บทที่ 5
สหสัมพันธ์และการถดถอย
CORRELATION AND REGRESSION

การศึกษาสัมพันธภาพระหว่างตัวแปร (Relationship between variables)

เท่าที่ผ่านมาเราได้พิจารณาข้อมูลแบบที่มีตัวแปรในข้อมูลนั้นเพียงตัวเดียว ในบทนี้เราจะเริ่มศึกษาสัมพันธภาพหรือความเกี่ยวพันกันระหว่างตัวแปรสองตัวในชุดข้อมูล ตัวแปรในที่นี้เราใช้หมายถึงสิ่งที่เป็นตัวแทนค่าสังเกตจากการวัดสิ่งที่เราสนใจ เช่น เราอาจให้ตัวแปร x หมายถึงความสูงของนักศึกษาแต่ละคน ตัวแปร y หมายถึงน้ำหนักของนักศึกษาแต่ละคน ถ้าเราหาตัวอย่างนักศึกษาในมหาวิทยาลัยมาสัก 50 คน แล้ววัดความสูงและน้ำหนักตัวของแต่ละคน ชุดข้อมูลที่เราได้ก็จะประกอบด้วยตัวแปรสองตัว คือความสูง (x) กับน้ำหนักตัว (y) เป็นของคู่กันสำหรับนักศึกษาแต่ละคน มีจำนวน 50 คู่ สิ่งที่เราจะสนใจในขั้นนี้ก็คือว่าตัวแปรสองตัวที่ได้มานี้มีความเกี่ยวพันกันหรือเราจะเรียกต่อไปว่ามีสัมพันธภาพระหว่างกันอย่างไร ในแง่ที่ว่าถ้าคนตัวสูงมักมีน้ำหนักตัวมาก หรือว่าตัวสูงมักมีน้ำหนักตัวน้อย หรือว่าไม่มีสัมพันธภาพต่อกันเลย คือว่าไม่อาจกล่าวได้ว่าตัวสูงแล้วมักจะน้ำหนักตัวมากหรือน้อย (ในทางกลับกันว่าน้ำหนักตัวมากแสดงว่าตัวสูงหรือตัวเตี้ย)

กรณีเช่นนี้เราจะเริ่มพิจารณาได้โดยการใช้กราฟแสดงการกระจายของจุดที่เรียกว่า สแกตเตอร์แกรม (scattergram or scatter diagram or correlation chart) เพื่อดูแบบแผนของจุดค่าสังเกตในข้อมูลว่าเป็นรูปใด โดยการใช้แกนนอน (แกน x) ของกราฟแทนตัวแปรตัวหนึ่งและแกนตั้ง (แกน y) แทนตัวแปรอีกตัวหนึ่ง แล้วเขียนจุดบนกราฟตามค่าพิกัด (x, y) ของค่าสังเกตแต่ละคู่

ขั้นต่อไปเราจะใช้ตัวสถิติที่เรียกว่าสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient) วัดระดับหรือขนาดของสัมพันธภาพระหว่างกันและกันของตัวแปรสองตัวซึ่งเราใช้คำว่าสหสัมพันธ์ โดยการคำนวณจากข้อมูลที่ได้แล้วแทนค่าลงในสูตรของสัมประสิทธิ์สหสัมพันธ์ นอกจากนี้เราจะทดสอบดูว่ามีสหสัมพันธ์อยู่จริงหรือไม่ในประชากรที่เราหาตัวอย่างมาใช้คำนวณสัมประสิทธิ์-สหสัมพันธ์ เช่นทดสอบว่าความสูงและน้ำหนักตัวมีสหสัมพันธ์ต่อกันจริงหรือไม่ในประชากร

นักศึกษาทั้งมหาวิทยาลัย ถ้าจากตัวอย่างนักศึกษา 50 คนที่เอ่ยถึงข้างต้น พบว่ามีสหสัมพันธ์อยู่ ทั้งนี้เพราะว่าค่าของสัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้จากตัวอย่างอาจได้มาโดยบังเอิญ ขณะที่ในประชากรแล้วอาจไม่มีสหสัมพันธ์ระหว่างตัวแปรสองตัว ที่เราสนใจพิจารณาอยู่ก็เป็นได้

คราวนี้ถ้าตัวแปรสองตัวที่เราพิจารณาอยู่มีสหสัมพันธ์ต่อกันอยู่มากเพียงพอ เราก็อาจจะใช้ค่าของตัวแปรตัวหนึ่งมาพยากรณ์ว่าตัวแปรอีกตัวหนึ่งจะมีค่าเท่าไร เช่นถ้าเราได้ว่าความสูงกับน้ำหนักตัวมีสหสัมพันธ์กันอย่างมาก เราก็อาจพยากรณ์น้ำหนักตัวของนักศึกษาคนหนึ่งได้ถ้าทราบความสูงของเขา หรือกลับกันอาจจะพยากรณ์ความสูงถ้าทราบน้ำหนักตัวของเขา วิธีการที่เราจะใช้ในการพยากรณ์ (prediction) นี้เรียกว่าการวิเคราะห์ความถดถอย (regression analysis) การพยากรณ์จะดีเพียงไรนั้นก็ขึ้นอยู่กับว่าสหสัมพันธ์ระหว่างตัวแปรสองตัวนั้นมีอยู่มากเพียงใด วิธีพยากรณ์ก็เพียงแต่สร้างเส้นตรงที่จะใช้แสดงแนวกลุ่มจุดของค่าสังเกตที่ได้ในสะแกตเตอร์แกรมเท่านั้นเอง

นอกจากนี้เราจะพิจารณาตัวแปรประเภทต่าง ๆ และวิธีการหาสหสัมพันธ์ระหว่างตัวแปรในแบบอื่น ๆ นอกเหนือจากที่กล่าวถึงในขั้นต้น พอเป็นพื้นฐานของการศึกษาการวิเคราะห์สหสัมพันธ์และการวิเคราะห์การถดถอยในชั้นสูงต่อไป

สรุปแล้วในบทนี้ นักศึกษาจะได้เรียนรู้ถึง

ก) การสร้างและการตีความหมายของสะแกตเตอร์แกรม

ข) การคำนวณหาค่าและการตีความหมายของสัมประสิทธิ์สหสัมพันธ์ r (product moment correlation coefficient)

ค) การทดสอบสมมติฐานเกี่ยวกับนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์ในประชากรของข้อมูล

ง) การคำนวณหาสัมประสิทธิ์การถดถอย (regression coefficient) และการใช้สมการถดถอยเพื่อการพยากรณ์ค่าของตัวแปรตาม (dependent variable) เมื่อทราบค่าของตัวแปรอิสระ (independent variable)

จ) ประเภทของตัวแปรจำแนกตามมาตรการวัดที่ใช้กับตัวแปรนั้น และวิธีการหาค่าสหสัมพันธ์ระหว่างตัวแปรแบบที่ต่างจากที่กล่าวมาแล้ว รวมทั้งการทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์เหล่านั้นด้วย

5.2 ตัวอย่างการศึกษาสัมพันธภาพระหว่างตัวแปร

ตัวอย่าง 5.1 นักวิจัยทางการศึกษาอาจสนใจสัมพันธภาพระหว่างความสามารถทางดนตรีกับความสามารถทางคณิตศาสตร์ของนักเรียนชั้นประถม เพื่อพิจารณาว่าความสามารถสองอย่างนี้เกี่ยวพันกันหรือไม่ อย่างไร (จริงหรือไม่ที่นักเรียนที่เก่งดนตรีมักจะเก่งคณิตศาสตร์ด้วย หรือ

ว่ากลับกันว่าถ้าเก่งดนตรีมักจะไม่เก่งคณิตศาสตร์ หรือว่าไม่มีสหสัมพันธ์ระหว่างตัวแปรทั้งสองนี้เลย)

วิธีหนึ่งที่เราจะวัดสหสัมพันธ์ระหว่างความสามารถสองอย่างนี้ ก็คือโดยการเลือกตัวอย่างนักเรียนมาโดยวิธีสุ่ม แล้วให้นักเรียนที่เลือกมาในกลุ่มนั้นแต่ละคนทำข้อสอบวิชาดนตรีและวิชาคณิตศาสตร์ (โดยสันนิษฐานว่าคะแนนสอบที่ได้จะใช้เป็นเครื่องวัดความสามารถของนักเรียน ที่เชื่อถือได้) จากนั้นก็อาจคำนวณหาสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบสองวิชานี้ เพื่อดูว่าคะแนนสอบวิชาคณิตศาสตร์และคะแนนสอบวิชาดนตรีมีสัมพันธภาพต่อกันในรูปใด และมีมากน้อยเพียงใด ถ้าสัมพันธภาพระหว่างคะแนนที่ใช้แสดงความสามารถสองอย่างนี้สูงมากพอ เราก็อาจจะใช้คะแนนในวิชาหนึ่งพยากรณ์คะแนนที่นักเรียนคนหนึ่งจะได้ในอีกวิชาหนึ่ง โดยวิธีการวิเคราะห์ความถดถอย

ตัวอย่าง 5.2 บริษัทขายแผ่นเสียงสนใจว่ามีสหสัมพันธ์ระหว่างจำนวนแผ่นเสียงที่ขายได้ในจังหวัดหนึ่ง กับจำนวนเด็กวัยรุ่นในจังหวัดนั้น นอกจากนั้นบริษัทอาจจะต้องการพยากรณ์ว่าจะขายแผ่นเสียงได้กี่แผ่น ถ้าทราบจำนวนเด็กวัยรุ่นในจังหวัดนั้น ปัญหาสองอย่างนี้สามารถจะแก้ไขได้โดยการใช้วิธีวิเคราะห์สหสัมพันธ์และการถดถอย

ตัวอย่าง 5.3 ทีมวิจัยทางการแพทย์ต้องการศึกษาสัมพันธภาพระหว่างอายุกับความดันโลหิตของบุคคล ถ้าพบว่าตัวแปรสองอย่างนี้ เกี่ยวพันกันมากพอ ก็จะทำให้สามารถพยากรณ์ความดัน-โลหิตได้ใกล้เคียงเมื่อทราบอายุของบุคคลนั้น

ตัวอย่าง 5.4 นักวิจัยทางจิตวิทยาสนใจว่าระดับเชาวน์ปัญญา (IQ) ในวัยเด็กกับระดับเชาว์ปัญญาเมื่อโตเป็นผู้ใหญ่แล้วมีสัมพันธภาพกันอย่างไร และอาจต้องการใช้ระดับเชาวน์ปัญญาอันหนึ่งเพื่อพยากรณ์อีกอันหนึ่ง ถ้าพบว่าตัวแปรสองอย่างนี้มีสหสัมพันธ์ต่อกันสูงพอ

ตัวอย่าง 5.5 นักรัฐศาสตร์สนใจว่าจำนวนเปอร์เซนต์ของประชาชนที่ลงคะแนนให้พรรค ฎ กับเปอร์เซนต์ที่นั่งในสภาผู้แทนราษฎรของพรรค ฎ มีสัมพันธภาพอย่างไรดูจากการเลือกตั้งทุกครั้งที่แล้วมา และสำหรับการเลือกตั้งครั้งต่อไปอาจจะใช้จำนวนเปอร์เซนต์ของประชาชนที่คาดว่าจะลงคะแนนให้พรรค เพื่อพยากรณ์จำนวนเปอร์เซนต์ที่นั่งในสภาของพรรคนี้ได้

ตัวอย่าง 5.6 นักสังคมวิทยาศึกษาว่าชีวิตวัยเด็กที่มีความสุขมีอิทธิพลอย่างไรต่อการประสบความสำเร็จในชีวิตสมรสเมื่อโตเป็นผู้ใหญ่ โดยวัดระดับสหสัมพันธ์ระหว่างจำนวนคนที่ได้คำตอบจากการสัมภาษณ์ ว่ามีวัยเด็กที่มีความสุข กับจำนวนคนที่ตอบว่าประสบความสำเร็จในชีวิตสมรส

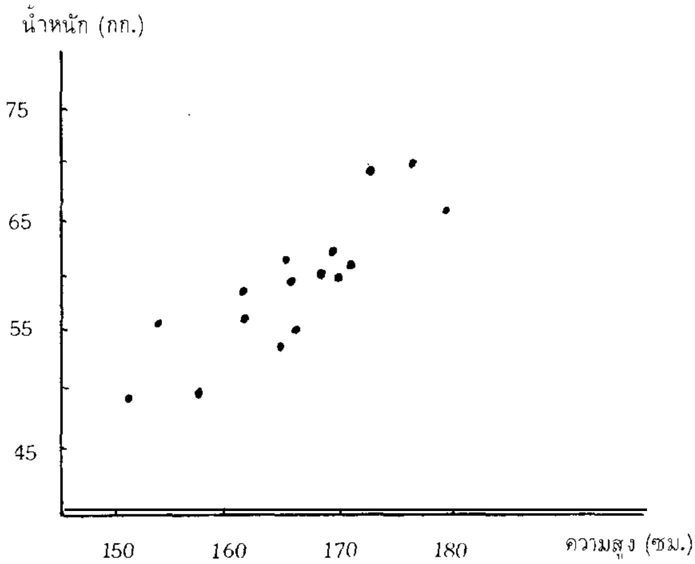
5.3 สแกตเตอร์แกรม (scattergram)

บ่อยครั้งที่เราสนใจสัมพันธภาพระหว่างการวัดหรือค่าสังเกตในตัวอย่างที่เลือกมา เมื่อได้ค่าสังเกตที่ประกอบด้วยตัวแปรเป็นคู่ ๆ ตัวอย่างเช่นปัญหาที่ว่าคะแนนสอบกลางภาคกับคะแนนสอบปลายภาคในวิชาหนึ่งมีสัมพันธภาพต่อกันหรือไม่ หรือปัญหาที่ว่าครอบครัวชาวกรุงเทพฯ ที่ร่ำรวยมักจะมีลูกมากหรือว่ามักจะมีลูกน้อย นั่นคือว่าสัมพันธภาพระหว่างความร่ำรวยกับจำนวนบุตรในครอบครัวเป็นในลักษณะใด ปัญหาประเภทนี้เป็นเรื่องที่เกี่ยวกับสหสัมพันธ์ระหว่างตัวแปรสองตัว วิธีการเบื้องต้นง่าย ๆ ที่เราจะทำได้ในการศึกษาสหสัมพันธ์ก็โดยการใช้สแกตเตอร์แกรม หรือเขียนกราฟแสดงการกระจายของจุดค่าสังเกต โดยให้แกนนอนแทนตัวแปรหนึ่ง และแกนตั้งแทนตัวแปรอีกตัวหนึ่ง นั่นเอง

ในที่นี้เราจะดูจากตัวอย่างง่าย ๆ ถ้าเราเลือกตัวอย่างนักศึกษาผู้ชายในมหาวิทยาลัยมาโดยสุ่มจำนวน 16 คน ได้ค่าสังเกตสำหรับความสูงและน้ำหนักตัวของแต่ละคนดังนี้

นักศึกษาคนที่	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
ความสูง (ซม.)	173	152	162	171	170	158	166	170	180	165	168	165	176	162	165	153
น้ำหนัก (กก.)	69	48	56	61	62	49	55	61	67	53	60	58	71	58	60	56

ถ้าเราเขียนกราฟให้แกนนอนแทนความสูง และแกนตั้งแทนน้ำหนักตัว สำหรับข้อมูลชุดนี้เราจะได้สแกตเตอร์แกรมดังนี้



จากตารางค่าข้อมูลเราอาจบอกไม่ได้มากนักว่าความสูงกับน้ำหนักตัวมีสัมพันธภาพกันอย่างไร แต่จากสแกตเตอร์แกรมจะเห็นได้ว่า คนที่ตัวสูงมักจะมีน้ำหนักตัวมากด้วย หรือพูดได้อีกอย่างว่าความสูงกับน้ำหนักตัวในข้อมูลที่ได้มานี้จะมีการเปลี่ยนแปลงไปในทางเดียวกัน คือเมื่อค่าของตัวแปรหนึ่งเพิ่มขึ้น ค่าของตัวแปรอีกตัวหนึ่งก็เพิ่มขึ้นด้วยโดยการสังเกตจากแบบแผนของจุดบนสแกตเตอร์แกรมที่เป็นแนวเอียงไปทางขวา

แบบฝึกหัดที่ 5.1

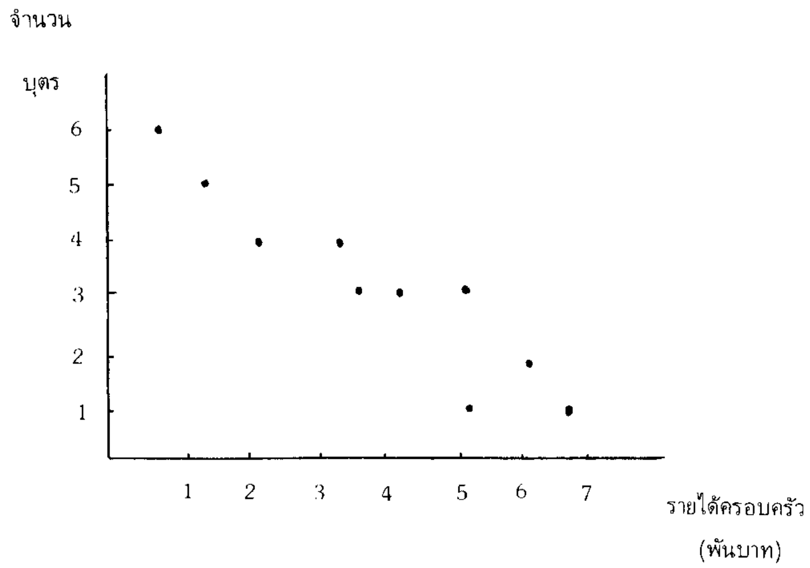
1) จากสแกตเตอร์แกรมข้างต้นนี้ จงวงรอบจุดที่แสดงค่าสังเกตของนักศึกษาคนที่ 10
2) จากสแกตเตอร์แกรมข้างต้นนี้ นักศึกษาคิดว่าความสูงกับน้ำหนักตัวของบุคคลมีสหสัมพันธ์ต่อกันหรือไม่ ตอบ (มี)
3) ถ้า x เป็นตัวแปรที่แทนความสูง และ y เป็นตัวแปรที่แทนน้ำหนักตัว เราจะพูดได้ว่า ตัวแปร x กับ ตัวแปร y มี ... ต่อกันและกัน ตอบ (สหสัมพันธ์)

โดยปกติแล้วเมื่อตัวแปรมีสหสัมพันธ์ต่อกันและกัน เราอาจใช้ตัวแปรหนึ่งพยากรณ์ค่าอีกตัวหนึ่งได้ ตัวแปรที่ใช้พยากรณ์อีกตัวหนึ่งนั้นเรียกว่าตัวแปรอิสระ มักจะแทนด้วยแกนนอน (x) ในสแกตเตอร์แกรม ส่วนตัวแปรที่ถูกพยากรณ์นั้นเรียกว่าตัวแปรตาม มักจะแทนด้วยแกนตั้ง (y) ดังตัวอย่างที่แล้วมานี้ เราเลือกให้ความสูงเป็นตัวแปรอิสระ และแทนความสูงด้วยแกน x แต่ที่จริงแล้วเราอาจให้น้ำหนักตัวเป็นตัวแปรอิสระก็ได้ ขึ้นอยู่กับว่าเราจะสนใจพยากรณ์ตัวแปรใดก็ให้ตัวแปรนั้นเป็นตัวแปรตาม

ตัวแปรที่เราสนใจนั้นอาจไม่ได้เปลี่ยนแปลงไปทางเดียวกันก็ได้ ดังที่จะเห็นได้จากกรณีต่อไปนี้

นักสังคมศาสตร์สนใจว่าจำนวนบุตรในครอบครัวกับรายได้ของครอบครัวนั้นมีสัมพันธภาพต่อกันอย่างไร จากตัวอย่างครอบครัวที่ไปสัมภาษณ์มา 10 ครอบครัว ได้ข้อมูลมาดังนี้

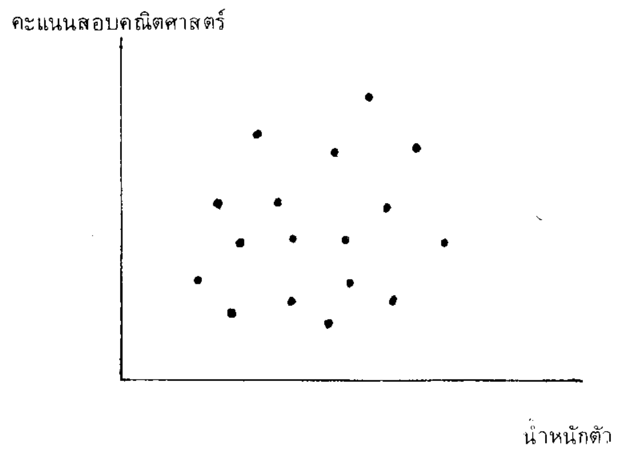
ครอบครัวที่	1	2	3	4	5	6	7	8	9	10
รายได้	750	1200	3200	5000	4000	1800	5000	6700	5800	3500
จำนวนบุตร	6	5	4	3	3	4	1	1	2	3



