



การวิเคราะห์จำแนกประเภทกับการวิเคราะห์ถดถอยแบบโลจิสติกทวิ สำหรับทำนายปริมาณน้ำนมของฟาร์มโคนม

Discriminant Analysis and Binary Logistic Regression for Prediction of Milk Production of Dairy Cattle Farms

อรอุมา ทองหล่อ^{1*} สุภาวดี มานะไตรนนท์¹ และ อนันท์ เชาว์เครือ¹

บทคัดย่อ

วัตถุประสงค์ของงานวิจัยในครั้งนี้เพื่อศึกษาปัจจัยที่มีผลต่อปริมาณผลผลิตน้ำนมของการเลี้ยงโคนม โดยทำการเปรียบเทียบวิธีการวิเคราะห์จำแนกประเภท (ทั้งแบบ equal priors และ proportional priors) กับการวิเคราะห์ถดถอยแบบโลจิสติกทวิในการทำนายปริมาณน้ำนมของฟาร์มโคนมว่าเป็นฟาร์มที่ให้น้ำนมจัดอยู่ในกลุ่มฟาร์มที่ให้น้ำนมสูงหรือกลุ่มฟาร์มที่ให้น้ำนมต่ำ โดยใช้ค่าเฉลี่ยของน้ำนมรวมเป็นเกณฑ์ (13 กิโลกรัม/ตัว/วัน) ซึ่งมีตัวแปรต้นที่ทำการศึกษาเป็นตัวแปรเชิงกลุ่มประกอบไปด้วย ระดับเลือดโคพันธุ์ Holstein Friesian (3 กลุ่ม) อายุของแม่โค (4 กลุ่ม) ปีเกิดของแม่โค (5 กลุ่ม) เพอร์เซ็นต์อาหารชั้น (3 กลุ่ม) และแหล่งฟาร์ม (2 กลุ่ม) เก็บข้อมูลโดยใช้แบบสอบถามจำนวน 120 ฟาร์ม แบ่งเป็นฟาร์มโคนมในเขตตำบลสามพระยา อำเภอลำทะเมนชัย จังหวัดนครราชสีมา จำนวน 60 ฟาร์ม และฟาร์มโคนมในเขตตำบลห้วยสัตว์ใหญ่ อำเภอบางขัน จังหวัดนครศรีธรรมราช จำนวน 60 ฟาร์ม ผลการวิเคราะห์จำแนกประเภทพบว่าตัวแปรที่มีน้ำหนักหรือมีอิทธิพลในการจำแนกกลุ่มระหว่างกลุ่มฟาร์มที่มีน้ำนมสูงและน้ำนมต่ำสูงสุดเรียงตามลำดับคือ อายุของแม่โค (1.573) ระดับเลือด HF (1.092) ปีเกิดของแม่โค (0.736) เพอร์เซ็นต์อาหารชั้น (0.521) และแหล่งฟาร์ม (0.179) โดยสมการทำนายทั้งสามแบบสามารถใช้ในการจำแนกกลุ่มได้ผลใกล้เคียงกัน โดยสมการวิเคราะห์จำแนกประเภทแบบ equal priors สามารถจำแนกกลุ่มถูกต้อง 65.0 % ในขณะที่สมการจำแนกประเภทแบบ proportional priors และสมการถดถอยแบบโลจิสติกทวิสามารถจำแนกกลุ่มได้ถูกต้องเท่ากับ 64.2 % ส่วนความผิดพลาดในการจำแนกแต่ละกลุ่มระหว่างกลุ่มฟาร์มที่มีปริมาณน้ำนมต่ำและกลุ่มฟาร์มที่มีปริมาณน้ำนมสูง พบว่าสมการวิเคราะห์จำแนกประเภทแบบ equal priors มีค่าความผิดพลาดในการจัดกลุ่มผิดพลาดเท่ากับ 43.5 % และ 25.9 % ตามลำดับ ส่วนสมการวิเคราะห์จำแนกประเภทแบบ proportional priors และสมการถดถอยแบบโลจิสติกทวิมีค่าความผิดพลาดในการจัดกลุ่มผิดพลาด

¹คณะสัตวศาสตร์และเทคโนโลยีการเกษตร มหาวิทยาลัยศิลปากร วิทยาเขตสารสนเทศเพชรบุรี จ.เพชรบุรี 76120

*Corresponding Author, E-mail: on_uma2523@yahoo.com

เท่ากันคือเท่ากับ 43.5 % และ 27.6 % ตามลำดับ จากการเปรียบเทียบวิธีการวิเคราะห์จำแนกกลุ่มฟาร์มโคนมใน ครั้งนี้จะเห็นได้ว่าการวิเคราะห์จำแนกประเภท (ทั้งแบบ equal priors และ proportional priors) และการ วิเคราะห์ถดถอยแบบโลจิสติกทวิมัลติพริสตีทิภาพการจำแนกจัดกลุ่มฟาร์มตามปริมาณน้ำนมในกลุ่มต่ำกับกลุ่มสูงได้ ใกล้เคียงกัน

