

บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ในส่วนของวรรณคดีที่เกี่ยวข้อง นำเสนอเป็น 6 ตอน คือ

- ตอนที่ 1 ประเภทของการวิจัยทางการศึกษา
- ตอนที่ 2 ประเภทของปัญหาวิจัย
- ตอนที่ 3 มาตรฐานประมาศ่า
- ตอนที่ 4 มาตรฐานการวัด
- ตอนที่ 5 การเลือกใช้เทคนิคทางสถิติเพื่อวิเคราะห์ข้อมูล
- ตอนที่ 6 บทความที่เกี่ยวข้อง

ตอนที่ 1 ประเภทของการวิจัยทางการศึกษา

ประเภทของการวิจัยทางการศึกษาแบ่งออกเป็น 3 ชนิดใหญ่ ๆ

(John W Best ,1981; Walter R borg และ M K Gall,1979 อ้างถึงใน บุญเรียง ขจรศิลป์, 2533) โดยยึดลักษณะวิธีวิจัยเป็นหลักในการแบ่ง ดังนี้

1. การวิจัยเชิงประวัติศาสตร์ (Historical Research) เป็นการวิจัยที่เน้นถึงการศึกษา ค้นคว้า รวบรวมข้อมูล หรือเหตุการณ์ต่าง ๆ ที่เกิดขึ้นมาแล้วในอดีต (What was?) ประโยชน์ของการวิจัยชนิดนี้ก็คือจากผลการวิจัยที่ได้จากการวิจัยเชิงประวัติศาสตร์ ซึ่งเป็นข้อสรุปจากข้อมูลหรือข้อเท็จจริงต่าง ๆ ที่เกิดขึ้นมาแล้วในอดีต สามารถนำมาใช้เป็นแนวทางการศึกษาเหตุการณ์ต่าง ๆ ในปัจจุบัน หรือสามารถนำมาใช้ประกอบการตัดสินใจเพื่อแก้ปัญหาต่าง ๆ ที่เกิดขึ้นในปัจจุบันได้ด้วย นอกจากนั้น ผลจากการวิจัยเชิงประวัติศาสตร์ยังทำให้ทราบถึงแนวโน้มของเหตุการณ์ต่าง ๆ

ที่เกิดขึ้นมาแล้ว การรวบรวมข้อมูลการวิจัยเชิงประวัติศาสตร์มักจะรวบรวมจากเอกสารหรือหลักฐานต่าง ๆ ที่เกิดขึ้นในอดีต ตัวอย่างของการวิจัยเชิงประวัติศาสตร์ได้แก่ การศึกษาแนวโน้มของปริมาณนักศึกษาที่มาใช้บริการการศึกษาผู้ใหญ่ในช่วง 10 ปี ที่ผ่านมา วิทยาการของเครื่องปั้นดินเผาตั้งแต่สมัยสุรพัตย์จนถึงปัจจุบัน ฯลฯ

2. การวิจัยเชิงบรรยาย (Descriptive Research) เป็นการวิจัยที่เน้นถึงการศึกษารวบรวมข้อมูลต่าง ๆ ที่เกิดขึ้นในปัจจุบัน (What is?) ในการดำเนินงานการวิจัย นักวิจัยไม่สามารถที่จะไปจัดสร้างสถานการณ์หรือควบคุมตัวแปรต่าง ๆ ได้ตามใจชอบ การวิจัยแบบนี้เป็นการค้นหาข้อเท็จจริงหรือเหตุการณ์ต่าง ๆ ที่เกิดขึ้นอยู่แล้ว เช่น นักวิจัยจะต้องการศึกษาความสัมพันธ์ระหว่างเพศและความสนใจต่อการเมือง ซึ่งตัวแปรเพศและความสนใจต่อการเมืองนั้นเป็นสิ่งที่มียู่แล้ว นักวิจัยเพียงแต่ไปรวบรวมข้อมูลมาเท่านั้น ไม่สามารถที่จะไปจัดการเปลี่ยนแปลงค่าต่าง ๆ ของตัวแปรอิสระได้ตามชอบใจ มีการวิจัยหลายชนิดที่จัดได้ว่าเป็นการวิจัยเชิงบรรยาย ได้แก่

(1) การวิจัยเชิงสำรวจ (Survey Research) เป็นการวิจัยที่มุ่งจะศึกษาถึงลักษณะความเป็นจริงตามสถานการณ์ต่าง ๆ ที่เป็นอยู่ ตลอดจนพิจารณาถึงความเหมาะสมของสถานภาพ โดยนำไปเปรียบเทียบกับสถานภาพที่เป็นมาตรฐานเพื่อนำผลการวิจัยไปปรับปรุงสถานภาพเดิมให้ดีขึ้น ขอบเขตของการวิจัยเชิงสำรวจนั้นอาจจะกว้างหรือแคบก็ได้แล้วแต่ขอบเขตของปัญหาในการวิจัย การวิจัยเชิงสำรวจนั้นอาจจะแยกย่อยได้หลายประเภทตามลักษณะของสิ่งที่จะศึกษา เช่น การสำรวจเจตคติ การสำรวจโรงเรียน การสำรวจชุมชน การสำรวจประชามติ การวิเคราะห์เอกสาร การสำรวจปัญหาสังคม เป็นต้น การรวบรวมข้อมูลสำหรับการวิจัยเชิงสำรวจนั้น ส่วนใหญ่จะใช้แบบสอบถาม ในบางครั้งใช้การสัมภาษณ์ หรือใช้ทั้งการส่งแบบสอบถามและการสัมภาษณ์ประกอบกัน การส่งแบบสอบถามส่วนใหญ่ใช้รวบรวมข้อมูลในกรณีที่กลุ่มตัวอย่างมีจำนวนมาก ส่วนการสัมภาษณ์นั้นใช้รวบรวมข้อมูลในกรณีที่จำนวนสมาชิกในกลุ่มตัวอย่างไม่มากนักและต้องการรายละเอียดมากกว่าการส่งแบบสอบถาม ในกรณีที่ใช้ทั้งสองอย่างประกอบกันนั้นจะใช้การสัมภาษณ์กับตัวอย่างที่ไม่ได้ตอบแบบสอบถามที่ส่งไปรอบแรก หรือใช้ในการรวบรวมข้อมูลเพิ่มเติม

(2) การวิจัยเชิงสังเกต (Observational Research) เป็นการวิจัยที่ใช้วิธีสังเกตในการรวบรวมข้อมูล งานวิจัยประเภทนี้ให้ผลดีในแง่ที่สามารถรวบรวมข้อมูลได้โดยตรงจากต้นตอซึ่งสามารถใช้ได้อย่างมีประสิทธิภาพ ในกรณีที่นักวิจัยต้องการจะศึกษาพฤติกรรมต่าง ๆ ได้โดยละเอียด เช่น การศึกษาว่าวิธีการสอนแบบใดที่จะมีประสิทธิภาพ ในการสอนระบบเลขจำนวนให้กับเด็กปฐมวัยอ่อน ปฏิกริยาของเด็กอายุ 5 - 12 ปี ที่มีต่อรายการโทรทัศน์ที่เกี่ยวกับความรุนแรงหรือการกระทบกระเทือนใจ อย่างไรก็ตาม การวิจัยแบบนี้ผู้สังเกตจะต้องระวังไม่ให้เกิดความลำเอียง ความคลาดเคลื่อนที่อาจจะเกิดขึ้นแล้วทำให้ลดความตรงของการวิจัยแบบนี้ก็คือ การให้คำจำกัดความของพฤติกรรมที่จะสังเกตกว้างมากเกินไป ควรจะเฉพาะเจาะจงให้ชัดเจน การรวบรวมข้อมูลโดยใช้เทคนิคของการสังเกตนี้ ถ้าจะทำให้ผลเที่ยงตรงและมีความเชื่อถือได้ควรจะมีการฝึกผู้สังเกตก่อนที่จะไปทำการสังเกต

(3) การวิจัยเชิงเหตุผลเปรียบเทียบ (Causal Comparative Research) เป็นการวิจัยที่ต้องการศึกษาเปรียบเทียบ และสรุปความเป็นเหตุเป็นผลซึ่งกันและกัน แต่ในทางปฏิบัติไม่สามารถที่จะทำการวิจัยเชิงการทดลองได้ เป็นการศึกษาจากผลไปหาเหตุ เช่นนักวิจัยต้องการที่จะศึกษาถึงผลที่จะเกิดขึ้นตามมา ในกรณีอย่างนี้มีนักวิจัยเชิงทดลองบางคนเลี้ยงไปศึกษากับสัตว์แทน เช่นอาจจะทำการทดลองกับหนู เพราะถือว่าหนูเป็นสัตว์เลี้ยงลูกด้วยนมเช่นเดียวกับคน โดยเริ่มต้นจากการสุ่มตัวอย่างหนูมากลุ่มหนึ่งวัดความสามารถในการแก้ปัญหาของหนูเหล่านั้น เช่นวัดจากระยะเวลาในการที่หนูสามารถค้นหาทางออกจากกรงกล ขึ้นต่อไปแบ่งเป็นสองกลุ่ม โดยการสุ่มกลุ่มหนึ่งจับมาทำลายระบบสมองบางส่วนและวัดความสามารถในการแก้ปัญหาของหนูทั้งสองกลุ่มนั้นอีกครั้งเพื่อที่จะเปรียบเทียบดูว่า หลังจากสมองถูกทำลายบางส่วนแล้วความสามารถในการแก้ปัญหาของหนูทั้งสองกลุ่มนั้นจะต่างกันหรือไม่ ซึ่งผลของการวิจัยเชิงทดลองในลักษณะนี้ต้องระวังอย่างมากในการที่จะอ้างอิงไปใช้กับมนุษย์ ถึงแม้ว่าหนูจะเป็นสัตว์ที่เลี้ยงลูกด้วยนมเช่นเดียวกับมนุษย์ แต่ยังมีสิ่งต่าง ๆ อีกมากมายที่แตกต่างไปจากมนุษย์ ในกรณีแบบนี้การวิจัยเชิงเหตุผลจะเหมาะสมมากกว่า โดยที่นักวิจัยไปขอศึกษาจากทะเบียนประวัติของคนไข้ที่ได้รับอุบัติเหตุกระทบกระเทือนทางสมอง เมื่อได้

จำนวนตัวอย่างพอสมควรแล้วก็หากกลุ่มตัวอย่างอีกกลุ่มหนึ่งที่สมองมิได้รับความกระทบกระเทือนและอยู่ในสภาวะแวดล้อมต่าง ๆ ใกล้เคียงกับกลุ่มแรก ขึ้นต่อไปนักวิจัยศึกษาความสามารถในการเรียนของตัวอย่างทั้งสองกลุ่มในช่วงก่อนที่สมองจะได้รับความกระทบกระเทือนและหลังจากที่สมองได้รับความกระทบกระเทือนแล้ว การวิจัยนี้พอจะสรุปได้ว่าความแตกต่างทั้งหลายที่เกิดขึ้นอาจจะเนื่องมาจากการที่สมองได้รับความกระทบกระเทือนแต่ในขณะที่เดียวกันควรจะศึกษาจากองค์ประกอบอื่น ๆ ด้วยเพื่อให้ได้ข้อสรุปที่เชื่อถือได้มากยิ่งขึ้น งานวิจัยที่เหมาะสมที่จะใช้วิจัยเชิงเหตุผลเปรียบเทียบ ได้แก่ การศึกษาเปรียบเทียบบุคลิกภาพของวัยรุ่นที่มีปัญหาเกี่ยวกับวัยรุ่นที่ไม่มีปัญหาทางสังคม การศึกษาเปรียบเทียบการเป็นแม่แรงที่บอกของคนที่สูงหูรูดกับคนที่ไม่สูง การศึกษาเปรียบเทียบพฤติกรรมก้าวร้าวของเด็กที่ดูภาพยนตร์ทัศนคติที่มีความก้าวร้าวรุนแรง 5 ชั่วโมงต่อวัน กับเด็กที่ดูภาพยนตร์ทัศนคติเดียวกันวันละหนึ่งชั่วโมง การศึกษาเปรียบเทียบพฤติกรรมทางการเรียนของเด็กนักเรียนชั้นประถมศึกษาปีที่ 6 ที่ขาดอาหารกับนักเรียนในระดับเดียวกันที่ไม่ขาดอาหาร แต่อย่างไรก็ตามผลของการวิจัยประเภทนี้ไม่สามารถที่จะใช้ยืนยันได้อย่างแน่นอนว่า ตัวแปรตัวหนึ่งเป็นสาเหตุของตัวแปรอีกตัวหนึ่ง เนื่องมาจากมีตัวแปรเกินที่ควบคุมไม่ได้มากมาย เช่น ผลการวิจัยพบว่านักเรียนที่มีปัญหาทางสังคมมีพฤติกรรมความก้าวร้าวมากกว่านักเรียนที่ไม่มีปัญหาทางสังคม ซึ่งอาจสรุปได้ว่าระดับความก้าวร้าวเป็นสาเหตุของการที่เด็กมีปัญหาทางสังคม และในทางกลับกันก็อาจจะกล่าวได้ว่าเนื่องจากเด็กมีปัญหาทางสังคมจึงมีพฤติกรรมก้าวร้าว ดังนั้นการวิจัยที่ต้องศึกษาความเป็นเหตุเป็นผลควรจะมีการวิจัยเชิงทดลองควบคู่ไปด้วย ซึ่งอาจจะใช้สัตว์ชั้นสูงที่ใกล้เคียงกับมนุษย์เข้าในการทดลอง

(4) การวิจัยเชิงสหสัมพันธ์ (Correlational Research) เป็นงานวิจัยที่ศึกษาถึงความสัมพันธ์ระหว่างตัวแปรตั้งแต่สองตัวขึ้นไป รวมถึงการพยากรณ์ค่าของตัวแปรด้วย การวิจัยเชิงสหสัมพันธ์ต่างจากการวิจัยเชิงเหตุผลเปรียบเทียบในแง่ที่ว่าการศึกษาวิจัยเชิงเหตุผลเปรียบเทียบเน้นที่ความแตกต่าง ส่วนการวิจัยเชิงสหสัมพันธ์เน้นความสัมพันธ์ระหว่างตัวแปร เช่น การศึกษาความสัมพันธ์ระหว่างแรงจูงใจใฝ่สัมฤทธิ์กับการเรียน การศึกษาองค์ประกอบที่สามารถใช้ในการพยากรณ์

ผลสัมฤทธิ์ในการเรียนคณิตศาสตร์

(5) การศึกษาเฉพาะกรณี (Case Study) การศึกษาเฉพาะกรณีเป็นการศึกษาค้นคว้าอย่างละเอียดลึกซึ้งเฉพาะเจาะจงในเรื่องเรื่องหนึ่ง เพื่อหารายละเอียดของเรื่องที่เกี่ยวข้องมากที่สุดเท่าที่จะมากได้ โดยมีวัตถุประสงค์ที่จะทราบธรรมชาติของสภาพการณ์อันสลับซับซ้อนของเรื่อง ตลอดจนทราบถึงเรื่องขององค์ประกอบต่าง ๆ อันเป็นสาเหตุที่ทำให้เกิดสภาพการณ์เช่นนั้น เช่นครูแนะแนวต้องการทราบสาเหตุที่นักเรียนบางคนมีพฤติกรรมก้าวร้าวไม่สนใจต่อการเรียน ครูแนะแนวท่านนั้นควรจะทำการศึกษาเฉพาะกรณี เพื่อที่จะได้ทราบสาเหตุต่าง ๆ อย่างละเอียด

3. การวิจัยเชิงทดลอง (Experimental Research) เป็นการวิจัยที่มุ่งสร้างสถานการณ์ขึ้นแล้วศึกษาผลที่ตามมา (What will be ?) ผู้วิจัยสามารถที่จะกำหนดค่าต่าง ๆ ของตัวแปรอิสระได้ตามต้องการ เพื่อที่จะศึกษาถึงผลที่เกิดขึ้นกับตัวแปรตาม งานวิจัยเชิงทดลองนั้นเป็นที่นิยมใช้กันมากทางด้านวิทยาศาสตร์ สำหรับการศึกษาค้นคว้าอย่างกว้างขวาง ในแง่ของการควบคุมตัวแปรเกินที่จะเข้ามามีบทบาทต่อตัวแปรตามทำให้ผลสรุปของงานวิจัยคลาดเคลื่อนไป ตัวอย่างของการวิจัยเชิงทดลองทางการศึกษา ได้แก่ การศึกษาเปรียบเทียบผลสัมฤทธิ์ในการเรียนคณิตศาสตร์ระหว่างกลุ่มที่สอนโดยวิธีค้นพบกับกลุ่มที่สอนแบบบรรยาย ข้อแตกต่างระหว่างงานวิจัยเชิงทดลองกับงานวิจัยเชิงสหสัมพันธ์และงานวิจัยเชิงเหตุผลเปรียบเทียบอยู่ที่ว่างานวิจัยเชิงทดลองสามารถให้ข้อสรุปที่แสดงถึงความสัมพันธ์ระหว่างตัวแปรในลักษณะความเป็นเหตุเป็นผลได้ ลักษณะที่สำคัญของการวิจัยเชิงทดลองคือ ควบคุมตัวแปรเกินไว้ได้ (Control) จัดการเปลี่ยนแปลงค่าของตัวแปรอิสระได้ (Manipulation) สังเกตได้ (Observation) และทำซ้ำได้ (Replication) การวิจัยเชิงทดลองแบ่งได้เป็น 2 ชนิดดังนี้

(1) การวิจัยเชิงทดลองอย่างแท้จริง (True Experimental Research) เป็นการวิจัยเชิงทดลองที่สามารถที่ควบคุมตัวแปรเกินต่าง ๆ ได้ทั้งหมดเพื่อที่จะวัดผลที่เกิดจากตัวแปรอิสระเท่านั้น ซึ่งในการควบคุมตัวแปรนั้นใช้วิธีจัดสมาชิกเข้ากลุ่มโดยการสุ่ม (Randomly Assignment) ซึ่งการจัดสมาชิกเข้ากลุ่มโดยการสุ่มนั้น Ary และคณะ (อ้างถึงใน Koul, 1984) ระบุว่า "การจัดสมาชิกโดยการสุ่มไม่เพียงแต่

จะกำจัดตัวแปรที่นักวิจัยระบุไว้เท่านั้น แต่ยังสามารถกำจัดตัวแปรเกินอื่น ๆ ที่เกี่ยวข้องกับที่นักวิจัยนึกไม่ถึงออกไปด้วย" เช่นในกรณีที่ต้องการศึกษาเปรียบเทียบผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6 ระหว่างกลุ่มที่สอนโดยวิธีบรรยายและวิธีค้นพบ ซึ่งใช้ผู้สอนคนเดียวกัน ถ้าผู้วิจัยทำการทดลองโดยชักกลุ่มตัวอย่างทั้งหมด 100 คน และจัดสมาชิก 100 คนนี้เข้ากลุ่ม ๆ ละ 50 คนโดยการสุ่ม นั่นคือทุก ๆ คนมีโอกาสเท่า ๆ กัน ที่จะอยูในกลุ่มที่จะสอนโดยวิธีค้นพบหรือกลุ่มที่สอนโดยวิธีบรรยาย การวิจัยลักษณะนี้ถือว่าเป็นการวิจัยเชิงทดลองอย่างแท้จริง ซึ่งในทางปฏิบัติทำได้ลำบากมาก

