บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้ผู้วิจัย ได้แยกการทบทวนวรรณกรรมออกเป็น 2 ส่วน ส่วนแรกเป็นส่วนของเป็นส่วนผลงานวิจัยที่เกี่ยวข้องและส่วนที่สอง ทฤษฎีต่างๆ ที่เกี่ยวข้อง

2.1 งานวิจัยที่เกี่ยวข้อง

ลองเลย์ (Longley, 1967) ศึกษาโปรแกรมสำเร็จรูปที่ใช้ในการคำนวณสถิติ 4 โปรแกรมได้แก่ โปรแกรม NIPD, โปรแกรม ORTHO, โปรแกรม Dartmouth Time–Sharing และโปรแกรม IBM Statistical Subroutine โดยเปรียบเทียบผลจากการวิเคราะห์การถดถอย และเปรียบเทียบกับผลที่ได้จากการคำนวณโดยใช้เครื่องคำนวณ ซึ่งข้อมูลที่ใช้ในการคำนวณนำมาจากผลผลิตมวลรวมประชาชาติ ผลการศึกษาพบว่าโปรแกรม ORTHO เป็นโปรแกรมซึ่งให้ผลลัพธ์การวิเคราะห์ที่มีความแม่นมากที่สุด อีกทั้งลองเลย์ได้เสนอให้มีการพัฒนาโปรแกรมสำเร็จรูปทางสถิติทางด้านอัลกอริธึม (Algorithm) เพื่อให้ได้ผลการวิเคราะห์ที่มีความถูกต้องแม่นพอเพียงสำหรับใช้ประโยชน์ต่อไป

แวมเพลอร์ (Wampler, 1970) ศึกษาโปรแกรมสำเร็จรูปที่ใช้ในการคำนวณสถิติ 18 โปรแกรม โดยเปรียบเทียบผลจากการวิเคราะห์การถดถอยด้วยวิธีกำลังสองน้อยที่สุด (Least Square Method) ซึ่งทดสอบตามทฤษฎี 3 ทฤษฎี คือ Orthogonal House – Holder Transformation, Classical Gram – Schmidt Orthogonalization และ Elimination Algorithms ผลการศึกษาพบว่าทฤษฎีของ Orthogonal House – Holder Transformation และ Classical Gram – Schmidt Orthogonalization จะให้ผลของการวิเคราะห์ที่มีความแม่นสูงกว่าที่ใช้ทฤษฎีของ Elimination Algorithms

ลาเชนบรัช (Lachenbrush, 1983) ศึกษาโปรแกรมสำเร็จรูปที่ใช้ในการคำนวณสถิติ 5 โปรแกรม โดยเปรียบเทียบการวิเคราะห์สถิติตัวแปรเดียว (Univariate Statistics) ได้แก่ โปรแกรม DAISY รุ่น 1.2.2 และ 2.0, โปรแกรม AIDA, โปรแกรม A – STAT รุ่น 79.6 และ 83.1, โปรแกรม GLIM และโปรแกรม HSD ผลการศึกษาพบว่า เมื่อข้อมูลมีขนาด 7 – 8 หลัก (Digit) โปรแกรม A – STAT รุ่น 83.1 และโปรแกรม GLIM คำนวณค่าส่วนเบี่ยงเบนมาตรฐานได้ ในขณะ

ที่ โปรแกรม DAISY รุ่น 1.2.2, โปรแกรม HSD และโปรแกรม A – STAT รุ่น 79.6 ไม่สามารถคำนวณค่าส่วนเบี่ยงเบนมาตรฐานได้ กรณีที่ข้อมูลมีขนาด 5 หลัก (Digit) โปรแกรม HSD ไม่สามารถคำนวณค่าเฉลี่ยได้ถูกต้อง นอกจากนี้เมื่อใช้ข้อมูลของลองเลย์ (Longley) สำหรับทดสอบผลการวิเคราะห์ถดถอยด้วยโปรแกรมข้างต้น พบว่าโปรแกรม AIDA และ A – STAT รุ่น 79.6 ให้ผลวิเคราะห์ที่มีความแม่นน้อยที่สุด

พีสและคณะ (Pease et al, 1984) ศึกษาโปรแกรมสำเร็จรูปที่ใช้ในการคำนวณสถิติ 6 โปรแกรม ได้แก่ โปรแกรม ABSTAT, โปรแกรม AIDA, โปรแกรม A–STAT, โปรแกรม MICROSTAT, โปรแกรม MICROSTAT NUMBER CRUNCHER และ โปรแกรม SPS โดยศึกษาจากการสอบถามไปยังผู้ใช้โปรแกรม เพื่อจัดลำดับความสามารถของแต่ละโปรแกรมพบว่า โปรแกรม ABSTAT มีลักษณะโดยทั่วไปสมบูรณ์ ในขณะที่โปรแกรม MICROSTAT แสดงผลลัพธ์การวิเคราะห์ได้ดีกว่า และสำหรับโปรแกรม ABSTAT สามารถวิเคราะห์สถิติพื้นฐานได้สมบูรณ์กว่าโปรแกรมอื่นๆ และโปรแกรม A-STAT เป็นโปรแกรมที่ผู้ใช้โดยมากไม่ใคร่พอใจต่อลักษณะรายละเอียดต่างๆมากที่สุด นอกจากนี้พีสได้ทำการเปรียบเทียบโปรแกรมทั้งหมดกับข้อมูลสำหรับการทดสอบของแวมเพลอร์ (Wampler) และข้อมูลของลาเพจ (Lapage) พบว่าในกรณีที่ใช้ข้อมูลแวมเพลอร์ (Wampler) ผลการวิเคราะห์จากโปรแกรม AIDA มีความคลาดเคลื่อนสูงสำหรับการวิเคราะห์สถิติพื้นฐาน และเมื่อใช้ข้อมูลของลาเพจ (Lapage) พบว่า ทุกโปรแกรมสามารถวิเคราะห์ผลได้ถูกต้อง

