



รายงาน

การวิเคราะห์เปอร์เซ็นต์ไขมันในร่างกาย

โดย

นายคุณานนท์	ศรีสันติโรจน์	59070022
นางสาวธัญพิชชา	อุบลวิรัตน์	59070078
นายพรพรม	เขี้ยวจักร์	59070113
นางสาวเพ็ญวิสา	โลหิรัญญานนท์	59070124
นางสาวรวิภา	ตั้งคำ	59070145

เสนอ

ดร.บัณฑิต ฐานะโสภณ

รายงานนี้เป็นส่วนหนึ่งของวิชา 06016239 DATA ANALYSIS FOR BUSINESS

ภาคเรียนที่ 1 ปีการศึกษา 2561

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

คำนำ

รายงานเล่มนี้จัดทำขึ้นเพื่อศึกษาวิเคราะห์เปอร์เซ็นต์ไขมันในร่างกายและศึกษาขั้นตอนกระบวนการในการวิเคราะห์ข้อมูล ทางคณะผู้จัดทำจึงเห็นความสำคัญและต้องการที่จะศึกษาข้อมูล ด้วยกระบวนการการวิเคราะห์เพื่อให้ได้ประสิทธิภาพจากการวิเคราะห์มากที่สุด และสามารถใช้ประโยชน์จากผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูล

ทางคณะผู้จัดทำหวังว่ารายงานฉบับนี้จะเป็นความรู้ต่อผู้ที่ได้ศึกษาและนำไปใช้ประโยชน์ต่อไปไม่มากนักขอ

คณะผู้จัดทำ

เรื่อง	สารบัญ	หน้า
คำนำ		ก
สารบัญ		ข
1. ทฤษฎีที่เกี่ยวข้องและสมมติฐาน		1
2. การออกแบบการทดลองและการเก็บข้อมูล		1
2.1 การออกแบบการทดลอง		1
2.2 การเก็บข้อมูล		2
3. ผลการวิเคราะห์ข้อมูล		3
4. สรุปผลการทดลอง		5
ภาคผนวก		7

ทฤษฎีที่เกี่ยวข้องและสมมติฐาน

ในปัจจุบันนี้คนในสังคมส่วนใหญ่รวมทั้งภายในคณะเทคโนโลยีและสารสนเทศเริ่มที่จะรักสุขภาพกันมากขึ้น มีการจับกลุ่มวิ่งกันเป็นประจำสัปดาห์และเล่นกีฬาในโรงยิมหรือกลางแจ้ง เราจึงเห็นว่าประโยชน์ที่จะสร้างสมการเพื่อทำนายว่า “ร่างกายของเรามีเปอร์เซ็นต์ไขมันเป็นเท่าไร” เพื่อที่จะให้คนที่ต้องการลดปริมาณไขมันในร่างกายได้ทราบว่าตัวเองมีไขมันในร่างกายเท่าไรก่อนที่จะออกกำลังกาย และเมื่อออกกำลังกายไปสักระยะมีไขมันในร่างกายเปลี่ยนแปลงมาน้อยเพียงใด เพื่อที่จะสังเกตการณ์และปรับปรุงวิธีออกกำลังกายได้อย่างเหมาะสม

จากที่กล่าวว่านั้นทางผู้จัดทำจึงได้พิจารณาจากปัจจัยต่าง ๆ เหล่านี้เพื่อมาสร้างสมการหรือโมเดลในการทำนาย

- ส่วนสูง (เซนติเมตร)
- ความยาวรอบส่วนต่าง ๆ ของร่างกาย (เซนติเมตร)
- น้ำหนัก (กิโลกรัม)

โดยมีการตั้งสมมติฐานดังนี้คือ “ส่วนสูง ความยาวรอบข้อมือ ความยาวรอบต้นขา ความยาวรอบคอ และ น้ำหนัก จะมีความสัมพันธ์กับเปอร์เซ็นต์ไขมันในร่างกาย”

การออกแบบการทดลองและการเก็บข้อมูล

การทดลองนี้เราได้ใช้การวิเคราะห์การถดถอยพหุคูณ (Multiple Regression) เป็นวิธีการทางสถิติที่ใช้เรื่องความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตามที่มีความสัมพันธ์เชิงเส้นตรง โดยที่มีตัวแปรต้น (Independent Variable) มากกว่าหนึ่งตัวโดยที่สามารถเป็นระดับการวัดแบบนามบัญญัติที่มีสองหมวดหมู่ขึ้นไป (Dichotomous) กับตัวแปรตาม (Dependent Variable) หนึ่งตัวที่เป็นระดับการวัดแบบอันตรภาคขึ้นไป (Interval Scale) ซึ่งการถดถอยพหุคูณนั้นมีข้อตกลงเบื้องต้น (Assumption) ในการวิเคราะห์ 5 อย่างหลัก ๆ ได้แก่

- 1.) ตัวแปรอิสระแต่ละตัวมีความสัมพันธ์เชิงเส้นตรงกับตัวแปรตาม (Linearity)
- 2.) ตัวแปรอิสระไม่ควรมีความสัมพันธ์กันเพราะจะทำให้เกิดพหุสหสัมพันธ์ขึ้น (Multicollinearity) ซึ่งทำให้ตัวแปรอิสระมีความสัมพันธ์ (Correlation) กันมากและทำให้มีผลกระทบต่ค่าสัมประสิทธิ์การตัดสินใจ (R Squared) สูงเกินความเป็นจริง
- 3.) การแจกแจงของตัวแปรตามเป็นแบบโค้งปกติ (Normal Distribution / Normally Distributed) ที่ทุกค่าของตัวแปรต้น
- 4.) ค่าของตัวแปรตามมีความแปรปรวนเท่ากันในทุกค่าของตัวแปรต้น (Independent Errors)
- 5.) ความแปรปรวนของค่าความคลาดเคลื่อนจากการพยากรณ์ (Residual) ทุกจุดบนเส้นถดถอยมีค่าเท่ากัน (Homoscedasticity)

