

ความแปรปรวนของ Residual และ R-square

ถ้าค่าความแปรปรวนของค่า *Residual* รอบๆเส้นถดถอยเมื่อเทียบกับความแปรปรวนดั้งเดิม¹ มีค่าน้อยแล้ว ก็จะทำให้ความสามารถในการทำนายยิ่งดีขึ้น ตัวอย่างเช่น ถ้าไม่มีความสัมพันธ์ระหว่าง X กับ Y ดังนั้น อัตราส่วนของความแปรปรวนของค่า *Residual* ของ Y ต่อค่าความแปรปรวนดั้งเดิมจะเท่ากับ 1 ถ้า X กับ Y มีความสัมพันธ์กันอย่างสมบูรณ์ ดังนั้นจะไม่มีค่า *Residual* อัตราส่วนค่าความแปรปรวนจะเท่ากับ 0 โดยทั่วไปค่านี้จะอยู่ระหว่าง 0 กับ 1 และ ถ้า 1 ลบด้วยอัตราส่วนนี้ก็คือค่า *R-square* หรือ the *coefficient of determination* ถ้าค่านี้เท่ากับ 0.40 หมายความว่า ความแปรปรวนของค่า Y รอบๆ เส้นถดถอยเท่ากับ 0.60 เท่าของค่าความแปรปรวนทั้งหมดหรือสมการถดถอยสามารถอธิบายได้ร้อยละ 40 ของความแปรปรวนดั้งเดิม ส่วนร้อยละ 60 เป็นความแปรปรวนอันเกิดจากค่า *Residual* ต่อค่าความแปรปรวนดั้งเดิม *R-square* เป็นตัวชี้ว่าตัวแบบของเรามีความเหมาะสม (Fit) กับข้อมูลมากน้อยเพียงใดหรือสมการถดถอยนั้น

Adjusted R-Squared

ตามความหมายของ R Square (the Coefficient of Determination) ที่กล่าวข้างต้นนั้นมีจุดอ่อนก็คือ การคำนวณดังกล่าวอยู่ในรูปร้อยละ กล่าวคือ ตัวหารนั้นมีค่าคงที่ (TSS) แต่ตัวเศษมีค่าลดลง (RSS) ดังนั้น การเพิ่มตัวแปรอิสระในสมการทำให้ตัวเศษลดลง เป็นผลให้ R Square มีค่าสูงขึ้น หรือ การเพิ่มตัวแปรใหม่เข้าไปทำให้สมการถดถอยมีประสิทธิภาพ (Efficient) ลดลง

ในทางทฤษฎีการเพิ่มตัวแปรอิสระแบบไม่จำกัด (Infinity) จะทำให้ R Square มีค่าเท่ากับ 1 Adjusted R Squares หรือ $\bar{R}^2$ เป็นความพยายามทำให้ข้อบกพร่องนี้ถูกต้องด้วยการปรับทั้งตัวเศษและตัวส่วนด้วยค่าองศาอิสระ (degrees of freedom) ซึ่งมีค่าเท่ากับ

$$1 - (1 - R^2) \frac{n-1}{n-k-1}$$

where R^2 = Coefficient of Determinant

$\bar{R}^2$ = Adjusted Coefficient of Determination

n = number of observations

k = number of Independent Variables

for example: when $\bar{R}^2 = .9$; n=100; and k=5; then

$$\bar{R}^2 = 1 - (1 - .9)((100 - 1)/(100 - 5 - 1))$$

$$= 1 - (1 - .9)(99/94)$$

$$= 1 - (.1)(1.05319)$$

¹ ความแปรปรวนดั้งเดิมหรือความแปรปรวนของข้อมูล เท่ากับ $\sum_{i=1}^n (y_i - \bar{y})^2$, the explained sum of squares หรือ ESS เท่ากับ $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$, the residual sum of squares หรือ RSS เท่ากับ $\sum_{i=1}^n (y_i - \hat{y}_i)^2$, $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$

$$= 1 - .105319$$

$$= .89468$$

Adjusted R Square สามารถมีค่าลดลงถ้าผลของ explained deviation จากการเพิ่มตัวแปรอิสระ มีค่าน้อยกว่าผลของ degrees of freedom หมายความว่า Adjusted R Square จะมีการตอบสนองต่อตัวแบบต่างๆที่มีตัวแปรตามเดียวกันในลักษณะคล้ายคลึงกับ Standard Error of the Estimate นั่นคือ สมการที่มี Standard Error of the Estimate น้อยส่วนใหญ่ จะมี Adjusted R Square ที่สูง นอกจากนี้ที่ควรระวังก็คือ R Square อยู่ในรูปร้อยละ แต่ Adjusted R Square ไม่ใช่ซึ่งควร จะมองเป็นค่าดัชนีชนิดหนึ่ง

P, t and standard error

The *t statistic* ก็คือ ค่าสัมประสิทธิ์หารด้วย *standard error* ของค่าสัมประสิทธิ์ ค่า *standard error* เป็นตัวประมาณของ *standard deviation* ของค่าสัมประสิทธิ์ค่ามันจึงเปลี่ยนไปแล้วแต่กรณี ถ้าค่าสัมประสิทธิ์มีค่าใหญ่มากเมื่อเทียบกับค่า *standard error* ของมันเอง ก็เป็นไปได้ว่าค่าสัมประสิทธิ์นั้นจะไม่เท่ากับศูนย์

F-Statistic

ใช้ทดสอบว่าค่าสัมประสิทธิ์ของตัวแปรทุกตัวเท่ากับศูนย์หรือไม่ หรือ ตัวแปรทุกตัวมีความสัมพันธ์กับตัวแปรตามหรือไม่ ถ้าเราปฏิเสธ *null hypothesis* หมายความว่าอย่างน้อยหนึ่งตัวแปรที่มีความสัมพันธ์กับตัวแปรตาม

Standard Error of Regression

มีค่าเท่ากับ $\hat{\sigma} = \sqrt{\frac{\sum \hat{u}^2}{n-k}}$ หรือ S_e หมายถึงเป็นตัวประมาณ *standard deviation* ของค่า Residual