ABSTRACT

There were 2 objectives in this study. The first objective was to evaluate the factors influencing on the milk production in dairy cows by using discriminant analysis (equal priors/proportional priors) and binary logistic regression. The second objective was to compare the classification efficiency among three multivariate analysis techniques in order to classified dairy farms into 2 groups that were high and low milk production based on the average milk yield for all studied dairy farms (13 kg/cow/day). The three methods were included discriminant analysis with equal priors, discriminant analysis with proportional priors and binary logistic regression. The attribute variables, blood level of Holstein Friesian cows (3 groups), age of cows (4 groups), year of birth of cows (5 groups), percentage of concentrate feed (3 groups) and location of farm (2 groups) were used to identify the group of farms. All information of dairy cattle farms were surveyed by using 120 questionnaires from dairy cattle farms in Tambon Sam Pra Ya, Cha-Am, Phetchaburi (60 farms) and in Tambon Huay Sat Yai, Hua-Hin, Prachuabkirikhan (60 farms). The results based on discriminant analysis revealed that the age of cows showed the highest classification scores (1.573). The classification scores were used to classify any farm into the group of farms with high or low milk yield. While the classification scores of other variables were ranged from 1.092 for blood level of Holstein Friesian, 0.736 for year of birth of cows, 0.521 for percentage of concentrate feed and 0.179 for location of farm, respectively. The study found that all 3 models showed the similar results. The correct classifications were 65 % for discriminant analysis with equal priors model while the results from discriminant analysis with proportional priors model was equal to the results from binary logistic regression model that was 64.2 %. The misclassification of dairy farms into 2 groups between groups of farm that showed low milk production and group of farms that showed high milk production for discriminant analysis with equal priors model were 43.5 % and 25.9 %, respectively, while the results from discriminant analysis with proportional priors model was equal to the results from binary logistic regression model that were 43.5 % and 27.6%, respectively. The results can be concluded that both types of discriminant analysis (with equal priors or with proportional priors)

and logistic regression have similar efficiency for classification the farms into 2 groups according to the level of milk production.

คำสำคัญ: โคนม ผลผลิตน้ำนม การทำนาย การวิเคราะห์จำแนกประเภท วิเคราะห์ถดถอยแบบโลจิสติก

Keywords: Dairy cattle, Milk production, Prediction, Discriminant analysis, Logistic regression

บทนำ

ปัจจัยหลักที่มีผลต่อประสิทธิภาพการผลิตและความสำเร็จในการเลี้ยงโคนมที่สำคัญ ได้แก่ ปัจจัยทางด้านพันธุ์สัตว์ อาหาร และการจัดการฟาร์ม ซึ่งผู้เลี้ยงจะต้องให้ความสำคัญเป็นอย่างมากเพราะปัจจัยต่าง ๆ เหล่านี้สามารถส่งผลกระทบต่อทั้งทางตรงและทางอ้อมต่อปริมาณผลผลิตน้ำนมของฟาร์มได้ โดยการจัดการฟาร์มในแต่ละพื้นที่หรือแต่ละประเทศจะมีพื้นฐานของปัจจัยเหล่านี้แตกต่างกันไป จึงจำเป็นต้องมีการศึกษาหาปัจจัยที่มีอิทธิพลต่อประสิทธิภาพการผลิตที่แท้จริง

การศึกษาปัจจัยที่มีผลต่อปริมาณผลผลิตน้ำนม โดยนำปัจจัยทางด้านพันธุ์สัตว์ ปัจจัยทางด้านอาหาร และปัจจัยทางการจัดการฟาร์ม มาใช้ในการทำนายปริมาณผลผลิตน้ำนม ซึ่งแบ่งเป็นกลุ่มฟาร์มที่มีปริมาณน้ำนมต่ำกับสูง สามารถใช้เทคนิคการวิเคราะห์ตัวแปรหลายตัวได้ทั้ง 2 วิธี วิธีหนึ่งคือวิเคราะห์ถดถอยแบบโลจิสติก (logistic regression, LR) โดยในการศึกษาครั้งนี้ใช้แบบทวิ (binary logistic regression, BLR) เนื่องจากเป็นการจำแนกออกเป็น 2 กลุ่ม อีกวิธีหนึ่งคือการวิเคราะห์จำแนกประเภท (discriminant analysis, DA) เนื่องจากทั้งสองวิธีนี้มีคุณสมบัติคล้ายกัน โดยเฉพาะอย่างยิ่งในด้านตัวแปรตามและในรูปแบบของสมการ จึงมีงานวิจัยหลายงานที่นำวิธีการจำแนกทั้งสองวิธีนี้มาใช้เพื่อทำการเปรียบเทียบ อาทิ Halperin et al. (1971) สรุปไว้ว่า