(2) การวิจัยกึ่งทดลอง(Quasi Experimental Research) เป็นงานวิจัยเชิงทดลองที่ไม่สามารถควบคุมตัวแปรเกินในงานวิจัยได้ เนื่องจากการวิจัยเชิงทดลองบางสถานการณ์โดยเฉพาะการวิจัยทางการศึกษา หรือการวิจัยทางสังคมศาสตร์ ส่วนใหญ่ไม่สามารถที่จะจัดสมาชิกเข้ากลุ่มโดยการสุ่ม ตัวอย่างเช่น การวิจัยในลักษณะดังกล่าวแล้วในหัวข้อ 3.1 ถ้าผู้วิจัยไปขอทำการทดลองสอนตามโรงเรียนซึ่งทางโรงเรียนได้จัดนักเรียนไว้เป็นห้อง ๆ แล้ว และทางโรงเรียนมอบนักเรียนให้ 2 ห้องเรียนในการที่จะทำการทดลองสอน โดยที่ผู้วิจัยไม่สามารถจะนำเด็ก 2 ห้องเรียนนั้นมาจัดใหม่โดยการสุ่ม อย่งดีที่สุดผู้วิจัยทำได้แต่จับฉลากว่าห้องใดจะสอนวิธีไหนเท่านั้น ในลักษณะเช่นนี้ ผู้วิจัยไม่สามารถที่จะกำจัดหรือควบคุมตัวแปรเกินได้ เช่น นักเรียนทั้งสองกลุ่มนั้นอาจจะมีความแตกต่างกันอยู่ตั้งแต่แรกก่อนทำการทดลองซึ่งจะส่งผลทำให้ผลการวิจัยที่สรุปได้นั้นคลาดเคลื่อนจากความเป็นจริง ดังนั้นผู้วิจัยจำเป็นต้องทราบว่ามีตัวแปรเกินอะไรบ้างที่ควบคุมไม่ได้ เพื่อที่จะได้นำมาพิจารณาในการแปลผลการวิจัย การวิจัยกึ่งทดลองนั้นจำเป็นต้องใช้วิธีการทางสถิติเข้ามาช่วยในการวิเคราะห์ข้อมูลเพื่อลดความคลาดเคลื่อนอันเป็นผลเนื่องมาจากตัวแปรเกินที่ไม่สามารถควบคุมได้

ตอนที่ 2 ประเภทของปัญหาวิจัย

ประเภทของปัญหาวิจัยแบ่งเป็น 3 ประเภท (Jean Royer Dyer, 1979) คือ

1. ประเภทบรรยายลักษณะของตัวแปร (Descriptive Question) เป็นการวิจัยที่พยายามบรรยายลักษณะหรือสภาพของตัวแปร หรือบรรยายลำดับของเหตุการณ์ในสภาพธรรมชาติ

2. ประเภทบรรยายความสัมพันธ์แบบเกี่ยวข้องกับระหว่างตัวแปร (Association Question) เป็นการวิจัยที่มุ่งบรรยายความสัมพันธ์ระหว่างตัวแปร หรือศึกษาว่า มีตัวแปรใดบ้างที่มีความสัมพันธ์กับตัวแปรที่เรามุ่งศึกษา ซึ่งอาจเป็นความสัมพันธ์ระหว่างตัวแปร 2 ตัวหรือมากกว่า

3. ประเภทอธิบายความสัมพันธ์เชิงเหตุผลระหว่างตัวแปร (Causal Question) เป็นการวิจัยที่พยายามอธิบายความสัมพันธ์ระหว่างตัวแปรโดยที่มีตัวแปรหนึ่งเป็นตัวแปรเหตุที่ทำให้เกิดตัวแปรอื่น ๆ ปัญหาวิจัยลักษณะนี้มี 2 แบบ คือ แบบที่ 1 มีตัวแปรทดลองหรือตัวแปรจัดกระทำซึ่งแบบวิจัยมักจะเป็นการวิจัยเชิงทดลอง แบบที่ 2 เป็นปัญหาวิจัยซึ่งเป็นแบบของการวิจัยเชิงบรรยายแต่มีค่าที่มีความหมายของความเป็นเหตุเป็นผล เช่น ผลของ... ผลกระทบของ... สาเหตุของ... เป็นต้น

ตอนที่ 3 มาตรฐานค่า

ความหมายของมาตรฐานค่า

มาตรฐานค่ามีชื่อเรียกภาษาไทยหลายชื่อ แต่ที่ชั้บ่อย ๆ ได้แก่ มาตรฐานส่วนประมาณค่า มาตรฐานค่า แบบประมาณค่า แบบประเมินค่า โดยที่ทุกคำมาจากภาษาอังกฤษว่า rating scale ในที่นี้จะใช้คำว่า มาตรฐานค่า ซึ่งได้มีผู้ให้ความหมายไว้ต่าง ๆ กัน ดังตัวอย่างต่อไปนี้

Mchsens (1984) กล่าวว่า มาตรฐานค่า เป็นกระบวนการที่จัดเตรียม ขึ้นอย่างมีระบบ เพื่อใช้ในการรายงานหรือจัดบันทึกพฤติกรรมที่สังเกตได้ โดยที่สามารถ บอกตำแหน่งหรือปริมาณของสิ่งที่สังเกตได้ ซึ่งจะใช้ก็ต่อเมื่อไม่สามารถหาเครื่องมืออื่น ที่เหมาะสมกว่าได้

Good (1959) ให้ความหมาย มาตรฐานค่าว่า หมายถึงเครื่องมือ ที่ใช้สำหรับประเมินผลงาน ความคิดเห็น หรือลักษณะอื่น ๆ ของบุคคล โดยเฉพาะลักษณะ ที่วัดได้ยาก ซึ่งผู้ประเมินจะตัดสินให้ค่าตัวเลขเพื่อแสดงปริมาณของลักษณะที่ถูกรประเมิน

วิเชียร เกตุสิงห์ (2530) ให้ความหมายของมาตรฐานค่า ว่า หมายถึงแบบวัดที่มีข้อความหรือข้อความเกี่ยวกับเรื่องที่จะให้ผู้เข้าพิจารณา พร้อมกับ มีคำตอบที่แสดงความเข้มของความเห็นเกี่ยวกับเรื่องนั้น ๆ

บุญธรรม กิจปรีดาบริสุทธิ์ (2531) ให้ความหมายของมาตรฐานค่าว่า เป็นเครื่องมือที่ใช้วัดสิ่งที่เป็นนามธรรมด้วยการแปลงเป็นปริมาณเชิงเปรียบเทียบ นิยม ใช้วัดพฤติกรรมหรือสิ่งต่าง ๆ ที่ไม่สามารถวัดออกมาเป็นตัวเลขโดยตรงได้ เช่น ความดี ความซื่อสัตย์ ค่านิยม เจตคติ ความเชื่อ ความสะอาด และการปฏิบัติ ต่าง ๆ เป็นต้น ลักษณะสำคัญของมาตรฐานค่าอยู่ที่คำตอบซึ่งมีลักษณะเป็นการ เปรียบเทียบปริมาณมากน้อย ผู้ตอบจะตอบด้วยการประเมิน(Evaluate)สถานการณ์ที่เป็น สิ่งเร้าที่กำหนดค่าให้และ เลือกตอบเพียงคำตอบเดียว

จากความหมายที่มีผู้ให้ไว้ดังกล่าวสรุปได้ว่า มาตรฐานค่า หมายถึง เครื่องมือที่ช่วยในการตีค่าคุณลักษณะทางจิตวิทยาอันเป็นสิ่งที่จับต้องและเป็นนามธรรม วัดได้ยาก ออกมาเป็นค่าเชิงปริมาณอย่างเป็นระบบ ซึ่งสามารถใช้ในการประเมิน กระบวนการ ผลผลิต และบุคลิกภาพทางสังคมของมนุษย์

มาตรฐานค่าประกอบไปด้วย 2 ส่วน คือ ส่วนที่เป็นข้อความ หรือสิ่งเร้า กับส่วนที่เป็นคำตอบหรือสเกลซึ่งแสดงถึงระดับความเข้มของพฤติกรรมหรือคุณลักษณะนั้น ๆ



รูปแบบคำตอบของมาตราประมาณค่า

การอธิบายคำตอบของมาตราประมาณค่าหรือรูปแบบคำตอบของมาตราประมาณค่า (Response Rating Scale Format) มีหลายรูปแบบตั้งแต่นั้นไม่มีคำอธิบายจนกระทั่งมีคำอธิบายทุกสเกลของคำตอบ และได้รับการพัฒนาและปรับปรุงให้แตกต่างกันไป ดังเช่น Smith และ Kendal (1963) ได้พัฒนารูปแบบการประเมินพฤติกรรม และได้สนับสนุนให้มีการสร้างรูปแบบคำตอบที่ทุกสเกลมีคำอธิบายเพื่อบอกค่าพฤติกรรมที่ใกล้เคียงมากที่สุด แต่ไม่ว่ามาตราประมาณค่าจะสร้างได้หลายรูปแบบแตกต่างกันไป ยังมีสิ่งสำคัญที่ต้องคงลักษณะที่ร่วมกันไว้คือ จะต้องมีส่วนที่เป็นลักษณะที่ต้องการให้ประเมินหรือข้อความ และส่วนคำตอบที่เป็นสเกล ซึ่งส่วนหลังนี้จะเปลี่ยนแปลงไปตามความต้องการของผู้สร้าง ซึ่งรูปแบบที่ต่างกันอาจกำหนดสเกลเป็นตัวเลข ตัวอักษร คำอธิบาย หรือคำคุณศัพท์ก็ได้ นอกจากนี้ยังมีผู้พัฒนารูปแบบคำตอบที่แตกต่างกันมากมาย ดังที่ Guilford (1954) กล่าวว่า มาตราประมาณค่าแต่ละแบบจะเหมือนกันตรงที่ทำให้ประเมินสิ่งต่าง ๆ โดยพิจารณาจากระดับของสเกล และผลสุดท้ายจะเป็นผลรวมของจำนวนนับตามที่ประเมินไว้ ความแตกต่างจะอยู่ที่วิธีการจัดตำแหน่ง ชนิดและจำนวนสเกล ตลอดจนความละเอียดของระดับ ฉะนั้นรูปแบบคำตอบ จึงมีความแตกต่างกันไปตามความต้องการของผู้สร้าง และรูปแบบคำตอบของมาตราประมาณค่าแบ่งได้เป็น 5 แบบ ดังนี้

1. แบบตัวเลข (Numerical Scales) แบบนี้จะกำหนดตัวเลขที่เรียงลำดับมาที่ 1 ชุด พร้อมกับคำนิยามหรือคำอธิบายไว้ให้ สำหรับให้ผู้ประเมินได้เลือกตอบหรือบางครั้งอาจมีเพียงคำนิยามอย่างเดียวโดยไม่มีตัวเลขกำกับอยู่ก็ได้ เช่น

- 5 หมายถึง หนักมาก
- 4 หมายถึง หนัก
- 3 หมายถึง ปานกลาง
- 2 หมายถึง เบา
- 1 หมายถึง เบามาก

2. แบบกราฟิก (Graphic Scales) เป็นแบบที่นิยมใช้กันมาก เน้นการกำหนดข้อความมาให้เข้ากับเส้นตรง ใช้ได้ทั้งเส้นตรงแนวนอนและแนวตั้ง ดังตัวอย่างในวงสนทนา คุณเป็นอย่างไร _____

ข้างพูด พูดน้อย พูดเมื่อ ชอบฟัง ไม่เคย
 จำเป็น มากกว่า พูดเลย

3. แบบสเกลมาตรฐาน (Standard Scales) แบบนี้จะต้องกำหนดค่าของสเกลมาตรฐานไว้ก่อน แล้วจึงนำสิ่งที่ต้องการประเมินไปเปรียบเทียบกับสเกลมาตรฐาน วิธีนี้จัดทำลำบาก จึงไม่นิยมใช้ มีที่ใช้คือการเปรียบเทียบคัดลายมือ

4. แบบแต้มสะสม (Rating by Cumulated Points) แบบนี้แตกต่างจากแบบอื่น ๆ ที่วิธีการให้คะแนน ตัวอย่างหนึ่งของแบบนี้คือ วิธีการตรวจรายการ (Check - list Method) ซึ่งใช้ในการประเมินคุณลักษณะของเด็กโรคยุงการกำหนดรายการที่มีทั้งคุณลักษณะที่ดีและไม่ดีมาให้ผู้ประเมินเลือก ผู้ประเมินแต่ละคนจะเลือกทุก ๆ คุณลักษณะที่ตรงกับเด็กคนนี้ แล้วจึงนำคะแนนของผู้ประเมินทั้งหมดมารวมกันโดยกำหนดคะแนน +1 สำหรับคุณลักษณะที่ดี -1 สำหรับคุณลักษณะที่ไม่ดี

5. แบบบังคับให้เลือก (Force-choice Ratings) ใช้สำหรับประเมินบุคคลโดยกำหนดคุณลักษณะของบุคคลมาให้เป็นคู่ ผู้ประเมินเพียงแต่ประเมินว่าบุคคลนั้นมีลักษณะใดมากกว่ากัน

พวงรัตน์ ทวีรัตน์ (2523) แบ่งมาตรฐานประมาณค่าเป็น 3 แบบ คือ

1. แบบจัดประเภท เป็นการกำหนดลักษณะพฤติกรรมที่เข้ากับพฤติกรรมไว้ให้ผู้ประเมินเลือก เช่น

เขาเป็นคนขยันเพียงใด

..... ขยันมากที่สุด

..... ขยันมาก

..... ขยันปานกลาง

..... ขยันน้อย

..... ขยันน้อยมาก

2. แบบกำหนดเป็นตัวเลข แบบนี้ใช้รหัสตัวเลขแทนลักษณะต่าง ๆ ของพฤติกรรม แล้วให้ผู้ประเมินทำเครื่องหมาย ลงบนตัวเลข เช่น

เขาเป็นคนคล่องแคล่วเพียงใด

1 2 3 4

1 = ไม่คล่องแคล่วเลย 2 = คล่องแคล่วน้อย

3 = คล่องแคล่วมาก 4 = คล่องแคล่วมากที่สุด

3. แบบกำหนดเป็นกราฟ กำหนดให้ผู้สังเกตทำเครื่องหมายถูกตามลักษณะที่เห็นว่าตรงตามสิ่งที่สังเกต ลงบนเส้นกราฟ เช่น

เขาเป็นคนพูดมากเพียงใด



พูดมากที่สุด พูดมากปานกลาง พูดน้อย ไม่พูดเลย

วิเชียร เกตสิงห์ (2530) ได้แบ่งรูปแบบของมาตราประมาณค่า เป็น 3 รูปแบบ คือ

1. แบบใช้ตัวเลข (Numerical Scales) เป็นวิธีกำหนดความหนักเบาหรือระดับของความเห็นเป็นตัวเลข

2. แบบเส้นกราฟ (Graphic Scales) ใช้เส้นแสดงทิศทาง และความคิดเห็นบนเส้นกราฟ

3. แบบมาตรฐาน (Standard Scales) เป็นแบบเพื่ออาศัยการวิเคราะห์การตอบจากการทดลองใช้ หรือจากการนำแบบสอบถามไปใช้มาแล้ว เพื่อกำหนดว่าการเห็นด้วยกับข้อนั้นจะมีคะแนนเท่าใด เรียกคะแนนว่า "ค่าประจำข้อ" ใครตอบข้อใดจะได้คะแนนเท่ากับค่าประจำข้อนั้น ๆ

Turney & Robb (1971) และ Schwartz & Tiedman (1957) มีความเห็นตรงกันว่า ถึงแม้มาตราประมาณค่าจะมีหลายรูปแบบก็ตาม แต่แบบที่เป็นพื้นฐานมี 3 แบบ คือ

1. แบบตัวเลข (Numerical Scale) ใช้ตัวเลขบ่งชี้ระดับของลักษณะหรือพฤติกรรมที่ต้องการประเมิน เช่น ให้ประเมินพฤติกรรมของนักเรียนโดยวงกลมล้อมรอบตัวเลข ตัวเลข 1 เมื่ออยู่ในระดับต่ำ 2 เมื่ออยู่ในระดับต่ำกว่าปานกลาง 3 เมื่ออยู่ในระดับปานกลาง 4 เมื่ออยู่ในระดับสูงกว่าปานกลาง และ 5 เมื่ออยู่ในระดับสูง

ความร่วมมือกับเพื่อน ๆ 1 2 3 4 5

ความร่วมมือกับครู 1 2 3 4 5

2. แบบกราฟิก (Graphic Scale) เป็นวิธีที่เขียนระดับพฤติกรรมบนเส้นตรง โดยผู้ประเมินจะขีดเครื่องหมายลงบนเส้นตรงในตำแหน่งที่เหมาะสมกับพฤติกรรมที่วัดในแต่ละพฤติกรรม ตรงตำแหน่งใดก็ได้ เพื่อชี้ค่าประเมิน เช่น ให้ทำเครื่องหมายลงบนเส้นตรงในตำแหน่งที่ตรงกับลักษณะของนักเรียนมากที่สุด

ความร่วมมือ

จัดขวาง

เต็มใจเสมอ

3. แบบบรรยาย (Descriptive Scale) แบบนี้แต่ละคุณลักษณะจะอธิบายออกมาเป็นระดับความเข้มของคุณลักษณะนั้น ๆ แล้วให้ผู้ประเมินเลือกระดับพฤติกรรมที่ใกล้เคียงกับบุคคลหรือสิ่งที่ประเมินมากที่สุด แล้วทำเครื่องหมายลงในช่วงระดับที่เลือกไว้ เช่น ข้อความที่บรรยายพฤติกรรมของนักเรียนตามที่สังเกตได้

นักเรียนคนนี้เข้มแข็งหรือไม่

..... เข้มแข็งและคล่องแคล่ว

..... ไม่ใคร่เหนียว

..... มีความอดทน

..... เจื่องหงอย, เหนื่อยหน่าย

อย่างไรก็ตามสามารถนำทั้ง 3 รูปแบบมาใช้ร่วมกันได้แล้วแต่ความต้องการของผู้สร้าง และลักษณะของสิ่งที่ต้องการประเมิน เช่น อาจมีทั้งตัวเลขและคำบรรยาย หรืออาจมีเส้นกราฟพร้อมทั้งตัวเลขและคำอธิบาย โดยการอธิบายยังสามารถปรับเปลี่ยนไปได้อีกหลายแบบตั้งแต่เป็นคำวิเศษณ์ง่าย ๆ จนถึงอธิบายด้วยตัวอย่างพฤติกรรม

จำนวนช่วงคำตอบของมาตรฐานค่า

มาตรฐานค่าโดยมากกำหนดจำนวนช่วงคำตอบไว้ไม่ต่ำกว่า 3 ช่วงและไม่เกิน 11 ช่วง ส่วนใหญ่จะใช้กันเพียง 5 ช่วง (สุภาพ วาดเขียน, 2524) การแบ่งช่วงยิ่งละเอียดมากยิ่งขึ้นแสดงความแตกต่างได้มาก แต่ไม่ควรมากเกินไปจนแยกความแตกต่างไม่ได้ นอกจากนี้ได้มีการศึกษาที่มีผลตรงกันว่าจำนวนช่วงคำตอบที่เหมาะสมและลดความคลาดเคลื่อนได้คือ จำนวน 3 - 7 ช่วง (Gronlund, 1981)

