บทที่ 3

การนำค่าสถิติทีและเอฟมาใช้ในการวิจัยทางการศึกษา

การแจกแจงที (t-Distribution)

ประวัติการพัฒนาของการแจกแจงที

วิลเลียม ซีลี กอสเซท (William Sealy Gosset 1876 - 1937) ผู้ใช้นามแฝงว่าสติวเดนท์ (Student) นักสถิติชาวอังกฤษได้เสนอการแจกแจงทีเป็นคนแรกในปี ค.ศ.1908 โดยเริ่มแรกได้สนใจศึกษาการแปรผันของส่วนเบี่ยงเบนมาตรฐานจากกลุ่มตัวอย่างขนาดเล็กหลาย ๆ กลุ่ม (Sampling Fluctuation of σ) พบว่าไม่เป็นโค้งปกติ จึงได้ค้นคว้าหาขนาดที่เหมาะสมของกลุ่มตัวอย่างขนาดใหญ่ และค้นหาวิธีปฏิบัติเมื่อกลุ่มตัวอย่างมีขนาดเล็ก สติวเดนท์พบว่า การแจกแจงตัวอย่างการสุ่มของความแปรปรวน (Sampling Distribution of S^2) นั้น ค่าโมเมนต์ 4 โมเมนต์แรกมีลักษณะเหมือนโค้งชนิดที่ 3 ของเพียร์สัน[1] (Pearson Type III Curve) ต่อมาจึงศึกษาลักษณะการแจกแจงอัตราส่วน $\frac{\bar{X} - \mu}{S}$ โดยเรียกว่าการแจกแจงซีของสติวเดนท์[2] (Student's Z Distribution) ข้อค้นพบได้ลงพิมพ์เป็นบทความเรื่อง The Probable Error of a Mean ในวารสาร Biometrika[3] บทความนี้ได้กล่าวถึงวิธีปฏิบัติเพื่อหาข้อสรุปเกี่ยวกับมัชฌิมเลขคณิตของประชากรจากกลุ่มตัวอย่างขนาดเล็ก และกล่าวถึงความแปรปรวนของกลุ่มตัวอย่างขนาดเล็กด้วย

---

[1] William Edwards Reming, Some Theory of Sampling, (New York : John Wiley & Sons, Inc., 1955), pp. 537 - 539.

[2] Udny G. Yule and Maurice G. Kendall, An Introduction to the Theory of Statistics, (New York : Hofter Publishing Company, 1950), p. 485.

[3] Taro Yamane, Statistics : An Introductory Analysis, (2d ed. New York : A Farper International Edition, 1967), p. 647.

ในระยะแรก การแจกแจงทีใช้ทดสอบความมีนัยสำคัญของมัชฌิมเลขคณิตเท่านั้น ต่อมา ค.ศ.1926 เซอร์ โรนัลด์ เอ.ฟิชเชอร์ (Sir Ronald A. Fisher) ได้ปรับปรุงแก้ไขให้ชื่อว่า การแจกแจงทีของฟิชเชอร์ (Fisher's t Distribution) หรือบางครั้งเรียกว่า การแจกแจงของที (Distribution of t) นำมาใช้ทดสอบความมีนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์[4] ในปี ค.ศ.1940 โฮเทลลิ่ง (Hotelling) ใช้ทดสอบความแตกต่างระหว่างสัมประสิทธิ์สหสัมพันธ์ของประชากร 2 กลุ่ม[5] ในปี ค.ศ. 1947 เวลช (Welch) ได้พัฒนาการแจกแจงทีเพื่อใช้กับประชากรที่มีความแปรปรวนไม่เท่ากัน โดยกำหนดสูตรสำหรับคำนวณชั้นความเป็นอิสระ (Degree of Freedom) ที่เหมาะสมขึ้น และในปี ค.ศ.1950 คอแครน กับคอก (Cochran and Cox) ได้พัฒนาการแจกแจงทีสำหรับใช้กับประชากรที่มีความแปรปรวนไม่เท่ากัน โดยกำหนดสูตรสำหรับคำนวณค่าวิกฤตของทีที่ระดับความมีนัยสำคัญ .05[6]

การแจกแจงที คือการแจกแจงตามทฤษฎีของค่าที (t) ของตัวอย่างทั้งหมดเท่าที่จะเป็นได้ การแจกแจงนี้มีความถี่สูงสุดรวมอยู่ที่จุดศูนย์กลางของโค้งการแจกแจงและกระจายออกไปทางด้านค่าบวกและค่าลบอย่างสม่ำเสมอกันทั้งสองด้าน ปลายของโค้งการแจกแจงจะเข้าใกล้ค่าอนันต์ทั้งทางค่าบวก $(+\infty)$ และทางค่าลบ $(-\infty)$[7]

---

[4] George ". Snedecor and William G. Cochran, Statistical Method. (Iowa : The Iowa State University Press Ames, 1968),p.60.

[5] John G. Peatman, Introduction to Applied Statistics. (New York : Parper & Row, 1964).

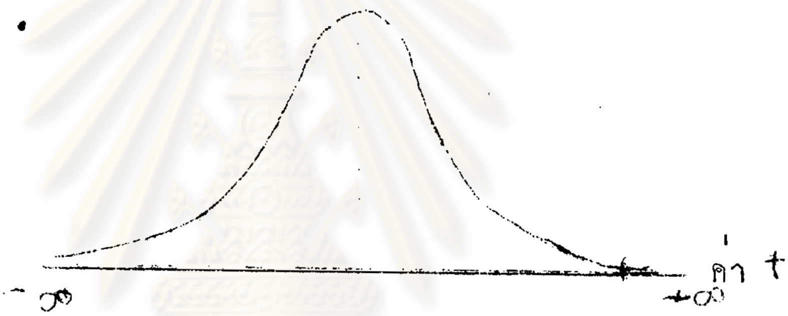
[6] William G. Cochran and Gertrude M. Cox, Experinental Design. (New York : Wiley & Sons Inc., 1960), p. 101.

[7] E.S. Keeping, Introduction to Statistical Inference. (Princeton : Dvan Nostrand Company, Inc., 1962), p. 181.

ถ้า X_k (k = 1, 2,---n) เป็นค่าของตัวแปรสุ่มอิสระมาจากประชากร ที่มีการแจกแจงปกติ โดยมีมัชฌิมเลขคณิตเท่ากับ μ และความแปรปรวน σ^2 ถ้า $\bar{X} = \frac{1}{n}\cdot \sum_{k=1}^{n} X_k$ เป็นมัชฌิมเลขคณิตของกลุ่มตัวอย่าง $S^2 = \frac{1}{n-1}\cdot \sum_{k=1}^{n} (X_k-\bar{X})^2$ เป็นความแปรปรวนของกลุ่มตัวอย่าง อัตราส่วน

$$t_i = \frac{\bar{X}_1 - \mu}{S_i/\sqrt{n_i}}$$

รูปที่ 1 การแจกแจงที



จะมีการแจกแจงทีโดยมีขั้นความเป็นอิสระ ν เท่ากับ n - 1[8]

คุณสมบัติทางคณิตศาสตร์ของการแจกแจงที

การแจกแจงที เมื่อมีขั้นความเป็นอิสระ ν = n-1 มีฟังก์ชันเดนซิตี้ (Density Function) ดังนี้

$$G(t) = \frac{1}{\sqrt{n-1}\;\beta\left(\frac{1}{2};\ \frac{1}{2}(n-1)\right)} \cdot \frac{1}{\sqrt{\left(1+t^2/n-1\right)^n}}, \quad -\infty < t < \infty$$ [9]

เมื่อ n = ขนาดของกลุ่มตัวอย่าง

---

[8] Marek Fisz, Theory and Mathematical Statistics (New York: John Wiley & Sons, 1967), p. 348.

[9] ดูภาคผนวก ก.

ν = n-1 คือ ชั้นความเป็นอิสระ[10]

β =เบ้ตาฟังก์ชัน[11] (Beta Function)

t =ตัวแปรสุ่ม t

จากสูตรฟังก์ชันเดนซิตี้ (Density Function) จะเห็นว่า การแจกแจงทีเป็นอิสระจากส่วนเบี่ยงเบนมาตรฐาน (σ) เพราะการแจกแจงที่ประกอบด้วยพารามีเตอร์เพียงตัวเดียวคือชั้นความเป็นอิสระ (ν= n-1) ดังนั้นการแจกแจงที่จะขึ้นอยู่กับชั้นความเป็นอิสระ ถ้าชั้นความเป็นอิสระมีค่ามากหรือเข้าใกล้อนันต์ (∞) ค่าประมาณความแปรปรวนของประชากร (s^2) จะเท่ากับความแปรปรวนของประชากร (σ^2) การแจกแจงที่ก็จะเข้าใกล้การแจกแจงปกติ[12] และเนื่องจากค่า t ในฟังก์ชันเดนซิตี้อยู่ในลักษณะกำลังสอง สัดส่วนทางด้านบวกของการแจกแจงทีมีค่าเท่ากับสัดส่วนทางด้านลบ ทำให้การแจกแจงทีมีลักษณะสมมาตร (Symmetry) ที่จุด t = 0 ดังนั้นโมเมนต์อันดับเลขคี่ (Odd-order Moment) มีค่าเท่ากับ 0 และโมเมนต์อันดับเลขคู่ (Even-Order Moment) เมื่อจำนวนอันดับมีค่าน้อยกว่าชั้นความเป็นอิสระจะคำนวณได้จาก

$$\mu_{2r} = \frac{(n-1)^r(2r-1)(2r-3)---1}{(n-3)(n-5)---(n-1-2r)} \quad {}^{23}$$

เมื่อ μ_{2r} คือโมเมนต์อันดับเลขคู่ โดยที่ r = 1,2,3--- และค่า 2r จะต้องน้อยกว่า n-1 เมื่อ n คือขนาดของกลุ่มตัวอย่าง, n-1 คือชั้นความเป็นอิสระ

---

[10]ดูภาคผนวก ช.

[11]ดูภาคผนวก ฌ.

[12]ดูภาคผนวก ค.

[13]Harold Cramer, Methematical of Statistics (Princeton: Princeton University Press, 1946), p. 239.

อัตราส่วน t ของการแจกแจงทีเป็น Non Linear Transformation[14] เพราะว่าค่า $\frac{X}{S/\sqrt{n}} - \frac{\mu}{S/\sqrt{n}}$ ไม่คงที่ เนื่องจากค่า $S/\sqrt{n}$ เปลี่ยนแปลงไปตามขนาดของกลุ่มตัวอย่าง และนอกจากนี้อัตราส่วน t เป็นตัวแปรแบบต่อเนื่อง (Continuous Variable) สามารถกำหนดค่าได้ทุกค่าในช่วงใดช่วงหนึ่ง ดังนั้นความน่าจะเป็นของ t ในช่วง (t_1, t_2) คำนวณได้จาก

$$P_r(t_1 < t < t_2) = \frac{1}{\sqrt{n-1}\; B\left(\frac{1}{2}, \frac{1}{2}(n-1)\right)} \cdot \int_{t_1}^{t_2} \frac{dt}{\sqrt{\left(1+t^2/n-1\right)^n}}$$

คุณสมบัติทางสถิติของการแจกแจงที

โค้งการแจกแจงทีมีลักษณะสมมาตรและเป็นรูประฆัง (Bell-Sahped) ที่มีศูนย์กลาง (Center) อยู่ที่ t = 0 ดังนั้นมัชฌิมเลขคณิต (mean) มัธยฐาน (median) และฐานนิยม (Mode) ของการแจกแจงทีจึงมีค่าเท่ากันคือ o[15] ค่าควอไทล์ (Quatile) ที่ 1 และควอไทล์ที่ 3 ของการแจกแจงทีจะห่างจากมัธยฐานเป็นระยะห่างเท่ากันด้วย ความน่าจะเป็นสะสมหรือพื้นที่ใต้โค้งการแจกแจงทีทั้งหมดเท่ากับ 1 การแจกแจงทีมีความแปรปรวนเท่ากับ $\nu/\nu-2$ เมื่อ ν มีค่าน้อยกว่า 2 การแจกแจงจะไม่มีความแปรปรวน ดังนั้นความแปรปรวนจะมีพิสัยตั้งแต่ 1 ถึง 3 เสมอ[16] เพราะความแปรปรวนนี้จะขึ้นอยู่กับชั้นความเป็นอิสระ (Degree of Freedom) เมื่อชั้นความเป็น

---

[14] Sheldon Lorey, *Inferential Statistic in Behavioral Science* (New York : Holt, Rinehart and Winston, Inc., C1968), p.92.

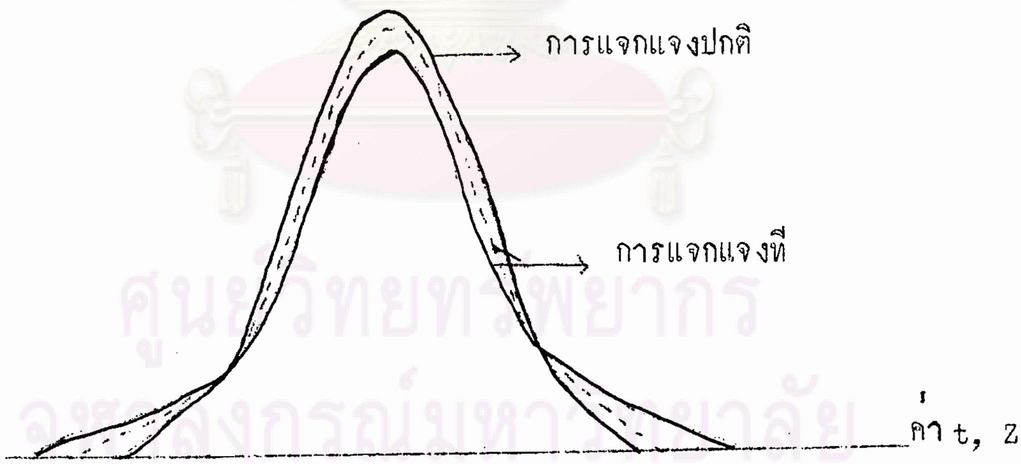
[15] Samuel B. Richmond, *Statistical Analysis*. (New York : The Ronald Press Company, 1964), p. 187.

[16] Daniel E. Bailey, *Probability and Statistics Models for Research* (New York : John Wiley & Sons, Inc., 1971), pp.293-295.

เป็นอิสระ $\vee = 3$ ความแปรปรวนมีค่าเท่ากับ 3 และถ้าขั้นความเป็นอิสระมากหรือเข้าใกล้ค่าอนันต์ (∞) ความแปรปรวนของการแจกแจงทีจะลดลงจนเข้าใกล้ 1 เป็นต้น

เนื่องจากการแจกแจงทีมีลักษณะสมมาตร จึงไม่มีความเบ้ (Skewed) ดังนั้นดัชนีความเบ้หรือโมเมนต์อันดับที่ 3 ของการแจกแจงทีจึงเป็น 0 ซึ่งเหมือนกับการแจกแจงปกติ แต่ดัชนีความโด่ง (Kurtosis) ของการแจกทีมีค่ามากกว่า 3 นั่นคือมีค่ามากกว่าดัชนีความโด่งของการแจกแจงปกติ ดังนั้นการแจกแจงทีจะมียอดแหลมกว่าการแจกแจงปกติ ซึ่งคีปปิง (E.S. Keeping) เรียกลักษณะการแจกแจงชนิดนี้ว่าเลปโตเคอร์ติก[17] (Leptokurtic) แต่เลวิส (Edward E. Lewis) เรียกลักษณะการแจกแจงชนิดนี้ว่าแพลทตีเคอร์ติก[18] (Platykurtic) เพราะโค้งการแจกแจงทีต่ำกว่าโค้งการแจกแจงปกติ และแม้ว่ายอดของการแจกแจงทีจะแหลมกว่ายอดของการแจกแจงปกติ แต่ปลายทั้งสองข้างของโค้งการแจกแจงทีจะสูงกว่าปลายของโค้งการแจกแจงปกติ ดังรูปที่ 2 ดังนั้นเมื่อกำหนดค่าใด ๆ ของ t และ Z พื้นที่ใต้โค้งการแจกแจงทีจะมากกว่าพื้นที่ใต้โค้งการแจกแจงปกติ

รูปที่ 2 การแจกแจงทีและการแจกแจงปกติเปรียบเทียบกัน



---

17 Keeping, op.cit; P. 180

18 Edward E. Lewis, Method of Analysis in Economics and Business (Boston : Haughton Miffin Company; 1963) P. 229

ความสูงหรือออดิเนต (Ordinate) ของโค้งการแจกแจงทีเมื่อกำหนดค่าทีและขั้นความเป็นอิสระ (ν) จะเท่ากับ

$$Y = \frac{\Gamma\frac{n}{2}}{\Gamma\left(\frac{n-1}{2}\right)\sqrt{2\pi}} \cdot (1+t^2/n-1)^{-n/2}$$ [19]

เมื่อ Γ คือ แกมมาฟังก์ชัน[20] (Gamma Function)

n คือ ขนาดของกลุ่มตัวอย่าง

$n-1$ คือ ขั้นความเป็นอิสระ (ν)

ตารางที่ 1 ค่าออดิเนตของการแจกแจงที[21]

	ขั้นความเป็นอิสระ (ν)			
t	3	15	29	
0	.368	.392	.396	.399
± 0.5	.313	.344	.348	.352
± 1.0	.207	.234	.238	.242
± 1.5	.120	.128	.129	.130
± 2.0	.068	.059	.058	.054
± 2.5	.039	.024	.021	.018
± 3.0	.023	.009	.007	.004
± 3.5	.014	.003	.002	.001
± 4.0	.009	.001	.001	.000

---

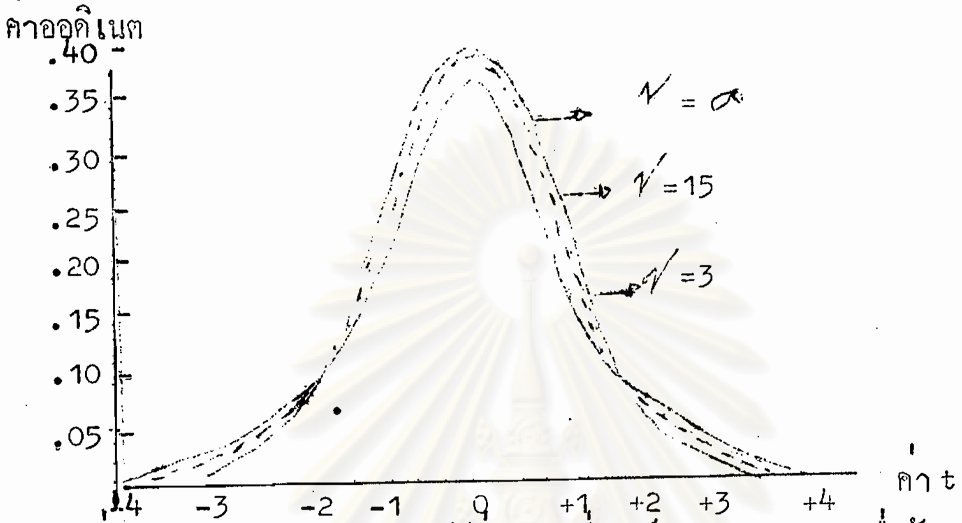
[19] Frederick E. Croxton and Dudley J. Cowden, Applied General Statistics (Newdelhi : Prentice Hall of India Private Limited, 1969), p,556.

[20] ดูภาคผนวก ฌ.

[21] Paul Blommers and E.F. Lindquist, Elementary Statistical Methods. (Calcutta : Oxford Book Company, C 1960), p. 339.

จากค่าออดิเนตในตารางที่ 1 อาจแสดงได้ด้วยรูปกราฟ ดังนี้

รูปที่ 3 การแจกแจงทีที่ขั้นความมีอิสระ ∞ , 15 และ 3



ค่าออดิเนตของการแจกแจงทีมีค่ามากที่สุดเป็น 1[22] เพราะเมื่อขั้นความเป็นอิสระเข้าใกล้ค่าอนันต์ (∞) โค้งการแจกแจงทีจะเข้าใกล้การแจกแจงปกติมาตรฐาน ซึ่งมีมัชฌิมเลขคณิตเท่ากับศูนย์ และความแปรปรวนเท่ากับหนึ่ง[23] หรืออาจกล่าวได้ว่า เมื่อขั้นความเป็นอิสระมาก ค่า t จะใกล้เคียงกับค่า Z ดังนั้นการแจกแจงทีจะมีพิสัย (Range) ตั้งแต่ $-\infty$ ถึง $+\infty$

ความสำคัญของการแจกแจงที

เนื่องจากการใช้การแจกแจงปกติมาตรฐานทดสอบสมมุติฐานเกี่ยวกับมัชฌิมเลขคณิตของประชากร (μ) นั้นจำเป็นจะต้องทราบค่าความแปรปรวนของประชากร (σ^2) ก่อนจึงจะใช้สถิติ $Z = (\bar{X} - \mu) / \sigma/\sqrt{n}$ และในกรณีที่ไม่ทราบค่าความแปรปรวน

---

[22] Croxton and Cowden, Loc.cit.

[23] ดูภาคผนวก ค.

ของประชากร แต่กลุ่มตัวอย่างมีขนาดใหญ่ก็สามารถใช้ความแปรปรวนของกลุ่มตัวอย่าง (s^2) ประมาณค่าความแปรปรวนของประชากรได้[24] จึงยังสามารถทดสอบด้วยสถิติ z ได้ด้วย

ในทางภาคปฏิบัติในงานวิจัยโดยทั่วไปแล้ว เรามักจะไม่ทราบค่าส่วนเบี่ยงเบนมาตรฐานของประชากร (σ) และกลุ่มตัวอย่างที่ใช้ก็มีขนาดเล็กจึงต้องประมาณค่าส่วนเบี่ยงเบนมาตรฐานของประชากร (σ) ค่าประมาณที่ใช้คือส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง (S) ซึ่งในกรณีเช่นนี้ กอสเซท (Gosset) ได้พิสูจน์ว่าจะใช้การแจกแจงปกติไม่ได้ เพราะข้อสรุปที่ได้มาจะเชื่อถือไม่ได้ ต้องใช้การแจกแจงทีจึงจะได้ข้อสรุปที่น่าเชื่อถือ เมื่อกลุ่มตัวอย่างมีขนาดเล็ก[25] ดังนั้นการแจกแจงทีจึงช่วยลดข้อจำกัดของการใช้การแจกแจงปกติมาตรฐาน (Standard Normal Distribution) ซึ่งต้องอาศัยค่าส่วนเบี่ยงเบนมาตรฐานของประชากร (σ) ทำให้การแจกแจงทีสามารถนำไปใช้ได้อย่างกว้างขวางในวิชาสถิติสรุปอ้างอิง (Inference Statistics) และนอกจากนี้การแจกแจงทีก็นำไปใช้กับกลุ่มตัวอย่างขนาดใหญ่ได้[26] เพราะการแจกแจงทีจะเข้าใกล้การแจกแจงปกติมาตรฐานที่มีมัชฌิมเลขคณิต $\mu = 0$ ความแปรปรวน $= 1$ เมื่อขั้นความเป็นอิสระเข้าใกล้ค่าอนันต์ (∞)[27]

---

[24]Hurbert M. Blalock, Social Statistics. (New York : McGraw-Hill Book Company Inc., 1960), pp. 147 - 149.

[25]Lawrence J. Kaplan, Elementary Statistics for Economics and Business. (New York : Pitman Publishing Cooperation, C 1966), p.128.

[26]Everett Vernon Lewis, Statistical Analysis. (New York : Dvan Nostrand Company, Inc., 1963), p. 183.

[27]William Mendenhall, Introduction to Probability and Statistics (2nd ed., Belmont, California : Wadsworth Publishing Company, Inc., 1969), p.191.

ข้อตกลงเบื้องต้นโดยทั่วไปในการใช้การแจกแจงที

การใช้การแจกแจงทีจะต้องมีข้อตกลงเบื้องต้นดังนี้

1. กลุ่มตัวอย่างจะต้องสุ่มมาจากประชากรที่มีการแจกแจงปกติซึ่งมีค่ามัชฌิมเลขคณิต μ และความแปรปรวน σ^2 จึงจะทำให้การกระจายของค่าที่เป็นไปได้ทุก ๆ ค่า (All Possible Values) ของมัชฌิมเลขคณิตของกลุ่มตัวอย่าง ($\bar{X}_i$) มีการแจกแจงปกติด้วย นอกจากนี้ยังมีผลให้มัชฌิมเลขคณิตของกลุ่มตัวอย่าง ($\bar{X}_i$) กับความแปรปรวนของกลุ่มตัวอย่าง (s_i^2) ต่างก็เป็นอิสระ

2. ส่วนเบี่ยงเบนมาตรฐานของประชากรจะได้จากการประมาณค่าส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง (s_i) ซึ่งเป็นค่าไม่คงที่ เปลี่ยนแปลงไปตามกลุ่มตัวอย่าง เมื่อกลุ่มตัวอย่างมีขนาดใหญ่ ค่าเหล่านี้จะเข้าใกล้ค่าคงที่

3. กลุ่มตัวอย่างมีขนาดเล็กตามทฤษฎี Central Limit Theorem อธิบายว่า ไม่ว่าการแจกแจงตัวอย่างตามทฤษฎีจะเป็นปกติหรือไม่ ถ้ากลุ่มตัวอย่างมีขนาดใหญ่ การแจกแจงตัวอย่างการสุ่ม (Sampling Distribution) ของค่าสถิติจะเป็นปกติหรือเข้าใกล้ปกติ จึงเกิดปัญหาว่ากลุ่มตัวอย่างขนาดเท่าใดจึงจะถือว่าเป็นกลุ่มตัวอย่างขนาดใหญ่พอที่มีการแจกแจงปกติได้ นักสถิติส่วนมากให้ความเห็นว่ากลุ่มตัวอย่างควรมีขนาดมากกว่า 30[28] โรเบอร์ต กับ วอลลิส (Roberts and Wallis) ใช้เมื่อขนาดของกลุ่มตัวอย่างมากกว่า 20[29] และ วอลเกอร์ กับ เลฟ (Walker and Lev) เสนอว่าขนาดของกลุ่มตัวอย่างควรจะมากกว่า 40[30] เฮย์ (Hays) ให้ใช้เมื่อขนาดของกลุ่ม

---

[28] Dick A. Leabo, Basic Statistics, (3[nd] ed., Illinois : Richard D-Irwin, Inc., 1973), p. 190.

[29] Henry V. Roberts and Allen W. Wallis, Statistic a New Approach, (New York : The Free Press of Glencoe, 1964), p.418.

[30] Helen M. Walker and Joseph Lev, Elementary Statistical Methods (New York : Holt, Rinehart and Winston, C.1958).

ตัวอย่างมากกว่า 100 ขึ้นไป[31] ปีเตอร์ กับ ซัมเมอร์ (Peters and Summers) แนะนำว่าถ้าไม่ทราบการแจกแจงของประชากรต้องใช้เมื่อขนาดของกลุ่มตัวอย่างมากกว่า 100 ขึ้นไปเสมอ[32]

การใช้การแจกแจงทีในงานวิจัยการศึกษา.

ในงานวิจัยทางการศึกษา ได้นำการแจกแจงทีมาใช้ในกรณีต่อไปนี้

1. การประมาณค่าพารามีเตอร์ภายใต้ความเชื่อมั่นที่กำหนด

วัตถุประสงค์ในการสุ่มตัวอย่างก็เพื่อที่จะนำค่าของกลุ่มตัวอย่างหรือค่าสถิติ (Statistic) ที่ได้ไปประมาณค่าของประชากร (Parameter) หรือนำไปทดสอบสมมุติฐานบางประการเกี่ยวกับลักษณะของประชากรที่เราตั้งไว้ หรืออาจกระทำทั้งสองอย่างพร้อมกัน ในการประมาณค่าพารามีเตอร์ภายใต้ความเชื่อมั่นที่กำหนดนั้น สามารถใช้ประโยชน์จากการแจกแจงทีช่วยประมาณค่ามัชฌิมเลขคณิตของประชากร (μ) ความแตกต่างของมัชฌิมเลขคณิตของประชากร 2 กลุ่ม ($\mu_1-\mu_2$) สัมประสิทธิ์ความถดถอยของประชากร (β) ความแตกต่างระหว่างความถดถอยของประชากร 2 กลุ่ม ($\beta_1 - \beta_2$) และประมาณค่าความถดถอยเชิงเส้นตรงของประชากร (Linear Regression) ซึ่งจะได้กล่าวแต่ละเรื่องดังต่อไปนี้

1.1 การประมาณค่ามัชฌิมเลขคณิตของประชากร (μ)

ค่าที่ใช้ประมาณมัชฌิมเลขคณิตของประชากร (μ) ได้ดีที่สุด (The Best Estimate) คือ มัชฌิมเลขคณิตของกลุ่มตัวอย่าง ($\bar{X}_i$) [ค่า $\bar{X}_i$ นั้น

---

[31] William L. Hays, Statistical for Psychologist (New York: Holt, Rinehart and Winston, 1965).

