


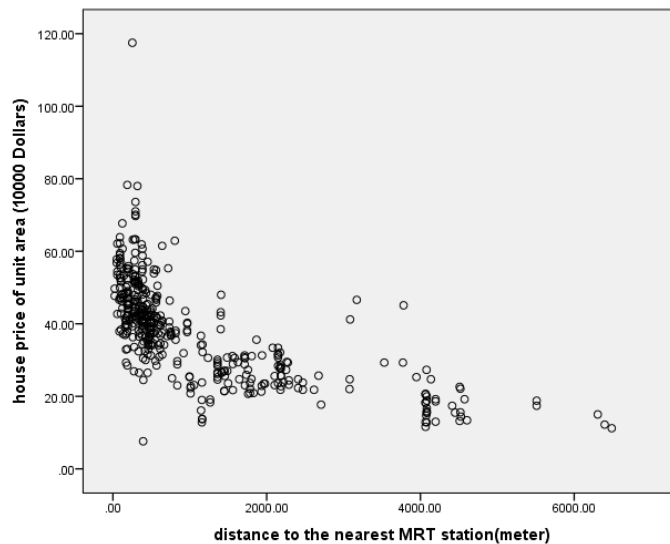
## ใบงานการวิเคราะห์ความสัมพันธ์ของตัวแปรเชิงปริมาณด้วยโปรแกรม SPSS For Window

1. การวิเคราะห์ความสัมพันธ์ของตัวแปรเชิงปริมาณที่มีเพียง 2 ตัว สามารถใช้เทคนิค 2 วิธีดังนี้
  - a. การวิเคราะห์ถดถอยอย่างง่าย
  - b. การวิเคราะห์สหสัมพันธ์

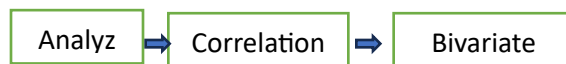
ก่อนวิเคราะห์จะต้องตรวจสอบรูปแบบของความสัมพันธ์ ว่าเป็นรูปแบบใด เช่น ความสัมพันธ์ในรูปเชิงเส้น เส้นโค้ง พาราโบลาหรือเอ็กโพเนนเชียล เป็นต้น จึงจะสามารถเลือกเทคนิคสำหรับการวิเคราะห์ต่อไปได้อย่างถูกต้อง ตรวจสอบโดยใช้การพิจารณาจากกราฟแสดงความสัมพันธ์ (Scatterplot)

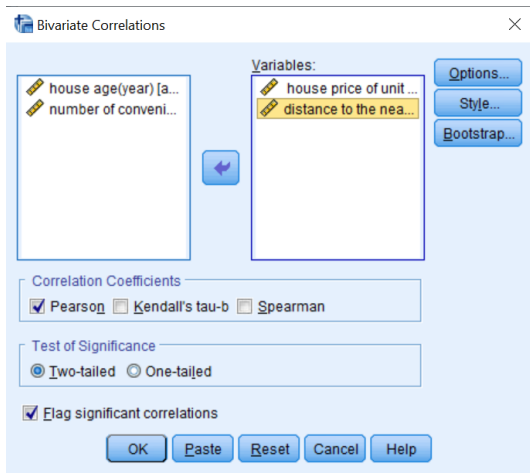
การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย โดยใช้ชุดข้อมูล real estate เพื่อวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรราคากับระยะทางเป็นเมตรไปถึง MRT ที่ใกล้ที่สุด

