Research Article

Received: July 18, 2018; Accepted: August 3, 2018

การเปรียบเทียบการใช้สถิติการวิเคราะห์ถดถอยโลจิสติกกับสถิติการวิเคราะห์จำแนกกลุ่มในการระบุเพศโดยการใช้ภาพเอกซเรย์คอมพิวเตอร์ระบบ 3 มิติ ของกระดูกสะบักในคนไทย

Comparison of Logistic Regression Analysis and Discriminant Analysis in Sex Identification Using 3D Computed Tomography Image of Thai Scapula

ศราวุฒิ สุจริตธรรม*

กลุ่มงานนิติเวชคลินิก กองนิติวิทยาศาสตร์บริการ

สถาบันนิติวิทยาศาสตร์ แขวงทุ่งสองห้อง เขตหลักสี่ กรุงเทพมหานคร 10210

Sarawut Sujarittham*

Group of Clinical Forensic Medicine, Forensic Science Service Division,

Central Institute of Forensic Science, Thung Song Hong, Lak Si, Bangkok 10210

บทคัดย่อ

การศึกษาเปรียบเทียบการใช้สถิติการวิเคราะห์ถดถอยโลจิสติกกับสถิติการวิเคราะห์จำแนกกลุ่มในการระบุเพศด้วยการใช้ภาพถ่ายรังสีจากเครื่องเอกซเรย์คอมพิวเตอร์ระบบ 3 มิติ ของกระดูกสะบัก จำนวนทั้งหมด 100 ตัวอย่าง แบ่งเป็นเพศหญิง 50 ตัวอย่าง และเพศชาย 50 ตัวอย่าง ในช่วงอายุ 20-70 ปี โดยภาพถ่ายรังสีดังกล่าวเกิดขึ้นจากกระบวนการผ่าชันสูตรศพของสถาบันนิติวิทยาศาสตร์ในช่วงปี พ.ศ. 2559 ถึง พ.ศ. 2560 โดยวัดขนาดของกระดูกสะบักด้านขวาจำนวน 7 ตัวแปร ด้วยโปรแกรม Vitrea version 4.1.52 (Vital Images, Inc.) วิเคราะห์ข้อมูลด้วยสถิติการวิเคราะห์ถดถอยโลจิสติกและสถิติการวิเคราะห์จำแนกกลุ่ม เพื่อคำนวณหาสมการและร้อยละความแม่นยำในการระบุเพศของแต่ละตัวแปรในการวัด การศึกษาการใช้ตัวแปรการวัดตัวแปรเดียว (univariate variable) พบว่าความสามารถในการแยกเพศของแต่ละตัวแปรต่างกันเล็กน้อยโดยที่สถิติการวิเคราะห์จำแนกกลุ่มให้ความแม่นยำในการระบุเพศร้อยละ 82-94 เทียบกับสถิติการวิเคราะห์ถดถอยโลจิสติกที่ให้ความแม่นยำร้อยละ 82-95 และตัวแปรที่ให้ความแม่นยำในการระบุเพศสูงสุด คือ GL (length of the glenoid cavity) ส่วนการใช้ตัวแปรการวัดหลายตัวแปร (multivariate variable) พบว่าสถิติการวิเคราะห์จำแนกกลุ่มให้ความแม่นยำในการระบุเพศที่ดีกว่าและใช้ตัวแปรเพียงสองตัวแปร คือ length of the glenoid cavity (GL) และ projection length of scapula spine (SS) มีสมการในการระบุเพศ คือ *-27.286 + 0.425GL + 0.09SS (cut value = 0)* โดยให้ความแม่นยำในการระบุเพศเฉลี่ยร้อยละ 97

***ผู้รับผิดชอบบทความ :** qforensic@gmail.com **doi:** 10.14456/tstj.2020.11

คำสำคัญ : กระดูกสะบัก; นิติมานุษยวิทยา; สถิติการวิเคราะห์จำแนกกลุ่ม; สถิติการวิเคราะห์ถดถอยโลจิสติกส์; ภาพเอกซเรย์คอมพิวเตอร์

Abstract

Comparison study of statistical application of logistic regression with discriminant analysis in sex identification by using 3-dimension computed tomography image was done in the samples of 100 scapula (from 50 females and 50 males, with ages between 20-70 years), and the mentioned images happened from the process of autopsy of Central Institute of Forensic Science during the year 2016 to 2017. The process had measured sizes of right scapula in a total of 7 positions with Vitrea version 4.1.52 program (Vital Images, Inc.), and data were analyzed with statistics of discriminant analysis and logistic regression in order to calculate for equation and percentage accuracy for identifying sex of variables measurement. The study had applied univariate variable measurement and found that the ability in sex distinguish of each variable is slightly different. By statistic discrimination analysis, it gave accuracy in sex identification for 82-94 %, compared with statistic logistic regression that gave accuracy of 82-95 %. The variable that gives the most accuracy in sex identification is GL (length of the glenoid cavity). However, for the application of multivariate variable, it was found that statistic discriminant analysis gives better accuracy in sex identification. Using only 2 variables of length of the glenoid cavity (GL) and projection length of scapula spine (SS) with sex identification equation of *-27.286 + 0.425GL + 0.09SS (cut value = 0)* gives accuracy in sex identification in an average of 97 %.