LR กับ DA ให้ผลต่างกันเพียงเล็กน้อยในการจำแนกกลุ่ม Montgomery et al. (1987) ให้ผลสรุปว่าทั้งสองวิธีให้ผลใกล้เคียงกัน แต่ LR มีคุณสมบัติคงทน (robustness) ต่อการฝ่าฝืนข้อกำหนดของข้อมูลและผลการจำแนกดีกว่า DA เพียงเล็กน้อยเมื่อข้อมูลไม่สอดคล้องตามข้อกำหนด เช่นเดียวกับ Press and Wilson (1978) คือถ้าข้อกำหนดของการแจกแจงแบบปกติและความแปรปรวนหรือเมทริกซ์ความแปรปรวนเท่ากันแล้ว DA จะให้ผลการจำแนกดีกว่า แต่ถ้าข้อกำหนดดังกล่าวไม่สอดคล้อง LR จะจำแนกได้ดีกว่า Efron (1975) กล่าวว่า LR มี 1/2 ถึง 2/3 ครั้งที่มีประสิทธิภาพในการจำแนกเทียบเท่ากับ DA กรณีที่ตัวแปรตามจำแนกออกเป็น 2 กลุ่ม และตัวแปรต้นในแต่ละกลุ่มมีการแจกแจงแบบปกติ และการศึกษาของ Bull and Donner (1987) ก็ให้ผลเช่นเดียวกันในกรณีที่ตัวแปรตามจำแนกออกเป็น 3 กลุ่ม ดังนั้นจึงเกิดทางเลือกว่าควรเลือกใช้วิธีใดจึงจะเหมาะสมกว่า

DA เป็นการวิเคราะห์เพื่อศึกษาตัวแปรที่มีอิทธิพลต่อการจำแนกประชากรออกเป็นกลุ่ม ๆ โดยทดสอบว่า ตัวแปรต้นมีความสัมพันธ์กับตัวแปรตามจนถึงขั้นที่จะสามารถจำแนกประชากรออกเป็นกลุ่ม ๆ ได้หรือไม่ ทำให้ได้สมการจำแนกประเภท เพื่อนำไปใช้ในการทำนายการเป็นสมาชิกของกลุ่มของตัวแปรตาม โดยสมการในการจำแนกมีรูปแบบดังนี้ $Z_{jk} = \beta_0 + \beta_1 X_{1k} + \beta_2 X_{2k} + \dots + \beta_n X_{nk}$ เมื่อ Z_{jk} คือ คะแนนมาตรฐานของการจำแนก, β_0 คือ จุดตัด โดยค่าของ

จุดตัดจะขึ้นกับความน่าจะเป็นที่เกิดขึ้นก่อน (prior probabilities) ของแต่ละกลุ่ม ดังนั้นสมการของ DA จึงมี 2 รูปแบบคือ กรณีที่ใช้ความน่าจะเป็นที่เกิดขึ้นก่อนเท่ากัน (DA with Equal Priors) และใช้ความน่าจะเป็นที่เกิดขึ้นก่อนจากค่าสัดส่วน (DA with Proportional Priors) และ β_1 คือ สัมประสิทธิ์ของตัวแปรต้น ซึ่งการประมาณค่าไม่ขึ้นกับความน่าจะเป็นที่เกิดขึ้นก่อน Hossain et al. (2002)

ข้อกำหนดของ DA คือ เมทริกซ์ความแปรปรวนของแต่ละกลุ่มของตัวแปรตามจะต้องเท่ากัน และแต่ละกลุ่มจะต้องมีการแจกแจงแบบปกติ แต่ข้อกำหนดจะถูกฝ่าฝืนเมื่อตัวแปรต้นเป็นข้อมูลเชิงกลุ่ม ส่วนการวิเคราะห์ถดถอยแบบโลจิสติกมีตัวแปรตามเป็นค่าล็อกของอัตราส่วนโอกาสในการเป็นสมาชิกของกลุ่มหนึ่งเหนืออีกกลุ่มหนึ่ง (log odds of one group over the other) ดังนี้

$$\ln \left(\frac{\text{Pr ob(even)}}{\text{Pr ob(no even)}} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

หรือ

$$\text{odds}_i = \frac{\text{Pr ob(even)}}{\text{Pr ob(no even)}} = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n}$$

β_0 และ β_1 คือ จุดตัดและสัมประสิทธิ์ของตัวแปรต้น เช่นเดียวกัน ซึ่งประมาณด้วยวิธี maximum likelihood (McCullagh and Nelder, 1989)

งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อศึกษาปัจจัยที่มีผลต่อปริมาณผลผลิตน้ำนมของการเลี้ยงโคนมในฟาร์มเกษตรกรรายย่อยในเขตพื้นที่ภาคกลางตอนล่าง จังหวัดเพชรบุรีและประจวบคีรีขันธ์ โดยทำการ