ลักษณะที่ดีของมาตรฐานค่า

1. ส่วนของชุดสิ่งเร้าที่เป็นลักษณะที่จะให้ประเมินหรือพฤติกรรมที่เลือกมาควรมีลักษณะ ดังนี้

- (1) เป็นสิ่งที่มีความสำคัญตรงกับสิ่งที่ต้องการวัดมากที่สุด
- (2) ต้องเป็นอิสระไม่ก้ำกึ่งกับลักษณะอื่น
- (3) เป็นสิ่งที่สามารถสังเกตเห็นได้โดยตรง
- (4) มีความเป็นเอกภาพ ไม่ซ้ำซ้อนประกอบของลักษณะที่มีความสัมพันธ์กันต่ำ

สัมพันธ์กันต่ำ

- (5) คุณลักษณะที่จะให้ประเมินไม่ควรมีมากเกินไป

2. ส่วนของคำตอบที่ให้ความหมายระดับพฤติกรรมจะต้องสร้าง ดังนี้

(1) ให้คำนิยามหรือบรรยายระดับพฤติกรรมของแต่ละลักษณะที่จะประเมินให้ชัดเจน เพื่อให้เกิดความชัดเจนเป็นที่เข้าใจตรงกันของผู้ประเมิน จะทำให้สามารถตัดสินใจได้อย่างเป็นปรนัยและมีความเที่ยง

- (2) สิ่งที่อยู่อธิบายระดับพฤติกรรมจะต้องมีความต่อเนื่องกันตามลำดับ

ทุกสเกล

- (3) คำอธิบายสเกลแต่ละพฤติกรรมจะต้องสอดคล้องกับแต่ละลักษณะ

- (4) ควรเป็นคำอธิบายที่เข้าใจง่ายและตรงกันสามารถมองเห็น

เป็นรูปธรรม มากกว่าที่จะใช้คำวิเศษณ์ง่าย ๆ ทั่ว ๆ ไป

(5) การอธิบายในทุกสเกลจะต้องเป็นเรื่องเดียวกัน

(6) จำนวนช่วงหรือสเกลที่จะประเมินควรสร้างให้เหมาะสม

ปกติมักแบ่งช่วงออกเป็น 3-7 ช่วง มีการนิยามคำตอบของสเกลอย่างชัดเจนและควรมีการสลับช่วงหรือลำดับพฤติกรรมจากมากไปหาน้อย และจากน้อยไปหามาก เพื่อที่จะแก้ปัญหาการประเมินโดยอัตโนมัติ

ตอนที่ 4 มาตรการวัด

ในการใช้สถิติเพื่อวิเคราะห์ข้อมูลในการวิจัยนั้น ผู้วิจัยจำเป็นต้องทราบลักษณะของข้อมูลที่ใช้ในการวิจัยเสียก่อน เพื่อเป็นเครื่องวินิจฉัยว่าจะใช้วิธีการทางสถิติแบบใดในการวิเคราะห์จึงจะถูกต้องเหมาะสม ดังนั้นผู้วิจัยจึงควรทราบถึงมาตรการวัดของข้อมูล โดย S.S.Stevens (1946) ได้จำแนกมาตรการวัดออกเป็น 4 ระดับ ดังนี้

1. มาตรการแบ่งกลุ่ม (Nominal Scales) เป็นมาตรการวัดในระดับต่ำสุด เป็นการวัดแบบง่าย ๆ โดยการจำแนกหรือแยกประเภท (Classification) ตามคุณลักษณะที่ไม่เหมือนกัน อาจเป็นเพียงการเรียกชื่อ ดังนั้นการวัดในระดับนี้บางทีจึงไม่เป็นที่ยอมรับว่าเป็น "การวัด" เพราะไม่สามารถบอกปริมาณมากน้อยได้ เป็นแต่เพียงแสดงให้เห็นความแตกต่างของสิ่งต่าง ๆ เท่านั้น ตัวอย่างการวัดในระดับนี้ เช่น การจำแนกคนตามเพศ เป็นเพศชาย - หญิง หรือจำแนกคนตามศาสนาที่นับถือ เป็นคนที่นับถือศาสนาพุทธ คริสต์ อิสลาม เป็นต้น

การกำหนดตัวเลขให้กับสิ่งต่าง ๆ ในมาตรการแบ่งกลุ่มจะเป็นเพียงตัวแทนประเภทของที่ถูกวัด หรือเพื่อใช้ในการสื่อความหมาย ให้จำง่ายหรือสะดวกในการนับ โดยที่ตัวเลขดังกล่าวไม่มีความหมายในเชิงปริมาณแต่ประการใด ดังนั้นจึงไม่สามารถนำตัวเลขเหล่านั้นมาบวก ลบ คูณ หรือหารกันได้ นอกจากกระทำได้แต่เพียงการนับจำนวนหรือหาความถี่ของลักษณะที่มีเหมือนกันเท่านั้น เช่น กำหนดค่าให้เลข 1 แทนเพศชาย และ เลข 2 แทนเพศหญิง

2.มาตราจัดอันดับ (Ordinal Scales) การวัดตามมาตรานี้จะมีระดับการวัดสูงกว่ามาตราแบ่งกลุ่ม ลักษณะการวัดเป็นการจัดอันดับ (rank order) ให้ความสัมพันธ์บางอย่างว่ามากกว่าหรือน้อยกว่าลักษณะอื่น ๆ ซึ่งสามารถจัดอันดับข้อมูลตามตำแหน่งได้จากมากที่สุดไปหาน้อยที่สุด หรือจากน้อยที่สุดไปหามากที่สุดได้ เช่น นักเรียนที่สอบได้ที่ 1, 2, 3, ... เป็นต้น ผลการวัดตามมาตรานี้ช่วงห่างระหว่างอันดับจะไม่เท่ากัน และไม่ทราบว่าห่างกันเป็นปริมาณเท่าใด เช่น จะบอกว่าการเรียนที่สอบได้ที่ 1 เก่งกว่านักเรียนที่สอบได้ที่ 2 เท่ากับนักเรียนที่สอบได้ที่ 3 เก่งกว่านักเรียนที่สอบได้ที่ 4 ไม่ได้ เป็นต้น ดังนั้นข้อมูลที่ปรากฏออกมาเป็นตัวเลขในระดับนี้ไม่สามารถนำมาบวก ลบ คูณ หรือหารกันได้ จะบอกได้แค่เพียงว่าต่างกันไปในทางไหน หรือทิศทางที่ตรงกันเท่านั้น

3.มาตราอันตรภาค (Interval Scales) มาตราการวัดตามมาตรานี้มีลักษณะเหมือนมาตราจัดอันดับทุกอย่าง แต่มีคุณสมบัติเพิ่มเติมมากกว่าตรงที่แต่ละหน่วยของการวัดมีระยะห่างเท่า ๆ กัน จึงสามารถเปรียบเทียบกันได้ว่ามากหรือน้อยกว่ากันเท่าใด แต่ไม่สามารถบอกได้ว่าเป็นกี่เท่าของกันและกัน เพราะมาตรานี้ไม่มีศูนย์สมบูรณ์ (true zero or absolute zero) เช่น ความร้อนที่มีหน่วยการวัดเป็นฟาเรนไฮต์ และเซลเซียส ณ ที่อุณหภูมิ 0 องศาเซลเซียส ไม่ได้หมายความว่า ปราศจากความร้อนเลย ซึ่ง 0 องศาเซลเซียสนี้เท่ากับ 32 องศาฟาเรนไฮต์ หรือผู้ที่สอบได้ศูนย์คะแนน ไม่ได้หมายความว่าไม่มีความรู้ในวิชานั้น เป็นแค่เพียงผู้สอบคนนั้นทำข้อสอบฉบับนั้นไม่ได้เลย เป็นต้น ด้วยเหตุนี้ข้อมูลที่วัดได้ตามมาตรานี้จึงไม่สามารถนำมา บวก ลบ คูณ หรือหารกันได้ เฉพาะ บวก หรือ ลบ เท่านั้น

4.มาตราอัตราส่วน (Ratio Scales) เป็นมาตราการวัดในระดับสูงที่สุดกล่าวคือ ครอบคลุมคุณสมบัติทุกอย่างของมาตราอันตรภาค แต่ดีกว่ามาตราอันตรภาคตรงที่มีศูนย์แท้ ข้อมูลในระดับนี้สามารถบวก ลบ คูณ หรือหารกันได้ มาตราการวัดที่จัดอยู่ในระดับนี้มักจะเป็นการวัดทางฟิสิกส์ เช่น ความยาว ความสูง น้ำหนัก เวลา เป็นต้น

ตอนที่ 5 การเลือกใช้เทคนิคทางสถิติเพื่อวิเคราะห์ข้อมูล

การเลือกใช้เทคนิคทางสถิติเพื่อวิเคราะห์ข้อมูล เป็นขั้นตอนที่สำคัญขั้นตอนหนึ่งของการวิจัยเพราะถ้าเลือกใช้เทคนิคทางสถิติไม่ถูกต้องหรือไม่เหมาะสมกับข้อมูลที่จะทำการวิเคราะห์ก็จะทำให้ผลการวิจัยผิดพลาดได้ ซึ่งในเรื่องนี้ตามคู่มือการเลือกใช้เทคนิคทางสถิติเพื่อวิเคราะห์ข้อมูลทางสังคมศาสตร์ (อุทุมพร (ทองอุทัย) จามรมาน, 2531) ระบุว่า การเลือกใช้สถิติในการวิเคราะห์ข้อมูลในงานวิจัยทางสังคมศาสตร์มีประเด็นในการพิจารณาอยู่ 4 ข้อคือ

1. จำนวนตัวแปร (Variable) กับประเภทของตัวแปร
2. ตัวแปรนั้น ๆ ใดที่ข้อมูลประเภทใด (Scale of Measurement)
3. แหล่งข้อมูลมาจากประชากรหรือกลุ่มตัวอย่าง
4. ผู้วิจัยต้องการบรรยายสรุปข้อมูลที่ได้มาเท่านั้น หรือสรุปอ้างอิงไปยังแหล่งข้อมูลทั้งหมด (ประชากร)

การวิเคราะห์ข้อมูลตัวแปรเดียว

ตัวอย่างงานวิจัยที่มีตัวแปรเดียว เช่น การศึกษาความสามารถของผู้เรียนที่สำเร็จการศึกษาภาคบังคับในวิชาคณิตศาสตร์ หรือการศึกษาความแข็งแรงของร่างกายหญิงมีครรภ์ หรือการศึกษาแนวโน้มของจำนวนผู้สมัครสอบเข้ามหาวิทยาลัยปิดของรัฐ เป็นต้น การเลือกใช้สถิติที่เหมาะสมในกรณีที่เป็นงานวิจัยที่มีตัวแปรเดียว จำนวนดังนี้

- ก. ต้องการบรรยายลักษณะของข้อมูล

ดูแผนภูมิที่ 1

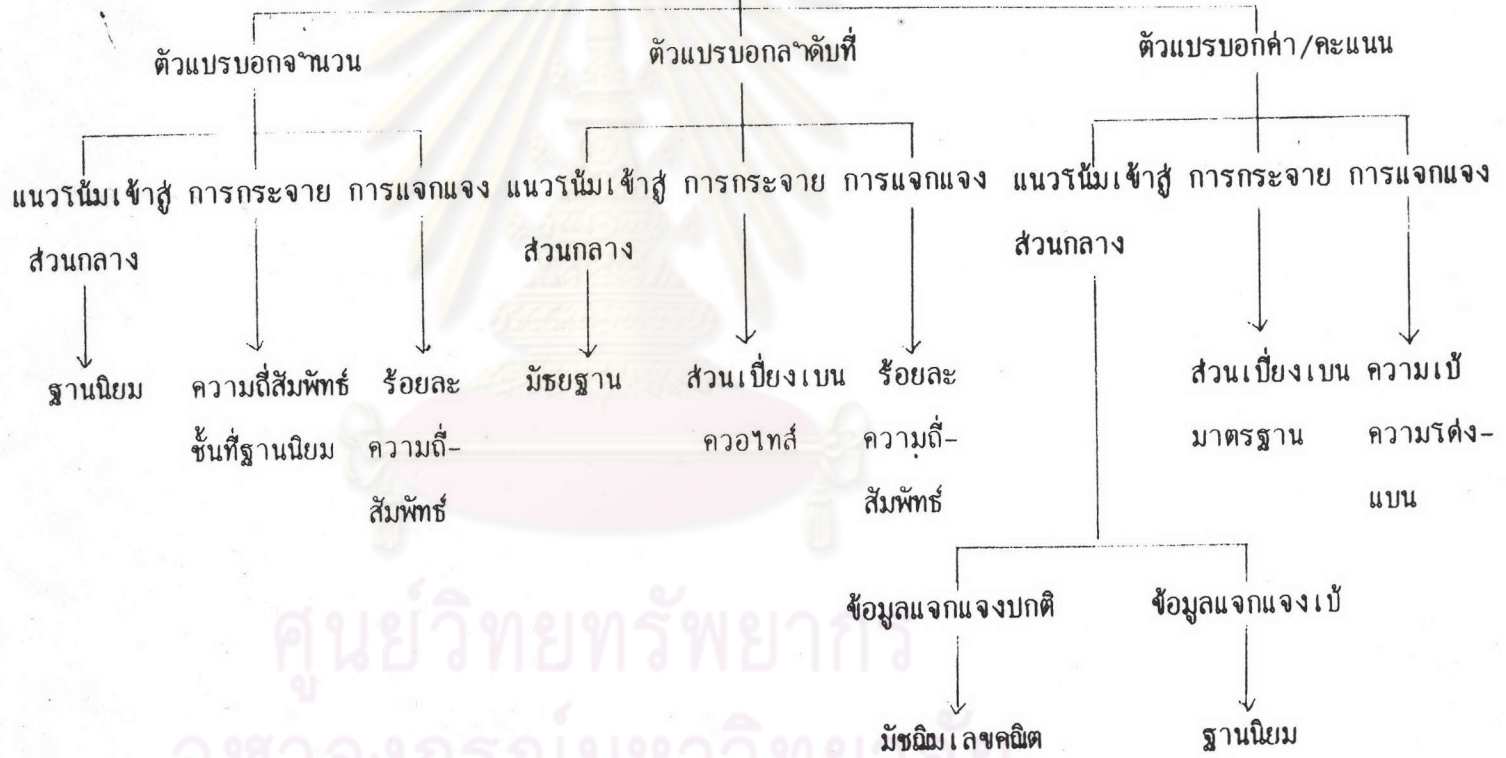
แผนภูมิที่ 1

ต้องการบรรยายลักษณะข้อมูล

ประเภทของตัวแปร

วัตถุประสงค์ของการวิเคราะห์

สถิติที่เหมาะสม



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

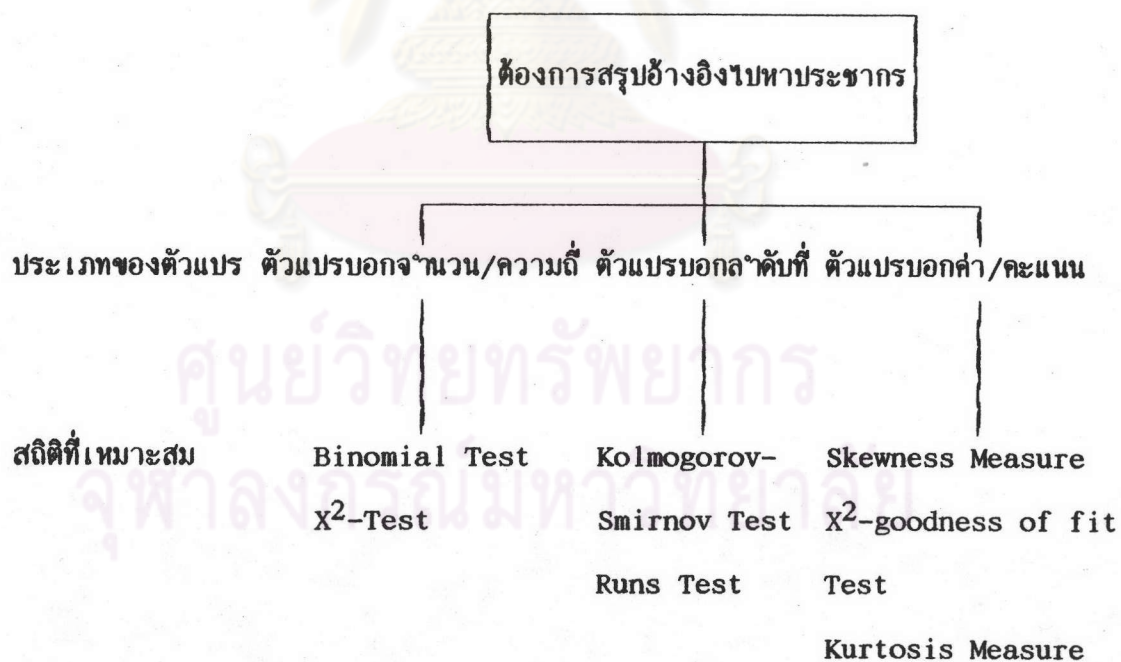
ข. ต้องการสรุปอ้างอิงข้อมูลจากกลุ่มตัวอย่างไปหาประชากร

ก่อนที่จะทำการสรุปอ้างอิงข้อมูลจากกลุ่มตัวอย่างไปหาข้อมูลในประชากร ซึ่งบางคนเรียกว่า การทดสอบสมมติฐาน / การทดสอบนัยสำคัญ / การประมาณค่า พารามิเตอร์ มีคำถามนาคือ

ข้อมูลที่รวบรวมมาค่านั้นเป็นข้อมูลจาก

- ประชากร (การวิเคราะห์ค่าใช้ตามแผนภูมิที่ 1)
- กลุ่มตัวอย่าง
 - ต้องการบรรยายลักษณะข้อมูลกลุ่มตัวอย่าง (ดูแผนภูมิที่ 1)
 - ต้องการสรุปอ้างอิงข้อมูลจากกลุ่มตัวอย่างไปหาประชากร (ดูแผนภูมิที่ 2)

แผนภูมิที่ 2



รายละเอียดของแผนภูมิที่ 2 มีดังนี้

1. ข้อมูลตัวแปรที่เป็นจำนวนหรือความถี่ (Nominal Data)

1.1 ทดสอบเกี่ยวกับการแจกแจงแบบไบนอมิยัล ๒



The Binomial Test

1.2 ทดสอบเกี่ยวกับจำนวนที่คาดหวังกับจำนวนที่ได้รับว่าแตกต่างกันหรือไม่ ใช้ The X^2 -Test

2. ข้อมูลตัวแปรที่เป็นลำดับที่ (Ordinal Data)

2.1 ทดสอบความสอดคล้องของการแจกแจงของข้อมูลกลุ่มตัวอย่างกับข้อมูลเชิงทฤษฎี ใช้ The Kolmogorov-Smirnov Test

