บทที่ 3

วิธีการดำเนินการศึกษา

3.1 การสร้างแบบจำลอง

ในการสร้างแบบจำลองการประมาณต้นทุนของงานอาคารซึ่งจะเป็นฟังค์ชั่นของตัวแปรอิสระหลายตัวนี้ สิ่งแรกที่จะต้องทำก็คือการกำหนดตัวแปรอิสระที่คิดว่าจะมีอิทธิพลต่อตัวแปรตามหรือต้นทุน นอกจากนี้การกำหนดลักษณะแนวโน้มของข้อมูลหรือฟังค์ชั่นที่จะใช้ในการทำนายตัวแปรตามว่าจะอยู่ในรูปใด

3.1.1 ตัวแปรอิสระ

ในการศึกษาวิจัยครั้งนี้ได้เลือกตัวแปรอิสระที่คิดว่ามีอิทธิพลต่อต้นทุน 9 ตัวแปร ดังนี้

1. ความสูงของอาคาร
2. ความสูงระหว่างชั้นเฉลี่ย
3. จำนวนชั้น
4. เส้นรอบรูปเฉลี่ย
5. พื้นที่ใช้งานรวม
6. พื้นที่หลังคา
7. พื้นที่ห้องน้ำ
8. พื้นที่พื้นที่วางบนดิน
9. พื้นที่ช่องเปิด

แต่ในการหาแบบจำลองและการนำแบบจำลองที่ได้ไปใช้กับโครงการใหม่ต้องทราบถึงข้อกำหนด และวิธีการหาตัวแปรอิสระต่าง ๆ จากแบบดังนี้

3.1.1.1 ความสูงของอาคาร

ความสูงของอาคารมีค่าเท่ากับระยะจากระดับพื้นชั้นล่างสุดถึงเพดานห้องเครื่อง หรือถังเก็บน้ำ หรือสันหลังคา มีหน่วยเป็นเมตร

3.1.1.2 ความสูงระหว่างชั้นเฉลี่ย

ความสูงระหว่างชั้นเฉลี่ยคำนวณจากผลรวมของระยะจากระดับพื้นถึงเพดานของทุกชั้นหารด้วยจำนวนชั้น ยกเว้นชั้นดาดฟ้า มีหน่วยเป็นเมตร

3.1.1.3 จำนวนชั้น

จำนวนชั้นมีค่าเท่ากับจำนวนชั้นทั้งหมดซึ่งรวมชั้นใต้ดินแต่ไม่รวมชั้นดาดฟ้า มีหน่วยเป็นชั้น

3.1.1.4 เส้นรอบรูปเฉลี่ย

เส้นรอบรูปเฉลี่ยคำนวณได้จากผลรวมของเส้นรอบรูปทุกชั้นหารด้วยจำนวนชั้น แต่ในการวัดเส้นรอบรูปจะวัดรอบผนังอาคารเท่านั้น มีหน่วยเป็นเมตร

3.1.1.5 พื้นที่ใช้งานรวม

พื้นที่ใช้งานรวมมีค่าเท่ากับผลรวมของพื้นที่ใช้งานของทุกชั้นยกเว้น พื้นดาดฟ้า และพื้นที่วางบนดิน มีหน่วยเป็นตารางเมตร

3.1.1.6 พื้นที่หลังคา

พื้นที่หลังคามีค่าเท่ากับพื้นที่ที่มีการมุงหลังคา ทั้งกระเบื้องหรือวัสดุอื่น ๆ มีหน่วยเป็นตารางเมตร

3.1.1.7 พื้นที่ห้องน้ำ

พื้นที่ห้องน้ำมีค่าเท่ากับผลรวมของพื้นที่ห้องน้ำทั้งหมดของอาคาร มีหน่วยเป็นตารางเมตร