เปรียบเทียบผลการจำแนกหรือทำนายกลุ่มฟาร์มตามปริมาณน้ำนมทั้งหมด 3 วิธีเปรียบเทียบกัน ประกอบด้วยการวิเคราะห์ถดถอยแบบโลจิสติกทวิ (BLR) การวิเคราะห์จำแนกประเภทแบบ equal priors (DA with Equal Priors) และการวิเคราะห์จำแนกประเภทแบบ proportional priors (DA with proportional priors) โดยแบ่งกลุ่มฟาร์มเป็น 2 กลุ่มคือ กลุ่มฟาร์มที่มีปริมาณน้ำนมต่ำกว่ากลุ่มฟาร์มที่มีปริมาณน้ำนมสูง และใช้ตัวแปรทางด้านพันธุ์สัตว์อาหาร และการจัดการฟาร์ม ในการจำแนก โดยทุกตัวแปรอยู่ในรูปตัวแปรเชิงกลุ่ม แล้วหาข้อสรุปว่าวิธีไหนมีประสิทธิภาพในการจำแนกกลุ่มได้ดีกว่ากันสำหรับข้อมูลชุดนี้

วิธีดำเนินงาน

1. การรวบรวมข้อมูล

ประชากรในการศึกษา คือ กลุ่มเกษตรกรรายย่อยที่เลี้ยงโคนมในเขตพื้นที่ภาคกลางตอนล่าง ตำบลสามพระยาอำเภอชะอำจังหวัดเพชรบุรี และตำบลห้วยสัตว์ใหญ่ อำเภอหัวหิน จังหวัดประจวบคีรีขันธ์ โดยทำการสุ่มกลุ่มตัวอย่างโดยวิธีคัดเลือก เกษตรกรที่เลี้ยงโคนมในเขตพื้นที่จำนวน 120 ฟาร์ม แบ่งเป็นพื้นที่ละ 60 ฟาร์ม โดยใช้การสัมภาษณ์ (interview schedule) เป็นเครื่องมือในการเก็บข้อมูล และใช้วิธีการสังเกต (observation) การจดบันทึกสภาพแวดล้อมภายในฟาร์มของเกษตรกร โดยเก็บข้อมูลตั้งแต่วันที่ 4-17 มีนาคม พ.ศ. 2554

2. การวิเคราะห์ข้อมูล

สร้างแบบจำลองเพื่อทำนายปริมาณผลผลิตน้ำนม โดยใช้สมการจำแนกกลุ่ม 3 แบบ คือ BLR, DA with equal priors (กำหนดให้ความน่าจะเป็นแต่ละกลุ่มเท่ากับ 0.5) และ DA with proportional priors

(กำหนดความน่าจะเป็นเป็นตามสัดส่วนจำนวนฟาร์ม โดยกลุ่มปริมาณน้ำนมต่ำเท่ากับ 0.517 และกลุ่มปริมาณน้ำนมสูงเท่ากับ 0.483) และตรวจสอบความเหมาะสมของสมการทั้งสามแบบที่ความเชื่อมั่น 80 % โดยทำการวิเคราะห์ข้อมูลทางสถิติตามวิธีการอ้างอิงโดยอภิญา และพูลพงษ์ (2554)

ตัวแปรในงานวิจัย มีดังนี้ ตัวแปรตาม คือ กลุ่มฟาร์มที่มีปริมาณน้ำนมต่ำและกลุ่มฟาร์มที่มีปริมาณน้ำนมสูง โดยใช้ค่าเฉลี่ยของปริมาณน้ำนมเป็นเกณฑ์แบ่งกลุ่ม (ปริมาณน้ำนมเฉลี่ยเท่ากับ 13 กิโลกรัม/ตัว/วัน) ดังนั้นกลุ่มฟาร์มที่มีปริมาณน้ำนมต่ำคือ ฟาร์มที่มีปริมาณน้ำนมต่ำกว่า 13 กิโลกรัม/ตัว/วัน มีจำนวน 62 ฟาร์ม และกลุ่มฟาร์มที่มีปริมาณน้ำนมสูงคือ ฟาร์มที่มีปริมาณน้ำนมมากกว่าหรือเท่ากับ 13 กิโลกรัม/ตัว/วัน มีจำนวน 58 ฟาร์ม

ตัวแปรต้น คือ ปัจจัยทางด้านพันธุ์สัตว์ ปัจจัยทางด้านอาหาร และปัจจัยทางการจัดการฟาร์ม ดังนี้

1. ระดับเลือด Holstein Friesian (HF) แบ่งเป็น 3 กลุ่ม คือ ระดับเลือดต่ำกว่าหรือเท่ากับ 87.5 สูงกว่า 87.5 และสัดส่วนกลุ่มต่ำกว่าหรือเท่ากับ 87.5:สูงกว่า 87.5 เป็น 1:1