จะเห็นได้จากสแกตเตอร์แกรมว่าในข้อมูลชุดนี้สัมพันธภาพระหว่างรายได้ครอบครัวกับจำนวนบุตรในครอบครัวนั้นมีอยู่ ตัวแปรทั้งสองดูจะเปลี่ยนแปลงไปในทางตรงกันข้าม คือ ครอบครัวที่ร่ำรวย ดูจะมีบุตรน้อย ขณะที่ครอบครัวที่จนกว่าดูจะมีบุตรมาก

การแสดงสัมพันธภาพระหว่างตัวแปรนี้ไม่ได้แสดงว่าตัวแปรหนึ่งเป็นสาเหตุของอีกตัวแปรหนึ่งแต่อย่างใด สหสัมพันธ์ระหว่างตัวแปรทั้งสองตัวที่เห็นได้จากสแกตเตอร์แกรมเพียงแต่บอกให้เห็นว่าตัวแปรสองตัวเปลี่ยนแปลงไปในทางเดียวกัน หรือในทางตรงกันข้าม การที่เรียกตัวแปรตัวหนึ่งเป็นตัวแปรอิสระ และอีกตัวหนึ่งเป็นตัวแปรตาม ไม่ได้เกี่ยวกับการเป็นเหตุและผลของกันแต่อย่างใด

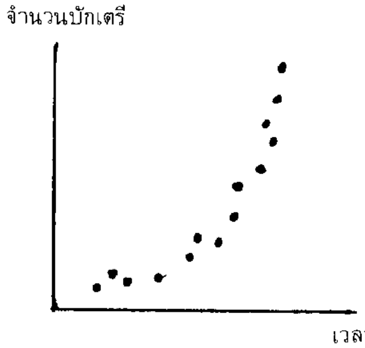
นอกจากนี้ในบางกรณีเราอาจพบว่าตัวแปรสองตัวที่เราพิจารณาไม่ได้มีการเปลี่ยนแปลงร่วมกันแต่อย่างใด นั่นคือไม่ได้มีสัมพันธภาพต่อกัน เช่นในกรณีของคะแนนสอบคณิตศาสตร์กับน้ำหนักตัว เราอาจได้สแกตเตอร์แกรมที่มีลักษณะดังนี้



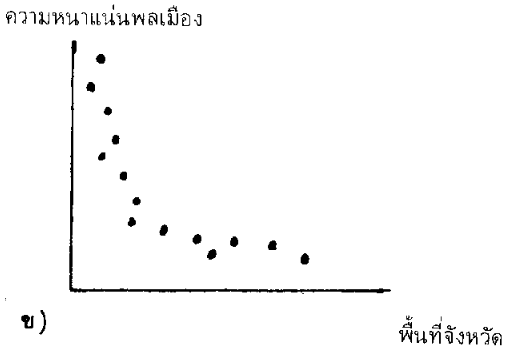
จากสแกตเตอร์แกรมจะเห็นว่ากลุ่มจุดที่ได้ไม่ได้แสดงแบบแผนเป็นแนวหรือเป็นรูปร่างที่แสดงแนวโน้มระหว่างค่าของตัวแปรทั้งสองแต่อย่างใด เราอาจกล่าวได้จากการพิจารณาสแกตเตอร์แกรมว่าตัวแปรทั้งสองไม่มีสัมพันธภาพต่อกัน

เท่าที่แล้วมาเราได้เห็นสแกตเตอร์แกรมแสดงสัมพันธภาพระหว่างตัวแปรที่มีรูปร่างของกลุ่มจุดอยู่ในแนวค่อนข้างจะเป็นเส้นตรง และที่จริงเราอาจลากเส้นตรงแสดงแบบแผนของกลุ่มจุดที่ผ่านมาแล้วได้ ดังนั้นสัมพันธภาพระหว่างตัวแปรที่สามารถแสดงได้ด้วยเส้นตรงนี้เราจะเรียกว่า สัมพันธภาพแบบเส้นตรง และพูดถึงสหสัมพันธ์ระหว่างตัวแปรว่า เป็นสหสัมพันธ์แบบเส้นตรง หรือสหสัมพันธ์เชิงเส้นตรง (linear correlation)

ในกรณีอื่น สัมพันธภาพระหว่างตัวแปรอาจจะมีอยู่จริงแต่ไม่ได้อยู่ในรูปเส้นตรง เช่นสัมพันธภาพระหว่างจำนวนบักเตรีในภาชนะทดลองกับเวลาที่ใช้เพาะเชื้อ ดังรูป ก) หรือสัมพันธภาพระหว่างพื้นที่ของจังหวัดกับความหนาแน่นของพลเมืองในจังหวัดนั้นต่อพื้นที่ ดังรูป ข)



ก)



ข)

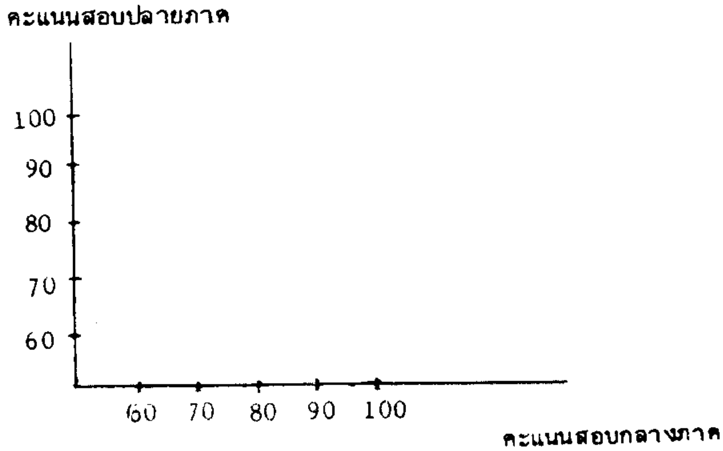
กรณีเช่นนี้สัมพันธภาพระหว่างตัวแปรเรียกว่าเป็นสัมพันธภาพแบบเส้นโค้ง (curvilinear relationship) แต่ในขั้นนี้เราจะสนใจแต่เพียงกรณีที่ตัวแปรมีสัมพันธภาพต่อกันเป็นแบบเส้นตรงเท่านั้น

ขอให้สังเกตว่าเมื่อพิจารณาจากสแกตเตอร์แกรมนั้น กลุ่มจุดที่ใกล้เคียงกับแนวเส้นตรงหรือเส้นโค้งแสดงว่าสัมพันธภาพระหว่างตัวแปรทั้งสองตัวนั้นมีมากกว่ารูปที่กลุ่มจุดกระจัดกระจายไม่เข้าแนว

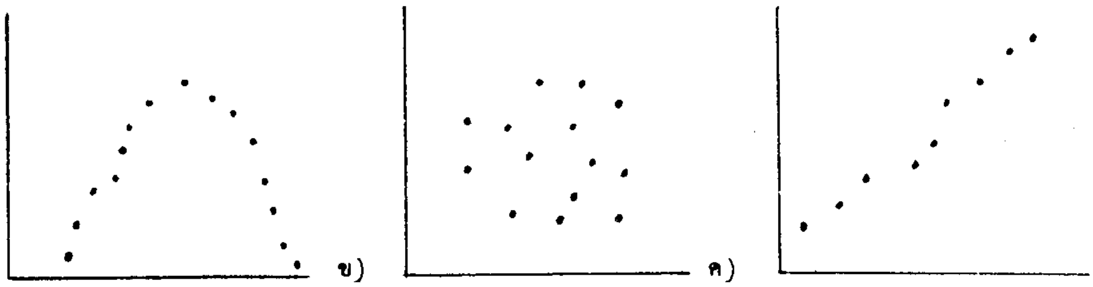
แบบฝึกหัดที่ 5.2

1) จากข้อมูลเกี่ยวกับคะแนนสอบกลางภาค กับคะแนนสอบปลายภาคของนักศึกษา 10 คนนี้ จงเขียนสแกตเตอร์แกรมแสดงสัมพันธภาพระหว่างตัวแปรทั้งสองและบรรยายว่าจากแบบแผนของกลุ่มจุดที่ได้ จะกล่าวได้อย่างไรเกี่ยวกับสัมพันธภาพของคะแนนสอบกลางภาคกับคะแนนสอบปลายภาค

นักศึกษาคนที่	1	2	3	4	5	6	7	8	9	10
คะแนนกลางภาค	65	68	71	75	75	86	88	95	97	99
คะแนนปลายภาค	74	73	71	80	83	85	90	94	99	98



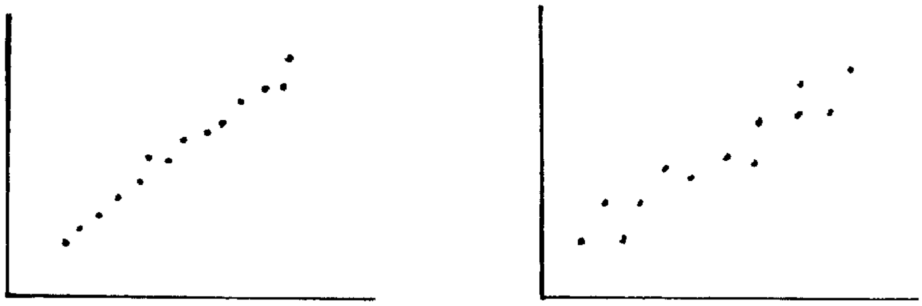
2) สัมพันธภาพระหว่างตัวแปรที่แสดงแบบแผนได้ด้วยเส้นตรงนั้นเรียกว่าสัมพันธภาพแบบ...
3) สัมพันธภาพระหว่างตัวแปรที่แสดงแบบแผนได้ด้วยเส้นโค้งนั้นเรียกว่าสัมพันธภาพแบบ...
4) จากสแกตเตอร์แกรมต่อไปนี้ รูปใดแสดงสัมพันธภาพแบบเส้นตรงระหว่างตัวแปร x กับตัวแปร y



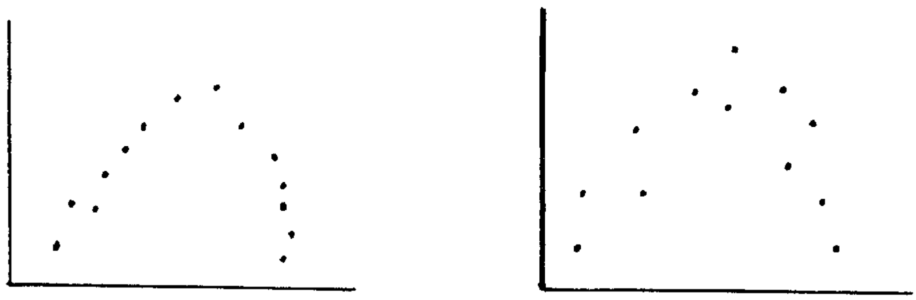
ก) ข) ค)

5) จากข้อ 4 รูปใดแสดงสัมพันธภาพแบบเส้นโค้งระหว่างตัวแปร x กับตัวแปร y

6) จากสแกตเตอร์แกรมต่อไปนี้ รูปใดแสดงสหสัมพันธ์ระหว่างตัวแปรมากกว่าอีกรูปหนึ่ง



7) จากสแกตเตอร์แกรมต่อไปนี้ รูปใดแสดงสหสัมพันธ์ระหว่างตัวแปรมากกว่าอีกรูปหนึ่ง



สัมประสิทธิ์สหสัมพันธ์ (correlation coefficient)

ถ้าตัวแปรที่เราสนใจมีสัมพันธภาพต่อกันและกันเป็นแบบเส้นตรง เราสามารถที่จะวัดว่าสัมพันธภาพนั้นมีระดับมากน้อยเพียงใด โดยอาศัยตัวสถิติที่เรียกว่าสัมประสิทธิ์สหสัมพันธ์ ซึ่งใช้สัญลักษณ์แทนด้วยตัว r สำหรับสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรที่คำนวณได้จากตัวอย่าง และใช้สัญลักษณ์ ρ สำหรับสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรในประชากรทั้งหมด (สัญลักษณ์ ρ นี้มาจากอักษรกรีก ออกเสียงว่า "โร")

ที่จริงแล้วสัมประสิทธิ์สหสัมพันธ์นั้นมีอยู่ด้วยกันหลายแบบ สำหรับข้อมูลประเภทต่างกันไป แต่ในขั้นนี้เราจะศึกษาสัมประสิทธิ์สหสัมพันธ์แบบที่เรียกว่า productmoment correlation coefficient

ค่าของ r ซึ่งเป็นตัววัดขนาดหรือระดับ (ดัชนี) ของสหสัมพันธ์ระหว่างตัวแปรทั้งสอง จะอยู่ระหว่าง -1 ถึง +1 เท่านั้น โดยที่ r จะมีค่าและมีความหมายดังต่อไปนี้

ก) r มีค่าเป็น + ถ้าตัวแปรสองตัวเปลี่ยนแปลงไปในทางเดียวกัน คือเมื่อตัวหนึ่งเพิ่มขึ้นอีกตัวหนึ่งก็เพิ่มขึ้นด้วย เรียกว่ามีสหสัมพันธ์ทางบวกระหว่างตัวแปร

ข) r มีค่าเป็น - ถ้าตัวแปรสองตัวเปลี่ยนแปลงไปในทางตรงกันข้าม คือเมื่อตัวหนึ่งเพิ่มขึ้นอีกตัวหนึ่งก็จะลดลง เรียกว่ามีสหสัมพันธ์ทางลบระหว่างตัวแปร

ค) r มีค่าเป็น 0 ถ้าตัวแปรสองตัวไม่มีการเปลี่ยนแปลงร่วมกันเลย เรียกว่ามีค่าสหสัมพันธ์เป็นศูนย์ หรือไม่มีสัมพันธภาพต่อกันและกันเลย

ถ้าค่า r เท่ากับ +1 เราเรียกว่ามีสหสัมพันธ์ทางบวกแบบสมบูรณ์ (perfect positive correlation) ระหว่างตัวแปร จุดค่าสังเกตบนสแกตเตอร์แกรมจะเรียงเป็นเส้นตรงพอดี (เอียงขวา)

ถ้าค่า r เท่ากับ -1 เราเรียกว่ามีสหสัมพันธ์ทางลบแบบสมบูรณ์ (perfect negative correlation) ระหว่างตัวแปร จุดค่าสังเกตบนสแกตเตอร์แกรมจะเรียงเป็นเส้นตรงพอดี (เอียงซ้าย)

ถึงแม้จะไม่กำหนดตายตัวว่า r จะต้องมีค่าเท่าใดจึงจะเรียกว่ามีสหสัมพันธ์สูงหรือสหสัมพันธ์ต่ำ เพราะขึ้นอยู่กับประเภทของงานที่จะใช้ค่าสหสัมพันธ์นั้นว่าต้องการความละเอียดมากเพียงใด แต่เกณฑ์ที่ใช้กันโดยคร่าว ๆ มักจะถือเอาค่า $|r|$ (ค่า r ที่ไม่คิดเครื่องหมาย) ระหว่าง 1.0 ถึง 0.7 เป็นสหสัมพันธ์ระดับสูง ระหว่าง 0.7 ถึง 0.4 เป็นสหสัมพันธ์อย่างต่ำ และระหว่าง 0.4 ถึง 0 เป็นสหสัมพันธ์อย่างต่ำมากหรืออาจไม่มีสหสัมพันธ์เลย

ข้อพึงระวังเกี่ยวกับการศึกษาเรื่องสหสัมพันธ์นี้สำคัญยิ่งก็คือ ค่าของสัมประสิทธิ์สหสัมพันธ์ที่ได้นั้นไม่ได้บอกถึงการเป็นเหตุหรือเป็นผลกันระหว่างตัวแปร เพียงแต่บอกว่าตัวแปรสองตัวนั้นมีสัมพันธภาพต่อกันในลักษณะที่มีการเปลี่ยนแปลงร่วมกันมากน้อยเพียงใดเท่านั้นเอง บ่อยครั้งที่เราจะพบสหสัมพันธ์ระหว่างตัวแปรสองตัวที่ไม่ได้มีความเกี่ยวข้องกันโดยตรงเลย แต่อาจมีสัมพันธภาพต่อตัวแปรที่สามร่วมกันทำให้พบว่ามีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างกันค่อนข้างสูง ทำให้บางครั้งเราก็ไม่อาจอธิบายได้ว่าทำไมตัวแปรทั้งสองตัวจึงมีสหสัมพันธ์ต่อกันสูง ทั้ง ๆ ที่ไม่มีความเกี่ยวข้องกันแต่อย่างใด พวกนี้เราเรียกว่าสหสัมพันธ์ลวง (*spurious correlation*) ดังเช่นว่า เราอาจพบว่าสหสัมพันธ์ระหว่างความสูงของเด็กนักเรียนในชั้นประถมกับความสามารถในการอ่าน มีอยู่ค่อนข้างสูง แต่ทั้งนี้เป็นเพราะว่าทั้งความสูงและความสามารถในการอ่านเปลี่ยนแปลงไปตามอายุ นั่นคืออายุตัวกลางเชื่อมระหว่างความสูงกับความสามารถในการอ่าน การสรุปว่าความสูงกับการอ่านเป็นเหตุเป็นผลต่อกันโดยตรง หรือว่าเด็กสูงมักอ่านเก่งนั้นจึงไม่ถูกต้อง ในทำนองเดียวกันสหสัมพันธ์ระหว่างความยาวของแขนขวากับความยาวของขาซ้ายที่มีอยู่ค่อนข้างสูงในคนทั่วไป ก็ไม่ได้หมายความว่าความยาวของแขนเป็นเหตุหรือเป็นผลจากการมีขายาวแต่อย่างใด

นอกจากนี้สัมประสิทธิ์สหสัมพันธ์ r ที่เรากำลังศึกษาอยู่นี้ใช้วัดสหสัมพันธ์ระหว่างตัวแปรที่เป็นแบบเส้นตรงเท่านั้น จึงเป็นไปได้ว่าตัวแปรสองตัวที่มีสัมพันธภาพระหว่างกันค่อนข้างมากเป็นแบบเส้นโค้ง จะคำนวณค่า r ออกมาได้ค่อนข้างต่ำหรือใกล้ศูนย์

แบบฝึกหัดที่ 5.3

1) ในการศึกษาการไหลของเลือดผ่านไตของคนที่อายุต่าง ๆ กัน พบว่าอัตราการไหลของเลือดกับอายุของบุคคลมีค่าสัมประสิทธิ์สหสัมพันธ์วัดได้ -0.85 จะอธิบายได้ว่า เมื่ออายุมากขึ้นอัตราการไหลของเลือดผ่านไตจะ...... ตอบ (ช้าลง)
2) ถ้าจำนวนครูโรงเรียนในจังหวัด และปริมาณการขายสุราในจังหวัดต่าง ๆ ในภาคกลาง มีสหสัมพันธ์ระหว่างกันวัดได้ค่าสัมประสิทธิ์สหสัมพันธ์ 0.75 เราจะกล่าวได้ไหมว่าครูโรงเรียนเป็นพวกที่ดื่มสุรามาก เมื่อมีครูมากก็ทำให้มีการขายสุรามากไปด้วย ตอบ (ไม่ได้)

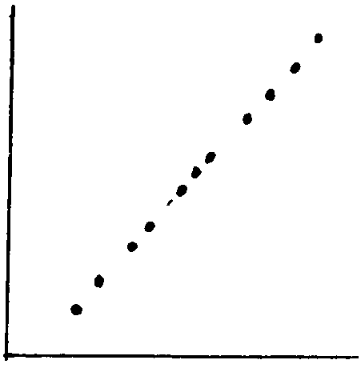
(อย่าลืมว่าสัมประสิทธิ์สหสัมพันธ์ไม่ได้แสดงเหตุผลหรือผลระหว่างตัวแปรในกรณีนี้อาจเป็นได้ว่าทั้งจำนวนครู และปริมาณการขายสุราในแต่ละจังหวัดนั้นขึ้นอยู่กับจำนวนพลเมืองในจังหวัดเหล่านั้น ถ้าเป็นจังหวัดใหญ่มีพลเมืองมากก็มีครูมาก และมีการขายสุรามากเป็นธรรมดา)