[32] William S. Peters and George W. Summers, Statistical Analysis for Business Decisions (Englewood Cliffs, Newjersey : Prentice Hall, Inc., 1965), pp. 161 - 162.

คำนวณได้จาก $\sum_{i=1}^{n} X_i/n, (i = 1, 2, --- n)$, X_i คือ ค่าในมาตราอันตรภาค (Interval Scale) และ n คือขนาดของกลุ่มตัวอย่าง] แต่เมื่อมีกลุ่มตัวอย่างหลายกลุ่มจะได้ค่าประมาณ μ หลายค่า จึงมีคำถามว่าค่าใดจึงเป็นค่าประมาณ μ ที่ดีที่สุด นักสถิติจึงหันมาพิจารณาประมาณ μ ในรูปของอันตรภาคแห่งความเชื่อมั่น (Confidence Interval Estimate) ที่กำหนด เช่นค่า μ เป็นค่าหนึ่งอยู่ระหว่างค่า a และ b ภายใต้ความเชื่อมั่น $(1 - \alpha)$ 100 % โดยที่ค่า a เป็นค่าที่ต่ำสุด และ b เป็นค่าสูงสุด ค่า μ จะมีเพียงค่าเดียวที่อยู่ในช่วง a และ b ซึ่งอาจเขียนได้ดังนี้

$$\text{Conf } (a \leq \mu \leq b) = (1 - \alpha)$$

ซึ่งสมการข้างบนนี้หมายความว่า ความน่าจะเป็นที่ μ จะอยู่ในอันตรภาค a และ b เป็น $1 - \alpha$ นั่นคือมีโอกาส $(1 - \alpha)$ 100 จากการทดลอง 100 ครั้งที่ μ จะอยู่ระหว่าง a ถึง b

แต่ในทางปฏิบัติ เราไม่ได้สุ่มตัวอย่างมาหลาย ๆ กลุ่ม เราสุ่มมาเพียงกลุ่มเดียวแล้วคำนวณ $\bar{X}_i$ ค่า $\bar{X}_i$ ที่คำนวณได้จากกลุ่มตัวอย่างเดียวนี้ย่อมเป็นค่าคงที่ไม่ใช่ตัวแปร เมื่อนำมาสร้างช่วงหรืออันตรภาคแห่งความเชื่อมั่น อันตรภาคที่ได้ก็เป็นอันตรภาคที่คงที่ จึงอาจจะรวมหรือไม่รวม μ ไว้ก็ได้

ในการประมาณค่า μ ในอันตรภาคที่กำหนดนั้นอาศัยการกระจายทางทฤษฎีของค่ามัชฌิมเลขคณิตของกลุ่มตัวอย่างทุกกลุ่มที่จะเป็นไปได้ (Sampling Distribution of Means) เนื่องจากการกระจายทางทฤษฎีของค่ามัชฌิมเลขคณิตเป็นโค้งปกติที่มีค่าคาดหวัง (Expected Value) เป็น $E(\bar{X}) = \mu$ และส่วนเบี่ยงเบนมาตรฐานที่เรียกว่าความคลาดเคลื่อนมาตรฐาน (Standard Error) นั้นเท่ากับ $\sigma/\sqrt{n}$ กรณีสุ่มตัวอย่างแบบมีการแทนที่ (Sampling with Replacement) จากประชากรขนาดจำกัดหรือสุ่มตัวอย่างจากประชากรขนาดไม่จำกัดและเท่ากับ $\frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$ กรณีสุ่มตัวอย่างแบบไม่แทนที่ (Sampling without Replacement) จากประชากรขนาดจำกัด [N คือ ขนาดของประชากร, $\sqrt{\frac{N-n}{N-1}}$ คือค่าแก้สำหรับประชากรขนาดจำกัด (Finite Population Correction หรือ fpc)] ถ้าขนาดของ n เมื่อเทียบกับ N เล็ก

(เล็กกว่า 5 %) จะทำให้ค่า $\frac{N-n}{N-1}$ มีค่าเข้าใกล้ 1 ก็ตัด fpc ทิ้งได้

ดังได้กล่าวมาแล้วข้างต้นว่า ในทางภาคปฏิบัติ ผู้วิจัยมักจะไม่ทราบค่าความแปรปรวนของประชากร (σ^2) การแจกแจงทีได้เข้ามามีบทบาทและเป็นประโยชน์ในการประมาณค่าพารามีเตอร์เพราะการแจกแจงที่ยอมให้ใช้ความแปรปรวนของกลุ่มตัวอย่าง (s^2) ประมาณค่า σ^2 ได้ ดังนั้นการประมาณค่ามัชฌิมเลขคณิตของประชากร (μ) ภายใต้ความเชื่อมั่นที่กำหนดจึงทำได้โดยคำนวณความแปรปรวนของกลุ่มตัวอย่าง (s^2) ดังนี้คือ

$$\text{Conf } (t_a \leq t \leq t_b) = (1 - \alpha)$$

โดยที่ $t = \frac{\bar{X} - \mu}{S/\sqrt{n}}$ เมื่อมีการสุ่มตัวอย่างแบบแทนที่

$t = \frac{\bar{X} - \mu}{S/\sqrt{n} \cdot \sqrt{\frac{N-n}{N-1}}}$ เมื่อมีการสุ่มตัวอย่างแบบไม่แทนที่

t_a = ค่าต่ำสุดของอันตรภาคความเชื่อมั่นที่กำหนด

t_b = ค่าสูงสุดของอันตรภาคความเชื่อมั่นที่กำหนด

ดังนั้นสมการข้างบนหมายความว่า ที่ระดับความเชื่อมั่น ($1 - \alpha$) ค่า t จะอยู่ระหว่าง t_a และ t_b

แทนค่า t ในสมการ จะได้

$$\text{Conf } \left(t_a < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_b\right) = (1 - \alpha)$$

นั่นคือ $\text{Conf } \left(\bar{X} - \frac{S}{\sqrt{n}} \cdot t_a < \mu < \bar{X} + \frac{S}{\sqrt{n}} \cdot t_b\right) = 1 - \alpha$

ซึ่งอันตรภาคแห่งความเชื่อมั่นที่กำหนดของมัชฌิมเลขคณิตที่ประมาณได้นี้มีความหมายว่าถ้าสุ่มตัวอย่างขนาด n มา 100 กลุ่ม จะมีโอกาสร้อยละ ($1 - \alpha$) 100 ที่มัชฌิมเลขคณิตของประชากรจะมีค่าอยู่ระหว่าง $\bar{X} - \frac{S}{\sqrt{n}} \cdot t_a$ กับ $\bar{X} + \frac{S}{\sqrt{n}} \cdot t_b$ โดยที่

μ คือ มัชฌิมเลขคณิตของประชากร

$\bar{X}$ คือ มัชฌิมเลขคณิตของกลุ่มตัวอย่าง

S คือ ส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง

n คือ ขนาดของกลุ่มตัวอย่าง

$1-\alpha$ คือ ระดับความเชื่อมั่นหรือสัมประสิทธิ์ความเชื่อมั่น

ตัวอย่าง สมมุติว่าต้องการทราบอายุเฉลี่ยของคู่สมรสชาวไทยภายใต้ความเชื่อมั่น 95 % โดยสุ่มตัวอย่างชาวไทยมา 242 คน ปรากฏว่าอายุเฉลี่ยของคู่สมรสเป็น 28.75 ปี และส่วนเบี่ยงเบนมาตรฐานเป็น 6.27[33]

ถ้าถือว่าการแจกแจงอายุของคู่สมรสเป็นโค้งปกติ แต่ไม่ทราบค่าส่วนเบี่ยงเบนมาตรฐานของประชากร จึงใช้การแจกแจงที ประมาณค่าอายุเฉลี่ยของคู่สมรสได้

กำหนดระดับความเชื่อมั่น เท่ากับ .95 ดังนั้นค่า $t_{\alpha/2}$ จากตารางทีเป็น ± 1.96 ขั้นความเป็นอิสระ 241

เนื่องจากขนาดของกลุ่มตัวอย่างใหญ่ (n = 242) จึงอาจประมาณค่าส่วนเบี่ยงเบนมาตรฐานของประชากร (σ) ด้วยส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง (S) ได้ ดังนั้น

$$\text{Conf}(\bar{X} - \sigma/\sqrt{n} \cdot t_a < \mu < \bar{X} + \sigma/\sqrt{n} \cdot t_b) = 95\ \%$$

หรือ $$\text{Conf}\ (\bar{X} - S/\sqrt{n} \cdot t_a < \mu < \bar{X} + S/\sqrt{n} \cdot t_b) = 95\ \%$$

นั่นคือ $$\text{Conf}\ (27.97 < \mu < 29.53) = 95\ \%$$

ซึ่งหมายความว่ามีโอกาสร้อยละ 95 ที่อายุเฉลี่ยของคู่สมรสจะมีค่าอยู่ระหว่าง 27.97 ถึง 29.53

---

[33] A.L.O' Toole, Elementary Practical Statistics (New York: The Mcmillan Company, C 1964), p. 219.

1.2 การประมาณค่าความแตกต่างของมัชฌิมเลขคณิตของประชากร 2 กลุ่ม $(\mu_1 - \mu_2)$

ค่าที่ประมาณความแตกต่างระหว่างมัชฌิมเลขคณิตของประชากร 2 กลุ่มคือ ความแตกต่างของมัชฌิมเลขคณิตของกลุ่มตัวอย่าง 2 กลุ่ม $(\bar{X}_1 - \bar{X}_2)$ เมื่อ $\bar{X}_1$, $\bar{X}_2$ คือ มัชฌิมเลขคณิตของกลุ่มตัวอย่างขนาด n_1 , n_2 ตามลำดับ เมื่อมีกลุ่มตัวอย่างหลายกลุ่มจะได้ค่าความแตกต่างหลายค่า แล้วแต่ค่าของ $\bar{X}_1$ และ $\bar{X}_2$ ดังนั้นค่าประมาณของ $\mu_1 - \mu_2$ จึงมีหลายค่า นักสถิตินิยมประมาณค่า $\mu_1 - \mu_2$ ให้อยู่ในช่วงหรืออันตรภาคภายใต้ความเชื่อมั่นที่กำหนด เช่น ค่า $\mu_1 - \mu_2$ อยู่ระหว่างค่า a และ b โดยที่ค่า a เป็นค่าต่ำสุด และ b เป็นค่าสูงสุด ค่า $\mu_1 - \mu_2$ จะมีเพียงค่าเดียวที่อยู่ภายในช่วง a และ b ซึ่งอาจเขียนได้ดังนี้

$$\text{Conf } (a < \mu_1 - \mu_2 < b) = (1 - \alpha)$$

ซึ่งหมายความว่ามีโอกาสร้อยละ $(1-\alpha)100$ ที่ความแตกต่างของมัชฌิมเลขคณิตของประชากร 2 กลุ่ม จะมีค่าอยู่ระหว่าง a ถึง b

ในการประมาณค่าความแตกต่างของมัชฌิมเลขคณิตของประชากร 2 กลุ่ม $(\mu_1 - \mu_2)$ ภายใต้ความเชื่อมั่นที่กำหนดนั้นอาศัยการแจกแจงการสุ่มตัวอย่างของค่าความแตกต่างของมัชฌิมเลขคณิตของกลุ่มตัวอย่างทั้งหมดที่จะเป็นไปได้ (Sampling Distribution of Difference Between Sample Means) เนื่องจากการแจกแจงการสุ่มตัวอย่างของความแตกต่างของมัชฌิมเลขคณิตเป็นโค้งปกติ เมื่อประชากรแจกแจงปกติ ดังนั้นการแจกแจงความแตกต่างระหว่างมัชฌิมเลขคณิตของกลุ่มตัวอย่าง $(\bar{X}_1 - \bar{X}_2)$ จะเป็นโค้งปกติด้วย โดยมีมัชฌิมเลขคณิต $\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$ และความคลาดเคลื่อนมาตรฐาน (Standard Error) $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ เมื่อกลุ่มตัวอย่างขนาด n_1 และกลุ่มตัวอย่างขนาด n_2 ต่างก็สุ่มมาจากประชากรที่มีการแจกแจงปกติ ซึ่งทำให้กลุ่มตัวอย่างทั้งสองกลุ่มเป็นอิสระ

แต่ในทางปฏิบัติเราไม่ทราบค่าความแปรปรวนของประชากร (σ^2) ทั้งสองกลุ่ม ดังนั้นการแจกแจงที่จึงนำมาใช้ได้อย่างเหมาะสม โดยใช้ความแปรปรวนของกลุ่มตัวอย่าง (s^2) แทน σ^2 ดังนั้นอันตรภาคของการประมาณค่าความ

แตกต่างของมัชฌิมเลขคณิตของประชากร ($\mu_1 - \mu_2$) ภายใต้ความเชื่อมั่นที่กำหนดจะมี 2 กรณี คือ

กรณีที่ 1 เมื่อกลุ่มตัวอย่างทั้งสองเป็นอิสระ นั่นคือ กลุ่มตัวอย่างขนาด n_1 และกลุ่มตัวอย่างขนาด n_2 ต่างก็สุ่มมาจากประชากรที่แจกแจงปกติ จึงจะทำให้กลุ่มตัวอย่างทั้งสองกลุ่มเป็นอิสระ และถ้าความแปรปรวนของประชากร 2 กลุ่มเท่ากันหรือสมมุติได้ว่าเท่ากัน อันตรภาคของการประมาณค่าความแตกต่างของมัชฌิมเลขคณิตของประชากร ($\mu_1 - \mu_2$) ภายใต้ความเชื่อมั่นที่กำหนด คือ

$$\text{Conf } (t_a < t < t_b) = (1-\alpha) \qquad (1)$$

เมื่อ t_a คือ ค่าต่ำสุดของอันตรภาคความเชื่อมั่นที่กำหนด

t_b คือ ค่าสูงสุดของอันตรภาคความเชื่อมั่นที่กำหนด

t คือ $\dfrac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

S_P^2 คือ $\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

นั่นคือ แทนค่าใน (1) จะได้

$$\text{Conf } \left((\bar{X}_1-\bar{X}_2)-S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\cdot t_a < \mu_1-\mu_2 < (\bar{X}_1-\bar{X}_2)+S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\cdot t_b\right) = 1-\alpha$$

ซึ่งหมายความว่า ความน่าจะเป็นที่ $\mu_1 - \mu_2$ จะอยู่ในอันตรภาค $(\bar{X}_1-\bar{X}_2)$ $S_P \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \cdot t_a$ ถึง $(\bar{X}_1-\bar{X}_2)+S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \cdot t_b$ เป็น $1-\alpha$ ซึ่งแสดงว่ามีโอกาส $(1-\alpha)100$ จาก 100 ครั้งที่ $\mu_1 - \mu_2$ จะอยู่ระหว่าง $(\bar{X}_1-\bar{X}_2) - S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \cdot t_a$ ถึง $(\bar{X}_1-\bar{X}_2) + S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \cdot t_b$

ตัวอย่าง ในการศึกษาวิธีการสอนแบบTest Wisenessผู้วิจัยสุ่มตัวอย่างนักศึกษามหาวิทยาลัยมาจำนวนหนึ่ง แล้วแบ่งออกเป็น 2 กลุ่ม ในกลุ่มแรกได้รับการสอนแบบTest Wiseness กลุ่มที่สองได้รับการสอนแบบธรรมดา ภายหลังการทดลองสอน 3 วัน จึงทดสอบผลสัมฤทธิ์ทางการเรียนปรากฏว่า

	กลุ่มที่ 1	กลุ่มที่ 2
n	30	28
$\bar{X}$	63	42
S^2	100	200

จากช่วงแห่งความเชื่อมั่นของความแตกต่างระหว่างมัชฌิมเลขคณิตของประชากร ภายใต้ความเชื่อมั่น 95 %[34]

กำหนดความเชื่อมั่น 95 % ดังนั้นค่าทีจากตารางการแจกแจงทีเมื่อชั้นความเป็นอิสระ $\nu = 56$ คือ 2.00

แต่เนื่องจากไม่ทราบค่าส่วนเบี่ยงเบนมาตรฐานของประชากร (σ) ทั้งสองกลุ่ม จึงอาจประมาณค่าได้ด้วยส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง ดังนั้นช่วงแห่งความเชื่อมั่นของ $\mu_1 - \mu_2$ ที่ระดับ 95 % คือ

$$\text{conf}\left[(\bar{X}_i-\bar{X}_j)-S_P\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}\cdot t_a < \mu_1-\mu_2 < (\bar{X}_i-\bar{X}_j)+S_P\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}\cdot t_b\right] = 95\ \%$$

นั่นคือ $\text{Conf}\ (10.25 < \mu_1-\mu_2 < 31.75) = 95\ \%$

ซึ่งหมายความว่าค่าความน่าจะเป็นที่ $\mu_1 - \mu_2$ จะอยู่ในช่วง 10.25 ถึง 31.75 เป็น .95 นั่นคือมีโอกาส 95 จาก 100 ที่ $\mu_1 - \mu_2$ จะอยู่ในช่วง 10.25 ถึง 31.75

กรณีที่ 2 เมื่อกลุ่มตัวอย่างมีความสัมพันธ์กัน นั่นคือ กลุ่มตัวอย่างสองกลุ่มซึ่งมีขนาดเท่ากัน และสมาชิกของกลุ่มตัวอย่างหนึ่งมีลักษณะเป็นคู่กับสมาชิกของกลุ่มตัวอย่างที่สอง เช่น สมาชิก X_1 X_2 ---X_n และ สมาชิก Y_1, Y_2---Y_n มีความสัมพันธ์กัน คือ X_1 กับ Y_1, X_2 กับ Y_2, --- X_n กับ Y_n หรืออาจกล่าวได้ว่าตัวแปร X กับตัวแปร Y มีความสัมพันธ์กัน หรือ ในบางครั้งอาจมีการทดลองซ้ำกับกลุ่มตัวอย่างเดิม

---

[34] Malcolm J. Slakter, Statistical Inference for Educational Researchs. (Massachusett : Addison-Wesley Publishing Company, 1972), p. 266.

เราก็เรียกว่ากลุ่มตัวอย่างมีความสัมพันธ์กัน (Dependent Sample) เช่น มีการทดสอบ 2 ครั้ง กับกลุ่มตัวอย่างกลุ่มเดียวกัน

การประมาณค่าความแตกต่างของมัชฌิมเลขคณิตของประชากรจะอาศัยการแจกแจงที เพราะว่าเราไม่ทราบค่าความแปรปรวนของความแตกต่างระหว่างมัชฌิมเลขคณิตของประชากร จึงต้องใช้ความแปรปรวนของความแตกต่างของมัชฌิมเลขคณิตของกลุ่มตัวอย่าง โดยที่ $S_D^2 = \sqrt{\sum_{i=1}^{n} \frac{(D_i - \bar{D})}{n-1}/n}$ เมื่อใช้วิธีที่เรียกว่า Difference Method โดยที่ D คือ ความแตกต่างระหว่างสมาชิกที่เข้าคู่กัน n คือ ขนาดของกลุ่มตัวอย่าง หรือ $S_D^2 = S^2_{\bar{X}_i} + S^2_{\bar{X}_j} - 2r_{ij}S_{\bar{X}_i} \cdot S_{\bar{X}_j}$ โดยที่ r_{ij} คือ สหสัมพันธ์ระหว่างลักษณะที่เข้าคู่กันของกลุ่มตัวอย่างทั้งสอง ดังนั้นอันตรภาคแห่งความเชื่อมั่นของ $\mu_i - \mu_j$ ภายใต้ความเชื่อมั่นที่กำหนด คือ

$$\text{Conf}\left[(\bar{X}_i - \bar{X}_j) - S_D \cdot t_a < \mu_i - \mu_j < (\bar{X}_i - \bar{X}_j) + S_D \cdot t_b\right] = 1 - \alpha$$

ซึ่งหมายความว่า ความน่าจะเป็นที่ $\mu_i - \mu_j$ จะอยู่ในช่วง $(\bar{X}_i - \bar{X}_j) - S_D \cdot t_a$ กับ $(\bar{X}_i - \bar{X}_j) + S_D \cdot t_b$ เป็น $1 - \alpha$ ซึ่งแสดงว่า มีโอกาส $(1 - \alpha)$ 100 จาก 100 ครั้งที่ $\mu_i - \mu_j$ จะอยู่ในอันตรภาค $(\bar{X}_i - \bar{X}_j) - S_D \cdot t_a$ ถึง $(\bar{X}_i - \bar{X}_j) + S_D \cdot t_b$ นั่นแสดงว่าเมื่อสร้างอันตรภาคหรือช่วงแห่งความเชื่อมั่น เราอาจคาดหวังว่าจะมี $(1 - \alpha)100$ ช่วง จาก 100 ช่วงที่รวมเอาค่า $\mu_i - \mu_j$ ไว้

ตัวอย่าง การศึกษาทัศนคติของนิสิตครุศาสตร์ที่มีต่อการฝึกสอนนั้น ผู้วิจัยสุ่มตัวอย่างนิสิตชั้นปีที่ 3 ที่จะออกฝึกสอนในปีการศึกษาต่อไป จำนวน 9 คน ใช้แบบวัดทัศนคติวัดก่อนออกฝึกสอนครั้งหนึ่ง และภายหลังที่นิสิตจำนวนนี้ฝึกสอนแล้วก็วัดทัศนคติอีกครั้ง ปรากฏว่าความแตกต่างระหว่างมัชฌิมเลขคณิตของคะแนนทัศนคติทั้งสองครั้งเป็น 5 และความแปรปรวนของความแตกต่างระหว่างมัชฌิมเลขคณิตของคะแนนทัศนคติเป็น 6 จงหาช่วงความเชื่อมั่นของความแตกต่างระหว่าง

มัชฌิมเลขคณิต เมื่อกำหนดระดับความเชื่อมั่น 95%

กำหนดระดับความเชื่อมั่น 95 % ∴ ค่าที่จากตารางการแจกแจงทีเมื่อ $\nu = 7$ เท่ากับ 2.306 ดังนั้นช่วงแห่งความเชื่อมั่นของ $\mu_i - \mu_j$ ภายใต้ความเชื่อมั่น 95 % คือ

$$\text{Conf}\left[(\bar{X}_i-\bar{X}_j)-S_D \cdot t_a < \mu_i-\mu_j < (\bar{X}_i-\bar{X}_j)+S_D \cdot t_b\right] = 95\%$$

นั่นคือ $\text{Conf}\ (3.4 < \mu_i - \mu_j < 6.6) = 95\%$

ซึ่งหมายความว่าความน่าจะเป็นที่ความแตกต่างของมัชฌิมเลขคณิตของประชากร 2 กลุ่ม จะอยู่ในช่วง 3.4 ถึง 6.6 เป็น .95 นั่นคือ มีโอกาส 95 % ที่ $\mu_i - \mu_j$ จะอยู่ในช่วง 3.4 ถึง 6.6

1.3 การประมาณค่าความถดถอยเชิงเส้นตรงของประชากร (Linear Regression)

สมการถดถอย (Regression Equation) คือสมการที่แสดงความสัมพันธ์ของตัวแปร 2 ตัว ตัวแปรหนึ่งคือตัวแปรที่ไม่มีความคลาดเคลื่อน กำหนดค่าได้หรือวัดได้แน่นอน และมีอิทธิพลต่ออีกตัวแปรหนึ่งโดยทั่วไป มักเรียกว่าตัวแปรอิสระ (X) ส่วนอีกตัวแปรหนึ่งเป็นตัวแปรที่วัดค่าได้ภายหลังจากกำหนดหรือวัด X และจะเปลี่ยนแปลงกระทบกระเทือนเมื่อ X เปลี่ยนไป

ถ้าตัวแปรทั้ง 2 ตัวนี้มีความสัมพันธ์เชิงเส้นตรง (Linear) ลักษณะของความสัมพันธ์แทนได้ด้วยเส้นตรง (Straight Line) $Y_i = \alpha + \beta X_i + \epsilon_i$ โดยที่ α, β คือ พารามีเตอร์ของประชากร ϵ_i คือ ตัวแปรสุ่ม (Random Variable) ซึ่งมีมัชฌิมเลขคณิตเท่ากับศูนย์, X คือ ตัวแปรอิสระที่ทราบค่า, Y คือ ตัวแปรตามที่สามารถประมาณค่าได้จากสมการถดถอย ในกรณีที่กำหนดค่าตัวแปรอิสระ (X) ให้คงที่ (For a Fixed X) ตัวแปรตาม Y จะแจกแจงปกติ โดยมีมัชฌิมเลขคณิต $\mu_{Y.X} = \alpha + \beta X$ และความแปรปรวน $\sigma^2_{Y.X}$ ซึ่งสมการ $\mu_{Y.X} = \alpha + \beta X$ คือ เส้นความสัมพันธ์เฉลี่ยระหว่างตัวแปรอิสระ X กับตัวแปรตาม Y ของประชากร ซึ่งเรียกว่าเส้นถดถอยของ Y บน X (Line of Regression of Y on X)

โดยทั่วไป เรามีจุดมุ่งหมายจะประมาณค่าเส้นถดถอยของประชากร นั่นคือจะต้อง

ประมาณค่า α และ β โดยมีตัวประมาณค่าคือ a และ b ดังนั้นแต่ละค่าที่สังเกตได้จะแทนด้วยสมการ $Y_i = a + bX + ei$ โดยที่ $ei = Y_i - (a + bX_i)$ หรือ $Y_i - \hat{Y}$ และเนื่องจาก $\Sigma ei = 0$ จะได้สมการถดถอยของกลุ่มตัวอย่างที่ใช้ประมาณค่าสมการถดถอยของประชากร $\mu_{Y.X}$ คือ $\hat{Y} = a + bX$ ซึ่งจะมีหลายค่าหลายสมการขึ้นอยู่กับค่าตัวแปรอิสระ X เช่น X_0 ก็จะได้สมการถดถอย $\hat{Y} = a + bX_0$ ดังนั้นการพยากรณ์หรือการทำนายโดยใช้สมการถดถอย จึงมีความคลาดเคลื่อนเกิดขึ้นเสมอ เรียกว่าความคลาดเคลื่อนในการประมาณค่า (Error of Estimate) ซึ่งเท่ากับความแตกต่างระหว่างค่าที่สังเกต (Observed Value) และค่าพยากรณ์ (predicted Value) เมื่อความแตกต่างมีค่าน้อย แสดงว่าการพยากรณ์นั้นใกล้เคียงความจริง การวัดความถูกต้องของการประมาณค่า (Accuracy of Estimate) ก็คือการวัดความแปรปรวนของความคลาดเคลื่อนในการประมาณค่า[36] (Variance of the Errors of Estimate) ซึ่งเท่ากับ $\sigma^2_{Y.X}\left[\frac{1}{n} + \frac{(X_i-\bar{X})^2}{\Sigma(X-\bar{X})}\right]$ เมื่อต้องการพิจารณาว่า $\hat{Y}$ แตกต่างจาก E(Y) และเท่ากับ $\sigma^2_{Y.X}\left[1+\frac{1}{n} + \frac{(X_i-\bar{X})}{\Sigma(X-\bar{X})^2}\right]$ เมื่อต้องการพิจารณาว่า $\hat{Y}$ แตกต่างจาก Y หรือไม่

ในทางปฏิบัติ เราไม่ทราบค่า $\sigma^2_{Y.X}$ ดังนั้นจึงใช้การแจกแจงทีในการประมาณค่าสมการถดถอยเชิงเส้นตรงได้ การประมาณค่านี้นักสถิตินิยมพิจารณาเป็นช่วงหรืออันตรภาค เรียกว่าอันตรภาคพยากรณ์ (prediction Interval) ซึ่งเหมือนกับอันตรภาคความเชื่อมั่น การหาอันตรภาคพยากรณ์มี 2 วิธี คือ[37]

---

[36] Robert G.D.Steel and H. Ronie, Principles and Procedures of Statistics. (New York : McGraw-Hill Book Company, Inc., C1960), p. 169.