Keywords: scapula; forensic anthropology; discriminant analysis; logistic regression; computed tomography image

1. บทนำ

การระบุเพศถือเป็นหนึ่งในขั้นตอนที่มีความสำคัญในการพิสูจน์เอกลักษณ์บุคคล ซึ่งอาจทำได้ยากหากศพอยู่ในลักษณะของโครงกระดูกหรือพบเพียงกระดูกบางชิ้น โดยทั่วไปพบว่ากระดูกเพศหญิงและเพศชายนั้นจะมีความแตกต่างกันทั้งในรูปร่างและขนาด เนื่องจากความแตกต่างระหว่างเพศ (sexual dimorphism) ซึ่งเป็นผลมาจากการสร้างฮอร์โมน testosterone [1] ที่มีปริมาณสูงมากกว่าในเพศชาย จึงทำให้กระดูกในเพศชายจะมีขนาดที่ใหญ่กว่าเพศหญิงในหลาย ๆ ตำแหน่ง ดังนั้นจึงสามารถระบุเพศจากชิ้นส่วนกระดูกที่พบ โดยกระดูกที่มีความน่าเชื่อถือและมักนำมาใช้ในการระบุเพศ ได้แก่ กระดูกเชิงกรานและกะโหลกศีรษะ [2-4] แต่ในบางกรณีที่กระดูกเชิงกราน และกะโหลกศีรษะอยู่ในสภาพที่แตกหักไม่สามารถนำมาระบุเพศ อาจต้องนำกระดูกชิ้นอื่นมาใช้ระบุเพศแทน โดยที่กระดูกสะบักเป็นหนึ่งในกระดูกที่สามารถนำมาใช้ในการระบุเพศและเป็นกระดูกที่แตกหักได้ยาก เนื่องจากมีชั้นกล้ามเนื้อปกป้องอยู่ [5,6]

การระบุเพศโดยใช้ความแตกต่างของขนาดกระดูกด้วยการวัด (osteometry) จำเป็นต้องใช้สถิติที่มีความสามารถในการจำแนกกลุ่มของขนาดกระดูกที่มีความแตกต่างระหว่างเพศหญิงและเพศชายออกจากกัน โดยสถิติ logistic regression และ discriminant analysis เป็นสถิติที่สามารถนำมาใช้วิเคราะห์จำแนกขนาดของกระดูกในกลุ่มเพศหญิงและเพศชายออกจากกันด้วยการวิเคราะห์จากตัวแปรอิสระ [7] คือ ขนาด ตำแหน่งของกระดูกที่วัด และยังเป็นสถิติที่มักนำมาใช้ในงานวิจัยต่าง ๆ ทางนิติมานุษยวิทยาอีกด้วย [8-11] มีงานวิจัยหลายงานที่นำสถิติ logistic regression และ discriminant analysis มาใช้เปรียบเทียบกัน ซึ่งพบว่าความสามารถในการจำแนกกลุ่มมีความแตกต่างกันไปในแต่ละการวิจัยขึ้นอยู่กับลักษณะของชุดข้อมูลที่นำมาวิเคราะห์ ลักษณะการแจกแจงปรกติของข้อมูล (normal distribution) และลักษณะความเท่ากันของเมททริกซ์ความแปรปรวนร่วม (equality of covariance matrix) โดยพบว่าหากลักษณะชุดข้อมูลที่นำมาวิเคราะห์เป็นไปตามข้อกำหนดดังกล่าว การใช้ discriminant analysis จะให้ความแม่นยำในการจำแนกกลุ่มได้ดีกว่า แต่หากลักษณะของชุดข้อมูลมีลักษณะที่ไม่เป็นไปตามข้อกำหนด การใช้ logistic regression จะให้ความแม่นยำในการจำแนกกลุ่มที่ดีกว่า [12-15]

เนื่องจากหลักเกณฑ์ในการระบุเพศด้วยวิธีการวัดกระดูก (osteometry) ในแต่ละกลุ่มประชากรมีความแตกต่างกัน [2] และยังไม่พบว่ามีหลักเกณฑ์ในการระบุเพศโดยใช้ภาพเอกซเรย์คอมพิวเตอร์ระบบ 3 มิติ ของกระดูกสะบักในคนไทย งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อศึกษาหาหลักเกณฑ์ในการระบุเพศของกระดูกสะบักในคนไทยโดยการใช้ภาพเอกซเรย์คอมพิวเตอร์ระบบ 3 มิติ ที่เกิดจากกระบวนการผ่าชันสูตรศพ และเปรียบเทียบความแม่นยำในการจำแนกเพศระหว่างสถิติ logistic regression กับสถิติ discriminant analysis รวมทั้งหาว่าวิธีการใดมีความแม่นยำในการระบุเพศได้ดีที่สุด เพื่อนำมาใช้ในงานพิสูจน์เอกลักษณ์บุคคลของสถาบันนิติวิทยาศาสตร์ต่อไป

