

การแปลงข้อมูลที่มีการแจกแจงไม่ปกติ และความแปรปรวนไม่คงที่

[โดย ผศ.ดร.นิติพงษ์ ส่งศรีโรจน์ 16 ตุลาคม 2557]

การแปลงค่า Y โดยพิจารณาจากความสัมพันธ์ของความแปรปรวนกับ \hat{Y} ดังนี้

ความสัมพันธ์ระหว่าง σ^2 กับ $E(Y)$	รูปแบบการแปลง
$\sigma^2 \propto$ ค่าคงที่	$Y' = Y$ (ไม่ต้องแปลงข้อมูล)
$\sigma^2 \propto E(Y)$	$Y' = \sqrt{Y}$ (กรณี y แจกแจงแบบปัวซอง)
$\sigma^2 \propto E(Y)[1 - E(Y)]$	$Y' = \sin^{-1}(\sqrt{Y})$ (กรณี $0 \leq Y \leq 1$)
$\sigma^2 \propto [E(Y)]^2$	$Y' = \ln(Y)$
$\sigma^2 \propto [E(Y)]^3$	$Y' = Y^{-\frac{1}{2}}$
$\sigma^2 \propto [E(Y)]^4$	$Y' = Y^{-1}$

ที่มา: Montgomery and Peck (1992:98)

การแปลงข้อมูลโดยวิธี Box-Cox transformation

รูปแบบการแปลงข้อมูลวิธีนี้ คือ

$$y' = y^\lambda$$

โดยที่ λ คือ ค่าพารามิเตอร์ที่ได้มาจากข้อมูล , $-1 \leq \lambda \leq 1$

ถ้ากำหนด $\lambda = 2$ จะได้ $y' = y^2$

$\lambda = 0.5$ จะได้ $y' = \sqrt{y}$

$\lambda = 0$ จะได้ $y' = \ln Y$

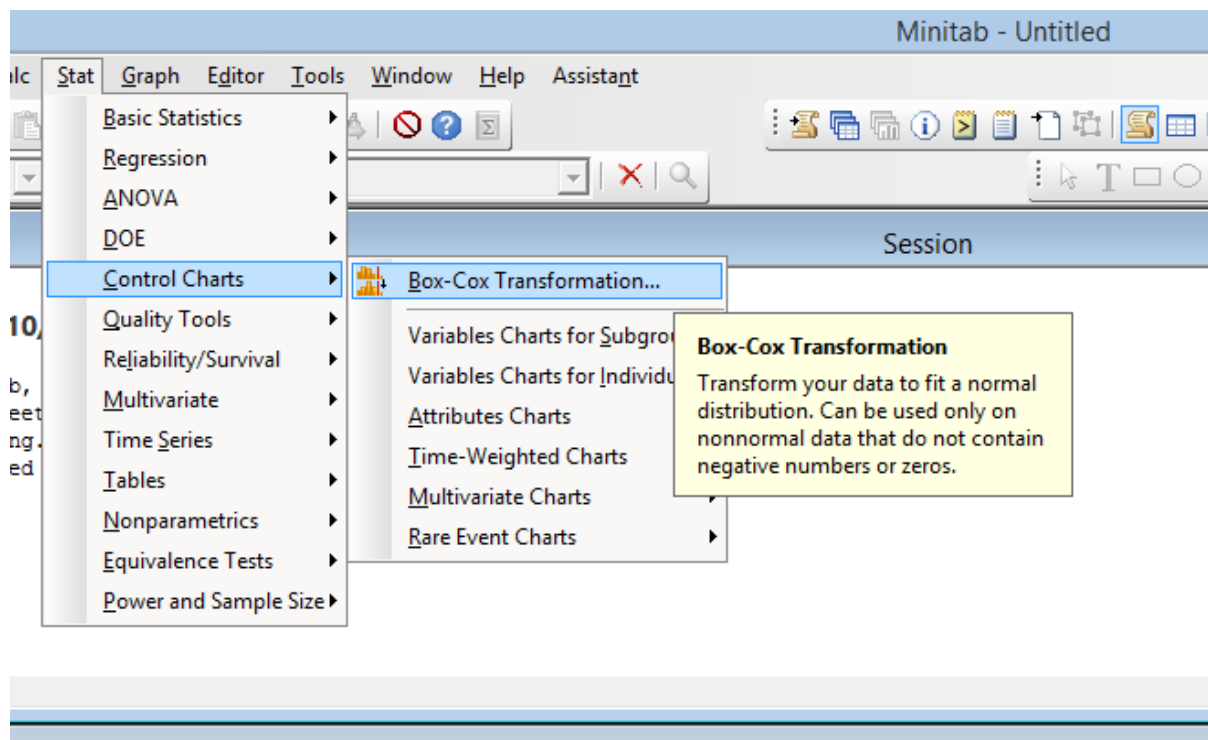
$\lambda = -0.5$ จะได้ $y' = \frac{1}{\sqrt{y}}$