ข้อสังเกต อย่างไรก็ตามการที่ตัวแปรสองตัวมีสหสัมพันธ์อย่างมากต่อกัน อาจเป็นเครื่องช่วยในการพิจารณาตัดสินว่าตัวแปรสองตัวนั้นเป็นเหตุหรือเป็นผลของอีกตัวหนึ่งหรือไม่โดยอาศัยข้อมูลแวดล้อมและการวิเคราะห์โดยเหตุผลเป็นเครื่องประกอบการตัดสินใจ เช่นว่า สหสัมพันธ์อย่างสูงระหว่างปริมาณปุ๋ยที่ให้ต่อพื้นที่เพาะปลูกกับผลผลิตข้าวที่ได้ต่อพื้นที่ ก็อาจเห็นได้ว่าผลผลิตที่ได้เพิ่มขึ้นเป็นผลจากปริมาณปุ๋ยที่ให้ (ไม่ใช่ปริมาณปุ๋ย เป็นผลมาจากปริมาณผลผลิตข้าวที่เพิ่มขึ้น) หรือตัวอย่างในทางการแพทย์ที่พบว่าสหสัมพันธ์ระหว่างการสูบบุหรี่กับการเป็นมะเร็งในปอดทำให้มีการศึกษาค้นคว้าลึกซึ้งต่อไปที่ทำให้แพทย์ส่วนมากเชื่อว่าการสูบบุหรี่เป็นสาเหตุหนึ่งของการนำไปสู่การเป็นมะเร็งในปอด หรือการที่พบสหสัมพันธ์ทางบวกอย่างสูงมากของการที่เด็กเกิดมาผิดปกติกับการที่มารดาเป็นโรคหัดเยอรมันระหว่างตั้งครรภ์ก็นำไปสู่การสรุปที่ว่าโรคหัดเยอรมัน เป็นสาเหตุของการผิดปกติของทารก

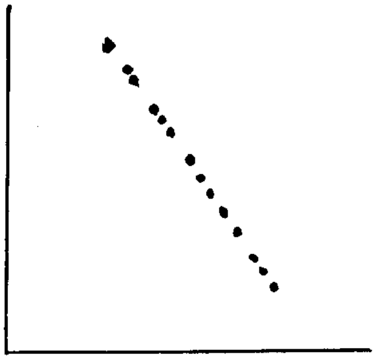
ทั้งนี้พึงสังเกตว่าในบางกรณี เช่นการที่พบสหสัมพันธ์ทางลบค่อนข้างสูงระหว่างจำนวนบุตรกับรายได้ครอบครัว ไม่ได้หมายความว่าจะสรุปได้ง่าย ๆ ว่าลูกมากทำให้ยากจน เพราะในทางตรงกันข้ามจะพบว่ามีเหตุผลแวดล้อมอย่างอื่นอีกมากที่จะแสดงให้เห็นได้ว่าความยากจนนำไปสู่การมีลูกมาก

5.5 สแกตเตอร์แกรมแสดงสหสัมพันธ์แบบต่าง ๆ กัน

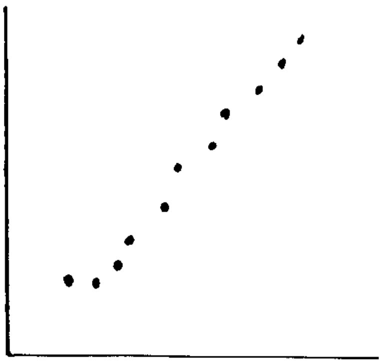
1) สหสัมพันธ์ทางบวกแบบสมบูรณ์



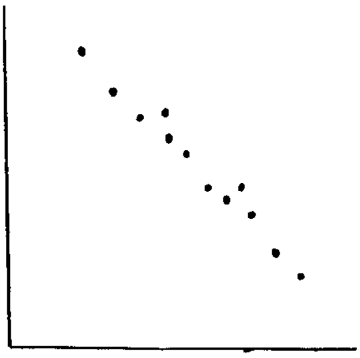
2) สหสัมพันธ์ทางลบแบบสมบูรณ์



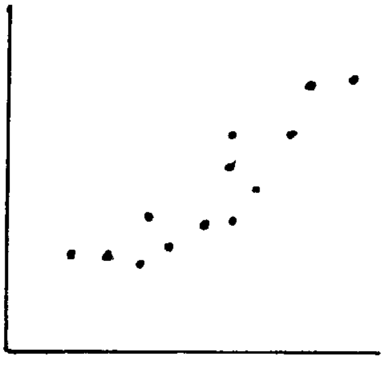
3) สหสัมพันธ์ทางบวกอย่างสูง



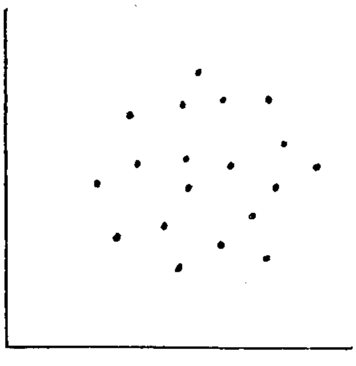
4) สหสัมพันธ์ทางลบอย่างสูง



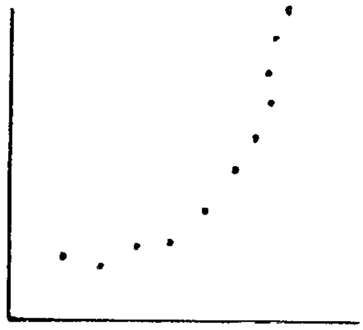
5) สหสัมพันธ์ทางบวกอย่างต่ำ

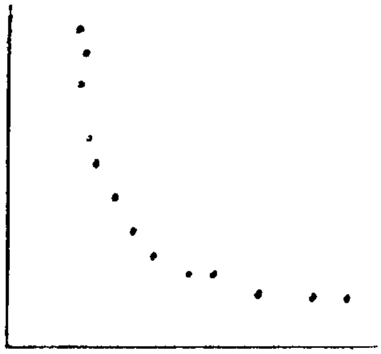


6) สหสัมพันธ์เป็นศูนย์



7) สหสัมพันธ์แบบเส้นโค้ง (จะวัดได้ r ต่ำ เพราะสัมพันธภาพระหว่างตัวแปร ไม่เป็นแบบเส้นตรง)





การคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์

สูตรในการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ คือ

$$r = \frac{n\,\Sigma xy - \Sigma x \Sigma y}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

ในเมื่อ r คือสัมประสิทธิ์สหสัมพันธ์

x และ y คือตัวแปรสองตัวที่เราต้องการวัดสหสัมพันธ์ระหว่างกันและกัน

n คือจำนวนคู่ค่าสังเกต

ขั้นตอนการคำนวณ

ขั้นที่ 1 ยกกำลังสองแต่ละค่าสังเกตของ x และ y

ขั้นที่ 2 หาผลคูณของ x กับ y แต่ละคู่

ขั้นที่ 3 หาผลบวกของ x, y, x^2, y^2 และ xy^2

ขั้นที่ 4 แทนค่าผลบวกขั้นที่ 3 ในรูป Σx, Σy, Σx^2, Σy^2 และ Σxy ลงในสูตร

ตัวอย่าง 5.7 ข้อมูลเกี่ยวกับคะแนนสอบวิชา ก. กับคะแนนสอบวิชา ข. ของนักศึกษา 5 คน เป็นดังนี้

คนที่	คะแนนวิชา ก. (x)	คะแนนวิชา ข. (y)	x^2	y^2	xy
1	1	2	1	4	2
2		4	4	16	8
3	4	5	16	25	20
4	5	7	25	49	35
5	5	8	25	64	40
ผลบวก	17	26	71	158	105

ดังนี้เราจะได้ค่า $\Sigma x = 17, \Sigma y = 26$

$\Sigma x^2 = 71, \Sigma y^2 = 158, \Sigma xy = 105$

แทนค่าลงในสูตรจะได้

$$r = \frac{5(105) - (17)(26)}{\sqrt{\{5(71) - (17)^2\}\{5(158) - (26)^2\}}}$$

$$= \frac{525-442}{\sqrt{\{355 - 289\}\{790 - 676\}}}$$

$$= \frac{83}{\sqrt{(66)(114)}} = \frac{83}{\sqrt{7524}} = \frac{83}{86.74} = 0.96$$

แบบฝึกหัดที่ 5.4

1) จากข้อมูลเกี่ยวกับคะแนนสอบวิชาภาษาอังกฤษ (x) กับคะแนนสอบวิชาคณิตศาสตร์ (y) ของนักเรียน 5 คน ข้างล่างนี้ จงคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ x และอธิบายด้วยว่าค่าสัมประสิทธิ์ที่ได้แสดงว่าคะแนนสอบของวิชานี้มีสัมพันธภาพกันอย่างไร

คนที่	x	y	x^2	y^2	xy
1	5	4	25	16	20
2	6	3	36	9	18
3	1	2	1	4	2
4	4	6	16	36	24
5	2	3	4	9	6
ผลบวก	18	18	82	74	70

5.6 วิธีคำนวณหาค่า r อย่างง่าย

ถ้านักศึกษาไม่มีเครื่องคิดเลขช่วยในการคำนวณ หรือเมื่อค่าของ x และ y มากเกินไป มีวิธีที่จะช่วยทำให้ค่าตัวเลขของตัวแปรทั้งสองน้อยลง และทำให้คำนวณได้สะดวกขึ้น โดยการลบค่าคงที่ออกจากค่าสังเกตของ x และค่าสังเกตของ y ซึ่งทำให้ได้ค่าใหม่ของตัวแปรทั้งสองน้อยลง เหมาะแก่การคำนวณ โดยยังให้ผลลัพธ์ในการคำนวณค่า r เหมือนเดิม

ค่าคงที่สำหรับที่จะลบออกจาก x และที่จะลบออกจาก y นั้นไม่จำต้องเท่ากัน โดยทั่วไปก็จะเลือกค่าที่เมื่อลบออกจากค่าสังเกตแล้วทำให้ได้ตัวเลขที่สะดวกแก่การคำนวณ

ถ้าเราเรียกตัวแปรสองตัวที่ถูกลบค่าคงที่ออกแล้วเป็น x* และ y* เราจะได้ว่า

x* = x - ค่าคงที่สำหรับ x

y* = y - ค่าคงที่สำหรับ y

ขั้นตอนการคำนวณเหมือนเดิมทุกอย่าง เพียงแต่ใช้ x* กับ y* แทน x กับ y

ตัวอย่าง 5.8 จากตัวอย่างที่แล้วมา ถ้าเราเลือกค่าคงที่สำหรับลบออกจาก x ให้เป็น 5 และค่าคงที่สำหรับลบออกจาก y เป็น 4 เราจะได้ค่า x* = x - 5 และ y* = y - 4

ตารางสำหรับการคำนวณค่า r โดยที่จะใช้ค่า x* และ y* แทน x และ y ในสูตรเป็นดังนี้ (เพียงแต่ลบ 5 ออกจากทุกค่าของ x และลบ 4 ออกจากทุกค่าของ y)

คนที่	x	y	x* = x - 5	y = y - 4	x^{*2}	y^{*2}	x*y*
1	1	2	-4	-2	16	4	-8
2	2	4	-3	0	9	0	0
3	4	5	-1	1	1	1	-1
4	5	7	0	3	0	9	0
5	5	8	0	4	0	16	0
ผลบวก			-8	6	16	30	7

ดังนั้นเราได้ค่า $\Sigma x^* = -8$ $\Sigma y^* = -6$

$\Sigma x^{*2} = 26$ $\Sigma y^{*2} = 30$ $\Sigma x^*y^* = 7$

แทนค่าลงในสูตรจะได้

$$r = \frac{5(7) - (-8)(6)}{\sqrt{[5(26) - (8)^2]\,[5(3) \quad - (6)^2]}}$$

$$r = \frac{35 + 48}{\sqrt{(66)(114)}} = \frac{83}{\sqrt{7254}} = \frac{83}{86.74} = 0.96$$

จะเห็นได้ว่าตัวเลขที่ใช้ในการคำนวณคราวนี้ทำให้คำนวณด้วยมือได้ง่ายขึ้นกว่าเดิม เพราะค่าตัวเลขเล็กลงกว่าในตัวอย่างที่แล้วมา แต่ยังให้คำตอบเท่าเดิม

แบบฝึกหัดที่ 5.5

1) จากข้อมูลคะแนนสอบวิชาภาษาอังกฤษ และคะแนนสอบวิชาคณิตศาสตร์ในแบบฝึกหัด 5.4 จงคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ เปรียบเทียบว่าได้คำตอบเหมือนเดิมหรือไม่ ถ้าใช้วิธีคำนวณแบบง่าย โดยนักศึกษาเลือกค่าคงที่สำหรับลบออกจากตัวแปรทั้งสอง
2) จากข้อมูลน้ำหนักตัวและความสูงของนักศึกษา 16 คนในตัวอย่างเกี่ยวกับสแกตเตอร์แกรม (ในตอน 5.3) จงคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างความสูงกับน้ำหนักตัว (เลือกวิธีคำนวณตามแต่ว่าอันไหนสะดวกสำหรับนักศึกษา เช่นถ้ามีเครื่องคิดเลขก็ไม่จำต้องใช้วิธีง่าย) ($r = 0.86$)

   จากข้อมูลเกี่ยวกับรายได้และจำนวนบุตรในครอบครัว (ท้ายแบบฝึกหัด 5.1) จงคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรทั้งสองนี้ แล้วตีความหมายของค่า r ที่ได้ว่าเหมือนกับที่ได้จากการพิจารณาสแกตเตอร์แกรมของตัวแปรทั้งสองตัวนี้ที่กล่าวมาข้างต้นหรือไม่ ($r = -0.93$)

5.7 สรุปความเข้าใจเกี่ยวกับสหสัมพันธ์ระหว่างตัวแปร

1) การที่ตัวแปรสองตัวมีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างกันสูง เป็นการแสดงว่า**ตัวแปรทั้งสองมีการเปลี่ยนร่วมกัน** (อาจเป็นไปในทางเดียวกันหรืออาจเป็นไปในทางตรงข้ามกัน) **แต่มิได้หมายความว่าตัวแปรหนึ่งจะต้องเป็นเหตุหรือเป็นผลของอีกตัวแปรหนึ่ง**

2) **สัมประสิทธิ์สหสัมพันธ์เป็นเพียงดัชนีหรือเครื่องวัด ที่บอกระดับสัมพันธภาพระหว่างตัวแปร แต่ไม่ได้บอกอะไรเกี่ยวกับสัดส่วนของสัมพันธภาพ** ถึงแม้เราจะบอกได้ว่าค่า $r = 8$ **แสดงสหสัมพันธ์ในระดับสูงกว่าค่า** $r = 0.4$ แต่เราไม่อาจกล่าวได้ว่าสูงกว่าเป็นสองเท่า **และเราไม่อาจกล่าวได้ว่า** การเพิ่มของสหสัมพันธ์จาก $r = 0.4$ เป็น $r = 0.6$ นั้น เท่ากับการเพิ่มจาก $r = 0.7$ เป็น $r = 0.9$ อย่างไรก็ตามสังเกตว่า $r = 0.6$ กับ $r = -0.6$ แสดงสัมพันธภาพที่มีระดับเท่ากันเพียงแต่มีทิศทางตรงกันข้าม

สัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้นั้นจะเปลี่ยนแปลงไปตามกลุ่มตัวอย่าง ในตัวอย่างชุดต่างกันที่มาจากประชากรเดียวกันก็จะมีค่า r ต่างกัน นอกจากนี้ก็ยังขึ้นกับขนาดตัวอย่าง (จำนวนคู่ค่าสังเกตในตัวอย่างนั้น) ว่ามีมากหรือน้อยเพียงใด เพราะเป็นไปได้ว่า ค่า r ที่คำนวณได้

นั้นอาจได้มาโดยความบังเอิญโดยไม่ได้แสดงว่าตัวแปรทั้งสองนั้นมีสัมพันธภาพต่อกันในระดับที่คำนวณได้จากตัวอย่าง ซึ่งอาจเป็นว่าสหสัมพันธ์ระหว่างตัวแปรทั้งสองนั้นอาจจะมีมากกว่าหรือน้อยกว่าที่คำนวณได้จากตัวอย่างก็เป็นได้ ดังนั้นการตีความหมายของสัมประสิทธิ์สหสัมพันธ์ในลักษณะของค่าสถิติที่ได้จากตัวอย่าง ทำนองเดียวกับค่า $\bar{x}$ และค่าสถิติอื่น ๆ จากตัวอย่าง จึงต้องอาศัยการพิจารณาขนาดของตัวอย่าง และมีการทดสอบว่าค่า r ที่ได้มีค่าเช่นนั้นโดยบังเอิญ หรือว่ามีสหสัมพันธ์อยู่จริงในประชากรของตัวแปรทั้งสอง (นั่นคือการทดสอบสมมติฐานเกี่ยวกับสัมประสิทธิ์สหสัมพันธ์ในประชากร หรือการทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์)

5.8 การตีความหมายของสหสัมพันธ์

ถ้าเราเลือกตัวอย่างของคู่ค่าสังเกตมาจากประชากร เราอาจจะได้ตัวอย่างที่ให้ค่า r ที่ไม่เป็นศูนย์ ทั้ง ๆ ที่ตัวแปรทั้งสองที่เราเลือกมาวัดสหสัมพันธ์นั้นไม่ได้สัมพันธภาพต่อกันเลย เช่น ถ้านักศึกษาลองคำนวณหาสัมประสิทธิ์สหสัมพันธ์ระหว่างเลขท้ายสองตัวของรหัสประจำตัวนักศึกษา กับน้ำหนักตัวเป็นกิโลกรัมของนักศึกษา โดยเลือกตัวอย่างเพื่อน ๆ นักศึกษามาสัก 5 คน ก็จะพบว่าค่า r ที่ได้นั้นมักจะไม่เป็น 0 และถ้าเลือกตัวอย่างเพื่อนนักศึกษามาอีกชุดหนึ่งก็อาจจะได้ค่า r ที่แตกต่างออกไป ดังนี้เราจึงจำเป็นต้องมีวิธีการทดสอบว่าค่าสหสัมพันธ์ที่ได้จากตัวอย่างนั้นแสดงว่ามีสหสัมพันธ์อยู่จริงหรือไม่ในประชากรของตัวแปรทั้งสอง วิธีการทดสอบนี้เรียกว่าการทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์ ซึ่งก็เป็นไปในทำนองเดียวกันกับการทดสอบสมมติฐานเกี่ยวกับค่าสถิติอื่น ๆ ในประชากรที่แล้วมานั่นเอง

ข้อสันนิษฐานเบื้องต้นในการตีความหมายของค่า r ก็คือ

1) ข้อมูลที่เป็นตัวอย่างได้มาเป็นคู่ ๆ โดยแต่ละคู่ได้มาโดยวิธีหาตัวอย่างแบบสุ่ม (random sampling) ซึ่งหมายความว่าแต่ละคู่ค่าสังเกตที่ได้มาเป็นอิสระจากคู่อื่น ๆ

2) ตัวแปรทั้งสองตัวที่จะหาสหสัมพันธ์เป็นตัวแปรที่มีค่าต่อเนื่อง หรือมาจากการวัดโดยใช้มาตราวัดแบบช่วง (กล่าวถึงภายหลัง)

3) สัมพันธภาพระหว่างตัวแปรทั้งสองเป็นแบบเส้นตรง (linear relationship) หรือเป็นแบบใกล้เคียงกับเส้นตรง

ข้อ 3) นี้เป็นข้อที่สำคัญที่สุดในการใช้และตีความหมายของสัมประสิทธิ์สหสัมพันธ์ r เพราะสัมประสิทธิ์สหสัมพันธ์ตัวนี้ใช้วัดสหสัมพันธ์แบบเส้นตรง ดังนั้นถ้าตัวแปรที่มีสัมพันธภาพระหว่างกันเป็นเส้นโค้ง จะพบว่าคำนวณค่า r ได้ต่ำ ซึ่งไม่ได้เป็นเพราะตัวแปรทั้งสองไม่มีสหสัมพันธ์ต่อกัน แต่เป็นเพราะสหสัมพันธ์ที่อยู่ต่อกันนั้นไม่อาจวัดได้โดยการใช้สัมประสิทธิ์ r