1. Plot graph ดูรูปแบบความสัมพันธ์ ใช้คำสั่ง 



2. ตรวจสอบความสัมพันธ์ของตัวแปรทั้งสองว่ามีความสัมพันธ์เชิงเส้นกันจริงหรือไม่

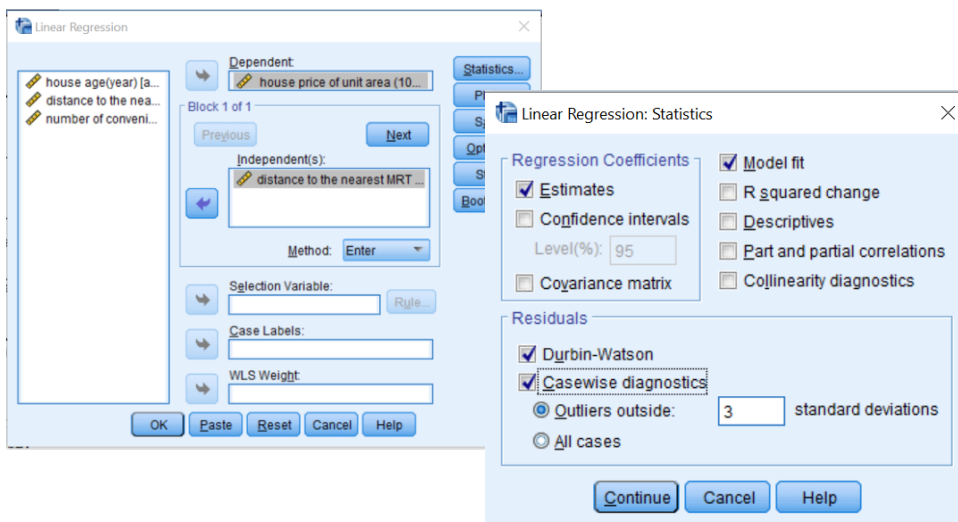
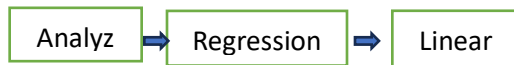




### Correlations

	house price of unit area (10000 Dollars)	distance to the nearest MRT station (meter)
house price of unit area (10000 Dollars)	1	-.674**
		.000
	N	414

3. หาค่า a และ b ซึ่งเป็นค่าประมาณพารามิเตอร์  $\beta_0$  และ  $\beta_1$  ตามลำดับ



### Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.674 <sup>a</sup>	.454	.452	10.06854	2.151

a. Predictors: (Constant), distance to the nearest MRT station(meter)

b. Dependent Variable: house price of unit area (10000 Dollars)

Durbin watson ใช้วัดความคลาดเคลื่อนว่าเป็นอิสระกันหรือไม่ซึ่งเป็นข้อตกลงของการวิเคราะห์ถดถอยเชิงเส้นถ้าค่า Durbin watson เข้าใกล้ 2 แสดงว่าความคลาดเคลื่อนเป็นอิสระกัน

### การทดสอบความเหมาะสมของสมการถดถอยเชิงเส้นอย่างง่าย

เพื่อทดสอบว่าสมการที่ได้จากการวิเคราะห์เป็นสมการที่แสดงความสัมพันธ์ของตัวแปรต้นและตัวแปรตามจริงหรือไม่ ใช้ทดสอบสมมติฐาน  $H_0: \beta_1 = 0$  โดยใช้ F-test จากตาราง ANOVA หรือใช้  $H_1: \beta_1 \neq 0$  โดยใช้ T-test จากตาราง Coefficients อย่างใดอย่างหนึ่งซึ่งผลจะเหมือนกัน

### ANOVA<sup>a</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	34694.677	1	34694.677	342.239	.000 <sup>b</sup>
	Residual	41766.700	412	101.375		
	Total	76461.378	413			

a. Dependent Variable: house price of unit area (10000 Dollars)

b. Predictors: (Constant), distance to the nearest MRT station(meter)

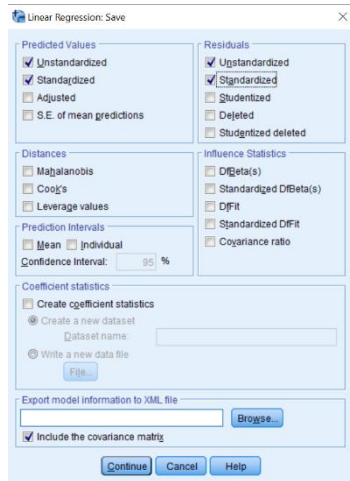
### Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	45.851	.653		70.258	.000
	distance to the nearest MRT station(meter)	-.007	.000	-.674	-18.500	.000

a. Dependent Variable: house price of unit area (10000 Dollars)