2. อายุของแม่โค แบ่งเป็น 4 กลุ่ม คือ ต่ำกว่า 3 ปี ตั้งแต่ 3 ถึง 7 ปี สูงกว่า 7 ปี และสัดส่วนกลุ่มตั้งแต่ 3 ถึง 7 ปี:สูงกว่า 7 ปีเป็น 1:1

3. ปีเกิดของแม่โค แบ่งเป็น 5 กลุ่ม คือ ก่อนปี พ.ศ. 2544 ตั้งแต่ปี พ.ศ. 2544 ถึง 2548 หลังปี พ.ศ. 2548 สัดส่วนกลุ่มตั้งแต่ปี พ.ศ. 2544 ถึง 2548:หลังปี พ.ศ. 2548 เป็น 1:1 และสัดส่วนกลุ่มก่อนปี พ.ศ. 2544: ตั้งแต่ปี พ.ศ. 2544 ถึง 2548: หลังปี พ.ศ. 2548 เป็น 1:1:1

4. เปอร์เซ็นต์อาหารชั้น (เปอร์เซ็นต์อาหารชั้นเมื่อเทียบกับปริมาณอาหารหยาบ) แบ่งเป็น 3 กลุ่ม คือ ต่ำกว่า 15% ตั้งแต่ 15% ถึง 23% และสูงกว่า 23%

5. แหล่งฟาร์ม แบ่งเป็น 2 กลุ่ม คือ สามพระยา และห้วยสัตว์ใหญ่

ทุกตัวแปรอยู่ในรูปแบบตัวแปรหุ่น (dummy variable) ที่มีค่าเป็น 0 แทน ไม่ใช่ และ 1 แทนใช่เท่านั้น ดังนั้นในแต่ละปัจจัยจะมีจำนวนตัวแปรในสมการเท่ากับจำนวนกลุ่มลบบอกด้วยหนึ่ง โดยมีโครงสร้างของข้อมูล (ในรูป %) แต่ละตัวแปรที่ใช้ในการวิเคราะห์ ดังแสดงรายละเอียดในตารางที่ 1

ตารางที่ 1 แสดงโครงสร้างของข้อมูล(ในรูป %) แต่ละตัวแปรที่ใช้ในการวิเคราะห์

ปัจจัยการจัดการฟาร์ม	กลุ่มฟาร์มแยกตามปริมาณน้ำนม		χ^2 test (P-Value)
	ต่ำ	สูง	
1. ระดับเลือด HF			0.199
- ต่ำกว่าหรือเท่ากับ 87.5	62.3	75.9	
- สูงกว่า 87.5	27.9	20.7	
- สัดส่วนกลุ่มต่ำกว่าหรือเท่ากับ 87.5:สูงกว่า 87.5 เป็น 1:1	9.8	3.4	

ตารางที่ 1 แสดงโครงสร้างของข้อมูล (ในรูป %) แต่ละตัวแปรที่ใช้ในการวิเคราะห์ (ต่อ)

ปัจจัยการจัดการฟาร์ม	กลุ่มฟาร์มแยกตามปริมาณน้ำนม		χ^2 test (P-Value)
	ต่ำ	สูง	
2. อายุของแม่โค			0.038
- ต่ำกว่า 3 ปี	0.0	3.4	
- ตั้งแต่ 3 ถึง 7 ปี	77.0	89.7	
- สูงกว่า 7 ปี	18.0	6.9	
- สัดส่วนกลุ่มตั้งแต่ 3 ถึง 7 ปี : สูงกว่า 7 ปีเป็น 1:1	4.9	0.0	
3. ปีเกิดของแม่โค			0.310
- ก่อนปี พ.ศ. 2544	4.9	3.4	
- ตั้งแต่ปี พ.ศ. 2544 ถึง 2548	65.6	48.3	
- หลังปี พ.ศ. 2548	18.0	32.8	
- สัดส่วนกลุ่มตั้งแต่ปี พ.ศ. 2544 ถึง 2548: หลังปี พ.ศ. 2548 เป็น 1:1	9.8	12.1	
- สัดส่วนกลุ่มก่อนปี พ.ศ. 2544 : ตั้งแต่ปี พ.ศ. 2544 ถึง 2548: หลังปี พ.ศ. 2548 เป็น 1:1:1	1.6	3.4	
4. เพอร์เซ็นต์อาหารชั้น			0.167
- ต่ำกว่า 15%	53.2	50.0	
- ระหว่าง 15% ถึง 23%	25.8	15.5	
- สูงกว่า 23%	21.0	34.5	
5. แหล่งฟาร์ม			0.201
- สามพระยา	56.5	43.1	
- ห้วยสัตว์ใหญ่	43.5	56.9	

ผลการวิเคราะห์

1. รูปแบบสมการจำแนกกลุ่ม

จากการสร้างสมการในการจำแนกกลุ่มระหว่างฟาร์มที่มีปริมาณน้ำนมต่ำกับสูง โดยใช้ BLR, DA with equal priors และ DA with proportional

priors พบว่าค่าจุดตัดของสมการทั้ง 3 สมการต่างกันเพียงเล็กน้อย ส่วนค่าสัมประสิทธิ์ของแต่ละตัวแปรสมการ DA with equal priors กับสมการ DA with proportional priors มีค่าเท่ากัน ดังตารางที่ 2