2. วิธีการวิจัย

โดยศึกษาภาพถ่ายรังสีจากเครื่องเอกซเรย์คอมพิวเตอร์ระบบ 3 มิติ Toshiba Aquilion Lightning, 16-row helical CT system (120 kVp, 50-150 mAs, 1-2 mm. slice thickness) จำนวน 100 ตัวอย่าง แบ่งเป็นเพศหญิง 50 ตัวอย่าง และเพศชาย 50 ตัวอย่าง ซึ่งมีอายุ 20-70 ปี ซึ่งภาพถ่ายรังสีดังกล่าวเกิดขึ้นจากกระบวนการผ่าชันสูตรศพของสถาบันนิติวิทยาศาสตร์ในช่วงปี พ.ศ. 2559 ถึง พ.ศ. 2560 ข้อมูลเพศและอายุของตัวอย่างบันทึกจากข้อมูลรายงานผ่าชันสูตร ตัวอย่างที่คัดเลือกเพื่อนำมาวัดขนาด โดยจะต้องไม่มีลักษณะที่ผิดรูป แตกหัก หรือมีพยาธิสภาพที่ผิดปกติของกระดูกสะบักขวา วัดขนาดของกระดูกสะบักด้านขวาจำนวน 7 ตำแหน่ง ด้วยโปรแกรม Vitrea version 4.1.52 (Vital Images, Inc.) ดังรูปที่ 1 และ 2 โดยมาตรฐานตำแหน่งของการวัดอ้างอิงจาก Polguj และคณะ [16] กำหนดตัวย่อของตัวแปรการวัดเพื่อนำไปใช้ในสมการเพื่อการระบุเพศ (ตารางที่ 1) แล้วนำข้อมูลที่ได้มาวิเคราะห์โดยใช้โปรแกรม SPSS version 24 (SPSS Inc., Chicago, IL, USA) สถิติที่นำมาใช้ได้แก่ สถิติเชิงพรรณนา (descriptive statistic) เพื่อหาค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐาน (standard deviation) ของแต่ละตำแหน่งของการวัด สถิติ independent t-test และ Mann-Whitney U test เพื่อเปรียบเทียบค่าเฉลี่ยของการวัดระหว่างเพศหญิงกับชาย และเปรียบเทียบค่าความแม่นยำในการระบุเพศโดยใช้สถิติ logistic regression และ discriminant

analysis ของแต่ละตัวแปรการวัด (original) สุดท้ายทดสอบความแม่นยำในการระบุเพศด้วย 5-fold cross validation ของตัวแปรแบบ multivariate เพื่อหาสมการที่มีความแม่นยำและมีความเหมาะสมมากที่สุดที่จะนำมาใช้ในการระบุเพศ

3. ผลการวิจัย

สถิติเชิงพรรณนาของตัวแปรการวัดกระดูกสะบักด้านขวาทั้งหมด 7 ตำแหน่ง จำนวน 100 ตัวอย่าง แสดงดังตารางที่ 2 เมื่อทดสอบการแจกแจงปรกติด้วยสถิติทดสอบ Kolmogorov-Smirnov พบว่าเกือบทุกตัวแปรมีการแจกแจงปรกติ ยกเว้นตัวแปร projection length of scapula spine (SS) และเมื่อทดสอบความแตกต่างของค่าเฉลี่ยตัวแปรการวัดด้วยสถิติ independent t-test และ Mann-Whitney U test พบว่ามีความแตกต่างอย่างมีนัยทางสถิติทุกตัวแปร (p-value = 0.000) โดยค่าเฉลี่ยของตัวแปรการวัดทั้งหมดของเพศชายมีขนาดใหญ่กว่าเพศหญิง

ข้อมูลในตารางที่ 3 แสดงค่า Wilk's lamda และความแม่นยำในการระบุเพศจากผลการวิเคราะห์ของสถิติ discriminant analysis และแสดงค่า -2 Log likelihood และความแม่นยำในการระบุเพศจากผลการวิเคราะห์ของสถิติ logistic regression ด้วยตัวแปรการวัดตัวแปรเดียว (univariate variable) โดยตัวแปรการวัดที่มีค่า Wilk's lamda และค่า -2 Log

