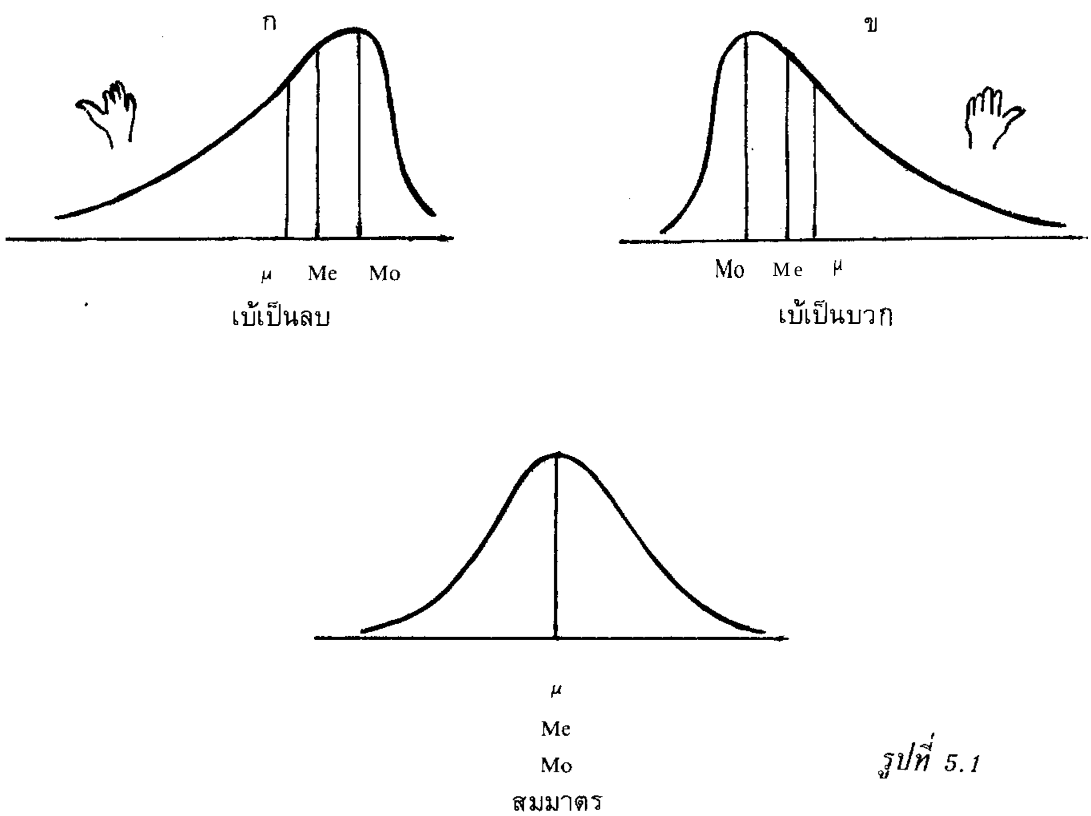
บทที่ 5
ความเบ้และความสูงของยอด

5.1 ความเบ้ (Skewness)

ข้อมูลหรือค่าสังเกตที่มีการแจกแจงสมมาตร มัชฌิมเลขคณิต มัธยฐานและฐานนิยมจะมีค่าเท่ากันหรือทับกันพอดี แต่ถ้าข้อมูลที่มีการแจกแจงไม่สมมาตรคือเบ้ไปข้างใดข้างหนึ่ง มัชฌิมเลขคณิต มัธยฐานและฐานนิยมก็จะมีค่าต่างกัน ในกรณีข้อมูลที่มีการแจกแจงเบ้ไปทางขวา ค่าของมัชฌิมเลขคณิตจะมากกว่ามัธยฐานและมัธยฐานจะมีค่ามากกว่าฐานนิยม (μ > Me > Mo) รูปที่ 5.1 ก. แต่ข้อมูลที่มีการแจกแจงเบ้ไปทางซ้าย ค่าฐานนิยมก็จะมากกว่าค่ามัธยฐานและค่ามัธยฐานก็จะมากกว่าค่ามัชฌิมเลขคณิต (Mo > Me > μ) รูปที่ 5.1 ข.



