บทที่ 2

ทฤษฎีที่เกี่ยวข้องในการวิจัย

ตัวสถิติต่าง ๆ ที่จำเป็นต้องใช้ในการวิจัยครั้งนี้มีดังต่อไปนี้

2.1 ตัวสถิติที่ใช้สำหรับทดสอบความเป็นอิสระกันของข้อมูลแบบแบ่งกลุ่มในตารางการณ์จร

ส่วนมากแล้วตัวสถิติที่นิยมใช้ทดสอบความเป็นอิสระกันของข้อมูลแบบแบ่งกลุ่มในตารางการณ์จรคือตัวสถิติ χ^2 โดยที่การแจกแจงแบบไคสแควร์ที่ใช้ทดสอบนั้น จะมีองศาความเป็นอิสระ (degree of freedom) เท่ากับ $(r-1)(c-1)$ ค่า χ^2 คำนวณหาได้ตามสูตรในบทที่ 1 โดยค่าต่ำสุดของ χ^2 ที่คำนวณได้จะมีค่าเท่ากับศูนย์ และค่าสูงสุดของ χ^2 นั้นจะมีค่าได้เท่ากับ $n \min\{(r-1),(c-1)\}$[1] กล่าวคือ

$$\chi^2_{max} = n(r-1) \quad \text{เมื่อ} \quad r \leq c$$
$$\qquad = n(c-1) \quad \text{เมื่อ} \quad c \leq r$$

ส่วนผลการทดสอบนั้นทราบได้โดยการนำค่า χ^2 ที่คำนวณได้ไปเทียบกับค่าวิกฤต ถ้าค่าที่คำนวณได้ไม่ตกในเขตวิกฤต จะสรุปผลว่าตัวแปรทั้งสองเป็นอิสระต่อกัน แต่ถ้าค่าที่คำนวณได้ไม่ตกในเขตวิกฤต จะสรุปผลว่าตัวแปรทั้งสองเป็นอิสระต่อกัน

2.2 สัมประสิทธิ์ฟาย

ค่าสัมประสิทธิ์ฟาย ซึ่งเท่ากับ $\sqrt{\chi^2/n}$ ที่ใช้วัดความสัมพันธ์ระหว่างตัวแปรนั้น จะมีค่าต่ำสุดเท่ากับศูนย์ (ตามค่าของ χ^2) แต่ค่าสูงสุดของ ϕ นั้นยังไม่แน่นอน ทั้งนี้ขึ้นอยู่กับขนาดของตารางการณ์จรที่ใช้ทดสอบว่ามีขนาดเท่าใด และเนื่องจากในการวิจัยครั้งนี้เป็นการเปรียบเทียบ

---

[1] Lothar Sachs "Applied Statistics" p. 479.

ตัวสถิติต่าง ๆ ที่ใช้วัดความสัมพันธ์ของข้อมูลแบบแบ่งกลุ่ม ดังนั้นจึงต้องทำการปรับค่าของ ϕ ให้มีค่าอยู่ในช่วง $[0, 1]$ โดยค่าของ ϕ นี้ไม่ว่าตารางการณ์จรจะมีขนาดเท่าใดก็ตาม ค่าต่ำสุดจะเท่ากับศูนย์ แต่ค่าสูงสุดนั้นจะขึ้นอยู่กับค่าของ χ^2 ซึ่งจะเท่ากับ $n \min\{(r-1),(c-1)\}$ ดังนั้น ค่าสูงสุดของ ϕ ในแต่ละกรณีจะมีค่าดังนี้

ตารางการณ์จรขนาด 2x2 ค่าสูงสุดของ $\phi = \sqrt{n/n} = 1$

ตารางการณ์จรขนาด 2x3 ค่าสูงสุดของ $\phi = \sqrt{n/n} = 1$

ตารางการณ์จรขนาด 2x4 ค่าสูงสุดของ $\phi = \sqrt{n/n} = 1$

ตารางการณ์จรขนาด 3x3 ค่าสูงสุดของ $\phi = \sqrt{2n/n} = \sqrt{2}$

ตารางการณ์จรขนาด 3x4 ค่าสูงสุดของ $\phi = \sqrt{2n/n} = \sqrt{2}$

ตารางการณ์จรขนาด 3x5 ค่าสูงสุดของ $\phi = \sqrt{2n/n} = \sqrt{2}$

ตารางการณ์จรขนาด 4x4 ค่าสูงสุดของ $\phi = \sqrt{3n/n} = \sqrt{3}$

ตารางการณ์จรขนาด 4x5 ค่าสูงสุดของ $\phi = \sqrt{3n/n} = \sqrt{3}$

ตารางการณ์จรขนาด 5x5 ค่าสูงสุดของ $\phi = \sqrt{4n/n} = \sqrt{4}$

ดังนั้นจะเห็นได้ว่าตารางการณ์จรขนาด 2x2 2x3 และ 2x4 นั้นค่าสูงสุดของ ϕ จะเท่ากับ 1 แต่ถ้าเป็นตารางการณ์จรขนาด 3x3 3x4 และ 3x5 ค่าสูงสุดของ ϕ จะเท่ากับ $\sqrt{2}$ และตารางการณ์จรขนาด 4x4 4x5 ค่าสูงสุดของ ϕ จะเท่ากับ $\sqrt{3}$ และตารางการณ์จรขนาด 5x5 ค่าสูงสุดของ ϕ จะเท่ากับ 2 จะเห็นได้ว่าค่าสูงสุดของ ϕ จะยังไม่เท่ากับ 1 ในหลาย ๆ กรณี เมื่อทำการคำนวณหาค่าสัมประสิทธิ์ฟายได้แล้ว จะต้องทำการปรับค่าของ ϕ ให้มีค่าอยู่ในช่วง $[0, 1]$ โดยนำค่าสูงสุดของ ϕ ไปหารค่า ϕ เดิม กล่าวคือ

$$\phi_{\text{ใหม่}} = \frac{\phi_{\text{เดิม}}}{\phi_{\text{สูงสุด}}}$$

2.3 สัมประสิทธิ์เงื่อนไขของเพียร์สัน

สัมประสิทธิ์เงื่อนไขของเพียร์สัน ซึ่งมีค่าเท่ากับ $\sqrt{\dfrac{\chi^2}{\chi^2+n}}$ มีค่าต่ำสุดเท่ากับศูนย์ แต่ค่าสูงสุดมีค่าไม่แน่นอนขึ้นอยู่กับขนาดของตารางการณ์จรที่ใช้ทดลอบ ว่ามีขนาดเท่าใดเนื่องจากในการวิจัยครั้งนี้ เป็นการเปรียบเทียบตัวสถิติต่าง ๆ ที่ใช้วัดความสัมพันธ์ของข้อมูลแบบแบ่งกลุ่ม ดังนั้นจึงต้องทำการปรับค่าของ C ให้มีค่าอยู่ในช่วง $[0, 1]$ โดยค่าของ C นี้ไม่ว่าตารางการณ์จรจะ