3.1.1.8 พื้นที่พื้นที่วางบนดิน

พื้นที่พื้นที่วางบนดินมีค่าเท่ากับพื้นที่รวมของพื้นที่วางบนดินส่วนมากจะเป็นพื้นชั้นล่างสุดที่ติดกับผิวดิน มีหน่วยเป็นตารางเมตร

3.1.1.9 พื้นที่ช่องเปิด

พื้นที่ช่องเปิดมีค่าเท่ากับผลรวมของพื้นที่ช่องเปิดต่าง ๆ เช่น ช่องแสง หน้าต่าง ประตู มีหน่วยเป็นตารางเมตร

หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

3.1.2 ตัวแปรตาม

ในการประยุกต์ใช้การวิเคราะห์การถดถอย (Regression) กับการประมาณราคาอาคารนั้น สำหรับการวิจัยที่ผ่าน ๆ มามักจะให้ราคารวมเป็นตัวแปรตามหรือตัวที่ต้องการจะทำนาย แต่ในการศึกษาวิจัยครั้งนี้ตัวแปรตามที่นำมาพิจารณานอกเหนือจากราคารวมคือ ต้นทุนหมวดโครงสร้าง สถาปัตย์ ไฟฟ้า สุขาภิบาล และต้นทุนรวม โดยในแต่ละหมวดจะไม่รวมต้นทุนที่เกี่ยวกับ เช่น ถนน สวนหย่อม และลานจอดรถ

3.1.3 รูปแบบฟังค์ชั่น

ต้นทุนต่าง ๆ หรือตัวแปรตามจะถูกทำนายจากฟังค์ชั่นต้นทุนซึ่งเขียนในรูปทั่วไปดังนี้

$$Y = f(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9)$$

โดยที่ Y = ต้นทุนค่าก่อสร้าง

X_1 = ความสูงของอาคาร

X_2 = ความสูงระหว่างชั้นเฉลี่ย

X_3 = จำนวนชั้น

X_4 = เส้นรอบรูปเฉลี่ย

X_5 = พื้นที่ใช้งานรวม

X_6 = พื้นที่หลังคา

X_7 = พื้นที่ห้องน้ำ

X_8 = พื้นที่พื้นที่วางบนดิน

X_9 = พื้นที่ช่องเปิด

ในการวิจัยครั้งนี้ได้เลือกฟังค์ชั่นที่จะใช้เป็นแบบจำลองทางคณิตศาสตร์ ดังนี้

ฟังค์ชั่นเส้นตรง (Linear function)

$$Y = a + b_1X_1 + b_2X_2 + \ldots + b_9X_9$$

ฟังค์ชั่นตัวแปรยกกำลัง (Power function)

$$Y = aX_1^{b1}X_2^{b2} \ldots X_9^{b9}$$

ส่วนการพิจารณาว่าฟังค์ชั่นใดจะเหมาะสมกับข้อมูลดีที่สุด ได้กล่าวในหัวข้อการเลือกฟังค์ชันเป็นแบบจำลอง แล้วว่าจะพิจารณาจากค่าสัมประสิทธิ์ตัวกำหนดหรือ r^2 ถ้า

016551

ฟังค์ชั่นใดมีค่า r^2 สูงสุดจะเป็นฟังค์ชั่นที่เหมาะสมกับข้อมูลที่ดีที่สุด และพิจารณาจากค่าเบี่ยงเบนมาตรฐานหรือค่า S ถ้าฟังค์ชั่นใดมีค่า S ต่ำ จะเป็นฟังค์ชั่นที่เหมาะสมกับข้อมูลดีที่สุด

3.1.4 การวิเคราะห์หาแบบจำลองที่เหมาะสมที่สุด

ในการศึกษาหาแบบจำลองการประมาณต้นทุนโดยวิธีการวิเคราะห์การถดถอยนี้ เนื่องจากว่าข้อมูลที่เก็บรวบรวมได้ ผู้วิจัยไม่สามารถรู้ลักษณะความสัมพันธ์ของข้อมูลหรือรู้ว่าแบบจำลองใดจะดีที่สุดและใช้แบบจำลองใดในการทำนายต้นทุนของโครงการใหม่ได้แม่นยำที่สุด ซึ่งในการวิจัยครั้งนี้ได้มีวิธีการสร้างและวิเคราะห์หาแบบจำลองที่เหมาะสมที่สุดดังนี้