2.2 การทดสอบลักษณะการแจกแจงว่าเป็นการสุ่มจากประชากรหรือไม่ ใช้ The One Sample Runs Test

3. ข้อมูลตัวแปรที่เป็นค่าหรือคะแนน (Interval Data)

3.1 ทดสอบว่าการแจกแจงของข้อมูลแตกต่างไปจากการแจกแจงแบบปกติหรือไม่

ถ้าขนาดของกลุ่มตัวอย่างมากกว่า 150 ราย ให้ใช้ค่าอัตราส่วนวิกฤต (Critical Ratio) ของค่าความเบ้ (Skewness Measure) เทียบกับค่าในตารางรังผึ้งปกติ

ถ้าขนาดของกลุ่มตัวอย่างอยู่ระหว่าง 25-150 ราย ใช้การทดสอบค่าความเบ้

3.2 ทดสอบว่า การแจกแจงของข้อมูลแตกต่างจากการแจกแจงแบบปกติหรือไม่

ถ้าขนาดของกลุ่มตัวอย่างมากกว่า 1,000 ราย ให้ใช้อัตราส่วนวิกฤตของค่าความโค้ง-แบน (Kurtosis Measure) เทียบกับค่าในตารางรังผึ้งปกติ

ถ้าขนาดของกลุ่มตัวอย่างอยู่ระหว่าง 200-1,000 ราย ให้ทดสอบค่าความโค้ง-แบน

ถ้าขนาดของกลุ่มตัวอย่างน้อยกว่า 200 ราย ใช้ Geary's Criterion

3.3 ทดสอบการแจกแจงของข้อมูลว่าแตกต่างจากการแจกแจงแบบปกติหรือไม่ ใช้

(1) Kolmogorov-Smirnov One Sample Test

(2) Lilliefors extension of the Kolmogorov-Smirnov Test

(3) Chi-Square goodness of fit Test

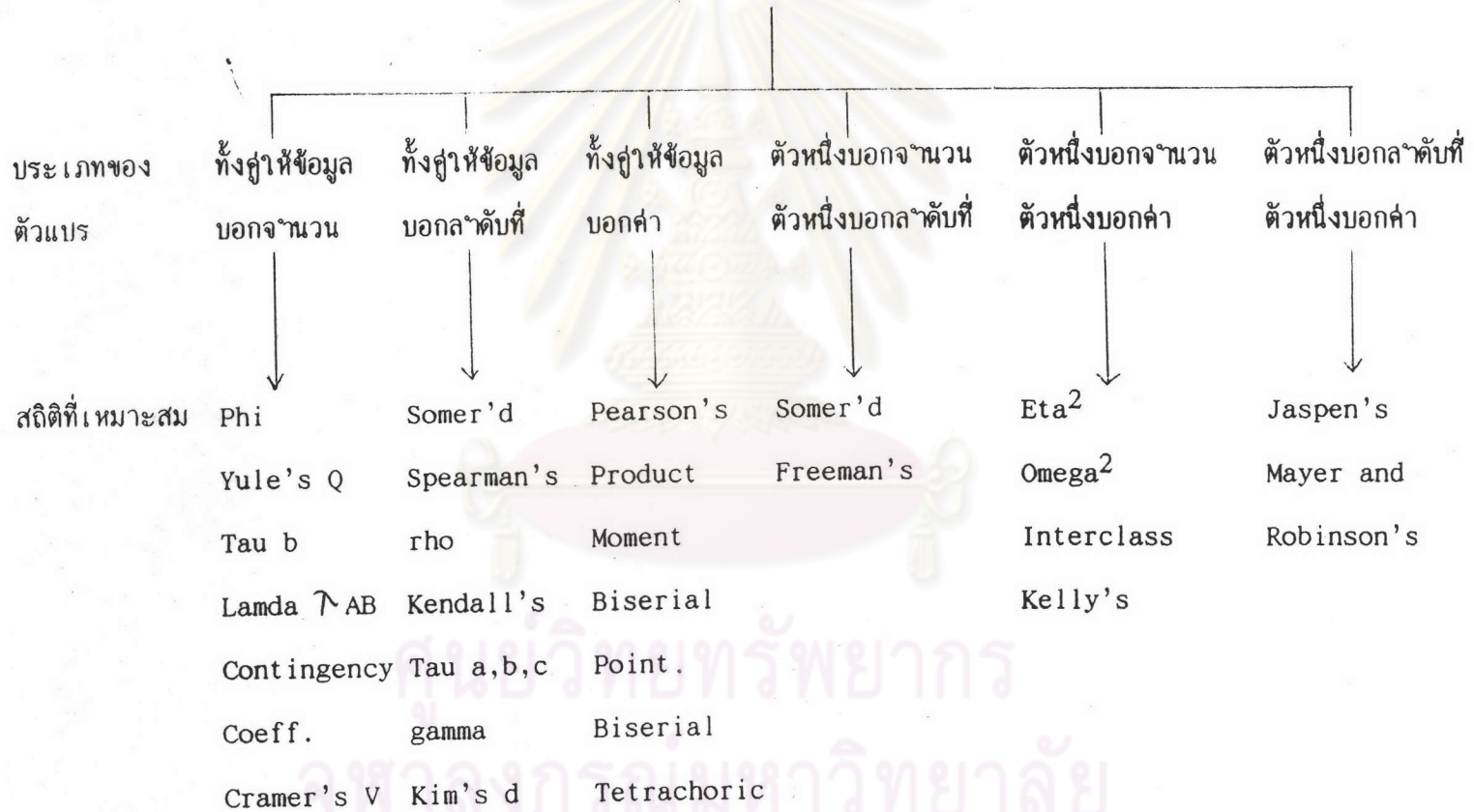
การวิเคราะห์ข้อมูลตัวแปร 2 ตัว

งานกรณีที่มีตัวแปร 2 ตัว เช่นศึกษาน้ำหนักและส่วนสูงของคนไทยพร้อมกัน การวิเคราะห์ข้อมูลมีความยุ่งยากมากขึ้น ก่อนเลือกสถิติที่เหมาะสมมาใช้ ผู้ใช้ควรตอบคำถามเบื้องต้นดังนี้

- (1) ตัวแปรให้ข้อมูลแบบใด
 - [] บอกจำนวน / ความถี่
 - [] บอกลำดับที่
 - [] บอกค่า / คะแนน
 - [] ผสม
- (2) สามารถระบุเป็นตัวแปรอิสระและตัวแปรตามได้หรือไม่
 - [] ได้ (เช่น การวิจัยเชิงทดลอง)
 - [] ไม่ได้ (เช่น การวิจัยเชิงสำรวจ)
- (3) การวิเคราะห์มุ่งที่ (ตอบได้มากกว่า 1 คำตอบ)
 - [] การบรรยายลักษณะข้อมูลของแต่ละตัวแปรแยกกัน (ดูแผนภูมิที่ 1)
 - [] การบรรยายลักษณะข้อมูลของตัวแปรทั้งสองพร้อมกันโดยใช้ตาราง 2 มิติ หรือคำนวณค่าร้อยละเทียบจากจำนวนรวมตามแถว คอลัมน์ และทั้งหมด
 - [] การบรรยายความสัมพันธ์ของตัวแปรทั้งสองแล้วทดสอบความมีนัยสำคัญ (ดูแผนภูมิที่ 3)
 - [] การเปรียบเทียบความแตกต่างของตัวแปรทั้งสอง แล้วทดสอบความมีนัยสำคัญ (ดูแผนภูมิที่ 4)

แผนภูมิที่ 3

การเลือกใช้เทคนิคที่ต้องการบรรยายความสัมพันธ์ของตัวแปรสองตัว



รายละเอียดของแผนภูมิที่ 3 มีดังนี้

(ข้อ 1) ตัวแปรทั้งคู่ให้ข้อมูลบอกจำนวน (Two Nominal Variables)

1.1 ถ้าต้องการวัดความสมมาตร (Symmetry) ของการแจกแจงความถี่ใช้

McNemar's Test of Symmetry

1.2 ถ้าต้องการวัดความสอดคล้อง (Agreement) ของตัวแปรทั้งคู่

1.2.1 Scott's Coefficient of agreement (π)

1.2.2 Cohen's agreement coefficient (Kappas X) แล้วทดสอบ

อัตราส่วนวิกฤตกับตารางร้อยละ

1.3 ถ้าต้องการวัดความเกี่ยวข้อง (Covariation) ของตัวแปรทั้งคู่

1.3.1 Yule's Q

1.3.2 Phi coefficient (ϕ) แล้วทดสอบความมีนัยสำคัญกับตาราง

ร้อยละ

1.3.3 Tschuprow's T, Contingency coefficient (c) และ

Cramer's V แล้วทดสอบความมีนัยสำคัญด้วย Pearson's Chi-Square Test

1.3.4 Symmetric lamda (τ_{AB}) แล้วทดสอบค่าอัตราส่วนวิกฤตกับ

ตารางร้อยละ

1.4 ถ้ากำหนดได้ว่าตัวแปรใดเป็นตัวแปรอิสระ และตัวแปรใดเป็นตัวแปรตาม

1.4.1 Goodman and Kruskal's tau b แล้วทดสอบค่า tau b กับ

ตารางร้อยละ

1.4.2 Asymmetric lamda (π_a, π_b) แล้วทดสอบค่า lamda กับ

ตารางร้อยละ

(ข้อ 2) ตัวแปรทั้งคู่ให้ข้อมูลลำดับที่ (Two Ordinal Variables)

2.1 ต้องการวัดความเกี่ยวข้อง (Covariation)

2.1.1 Spearman's rho

ถ้าขนาดของกลุ่มตัวอย่างเท่ากับหรือมากกว่า 10 คู่ เทียบกับค่าใน

ตารางแจกแจง t

ถ้าขนาดของกลุ่มตัวอย่างน้อยกว่า 10 คู่ เทียบกับตารางของ rho

2.1.2 Kendall's tau a, tau b, tau c Goodman and
Kruskal's gamma, Kim's d

อิงตาราง s ถ้าต้องการทดสอบความมีนัยสำคัญ

2.2 ถ้ากำหนดค่าตัวแปรใดเป็นตัวแปรอิสระและตัวใดเป็นตัวแปรตาม

Somers' d

ถ้าขนาดของกลุ่มตัวอย่างมากกว่า 10 คู่ อิงตาราง รังนกพิราบและถ้าน้อยกว่า
10 คู่ อิงตาราง s ในการทดสอบความมีนัยสำคัญ

(ข้อ 3) ตัวแปรทั้งคู่ให้ข้อมูลบอกค่า / คะแนน (Two Interval Variables)

การพิจารณาที่สำคัญของเรื่องนี้ คือ

ก. ความสัมพันธ์ของตัวแปรทั้งคู่เป็นเส้นตรง (Linear) หรือเส้นโค้ง
(Curvilinear)

ข. ต้องการวัดความสอดคล้อง (agreement) หรือความเกี่ยวข้อง
(covariation)

ค. ตัวแปรที่ให้ค่าหรือคะแนนคือตัวแปรต่อเนื่อง (Continuous Variable)
แต่อาจมีการจัดกระทำให้เป็นตัวแปรแบ่งครึ่ง (Dichotomus Variable) ได้ เช่น
ทำตัวแปรคะแนนความสามารถภาษาอังกฤษเป็นตัวแปรแบ่งสองคือ เก่งกับไม่เก่ง

3.1 ความสัมพันธ์ของตัวแปรทั้งสองเป็นเส้นตรง

3.1.1 ตัวแปรต่อเนื่องทั้งคู่ การวัดความเกี่ยวข้อง (Covariation)

ใช้ Pearson's Product Moment แล้วทดสอบความมีนัยสำคัญด้วย t-test หรือแปลง
เป็นค่า Z แล้วเทียบกับตารางรังนกพิราบ

3.1.2 ตัวแปรหนึ่งเป็นตัวแปรต่อเนื่อง อีกตัวหนึ่งแปลงเป็นตัวแปรแบ่งสอง
การวัดความเกี่ยวข้องใช้ Biserial r แล้วทดสอบความมีนัยสำคัญกับตารางรังนกพิราบ

3.1.3 ตัวแปรหนึ่งเป็นตัวแปรต่อเนื่อง อีกตัวหนึ่งแปลงเป็นตัวแปรแบ่งสอง
แต่ยังต้องการประมาณค่าความเกี่ยวข้อง ถ้าตัวแปรทั้งคู่ยังคงเป็นตัวแปรต่อเนื่องใช้ Point
Biserial แล้วทดสอบความมีนัยสำคัญกับตารางรังนกพิราบ

3.1.4 ตัวแปรทั้งสองแปลงเป็นตัวแปรแบ่งครึ่ง แต่ต้องการประมาณค่าความเกี่ยวข้องของตัวแปรทั้งสอง ถ้าตัวแปรทั้งสองยังคงเป็นตัวแปรต่อเนื่องใช้ Tetrachoric r แล้วทดสอบความมีนัยสำคัญเทียบกับตารางรังปกติ

3.1.5 ตัวแปรทั้งสองเป็นตัวแปรต่อเนื่อง การวัดความสอดคล้อง (Agreement) ใช้ Robinson's A

Intraclass Correlation (r_i) แล้วทดสอบความมีนัยสำคัญด้วย F-Test

Krippendorff's Coefficient of agreement (f)

3.1.6 ถ้าระบุได้ว่าตัวแปรใดเป็นตัวแปรอิสระ และตัวใดเป็นตัวแปรตามต้องการสร้างสมการถดถอย ใช้

Regression coefficient (b , Beta) แล้วทดสอบความมีนัยสำคัญด้วย F-Test

3.2 ความสัมพันธ์ของตัวแปรทั้งสองเป็นเส้นโค้ง

3.2.1 ต้องการสร้างสมการถดถอย ใช้ Regression Coefficient from Curvilinear (b , Beta) แล้วทดสอบด้วยความมีนัยสำคัญ ด้วย F-Test (ข้อ 4) ตัวแปรหนึ่งให้ข้อมูลบอกจำนวน อีกตัวหนึ่งให้ข้อมูลบอกลำดับที่ (Two Variables: One Nominal, One Ordinal)

4.1 ความสัมพันธ์ ใช้

Somers'd การทดสอบความมีนัยสำคัญของค่า d ขึ้นอยู่กับขนาดของกลุ่มตัวอย่าง ถ้ากลุ่มตัวอย่างมีขนาดใหญ่กว่า 10 ให้แปลงค่า d เป็น s แล้วเทียบกับตารางรังปกติ ถ้ากลุ่มตัวอย่างมีขนาดเล็กกว่า 10 ให้แปลงค่า d เป็น s แล้วทดสอบด้วย ตาราง S

Sign Test

Wilcoxon signed rank Test

Median Test

Mann-Whitney U Test

Kolmogorov-Smirnov two sample Test

Runs Test

4.2 ความแตกต่าง ๖

Freeman's coefficient of differentiation (θ)

Friedman Test

Kruskal-Wallis Test

Median Test (สำหรับกลุ่มมากกว่า 2 กลุ่ม)

(ข้อ 5) ตัวแปรหนึ่งให้ข้อมูลบอกจำนวน และ อีกตัวหนึ่งให้ข้อมูลบอกค่า (Two Variables: One Nominal, One Interval) สัมประสิทธิ์ที่บอกความสัมพันธ์คือ

$\text{Eta}^2 (y^2)$

$\text{Omega}^2 (w^2)$

Intraclass correlation coefficient (r_i)

Kelly's epsilon² (E^2) แล้วทดสอบความมีนัยสำคัญด้วย F-Test

(ข้อ 6) ตัวแปรหนึ่งให้ข้อมูลบอกลำดับที่และอีกตัวหนึ่งให้ข้อมูลบอกค่า (Two Variables: One Ordinal, One Interval) สถิติที่เหมาะสมกรณีที่มีข้อตกลงเบื้องต้นของข้อมูลตัวแปรบอกลำดับที่ว่าการแจกแจงปกติ

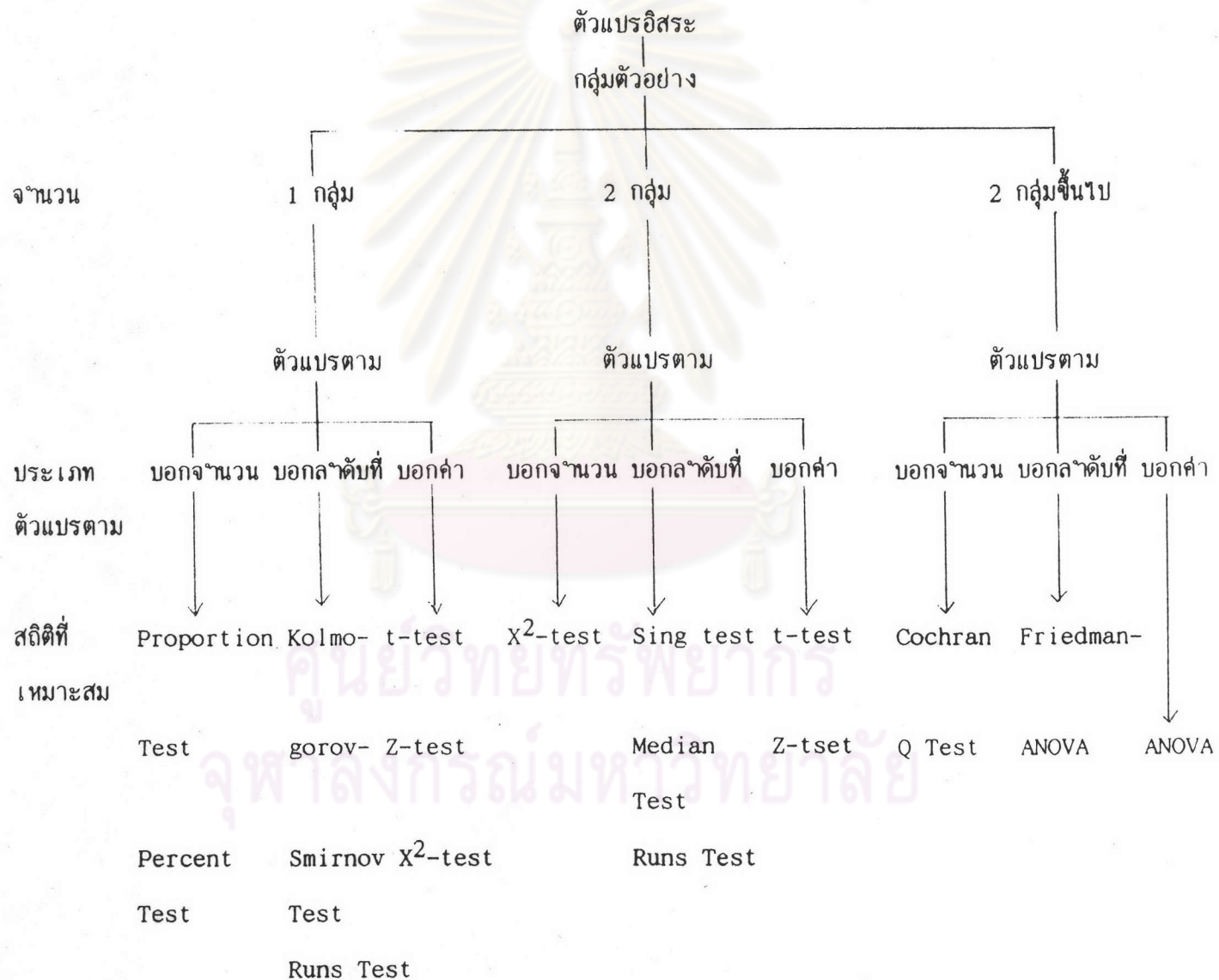
(1) Jansen's Coefficient of multiserial correlation แล้วแปลงค่าเป็น Z ทดสอบค่า Z เทียบกับตารางรังปกติ