มีขนาดเท่าใดก็ตาม ค่าต่ำสุดจะเท่ากับศูนย์ แต่ค่าสูงสุดขึ้นอยู่กับค่าของ X^2 ซึ่งจะเท่ากับ $n \min\{(r-1),(c-1)\}$ ดังนั้นค่าสูงสุดของ C ในแต่ละกรณีจะเป็นดังนี้

ตารางการณ์จรขนาด 2x2 ค่าสูงสุดของ C $= \sqrt{\dfrac{n}{n+n}} = \sqrt{1/2}$

ตารางการณ์จรขนาด 2x3 ค่าสูงสุดของ C $= \sqrt{\dfrac{n}{n+n}} = \sqrt{1/2}$

ตารางการณ์จรขนาด 2x4 ค่าสูงสุดของ C $= \sqrt{\dfrac{n}{n+n}} = \sqrt{1/2}$

ตารางการณ์จรขนาด 3x3 ค่าสูงสุดของ C $= \sqrt{\dfrac{2n}{2n+n}} = \sqrt{2/3}$

ตารางการณ์จรขนาด 3x4 ค่าสูงสุดของ C $= \sqrt{\dfrac{2n}{2n+n}} = \sqrt{2/3}$

ตารางการณ์จรขนาด 3x5 ค่าสูงสุดของ C $= \sqrt{\dfrac{2n}{2n+n}} = \sqrt{2/3}$

ตารางการณ์จรขนาด 4x4 ค่าสูงสุดของ C $= \sqrt{\dfrac{3n}{3n+n}} = \sqrt{3/4}$

ตารางการณ์จรขนาด 4x5 ค่าสูงสุดของ C $= \sqrt{\dfrac{3n}{3n+n}} = \sqrt{3/4}$

ตารางการณ์จรขนาด 5x5 ค่าสูงสุดของ C $= \sqrt{\dfrac{4n}{4n+n}} = \sqrt{4/5}$

ดังนั้นจะเห็นได้ว่าตารางการณ์จรขนาด 2x2 2x3 และ 2x4 ค่าสูงสุดของ C จะเท่ากับ $\sqrt{1/2}$ แต่ถ้าเป็นตารางขนาด 3x3 3x4 และ 3x5 ค่าสูงสุดของ C จะเท่ากับ $\sqrt{2/3}$ ตารางการณ์จรขนาด 4x4 และ 4x5 ค่าสูงสุดของ C จะเท่ากับ $\sqrt{3/4}$ และตารางขนาด 5x5 ค่าสูงสุดของ C จะเท่ากับ $\sqrt{4/5}$ ดังนั้นจะเห็นได้ว่าค่าสูงสุดของ C จะไม่เท่ากับ 1 เลย นั่นคือเมื่อคำนวณหาค่าสัมประสิทธิ์เงื่อนไขของเพียร์สันได้แล้ว จะต้องทำการปรับค่าของ C ให้มีค่าอยู่ในช่วง $[0, 1]$ โดยนำค่าสูงสุดของ C ไปหารค่า C เดิม กล่าวคือ

$$C_{\text{ใหม่}} = \frac{C_{\text{เดิม}}}{C_{\text{สูงสุด}}}$$

2.4 <u>สัมประสิทธิ์เงื่อนไขของชูโพร</u>

สัมประสิทธิ์เงื่อนไขของชูโพร ซึ่งมีค่าเท่ากับ $\sqrt{\dfrac{\chi^2/n}{\sqrt{(r-1)(c-1)}}}$ จะมีค่าต่ำสุดเท่ากับศูนย์ แต่ค่าสูงสุดมีค่าไม่แน่นอนขึ้นอยู่กับขนาดของตารางการณ์จรที่ใช้ทดลองว่ามีขนาดเท่าใดและเนื่องจากในการวิจัยครั้งนี้ เป็นการเปรียบเทียบตัวสถิติต่าง ๆ ที่ใช้วัดความสัมพันธ์ของข้อมูลแบบแบ่งกลุ่ม ดังนั้นจึงต้องทำการปรับค่าของ T ให้มีค่าอยู่ในช่วง $[0, 1]$ โดยค่าของ T นี้ไม่ว่าตารางการณ์จรจะมีขนาดเท่าใดก็ตามค่าต่ำสุดจะเท่ากับศูนย์ แต่ค่าสูงสุดนั้นจะขึ้นอยู่กับค่าของ X^2 ซึ่งเท่ากับ $n \min\{(r-1),(c-1)\}$ ดังนั้นค่าสูงสุดของ T ในแต่ละกรณีจะเป็นดังนี้

ตารางการณ์จรขนาด 2x2 ค่าสูงสุดของ T $= \sqrt{\dfrac{n/n}{\sqrt{(2-1)(2-1)}}} = 1$

ตารางการณ์จรขนาด 2x3 ค่าสูงสุดของ T $= \sqrt{\dfrac{n/n}{\sqrt{(2-1)(3-1)}}} = \sqrt{\dfrac{1}{\sqrt{2}}}$

ตารางการณ์จรขนาด 2x4 ค่าสูงสุดของ T $= \sqrt{\dfrac{n/n}{\sqrt{(2-1)(4-1)}}} = \sqrt{\dfrac{1}{\sqrt{3}}}$

ตารางการณ์จรขนาด 3x3 ค่าสูงสุดของ T $= \sqrt{\dfrac{2n/n}{\sqrt{(3-1)(3-1)}}} = 1$

ตารางการณ์จรขนาด 3x4 ค่าสูงสุดของ T $= \sqrt{\dfrac{2n/n}{\sqrt{(3-1)(4-1)}}} = \sqrt{2/\sqrt{6}}$

ตารางการณ์จรขนาด 3x5 ค่าสูงสุดของ T $= \sqrt{\dfrac{2n/n}{\sqrt{(3-1)(5-1)}}} = \sqrt{2/\sqrt{8}}$

ตารางการณ์จรขนาด 4x4 ค่าสูงสุดของ T $= \sqrt{\dfrac{3n/n}{\sqrt{(4-1)(4-1)}}} = 1$

ตารางการณ์จรขนาด 4x5 ค่าสูงสุดของ T $= \sqrt{\dfrac{3n/n}{\sqrt{(4-1)(5-1)}}} = \sqrt{3/\sqrt{12}}$