ชัชพงศ์ (2531) ศึกษาโปรแกรมสำเร็จรูปที่ใช้ในการคำนวณสถิติสำหรับไมโครคอมพิวเตอร์ 4 โปรแกรม คือโปรแกรม SPSS/PC+, โปรแกรม SAS for PC, โปรแกรม SYSTAT และ โปรแกรม Statpro โดยเปรียบเทียบความสามารถด้านการวิเคราะห์สถิติของแต่ละโปรแกรม เปรียบเทียบความแม่นของผลการวิเคราะห์สถิติของแต่ละโปรแกรม และเปรียบเทียบเวลาที่ใช้สำหรับการวิเคราะห์ของแต่ละโปรแกรม ในการวิเคราะห์ความสามารถด้านการวิเคราะห์สถิติของแต่ละโปรแกรม จะแจกแจงความสามารถต่างๆ ออกมาแล้วดูว่าโปรแกรมสามารถวิเคราะห์ได้หรือไม่ เช่น ค่าเฉลี่ย ความแปรปรวน ผลรวมกำลังสอง เป็นต้น แล้วทำการนับความสามารถ 1 ครั้งเป็น 1 เครื่องหมาย พร้อมจัดอันดับของแต่ละโปรแกรมตามหน่วยความสามารถที่ได้ และในการวิเคราะห์ความแม่นของแต่ละโปรแกรมใช้ข้อมูลในการวิเคราะห์ประกอบด้วย ข้อมูลตัวอย่างทั่วไป เป็นข้อมูลที่ได้จากการวิจัยทั่วไป และข้อมูลเพื่อการทดสอบโดยเฉพาะ คือข้อมูลแวมเพลอร์ (Wampler) ข้อมูลลองเลย์ (Longley) และข้อมูลลาเพจ (Lapage) โดยพิจารณาผลลัพธ์ย่อยของโปรแกรมนั้นกับผลลัพธ์เปรียบเทียบที่ตรงกันในลักษณะที่

ยึดหลักซ้ายสุดของผลลัพธ์นั้น เป็นการเริ่มให้หน่วยความแม่น โดยถ้าเลขในหลักของผลลัพธ์เปรียบเทียบตรงกับผลลัพธ์ย่อยนั้นๆแล้ว ผลลัพธ์ย่อยดังกล่าวจะได้หน่วยความแม่น 1 หน่วย และเทียบในหลักต่อมาเรื่อยๆ ตามเท่าที่เลขหลักต่อมายังตรงกัน โดยจะหยุดให้หน่วยความแม่น เมื่อมีการไม่ตรงกันในหลักใดเป็นหลักแรกของผลลัพธ์ทั้งคู่ และจำนวนหน่วยความแม่น สำหรับการวิเคราะห์สถิติแต่ละประเภท ได้จากผลรวมหน่วยความแม่นของทุกผลลัพธ์ย่อย เมื่อได้ความแม่นแล้วจึงทำการทดสอบสมมติฐาน และในการทดสอบนั้นเนื่องจากข้อมูลที่ได้นั้นอยู่ในมาตราเรียงลำดับ (Ordinal Scale) จึงใช้การทดสอบที่ไม่อิงพารามิเตอร์ (Nonparametric Test) คือการวิเคราะห์ความแปรปรวนตามวิธีของฟรีดแมน (The Friendman Two-Way Analysis of Varience) และถ้าปฏิเสธสมมติฐาน H_0 จึงทำการทดสอบรายคู่ โดยใช้การทดสอบเครื่องหมาย (Sign Test) พบว่าโปรแกรม SPSS/PC+ ให้ความแม่นสูงกว่าโปรแกรมอื่นๆ กรณีที่จำนวนตัวอย่างมีขนาดปานกลางถึงมาก ในขณะที่โปรแกรม SAS for PC, โปรแกรม SYSTAT และโปรแกรม Statpro ให้ความแม่นไม่แตกต่างกัน แต่เมื่อพิจารณากรณีที่ตัวอย่างมีขนาดเล็ก พบว่าโปรแกรมทั้ง 4 ให้ความแม่นไม่แตกต่างกันในทุกระดับของข้อมูล แต่กรณีการวิเคราะห์การถดถอยของข้อมูลที่มีปัญหาพบว่าโปรแกรม SAS for PC จะมีประสิทธิภาพดีกว่าโปรแกรมอื่นๆ เมื่อข้อมูลที่ใช้สำหรับการวิเคราะห์การถดถอยที่เป็นข้อมูลที่มีความยากระดับสูง

แมคคอลเลาจ์ (McCullough, 1999) ศึกษาโปรแกรมสำเร็จรูปที่ใช้ในการคำนวณสถิติ 3 โปรแกรมได้แก่ โปรแกรม SPSS 7.5 for WINDOWS, โปรแกรม SAS 6.12 และ โปรแกรม S – PLUS 4.0 โดยเปรียบเทียบความน่าเชื่อถือจากความแม่นของค่าประมาณทางสถิติ ซึ่งแบ่งระดับความยากของข้อมูลเป็น 3 ระดับ คือระดับต่ำ ระดับปานกลาง และ ระดับสูง ตามเกณฑ์ของ The National Institute of Standards and Technology (NIST) มีการกำหนดเลขนัยสำคัญของผลลัพธ์ 15 ตำแหน่ง และศึกษากระบวนการวิเคราะห์ทางสถิติ 4 วิธี คือ การวิเคราะห์สถิติตัวแปรเดียว, วิเคราะห์การถดถอยเชิงเส้น, วิเคราะห์การถดถอยไม่เชิงเส้น และ วิเคราะห์ความแปรปรวน ผลการศึกษาพบว่า ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานที่ระดับความยากของข้อมูลทั้ง 3 ระดับ พบว่า โปรแกรมทางสถิติทั้ง 3 โปรแกรม จะให้ความแม่นไม่แตกต่างกัน และค่าอัตสหสัมพันธ์ของโปรแกรม SAS 6.12 มีความแม่นสูงสุด การวิเคราะห์ถดถอยเชิงเส้นและวิเคราะห์ถดถอยไม่เชิงเส้นพบว่าโปรแกรมทั้ง 3 คำนวณได้แม่นไม่แตกต่างกันทุกๆ ระดับความยากของข้อมูล ส่วนการวิเคราะห์ความแปรปรวน, โปรแกรม SPSS 7.5 for WINDOWS และ โปรแกรม SAS 6.12 มีความแม่นน้อยที่ความยากของข้อมูลระดับปานกลางและที่ความยากของข้อมูลระดับสูงพบว่าโปรแกรม S – PLUS 4.0 ไม่สามารถคำนวณความแม่นได้