(2) Mayer and Robinson's M_{yu} แล้วแปลงเป็นค่า Z ทดสอบค่า Z เทียบตารางรังปกติ

จุฬาลงกรณ์มหาวิทยาลัย

แผนภูมิที่ 4

การเลือกใช้เทคนิคสถิติที่ต้องการ เปรียบเทียบความแตกต่างของตัวแปรสองตัว



รายละเอียดของแผนภูมิที่ 4 มีดังนี้

(ข้อ 7) ผู้ให้ข้อมูล 1 กลุ่ม

7.1 ตัวแปรตามให้ข้อมูลบอกจำนวน สถิติที่เหมาะสม คือ

(1) Proportion Test

(2) Percent Test

7.2 ตัวแปรตามให้ข้อมูลบอกลำดับที่ สถิติที่เหมาะสม คือ

(1) Kolmogorov-Smirnov Test

(2) One Sample Runs Test

7.3 ตัวแปรตามให้ข้อมูลบอกค่า สถิติที่เหมาะสม คือ

(1) ทดสอบความแตกต่างของค่าเฉลี่ยของกลุ่มตัวอย่างกับค่าเฉลี่ยของ

ประชากร ใช้ t-test one-group

Z-test one-group

(2) ทดสอบความแตกต่างของค่าความแปรปรวนของกลุ่มตัวอย่าง

กับค่าความแปรปรวนของประชากร ใช้ X^2 -Test (parametric)

(ข้อ 8) กลุ่มผู้ให้ข้อมูล 2 กลุ่ม

8.1 กลุ่มทั้ง 2 เหมือนกันหรือเป็นกลุ่มเดียวกันแต่วัด 2 ครั้ง

8.1.1 ตัวแปรตามให้ข้อมูลบอกจำนวน

(1) McNemar test for the significance of change

8.1.2 ตัวแปรตามให้ข้อมูลบอกลำดับที่

(1) Sign Test

(2) Wilcoxon matched-pairs signed-ranks Test

8.1.3 ตัวแปรตามให้ข้อมูลบอกค่า

(1) Walsh Test

(2) Randomization Test for matched pairs

(3) t-test for paired observations

8.2 กลุ่มทั้งสองไม่เหมือนกัน เป็นอิสระต่อกัน

8.2.1 ตัวแปรตามให้ข้อมูลบอกจำนวน ใช้ X^2 -Test for two Independent samples

8.2.2 ตัวแปรตามให้ข้อมูลบอกลำดับที่

- (1) Median Test
- (2) Mann-Whitney U Test
- (3) Kolmogorov-Smirnov Two Sample Test
- (4) Wald-Wolfowitz runs Test
- (5) Moses Test of extreme reactions
- (6) Wilcoxon signed rank Test
- (7) Runs Test
- (8) Friedman Test
- (9) Kruskal-Wallis Test

8.2.3 ตัวแปรตามให้ข้อมูลบอกค่า

(1) ถ้าตั้งข้อตกลงเบื้องต้นว่าลักษณะการแจกแจงของข้อมูลเป็น
รูปร่างปกติ การเปรียบเทียบความแตกต่างของค่าเฉลี่ย (มีชนิดเลขคณิต) และ/หรือ ค่า
ความแปรปรวน ใช้ t-test

Z-test

ANOVA

Welch statistic

Brown-Forsythe statistic

Bartlett's test

(2) ถ้าไม่ได้ตั้งข้อตกลงเบื้องต้นว่าลักษณะการแจกแจงของข้อมูล
เป็นรูปร่างปกติ การเปรียบเทียบความแตกต่างของค่าเฉลี่ยหรือค่าความแปรปรวน ใช้

ANOVA

Levene's W

(ข้อ 9) ผู้ให้ข้อมูลมากกว่า 2 กลุ่มขึ้นไป



9.1 กลุ่มผู้ให้ข้อมูลเป็นกลุ่มเดียวกันหรือเหมือนกันวัดหลายครั้ง

9.1.1 ตัวแปรตามให้ข้อมูลบอกจำนวน ใช้ Cochran Q Test

9.1.2 ตัวแปรตามให้ข้อมูลบอกลำดับที่ใช้ Friedman two-way ANOVA

9.2 กลุ่มผู้ให้ข้อมูลไม่เกี่ยวข้องกัน

9.2.1 ตัวแปรตามให้ข้อมูลบอกจำนวน ใช้ χ^2 -test for independent samples

9.2.2 ตัวแปรตามให้ข้อมูลบอกลำดับที่ใช้

Extension of the median test

Kruskal-Wallis One-Way ANOVA

9.2.3 ตัวแปรตามให้ข้อมูลบอกค่า ใช้

One-Way ANOVA

Least Significant Difference Test (LSD)

Honest Significant Difference Test (HSD)

Scheffe's Method

Newman-Keuls Test

Duncan's New Multiple Range Test

Dunnett's Test for Comparison Involving a Control

Mean

การวิเคราะห์ข้อมูลตัวแปรหลายตัว

แม้ว่าในงานวิจัยจะมีจำนวนตัวแปรมากกว่า 2 ตัว แต่การวิเคราะห์ก็ทำได้ยุ่งยากไปกว่าการวิเคราะห์ตัวแปร 2 ตัว ถ้าผู้วิจัยสามารถตอบคำถามเหล่านี้ได้

1. มีการจำแนกตัวแปรทั้งหลายออกเป็นตัวอิสระ ตัวแปรตาม ตัวแปรแทรกซ้อน

ไม่จำแนก (ดูแผนภูมิที่ 1)

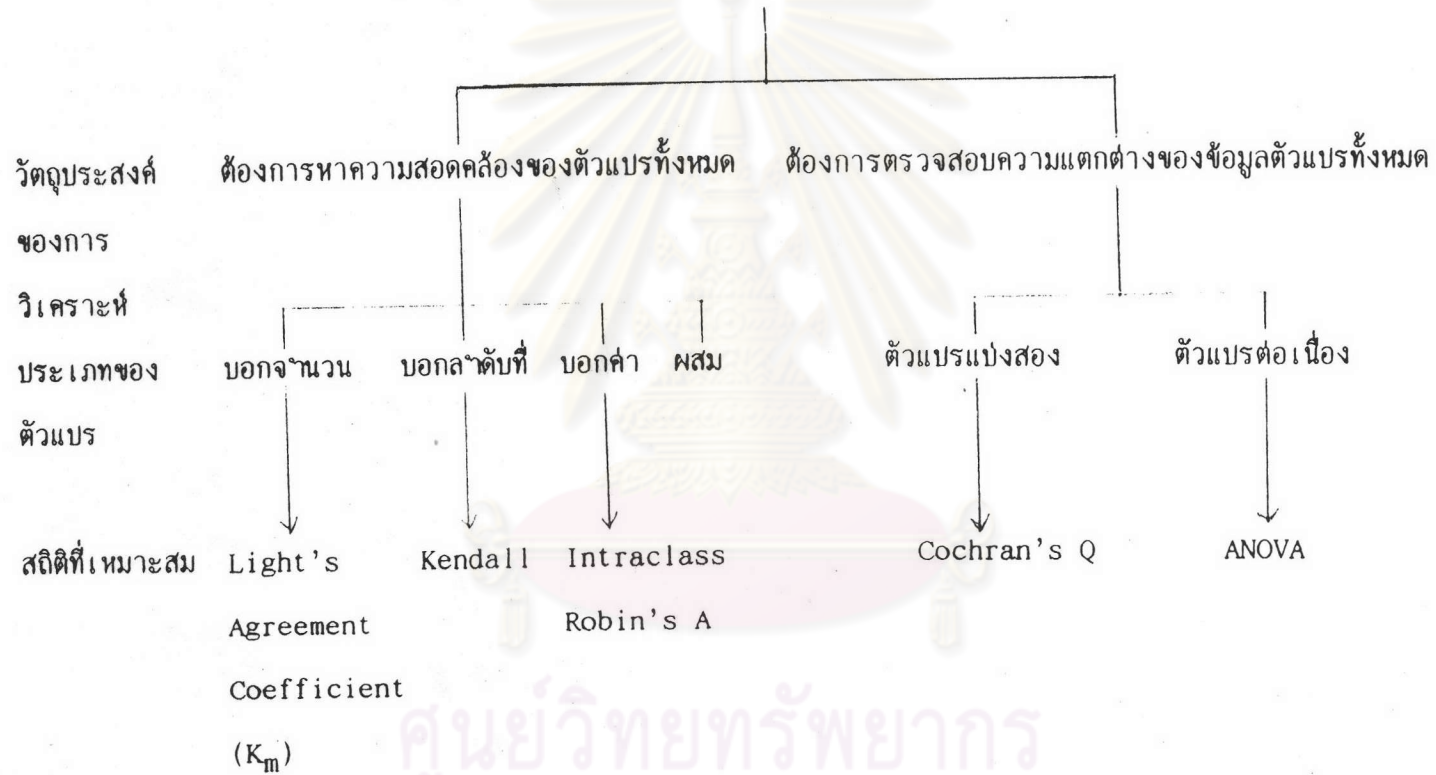
จำแนกได้ (ดูแผนภูมิที่ 7)

2. ตัวแปรทั้งหลายที่อยู่ในงานวิจัยนั้น

มุ่งศึกษาหาความสัมพันธ์ (ดูแผนภูมิที่ 5)

แผนภูมิที่ 5

การเลือกใช้เทคนิคสถิติที่ต้องการหาความสัมพันธ์ของข้อมูลตัวแปรหลายตัว



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

[] มุ่งศึกษาหาความแตกต่างกัน (ดูแผนภูมิที่ 5)

[] มุ่งศึกษาหาแบบแผนที่ซ่อนอยู่ในความสัมพันธ์ของตัวแปร
(ดูแผนภูมิที่ 6)

รายละเอียดดังนี้

รายละเอียดของแผนภูมิที่ 5

ในกรณีที่ผู้วิจัยไม่แน่ใจว่าตัวแปรใดเป็นตัวอิสระ ตัวแปรตาม ตัวแปรแทรกซ้อน จุดมุ่งหมายในการวิเคราะห์จึงแยกได้ 2 ประเด็นคือ

1. เพื่อหาค่าความสัมพันธ์ของตัวแปรทั้งหลาย กับ 2. การตรวจสอบความแตกต่างของข้อมูลตัวแปรทั้งหลาย

จึงอธิบายได้ ดังนี้

1. การศึกษาค่าความสัมพันธ์ของตัวแปรทั้งหลาย ความสัมพันธ์ที่จะหาได้ในที่นี้ คือ ความสอดคล้อง (Agreement) ระหว่างตัวแปร

(1) ถ้าตัวแปรทั้งหมดให้ข้อมูลบอกจำนวน ใช้ Light's Agreement Coefficient (K_m) แล้วเปรียบเทียบค่ากับตารางโค้งปกติ

(2) ถ้าตัวแปรทั้งหมดให้ข้อมูลลำดับที่ ใช้

Kendall's coefficient of Concordance (W) การทดสอบนัยสำคัญพิจารณาจากขนาดของกลุ่มตัวอย่าง ถ้ากลุ่มตัวอย่างมีขนาดใหญ่มากกว่า 7 ใช้การทดสอบด้วย χ^2 สำหรับค่า W ถ้ากลุ่มตัวอย่างมีขนาดเล็กกว่าหรือเท่ากับ 7 ให้ใช้ตาราง S ของ Kendall

(3) ถ้าตัวแปรทั้งหมดให้ข้อมูลบอกค่า ใช้

Intraclass Correlation Coefficient (r_i)

Robinson's A แล้วทดสอบความมีนัยสำคัญด้วย F-test

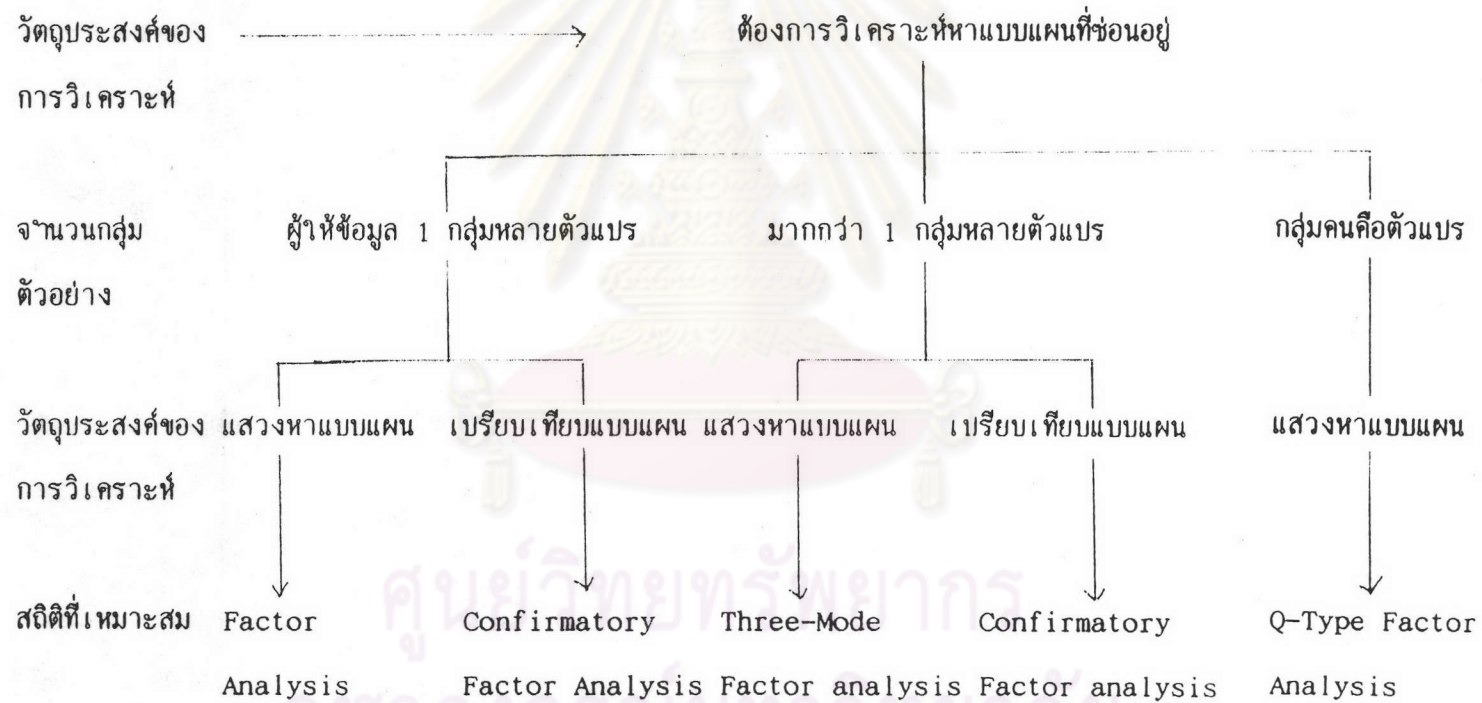
2. การตรวจสอบความแตกต่างของข้อมูลตัวแปรทั้งหลาย

(1) ในกรณีที่เป็นตัวแปรแบ่งสอง (Dichotomous Variable) ใช้ Cochran's Q

(2) ในกรณีที่เป็นตัวแปรต่อเนื่อง (Continuous Variable) ใช้การวิเคราะห์ความแปรปรวนแบบซ้ำ (Repeated Measure)

แผนภูมิที่ 6

การเลือกใช้เทคนิคสถิติที่มุ่งศึกษาแบบแผนที่ซ่อนอยู่ในความสัมพันธ์ของตัวแปร



รายละเอียดของแผนภูมิที่ 6

ในกรณีที่ไม่มีงานแก้ตัวแปรออกเป็นตัวแปรอิสระ ตัวแปรตาม ตัวแปรแทรกซ้อน นอกจากนี้ตัวแปรที่เลือกมานั้น เป็นตัวแปรที่มีความเกี่ยวข้องกันอยู่ ก่อนเลือกใช้สถิติให้เหมาะสม ควรตอบคำถามต่อไปนี้ก่อน

- [] เป็นการศึกษาหาลักษณะร่วมของคน (ข้อ 1)
- [] กลุ่มตัวอย่าง 1 กลุ่ม หลายตัวแปร (ข้อ 2)
- [] กลุ่มตัวอย่างหลายกลุ่ม หลายตัวแปร (ข้อ 3)

ประเด็นการวิเคราะห์คือ

(ข้อ 1) ในการวิเคราะห์หาแบบแผนลักษณะร่วมกันของคนหลายคน โดยคำนวณค่าความสัมพันธ์ระหว่างคนที่เรียกว่า Q-correlation การวิเคราะห์หาแบบแผนจึงใช้วิธี Q-Type Factor Analysis

(ข้อ 2) กลุ่มตัวอย่าง 1 กลุ่ม มีหลายตัวแปร โดยที่ตัวแปรดังกล่าวให้ข้อมูลบอกค่า/คะแนนและความสัมพันธ์ระหว่างตัวแปรเป็นเส้นตรงที่เรียกว่า R-Correlation การวิเคราะห์เพื่อแสวงหาแบบแผนที่ซ่อนอยู่จึงทำได้ดังนี้

(2.1) เพื่อแสวงหาแบบแผน

(2.1.1) ข้อมูลเป็นคะแนนมาตรฐานใช้ Factor Analysis of Correlation Matrix

(2.1.2) ข้อมูลเป็นคะแนนดิบใช้ Factor Analysis of Variance-covariance matrix

(2.2) เพื่อเปรียบเทียบแบบแผนที่มีอยู่แล้ว

(2.2.1) ข้อมูลเป็นคะแนนมาตรฐาน ใช้ Confirmatory Factor Analysis of a standardized variance-covariance matrix

(2.2.2) ข้อมูลเป็นคะแนนดิบใช้ Confirmatory Factor Analysis of variance-covariance matrix

ไม่ว่าจะเป็นแบบ (2.2.1) หรือ (2.2.2) ทดสอบได้ด้วย Maximum Likelihood

Chi-Square Test

(ข้อ 3) กลุ่มตัวอย่างหลายกลุ่ม หลายตัวแปรโดยที่ตัวแปรดังกล่าวให้ข้อมูลบอกค่า/คะแนนและความสัมพันธ์ระหว่างตัวแปรดังกล่าวเป็นเส้นตรง การวิเคราะห์จำแนกได้เป็น 2 ประเด็น คือ

(3.1) เพื่อแสวงหาแบบแผนของแต่ละกลุ่มพร้อมกัน

(3.1.1) Three-Mode Factor Analysis

(3.1.2) ในกรณีที่มีตัวแปรดังกล่าวไม่ได้ให้ข้อมูลบอกค่าและความสัมพันธ์ระหว่างตัวแปรไม่ใช่เส้นตรงใช้ Three-Way non-metric multi-dimensional scaling technique