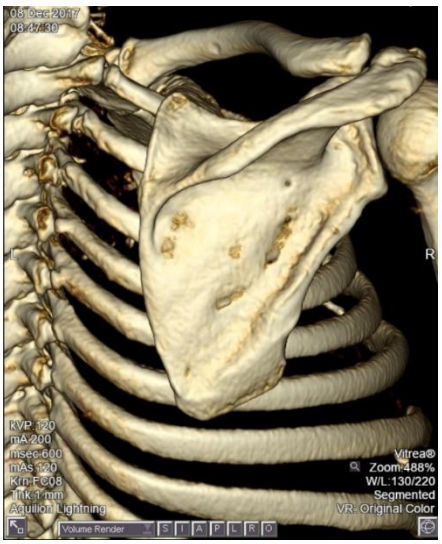

Figure 1 Scapula measurement by using Vitrea

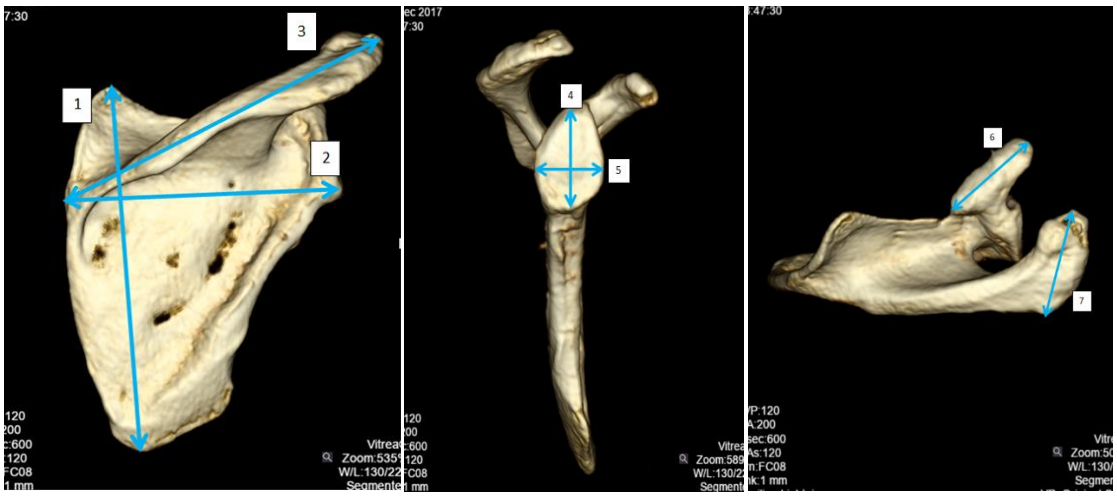

Figure 2 (A) Measurement of posterior landmark, (B) Measurement of glenoid cavity, (C) Measurement of acromion and coracoid process

Table 1 Description of each scapula measurement and abbreviations

No.	Measurements	Description	Abbr.
1	Morphological length	The distance from superior angle to inferior angle	ML
2	Morphological width	The distance from medial border to infraglenoid tubercle	MW
3	Projection length of Scapula spine	The length from spine of scapula to acromion process	SS
4	Length of the glenoid cavity	The maximum length of glenoid cavity	GL
5	Width of the glenoid cavity	The maximum width of glenoid cavity	GW
6	Maximal length of the coracoid process	The length of coracoid process	CP
7	Length of acromion	The length of acromion process	AP

Table 2 Descriptive statistics, Kolmogorov-Smirnov and compared means in females and males

Variables	Female (N=50)		Male (N=50)		Kolmogorov-Smirnov (* = p<0.05)	Compared means (* = p<0.05)
	Mean (mm.)	SD	Mean (mm.)	SD		
ML	132.0500	7.69726	151.4460	9.91378	0.083	t = -10.927*
MW	95.7300	4.96298	106.6640	5.30413	0.063	t = -10.644*
SS	122.1800	5.82307	138.3480	6.66435	0.093*	U = 89.000*
AP	38.5300	3.80844	46.1160	3.82763	0.054	t = -9.934*
CP	37.0300	3.34415	41.7500	3.47159	0.058	t = -6.924*
GL	34.1100	1.55999	39.3960	1.85318	0.082	t = -15.430*
GW	24.2640	1.86643	28.2120	1.77702	0.062	t = -10.833*

t = independent t-test; U = Mann-Whitney U test

likelihood ต่ำที่สุด คือ GL (length of the glenoid cavity) และเป็นตัวแปรที่มีความแม่นยำในการระบุเพศเฉลี่ยมากที่สุด (original) คือร้อยละ 94 (discriminant analysis) และ 95 (logistic regression) เมื่อเปรียบเทียบความแม่นยำในการระบุเพศเฉลี่ยทั้งเพศหญิงและชายระหว่างสถิติทั้งสอง (average accuracy for discriminant analysis and logistic regression) พบว่าตัวแปรการวัด SS (projection length of scapula spine) ให้ค่าความแม่นยำในการระบุเพศด้วยการใช้สถิติ discriminant analysis สูงกว่าสถิติ logistic regression เล็กน้อย ส่วนตัวแปรการวัด ML (morphological length) และ GL (length of the glenoid cavity) ให้ค่าความแม่นยำในการระบุเพศด้วยการใช้สถิติ logistic regression สูงกว่าสถิติ discriminant analysis ส่วนตัวแปรการวัดที่เหลือมีค่าความแม่นยำไม่แตกต่างกันระหว่างสถิติทั้งสอง

