บทที่ ๔

ผลการวิเคราะห์และการทดสอบโมเดล

๔.๑ การเก็บรวบรวมข้อมูล

ข้อมูลที่จะนำมาเพื่อใช้ในการวิเคราะห์ และการพยากรณ์ปริมาณผลผลิต น้ำตาลทรายและปริมาณการบริโภคน้ำตาลทรายภายในประเทศ สามารถจะเก็บรวบรวมข้อมูลได้จากหลายแหล่งด้วยกัน อาทิเช่น จากสำนักงานอ้อย และน้ำตาล กระทรวงอุตสาหกรรม สมาคมกลุ่มชาวไร่อ้อย สมาคมโรงงานน้ำตาล ฯลฯ สำหรับการวิเคราะห์ครั้งนี้ข้อมูลส่วนใหญ่ได้มาจาก สำนักงานอ้อยและน้ำตาล มีลักษณะข้อมูลเป็นแบบทุติยภูมิ (Secondary Data) ซึ่งทำการเก็บรวบรวมในระยะ ๑๔ จากปี ๒๕๐๕ ถึงปี ๒๕๑๘ รวม ๑๔ ปี ยกเว้นข้อมูลของการสต็อคน้ำตาล สามารถรวบรวมจากข้อมูลได้เพียง ๑๐ ปี ดังรายละเอียดจากแหล่งที่มาของข้อมูลดังนี้ :-

๔.๑.๑ ข้อมูลเกี่ยวกับพื้นที่ปลูกอ้อย

ได้จากสำนักงานอ้อยและน้ำตาล ซึ่งรวบรวมจากแหล่งปลูกอ้อยใน ๔ ภาค คือ ภาคเหนือ ภาคกลาง ภาคตะวันออก และภาคตะวันออกเฉียงเหนือ ดังรายละเอียดได้กล่าวแล้วในหัวข้อ ๒.๑.๒ ส่วนปริมาณพื้นที่ปลูกอ้อยมีรายละเอียดดังตารางที่ ๑

๔.๑.๒ ข้อมูลเกี่ยวกับผลผลิตอ้อยในภาคต่าง ๆ

ได้จากสำนักงานอ้อยและน้ำตาล ดังรายละเอียดในตารางที่ ๓

๔.๑.๓ ข้อมูลเกี่ยวกับราคารับซื้ออ้อยในภาคต่าง ๆ ๔

มีรายละเอียดดังตารางที่ ๕ ได้จากสำนักงานอ้อยและน้ำตาล ยกเว้นราคาเฉลี่ยซึ่งคำนวณได้จากการนำเอาผลบวกราคาอ้อยใน ๔ ภาค หารด้วย ๔

๔.๑.๔ ข้อมูลเกี่ยวกับปริมาณน้ำตาลทรายที่ผลิตได้

มีรายละเอียดดังตารางที่ ๖ ได้มาจากหน้า ๗๒ Industrial Statistics ๑๙๗๕ ของกระทรวงอุตสาหกรรม ยกเว้นปี ๒๕๑๘ ได้มาจากสำนักงานอ้อยและน้ำตาล

๔.๑.๕ ข้อมูลเกี่ยวกับผลผลิตน้ำตาลทรายชนิดต่าง ๆ

มีรายละเอียดดังตารางที่ ๗ ได้มาจากหน้า ๗๓ Industrial Statistics

๑๙๗๕ ของกระทรวงอุตสาหกรรมยกเว้นปี ๒๕๑๘ ได้มาจากสำนักงานอ้อยและน้ำตาล

๔.๑.๖ ข้อมูลเกี่ยวกับปริมาณการส่งน้ำตาลจำหน่ายต่างประเทศ

มีรายละเอียดดังตารางที่ ๑๒ ได้จากสำนักงานอ้อยและน้ำตาล ยกเว้นปี ๒๕๑๘ ได้จากหน้า ๔ รายงานเศรษฐกิจเดือนมกราคม ๒๕๑๙ ธนาคารกรุงไทย

๔.๑.๗ ข้อมูลเกี่ยวกับปริมาณการบริโภคน้ำตาลทรายภายในประเทศ

มีรายละเอียดดังตารางที่ ๑๔ ได้มาจากสำนักงานอ้อยและน้ำตาล

๔.๑.๘ ข้อมูลเกี่ยวกับราคาน้ำตาลขายปลีกและส่ง

มีรายละเอียดดังตารางที่ ๑๕ ได้มาจาก กรมการค้าภายใน กระทรวงพาณิชย์

๔.๑.๙ ข้อมูลเกี่ยวกับจำนวนประชากร

มีรายละเอียดดังตารางที่ ๑๗ ได้มาจากหลายแหล่งดังนี้

ก. จากหน้า ๑๐ สถิติที่สำคัญของประเทศไทย ปี ๒๕๑๑ - ๒๕๑๒

ข. จากหน้า ๒ ECONOMIC PROGRESS OF THAILAND ; June 1973

ค. จากหน้า ๒๐๓ วารสารประชากรศึกษา เดือนกรกฎาคม ๒๕๑๘ มหาวิทยาลัยมหิดล

ง. ปี ๒๕๑๘ ได้จากการเพิ่มประชากร ๓% ตามแผนพัฒนาเศรษฐกิจแห่งชาติระยะที่ ๓

๔.๑.๑๐ ข้อมูลเกี่ยวกับรายได้ประชาชาติต่อคน

มีรายละเอียดดังตารางที่ ๑๗ ได้มาจากหลายแหล่งดังนี้

ก. จากหน้า ๘๕ รายได้ประชาชาติ ฉบับที่ ๒๕๐๙

ข. จากหน้า ๑๓ รายได้ประชาชาติ ฉบับปี ๒๕๑๕ - ๑๖

ค. จากหน้า ๒ ECONOMIC PROGRESS OF THAILAND, June 1972

ง. จากหน้า ๒๖ ECONOMIC PROGRESS OF THAILAND, June 1973

๔.๑.๑๒ ข้อมูลเกี่ยวกับปริมาณการสต็อคน้ำตาล

มีรายละเอียดดังตารางที่ **๑๙** ได้มาจากสำนักงานอ้อยและน้ำตาล

๔.๒ ผลการคำนวณโมเดลโดยใช้ **Time Series**

ข้อมูลที่ใช้ในการคำนวณมี ๑๔ ปี เพื่อการสดวกในการคำนวณโดยการเลือก ให้ปี พ.ศ. ๒๕๐๕ มีค่า X = − ๑๓ ; เมื่อ X เป็นตัวแปรอิสระ (แทนปี พ.ศ.) ดังตารางต่อไปนี้

ปี พ.ศ.	X	X^2	X^3	X^4
๒๕๐๕	− ๑๓	๑๖๙	− ๒๑๙๗	๒๘,๕๖๑
๒๕๐๖	− ๑๑	๑๒๑	− ๑๓๓๑	๑๔,๖๔๑
๒๕๐๗	− ๙	๘๑	− ๗๒๙	๖,๕๖๑
๒๕๐๘	− ๗	๔๙	− ๓๔๓	๒,๔๐๑
๒๕๐๙	− ๕	๒๕	− ๑๒๕	๖๒๕
๒๕๑๐	− ๓	๙	− ๒๗	๘๑
๒๕๑๑	− ๑	๑	− ๑	๑
๒๕๑๒	๑	๑	๑	๑
๒๕๑๓	๓	๙	๒๗	๘๑
๒๕๑๔	๕	๒๕	๑๒๕	๖๒๕
๒๕๑๕	๗	๔๙	๓๔๓	๒,๔๐๑
๒๕๑๖	๙	๘๑	๗๒๙	๖,๕๖๑
๒๕๑๗	๑๑	๑๒๑	๑๓๓๑	๑๔,๖๔๑
๒๕๑๘	๑๓	๑๖๙	๒๑๙๗	๒๘,๕๖๑
$\sum$	๐	๙๑๐	๐	๑๐๕,๗๔๒

โมเดลของแนวโน้มเส้นตรง

$$\hat{Y}_T = a + b X$$

จาก (2) ใน 3.2.1 จะได้

$$\left.\begin{aligned} a &= \Sigma Y / 14 \\ b &= \Sigma XY / 910 \end{aligned}\right\} \text{-----} (2)'$$

โมเดลของแนวโน้มโพลีโนเมี่ยลกำลังสอง

$$\hat{Y}_T = a + b X + c X^2$$

จาก (7)' ใน 3.2.2 จะได้

$$\left.\begin{aligned} b &= \Sigma XY / 910 = 0.001099 \Sigma XY \\ a &= 0.162199 \Sigma XY - 0.001395 \Sigma X^2 Y \\ c &= 0.0000214 \Sigma X^2 Y - 0.001395 \Sigma Y \end{aligned}\right\} (7)'$$

โมเดลของแนวโน้มเอกโพเนนเชี่ยล

$$\hat{Y}_T = ab^X$$

หรือ $\log \hat{Y}_T = \log a + X \log b$

จาก (8) ใน 3.2.2 จะได้

$$\left.\begin{aligned} \log a &= (\Sigma \log Y)/14 \\ \log b &= (\Sigma X \log Y) / 910 \end{aligned}\right\} \text{----} (8)$$

๔.๒.๑ ผลการคำนวณโมเดลเกี่ยวกับปริมาณการบริโภคน้ำตาลทรายรวมภายในประเทศ

จากลักษณะของข้อมูลดิบที่นำมาทดลองเขียนกราฟดูจะเห็นมีลักษณะเป็นเส้นตรง หรือโค้งเล็กน้อย ดังนั้นจึงพิจารณาลักษณะสมการเส้นตรงเสียก่อนดังนี้

พิจารณา โมเดลของ $\hat{Y}_T = a + b X$ เทียบกับ $\hat{Y}_T = ab^X$ หรือเมื่อเปลี่ยนรูปให้อยู่ในลักษณะเส้นตรงคือ $\log \hat{Y}_T = \log a + (\log b) X$

ผลการคำนวณโมเดล $\hat{Y}_T = a + b X$

จะได้ค่า $a = 265.6002$

$b = 12.0787$

$r = 0.9418$, $r^2 = 0.8869$

$\sum (Y_i - \hat{Y}_i)^2 = 16,893.2447$

% ของการกระจาย = 88.69 %

ผลการคำนวณโมเดล $\log \hat{Y}_T = \log a + (\log b) X$

จะได้ค่า $\log a = 2.3804$

$\log b = 0.0234$

$r = 0.8988$, $r^2 = 0.8078$

จะเห็นได้ว่าลักษณะของโมเดล น่าจะอยู่ในรูปสมการเส้นตรงมากกว่าอยู่ในรูปสมการเอกโพเนนเชียล เพราะค่าของ r จากโมเดลเส้นตรงมีค่ามากกว่า แสดงว่าโมเดลที่สร้างขึ้นโดย $\hat{Y}_T = a + b X$ มีความสัมพันธ์ดีกว่าโมเดลเอกโพเนนเชียล

ผลการคำนวณโมเดล $\hat{Y} = a + b X + c X^2$

จะได้ค่า $a = 281.7442$

$b = 12.0798$

$c = -0.2622$

ได้ $R^2 = 0.9061$

$\sum (Y_i - \hat{Y}_i)^2 = 14,046.5676$

% ของการกระจาย = 90.61 %

เมื่อนำผลการคำนวณจากโมเดลนี้ เทียบกับโมเดล $\hat{Y}_T = a + b X$ จะเห็นได้ว่าโมเดล $\hat{Y}_T = a + bX + c X^2$ มีค่า R^2 สูงกว่า และค่า $\sum (Y_i - \hat{Y}_i)^2$ น้อยกว่า ดังนั้นแสดงว่าโมเดลที่จะใช้แสดงปริมาณการบริโภคน้ำตาลทรายรวมจะมีลักษณะ