5.9 การทดสอบสมมติฐานเกี่ยวกับสัมประสิทธิ์สหสัมพันธ์ r

ข้อสันนิษฐานทางทฤษฎีเกี่ยวกับการทดสอบสัมประสิทธิ์สหสัมพันธ์ r ที่สำคัญก็คือว่ากลุ่มตัวอย่างที่ใช้ในการคำนวณหาค่า r จะต้องมาจากประชากรที่มีการแจกแจงตัวแปรที่สองที่ใช้คำนวณ r เป็นแบบที่เรียกว่า Bivariate Normal Distribution มิฉะนั้นก็ควรใช้วิธีอื่นในการวัดสหสัมพันธ์ แทนที่จะใช้ r ข้อสันนิษฐานนี้ใช้ได้โดยทั่วไป ถ้าตัวอย่างที่ใช้คำนวณค่า r มีขนาดใหญ่พอ (โดยทั่วไปถือว่าตัวอย่างที่มีค่าสังเกต 30 คู่ขึ้นไปเป็นตัวอย่างขนาดใหญ่) อย่างไรก็ตามเราจะถือว่าในการทดสอบที่จะกล่าวถึงต่อไปนี้ข้อสันนิษฐานที่ว่านี้เป็นจริง

ดังที่ได้กล่าวถึงแล้ว (ในตอน 5.7) ว่าแม้สหสัมพันธ์ในประชากรของตัวแปรทั้งสองจะไม่มีอยู่ หรือพูดอีกอย่างหนึ่งว่าสัมประสิทธิ์สหสัมพันธ์ในประชากร(ซึ่งเราใช้แทนด้วยสัญลักษณ์ ρ) จะเป็นศูนย์ก็ตาม ในการคำนวณค่า r จากตัวอย่างเรามักพบว่า $r \neq 0$ สิ่งที่เราต้องการทดสอบก็คือว่าค่า r ที่ได้จากตัวอย่างนั้น แตกต่างจากศูนย์มากพอที่จะเชื่อได้ว่า ค่าสัมประสิทธิ์สหสัมพันธ์ในประชากร (ρ) ไม่เป็นศูนย์ได้หรือไม่

การทดสอบนัยสำคัญของ r มีวิธีทดสอบได้หลายแบบ แต่ขั้นตอนที่ใช้เหมือนกันก็คือ

ขั้นที่ 1 ตั้งสมมุติฐานเกี่ยวกับสหสัมพันธ์ในประชากรระหว่างตัวแปรทั้งสองที่เราสนใจว่าไม่มีอยู่ นั่นคือ กล่าวว่า ρ เป็นศูนย์ และตั้งสมมุติฐานทางเลือกอีกอย่างว่ามีอยู่นั่นคือ ρ ไม่เป็นศูนย์ ได้สมมุติฐานในรูป $H_0 : \rho = 0$ และ $H_1 : \rho \neq 0$

ขั้นที่ 2 การทดสอบสมมุติฐานในขั้นต่อไปก็เป็นการตัดสินใจว่า จะเชื่อว่าสมมุติฐานอย่างใดถูก เมื่อกำหนดความน่าจะเป็นของการตัดสินใจผิดไว้ให้ นั่นคือเมื่อกำหนดระดับนัยสำคัญของการทดสอบ

ระดับนัยสำคัญของการทดสอบใช้สัญลักษณ์ α (อัลฟา) เป็นความน่าจะเป็นของการตัดสินใจผิด โดยปฏิเสธสมมุติฐานหลัก (H_0) เมื่อที่จริงแล้ว H_0 เป็นจริง (ปกติจะตั้งไว้ให้มีโอกาสตัดสินใจได้น้อยเช่นตั้งไว้ 0.05 (5%) หรือ 0.01 (1%) เป็นต้น

ขั้นที่ 3 หาค่าวิกฤตของตัวสถิติที่จะใช้เป็นเครื่องทดสอบ จากตารางค่าวิกฤตของตัวสถิติชนิดที่ใช้ (อาจเป็นตารางค่าวิกฤตของ r หรือของ t แล้วแต่กรณี)

ขั้นที่ 4 คำนวณหาค่าสถิติจากข้อมูลที่ต้องการทดสอบสมมุติฐาน ตามสูตรของตัวสถิติที่เลือกใช้ทดสอบ

ขั้นที่ 5 ตัดสินใจว่าจะยอมรับหรือจะปฏิเสธสมมุติฐานหลัก (H_o) ที่กล่าวว่าไม่มีสหพันธ์ระหว่างตัวแปรทั้งสองในประชากร โดยจะปฏิเสธ H_o ถ้าค่าสถิติที่คำนวณได้จากข้อมูล (ไม่คิดเครื่องหมาย) มากกว่าค่าวิกฤตของตัวสถิตินั้น (ที่ได้ในขั้นที่ 3)

ในบทนี้จะได้กล่าวถึงการทดสอบสมมติฐานเกี่ยวกับสัมประสิทธิ์สหสัมพันธ์โดยวิธีมาตรฐาน 3 วิธี

5.9.1. การทดสอบสหสัมพันธ์ โดยใช้ตารางค่าวิกฤตของ r

ตารางค่าวิกฤตของ r (ดูตารางที่ 10 ในภาคผนวก) ได้มาโดยการคำนวณความน่าจะเป็นของการคำนวณได้ค่า r ต่าง ๆ กับเมื่อกำหนดจากจำนวนตัวอย่างที่ใช้ วิธีนี้เป็นวิธีง่ายที่สุดใน 3 วิธีที่ใช้

วิธีการอ่านตารางค่าวิกฤต r หาค่า r วิกฤตในตารางที่ตรงกับระดับนัยสำคัญ (ช่องแนวตั้ง) กับจำนวนตัวอย่างที่ใช้ (ช่องแนวนอน)

ตัวอย่าง 5.9 จากตัวอย่าง 5.8 เราคำนวณค่าสหสัมพันธ์ r ระหว่างคะแนนสอบวิชา ก. กับคะแนนสอบวิชา ข. ของนักศึกษา 5 คน ได้ $r = 0.96$ จะพูดได้ไหมว่า สหสัมพันธ์ที่คำนวณได้มีนัยสำคัญ หรือว่ามีสหสัมพันธ์ระหว่างคะแนนสอบสองวิชานี้ในประชากรนักศึกษาที่สอบทั้งสองวิชา โดยจะทดสอบที่ระดับนัยสำคัญ 5% (ซึ่งหมายความว่าเราจะยอมเชื่อว่ามีสหสัมพันธ์ในประชากรนักศึกษาที่ไม่เป็นศูนย์อยู่จริง ถ้าความน่าจะเป็นของการได้ค่า $r = 0.96$ จากตัวอย่างขนาด $n = 5$ โดยความบังเอิญน้อยกว่า 0.05 หรือเกิดขึ้นได้น้อยกว่า 5 ครั้งใน 100 ครั้งนั่นเอง

ดำเนินการทดสอบตามขั้นตอนที่แสดงไว้ข้างต้น

ขั้นที่ 1 สมมุติฐานหลักคือ สมมุติฐานที่กล่าวว่าไม่มีสหสัมพันธ์ในประชากร

$H_o : \rho = 0$

สมมุติฐานเลือกคือ สมมุติฐานที่กล่าวว่ามีสหสัมพันธ์ในประชากร

$H_1 = \rho \neq 0$

ขั้นที่ 2 ระดับนัยสำคัญของการทดสอบ กำหนดให้ $\alpha = 0.05$

ขั้นที่ 3 ค่าวิกฤตของ r จากตารางค่าวิกฤต เมื่อ $\alpha = 0.05$ และ $n = 5$ คือ 0.878

ขั้นที่ 4 ค่า r ในตัวอย่างนี้ ถูกคำนวณแล้วเท่ากับ 0.96

ขั้นที่ 5 เพราะว่าค่า $|r|$ (r ที่ไม่คิดเครื่องหมาย) มากกว่าค่าวิกฤตของ r ที่ได้จากตารางค่าวิกฤต $(0.96 > 0.878)$

ดังนั้นเราจึงตัดสินใจปฏิเสธสมมุติฐานหลัก (H_o) ที่ว่าไม่มีสหสัมพันธ์ในประชากร นั่นคือ เราจะกล่าวได้ว่าค่าสหสัมพันธ์ที่คำนวณได้จากตัวอย่าง มีนัยสำคัญที่ระดับ 5% (หรือว่าค่า r ที่ได้จากตัวอย่างไม่น่าจะได้มาโดยบังเอิญ เพราะความน่าจะเป็นมีเพียง 5% ในการได้ โดยบังเอิญแล้วเท่ากับหรือมากกว่า % ค่าวิกฤตนี้

จากการทดสอบเราสรุปได้ว่า มีสหสัมพันธ์ระหว่างคะแนนสอบของวิชา ก. กับคะแนนสอบวิชา ข. อยู่ในประชากรของนักศึกษาที่สอบทั้งสองวิชา ที่ระดับนัยสำคัญ 5%

แบบฝึกหัดที่ 5.6

1) จากแบบฝึกหัด 5.4 จงทดสอบสมมุติฐานเกี่ยวกับสหสัมพันธ์ในประชากรตัวอย่างที่ใช้คำนวณค่า r ว่ามีอยู่จริงหรือไม่ ใช้ระดับนัยสำคัญ 5% และ 1% อธิบายผลการทดสอบมาด้วยว่าจะสรุปได้อย่างไรเกี่ยวกับสหสัมพันธ์ที่ได้จากตัวอย่าง

ข้อสังเกต

ก) ถ้าเราพบในการทดสอบว่า ค่าสถิติที่คำนวณได้นั้นน้อยกว่าค่าวิกฤตจากตาราง (อย่าลืมว่าไม่คิดเครื่องหมาย) เราไม่อาจจะปฏิเสธสมมุติฐานนั้นที่ว่าไม่มีสหสัมพันธ์ในประชากร นั่นคือ เราจะกล่าวได้ว่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้ไม่มีนัยสำคัญ (ที่ระดับนัยสำคัญที่กำหนด) ซึ่งหมายความว่าค่าที่คำนวณได้จากตัวอย่างอาจจะได้มาโดยบังเอิญด้วยความน่าจะเป็นมากกว่าระดับนัยสำคัญที่เรากำหนดไว้เป็นเกณฑ์ ทั้ง ๆ ที่ในประชากรแล้วไม่มีสหสัมพันธ์อยู่เลยก็เป็นได้

ข) อย่าลืมว่าความแม่นยำของการทดสอบนั้นขึ้นอยู่กับขนาดของตัวอย่างที่ใช้ในการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ เช่นเดียวกับการประมาณค่าสถิติของประชากรอื่น ๆ (เช่น ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน) ตัวอย่างที่มีขนาดใหญ่ขึ้นทำให้ประมาณค่าจริงในประชากรได้ใกล้เคียงมากขึ้น

5.9.2 การทดสอบสหสัมพันธ์โดยใช้ตัวสถิติ t

วิธีนี้เป็นวิธีดีเท่ากับการใช้ตารางค่าวิกฤตของ r ใช้เมื่อเราไม่มีตารางค่าวิกฤตของ r โดยใช้ตัวสถิติ t เปรียบเทียบกับตารางค่าวิกฤตของ t สูตรที่ใช้คือ $t = \frac{r\sqrt{(n-2)}}{\sqrt{1-r^2}}$

ค่าวิกฤตของ t หาได้จากตารางค่าวิกฤตของ t (ดูตารางที่ 9 ในภาคผนวก) ตามระดับนัยสำคัญที่ต้องการ และตามขนาดตัวอย่างที่ใช้ (n) หรือถ้าไม่มีช่องแสดงขนาดตัวอย่าง ก็ดูจากช่องที่แสดงองศาความเป็นอิสระ (degree of freedon = df) โดยที่ df = n-2 นั่นคือ ถ้าเราใช้ตัวอย่างขนาด n=5 เราจะได้ df = 5-2 = 3

การทดสอบเป็นไปตามขั้นตอนเช่นที่แล้วมา

ตัวอย่าง 5.10 จากตัวอย่างที่แล้วมา ถ้าเราต้องการทดสอบนัยสำคัญของค่า r=0.96 ที่ได้จากตัวอย่างขนาด n=5 ที่ระดับนัยสำคัญ 5% โดยใช้ตัวสถิติ t

ขั้นที่ 1 $H_o : \rho=0$

$H_1 : \rho\neq 0$

ขั้นที่ 2 ระดับนัยสำคัญที่กำหนดคือ $\alpha=0.05$

ขั้นที่ 3 ค่าวิกฤตของ t (เมื่อ $\alpha=0.05$ และ df = 5-2 = 3 จากตารางค่าวิกฤตของ t) คือ 2.35

ขั้นที่ 4 คำนวณค่าสถิติ t จากสูตร

$$t = \frac{0.96\sqrt{(5-2)}}{\sqrt{1-(0.96)^2}} = \frac{1.66}{\sqrt{0.08}} = \frac{1.66}{0.28} = 5.39$$

ขั้นที่ 5 เพราะว่าค่า |t| (t ที่ไม่คิดเครื่องหมาย) มากกว่าค่าวิกฤตของ t จากตาราง (5.93>2.35)

ดังนั้น เราจะปฏิเสธสมมุติฐานหลัก (H_o) ที่ว่าไม่มีสหสัมพันธ์ในประชากรนั่นคือ เราจะกล่าวได้ว่าค่าสหสัมพันธ์ที่คำนวณได้จากตัวอย่างมีนัยสำคัญที่ระดับ 5% (หรือว่า ค่า r ที่ได้จากตัวอย่างไม่น่าจะได้มาโดยบังเอิญ เพราะความน่าจะเป็นของการได้ r สูงกว่าค่าวิกฤตนั้นน้อยกว่า 5%)

จากการทดสอบเราสรุปได้ว่ามีสหสัมพันธ์ระหว่างคะแนนสอบของวิชา ก. กับคะแนนสอบวิชา ข. อยู่ในประชากรของนักศึกษาที่สอบทั้งสองวิชา ที่ระดับนัยสำคัญ 5%

แบบฝึกหัด 5.7

1) จากแบบฝึกหัด 5.6 จงทดสอบนัยสำคัญของสัมประสิทธิสหสัมพันธ์ r ที่ได้โดยใช้ตัวสถิติ t ที่ระดับนัยสำคัญ 5% ได้ผลการทดสอบเหมือนกับการใช้ค่าวิกฤตของ r ที่แล้วมาหรือไม่

5.10 การวิเคราะห์ความถดถอยและการพยากรณ์

(Regression Analysis and Prediction)

บ่อยครั้งที่เราต้องการพยากรณ์ค่าตัวแปรหนึ่งโดยอาศัยค่าที่ทราบของตัวแปรอีกตัวหนึ่ง เช่นเราอาจต้องการพยากรณ์ระดับเชาวน์ปัญญาเมื่อเป็นผู้ใหญ่ โดยอาศัยเชาวน์ปัญญาในวัยเด็กเป็นเครื่องพยากรณ์ หรือพยากรณ์ปริมาณผลผลิตข้าวโพดจากปริมาณปุ๋ยที่ใช้ ฯลฯ วิธีการที่เราใช้ในการพยากรณ์นี้เรียกว่าการวิเคราะห์ความถดถอย

ในขั้นนี้เราจะใช้วิธีการพยากรณ์ตัวแปรโดยอาศัยการสร้างสมการถดถอย (regression equation) ซึ่งเป็นสมการที่บรรยายเส้นตรงที่แสดงสัมพันธภาพระหว่างตัวแปรสองตัว แทนที่กลุ่มจุดในสแกตเตอร์แกรมที่เราได้เห็นมาแล้ว

จากที่แล้วมาเราได้เห็นว่า ถ้าสหสัมพันธ์ระหว่างตัวแปรเป็นแบบสมบูรณ์ (-1 หรือ +1) กลุ่มจุดค่าสังเกตบนสแกตเตอร์แกรมจะอยู่บนเส้นตรงพอดี ดังนั้นถ้าเราสามารถหาสมการแสดงเส้นตรงนั้นได้ เราก็จะสามารถบอกค่าของตัวแปรตามได้เมื่อทราบค่าของตัวแปรอิสระ (ขึ้นอยู่กับว่าเราจะให้ตัวแปรใดเป็นตัวแปรตาม) แต่ถ้าสหสัมพันธ์ระหว่างตัวแปรไม่เป็นแบบสมบูรณ์ ก็ขึ้นอยู่กับว่าสหสัมพันธ์ระหว่างตัวแปรทั้งสองนั้นสูงเพียงใด (ใกล้-1 หรือ +1 เพียงใด) ที่จะทำให้เราหาเส้นตรงแสดงกลุ่มจุดบนได้ดีหรือใกล้เคียงกับแนวของกลุ่มจุดเหล่านั้น ตามไปด้วย

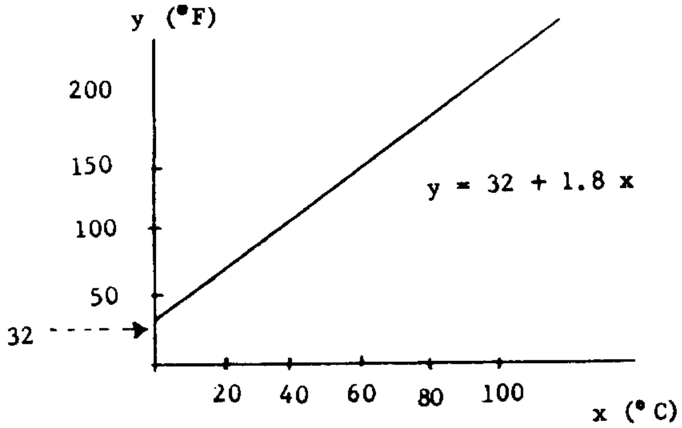
5.10.1 สมการเส้นตรง

สมการที่ใช้บรรยายเส้นตรงบนกราฟนั้นคือสมการเส้นตรง (linear equation) ซึ่งนักศึกษาเคยรู้จักมาแล้วในชั้นมัธยม ตัวอย่างของตัวแปรสองตัวที่มีสหสัมพันธ์ต่อกันแบบสมบูรณ์ที่เห็นได้ง่ายได้แก่สัมพันธภาพระหว่างอุณหภูมิที่วัดเป็นองศาเซ็นติเกรด กับอุณหภูมิที่วัดเป็นองศาฟาเรนไฮต์ ซึ่งมีสูตรหรือสมการแสดงความสัมพันธ์ระหว่างกันดังนี้

$$y = 32 + 1.8x$$

เมื่อ x คืออุณหภูมิวัดเป็น C และ y คืออุณหภูมิวัดเป็น F

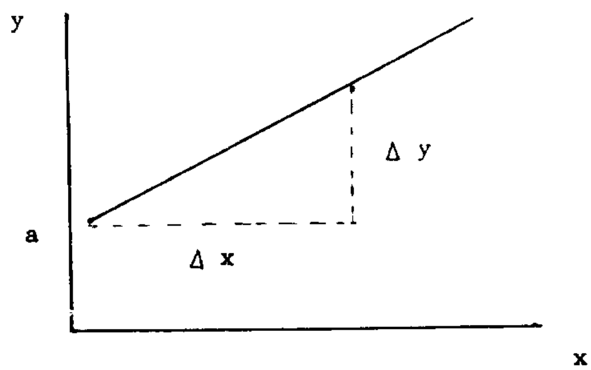
จากสมการนี้เราอาจเขียนกราฟเส้นตรงแสดงสัมพันธภาพระหว่าง x กับ y ได้ดังนี้



และเราสามารถจะบอกค่าของ y เมื่อทราบค่าของ x โดยดูจากกราฟหรือโดยการแทนค่า x ลงในสมการเส้นตรงข้างต้น เช่น เมื่อ x = 10 °C จะได้ y = 50 °F

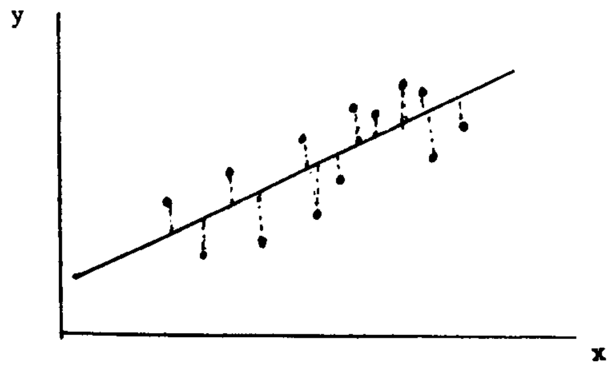
นักศึกษาจะเห็นแล้วว่าสมการเส้นตรงนั้นอยู่ในรูป $y = a + bx$

เมื่อ a คือค่าจุดตัดบนแกน y ของเส้นตรง แสดงค่า y เมื่อ x = 0 และ b คือค่าความลาดของเส้นตรง (จากกราฟ $b = \frac{\Delta x}{\Delta y}$) แสดงอัตราการเปลี่ยนแปลงของ y ต่อหนึ่งหน่วยของ x (นั่นคือเมื่อ x เปลี่ยนไป 1 หน่วย y จะเปลี่ยนไป b หน่วย)



