แทนค่า $b\Sigma(X_i - \bar{X})^2$ ในเทอมสุดท้ายขวามือของสมการ

$$\Sigma e_i^2 = \Sigma(Y_i - \bar{Y})^2 - 2b\Sigma(X_i-\bar{X})(Y_i-\bar{Y}) + b\Sigma(X_i - \bar{X})(Y_i - \bar{Y})$$

$$= \Sigma(Y_i-\bar{Y})^2 - b\Sigma(X_i-\bar{X})(Y_i-\bar{Y}) \qquad \ldots\ldots(1)$$

หรือถ้าเขียนในรูปเบี่ยงเบนจากmean จะได้

$$\Sigma e_i^2 = \Sigma y^2 - b\Sigma xy$$

ถ้าX และY ไม่มีความสัมพันธ์กัน, $b = 0$ แล้ว Σe_i^2 จะ $= \Sigma(Y_i-\bar{Y})^2$

ฉะนั้น $\Sigma(Y_i-\bar{Y})^2$ = ความแปรปรวนทั้งหมด ซึ่งถูกอธิบาย

ถ้าX มีความสัมพันธ์กับ Y อย่างสมบูรณ์ เส้นที่ดีที่สุดจะผ่านทุกจุดของข้อมูล ความคลาดเคลื่อนจะเป็น 0 ทั้งหมด *i.e.* $\Sigma e_i^2 = 0$, $\Sigma(Y_i-\bar{Y})^2 = b\Sigma(X_i-\bar{X})(Y_i-\bar{Y})$

ฉะนั้น $b\Sigma(X_i - \bar{X})(Y_i-\bar{Y})$ เป็นความแปรปรวนใน Y ซึ่งถูกอธิบายโดย X

อัตราส่วน $\dfrac{b\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(Y_i - \bar{Y})^2}$ เป็นเศษส่วนของความแปรปรวนใน Y ซึ่ง "ถูกอธิบายโดย X"

อัตราส่วน $\dfrac{\Sigma e_i^2}{\Sigma(Y_i - \bar{Y})^2}$ เป็นเศษส่วนของความแปรปรวนใน Y ซึ่ง "ไม่ถูกอธิบาย" หรือไม่อาจอธิบายได้

ผลบวกของความแปรปรวนที่ถูกอธิบายกับไม่ถูกอธิบายจะเท่ากับ 1 เพราะว่าจาก สมการ (1)

$$\Sigma(Y_i-\bar{Y})^2 = b\Sigma(X_i-\bar{X})(Y_i-\bar{Y}) + \Sigma e_i^2$$

เอา $\Sigma(Y_i-\bar{Y})^2$ หารตลอด

$$1 = \underbrace{\frac{b\Sigma(X_i-\bar{X})(Y_i-\bar{Y})}{\Sigma(Y_i-\bar{Y})^2}}_{\text{Explained Vaxiation}} + \underbrace{\frac{\Sigma e_i^2}{\Sigma(Y_i-\bar{Y})^2}}_{\text{Unexplained Variation}} \quad \text{.........(2)}$$

อัตราส่วนของความแปรปรวนที่ถูกอธิบายโดย X (Explained variation) เรียกว่า สัมประสิทธิการตัดสินใจ (Coefficient of Determination) ใช้สัญลักษณ์ R^2

* ดังนั้น (i) $R^2 = \frac{b\Sigma(X_i-\bar{X})(Y_i-\bar{Y})}{\Sigma(Y_i-\bar{Y})^2} = \frac{b\Sigma xy}{\Sigma y^2}$

แทนค่า R^2 ในสมการ (2)

** (ii) $R^2 = 1 - \frac{\Sigma e_i^2}{\Sigma(Y_i-\bar{Y})^2} = 1 - \frac{\Sigma e_i^2}{\Sigma y^2}$

โดยที่ $\Sigma e_i^2 = \Sigma y^2 - b\Sigma xy$

หรือ $\Sigma e_i^2 = \Sigma y^2 - b^2\Sigma x^2$ (หาอีกวิธีหนึ่งซึ่งไม่ได้แสดงให้ดู)

นอกจากนี้ สูตร R^2 อาจจะหาโดยวิธีอื่นจะได้

* (iii) $R^2 = \frac{(\Sigma xy)^2}{\Sigma x^2 \Sigma y^2}$

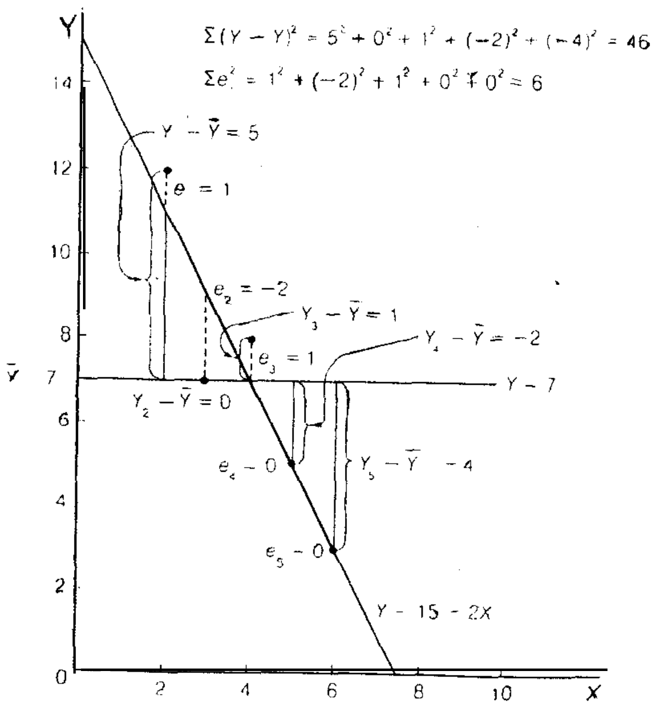
* (iv) $R^2 = \frac{b^2\Sigma x^2}{\Sigma y^2}$

* (v) $R^2 = \frac{\Sigma(\hat{Y} - \bar{Y})^2}{\Sigma(Y_i - \bar{Y})^2}$

โดย $\Sigma(\hat{Y}_i-\bar{Y})^2$ คือ ความแปรปรวนที่ถูกอธิบาย (explained variation) เป็นความแปรปรวนของ Y ที่สามารถจะวัดหรืออธิบายได้ด้วยเส้น regression อาจจะเรียกในเชิงสถิติอีกชื่อหนึ่งว่า Regression Sum of Square (RSS)