อยู่ในรูปโพลี่โนเมี่ยลกำลังสอง

$$\hat{Y}_T = a + b\,X + c\,X^2$$

หรือ

$$\hat{Y}_T = 281.7442 + 12.0798\,X - 0.2622\,X^2$$

Y_T คือปริมาณการบริโภครวมน้ำตาลทรายมีหน่วย : พันตัน

a เป็นค่า $\hat{Y}$ ที่ X มีค่า = ๐

b เป็นค่าความชันของเส้นแนวโน้ม

c เป็นอัตราการเปลี่ยนแปลงของความชัน

$\hat{Y}_T$ เป็นค่าประมาณการบริโภครวมน้ำตาลทราย หน่วย : พันตัน

X เป็นค่าของปีที่มีการบริโภค ให้ปี ๒๕๐๕ มีค่า $X = -13$

X_i	Y_i	$X_i Y_i$	$X_i^2 Y_i$	$\hat{Y}_i$	$(Y_i - \hat{Y}_i)^2$
– ๑๓	๑๒๕.๐๗๗	–๑,๖๒๖.๐๐๑	๒๑,๑๓๘.๐๑๓	๘๐.๓๙๕๐	๑,๙๙๖.๔๘๑๑
– ๑๑	๘๑.๙๔๕	– ๙๐๑.๓๙๕	๙,๙๑๕.๓๔๕	๑๑๗.๑๔๐๒	๑,๒๓๘.๗๐๒๑
– ๙	๑๑๗.๘๕๔	–๑,๐๖๐.๖๘๖	๙,๕๔๖.๑๗๔	๑๕๑.๗๘๗๘	๑,๑๕๑.๕๐๒๗
– ๗	๒๓๓.๐๖๗	–๑,๖๓๑.๔๖๙	๑๑,๔๒๐.๒๘๓	๑๘๔.๓๓๗๘	๒,๓๗๔.๕๓๔๙
– ๕	๑๘๔.๓๑๗	– ๙๒๑.๕๘๕	๔,๖๐๗.๙๒๕	๒๑๔.๗๙๐๒	๙๒๘.๖๑๕๙
– ๓	๒๓๓.๗๒๖	– ๗๐๑.๑๗๘	๒,๑๐๓.๕๓๔	๒๔๓.๑๔๕๐	๘๘.๗๑๗๖
– ๑	๒๓๘.๘๐๔	– ๒๓๘.๘๐๔	๒๓๘.๘๐๔	๒๖๙.๔๐๒๒	๙๓๖.๒๔๙๘
๑	๓๒๔.๔๕๙	๓๒๔.๔๕๙	๓๒๔.๔๕๙	๒๙๓.๕๖๑๘	๙๕๔.๖๓๖๘
๓	๓๕๗.๕๕๙	๑,๐๗๒.๖๗๗	๓,๒๑๘.๐๓๑	๓๑๕.๖๒๓๘	๑,๗๕๘.๕๖๐๘
๕	๓๖๔.๖๘๐	๑,๘๒๓.๔๐๐	๙,๑๑๗.๐๐๐	๓๓๕.๕๘๘๒	๘๔๖.๓๓๒๘
๗	๓๑๒.๓๐๒	๒,๑๘๖.๑๑๔	๑๕,๓๐๒.๗๙๘	๓๕๓.๔๕๕๐	๑,๖๙๓.๕๖๙๔
๙	๓๖๒.๓๑๖	๓,๒๖๐.๘๔๔	๒๙,๓๔๗.๕๙๖	๓๖๙.๒๒๔๒	๔๗.๗๒๓๒
๑๑	๓๘๒.๒๙๘	๔,๒๐๕.๒๗๘	๔๖,๒๕๘.๐๕๘	๓๘๒.๘๙๕๘	๐.๓๕๗๓
๑๓	๔๐๐.๐๐๐	๕,๒๐๐.๐๐๐	๖๗,๖๐๐.๐๐๐	๓๙๔.๔๖๙๘	๓๐.๕๘๓๑
$\sum$	๓,๗๑๘.๔๐๔	๑๐,๙๙๑.๖๕๔	๒๓๐,๑๓๘.๐๒๐	๓,๗๐๕.๘๑๖๘	๑๔,๐๔๖.๕๖๗๖

$$\overline{Y} = \Sigma Y/14 = 265.60028$$

$$\Sigma Y_i^2 = 1,137,267.7476 \quad , \Sigma \hat{Y}_i^2 = 1,116,925.9263$$

$$\Sigma (Y_i - \overline{Y})^2 = 149,658.5828$$

เปอร์เซ็นต์ของการกระจาย Y = 90.61 %

ค่า F_c ที่ได้จากการคำนวณในตาราง ANOVA

$$F_c = \frac{MSR}{MSE} = \frac{SSR / k}{SSE / (n-k-1)}$$

k = d.f ของ Regression ; k = 2

n-k-1 = d.f ของ Residual ; n-k-1 = 14-2-1 = 11

$$F_c = \frac{11}{2} \left\{ \frac{SSR}{SSE} \right\} = 5.5 \left\{ \frac{SSR}{SSE} \right\}$$

$$SSE = \Sigma (Y_i - \hat{Y}_i)^2$$

$$SSR = SST - SSE = \Sigma (Y - \overline{Y})^2 - \Sigma (Y_i - \hat{Y}_i)^2$$

$$F_c = 5.5 \left\{ \frac{\Sigma Y_i^2 - (\Sigma Y_i)^2/n - \Sigma (Y_i - \hat{Y}_i)^2}{\Sigma (Y_i - \hat{Y}_i)^2} \right\}$$

$$= 5.5 \left\{ \frac{(1,137,267.7476 - 987,609.1648) - 14,046.5676}{14,046.5676} \right\}$$

$$= 5.5 \left\{ \frac{149,658.5828 - 14,046.5676}{14,046.5676} \right\}$$

$$F_c = 53.0992 \quad , \quad d.f\ (2,11)$$

การตรวจสอบความมีนัยสำคัญของโมเดล

ค่า F จากตารางที่ d.f (2,11)

ณ. ระดับนัยสำคัญ α = ๐.๐๕ ค่า F_t = ๓.๘๘

= ๐.๐๑ F_t = ๗.๒๐

จาก Ho : $B = C = 0$

เมื่อคำนวณได้ $F_c > F_t$ ที่ α =๐.๐๕ และ .๐๑ แสดง ว่า เรา Reject Ho นั่นคือ ยอมรับว่า โมเดลที่ใช้อยู่ในรูปโพลีโนเมียลอันดับสอง น่าจะใช้ได้กับการพยากรณ์ปริมาณการบริโภคน้ำตาลทรายรวมภายในประเทศ

แต่เมื่อนำโมเดลนี้ไปทำการทดลองพยากรณ์หาค่าประมาณในอนาคตจะพบ ว่า มีลักษณะ Converge เร็วดังนั้นจึงไม่เหมาะสมที่จะใช้ พยากรณ์ต่อไป ผลที่ได้จากการพยากรณ์ใน ๗ ปีข้างหน้ามีดังนี้

ปี	ค่าพยากรณ์ปริมาณการบริโภค	
๒๕๑๙	๕๐๓.๘๘๖๒	พันตัน
๒๕๒๐	๕๑๑.๗๒๕๐	พันตัน
๒๕๒๑	๕๑๖.๖๐๖๒	พันตัน
๒๕๒๒	๕๑๙.๗๘๘๘	พันตัน
๒๕๒๓	๕๒๐.๘๗๕๗	พันตัน
๒๕๒๔	๕๑๙.๘๖๔๒	พันตัน
๒๕๒๕	๕๑๖.๗๕๕๐	พันตัน

ค่าพยากรณ์ปี ๒๕๒๓ จะให้ค่าพยากรณ์สูงสุด แต่หลังจากนั้นปริมาณการบริโภคจะเริ่มลดลงต่อไปเรื่อย ซึ่งขัดกับข้อเท็จจริงที่ว่าผู้บริโภคน้ำตาลทรายได้แก่ประชากร ในประเทศและโรงงานอุตสาหกรรมบางประเภท จำนวนประชากรนับวันจะมีเพิ่มขึ้นทุกปี ดังตารางที่ ๑๗ประกอบกับปริมาณการบริโภคของประชาชนส่วนใหญ่เป็นการบริโภคโดย ทางตรง ส่วนปริมาณการบริโภคทางอ้อมซึ่งเป็นผลเนื่องจากการใช้น้ำตาลในโรงงานอุตสาหกรรม มีปริมาณประมาณ ๑/๓ ของปริมาณการบริโภครวมภายในประเทศดังรายละเอียดในตารางที่ ๑๗

การที่นำข้อมูลปริมาณการบริโภคน้ำตาล มาทำการทดลองหาเส้นแนวโน้มตามลำดับเวลา แล้วทดสอบโมเดลที่คาดว่าจะใช้เป็นตัวแทนของข้อมูลได้โดยใช้ F test ปรากฏว่าเส้นแนวโน้มจะมีลักษณะของโพลิโนเมี่ยลกำลังสอง และถ้าขอบเขตของการวิจัยครั้งนี้มิได้กำหนดโมเดลที่จะใช้ทดสอบเพียง ๓ โมเดลแล้ว เราก็ควรที่จะพิจารณาทำการทดสอบโมเดลโพลิโนเมี่ยลกำลังที่สูงกว่านี้ขึ้นไปอีก ดังวิธีการทดสอบโมเดลโพลิโนเมี่ยลว่าข้อมูลที่ใช้ควรจะอยู่ในรูปของโมเดลโพลิโนเมี่ยลกำลังเท่าใด ตามภาคผนวก

แต่เมื่อกำหนดโมเดลให้เลือกใช้เพียง ๓ โมเดล ดังนั้นจึงควรทำการพิจารณาเลือกโมเดลที่จะใช้เป็นตัวแทนของข้อมูล โดยเลือกโมเดลที่ให้ค่า R^2 ต่ำกว่าเดิม นั่นคือสมควรพิจารณาทดสอบโมเดลของเส้นตรง

$$\hat{Y}_T = a + bX$$

หรือ $\hat{Y}_T = 265.6002 + 12.0787 X$

X_i เป็นปีที่มีการบริโภคน้ำตาล โดยให้ปี ๒๕๐๕ มีค่า $X = -13$ ปีที่เป็นจุดเริ่มต้นคือกลางปี ๒๕๑๑ และค่าของ X เพิ่มขึ้นหรือลดปีละ ๒

$\hat{Y}_i$ เป็นค่าประมาณของปริมาณการบริโภคน้ำตาล มีหน่วยเป็นพันตัน

$$R^2 = 0.8869$$

$$S.E = 37.5202 \quad ; \; F_c = 93.308$$

$$\Sigma (Y_i - \hat{Y}_i)^2 = 16,893.2447$$

เปอร์เซนต์การกระจายของ $Y = 88.69\ \%$

$$\Sigma (Y_i - \bar{Y})^2 = 16,893.2447$$

รายละเอียดในการคำนวณมีดังนี้

ปี (X_i)	Y_i	$\hat{Y}_i$	$(Y_i - \hat{Y}_i)^2$
− ๑๓	๑๒๕.๐๗๗	๑๐๘.๕๗๗๑	๒๗๒.๒๔๖๗
− ๑๑	๘๑.๙๔๕	๑๓๒.๗๓๔๕	๒๕๗๙.๕๗๒๓
− ๙	๑๑๗.๘๕๔	๑๕๖.๘๙๑๙	๑๕๒๓.๙๕๗๖
− ๗	๒๓๓.๐๖๗	๑๘๑.๐๔๙๓	๒๗๐๕.๘๔๑๑
− ๕	๑๘๔.๓๑๗	๒๐๕.๒๐๖๗	๔๓๖.๓๗๙๕
− ๓	๒๓๓.๗๒๖	๒๒๙.๓๖๔๑	๑๙.๐๒๖๑
− ๑	๒๓๘.๘๐๔	๒๕๓.๕๒๑๕	๒๑๖.๖๐๔๘
๑	๓๒๔.๔๕๙	๒๗๗.๖๗๘๙	๒๑๘๘.๓๗๗๗
๓	๓๕๗.๕๕๙	๓๐๑.๘๓๖๓	๓๑๐๕.๐๑๙๒
๕	๓๖๔.๖๘๐	๓๒๕.๙๙๓๗	๑๔๙๖.๖๒๘๘
๗	๓๑๒.๓๐๒	๓๕๐.๑๕๑๑	๑๔๓๒.๕๕๔๓
๙	๓๖๒.๓๑๖	๓๗๔.๓๐๘๕	๑๔๓.๘๒๐๐
๑๑	๓๘๒.๒๙๘	๓๙๘.๔๖๕๙	๒๖๑.๔๐๐๙
๑๓	๔๐๐.๐๐๐	๔๒๒.๖๒๓๓	๕๑๑.๘๑๓๗
$\sum$	๓,๗๑๘.๔๐๔	๓,๗๑๘.๔๐๒๘	๑๖,๘๙๓.๒๘๘๗

$$\bar{Y} = \sum Y/14 = \frac{3,718.404}{14} = 265.60028$$

$$\sum \hat{Y}_i^2 = 1,120,372.0062$$

$$\sum Y_i^2 = 1,137,267.7476$$

$$\sum X\,Y = 10,991.6540$$

% ของการกระจาย $= \frac{\sum \hat{Y}_i^2 - (\sum Y_i)^2/n}{\sum (Y_i - \bar{Y})^2} \times 100$