5.11 สมการถดถอยและเส้นถดถอย (Regression Equation and Regression Line)

จากสแกตเตอร์แกรมแสดงจุดค่าสังเกตระหว่างคะแนนสอบวิชา ก. และคะแนนสอบวิชา ข. จะเห็นว่าเราสามารถลากเส้นตรงแสดงแทนกลุ่มจุดค่าสังเกตที่จะทำให้เส้นตรงนั้นเข้ากับกลุ่มจุดได้หลายวิธี รวมทั้งการลากเส้นตรงโดยกะเอาด้วยสายตา แต่วิธีการที่เราจะใช้ต่อไปนี้ทำโดยการใช้วิธีทางคณิตศาสตร์เข้าช่วย เพื่อให้ได้เส้นตรงที่เข้ากับกลุ่มจุดค่าสังเกตได้ดีที่สุด วิธีที่เราจะใช้นี้เรียกว่า วิธีกำลังสองน้อยที่สุด (Least Squares Method) ซึ่งจะทำให้ผลบวกของกำลังสองของระยะทางระหว่างจุดค่าสังเกตกับเส้นตรง (วัดแนวตั้ง) มีค่าน้อยที่สุด โดยวิธีนี้เราจะได้สมการเส้นตรงที่จะใช้สร้างเส้นตรงที่มีสมบัติดังกล่าวนี้กับข้อมูลที่เราได้ เรียกว่าเส้นถดถอย



โดยวิธีกำลังสองน้อยที่สุดเราจะได้สูตรสำหรับคำนวณหาค่า b และ a ซึ่งจะเรียกว่าสัมประสิทธิ์การถดถอย (Regression Coefficient) สำหรับสมการถดถอยได้ดังนี้

$$b = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\ \Sigma x^2 - (\Sigma x)^2}$$

$$a = \bar{y} - b\bar{x}$$

$$(\bar{x} = \Sigma x/n;\ \bar{y} = \Sigma y/n)$$

5.11.1 การคำนวณค่าสัมประสิทธิ์การถดถอย

วิธีการคำนวณใช้ขั้นตอนเช่นเดียวกับการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ แล้วแทนค่าลงในสูตร โดยที่การคำนวณค่า b นั้นอาจใช้วิธีคำนวณจากข้อมูลโดยตรง หรือจะใช้วิธีแปลงค่าตัวแปรแบบเดียวกับการคำนวณค่า r อย่างง่ายก็ได้ แต่ในการคำนวณค่า a นั้น ต้องใช้ค่า $\bar{x}$ และ $\bar{y}$ จากตัวแปรดั้งเดิมที่ไม่ได้ถูกแปลงค่าให้เล็กลง

5.11.2 สมบัติของเส้นถดถอยแบบกำลังสองน้อยที่สุด

1) เป็นเส้นตรงที่ทำให้ผลบวกของกำลังสองของระยะทางแนวตั้งระหว่างจุดค่าสังเกตกับเส้นตรงนั้นมีค่าน้อยที่สุด

2) เป็นเส้นตรงที่ผ่านจุดพิกัด $(\bar{x}, \bar{y})$ บนสแกตเตอร์แกรม

3) ถ้าสหสัมพันธ์ระหว่างตัวแปร x กับ y เป็นศูนย์ เส้นถดถอยจะเป็นเส้นตรงที่ขนานกับแกนนอน (ค่าความลาด หรือ $b = 0$ ไปด้วย) และจะตัดแกนตั้งที่ $\bar{y}$ (นั่นคือ $a = \bar{y}$)

5.11.3 วิธีการเขียนเส้นถดถอย

จากสมบัติของเส้นถดถอย เรารู้ว่าเส้นถดถอยจะต้องตัดแกนตั้งที่ a และต้องผ่านจุด $(\bar{x}, \bar{y})$ ดังนั้นเราก็เพียงหาจุดพิกัด (0, a) กับ $(\bar{x}, \bar{y})$ เราก็จะได้เส้นถดถอยที่ต้องการ (ทั้งนี้เราต้องคำนวณค่า b ก่อน เพื่อให้ได้ค่า a รวมทั้งคำนวณค่า $\bar{x}$ และ $\bar{y}$ จากข้อมูลดั้งเดิมด้วย)

ตัวอย่าง 5.11 จากตัวอย่าง 5.10 ดำเนินขั้นตอนดังที่ทำมาแล้วในการคำนวณค่า r ได้ค่าผลบวก (จากข้อมูลดั้งเดิมที่ไม่ได้แปลง) คือ

$\Sigma x = 17 \quad \Sigma y = 26$

$\Sigma x^2 = 71 \quad \Sigma y^2 = 158 \quad \Sigma xy = 105$

แทนค่าลงในสูตรจะได้

$$= \frac{5(105) - (17)(26)}{5(71) - (17)^2} = \frac{83}{66} = 1.26$$

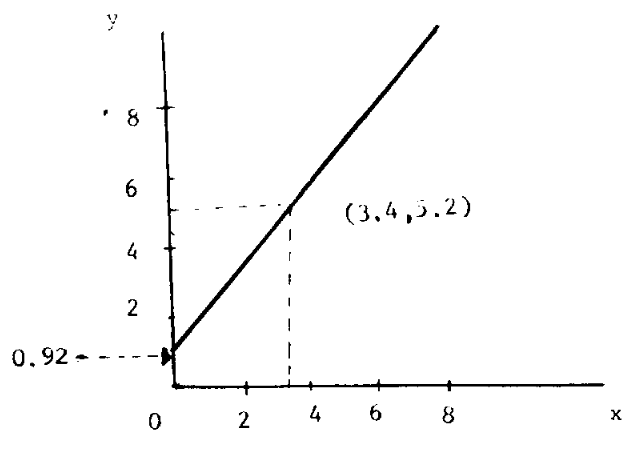
และได้

$$\bar{x} = \frac{17}{5} = 3.4 \; ; \; \bar{y} = \frac{26}{5} = 5.2$$

ดังนั้น $a = 5.2 - 1.26\,(3.4) = 5.2 - 4.28 = 0.92$

สมการถดถอยของข้อมูลชุดนี้คือ $y = 0.92 + 1.26x$

เราจะเขียนเส้นถดถอยของข้อมูลชุดนี้บนสแกตเตอร์แกรมได้ดังนี้



5.12 วิธีพยากรณ์ค่าตัวแปรตามเมื่อกำหนดค่าตัวแปรอิสระ

ดังที่กล่าวมาแล้วข้างต้นว่าเราอาจเลือกตัวแปรใดเป็นตัวแปรตาม (ซึ่งแทนด้วยแกนตั้งในสแกตเตอร์แกรม) หรือเป็นตัวแปรอิสระ (ซึ่งแทนด้วยแกนนอน) ขึ้นอยู่กับว่าเราสนใจจะพยากรณ์ค่าตัวแปรใดก็เลือกให้ตัวแปรนั้นเป็นตัวแปรตาม ซึ่งแทนด้วย y ในการคำนวณแล้วก็คำนวณหาค่าสัมประสิทธิ์การถดถอย เพื่อสร้างสมการถดถอยดังที่แล้วมา ถ้าเราทราบค่าของตัวแปรอิสระแล้วต้องการพยากรณ์ว่าตัวแปรตามจะมีค่าเท่าไร เราก็เพียงแต่แทนค่าตัวแปรอิสระ (x) นั้นลงในสมการถดถอย ก็จะได้คำตอบเป็นค่าพยากรณ์ของตัวแปรตาม (y)

ถ้าเราให้ x' เป็นค่าตัวแปรอิสระที่เรากำหนด

และ Y' เป็นค่าตัวแปรตามที่ต้องการพยากรณ์

ค่าตัวแปรตามที่เราต้องการพยากรณ์จะอยู่ในรูป

$$Y' = a + bx'$$

ตัวอย่าง 5.12 จากตัวอย่าง 5.11 ถ้าเราทราบคะแนนสอบวิชา ก. ของนักเรียนกลุ่มหนึ่งว่าได้คะแนน และต้องการพยากรณ์ว่าถ้าเขาสอบวิชา ข. จะสอบได้คะแนนเท่าไร

ในที่นี้เรากำหนด $x' = 3$ แทนค่าลงในสมการถดถอยที่เราทราบมาแล้ว

$$Y' = 0.92 + 1.26\,(3) = 0.92 + 3.78 = 4.70$$

ดังนั้นเราจะพยากรณ์ว่านักเรียนกลุ่มนี้จะสอบวิชา ข. ได้ 4.70 คะแนน

อย่าลืมว่าค่าการพยากรณ์นี้เป็นค่าจากสมการถดถอย และเนื่องจากสหสัมพันธ์ระหว่างตัวแปรไม่เป็นแบบสมบูรณ์ ค่าสังเกตที่ได้บนสแกตเตอร์แกรมไม่ได้อยู่บนเส้นถดถอยทั้งหมด ค่าพยากรณ์นั้นจึงไม่จำเป็นต้องมีค่าตรงกับค่าสังเกตที่ได้ ค่าพยากรณ์นั้นเป็นค่าประมาณค่าเฉลี่ยของคะแนนสอบวิชา ข. ของกลุ่มนักศึกษาในประชากรพวกที่สอบวิชา ก. ได้คะแนนเท่ากับ 3 ไม่ใช่เป็นค่าพยากรณ์ของนักศึกษาคนหนึ่งคนใดโดยเฉพาะ

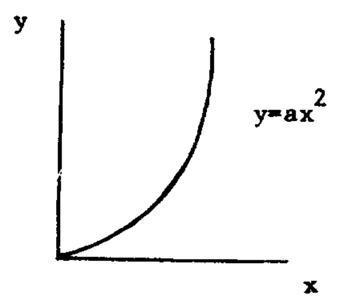
5.13 ความเข้าใจเกี่ยวกับการวิเคราะห์ความถดถอยและการพยากรณ์

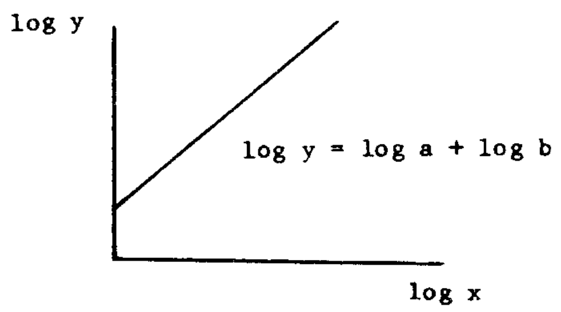
เมื่อเราใช้เส้นถดถอยในการพยากรณ์ค่าของตัวแปรตามจากค่าของตัวแปรอิสระ เราใช้ข้อสันนิษฐานที่ว่าตัวแปรทั้งสองมีสหสัมพันธ์ต่อกันอย่างมีนัยสำคัญ มิฉะนั้นแล้วการพยากรณ์ก็จะไม่มีความหมายแต่อย่างใด ถ้าตัวแปรทั้งสองไม่สหสัมพันธ์เส้นตรงต่อกัน ดังนั้น เราจึงควรทดสอบนัยสำคัญของสหสัมพันธ์ระหว่างตัวแปรก่อนว่ามีอยู่หรือไม่ แล้วจึงใช้สมการถดถอยหรือเส้นถดถอยเพื่อการพยากรณ์ นอกจากนี้การทดสอบว่าสมการถดถอยของเราใช้พยากรณ์ได้ดีหรือไม่ ก็อาจทำให้โดยการหาตัวอย่างอีกชุดหนึ่งนอกเหนือไปจากตัวอย่างชุดที่ใช้ในการคำนวณหาสัมประสิทธิ์การถดถอยเพื่อเปรียบเทียบค่าพยากรณ์กับค่าสังเกตจริงของตัวแปรตามว่าใกล้เคียงกันมากน้อยเพียงไร

ข้อสันนิษฐานที่สำคัญอีกอันหนึ่งในการพยากรณ์ ก็คือว่าค่าของตัวแปรอิสระที่ใช้ในการพยากรณ์ค่าของตัวแปรตามนั้นควรจะอยู่ในช่วงค่าสังเกตที่ใช้ในการคำนวณหาสัมประสิทธิ์สหสัมพันธ์ และสัมประสิทธิ์การถดถอยของข้อมูล เพราะเส้นถดถอยที่เราสร้างขึ้นสำหรับข้อมูลชุดนั้นอาศัยสมมุติฐานที่ว่าตัวแปรทั้งสองในช่วงข้อมูลที่คำนวณมีสัมพันธภาพต่อกันเป็นแบบเส้นตรงหรือใกล้เคียงเส้นตรง แต่ถ้าเราออกนอกช่วงค่าสังเกตของตัวแปรอิสระในข้อมูลที่ใช้สร้างเส้นถดถอยนั้น เราไม่อาจทราบได้ว่าสัมพันธภาพระหว่างตัวแปรทั้งสองยังจะอยู่ในรูปเดิมหรือไม่ เช่นถ้าเราพยากรณ์ความสูงของกลุ่มคนที่มีน้ำหนักตัว 150 กก. (ซึ่งเกินกว่าช่วงค่าสังเกตที่เราใช้ในการสร้างเส้นถดถอยระหว่างความสูงและน้ำหนักตัว) แล้วได้ค่าความสูงออกมาเท่ากับ 5.6 เมตร ก็จะเห็นได้ว่าห่างไกลจากความเป็นจริงอย่างยิ่ง ทั้งนี้เป็นเพราะว่าเมื่อน้ำหนักตัวอยู่ในระดับสูงมาก ๆ นั้น สหสัมพันธ์ระหว่างน้ำหนักตัวไม่อยู่ในรูปเส้นตรง ดังที่เราได้ในช่วงน้ำหนักตัวปกติของคนทั่วไป (อย่างไรก็ตามในการพยากรณ์บางอย่าง เราจะสันนิษฐานว่าสัมพันธภาพเป็นแบบเส้นตรงตลอด และพยากรณ์เลยไปจากช่วงข้อมูลที่ใช้สร้างสมการถดถอย)

ข้อสังเกตอย่างหนึ่งก็คือว่าเมื่อตัวแปรทั้งสองไม่มีสหสัมพันธ์ต่อกันนั้น เราได้ $r = 0$ และได้ค่า $b = 0$ ด้วย การทดสอบนัยสำคัญว่า r แตกต่างจากศูนย์อย่างมีนัยสำคัญหรือไม่ จึงเป็นการทดสอบด้วยว่า b แตกต่างจากศูนย์อย่างมีนัยสำคัญหรือไม่

นอกจากที่ได้กล่าวมาแล้วเกี่ยวกับตัวแปรที่มีสัมพันธภาพต่อกันแบบเส้นตรง ในหลายกรณี เราจะพบว่าตัวแปรสองตัวที่มีสัมพันธภาพต่อกันเป็นแบบเส้นโค้งนั้น เราสามารถจะแปลงค่าตัวแปรตัวหนึ่งหรือทั้งสองตัวเพื่อให้ได้สัมพันธภาพระหว่างตัวแปรที่แปลงค่าแล้วนี้อยู่ในแบบเส้นตรงได้ ตัวอย่างเช่น ถ้าเราทราบว่าสัมพันธภาพระหว่างตัวแปร x กับ y อยู่ในรูปของสมการ $y = ax^2$ ซึ่งเป็นสมการเส้นโค้ง แต่ถ้าเราแปลงค่าตัวแปรทั้งสองเป็นค่า log โดยการหาค่า log ของสมการ เราจะได้สมการใหม่ในรูป $\log y = \log a + 2 \log x$ ซึ่งเป็นสมการเส้นตรง นั่นคือสัมพันธภาพระหว่าง $\log x$ กับ $\log y$ จะอยู่ในรูปเส้นตรง





แบบฝึกหัดที่ 5.7

1. จากตัวอย่าง 5.12 ลองคำนวณหาค่า b โดยใช้ตัวแปรที่แปลงให้เล็กลงตามวิธีการคำนวณค่า r อย่างง่าย เพื่อดูว่าจะได้ค่า b เท่าเดิมหรือไม่ (อย่าลืมว่าค่า $\bar{x}, \bar{y}$ ที่ใช้ในการคำนวณ a ต้องมาจากตัวแปรดั้งเดิมที่ไม่ได้แปลงค่า)
2. จากคำถามที่แล้ว ๆ มา จงคำนวณหาสัมประสิทธิ์การถดถอย (เขียนสมการถดถอยของข้อมูลชุดนั้น ๆ และเขียนเส้นถดถอยของข้อมูลบนสแกตเตอร์แกรมของข้อมูลเหล่านั้น เปรียบเทียบดูว่า เส้นถดถอยเข้ากับจุดค่าสังเกตได้ดีมากน้อยเพียงใด เมื่อค่าสัมประสิทธิ์สหสัมพันธ์ต่างกัน

5.14 ความคลาดเคลื่อนมาตรฐานของค่าประมาณ (Standard Error of Estimate)

เราอาจจะมอบค่าพยากรณ์ของตัวแปรตาม (y') ในแง่ของค่าประมาณสำหรับตัวแปรตาม (y) และเนื่องจากค่าพยากรณ์นั้นไม่ตรงกับค่าสังเกตเสมอไป เราจะวัดได้ว่าเส้นถดถอยและสมการพยากรณ์ค่าตัวแปรตามได้ดีเพียงไร โดยการวัดความเปลี่ยนแปรจากค่าสังเกตของค่าพยากรณ์เมื่อตัวแปรอิสระมีค่าต่าง ๆ กัน (ในทำนองเดียวกับการวัดความเบี่ยงเบนมาตรฐานของตัวแปรเดียวที่เราทำมาแล้ว) ในที่นี้เราเรียกตัวสถิตินี้ว่าความคลาดเคลื่อนมาตรฐาน เพราะเกี่ยวกับการพยากรณ์ค่าของ y โดยใช้ x สูตรในการคำนวณ เป็นดังนี้

$$S_{yx} = S_y \sqrt{(1-r^2)}$$

เมื่อ S_{yx} = ความคลาดเคลื่อนมาตรฐานของค่าประมาณ โดยใช้

S_{yx} = ความเบี่ยงเบนมาตรฐานของ คำนวณตามสูตรที่เคยใช้คือ

$$S_y = \sqrt{\frac{\Sigma y^2 - (\Sigma y)^2/n}{n-1}}$$

r = สหสัมพันธ์ระหว่าง x กับ y

จะเห็นได้จากสูตรว่าถ้าตัวแปรมีสหสัมพันธ์แบบสมบูรณ์ต่อกัน $(r = -1$ หรือ $+1)$ ค่าความคลาดเคลื่อนมาตรฐานจะเป็นศูนย์ และยิ่งตัวแปรมีสหสัมพันธ์ต่อกันมากขึ้นเท่าไร ก็จะได้คลาดเคลื่อนมาตรฐานน้อยลงเท่านั้น ซึ่งเป็นเครื่องแสดงให้เห็นว่าค่าประมาณของ y หรือค่าพยากรณ์ที่ได้จากสมการถดถอยจะยิ่งดีขึ้นถ้าตัวแปรทั้งสองมีสหสัมพันธ์สูงขึ้น

นอกจากนี้เราอาจใช้ค่า S_{yx} ในการสร้างช่วยความเชื่อมั่นของค่าประมาณ (Confidence Interval of Estimate) ซึ่งใช้ในการบอกให้เราทราบช่วงของค่าพยากรณ์ ที่เราจะเชื่อได้ว่าจะครอบคลุมค่าสังเกตของตัวแปรตาม ด้วยความน่าจะเป็นที่เรากำหนด (เป็นเรื่องนอกเหนือจากหลักสูตรนี้ขึ้นไป)

ตัวอย่าง 5.13 จากตัวอย่าง 5.12 คำนวณหาความคลาดเคลื่อนมาตรฐานของค่าประมาณ y (S_{yx}) ได้ดังนี้

เราทราบ $\Sigma y = 26, \Sigma y^2 = 158, r = 0.96, n = 5$

คำนวณ $S_y = \sqrt{\frac{158 - (26)^2/5}{5-1}} = \sqrt{\frac{158 - 135.2}{4}} = \sqrt{\frac{21.8}{4}}$