ตารางการณ์จรขนาด 5x5 ค่าสูงสุดของ T $= \sqrt{\dfrac{4n/n}{\sqrt{(5-1)(5-1)}}} = 1$

ดังนั้นจะเห็นได้ว่าตารางการณ์จรขนาด 2x2 3x3 4x4 5x5 นั้น ค่าสูงสุดของ C จะเท่ากับ 1 ตารางการณ์จรขนาด 2x3 ค่าสูงสุดของ T เท่ากับ $\sqrt{1/\sqrt{2}}$ ตารางการณ์จรขนาด 2x4 ค่าสูงสุดของ T เท่ากับ $\sqrt{1/\sqrt{3}}$ ตารางการณ์จรขนาด 3x4 ค่าสูงสุดของ T เท่ากับ $\sqrt{2/\sqrt{6}}$ ตารางการณ์จรขนาด 3x5 ค่าสูงสุดของ T เท่ากับ $\sqrt{2/\sqrt{8}}$ ตารางขนาด 4x5 ค่าสูงสุดของ T เท่ากับ $\sqrt{3/\sqrt{12}}$ จะเห็นได้ว่าค่าสูงสุดของ T จะไม่เท่ากับ 1 ในหลาย ๆ กรณี นั่นคือเมื่อคำนวณหาค่าสัมประสิทธิ์เงื่อนไขชูโพรได้แล้ว จะต้องปรับค่าของ T ให้มีค่าอยู่ในช่วง $[0, 1]$ โดยนำค่าสูงสุด T ไปหารค่า T เดิม กล่าวคือ

$$T_{\text{ใหม่}} = \frac{T_{\text{เดิม}}}{T_{\text{สูงสุด}}}$$

2.5 <u>สัมประสิทธิ์เงื่อนไขของคราแมร์</u>

สัมประสิทธิ์เงื่อนไขของคราแมร์ซึ่งเท่ากับ $\sqrt{\dfrac{\chi^2}{n\ \min\{(r-1),(c-1)\}}}$

เป็นสัมประสิทธิ์ที่ได้แก้ไขข้อบกพร่องบางประการของสัมประสิทธิ์ ϕ C และ T แล้ว ดังจะเห็นได้จาก ค่า V ที่คำนวณได้จะมีค่าอยู่ในช่วง $[0,1]$ ดังนั้นจึงไม่จำเป็นที่จะต้องปรับค่าของ V อีก

2.6 <u>มาตราวัดเชิงลดสัดส่วนความคลาดเคลื่อน</u> (Proportional-Reduction-in-Error Measures, PRE)

เพื่อที่จะหลีกเลี่ยงจุดอ่อนของดัชนี หรือมาตราวัดที่ขึ้นอยู่กับค่าไคสแควร์ นักสถิติจึงได้พัฒนาวิธีการต่าง ๆ ขึ้นมาใช้ วิธีหนึ่งซึ่งเป็นที่รู้จักกันมากอาศัยแนวคิดง่าย ๆ ของความสัมพันธ์ดังนี้ สมมติแบ่งตัวแปรออกเป็นตัวแปรX และตัวแปร Y โดยเมื่อต้องการเดาค่าของตัวแปร X ซึ่งเป็นตัวแปรตาม การทำนายทำได้ตามกฎ 2 ข้อ ภายใต้กฎข้อแรกนั้นจะไม่มีข้อมูลข่าวสารเพื่อใช้ทำนายค่าของตัวแปรใน X ส่วนภายใต้กฎข้อที่สองจะพิจารณาแต่ละประเภทในตัวแปร Y และอาศัยข้อมูลข่าวสารนั้นไปช่วยทำนายค่าตัวแปร X เมื่อเปรียบเทียบความน่าจะเป็นของการที่ความคลาดเคลื่อนแบบแยกประเภทผิดภายใต้กฎสองข้อนั้นแล้ว จะได้ว่ามาตราวัดความเกี่ยวพันแบบลดความคลาดเคลื่อนเนื่องจากใช้กฎที่สองที่ตรงข้ามกฎข้อแรกจึงเป็นดังนี้

$$PRE = \frac{P(1)-P(2)}{P(1)}$$

ในเมื่อ P(1)และ P(2) เป็นความน่าจะเป็นของการแยกประเภทหน่วยทดลองที่ผิดตามกฎข้อแรกและกฎข้อสองตามลำดับ

จะเห็นได้ว่ามาตราวัด PRE จะให้ค่าอยู่ระหว่าง 0 ถึง 1 ถ้าตัวแปรเป็นอิสระกันแล้ว PRE จะเท่ากับศูนย์ ซึ่งหมายความว่าความน่าจะเป็นของการแยกประเภทหน่วยทดลองที่ผิดภายใต้กฎข้อแรกนั้น เท่ากับความน่าจะเป็นของการทำความคลาดเคลื่อนภายใต้กฎข้อที่สองและ PRE เท่ากับ 1 ถ้า P(2) เท่ากับศูนย์ ซึ่งเป็นกรณีที่จะมีความรู้ในตัวแปร Y จะทำให้การทำนายที่ถูกต้องต่อตัวแปร X ถ้า P(1) เท่ากับ 0 แล้ว ค่า PRE จะกำหนดไม่ได้ (Undefined) แต่กรณีเช่นนี้จะไม่เกิดขึ้นเพราะถ้าไม่มีความเป็นไปได้ในการแยกประเภทผิดตามกฎข้อแรกแล้ว หน่วยทดลองทั้งหมดจะต้องอยู่ในประเภทเดียวกัน และไม่มีความผันแปรในตัวแปร X

เนื่องจากเหตุผลเชิงลดสัดส่วนความคลาดเคลื่อนเป็นที่ใช้กันอย่างกว้างขวาง มาตรวัดแบบแบ่งกลุ่มต่าง ๆ ที่เกี่ยวกับความเกี่ยวพัน จึงขึ้นอยู่กับเหตุผลนั้น ความหมายและการคำนวณของมาตราวัดต่าง ๆ เหล่านั้นก็อาศัยนิยามของความคลาดเคลื่อนนั่นเอง

2.7 สัมประสิทธิ์การทำนายของกัทแมน

สัมประสิทธิ์การทำนายของกัทแมนซึ่งมีค่าเท่ากับ $\dfrac{(\Sigma f_r - \Sigma f_c)-(F_r+F_c)}{2n-(F_r+F_c)}$