และยังคงมีสมมติฐานที่ 2 อย่างดังนี้

- 1.) ค่าสัมประสิทธิ์แต่ละตัวต่างจาก 0 อย่างมีนัยสำคัญหรือไม่ โดยใช้สถิติทดสอบ

$$H_0 : \beta_j = 0, j = 1, 2, \dots, k$$
$$H_a : \beta_j \neq 0, j = 1, 2, \dots, k$$

- 2.) โมเดลการถดถอยพหุคูณสามารถอธิบายค่าที่ได้แตกต่างจากโมเดลพื้นฐาน กล่าวคือสามารถทำนายการเปลี่ยนแปลงของตัวแปรตามอย่างมีนัยสำคัญหรือไม่ โดยใช้สถิติทดสอบ

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$
$$H_a : \text{อย่างน้อย 1 พารามิเตอร์มีค่าไม่เท่ากับ 0 ; } \beta_j \neq 0, j = 1, 2, \dots, k$$

โดยที่สมการการถดถอยพหุคูณมีลักษณะเป็นดังนี้

$$\gamma = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

และเราพยายามที่จะจำลองโมเดลดังนี้โดยมีสมการดังนี้

$$\gamma = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

โดยมีตัวแปรต้นและตัวแปรตามดังต่อไปนี้

- ตัวแปรต้น
 - ส่วนสูง (Height_CM)
 - ความยาวรอบรอบข้อมือ (Wrist)
 - ความยาวรอบรอบต้นขา (Thigh)
 - ความยาวรอบรอบคอ (Neck)
 - น้ำหนัก (Weight_KG)
- ตัวแปรตาม
 - เปอร์เซนต์ไขมันในร่างกาย (bodyfat)

การเก็บรวบรวมข้อมูลที่ใช้ในการสร้างสมการนั้นได้เลือกใช้วิธีการเก็บแบบทุติยภูมิโดยการใช้ข้อมูลจากเว็บไซต์ <http://staff.pubhealth.ku.dk> ซึ่งจำนวนตัวอย่างทั้งหมดมี 252 ตัวอย่างและมีลักษณะของข้อมูลดังนี้

- Density determined from underwater weighing
- Percent body fat from Siri's (1956) equation
- Age (years)
- Weight (lbs)
- Height (inches)
- Neck circumference (cm)
- Chest circumference (cm)
- Abdomen 2 circumference (cm)
- Hip circumference (cm)
- Thigh circumference (cm)
- Knee circumference (cm)
- Ankle circumference (cm)
- Biceps (extended) circumference (cm)
- Forearm circumference (cm)
- Wrist circumference (cm)

โดยที่เราสร้างตัวแปรใหม่ขึ้นมาจากตัวแปรเดิมมา 2 ตัวแปรคือ น้ำหนักที่หน่วยเป็นกิโลกรัม (Weight_KG) และส่วนสูงที่หน่วยเป็นเซนติเมตร (Height_CM) โดยใช้ตัวแปร Weight (lbs) และตัวแปร Height (inches)

ผลการวิเคราะห์ข้อมูล

1.) ขั้นตอนลดความลำเอียงของข้อมูล

Casewise Diagnostics^a

Case Number	Std. Residual	bodyfat	Predicted Value	Residual
39	-3.416	35.2	56.108	-20.9080
42	-3.344	32.9	53.370	-20.4700

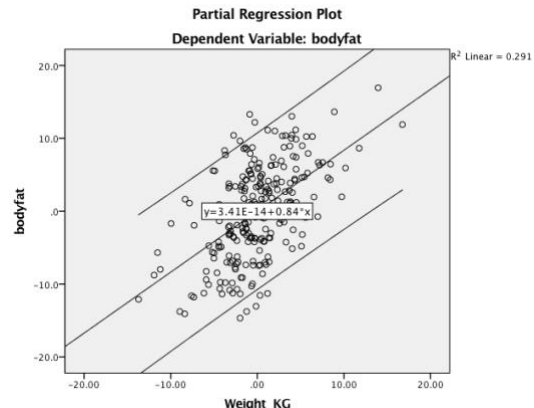
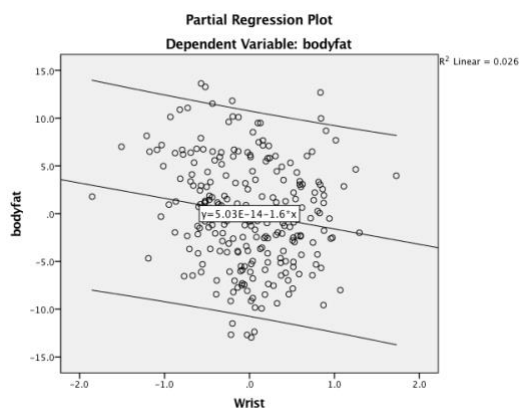
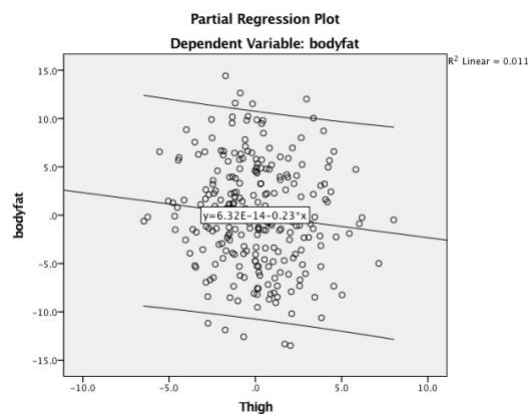
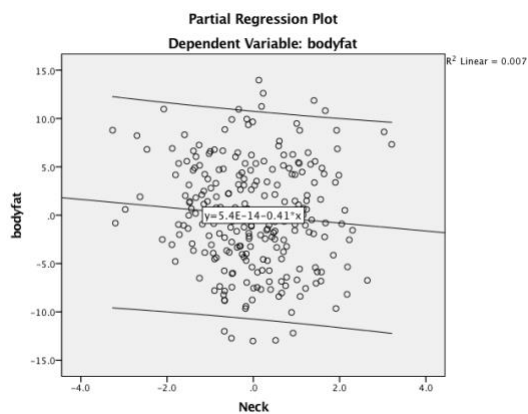
a. Dependent Variable: bodyfat