$\Sigma(Y_i-\bar{Y})^2$ คือ ความแปรปรวนทั้งหมด (Total Variation) เป็นการกระจายหรือความแปรปรวนของ Y ทั้งหมดจากข้อมูล เรียกอีกอย่างหนึ่งว่า Total Sum of Square (TSS)

จากรูปที่ 2.7 จะช่วยให้นักศึกษาเข้าใจชัดเจนขึ้น ถ้าเราเปรียบเทียบเส้นกำลังสองน้อยที่สุดซึ่งเป็นเส้นที่เราพยากรณ์หาค่า Y กับเส้นนอนซึ่งเป็นเส้นที่มีค่า $(\bar{Y})$ เฉลี่ย ถ้าไม่มีความสัมพันธ์ระหว่าง Y กับ X เส้นที่ดีที่สุด คือเส้นแนวนอนที่มีค่า Y เฉลี่ย ผลบวกของส่วนเบี่ยงเบนกำลังสองจากเส้นนอนนี้ คือ $\Sigma(Y_i-\bar{Y})^2$ ผลบวกนี้ก็คือ จำนวนแปรปรวนใน Y ซึ่งถูกอธิบาย แม้ว่า X และ Y ไม่มีความสัมพันธ์กัน ผลบวกของส่วนเบี่ยงเบนกำลังสอง (Σe_i^2) จะไม่มากไปกว่า $\Sigma(Y_i-\bar{Y})^2$ ถ้า $\Sigma e_i^2 = \Sigma(Y_i-\bar{Y})^2$ โดยที่ X กับ Y ไม่มีความสัมพนธ์กันแล้ว R^2 จะเท่ากับ 0 ในทางตรงข้าม ถ้า X และ Y มีความสัมพันธ์กันอย่างสมบูรณ์ ซึ่งมี $\Sigma e_i^2 = 0$ R^2 จะ = 1 เมื่อ $R^2 = 0$ เส้นถดถอยจะอธิบายอะไรไม่ได้เลย เมื่อ $R^2 = 1$ เส้นถดถอยที่คำนวณได้จะเป็นเส้นที่ดีที่สุด ฟิตกับตัวเลขที่สุดคือ ข้อมูลทุกค่าจะอยู่บนเส้นถดถอย



รูปที่ 2.7

ฉะนั้น สรุปค่าของ R^2 ได้ว่า

1. ถ้า Σe^2 ใกล้ 0 , R^2 จะมีค่าใกล้ 1

2. ถ้า Σe_2^2 ใกล้ $\Sigma(Y_i-\bar{Y})^2$, R^2 จะมีค่าใกล้ 0

3. R^2 จะมีค่าตั้งแต่ 0 ถึง 1 คือ $0 \leq R^2 \leq 1$ หมายความว่า R^2 มีค่าใกล้ 1 มากเท่าใดแสดงว่าเส้นถดถอยฟิตกับข้อมูลที่เป็นจริงมากเท่านั้น หรือ เปอร์เซ็นต์ของ Y ซึ่งจะถูกอธิบายโดย X มากเท่านั้น

ค่า $1 - R^2$ เรียกว่า Coefficient of Nondetermination คือ เปอร์เซ็นต์ของ Y ซึ่งเกิดจากอิทธิพลอื่นที่ไม่ใช่ตัวแปรอิสระ X

จากตัวอย่างที่ 2.1 หน้า ๒๔ ค่าของสัมประสิทธิการตัดสินใจของสมการเส้น demand จะหาได้คือ

$$R^2 = \frac{b\Sigma xy}{\Sigma y^2}$$

$$= \frac{(-2)(-20)}{46} = \frac{40}{46} = 0.87$$

หรืออีกสูตรหนึ่ง

$$R^2 = 1 - \frac{e_i^2}{\Sigma(Y_i - \bar{Y})^2} = 1 - \frac{\Sigma e_i^2}{\Sigma y_i^2}$$

$$= 1 - \frac{6}{46} = \frac{40}{46} = 0.87$$

ซึ่งแสดงว่า ข้อมูลฟิตกับเส้น regression ถึง 87% หรือนัยหนึ่งตัวแปรอิสระ P หรือราคามีอิทธิพลต่อปริมาณซื้อ Q_d ถึง 87% นอกนั้นเป็นอิทธิพลจากภายนอก

2. <u>สัมประสิทธิการตัดสินใจปรับปรุง</u> (Adjusted Coefficient of Determination : $\bar{R}^2$)

$\bar{R}^2$ คือ R^2 ที่ได้ทำการปรับปรุงชั้นความเป็นอิสระ (degree of freedom) จนทำให้อัตราส่วนซึ่งความแปรปรวนของตัวแปรตาม (Y) ลดลงเมื่อพิจารณาตัวแปรอิสระ (X) สาเหตุที่ปรับปรุง R^2 เนื่องจากจำนวนค่าสังเกตหรือข้อมูล (n) ที่เรานำมาหาความสัมพันธ์มีน้อย ค่า R^2 ที่คำนวณได้อาจจะให้ค่าที่ลำเอียง, กล่าวคือ R^2 ซึ่ง $= 1- \frac{\Sigma e^2}{\Sigma y^2}$ มีแนวโน้มที่จะให้ค่า "สูงเกินไป" ดังนั้น ความแปรปรวนของตัวแปรอิสระจะอธิบายความแปรปรวนของตัวแปรตามผิดพลาดได้