ครีทตันและดิง (Creighton and Ding, 2002) ศึกษาโปรแกรมสำเร็จรูปที่ใช้ในการคำนวณสถิติ 3 โปรแกรมได้แก่ โปรแกรม JMP โปรแกรม Excel และโปรแกรม SAS โดยเปรียบเทียบความน่าเชื่อถือจากความแม่นของค่าประมาณทางสถิติเช่นเดียวกับแมคคอลเลาจ์ (McCullough, 1999) ผลการศึกษาพบว่า การวิเคราะห์สถิติตัวแปรเดียว ที่ทุกระดับความยากของข้อมูลให้ค่าความแม่นไม่แตกต่างกัน แต่กรณีที่ระดับความยากปานกลางและสูง โปรแกรม JMP ให้ความแม่นน้อยกว่า โปรแกรม Excel และโปรแกรม SAS สำหรับการวิเคราะห์ถดถอยเชิงเส้นและการวิเคราะห์ถดถอยไม่เชิงเส้นนั้นทุกระดับความยากจะให้ความแม่นไม่แตกต่างกัน การวิเคราะห์ความแปรปรวนที่ความยากของข้อมูลระดับต่ำให้ความแม่นไม่แตกต่างกัน แต่ความยากระดับปานกลางและระดับสูงโปรแกรม SAS และ โปรแกรม JMP ไม่สามารถคำนวณค่าความแม่นได้และโปรแกรม Excel มีความแม่นน้อยที่สุดเมื่อเทียบกับโปรแกรม SAS และ JMP

กิตติกรณ์ (2545) ศึกษาโปรแกรมสำเร็จรูปที่ใช้ในการคำนวณสถิติ 3 โปรแกรม คือ โปรแกรม Statistica 5.5 โปรแกรม SPSS 10.0 for windows และ โปรแกรม S-plus 2000 ศึกษาการประมาณค่าทางสถิติ 3 ประเภท คือ สถิติตัวแปรเดียว การวิเคราะห์การถดถอย และการวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียว โดยจะทดสอบความแม่นของการประมาณค่าทางสถิติ เมื่อให้โปรแกรมทั้ง 3 ประมาณค่าทางสถิติประเภทเดียวกัน เพื่อพิจารณาว่าโปรแกรมใดมีความน่าเชื่อถือสูงสุดในการประมาณค่าทางสถิติ แบ่งระดับความยากของข้อมูลเป็น 3 ระดับคือ ระดับต่ำ ระดับปานกลาง และระดับสูง และใช้ข้อมูลจาก The National Institute of Standards and Technology (NIST) โดยพิจารณาจากค่า LRE (The base-10 Logarithm of the Relative Error) หรือ LAR (The base – 10 Logarithm of the Absolute Error) และได้จัดลำดับความแม่นของโปรแกรมที่ได้ และนำไปทดสอบความแตกต่างของความแม่นอีกครั้ง เนื่องจากข้อมูลที่ได้นั้นอยู่ในมาตราเรียงลำดับ (Ordinal Scale) จึงใช้การทดสอบที่ไม่อิงพารามิเตอร์ (Nonparametric Test) คือการวิเคราะห์ความแปรปรวนตามวิธีของฟรีดแมน (The Friendman Two-Way Analysis of Varience) และถ้าปฏิเสธสมมติฐาน H_0 จึงทำการทดสอบรายคู่ โดยใช้การทดสอบเครื่องหมาย (Sign Test) ผลการศึกษาพบว่า การวิเคราะห์สถิติตัวแปรเดียว ให้ค่าความแม่นไม่แตกต่างกัน การวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียว ผลการศึกษาพบว่า โปรแกรม Statistica 5.5 กับ โปรแกรม S-plus 2000 ให้ค่าความแม่นไม่แตกต่างกัน และมีความแม่นสูงกว่า โปรแกรม SPSS 10.0 for windows การวิเคราะห์ถดถอยเชิงเส้น ผลการศึกษาพบว่า โปรแกรม S-plus 2000 ให้ความแม่นสูงกว่า โปรแกรม SPSS 10.0 for windows และ Statistica 5.5

สถาบัน SAS (2003) ได้ศึกษาเปรียบเทียบโปรแกรม SAS ในรุ่น 6.12 และ 8.02 โดยใช้ข้อมูลมาตรฐานจาก NIST ในการเปรียบเทียบความแม่นของตัวเลขในทางสถิติด้านขั้นตอนวิธีในการวิเคราะห์ โดยพิจารณาจากค่า LRE (The base-10 Logarithm of the Relative Error) ผลการศึกษาพบว่า ในการวิเคราะห์สถิติตัวแปรเดียว SAS ซึ่งใช้คำสั่ง PROC UNIVARIATE ให้ค่าความแม่นสูงมาก ส่วนในการวิเคราะห์ความแปรปรวนและการวิเคราะห์การถดถอยเชิงเส้น ในชุดข้อมูลที่มีระดับความยากระดับปานกลางและสูงนั้น การใช้คำสั่ง PROC ORTHOREG จะเหมาะสมกว่าการใช้คำสั่ง PROC ANOVA และ PROC REG

เคลลี่และโรเบิร์ท (Kellie and Robert, 2007) ศึกษาโปรแกรมสำเร็จรูปที่ใช้ในการคำนวณสถิติ 9 โปรแกรมได้แก่ โปรแกรม Excel 2000/XP, Excel 2003, JMP5.0, Minitab14.0, R 1.9.1, SAS 9.1, SPSS 12.0, Splus 6.2, Stata 8.1 และ StatCrunch3.0 ศึกษาการประมาณค่าทางสถิติ 4 ประเภท คือ สถิติตัวแปรเดียว การวิเคราะห์การถดถอยเชิงเส้น การวิเคราะห์การถดถอยไม่เชิงเส้น และการวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียว โดยดูจากกราฟลักษณะที่สำคัญที่ศึกษา คือ การประมาณค่าทางสถิติของแต่ละโปรแกรม และดูถึงการพัฒนาของบริษัทที่รับผิดชอบในแต่ละโปรแกรมว่า ในรุ่นที่ใหม่ขึ้นได้ทำการปรับปรุงแก้ไขในจุดบกพร่องของโปรแกรมอะไรบ้าง โดยรวมแล้วผลการศึกษาพบว่า โปรแกรม StatCrunch3.0 ยังให้ผลได้ไม่ดีเมื่อเทียบกับโปรแกรมทั่วไป Excel 2000/XP มีความแม่นน้อย ส่วน Excel 2003 ได้ทำการพัฒนาขึ้นมาจนทำงานได้ผลเกือบเท่า Splus6.2 Minitab14.0 มีความแม่นสูงในหลายๆชุดข้อมูลและทำงานได้ใกล้เคียงกับ R1.9.1 และSAS เป็นโปรแกรมที่มีความแม่นสูง

2.2 ทฤษฎีที่เกี่ยวข้อง

การวิจัยครั้งนี้จะทำการเปรียบเทียบค่าประมาณทางสถิติ แบ่งเป็น 3 ประเภท คือ

2.2.1 สถิติตัวแปรเดียว (Univariate Statistics)

การวิจัยครั้งนี้จะทำการเปรียบเทียบค่าประมาณ สำหรับข้อมูลเชิงปริมาณ แบ่งเป็น 2 ชนิด คือ