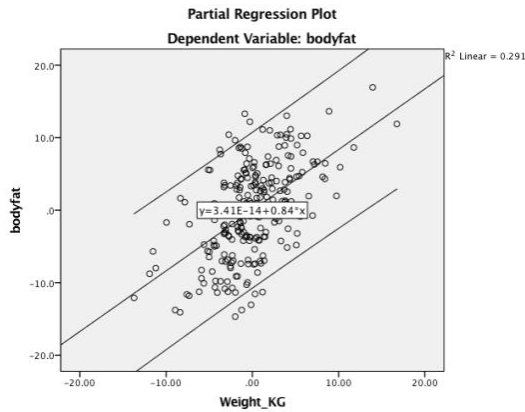
จากตาราง Casewise Diagnostics ตามภาพด้านบนนั้นคือผลลัพธ์ที่ SPSS แสดงมาให้ว่า Case Number 39 และ 42 เป็น Extreme Case เพราะมีค่า Standardize Residual ที่มีขนาดมากกว่า 3 S.D. จึงลบออกไปจากชุดข้อมูล

นอกจากนี้ยังทำการตัดข้อมูลแถวที่มี Cook's distance มากกว่า 1 หรือ Centered Leverage Value มากกว่า 3 เท่าของ $(k + 1 / n)$ หรือ Mahalanobis distance เกิน 15 หรือกล่าวคือเป็นข้อมูลที่น่าจะเป็น Influential Point

2.) ตรวจสอบสมมติฐาน

2.1) Linearity





จาก Partial Regression Plot ของแต่ละตัวแปรต้นเทียบกับตัวแปรตาม แสดงความสัมพันธ์กันเป็นเส้นตรงถือว่ายังคงผ่านสมมติฐาน

2.2) Multicollinearity

Correlations

	bodyfat	Neck	Thigh	Wrist	Weight_KG	Height_CM
Pearson Correlation	bodyfat	.487	.541	.324	.610	-.025
	Neck		.663	.747	.815	.332
	Thigh			.530	.847	.360
	Wrist				.715	.419
	Weight_KG					.531
	Height_CM					
Sig. (1-tailed)	bodyfat	.000	.000	.000	.000	.345

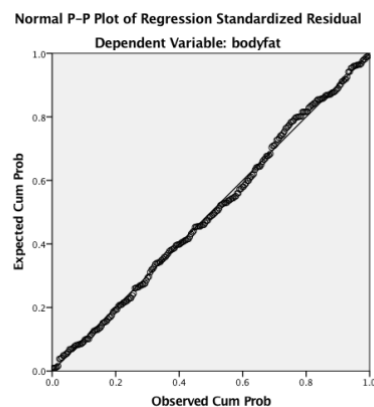
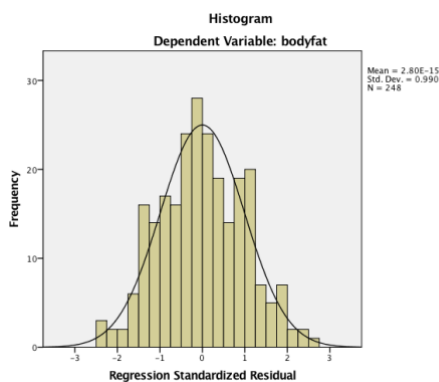
Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients		Sig.	95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta	t		Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	123.517	16.221		7.614	.000	91.564	155.470					
	Neck	-.408	.306	-.110	-1.333	.184	-1.012	.195	.487	-.085	-.056	.261	3.831
	Thigh	-.233	.141	-.137	-1.646	.101	-.511	.046	.541	-.105	-.070	.260	3.853
	Wrist	-1.605	.627	-.174	-2.559	.011	-2.840	-.369	.324	-.162	-.108	.388	2.576
	Weight_KG	.836	.084	1.210	9.971	.000	.671	1.002	.610	.540	.421	.121	8.252
	Height_CM	-.633	.065	-.510	-9.663	.000	-.761	-.504	-.025	-.528	-.408	.642	1.557

a. Dependent Variable: bodyfat

จากตารางดังกล่าวนี้จะเห็นว่าจุดขีดเส้นใต้ไว้เป็นลักษณะของปัญหา Multicollinearity กล่าวคือมีค่าความสัมพันธ์ (Correlation) ระหว่างตัวแปรต้นด้วยกันมากกว่า 0.8 หรือมีค่า VIF ที่มากกว่า 5 และมี Tolerance น้อยกว่า 0.2 ซึ่งตัวแปรต้น Weight_KG ไม่ผ่านสมมติฐานดังกล่าว สรุปว่าสมมติฐาน Multicollinearity ไม่ผ่าน