ในการวิจัยครั้งนี้ได้สร้างแบบจำลองจากกลุ่มตัวอย่างหรือข้อมูลที่เก็บรวบรวมมาของอาคารแต่ละประเภท โดยจำนวนข้อมูลที่ใช้ในการสร้างแบบจำลองจะเท่ากับจำนวนข้อมูลที่เก็บรวบรวมมาทั้งหมดตัดออก 1 ข้อมูล และจะใช้ข้อมูลที่ตัดออกไปนี้ในการตรวจสอบความแม่นยำของแบบจำลองที่สร้างขึ้นจากข้อมูลอื่นและจะสร้างแบบจำลองโดยเปลี่ยนข้อมูลตัวที่ตัดออกไป 1 ข้อมูลจนครบทุกตัวอย่าง ซึ่งจะได้จำนวนแบบจำลองทั้งหมดเท่ากับจำนวนข้อมูลทั้งหมด นำแบบจำลองทั้งหมดที่สร้างได้ มาพิจารณาเปรียบเทียบกับแบบจำลองที่สร้างขึ้นจากจำนวนข้อมูลทั้งหมดเพื่อพิจารณาเลือกแบบจำลองที่เหมาะสมที่สุด

ในการพิจารณาว่าแบบจำลองใดจะเหมาะสมกับข้อมูลดีที่สุด โดยจะพิจารณาค่าเบี่ยงเบนมาตรฐานหรือค่า S ของการทำนายของข้อมูลทั้งหมด ซึ่งค่าเบี่ยงเบนมาตรฐานเป็นค่าที่แสดงการกระจัดกระจายของข้อมูล ในการพิจารณาถ้าแบบจำลองใดมีค่า S มากแสดงว่าแบบจำลองนั้นจะมีข้อมูลกระจัดกระจายห่างจากเส้นถดถอยมาก นั่นคือเมื่อนำแบบจำลองนี้มาใช้ทำนายต้นทุนแล้วจะมีโอกาสผิดพลาดสูง แต่ถ้าค่า S มีค่าน้อยแสดงว่าแบบจำลองนั้นจะมีข้อมูลกระจัดกระจายใกล้เส้นถดถอยที่สร้างขึ้นมาก และเมื่อนำแบบจำลองนี้มาใช้ทำนายต้นทุนจะมีโอกาสถูกต้องมาก สรุปได้ว่าแบบจำลองใดที่มีค่าเบี่ยงเบนมาตรฐานหรือค่า S ของข้อมูลทั้งหมดต่ำสุด จะเป็นแบบจำลองที่เหมาะสมกับข้อมูลและมีความน่าเชื่อถือสูง

3.1.5 ดัชนีราคา (Price Index)

เนื่องจากว่าต้นทุนที่จะนำมาใช้ในการสร้างแบบจำลองจำเป็นต้องเป็นค่าในปีเดียวกัน ดังนั้นจึงต้องอาศัยดัชนีราคา มาเปลี่ยนต้นทุนที่อยู่ในปีต่างกันให้เป็นค่าในปีเดียวกันหรือปีฐาน แต่เนื่องจากว่าดัชนีราคาซึ่งเป็นตัวบอกความเปลี่ยนแปลงราคาวัสดุและค่าแรงนั้น ในประเทศไทยยังไม่มีการสร้างเพื่อนำมาใช้ในการปรับราคาค่าก่อสร้าง ดังนั้นในการวิจัยครั้งนี้จะใช้ดัชนีราคาวัสดุก่อสร้างราคาขายส่ง (แสดงในตารางที่ 3.1) ซึ่งเป็นตัวบอกความเปลี่ยนแปลงราคาวัสดุเท่านั้น ซึ่งจัดทำโดยกรมเศรษฐกิจการพาณิชย์มาเปลี่ยนต้นทุนในปีต่าง ๆ ให้อยู่ในปีฐานคือ พ.ศ. 2519 ก่อนที่จะนำไปวิเคราะห์หาแบบจำลอง และเมื่อนำแบบจำลองที่ได้มาใช้ทำนายต้นทุนของโครงการใหม่ ต้นทุนที่ทำนายได้จะเป็นค่า ในปี พ.ศ.2519 ดังนั้นเมื่อจะหาค่าในปีใดก็ให้ใช้ดัชนีราคาของปีนั้น ๆ คูณกับต้นทุนที่คำนวณได้จากแบบจำลอง