เพื่อแก้ปัญหานี้ จะต้องปรับ R^2 เสียใหม่ด้วยชั้นความเป็นอิสระ R^2 ที่ปรับปรุงนี้เขียนสัญลักษณ์เป็น $\bar{R}^2$ $\bar{R}^2$ จึงเป็นอัตราส่วนซึ่งความแปรปรวนของ Y ลดลงเมื่อพิจารณา X

$$\bar{R}^2 = 1-\frac{\Sigma e_i^2/n-2}{\Sigma y_i^2/n-1}$$

$$* \qquad = 1- \frac{(n-1)}{(n-2)}\frac{\Sigma e_i^2}{\Sigma y_i^2}$$

ถ้าเราไม่พิจารณา X แล้วความแปรปรวนของ Y จะ = $\Sigma y_i^2/n-1$ แต่ถ้าพิจารณา X โดยเส้นถดถอยกำลังสองน้อยที่สุดความแปรปรวนของY จากเส้นถดถอยจะเท่ากับ $\Sigma e_i^2/n-2$ ถ้า Y ถูกอธิบายโดย X อย่างสมบูรณ์ $\Sigma e_i^2 = 0$ และ $\bar{R}^2$ จะ = 1 ถ้า X และ Y ไม่มีความสัมพันธ์กัน (b = 0) แล้วความแปรปรวนของ Y และความแปรปรวนของ Y โดยเส้นถดถอยจะเป็นเลขจำนวนเดียวกัน ทำให้อัตราส่วนของมันเป็น 1 แต่อาจเป็นไปได้ว่าความแปรปรวนของY โดยเส้นถดถอย ($\Sigma e_i^2/n-2$) จะมากกว่าความแปรปรวนของ Y ($\Sigma y_i^2/n-1$) ดังนั้น $\bar{R}^2$ อาจจะเป็นลบ (negative) จะเห็นได้ชัดเจนเมื่อ $\frac{(n-1)}{(n-2)} > 1$ $\bar{R}^2$ จะเป็น negative ถ้า $\Sigma e_i^2/\Sigma y_i^2 = 1$ หรือใกล้ 1 (แต่ R^2 มีค่าระหว่าง 0 ถึง 1 และไม่เป็นnegative)

จากตัวอย่าง 2.1 จะหา $\overline{R}^2$ ได้ดังนี้

$$\overline{R}^2 = 1 - \frac{(n-1)}{(n-2)}\frac{\sum e_i^2}{\sum y_i^2}$$

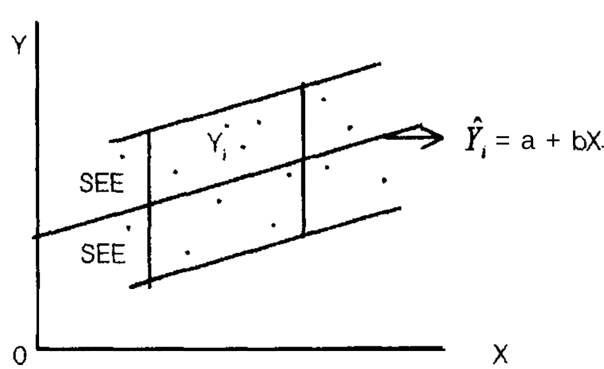
$$= 1 - \frac{(5-1)}{(5-2)}\frac{6}{46}$$

$$= 0.83$$

ถ้า n น้อย ค่า $\overline{R}^2$ จะน้อยกว่าค่า R^2 ถ้า n มากค่า $\overline{R}^2$ จะใกล้เคียงกับค่า R^2 ในการนำเสนอสมการถดถอย จึงควรนำเสนอทั้ง R^2 และ $\overline{R}^2$ ไปพร้อมๆกัน

3.<u>ความคลาดเคลื่อนมาตรฐานของการกะประมาณ</u> (Standard Error of Estimate: SEE , S , $S_{Y,X}$)

ความคลาดเคลื่อนมาตรฐานของการกะประมาณ (Standard Error of Estimate) คือ การวัดการกระจายของข้อมูลตัวแปรตามจากเส้นถดถอยที่ใช้ในการประมาณค่าตัวแปรนั้น การวัดนี้จะเป็นฐานในการตัดสินว่า ค่าของตัวแปรตาม (Y_i) จะกระจายห่างจากค่าที่ประมาณได้ $(\hat{Y}_i)$ เพียงใด ถ้าข้อมูลกระจายห่างจากเส้นถดถอยมาก ค่า $\hat{Y}_i$ จะแตกต่างหรือคลาดเคลื่อนจาก Y_i จริงมาก ตรงกันข้ามถ้าข้อมูลกระจายห่างจากเส้นถดถอยน้อยนั่นคืออยู่ใกล้เส้นถดถอย ค่า $\hat{Y}_i$ จะมีความแตกต่างจากค่า Y_i จริงน้อย เส้นถดถอยที่คำนวณได้จะเป็นเส้นถดถอยที่น่าเชื่อถือได้ ค่าพยากรณ์จะมีความคลาดเคลื่อนน้อย



รูปที่ 2.8 แสดงการกระจายของข้อมูลจากเส้นถดถอย

ความคลาดเคลื่อนมาตรฐานของการกะประมาณ (SEE หรือ S) จะใช้เป็นตัวประมาณค่าแทนความเบี่ยงเบนมาตรฐานของประชากร (σ) ที่เราไม่ทราบค่า SEE จะอธิบายได้ดังสูตร

$$\text{SEE (หรือ S)} = \sqrt{S^2} = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n-2}} = \sqrt{\frac{\sum e_i^2}{n-2}}$$

โดยที่ $\sum e_i^2 = \sum y^2 - b^2 \sum x^2$

หรือ $\sum e_i^2 = \sum y^2 - b \sum xy$

อีกสูตรหนึ่งของ SEE ที่ใช้บ่อยคือ

$$\text{SEE} = \sqrt{\frac{\sum Y_i^2 - a\sum Y_i - b\sum X_i Y_i}{n-2}}$$

จากตัวอย่าง 2.1 จะหา SEE ได้โดยใช้สูตรใดสูตรหนึ่งข้างต้นได้เช่น

$$\text{SEE} = \sqrt{\frac{\sum e^2}{n-2}}$$

แต่ $\sum e^2 = \sum y^2 - b^2 \sum x^2$

$= 46 - (-2)^2(10) = 6$

$$\text{SEE} = \sqrt{\frac{6}{5-2}} = 1.4142$$

การตีความ : ความคลาดเคลื่อนมาตรฐานของการกะประมาณการเสนอซื้อส้มเท่ากับ 1.4142 กิโลกรัม SEE มีค่าไม่สูง แสดงว่ามีการกระจายของข้อมูลน้อยโดยข้อมูลกระจายอยู่ใกล้เส้นถดถอย สมการจะมีความน่าเชื่อถือได้ ถ้าหากเราทราบค่าของ α และ β จากสมการถดถอยที่แท้จริง ($\mu_{y.x} = \alpha + \beta X$) ส่วนเบี่ยงเบนมาตรฐาน($\sigma$)ของความคลาดเคลื่อนในการพยากรณ์การเสนอซื้อส้มที่อยู่บนพื้นฐานของเส้นถดถอยที่แท้จริงจะเท่ากับ 1.4142 กิโลกรัม