โดยเป็นมาตราวัดที่ใช้นิยามของความคลาดเคลื่อนของการทำนายโดยตรง

โดยกฎข้อแรก ถ้าทำนายประเภทของตัวแปร X โดยไม่อาศัยความรู้ในการแยกประเภทของตัวแปร Y แล้ว การทำนายก็คือ เดาประเภทของตัวแปร X ที่เกิดขึ้นบ่อยที่สุด ซึ่งจะให้สัดส่วนสูงสุด เพราะค่าสังเกตส่วนมากจะอยู่ในประเภทนี้ และในระยะยาวของความคลาดเคลื่อนจะเกิดขึ้นน้อย สำหรับตารางการณ์จรขนาด rxC ถ้าให้ $P_{m.}$ แทนความน่าจะเป็นหรือสัดส่วนตามแนวนอนริมสุดที่มากที่สุด (Maximum marginal row Probability) เมื่อไม่ทราบค่าตัวแปร Y ควรจะเดาประเภทที่สอดคล้องกับความน่าจะเป็น $P_{m.}$ ความน่าจะเป็นที่จะทำการทำนายถูกต้องคือ $P_{m.}$ ในเมื่อความน่าจะเป็นของความคลาดเคลื่อนคือ

$$P(1) = 1-P_{m.}$$
$$= 1-Fr/n$$

ตามกฎข้อสองผู้สำรวจจะเลือกหน่วยแบบสุ่ม และพิจารณาประเภทของหน่วยในตัวแปร Y แล้วจึงทำนายประเภทของตัวแปร X ในการทำนายนั้นก็ต้องพิจารณาในแถวตั้ง (หรือแต่ละประเภทของตัวแปร Y) แล้วผู้สำรวจก็จะเลือกประเภทของตัวแปร Xที่เกิดขึ้นบ่อยที่สุด หรืออาจจะกล่าวได้ว่า เมื่อกำหนดประเภทในตัวแปร Y ให้ ก็จะเลือกประเภทของตัวแปรX ที่มีสัดส่วนสูงสุด ความคลาดเคลื่อนจะเกิดขึ้นได้ แต่ก็จะน้อยกว่าเลือกประเภทอื่นจากตัวแปร X

ในตารางการณ์จรขนาด rxC ให้ P_{mj} แทนความน่าจะเป็นหรือสัดส่วนที่มากที่สุดของแถวที่ j ความน่าจะเป็นของความคลาดเคลื่อนภายใต้กฎข้อที่สอง จะเป็น

$$P(2) = 1-\sum_j P_{mj} \quad ; \quad j = 1,2...,c$$
$$= 1-\Sigma fc/n$$

ดังนั้นจะได้มาตราวัดแบบลดความคลาดเคลื่อนหรือ λ_X โดยที่

$$\lambda_X = \frac{(1-P_{m.})-(1-\sum_j P_{mj})}{(1-P_{m.})} \quad ; \quad j = 1,2,...,c$$

$$= \frac{(1-F_{r/n})-(1-\Sigma fc/n)}{(1-F_{r/n})}$$

$$= \frac{\Sigma f_c - F_r}{n-F_r}$$

โดยที่ λ_X จะแสดงประเภทของตัวแปร X ที่ได้รับการทำนายจากข้อมูลข่าวสารของตัวแปร Y ดังนั้นด้วยเหตุเดียวกัน เมื่อเราให้ตัวแปร Y เป็นตัวแปรตามแล้ว เราจะได้มาตราวัด λ_Y โดย

$$\lambda_Y = \frac{(1-P_{.m})-(1-\sum_i P_{im})}{(1-P_{.m})} \quad ; i=1,2,...,r$$

$$= \frac{(1-F_{c/n})-(1-\Sigma f_{r/n})}{(1-F_{c/n})}$$

$$= \frac{\Sigma f_r - F_c}{n-F_c}$$

ในเมื่อ $P_{.m}$ เป็นสัดส่วน หรือความน่าจะเป็นที่สูงสุดของด้านข้างของแถวตั้ง และ P_{im} เป็นความน่าจะเป็นที่สูงสุดในเซลของแถวนอนที่ i

ซึ่งมาตราวัด λ_X และ λ_Y นี้ถือว่าเป็นมาตราวัดแบบไม่สมมาตร (asymmetric measure) ดังนั้น λ_X จะไม่เท่ากับ λ_Y

ในบางครั้งผู้สำรวจไม่ทราบ หรือไม่ตั้งใจจะรวมค่าตัวแปรโดยเป็นตัวแปรควบ (ตัวแปรที่จะทำนาย) ในกรณีนี้ก็จำเป็นต้องใช้มาตราวัดแบบสมมาตร คือ λ ซึ่งได้จากการปรับปรุงเหตุผลเชิงลดสัดส่วนความคลาดเคลื่อนนั่นเอง

สัมประสิทธิ์แบบสมมาตร λ จะรวมเหตุผลของการคำนวณทั้ง λ_X และ λ_Y ดังนั้นเราจะได้ P(1) และ P(2)ดังนี้

$$P(1) = 1-(P_{m.}+P_{.m})/2$$

$$= 1-(F_{r/n}+F_{c/n})/2$$

$$= \frac{2n-(F_r+F_c)}{2n}$$

$$P(2) \quad = \quad 1-(\sum_{j} P_{mj}+\sum_{i} P_{im})/2$$

$$= \quad 1-(\Sigma f_{c/n}+\Sigma f_{r/n})/2$$

$$= \quad \frac{2n-(\Sigma f_c+\Sigma f_r)}{2n}$$

นั่นคือจะได้

$$\lambda \quad = \quad \frac{P(1)-P(2)}{P(1)}$$

$$= \quad \frac{\left[\frac{2n-(F_r+F_c)}{2n}\right] - \left[\frac{2n-(\Sigma f_c+\Sigma f_r)}{2n}\right]}{\frac{2n-(F_r+F_c)}{2n}}$$

$$= \quad \frac{(\Sigma f_c+\Sigma f_r)-(F_r+F_c)}{2n-(F_r+F_c)}$$

2.8 สัมประสิทธิ์แบบกูดแมนและครัสคัล

สัมประสิทธิ์แบบกูดแมนและครัสคัลมีค่าเท่ากับ

$$\frac{\{\sum_i\sum_j f^2_{ij}/f_{.j} - \sum_i f^2_{i./n}\}+\{\sum_i\sum_j f^2_{ij}/f_{i.} - \sum_j f^2_{.j/n}\}}{2n-\{\sum_i f^2_{i./n}+\sum_j f^2_{.j/n}\}}$$ มาตราวัดนี้เป็น

การแก้ไขการเดาเกมที่สมมติไว้ตามวิธี PRE โดยเดาสู่หน่วยทดลองที่กำหนดให้แก่ตัวแปรตาม X ด้วย การทราบหรือไม่ทราบตัวแปรอิสระ แต่การกำหนดจะคงรักษาการแจกแจงเดิมอยู่