เปอร์เซ็นต์การกระจายของ Y = $\frac{132,765.3381}{149,658.5828}$ X 100 = 88.69 %

F_c จาก ANOVA = $\frac{(n-k-1)}{k}\left\{\frac{SSR}{SSE}\right\} = \frac{n-k-1}{k}\left\{\frac{SST - SSE}{SSE}\right\}$

แทนค่า n = 14 , k = 1

$$F_c = 12\left\{\frac{149,658.5828 - 16,893.2447}{16,893.2447}\right\}$$

$$= 12\left\{\frac{132,765.3381}{16,893.2447}\right\}$$

$$= 12\ (7.859)$$

$$= 94.308$$

การตรวจสอบความมีนัยสำคัญของโมเดล

ค่า F จากตารางที่ d.f (1,12) ณ. ระดับนัยสำคัญ α = 0.01

ได้ F_t = 9.33

$\therefore\ F_c > F_t$

หมายความว่าเราจะยอมรับโมเดลเส้นตรงและต่อไปจะทำการทดสอบสัมประสิทธิ์ของตัวแปรอิสระคือ โดยใช้ t - test

$$t_c = \frac{b - \beta}{S_b} = \frac{b - 0}{S_b}$$

S_b = Standard Error ของ b

จากสูตร $V(b) = \frac{V(Y)}{\Sigma(X - \overline{X})^2} = \frac{V(Y)}{\Sigma X^2}$

V (b) = Variance b

$$S(b) = \frac{S(Y)}{\sqrt{\Sigma X^2}}, \quad \therefore\ t_c = \frac{b\sqrt{\Sigma X^2}}{S(Y)} = \frac{12.0787\sqrt{910}}{37.5202}$$

$$= 9.711$$

จากตาราง t - distribution ที่ d.f (n-k-1) = 12

มีค่า $t = 3.055$ ที่ $\alpha = .01$

ปรากฏว่า $t_c > t_t$

หมายความว่าจะปฏิเสธสมมติฐาน $H_0 : B = 0$

ดังนั้นจึงยอมรับว่าโมเดลเส้นตรงจะนำไปใช้เป็นตัวแทนในการพยากรณ์ได้ และมีผลของการประมาณค่าปริมาณการบริโภคไว้เรียบร้อยแล้ว

๔.๒.๒ ผลการคำนวณโมเดลเกี่ยวกับปริมาณการผลิตน้ำตาลทรายรวม ในประเทศ

พิจารณาจากลักษณะของข้อมูลดิบที่นำมาลองเขียนกราฟดู น่าจะได้โมเดลที่ อยู่ในรูปโพลีโนเมียลอันดับสองมากกว่าลักษณะอื่น ดังผลในการคำนวณเปรียบเทียบดังนี้

ผลการคำนวณโมเดลในรูป $\hat{Y}_T = a + bX$

จะได้ค่า $a = 415.7620$

$b = 30.5915$

$r = 0.8972$; $r^2 = 0.8049$

$\Sigma (Y_i - \hat{Y}_i)^2 = 206,309.6845$

% การกระจาย = 80.49 %

ผลการคำนวณโมเดลในรูป $\hat{Y}_T = ab^X$ หรือ $\log \hat{Y}_T = \log a + (\log b)(X)$

จะได้ค่า $\log a = 2.5292$

$\log b = 0.0324$

$r = 0.9394$; $r^2 = 0.8824$

$\Sigma (Y_i - \hat{Y}_i)^2 = 86,919.3434$

% การกระจาย = 88.24 %

จะเห็นว่าในการผลิตน้ำตาลโมเดลในรูป $\hat{Y}_T = ab^X$ น่าจะใช้เป็นตัวแทนได้ดีกว่า $\hat{Y}_T = a + bX$ เนื่องจากให้ค่า r^2 ที่สูงกว่าและ $\Sigma(Y_i - \hat{Y}_i)^2$ ต่ำกว่า

ผลการคำนวณโมเดลในรูป $\hat{Y}_T = a + bX + cX^2$

จะได้ค่า $a = 301.7361$

$b = 30.5943$

$c = 1.7264$

ได้ค่า $R^2 = 0.9405$

$\Sigma(Y_i - \hat{Y}_i)^2 = 62,914.1943$

เปอร์เซนต์การกระจายของ Y = 94.05 %

เมื่อนำผลที่ได้จากโมเดลนี้ เปรียบเทียบกับโมเดล $\hat{Y} = ab^X$ จะเห็นได้ว่ามีค่า R^2 สูงกว่า และค่า $\Sigma(Y_i - \hat{Y}_i)^2$ น้อยกว่า ดังนั้นโมเดลที่จะใช้เป็นตัวแทนเกี่ยวกับผลผลิตน้ำตาลทรายรวมภายในประเทศน่าจะอยู่ในรูปโพลีโนเมี่ยลกำลังสองดังนี้ :-

$$\hat{Y}_T = a + bX + cX^2$$

หรือ $$\hat{Y}_T = 301.7361 + 30.5943\,X + 1.7264\,X^2$$

$\hat{Y}$ ค่าประมาณปริมาณการผลิตน้ำตาลทรายรวมภายในประเทศมีหน่วย : พันตัน

<u>หมายเหตุ</u> การที่ไม่เลือกพิจารณากรณีโมเดลโพลิโนเมี่ยลกำลังที่สูงกว่า ก็มีเหตุผลในทำนองเดียวกันกับกรณีของปริมาณการบริโภคน้ำตาล

ปริมาณการผลิตน้ำตาลทราย

X_g	Y_i	X_iY_i	$X_i^2Y_i$	Y_i^2	$\hat{Y}_i$	$(Y_i - \hat{Y}_i)^2$
- ๑๓	๑๕๑.๓๔๓	-๑,๙๖๗.๔๕๙	๒๕,๕๗๖.๙๖๗	๒๒,๙๐๔.๗๐๓๖	๑๙๕.๗๗๑๘	๑,๙๗๓.๙๑๘๒
- ๑๑	๑๒๕.๐๓๑	-๑,๓๗๕.๓๔๑	๑๕,๑๒๘.๗๕๑	๑๕,๖๓๒.๗๕๐๙	๑๗๔.๐๙๓๒	๒,๔๐๗.๐๙๙๔
- ๙	๑๖๗.๙๗๓	-๑,๕๑๑.๗๕๗	๑๓,๖๐๕.๘๑๓	๒๘,๒๑๔.๙๒๘๗	๑๖๖.๒๒๕๘	๓.๐๕๒๗
- ๗	๓๑๙.๙๗๖	-๒,๒๓๙.๘๓๒	๑๕,๖๗๘.๘๒๔	๑๐๒,๓๘๔.๖๔๐๕	๑๗๒.๑๖๙๖	๒๑,๘๔๖.๗๓๑๘
- ๕	๒๖๙.๑๖๘	-๑,๓๔๕.๘๔๐	๖,๗๒๙.๒๐๐	๗๒,๔๕๑.๔๑๒๒	๑๙๑.๙๒๔๖	๕,๙๖๖.๕๔๒๘
- ๓	๒๓๒.๔๑๒	- ๖๙๗.๒๓๖	๒,๐๙๑.๗๐๘	๕๔,๐๑๕.๓๓๗๗	๒๒๕.๔๙๐๘	๔๗.๙๐๓๐
- ๑	๑๘๘.๗๗๗	- ๑๘๘.๗๗๗	๑๘๘.๗๗๗	๓๕,๖๓๖.๗๕๕๗	๒๗๒.๘๖๘๒	๗,๐๗๑.๓๒๙๙
๑	๓๑๘.๑๑๙	๓๑๘.๑๑๙	๓๑๘.๑๑๙	๑๐๑,๑๙๙.๖๙๘๑	๓๓๔.๐๕๖๘	๒๕๔.๐๑๓๔
๓	๔๐๖.๖๓๙	๑,๒๑๙.๙๑๗	๓,๖๕๙.๗๕๑	๑๖๕,๓๕๕.๒๗๖๓	๔๐๙.๐๕๖๖	๕.๘๔๔๗
๕	๕๓๒.๔๒๕	๒,๖๖๒.๑๒๕	๑๓,๓๑๐.๖๒๕	๒๘๓,๔๘๐.๖๔๐๐	๔๙๗.๘๖๗๖	๑,๑๙๔.๔๙๐๓
๗	๕๐๑.๗๙๔	๓,๕๑๒.๔๑๘	๒๔,๕๘๖.๙๒๖	๒๕๑,๗๗๗.๑๔๗๐	๖๐๐.๔๘๙๘	๙,๗๔๔.๘๐๑๙
๙	๖๔๘.๔๓๗	๕,๘๓๕.๙๓๓	๕๒,๕๒๓.๓๙๗	๔๒๐,๔๗๐.๕๔๒๙	๗๑๖.๙๒๓๒	๔,๖๙๐.๓๕๕๕
๑๑	๙๒๒.๘๒๖	๑๐,๑๕๑.๐๘๖	๑๑๑,๖๖๑.๙๔๖	๘๕๑,๖๐๗.๘๒๖๒	๘๔๗.๑๖๗๘	๕,๗๒๔.๑๖๓๒
๑๓	๑,๐๓๕.๗๖๕	๑๓,๔๖๔.๙๔๕	๑๗๕,๐๔๔.๒๘๕	๑,๐๗๒,๘๐๙.๑๓[illegible]	๙๙๑.๒๒๓๖	๑,๙๘๓.๙๓๖๓
$\sum$	๕,๘๒๐.๖๖๙	๒๗,๘๓๘.[illegible]	๔๖๐,๑๐๕.[illegible]	๓๘๗๗,๙๔๐.๗๙๕	๕,๗๙๕.๓๖๔๔	๖๒,๙๑๔.๑๙๔๓

$$\bar{Y} = \Sigma\ Y_i/14 = 415.76207\ ,\ \Sigma Y_i^2 = 3,389,624.4103$$

$$\Sigma (Y_i - \bar{Y})^2 = 1,057,927.3945 = SST$$

เปอร์เซ็นต์การกระจายของ Y = 94.05 %

F_c ที่ได้จากการคำนวณในตาราง ANOVA

$$= 5.5 \left\{ \frac{SSR}{SSE} \right\} = 5.5 \left\{ \frac{SST - SSE}{SSE} \right\}$$

$$= 5.5 \left\{ \frac{1,057,927.3945 - 62,914.1943}{62,914.1943} \right\}$$

$$= 86.9847$$

การตรวจสอบความมีนัยสำคัญของโมเดล

ค่า F จากตารางที่ d.f (2,11)

ณ ระดับนัยสำคัญ $\alpha = .01$ ค่า $F_t = 7.20$

จาก Ho : B = C = 0

ค่า $F_c > F_t$ แสดงว่าเรา Reject Ho

คือยอมรับว่าโมเดลที่อยู่ในรูปโพลีโนเมียลอันดับสองน่าจะใช้ได้กับการพยากรณ์ปริมาณการผลิตน้ำตาลทรายรวมภายในประเทศ

$$\hat{Y}_T = 301.7361 + 30.5943X + 1.7264 X^2$$

$$R^2 = 0.9405$$

Standard Error of Estimate = 72.4075

$\hat{Y}_T$ ค่าประมาณการผลิตน้ำตาลทรายรวมภายในประเทศ

X เป็นค่าของปีการผลิตน้ำตาลทรายโดยให้ปี ๒๕๐๕ มีค่า X = - 13

และปีเริ่มต้น X = 0 คือกลางปี ๒๕๑๑

ค่าของ X เพิ่มหรือลดปีละ ๒

๔.๓ ผลการคำนวณโดยใช้ Multiple Regression

ในการวิเคราะห์สมการถดถอยเชิงซ้อน เกี่ยวกับปริมาณการบริโภคและการผลิตน้ำตาลทรายภายในประเทศ โดยอาศัยเครื่องคอมพิวเตอร์ช่วยในการคำนวณตามหลัก วิธีของ Stepwise Multiple Regression ได้ผลการคำนวณดังนี้

๔.๓.๑ ผลการคำนวณโม กลปริมาณการบริโภคน้ำตาลทรายภายในประเทศ

โมเดลที่ใช้ในการทดสอบ นขั้นต้นมีลักษณะดังนี้ :-

$$Y_{R} = A + BX_1 + CX_2 + DX_3 + FX_4 + E$$

เมื่อ Y_R แทนปริมาณการบริโภคน้ำตาลทรายรวม (มีหน่วยเป็น พันตัน)

X_1 แทนจำนวนประชากร (มีหน่วยเป็น พันคน)