[37] R.L. Anderson and T.A. Bancroft, Statistic Theory in Research (New York : McGraw-Hill Book Company, 1952), pp.158-160.

1. อันตรภาคพยากรณ์ของ E(Y) ภายใต้ความเชื่อมั่นที่กำหนดคือ

$$\text{Conf}\left(\hat{Y}-t_a\cdot\sqrt{\frac{1}{n}+\frac{(X_i-\bar{X})^2}{\Sigma(X-\bar{X})^2}}\cdot S_{YX} < \mu_{Y.X} < \hat{Y}+t_b\cdot\sqrt{\frac{1}{n}+\frac{(X_i-\bar{X})^2}{\Sigma(X-\bar{X})^2}}\cdot S_{Y.X}\right) = 1-\alpha$$

2. อันตรภาคพยากรณ์ของ Y (Single Predicted Value) ภายใต้ความเชื่อมั่นที่กำหนดคือ

$$\text{Conf}\left(\hat{Y}-t_a\cdot\sqrt{1+\frac{1}{n}+\frac{(X_i-\bar{X})^2}{\Sigma(X-\bar{X})^2}}\cdot S_{Y.X} < Y < \hat{Y}+t_b\cdot\sqrt{1+\frac{1}{n}+\frac{(X_i-\bar{X})^2}{\Sigma(X-\bar{X})^2}}\cdot S_{Y.X}\right) = 1-\alpha$$

ซึ่งหมายความว่าความน่าจะเป็นที่ Y จะอยู่ในช่วง $Y-t_a\cdot\sqrt{1+\frac{1}{n}+\frac{(X_i-\bar{X})^2}{\Sigma(X-\bar{X})^2}}\cdot S_{Y.X}$ ถึง $Y+t_b\cdot\sqrt{1+\frac{1}{n}+\frac{(X_i-\bar{X})^2}{\Sigma(X-\bar{X})^2}}\cdot S_{Y.X}$ เป็น $(1-\alpha)$

ตัวอย่าง ในการศึกษาพัฒนาการของนักเรียน อาจารย์พลศึกษาพบว่า ความยาวของแขนและขนาดของรอบคอมีความสัมพันธ์เชิงเส้นตรงต่อกัน นั่นคือถ้าทราบความยาวของช่วงแขนก็จะพยากรณ์หรือทำนายขนาดของรอบคอได้ ดังนั้นเลือกกลุ่มตัวอย่างมา 20 คน ปรากฏว่าความยาวเฉลี่ยของแขนเป็น 33.15 นิ้ว ความแปรปรวน 1.0275 นิ้ว ถ้าความคลาดเคลื่อนในการประมาณค่าเท่ากับ 0.615 และความยาวของช่วงแขนเท่ากับ 31 นิ้ว ปรากฏว่าขนาดของรอบคอเท่ากับ 14.61 จงหาอันตรภาคความเชื่อมั่นของขนาดรอบคอของประชากรที่ระดับความเชื่อมั่น 95 %[38]

กำหนดระดับความเชื่อมั่น 95 % ดังนั้นค่า t จากตารางการแจกแจงทีเมื่อขั้นความเป็นอิสระ $\nu = 18$ คือ 2.101

$\therefore$ อันตรภาคความเชื่อมั่นของ $\mu_{Y.X}$ รอบ $\hat{Y}$ ที่ระดับความเชื่อมั่น 95 % คือ

---

[38] O'Tode, Op.cit., p. 292.

$$\text{Conf}\left(\hat{\bar{Y}}-t_a\cdot\sqrt{\frac{1}{n}+\frac{(X_i-\bar{X})^2}{\Sigma(X-\bar{X})^2}}\cdot S_{Y.X} < \mu_{Y.X} < \hat{\bar{Y}}+t_b\cdot\sqrt{\frac{1}{n}+\frac{(X_i-\bar{X})^2}{\Sigma(X-\bar{X})^2}}\cdot S_{Y.X}\right) = 95\%$$

หรือ $$\text{Conf}\left(\hat{\bar{Y}}-t_a\cdot\sqrt{\frac{1}{n}+\frac{(X_i-\bar{X})^2}{n\ S_X^2}}\cdot S_{Y.X} < \mu_{Y.X} < \hat{\bar{Y}}+t_b\cdot\sqrt{\frac{1}{n}+\frac{(X_i-\bar{X})^2}{n.S_X^2}}\cdot S_{YX}\right) = 95\%$$

นั่นคือ $$\text{Conf}\left(14.61\pm 2.101(0.615)\sqrt{\frac{1}{20}+\frac{(31-33.15)^2}{20(1.0275)}}\right) = (13.93,\ 15.29)$$

และเมื่อสุ่มตัวอย่างนักเรียนมาเพียงคนเดียว วัดความยาวของช่วงแขนได้ 31 นิ้ว การประมาณค่าความยาวของขนาดรอบคอก็ย่อมกระทำได้ นั่นคือ อันตรภาคความเชื่อมั่นของ Y รอบ $\hat{Y}$ ที่ระดับความเชื่อมั่น 95 % คือ

$$\text{Conf}\left(\hat{Y}-t_a\cdot\sqrt{1+\frac{1}{n}+\frac{(X_i-\bar{X})^2}{n\ S_X^2}}\cdot S_{Y.X} < Y < \hat{Y}+t_b\sqrt{1+\frac{1}{n}+\frac{(X_i-\bar{X})^2}{n\ S_X^2}}\cdot S_{Y.X}\right) = 95\%$$

นั่นคือ $$\text{Conf}\left(14.61\pm 2.101(0.615)\sqrt{1+\frac{1}{20}+\frac{(31-33.15)^2}{20(1.0275)}}\right) = (13.15, 16.07)$$

ซึ่งหมายความว่ามีโอกาสร้อยละ 95 ที่ขนาดรอบคอของประชากรอยู่ในช่วง 13.15 ถึง 16.07

การหาอันตรภาคพยากรณ์ทั้ง 2 วิธีนี้มีข้อที่น่าสังเกตคือ ถ้ามี $\hat{Y}$ เพียงค่าเดียว อันตรภาคพยากรณ์ที่ได้จะกว้างกว่าอันตรภาคพยากรณ์ที่ได้จาก $\hat{\bar{Y}}$ เสมอ เพราะว่า $\hat{Y}$ จะมีความคลาดเคลื่อน 2 ส่วนด้วยกันคือ 1) การเบี่ยงเบนของเส้นประมาณการถดถอยจากเส้นถดถอยที่แท้จริงของประชากร (population) และ 2) การเบี่ยงเบนของจุดตัวอย่าง (Sample Point, X_iY_i) จากเส้นถดถอยที่แท้จริง

1.4 การประมาณค่าสัมประสิทธิ์การถดถอยของประชากร (β)

สัมประสิทธิ์การถดถอยของประชากร (β) คือ อัตราการเปลี่ยนแปลงของตัวแปรตาม (Y) ต่อ 1 หน่วยการเปลี่ยนแปลงของตัวแปรอิสระ (X) หรือ อาจกล่าวได้ว่า β คือ ความชัน (Slope) ของเส้นถดถอยเชิงเส้นตรง

ตัวประมาณค่าที่ใช้ประมาณสัมประสิทธิ์การถดถอยของประชากร (β)

คือสัมประสิทธิ์การถดถอยของกลุ่มตัวอย่าง (b) ซึ่งคำนวณได้จาก $\frac{\Sigma XY-(\Sigma X \Sigma Y/N)}{\Sigma X^2 - [(\Sigma X^2)/N]}$

เมื่อมีกลุ่มตัวอย่างหลายกลุ่มจะได้ค่าสัมประสิทธิ์การถดถอยหลายค่า ดังนั้นค่าประมาณสัมประสิทธิ์การถดถอยจึงมีหลายค่า นักสถิติจึงนิยมประมาณค่า β ให้อยู่ในช่วงหรืออันตรภาคความเชื่อมั่นที่กำหนด เช่นค่า β อยู่ระหว่างค่า a และ b โดยที่ค่า a เป็นค่าต่ำสุด และ b เป็นค่าสูงสุด

ในการประมาณค่าสัมประสิทธิ์การถดถอยของประชากรในอันตรภาคที่กำหนดนั้นอาศัยการกระจายตัวอย่างการสุ่มของสัมประสิทธิ์การถดถอยของกลุ่มตัวอย่างทุกกลุ่มที่จะเป็นไปได้ (Sampling Distribution of Regression Coefficient) เนื่องจากการกระจายของสัมประสิทธิ์การถดถอยของประชากรเป็นโค้งปกติ เมื่อประชากรมีการแจกแจงปกติชนิดสองตัวแปร (Bivariate Normal Distribution)ดังนั้นสัมประสิทธิ์การถดถอยของกลุ่มตัวอย่างก็จะมีการแจกแจงเป็นโค้งปกติด้วย ซึ่งมีค่า $E(b) = \beta$ และความคลาดเคลื่อนมาตรฐาน $\sigma_b^2 = \frac{\sigma^2_{Y.X}}{\Sigma(X-\bar{X})^2} = \frac{\sigma^2_{Y.X}}{\Sigma x^2}$ แต่ในทางปฏิบัติ เราไม่ทราบค่าความแปรปรวน $\sigma^2_{Y.X}$ จึงใช้ $S^2_{Y.X}$ ประมาณค่า $\sigma^2_{Y.X}$ การประมาณค่าสัมประสิทธิ์การถดถอยโดยใช้การแจกแจงที จึงทำได้[39] ดังนั้นอันตรภาคความเชื่อมั่นของ β ภายใต้ความเชื่อมั่นที่กำหนด คือ

$$\text{Conf}\left(t_a < \frac{b-\beta}{\sqrt{\frac{S^2_{Y.X}}{\Sigma x^2}}} < t_b \right) = 1-\alpha$$

นั่นคือ $\text{Conf}\left(b-t_a\cdot\frac{S_{Y.X}}{\sqrt{\Sigma x^2}} < \beta < b+t_b\cdot\frac{S_{Y.X}}{\sqrt{\Sigma x^2}} \right) = 1-\alpha$

ซึ่งหมายความว่ามีโอกาส $(1-\alpha)100$ ช่วงจาก 100 ที่ค่าสัมประสิทธิ์การถดถอยของประชากรจะอยู่ในอันตรภาค $b-t_a\cdot\frac{S_{Y.X}}{\sqrt{\Sigma x^2}}$ ถึง $b+t_b\cdot\frac{S_{Y.X}}{\sqrt{\Sigma x^2}}$

---

[39] Snedecor and Cochran, Op.cit., p. 153

ตัวอย่าง ข้อมูลชุดหนึ่งมีรายละเอียดดังนี้คือ n=13, $S_b^2 = 0.1584$, b=7.478 จงหาอันตรภาคความเชื่อมั่นของ β ภายใต้ความเชื่อมั่น 95 %[40]

กำหนดระดับความเชื่อมั่น 95 % ดังนั้นค่า t จากตารางการแจกแจงที่คือ 2.201

∴ อันตรภาคความเชื่อมั่นของ β ภายใต้ความเชื่อมั่น 95 % คือ

$$\text{Conf}\left(b - t_a \cdot \frac{S_{Y.X}}{\sqrt{\Sigma x^2}} < \beta < b + t_b \cdot \frac{S_{Y.X}}{\sqrt{\Sigma x^2}} \right) = 95\%$$

นั่นคือ $\text{Conf}\ (7.478 \pm (2.201)(0.1584)) = (6.602, 8.352)$

แสดงว่าค่าความน่าจะเป็นที่สัมประสิทธิ์การถดถอยของประชากรจะอยู่ในช่วง 6.602 ถึง 8.352 คือ .95 นั่นคือมีโอกาสร้อยละ 95 ที่สัมประสิทธิ์การถดถอยของประชากรจะอยู่ในช่วง 6.602 ถึง 8.352

1.5 การประมาณค่าความแตกต่างระหว่างสัมประสิทธิ์การถดถอยของประชากร 2 กลุ่ม $(\beta_1 - \beta_2)$

เมื่อมีสมการถดถอยมากกว่าหนึ่งสมการขึ้นไป ในบางครั้งผู้วิจัยต้องการประมาณค่าความแตกต่างระหว่างสัมประสิทธิ์การถดถอยของประชากร ค่าที่ใช้ประมาณค่าความแตกต่างระหว่างสัมประสิทธิ์การถดถอยของประชากรคือความแตกต่างระหว่างสัมประสิทธิ์การถดถอยของกลุ่มตัวอย่าง ซึ่งในการประมาณค่านี้อาศัยการกระจายตัวอย่างการสุ่มของความแตกต่างระหว่างสัมประสิทธิ์การถดถอย ซึ่งการกระจายนี้จะเป็นโค้งปกติ ดังนั้นความแตกต่างระหว่างสัมประสิทธิ์การถดถอยของกลุ่มตัวอย่างก็จะเป็นปกติด้วย โดยมีค่า $E(b_1 - b_2) = \beta_1 - \beta_2$ และความคลาดเคลื่อนมาตรฐาน $\sigma_{b_1 - b_2}$ เท่ากับ

$$\sqrt{\sigma^2_{y.x}\left[\frac{1}{SS_1} + \frac{1}{SS_2}\right]} \quad \text{หรือ} \quad \sqrt{\sigma^2_{Y.X}\left[\frac{1}{\Sigma x_1^2} + \frac{1}{\Sigma x_2^2}\right]}$$

เมื่อ $\sigma^2_{y.x}$ คือ ความแปรปรวนรวม (Pooled Variance)

---

[40] Bernard Ostle, Statistics in Research (Iowa: The Iowa State University Press, 1966), p.171.

ในทางปฏิบัติเราไม่ทราบค่า $\sigma^2_{Y.X}$ ดังนั้นจึงนำการแจกแจงทีมาใช้ ดังนั้นความคลาดเคลื่อนมาตรฐานของค่าสถิติความแตกต่างของสัมประสิทธิ์การถดถอยจึงเป็น

$$\sqrt{S^2_{Y.X.}\left(\frac{1}{SS_1} + \frac{1}{SS_2}\right)}$$

โดยที่ $S^2_{Y.X} = \frac{SS_1+SS_2}{(n_1-2)(n_2-2)} = \frac{\Sigma x_1^2 + \Sigma x_2^2}{n_1+n_2-4}$

ดังนั้นความคลาดเคลื่อนมาตรฐานจึงเป็น $\sqrt{\frac{\Sigma x_1^2 + \Sigma x_2^2}{n_1+n_2-4}\left(\frac{1}{\Sigma x_1^2} + \frac{1}{\Sigma x_2^2}\right)}$ เมื่อต้องการประมาณค่าความแตกต่างระหว่างสัมประสิทธิ์การถดถอยของประชากรจะได้อันตรภาคความเชื่อมั่นภายใต้ความเชื่อมั่นที่กำหนดคือ

$$\text{Conf}\left[(b_1-b_2)-t_a\cdot\sqrt{\frac{\Sigma x_1^2+\Sigma x_2^2}{n_1+n_2-4}\left(\frac{1}{\Sigma x_1^2}+\frac{1}{\Sigma x_2^2}\right)} < \beta_1-\beta_2 < (b_1-b_2)+t_a\cdot\sqrt{\frac{\Sigma x_1^2+\Sigma x_2^2}{n_1+n_2-4}\cdot\left(\frac{1}{\Sigma x_1^2}+\frac{1}{\Sigma x_2}\right)}\right] = 1-\alpha$$

นั่นคือมีโอกาส $(1-\alpha)100$ จาก 100 ช่วงที่ความแตกต่างระหว่างสัมประสิทธิ์การถดถอยของประชากรจะอยู่ในช่วง $(b_1-b_2)-ta\sqrt{\frac{\Sigma x_1^2+\Sigma x_2^2}{n_1+n_2-4}\left(\frac{1}{\Sigma x_1^2}+\frac{1}{\Sigma x_2^2}\right)}$ ถึง

$$(b_1-b_2)+t_b\cdot\sqrt{\frac{\Sigma x_1^2+\Sigma x_2^2}{n_1+n_2-4}\left(\frac{1}{\Sigma x_1^2}+\frac{1}{\Sigma x_2^2}\right)}$$

ตัวอย่าง กลุ่มตัวอย่าง 2 กลุ่ม มีข้อมูลดังต่อไปนี้

	กลุ่มที่ 1	กลุ่มที่ 2
n	5	5
b	2.5	2.0
SS	8	16

จากช่วงความเชื่อมั่นของความแตกต่างระหว่างสัมประสิทธิ์การถดถอยภายใต้ความเชื่อมั่น 95 %

กำหนดระดับความเชื่อมั่น 95 % ดังนั้นค่า t จากตารางการแจกแจงทีคือ 2.447

$\therefore$ ช่วงความเชื่อมั่นของ $\beta_1 - \beta_2$ ที่ระดับความเชื่อมั่น 95 % คือ

$$\text{Conf}\left[(b_1-b_2)\pm(2.447)\sqrt{\frac{\Sigma x_1^2+\Sigma x_2^2}{n_1+n_2-4}\left(\frac{1}{\Sigma x_2^2}+\frac{1}{\Sigma x_2^2}\right)}\right]=(-2.50,3.50)$$

นั่นคือมีโอกาสร้อยละ 95 ที่ความแตกต่างของสัมประสิทธิ์การถดถอยของประชากรจะอยู่ในช่วง -2.50 ถึง 3.50

2. การทดสอบสมมุติฐานทางสถิติ

ในปัญหาวิจัยทางการศึกษาและจิตวิทยานิยมใช้การแจกแจงที่ช่วยในการทดสอบสมมุติฐานที่เกี่ยวกับพารามีเตอร์ดังต่อไปนี้

2.1 มัชฌิมเลขคณิตของประชากร (μ) 1 กลุ่ม

2.2 ความแตกต่างของมัชฌิมเลขคณิตของประชากร 2 กลุ่ม ($\mu_1 - \mu_2$)

2.3 การทดสอบความแตกต่างระหว่างความแปรปรวนของประชากร 1 กลุ่ม

2.4 การทดสอบสัมประสิทธิ์สหสัมพันธ์แบบโพรดักโมเมนต์ (ρ)

2.5 ความแตกต่างระหว่างสัมประสิทธิ์สหสัมพันธ์แบบโพรดักโมเมนต์ ($\rho_{XY} - \rho_{XZ}$)

2.6 สัมประสิทธิ์สหสัมพันธ์แบบอื่น ๆ (r_s, r_{p_b}, $r_{XY.Z}$)

2.7 การถดถอยเชิงเส้นตรง (linear Regression)

2.8 สัมประสิทธิ์การถดถอยของประชากร (β)

2.9 ความแตกต่างระหว่างสัมประสิทธิ์การถดถอย ($\beta_1-\beta_2$)

ซึ่งจะได้กล่าวถึงแต่ละข้ออย่างละเอียดดังนี้

2.1 การทดสอบสมมุติฐานของประชากร (μ)

ในการทดสอบสมมุติฐานของประชากร (μ) นี้มีข้อกำหนดดังนี้คือ กลุ่มตัวอย่างขนาด n จะต้องสุ่มมาจากประชากรที่แจกแจงปกติ ซึ่งมีมัชฌิมเลขคณิต μ และความแปรปรวน σ^2 อันจะเป็นผลให้การกระจายตัวอย่างการสุ่มของมัชฌิมเลขคณิตของกลุ่มตัวอย่างทั้งหมดเป็นปกติ โดยมีมัชฌิมเลขคณิต $\mu_{\bar{X}} = \mu$ และความคลาดเคลื่อน

มาตรฐาน $\sigma_{\bar{X}} = \sigma/\sqrt{n}$

เนื่องจากไม่ทราบค่าส่วนเบี่ยงเบนมาตรฐานของประชากรคือ σ จึงต้องประมาณค่าจากส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง S การทดสอบสมมุติฐานทางสถิติจึงใช้สถิติ

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \qquad (1)$$

ซึ่งถ้าสมมุติฐานสูญที่ว่า $H_0 : \mu = {}_0$ จริง สถิติ t ในสมการ (1) จะมีการแจกแจงทีที่มีขั้นความเป็นอิสระ $\nu = n-1$

ตัวอย่าง การวิจัยทางจิตวิทยาพบว่า นักเรียนที่มีปัญหาทางครอบครัวและปัญหาทางเพื่อนบ้านจะมีความถนัดทางคณิตศาสตร์ต่ำ โดยมีคะแนนเฉลี่ยเป็น 50 เมื่อสุ่มตัวอย่างนักเรียนที่มีปัญหาดังกล่าวมา 9 คน สอนความรู้ทางด้านวิชาการทางคณิตศาสตร์ให้ก่อน แล้วจึงทดสอบความถนัดทางคณิตศาสตร์ปรากฏว่าได้คะแนนดังนี้ 50 56 48 48 53 52 55 55 และ 51 ความแปรปรวนเป็น 9 อยากทราบว่าความถนัดทางคณิตศาสตร์ของนักเรียนเพิ่มมากขึ้นจริงหรือไม่ โดยกำหนดระดับความมีนัยสำคัญ α = .01[41]

สมมุติฐานที่ทดสอบ ; $H_0 : \mu = 50$

$H_1 : \mu \neq 50$

ค่าวิกฤตที่ระดับความมีนัยสำคัญ .01 จากตารางการแจกแจงทีเมื่อ $\nu = 8$ คือ 2.836

แทนค่าในสมการ (1) ได้ $t = \frac{52-50}{3/\sqrt{9}} = 2.0$ ซึ่งน้อยกว่า 2.836

$\therefore$ จะคงสมมุติฐานสูญไว้เพราะไม่มีหลักฐานเพียงพอที่จะปฏิเสธ แสดงว่าความถนัดทางคณิตศาสตร์ของนักเรียนไม่เพิ่มมากขึ้น

---

41 Hashbuger, Op.cit., p.252.

2.2 การทดสอบความแตกต่างระหว่างมัชฌิมเลขคณิตของประชากร 2 กลุ่ม ($\mu_1 - \mu_2$)

ในการทดสอบสมมุติฐานทางสถิติเกี่ยวกับความแตกต่างระหว่างมัชฌิมของประชากรนั้นจะมีข้อกำหนดดังนี้คือ

ถ้ากลุ่มตัวอย่างขนาด n_1 สุ่มมาจากประชากรที่แจกแจงปกติ ซึ่งมีมัชฌิมเลขคณิต μ_1 ความแปรปรวน σ_1^2 และกลุ่มตัวอย่างขนาด n_2 สุ่มมาจากประชากรที่แจกแจงปกติ ซึ่งมีมัชฌิมเลขคณิต μ_2 ความแปรปรวน σ_2^2 ดังนั้นการแจกแจงตัวอย่างสุ่มของค่าที่จะเป็นไปได้ทุกค่าของมัชฌิมเลขคณิตของกลุ่มตัวอย่าง $\bar{X}_1$ และ $\bar{X}_2$ จะเป็นโค้งปกติ โดยไม่ต้องคำนึงถึงขนาดของกลุ่มตัวอย่าง โดยมีมัชฌิมเลขคณิต $\mu_{\bar{X}_1-\bar{X}_2} = \mu_1 - \mu_2$ ความแปรปรวน $\sigma^2_{\bar{X}_1-\bar{X}_2} = \sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

เมื่อไม่ทราบค่าความเบี่ยงเบนมาตรฐานของประชากร (σ) จะต้องประมาณค่าจากส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง (S) ซึ่งจะได้

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

ดังนั้นการทดสอบสมมุติฐานจึงใช้สถิติ

$$t = \frac{(\bar{X}_1-\bar{X}_2) - (\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \qquad (2)$$

เมื่อ $\bar{X}_1, \bar{X}_2$ คือ มัชฌิมเลขคณิตของกลุ่มตัวอย่างที่ 1 และที่ 2 ตามลำดับ

S_1^2, S_1^2 คือ ความแปรปรวนของกลุ่มตัวอย่างที่ 1 และที่ 2 ตามลำดับ

n_1, n_2 คือ ขนาดของกลุ่มตัวอย่างที่ 1, ที่ 2 ตามลำดับ

เมื่อสมมุติฐานสูญ ; $H_0 : \mu_1 = \mu_2$ จริง การแจกแจงของสถิติในสมการ (2) จะเป็นการแจกแจงทีโดยมีชั้นความเป็นอิสระ $\nu = n_1+n_2-2$

ตัวอย่าง ในการทดลองศึกษาความสามารถในการเรียนรู้ของลิง 2 ชนิด คือ ลิง Cebus และลิง Rhesus ผู้ทดลองเลือกลิงชนิดแรกมา 6 ตัว และชนิดที่ 2 มา 17 ตัว ทดลองในห้องทดลองที่มีกรง 2 กรง กรงหนึ่งบรรจุอาหาร อีกกรงหนึ่งเป็นกรงเปล่า ทั้งสองกรงนี้อยู่ในกล่อง

ขนาดใหญ่ ซึ่งมีสวิตไฟสำหรับปิด-เปิดอยู่ภายนอกกล่อง ถ้ากดสวิตไฟเปิด ลิงจะเข้าไปในกรงที่มีอาหาร แต่ถ้ากดสวิตแล้วไฟไม่เปิดจะพบกรงเปล่า ดังนั้นลิงทั้งสองชนิดจะต้องเรียนรู้การปิด-เปิดสวิตไฟ ผลการทดลองมีดังนี้

ลิง Rhesus $\bar{X} = 162.5$, $S = 96.23$, $S^2 = 9460.30$

ลิง Cebus $\bar{X} = 137.2$, $S = 118.72$, $S^2 = 14103.28$

จะสรุปได้ไหมว่าลิง Rhesus ฉลาดกว่าลิง Cebus[42]

สมมุติฐานที่ทดสอบ ; $H_0 : \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 - \mu_2 \neq 0$

ค่าวิกฤต ที่ระดับความมีนัยสำคัญ .01, ขั้นความเป็นอิสระ $\nu = 21$ ค่า t จากตารางที คือ $\pm$ 2.8 แทนค่าในสมการ (2) ได้

$$t = \frac{(162.5 - 137.2)}{\sqrt{\frac{(17-1)(9460.30) + (6-1)(14,103.03)}{(17+6-2)}\left(\frac{1}{17}+\frac{1}{6}\right)}}$$

$$= .52 < 2.8$$

∴ จะคงสมมุติฐานสูญไว้ เพราะไม่มีหลักฐานเพียงพอที่จะปฏิเสธ ดังนั้นความสามารถในการเรียนรู้ของลิงทั้งสองชนิดไม่แตกต่างกัน

2.3 การทดสอบความแตกต่างระหว่างมัชฌิมเลขคณิตของประชากร 2 กลุ่ม $(\mu_1 - \mu_2)$ เมื่อกลุ่มตัวอย่างมีความสัมพันธ์กัน

ในบางครั้งสิ่งที่เราต้องการศึกษามีลักษณะเป็นคู่กัน เช่น ฝาแฝดคู่หนึ่ง หรืออาจเป็นกลุ่มตัวอย่างกลุ่มเดียวกัน แต่มีการทดสอบหลายครั้ง เช่นทดสอบก่อนการ

---

[42]John G. Peatman, Introduction to Applied Statistics (New York : Harper & Row, 1964), pp. 294 - 295.