1. ค่าเฉลี่ย (Mean)

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} \qquad (2.1)$$

โดยที่ $\bar{Y}$ คือ ตัวประมาณไม่เอนเอียงของ μ

n คือ จำนวนค่าสังเกต

Y_i คือ ค่าสังเกตจำนวนที่ i

2. ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation)

$$S = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}} \qquad (2.2)$$

โดยที่ S คือ ตัวประมาณไม่เอนเอียงของ σ

n คือ จำนวนค่าสังเกต

$\bar{Y}$ คือ ตัวประมาณไม่เอนเอียงของ μ

Y_i คือ ค่าสังเกตจำนวนที่ i

2.2.2 การวิเคราะห์การถดถอยเชิงเส้น (Linear Regression Analysis)

การวิเคราะห์การถดถอย คือ การศึกษาหาตัวแบบทางคณิตศาสตร์เพื่ออธิบายความสัมพันธ์ระหว่างตัวแปรอิสระ (Independent Variable) และตัวแปรตาม (Dependent Variable) จะมีทั้งความสัมพันธ์เชิงเส้นตรง (Linear) และไม่เป็นเส้นตรง (Nonlinear) ซึ่งถ้าความสัมพันธ์จากข้อมูลดังกล่าวประกอบด้วยตัวแปรอิสระ (X) 1 ตัว และตัวแปรตาม (Y) 1 ตัว จะเรียกว่า การวิเคราะห์ การถดถอยเชิงเส้นอย่างง่าย (Simple Regression) แต่ถ้ามีตัวแปรอิสระ (X) หลายตัวซึ่งมีความสัมพันธ์กับตัวแปรตาม (Y) จะเรียกว่าการวิเคราะห์การถดถอยพหุคูณ (Multiple Regression) ซึ่งการวิจัยครั้งนี้ได้ทำการศึกษาตัวแบบการถดถอยเชิงเส้น ดังต่อไปนี้

1. การถดถอยเชิงเส้นอย่างง่าย (The Simple Linear Regression)

ก) วัตถุประสงค์ของการวิเคราะห์การถดถอย

- เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรว่ามีความสัมพันธ์กันมากน้อยเพียงใด

- ใช้ความสัมพันธ์ที่วิเคราะห์ได้มาประมาณค่าหรือพยากรณ์ค่า Y ในอนาคตเมื่อกำหนด X

ข) สมมติฐานของการวิเคราะห์การถดถอย

- ความคลาดเคลื่อน ε_i เป็นตัวแปรที่มีค่าเฉลี่ยเป็นศูนย์หรือ $\mathrm{E}(\varepsilon_i) = 0$ ค่าความแปรปรวนของ ε_i มีค่าเท่ากันทุกค่าของ i และมีค่าเท่ากับความแปรปรวนของ Y $V(\varepsilon_i) = V(Y) = \sigma^2_{y.x} = \sigma^2$

- ε_i และ ε_j เป็นอิสระกัน นั่นคือ $Cov(\varepsilon_i, \varepsilon_j) = \mathrm{E}(\varepsilon_i \varepsilon_j) = 0\,;\ i \neq j$

- ε_i มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็นศูนย์และความแปรปรวนเท่ากับ σ^2

ซึ่งจะพิจารณาตัวแบบการถดถอยเชิงเส้นอย่างง่ายของตัวแปรอิสระ (X) และตัวแปรตาม (Y) ซึ่งเป็นความสัมพันธ์เชิงเส้นตรง

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad (2.3)$$

โดยที่ β_0 คือ ส่วนตัดบนแกน Y

β_1 คือ ค่าความชัน (X) ซึ่งทั้ง β_0 และ β_1 เป็นค่าคงตัวไม่ทราบค่า

และ ε คือ ค่าความคลาดเคลื่อนสุ่ม (Random Error Component)

ค) ค่าประมาณพารามิเตอร์ β_0 , β_1

ซึ่งในการวิจัยนี้จะใช้วิธีกำลังสองน้อยที่สุด (Least Square Estimation) ในการประมาณค่าพารามิเตอร์ β_0 , β_1 โดยพิจารณาจากค่าผลรวมกำลังสองของความคลาดเคลื่อนเป็นผลต่างระหว่างค่าสังเกต Y_i ใดๆ กับค่า $\hat{Y}_i$ ที่ถูกประมาณจากสมการ $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ ให้มีค่าต่ำที่สุด

ซึ่งกำหนดให้ e_i = ค่าความคลาดเคลื่อน (Error)

ดังนั้น $$e_i = Y_i - \hat{Y}_i \qquad (2.4)$$

ค่าประมาณพารามิเตอร์ β_0 และ β_1 คือ

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} \qquad (2.5)$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}$$

หรือ
$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} Y_i X_i - n\overline{X}\overline{Y}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2} \qquad (2.6)$$

โดยที่ $\overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$ และ $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$

ค่า $\hat{\beta}_0$, $\hat{\beta}_1$ ในสมการ (2.5) และ (2.6) เป็นตัวประมาณกำลังสองน้อยที่สุดของส่วนตัด (Intercept) และความชัน (Slope) ซึ่งตัวประมาณกำลังสองน้อยที่สุดโดยทฤษฎีของเกาส์ – มาร์คอฟซึ่งแสดงให้เห็นว่า ตัวแบบการถดถอยเชิงเส้น (2.3) ภายใต้ข้อตกลง $\mathrm{E}(\varepsilon) = 0$ และ $V(\varepsilon) = \sigma^2$ และค่าคลาดเคลื่อนไม่มีสหสัมพันธ์กันจะเป็นตัวประมาณที่ไม่เอนเอียงและมีความแปรปรวนต่ำสุด เมื่อเปรียบเทียบกับตัวประมาณที่ไม่เอนเอียงทั้งหมด ซึ่งโดยทั่วไปจะเรียกตัวประมาณโดยวิธีกำลังน้อยที่สุดว่า “BLUE” (Best Linear Unbiased Estimators)

งง) สถิติทดสอบ และตาราง ANOVA

การพิจารณาการทดสอบสมมติฐาน $\mathrm{H}_0 : \beta_1 = 0$ โดยใช้การวิเคราะห์ความแปรปรวนเข้าช่วยโดยเลือกสถิติทดสอบ

$$F = \frac{SS_R / 1}{SS_E / n-2} = \frac{MS_R}{MS_E} \qquad (2.7)$$

โดยที่ MS_R คือ ค่าเฉลี่ยกำลังสองของการถดถอย (Regression Mean Square)

MS_E คือ ค่าเฉลี่ยคลาดเคลื่อนกำลังสอง (Residual Mean Square)