รูปที่ 5.1

5.2 มาตราวัดความเบ้ (Measures of skewness)

ข้อมูลกลุ่มหนึ่งมีความเบ้มากน้อยเพียงใด เราอาจคำนวณหาได้โดยหาผลต่างระหว่างค่าของส่วนเฉลี่ยเลขคณิตกับฐานนิยม (Mean – mode) ค่าที่ได้นี้นอกจากจะบอกให้ทราบขนาดความเบ้ของการแจกแจง ยังจะแสดงลักษณะของการเบ้อีกว่าเป็นแบบบวกหรือลบ ถ้าเป็นแบบบวกก็แสดงว่าข้อมูลมีการแจกแจงที่เบ้ไปทางขวา การวัดความเบ้แบบนี้ มีข้อเสียอยู่คือใช้เปรียบเทียบกับความเบ้ของข้อมูล 2 กลุ่มที่มีหน่วยต่างกัน และการกระจายต่างกันไม่ได้

เราอาจหาความเบ้สัมพันธ์แทน โดยหาผลต่างระหว่างมัชฌิมเลขคณิตกับฐานนิยมแล้วหารด้วยส่วนเบี่ยงเบนมาตรฐานตามสูตรของ Carl Pearson เขียนได้

$$\text{ความเบ้ Sk} = \frac{\mu - \text{Mo}}{\sigma} \qquad \text{..........(5.2.1)}$$

แต่ค่าฐานนิยมเป็นค่าโดยประมาณในกรณีข้อมูลที่มีการแจกแจงความถี่ ซึ่งไม่ดีเท่ากับค่ามัธยฐาน จึงนิยมใช้ค่ามัธยฐานแทนใช้ความสัมพันธ์ระหว่างมัชฌิมเลขคณิต มัธยฐานและฐานนิยม มัธยฐานจะอยู่ระหว่างฐานนิยมกับมัชฌิมเลขคณิตเขียนเป็นสูตรได้

$$\text{Mo} - \text{Me} = 2(\text{Me} - \mu)$$

$$\text{Mo} = 3\,\text{Me} - 2\mu$$

แทนค่าลงในสูตรได้

$$\text{ความเบ้ของ Sk} = \frac{\mu - (3\,\text{Me} - 2\mu)}{\sigma}$$

$$= \frac{3(\mu - \text{Me})}{\sigma} \qquad \text{..........(5.2.2)}$$

ตัวอย่าง 5.2.1 จากรูปที่ 2.5 หาความเบ้ของคะแนนนักศึกษา 50 คน ในเมื่อ $\mu = 65.1$;

$$\text{Me} = 65.33;\ \sigma = 16.78$$

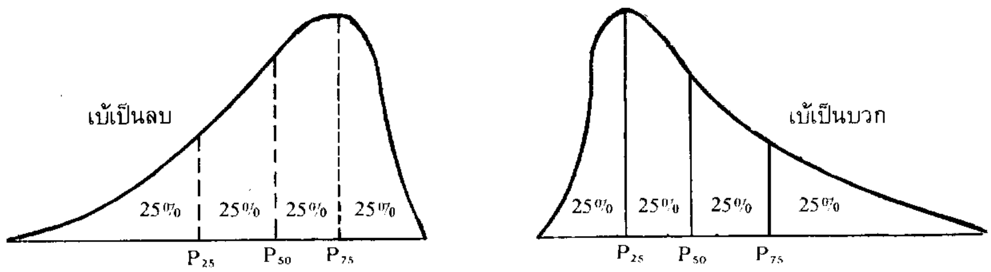
$$\text{ความเบ้ Sk} = \frac{3(\mu - \text{Me})}{\sigma}$$

$$= \frac{3(65.1 - 65.33)}{16.78}$$

$$= \frac{-.69}{16.78} = -0.041$$

แสดงว่า คะแนนของนักศึกษาเบ้ไปทางซ้ายเล็กน้อยเพราะได้ค่า -0.041 ไม่ถึง ± 1 การวัดความเบ้อาจวัดได้โดยใช้ควอร์ไทล์หรือ เปอร์เซ็นไทล์