3.1.6 การวิเคราะห์หาเส้นการถดถอยโดยใช้ไมโครคอมพิวเตอร์

เนื่องจากการหาเส้นการถดถอยหลายตัวแปร มีการคำนวณที่ยุ่งยากและใช้เวลานาน ดังนั้นจึงใช้เครื่องคอมพิวเตอร์ช่วย ซึ่งสามารถให้ผลได้ถูกต้องและน่าเชื่อถือยิ่งขึ้น แต่การใช้คอมพิวเตอร์ จำเป็นต้องมีโปรแกรมสำเร็จรูป และในการศึกษาวิจัยครั้งนี้ได้เลือกใช้โปรแกรมการฟิตสมการเชิงเส้นหลายตัวแปร (ก้องเกียรติ, 2532)

3.2 สถิติที่ใช้ในการวิเคราะห์

หลังจากสร้างแบบจำลองการประมาณต้นทุนแล้วได้ทำการทดสอบความแม่นยำในการทำนายต้นทุน วิธีทางสถิติที่ใช้ในการทดสอบแบบจำลอง มีดังนี้

3.2.1 ค่าสัมประสิทธิ์ของตัวกำหนด (Coefficient of Determination)

ค่าสัมประสิทธิ์ของตัวกำหนดจะเป็นตัวบอกให้ทราบว่า การกระจายทั้งหมดของค่า Y นั้นสามารถอธิบายได้จากเส้นการถดถอยกี่เปอร์เซนต์ หรือเป็นตัวบอกความแม่นยำในการทำนาย Y จากเส้นการถดถอยกี่เปอร์เซนต์

จาก $Y_i-\bar{Y} = (\hat{Y}_i-\bar{Y})+(Y_i-\hat{Y}_i)$

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

แก้สมการได้ว่า

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

ค่า $\Sigma (Y_i-\bar{Y})^2$ เรียกว่าผลรวมกำลังสองของการกระจายทั้งหมด (Total sum square)

ค่า $\Sigma (\hat{Y}_i-\bar{Y})^2$ เรียกว่าผลรวมกำลังสองของการกระจายที่อธิบายได้ (Explained sum square)

ค่า $\Sigma (Y_i-\hat{Y}_i)^2$ เรียกว่าผลรวมกำลังสองของการกระจายที่อธิบายไม่ได้ (Unexplained sum square) ซึ่งเท่ากับ Σe_i^2 นั่นเอง

เพราะฉะนั้น Total ss = Explained ss + Unexplained ss นั่นคือ Total ss เป็นค่าการกระจายของ Y ที่เกิดขึ้นในข้อมูลทั้งหมด จะเท่ากับผลบวกของ Explained ss เป็นค่าการกระจายของค่า Y ที่คาดคะเนได้จากเส้นการถดถอย กับ Unexplained ss ซึ่งเป็นค่าการกระจายของ Y ที่เกิดจากผลต่างระหว่าง Y ที่แท้จริงกับ Y ที่ทำนายได้ แต่ไม่สามารถอธิบายได้จากเส้นการถดถอย