2.3) Normal Distribution



จากกราฟการกระจายตัวของ Residual และ P-P Plot นั้นจะเห็นได้ว่าเป็นไปตามสมมติฐานทั้งคู่

2.4) Independent of Errors

Model Summary^b

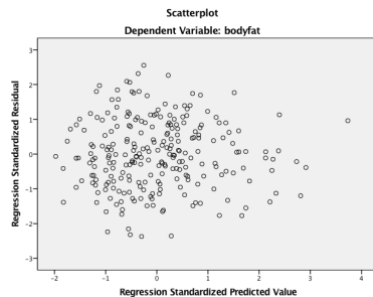
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.754 ^a	.568	.559	5.4915	.568	63.600	5	242	.000	1.565

a. Predictors: (Constant), Height_CM, Neck, Thigh, Wrist, Weight_KG

b. Dependent Variable: bodyfat

จากตารางด้านล่างค่า Durbin-Watson ยังอยู่ในช่วง 1.5 – 2.5 ซึ่งรับได้ว่าสมมติฐานนี้ผ่าน

2.5) Homoscedasticity



จากกราฟนี้เห็นได้ว่าการกระจายแบบ Homoscedasticity เพราะเนื่องจากว่าต่อให้ค่า Predicted Values เปลี่ยนก็ไม่ได้ส่งผลให้ Residual มีแนวโน้มและลักษณะการกระจายตัวที่เปลี่ยนไปอย่างชัดเจน สมมติฐานสุดท้ายจึงผ่าน

3.) ผลลัพธ์ของโมเดล

ANOVA ^a					
Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	9589.762	5	1917.952	63.600	.000 ^b
Residual	7297.882	242	30.157		
Total	16887.644	247			

a. Dependent Variable: bodyfat

b. Predictors: (Constant), Height_CM, Neck, Thigh, Wrist, Weight_KG

Coefficients ^a										
Model	Unstandardized Coefficients		Standardized Coefficients		t	Sig.	95.0% Confidence Interval for B		Correlations	
	B	Std. Error	Beta				Lower Bound	Upper Bound	Zero-order	Partial
1 (Constant)	123.517	16.221			7.614	.000	91.564	155.470		
Neck	-.408	.306	-.110	-1.333	.184		-1.012	.195	.487	-.056
Thigh	-.233	.141	-.137	-1.646	.101		-.511	.046	.541	-.105
Wrist	-1.605	.627	-.174	-2.559	.011		-2.840	-.369	.324	-.108
Weight_KG	.836	.084	1.210	9.971	.000		.671	1.002	.610	.421
Height_CM	-.633	.065	-.510	-9.663	.000		-.761	-.504	-.025	-.408

a. Dependent Variable: bodyfat

สรุปผลว่า “โมเดลที่ประกอบไปด้วยส่วนสูง ความยาวรอบข้อมือ ความยาวรอบต้นขา ความยาวรอบคอ และ น้ำหนัก สามารถทำนายเปอร์เซ็นต์ไขมันในร่างกายได้อย่างมีนัยสำคัญทางสถิติ (Sig. < 0.05)”

แต่สัมประสิทธิ์ตัวแปรต้น ความยาวรอบต้นขา และความยาวรอบคอไม่เท่ากับ 0 อย่างไม่มีนัยสำคัญทางสถิติ (Sig. > 0.05)

ในขณะที่ตัวแปรต้นตัวที่เหลือไม่เท่ากับ 0 อย่างมีนัยสำคัญทางสถิติ (Sig. < 0.05)

รายการเอกสารอ้างอิง

- <http://staff.pubhealth.ku.dk/~tag/Teaching/share/data/Bodyfat.htm>
- Data Analysis for business story of Regression Analysis by Dr Bundit Thanasopon