ข้อมูลข่าวสารเกี่ยวกับความสมมาตรอาจคำนวณหาได้จากการพิจารณาเปรียบเทียบที่ตั้งของ P_{25}, P_{50} และ P_{75} ทั้งสามค่าของเปอร์เซ็นต์ไทล์เหล่านี้แบ่งพื้นที่ของการแจกแจงออกเป็นสี่ส่วนเท่า ๆ กัน ในการแจกแจงที่เบ้เป็นลบ ระยะทางระหว่าง P_{25} กับ P_{50} มีค่ามากกว่าระยะทาง P_{50} กับ P_{75} ในการแจกแจงที่เบ้เป็นบวก ระยะทางระหว่าง P_{25} กับ P_{50} มีค่าน้อยกว่าระยะทาง P_{50} กับ P_{75} ดังรูปแสดงถึงคุณลักษณะเหล่านี้

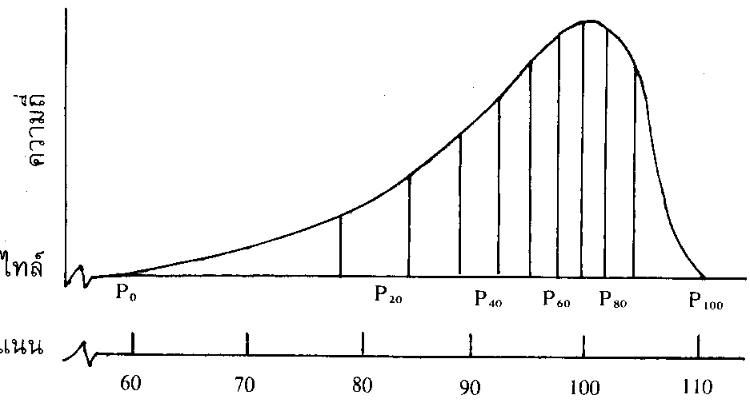


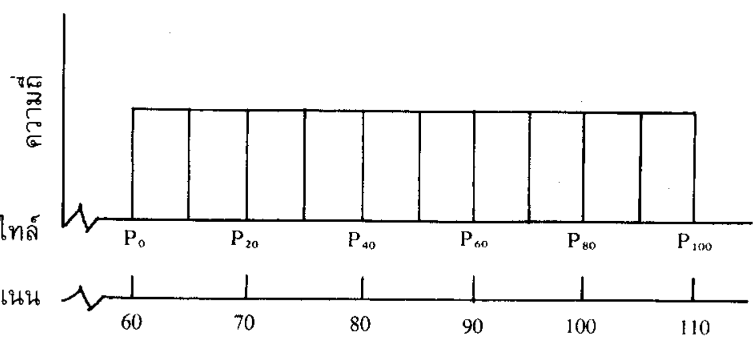
รูปที่ 5.2 การเปรียบเทียบตำแหน่งของ P_{25}, P_{50} และ P_{75} ในการแจกแจงที่เบ้เป็นลบและบวก

5.3 คะแนนเปอร์เซ็นต์ไทล์

คุณลักษณะพื้นฐานของคะแนนเปอร์เซ็นต์ไทล์นักศึกษาก็ได้ทราบมาบ้างแล้วในหัวข้อก่อน แต่คุณสมบัติที่แตกต่างในบางส่วนไปจากคะแนนมาตรฐาน คือ ตำแหน่งเปอร์เซ็นต์ไทล์ของคะแนนอธิบายถึงที่ตั้งเพื่อเปรียบเทียบกับคะแนนอื่น ๆ ของการแจกแจง ตำแหน่งเปอร์เซ็นต์ไทล์มีข้อดีในเทอมของทิศทางของความหมายและต้องการคุณสมบัติเสมอ ตัวอย่างเช่นตำแหน่งเปอร์เซ็นต์ไทล์ของนายแดงเป็น 75 ในการสอบวิชาหลักสถิติ หมายความว่านายแดงสอบได้ดีกว่านักศึกษาถึง 75% ซึ่งเป็นการเข้าใจที่ง่ายโดยที่บุคคลนั้นไม่จำเป็นได้รับการฝึกอบรมวิชาสถิติ