$= \sqrt{5.70} = 2.39$

ดังนั้น $= 2.39\sqrt{(1 - (0.96)^2)}$

$= 2.39\sqrt{(1 - (0.92))}$

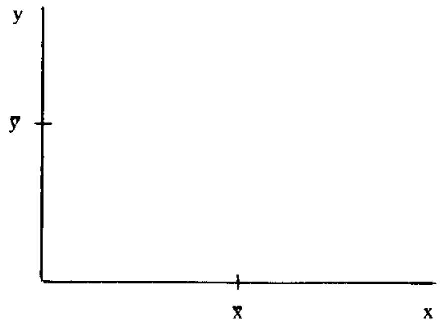
$= 2.39\sqrt{(0.078)}$

$= 2.39\ (0.28)$

$= 0.67$

แบบฝึกหัดที่ 5.8

1) ถ้าเราพบว่าสหสัมพันธ์ระหว่างตัวแปร x กับตัวแปร y เป็นศูนย์ จงเขียนเส้นถดถอยของข้อมูลชุดนี้ลงในสแกตเตอร์แกรมต่อไปนี้



2) เมื่อเราใช้เส้นถดถอย เราสันนิษฐานว่าสัมพันธภาพระหว่างตัวแปรทั้งสองเป็นแบบ...(ก) ถ้าสัมพันธ์ที่ได้มีค่าสูง แสดงว่าตัวแปรทั้งสองมีสัมพันธภาพต่อกันแบบเส้นตรงหรือใกล้เคียงกันแบบเส้นตรงใช่หรือไม่...(ข) แต่ถ้าสหสัมพันธ์ที่คำนวณได้มีค่าต่ำ ซึ่งอาจเป็นไปได้ว่าไม่มีสหพันธ์ระหว่างตัวแปรในประชากร เป็นไปได้หรือไม่ว่าที่จริงแล้ว ตัวแปรทั้งสองอาจมีสัมพันธภาพต่อกันในแบบที่ไม่ใช่เส้นตรง...(ค) เท่าที่แล้วมาในบทนี้เราบอกได้อย่างไรว่าตัวแปรทั้งสองอาจมีสัมพันธภาพต่อกันแบบเส้นโค้ง...(ง) และถ้าเราเห็นว่าตัวแปรทั้งสองมีสัมพันธภาพต่อกันแบบเส้นโค้ง เราจะใช้สมการถดถอยแบบเส้นตรงพยากรณ์ได้แม่นยำหรือไม่....(จ)

ตอบ ก) เส้นตรง ข) ใช่ ค) เป็นไปได้ ง) ดูสแกตเตอร์แกรม จ) ไม่

3) ถ้านักทดลองคำนวณสมการถดถอยของคะแนนการจำแนกเสียง (x) กับคะแนนความเข้าใจภาษา (y) ของเด็กนักเรียน ในการทดลองทางจิตวิทยา ได้สมการถดถอยดังนี้

$Y = 50 + 0.4y$; $r = 0.96$; $S_y = 22$

ก) เพื่อทดสอบดูว่าสมการถดถอยนี้จะเชื่อถือได้หรือไม่ในการใช้พยากรณ์ตัวแปรตามนักทดลองเลือกตัวอย่างนักเรียนมาโดยสุ่มอีกชุดหนึ่ง จำนวน 5 คน จงคำนวณค่าพยากรณ์ของ y (y′) จากค่าของ x ดังต่อไปนี้ และเปรียบเทียบ y′ กับค่าสังเกตจริงของตัวแปรตาม

x	5	6	8	12	15
y′					

ถ้าสังเกตในการสอบความเข้าใจภาษา (y) จริง ๆ เป็นดังนี้

y	23	35	41	58	75

ข. จงคำนวณหาความคลาดเคลื่อนมาตรฐานของค่าประมาณของข้อมูลชุดนี้ (ทำโดยคำนวณค่า r ระหว่างค่าสังเกตของ x ในข้อ ก) และ y ในข้อนี้ แล้วคำนวณหาค่า S_y แล้วแทนลงในสูตร) เปรียบเทียบว่าสมการถดถอยที่จะคำนวณได้จากข้อมูล ชุดใหม่นี้ มีความคลาดเคลื่อนมาตรฐานมากหรือน้อยกว่าความคลาดเคลื่อนมาตรฐานของสมการถดถอยจากข้อมูลชุดก่อน

!5 สหสัมพันธ์ระหว่างตัวแปรแบบต่าง ๆ

วิธีการหาสหสัมพันธ์ระหว่างตัวแปรนั้นมีอีกหลายแบบ ขึ้นอยู่กับประเภทของตัวแปรหรือมาตรฐานการวัดที่ใช้กับตัวแปรนั้น ในบทนี้เราจะศึกษาวิธีการที่สำคัญแต่ง่ายอีกเพียงสองวิธี สำหรับใช้กับตัวแปรประเภทที่ต่างไปจากประเภทที่แล้วมา แต่ก่อนอื่นเราต้องทำความรู้จักกับมาตรการวัด (scale) แบบต่าง ๆ ที่ใช้บอกประเภทของตัวแปร

1) มาตรการวัดแบบจำแนกประเภท (Nominal Scale) มาตราวัดแบบนี้จำแนกข้อมูลออกเป็นประเภทหรือเป็นกลุ่ม เช่นตัวแปร x อาจใช้แทนเพศ ซึ่งจำแนกออกเป็นชายกับหญิง ตัวแปร y อาจใช้แสดงระดับการศึกษา ซึ่งจำแนกออกเป็น ประถม, มัธยม และอุดมศึกษา ค่าของตัวแปร x และตัวแปร y มักอยู่ในรูปตารางจำแนกความถี่ เช่น ถ้าเราจำแนกพนักงานในบริษัทหนึ่งออกตามลักษณะของตัวแปรทั้งสองนี้ เราอาจได้ตารางดังนี้

x \ y	ประถม	มัธยม	อุดมศึกษา	รวม
ชาย	2	7	3	12
หญิง	1	5	2	8
รวม	3	12	5	20

2) มาตรการวัดแบบลำดับ (Ordinal Scale) มาตรวัดแบบนี้ค่าของตัวแปรคือลำดับที่ของค่าสังเกตในข้อมูล เช่นในการประกวดนางงาม ถ้ามีกรรมการ 2 คน ษ กับ ฒ ตัดสินให้ลำดับที่ผู้เข้าประกวดจากที่หนึ่งไปถึงที่สุดท้าย ตามความเห็นของแต่ละคน เราอาจให้ตัวแปร x แทนลำดับที่ตามกรรมการ ษ และตัวแปร y เป็นลำดับที่ตามกรรมการ ฒ และได้ค่าสังเกตของตัวแปรทั้งสองนี้

ผู้เข้าประกวด	ก	ข	ค	ง	จ
กรรมการ ษ (x)	2	1	3	5	4
กรรมการ ฒ (y)	1	4	2	3	5

จะเห็นได้ว่าลำดับที่อาจได้มาจากคะแนนที่กรรมการให้ผู้เข้าประกวดแต่ละคน แล้วมาเรียงลำดับที่ว่าคนไหนได้ที่เท่าไรตามความเห็นของกรรมการแต่ละคน ไม่ได้หมายความว่าคนที่ได้ที่ 1 ได้คะแนนสองเท่าของคนที่ 2 เพียงแต่บอกว่าได้คะแนนมากกว่าเท่านั้น

3) มาตราการวัดแบบช่วง (Interval Scale) การวัดแบบนี้คือการวัดที่หน่วยวัดมีช่วงเท่า ๆ กัน เช่นหน่วยวัดองศาอุณหภูมิ (ที่แต่ละองศามีช่วงเท่ากัน) หน่วยวัดเวลา หรือคะแนนสอบ เป็นต้น ค่าของตัวแปรที่ใช้การวัดแบบนี้สามารถนำมาบวกลบกันได้

4) มาตราการวัดแบบอัตราส่วน (Ratio Scale) การวัดแบบนี้ก็คือการวัดแบบช่วงที่ค่าของการวัดเริ่มต้นจากศูนย์เป็นต้นไป ทำให้สามารถเปรียบเทียบค่าของการวัด เป็นอัตราส่วนได้ เช่นว่า ผ้าที่ยาว 1 เมตร มีความยาวเป็น 1/2 เท่าของผ้าผืนที่มีความยาว 2 เมตร หรือว่าของที่หนัก 5 กก. มีน้ำหนักเป็น 2.5 เท่าของของที่หนัก 2 กก. เป็นต้น ดังนั้นค่าของตัวแปรใช้บวกลบคูณหารกันได้

(สังเกตเปรียบเทียบกับมาตราการวัดแบบช่วง ถ้าหน่วยวัดไม่ได้เริ่มจาก 0 เราไม่อาจใช้ค่าตัวแปรคูณหารกัน เช่นอุณหภูมิเป็น °C ซึ่งลดลงไปได้ต่ำกว่า 0 เราไม่อาจพูดได้ว่า น้ำร้อน 50°C มีอุณหภูมิเป็น 1/2 เท่าของน้ำที่มีอุณหภูมิ 100°C เพราะอุณหภูมิไม่ได้เริ่มจาก 0)

เมื่อเราเข้าใจถึงมาตราการวัดที่ใช้แบบต่าง ๆ นี้แล้ว จะเห็นได้ว่าในเรื่องของสหสัมพันธ์และการถดถอยที่แล้วมา เราใช้ตัวแปรที่ใช้มาตราวัดแบบช่วงและมาตราวัดแบบอัตราส่วนเท่านั้น ต่อไปนี้เราจะมาสนใจกับวิธีการหาสหสัมพันธ์เมื่อตัวแปรมีการวัดแบบจำแนกประเภทหรือมีการวัดแบบลำดับ จะเห็นได้ด้วยว่าเราอาจแปลงตัวแปรที่มีการวัดแบบช่วงหรือแบบอัตราส่วนให้กลายเป็นตัวแปรแบบลำดับของข้อมูล ทำนองเดียวกันเราอาจจำแนกประเภทข้อมูลตามค่าของตัวแปรแล้วทำให้ได้ตัวแปรแบบจำแนกประเภทขึ้นมา

ก่อนที่จะเริ่มรายละเอียดของวิธีการหาสหสัมพันธ์แบบอื่นนอกเหนือจากที่ได้ทำมาแล้ว อาจจะบรรยายโดยสังเขปถึงวิธีการหาสหสัมพันธ์ที่สำคัญสำหรับตัวแปรทุกแบบได้ดังนี้

1) ถ้าตัวแปรมีการวัดแบบช่วงหรือแบบอัตราส่วน

ก) Simple Correlation สหสัมพันธ์แบบนี้คือแบบที่เราศึกษาไปแล้วเรียกว่าเป็นสหสัมพันธ์อย่างง่ายเพราะมีตัวแปรเพียงสองตัว วิธีพยากรณ์คือ Simple Regression

ข) Multiple Correlation วิธีนี้ใช้หาสหสัมพันธ์ระหว่างตัวแปรตัวหนึ่งกับตัวแปรอีกหลายตัว เพราะที่จริงแล้วเราอาจมีตัวแปรหลายตัวที่มีสหสัมพันธ์ร่วมกัน เช่น เราอาจต้องการศึกษาสัมพันธภาพระหว่างความสูงของเด็ก (x) น้ำหนักตัว (y) และอายุของเด็ก (z) ซึ่งเราต้องการทราบว่าตัวแปรทั้ง 3 นี้มีสหสัมพันธ์ต่อกันอย่างไร เพื่อดูอิทธิพลของตัวแปรตัวหนึ่งต่อตัวแปรตัวอื่น ๆ (คือเราอาจต้องการคำนวณสหสัมพันธ์ร่วมกันของตัวแปรทั้งสาม)

ค) Partial Correlation วิธีนี้ใช้เมื่อมีสหสัมพันธ์ระหว่างตัวแปรหลายตัว แต่เราต้องการพิจารณาสหสัมพันธ์ ระหว่างตัวแปรสองตัวใด ๆ เมื่อกำหนดให้ตัวแปรตัวอื่น ๆ มีค่าคงที่ เช่น ถ้าเรารู้ว่าอายุของเด็กมีอิทธิพลต่อความสูงและน้ำหนักตัวของเด็ก เราอาจต้องการรู้สหสัมพันธ์บางส่วนของความสูง (x) กับ น้ำหนักตัว (y) สำหรับเด็กที่มีอายุเท่ากัน นั่นคือกำหนดให้อายุ (z) คงที่

วิธีพยากรณ์ค่าตัวแปรตามเมื่อมีตัวแปรอิสระหลายตัวนี้คือ Multiple Regression เช่น เราอาจใช้ทั้งน้ำหนักตัวและอายุประกอบกันเพื่อพยากรณ์ความสูง วิธีการวิเคราะห์ ความถดถอยเชิงซ้อนนี้มีที่ใช้อย่างมากในงานวิจัยทางสังคมศาสตร์ทุกสาขา แต่เราจะยังไม่ศึกษาในขั้นนี้

2) ถ้าตัวแปรมีการวัดแบบลำดับ วิธีที่สำคัญที่สุดคือการหาสหสัมพันธ์ลำดับ (Rank Correlation) ซึ่งเราจะได้ศึกษารายละเอียดต่อไป

3) ถ้าตัวแปรมีการวัดแบบจำแนกประเภท ถ้าตัวแปรมีการจำแนกประเภทได้เพียงสองอย่างทั้งสองตัว เราอาจใช้สัมประสิทธิ์ ϕ (ไฟ) เป็นเครื่องวัดสหสัมพันธ์ระหว่างตัวแปรทั้งสอง ถ้าตัวแปรหนึ่งในสองตัวหรือทั้งสองตัวสามารถจำแนกได้มากกว่าสองอย่าง เราอาจใช้ Contingency Coefficient เพื่อวัดสหสัมพันธ์ระหว่างตัวแปรทั้งสอง

4) ถ้าตัวแปรสองตัวมีการวัดต่างประเภทกัน

ก) Biserial Correlation ใช้วัดสหสัมพันธ์เมื่อเราจำแนกตัวแปรหนึ่งที่เป็นการวัดแบบช่วงออกเป็นสองประเภท (เช่นจำแนกคะแนนสอบนักเรียนออกเป็นสองพวกคือสอบได้กับสอบตก) แล้วต้องการวัดสหสัมพันธ์กับตัวแปรตัวที่สองที่มีการวัดแบบช่วง (เช่นคะแนนปลายปี)

ข) Point-biserial Correlation ใช้วัดสหสัมพันธ์เมื่อตัวแปรตัวหนึ่ง มีการวัดแบบจำแนกประเภทแน่นอนโดยมีได้สองอย่างเท่านั้น (เช่นเป็นผู้หญิง หรือผู้ชาย) และอีกตัวแปรหนึ่งมีการวัดแบบช่วง (เช่นความสูง)

แบบฝึกหัดที่ 5.9

1) จงอธิบายถึงประเภทของตัวแปรที่ใช้สัมประสิทธิ์สหสัมพันธ์ต่อไปนี้ วัดสัมพันธภาพระหว่างตัวแปรสองตัวหรือมากกว่า

ก) Point-biserial Correlation Coefficient

ข) Phi Coefficient

ค) Biserial Correlation Coefficient

ง) Contingency Coefficient

จ) Rank Correlation Coefficient

ฉ) Product Moment Correlation Coefficient

ช) Multiple Correlation Coefficient

5.16 สหสัมพันธ์แบบลำดับ (Rank Correlation)

สัมประสิทธิ์สหสัมพันธ์แบบลำดับนี้เรียกว่า Spearman Rank Correlation Coefficient (ตามชื่อผู้คิดวิธี) วิธีการคำนวณหาสัมประสิทธิ์สหสัมพันธ์ตัวนี้ (ซึ่งใช้สัญลักษณ์ r_s สำหรับตัวอย่างและสัญลักษณ์ p_s สำหรับค่าในประชากร) ยิ่งง่ายกว่าการคำนวณค่า r ที่แล้วมาด้วยซ้ำ) ข้อมูลที่เราต้องการหาสหสัมพันธ์แบบลำดับระหว่างตัวแปรที่มีการวัดแบบลำดับ ก็เช่นว่าเราอาจต้องการวัดดูว่าความเห็นของกรรมการสองคนในการตัดสินให้ตำแหน่งนางงามในตัวอย่างที่ผ่านมา มีสหสัมพันธ์ต่อกัน หรือว่ามีความสอดคล้องกันหรือไม่เพียงใด

สูตรในการคำนวณหาสัมประสิทธิ์สหสัมพันธ์แบบลำดับคือ

$$r_s = 1 - \frac{6(\Sigma d^2)}{n\,(n^2-1)}$$

เมื่อ d = ผลต่างของลำดับที่ระหว่างคู่ค่าสังเกต

n = จำนวนคู่ค่าสังเกตในตัวอย่าง

ค่าของสัมประสิทธิ์สหสัมพันธ์ลำดับ r_s จะอยู่ระหว่าง -1 กับ $+1$ เช่นเดียวกับค่าของ r และมีความหมายเช่นเดียวกับค่า r ที่ได้บรรยายไปแล้ว

5.16.1 วิธีการคำนวณค่า

ขั้นที่ 1 ถ้าข้อมูลอยู่ในรูปที่ตัวแปรทั้งสองเป็นแบบลำดับแล้ว เริ่มโดยการหาผลต่างระหว่างคู่ลำดับ (เช่นถ้าให้ R_x แสดงลำดับของตัวแปร x และ R_y แสดงลำดับของตัวแปร y เราจะได้ว่า ค่า $d = R_x - R_y$)

ขั้นที่ 2 ยกกำลังสองของผลต่าง d แต่ละอัน

ขั้นที่ 3 หาผลบวกของ d^2 ทั้งหมด (Σd^2)

ขั้นที่ 4 แทนค่าของ Σd^2 และ n ลงในสูตร

ตัวอย่าง 5.14 จากตัวอย่างข้างต้นเรื่องการประกวดนางงาม ถ้าเราให้ R_x แทนลำดับที่นางงามตามความเห็นของกรรมการ ษ R_y แทนลำดับที่ตามความเห็นของกรรมการ ฒ เราจะได้ตารางการคำนวณสหสัมพันธ์ลำดับดังนี้

ผู้เข้าประกวด	ก	ข	ค	ง	จ	
กรรมการ ษ (R_x)	2	1	3	5	4	
กรรมการ ฒ (R_y)	1	4	2	3	5	
d	1	-3	1	2	-1	
d^2	1	9	1	4	1	$\Sigma d^2 = 16$

เราได้ค่า $\Sigma d^2 = 16$ และ $n = 5$

แทนค่าลงในสูตร

$$r_s = 1 - \frac{6x\,(16)}{5\,(25-1)} = 1 - \frac{96}{120}$$

$$= 1 - 0.80 = 0.20$$

สหสัมพันธ์ระหว่างค่าลำดับนางงามในความเห็นของกรรมการ ษ และกรรมการ ฒ มีอยู่ค่อนข้างต่ำ แสดงว่าความเห็นของกรรมการทั้งสองนี้ไม่สอดคล้องกันเท่าไร

5.16.2 ข้อได้เปรียบของสหสัมพันธ์แบบลำดับ

เนื่องจากสหสัมพันธ์ลำดับเป็นการคำนวณจากลำดับที่ของตัวแปร แม้ว่าตัวแปรนั้นจะมีการวัดแบบช่วงก็ตาม เราไม่ต้องอาศัยข้อสันนิษฐานว่าตัวแปรทั้งสองจะต้องมีสัมพันธภาพต่อกันเป็นแบบเส้นตรง ดังนั้นถ้าเราไม่แน่ใจว่าตัวแปรที่จะหาสหสัมพันธ์มีสัมพันธภาพต่อกันแบบใด เราก็อาจแปลงข้อมูลเดิมของตัวแปรทั้งสองให้อยู่ในรูปลำดับที่ของตัวแปร และคำนวณ

หาสัมพันธ์แบบลำดับแทนที่จะใช้สหสัมพันธ์แบบ Product Moment (r) ซึ่งวัดเฉพาะสหสัมพันธ์ที่เป็นแบบเส้นตรง

ข้อสังเกตเมื่อใช้สหสัมพันธ์ลำดับก็คือว่า ถ้าเราแปลงตัวแปรแบบช่วงเป็นตัวแปรลำดับแล้วต้องการหาสมการถดถอยของข้อมูลเดิม สหสัมพันธ์แบบลำดับที่คำนวณได้จะไม่บอกว่าสมการถดถอยจะใช้พยากรณ์ได้ดีหรือไม่ เพราะสหสัมพันธ์แบบลำดับไม่ได้วัดสัมพันธภาพแบบเส้นตรง ซึ่งเป็นข้อสันนิษฐานสำหรับการพยากรณ์ด้วยเส้นถดถอยหรือสมการถดถอย