เส้นการถดถอยหรือแบบจำลองที่ดีที่สามารถนำไปคาดคะเนค่า Y ในอนาคตโดยมีโอกาสที่จะถูกต้องตรงกับค่า Y ที่แท้จริงได้มากนั้น จะต้องสามารถอธิบายการกระจายที่เกิดขึ้นในค่า Y ได้มากที่สุดนั่นคือ ผลรวมกำลังสองของการกระจายที่อธิบายได้จะต้องมีค่าใกล้เคียงกับผลรวมกำลังสองของการกระจายทั้งหมด และถ้าผลรวมกำลังสองของการกระจายที่อธิบายได้เท่ากับผลรวมกำลังสองของการกระจายทั้งหมดแล้ว แบบจำลองสามารถอธิบายการกระจายของ Y ทั้งหมดได้ถึง 100% ถ้านำแบบจำลองไปคาดคะเนค่า Y อนาคตจะได้ค่า Y ถูกต้องกับความเป็นจริง 100% เช่นกัน

โดยเหตุผลดังกล่าวข้างต้น จึงได้มีการวัดอัตราส่วนระหว่างผลรวมกำลังสองของการกระจายที่อธิบายได้กับผลรวมกำลังสองของการกระจายทั้งหมด ซึ่งเรียกว่า สัมประสิทธิ์ของตัวกำหนด (coefficient of Determination) (r^2) นั่นคือ

$$r^2 = \frac{\Sigma (\hat{Y}_i-\bar{Y})^2}{\Sigma (Y_i-\bar{Y})^2}$$

หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

ค่า r^2 ในทางปฏิบัติเรามักคูณด้วย 100 เพื่อคำนวณออกมาในรูปของเปอร์เซ็นต์ ดังนั้นค่า r^2 จะมีค่าตั้งแต่ 0 ถึง 100% และใช้ r^2 เป็นเครื่องแสดงอิทธิพลของตัวแปรอิสระ (X) ที่มีต่อตัวแปรตาม (Y) ทั้งนี้เพราะว่าจากค่า r^2 ที่คำนวณได้จะบอกให้ทราบว่าการกระจายทั้งหมดของค่า Y นั้น สามารถอธิบายได้จากเส้นการถดถอยกี่เปอร์เซ็นต์ หรืออาจพูดได้อีกอย่างว่า r^2 นั้นเป็นค่าบอกให้เราทราบว่า การกระจายของค่า Y ทั้งหมดนั้นจะสามารถอธิบายได้จากค่า X ได้กี่เปอร์เซ็นต์ หรือ X มีอิทธิพลต่อ Y กี่เปอร์เซนต์ นอกจากนี้ยังใช้ค่า r^2 ในการพิจารณาเลือกเส้นการถดถอยหรือแบบจำลองที่เหมาะสมกับข้อมูลที่ดีที่สุด นั่นคือถ้าแบบจำลองใดมีค่า r^2 สูงสุดจะเป็นแบบจำลองที่เหมาะสมกับข้อมูลดีที่สุด

3.2.2 <u>ส่วนเบี่ยงเบนมาตรฐานของการประมาณ</u> (Standard Error of Estimate)

การใช้เส้นการถดถอยนำไปคาดคะเนค่า Y ในอนาคตนั้นจะถูกต้องแค่ไหนขึ้นอยู่กับข้อมูลที่รวบรวมมาได้นั้น จะกระจัดกระจายไปจากเส้นการถดถอยที่หาได้มากน้อยเพียงใด กล่าวคือ ถ้าข้อมูลกระจัดกระจายห่างจากเส้นการถดถอยมาก การคาดคะเน Y จะมีโอกาสที่จะแตกต่างไปจาก Y ที่แท้จริงได้ง่าย ในทางตรงข้ามถ้าค่าของข้อมูลกระจัดกระจายห่างจากเส้นการถดถอยน้อย การคาดคะเน Y จะมีโอกาสถูกต้องกับ Y ที่แท้จริงได้มากที่สุด มาตรการที่ใช้วัดการกระจายของข้อมูล Y ที่รวบรวมได้นี้จะวัดอยู่ในรูปของความแตกต่างระหว่าง Y กับ $\hat{Y}$ หรือค่า e นั่นเอง