ข้อเสียของตำแหน่งเปอร์เซ็นต์ไทล์ก็ถือว่า การเปลี่ยนแปลงในคะแนนดิบไม่จำเป็นที่จะสะท้อนไปถึงการเปลี่ยนแปลงที่เป็นสัดส่วนกันในตำแหน่งเปอร์เซ็นต์ไทล์ เมื่อไรที่ตำแหน่งเปอร์เซ็นต์ไทล์หนึ่งสูงกว่าตำแหน่งอื่น ๆ คะแนนดิบที่สมนัยกันของตำแหน่งเปอร์เซ็นต์ไทล์ที่สูงกว่าตำแหน่งอื่น ๆ ก็จะเปลี่ยนในคะแนนดิบไปกับการเปลี่ยนแปลงที่เป็นสัดส่วนกันในตำแหน่งเปอร์เซ็นต์ไทล์เมื่อไรที่การแจกแจงของคะแนนเป็นรูปสี่เหลี่ยมผืนผ้าเท่านั้นดังรูปที่ 5.3

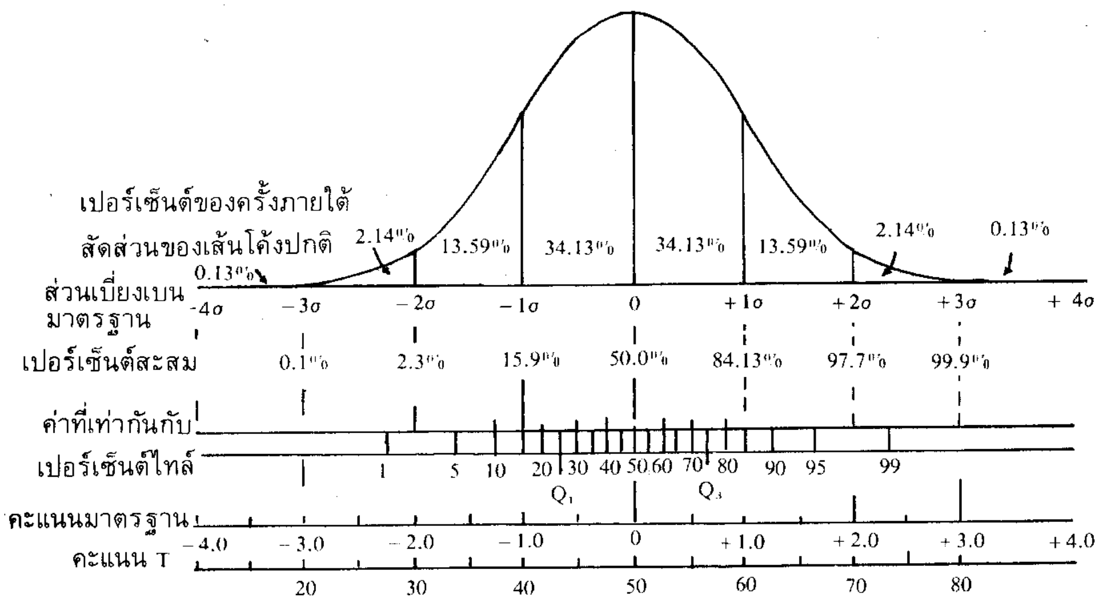




รูปที่ 5.3 การเปรียบเทียบที่ตั้งของตำแหน่งเปอร์เซ็นต์ไทล์ของสองการแจกแจง

แสดงถึงการแจกแจงแบบสี่เหลี่ยมผืนผ้า การเปลี่ยนแปลงสิบจุดในตำแหน่งเปอร์เซ็นต์ไทล์สะท้อนการเปลี่ยนแปลงคะแนนห้าคะแนน ในการแจกแจงอื่น ๆ การเปลี่ยนแปลงสิบจุดในตำแหน่งเปอร์เซ็นต์ไทล์สะท้อนการเปลี่ยนแปลงในคะแนนซึ่งอาจจะมากกว่าหรือน้อยกว่าห้าจุดขึ้นอยู่กับที่ตั้งของการแจกแจง

ในการแจกแจงปกติ ความถี่ของคะแนนมากที่สุดอยู่ใกล้ศูนย์กลางของการแจกแจง ระยะทางระหว่างสองค่าของคะแนนใกล้ศูนย์กลางแทนได้ด้วยการเปรียบเทียบของผลต่างที่มีค่ามากในตำแหน่งเปอร์เซนต์ไทล์ สำหรับอีกด้านหนึ่งที่ปลายทั้งสองของการแจกแจงที่ระยะทางระหว่างคะแนนที่คล้ายคลึงกันใช้แทนได้ด้วยผลต่างที่มีค่าน้อยกว่าในเปอร์เซ็นต์ไทล์ ดังแสดงในรูป 5.4



รูปที่ 5.4 คะแนนมาตรฐานและเปอร์เซ็นต์ไทล์ในการแจกแจงปกติ