(3.2) เพื่อเปรียบเทียบแบบแผนของแต่ละกลุ่มหรือแต่ละแบบแผน

(3.2.1) ข้อมูลเป็นคะแนนมาตรฐานใช้ Confirmatory Factor Analysis of Standardized variance-covariance matrix

(3.2.2) ข้อมูลเป็นคะแนนดิบ ใช้ Confirmatory Factor Analysis of variance-covariance matrix

ไม่ว่าในกรณี (3.2.1) หรือ (3.2.2) ทดสอบด้วย Maximum-Likelihood Chi-Square Test

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

แผนภูมิที่ 7



รายละเอียดแผนภูมิที่ 7

ถ้าการวิจัยมีการจำแนกตัวแปรออกเป็นตัวแปรอิสระ ตัวแปรตาม และตัวแปรแทรกซ้อน และถ้าจะวิเคราะห์ข้อมูลได้อย่างเหมาะสม ก็ต้องพิจารณาจำนวนตัวแปรตามดังนี้

- [] จำนวนตัวแปรตาม 1 ตัว (ข้อ 4)
- [] จำนวนตัวแปรตามมากกว่า 1 ตัว (ข้อ 5)
- (ข้อ 4) วนกรณีที่มีตัวแปรตาม 1 ตัว และมีตัวแปรอิสระหลายตัว

(4.1) การเปรียบเทียบค่าเฉลี่ยของตัวแปรตามอันเกิดจากตัวแปรอิสระหลายตัว โดยที่ตัวแปรอิสระอาจให้ข้อมูลบอกจำนวน แต่ตัวแปรตามต้องให้ข้อมูลบอกค่า

ใช้การวิเคราะห์ความแปรปรวนหลายทาง (Many-Way ANOVA) หรือการวิเคราะห์ความแปรปรวนร่วม (ANCOVA) หรือใช้ Profile Analysis แล้วทดสอบด้วย Wilk's Lamda, Roy's greatest root Criterion, Pillai-Bartlett's V

(4.2) การหาค่าความสัมพันธ์ระหว่างตัวแปรตาม 1 ตัว กับชุดของตัวแปรอิสระ ใช้ Multiple Correlation แล้วทดสอบด้วย F-test

(4.3) การสร้างสมการถดถอยทำนายตัวแปรตาม 1 ตัว ใช้ Multiple Regression Analysis แล้วทดสอบด้วย F-test

(4.4) ถ้าต้องการหาค่าความแปรปรวนในตัวแปรตาม เพื่ออธิบายว่าเป็นผลมาจากการเพิ่มตัวแปรอิสระแต่ละตัว ใช้ Part Correlation² แล้วทดสอบด้วย F-test

(4.5) ถ้าต้องการหาค่าความแปรปรวนในตัวแปรตาม อันเป็นผลจากการลดตัวแปรอิสระทีละตัว โดยที่ค่าความแปรปรวนที่เหลือเป็นค่าที่ไม่สามารถอธิบายได้ว่าเกิดจากตัวแปรอิสระตัวใดบ้าง ใช้ Partial Correlation² แล้วทดสอบด้วย F-test

(4.6) ในกรณีที่ต้องการวิเคราะห์ขั้นตอน (hierarchical) และถ้าตัวแปรทั้งหลายให้ข้อมูลบอกจำนวนใช้ Multidimensional contingency table analysis แล้วทดสอบด้วย χ^2 -Test

ถ้าตัวแปรทั้งหลายเป็นตัวแปรที่ค่าใช้ Multidimensional Scaling

(4.7) ถ้าตัวแปรอิสระให้ข้อมูลบอกค่า ตัวแปรตามให้ข้อมูลบอกจำนวน ความสัมพันธ์ของตัวแปรอิสระเป็นเส้นตรง และต้องการวิเคราะห์เพื่อจำแนกลักษณะต่าง ๆ ออกจากกันใช้ Multiple Discriminant Function แล้วทดสอบด้วย Wilk's lamda, Roy's greatest root criterion, Pillai-Bartlett's V

(4.8) ถ้าตัวแปรตามให้ข้อมูลบอกค่าใช้ Multiple Classification-Analysis

(4.9) ถ้าตัวแปรอิสระไม่ได้มีความสัมพันธ์เชิงเส้นตรง ใช้ Multiple curvilinear regression

(ข้อ 5) ในกรณีที่มีตัวแปรตามหลายตัว และตัวแปรอิสระหลายตัว

(5.1) ถ้าตัวแปรตามและตัวแปรอิสระให้ข้อมูลบอกค่า และมีตัวแปรแทรกซ้อนอย่างน้อย 1 ตัว ใช้ Path Analysis หรือ Structural models with latent variables

(5.2) ถ้ามีเฉพาะตัวแปรอิสระหลายตัวแปรตามเท่านั้น และต้องการหาความสัมพันธ์ระหว่างชุดของตัวแปรทั้งสองใช้ Canonical Correlation แล้วทดสอบด้วย Wilk's lamda, Roy's greatest root criterion หรือ Pillai-Bartlett's V

(5.3) ถ้าต้องการเปรียบเทียบค่าเฉลี่ย (Centroid) ของตัวแปรตามทั้งหลายอันเป็นผลมาจากตัวแปรอิสระทั้งหลาย โดยที่ตัวแปรอิสระให้ข้อมูลบอกจำนวนและตัวแปรตามให้ข้อมูลบอกค่า ใช้ Multivariate Analysis หรือ Profile analysis แล้วทดสอบด้วย Wilk's lamda หรือ Roy's greatest root Criterion หรือ Pillai-Bartlett's V

ตอนที่ 6 บทความที่เกี่ยวข้อง

John Gaito (1960) ได้เขียนบทความเรื่อง การแบ่งสเกลกับวิธีการทางสถิติ (Scales Classification and Statistics) สรุปได้ว่า

การแบ่งสเกลของข้อมูลทางจิตวิทยาเป็นมาตรา nominal, ordinal, interval และ ratio โดย Steven (1951) ได้รับการสนับสนุนอย่างมากในทางจิตวิทยาและทฤษฎีของการวัด (measurement) การแบ่งสเกลแบบนี้ได้รับการยอมรับอย่างกว้างขวางและได้รับการนำไปประยุกต์ใช้โดยนักจิตวิทยาในการบรรยาย การสอน และการพิมพ์ออกเผยแพร่ (เช่น Edwards, 1958; Guilford, 1954; Sender, 1958; Siegel, 1956) ในการพิจารณาสเกลเหล่านี้ Steven ได้กำหนดสถิติที่เหมาะสมสำหรับการนำไปใช้กับมาตราวัดแต่ละสเกล ซึ่งเกณฑ์นี้ต่อมาได้มีผลกระทบมากมายต่อผู้เขียนตำราทางสถิติ ตัวอย่างเช่น จากพื้นฐานรูปแบบการแบ่งสเกลเหล่านี้ในตำราของ Sender

(1958) และ Siegel (1956) ได้มีการนำไปใช้อย่างกว้างขวางเป็นทฤษฎีที่มีคุณค่าอย่างไรก็ตามเรารู้สึกว่าการประยุกต์ใช้การแบ่ง scale นั้นนำไปสู่การไม่เข้าใจบางประการในการพิจารณาเลือกใช้เทคนิคสถิติอย่างเหมาะสม และเป็นการทำให้อهميةอย่างมากในการใช้เทคนิคนี้พาราเมตริกในการวิจัยทางจิตวิทยา

Siegel (1956) ยืนยันว่าสถิติทดสอบพาราเมตริกควรใช้กับข้อมูลกึ่งอันตรภาค (sub interval) โดยยกเอาข้อกำหนดของ interval scale เป็นข้อตกลงเบื้องต้นข้อหนึ่งในการวิเคราะห์ความแปรปรวน อย่างไรก็ตามข้อตกลงนี้ก็ไม่มีพบในพื้นฐานทางคณิตศาสตร์ของข้อตกลงเบื้องต้น (Eisenhart, 1947) ในข้อสังเกตของนักสถิติ Kempthorne (1955) หลังจากแสดงหลักการทางคณิตศาสตร์ว่าด้วยการวิเคราะห์ความแปรปรวนสามารถประมาณค่าการทดสอบแบบสุ่มโดยมีข้อสังเกตบางประการจากนักสถิติส่วนใหญ่ที่ระบุว่า ระดับความมีนัยสำคัญของการวิเคราะห์ความแปรปรวนซึ่งทดสอบความแตกต่างระหว่าง treatment มีผลกระทบเล็กน้อยจากการวัดสเกลและการพิจารณาที่สำคัญในการใช้สำหรับการวิเคราะห์ความแปรปรวนซึ่งไม่แน่ใจในคุณสมบัติของ scale แต่ข้อมูลนั้นสามารถทำให้มีการแจกแจงแบบปกติ รวมกับการจัดทำให้ใกล้เคียงกับข้อตกลงอื่น ๆ ของการเป็นอิสระและ homogeneity ของความคลาดเคลื่อน ยิ่งกว่านั้นจะต้องเน้นว่าการแจกแจงแบบปกติจะสามารถอ้างอิงไปถึงการแจกแจงส่วนต่าง ๆ ของข้อมูลซึ่งเป็นการใช้การประมาณค่าความคลาดเคลื่อนที่เหมาะสม การพิจารณาเช่นนี้เป็นสิ่งสำคัญ เพราะว่าการออกแบบการวิเคราะห์ความแปรปรวนภายในกลุ่ม แบบ Multivariate จะไม่มีการแจกแจงแบบปกติ แต่การพิจารณาปฏิสัมพันธ์ของส่วนย่อยอาจจะเป็นการแจกแจงแบบปกติได้

ดูเหมือนว่าความคิดเห็นเกี่ยวกับความสัมพันธ์ระหว่างสเกลในหลายรูปแบบกับกระบวนการทางสถิติที่เหมาะสมสามารถใช้เป็นแนวทางที่ใช้ในการพิจารณาได้ในบางกรณีหลักเกณฑ์เกี่ยวกับข้อมูลที่กำหนดขึ้นไว้ อาจจะเป็นตัวชี้ว่าการยึดมั่นอย่างไม่ยืดหยุ่นกับวิธีการทางสถิติไม่ใช่ว่าจะเป็น และไม่ก่อให้เกิดผลเสียหายกับข้อมูลด้วย ในกรณีอื่น ๆ อาจจะมีการแสดงความคิดเห็นและอ้างความสัมพันธ์กันของข้อมูลเกี่ยวกับวิธีการวิเคราะห์ความแปรปรวน



นอกจากนี้ข้อมูลที่เหมือนกันอาจจะพิจารณาว่ามีคุณสมบัติเป็นข้อมูลในสองระดับหรือมากกว่าก็ได้ ขึ้นอยู่กับหลักการที่ใช้ในการพิจารณา ตัวอย่างเช่น ถ้าเราพิจารณาการตอบของคำถามข้อหนึ่ง คุณสมบัติของข้อมูลอาจจัดเป็นมาตราแบ่งกลุ่ม (nominal scale) และอื่น ๆ ซึ่งอาจจะถูกหรือผิดก็ได้ อย่างไรก็ตามถ้าเรามุ่งพิจารณาข้อมูลรวมของข้อคำถามหนึ่ง หรือ ข้อมูลรวมของข้อคำถามทั้งหมด เราจะมีข้อมูลอย่างน้อยที่สุดคือมาตราจัดอันดับ (ordinal scale) ลักษณะนี้คล้ายกับสถานะภาพของการประมาณค่าเป็นการแจกแจงแบบไบโนเมียล (ซึ่งเป็นการแจกแจงที่ข้อมูลมีคุณสมบัติเป็นมาตราแบ่งกลุ่มหรือมาตราจัดอันดับ) โดยจะมีการแจกแจงแบบปกติ (ซึ่งเป็นการแจกแจงที่ข้อมูลมีคุณสมบัติอย่างน้อย เป็นมาตราจัดอันดับ หรืออาจจะเป็นมาตราอันตรภาคก็ได้) เมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้นและระดับความมีนัยสำคัญ (p) เข้าใกล้ 0.5 มีตัวอย่างอื่น ๆ อีก เช่น ความสัมพันธ์ระหว่าง sign test กับ binomial distribution test สำหรับข้อมูลที่อาจจะวัดเป็น 2 ระดับได้ ก็ให้เหตุผลเช่นเดียวกัน

ปรากฏว่าในข้อตกลงของสถิติบรรยาย อาจทำให้เกิดการใช้สถิติแบบพาราเมตริกได้ สำหรับข้อมูลกึ่งมาตราอันตรภาค (sub interval type data) ซึ่งเป็นสิ่งที่ทำได้เหมาะสม อย่างไรก็ตามข้อตกลงของการทดสอบความมีนัยสำคัญบางประการก็ไม่เหมาะสม ถึงแม้ว่าสถิติบรรยายจะไม่เกี่ยวข้องกับกระบวนการทางสถิตินั้นก็ตาม สำหรับการทดสอบบางอย่าง หลักฐานทางคณิตศาสตร์เป็นสิ่งแสดงข้อตกลงเบื้องต้นแต่ไม่มีสิ่งเกี่ยวข้องกับหลักฐานของคุณสมบัติเฉพาะสเกล ถึงแม้ว่าหลักเกณฑ์การวิจัยที่ตั้งขึ้นจะทำให้เกิดผลเล็กน้อยเกี่ยวกับข้อมูลเมื่อข้อตกลงเบื้องต้นไม่ครบถ้วน แต่ผู้ทำการวิจัยอาจใช้สถิติพาราเมตริก และยอมรับผลที่เกิดจากการเบี่ยงเบนจากข้อตกลงเบื้องต้นในการแปลความหมายได้

มีข้อสังเกตว่าบางคนอาจใจที่จะยอมรับวิธีการของสถิติแบบพาราเมตริก (เช่น Guilford, 1954; McNemar, 1957, 1958; Savage, 1957) โดย Grant (1959; หน้า 137) สรุปไว้ในหนังสือ Annual Review of Psychology ว่า "มีกลุ่มคนที่มีแนวความคิดในทางนิเสธที่จะใช้วิธีการสถิติแบบพาราเมตริก"

John Gaito (1960) ได้กล่าวไว้ว่า เรื่องราวเกี่ยวกับคุณสมบัติของสเกลและ

วิธีการทางสถิติ ควรจะกำหนดไว้เป็นแนวทางในการเลือกทำให้เหมาะสม และเนื้อหาเกี่ยวกับข้อตกลงเบื้องต้นของวิธีการทางสถิติและผลของการวิจัยที่คำนึงถึงความไม่สอดคล้องของข้อตกลงเบื้องต้น จะเป็นตัวกำหนดในการตัดสินใจเลือกใช้วิธีการทางสถิติ

Bela O. Baker, Gurtis O. Hardyck และ Petrinovick (1966) ได้เขียนบทความเรื่อง การเปรียบเทียบระดับการวัดข้อมูลที่อ่อน กับ สถิติที่แข็งแกร่ง (Weak Measurements VS. Strong Statistics : An Empirical Critique of S.S. Stevens' Proscription on Statistics) สรุปได้ว่า

ความขัดแย้งกันระหว่างพวกที่ Lubin (1962) เรียกว่า "พวกนักทฤษฎี weak measurement" และพวกที่อาจถูกเรียกว่าเป็น "พวก strong statistics" ได้มีอยู่เป็นเวลานานหลายปีและแทบจะไม่มีที่ท่าว่าแต่ละฝ่ายจะเปลี่ยนแปลงทัศนคติของตน Steven (1951) กล่าวถึงทัศนคติที่ว่า การวัดระดับข้อมูลเป็นแบบอย่างของความสัมพันธ์ของวัตถุ และโดยส่วนใหญ่มักจะเป็นแบบอย่างที่ไม่ดีซึ่งจะทำให้บิดเบือนไปจากความจริง หากว่าคะแนนที่ได้ถูกนำมารวมกันเมื่อควรจะเพียงถูกนับเท่านั้น อย่างน้อยที่สุดตำราสถิติ 2 เล่ม ในปัจจุบันที่เขียนขึ้นโดยนักจิตวิทยา (Senders, 1958; Siegel, 1956) ได้เสนอข้อคิดเห็นนี้ว่าเป็นกฎเกณฑ์

บรรดานักจิตวิทยาที่ไม่เห็นด้วยกับแนวความคิดนี้ คือ Lord (1953), Burke (1953), Anderson (1961), Mc Nemor (1962) และ Heys (1963) ต่างก็โต้แย้งว่า วิชาสถิติเกี่ยวข้องกับตัวเลขมากกว่าสรรพสิ่ง และคุณสมบัติต่าง ๆ ที่เป็นทางการของการวัดระดับข้อมูลไม่น่าที่จะมีอิทธิพลต่อการเลือกสถิติ Savage (1957) ซึ่งเป็นนักสถิติได้สนับสนุนแนวความคิดนี้ โดยระบุว่า "ข้าพเจ้าไม่เห็นว่าจะมีเหตุผลใดที่จะจำกัดขอบเขตทางสถิติอยู่เพียงแต่สิ่งที่พัวพันกับการคำนวณตัวเลข ซึ่งสอดคล้องกันกับคุณสมบัติต่าง ๆ ของการวัดระดับข้อมูลของจำนวนที่ได้ทำการสังเกต ในอีกนัยหนึ่ง การทดสอบทางสถิติสามารถตอบปัญหาที่ได้ถูกออกแบบมาให้ตอบ ไม่ว่าการวัดจะเป็นแบบ weak หรือ strong"

ในการกล่าวถึงการวัดระดับข้อมูลอย่างมากมาย Stevens (1951) ได้แบ่งมาตราออกเป็น 4 ประเภท คือ มาตรานับกลุ่ม มาตรารัดอันดับ มาตรานันตรภาค และ