4. <u>ความคลาดเคลื่อนมาตรฐานของ a และb</u> (Standard Error of Estimate of a and b S_a, S_b)

ในการวิเคราะห์ถดถอย สิ่งที่เกี่ยวข้องสำคัญมิใช่แค่การพยากรณ์ หากแต่เป็นลักษณะในตัวของสมการเองได้แก่ความลาดชัน (Slope: β) และจุดตัด(Intercept : α)ของเส้นถดถอยที่แท้จริง ($\mu_{y.x} = \alpha + \beta X$) จนกระทั่งประมาณค่า (estimate)ได้และทดสอบ(test)ได้ ทำให้วินิจฉัยเกี่ยวกับ α และ β ได้

จากการที่เราศึกษาสมการถดถอยจากตัวอย่างไปหาประชากร S^2_a เป็นตัวประมาณค่าแทน σ^2_a และ S^2_b เป็นตัวประมาณค่าแทน σ^2_b โดยวิธีกำลังสองน้อยที่สุด S^2_a S^2_b จะให้ค่าไม่ลำเอียง มีความแปรปรวนน้อย เมื่อถอด square root S^2_a S^2_b จะให้ความคลาดเคลื่อนมาตรฐานของ a และ b S_a, S_b จะทำให้เราทราบการกระจายหรือความคลาดเคลื่อนของค่า a และ b

4.1) ความคลาดเคลื่อนมาตรฐานของ a มีสูตรดังนี้

$$S_a = \sqrt{S_a^2} = \sqrt{\frac{S^2 \sum X^2}{n \sum x^2}}$$

สูตรหนึ่งที่ใช้ได้ง่ายคือ

$$S_a = \sqrt{\frac{S^2 \sum X^2}{n(\sum X^2 - n\overline{X}^2)}}$$

4.2) ความคลาดเคลื่อนมาตรฐานของ b มีสูตรดังนี้

$$S_b = \sqrt{S_b^2} = \sqrt{\frac{S^2}{\sum x^2}} = \frac{S}{\sqrt{\sum x^2}}$$

อีกสูตรหนึ่งคือ $$S_b = \frac{S}{\sqrt{\sum X^2 - n\overline{X}^2}}$$

<u>ช่วงความเชื่อมั่นของการประมาณค่าพารามิเตอร์</u>

นักสถิติจะใช้ความคลาดเคลื่อนมาตรฐานช่วยในการประมาณช่วงความเชื่อมั่น(Confidence interval)ของค่าพารามิเตอร์ เพื่อให้การคาดคะเนหรือการพยากรณ์ค่าของประชากรอยู่ในช่วงที่ถูกต้องมากยิ่งขึ้น

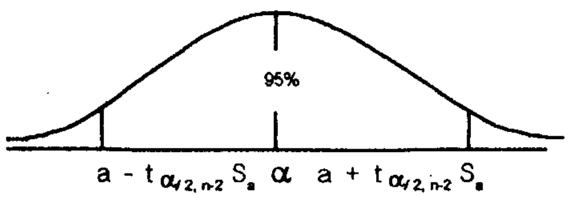
1. ช่วงความเชื่อมั่นของการคาดคะเน α

S_a จะช่วยการคำณวนช่วงการคาดคะเน α ณ ระดับความเชื่อมั่นที่กำหนดดังนี้

$$\alpha = a \pm t_{\alpha/2,\, n-2} S_a$$

หรือเขียนกระจายได้คือ

$$a - t_{\alpha/2,\, n-2} S_a < \alpha < a + t_{\alpha/2,\, n-2} S_a$$



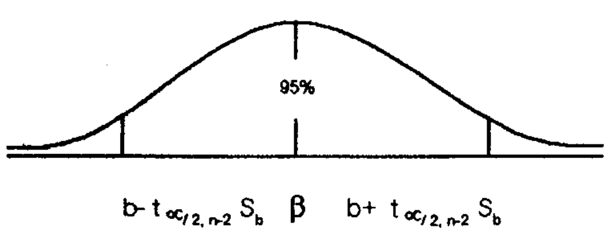
รูปที่ 2. 9 แสดง 95% ช่วงความเชื่อมั่นในการคาดคะเน α

2. ช่วงความเชื่อมั่นของการคาดคะเน β

เมื่อ S_b เป็นความคลาดเคลื่อนมาตรฐานของ b ช่วงคาดคะเน β ณ ระดับความเชื่อมั่นที่กำหนดจะหาได้ดังนี้

$$\beta = b \pm t_{\alpha/2,\, n-2} S_b$$

หรือเขียนกระจายได้คือ $b - t_{\alpha/2,\, n-2} S_b < \beta < b + t_{\alpha/2,\, n-2} S_b$



รูปที่ 2.10 แสดง 95% ช่วงความเชื่อมั่นในการคาดคะเน β

3. ช่วงความเชื่อมั่นของการพยากรณ์ค่าเฉลี่ย $\mu_{y.x}$

เมื่อ SEE เป็นความคลาดเคลื่อนมาตรฐานของการกะประมาณช่วงความเชื่อมั่นของการคาดคะเนค่าเฉลี่ย $\mu_{y.x}$ จะหาได้ดังนี้

$$\mu_{y.x} = \hat{Y} \pm t_{\alpha/2,n-2} SEE \sqrt{\frac{1}{n} + \frac{(X-\overline{X})^2}{\sum X^2 - n\overline{X}^2}}$$

หรืออีกสูตรหนึ่ง

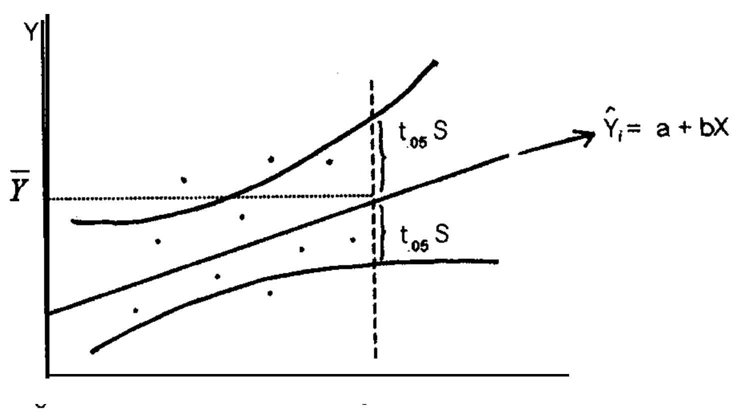
$$\mu_{y.x} = \hat{Y} \pm t_{\alpha/2,n-2} SEE \sqrt{\frac{1}{n} + \frac{x^2}{\sum x^2}}$$