ทดลองและทดสอบหลังการทดลอง ในกรณีเช่นนี้การทดสอบสมมุติฐานทางสถิติมีข้อกำหนดดังนี้

ถ้ากลุ่มตัวอย่าง 2 กลุ่ม มีขนาดเท่ากันคือ n และแต่ละสมาชิกของกลุ่มตัวอย่างที่ 1 มีลักษณะเป็นคู่กับสมาชิกของกลุ่มตัวอย่างที่ 2 เช่น สมาชิก X_1, X_2 --- X_n มีลักษณะเป็นคู่กับ Y_1, Y_2 --- Y_n ตามลำดับ หรืออาจกล่าวได้ว่าตัวแปร X และตัวแปร Y มีความสัมพันธ์กัน ดังนั้นถ้าประชากรมีการแจกแจงปกติ โดยมีมัชฌิมเลขคณิต μ ความแปรปรวน σ^2 ความแตกต่างระหว่างมัชฌิมเลขคณิตของกลุ่มตัวอย่างจะเป็นโค้งปกติ โดยมี $\mu_D = \mu_X - \mu_Y$ และความแปรปรวน $\sigma_D^2 = \sigma_X^2 + \sigma_Y^2 - 2r_{XY} \cdot \sigma_X \sigma_Y$ เมื่อมีกลุ่มตัวอย่างเพียงกลุ่มเดียวหรือ $\sigma_{\bar{D}}^2 = \sum_{i=1}^{n} \frac{(D_i-\bar{D})}{n-1} / n$ เมื่อมีกลุ่มตัวอย่างสองกลุ่ม

เมื่อไม่ทราบค่าความแปรปรวนของประชากร จะต้องใช้การประมาณค่าจากความแปรปรวนของกลุ่มตัวอย่าง นั่นคือ $S_D^2 = S_X^2 + S_Y^2 - 2r_{XY}\, S_X \cdot S_Y$ ดังนั้นการทดสอบสมมุติฐานทางสถิติจึงใช้สถิติ

$$t = \frac{(\bar{X}_1-\bar{X}_2) - (\mu_1-\mu_2)}{S_D} \qquad (3)$$

หรือ

$$t = \frac{\bar{D} - \mu_D}{S_{\bar{D}}} \qquad (4)$$

เมื่อ $\bar{D} = \sum_{i=1}^{n} D_{i/n} = \sum_{i=1}^{n} \frac{(X_i-Y_i)}{n}$

เมื่อสมมุติฐานสูญ ; $H_0 : \mu_1 = \mu_2 = \mu_D = 0$ จริง การแจกแจงของสถิติในสมการ (3) และสมการ (4) จะเป็นการแจกแจงที โดยมีขั้นความเป็นอิสระ $\nu = n-1$

ตัวอย่าง นักวิจัยการศึกษาต้องการเลือกแบบเรียนคณิตศาสตร์สำหรับนักเรียนชั้นปีที่ 9 จึงสุ่มตัวอย่างนักเรียนมา 60 คน โดยเรียงลำดับที่ของตัวอย่างตามความถนัดทางคณิตศาสตร์แล้วแบ่งออกเป็น 2 กลุ่มเท่า ๆ กัน ตามลำดับเลขคู่ คี่ ให้ตัวอย่างของแต่ละคู่เลือกแบบเรียน ถ้าคนหนึ่งเลือกแบบเรียน A อีกคนหนึ่งต้องใช้แบบเรียน B เมื่อทดสอบผลสัมฤทธิ์

ของแต่ละคู่ได้ความแตกต่างระหว่างมัชฌิมเลขคณิตเท่ากับ 5 ความแปรปรวน 64 แบบเรียนทั้งสองนี้มีประสิทธิภาพเท่ากันหรือไม่[43]

สมมุติฐานที่ทดสอบ ; $H_0 \quad : \quad \mu_D = 0$

$H_1 \quad : \quad \mu_D \neq 0$

ค่าวิกฤต เมื่อกำหนดระดับความมีนัยสำคัญ .05, $\nu = 29$ ค่า t จากตารางที่คือ $\pm$ 2.045 แทนค่าในสมการ (4) ได้ $t = \frac{5 - 0}{\sqrt{64/29}} = 3.42$ มากกว่า 2.045

$\therefore$ จะปฏิเสธสมมุติฐานสูญ แสดงว่าแบบเรียนทั้งสองมีคุณภาพแตกต่างกันอย่างมีนัยสำคัญ

2.4 การทดสอบความแตกต่างระหว่างมัชฌิมเลขคณิตของประชากร 2 กลุ่ม ($\mu_1 - \mu_2$) เมื่อความแปรปรวนของประชากรไม่เท่ากัน

การทดสอบความแตกต่างระหว่างมัชฌิมเลขคณิตของประชากรนั้น ถ้าความแปรปรวนของประชากรทั้งสองไม่เท่ากัน จะทำให้ค่าความคลาดเคลื่อนชนิดที่ 1 เพิ่มมากกว่าที่กำหนดไว้ ดังนั้นนักสถิติจึงได้แก้ไขข้อบกพร่องอันนี้โดยได้เสนอว่าควรใช้ขนาดของกลุ่มตัวอย่างเท่ากัน และใช้ขั้นความเป็นอิสระ $\nu = \frac{1}{2}(n_1+n_2 - 2)$ จึงจะมีผลกระทบกระเทือนต่อข้อสรุป เมื่อใช้การแจกแจงทีเพียงเล็กน้อยเท่านั้น[44] และถ้าขนาดของกลุ่มตัวอย่างไม่เท่ากันก็ควรจะใช้การทดสอบสมมุติฐานชนิดทางเดียวก็จะมีผลกระทบกระเทือนต่อข้อสรุปเพียงเล็กน้อย แต่การเลือกกลุ่มตัวอย่างบางครั้งก็ก่อให้เกิดความยุ่งยาก นักสถิติจึงได้คิดค่าแก้ (Correction) เมื่อความแปรปรวนของประชากรไม่เท่ากันขึ้นมา

---

[43] Slakter, *Op. cit.*, p. 260.

[44] Henry E. Garrett, *Statistics in Psychology and Educations* (Bombays: Vakils, Feffer and Simons Privated Ltd., C 1966), p. 454.

เวลช[45](Welch) ได้ปรับปรุงขั้นความเป็นอิสระที่เหมาะสม สำหรับใช้ในการทดสอบสมมุติฐานโดยที่ $\nu = \dfrac{(S_1^2 + S_2^2)}{(S_1)^2/_{n_1+1} + (S_2)^2/_{n_2+1}} - 2$ (5)

เมื่อ S_1^2, S_2^2 = ความแปรปรวนของกลุ่มตัวอย่างที่ 1 และที่ 2

n_1 , n_2 = ขนาดของกลุ่มตัวอย่างที่ 1 และที่ 2

คอแครน กับค็อก[46] (Cochran and Cox) ได้ปรับปรุงค่าวิกฤตที่เหมาะสมของระดับความมีนัยสำคัญ .05 ขึ้นมาคือ

$$t'_{.05} = \frac{S_1^2 t_1 + S_2^2 t_2}{S_1^2 + S_2^2} \qquad (6)$$

เมื่อ $t_1 = t_{.05} \cdot \nu_1$

$t_2 = t_{.01} \cdot \nu_2$

$\nu_1 = n_1 - 1$

$\nu_2 = n_2 - 1$

ในกรณีของเวลช (Welch) และ คอแครน กับค็อก (Cochran and Cox) การทดสอบสมมุติฐานจะใช้สถิติ

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_{\bar{X}_1 - \bar{X}_2}} \qquad (7)$$

เมื่อ $S^2_{\bar{X}_1 - \bar{X}_2} = S_1^2/n_1 + S_2^2/n_2$

เมื่อความแปรปรวนของประชากรไม่เท่ากัน ($\sigma_1^2 \neq \sigma_2^2$) จะใช้วิธีของเวลช(Welch)

---

[45] George A. Ferguson, Statistical Analysis in Psychology and Education. (New York : McGraw-Hill Book Company, Inc., 1959), p. 45.

[46] Cochran and Cox, Op.cit., p. 101.

หรือวิธีของคอแครน กับ ค็อก (Cochran and Cox) ก็จะให้ผลเหมือนกัน[47]

ตัวอย่าง ข้อมูลชุดหนึ่งมาจากกลุ่มตัวอย่างขนาด 20, 18 ตามลำดับ กลุ่มที่หนึ่งมีมัชฌิมเลขคณิตเป็น 49 กลุ่มที่สองมีมัชฌิมเลขคณิตเป็น 46.28 ถ้าความแปรปรวนของกลุ่มที่หนึ่งเป็น 10.105 และกลุ่มที่สองเป็น 13.154 จากข้อมูลนี้จะสรุปได้หรือไม่ว่ามัชฌิมเลขคณิตของประชากรทั้งสองเท่ากัน[48]

สมมุติฐานที่ทดสอบ ; $H_0 : \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 - \mu_2 \neq 0$

ค่าวิกฤติ กำหนดระดับความมีนัยสำคัญ .05 และขั้นความเป็นอิสระตามสมการ (5)

คือ $\nu = 32$

$\therefore$ ค่า t จากตารางที่คือ $\pm$ 2.042

แทนค่าในสมการ (7) ได้ $t = \dfrac{49 - 46.28}{\sqrt{\dfrac{10.105}{21} + \dfrac{13.154}{19}}}$

$= 2.45$ มากกว่า 2.042

แทนค่าในสมการ (6) ได้ $t'_{.05} = \dfrac{(10.105)(2.093) + (13.154)(2.898)}{10.105 + 13.154}$

$= 2.072 < 2.45$

ดังนั้นจะปฏิเสธสมมุติฐานสูญ แสดงว่ามัชฌิมเลขคณิตของประชากรทั้งสองแตกต่างกันอย่างมีนัยสำคัญ

2.5 การทดสอบความแตกต่างของความแปรปรวนของประชากร ($\sigma_1^2 = \sigma_2^2$) เมื่อกลุ่มตัวอย่างมีความสัมพันธ์กัน

เมื่อกลุ่มตัวอย่างมีความสัมพันธ์กัน ความแปรปรวนของกลุ่มตัวอย่างไม่

---

[47]Hashbarger, Op.cit., p.271.

[48]Bailey, Op.cit., p. 383.

เป็นอิสระ การทดสอบความแตกต่างระหว่างความแปรปรวนของประชากรจะใช้การแจกแจงเอฟทดสอบไม่ได้ ดังนั้นการทดสอบสมมุติฐานจึงมีข้อกำหนดดังนี้

กลุ่มตัวอย่างที่มีความสัมพันธ์กัน 2 กลุ่ม มีขนาดเท่ากันคือ n สุ่มมาจากประชากรที่แจกแจงปกติ มีมัชฌิมเลขคณิต μ_1 ความแปรปรวน σ_1^2 และกลุ่มตัวอย่างที่ 2 สุ่มมาจากประชากรที่แจกแจงปกติ มีมัชฌิมเลขคณิต μ_2 ความแปรปรวน σ_2^2 ถ้า $\sigma_1^2 = \sigma_2^2 = \sigma^2$ ในกรณีที่ไม่ทราบค่าความแปรปรวนของประชากรจะได้ s_1^2 และ s_2^2 เป็นค่าประมาณที่ไม่เอนเอียงของ σ^2 การทดสอบสมมุติฐานจะใช้สถิติ

$$t = \frac{s_1^2 - s_2^2}{\sqrt{4s_1^2\, s_2^2\,(1 - r_{12}^2)/n-2}} \qquad (8)$$

เมื่อ s_1^2, s_2^2 คือ ความแปรปรวนของกลุ่มตัวอย่างที่ 1 และที่ 2 ตามลำดับ

n คือ ขนาดของกลุ่มตัวอย่าง

r_{12} คือ สหสัมพันธ์แบบโพรดักโมเมนต์ระหว่างกลุ่มตัวอย่างที่ 1 และที่ 2

ถ้าสมมุติฐานสูญ ; $H_0 : \sigma_1^2 = \sigma_2^2$ จริง การแจกแจงของสถิติในสมการ (8) เป็นการแจกแจงที โดยมีชั้นความเป็นอิสระ $\nu = n - 2$

ตัวอย่าง นักวิจัยต้องการศึกษาว่า Performance ของนักเรียนจะคงที่หรือไม่ จึงเลือกกลุ่มตัวอย่างนักเรียนชั้นปีที่ 7 มาจำนวน 95 คน ทดสอบ Performance โดยใช้แบบทดสอบ Standford Achievement Test บันทึกคะแนนไว้อีก 1 ปี ต่อมาเมื่อนักเรียนจำนวน 95 คนนี้เลื่อนชั้นไปเรียนในชั้นปีที่ 8 จึงทดสอบด้วย Standford Achievement Test อีกครั้ง ผลปรากฏดังนี้[49]

	ชั้นปีที่ 7	ชั้นปีที่ 8
n	95	95
s^2	134.56	201.64
r_{12}	.876	

---

[49] Gene V. Glass and Julian C. Stanley, Statistical Methods in Education and Psychology. (Englewood Cliffs, Newjersey: Prentice Hall, C 1970).

สมมุติฐานที่ทดสอบ ; $H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 \neq \sigma_2^2$

ค่าวิกฤต กำหนดระดับความมีนัยสำคัญ .10 ขั้นความเป็นอิสระ $\nu = 93$ ค่า t จากตารางที่คือ ± 1.66 แทนค่าในสมการ ฤ(8) ได้

$$t = \frac{201.64 - 134.56}{\sqrt{4(201.64)(134.56)(1-.876)^2/(95-2)}} = 4.077 > 1.66$$

∴ จะปฏิเสธสมมุติฐานสูญ แสดงว่า Performance ของนักเรียนเปลี่ยนแปลงไปอย่างมีนัยสำคัญ

2.6 การทดสอบค่าสัมประสิทธิ์สหสัมพันธ์โพรดักโมเมนต์ของประชากร

ในการวิจัยทางการศึกษา บางครั้งผู้วิจัยต้องการจะศึกษาความสัมพันธ์ระหว่างตัวแปร 2 ตัวแปร เช่น ความสัมพันธ์ระหว่างคะแนนสอบไล่ชั้นมัธยมศึกษาปีที่ 5 กับคะแนนสอบเข้ามหาวิทยาลัยของนิสิต น้ำหนักกับความสูงของคน รายรับกับรายจ่ายของครอบครัว คะแนนในวิชาสถิติ กับเกรดเฉลี่ยของนักศึกษา เป็นต้น ค่าความสัมพันธ์ระหว่างสองตัวแปรจะมีมากน้อยเพียงใดนั้นเราวัดได้จากการคำนวณสัมประสิทธิ์สหสัมพันธ์โพรดักโมเมนต์ (ρ)

การแจกแจงตัวอย่างการสุ่มของสัมประสิทธิ์สหสัมพันธ์โพรดักโมเมนต์จะขึ้นอยู่กับขนาดของสัมประสิทธิ์สหสัมพันธ์โพรดักโมเมนต์ของประชากร(ρ) และขนาดของกลุ่มตัวอย่าง[50]

ถ้าค่า ρ แตกต่างจากศูนย์ การแจกแจงตัวอย่างการสุ่มของสัมประสิทธิ์สหสัมพันธ์โพรดักโมเมนต์ จะมีความเบ้ (Skewed) นั่นคือจะมีความเบ้ไปทางด้านขวา ถ้า ρ มีค่าบวกมาก ๆ และเบ้ไปทางด้านซ้าย ถ้า ρ มีค่าลบมาก ๆ[51] และ

---

[50] Walker and Lev, Op.cit., p.258.

[51] Lincoln L. Chao, Statistics Methods and Analysis (New York : McGraw-Hill Book Company, 1969), p.354.

ถ้าค่า ρ เข้าใกล้ศูนย์ เมื่อกลุ่มตัวอย่างมีขนาดใหญ่กว่า 30 การแจกแจงตัวอย่างการสุ่มของสัมประสิทธิ์สหสัมพันธ์โพรดักโมเมนต์จะเป็นโค้งปกติ โดยมีมัชฌิมเลขคณิต $\mu_r = 0$ และความคลาดเคลื่อนมาตรฐาน $\sigma_r = \sqrt{\frac{1 - \rho^2}{n-1}}$ ซึ่ง $\sigma_r = \frac{1}{\sqrt{n-1}}$ เมื่อ $\rho = 0$

ถ้าค่า ρ เข้าใกล้ศูนย์ แต่กลุ่มตัวอย่างมีขนาดเล็ก การแจกแจงตัวอย่างการสุ่มของ ρ ไม่ใช่โค้งปกติ[52] ดังนั้นการทดสอบสมมุติฐานจะมีข้อกำหนดดังนี้

กลุ่มตัวอย่างขนาด n สุ่มมาจากประชากรที่แจกแจงปกติชนิดสองตัวแปร (Bivariate Normal Distribution)นั่นคือ ข้อมูลที่จะใช้ในการศึกษาความสัมพันธ์จะต้องประกอบด้วยตัวแปร 2 ตัวแปร วัดเป็นคู่ ๆ กัน เช่น ความสูงของพี่ชาย กับความสูงของน้องสาว เมื่อคำนวณสัมประสิทธิ์สหสัมพันธ์โพรดักโมเมนต์ระหว่างตัวแปรทั้งสองในประชากร คือ ρ การทดสอบสมมุติฐานจะใช้สถิติ

$$t = \frac{r - \rho}{\sqrt{\frac{1 - r^2}{n - 2}}} \qquad (9)$$

เมื่อสมมุติฐาน ; $H_0 : \rho = 0$ จริง การแจกแจงของสถิติ t ในสมการ (9) จะเป็นการแจกแจงที โดยมีชั้นความเป็นอิสระ $\nu = n - 2$

ตัวอย่าง อาจารย์ผู้สอนวิชาสถิติเบื้องต้น ต้องการทราบว่าคะแนนสอบกลางปีและคะแนนสอบปลายปี มีความสัมพันธ์กันหรือไม่ จึงสุ่มตัวอย่างนักเรียนที่เรียนวิชาสถิติเบื้องต้นมาจำนวน 20 คน คำนวณสัมประสิทธิ์สหสัมพันธ์โพรดักโมเมนต์ระหว่างคะแนนทั้งสองได้ $r = 0.642$ อยากทราบว่าคะแนนสอบกลางปีและคะแนนสอบปลายปีของวิชาสถิติเบื้องต้น มีความสัมพันธ์กันจริงหรือไม่[53]

---

[52] ดูภาคผนวก ช.

[53] Edward G. Bryant, Statistical Analysis. (New York : Mc Graw-Hill Book Company, 1966), pp. 139-141.

สมมุติฐานที่ทดสอบ ; $H_0 : \rho = 0$

$H_1 : \rho \neq 0$

ค่าวิกฤต กำหนดระดับความมีนัยสำคัญ .01 ชั้นความเป็นอิสระ = 18 ค่า

จากตารางที คือ ± 2.101

แทนค่าในสมการ (9) ได้ $t = \frac{0.642 - 0}{\sqrt{\frac{1 - 0.642^2}{20 - 2}}} = 3.55 > 2.101$

∴ จะปฏิเสธสมมุติฐานสูญ แสดงว่า คะแนนสอบกลางปีและคะแนนสอบปลายปีของวิชาสถิติเบื้องต้น มีความสัมพันธ์กันอย่างมีนัยสำคัญ

2.7 การทดสอบความแตกต่างของสัมประสิทธิ์สหสัมพันธ์โพรดักโมเมนต์ ($\rho_{XY} - \rho_{XZ}$) เมื่อใช้กลุ่มตัวอย่างที่สัมพันธ์กัน

ในกรณีที่ต้องการทดสอบว่าตัวแปร X มีความสัมพันธ์กับตัวแปร Y และตัวแปร Z เท่ากันหรือไม่ เช่น การทำนายความสำเร็จในการเรียนมีตัวทำนายที่มีอำนาจ 2 ตัว คือ คะแนนวิชาคณิตศาสตร์ และ คะแนนวิชาภาษาอังกฤษ ถ้าต้องการตัวทำนายเพียงตัวเดียวจะต้องเลือกเอาตัวทำนายตัวใดตัวหนึ่งเท่านั้น จึงต้องคำนวณค่าความสัมพันธ์ระหว่าง X กับ Y และ X กับ Z ซึ่งจะได้ค่า r_{XY} กับ r_{XZ} แล้วจึงทดสอบความแตกต่างของ ρ_{XY} กับ ρ_{XZ} ถ้าผลสรุปได้ว่า $\rho_{XY} = \rho_{XZ}$ เราอาจจะใช้ตัวทำนาย Y หรือ Z ก็ได้ แต่ถ้า $\rho_{XY} \neq \rho_{XZ}$ ต้องเลือกใช้ตัวทำนายที่มีความสัมพันธ์กับ X มากที่สุด การทดสอบสมมุติฐานในกรณีนี้จะมีข้อกำหนดดังนี้

กลุ่มตัวอย่างขนาด n สุ่มมาจากประชากรที่มีตัวแปร X , Y และ Z โดยที่ X กับ Y , X กับ Z และ Y กับ Z แต่ละคู่มีการแจกแจงปกติแบบสองตัวแปร (Bivariate Normal Distribution) ดังนั้น สัมประสิทธิ์สหสัมพันธ์ r_{XY} r_{XZ} และ r_{YZ} ที่คำนวณได้จากกลุ่มตัวอย่างจะไม่เป็นอิสระ การทดสอบสมมุติฐานจึงใช้สถิติ

$$t = \frac{(r_{XY} - r_{XZ}) \sqrt{(n-3)(1+r_{YZ})}}{\sqrt{2(1 - r_{XY}^2 - r_{XZ}^2 - r_{YZ}^2 + 2r_{XY} r_{XZ} r_{YZ})}} \qquad (10)$$

เมื่อ n คือ ขนาดของกลุ่มตัวอย่าง

r_{XY} คือ สหสัมพันธ์โพรดักโมเมนต์ระหว่างตัวแปร X กับ Y

r_{XZ} คือ สหสัมพันธ์โพรดักโมเมนต์ระหว่างตัวแปร X กับ Z

r_{YZ} คือ สหสัมพันธ์โพรดักโมเมนต์ระหว่างตัวแปร Y กับ Z

ตัวสมมุติฐานสูญ ; $H_0 : \rho_{XY} = \rho_{XZ}$ จริง การแจกแจงของสถิติ t ในสมการ (10) เป็นการแจกแจงที โดยมีขั้นความเป็นอิสระ $\nu = n-3$

ตัวอย่าง ความสัมพันธ์ระหว่างคะแนนวิชาต่าง ๆ เมื่อนักศึกษาเรียนชั้นปีที่ 1 กับคะแนนเกรดเฉลี่ย 4 ปี ของนักศึกษา เมื่อสุ่มตัวอย่างนักศึกษามา 79 คน ได้ข้อมูลดังนี้

วิชาที่เรียนปี 1	ค่าสัมประสิทธิ์สหสัมพันธ์โพรดักโมเมต์กับเกรดเฉลี่ย (A)
ภาษาอังกฤษ (E)	.45
คณิตศาสตร์ (M)	.67
สังคมศาสตร์ (S)	.49
ภาษาต่างประเทศ (F)	.56
วิทยาศาสตร์ (S)	.58

อยากทราบว่าคะแนนคณิตศาสตร์จะมีความสัมพันธ์กับคะแนนเกรดเฉลี่ย 4 ปี มากกว่าคะแนนภาษาอังกฤษจริงหรือไม่เมื่อค่าสหสัมพันธ์ $r_{ME} = .31$[54]

สมมุติฐานที่ทดสอบ ; $H_0 : \rho_{MA} = \rho_{EA}$

$H_1 : \rho_{MA} \neq \rho_{EA}$

ค่าวิกฤต กำหนดระดับความมีนัยสำคัญ .05 เมื่อขั้นความเป็นอิสระ $\nu = 76$ ค่า t จากตารางที่จะเท่ากับ 2.00

แทนค่าในสมการ (10) ได้ $t = \dfrac{(.67 - .45)\sqrt{(79-3)(1+.31)}}{\sqrt{2(1-.45^2-.67^2-.31^2+2(.45)(.67)(.31)}}$

$= 2.34 > 2.00$

---

[54]Peatman, Op.cit., p.310.

∴ จะปฏิเสธสมมุติฐานสูญ แสดงว่าคะแนนคณิตศาสตร์ทำนายความสำเร็จในการเรียนได้ดีกว่าคะแนนภาษาอังกฤษ

2.8 การทดสอบสัมประสิทธิ์สหสัมพันธ์อื่น ๆ

ก. สัมประสิทธิ์สหสัมพันธ์แบบลำดับที่ของสเปียร์แมน (r_s)

สัมประสิทธิ์สหสัมพันธ์แบบลำดับที่ของสเปียร์แมน (r_s) จะวัดความสัมพันธ์ระหว่างตัวแปร X กับตัวแปร Y เมื่อเรียงลำดับที่ของตัวแปร X ตัวแปร Y จาก 1 ถึง n ของกลุ่มตัวอย่างขนาด n[55] เช่น ลำดับที่ของนักเรียนที่เรียนตามคะแนนวิชาที่สอบได้ 2 วิชา ค่า r_s คำนวณได้จากสูตร $1-\frac{6\Sigma d^2}{n(n^2-1)}$ เมื่อ n คือ ขนาดของกลุ่มตัวอย่าง, d คือ ผลต่างระหว่างลำดับที่ในแต่ละคู่ การทดสอบสมมุติฐานมีข้อกำหนดดังนี้

กลุ่มตัวอย่างขนาด n สุ่มมาจากประชากรที่แจกแจงปกติแบบสองตัวแปร (Bivariate Normal Distribution) เมื่อเรียงลำดับที่ของตัวแปร X เป็น $X_1, X_2 --- X_n$ และตัวแปร Y เป็น $Y_1, Y_2 --- Y_n$ วัดความสัมพันธ์ระหว่างตัวแปร X และตัวแปร Y ในประชากรเป็น ρ_s ซึ่ง ρ_s เป็นสัมประสิทธิ์สหสัมพันธ์โพรดักโมเมนต์ ดังนั้นการทดสอบสมมุติฐานจึงใช้สถิติ

$$t = \frac{r_s - \rho_s}{\sqrt{\frac{1 - r_s^2}{n-2}}} \qquad (11)$$

เมื่อขนาดของกลุ่มตัวอย่าง $10 < n < 25$ และ $H_0 : \rho_s = 0$ จริง การแจกแจงของสถิติ t ในสมการ(11) เป็นการแจกแจงที โดยมีขั้นความเป็นอิสระ $\nu = n-2$[56]

---

[55] Sidney Siegel, Nonparametric Statistics for the Behavioral Sciences (New York : McGraw-Hill Book Company, Inc., C.1956), p.202.

[56] Quin Mcnemar , Psychological Statistics (New York : John Wiley & Sons, Inc., C 1962), p.204.

ตัวอย่าง ค่าสัมประสิทธิ์สหสัมพันธ์แบบลำดับที่ของสเปียร์แมนของกลุ่มตัวอย่าง จำนวน 17 คน ภายหลังการทดสอบด้วยแบบทดสอบวัดเชาวน์ปัญญา ของ Minnesota Paper Form Board และ Otis Self-Administering Test of Mental Ability มีค่าเท่ากับ .53 เราจะเชื่อได้หรือไม่ว่าคะแนนจากแบบทดสอบทั้งสองฉบับมีความสัมพันธ์กันจริง[57]

สมมุติฐานที่ทดสอบ ; H_0 : $\rho_s = 0$

H_1 : $\rho_s \neq 0$

ค่าวิกฤต กำหนดระดับความมีนัยสำคัญ .05 ขั้นความเป็นอิสระ $\nu = 15$ ค่า t จากตารางที่เท่ากับ $\pm$ 2.131

แทนค่าในสมการ (11) ได้ $t = \dfrac{.50 - 3}{\sqrt{\dfrac{1 - .53^2}{17 - 2}}} = 2.41 > 2.131$

$\therefore$ จะปฏิเสธสมมุติฐานสูญ แสดงว่าคะแนนจากแบบทดสอบทั้งสองมีความสัมพันธ์กันอย่างมีนัยสำคัญ

ข. สัมประสิทธิ์สหสัมพันธ์แบบพอยน์ไบซีเรียล (Point Biserial r_{P_b})

การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์นั้น ถ้าตัวแปรมีคุณสมบัติเฉพาะตัวเราไม่จำเป็นจะต้องคำนวณหาสัมประสิทธิ์สหสัมพันธ์โพรดักโมเมนต์ เช่น ในกรณีที่ตัวแปรมีคุณสมบัติแบ่งออกเป็น 2 พวก ได้อย่างเด็ดขาด (True Dichotomy) เช่น + -, ชาย หญิง, 0 1, ได้ ตก ฯลฯ เราสามารถคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบพอยน์ไบซีเรียลได้โดยไม่คำนึงว่า ตัวแปรที่สามารถแบ่งออกเป็น 2 พวก ได้โดยเด็ดขาดนั้นมี

---

[57] N.M. Downie and R.W.Heath, Basic Statistical Methods. (New York : Harper & Row, Publishers, 3rd ed., C 1970), p.122.