มาตราอัตราส่วน และระบบการคำนวณตัวเลข (และสถิติ) ซึ่งสามารถนำมาใช้กับมาตราแต่ละชนิดได้ มาตราแบ่งกลุ่มเป็นเพียงบอกชื่อประเภทต่าง ๆ และสามารถนำมาใช้ได้กับการนับและการแจกแจงความถี่เท่านั้น มาตราจัดอันดับได้ถูกพัฒนาขึ้นโดยการสาธิตว่า วัตถุบางอย่างมีคุณสมบัติเฉพาะตัวมากกว่าวัตถุชนิดอื่น และมีค่าเป็นตัวเลขตามลำดับเท่านั้นโดยในบรรดาวัตถุเหล่านั้นไม่มีหน่วย ตัวเลขของมาตราจัดอันดับไม่สามารถนำมาบวก ลบ คูณ หรือหารได้ แต่ทว่าตัวเลขของมาตราเหล่านี้สามารถนำมาใช้ได้ในสถิติว่าด้วยลำดับ เช่น ค่ามัธยฐาน หรือสหสัมพันธ์ลำดับชั้น (rank - order correlation) มาตราอันตรภาคชี้ให้เห็นขนาดที่เพิ่มขึ้นเท่า ๆ กัน ของคุณสมบัติของวัตถุที่เป็นการเพิ่มขึ้นที่เท่ากัน การเพิ่มขึ้นจำนวน 1 หน่วย ในแต่ละบริเวณของมาตรที่บอกช่วงคือการเพิ่มขึ้นที่เท่ากันของคุณสมบัติของวัตถุ ดังเช่นการเพิ่มขึ้นของ 1 หน่วย ในแต่ละบริเวณอื่นของมาตร เราสามารถนำคะแนนจากมาตราอันตรภาคมาบวกและลบกันได้ และก็สามารถใช้สถิติ เช่น ค่ามัธยฐาน เลขคณิตส่วนเบี่ยงเบนมาตรฐานและสหสัมพันธ์ที่เป็น product - moment ได้ด้วย มาตราอัตราส่วน เมื่อเติมศูนย์ที่แท้จริงจะเข้าไปที่ช่วงต่าง ๆ ที่เท่ากัน และสามารถนำมา คูณ หาร และกระทำการสถิติที่ซับซ้อนได้ ถึงแม้ Steven จะพัฒนาหลักเหตุผลของการเชื่อมโยงการวัดและสถิติเข้าด้วยกัน ซึ่งเกือบเป็นไปในรูปแบบของสถิติบรรยาย Steven ได้แนะนำถึงเรื่องการทดสอบสมมติฐานในการพูดถึงความไม่แปรปรวนของผลลัพธ์ภายใต้การแปลงข้อมูล (1951, 1959) Steven กล่าวว่า "กฎพื้นฐานก็คือ เมื่อได้วัดวัตถุ 1 เซต โดยการระบุหมายเลขตามกฎเกณฑ์ เรามีอิสระที่จะเปลี่ยนการระบุตัวเลขนั้นโดยการแปลงข้อมูล โดยขึ้นอยู่กับว่าจะจัดอยู่ในกลุ่มไหน จะมีผลกระทบต่อทางสถิติแต่จะไม่มีผลกระทบต่อสิ่งอื่น ๆ"

หากมีการทดสอบนัยสำคัญตัวพารามิเตอร์ (parameter) ดังเช่น t หรือ F การแปลงข้อมูลที่ยินยอมได้ก็คือ แบบชนิดเส้นตรง (linear transformation) โดยวิธีการเช่นนี้เท่านั้นที่จะได้ผลลัพธ์อันไม่แปรปรวนในกลุ่มที่ทำการเปรียบเทียบ สิ่งที่เป็นนัยแห่งความคิดเห็นนี้ ซึ่ง Steven ไม่สามารถแสดงให้เห็นชัดเจนได้ก็คือ หากว่าเรามองดูมาตราว่าเป็นตัวอย่างของความสัมพันธ์ต่าง ๆ ของวัตถุ เมื่อนั้นการแปลงข้อมูลก็คือ การแปลงความสัมพันธ์เหล่านั้น นั่นก็คือปัญหาแห่งความ

ใหม่แปรปรวนของผลลัพธ์ภายใต้การแปลงข้อมูล ทำให้เราตั้งคำถามว่า เราสามารถกระทำการตัดสินใจที่ถูกต้องเกี่ยวกับลักษณะของความเป็นจริงหรือไม่ หากว่าเราไม่สนใจต่อลักษณะของมาตราที่ใช้วัด เมื่อเราใช้การทดสอบทางสถิติใช่หรือไม่

นักวิจารณ์หลายคนไม่ได้คำนึงถึงแนวความคิดของ Steven ในแง่ Anderson (1961) ได้เลิกเกี่ยวข้องกับความสัมพันธ์ระหว่างการวัดการใช้สถิติ t แต่การอธิบายกันถึงคำถามที่เกี่ยวกับความใหม่แปรปรวนของผลลัพธ์ภายใต้การแปลงข้อมูลที่รุนแรงและได้อธิบายกันยาวนานก่อนที่จะสรุปว่า "ปัญหาของการฝึกฝนของความใหม่แปรปรวนอยู่ในเหนือระบบการแบ่งประเภทของมาตราที่ใช้วัด"

แง่มุมต่าง ๆ ของปัญหาที่เกี่ยวข้องกันกับสถิติบรรยายและสถิติอ้างอิงมีดังต่อไปนี้ ปัญหาสำหรับสถิติบรรยายดังที่เสนอโดย Steven (1951, 1959, 1960) เป็นเรื่องของความสัมพันธ์ของค่าสถิติโดยเฉพาะที่คำนวณบนการวัดไปยังค่าของสถิติอื่นเดียวกันที่คำนวณภายใต้เงื่อนไขของการวัดที่สมบูรณ์ การถกเถียงกันได้พูดกันไปถึงรูปแบบของการวัด นอกเหนือไปจากภายใต้คุณสมบัติของวัดอุปสรรคที่จะวัด อย่างน้อยที่สุดก็คือความถูกต้องของสถิติหรือพูดอีกอย่างหนึ่งก็คือ ลักษณะนี้เกี่ยวข้องกับความจริงในการวัด

โดยสรุปแล้วในทัศนะของ Steven (1951, 1959, 1962) และผู้สนับสนุนคือ Senders (1958), Siegel (1956) และ Stake (1960) ซึ่งว่า เมื่อใช้ t -test รูปแบบการวัดก็ควรจะมีช่วงที่เท่ากัน ซึ่งเป็นการแปลงค่าในเชิงเส้นของตัวแปรที่ทำการวัด มิฉะนั้นสถิติจะได้รับผลกระทบ ทัศนะนี้อาจจะถูกถ้ามีการกำหนดการเลือกใช้สถิติบรรยายเฉพาะแบบใดแบบหนึ่ง ซึ่งดูเหมือนจะมีนัยสำคัญในการประมาณค่าความคลาดเคลื่อนมาตรฐาน แต่ไม่ถูกต้องเมื่อใช้กับสถิติอ้างอิง

Bela O. Baker และคณะ สรุปว่าในปัจจุบันนี้ strong statistics เช่น การทดสอบค่าของ t ก็เกินพอที่จะรับมือกับ weak measurement ได้ และโดยมีข้อสงวนบางส่วนที่เล็กน้อยที่ความน่าจะเป็นที่ประเมินจากการแจกแจงค่าของ t อาจได้รับผลกระทบเพียงเล็กน้อยจากชนิดของมาตราการวัดที่ใช้

Paul Leslie Gardner (1975) ได้เขียนบทความเรื่อง สเกลและวิธีการทางสถิติ (Scales and Statistics) สรุปได้ว่า

เป็นเวลากว่า 25 ปีที่ผ่านมาตั้งแต่เอกสารเกี่ยวกับความสัมพันธ์ระหว่างสเกลในการวัดและสถิติที่เหมาะสมของ Steven พิมพ์ออกเผยแพร่ การถกเถียงโต้แย้งกันเกี่ยวกับลักษณะของความสัมพันธ์กันระหว่างมาตราจัดอันดับกับมาตราอันตรภาค (Ordinal and Interval Scales) และสถิติทดสอบพารามेटริกกับนันทพารามेटริกก็ยังไม่ยุติลงได้ ความแตกต่างระหว่างมาตราจัดอันดับกับมาตราอันตรภาค ไม่ใช่ความแตกต่างในลักษณะค่ากับขาว มีหลายครั้งที่การสรุประดับการวัดสเกลเป็นไปในลักษณะคลุมเครือ และSteven (1946) แบ่งสเกลการวัดเป็น 4 ระดับ คือ มาตราแบ่งกลุ่ม (Nominal) มาตราจัดอันดับ (Ordinal) มาตราอันตรภาค (Interval) และมาตราอัตราส่วน (Ratio) ความแตกต่างระหว่างมาตราจัดอันดับกับมาตราอันตรภาค เป็นปัญหาที่มีการอภิปรายมากที่สุดในการวัดทางจิตวิทยา มาตราจัดอันดับได้รับการพิจารณาว่า เป็นการวัดลำดับ เช่นเดียวกับวิธีที่กำหนดรูปแบบตัวเลข การวัดระยะห่างของมาตราจัดอันดับเป็นการกำหนดลักษณะที่เป็นลำดับชั้น (Rank Order) ซึ่งไม่สามารถบอกระยะห่างระหว่างช่วงได้ และไม่สามารถบอกความแตกต่างของความถี่ในสิ่งที่วัดได้ การวัดระยะห่างของมาตราอันตรภาคมีความหมายมากกว่าการจัดลำดับของเรื่องที่ถูกถาม มาตราอันตรภาคมีช่วงห่างแต่ละหน่วยเท่ากัน ซึ่งแตกต่างจากมาตราจัดอันดับ ระยะห่างที่สัมพันธ์กันของช่วงระหว่างมาตราจัดอันดับและมาตราอันตรภาคอาจจะแปลความหมายได้ ซึ่งความหมายในที่นี้ให้พิจารณาการแจกแจงของความถี่

มาตราการวัดแต่ละชนิดมีลักษณะที่สามารถใช้หลักการทางคณิตศาสตร์เป็นมาตราการวัดอีกแบบหนึ่งได้ เช่น ใช้วิธีการแปลงเชิงเส้น (Linear Transformations) กับมาตราอันตรภาคได้ เนื่องจากคุณสมบัติของช่วงของมาตราไม่ได้รับการกระทบกระเทือน และใช้การแปลงที่ไม่ใช่เชิงเส้น (Nonlinear Transformation) กับมาตราจัดอันดับได้แต่ไม่สามารถใช้กับมาตราอันตรภาคได้ โดยทั่วไปมาตราจัดอันดับสามารถแปลงได้โดยอาศัยฟังก์ชันเพิ่มและมาตราอันตรภาคใช้การแปลงเชิงเส้น

จุดที่สร้างความแตกต่างกันระหว่างมาตราจัดอันดับกับมาตราอันตรภาคในทัศนะของ Steven คือการใช้สถิติที่เหมาะสมที่แตกต่างกัน Steven พิจารณาว่า ควรใช้สถิติที่เหมาะสมกับระดับการวัดของข้อมูล นอกเสียจากว่าได้มีการแปลงคะแนนซึ่งทำให้ระดับการวัดไม่แตกต่างกัน Anderson (1961) ได้อธิบายว่า การใช้วิธีการทางสถิติกับค่าของมาตราที่มาจากการแปลงนั้นสามารถใช้ได้

ข้อโต้แย้งนี้นำไปสู่การจำแนกวิธีการทางสถิติเป็น 2 ชนิด คือ พาราเมตริก กับนอปปาราเมตริก สถิติพาราเมตริก เช่น t - test และ F - test ที่เรียกเช่นนี้ เพราะมีข้อกำหนดว่าต้องใช้กับค่าพารามิเตอร์อย่างน้อย 1 ตัว เช่น ค่าของประชากร ซึ่งตามปกติแล้วสถิติตกลงว่า กลุ่มตัวอย่างที่จะเปรียบเทียบเป็นกลุ่มตัวอย่างจากประชากรที่มีการแจกแจงแบบปกติ (Normal Distribution) และระดับการวัดข้อมูลต้องอยู่ในมาตราอันตรภาค สถิตินอปปาราเมตริก เช่น ANOVA by ranks ไม่ต้องใช้พารามิเตอร์ ไม่มีข้อตกลงว่า ระยะห่างระหว่างสเกลต้องเท่ากันและไม่กำหนดรูปร่างการแจกแจงของประชากร จึงอาจจะเรียกชื่อได้ 2 ชื่อ คือ การแจกแจงอิสระกับการทดสอบลำดับ

ความเกี่ยวพันกันของมาตราอันตรภาคกับสถิติพาราเมตริกได้รับการอธิบายในหนังสือของ Siegel (1956) ซึ่งแสดงอย่างแน่ชัดว่า ห้ามใช้สถิติพาราเมตริกกับข้อมูลมาตราจัดอันดับ ถ้ายอมรับความเห็นของ Siegel สถิติพาราเมตริก เช่น t - test, product - moments correlation, การวิเคราะห์องค์ประกอบ (Factor Analysis) และ ANOVA จะใช้กับข้อมูลที่ต่ำกว่ามาตราอันตรภาคไม่ได้ ซึ่งจะมีผลกระทบกับการวิจัยในด้านพฤติกรรมศาสตร์ว่า ไม่เหมาะที่จะใช้สถิติพาราเมตริก

ระหว่างหลายทศวรรษที่ผ่านมา วรรณกรรมที่มีขอบข่ายกว้างขวางได้มีการพัฒนาเคียงคู่ไปกับปัญหาความกว้างของสเกลและสถิติที่เหมาะสม การโต้แย้งจุดยืนของ Steven และ Siegel ก็วนเวียนอยู่กับประเด็น 2 ประเด็น คือ

1. ความแตกต่างระหว่างมาตราจัดอันดับกับมาตราอันตรภาค
 2. ความต้องการของข้อตกลงเบื้องต้นที่จำเป็นสำหรับสถิติพาราเมตริก
- ในประเด็นที่ 1 Gaito (1960) ชี้แจงว่าข้อมูลแบบเดียวกัน อาจพิจารณาว่ามีคุณสมบัติของ 2 สเกลหรือมากกว่า ขึ้นอยู่กับเนื้อหาที่จะนำมาใช้พิจารณา

การตอบข้อคำถามอย่างใดอย่างหนึ่ง (ถูกหรือผิด) ย่อมเป็นมาตราแบ่งกลุ่ม (Nominal Scale) โดยแท้ แต่เมื่อรวมหลาย ๆ ข้อคำถามเข้าด้วยกัน คือการพิจารณาข้อมูลรวม จะเป็นข้อมูลในมาตราจัดอันดับ ซึ่งไกรดี กล่าวไว้ว่า "เป็นการจัดสถานการณ์ทำให้เป็นการแจกแจงแบบไบนอมิเยล (ซึ่งข้อมูลมีคุณสมบัติเป็นมาตราแบ่งกลุ่มหรือมาตราจัดอันดับเป็นอย่างน้อย) ใกล้เคียงกับการแจกแจงแบบปกติ (ซึ่งข้อมูลมีคุณสมบัติเป็นมาตราจัดอันดับหรือมาตราอันตรภาคเป็นอย่างน้อย) เมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้นและ p เข้าใกล้ 0.5 ตัวอย่างอื่น ๆ เช่น ความสัมพันธ์ระหว่างการทดสอบเครื่องหมาย (Sign Test) กับการทดสอบการแจกแจงแบบไบนอมิเยล (Binomial Distribution Test) ซึ่งให้ข้อมูลเหมือนกัน การทดสอบทั้งสองวิธีดังกล่าวจะให้ผลเหมือนกัน"

ในประเด็นที่ 2 ความต้องการเกี่ยวกับข้อตกลงเบื้องต้นของการใช้สถิติแบบพาราเมตริกเป็นความจำเป็นที่ยังเป็นจริงอยู่ ประเด็นที่ 2 นี้ มีความสำคัญมาก Siegel(1956) ได้กำหนดข้อตกลงเบื้องต้น 5 ประการ ในการใช้สถิติพาราเมตริก เช่น t - test และ F - test ในการเปรียบเทียบข้อมูลจากการทดลอง ข้อตกลงเบื้องต้น 5 ประการมีดังนี้

1. การสังเกตต่าง ๆ จะต้องเป็นอิสระ นั่นคือ การสุ่มตัวอย่างจากประชากรจะต้องไม่ลาเอียงเข้ากลุ่มใดกลุ่มหนึ่งโดยเฉพาะ
2. การสังเกตต่าง ๆ จะต้องมาจากประชากรที่มีการแจกแจงแบบปกติ
3. ประชากรจะต้องมีความแปรปรวนเท่ากัน
4. ตัวแปรที่เกี่ยวข้องจะต้องอยู่ในมาตราอันตรภาคเป็นอย่างน้อย ซึ่งสามารถใช้วิธีการทางคณิตศาสตร์ (เช่น การบวก การหาร การหาค่าเฉลี่ย และอื่น ๆ) กับข้อมูลได้
5. ค่าเฉลี่ยของประชากรจะต้องเป็นเชิงเส้น ซึ่งข้อตกลงเบื้องต้นข้อ 5 นี้ใช้เฉพาะการออกแบบชนิด Fractional Anova ซึ่งไม่ใช้กับ t - test หรือ การวิเคราะห์ความแปรปรวนทางเดียว (One - way Anova)

Gaito (1959) วิเคราะห์ว่า ข้อตกลงเบื้องต้นข้อ 1 เป็นความจำเป็น และเป็นความต้องการของสถิติพาราเมตริกด้วย ข้อตกลงเบื้องต้นข้อ 5 ไม่มีความจำเป็น

และยังมีความสับสนเกี่ยวกับการตีความหมายของข้อตกลงเบื้องต้นนี้ การอธิบายเกี่ยวกับความตรงของข้อตกลงเบื้องต้นข้อ 2 และ 3 ยังเกี่ยวข้องกับความแกร่ง (Robustness) ตามที่ Labovitz (1967) กล่าวไว้ว่า "ความแกร่งนั้นเป็นความสามารถของสถิติทดสอบที่จะทดสอบข้อมูลเพื่อให้บรรลุถึงข้อสรุปที่ได้อนุมานไว้เมื่อมีการฝ่าฝืน (Violate) ข้อตกลงเบื้องต้นบางข้อหรือหลายข้อ" สำหรับข้อตกลงเบื้องต้นข้อ 4 หมายถึงความต้องการเกี่ยวกับการวัดระดับข้อมูลมาได้เป็นส่วนหนึ่งของการใช้สถิติแบบพาราเมตริก แต่ตามความเห็นของ Siegel แล้ว ข้อนี้ย่อมเป็นข้อตกลงเบื้องต้นที่สำคัญข้อหนึ่ง จากข้อตกลงเบื้องต้นข้อนี้ที่ทำให้ Siegel โต้แย้งว่าสถิติแบบนี้พาราเมตริกจะต้องใช้กับข้อมูลแบบจัดอันดับ และไกรดี (1959) ได้ชี้ให้เห็นว่าข้อตกลงเบื้องต้นข้อนี้เป็นอุปสรรคต่อการใช้สถิติพาราเมตริกในการวิจัยทางจิตวิทยา ซึ่งข้อมูลส่วนใหญ่เป็นแบบกึ่งอันตรภาค (sub - interval)

Baker, Hardyck and Petrinovich (1966) กล่าวว่า "สถิติใช้กับจำนวนมากกว่าใช้กับสิ่งของ คุณสมบัติของการวัดระดับข้อมูลจึงไม่ควรมีอิทธิพลต่อการเลือกใช้สถิติ" ตามนัยการโต้เถียงกันนี้ ข้อตกลงเบื้องต้นข้อ 4 เริ่มจะสูญอำนาจลงในการนำไปปฏิบัติ เพราะว่าสถิติพาราเมตริกมีความแกร่งสามารถนำไปใช้กับข้อมูลแบบจัดอันดับซึ่งปฏิบัติเสมือนว่าเป็นข้อมูลแบบอันตรภาคได้