Table 3 Accuracy of classification of the original group by using discriminant analysis and logistic regression for the univariate

Variables	Wilk's lamda (* = p<0.05)	-2 Log likelihood	Predicted group membership for discriminant function (%)		Predicted group membership for logistic regression (%)		Average accuracy for discriminant function (%)	Average accuracy for logistic regression (%)
			Female	Male	Female	Male		
ML	0.451*	60.200	92	80	92	88	86	90
MW	0.464*	63.826	88	88	88	88	88	88
SS	0.370*	47.653	94	92	90	92	93	91
AP	0.498*	67.574	86	86	86	86	86	86
CP	0.672*	97.935	86	78	86	78	82	82
GL	0.292*	20.023	98	90	96	94	94	95
GW	0.455*	63.836	90	88	90	88	89	89

Table 4 Accuracy of classification of the original group by using discriminant analysis and logistic regression for the multivariate

Variables	Wilk's lamda (* = p<0.05)	-2 Log likelihood	Predicted group membership for discriminant function (%)		Predicted group membership for logistic regression (%)		Average accuracy for discriminant function (%)	Average accuracy for logistic regression (%)
			Female	Male	Female	Male		
GL/SS	0.223*	11.787	98	96	96	96	97	96
GL/SS/GW	0.222*	11.768	98	96	96	96	97	96
GL/SS/ML	0.222*	9.670	98	96	98	98	97	98
GL/SS/GW/ML	0.221*	9.669	98	96	98	98	97	98
GL/SS/GW/MW	0.219*	11.704	98	96	96	96	97	96
GL/SS/GW/MW/ML	0.217*	9.393	98	96	98	96	97	97
GL/SS/GW/MW/AP	0.217*	0.000	98	96	100	100	97	100
GL/SS/GW/MW/ML/AP	0.216*	0.000	98	96	100	100	97	100
GL/SS/GW/MW/ML/AP/CP	0.211*	0.000	96	96	100	100	96	100

ข้อมูลในตารางที่ 4 แสดงค่า Wilk's lamda และความแม่นยำในการระบุเพศของสถิติ discriminant analysis และค่า -2 Log likelihood และความแม่นยำในการระบุเพศของสถิติ logistic regression โดยใช้ตัวแปรการวัดหลายตัวแปร (mulivariate variable) โดยหากเพิ่มจำนวนตัวแปรการวัดค่า Wilk's lamda จะมีค่าลดลงตามจำนวนตัวแปรที่เพิ่มเข้าไป และค่า -2 Log likelihood สามารถลดลดจนมีค่าเป็น 0.000 ได้ ตัวแปรการวัด GL/SS เป็นตัวแปรที่มีความแม่นยำในการระบุเพศเฉลี่ยมากที่สุดในกลุ่มต้นแบบ (original) และใช้จำนวนตัวแปรน้อยที่สุดด้วยสถิติ discriminant analysis (ร้อยละ 97) การเพิ่มตัวแปรการวัดไม่ได้ทำให้ความแม่นยำในการระบุเพศเพิ่มมากขึ้นและอาจทำให้ความแม่นยำลดลง (เหลือร้อยละ 96 หากใช้ทุกตัวแปร) ส่วนการเพิ่มจำนวนตัวแปรการวัดในสถิติ logistic regression จะมีแนวโน้มทำให้ความแม่นยำในการระบุเพศเพิ่มมากขึ้น โดยที่ตัวแปรการวัด GL/SS/GW/MW/AP เป็นตัวแปรที่มีความแม่นยำในการระบุเพศเฉลี่ยมากที่สุดในกลุ่มต้นแบบ (original) และใช้จำนวนตัวแปรน้อยที่สุด (ร้อยละ 100) เมื่อเปรียบเทียบความแม่นยำในการระบุเพศเฉลี่ยระหว่างทั้งสองสถิติในกลุ่มต้นแบบ (original) พบว่าสถิติ logistic regression มีความแม่นยำมากกว่าสถิติ discriminant analysis เล็กน้อย (ร้อยละ 96-97 กับ ร้อยละ 96-100)

Table 5 Accuracy of classification of the 5-fold cross validation by using discriminant analysis and logistic regression for the multivariate