ภายใต้สมมติฐาน $\mathrm{H}_0 : \beta_1 = 0$ เป็นจริงและ ณ ระดับนัยสำคัญที่กำหนด คือ α โดยใช้ทฤษฎีบทของ คอกแคร็น (Chochran) จะได้

$$F = \frac{\chi_1^2 / 1}{\chi_{n-2}^2 / n-2} \sim F_{\alpha,(1,n-2)}$$

และ $E(MS_E) = \sigma^2$, $E(MS_R) = \sigma^2 + \beta_1^2 S_{XX}$

ตารางวิเคราะห์ความแปรปรวนสำหรับการถดถอยเชิงเส้นอย่างง่าย ซึ่งแสดงได้ดังตารางที่ 2.1

ตารางที่ 2.1

แสดงการวิเคราะห์ความแปรปรวนสำหรับการถดถอยเชิงเส้นอย่างง่าย

ความผันแปร (SOV)	องศาอิสระ (d.f)	ผลรวมกำลังสอง (SS)	ค่ากำลังสองเฉลี่ย (MS)	ค่าสถิติ F
การถดถอย	1	$SS_R = \hat{\beta}_1 S_{XY}$	$MS_R = \frac{SS_R}{1}$	$F = \frac{MS_R}{MS_E}$
ค่าคลาดเคลื่อน	n-2	$SS_E = SS_Y - \hat{\beta}_1 S_{XY}$	$MS_E = \frac{SS_E}{n-2}$	
รวม	n-1	$SS_T = S_{YY}$		

2 การวิเคราะห์การถดถอยเชิงเส้นโค้งแบบโพลีโนเมียล (Polynomial Regression analysis)

ในกรณีที่ตัวแปรเกณฑ์ Y มีค่าเปลี่ยนไปตามตัวแปรอิสระ X หนึ่งตัว (Y เป็นฟังก์ชันกับกำลังของ X) การเปลี่ยนแปลงของค่า Y ตามกำลังต่างๆของค่า X นี้เรียกว่า การถดถอยแบบโพลีโนเมียล ซึ่งเขียนเป็นตัวแบบได้ดังนี้

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \ldots + \beta_p X^p + \varepsilon \qquad (2.8)$$

เมื่อ

$\beta_0, \beta_1, \beta_2, \beta_3, \ldots, \beta_p$ แทน ค่าสัมประสิทธิ์การถดถอยของประชากร (Population Regression Coefficients)

2.2.3 การวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียว (One – Way Analysis of Variance)

การวิเคราะห์ความแปรปรวนแบบมีปัจจัยเดียว เป็นการจำแนกข้อมูลด้วยตัวแปรหรือปัจจัยเพียงตัวเดียว วัตถุประสงค์ของการวิเคราะห์ความแปรปรวนแบบมีปัจจัยเดียว คือ การทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากรที่ได้รับปัจจัยที่ต่างระดับกันตั้งแต่ 3 ระดับขึ้นไป นั่นคือ เพื่อเปรียบเทียบค่าเฉลี่ยประชากรตั้งแต่ 3 ประชากรขึ้นไป โดยถือว่าหน่วยที่ได้รับปัจจัยระดับเดียวกันเป็นประชากรหนึ่งๆ และหน่วยที่ได้รับปัจจัยคนละระดับเป็นคนละประชากร

การเปรียบเทียบค่าเฉลี่ยประชากรโดยการทดสอบสมมติฐานนั้น จะต้องเก็บข้อมูลตัวอย่างหรือทำการทดลองโดยการกำหนดระดับของปัจจัยให้แก่หน่วยทดลอง(Experimental Unit) อย่างสุ่ม โดยที่จะเรียกปัจจัยว่า สิ่งทดลองหรือ ทรีตเมนต์ (Treatment) ดังนั้น ทรีตเมนต์หมายถึง วิธีการหรือลักษณะต่างๆ ที่ต้องการเปรียบเทียบ โดยข้อมูลที่นำมาใช้ในการเปรียบเทียบจะวัดได้จากหน่วยทดลอง

สมมติฐานการทดสอบ

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$$

$$H_1 : \mu_i \neq \mu_j \text{ อย่างน้อย 1 คู่ } ; i \neq j$$

เมื่อ $Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad ; i = 1,2,...,t; \; j = 1,2,...,r_i$

โดยที่ μ คือ ค่าเฉลี่ยรวม

τ_i คือ อิทธิพลของทรีตเมนต์ที่ $i, i = 1,2,...,t$

ε_{ij} คือ ความคลาดเคลื่อนสุ่ม $\varepsilon_{ij} \sim NID(0, \sigma^2)$

สมมติฐานของการวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียว

1. ประชากรทั้ง k ประชากร ต้องมีการแจกแจงปกติ
2. การสุ่มตัวอย่างแต่ละชุดจากประชากร จะต้องเป็นอิสระกัน
3. ความแปรปรวนของแต่ละประชากรต้องเท่ากัน

ความผันแปรทั้งหมด = ความผันแปรระหว่างทรีตเมนต์ + ความผันแปรในทรีตเมนต์เดียวกัน

$$SS_T = SS_{trt} + SS_E \qquad (2.9)$$

โดยที่ SS_T คือ ความผันแปรทั้งหมด $\sum_{i=1}^{t}\sum_{j=1}^{r_i}(Y_{ij}-\bar{Y}..)^2$

SS_{trt} คือ ความผันแปรระหว่างทรีตเมนต์ $\sum_{i=1}^{t}\sum_{j=1}^{r_i}(\bar{Y}_{i.}-\bar{Y}..)^2$

SS_E คือ ความผันแปรในทรีตเมนต์เดียวกัน $\sum_{i=1}^{t}\sum_{j=1}^{r_i}(Y_{ij}-\bar{Y}_{i.})^2$