การแจกแจงเป็นเช่นใด[58] ซึ่งผลที่ได้จะเหมือนกับสัมประสิทธิ์สหสัมพันธ์โพรดักโมเมนต์[59]

สัมประสิทธิ์สหสัมพันธ์แบบพอยน์ไบซีเรียลนิยมใช้สำหรับวิเคราะห์ข้อกระทง (Item Analysis) ของแบบทดสอบต่าง ๆ เมื่อต้องการแก้ไขข้อสอบ (Item) ที่เราออกไปแล้วว่ามีอำนาจจำแนกที่จะแยกคนเก่งและคนอ่อนออกจากกันได้ตามความต้องการหรือไม่ การทดสอบสมมุติฐานมีข้อกำหนดดังนี้

กลุ่มตัวอย่างขนาด n ประกอบด้วยตัวแปร 2 ตัวแปร โดยตัวแปรหนึ่งมีการแจกแจงปกติและต่อเนื่อง (Continuous) อีกตัวแปรหนึ่งสามารถแบ่งออกเป็น 2 พวก ได้โดยเด็ดขาด (True Dichotomy) คำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรทั้งสองในประชากรเป็น ρ_{P_b} การทดสอบสมมุติฐานจึงใช้สถิติ

$$t = \frac{r_{P_b} - \rho_{P_b}}{\sqrt{\frac{1 - r^2_{P_b}}{n - 2}}} \qquad (12)$$

เมื่อ r_{P_b} คือ สหสัมพันธ์แบบพอยน์ไบซีเรียล

n คือ ขนาดของกลุ่มตัวอย่าง

เนื่องจากสัมประสิทธิ์สหสัมพันธ์แบบพอยน์ไบซีเรียลเหมือนกับสัมประสิทธิ์สหสัมพันธ์โพรดักโมเมนต์ เมื่อสมมุติฐานสูญ ; $H_0 : \rho_{P_b} = 0$ จริง การแจกแจงของสถิติ t ในสมการ (12) เป็นการแจกแจงที่ โดยมีชั้นความเป็นอิสระ $\nu = n - 2$

ตัวอย่าง การวิเคราะห์ข้อสอบข้อที่ 17 ของแบบทดสอบวิชาภาษาอังกฤษฉบับหนึ่ง เมื่อสุ่มตัวอย่างคะแนนของนักเรียนมา 5 คน ปรากฏว่ามีนักเรียนทำถูกเพียง 3 คน คำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบพอยน์ไบซีเรียลได้ .80

---

[58] Garrett, Op.cit., p. 383.

[59] N.H. Downie and R.W. Heath, Basic Statistical Methods. (3rd ed., New York : Harper & Row, Publishers, 1970), p. 234.

อยากทราบว่าข้อสอบข้อนี้สามารถแยกคนอ่อนและคนเก่งออกจากกันได้หรือไม่[60]

สมมุติฐานที่ทดสอบ ; H_0 : $\rho_{P_b} = 0$

H_1 : $\rho_{P_b} \neq 0$

ค่าวิกฤต กำหนดระดับความมีนัยสำคัญ .05 ชั้นความเป็นอิสระ $\nu = 3$ ค่า จากตารางที่เท่ากับ $\pm$ 3.182

แทนค่าในสมการ (12) ได้ค่า $t = \dfrac{.80 - 0}{\sqrt{\dfrac{1 - .80^2}{5 - 2}}} = 2.3 < 3.182$

∴ จะคงสมมุติฐานสูญไว้ แสดงว่าข้อสอบข้อนี้ไม่สามารถแยกคนเก่งและคนอ่อนออกจากกันได้

ค. สัมประสิทธิ์สหสัมพันธ์แบบพาร์เชียล (Partial Correlation)

ค่าสหสัมพันธ์ระหว่างตัวแปร 2 ตัวแปร ที่ขจัดอิทธิพลของตัวแปรที่สามออกไปคือ สัมประสิทธิ์สหสัมพันธ์แบบพาร์เชียล ($r_{XY.Z}$) ซึ่งคำนวณได้จาก $(r_{XY}-r_{XZ}-r_{YZ})/\sqrt{(1-r_{XZ}^2)(1-r_{YZ}^2)}$ เมื่อ r_{XY}, r_{XZ}, r_{YZ} คือ ค่าสหสัมพันธ์โพรดักโมเมนต์ระหว่างตัวแปร X กับ ตัวแปร Y, ตัวแปร X กับตัวแปร Z และตัวแปร Y กับตัวแปร Z ตามลำดับ การทดสอบสมมุติฐานจะมีข้อกำหนดดังนี้

กลุ่มตัวอย่างขนาด n สุ่มมาจากประชากรที่ประกอบด้วยตัวแปร X, Y และ Z โดยที่ตัวแปร X กับตัวแปร Y, ตัวแปร X กับตัวแปร Z และตัวแปร Y กับตัวแปร Z แต่ละคู่มีการแจกแจงปกติ แบบสองตัวแปร (Bivariate Normal Distribution) คำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X และตัวแปร Y เมื่อขจัดอิทธิพลของตัวแปร Z (Eliminate Z) ในประชากรได้ $\rho_{XY.Z}$ การทดสอบ

---

[60]Hashbarger, *Op.cit.*, p.389.

สมมุติฐานจะใช้สถิติ

$$t = \frac{r_{XY.Z} - r_{XY.Z}}{\sqrt{\frac{1 - r^2_{XY.Z}}{n-3}}} \qquad (13)$$

เมื่อสมมุติฐานสูญ ; $H_0 : \rho_{XY.Z} = 0$ จริง การแจกแจงตัวอย่างของสถิติ t ในสมการ (13) เป็นการแจกแจงที โดยมีชั้นความเป็นอิสระ $\nu = n-3$

ตัวอย่าง สุ่มตัวอย่างนักเรียนมา 15 คน คำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนเชาวน์ปัญญาและคะแนนความสามารถในการเคลื่อนไหวโดยให้อายุของนักเรียนคงที่ ได้ค่าสหสัมพันธ์ $r_{XY.Z} = .36$ เชาวน์ปัญญาและความสามารถในการเคลื่อนไหวมีความสัมพันธ์กันจริงหรือไม่[61]

สมมุติฐานที่ทดสอบ ; $H_0 : \rho_{XY.Z} = 0$

$H_1 : \rho_{XY.Z} \neq 0$

ค่าวิกฤต เมื่อกำหนดระดับความมีนัยสำคัญ .05 ชั้นความเป็นอิสระ $\nu = 12$ ค่า t จากตารางการแจกแจงทีคือ $\pm$ 2.179

แทนค่าในสมการ (13) ได้ค่า $t = \frac{.36 - 0}{\sqrt{\frac{1-(.36)^2}{15-3}}} = 1.34 < 2.179$

$\therefore$ จะคงสมมุติฐานสูญไว้ แสดงว่าเชาวน์ปัญญาและความเคลื่อนไหวไม่มีความสัมพันธ์กัน

2.9 การทดสอบความถดถอยเชิงเส้นตรง (Linearity Regression)

การทดสอบความถดถอยเชิงเส้นตรงมีข้อกำหนดดังนี้

กลุ่มตัวอย่างขนาด n สุ่มมาจากประชากรที่แจกแจงปกติแบบสองตัวแปร (Bivariate Normal Distribution) โดยที่ตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นตรงต่อกัน ดังนั้นการแจกแจงตัวอย่างการสุ่มของค่าสถิติที่จะเป็นไปได้ทุก ๆ ค่า

---

[61] Ferguson, Op.cit., p.190.

ของ $\bar{Y}$ จะเป็นโค้งปกติโดยมีมัชฌิมเลขคณิต $\mu_{\bar{Y}} = \mu_{YX} = \alpha + \beta_X$ และความคลาดเคลื่อนมาตรฐาน $\sigma_{Y.X}\sqrt{\frac{1}{n} + \frac{(X_i - \bar{X})^2}{\Sigma(X-\bar{X})^2}}$ ในทางปฏิบัติเรามักไม่ทราบค่า $\sigma_{Y.X}$

ดังนั้นจึงต้องประมาณค่าจากส่วนเบี่ยงเบนมาตรฐานของมัชฌิมเลขคณิต ความถดถอยเมื่อกำหนดค่า X[62] (Standard Error of the Average Value of Y for a Fix X) ของกลุ่มตัวอย่าง ดังนั้นการทดสอบสมมุติฐานจึงใช้สถิติ

$$t = \frac{\hat{\bar{Y}} - \mu_{Y.X}}{S_{Y.X}\sqrt{\frac{1}{n} + \frac{(X_i-\bar{X})^2}{\Sigma(X-\bar{X})^2}}} \qquad (14)$$

ซึ่งสถิติ t ในสมการ (14) จะมีการแจกแจงทีโดยมีขั้นความเป็นอิสระ $\nu = n - 2$

ตัวอย่าง ข้อมูลชุดหนึ่งมีส่วนประกอบดังนี้คือ $n=10$, $\bar{X} = 2.5$, $\Sigma(X-\bar{X})^2=14.5$ และถ้า $X = 2$ จะได้ $\hat{\bar{Y}}_2 = 7.2$, $S^2_{Y.2} = 5.2$ อยากทราบว่าความถดถอยที่แท้จริงของประชากรเท่ากับ 4 หรือไม่[63]

สมมุติฐานที่ทดสอบ ; $H_0 : \mu_{Y.2} = 4$

$H_1 : \mu_{Y.2} \neq 4$

ค่าวิกฤต เมื่อกำหนดระดับความมีนัยสำคัญ .05 ขั้นความเป็นอิสระ $\nu = 8$ ค่า t จากตารางที่ = 2.306

แทนค่าในสมการ (14)ได้ $t = (7.2-4)/\sqrt{5.2\left(\frac{1}{10}+\frac{(2-2.5)^2}{14.5}\right)}$

$= 4.10 > 2.306$

∴ จะปฏิเสธสมมุติฐานสูญ แสดงว่ามัชฌิมเลขคณิตความถดถอยหรือความถดถอยที่แท้จริงของประชากรแตกต่างจาก 4 อย่างมีนัยสำคัญ

ในบางครั้งเรามีความถดถอยเชิงเส้นตรงจากกลุ่มตัวอย่างเพียงค่าเดียว

---

[62] Bryant, Op.cit., p.136.

[63] Jerome G.R.Li, Statistical Inference. (Michigan : Edwards Brothers, Inc., 1968), p.320.

(Single Predicted Value) การทดสอบสมมุติฐานจะใช้สถิติ

$$t = \frac{\hat{Y} - Y}{s_{Y.X}\sqrt{1 + \frac{1}{n} + \frac{(X-\bar{X})^2}{\Sigma(X-\bar{X})^2}}} \qquad (15)$$

เมื่อ $\hat{Y}$ คือ ค่าพยากรณ์ของจากกลุ่มตัวอย่างเมื่อกำหนดค่า X

Y คือ ค่าพยากรณ์ของประชากรเมื่อกำหนดค่า X

สถิติ t ในสมการ (15) จะเป็นการแจกแจงทีโดยมีชั้นความเป็นอิสระ $\nu = n-2$

ตัวอย่าง นักวิจัยต้องการศึกษาความสัมพันธ์ระหว่างความสูงของบิดากับความสูงของ**บุตรคนโต** โดยเลือกกลุ่มตัวอย่างบุตรคนโตมา 12 คน ได้ข้อมูลดังนี้

ความสูงของบิดา 65 68 67 64 68 62 70 66 66 67 69 71

ความสูงของบุตร 68 66 68 65 69 66 688 65 71 67 68 70

นักวิจัยพบว่า ความสูงของบิดาและความสูงของบุตรมีความสัมพันธ์กันดังนี้ $\hat{Y} = 35.82 + .470X$ โดยมีความคลาดเคลื่อนมาตรฐานในการประมาณค่า $s_{Y.X} = 1.28$ ถ้าความสูงของบิดาเท่ากับ 65.0, $\hat{Y} = 66.76$ และ $(X_i-\bar{X})^2 = 2.78$ และส่วนเบี่ยงเบนมาตรฐาน $s_X = 2.66$ อยากทราบว่าค่าพยากรณ์ของประชากรเท่ากับ 68 หรือไม่[64]

สมมุติฐานที่ทดสอบ ; $H_0 : Y = 68$

$H_1 : Y \neq 68$

ค่าวิกฤต เมื่อกำหนดระดับความมีนัยสำคัญ .05, ชั้นความเป็นอิสระ = 10 ค่า t จากตารางที คือ ± 2.23

แทนค่าในสมการ (15) ได้ $t = \dfrac{66.76 - 68}{1.28\sqrt{12 + 1 + \frac{2.78}{(2.66)^2}}}$

---

[64] Murray R. Spiegel, Theory and Problems of Statistics. (New York : McGraw-Hill Book Company, C 1961), p. 265.

$= .26 < 2.23$

$\therefore$ จะคงสมมุติฐานสูญไว้ แสดงว่าความสูงของประชากรไม่เท่ากับ 68 นิ้ว

2.10 การทดสอบสัมประสิทธิ์ความถดถอยของประชากร (β)

การทดสอบสัมประสิทธิ์ความถดถอยของประชากร (β) จะได้ผลเหมือนกับการทดสอบความมีนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์โพรดักโมเมนต์[65] เพราะว่าสมการถดถอยจะเกิดขึ้นได้เพราะตัวแปร 2 ตัวแปรจะต้องมีความสัมพันธ์เชิงเส้นตรงต่อกัน การทดสอบสมมุติฐาน ; $H_0 : \beta = 0$ นี้ ฟิชเชอร์ (Fisher) ได้แสดงว่าจะให้ผลเหมือนกับการทดสอบสมมุติฐาน ; $H_0 : \rho = 0$[66] ดังนั้นการทดสอบสมมุติฐานจะมีข้อกำหนดดังนี้

กลุ่มตัวอย่างขนาด n สุ่มมาจากประชากรที่แจกแจงปกติแบบสองตัวแปร และตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นตรงซึ่งกันและกัน และตัวแปรอิสระ X มักจะมีอิทธิพลทำให้ตัวแปรตาม Y มีความแตกต่างกัน โดยคำนวณสัมประสิทธิ์ความถดถอยของประชากรคือ β

เนื่องจากการแจกแจงตัวอย่างการสุ่มของค่าทุก ๆ ค่าที่จะเป็นไปได้ของสัมประสิทธิ์ความถดถอยของกลุ่มตัวอย่าง (b) จะเป็นโค้งปกติซึ่งมี $E(b) = \beta$ และความคลาดเคลื่อนมาตรฐาน $\sigma_b = \sigma_{Y.X}/\sqrt{\Sigma(X-\bar{X})^2}$ เมื่อ $\sigma^2_{Y.X}$ คือความแปรปรวนของ Y เมื่อ X คงที่ แต่เรามักไม่ทราบค่า σ_b ดังนั้นจึงต้องประมาณค่าจาก s_b ดังนั้นการทดสอบสมมุติฐานจึงใช้สถิติ

$$t = \frac{b - \beta}{s_b} \qquad (16)$$

เมื่อสมมุติฐานสูญ ; $H_0 : \beta = 0$ จริง การแจกแจงของสถิติ t ในสมการ (16) จะเป็นการแจกแจงที่โดยมีชั้นความเป็นอิสระ $\nu = n - 2$

---

[65] Hays, Op.cit., p.520.

[66] R.L. Anderson and T.A.Bancroft, Statistic Theory in Research (New York : McGraw-Hill Book Company, 1952),pp.158-160.

ตัวอย่าง การพยากรณ์รายได้สุทธิของเกษตรกรจำนวน 909 ครัวเรือน โดยสังเกตจากฐานะทางเศรษฐกิจสังคม ปรากฏผลดังนี้ $b = .06824$, $S^2_{Y.X} = 1.704$, $S_{Y.X} = 1.305$ และ $\Sigma(X-\bar{X})^2 = 96.019$ อยากทราบว่าสัมประสิทธิ์ความถดถอยแตกต่างจากศูนย์อย่างมีนัยสำคัญหรือไม่[67]

สมมุติฐานที่ทดสอบ ; $H_0 \;:\; \beta = 0$

$H_1 \;:\; \beta \neq 0$

ค่าวิกฤติ เมื่อกำหนดระดับความมีนัยสำคัญ .05, ขั้นความเป็นอิสระ $\nu = 907$ ค่าที่จากตารางที่เท่ากับ $\pm$ 1.96

แทนค่าในสมการ (15) ได้ค่า $t = \dfrac{.06824}{1.305/\sqrt{96.019}} = 16.2 > 1.96$

$\therefore$ จะปฏิเสธสมมุติฐานสูญ แสดงว่ามีสัมประสิทธิ์ความถดถอยของประชากรแตกต่างจากศูนย์อย่างมีนัยสำคัญ

2.11 การทดสอบความแตกต่างระหว่างสัมประสิทธิ์ความถดถอยของประชากร 2 กลุ่ม ($\beta_1 - \beta_2$)

ในกรณีที่เราเชื่อว่าสัมประสิทธิ์ความถดถอยของสมการถดถอย 2 สมการเท่ากัน การทดสอบสมมุติฐานมีข้อกำหนดดังนี้

กลุ่มตัวอย่างขนาด n_1 สุ่มมาจากประชากรที่แจกแจงปกติแบบสองตัวแปร (Bivariate Normal Distribution) ค่าสัมประสิทธิ์ความถดถอยในประชากรคือ β_1 และสัมประสิทธิ์ความถดถอยในกลุ่มตัวอย่างคือ b_1

กลุ่มตัวอย่างขนาด n_2 สุ่มมาจากประชากรที่แจกแจงปกติ แบบสองตัวแปรโดยมีสัมประสิทธิ์ความถดถอยในประชากร คือ β_2 และค่าสัมประสิทธิ์ความถดถอยในกลุ่มตัวอย่าง b_2 การทดสอบสมมุติฐานจะใช้สถิติ

---

[67] Snedecor and Cochran, Op.cit., p.184.

$$t = \frac{(b_1 - b_2) - (\beta_1 - \beta_2)}{\sqrt{\left(\frac{\sum x_1^2 + \sum x_2^2}{n_1+n_2-4}\right)\left(\frac{1}{\sum x_1^2} + \frac{1}{\sum x_2^2}\right)}} \qquad (17)$$

เมื่อ $\sum x_1^2$, $\sum x_2^2$ คือ Sum Square (SS) ของ (Y_1/X_1) และ (Y_2/X_2) ถ้าสมมุติฐานสูญ ; $H_0 : \beta_1-\beta_2 = \beta_0$ จริง การแจกแจงของสถิติ t ในสมการ (17) เป็นการแจกแจงทีโดยมีขั้นความเป็นอิสระ $\nu = n_1+n_2-4$

ตัวอย่าง ข้อมูลชุดหนึ่งประกอบด้วยรายละเอียดดังต่อไปนี้ $\sum x_1^2 = 8$, $b_1 = 2.50$ $n_1 = 5$ และข้อมูลชุดที่สองประกอบด้วย $\sum x_2^2 = 16$, $b_2 = 2.00$ $n_2 = 5$ อยากทราบว่าสัมประสิทธิ์การถดถอยของประชากรแตกต่างกัน 1 จริงหรือไม่[68]

สมมุติฐานที่ทดสอบ ; $H_0 : \beta_1 - \beta_2 = 1$

$H_1 : \beta_1 - \beta_2 \neq 1$

ค่าวิกฤต เมื่อกำหนดระดับความมีนัยสำคัญ .05 ขั้นความเป็นอิสระ $\nu = 6$ ค่า t จากตารางการแจกแจงที คือ $\pm$ 2.447

แทนค่าในสมการ(16) ได้ $t = \dfrac{(2.50 - 2.00) - 1}{\sqrt{\left(\frac{8+16}{5+5-4}\right)\left(\frac{1}{8} + \frac{1}{16}\right)}} = 0.408 < 2.447$

∴ จะคงสมมุติฐานสูญไว้ แสดงว่าสัมประสิทธิ์ความถดถอยของประชากรแตกต่างกันเท่ากับ 1 จริง

การแจกแจงเอฟ

ประวัติการพัฒนาการแจกแจงเอฟ

เซอร์โรนัลด์ เอ. ฟิชเชอร์ (Sir Ronald A. Fisher) เป็นบุคคลแรกที่

---

68 Li, Op.cit., p. 400.

เสนอการแจกแจงเอฟ เมื่อปี ค.ศ.1924 เดิมให้ชื่อว่า การแจกแจงซี (Z-Distribution) หรือการแจกแจงซีของฟิชเชอร์ (Fisher's Z Distribution) โดยมีสูตร $Z = \frac{1}{2} \log_e S_1^2/S_2^2$ ซึ่งค่าลอการิธึมทำให้ยุ่งยากแก่การคำนวณ[69] ดังนั้นมะหะโลโนบิส (Mahalonobis) ได้ปรับปรุงแก้ไขโดยเรียกว่าเอ็กซ์ (X) โดยใช้สูตร $X = S_1^2/S_2^2$ แต่ไม่เป็นที่นิยมแพร่หลาย ต่อมาจอร์จ ดับบลิว สเนดคีเคอร์ (George W. Snedecor) ได้ปรับปรุงใหม่ โดยได้สูตร $F = S_1^2/S_2^2$ หรือ เอฟ คือ อัตราส่วนระหว่างความแปรปรวนของกลุ่มตัวอย่าง 2 กลุ่ม สเนดคีเคอร์ให้ชื่อว่าการแจกแจงเอฟ (F-Distribution) เพื่อเป็นเกียรติแก่ เซอร์โรนัลด์ เอ-ฟิชเชอร์ แต่ในบางครั้งอาจเรียกว่าการแจกแจงเอฟของสเนดคีเคอร์ (Snedecor's F Distribution) เพื่อเป็นเกียรติแก่จอร์จดับบลิว สเนดคีเคอร์[70]

ในปี ค.ศ.1946 สเนดคีเคอร์ คำนวณตารางเอฟ (F-Table) ที่ระดับความมีนัยสำคัญ .05 และฟิชเชอร์ กับ เยท (Fisher and Yate) ได้คำนวณตารางที่ระดับความมีนัยสำคัญ .01 ปี ค.ศ.1953 เมอริงตัน กับ ทอมป์สัน (Menington and Thompson) ได้คำนวณตารางเอฟที่ระดับความมีนัยสำคัญ .25, 10, .025 และ .005[71] ต่อมา นอร์ตัน (Norton) ได้คำนวณที่ระดับความมีนัยสำคัญ .20 และแคลคอร์ด กับ เดมิง (Calcord and Deming) คำนวณที่ระดับความมีนัยสำคัญ .001[72].

---

[69] A.G. Rosander, Elementary Principle of Statistics (Princeton : Dvan Nostrand Company, Inc., 1957), p.506.

[70] Robert A. Hulguist, Introduction to Statistics. (New York : Holt, Rinehart and Winston, Inc., C 1965), p.63.

[71] Allen E. Edwards, Statistical Analysis (3rd ed., New York : Holt, Rinehart and Winston, Inc., C 1965), p. 105.

[72] Andersen and Barcroft, Op.cit., p.85.

ดังนั้นในปัจจุบันจึงมีค่าเอฟที่ระดับความมีนัยสำคัญต่าง ๆ กัน คือ .001, .005, .025, .01, .05, .20, .25 และ .10

การแจกแจงเอฟคือการแจกแจงตัวอย่างการสุ่มของค่าที่จะเป็นไปได้ทั้งหมดของอัตราส่วนระหว่างความแปรปรวนของกลุ่มตัวอย่าง 2 กลุ่ม

จากการสุ่มกลุ่มตัวอย่างขนาด n_1 จากประชากรที่แจกแจงปกติซึ่งมีมัชฌิมเลขคณิต μ_1 และ ความแปรปรวน σ_1^2 เมื่อคำนวณค่าความแปรปรวนของกลุ่มตัวอย่างเท่ากับ s_1^2 โดยที่ s_1^2 เป็นค่าประมาณที่ไม่มีความลำเอียงของ σ_1^2 เราจะได้ความสัมพันธ์

$$s_1^2 = \frac{\sigma_1^2 \chi^2_{\nu_1}}{n_1-1}$$

เมื่อ χ^2_ν คือการแจกแจงไคสแควร์ โดยมีขั้นความเป็นอิสระ $\nu = n-1$ และในทำนองเดียวกัน ถ้าสุ่มกลุ่มตัวอย่างขนาด n_2 จากประชากรที่แจกแจงปกติ ก็จะได้ความสัมพันธ์

$$s_2^2 = \frac{\sigma_2^2 \chi^2_{\nu_2}}{n_2-1}$$

ซึ่งจะนิยามตัวแปร F โดยให้ $F = s_1^2/s_2^2$ ถ้า $\sigma_1^2 = \sigma_2^2 = \sigma^2$ ดังนั้น

$$F = \frac{\chi^2 (n_1-1)/(n_1-1)}{\chi^2 (n_2-1)/(n_2-1)}$$

จะมีการแจกแจงเอฟ โดยมีขั้นความเป็นอิสระ $\nu_1 = n_1-1$, $\nu_2 = n_2-1$

คุณสมบัติทางคณิตศาสตร์ของการแจกแจงเอฟ

การแจกแจงเอฟมีพัฒนาการมาจากการแจกแจงซี[73] (Z-Distribution) ดังนั้นฟังก์ชันต์เดนซิตี้ของการแจกแจงเอฟก็ได้ (Derived) มาจากฟังก์ชันต์เดนซิตี้ของการแจกแจงซีของฟิชเชอร์ ซึ่งฟังก์ชันต์เดนซิตี้ของการแจกแจงเอฟ ได้แก่

---

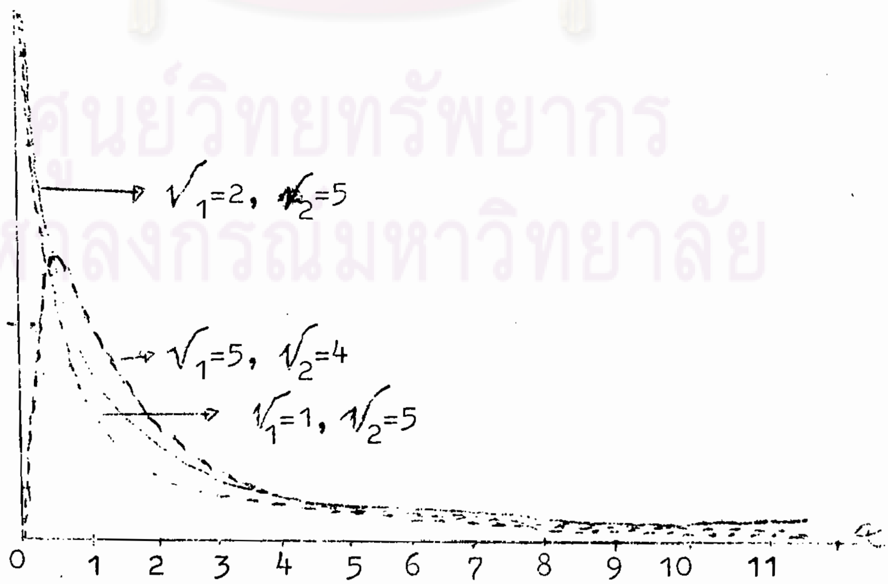
[73] ดูภาคผนวก ข.

$$g(F) = \frac{\Gamma\left[\frac{1}{2}(\nu_1+\nu_2)\right]}{\Gamma\left[\frac{1}{2}\nu_2\right]\Gamma\left[\frac{1}{2}\nu_1\right]} \cdot \left(\frac{\nu_1}{\nu_2}\right) \cdot \frac{F^{\frac{1}{2}(\nu_1+1)}}{\left(1+\frac{\nu_1}{\nu_2}F\right)^{\frac{1}{2}(\nu_1+\nu_2)}}$$

โดยที่ ν_1, ν_2 คือขั้นความเป็นอิสระ และ Γ คือ แกมมาฟังก์ชันต์

จากฟังก์ชันต์เดนซิตี้จะเห็นได้ว่าการแจกแจงเอฟเป็นอิสระจากความเบี่ยงเบนมาตรฐานของประชากร (σ) แต่การแจกแจงจะขึ้นอยู่กับพารามีเตอร์ 2 ตัว คือ ν_1, ν_2 จึงมีแบบของการแจกแจงมากมาย การแจกแจงเอฟไม่สมมาตร(Non-Symmetry) โดยจะมีความเบ้ไปทางด้านบวก (Positive Sk eweness) รูปลักษณะของการแจกแจงเป็นรูปกลับของอักษร J (Reversed J. Shape) เมื่อ $\nu_1 = 1, 2$ และถ้า $\nu_1 > 3$ การแจกแจงจะเบ้ไปทางด้านขวา จนกระทั่งขั้นความเป็นอิสระเข้าใกล้ค่าอนันต์ (∞) การแจกแจงจะมีลักษณะสมมาตรดังรูปที่ 4

รูปที่ 4 การแจกแจงเอฟที่ขั้นความเป็นอิสระ ($\nu_1 = 1, \nu_2 = 5$), ($\nu_1 = 2, \nu_2 = 5$) และ ($\nu_1 = 5, \nu_2 = 4$)



การแจกแจงเอฟไม่มีค่าลบ (Nonnegative) เพราะว่าทั้งตัวตั้งและตัวหารอยู่ในลักษณะกำลังสองเสมอ ค่าเอฟในการทดสอบความแตกต่างระหว่างความแปรปรวนของประชากร 2 กลุ่ม จะต้องไม่น้อยกว่า 1 เพราะว่าเรากำหนดให้ค่าประมาณความแปรปรวนที่มีค่ามากเป็นตัวตั้งและในกรณีที่ค่าประมาณความแปรปรวน s_1^2 และ s_2^2 มาจากประชากรเดียวกัน ซึ่งแสดงว่า $\sigma_1^2 = \sigma_2^2 = \sigma^2$ ค่าเอฟจะเท่ากับ 1 หรือใกล้เคียง 1

คุณสมบัติทางสถิติของการแจกแจงเอฟ

พื้นที่ใต้โค้งการแจกแจงเอฟทั้งหมดจะมีค่าเท่ากับ 1 และค่าที่คาดหวัง (Expected Value) ของเอฟ ซึ่งคำนวณได้จาก $E(F) = \nu_{2}/\nu_2 - 1$ โดยที่ ν_2 จะต้องมากกว่า 2 นั้น จะเท่ากับ 1 หรือเข้าใกล้ 1 และความแปรปรวนของการแจกแจงเอฟ $E\left\{\left[F-E(F)\right]^2\right\}$ จะเท่ากับ $\frac{2\nu_1(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}$ แต่ในกรณีที่ $\nu_2 \leq 4$ การแจกแจงเอฟจะไม่มีความแปรปรวน และถ้า $\nu_1 > 2$ โค้งการแจกแจงจะมีความสูง (Ordinate) เป็น $\frac{\nu_1-2}{\nu_1} \cdot \frac{\nu_2}{\nu_2-1}$ และมีค่ามากที่สุดเท่ากับ 1 เมื่อ ν_1, ν_2 เข้าใกล้ค่าอนันต์[74] (∞) การแจกแจงเอฟจะมีพิสัยตั้งแต่ 0 ถึง ค่าอนันต์ (∞) และถ้า $F^{\nu_1}_{\nu_2}$ มีการแจกแจงเอฟ ดังนั้น $F^{\nu_2}_{\nu_1} = \frac{1}{F^{\nu_1}_{\nu_2}}$ จะต้องมีการแจกแจงเอฟด้วย[75]

---

[74] Wilford L.L' Esperance, Modern Statistics for Business and Economics. (New York : The Mcmillan Company, C 1971), p.211.