4. ช่วงความเชื่อมั่นของการพยากรณ์ค่า Y แต่ละตัว

$$\mu_{y.x} = \hat{Y} \pm t_{\alpha/2,n-2} SEE \sqrt{\frac{1}{n} + \frac{(X-\overline{X})^2}{\sum X^2 - n\overline{X}^2} + 1}$$

หรืออีกสูตรหนึ่ง

$$\mu_{y.x} = \hat{Y} \pm t_{\alpha/2,n-2} SEE \sqrt{\frac{1}{n} + \frac{x^2}{\sum x^2} + 1}$$



รูปที่ 2.11 แสดง 95 เปอร์เซนต์ช่วงความเชื่อมั่นของการพยากรณ์

<u>ตัวอย่างที่ 2.3</u> จากข้อมูลปริมาณการเสนอซื้อส้มและราคาส้มในตัวอย่างที่ 2.1

ก. จงหาค่าความคลาดเคลื่อนมาตรฐานของ a และ b

ข. จงหาช่วงความเชื่อมั่น 95 เปอร์เซนต์ของ α และ β

ค. จงหาช่วงความเชื่อมั่น 95เปอร์เซนต์ของปริมาณการเสนอซื้อส้มที่แท้จริงเฉลี่ย ($\mu_{y.x}$) เมื่อส้มราคากิโลกรัมละ 10 บาท

ง. จงหาช่วงความเชื่อมั่น 95 เปอร์เซนต์ของปริมาณการเสนอซื้อส้มที่แท้จริง(Y) เมื่อส้มราคากิโลกรัมละ 10 บาท

<u>วิธีทำ</u> ก. หาค่าความคลาดเคลื่อนมาตรฐานของ a และ b

$$S_a = \sqrt{S_a^2}$$

$$= \sqrt{\frac{S^2\sum X^2}{n\sum x^2}}$$

$$= \sqrt{\frac{(1.4142)^2(90)}{5(10)}} \quad \text{เมื่อ } X^2 = 4,9,16,25,36 \ \sum X^2 = 90$$

$$= 1.8974$$

$$S_b = \sqrt{S_b^2}$$

$$= \sqrt{\frac{S^2}{\sum x^2}} = \frac{S}{\sqrt{\sum x^2}}$$

$$= \frac{1.4142}{\sqrt{10}}$$

$$= 0.4472$$

ข. จงหาช่วงความเชื่อมั่น 95 เปอร์เซนต์ของ α และ β

$$\alpha = a \pm t_{.05/2,\,5-2}\, S_a$$

$$= 15 \pm 3.182(1.8974)$$

$$= 15 \pm 6.0375$$

$$= [\ 9.9625\ ,\ 21.0375\ ]$$

การตีความหมายค่าช่วงของ α :

α มีค่าอยู่ระหว่าง 9.69 กิโลกรัม ถึง 21.04 กิโลกรัม ด้วยความเชื่อมั่น 95 เปอร์เซ็นต์ หมายความว่า ผลการคาดคะเน α ครั้งนี้ ถูกต้องกับความเป็นจริง 95 เปอร์เซ็นต์

$$\beta = b \pm t_{\alpha/2,\, n-2} S_b$$

$= -2 \pm 3.182\,(0.44720$

$= -2$ il.4230

$= [\ -0.57\ ,\ -3.43\ ]$

การตีความหมายค่าช่วงของ β :

β มีค่าอยู่ระหว่าง –0.57 กิโลกรัม ถึง –3.43 กิโลกรัม ด้วยความเชื่อมั่น 95 เปอร์เซ็นต์

ค. จงหาช่วงความเชื่อมั่น 95 เปอร์เซนต์ของปริมาณการเสนอซื้อส้มที่แท้จริงเฉลี่ย ($\mu_{y.x}$) เมื่อส้มราคากิโลกรัมละ 10 บาท

$$\mu_{y.x} = \hat{Y} \pm t_{\alpha/2,n-2} SEE \sqrt{\frac{1}{n} + \frac{x^2}{\sum x^2}}$$

$$\hat{Y} = 15 - 2(1) = 13$$

และ $x = X - \overline{X} = 1\text{-}4 = -3 \qquad x^2 = 9$

$$\mu_{y.x} = 13 \pm 3.182(\ 1.4142\ )\sqrt{\frac{1}{5} + \frac{9}{10}} = 13 \pm 4.95$$

$= [\ 8.05\ ,\ 17.95\ ]$

การตีความหมายของการคาดคะเน $\mu_{y.x}$

คาดคะเนได้ว่า เมื่อส้มราคากิโลกรัมละ 10 บาท จะมีปริมาณการซื้อเฉลี่ยอยู่ระหว่าง 8.05 กิโลกรัม ถึง 17.95 กิโลกรัม ด้วยระดับความมั่นใจ 95 เปอร์เซนต์

ง. จงหาช่วงความเชื่อมั่น 95 เปอร์เซนต์ของปริมาณการเสนอซื้อส้มที่แท้จริง(Y) เมื่อส้มราคากิโลกรัมละ 10 บาท

$$\mu_{y.x} = \hat{Y} \pm t_{\alpha/2,n-2} SEE \sqrt{\frac{1}{n} + \frac{x^2}{\sum x^2} + 1}$$

$$= 13 \pm 1.382(\ 1.4142\ )\sqrt{\frac{1}{5} + \frac{9}{10} + 1}$$

$= 13 \pm 5.4407$

$= [\ 7.5593\ ,\ 18.4407\ ]$

การตีความหมายของการคาดคะเน Y เมื่อส้มราคากิโลกรัมละ 10 บาท การเสนอซื้อส้มจะอยู่ระหว่างช่วง 7.6 กิโลกรัมถึง 18.4 กิโลกรัม ด้วยระดับความเชื่อมั่น 95 เปอร์เซ็นต์

5. <u>ค่าสถิติ t และการทดสอบ (t – test</u>)

การทดสอบค่าพารามิเตอร์ α และ β ที่ไม่ทราบค่า ต้องใช้ค่า a และ b ของสมการถดถอยจากตัวอย่างซึ่งมีการกระจายแบบ t ช่วยในการทดสอบแทนที่จะใช้การกระจายปกติ (Z)