ภาคผนวก

1. Influential Point

	WeightKG	HeightCM	PRE_1	RES_1	ZPR_1	ZRE_1	MAH_1	COO_1	LEV_1	DFF_1	SDF_1
1	92.99	74.93	53.41412	-20.51412	6.01154	-3.33694	149.27934	7.01083	.59712	-30.91279	-6.87607
2	164.72	183.52	56.35530	-21.15530	6.52679	-3.44124	46.31360	.56820	.18525	-4.93782	-1.90014
3	75.07	173.99	16.45893	1.54107	-.46239	-.25068	31.64793	.00181	.12659	.23145	.10399
4	119.18	174.63	35.01644	-.51644	2.78857	-.08401	16.33180	.00009	.06533	-.03846	-.02372
5	101.15	177.17	31.40257	3.39743	2.15548	.55265	14.98880	.00371	.05996	.23207	.14908
6	94.12	177.80	22.58468	9.31532	.61074	1.51528	13.61489	.02523	.05446	.57821	.39021
7	85.16	176.53	20.64160	1.45840	.27034	-.23723	13.45266	.00061	.05381	.08946	.06041
8	101.83	173.36	30.04663	4.95337	1.91795	.80574	13.19290	.00690	.05277	.29805	.20338
9	99.34	162.56	32.97612	14.52388	2.43115	2.36254	11.96668	.05365	.04787	.79426	.57316
10	92.19	188.60	21.80065	-3.50065	.47339	-.56944	11.46841	.00298	.04587	-.18369	-.13364
11	103.53	176.53	29.30505	4.99495	1.78803	.81251	11.12816	.00589	.04451	.25459	.18793
12	66.56	174.63	11.90113	-3.10113	-1.26084	-.50445	10.58460	.00216	.04234	-.15063	-.11367
13	94.57	184.79	22.59951	-3.39951	.61333	-.55298	10.28876	.00252	.04116	-.16070	-.12287
14	80.85	177.80	16.37479	5.72521	-.47713	.93129	9.90496	.00689	.03962	.26102	.20330
15	67.13	171.45	15.22962	7.67038	-.67775	1.24771	9.70198	.01212	.03881	.34290	.26999
16	83.57	180.98	19.92709	8.77291	.14517	1.42705	9.60473	.01569	.03842	.38847	.30758
17	83.57	174.63	18.69343	-5.69343	-.07094	-.92613	9.52049	.00655	.03808	-.25002	-.19825
18	101.83	197.49	22.01624	-4.61624	.51116	-.75090	9.37637	.00424	.03751	-.19982	-.15944
19	71.55	171.45	16.23078	9.56922	-.50236	1.55658	9.28906	.01807	.03716	.41057	.33030
20	109.43	181.61	32.51155	-2.61155	2.34976	-.42481	9.02897	.00131	.03612	-.10910	-.08847
21	56.81	172.72	11.27830	-3.57830	-1.36995	-.58207	8.67732	.00236	.03471	-.14403	-.11895
22	60.89	170.18	10.78660	.21340	-1.45609	.03471	8.55080	.00001	.03420	.00847	.00704
23	89.24	173.36	22.43472	5.96528	.58447	.97035	8.51285	.00645	.03405	.23586	.19672
24	98.43	177.80	27.53018	4.06982	1.47711	.66202	8.30707	.00293	.03323	.15730	.13250
25	70.42	176.53	13.21379	-.81379	-1.03089	-.13238	8.26627	.00012	.03307	-.03131	-.02641
26	61.80	171.45	10.71628	-6.81628	-1.46841	-1.10878	8.23214	.00815	.03293	-.26125	-.22132

ภาพที่ 1: ตารางแสดงข้อมูลความผิดปกติของ Influential point

-ค่า Cook's Distance

สถิติชนิดหนึ่งที่ใช้ในการคำนวณค่าผิดปกติของข้อมูลที่มีผลต่อการวิเคราะห์การถดถอย

เป็นสถิติที่ใช้กับค่าของตัวแปรอิสระและค่าของตัวแปรตาม โดยข้อมูลที่มีความผิดปกตินั้นคือข้อมูลที่มีค่า Cook's Distance เกิน 1

จะเห็นว่า มีค่า Cook's Distance ที่เกิน 1

-ค่า Centered Leverage Value ถ้าสถิติ Centered Leverage Value มีค่ามากกว่าค่าที่ได้จากการคำนวณ $[(k+1)/n]$ มากถึง 3

เท่า จะถือว่ามีความผิดปกติของข้อมูล

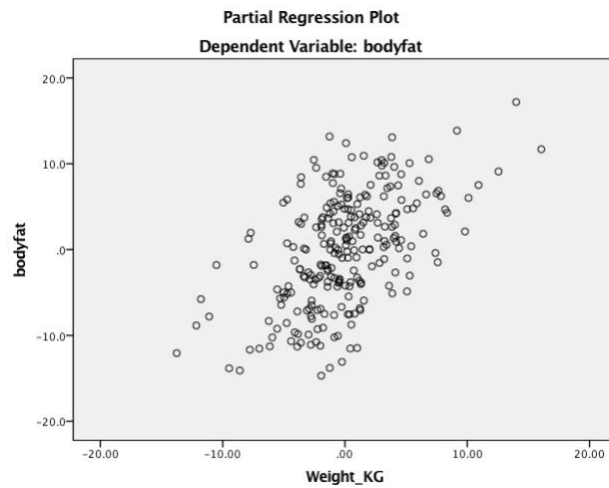
จะเห็นว่า มีค่า Centered Leverage Value ที่มีค่าเกิน 3 เท่า จากค่าปกติ

-ค่า Mahalanobis Distance ถ้าค่าสถิติ Mahalanobis Distance ในข้อมูลที่มีขนาดใหญ่มีค่าเกิน 25 และในข้อมูลขนาดเล็กมีค่า

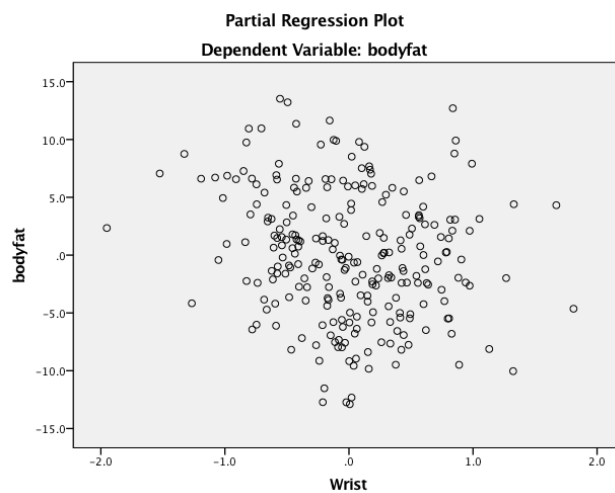
เกิน 15 จะถือว่ามีความผิดปกติของข้อมูล

จะเห็นว่า ข้อมูลที่เก็บมาทั้งหมดมีจำนวน 252 ตัว ซึ่งถือว่าอยู่ในข้อมูลที่มีขนาดเล็ก มีค่าที่เกิน 15