สำหรับองศาอิสระจะเป็นสองส่วนเช่นเดียวกับความผันแปรทั้งหมด ดังนี้

องศาอิสระทั้งหมด = องศาอิสระระหว่างทรีตเมนต์ + องศาอิสระ (ภายในทรีตเมนต์)

df : $(N-1) = (t-1)+\sum_{i=1}^{t}(r_i-1)$ เมื่อ $N=\sum_{i-1}^{t}r_i$

สูตรที่ง่ายต่อการคำนวณ ดังนี้

$$SS_T = \sum_{i=1}^{t}\sum_{j=1}^{r_i}Y_{ij}^2-\frac{\bar{Y}_{..}^2}{N} \quad (2.10)$$

$$SS_{trt} = \sum_{i=1}^{t}\frac{Y_{i.}^2}{r_i}-\frac{Y_{..}^2}{N} \quad (2.11)$$

$$SS_E = SS_T - SS_{trt} \quad (2.12)$$

เมื่อกำหนดให้ $CT=\frac{X_{..}^2}{N}$ ซึ่งแทนในสมการที่ (2.10), (2.11) และ (2.12) จะได้สูตรใหม่ คือ

$$SS_T = \sum_{i=1}^{t}\sum_{j=1}^{r_i}X_{ij}^2-CT \quad (2.13)$$

$$SS_{trt} = \sum_{i=1}^{t}\frac{X_{i.}^2}{r_i}-CT \quad (2.14)$$

การวิเคราะห์ความแปรปรวนหลังจากที่คำนวณค่าผลรวมกำลังสองและองศาอิสระ แล้วนำมาสร้างตารางวิเคราะห์ความแปรปรวนซึ่งมีลักษณะดัง ตารางที่ 2.2

ตารางที่ 2.2

แสดงการวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียว

ความผันแปร (SOV)	องศาอิสระ (d.f)	ผลรวมกำลังสอง (SS)	ค่ากำลังสองเฉลี่ย (MS)	ค่าสถิติ F
Treatment	$t-1$	$SS_{trt}=\sum_{i=1}^{t}\frac{Y_{i.}^2}{r_i}-\frac{Y_{..}^2}{N}$	$MS_{trt}=\frac{SStrt}{t-1}$	$F=\frac{MS_{trt}}{MS_E}$
Error	$N-t$	$SS_E=SS_T-SS_{trt}$	$MS_E=\frac{SS_E}{N-t}$	
Total	$N-1$	$SS_T=\sum_{i=1}^{t}\sum_{j=1}^{r_i}Y_{ij}^2-\frac{\overline{Y}_{..}^2}{N}$		

2.2.4 ลักษณะทั่วไปของตัวเลข

ลักษณะทั่วไปของตัวเลขที่ใช้ในการคำนวณและการเก็บค่าของหน่วยความจำของเครื่องคอมพิวเตอร์จะมี 2 แบบ (สุวรรณ: 2531) คือ

1. เลขแบบจุดทศนิยมลอย (Foating Point Arithmetic) คือ การเก็บค่าต่างๆ ที่ได้จากการคำนวณของเครื่องคอมพิวเตอร์ ซึ่งการเก็บค่าตัวเลขนั้นจะอยู่ในรูปของ

$$a=\pm .d_1d_2d_3...\times 10^q \qquad (2.15)$$

เมื่อ a เป็นเลขใดๆในระบบเลขฐานสิบ

q เป็นจำนวนเต็ม

โดยที่ $1\le d_1\le 9$ และ $0\le d_i\le 9, i\ge 2$

ถ้าให้ $m=.d_1d_2d_3...$

m จะมีค่าอยู่ระหว่าง $0.1\le m\le 1(.999...=1)$ ทำให้เราสามารถเขียน a ใหม่ได้

$$a=\pm m\times 10^q \qquad (2.16)$$

ซึ่ง m เรียกว่า แมนทิสซา (mantissa)

q เรียกว่า เลขชี้กำลัง (exponent)

จำนวนตัวเลขที่ใช้สำหรับค่าของ m และ q มีจำนวนจำกัด แล้วแต่ชนิดของเครื่องคอมพิวเตอร์ จำนวน a ซึ่งเขียนได้ $a=\pm m\times 10^q$ นี้จะถูกประมาณค่าโดยเครื่องคอมพิวเตอร์ด้วย $\overline{a}=\pm\hat{m}\times 10^q$ ซึ่ง $\hat{m}$ เป็นตัวประมาณของ m

การเก็บค่าตัวเลขแบบจุดทศนิยมลอย จะมีการตัดทอนเลขออก (Round Off Errors) คือ เป็นความคลาดเคลื่อนเนื่องมากจากเครื่องคอมพิวเตอร์ทำการตัดตัวเลขเพียงจุดทศนิยมที่ต้องการเท่านั้น มี 2 วิธี คือ

(1) การตัดออกแบบมีกฎเกณฑ์ (rounding)

1.1 ถ้า $.d_{t+1}d_{t+2}... < .5$ ให้ตัดทิ้งไป

ดังนั้น $s_i = d_i \quad i = 1,2,3,....t$

และ $s_i = 0$ เมื่อ $i = t+1, t+2, \ldots$

เช่น

$$D = 1.23456432$$
$$\downarrow$$
$$S = 1.23456$$

1.2 ถ้า $.d_{t+1}d_{t+2}... > .5$ ให้ปัดเศษขึ้นไป 1 แล้วตัดทิ้ง

ดังนั้น $s_i = d_i \quad i = 1,2,3,....t-1$

และ $s_t = d_t + 1 \quad s_i = 0$ เมื่อ $i = t+1, t+2, \ldots$

เช่น

$$D = 1.23456789$$
$$\downarrow$$
$$S = 1.23457$$

1.3 ถ้า $.d_{t+1}d_{t+2}... = .5$

1.3.1 ถ้า $.d_{t+1}d_{t+2}... = .5$ และ d_t เป็นเลขคู่ (even) ให้ตัดทิ้ง

ดังนั้น $s_i = d_i \quad i = 1,2,3,....t$

1.3.2 ถ้า $.d_{t+1}d_{t+2}... = .5$ และ d_t เป็นเลขคี่ (odd) ให้ปัดเศษไป 1

ดังนั้น $s_i = d_i \quad i = 1,2,3,....t-1$ และ $s_t = d_t + 1$

(2) การตัดออกเลย (chopping)

$$s_i = d_i, \quad i = 1,2,3,....t$$

โดยตัดเทอมที่ d_i เมื่อ $i = t+1,...$ ออก (เหมือนมีค่าเป็นศูนย์)

เช่น

$$D = 8.123456789$$
$$\downarrow$$
$$S = 8.12345$$

แม้การเก็บค่าของเครื่องคอมพิวเตอร์ในลักษณะเลขแบบจุดทศนิยมลอยจะยุ่งยากและมีความคลาดเคลื่อน (Error) เกิดขึ้นก็ตาม แต่ข้อดีคือสามารถเก็บค่าในช่วงที่กว้างเมื่อใช้จำนวนบิตเท่ากับเลขแบบจุดทศนิยมคงที่