[75] Henry L. Alder and Edward B. Roessler, Introduction to Probability and Statistics (3rd ed., Sanfrancisco : W.H. Freeman and Company, 1969), p.232.

ความสำคัญของการแจกแจงเอฟ

การเปรียบเทียบความแตกต่างระหว่างมัชฌิมเลขคณิตของกลุ่มตัวอย่างสองกลุ่มสามารถทดสอบความมีนัยสำคัญได้โดยใช้อัตราส่วน t หรือ z ซึ่งขึ้นอยู่กับการทราบหรือไม่ทราบค่าความแปรปรวนของประชากร ในกรณีมีจำนวนกลุ่มที่จะเปรียบเทียบมากกว่า 2 กลุ่มขึ้นไป ย่อมเกิดความยุ่งยากซับซ้อน ถ้าใช้วิธีเปรียบเทียบของกลุ่มตัวอย่าง 2 กลุ่ม จะทำให้เสียเวลา เพราะต้องทดสอบทีละคู่ นอกจากนี้ยังเพิ่มความคลาดเคลื่อนชนิดที่ 1 ให้มากกว่ากำหนดด้วย ดังนั้นจึงนิยมใช้เทคนิควิธีที่ฟิชเชอร์ (Fisher) เรียกว่า การวิเคราะห์ความแปรปรวน (Analysis of Variance) ซึ่งจะต้องใช้อัตราส่วน F ทดสอบความมีนัยสำคัญเพื่อสรุปอ้างอิง (Statistical Inference)จากการใช้กลุ่มตัวอย่างมากกว่า 2 กลุ่มขึ้นไปได้ในเวลาเดียวกัน

นอกจากนี้การแจกแจงเอฟสามารถนำไปใช้ประโยชน์ในการสมมุติฐานทางสถิติได้อีกมากมาย เช่น การทดสอบความแตกต่างระหว่างความแปรปรวยของประชากร 2 กลุ่ม การทดสอบสัมประสิทธิ์สหสัมพันธ์พหุคูณ การทดสอบภาวะเชิงเส้น (Linearity Test) ฯลฯ

ข้อตกลงเบื้องต้นโดยทั่วไปในการใช้การแจกแจงเอฟ

การใช้การแจกแจงเอฟมีข้อตกลงเบื้องต้นดังนี้คือ

1. กลุ่มตัวอย่างจะต้องสุ่มมาจากประชากรที่แจกแจงปกติซึ่งมีมัชฌิมเลขคณิต μ และความแปรปรวน σ^2 จึงจะทำให้กลุ่มตัวอย่างเป็นอิสระ และค่าความแปรปรวนของกลุ่มตัวอย่างก็จะเป็นอิสระด้วย

2. ความแปรปรวนของกลุ่มตัวอย่างทั้งสองกลุ่มจะต้องเท่ากัน (Homocedasticity) นั่นคือ $\sigma_1^2 = \sigma_2^2 = \sigma^2$

3. ค่าประมาณความแปรปรวน s_1^2 และ s_2^2 จะต้องมาจากตัวแปรสุ่มอิสระ (Random Variable) นั่นคือ กลุ่มตัวอย่างจะต้องได้มาโดยวิธีสุ่ม

การใช้การแจกแจงเอฟในการวิจัยทางการศึกษา

การแจกแจงเอฟได้นำมาใช้ในการวิจัยทางการศึกษา ดังนี้

1. การประมาณค่าพารามีเตอร์ (σ_1^2/σ_2^2) ภายใต้ความเชื่อมั่นที่กำหนด

การใช้การแจกแจงเอฟประมาณค่าพารามีเตอร์นั้น คือการประมาณค่าอัตราส่วนระหว่างความแปรปรวนของประชากร (σ_1^2/σ_2^2) เพียงอย่างเดียวเท่านั้น

การประมาณค่าอัตราส่วนระหว่างความแปรปรวนของประชากร (σ_1^2/σ_2^2) นั้น ค่าที่ใช้ในการประมาณคือ อัตราส่วนระหว่างความแปรปรวนของกลุ่มตัวอย่าง (s_1^2/s_2^2) เมื่อ s_1^2 คำนวณได้จาก $\sum_{i=1}^{n_1}\frac{(x_{i1}-\bar{x}_1)^2}{n_1-1}$ และ s_2^2 คำนวณจาก $\sum_{i=1}^{n_2}\frac{(x_{i2}-\bar{x}_2)^2}{n_2-1}$, ดังนั้นเมื่อมีกลุ่มตัวอย่างหลายกลุ่ม อัตราส่วนระหว่างความแปรปรวนก็จะมีหลายค่า จึงนิยมพิจารณา และประมาณค่า σ_1^2/σ_2^2 ให้อยู่ในช่วงหรืออันตรภาคภายใต้ความเชื่อมั่นที่กำหนด

ในการประมาณค่า σ_1^2/σ_2^2 นั้นจะต้องอาศัยการแจกแจงตัวอย่างการสุ่มของอัตราส่วนระหว่างความแปรปรวนของกลุ่มตัวอย่างที่จะเป็นไปได้ทุก ๆ ค่า ดังต่อไปนี้

ถ้าเลือกกลุ่มตัวอย่างขนาด n_1 จากประชากร x_i ; $i = 1,2---n_i$ ที่แจกแจงปกติ โดยมีมัชฌิมเลขคณิต μ_1 ความแปรปรวน σ_1^2 และกลุ่มตัวอย่างขนาด n_2 จากประชากร x_i ; $i=1,2---n_2$ ที่แจกแจงปกติ โดยมีมัชฌิมเลขคณิต μ_2 ความแปรปรวน σ_2^2 ดังนั้นตัวแปรสุ่ม $\frac{s_1^2(n_1-1)}{\sigma_1^2}$ และ $\frac{s_2^2(n_2-1)}{\sigma_2^2}$ เมื่อ $s_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_i-\bar{x})^2$ และ $s_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(x_i-\bar{x})^2$ มีการแจกแจงไคสแควร์ โดยมีขั้นความเป็นอิสระ $n_1 - 1$ และ $n_2 - 1$ ตามลำดับ

เนื่องจาก F คืออัตราส่วนระหว่างตัวแปรอิสระไคสแควร์หารด้วยขั้นความเป็นอิสระกับตัวแปรอิสระไคสแควร์หารด้วยขั้นความเป็นอิสระ ดังนั้น

$$\therefore F = \frac{(n_1-1)(s_1^2/\sigma_1^2)(n_2-1)}{(n_2-1)(s_2^2/\sigma_2^2)(n_1-1)} = \frac{(s_1\ \sigma_2)^2}{(s_2\ \sigma_1)^2}$$ [75]

จะมีการแจกแจงเอฟ โดยมีขั้นความเป็นอิสระ $\nu_1=n_1-1, \nu_2=n_2-1$ ดังนั้นความเชื่อมั่นที่ F จะอยู่ในช่วงใดช่วงหนึ่งคือ

$$\text{Conf } (F_a < F < F_b) = 1 - \alpha$$

ซึ่ง F_a, F_b คือ ค่าต่ำสุดและสูงสุดภายในช่วงความเชื่อมั่นที่กำหนด จากสมการข้างบนจึงให้ความหมายว่า ด้วยความเชื่อมั่น $1 - \alpha$ ค่า F จะอยู่ระหว่าง F_a และ F_b ซึ่งสามารถแทนค่า F แล้ว ประมาณค่า $\frac{\sigma_1^2}{\sigma_2^2}$ ได้ดังนี้

$$\text{Conf } \left(F_a < \frac{s_1^2\ \sigma_2^2}{s_2^2\ \sigma_1^2} < F_b\right) = 1 - \alpha$$

$$\text{Conf } \left(F_a \cdot \frac{s_2^2}{s_1^2} < \frac{\sigma_2^2}{\sigma_1^2} < F_b \cdot \frac{s_2^2}{s_1^2}\right) = 1 - \alpha$$

โดยที่ $F_a = \dfrac{1}{F^{\nu_2}_{\nu_1(\alpha/2)}}$ และ $F_b = F^{\nu_1}_{\nu_2(1-\alpha/2)}$

ตัวอย่าง กลุ่มตัวอย่างอิสระ 2 กลุ่ม ขนาด 8, 10 มีความแปรปรวน $s_1^2 = 7.14$ และ $s_2^2 = 3.21$ จงหาช่วงความเชื่อมั่นของอัตราส่วนระหว่างความแปรปรวนของประชากรภายใต้ความเชื่อมั่นร้อยละ 90[76]

กำหนดความเชื่อมั่น 90 % ดังนั้นค่า F จากตารางเอฟ คือ

$F^{9}_{7} = 3.68$ และ $F^{7}_{9} = 3.29$

---

[75] Henry Lass and Peter Gottleib, Probability and Statistics (Reading, Massachusetts : Addison-Wesley Publishing Company, C1971), p.287.

[76] Mendenhall, Op.cit., p. 211.

ดังนั้นช่วงความเชื่อมั่นของ σ_2^2/σ_1^2 เมื่อกำหนดความเชื่อมั่น 90 % คือ

$$\text{Conf}\ (F_a \cdot S_2^2/S_1^2 < \sigma_2^2/\sigma_1^2 < F_b \cdot S_2^2/S_1^2) = 1 - \alpha$$

$$\text{Conf}\ (.60 < \sigma_2^2/\sigma_1^2 < 7.32) = 90\ \%$$

หมายความว่ามีโอกาสร้อยละ 90 ที่ σ_2^2/σ_1^2 จะอยู่ในช่วง .60 ถึง 7.32

2. การทดสอบสมมุติฐานทางสถิติ

ในปัญหาวิจัยการศึกษาแนิยมใช้การแจกแจงเอฟช่วยในการทดสอบสมมุติฐานที่เกี่ยวกับ

2.1 ความแตกต่างระหว่างความแปรปรวนของประชากร 2 กลุ่ม
($\sigma_1^2 = \sigma_2^2$)

2.2 ความแตกต่างระหว่างความแปรปรวนของประชากรหลายกลุ่ม
($\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \ldots$)

2.3 สัดส่วนของประชากร (p) 1 กลุ่ม

2.4 สหสัมพันธ์พหุคูณ

2.5 ความแตกต่างระหว่างสัมประสิทธิ์สหสัมพันธ์พหุคูณ (R_1-R_2)

2.6 อัตราส่วนสหสัมพันธ์ (η)

2.7 ภาวะเชิงเส้น (Linearity Test)

2.8 ความแตกต่างระหว่างมัชฌิมเลขคณิตของประชากรหลายกลุ่ม
($\mu_1 = \mu_2 = \ldots$)

ซึ่งจะได้กล่าวถึงแต่ละข้ออย่างละเอียดดังนี้

2.1 การทดสอบความแตกต่างระหว่างความแปรปรวนของประชากร 2 กลุ่ม ($\sigma_1^2 - \sigma_2^2$)

ปัญหาในด้านวิจัยการศึกษานั้น ยากที่จะเกี่ยวข้องกับความแปรปรวนเพียง

ค่าเดียว โดยมากมักจะเกี่ยวกับปัญหาที่จะต้องนำความแปรปรวนของประชากรทั้งสองมาเปรียบเทียบกัน เช่น ความแปรปรวนของประชากรหนึ่งเท่ากับความแปรปรวนของอีกประชากรหนึ่งหรือไม่ นั่นคือ $H_0 : \sigma_1^2 = \sigma_2^2$ การทดสอบสมมุติฐานทางสถิติมีข้อกำหนดดังนี้คือ

กลุ่มตัวอย่างขนาด n_1 สุ่มมาจากประชากรที่มีการแจกแจงปกติ ซึ่งมีมัชฌิมเลขคณิต μ_1 ความแปรปรวน σ_1^2 และกลุ่มตัวอย่างขนาด n_2 สุ่มมาจากประชากรที่แจกแจงปกติ ซึ่งมีมัชฌิมเลขคณิต μ_2 ความแปรปรวน σ_2^2 ถ้า s_1^2 และ s_2^2 เป็นความแปรปรวนของกลุ่มตัวอย่างขนาด n_1 และ n_2 ตามลำดับ ดังนั้น s_1^2 และ s_2^2 ต่างก็เป็นอิสระ และถ้า $\sigma_1^2 = \sigma_2^2 = \sigma^2$ ค่า s_1^2 และ s_2^2 จะเป็นค่าประมาณความแปรปรวนของประชากรกลุ่มเดียวกัน ดังนั้นการทดสอบสมมุติฐานจะใช้สถิติ

$$F = s_1^2 / s_2^2 \qquad (1)$$

ถ้าสมมุติฐานสูญ ; $H_0 : \sigma_1^2 = \sigma_2^2 = 0$ จริง การแจกแจงของสถิติ F ในสมการ (1) เป็นการแจกแจงเอฟ โดยมีขั้นความเป็นอิสระ $\nu_1 = n - 1$, $\nu_2 = n_2 - 1$

ตัวอย่าง การเปรียบเทียบประสิทธิภาพการทำงานของพนักงานเจาะบัตร ระหว่างบริษัท 2 บริษัท โดยสุ่มตัวอย่างพนักงานของบริษัทแรกมา 10 คน ปรากฏว่าความสามารถเฉลี่ยเท่ากับ 200 บัตร/ชั่วโมง โดยมีความแปรปรวนเท่ากับ 225 เมื่อสุ่มตัวอย่างพนักงานของบริษัทที่ 2 มาจำนวน 16 คน ปรากฏว่า ความสามารถเฉลี่ยประมาณ 205 บัตร/ชั่วโมง ความแปรปรวนเท่ากับ 361 อยากทราบว่าพนักงานทั้งสองบริษัทนี้มีความแปรปรวนในการทำงานเท่ากันหรือไม่[77]

---

[77] Ann Hugkes and Dennis Grzwoig, Statistics : A Foundation for Analysis (Reading Massachusetts : Addison-Wesley Publishing Company, C 1971), p. 257.

สมมุติฐานที่ทดสอบ ; $H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 \neq \sigma_2^2$

ค่าวิกฤต เมื่อกำหนดระดับความมีนัยสำคัญ $\alpha = .02$ และชั้นความเป็นอิสระ

$\nu_1 = 9,\ \nu_2 = 15$ คือ $F_{9,15} = 4.63$ และ $F_{15,9} = 0.19$

แทนค่าในสมการ (1) ได้ $F = \frac{225}{361} = 0.623 > 0.19$

ดังนั้นจะปฏิเสธสมมุติฐานสูญ แสดงว่าประสิทธิภาพการทำงานของพนักงานทั้งสองบริษัท แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

2.2 การทดสอบความแตกต่างระหว่างความแปรปรวนของประชากรหลายกลุ่ม

การทดสอบความแตกต่างระหว่างความแปรปรวนของกลุ่มตัวอย่างตั้งแต่ 2 กลุ่มขึ้นไปนั้น ถ้าจะเปรียบเทียบทีละคู่จะทำให้เสียเวลาและเสียแรงงาน และสิ่งสำคัญคือทำให้ความคลาดเคลื่อนชนิดที่ 1 (Type One Error) เพิ่มมากขึ้นกว่าที่กำหนดไว้ ดังนั้นในปี ค.ศ.1950 H.O.Hartley ได้แก้ไขข้อบกพร่องอันนี้ ดังนั้นการทดสอบสมมุติฐานทางสถิติจะมีข้อกำหนดดังนี้

กลุ่มตัวอย่าง m กลุ่ม แต่ละกลุ่มขนาด n สุ่มมาจากประชากรที่แจกแจงปกติโดยมีมัชฌิม μ_i ความแปรปรวน σ_i^2 ถ้า $\sigma_1^2 = \sigma_2^2 = \dots \sigma_m^2$ แล้วละก็การเปรียบเทียบความแตกต่างระหว่างความแปรปรวนของประชากรจะใช้สถิติ

$$F = \frac{\hat{S}^2_{max}}{\hat{S}^2_{min}} \qquad (2)$$

เมื่อ $\hat{S}^2_{max}$ คือ ความแปรปรวนของกลุ่มตัวอย่างที่มีค่ามากที่สุด

$\hat{S}^2_{min}$ คือ ความแปรปรวนของกลุ่มตัวอย่างที่มีค่าน้อยที่สุด

ถ้าสมมุติฐานสูญ ; $H_0 : \sigma_1^2 = \sigma_2^2 = \dots \sigma_m^2$ จริง การแจกแจงของสถิติ F ในสมการ (2) จะมีการแจกแจงเป็นเอฟ โดยมีชั้นความเป็นอิสระ $\nu_1 = n-2,\ \nu_2 = n$

ตัวอย่าง จงเปรียบเทียบความแปรปรวนของกลุ่มตัวอย่าง 6 กลุ่ม ซึ่งมีลักษณะดังต่อไปนี้

ตารางที่ 2 ตัวอย่างข้อมูลสมมุติ[78]

กลุ่มตัวอย่าง	SS	df	s^2
1	26	4	6.5
2	16	4	4.0
3	54	4	13.5
4	34	4	8.5
5	50	4	12.5
6	14	4	3.5

สมมุติฐานที่ทดสอบ ; $H_0 : \sigma_i^2 = \sigma^2$; $i = 1, 2 \text{ --- } 6$

$H_1 : \sigma_i^2 \neq \sigma^2$

ค่าวิกฤต เมื่อกำหนดระดับความมีนัยสำคัญที่ .05 ได้ค่า F จากตารางเอฟเท่ากับ 29.5

แทนค่าในสมการ (2) ได้ $F = \frac{13.5}{3.5} = 3.86 < 29.5$

ดังนั้นจะคงสมมุติฐานสูญไว้ แสดงว่าความแปรปรวนไม่แตกต่างกัน

2.3 การทดสอบความแตกต่างระหว่างมัชฌิมเลขคณิตของประชากรหลายกลุ่ม ($\mu_1 = \mu_2 = \mu_3 = \ldots \mu_k$)

เมื่อต้องการทดสอบว่ากลุ่มตัวอย่าง k กลุ่ม เป็นกลุ่มตัวอย่างสุ่มมาจากประชากรเดียวกันหรือต่างกัน แต่มีค่าความแปรปรวนเท่ากัน จะต้องใช้เทคนิควิธีที่เรียกว่าการวิเคราะห์ความแปรปรวน (Analysis of Variance)

การวิเคราะห์ความแปรปรวน (Analysis of Variance) คือการจำแนกหรือแยกการแปรผันทั้งหมด หรือการแปรผันรวมยอด (Total Variation) อันได้จากการสังเกตชุดหนึ่งตามที่วัดได้โดยใช้ผลรวมกำลังสองของการเบี่ยงเบนจากมัชฌิมเลขคณิตออกเป็นส่วนประกอบต่าง ๆ ตามแหล่งกำเนิดของการแปรผันที่นิยามไว้หรือตามตัวแปรอิสระ

78 Peatman, Op. cit., p.336.

ที่มีผลต่อการตอบสนองของตัวอย่าง (Source of Variation)

การวิเคราะห์ความแปรปรวนจะต้องใช้การประมาณความแปรปรวนจากข้อมูลของกลุ่มตัวอย่าง ตามวิธีดังนี้คือ

1. ประมาณค่าความแปรปรวนของประชากรจากการเบี่ยงเบนของมัชฌิมเลขคณิตของกลุ่มตัวอย่าง (Sample Mean) จากมัชฌิมเลขคณิตรวม (Grand Mean) ซึ่งบางครั้งเรียกว่า การแปรผันระหว่างกลุ่ม (Between Group Variation)

2. ประมาณค่าความแปรปรวนของประชากรจากการเบี่ยงเบนของตัวอย่าง (Observation) จากมัชฌิมเลขคณิตของกลุ่มตัวอย่าง (Sample Mean) ในแต่ละกลุ่มตัวอย่าง ซึ่งเรียกว่าการแปรผันภายในกลุ่ม (With in Group Variation)

3. ประมาณค่าความแปรปรวนของประชากรจากการเบี่ยงเบนของตัวอย่าง (Observation)จากมัชฌิมเลขคณิตรวม (Grand Mean) ซึ่งเรียกว่าการแปรผันทั้งหมดหรือการแปรผันรวมยอด (Total Variation)

ค่าประมาณความแปรปรวนของแต่ละวิธีคือ ผลหารระหว่างการแปรผันของแต่ละวิธี กับขั้นความเป็นอิสระ (Degree of Freedom) การทดสอบสมมุติฐานจะมีข้อกำหนดดังนี้คือ

กลุ่มตัวอย่างสุ่มอิสระจำนวน k กลุ่ม มาจากประชากรที่แจกแจงปกติ โดยมีมัชฌิมเลขคณิต ($\mu_1, \mu_2, \ldots \mu_k$) ความแปรปรวน ($\sigma_1^2 = \sigma_2^2 = \ldots \sigma_k^2$) และแต่ละตัวอย่างในกลุ่มตัวอย่างอิสระนั้น สามารถเขียนแทนได้ด้วย Mathematical Model[79] ดังนั้นการประมาณค่าความแปรปรวนของประชากรจะประมาณได้ตามแหล่งต้นกำเนิดของความแปรผัน (Source of Variation) ใน Model นั้น ๆ

---

[79]Janet T. Spence, Benton J. Underwood, Carl p.Duncan and John W. Cottan, Element Statistics (New York : Appleton-Century Crafts, C 1965), p.182.

เนื่องจากกลุ่มตัวอย่างสุ่มอิสระ ดังนั้นค่าประมาณความแปรปรวนของประชากรแต่ละจำนวนจะต้องเป็นอิสระด้วย ดังนั้นการทดสอบสมมุติฐานจึงใช้สถิติ F เพราะเป็นการเปรียบเทียบระหว่างค่าความแปรปรวนที่เป็นอิสระของประชากรจำนวนเดียวกัน

ถ้ากำหนดให้ i = กลุ่มตัวอย่างที่ i

k = จำนวนกลุ่มตัวอย่าง

n_i = จำนวนตัวอย่างในกลุ่มตัวอย่างที่ i

T = จำนวนตัวอย่างทั้งหมด

$\bar{X}_i$ = มัชฌิมเลขคณิตของกลุ่มตัวอย่างที่ i

X_{ij} = ค่าที่ได้จากตัวอย่าง ในกลุ่มตัวอย่างที่ i

$\bar{\bar{X}}$ = มัชฌิมเลขคณิตรวม

ดังนั้นค่าประมาณความแปรปรวนที่ได้จากการประมาณค่าตามวิธีดังกล่าวคือ

1. เมื่อประมาณค่าจากการแปรผันระหว่างกลุ่มจะได้

$$M.S. = \sum_{i=1}^{k} n_i \frac{(\bar{X}_i-\bar{\bar{X}})^2}{k-1}$$

2. เมื่อประมาณค่าจากการแปรผันภายในกลุ่ม จะได้

$$M.S. = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{(X_{ij}-\bar{X}_i)^2}{T-k}$$

3. เมื่อประมาณค่าจากการแปรผันทั้งหมด จะได้

$$M.S. = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{(X_{ij}-\bar{\bar{X}})^2}{(k-1)+(T-k)}$$

หรืออาจจะแสดงได้ด้วยตาราง

ตารางที่ 3 ตารางวิเคราะห์ความแปรปรวนตามแหล่งต้นกำเนิด

แหล่งต้นกำเนิด	ชั้นความเป็นอิสระ	ความแปรผัน (SS)	ค่าประมาณความแปรปรวน (MS)	F
ระหว่างกลุ่ม	$k-1$	$\sum_{i=1}^{k} n_i(\bar{X}_i-\bar{X})^2$	$SS_{B/df}$	$M.S_B$
ภายในกลุ่ม	$T-k$	$\sum_{i=1}^{k} \sum_{j=1}^{n_1}(X_{ij}-\bar{X}_i)^2$	$SS_{W/df}$	$M.S_W$
รวมยอด	$(k-1)(T-k)=T-1$	$\sum_{i=1}^{k} \sum_{j=1}^{n_i}(X_{ij}-\bar{\bar{X}})^2$		

ตัวอย่างที่ 4 ตัวอย่างข้อมูลสมมุติ[80]

	กลุ่มตัวอย่าง			
	1	2	3	4
ตัวอย่าง	45	35	34	41
	46	33	34	41
	49		35	44
	44		34	43
			33	41
				42
				44
				41
				41
รวม	184	68	170	378
มัชฌิมเลขคณิต	46	34	34	42

สมมุติฐานที่ทดสอบ $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

$H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$

ค่าวิกฤต เมื่อกำหนดระดับความมีนัยสำคัญ $\alpha = .01$; ขั้นความเป็นอิสระ $\nu_1 = 3$

$\nu_2 = 16$ คือ $F^{3}_{16} = 5.29$

เมื่อแทนค่าจะได้ตารางวิเคราะห์ความแปรปรวนดังนี้

---

[80] Ostle, Op.cit., pp. 135-136.

