อพัฒนาการเรียนการสอน”. หน่วยที่ 6 ในเอกสารการวิจัยเพื่อพัฒนาการเรียนการสอน กรุงเทพฯ: เจริญดีมั่นคงการพิมพ์.
- Bartz, A.E. (1999). **Basics Statistical Concepts**. 4th ed. New Jersey: Prentice Hall.



C. Muller. (1973). **Introduction to methods of linguistic statistics**. Retrieved on October, 1st, 2020, from <https://wikilean.com/en/different-types-of-data>