จาก $e_i = Y_i - \hat{Y}_i$

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

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

$\frac{\Sigma e_i^2}{df} = \frac{\Sigma (Y_i - \hat{Y}_i)^2}{n-k-1}$ ซึ่งเรียกว่า ความแปรปรวนของการประมาณ (Variance of estimate)

$\sqrt{\frac{\Sigma e_i^2}{df}} = \sqrt{\frac{\Sigma (Y_i - \hat{Y}_i)^2}{n-k-1}}$ เรียกว่าส่วนเบี่ยงเบนมาตรฐานของการประมาณ (Standard error of estimate)

ให้ $S_{Y/X1,\ldots,Xk}$ เป็นส่วนเบี่ยงเบนมาตรฐานของการประมาณค่า Y บนค่า X

ได้ว่า $S_{y/x1,\ldots,xk} = \sqrt{\frac{\Sigma (Y_i - \hat{Y}_i)^2}{n-k-1}}$

นอกจากนี้การหาค่า $S_{Y/X1,\ldots,Xk}$ ยังสามารถหาได้โดยวิธีแยกการกระจายของข้อมูลออกเป็นหลาย ๆ ส่วนดังนี้

จาก Total ss = Explained ss + Unexplained ss

หรือ Unexplained ss = Total ss - Explained ss

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

แต่ $\Sigma (Y_i-\bar{Y})^2 = \Sigma Y_i^2$

และ $\Sigma (\hat{Y}_i-\bar{Y})^2 = \sum_{j=1}^{k} b_j (\sum_{i=1}^{n} X_{ji} Y_i)$

$j = k$ (จำนวนตัวแปรอิสระ)

$i = n$ (จำนวนข้อมูลที่สุ่มมา)

ดังนั้น $$S_{y/x1,\ldots,xk} = \sqrt{\frac{\Sigma Y_i^2 - \sum_{j=1}^{k} b_j (\sum_{i=1}^{n} X_{ji} Y_i)}{n-k-1}}$$

3.2.3 <u>การวิเคราะห์ความแปรปรวน</u> (Analysis of Variance)

การทดสอบความแปรปรวนของเส้นการถดถอยเชิงซ้อนหรือมีตัวแปรอิสระหลายตัวนั้น จะเป็นการทดสอบดูความมีนัยสำคัญของตัวแปรอิสระว่ามีอิทธิพลต่อค่า Y หรือไม่ หรือกล่าวอย่างง่าย ๆ คือ ทดสอบดูว่าตัวแปรอิสระที่ประกอบกันเป็นแบบจำลองสามารถนำมาใช้ทำนายค่า Y ได้หรือไม่ ในการทดสอบความแปรปรวนของเส้นการถดถอยนี้จะทดสอบความมีนัยสำคัญของตัวแปรอิสระทุกตัวรวมกัน และตารางการวิเคราะห์ความแปรปรวนเป็นดังนี้

ตารางการวิเคราะห์ความแปรปรวน

Source	Degree of freedom (df)	Sum of Square (ss)	Mean Square (MS)	F_{cal}	$F_{\alpha,df}$
Regression	k	$\sum_{j=1}^{k} b_j (\sum_{i=1}^{n} X_{ji} Y_i)$	ss/df	$\frac{MSR}{MSE}$	
Residual (Error)	n-k-1	$\sum Y_i^2 - \sum_{j=1}^{k} b_j (\sum_{i=1}^{n} X_{ji} Y_i)$	ss/df		
Total	n-1	$\sum Y_i^2$			

โดย $F_{\alpha,df,k}$ เป็นค่า F ที่ได้จากการเปิดตารางสถิติค่า F ที่ df = n-k-1 และที่ระดับความเชื่อมั่น $1-\alpha$