มาตราวัด τ จะเป็นสัดส่วนกับจำนวนหน่วยทดลองในประเภทต่าง ๆ ของตัวแปร โดยให้ประเภทของตัวแปร X และ Y เป็น X_1, X_2, $X_3 \ldots, X_r$ และ $Y_1, Y_2, \ldots, Y_c$ ภายใต้กฎข้อแรก จะเดาเป็นประเภทแรกของตัวแปร X หรือ X_1 ด้วยความน่าจะเป็น $P_{1.}$ ประเภทที่สองด้วยความน่าจะเป็น $P_{2.}$ และเรื่อย ๆ ไปสำหรับ r ประเภทของตัวแปร X อัตราความคลาดเคลื่อนหวังระยะยาว (long-run expected error rate) จะเป็น P(1) โดย

$$P(1) \quad = \quad 1-\sum_i P_i^2$$

$$= \quad 1-\left[\sum_i (f_{i/n})^2\right]$$

$$= \quad \frac{n-\sum_i f^2_{i/n}}{n}$$

ภายใต้กฎข้อที่สองจะเดาตัวแปร X_1 ด้วยความน่าจะเป็น $P_{1j}/P_{.j}$ (ความน่าจะเป็นเงื่อนไขของ X_1 เมื่อกำหนด Y_1 ให้) X_2 ด้วยความน่าจะเป็น $P_{2j}/P_{.j}$ และต่อ ๆ ไป สำหรับทุกค่าของ r และ c อัตราความคลาดเคลื่อนคาดหวังสำหรับวิธีการคือ P(2) โดย

$$P(2) \quad = \quad 1-\sum_i\sum_j P^2_{ij}/P_{.j}$$

$$= \quad 1-\left[\sum_i\sum_j (f_{ij/n})^2/(f_{.j/n})\right]$$

$$= \quad 1-\left[\sum_i\sum_j f^2_{ij}/nf_{.j}\right]$$

$$= \quad \frac{n-\sum_i\sum_j f^2_{ij}/f_{.j}}{n}$$

ดังนั้นการลดแบบสัดส่วนในความคลาดเคลื่อนที่ขึ้นอยู่กับเงื่อนไขก็จะเป็น τ_X โดย

$$\tau_X \quad = \quad \frac{P(1)-P(2)}{P(1)}$$

$$= \quad \frac{\left[\frac{n-\sum_i f^2_{i/n}}{n}\right] - \left[\frac{n-\sum_i\sum_j f^2_{ij}/f_{.j}}{n}\right]}{\left[\frac{n-\sum_i f^2_{i/n}}{n}\right]}$$

$$= \quad \frac{\sum_i\sum_j f^2_{ij}/f_{.j}-\sum_i f^2_{i/n}}{n-\sum_i f^2_{i/n}}$$

ในทำนองเดียวกัน τ_Y ก็หาได้เป็น

$$\tau_Y \quad = \quad \frac{\sum_i\sum_j f^2_{ij}/f_{i.}-\sum_j f^2_{.j/n}}{n-\sum_j f^2_{.j/n}}$$

ดังนั้น

$$\tau \quad \frac{\{\sum_i\sum_j f_{ij}^2/f_{.j}-\sum_i f_{i.}^2/n\}+\{\sum_i\sum_j f_{ij}^2/f_{i.}-\sum_j f_{.j}^2/n\}}{2n-\{\sum_i f_{i.}^2/n+\sum_j f_{.j}^2/n\}}$$

ในเมื่อ $f_{i.}$ $f_{.j}$ และ f_{ij} เป็นความถี่ที่สอดคล้องกับ $P_{i.}$, $P_{.j}$ และ P_{ij} ตามลำดับ สำหรับ τ นั้นเป็นมาตราวัดแบบสมมาตร โดยค่าของ τ จะอยู่ระหว่าง 0 กับ 1 เช่นเดียวกับค่าของ τ_X และ τ_Y แต่ค่าของ τ_X และ τ_Y นั้นเป็นมาตราวัดแบบไม่สมมาตร

2.9 ความสัมพันธ์ระหว่างสัมประสิทธิ์ฟาย สัมประสิทธิ์เงื่อนไขของชูโพร และสัมประสิทธิ์เงื่อนไขของคราแมร์

ตัวสถิติทั้ง 3 นี้ เมื่อคำนวณหาค่าและปรับค่าของตัวสถิติทั้ง 3 ให้มีค่าอยู่ในช่วง 0 ถึง 1 โดยนำค่าสูงสุดของตัวสถิตินั้น ๆไปหารแล้ว จะพบว่าตัวสถิติทั้ง 3 จะให้ค่าที่ไม่แตกต่างกัน ไม่ว่าขนาดของตารางการณ์จรที่ใช้สำหรับคำนวณหาตัวสถิติทั้ง 3 จะมีขนาดใดก็ตาม เช่น

1) ตารางการณ์จรขนาด 2x2 ตัวสถิติทั้ง 3 จะมีค่าดังนี้

เนื่องจากตัวสถิติ ϕ สำหรับตารางการณ์จรขนาด 2x2 มีค่าสูงสุดเป็น 1 ดังนั้น

$$\phi = \sqrt{\chi^2/n}$$

เนื่องจากตัวสถิติ T สำหรับตารางการณ์จรขนาด 2x2 มีค่าสูงสุดเป็น 1 ดังนั้น

$$T = \sqrt{\frac{\chi^2/n}{\sqrt{(2-1)(2-1)}}} = \sqrt{\chi^2/n}$$

เนื่องจากตัวสถิติ V สำหรับตารางการณ์จรขนาด 2x2 จะมีค่าสูงสุดเป็น 1 ดังนั้น

$$V = \sqrt{\frac{\chi^2}{n\min\{(2-1),(2-1)\}}} = \sqrt{\chi^2/n}$$

ดังนั้นสามารถสรุปได้ว่า กรณีตารางการณ์จรขนาด 2x2 ตัวสถิติทั้ง 3 จะมีค่าเท่ากันเป็น $\sqrt{\chi^2/n}$

2) ตารางการณ์จรขนาด 2x3 ตัวสถิติทั้ง 3 จะมีค่าดังนี้

เนื่องจากตัวสถิติ ϕ สำหรับตารางการณ์จรขนาด 2x3 มีค่าสูงสุดเป็น 1 ดังนั้น

$$\phi = \sqrt{\chi^2/n}$$

เนื่องจากตัวสถิติ T สำหรับตารางการณ์จรขนาด 2x3 มีค่าสูงสุดเป็น $\sqrt{\frac{1}{\sqrt{2}}}$

