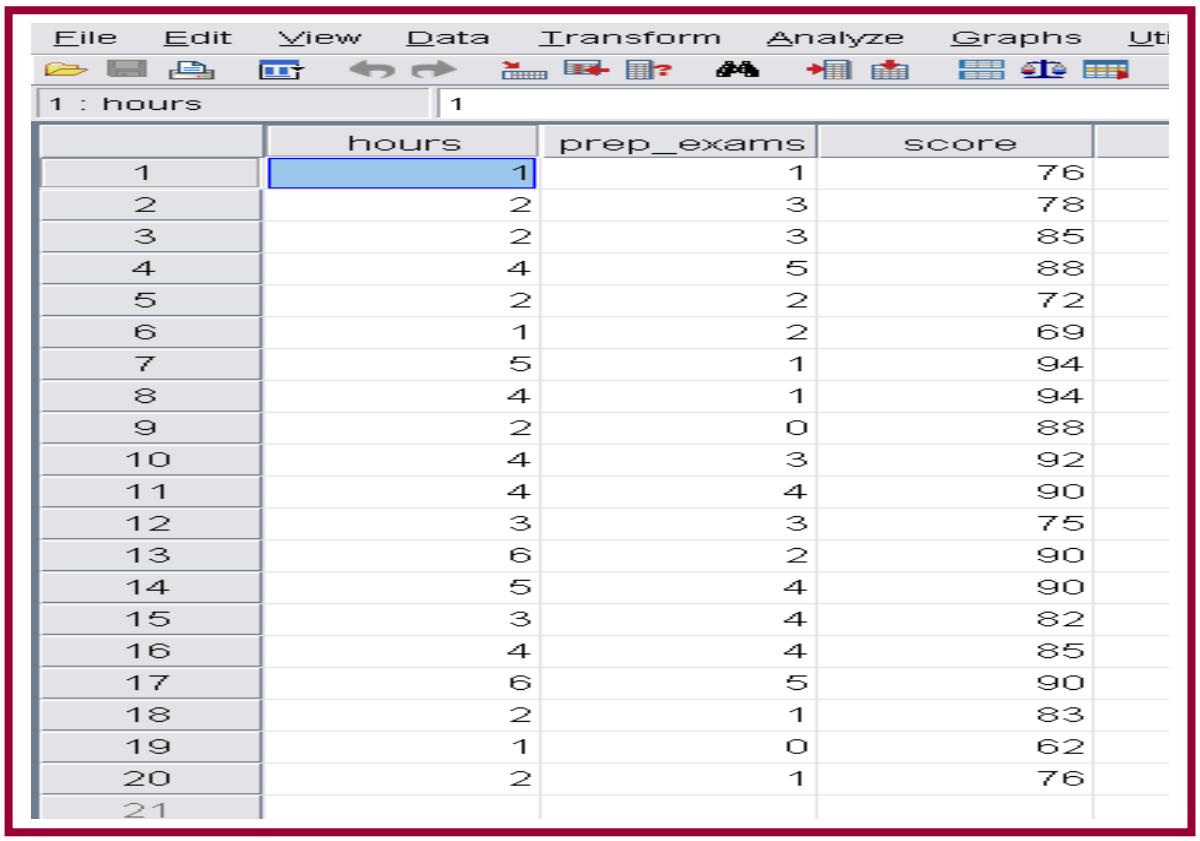
Multiple Regression Workshop #5 -Detecting heteroscedasticity using Breusch-Pagan Procedure

ตารางด้านล่างแสดงข้อมูลจำนวนชั่วโมงที่ใช้ในการศึกษาทบทวนเพื่อเตรียมสอบ(hours) จำนวนครั้งที่เข้าร่วมการทดสอบย่อยก่อนจะมีการสอบปลายภาค(prep exams) และคะแนนผลทดสอบปลายภาคของนักศึกษา(final exam score) ในที่นี้สมการถดถอยพหุคูณที่เราต้องการอยู่ในรูป

Final exam score = $\beta_0 + \beta_1$ (hours) + β_2 (prep exams)