5.1) การทดสอบค่า α

ตั้งสมมติฐาน Ho : $\alpha = 0$

Ha : $\alpha \neq 0$

สูตร $t_a = \dfrac{a}{S_a}$

เขตวิกฤติเมื่อ $t_a > t_{\alpha/2,\, n-2}$
$-t_a < -t_{\alpha/2,\, n-2}$

5.2) การทดสอบค่า β

การทดสอบที่สำคัญคือการทดสอบค่า β เนื่องจากค่า β เป็นค่าซึ่ง X อธิบายความผันแปรของ Y ทำให้ทราบถึงความสัมพันธ์ระหว่าง X และ Y ถ้า X และ Y มีความสัมพันธ์กัน เส้นถดถอยจะเป็นเส้นที่มีความลาดชัน(Slope) ถ้า X และ Y ไม่มีความสัมพันธ์กัน ค่า β จะเท่ากับ 0 เส้นถดถอยจะอยู่ในแนวนอน ไม่ว่า Xจะเปลี่ยนไปเท่าใดก็ตาม Y ยังอยู่คงที่

ตั้งสมมุติฐาน Ho : $\beta = 0$ (X อธิบายความผันแปรใน Y ไม่ได้)

Ha: $\beta \neq 0$ (X อธิบายความผันแปรใน Y ได้)

สูตร $t_b = \dfrac{b}{S_b}$

ถ้า **Ho : β = (มีค่าเป็นตัวเลข)**

Ha : $\beta \neq$ (มีค่าเป็นตัวเลข)

สูตร $t_b = \dfrac{b-\beta}{\sigma}$

เขตวิกฤติ เมื่อ $t_{\hat{b}} > t_{\alpha/2,\, n-2}$
$-t_{\hat{b}} < -t_{\alpha/2,\, n-2}$

การสรุปผล :

1. ถ้า t_b อยู่ในเขตวิกฤติ(Critical Region) จะปฏิเสธ Ho และยอมรับ Ha สรุปได้ว่า β มีนัยสำคัญทางสถิติตามระดับความเชื่อมั่นที่กำหนด
2. ถ้า t_b อยู่ในเขตยอมรับ(Accepted Region) จะยอมรับ Ho และปฏิเสธ Ha สรุปได้ว่า β ไม่มีนัยสำคัญทางสถิติ

6. ค่าสถิติ F และการทดสอบ (F-test)

การทดสอบสมมุติฐานของสัมประสิทธิ์เส้นถดถอย (β) ซึ่ง X อธิบายความผันแปรของ Y ได้อีกทางหนึ่งคือ ทดสอบด้วย F หรือ นิยมเรียกว่า การวิเคราะห์ความแปรปรวน (Analysis of Variance : ANOVA) ได้กล่าวถึงความแปรปรวนมาบ้างแล้วในหัวข้อข้างต้น ความแปรปรวนประกอบด้วย :

1. $\sum(Y_i - \overline{Y})^2$ คือ ความแปรปรวนทั้งหมด (Total Variation) เรียกอีกชื่อหนึ่งว่า Total Sum of Square (SST)
2. $\sum(\hat{Y}_i - \overline{Y})^2$ คือ ความแปรปรวนที่อธิบายได้ (Explained Variation) อีกชื่อหนึ่งคือ Regression Sum of Square (SSR)
3. $\sum(Y - \hat{Y}_i)^2$ คือ ความแปรปรวนที่อธิบายไม่ได้ (Unexplained Variation) อีกชื่อหนึ่งคือ *Error Sum of Square (SSE)*

จะเขียนในรูปสมการได้คือ SST = SSR + SSE

ค่าสถิติ F เป็นการเปรียบเทียบความเท่ากันของ 2 ความแปรปรวน ในรูปของ Mean of *Squares หรือ เปรียบเทียบ Variance ที่อธิบายได้ด้วย regression กับ Variance ที่อธิบายไม่ได้ด้วย* regression (เนื่องจาก error หรือ residual)อัตราส่วนของความแปรปรวนทั้งสองจะมีการแจกแจงแบบ F ซึ่งมี degree of freedom. ของความแปรปรวนที่เป็นเศษและความแปรปรวนที่เป็นส่วนคือ 1 และ n-2 ค่า F เขียนได้ดังนี้ :-

$$F = \frac{\text{Variance ที่อธิบายได้ด้วย regression}}{\text{Variance ที่อธิบายไม่ได้ด้วยregression}}$$

$$= \frac{S_1^2}{S_2^2}$$

วิธีทดสอบ

ตั้งสมมติฐาน Ho : $\beta = 0$ (x อธิบายความผันแปรใน Y ไม่ได้)

Ha : $\beta \neq 0$ (X อธิบายความผันแปรใน Y ได้)

สูตร $F = \frac{S_1^2}{S_2^2}$

เขตวิกฤติ	F > Fdf l,n-2

ตารางที่ 2.4 การวิเคราะห์ความแปรปรวน เพื่อทดสอบ Ho : $\beta = 0$

Source of Variation	d.f.	Sum of Square S.S.	Mean Square M.S.	F
Explained Variation	1	$\sum(\hat{Y}_i - \bar{Y})^2$	$\frac{\sum(\hat{Y}_i - \bar{Y})^2}{1} = S_1^2$	$\frac{S_1^2}{S_2^2}$
Unexplained Variation	n-2	$\sum(Y_i - \hat{Y}_i)^2$	$\frac{\sum(Y_i - \hat{Y}_i)^2}{n-2} = S_2^2$	
Total Variation	n-l	$\sum(Y_i - \bar{Y})^2$		

การสรุปผล : ถ้า F > F df 1,n-2 สรุปได้ว่า ปฏิเสธ Ho ยอมรับ Ha ตามระดับความเชื่อมั่นที่กำหนด

ถ้า F < F df 1,n-2 สรุปได้ว่า ยอมรับ Ho ปฏิเสธ Ha ตามระดับความเชื่อมั่นที่กำหนด

<u>ห</u>รืออีกวิธีหนึ่ง ตารางที่ 2.5 A NOVA

S V	df.	S S	M S	F
EV	1	$b\sum xy$ หรือ $b^2\sum x^2$	$b\sum xy/1$	
uv	n - 2	$\sum y^2 - b\sum xy$ หรือ $\sum e^2$.	$\sum e^2/n-2$	S_1^2/S_2^2
TV	n-l	$\sum y^2$		