ดังนั้น

$$T = \sqrt{\frac{\chi^2/n}{\sqrt{(2-1)(3-1)}}} \Big/ \sqrt{\frac{1}{\sqrt{2}}} = \sqrt{\chi^2/n}$$

เนื่องจากตัวสถิติ V สำหรับตารางการณ์จรขนาด 2x3 มีค่าสูงสุดเป็น 1 ดังนั้น

$$V = \sqrt{\frac{\chi^2}{n.\ \mathrm{Min}\{(2-1),(3-1)\}}} = \sqrt{\chi^2/n}$$

ดังนั้นสามารถสรุปได้ว่า กรณีตารางการณ์จรขนาด 2x3 ตัวสถิติทั้ง 3 จะมีค่าเท่ากันเป็น $\sqrt{\chi^2/n}$

3) ตารางการณ์จรขนาด 2x4 ตัวสถิติทั้ง 3 จะมีค่าดังนี้

เนื่องจากตัวสถิติ ϕ สำหรับตารางการณ์จรขนาด 2x4 มีค่าสูงสุดเป็น 1 ดังนั้น

$$\phi = \sqrt{\chi^2/n}$$

เนื่องจากตัวสถิติ T สำหรับตารางการณ์จรขนาด 2x4 มีค่าสูงสุดเป็น $\sqrt{1/\sqrt{3}}$

$$T = \sqrt{\frac{\chi^2/n}{\sqrt{(2-1)(4-1)}}} \Big/ \sqrt{1/\sqrt{3}} = \sqrt{\chi^2/n}$$

เนื่องจากตัวสถิติ V สำหรับตารางการณ์จรขนาด 2x4 มีค่าสูงสุดเป็น 1 ดังนั้น

$$V = \sqrt{\frac{\chi^2}{n\ \min\{(2-1),(4-1)\}}} = \sqrt{\chi^2/n}$$

ดังนั้นสามารถสรุปได้ว่า กรณีตารางการณ์จรขนาด 2x4 ตัวสถิติทั้ง 3 จะมีค่าเท่ากันเป็น $\sqrt{\chi^2/n}$

4) ตารางการณ์จรขนาด 3x3 ตัวสถิติทั้ง 3 จะมีค่าดังนี้

เนื่องจากตัวสถิติ ϕ สำหรับตารางการณ์จรขนาด 3x3 มีค่าสูงสุดเป็น $\sqrt{2}$ ดังนั้น

$$\phi = \sqrt{\chi^2/n}\Big/\sqrt{2} = \sqrt{\chi^2/2n}$$

เนื่องจากตัวสถิติ T สำหรับตารางการณ์จรขนาด 3x3 มีค่าสูงสุดเป็น 1 ดังนั้น

$$T = \sqrt{\frac{\chi^2/n}{\sqrt{(3-1)(3-1)}}} = \sqrt{\chi^2/2n}$$

เนื่องจากตัวสถิติ V สำหรับตารางการณ์จรขนาด 3x3 มีค่าสูงสุดเป็น 1 ดังนั้น

$$V = \sqrt{\frac{\chi^2}{n \min\{(3-1),(3-1)\}}} \quad \sqrt{\chi^2/2n}$$

ดังนั้นสามารถสรุปได้ว่า กรณีตารางการณ์จรขนาด 3x3 ตัวสถิติทั้ง 3 จะมีค่าเท่ากันเป็น $\sqrt{\chi^2/2n}$

5) ตารางการณ์จรขนาด 3x4 ตัวสถิติทั้ง 3 จะมีค่าดังนี้

เนื่องจากตัวสถิติ ϕ สำหรับตารางการณ์จรขนาด 3x4 มีค่าสูงสุดเป็น $\sqrt{2}$ ดังนั้น

$$\phi = \sqrt{\chi^2/n}\Big/\sqrt{2} = \sqrt{\chi^2/2n}$$

เนื่องจากตัวสถิติ T สำหรับตารางการณ์จรขนาด 3x4 มีค่าสูงสุดเป็น $\sqrt{2/\sqrt{6}}$ ดังนั้น

$$T = \sqrt{\frac{\chi^2/n}{\sqrt{(3-1)(4-1)}}}\Big/\sqrt{2/\sqrt{6}} = \sqrt{\chi^2/2n}$$

เนื่องจากตัวสถิติ V สำหรับตารางการณ์จรขนาด 3x4 มีค่าสูงสุดเป็น 1 ดังนั้น

$$V = \sqrt{\frac{\chi^2}{n \min\{(3-1),(4-1)\}}} = \sqrt{\chi^2/2n}$$

ดังนั้นสามารถสรุปได้ว่า กรณีตารางการณ์จรขนาด 3x4 ตัวสถิติทั้ง 3 จะมีค่าเท่ากันเป็น $\sqrt{\chi^2/2n}$

6) ตารางการณ์จรขนาด 3x5 ตัวสถิติทั้ง 3 จะมีค่าดังนี้

เนื่องจากตัวสถิติ ϕ สำหรับตารางการณ์จรขนาด 3x5 มีค่าสูงสุดเป็น $\sqrt{2}$ ดังนั้น

$$\phi = \sqrt{\chi^2/n}\Big/\sqrt{2} = \sqrt{\chi^2/2n}$$

เนื่องจากตัวสถิติ T สำหรับตารางการณ์จรขนาด 3x5 มีค่าสูงสุดเป็น $\sqrt{2/\sqrt{8}}$ ดังนั้น

$$T = \sqrt{\frac{\chi^2/n}{\sqrt{(3-1)(5-1)}}}\Bigg/\sqrt{2/\sqrt{8}} = \sqrt{\chi^2/2n}$$

เนื่องจากตัวสถิติ V สำหรับตารางการณ์จรขนาด 3x5 มีค่าสูงสุดเป็น 1 ดังนั้น

$$V = \sqrt{\frac{\chi^2}{n \min\{(3-1),(5-1)\}}} = \sqrt{\chi^2/2n}$$

ดังนั้นสามารถสรุปได้ว่า กรณีตารางการณ์จรขนาด 3x5 ตัวสถิติทั้ง 3 จะมีค่าเท่ากันเป็น $\sqrt{\chi^2/2n}$