X_2 แทนราคาขายปลีกน้ำตาลทราย (มีหน่วยเป็น บาท)

X_3 แทนรายได้ต่อคนของประชากร (มีหน่วยเป็น บาท)

X_4 แทนปริมาณการผลิตน้ำตาลทราย (มีหน่วยเป็น พันตัน)

ผลที่คำนวณได้จากเครื่องคอมพิวเตอร์มีดังนี้ :-

<u>ขั้นที่ ๑</u>

Variable	Mean	Standard deviation
ประชากร (X_1)	๓๔๕๗๑.๘๕๕๔๗	๖๖๔๕.๙๕๓๑๓
ราคาน้ำตาล (X_2)	๔.๒๕๒๑๔	๐.๗๒๘๕๒
รายได้ (X_3)	๓๕๖๐.๖๕๒๘๒	๑๑๗๖.๘๑๑๗๗
ปริมาณการผลิต (X_4)	๔๑๕.๗๖๑๗๒	๒๘๕.๒๖๙๗๘
ปริมาณการบริโภค (Y)	๒๖๔.๘๘๕๗๔	๑๐๗.๕๖๐๙๙

ขั้นที่ ๒

เครื่องคอมพิวเตอร์จะคำนวณหา Correlation Matrix ได้ผลดังนี้

	X_1	X_2	X_3	X_4	Y
X_1	๑.๐๐๐๐๐	๐.๑๒๗๘๔	๐.๙๖๒๓๕	๐.๙๓๓๘๐	๐.๙๒๐๘๘
X_2	๐.๑๒๗๘๔	๑.๐๐๐๐๐	๐.๒๐๓๐๑	๐.๑๘๒๗๐	-๐.๑๗๔๗๓
X_3	๐.๙๖๒๓๕	๐.๒๐๓๐๑	๑.๐๐๐๐๐	๐.๙๕๖๐๒	๐.๘๕๒๖๔
X_4	๐.๙๓๓๘๐	๐.๑๘๒๗๐	๐.๙๕๖๐๒	๑.๐๐๐๐๐	๐.๘๒๔๙๗
Y	๐.๙๒๐๘๘	-๐.๑๗๔๗๓	๐.๘๕๒๖๔	๐.๘๒๔๙๗	๑.๐๐๐๐๐

ขั้นที่ ๓

จาก Correlation Matrix ที่ได้จะพบว่า Y กับ X_1 ให้ค่าสัมประสิทธิ์สหสัมพันธ์(Correlation Coefficient) สูงสุดคือ r_{YX_1} มีค่า ๐.๙๒๐๘๘ ดังนั้น X_1 จึงถูกเลือกเข้าในโมเดลเป็นอันดับแรก ตามหลักของ Stepwise จะได้สมการเป็น $\hat{Y}_k = a + b X_1$ และได้ผลลัพธ์ดังนี้

ค่าMultiple Correlation Coefficient (M.C.C.) or R	๐.๙๒๑
ค่าF_c จาก ANOVA	๖๖.๙๕๙
ค่าStandard Error of Estimate (S.E)	๔๓.๖๔๔

Variable	Regression Coefficient	Std. Error of R.C.	t_c
X_1	๐.๐๒๑๓๒	๐.๐๐๒๖๑	๘.๑๘๓
Intercept	-๔๗๒.๑๘๐๑๘		

ดังนั้นจะได้สมการเป็น $\hat{Y} = -472.18018 + 0.02132\, X_1$

(0.00261)

ขั้นที่ ๔

เป็นการเลือกค่าตัวแปรอิสระต่อไปเข้าในโมเดล และพบว่า X_2 เป็นตัวที่เหมาะสมถัดจาก X_1 ซึ่งได้เลือกไปแล้ว ได้ผลดังนี้

ค่า M.C.C. or R	.... =	๐.๙๖๗
ค่า F_c จาก ANOVA	.... =	๗๕.๐๘๗
ค่า S.E	.... =	๒๕.๘๑๗

Variable	Regression Coefficient	Std. Error of R.C.	t_c
X_1	๐.๐๒๒๒๐	๐.๐๐๑๗๙	๑๒.๓๗๐
X_2	−๔๓.๘๙๗๕๔	๑๑.๔๔๕๒๗	−๓.๘๓๕
a	−๓๑๕.๙๔๔๘๒		

ดังนั้นจะได้สมการเป็น

$$\hat{Y}_R = -315.94482 + \underset{(0.00179)}{0.0222\,X_1} - \underset{(11.44527)}{43.89754\,X_2}$$

ขั้นที่ ๕

เป็นการเลือกตัวแปรอิสระที่เหลือเข้าในโมเดลต่อไป และพบว่า X_4 เป็นตัวที่เหมาะสมถัดจากการเลือก X_1 และ X_2 ซึ่งได้เลือกไปแล้ว ได้ผลดังนี้ :-

ค่า M.C.C. or R	.... =	๐.๙๖๘
ค่า F_e จาก ANOVA	.... =	๔๕.๖๓๕
ค่า S.E	.... =	๓๐.๗๖๔

Variable	Regression Coefficient	Std. Error of R.C.	t_c
X_1	๐.๐๒๔๙๙	๐.๐๐๕๑๗	๔.๘๓๒
X_2	-๔๒.๖๖๑๓๐	๑๒.๐๐๑๔๒	-๓.๕๕๕
X_4	-๐.๐๔๙๐๔	๐.๐๘๔๙๖	-๐.๕๗๗
a	-๓๙๗.๑๖๕๕๓		

ดังนั้นจะได้สมการเป็น

$$\hat{Y}_R = -397.16553 + 0.02499\,X_1 - 42.66130\,X_2 - 0.04904\,X_4$$
$$(0.00517) \quad (12.00142) \quad (0.08496)$$

ขั้นที่ ๖

เป็นขั้นสุดท้ายในการนำตัวแปรอิสระที่เหลือได้แก่ X_3 เข้าในโมเดล (ถ้ามีตัวแปรอิสระยังคงเหลืออยู่อีกหลายตัว ก็จัดทำการเลือกต่อไปอีก) มีผลดังนี้

ค่า M.C.C. หรือ R ๐.๙๖๘

ค่า F_c จาก ANOVA ๓๓.๖๑๙

ค่า S.E ๓๒.๓๗๗

Variable	Regression Coefficient	Std. Error of R.C.	t_c
X_1	๐.๐๒๕๘๕	๐.๐๐๗๔๖	๓.๔๖๖
X_2	-๔๒.๑๓๒๖๑	๑๓.๐๑๓๙๖	-๓.๒๓๗
X_4	-๐.๐๓๘๕๔	๐.๑๐๘๙๖	-๐.๓๕๔
X_3	-๐.๐๐๖๐๔	๐.๐๓๕๘๒	-๐.๑๖๙
a	-๔๑๑.๙๙๔๘๗		

ดังนั้นจะได้สมการสุดท้าย ซึ่งมีตัวแปรอิสระครบตามโมเดลที่จะใช้ทดสอบในเบื้องต้น

$$\hat{Y}_R = -411.99487 + \underset{(0.00746)}{0.02585}\,X_1 - \underset{(13.01396)}{42.13261}\,X_2 - \underset{(0.10896)}{0.03854}\,X_4 - \underset{(0.03582)}{0.00604}\,X_3$$

<u>ผลพิจารณาเลือกโมเดลที่จะใช้เป็นตัวแทนในการพยากรณ์ปริมาณการบริโภคน้ำตาลโดย **Multiple Regression**</u>

พิจารณาตามหลักของ Step-wise Multiple Regression การเลือกตัวแปรอิสระเข้าในโมเดล จะหยุดเลือกต่อเมื่อได้นำตัวแปรอิสระนั้นเข้าไปแล้ว ทำให้ Standard Error เพิ่มขึ้น และค่าของ R มิได้เปลี่ยนแปลงหรือเริ่มลดลง

จากผลการคำนวณที่ได้จะพบว่าเมื่อนำตัวแปรอิสระตัวแรกเข้าในโมเดลได้แก่

	X_1	พบว่า	R	มีค่า 0.921	และ SE	= 43.644
เมื่อเพิ่มตัวที่ ๒ เข้าได้แก่	X_2	พบว่า	R	มีค่า 0.967	"	= 29.817
เมื่อเพิ่มตัวที่ ๓ เข้าได้แก่	X_4	พบว่า	R	มีค่า 0.968	"	= 30.764
เมื่อเพิ่มตัวที่ ๔ เข้าได้แก่	X_3	พบว่า	R	มีค่า 0.968	"	= 32.377

ดังนั้นในโมเดลที่ใช้สมควรจะมีเฉพาะ X_1 กับ X_2 เท่านั้น เพราะว่าเมื่อนำ X_4 เข้าเป็นตัวที่สาม จะทำให้ Standard Error of Estimate มีค่าเพิ่มขึ้น นอกจากนี้ยังทำให้เสียเวลาในการคำนวณมากขึ้นโดยไม่จำเป็น แต่ทั้งนี้ X_3 กับ X_4 ที่ไม่ได้นำเข้าในโมเดล มิได้หมายความว่า X_4 กับ X_3 จะไม่มีความหมายต่อปริมาณการบริโภค (Y) ก็หาไม่ การที่สามารถนำออกไปจากโมเดลได้ก็เนื่องจาก X_4 และ X_3 อาจมีความสัมพันธ์กับตัวแปรอิสระที่เลือกเข้าไปแล้วโดยมีค่าสัมประสิทธิ์สหสัมพันธ์ค่อนข้างสูงแต่ก็ต่ำกว่าตัวแปรอิสระที่เลือกเข้าไปแล้ว จาก Correlation Matrix จะพบว่า $r_{X_4X_1}$ มีค่า ๐.๙๓๓๘๐ และ $r_{X_3X_1}$ มีค่า ๐.๙๖๒๓๕ นั่นคือ X_4 และ X_3 ที่เราทำการตัดออกไป ต่างก็ มีความสัมพันธ์กับค่า X_1 ที่ค่อนข้างสูงและยังพบอีกว่า X_3 , X_4 ต่างก็มีความสัมพันธ์กับ Y

อยู่ในอัตราที่สูงแต่ว่ายังน้อยกว่า $r_{YX_1} = 0.92088$

ดังนั้นโมเดลที่จะใช้ในการพยากรณ์ก็จะลดรูปเหลือ $\hat{Y} = a + b X_1 + c X_2$

หรือ $\hat{Y}_R = -315.94482 + 0.0222 X_1 - 43.89754 X_2$ - - - - - ⊛

$$(0.00179) \qquad (11.44527)$$

โดยมีค่า $R = 0.967$, F_c จาก ANOVA มีค่า 79.087

S.E = 29.817 และ t_c ของ *b* มีค่า 12.37

t_c ของ *c* มีค่า -3.835

เมื่อนำไปทำการทดสอบโมเดลโดยใช้ F - test

ได้ค่า F_t จากตารางที่ $d.f = (k, n-k-1)$

k = จำนวนตัวแปรอิสระที่ใช้ในโมเดล = 2

n = จำนวนข้อมูล = 14

ระดับนัยสำคัญที่ใช้ (α) ก็ควรจะเป็นระดับเดียวกับในการหาโมเดลที่ใช้ในการพยากรณ์โดยวิธี Time Series เพื่อจะได้เลือกข้อสรุปในการเปรียบเทียบว่าวิธีการใดระหว่าง Time Series กับ Multiple Regression จะให้ผลในการพยากรณ์ ได้ดีกว่ากัน ค่าของ α ที่ใช้คือ ๐.๐๑

ค่า F_t ที่ได้จากตาราง F - distribution ที่ =๐.๐๑ df = (๒,๑๑)

มีค่า F_t = ๗.๒๐

ดังนั้น F_c = ๗๙.๐๘๗ > F_t = ๗.๒๐

นั่นคือเราจะยอมรับว่าโมเดล ⊛ ใช้เป็นตัวแทนในการพยากรณ์ค่าในอนาคตได้

จากนั้นจึงทำการทดสอบค่าสัมประสิทธิ์ของตัวแปรอิสระโดยใช้ t - test ที่ระดับนัยสำคัญ

α = ๐.๐๑ , d.f =(n-k-1) = ๑๑ และจากตาราง t - Distribution จะได้

t_t(๑๑) = ๓.๑๐๖ หรือ − ๓.๑๐๖

พิจารณาค่าของ β ได้ $t_c = 12.37 > t_t = 3.106$

นั่นคือเราจะปฏิเสธสมมติฐาน $H_o : \beta = 0$ หรือยอมรับว่าค่า b นั้นใช้ได้ในโมเดลเพื่อการพยากรณ์