ตารางที่ 2 แสดงค่าสัมประสิทธิ์ของสมการในการจำแนกกลุ่ม

ปัจจัยการจัดการฟาร์ม	BLR	DA with equal priors		DA with proportional priors	
		น้ำหนัก ¹	น้ำหนัก ²	น้ำหนัก ¹	น้ำหนัก ²
จุดตัด (Intercept)	-	-29.592	-31.746	-29.559	-31.780
1. ระดับเลือด HF	21.254				
ดื่มน้ำกลุ่มต่ำกว่าหรือเท่ากับ 87.5	1.440	11.102	12.606	11.102	12.606
ดื่มน้ำกลุ่มสูงกว่า 87.5	0.622	9.445	10.113	9.445	10.113
2. อายุของแม่โค					
ดื่มน้ำกลุ่มต่ำกว่า 3 ปี	41.874	26.469	30.180	26.469	30.180
ดื่มน้ำกลุ่มตั้งแต่ 3 ถึง 7 ปี	21.280	26.381	28.506	26.381	28.506
ดื่มน้ำกลุ่มสูงกว่า 7 ปี	19.435	24.362	24.797	24.362	24.797
3. ปีเกิดของแม่โค					
ดื่มน้ำกลุ่มก่อนปี พ.ศ. 2548	0.840	23.003	23.503	23.003	23.503
ดื่มน้ำกลุ่มตั้งแต่ปี พ.ศ. 2544 ถึง 2548	-0.849	19.127	18.329	19.127	18.329
ดื่มน้ำกลุ่มหลังปี พ.ศ. 2548	-0.053	18.932	18.971	18.932	18.971
ดื่มน้ำกลุ่มสัดส่วนกลุ่มตั้งแต่ปี พ.ศ. 2544 ถึง 2548:หลังปี พ.ศ. 2548 เป็น 1:1	-0.586	18.711	18.224	18.711	18.224
4. เพอร์เซ็นต์อาหารชั้น					
ดื่มน้ำกลุ่มต่ำกว่า 15%	-0.194	5.032	4.872	5.032	4.872
ดื่มน้ำกลุ่มระหว่าง 15% ถึง 23%	-0.935	4.854	3.896	4.854	3.896
5. แหล่งฟาร์ม					
- ดื่มน้ำกลุ่มสามพระยา	-0.332	2.710	2.387	2.710	2.387

หมายเหตุ ^{1,2} Fisher's linear discriminant functions สำหรับจำแนกกลุ่มฟาร์มที่มีปริมาณน้ำนมต่ำและสูง ตามลำดับ

ผลการทดสอบความเหมาะสมของรูปแบบสมการจำแนก BLR พบว่าตัวแปรต้นมีความสัมพันธ์กับตัวแปรตาม ($p < 0.20$) (Omnibus test of model coefficients; Chi-square model = 24.448, df. = 12, P-value = 0.018, Nagelkerke R-Square = 0.246) และสมการมีความสามารถในการจำแนกกลุ่ม (Chi-square; Hosmer and Lemeshow test =

3.47, df. = 8, P-value = 0.901) และสมการของ DA with equal priors กับสมการ DA with proportional priors ก็มีความสามารถในการจำแนกกลุ่มเช่นเดียวกัน ($p < 0.20$) (Wilks' Lambda = 0.829, Chi-square = 21.019, df. = 12, P-value = 0.050, canonical correlation = 0.414) และจาก standardized canonical discriminant function

coefficient ได้ว่าตัวแปรที่มีน้ำหนักในการจำแนกสูงสุดหรือมีอิทธิพลสูงสุดเรียงตามลำดับคือ อายุของแม่โค (1.573) ระดับเลือด HF (1.092) ปีเกิดของแม่โค (0.736) เพอร์เซ็นต์อาหารชั้น (0.521) และแหล่งฟาร์ม (0.179)

2. ผลการจำแนกกลุ่ม

ผลการจำแนกกลุ่มพบว่า กลุ่มฟาร์มที่มีปริมาณน้ำนมต่ำ (62 ฟาร์ม) สมการทั้ง 3 รูปแบบ จำแนกผิดพลาด 43.5% เท่ากัน แต่กลุ่มฟาร์มที่มีปริมาณน้ำนมสูง (58 ฟาร์ม) ซึ่งมีจำนวนฟาร์มหรือ

จำนวนสมาชิกน้อยกว่า สมการของ DA with proportional priors และ BLR จะจำแนกกลุ่มผิดพลาด 27.6% ซึ่งสูงกว่าสมการ DA with equal priors (25.9%) แต่โดยรวมแล้วพบว่าความผิดพลาดในการจำแนกกลุ่มจากสมการทั้ง 3 รูปแบบมีความใกล้เคียงกัน โดยสมการ DA with equal priors (35.0%) มีความผิดพลาดต่ำกว่าสมการ DA with proportional priors (35.8%) และ BLR (35.8%) ซึ่งมีค่าเท่ากัน