2. เลขแบบจุดทศนิยมคงที่ (Fixed Point Arithmetic) คือ การเก็บตัวเลขมีลักษณะง่ายมีการกำหนดจุดทศนิยมให้มีตำแหน่งคงที่ คือ อยู่ที่ตำแหน่งขวามือของหลักขวาสุด ดังนั้นการกำหนดให้เครื่องคอมพิวเตอร์เก็บค่าเลขแบบจุดทศนิยมคงที่ จึงไม่สามารถเก็บเป็นทศนิยมได้ ดังนั้นจะเก็บได้เฉพาะเลขที่เป็นจำนวนเต็ม เช่น +4500501502 ซึ่งไม่มีส่วนของแมนทิสซา และเลขชี้กำลังให้ยุ่งยาก ค่าตัวเลขที่จะเก็บได้จึงมีค่าจำกัดอยู่ระหว่าง -99999999999 ถึง +99999999999 (เป็นการเปรียบเทียบในระบบฐานสิบ) ค่าตัวเลขที่ต้องการใช้หลักมากกว่านี้เรียกว่า โอเวอร์โฟลว์ (Overflow) และค่าตัวเลขที่ต้องการใช้หลักน้อยกว่านี้เรียกว่า อันเดอร์โฟลว์ (Underflow)

2.2.5 ความเป็นมาของแหล่งข้อมูลที่ใช้ในการวิเคราะห์จาก The National Institute of Standards and Technology (NIST) มีดังนี้

ข้อมูลและค่ามาตรฐานได้เตรียมไว้เพื่อประเมินความแม่นของโปรแกรม สำหรับ สถิติตัวแปรเดียว, การถดถอยเชิงเส้น, การถดถอยไม่เชิงเส้น และการวิเคราะห์ความแปรปรวน การรวบรวมข้อมูลนี้ประกอบด้วยทั้ง ข้อมูลที่ถูกสร้างขึ้น(generated) และ ข้อมูลจริง (real-world) ของหลายๆระดับความยาก ชุดข้อมูลที่ถูกสร้างถูกออกแบบเพื่อเปรียบเทียบความแม่นในการคำนวณโดยเฉพาะ ข้อมูลเหล่านี้รวมไปถึงชุดข้อมูล Wampler สำหรับทดสอบขั้นตอนวิธีการวิเคราะห์การถดถอย และชุดข้อมูล Simon & Lesage สำหรับทดสอบขั้นตอนวิธีการวิเคราะห์ความแปรปรวนจำแนกทางเดียว ข้อมูลจริง (real-world) รวมไปถึงข้อมูลชุดสำคัญ เช่น ข้อมูล Longley สำหรับ linear regression และค่ามาตรฐาน คือ การแก้ปัญหาที่ดีที่สุดที่สามารถทำได้

ข้อมูลถูกเรียงลำดับตามความยาก (ต่ำ,ปานกลาง,สูง) ระดับความยากของข้อมูลชุดนี้ขึ้นอยู่กับขั้นตอนวิธีการวิเคราะห์ ข้อมูลเหล่านี้เป็นการแนะแนวทางแบบคร่าวๆสำหรับผู้ใช้ และการได้ผลลัพธ์ที่ถูกต้องในทุกชุดข้อมูลที่ระดับความยากสูง ไม่ได้หมายความว่าโปรแกรมนั้นจะให้

ผลลัพธ์ที่ถูกต้องของทุกชุดข้อมูลที่ระดับความยากปานกลางหรือแม้แต่ระดับความยากระดับต่ำในทำนองเดียวกัน การให้ผลลัพธ์ที่ถูกต้องในข้อมูลทุกชุดไม่ได้หมายความว่าโปรแกรมนั้นๆจะให้ค่าที่ถูกต้องในข้อมูลที่มีลักษณะเฉพาะอื่นๆ อย่างไรก็ตามก็เป็นการให้ความเชื่อมั่นในด้านการให้ผลลัพธ์ที่ถูกต้อง เพื่อที่จะรู้ว่าบางโปรแกรมให้ผลลัพธ์ไม่ถูกต้อง เนื่องด้วยเกิดความคลาดเคลื่อนจากสาเหตุต่างๆ และNIST ได้วางแผนที่จะปรับปรุงและพัฒนาการรวบรวมข้อมูลในการคำนวณสถิติที่มากขึ้นและรับฟังการตอบรับในด้านวิธีการทางสถิติ เพื่อจัดข้อมูล และบริการให้ดีขึ้น

2.2.6 แหล่งที่ทำให้เกิดความคลาดเคลื่อน

การคำนวณด้วยตัวเลขสำหรับเครื่องคอมพิวเตอร์นั้นสิ่งที่หลีกเลี่ยงไม่ได้ คือ ผลลัพธ์ที่เกิดความคลาดเคลื่อนขึ้นซึ่งความคลาดเคลื่อนที่เกิดขึ้นมาจากแหล่งต่างๆ กัน ซึ่งแยกประเภทได้ดังนี้ (McCullough: 1998)

1. การแสดงเลขในระบบฐานสองและความเที่ยงตรงที่จำกัด (Binary Representation and Finite Precision)

สำหรับคอมพิวเตอร์นั้นจะทำการแปลงข้อมูลที่เก็บเป็นเลขฐานสอง ไว้ในหน่วยความจำของคอมพิวเตอร์ คือ 0 และ 1 ซึ่งในระบบเลขฐานสอง แต่ละตัว เรียกว่า บิต (bit) ย่อมาจากคำว่า Binary Digit ซึ่งเป็นหน่วยที่เล็กที่สุดในการเก็บข้อมูลด้วยคอมพิวเตอร์ แต่เนื่องจากบิตเพียง 1 ตัว ไม่สามารถเก็บข้อมูลตัวเลข ตัวอักษรและสัญลักษณ์ต่างๆ ได้ครบ ดังนั้นจึงรวมบิตหลายบิตเข้าเป็นกลุ่มเรียกว่า ไบต์ (byte) ซึ่งแต่ละไบต์จะใช้แทนอักขระหนึ่งตัว โดยปกติแล้วจะใช้ 8 บิตรวมกันเป็น 1 ไบต์ นอกจากนี้ยังมีหน่วยความจุเป็น เมกะไบต์ ใช้สัญลักษณ์ MB ซึ่งประมาณความจุ 1,048,579 ไบต์ แต่ปัจจุบันนี้หน่วยความจำมีความจุอยู่ในหน่วย จิกะไบต์ (Gigabyte) ใช้สัญลักษณ์ GB ซึ่งประมาณความจุ 1,073,741,824 ไบต์ ซึ่งจะเห็นว่าปัจจุบันมีการพัฒนาหน่วยความจำให้เก็บค่าได้มากขึ้นซึ่งการเก็บค่าต่างๆ จะต้องมีการกำหนดให้แน่นอนซึ่งเราเรียกว่า ความเที่ยงตรงที่จำกัด (Finite Precision)