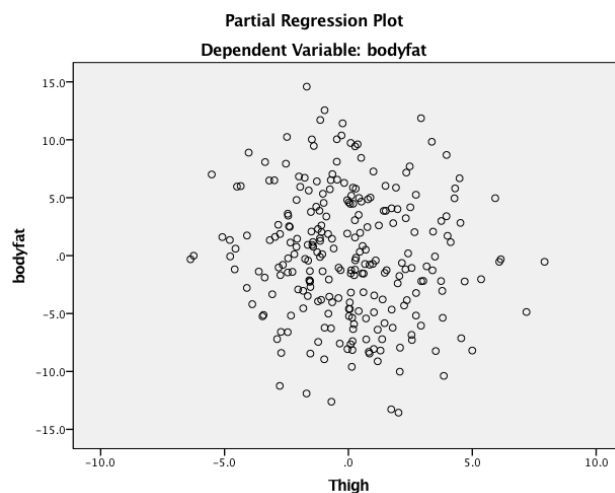
ภาพ Partial Regression Plot (each independent variables with dependent variable) ก่อนที่จะทำการตัด Influential Point



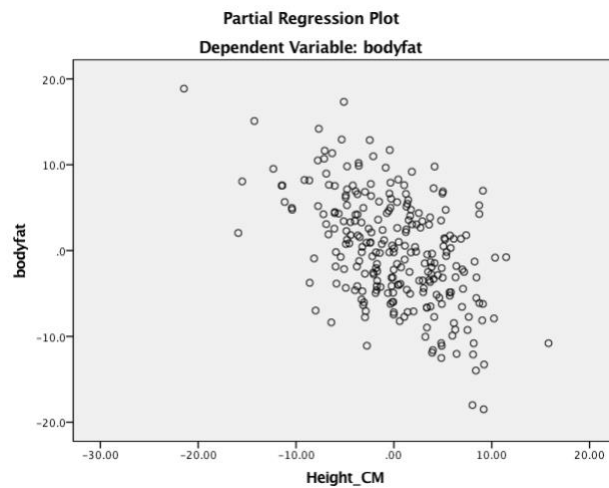
ภาพที่ 2-1: Partial Regression Plot ระหว่างเปอร์เซ็นต์ไขมันในร่างกายและน้ำหนัก



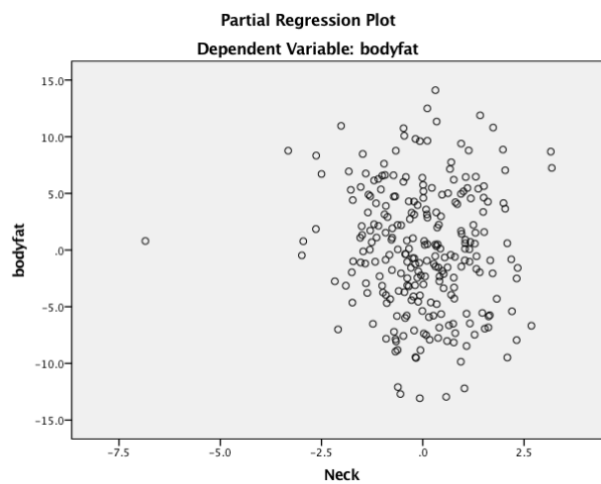
ภาพที่ 2-2: Partial Regression Plot ระหว่าง เปอร์เซ็นต์ไขมันในร่างกายและความยาวรอบรอบข้อมือ



ภาพที่ 2-3: Partial Regression Plot ระหว่าง เปอร์เซ็นต์ไขมันในร่างกายและความยาวรอบรอบต้นขา