Variables	Function of Discriminant analysis (DA) and Logistic regression model (LR)	Predicted group membership For 5-fold cross validation (%)		Average accuracy for 5-fold cross validation (%)	Box's M * = p<0.05
		Female	male		
GL/SS (DA)	-27.286 + 0.425GL + 0.09SS	98	96	97	0.127
GL/SS/GW (DA)	-27.223 + 0.406GL + 0.086SS + 0.044GW	98	96	97	6.978
GL/SS/ML (DA)	-27.335 + 0.407GL + 0.082SS + 0.012ML	98	96	97	9.535
GL/SS/GW/ML (DA)	-27.276 + 0.392GL + 0.079SS + 0.038GW + 0.011ML	98	94	96	11.410
GL/SS/GW/MW (DA)	-26.695 + 0.4GL + 0.117SS + 0.073GW - 0.051MW	96	96	96	8.123
GL/SS/ML (LR)	-226.755 + 4.778GL + 0.208SS + 0.184ML	96	92	94	9.535
GL/SS/GW/ML (LR)	-225.945 + 4.778GL + 0.206SS - 0.026GW + 0.185ML	96	92	94	11.410
GL/SS/GW/MW /AP (LR)	-6195.758 + 106.683GL + 17.492SS + 1.296GW - 5.757MW + 13.784AP	94	96	95	15.277
GL/SS/GW/MW /ML/AP (LR)	-3246.708 + 37.011GL + 16.456SS - 37.465GW - 6.722MW + 3.993ML + 20.014AP	92	94	93	19.308
GL/SS/GW/MW /ML/AP/CP (LR)	-3243.871 + 36.864GL + 16.608SS - 37.787GW - 6.887MW + 3.911ML + 20.2AP + 0.291CP	92	92	92	31.041

ข้อมูลในตารางที่ 5 จากการทำ 5-fold cross validation ของตัวแปรแบบ multivariate พบว่าตัวแปรที่มีความแม่นยำในการระบุเพศเฉลี่ยมากที่สุดด้วยการใช้สถิติ discriminant analysis คือ GL/SS (DA) ให้ความแม่นยำร้อยละ 97 (ตารางที่ 5) โดยมีสมการในการระบุเพศ คือ *-27.286 + 0.425GL + 0.09SS (cut value = 0)* ส่วนตัวแปรที่มีความแม่นยำในการระบุเพศเฉลี่ยมากที่สุดด้วยการใช้สถิติ logistic regression คือ GL/SS/GW/MW/AP ให้ความแม่นยำร้อยละ 95 โดยมีสมการในการระบุเพศ คือ *logit = -6195.758 +106.683GL +17.492SS +1.296GW -5.757MW + 13.784AP*

4. วิจารณ์

กระดูกเชิงกรานและกะโหลกศีรษะมักเป็นตัวเลือกแรก ๆ ที่นำมาใช้ระบุเพศ การศึกษาของ Đurić และคณะ [17] พบว่าการใช้กระดูกเชิงกรานและกะโหลกศีรษะสามารถให้ความแม่นยำถึงร้อยละ 100 แต่ในบางกรณีอาจพบสภาพศพหรือโครงกระดูกที่ไม่สมบูรณ์ การระบุเพศจึงจำเป็นต้องทำโดยการตรวจลักษณะของกระดูกชิ้นอื่น ๆ ที่ยังคงเหลืออยู่ การศึกษานี้จึงเลือกนำกระดูกสะบักมาเป็นตัวเลือกในการระบุเพศหากกระดูกเชิงกรานและกะโหลกศีรษะไม่สามารถนำมาใช้

ข้อมูลสถิติเชิงพรรณนาที่แสดงดังตารางที่ 1 พบว่าตัวแปรการวัดทั้งหมด 8 ตัวแปร มีค่าเฉลี่ยขนาดของกระดูกในเพศชายที่ใหญ่กว่าเพศหญิงทุกตัวแปร แสดงให้เห็นว่าสามารถนำความแตกต่างระหว่างเพศ (sexual dimorphism) นี้มาใช้ในการแยกแยะกระดูกสะบักระหว่างเพศหญิงและเพศชาย การใช้ตัวแปรการวัดตัวแปรเดียว (univariate variable) พบว่าความสามารถในการแยกเพศของแต่ละตัวแปรต่างกันเล็กน้อย โดยสถิติ discrimination analysis ให้ความแม่นยำในการระบุเพศร้อยละ 82-94 เมื่อเปรียบเทียบกับสถิติ logistic regression ที่ให้ความแม่นยำร้อยละ 82-95 และตัวแปรที่ให้ความแม่นยำในการระบุเพศมากที่สุด คือ GL (length of the glenoid cavity) สอดคล้องกับงานวิจัยของ Zhang และคณะ [18] ที่พบว่าการใช้ตัวแปรการวัดตัวแปรเดียวสามารถให้ความแม่นยำในการระบุเพศมากกว่าร้อยละ 80 และสอดคล้องกับงานวิจัยของ Di Vella และคณะ [19] และ Özer และคณะ [20] ที่พบว่า length of the glenoid cavity (GL) เป็นตัวแปรที่มีความแม่นยำสูงสุดในการระบุเพศ โดยบางงานวิจัยพบว่าตัวแปร morphological width (MW) ให้ความแม่นยำในการระบุเพศสูงสุด [19] แต่งานวิจัยนี้พบว่าให้ความแม่นยำเพียงร้อยละ 88