ตัวอย่าง 5.15 จากตัวอย่างที่แล้วมาเกี่ยวกับคะแนนสอบวิชา ก. และคะแนนสอบวิชา ข. เราอาจคำนวณสหสัมพันธ์แบบลำดับระหว่างคะแนนทั้งสองวิชา โดยหาลำดับที่ของคะแนนทั้งเรียงจากมาไปน้อยหรือจากน้อยไปมากก็ได้ แต่ต้องเรียงเหมือนกันทั้งสองตัวแปร

5.16.3 วิธีการเรียงลำดับ

ถ้าข้อมูลยังไม่อยู่ในรูปตัวแปรแบบลำดับ เราต้องแปลงให้อยู่ในรูปลำดับเสียก่อน โดยเรียงลำดับทั้งสองตัวแปร จากมากไปน้อยหรือจากน้อยไปมากก็ได้ ถ้าค่าตัวแปรเดิมมีซ้ำกันทำให้ได้ลำดับที่เท่ากัน ให้หาค่าเฉลี่ยของลำดับที่ แล้วให้ลำดับเฉลี่ยนั้นกับตัวแปรที่เดิมได้ลำดับที่เท่ากัน เช่นถ้าในชั้นเรียนมีนักเรียนสอบได้คะแนนเป็นที่ 3 เท่ากันถึง 4 คน เราก็หาลำดับที่เฉลี่ยโดยบวกลำดับที่ที่ 4 คนนี้จะได้คะแนนเรียงกัน แล้วหารด้วยจำนวนคนที่ได้คะแนนเท่ากัน นั่นคือจะได้ลำดับที่เฉลี่ยสำหรับนักเรียนทั้ง 4 คนนี้เท่ากับ $(3+4+5+6)/4 = 18/4 = 4.5$ หรือถ้ามีนักเรียนสอบได้ที่ 5 เท่ากัน 3 คน ก็จะได้ลำดับที่เฉลี่ยของทั้งสามคนเท่ากับ $(5+6+7)/3 = 18/3 = 6$

คนที่	คะแนนวิชา ก. (x)	คะแนนวิชา ข. (y)	ลำดับที่ของ x (R_x)	ลำดับที่ของ y (R_y)	d	d^2
1	1	2	1	1	0	0
2	2	4	2	2	0	0
3	4	5	3	3	0	0
4	5	7	4.5	4	0.5	0.25
5	5	8	4.5	5	-0.5	0.25

$\Sigma d^2 = 0.50$

แทนค่า Σd^2 และ $n = 5$ ลงในสูตรจะได้

$$r_s = 1 - \frac{6(0.50)}{5(25-1)} = 1 - \frac{3}{120}$$

$$= 1 - 0.025 = 0.975$$

จะเห็นว่าค่าของ r_s แสดงความหมายเหมือนกับค่าสหสัมพันธ์ r ของตัวอย่างนี้ที่เราคำนวณได้ $r = 0.96$ ถึงแม้ค่าของ r_s และ r ที่คำนวณจากข้อมูลชุดเดียวกันจะไม่เท่ากัน แต่ก็จะให้ความหมายสหสัมพันธ์ได้คล้ายกัน ถ้าสหสัมพันธ์ระหว่างตัวแปรเป็นแบบเส้นตรง (แต่ถ้าสหสัมพันธ์ระหว่างตัวแปรไม่เป็นแบบเส้นตรง เราจะได้ค่า r ที่ต่ำ ขณะที่ค่าของ r_s ยังแสดงว่ามีสหสัมพันธ์ระหว่างตัวแปรทั้งสองอยู่)

ค่าของ r_s ที่คำนวณได้ในตัวอย่างนี้สรุปได้ว่ามีสหสัมพันธ์ระหว่างคะแนนสอบวิชา ก. กับคะแนนสอบวิชา ข. อย่างสูง แสดงว่านักเรียนที่สอบได้คะแนนดีในวิชาหนึ่งก็มักสอบได้คะแนนดีในอีกวิชาหนึ่งด้วย

5.16.4 การทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์แบบลำดับ

ทำได้โดยวิธีการเช่นเดียวกับที่ใช้ทดสอบสัมประสิทธิ์ r ที่แล้วมา โดยการใช้ค่าวิกฤตของ r_s (ดูตารางที่ 11 ในภาคผนวก) หรือถ้าไม่มีตารางค่าวิกฤตของ r_s ก็อาจใช้ตัวสถิติ t และเปรียบเทียบกับตารางค่าวิกฤตของ t เช่นที่แล้วมา

ตัวอย่าง 5.16 ทดสอบนัยสำคัญของ r_s โดยใช้ ค่าวิกฤตของ r_s

จากตัวอย่างที่คำนวณหาค่า r_s ได้ข้างต้น ถ้าเราต้องการทดสอบว่าค่า r_s ที่ได้มีนัยสำคัญหรือไม่ นั่นคือต้องการทดสอบสมมุติฐานว่าสัมประสิทธิ์สหสัมพันธ์แบบลำดับในประชากร P_s เป็นศูนย์หรือไม่ ถ้ากำหนดให้ทดสอบที่ระดับนัยสำคัญ 5% เราจะดำเนินการทดสอบแบบเดียวกันกับที่แล้ว ๆ มาได้ดังนี้

ขั้นที่ 1 $H_0 : \rho_s = 0$ (สหสัมพันธ์แบบลำดับในประชากรเป็นศูนย์)

$H_1 : \rho_s \neq 0$

ขั้นที่ 2 ระดับนัยสำคัญ $\alpha = 0.05$

ขั้นที่ 3 ค่าวิกฤต r_s (จากตารางค่าวิกฤต เมื่อ $\alpha = 0.05$ และ $n = 5$) คือ 0.900

ขั้นที่ 4 ค่าสัมประสิทธิ์สหสัมพันธ์ลำดับจากตัวอย่าง คำนวณได้ $r_s = 0.975$

ขั้นที่ 5 เพราะว่าค่า $|r_s|$ (r_s ที่ไม่คิดเครื่องหมาย) มากกว่าค่าวิกฤตจากตาราง ($0.975 > 0.900$)

ดังนั้น เราจะปฏิเสธสมมุติฐานหลัก (H_0) ที่ว่าไม่มีสหสัมพันธ์ในประชากรของตัวแปร นั่นคือ เราจะกล่าวได้ว่าค่าสหสัมพันธ์แบบลำดับที่คำนวณได้มีนัยสำคัญที่ระดับนัย

สำคัญ 5% (เปรียบเทียบกับผลการทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์ r ที่แล้วมาว่าได้ผลเหมือนกัน)

ตัวอย่าง 5.17 ทดสอบนัยสำคัญของ r_s โดยใช้ตัวสถิติ t

จากตัวอย่าง 5.15 ถ้าเราไม่ตารางค่าวิกฤตของ r_s เราอาจทดสอบนัยสำคัญของ r_s โดยตัวสถิติ t เปรียบเทียบค่าวิกฤตของ t สูตรที่ใช้เหมือนกับสูตรที่ใช้ทดสอบ r คือ

$$t = \frac{r_s \sqrt{(n-2)}}{\sqrt{1-r^2_s}}$$

ขั้นที่ 1 $H_0 : \rho_s = 0$

$H_1 : \rho_s \neq 0$

ขั้นที่ 2 ระดับนัยสำคัญที่กำหนดคือ $\alpha = 0.05$

ขั้นที่ 3 ค่าวิกฤตของ t เมื่อ $\alpha = 0.05$ และ $df = n-2 = 5-2 = 3$

จากตารางค่าวิกฤตของ t คือ 2.35

ขั้นที่ 4 คำนวณค่าสถิติ t จากสูตร

$$t = \frac{0.975\sqrt{(5-2)}}{\sqrt{1-(0.975)^2}} = \frac{1.69}{0.22} = 7.61$$

ขั้นที่ 5 เพราะว่าค่า $|t|$ มากกว่าค่าวิกฤตของ t จากตาราง $(7.61 > 2.35)$

ดังนั้นเราจะปฏิเสธสมมุติฐานหลัก (H_0) ที่ว่าไม่มีสหสัมพันธ์แบบลำดับในประชากร นั่นคือ เราจะกล่าวได้ว่าค่าสหสัมพันธ์แบบลำดับที่คำนวณได้มีนัยสำคัญที่ระดับสำคัญ 5%

แบบฝึกหัดที่ 5.10

1) กรรมการตัดสินการประกวดภาพยนตร์สองคน ให้ลำดับที่ของภาพยนตร์ที่ตนชอบมากที่สุดไปถึงน้อยที่สุด ดังปรากฏในตารางต่อไปนี้ จงคำนวณหาสัมประสิทธิ์สหสัมพันธ์แบบลำดับของข้อมูลชุดนี้ และทดสอบนัยสำคัญของสัมประสิทธิ์ r_s ที่ได้ที่ระดับนัยสำคัญ 5% จะกล่าวได้ไหมว่ากรรมการสองคนนี้มีความเห็นไม่สอดคล้องกันเลย (ตอบ $r_s = 0.13$)

ภาพยนต์เรื่อง	A	B	C	D	E	F	G	H	I
กรรมการ ก.	1	3	2	6	4	5	9	8	7
กรรมการ ข.	6	5	2	1	8	3	4	7	9

2) จงให้ลำดับตัวแปรทั้งสองในข้อมูลต่อไปนี้ เมื่อ x คือคะแนนสอบวิชาสังคมศึกษา และ y คือคะแนนสอบวิชาวิทยาศาสตร์ทั่วไป ของนักเรียนชั้นประถมปลายจำนวน 10 คน แล้วคำนวณหาสัมประสิทธิ์สหสัมพันธ์อันดับของคะแนนสอบสองวิชานี้ พร้อมกับทดสอบนัยสำคัญของ r_s ที่ระดับนัยสำคัญ 5% จะกล่าวได้อย่างไรเกี่ยวกับสหสัมพันธ์ของคะแนนสอบสองวิชานี้ ($r_s = 0.34$)

นักเรียนคนที่	x	y
1	8	4
2	13	14
3	13	6
4	18	13
5	14	8
6	19	12
7	8	10
8	4	7
9	17	6
10	15	7

5.17 สหสัมพันธ์ของตัวแปรแบบจำแนกประเภท
(Correlation of Nominal Scale Variables)

ในชั้นนี้เราจะพิจารณาถึงวิธีวัดสหสัมพันธ์ระหว่างตัวแปรที่มีการวัดแบบจำแนกประเภทเฉพาะกรณีที่ง่ายที่สุดคือเมื่อ ตัวแปรแบบจำแนกประเภททั้งสองตัวจำนวนออกได้เพียงสองอย่างเท่านั้น ตารางข้อมูลจะเป็นแบบที่มี 4 ช่อง ตัวสถิติที่ใช้วัดสหสัมพันธ์คือสัมประสิทธิ์ ϕ (Phi Coefficient) (ตัว ϕ ออกเสียงว่า "ไฟ")

ถ้าตัวแปร x มีค่าได้เพียงสองอย่างคือ 0 กับ 1 (ในเมื่อ 0 อาจแทนเพศหญิง, 1 แทนเพศชาย ถ้า x เป็นตัวแปรแสดงเพศ) และตัวแปร y มีค่าได้เพียงสองอย่างคือ 0 กับ 1 (ในเมื่อ 0 อาจแทนการไม่สูบบุหรี่, 1 แทนการสูบบุหรี่ ถ้า Y เป็นตัวแปรแสดงบุคลิกที่เกี่ยวกับการสูบบุหรี่) ตารางข้อมูลที่เราได้จะอยู่ในลักษณะดังนี้

x \ y	0	1	รวม
1	a	b	a + b
0	c	d	c + d
รวม	a + c	b + d	N

ถ้าเราทำการสัมภาษณ์ตัวอย่างนักศึกษาในมหาวิทยาลัยทั้งชายและหญิงรวม 150 คน ถามว่าสูบบุหรี่หรือไม่ แล้วมาจำแนกข้อมูลที่ได้ตามตัวแปรแสดงเพศและบุคลิกการสูบบุหรี่ เราอาจได้ตารางดังนี้

x \ y	ไม่สูบ (0)	สูบ (1)	รวม
ชาย (1)	15	75	90
หญิง (0)	55	5	60
รวม	70	80	150

ถ้าสิ่งที่เราสนใจจะวัดคือสัมพันธภาพระหว่างเพศกับการสูบบุหรี่ นั่นคือ เราต้องการศึกษาว่า ถ้าเป็นผู้ชายมักสูบบุหรี่ ถ้าเป็นผู้หญิงมักจะไม่สูบบุหรี่ เป็นคำกล่าวที่ใช้ได้หรือไม่กับนักศึกษามหาวิทยาลัย เราอาจจะใช้สัมประสิทธิ์ ϕ (ไฟ) เป็นเครื่องวัดสัมพันธภาพที่ว่านี้

สัมประสิทธิ์ ϕ คำนวณได้ตามสูตรต่อไปนี้ (เมื่อ a, b, c, d คือค่าในช่องต่าง ๆ จากตารางข้างต้น)

$$\phi = \frac{bc - ad}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

สัมประสิทธิ์ ϕ นี้บางตัวก็เรียกว่า Fourfold Point Correlation Coefficient หรือ Tetrachoric Correlation Coefficient

ค่าของสัมประสิทธิ์ ϕ จะอยู่ระหว่าง -1 กับ $+1$ โดยที่ ϕ จะมีค่าเป็นศูนย์ ถ้าตัวแปรจำแนกประเภททั้งสองไม่มีสัมพันธภาพต่อกันเลย แต่เครื่องหมายของ ϕ **นั้นไม่ได้ใช้ใน**การตีความหมายเท่าไรนักเพราะเครื่องหมายของ ϕ เปลี่ยนแปลงได้เพียงแต่สลับค่า**ในช่องที่เป็น** 0 กับช่องที่เป็น 1 ถึงแม้ข้อมูลจะไม่เปลี่ยนแปลงก็ตาม ดังนั้นการสรุปว่าสหสัม**พันธ์ระหว่าง**ตัวแปรจำแนกประเภทในตารางสี่ช่องนั้นอยู่ในลักษณะใด เราพูดได้เพียงว่าตัวแปรสองตัวนี้มีสัมพันธภาพต่อกันหรือเป็นอิสระต่อกันเท่านั้น เราไม่อาจพูดถึงสหสัมพันธ์ทางบวกหรือทางลบเช่นที่แล้วมาได้

ตัวอย่างที่ 5.18 จากตัวอย่างการสัมภาษณ์นักศึกษาข้างต้น ถ้าเราต้องการวัดสหสัมพันธ์ระหว่างเพศกับการสูบบุหรี่ จะคำนวณค่าสัมประสิทธิ์ ϕ ได้ดังนี้

$$\phi = \frac{75(55) - 15(5)}{\sqrt{(90)(60)(70)(80)}} = \frac{25(165-3)}{25\,(\sqrt{48384})}$$

$$= \frac{162}{219.96} = 0.74$$

ค่าสัมประสิทธิ์ ϕ ที่ได้แสดงว่ามีสหสัมพันธ์อยู่ค่อนข้างสูงระหว่างเพศกับการสูบบุหรี่ โดยการพิจารณาตารางแสดงเพศกับบุคลิกการสูบบุหรี่ เราอาจกล่าวว่าคำกล่าวที่เอ่ยถึงข้างต้นเป็นจริงหรือไม่ แต่ก่อนอื่นเราควรทดสอบนัยสำคัญของสัมประสิทธิ์ ϕ ที่ได้เสียก่อนเพื่อให้แน่ใจว่าค่าที่ได้ไม่ได้เกิดขึ้นโดยบังเอิญทั้งที่ในประชากรแล้วตัวแปรทั้งสองอาจไม่มีสหสัมพันธ์แต่เป็นอิสระต่อกัน

15.7.1 การทดสอบนัยสำคัญของสัมประสิทธิ์ ϕ

การทดสอบนัยสำคัญของสัมประสิทธิ์ ϕ คือการทดสอบว่าค่า ϕ ที่ได้แตกต่างจากศูนย์มากพอที่จะทำให้เชื่อถือได้ว่ามีสหสัมพันธ์อยู่จริงในประชากรของตัวแปรทั้งสอง และค่า ϕ ที่ได้ ไม่ได้เป็นไปโดยบังเอิญ (ระดับนัยสำคัญของการทดสอบ คือความน่าจะเป็นที่ใช้ตัดสินว่าค่าที่คำนวณได้จะเกิดขึ้นโดยบังเอิญหรือไม่)

การทดสอบทำได้ตามขั้นตอนเช่นที่แล้ว ๆ มา โดยอาศัยตัวสถิติ Z ตามสูตร

$$Z = \frac{|a-d| - 1}{\sqrt{(a+d)}} \quad (|\;| \text{ คือไม่คิดเครื่องหมาย})$$

ค่าวิกฤตที่ใช้เปรียบเทียบ ได้จากตารางพื้นที่ใต้โค้งปกติ ที่ระดับนัยสำคัญที่กำหนด

ตัวอย่าง 5.18 จากตัวอย่างที่แล้วมา ค่าสัมประสิทธิ์ ϕ ที่ได้ มีนัยสำคัญพอที่จะทำให้เรากล่าวได้ว่า เพศกับการสูบบุหรี่มีสหสัมพันธ์กัน ตามคำกล่าวที่ว่าข้างต้น หรือ ให้ทดสอบที่ระดับนัยสำคัญ 5%

ขั้นที่ 1 $H_o : \Phi = 0$ (สัมประสิทธิ์ Φ สำหรับประชากรเป็นศูนย์)

$H_1 : \Phi \neq 0$

ขั้นที่ 2 ระดับนัยสำคัญที่กำหนด $\alpha = 0.05$

ขั้นที่ 3 ค่าวิกฤต Z ที่ $\alpha = 0.05$ คือ 1.96

ขั้นที่ 4 คำนวณค่าสถิติทดสอบ $Z = \frac{|15-5| - 1}{\sqrt{(15+5)}} = \frac{9}{20} = \frac{9}{4.47} = 2.01$

ขั้นที่ 5 เพราะว่าค่า Z ที่คำนวณได้มากกว่า ค่าวิกฤต $(2.01 > 1.96)$

ดังนั้นเราจะปฏิเสธสมมติฐานหลัก (H_o) ที่ว่าไม่มีสหสัมพันธ์ในประชากร

นั่นคือ เราจะสรุปได้ว่าค่าสัมประสิทธิ์ ϕ ที่คำนวณได้มีนัยสำคัญที่ระดับนัยสำคัญ 5% ทำให้เรากล่าวได้ตามข้อมูลที่ได้ในตัวอย่างชุดนี้ว่า เพศกับการสูบบุหรี่มีสหสัมพันธ์กันในลักษณะที่ผู้ชายมักสูบบุหรี่ และผู้หญิงมักไม่สูบบุหรี่ ในกรณีของนักศึกษามหาวิทยาลัยที่เราไปหาตัวอย่างสัมภาษณ์มา

แบบฝึกหัดที่ 5.11

1) ในการสัมภาษณ์นักศึกษามหาวิทยาลัย จำนวน 100 คน มีคำถามเกี่ยวกับความเห็นในการอนุญาตให้มีการทำแท้งโดยถูกกฎหมาย ตามคำเรียกร้องของผู้ตั้งครรภ์ โดยไม่ต้องขอความเห็นจากแพทย์ว่ามีความจำเป็นทางการแพทย์ในการทำแท้งหรือไม่ จำแนกคำตอบได้ดังนี้

เพศผู้ตอบ / คำตอบ	หญิง	ชาย
เห็นด้วย	10	30
ไม่เห็นด้วย	40	20

จงคำนวณหาสัมประสิทธิ์ ϕ จากข้อมูลชุดนี้ และทดสอบนัยสำคัญของค่าสัมประ-ประสิทธิ์ ϕ ที่ได้ที่ระดับนัยสำคัญ 5% จะกล่าวได้ไหมว่าเพศของผู้ตอบมีอิทธิพลกับคำตอบที่ได้ (นั่นคือว่ามีสหสัมพันธ์ระหว่างเพศผู้ตอบกับความเห็นเกี่ยวกับการทำแท้ง)

(ตอบ $\phi = 0.41$)

$Z = 1.64$)