1 : hours | 1

	hours	prep_exams	score
1	1	1	76
2	2	3	78
3	2	3	85
4	4	5	88
5	2	2	72
6	1	2	69
7	5	1	94
8	4	1	94
9	2	0	88
10	4	3	92
11	4	4	90
12	3	3	75
13	6	2	90
14	5	4	90
15	3	4	82
16	4	4	85
17	6	5	90
18	2	1	83
19	1	0	62
20	2	1	76
21			

เนื่องจากการหาสมการถดถอยพหุคูณจำเป็นต้องคำนึงถึงปัญหา heteroscedasticity คือ ปัญหาที่variance ของ error term มีค่าไม่สม่ำเสมอตลอดค่าของตัวแปรอิสระ

ในเอกสารวิชาการนี้เราจะใช้ Breusch-Pagan Test procedure ในการตรวจสอบการมีอยู่/หรือไม่มีของ heteroscedasticity โดยมีสมมติฐานทางสถิติดังต่อไปนี้

H_O : Var(ε_i) ของแบบจำลองการถดถอยมีค่าคงที่ หรืออีกนัยหนึ่ง error terms มีลักษณะ homoscedastic

H_a : Var(ε_i) ของแบบจำลองการถดถอยขึ้นอยู่กับตัวแปรอิสระ หรืออีกนัยหนึ่ง error terms มีลักษณะ heteroscedastic

Breusch-Pagan test Procedure:

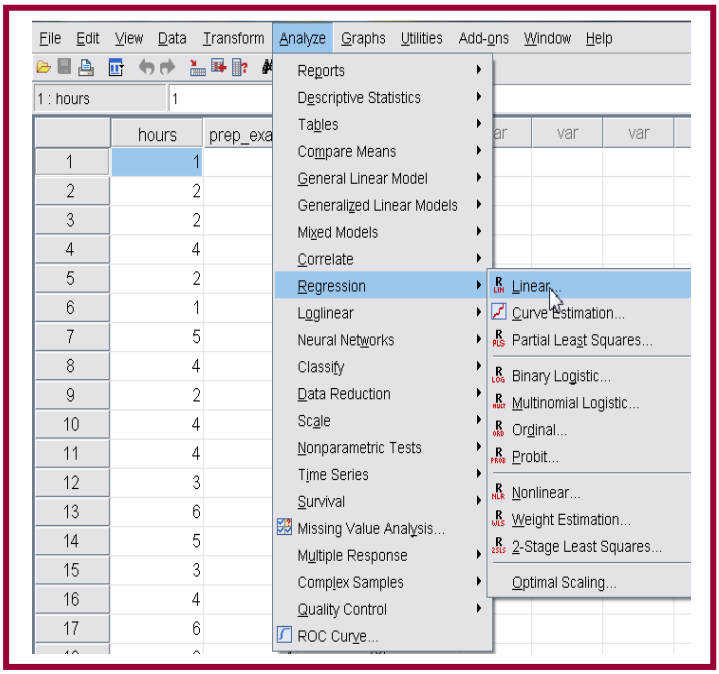
1.จากข้อมูลให้หาสมการถดถอยและ error terms(residuals)

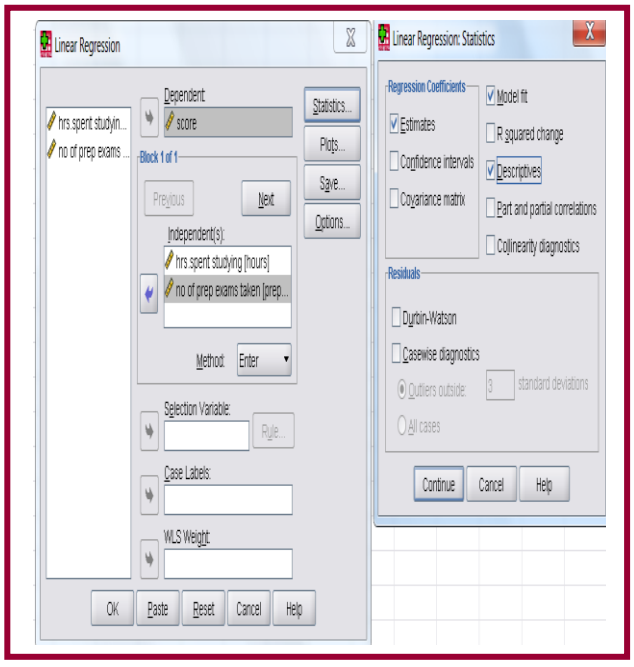
2.หาค่า error terms ยกกำลังสองโดยถือเป็นตัวแปรตัวหนึ่งที่สร้างขึ้น