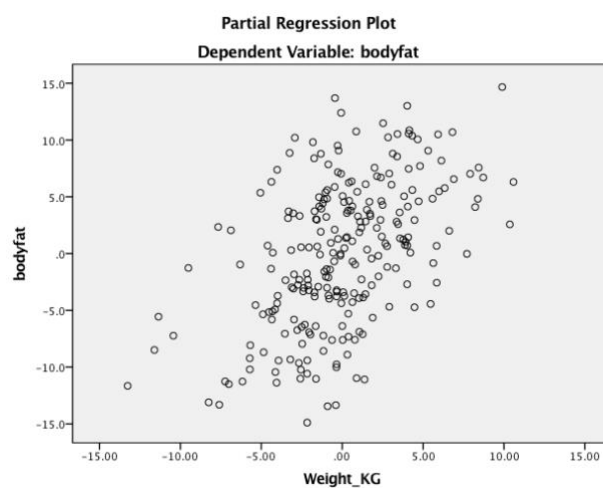


ภาพที่ 2-4: Partial Regression Plot ระหว่าง เปอร์เซ็นต์ไขมันในร่างกายและส่วนสูง

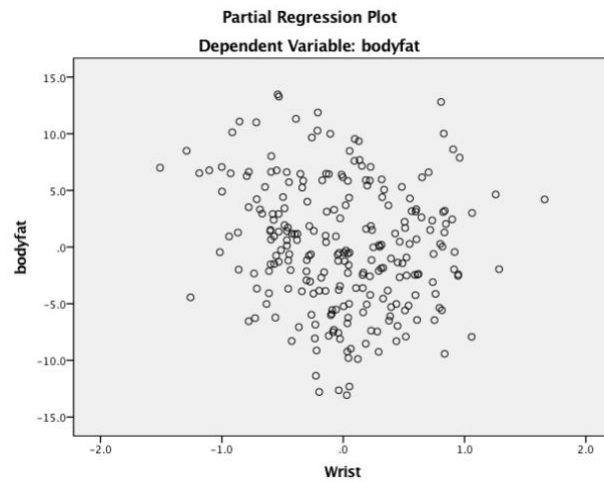


ภาพที่ 2-5: Partial Regression Plot ระหว่าง เปอร์เซ็นต์ไขมันในร่างกายและความยาวรอบคอ

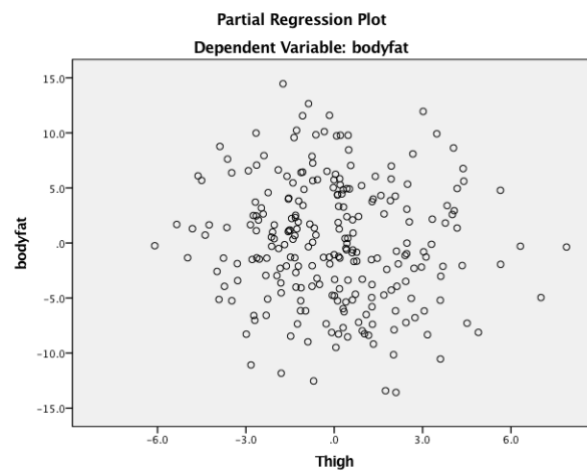
ภาพ Partial Regression Plot (each independent variables with dependent variable) หลังทำการตัด Influential Point



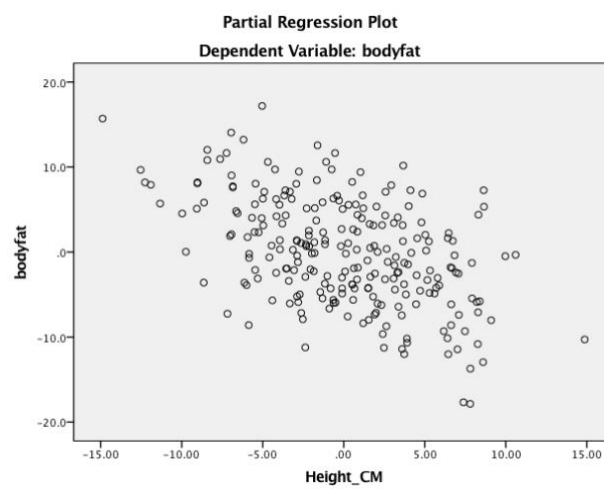
ภาพที่ 2-6: Partial Regression Plot ระหว่างเปอร์เซ็นต์ไขมันในร่างกายและน้ำหนัก



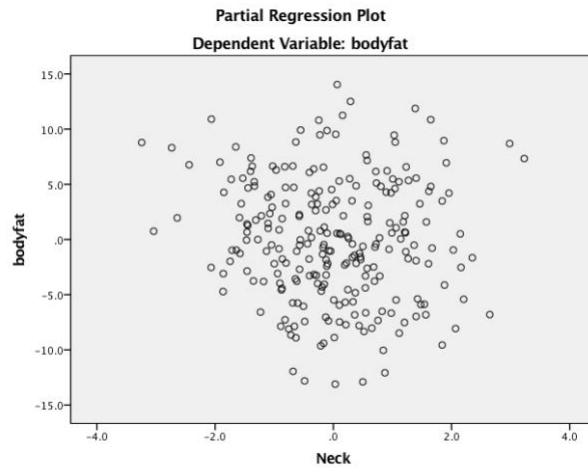
ภาพที่ 2-7: Partial Regression Plot ระหว่าง เปอร์เซ็นต์ไขมันในร่างกายและความยาวรอบรอบข้อมือ



ภาพที่ 2-8: Partial Regression Plot ระหว่าง เปอร์เซ็นต์ไขมันในร่างกายและความยาวรอบรอบต้นขา



ภาพที่ 2-9: Partial Regression Plot ระหว่าง เปอร์เซ็นต์ไขมันในร่างกายและส่วนสูง



ภาพที่ 2-10: Partial Regression Plot ระหว่าง เปอร์เซ็นต์ไขมันในร่างกายและความยาวรอบคอ

2. Multicollinearity

โมเดลที่แก้ไขสมมติฐาน Multicollinearity ด้วยการนำตัวแปรต้น Weight_KG ออกจากโมเดล

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-3.009	11.978		-.251	.802	-26.604	20.586					
	Neck	1.078	.317	.291	3.397	.001	.453	1.702	.487	.213	.170	.342	2.924
	Thigh	.782	.116	.459	6.723	.000	.553	1.011	.541	.396	.337	.539	1.857
	Wrist	-.188	.724	-.020	-.260	.795	-1.615	1.238	.324	-.017	-.013	.409	2.443
	Height_CM	-.346	.070	-.279	-4.963	.000	-.483	-.209	-.025	-.303	-.249	.795	1.257

a. Dependent Variable: bodyfat

ภาพที่ 3 : ตาราง Coefficients

เป็นเครื่องมือที่แสดงปัจจัยหรือผลกระทบของตัวแปรอิสระตัวหนึ่งว่าจะมีอิทธิพลต่อค่าความแปรปรวน (variance) ของสัมประสิทธิ์ของตัวแปรอิสระต่างๆ มากน้อยเพียงใด ซึ่งถ้ามีค่ามากเกินไป 5 จะหมายความว่าตัวแปรนั้นมีปัญหาเรื่อง Multicollinearity คือตัวแปรมีความสัมพันธ์กันสูงมากเกินไป