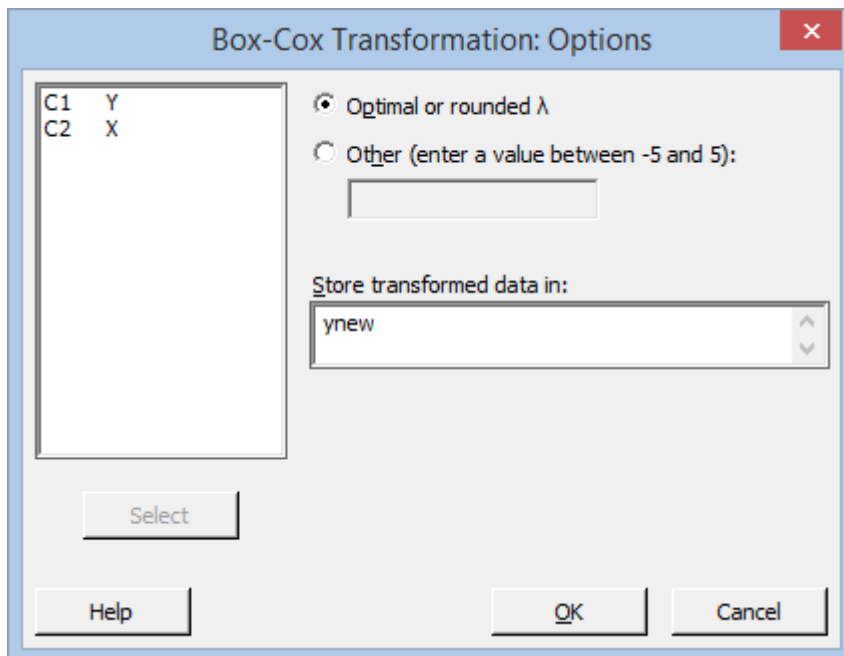
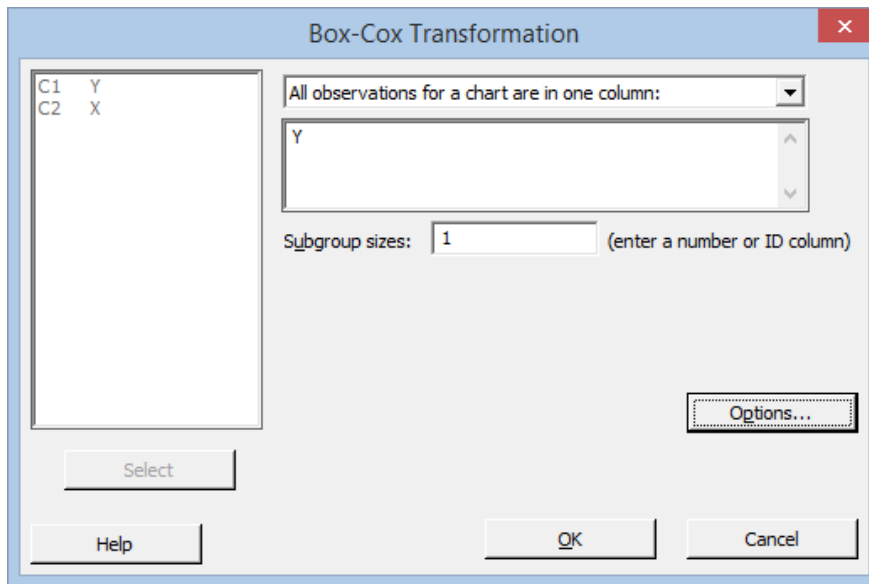
$\lambda = -1.0$ จะได้ $y' = \frac{1}{y}$