7) ตารางการณ์จรขนาด 4x4 ตัวสถิติทั้ง 3 จะมีค่าดังนี้

เนื่องจากตัวสถิติ ϕ สำหรับตารางการณ์จรขนาด 4x4 มีค่าสูงสุดเป็น $\sqrt{3}$ ดังนั้น

$$\phi = \sqrt{\chi^2/n}\Big/\sqrt{3} = \sqrt{\chi^2/3n}$$

เนื่องจากตัวสถิติT สำหรับตารางการณ์จรขนาด 4x4 มีค่าสูงสุดเป็น 1 ดังนั้น

$$T = \sqrt{\frac{\chi^2/n}{\sqrt{(4-1)(4-1)}}} = \sqrt{\chi^2/3n}$$

เนื่องจากตัวสถิติ V สำหรับตารางการณ์จรขนาด 4x4 มีค่าสูงสุดเป็น 1 ดังนั้น

$$V = \sqrt{\frac{\chi^2}{n \min\{(4-1),(4-1)\}}} = \sqrt{\chi^2/3n}$$

ดังนั้นสามารถสรุปได้ว่า กรณีตารางการณ์จรขนาด 4x4 ตัวสถิติทั้ง 3 จะมีค่าเท่ากัน

เป็น $\sqrt{\chi^2/3n}$

8) ตารางการณ์จรขนาด 4x5 ตัวสถิติทั้ง 3 จะมีค่าดังนี้

เนื่องจากตัวสถิติ ϕ ในตารางการณ์จรขนาด 4x5 มีค่าสูงสุดเป็น $\sqrt{3}$ ดังนั้น

$$\phi = \sqrt{\chi^2/n}\Big/\sqrt{3} = \sqrt{\chi^2/3n}$$

เนื่องจากตัวสถิติ T ในตารางการณ์จรขนาด 4x5 มีค่าสูงสุดเป็น $\sqrt{3/\sqrt{12}}$

ดังนั้น

$$T = \sqrt{\frac{\chi^2/n}{\sqrt{(4-1(5-1)}}}\Big/\sqrt{3/\sqrt{12}} = \sqrt{\chi^2/3n}$$

เนื่องจากตัวสถิติ V ในตารางการณ์จรขนาด 4x5 มีค่าสูงสุดเป็น 1 ดังนั้น

$$V = \sqrt{\frac{\chi^2}{n\min\{(4-1),(5-1)\}}} = \sqrt{\chi^2/_{3n}}$$

ดังนั้น สามารถสรุปได้ว่า กรณีตารางการณ์จรขนาด 4x5 ตัวสถิติทั้ง 3 จะมีค่า

เท่ากันเป็น $\sqrt{\chi^2/3n}$

9) ตารางการณ์จรขนาด 5x5 ตัวสถิติทั้ง 3 จะมีค่าดังนี้

เนื่องจากตัวสถิติ ϕ ในตารางการณ์จรขนาด 5x5 มีค่าสูงสุดเป็น 2 ดังนั้น

$$\phi = \sqrt{\chi^2/n}\Big/2 = \sqrt{\chi^2/4n}$$

เนื่องจากตัวสถิติ T ในตารางการณ์จรขนาด 5x5 มีค่าสูงสุดเป็น 1 ดังนั้น

$$T = \sqrt{\frac{\chi^2/n}{\sqrt{(5-1)(5-1)}}} = \sqrt{\chi^2/4n}$$

เนื่องจากตัวสถิติ V ในตารางการณ์จรขนาด 5x5 มีค่าสูงสุดเป็น 1 ดังนั้น

$$V = \sqrt{\frac{\chi^2}{n\min\{(5-1),(5-1)\}}} = \sqrt{\chi^2/4n}$$

ดังนั้นสามารถสรุปได้ว่า กรณีตารางการณ์จรขนาด 5x5 ตัวสถิติทั้ง 3 จะมีค่าเท่ากันเป็น $\sqrt{\chi^2/4n}$

2.10 ความคลาดเคลื่อนกำลังสองเฉลี่ย (mean squared error)

ในการพิจารณาคุณสมบัติของตัวประมาณ มักจะพบว่าคุณสมบัติบางประการมักจะไม่เหมือนกัน เช่น ความไม่เอนเอียงของตัวประมาณ กับความคลาดเคลื่อนต่ำสุดของตัวประมาณ

ถ้า $\hat{\theta}$ เป็นตัวประมาณของพารามิเตอร์ θ ความคลาดเคลื่อนกำลังสองเฉลี่ยหรือ MSE ของ $\hat{\theta}$ คือ

$$\begin{aligned} MSE &= E(\hat{\theta}-\theta)^2 \\ &= V(\hat{\theta})+(Bias)^2 \\ &= V(\hat{\theta})+(E(\hat{\theta})-\theta)^2 \end{aligned}$$

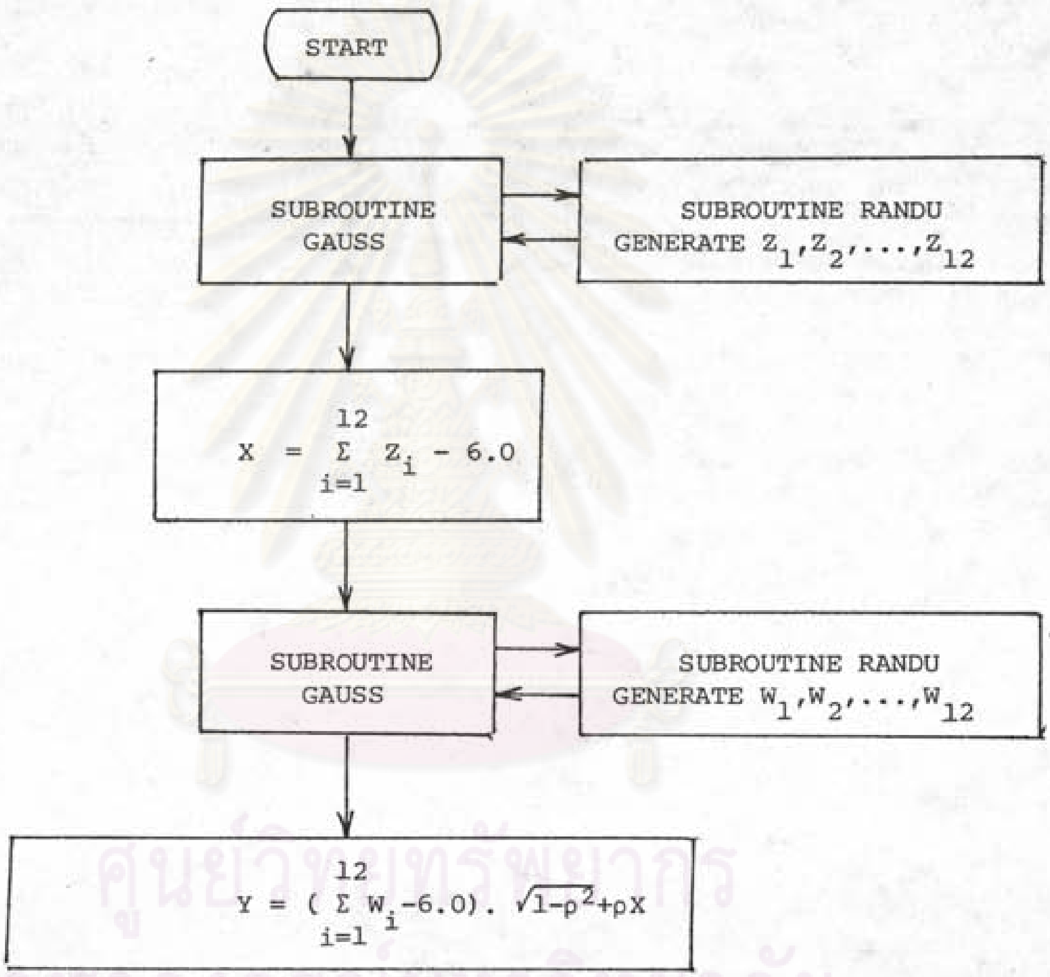
ตัวประมาณ $\hat{\theta}$ เป็นตัวประมาณที่มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด ถ้า $\hat{\theta}$ มีค่า $E(\hat{\theta}-\theta)^2$ ต่ำที่สุด หรือ