ตารางที่ 3 แสดงความผิดพลาด (%) ในการจำแนกกลุ่ม

กลุ่ม	BLR (%)	DA with equal priors (%)	DA with proportional priors (%)
กลุ่มฟาร์มที่มีปริมาณน้ำนมต่ำ	43.5	43.5	43.5
กลุ่มฟาร์มที่มีปริมาณน้ำนมสูง	27.6	25.9	27.6
รวม	35.8	35.0	35.8

วิจารณ์ผลการทดลอง

จากการใช้สมการ BLR, DA with equal priors และ DA with proportional priors ในการจำแนกกลุ่มฟาร์มตามปริมาณน้ำนมซึ่งแบ่งเป็นกลุ่มฟาร์มที่มีปริมาณน้ำนมต่ำ (มีปริมาณน้ำนมต่ำกว่า 13 กิโลกรัม/ตัว/วัน) และกลุ่มฟาร์มที่มีปริมาณน้ำนมสูง (มีปริมาณน้ำนมมากกว่าหรือเท่ากับ 13 กิโลกรัม/ตัว/วัน) โดยตัวแปรที่นำมาใช้ในการจำแนกกลุ่มประกอบไปด้วย 5 ตัวแปร คือ ระดับเลือด HF อายุของแม่โค ปีเกิดของแม่โค เพอร์เซ็นต์อาหารชั้น และแหล่งฟาร์ม เนื่องจากทุกตัวแปรเป็นตัวแปรเชิงกลุ่ม จึงทำให้ข้อกำหนดของ DA ไม่สามารถตรวจสอบได้ ผลการวิเคราะห์ที่ได้ว่าตัวแปรต้นทั้ง 5 ตัวแปรมีความสัมพันธ์กับตัวแปรตาม (กลุ่มฟาร์มแยกตามปริมาณน้ำนม) จนถึงขั้นที่สามารถจำแนกฟาร์มตามปริมาณน้ำนมออกเป็น 2

กลุ่มได้ ทำให้สมการทั้ง 3 รูปแบบมีความสามารถในการจำแนกกลุ่ม และนำไปใช้ในการทำนายการเป็นสมาชิกของกลุ่มของตัวแปรตามได้ จากการตรวจสอบประสิทธิภาพของสมการด้วย Hosmer and Lemeshow test (P-value = 0.901) สำหรับสมการ BLR และทดสอบด้วย Wilks' Lambda (P-value = 0.050) สำหรับ DA with equal priors และ DA with proportional priors โดยตัวแปรที่มีอิทธิพลสูงสุดในการจำแนกฟาร์มว่าควรจัดอยู่ในกลุ่มฟาร์มที่ให้ น้ำนมสูงหรือต่ำเรียงตามลำดับคือ อายุของแม่โค ระดับเลือด HF ปีเกิดของแม่โค เพอร์เซ็นต์อาหารชั้น และแหล่งฟาร์ม

จากผลการจำแนกได้ว่าสมการ DA with proportional priors มีความผิดพลาดในการจำแนกเท่ากับสมการ BLR คือ 35.8 % แต่สูงกว่าสมการ DA

with equal priors ที่ผิดพลาด 35.0 % จะเห็นว่า สมการของ DA with proportional priors กับ DA with equal priors ให้ผลการจำแนกต่างกันทั้งนี้เป็นผลเนื่องมาจากรูปแบบของสมการมีจุดตัด (intercept) ต่างกัน เพราะความน่าจะเป็นที่เกิดขึ้นก่อน (prior probabilities) จะมีผลต่อค่าจุดตัดแต่ไม่มีผลต่อค่าสัมประสิทธิ์ของตัวแปรต้น อย่างไรก็ตามจะเห็นว่าค่าจุดตัดต่างกันเพียงเล็กน้อยเนื่องจากค่า prior probabilities แต่ละกลุ่มในสมการทั้งสองรูปแบบใกล้เคียงกัน จึงส่งผลให้ความผิดพลาดในการจำแนกต่างกันเพียง 0.8 %

เมื่อพิจารณาความผิดพลาดในการจำแนกแต่ละกลุ่ม จะเห็นว่ากลุ่มฟาร์มที่มีปริมาณน้ำนมต่ำ (62 ฟาร์ม) มีค่าเท่ากับในสมการทั้ง 3 รูปแบบ คือ 43.5 % แต่กลุ่มฟาร์มที่มีปริมาณน้ำนมสูง (58 ฟาร์ม) ซึ่งมีจำนวนสมาชิกในกลุ่มน้อยกว่า สมการของ DA with equal priors จำแนกกลุ่มผิดพลาดต่ำกว่า สมการของ DA with proportional priors และ BLR ซึ่งสอดคล้องกับ Hossain et al. (2002) ที่ว่าในกลุ่มที่มีจำนวนสมาชิกน้อยกว่า สมการ DA with equal priors จะมีความแม่นยำมากกว่า