ในการพิจารณาข้อโต้แย้งทุก ๆ ข้อด้วยกันแล้ว ดูเหมือนว่าจะมีความชัดเจนว่าปัญหาของพาราเมตริกและนินพาราเมตริกไม่ควรวิพากษ์วิจารณ์กันอีกเหมือนกับที่คิดกันเมื่อสองทศวรรษที่แล้ว Heerman และ Braskamp (1970) ได้พิจารณาข้อโต้แย้งต่าง ๆ ของปัญหานี้ และกล่าวว่า "โดยสรุปแล้วจุดสำคัญของการโต้แย้งในเรื่องการใช้สถิตินินพาราเมตริกในที่ควรใช้สถิติพาราเมตริกไม่มีข้อบังคับ ผู้วิจารณ์และผู้พิจารณาเกี่ยวกับเรื่องนี้ดูจะเห็นด้วยว่า แบบของสเกลไม่เกี่ยวข้องกับการเลือกวิธีการทางสถิติ ถึงแม้ว่าสถิติพาราเมตริกต้องการข้อตกลงเบื้องต้นมากกว่าสถิตินินพาราเมตริกก็ตาม การล้มเหลวที่จะพบกับข้อตกลงเบื้องต้นของสถิติพาราเมตริกก็ปรากฏว่าไม่มีผลกระทบอย่างจริงจัง" McNemar (1969) เห็นด้วยและกล่าวว่า ไม่ว่าจะ เป็นสถิติทดสอบ F, t หรือ Z ก็ตาม เมื่อพิจารณาถึงการใช้ค่าเฉลี่ย เลขคณิตและความแปรปรวน

จะสามารถเข้าใจได้อย่างปลอดภัย เมื่อสเกลที่ใช้วัดเป็นสเกลทางจิตวิทยา ซึ่งบางครั้งจะอยู่ระหว่างมาตราจัดอันดับกับมาตราอันตรภาค ปัญหาที่จะมีว่าสถิติทดสอบ F , t และ z จะเป็นไปตามการแจกแจงของกลุ่มตัวอย่างเมื่อข้อมูลไม่ได้อยู่ในมาตราอันตรภาคได้หรือไม่ คำตอบก็คือ "ได้" โดยกำหนดว่า การแจกแจงของข้อมูลไม่ได้แตกต่างจากรูปแบบปกติ

ข้อโต้แย้งต่าง ๆ ที่เสนอไว้ในบทความนี้ Gardner (1975) สรุปเป็นประเด็นไว้ดังนี้

1. คำถามไม่ว่าจะเป็นการกำหนดข้อมูลเป็นแบบจัดอันดับหรืออันตรภาคย่อมมีความสำคัญ ถ้าเราปรารถนายืดความหมายไว้กับชุดของตัวเลข แต่จะไม่มีผลสำคัญ ถ้าเราปรารถนาที่จะประเมินว่า ตัวเลขสองชุดนั้นมาจากประชากรกลุ่มเดียวกัน
2. ความแตกต่างระหว่างมาตราจัดอันดับกับมาตราอันตรภาคยังไม่เด่นชัด
3. ข้อโต้แย้งบางข้อซึ่งเน้นการยอมรับว่า วิธีการของพาราเมตริกต้องการข้อมูลในระดับมาตราอันตรภาค เป็นปรากฏการณ์ในทางปฏิบัติที่ยังไม่ชัดเจน
4. สถิติพาราเมตริกย่อมมีความแกร่งและให้ข้อสรุปที่สมเหตุสมผลแม้ว่าข้อมูลจะบิดเบือนเล็กน้อย ยิ่งไปกว่านั้น ถ้าการบิดเบือนเป็นแบบจริงจัง เทคนิควิธีการแปลงข้อมูลก็สามารถนำมาประยุกต์ใช้กับข้อมูลได้
5. สำหรับการออกแบบการวิจัยบางอย่าง วิธีการของสถิติทดสอบประเภทพาราเมตริกมีประโยชน์หลายอย่างคือ วิธีการของสถิติพาราเมตริกจะมีความไวมากกว่า สถิตินั้นพาราเมตริก และมีความสามารถในการค้นหาผลของปฏิสัมพันธ์ในการทดลองแบบแฟกตอเรียล (factorial) และมีความสามารถที่จะประมาณค่าขนาดของผลกระทบของการจัดกระทำได้

Thomas R. Knapp (1990) ได้เขียนบทความเรื่อง การจัดกระทำกับข้อมูลในมาตราจัดอันดับให้เป็นมาตราอันตรภาค : ความพยายามในการแก้ปัญหาข้อโต้แย้ง (Treating Ordinal Scales as Interval Scales : An Attempt to Resolve the Controversy) สรุปได้ว่า

การที่ S.S Steven (1946) ได้เสนอแนวความคิดเกี่ยวกับความเกี่ยวข้อง

ระหว่างระดับการวัดข้อมูลและการเลือกใช้สถิติในการวิเคราะห์ข้อมูล ได้มีข้อถกเถียงกันเรื่อยมาในเรื่องของความเชื่อในการใช้สถิติบรรยายและสถิติทดสอบ สำหรับข้อมูลในมาตราจัดอันดับ (ordinal level) ในทางจิตวิทยาประเด็นที่ถือว่าเป็นข้อขัดแย้งอันอับแรก คือ การตกลงใจที่แน่นอนเกี่ยวกับระดับของสเกลที่ควรจะเป็นซึ่ง Steven (1946) ได้แสดงให้เห็นถึงความแตกต่างระหว่างมาตราจัดอันดับ กับมาตราอันตรภาค โดยระบุว่า ตัวแปรที่เป็นมาตราอันตรภาคพิจารณาได้จากสมการเชิงเส้น $y = bx+a$ เช่น การแปลงมาตราวัดอุณหภูมิจากเซลเซียสเป็นฟาเรนไฮต์ ($F = 1.8 C + 32$) สำหรับตัวแปรมาตราจัดอันดับ การจัดลำดับจะต้องเป็นอิสระจากกัน (Mutually exclusive) และเรียงตามลำดับ นอกจากนี้ได้มีการเพิ่มคุณลักษณะเฉพาะสำหรับมาตราอันตรภาคขึ้น โดยกำหนดว่ามาตราอันตรภาคมีจุดศูนย์ (0) ที่กำหนดขึ้นเท่านั้น ซึ่งไม่ใช่ค่าศูนย์ที่แท้จริง แต่ในการที่จะชี้ชัดลงไปว่ารูปแบบใดเป็นมาตราในระดับใดไม่ใช่เรื่องง่าย ซึ่ง Gaito(1980), Borgatta และ Bohrnstedt (1980) และคนอื่น ๆ ได้กล่าวว่า ถ้าตัวแปรมีการแจกแจงแบบปกติข้อมูลนั้นมักจะเป็นมาตราอันตรภาค แต่ Thomas(1982)ไม่เห็นด้วย และได้พิสูจน์ข้อยืนยันดังกล่าว และระบุว่า "ถ้าคะแนนมีการแจกแจงปกติแล้วเป็นมาตราอันตรภาค" เป็นข้อยืนยันที่ไม่ถูกต้อง

Michell (1986) ไม่เห็นด้วยอย่างมากในการพิจารณาประเภทสเกล (scale type) และข้อโต้แย้งที่เกี่ยวกับสถิติที่เหมาะสม และกล่าวว่า นักวิทยาศาสตร์ได้ให้การสนับสนุนในการกำหนดความแตกต่างที่แน่นอนตามแนวคิดของ Steven แต่นักจิตวิทยาสนใจเฉพาะตัวเลขที่ใช้แทนลำดับ ซึ่งเป็นการศึกษาความสัมพันธ์เชิงปริมาณระหว่างตัวแปรหลาย ๆ ตัว

Marcus - Roberts (1987) ได้พิจารณาทั้งการวัดทางจิตวิทยาและความสำคัญของสถิติ และได้ให้ข้อสังเกตเกี่ยวกับประเด็นปัญหาที่สำคัญที่กล่าวถึงกันอยู่ และการใช้สถิติที่ไม่เหมาะสม (Meaningless Statistics) ว่า การวัดทางจิตวิทยาซึ่งมีข้อขัดแย้งระหว่างการวัดข้อมูลในระดับมาตราจัดอันดับ และมาตราอันตรภาค กับประเด็นปัญหาที่เกี่ยวกับสถิติบรรยายและสถิติอ้างอิง ควรแยกจากกันไม่ควรนำมาเกี่ยวข้องกัน



ข้อขัดแย้งส่วนใหญ่ระหว่างผู้สนับสนุน Steven (เรียกว่า กลุ่มอนุรักษนิยม) กับผู้คัดค้าน Steven (เรียกว่า กลุ่มเสรีนิยม) เริ่มเกิดขึ้นหลังจากที่ทั้งสองฝ่ายตกลงกันว่า ตัวแปรของการวัดทางจิตวิทยาเป็นมาตราจัดอันดับ แต่เมื่อมีการวิเคราะห์ข้อมูล กลุ่มเสรีนิยมกลับไม่คำนึงถึงระดับการวัดของข้อมูล

กลุ่มที่คัดค้าน Steven โต้แย้งว่าถึงแม้ว่าเขาไม่สามารถวัดระดับตัวแปรได้ถึงมาตราอันตรภาคแต่เขาดตกลงว่า ความแตกต่างระหว่างการจัดลำดับจาก A ไป B และจาก B ไป C และต่อ ๆ ไป เป็นความแตกต่างที่เท่ากัน ซึ่งปรากฏให้เห็นในรายงานของ Baker, Hardyk, Petrinovich (1966) และ Labovitz (1967) ซึ่งแสดงให้เห็นว่ามีความแตกต่างเพียงเล็กน้อย ถ้าเราจัดกระทำข้อมูลในมาตราจัดอันดับให้เหมือนกับข้อมูลในมาตราอันตรภาคในการใช้สถิติบรรยาย

กลุ่มที่สนับสนุน Steven กล่าวว่า มีผลการวิจัยที่แปลกประหลาดว่า ถ้าใช้ค่าเฉลี่ยเลขคณิต ส่วนเบี่ยงเบนมาตรฐานและสหสัมพันธ์แบบเพียร์สันกับข้อมูลในมาตราจัดอันดับ (ผู้วิจัยประกอบด้วย Steven, 1955; O'Brien, 1979; Marcus-Roberts & Roberts, 1987) ค่าเฉลี่ยของกลุ่มหนึ่งจะสูงกว่าค่าเฉลี่ยของกลุ่มหนึ่งในข้อมูลจากมาตรเดิม แต่ค่าเฉลี่ยของกลุ่มหนึ่งจะต่ำกว่าค่าเฉลี่ยของกลุ่มหนึ่งถ้ามาตรเดิมได้รับการแปลง ในส่วนของค่าสหสัมพันธ์ เครื่องหมายของค่าสหสัมพันธ์แบบเพียร์สัน (Pearson r) ของตัวแปรแบบจัดอันดับสองตัวแปรจะมีเครื่องหมายตรงกันข้ามกับค่าสหสัมพันธ์แบบสเปียร์แมนไร (Spearman's ρ)

Marcus - Roberts และ Roberts (1987) โต้แย้งว่าต้องพิจารณาเลือกใช้ค่าเฉลี่ยที่เหมาะสมในการคำนวณค่าเฉลี่ยของข้อมูลมาตราจัดอันดับ และจากรายงานของ Suppes (1959) และคณะ ได้แสดงให้เห็นว่า ข้อกล่าวที่ว่า ค่าเฉลี่ยของตัวแปรของกลุ่มที่คัดค้าน Steven มีค่ามากกว่าค่าเฉลี่ยของกลุ่มที่สนับสนุน Steven เป็นความสำคัญอย่างมากสำหรับข้อมูลในมาตราอันตรภาค แต่ไม่มีความสำคัญสำหรับข้อมูลในมาตราจัดอันดับ โดยมีเหตุผลในการยึดหลักของการยอมรับการแปลงข้อมูลในมาตราอันตรภาค แต่ไม่สามารถแปลงข้อมูลในมาตราจัดอันดับเป็นมาตราอันตรภาคได้

ข้อโต้แย้งของข้อมูลมาตราจัดอันดับและมาตราอันตรภาคมีความเกี่ยวข้องกับ

การสรุปผลของสถิติอ้างอิง กลุ่มที่สนับสนุน Steven ค่อนข้างจะเชื่อว่าครั้งแรกเรา เคยยอมรับการใช้สถิติบรรยายในข้อมูลมาตราจัดอันดับซึ่งเป็นการลงความเห็นโดยอัตโนมัตินี้เกี่ยวกับค่ามัธยฐาน และฐานนิยมของประชากร มากกว่าค่าเฉลี่ยเลขคณิต และยอมรับ rank correlation (หรือตัวอื่น ๆ) ที่คล้ายกัน มากกว่า pearson r's ซึ่งจะต้องใช้สถิติประเภทนพารามетริกมากกว่าใช้พารามетริก โดยยอมรับการมีอำนาจในการทดสอบน้อยกว่า แต่กลุ่มที่คัดค้าน Steven กล่าวว่าไม่เป็นการผิดที่จะใช้ค่าเฉลี่ยและสหสัมพันธ์ของเพียร์สันกับข้อมูลมาตราจัดอันดับ กลุ่มนี้ใช้สถิติประเภทพารามетริกกับทั้งข้อมูลมาตราจัดอันดับและมาตราอันตรภาค โดยอ้างว่าชนิดของสเกลไม่อยู่ในข้อตกลงสำหรับความตรง (Validity) ของการแจกแจงแบบ t และ F และได้อ้างว่าสถิติพารามетริกมีอำนาจในการทดสอบมากกว่า แต่กลุ่มที่คัดค้าน Steven กับกลุ่มที่สนับสนุน Steven ก็ยังคงผิดพลาดเกี่ยวกับการอ้างอำนาจในการทดสอบว่า สถิติทดสอบประเภทพารามетริกมีอำนาจในการทดสอบเหนือกว่าสถิติทดสอบประเภทนพารามетริกตัวอย่าง เช่น สถิติทดสอบวิลคอกซอน (ซึ่งเป็นสถิติทดสอบประเภทนพารามетริก) ที่ใช้ทดสอบกลุ่มตัวอย่างที่เป็นอิสระต่อกันและกลุ่มตัวอย่างที่เป็นคู่ ไม่ได้มีอำนาจการทดสอบน้อยกว่าสถิติทดสอบ t เลย และเมื่อประชากรไม่ได้มีการแจกแจงแบบปกติ (ทั้งข้อมูลมาตราจัดอันดับและมาตราอันตรภาค) สถิติทดสอบทั้งสองก็จะมีอำนาจการทดสอบมากขึ้น (Blair & Higgins, 1980) แต่ Maxwell และ Delaney (1985) แสดงให้เห็นว่า ในกลุ่มตัวอย่างที่เป็นอิสระต่อกัน การใช้ t - test สำหรับตัวแปรสังเกตทำให้สามารถสรุปอ้างถึงประชากรได้ถูกต้องเหมาะสม นอกจากนี้ Knapp (1990) ได้ให้คำแนะนำในการวางแผนแก่นักวิจัยซึ่งสนใจเกี่ยวกับการวัดระดับข้อมูลแบบสเกลและความสอดคล้องของสถิติในการวิเคราะห์ข้อมูลมีดังนี้

1. เลือกการวัดตามที่ผู้วัดคาดคิดว่าจะเป็นที่ ทฤษฎีของ Steven และคนอื่น ๆ เป็นทฤษฎีที่ใช้ได้ ถ้าเป้าหมายของข้อสรุปของสเกลอิสระ (Scale-free) ซึ่งเทียบกันได้กับปรัชญาของศาสตร์ของผู้วัด แต่ถ้าเทียบกันไม่ได้และผู้วัดพอใจทฤษฎีปฏิบัติการ หรือทฤษฎีคลาสสิก (Operational or Classical Theories) ข้อโต้แย้งส่วนใหญ่หรือทั้งหมดระหว่างสเกลและวิธีการทางสถิติที่ระบุไว้ในข้อเขียนหรือรายงานต่าง ๆ ก็ไม่เหมาะสม
2. อ่านข้อเขียนครั้งแรกของ Steven (1946) ด้วยความพินิจพิจารณา

ซึ่งเขียนไว้ดีมากและมีความหมายมาก โดยเฉพาะอย่างยิ่งในเนื้อหาที่มีการพัฒนาเมื่อเร็ว ๆ นี้ ซึ่งคำนึงถึงมโนทัศน์ (Concept) ที่มีความหมายในประเด็นหนึ่งงานข้อเขียนนั้น เขาทำห้กลุ่มที่คัดค้าน Steven ยอมรับเมื่อเขากล่าวว่า "คุณสมบัติที่เข้มงวดที่สุดของสถิติพรรณนาเกี่ยวกับค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานไม่ควรใช้กับมาตราจัดอันดับ สถิติเหล่านี้ให้ความรู้บางสิ่งมากกว่าการมีความเกี่ยวพันกันแบบลำดับที่ของข้อมูล"

3. เมื่อมีการสร้างสเกลจะต้องพิจารณาอย่างซื่อสัตย์ต่อการจัดจำพวกตามรูปแบบของ Steven และจะต้องชี้เฉพาะลงไปว่า สเกลนั้นอยู่ในระดับการวัดใด จำนวนของ Categories ที่ประกอบเป็นสเกลอาจมีความสำคัญมาก เช่น กำหนดค่าให้สเกลมี 10 จุด จะทำให้สเกลมีความต่อเนื่องมากกว่า 5 จุด

4. เมื่อมีการใช้สถิติบรรยายมาเกี่ยวข้อง จะต้องคำนึงถึงการตีความหมายก่อนที่จะสรุปข้อมูล ดูว่าใช้ประชากรทั้งหมดหรือว่าสุ่มตัวอย่าง หรือว่าเป็นตัวอย่างตามโอกาส (convenience sample) ถ้าจะต้องใช้สถิติ เช่น ค่าเฉลี่ยเลขคณิต (means) ส่วนเบี่ยงเบนมาตรฐาน (standard deviation) และสหสัมพันธ์แบบเพียร์สัน (pearson r's) และมีวิธีการวิเคราะห์แบบ Exploratory Data Analysis (EDA) ของ Tukey's (1977) และ Agresti (1984) ได้เขียนหนังสือเกี่ยวกับการวิเคราะห์ข้อมูลแบบจัดอันดับว่า ถ้ามั่นใจว่าข้อมูลเป็นมาตราจัดอันดับจะไม่สามารถศึกษาปฏิสัมพันธ์ (interaction) ในแบบดั้งเดิมซึ่งมีสถิติน้อยมากในการทดสอบความมีนัยสำคัญของความชัดเจนของปฏิสัมพันธ์ซึ่งโดยปกติสถิติเหล่านี้ไม่เหมาะสมสำหรับข้อมูลในมาตราจัดอันดับ

5. จะต้องมีการตัดสินใจที่แน่นอนว่าเป็นสเกลแบบใด ต้องการอธิบายข้อมูลอย่างไร ซึ่งมีทางเลือกของกระบวนการอ้างอิงหลายกระบวนการ ถ้าตัดสินใจว่าสเกลอยู่ในมาตราจัดอันดับจะต้องใช้สถิติทดสอบประเภทนพาราเมตริก แต่ถ้าข้อมูลอยู่ในมาตราอันดับจะต้องใช้สถิติทดสอบประเภทพาราเมตริก แต่ถ้าคำนึงถึงการแจกแจงแบบปกติและความเหมือนกันของความแปรปรวนแล้ว เลือกใช้สถิตินพาราเมตริกที่เทียบเคียงกับพาราเมตริกก็ไม่ต้องกลัวว่าอำนาจในการทดสอบจะลดลงอาจจะสูงกว่าก็ได้