รูปนี้แสดงถึงการเปรียบเทียบมาตราต่าง ๆ ที่ใช้คำนวณหาของคะแนนในการแจกแจงปกติ

โดยใช้ควอร์ไทล์

ความเบ้ $Sk = \dfrac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} = \dfrac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$ ....(5.2.3)

ตัวอย่าง 5.2.2 จากตารางที่ 3.4 ของคะแนนนักศึกษา 50 คน

ในเมื่อ $Q_1 = 52.625$; $Q_2 = 65.33$; $Q_3 = 77.822$

แทนค่าในสูตร

$$Sk = \frac{77.822 - 2 \times 65.33 + 52.625}{77.822 - 52.625}$$

$$= \frac{-.213}{25.197} = -0.008$$

โดยใช้เปอร์เซ็นต์ที่ 10 และ 90

สูตร $Sk = \dfrac{(P_{90} - P_{50}) - (P_{50} - P_{10})}{P_{90} - P_{10}}$

หรือ $= \dfrac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}}$ ....(5.2.4)

วิธีคำนวณแบบควอร์ไทล์ และเปอร์เซ็นต์ไทล์ ไม่เป็นที่นิยมกันเพราะมีข้อเสียเกี่ยวกับการใช้มาตรการวัดการกระจายแบบควอร์ไทล์และเปอร์เซ็นต์ไทล์

วิธีที่นิยมมากที่สุดคือ ค่า มัชฌิมเลขคณิตของส่วนเบี่ยงเบนไปจากมัชฌิมเลขคณิตยกกำลังสามหารด้วยส่วนเบี่ยงเบนมาตรฐานยกกำลังสาม เขียนเป็นสูตรได้

(ข้อมูลที่ไม่มีการจัดเป็นกลุ่ม) $$\alpha_3 = \frac{\sum_{i=1}^{N} (X_i - \mu)^3}{N\sigma^3} \qquad \text{..........(5.2.5)}$$

หรือ (ข้อมูลที่มีการจัดเป็นกลุ่ม) $$\alpha_3 = \frac{\sum_{i=1}^{k} f_i(X_i - \mu)^3}{N\sigma^3} \qquad \text{..........(5.2.6)}$$