ในที่นี้พบว่า ไม่มีค่า VIF ของตัวแปรใด ที่มากกว่า 5 จึงไม่มีปัญหาเรื่อง Multicollinearity

3. ข้อมูลดิบ

Density	bodyfat	Age	Weight	Height	Neck	Chest	Abdomen	Hip	Thigh	Knee	Ankle	Biceps	Forearm	Wrist
1.0708	12.3	23	154.25	67.75	36.2	93.1	85.2	94.5	59.0	37.3	21.9	32.0	27.4	17.1
1.0853	6.1	22	173.25	72.25	38.5	93.6	83.0	98.7	58.7	37.3	23.4	30.5	28.9	18.2
1.0414	25.3	22	154.00	66.25	34.0	95.8	87.9	99.2	59.6	38.9	24.0	28.8	25.2	16.6
1.0751	10.4	26	184.75	72.25	37.4	101.8	86.4	101.2	60.1	37.3	22.8	32.4	29.4	18.2
1.0340	28.7	24	184.25	71.25	34.4	97.3	100.0	101.9	63.2	42.2	24.0	32.2	27.7	17.7
1.0502	20.9	24	210.25	74.75	39.0	104.5	94.4	107.8	66.0	42.0	25.6	35.7	30.6	18.8
1.0549	19.2	26	181.00	69.75	36.4	105.1	90.7	100.3	58.4	38.3	22.9	31.9	27.8	17.7
1.0704	12.4	25	176.00	72.50	37.8	99.6	88.5	97.1	60.0	39.4	23.2	30.5	29.0	18.8
1.0900	4.1	25	191.00	74.00	38.1	100.9	82.5	99.9	62.9	38.3	23.8	35.9	31.1	18.2
1.0722	11.7	23	198.25	73.50	42.1	99.6	88.6	104.1	63.1	41.7	25.0	35.6	30.0	19.2
1.0830	7.1	26	186.25	74.50	38.5	101.5	83.6	98.2	59.7	39.7	25.2	32.8	29.4	18.5
1.0812	7.8	27	216.00	76.00	39.4	103.6	90.9	107.7	66.2	39.2	25.9	37.2	30.2	19.0
1.0513	20.8	32	180.50	69.50	38.4	102.0	91.6	103.9	63.4	38.3	21.5	32.5	28.6	17.7
1.0505	21.2	30	205.25	71.25	39.4	104.1	101.8	108.6	66.0	41.5	23.7	36.9	31.6	18.8
1.0484	22.1	35	187.75	69.50	40.5	101.3	96.4	100.1	69.0	39.0	23.1	36.1	30.5	18.2
1.0512	20.9	35	162.75	66.00	36.4	99.1	92.8	99.2	63.1	38.7	21.7	31.1	26.4	16.9
1.0333	29.0	34	195.75	71.00	38.9	101.9	96.4	105.2	64.8	40.8	23.1	36.2	30.8	17.3
1.0468	22.9	32	209.25	71.00	42.1	107.6	97.5	107.0	66.9	40.0	24.4	38.2	31.6	19.3
1.0622	16.0	28	183.75	67.75	38.0	106.8	89.6	102.4	64.2	38.7	22.9	37.2	30.5	18.5
1.0610	16.5	33	211.75	73.50	40.0	106.2	100.5	109.0	65.8	40.6	24.0	37.1	30.1	18.2
1.0551	19.1	28	179.00	68.00	39.1	103.3	95.9	104.9	63.5	38.0	22.1	32.5	30.3	18.4
1.0640	15.2	28	200.50	69.75	41.3	111.4	98.8	104.8	63.4	40.6	24.6	33.0	32.8	19.9

ภาพที่ 4: ข้อมูลดิบ (Data view)

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
Density	Numeric	6	4		None	None	8	Right	Scale	Input
bodyfat	Numeric	4	1		None	None	8	Right	Scale	Input
Age	Numeric	2	0		None	None	8	Right	Scale	Input
Weight	Numeric	6	2		None	None	8	Right	Scale	Input
Height	Numeric	5	2		None	None	8	Right	Scale	Input
Neck	Numeric	4	1		None	None	8	Right	Scale	Input
Chest	Numeric	5	1		None	None	8	Right	Scale	Input
Abdomen	Numeric	5	1		None	None	8	Right	Scale	Input
Hip	Numeric	5	1		None	None	8	Right	Scale	Input
Thigh	Numeric	4	1		None	None	8	Right	Scale	Input
Knee	Numeric	4	1		None	None	8	Right	Scale	Input
Ankle	Numeric	4	1		None	None	8	Right	Scale	Input
Biceps	Numeric	4	1		None	None	8	Right	Scale	Input
Forearm	Numeric	4	1		None	None	8	Right	Scale	Input
Wrist	Numeric	4	1		None	None	8	Right	Scale	Input

ภาพที่ 5: ข้อมูล Variable view