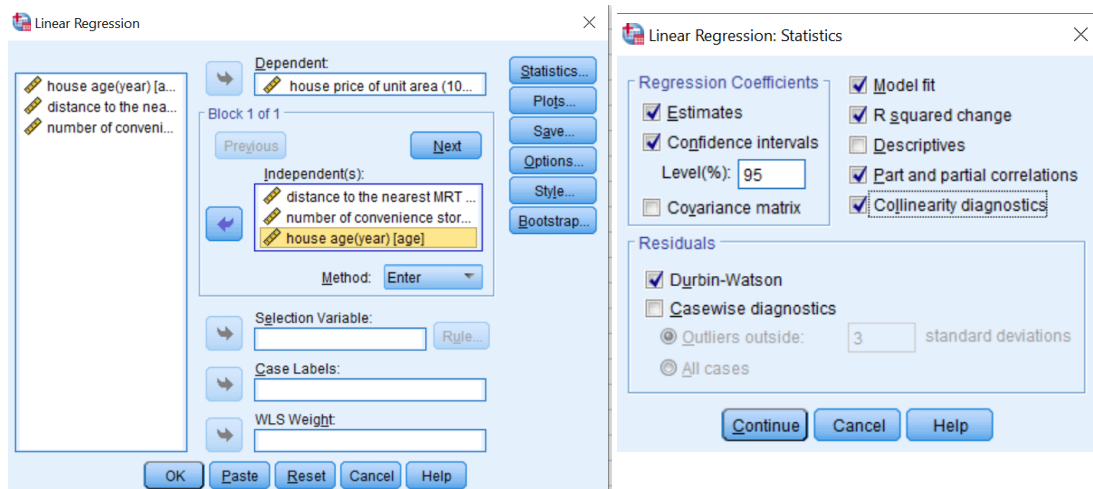
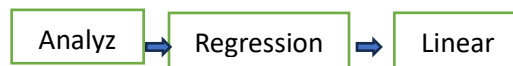
ตารางCoefficients แสดงค่าประมาณ และการทดสอบ t-test ของพารามิเตอร์ทั้ง 2 ตัว

เลือกบันทึกค่าประมาณ ( $\hat{Y}_i$ ) ค่าความคลาดเคลื่อนและสถิติที่ใช้ตรวจสอบ ซึ่ง  $\hat{Y}_i$  จะถูกเก็บไว้ในไฟล์ข้อมูลที่เปิดอยู่ในรูปตัวแปรใหม่



	age	distance	store	price	PRE_1	RES_1	ZPR_1	ZRE_1
1	32.00	84.88	10	37.90	45.23503	-7.33503	.79154	-.72851
2	19.50	306.59	9	42.20	43.62492	-1.42492	.61587	-.14152
3	13.30	561.98	5	47.30	41.77027	5.52973	.41351	.54921
4	13.30	561.98	5	54.80	41.77027	13.02973	.41351	1.29410
5	5.00	390.57	5	43.10	43.01510	.08490	.54933	.00843
6	7.10	2175.03	3	32.10	30.05625	2.04375	-.86454	.20298
7	34.50	623.47	7	40.30	41.32373	-1.02373	.36480	-.10168
8	20.30	287.60	6	46.70	43.76284	2.93716	.63091	.29172
9	31.70	5512.04	1	18.80	5.82272	12.97728	-3.50853	1.28889
10	17.90	1783.18	3	22.10	32.90188	-10.80188	-.55407	-1.07284
11	34.80	405.21	1	41.40	42.90875	-1.50875	.53773	-.14965
12	6.30	90.46	9	58.10	45.19453	12.90547	.78712	1.28176
13	13.00	492.23	5	39.30	42.27682	-2.97682	.46878	-.29566
14	20.40	2469.65	4	23.80	27.91674	-4.11674	-1.09797	-.40887
15	13.20	1164.84	4	34.30	37.39231	-3.09231	-.06414	-.30713
16	35.70	579.21	2	50.50	41.64519	9.85481	.39987	.87945
17	.00	293.00	6	70.10	43.72366	26.37634	.62664	2.61968
18	17.70	350.85	1	37.40	43.30353	-5.90353	.58080	-.58633
19	16.90	368.14	8	42.30	43.17800	-.87800	.56711	-.08720
20	1.50	23.38	7	47.70	45.68162	2.01838	.84026	.20046

- การวิเคราะห์ความสัมพันธ์ของตัวแปรเชิงปริมาณมากกว่า 2 ตัว ด้วยเทคนิคการวิเคราะห์ความถดถอยเชิงพหุ โดยใช้ชุดข้อมูล real estate เพื่อวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรราคา(Y)เป็นตัวแปรตามกับตัวแปรต้นได้แก่ ระยะทางเป็นเมตรไปถึง MRT ที่ใกล้ที่สุด( $x_1$ ) จำนวนร้านสะดวกซื้อในระยะเดินไปได้( $x_2$ ) และอายุการใช้งานของที่พัก( $x_3$ ) ได้ดังนี้