การใช้ตัวแปรการวัดหลายตัวแปร (multivariate variable) ดังตารางที่ 4 พบว่าความสามารถในการระบุเพศในกลุ่มต้นแบบ (original) ของสถิติ logistic regression ให้ความแม่นยำในการระบุเพศได้ดีกว่าสถิติ discriminant analysis โดยพบว่าบางสมการสามารถให้ความแม่นยำในการระบุเพศถึงร้อยละ 100 (-2 Log likelihood = 0.000) ซึ่งอาจเกิดจาก overfiiting [21] ของข้อมูลที่นำไปทดสอบกับกลุ่มต้นแบบ (original) จึงได้การทดสอบความแม่นยำของการระบุเพศระหว่างสถิติทั้งสองด้วย 5-fold cross validation (ตารางที่ 5) พบว่าสถิติ discriminant analysis (ร้อยละ 96-97) ให้ความแม่นยำในการระบุเพศดีกว่าสถิติ logistic regression (ร้อยละ 92-95) การทดสอบการแจกแจงปรกติของข้อมูลด้วย Kolmogorov-Smirnov (ตารางที่ 2) พบว่าข้อมูลเกือบทั้งหมดมีการแจกแจกปรกติ และเมื่อทดสอบความเท่ากันของเมททริกซ์ความแปรปรวนร่วมด้วย Box's M (ตารางที่ 5) พบว่าทุกสมการในการระบุเพศเป็นไปตามข้อกำหนด ซึ่งสอดคล้องกับข้อมูลของ

Pohar และคณะ [22] ที่พบว่าสถิติ discriminant analysis เป็นตัวเลือกที่ดีกว่าในการจำแนกข้อมูลหาก ข้อมูลดังกล่าวไม่มีการละเมิดข้อกำหนดข้างต้น

เมื่อพิจารณาความแม่นยำในการระบุเพศหลังจากทำ 5-fold cross validation ของสถิติทั้งสองแล้ว สมการจากตัวแปร GL/SS จากการใช้สถิติ discriminant analysis เป็นสมการที่เหมาะสมที่จะนำมาใช้ในการระบุเพศ โดยมีความแม่นยำในการระบุเพศสูงสุด (ร้อยละ 97) และใช้ตัวแปรในการวัดเพียง 2 ตัวแปร มีสมการในการระบุเพศ คือ *-27.286 + 0.425GL + 0.09SS (cut value = 0)* เมื่อเปรียบเทียบกับงานวิจัยในอดีตพบว่าสมการดังกล่าวให้ความแม่นยำใกล้เคียงกับงานวิจัยของ Hanihara [23] (ร้อยละ 96.8) ที่ศึกษาวิจัยกระดูกสะบักในกลุ่มประชากรญี่ปุ่น และให้ความแม่นยำที่ดีกว่าเล็กน้อยเมื่อเปรียบเทียบกับงานวิจัยของ Di Vella และคณะ [19] ในกลุ่มประชากรอิตาลี่ (ร้อยละ 95) งานวิจัยของ Murphy [24] ในกลุ่มประชากรนิวซีแลนด์ (ร้อยละ 93.6) และงานวิจัยของ Frutos [25] ในกลุ่มประชากรกัวเตมาลา (ร้อยละ 94.8) จะเห็นได้ว่าแม่นยำของการระบุเพศโดยการใช้วิธีการวัดขนาด (osteometry) จะต่างกันไปในแต่ละกลุ่มประชาการที่ต่างเชื้อชาติกัน โดยสมการที่ได้จะสามารถนำมาใช้เฉพาะในกลุ่มตัวอย่างของเชื้อชาติที่นำมาใช้สร้างสมการ ซึ่งจะให้ความคลาดเคลื่อนในการระบุเพศน้อยที่สุด [2]

5. สรุป

การเปรียบเทียบการใช้สถิติการวิเคราะห์ถดถอยโลจิสติก (logistic regression) กับสถิติการวิเคราะห์จำแนกกลุ่ม (discriminant analysis) ในการระบุเพศโดยการใช้ภาพเอกซเรย์คอมพิวเตอร์ระบบ 3 มิติ ของกระดูกสะบักในคนไทย เมื่อทำ 5-fold cross validation พบว่าสถิติ discriminant analysis ให้ความแม่นยำในการระบุเพศที่ดีที่สุดและใช้ตัวแปรเพียง 2 ตัวแปร คือ length of the glenoid cavity (GL) และ projection length of scapula spine (SS) โดยให้ความแม่นยำในการระบุเพศเฉลี่ยร้อยละ 97 การเพิ่มตัวแปรเข้าไปมากกว่า 2 ตัวแปร ไม่ช่วยให้ความแม่นยำในการระบุเพศเพิ่มมากขึ้นแต่อย่างใด

6. References

[1] Bardin, C.W. and Catterall, J.F., 1981, Testosterone: A major determinant of extragenital sexual dimorphism, Science 211: 1285-94.