MSR คือ Mean Square ของ Regression และ

MSE คือ Mean Square ของ Residual หรือ Error

ผลที่ได้จากการเปรียบเทียบค่า F ที่คำนวณได้หรือ F_{cal} กับ $F_{\alpha,df,k}$ ที่ได้จากการเปิดตารางสถิติจะมี 2 ลักษณะคือ

1) ถ้า F คำนวณน้อยกว่า F จากการเปิดตาราง ก็สรุปได้ว่า ตัวแปรอิสระที่มีอิทธิพลต่อ Y อย่างไม่มีนัยสำคัญ

2) ถ้า F คำนวณมากกว่า F จากการเปิดตาราง ก็สรุปได้ว่า ตัวแปรอิสระอย่างน้อยหนึ่งตัวที่มีอิทธิพลต่อ Y อย่างมีนัยสำคัญยิ่งขึ้น

3.3 การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลเป็นการถอดแบบเพื่อหาตัวแปรอิสระต่าง ๆ ดังที่กล่าวมา และคำนวณหาต้นทุนค่าก่อสร้างแต่ละหมวดคือ หมวดโครงสร้าง สถาปัตย์ ไฟฟ้า สุขาภิบาล ซึ่งรายละเอียดของงานแต่ละหมวดได้แสดงในภาคผนวก ข ต้นทุนที่ใช้นี้คำนวณหาทั้งจากราคากลางหรือราคาเจ้าของงานและจากราคาประมูลหรือราคาของผู้รับเหมาที่ประมูลงานนั้น ๆ ได้ นอกจากเก็บข้อมูลต้นทุนแล้ว ยังเก็บราคาประมูลของผู้รับเหมาที่ได้งาน นำมาวิเคราะห์หาเส้นการถดถอยเช่นกัน

ข้อมูลส่วนใหญ่จะเป็นข้อมูลของอาคารที่ทำการก่อสร้างเสร็จไปแล้ว และเนื่องจากมีเวลาในการวิจัยจำกัดประกอบกับแหล่งข้อมูลมีจำนวนจำกัด ดังนั้นจึงเก็บข้อมูลให้มีจำนวนเพียงพอที่สามารถนำมาใช้วิเคราะห์โดยวิธีการวิเคราะห์การถดถอยได้ของอาคารแต่ละประเภท โดยได้เก็บรวบรวมข้อมูลทั้งจากหน่วยงานเอกชนและหน่วยงานราชการ รวมทั้งสิ้น 6 หน่วยงานดังนี้

3.3.1 บริษัทสถาปนิก 110 จำกัด

3.3.2 บริษัทเอสคอน จำกัด

3.3.3 บริษัทซีด้าคอนซันแตนส์ จำกัด

3.3.4 จุฬาลงกรณ์มหาวิทยาลัย

3.3.5 กรมโยธาธิการ

3.3.6 กรุงเทพมหานคร

3.4 สมมุติฐานที่ใช้ในการวิเคราะห์ข้อมูล

ในการศึกษาวิจัยครั้งนี้จะมีสมมติฐานที่ใช้ในการวิเคราะห์ข้อมูลดังนี้

3.4.1 วิเคราะห์ข้อมูลโดยการรวบรวมข้อมูลทั้งหมดที่ได้จากหน่วยงานต่าง ๆ เข้าด้วยกันโดยไม่คำนึงว่าแต่ละหน่วยงานจะมีข้อแตกต่างในเรื่องระเบียบและกฎเกณฑ์ในการประมาณราคากลางหรือราคาเพื่อประมูลงานก่อสร้าง

3.4.2 ไม่คำนึงถึงข้อมูลว่าจะมีแนวโน้มเอียงหรือเบี่ยงเบนเนื่องจากการประมูลงานที่ตัดราคากัน หรือมีส่วนได้ส่วนเสียกัน เช่นการฮั้วงานกันเป็นต้น