3.หาสมการถดถอยโดยใช้ ค่าของตัวแปรตาม 2 เป็นตัวแปรตามส่วนตัวแปรอิสระได้แก่ตัวแปรอิสระเดิม

4.ดำเนินการทดสอบสถิติไคว์สแควร์ โดยใช้สูตร χ^2 = n. R^2ที่ได้จาก 3) ที่ degree of freedom=k โดย k คือจำนวนของตัวแปรอิสระ จากนั้นนำไปเปรียบเทียบกับ critical χ^2 (χ^2_c) โดยจะปฏิเสธ H_o หากค่า χ^2 ที่คำนวณได้มีค่าสูงกว่าχ^2_c

◉คลิ๊กAnalyze> Linear จะปรากฏ submenu-linear regressionดังภาพด้านขวาล่างต่อไปนี้





◉คลิ๊กเลือก score ให้ไปอยู่ในกล่อง Dependent และคลิ๊กเลือก hrs.spent studying และ no.of prep-exams ไปอยู่ในกล่อง Independent(s)

◉คลิ๊กที่ปุ่ม Statistics จะปรากฏ submenu ในส่วน Regression Coefficients ให้เลือก Estimates, Model fit และ Descriptivesจากนั้นคลิ๊กContinue คลิ๊กOK

ภาพด้านล่างแสดง computer output

Descriptive Statistics

	Mean	Std. Deviation	N
score	82.95	8.953	20
hrs.spent studying	3.15	1.599	20
no of prep exams taken	2.45	1.572	20

Correlations

		score	hrs.spent studying	no of prep exams taken
Pearson Correlation	score	1.000	.780	.331
	hrs.spent studying	.780	1.000	.516
	no of prep exams taken	.331	.516	1.000
Sig. (1-tailed)	score	.	.000	.077
	hrs.spent studying	.000	.	.010
	no of prep exams taken	.077	.010	.
N	score	20	20	20
	hrs.spent studying	20	20	20
	no of prep exams taken	20	20	20

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.785[a]	.616	.571	5.867

a. Predictors: (Constant), no of prep exams taken, hrs.spent studying

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	937.813	2	468.907	13.623	.000[a]
	Residual	585.137	17	34.420		
	Total	1522.950	19			

a. Predictors: (Constant), no of prep exams taken, hrs.spent studying

b. Dependent Variable: score

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	69.661	3.079		22.626	.000
	hrs.spent studying	4.654	.983	.831	4.733	.000
	no of prep exams taken	-.559	1.000	-.098	-.560	.583