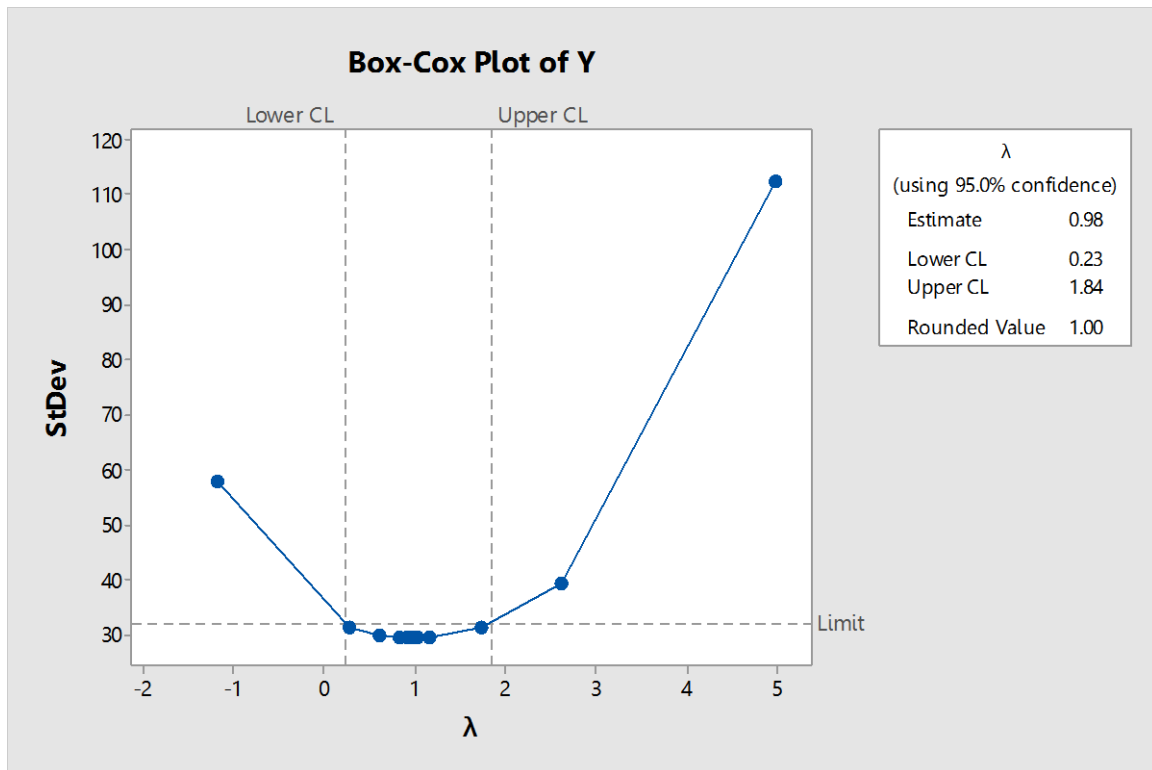
λ จะมีค่าเท่าใด และการแปลงข้อมูล Y จะอยู่ในรูปแบบใดนั้น จะต้องมีการลองผิดลองถูกโดยเลือก λ ที่ทำให้ SSE มีค่าต่ำสุด

การใช้โปรแกรม MINITAB สำหรับวิธีหาค่า λ ที่เหมาะสม มีวิธีดังนี้

- 1) เลือก Stat
- 2) เลือก Control Charts
- 3) เลือก Box-Cox Transformation
- 4) กำหนดค่าตัวแปร Y ที่ต้องการแปลงค่าลง Single column
- 5) กำหนดขนาดของกลุ่มย่อย (subgroup) ลงใน subgroup size หากไม่มีกลุ่มย่อยให้กำหนดเป็น 1
- 6) หากต้องการเก็บค่าที่แปลงแล้ว ให้เลือก Store transform data in single column







จากภาพพบว่าค่า λ ที่เหมาะสมคือ 0.98 และช่วงความเชื่อมั่นที่ 95% ของค่า λ คือ (0.23, 1.84) ซึ่งค่า λ มีค่าใกล้เคียงกับ 1 จึงไม่ต้องมีการแปลงค่า