Method คือวิธีการเลือกตัวแปรอิสระเข้าสมการถดถอยซึ่งมีหลายวิธี

Regression coefficient เป็นรายการสำหรับการประมาณค่าพารามิเตอร์ในสมการถดถอยพหุเชิงเส้นและค่าต่างๆ

Model fit เป็นรายการสำหรับแสดงตัวแปรนำเข้าและเอาออกจากสมการถดถอย

Multiple R ค่า  $R^2$  และ adjust  $R^2$  ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า  $\hat{y}$

ตารางANOVA เป็นการทดสอบสมการถดถอยเชิงพหุจากการวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียว

สำหรับการทดสอบ  $H_0: \beta_1 = \beta_2 = \beta_3$

$H_1: \beta_i \neq 0$  อย่างน้อย 1 ค่า  $i = 1, 2, 3$

สถิติทดสอบคือ  $F = \frac{MSR}{MSE}$

ผลการทดสอบ: ถ้า ยอมรับ  $H_0$  สรุปได้ว่า Y ไม่มีความสัมพันธ์กับ X ทั้ง 3 ตัว ในรูปเชิงเส้น

ถ้า ปฏิเสธ  $H_0$  สรุปได้ว่า มี X อย่างอย่างน้อย 1 ตัวที่มีความสัมพันธ์กับ Y ในรูปเชิงเส้น จึงต้องทดสอบต่อไปว่า X ตัวใดที่มีความสัมพันธ์กับ Y โดยใช้สถิติ t ในตาราง coefficient

R square change จะแสดงค่า  $R^2$  ที่เปลี่ยนแปลง(เพิ่มขึ้น)เมื่อมีตัวแปรอิสระเพิ่มขึ้นในสมการถดถอย และค่า F change เมื่อจำนวนตัวแปรอิสระเปลี่ยนแปลง

Part correlations จะแสดงค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรแต่ละคู่โดยไม่มีการควบคุมตัวแปรอื่นที่เกี่ยวข้อง

Partial correlations สัมประสิทธิ์สหสัมพันธ์เชิงส่วนของตัวแปรแต่ละคู่เมื่อมีการควบคุมอิทธิพลตัวแปรอื่น ๆ

Collinearity diagnostics จะแสดงค่าสำหรับการตรวจสอบความสัมพันธ์กันของตัวแปรอิสระ

- Tolerance ของตัวแปร  $x_i$  คือ  $1 - R_i^2$  เมื่อ  $R_i^2$  เป็นสัมประสิทธิ์สหสัมพันธ์เชิงพหุที่ได้จากสมการถดถอยของตัวแปรอิสระโดยมีตัวแปรอิสระ  $x_i$  เป็นตัวแปรตามและตัวแปรอิสระอื่นเป็นตัวแปรอิสระในสมการ ถ้า Tolerance ของตัวแปร  $x_i$  เข้าใกล้ศูนย์แสดงว่าตัวแปรอิสระ  $x_i$  มีความสัมพันธ์กันตัวแปรอิสระอื่นๆ แสดงว่าเกิด Multicollinearity ไม่เป็นไปตามเงื่อนไขของการวิเคราะห์การถดถอย

- VIF (Variance Inflation Factor) VFI ของตัวแปร  $x_i$  คำนวณจาก  $\frac{1}{1 - R_i^2}$

- Eigenvalue ถ้ามีค่าใกล้ศูนย์แสดงว่าตัวแปรอิสระนั้นมีความสัมพันธ์กับตัวแปรอิสระอื่นๆ ผลบวกของ Eigenvalue จะเท่ากับ k+1 เมื่อ k คือจำนวนตัวแปรอิสระ
- Condition Index =  $\sqrt{\frac{\text{Eigenvalue}_{\max}}{\text{Eigenvalue}_i}}$  ถ้ามีค่ามากแสดงว่า แสดงว่า ตัวแปรอิสระนั้นมีความสัมพันธ์กับตัวแปรอิสระอื่นๆอย่างมาก (มากกว่า 20 จะถือว่ามาก)

- ผลจากการวิเคราะห์

Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.736 <sup>a</sup>	.541	.538	9.25136	.541	161.123	3	410	.000	2.115

a. Predictors: (Constant), house age(year), distance to the nearest MRT station(meter), number of convenience stores in the living circle on foot

b. Dependent Variable: house price of unit area (10000 Dollars)