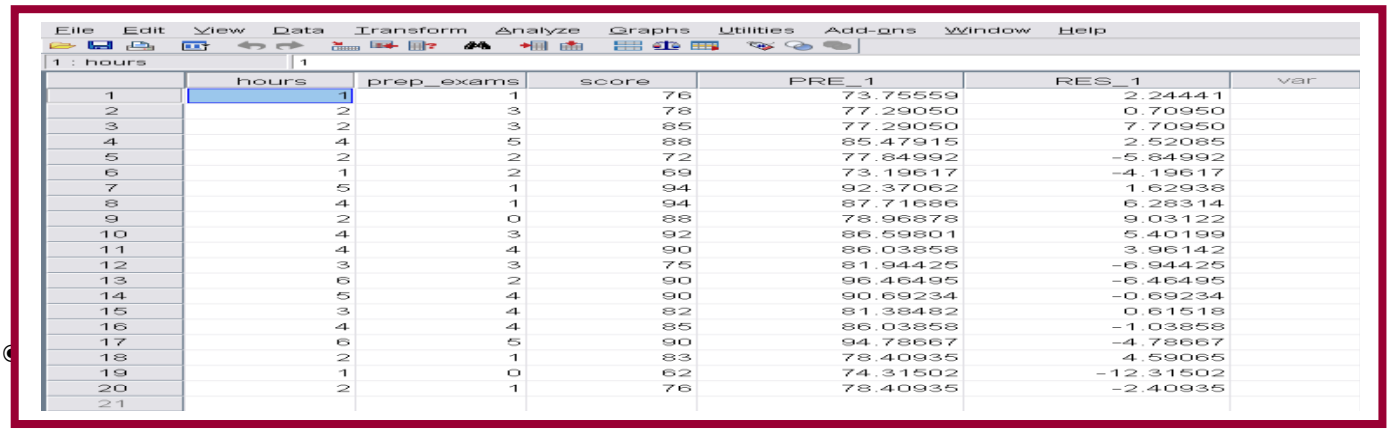
a. Dependent Variable: score

Pearson Correlation ระหว่าง score และ hrs. spent studying มีค่าเท่ากับ .780 และมีนัยสำคัญ ส่วนPearson Correlation ระหว่าง score และ no.of prep exams มีค่าเท่ากับ .331 อยู่ในระดับที่ไม่มีนัยสำคัญ แต่ข้อมูลพื้นฐานเหล่านี้ไม่แสดงให้เห็นปัญหา heteroscedasticity ที่อาจมีอยู่

R-square มีค่าเท่ากับ 0.618 แสดงว่าความผันผวนในตัวแปรอิสระช่วยอธิบายความผันผวนในตัวแปรตามได้ร้อยละ 61.8

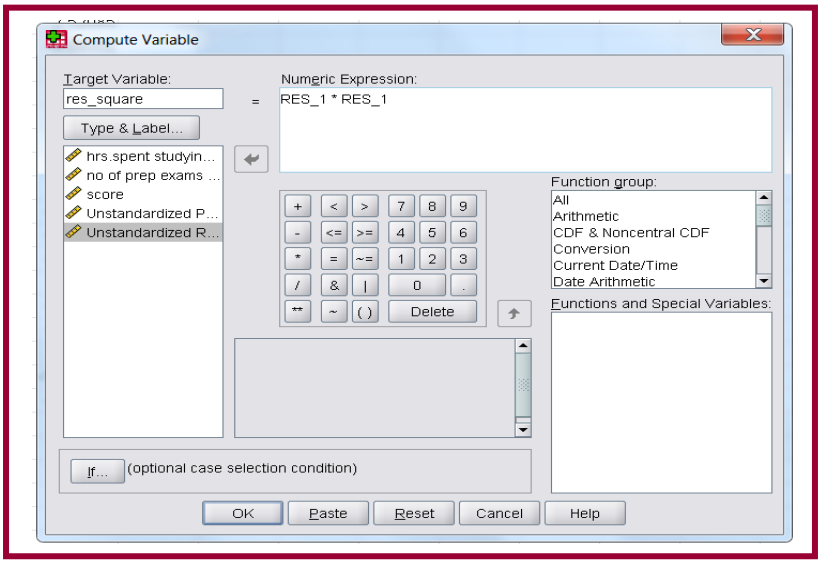
พิจารณาจาก F-test พบว่าสมการถดถอยพหุคูณที่ได้มีคุณค่าในการพยากรณ์ และรูปแบบสมการถดถอยพหุคูณก็คือ score = 69.661+ 4.654(hrs.spent studying) -.559(no.of prep exams taken) ซึ่งไม่มีข้อมูลใดๆที่จะบ่งบอกถึงปัญหา heteroscedasticity