<u>ความสัมพันธ์ระหว่าง F และ t</u>

การทดสอบ $\beta = 0$ โดยค่าสถิติ F และค่าสถิติ t จะให้คำตอบเหมือนกัน ทั้งนี้เพราะ F และ t มีความสัมพันธ์กันดังนี้

$$
\begin{aligned}
F &= \frac{b^2\sum x^2}{\dfrac{\sum y^2 - b\sum x^2}{n-2}} \\
&= \frac{b^2\sum x^2}{S^2} \\
&= \frac{b^2}{\dfrac{S^2}{\sum x^2}} \\
&= \left[\frac{b}{\dfrac{S}{\sqrt{\sum X^2}}}\right]^2 \\
&= \left[\frac{b}{S_b}\right] \\
&= t^2
\end{aligned}
$$

นั่นคือค่าวิกฤติของ F d.f. 1 และ n-2 จะเท่ากับค่า t^2 d.f. n-2 ตัวอย่างเช่น $F_{.05,1\text{ และ }3} = 10.13$ $t_{.025,3} = 3.182$ ดังนั้น $t^2 = (3.182)^2 = 10.13 = F$

ค่า F จะเท่ากับ t^2 ในกรณีถดถอยอย่างง่าย ซึ่งมีตัวแปรอิสระเพียง 1 ตัวเท่านั้น

<u>ความสัมพันธ์ระหว่างค่า F และ R^2</u>

$$F = \frac{b^2\sum x^2}{S^2}$$

$$= \frac{\hat{b}^2\sum x^2}{\frac{\sum e^2}{n-2}}$$

$$= \frac{(n-2)\hat{b}^2\sum x^2}{\sum e_i^2}$$

$$R^2 = 1-\frac{\sum e_i^2}{\sum y^2}$$

$$\sum e_i^2 = (I- R^2)\sum y^2$$

แทนค่า $\sum e_i^2$ ในสมการ

$$F = \frac{(n-2)\hat{b}^2\sum x^2}{(1-R^2)\sum y^2}$$

$\sum y^2$ หารทั้งเศษและส่วน

$$F = \frac{(n-2)\hat{b}^2\frac{\sum x^2}{\sum y^2}}{(1\text{-R'})}$$

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

ตัวอย่าง 2.4 จากสมการการเสนอซื้อส้ม Y = 15 – 2 X ในตัวอย่างที่ 2.1

ก. จงทดสอบ Ho : $\beta = 0$ กับ Ha : $\beta \neq 0$ ด้วยการทดสอบแบบ t ระดับนัยสำคัญ .05

ข. จงทดสอบ Ho : $\beta = -1.5$ กับ Ha : $\beta \neq -1.5$ ด้วยการทดสอบแบบ t ระดับนัยสำคัญ .05

ค. จงทดสอบ Ho : $\beta = 0$ กับ Ha : $\beta \neq 0$ ด้วยการทดสอบแบบ F ระดับนัยสำคัญ .05

วิธีทำ

ก. จงทดสอบ Ho : $\beta = 0$ กับ Ha : $\beta \neq 0$ ด้วยการทดสอบแบบ t ระดับนัยสำคัญ.05

Ho: $\beta = 0$ (ราคาส้ม อธิบายปริมาณการเสนอซื้อส้มไม่ได้)

Ha : $\beta \neq 0$

$$t = \frac{b}{S_b} = \frac{-2}{0.4472} = -4.472$$

สรุปผล : ค่า $t = -4.4723 < t_{.025,3} = -3.182$ เราปฏิเสธ Ho เป็นการชี้ชัดเจนว่า เส้นถดถอยที่แท้จริงมีความลาด (Slope) เป็นลบ ไม่ใช่ 0 เรียกว่า β มีนัยสำคัญทางสถิติ แสดงว่า ราคาส้มสามารถอธิบายความผันแปรของปริมาณการเสนอซื้อส้มได้ในทางกลับกัน นั่นคือ ราคาส้มกับการเสนอซื้อส้มมีความสัมพันธ์กันไปในทางตรงกันข้าม

ข. จงทดสอบ Ho : $\beta = -1.5$ กับ Ha : $\beta \neq -1.5$ ด้วยการทดสอบแบบ t ระดับนัยสำคัญ .05

Ho: $\beta = -1.5$

Ha : $\beta \neq -1.5$

$$t = \frac{b-\beta}{S_b} = \frac{-2-(-1.5)}{0.4472} = -1.1181$$

$t_{.025,3} = -3.182$ เรายอมรับ Ho แสดงว่า $\beta = -1.5$ ณ ระดับความเชื่อมั่น 95 เปอร์เซ็นต์ การสรุปผลเช่นเดียวกับข้างต้น

ค. . จงทดสอบ Ho : $\beta = 0$ กับ Ha : $\beta \neq 0$ ด้วยการทดสอบแบบ F ระดับนัยสำคัญ .05

Ho : $\beta = 0$

Ha : $\beta \neq 0$

เขตวิกฤติ F > F d.f. 1,n-2

ตารางที่ 2.6 การวิเคราะห์ความแปรแปรวน

SV	d.f	SS	M S	F-Ratio
EV	1	$b^2\sum x^2 = (-2)^2(10)$ $= 40$	$\frac{40}{1} = S_1^2$	$\frac{S_1^2}{S_2^2} = \frac{40}{2}$ $= 20$
UV	n-2 = 3	$\sum e^2 = 6$	$\frac{6}{3} = 2 = S_2^2$	
TV	n-1 = 4	$\sum y^2 = 46$		

$F_{1,3} = 10.13$

$F > F_{1,3}$ เราปฏิเสธ Ho : β มีนัยสำคัญทางสถิติ ณ ระดับความเชื่อมั่น 95 เปอร์เซ็นต์ แสดงว่า ด้วยความเชื่อมั่น 95 เปอร์เซ็น ราคาส้มสามารถจะอธิบายความผันแปรในปริมาณการเสนอซื้อส้มได้ โดยราคาส้มและปริมาณการเสนอซื้อส้มมีความสัมพันธ์กันอย่างผกผัน

<u>การนำเสนอเส้นถดถอย</u>

การเสนอเส้นถดถอยที่คำนวณได้ควรเสนอไปพร้อมกับค่าสถิติต่าง ๆ ดังรูปแบบข้างล่างนี้ คือ

$$\hat{Y} = a + bX$$
$$(S_{\hat{a}}) \quad (S_{\hat{b}})$$
$$t_{\hat{a}} \quad t_{\hat{b}}$$

R^2 = ... $\overline{R}^2$ = ... S= .... F =. ..