พิจารณาค่าของ $\mathfrak{C}$ ได้ $t_c = -3.835 < t_t = -3.106$

นั่นคือเราจะปฏิเสธสมมติฐาน $H_o : \mathfrak{C} = 0$ เช่นกันโดยการยอมรับว่าค่าของ $\mathfrak{C}$ จะมีอยู่ในโมเดลได้

นั่นคือโมเดล (*) จะใช้เป็นตัวแทนของ Multiple Regression ในการใช้พยากรณ์ปริมาณการบริโภคน้ำตาลได้ โดยมีค่าประมาณการพยากรณ์ จากโมเดล

$$\hat{Y}_R = -315.94482 + 0.0222\, X_1 - 43.89754\, X_2$$ ดังนี้ :-

$$(0.00179) \qquad (11.44527)$$

ปี	Y	ค่าประมาณ $\hat{Y}$
๒๕๐๕	๑๒๕.๐๗๗	๑๐๗.๕๕๙
๒๕๐๖	๘๑.๙๕๕	๙๗.๖๓๕
๒๕๐๗	๑๑๗.๘๕๕	๑๐๘.๑๑๔
๒๕๐๘	๒๒๓.๐๖๗	๒๓๖.๖๓๕
๒๕๐๙	๑๘๔.๓๑๗	๒๓๙.๘๑๖
๒๕๑๐	๒๓๓.๗๒๖	๒๔๒.๔๔๖
๒๕๑๑	๒๓๘.๘๐๔	๒๖๐.๕๘๙
๒๕๑๒	๓๒๔.๔๕๙	๒๗๔.๘๒๖
๒๕๑๓	๓๕๗.๕๕๙	๓๒๓.๕๗๔
๒๕๑๔	๓๖๔.๖๘๐	๓๒๓.๖๗๔
๒๕๑๕	๓๑๒.๓๐๒	๓๓๖.๗๒๙
๒๕๑๖	๓๖๒.๓๑๖	๓๗๐.๑๔๔
๒๕๑๗	๓๘๒.๒๙๘	๓๙๔.๐๓๔
๒๕๑๘	๔๐๐.๐๐๐	๔๐๘.๘๓๒

เมื่อจะนำโมเดล เพื่อใช้ค่าพยากรณ์ในอนาคตจำเป็นที่จะต้องหาค่าพยากรณ์ของตัวแปรอิสระ คือ X_1 และ X_2 ก่อน นั่นคือจำเป็นที่จะต้องทราบเส้นแนวโน้มตามลำดับเวลาของ

X_1 และ X_2ให้ได้โดยใช้ระดับนัยสำคัญ = ๐.๐๑ เหมือนเดิม ถ้าการทดสอบเส้นแนวโน้มของ X_1 หรือ X_2 แล้วปรากฏผลว่าไม่ยอมรับโมเดล X_1 หรือ X_2ก็ทำการตัด X_1 หรือ X_2 ออกจากโมเดล

ดังนั้นโมเดลที่ได้ใหม่อาจอยู่ในรูป

$\hat{Y}_R = a + b X_1$ หรือ $\hat{Y}_R = a + c X_2$

และถ้าผลการทดสอบโมเดลของ X_1 , X_2 โดยพิจารณาจากเส้นแนวโน้มตามลำดับเวลา ปรากฏว่าปฏิเสธโมเดลทั้งคู่ก็ย่อมแสดงว่า โดยวิธี Multiple Regressionไม่เหมาะสม ที่จะใช้ในการพยากรณ์หาค่าในอนาคต

การคำนวณหาเส้นแนวโน้มตามลำดับเวลาของX_1 (จำนวนประชากร)

จาก Correlation Matrix จะพบว่า r_{YX_1} มีค่า ๐.๙๒๘๘ แสดงให้เห็นว่าเมื่อ X มีค่าเพิ่มขึ้นY ก็จะมีค่าเพิ่มขึ้นตามไปด้วย และเมื่อ X มีค่าลดลงจะทำให้ค่า Y มีค่าลดลงเช่นกัน หรือในทำนองเดียวกันอาจกล่าวได้ว่าถ้า Y มีค่าเพิ่มขึ้น X ก็จะมีค่าเพิ่มขึ้นในทางเดียวกัน แต่ลักษณะโมเดลของ Y (ปริมาณการบริโภค) ซึ่งทำการคำนวณโดยใช้ Time Series พบว่ามีลักษณะอยู่ในรูปโพลิโนเมี่ยลอันดับสอง ดังนั้นลักษณะโมเดลที่จะนำมาทดสอบในเบื้องต้นควรจะเป็นรูปโพลิโนเมี่ยลอันดับสองเช่นกัน คือ

$$\hat{X}_1 = a + b X + c X^2$$

$\hat{X}_1$ คือค่าประมาณของประชากร (หน่วยเป็น พันคน)

X คือปี โดยใช้ปี ๒๕๐๕ มีค่า $X = -$ ๑๓ เช่นเดียวกับกรณีของ Time Series

การคำนวณหาค่าสัมประสิทธิ์ของตัวแปรอิสระ ยังคงใช้สมการ (7) เหมือนเดิม กล่าวคือ

$$a = 0.162109 \sum X_1 - 0.001395 \sum X^2 X_1$$

$$b = 0.001099 \sum X X_1$$

$$c = 0.0000214 \sum X^2 X_1 - 0.001395 \sum X_1$$

ข้อมูลของ X_1 (จำนวนประชากร) ดูรายละเอียดในตารางที่ ๑๙ ผลจากการคำนวณจะได้

$$\Sigma X_1 = 484{,}006.0 \quad , \quad \Sigma X_1^2 = 17{,}013{,}590{,}388.0$$

$$\Sigma X X_1 = 502{,}304.0 \quad \Sigma(X_1 - \hat{X}_1)^2 = 671{,}678.45$$

$$\Sigma X^2 X_1 = 31{,}838{,}358.0 \, , \, \Sigma(X_1 - \bar{X}_1)^2 = 280{,}604{,}099.7143$$

$$b = 552.0320$$

$$c = 6.1525$$

$$a = 34047.2192$$

$$\therefore \hat{X}_1 = 34047.2192 + 552.0320\,X + 6.1525\,X^2$$

$$F_c = 2292.2102$$

$$R^2 = 0.9976 \quad , \quad S.E = 236.5865$$

ได้ค่า Sum of Square of Residuals $= \sum e^2 = 671{,}678.45$

ผลการทดสอบนัยสำคัญของโมเดล

ใช้ α = ๐.๐๕ ค่า F_c ที่คำนวณได้จาก ANOVA มีค่า = ๒๒๙๒.๒๑๐๒

ค่า F_t จากตาราง F-distribution $d.f = (k, n-k-1) = (2, 11)$

$$F_t = 7.20$$

$$\therefore F_c > F_t$$

ดังนั้นเราจะยอมรับว่าโมเดลใช้ได้

<u>การคำนวณหาเส้นแนวโน้มตามลำดับเวลาของ</u> X_2 (ราคาขายปลีก)

จาก Correlation Matrix ค่าของ r_{YX_2} มีค่าเป็น – ๐.๑๗๘๗๓ แสดงให้เห็นว่า Y กับ X_2 มีความสัมพันธ์ต่อกันน้อยมาก และเป็นชนิดส่วนทางกัน

ดังนั้นการเพิ่ม X_2 เข้าไปในโมเดลก็ย่อมจะไม่ช่วยให้ได้ผลในการพยากรณ์ดีขึ้น

โมเดลของ X_2 ที่ควรจะเป็นก็น่าจะอยู่ในรูปของโพลีโนเมี่ยลกำลังสองในทำนองเดียวกันกับปริมาณการบริโภค

$$\hat{X}_2 = a + bX + cX^2$$

$\hat{X}_2$ ค่าประมาณราคาขายปลีกน้ำตาล

จากสมการ (7) จะได้ผลการคำนวณดังนี้

$b = 0.0053$

$c = 0.0076$

$a = 3.7388$

$\sum X_2 = 59.53$, $\sum X_2^2 = 260.0297$

$\sum X X_2 = 4.83$ $\quad \sum(X_2 - \hat{X}_2)^2 = 3.9723$

$\sum X^2 X_2 = 4237.69$ $\quad \sum(X_2 - \bar{X}_2)^2 = 6.8997$

$\hat{X}_2 = 3.7388 + 0.0053X + 0.0076X^2$

$F_c = 4.0529$

$R^2 = 0.4242$; $SE = 0.5753$

ผลการทดสอบโมเดลปรากฏว่า

$F_c = 4.0529 < F_t$ จากตาราง $= 7.20$

หมายความว่าเราจะไม่ยอมรับว่าโมเดลของราคาขายปลีกน้ำตาลอยู่ในรูปกำลังสองที่ระดับความเชื่อมั่น α = ๐.๑

ขณะเดียวกันเมื่อลองพิจารณาโมเดลในลักษณะเส้นตรงจะได้

$\hat{X}_2 = a + bX$

$a = 4.2521$

$b = 0.0053$

$\sum X_2 = 59.53$; $R = 0.0609$

$$\Sigma\ X_2 X = 4.83 \quad ; \quad R^2 = 0.0037$$

$$\Sigma\ X_2^2 = 260.0297$$

พิจารณาจากค่า R = ๐.๐๖๐๘ ซึ่งต่ำมากและแสดงว่าโมเดลของ X_2 ก็จะไม่อยู่ในลักษณะของเส้นตรง

เหลือโมเดลที่จะพิจารณาอีกรูปคือ เอกโพเนนเชี่ยล

$$\hat{X}_2 = a\, b^X$$

หรือ $\log \hat{X}_2 = \log a + X \log b$

จากผลการคำนวณจะได้

$$\log a = 0.6224$$

$$\log b = 0.0009$$

$$R = 0.0994$$

$$R^2 = 0.0098$$

พิจารณาจากค่า R ที่ได้ จะเห็นว่าโมเดลของ X_2 ก็จะไม่อยู่ในลักษณะของเอกโพเนนเชี่ยลเช่นกัน

ดังนั้นจากโมเดล (※) ที่คาดว่าจะใช้เป็นตัวแทนในการหาค่าพยากรณ์ปริมาณการบริโภคในขั้นแรก ก็จะเหลือเพียง

$$\hat{Y}_R = a + b X_1$$

หรือ $$\hat{Y}_R = -472.18018 + 0.02132\ X_1$$
$$(0.00261)$$

$$R = 0.921\ ,\ F_c = 66.959\ ,\ S.E = 43.644$$

ผลการทดสอบโมเดล โดย F - test

ค่า F_t จากตาราง F - distribution ที่ $d.f = (1,12)$ ที่ $\alpha = 0.01$

มีค่า $F_t = 9.33$

∴ $F_c > F_t$ แสดงว่าโมเดลที่ใช้ทดสอบนี้ใช้ได้

ผลการทดสอบสัมประสิทธิ์ของ X_1 โดยใช้ t - test

ค่า t จากตาราง t - distribution ที่ d.f = 11 ที่ $\alpha = 0.01$

มีค่า $t_t = 3.055$

และค่า t_c ของ $B = 8.183 > t_t = 3.055$

แสดงว่าเราปฏิเสธสมมติฐาน $H_o : B = 0$

นั่นคือโมเดลจะอยู่ในรูปของ $\hat{Y}_R = a + b X_1$ เป็นโมเดลที่จะนำไปใช้ในการหาค่าพยากรณ์ในอนาคตได้

<u>ค่าของการพยากรณ์ปริมาณการบริโภคน้ำตาลโดยวิธี Multiple Regression</u>

โมเดลที่ใช้ในการพยากรณ์มีดังนี้ :-

1. $\hat{Y}_R = -472.18018 + 0.02132 X_1$

(0.00261)

$R = 0.921$, $F_c = 66.959$

$S.E = 43.644$

$\hat{Y}_R$ ค่าพยากรณ์ปริมาณการบริโภคน้ำตาลทราย (หน่วยเป็น พันตัน)

X_1 จำนวนประชากร (มีหน่วยเป็น พันคน)