◉คลิ๊กที่ปุ่ม Save จะปรากฏ submenu ในส่วน Predicted Values ให้เลือก Unstandardized ในส่วน Residuals ให้เลือก Unstandardized เช่นกัน จากนั้นคลิ๊กContinue คลิ๊กOKจะปรากฏ column Pre_1 และ RES_1 ขึ้นใน data view



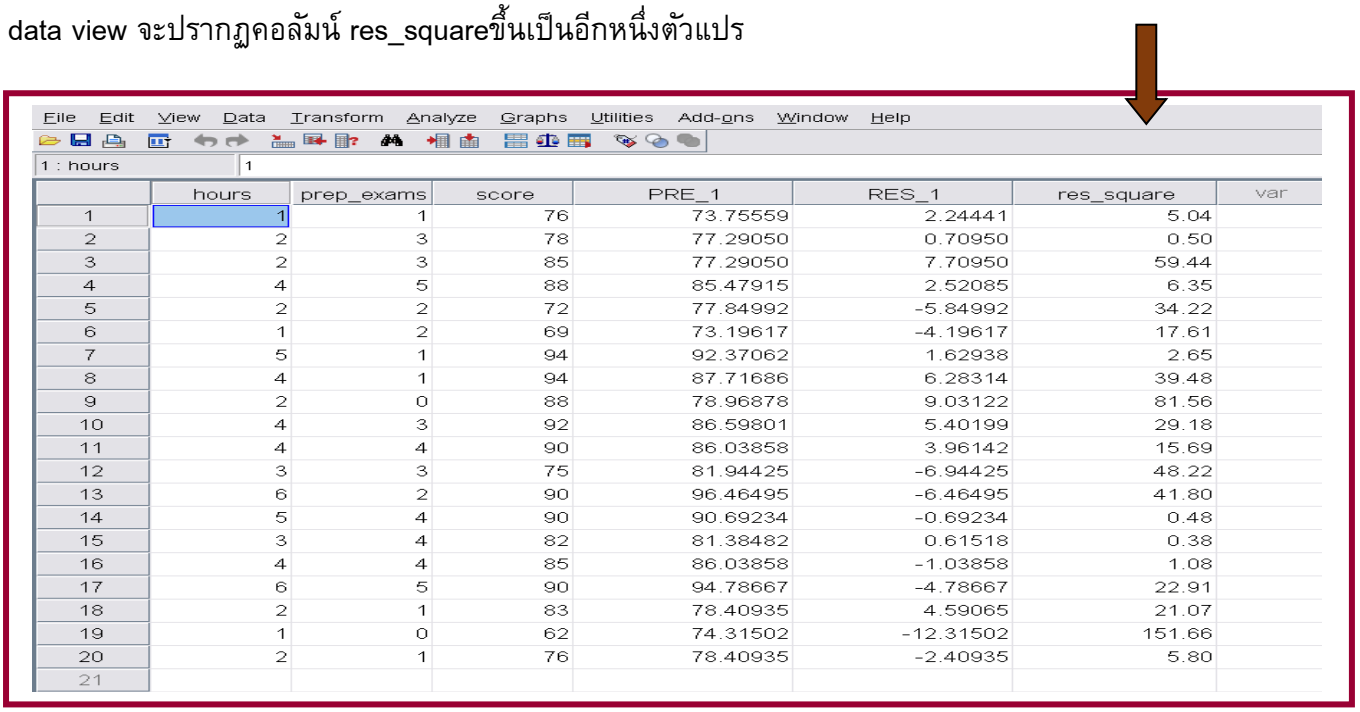
	hours	prep_exams	score	PRE_1	RES_1	var
1	1	1	76	73.75559	2.24441	
2	2	3	78	77.29050	0.70950	
3	2	3	85	77.29050	7.70950	
4	4	5	88	85.47915	2.52085	
5	2	2	72	77.84992	-5.84992	
6	1	2	69	73.19617	-4.19617	
7	5	1	94	92.37062	1.62938	
8	4	1	94	87.71686	6.28314	
9	2	0	88	78.96878	9.03122	
10	4	3	92	86.59801	5.40199	
11	4	4	90	86.03858	3.96142	
12	3	3	75	81.94425	-6.94425	
13	6	2	90	96.46495	-6.46495	
14	5	4	90	90.69234	-0.69234	
15	3	4	82	81.38482	0.61518	
16	4	4	85	86.03858	-1.03858	
17	6	5	90	94.78667	-4.78667	
18	2	1	83	78.40935	4.59065	
19	1	0	62	74.31502	-12.31502	
20	2	1	76	78.40935	-2.40935	
21						