ตารางที่ 5 ตัวอย่างตารางวิเคราะห์ความแปรปรวนจากข้อมูลสมมุติ

แหล่งต้นกำเนิด	ชั้นความเป็นอิสระ	การแปรผัน	ค่าประมาณความแปรปรวน	F
ระหว่างกลุ่ม	3	432	144	72
ภายในกลุ่ม	16	32	2	
รวมทั้งหมด	19	464	164	

ดังนั้นจะปฏิเสธสมมุติฐานสูญ แสดงว่ามัชฌิมเลขคณิตของกลุ่มตัวอย่างแตกต่างกันอย่างมีนัยสำคัญ

นั่นคือ มัชฌิมเลขคณิตของกลุ่มตัวอย่างไม่เท่ากัน

2.4 การทดสอบสัดส่วนของประชากร

การแจกแจงการสุ่มตัวอย่างที่จะเป็นไปได้ทุกค่าของสัดส่วนของกลุ่มตัวอย่าง (p) ถ้ากลุ่มตัวอย่างสุ่มมาอย่างอิสระ คือสุ่มแบบแทนที่จากประชากรขนาดจำกัด หรือสุ่มมาจากประชากรขนาดไม่จำกัด p จะแจกแจงไบโนเมียล โดยมีมัชฌิมเลขคณิต $\mu_p = P$ และความคลาดเคลื่อนมาตรฐาน $\sigma_p = \sqrt{\frac{(1-P)}{n}}$ ดังนั้นถ้ากลุ่มตัวอย่างที่ขนาดใหญ่พอ การแจกแจงการสุ่มตัวอย่างของ p ก็จะเข้าใกล้การแจกแจงปกติ ดังนั้นการทดสอบสมมุติฐานทางสถิติก็จะใช้การแจกแจงปกติได้

ในกรณีที่กลุ่มตัวอย่างขนาดเล็กและความน่าจะเป็นของความสำเร็จค่าน้อยจะใช้การแจกแจงปกติไม่ได้ การทดสอบสมมุติฐานจะมีข้อกำหนดดังนี้คือ

มีการทดลองแบบเบอร์นูลีชุดหนึ่ง (n) ซึ่งมีความสำเร็จ k ครั้ง ให้ p เป็นความน่าจะเป็นของความสำเร็จ และความน่าจะเป็นของความสำเร็จที่รวมตั้งแต่ k ครั้งขึ้นไป คิดจากการแจกแจงไบโนเมียล

$$p\ (x \geq k) = \sum_{x=k}^{n} \binom{n}{x} p^x (1-p)^{n-x}$$

การทดสอบสมมุติฐานจะแบ่งได้เป็น 3 กรณี คือ ถ้าสามารถแสดงได้ว่า

1) $P(x \geq k) = (F > F_0)$

การทดสอบสมมุติฐานจะใช้สถิติ

$$F_0 = \frac{\nu_2(1-p)}{\nu_1 p} \qquad (3)$$

ซึ่งมีการแจกแจงเอฟ โดยมีขั้นความเป็นอิสระ $\nu_1 = 2(n-k+1)$ และ $\nu_2 = 2k$ ซึ่งนำมาใช้ในการทดสอบทางขวา (Right-Tail Test) ที่ระดับความมีนัยสำคัญ

2) $P(x \leq k) = 1 - P(x > k)$

และสามารถแสดงได้ว่า

$$P(x \leq k) = P(F < F_0).$$

การทดสอบสมมุติฐานจะใช้สถิติ

$$F_0 = \frac{\nu_2 p}{\nu_1(1-p)} \qquad (4)$$

ซึ่งจะมีการแจกแจงเอฟโดยมีขั้นความเป็นอิสระ $\nu_1 = 2(k+1)$, $\nu_2 = 2(n-k)$ จะนำมาใช้ในการทดสอบทางซ้าย (Left-Tail Test) ที่ระดับความมีนัยสำคัญ α

3) สำหรับการทดสอบสองทาง (Two-Tail Test) ก็ใช้ผลจากที่กล่าวมาแล้วข้างต้น คือ จากข้อมูลที่ได้มาให้หาสัดส่วนของกลุ่มตัวอย่าง $p = \frac{k}{n}$ ถ้า $p > P_0$ ใช้สถิติในสมการ (3) และหาค่าวิกฤตจากตารางที่ระดับความมีนัยสำคัญ $\alpha/2$ ถ้า $p < P_0$ ใช้สถิติในสมการ (4) และหาค่าวิกฤตจากตารางเอฟที่ระดับความมีนัยสำคัญ $\alpha/2$

ตัวอย่าง ในการทอดลูกเต๋าจำนวน 36 ครั้ง ปรากฏว่าหน้า 1 ขึ้น จำนวน 12 ครั้ง จะเป็นการยืนยันได้ไหมว่าความน่าจะเป็นที่จะปรากฏหน้า 1 มีมากกว่า $\frac{1}{4}$ [81]

สมมุติฐานที่ทดสอบ ; $H_0 : P = \frac{1}{4}$

$H_1 : P > \frac{1}{4}$

ค่าวิกฤต ที่ระดับความมีนัยสำคัญ $\alpha = .05$; ขั้นความเป็นอิสระ $\nu_1 = 24$, $\nu_2 = 50$

---

[81] Yamane, Op.cit., pp. 661-664.

คือ $F^{24}_{50} = 1.86$

แทนค่าในสมการ (3) ได้ $F = \frac{24(1-\frac{1}{4})}{50(\frac{1}{4})} = 1.44 < 1.86$

ดังนั้นจะคงสมมุติฐานสูญไว้ นั่นคือไม่สามารถยืนยันได้ว่ามีสัดส่วนที่จะปรากฏหน้า 12 มากกว่า $\frac{1}{4}$

ตัวอย่าง สมมุติว่าเปอร์เซนต์ของจำนวนข้อสอบที่ไม่ดีในการออกข้อสอบครั้งหนึ่งเป็น 5% เมื่อเลือกข้อสอบเป็นตัวอย่างแบบสุ่มมา 100 ข้อ และพบว่ามีข้อสอบที่ไม่ดีอยู่เพียง 3 ข้อ จะมีการลดเปอร์เซนต์ของข้องสอบที่เคยเชื่อว่าไม่ดีลงอีกหรือไม่

สมมุติฐานที่ทดสอบ ; $H_0 \; : \; p = 0.05$

$H_1 \; : \; P < 0.05$

ค่าวิกฤต ที่ระดับความมีนัยสำคัญ $\alpha = .05$; ขั้นความเป็นอิสระ $\nu_1 = 8, \nu_2 = 194$

คือ $F^{8}_{194} = 1.98$

แทนค่าในสมการ (4) $F = \frac{194(0.5)}{8(.95)} = 1.28 < 1.98$

ดังนั้นจะคงสมมุติฐานสูญไว นั่นคือยังไม่มีการลดเปอร์เซนต์ของข้อสอบที่ไม่ดีลงอีก

ตัวอย่าง จากข้อมูลที่ผ่านมาพบว่า นักเรียนร้อยละ 30 สนับสนุนพรรครีพับริกัน ต่อมามีเหตุการณ์บางอย่างที่เชื่อว่าจำนวนร้อยละนี้ควรจะเปลี่ยนไป จึงสุ่มตัวอย่างนักเรียนมา 50 คน ปรากฏว่าสนับสนุนพรรครีพับริกันเพียง 10 คน เท่านั้น

สมมุติฐานที่ทดสอบ ; $H_0 \; : \; P = 30\ \%$

$H_1 \; : \; p \neq 30\ \%$

ค่าวิกฤติ ที่ระดับความมีนัยสำคัญ $\alpha = .05$; ขั้นความเป็นอิสระ $\nu_1 = 22, \nu_2 = 80$

คือ $F^{22}_{80} = 1.75$

เนื่องจาก $p = \frac{k}{n} = \frac{10}{50} = 20\ \%$ ซึ่งน้อยกว่า 30 % ต้องใช้สูตร (3)

แทนค่าในสูตร (3) ได้ $F = \frac{80\ (0.3)}{22\ (0.7)} = 1.56 < 1.76$

ดังนั้นจะคงสมมุติฐานสูญไว้ นั่นคือจำนวนร้อยละของนักเรียนที่สนับสนุนพรรครีพับริกันยังไม่เปลี่ยนแปลง

2.5 การทดสอบสัมประสิทธิ์สหสัมพันธ์แบบพหุคูณของประชากร (R)

สัมประสิทธิ์สหสัมพันธ์แบบพหุคูณ (Multiple Correlation Coeeficient) คือสหสัมพันธ์แบบผลคูณระหว่างค่าต่าง ๆ ของตัวแปรอิสระ (Independent Variable) ซึ่งเกิดขึ้นในการถดถอยแบบพหุคูณ (Multiple Regression) กับค่าที่ได้จากสมการถดถอยแบบพหุคูณ หรือค่าตัวแปรตาม (Dependent Variable) สัมประสิทธิ์สหสัมพันธ์แบบพหุคูณเป็นสัมประสิทธิ์สหสัมพันธ์ในระหว่างตัวแปรตาม (Dependent Variable)กับชุดของตัวแปรอิสระที่มีจำนวนตัวแปรอิสระในแต่ละชุดตั้งแต่ 2 ตัวขึ้นไป ดังนั้นการทดสอบสมมุติฐานจะมีข้อกำหนดดังนี้

กลุ่มตัวอย่างขนาด n สุ่มมาจากประชากรที่แจกแจงปกติ ซึ่งประกอบด้วยตัวแปร Y และตัวแปร X_i ; (i=1,2---n) โดยที่ตัวแปร มีความสัมพันธ์กับชุดของตัวแปร X_i ซึ่งความสัมพันธ์นี้วัดได้ด้วยสัมประสิทธิ์สหสัมพันธ์แบบพหุคูณ $R_{Y(X_1,X_2\,---X_n)}$ การทดสอบสมมุติฐานจะใช้สถิติ

$$\frac{R^2/m}{(1-R^2)/(n-m-1)} \qquad (5)$$

เมื่อ R คือ สัมประสิทธิ์สหสัมพันธ์แบบพหุคูณ

m คือ จำนวนตัวแปรอิสระ

n คือ ขนาดของกลุ่มตัวอย่าง

ถ้าสมมุติฐานสูญ ; H_0 : R=0 จริง การแจกแจงของสถิติเอฟ สมการ (5) จะเป็นการแจกแจงเอฟ โดยมีชั้นความเป็นอิสระ $\nu_1 = m$, $\nu_2 = n-m-1$

ตัวอย่าง สัมประสิทธิ์สหสัมพันธ์แบบพหุคูณของคะแนนสอบไล่ คะแนนจิตวิทยา และคะแนนคณิตศาสตร์ ของนักเรียนจำนวน 15 คน เป็น .66 อยากทราบว่าคะแนนสอบไล่ คะแนนจิตวิทยาและคะแนนคณิตศาสตร์มีความสัมพันธ์กันจริงหรือไม่[82]

---

[82] Norman J. Bailey, Statistical Methods in Biology (London : The English University Press, Ltd.), p.121.

สมมุติฐานที่ทดสอบ $H_0 : R = 0$

$H_1 : R \neq 0$

ค่าวิกฤต ที่ระดับความมีนัยสำคัญ = .05 , ขั้นความเป็นอิสระ $\nu_1 = 2$, $\nu_2 = 12$

คือ $F^2_{12} = 3.88$

แทนค่าในสมการ (5) ได้ค่า $F = \frac{(.66)^2/2}{(1-.66^2)(15-2-1)} = 1.17 < 3.84$

ดังนั้นจะคงสมมุติฐานสูญ แสดงว่าคะแนนสอบไล่ คะแนนจิตวิทยาและคะแนนคณิตศาสตร์ ไม่มีความสัมพันธ์กันจริง

2.6 การทดสอบความแตกต่างระหว่างสัมประสิทธิ์สหสัมพันธ์แบบพหุคูณ ($R_1 - R_2$)

การทำนายเกณฑ์ (Criteria) ที่ใช้ตัวแปรอิสระทำนายร่วมกันมากกว่า 2 ตัวแปรขึ้นไปนั้น เราอาจจะลดตัวแปรอิสระที่ทำนายร่วมกันนี้ให้น้อยลงได้ ถ้าทราบว่าตัวแปรอิสระใดที่จำเป็นจริง ๆ ต่อการทำนายเกณฑ์นั้น ๆ เช่น แบบทดสอบทั้งฉบับ (Battery) จะมีแบบทดสอบย่อย ๆ (Sub Test) หลายฉบับด้วยกัน การทำแบบทดสอบทั้งฉบับจะต้องใช้เวลามาก บางครั้งเราไม่มีความจำเป็นที่จะต้องใช้แบบทดสอบทั้งฉบับ เราอาจจะตัดแบบสอบย่อยบางฉบับออกไปได้ ถ้าการตัดออกไปนี้ไม่ทำให้การทำนายเกณฑ์นั้นเปลี่ยนแปลงไป ดังนั้นการเปรียบเทียบความแตกต่างระหว่างการทำนายเกณฑ์เมื่อก่อนและหลังจากที่ลดจำนวนตัวแปรอิสระออกไป มีข้อกำหนดสำหรับการทดสอบสมมุติฐานทางสถิติดังนี้

กลุ่มตัวอย่างขนาด n สุ่มมาจากประชากรที่แจกแจงปกติซึ่งประกอบด้วยตัวแปรตาม Y และตัวแปรอิสระ $X_1, X_2 --- X_m$ ค่าสหสัมพันธ์ระหว่าง Y และชุดของตัวแปรอิสระ $X_1, X_2 --- X_m$ เป็น R_1 และค่าสหสัมพันธ์ระหว่าง Y กับชุดของตัวแปรอิสระ $X_1, X_2 --- X_m$ เป็น R_2 การเปรียบเทียบความแตกต่างระหว่างสัมประสิทธิ์สหสัมพันธ์แบบพหุคูณ ($R_1 - R_2$) จะใช้สถิติ

$$F = \frac{(R_1^2 - R_2^2)/k}{(1-R_1^2)/(n-m-1)} \qquad (6)$$

เมื่อ R_1, R_2 คือ สหสัมพันธ์แบบผลคูณก่อนลดจำนวนตัวแปรอิสระและภายหลังลดจำนวนตัวแปรอิสระ

k คือ จำนวนตัวแปรอิสระที่ลดลง

n คือ ขนาดของกลุ่มตัวอย่าง

m คือ จำนวนตัวแปรอิสระก่อนที่จะลดจำนวนลง

ถ้าความสามารถในการทำนายคงเดิม การแจกแจงของสถิติในสมการ (6) เป็นการแจกแจงเอฟ โดยมีชั้นความเป็นอิสระ $\nu_1 = k$, $\nu_2 = n-m-1$

ตัวอย่าง แบบสอบวิชาภาษาอังกฤษฉบับหนึ่งประกอบด้วยแบบสอบย่อย 4 ฉบับ นำไปทดสอบนักเรียนจำนวน 30 คน วัดค่าสัมประสิทธิ์สหสัมพันธ์แบบพหุคูณเป็น .70 และเนื่องจากเวลามีน้อย การทดสอบต้องลดจำนวนแบบสอบย่อย 1 ฉบับ วัดค่าสัมประสิทธิ์สหสัมพันธ์แบบพหุคูณได้ .62 อำนาจการพยากรณ์ของแบบสอบคงเดิมหรือไม่

สมมุติฐานที่ทดสอบ ; $H_0 : R_1 = R_2$

$H_1 : R_1 \neq R_2$

ค่าวิกฤตที่ระดับความมีนัยสำคัญ = .05 ชั้นความเป็นอิสระ $\nu_1 = 1$, $\nu_2 = 15$

คือ $F^{1}_{,15} = 4.24$

แทนค่าในสมการ (6) ได้ $F = \dfrac{(.49-.38)/1}{(.51)/25} = 5.4 > 4.24$

ดังนั้นเราจะปฏิเสธสมมุติฐานสูญ แสดงว่าความสามารถในการทำนายของแบบสอบเปลี่ยนไปอย่างมีนัยสำคัญ

2.7 การทดสอบอัตราส่วนสหสัมพันธ์ (Correlation Ratio)

ถ้าตัวแปร 2 ตัวแปร ไม่มีความสัมพันธ์ในเชิงเส้นตรง ตัวแปรทั้งสองอาจมีความสัมพันธ์เชิงเส้นโค้งได้ การวัดความสัมพันธ์ระหว่างตัวแปรจึงใช้อัตราส่วนสหสัมพันธ์ (Correlation Ratio) ซึ่งมีสัญญลักษณ์ η (eta)

อัตราส่วนสหสัมพันธ์ (η) คืออัตราส่วนในระหว่างความแปรปรวนภายในแถวลำดับต่าง ๆ และความแปรปรวนรวมยอดในตารางแจกแจงความถี่ชนิดตัวแปรผัน 2 ตัวแปร (Bivariate Frequency Table) ถ้าให้ X เป็นตัวแปรตาม Y เป็นตัวแปร

อิสระ อัตราส่วนสหสัมพันธ์ของ X จาก (X on Y) อาจนิยามได้ดังนี้

$$\eta^2_{XY} = \frac{\sum(\bar{X}_i - \bar{X})^2}{\sum(X-\bar{X})^2}$$

เมื่อ $\bar{X}_i$ เป็นมัชฌิมเลขคณิตของแถวลำดับที่ i และ $\bar{X}$ เป็นมัชฌิมเลขคณิตของ X ในการแจกแจงทั้งหมด การหาผลรวมในส่วน คิดจากค่าของ X ทั้งหมด แต่การหาผลรวมในเศษคิดเฉพาะค่า X ที่อยู่ในแถวลำดับ Y เป็นแถว ๆ ไป ในทำนองเดียวกัน ถ้าให้ Y เป็นตัวแปรตาม X เป็นตัวแปรอิสระ อัตราส่วนสหสัมพันธ์ของ Y จาก X คือ

$$\eta^2_{YX} = \frac{\sum(\bar{Y}_j - \bar{Y})^2}{\sum(Y-\bar{Y})^2}$$

เมื่อ $\bar{Y}_j$ เป็นมัชฌิมเลขคณิตของแถวลำดับที่ j และ $\bar{Y}$ เป็นมัชฌิมเลขคณิตของ Y ในการแจกแจงทั้งหมด การหาผลรวมในส่วนคิดจากค่าของ Y ทั้งหมด แต่การหาผลรวมในเศษคิดจากค่าของ Y เฉพาะที่อยู่ในแถวลำดับ X เป็นแถว ๆ ไป ซึ่งการทดสอบความมีนัยสำคัญทางสถิติมีข้อกำหนดดังนี้

กลุ่มตัวอย่างขนาด n สุ่มมาจากประชากรที่แจกแจงปกติแบบสองตัวแปร (Bivariate Normal Distribution) วัดความสัมพันธ์ระหว่างตัวแปรในประชากรเป็น η และวัดความสัมพันธ์ระหว่างตัวแปรในกลุ่มตัวอย่างเป็น E ข้อมูลที่ได้จากตัวแปรทั้งสองจะต้องเป็นหมวดหมู่ (Groupted Data) หรือ อย่างน้อยที่สุดตัวแปรอิสระจะต้องจัดเป็นหมวดหมู่ได้[83] การทดสอบสมมุติฐานใช้สถิติ

$$F = \frac{E^2/k-1}{(1-E^2)/n-k} \qquad (7)$$

เมื่อ E คือ อัตราส่วนสหสัมพันธ์ระหว่างตัวแปรในกลุ่มตัวอย่าง

n คือ ขนาดของกลุ่มตัวอย่าง

k คือ จำนวนแถว หรือ จำนวนคอลัมน์

---

[83] Croxton and Cowden, Op. cit., p.452.

ถ้าสมมุติฐานสูญ ; $H_0 : \eta = 0$ จริง การแจกแจงของสถิติเอฟ ในสมการ (7) จะเป็นการแจกแจงเอฟโดยมีขั้นความเป็นอิสระ $\nu_1 = k-1$, $\nu_2 = n-k$

ตัวอย่าง จากข้อมูลที่แจกแจงความถี่แบบสองตัวแปร พบว่าความสัมพันธ์ระหว่างความสูงและน้ำหนักของนักเรียนจำนวน 106 คน มีค่าอัตราส่วนสหสัมพันธ์เป็น 0.117 จงทดสอบความมีนัยสำคัญของอัตราส่วนสหสัมพันธ์ของประชากร เมื่อ $k = 5$[84]

สมมุติฐานที่ทดสอบ ; $H_0 : \eta = 0$

$H_1 : \eta \neq 0$

ค่าวิกฤติ เมื่อกำหนดระดับความมีนัยสำคัญ $\alpha = .05$ ขั้นความเป็นอิสระ $\nu_1 = 4$, $\nu_2 = 101$ คือ $F^{4}_{106} = 2.47$

แทนค่าในสมการ (7) ได้ $F = \dfrac{(0.117)(101)}{0.883(4)} = 3.35 > 2.47$

ดังนั้นปฏิเสธสมมุติฐานสูญ แสดงว่าอัตราส่วนสหสัมพันธ์ของประชากรแตกต่างจากศูนย์อย่างมีนัยสำคัญ นั่นคือ ความสูงและน้ำหนักมีความสัมพันธ์เชิงเส้นโค้ง

2.8 การทดสอบภาวะเชิงเส้นของความถดถอย (Linearity of Regression)

การทดสอบภาวะเชิงเส้น (Linearity Test) คือการทดสอบดูว่าความเกี่ยวข้องสัมพันธ์ในระหว่างค่าต่าง ๆ ของตัวแปรชุดหนึ่งในพิสัยที่กำหนดให้จะเป็นความเกี่ยวข้องสัมพันธ์เชิงเส้นตรงหรือไม่ ในกรณีที่ตัวเปลี่ยนแปรมีเพียง 2 ตัว

เนื่องจากอัตราส่วนสหสัมพันธ์ซึ่งบอกถึงความมากน้อยของสหสัมพันธ์ (Degree of Correlation) และสัมประสิทธิ์สหสัมพันธ์เป็นตัวชี้ระดับความสัมพันธ์เชิงเส้นตรง (Degree of Linear Relationship) แต่ในบางครั้งแม้ว่าความถดถอยเชิงเส้นของ Y จาก X (Y on X) ในประชากรจะมีลักษณะเป็นเส้นตรง แต่มัชฌิมเลขคณิตของแถวหรือของคอลัมน์ก็ยังเบี่ยงเบนจากเส้นตรง ด้วยเหตุนี้ อัตราส่วนสหสัมพันธ์มักจะมีค่า

---

[84] Theodore R. Anderson and Monis Zelditch, Jr. A Basic Course in Statistics with Sociological Applications (New York : Holt, Rinehart and Winston, Inc., C 1968), p. 279.

มากกว่าสัมประสิทธิ์สหสัมพันธ์ ($\eta^2 > \rho^2$) ดังนั้นความแตกต่างระหว่างอัตราส่วนสหสัมพันธ์กับสัมประสิทธิ์สหสัมพันธ์ ($\eta^2 - \rho^2$) จะชี้ให้เห็นว่า ความสัมพันธ์นั้นเป็นความสัมพันธ์เชิงเส้นตรง (Linearity) หรือไม่ โดยที่ $\eta^2 - \rho^2 = 0$ แสดงว่าความสัมพันธ์นั้นเป็นความสัมพันธ์เชิงเส้นตรง

การทดสอบสมมุติฐานทางสถิติจะมีข้อกำหนดดังนี้คือ

กลุ่มตัวอย่างขนาด n สุ่มมาจากประชากรที่แจกแจงปกติแบบสองตัวแปร โดยที่ข้อมูลทั้งสองจะต้องจัดเป็นหมวดหมู่ (Groupted Data) เมื่อคำนวณค่าอัตราส่วนสหสัมพันธ์ของกลุ่มตัวอย่าง เป็น E และสหสัมพันธ์ของกลุ่มตัวอย่างเป็น r การทดสอบสมมุติฐานจะใช้สถิติ

$$F = \frac{(E^2 - r^2)/(k-R)}{(1-E^2)/(n-k)} \qquad (8)$$

เมื่อ E และ r คืออัตราส่วนสหสัมพันธ์ และสหสัมพันธ์โพรดักโมเมนต์ของตัวแปร

k คือ จำนวนแถวหรือจำนวนคอลัมน์

n คือ ขนาดของกลุ่มตัวอย่าง

ถ้าสมมุติฐานสูญ ; $H_0 : \eta^2 - \rho^2 = 0$ จริง การแจกแจงของสถิติเอฟในสมการ (8) จะเป็นการแจกแจงเอฟ โดยมีชั้นความเป็นอิสระ $\nu_1 = k-2$, $\nu_2 = n-k$

ตัวอย่าง การแจกแจงแบบสองตัวแปร (Bivariate Distribution) ของคะแนนสอบไล่และคะแนนความคิดริเริ่มของนักเรียนจำนวน 92 คน ประกอบด้วยจำนวนคอลัมน์ของตารางการแจกแจงเท่ากับ 7 วัดค่าอัตราส่วนสหสัมพันธ์เป็น .639 และค่าสัมประสิทธิ์สหสัมพันธ์เป็น .5687 การทำนายหรือพยากรณ์คะแนนสอบไล่ เมื่อใช้ค่าความคิดริเริ่มจะเป็นเส้นตรงหรือไม่

สมมุติฐานที่ทดสอบ ; $H_0 : \eta^2 - \rho^2 = 0$

$H_1 : \eta^2 - \rho^2 \neq 0$

ค่าวิกฤต ที่ระดับความมีนัยสำคัญ .05 ชั้นความเป็นอิสระ $\nu_1 = 5$, $\nu_2 = 85$ จะได้ค่าเอฟจากตารางเอฟเท่ากับ $F^{5}_{,85} = 4.42$

แทนค่าในสมการ (8) ได้ค่า $F = \frac{(.639^2 - .568^2)/(7-2)}{(1-.639^2)/(92-7)} = 2.5 < 4.42$

ดังนั้นจะ คง เสมมุติฐานสูญ แสดงว่าความสัมพันธ์ระหว่างคะแนนสอบไล่และคะแนนความคิดริเริ่มเป็นเส้นตรง

การวิเคราะห์แนวโน้ม (Trend Analysis)

การวิเคราะห์แนวโน้มของข้อมูลใด ๆ นั้น มักจะต้องอาศัยโพลีโนเมียลฟังก์ชัน (Poly Nomial Function) ดังเช่น

แอนเดอร์สัน[85] (T.W.Anderson) กล่าวว่าเมื่อไม่มีทฤษฎีใดมาบ่งชัดได้ว่าแนวโน้มของข้อมูลนั้นเป็นแบบใดแล้ว เราก็สามารถที่จะประมาณได้โดย ทำเป็นแบบโพลีโนเมียลกำลังต่ำ ๆ แบบที่ง่ายที่สุดคือสมการโพลีโนเมียลกำลังหนึ่ง ซึ่งเป็นแบบที่มีการเปลี่ยนแปลงคงที่โดยตลอด

โฮล[86](Paul G. Hoel) ก็ได้กล่าวในทำนองเดียวกันว่า ถ้าไม่มีเหตุผลจากทฤษฎีใด ๆ มากล่าวได้ว่าโค้งความสัมพันธ์เป็นรูปใดได้อย่างแน่นอนแล้วก็มักจะเลือกใช้ในแบบโพลีโนเมียล เนื่องจากสามารถทำได้ง่ายกว่าและมีความยืดหยุ่นได้มากกว่าวิธีอื่น ๆ ส่วนการที่จะใช้สมการโพลีโนเมียลกำลังเท่าใดนั้นจะตัดสินได้โดยสังเกตจากการทำตารางกระจัดกระจาย (Scatter Diagram) และเมื่อเลือกกำลังของโพลีโนเมียลได้แล้ว ก็นำวิธีกำลังสองน้อยที่สุด (Least Square Method) มาใช้ในการประมาณค่าสัมประสิทธิ์ของสมการกำลังนั้น ๆ ได้

สมการโดยทั่วไปของโพลีโนเมียลกำลัง k คือ

$$Y = b_0 + b_1X + \cdots + b_kX^k$$

ถ้าเป็นโพลีโนเมียลกำลังหนึ่ง หมายถึงค่าสัมประสิทธิ์ตั้งแต่ $b_2, b_3, \cdots, b_k$

---

[85] T.W. Anderson, "Trends and Smoothing", The Statistical Analysis of Time Series, (New York : John Wiley and Sons, Inc., C1971), p. 31.

[86] Paul G. Hoel, "Curvilinear Regression", Introdcution to Mathematical Statistics (New York : John Wiley and Sons, Inc., c 1962), p. 175.