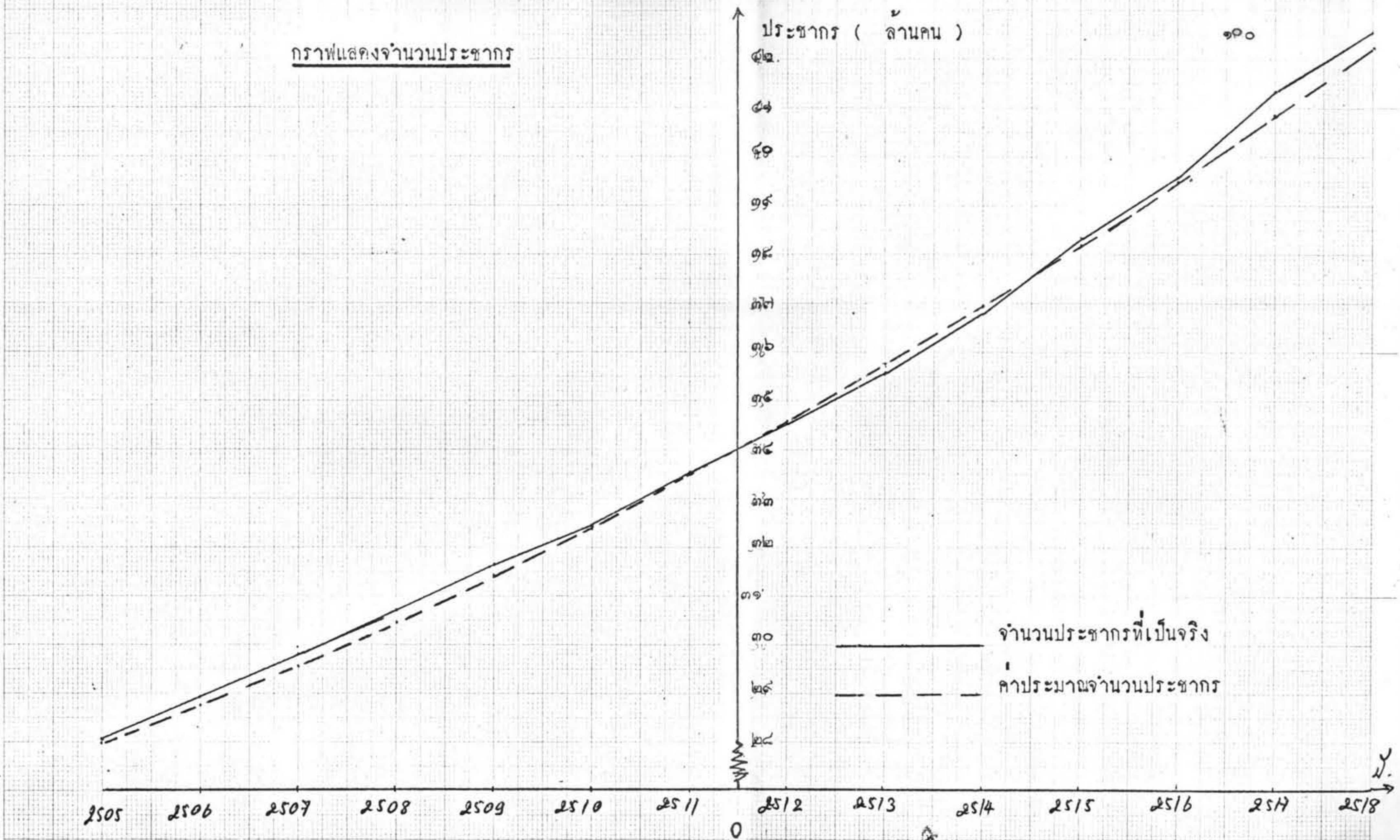
2. $\hat{X}_1 = 34047.2192 + 552.0320 X + 6.1525 X^2$

$R^2 = 0.9976$, $S.E = 236.5865$

$F_c = 2292.2102$

X คือระยะเวลาเป็นปี โดยใช้ปี ๒๕๐๕ ให้ X = - 13 , 2506 = - 11

ปีเริ่มต้นคือ X = 0 คือกลางปี ๒๕๑๑


กราฟแสดงจำนวนประชากร
ประชากร (ล้านคน)
๔๒
๔๑
๔๐
๓๙
๓๘
๓๗
๓๖
๓๕
๓๔
๓๓
๓๒
๓๑
๓๐
๒๙
๒๘
0
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
ปี.
จำนวนประชากรที่เป็นจริง
ค่าประมาณจำนวนประชากร

พยากรณ์ปี	ค่าพยากรณ์ประชากร		ค่าพยากรณ์ปริมาณการบริโภค	
๒๕๑๙ (X = 15)	๔๓๗๑๒.๐๑๑	พันคน	๔๔๙.๗๕๙๙	พันตัน
๒๕๒๐ (X = 17)	๔๕๒๐๙.๘๓๕	พันคน	๔๙๑.๖๙๓๕	พันตัน
๒๕๒๑ (X = 19)	๔๖๗๕๖.๘๗๙	พันคน	๕๒๔.๖๗๖๕	พันตัน
๒๕๒๒ (X = 21)	๔๘๓๕๓.๑๘๓	พันคน	๕๕๘.๗๐๘๙	พันตัน
๒๕๒๓ (X = 23)	๔๙๙๙๘.๖๒๗	พันคน	๕๙๓.๗๙๐๖	พันตัน

<u>ค่าของการพยากรณ์ปริมาณการบริโภคน้ำตาลทรายโดยวิธี Time Series</u>

โมเดลที่ใช้เป็นตัวแทนในการพยากรณ์ที่เหมาะสม คือ

$$\hat{Y}_T = a + b\,X$$

หรือ $$\hat{Y}_T = 265.6002 + 12.0787\,X$$

$R^2 = 0.8869$, $R = 0.9418$, $\sum e^2 = 16{,}893.2447$

$S.E = 37.5202$, $F_c = 94.308$

% การกระจาย = 88.69 %

มีค่าพยากรณ์ดังนี้ :-

ปี	ค่าพยากรณ์ปริมาณการบริโภค	
๒๕๑๙	๔๔๖.๗๘๐๗	พันตัน
๒๕๒๐	๔๗๐.๙๓๑๘	พันตัน
๒๕๒๑	๔๙๕.๐๙๕๕	พันตัน
๒๕๒๒	๕๑๙.๒๕๒๙	พันตัน
๒๕๒๓	๕๔๓.๔๑๐๓	พันตัน

๔.๓.๒ ผลการคำนวณโมเดลปริมาณการผลิตน้ำตาลทรายรวมภายในประเทศ

โมเดลที่ใช้ในการทดสอบในขั้นต้นมีลักษณะดังนี้ :-

$$Y_R = A + B\,X_1 + C\,X_2 + D\,X_3 + F\,X_4 + G\,X_5 + E$$

เมื่อ Y	แทนปริมาณการผลิตน้ำตาลทรายรวม	(หน่วยเป็น	พันตัน)
X_1	แทนจำนวนไร่อ้อย	(หน่วยเป็น	พันไร่)
X_2	แทนราคาอ้อย	(หน่วยเป็น	บาท/ตัน)
X_3	แทนราคาขายส่งน้ำตาลทราย	(หน่วยเป็น	บาท/กก)
X_4	แทนราคาน้ำตาลในตลาดต่างประเทศ	(หน่วยเป็น	บาท/กก)
X_5	แทนปริมาณอ้อยที่ผลิตได้	(หน่วยเป็น	พันตัน)

ผลที่คำนวณได้จากเครื่องคอมพิวเตอร์มีดังนี้ :-

ขั้นที่ ๑

	Variable	Mean	Standard deviation
X_1	(ไร่อ้อย)	๗๗๘.๙๖๔๘๔	๔๘๑.๕๗๙๑๐
X_2	(ราคาอ้อย)	๑๕๖.๕๙๐๕๘	๕๖.๘๘๔๕๗
X_3	(ราคาขายส่งน้ำตาล)	๓.๘๖๖๗๒	๐.๗๙๔๔๑
X_4	(ราคาน้ำตาลเมืองนอก)	๔.๕๓๕๐๐	๒.๗๙๗๘๐
X_5	(ปริมาณอ้อย)	๕,๓๘๗.๔๘๐๔๗	๓,๘๒๐.๑๑๗๖๘
Y	(ปริมาณการผลิตน้ำตาล)	๔๑๕.๗๖๑๗๒	๒๘๕.๒๖๙๗๘

ขั้นที่ ๒ จะได้ Correlation Matrix ดังนี้

	X_1	X_2	X_3	X_4	X_5	Y
X_1	๑.๐๐๐๐๐	๐.๗๓๙๖๖	๐.๒๔๗๒๖	๐.๘๓๙๔๘	๐.๙๘๘๔๖	๐.๙๘๗๓๔
X_2	๐.๗๓๙๖๖	๑.๐๐๐๐๐	๐.๔๑๖๙๐	๐.๕๕๑๐๖	๐.๖๕๖๕๓	๐.๖๕๙๒๐
X_3	๐.๒๔๗๒๖	๐.๔๑๖๙๐	๑.๐๐๐๐๐	๐.๓๐๘๖๒	๐.๑๙๕๗๔	๐.๑๕๗๔๒
X_4	๐.๘๓๙๔๘	๐.๕๕๑๐๖	๐.๓๐๘๖๒	๑.๐๐๐๐๐	๐.๘๔๔๑๓	๐.๘๓๔๙๐
X_5	๐.๙๘๘๔๖	๐.๖๕๖๕๓	๐.๑๙๕๗๔	๐.๘๔๔๑๓	๑.๐๐๐๐๐	๐.๙๙๒๕๙
Y	๐.๙๘๗๓๔	๐.๖๕๙๒๐	๐.๑๕๗๔๒	๐.๘๓๔๙๐	๐.๙๙๒๕๙	๑.๐๐๐๐๐

ขั้นที่ ๓

จาก Correlation Matrix ที่ได้จะพบว่า Y กับ X_5 ให้ค่า R สูงสุด คือ R_{YX_5} มีค่า 0.99259 ดังนั้น X_5 จึงถูกเลือกเข้าในโมเดลเป็นตัวแรก จะได้สมการเป็น $\hat{Y}_R = a + g\ X_5$

และได้ผลลัพธ์ดังนี้

ค่า Multiple Correlation Coefficient (M.C.C) หรือ R = 0.993

ค่า F_c จาก ANOVA = 800.833

ค่า Standard Error of Estimate (S.E) = 36.076

Variable	Regression Coefficient	Std.Error of R.C	t_c
X_5	0.07412	0.00262	28.300
a	16.42896		

ดังนั้นจะได้สมการเป็น $\hat{Y}_R = 16.42896 + 0.07412\ X_5$

ขั้นที่ ๔

การเลือกตัวแปรอิสระตัวที่ ๒ เข้าในโมเดล จะพบว่า X_1 เป็นตัวที่เหมาะสมที่จะถูกเลือกเข้าในโมเดลเป็นตัวถัดไปจาก X_5 ได้ผลดังนี้

ค่า M.C.C หรือ R = 0.994

ค่า F_c จาก ANOVA = 469.453

ค่า S.E = 33.372

Variable	Regression Coefficient	Std.Error of R.C	t_c
X_5	0.05019	0.01398	3.519
X_1	0.19276	0.11087	1.739
a	-4.79027		

ดังนั้นจะได้สมการใหม่เป็น

$$\hat{Y}_R = -4.79027 + 0.05019 \ X_5 + 0.19276 \ X_1$$

ขั้นที่ ๕

การเลือกตัวแปรอิสระตัวที่ ๓ เข้าในโมเดล จะพบว่า X_3 เป็นตัวที่เหมาะสม หลังจากได้เลือก X_5 และ X_1 ไปแล้ว ผลจากการคำนวณมีดังนี้

ค่า M.C.C	หรือ R	0.996
ค่า F_c	จาก ANOVA	406.858
ค่า S.E		29.321

Variable	Regression Coefficient	Std.Error of R.C	t_c
X_5	0.04268	0.01281	3.332
X_1	0.26072	0.10283	2.535
X_3	-23.92126	11.60274	-2.062
a	75.23653		

จะได้สมการใหม่คือ

$$\hat{Y}_R = 75.23653 + 0.04268 \ X_5 + 0.26072 \ X_1 - 23.92126 \ X_3$$

ขั้นที่ ๖

การเลือกตัวแปรอิสระตัวที่ ๔ เข้าในโมเดล จะพบว่า X_2 เป็นตัวที่เหมาะสม ผลที่ได้จากการคำนวณมีดังนี้

ค่า M.C.C หรือ R = 0.996

ค่า F_c จาก ANOVA = 307.931

ค่า S.E = 29.200

Variable	Regression Coefficient	Std.Error of R.C	t_c
X_5	0.03335	0.01559	2.139
X_1	0.35762	0.13841	2.584
X_3	-20.90771	11.91246	-1.755
X_2	-0.34976	0.33607	-1.041
a	93.12996		

$$\therefore \hat{Y}_R = 93.12996 + 0.03335\, X_5 + 0.35762\, X_1 - 20.90771\, X_3 - 0.34976\, X_2$$