สรุป

การศึกษาปัจจัยที่มีผลต่อปริมาณผลผลิตน้ำนมของการเลี้ยงโคนมในฟาร์มเกษตรกรรายย่อยในเขตพื้นที่ภาคกลางตอนล่างจังหวัดเพชรบุรีและประจวบคีรีขันธ์ โดยใช้วิธีการจำแนกกลุ่มฟาร์มตามปริมาณน้ำนม ด้วยการวิเคราะห์ถดถอยแบบโลจิสติก (BLR) และการวิเคราะห์จำแนกประเภท (DA with equal priors และ DA with proportional priors) โดยใช้ข้อมูลจากการสำรวจจากฟาร์มเกษตรกรรายย่อยที่เลี้ยงโคนม โดยแบ่งฟาร์มออกเป็น 2 กลุ่มคือ กลุ่ม

ฟาร์มที่มีปริมาณน้ำนมต่ำและกลุ่มฟาร์มที่มีปริมาณน้ำนมสูง สรุปได้ว่าจากการวิเคราะห์จำแนกประเภท (DA) ตัวแปรที่มีอิทธิพลสูงสุดในการจำแนกฟาร์มว่าควรจัดอยู่ในกลุ่มฟาร์มที่ให้ปริมาณน้ำนมสูงหรือต่ำเรียงตามลำดับคือ อายุของแม่โค ระดับเลือด HF ปีเกิดของแม่โค เบอร์เซ็นต์อาหารชั้น และแหล่งฟาร์ม และจากการเปรียบเทียบผลการจำแนกกลุ่มฟาร์มตามปริมาณน้ำนมของสมการทั้ง 3 รูปแบบดังกล่าว สรุปว่า สมการทั้ง 3 รูปแบบมีประสิทธิภาพในการจำแนกใกล้เคียงกัน โดยสมการ DA with equal priors มีความถูกต้องในการจำแนก 65.0 % DA with proportional priors เท่ากับ 64.2 % ซึ่งเท่ากับสมการ BLR แสดงให้เห็นว่าสามารถเลือกใช้สมการได้ทั้ง 3 รูปแบบในการจำแนกกลุ่มฟาร์มตามปริมาณน้ำนมสำหรับข้อมูลชุดนี้

กิตติกรรมประกาศ

คณะผู้วิจัยขอขอบคุณคณะสัตวศาสตร์และเทคโนโลยีการเกษตร มหาวิทยาลัยศิลปากรที่สนับสนุนงบประมาณในการทำวิจัย ขอขอบคุณองค์การส่งเสริมกิจการโคนมแห่งประเทศไทย จังหวัดประจวบคีรีขันธ์ และสหกรณ์โคนม ชะอำห้วยทราย และศูนย์รับนมในเขตพื้นที่ตำบลสามพระยา อำเภอชะอำ จังหวัดเพชรบุรี ที่เอื้อเฟื้อในการให้ข้อมูลปริมาณน้ำนม และขอขอบคุณเกษตรกรฟาร์มโคนมที่กรุณาสละเวลาในการให้ข้อมูลเพื่อประกอบการศึกษาวิจัยในครั้งนี้

เอกสารอ้างอิง

- อภิญา อิงอาจ และ พูลพงศ์ สุขสว่าง. (2554). การวิเคราะห์สถิติขั้นสูง. เอกสารประกอบการอบรมเชิงปฏิบัติการ. วิทยาลัยวิทยาการวิจัยและวิทยาการทางปัญญา มหาวิทยาลัยบูรพา, ชลบุรี. 52 หน้า.
- Bull, S.B. and Donner, A. (1987). The efficiency of multinomial logistic regression compared

- with multiple group discriminant analysis. *J Am Stat Assoc.* 82: 1118-1122.
- Efron, B. (1975). The efficiency of logistic regression compared to normal discriminant analysis. *J Am Stat Assoc.* 70: 892-898.
- Halperin, M., Blackwelder, W.E. and Verter, J.I. (1971). Estimation of the multivariate logistic risk function: A comparison of the discriminant function and maximum likelihood approaches. *J Chron Dis.* 24: 125-158.
- Hossian, M., Wright, S. and Petersen, L.A. (2002). Comparing performance of multinomial logistic regression and discriminant analysis for monitoring access to care for acute myocardial infarction. *Journal of Clinical Epidemiology* 55: 400-406.
- McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*. 2nd ed. Monographs on Statistics and Applied Probability. London: Chapman and Hall. 511 pp.
- Montgomery, M.E., White, M.E. and Martin, S.W. (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.
- Press, S.J. and Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. *J Am Stat Assoc.* 73: 699-705.