- R Square = 0.541 หมายถึง  $X_1, X_2, X_3$  รวมกันอธิบายการเปลี่ยนแปลงของ Y ได้ 54.1 % อีก 45.9 % เกิดจากอิทธิพลของตัวแปรอื่นที่ไม่ได้นำมาพิจารณา
- Durbin-Watson ใกล้ 2 แสดงว่าความคลาดเคลื่อนอิสระกัน

ANOVA<sup>a</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	41370.445	3	13790.148	161.123	.000 <sup>b</sup>
	Residual	35090.932	410	85.588		
	Total	76461.378	413			

$H_0: \beta_1 = \beta_2 = \beta_3$

$H_1: \beta_i \neq 0$

Sig ของ F =0.00 ปฏิเสธ  $H_0$  มี  $\beta_i$  อย่างน้อย 1 ค่าที่ไม่เป็น 0

Coefficients<sup>a</sup>

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
	B	Std. Error				Beta	Zero-order	Partial	Part	Tolerance
	1 (Constant)	42.977	1.385				31.041	.000		
distance to the nearest MRT station(meter)	-.005	.000	-.499	-11.874	.000	-.674	-.506	-.397	.634	1.578
number of convenience stores in the living circle on foot	1.297	.194	.281	6.678	.000	.571	.313	.223	.633	1.580
house age(year)	-.253	.040	-.212	-6.305	.000	-.211	-.297	-.211	.993	1.007

ตาราง Coefficients ใช้สำหรับทดสอบตัวแปรอิสระแต่ละตัวว่าเหมาะสมที่จะอยู่ในสมการพยากรณ์หรือไม่ โดย

ใช้การทดสอบ t-test สำหรับสมมติฐาน  $H_0 : \beta_i = 0$

$H_1 : \beta_i \neq 0$

ถ้าผลการทดสอบยอมรับ  $H_0$  แสดงว่าตัวแปร  $x_i$  ไม่มีผลต่อการเปลี่ยนแปลงค่า  $\hat{y}_i$  นั่นคือ  $x_i$  จึงไม่ควรอยู่ในสมการพยากรณ์สะท้อนให้ทราบว่า  $x_i$  มีอิทธิพลไม่มากพอต่อการเปลี่ยนแปลงค่า  $\hat{y}_i$

จากตารางวิเคราะห์พบว่า  $x_1, x_2, x_3$  มีผลต่อการเปลี่ยนแปลงค่า  $\hat{y}_i$  อย่างมีนัยสำคัญทางสถิติ และ ควรมีค่าคงที่ในสมการพยากรณ์

สมการพยากรณ์คือ  $\hat{y} = 42.977 - 0.005x_1 + 1.297x_2 + 0.253x_3$

#### Collinearity Diagnostics<sup>a</sup>

Mode	Dimension	Eigenvalue	Condition Index	Variance Proportions			
				(Constant)	distance to the nearest MRT station (meter)	number of convenience stores in the living circle on foot	house age (year)
1	1	2.951	1.000	.01	.02	.02	.03
	2	.746	1.989	.00	.32	.10	.00
	3	.232	3.569	.02	.12	.18	.86
	4	.071	6.453	.97	.54	.70	.11

a. Dependent Variable: house price of unit area (10000 Dollars)

#### Residuals Statistics<sup>a</sup>

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	2.0550	53.0305	37.9802	10.00852	414
Residual	-37.30428	77.31479	.00000	9.21770	414
Std. Predicted Value	-3.589	1.504	.000	1.000	414
Std. Residual	-4.032	8.357	.000	.996	414

a. Dependent Variable: house price of unit area (10000 Dollars)

**Residuals Statistics** เป็นค่าสถิติของค่าความคลาดเคลื่อน ช่วยในการตรวจสอบค่าผิดปกติ

- **Predicted Value** แสดงค่าประมาณของตัวแปรตาม
- **Residual** แสดงค่าความคลาดเคลื่อนที่เกิดจากการประมาณค่าตัวแปรตาม(ผลต่างของค่าจริงกับค่าประมาณที่ได้จากสมการถดถอย)
- **Std. Predicted Value** แสดงค่าพยากรณ์ในรูปมาตรฐาน
- **Std. Residual** ค่าคลาดเคลื่อนมาตรฐาน (ถ้าไม่มากเกินไป 3 ถือว่าไม่มีค่าผิดปกติ)