ขั้นที่ ๗

คงเหลือตัวแปรอิสระตัวสุดท้ายคือ X_4 เมื่อนำเข้าไปในโมเดล จะได้ผลจากการคำนวณดังนี้

ค่า M.C.C หรือ R = 0.996

ค่า F_c จาก ANOVA = 218.973

ค่า S.E = 30.972

Variable	Regression Coefficient	Std. Error of R.C	t_c
X_5	0.03337	0.01665	2.004
X_1	0.035777	0.14784	2.420
X_3	-20.87416	13.21888	-1.579
X_2	-0.35030	0.36198	-0.968
X_4	-0.05220	6.04462	-0.009
a	93.11557		

ดังนั้นจะได้สมการสุดท้าย ซึ่งมีตัวแปรอิสระครบตามโมเดลที่จะใช้ทดสอบในเบื้องต้น

$$\hat{Y}_R = 93.11557 + 0.03337\,X_5 + 0.35777\,X_1 - 20.87416\,X_3 - 0.35030\,X_2 - 0.05220\,X_4$$

$$\sum e^2 = 7673.5101$$

<u>ผลการพิจารณาเลือกโมเดลที่จะใช้เป็นตัวแทนในการพยากรณ์ปริมาณการผลิตน้ำตาลทรายโดย Multiple Regression</u>

การพิจารณาจะเป็นไปในทำนองเดียวกันกับการพยากรณ์ปริมาณการบริโภคมีผลที่ได้จากการคำนวณเมื่อเลือกตัวแปรอิสระเข้าในโมเดลดังนี้ :-

ตัวแปรอิสระตัวแรกคือ X_5 พบว่าให้ค่า R มีค่า ๐.๙๙๒๕๙ และSE=๓๖.๐๓๖
เมื่อเพิ่มตัวแปรอิสระตัวที่สอง คือX_1 พบว่าให้ค่า มีค่า ๐.๙๙๔ =๓๓.๓๗๒
เมื่อเพิ่มตัวแปรอิสระตัวที่สาม คือ X_3พบว่าให้ค่า มีค่า ๐.๙๙๖ = ๒๙.๗๒๑
เมื่อเพิ่มตัวแปรอิสระตัวที่สี่ คือ X_2 พบว่าให้ค่า มีค่า ๐.๙๙๖ = ๒๙.๒๐๐
เมื่อเพิ่มตัวแปรอิสระตัวที่ห้า คือ X_4พบว่าให้ค่า มีค่า ๐.๙๙๖ = ๓๐.๙๗๒

ถ้าพิจารณาจากS.E จะพบว่าในโมเดลที่มี X_5, X_1, X_3และ X_2 ด้วย จะให้ค่าต่ำสุด S.E ๒๙.๒๐๐ แต่ถ้าเพิ่ม X_4เข้าในโมเดลด้วยค่า S.E จะมีค่าสูงขึ้น แสดงว่า X_4 ควรจะเอาออกจากโมเดลได้ นอกจากเหตุผลเกี่ยวกับ S.E เพิ่ม

แล้วยังปรากฏว่า X_4 มีความสัมพันธ์กับค่า X_1 โดยให้ค่า $r_{X_4X_1}$ สูงถึง ๐.๘๓๙๘๘ (ดูจาก Correlation Matrix) ซึ่ง X_1 ได้ถูกเลือกเข้าในโมเดลไปก่อนหน้า นี้แล้ว ถ้าพิจารณาจากค่า R จะเห็นว่าในโมเดลที่มีค่า X_5, X_1 และ X_3 จะให้ค่า R สูงสุดคือ ๐.๙๙๖ ต่อจากนั้นค่า R ก็มิได้สูงขึ้นอีก แสดงว่าการเพิ่มค่า X_2 เข้าไปก็ไม่ช่วยให้การพยากรณ์ได้ดีขึ้น ถึงแม้ว่า ค่าของ S.E จะลดลงบ้างก็ตาม คือ S.E จะลดลงประมาณ ๐.๑๒๑ ดังนั้นจึงควรพิจารณาในการทดสอบโมเดล ดูเฉพาะตัวแปรอิสระ ๓ ตัวแรกที่เข้าไปคือ X_5, X_1 และ X_3 ถ้าผลการทดสอบโมเดลปรากฏว่ายอมรับโมเดลดังกล่าว และผลการทดสอบสัมประสิทธิ์ของตัวแปรอิสระเมื่อเป็นที่ยอมรับเช่นกัน ก็ควรจะลองพิจารณาทดสอบกรณีที่เพิ่ม X_2 เข้าไปแล้วทำการทดสอบโดยทำนองเดียวกัน

โมเดลที่น่าจะใช้ในการนำไปทดสอบเพื่อการพยากรณ์ในขั้นแรกควรจะอยู่ในรูป

$$\hat{Y}_R = a + g\,X_5 + b\,X_1 + d\,X_3$$

หรือ $\hat{Y}_R = 75.23653 + 0.04268\,X_5 + 0.26072\,X_1 - 23.92126\,X_3$ (**)

มีค่า $R = 0.996$, F_c จาก ANOVA มีค่า 406.858, S.E = 29.321

และ t_c ของ $\mathcal{G}$ มีค่า 3.332

B มีค่า 2.535

D มีค่า -2.062

ค่า F จากตารางที่ d.f (k, n-k-1) = (3,10)

ได้ $F_{t(3,10)}$ ที่ α = .01 มีค่า = 6.55

$F_c > F_t$ นั่นคือเราจะยอมรับว่าโมเดล (**) ใช้ได้อย่างมีนัยสำคัญ ขั้นต่อไปจะทำการทดสอบค่าสัมประสิทธิ์ของตัวแปรอิสระ โดยการเปิดตาราง t ที่ α = 0.101, d.f = n-k-1

จากตารางพบว่า t_{10} ที่ α =.01	มีค่า 3.169	หรือ- 3.169
= .05	2.228	หรือ-2.228
=.10	1.812	หรือ-1.812

แต่เมื่อต้องการนำโมเดลที่ได้ไปเปรียบเทียบกับโมเดลที่ได้จาก Time Series ก็ควรจะใช้ระดับ α ที่ ๐.๐๑ เหมือนกัน ดังนั้นจะได้ ผลการพิจารณาค่า t ดังนี้ :-

สำหรับค่า δ ได้ $t_c = 8.332 > t_t = 3.169$

ดังนั้นเราจะยอมรับว่าค่าของ X_4 มีอิทธิพลต่อโมเดล

ค่า β ได้ $t_c = 2.535 < t_t = 3.169$

ดังนั้นเราจะไม่ยอมรับ X_1 เมื่อ $\alpha = .01$

และค่า D ได้ $t_c = -2.062 > -3.169 = t_t$

ดังนั้นเราจะไม่ยอมรับ X_3 เข้าในโมเดล

เมื่อผลการทดสอบสัมประสิทธิ์ของตัวแปรอิสระในโมเดล (**) ปรากฏว่าค่าของ X_1 และ X_3 ก็ไม่เหมาะสมที่จะอยู่ในโมเดล เราก็ควรพิจารณาโดยการตัดตัวแปรอิสระ ตัวสุดท้ายที่เข้าในโมเดล (**) ซึ่งได้แก่ X_3 ออกไปก่อน แล้วทำการทดสอบโมเดลใหม่ซึ่ง จะลดรูปเหลือ $\hat{Y}_R = a + g X_5 + b X_1$

หรือ $\hat{Y}_R = -4.79027 + 0.05019 X_5 + 0.19276 X_1$

มีค่า $R = 0.994$, F_c ANOVA 469.453 , S.E = 33.372

ค่า t_c ของ δ มีค่า $= 3.591$

$\beta = 1.739$

พิจารณาค่า F_t ที่ d.f (2,11), $\alpha = .01$ ได้ $F_t = 7.20$

t_t ที่ d.f = 11 , $\alpha = .01$ ได้ $t_t = 3.106$

ควรจะเลือกพิจารณาค่าของสัมประสิทธิ์ของ X_1 คือ β ก่อน เพราะ X_1 กับ X_3 ในโมเดล (**) ไม่ยอมรับ แสดงว่าค่าของ β ก็น่าจะเป็นศูนย์จริง ค่า t_c ของ β มีค่า = ๑.๗๓๙ < t_t จากตาราง = ๓.๑๐๖ แสดงว่าเราจะ Accept Ho : $\beta = 0$ เป็นจริง นั่นคือเราจะไม่ยอมรับให้ X_1 อยู่ในโมเดล เพื่อใช้ในการพยากรณ์ ในที่สุดโมเดลของ ปริมาณการผลิตน้ำตาลทรายในประเทศก็จะเหลือตัวแปรอิสระเพียงตัวเดียวคือ X_5 ที่จะทำการทดสอบต่อไป

$$\hat{Y}_R = a + g\,X_5$$

หรือ $\hat{Y}_R = 16.42896 + 0.07412\,X_5$

มีค่า $R = 0.993$, $F_c = 800.88$, $S.E = 36.076$

$t_c = 28.300$

เปิดค่า $F_t = 9.33$ จาก $d.f\ (1,12)$, $\alpha = .01$

และ $t_t = 3.055$ $d.f = 12$, $\alpha = .01$

เปรียบเทียบค่า $F_c = 800.883 > F_t = 9.33$

แสดงว่าเรายอมรับโมเดลที่มีเฉพาะ X_5

และเมื่อทดสอบสัมประสิทธิ์ของ X_5 คือ δ

จะได้ $t_c = 28.3 > t_t = 3.055$

แสดงว่าเรา Reject $H_0 : \delta = 0$

นั่นคือยอมรับว่า $\delta \neq 0$ เป็นจริง

ดังนั้นโมเดลที่ใช้เป็นตัวแทนเพื่อการพยากรณ์ปริมาณการผลิตน้ำตาลก็จะขึ้นอยู่กับ X_5 (คือปริมาณอ้อยที่ผลิตได้) อย่างเดียวตามวิธี Multiple Regression

$$\hat{Y}_R = 16.42896 + 0.07412\,X_5$$

ได้ $R = 0.993$ และ $S.E = 36.076$

โดยมีค่าประมาณดังนี้

ปี	Y	$\hat{Y}_R$
๒๕๐๕	๑๕๑.๓๔๓	๑๗๙.๑๘๕
๒๕๐๖	๑๒๕.๐๓๑	๑๔๖.๐๒๗
๒๕๐๗	๑๖๗.๙๗๒	๑๙๓.๓๖๗
๒๕๐๘	๓๑๙.๙๗๖	๓๐๖.๘๔๔
๒๕๐๙	๒๖๙.๑๖๘	๒๔๖.๑๑๓

ปี	Y	$\hat{Y}_R$
๒๕๑๐	๒๓๒.๔๑๒	๒๐๔.๒๘๘
๒๕๑๑	๑๘๘.๗๗๗	๑๙๒.๗๙๒
๒๕๑๒	๓๑๘.๑๑๘	๓๙๑.๔๘๗
๒๕๑๓	๔๐๖.๒๓๙	๓๘๔.๒๐๘
๒๕๑๔	๕๓๒.๘๒๙	๕๐๔.๕๗๓
๒๕๑๕	๕๐๑.๗๗๔	๔๕๕.๒๓๒
๒๕๑๖	๖๘๘.๔๓๗	๗๑๐.๑๕๑
๒๕๑๗	๙๒๒.๘๒๖	๘๕๓.๓๓๖
๒๕๑๘	๑๐๓๕.๗๖๕	๘๘๘.๑๐๗

$\sum e^2$ = ๑๕,๕๒๓.๕๓๑

เมื่อนำโมเดล (* *) ไปใช้ในการพยากรณ์ค่าในอนาคต จำเป็นที่จะต้องทราบค่าแนวโน้มตามลำดับเวลาของ X_5 (ปริมาณอ้อยที่ผลิตได้) เสียก่อนว่ามีลักษณะเส้นแนวโน้มเป็นเช่นใด แล้วจึงจะบอกค่าพยากรณ์ในอนาคตของปริมาณการผลิตน้ำตาลทรายได้