◉



เราจะสร้างตัวแปรขึ้นใหม่ให้ชื่อว่า res_squareซึ่งได้จากการนำ RES_1 คูณกันสองครั้ง จากนั้นคลิ๊กOK

ใน data view จะปรากฏคอลัมน์ res_squareขึ้นเป็นอีกหนึ่งตัวแปร

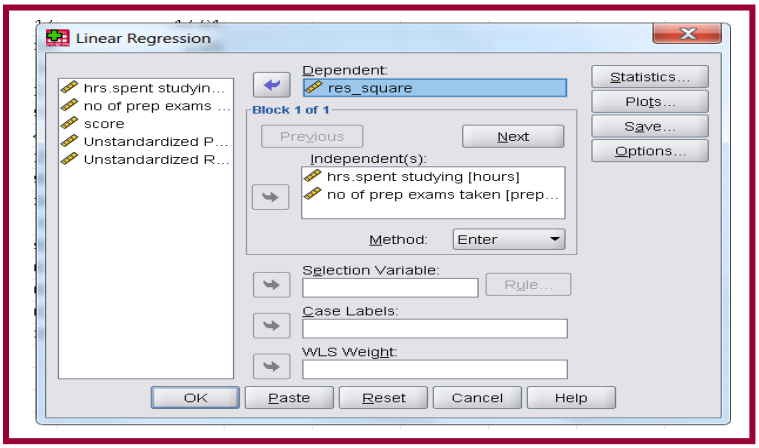


File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : hours 1

	hours	prep_exams	score	PRE_1	RES_1	res_square	var
1	1	1	76	73.75559	2.24441	5.04	
2	2	3	78	77.29050	0.70950	0.50	
3	2	3	85	77.29050	7.70950	59.44	
4	4	5	88	85.47915	2.52085	6.35	
5	2	2	72	77.84992	-5.84992	34.22	
6	1	2	69	73.19617	-4.19617	17.61	
7	5	1	94	92.37062	1.62938	2.65	
8	4	1	94	87.71686	6.28314	39.48	
9	2	0	88	78.96878	9.03122	81.56	
10	4	3	92	86.59801	5.40199	29.18	
11	4	4	90	86.03858	3.96142	15.69	
12	3	3	75	81.94425	-6.94425	48.22	
13	6	2	90	96.46495	-6.46495	41.80	
14	5	4	90	90.69234	-0.69234	0.48	
15	3	4	82	81.38482	0.61518	0.38	
16	4	4	85	86.03858	-1.03858	1.08	
17	6	5	90	94.78667	-4.78667	22.91	
18	2	1	83	78.40935	4.59065	21.07	
19	1	0	62	74.31502	-12.31502	151.66	
20	2	1	76	78.40935	-2.40935	5.80	
21							

◉คลิ๊กAnalyze>regressionจะปรากฏ submenu ขึ้น



คลิ๊กเลือก res_squareให้ไปอยู่ในช่อง Dependent คลิ๊กเลือก hrs.spent studying และ no.of prep exams ไปอยู่ในช่อง Independent(s) คลิ๊กOk

ต่อไปนี้เป็น computer output

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	no of prep exams taken, hrs.spent studying[a]	.	Enter

a. All requested variables entered.
b. Dependent Variable: res_squared

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.502[a]	.252	.164	33.39438

a. Predictors: (Constant), no of prep exams taken, hrs.spent studying
b. Dependent Variable: res_squared