ความคลาดเคลื่อนจากการแสดงเลขในระบบฐานสองเลขและความเที่ยงตรงที่จำกัดจะแสดงให้เห็นตามตัวอย่าง ดังนี้

0.1 แปลงเป็นเลขฐานสองจะได้ .0001100110<u>0011</u> จะเห็นว่าตัวเลขที่ขีดเส้นใต้นั้นเป็นจำนวนซ้ำที่ไม่รู้จบ ซึ่งถ้ากำหนดการเก็บข้อมูลเป็นแบบดับเบิลพรีซิสชันกำหนด 9 ตำแหน่ง เมื่อคอมพิวเตอร์แสดงผลลัพธ์เป็นเลขฐานสิบ จะได้ 0.09999996 จะเห็นว่าค่าที่ได้ไม่ตรงกันกับ

ค่าแรก ดังนั้นจึงทำให้เกิดค่าคลาดเคลื่อน เรียกว่า ความคลาดเคลื่อนจากการแทนเลขฐานสอง (Binary Representation error)

2. ขั้นตอนการดำเนินการทางคณิตศาสตร์ (Mathematical Operations)

การดำเนินการทางคณิตศาสตร์ซึ่งเป็นการใช้มือในการคำนวณและเป็นพื้นฐานของการคำนวณด้วยคอมพิวเตอร์ ซึ่งการใช้คอมพิวเตอร์ช่วยในการคำนวณต้องควบคุมขั้นตอนในการคำนวณให้ถูกต้อง เพราะคอมพิวเตอร์จะแปลงค่าตัวเลขเป็นรูปแบบจุดทศนิยมลอยและทำให้ขั้นตอนการคำนวณแตกต่างกันไป ซึ่งจะทำให้เกิดค่าคลาดเคลื่อนได้ซึ่งจะแสดงให้เห็นตามตัวอย่างดังนี้ (Roger,et al.:1998)

(2.5 – 2.4) / 0.1 = 1.0000000000000008881…65625

ขณะที่

$$\frac{2.5}{0.1} - \frac{2.4}{0.1} = 1.0000000000000035527\ldots90625$$

3. ความคลาดเคลื่อนการปัดเศษ (Rounding Error)

ความคลาดเคลื่อนการปัดเศษ เป็นความคลาดเคลื่อนเนื่องมากจากเครื่องคอมพิวเตอร์และการเก็บค่าของหน่วยความจำที่แน่นอนของเครื่องคอมพิวเตอร์ที่ใช้ ไม่สามารถเก็บค่าตัวเลขที่ได้จากการคำนวณครบทุกตำแหน่งระหว่างที่มีการดำเนินการทางคณิตศาสตร์เกิดขึ้นสำหรับข้อจำกัดของเครื่องคำนวณที่ทำให้ต้องตัดเอาเลขทศนิยมบางตำแหน่งทิ้งไปซึ่งจะแสดงให้เห็นตามตัวอย่างดังนี้ (McCullough: 1998)

X = 1,000,000 y = 0.000001 ดังนั้น x + y = z

สำหรับค่า z เมื่อทำการกำหนดเป็นแบบซิงเกิลพรีซิสชัน จะได้ z = 1,000,000 แต่ถ้ากำหนดเป็นแบบดับเบิลพรีซิสชัน จะได้ z = 1,000,000.000001 จะเห็นว่าการกำหนดค่าเป็นแบบซิงเกิลพรีซิสชัน ในตัวอย่างนี้จะคลาดเคลื่อนขึ้น เพราะว่ามีการปัดเศษทิ้ง

4. ความคลาดเคลื่อนจากการตัดปลาย (Truncation Errors)

ความคลาดเคลื่อนจากการตัดปลาย เป็นความคลาดเคลื่อนเนื่องจากการหยุดขบวนการคำนวณก่อนที่ขบวนการคำนวณนั้นสิ้นสุดลง เช่น การคำนวณหาค่าของอนุกรมอนันต์ (Infinite Series)

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots$$

ซึ่งจากการคำนวณนี้จำเป็นที่จะต้องหยุดลงเมื่อการคำนวณผ่านไปเป็นจำนวนเทอมที่แน่นอนจำนวนใดจำนวนหนึ่งเท่านั้น ทั้งนี้เป็นเพราะคอมพิวเตอร์ไม่อาจทำการคำนวณต่อไปแบบไม่มีที่สิ้นสุดได้

5. ความคลาดเคลื่อนสะสม (Accumulation Error)

ความคลาดเคลื่อนสะสม เป็นความคลาดเคลื่อนเนื่องมาจากเครื่องคำนวณที่ใช้ไม่สามารถเก็บค่าตัวเลขที่ได้จากการคำนวณครบทุกตำแหน่ง ทำให้เกิดความคลาดเคลื่อนที่เปลี่ยนไปในลักษณะเพิ่มขึ้นหรือลดลง แล้วแต่ลักษณะการคำนวณนั้นๆ ค่าคลาดเคลื่อนในข้อมูลนี้อาจเกิดจากการปัดเลขก่อนและหลังการคำนวณมักจะเป็นสิ่งสำคัญที่ต้องให้ความระมัดระวังอย่างมากหากคำนวณต่อเนื่องกันหลายครั้ง ค่าคลาดเคลื่อนอาจจะมากพอจะรบกวนค่าที่คำนวณได้จนหมดความหมายในที่สุดจะแสดงให้เห็นตามตัวอย่างดังนี้ (Roger,et al.:1998)

$$\sum_{i=1}^{100} 0.11111 = 1.1111000000000002430...875 \times 10^{1}$$

$$\sum_{i=1}^{10000} 0.11111 = 1.111100000002723027...125 \times 10^{3}$$

6. ความคลาดเคลื่อนการตัดออก (Cancellation Error)

ความคลาดเคลื่อนการตัดออก เป็นความคลาดเคลื่อนที่เกิดจาก จำนวน 2 จำนวนที่มีค่าใกล้เคียงกันมาลบกันจะทำให้เหลือผลลัพธ์ของหลังตำแหน่งทศนิยมที่แตกต่างกันและเกิดการปัดเศษผลลัพธ์ทำให้ผลที่ได้เกิดคลาดเคลื่อนขึ้น ซึ่งจะแสดงให้เห็นตามตัวอย่างดังนี้

a = 97233452898.75 = 97233452898.75

b = -97233452898.60 = -97233452898.6000006103515625

c = -0.15 = -0.1499999999999999944…75

ดังนั้น

$$(a+b)+c = -6.103...984375 \times 10^{-6}$$

$$a+(b+c) = -0.000...000000 \times 10^{0}$$