[2] İşcan, M.Y., 2005, Forensic anthropology of sex and body size, Forensic Sci. Int. 147: 107-12.

[3] Patriquin, M., Loth, S. and Steyn, M., 2003, Sexually dimorphic pelvic morphology in South African whites and blacks, Homo 53: 255-62.

[4] Steyn, M. and İşcan, M.Y., 1998, Sexual dimorphism in the crania and mandibles of South African whites, Forensic Sci. Int. 98: 9-16.

[5] Papaioannou,V.A.,Kranioti,E.F.,Joveneaux, P., Nathena, D. and Michalodimitrakis, M., 2012, Sexual dimorphism of the scapula and the clavicle in a contemporary Greek population: Applications in forensic identification, Forensic Sci. Int. 217: 231.e1-231.e7.

[6] Dabbs, G.R. and Moore-Jansen, P.H., 2010, A method for estimating sex using metric analysis of the scapula, J. Forensic Sci. 55:

149-152.

[7] Yarnold, P.R., Hart, L.A. and Soltysik, R.C., 1994, Optimizing the classification performance of logistic regression and Fisher's discriminant analyses, Educ. Psychol. Meas. 54: 73-85.

[8] Patil, K.R. and Mody, R.N., 2005, Determination of sex by discriminant function analysis and stature by regression analysis: A lateral cephalometric study, Forensic Sci. Int. 147: 175-180.

[9] Johnson, D., O'higgins, P., Moore, W. and McAndrew, T., 1989, Determination of race and sex of the human skull by discriminant function analysis of linear and angular dimensions, Forensic Sci. Int. 41: 41-53.

[10] Júnior, L.F., Francesquini, M.A., De La Cruz, B.M., Pereira, S.D., Ambrosano, G.M., Barbosa, C.M., Junior, E.D., Cury, A. and Daruge, E., 2007, Identification of sex using cranial base measurements, J. Forensic Odontostomatol. 25: 7-11.

[11] Acharya, A.B., Prabhu, S. and Muddapur, M.V., 2011, Odontometric sex assessment from logistic regression analysis, Int. J. Legal Med. 125: 199-204.

[12] Amores-Ampuero, A. and Alemán, I., 2016, Comparison of cranial sex determination by discriminant analysis and logistic regression, Anthropol. Anz. 73: 207-214.

[13] Santos, F., Guyomarc'h, P. and Bruzek, J., 2014, Statistical sex determination from craniometrics: Comparison of linear discriminant analysis, logistic regression, and support vector machines, Forensic Sci. Int. 245: 204.e1-204.e8.

[14] Kleinbaum, D.G., Kupper, L.L. and Morgenstern, H., 1982, Epidemiologic research: Principles and quantitative methods, John Wiley & Sons, New York.

[15] Montgomery, M., White, M. and Martin, S., 1987, A comparison of discriminant analysis and logistic regression for the prediction of coliform mastitis in dairy cows, Can. J. Vet. Res. 51: 495-498.

[16] Polguj, M., Jedrzejewski, K., Podgórski, M. and Topol, M., 2011, Correlation between morphometry of the suprascapular notch and anthropometric measurements of the scapula, Folia. Morphol. 70: 109-115.

[17] Đurić, M., Rakočević, Z. and Đonić, D., 2005, The reliability of sex determination of skeletons from forensic context in the Balkans, Forensic Sci. Int. 147: 159-164.

[18] Zhang, K., Cui, J.H., Luo, Y.Z., Fan, F., Yang, M., Li, X.H., Zhang, W., Deng, Z.H., 2016, Estimation of stature and sex from scapular measurements by three-dimensional volume-rendering technique using in Chinese, Leg. Med. 21: 58-63.

[19] Di Vella, G., Campobasso, C.P., Dragone, M. and Introna, F., 1994, Skeletal sex determination by scapular measurements, Boll. Soc. Ital. Biol. Sper. 70: 299-305.

[20] Özer, I., Katayama, K., Sahgir, M. and Güleç, E., 2006, Sex determination using the scapula in medieval skeletons from East Anatolia, Coll. Antropol. 30: 415-419.

[21] Cawley, G.C. and Talbot, N.L., 2010, On over-fitting in model selection and subsequent selection bias in performance evaluation, J. Mach. Learn Res. 11: 2079-2107.

[22] Pohar, M., Blas, M. and Turk, S., 2004, Comparison of logistic regression and linear discriminant analysis: A simulation study, Metodoloski zvezki 1: 143-161.

[23] Hanihara, K., 1959, Sex Diagnosis of Japanese Skulls and Scapulae by means of discriminant function, J. Anthropol. Soc. Nippon. 67: 191-197.

[24] Murphy, A., 1995, Sex determination of prehistoric New Zealand Polynesian scapulae, NZ J. Archaeol. 17: 29-34.

[25] Frutos, L.R., 2002, Determination of sex from the clavicle and scapula in a Guatemalan contemporary rural indigenous population, Am. J. Forensic Med. Pathol. 23: 284-288.