เช่น การเสนอสมการ demand จากตัวอย่างที่ 2.1

Q_d = **15 - 2P**

(1.8974) (0.4472)

t = 7.9056 **5.554**

R^2 = **0.87** $\overline{R}^2$ **= 0.83** S = **1.4142** **F = 20**

แบบฝึกหัด

1.จากตัวอย่างที่ 2.1 ให้คำณวนหาเส้น demand โดยไม่ต้องเปลี่ยนตัวแปรให้อยู่ในรูปเบี่ยงเบนจาก mean

2. กำหนดราคาและปริมาณการเสนอขายสินค้า Y ดังต่อไปนี้

ราคา (บาท /ก.ก.)	15	25	20	40	35	60	50	60
ปริมาณ (ก.ก.)	44	60	40	45	85	100	90	110

2.1 จงคำณวนสมการถดถอยโดยวิธีกำลังสองน้อยที่สุด

2.2 จงคำณวณความคลาดเคลื่อนมาตรฐานของการกะประมาณ

2.3 จงทดสอบสัมประสิทธิ์ถดถอยด้วย t ณ ระดับนัยสำคัญ .05 และ .01

2.4 จงสร้างตารางความแปรปรวนเพื่อทดสอบสัมประสิทธิ์ถดถอยด้วย F ณ ระดับนัยสำคัญ .05

2.5 จงพยากรณ์ช่วงการเสนอขายเฉลี่ยของสินค้า Y เมื่อราคาสินค้า Y ก.ก. ละ 80 บาท ในช่วงความเชื่อมั่น 95 เปอร์เซ็นต์

3.สมมติสุ่มตัวอย่างครอบครัวมา 5 ครอบครัวซึ่งมีรายได้และการออมดังนี้

รายได้ (1,000 บาท)	8	11	9	6	6
การออม (1.000 บาท)	0.6	1.2	1	0.7	0.3

3.1 จงประมาณสมการถดถอย $S = a + by$ และ $S = a + bY$

3.2 จงตีความหมายค่า a และ b จากทั้ง 2 สมการ

3.3 ค่า b ในสมการหลังคือค่าอะไร

3.4 จงคำณวนสมการถดถอยของการบริโภค

4.สมมติจากการทดลองใช้ปุ๋ยในที่นาแปลงหนึ่ง ปรากฏผลผลิตข้าวและการใช้ปุ๋ยดังตารางข้างล่างนี้

ผลผลิต (ถัง /ไร่)	40	45	50	65	70	70	80
ปุ๋ย (ก.ก./ ไร่)	100	200	300	400	500	600	700

4.1 จงคำณวนหาความสัมพันธ์ของผลผลิตข้าวกับการใช้ปุ๋ยโดยวิธีกำลังสองน้อยที่สุด

4.2 ผลผลิตเพิ่ม (Marginal Physical Product) ของการใช้ปุ๋ยเป็นเท่าใด ถ้าราคาข้าวถังละ 50 บาท รายได้เพิ่มจะเป็นเท่าใด

4.3 ถ้าใส่ปุ๋ยไร่ละ 800 ก.ก. ผลผลิตข้าวทั้งหมดจะเป็นเท่าใด

4.4 จงประมาณช่วงผลผลิตข้าวเฉลี่ย (μ_{yx}) เมื่อใส่ปุ๋ยไร่ละ 800 กิโลกรัม โดยใช้ช่วงความเชื่อมั่น 95 เปอร์เซ็นต์

5. ผู้จัดการบริษัทแปรรูปอาหารได้ใช้สมการถดถอยกำลังสองน้อยที่สุดเพื่อพยากรณ์ต้นทุนการผลิตอาหารสำเร็จรูปดังนี้

ผลผลิต (ตัน)	500	300	100	150	150	220	350
ต้นทุนทั้งหมด (ล้านบาท)	200	140	65	70	65	125	190

1. จงคำณวนสมการถดถอยโดยวิธีกำลังสองน้อยที่สุด
2. อธิบายความหมายของความลาดชัน b และจุดตัด a ของเส้นถดถอย
3. พยากรณ์ต้นทุนทั้งหมดต่อการผลิตจำนวน 400 ตัน

6. สมการถดถอยของการประมาณการผลผลิตข้าวโพดต่อไร่ Y = 150 + 0.1 X เมื่อใส่ปุ๋ย X กิโลกรัม สมการคำณวนมาจากตัวอย่าง 10 ตัวอย่าง ซึ่งมีความคลาดเคลื่อนมาตรฐาน 5 กิโลกรัม จงพยากรณ์ช่วงความเชื่อมั่น 95 เปอร์เซ็นต์ของผลผลิตข้าวโพดต่อไร่เฉลี่ย (μ_{yx}) และผลผลิตข้าวโพดต่อไร่ (Y) เมื่อ X เท่ากับ 50 กิโลกรัม สมมติ $\sum X^2$ = 650 และ $\overline{X}$ =15 กิโลกรัม

7. ความสัมพันธ์ของน้ำหนักของกระเป๋าเดินทาง Y กิโลกรัม ซึ่งเก็บไว้ในห้องเก็บกระเป๋าของเครื่องบิน และจำนวนผู้โดยสารเครื่องบิน X คน มีสมการดังนี้ Y = 250 + 27 X ผู้ควบคุมสนามบินจะใช้สมการนี้ตัดสินว่าจะเพิ่มน้ำหนักบนเครื่องบินอย่างปลอดภัยได้อีกเท่าไร หลังจากพิจารณาน้ำหนักบรรทุกน้ำมันและน้ำหนักผู้โดยสาร ข้อมูลได้รับจากตัวอย่างเครื่องบิน 25 ลำ ผลของตัวอย่างมีความคลาดเคลื่อนมาตรฐาน 100 กิโลกรัม $\sum X^2$ = 64000 $\overline{X}$ = 50 จงพยากรณ์ช่วงความเชื่อมั่น 95เปอร์เซ็นต์ของน้ำหนักกระเป๋าเดินทางเฉลี่ย (μ_{yx}) และช่วงของน้ำหนักกระเป๋า (Y) เมื่อจำนวนผู้โดยสาร 100 คน