<u>การคำนวณหาเส้นแนวโน้มตามลำดับเวลาของ</u> X_5 <u>(ปริมาณผลผลิตอ้อย)</u>

จาก Cerrelation Matrix พบว่า r_{YX_5} มีค่า ๐.๙๙๒๘๙ (เมื่อ Y คือปริมาณการผลิตน้ำตาลทรายภายในประเทศ) แสดงว่า Y กับ X_5 มีความสัมพันธ์ต่อกันสูงมากในลักษณะเส้นตรงและการพยากรณ์จากค่า Y (ปริมาณการผลิตน้ำตาล) เราทราบว่าโมเดลที่ใช้เป็นตัวแทนอยู่ในลักษณะของโพลี่โนเมี่ยลอันดับสอง ดังนั้นจึงเป็นแนวทางที่ชี้ให้เห็นว่าโมเดลที่จะบอกลักษณะของ X_5 ก็ควรจะอยู่ในลักษณะเดียวกัน ประกอบกับการพิจารณาดู จากลักษณะกราฟจาก Scakter Diagram ก็จะเห็นว่าลักษณะของ Trend ของ X_5 ก็ควรจะเป็นเส้นโค้ง ฉะนั้นจึงสมควรทำการทดสอบโมเดลของ X_5 เริ่มด้วยในลักษณะของโพลี่โนเมี่ยลอันดับสอง

$$\hat{X}_5 = a + bX + cX^2$$

เมื่อ X_5 คือปริมาณผลผลิตอ้อย

X คือปีที่ทำการผลิตให้ปี ๒๕๐๕ มีค่า =13 เช่นเดียวกับกรณีของ Time Series ที่กล่าวมาแล้ว

วิธีการคำนวณในทำนองเดียวกันโดยอาศัยสมการ (7)

$$a = 0.162109 \Sigma X_5 - 0.001395 \Sigma X^2 X_5$$

$$b = 0.001099 \Sigma X X_5$$

$$c = 0.0000214 \Sigma X^2 X_5 - 0.001395 \Sigma X_5$$

ข้อมูลของ X_5 (ผลผลิตอ้อย) ดูรายละเอียดในตารางที่ 3 ผลจาก การคำนวณ

ได้ $\Sigma X_5 = 75,424.804$ $\Sigma X_5^2 = 596,063,148.3259$

$\Sigma X X_5 = 367,930.010$ $\Sigma (X_5 - \hat{X}_5)^2 = 11,042,043.3449$

$\Sigma X^2 X_5 = 6,083,395.572$ $\Sigma (X_5 - \bar{X}_5)^2 = 189,713,072.7226$

$$b = 404.3550$$

$$c = 24.9670$$

$$a = 3,740.7027$$

$$\hat{X}_5 = 3740.7027 + 404.355 X + 24.967 X^2 \text{ - - - - } (A*)$$

$$F_c = 110.5016$$

$$R^2 = 0.9525$$

$$S.E = 865.7777$$

และได้ค่าประมาณผลผลิตอ้อยดังนี้

ปี	X_5	$\hat{X}_5$
๒๕๐๕	๒๑๙๕.๘๕๓	๒๗๐๓.๕๑๐๗
๒๕๐๖	๑๖๙๔.๕๓๓	๒๓๑๓.๘๐๔๗
๒๕๐๗	๒๓๘๗.๑๘๕	๒๑๒๓.๘๓๔๗
๒๕๐๘	๓๙๑๒.๗๘๘	๒๑๓๓.๖๐๐๗
๒๕๐๙	๓๐๔๔.๘๕๐	๒๓๔๓.๑๐๒๗
๒๕๑๐	๒๕๓๔.๖๖๐	๒๗๕๒.๓๔๐๗
๒๕๑๑	๒๓๗๙.๔๓๐	๓๓๖๑.๓๑๔๗
๒๕๑๒	๔๓๙๙.๐๖๗	๔๑๗๐.๐๒๔๗
๒๕๑๓	๕๑๐๒.๒๖๘	๕๑๗๘.๔๗๐๗
๒๕๑๔	๖๕๘๕.๘๕๙	๖๓๘๖.๖๕๒๗
๒๕๑๕	๕๙๒๕.๕๖๖	๗๗๙๔.๕๗๐๗
๒๕๑๖	๙๕๑๒.๗๙๔	๙๔๐๒.๖๒๔๗
๒๕๑๗	๑๒๖๔๐.๔๑๗	๑๑๒๐๙.๖๑๔๗
๒๕๑๘	๑๓๑๐๙.๕๓๔	๑๗๒๑๖.๗๕๐๗

$\sum e^2$ = ๑๑,๐๕๒,๐๔๙.๓๔๔๙

ผลการทดสอบนัยสำคัญของโมเดล

โดยใช้ $\alpha = .01$

ค่า F_c ที่คำนวณได้จาก ANOVA มีค่า $= 110.5016$

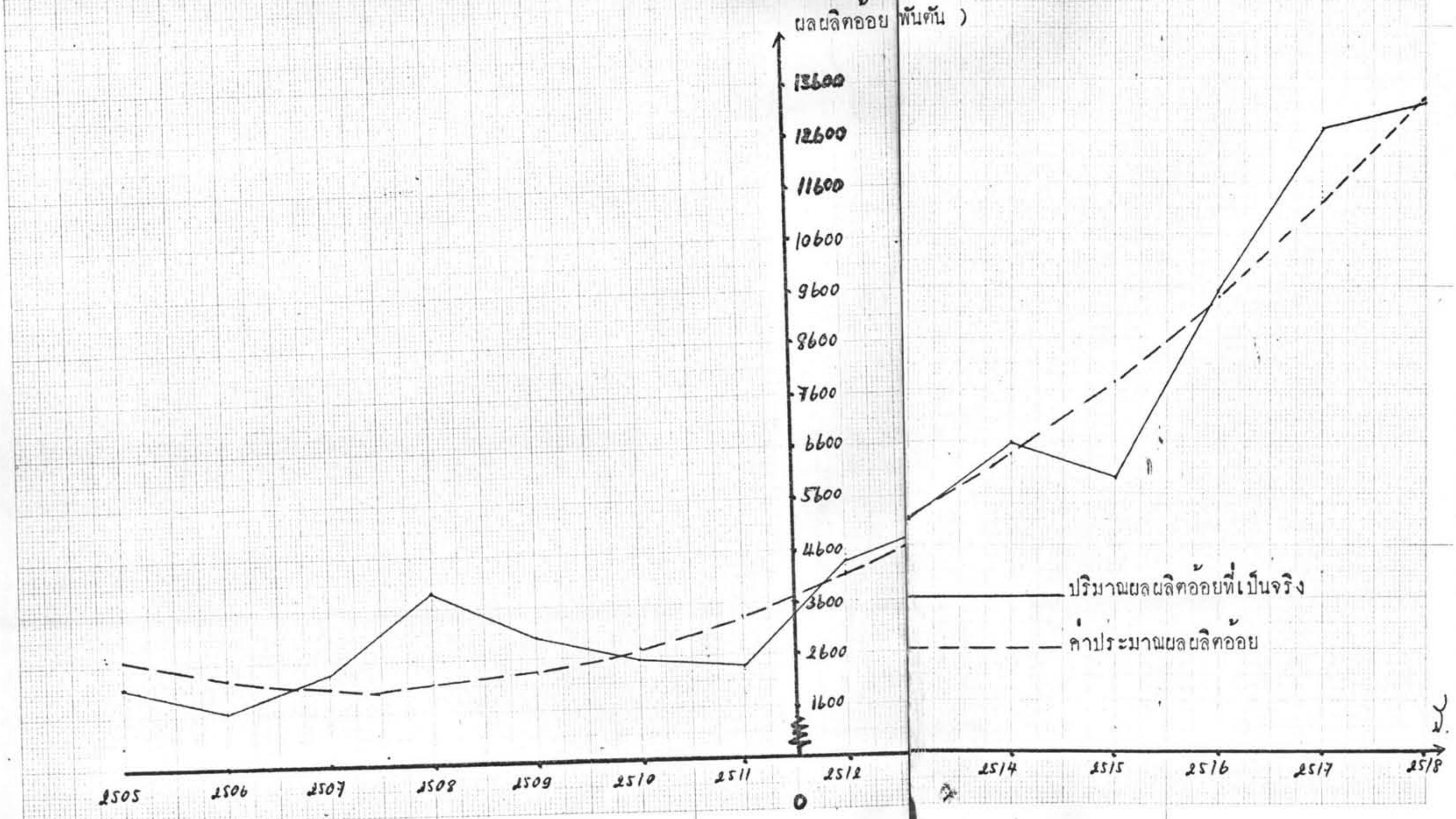
ค่า F_t จากตาราง F-distribution d.f.=(k,n-k-1) = (2,11)

ได้ค่า $F_t = 7.20$

$F_c > F_t$

นั่นคือเราจะยอมรับ Model ว่าใช้ได้

กราฟแสดงปริมาณผลผลิตอ้อย



<u>ค่าของการพยากรณ์ผลผลิตน้ำตาลโดยวิธี Multiple Regression</u>

โมเดลที่ใช้ในการพยากรณ์มีดังนี้ :-

1. $\hat{Y}_R = 16.4289 + 0.07412 \; X_5$

$R = 0.993 \qquad S.E = 36.076 \; , \; \sum e^2 = 15,563.531$

$\hat{Y}_R$ คือค่าพยากรณ์ผลผลิตน้ำตาล ,(การหาค่า **Var** ($\hat{Y}$) มีดังผนวก ค)

X_5 คือผลผลิตอ้อยจริง มีหน่วย : พันตัน

2. $\hat{X}_5 = 3740.7027 + 404.355 \; X + 24.967 \; X^2$

$R^2 = 0.9525 \quad , \; S.E = 865.7777 \; ,$

$\sum e^2 = 11,042,049.3449$

$\hat{X}_5$ คือค่าพยาพรณ์ผลผลิตอ้อย

X คือการระยะเวลาเป็นปี โดยใช้ปี ๒๕๐๕ ให้ X = − ๑๓ , ๒๕๐๖ = − ๑๑

พยากรณ์ปี	ค่าพยากรณ์ผลผลิตอ้อย	ค่าพยากรณ์ผลผลิตน้ำตาล
๒๕๑๙ (X = 15)	๑๕,๔๒๓.๖๐๒	๑๑๕๙.๖๒๖๒
๒๕๒๐ (X = 17)	๑๗,๘๓๐.๒๐๐	๑๓๓๘.๐๐๓๓
๒๕๒๑ (X = 19)	๒๐,๔๓๖.๕๓๔	๑๕๓๑.๑๘๔๘
๒๕๒๒ (X = 21)	๒๓,๒๔๒.๖๐๔	๑๗๓๙.๑๗๐๗
๒๕๒๓ (X =23)	๒๖,๒๔๘.๔๑๐	๑๙๖๑.๙๖๑๐

<u>ค่าของการพยากรณ์ผลผลิตน้ำตาลโดยวิธี Time Series</u>

ได้โมเดลที่ใช้เป็นตัวแทนในการพยากรณ์ดังนี้

$$\hat{Y}_T = a + b\,X + c\,X^2$$

หรือ $$\hat{Y}_T = 301.7361 + 30.5943\,X + 1.7264\,X^2$$

$r^2 = 0.9045$, $S.E = 72.4075$, $F_c = 86.9847$

% การกระจาย = 90.45 % .

$$\sum e^2 = 86{,}919.3434$$

มีค่าสำหรับการพยากรณ์ตั้งแต่ปี ๒๕๑๙ ดังนี้

พยากรณ์ปี		ค่าพยากรณ์ผลผลิตน้ำตาล
๒๕๑๙	(X=15)	๑,๑๔๘.๐๙๐๖
๒๕๒๐	(X=17)	๑,๓๒๐.๗๖๘๘
๒๕๒๑	(X=19)	๑,๕๐๖.๒๕๘๒
๒๕๒๒	(X=21)	๑,๗๐๕.๕๕๘๘
๒๕๒๓	(X=23)	๑,๙๑๘.๖๗๐๖

<u>เปรียบเทียบค่าพยากรณ์ผลผลิตน้ำตาลทรายในประเทศไทย</u>

ปี	โดย Multiple Regression	โดย Time Series	ผลต่าง
๒๕๑๙	๑,๑๕๘.๖๒๖๒	๑,๑๔๘.๐๙๐๖	๑๐.๕๓๕๖
๒๕๒๐	๑,๓๓๘.๐๐๓๓	๑,๓๒๐.๗๖๘๘	๑๗.๒๓๔๕
๒๕๒๑	๑,๕๓๑.๑๘๔๘	๑,๕๐๖.๒๕๘๒	๒๔.๙๒๖๖
๒๕๒๒	๑,๗๓๙.๑๗๐๗	๑,๗๐๕.๕๕๘๘	๓๓.๖๑๑๙
๒๕๒๓	๑,๙๖๑.๙๖๑๐	๑,๙๑๘.๖๗๐๖	๔๓.๒๙๐๔