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6395.629	2	3197.814	2.868	.085[a]
	Residual	18958.141	17	1115.185		
	Total	25353.770	19			

a. Predictors: (Constant), no of prep exams taken, hrs.spent studying
b. Dependent Variable: res_squared

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	60.832	17.525		3.471	.003
	hrs.spent studying	-1.727	5.596	-.076	-.309	.761
	no of prep exams taken	-10.667	5.691	-.459	-1.874	.078

a. Dependent Variable: res_squared

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-2.8673	59.1052	29.2568	18.34699	20
Residual	-4.3400E1	92.55453	.00000	31.58792	20
Std. Predicted Value	-1.751	1.627	.000	1.000	20
Std. Residual	-1.300	2.772	.000	.946	20

a. Dependent Variable: res_squared

ค่าสถิติไคว์สแควร์ตามB-P procedure สามารถคำนวณได้โดยใช้สูตร

$$\text{B-P statistic} = n^*R^2$$

$= 20^* \ 0.252$

$= 5.04$

ที่ α=0.05 χ^2_C =5.991

$\therefore$ ยอมรับ H_O นั่นก็คือ Variance ของ error terms มีค่าคงที่ หรือ error terms มีลักษณะ homoscedastic

กล่าวโดยสรุป B-P Procedure ไม่สามารถ detect ว่า error terms มีลักษณะ heteroscedastic

Contribution this issue: ดร. ดนัย ปัตตพงศ์*

อยากเรียนรู้การนำสถิติข้างต้นนี้ไปใช้ในการวิจัยระดับสารนิพนธ์ (independent study)
วิทยานิพนธ์ (thesis) ดุษฎีนิพนธ์(dissertation) ปรึกษาได้ที่ dpattaphongse@gmail.com

*ผู้แต่ง MBA's Made Easy(160+ issues) เอกสารวิชาการด้านศาสตร์การบริหารธุรกิจที่ช่วยให้ธุรกิจสามารถยืนหยัดและอยู่รอดได้ในภาวะที่โลกเปลี่ยนแปลงอยู่ตลอดเวลา

*ผู้พัฒนา FINALYSIS… a dedicated software สำหรับให้บริการนักธุรกิจที่ต้องการวิเคราะห์ความเป็นไปได้ทางการเงินของโครงการพัฒนาอสังหาริมทรัพย์(บ้านจัดสรร/จัดสรรที่ดินเพื่อการอุตสาหกรรม/อาคารชุด/อาคารสำนักงานให้เช่า) โรงแรม โรงพยาบาลเอกชน ห้างสรรพสินค้า โรงงานน้ำตาล โรงงานกระดาษ โรงไฟฟ้าชีวมวล ฯลฯได้เห็นตัวเลขก่อนโครงการเกิด หลีกเลี่ยงความผิดพลาดเป็นร้อยเป็นพันล้านหากเกิดการลงทุนจริง(กำหนด DEBUT 1 เมษายน 2569)

*ผู้แต่งหนังสือ"การวิเคราะห์ความเป็นไปได้ทางการเงินและการจัดวงเงินเครดิตของโครงการลงทุน"ประกอบด้วยตัวอย่างของธุรกิจจริงที่ไม่เปิดเผยชื่อนับ 100 บริษัท ครอบคลุมอุตสาหกรรม 24 อุตสาหกรรม

*Co-developer ซอฟแวร์en◉gex◉cel®สำหรับใช้ทดสอบ/เรียนรู้ศัพท์(ประกอบด้วยแบบฝึกหัดและเฉลยกว่า 90 บทครอบคลุมศัพท์ระดับ SAT/IELTS/TOEFL กว่า 12,000 คำ) และไวยากรณ์อังกฤษ (ประกอบด้วยแบบฝึกหัดและเฉลยกว่า 160 บทหรือกว่า 10,000 ข้อครอบคลุมเนื้อหาระดับอุดมศึกษาและTOEFL) มาพร้อมกับไฟล์เสียง/ไฟล์ข้อมูล/ฯลฯอีกมาก(กำหนด DEBUT 1 เมษายน 2569)