เพื่อให้สะดวกในการคำนวณและรวดเร็วขึ้นจึงเปลี่ยนสเกลให้มีตัวเลขน้อยเข้า อย่างเดียวกับการคำนวณหามัชฌิมและส่วนเบี่ยงเบนมาตรฐาน ดังสูตร

$$\alpha_3 = \frac{I^3}{\sigma^3}\left[\frac{\sum_{i=1}^{k} f_i d_i^3}{N} - 3\left[\frac{\sum_{i=1}^{k} f_i d_i^2}{N}\right]\left[\frac{\sum_{i=1}^{k} f_i d_i}{N}\right] + 2\left[\frac{\sum_{i=1}^{k} f_i d_i}{N}\right]^3\right] \qquad \text{..........(5.2.7)}$$

สูตรนี้อาจจะยาวแต่ใช้ไม่ยาก เพราะค่า $\sum_{i=1}^{k} f_i d_i$ และ $\sum_{i=1}^{k} f_i d_i^2$ ก็มีอยู่แล้วในการคำนวณหาส่วนเบี่ยงเบนมาตรฐาน เพียงแต่เพิ่มอีกช่องเพื่อคำนวณหา $\sum_{i=1}^{k} f_i d_i^3$ เท่านั้น

ตัวอย่าง 5.2.3 คำนวณหา α_3 จากตารางที่ 3.3

ชั้นคะแนน	f	x	d	df	d^2f	d^3f
30-39	4	34.5	- 3	-12	36	104
40-49	6	44.5	- 2	- 12	24	48
50-59	8	54.5	- 1	- 8	8	- 8
60-69	12	64.5	0	0	0	0
70-79	9	74.5	l	9	9	9
80-89	7	84.5	2	14	28	56
90-99	4	94.5	3	12	36	104
	50			3	141	9

แทนค่าลงในสูตรได้ $\alpha_3 = \frac{10_3}{(16.78)^3}\left[\frac{9}{50} - 3\left(\frac{141}{50}\right)\left(\frac{3}{50}\right) + 2\left(\frac{3}{50}\right)^3\right]$

$$= \frac{1000}{(16.78)^3}\left[.18 - 3(2.82)(.06) + 2(.000216)\right]$$

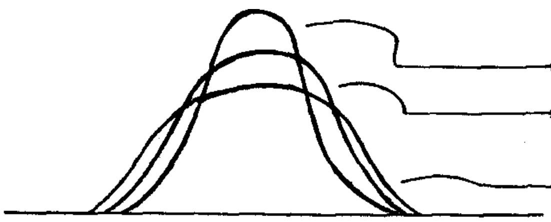
$$= \frac{1000}{4724.7177}(-\ .3272) = -0.069$$

แสดงว่าการแจกแจงมีความเบ้ไปทางซ้ายเล็กน้อย ถ้าหากว่าเป็นการแจกแจงสมมาตร เราจะได้ $\alpha_3 = 0$

ค่า α_3 นาน ๆ จะพบสักครั้ง เว้นแต่ N มีค่ามาก สูตรคำนวณก็ใช้สูตรเดียวกันไม่ว่าตัวอย่างหรือประชากร

5.4 มาตรวัดส่วนสูงของยอด (Measures of Peakedness)

ข้อมูลที่มีการแจกแจงอย่างสมมาตร นอกจากจะต้องมีลักษณะทั้งสองด้านนับจาก ฐานนิยมเหมือนกันทุกอย่าง ยังต้องมีส่วนสูงได้พอเหมาะอีกด้วย ถ้าการแจกแจงที่มีส่วนสูงมากเรียกว่า Leptokurtic (หมายถึงยอดแหลมมีปลายทั้งสองข้างกว้าง) ถ้าการแจกแจงที่มียอดค่อนข้าง ราบปลายทั้งสองบานเรียกว่า platykurtic ส่วนการแจกแจงปกติเรียกว่า Mesokurtic ดังรูป 5.5