มีค่าเป็น o หมด ลักษณะของเส้นกราฟที่ได้จะเป็นเส้นตรง

ถ้าเป็นโพลีโนเมียลกำลังสอง หมายถึงค่าสัมประสิทธิ์ตั้งแต่ $b_3, b_4, ---, b_k$ มีค่าเป็น o หมด ลักษณะของเส้นกราฟที่ได้จะเป็นเส้นโค้งครึ่งหนึ่ง เรียกว่า พาราโบลา (Parabola)

ถ้าเป็นโพลีโนเมียลกำลังสามจะโค้งสองครั้ง กำลังสี่จะโค้งสามครั้ง เพิ่มขึ้นเป็นลำดับไป สรุปได้ว่าจำนวนโค้งจะน้อยกว่ากำลังของสมการอยู่ 1 เสมอ ดังรูป

$Y = b_0 + b_1X$ $\quad$ $Y = b_0 + b_1X + b_2X^2$ $\quad$ $Y = b_0 + b_1X + b_2X^2 + b_3X^3$ $\quad$ $Y = b_0 + b_1X + b_2X^2 + b_3X^3 + b_4X^4$

ในการนำรูปแบบโพลีโนเมียล (Polynomial Model) มาใช้นั้นแฟรงกลิน เอ เกรย์บิลล์[87] (Franklin A.Grayhill) กล่าวว่ามีอยู่ 2 กรณี คือ

1. เมื่อต้องการประมาณค่าสัมประสิทธิ์ของสมการโพลีโนเมียลกำลัง P หรือต่ำกว่า โดยใช้วิธีกำลังสองน้อยที่สุด (Least-Squares Technique) หรือวิธีแมกซิมั่มไลค์ลิฮูด (Maximum Likelihood Methods) ในเมื่อทราบโดยทางทฤษฎีหรือด้วยเหตุผลอื่น ๆ ว่าข้อมูลนั้นมีลักษณะเป็นโพลีโนเมียลกำลัง P หรือต่ำกว่า

2. ในกรณีที่ไม่ทราบว่าข้อมูลนั้นจะใช้ฟังก์ชันกำลังใดจึงจะเหมาะสมก็อาจจะหาได้โดยทำเป็นโพลีโนเมียลกำลังต่ำ ๆ ที่เหมาะสมกับข้อมูลนั้น

การหาโพลีโนเมียลฟังก์ชันของข้อมูลใดมีวิธีการอยู่ 3 ขั้นดังนี้

1. หากำลังของโพลีโนเมียลฟังก์ชันที่เหมาะสมกับข้อมูลนั้น

2. ประมาณค่าสัมประสิทธิ์ของโพลีโนเมียลกำลังนั้น ๆ

---

[87] Franklin A. Grayhill, Polynomial or Curvilinear Models, An Introduction to Liniar Statistical Models (New York : McGraw-Hill Book Company, Inc., 1961), p. 165.

3. ทดสอบค่าสัมประสิทธิ์ค่าต่าง ๆ ในโพลีโนเมียลกำลังนั้น ๆ

การหากำลังของโพลีโนเมียลฟังก์ชันที่เหมาะสมกับข้อมูลนั้นกระทำได้ 2 วิธี วิธีที่หนึ่งคือลองทำโพลีโนเมียลกำลังหนึ่งคือ $Y = \alpha_0 + \alpha_1 X + e$ แล้วทดสอบสมมุติฐานว่า $\alpha_1 = 0$ โดยวิธีวิเคราะห์ความแปรปรวน (Analysis of Variance) ถ้าทดสอบได้ว่า $\alpha_1 = 0$ จะสรุปได้ว่า $Y = \alpha_0 + e$ นั้นเป็นฟังก์ชันอันที่เหมาะสมกับข้อมูลมากที่สุด แต่ ถ้าได้ผลว่า $\alpha_1 \neq 0$ จะต้องทำโพลีโนเมียลกำลังสอง คือ $Y = \alpha_0 + \alpha_1 X + \alpha_2 X^2 + e_2$ แล้วทดสอบสมมุติฐานว่า $\alpha_2 = 0$ โดยวิธีวิเคราะห์ความแปรปรวน ถ้าทดสอบได้ว่า $\alpha_2 = 0$ จะสรุปได้ว่า $Y = \alpha_0 + \alpha_1 X + e$ เหมาะสมกับข้อมูลมากที่สุด แต่ถ้าได้ผลได้ $\alpha_2 \neq 0$ จะต้องทำโพลีโนเมียลกำลังสามต่อไปอีก ทำดังนี้จนกว่าจะสรุปได้ว่า ฟังก์ชันใดเหมาะสมกับข้อมูลมากที่สุด การใช้วิธีนี้พบว่าการคำนวณจะยุ่งยากมากขึ้นตามลำดับ เมื่อกำลังของฟังก์ชันมากขึ้น ดังนั้นจึงมีวิธีการที่ทำให้การคำนวณง่ายและสั้นเข้าโดยวิธีออโทโกนัลโพลีโนเมียล (Orthogonal Polynomial) ซึ่งเป็นวิธีการหากำลังของโพลีโนเมียลอีกวิธีหนึ่ง ซึ่งสูตรและวิธีวิเคราะห์ความแปรปรวนแบบออโทโกนัลโพลีโนเมียล แสดงไว้ในตารางที่ 4

ตารางที่ 4 ตารางการวิเคราะห์ความแปรปรวนด้วยวิธีออโธโกนัล
โพลีโนเมียล

SV	D.F	SS	M.S	F
Mean	1	$(T)^2/t_n$	M	...
Treatments	t-1	$\sum T_X^2/t$	T	T/E
T_L	1	$(T_L)_{YY}$	T_L	$T_{L/E}$
T_Q	1	$(T_Q)_{YY}$	T_Q	$T_{Q/E}$
T_C	1	$(T_C)_{YY}$	T_C	$T_{C/E}$
T_{DEV}	t-4	$(T_{DEV})_{YY}$	T_{DEV}	$T_{DEV/E}$
Experimental Error	$t_{(n-1)}$	E_{YY}	E	.....
Total	t_n	$\sum Y_i^2$		

เมื่อ SV = แหล่งของความแปรปรวน (Source of Variation)

DF = ขั้นความเป็นอิสระ

SS = ผลบวกของส่วนเบี่ยงเบนกำลังสอง (Sum of Square)

MS = ส่วนเบี่ยงเบนกำลังสองเฉลี่ย (Mean Square) = SS/df.

T = ผลรวมตัวอย่างทั้งหมด

T_i = ผลรวมในกลุ่มตัวอย่างที่ i

t = จำนวน Treatment

n_i = จำนวนตัวอย่างในกลุ่มตัวอย่างที่ i

i = กลุ่มตัวอย่างที่ i

Y_i = ตัวอย่างที่ i ใน Treatment

ดังนั้นเราจะได้ผลบวกของส่วนเบี่ยงเบนกำลังสองจากแหล่งความแปรปรวนอันเนื่องมาจาก

กำลังต่าง ๆ ของฟังก์ชัน เท่ากับ $\dfrac{\sum_{i=1}^{t} T_i \quad P_{iq}^{2}}{n \sum_{i=1}^{t} P_{iq}^{2}}$, $q = 1, --- t-1$

P_{iq} = ค่าสัมประสิทธิ์ของออโธโกนัลโพลีโนเมียลซึ่งหาได้จากตาราง

T_{DEV} = ผลบวกของส่วนเบี่ยงเบนกำลังสองจากแหล่งความแปรปรวนที่เนื่องมาจากความคลาดเคลื่อนกำลังต่าง ๆ ของฟังก์ชัน $\left[\sum X_i^2 - T^2/t_n\right]$ – (ผลบวกของส่วนเบี่ยงเบนกำลังสองจากแหล่งความแปรปรวนอันเนื่องมาจากกำลังต่าง ๆ ของฟังก์ชัน)

F = ค่าทดสอบเอฟ (F-Test)

ในการทดสอบค่า F ที่คำนวณได้กับค่าใ F ในตาราง เมื่อพบว่าค่า F มีนัยสำคัญที่ฟังก์ชันใด จะสรุปได้ว่าฟังก์ชันกำลังนั้น ๆ เหมาะสมกับข้อมูลมากที่สุด แต่การเป็นค่านัยสำคัญที่มีค่า F ของฟังก์ชันกำลังถัดขึ้นไปนั้นไร้นัยสำคัญติดต่อกัน ๒ ค่า จึงจะสรุปได้ว่าฟังก์ชันกำลังนั้น ๆ เหมาะสมกับข้อมูลมากที่สุด การทดสอบสมมุติฐานมีข้อกำหนดว่า ระดับแต่ละระดับของ Treatment จะต้องมีลักษณะเป็น Equally Spaced คือมีช่วงเท่ากัน

ตัวอย่าง การทดลองสอนวิชาวิทยาศาสตร์โดยใช้สไลด์ประกอบการสอนเป็นระยะเวลาต่าง ๆ กัน ปรากฏผลดังนี้

	ระยะเวลาที่ใช้สไลด์ประกอบการสอน				
	0	5	10	15	20
	20	25	36	35	43
	25	29	37	39	40
	23	31	29	31	36
	27	30	40	42	48
	19	27	33	44	47
คะแนนรวม	114	142	175	191	214
มัชฌิมเลขคณิต	22.8	28.4	35	38.2	42.8

H_0 : ลักษณะความสัมพันธ์ของข้อมูลเป็นลักษณะใด

H_1 : ลักษณะความสัมพันธ์ของข้อมูลเป็น Linear, Quadratic, Cubic----

ดังนั้นเราจะได้ผลของส่วนเบี่ยงเบนกำลังสองจากแหล่งความแปรปรวนอันเนื่องมาจากกำลังต่าง ๆ ของฟังก์ชันคือ

$$T_L = \frac{(-2)(114)+(-1)(142)+(0)(175)+ (1)(191) +(2)(214)^{\ 2}}{5\ (-2)^2+ (-1)^2+ (0)^2+ (-1)^2+ (2)^2}$$

$$= 1240.02$$

$$T_Q = \frac{(2)(114)+ (-1)(142)+ (-2)(175) +(-1)(191) +(2)(214)^{\ 2}}{5\ (2)^2+(-1)^2+(-2)^2+(-1)^2 +(2)^2}$$

$$= 10.41$$

$$\text{Mean} = (836)^2/25 = 27,955.84$$

$$\text{Treatment} = (114)^2+(142)^2+(175)^2 +(191)^2 +(214)^2 \ /5-27,955.84$$

$$= 1256.56$$

$$\text{Experimental Error} = 29,560-27,955.85/1256.56 = 347.60$$

$$\therefore T_{DEV} = 1256.56-1240.02 - 10.41 = 6.13$$

ซึ่งแสดงด้วยตารางได้ดังนี้คือ

SV	D F	Sum of Square	Mean Square	F
Total	25	29,560.00	29,560.00	....
Mean	1	27,955.84	27,955.84	....
Treatment	4	1,256.56	314.14	18.07
T_L	1	1,240.02	1,240.02	71.35*
T_Q	1	10.41	10.41	0.60
T_{DEV}	2	6.13	3.07	0.18
Experimental Error	20	347.60	17.38	

ค่า F จากตารางมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าความสัมพันธ์มีลักษณะเป็นเส้นตรง นั่นคือ ข้อมูลชุดนี้มีลักษณะการเปลี่ยนแปลงเป็นเส้นตรง

ความสัมพันธ์ระหว่างการแจกแจงทีและการแจกแจงเอฟ

เนื่องจากการแจกแจงทีและการแจกแจงเอฟ เป็นการแจกแจงซึ่งได้มาจากการแจกแจงปกติ ดังนั้นก่อนที่จะศึกษาความสัมพันธ์ระหว่างการแจกแจงทีและการแจกแจงเอฟ ก็ควรจะศึกษาถึงการแจกแจงที่ได้มาจากการแจกแจงปกติ อันได้แก่ การแจกแจงที การแจกแจงเอฟ และการแจกแจงไคสแควร์

1. การแจกแจงไคสแวร์

ถ้าให้ x เป็นตัวแปรสุ่มซึ่งมีการแจกแจงปกติโดยมีมัชฌิมเลขคณิต μ และความแปรปรวน σ^2 ดังนั้นตัวแปร $z = \frac{\bar{x} - \mu}{\sigma}$ จะแจกแจงปกติมาตรฐานโดยมีมัชฌิมเลขคณิต $\mu = 0$ และความแปรปรวน $\sigma^2 = 1$ ดังนั้น $z^2 = \frac{(\bar{x} - \mu)^2}{\sigma^2}$ จะมีการแจกแจงไคสแควร์ โดยมีขั้นความเป็นอิสระ $\nu = 1$ เราจะนิยามตัวแปร $\chi^2_{(\nu)}$ โดยให้

$$\chi^2_{(1)} = z^2 = \frac{(x - \mu)^2}{\sigma^2} \quad \text{เมื่อ } \nu = 1$$

$$\chi^2_{(2)} = z_1^2 + z_2^2 = \frac{(x_1 - \mu)^2}{\sigma^2} + \frac{(x_2 - \mu)^2}{\sigma^2} \quad \text{เมื่อ } \nu = 2$$

$$\text{ดังนั้น } \chi^2_{(n)} = \sum_{i=1}^{n} z_i^2 = \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^2} \quad \text{เมื่อ } \nu = n$$

ในทางปฏิบัติ เรามักไม่ทราบค่า μ ดังนั้นถ้าใช้ $\bar{x}$ แทนจะได้

$$\chi^2_{(n-1)} = \frac{(n-1)s^2}{\sigma^2} \quad \text{เมื่อ } \nu = n-1$$

2. การแจกแจงที

ถ้าให้ z มีการแจกแจงปกติมาตรฐาน (Standard Normal Distribution) และ $\chi^2_{(\nu)}$ มีการแจกแจงไคสแควร์ (χ^2-Distribution) ซึ่งมีขั้นความเป็นอิสระ (ν) เราจะนิยามตัวแปร t ได้โดยให้

$$t = \frac{z}{\sqrt{\chi^2_{(\nu)}/\nu}}$$

---

[88] ดูภาคผนวก จ.

นั่นคือ t คืออัตราส่วนระหว่างการแจกแจงปกติมาตรฐาน กับรากที่สองของผลหารระหว่างการแจกแจงไคสแควร์กับขั้นความเป็นอิสระ (ν)

ความสัมพันธ์ระหว่างการแจกแจงทีและการแจกแจงไคสแควร์

เนื่องจากการแจกแจงทีคืออัตราส่วนระหว่างการแจกแจงปกติ กับรากที่สองของผลหารระหว่างการแจกแจงอิสระไคสแควร์ กับขั้นความเป็นอิสระ ดังนั้น

$$t_{(\nu)} = \frac{Z}{\sqrt{\chi^2(\nu)/\nu}} \quad , \quad \nu = \text{ขั้นความเป็นอิสระ}$$

แต่ $Z^2 = \chi^2_{(1)}$ $\therefore t_{(\nu)} = \dfrac{\chi^2(1)}{\sqrt{\chi^2(\nu)/\nu}}$

นั่นคือ $t_{(\infty)} = \dfrac{\chi^2(1)}{\sqrt{\chi^2(\infty)/\infty}}$

แต่ $\lim\limits_{\nu \to \infty} \chi^2(\nu)/\nu = 1$

$$\therefore t_{(\infty)} = \chi^2_{(1)}$$

นั่นคือการแจกแจงทีโดยมีขั้นความเป็นอิสระ $\nu = \infty$ จะเท่ากับการแจกแจงไคสแควร์ซึ่งมีขั้นความเป็นอิสระ $\nu = 1$

ความสัมพันธ์ระหว่างการแจกแจงทีและการแจกแจงปกติ

เนื่องจากการแจกแจงทีคืออัตราส่วนระหว่างการแจกแจงปกติ กับรากที่สองของผลหารระหว่างการแจกแจงอิสระไคสแควร์ กับขั้นความเป็นอิสระ นั่นคือ

$$t_{(\nu)} = \frac{Z}{\sqrt{\chi^2(\nu)/\nu}}$$

แต่ $\lim\limits_{\nu \to \infty} \chi^2(\nu)/\nu = 1$

ดังนั้น $t_{(\infty)} = \dfrac{Z}{\sqrt{\chi^2(\infty)/\infty}} = Z$

นั่นคือ การแจกแจงทีโดยมีขั้นความเป็นอิสระ $\nu = \infty$ จะเท่ากับการแจกแจง

ปกติมาตรฐาน

3. การแจกแจงเอฟ (F-Distribution)

ถ้าตัวแปรสุ่ม X_1 และ X_2 มีการแจกแจงปกติโดยมีมัชฌิมเลขคณิต μ_1 และ μ_2 ความแปรปรวน σ_1^2 และ σ_2^2 สุ่มกลุ่มตัวอย่างขนาด n_1 และ n_2 มาจาก X_1 และ X_2 โดยที่ความแปรปรวนของกลุ่มตัวอย่าง s_1^2 และ s_2^2 เป็นค่าประมาณที่ไม่เอนเอียงของ σ_1^2 และ σ_2^2 ตามลำดับ จะได้ความสัมพันธ์

$$s_1^2 = \frac{\sigma_1^2 \chi^2(n_1-1)}{n_1-1}$$

$$s_2^2 = \frac{\sigma_2^2 \chi^2(n_2-1)}{n_2-1}$$

ซึ่งจะนิยามตัวแปร F โดยให้ $F = s_1^2/s_2^2$ ภายใต้สมมุติฐาน $\sigma_1^2 = \sigma_2^2$ ดังนั้น

$$F = \frac{\chi^2(n_1-1)/(n_1-1)}{\chi^2(n_2-1)/(n_2-1)} \quad \text{หรือ} \quad F_{\nu_1,\nu_2} = \frac{\chi^2(\nu_1)/\nu_1}{\chi^2(\nu_2)/\nu_2}$$

โดยที่ $\nu_1 = n_1 - 1$, $\nu_2 = n_2 - 1$

ความสัมพันธ์ระหว่างการแจกแจงเอฟและการแจกแจงไคสแควร์

เนื่องจากการแจกแจงเอฟคือ อัตราส่วนระหว่างความแปรปรวนของกลุ่มตัวอย่าง 2 กลุ่ม และถ้า s_2^2 ซึ่งเป็นความแปรปรวนของกลุ่มตัวอย่างที่สอง มีขั้นความเป็นอิสระ $\nu_2 = \infty$ ทำให้ $s_2^2 = \sigma_2^2$ ดังนั้นเราจะได้

$$F = \frac{s_1^2}{\sigma_2^2}$$

และเพราะว่า s_1^2 มีการแจกแจงไคสแควร์ซึ่งมีขั้นความเป็นอิสระ $\nu_1 = n_1 - 1$ เราจะได้ความสัมพันธ์ $s_1^2 = \sigma_1^2 \chi^2(\nu_1)/\nu_1$

ดังนั้น
$$F = \frac{\sigma_1^2 \chi^2(\nu_1)/\nu_1}{\sigma_2^2}$$

จากข้อตกลงที่ว่า $\sigma_1^2 = \sigma_2^2$ จะทำให้

$$F = \chi^2(\nu_1)/\nu_1$$

$$\therefore \quad F_{(\nu_1,\infty)} = \frac{\chi^2(\nu_1)}{\nu_1}$$

นั่นคือ $F_{(1,\infty)} = \chi^2_{(1)}$

นั่นคือ การแจกแจงเอฟโดยมีขั้นความเป็นอิสระ $\nu_1 = 1$, $\nu_2 = \infty$ จะเท่ากับการแจกแจงไคสแควร์ ซึ่งมีขั้นความเป็นอิสระ $\nu = 1$

ความสัมพันธ์ระหว่างการแจกแจงเอฟและการแจกแจงปกติ

เนื่องจากการแจกแจงเอฟ โดยมีขั้นความเป็นอิสระ $\nu_1 = 1$, $\nu_2 = \infty$ เท่ากับการแจกแจงไคสแควร์ เมื่อขั้นความเป็นอิสระ $\nu = 1$ นั่นคือ

$$F_{(1,\infty)} = \chi^2_{(1)}$$

แต่การแจกแจงไคสแควร์โดยมีขั้นความเป็นอิสระ $\nu = 1$ จะเท่ากับกำลังสองของการแจกแจงปกติมาตรฐาน ดังนั้นเราจะได้

$$F_{(1,\infty)} = \chi^2_{(1)} = Z^2$$

นั่นคือ การแจกแจงเอฟ โดยมีขั้นความเป็นอิสระ $\nu_1 = 1$, $\nu_2 = \infty$ จะเท่ากับกำลังสองของการแจกแจงปกติมาตรฐาน

ความสัมพันธ์ระหว่างการแจกแจงทีและการแจกแจงเอฟ

เนื่องจากการแจกแจงทีมีความสัมพันธ์กับการแจกแจงปกติและการแจกแจงไคสแควร์

นั่นคือ $$t_{(\nu)} = \frac{Z}{\sqrt{\chi^2(\nu)/\nu}}$$

ดังนั้น $$t^2_{(\nu)} = \frac{Z^2}{\chi^2(\nu)/\nu}$$

$$\therefore \quad Z^2 = \chi^2_{(1)}$$

$$\therefore \quad t^2_{(\nu)} = \frac{\chi^2(1)/1}{\chi^2(\nu)/\nu} = F^1_{\nu}$$

นั่นคือการแจกแจงเอฟ โดยมีขั้นความเป็นอิสระ $\nu_1 = 1$, $\nu_2 = \nu$ จะเท่ากับกำลังสองของการแจกแจงทีซึ่งมีขั้นความเป็นอิสระ $\nu = n - 1$

ความสัมพันธ์ของการแจกแจงปกติ การแจกแจงที การแจกแจงไคสแควร์ และการแจกแจงเอฟ

เนื่องจากการแจกแจงไคสแควร์ โดยมีขั้นความเป็นอิสระ $\nu = 1$ จะเท่ากับกำลังสองของการแจกแจงปกติมาตรฐาน และ

การแจกแจงที โดยมีขั้นความเป็นอิสระ $\nu = \infty$ จะเท่ากับการแจกแจงปกติมาตรฐาน และ การแจกแจงเอฟ โดยมีขั้นความเป็นอิสระ $\nu_1 = 1, \nu_2 = \infty$ จะเท่ากับกำลังสองของการแจกแจงปกติมาตรฐาน

ดังนั้นเราจะได้ความสัมพันธ์

$$\chi^2_{(1)} = Z^2 = t^2_{(\infty)} = F^1_{\infty}$$

นั่นคือการแจกแจงเอฟโดยมีขั้นความเป็นอิสระ $\nu_1 = 1, \nu_2 = \infty$ จะเท่ากับการแจกแจงไคสแควร์ที่มีขั้นความเป็นอิสระ $\nu = 1$ กำลังสองของการแจกแจงปกติมาตรฐาน และเท่ากับกำลังสองของการแจกแจงที โดยมีขั้นความเป็นอิสระ $\nu = \infty$

การเปรียบเทียบการแจกแจงที (t-Distribution) และการแจกแจงเอฟ (F-Distribution)

การเปรียบเทียบการแจกแจงทีและการแจกแจงเอฟ จะแบ่งเป็น 2 กรณี คือ

ก) ความคล้ายคลึงกันของการแจกแจงทีและการแจกแจงเอฟ

ข) ความแตกต่างกันของการแจกแจงทีและการแจกแจงเอฟ

ก.ความคล้ายคลึงกันระหว่างการแจกแจงทีและการแจกแจงเอฟ

1. ข้อตกลงเบื้องต้น : การแจกแจงทีและการแจกแจงเอฟ ต่างก็เป็น Derived Distribution ได้มาจากการแจกแจงปกติ(Normal Distribution) ดังนั้นจึงต้องมีข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงปกติ คือกลุ่มตัวอย่างจะต้องสุ่มมาจากประชากรที่แจกแจงปกติ ซึ่งกลุ่มตัวอย่างมีขนาดใหญ่ การแจกแจงทีและการแจกแจงเอฟ จะยิ่งเข้าใกล้การแจกแจงปกติ และความแปรปรวนของประชากรจะต้องได้จากการประมาณค่าความแปรปรวนของกลุ่มตัวอย่างเสมอ ซึ่งในการทดสอบสมมุติฐาน และ/หรือ ประมาณค่าความแตกต่างระหว่างพารามิเตอร์ของประชากร 2 กลุ่มนั้น ความแปรปรวนของประชากร

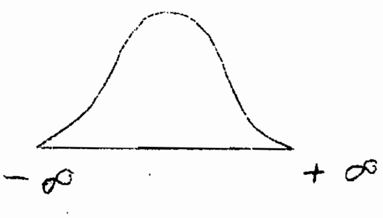
จะต้องเท่ากันด้วย

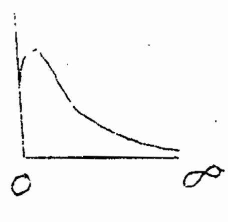
2. คุณสมบัติการแจกแจง : ทั้งการแจกแจงทีและการแจกแจงเอฟต่างก็เป็นอิสระจากความแปรปรวนของประชากร เพราะการแจกแจงทั้งสองต่างขึ้นอยู่กับพารามีเตอร์ตัวเดียวกันคือขั้นความเป็นอิสระเท่านั้น จึงทำให้โค้งการแจกแจงมีการกระจายหลายแบบตามจำนวนขั้นความเป็นอิสระนั้น ๆ แต่พื้นที่ใต้โค้งการแจกแจงทีและการแจกแจงเอฟจะเท่ากับ 1 เสมอ

3. การนำไปใช้ : การแจกแจงเอฟที่มีขั้นความเป็นอิสระ ($\nu_1 = 1$, $\nu_2 = n-1$) จะเท่ากับกำลังสองของการแจกแจงทีที่มีขั้นความเป็นอิสระ $\nu = n - 1$ ดังนั้นในการทดสอบสมมุติฐานเกี่ยวกับพารามีเตอร์จึงสามารถใช้แทนกันได้ และผลที่ได้ก็จะตรงกัน เช่น การทดสอบสมมุติฐานเกี่ยวกับสัมประสิทธิ์สหสัมพันธ์ของประชากร (ρ_{XY}) สัมประสิทธิ์ความถดถอย (β) ความแตกต่างระหว่างสัมประสิทธิ์ความถดถอย ($\beta_1 - \beta_2$) ความแตกต่างระหว่างมัชฌิมเลขคณิตของประชากร ($\mu_1 - \mu_2$) 2 กลุ่ม

ข. ความแตกต่างระหว่างการแจกแจงทีและการแจกแจงเอฟ

	การแจกแจงที	การแจกแจงเอฟ
ข้อตกลงเบื้องต้น :	กำหนดขนาดของกลุ่มตัวอย่าง นั่นคือ ขนาดของกลุ่มตัวอย่างจะน้อยกว่า 30	ไม่กำหนดขนาดของกลุ่มตัวอย่าง
คุณสมบัติ :	มีลักษณะสมมาตรที่ t=0 และมีพิสัยตั้งแต่ $-\infty$ ถึง $+\infty$ ดังนั้นค่า t จึงมีทั้งค่าบวกและค่าลบ ซึ่งโค้งการแจกแจงจะเป็นรูประฆัง (Bell Shaped) ดังรูป	มีความเบ้ทางด้านบวก และมีพิสัยตั้งแต่ 0 ถึง ∞ จึงไม่มีค่าลบ และโค้งการแจกแจงจะเป็นรูปกลับของอักษรตัว J (Reversed J Shape) ดังรูป





การนำไปใช้ :	ในการทดสอบสมมุติฐานสถิติจะมีการทดสอบชนิดทางเดียวและชนิดสองทาง การแจกแจงที่ใช้ประมาณค่าและ/หรือทดสอบสมมุติฐานเกี่ยวกับมัชฌิมเลขคณิตของประชากร 1 กลุ่ม, ความแตกต่างระหว่างมัชฌิมเลขคณิตของประชากร 2 กลุ่ม,ความแตกต่างระหว่างความแปรปรวนของประชากรที่สัมพันธ์กัน สัมประสิทธิ์สหสัมพันธ์แบบต่าง ๆ ความแตกต่างระหว่างสัมประสิทธิ์สหสัมพันธ์โพรดักโมเมนต์ สัมประสิทธิ์ความถดถอย ความแตกต่างระหว่างสัมประสิทธิ์ความถดถอย และความถดถอยเชิงเส้นตรงของประชากร (Linear Regression)	ในการทดสอบสมมุติฐานทางสถิติมักนิยมทดสอบชนิดทางเดียว คือทางด้านขวาหรือทางด้านที่มีค่ามาก การแจกแจงเอฟใช้ประมาณค่าและ/หรือทดสอบสมมุติฐานเกี่ยวกับความแตกต่างระหว่างมัชฌิมเลขคณิตของประชากรหลายกลุ่ม ความแตกต่างระหว่างความแปรปรวนของประชากรที่เป็นอิสระ สัมประสิทธิ์สหสัมพันธ์แบบพหุคูณ ความแตกต่างระหว่างสัมประสิทธิ์สหสัมพันธ์แบบพหุคูณ อัตราส่วนสหสัมพันธ์ สัดส่วนของประชากร ความแตกต่างของความแปรปรวนของประชากรหลายกลุ่ม และการทดสอบภาวะเชิงเส้น (Linearity Test)