15.8 ทบทวนสิ่งที่ทำไปในบทนี้

ถ้านักศึกษาเข้าใจบทนี้ดีแล้ว นักศึกษาจะสามารถทำสิ่งต่อไปนี้

1) เขียนสแกตเตอร์แกรมแสดงสัมพันธภาพของตัวแปรสองตัว
2) คำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ r
3) ทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์ r
4) อธิบายความหมายของสัมประสิทธิ์สหสัมพันธ์ r ที่คำนวณและทดสอบนัยสำคัญแล้ว
5) คำนวณหาสัมประสิทธิ์การถดถอย (a และ b) สำหรับสมการถดถอย
6) สร้างเส้นถดถอยบนสแกตเตอร์แกรมเพื่อแสดงสัมพันธภาพแบบเส้นตรงระหว่างตัวแปร
7) พยากรณ์ค่าของตัวแปรตามด้วยสมการถดถอยเมื่อกำหนดค่าของตัวแปรอิสระ
8) คำนวณหาความคลาดเคลื่อนมาตรฐานสำหรับสมการถดถอย
9) อธิบายความหมายของมาตรการวัดแบบต่าง ๆ ที่ใช้กับตัวแปรทางสถิติ
10) อธิบายวิธีการที่สำคัญ ๆ ในการวัดสหสัมพันธ์ระหว่างตัวแปรที่ใช้การวัดแบบต่าง ๆ
11) คำนวณหาสัมประสิทธิ์สหสัมพันธ์แบบลำดับ r_s
12) ทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์ r_s
13) อธิบายความหมายของ r_s ที่คำนวณได้และทดสอบนัยสำคัญแล้ว
14) คำนวณหาสัมประสิทธิ์ ϕ
15) ทดสอบนัยสำคัญของสัมประสิทธิ์ ϕ
16) อธิบายความหมายของค่า ϕ ที่คำนวณได้และทดสอบนัยสำคัญแล้ว

แบบฝึกหัดที่ 5.12

1) จะเขียนสแกตเตอร์แกรมของข้อมูลต่อไปนี้ เมื่อ x คือคะแนนสอบวิชา ก. และ y คือคะแนนสอบวิชา ข.

x	1	2	3	5	7	8	10	11
y	1	2	4	5	4	3	1	1

2) จากข้อ 1) จงอธิบายว่าสัมพันธภาพระหว่างตัวแปร x กับตัวแปร y เป็นแบบใด และสัมประสิทธิ์แบบเส้นตรงจะเป็นเครื่องวัดระดับของสัมพันธภาพระหว่างตัวแปรทั้งสองที่ดีหรือไม่ในกรณีนี้

3) จงคำนวณหาสัมประสิทธิ์สหสัมพันธ์ r จากข้อมูลต่อไปนี้ และอธิบายว่าตัวแปรทั้งสองมีสหสัมพันธ์กันแบบใด (จะใช้วิธีคำนวณโดยตรงหรือจะคำนวณแบบง่ายก็ได้)

x	2	3	3	4	5	5	6
y	7	6	5	4	3	2	1

4) จากข้อ 3) จงทดสอบว่าค่า r ที่คำนวณได้มีนัยสำคัญหรือไม่ที่ระดับนัยสำคัญ 1% และจะสรุปเกี่ยวกับค่า r ที่ได้ว่าอย่างไร (r = -0.98)

5) จากข้อ 3) จงคำนวณหาสัมประสิทธิ์การถดถอย เขียนสมการถดถอย และสร้างเส้นถดถอยในสแกตเตอร์แกรม ดูว่าเส้นถดถอยเข้ากับกลุ่มจุดบนสแกตเตอร์แกรมได้ดีหรือไม่ ถ้ากำหนด x = 5 จะพยากรณ์ค่า y ได้เท่าไร คำนวณค่าความคลาดเคลื่อนมาตรฐานของสมการถดถอยนี้มาด้วย

$$(a = 10,\ b = 1.50,\ s_{xy} = 0.41,\ \hat{y} = 2.50)$$

6) เด็กนักเรียนอายุแปดขวบ จำนวน 10 คน ทำการทดสอบทางจิตวิทยาเพื่อวัดคะแนนความคิดริเริ่ม กับคะแนนเชาว์ปัญญา (IQ.) ผลการทดสอบได้ข้อมูลดังนี้

เด็กคนที่	1	2	3	4	5	6	7	8	9	10
(y) คะแนนความคิดริเริ่ม	15	9	6	12	13	8	2	10	15	5
(x) คะแนนเชาวน์ปัญญา	95	120	138	116	106	115	127	110	100	132

ก) จงเขียนสแกตเตอร์แกรมของข้อมูลชุดนี้

ข) คำนวณหาสัมประสิทธิ์สหสัมพันธ์ r ตัวแปรทั้งสองมีสหสัมพันธ์กันแบบใด

($r = -0.88$)

ค) ทดสอบนัยสำคัญของค่า r ที่ได้ที่ระดับนัยสำคัญ 5% มีสหสัมพันธ์เส้นตรงจริงหรือไม่

ง) คำนวณหาสัมประสิทธิ์สหสัมพันธ์แบบลำดับ r_s เปรียบเทียบกับค่า r ในข้อ ข) ว่าได้ค่าใกล้เคียงกันหรือไม่ ($r_s = -0.88$)

จ) ทดสอบนัยสำคัญของค่า r_s ที่ได้ที่ระดับนัยสำคัญ 5% อธิบายผลการทดสอบมาด้วย

7) จากข้อ 6) เกี่ยวกับสมการถดถอยและการพยากรณ์

ก) ถ้าให้ x แทนคะแนนเชาว์ปัญญา และ y แทนคะแนนความคิดริเริ่ม จงคำนวณสัมประสิทธิ์การถดถอยสำหรับสมการถดถอยของข้อมูลชุดนี้

($a = 41.58$, $b = -0.28$)

ข) ถ้ากลับกันให้ x แทนคะแนนความคิดริเริ่ม และ y แทนคะแนนเชาว์ปัญญา จงคำนวณหาสัมประสิทธิ์การถดถอยสำหรับสมการถดถอยของข้อมูลชุดนี้

($a = 142.28$, $b = -2.78$)

ค) ถ้าได้นักเรียนอายุแปดขวบมาอีกกลุ่มหนึ่ง ทำสอบได้คะแนนเชาวน์ปัญญาเท่ากับ 122 จะพยากรณ์ว่านักเรียนกลุ่มนี้จะสอบได้คะแนนความคิดริเริ่มเท่าไร (ใช้สมการถดถอยที่มีคะแนนความคิดริเริ่มเป็นตัวแปรตาม) คำนวณความคลาดเคลื่อนมาตรฐานของค่าประมาณมาด้วย

($\hat{y} = 7.81$, $s_{yx} = 2.09$)

ง) ถ้าได้นักเรียนอายุแปดขวบมาอีกกลุ่มหนึ่ง ทำสอบได้คะแนนความคิดริเริ่มเท่ากับ 14 จะพยากรณ์ว่านักเรียนกลุ่มนี้จะสอบได้คะแนนเชาวน์ปัญญาเท่าไร (ใช้สมการถดถอยที่มีคะแนนเชาวน์ปัญญาเป็นตัวแปรตาม) คำนวณความคลาดเคลื่อนมาตรฐานของค่าประมาณมาด้วย ($\hat{y} = 103.40$, $s_{yx} = 6.63$)

8) ถ้าคำนวณพบว่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ -0.42 ระหว่างน้ำหนักตัว (กก.) และเวลาที่ใช้วิ่งระยะทาง 100 เมตรของผู้หญิง 22 คน จะสรุปได้ว่าน้ำหนักตัวกับเวลาที่ใช้วิ่งมีสหสัมพันธ์กันอย่างไร (ทดสอบก่อนว่า r ที่ได้นี้มีนัยสำคัญหรือไม่ที่ระดับ 5%)

9) ถ้าในโรงงานแห่งหนึ่งพบว่าสหสัมพันธ์แบบลำดับของ จำนวนอุบัติเหตุกับจำนวนคนงานในแต่ละแผนกของโรงงาน คำนวณได้เท่ากับ 0.50 โรงงานนี้มี 15 แผนก ค่า r_s ที่ได้นี้มีนัยสำคัญหรือไม่ที่ระดับ 5% และจะสรุปได้อย่างไรเกี่ยวกับสหสัมพันธ์ระหว่างจำนวนอุบัติเหตุกับจำนวนคนงานในแต่ละแผนก

10) นักพยากรณ์ม้าแข่งสองคน ถูกถามให้พยากรณ์ลำดับที่ม้าแข่ง 8 ตัวจะเข้าเส้นชัยลำดับที่เป็นดังนี้

ม้า	ก	ข	ค	ง	จ	ฉ	ช	ซ
นาย ม.	2	7	8	3	5	1	6	4
นาย ล.	1	4	8	5	7	3	5	2

ถามว่า นาย ม. กับนาย ล. มีความเห็นตรงกันมากน้อยเพียงใด (คำนวณหาสัมประสิทธิ์สัมพันธ์แบบลำดับ แล้วตีความหมายของค่า r_s ที่ได้) ($r_s = 0.68$)

11) จากข้อ 10) ถ้าผลการแข่งขันม้าในรอบนั้นปรากฏผลว่า ลำดับที่ม้าเข้าเส้นชัยจริง ๆ เป็นดังนี้

ม้า	ก	ข	ค	ง	จ	ฉ	ช	ซ
เข้าลำดับที่	1	8	6	4	3	2	7	5

ถามว่านาย ม. หรือนาย ล. พยากรณ์ผลการแข่งขันได้ใกล้เคียงความจริงกว่ากัน (คำนวณหา r_s ระหว่างลำดับที่แต่ละคนพยากรณ์ กับลำดับที่ม้าเข้าเส้นชัยจริง ๆ) แล้วเปรียบเทียบค่า r_s ที่ได้ว่าของคนไหนมากกว่ากัน)

(r_s นาย ม. = 0.83, r_s นาย ล. = 0.39)

12) ข้อมูลเกี่ยวกับจำนวนผลไม้ชนิดหนึ่ง (กก.) กับราคาผลไม้ต่อ กก. ที่ขายในตลาดแห่งหนึ่งในระยะเวลา 9 วัน เป็นดังนี้

วันที่	1	2	3	4	5	6	7	8	9
(y) จำนวนผลไม้ (กก.)	35	27	30	22	15	20	12	25	30
(x) ราคาต่อ กก.	10.60	14.80	10.20	18.80	13.60	18.00	20.00	15.00	12.00

ก) จงเขียนสแกตเตอร์แกรมของข้อมูลชุดนี้

ข) คำนวณหาค่า r และทดสอบนัยสำคัญของ r ที่ได้ สรุปได้ว่าจำนวนผลไม้ที่ขาย กับราคามีสัมพันธภาพต่อกันอย่างไร (กำหนดระดับนัยสำคัญที่ 5%) ($r = -0.76$)

ค) คำนวณหาสัมประสิทธิ์สหสัมพันธ์แบบลำดับของตัวแปรทั้งสอง และทดสอบนัยสำคัญของ r_s ที่ได้ เปรียบเทียบว่าสรุปความเกี่ยวกับสัมพันธภาพระหว่างตัวแปรทั้งสองได้เหมือนกับข้อ ข) หรือไม่ (ทดสอบที่ระดับนัยสำคัญ 5%) ($r_s = -0.78$)

ง. คำนวณหาสัมประสิทธิ์การถดถอยของจำนวนผลไม้ที่ขาย ต่อราคาเป็น กก. (ให้จำนวนผลไม้เป็นตัวแปรตาม) คำนวณความคลาดเคลื่อนมาตรฐานของค่าประมาณด้วย ($a = 47.73$, $b = -1.61$, $s_{xy} = 4.83$)

จ) ถ้าในวันต่อมาราคาผลไม้เป็น 16.00 บาทต่อ กก. จะพยากรณ์ว่าวันนั้นจะขายผลไม้ได้ กี่ กก. ($\hat{=}22.04$)

13) ในช่วง 8 ปีที่ผ่านมา ในอำเภอ ฆ. ปรากฏตัวเลขเกี่ยวกับจำนวนพลเมืองในอำเภอนั้น และจำนวนสุนัขไม่มีเจ้าของ เป็นดังนี้

ปี	พลเมือง	จำนวนสุนัขไม่มีเจ้าของ
2513	1051	2463
2514	1280	2677
2515	1562	2881
2516	1897	3427
2517	2051	3629
2518	2463	3629
2519	2463	3926
2520	3486	4164

ก) จงเขียนสแกตเตอร์แกรมของข้อมูลชุดนี้ ดูว่าสัมพันธภาพระหว่างจำนวนพลเมืองกับจำนวนสุนัขไม่มีเจ้าของในอำเภอนี้เป็นแบบเส้นตรงหรือไม่

ข) คำนวณหาค่า r ทดสอบนัยสำคัญและสรุปเกี่ยวกับสหสัมพันธ์ระหว่างตัวแปรทั้งสอง (กำหนดระดับนัยสำคัญที่ 5%) ($r = 0.94$)

ค) คำนวณหาค่า r_s และทดสอบนัยสำคัญที่ระดับนัยสำคัญ 5% สรุปผลการทดสอบได้เหมือนกับในข้อ ข) หรือไม่ ($r_s = 0.97$)

ง) ถ้าในปี 2521 จำนวนพลเมืองในอำเภอนี้ = 3380 จะพยากรณ์ว่ามีสุนัขไม่มีเจ้าของในอำเภอนี้กี่ตัว คำนวณความคลาดเคลื่อนมาตรฐานของค่าประมาณ และอธิบายความหมายของความคลาดเคลื่อนมาตรฐานของค่าประมาณมาด้วย ($\hat{y} = 4346$, $s_{yx} = 204.24$)

14) บริษัทสุราบางยี่เรือมีข้อมูลค่าใช้จ่ายในการโฆษณา กับรายได้จากการขาย ในช่วงแปดสัปดาห์ที่ผ่านมา เป็นดังนี้

สัปดาห์ที่	ค่าใช้จ่ายโฆษณา(x) (พันบาท)	รายได้จากการขาย(y) (พันบาท)
1	3.2	114.6
2	2.5	86.5
3	2.3	95.5
4	2.4	61.8
5	3.5	154.8
6	3.2	120.9
7	2.6	104.2
8	3.0	136.8

ก) ฝ่ายบริหารของบริษัทต้องการทราบว่า การลงทุนโฆษณาให้ผลในการเพิ่มรายได้จากการขายหรือไม่ (นั่นคือต้องการดูว่าก่อนอื่นค่าใช้จ่ายโฆษณากับรายได้จากการขายมีสหสัมพันธ์ทางบวกต่อกันหรือไม่ เพื่อจะอ้างได้ว่ารายได้ที่เพิ่มขึ้นในแต่ละสัปดาห์อาจเป็นผลมาจากการโฆษณา ถ้าไม่คิดถึงสาเหตุอื่น ๆ ที่อาจทำให้การขายเพิ่มขึ้นได้) จงคำนวณสัมประสิทธิ์สหสัมพันธ์ r ของข้อมูลชุดนี้ ทดสอบนัยสำคัญที่ระดับนัยสำคัญ 5% และสรุปผลที่ฝ่ายบริหารต้องการทราบ (r = 0.86), (3.4 พันล้านบาท)

ข) นอกจากนี้ฝ่ายบริหารต้องการทราบว่าถ้าในสัปดาห์ที่มีค่าใช้จ่ายโฆษณาเท่ากับ 3,400 บาท จะคาดว่าสัปดาห์นั้นจะมีรายได้จากการขายเท่าไร
(ค่าสัมประสิทธิ์การถดถอย เพื่อใช้สมการถดถอยพยากรณ์ค่ารายได้จากการขาย) และต้องการทราบว่าค่าพยากรณ์นี้จะใกล้เคียงกับค่าจริงมากเพียงไร (คำนวณความคลาดเคลื่อนมาตรฐานของค่าประมาณ) ($\hat{y}$ = 141.15, s_{xy} = 15.06
a = -50.82, b = 56.46)

15) ในบริษัทหนึ่ง ผู้จัดการและผู้ช่วยผู้จัดการได้ถูกขอให้เรียงลำดับความสำคัญของเจ้าหน้าที่ของบริษัทจำนวน 12 คน เพื่อใช้ตัดสินในการลดจำนวนพนักงานที่มีความสำคัญน้อยที่สุด ปรากฏว่าได้ผลดังนี้

พนักงาน	ก	ข	ค	ง	จ	ฉ	ช	ซ	ฌ	ญ	ฎ	ฏ
ผู้จัดการ	8	2	7	4	3	1	9	12	6	11	5	10
ผู้ช่วยผู้จัดการ	2	9	5	10	11	12	4	3	7	4	8	1

ถามว่าความเห็นของผู้จัดการและผู้ช่วยผู้จัดการตรงกันหรือไม่เพียงไร พอจะกล่าวได้ไหมว่าความเห็นของทั้งสองตรงกันในการเลือกพนักงานให้ออกจากงาน

(r_s = -0.94)

16) ในการวิจัยทางการศึกษาให้นักเรียน 150 คน ตอบคำถามเพื่อวัดความเข้าใจบทเรียน พบว่าคำตอบจำแนกเป็นถูกกับผิด สำหรับข้อที่ 1 กับข้อที่ 2 เป็นดังนี้

ข้อ 2 \ ข้อ 1	ตอบผิด	ตอบถูก
ตอบถูก	24	56
ตอบผิด	36	34

จงคำนวณหาสัมประสิทธิ์ ϕ ทดสอบนัยสำคัญที่ระดับ 5% และสรุปว่ามีสหสัมพันธ์ระหว่างการตอบผิดถูกกับคำถามที่ให้นักเรียนหรือไม่ (ϕ = 0.22, Z = 1.81)

17) นักจิตวิทยาศึกษาว่าบิคลิกของสามีกับบิดาของผู้หญิง 591 คน ในด้านที่เกี่ยวกับความสนใจทางวิทยาศาสตร์กับความสนใจทางศาสนา มีสหสัมพันธ์กันหรือไม่ ได้ข้อมูลจากแบบสอบถามที่จำแนกได้ตารางดังนี้

บิดา \ สามี	สนใจศาสนา	สนใจวิทยาศาสตร์
สนใจวิทยาศาสตร์	63	326
สนใจศาสนา	68	134

จงคำนวณสัมประสิทธิ์ ϕ ของสหสัมพันธ์ระหว่างบุคลิกของบิดากับสามี ทดสอบนัยสำคัญที่ระดับ 5% และสรุปผลการทดสอบมาให้ด้วย (ϕ = 0.20, Z = 4.99)

18) สถานีวิทยุ ททท. ทำการสำรวจความเห็นผู้ฟังรายการวิทยุรายการหนึ่งของสถานี ได้คำตอบแยกตามเพศของผู้ฟังได้ดังนี้

เพศ \ ความเห็น	ชอบ	ไม่ชอบ
ชาย	55	45
หญิง	10	90

จำคำนวณสัมประสิทธิ์ ϕ ของสหสัมพันธ์ระหว่างเพศกับความเห็นของผู้ฟังรายการวิทยุรายการนี้ ทดสอบนัยสำคัญที่ระดับ 5% และสรุปผลการสำรวจความเห็นผู้ฟังอันนี้มาด้วย

(ϕ = -0.48, Z = 2.82)

19) จากการสัมภาษณ์ผู้หญิง 200 คน ต่อคำถามที่ว่ามีชีวิตสมรสที่มีความสุขหรือไม่ กับคำถามที่ว่าชีวิตในวัยเด็กที่มีความสุขหรือไม่ ได้คำตอบดังปรากฏในตารางต่อไปนี้

ชีวิตวัยเด็ก \ ชีวิตสมรส	ไม่มีความสุข	มีความสุข
มีความสุข	40	70
ไม่มีความสุข	60	30

จงคำนวณสัมประสิทธิ์ ϕ ของสหสัมพันธ์ระหว่างชีวิตสมรส กับชีวิตวัยเด็กของผู้หญิงจากข้อมูลชุดนี้ ทดสอบนัยสำคัญของ ϕ ที่ระดับ 5% และสรุปผลที่ได้ด้วย(ϕ =0.30 Z = 1.08)

หนังสืออ้างอิง

1) Amos J.R., Brown F.L. and Mink O.G., **Statistical Concepts : A Basic Program,** Harper & Row 1965.
2) Blalock H.M., **Social Statistics,** McGraw-Hill 1972.
3) Chase C.I., **Elementary Statistical Procedures,** McGraw-Hill 1976.
4) Croucher J.S., **Elementary Statistics for Business** McGraw-Hill 1977.
5) Edwards A.L., **Statistical Methods** (second edition), Holt, Rinehart and Winston 1967.
6) Guildford J.P. and Fruchter B., **Fundamental Statistics in Psychology and Education** (fifth edition), McGraw-Hill 1973.
7) Hays W.L., **Statistics for the Social Science** (second edition). McGraw-Hill 1973.
8) Koosis D.J. and Coladarci A.P., **Statistics** : A Self Teaching Guide (second edition), John Wiley & Sons 1977.
9) Maxwell A.E., **Basic Statistics in Behaviorial Research** , Penguin 1972.
10) Tufte E.R., **Data Analysis for Politics and Policy** , Prentice-Hall 1974.