รูปที่ 5.5

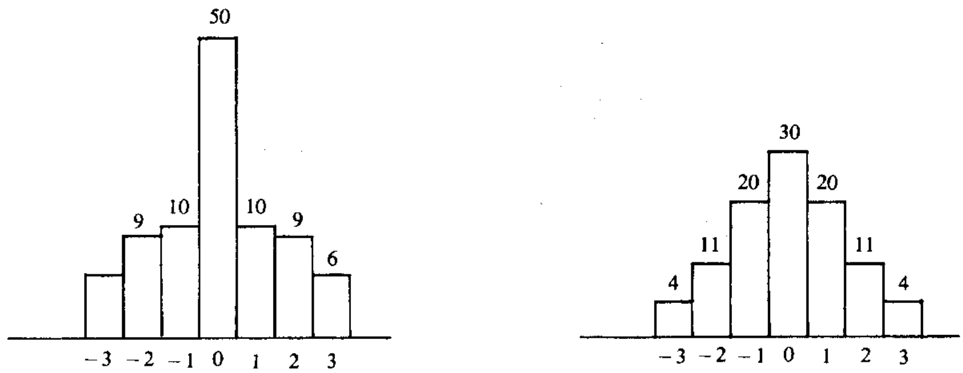
การวัดส่วนสูงของยอดวัดได้จากมัชฌิมเลขคณิตของส่วนเบี่ยงเบนไปจากมัชฌิมเลขคณิตยกกำลังสี่หารด้วยส่วนเบี่ยงเบนมาตรฐานยกกำลังสี่ เขียนเป็นสูตรได้

$$\alpha_4 = \frac{\sum_{i=1}^{k} f_i(x_i-\mu)^4}{N\sigma^4} \qquad \text{..........(5.4.1)}$$

สำหรับข้อมูลที่มีการแจกแจงความถี่ สูตรวิธีลัดเขียนได้คือ

$$= \frac{I^4}{\sigma^4}\left(\frac{\sum_{i=1}^{k} f_i d_i^4}{N} - 4\left[\frac{\sum_{i=1}^{k} f_i d_i^3}{N}\right]\left[\frac{\sum_{i=1}^{k} f_i d_i}{N}\right] + 6\left[\frac{\sum_{i=1}^{k} f_i d_i^2}{N}\right]\left[\frac{\sum_{i=1}^{k} f_i d_i}{N}\right]^2 - 3\left[\frac{\sum_{i=1}^{k} f_i d_i}{N}\right]^4\right) \text{...(5.4.2)}$$

ในกรณีการแจกแจงแบบ Mesokurtic ค่า α_4 ประมาณ 3 แต่ค่า α_4 ของการแจกแจงแบบ Leptokurtic มีค่ามากกว่า 3 α_4 ของ Platykurtic มีค่าน้อยกว่า 3 คือประมาณ 2.633 ก็อยู่ในจำพวก platykurtic การแจกแจงสองชนิดดังรูป 5.6 ซึ่งมีมัชฌิมเลขคณิต ส่วนเบี่ยงเบนมาตรฐานและความเบ้เท่ากับ คือ $\mu = 0; \sigma = \sqrt{2}$ และ $\alpha_3 = 0$ ตามลำดับ แต่การแจกแจงชนิดแรกเป็นแบบ Leptokurtic มี $\alpha_4 = 3.2$ ขณะที่การแจกแจงชนิดที่สองเป็นแบบ platykurtic มี $\alpha_4 = 2.6$



รูปที่ 5.6

แบบฝึกหัดที่ 5

1. จากแบบฝึกหัดที่ 4 ข้อ 5 จงหาค่าสัมประสิทธิ์ของความเบ้สำหรับการแจกแจงคะแนนสอบไล่นักศึกษา 600 คนแบบ Carl Pearson; วัดโดยใช้ควอร์ไทล์; วัดโดยใช้เปอร์เซ็นต์ไทล์ (0.0343, − 0.02556)
2. จงคำนวณหา α_3 โดยใช้สูตรวิธีลัดจากข้อ 5 ของแบบฝึกหัดที่ 4 (0.3497)
3. จงคำนวณหา α_4 จากสูตรวิธีลัดของแบบฝึกหัดที่ 4 ข้อ 5 (3.8334)