$$MSE \simeq \frac{\Sigma(\hat{\theta}-\theta)^2}{n}$$

2.11 การสร้างข้อมูลที่มีการแจกแจงแบบปกติสองตัวแปรโดยวิธีซิมูเลชั่น (simulation)

ในการวิจัยครั้งนี้ผู้วิจัยต้องสร้างตัวแปรที่มีการแจกแจงปกติสองตัวแปร โดยการแทนค่า ρ ระดับต่าง ๆ แล้วหาตัวแปร X และ Y ที่มีความสัมพันธ์กันในระดับของ ρ ที่กำหนดขึ้น โดยเริ่มจากการสร้างตัวเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอ (uniform distribution) ขึ้นมาก่อน โดยใช้โปรแกรม SUBROUTINE RANDU จากนั้นนำผลที่ได้จากการสร้างตัวเลขสุ่มแบบสม่ำเสมอไปสร้างตัวแปรแบบปกติมาตรฐานโดยใช้โปรแกรม SUBROUTINE GAUSS โดยค่าที่ได้จากโปรแกรมนี้เราแทนด้วยตัวแปร X จากนั้นสร้างตัวเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอขึ้นมาอีกชุดหนึ่ง (โดยไม่ขึ้นอยู่กับตัวเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอชุดก่อน) โดยใช้โปรแกรม SUBROUTINE RANDU จากนั้นนำค่าตัวเลขสุ่มมาสร้างตัวแปรที่มีการแจกแจงแบบปกติ โดยเรียกค่าที่ได้ว่า Y โดยค่าของ Y นี้จะมีลักษณะการแจกแจงที่ขึ้นอยู่กับค่าของ X ที่ได้จากครั้งแรก

การสร้างตัวแปรปกติมาตรฐาน 1 ตัว โดยใช้โปรแกรม SUBROUTINE GAUSS นี้ โปรแกรม SUBROUTINE GAUSS จะเรียกตัวเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอ 12 ตัว มาใช้ (ดูโปรแกรมในภาคผนวก ก) ผลลัพธ์ที่ได้จะเป็นค่า (X,Y) ที่เป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติสองตัวแปร ที่มีความสัมพันธ์ในระดับของ ρ ตามที่กำหนดดังแผนผังต่อไปนี้



<u>แผนผังที่ 1</u> การสร้างตัวแปรสุ่มที่มีการแจกแจงแบบปกติสองตัวแปร

ในการสร้างตัวแปรสุ่มที่มีการแจกแจงแบบปกติสองตัวแปร โดยใช้โปรแกรม SUBROUTINE GAUSS ได้อาศัยทฤษฎีลิมิตส่วนกลาง (Central limit theorem) ซึ่งมีสาระที่สำคัญดังนี้

ถ้า $Y_1\ Y_2\ Y_3,\ldots,Y_n$ แทนตัวอย่างสุ่มขนาด n ที่เป็นอิสระต่อกัน ซึ่งเลือกมาจากประชากรที่มีการแจกแจงแบบปกติ โดยมีค่าเฉลี่ย μ และค่าความแปรปรวน σ^2 แล้ว การแจกแจงของตัวสถิติ

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

จะประมาณได้ด้วยการแจกแจงแบบปกติที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2/n

เนื่องจากโปรแกรม GAUSS ใช้ตัวแปรที่มีการแจกแจงแบบสม่ำเสมอในช่วง $[0,1]$ 12 ตัว โดยมีค่าเฉลี่ยเท่ากับ 1/2 และค่าความแปรปรวนเท่ากับ 1/144

จะได้ $$\bar{Y} = \frac{1}{12}\sum_{i=1}^{12} Y_i \sim N(1/2,\ 1/144)$$

ดังนั้น

$$X = \frac{\bar{Y}-\mu_{\bar{Y}}}{\sigma_{\bar{Y}}}$$

$$= \frac{\bar{Y}-1/2}{\sqrt{1/144}}$$

$$= \sum_{i=1}^{12} Y_i - 6.0$$

จะมีการแจกแจงโดยประมาณเป็นแบบปกติมาตรฐาน

2.12 การแจกแจงแบบปกติสองตัวแปร (bivariate normal distribution)

ถ้า (X,Y) เป็นตัวแปรที่มีการแจกแจงเป็นแบบปกติฟังก์ชั่นการแจกแจงความน่าจะเป็นของตัวแปรทั้งสองจะเป็น

$$f(x,y) = \frac{1}{2\Pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{X-\mu_X}{\sigma_X}\right) - 2\rho\left(\frac{X-\mu_X}{\sigma_X}\right)\left(\frac{Y-\mu_Y}{\sigma_Y}\right) + \left(\frac{Y-\mu_Y}{\sigma_Y}\right)^2\right\}\right]$$

เมื่อ $-\infty<X<\infty$ $-\infty<Y<\infty$ และ $-1\le\rho\le 1$ $\sigma_X^2 \ge 0$ $\sigma_Y^2 \ge 0$

$-\infty<\mu_X<\infty$ $-\infty<\mu_Y<\infty$

โดยมีค่าเฉลี่ยและความแปรปรวนดังนี้

$$\underset{\sim}{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sigma^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$$

เมื่อ ρ คือค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปร X และ Y

μ_X คือค่าเฉลี่ยของตัวแปร X

μ_Y คือค่าเฉลี่ยของตัวแปร Y

σ_X^2 คือความแปรปรวนของตัวแปร X

σ_Y^2 คือความแปรปรวนของตัวแปร Y